JMIR Formative Research

Impact Factor (2023): 2.0
Volume 9 (2025) ISSN 2561-326X Editor in Chief: Amaryllis Mavragani, PhDc

Contents

Original Papers

Evaluating Feasibility and Acceptability of the “My HeartHELP” Mobile App for Promoting Heart-Healthy Lifestyle Behaviors: Mixed Methods Study (e66108)
Jina Choo, Songwhi Noh, Yura Shin. 28

School-Based Virtual Reality Programming for Obtaining Moderate-Intensity Exercise Among Children With Disabilities: Pre-Post Feasibility Study (e65801)
Byron Lai, Ashley Wright, Bailey Hutchinson, Larsen Bright, Raven Young, Drew Davis, Sultan Malik, James Rimmer, Pelham High Community Engagement Group. 44

Discussions of Antibiotic Resistance on Social Media Platforms: Text Mining and Mixed Methods Content Analysis Study (e37160)
Jocelyne Arquembourg, Philippe Glaser, France Roblot, Isabelle Metzler, Mélanie Gallant-Dewavrin, Hugues Nanguem, Adel Mebarki, Paméla Voillot, Stéphane Schück. 53

The Motivations of Citizens to Attend an eHealth Course in the Public Library: Qualitative Interview Study (e60612)
Lucille Standaar, Adriana Israel, Rosalie van der Vaart, Brigitta Keij, Frank van Lenthe, Roland Friele, Mariëlle Beenackers, Lilian van Tuyl. . . 6 3

Adolescent Emoji Use in Text-Based Messaging: Focus Group Study (e59640)
Matt Minich, Bradley Kerr, Megan Moreno. 76

Impact of the Kidney Score Platform on Communication About and Patients’ Engagement With Chronic Kidney Disease Health: Pre–Post Intervention Study (e56855)
Delphine Tuot, Susan Crowley, Lois Katz, Joseph Leung, Delly Alcantara-Cadillo, Christopher Ruser, Elizabeth Talbot-Montgomery, Joseph Vassalotti. 86

Effect of a Digital Health Exercise Program on the Intention for Spinal Surgery in Adult Spinal Deformity: Exploratory Cross-Sectional Survey (e66889)
Marsalis Brown, Christopher Lin, Christopher Jin, Matthew Rohde, Brett Rocos, Jonathan Belding, Barrett Woods, Stacey Ackerman. 95

Assessing the Utilization of Electronic Consultations in Genetics: Seven-Year Retrospective Study (e63028)
Sawona Biswas, Joyce So, Robert Wallerstein, Ralph Gonzales, Delphine Tout, Lisa DeAngelis, Aleksandar Rajkovic. 103

Predicting Transvaginal Surgical Mesh Exposure Outcomes Using an Integrated Dataset of Blood Cytokine Levels and Medical Record Data: Machine Learning Approach (e59631)
Mihyun Waugh, Tyler Mills, Nicholas Boltin, Lauren Wolf, Patti Parker, Ronnie Horner, Thomas Wheeler II, Richard Goodwin, Melissa Moss. . . 1 1 0

XSL•FO
RenderX

Integration of a Mental Health App (e-MICHI) Into a Blended Treatment of Depression in Adolescents: Single-Group, Naturalistic Feasibility Trial (e58427)
Lena Lincke, Tim Martin-Döring, Andrea Daunke, Antonia Sadkowiak, Daria Nolkemper, Nina Sproeber-Kolb, Stefanie Bienioschek, Olaf Reis, Michael Kölch. . . . 121

Identification of Major Bleeding Events in Postoperative Patients With Malignant Tumors in Chinese Electronic Medical Records: Algorithm Development and Validation (e66189)
Hui Li, Haiyang Yao, Yuxiang Gao, Hang Luo, Changbin Cai, Zhou Zhou, Muhan Yuan, Wei Jiang. . . . 137

Mobile Health App for Adolescent Asthma Self-Management: Development and Usability Study of the Pulmonary Education and Knowledge Mobile Asthma Action Plan (e64212)
Xing He, Jiang Bian, Ariel Berlinski, Yi Guo, A Simmons, S Marshall, Carolyn Greene, Rita Brown, Jessica Turner, Tamara Perry. . . . 147

Tailoring Rheumatoid Arthritis Visit Timing Based on mHealth App Data: Mixed Methods Assessment of Implementation and Usability (e60854)
Robert Rudin, Leah Santacroce, Ishani Ganguli, Daniel Solomon. . . . 161

Developing an Online Community Advisory Board (CAB) of Parents From Social Media to Co-Design an Human Papillomavirus Vaccine Intervention: Participatory Research Study (e65986)
Regan Murray, Shawn Chiang, Ann Klassen, Jennifer Manganello, Amy Leader, Wen-Juo Lo, Philip Massey. . . . 174

Machine Learning Approach to Identifying Empathy Using the Vocals of Mental Health Helpline Counselors: Algorithm Development and Validation (e67835)
Ruvini Sanjeewa, Ravi Iyer, Pragalathan Apputhurai, Nilmini Wickramasinghe, Denny Meyer. . . . 188

Acoustic and Natural Language Markers for Bipolar Disorder: A Pilot, mHealth Cross-Sectional Study (e65555)
Cristina Crocamo, Riccardo Cioni, Aurelia Canestro, Christian Nasti, Dario Palpella, Susanna Piacenti, Alessandra Bartoccetti, Martina Re, Valentina Simonetti, Chiara Barattieri di San Pietro, Maria Bulgheroni, Francesco Bartoli, Giuseppe Carrà. . . . 203

Using the Healthy Native Youth Implementation Toolbox to Provide Web-Based Adolescent Health Promotion Decision Support to American Indian and Alaska Native Communities: Implementation Study (e67885)
Amrita Sidhu, Ross Shegog, Stephanie Craig-Rushing, Nicole Trevino, Michelle Singer, Cornelia Jessen, Gwenda Gorman, Sean Simpson, Melissa Peskin, Belinda Hernandez, Christine Markham. . . . 222

Evaluation of a Curriculum-Based Nutrition Education Intervention Protocol in Elementary Schools: Nonrandomized Feasibility Study (e69242)
Jacqueline Brown, Nicholas Rita, Beatriz Franco-Arellano, Ann LeSage, Joanne Arcand. . . . 234

Evaluating the Acceptability of a Brief Web-Based Alcohol Misuse Prevention Program Among US Military Cadets: Mixed Methods Formative Evaluation (e67637)
Emily Schmied, Lauren Hurtado, W Robinson, Cynthia Simon-Arndt, Richard Moyer III, Leslie Wilson, Mark Reed, Shannon Blakey, Marni Kan. . . . 249

Impact of a Virtual Reality Video ("A Walk-Through Dementia") on YouTube Users: Topic Modeling Analysis (e67755)
Xiaoli Li, Xiaoyu Liu, Cheng Yin, Sandra Collins, Eman Alanazi. . . . 260

Feasibility, Acceptability, and Effectiveness of a Smartphone App to Increase Pretransplant Vaccine Rates: Usability Study (e68855)
Amy Feldman, Brenda Beaty, Susan Moore, Sheana Bull, Kumanan Wilson, Katherine Atkinson, Cameron Bell, Kathryn Denize, Allison Kempe.
2 7 4

XSL•FO
RenderX

Development of an eHealth Mindfulness-Based Music Therapy Intervention for Adults Undergoing Allogeneic Hematopoietic Stem Cell Transplantation: Qualitative Study (e65188)
Sara Fleszar-Pavlovic, Blanca Noriega Esquives, Padideh Lovan, Arianna Brito, Ann Sia, Mary Kauffman, Maria Lopes, Patricia Moreno, Tulay Koru-Sengul, Rui Gong, Trent Wang, Eric Wieder, Maria Rueda-Lara, Michael Antoni, Krishna Komanduri, Teresa Lesiuk, Frank Penedo. . . . 2 8 3

Evaluating User Experience and Satisfaction in a Concussion Rehabilitation App: Usability Study (e67275)
Michael Hutchison, Alex Di Battista, Kyla Pyndiura. . . . 296

Oxidative Stress Markers and Prediction of Severity With a Machine Learning Approach in Hospitalized Patients With COVID-19 and Severe Lung Disease: Observational, Retrospective, Single-Center Feasibility Study (e66509)
Olivier Raspado, Michel Brack, Olivier Brack, Mélanie Vivancos, Aurélie Esparcieux, Emmanuelle Cart-Tanneur, Abdellah Aouifi. . . . 306

Association of Social Media Recruitment and Depression Among Racially and Ethnically Diverse Metabolic and Bariatric Surgery Candidates: Prospective Cohort Study (e58916)
Jackson Francis, Sitapriya Neti, Dhatri Polavarapu, Folefac Atem, Luyu Xie, Olivia Kapera, Matthew Mathew, Elisa Marroquin, Carrie McAdams, Jeffrey Schellinger, Sophia Ngenge, Sachin Kukreja, Benjamin Schneider, Jaime Almandoz, Sarah Messiah. . . . 317

Encouraging the Voluntary Mobilization of Mental Resources by Manipulating Task Design: Explorative Study (e63491)
Lina-Estelle Louis, Saïd Moussaoui, Sébastien Ravoux, Isabelle Milleville-Pennel. . . . 327

Support of Home-Based Structured Walking Training and Prediction of the 6-Minute Walk Test Distance in Patients With Peripheral Arterial Disease Based on Telehealth Data: Prospective Cohort Study (e65721)
Fabian Wiesmüller, Andreas Prenner, Andreas Ziegl, Gihan El-Moazen, Robert Modre-Osprian, Martin Baumgartner, Marianne Brodmann, Gerald Seinost, Günther Silbernagel, Günter Schreier, Dieter Hayn. . . . 339

Web-Based Explainable Machine Learning-Based Drug Surveillance for Predicting Sunitinib- and Sorafenib-Associated Thyroid Dysfunction: Model Development and Validation Study (e67767)
Fan-Ying Chan, Yi-En Ku, Wen-Nung Lie, Hsiang-Yin Chen. . . . 352

Assessment of Gait Parameters Using Wearable Sensors and Their Association With Muscle Mass, Strength, and Physical Performance in Korean Older Adults: Cross-Sectional Study (e63928)
Jinyoung Shin, Hyuk Kweon, Jaekyung Choi. . . . 365

Centering Youth Voice in the Adaptation of an mHealth Intervention for Young Adults With HIV in South Texas, United States: Human-Centered Design Approach (e60531)
Nhat Ho, Catherine Johnson, Autumn Chidester, Ruby Viera Corral, Jacundo Ramos, Miguel Garcia, Rishi Gonuguntla, Cyrena Cote, Divya Chandramohan, Hueylie Lin, Anna Taranova, Ank Nijhawan, Susan Kools, Karen Ingersoll, Rebecca Dillingham, Barbara Taylor. . . . 374

Chatbot to Support the Mental Health Needs of Pregnant and Postpartum Women (Moment for Parents): Design and Pilot Study (e72469)
Kelsey McAlister, Lara Baez, Jennifer Huberty, Marianna Kerppola. . . . 386

Anticipated Acceptability of Blended Learning Among Lay Health Care Workers in Malawi: Qualitative Analysis Guided by the Technology Acceptance Model (e62741)
Tiwonge Mbeya-Munkhondya, Caroline Meek, Mtisunge Mphande, Tapiwa Tembo, Mike Chitani, Milenka Jean-Baptiste, Caroline Kumbuyo, Dhrutika Vansia, Katherine Simon, Sarah Rutstein, Victor Mwapasa, Vivian Go, Maria Kim, Nora Rosenberg. . . . 403

The Impact of a Human Papillomavirus Facebook-Based Intervention (#HPVVaxTalks) Among Young Black (African American and Sub-Saharan African Immigrants) Adults: Pilot Pre- and Poststudy (e69609)
Adebola Adegboyega, Amanda Wiggins, Abubakari Wuni, Melinda Ickes. . . . 414

Medical Device Based on a Virtual Reality–Based Upper Limb Rehabilitation Software: Usability Evaluation Through Cognitive Walkthrough (e68149)
Seojin Hong, Hyun Choi, Hyosun Kweon. . . . 427

XSL•FO
RenderX

Development of a Comprehensive Decision Support Tool for Chemotherapy-Cycle Prescribing: Initial Usability Study (e62749)
Sanna Iivanainen, Reetta Arokoski, Santeri Mentu, Laura Lang, Jussi Ekström, Henri Virtanen, Vesa Kataja, Jussi Koivunen. 438

Barriers and Determinants of Referral Adherence in AI-Enabled Diabetic Retinopathy Screening for Older Adults in Northern India During the COVID-19 Pandemic: Mixed Methods Pilot Study (e67047)
Anshul Chauhan, Anju Goyal, Ritika Masih, Gagandeep Kaur, Lakshay Kumar, Neha, Harsh Rastogi, Sonam Kumar, Bidhi Singh, Preeti Syal, Vishali Gupta, Luke Vale, Mona Duggal. 447

Personalized mHealth Intervention (StepAdd) for Increasing Physical Activity in Japanese Patients With Type 2 Diabetes: Secondary Analysis of Social Cognitive Theory Measurements of a Single-Arm Pilot Study (e60221)
Kayo Waki, Syunpei Enomoto, Toshimasa Yamauchi, Masaomi Nangaku, Kazuhiko Ohe. 462

Accuracy of Smartphone-Mediated Snore Detection in a Simulated Real-World Setting: Algorithm Development and Validation (e67861)
Jeffrey Brown, Zachary Mitchell, Yu Jiang, Ryan Archdeacon. 470

Mobile Health Tool to Capture Social Determinants of Health and Their Impact on HIV Treatment Outcomes Among People Who Use Drugs: Pilot Feasibility Study (e59953)
Rachel Gicquelais, Caitlin Conway, Olivia Vjorn, Andrew Genz, Gregory Kirk, Ryan Westergaard. 479

Mobile App-Based Interactive Care Plan for Migraine: Survey Study of Usability and Improvement Opportunities (e66763)
Nathan Young, Jennifer Stern, Stephanie Steel, Jon Ebbert. 490

Changes in Health Education Literacy After Structured Web-Based Education Versus Self-Directed Online Information Seeking in Patients Undergoing Carpal Tunnel Release Surgery: Nonrandomized, Controlled Study (e65114)
Mariella Seel, Julian Mihalic, Stefan Froschauer, Bernhard Holzner, Jens Meier, Tobias Gotterbarm, Matthias Holzbauer. 501

The Impact of ChatGPT Exposure on User Interactions With a Motivational Interviewing Chatbot: Quasi-Experimental Study (e56973)
Jiading Zhu, Alec Dong, Cindy Wang, Scott Veldhuizen, Mohamed Abdelwahab, Andrew Brown, Peter Selby, Jonathan Rose. 519

A Brief Cognitive Behavioral Therapy–Based Digital Intervention for Reducing Hazardous Alcohol Use in South Korea: Development and Prospective Pilot Study (e64459)
Manjae Kwon, Daa Moon, Minjae Kang, Young-Chul Jung. 530

Exploring the Use of an Augmented Reality Device Learning Tool for Multidisciplinary Staff Training on Domestic Abuse and Sexual Violence: Postintervention Qualitative Evaluation (e60075)
Dilroshini Karunaratne, Jessica Whittock, Amber Moore, Krishna Dasigan, Jasmine Chevolleau, Brent Bartholomew, Nikki Kelly, Charlotte Cohen. 547

Optimizing Testimonials for Behavior Change in a Digital Intervention for Binge Eating: Human-Centered Design Study (e59691)
Isabel Rooper, Adrian Ortega, Thomas Massion, Tanvi Lakhtakia, Macarena Kruger, Leah Parsons, Lindsay Lipman, Chidiebere Azubuike, Emily Tack, Katrina Obleada, Andrea Graham. 562

The Interactive Care Coordination and Navigation mHealth Intervention for People Experiencing Homelessness: Cost Analysis, Exploratory Financial Cost-Benefit Analysis, and Budget Impact Analysis (e64973)
Hannah McCullough, Leticia Moczygemba, Anton Avanceña, James Baffoe. 574

XSL•FO
RenderX

Digital Health Literacy and Attitudes Toward eHealth Technologies Among Patients With Cardiovascular Disease and Their Implications for Secondary Prevention: Survey Study (e63057)
Greta Ullrich, Alexander Bäuerle, Hannah Vogt, Amir Mahabadi, Katrin Paldán, Daniel Messiha, Lisa Jahre, Christos Rammos, Tienush Rassaf, Julia Lortz. . . . 586

Interprofessional Discussion for Knowledge Transfer in a Digital “Community of Practice” for Managing Pneumoconiosis: Mixed Methods Study (e67999)
Varinn Sood, Heidi Rishel Brakey, Orrin Myers, Xin Shore, Akshay Sood. . . . 598

Telemedicine Prescribing by US Mental Health Care Providers: National Cross-Sectional Survey (e63251)
Mollie Cummins, Julia Ivanova, Hiral Soni, Zoe Robbins, Brian Bunnell, Esteban López, Brandon Welch. . . . 611

Supplementing Consent for a Prospective Longitudinal Cohort Study of Infants With Antenatal Opioid Exposure: Development and Assessment of a Digital Tool (e59954)
Jamie Newman, Leslie Clarke, Pranav Athimuthu, Megan Dhawan, Sharon Owen, Traci Beiersdorfer, Lindsay Parlberg, Ananta Bangdiwala, Taya McMillan, Sara DeMauro, Scott Lorch, Myriam Peralta-Carcelen, Deanne Wilson-Costello, Namasivayam Ambalavanan, Stephanie Merhar, Brenda Poindexter, Catherine Limperopoulos, Jonathan Davis, Michele Walsh, Carla Bann. . . . 624

A Digital Pornography Education Prototype Co-Designed With Young People: Formative Evaluation (e65859)
Jake Turvey, Michelle Raggatt, Cassandra Wright, Angela Davis, Meredith Temple-Smith, Megan Lim. . . . 635

Purposes of Smartphone Health Care Apps and the Practicality of Their Functions in Disaster Situations: Qualitative Function Assessment Study (e56862)
Sayuri Nonaka, Susumu Fujii, Kosuke Yamada. . . . 646

Assessment of the Sensitivity of a Smartphone App to Assist Patients in the Identification of Stroke and Myocardial Infarction: Cross-Sectional Study (e60465)
Amar Dhand, Rama Mangipudi, Anubodh Varshney, Jonathan Crowe, Andria Ford, Nancy Sweitzer, Min Shin, Samuel Tate, Haissam Haddad, Michael Kelly, James Muller, Jay Shavadia. . . . 660

Text Messaging Interventions for Unhealthy Alcohol Use in Emergency Departments: Mixed Methods Assessment of Implementation Barriers and Facilitators (e65187)
Megan O'Grady, Laura Harrison, Adekemi Suleiman, Morica Hutchison, Nancy Kwon, Frederick Muench, Sandeep Kapoor. . . . 677

Patients’ Experience With Evaluation by Both a Musculoskeletal Physician and Physical Therapist in the Same Digital Visit: Survey Study (e66744)
Mary O'Connor, Carolyn Chudy, Kaitlyn Peters, Megan Ribaudo, Carrie McCulloch, Jared Aguilar, Trista Taylor, Ryan Grant. . . . 691

App- and Wearable-Based Remote Monitoring for Patients With Myasthenia Gravis and Its Specialists: Feasibility and Usability Study (e58266)
Maike Stein, Regina Stegherr, Pushpa Narayanaswami, David Legg, Meret Herdick, Andreas Meisel, Lea Gerischer, Sophie Lehnerer. . . . 701

Enhancing Distress Tolerance Skills in Adolescents With Anorexia Nervosa Through the BALANCE Mobile App: Feasibility and Acceptability Study (e70278)
Christina Miranda, Brittany Matheson, Nandini Datta, Aileen Whyte, Hyun-Joon Yang, Paul Schmiedmayer, Vishnu Ravi, Oliver Aalami, James Lock. . . . 717

Probing Public Perceptions of Antidepressants on Social Media: Mixed Methods Study (e62680)
Jianfeng Zhu, Xinyu Zhang, Ruoming Jin, Hailong Jiang, Deric Kenne. . . . 724

Novel Evaluation Metric and Quantified Performance of ChatGPT-4 Patient Management Simulations for Early Clinical Education: Experimental Study (e66478)
Riley Scherr, Aidin Spina, Allen Dao, Saman Andalib, Faris Halaseh, Sarah Blair, Warren Wiechmann, Ronald Rivera. . . . 736

Capturing Everyday Parental Feeding Practices and Eating Behaviors of 3- to 5-Year-Old Children With Avid Eating Behavior: Ecological Momentary Assessment Feasibility and Acceptability Study (e66807)
Abigail Pickard, Katie Edwards, Claire Farrow, Emma Haycraft, Jacqueline Blissett. . . . 747

XSL•FO
RenderX

Mobilizing Health Promotion Through Canada's Student Mental Health Network: Concurrent, Mixed Methods Process Evaluation (e58992)
Amy Ecclestone, Brooke Linden, Jessica Rose, Kiran Kullar. . . . 759

Evaluation of Financial Support Workshops for Patients Under State Pension Age With Degenerative Cervical Myelopathy: Survey Study (e59032)
Tanzil Rujeedawa, Zahabiya Karimi, Helen Wood, Irina Sangeorzan, Roy Smith, Iwan Sadler, Esther Martin-Moore, Adrian Gardner, Andreas K Demetriades, Rohitashwa Sinha, Gordan Grahovac, Antony Bateman, Naomi Deakin, Benjamin Davies. . . . 770

Snowball Group Usability Testing for Rapid and Iterative Multisite Tool Development: Method Development Study (e55316)
Katherine Dauber-Decker, David Feldstein, Rachel Hess, Devin Mann, Eun Kim, Pranisha Gautam-Goyal, Jeffrey Solomon, Sundas Khan, Fatima Malik, Lynn Xu, Ainsley Huffman, Paul Smith, Wendy Halm, Alice Yuroff, Safiya Richardson. . . . 777

Exploring the Effect of an 8-Week AI-Composed Exercise Program on Pain Intensity and Well-Being in Patients With Spinal Pain: Retrospective Cohort Analysis (e57826)
Annika Griefahn, Florian Avermann, Christoff Zalpour, Robert Marshall, Inés Cordon Morillas, Kerstin Luedtke. . . . 786

Involving End Users in the Development and Usability Testing of a Smartphone App Designed for Individuals With Prediabetes: Mixed-Methods Focus Group Study (e59386)
Natalie Grieve, Kyra Braaten, Megan MacPherson, Sam Liu, Mary Jung. . . . 799

Activity Trackers in Physical Therapy for People With Chronic Obstructive Pulmonary Disease in the Netherlands: Cross-Sectional Study on Current Use and Implementation Determinants (e59533)
Darcy Ummels, Esther Bols, Roel Frantzen, Tim Frantzen, Levy Robeerts, Emmylou Beekman. . . . 809

Understanding Providers' Attitude Toward AI in India's Informal Health Care Sector: Survey Study (e54156)
Sumeet Kumar, Snehil Rayal, Raghuram Bommaraju, Navya Varasala, Sirisha Papineni, Sarang Deo. . . . 823

Developing an ICD-10 Coding Assistant: Pilot Study Using RoBERTa and GPT-4 for Term Extraction and Description-Based Code Selection (e60095)
Sander Puts, Catharina Zegers, Andre Dekker, Iñigo Bermejo. . . . 834

Web-Based, Algorithm-Guided Insulin Titration in Insulin-Treated Type 2 Diabetes: Pre-Post Intervention Study (e68914)
Nishanth Thiagarajan, Hong Tan, Suresh Rama Chandran, Phong Lee, Yun Chin, Wanling Zeng, Emily Ho, David Carmody, Su-Yen Goh, Yong Bee. . . . 846

Assessment of Digital Capabilities by 9 Countries in the Alliance for Healthy Cities Using AI: Cross-Sectional Analysis (e62935)
Hocheol Lee. . . . 856

A Web-Based Resource Informed by Cognitive Behavioral Therapy and Positive Psychology to Address Stress, Negative Affect, and Problematic Alcohol Use: A Usability and Descriptive Study (e63819)
Ingrid Serck-Hanssen, Marit Solheim-Witt, Justin Anker, Dawn Sugarman. . . . 869

Step Count Accuracy of the Life Plus Connected Watch at Different Localizations and Speeds in Healthy Adults, Patients With Cardiovascular Disease, and Patients With Peripheral Artery Disease: Step Count Validation Study in Laboratory Settings (e58964)
Anne-Noëlle Heizmann, Edouard Ollier, Pierre Labeix, Ivan Goujon, Frédéric Roche, Claire Le Hello. . . . 881

Mediating Effect of Social Capital on the Association Between Digital Literacy and Life Satisfaction Among Older Adults in South Korea: Cross-Sectional Study (e68163)
Hyein Jung, Hocheol Lee, Eun Nam. . . . 890

XSL•FO
RenderX

Addressing the “Black Hole” of Low Back Pain Care With Clinical Decision Support: User-Centered Design and Initial Usability Study (e66666)
Robert Rudin, Patricia Herman, Robert Vining. . . . 902

Summer Research Internship Curriculum to Promote Self-Efficacy, Researcher Identity, and Peer-to-Peer Learning: Retrospective Cohort Study (e54167)
Yulia Levites Strekalova, Rachel Liu-Galvin, Samuel Border, Sara Midence, Mishal Khan, Maya VanZanten, John Tomaszewski, Sanjay Jain, Pinaki Sarder. . . . 915

Web-Based Exercise and Nutrition Intervention to Improve Leg Muscle Strength and Physical Functioning in Older Adults: Pre-Post Pilot Study (e54392)
Berber Dorhout, Nick Wezenbeek, Lisette de Groot, Pol Grootswagers. . . . 926

Identifying Digital Markers of Attention-Deficit/Hyperactivity Disorder (ADHD) in a Remote Monitoring Setting: Prospective Observational Study (e54531)
Heet Sankesara, Hayley Denyer, Shaoxiong Sun, Qigang Deng, Yatharth Ranjan, Pauline Conde, Zulqarnain Rashid, Philip Asherson, Andrea Bilbow, Madeleine Groom, Chris Hollis, Richard Dobson, Amos Folarin, Jonna Kuntsi. . . . 938

The Development of an Instagram Reel-Based Bystander Intervention Message Among College Students: Formative Survey and Mixed Methods Pilot Study (e66769)
Leticia Couto. . . . 952

Evaluating Older Adults’ Engagement and Usability With AI-Driven Interventions: Randomized Pilot Study (e64763)
Marcia Shade, Changmin Yan, Valerie Jones, Julie Boron. . . . 970

Characterizing Telehealth Barriers and Preferences to Promote Acceptable Implementation Strategies in Central Uganda: Multilevel Formative Evaluation (e60843)
Michael Kizito, Erina Mugabi, Sabrina Ford, Bree Holtz, Kelly Hirko. . . . 982

Testing an Electronic Patient-Reported Outcome Platform in the Context of Traumatic Brain Injury: PRiORiTy Usability Study (e58128)
Christel McMullan, Grace Turner, Ameeta Retzer, Antonio Belli, Elin Davies, Laura Nice, Luke Flavell, Jackie Flavell, Melanie Calvert. . . . 995

The Association of Psychological Factors With Willingness to Share Health-Related Data From Technological Devices: Cross-Sectional Questionnaire Study (e64244)
Marijn Eversdijk, Emma Douma, Mirela Habibovic, Willem Kop. . . . 1008

AI Machine Learning–Based Diabetes Prediction in Older Adults in South Korea: Cross-Sectional Analysis (e57874)
Hocheol Lee, Myung-Bae Park, Young-Joo Won. . . . 1018

Insights From Diverse Perspectives on Social Media Messages to Inform Young Adults With Cancer About Clinical Trials: Focus Group Study (e64265)
Melissa Beauchemin, Desiree Walker, Allison Rosen, Maria Frazer, Meital Eisenberger, Rhea Khurana, Edward Bentlyewski, Victoria Fedorko, Corey Basch, Grace Hillyer. . . . 1027

Sentiments of Individuals with Interstitial Cystitis/Bladder Pain Syndrome Toward Pentosan Polysulfate Sodium: Infodemiology Study (e54209)
Yulin Hswen, Qiuyuan Qin, Pressley Smith, Alison Swierczynski, Stuart Bauer, Erika Ladson, Amanda Garrett, Catherine Brownstein. . . . 1036

Using Health Information Resources for People With Cognitive Impairment (digiDEM Bayern): Registry-Based Cohort Study (e54460)
Florian Weidinger, Nikolas Dietzel, Elmar Graessel, Hans-Ulrich Prokosch, Peter Kolominsky-Rabas. . . . 1046

XSL•FO
RenderX

Mobile Phone App to Promote Lifestyle Change in People at Risk of Type 2 Diabetes: Feasibility 3-Arm Randomized Controlled Trial (e63737)
Gyri Skoglund, Gunvor Hilde, Pernille Lunde, Venessa Cruz Naceno, Cecilie Fromholt Olsen, Birgitta Blakstad Nilsson. 1056

Considering Comorbidities and Individual Differences in Testing a Gaming Behavioral Activation App for Perinatal Depression and Anxiety: Open Trial Pilot Intervention Study (e59154)
Gabriella Hamlett, Chloe Schrader, Craig Ferguson, Lauren Kobylski, Rosalind Picard, Joseph Locascio, Richard McNally, Lee Cohen, Rachel Vanderkruik. 1068

Verification of the Reliability of an Automated Urine Test Strip Colorimetric Program Using Colorimetric Analysis: Survey Study (e62772)
Keigo Inagaki, Daisuke Tsuriya, Takuya Hashimoto, Katsumasa Nakamura. 1079

Determinants of Dropping Out of Remote Patient-Reported Outcome–Based Follow-Up Among Patients With Epilepsy: Prospective Cohort Study (e58258)
Sofie Vestergaard, Mette Roost, David Christiansen, Liv Schougaard. 1089

Impact of a Virtual Care Navigation Service on Member-Reported Outcomes Among Lesbian, Gay, Bisexual, Transgender, and Queer Populations: Case Study (e64137)
Seul Choi, Jaclyn Marshall, Patrina Sexton Topper, Andrew Pregnall, José Bauermeister. 1103

A Mobile App for Promoting Breastfeeding-Friendly Communities in Hong Kong: Design and Development Study (e64191)
Heidi Fan, Emily Leung, Ka Lau, Janet Wong, Edmond Choi, Christine Lam, Marie Tarrant, Hextan Ngan, Patrick Ip, Chia Lin, Kris Lok. 1115

Predicting Age and Visual-Motor Integration Using Origami Photographs: Deep Learning Study (e58421)
Chien-Yu Huang, Yen-Ting Yu, Kuan-Lin Chen, Jenn-Jier Lien, Gong-Hong Lin, Ching-Lin Hsieh. 1130

Comparative Evaluation of Consumer Wearable Devices for Atrial Fibrillation Detection: Validation Study (e65139)
Femke Wouters, Henri Gruwez, Christophe Smeets, Anessa Pijalovic, Wouter Wilms, Julie Vranken, Zoë Pieters, Hugo Van Herendael, Dieter Nuyens, Maximo Rivero-Ayerza, Pieter Vandervoort, Peter Haemers, Laurent Pison. 1138

Real-Time Analytics and AI for Managing No-Show Appointments in Primary Health Care in the United Arab Emirates: Before-and-After Study (e64936)
Yousif AlSerkal, Naseem Ibrahim, Aisha Alsereidi, Mubaraka Ibrahim, Sudheer Kurakula, Sadaf Naqvi, Yasir Khan, Neema Oottumadathil. . .
1 1 5 1

Effects of Smart Goggles Used at Bedtime on Objectively Measured Sleep and Self-Reported Anxiety, Stress, and Relaxation: Pre-Post Pilot Study (e58461)
Sharon Danoff-Burg, Elie Gottlieb, Morgan Weaver, Kiara Carmon, Duvia Lara Ledesma, Holly Rus. 1159

The Use of Social Media on Enhancing Dental Care and Practice Among Dental Professionals: Cross-Sectional Survey Study (e66121)
Joseph Acosta, Palinee Detsomboonrat, Pagaporn Pisarnturakit, Nipaporn Urwannachotima. 1167

Artificial Intelligence–Powered Training Database for Clinical Thinking: App Development Study (e58426)
Heng Wang, Danni Zheng, Mengying Wang, Hong Ji, Jiangli Han, Yan Wang, Ning Shen, Jie Qiao. 1177

The Physical Activity at Work (PAW) Program in Thai Office Workers: Mixed Methods Process Evaluation Study (e57604)
Katika Akksilp, Thomas Rouyard, Wanrudee Isaranuwatchai, Ryota Nakamura, Falk Müller-Riemenschneider, Yot Teerawattananon, Cynthia Chen. 1188

XSL•FO
RenderX

ChatGPT's Attitude, Knowledge, and Clinical Application in Geriatrics Practice and Education: Exploratory Observational Study (e63494)
Huai Cheng. . . . 1205

A Digital Parenting Intervention With Intimate Partner Violence Prevention Content: Quantitative Pre-Post Pilot Study (e58611)
Moa Schafer, Jamie Lachman, Paula Zinser, Francisco Calderón Alfaro, Qing Han, Chiara Facciola, Lily Clements, Frances Gardner, Genevieve Haupt Ronnie, Ross Sheil. . . . 1220

Differences in eHealth Access, Use, and Perceived Benefit Between Different Socioeconomic Groups in the Dutch Context: Secondary Cross-Sectional Study (e49585)
Lucille Standaar, Lilian van Tuyl, Anita Suijkerbuijk, Anne Brabers, Roland Friele. . . . 1251

The Trifecta of Industry, Academic, and Health System Partnership to Improve Mental Health Care Through Smartphone-Based Remote Patient Monitoring: Development and Usability Study (e57624)
C Epperson, Rachel Davis, Allison Dempsey, Heinrich Haller, David Kupfer, Tiffany Love, Pamela Villarreal, Mark Matthews, Susan Moore, Kimberly Muller, Christopher Schneck, Jessica Scott, Richard Zane, Ellen Frank. . . . 1265

Co-Design of an Escape Room for e-Mental Health Training of Mental Health Care Professionals: Research Through Design Study (e58650)
Joyce Bierbooms, Wouter Sluis-Thiescheffer, Milou Feijt, Inge Bongers. . . . 1277

Measuring Bound Attention During Complex Liver Surgery Planning: Feasibility Study (e62740)
Tim Schneider, Timur Cetin, Stefan Uppenkamp, Dirk Weyhe, Thomas Muender, Anke Reinschluessel, Daniela Salzmann, Verena Uslar. . . .
1 2 9 4

Development of Personas and Journey Maps for Artificial Intelligence Agents Supporting the Use of Health Big Data: Human-Centered Design Approach (e67272)
Yoon Lee, Hanna Choi, Soo-Kyoung Lee. . . . 1311

Effect of a Narrative-Based Online Course Aimed at Reducing Stigma Toward Transgender Children and Adolescents: Longitudinal Observational Study (e59605)
Merlin Greuel, Van Nguyen, Doron Amsalem, Maya Adam, Till Bärnighausen. . . . 1329

User Outcomes for an App-Delivered Hypnosis Intervention for Menopausal Hot Flashes: Retrospective Analysis (e63948)
Katherine Scheffrahn, Claire Hall, Vanessa Muñiz, Gary Elkins. . . . 1343

Development and Beta Validation of an mHealth-Based Hearing Screener (SRESHT) for Young Children in Resource-Limited Countries: Pilot Validation Study (e53460)
Vidya Ramkumar, Deepashree Joshi B, Anil Prabhakar, James Hall, Ramya Vaidyanath. . . . 1357

Healthy Kai (Food) Checker Web-Based Tool to Support Healthy Food Policy Implementation: Development and Usability Study (e60447)
Magda Rosin, Cliona Ni Mhurchu, Elaine Umali, Sally Mackay. . . . 1373

Exploring the General Acceptability and User Experience of a Digital Therapeutic for Cognitive Training in a Singaporean Older Adult Population: Qualitative Study (e63568)
Siong Kwek, Qiao Leong, V Lee, Ni Lau, Smrithi Vijayakumar, Wei Ng, Bina Rai, Marlena Raczkowska, Christopher Asplund, Alexandria Remus, Dean Ho. . . . 1388

A Proposed mHealth Intervention to Address Patient Barriers to Colposcopy Attendance: Qualitative Interview Study of Clinic Staff and Patient Perspectives (e55043)
Jennifer Hemler, Rachel Wagner, Brittany Sullivan, Myneka Macenat, Erin Tagai, Jazmarie Vega, Enrique Hernandez, Suzanne Miller, Kuang-Yi Wen, Charletta Ayers, Mark Einstein, Shawna Hudson, Racquel Kohler. . . . 1400

XSL•FO
RenderX

Acceptance, Safety, and Effect Sizes in Online Dialectical Behavior Therapy for Borderline Personality Disorder: Interventional Pilot Study (e66181)
Ruben Vonderlin, Tali Boritz, Carola Claus, Büsra Senyüz, Saskia Mahalingam, Rachel Tennenhouse, Stefanie Lis, Christian Schmahl, Jürgen Margraf, Tobias Teismann, Nikolaus Kleindienst, Shelley McMain, Martin Bohus. 1414

Cardiomeds, an mHealth App for Self-Management to Support Swiss Patients With Heart Failure: 2-Stage Mixed Methods Usability Study (e63941)
Lisa Simioni, Elena Tessitore, Hamdi Hagberg, Aurélie Schneider-Paccot, Katherine Blondon, Liliane Gschwind, Philippe Meyer, Frederic Ehrler. 1431

Teaching in the Digital Age—Developing a Support Program for Nursing Education Providers: Design-Based Research (e66109)
Stefan Walzer, Carolin Barthel, Ronja Pazouki, Helga Marx, Sven Ziegler, Peter Koenig, Christiane Kugler, Stefan Jobst. 1460

Sexual and Gender Minority Adolescents' Preferences for HIV Pre-Exposure Prophylaxis Social Marketing Campaigns: Qualitative Preimplementation Study (e60398)
Kathryn Macapagal, Juan Zapata, Junye Ma, Jacob Gordon, Christopher Owens, Silvia Valadez-Tapia, Peter Cummings, Nathan Walter, Jim Pickett. 1473

A Mobile Health Intervention to Support Collaborative Decision-Making in Mental Health Care: Development and Usability (e57614)
Kristin Romm, Mari Skoge, Elizabeth Barrett, Lars-Christian Berentzen, Dagfinn Bergsager, Pål Fugelli, Thomas Bjella, Erlend Gardsjord, Kristine Kling, Sindre Kruse, Kari Kværner, Ingrid Melle, Erlend Mork, Henrik Ihler, Eline Rognli, Carmen Simonsen, Tor Værnes, Sofie Aminoff. 1493

A Mindfulness-Based App Intervention for Pregnant Women: Qualitative Evaluation of a Prototype Using Multiple Case Studies (e58265)
Silvia Rizzi, Maria Pavesi, Alessia Moser, Francesca Paolazzi, Michele Marchesoni, Stefania Poggianella, Erik Gadotti, Stefano Forti. 1508

Reflections of Foster Youth Engaging in the Co-Design of Digital Mental Health Technology: Duoethnography Study (e53231)
Ifunanya Ezimora, Tylia Lundberg, Dylan Miars, Jeruel Trujeque, Ashley Papias, Margareth Del Cid, Johanna Folk, Marina Tolou-Shams. . . .
1 5 2 6

Developing a Machine Learning–Based Automated Patient Engagement Estimator for Telehealth: Algorithm Development and Validation Study (e46390)
Pooja Guhan, Naman Awasthi, Kathryn McDonald, Kristin Bussell, Gloria Reeves, Dinesh Manocha, Aniket Bera. 1541

Psychological and Behavioral Insights From Social Media Users: Natural Language Processing–Based Quantitative Study on Mental Well-Being (e60286)
Xingwei Yang, Guang Li. 1559

Attitudes of German General Practitioners Toward eHealth Apps for Dementia Risk Reduction: Qualitative Interview Study (e56310)
Adrian Schultz, Melanie Luppa, Markus Bleckwenn, Steffi Riedel-Heller, Andrea Zuelke. 1581

Discrimination of Radiologists' Experience Level Using Eye-Tracking Technology and Machine Learning: Case Study (e53928)
Stanford Martinez, Carolina Ramirez-Tamayo, Syed Akhter Faruqui, Kal Clark, Adel Alaeddini, Nicholas Czarnek, Aarushi Aggarwal, Sahra Emamzadeh, Jeffrey Mock, Edward Golob. 1595

Use of Go-Beyond as a Self-Directed Internet-Based Program Supporting Veterans' Transition to Civilian Life: Preliminary Usability Study (e60868)
Karolina Alichniewicz, Sarah Hampton, Madeline Romaniuk, Darcy Bennett, Camila Guindalini. 1611

XSL•FO
RenderX

Usability and Usefulness of a Symptom Management Coaching System for Patients With Cancer Treated With Immune Checkpoint Inhibitors: Comparative Mixed Methods Study (e57659)
Savannah Glaser, Itske Fraterman, Noah van Brummelen, Valentina Tibollo, Laura Del Campo, Henk Mallo, Sofie Wilgenhof, Szymon Wilk, Vitali Gisko, Vadzim Khadakou, Ronald Cornet, Manuel Ottaviano, Stephanie Medlock. . . . 1625

Improving Mental Health and Well-Being Through the Paradym App: Quantitative Study of Real-World Data (e68031)
Athina Metaxa, Shaun Liverpool, Mia Eisenstadt, John Pollard, Courtney Carlsson. . . . 1642

Digital Mindfulness Training for Burnout Reduction in Physicians: Clinician-Driven Approach (e63197)
Lia Antico, Judson Brewer. . . . 1655

Factors Influencing Telemedicine Adoption Among Health Care Professionals: Qualitative Interview Study (e54777)
Fiona Schürmann, Daniel Westmattelmann, Gerhard Schewe. . . . 1683

A Digital Tool (Technology-Assisted Problem Management Plus) for Lay Health Workers to Address Common Mental Health Disorders: Co-production and Usability Study in Pakistan (e59414)
Maham Saleem, Shamsa Zafar, Thomas Klein, Markus Koesters, Adnan Bashir, Daniela Fuhr, Siham Sikander, Hajo Zeeb. . . . 1702

Technological Adjuncts to Streamline Patient Recruitment, Informed Consent, and Data Management Processes in Clinical Research: Observational Study (e58628)
Jodie Koh, Stacey Caron, Amber Watters, Mahesh Vaidyanathan, David Melnick, Alyssa Santi, Kenneth Hudson, Catherine Arguelles, Priyanka Mathur, Mozziyar Etemadi. . . . 1720

Understanding Individual Differences in Happiness Sources and Implications for Health Technology Design: Exploratory Analysis of an Open Dataset (e65658)
Edel Ennis, Raymond Bond, Maurice Mulvenna, Colm Sweeney. . . . 1728

Experiences of Peer Mentoring Sexual and Gender Minority Emerging Adults Who Are at Risk for Suicide: Mixed Methods Study (e67814)
Jennifer Tran, Jessica Webster, James Wolfe, Jennifer Ben Nathan, Lindiwe Mayinja, Marin Kautz, Maria Oquendo, Gregory Brown, David Mandell, Danielle Mowery, José Bauermeister, Lily Brown. . . . 1740

Self-Guided Smartphone App (Vimbo) for the Reduction of Symptoms of Depression and Anxiety in South African Adults: Pilot Quantitative Single-Arm Study (e54216)
Sherrie Steyn, Meggan Slabbert. . . . 1751

Co-Designing a Digital Brief Intervention to Reduce the Risk of Prescription Opioid–Related Harm Among People With Chronic Noncancer Pain: Qualitative Analysis of Patient Lived Experiences (e57208)
Rachel Elphinston, Sue Pager, Kelly Brown, Michele Sterling, Farhad Fatehi, Paul Gray, Linda Hipper, Lauren Cahill, Jason Connor. . . . 1764

Human-Centered Design of an mHealth Tool for Optimizing HIV Index Testing in Wartime Ukraine: Formative Research Case Study (e66132)
Nancy Puttkammer, Elizabeth Dunbar, Myroslava Germanovych, Mariia Rosol, Matthew Golden, Anna Hubashova, Vladyslav Fedorchenko, Larisa Hetman, Liudmyla Legkostup, Jan Flowers, Olena Nesterova. . . . 1783

Evaluating the Impact of Pharmacotherapy in Augmenting Quit Rates Among Hispanic Adults in an App-Delivered Smoking Cessation Intervention: Secondary Analysis of a Randomized Controlled Trial (e69311)
Margarita Santiago-Torres, Kristin Mull, Brianna Sullivan, Ana Cupertino, Ramzi Salloum, Matthew Triplette, Michael Zvolensky, Jonathan Bricker. . . . 1802

Effectiveness of Cognitive Behavioral Therapy Provided Through a Web Application for Subthreshold Depression, Subthreshold Insomnia, and Subthreshold Panic: Open-Labeled 6-Arm Randomized Clinical Trial Pilot Study (e63139)
Kayoko Taguchi, Mirai Miyoshi, Yoichi Seki, Shiori Baba, Eiji Shimizu. . . . 1815

XSL•FO
RenderX

Effectiveness of Electronic Quality Improvement Activities to Reduce Cardiovascular Disease Risk in People With Chronic Kidney Disease in General Practice: Cluster Randomized Trial With Active Control (e54147)
Jo-Anne Manski-Nankervis, Barbara Hunter, Natalie Lumsden, Adrian Laughlin, Rita McMorrow, Douglas Boyle, Patty Chondros, Shilpanjali Jesudason, Jan Radford, Megan Prictor, Jon Emery, Paul Amores, An Tran-Duy, Craig Nelson. 1831

Nurses' Perspectives and Experiences of Using a Bed-Exit Information System in an Acute Hospital Setting: Mixed Methods Study (e64444)
Stefan Walzer, Isabel Schön, Johanna Pfeil, Sam Klemm, Sven Ziegler, Claudia Schmoor, Christophe Kunze. 1851

Status of Digital Health Technology Adoption in 5 Vietnamese Hospitals: Cross-Sectional Assessment (e53483)
Duc Tran, Nguyen Thanh Dung, Chau Minh Duc, Huynh Ngoc Hon, Le Minh Khoi, Nguyen Phuc Hau, Duong Thi Thu Huyen, Huynh Thi Le Thu, Tran Van Duc, VITAL (Vietnam ICU Translational Applications Laboratory) Investigators, Lam Minh Yen, C Thwaites, Chris Paton. 1864

Adapting a Mobile Health App for Smoking Cessation in Black Adults With Anxiety Through an Analysis of the Mobile Anxiety Sensitivity Program Proof-of-Concept Trial: Qualitative Study (e53566)
Marshall Cheney, Adam Alexander, Lorra Garey, Matthew Gallagher, Emily Hébert, Anka Vujanovic, Krista Kezbers, Cameron Matoska, Michael Zvolensky, Michael Businelle. 1880

Improving the Utility, Safety, and Ethical Use of a Passive Mood-Tracking App for People With Bipolar Disorder Using Coproduction: Qualitative Focus Group Study (e65140)
Laurence Astill Wright, Matthew Moore, Stuart Reeves, Elvira Vallejos, Richard Morriss. 1897

Social Media Use and Oral Health–Related Misconceptions in Saudi Arabia: Cross-Sectional Study (e70071)
Rahaf BinHamdan, Salwa Alsadhan, Arwa Gazzaz, AlBandary AlJameel. 1910

mHealth App to Promote Healthy Lifestyles for Diverse Families Living in Rural Areas: Usability Study (e60495)
Alejandra Perez Ramirez, Adrian Ortega, Natalie Stephenson, Angel Muñoz Osorio, Anne Kazak, Thao-Ly Phan. 1926

Social Acceptance of Smart Glasses in Health Care: Model Evaluation Study of Anticipated Adoption and Social Interaction (e49610)
Niek Zuidhof, Oscar Peters, Peter-Paul Verbeek, Somaya Ben Allouch. 1941

Designing Patient-Centered Interventions for Emergency Care: Participatory Design Study (e63610)
Woosuk Seo, Shruti Jain, Vivian Le, Jiaqi Li, Zhan Zhang, Hardeep Singh, Kalyan Pasupathy, Prashant Mahajan, Sun Park. 1958

The Effect of the Virtual Reality–Based Biofeedback Intervention DEEP on Stress, Emotional Tension, and Anger in Forensic Psychiatric Inpatients: Mixed Methods Single-Case Experimental Design (e65206)
Lisa Klein Haneveld, Tessa Dekkers, Yvonne Bouman, Hanneke Scholten, Joanneke Weerdmeester, Saskia Kelders, Hanneke Kip. 1974

Machine Learning Approach to Identifying Wrong-Site Surgeries Using Centers for Medicare and Medicaid Services Dataset: Development and Validation Study (e68436)
Yuan-Hsin Chen, Ching-Hsuan Lin, Chiao-Hsin Fan, An Long, Jeremiah Scholl, Yen-Pin Kao, Usman Iqbal, Yu-Chuan Li. 1991

Exploring the Acceptance of Just-in-Time Adaptive Lifestyle Support for People With Type 2 Diabetes: Qualitative Acceptability Study (e65026)
Eclaire Hietbrink, Anouk Middelweerd, Wendy d'Hollosy, Laura Schrijver, Gozewijn Laverman, Miriam Vollenbroek-Hutten. 2001

Nomophobia, Psychopathology, and Smartphone-Inferred Behaviors in Youth With Depression: Longitudinal Study (e57512)
Tianyi Zhang, Andres Camargo, Lianne Schmaal, Vassilis Kostakos, Simon D'Alfonso. 2019

XSL•FO
RenderX

Adapting the Germ Defence Web-Based Intervention to Improve Infection Prevention and Control in Care Homes: Interview Study Among Care Home Staff (e66706)
Alex Hall, Johanna Aguilera-Muñoz, Lisa McGarrigle, Charlotte Eost-Telling, James Denison-Day, Christie Cabral, Merlin Willcox, Chris Todd. 2026

Designing Values Elicitation Technologies for Mental Health and Chronic Care Integration: User-Centered Design Approach (e68419)
Isabel Rooper, William Liem, Martha Burla, Jacob Gordon, Lara Baez, Rachel Kornfield, Andrew Berry. . . . 2052

A Classroom-Based Intervention for Reducing Sedentary Behavior and Improving Spinal Health: Pragmatic Stepped-Wedge Feasibility Randomized Controlled Trial (e65169)
Dominic Fisher, Rentia Maart, Lehana Thabane, Quinette Louw. . . . 2068

Formative Evaluation of Suicide Prevention Websites for Men: Qualitative Study with Men at Risk of Suicide and with Potential Gatekeepers (e59829)
Doreen Reifegerste, Anna Wagner, Lisa Huber, Manuel Fastuca. . . . 2081

Automated Digital Safety Planning Interventions for Young Adults: Qualitative Study Using Online Co-design Methods (e69602)
Jonah Meyerhoff, Sarah Popowski, Tanvi Lakhtakia, Emily Tack, Rachel Kornfield, Kaylee Kruzan, Charles Krause, Theresa Nguyen, Kevin Rushton, Anthony Pisani, Madhu Reddy, Kimberly Van Orden, David Mohr. . . . 2095

Evaluation of the Clear Fear Smartphone App for Young People Experiencing Anxiety: Uncontrolled Pre–and Post–Follow-Up Study (e55603)
Chiara Samele, Norman Urquia, Rachel Edwards, Katie Donnell, Nihara Krause. . . . 2110

A Digital Program for Daily Life Management With Endometriosis: Pilot Cohort Study on Symptoms and Quality of Life Among Participants (e58262)
Zélia Breton, Emilie Stern, Mathilde Pinault, Delphine Lhuillery, Erick Petit, Pierre Panel, Maïa Alexaline. . . . 2127

Daily Treatment Monitoring for Patients Receiving Home-Based Peritoneal Dialysis and Prediction of Heart Failure Risk: mHealth Tool Development and Modeling Study (e56254)
Jia Wu, Youjia Zeng, Jun Yang, Yutong Yao, Xiuling Xu, Gaofeng Song, Wuyong Yi, Taifen Wang, Yihou Zheng, Zhongwei Jia, Xiangyu Yan. . . 2146

Experiences With an In-Bed Real-Time Motion Monitoring System on a Geriatric Ward: Mixed Methods Study (e63572)
Stefan Walzer, Isabel Schön, Johanna Pfeil, Nicola Merz, Helga Marx, Sven Ziegler, Christophe Kunze. . . . 2165

Patient Adoption of Digital Use Cases in Family Medicine and a Nuanced Implementation Approach for Family Doctors: Quantitative Web-Based Survey Study (e58867)
Julian Beerbaum, Sibylle Robens, Leonard Fehring, Achim Mortsiefer, Sven Meister. . . . 2178

Feasibility and Links Between Emotions, Physical States, and Eating Behavior in Patients After Metabolic Bariatric Surgery: Experience Sampling Study (e60486)
Ellen Kuipers, Josien Timmerman, Marc van Det, Miriam Vollenbroek-Hutten. . . . 2195

FRAILSURVEY—an mHealth App for Self-Assessment of Frailty Based on the Portuguese Version of the Groningen Frailty Indicator: Validation and Reliability Study (e51975)
Luis Midao, Mafalda Duarte, Rute Sampaio, Marta Almada, Cláudia Dias, Constança Paúl, Elísio Costa. . . . 2211

Evaluation of the MyFertiCoach Lifestyle App for Subfertile Couples: Single-Center Evaluation of Augmented Standard Care (e64239)
Jesper Smeenk, Ellen Smit, Marc Jacobs, Ilse van Rooij. . . . 2227

XSL•FO
RenderX

Development of a Voice-Activated Virtual Assistant to Improve Insomnia Among Young Adult Cancer Survivors: Mixed Methods Feasibility and Acceptability Study (e64869)
Hunter Groninger, Hannah Arem, Lylian Ayangma, Lisa Gong, Eric Zhou, Daniel Greenberg. 2245

Zoom-Delivered Empowered Relief for Chronic Pain: Observational Longitudinal Pilot Study Exploring Feasibility and Pain-Related Outcomes in Patients on Long-Term Opioids (e68292)
Karlyn Edwards, Troy Dildine, Dokyoung You, Ashley Herrick, Beth Darnall, Sean Mackey, Maisa Ziadni. 2258

Digital Assessment of Cognitive Health in Outpatient Primary Care: Usability Study (e66695)
Adam Doerr, Taylor Orwig, Matthew McNulty, Stephanie Sison, David Paquette, Robert Leung, Huitong Ding, Stephen Erban, Bruce Weinstein, Yurima Guilarte-Walker, Adrian Zai, Allan Walkey, Apurv Soni, David McManus, Honghuang Lin. 2277

Teen Perspectives on Integrating Digital Mental Health Programs for Teens Into Public Libraries ("I Was Always at the Library"): Qualitative Interview Study (e67454)
Ashley Knapp, Katherine Cohen, Kaylee Kruzan, Rachel Kornfield, Miguel Herrera, Aderonke Pederson, Sydney Lee, Kathryn Macapagal, Chantelle Roulston, Kaleigh Clarke, Clarisa Wijaya, Robert Simmons, Latonia Jackson, Simrandeep Kour, Sandra Franco, David Mohr. 2286

Use of Video Consultation Between 2017 and 2020 in Outpatient Medical Care in Germany and Characteristics of Their User Groups: Analysis of Claims Data (e60170)
Theresa Hüer, Anke Walendzik, Lara Kleinschmidt, Klemens Höfer, Beatrice Nauendorf, Juliane Malsch, Matthias Brittner, Paul Brandenburg, André Aeustergerling, Udo Schneider, Anja Wadeck, Sebastian Liersch, Stephanie Sehlen, Katharina Schwarze, Jürgen Wasem. 2302

Development of a Clinical Decision Support Tool to Implement Asthma Management Guidelines in Pediatric Primary Care: Qualitative Study (e65794)
David Fedele, Jessica Ray, Jaya Mallela, Jiang Bian, Aokun Chen, Xiao Qin, Ramzi Salloum, Maria Kelly, Matthew Gurka, Jessica Hollenbach.
2 3 1 5

Designing a Digital Intervention to Increase Human Milk Feeding Among Black Mothers: Qualitative Study of Acceptability and Preferences (e67284)
Loral Patchen, Jeannette Tsuei, Donna Sherard, Patricia Moriarty, Zoe Mungai-Barris, Tony Ma, Elina Bajracharya, Katie Chang, William Evans. 2327

Exploring Laypersons' Experiences With a Mobile Symptom Checker App as an Interface Between eHealth Literacy, Health Literacy, and Health-Related Behavior: Qualitative Interview Study (e60647)
Roland Koch, Marie-Theres Steffen, Anna-Jasmin Wetzel, Christine Preiser, Malte Klemmt, Hans-Jörg Ehni, Regina Mueller, Stefanie Joos. .
2 3 4 0

Exploring Physicians' Dual Perspectives on the Transition From Free Text to Structured and Standardized Documentation Practices: Interview and Participant Observational Study (e63902)
Olga Golburean, Rune Pedersen, Line Melby, Arild Faxvaag. 2358

Co-Design of a Depression Self-Management Tool for Adolescent and Young Adult Cancer Survivors: User-Centered Design Approach (e67175)
Karly Murphy, Rachel Glock, David Victorson, Madhu Reddy, Sarah Birken, John Salsman. 2374

Safe Listening Beliefs, Attitudes, and Practices Among Gamers and Esports Participants: International Web-Based Survey (e60476)
Nicola Diviani, Shelly Chadha, Peter Mulas, Sara Rubinelli. 2391

Perceived Trust and Professional Identity Threat in AI-Based Clinical Decision Support Systems: Scenario-Based Experimental Study on AI Process Design Features (e64266)
Sophia Ackerhans, Kai Wehkamp, Rainer Petzina, Daniel Dumitrescu, Carsten Schultz. 2406

XSL•FO
RenderX

Stigma and Behavior Change Techniques in Substance Use Recovery: Qualitative Study of Social Media Narratives (e57468)
Annie Chen, Lexie Wang, Shana Johnny, Sharon Wong, Rahul Chaliparambil, Mike Conway, Joseph Glass. 2423

User Experience With a Personalized mHealth Service for Physical Activity Promotion in University Students: Mixed Methods Study (e64384)
Silke Wittmar, Tom Frankenstein, Vincent Timm, Peter Frei, Nicolas Kurpiers, Stefan Wölwer, Axel Schäfer. 2439

Convergent and Known-Groups Validity and Sensitivity to Change of the Virtual Performance Measure in Patients With Hip and Knee Osteoarthritis: Longitudinal Study (e69001)
Helen Razmjou, Suzanne Denis, Susan Robarts, Amy Wainwright, Patricia Dickson, John Murnaghan. 2456

Ambulance Commanders' Reluctance to Enter Road Tunnels in Simulated Incidents and the Effects of a Tunnel-Specific e-Learning Course on Decision-Making: Web-Based Randomized Controlled Trial (e58542)
Johan Hylander, Lina Gyllencreutz, Michael Haney, Anton Westman. 2463

Videoconference-Delivered Acceptance and Commitment Therapy for Family Caregivers of People With Dementia: Pilot Randomized Controlled Trial (e67545)
Areum Han, Robert Oster, Hon Yuen, Jeremy Jenkins, Jessica Hawkins, Lauren Edwards. 2480

Personalized Physician-Assisted Sleep Advice for Shift Workers: Algorithm Development and Validation Study (e65000)
Yufei Shen, Alicia Choto Olivier, Han Yu, Asami Ito-Masui, Ryota Sakamoto, Motomu Shimaoka, Akane Sano. 2496

Examining the Client Experience of Digital Tools in Blended Care Therapy: Qualitative Interview Study (e68249)
Emily Lattie, Miranda Beltzer, Alethea Varra, Connie Chen, Anita Lungu. 2518

A Remote Intervention Based on mHealth and Community Health Workers for Antiretroviral Therapy Adherence in People With HIV: Pilot Randomized Controlled Trial (e67997)
Shivesh Shourya, Jianfang Liu, Sophia McInerney, Trinity Casimir, James Kenniff, Trace Kershaw, David Batey, Rebecca Schnall. 2529

Assessing the Cultural Fit of a Digital Sleep Intervention for Refugees in Germany: Qualitative Study (e65412)
Maja Blomenkamp, Andrea Kiesel, Harald Baumeister, Dirk Lehr, Josef Unterrainer, Lasse Sander, Kerstin Spanhel. 2552

Digital Health Platform for Maternal Health: Design, Recruitment Strategies, and Lessons Learned From the PowerMom Observational Cohort Study (e70149)
Toluwalase Ajayi, Jacqueline Kueper, Lauren Ariniello, Diana Ho, Felipe Delgado, Matthew Beal, Jill Waalen, Katie Baca Motes, Edward Ramos. 2567

The Use of Mobile Apps in Adolescent Psychotherapy: Assessment of Psychotherapists' Perspectives (e65788)
Sarah Wüllner, Katharin Hermenau, Tobias Hecker, Michael Siniatchkin. 2583

Developing an Internet-Based Cognitive Behavioral Therapy Intervention for Adolescents With Anxiety Disorders: Design, Usability, and Initial Evaluation of the CoolMinds Intervention (e66966)
Nikita Sørensen, Helene Skaarnes, Kim Mathiasen, Mikael Thastum, Johanne Lomholt. 2599

Opportunities and Challenges Surrounding the Use of Wearable Sensor Bracelets for Infectious Disease Detection During Hajj: Qualitative Interview Study (e60484)
Noha Maddah, Arpana Verma, John Ainsworth. 2615

A New Mobile App to Train Attention Processes in People With Traumatic Brain Injury: Logical and Ecological Content Validation Study (e64174)
Roxanne Laverdière, Philip Jackson, Frédéric Banville. 2634

XSL•FO
RenderX

Cost-Effectiveness Analysis of a Machine Learning–Based eHealth System to Predict and Reduce Emergency Department Visits and Unscheduled Hospitalizations of Older People Living at Home: Retrospective Study (e63700)
Charlotte Havreng-Théry, Arnaud Fouchard, Fabrice Denis, Jacques-Henri Veyron, Joël Belmin. 2647

The EmpkinS-EKSpression Reappraisal Training Augmented With Kinesthesia in Depression: One-Armed Feasibility Study (e65357)
Marie Keinert, Lena Schindler-Gmelch, Lydia Rupp, Misha Sadeghi, Robert Richer, Klara Capito, Bjoern Eskofier, Matthias Berking. 2656

Evaluating User Engagement With a Real-Time, Text-Based Digital Mental Health Support App: Cross-Sectional, Retrospective Study (e66301)
Edward Coffield, Khadeja Kausar. 2672

Acceptability, Usability, and Insights Into Cybersickness Levels of a Novel Virtual Reality Environment for the Evaluation of Depressive Symptoms: Exploratory Observational Study (e68132)
Sara Sutori, Emma Eliasson, Francesca Mura, Victor Ortiz, Vincenzo Catrambonephd, Gergö Hadlaczky, Ivo Todorov, Antonio Alfeo, Valentina Cardi, Mario Cimino, Giovanna Mioni, Mariano Raya, Gaetano Valenza, Vladimir Carli, Claudio Gentili. 2681

A Digital Photo Activity Intervention for Nursing Home Residents With Dementia and Their Carers: Mixed Methods Process Evaluation (e56586)
Josephine Tan, David Neal, Maria Vilmen, Petra Boersma, Teake Ettema, Robbert Gobbens, Sietske Sikkes, Rose-Marie Dröes. 2702

Creating a Usable and Effective Digital Intervention to Support Men to Test for HIV and Link to Care in A Resource-Constrained Setting: Iterative Design Based on A Person-Based Approach and Human Computer Interaction Methods (e65185)
Anya Zeitlin, Thulile Mathenjwa, Thembelihle Zuma, Sally Wyke, Philippa Matthews, Nuala McGrath, Janet Seeley, Maryam Shahmanesh, Ann Blandford. 2725

Development of a GPT-4–Powered Virtual Simulated Patient and Communication Training Platform for Medical Students to Practice Discussing Abnormal Mammogram Results With Patients: Multiphase Study (e65670)
Dan Weisman, Alanna Sugarman, Yue Huang, Lillian Gelberg, Patricia Ganz, Warren Comulada. 2740

User Experience of and Adherence to a Smartphone App to Maintain Behavior Change and Self-Management in Patients With Work-Related Skin Diseases: Multistep, Single-Arm Feasibility Study (e66791)
Nele Ristow, Annika Wilke, Christoph Skudlik, Swen John, Michaela Ludewig. 2757

Designing Health Care Provider–Centered Emergency Department Interventions: Participatory Design Study (e68891)
Woosuk Seo, Jiaqi Li, Zhan Zhang, Chuxuan Zheng, Hardeep Singh, Kalyan Pasupathy, Prashant Mahajan, Sun Park. 2776

Compatibility of the CEN-ISO/TS 82304-2 Health App Assessment Framework With Catalan and Italian Health Authorities' Needs: Qualitative Interview Study (e67855)
Petra Hoogendoorn, Mariam Shokralla, Romy Willemsen, Nick Guldemond, María Villalobos-Quesada. 2792

A Mobile Health App Informed by the Multi-Process Action Control Framework to Promote Physical Activity Among Inactive Adults: Iterative Usability Study (e59477)
Heather Hollman, Wuyou Sui, Haowei Zhang, Ryan Rhodes. 2810

XSL•FO
RenderX

Applying Patient and Health Professional Preferences in Co-Designing a Digital Brief Intervention to Reduce the Risk of Prescription Opioid–Related Harm Among Patients With Chronic Noncancer Pain: Qualitative Analysis (e57212)
Rachel Elphinston, Sue Pager, Farhad Fatehi, Michele Sterling, Kelly Brown, Paul Gray, Linda Hipper, Lauren Cahill, Maisa Ziadni, Peter Worthy, Jason Connor. . . . 2829

Acceptability and Usability of a Digital Behavioral Health Platform for Youth at Risk of Suicide: User-Centered Design Study With Patients, Practitioners, and Business Gatekeepers (e65418)
Trinity Tse, Lauren Weiner, Carter Funkhouser, Danielle DeLuise, Colleen Cullen, Zachary Blumkin, Casey O'Brien, Randy Auerbach, Nicholas Allen. . . . 2848

Assessing the Causal Association Between COVID-19 and Graves Disease: Mendelian Randomization Study (e66003)
Hui Nian, Yu Bai, Hua Yu. . . . 2863

COVID-19 Perceptions Among Communities Living on Ground Crossings of Somali Region of Ethiopia: Community Cross-Sectional Survey Study (e66751)
Alinoor Farah, Abdifatah Abdulahi, Abdulahi Hussein, Ahmed Hussein, Abdi Osman, Mohamed Mohamud, Hasan Mowlid, Girum Hailu, Fathia Alwan, Ermiyas Bizuneh, Ahmed Ibrahim, Elyas Abdulahi. . . . 2872

Synthetic Data-Driven Approaches for Chinese Medical Abstract Sentence Classification: Computational Study (e54803)
Jiajia Li, Zikai Wang, Longxuan Yu, Hui Liu, Haitao Song. . . . 2890

Identifying High-Priority Ethical Challenges for Precision Emergency Medicine: Nominal Group Study (e68371)
Christian Rose, Emily Shearer, Isabela Woller, Ashley Foster, Nicholas Ashenburg, Ireh Kim, Jennifer Newberry. . . . 2903

Impact of Psychological Resilience on the Fear of Pain and Activity Recovery in Postsurgical Patients: Observational Cohort Study (e63556)
Yang Luo, Sisi Li, Lijuan Feng, Junyi Zheng, Chunfen Peng, Lihong Bao. . . . 2915

Impact of Preventive Intravenous Amiodarone on Reperfusion Ventricular Fibrillation in Patients With Left Ventricular Hypertrophy Undergoing Open-Heart Surgery: Randomized Controlled Clinical Trial (e64586)
Chan-Juan Gong, Xiao-Kai Zhou, Zhen-Feng Zhang, Yin Fang. . . . 2924

Identification of Behavioral, Clinical, and Psychological Antecedents of Acute Stimulant Poisoning: Development and Implementation of a Mixed Methods Psychological Autopsy Study (e64873)
Marley Antolin Muñiz, Vanessa McMahan, Xochitl Luna Marti, Sarah Brennan, Sophia Tavasieff, Luke Rodda, James Knoll, Phillip Coffin. . . 2 9 3 5

Resilience Informatics in Public Health: Qualitative Analysis of Conference Proceedings (e63217)
Maiya Block Ngaybe, Lidia Azurdia Sierra, Andrew McNair, Myla Gonzalez, Mona Arora, Kacey Ernst, Enrique Noriega-Atala, M Iyengar. . . 2 9 4 4

Assessment, Decision, Adaptation, Production, Topical Experts-Integration, Training, and Testing (ADAPT-ITT) Framework to Tailor Evidence-Based Posttraumatic Stress Disorder Treatment for People With HIV to Enhance Engagement and Adherence: Qualitative Results from a Feasibility Randomized Controlled Trial (e64258)
Cristina Lopez, Angela Moreland, Stephanie Amaya, Erin Bisca, Christin Mujica, Tayler Wilson, Nathaniel Baker, Lauren Richey, Allison Eckard, Patricia Resick, Steven Safren, Carla Danielson. . . . 2960

Designing the First Pregnancy Guaranteed Income Program in the United States: Qualitative Needs Assessment and Human-Centered Design to Develop the Abundant Birth Project (e60829)
Deborah Karasek, Jazzmin Williams, Michaela Taylor, Monica De La Cruz, Stephanie Arteaga, Sabra Bell, Esperanza Castillo, Maile Chand, Anjeanette Coats, Erin Hubbard, Latriece Love-Goodlett, Breezy Powell, Solaire Spellen, Zea Malawa, Anu Gomez. . . . 2974

XSL•FO
RenderX

Tobacco Use, Risk Perceptions, and Characteristics of Adults Who Used a Heated Tobacco Product (IQOS) in the United States: Cross-Sectional Survey Study (e57398)
Hui Cheng, Brendan Noggle, Andrea Vansickel, Edward Largo, Pierpaolo Magnani. . . . 2988

Predictors of the Intention to Stop Using Smart Devices at Bedtime Among University Students in Saudi Arabia: Cross-Sectional Survey (e67223)
Manal Almalki. . . . 3003

Impact of a Sensorimotor Integration and Hyperstimulation Program on Global Motor Skills in Moroccan Children With Autism Spectrum Disorder: Exploratory Clinical Quasi-Experimental Study (e65767)
Rachid Touali, Jamal Zerouaoui, El Chakir, Hung Bui, Mario Leone, Maxime Allisse. . . . 3020

The Norwegian Version of the Self-Efficacy in Clinical Performance Scale (SECP): Psychometric Validation Study (e68173)
Camilla Olaussen, Marko Stojiljkovic, Jaroslav Zlamal, Tone Fløtlo, Andréa Nes. . . . 3035

Increasing the Uptake of Breast and Cervical Cancer Screening Via the MAwar Application: Stakeholder-Driven Web Application Development Study (e65542)
Nurfarhana Nasrudin, Shariff-Ghazali Sazlina, Ai Cheong, Ping Lee, Soo-Hwang Teo, Abdul Aneesa, Chin Teo, Fakhrul Rokhani, Nuzul Haron, Noor Harrun, Bee Ho, Salbiah Mohamed Isa. . . . 3053

Using a Hybrid of AI and Template-Based Method in Automatic Item Generation to Create Multiple-Choice Questions in Medical Education: Hybrid AIG (e65726)
Yavuz Kıyak, Andrzej Kononowicz. . . . 3066

Using Large Language Models to Automate Data Extraction From Surgical Pathology Reports: Retrospective Cohort Study (e64544)
Denise Lee, Akhil Vaid, Kartikeya Menon, Robert Freeman, David Matteson, Michael Marin, Girish Nadkarni. . . . 3076

Moving Standard Deviation of Trunk Acceleration as a Quantification Index for Physical Activities: Validation Study (e63064)
Takuya Suzuki, Yuji Kono, Takayuki Ogasawara, Masahiko Mukaino, Yasushi Aoshima, Shotaro Furuzawa, Yurie Fujita, Hirotaka Matsuura, Masumi Yamaguchi, Shingo Tsukada, Yohei Otaka. . . . 3085

Provider Perspectives on Implementing an Enhanced Digital Screening for Adolescent Depression and Suicidality: Qualitative Study (e67624)
Morgan Coren, Oliver Lindhiem, Abby Angus, Emma Toevs, Ana Radovic. . . . 3095

Reliability of Average Daily Steps Measured Through a Consumer Smartwatch in Parkinson Disease Phenotypes, Stages, and Severities: Cross-Sectional Study (e63153)
Edoardo Bianchini, Domiziana Rinaldi, Lanfranco De Carolis, Silvia Galli, Marika Alborghetti, Clint Hansen, Antonio Suppa, Marco Salvetti, Francesco Pontieri, Nicolas Vuillerme. . . . 3104

Patient and Provider Perspectives of a Web-Based Intervention to Support Symptom Management After Radioactive Iodine Treatment for Differentiated Thyroid Cancer: Qualitative Study (e60588)
Alaina Carr, Angela Jenkins, Jacqueline Jonklaas, Kate Gabriel, Kristen Miller, Kristi Graves. . . . 3115

Customizing Computerized Adaptive Test Stopping Rules for Clinical Settings Using the Negative Affect Subdomain of the NIH Toolbox Emotion Battery: Simulation Study (e60215)
Saki Amagai, Aaron Kaat, Rina Fox, Emily Ho, Sarah Pila, Michael Kallen, Benjamin Schalet, Cindy Nowinski, Richard Gershon. . . . 3131

“Your Life, Your Health: Tips and Information for Health and Well-Being”: Development of a World Health Organization Digital Resource to Support Universal Access to Trustworthy Health Information (e57881)
Danielle Muscat, Rachael Hinton, Shyama Kuruvilla, Don Nutbeam. . . . 3141

XSL•FO
RenderX

HUMAn, a Real-Time Evolutive Patient Model for Major Incident Simulation: Development and Validation Study (e66201)
Maxence Laurent, Arnaud Jaccard, Laurent Suppan, Elio Erriquez, Xavier Good, Eric Golay, Dominique Jaccard, Mélanie Suppan. 3151

Exploratory Co-Design on Electronic Health Record Nursing Summaries: Case Study (e68906)
Suhyun Park, Jenna Marquard, Robin Austin, Christie Martin, David Pieczkiewicz, Connie Delaney. 3166

Improving the User Interface and Guiding the Development of Effective Training Material for a Clinical Research Recruitment and Retention Dashboard: Usability Testing Study (e66718)
Leah Gardner, Pezhman Raeisian Parvari, Mark Seidman, Richard Holden, Nicole Fowler, Ben Zarzaur, Diana Summanwar, Cristina Barboi, Malaz Boustani. 3177

Exploring the Relationship Between Smartphone GPS Patterns and Quality of Life in Patients With Advanced Cancer and Their Family Caregivers: Longitudinal Study (e59161)
Kyungmi Lee, Andres Azuero, Sally Engler, Sidharth Kumar, Frank Puga, Alexi Wright, Arif Kamal, Christine Ritchie, George Demiris, Marie Bakitas, J Odom. 3186

eHealth Literacy and Cyberchondria Severity Among Undergraduate Students: Mixed Methods Study (e63449)
Wan -Chen Hsu. 3202

Estimating the Prevalence of Schizophrenia in the General Population of Japan Using an Artificial Neural Network–Based Schizophrenia Classifier: Web-Based Cross-Sectional Survey (e66330)
Pichsinee Choomung, Yupeng He, Masaaki Matsunaga, Kenji Sakuma, Taro Kishi, Yuanying Li, Shinichi Tanihara, Nakao Iwata, Atsuhiko Ota. 3213

Measuring Self-Reported Well-Being of Physicians Using the Well-Being Thermometer: Cohort Study (e54158)
Marios Adamou, Sarah Jones, Niki Kyriakidou, Andrew Mooney, Shriti Pattani, Matthew Roycroft. 3224

Recruitment Challenges and Strategies in a Technology-Based Intervention for Dementia Caregivers: Descriptive Study (e59291)
Eunjung Ko, Ye Gao, Peng Wang, Lahiru Wijayasingha, Kathy Wright, Kristina Gordon, Hongning Wang, John Stankovic, Karen Rose. 3234

Analysis of Social Performance and Action Units During Social Skills Training: Focus Group Study of Adults With Autism Spectrum Disorder and Schizophrenia (e59261)
Hiroki Tanaka, Kana Miyamoto, Jennifer Hamet Bagnou, Elise Prigent, Céline Clavel, Jean-Claude Martin, Satoshi Nakamura. 3246

Unified Mobile App for Streamlining Verbal Autopsy and Cause of Death Assignment in India: Design and Development Study (e59937)
Harleen Kaur, Stuti Tripathi, Manjeet Chalga, Sudhir Benara, Amit Dhiman, Shefali Gupta, Saritha Nair, Geetha Menon, B Gulati, Sandeep Sharma, Saurabh Sharma. 3260

Public Understanding and Expectations of Digital Health Evidence Generation: Focus Group Study (e56523)
Paulina Bondaronek, Jingfeng Li, Henry Potts. 3269

International Expert Consensus on Relevant Health and Functioning Concepts to Assess in Users of Tobacco and Nicotine Products: Delphi Study (e58614)
Vivienne Law, Esther Afolalu, Linda Abetz-Webb, Lee Wemyss, Andrew Turner, Christelle Chrea. 3280

Development of an eHealth Intervention Including Self-Management for Reducing Sedentary Time in the Transition to Retirement: Participatory Design Study (e63567)
Lisa Hultman, Caroline Eklund, Petra von Heideken Wågert, Anne Söderlund, Maria Lindén, Magnus Elfström. 3300

XSL•FO
RenderX

Methods to Adjust for Confounding in Test-Negative Design COVID-19 Effectiveness Studies: Simulation Study (e58981)
Elizabeth Rowley, Patrick Mitchell, Duck-Hye Yang, Ned Lewis, Brian Dixon, Gabriela Vazquez-Benitez, William Fadel, Inih Essien, Allison Naleway, Edward Stenehjem, Toan Ong, Manjusha Gaglani, Karthik Natarajan, Peter Embi, Ryan Wiegand, Ruth Link-Gelles, Mark Tenforde, Bruce Fireman. . . . 3315

Multimodal Pain Recognition in Postoperative Patients: Machine Learning Approach (e67969)
Ajan Subramanian, Rui Cao, Emad Naeini, Seyed Aqajari, Thomas Hughes, Michael-David Calderon, Kai Zheng, Nikil Dutt, Pasi Liljeberg, Sanna Salanterä, Ariana Nelson, Amir Rahmani. . . . 3333

Evaluating Reaction Videos of Young People Watching Edutainment Media (MTV Shuga): Qualitative Observational Study (e55275)
Venetia Baker, Sarah Mulwa, David Khanyile, Georgia Arnold, Simon Cousens, Cherie Cawood, Isolde Birdthistle. . . . 3349

Proficiency, Clarity, and Objectivity of Large Language Models Versus Specialists' Knowledge on COVID-19's Impacts in Pregnancy: Cross-Sectional Pilot Study (e56126)
Nicola Bragazzi, Michèle Buchinger, Hisham Atwan, Ruba Tuma, Francesco Chirico, Lukasz Szarpak, Raymond Farah, Rola Khamisy-Farah.
3 3 6 2

A 4-Site Public Deliberation Project on the Acceptability of Youth Self-Consent in Biomedical HIV Prevention Trials: Assessment of Facilitator Fidelity to Key Principles (e58451)
Claire Draucker, Andrés Carrión, Mary Ott, Ariel Hicks, Amelia Knopf. . . . 3374

Designing a Smartphone-Based Virtual Reality App for Relaxation: Qualitative Crossover Study (e62663)
Amandine Verstegen, Tom Van Daele, Bert Bonroy, Glen Debard, Romy Sels, Marlon van Loo, Sylvie Bernaerts. . . . 3391

Using an Interactive Voice Response Survey to Assess Patient Satisfaction in Ethiopia: Development and Feasibility Study (e67452)
Dessalegn Shamebo, Anagaw Derseh Mebratie, Catherine Arsenault. . . . 3406

Exploring Metadata Catalogs in Health Care Data Ecosystems: Taxonomy Development Study (e63396)
Simon Scheider, Mostafa Mallick. . . . 3419

The Design and Evaluation of a Simulation Tool for Audiology Screening Education: Design Science Approach (e47150)
John Gerdes, Benjamin Schooley, Dakota Sharp, Juliana Miller. . . . 3441

Effects of Missing Data on Heart Rate Variability Measured From A Smartwatch: Exploratory Observational Study (e53645)
Hope Davis-Wilson, Meghan Hegarty-Craver, Pooja Gaur, Matthew Boyce, Jonathan Holt, Edward Preble, Randall Eckhoff, Lei Li, Howard Walls, David Dausch, Dorota Temple. . . . 3456

Development and Implementation of MyPainHub, a Web-Based Resource for People With Musculoskeletal Conditions and Their Health Care Professionals: Mixed Methods Study (e63780)
Kerrie Evans, Jonathan Ko, Dragana Ceprnja, Katherine Maka, Darren Beales, Michele Sterling, Kim Bennell, Gwendolen Jull, Paul Hodges, Marnee McKay, Trudy Rebbeck. . . . 3474

The Feasibility of Large Language Models in Verbal Comprehension Assessment: Mixed Methods Feasibility Study (e68347)
Dorit Hadar-Shoval, Maya Lvovsky, Kfir Asraf, Yoav Shimoni, Zohar Elyoseph. . . . 3490

An Integrated Platform Combining Immersive Virtual Reality and Physiological Sensors for Systematic and Individualized Assessment of Stress Response (bWell): Design and Implementation Study (e64492)
Budhachandra Khundrakpam, Melanie Segado, Jesse Pazdera, Vincent Gagnon Shaigetz, Joshua Granek, Nusrat Choudhury. . . . 3504

XSL•FO
RenderX

Long-Term Experiences of Health Care Providers Using Iris Scanning as an Identification Tool in a Vaccine Trial in the Democratic Republic of the Congo: Qualitative Study (e54921)
Trésor Zola Matuvanga, Antea Paviotti, Freddy Bikioli Bolombo, Gwen Lemey, Ynke Larivière, Maha Salloum, Bernard Isekah Osang'ir, Emmanuel Esanga Longomo, Solange Milolo, Junior Matangila, Vivi Maketa, Patrick Mitashi, Pierre Van Damme, Hypolite Muhindo-Mavoko, Jean-Pierre Van geertruyden. . . . 3522

Assessment of Fraud Deterrence and Detection Procedures Used in a Web-Based Survey Study With Adult Black Cisgender Women: Description of Lessons Learned and Recommendations (e59955)
Amber Sophus, Jason Mitchell. . . . 3534

A Trauma Support App for Young People: Co-design and Usability Study (e57789)
Maria Thell, Kerstin Edvardsson, Reem Aljeshy, Kalid Ibrahim, Georgina Warner. . . . 3546

Kinetic Pattern Recognition in Home-Based Knee Rehabilitation Using Machine Learning Clustering Methods on the Slider Digital Physiotherapy Device: Prospective Observational Study (e69150)
Clement Twumasi, Mikail Aktas, Nicholas Santoni. . . . 3560

Smartwatch-Based Ecological Momentary Assessment for High-Temporal-Density, Longitudinal Measurement of Alcohol Use (AlcoWatch): Feasibility Evaluation (e63184)
Chris Stone, Sally Adams, Robyn Wootton, Andy Skinner. . . . 3578

Using Social Media to Engage and Enroll Underrepresented Populations: Longitudinal Digital Health Research (e68093)
Christiana Harry, Sarah Goodday, Carol Chapman, Emma Karlin, April Damian, Alexa Brooks, Adrien Boch, Nelly Lugo, Rebecca McMillan, Jonell Tempero, Ella Swanson, Shannon Peabody, Diane McKenzie, Stephen Friend. . . . 3594

The Influence of Medical Expertise and Information Search Skills on Medical Information Searching: Comparative Analysis From a Free Data Set (e62754)
Aline Chevalier, Cheyenne Dosso. . . . 3611

Application of an Innovative Methodology to Build Infrastructure for Digital Transformation of Health Systems: Developmental Program Evaluation (e53339)
M Buchan, Tarun Katapally, Jasmin Bhawra. . . . 3627

Ecological Momentary Assessment of Parental Well-Being and Time Use: Mixed Methods Compliance and Feasibility Study (e67451)
Laura Altweck, Samuel Tomczyk. . . . 3644

A Portal-Based Intervention (PATTERN) Designed to Support Medication Use Among Older Adults: Feasibility and Acceptability Study (e71676)
Allison Pack, Stacy Bailey, Rachel O'Conor, Evelyn Velazquez, Guisselle Wismer, Fangyu Yeh, Laura Curtis, Kenya Alcantara, Michael Wolf. . 3 6 7 7

Cardiac Self-Efficacy Improvement in a Digital Heart Health Program: Secondary Analysis From a Feasibility and Acceptability Pilot Study (e60676)
Kimberly Lockwood, Priya Kulkarni, OraLee Branch, Sarah Graham. . . . 3685

Patient Acceptability and Technical Reliability of Wearable Devices Used for Monitoring People With Parkinson Disease: Survey Study (e63704)
Tasmin Rookes, Amit Batla, Megan Armstrong, Gareth Ambler, Kate Walters, Anette Schrag. . . . 3697

Using Text Messaging Ecological Momentary Assessment to Record Changes in e-Cigarette and Combustible Cigarette Use: Pilot Randomized Clinical Trial (e66709)
Tucker Morgan, Michelle He, Andrew Nicholson, Omar El Shahawy, Scott Sherman, Elizabeth Stevens. . . . 3705

XSL•FO
RenderX

Development and Implementation of Strong Foundations, a Digitally Delivered Fall Prevention Program: Usability and Feasibility Pilot Exercise Cohort Study (e67406)
Ryan Moran, David Wing, Hope Davey, Hava Barkai, Jeanne Nichols. . . . 3720

Evaluating the Knowledge Level, Practice, and Behavioral Change Potential of Care Managers in Pressure Injury Prevention Using a Mobile App Prototyping Model in the Home-Care Setting: Single-Arm, Pre-Post Pilot Study (e57768)
Masushi Kohta, Mayumi Takahashi, Hiroe Koyanagi, Junko Sugama. . . . 3730

Assessing Dynamic Cognitive Function in the Daily Lives of Youths With and Without Type 1 Diabetes: Usability Study (e60275)
Mary Ray, Jorie Fleming, Andrew Aschenbrenner, Jason Hassenstab, Brooke Redwine, Carissa Burns, Ana Arbelaez, Mary Vajravelu, Tamara Hershey. . . . 3742

Use of a Technology-Based Fall Prevention Program With Visual Feedback in the Setting of Early Geriatric Rehabilitation: Controlled and Nonrandomized Study (e66692)
Carolin Steinmetz, Christina Stenzel, Maj Sylvester, Denis Glage, Anne Linke, Monika Sadlonova, Christine von Arnim, Marlena Schnieder, Miroslava Valentová, Stephanie Heinemann. . . . 3758

Acceptance of Unsupervised App-Based Cognitive Assessment in Outpatient Care: An Implementation Study (e62706)
Iris Blotenberg, Melanie Boekholt, Nils Lieberknecht, Paula Säring, Jochen Thyrian. . . . 3774

Users' Perspectives of Direct-to-Consumer Telemedicine Services: Survey Study (e68619)
Kate Churruca, Darran Foo, Andrew Turner, Emily Crameri, Maree Saba, Samantha Spanos, Matthew Vickers, Jeffrey Braithwaite, Louise Ellis. . . . 3783

Feasibility, User Acceptance, and Outcomes of Using a Cancer Prehabilitation App for Exercise: Pilot Cohort Study (e64427)
Fuquan Zhang, Deepali Bang, Christine Visperas, Mon Tun, San Tay. . . . 3793

Longitudinal Changes in Pitch-Related Acoustic Characteristics of the Voice Throughout the Menstrual Cycle: Observational Study (e65448)
Jaycee Kaufman, Jouhyun Jeon, Jessica Oreskovic, Anirudh Thommandram, Yan Fossat. . . . 3807

Postpartum Remote Health Coaching Intervention for Individuals With a Hypertensive Disorder of Pregnancy: Proof-of-Concept Study (e65611)
Jaclyn Borrowman, Lucas Carr, Gary Pierce, William Story, Bethany Gibbs, Kara Whitaker. . . . 3818

Testing the Recruitment Frequency, Implementation Fidelity, and Feasibility of Outcomes of the Heart Failure Activity Coach Study (HEALTHY): Pilot Randomized Controlled Trial (e62910)
Andreas Blomqvist, Maria Bäck, Leonie Klompstra, Anna Strömberg, Tiny Jaarsma. . . . 3831

Effectiveness of a Videoconference-Based Cognitive Behavioral Therapy Program for Patients with Schizophrenia: Pilot Randomized Controlled Trial (e59540)
Masayuki Katsushima, Hideki Nakamura, Yuki Shiko, Hideki Hanaoka, Eiji Shimizu. . . . 3847

Testing a Web-Based Interactive Comic Tool to Decrease Obesity Risk Among Racial and Ethnic Minority Preadolescents: Randomized Controlled Trial (e58460)
May Leung, Katrina Mateo, Marlo Dublin, Laura Harrison, Sandra Verdaguer, Katarzyna Wyka. . . . 3864

Exploring Web-Based Support for Suicidal Ideation in the Scottish Population: Usability Study (e55932)
Heather McClelland, Rory O'Connor, Laura Gibson, Donald MacIntyre. . . . 3883

XSL•FO
RenderX

Engagement With an Internet-Administered, Guided, Low-Intensity Cognitive Behavioral Therapy Intervention for Parents of Children Treated for Cancer: Analysis of Log-Data From the ENGAGE Feasibility Trial (e67171)
Christina Reuther, Louise von Essen, Mudassir Mustafa, Markus Saarijärvi, Joanne Woodford. . . . 3896

Women Empowered to Connect With Addiction Resources and Engage in Evidence-Based Treatment (WE-CARE)—an mHealth Application for the Universal Screening of Alcohol, Substance Use, Depression, and Anxiety: Usability and Feasibility Study (e62915)
Krystyna Isaacs, Autumn Shifflett, Kajal Patel, Lacey Karpisek, Yi Cui, Maayan Lawental, Golfo Tzilos Wernette, Brian Borsari, Katie Chang, Tony Ma. . . . 3916

Digital Solution to Support Medication Adherence and Self-Management in Patients with Cancer (SAMSON): Pilot Randomized Controlled Trial (e65302)
Thu Dang, Nilmini Wickramasinghe, Prem Jayaraman, Kate Burbury, Marliese Alexander, Ashley Whitechurch, Mitchell Dyer, Stephen Quinn, Abdur Forkan, Penelope Schofield. . . . 3931

Assessing the Feasibility and Utility of Patient-Specific 3D Advanced Visualization Modeling in Cerebrovascular Disease: Retrospective Analysis and Prospective Survey Pilot Study (e51939)
Korak Sarkar, Vishal Bhimarasetty, Abdul Rahim, Colin Curtis, Kimberly Hughes. . . . 3944

Effect on Response Rates of Adding a QR Code to Patient Consent Forms for Qualitative Research in Patients With Cancer: Pilot Randomized Controlled Trial (e66681)
Rebecca Wyse, Erin Forbes, Grace Norton, Priscilla Viana Da Silva, Kristy Fakes, Sally Johnston, Stephen Smith, Alison Zucca. . . . 3957

Testing and Iterative Improvement of the CEN ISO/TS 82304-2 Health App Quality Assessment: Pilot Interrater Reliability Study (e64565)
Anna-Lena Frey, Diana Matei, Ben Phillips, Adam McCabe, Rachel Fuller, Begoña Laibarra, Laura Alonso, Victor de la Hoz, Carme Pratdepadua Bufill, Berta Llebot Casajuana, Giuseppe D'Avenio, Pier Sottile, Laura Rocchi, Matteo Errera, Yasmine Laaissaoui, Michael Cardinal, Menno Kok, Petra Hoogendoorn. . . . 3964

Digital Translation Platform (Translatly) to Overcome Communication Barriers in Clinical Care: Pilot Study (e63095)
Victor Olsavszky, Mutaz Bazari, Taieb Dai, Ana Olsavszky, Fabian Finkelmeier, Mireen Friedrich-Rust, Stefan Zeuzem, Eva Herrmann, Jan Leipe, Florian Michael, Hans Westernhagen, Olivier Ballo. . . . 3982

Multilevel Intervention to Increase Patient Portal Use in Adults With Type 2 Diabetes Who Access Health Care at Community Health Centers: Single Arm, Pre-Post Pilot Study (e67293)
Robin Whittemore, Sangchoon Jeon, Samuel Akyirem, Helen Chen, Joanna Lipson, Maritza Minchala, Julie Wagner. . . . 3996

Patients' and Health Care Professionals' Expectations of Virtual Therapeutic Agents in Outpatient Aftercare: Qualitative Survey Study (e59527)
Diana Immel, Bernhard Hilpert, Patricia Schwarz, Andreas Hein, Patrick Gebhard, Simon Barton, René Hurlemann. . . . 4012

AI-Enhanced Virtual Reality Self-Talk for Psychological Counseling: Formative Qualitative Study (e67782)
Moreah Zisquit, Alon Shoa, Ramon Oliva, Stav Perry, Bernhard Spanlang, Anat Brunstein Klomek, Mel Slater, Doron Friedman. . . . 4032

Factors Impacting Mobile Health Adoption for Depression Care and Support by Adolescent Mothers in Nigeria: Preliminary Focus Group Study (e42406)
Lola Kola, Tobi Fatodu, Manasseh Kola, Bisola Olayemi, Adeyinka Adefolarin, Simpa Dania, Manasi Kumar, Dror Ben-Zeev. . . . 4045

The Monthly Cycling of Food Insecurity in Latinas at Risk for Diabetes: Methods, Retention, and Sample Characteristics for a Microlongitudinal Design (e66970)
Angela Bermúdez-Millán, Rafael Pérez-Escamilla, Sofia Segura-Pérez, James Grady, Richard Feinn VI, Hanako Agresta, Dean Kim, Julie Wagner. . . . 4054

XSL•FO
RenderX

Improving Health and Well-Being of People With Post–COVID-19 Consequences in South Africa: Situation Analysis and Pilot Intervention Design (e58436)
Nicole Glover, Farzana Sathar, Pride Mokome, Nkululeko Mathabela, Sipokazi Taleni, Sarah van Blydenstein, Anna-Maria Mekota, Salome Charalambous, Andrea Rachow, Olena Ivanova. . . . 4070

Impact of Acute Respiratory Infections on Medical Absenteeism Among Military Personnel: Retrospective Cohort Study (e69113)
Premikha M, Jit Goh, Jing Ng, Adeliza Mutalib, Huai Lim. . . . 4082

Perceptions and Experiences of Developing Prediabetes Awareness and Participating in the National Diabetes Prevention Program Among Hispanic US Adults: Qualitative Study (e66964)
Kevin Joiner, Mackenzie Adams, Edith Minehan, Matthew O'Brien. . . . 4091

Disease Awareness in Patients With Type 2 Diabetes: Analysis of Baseline Data From the SMART-Finder Observational Study (e60246)
Christian Mueller, Thomas Neusser, Inga Thate-Waschke, Julia Nowicki, Tomasz Plominski, Regine Griesinger, Stefanie Kessner, Stephan Martin. . . . 4101

A Very Low–Carbohydrate Program in Adults With Metabolic Dysfunction–Associated Steatotic Liver Disease and Phospholipase Domain–Containing Protein 3 Risk Genotype: Pre-Post Intervention Study (e60051)
Laura Saslow, Jamie Krinock, Alison O'Brien, Kaitlyn Raymond, Hovig Bayandorian, Judith Moskowitz, Jennifer Daubenmier, Antonino Oliveri, Deanna Marriott, Dina Griauzde, Elizabeth Speliotes. . . . 4109

Continuous Glucose Monitoring Among Patients With Type 1 Diabetes in Rwanda (CAPT1D) Phase I: Prospective Observational Feasibility Study (e64585)
Jason Baker, Giacomo Cappon, Jean Habineza, Corey Basch, Steven Mey, Diana Malkin-Washeim, Christian Schuetz, Niyonsenga Simon Pierre, Etienne Uwingabire, Alvera Mukamazimpaka, Paul Mbonyi, Sandhya Narayanan. . . . 4125

Developing Assessments for Key Stakeholders in Pediatric Congenital Heart Disease: Qualitative Pilot Study to Inform Designing of a Medical Education Toy (e63818)
Neda Barbazi, Ji Shin, Gurumurthy Hiremath, Carlye Lauff. . . . 4133

Effect of the Reassured Self-Compassion–Based School Program on Anxiety, Video Game Addiction, and Body Image Among Rural Female Adolescents: Retrospective Study (e68840)
Areeg Zuair. . . . 4147

Machine Learning–Based Prediction of Delirium and Risk Factor Identification in Intensive Care Unit Patients With Burns: Retrospective Observational Study (e65190)
Ryo Esumi, Hiroki Funao, Eiji Kawamoto, Ryota Sakamoto, Asami Ito-Masui, Fumito Okuno, Toru Shinkai, Atsuya Hane, Kaoru Ikejiri, Yuichi Akama, Arong Gaowa, Eun Park, Ryo Momosaki, Ryuji Kaku, Motomu Shimaoka. . . . 4159

Crisis Communication About the Maui Wildfires on TikTok: Content Analysis of Engagement With Maui Wildfire–Related Posts Over 1 Year (e67515)
Jim Stimpson, Aditi Srivastava, Ketan Tamirisa, Joseph Kaholokula, Alexander Ortega. . . . 4179

Fast Healthcare Interoperability Resources (FHIR)–Based Interoperability Design in Indonesia: Content Analysis of Developer Hub's Social Networking Service (e51270)
Lukman Heryawan, Yukiko Mori, Goshiro Yamamoto, Naoto Kume, Lutfan Lazuardi, Anis Fuad, Tomohiro Kuroda. . . . 4191

Informal Caregivers Connecting on the Web: Content Analysis of Posts on Discussion Forums (e64757)
Michelle Foster, Chinenye Egwuonwu, Erin Vernon, Mohammad Alarifi, M Hughes. . . . 4204

XSL•FO
RenderX

Planned Behavior in the United Kingdom and Ireland Online Medicine Purchasing Context: Mixed Methods Survey Study (e55391)
Bernard Naughton. . . . 4215

Quantifying the Regional Disproportionality of COVID-19 Spread: Modeling Study (e59230)
Kenji Sasaki, Yoichi Ikeda, Takashi Nakano. . . . 4274

Increasing COVID-19 Testing and Vaccination Uptake in the Take Care Texas Community-Based Randomized Trial: Adaptive Geospatial Analysis (e62802)
Kehe Zhang, Jocelyn Hunyadi, Marcia de Oliveira Otto, Miryoung Lee, Zitong Zhang, Ryan Ramphul, Jose-Miguel Yamal, Ashraf Yaseen, Alanna Morrison, Shreela Sharma, Mohammad Rahbar, Xu Zhang, Stephen Linder, Dritana Marko, Rachel Roy, Deborah Banerjee, Esmeralda Guajardo, Michelle Crum, Belinda Reininger, Maria Fernandez, Cici Bauer. . . . 4281

Investigating Reddit Data on Type 2 Diabetes Management During the COVID-19 Pandemic Using Latent Dirichlet Allocation Topic Modeling and Valence Aware Dictionary for Sentiment Reasoning Analysis: Content Analysis (e51154)
Meghan Nagpal, Niloofar Jalali, Diana Sherifali, Plinio Morita, Joseph Cafazzo. . . . 4296

COVID-19 Public Health Communication on X (Formerly Twitter): Cross-Sectional Study of Message Type, Sentiment, and Source (e59687)
Sana Parveen, Agustin Pereira, Nathaly Garzon-Orjuela, Patricia McHugh, Aswathi Surendran, Heike Vornhagen, Akke Vellinga. . . . 4307

Medical Misinformation in AI-Assisted Self-Diagnosis: Development of a Method (EvalPrompt) for Analyzing Large Language Models (e66207)
Troy Zada, Natalie Tam, Francois Barnard, Marlize Van Sittert, Venkat Bhat, Sirisha Rambhatla. . . . 4328

YouTube User Traffic to Paired Epilepsy Education Videos in English and Spanish: Comparative Study (e56720)
Luna Varela, Stephanie Horton, Ahmed Abdelmoity, Jean-Baptiste Le Pichon, Mark Hoffman. . . . 4339

Privacy Concerns Versus Personalized Health Content—Pregnant Individuals' Willingness to Share Personal Health Information on Social Media: Survey Study (e60862)
Haijing Hao, Yang Lee, Marianne Sharko, Qilu Li, Yiye Zhang. . . . 4352

Instagram Posts Promoting Colorectal Cancer Awareness: Content Analysis of Themes and Engagement During Colorectal Cancer Awareness Month (e63344)
Aditi Srivastava, Jim Stimpson. . . . 4362

Quality Assessment of Medical Institutions' Websites Regarding Prescription Drug Misuse of Glucagon-Like Peptide-1 Receptor Agonists by Off-Label Use for Weight Loss: Website Evaluation Study (e68792)
Rie Oyama, Tsuyoshi Okuhara, Emi Furukawa, Hiroko Okada, Takahiro Kiuchi. . . . 4370

Public Health Discussions on Social Media: Evaluating Automated Sentiment Analysis Methods (e57395)
Lisa Gandy, Lana Ivanitskaya, Leeza Bacon, Rodina Bizri-Baryak. . . . 4384

Understanding Health-Related Discussions on Reddit: Development of a Topic Assignment Method and Exploratory Analysis (e55309)
Garrett Chan, Mark Fung, Jill Warrington, Sarah Nowak. . . . 4398

Person-Specific Analyses of Smartphone Use and Mental Health: Intensive Longitudinal Study (e59875)
Merve Cerit, Angela Lee, Jeffrey Hancock, Adam Miner, Mu-Jung Cho, Daniel Muise, Anna-Angelina Garròn Torres, Nick Haber, Nilam Ram, Thomas Robinson, Byron Reeves. . . . 4420

XSL•FO
RenderX

Short Papers

Validation of Sleep Measurements of an Actigraphy Watch: Instrument Validation Study (e63529)
Mari Waki, Ryohei Nakada, Kayo Waki, Yuki Ban, Ryo Suzuki, Toshimasa Yamauchi, Masaomi Nangaku, Kazuhiko Ohe. 1243

CoVimmune COVID-19 Immunity Calculator: Web Application Development and Validation Study (e59467)
Rebecca Slotkin, Tassos Kyriakides, Vinni Yu, Xien Chen, Anupam Kundu, Shaili Gupta. 3047

Identifying the Minimal Clinically Important Difference in Emotion Regulation Among Youth Using the JoyPop App: Survey Study (e64483)
Jaidyn Charlton, Ishaq Malik, Angela Ashley, Amanda Newton, Elaine Toombs, Fred Schmidt, Janine Olthuis, Kristine Stasiuk, Tina Bobinski, Aislin Mushquash. 4461

Assessing the Impact of a Virtual Reality Cognitive Intervention on Tennis Performance in Junior Tennis Players: Pilot Study (e66979)
Joaquin Anguera, Aleem Choudhry, Michael Seaman, Dominick Fedele. 4468

Protocol

Incorporating Patient-Reported Outcome Measures and Patient-Reported Experience Measures in Addiction Treatment Services in Belgium: Naturalistic, Longitudinal, Multicenter Cohort Study (e65686)
Amine Zerrouk, Charlotte Migchels, Clara De Ruysscher, Kim Fernandez, Jerome Antoine, Florian De Meyer, Frieda Matthys, Wim van den Brink, Cleo Crunelle, Wouter Vanderplasschen. 3659

Viewpoints

Applications of Self-Driving Vehicles in an Aging Population (e66180)
Sara Shu, Benjamin Woo. 4227

Assessing the Current Limitations of Large Language Models in Advancing Health Care Education (e51319)
JaeYong Kim, Bathri Vajravelu. 4232

Promoting Digital Health Data Literacy: The Datum Project (e60832)
Daniel Powell, Laiba Asad, Elissa Zavaglia, Manuela Ferrari. 4246

Development of the Big Ten Academic Alliance Collaborative for Women in Medicine and Biomedical Science: “We Built the Airplane While Flying It” (e65561)
Maya Iyer, Aubrey Moe, Susan Massick, Jessica Davis, Megan Ballinger, Kristy Townsend. 4252

Applying Critical Discourse Analysis to Cross-Cultural Mental Health Recovery Research (e64087)
Yasuhiro Kotera, Riddhi Daryanani, Oliver Skipper, Jonathan Simpson, Simran Takhi, Merly McPhilbin, Benjamin-Rose Ingall, Mariam Namasaba, Jessica Jepps, Vanessa Kellermann, Divya Bhandari, Yasutaka Ojio, Amy Ronaldson, Estefania Guerrero, Tesnime Jebara, Claire Henderson, Mike Slade, Sara Vilar-Lluch. 4263

Are Dating App Algorithms Making Men Lonely and Does This Present a Public Health Concern? (e70594)
Eric Balki. 4270

XSL•FO
RenderX

Corrigenda and Addendas

Correction: Latino Parents' Reactions to and Engagement With a Facebook Group–Based COVID-19 Vaccine Promotion Intervention: Mixed Methods Pilot Study (e76107)
Anna González-Salinas, Elizabeth Andrade, Lorien Abroms, Kaitlyn Gómez, Carla Favetto, Valeria Gómez, Karen Collins. 4321

Correction: Evaluating Bard Gemini Pro and GPT-4 Vision Against Student Performance in Medical Visual Question Answering: Comparative Case Study (e71664)
Jonas Roos, Ron Martin, Robert Kaczmarczyk. 4323

Correction: Acceptability and Initial Adoption of the Heart Observation App for Infants With Congenital Heart Disease: Qualitative Study (e70280)
Elin Hjorth Johansen, Elin Børøsund, Ingeborg Martinsen Østen, Henrik Holmstrøm, Anne Moen. 4324

Correction: Assessing and Improving Data Integrity in Web-Based Surveys: Comparison of Fraud Detection Systems in a COVID-19 Study (e76462)
Stephen Bonett, Willey Lin, Patrina Sexton Topper, James Wolfe, Jesse Golinkoff, Aayushi Deshpande, Antonia Villarruel, José Bauermeister.
4 3 2 6

Research Letters

The AI Reviewer: Evaluating AI's Role in Citation Screening for Streamlined Systematic Reviews (e58366)
Jamie Ghossein, Brett Hryciw, Tim Ramsay, Kwadwo Kyeremanteng. 4436

Engaging Stakeholders With Professional or Lived Experience to Improve Firearm Violence Lexicon Development (e68105)
Nicole Cook, Frances Biel, Kerry Bet, Marion Sills, Ali Al Bataineh, Pedro Rivera, Anna Templeton, Natalie Cartwright. 4439

Guideline-Incorporated Large Language Model-Driven Evaluation of Medical Records Using MedCheckLLM (e53335)
Marc Schubert, Stella Soyka, Wolfgang Wick, Varun Venkataramani. 4445

Exploring Public Sentiment on the Repurposing of Ivermectin for COVID-19 Treatment: Cross-Sectional Study Using Twitter Data (e50536)
Angga Kautsar, Rano Sinuraya, Jurjen van der Schans, Maarten Postma, Auliya Suwantika. 4449

Case Report

Monitoring Nystagmus in a Patient With Vertigo Using a Commercial Mini-Infrared Camera and 3D Printer: Cost-Effectiveness Evaluation and Case Report (e70015)
Hiroyuki Sakazaki, Masao Noda, Yumi Dobashi, Tatsuaki Kuroda, Reiko Tsunoda, Hiroaki Fushiki. 4453

XSL•FO
RenderX

Evaluating Feasibility and Acceptability of the "My HeartHELP" Mobile App for Promoting Heart-Healthy Lifestyle Behaviors: Mixed Methods Study

Jina Choo[1,2], PhD, DrPH; Songwhi Noh[1,2], MSN; Yura Shin[1,2], MSN

[1]College of Nursing, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul, Republic of Korea

[2]Transdisciplinary Major in Learning Health Systems, Department of Health Sciences, Korea University Graduate School, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul, Republic of Korea

Corresponding Author:
Jina Choo, PhD, DrPH
College of Nursing, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul, Republic of Korea

Abstract

Background: Few mobile apps have strategies for self-monitoring multiple heart-healthy behaviors simultaneously, as well as automated and tailored feedback on individual behavioral outcomes for cardiovascular health. An app named "My HeartHELP" was developed for the general adult population to promote 6 heart-healthy lifestyle behaviors—physical activity, nonsedentary behaviors, healthy eating behaviors, nonsmoking, no alcohol binge drinking, and self-assessment of body weight. Three behavioral strategies were used: (1) text messaging the users for information on cardiovascular health, (2) self-monitoring of 6 heart-healthy behaviors to fill out the blanks of behavioral items, and (3) automated and tailored feedback messaging to users for behavioral outcomes obtained from self-monitoring.

Objectives: This study aimed to evaluate the feasibility and acceptability of the "My HeartHELP" app.

Methods: The participants were 29 community residents in Seoul, South Korea, who met at least 1 criterion of metabolic syndrome. To evaluate the feasibility, we assessed 3 records, which are as follows: First, the "record for self-monitoring" was determined as feasible if an average percentage for each of the 6 behaviors over 4 weeks was 75% or higher based on percentages of participants who completed to record each of 6 heart-healthy behaviors. Second, the "record for access to the app" was determined as feasible if users accessed at least once a day on average per week. Third, "records for behavioral changes" over 4 weeks were collected via a self-reported questionnaire. To evaluate acceptability, we used an assessment tool comprising 12 items that included subscales for comprehensibility, ease, health benefits, technical completeness, overall satisfaction, and recommendation to others on a 5-point Likert scale. Acceptability was determined as acceptable if the average scores for the total scale and each subscale were 3.5 points or greater. Second, qualitative data were collected through 2 focus groups, each consisting of 14 or 15 participants. All data were collected in June and July 2022.

Results: During the 4 weeks, 95.6% (range: 85.8%-97.4%) of the participants adhered to more than 75% of "completion of daily self-monitoring of each heart-healthy behavior," having met the criterion. The participants accessed the app on average 1.8 (SD 1.70) times per day, meeting the criteria. Participants had positive behavioral changes in all 6 behaviors, of which nonsedentary behavior (10%-28%; χ^2_1=1.76; *P*<.001) and non–fast-food intake were especially statistically significant (72%-93%; χ^2_1=5.64; *P*=.03) over 4 weeks. Participants reported 3.8 points for a total score of acceptability and more than 3.5 points for all subscales, which met the criterion. Qualitative data obtained from focus groups indicated that automated and tailored feedback messages motivated participants to promote healthy lifestyles.

Conclusions: The "My HeartHELP" app may be a feasible and acceptable mobile app to promote self-monitoring and possibly behavioral changes in heart-healthy lifestyle behaviors.

(JMIR Form Res 2025;9:e66108) doi:10.2196/66108

KEYWORDS

healthy lifestyle; primary prevention; cardiovascular diseases; community health nursing; mobile app

Introduction

Cardiovascular disease is preventable through the promotion of heart-healthy behaviors, such as physical activity, nonsmoking, nonsedentary, and healthy eating behaviors, refraining from alcohol binge drinking, and self-assessment of body weight [1,2]. Each additional behavioral modification may reduce the risk of cardiovascular disease [1,3]. However, only a small percentage of the population is known to adhere to these

XSL•FO
RenderX

behaviors [4,5]. In this context, strategies are needed to effectively change multiple heart-healthy behaviors that are accessible to everyone, even before the onset of heart disease, and in community settings.

An app of mobile health technologies that allow community-dwelling individuals to monitor their own health behaviors and promote health behaviors in their daily lives may be needed to effectively change heart-healthy behaviors [6]. Mobile apps have the advantages of self-directedness, instant connectivity, ubiquity, personalization, and interactive learning through social engagement [7], which may enhance self-monitoring and practice of health behaviors without time and space constraints. However, few mobile apps are equipped with strategies for a 1-stop approach to monitor multiple heart-healthy behaviors that may correlate with each other [8] and target individuals without cardiovascular disease in a community setting. However, evidence indicating that mobile health technologies can effectively improve heart-healthy behaviors is lacking. This may be due to a lack of empathy and ineffective educational effects through active and persistent interactions with cardiovascular experts [9]. In this context, mobile technologies should be designed to enhance the self-monitoring of their own behaviors and interactions between mobile app users and experts in the context of cardiovascular prevention and promotion.

Several mobile apps have been developed to monitor health behavior changes [10,11] in the context of disease management for hypertension, diabetes, cardiovascular disease, cancer, and obesity [12-15]. However, the monitoring strategies used in each study were inconsistent [16]. In addition, the design of such mobile apps for health behavioral interventions has been confined to the process of self-reporting individual health behaviors in the form of a checklist [17]. Finally, few health-promoting mobile apps have automated feedback text messages for individual behavioral outcomes achieved with the self-monitoring of multiple lifestyle behaviors in the context of cardiovascular health.

The Information-Motivation-Behavioral Skills model (IMB model) is widely used as a theoretical framework for predicting health behavior changes [18]. The IMB model addresses the core constructs of information, motivation, and behavioral skills as prerequisites for health behavior change and assumes that when an individual acquires sufficient information related to health behavior, is motivated to change behavior, and improves behavioral skills, their behavior change and maintenance are promoted, which in turn can improve health levels [18]. In this theoretical background, we designed a mobile app to be embedded with behavioral strategies to encourage individuals to be informed, motivated, and skilled in behavior change by using cognitive-behavioral strategies addressed by the American Heart Association [19].

One such app is the “My HeartHELP” app, which was developed for the general adult population to promote heart-healthy lifestyle behaviors based on the constructs of information, motivation, and behavioral skill in the IMB model [20]. The “My HeartHELP” app is an abbreviation derived from the Heart-Healthy Lifestyle-Promoting app [21]. This app targets 6 heart-healthy lifestyle behaviors, such as physical activity, nonsedentary behaviors, healthy eating behaviors, nonsmoking, refraining from alcohol binge drinking, and self-assessment of body weight [22,23]. Of the 6 behaviors, 5 were selected based on the rationale that the US Centers for Disease Control and Prevention emphasizes key preventive behaviors for heart health, including healthy food and beverage choices, maintaining a healthy weight, engaging in regular physical activity, and nonsmoking [23]. An additional behavior, nonsedentary behavior, was incorporated due to the established association between sedentary behavior and increased cardiovascular risk [1]. The “My HeartHELP” app was designed with three behavioral strategies : (1) text messaging users for information on cardiovascular health, (2) self-monitoring of 6 lifestyles, and (3) automated and tailored feedback text messaging to users for behavioral outcomes obtained from self-monitoring. To increase the construct of information in the IMB model, texts on knowledge about cardiovascular prevention [24] was provided via text messaging. To increase the constructs of motivation and behavioral skills in the IMB model, goal setting [25,26], self-reflection via self-monitoring of behaviors [27], and positive reinforcement and tailored real-time feedback of behavioral performance via text messages [28,29] were installed in the “My HeartHELP” app. In particular, providing feedback is a well-established strategy in behavioral interventions. A Cochrane review underscored its importance in facilitating behavior change, particularly in contexts where substantial improvement is needed [30]. This aligns with the American Heart Association’s cognitive-behavioral strategies, which also emphasize feedback as a key component in sustaining behavioral improvements [19].

In this context, we aimed to assess the feasibility and acceptability of the “My HeartHELP” app in this study to conduct a future randomized controlled trial (ISRCTN83643383).

Methods

Study Design and Sample

This study used a mixed methods design [31] with a single-group approach to evaluate the feasibility and acceptability of the “My HeartHELP” app using both quantitative and qualitative data. This study design was chosen to incorporate qualitative data, in addition to quantitative data, to facilitate a more comprehensive evaluation of app users’ perceptions of its acceptability, including its health benefits and overall usefulness. The app was implemented over a 4-week period. Data were collected at both pretest and posttest stages. At each stage, a questionnaire was administered, and at the posttest stage, focus group discussions were conducted.

The participants were 29 adults. They were recruited by the following ways: First, we posted a recruitment notice on the bulletin boards of the university affiliated with the researchers. Second, we additionally posted it on a community-based mobile commercial and life platform (ie, Karrot app). In this way, the Karrot automatically displays the notice to residents in 174 neighborhoods across 5 municipal districts surrounding the university in Seoul, South Korea. Individuals interested in

XSL•FO
RenderX

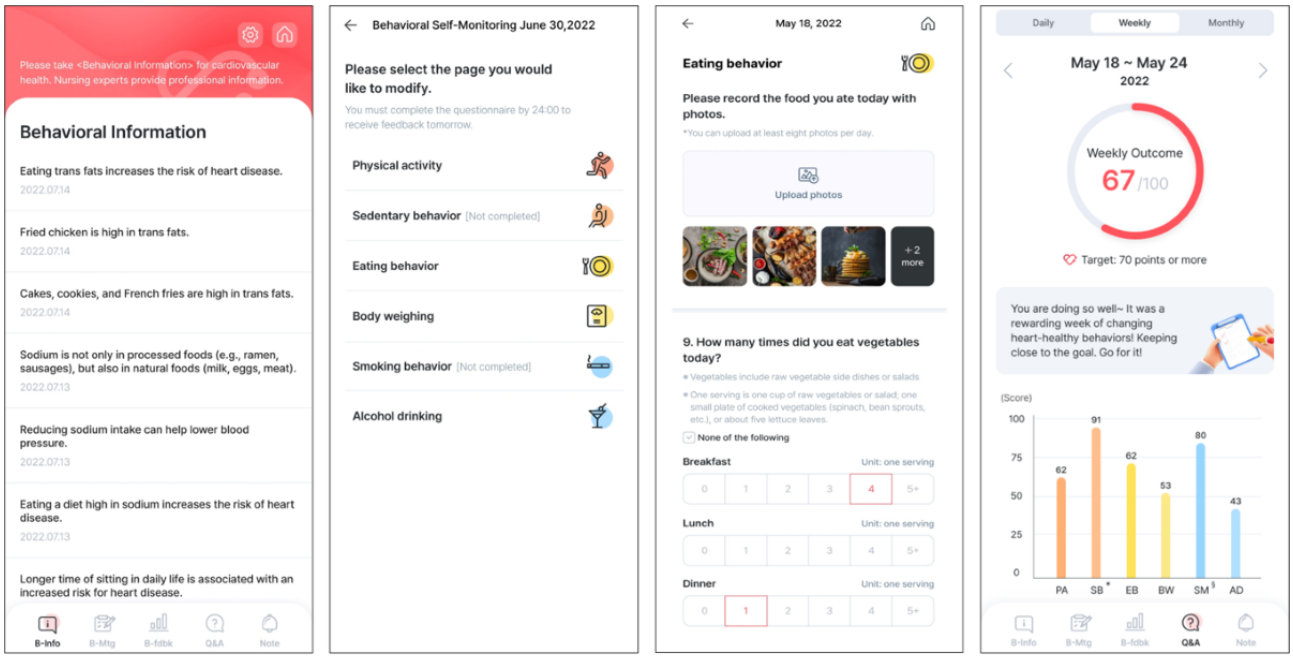
participating in the study contacted the researchers via mobile communication platforms (eg, the KakaoTalk app) or by phone, after which an eligibility screening was conducted based on the following criteria. The inclusion criteria were adults (1) aged between 20 and 64 years who had at least one of the metabolic syndrome components (ie, abdominal obesity, hypertension, dyslipidemia, or glucose intolerance). The exclusion criteria were individuals diagnosed with (1) cardiovascular disease and undergoing treatment, (2) major depression or anxiety disorder, (3) physical impairment, (4) cognitive impairment, and (5) difficulties in using a mobile app.

The sample size was determined to be 30 adults based on a study by Symer et al [32] that tested the validity of a mobile app. Of the 30 participants, 1 lost contact during the study period of 4 weeks and 29 were included in the final analysis of quantitative and qualitative data.

Overview of the "My HeartHELP" App

The "My HeartHELP" app was designed with 3 behavioral strategies, that is, behavioral information, behavioral self-monitoring, and behavioral performance/feedback. The structure and content of the app are described in detail in Figure 1 and Table 1.

Figure 1. Overview of the "My HeartHELP" app. This mobile app was developed in Korean and designed based on 3 behavioral strategies: behavioral information, behavioral self-monitoring, and behavioral performance/feedback. In this study, the app was applied over a 4-week period to 29 participants residing in Seoul, South Korea, who were at risk for cardiovascular disease. Pre- and posttests were performed to evaluate the feasibility and acceptability of the app.



XSL•FO
RenderX

Table . The structure of the "My HeartHELP" app: its strategies and contents.

Strategies	Contents
Behavioral information	• Behavioral knowledge • A total of 43 text messages for symptoms and characteristics of cardiovascular disease, cardiovascular risk factors, and heart-healthy behaviors • 8 - 11 texts delivered per week to participants • From Monday to Friday at 1:00 PM over 4 weeks
Behavioral self-monitoring	• Ensuring that the participants recorded the extent to which they practiced each of the 6 heart-healthy behaviors during the entire day • Physical activity (steps of walking and types and minutes of moderate-intensity exercise) • Sedentary behavior (time, hours) • Eating behavior (fruit and vegetable, sugar-sweetened beverage, fast foods, foods high in saturated fat or trans fat, whole grain, and sodium intake) • Self-measurement of body weight (kg per day) • Smoking behavior (cigarettes per day) • Alcohol-drinking behavior (drinks per day)
Behavioral performance/feedback	• Interface displays according to the level of heart-healthy behavior achieved over the entire day, 7 days (a week), and 30 days (a month) • Based on the algorithm for behavioral target goals that the authors had installed for each of 6 heart-healthy behaviors • Automated and tailored feedback text messaging • Daily at 9:30 AM from Monday to Friday • Weekly at 9:30 AM on Monday • Monthly at 9:30 AM on the first Tuesday • Example: "You did a good job today" or "You almost reached your goal today. Just a little more effort, and you will reach your goal. Cheer up!"

Behavioral Information

"Behavioral information" refers to text messaging of heart-health–related information on the characteristics of cardiovascular diseases and heart-health–related behaviors (Figure 1 and Table 1). Text messages were delivered as push notifications 5 times per week at 1:00 PM over a course of 4 weeks.

Behavioral Self-Monitoring

"Behavioral self-monitoring" refers to the participants' recording the extent to which they achieved 6 heart-healthy behaviors (Figure 1 and Table 1). Physical activity was recorded into 2 types: daily step counts for walking and daily time duration for moderate-intensity aerobic exercise; sedentary behavior recorded screen time (hours); and eating behavior recorded frequencies and types of whole grain intake, fruit and vegetable intake, sugar-sweetened beverage intake, sodium intake, fast-food intake, and intake of foods high in saturated fat. The participants were instructed to weigh body weights to measure and record. Participants were asked to daily record whether they smoked cigarettes and, if so, how many cigarettes they smoked. The participants were requested to record the frequency of alcohol consumption with a standard glass [33,34].

Behavioral Performance/Feedback

"Behavioral performance/feedback" refers to automatized and tailored interface displays including diagram and texts (Figure 1 and Table 1) according to outcomes adhered to 6 heart-healthy behaviors achieved over 1 day, 7 days (a week), and 30 days (a month), based on the algorithm for behavioral target goals that the authors had installed for the 6 heart-healthy behaviors.

The frequency of feedback delivery was structured as daily, weekly, and monthly, based on the rationale that frequent feedback may enhance motivation, self-monitoring, and goal attainment for behavioral change [35], and that daily, weekly, and monthly self-monitoring may support positive health outcomes [35-37]. Push notifications were delivered to request access to the interface display.

Daily outcomes are displayed as achieved scores out of 10 points (eg, 6/10) (Figure 1), weekly outcomes as achieved scores out of 100 points and bar graphs per behavior (Figure 1), and monthly outcomes as achieved scores out of 100 points and the symbol of the heart if achieved. Specifically, the daily outcomes were obtained by achieved scores from the components of 6 behaviors. Physical activity was evaluated by awarding 1 point for achieving a minimum of 8000 steps per day [38] or engaging in at least 30 minutes of moderate-intensity exercise per day [39]. Nonsedentary behavior was assessed by awarding 1 point for limiting sedentary time to less than 2 hours per day [40]. Eating behaviors were evaluated across multiple dietary components. Fruit and vegetable intake was assessed by awarding 1 point for consuming a minimum of 5 servings per day [41]. Similarly, 1 point was awarded for abstaining from sugar-sweetened beverages [42], avoiding fast-food consumption [43], refraining from consuming foods high in trans fats [44], and maintaining sodium intake below 1500 mg per day [45]. Self-monitoring of body weight was assessed by awarding 1 point for weighing oneself daily [46,47]. Smoking behavior was

XSL•FO
RenderX

evaluated by awarding 1 point for complete abstinence from smoking [48]. Finally, alcohol consumption was assessed by awarding 1 point for limiting intake to fewer than 5 drinks per day for men and fewer than 4 drinks per day for women [49].

Weekly outcomes were derived from the average scores obtained over 7 days, which were then converted to a 100-point scale and categorized into 3 ranges: ≥70, 50 - 69, and <50. Feedback texts were generated based on these score ranges. Similarly, a monthly outcome was calculated from the average scores over the previous 4 weeks, converted to a 100-point scale, and classified into 2 categories: ≥70 and <70. Corresponding feedback texts were generated accordingly.

To account for users' baseline characteristics and minimize redundant data entry, participants who self-identified as nonsmokers and nondrinkers were provided with a tailored interface that excluded smoking- and alcohol-related inputs. In addition, they received automated feedback reinforcing their adherence to these behaviors.

Study Procedure and Data Collection

The study procedure and data collection were conducted in 3 sequential stages between June and July 2022 in Seoul, South Korea. Specifically, the pretest was conducted on June 16 - 17 (stage 1), followed by the 4-week implementation of the "My HeartHELP" app from June 20 to July 17 (stage 2). The posttest and focus group sessions took place on July 21 - 22 (stage 3).

Stage 1: The pretest took place on the same day using a self-reported questionnaire on general characteristics and heart-healthy behaviors. After the pretest, the pre-education session for informing the usage of the "My HeartHELP" app was conducted as a face-to-face group session. At this session, the principal investigator (JC) presented a lecture to inform the characteristics of heart disease and the importance of promoting heart-healthy behaviors for 20 minutes. This lecture aimed to provide background knowledge on the variables measured by the mobile app and to facilitate understanding of the app's usage. Following the lecture, the researchers provided a detailed explanation of the structure and usage of the "My HeartHELP" app, along with an overview of the overall research process. Stage 2: The implementation of the "My HeartHELP" app took place for 4 weeks without contact with the researchers in between. Stage 3: Postsurvey and focus group session took place on the same day after the end of the 4-week implementation. The postsurvey session was administered using the self-reported questionnaire with heart-healthy behaviors (ie, repeated measures) and the acceptability of the app. Focus group sessions, each lasting 60 minutes, were conducted with 2 groups comprising of 14 - 15 participants per group to assess the acceptability of "My HeartHELP" app. The principal investigator (JC) facilitated the discussion by using open-ended questions, probing questions, and laddering techniques [50] to encourage participants to freely express their thoughts and provide in-depth feedback. Key topic areas included overall impressions, strengths and weaknesses, comprehensibility, perceived health benefits, and usability (ie, ease of use and convenience in navigation and operation). In addition, challenges such as system errors, data input difficulties, and technical issues were explored. Participants also suggested improvements related to system errors, data input difficulties, and readability concerns. All sessions were audio-recorded and transcribed verbatim for analysis.

Study Variables and Metrics

General characteristics included sex, age, college-educated status, employment status, exposure to heart health education, history of chronic diseases (ie, hypertension, diabetes mellitus, or dyslipidemia), and self-reported height and body weight.

The Feasibility of "My HeartHELP"

To evaluate the feasibility of "My HeartHELP," we assessed the following 3 behavioral records: First, the "record for self-monitoring" was determined based on the data of the percentage of participants who completed recording each of 11 behavioral components (Figure 1) of the 6 heart-healthy behaviors per day, and their average percentage over 4 weeks was assessed. We determined that this was feasible if the average percentage was 75% or higher [32]. Second, the "record for access to the app" was determined based on the data of a mean of daily access counts per week. Next, we determined whether it was feasible for the users to access the app at least once a day on average per week. Third, the "record for behavioral changes" was determined as feasible if participants showed improvements in each of the 6 behaviors when comparing pretest and posttest. Among the 6 behaviors, the following behaviors were included in the analysis : Physical activity (2 items) was assessed as moderate-intensity aerobic exercise for 150 minutes per week [39]. Walking (2 item) was also assessed for 150 minutes per week [38]. Nonsedentary behavior (1 item) was assessed daily, defined as having less than 2 hours of screen time (screen time included activities such as watching TV and using a mobile phone, computer, or tablet, excluding work- or study-related activities) [40]. The intake of fruits and vegetables (1 item) was assessed as 2 or more servings per day [41]. Non—sugar-sweetened beverage intake (1 item) was assessed none per day [42]. Non—fast-food intake (1 item) was assessed as consumption less than 2 times per week [43]. Whole grain intake (1 item) was assessed as consuming 2 or more servings per day, where 1 serving equals 1 slice of whole grain bread [51]. High-risk alcohol consumption (2 items) was defined as 5 or more drinks per session for men (including 4 or more drinks per session for women) [49]. Maintaining healthy body weight (1 item) was assessed by having participants report their current weight.

The Acceptability of "My HeartHELP"

To evaluate the acceptability of "My HeartHELP," we used quantitative data from an assessment tool and qualitative data from focus groups. First, we developed the assessment tool with 12 items that consisted of the following subscales: comprehensibility (2 items), ease (2 items), health benefits (5 items), technical completeness (1 item), overall satisfaction (1 item), and recommendation to others (1 item). Health benefits were categorized into "behavioral information," "behavioral monitoring," and "behavioral performance/feedback." The items were scored on a 5-point Likert scale (1=not at all, 2=quite not, 3=somewhat, and 4=very much). The subscales were considered acceptable if the mean score for each subscale was 3.5 points

XSL•FO
RenderX

or greater. Second, qualitative data obtained from focus groups were used. The following questions [52] were addressed: "What is your overall impression of the My HeartHELP?" "Please tell us three pros and three cons of the My HeartHELP." "Please tell us about My HeartHELP based on (1) comprehensibility (2) health benefits (3) usability (4) challenges." The qualitative data were considered acceptable if researchers identified meaningful quotations highlighting the benefits and usefulness of the "My HeartHELP" app in facilitating positive heart-healthy behavioral changes and improving health outcomes. While the quantitative data provided structured acceptability scores, they were limited in capturing the depth of users' perceptions and the specific challenges encountered during app usage. To address this limitation, qualitative findings from focus group discussions were integrated, offering a more nuanced understanding of user perceptions regarding comprehensibility, perceived benefits, and challenges. This approach enabled a more comprehensive evaluation that extends beyond numerical assessments alone.

Data Analysis

SPSS/WIN (version 28.0; SPSS, Inc) was used [53]. The significance level was set at less than .05. Participants' baseline measures (ie, general characteristics and heart-healthy behaviors) were analyzed using descriptive statistics (frequencies, percentages, means, and SDs). To compare the changes in heart-healthy behaviors between pre- and postsurvey, McNemar's chi-square test for dichotomous variables and the 2-tailed paired *t* test for continuous variables were conducted. The measures of feasibility and acceptability were analyzed using descriptive statistics for the quantitative data and content analysis for the qualitative data. Qualitative data obtained from focus groups were analyzed using Graneheim and Lundman's qualitative content analysis [54]. The interview recordings were repeatedly reviewed, transcribed, and thoroughly examined to capture participants' perspectives comprehensively. The transcribed data were then segmented into meaningful units, systematically coded, and organized into categories and subcategories to ensure a structured interpretation. Continuous comparisons were conducted across data points to refine themes and establish thematic classifications. The 2 sets of qualitative data were independently analyzed by 3 authors (JC, SN, and YS) and consolidated into meaningful themes and categories [55,56].

Ethical Considerations

This study was approved by the institutional review board of Korea University (KUIRB-2022-0147-01). All participants provided written informed consent for their participation in the study, including permission for audio recording during the focus group discussions. All procedures were performed in accordance with the ethical standards of the institutional research committee and the 2013 Declaration of Helsinki. To ensure the privacy and confidentiality of participants, all data were deidentified and securely stored on encrypted servers, with access restricted to authorized members of the research team. Participants who successfully completed all stages of the study received a monetary compensation of KR ₩200,000 (US $137) [57]. Participants who completed only the pretest received a KR ₩10,000 (US $7) gift voucher. All details related to compensation were fully disclosed to participants during the informed consent process, prior to their enrollment in the study.

Results

Participant Characteristics

The participants (N=29) had a mean age of 41.1 (SD 13.03) years (Table 2). Of all the participants, 62% (18/29) were women; 79% (23/29) were college-educated; 66% (19/29) were employed; 45% (13/29) had a previous history of hypertension, diabetes mellitus, or dyslipidemia; and 69% (20/29) fell into the category of more than 25 kg/m^2 of BMI. Moreover, 7% (2/29) of the participants had undergone a heart-health educational session.

XSL•FO
RenderX

Table . General characteristics of participants (N=29).

Participant characteristics	Values
Age (years), mean (SD)	41.1 (13.03)
Gender, n (%)	
Women	18 (62)
Men	11 (38)
College-educated, n (%)	
Yes	23 (79)
No	6 (21)
Employed, n (%)	
Yes	19 (66)
No	10 (34)
Having diseases[a], n (%)	
Yes	13 (45)
No	16 (55)
BMI, kg/m^2, n (%)	
≥25	20 (69)
<25	9 (31)
Exposed to heart-health educational sessions, n (%)	
Yes	2 (7)
No	27 (93)

[a]Having diseases indicate hypertension, diabetes mellitus, or dyslipidemia.

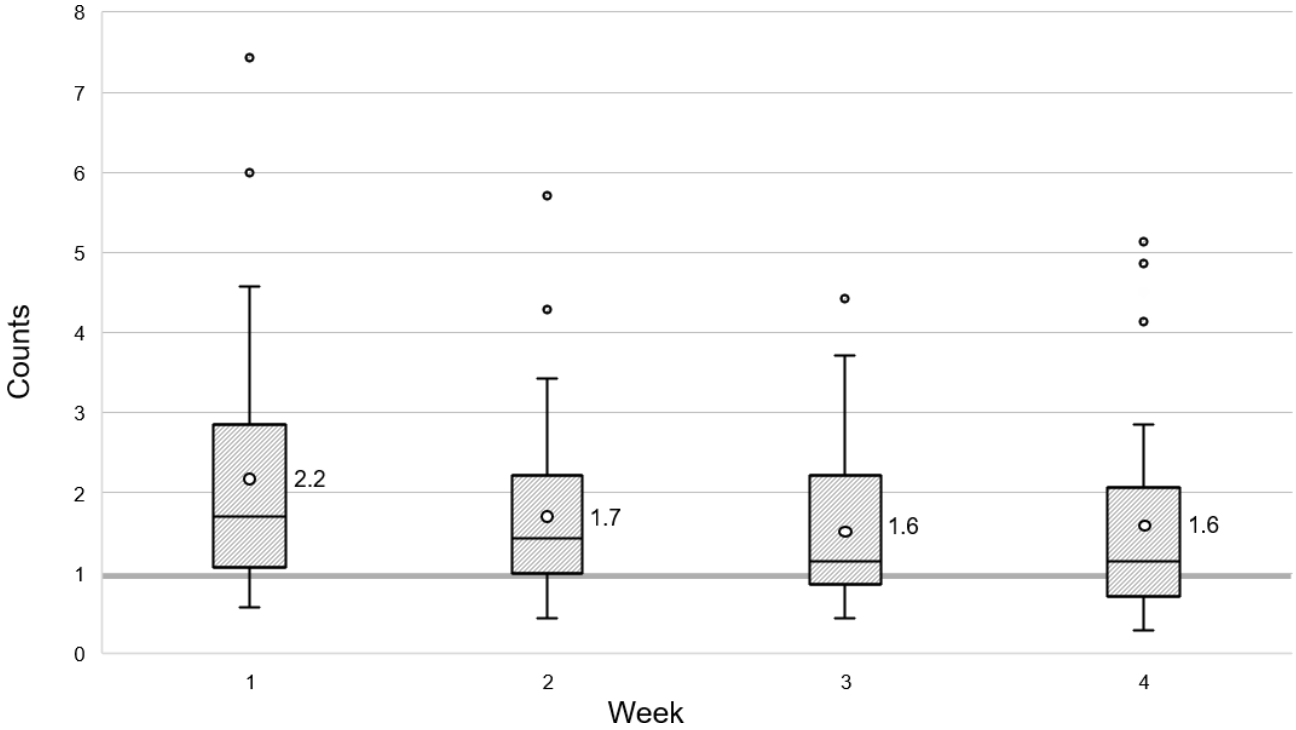
The Feasibility of "My HeartHELP"

The feasibility for the "record of self-monitoring" was determined to be feasible. The record of self-monitoring for 11 behavioral components of the 6 heart-healthy lifestyle behaviors is shown in Multimedia Appendix 1. The 4-week percentages of participants who adhered to more than 75% of the completion of daily self-monitoring for each heart-healthy behavior are reported in Multimedia Appendix 1. The average percentage was 95.6%, ranging from a minimum of 85.8% (recording body weight) to a maximum of 97.4% (recording physical activity).

The feasibility for "record of access to the app" was determined as feasible. Mean daily access counts per week showed 2.2 (SD 2.10) at week 1, 1.7 (SD 1.58) at week 2, 1.6 (SD 1.34) at week 3, and 1.6 (SD 1.62) at week 4. The grand mean was 1.8 (SD 1.70), which exceeded the feasibility criteria (Figure 2).

XSL•FO
RenderX

Figure 2. Feasibility of the "My HeartHELP" app: behavioral record of access to the app. The box plot represents the mean values of daily app access counts per week over a 4-week period based on data from 29 participants at risk for cardiovascular disease residing in Seoul, South Korea. The feasibility criterion was set at an average of at least 1 access per week.



The feasibility for "record for behavioral changes" is shown in Table 3. Numbers of participants for physical activity increased from 24% (7/29) to 34% (10/29) for sufficient moderate-intensity aerobic exercise, from 59% (17/29) to 62% (18/29) for sufficient walking exercise, and from 10% (3/29) to 28% (8/29) for nonsedentary behavior (χ^2_1=1.76; *P*<.001). Regarding eating behavior, numbers of participants increased from 7% (2/29) to 21% (6/29) for intake of fruits and vegetables, from 86% (25/29) to 90% (26/29) for the intake of non–sugar-sweetened beverages, from 72% (21/29) to 93% (27/29) for the intake of non–fast foods (χ^2_1=5.64; *P*=.03), and from 62% (18/29) to 66% (19/29) for intake of whole grains. The number of high-risk alcohol drinkers decreased from 38% (11/29) to 34% (10/29). Finally, body weight changed from 76.2 kg to 76.0 kg.

XSL•FO
RenderX

Table . Changes in heart-healthy behaviors after the application of "My HeartHELP" (N=29).

Characteristics	Pretest	Posttest	Chi-square or *t* test (*df*)	*P* value
Physical activity				
Sufficient moderate-intensity aerobic exercise, n (%)			10.72 (1)[a]	.38
Yes	7 (24)	10 (34)		
No	22 (76)	19 (66)		
Sufficient walking exercise (≥150 minutes), n (%)			14.78 (1)[a]	.99
Yes	17 (59)	18 (62)		
No	12 (41)	11 (38)		
Nonsedentary behavior (<2 hours/day), n (%)			1.76 (1)[a]	<.001
Yes	3 (10)	8 (28)		
No	26 (90)	21 (72)		
Fruit and vegetable intake (≥2 servings/day), n (%)			1.13 (1)[a]	.22
Yes	2 (7)	6 (21)		
No	27 (93)	23 (79)		
Non–sugar-sweetened beverage intake (less than once per day), n (%)			7.87 (1)[a]	.99
Yes	25 (86)	26 (90)		
No	4 (14)	3 (10)		
Non–fast foods (<2 servings/week), n (%)			5.64 (1)[a]	.03
Yes	21 (72)	27 (93)		
No	8 (28)	2 (7)		
Whole grains (≥2 servings/day), n (%)			11.47 (1)[a]	.99
Yes	18 (62)	19 (66)		
No	11 (38)	10 (34)		
High-risk alcohol drinking[b], n (%)			24.98 (1)[a]	.99
Yes	11 (38)	10 (34)		
No	18 (62)	19 (66)		
Body weight (kg), mean (SD)	76.2 (15.36)	76.0 (15.21)	1.22 (28)[c]	.23

[a]Chi-square test.

[b]High-risk drinking indicates ≥5 drinks per session for men (≥4 drinks per session for women).

[c]*t* test.

The Acceptability of "My HeartHELP"

Acceptability using quantitative data was determined as acceptable if the scores for all the subscales of acceptability were 3.5 points (the criteria of acceptability) or greater (Table 4). The scores ranged from a minimum of 3.6 points (health benefits for information, health benefits for performance/feedback, and technical completeness) to a maximum of 4.1 points (comprehensibility).

XSL•FO
RenderX

Table . Evaluation of the acceptability for the My HeartHELP (N=29).

	Values, mean (SD)
Total score of the acceptability	3.8 (0.61)
Comprehensibility	4.1 (0.81)
Ease	3.9 (0.84)
Health benefits	
Overall	3.8 (0.87)
Behavioral information	3.6 (1.02)
Behavioral monitoring	3.8 (0.90)
Behavioral performance/feedback	3.6 (0.70)
Completeness, technical	3.6 (0.87)
Satisfaction, overall	3.8 (0.73)
Recommendation to others	3.8 (0.87)

The acceptability using the qualitative data may be determined as acceptable, as shown in Table 5. Participants mentioned that reflecting on their own heart-healthy behaviors via self-monitoring, adjusting their own heart-healthy behaviors via receiving feedback messages, and becoming motivated for weight control were health benefits. The participants mentioned that easily understood phrases were acceptable in terms of the use of easy language in the app. However, the participants suggested the following challenges for the structure and interface: differentiating between good and bad behaviors by color, including options for a variety of dietary behavior options, and installing additional reminder alarms (Table 5).

XSL•FO
RenderX

Table . Themes and quotations extracted using focus group discussion (N=29).

Themes and categories	Quotations
Comprehensibility	
Understanding phrases easily	"Overall, I think it was easy to understand. The language used was everyday language that we encounter in our daily lives, so it was not difficult."
Benefits	
Reflecting on their own heart-healthy behaviors via self-monitoring	"I habitually consume large yogurt drinks several times a day. However, as I continued to record them under the sweetened beverage category, I realized that just two of these large yogurts exceeded my daily sugar intake limit, leading me to stop consuming them."
	"I used to drink alcohol quite often even when dining alone, but after seeing the records I entered, I found myself not ordering alcohol even when getting delivery food."
Adjusting their own heart-healthy behaviors via receiving feedback messages	"When I woke up each day, I first checked the daily achievement scores in 'My HeartHELP' After that, I started paying more attention to what I ate and definitely did more exercise throughout the day."
Getting motivated for weight control	"I'm maintaining my 3 kg weight loss because I have been using it continuously to motivate me to lose weight."
Challenges	
Differentiating good and bad behaviors by colors	"It would be good to distinguish unhealthy foods, like salty ones, by marking them with red or orange, and to mark healthy foods with yellow."
Including options of a variety of dietary behavior options	"I think it would have been better if I could enter other foods not listed. Or, it would be nice if there was an explanation that allowed me to include foods not presented based on my own judgment."
Installing additional reminder alarms	"I think it would be good if push notifications could come several times a day. Feedback messages arrive during the day, and I want to enter them all at once in the evening, but sometimes I forget and it is already past midnight."

Discussion

Principal Findings

We found that the "My HeartHELP" app may be feasible and acceptable for individuals at risk of cardiovascular disease to be informed of their cardiovascular health, to self-monitor their heart-healthy behaviors, and to adhere to heart-healthy behaviors. Moreover, our data revealed that positive changes in heart-healthy behaviors may be achieved using the "My HeartHELP" app when included as a strategy for heart-healthy behavioral lifestyle interventions.

The "My HeartHELP" app met the criteria of feasibility and acceptability for the validity test of a mobile app. The vast majority of participants (95.6%) completed daily self-monitoring of their heart-healthy behaviors and accessed the app more than once per day (1.8 times). Moreover, the participants perceived the "My HeartHELP" app as comprehensible, easy to use, beneficial to their health, highly usable, and technically complete. Overall, the participants were satisfied with the My HeartHELP app and agreed to recommend it to others. Meanwhile, monitoring one's own behavior may be directly linked to adherence to healthy behavior [58,59] through the behavioral actions of participants, such as opening the app and entering their daily behavioral outcome for the items. Logging food in real time on mobile devices appears to encourage individuals to become more aware of their food consumption, which may have supported their efforts to achieve dietary goals and weight loss [60,61]. In this context, the action of accessing the app itself may induce motivation to monitor one's own behavior, which can lead to self-awareness of one's own behavior and thus to desirable heart-healthy behavior [61].

Our findings revealed that heart-healthy behaviors changed positively over 4 weeks, although significant changes were not verified for all 6 behaviors. Mobile health interventions have

XSL•FO
RenderX

revealed inconsistent results for improving heart-healthy behaviors and have reported positive effects limited to a single behavior, that is, physical activity. In this regard, the "My HeartHELP" app is equipped with behavioral strategies that allow individuals to monitor and practice complex and multiple heart-healthy behaviors using a 1-stop approach. Six heart-healthy behaviors via the app were positively changed over the course of 4 weeks, with the limitation of a short duration. These positive results may be explained by installing behavioral strategies for promoting motivation and self-efficacy based on the IMB model [18], that is, goal setting, self-monitoring, and feedback provision. A systematic review reported that behavioral counseling had a small to moderate effect on behavioral interventions when tailored to advice, goal setting, and written material provision [62]. Therefore, we can determine the feasibility and acceptability of the "My HeartHELP" app as a critical strategy for heart-healthy behavioral lifestyle interventions.

We found that the health benefits scores (behavioral information and performance/feedback) and technical completeness score in the acceptability presented the lowest score (3.6 points each), although the acceptability criteria were met (3.5 points). Linked to this finding, qualitative data from the focus group revealed that the participants had requested an additional alarm that could remind them to practice heart-healthy behaviors. Moreover, direct interaction with experts is necessary beyond self-monitoring and the performance interface of healthy behavioral performance [63]. In the future, feedback text messaging in which experts can provide sympathy for and support individual behavioral performance is needed to initiate intervention studies. Finally, the participants mentioned that the information messages were perceived as not individualized to an individual's lifestyle and situation. This finding indicates that the use of mHealth alone may have limitations in providing information tailored to an individual's situation and that face-to-face education is needed to provide information tailored to specific individual needs.

Limitations

This study outlined a few limitations. First, the generalizability of the results to ethnic groups other than Koreans is limited. In addition, to apply it to more people, the app developed in Korean needs to be translated and developed in other languages for linguistic inclusiveness. Second, because it is based on the diet of Koreans, in terms of monitoring eating behavior, its scope of use may be limited. However, because the behavioral strategies of "My HeartHELP" are key skills for behavioral change, the application of "My HeartHELP" to other languages may be necessary. Third, although "My HeartHELP" establishes behavioral goal setting based on standard criteria for adults [38-43,49,51], future studies should further incorporate tailored goal-setting strategies that promote individuals' autonomy, as emphasized in self-determination theory. Enhancing autonomy in goal setting may foster intrinsic motivation, which is essential for sustaining long-term behavior change [29,64]. Fourth, the 4-week follow-up period may have been too short to capture sustained behavior changes, highlighting the need for longer validation studies. Finally, the use of focus group discussions instead of anonymous methods may introduce social desirability bias, as participants may adjust their feedback based on group dynamics rather than expressing independent perspectives [65].

Conclusions

"My HeartHELP," which includes information, self-monitoring, and real-time feedback messaging strategies that promote the practice of complex heart-healthy behaviors, could be feasible and acceptable as an app for promoting cardiovascular health and could contribute to achieving optimal heart health for populations at risk for cardiovascular disease in the community. Practically, community settings, unlike hospital settings, present challenges and limitations in providing intensive professional care for comprehensive heart-healthy lifestyle modification. Therefore, the utilization of the "My HeartHELP" app in health promotion programs aimed at cardiovascular disease prevention at community settings may offer a potential way to solve the limitations.

Acknowledgments

This research was supported by the National Research Foundation of Korea (NRF) grant from the Korean Government (NRF-2019R1A2C1004116) and NRF grant from the Ministry of Science and ICT, Korean Government (RS-2024 - 00336847).

Data Availability

The data analyzed in this study may be available from the corresponding author upon reasonable request.

Authors' Contributions

JC conceptualized and designed this study and prepared the original manuscript draft; JC, SN, and YS were involved in data collection; SN and YS performed the data curation and analyses; and JC provided expert consultation on data interpretation and helped draft the manuscript. All authors were involved in the review and revision of the manuscript and gave approval for the final version to be published.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Feasibility of the "My HeartHELP" app: Record of self-monitoring for 11 behavioral components of the 6 heart-healthy lifestyle behaviors. A total of 29 participants at risk for cardiovascular disease residing in Seoul, South Korea, took part in the study. The feasibility of the self-monitoring record was assessed based on participant adherence, which was measured by calculating the percentage of participants who adhered to at least 75% of daily self-monitoring completion for each heart-healthy behavior over 4 weeks.
[PPTX File, 48 KB - formative_v9i1e66108_app1.pptx]

Checklist 1
Good reporting of a mixed-methods study (GRAMMS) checklist.
[DOCX File, 18 KB - formative_v9i1e66108_app2.docx]

References

1. Lavie CJ, Ozemek C, Carbone S, Katzmarzyk PT, Blair SN. Sedentary behavior, exercise, and cardiovascular health. Circ Res 2019 Mar;124(5):799-815. [doi: 10.1161/CIRCRESAHA.118.312669] [Medline: 30817262]
2. Spring B, Ockene JK, Gidding SS, et al. Better population health through behavior change in adults: a call to action. Circulation 2013 Nov 5;128(19):2169-2176. [doi: 10.1161/01.cir.0000435173.25936.e1] [Medline: 24100544]
3. Barbaresko J, Rienks J, Nöthlings U. Lifestyle indices and cardiovascular disease risk: a meta-analysis. Am J Prev Med 2018 Oct;55(4):555-564. [doi: 10.1016/j.amepre.2018.04.046] [Medline: 30241617]
4. Nielsen JB, Leppin A, Gyrd-Hansen DE, Jarbøl DE, Søndergaard J, Larsen PV. Barriers to lifestyle changes for prevention of cardiovascular disease - a survey among 40-60-year old Danes. BMC Cardiovasc Disord 2017 Sep 12;17(1):245. [doi: 10.1186/s12872-017-0677-0] [Medline: 28899356]
5. van Trier TJ, Mohammadnia N, Snaterse M, Peters RJG, Jørstad HT, Bax WA. Lifestyle management to prevent atherosclerotic cardiovascular disease: evidence and challenges. Neth Heart J 2022 Jan;30(1):3-14. [doi: 10.1007/s12471-021-01642-y] [Medline: 34762283]
6. Gold N, Yau A, Rigby B, Dyke C, Remfry EA, Chadborn T. Effectiveness of digital interventions for reducing behavioral risks of cardiovascular disease in nonclinical adult populations: systematic review of reviews. J Med Internet Res 2021 May 14;23(5):e19688. [doi: 10.2196/19688] [Medline: 33988126]
7. Goundar MS, Kumar BA. The use of mobile learning applications in higher education institutes. Educ Inf Technol 2022 Jan;27(1):1213-1236. [doi: 10.1007/s10639-021-10611-2]
8. Noh S, Choo J. Gender-specific clustering of lifestyle behaviors and its impacts on cardiovascular health. J Cardiovasc Nurs 2024;39(4):E115-E125. [doi: 10.1097/JCN.0000000000001006] [Medline: 37249529]
9. Svavarsdóttir MH, Sigurðardóttir Á, Steinsbekk A. How to become an expert educator: a qualitative study on the view of health professionals with experience in patient education. BMC Med Educ 2015 May 13;15(1):87. [doi: 10.1186/s12909-015-0370-x] [Medline: 25966846]
10. Frederix I, Van Driessche N, Hansen D, et al. Increasing the medium-term clinical benefits of hospital-based cardiac rehabilitation by physical activity telemonitoring in coronary artery disease patients. Eur J Prev Cardiol 2015 Feb;22(2):150-158. [doi: 10.1177/2047487313514018] [Medline: 24249840]
11. Burke LE, Sereika SM, Parmanto B, et al. Effect of tailored, daily feedback with lifestyle self-monitoring on weight loss: the SMARTER randomized clinical trial. Obesity (Silver Spring) 2022 Jan;30(1):75-84. [doi: 10.1002/oby.23321] [Medline: 34898011]
12. Kumar N, Khunger M, Gupta A, Garg N. A content analysis of smartphone-based applications for hypertension management. J Am Soc Hypertens 2015 Feb;9(2):130-136. [doi: 10.1016/j.jash.2014.12.001] [Medline: 25660364]
13. García-Gómez JM, de la Torre-Díez I, Vicente J, Robles M, López-Coronado M, Rodrigues JJ. Analysis of mobile health applications for a broad spectrum of consumers: a user experience approach. Health Informatics J 2014 Mar;20(1):74-84. [doi: 10.1177/1460458213479598]
14. Pandey A, Hasan S, Dubey D, Sarangi S. Smartphone apps as a source of cancer information: changing trends in health information-seeking behavior. J Cancer Educ 2013 Mar;28(1):138-142. [doi: 10.1007/s13187-012-0446-9] [Medline: 23275239]
15. Ubhi HK, Kotz D, Michie S, et al. Comparative analysis of smoking cessation smartphone applications available in 2012 versus 2014. Addict Behav 2016 Jul;58:175-181. [doi: 10.1016/j.addbeh.2016.02.026] [Medline: 26950256]
16. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: a systematic review. J Telemed Telecare 2018 Jan;24(1):22-30. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
17. Widmer RJ, Collins NM, Collins CS, West CP, Lerman LO, Lerman A. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. Mayo Clin Proc 2015 Apr;90(4):469-480. [doi: 10.1016/j.mayocp.2014.12.026] [Medline: 25841251]

XSL•FO
RenderX

18. Fisher WA, Fisher JD, Harman J. The information-motivation-behavioral skills model: a general social psychological approach to understanding and promoting health behavior. In: Suls J, Wallston KA, editors. Social Psychological Foundations of Health and Illness: Blackwell Publishing; 2003:82-106. [doi: 10.1002/9780470753552.ch4]
19. Artinian NT, Fletcher GF, Mozaffarian D, et al. Interventions to promote physical activity and dietary lifestyle changes for cardiovascular risk factor reduction in adults: a scientific statement from the American Heart Association. Circulation 2010 Jul 27;122(4):406-441. [doi: 10.1161/CIR.0b013e3181e8edf1] [Medline: 20625115]
20. Choo J, Noh S, Shin Y. Abstract P640: evaluating feasibility and acceptability of “My HeartHELP” mobile application for promoting heart-healthy lifestyle behaviors. Circulation 2023 Feb 28;147(Suppl_1). [doi: 10.1161/circ.147.suppl_1.P640]
21. Choo J, Noh S, Shin Y. Abstract P147: a community-based, heart-healthy lifestyle-promoting program using hybrid behavioral strategies: study protocol of a three-group randomized controlled trial. Circulation 2023 Feb 28;147(Suppl 1). [doi: 10.1161/circ.147.suppl_1.P147]
22. Li Y, Pan A, Wang DD, et al. Impact of healthy lifestyle factors on life expectancies in the US population. Circulation 2018 Jul 24;138(4):345-355. [doi: 10.1161/CIRCULATIONAHA.117.032047] [Medline: 29712712]
23. Preventing heart disease. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/heart-disease/prevention/index.html [accessed 2025-04-09]
24. Choo J, Noh S, Moon J, Shin Y. Development and psychometric testing of the Heart-Healthy Information Questionnaire. Eur J Cardiovasc Nurs 2023 Apr 12;22(3):299-310. [doi: 10.1093/eurjcn/zvac055] [Medline: 35766172]
25. Mann T, de Ridder D, Fujita K. Self-regulation of health behavior: social psychological approaches to goal setting and goal striving. Health Psychol 2013 May;32(5):487-498. [doi: 10.1037/a0028533] [Medline: 23646832]
26. Miller CK, Bauman J. Goal setting: an integral component of effective diabetes care. Curr Diab Rep 2014 Aug;14(8):509. [doi: 10.1007/s11892-014-0509-x] [Medline: 24907180]
27. Western MJ, Standage M, Peacock OJ, Nightingale T, Thompson D. Supporting behavior change in sedentary adults via real-time multidimensional physical activity feedback: mixed methods randomized controlled trial. JMIR Form Res 2022 Mar 2;6(3):e26525. [doi: 10.2196/26525] [Medline: 35234658]
28. Prestwich A, Kellar I, Parker R, et al. How can self-efficacy be increased? Meta-analysis of dietary interventions. Health Psychol Rev 2014;8(3):270-285. [doi: 10.1080/17437199.2013.813729] [Medline: 25053214]
29. Self-determination theory. Wikipedia. 2025. URL: https://en.wikipedia.org/wiki/Self-determination_theory [accessed 2025-04-09]
30. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev 2012 Jun 13;2012(6):CD000259. [doi: 10.1002/14651858.CD000259.pub3] [Medline: 22696318]
31. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013 Dec;48(6 Pt 2):2134-2156. [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
32. Symer MM, Abelson JS, Milsom J, McClure B, Yeo HL. A mobile health application to track patients after gastrointestinal surgery: results from a pilot study. J Gastrointest Surg 2017 Sep;21(9):1500-1505. [doi: 10.1007/s11605-017-3482-2] [Medline: 28685388]
33. What is a standard drink? National Institute on Alcohol Abuse and Alcoholism. 2025. URL: https://www.niaaa.nih.gov/alcohols-effects-health/overview-alcohol-consumption/what-standard-drink [accessed 2025-04-09]
34. Guidelines for alcohol consumption in cancer prevention. National Cancer Center Korea. 2023. URL: https://www.cancer.go.kr/download.do?uuid=68379447-9a18-4ac0-a601-8e32ee918674.pdf [accessed 2025-04-09]
35. Ambeba EJ, Ye L, Sereika SM, et al. The use of mHealth to deliver tailored messages reduces reported energy and fat intake. J Cardiovasc Nurs 2015;30(1):35-43. [doi: 10.1097/JCN.0000000000000120] [Medline: 24434827]
36. Alshagrawi S, Abidi ST. Efficacy of an mHealth behavior change intervention for promoting physical activity in the workplace: randomized controlled trial. J Med Internet Res 2023 Apr 27;25:e44108. [doi: 10.2196/44108] [Medline: 37103981]
37. Elbert SP, Dijkstra A, Oenema A. A mobile phone app intervention targeting fruit and vegetable consumption: the efficacy of textual and auditory tailored health information tested in a randomized controlled trial. J Med Internet Res 2016 Jun 10;18(6):e147. [doi: 10.2196/jmir.5056] [Medline: 27287823]
38. Lee LL, Mulvaney CA, Wong YKY, Chan ES, Watson MC, Lin HH. Walking for hypertension. Cochrane Database Syst Rev 2021 Feb 24;2(2):CD008823. [doi: 10.1002/14651858.CD008823.pub2] [Medline: 33630309]
39. Jabbarzadeh Ganjeh B, Zeraattalab-Motlagh S, Jayedi A, et al. Effects of aerobic exercise on blood pressure in patients with hypertension: a systematic review and dose-response meta-analysis of randomized trials. Hypertens Res 2024 Feb;47(2):385-398. [doi: 10.1038/s41440-023-01467-9] [Medline: 37872373]
40. Choo J, Kim HJ, Park S. Neighborhood environments: links to health behaviors and obesity status in vulnerable children. West J Nurs Res 2017 Aug;39(8):1169-1191. [doi: 10.1177/0193945916670903] [Medline: 27753629]
41. Wang X, Ouyang Y, Liu J, et al. Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of prospective cohort studies. BMJ 2014 Jul 29;349:g4490. [doi: 10.1136/bmj.g4490] [Medline: 25073782]

XSL•FO
RenderX

42. Huang Y, Chen Z, Chen B, et al. Dietary sugar consumption and health: umbrella review. BMJ 2023 Apr 5;381:e071609. [doi: 10.1136/bmj-2022-071609] [Medline: 37019448]
43. Jiao J, Moudon AV, Kim SY, Hurvitz PM, Drewnowski A. Health implications of adults' eating at and living near fast food or quick service restaurants. Nutr Diabetes 2015 Jul 20;5(7):e171. [doi: 10.1038/nutd.2015.18] [Medline: 26192449]
44. Sacks FM, Lichtenstein AH, Wu JHY, et al. Dietary fats and cardiovascular disease: a presidential advisory from the American Heart Association. Circulation 2017 Jul 18;136(3):e1-e23. [doi: 10.1161/CIR.0000000000000510] [Medline: 28620111]
45. How much sodium should I eat per day? American Heart Association. 2024. URL: https://www.heart.org/en/healthy-living/healthy-eating/eat-smart/sodium/how-much-sodium-should-i-eat-per-day [accessed 2025-04-09]
46. Sakane N, Suganuma A, Domichi M, et al. The effect of a mHealth App (KENPO-app) for specific health guidance on weight changes in adults with obesity and hypertension: pilot randomized controlled trial. JMIR Mhealth Uhealth 2023 Apr 12;11:e43236. [doi: 10.2196/43236] [Medline: 37043287]
47. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102. [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
48. Tobacco. World Health Organization. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/tobacco [accessed 2025-04-09]
49. Alcohol's effects on health. National Institute on Alcohol Abuse and Alcoholism. 2024. URL: https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking [accessed 2025-04-09]
50. Drury VB, Saw SM, Finkelstein E, Wong TY, Tay PK. A new community-based outdoor intervention to increase physical activity in Singapore children: findings from focus groups. Ann Acad Med Singap 2013 May;42(5):225-231. [Medline: 23771109]
51. Aune D, Keum N, Giovannucci E, et al. Whole grain consumption and risk of cardiovascular disease, cancer, and all cause and cause specific mortality: systematic review and dose-response meta-analysis of prospective studies. BMJ 2016 Jun 14;353:i2716. [doi: 10.1136/bmj.i2716] [Medline: 27301975]
52. Vorrink S, Huisman C, Kort H, Troosters T, Lammers JW. Perceptions of patients with chronic obstructive pulmonary disease and their physiotherapists regarding the use of an eHealth intervention. JMIR Hum Factors 2017 Sep 19;4(3):e20. [doi: 10.2196/humanfactors.7196] [Medline: 28928110]
53. IBM SPSS statistics for Windows. Version 28.0 ed. : IBM Corp; 2021.
54. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
55. McDowell JRS, McPhail K, Halyburton G, Brown M, Lindsay G. Perceptions of a service redesign by adults living with type 2 diabetes. J Adv Nurs 2009 Jul;65(7):1432-1441. [doi: 10.1111/j.1365-2648.2009.05003.x]
56. Russell CK, Gregory DM. Evaluation of qualitative research studies. Evid Based Nurs 2003 Apr;6(2):36-40. [doi: 10.1136/ebn.6.2.36] [Medline: 12710415]
57. Currency converter. OANDA. 2025. URL: https://www.oanda.com/currency-converter/en/?from=KRW&to=USD&amount=1 [accessed 2025-04-09]
58. Kassavou A, Wang M, Mirzaei V, Shpendi S, Hasan R. The association between smartphone app-based self-monitoring of hypertension-related behaviors and reductions in high blood pressure: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022 Jul 12;10(7):e34767. [doi: 10.2196/34767] [Medline: 35819830]
59. MacPherson MM, Merry KJ, Locke SR, Jung ME. How can we keep people engaged in the behavior change process? An exploratory analysis of two mHealth applications. J Technol Behav Sci 2022 Sep;7(3):337-342. [doi: 10.1007/s41347-022-00253-7]
60. Sevick MA, Zickmund S, Korytkowski M, et al. Design, feasibility, and acceptability of an intervention using personal digital assistant-based self-monitoring in managing type 2 diabetes. Contemp Clin Trials 2008 May;29(3):396-409. [doi: 10.1016/j.cct.2007.09.004] [Medline: 17997364]
61. Burke LE, Styn MA, Sereika SM, et al. Using mHealth technology to enhance self-monitoring for weight loss: a randomized trial. Am J Prev Med 2012 Jul;43(1):20-26. [doi: 10.1016/j.amepre.2012.03.016] [Medline: 22704741]
62. Patnode CD, Redmond N, Iacocca MO, Henninger M. Behavioral counseling interventions to promote a healthy diet and physical activity for cardiovascular disease prevention in adults without known cardiovascular disease risk factors: updated evidence report and systematic review for the US preventive services task force. JAMA 2022 Jul 26;328(4):375-388. [doi: 10.1001/jama.2022.7408] [Medline: 35881116]
63. Jakob R, Harperink S, Rudolf AM, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
64. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp Educ Psychol 2000 Jan;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]
65. Smithson J. Using and analysing focus groups: limitations and possibilities. Int J Soc Res Methodol 2000 Jan;3(2):103-119. [doi: 10.1080/136455700405172]

XSL•FO
RenderX

Abbreviations

IMB: Information-Motivation-Behavioral Skills

Edited by A Mavragani; submitted 04.09.24; peer-reviewed by D Lauver, E Deja; revised version received 10.03.25; accepted 11.03.25; published 02.05.25.

Please cite as:
Choo J, Noh S, Shin Y
Evaluating Feasibility and Acceptability of the "My HeartHELP" Mobile App for Promoting Heart-Healthy Lifestyle Behaviors: Mixed Methods Study
JMIR Form Res 2025;9:e66108
URL: https://formative.jmir.org/2025/1/e66108
doi:10.2196/66108

© Jina Choo, Songwhi Noh, Yura Shin. Originally published in JMIR Formative Research (https://formative.jmir.org), 2.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

School-Based Virtual Reality Programming for Obtaining Moderate-Intensity Exercise Among Children With Disabilities: Pre-Post Feasibility Study

Byron Lai[1], PhD; Ashley Wright[1*], MS; Bailey Hutchinson[1*], BS; Larsen Bright[1*]; Raven Young[1*], BS; Drew Davis[1*], MD; Sultan Ali Malik[2*], MBBS; James H Rimmer[3*], PhD; Pelham High Community Engagement Group[4*]

[1]Division of Pediatric Rehabilitation Medicine, Department of Pediatrics, University of Alabama at Birmingham, 1600 7th Avenue South, Birmingham, AL, United States

[2]CMH Kharian Medical College, National University of Medical Sciences, Rawalpindi, Pakistan

[3]Dean's Office, School of Health Professions, University of Alabama at Birmingham, Birmingham, AL, United States

[4]

[*]these authors contributed equally

Corresponding Author:
Byron Lai, PhD
Division of Pediatric Rehabilitation Medicine, Department of Pediatrics, University of Alabama at Birmingham, 1600 7th Avenue South, Birmingham, AL, United States

Abstract

Background: Children have busy daily schedules, making school an ideal setting for promoting health-enhancing exercise behavior. However, children with mobility disabilities have limited exercise options to improve their cardiorespiratory fitness and cardiometabolic health.

Objective: This study aims to test the feasibility of implementing a virtual reality (VR) exercise program for children with mobility disabilities in a high school setting.

Methods: A pre- to posttrial single-group design with a 6-week exercise intervention was conducted at a high school. The study aimed to enroll up to 12 students with a disability. Participants were given the option of exercising at home or school. The exercise prescription was three 25-minute sessions per week at a moderate intensity, using a head-mounted VR display. School exercise sessions were supervised by research staff. Home exercise sessions were performed autonomously. Several implementation metrics of feasibility were recorded, including exercise attendance, volume, adverse events or problems, and benefits related to health-related fitness (walking endurance and hand-grip strength). The study also included a qualitative evaluation of critical implementation factors and potential benefits for participants that were not included in the study measures. Outcomes were descriptively analyzed, and 2-tailed *t* tests were used as appropriate.

Results: In total, 10 students enrolled in the program and 9 completed the study (mean age 17, SD 0.6 y). In total, 5 (56%) participants exercised at school, and 4 (44%) exercised at home; 1 participant dropped out prior to exercise. The mean attendance for all 9 completers was 61.1% (11/18 sessions). The mean exercise minutes per week was 35.5 (SD 22) minutes. The mean move minutes per session was 17.7 (SD 11) minutes. The mean minutes per session was 18 (SD 1.4) minutes for school exercisers and 17 (SD 18) minutes for home exercisers, indicating variable responses from home exercisers. The mean rating of perceived exertion per exercise session was 4.3 (SD 2), indicating a moderate intensity that ranged from low to hard intensity. No adverse events or problems were identified. No improvements in walking endurance or hand-grip strength were observed. School exercisers achieved a higher attendance rate (83%) than home exercisers (27%; $P<.001$) and seemingly had a 2-fold increase in the volume of exercise achieved (school: mean 279, SD 55 min; 95% CI 212 - 347; home: mean 131, SD 170 min; 95% CI –140 to 401; $P=.10$). Qualitative themes relating to implementation factors and benefits to participant well-being were identified.

Conclusions: This study identified factors to inform an optimal protocol for implementing a high school–based VR exercise program for children with disabilities. Study findings demonstrated that moderate exercise at school is feasible in VR, but simply providing children with VR exergaming technology at home, without coaching, will not successfully engage them in exercise.

Trial Registration: ClinicalTrials.gov NCT05935982; https://clinicaltrials.gov/study/NCT05935982

(JMIR Form Res 2025;9:e65801) doi:10.2196/65801

XSL•FO
RenderX

KEYWORDS

disability; adapted physical activity; leisure-time exercise; exercise; telehealth; tele-exercise; cerebral palsy; pediatric rehabilitation; intellectual disability; developmental disability; child; high school; exercise; mobility; mobility disability; cardio; cardiorespiratory; cardiometabolic; feasibility; virtual reality; controlled trial; t test; VR; exergame; mixed method

Introduction

Children with disabilities generally participate in low volumes of physical activity [1,2], which places them at high risk for cardiometabolic disorders (eg, cardiovascular and metabolic diseases) [3-6]. Cardiometabolic disorders among people with disabilities seem to onset during the early adult years and manifest around the mid-30s at higher prevalences than people without disabilities [6-8]. Age-related trends in disease onset are believed to be linked with gradual reductions in health-enhancing doses of physical activity and accelerated physical deconditioning that is unique to people with disabilities [5,8,9]. Reduced participation in physical activity is suggested to be primarily caused by a lack of accessible, usable, and available opportunities and support within the community [10]. Given that children who participate in physical activity are more likely to participate as adults [11,12], there is a strong need to engage children with disabilities in regular physical activity behavior as a means of health prevention.

Outside of formal rehabilitation from physical and occupational therapy, the high school setting is often the last structured opportunity for physical activity in a person's lifetime. In the United States, physical activity opportunities in high school are mandated by federal law to be provided to all children with disabilities. Nevertheless, mandated opportunity does not necessarily mean that a child will have accessible and usable exercises that can lead to meaningful improvements in health. A systematic review has demonstrated that school-based exercise studies generally have not elicited favorable effects on cardiometabolic health among children with disabilities, but improvements can be observed in cardiorespiratory and muscular fitness and other components of health-related physical fitness [13]. These findings are also supported by more overarching reviews of disability exercise studies outside of the school setting [2]. Children with mobility disabilities who cannot walk, run, or cycle for prolonged periods have very few evidenced aerobic exercise modalities for improving their cardiometabolic health and cardiorespiratory fitness [2,14,15].

Active video gaming is one of the few evidenced aerobic exercise modalities that can improve health among children with disabilities [14,15]. Past active video gaming studies typically used console-based devices such as the Nintendo Wii or X-Box Connect. However, those devices have been discontinued by their manufacturer, and the market is dominated solely by virtual reality (VR) head-mounted displays (HMDs). HMDs emerged during the COVID-19 pandemic in 2019, with the release of the Oculus Quest (now referred to as Meta Quest). The Quest was the first all-in-one HMD that allowed a high-quality gaming experience, and it could be purchased at any major retailer at a price comparable to gaming consoles. In 2019, the first study to use a VR HMD among children with disabilities demonstrated that 2 children with spina bifida could obtain health-enhancing doses of moderate-intensity exercise at home with behavioral telehealth coaching. Since then, HMDs have been used in a variety of contexts, such as pediatric rehabilitation, pain management, and mindfulness training [16-19]. One trial is currently testing the preliminary efficacy of VR HMD exercise training on cardiometabolic health in children with cerebral palsy [20]. However, to the best of our knowledge, few studies have tested the implementation of a VR HMD exercise program for children with disabilities in high school, particularly using the latest HMD, the Meta Quest 3.

The purpose of this 1-year feasibility study was to work with a community engagement group to develop and test the implementation of a VR HMD exergaming program for children with special needs at a high school. The overall purpose of this study was to identify a replicable protocol that could be implemented within the daily operations of a high school through the following research questions:

- How much exercise can students feasibly fit within a school-based program?
- How well would students attend the program?
- How many students can be anticipated to be enrolled from a single school?
- What payment incentive would be required to satisfy enrollment targets and ensure adherence to data collection?
- What were the potential effect estimates of the program on physical fitness tests?
- How do outcomes compare between exercising in school or exercising at home by using a loaned headset by the school?
- Qualitatively, what critical factors should be considered in designing a larger trial?

Methods

Design

This feasibility study was a community-based participatory research design with 2 phases, namely, a development phase and an implementation phase. The development phase included a 4-month preparation phase, where study staff worked with a community engagement group to modify a working VR randomized controlled trial home-based protocol to a local public high school setting. The 9-month implementation phase used a "learn-as-you-go" approach with a single group pre- to posttrial design to test the implementation of the developed VR HMD school-based protocol. The study was implemented at a public high school within the United States, which had adapted education support for children with special needs.

Community Engagement Group

The community engagement group included a case manager of the high school, a young adult with cerebral palsy (male; 19 y of age) with mild intellectual disability and mobility disability who was a student at the school, and his caregiver. The group was involved in the study, starting from the design in the

XSL•FO
RenderX

development phase to the completion of the study in the implementation phase. The group met weekly with the research staff during the development phase, then biweekly during the first half of the implementation phase, and monthly thereafter. The purpose of the community engagement group was to make protocol modifications to intervention and data collection procedures of an existing clinical intervention [20] that would best fit in real-world, daily operations at the school.

Participants

This feasibility study aimed to recruit a convenient sample of 12 participants, which is a general recommendation for exploratory studies targeting implementation process metrics [21], as opposed to group-effect comparisons that would need larger samples [22]. Eligibility criteria are listed below. Participants were recruited through word-of-mouth and flyers by the community engagement group (case manager and young adult).

Inclusion criteria for the study included (1) between 10 and 19 years of age, (2) enrolled in school special education or self-reported disability, (3) a note from a physician permitting participation in moderate-intensity exercise or cleared for sports participation through physical examination that was on file by the school, and (4) ability to communicate in English (or a caregiver).

Exclusion criteria for the study included (1) complete blindness or deafness and (2) conditions that may make participation unsafe.

Supplies

Participants used a VR HMD for exercise (Quest 3, Meta, United States), which came with 2 hand-held controllers. Active video games were purchased through the Meta Oculus digital cloud server. The games were purchased for a standard Meta account created for the study. Games included Beat Saber, Les Mills Body Combat, Creed: Rise to Glory, Smash Drums, Power Beats VR, and Thrill of the Fight.

Intervention Development and Protocol

During the development phase of the study, the community engagement group worked with the research team to modify a home-based VR HMD clinical trial protocol [20] to be more easily implemented in a high school setting. In summary, the original intervention prescription included 150 minutes of moderate-intensity activity per week using the Quest 3 at home, for a total of 12 weeks. The home exercise was supported by theory-driven behavioral telecoaching calls [20]. In that study, caregivers were the primary intervention agents who managed their child's exercise schedules and study activities.

Through the 4-month development phase for this study, meetings with the community engagement group resulted in the following finalized exercise protocol to test in high school:

- Frequency: 3 times per week.
- Intensity: moderate intensity exercise as indicated by a rating of perceived exertion of 5 to 7 on the Borg 0 - 10 scale.
- Time: as many minutes of exercise as possible that could be fit within students' school schedules (eg, either within their physical education period or a study hall or gap period).
- Type: VR exercise at home or in the school setting, at the choice of the caregiver and participant. At enrollment, caregivers were asked whether they preferred the child to perform the exercise at school or home with a borrowed headset. Home-based participants were not given more attention than those who exercised at school. For example, no behavioral coaching calls were given to home participants as is done in the protocol from which this study was modified [20].

Intervention Procedures

Participants and their caregivers who were interested in the program were referred to study staff by the community engagement case manager. Participants were contacted by research staff and screened through phone calls. Eligible participants were directed to a web-based electronic data capture platform (REDCap [Research Electronic Data Capture]; Vanderbilt University) to provide informed consent and assent, participant demographics, and complete the physical activity survey. Next, they were scheduled for an on-site visit at the school to complete the physical fitness tests. Next, study staff scheduled the exercise sessions within a gap period or physical education period for participants who chose to exercise at school. Exercise sessions were performed at the school for 6 weeks, and all sessions were supervised by at least 1 research team member. For participants who chose to exercise at home, they were provided with a Quest 3 with the preinstalled games in a convenient carrying case. They were provided with the same exercise prescription as those who chose to exercise in the school. Home-based participants were not provided with additional attention or behavioral coaching. Quest 3 headsets were returned to the school once the home-based participant completed the 6-week program.

Measures

Participant Characteristics

Age, sex, height, weight, and ethnicity were descriptively reported for the enrolled participants. The use of a mobility device and wheelchair was recorded. As a gauge for how physically active participants were upon joining the study, physical activity level was measured using the National Institutes of Health Patient-Reported Outcomes Measurement Information System Pediatric Physical Activity Short Form 8-item survey [23]. The survey results in a total T-score. T-scores range from 20 to 80, with a mean of 50 (SD 10). A higher T-score indicates high levels of participation in physical activity. Scores that were 0.5 - 1 SD below the mean indicate "mild" difference or impairment; scores 1.0 - 2.0 SD below the mean indicate "moderate" impairment; and scores ≥2.0 SD below the mean indicate "severe" impairment.

Research Question 1: Exercise Output

Exercise volume was measured weekly by 2 variables, namely, mean "move" minutes and rating of perceived exertion. Move minutes were automatically recorded by the Quest 3. The Quest

XSL•FO
RenderX

3 includes a built-in fitness tracker app, referred to as Meta Quest Move. The app tracks calories and move minutes, and these data are stored in cloud-based Meta servers through end-to-end encryption. Meta does not provide specifics on how these data are calculated, but the calculations are reported by Meta to be estimated through an algorithm that includes basal metabolic rate along with the headset, controller, and hand movements. At school, the maximum possible number of active minutes that a participant could exercise each week was 75 minutes because of designated 3-hour class periods. Home-based exercise participants were asked to match the school prescription (3 sessions per week for a total of 75 min). Mean move minutes were averaged each week across the 6-week program. To provide a measure of the intensity of the exercise, participants were asked immediately after exercise to provide a single rating of perceived exertion score that summarized each session, using the Borg 0 - 10 scale [24]. In summary, a score of 0 represents "nothing at all," a score of 3 is indicative of moderate intensity, a score of 5 indicates "strong" exercise, a score of 7 indicates "very strong," and 10 indicates "extremely strong." The mean calories were also used as an indicator of exercise output. Adverse events or problems were reported and documented in accordance with the policies established by the university.

Research Question 2: Student Attendance

Student attendance was expressed as a percentage value, measured by the number of exercise sessions attended, divided by the total sessions prescribed. The exercise prescription was a total of 18 sessions, broken up into 3 sessions per week for 6 weeks.

Research Questions 3 and 4: Student Enrollment and Payment Incentives

Student enrollment was measured by the number of students enrolled per school semester. The program lasted a total of 2 school semesters and was completed prior to summer break. Payment incentives were planned to be increased if the target sample size of 12 was not met from the first school semester. Payment incentives started at US $50 for each of the 2 data collection visits, for a total of US $100. Because of inadequate enrollment in the first semester, the payment incentive for the second semester was increased to US $100 per data collection visit, for a total of US $200.

Research Question 5: Potential Effect Estimates of the Program on Physical Fitness Tests

Physical fitness was measured by walking endurance and hand-grip strength. Walking endurance was measured using a 6-minute walk test (6MWT). A 6-minute push test was planned to be used for students who used wheelchairs, but this was never used. The 6MWT measures the distance walked within 6 minutes around a designated course. The course was a long unobstructed hallway that allowed the participant to freely walk for the duration of the test. When participants reached the end of the hall, they were instructed to turn and continue down the hallway. The rationale for choosing this location was that the study staff found there was limited space to set up and conduct a larger circular track during school hours, particularly without interference from other students. The 6MWT is suggested to be valid and reliable among children without disabilities [25]. However, the 6MWT has conflicting evidence regarding validity for various pediatric clinical groups [26,27]. Hand-grip strength was measured by the mean value of 3 maximal grip trials on a hand-held dynamometer (Camry Digital Hand Dynamometer, Camry, United States), while the participant was in a seated position. Grip-strength has evidence to support its use among children with intellectual disabilities [27].

Research Question 6: Ideal Exercise Environment

To identify the ideal method for how schools can implement the exercise program for students in a future trial, outcomes were compared between exercise settings (home vs school). Compared outcomes included enrollment, exercise volume, attendance, and fitness tests.

Research Question 7: Qualitative Evaluation of Factors for Future Trials

The qualitative component of this study used a mini-ethnographic design, whereby the interventionists were immersed in the culture at school with participants. At postintervention, the interventionists were interviewed (semistructured) to gain insight into feasibility issues and possible programs that should be considered in a future trial. Specific questions were: (1) What were critical issues that impeded or facilitated the implementation of the program at school? and (2) What potential health or fitness-related benefits did participants experience throughout the program (particularly, benefits that were not measured)? When considering these questions, the interventionists were asked to recall and talk through their experience with each of the on-site exercise students. Interventionists were asked to take notes throughout the study. Interviews were audio recorded and analyzed by the lead interventionist, a qualitative analyst (BL), who had conducted more than 400 interviews related to disability and health.

Analyses

Descriptive statistics, including means, SDs, and CIs, were obtained for all study variables. Changes in means were compared primarily using descriptive comparisons with 95% CIs, with 2-tailed *t* tests as appropriate. Analyses were performed using SPSS Statistics (version 29; IBM Corp). Qualitative data were analyzed using a thematic analysis approach [19].

Ethical Considerations

The study was approved by the University of Alabama at Birmingham Institutional Review Board for Human Use prior to enrolling participants (FWA00005960). Participants had to provide informed consent and assent documentation prior to joining the study. The principal of the high school approved and supported the study. Participants were compensated for participation in the data collection (US $100 per data collection, total incentive of US $200). All data were deidentified.

XSL•FO
RenderX

Results

Participant Characteristics

A total of 10 participants enrolled in the program over the 2 semesters. One participant lost interest in joining the program and dropped out prior to starting the program. A total of 9 participants completed the data collection. A total of 5 (56%) participants chose to exercise at school, and 4 (44%) participants chose to exercise at home. Characteristics for the 9 completers are provided in Table 1. The mean age of completers was 17 (SD 0.6) years (6 male and 3 female participants). The mean total Patient-Reported Outcomes Measurement Information System physical activity T-score was 41.9 (SD 7.2), indicating that physical activity levels were mild to moderately below average. Only 1 participant used a mobility device, ankle-foot orthotics. None of the participants used a wheelchair.

Table . Participant characteristics.

ID	Age (years)	Sex	Hispanic or Latino	Ethnicity	Height (inches)	Weight (lbs)	Primary disability	Physical activity (T-score)	Exercise setting
1	17	Female	No	White	56	225	Down syndrome	43.1	School
2	18	Male	No	White	60	190	Autism	28.8	School
3	17	Male	No	White	68	195	Brain tumor	48.8	School
4	16	Male	No	White	61	145	Autism or ADHD[a]	44.6	School
5	16	Female	No	White	62	105	Unknown	43.1	School
6	17	Male	No	White	68	130	Unknown	40.4	Home
7	17	Female	No	White	62	125	ADHD	48.3	Home
8	17	Male	Yes	White	68	126	ADHD	31.8	Home
9	17	Male	No	White	67	114	Cerebral palsy	48.4	Home

[a]ADHD: attention-deficit/hyperactivity disorder.

Implementation Observations

At the high school, all classes within students' schedules were at the longest, 50-minute periods, and the shortest, 25-minute periods. All students exercised during their study hall period. The community engagement group and research staff chose to structure a 25-minute exercise session within the students' 1-hour study hall to allow the students to have 25 minutes to spend on completing academic work. Students did not have physical education built into their class schedule, because they had completed the 1-year physical education requirement prior to joining the study. The education guidelines of the state required only 1 year of physical education across the students' 4 years of education. The exercise setting was typically conducted in an available classroom, which sometimes required desks and chairs to be moved to allow adequate space for gameplay. The school information technology staff created a secure Wi-Fi for the headsets to log in to. School Wi-Fi was generally protected to prevent inappropriate internet use. The VR headsets required Meta accounts with an active email address. Research staff created the accounts and preinstalled games onto the accounts prior to the intervention.

Exercise Volume Results

From the 9 completers (people who completed the data collections), the total mean move minutes across the 6 weeks was 213 (SD 136), ranging from 0 to 377 minutes. The mean move minutes per week was 35.5 (SD 22) minutes. The mean move minutes per session was 17.7 (SD 11) minutes. Excluding a person from the home exercise group who attended no sessions, the mean minutes per session was 19.9 (SD 11) minutes. The mean rating of perceived exertion per exercise session was 4.3 (SD 2), indicating that most exercise sessions were performed at a moderate intensity, which ranged from low to hard intensity. The mean calories expended per attended session was 90.9 (SD 53.8), ranging from 33 to 196. No adverse events or problems (eg, seizures, injuries, or incidental behavioral issues between participants and staff) occurred throughout the study.

Attendance and Enrollment Results

The mean attendance for all 9 completers was 61.1% (mean of 11 sessions attended out of 18 total prescribed sessions). One completer attended zero sessions but completed the pre- and postdata collections. Two participants were enrolled during the first school semester, with a payment incentive of US $50 per data collection (total incentive of US $100). The remaining 8 participants were enrolled during the second school semester after the payment incentive was increased to US $100 per data collection (total incentive of US $200). Recruitment efforts by the research team and community engagement group were consistent between semesters.

Physical Fitness Results

Mean 6MWT distances for the 9 completers were 447 (SD 94; range: 305 - 567 meters; 95% CI 375 - 518) and 469 (SD 78; range: 350 - 568 meters; 95% CI 408 - 529) at pre- and postintervention, respectively. The mean difference in 6MWT

XSL•FO
RenderX

distance from pre to post was 22 (SD 52; t_8=1.3; P=.24; 95% CI −17.7 to 61.8), indicating a possible trend toward improvement but no statistically significant change across time. Left grip strength at pre and postintervention was 24.9 (SD 8; range: 9 - 35; 95% CI 19 - 32) and 24 (SD 9; range: 8 - 39; 95% CI 17 - 31), respectively. Right grip strength at pre and postintervention was 26.7 (SD 8; range: 12 - 40 kg; 95% CI 20 - 33) and 27.7 (SD 10; range: 11 - 47 kg; 95% CI 20 - 35), respectively. These data indicated no difference in grip strength from pre- to postintervention for both hands.

Comparing School and Home Settings

There did not appear to be a difference in baseline levels of physical activity participation between groups (school exercisers: mean T-score 41, SD 8; n=5; home exercisers: mean T-score 42.2, SD 7; n=4). Intervention attendance was higher among those who exercised at school (83% attendance; mean sessions attended 15, SD 2; 95% CI 13 - 18) compared with home exercisers (27% attendance; mean sessions attended 4.8, SD 4; 95% CI –1 to 11), and 95% CIs indicated a difference between settings of 6 - 15 sessions that were statistically significant (P<.001; mean difference=10.7 sessions). Reasons for nonattendance were mostly due to illness and detention, in response to poor behavior at school. The mean total move minutes among the school exercisers (279, SD 55 minutes; 95% CI 212 - 347) was approximately a 2-fold increase over the home exercisers (131, SD 170 min; 95% CI –140 to 401). The 95% CIs indicated a difference between settings of –39 to 337 minutes in favor of the home exercisers, but this difference was not statistically significant (P=.10; mean difference=149 min). The mean minutes per session was 18 (SD 1.4) minutes for school exercisers and 17 (SD 18) minutes for home exercisers, indicating variable responses from home exercisers. There did not appear to be any visual or statistical changes between the home and school settings for physical fitness tests.

Qualitative Evaluation of School Program Benefits

Given the small number of on-site school participants, only 4 themes resulted from thematic analysis. Two themes were related to program benefits, and 2 themes were related to implementation. First, gaming at school aided well-being. Students reported that an exergaming break amidst a full school schedule made them feel better. Specifically, gaming provided a mental break from schoolwork and social interaction stressors (eg, negatively perceived interactions with peers without disabilities). Second, participation in moderate-intensity exergaming increased energy levels throughout the program. For example, halfway through the program, 1 participant who was heavily engaged in the exercise was notably less exhausted when walking in between classrooms, and the participant took less rest in between exercise bouts. This benefit was observed halfway through the 6-week program. Moreover, participants noticeably increased their exercise intensity and gaming skills across the program.

Third, interventionists reported that supervision was necessary for the exercise sessions. Students sometimes require setup and management of their exercise space. For example, some students exercised in a classroom, which required assistance in moving around chairs to make sufficient space for gaming. In 1 case, a participant who was highly engaged in the exercise needed to be prevented from overexertion. The participant was demonstrating overexertion symptoms (eg, face turning red, breathing heavily, and balance becoming unstable), and this required interventionists to urge the participant to sit down and slow down on the exergaming movements. Fourth, designated school resources are critical factors for successful implementation. Space for exercise was limited within the high school. The school gymnasium was often not a suitable place for exergaming, considering that the space often had verbal and physical interference (accidental balls and objects rolling or flying into the play space) from other students. Empty classrooms were often the primary exercise scenarios. Interventionists noted that a designated space for exergaming would allow more time for exercise by reducing the time needed to set up a play space. Moreover, having a designated wireless internet connection specifically for the headsets aided implementation. The school staff assisted with creating a wireless network access log-in that was made specifically for the project. Although internet access was not necessary to play the games in the Quest 3 headset, games often required software updates that would alleviate bugs or issues in gameplay.

Discussion

Principal Findings

This study developed and piloted a school-modified VR active video gaming program in a high school among students with disabilities. Several key findings were identified that will inform a larger clinical trial. First, this study identified a feasible dose of exercise that can be obtained at school. Without making substantial changes to a student's school schedule, 30 minutes in a single day is likely the most that can be inserted into a student's schedule, specifically in their study hall period. This dose is similar to other feasibility studies attempting to insert external exercise programs into school settings [13,28,29]. Most schools in the United States require physical education, but requirements vary widely. The high school in this study required only 1 year of physical education, in accordance with requirements by the state. Considering these constraints and exercise setup (eg, equipping the headset, clearing space to play, and starting a game) and cooldown allowed students at school to achieve 18 minutes of moderate-intensity aerobic exercise each session. It is important to note that 18 minutes, done 3 times a week or even daily, is far lower than the national guidelines of 60 minutes of moderate to vigorous activity per day for children or even 150 minutes of moderate aerobic exercise per week for adults [30]. Nevertheless, 18 minutes of moderate exercise done 3 times per week may elicit health benefits among people with disabilities. A review of meta-analyses demonstrated that people with disabilities can obtain health benefits at far lower doses of exercise than those noted in adult guidelines [2].

Another notable finding was that the school was an ideal setting for both recruitment and exercise. Obtaining 9 students from a single high school is encouraging, considering that the mean sample size for randomized controlled trials of exercise among children with disabilities is approximately 30 children [14]. To

XSL•FO
RenderX

foster strong recruitment, study findings suggest a minimum of a US $200 payment incentive for a total of 2 data collections. The US $200 payment incentive resulted in 77% (7/9) of participants from a single school semester, whereas a US $100 incentive resulted in only 2 enrolled participants for a semester. Moreover, students who exercised with supervision at school achieved strong attendance (83%; 15 out of 18 sessions), particularly when compared with nonsupervised home-based programs (27%; 4.8 out of 18 sessions). Although supervision places a greater burden on school or research staff, the differences in attendance likely justify the burden. Study findings demonstrate that simply providing children with VR exergaming technology will not successfully engage them in exercise volumes that can likely lead to health benefits.

Limitations

Regarding health and well-being benefits, a limitation of this study was the inclusion of physical performance tests instead of more concrete measures of health. Results on physical performance tests can vary due to mobility, intellectual, and behavioral differences among children with disabilities. Health assessments with stronger psychometric properties will be necessary to further investigate the health benefits of VR active gaming in school settings. Examples of more robust assessments could include a graded exercise test to measure peak oxygen consumption as an indicator of aerobic fitness, or the inclusion of biomarkers (eg, blood spot tests or blood draws) to measure changes in cardiometabolic health. Measures to support the qualitative findings could include surveys or biomarkers of well-being, stress, and anxiety. Interviews could have been conducted directly with students for more in-depth data. Another study limitation was the short 6-week duration of the intervention. Although 6 weeks of structured exercise may be sufficient to improve cardiorespiratory fitness, changes in cardiometabolic health are much harder to influence from exercise and generally require higher durations lasting several months with longer volumes of weekly exercise [2,13,29]. Future trials aiming to incorporate longer intervention durations in school settings should start participants in exercise as early as possible in a school semester, which typically lasts approximately 18 weeks. Another limitation was the small sample size and quasi-experimental design. Study findings warrant confirmation in larger trials that are randomized.

Conclusions

This study identified a feasible VR program dose that could be fit within the daily activities of a public high school in the Southeast United States. Future research is needed to identify optimal exercise doses that can improve the physical and cardiometabolic health of children with disabilities at school, given the structural and logistical constraints of implementation.

Acknowledgments

This study was supported by grant funding provided by the Forge AHEAD Center and the Center for Engagement in Disability Health and Rehabilitation Sciences (CEDHARS), both of which were located at the University of Alabama at Birmingham. Research reported in this publication was supported by the National Institute on Minority Health and Health Disparities of the National Institutes of Health under award P50MD017338. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We would like to thank the community engagement group for their contribution to this project and give special thanks to Pelham High School for supporting the study. No artificial intelligence was used in the writing or development of this manuscript.

The Pelham High Community Engagement Group includes Dorothy Coogan, Jason Harchuck, and Larsen Bright.

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

References

1. Carlon SL, Taylor NF, Dodd KJ, Shields N. Differences in habitual physical activity levels of young people with cerebral palsy and their typically developing peers: a systematic review. Disabil Rehabil 2013 Apr;35(8):647-655. [doi: 10.3109/09638288.2012.715721] [Medline: 23072296]
2. Ginis KAM, van der Ploeg HP, Foster C, et al. Participation of people living with disabilities in physical activity: a global perspective. Lancet 2021 Jul 31;398(10298):443-455. [doi: 10.1016/S0140-6736(21)01164-8] [Medline: 34302764]
3. Messiah SE, Vidot DC, Somarriba G, et al. Obesity and cardiometabolic disease risk factors among US adolescents with disabilities. World J Diabetes 2015 Feb 15;6(1):200-207. [doi: 10.4239/wjd.v6.i1.200] [Medline: 25685291]
4. Dhanasekara CS, Ancona D, Cortes L, et al. Association between autism spectrum disorders and cardiometabolic diseases: a systematic review and meta-analysis. JAMA Pediatr 2023 Mar 1;177(3):248-257. [doi: 10.1001/jamapediatrics.2022.5629] [Medline: 36716018]

XSL•FO
RenderX

5. Ryan JM, Crowley VE, Hensey O, Broderick JM, McGahey A, Gormley J. Habitual physical activity and cardiometabolic risk factors in adults with cerebral palsy. Res Dev Disabil 2014 Sep;35(9):1995-2002. [doi: 10.1016/j.ridd.2014.03.051] [Medline: 24864052]
6. Heyn PC, Tagawa A, Pan Z, Thomas S, Carollo JJ. Prevalence of metabolic syndrome and cardiovascular disease risk factors in adults with cerebral palsy. Develop Med Child Neuro 2019 Apr;61(4):477-483. [doi: 10.1111/dmcn.14148]
7. Peterson MD, Kamdar N, Hurvitz EA. Age-related trends in cardiometabolic disease among adults with cerebral palsy. Dev Med Child Neurol 2019 Apr;61(4):484-489. [doi: 10.1111/dmcn.13777] [Medline: 29704244]
8. McPhee PG, MacDonald MJ, Cheng JL, Dunford EC, Gorter JW. Emerging evidence for accelerated ageing and cardiovascular disease in individuals with cerebral palsy. J Rehabil Med 2019 Jul 8;51(7):525-531. [doi: 10.2340/16501977-2564] [Medline: 31120542]
9. Rimmer JH, Schiller W, Chen MD. Effects of disability-associated low energy expenditure deconditioning syndrome. Exerc Sport Sci Rev 2012 Jan;40(1):22-29. [doi: 10.1097/JES.0b013e31823b8b82] [Medline: 22016146]
10. Ginis KAM, Ma JK, Latimer-Cheung AE, Rimmer JH. A systematic review of review articles addressing factors related to physical activity participation among children and adults with physical disabilities. Health Psychol Rev 2016 Dec;10(4):478-494. [doi: 10.1080/17437199.2016.1198240] [Medline: 27265062]
11. Dohle S, Wansink B. Fit in 50 years: participation in high school sports best predicts one's physical activity after age 70. BMC Public Health 2013 Dec 1;13:1100. [doi: 10.1186/1471-2458-13-1100] [Medline: 24289060]
12. Telama R, Yang X, Viikari J, Välimäki I, Wanne O, Raitakari O. Physical activity from childhood to adulthood: a 21-year tracking study. Am J Prev Med 2005 Apr;28(3):267-273. [doi: 10.1016/j.amepre.2004.12.003] [Medline: 15766614]
13. Manojlovic M, Roklicer R, Trivic T, et al. Effects of school-based physical activity interventions on physical fitness and cardiometabolic health in children and adolescents with disabilities: a systematic review. Front Physiol 2023;14:1180639. [doi: 10.3389/fphys.2023.1180639] [Medline: 37362446]
14. Lai B, Lee E, Wagatsuma M, et al. Research trends and recommendations for physical activity interventions among children and youth with disabilities: a review of reviews. Adapt Phys Activ Q 2020 Apr 1;37(2):211-234. [doi: 10.1123/apaq.2019-0081] [Medline: 32131052]
15. Lai B, Lee E, Kim Y, et al. Leisure-time physical activity interventions for children and adults with cerebral palsy: a scoping review. Dev Med Child Neurol 2021 Feb;63(2):162-171. [doi: 10.1111/dmcn.14751] [Medline: 33241561]
16. Tobaiqi MA, Albadawi EA, Fadlalmola HA, Albadrani MS. Application of virtual reality-assisted exergaming on the rehabilitation of children with cerebral palsy: a systematic review and meta-analysis. J Clin Med 2023 Nov 14;12(22):7091. [doi: 10.3390/jcm12227091] [Medline: 38002703]
17. Lai B, Powell M, Clement AG, Davis D, Swanson-Kimani E, Hayes L. Examining the feasibility of early mobilization with virtual reality gaming using head-mounted display and adaptive software with adolescents in the pediatric intensive care unit: case report. JMIR Rehabil Assist Technol 2021 May 27;8(2):e28210. [doi: 10.2196/28210] [Medline: 34042602]
18. Lai B, Davis D, Narasaki-Jara M, et al. Feasibility of a commercially available virtual reality system to achieve exercise guidelines in youth with spina bifida: mixed methods case study. JMIR Serious Games 2020 Sep 3;8(3):e20667. [doi: 10.2196/20667] [Medline: 32880577]
19. Lai B, Davis D, Young R, et al. Group telegaming through immersive virtual reality to improve mental health among adolescents with physical disabilities: pre- and posttrial protocol. JMIR Res Protoc 2022 Oct 13;11(10):e42651. [doi: 10.2196/42651] [Medline: 36194864]
20. Lai B, Davis D, Young R, et al. The effects of virtual reality tele-exergaming on cardiometabolic indicators of health among youth with cerebral palsy: protocol for a pilot randomized vontrolled trial. JMIR Res Protoc 2022 Aug 17;11(8):e40708. [doi: 10.2196/40708] [Medline: 35976192]
21. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharm Stat 2005 Oct;4(4):287-291. [doi: 10.1002/pst.185]
22. Teresi JA, Yu X, Stewart AL, Hays RD. Guidelines for designing and evaluating feasibility pilot studies. Med Care 2022 Jan 1;60(1):95-103. [doi: 10.1097/MLR.0000000000001664] [Medline: 34812790]
23. Tucker CA, Bevans KB, Becker BD, Teneralli R, Forrest CB. Development of the PROMIS pediatric physical activity item banks. Phys Ther 2020 Aug 12;100(8):1393-1410. [doi: 10.1093/ptj/pzaa074] [Medline: 32313952]
24. Borg E, Kaijser L. A comparison between three rating scales for perceived exertion and two different work tests. Scand J Med Sci Sports 2006 Feb;16(1):57-69. [doi: 10.1111/j.1600-0838.2005.00448.x] [Medline: 16430682]
25. Li K, Cardoso C, Moctezuma-Ramirez A, Elgalad A, Perin E. Heart rate variability measurement through a smart wearable device: another breakthrough for personal health monitoring? Int J Environ Res Public Health 2023 Dec 6;20(24):7146. [doi: 10.3390/ijerph20247146] [Medline: 38131698]
26. Bartels B, de Groot JF, Terwee CB. The six-minute walk test in chronic pediatric conditions: a systematic review of measurement properties. Phys Ther 2013 Apr;93(4):529-541. [doi: 10.2522/ptj.20120210] [Medline: 23162042]
27. Wouters M, Evenhuis HM, Hilgenkamp TIM. Systematic review of field-based physical fitness tests for children and adolescents with intellectual disabilities. Res Dev Disabil 2017 Feb;61:77-94. [doi: 10.1016/j.ridd.2016.12.016] [Medline: 28064026]

XSL•FO
RenderX

28. Bellamy J, Broderick C, Hardy LL, et al. Feasibility of a school-based exercise intervention for children with intellectual disability to reduce cardio-metabolic risk. J Intellect Disabil Res 2020 Jan;64(1):7-17. [doi: 10.1111/jir.12690] [Medline: 31625658]
29. Wallén EF, Müllersdorf M, Christensson K, Marcus C. A school-based intervention associated with improvements in cardiometabolic risk profiles in young people with intellectual disabilities. J Intellect Disabil 2013 Mar;17(1):38-50. [doi: 10.1177/1744629512472116] [Medline: 23322342]
30. Piercy KL, Troiano RP, Ballard RM, et al. The physical activity guidelines for Americans. JAMA 2018 Nov 20;320(19):2020-2028. [doi: 10.1001/jama.2018.14854] [Medline: 30418471]

Abbreviations

6MWT: 6-minute walk test
HMD: head-mounted display
REDCap: Research Electronic Data Capture
VR: virtual reality

Edited by A Mavragani; submitted 26.08.24; peer-reviewed by K Yuenyongchaiwat, N Charususin, S DeLuca; revised version received 12.03.25; accepted 14.03.25; published 25.04.25.

Please cite as:
Lai B, Wright A, Hutchinson B, Bright L, Young R, Davis D, Malik SA, Rimmer JH, Pelham High Community Engagement Group School-Based Virtual Reality Programming for Obtaining Moderate-Intensity Exercise Among Children With Disabilities: Pre-Post Feasibility Study
JMIR Form Res 2025;9:e65801
URL: https://formative.jmir.org/2025/1/e65801
doi:10.2196/65801

© Byron Lai, Ashley Wright, Bailey Hutchinson, Larsen Bright, Raven Young, Drew Davis, Sultan Ali Malik, James H Rimmer, Pelham High Community Engagement Group. Originally published in JMIR Formative Research (https://formative.jmir.org), 25.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Discussions of Antibiotic Resistance on Social Media Platforms: Text Mining and Mixed Methods Content Analysis Study

Jocelyne Arquembourg[1*], PhD; Philippe Glaser[2*], PhD; France Roblot[3*], PhD, MD; Isabelle Metzler[4], MSc; Mélanie Gallant-Dewavrin[5], MSc; Hugues Feutze Nanguem[6,7], MPA; Adel Mebarki[8], MBA; Paméla Voillot[8*], MSc; Stéphane Schück[8], MD

[1]Laboratoire l3, Telecom Paris, Telecom Paris, Palaiseau, France

[2]Unité EERA Institut Pasteur, CNRS UMR3525, Université de Paris, Paris, France

[3]CHRU Poitiers, INSERM U1070, CHRU Poitiers, Poitiers, France

[4]Association France Spondyloarthrites, Association France Spondyloarthrites, Tulle, France

[5]Association HTaPFrance, Association HTaPFrance, Meursault, France

[6]CREDIMI: Center for Research on International Market and Investment Law, University of Burgundy, Dijon, France

[7]Pfizer, Paris, France

[8]Kap Code, 146 Rue Montmartre, Paris, France

[*]these authors contributed equally

Corresponding Author:
Paméla Voillot, MSc
Kap Code, 146 Rue Montmartre, Paris, France

Abstract

Background: With the increasing popularity of web 2.0 apps, social media has made it possible for individuals to post messages on antibiotic ineffectiveness. In such online conversations, patients discuss their quality of life (QoL). Social media have become key tools for finding and disseminating medical information.

Objective: To identify the main themes of discussion, the difficulties encountered by patients with respect to antibiotic ineffectiveness and the impact on their QoL (physical, psychological, social, or financial).

Methods: A noninterventional retrospective study was carried out by collecting social media posts in French language written by internet users mentioning their experience with antibiotics, and the impact of their ineffectiveness on their QoL. Messages posted between January 2014 and July 2020 were extracted from French-speaking publicly available online forums.

Results: A total of 3773 messages were included in the analysis corpus after extraction and filtering. These messages were posted by 2335 individual web users, most of them being women around 35 years of age. Inefficacy of treatment options and the lack of information regarding the use of antibiotics were among the most discussed topics. QoL was discussed in 63% of the 3773 messages posted. The most common is the physical impact (78%). Patients discussed the persistence of symptoms and adverse effects. The second kind of impact is psychological (65%), characterized by feelings of anxiety or despair about the situation.

Conclusions: This social media analysis allowed us to identify a strong impact of the perceived ineffectiveness of antibiotic therapy on patients' daily life particularly in terms of physical and psychological consequences. These results provide health care experts information directly generated by patients regarding their own experiences. Social media studies constitute a complementary source of evidence that could be used to optimize messages to the public about appropriate use of antibiotics.

(JMIR Form Res 2025;9:e37160) doi:10.2196/37160

KEYWORDS

antibiotic ineffectiveness; antibiotic resistance; health-related quality of life; real-world; social media; quality of life; quantitative; qualitative; app; application; online; medical information; French; users; antibiotic; social media use

Introduction

When antibiotics were first introduced in the middle of the 20th century, they were hailed as wonder drugs. Patients and physicians alike were amazed at the almost miraculous effect of these drugs on serious bacterial infections. For the past 70 years, physicians have come to expect that antibiotics would cure almost all their patient's bacterial infections, and patients expect that the miracle drugs will still work wonders. Despite an increasing rate of antibiotic resistance globally, they remain

XSL•FO
RenderX

one of the most cost-effective, life-saving medicines contributing to an extended lifespan [1].

Commonly prescribed antibiotics are becoming less effective in everyday populations over time as a result of inappropriate prescriptions, unnecessary use, or their overuse in both humans and animals. At the same time, research and development on new antibiotics has slowed down [2]. It is to note that almost all the new antibiotics that have been brought to market in recent decades are variations of antibiotic drugs classes that had been discovered by the 1980s.

The European Surveillance of Antimicrobial Consumption Network showed that large variations in antibiotic use and antibiotic resistance rates exist across Europe, with higher use in Southern Europe including France and lower in Northern Europe [3].

In general, social media refers to forms of electronic communication (such as websites for social networking and microblogging) through which users create online communities to share information, opinions, personal messages, photos, videos, and other contents within internet apps [4]. Laudon and Traver [5] describe online social networking as an online social area for people who share common ties that can interact with one another. Organizations are able to use the internet to connect with consumers in the health care field. Consumers heavily rely on information found online and use internet to gather health care information and connect with other patients to garner support and learn about similar conditions. Individuals will use social media to post reviews or other comments that support or possibly deter others from choosing that type of health care in the future. It is essential for providers to be active on social media and provide accurate information and connect with readers. Certain groups use online channels more often depending on the topic they are searching for. And people responsible for others, such as parents or caregivers, search health topics even more. Almost 90% of adult users accessed popular social media sites to find and share health information [6]. People aged 18 to 24 years are twice as likely to use social media to discuss health issues than people aged between 45 to 54 years [7].

The digitizing of health care data, as well as advancements in computer processing and data storage, has enabled the development of advanced algorithms in the form of artificial intelligence (AI) [8]. AI tools are strongly related with data mining and AI is nowadays ranked among the top-10 technology, whichever the application [9]. Despite their limitations, AI tools and techniques that are still in their infancy already provide substantial benefits in providing in-depth knowledge on individuals' health and predicting population health risks. Their use for medicine and public health is expected to increase substantially in the near future [10].

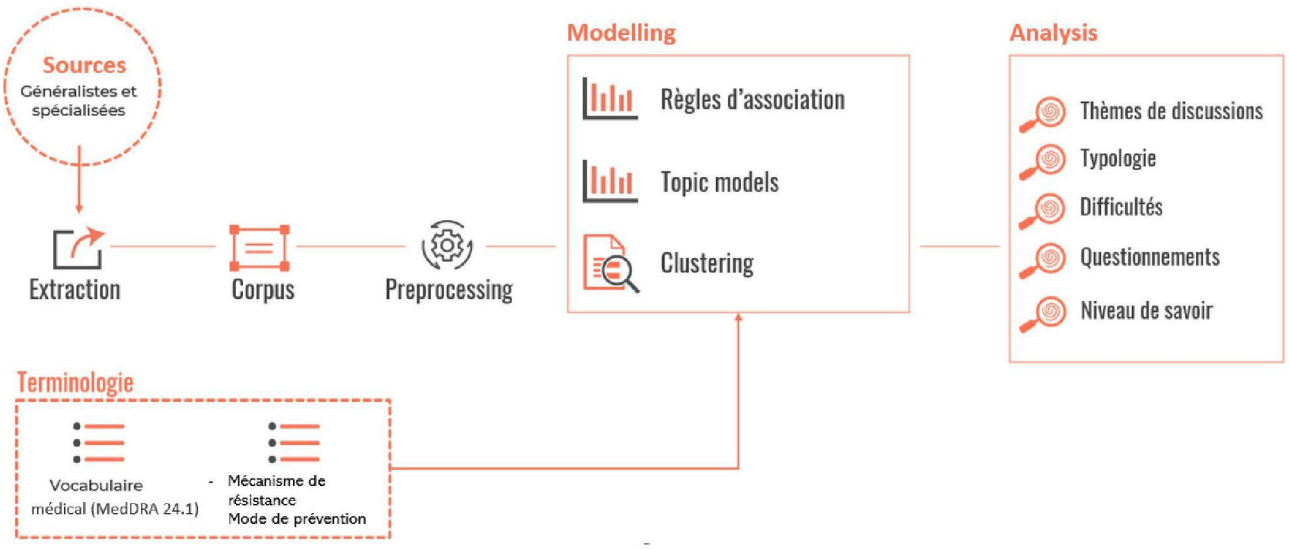
Health is a multifaceted concept and there is no single indicator that can adequately assess its impact in relation to quality of life (QoL). The framework for measuring the QoL includes, among others, information on the share of the population with unmet needs for medical and dental care and a contrasting set of health determinant indicators [11]. In this study, we collected on the web and sought to analyze the messages and complaints of patients with antibiotic treatment failure using innovative tools, without prejudging the cause of the failure (misuse or antibiotic resistance) in order to identify the main topics of discussion, and to identify the difficulties encountered by patients regarding antibiotic ineffectiveness and the impact on their quality of life (physical, psychological, social, or financial).

Methods

Study Design and Population

This was a noninterventional retrospective study using a text mining approach to retrieve information from social media posts (data available in the public domain) written by French speaking internet users between 2014 and 2020 (Figure 1). The study was conducted in 2 phases, data collection using the published Detec't webcrawler [12,13] developed by Kap Code (Paris, France) to collect antibiotic resistance-related posts and quantitative and qualitative analyses to identify trends and characterize key topics discussed by French speaking users.

Figure 1. Study framework.



XSL•FO
RenderX

Data Extraction

A web crawler is an engine that browses through hyperlinks and stores them for future download of associated web pages (identified by the visited hyperlinks) [14]. Scraping of messages was performed according to the HTML structure of each forum. All discussions containing at least one of the keywords or one of their synonyms, coupled with the search for a lexical field of inefficiency within the same message (eg, inefficient or does not work), were automatically retrieved with all the associated metadata and anonymized before being stored in a study-specific database.

Messages associated with antibiotic's supposed ineffectiveness were retrieved from 46 French sources (such as Doctissimo [Reworld Media], Twitter [rebranded as X], and listed in Table 1) dating from 2014 to 2020.

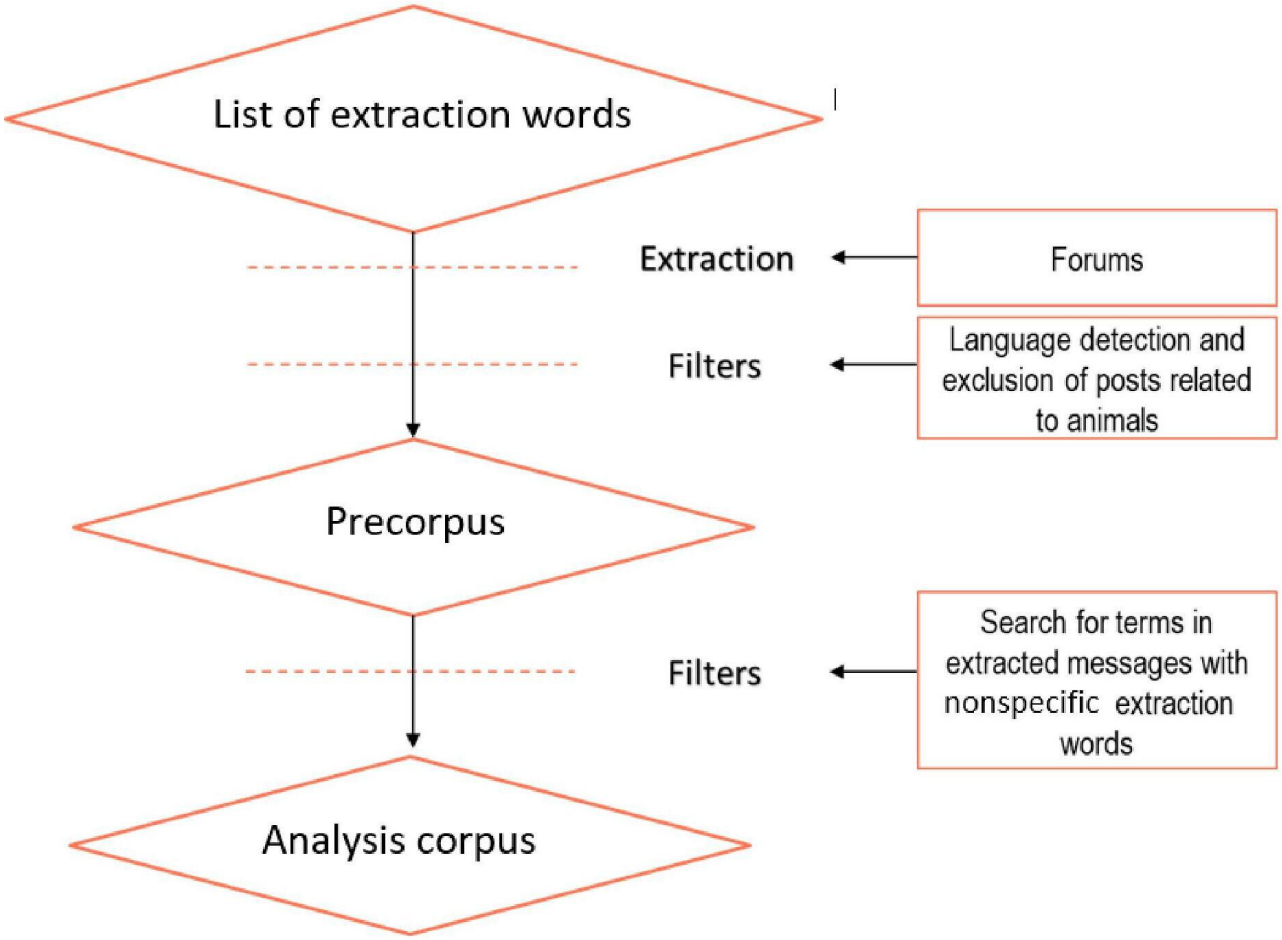
Table . Data sources. Messages about lack of efficacy of antibiotics and antibiotic resistance were retrieved from 46 French sources.

Forums	Posts, n	Users, n
Twitter (X)	888	475
Doctissimo	885	621
Atoute	396	131
Futura-Sciences	260	135
Au féminin	171	126
Meamedica	125	124
Les impatientes	123	54
Psychologies	94	43
Onmeda	94	85
Journal des femmes	79	62
Maladies rares	77	67
Thyroide	72	46
YouTube	71	68
Psychoactif	68	42
Lymphome espoiri	61	41
Vinted	53	40
Vulgaris Medical	49	38
Ligue contre le cancer	47	43
Andlil	39	18
Alarm.asso	38	15
Maman pour la vie	37	29
Pedia Blog	13	7
Forum HardWare	10	7
Other sites	23	18
Total	3773	2335

The analysis corpus (Figure 2) consisted of the corpus filtered after the removal of messages containing predetermined keywords written in a language other than French, post containing animal-related vocabulary, and messages containing at least 1 of the study-specific exclusion words listed in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 2. Flowchart presenting the steps for creating the analysis corpus.



Data Analysis

Age and Gender

Web users' gender is determined through the identification of regular expressions for each gender. First names and gender-associated suffix and prefix are first searched in the username. Then, content of all available messages is screened for gender-specific lexical fields and gender agreements of adjectives and verbs. Finally, a score is computed for each gender and a prediction for the user is obtained by comparing them.

Web users' ages are identified by a 2-fold method. First, regular expressions of age are identified in the total of the user posts. Then, if no expressions were found, a probabilistic machine learning model predicts an age based on several features. Among them, the model considers syntactic aspects of posts as well as expressed feeling and the source on which the user expresses himself.

Topic Model

A topic model was applied to identify the themes addressed in each message. Topic models consist of text mining approach aiming to automatically identify the abstract themes addressed in a collection of documents. Such models are based on the hypothesis that each document in the corpus corresponds to a distribution of several topics. A biterm topic model (BTM) was used to identify the topics without previous knowledge. A topic is defined as a subject of discussion, which amounts to tokens that frequently appear together in a corpus. The BTM considers the whole corpus as a mixture of topics, where each posts' co-occurring pair of tokens (the biterm) is drawn from a specific topic independently, and modeled topics are probability distributions over the biterms [15].

Topics being probability distributions over tokens of the corpus of study, they can be characterized by the highest per-topic probability tokens. Weighting these probabilities through term-frequency inverse document frequency weighting allows to allocate a higher importance to topic-specific tokens. In this case, the per-topic probability of a token was weighted by the inverse of the probabilities of this token in other topics. For each topic, tokens were ranked from highest to lowest weighted probabilities. The first 15 tokens are designated as the set of characteristic tokens and used to manually name the topic.

QoL

QoL impacts were identified; thanks to an algorithm previously developed [16] on the basis of 2 standardized questionnaires measuring health-related QoL (EQ5D and SF36). The algorithm is 2-fold. First, it indicates if an impact is expressed and, second, thanks to 5 specific models, it indicates the nature of the impact (physical, psychic, activity-related, relational, or financial). The 6 models (1 for the first step and 5 for the second) are extreme gradient boosting models trained on annotated patients posts related various pathologies. Features involved in both steps of

XSL•FO
RenderX

the algorithm describe expressed feelings, grammar, conjugation, and lexical fields of health-related QoL-related features, integrating semantic fields specific to the social media domain, allowing it to capture the patient experience as closely as possible. This algorithm was used to automatically create groups of posts as per type of impact, which were subsequently manually reviewed.

Ethical Considerations

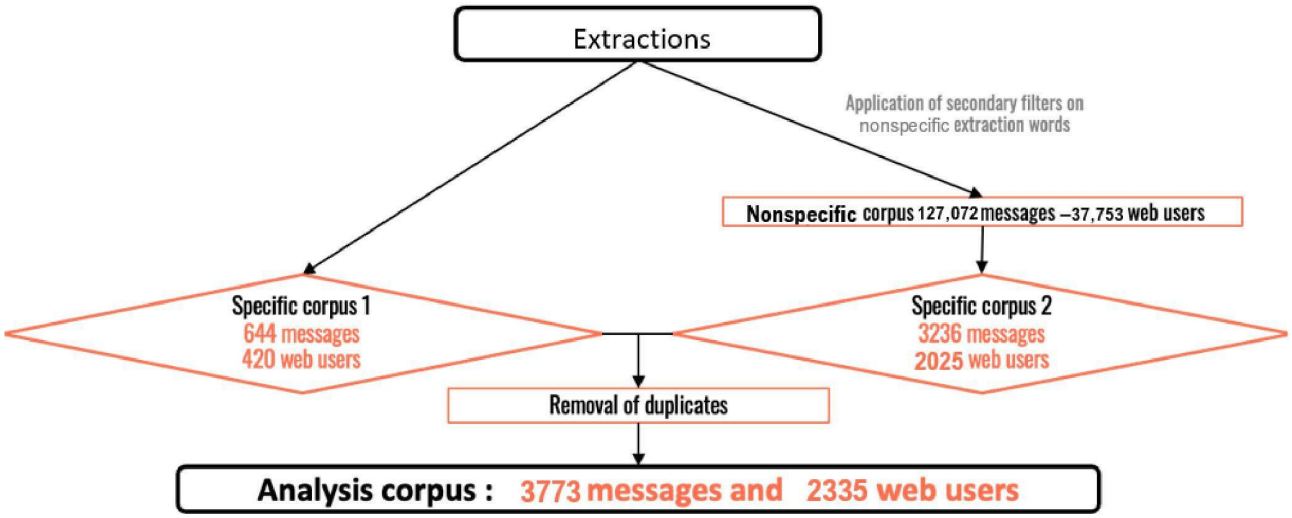
This study included data from publicly available sources; private groups and web pages were thus excluded from our data extraction process. We did not seek approval as users automatically grant their consent for the reuse of their data when they post on public platforms. Furthermore, the results of this study do not contain any identifiable information and are presented in aggregate. Information such as the name, username or handle, geographic locations, or any other sensitive data were not included.

Results

Description of the Population and Posts

After filtering and formatting, the analysis corpus contained a total of 3773 messages corresponding to 2335 different web users, with a mean of 1.62 posts per user (Figure 3).

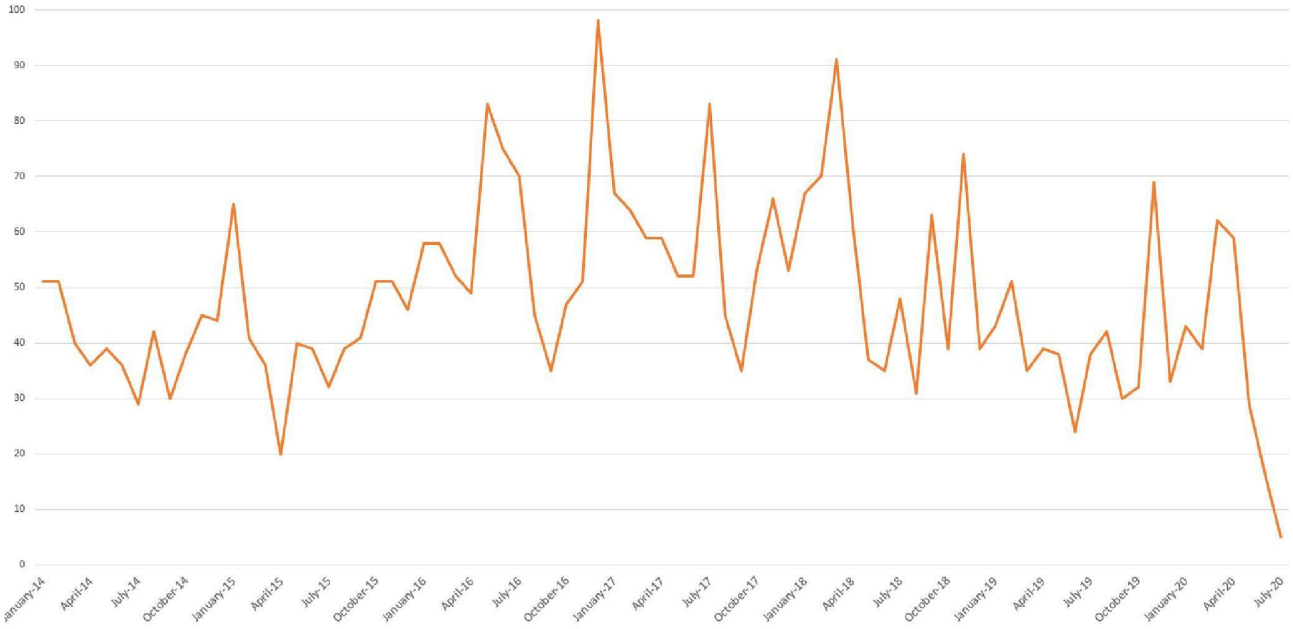
Figure 3. Methodology.



Data showed that there is no specific forum for the sharing of information on activities relating to lack of efficacy of antibiotics and antibiotic resistance and a disparity of sources with a significant proportion of extracted data coming from X and Doctissimo with 888 (23.5%) and 885 (23.5%) posts out of 3773, respectively (Table 1). Furthermore, there was no connection between sources, suggesting a lack of structuring and of visibility of information networks.

We observed fluctuations in the evolution of the volume of messages between January 2014 and July 2020 (Figure 4). We observe a seasonality effect mainly during the “cold” and flu periods, from November to March.

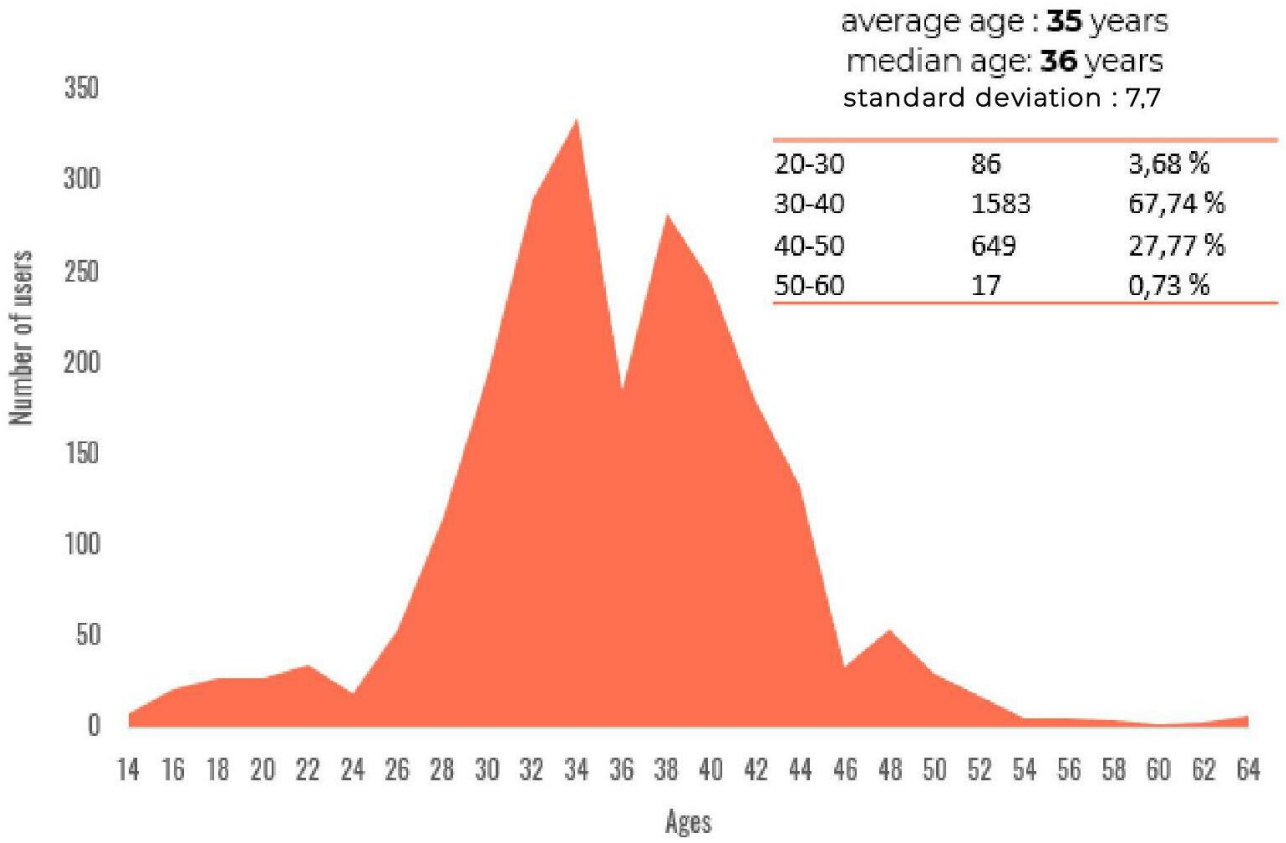
Figure 4. Fluctuations in the evolution of the volume of messages between January 2014 and July 2020 .



XSL•FO
RenderX

Most users were determined as being women (2015/2337, 86.22%), and the average age was estimated of about 35 years (Figure 5). This is consistent with the results of many studies that point out that women express more personal issues in social networks [17,18]. Furthermore, mothers are more likely to seek advice and help about their infants' health from multiple sources, including social media [19].

Figure 5. Age distribution.



Discussion Themes and Topics

As shown in Table 2, the BTM identified 13 main themes of discussion. The largest category of topics was related to the inefficacy of treatment options ("ineffective antibiotics" [597/3770, 15.84%]), especially for ambulatory diseases such as cystitis (356/3770, 9.44%), acne (226/3770, 5.99%), or dental disorders (109/3770, 2.89%), followed closely by the lack of information regarding their use ("antibiotics and pregnancy," 402/3770, 10.66%). Furthermore, our findings showed a certain degree of superficiality in the knowledge regarding antibiotics resistance (351/3770, 9.31%) and antibiotics overuse (301/3770, 7.98%).

XSL•FO
RenderX

Table . Distribution of posts by overall topic category. Application of a biterm topic model that automatically identifies the different topics addressed within the message corpus. Each theme is associated with a set of characteristic words which allows to define them. A message can only be associated with 1 theme.

Topics	n (%)
Ineffective antibiotics	597 (15.84)
Antibiotics and ear, nose, and throat disorder	461 (12.23)
Antibiotics and pregnancy	402 (10.66)
Persistent urinary infections	356 (9.44)
Mechanisms of resistance	351 (9.31)
Antibiotics overuse	301 (7.98)
Inappropriate use of antibiotics	272 (7.21)
Medical data exchanges	238 (6.31)
Antibiotics and acne	226 (5.99)
Gastrointestinal impact of antibiotics	137 (3.63)
Dental disorders and antibiotics	109 (2.89)
Sensitivity analysis antibiogram	77 (2.04)
Nosocomial infection	66 (1.75)

QoL

QoL is an important criterion for evaluating the effects of a disease and of treatment interventions. In this study, the personal experience of the users in terms of impact on QoL was assessed as physical well-being, social well-being, financial well-being, emotional or psychological well-being, and day-to-day activities. Among the 3773 messages posted, 63% (2384) contained at least 1 impact related to the ineffectiveness of antibiotic treatment on their daily life (Table 3). The most reported impact is physical (1866, 78% of messages), with patients discussing the persistence of symptoms (pain and fatigue), and adverse effects (diarrhea and joint pain). Psychological impact (1551, 65% of messages) is characterized by feelings of anxiety, fear, or despair, with some messages even mentioning depression. Current activities (745, 31% of messages) were also impacted, with sick leave, professional difficulties, or a decrease in regular physical activities. Social impact (602, 25% of messages) was marked by family or sexual problems and feelings of social isolation. Financial impact was less discussed in only 1% (28) of the messages, with messages essentially related to unnecessary expenses or care coverage.

Table . Impact of antibiotic treatment ineffectiveness on patients' quality of life.

Impact	n (%)
Physical	1866 (39)
Psychological	1551 (32)
Day-to-day activities	745 (16)
Social	602 (13)
Financial	28 (1)

Discussion

Principal Findings

Using a web-based tool, we have been able to assess retrospectively how (web user) patients perceive lack of efficacy of antibiotics in France over the last 7 years. Most users were women, and the average age was of about 35 years (Figure 5). Discussions were mainly about therapeutic wandering, without a strong knowledge about antibiotics mechanism of action and antibiotic resistance. This situation might be due to some lack of knowledge and interest on the subject, highlighting the wandering without addressing the essential points at the origin of the lack of efficacy like prescription of antibiotics in case of viral infection. Furthermore, we showed that there is a strong impact of perceived ineffectiveness of antibiotic therapy on the QoL of the patients, especially a strong physical and psychological impact for diverse infections and situations.

Limitations of the Study

This study is to our knowledge the first analysis of messages posted on social media and related to antibiotic resistance in human medicine. It also shows several limitations. First, by focusing only on terms in the French language, we will have underreported the total volume of messages related to antibiotics resistance, and thus, these results are not generalizable at a worldwide scale. Furthermore, using social media to analyze patients' reactions excludes patients who do not have access to the internet or who are not familiar with the use of online discussions.

XSL•FO
RenderX

Extraction bias and a limited amount of data compared with similar studies related to other patient groups (such as breast cancer) were the second limitation. Even if a reasonable sized set of keywords was used, keyword selection in social media studies can induce varying levels of extraction bias, accentuated by a low volume of data.

Another limitation of using social media is the lack of complete information about individual cases. There is also the problem of discovering demographic information—only limited or no information regarding individual user demographics (such as age and origin) may be available.

Despite these limitations, social media represent an ideal place where patients can freely and spontaneously discuss their experiences along with their therapy, thus providing valuable information on their QoL difficult to access by other means. Nevertheless, these observations should be interpreted cautiously, since social media data may include a higher frequency of erroneous information, and patients posting on social media forums may not be representative of the wider patient population.

Implications and Future Research

Messages published on social networks should be integrated into the assessment of patient's QoL, as they can help to characterise the patient's experience in a more individualized and spontaneous way. Furthermore, it seems important to further explore certain subpopulations (eg, pregnant women, children, or even older adults via their parents or caregivers).

To the extent that the most dominant topics can be interpreted as unmet informational needs, our study highlights the refinement of practical implications such as the improvement of existing tools, and further reinforcement of the use of available tools (such as antibiotic susceptibility testing or rapid detection tests), and as well as the communication around them.

The ability of internet users to attribute this perceived ineffectiveness to inadequate prescribing or antibiotic resistance is not perceptible in this study and could be the subject of further analysis.

This study suggested to carry out a study segmented by year to examine the evolution of the different subjects (QoL, topics, and treatments) over the years and to observe the global evolutions. Furthermore, similar studies in another country, such as Germany or the United Kingdom, should be planned to compare messages posted between three countries with similar way of life and growth domestic product.

Conclusion

Given that health information is shared extensively on social networks, such services can potentially be used to gather important real-time health data and may provide a venue to identify potential misuse or misunderstanding of antibiotics, promote positive behavior change, and disseminate valid information related to QoL and to bacterial and nonbacterial infections.

One of the most important challenge for health care professionals is to educate people about antibiotics, their adverse effects, and to encourage them to stop the misuse of such drugs. Our study illustrates the current situation concerning the lack of efficacy of antibiotics and the paucity of valuable information and treatment options for wandering patients. In our case, nothing allows us to affirm that these are therapeutic failures related to antibiotic resistance, inappropriate prescriptions or misuse by patients. However, it can provide health care experts with direct patient input on their experience in real-life conditions. This web analysis made it possible to collect, for the first time, the expression of patients who have failed antibiotic therapy. While the extent of these failures is difficult to demonstrate, these patients' experiences tell us how they feel. This research deserves to be continued to better understand the evolution of this phenomenon.

Acknowledgments

Medical writing support was provided by Nathalie Baudry of the microenterprise NB-Pharmaconsult. The study was funded by Pfizer SAS.

Conflicts of Interest

LO has disclosed that he is employee of Pfizer. FNH, PV, and SS have disclosed that they are employees of Kap Code, a CRO, which contracted with Pfizer to carry out this study.
PG, JA, IM, and MG-D have nothing to declare.

Multimedia Appendix 1
List of exclusion words used to filter the extraction of messages.
[DOCX File, 25 KB - formative_v9i1e37160_app1.docx]

References

1. Sengupta S, Chattopadhyay MK, Grossart HP. The multifaceted roles of antibiotics and antibiotic resistance in nature. Front Microbiol 2013;4:47. [doi: 10.3389/fmicb.2013.00047] [Medline: 23487476]
2. Ventola CL. The antibiotic resistance crisis: part 1: causes and threats. P T 2015 Apr;40(4):277-283. [Medline: 25859123]

XSL•FO
RenderX

3. European Centre for Disease Prevention and Control. Antimicrobial resistance in the EU/EEA (EARS-net) - annual epidemiological report for 2020. 2022.
4. von Muhlen M, Ohno-Machado L. Reviewing social media use by clinicians. J Am Med Inform Assoc 2012;19(5):777-781. [doi: 10.1136/amiajnl-2012-000990] [Medline: 22759618]
5. Laudon KC, Traver CG. E-Commerce: Business, Technology, Society, 4th edition: Pearson; 2008.
6. Tennant B, Stellefson M, Dodd V, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar 17;17(3):e70. [doi: 10.2196/jmir.3992] [Medline: 25783036]
7. Harger howe advertising. Top healthcare recruitment social media secrets. URL: https://inbound.hargerhowe.com/blog/top-healthcare-recruitment-social-media-secrets [accessed 2025-03-03]
8. AHRQ. Medal.org (Medical Algorithms): Agency for Healthcare Research and Quality, Innovation Exchange. 2011. URL: https://innovations.ahrq.gov/qualitytools/medalorg-medical-algorithms [accessed 2025-03-10]
9. 10 Tech & IT Buzzwords For 2021 You Won't Be Able To Avoid BI Blog | Data Visualization & Analytics Blog | datapine. 2020 Dec 1. URL: https://cloudwars.com/reporting/10-tech-it-buzzwords-for-2021-you-wont-be-able-to-avoid/ [accessed 2021-02-10]
10. Shaban-Nejad A, Michalowski M, Buckeridge DL. Health intelligence: how artificial intelligence transforms population and personalized health. NPJ Digit Med 2018;1:53. [doi: 10.1038/s41746-018-0058-9] [Medline: 31304332]
11. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Archive:Quality_of_life_indicators_-_health.
12. Abdellaoui R, Schück S, Texier N, Burgun A. Filtering entities to optimize identification of adverse drug reaction from social media: how can the number of words between entities in the messages help? JMIR Public Health Surveill 2017 Jun 22;3(2):e36. [doi: 10.2196/publichealth.6577] [Medline: 28642212]
13. Kürzinger ML, Schück S, Texier N, et al. Web-based signal detection using medical forums data in France: comparative analysis. J Med Internet Res 2018 Nov 20;20(11):e10466. [doi: 10.2196/10466] [Medline: 30459145]
14. Najork M. Web crawler architecture. In: LIU L, ÖZSU MT, editors. Encyclopedia of Database Systems: Springer US; 2009:3462-3465. [doi: 10.1007/978-0-387-39940-9_457]
15. Yan X, Guo J, Lan Y, Cheng X. A biterm topic model for short texts. Presented at: WWW '13 Proceedings of the 22nd international conference on World Wide Web; May 13-17, 2013; Rio de Janeiro Brazil p. 1445-1456 URL: https://dl.acm.org/doi/proceedings/10.1145/2488388
16. Renner S, Marty T, Khadhar M, et al. A new method to extract health-related quality of life data from social media testimonies: algorithm development and validation. J Med Internet Res 2022 Jan 28;24(1):e31528. [doi: 10.2196/31528] [Medline: 35089152]
17. Rollero C, Daniele A, Tartaglia S. Do men post and women view? The role of gender, personality and emotions in online social activity. Cyberpsychology (Brno) 2019;13(1). [doi: 10.5817/CP2019-1-1]
18. Social Media Today. Gender-Specific Behaviors on Social Media and What They Mean for Online Communications. URL: https://www.socialmediatoday.com/social-networks/gender-specific-behaviors-social-media-and-what-they-mean-online-communications [accessed 2021-02-17]
19. Duggan M, Lenhart A, Lampe C, Ellison N. Parents and social media. Pew Research Center. 2015 URL: https://www.pewinternet.org/2015/07/16/parents-and-social-media/

Abbreviations

AI: artificial intelligence
BTM: biterm topic model
QoL: quality of life

Edited by A Mavragani; submitted 24.02.22; peer-reviewed by J Cai, T Delory; revised version received 26.09.22; accepted 27.09.22; published 25.04.25.

Please cite as:
Arquembourg J, Glaser P, Roblot F, Metzler I, Gallant-Dewavrin M, Nanguem HF, Mebarki A, Voillot P, Schück S
Discussions of Antibiotic Resistance on Social Media Platforms: Text Mining and Mixed Methods Content Analysis Study
JMIR Form Res 2025;9:e37160
URL: https://formative.jmir.org/2025/1/e37160
doi:10.2196/37160

© Jocelyne Arquembourg, Philippe Glaser, France Roblot, Isabelle Metzler, Mélanie Gallant-Dewavrin, Hugues Feutze Nanguem, Adel Mebarki, Paméla Voillot, Stéphane Schück. Originally published in JMIR Formative Research (https://formative.jmir.org), 25.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License

XSL•FO
RenderX

(https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Motivations of Citizens to Attend an eHealth Course in the Public Library: Qualitative Interview Study

Lucille Standaar[1,2*], MSc; Adriana Margje Israel[3*], MD; Rosalie van der Vaart[4], PhD; Brigitta Keij[1], PhD; Frank J van Lenthe[3], Prof Dr; Roland Friele[2,5], Prof Dr; Mariëlle A Beenackers[3], PhD; Lilian Huibertina Davida van Tuyl[5], MBA, PhD

[1]Centre for Public Health, Healthcare and Society, Department of National Public Health and Healthcare, National Institute for Public Health and the Environment, Postbus 1, Bilthoven, The Netherlands

[2]Department of Tranzo, School of Social and Behavioral Sciences, Tilburg University, Tilburg, The Netherlands

[3]Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

[4]Research Group Technology for Healthcare, The Hague University of Applied Sciences, The Hague, The Netherlands

[5]Netherlands Institute for Health Services Research, Utrecht, The Netherlands

[*]these authors contributed equally

Corresponding Author:
Lucille Standaar, MSc
Centre for Public Health, Healthcare and Society, Department of National Public Health and Healthcare, National Institute for Public Health and the Environment, Postbus 1, Bilthoven, The Netherlands

Abstract

Background: There is worldwide recognition of the potential increase of digital health inequity due to the increased digitalization of health care systems. Digital health skill development may prevent disparities in eHealth access and use. In the Dutch context, the public library has started to facilitate support in digital health skill development by offering public eHealth courses. Understanding the motivations of people to seek support may help to further develop this type of public service.

Objective: This is a qualitative study on the motivations of citizens participating in an eHealth course offered by public libraries. The study aimed to explore why citizens were motivated to seek nonformal support for eHealth use.

Methods: A total of 20 semistructured interviews with participants who participated in an eHealth course were conducted in 7 public libraries across the Netherlands. The interviews were conducted between April and June 2022. Purposive sampling took place in the public library during the eHealth course. The interviews covered participants' motivations, attitudes, and experiences with eHealth use and their motivations to seek help with eHealth use. Interviews were audio-recorded and transcribed. Themes were identified via a comprehensive thematic data analysis.

Results: The participants were 51 to 82 years of age (average 73.5, SD 6.6 y) and 14 (70%) participants were female. Three motivational themes were identified: (1) adapting to an increasingly digital society, (2) sense of urgency facilitated by prior experience in health care, and (3) a need for self-reliance and autonomy. Additionally, participants expressed a general desire for social contact and lifelong learning. A lack of adequate informal support by friends and family for digital skills and positive experiences with formal support from public libraries stimulated the participants to seek formal support for eHealth use.

Conclusions: We show that the participants had a feeling of urgency that sparked their motivation to seek nonformal support in the public library. Motivations to participate in the eHealth course stemmed from the need to adapt to the digital society, being a patient or a caregiver, or the need or wish to be independent from others. Participants of the study were mainly older female adults who had native language abilities, up-to-date digital devices, and time. It is likely that other populations experience similar feelings of urgency but have other support needs. Future research should explore the needs and attitudes of nonusers and other users of digital health toward seeking support in eHealth access and use.

(JMIR Form Res 2025;9:e60612) doi:10.2196/60612

KEYWORDS

digital health; help-seeking behavior; socioeconomic factors; community health services; older people; eHealth; older adults; geriatric; support; eHealth literacy; interview; Netherlands; thematic data analysis; health literacy; mobile phone

XSL•FO
RenderX

Introduction

With the growing demand for health care services, eHealth is increasingly recognized as a critical tool for efficiently managing and delivering health care [1]. Through the adoption of patient portals, telemonitoring tools, and smartphone apps, patients can experience improved access to health information, shared decision-making, and self-management [2,3]. However, not everyone within society is able to access and use digital health care [4-12]. The increased implementation of digital tools in health services is linked to the widening disparities in health [13].

The need for more emphasis on equity in the access and use of digital health care systems is recognized worldwide. Research shows that disparities in the use of digital health are common in member states of the European Union [14,15]. The causes of digital health inequities derive from unequal access to resources, language, and digital health skills required for effective eHealth utilization [8-10,16-22]. World Health Organization and Organization for Economic Co-operation and Development encourage countries to adopt digital health literacy policies on citizen and health system levels [23,24]. Yet, only 27 out of 52 European Union member states have developed policies on digital health literacy (2023) [23]. Digital health literacy is defined by Norman and Skinner [25] as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem." Enhancement of digital health skills is one of the avenues for achieving a more equitable digital health care system [23,24,26].

In The Netherlands, the COVID-19 pandemic created awareness about the lack of digital health literacy which resulted in a collaborative effort between the government and the public library organization to facilitate support and education in the development of digital skills, including digital health skills [27-29]. As a result of this collaboration, the public library in the Netherlands has evolved into an organization that aims to support "those who need help" [30,31]. The public library offers a nonformal free-of-charge eHealth course, designed to enhance digital health skills [31]. In the literature, 3 categories of support are defined; formal support: institutionalized education with officially recognized awards, nonformal support: institutionalized support and education without officially recognized awards, and informal support: support via social networks [32]. Support via the public library is a form of nonformal support. As the public library is freely accessible to the whole population, costs or resources would not deter the use of this type of support. However, the first step to participate remains to be taken by the individual and depends on their motivation to improve their digital health skills [17,19,31].

A better understanding of the motivations of people who seek nonformal support can create useful insights into the sentiments that drive eHealth adoption and support-seeking behavior. These insights can be used to improve eHealth implementation and the design of eHealth support strategies. This study aims to explore the motivations that drive participants in a freely accessible eHealth course to seek nonformal support for eHealth use. Building on our prior research and other findings, which demonstrated that education was associated with motivation to use eHealth [18], this study's findings will also be explored through the lens of the educational background of people who seek support.

Methods

Study Design

This study aims to understand and explore the motivations of participants to seek support with digital health skill development. A qualitative study design with semistructured interviews was deemed most appropriate as this study aims to explore and describe the motivations to seek help in the eHealth use of participants of an eHealth course [33]. Due to the explorative nature of the research question, this study follows the interpretivism approach [34]. As this study has an explorative nature with the aim to create a deep understanding of the motivations of the participants to seek support, the results of this study will be unique to the study setting and the study sample. This paper adhered to the SRQR (Standards for Reporting Qualitative Research) checklist (Checklist 1) [35].

Ethical Considerations

The study was declared to fall outside the scope of the Dutch Medical Research Involving Human Subjects Act by the Clinical Expertise Center of the Dutch National Institute for Public Health and the Environment (VPZ-559). Transcripts of the interviews and the informed consent forms are stored and protected at the Dutch National Institute of Public Health and the Environment. All participants provided their informed consent prior to conducting the interviews. Participants received a 15€ (US $16.047) voucher for their participation. Transcripts were deidentified.

Setting

Digital health care is widely implemented within the Dutch health care system and involves data storage, interaction between (health care) professionals, interaction between health care professionals and patients, and freely accessible health information websites and apps [36]. Implementation of digital health care has been experimented with over the last decades within the Dutch health care system and is stimulated by the government [37-39]. The Netherlands is one of the best-connected countries in the European Union, with 98% of its citizens connected to a type of internet connection [40]. The history of experimentation with eHealth implementation, governmental support, and the wide availability of the internet is likely to support the digitalization of the health care system further.

The eHealth course "DigiVitaler" is part of the web-based educational material from the National Public Library and available to all public libraries since 2021 [41]. In 2022, 60 out of 137 public libraries offered the eHealth course [41]. This course is structured into multiple classes, each covering a specific eHealth topic, such as the patient portal, video consultations, and web-based health information. Participants are introduced to eHealth by providing hands-on experience with eHealth apps in simulated digital environments. The course

XSL•FO
RenderX

consists of 1 up to 8 classes, each with a duration of 2 - 3 hours, and accommodation of 2 - 12 participants. Public libraries are free to choose the number of classes they offer. Public libraries used printed, word-of-mouth, and social media advertisement strategies to announce the course, disseminated via libraries' own platforms, local newspapers, and local organizations.

Recruitment

Purposive sampling of participants who had experience with the eHealth course was performed by 3 researchers (LS, AMI, and Anita Suijkerbuijk) in the context of 7 public libraries. Researchers aimed to recruit a representative sample of all participants of the eHealth course. Through contact with the developer (Stichting DigiSterker), the researchers retrieved contact information of the libraries offering the eHealth course. The course developer was not involved in this study. Libraries were contacted via email and received information about the study with the request to allow the researchers to be present during the eHealth course for participant recruitment. A total of 7 out of the 26 contacted libraries agreed to participate. Reasons for libraries to decline were course cancellations due to a low number of course participants, no course dates in the data collection period, no response, or unwillingness to participate in the study. After obtaining consent from the librarians, the researchers (LS, AMI, and Anita Suijkerbuijk) attended 9 classes in 7 libraries located in rural and urban areas in the Netherlands. At the start of each class, potential participants were introduced to the study through a presentation and a flyer. With the approval of the course trainer, researchers would play an active role during the class, providing support and answering the course participants' questions about the course material. At the end of the class, participants were invited for an interview. Participants who expressed interest were provided with detailed study information, a consent form, and the opportunity to ask questions. Recruitment and interview procedures were designed to limit the participants' potential efforts to participate in the study. Procedures were conducted in a familiar context for the participants and required limited travel time and no digital skills. All participants were interviewed at the library, either immediately following the course or at a preferred later date. Telephone interviews were an option but were not requested by the participants. Digital interview formats were not considered as it was highly likely that the participants had limited digital skills, which would potentially pose a barrier to recruitment.

Participants

Participants were eligible to participate if they were older than 18 years of age and participated in the eHealth course. Educational background was categorized according to the International Standard Classification of Education (ISCED), where a low educational background is categorized as ISCED 0 - 2, intermediate as ISCED 3 - 4, and high as ISCED 5 - 8 [42]. The rurality of the participants' living area was derived from the area where libraries were situated. Rurality was determined via the rurality classification from Statistics Netherlands [43]. A total of 48 participants attended the classes visited by the researchers, of which 20 (20/48, 42%) participants participated in the study. Reasons for nonparticipation were lack of time or interest or the participant's own perception of lack of experience with eHealth.

Data Collection

In this study, motivations related to seeking help and the use of eHealth were studied. The themes of the interview guide covered: the demographics of the participant; the motivations, self-efficacy, and attitudes for eHealth use; motivations regarding seeking and receiving support for eHealth use; and social context of support and eHealth use. The interview guide can be found in Multimedia Appendix 1.

Multiple theories describe the transition from motivation to behavior [44-46]. The 3 concepts, attitude, social norms, and self-efficacy, are often described to result in motivation [44-46]. In this study, these theories were used as a means to develop the interview guide that covers the main concepts of motivation. Attitude refers to positive and negative beliefs about the consequences of certain behaviors [44-46]. Social norms include the beliefs of others about the behavior [44-46]. Self-efficacy expectations refer to the beliefs of one's personal capacities to actually change behavior [44-46]. Questions that cover these concepts entailed questions such as: "What or who caused you to seek support with eHealth use and why?" (attitude and social norm); "What was the role of family; friends; health care professionals; the library; or the municipality to use eHealth?" (social norm); "Would you be able to use eHealth yourself after the course?" (self-efficacy). The interview guide was developed by LS and AMI and reviewed by LT, AS, and MAB.

Semistructured interviews were conducted face-to-face by researchers LS, Anita Suijkerbuijk, and AMI, all experienced in conducting qualitative research. None of the researchers knew the participants prior to this study. The data collection period lasted from April until June 2022. All interviews were audio-recorded with participants' permission and lasted 30 - 60 minutes. After each cycle of 6 interviews, the content of the interviews was discussed by LS and AMI. After 20 interviews, no new information regarding motivation to seek help with eHealth use was heard. At this point, it was concluded that data saturation was reached.

Data Analysis

Deidentified verbatim transcriptions of the interviews were analyzed with the "Codebook Thematic Analysis" approach [47]. This approach allows for a structured process of coding, theme development, and conceptualization in a multidisciplinary team while maintaining the open, inductive approach of reflective thematic analysis to explore the motivations of the participants in-depth [48]. First, transcripts were read and summarized for data familiarization. A deductive codebook was derived from the interview guide. The first 2 interviews were double-coded by LS, AS, LT, and AMI, who then collaboratively defined and fine-tuned the codebook. After consensus, the remaining interviews were coded and cross-checked by LS and AMI, using MaxQDA (VERBI Software) for data management. This deductive analysis was performed to create an overview of the data gathered, followed by an inductive analysis to identify underlying patterns. The following topics were analyzed inductively: motivation for

XSL•FO
RenderX

eHealth use, motivation for seeking support in eHealth use, and motivation to participate in the eHealth course. The first 20% of the transcript segments per topic were inductively coded by LS and AMI separately. After reaching a consensus, the rest of the transcripts were coded and cross-checked by LS and AMI. The inductive codes were arranged into candidate themes. Candidate themes were then further defined through discussion between all researchers to determine the final themes. After analysis, all the transcript summaries were reread and the presence of the themes for every participant was evaluated through the lens of their life course with the aim of identifying possible differences according to educational background. Quotes were translated using the translation software DeepL (DeepL SE) and these were checked by a native English speaker with limited Dutch fluency.

Results

Overview

A total of 5 men and 15 women with an average age of 73.5 (SD 6.6; range: 51 - 82) years participated in this study. All participants were White and fluent in the Dutch language. Most participants (n=13) lived in an urban area. Seven participants were categorized with a low educational background (LE), 6 participants with an intermediate (IE), and 7 participants with a high educational background (HE). All participants made use of 1 up to 4 digital devices and 14 participants had prior experience with digital health services. The study sample's (n=20) demographic overview can be found in Multimedia Appendix 2. Participants mentioned three types of support to develop their digital skills: (1) family, friends, and acquaintances; (2) the public library and other public or social organizations; and (3) commercial resources. Participants were either introduced to the eHealth course by library staff during participation in other courses offered by the library, advertisements in the local newspaper, or via their church, social organization, or computer club.

General Experience With the Digitalizing Society

The participants' experiences with and attitudes toward the digital world varied, yet all expressed some degree of difficulty using digital devices and services due to a perceived lack of adequate digital skills. Many pointed out that it is difficult for "their generation" to keep up with the latest developments within an ever-changing field. "Their generation" pertains to those who did not grow up with computers or the internet and only learned this later in life, even though some had multiple years of digital experience.

> *I think it's a shame that there's no choice anymore. There's obviously a generation of us that needs to have a transition that's a little bit smooth, and not like: bam, next year everything's just different. I think a lot of people my age have problems with that.* [Participant 5, female, 72 years of age, LE]

For most, this digital experience was work-related where they were trained in specific, mostly simple computer programs useful for their jobs. However, this was different from their experience in this day and age where every aspect of life has become infused with digitalization and there often is little to no alternative left.

Themes

Overview

Three themes were identified to describe individuals' motivations to seek help for eHealth use: (1) adapting to an increasingly digital society, (2) a sense of urgency facilitated by health care experience, and (3) a need for self-reliance and autonomy. Additionally, a desire for social contact and lifelong learning were general reasons to participate in courses offered by the public library. An overview of the themes, the topics, and the number of codes linked to a topic is presented in Multimedia Appendix 3.

Theme 1—Adapting to an Increasingly Digital Society

Overview

The majority of the participants experienced that society is becoming more digital and they feel the urgency to try and keep up with the digital transition to be able to function as participating citizens. Three aspects encompass this motivation: (1) realizing the lack of one's own digital skills, (2) feeling forced by society to keep up with digitalization, and (3) a desire to feel included in society and be prepared for the future. This theme was identified among all participants, regardless of educational level. It seems in this respect that among the participants of this course, their age and the fact they have little experience with digital technology played a bigger role than their educational background.

A Lack of Digital Skills

The experience and expectation of lacking digital skills created a sense of urgency for most participants to learn how to use eHealth. Participants found it uneasy to navigate through the digital world in general. Many participants reported that developing and maintaining digital skills requires continuous effort.

> *But it's my experience, if you don't do it regularly, it disappears again. Then you think, "Oh dear, how did I do this again?" And just figuring that out, it's very frustrating.* [Participant 11, female, 73 years of age, HE]

Some participants experienced fear of using the computer since they were afraid to make (irreversible) mistakes.

> *The fear of buttons for me is that I think I'm pressing such wrong buttons that everything'll disappear.* [Participant 15, female, 74 years of age, HE]

Some participants tried to use eHealth but failed due to a lack of digital skills. Others found that their lack of digital skills limited them in using eHealth to its full potential or impeded their preferred option in health care services. This motivated them to start learning.

Feeling Forced by Society to Keep Up With Digitalization

As services and goods become increasingly digitalized without alternatives, participants felt compelled to adapt and commit to them.

XSL•FO
RenderX

> *Then I think: something is being shoved down my throat again. I don't want that at all. But still, I get it: you have to, whether you want to or not.* [Participant 19, female, 74 years of age, LE]

Anxiety about falling behind and becoming isolated from society or losing part of their autonomy was common. Feelings of insecurity, shame, or feeling scared using new digital technology were no exception, as previous experiences with or news warnings about scamming or loss of privacy raised their concerns.

> *A copy of my passport, I find that very uncomfortable when asked for it. […] I don't really dare to [upload] ... because I'm afraid, yes it's very farfetched but it does happen. I'm still afraid of being taken over by cybercriminals.* [Participant 17 (female, 51 years of age, HE]

Many participants foresaw that health care services would further digitalize. They feared diminished access to care if they were unable to use eHealth, which resulted in beliefs that they needed to forcefully commit. However, the barriers participants experienced for the use of eHealth were numerous. Negative experiences contributed to the reluctance participants felt to use eHealth: lack of uniformity between eHealth apps, difficulty identifying appropriate health information, and difficulty in phrasing health problems in digital form.

> *Too complicated for us at this age [asking questions via e-consultation]. You have to do too many steps. You have to log in, you have to type, you have to explain what it is, and everything has to be short.* [Participant 10, female, 78 years of age, HE]

Desire to Feel Included in Society and Be Prepared for the Future

The desire to keep up with digital developments led to an intrinsic motivation to learn how to use eHealth. The participants felt that by keeping up, they stay part of society and are prepared for possible (future) encounters with eHealth. Some participants felt the urgency to learn how to use eHealth now because they foresaw that this would become harder as they age.

> *It is the future where we are headed, soon you have to do a lot by computer and by smartphone. Then it's really nice if I know how to use it, because if I'm approaching 70 and I still have to learn, well then it is too late.* [Participant 6, female, 66 years of age, LE]

These participants noticed that they were the early adopters of eHealth in their social network. Friends and relatives did not share the same interests or understand why the participant was willing to learn digital skills to use eHealth. This is part of the reason why nonformal support in the form of the eHealth course is a particularly attractive option, as the informal network of the participants is not as knowledgeable.

Theme 2—Sense of Urgency Facilitated by Prior Health Care Experience

Overview

The second theme depicts that the participants felt motivated to seek help to use eHealth through their prior health care experience. Three aspects encompass this theme: (1) the responsibility as a patient or caregiver, (2) experience with the benefits of eHealth, and (3) the attitude of health care professionals toward eHealth. This theme was mentioned by all participants from the low-educated group, a few participants in the intermediate-education group, and half of the high-educated group. Regardless of their educational background, the participants who have had experience in the health care system felt more urgency to use eHealth.

The Responsibility as a Patient or Caregiver

Many participants had some form of experience with health care, although not all with eHealth. Being a patient or caregiver to their partner or child stimulated a feeling of responsibility to take an active role in the care process. For these participants, this active role encompassed the ability to use eHealth. Participants illustrated this with examples of preparing hospital appointments or looking up information in their electronic patient records. They are able to take more time to think about the medical information before a consultation to prepare questions, and also afterward to reread the information given during a consultation. eHealth allowed participants to better process the information and use this to their advantage.

> *My husband was in the hospital with a heart attack [...] and I saw everything was digital there. I had already looked up his medical history once and printed it out because my husband forgets a lot. [...] And I think, now they are going to start something else [a new digital system] in the hospital again. I have to learn this from the start, and I think my husband has to go along with that.* [Participant 13, female, 70 years of age, LE]

A few participants expressed that they feel this type of active role is expected of them by medical professionals to be able to partake in shared decision-making. These experiences motivated the participants to learn more about eHealth when they were presented with the opportunity.

Experience With the Benefits of eHealth

Prior experience with health care created perceptions that eHealth can be beneficial and can be a necessity in health care. The benefits participants mentioned were less travel time, the opportunity for digital health care appointments at home, improved exchange of health information between health professionals and organizations, and an improved overview and insight into their personal health information.

> *The next time I visit my doctor and I need information, then beforehand I'll read [my file]. [...] I'd rather know what's on the record about me first. Maybe I can ask another sensible question. Because otherwise you're just a clueless patient.* [Participant 14, male, 76 years of age, IE]

XSL•FO
RenderX

Some participants also believed that eHealth could contribute to the efficiency of the health care system, which they deemed positive for themselves and society. Most participants were not familiar with the possibilities of eHealth before their health care experience. After experiencing these benefits, participants felt eager to learn about the possibilities of digital care. As a result, they sought nonformal support.

Attitude of Health Care Professionals Toward eHealth

A positive attitude toward eHealth or regular use of eHealth by health professionals motivated participants to use eHealth and learn skills that enable eHealth use. Some participants decided to do the course after a health professional suggested or used eHealth for their care.

> *Yes, Thuisarts.nl [a website that provides easily readable patient information on primary health care issues, issued by the Dutch College of General Practitioners], I consult that often. [...] That's what I agreed with the general practitioner at the time. He said, "Did you look on Google again?" [...] And then, "If you go to Thuisarts.nl, that's very reliable." And I just do that.* [Participant 12, female, 80 years of age IE]

Most participants indicated they never received any information about the possibilities of eHealth via their general practitioner or other health care professionals. Some participants argued that better promotion of eHealth is needed to spread awareness and stimulate use.

Theme 3—Need for Self-Reliance and Autonomy

Overview

Participants' motivation to seek help for eHealth use was also tied to their wish for independence. The need for self-reliance and autonomy was described by three aspects: (1) limited informal support, (2) discomfort with dependency on others, and (3) discomfort with loss of their own autonomy. This theme was identified equally among all participants, regardless of educational level. Thus, again, the social circumstances such as having children and the existence or absence of a social network seemed to affect the participants more directly than their educational background.

Limited Informal Support

Limited or inadequate informal support resulted in participants feeling forced to become independent users of eHealth. In a few cases, limited support meant that participants were childless, whereas they felt that society often wrongly assumed that their family would take care of them. This can create a sense of loneliness, but also motivation to learn these skills.

> *I do feel more or less obliged [to become independent in eHealth use]. [...] because I have nobody to fall back on. Everyone says: oh, you just ask your children. Well, I don't have any.* [Participant 1, female, 73 years of age, HE]

Other participants expressed they never felt a need to learn because their spouse would take care of all things digital. However, after their partner falls ill or has passed away, they are on their own and feel they have no other choice but to learn these skills themselves. In addition to the absence of informal support, some participants also experienced inadequate informal support, when the support they receive does not match their needs and wishes. This is especially true for the help the participants receive from their children or grandchildren, who are digital natives, and proceed to quickly resolve any issues. This is often done in such a way that the participants are left clueless about what the solution entails and are not stimulated to learn how to tackle these problems themselves. They continue to be dependent on others, which frustrates them. To be able to develop these skills on their own, they feel they need more structured explanations, time to practice, and patient teachers which the course offers them.

> *Otherwise, I always have to ask one of the grandchildren. […] They'll help me, but it's all quick, quick, quick. Well, that doesn't work. […] That's why I'm here taking this course.* [Participant 6, female, 66 years of age, LE]

Discomfort With Dependency on Others

To lean on others for help in eHealth use was uncomfortable for some participants due to their desire for independence, leading to their wish for independent eHealth use. Usually requesting help with small or acute matters within close ties such as family is acceptable, but depending on others outside of the family such as neighbors or friends feels uncomfortable and like they are burdening others. Some participants expressed that even asking family for help can feel like they are asking too much of their children, especially if they feel their children are busy with their own household, young children to take care of, or demanding jobs. The course offers a welcome alternative, where the participants do not feel like a burden.

> *Because I wanted to deal with it [eHealth] more easily. I do have children who want to help, but everyone still has their own thing. I don't want to become dependent on my kids.* [Participant 10, female, 78 years of age, HE]

Some also reported that they are afraid or find it difficult to ask for help from formal institutions such as the municipality. This, in turn, also has to do with a lack of trust and safety. Participants feel discouraged to ask for support from these types of organizations because they think they will not get sufficient guidance. Participants found commercial services to be too expensive, incompatible with their own equipment, insufficiently informative, or too time-consuming.

Discomfort With the Loss of Their Own Autonomy

Asking for help or being unable to use eHealth meant for some participants that they could not manage their own life and health. Their discomfort lies in their dependency on others rather than being uncomfortable burdening others with their need for help. It means the loss of autonomy which many of the participants experienced their whole lives through their employment, role in the household, and management of daily life. But as they get older, they feel they lose this freedom as it becomes harder to keep up with digital innovation. Many participants feel they want to be able to solve problems on their own, without asking

XSL•FO
RenderX

for help. Attending the course feel like they are in control and actively solving the problem by learning new skills and by doing something themselves. They experience this as more empowering than asking someone else to solve their problems for them.

> *Of course, I always wanted to do everything myself. Now you have to [ask for help] every time... That's the thing. Every time you have to give up a piece of your personality. That's how it feels.* [Participant 19, female, 74 years of age, LE]

In addition to these 3 themes related specifically to the motivation for eHealth use support, the participants expressed a general desire for social contact and life-long learning as a reason to participate in the eHealth course. With regards to social contact, participants reported finding solace and support to tackle digital (health) skills together with likeminded people, where they are not the only ones experiencing difficulty with new technology and have the opportunity to learn from each other's questions and mistakes without feeling like the odd one out. This feeling of kindred spirits can contribute to alleviating feelings of shame that participants may feel due to their lack of adequate digital skills. With regard to life-long learning, some participants proposed their curious personality as a reason to participate in all kinds of different courses offered to them by the public library. Most of these participants initially started on a path to improve their digital skills and this course was a natural continuation of that process to develop these further. Especially those participants who already had positive experiences with formal support at the public library were stimulated to continue with this type of support.

Discussion

Principal Findings

In this qualitative study, including participants of older age and with varying educational backgrounds, three motivational themes were identified for seeking help with eHealth use: (1) adapting to an increasingly digital society, (2) a sense of urgency facilitated by health care experience, and (3) a need for self-reliance and autonomy. A lack of adequate informal support from friends and family and positive experiences with nonformal support stimulated the participants to seek nonformal support in the public library for eHealth use. In addition to the 3 themes, participants expressed a more general desire for social contact with peers and lifelong learning.

The first theme "adapting to an increasingly digital society" explores the motivation to seek support driven by a wish or need for digital inclusion and societal participation in a digitalized society. It shows that the exponential development of digital technology can seem daunting and therefore insurmountable for some older adults who have little to no prior experience with this [49,50]. However, for others, it acts as a stimulus to stay engaged with societal changes, fueling the desire to adopt new technology [49,51]. In this study, this latter attitude also emerged as an important reason to seek support for eHealth use. While computer anxiety and low computer self-efficacy have been named important barriers to technology and eHealth adoption in older adults [52-54], this study shows that awareness of one's own limited self-efficacy in eHealth use can be an important incentive to seek support when individuals see the potential usefulness for themselves as "preparation for the future." Associated concepts such as healthy aging, aging in place (eg, living longer at home), and older person's empowerment have all been positively linked to the use of digital (health) technology [55-58] and are reflected in the third theme.

The second theme "a sense of urgency facilitated by health care experience" revolves around participants' prior experience with eHealth, health care, and health professionals. In this study, some participants experienced this as a necessity, while for others, this stimulated a positive attitude toward the potential benefits. Both acted as a facilitator for seeking support. Ware et al [55] equally found that a desire for ownership, access, and individual responsibility for personal medical information were motivators to use eHealth. Other research also underscores the importance of perceived eHealth benefits, positive prior eHealth experiences, and endorsement of health care professionals to motivate individual use [53,54]. This "window of opportunity" could be used to motivate others to seek nonformal support for their digital skills, for example, when introduced by trusted health professionals. However, this type of experience as a motivator for older adults to seek nonformal support for the development of digital (health) skills seems to be a novel finding.

The third theme "a need for self-reliance and autonomy" emphasizes that participants want to be self-sufficient in their abilities. This attitude toward self-reliance and autonomy illustrated how participants want to live and manage their health independently. It can support aging in place, encouraging older adults to live independently for as long as possible, a desire of most older adults and stimulated by policy makers [59,60]. Informal support by family and friends can both be a facilitator or barrier to the adoption of new digital technology [51,53,54]. In this study, part of the participants showed a dislike for dependency on family and friends and did not want to burden others turning informal support into a barrier, a sentiment also observed by others [51,53]. In this study, the help offered by family members did not lead to the development of more digital skills, as the participants desired. Therefore, they experienced limited and inadequate informal support. This particular phenomenon is seen elsewhere in informal support for internet use among older adults [61]. In addition, our findings on limited informal support are in line with others who find that intergenerational support can lack sufficient patience and understanding for the level of digital skills of older adults [51,53]. This may act as a barrier if it creates a learning environment characterized by judgmental delivery, a fast pace, and the use of unfamiliar jargon [54].

Theoretical reflection from the perspective of Maslow's Hierarchy of Needs provides an opportunity to create a deeper understanding of the origin of the identified motivations. Maslow's [62] theory focuses on human needs, theorizing 5 different types of needs. When these needs are not met, motivations to change behavior can emerge [62]. Furthermore, the theory implies a hierarchy of the 5 needs [62]. The basic needs are physiological needs (eg, water and food) and safety

XSL•FO
RenderX

needs (eg, a roof over one's head and the heating on) [62]. Then there are psychological needs, starting with the social need for love and belonging and moving up to the need for self-esteem [62]. Finally, there are the self-fulfillment needs, or the need for self-actualization, for example, developing one's full potential [62]. Themes 1 and 2 both reflect a safety need. The urgency to adapt to an increasingly digital society and the urgency of being a patient or caregiver show a need for security in access to (digital) health care now or in the future. The motivations from theme 3 can be interpreted as esteem needs, as this theme can be interpreted as a need for independence in access and use of the digital world and independence from support. The motivation stemming from the desire for lifelong learning can be interpreted as a need for self-actualization. The need for social connection can be interpreted as a social need. The perspective of Maslow's theory provides the insight that independent access to digital health care was perceived as a basic safety need and an esteem need by the participants.

Motivation to use eHealth and to seek support in eHealth use are 2 aspects of many when discussing digital health inequity. The digital divide model by van Dijk [10] describes how personal and contextual factors are related to the resources one has available to have motivation, physical access, use, and benefit from digital media, or in this case, digital health. Resources to improve the likelihood of accessing digital media involve cultural, social, mental, material, and time-related factors [10]. The study sample expressed a lack of digital skills and knowledge about digital health but also expressed access to up-to-date devices, language skills, time, and feeling comfortable in a public library setting. Our prior research and other research found that other populations experience barriers to seeking support in the factors that were facilitative for this study population [63-65]. Standaar et al [64], Choudhary and Bansal [63], and Goedhart et al [65] reported that other populations that need and potentially seek help in digital skills development experience barriers in language skills, digital skills, stigma, time, logistics, and feelings of belonging in the context of institutions. A literature review by Choudhary and Bansal [63] allowed for the categorization of the barriers to participate in training programs on the administrative level, training level, learner level, and community level. This categorization acknowledges that barriers are not only individual-bound but also bound to the training content, training context, and community in which training is provided [63]. Our findings suggest that access to up-to-date devices, language skills, time, and feeling comfortable in a public library setting were required to perform support-seeking behavior in the public library context. These reflections suggest that access to support for digital health skill development is also dependent on personal factors, contextual factors, and resources.

We explored the motivations for seeking eHealth support among participants. Additionally, we explored if differences in educational backgrounds led to differences in motivation. We found that other life events like losing connection with (digital) society, becoming a patient or caregiver, and becoming dependent on others later in life appeared to be more important in the emergence of their motivations to participate in an eHealth course. Even prior computer experience was not always an indication of the well-developed digital skills that are needed in the current digital age. Furthermore, exploration of the results through the perspective of education is complex in this older, mostly female sample as it may not clearly indicate educational differences due to historical cultural norms and educational opportunities for women in the 20th century [66]. Therefore, the results of this study provide very limited and nongeneralizable insights on different motivations to seek support in eHealth use in people from different educational backgrounds.

Strengths and Limitations

This study has several strengths. First, participants come from diverse backgrounds, including rural and urban areas enabling exploration of a wide range of experiences. A large percentage of the potential participants (20/48, 42%) participated in the study, ensuring a broad representation of the study population. Second, interviews continued until data saturation was reached, ensuring robust findings. Third, the eHealth course was not part of a research program, which allowed us to evaluate participants' motivations in a real-world setting. The on-site visits enhanced the understanding of the eHealth course context and allowed the interviews to take place in familiar surroundings for the participants. Our participant selection resulted in a culturally nondiverse older group with access to digital devices, motivation, and the capability to seek nonformal support, as a result of this specific research context. Next to this, it is likely that the people attending an eHealth course have some prior knowledge or experience with eHealth, the majority (n=14) of this study sample had some experience with eHealth. Other research in the same context confirms from digital health skill development trainers' perspective that the population attending the eHealth course is homogenous and similar to this study sample [64]. Asmar et al [67] refer to this specific group as "community supported." However, their and other research indicates that older adults are far from homogenous in their use of digital technology or support-seeking behavior [67-69]. Therefore, it is likely that in other settings, different typologies and characteristics of older adults might result in other motivations.

Implications and Future Research

This study highlights the importance of nonformal support in using eHealth, for those lacking adequate informal support. Kebede et al [54] found that attainment of digital competence among older adults depends on awareness of existing technology and the availability of support and instruction. They stated that personalized training and opportunities for need-based learning in a safe environment facilitate skill acquisition and initial digital engagement if the learning environment is accessible and inclusive for older adults [54]. Our findings indicate that public libraries can provide this type of support for digital health skill development. The results shed light on the motivations of participants, mainly older adults, to engage in nonformal support for eHealth use, advancing digital health literacy. While existing literature has focused on facilitators and barriers of digital (health) technology adoption by older adults, measuring eHealth or digital health literacy and the effectiveness of such interventions, often in research settings [53,54,70-72], this study

XSL•FO
RenderX

reveals motivations for participation in the real world. It highlights how motivations are associated with their specific needs for support and training, a need for security in access, and a desire for independent use of digital health care.

It would be interesting to explore the group of nonusers and users who do not seek nonformal support. Earlier reflections on our results suggested that this study population possessed certain characteristics, resources, and needs that align with seeking support in the public library context. Research shows that eHealth nonusers are common among other age and socioeconomic groups and that their needs for support and eHealth use are diverse [4,8,64,65,73-75]. Scholars find that populations rely on different types of support, it being formal, nonformal, and informal support, to gain access to digital skill development and eHealth [76-78]. Findings from Nygren and Hayat underline the importance of informal support to overcome digital literacy barriers. However, other findings indicate that digital inequality can manifest itself in informal support, where those who experience the most difficulties in accessing and use of digital tools are also those who have fewer resources for high-quality informal support [65,74]. Furthermore, van Deursen et al [79] show that in comparison to formal support, informal support might be insufficient to overcome all skill and content barriers for beneficial internet use. The barriers of limited access and poor quality of informal support were also found in this study, additionally, we found that discomfort with being dependent on others was also important. Smit et al [78] showed that Dutch citizens with low literacy overcome digital and literacy barriers by support from family, friends, community, and volunteers and base decisions on who they ask for support based upon the level of privacy surrounding a topic, the affective dimension involved, and the depth of knowledge possessed by the supportive actors [78]. These and our findings implicate that different populations are likely to seek support from different sources that fit their needs and personal circumstances. Further research is needed to understand the motivations, experiences, and needs in support of eHealth use by both digital technology users and nonusers. Findings can provide insights into what qualities eHealth support should entail to ensure equitable access to support and consequently to eHealth.

Conclusions

This study indicates that participants seeking nonformal support for eHealth use were driven by a need to secure independent access and use of digital health care. This need is derived from the necessity to adapt to a digital society, the expected or current need to use digital health, and a desire to independently engage in a digital society. In general, participants were motivated to develop digital health skills they feel are necessary for societal participation. Additionally, the eHealth course met the need for social connection and lifelong learning. The study sample was a predominantly homogenous group of older female adults with language skills, access to up-to-date devices, and time. It is likely that the need for support to gain independent access and use of digital health care is present in other populations. Future research should explore the needs and attitudes of users and nonusers of digital health toward support in eHealth access.

Acknowledgments

We would like to thank Anita Suijkerbuijk for her contributions during the set-up and execution of the study. We would also like to thank Sumit Mehra for his feedback and contributions to finalizing the article. This study was financially supported by the Dutch Ministry of Health for scientific publication and the Netherlands Organization for Health Research and Development (ZonMw project 555002013).

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide used to conduct semistructured face-to-face interviews with participants of the eHealth course DigiVitaler in the context of Dutch public libraries about their motivations to seek support in eHealth use.
[DOCX File, 16 KB - formative_v9i1e60612_app1.docx]

Multimedia Appendix 2
Demographics of the study participants interviewed between April and June 2022. Participants were categorized in educational level based upon the international standard classification of education level (ISCED). Rurality of the public library the participants attended the eHealth course was determined based upon zip code of the library and the indexed rurality by Netherlands Statistics.
[DOCX File, 19 KB - formative_v9i1e60612_app2.docx]

Multimedia Appendix 3

XSL•FO
RenderX

Overview of inductive themes and inductive topics and number of codes per topic derived from twenty interviews that were conducted with participants from an eHealth course in Dutch public library settings.
[DOCX File, 14 KB - formative_v9i1e60612_app3.docx]

Checklist 1
Standards for Reporting Qualitative Research checklist
[DOCX File, 16 KB - formative_v9i1e60612_app4.docx]

References

1. da Fonseca MH, Kovaleski F, Picinin CT, Pedroso B, Rubbo P. E-health practices and technologies: a systematic review from 2014 to 2019. Healthcare (Basel) 2021;9(9):1192. [doi: 10.3390/healthcare9091192]
2. Shaw T, McGregor D, Brunner M, Keep M, Janssen A, Barnet S. What is eHealth (6)? Development of a conceptual model for eHealth: qualitative study with key informants. J Med Internet Res 2017 Oct 24;19(10):e324. [doi: 10.2196/jmir.8106] [Medline: 29066429]
3. Abernethy A, Adams L, Barrett M, et al. The promise of digital health: then, now, and the future. NAM Perspect 2022;2022. [doi: 10.31478/202206e] [Medline: 36177208]
4. Goedhart NS, Verdonk P, Dedding C. “Never good enough.” A situated understanding of the impact of digitalization on citizens living in a low socioeconomic position. Policy Internet 2022 Dec;14(4):824-844. [doi: 10.1002/poi3.315]
5. Goedhart NS, Zuiderent-Jerak T, Woudstra J, Broerse JEW, Betten AW, Dedding C. Persistent inequitable design and implementation of patient portals for users at the margins. J Am Med Inform Assoc 2021 Feb 15;28(2):276-283. [doi: 10.1093/jamia/ocaa273] [Medline: 33463691]
6. Chesser A, Burke A, Reyes J, Rohrberg T. Navigating the digital divide: a systematic review of eHealth literacy in underserved populations in the United States. Inform Health Soc Care 2016;41(1):1-19. [doi: 10.3109/17538157.2014.948171] [Medline: 25710808]
7. Xu RH, Zhou LM, Wong EY, Wang D. The association between patients’ eHealth literacy and satisfaction with shared decision-making and well-being: multicenter cross-sectional study. J Med Internet Res 2021 Sep 24;23(9):e26721. [doi: 10.2196/26721] [Medline: 34559062]
8. Deursen A, Helsper EJ. The third-level digital divide: who benefits most from being online? In: Communication and Information Technologies Annual: Emerald Group Publishing Limited; 2015, Vol. 10:29-52. [doi: 10.1108/S2050-206020150000010002]
9. Al-Dhahir I, Reijnders T, Faber JS, et al. The barriers and facilitators of eHealth-based lifestyle intervention programs for people with a low socioeconomic status: scoping review. J Med Internet Res 2022 Aug 24;24(8):e34229. [doi: 10.2196/34229] [Medline: 36001380]
10. Van Dijk J. The Digital Divide: John Wiley & Sons; 2020.
11. Livingood WC, Bautista MAB, Smotherman C, et al. Comparative study of different SES neighborhood clinics for health literacy and internet access. Digital Health 2022;8:20552076221123715. [doi: 10.1177/20552076221123715] [Medline: 36081750]
12. Lythreatis S, Singh SK, El-Kassar AN. The digital divide: a review and future research agenda. Technol Forecast Soc Change 2022 Feb;175:121359. [doi: 10.1016/j.techfore.2021.121359]
13. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017 Apr 27;19(4):e136. [doi: 10.2196/jmir.6731] [Medline: 28450271]
14. Alvarez-Galvez J, Salinas-Perez JA, Montagni I, Salvador-Carulla L. The persistence of digital divides in the use of health information: a comparative study in 28 European countries. Int J Public Health 2020 Apr;65(3):325-333. [doi: 10.1007/s00038-020-01363-w] [Medline: 32277246]
15. Elena-Bucea A, Cruz-Jesus F, Oliveira T, Coelho PS. Assessing the role of age, education, gender and income on the digital divide: evidence for the European union. Inf Syst Front 2021 Aug;23(4):1007-1021. [doi: 10.1007/s10796-020-10012-9]
16. Hilbert M. The end justifies the definition: the manifold outlooks on the digital divide and their practical usefulness for policy-making. Telecomm Policy 2011 Sep;35(8):715-736. [doi: 10.1016/j.telpol.2011.06.012]
17. van Deursen AJ, van Dijk JA. The first-level digital divide shifts from inequalities in physical access to inequalities in material access. New Media Soc 2019 Feb;21(2):354-375. [doi: 10.1177/1461444818797082] [Medline: 30886536]
18. Standaar L, van Tuyl L, Suijkerbuijk A, Brabers A, Friele R. Differences in eHealth access, use, and perceived benefit between different socioeconomic groups in the Dutch context: secondary cross-sectional study. JMIR Form Res 2025:e49585. [doi: 10.2196/49585] [Medline: 39773883]
19. Faber JS, Al-Dhahir I, Reijnders T, et al. Attitudes toward health, healthcare, and eHealth of people with a low socioeconomic status: a community-based participatory approach. Front Digital Health 2021;3:690182. [doi: 10.3389/fdgth.2021.690182] [Medline: 34713165]

XSL•FO
RenderX

20. Frishammar J, Essén A, Simms C, Edblad R, Hardebro V. Older individuals and digital healthcare platforms: usage motivations and the impact of age on postadoption usage patterns. IEEE Trans Eng Manage 2023;70(8):2903-2919. [doi: 10.1109/TEM.2022.3187792]
21. Freeman T, Fisher M, Foley K, et al. Barriers to digital health services among people living in areas of socioeconomic disadvantage: research from hospital diabetes and antenatal clinics. Health Promot J Austr 2022 Jul;33(3):751-757. [doi: 10.1002/hpja.540] [Medline: 34510601]
22. Scheerder A, van Deursen A, van Dijk J. Determinants of internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. Telemat Inform 2017 Dec;34(8):1607-1624. [doi: 10.1016/j.tele.2017.07.007]
23. The ongoing journey to commitment and transformation: digital health in the WHO European region. World Health Organization. 2023. URL: https://www.who.int/europe/publications/m/item/digital-health-in-the-who-european-region-the-ongoing-journey-to-commitment-and-transformation [accessed 2025-01-23]
24. OECD skills outlook 2023: skills for a resilient green and digital transition. : OECD; 2023. [doi: 10.1787/27452f29-en]
25. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006;8(2):e9. [doi: 10.2196/jmir.8.2.e9]
26. Azzopardi-Muscat N, Sørensen K. Towards an equitable digital public health era: promoting equity through a health literacy perspective. Eur J Public Health 2019 Oct 1;29(Supplement_3):13-17. [doi: 10.1093/eurpub/ckz166] [Medline: 31738443]
27. Advice stimulating appropriate care and digital care [Web page in Dutch]. Rijksoverheid. 2022. URL: https://www.rijksoverheid.nl/documenten/rapporten/2020/07/16/advies-stimuleren-van-passende-zorg-en-digitale-zorg [accessed 2025-01-22]
28. Bibliotheekconvenant 2020-2023 [Web page in Dutch]. Bibliotheekconvenant. 2023. URL: https://www.bibliotheeknetwerk.nl/sites/default/files/documents/Bibliotheekconvenant%202020-2023_dt.pdf [accessed 2023-10-13]
29. Digital government information point [Web page in Dutch]. informatiepunt. 2024. URL: https://www.informatiepuntdigitaleoverheid.nl/?lang=en [accessed 2023-10-13]
30. Giest S, Samuels A. Administrative burden in digital public service delivery: the social infrastructure of library programs for e‐inclusion. Rev Policy Res 2023 Sep;40(5):626-645. [doi: 10.1111/ropr.12516]
31. van de Burgt A, Klaren M. Public library services related to basic skills for adults [Web page in Dutch]. Bibliotheeknetwerk. 2022. URL: https://www.bibliotheeknetwerk.nl/publicaties/basisvaardigheden-voor-volwassenen-2021 [accessed 2023-10-13]
32. Malcolm J. Informality and Formality in Learning: A Report for the Learning and Skills Research Centre: Learning and Skills Research Centre; 2003. [doi: 10.13140/RG.2.1.1297.3924]
33. Dicicco-Bloom B, Crabtree BF. The qualitative research interview. Med Educ 2006 Apr;40(4):314-321. [doi: 10.1111/j.1365-2929.2006.02418.x] [Medline: 16573666]
34. Bhattacherjee A. Social science research: principles, methods, and practices, global text project. Digital Commons @ University of South Florida. 2012. URL: https://digitalcommons.usf.edu/oa_textbooks/3 [accessed 2025-01-30]
35. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251. [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
36. Keij B, Versluis A, Alblas EE, Keuper JJ, van Tuyl LHD, van der Vaart R. E-healthmonitor 2023 stand van zaken digitale zorg [Report in Dutch]. : Bilthoven, Rijksinstituut voor Volksgezondheid en Milieu; 2023. [doi: 10.21945/RIVM-2024-0008]
37. Ministery of Economic Affairs and Climate Policy. Dutch Digitisation Strategy 2.0. Nederland Digitaal. 2019. URL: https://www.nederlanddigitaal.nl/documenten/publicaties/2019/11/13/english-version-of-the-dutch-digitalisation-strategy-2.0 [accessed 2023-05-05]
38. Regeerprogramma hoofdstuk 6a [Web page in Dutch]. Rijksoverheid.: Zorg; 2024. URL: https://www.rijksoverheid.nl/regering/regeerprogramma/6a-zorg [accessed 2025-01-30]
39. Peeters JM, Krijgsman JW, Brabers AE, Jong JDD, Friele RD. Use and uptake of eHealth in general practice: a cross-sectional survey and focus group study among health care users and general practitioners. JMIR Med Inform 2016 Apr 6;4(2):e11. [doi: 10.2196/medinform.4515] [Medline: 27052805]
40. Digital Economy and Society Index (DESI) 2021. European Commission. 2021. URL: https://ec.europa.eu/newsroom/dae/redirection/document/80493 [accessed 2023-10-31]
41. Basic skills dashboard for adults [Web page in Dutch]. Bibliotheek. URL: https://www.bibliotheeknetwerk.nl/dashboard/dashboard-basisvaardigheden-voor-volwassenen [accessed 2024-11-21]
42. Institute for Statistics International Standard Classification of Education: ISCED 2011. UNESCO. 2012. URL: https://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf [accessed 2024-05-16]
43. Population on January 1 and average; gender, age and region [Web page in Dutch]. Centraal Bureau voor de Statistiek. 2024. URL: https://www.cbs.nl/nl-nl/cijfers/detail/03759ned?dl=39E0B [accessed 2024-06-19]
44. de Vries H, Dijkstra M, Kuhlman P. Self-efficacy: the third factor besides attitude and subjective norm as a predictor of behavioural intentions. Health Educ Res 1988;3(3):273-282. [doi: 10.1093/her/3.3.273]
45. Fishbein M, Ajzen I. Predicting and Changing Behavior: The Reasoned Action Approach: Psychology Press; 2011.
46. de Vries H. An integrated approach for understanding health behavior; the I-change model as an example. Psychol Behav Sci Int J 2017;2(2):555-585. [doi: 10.19080/PBSIJ.2017.02.555585]

XSL•FO
RenderX

47. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019 Aug 8;11(4):589-597. [doi: 10.1080/2159676X.2019.1628806]
48. Braun V, Clarke V. Conceptual and design thinking for thematic analysis. Qual Psychol 2022;9(1):3-26. [doi: 10.1037/qup0000196]
49. Köttl H, Gallistl V, Rohner R, Ayalon L. "But at the age of 85? Forget it!": internalized ageism, a barrier to technology use. J Aging Stud 2021 Dec;59:100971. [doi: 10.1016/j.jaging.2021.100971] [Medline: 34794716]
50. Schwab K. The Fourth Industrial Revolution: Crown Currency; 2017.
51. Luijkx K, Peek S, Wouters E. "Grandma, you should do it—It's cool" older adults and the role of family members in their acceptance of technology. Int J Environ Res Public Health 2015 Dec 5;12(12):15470-15485. [doi: 10.3390/ijerph121214999] [Medline: 26690188]
52. Czaja SJ, Charness N, Fisk AD, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352. [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
53. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021 Aug 17;21(1):1556. [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
54. Kebede AS, Ozolins LL, Holst H, Galvin K. Digital engagement of older adults: scoping review. J Med Internet Res 2022 Dec 7;24(12):e40192. [doi: 10.2196/40192] [Medline: 36477006]
55. Ware P, Bartlett SJ, Paré G, et al. Using eHealth technologies: interests, preferences, and concerns of older adults. Interact J Med Res 2017 Mar 23;6(1):e3. [doi: 10.2196/ijmr.4447] [Medline: 28336506]
56. Ollevier A, Aguiar G, Palomino M, Simpelaere IS. How can technology support ageing in place in healthy older adults? A systematic review. Public Health Rev 2020 Nov 23;41(1):26. [doi: 10.1186/s40985-020-00143-4] [Medline: 33292707]
57. Robbins TD, Lim Choi Keung SN, Arvanitis TN. E-health for active ageing: a systematic review. Maturitas 2018 Aug;114:34-40. [doi: 10.1016/j.maturitas.2018.05.008] [Medline: 29907244]
58. Mcmellon CA, Schiffman LG. Cybersenior empowerment: how some older individuals are taking control of their lives. J Appl Gerontol 2002 Jun;21(2):157-175. [doi: 10.1177/07364802021002002]
59. Wiles JL, Leibing A, Guberman N, Reeve J, Allen RES. The meaning of "aging in place" to older people. Gerontologist 2012 Jun;52(3):357-366. [doi: 10.1093/geront/gnr098] [Medline: 21983126]
60. Longer at home program [Report in Dutch]. : Den Haag: Ministerie van Volksgezondheid, Welzijn en Sport; 2018 URL: https://www.rijksoverheid.nl/documenten/rapporten/2018/06/15/programma-langer-thuis [accessed 2025-04-20]
61. Helsper EJ, van Deursen A. Do the rich get digitally richer? Quantity and quality of support for digital engagement. Inf Commun Soc 2017 May 4;20(5):700-714. [doi: 10.1080/1369118X.2016.1203454]
62. Maslow AH. A theory of human motivation. Psychol Rev 1943;50(4):370-396. [doi: 10.1037/h0054346]
63. Choudhary C, Nidhi B. Barriers affecting the effectiveness of digital literacy training programs (DLTPs) for marginalised populations: a systematic literature review. J Tech Educ Train 2022 Jun 30;14(1):110-127. [doi: 10.30880/jtet.2022.14.01.010]
64. Standaar LMB, Israel AMC, van der Vaart R, et al. Support in digital health skill development for vulnerable groups in a public library setting: perspectives of trainers. Front Digital Health 2024;6:1519964. [doi: 10.3389/fdgth.2024.1519964] [Medline: 39872241]
65. Goedhart NS, Broerse JEW, Kattouw R, Dedding C. ' Just having a computer doesn't make sense': the digital divide from the perspective of mothers with a low socio-economic position. New Media Soc 2019 Nov;21(11-12):2347-2365. [doi: 10.1177/1461444819846059]
66. Leyenaar MH. From pioneers to players: women in the netherlands. women and politics around the world. In: Gelb J, Palley ML, editors. A Comparative History and Surve: Bloomsbury Academic; 2009, Vol. 2:485-502.
67. Asmar A, Van Audenhove L, Mariën I. Social support for digital inclusion: towards a typology of social support patterns. Soc Incl 2020 May;8(2):138-150. [doi: 10.17645/si.v8i2.2627]
68. Quan-Haase A, Williams C, Kicevski M, Elueze I, Wellman B. Dividing the grey divide: deconstructing myths about older adults' online activities, skills, and attitudes. Am Behav Sci 2018 Aug;62(9):1207-1228. [doi: 10.1177/0002764218777572]
69. Petrovčič A, Reisdorf BC, Grošelj D, Prevodnik K. A typology of aging internet users: exploring digital gradations in internet skills and uses. Soc Sci Comput Rev 2023 Oct;41(5):1921-1940. [doi: 10.1177/08944393221117753]
70. Karnoe A, Kayser L. How is eHealth literacy measured and what do the measurements tell us? A systematic review. Knowl Manag E-Learn 2015 Dec 15;7(4):576-600. [doi: 10.34105/j.kmel.2015.07.038]
71. Kim H, Xie B. Health literacy in the eHealth era: a systematic review of the literature. Patient Educ Couns 2017 Jun;100(6):1073-1082. [doi: 10.1016/j.pec.2017.01.015] [Medline: 28174067]
72. Xie L, Zhang S, Xin M, Zhu M, Lu W, Mo PH. Electronic health literacy and health-related outcomes among older adults: a systematic review. Prev Med 2022 Apr;157:106997. [doi: 10.1016/j.ypmed.2022.106997]
73. Helsper EJ, Eynon R. Distinct skill pathways to digital engagement. Eur J Commun 2013 Dec;28(6):696-713. [doi: 10.1177/0267323113499113]
74. Courtois C, Verdegem P. With a little help from my friends: an analysis of the role of social support in digital inequalities. New Media Soc 2016 Sep;18(8):1508-1527. [doi: 10.1177/1461444814562162]

XSL•FO
RenderX

75. Kaihlanen AM, Virtanen L, Buchert U, et al. Towards digital health equity—a qualitative study of the challenges experienced by vulnerable groups in using digital health services in the COVID-19 era. BMC Health Serv Res 2022 Feb 12;22(1):188. [doi: 10.1186/s12913-022-07584-4] [Medline: 35151302]
76. Nygren H, Nissinen K, Hämäläinen R, De Wever B. Lifelong learning: formal, non‐formal and informal learning in the context of the use of problem‐solving skills in technology‐rich environments. Brit J Educ Tech 2019 Jul;50(4):1759-1770. [doi: 10.1111/bjet.12807]
77. Hayat TZ, Brainin E, Neter E. With some help from my network: supplementing eHealth literacy with social ties. J Med Internet Res 2017 Mar 30;19(3):e98. [doi: 10.2196/jmir.6472] [Medline: 28360024]
78. Smit A, Swart J, Broersma M. Bypassing digital literacy: marginalized citizens' tactics for participation and inclusion in digital societies. New Media Soc 2024;0. [doi: 10.1177/14614448231220383]
79. van Deursen AJAM, Courtois C, van Dijk JAGM. Internet skills, sources of support, and benefiting from internet use. Int J Hum Comput Interact 2014 Apr 3;30(4):278-290. [doi: 10.1080/10447318.2013.858458]

Abbreviations

ISCED: International Standard Classification of Education
SRQR: Standards for Reporting Qualitative Research

Edited by A Mavragani; submitted 17.05.24; peer-reviewed by N Goedhart, T Huang; revised version received 30.01.25; accepted 05.03.25; published 28.04.25.

Please cite as:
Standaar L, Israel AM, van der Vaart R, Keij B, van Lenthe FJ, Friele R, Beenackers MA, van Tuyl LHD
The Motivations of Citizens to Attend an eHealth Course in the Public Library: Qualitative Interview Study
JMIR Form Res 2025;9:e60612
URL: https://formative.jmir.org/2025/1/e60612
doi:10.2196/60612

© Lucille Standaar, Adriana Margje Israel, Rosalie van der Vaart, Brigitta Keij, Frank J van Lenthe, Roland Friele, Mariëlle A Beenackers, Lilian Huibertina Davida van Tuyl. Originally published in JMIR Formative Research (https://formative.jmir.org), 28.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Adolescent Emoji Use in Text-Based Messaging: Focus Group Study

Matt Minich, PhD; Bradley Kerr, MS; Megan Moreno, MD, MPH, MSED

Department of Pediatrics, School of Medicine and Public Health, University of Wisconsin-Madison, 2870 University Ave, Madison, WI, United States

Corresponding Author:
Matt Minich, PhD
Department of Pediatrics, School of Medicine and Public Health, University of Wisconsin-Madison, 2870 University Ave, Madison, WI, United States

Abstract

Background: Adolescents increasingly communicate through text-based messaging platforms such as SMS and social media messaging. These are now the dominant platforms for communication between adolescents, and adolescents use them to obtain emotional support from parents and other adults. The absence of nonverbal cues can make it challenging to communicate emotions on these platforms, however, so users rely on emojis to communicate sentiment or imbue messages with emotional tone. While research has investigated the functions of emojis in adult communication, less is known about adolescent emoji use.

Objective: This study sought to understand whether the pragmatic functions of adolescent emoji use resemble those of adults, and to gain insight into the semantic meanings of emojis sent by adolescents.

Methods: Web-based focus groups were conducted with a convenience sample of adolescents, in which participants responded to questions about their use and interpretation of emojis and engaged in unstructured interactions with one another. Two trained coders analyzed transcripts using a constant comparative coding procedure to identify themes in the discussion.

Results: A total of 6 focus groups were conducted with 31 adolescent participants (mean age 16.2, SD 1.5 years). Discussion in the groups generally fell into 4 themes: emojis as humorous or absurd, emokis as insincere or complex expressions of setiment, emojis as straightforward experssions of sentiment, and emojis as having context-dependent meanings. Across themes, participants often described important differences between their own emoji use and emoji use by adults.

Conclusions: Adolescent focus group participants described patterns of emoji use that largely resembled those observed in studies of adults. Like adults, our adolescent participants described emojis' semantic meanings as being highly flexible and context-dependent. They also described both phatic and emotive functions of emoji use but described both functions in ways that differed from the patterns of emoji use described in adult samples. Adolescents described their phatic emoji use as absurd and described their emotive emoji use as most often sarcastic. These findings suggest that emoji use serves similar pragmatic functions for both adolescents and adults, but that adolescents see their emoji use as more complex than adult emoji use. This has important implications for adults who communicate with adolescents through text-based messaging and for researchers interested in adolescents' text-based communication.

(JMIR Form Res 2025;9:e59640) doi:10.2196/59640

KEYWORDS

communication; text messaging; smartphones; emoji; focus groups; adolescent; teen; youth; teenagers; text; phone; messaging; text communication; emotion

Introduction

Overview

Adolescent communication is increasingly mediated by text-based communication platforms like SMS or social media messaging. About 95% of US adolescents aged 13 - 17 years have a smartphone for personal use [1], and adolescents in this age group send and receive an average of 67 text messages each day [2]. Text messaging is now the dominant mode of communication between adolescents [2] and is increasingly important to adolescents' relationships with parents and other adults [3]. For example, text-based communication is an effective way for healthcare practitioners to reach young or adolescent clients [4] and can help therapists discuss sensitive topics with younger clients [5]. Many parents also surveil the text exchanges of their adolescent children—in a nationwide survey, 64% of parents said they regularly checked the contents of their child's phone, including text message logs [6]. Thus, many adolescents' most important relationships are affected in some way by the features of text-based messaging platforms.

One feature that applies to text-based messaging of all kinds is the absence of nonverbal cues, such as body language or tone

XSL•FO
RenderX

of voice, that people use to convey emotional context [7]. Users often address this deficit with emojis, which can represent emotional states through pictograms [8]. However, emojis can create confusion because users often interpret them differently [9]. Past research has addressed this confusion both by identifying patterns in emojis' pragmatic functions [10] and by reducing ambiguity around emojis' semantic meanings [11,12], but that work has largely relied on samples of adult emoji users. Because emoji use and interpretation differ across age groups [13,14], it is unclear whether the insights from this research can be extended to adolescents. Given the emerging importance of text-based messaging to adolescent communication, this study explores adolescents' perspectives on their own emoji use. Its aims are twofold: to understand whether the pragmatic functions of adolescent emoji use resemble those of adults and to gain insight into the semantic meanings of emojis sent by adolescents.

Pragmatic Functions of Adult Emoji Use

The earliest known use of emotive pictograms in computer-mediated communication was by Carnegie Mellon Professor Scott Fahlman, who proposed in 1982 that users of an intradepartmental message board distinguish serious posts from humorous ones by marking them with the symbols :-) or :-([8]. Though similar symbols (called emoticons) are still used, they have since been largely supplanted by the Unicode emoji library, which gives users access to a set of more than 3600 pictographs that are mostly standardized across devices [15]. The use of emojis is now nearly ubiquitous—more than 5 billion emojis are exchanged each day on Facebook alone [15], and the "face with tears of joy" emoji ( ) was chosen as the Oxford English Dictionary's Word of the Year in 2015.

When asked about their own emoji use, adults say they use the pictograms both to express emotions directly and to modify the tone of messages. For example, surveyed English-speaking adults said they understand emojis as tools for tone modification, in which the emoji modifies text and clarifies its interpretation [16]. US adults have reported using emojis primarily to express sentiment, to strengthen expressions, or to adjust the tone of messages [17]. Thus, the utility of emojis for adults largely overlaps with that of emoticons, which adults also use to communicate feelings and reduce the ambiguity of text [18]. For emojis and emoticons alike, surveyed adults have said they send pictograms primarily to express their feelings, to strengthen the content of a message, or to imbue a message with an element of comedy or fun [19]. Notably, adults' reasons for sending emojis differ slightly across age groups, with younger adults reporting a more diverse set of motives (including, for example, a desire to make a message ironic or sarcastic) than older adults [19]. Research has also found that older adults use a less diverse emoji vocabulary than younger adults and are generally less likely to use emojis in text-based communication [20].

Theorists have suggested adult emoji use can be broadly categorized as serving 1 of 2 pragmatic functions: a phatic function or an emotive function [10]. The phatic function of emojis describes their use as a fill-in for "small talk:" semiverbal gestures that maintain connection and keep communication light and friendly but lack specific semantic meanings. For example, a user might send a simple smile emoji ( ) to open or close a conversation or to fill an uncomfortable period of silence. This aligns with research finding that emojis are usually associated with positive sentiments [21], and that they can be used to facilitate connections in a way that is agnostic to a conversation's emotional content [22]. The emotive function of emojis describes their use to convey the sender's emotional state or to imbue a message with emotional context. This may be done by using emojis in place of words describing emotions (for example, by sending in place of the words "I'm angry"), or by adding emojis to complete statements (eg, "I saw a car accident today ").

Semantic Ambiguity of Emojis

Though users often send emojis to reduce ambiguity about messages' emotional context, emojis can create confusion because receivers interpret them differently. For example, when experimental participants were asked to interpret expressive face emojis, they disagreed on even basic sentiment assignments (positive, negative, or neutral) 25% of the time [23]. This confusion persisted even when the emojis were accompanied by relevant text. Similarly, experimental participants tended to rate text messages paired with face emojis as more emotionally ambiguous than messages paired with nonface emojis [21]. When experimenters asked participants to interpret emojis contained within text messages, participants offered a wide variety of interpretations. For example, some participants understood the loudly crying emoji ( ) to signify sadness while others perceived this emoji as sarcastic or exaggerated [24].

Research has shown that the interpretation of emojis can be affected by the ages of the sender and receiver. For example, older adults were found more likely to interpret emojis as having positive meanings [14], and younger Chinese adults were more likely to use "positive emojis" (such as the simple smile) to express negative sentiments [25]. Older adults have also proven less likely than younger adults to recognize emojis as indicators of sarcasm [26], and more likely to express confusion about emojis or choose literal interpretations of their meaning [13]. Older adults have reported less confidence in their interpretations of emojis and perceptions that emojis are difficult to use correctly [20].

Understanding Adolescent Emoji Use

Among adults, research shows that both the pragmatic functions and semantic interpretations of emojis differ as a factor of age. Older adults have reported less diverse motivations for using emojis [19] than younger adults, for example, and more difficulty interpreting emojis they receive in messages from others [13,20]. This suggests that adolescent emoji use might differ from adult emoji use in meaningful ways, which can create challenges both for adults seeking to communicate with adolescents through text-based messaging and for researchers interested in text-based communication between adolescents.

Understanding how adolescents use and interpret emojis is particularly important because text-based messaging has emerged as an essential platform for adolescent communication. Text messaging behavior peaks around 11th grade [27] and slows in young adulthood [28], leading some scholars to suggest that the affordances of this medium may make it particularly

XSL•FO
RenderX

well-suited for adolescents' developmental needs. Indeed, the asynchronous and socially distanced nature of text messaging may facilitate adolescents' desires for autonomy, identity development, and the formation of intimate relationships [27]. Text-based messaging is also increasingly important to communication between adolescents and adults. Adolescents often use these platforms to request emotional support from parents, for example, and have reported improvements in their offline relationships with parents on days when they exchange more text messages with them [3]. Adolescents have expressed desires to use text-based messaging platforms for communication with adults about emotional support [4] and have said text-based platforms provide them with a greater sense of authority and control in interactions with counselors and therapists [29].

Emotive pictographs like emojis have long been used to communicate emotions through text [8], but age-related differences in the function and interpretation of emojis could facilitate miscommunication. This could exacerbate existing problems with communication between adolescents and adults, such as communication between adolescents and their parents. Parents and adolescents often fail to accurately understand one another's cognitive and affective states even when communicating face-to-face [30,31]. This miscommunication has been described as a failure of empathic accuracy [32], or the accurate prediction of a conversation partner's thoughts and feelings. Adults often use emojis to improve empathic accuracy [33] in text-based exchanges, but it is unclear whether adolescents use them for a similar purpose. Thus, this study seeks to better understand how adolescents use emojis in these exchanges, and how they understand the semantic meanings of those emojis. Given the persistence of emotive pictographs over decades of computer-mediated communication [8], these insights should remain useful even as communication platforms evolve.

Methods

Participants

Adolescent focus group participants were recruited through a node distribution method, in which contacts of the research team (staff at youth-serving organizations and participants in previous studies) used word-of-mouth to inform potential participants about the study. All participants provided assent and parental consent and then completed a demographic survey that assessed participants' age, gender, race, and ethnicity.

A total of 31 adolescents participated in the focus groups. A total of 48% of participants were female, and the average participant age was 16.2 (SD 1.5) years. Complete demographic information is presented in Table 1 below.

Table . Participant demographics. Self-reported demographic data of participants in 6 web-based focus groups conducted to identify themes in adolescents' understandings of their own use of emojis in text-based messaging.

Characteristic	Values, n (%)
Gender	
Cisgender female	14 (45.2)
Cisgender male	10 (32.3)
Non-binary	2 (6.5)
Transgender female	1 (3.2)
Transgender male	1 (3.2)
Other	1 (3.2)
Did not disclose	2 (6.5)
Ethnicity	
Non-Hispanic or non-Latino	26 (83.9)
Hispanic or Latino	3 (9.7)
Did not disclose	2 (6.5)
Race	
Caucasian or White	16 (51.6)
Asian or Asian-American	6 (19.4)
American Indian or Alaska Native	1 (3.2)
Black or African-American	1 (3.2)
Native Hawaiian or other Pacific Islander	0 (0.0)
Other	0 (0.0)
More than one race	5 (16.1)
Did not disclose	2 (6.5)

XSL•FO
RenderX

Data Collection Procedure

Adolescent perspectives were collected through 6 web-based focus groups, which allowed participants to express their experiences without restriction and encouraged them to build on one another's contributions. Participants completed a brief demographic questionnaire before joining the group. Groups were conducted using a digital conference call service by 2 trained facilitators (the first author and second author) and lasted up to 90 minutes. Both facilitators were White males employed at a large midwestern research institution, who both possessed MS degrees at the time of data collection. In focus groups, facilitators initially asked open-ended questions about emoji use and interpretation (eg, "What are some ways you use emojis in your communication?"), as well as specific questions about styles of emoji use (eg, "Do you usually use emojis to replace words or phrases, or to add emphasis?"). Participants were also shown several emojis, including the simple smile ( ) and 5 randomly selected face emojis (in the iPhone iOS style) asked to describe their interpretations of those emojis and the circumstances in which they might share them.

Ethical Considerations

The procedure used to collect data for this study was approved by the University of Wisconsin-Madison Health Sciences Institutional Review Board (#2019 - 0839), and no adverse events were reported. Informed consent from parents or guardians and informed assent from adolescent participants were obtained through a digitally delivered form before focus group participation. Informed consent documents clarified that deidentified focus group data could be made available to investigators outside the original study team. Privacy and confidentiality of participant data were protected by redacting all identifying information from focus group transcripts before analysis. All documents containing identifiable data (eg, informed consent documents and original focus group recordings) were stored on a secure server only accessible by members of the study team. Participants were compensated US $40 for participation in the study, regardless of whether they completed their focus group participation (no participants chose to leave focus groups early).

Analytical Approach

Analysis was performed using a constant comparative procedure, in which 2 trained coders applied inductive reasoning to identify themes [34]. Originally developed for use in Grounded Theory methodologies [35], this procedure allows analysts to identify themes from complex qualitative data sets without relying on a priori hypotheses. This approach was appropriate for this study both because of the dearth of past research into adolescent emoji use and because of the relatively unstructured nature of the data, which included not only responses to researchers' prompts but also unplanned interactions between participants.

After 3 focus groups had been conducted, coders began reviewing focus groups for theoretical saturation: a point at which new ideas or themes no longer emerged from discussions. Coders identified theoretical saturation after the completion of 6 focus groups, signaling the end of data collection and the beginning of thematic analysis.

Thematic analysis was conducted in 3 stages. First, coders independently read transcripts from 2 focus groups, each developing their own draft codebook. The coders then compared and discussed these codebooks, consolidating them into a single codebook that both coders agreed upon. This codebook contained a set of superordinate "branch" codes, each of which encompassed a set of more specific, subordinate "leaf" codes.

In the second stage of thematic analysis, coders refined the codebook by independently coding the same 2 additional transcripts and then meeting to compare results. Coders identified and discussed any discrepancies in their coding of these transcripts, then agreed on codebook revisions to resolve them. When coders agreed that all areas of confusion had been addressed, they divided and independently coded the remaining transcripts, so that each transcript was coded by a single coder. While coding this final set of transcripts, coders held regular meetings to discuss any further need for codebook revisions and to reach an agreement on the coding of any ambiguous statements.

In the third stage of thematic analysis, coders met to review the coded transcripts and identify overarching themes and subthemes. Coders reread excerpts assigned to each superordinate "branch" and subordinate "leaf" code, then independently collapsed these codes into larger "theme" groups. Coders then met to discuss the similarities and differences of the 2 sets of themes, consolidating them into a single set of shared themes. These themes were then reviewed by an independent youth advisory board consisting of 6 adolescents, who affirmed these themes as a reasonable summary of transcript content.

Finally, after thematic analysis was complete, coders considered how the agreed-upon themes aligned with their a priori areas of theoretical interest: identifying pragmatic functions of emoji use (phatic and emotive) and resolving the semantic ambiguity of emojis.

Results

Overview

Coders identified 4 emergent themes in focus group transcripts. One described the phatic function of emojis: "emojis as humorous or absurd", and 2 described their emotive functions: "emojis as insincere or complex expressions of sentiment and emojis as straightforward expressions of sentiment". Finally, one theme spoke to the semantic ambiguity of emojis: "emojis as having context-dependent meanings".

Notably, some statements made by participants were classified as pertaining to more than one theme. The extent of this crossover, as well as the total number of statements pertaining to each theme, is recorded in Table 2 below.

XSL•FO
RenderX

Table . Distribution of participant statements across themes. A matrix denoting the number of statements made by adolescent focus group participants coded for each of the 4 themes identified during analysis. Intersecting columns and rows represent the number of statements coded as 2 themes (eg, column 1 of row 2 describes the number of statements coded as both theme 1 and theme 2).

	Theme 1	Theme 2	Theme 3	Theme 4
Theme 1	39	6	2	11
Theme 2	6	81	4	11
Theme 3	2	4	57	5
Theme 4	11	11	5	46

Theme 1: Emojis as Humorous or Absurd

Overview

When participants described phatic uses of emojis (ie, use of emojis to maintain connection without giving them semantic meaning), they described doing so in the context of humor or absurdity. In these cases, it was understood that the sender or receiver found humor in the act of using emojis itself.

Theme 1a: Emojis as Innately Humorous

Several participants said they understood emojis as being inherently humorous and described sending emojis for the intention only of communicating humor or social connection. The emojis themselves in these communications were either divorced from their original semantic meanings or were assumed to have no semantic meaning. In most cases, participants described the absence of semantic meaning as essential to the emoji's communicative function.

> *I think, like there's one, which is just like, as a red background and like a B [ ], and they were, I mean, they'd seemingly like have no meaning. And they still don't really like, take on that big of a meaning, but they were like the randomness for like, used for, um, like comedic purpose or for like an inside joke kind of thing.*

> *…The sparkle [ ], or the pixie, dust, I think was originally used to like enhance the meaning of something, or like making it aesthetic in some way. But now I'm just using it for random words that it doesn't really make sense in—but it's like funny in a way (laughs). It's kind of like ironically…*

> *…there was like the chair emoji [ ], which like people would just use it to be funny and like the point of it was that it had no meaning, but like it was just supposed to be like a random emoji with no meaning.*

Theme 1b: Emoji Spam

Participants also described engaging in a specific pattern of absurd emoji use called "Emoji spamming," in which they sent long strings of randomly selected emojis. Again, participants described these strings as not only lacking any specific semantic meaning but described their communicative function as being innately tied to their absence of meaning.

> *…Emoji spamming often just signifies, like, chaos. Like, I think it's often used to, like, maybe you saw something outside and you wanna say something about it sarcastically, so you take a picture of it... Maybe, like, let's say you take a picture of a car crash, then you're like, "Oh, this is such a pretty car," and then you emoji spam. It's just, like, used to be sarcastic and funny.*

> *…I know definitely some of my friends do this, like, if they're typically experiencing, like, a negative emotion, they'll like emoji spam and do a bunch of random, like, really random ones…*

> *…People, like, sometimes, like, spam emojis because they're, like, angry that they lost [a mobile game] or they're excited that they won. I really think it goes both ways…*

Theme 2: Emojis as Insincere or Complex Expressions of Sentiment

Overview

Participants tended to describe emojis as inappropriate for communicating sincere emotions and mood, though some did acknowledge using them for this purpose (see theme 3). Rejection of sincere emoji use was generally framed as (1) a broad aversion to the use of emojis, (2) a preference for the ironic or sarcastic use of emojis, or (3) the use of apparently positive emojis to communicate negative emotions or moods.

Theme 2a: Aversion to the Use of Emojis

Many participants said they did not use emojis frequently, and they often specified that they were less likely to use emojis to communicate negative emotions or in emotionally negative situations, when they described emojis as socially inappropriate:

> *…When I'm, like, in a serious conversation I don't usually use emojis because I feel, like, it's not as serious if I were to use emojis.*

> *I don't really use emojis all that much. So if I am feeling those [negative] emotions, I'll probably just verbalize and like text them instead of using emojis for it, since it's kind of probably more of a serious conversation.*

Theme 2b: Preference for Ironic or Sarcastic Use of Emojis

Many participants said they typically used emojis in ironic or sarcastic ways so that their intended meaning was the opposite of a literal interpretation. For example, participants might use a laughing emoji to communicate that a joke had failed to land or might use a "cool" emoji [ ] to lampoon a person or message:

XSL•FO
RenderX

> *..With the sunglasses, I kind of use that sarcastically, kind of, like, saying, "Oh, I'm so cool I did that." Like, "Oh, that's so cool I'm doing something."*
>
> *I am one of those people that use this emoji [ ] sarcastically because, like, I feel like if I wanted to show that I was happy through emoji I would put, like, the closed eye smiling [ ] or something but this one seems, like, kind of creepy when you look at it so I usually use it sarcastically.*
>
> *…Just people in general, like, they'll use an emoji and me and my friends will, like, use that emoji [ ] to make fun of the people who use them, like, genuinely. Like [we'll] use it as, like, satire I guess you could say because we think it's funny that people actually use them.*

Mentions of ironic/sarcastic emoji use were notably more frequent in most focus groups than mentions of sincere emoji use.

Theme 2c: Communicating Negative Moods With Positive Emoji

When prompted to describe their use of traditionally positive emojis, participants often said they used these emojis to communicate negative emotions. Participants described positive emojis as adding layers of irony or insincerity to negative emotions, or to communicate a reluctant acceptance of negative circumstances.

> *I don't know if ironic is the right word, but I would use it [ ] in the opposite sense. Like, you're smiling through the pain or something. I know it's terrible, but that's just how I would see them.*
>
> *a lot of times, you could use that emoji [ ] as, like, I don't know, like, you're smiling through the pain, or, like, used, like, sarcastically, saying, like, "Oh, I'm so happy I messed up on something."*
>
> *I guess non-adolescents that don't really like use emojis ironically would definitely think that like [ ] is extremely like happy and positive. But, I mean, really, it's not. It's pretty lifeless.*
>
> *I feel like, it also goes with like a sarcastic tone too in kind of a joking manner. Like I texted my best friend last night and I was like "It's 5:00a.m. and I'm awake "*

Theme 3: Emojis as Straightforward Expressions of Sentiment

Though participants generally agreed that adults were more likely to use emojis to communicate sincere emotions, some described instances where they used emojis without irony or humor. In many such cases, they described the emoji as referring to a nameless, specific circumstance or emotional state:

> *Maybe I, I dropped my like pencil in the hallway and I had to like pick it up, like I would use that emoji [ ] like being like, "Oh yeah, so stressed about it."*
>
> *I'll probably just use like an emoji, like a heart or something, or a smiley. Or like something like that to just, you know, convey the message that, you know, it's how I'm feeling.*
>
> *I use emojis when I don't want to write out like everything. Like if I think something's funny I can just do like a laughing emoji [ ] and yeah, it's just easier to get the point across without having to write.*

Theme 4: Emojis as Having Context-Dependent Meanings

Finally, participants described emojis as having flexible meanings that shifted to suit the audience or conversational context. They cautioned against assigning specific meanings to emojis and said that they anticipated audience responses when choosing emojis.

> *... For the most part, there's no real universal meaning for most emojis. Like, I'd say there's a few that, like, most people always use to signify a few things, but for the most part, they all mean different things for different people in different scenarios.*
>
> *You have to think about and interpret how another person is going to receive that emoji. And so, when you are with your friends, or are texting your friends and you're using the emojis like ironically, you know that they're going to interpret those emojis ironically. But like with adults, I feel like, um, especially with adolescents, we, um, you know—that they're not going to interpret them the same way we do, so we take it more literally.*

Many participants stressed differences between adult and adolescent styles of emoji use, usually emphasizing how these groups differ in their preferences for sincerity or authenticity. Specifically, participants said they were more likely to use sarcasm or humor when sending emojis to peers, while they expected adults to interpret any emoji as a sincere expression of sentiment.

> *…Speaking of like older generations who use, who use like messaging or whatever, I like use the emojis in a different way than I do with my friends. Like with older people, I'll just use them in like an actual way of what they're like theoretically supposed to mean. Like a smiley face will actually mean a smiley face…*

Participants also said that certain emojis held meanings within their social groups that they knew were not shared by other groups. Though this was sometimes described as extending to entire geographic regions, these meanings were most often described as specific to groups of friends:

> *Almost everyone has different emojis that mean something to them with, like, another friend, like, an inside joke or, like, another meaning.*

Discussion

Principal Findings

This focus group study explored the ways adolescents use and interpret emojis, which has received less scholarly attention than adults' emoji use. By collecting data through focus groups and applying a coding strategy informed by Grounded Theory

XSL•FO
RenderX

[35], this work was the first to our knowledge to examine adolescent emoji use through adolescents' own words. The 4 themes that emerged in our analysis suggest that adolescent emoji use has many similarities with adult use but that it differs in some important ways. First, adolescents described the phatic functions of emojis that were informed by a sense of humor or absurdity (theme 1). Second, adolescents confirmed the emotive function of emojis, which they described as sometimes being achieved through a layer of sarcasm or irony (themes 2 and 3). Finally, they described emojis as having highly variable semantic meanings that are often understood differently by adolescents and adults (theme 4). Taken together, these findings suggest adolescents and adults use emojis for similar reasons (ie, phatic and emotive uses), but that adolescent emoji use is informed by humor, irony, and social dynamics that may make these messages difficult for adults to interpret.

Adolescents' Phatic Emoji Use is Humorous and Absurd

Focus group participants described emoji uses that seem to take advantage of emojis' phatic functions (ie, to maintain connection without providing any specific semantic meaning), but their use of this function was distinct from phatic use by adults. For example, adolescents said they frequently sent emojis that were meant to have no semantic meaning at all (theme 1a) or sent strings of randomly selected "emoji spam," or single emojis with no meaning outside of their repeated use (theme 1b). These phatic uses are notably different from the phatic uses observed in adult messaging [10], which include opening a conversation with a simple smile ( ) or closing a conversation with a wave ( ). Still, they fulfill the core phatic function of emojis: shoring up social connections without contributing substantial content.

Adolescents' Emotive Emoji Use is Often Sarcastic

Adolescents also described patterns of emoji use consistent with the emotive functions of emojis, but these patterns were informed by a preference for humor and a complex understanding of emojis' semantic meanings. For example, adolescents said they preferred not to use emojis in text communication (theme 2a) and were particularly averse to using emojis to express negative emotions. This aligns adolescent use with some research on adult use, which has found that adults tend to associate emoji use with expressions of more positive emotions [21] and sometimes see emojis as inappropriate in negative emotional contexts [36]. Though adolescents said they did sometimes use emojis to communicate sincere emotions (theme 3), much of the content from our focus groups suggested that adolescents understand the emotive functions of emojis differently than adults. While adults often use emojis to reduce emotional ambiguity [33] or to complement the emotional valence of text messages [18], adolescents reported that they usually sent emojis as sarcastic or ironic expressions of emotions (theme 2b) and used positive emojis to communicate negative emotions (theme 2c). Adolescents did not describe this sarcastic use of emojis as a source of confusion in communication with peers, however—instead such use was widely understood to communicate emotions in a sardonic or self-effacing tone. Thus, our results suggest adolescents also use emojis to reduce emotional ambiguity, but that they do so with the assumption that others can discern between straightforward and sarcastic expressions.

Adolescents Believe Adults Use Emojis Differently

Our overall interpretation of these findings—that adolescents use emojis for similar reasons as adults but perceive their use as more nuanced—is further supported by the fact that adolescents understood their own emoji use as different from use by adults (theme 4). Importantly, our focus group participants understood adults as favoring literal interpretations of emojis, which they often find risible (theme 2b). Participants said they anticipate these differences when communicating with adults, often suggesting adults would misunderstand the emojis they shared in peer-to-peer communication. Importantly, participants never said they were confused by the emoji choices made by adults. This suggests that adolescents do not understand adult emoji use as its own complex style of communication but instead view it as less developed than their own use.

Thus, adolescents in our focus group believed their use of emojis in their peer-to-peer communication was at least partially inaccessible to their parents and other adults. This might suggest that adolescents use emojis to facilitate one of the core "tasks" of their development: establishing autonomy from parents [37]. Coconstruction theory suggests that adolescents use digital media to achieve developmental tasks [38], and past research suggests that adolescents might gravitate to text messaging and other forms of text-based communication because these platforms help them develop relationships with peers while establishing independence and autonomy [27]. The apparent complexity of adolescent emoji use might serve a similar purpose—just as adolescents have long used slang, fashion, and other cultural codes to distinguish themselves from adults in physical spaces, they may use novel styles of emoji use to assert autonomy within text-based digital spaces.

Limitations

Study findings are limited in 3 important ways. First, while our purposeful sampling procedure recruited a diverse sample of focus group participants, it is likely that important subpopulations were not included in our sample. Thus, it may be that adolescents of certain demographic groups or from certain regions use emojis differently than the participants in this study. Second, the insights of this study depend on self-reflection by focus group participants who may not have been aware of or may have chosen not to disclose certain important aspects of their emoji use. Observation of real-world adolescent emoji use might reveal different patterns. Third, this study's comparisons between adolescent and adult emoji use were seen strictly from the perspective of adolescents because this study did not assess adults' motivations for emoji use. Future work could make stronger comparisons by assessing the emoji use of both adolescents and adults.

Practical Implications

These findings support the assertion [10] that emojis largely fill both emotive and phatic functions in interpersonal communication and suggest that adolescents see their emoji choices as informed by a complex set of shared norms they believe are inaccessible to adults. In addition to their

XSL•FO
RenderX

contribution to the theoretical understanding of adolescent emoji use, these findings also have implications for parents, who increasingly interact with adolescents through text-based platforms [3] and sometimes observe adolescents' text-based communications with peers [6]. These results suggest that parents should exercise caution when interpreting emojis sent by adolescents. While adolescents did say they often adjust their emoji use when communicating with adults, this may not be the case when other adolescents are also involved in the conversation (such as in a group chat), or if adolescents do not have the energy or motivation to adjust their habits for clarity of communication. These results also suggest that adults are likely to misinterpret the emojis used in adolescent-to-adolescent communication, which is an important consideration for parents, school administrators, or practitioners who may find themselves reading such conversations.

Acknowledgments

The authors confirm that generative artificial intelligence tools were not used at any point during the conceptualization or implementation of this study or the writing of this manuscript.

Data Availability

The datasets and generated analyzed during this study are not currently publicly available because they are currently in use in other ongoing projects. Updates and additional information are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

References

1. Anderson M, Jiang J. Teens, social media & technology. Analysis & Policy Observatory. 2018. URL: https://apo.org.au/node/175701 [accessed 2025-04-21]
2. Lenhart A. Teens, social media & technology overview. Pew Research Center. 2015. URL: https://www.pewresearch.org/internet/2015/04/09/teens-social-media-technology-2015/ [accessed 2025-04-21]
3. Jensen M, George MJ, Russell MA, Lippold MA, Odgers CL. Daily parent-adolescent digital exchanges. Res Child Adolesc Psychopathol 2021 Sep;49(9):1125-1138. [doi: 10.1007/s10802-020-00765-x] [Medline: 33821369]
4. Ersahin Z, Hanley T. Using text-based synchronous chat to offer therapeutic support to students: a systematic review of the research literature. Health Educ J 2017 Aug;76(5):531-543. [doi: 10.1177/0017896917704675]
5. Gibson K, Cartwright C. Young people's experiences of mobile phone text counselling: balancing connection and control. Child Youth Serv Rev 2014 Aug;43:96-104. [doi: 10.1016/j.childyouth.2014.05.010]
6. Lenhart A, Ling R, Campbell S, Purcell K. Teens and mobile phones: text messaging explodes as teens embrace it as the centerpiece of their communication strategies with friends. ERIC - Institute of Education Sciences. 2010. URL: https://eric.ed.gov/?id=ED525059 [accessed 2025-04-21]
7. Archer D, Akert RM. Words and everything else: verbal and nonverbal cues in social interpretation. J Pers Soc Psychol 1977;35(6):443-449. [doi: 10.1037//0022-3514.35.6.443]
8. Bai Q, Dan Q, Mu Z, Yang M. A systematic review of emoji: current research and future perspectives. Front Psychol 2019 Sep 15. [doi: 10.3389/fpsyg.2019.02221]
9. Pfeifer VA, Armstrong EL, Lai VT. Do all facial emojis communicate emotion? The impact of facial emojis on perceived sender emotion and text processing. Comput Human Behav 2022 Jan;126:107016. [doi: 10.1016/j.chb.2021.107016]
10. Danesi M. The Semiotics of Emoji: The Rise of Visual Language in the Age of the Internet: Bloomsbury Publishing; 2017. URL: https://www.bloomsbury.com/us/semiotics-of-emoji-9781474282000/ [accessed 2025-04-24]
11. Rodrigues D, Prada M, Gaspar R, Garrido MV, Emoji LDL, Database E. Norms for emoji and emoticons in seven evaluative dimensions. Behav Res Methods 2018 Feb;50(1):392-405. [doi: 10.3758/s13428-017-0878-6]
12. Kaneko D, Toet A, Ushiama S, Brouwer AM, Kallen V, van Erp JBF. EmojiGrid: a 2D pictorial scale for cross-cultural emotion assessment of negatively and positively valenced food. Food Res Int 2019 Jan;115:541-551. [doi: 10.1016/j.foodres.2018.09.049] [Medline: 30599977]
13. Herring SC, Dainas AR. Gender and age influences on interpretation of emoji functions. Trans Soc Comput 2020 Jun 30;3(2):1-26. [doi: 10.1145/3375629]
14. Weiß M, Bille D, Rodrigues J, Hewig J. Age-related differences in emoji evaluation. Exp Aging Res 2020;46(5):416-432. [doi: 10.1080/0361073X.2020.1790087] [Medline: 32662319]
15. Emoji statistics. Emojipedia. URL: https://emojipedia.org/stats [accessed 2025-04-21]
16. Dainas AR, Herring S. C. Interpreting emoji pragmatics. In: Approaches to Internet Pragmatics: Theory and Practice: John Benjamins Publishing Company; 2021:107-144. [doi: 10.1075/pbns.318.04dai]
17. Hu T, Guo H, Sun H, Nguyen TV, Luo J. Spice up your chat: the intentions and sentiment effects of using emojis. ICWSM 2017;11(1):102-111. [doi: 10.1609/icwsm.v11i1.14869]

XSL•FO
RenderX

18. Derks D, Bos AER, von Grumbkow J. Emoticons in computer-mediated communication: social motives and social context. Cyberpsychol Behav 2008 Feb;11(1):99-101. [doi: 10.1089/cpb.2007.9926] [Medline: 18275321]
19. Thompson D, Filik R. Sarcasm in written communication: emoticons are efficient markers of intention. J Comput-Mediat Comm 2016 Mar;21(2):105-120. [doi: 10.1111/jcc4.12156]
20. Boutet I, Goulet-Pelletier JC, Sutera E, Meinhardt-Injac B. Are older adults adapting to new forms of communication? A study on emoji adoption across the adult lifespan. Comput Hum Behav Rep 2024 Mar;13:100379. [doi: 10.1016/j.chbr.2024.100379]
21. Riordan MA. Emojis as tools for emotion work: communicating affect in text messages. J Lang Soc Psychol 2017 Oct;36(5):549-567. [doi: 10.1177/0261927X17704238]
22. Kelly R, Watts L. Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships. Presented at: Experiences of Technology Appropriation: Unanticipated Users, Usage, Circumstances and Design; Sep 20, 2015; Oslo, Norway.
23. Miller H, Kluver D, Thebault-Spieker J, Terveen L, Hecht B. Understanding emoji ambiguity in context: the role of text in emoji-related miscommunication. ICWSM 2017 May 3;11(1):152-161. [doi: 10.1609/icwsm.v11i1.14901]
24. Völkel ST, Buschek D, Pranjic J, Hussmann H. Understanding emoji interpretation through user personality and message context. Presented at: MobileHCI '19; Oct 1-4, 2019; Taipei, Taiwan. [doi: 10.1145/3338286.3340114]
25. An J, Li T, Teng Y, Zhang P. Factors influencing emoji usage in smartphone mediated communications. Presented at: Transforming Digital Worlds; Mar 25-28, 2018; Sheffield, UK. [doi: 10.1007/978-3-319-78105-1_46]
26. Cui J, Colston HL, Jiang G. Is that a genuine smile? Emoji-based sarcasm interpretation across the lifespan. Metaphor Symb 2024 Jul 2;39(3):195-216. [doi: 10.1080/10926488.2024.2314595]
27. Ehrenreich SE, George MJ, Burnell K, Underwood MK. Importance of digital communication in adolescents' development: theoretical and empirical advancements in the last decade. J Res Adolesc 2021 Dec;31(4):928-943. [doi: 10.1111/jora.12643] [Medline: 34820954]
28. Ling R, Bertel T, Sundsøy P. Texting among same-aged individuals: an analysis of traffic data. Presented at: International Communication Association Pre-Conference on Mobile Communication; Jun 21, 2010; Singapore.
29. Dhesi M, Sefi A, Donati M, Hayes J, Cooper M. Helpful and unhelpful elements of synchronous text‐based therapy: a thematic analysis. Couns Psychother Res 2022 Mar;22(1):157-165. [doi: 10.1002/capr.12414]
30. McLaren RM, Pederson JR. Relational communication and understanding in conversations about hurtful events between parents and adolescents. J Commun 2014 Feb;64(1):145-166. [doi: 10.1111/jcom.12072]
31. Sillars A, Smith T, Koerner A. Misattributions contributing to empathic (in)accuracy during parent-adolescent conflict discussions. J Soc Pers Relat 2010 Sep;27(6):727-747. [doi: 10.1177/0265407510373261]
32. Ickes W. Measuring empathic accuracy. In: Interpersonal Sensitivity: Lawrence Erlbaum Associates Publishers; 2001:219-241.
33. Kaye LK, Wall HJ, Malone SA. "Turn that frown upside-down": a contextual account of emoticon usage on different virtual platforms. Comput Human Behav 2016 Jul;60:463-467. [doi: 10.1016/j.chb.2016.02.088]
34. Glaser BG. The constant comparative method of qualitative analysis. Soc Probl 1965 Apr;12(4):436-445. [doi: 10.2307/798843]
35. Glaser BG, Strauss AL, Strutzel E. The discovery of grounded theory; strategies for qualitative research. Nurs Res 1968 Jul;17(4):364. [doi: 10.1097/00006199-196807000-00014]
36. Rodrigues D, Lopes D, Prada M, Thompson D, Garrido MV. A frown emoji can be worth a thousand words: perceptions of emoji use in text messages exchanged between romantic partners. Telemat Inform 2017 Dec;34(8):1532-1543. [doi: 10.1016/j.tele.2017.07.001]
37. Luyckx K, Goossens L, Soenens B, Beyers W. Unpacking commitment and exploration: preliminary validation of an integrative model of late adolescent identity formation. J Adolesc 2006 Jun;29(3):361-378. [doi: 10.1016/j.adolescence.2005.03.008] [Medline: 15878619]
38. Subrahmanyam K, Smahel D, Greenfield P. Connecting developmental constructions to the internet: identity presentation and sexual exploration in online teen chat rooms. Dev Psychol 2006 May;42(3):395-406. [doi: 10.1037/0012-1649.42.3.395] [Medline: 16756432]

Edited by A Mavragani; submitted 17.04.24; peer-reviewed by IA Lane, J Cui; revised version received 05.03.25; accepted 05.03.25; published 28.04.25.

Please cite as:
Minich M, Kerr B, Moreno M
Adolescent Emoji Use in Text-Based Messaging: Focus Group Study
JMIR Form Res 2025;9:e59640
URL: https://formative.jmir.org/2025/1/e59640
doi:10.2196/59640

XSL•FO
RenderX

© Matt Minich, Bradley Kerr, Megan Moreno. Originally published in JMIR Formative Research (https://formative.jmir.org), 28.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impact of the Kidney Score Platform on Communication About and Patients' Engagement With Chronic Kidney Disease Health: Pre–Post Intervention Study

Delphine Tuot[1], MAS, MDCM; Susan Crowley[2,3], MD, MBA; Lois Katz[4], MD; Joseph Leung[4], MD; Delly Alcantara-Cadillo[2,3], MD; Christopher Ruser[2,3], MD; Elizabeth Talbot-Montgomery[5], BS; Joseph Vassalotti[5,6], MD

[1]Division of Nephrology, University of California, San Francisco, Zuckerberg San Francisco General Hospital, 1001 Potrero Ave, Bldg 100, Room 342, San Francisco, CA, United States

[2]Veterans Administration Connecticut Healthcare System, New Haven, CT, United States

[3]Department of Medicine, Yale University, West Haven, CT, United States

[4]Veterans Administration NY Harbor Healthcare System, New York, NY, United States

[5]National Kidney Foundation, New York, NY, United States

[6]Department of Medicine, Icahn School of Medicine, Mt Sinai, New York, NY, United States

Corresponding Author:
Delphine Tuot, MAS, MDCM
Division of Nephrology, University of California, San Francisco, Zuckerberg San Francisco General Hospital, 1001 Potrero Ave, Bldg 100, Room 342, San Francisco, CA, United States

Abstract

Background: Chronic kidney disease (CKD) affects 14% of the US adult population, yet patient knowledge about kidney disease and engagement in their kidney health is low despite many CKD education programs, awareness campaigns, and clinical practice guidelines.

Objective: We aimed to examine the impact of the Kidney Score Platform (a patient-facing, risk-based online tool that provides interactive health information tailored to an individual's CKD risk plus an accompanying clinician-facing Clinical Practice Toolkit) on individual engagement with CKD health and CKD communication between clinicians and patients.

Methods: We conducted a pre-post intervention study in which English-speaking veterans at risk for CKD in two primary care settings interacted with the Kidney Score platform's educational modules and their primary care clinicians were encouraged to review the Clinical Practice Toolkit. The impact of the Kidney Score on the Patient Activation Measure (the primary outcome), knowledge about CKD, and communication with their clinician about kidney health was determined with paired *t* tests. Multivariable linear and logistic models were used to determine whether changes in outcomes after versus before intervention were influenced by age, race or ethnicity, sex, and diabetes status, accounting for baseline values.

Results: The study population (n=76) had a mean (SD) age of 64.4 (8.2) years, 88% (67/76) was male, and 30.3% (23/76) self-identified as African-American. Approximately 93% (71/76) had hypertension, 36% (27/76) had diabetes, and 9.2% (7/76) had CKD according to the laboratory criteria but without an *ICD-10* (International Classification of Diseases, 10th Edition) diagnosis. Patient interaction with the Kidney Score did not change the mean Patient Activation Measure (preintervention: 40.7%, postintervention: 40.2%, *P*=.23) but increased the mean CKD knowledge score (preintervention: 40.0%, postintervention 51.1%, *P*<.01), and changed the percentage of veterans who discussed CKD with their clinician (preintervention: 12.3%, postintervention: 31.5%, *P*<.01). Changes did not differ by age, sex, race, or diabetes status. Results were limited by the small sample size due to low recruitment and minimal clinician engagement with the Clinical Practice Toolkit during the COVID-19 pandemic.

Conclusions: One-time web-based tailored education for patients can increase CKD knowledge and encourage conversations about kidney health. Increasing patient activation for CKD management may require multilevel, longitudinal interventions that facilitate ongoing conversations about kidney health between patients and clinician teams.

(JMIR Form Res 2025;9:e56855) doi:10.2196/56855

KEYWORDS

chronic kidney disease; CKD; CKD communication; CKD knowledge; patient activation; kidney; kidney score platform; kidney health; United States; US; adult; aging; clinical practice; awareness campaign; health information; clinician; primary care; longitudinal intervention; web-based; mobile health; mHealth

XSL•FO
RenderX

Introduction

Chronic kidney disease (CKD) affects 37 million Americans [1] and is associated with high risks of emergency department visits, hospitalizations, cardiovascular events, and early mortality [2-5]. Yet, as many as half of all individuals with laboratory manifestations of kidney disease [6] and those at the highest risk of CKD progression to kidney failure [7] are unaware that they have kidney disease. CKD is usually asymptomatic. Individuals cannot readily know their disease status or risk for disease without risk recognition, testing, detection, and communication by clinicians [8]. A clinician diagnosis of CKD has been associated with increased delivery of evidence-based care, as well as increased patient awareness of their kidney disease [9-11]. However, clinician detection of CKD and communication about kidney disease in the United States are suboptimal [12,13].

Interventions to enhance patient CKD knowledge, increase patient engagement in kidney health, and promote effective communication about CKD between patients and clinicians are critical to facilitate the identification of high-risk populations that would benefit the most from aggressive management of CKD and nephrology referral. They are also important for individuals at a lower risk of CKD progression, to ensure accurate drug dosing, avoid nephrotoxic medications, and encourage the optimal management of diabetes and hypertension, the two most common etiologies of kidney failure in the United States [14].

Existing education programs, awareness campaigns, and clinical practice guidelines have minimally improved CKD awareness in the US population [15,16]. One reason may be because interventions have targeted either the patient or the provider, but rarely both at the same time. To bridge this gap, the National Kidney Foundation developed the Kidney Score Platform, which includes a patient-facing, risk-based online tool that provides interactive health information tailored to an individual's CKD risk as well as an accompanying clinician-facing Clinical Practice Toolkit [17]. The patient-facing elements aim to increase individual awareness of CKD and encourage patients with and at risk for CKD to initiate discussions about kidney disease with their clinicians. The Clinical Practice toolkit assists clinicians in discussing CKD with individuals with and at risk for kidney disease. In this manuscript, we describe the impact of the Kidney Score Platform on individual engagement with their CKD health and CKD communication between clinicians and patients.

Methods

Study Design, Population, and Settings

This was a pre-post–intervention study that examined the impact of the Kidney Score on CKD knowledge and participation in CKD self-management. The Kidney Score is a web-based educational platform designed to improve awareness and understanding of kidney disease among individuals at risk and living with CKD. The study took place in the primary care settings of the Veterans Affairs Medical Centers (VA New York Harbor Healthcare System and VA Connecticut Healthcare System) between September 2022 and March 2023. Eligible participants included English-speaking veterans between the ages of 18 and 75 years with diabetes or hypertension defined in the electronic health record (EHR). The exclusion criteria included an estimated glomerular filtration rate (eGFR) <15 ml/min/1.73 m^2 or individuals with end-stage kidney disease or kidney transplant recipients, as the Kidney Score is not geared towards individuals with severe CKD. Veterans enrolled in hospice services, those with vision impairment, and those with severe dementia identified in the EHR or during the consent process were also excluded from the study, as CKD awareness is much less important for this population's overall health.

Ethical Considerations

Eligible veterans with an upcoming primary care visit at one of two participating Veterans Affairs Medical Centers (VA New York Harbor Healthcare System and VA Connecticut Healthcare System) were identified by the study team, leveraging the EHR. These veterans were invited to participate in the study by mail. Individuals who did not opt out were contacted by phone; those that provided telephone assent were mailed/emailed an information sheet describing the program. During a subsequent phone call, verbal informed consent was obtained and participants were asked to complete an online preintervention survey administered by the study team. Study participants were given a unique study ID; all collected data were deidentified. Participants were given a US $30 gift card for participating. The study was approved by the Institutional Review Boards of the VA Connecticut Healthcare System (#02290) and VA NY Harbor Healthcare System (#01705).

Study Processes and Intervention

The Kidney Score is a free, online website that uses a rule engine and risk predictive analytics to provide interactive health information tailored to an individual's CKD risk and health status. In an anonymous fashion, individuals enter risk factor information as well as laboratory findings in the Kidney Score Platform's online interface and receive educational programming tailored to their clinical status and risk for CKD development or progression. Development of the Kidney Score has been previously described in detail and leveraged the Behavioral Change Wheel framework [18]. It was subsequently found to be acceptable to patients with and at risk for kidney disease [17].

After providing informed consent and baseline data, study participants were invited by the study team to engage with the Kidney Score Platform, which included educational modules tailored to their CKD risk as determined by the Kidney Score's rule engine. Veterans were given the option to print out questions to ask their providers about their kidney health and risk for kidney disease. Participants could access the tool in any location of their preference (ie, home, waiting room, or community center). Within a week after completing their subsequent primary care appointment, participants were asked to complete a postintervention closed survey electronically to assess the impact of the Kidney Score on their understanding of CKD and the quality of their conversation with their clinician about kidney health.

XSL•FO
RenderX

The pre- and postintervention surveys consisted of 4 instruments and 1 additional question. The question order was fixed for all the study participants; the surveys spanned 5 screens and did not include any adaptive components. Survey responses were all voluntary; there was no "completeness check" and participants were allowed to "go back" to change survey responses prior to submitting the final data.

Outcomes

The primary outcome was the change in self-efficacy for CKD management, ascertained by the Patient Activation Measure [19] after the participant engaged with the Kidney Score and completed a visit with their primary care clinician compared to preintervention. Secondary outcomes included self-reported changes in CKD knowledge [20], perceived risk of kidney disease [21], and communication quality with providers [22], all ascertained by validated instruments. The change in the percentage of individuals who discussed kidney disease with their provider was also assessed using the question "At your last visit with your primary care clinician, did you discuss your risk for chronic kidney disease or testing for kidney disease?"

Measurements

Patient demographic data (ie, age, sex, and race or ethnicity) were self-reported by participants. Prevalent comorbid conditions (diabetes, hypertension, and CKD) and the baseline eGFR were ascertained from the EHR.

Statistical Analysis

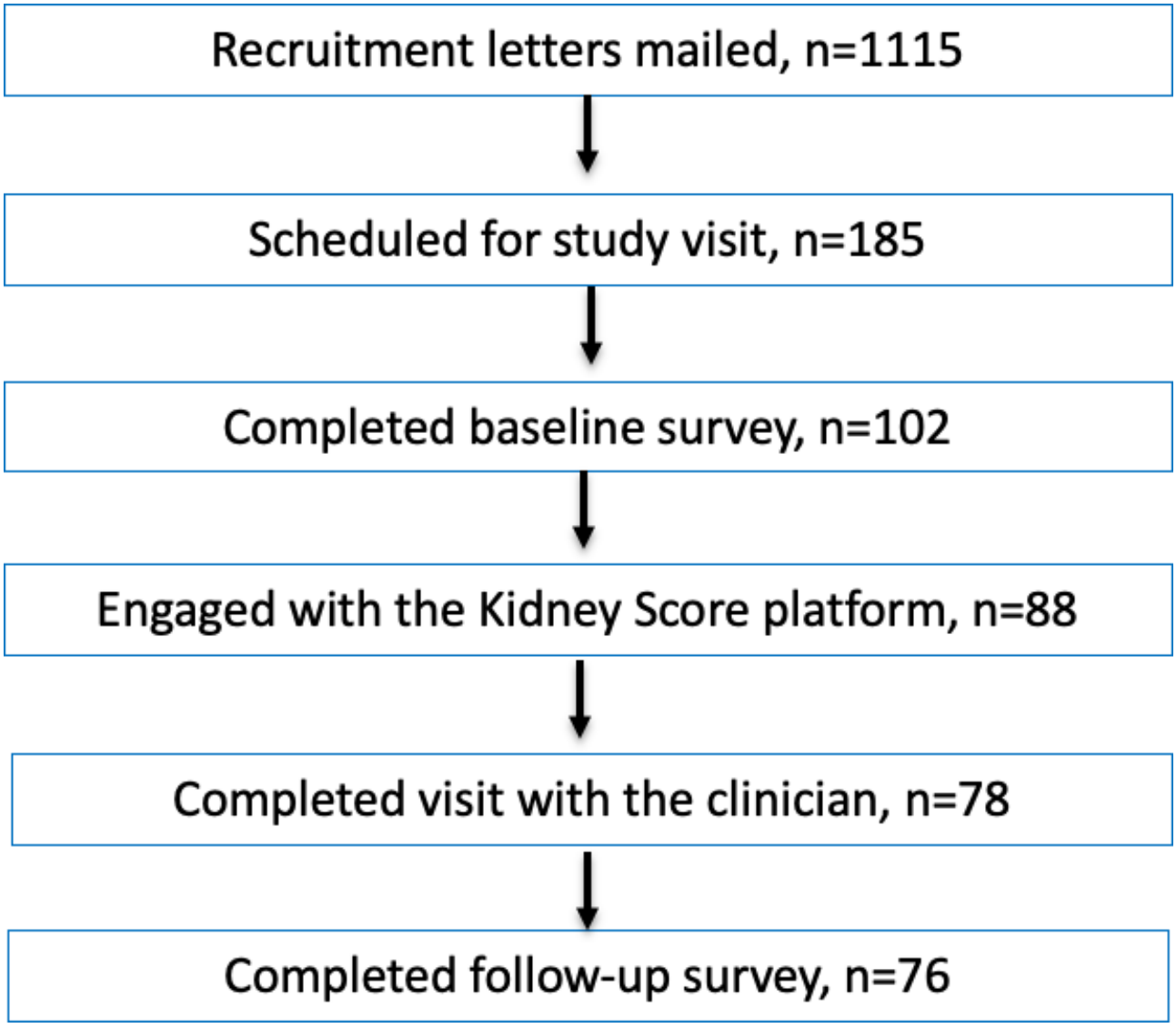
Power calculations performed before study activities suggested that 103 individuals would need to participate in the study to detect a clinically meaningful change in the Patient Activation Measure (the primary outcome). Due to challenges with study operations with the transition towards virtual care delivery during the COVID-19 pandemic, the study was stopped after the full participation of 76 individuals. We examined the characteristics of these study participants using descriptive statistics. Changes in the Patient Activation Measure, communication quality with providers, and kidney knowledge were determined with paired *t* tests. The McNemar test for binary paired data was used to ascertain changes in the perceived risk of kidney disease and the concern for developing kidney disease. Multivariable linear models (for continuous outcomes) and logistic models (for dichotomous outcomes) were used to determine whether changes in outcomes post versus preintervention were influenced by age, race or ethnicity, sex, and diabetes status, after accounting for baseline values. Statistical analyses were performed using Stata/SE, version 14.0 (StataCorp).

Results

Recruitment letters were sent to 1115 eligible veterans, of whom 87 (7.8%) opted out of receiving further communication. The study team was able to contact 206 veterans by phone, of whom 185 agreed to participate (Figure 1). Reasons for declining participation included being uncomfortable with online technology, mistrust of the study team, mistrust in the Veterans Affairs treatment center, not being interested in kidney disease, and not having the time to participate in a study. Approximately 41% (76/185) of the recruited veterans completed all aspects of the study, including completing the previsit survey, reviewing the web-based Kidney Score Platform, attending a scheduled visit with their primary care clinician, and completing the postvisit survey. The study population had a mean (SD) age of 64.4 (8.2) years and 88% (67/76) were male. Nearly one-third (23/76) of the study participants self-identified as African-American and 5% (4/76) self-identified as Native Hawaiian, Pacific Islander, or Asian. Approximately 93% (71/76) had hypertension and 36% (27/76) had diabetes; 9.2% (7/76) of the population had CKD defined by an eGFR 15 - 60 ml/min/1.73 m^2 but none had a diagnosis of CKD in their clinical record (Table 1). Characteristics of the study population were qualitatively similar to those describing the overall eligible primary care population in the participating clinics (Multimedia Appendix 1) .

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram for this implementation study assessing the impact of the kidney score on chronic kidney disease knowledge and participation in self-management.



XSL•FO
RenderX

Table . Characteristics of veterans with and at risk for chronic kidney disease (CKD) who participated in this study that examined the impact of the kidney score on CKD knowledge and participation in CKD self-management.

Characteristics	N=76
Age, years, mean (SD)	64.4 (8.2)
Male sex, n (%)	67 (88.2)
Race, n (%)	
American Indian or Alaskan Native	1 (1.3)
Asian	2 (2.6)
Black	23 (30.3)
Native Hawaiian or Pacific Islander	2 (2.63)
Declined to answer	5 (6.6)
White	43 (56.6)
Ethnicity, n (%)	
Hispanic or Latino	4 (5.3)
Not Hispanic or Latino	71 (93.4)
Declined to answer	1 (1.3)
Hypertension,[a] n (%)	71 (93.4)
Diabetes,[a] n (%)	27 (35.5)
Chronic kidney disease[b]	7 (9.2)

[a]Hypertension and diabetes are defined by the *ICD-10* (International Classification of Diseases, 10th Edition) code.

[b]Chronic kidney disease is defined by an estimated glomerular filtration rate of 15 - 60 ml/min/1.73 m^2.

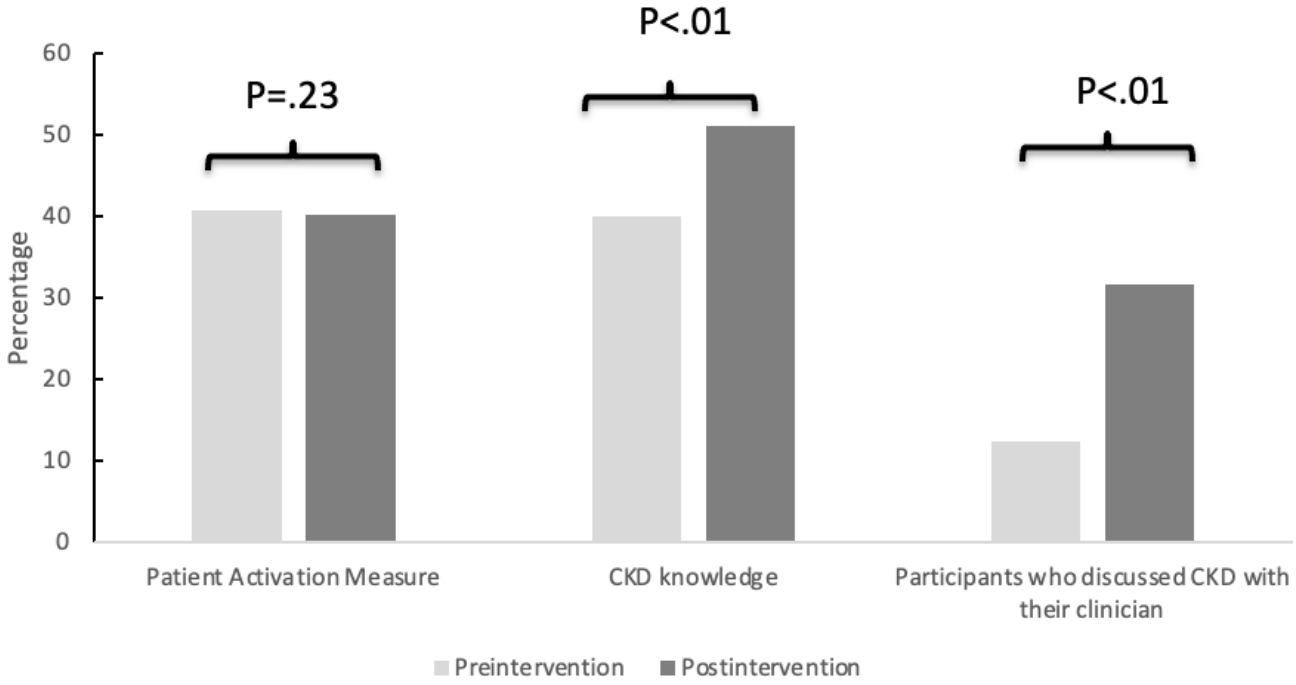
Among veterans with complete data, the mean (SD) Patient Activation Measure score at baseline was 40.7% (5.32%). After interacting with the Kidney Score Platform and attending a visit with their provider, the mean Patient Activation Measure score was 40.2% (5.48%), which was largely unchanged (*P*=.23). Similarly, the self-reported communication quality with providers did not change after the interaction with the Kidney Score Platform, from a mean (SD) baseline score of 2.77 (1.06) out of 3.00 to a mean score of 2.84 (0.94; *P*=.51). On multivariable regression models, age, sex, race or ethnicity, and diabetes status were not associated with changes in the Patient Activation Measure scores nor the communication quality.

The mean (SD) CKD knowledge score increased after the intervention from a baseline of 40% (25) to 51.1% (26.5), representing a statistically significant difference (*P*<.01). Similarly, the percentage of veterans who discussed CKD with their clinician in their prior visit increased from 12.3% at baseline to 31.5% after interacting with the Kidney Score Platform (*P*<.01). In multivariable regression models, change in CKD awareness and communication about CKD did not differ by age, sex, race, or diabetes status (Figure 2).

Consistent with the increased communication about CKD in a low-risk primary care population, the proportion of individuals with substantial concern for developing kidney problems or kidney disease in the next 10 years did not change significantly after participating in the study. At baseline, an estimated 52.1% of individuals reported feeling moderately or very concerned for developing CKD; after the study, this decreased to an estimated 42.3% (*P*=.06). Similarly, the estimated percentage of individuals who reported a moderate or high likelihood of developing kidney disease in the next 10 years did not change substantially before versus after the intervention (29.6% vs 31.0%; *P*=.7). Age, sex, race, and diabetes were not associated with greater concern or perceived risk of developing kidney disease.

XSL•FO
RenderX

Figure 2. Change in individual behaviors potentially associated with self-efficacy for chronic kidney disease (CKD) management after engagement with the kidney scope online platform.



Discussion

Principal Findings and Comparison With Previous Work

The Kidney Score Platform is one of the first interventions that has successfully increased patient knowledge about CKD and communication about kidney health between primary care clinicians and patients. This success can be attributed to its development based on the Behavioral Change Wheel [23], a well-recognized individual behavioral change theory. Theory-informed interventions are more likely to be effective and sustainable within the context of health care delivery than those that are not. The development of the Kidney Score Platform was based on extensive formative work, including qualitative interviews with clinicians and patients [18], as well as proto-typing, beta-testing, and usability testing with various stakeholders, culminating in a refined educational patient experience [17]. This process identified the importance of a tool with which both patients and clinicians could interact and strengthen their communication skills around kidney disease and kidney health.

Despite an increase in CKD conversations and CKD knowledge, the Kidney Score Platform did not increase patient activation for CKD management, which is a hallmark of self-efficacy. The baseline mean (SD) Patient Activation Measure score was 40.1% (5.3%), representing level 1 activation, characterized by individuals who may not believe that the patient's role is important in chronic disease management. This reflects a large opportunity for improvement, rendering this study's null primary outcome results disappointing. We speculate that the lack of change in the Patient Activation Measure score can be attributed in part to the suboptimal sample size, as well as the characteristics of the study population and the intervention itself. Veterans who participated in this study were primarily at risk for CKD (69/76) with only 9.2% (7/76) having laboratory evidence of CKD considering the eGFR. Engaging with the Kidney Score may be more motivating among those with CKD versus those at risk for CKD. Similarly, clinician discussions about CKD and how individuals can participate in self-management are likely more impactful for individuals with laboratory evidence of kidney disease compared to those at risk. The absence of a change in the perceived risk of kidney disease among study participants supports this lack of activation among those at risk for CKD. While we did not have the statistical power to assess changes in patient activation by CKD status, we suspect that a study population enriched with veterans with CKD (as opposed to those at risk for CKD) would have led to a greater change in patient activation.

The nature of the intervention may have also contributed to a lack of change in patient activation. While a handful of patient interactions with the Kidney Score Platform and one patient-initiated discussion with a clinician during a routine primary care visit may be sufficient to increase the broad understanding of kidney disease among individuals with or at risk for CKD, it is likely insufficient to motivate the need for behavior change to optimize kidney health. More than one conversation about kidney disease is likely needed to enhance participation in healthy behaviors and promote self-confidence in other aspects of CKD self-management. With that in mind, it is possible that patient activation was remeasured too early after the time-limited intervention. While speculative, it is plausible that an initial interaction with the Kidney Score Platform may have prompted future discussions about kidney disease and thus future changes in patient activation. Future interventions leveraging the Kidney Score Platform or other patient-facing educational modules will need to include multiple opportunities for discussions about kidney health over time, while also being mindful to not exacerbate potential health anxiety associated with greater knowledge about kidney disease.

XSL•FO
RenderX

This is consistent with Kolb's theory of experiential learning, which posits that most adult learning is gained through experience and engagement over time [24].

The application of Kolb's theory has been demonstrated for the CKD population with moderate to severe CKD depicted by an eGFR 15 - 60 ml/min/1.73 m^2. For example, Medicare's Kidney Disease Education program is designed to be delivered in 6 sessions for those with an eGFR 15 - 30 ml/min/1.73 m^2 and includes modules about knowledge about risk factor control, how the kidneys function, medication management, nutrition, and treatments for kidney failure [25]. Two recent studies demonstrated improvements in access to home dialysis and improved preparation for hemodialysis for the population that received kidney disease education versus eligible controls, although the education delivery was low, among approximately 1% (3469/369,968) of eligible patients [26,27]. A systematic review of educational interventions for patients with CKD with an eGFR range similar to our study participants found that more frequent interventions, (ie, weekly or monthly) were more effective [28]. The elements of the Kidney Score such as the website interaction could easily be scaled for more frequent interventions in future educational investigations.

Additionally, while this study was originally designed as a multilevel intervention directed at patients and clinicians, challenges during the COVID-19 pandemic mitigated the clinician focus of this study. Clinician input refined the Kidney Score Platform, but engagement with the Clinical Practice Toolkit was minimal due to clinician burn out and an emphasis on COVID-19 during the study period. The intervention was ultimately patient-focused with little emphasis on changing clinician behaviors with respect to discussing CKD with individuals with and at risk for kidney disease. Given the strong consensus that multilevel interventions are needed to address health communication [29], future research is needed on what types of tools can be integrated into clinician workflows without exacerbating burn out.

Strengths and Limitations

The strengths of this investigation include theory-informed patient-focused education to improve multiple domains of patient knowledge and engagement assessed using validated instruments. The limitations include low recruitment that may have biased results to the null, necessary study design refinement during the COVID-19 pandemic, absence of a quantitative measure identifying the number of times a participant engaged with the Kidney Score Platform, and limited generalizability to nonveterans considering the administration of the health care delivery systems.

Conclusions

Educating, engaging, and empowering individuals at risk for and living with CKD are critical. The findings of this study suggest that web-based tailored education for patients can increase CKD knowledge and encourage conversations about kidney health. Increasing patient activation and self-efficacy for CKD management will require multilevel, longitudinal interventions that facilitate ongoing conversations about kidney health between patients and clinician teams.

Acknowledgments

Gary Bronz and Anne Cordato provided operational and administrative support. Erin Madden provided primary care clinic demographic information. This study was funded by the Veteran Administration Center for Innovation (VA118-17-R-2144).

Data Availability

Datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of the primary care population potentially eligible for the study, March 1, 2020 to February 28, 2022.
[DOCX File, 17 KB - formative_v9i1e56855_app1.docx]

References

1. Centers for Disease Control and Prevention. Chronic kidney disease in the united states. : US Department of Health and Human Services, Centers for Disease Control and Prevention; 2019 URL: https://www.cdc.gov/kidney-disease/php/data-research/index.html [accessed 2020-02-26]
2. United States Renal Data System. 2018 USRDS annual data report: epidemiology of kidney disease in the united states. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases. 2018 URL: https://www.ajkd.org/article/S0272-6386(19)30009-5/fulltext [accessed 2020-02-26]
3. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu CY. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. N Engl J Med 2004 Sep 23;351(13):1296-1305. [doi: 10.1056/NEJMoa041031] [Medline: 15385656]

XSL•FO
RenderX

4. Sarnak MJ, Greene T, Wang X, et al. The effect of a lower target blood pressure on the progression of kidney disease: long-term follow-up of the modification of diet in renal disease study. Ann Intern Med 2005 Mar 1;142(5):342-351. [doi: 10.7326/0003-4819-142-5-200503010-00009] [Medline: 15738453]
5. Inker LA, Astor BC, Fox CH, et al. KDOQI US commentary on the 2012 KDIGO clinical practice guideline for the evaluation and management of CKD. Am J Kidney Dis 2014 May;63(5):713-735. [doi: 10.1053/j.ajkd.2014.01.416] [Medline: 24647050]
6. Tuot DS, Plantinga LC, Hsu CY, et al. Chronic kidney disease awareness among individuals with clinical markers of kidney dysfunction. Clin J Am Soc Nephrol 2011 Aug;6(8):1838-1844. [doi: 10.2215/CJN.00730111] [Medline: 21784832]
7. Chu CD, McCulloch CE, Banerjee T, et al. CKD Awareness Among US adults by future risk of kidney failure. Am J Kidney Dis 2020 Aug;76(2):174-183. [doi: 10.1053/j.ajkd.2020.01.007] [Medline: 32305206]
8. Vassalotti JA, Centor R, Turner BJ, et al. Practical approach to detection and management of chronic kidney disease for the primary care clinician. Am J Med 2016 Feb;129(2):153-162. [doi: 10.1016/j.amjmed.2015.08.025] [Medline: 26391748]
9. Rothberg MB, Kehoe ED, Courtemanche AL, et al. Recognition and management of chronic kidney disease in an elderly ambulatory population. J Gen Intern Med 2008 Aug;23(8):1125-1130. [doi: 10.1007/s11606-008-0607-z] [Medline: 18443883]
10. Allen AS, Forman JP, Orav EJ, Bates DW, Denker BM, Sequist TD. Primary care management of chronic kidney disease. J Gen Intern Med 2011 Apr;26(4):386-392. [doi: 10.1007/s11606-010-1523-6] [Medline: 20922494]
11. Szczech LA, Stewart RC, Su HL, et al. Primary care detection of chronic kidney disease in adults with type-2 diabetes: the ADD-CKD Study (awareness, detection and drug therapy in type 2 diabetes and chronic kidney disease). PLoS ONE 2014;9(11):e110535. [doi: 10.1371/journal.pone.0110535] [Medline: 25427285]
12. Plantinga LC, Tuot DS, Powe NR. Awareness of chronic kidney disease among patients and providers. Adv Chronic Kidney Dis 2010 May;17(3):225-236. [doi: 10.1053/j.ackd.2010.03.002] [Medline: 20439091]
13. Greer RC, Cooper LA, Crews DC, Powe NR, Boulware LE. Quality of patient-physician discussions about CKD in primary care: a cross-sectional study. Am J Kidney Dis 2011 Apr;57(4):583-591. [doi: 10.1053/j.ajkd.2010.08.027] [Medline: 21131116]
14. United States Renal Data System. 2022 USRDS annual data report: epidemiology of kidney disease in the United States. In: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases 2022.
15. Plantinga LC, Boulware LE, Coresh J, et al. Patient awareness of chronic kidney disease: trends and predictors. Arch Intern Med 2008 Nov 10;168(20):2268-2275. [doi: 10.1001/archinte.168.20.2268] [Medline: 19001205]
16. Management of Chronic Kidney Disease (CKD).: US Department of Veterans Affairs/Department of Defense URL: https://www.healthquality.va.gov/guidelines/CD/ckd [accessed 2020-02-26]
17. Tuot DS, Crowley ST, Katz LA, et al. Usability testing of the kidney score platform to enhance communication about kidney disease in primary care settings: qualitative think-aloud study. JMIR Form Res 2022 Sep 28;6(9):e40001. [doi: 10.2196/40001] [Medline: 36170008]
18. Tuot DS, Crowley ST, Katz LA, et al. The kidney score platform for patient and clinician awareness, communication, and management of kidney disease: protocol for a mixed methods study. JMIR Res Protoc 2020 Oct 19;9(10):e22024. [doi: 10.2196/22024] [Medline: 33074162]
19. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. Health Serv Res 2005 Dec;40(6 Pt 1):1918-1930. [doi: 10.1111/j.1475-6773.2005.00438.x] [Medline: 16336556]
20. Wright JA, Wallston KA, Elasy TA, Ikizler TA, Cavanaugh KL. Development and results of a kidney disease knowledge survey given to patients with CKD. Am J Kidney Dis 2011 Mar;57(3):387-395. [doi: 10.1053/j.ajkd.2010.09.018] [Medline: 21168943]
21. Boulware LE, Carson KA, Troll MU, Powe NR, Cooper LA. Perceived susceptibility to chronic kidney disease among high-risk patients seen in primary care practices. J Gen Intern Med 2009 Oct;24(10):1123-1129. [doi: 10.1007/s11606-009-1086-6] [Medline: 19711135]
22. Lorig K, Stewart A, Ritter P, Gonzalez V, Lauren D, Lynch J. Outcome Measures for Health Education and Other Health Care Interventions: Sage Publications; 1996. [doi: 10.4135/9781452232966]
23. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42. [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
24. Kolb DA. Experiential Learning: Experience as the Source of Learning and Development, 2nd edition: Pearson Education, Inc; 2015.
25. Centers for Medicare & Medicaid Services (CMS) H. Kidney Disease Education. URL: https://www.medicare.gov/coverage/kidney-disease-education [accessed 2024-01-06]
26. Shukla AM, Bozorgmehri S, Ruchi R, et al. Utilization of CMS pre-ESRD Kidney Disease Education services and its associations with the home dialysis therapies. Perit Dial Int 2021 Sep;41(5):453-462. [doi: 10.1177/0896860820975586] [Medline: 33258420]
27. Ruchi R, Bozorgmehri S, Chamarthi G, et al. Provision of kidney disease education service is associated with improved vascular access outcomes among US incident hemodialysis patients. Kidney360 2022 Jan 27;3(1):91-98. [doi: 10.34067/KID.0004502021] [Medline: 35368570]

XSL•FO
RenderX

28. Lopez-Vargas PA, Tong A, Howell M, Craig JC. Educational interventions for patients with ckd: a systematic review. Am J Kidney Dis 2016 Sep;68(3):353-370. [doi: 10.1053/j.ajkd.2016.01.022] [Medline: 27020884]
29. Oh AY, Rising CJ, Gaysynsky A, et al. Advancing multi-level health communication research: A Delphi study on barriers and opportunities. Transl Behav Med 2022 Dec 30;12(12):1133-1145. [doi: 10.1093/tbm/ibac068] [Medline: 36378100]

Abbreviations

CKD: chronic kidney disease
eGFR: estimated glomerular filtration rate
EHR: electronic health record

Edited by A Mavragani; submitted 28.01.24; peer-reviewed by N Alqurini, T Markossian; revised version received 06.03.25; accepted 07.03.25; published 29.04.25.

Please cite as:
Tuot D, Crowley S, Katz L, Leung J, Alcantara-Cadillo D, Ruser C, Talbot-Montgomery E, Vassalotti J
Impact of the Kidney Score Platform on Communication About and Patients' Engagement With Chronic Kidney Disease Health: Pre–Post Intervention Study
JMIR Form Res 2025;9:e56855
URL: https://formative.jmir.org/2025/1/e56855
doi:10.2196/56855

© Delphine Tuot, Susan Crowley, Lois Katz, Joseph Leung, Delly Alcantara-Cadillo, Christopher Ruser, Elizabeth Talbot-Montgomery, Joseph Vassalotti. Originally published in JMIR Formative Research (https://formative.jmir.org), 29.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effect of a Digital Health Exercise Program on the Intention for Spinal Surgery in Adult Spinal Deformity: Exploratory Cross-Sectional Survey

Marsalis Christian Brown[1*], MD; Christopher Quincy Lin[2*], MS; Christopher Jin[3], BS; Matthew Rohde[4*], BS; Brett Rocos[5], MD; Jonathan Belding[1], MD; Barrett I Woods[6*], MD; Stacey J Ackerman[7*], PhD

[1]Department of Orthopaedic Surgery, MetroHealth Medical Center, 2500 MetroHealth Dr, Cleveland, OH, United States

[2]Department of Medicine, The Larner College of Medicine, The University of Vermont, Burlington, VT, United States

[3]Department of Medical, Medical College of Wisconsin, Milwaukee, WI, United States

[4]The Department of Orthopaedics, Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, United States

[5]Duke Spine Center, Duke University Health System, Durham, NC, United States

[6]Department of Spine Surgery, Rothman Orthopaedic Institute, Thomas Jefferson University Rothman Institute, Philadelphia, PA, United States

[7]Department of Biomedical Engineering, Johns Hopkins University, San Diego, CA, United States

[*]these authors contributed equally

Corresponding Author:
Marsalis Christian Brown, MD
Department of Orthopaedic Surgery, MetroHealth Medical Center, 2500 MetroHealth Dr, Cleveland, OH, United States

Abstract

Background: Adult spinal deformity (ASD) is a prevalent condition estimated at 38%. Symptomatic ASD is associated with substantial health care costs. The role of nonoperative interventions in the management of ASD remains elusive. The National Scoliosis Clinic's (NSC) scoliosis realignment therapy (SRT) is a personalized digital health exercise program for the nonoperative management of ASD.

Objective: This exploratory study had two objectives: (1) to evaluate the effect of the SRT program on users' intention of having spinal fusion; and (2) from a US payer perspective, to estimate the annual cost savings per 100,000 beneficiaries by averting spinal surgery.

Methods: Individuals were enrolled in the SRT study from October 1, 2023 to September 1, 2024. Participants completed a web-based, cross-sectional survey about their history of prior scoliosis surgery and intent of having surgery before and after use of SRT (on a 4-point Likert scale, where 1 = "No Intent for Surgery" and 4 = "High Intent for Surgery"). Intent for surgery before and after participation in SRT was compared using a nonparametric Wilcoxon signed-rank test for paired data. Annual cost savings per 100,000 beneficiaries by averting spinal fusions were estimated separately for commercial payers and Medicare using published literature and public data sources. Payer expenditures were inflation-adjusted to 2024 US dollars using the Hospital Services component of the Consumer Price Index.

Results: A total of 62 NSC members (38.8%) responded to the survey and were enrolled in the SRT program for an average (SD) of 17 (12) weeks. The mean (SD) age was 65.3 (13.5) years, and the majority were female (47/48, 98%) and White (45/46, 98%). Among the SRT users who did not have prior scoliosis surgery (n=56), 14% (8/56) reported a decrease in intent for surgery (that is, a lower Likert score) with the use of SRT. The mean (SD) intent for surgery scores before compared to after SRT were 1.29 (0.53) and 1.14 (0.35), respectively (mean difference 0.15 [*P*=.006]). Participants with "No Intent for Surgery" pre- versus postuse of SRT (42/56 versus 48/56, respectively) corresponded to an absolute risk reduction of 11% and a number needed to treat of 9 to avert one spinal fusion. Among the 6 participants who transitioned to "No Intent" for spinal surgery with the use of SRT, 3 were aged <65 years and 3 were ≥65 years of age. The annual cost savings from averted spinal surgeries were estimated at US $415,000 per 100,000 commercially-insured beneficiaries and US $617,000 per 100,000 Medicare beneficiaries.

Conclusions: SRT is a personalized, scoliosis-specific digital health exercise program with the potential for averting 1 spinal surgery for every 9 participants, resulting in a substantial reduction in payer expenditures while improving the quality of care for commercial payers and Medicare beneficiaries.

(JMIR Form Res 2025;9:e66889) doi:10.2196/66889

KEYWORDS

adult spinal deformity; scoliosis; nonoperative management; spinal realignment therapy; digital health; program evaluation

XSL•FO
RenderX

Introduction

Adult spinal deformity (ASD) is a common condition with an estimated prevalence of 38% for primary (de novo) degenerative scoliosis [1,2]. ASD is a 3-dimensional deformity of the spine defined by a major curve magnitude angle of ≥10°, and women are more likely to have the diagnosis [1-4]. Patients with symptomatic ASD report increased back pain as well as lower health-related quality of life as demonstrated by worse SF-36 (36-Item Short Form Health Survey) scores compared to the general population [5,6].

As the US population continues to age, there is growing interest in understanding the most effective means to manage ASD. While nonoperative modalities are typically the first-line treatment for patients with symptomatic ASD, the role of nonoperative management has been questioned, with some studies reporting 2-year costs ranging from US $2041 to US $14,022 without improvements in patient outcomes [7-9]. In contrast, several studies have reported that patients who receive operative interventions for ASD have significantly reduced disability and pain and better improvement in clinical outcomes compared to nonoperative treatment [9-12]. However, spinal fusion is costly (Medicare reimbursement of $60,269 [13]) and carries significant risks; a multicenter database study of nearly 1000 adult surgical patients with ASD with 2-year follow-up reported an overall complication rate of 67.4% [14], and a separate meta-analysis demonstrated complication rates ranging from 17.0% to 71.5% [12]. Considering the substantial economic burden and complication rate associated with operative interventions for ASD—and noting that, other digital care programs in orthopedics (low back pain and chronic knee pain) have demonstrated significant reductions in surgical intent [15-17]—this study sought to further elucidate the role of a digital health program in the nonoperative management of ASD.

The National Scoliosis Clinic's (NSC) Scoliosis Realignment Therapy (SRT) is a remote, exercise-based, ScolioPilates® therapy specifically designed for individuals with scoliosis. ScolioPilates® is a scoliosis-specific exercise program using elongation, corrective breathing, strengthening, and activities of daily living to address pain associated with scoliosis. Additionally, members of SRT are enrolled in a supportive community with virtual group sessions, education, and AI-driven technology to assess spinal curvature with personalized therapy. While Clohisy et al [18] reported that the majority of patients crossover from nonoperative to operative treatment due to perceived worsening of symptoms, Rhode et al [19] recently reported that the use of SRT for 6 weeks resulted in significant improvement in pain as measured using the Scoliosis Research Society Health-Related Quality of Life Questionnaire (SRS-22r).

This exploratory study had two objectives: (1) to evaluate the effect of the SRT program on users' intention of having spinal fusion; and (2) from a US payer perspective, to estimate the annual cost savings per 100,000 beneficiaries by averting spinal surgery.

Methods

Recruitment

Individuals with scoliosis were recruited to the NSC through multiple mechanisms, including social media and the NSC website. NSC members were enrolled nationwide in the SRT study between October 1, 2023 and September 1, 2024. All NSC members using SRT during the aforementioned time period were invited by email to participate in the closed, web-based, cross-sectional survey about surgical intention. A formal sample size calculation was not performed, given that this was an exploratory study. In addition to demographics and clinical characteristics, the survey asked, "Has a surgeon ever told you you need scoliosis surgery, or offered you surgery?", "Have you ever intended to have scoliosis surgery?" and "Since you have started working with the NSC SRT, what is your current intention to have scoliosis surgery?" Responses to the questions about surgical intent were graded on a 4-point Likert scale, where 1=No Intent for Surgery, 2=Low Intent, 3=Moderate Intent, and 4=High Intent for Surgery. Given that this was an exploratory web-based survey, radiographic parameters were not collected.

Statistical Analysis

The primary outcome measure was the intent to pursue surgical correction of scoliosis pain. A nonparametric Wilcoxon signed-rank test for paired data was conducted to evaluate the difference in intention scores before and after SRT. The Wilcoxon test is appropriate for small samples that are not normally distributed. All analyses were performed using R software (version 4.4.1; R Foundation for Statistical Computing).

For potentially averted spinal surgeries, annual cost savings per 100,000 beneficiaries were estimated separately for US commercial payers and Medicare using published literature and multiple public data sources (Table 1). Payer expenditures were inflation-adjusted to US dollars in 2024 by using the Hospital Services component of the Consumer Price Index.

XSL•FO
RenderX

Table . Parameter values, data sources, and estimated cost savings from potentially averted spinal fusions among individuals with adult spinal deformity in the United States (2024, $US) [20].

Parameter	Value		Source
	Commercial	Medicare	
Medical resource use for scoliosis by adults, %	0.502	1.531	[20,21]
Adult scoliosis patients per 100,000 beneficiaries, n	502	1531	Derived
Convert from nonoperative care to spinal fusion (annual), %	21.7	21.7	[22]
Conversions to spinal fusion (annual), n	109	332	Derived
Absolute risk reduction in intent for surgery, %	11	11	SRT survey
SRT[a] participation rate (users), %	28	28	[23]
Spinal fusions averted annually per 100,000 beneficiaries, n	3.4	10.2	Derived
Spinal fusion insurer payment, US $[bc]	123,551[d]	60,269	[13,24,25]
Annual cost savings per 100,000 beneficiaries, $	414,534	616,504	Derived

[a]SRT: scoliosis realignment therapy.

[b]Weighted average of diagnosis-related group (DRG) 456 (spinal fusion with spinal curvature with major complication, 25%), DRG 457 (spinal fusion with spinal curvature with complication, 59%), and DRG 458 (spinal fusion with spinal curvature without complication or major complication, 16%).

[c]Inflated from 2022 to 2024 (half) US $ using the Consumer Price Index for hospital services; multiplier 1.097.

[d]Calculated using a Medicare-to-commercial payment multiplier of 2.05 for hospital inpatient services.

Data Exclusion

Users with prior spinal fusion surgery were excluded from the surgical intent analysis. Participants were given 2 weeks to respond to the survey, with subsequent reminders. Users who failed to respond to the surgical intent questions were excluded.

Ethical Considerations

This web-based survey evaluated a wellness program, not a medical treatment, and study data were deidentified for privacy and confidentiality protection. The datasets generated and analyzed during this study were restricted to the researchers conducting the analysis. NSC members who participated in the web-based survey study were provided US $50 gift cards for the completion of the survey.

Per the Office of Human Research Protections, under the US Department of Health and Human Services, research that (1) involves only survey procedures of adults and (2) is collected in a deidentified fashion is exempt from institutional review board (IRB) review and informed consent can be waived [26]. A retrospective IRB exemption was obtained for this study, confirming that both conditions are met, and therefore, this study was IRB exempt, and informed consent was not indicated.

Results

User Statistics

In total, 160 NSC members were invited to participate in the survey; 62 (38.8%) individuals nationwide voluntarily completed the survey. Six participants were excluded from the intention to pursue surgery analysis: 5 who had prior fusion surgery and 1 who did not respond. Among survey respondents who reported age (N=47), the mean (SD) age was 65.3 (13.5) years (age <65 years, 40%; age ≥65 years, 60%). The mean (SD) SRT duration of use was 118 (86.6 days); the majority were female (47/48, 98%) and White (45/46, 98%; Table 2).

XSL•FO
RenderX

Table . Demographics and clinical characteristics of scoliosis realignment therapy (SRT) cross-sectional survey participants with adult spinal deformity (n=62).

Characteristic	Value
Age in years, mean (SD)	65.3 (13.5)
Age <65 years, n (%)	19 (30.6)
Age ≥65 years, n (%)	28 (45.2)
Did not respond	15 (24.2)
Gender, n (%)	
Female	47 (75.8)
Male	1 (1.6)
Did not respond	14 (22.6)
Race, n (%)	
White	45 (72.6)
American Indian and Caucasian	1 (1.6)
Did not respond	16 (25.8)
SRT Duration of Use in days, mean (SD)	118 (86.6)
Offered scoliosis surgery in the past, n (%)	
Yes	15 (24.2)
No	26 (41.9)
Did not respond	21 (33.9)
Prior fusion surgery for scoliosis, n (%)[a]	
Yes	5 (8.1)
No	56 (90.3)
Did not respond	1 (1.6)

[a]Individuals with either prior fusion surgery or no response excluded from surgical intent analysis.

Evaluation Outcomes

Surgical Intent

Overall, 14% of participants (8/56) reported a decrease in intent for scoliosis surgery (ie, a lower Likert score) with the use of SRT, and no participants reported an increase in surgical intent. The mean (SD) intent for surgery scores before compared to after SRT were 1.29 (0.53) and 1.14 (0.35), respectively (mean difference 0.15 [*P*=.006]; Table 3). Participants with "No Intent" for spinal surgery pre- versus postuse of SRT (42/56 versus 48/56, respectively) corresponded to an absolute risk reduction of 11% and a number needed to treat of 9 to potentially avert 1 spinal fusion (1 divided by 0.11). Among the 6 participants who transitioned to "No Intent" for spinal surgery with the use of SRT, 3 were aged <65 years (ages 30, 62, and 64 years, respectively) and 3 were aged ≥65 years (ages 73, 73, and 82 years, respectively).

XSL•FO
RenderX

Table . Intentions for surgical intervention for adult spinal deformity prior to and after using scoliosis realignment therapy (SRT; cross-sectional survey; n=56).

Statement	Percentage of respondents				Mean (SD)[a]
	No intent	Low	Moderate	High	
Have you ever intended to have scoliosis surgery?	75.0	21.4	3.6	0.0	1.29 (0.530)
Since you have started working with the NSC[b] SRT, what is your current intention to have scoliosis surgery?	85.7	14.3	0.0	0.0	1.14 (0.353)

[a]Four-point Likert scale, where 1=No Intent for Surgery, 2=Low Intent, 3=Moderate Intent, and 4=High Intent. *P*=0.006, Wilcoxon signed-rank test.

[b]NSC: National Scoliosis Clinic.

Estimated Cost Savings by Averting Spinal Fusion

The percentage of adult patients using medical services for scoliosis was 0.502% (age 18 - 64 years) and 1.531% (age ≥65 years) for commercial and Medicare beneficiaries, respectively, or, 502 and 1531 per 100,000 beneficiaries, respectively (Table 1). An estimated 21.7% (41/189) of these patients converted from nonoperative care to spinal fusion annually. Applying the SRT-related 11% absolute risk reduction and an assumed 28% SRT participation rate based on digital health wellness intervention programs without incentives, a total of 3.4 fusions would potentially be averted annually per 100,000 commercially insured beneficiaries and 10.2 fusions potentially avoided annually per 100,000 Medicare beneficiaries. Applying a Medicare payment of US $60,269 for hospital inpatient spinal fusion (weighted average of diagnosis-related groups [DRGs] 456 - 458 in 2024 in US $) and multiplier for Medicare-to-commercial payment of 2.05 (ie, commercial payment of US $123,551), the annual cost savings from averted surgeries were estimated at US $415,000 per 100,000 commercially insured beneficiaries, and US $617,000 per 100,000 Medicare beneficiaries (Table 1).

Discussion

Principal Findings

This web-based, cross-sectional survey study explored whether the use of SRT, a personalized digital therapy program for adults with scoliosis, resulted in a reduction in intent for spinal surgery. After using SRT for an average of 17 weeks, an 11% absolute risk reduction in intent for surgery was observed, which corresponds to a number needed to treat of 9, suggesting that one spine surgery could potentially be averted for every 9 individuals using SRT. Hence, SRT offers a promising nonoperative therapy that may decrease spinal fusions with its associated risks and economic burden.

Comparison With Prior Work

The reduced intent in pursuing surgery with the use of SRT is consistent with other digital care programs in orthopedics, recognizing that scoliosis is a unique condition with its own clinical challenges. For a 12-week digital care program in low back pain, Shebib et al [15] demonstrated a significant reduction in surgical interest (*P*=.01). Similarly, Smittenaar and colleagues [16] reported a decrease in surgery intent at 3 months (*P*<.001) with a 12-week digital care program for chronic knee pain. Mecklenburg et al [17] also found that the self-reported likelihood of having surgery decreased over 1 year with the use of a digital care program for chronic knee pain (*P*=.01).

Investigators have reported conversion rates from nonoperative to operative treatment ranging from 12.7% (24/189) within 1 year to 31% (42/135) after 6 months [18,22]. Clohisy et al [18] further noted that the majority (90%) of patients crossed over from nonoperative to operative treatment due to perceived worsening of symptoms, and the remainder because the patients believed that the nonsurgical therapies were ineffective. Rhode et al [19] recently reported outcomes following 6 weeks of SRT using SRS-22r. Specifically, a significant improvement was observed in the pain (*P*<.001), self-image (*P*=.05), and mental health (*P*<.001) subdomains of the SRS-22r, with the improvement in the pain subdomain exceeding the minimal clinically important difference threshold. In addition, these investigators reported high satisfaction with the SRT program (9.5 out of 10, where 1=Extremely Dissatisfied and 10=Extremely Satisfied) [19]. Given that the SRT cohort reported in the study by Rhode et al (n=23) [19] is a subset of those reported herein (n=62), the statistically significant decrease in intent for surgery is consistent with the previously reported perceivable improvement in pain with the use of SRT. These consistent findings across SRT studies support the validity of using the question on surgical intent to extrapolate to the estimated cost savings.

In light of the aging population among other factors, Wadhwa and colleagues [27] reported that, among both commercial and Medicare beneficiaries, the rate of fusion for adult spinal deformity surgery doubled from 2007 to 2015. In addition, Passias et al [9] reported that the mean 2-year cost of nonoperative treatment was US $2041, and the mean 2-year cost for operative treatment was US $66,860 based on Medicare reimbursements. As such, the estimated cost savings for averted surgeries reported in the present study are likely underestimated

XSL•FO
RenderX

for multiple reasons, specifically, the percentage of beneficiaries seeking medical care for scoliosis is based on data from 2013 [20], the DRG reimbursement amounts reflect only the spinal surgery hospitalization (including the global period) [13], and incentives could increase the SRT participation rate [23].

The cost analyses are based on the best available published data, recognizing that more recent trends in insurer reimbursement and technological advancements may not be reflected. The cost-savings estimates rely on several assumptions, including a 28% participation rate for intervention-focused digital wellness programs. Mattke et al [23] described the relationship between incentive structure and participation rate, noting that the use of incentives increased wellness program participation. Our analysis assumed no incentives with an intervention-focused participation rate of 28% (versus 30% if any incentives). Others have reported a 32% participation rate for physical well-being programs [28]. Hence, assuming the absence of incentives offers a conservative estimate of the participation rate.

Limitations

There are several limitations of this formative research study. First, participants used SRT for an average of 17 weeks, which begs the question about whether the observed decrease in surgical intent would be sustained. That said, Smittenaar et al [16] found that the reduction in surgery interest continued for 6 months following the initiation of a digital health program for chronic knee pain (P<.001), and Mecklenburg et al [17] reported that the decreased likelihood of knee surgery with the use of a digital care program was sustained over 5 years (P=.002). Second, the self-reported survey data are subjective and participants completed both the pre- and post-SRT questions about surgical intention after initiating SRT, which potentially may be subject to recall bias. Third, the pre-post study design may be subject to confounding because it lacks an independent control or comparison group and is less rigorous than a randomized control trial; notwithstanding, each participant served as their own control. Furthermore, study participants were not blinded in that they were aware of using SRT, which may have introduced the possibility of performance bias. Lastly, although participants represented a nationwide sample of adults with scoliosis, the study included a relatively small cohort that did not permit the exploration of nuanced effects and was not ethnically diverse; therefore, the results may not be generalizable. Future studies will explore the durability of SRT use on surgical intent with a larger, more diverse cohort.

Conclusions

SRT is a personalized, scoliosis-specific digital health exercise program with the potential for averting 1 spinal fusion for every 9 participants, resulting in a substantial reduction in payer expenditures while improving the quality of care for commercial payer and Medicare beneficiaries. The SRT remote, digital care program holds promise as a feasible strategy to transform the nonoperative paradigm in adult spinal deformity. Digital care delivery is becoming more common with currently available health technology, and can increase patient agency and engagement, while improving overall outcomes and decreasing health care spending.

Acknowledgments

Generative AI was not used in any portion of the manuscript writing.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

MCB was involved in the conceptualization, data analysis, interpretation of findings, and manuscript preparation. CQL was involved in the conceptualization, interpretation of findings, and manuscript preparation. CJ was involved in the conceptualization, data acquisition, data analysis, and manuscript preparation. MR was involved in the conceptualization and manuscript preparation. BR, JB, and BIW were involved in the conceptualization, manuscript preparation, and critical manuscript review. SJA was involved in the conceptualization, data analysis, interpretation of findings, manuscript preparation, and critical manuscript review.

Conflicts of Interest

SJA declares to have an advisory contract with NSite Medical.
MCB declares consulting for NSite Medical.
All other authors have no conflicts of interest to disclose.

References

1. Bunnell WP. The natural history of idiopathic scoliosis before skeletal maturity. Spine (Phila Pa 1986) 1986 Oct;11(8):773-776. [doi: 10.1097/00007632-198610000-00003]
2. McAviney J, Roberts C, Sullivan B, Alevras AJ, Graham PL, Brown BT. The prevalence of adult de novo scoliosis: a systematic review and meta-analysis. Eur Spine J 2020 Dec;29(12):2960-2969. [doi: 10.1007/s00586-020-06453-0] [Medline: 32440771]
3. Schwab F, Dubey A, Gamez L, et al. Adult scoliosis: prevalence, SF-36, and nutritional parameters in an elderly volunteer population. Spine (Phila Pa 1986) 2005 May;30(9):1082-1085. [doi: 10.1097/01.brs.0000160842.43482.cd]

XSL•FO
RenderX

4. Francis RS. Scoliosis screening of 3,000 college-aged women. The Utah study--phase 2. Phys Ther 1988 Oct;68(10):1513-1516. [Medline: 3174833]
5. Jackson RP, Simmons EH, Stripinis D. Incidence and severity of back pain in adult idiopathic scoliosis. Spine (Phila Pa 1976) 1983 Oct;8(7):749-756. [doi: 10.1097/00007632-198310000-00011] [Medline: 6229884]
6. Bess S, Line B, Fu KM, et al. The health impact of symptomatic adult spinal deformity: comparison of deformity types to United States population norms and chronic diseases. Spine (Phila Pa 1976) 2016 Feb;41(3):224-233. [doi: 10.1097/BRS.0000000000001202] [Medline: 26571174]
7. Glassman SD, Carreon LY, Shaffrey CI, et al. The costs and benefits of nonoperative management for adult scoliosis. Spine (Phila Pa 1986) 2010 Mar;35(5):578-582. [doi: 10.1097/BRS.0b013e3181b0f2f8]
8. Teles AR, Mattei TA, Righesso O, Falavigna A. Effectiveness of operative and nonoperative care for adult spinal deformity: systematic review of the literature. Global Spine J 2017 Apr;7(2):170-178. [doi: 10.1177/2192568217699182] [Medline: 28507887]
9. Passias PG, Ahmad W, Dave P, et al. Economic burden of nonoperative treatment of adult spinal deformity. J Neurosurg Spine 2023 Dec 1;39(6):751-756. [doi: 10.3171/2023.7.SPINE23195] [Medline: 37728175]
10. Bridwell KH, Glassman S, Horton W, et al. Does treatment (nonoperative and operative) improve the two-year quality of life in patients with adult symptomatic lumbar scoliosis. Spine (Phila Pa 1986) 2009 Sep;34(20):2171-2178. [doi: 10.1097/BRS.0b013e3181a8fdc8]
11. Liu S, Schwab F, Smith JS, et al. Likelihood of reaching minimal clinically important difference in adult spinal deformity: a comparison of operative and nonoperative treatment. Ochsner J 2014;14(1):67-77. [Medline: 24688336]
12. Choi SH, Son SM, Goh TS, Park W, Lee JS. Outcomes of operative and nonoperative treatment in patients with adult spinal deformity with a minimum 2-year follow-up: a meta-analysis. World Neurosurg 2018 Dec;120:e870-e876. [doi: 10.1016/j.wneu.2018.08.179] [Medline: 30189313]
13. Data.CMS.gov centers for medicare and medicaid services. Medicare Inpatient Hospitals - by Geography and Service Data Dictionary | CMS Data. URL: https://data.cms.gov/resources/medicare-inpatient-hospitals-by-geography-and-service-data-dictionary-0 [accessed 2024-09-17]
14. Lafage R, Bass RD, Klineberg E, et al. Complication rates following adult spinal deformity surgery: evaluation of the category of complication and chronology. Spine (Phila Pa 1976) 2024;49(12):829-839. [doi: 10.1097/BRS.0000000000004969]
15. Shebib R, Bailey JF, Smittenaar P, Perez DA, Mecklenburg G, Hunter S. Randomized controlled trial of a 12-week digital care program in improving low back pain. NPJ Digit Med 2019;2(1):1. [doi: 10.1038/s41746-018-0076-7] [Medline: 31304351]
16. Smittenaar P, Erhart-Hledik JC, Kinsella R, Hunter S, Mecklenburg G, Perez D. Translating comprehensive conservative care for chronic knee pain into a digital care pathway: 12-week and 6-month outcomes for the hinge health program. JMIR Rehabil Assist Technol 2017 Apr 5;4(1):e4. [doi: 10.2196/rehab.7258] [Medline: 28582253]
17. Mecklenburg G, Smittenaar P, Erhart-Hledik JC, Perez DA, Hunter S. Effects of a 12-week digital care program for chronic knee pain on pain, mobility, and surgery risk: randomized controlled trial. J Med Internet Res 2018 Apr 25;20(4):e156. [doi: 10.2196/jmir.9667] [Medline: 29695370]
18. Clohisy JCF, Smith JS, Kelly MP, et al. Failure of nonoperative care in adult symptomatic lumbar scoliosis: incidence, timing, and risk factors for conversion from nonoperative to operative treatment. J Neurosurg Spine 2023 Oct 1;39(4):498-508. [doi: 10.3171/2023.5.SPINE2326] [Medline: 37327144]
19. Rhode M, Lin C, Jin C, et al. Does a digital care program for adult scoliosis improve outcomes? A pilot study. J Spine 2024;13(4). [doi: 10.2196/preprints.66889]
20. Daubs MD, Watkins-Castillo SI. Orthopedics research society (ORS), United States bone & joint initiative. Table 2B22: Healthcare visits for conditions contributing to curvature of spine disorders by age, United States. 2013. URL: https://www.bmus-ors.org/docs/bmus_e4_T2B.2.2.pdf [accessed 2024-09-08]
21. Projected population by five-year age group and sex for the united states. : U.S. Census Bureau, Population Division; 2023 Nov URL: https://www.census.gov/data/tables/2023/demo/popproj/2023-summary-tables.html [accessed 2024-09-08]
22. Passias PG, Jalai CM, Line BG, et al. Patient profiling can identify patients with adult spinal deformity (ASD) at risk for conversion from nonoperative to surgical treatment: initial steps to reduce ineffective ASD management. Spine J 2018 Feb;18(2):234-244. [doi: 10.1016/j.spinee.2017.06.044] [Medline: 28688984]
23. Mattke S, Kapinos KA, Caloyeras JP, et al. Workplace Wellness Programs: Services Offered, Participation, and Incentives.: RAND Corporation; 2015. URL: https://www.rand.org/pubs/research_reports/RR724.html [accessed 2024-09-17]
24. U.S. bureau of labor statistics. Consumer Price Index for All Urban Consumers, Hospital Services. URL: https://data.bls.gov/pdq/SurveyOutputServlet [accessed 2024-09-08]
25. Commercial reimbursement benchmarking: Commercial payment rates for medical services as percentage of Medicare fee-for-service rates. URL: https://www.milliman.com/en/insight/commercial-reimbursement-benchmarking-payment-rates-medicare-fee-for-service [accessed 2024-09-17]
26. U.S. department of health and human services. Human Subject Regulations Decision Charts: 2018 Requirements. URL: https://www.hhs.gov/ohrp/regulations-and-policy/decision-charts-2018/index.html#c1 [accessed 2025-04-22]

XSL•FO
RenderX

27. Wadhwa H, Leung C, Sklar M, et al. Utilization trends, cost, and payments for adult spinal deformity surgery in commercial and medicare-insured populations. Neurosurgery 2022 Dec 1;91(6):961-968. [doi: 10.1227/neu.0000000000002140] [Medline: 36136402]
28. Valencia C. How to get employees to (actually) participate in well-being programs. Harvard Business Review Published online October. 2021. URL: https://hbr.org/2021/10/how-to-get-employees-to-actually-participate-in-well-being-programs [accessed 2024-11-05]

Abbreviations

ASD: adult spinal deformity
DRG: diagnosis-related group
IRB: institutional review board
NSC: National Scoliosis Clinic
SF-36: 36-Item Short Form Health Survey
SRT: scoliosis realignment therapy

Edited by A Mavragani; submitted 25.09.24; peer-reviewed by RS Punekar; revised version received 08.04.25; accepted 10.04.25; published 29.04.25.

Please cite as:
Brown MC, Lin CQ, Jin C, Rohde M, Rocos B, Belding J, Woods BI, Ackerman SJ
Effect of a Digital Health Exercise Program on the Intention for Spinal Surgery in Adult Spinal Deformity: Exploratory Cross-Sectional Survey
JMIR Form Res 2025;9:e66889
URL: https://formative.jmir.org/2025/1/e66889
doi:10.2196/66889

© Marsalis Christian Brown, Christopher Quincy Lin, Christopher Jin, Matthew Rohde, Brett Rocos, Jonathan Belding, Barrett I Woods, Stacey J Ackerman. Originally published in JMIR Formative Research (https://formative.jmir.org), 29.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Assessing the Utilization of Electronic Consultations in Genetics: Seven-Year Retrospective Study

Sawona Biswas[1,2], MS; Joyce So[2], MD, PhD; Robert Wallerstein[2], MD; Ralph Gonzales[3], MD; Delphine Tout[4], MD; Lisa DeAngelis[5], BA; Aleksandar Rajkovic[1,6], MD, PhD

[1]Department of Pathology, University of California San Francisco, 2340 Sutter Street, San Francisco, CA, United States

[2]Department of Pediatrics, University of California, San Francisco, San Francisco, CA, United States

[3]Department of Medicine, University of California, San Francisco, San Francisco, CA, United States

[4]Division of Nephrology, University of California, San Francisco, San Francisco, CA, United States

[5]Ambulatory Patient Care Access, UCSF Health, San Francisco, CA, United States

[6]Department of Obstetrics, Gynecology & Reproductive Sciences, University of California, San Francisco, CA, United States

Corresponding Author:
Sawona Biswas, MS
Department of Pathology, University of California San Francisco, 2340 Sutter Street, San Francisco, CA, United States

Abstract

Background: Patient and health care provider access to genetic subspecialists is challenging owing to limited number of genetics experts across the United States. The University of California San Francisco (UCSF) Genetics electronic consultation (e-Consult) service was implemented along with the usual referral pathway to improve access to timely genetic expertise through robust asynchronous provider-to-provider communication.

Objectives: This study examined the impact of the UCSF Genetics e-Consult service on patient access to genetics expertise.

Methods: A retrospective chart review of 622 e-Consult requests was conducted. Data pertinent to e-Consult completion rates, provider response times, consultation content, and adherence to geneticist recommendations were abstracted.

Results: From October 2016 to March 2024, the UCSF Genetics e-Consult service received a total of 622 consultation orders, with yearly volumes increasing from 34 in 2017 to 144 in 2023. A total of 360/622 (57.8%) consultations were completed, of which 197/360 (54.6%) were resolved without requiring a specialty care visit. Of the 262/622 (42.1%) e-Consult orders declined by the geneticist reviewer, 184/262 (70.2%) were scheduled for a synchronous genetics visit due to case complexity precluding an appropriate e-Consult response and 29.8% (78/262) were recommended to be referred to a different and more appropriate specialty. Geneticists responded to 83.9% (522/622) of e-Consults within 3 days, with most spending between 5 and 20 minutes on their e-Consult response. Nearly half of the genetics e-Consult requests (69/144; 47.9%) came from primary care providers and pediatricians. Among the 144 e-Consult requests in 2023, 50.6% (73/144) were about diagnostic queries, 17% (25/144) were on symptom management, and 11% (16/144) were about test interpretation. Provider adherence to geneticists' recommendations was observed in 84% (116/144) of cases.

Conclusions: The UCSF Genetics e-Consult service has demonstrated efficiency in providing timely genetic consultations, with a high rate of provider adherence to recommendations. These findings support the potential of e-Consult frameworks as a viable strategy for enhancing access to genetic health care services.

(JMIR Form Res 2025;9:e63028) doi:10.2196/63028

KEYWORDS

genomic; e-Consult; genetic; utility; retrospective; assessment; effectiveness; electronic consultation; healthcare providers; genetic experts; university; consultations; e-Consult frameworks; accessibility; genetic testing; patient care

Introduction

The field of genetics and genomics has experienced significant advancements in recent years, leading to increased integration of genetic testing and considerations into various medical disciplines. However, the number of genetic subspecialists remains limited, resulting in challenges for patients and health care providers seeking timely access to genetics expertise [1,2]. The shortage of genetic specialists, coupled with the growing demand for their services, has led to prolonged waiting times for appointments and potential delays in patient care [3].

As genetic testing is increasingly incorporated into the diagnostic and treatment processes across medical specialties, there is a pressing need to improve communication and collaboration between genetic specialists and a wide range of health care providers [4]. Primary care clinicians, specialists in

XSL•FO
RenderX

other disciplines, and allied health professionals often require guidance and support when interpreting genetic test results, understanding their implications, and supporting patient informed decisions regarding their management [5].

To address these challenges, health care systems have explored innovative solutions, such as electronic consultations (e-Consults), to facilitate prompt access to specialist expertise [6]. e-Consults have emerged as a promising approach to bridge the gap between the limited number of genetic specialists and the growing demand for their service. Within genetics and genomics programs, e-Consult initiatives have demonstrated their effectiveness in addressing basic questions without the need for an in-person visit and providing educational opportunities for generalists to enhance their knowledge and skills in genetics care [7-10]. As genetics becomes increasingly integrated into mainstream medical practice, e-Consults serve as a valuable tool to foster collaboration, ensure knowledge sharing, and provide high-quality, genetics-informed care to patients.

The e-Consult service at the University of California San Francisco (UCSF) started in 2012 [11]. Initially, the program was grant funded and then financially supported by the health system. By 2015, the UCSF entered into agreements with three payers to receive reimbursement for specialist e-Consults, and by 2020, the Centers for Medicare and Medicaid Services (CMS) and other payers started providing reimbursement for e-Consults that are completed and not declined. The introduction of Genetics e-Consults occurred in June 2017, positioning the UCSF as an early adopter in the United States. Herein, we describe the establishment of our Genetics e-Consult service for providers within our institution and its utilization.

Methods

Study Structure

This study analyzed e-Consults submitted through the UCSF Health's electronic medical record system. Providers across multiple specialties, including primary care, pediatrics, and oncology, initiated consultations. The analysis included all Genetics e-Consults from October 2016 to March 2024, capturing utilization trends and consultation characteristics.

Process of e-Consult Services

Ordering providers within UCSF Health can initiate an e-Consult request through our hospital's electronic medical record (Epic Hyperspace 2023; developed by Epic Systems Corporation, Verona, WI, USA). Providers can choose from the Adult Genetics, Pediatrics Genetics, or Cancer Risk Genetics service. After selecting the service, the ordering providers will be prompted to choose one of the four diagnosis-specific SmartText templates: (1) Query genetic condition or syndrome, (2) Known genetic condition or syndrome, (3) Family history of genetic condition, or (4) Other. Each template will guide the provider through different open-ended questions that capture the reason for the e-Consult, relevant test results, family history information, and any other pertinent information. After submission, the e-Consult request is routed to the designated pool and the e-Consultant on service reviews the details and responds.

The Genetics e-Consultants are a group of board-certified geneticists practicing in the Pediatric or Adult Genetics service at the UCSF with more than 5 years of experience. The e-Consultant after accepting the referral summarizes or refines the consultative questions, offers an individualized recommendation along with supporting evidence or current best practices, and outlines a contingency plan (including guidelines for when to refer the patient for an in-person or telehealth specialty visit). The standard response time for an e-Consult is set at 3 business days. In cases where the clinical complexity exceeds the scope of an e-Consult, the e-Consultant may opt to recommend a traditional new patient appointment. The e-Consultant will cc: their specialty staff on the e-Consult response who then convert the e-Consult order to a referral order for routing to the appropriate scheduling team (upon submittal, the requester indicates okay to convert to referral or not okay to convert). Alternatively, if the consultation falls outside the consultant's area of expertise and is better suited for another specialty, the e-Consultant will decline the e-Consult with that recommendation. The e-Consultant self-reports the time spent in the electronic medical record before marking the consultation as complete. The median of these recorded times was then calculated. Within our department, a dedicated team of geneticists is responsible for addressing e-Consults. Once the e-Consult is completed or declined, it is returned to the referring provider's Epic In Basket (Epic Systems Corporation) an integrated electronic messaging and task management system within the Epic electronic health record, with either the clinical recommendations or reason for decline. A short 8-question provider survey was conducted to assess the utilization, satisfaction, and impact of the e-Consult service.

Retrospective Chart Review

A retrospective chart review was conducted to explore the details of e-Consults ordered from October 2016 to March 2024. The review did not require institutional review board approval at the UCSF. Our group characterized the ordering providers, departments, referral groups, descriptions of the consult, primary referral diagnosis, consult status, response times, e-Consult recommendations, number of related referrals to Genetics, and tests performed. Data were collected from EPIC reports and individual chart review, and descriptive statistical analysis was completed using R software (version 4.4.3; R Foundation for Statistical Computing). and Microsoft XML. Statistical reporting included all e-Consult activity originating from an order. Per UCSF Health e-Consult EPIC workflow, all e-Consult orders result in a completed and closed encounter for reporting, whether completed with a clinical recommendation or declined for the reason indicated.

Ethical Considerations

This retrospective quality control and quality improvement study involved a chart review of deidentified data from the electronic medical records at the UCSF. The study was conducted solely for the purpose of evaluating and improving the existing Genetics e-Consult service within the institution. As this project did not involve primary data collection and was

XSL•FO
RenderX

intended for internal quality improvement with no impact on patient care decisions or patient rights, it did not constitute human subjects research under the purview of the Department of Health and Human Services. Accordingly, this study did not require Institutional Review Board review or approval per UCSF's internal policies and federal regulations. All study data were deidentified and securely stored within UCSF's electronic health records system. No financial compensation was provided to study participants, as the research involved retrospective data analysis. No images or materials include identifiable patient information. This not a funded project. All authors are UCSF employees.

Results

From October 2016 to March 2024, the UCSF Genetics e-Consult service received a total of 622 consultation orders, with yearly volumes increasing from 34 in fiscal year (FY) 2017 to 144 in FY 2023 (Figure 1).

Of the 622 Genetics e-Consults, 360/622 (57.8%) were completed, while 262/622 (42.1%) were either declined or deferred to another specialty (Figure 2). Responses were provided on the same day for 263/622 consults (42.3%), within 3 days for 522/622 consults (83.9%), and within 7 days for 562/622 consults (90.4%; Multimedia Appendix 1).

A comprehensive case-level analysis was conducted for the 144 e-Consults ordered during the FY 2023 (June 2022 to July 2023), as this was the first year in which the e-Consult service was fully integrated, with standardized workflows, defined practices, and distinct divisions for Adult, Pediatric, and Cancer Genetics. Prior years lacked the necessary structure and documentation for a comparable analysis. Adult Genetics received the most consults (109/144, 75.7%), followed by Pediatric Genetics (25/144, 17.4%) and Adult Cancer Risk Genetics (10/144, 6.9%). A total of 88/144 (61.1%) of these e-Consults were completed with detailed responses, 46/144 (31.9%) were declined and converted to in-person Genetics appointments, and 10/144 (6.9%) were declined for other reasons such as insufficient information or referral to another specialty (Figure 3).

Nearly half (69/144, 47.9%) of the FY 2023 e-Consults originated from general practices such as primary care, general medicine, and women's health. The median time to answering an e-Consult was 15 (IQR 10 - 25) minutes. e-Consultants spent less than 5 minutes in all the declined cases and an average of 11 - 20 minutes on completed cases and mostly 5 - 10 minutes on cases that were declined and scheduled with Genetics (Figure 3). The most common types of e-Consult requests were queries about genetic diagnoses (73/144, 50.6%), assistance with test interpretation (16/144, 11.1%), and management recommendations for patients with known genetic diagnoses (13/144, 9%) or unknown genetic diagnoses (12/144, 8.3%; Multimedia Appendix 1). e-Consultants provided a range of services, with the most frequent being scheduling patients for outpatient visits (61/144, 42.3%), providing genetic counseling and management recommendations (35/144, 24.3%), making management recommendations alone (27/144, 18.5%), and offering guidance on genetic testing (17/144, 11.8%) and recommended genetics referrals (5/144, 3.4%; Multimedia Appendix 1). A chart review of 134 e-Consults (88 completed and 46 scheduled) revealed that 65/134 patients (48.5%) were referred to and seen by Genetics, 14/134 (12%) were referred with a pending appointment, 5/134 (3.7%) were referred but canceled their appointments, 31/134 (23.1%) had their care informed by the e-Consult recommendations, 5/134 (3.7%) had genetic testing ordered by their referring provider, 12/134 (9%) did not get a Genetics referral, and 2/134 (1.5%) did not have any genetic testing performed (Figure 3).

XSL•FO
RenderX

Figure 1. Total number of e-Consults received by genetics (Adult Genetics, Pediatrics Genetics, or Cancer Risk Genetics) at the University of California, San Francisco from Oct 01, 2016 till 01 March, 2024.

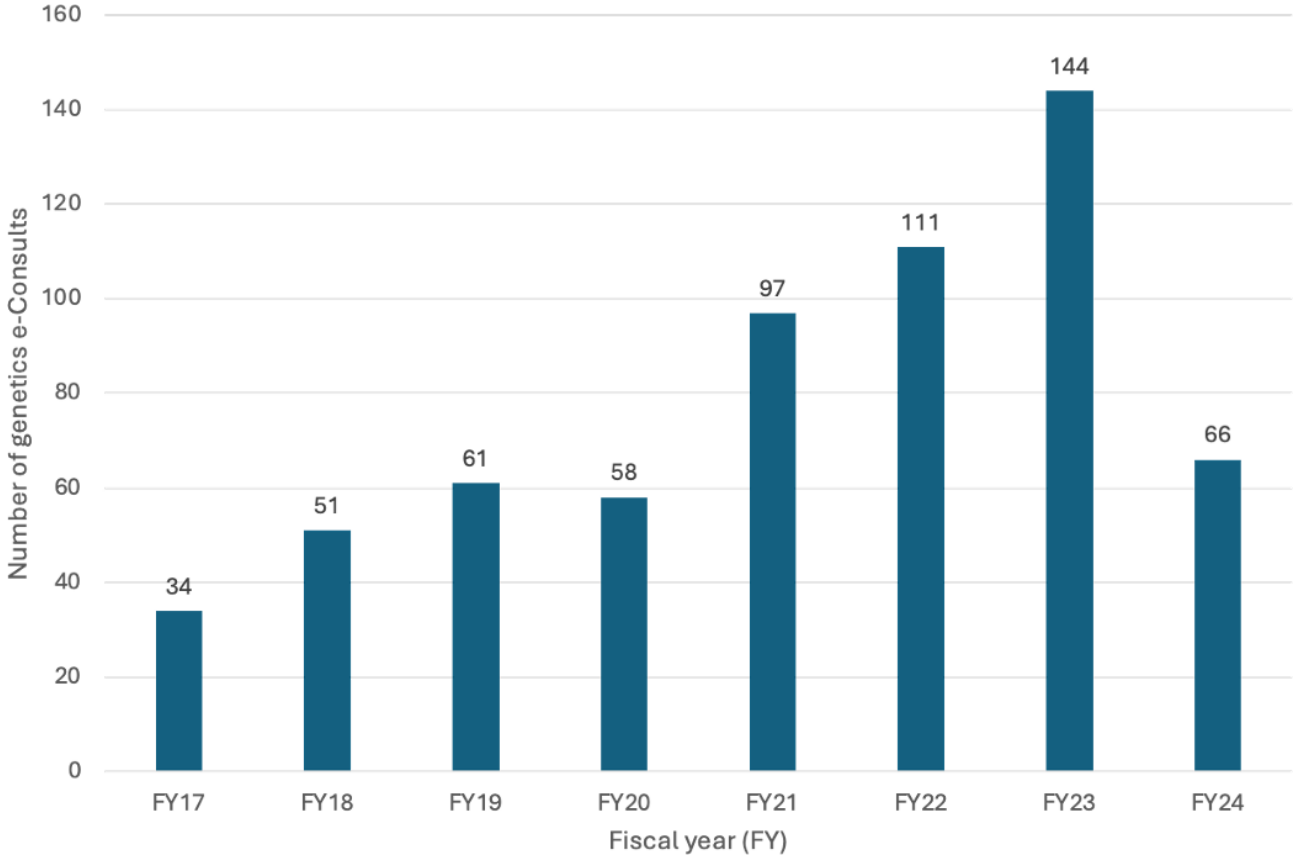
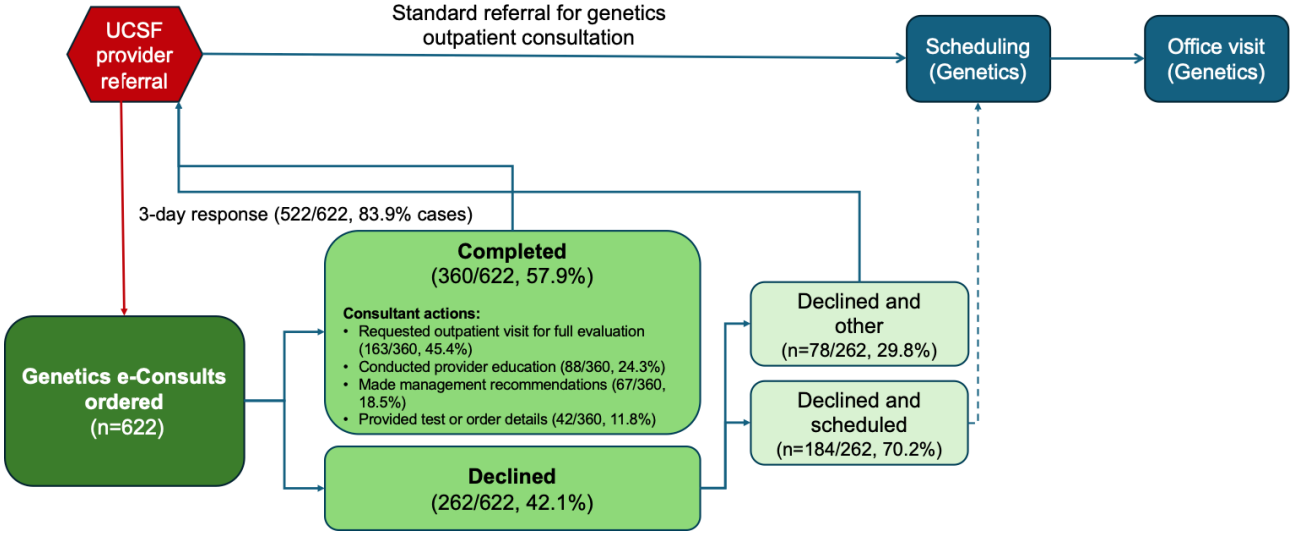
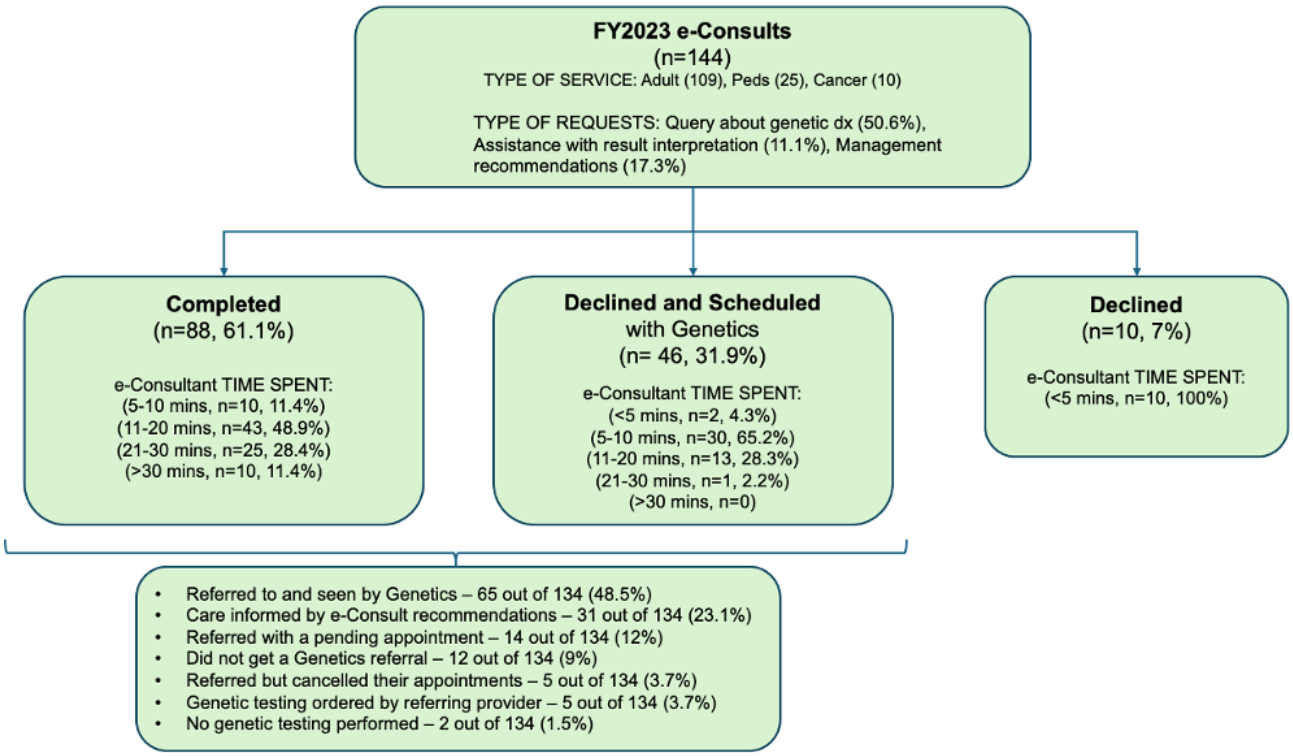


Figure 2. Flowchart and breakdown of e-Consults received and responded 2017 - 2024. UCSF: University of California, San Francisco



XSL•FO
RenderX

Figure 3. Breakdown of e-Consults received in the fiscal year (FY) 2023.



Discussion

Our study demonstrates the successful implementation and growth of an e-Consult service that connects general providers and specialists with genetics experts at a large academic medical center. The UCSF Genetics e-Consult service experienced a steady increase in volume from 34 consults in FY 2017 to 144 in FY 2023 (Figure 1), highlighting the growing demand for accessible genetics expertise.

The e-Consult service provided an efficient means to triage genetics questions, with 57.8% (360/622) of the submitted consults completed and 42.1% (262/622) declined, primarily to schedule in-person evaluations instead (184/262, 70.2% of declined cases). The average turnaround time of 3 days demonstrates the responsiveness of the service in addressing clinical questions.

An analysis of 144 e-Consults from FY 2023 revealed that Adult Genetics handled the majority 109/144 (75.7%), followed by Pediatric Genetics 25/144 (17.4%) and Adult Cancer Risk Genetics 10/144 (6.9%). While 88/144 (61.1%) of these consults were completed electronically, 46/144 (31.9%) were declined and converted to in-person visits (Figure 3), suggesting that e-Consults can help determine when face-to-face evaluations are needed. Nearly half (69/144, 47.9%) of FY 2023 e-Consults came from generalists in primary care, general medicine, and women's health, indicating broad adoption across nongenetics specialties. e-Consultants provided prompt responses, spending 20 minutes or less on 75% of the cases (Figure 3).

The most common reasons for e-Consults were questions about genetic diagnoses (73/144, 50.6%), test interpretations (16/144, 11.1%), and patient management (Multimedia Appendix 1). e-Consultants frequently recommended clinic visits (61/144, 42.3%), provided counseling and management advice (35/144, 24.3%), and offered guidance on genetic testing (17/144, 11.8%;). Referral outcomes data demonstrated that e-Consults resulted in 62% (84/134) of patients being seen by or scheduled with appointments at Genetics, while for 20% of patients (31/134), e-Consult advice influenced their care without necessitating a formal referral to Genetics.

These findings align with and expand upon the limited existing literature on Genetics e-Consult programs. Similar to studies by Bhola et al [9] and Folkerts et al [10], we observed efficient turnaround times, a wide range of clinical topics addressed, and high rates of guideline-concordant care when e-Consults provided actionable recommendations. Our higher e-Consult volume likely reflects the longer duration and wider scope of our program.

The high proportion of e-Consults completed electronically (360/622, 57.9%) highlights the potential for e-Consults to optimize genetics care delivery by addressing straightforward questions, triaging complex cases for in-person evaluation, and guiding appropriate previsit workups. The 84% adherence rate among providers to e-Consult recommendations underscores the educational impact of specialist guidance.

The limitations of our study include the single-center design, lack of long-term outcome data, and absence of referring provider and patient perspectives. Future research should compare e-Consult programs across diverse health care settings; incorporate feedback from key stakeholders; and evaluate the impact on care access, quality, and costs.

In conclusion, the UCSF Genetics e-Consult service demonstrates a scalable and efficient model for expanding access to genetics expertise. By providing timely, individualized guidance to nongenetics providers, e-Consults can help optimize

XSL•FO
RenderX

resource utilization, provider education, and patient care as genetics becomes increasingly integrated into routine clinical practice.

Acknowledgments

No Generative AI was used in any portion of the manuscript.

Data Availability

All data generated or analyzed during this study are included in this published article and in the supplementary information files. Any additional data can be requested from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File, 21 KB - formative_v9i1e63028_app1.docx]

References

1. Kieke MC, Conta JH, Riley JD, Zetzsche LH. The current landscape of genetic test stewardship: a multi-center prospective study. J Genet Couns 2021 Aug;30(4):1203-1210. [doi: 10.1002/jgc4.1403] [Medline: 33728729]
2. Maiese DR, Keehn A, Lyon M, Flannery D, Watson M, Working Groups of the National Coordinating Center for Seven Regional Genetics Service Collaboratives. Current conditions in medical genetics practice. Genet Med 2019 Aug;21(8):1874-1877. [doi: 10.1038/s41436-018-0417-6] [Medline: 30686822]
3. Dragojlovic N, Borle K, Kopac N, et al. The composition and capacity of the clinical genetics workforce in high-income countries: a scoping review. Genet Med 2020 Sep;22(9):1437-1449. [doi: 10.1038/s41436-020-0825-2] [Medline: 32576987]
4. Rubanovich CK, Cheung C, Mandel J, Bloss CS. Physician preparedness for big genomic data: a review of genomic medicine education initiatives in the United States. Hum Mol Genet 2018 Aug 1;27(R2):R250-R258. [doi: 10.1093/hmg/ddy170] [Medline: 29750248]
5. Salari K, Karczewski KJ, Hudgins L, Ormond KE. Evidence that personal genome testing enhances student learning in a course on genomics and personalized medicine. PLoS ONE 2013;8(7):e68853. [doi: 10.1371/journal.pone.0068853] [Medline: 23935898]
6. Carrillo de Albornoz S, Sia KL, Harris A. The effectiveness of teleconsultations in primary care: systematic review. Fam Pract 2022 Jan 19;39(1):168-182. [doi: 10.1093/fampra/cmab077] [Medline: 34278421]
7. Vidgen ME, Fowles LF, Istiko SN, et al. Evaluation of a genetics education program for health interpreters: a pilot study. Front Genet 2021;12:771892. [doi: 10.3389/fgene.2021.771892] [Medline: 35186003]
8. Vimalananda VG, Orlander JD, Afable MK, et al. Electronic consultations (e-consults) and their outcomes: a systematic review. J Am Med Inform Assoc 2020 Mar 1;27(3):471-479. [doi: 10.1093/jamia/ocz185] [Medline: 31621847]
9. Bhola PT, Liddy C, Afkham A, Keely E, Graham GE. A pilot eConsultation service in Eastern Ontario: bridging clinical genetics and primary care. Eur J Hum Genet 2019 Jul;27(7):1026-1032. [doi: 10.1038/s41431-019-0342-4] [Medline: 30778171]
10. Folkerts EK, Pelletier RC, Chung DC, et al. A pooled electronic consultation program to improve access to genetics specialists. Genet Med Open 2023;1(1):100819. [doi: 10.1016/j.gimo.2023.100819] [Medline: 39669243]
11. Gleason N, Prasad PA, Ackerman S, et al. Adoption and impact of an eConsult system in a fee-for-service setting. Healthc (Amst) 2017 Mar;5(1-2):40-45. [doi: 10.1016/j.hjdsi.2016.05.005] [Medline: 27469441]

Abbreviations

CMS: Centers for Medicare and Medicaid Services
e-Consult: electronic consultation
FY: fiscal year
UCSF: University of California San Francisco

XSL•FO
RenderX

Edited by A Mavragani; submitted 07.06.24; peer-reviewed by LE Hull, M Katapodi, M Jani; revised version received 10.03.25; accepted 11.03.25; published 30.04.25.

Please cite as:
Biswas S, So J, Wallerstein R, Gonzales R, Tout D, DeAngelis L, Rajkovic A
Assessing the Utilization of Electronic Consultations in Genetics: Seven-Year Retrospective Study
JMIR Form Res 2025;9:e63028
URL: https://formative.jmir.org/2025/1/e63028
doi:10.2196/63028

© Sawona Biswas, Joyce So, Robert Wallerstein, Ralph Gonzales, Delphine Tout, Lisa DeAngelis, Aleksandar Rajkovic. Originally published in JMIR Formative Research (https://formative.jmir.org), 30.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Predicting Transvaginal Surgical Mesh Exposure Outcomes Using an Integrated Dataset of Blood Cytokine Levels and Medical Record Data: Machine Learning Approach

Mihyun Lim Waugh[1], PhD; Tyler Mills[2], BS; Nicholas Boltin[1], PhD; Lauren Wolf[1], PhD; Patti Parker[3], BSN; Ronnie Horner[4], PhD; Thomas L Wheeler II[5], MD, MSPH; Richard L Goodwin[1], PhD; Melissa A Moss[1], PhD

[1]Department of Biomedical Engineering, University of South Carolina, 301 Main St, Rm 2C02, Columbia, SC, United States

[2]University of South Carolina School of Medicine, Columbia, SC, United States

[3]Prisma Health, Greenville, SC, United States

[4]Department of Health Services Research and Administration, University of Nebraska Medical Center, Omaha, NE, United States

[5]Department of Obstetrics and Gynecology, Spartanburg Regional Healthcare, Spartanburg, SC, United States

Corresponding Author:
Melissa A Moss, PhD
Department of Biomedical Engineering, University of South Carolina, 301 Main St, Rm 2C02, Columbia, SC, United States

Abstract

Background: Transvaginal insertion of polypropylene mesh was extensively used in surgical procedures to treat pelvic organ prolapse (POP) due to its cost-efficiency and durability. However, studies have reported a high rate of complications, including mesh exposure through the vaginal wall. Developing predictive models via supervised machine learning holds promise in identifying risk factors associated with such complications, thereby facilitating better informed surgical decisions. Previous studies have demonstrated the efficacy of anticipating medical outcomes by employing supervised machine learning approaches that integrate patient health care data with laboratory findings. However, such an approach has not been adopted within the realm of POP mesh surgery.

Objective: We examined the efficacy of supervised machine learning to predict mesh exposure following transvaginal POP surgery using 3 different datasets: (1) patient medical record data, (2) biomaterial-induced blood cytokine levels, and (3) the integration of both.

Methods: Blood samples and medical record data were collected from 20 female patients who had prior surgical intervention for POP using transvaginal polypropylene mesh. Of these subjects, 10 had experienced mesh exposure through the vaginal wall following surgery, and 10 had not. Standardized medical record data, including vital signs, previous diagnoses, and social history, were acquired from patient records. In addition, cytokine levels in patient blood samples incubated with sterile polypropylene mesh were measured via multiplex assay. Datasets were created with patient medical record data alone, blood cytokine levels alone, and the integration of both data. The data were split into 70% and 30% for training and testing sets, respectively, for machine learning models that predicted the presence or absence of postsurgical mesh exposure.

Results: Upon training the models with patient medical record data, systolic blood pressure, pulse pressure, and a history of alcohol usage emerged as the most significant factors for predicting mesh exposure. Conversely, when the models were trained solely on blood cytokine levels, interleukin (IL)-1β and IL-12 p40 stood out as the most influential cytokines in predicting mesh exposure. Using the combined dataset, new factors emerged as the primary predictors of mesh exposure: IL-8, tumor necrosis factor-α, and the presence of hemorrhoids. Remarkably, models trained on the integrated dataset demonstrated superior predictive capabilities with a prediction accuracy as high as 94%, surpassing the predictive performance of individual datasets.

Conclusions: Supervised machine learning models demonstrated improved prediction accuracy when trained using a composite dataset that combined patient medical record data and biomaterial-induced blood cytokine levels, surpassing the performance of models trained with either dataset in isolation. This result underscores the advantage of integrating health care data with blood biomarkers, presenting a promising avenue for predicting surgical outcomes in not only POP mesh procedures but also other surgeries involving biomaterials. Such an approach has the potential to enhance informed decision-making for both patients and surgeons, ultimately elevating the standard of patient care.

(JMIR Form Res 2025;9:e59631) doi:10.2196/59631

XSL•FO
RenderX

KEYWORDS

cytokines; inflammatory response; medical record; pelvic organ prolapse; polypropylene mesh; supervised machine learning models; polypropylene; mesh surgery; surgical outcome; cost-efficiency; risk factor; efficacy; health care data; female; informed decision-making; patient care; digital health

Introduction

Patients experiencing difficulty from symptoms of pelvic organ prolapse (POP) are prescribed nonsurgical option as a first-line treatment, including pelvic floor physical therapy to strengthen the muscles supporting the pelvic organs and the use of pessaries that are inserted into the vagina to provide structural support and alleviate symptoms. Lifestyle modifications are also suggested, such as weight management and avoidance of heavy lifting, to reduce symptoms. However, conservative treatments often fail to provide adequate relief, leading many patients to pursue surgical interventions. Notably, patients who undergo surgical treatments are more likely to report complete satisfaction 1 year after the procedure compared to those who opt for nonsurgical options [1]. The choice of surgical approach depends largely on the type of prolapse. Among these, transvaginal mesh surgery was once a widely used treatment when reinforcement of the pelvic floor structures was deemed necessary to treat severe or recurrent prolapse, though its use was later restricted following safety concerns raised by the Food and Drug Administration. This paper focuses on transvaginal mesh surgery to explore strategies for identifying patient populations that are more likely to experience favorable surgical outcomes, with the aim of minimizing the number of patients who face less successful results.

Surgical treatment via transvaginal mesh implantation uses polypropylene mesh due to its low cost, nontoxicity, and mechanical resilience. Its superior tensile strength and elasticity compared to native pelvic tissue enabled it to withstand the pressure changes within the pelvic cavity [2]. Despite these advantageous properties, the usage of polypropylene mesh in transvaginal repair of POP has been prohibited in multiple countries due to postsurgical complications, notably mesh exposure through the vaginal wall [2,3]. In fact, studies have reported mesh exposure rates ranging from 4% to 12.3% in POP patients following surgery [4-6]. Conversely, however, nearly 90% of patients experience successful POP treatment via transvaginal mesh. Consequently, accurately identifying risk factors and developing personalized predictive models can greatly benefit both patients and surgeons in making informed surgical decisions.

Building a machine learning model to predict postsurgical mesh exposure holds considerable promise, as it offers a robust framework that harnesses large datasets, identifies relevant patterns, and provides quantitative predictions that can be used as clinical decision-making tools. In fact, supervised machine learning models have demonstrated efficacy in predicting various surgical outcomes, including patient survival, symptom improvement, and disease progression [7-9]. This efficacy is particularly evident when using datasets that integrate patient health care information and biological assessments. For instance, Jung et al [7] developed an individualized predictive model incorporating patient health care information from medical records as well as pre- and postoperative laboratory measurements (eg, total bilirubin and sodium levels) to anticipate graft failure in pediatric liver transplant surgery. Similarly, Chowdhury et al [8] applied biomarkers, such as mucus cytokines, along with patient demographics and clinical characteristics to machine learning algorithms to predict chronic rhinosinusitis improvement when treated with endoscopic sinus surgery. Kawakita et al [9] developed a model to predict delayed graft function with kidney transplantation surgery by integrating donor and recipient health care characteristics (eg, BMI, race, age, diabetes diagnosis) and clinical laboratory data (eg, blood urea nitrogen levels, serum glutamic pyruvic transaminase). Despite these successful examples of incorporating demographic, clinicopathological, therapeutic, and biomolecular data into personalized clinical decision-making tools aimed at predicting poor clinical outcomes, such methods have yet to gain acceptance in the current clinical practice of mesh surgery.

In this preliminary study, we explored the benefit of integrating distinct datasets when developing supervised machine learning models to predict mesh exposure in patients who have undergone POP repair surgery involving transvaginal insertion of polypropylene mesh. Models were trained with 3 different datasets: patient medical record data (such as demographics, social history, and medical history), biomaterial-induced blood cytokine levels, and the integration of both datasets. Models trained with these 3 different datasets were able to identify pertinent health risk factors as well as important predictors among the cytokines. Moreover, models trained with the integrated dataset relative to individual datasets demonstrated superior performance. These outcomes illustrate the potential of leveraging integrated datasets with machine learning models to accurately predict transvaginal mesh implantation outcomes and identify crucial risk factors associated with such outcomes. Such predictive models offer invaluable guidance to patients and surgeons. This approach holds potential to enhance the decision-making process for both transvaginal and other mesh surgeries, leading to improved patient care and better surgical outcomes.

Methods

Study Population

As described previously [10], recruited patients with surgical intervention for POP using transvaginal polypropylene mesh included both those with mesh exposure, a postsurgical complication where the implanted mesh protrudes through the vaginal wall, and those without. A cohort of 20 healthy, nonpregnant female patients, divided evenly between the 2 categories, was determined necessary for statistical observation of a difference in blood cytokine biomarkers with a 10% chance of error. Briefly, this statistical analysis used a 1-tailed *t* test based on a priori power analysis, a Cohen *d* of 1, and a conservative study power of 0.80.

XSL•FO
RenderX

Participants were selected retrospectively from a list of patients who had undergone surgical intervention for POP using polypropylene mesh at Prisma Health Greenville Memorial Hospital, as detailed in Waugh et al [10]. The patients were randomly identified and recruited via phone call until 10 patients with mesh exposure and 10 without were enrolled for this study. Participants with POP recurrence or those taking medications that could affect inflammatory responses were excluded from the study.

Ethical Considerations

The study protocol was approved by the Institutional Review Board of Prisma Health (number 1984257). Informed consent was waived due to the retrospective nature of the data. All collected data were deidentified and stored in a secure, password-protected folder with restricted access. As this was a retrospective study with no subject participation, no compensation was provided.

Patient Medical Record Data Collection

Medical record data were collected from 20 female patients who had experienced prior surgical intervention for POP using transvaginal polypropylene mesh, selected retrospectively and randomly as described earlier. The data categories include vital signs (systolic blood pressure [BP], diastolic BP, and pulse pressure), previous medical history (hypertension, diabetes, renal disease, hyperlipidemia, hemorrhoids, endometriosis, gastroesophageal reflux, and hysterectomy), social history (alcohol and tobacco usage, sexual activity, marital status, and education level), and other relevant health statistics (age at surgery, BMI, BMI change since surgery, POP stage diagnosis at surgery, and number of births). Mean and SD values were calculated for continuous variables (systolic and diastolic BP, pulse pressure, age at surgery, BMI, and BMI changes) along with the range of the variables (Table 1). Unpaired *t* tests were conducted on each variable to determine significance ($P<.05$).

XSL•FO
RenderX

Table . Demographics of the patients with pelvic organ prolapse (POP) study population (N=20 subjects).

Variables	Mesh exposure (n=10 subjects)	No mesh exposure (n=10 subjects)
Vital signs		
Systolic blood pressure		
Mean (SD)	127.4 (17.9)	123.3 (19.3)
Range	102 - 162	98 - 152
Diastolic blood pressure		
Mean (SD)	77.9 (11.3)	76.9 (10.7)
Range	58 - 98	60 - 92
Pulse pressure		
Mean (SD)	49.5 (21.3)	46.4 (11.9)
Range	24 - 104	
Previous medical history, n (%)		
Hypertension	4 (40)	5 (50)
Diabetes	0 (0)	2 (20)
Renal disease	2 (20)	5 (50)
Hyperlipidemia	3 (30)	3 (30)
Hemorrhoids	3 (30)	2 (20)
Endometriosis	3 (30)	4 (40)
Gastroesophageal reflux	5 (50)	6 (60)
Hysterectomy	8 (80)	8 (80)
Social history, n (%)		
Alcohol	6 (60)	2 (20)
Tobacco	2 (20)	1 (10)
Sexual activity	9 (90)	5 (50)
Marital status		
Single or lives alone	0 (0)	1 (10)
Married	8 (80)	6 (60)
Separated or divorced	1 (10)	1 (10)
Widowed	1 (10)	2 (20)
Highest education level		
Primary	0 (0)	1 (10)
Secondary	2 (20)	2 (20)
College or higher	8 (80)	7 (70)
Other relevant health statistics		
Age at surgery (years)		
Mean (SD)	56.8 (1.2)	59.4 (8.0)
Range	43 - 77	47 - 75
BMI		
Mean (SD)	29.3 (7.0)	28.8 (6.5)
Range	19.3-42.3	21.5 - 44.9
BMI change, mean (SD)	−1.6 (4.5)	−1.1 (3.5)
POP stage preoperation, n (%)		
1	1 (10)	1 (10)

XSL•FO
RenderX

Variables	Mesh exposure (n=10 subjects)	No mesh exposure (n=10 subjects)
2	4 (40)	2 (20)
3	4 (40)	7 (70)
4	1 (10)	0 (0)
Number of births, n (%)		
1 - 2	6 (60)	5 (50)
3+	4 (40)	5 (50)

Patient Blood Cytokine Measurements

Patient blood samples were collected from the same 20 female patients and incubated with sterile polypropylene mesh (2 cm × 2 cm), as detailed in Waugh et al [10]. Plasma layers collected following centrifugation (1500xg, 4 °C) of blood samples were analyzed via multiplex assay to quantify cytokine levels, as in Waugh et al [10]. These cytokines include interleukin (IL)-1α, IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, IL-12 p40, IL-12 p70, IL-17A, interferon-γ, tumor necrosis factor-α (TNF-α), and granulocyte-macrophage colony-stimulating factor. In total, 3 independent measurements were observed to detect the cytokine levels, with each sample evaluated in duplicate.

Data Analysis

Raw datasets, including blood cytokine levels and various medical record data, were analyzed using the statistical programming language R (RStudio, Inc). The imported data contained 60 observations (20 subjects x 3 independent cytokine measurements) and 35 total variable fields (21 medical record variables+13 cytokines+1 target variable). The target variable of exposure referred to the presence or absence of surgical mesh exposure through the vaginal wall following POP surgery. Univariate and multivariate analysis was implemented to explore the dataset, including identifying missing values, analyzing outliers, and standardizing categorical variables.

Supervised machine learning models were created using the caret package (version 6.0 - 90) in the R programming language. The 4 models trained were decision tree (DT), logistic regression (LR), Naïve Bayes (NB), and artificial neural network (ANN). This approach focused on three different datasets: (1) patient medical record data, (2) blood cytokine levels following incubation with surgical mesh, and (3) an integrated dataset of both medical record data and blood cytokine levels. Prior to creating the predictive models, the original data were split using an industry standard of 70% for training and 30% for testing. Each group contained an equal distribution of subjects who did or did not experience postsurgical mesh exposure through the vaginal wall, the prediction target for each model. Each model was trained using the 70% (42/60) subset and a cross-validation training control. A 10-fold cross-validation with 25% (15/60) left out replicated 3 times was used on each model to avoid bias and overfitting. From this, training accuracies are reported. Additional testing was performed for the prediction accuracy of each model using the 30% (18/60) test data. Prediction accuracies are reported along with sensitivity and specificity for the prediction of subjects to experience postsurgical mesh exposure. After computing predictive statistics for each dataset, analysis using the varImp function was conducted to identify which variables contributed most to the predictive outcome within each model. This analysis allowed for visualization of individual variable importance regarding each predictive model.

Results

Descriptive Analysis of Patient Medical Record Data

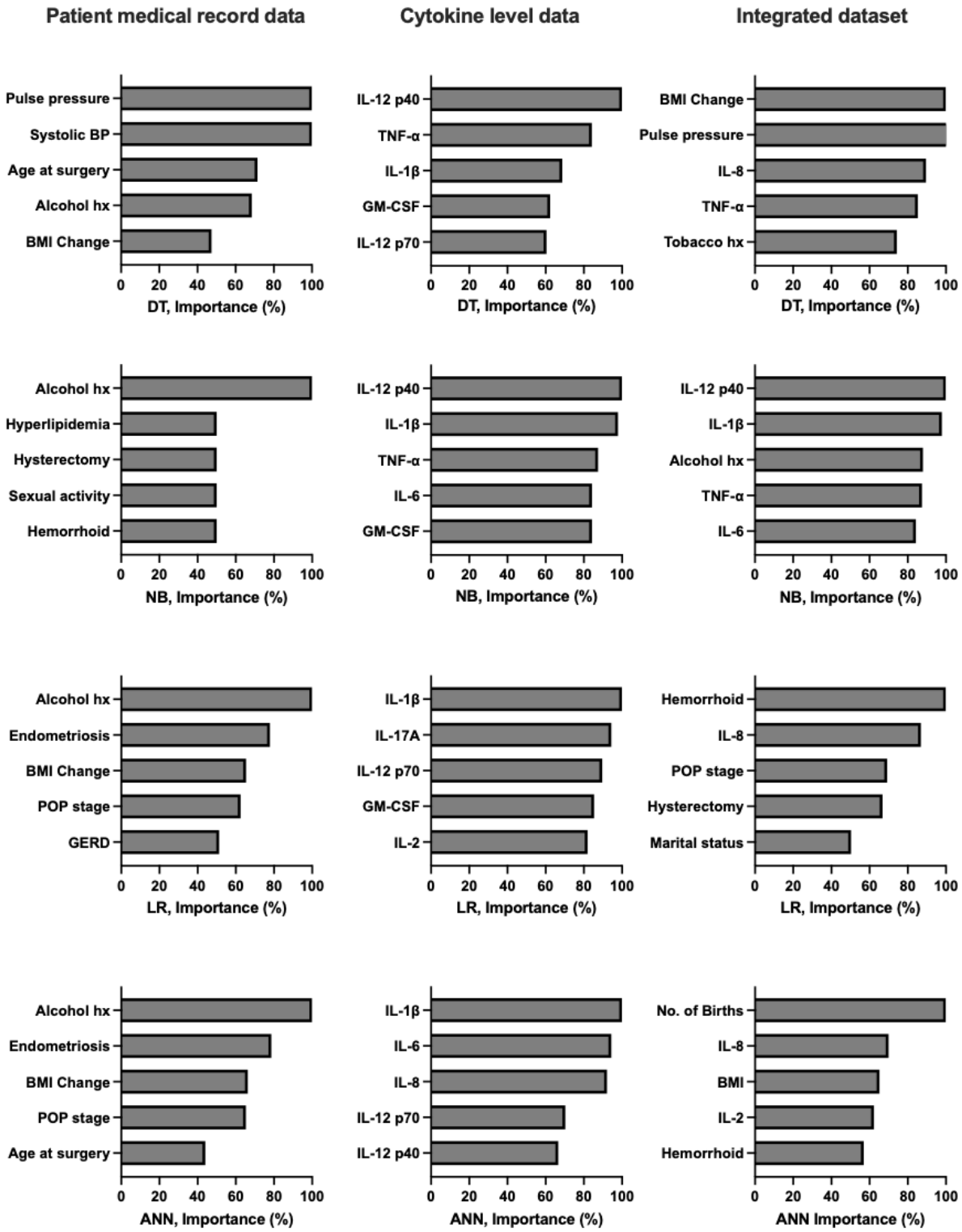
Health care data of patients with POP were collected from electronic medical records and analyzed in Table 1. The cohort included 20 patients with POP, among whom were the 10 that experienced vaginal mesh exposure and 10 that did not. Overall, no significant differences (P<.05) in systolic and diastolic BP, pulse pressure, age at surgery, BMI, and BMI changes were noted between patients with or without mesh exposure. Higher numbers of patients with no mesh exposure had previous medical diagnoses of diabetes and renal diseases. In contrast, higher numbers of patients with mesh exposure reported social histories of alcohol usage as well as sexual activity.

Predictive Analysis

Each of 4 supervised machine learning models, DT, NB, LR, and ANN, was implemented on 3 separate datasets: patient medical record data (21 variables), blood cytokine levels following incubation with surgical mesh (13 variables), and an integrated dataset of both medical data and blood cytokine levels (34 variables). The 5 most important factors that predict mesh exposure in each model are presented in Figure 1.

XSL•FO
RenderX

Figure 1. Four supervised machine learning models were used to identify important factors capable of yielding effective predictive analysis: decision tree (DT), Naïve Bayes (NB), logistic regression (LR), and artificial neural network (ANN). Each model was performed with patient medical record data (left), blood cytokine levels following incubation with surgical mesh (center), or an integrated dataset of both patient medical record data and blood cytokine levels (right). The importance of the 5 most important factors to predict mesh exposure in each model is indicated as a percentage. BP: blood pressure; GERD: gastroesophageal reflux disease; hx: history; IL: interleukin; TNF-α: tumor necrosis factor-α; GM-CSF: granulocyte-macrophage colony-stimulating factor.



When the models were trained with patient medical record data alone, history of alcohol usage was the most important factor for NB, LR, and ANN models, and it also accounted for greater than 60% of the importance in the DT model. BMI change was among the 5 most relevant factors in 3 of the models, DT, LR, and ANN. In total, 3 additional factors were among the 5 most relevant factors in 2 of the 4 models: medical history of endometriosis (LR and ANN, >77% importance), POP stage (LR and ANN, >60% importance), and age at surgery (DT and ANN, >40% importance).

XSL•FO
RenderX

When models were trained with blood cytokine levels alone, IL-1β was the most important cytokine to predict mesh exposure in LR and ANN models and among the 5 most important cytokines in NB and DT models, while IL-12 p40 was the most important cytokine in DT and NB models and among the 5 most important cytokines in the ANN model. In total, 2 cytokines exhibited>60% importance in 3 models: granulocyte-macrophage colony-stimulating factor in DT, NB, and LR models and IL-12 p70 in DT, LR, and ANN models. In total, 2 cytokines exhibited>80% importance in 2 models: TNF-α in DT and NB models and IL-6 in NB and ANN models.

Finally, the predictive models were also performed on an integrated dataset of both medical record data and blood cytokine levels. IL-8 was among the 5 most important factors (>70% importance) in 3 models, DT, LR, and ANN. In total, 2 factors were among the 5 most important in 2 models: TNF-α in DT and NB models and hemorrhoids in LR and ANN models. Factors of highest importance in all 4 models included a combination of health care characteristics and blood cytokine levels. For the NB model, the 5 most important factors in the integrated dataset matched those identified to be of highest importance in the individual datasets. In the DT model, 3 factors identified to be of high importance in individual datasets were also of the highest importance in the integrated dataset, while 2 previously unidentified factors rose to high importance within the integrated dataset. In contrast, ANN and LR models identified four unique factors to be of highest importance within the integrated dataset.

Statistics were generated for the performance of each supervised machine learning model (Table 2). Each model achieved the highest prediction accuracy when using the integrated dataset, ranging from 78% (14/18) to 94% (17/18), compared to either medical record data or blood cytokine levels alone, which scored ranges of 33% (6/18) to 50% (9/18) and 50% (9/18) to 83% (15/18), respectively. Similarly, the highest training accuracy for each model was observed when using the integrated dataset, as opposed to either medical record data or blood cytokine levels alone. When comparing different models, both LR and ANN achieved the highest training and prediction accuracies with the integrated dataset. In fact, when the LR and ANN models were trained with the integrated dataset to predict mesh exposure, a training accuracy of 91% (38/42) and a prediction accuracy of 94% (17/18) were observed in combination with sensitivity and specificity ranging from 89% (16/18) to 100% (18/18).

Table . Summary of supervised machine learning model statistics. Patient medical record data (Med), blood cytokine levels following incubation with surgical mesh (Cyto), and an integrated dataset of both patient medical record data and blood cytokine levels (Int) were used to predict the presence or absence of pelvic organ prolapse (POP) postsurgical mesh exposure; 70% (42/60) of observations were used for training, and 30% (18/60) of observations were used for testing.

Model and dataset		Training accuracy, n (%)	Prediction accuracy, n (%)	95% CI	Sensitivity, n (%)	Specificity, n (%)	Prediction κ
Naïve Bayes							
	Med	16 (37)	9 (50)	0.118-0.882	0 (0)	18 (100)	0
	Cyto	26 (62)	11 (61)	0.357-0.827	6 (33)	16 (89)	0.222
	Int	31 (73)	16 (89)	0.653-0.986	14 (78)	18 (100)	0.778
Decision tree							
	Med	18 (43)	6 (33)	0.0433-0.777	12 (67)	0 (0)	0.333
	Cyto	27 (64)	11 (61)	0.57-0.827	6 (33)	16 (89)	0.222
	Int	28 (67)	14 (78)	0.524-0.936	12 (67)	16 (89)	0.556
Logistic regression							
	Med	11 (27)	6 (33)	0.043-0.777	6 (33)	6 (33)	0.333
	Cyto	31 (73)	9 (50)	0.26-0.740	10 (56)	8 (44)	0.000
	Int	38 (90)	17 (94)	0.727-0.999	18 (100)	16 (89)	0.889
Artificial neural network							
	Med	17 (40)	9 (50)	0.118-0.882	6 (33)	12 (67)	0.000
	Cyto	33 (79)	15 (83)	0.586-0.964	14 (78)	16 (89)	0.667
	Int	38 (90)	17 (94)	0.727-0.999	16 (89)	18 (100)	0.889

Discussion

Summary

Machine learning is gaining acceptance as a tool for predicting medical conditions, including treatment prognosis, disease progression, and surgical outcomes [7,9,11,12]. In particular, literature reports have illustrated the utility of predictive models that incorporate both health care data and biomarkers [13,14]. These successes lay the groundwork for a pioneering method that integrates patient medical records and biomarker measurements for predicting mesh-related postsurgical outcomes. Patients with POP, in particular, could benefit from such a personalized approach. These patients currently make surgical decisions based solely on risk factors derived from

XSL•FO
RenderX

demographic statistics of the general population of surgical patients with POP, while individual medical history and personalized responses to the biomaterial used in the surgery could better inform their decision. Our previous study demonstrated the efficacy of predictive models that use biomaterial-induced blood cytokine-expression profiles to accurately predict mesh exposure following surgical intervention for POP using transvaginal polypropylene mesh [10]. In this exploratory study, we examined the benefit of combining continuous and categorical data from health care records and biomarker measurements to build supervised machine learning models and develop more inclusive and efficient predictive models with a higher rate of prediction accuracy, rendering such models more favorable as a novel, noninvasive, personalized clinical decision tool.

Principal Findings and Comparison to Prior Work

Comparative statistical analysis of medical record data of patients with POP with transvaginal mesh is illustrated in Table 1. Even though this study had a small population of 20 patients, the median outcomes in mesh-exposure and nonexposure groups were comparable to those reported in other studies with a larger study cohort [15,16]. The current dataset revealed that sexually active patients were more prone to experiencing mesh exposure, a trend similarly observed by Kaufman et al [4]. This observation could be attributed to the mechanical challenges posed by sexual activity, potentially leading to friction-induced issues with the mesh implant [17]. In contrast, within the current dataset, variables such as BP, medical diagnoses, education levels, BMI, age at surgery, and parity numbers were not significant risk factors when comparing patients with or without mesh exposure. Similarly, Long et al [18] found no disparity in BMI between the 2 groups. Additionally, Chavez et al [16] noted that risk factors including obesity, diabetes, and smoking status were not associated with vaginal mesh exposure, aligning with our result.

While significant associations were not evident within many individual health care data categories, implementing supervised machine learning models on such data both reinforced groups identified via comparative statistics and revealed new groups associated with patients who experienced mesh exposure. For instance, the history of alcohol usage, indicated as important via comparative statistics, was also considered the most important factor in NB, LR, and ANN models, while BP variables were newly identified as important in the DT model. From a clinical perspective, alcohol use could contribute to malnutrition and inflammation, which may affect the healing process and influence the outcomes of mesh surgery [19]. Similarly, changes in BP and pulse pressure could indicate cardiovascular or systemic issues that may impact surgical planning or postoperative recovery [20]. However, using supervised machine learning models on patient medical record data alone yielded the lowest prediction accuracy rate when compared to models trained with either cytokine levels or an integrated dataset (Table 2). Similarly, low predictive power was reported in Taneja et al [21] when health care data alone was employed in their predictive model [21]. Thus, relying solely on medical record data to construct predictive models does not provide a comprehensive representation. Given an individual's varying inflammatory response [10,22], as well as their race and genetic components [23,24], merely identifying commonalities among medical record data may not accurately represent predictors of the mesh exposure postsurgery.

Previous studies have indicated a relationship between elevated levels of pro-inflammatory cytokines and postsurgical complications [25-27]. When creating predictive models trained with 13 biomaterial-induced cytokines known to influence wound healing, the models predicted with a higher prediction accuracy of 50% (9/18) to 83% (15/18) compared to models trained with patient medical record data (Table 2). These supervised machine learning models demonstrated that IL-1β was one of the most important cytokines to predict mesh exposure across all four models. Elevated levels of IL-1β have been associated with various diseases, including obesity, cardiovascular diseases, cancer, and periodontitis [28-30]. Moreover, IL-1β was increased in a murine model following hernia polypropylene mesh implantation [31], mirroring our findings. Additionally, IL-12 p40 played a pivotal role in DT and NB models. This pro-inflammatory cytokine, which serves as a critical link between the innate and adaptive immune response, is observed in inflammatory conditions, such as pulmonary sarcoidosis and inflammatory bowel diseases [32-34].

All 4 supervised machine learning models achieved their highest prediction accuracies when trained with an integrated dataset comprising both patient medical record data and biomaterial-induced blood cytokine levels. This integrated dataset proved particularly effective when used in LR and ANN models, achieving 94% (17/18) prediction accuracy. The approach also exceeds predictive health care models for other surgical outcomes that used only cytokine-based [35] or electronic health record-based approaches [36]. In parallel, Taneja et al [21] demonstrated that their sepsis predictive models had higher predictive power when combined data were used, rather than either electronic medical record or biomarker data alone. Interestingly, this integrated approach also identifies new factors of high importance. In contrast to the models trained on cytokine expression alone, models trained with integrated data observed that IL-8 was among the important predictors in DT, LR, and ANN models, and TNF-α in DT and NB models. IL-8 and TNF-α are 2 prominent pro-inflammatory cytokines elevated in patients postsurgery, including mesh implantation and pelvic surgery [37,38]. In contrast to the models trained on medical record data alone, hemorrhoids were considered one of the important factors in LR and ANN models trained with the integrated dataset. Models trained with the integrated dataset also outperformed those trained with either dataset alone in terms of training accuracy, sensitivity, and specificity across all four models (Table 2).

Limitations and Future Directions

Despite employing rigorous research methodologies, it is imperative to acknowledge certain limitations inherent in this study. As a pilot study, the population size was limited to 20 patients within a single hospital system. Even though it was on a small scale, this study yielded a prediction accuracy as high as 94% (17/18), surpassing comparable predictive models that ranged from 71% to 86% in prediction accuracy across a

XSL•FO
RenderX

population size of 198-467 [35,39], thus affirming the validity of our approach. Future research will involve a larger participant pool drawn from multiple hospitals to further confirm these results. To mitigate potential confounders, patients with POP recurrence or taking medication that would alter inflammatory response were excluded. Future studies with a larger population will also enable matching participants with respect to these and other potentially confounding variables. In the analysis of the integrated dataset, medical record data were repeated 3 times to match the structure of the cytokine dataset, which contained 3 repeated measurements. This approach was employed to balance the contribution of both data types. This may represent to a potential limitation, although the resulting variable importance graphs for the integrated dataset demonstrated a relatively equal contribution from both the cytokine and medical record data, mitigating concerns that the repetition of the medical record data led to overfitting. Despite these limitations, the results of this pilot study highlight the importance of integrating inflammatory biomarkers with medical record data in the prediction of postsurgical mesh exposure.

Conclusions

Supervised machine learning models demonstrated significantly higher prediction accuracy when trained with an integrated dataset comprising both patient medical record data and biomaterial-induced blood cytokine levels, surpassing the performance of models trained using either dataset alone. These results demonstrate for the first time the advantage of an approach integrating medical record data and blood biomarkers. This innovative approach to predicting surgical outcomes following POP repair via transvaginal insertion of polypropylene mesh holds promise for both patients and surgeons. Additionally, such a method could be applied to the study of other mesh-related surgeries, such as those used to repair POP transabdominally or to treat hernia and stress urinary incontinence. By presenting a more personalized decision-making tool, this approach offers a more accurate depiction of potential outcomes, thereby enhancing informed decision-making and improving patient care.

Acknowledgments

This work was supported by a Transformative Research Grant from Greenville Health System Health Sciences Center to MAM, TLW II, RH, and RLG, by a South Carolina IDeA Networks of Biomedical Research Excellence (INBRE) Student Initiated Research Project (SIRP) to MLW, and by a South Carolina Research Authority Academic Matching Grant to MAM and NB. The authors wish to extend their sincere gratitude to Dr. Matthew Hermes, who contributed to the research insights.

Data Availability

The datasets generated or analyzed during this study are openly available in a public repository [40].

Conflicts of Interest

None declared.

References

1. Hullfish KL, Bovbjerg VE, Gurka MJ, Steers WD. Surgical versus nonsurgical treatment of women with pelvic floor dysfunction: patient centered goals at 1 year. J Urol 2008 Jun;179(6):2280-2285. [doi: 10.1016/j.juro.2008.01.147] [Medline: 18423762]
2. Seifalian A, Basma Z, Digesu A, Khullar V. Polypropylene pelvic mesh: what went wrong and what will be of the future? Biomedicines 2023 Mar 1;11(3):741. [doi: 10.3390/biomedicines11030741] [Medline: 36979721]
3. Syed KK, Consolo MJ, Gousse AE, Repair A. Anterior vaginal wall prolapse repair and the rise and fall of transvaginal mesh. Did we come full circle? A historical perspective. Urology 2021 Apr;150:110-115. [doi: 10.1016/j.urology.2020.08.015] [Medline: 32827535]
4. Kaufman Y, Singh SS, Alturki H, Lam A. Age and sexual activity are risk factors for mesh exposure following transvaginal mesh repair. Int Urogynecol J 2011 Mar;22(3):307-313. [doi: 10.1007/s00192-010-1270-6] [Medline: 20882269]
5. Akyol A, Akca A, Ulker V, et al. Additional surgical risk factors and patient characteristics for mesh erosion after abdominal sacrocolpopexy. J Obstet Gynaecol Res 2014 May;40(5):1368-1374. [doi: 10.1111/jog.12363] [Medline: 24754851]
6. Artsen AM, Liang R, Meyn L, Bradley MS, Moalli PA. Dysregulated wound healing in the pathogenesis of urogynecologic mesh complications. Sci Rep 2023 Dec 5;13(1):21437. [doi: 10.1038/s41598-023-48388-8] [Medline: 38052928]
7. Jung S, Park K, Ihn K, et al. Predicting graft failure in pediatric liver transplantation based on early biomarkers using machine learning models. Sci Rep 2022 Dec 27;12(1):22411. [doi: 10.1038/s41598-022-25900-0] [Medline: 36575218]
8. Chowdhury NI, Li P, Chandra RK, Turner JH. Baseline mucus cytokines predict 22-item sino-nasal outcome test results after endoscopic sinus surgery. Int Forum Allergy Rhinol 2020 Jan;10(1):15-22. [doi: 10.1002/alr.22449] [Medline: 31645085]
9. Kawakita S, Beaumont JL, Jucaud V, Everly MJ. Personalized prediction of delayed graft function for recipients of deceased donor kidney transplants with machine learning. Sci Rep 2020 Oct 27;10(1):18409. [doi: 10.1038/s41598-020-75473-z] [Medline: 33110142]

XSL•FO
RenderX

10. Waugh ML, Boltin N, Wolf L, et al. Prediction of pelvic organ prolapse postsurgical outcome using biomaterial-induced blood cytokine levels: machine learning approach. JMIR Perioper Med 2023 May 31;6:e40402. [doi: 10.2196/40402] [Medline: 37256676]
11. Li B, Shaikh F, Zamzam A, Syed MH, Abdin R, Qadura M. A machine learning algorithm for peripheral artery disease prognosis using biomarker data. iScience 2024 Mar;27(3):109081. [doi: 10.1016/j.isci.2024.109081]
12. Thorwarth RM, Scott DW, Lal D, Marino MJ. Machine learning of biomarkers and clinical observation to predict eosinophilic chronic rhinosinusitis: a pilot study. Int Forum Allergy Rhinol 2021 Jan;11(1):8-15. [doi: 10.1002/alr.22632] [Medline: 32472743]
13. Reddy BK, Delen D, Agrawal RK. Predicting and explaining inflammation in Crohn's disease patients using predictive analytics methods and electronic medical record data. Health Informatics J 2019 Dec;25(4):1201-1218. [doi: 10.1177/1460458217751015] [Medline: 29320910]
14. Ryan CT, Zeng Z, Chatterjee S, et al. Machine learning for dynamic and early prediction of acute kidney injury after cardiac surgery. J Thorac Cardiovasc Surg 2023 Dec;166(6):e551-e564. [doi: 10.1016/j.jtcvs.2022.09.045] [Medline: 36347651]
15. Niu K, Lu YX, Shen WJ, Zhang YH, Wang WY. Risk factors for mesh exposure after transvaginal mesh surgery. Chin Med J (Engl) 2016 Aug 5;129(15):1795-1799. [doi: 10.4103/0366-6999.186631] [Medline: 27453227]
16. El-Khawand D, Wehbe SA, O'Hare PGI, Arunachalam D, Vakili B. Risk factors for vaginal mesh exposure after mesh-augmented anterior repair. Female Pelvic Med Reconstr Surg 2014;20(6):305-309. [doi: 10.1097/SPV.0000000000000095]
17. Kowalik CR, Lakeman MME, de Kraker AT, Roovers JPWR. Effects of mesh-related complications in vaginal surgery on quality of life. Int Urogynecol J 2019 Jul;30(7):1083-1089. [doi: 10.1007/s00192-018-3680-9] [Medline: 29909555]
18. Long CY, Lo TS, Wang CL, Wu CH, Liu CM, Su JH. Risk factors of surgical failure following transvaginal mesh repair for the treatment of pelvic organ prolapse. Eur J Obstet Gynecol Reprod Biol 2012 Apr;161(2):224-227. [doi: 10.1016/j.ejogrb.2012.01.009] [Medline: 22326614]
19. Daher GS, Choi KY, Wells JW, Goyal N. A systematic review of oral nutritional supplement and wound healing. Ann Otol Rhinol Laryngol 2022 Dec;131(12):1358-1368. [doi: 10.1177/00034894211069437] [Medline: 35043693]
20. Mogoanta SS, Paitici S, Mogoanta CA. Postoperative follow-up and recovery after abdominal surgery. In: Zaghal A, Rifai AE, editors. Abdom Surg: IntechOpen; 2021. [doi: 10.5772/intechopen.97739]
21. Taneja I, Reddy B, Damhorst G, et al. Combining biomarkers with EMR data to identify patients in different phases of sepsis. Sci Rep 2017 Sep 7;7(1):10800. [doi: 10.1038/s41598-017-09766-1] [Medline: 28883645]
22. Pos Z, Selleri S, Spivey TL, et al. Genomic scale analysis of racial impact on response to IFN-alpha. Proc Natl Acad Sci U S A 2010 Jan 12;107(2):803-808. [doi: 10.1073/pnas.0913491107] [Medline: 20080756]
23. Hoffmann SC, Stanley EM, Cox ED, et al. Ethnicity greatly influences cytokine gene polymorphism distribution. Am J Transplant 2002 Jul;2(6):560-567. [doi: 10.1034/j.1600-6143.2002.20611.x] [Medline: 12118901]
24. Gökçe İ. Effects of human genetic factors (ethnicity and race) on clinical severity of SARS-CoV-2 (COVID-19). JEB Med Sci 2021;1(3):147-158. [doi: 10.5606/jebms.2020.75631]
25. Mathe Z, Serban RC, Pintilie I, Somkereki C, Hutanu A, Scridon A. Postoperative interleukin-8 levels are related to the duration of coronary artery bypass grafting surgery and predict in-hospital postsurgical complications. Rev Rom Med Lab 2018 Jul 1;26(3):293-303. [doi: 10.2478/rrlm-2018-0021]
26. Paruk F, Chausse JM. Monitoring the post surgery inflammatory host response. J Emerg Crit Care Med 2019;3:47-47. [doi: 10.21037/jeccm.2019.08.06]
27. Tsilimigras DI, Sigala F, Karaolanis G, et al. Cytokines as biomarkers of inflammatory response after open versus endovascular repair of abdominal aortic aneurysms: a systematic review. Acta Pharmacol Sin 2018 Jul;39(7):1164-1175. [doi: 10.1038/aps.2017.212] [Medline: 29770795]
28. Khafagy R, Dash S. Obesity and cardiovascular disease: the emerging role of inflammation. Front Cardiovasc Med 2021;8:768119. [doi: 10.3389/fcvm.2021.768119] [Medline: 34760952]
29. Pani P, Tsilioni I, McGlennen R, et al. IL-1B(3954) polymorphism and red complex bacteria increase IL-1β (GCF) levels in periodontitis. J Periodontal Res 2021 Jun;56(3):501-511. [doi: 10.1111/jre.12850] [Medline: 33638191]
30. Tulotta C, Ottewell P. The role of IL-1B in breast cancer bone metastasis. Endocr Relat Cancer 2018 Jul;25(7):R421-R434. [doi: 10.1530/ERC-17-0309] [Medline: 29760166]
31. Heymann F, von Trotha KT, Preisinger C, et al. Polypropylene mesh implantation for hernia repair causes myeloid cell-driven persistent inflammation. JCI Insight 2019 Jan 24;4(2):e123862. [doi: 10.1172/jci.insight.123862] [Medline: 30674727]
32. Abdi K. IL-12: the role of p40 versus p75. Scand J Immunol 2002 Jul;56(1):1-11. [doi: 10.1046/j.1365-3083.2002.01101.x] [Medline: 12100467]
33. Shigehara K, Shijubo N, Ohmichi M, et al. Increased circulating interleukin-12 (IL-12) p40 in pulmonary sarcoidosis. Clin Exp Immunol 2003 Mar 24;132(1):152-157. [doi: 10.1046/j.1365-2249.2003.02105.x]
34. Verstockt B, Salas A, Sands BE, et al. IL-12 and IL-23 pathway inhibition in inflammatory bowel disease. Nat Rev Gastroenterol Hepatol 2023 Jul;20(7):433-446. [doi: 10.1038/s41575-023-00768-1]

XSL•FO
RenderX

35. Chen Z, Chen L, Yao G, Yang W, Yang K, Xiong C. Novel blood cytokine-based model for predicting severe acute kidney injury and poor outcomes after cardiac surgery. J Am Heart Assoc 2020 Nov 17;9(22):e018004. [doi: 10.1161/JAHA.120.018004] [Medline: 33131359]
36. Weller GB, Lovely J, Larson DW, Earnshaw BA, Huebner M. Leveraging electronic health records for predictive modeling of post-surgical complications. Stat Methods Med Res 2018 Nov;27(11):3271-3285. [doi: 10.1177/0962280217696115] [Medline: 29298612]
37. Parpoudi S, Mantzoros I, Gkiouliava A, et al. Effect of N-acetyl-L-cysteine on inflammation after intraperitoneal mesh placement in a potentially contaminated environment: an experimental study in the rat. Asian J Surg 2022 Nov;45(11):2191-2196. [doi: 10.1016/j.asjsur.2021.11.001] [Medline: 34801356]
38. Shi JY, Bergerat-Thompson AM, Mitchell CM. The impact of the vaginal microbiome on vaginal epithelial wound healing. Research Square. Preprint posted online on Aug 25, 2023. [doi: 10.21203/rs.3.rs-3282980/v1]
39. Chu CS, Lee NP, Adeoye J, Thomson P, Choi SW. Machine learning and treatment outcome prediction for oral cancer. J Oral Pathol Med 2020 Nov;49(10):977-985. [doi: 10.1111/jop.13089] [Medline: 32740951]
40. IntegratedDataset. GitHub. URL: https://github.com/tomhjlim/IntegratedDataset [accessed 2025-03-17]

Abbreviations

ANN: artificial neural network
BP: blood pressure
DT: decision tree
IL: interleukin
LR: logistic regression
NB: Naïve Bayes
POP: pelvic organ prolapse
TNF-α: tumor necrosis factor alpha

Edited by A Mavragani; submitted 12.06.24; peer-reviewed by B Zhao, MF Kinci; revised version received 22.01.25; accepted 23.01.25; published 01.05.25.

Please cite as:
Waugh ML, Mills T, Boltin N, Wolf L, Parker P, Horner R, Wheeler II TL, Goodwin RL, Moss MA
Predicting Transvaginal Surgical Mesh Exposure Outcomes Using an Integrated Dataset of Blood Cytokine Levels and Medical Record Data: Machine Learning Approach
JMIR Form Res 2025;9:e59631
URL: https://formative.jmir.org/2025/1/e59631
doi:10.2196/59631

©Mihyun Lim Waugh, Tyler Mills, Nicholas Boltin, Lauren Wolf, Patti Parker, Ronnie Horner, Thomas L Wheeler II, Richard L Goodwin, Melissa A Moss. Originally published in JMIR Formative Research (https://formative.jmir.org), 1.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Integration of a Mental Health App (e-MICHI) Into a Blended Treatment of Depression in Adolescents: Single-Group, Naturalistic Feasibility Trial

Lena Lincke[1,2*], MSc; Tim Martin-Döring[1,2*], MSc; Andrea Daunke[1,2], Dipl-Psych; Antonia Sadkowiak[1]; Daria Alexandra Nolkemper[3,4], MA; Nina Sproeber-Kolb[5], Prof Dr; Stefanie Bienioschek[3,4,6], Dr Med; Olaf Reis[1,2], PhD, PD; Michael Kölch[1,2], Prof Dr Med

[1]Department of Child and Adolescent Psychiatry, Neurology, Psychosomatics and Psychotherapy, Rostock University Medical Center, Gehlsheimer Str. 20, Rostock, Germany

[2]Partner Site Greifswald/Rostock, German Center for Child and Adolescent Health (DZKJ), Rostock, Germany

[3]Brandenburg Medical School Theodor Fontane (MHB), Neuruppin, Germany

[4]Department of Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics, University Hospital Ruppin-Brandenburg (ukrb), Neuruppin, Germany

[5]School of Social Work, Baden-Wuerttemberg Cooperative State University, Stuttgart, Germany

[6]Faculty of Health Sciences, Joint Faculty of the University of Potsdam, the Brandenburg University of Technology Cottbus-Senftenberg and the Brandenburg Medical School Theodor Fontane, Potsdam, Germany

[*]these authors contributed equally

Corresponding Author:
Tim Martin-Döring, MSc
Department of Child and Adolescent Psychiatry, Neurology, Psychosomatics and Psychotherapy, Rostock University Medical Center, Gehlsheimer Str. 20, Rostock, Germany

Abstract

Background: Major depressive disorder is a common diagnosis among adolescents. Cognitive behavioral therapy is recommended as the first line of treatment. Digital health interventions, such as apps, could contribute to treatment. Advantages could be easy accessibility and availability, reduced time for face-to-face therapy, and the ability to intensify therapy by incorporating it into the patients' everyday lives. Challenges such as low adherence rates are common in digital health interventions. Therefore, they need to undergo rigorous testing for feasibility and effectiveness.

Objective: An evaluated, cognitive behavioral therapy–based face-to-face therapy program for depression in adolescents was transformed into an app called e-MICHI. This study examined its feasibility and efficacy for use in blended therapy in outpatient settings.

Methods: Adolescents aged 12 to 18 years with major depressive disorder receiving outpatient care were recruited from 2 university hospitals (n=36 included in analysis). The e-MICHI intervention combined daily app engagement over 6 weeks with 3 face-to-face sessions with a therapist. Feasibility was measured using various variables, including an adherence score (0=no or little patient engagement to 3=excellent engagement) and engagement rates (number of modules completed, number of messages sent by participants via the in-app messenger), satisfaction ratings from both participants and therapists, as well as participants' ratings of the usefulness of the antidepressant strategies covered in the app and the transfer of these strategies to everyday practice. Trends of efficacy were evaluated from multiple perspectives (participant self-rating, independent rater, or therapist), using the Beck Depression Inventory-II, the Children's Depression Rating Scale-Revised, and the Clinical Global Impressions-Severity Scale. Feasibility metrics were assessed by analyzing their central tendency and dispersion, efficacy data were analyzed using a repeated measures ANOVA.

Results: e-MICHI was positively evaluated by both participants and therapists (participants: mean 7.3, SD 1.2 and therapists: mean 7.3, SD 1.1, on a scale from 0=bad to 10=excellent). Participants demonstrated high adherence rates (nearly 80%, n=25, received a "good" or "excellent" adherence score) and showed overall good engagement (app modules completed [maximum 6]: mean 5.03, SD 1.27 and messages sent via messenger: mean 23, SD 22.1). Participants rated the psychoeducational content of the app as particularly useful and reported consistent practice of the e-MICHI strategies in everyday life. Use of the app was associated with a significant reduction of depressive symptoms (before app use vs 3-month follow-up, Beck Depression Inventory-II: mean −6.76, SD 11.49, P=.01; Children's Depression Rating Scale-Revised: mean −16.45, SD 16.76, P<.001; Clinical Global Impressions-Severity Scale: mean −1.1, SD 1.24, P<.001).

XSL•FO
RenderX

Conclusions: While acknowledging its limitations, such as the small number of participants and the limited validity concerning efficacy, this study confirms the feasibility of e-MICHI for treating adolescent depression in outpatient settings.

Trial Registration: German Clinical Trial Register DRKS00022420; https://drks.de/search/en/trial/DRKS00022420/details

(JMIR Form Res 2025;9:e58427) doi:10.2196/58427

KEYWORDS

adolescents; major depressive disorder (MDD); cognitive behavioral therapy (CBT); digital health interventions (DHIs); feasibility; adherence; outpatient setting; mobile application; accessibility; availability; efficacy; mobile phone

Introduction

Major depressive disorder (MDD) is a prevalent psychiatric diagnosis among adolescents and the primary cause of inpatient treatment of adolescents in Germany [1,2]. Cognitive behavioral therapy (CBT) is the first-line treatment [3,4].

Digital health interventions (DHIs) could contribute to facilitated access to psychotherapeutic treatment [5]. Adolescents and young adults show the greatest openness and acceptance toward mental e-health interventions [6]. However, there is a scarcity of e-health interventions for depression in adolescents. Furukawa et al [7] found that all forms of internet-based CBT were superior to a waiting list control condition. Similarly, meta-analytical evidence indicates that DHIs are moderately to considerably effective (standardized mean difference 0.16 to 0.76) when compared to primarily inactive control groups [8]. Numerous studies have reported high dropout rates and low adherence levels for nonguided e-health interventions [9,10]. DHIs that include some form of personalized support, for example, from psychotherapeutic professionals, seem to be more efficacious due to higher adherence rates [11]. In addition to adherence, transferability presents a further challenge. The transfer of the contents and skills learned in e-health interventions to everyday practice by users seems to be poor: within the SPARX-R program, an intervention for adolescent depression, only around 40% of users practiced few, if any, of the techniques taught in the program [12]. Incorporating engaging elements such as text, audiovisual aids, and gamification, alongside the real-world implementation and involvement of parents and professionals could enhance this transfer [13].

There is an increasing availability of various DHIs for adolescents [14]. Due to the lack of scientific evaluation, Christ et al [15] called for further research and long-term follow-up assessments of these interventions. Furthermore, Seiferth et al [13] established guidelines for evidence-based DHIs including individuals with suicidal thoughts, emphasizing robustness, comparability, crisis planning, and access to additional treatments. This study aimed to assess the feasibility and indicators for efficacy of a blended DHI application, based on an evaluated face-to-face therapy program for depression in adolescents. The target population was adolescents with clinically verified depressive symptoms who were receiving child and adolescent psychiatric and psychotherapeutic care.

Methods

Procedure

This study used a single-group design to investigate the feasibility of a 6-week blended DHI for adolescents with depression. To this end, treatment adherence and the effects on symptom severity during the intervention period and a 90-day follow-up were analyzed.

This study was conducted at the Departments of Child & Adolescent Psychiatry at Rostock University Medical Center and University Hospital Ruppin-Brandenburg, in association with Brandenburg Medical School, Neuruppin. Adolescents aged between 12 and 18 years who met the criteria for a MDD, according to the *ICD-10* (*International Statistical Classification of Diseases and Related Health Problems, 10th Revision*) [16] as diagnosed by a certified child and adolescent psychiatrist or psychotherapist and were receiving outpatient care were included in the trial. Patients who met these inclusion criteria were informed about this study and asked whether they wanted to participate. If they and their parents agreed to participate, MDD diagnoses were confirmed by independent raters using the Beck Depression Inventory-II (BDI-II) [17] and the Children's Depression Rating Scale-Revised (CDRS-R) [18]. Additionally, a comprehensive diagnostic assessment was performed (Table 1).

XSL•FO
RenderX

Table . Overview of assessments conducted at different time points in a feasibility trial integrating the e-MICHI mental health app into a blended treatment of adolescent depression.

	Screening or recruitment	Pretreatment (T1)	Interim (T2)	Posttreatment (T3)	Follow-up (T4)
Trial days	–7	1	21	42	T3 +90
Diagnosis of depressive disorder	✓	N/A[a]	N/A	N/A	N/A
Inclusion and exclusion criteria	✓	N/A	N/A	N/A	N/A
Patient information and consent	✓	N/A	N/A	N/A	N/A
Demographics	N/A	✓	N/A	N/A	N/A
Previous and concomitant medication	N/A	✓	N/A	N/A	N/A
Concomitant diseases	N/A	✓	N/A	N/A	N/A
Treatment expectancy	N/A	✓	N/A	N/A	N/A
Adherence score	N/A	N/A	N/A	N/A	✓
Satisfaction and usability[b]	N/A	N/A	N/A	✓	N/A
BDI-II[c]	N/A	✓	✓	✓	✓
CDRS-R[d]	N/A	✓	N/A	✓	✓
CGI-S[e]	N/A	✓	✓	✓	✓
Adverse events	N/A	N/A	✓	✓	✓

[a]N/A: not applicable.

[b]Once per participant and once per therapist.

[c]BDI-II: Beck Depression Inventory-II [17].

[d]CDRS-R: Children's Depression Rating Scale-Revised (clinician rating) [18].

[e]CGI-S: Clinical Global Impressions-Severity Scale [19].

Additional psychopharmacological antidepressant treatment was allowed, and treatment as usual was provided as needed, to ensure that this study's population was representative of "real world" patients in routine care. Exclusion criteria for participation in the trial were acute suicidality requiring acute inpatient treatment, severe other psychiatric diagnoses such as schizophrenia and bipolar-disorder, and IQ below 85 or inadequate German language skills, as this would hinder the use of the program. Assessment for these criteria was based on clinical judgement following the Kinder-DIPS diagnostic interview for mental disorders [20].

Intervention

The e-therapy was delivered through a newly developed app called "e-MICHI" (Figure 1). e-MICHI is based on the MICHI therapy program ("Manualized Intervention to Cope With Depressive Symptoms, Help Strengthen Resources and Improve Emotion Regulation") [21], which has been shown to be effective for the treatment of depression in adolescents [22]. MICHI incorporates the proven effective elements of CBT for depression: psychoeducation, behavioral activation, cognitive restructuring, emotion regulation, enhancement of self-esteem, problem-solving, and coping with sudden crises [23]. The e-MICHI app consisted of 6 modules containing psychoeducational texts, exercises, quizzes, and a total of 27 video clips (Table 2). Participants were encouraged to complete each module within 1 week. In addition to the modules, e-MICHI included other features such as a mood diary and an in-app messenger. The messenger enabled communication between participant and therapist via text message. Once a participant had completed a module, the therapist received an automatic notification via messenger (eg, "I completed lesson 1!") and subsequently contacted the participant to address any queries or challenges related to the module. Participants could contact their therapist via messenger at any time if they had difficulty understanding the content of the modules or if they experienced a decrease in depressive symptoms. The app was hosted by Zone 35, a digital solutions provider [24].

XSL•FO
RenderX

Figure 1. Screenshots of the e-MICHI app interface, illustrating the home screen (providing an overview of progress and tasks) and the module start screen (introducing structured activities and content).

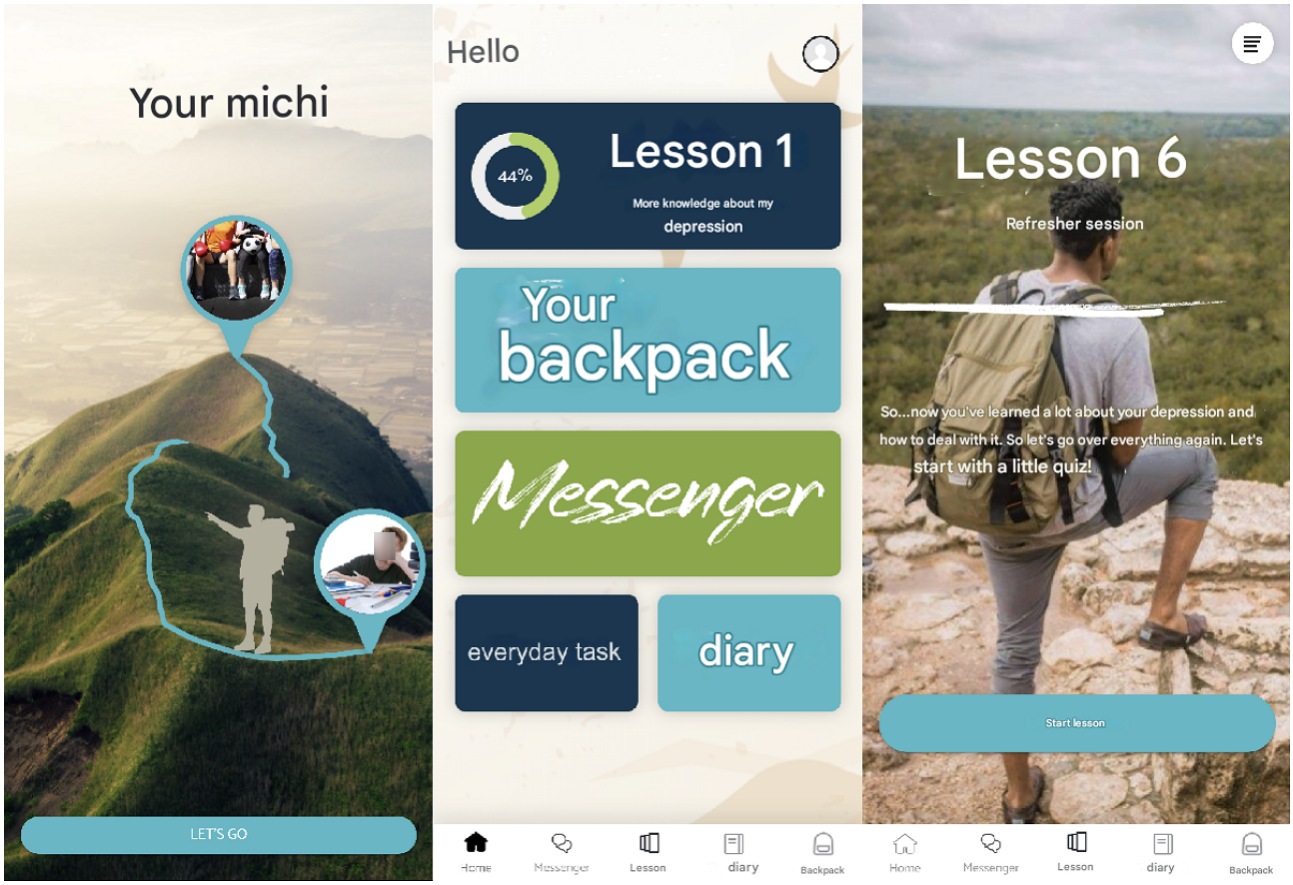


Table . Detailed summary of e-MICHI modules included in the feasibility trial. Each module's title, focus area, and corresponding therapeutic techniques are outlined.

	Content and psychoeducation	Exercises	Homework
1: "Knowledge of depression"	· Symptoms · Suicidality assessment	· Note personal takeaway[a] · Three quiz questions[a]	· Mood diary[b]
2: "Feelings"	· Depression causation · Mood improvement	· Personal disease model · Select pleasant activities · Identify strengths · Mind journey exercise	· Feedback on strengths · Pleasant activities · Mood diary[b] (including activities) · Feedback on strengths
3: "Influence on feelings"	· Positive self-esteem · Thoughts-behavior-feelings link	· Note compliments · Record thoughts, behaviors, or feelings in situations	· Mood diary[b] (behavior, thoughts, or feelings) · Practice complimenting
4: "Change thoughts"	· Crisis management · Cognition restructuring	· Note advice, link strategies to problems · Differentiate thoughts	· Personal emergency plan · Reflect on strategies with a therapist
5: "Support and health"	· Problem-solving skills	· Problem-solving exercise	· Appoint a "person of trust" · Share learned strategies
6: Refresher	· Quiz-based repetition	· Quiz on learned content	· No specific homework

[a]Repeats for lessons 1 - 5.

[b]7 days, 3 times per day.

Face-to-face contact was mandatory at baseline and after 3 and 6 weeks of using the app. Additional contact with the therapist was possible via messenger, phone, or as face-to-face contact, depending on the participant's needs. Every contact was recorded in a protocol.

XSL•FO
RenderX

As suicidality is a common symptom of depression, especially in adolescents, the first module included information about suicidal thoughts and how to seek immediate help at the hospital in case of acute suicidality. Additionally, when consent was obtained, information was provided about procedures for serious suicidal behavior and suicidal behavior was clarified as part of the diagnostic process (see "Ethical Considerations").

Measures

Numerous assessments were made at baseline (T1), 3 weeks (T2), and 6 weeks (T3), as well as at the 3-month follow-up (T4). A full list is provided in Table 2. The following estimates have been used as a basis for the assessment of feasibility: adherence was assessed by recording the number of modules completed and the number of messages sent by participants to their therapist via the in-app messenger. Therapists documented their weekly interactions with the patient, including any additional contacts initiated by the patient and conducted between the patient and therapist (via phone, video call, or in-person) as well as adherence, among other things. An adherence score was created, based on the therapists' weekly study documentation. For this purpose, 4 categories were used to rate each module, ranging from 0 ("little or no engagement") to 3 ("excellent engagement," meaning successful lesson completion, including the accomplishment of the everyday task). This rating was conducted by 2 independent raters; initially, each rater provided a rating per module, followed by reaching a consensus together, and then averaging over all evaluable modules.

Patient satisfaction was measured using a self-report questionnaire developed by the authors of the SPARX study [25], which was subsequently modified and translated for the purposes of this study. The questionnaire contained 37 quantitative and 5 qualitative items, which addressed various markers of user experience. The following section will focus exclusively on the items that were evaluated in the context of this paper. For an overall evaluation of the app the item "How would you rate the app overall on a scale of 1 - 10?" was asked (1=bad and 10=excellent). The impact of the app was measured by rating the usefulness of 7 antidepressant strategies addressed in the app using a 5-point Likert scale ranging from 0 ("very useless") to 4 ("very useful"). Transferability to daily life was measured by rating the frequency of use of 6 strategies on a 5-point Likert scale ranging from 0 ("never") to 4 ("nearly every day"). Total scores for the scales "usefulness of the strategies" and "frequency of use of the strategies" were then calculated by summing the individual items within each scale.

Therapists' overall satisfaction with the app was assessed using the item "How satisfied are you overall with the MICHI app?" on a scale from 1=not satisfied at all to 10=very satisfied. This item was part of a more extensive survey of therapists regarding their experiences with the app. A thorough evaluation of the therapists' perspectives on the app will be presented elsewhere [26].

Trends of efficacy were measured by changes in BDI-II [17], CDRS-R [18], and Clinical Global Impressions-Severity Scale [19] scores. The cutoff value of 29 in the BDI-II rating at T1 was used to classify the severity of depressive symptoms as mild or moderate versus severe [17].

Adverse events (AE) were coded according to MedDRA (Medical Dictionary for Regulatory Activities) terminology. Detailed information collected for each AE included a description of the event, duration, whether the AE was serious, intensity, relationship to trial treatment, action taken, and clinical outcome. Summary tables presented the number of participants observed with AEs by MedDRA System Organ Class and Preferred Term, along with the corresponding percentages. Additional subcategories were based on event intensity and relationship to trial treatment.

Ethical Considerations

The institutional review board of the Medical Faculty of the University of Rostock approved this study (registration number: A 2020 - 0093). Before enrollment, participants and parents were verbally informed about the objective and methodology of this study and provided with separate written information documents. Written consent was then obtained from both parents and participants. Participants were informed that they could withdraw from the study at any time without negative consequences for their treatment. Participants aged 18 years did not require parental consent. Participants were not financially compensated for their participation in this study, as they were allowed to take part in an innovative new treatment program.

Suicidality was assessed at baseline and in the first module of the e-MICHI app. Participants were queried about whether they experience thoughts of death. They were informed that if they do have thoughts of death, their primary therapist would be notified via messenger and would reach out to them.

All data were deidentified for storage and subsequent evaluation, ensuring participant confidentiality throughout the research process.

Procedures Against Bias

To reduce bias, independent raters, namely psychology students and research psychologists, assessed inclusion criteria and outcome measures, like the CDRS-R [18]. The raters underwent training in the administration of the diagnostic interviews, but were not acquainted with the participants' psychiatric and psychotherapeutic histories. Furthermore, the assessment of adherence scores was conducted independently by 2 authors (LL and AS). Discrepancies between their findings were discussed and resolved.

Data Analysis

Feasibility metrics were assessed by analyzing their central tendency and dispersion. Additionally, their association with the severity of depression was examined using appropriate statistical tests such as Pearson correlation coefficients, Mann-Whitney *U* tests, and 2-sample *t* tests.

Concerning trends of efficacy, 1-factor ANOVAs were conducted with 3 repeated measurements of the dependent variable (BDI-II and CDRS-R scores) within 1 sample, provided the prerequisite conditions were met. If the necessary conditions were not met, a nonparametric method was used for statistical

XSL•FO
RenderX

analysis. Previous research has shown that using DHIs to treat depression in children and teenagers has mild to moderate positive effects [27]. As a result, we expected a moderate effect size of Cohen *f*=0.25 for this investigation. With a priori of 28 participants in total and a 5% probability of type I error, as well as a power of 0.8 and a 20% probability of type II error, the sample size was required to observe significant changes in the sample as per the protocol. The G*Power (Heinrich-Heine-Universität Düsseldorf) tool was used to carry out the sample size calculation [28]. Considering an attrition rate of approximately one-third than that was reported in previous trials involving smartphone app interventions for depressive symptoms [29], it was deemed as a suitable anticipated dropout rate for the investigation, ending up with a total of 36 enrolled participants. This sample size enabled us to explore the impact of covariates and associations between different variables. Furthermore, violations beyond the per-protocol sample of 28 participants could be recorded.

The main analysis was conducted with the per-protocol group, consisting of all participants from the intention-to-treat population who did not violate the study protocol, by failing to meet all inclusion criteria, meeting at least 1 exclusion criterion or being noncompliant to therapy.

Subgroup analyses were performed to identify potential variances among predictive factors, such as adherence, comorbid disorders, and treatment expectancy. The statistical test interpretation was exploratory, except for the trends of efficacy outcomes for which this study was powered.

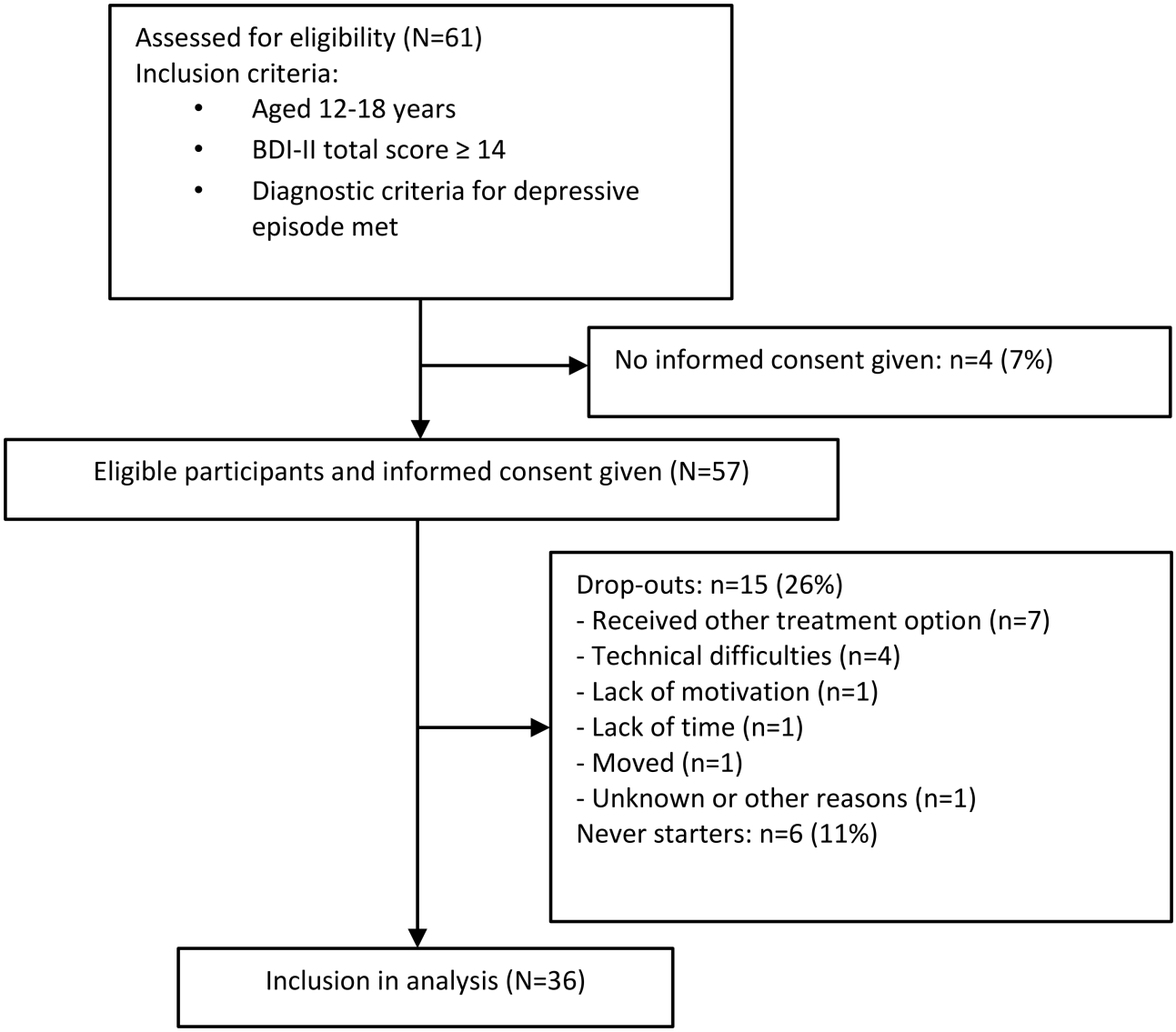
Results

Participants and Dropout

A total of 61 participants were assessed for eligibility in this study. While informed consent was obtained from 57 individuals, 6 (11%) participants did not commence with the program. A further 15 (26%) participants dropped out (Figure 2). Dropouts were defined as participants who completed at least 1 lesson within the app but stopped using it for a variety of reasons, such as becoming unreachable for further contact, starting external outpatient psychotherapy, or being admitted to a day clinic or inpatient facility. For demographic data on the participants, please refer to Table 3. There was no significant difference in age or severity of depressive symptoms at T1 between the dropout group and the included participants. Further dropout analysis was conducted using a chi-square test to examine differences by study center, gender, school type, education, treatment expectancy, and previous psychiatric or psychotherapeutic experience. The results show no significant differences between dropouts and other participants on these variables. Thus, dropouts were deemed irrelevant for further analyses.

XSL•FO
RenderX

Figure 2. Participant flowchart depicting the recruitment, eligibility screening, and retention rate in the feasibility trial integrating the e-MICHI mental health app into a blended treatment of adolescent depression. BDI-II: Beck Depression Inventory-II.



XSL•FO
RenderX

Table . Baseline demographic and clinical characteristics of participants (N=36) in the feasibility trial, including age, gender, distribution across outpatient centers, baseline depression severity, educational backgrounds, and prior treatment experience.

Characteristic		Values
Age, mean (SD)		
	Years	15 (1)
	Months	6 (7)
Age in years, n (%)		
	12	1 (3)
	13	7 (19)
	14	6 (17)
	15	10 (28)
	16	3 (8)
	17	8 (22)
	18	1 (3)
Gender, n (%)		
	Female	29 (83)
	Male	4 (11)
	Diverse	2 (6)
Center, n (%)		
	Neuruppin	17 (47)
	Rostock	19 (53)
Severity of depressive symptoms, n (%)		
	BDI-II<29	6 (17)
	BDI-II≥29	29 (83)
Education, n (%)		
	Grammar school	16 (46)
	Middle school	8 (23)
	Comprehensive school	4 (11)
	Vocational school	2 (6)
	Orientation stage	1 (3)
	Not attending school	4 (11)
Previous psychiatric or psychotherapy experience, mean (SD)		
	Years	1 (2)
	Months	8 (0)
Previous psychiatric or psychotherapy experience, n (%)		
	No previous experience	9 (26)
	≤1 year treatment	15 (43)
	>1 year treatment	11 (31)

Adherence and Overall Satisfaction

While 50% (n=17) of participants successfully completed all modules, 74% (n=25) of participants completed at least the first 5 modules covering all therapeutic content (Table 4). Regarding patient communication via the app's messaging feature, an average of 23 (SD 22.1) messages were sent per participant, with a high variance. According to the adherence score, almost 80% (n=25) of all participants showed good or excellent adherence (Table 4). On average, less than 1 additional contact initiated by the patient was documented by the therapist per week (mean 0.7, SD 0.31). There was no statistically significant difference in adherence score and messaging activity between participants with severe depressive symptoms and mild or moderate depressive symptoms (Table 4) and there was no

XSL•FO
RenderX

correlation between the number of messages sent by participants and the number of lessons completed (r=.33, P=.06).

Table . Feasibility measures, including completion rates for app modules, adherence to the intervention, engagement via messages, additional contacts, and relationships between feasibility outcomes and depression severity.

Variables	Values	*P* value
Modules completed (N=34), mean (SD)	5.03 (1.27)	N/A[a]
Modules completed, n (%)		N/A
1	34 (100)	
2	34 (100)	
3	31 (91)	
4	30 (88)	
5	25 (74)	
6	17 (50)	
Adherence score (N=32), n (%)		N/A
0 (no or little engagement)	0 (0)	
1 (moderate engagement)	7 (22)	
2 (good engagement)	10 (31)	
3 (excellent engagement)	15 (47)	
Messages sent (N=34), mean (SD)	23 (22.1)	
Additional contact per week, mean (SD)	0.7 (0.31)	
Severity of depression, *t* test *(df)*		
Perceived usefulness	0.732 (25)	.47
Strategy implementation	0.808 (24)	.43
Severity of depression, Mann-Whitney *U*		
Modules complete	56.5	.76
Messages sent	63	.73
Patient app rating	40	.1

[a]N/A: not applicable.

Further analysis also showed no correlation between adherence score and therapy expectation (r=.17, P=.34), the overall satisfaction with the app (r=–.07, P=.73), and changes in depressive symptoms, as measured by both BDI-II (r=.18, P=.40) and CDRS-R (r=−.24, P=.19) scales. There was no statistically significant difference between the adherence scores of the group with previous psychiatric or psychotherapeutic experience and the group with no previous experience (t_{30}=−1.22, P=.23).

Participants rated their overall satisfaction with the intervention as 7.3 on average (SD 1.2). Therapists indicated similar rates of overall satisfaction (mean 7.3, SD 1.1). Qualitative feedback from the therapists indicated that they particularly appreciated the inclusion of a mood diary due to its potential to monitor and reflect on emotional change.

Impact of the App and Transferability to Daily Life

A detailed analysis on the perceived usefulness of the therapeutic techniques addressed in the e-MICHI app is presented in Table 5. In general, a high rate of participants indicated using therapeutic techniques in everyday life, with variance between different techniques (Table 6).

XSL•FO
RenderX

Table . Participant feedback on the perceived usefulness of various therapeutic techniques addressed in the e-MICHI app based on self-reported relevance to managing depression symptoms (N=27).

	Very useless or useless, n (%)	Unsure, n (%)	Very useful or useful, n (%)
Psychoeducation	1 (4)	6 (22)	20 (74)
Behavioral change or activation	7 (26)	5 (19)	15 (56)
Identification of negative thoughts[a]	3 (11)	12 (44)	12 (44)
Changing negative thinking[a]	6 (22)	10 (37)	11 (41)
Problem solving	5 (19)	9 (33)	13 (48)
Feel better about myself[a]	10 (37)	6 (22)	11 (41)
Emotion regulation	6 (22)	9 (33)	12 (44)

[a]As part of cognitive restructuring.

Table . Participant-reported frequency of practicing therapeutic techniques addressed in the e-MICHI app at the third assessment point (T3), reflecting the integration of these strategies into participants' daily routines.

	N	Never, n (%)	A few times, n (%)	About once per week, n (%)	Several times per week, n (%)	Nearly every day, n (%)
Paid attention to basic needs	27	1 (4)	6 (22)	7 (26)	6 (22)	7 (26)
Spent time for activities that improve mood	27	1 (4)	3 (11)	7 (26)	11 (41)	5 (19)
Applied techniques of problem-solving	27	6 (22)	7 (26)	4 (15)	7 (26)	3 (11)
Identified positive things about yourself	27	2 (7)	6 (22)	9 (33)	7 (26)	3 (11)
Recognized negative automatic thoughts	27	3 (11)	6 (22)	6 (22)	9 (33)	3 (11)
Tried to change negative thoughts	26	2 (8)	7 (27)	6 (23)	9 (35)	2 (8)

A positive correlation was found between changes in BDI-II scores, reflecting a perceived decrease in depressive symptoms, and both the perceived overall usefulness of the techniques (r=.54, P=.004) and their total frequency of use (r=.62, P=.001). There was also a significant positive correlation between participants' overall satisfaction with the app and the perceived usefulness of the techniques (r=.64, P<.001), as well as the frequency of their implementation (r=.70, P<.001).

Based on the severity of depressive symptoms, the analysis revealed no significant differences in the perceived utility of the techniques or their implementation frequency (see Table 4).

A more detailed investigation was carried out to explore whether the perceived effectiveness of e-MICHI in "tackling problems or issues causing difficulty" was related to the severity of depression. Using a Mann-Whitney U test, no significant difference in the evaluation of the item's usefulness between participants with lower BDI-II scores and those with higher BDI-II scores was found (U=52.00, Z=−0.34, P=.76).

Trends of Efficacy of e-MICHI

Depressive symptoms decreased significantly during the blended therapy, regardless of the measure used (BDI-II, CDRS-R, and Clinical Global Impressions-Severity Scale scores), and remained stable at follow-up 3 months after using the app (Table 7).

XSL•FO
RenderX

Table . Efficacy measures evaluating changes in self-reported depression severity, clinician-rated changes in depressive symptoms, and severity of illness from baseline (T1) to follow-up (T4) in the feasibility trial of e-MICHI.

Variables		Values
BDI-II[a] (N=21)		
	T1, mean (SD)	37.71 (10.77)
	T2, mean (SD)	32.52 (12.89)
	T3, mean (SD)	30.62 (13.21)
	T4, mean (SD)	30.95 (15.38)
	F test (*df*)	4.10 (3, 60)
	P value	.01
CDRS-R[b] (N=27)		
	T1, mean (SD)	58.93 (11.51)
	T3, mean (SD)	48.37 (15.28)
	T4, mean (SD)	42.48 (16.61)
	F test (*df*)	16.39 (2, 52)
	P value	<.001
CGI-S[c] (N=29)		
	T1, mean (SD)	4.48 (0.79)
	T2, mean (SD)	3.97 (0.87)
	T3, mean (SD)	3.48 (0.95)
	T4, mean (SD)	3.38 (1.32)
	Chi-square (*df*)	33.45 (3)
	P value	<.001

[a]BDI-II: Beck Depression Inventory-II.

[b]CDRS-R: Children's Depression Rating Scale-Revised (clinician rating).

[c]CGI-S: Clinical Global Impressions-Severity Scale.

About AEs

There were 6 AEs of nonsuicidal self-injury, 1 of which was a serious adverse event (SAE) that required inpatient admission for crisis intervention. Thoughts of death were reported on 12 occasions, but there were no indications of acute suicidality. Another AE involved an outpatient procedure for a nonrelated medical condition. None of the 7 AEs, including the SAE, were deemed related to the intervention and this study's therapy continued unchanged.

Of the 7 dropouts who received other treatment options, 3 were inpatients and 2 were day patients (planned partial hospitalization), both categories coded as SAE. However, these SAEs were not considered to be related to the intervention. Additionally, there were 2 dropouts where study treatment was discontinued due to other outpatient treatment options.

Discussion

Principal Findings

This study aimed to investigate the feasibility of the DHI e-MICHI as a blended intervention to treat mild, moderate, and severe depression among adolescents. We found that adolescents adhered well to the intervention, indicated general satisfaction as well as a reduction in depressive symptoms from both patient and therapist perspectives. The use of the app was feasible in a population of adolescents with clinically significant depression and significant effects were observed in child and adolescent psychiatric outpatients. Improvement of depressive symptoms measured at the end of the intervention remained stable at the 3-month follow-up.

Crucial points in e-mental health applications are adherence and transferability to everyday practice [13]. We observed high adherence rates, which are comparable to other blended interventions [30] and substantially higher than unguided e-health applications [31,32]. In this study, participant satisfaction and reported transfer to daily practice were high. This differs from results of other studies, such as the SPARX-R program [12], which reported lower perceived usefulness and transfers in daily practice in adolescents excluded from mainstream education. A high transfer of psychotherapeutic exercises and skills into everyday life could be one of the most important benefits of using apps in psychotherapeutic practice. Apps could intensify therapy and with more individualized feedback, for example, based on ecological momentary

XSL•FO
RenderX

assessment or generated by AI, a tailored personalized psychotherapeutic intervention could be provided.

Concerning adherence, a substantial proportion of participants, about 3-quarters, completed the core modules of e-MICHI. Reasons for high adherence may include aesthetic design [33], relevant content [34], or the perceived immediacy of support through the app's messenger as stated by participants. These findings align with the principles highlighted by Domhardt et al [8]. Adolescents also reported applying the strategies they learned from the app in their daily lives, a promising sign of real-life transfer, even among those who rated their depressive symptoms as severe. The high adherence may be influenced by the fact that our study population consisted of adolescents seeking psychiatric-psychotherapeutic treatment who were clinically significantly depressed and impaired by their disorder. Interestingly, our findings indicate no significant differences in adherence between participants with long or short-term and no prior therapeutic experience. This highlights the possibility that even individuals with prior experience in therapy may derive benefits from e-MICHI suggesting the intervention could serve as a valuable adjunct to traditional therapeutic approaches, enhancing treatment outcomes and promoting sustained engagement across diverse patient populations. Additionally, the absence of an existing therapeutic relationship for adequate adherence would not limit the use of DHIs, for example, to bridge the gap for face-to-face psychotherapy or as a primary intervention. Notably, adolescents appreciated the close contact with therapists provided by the messaging function, highlighting the added benefit in blended approaches [35]. DHIs may therefore be more beneficial when embedded in conventional care settings.

The acceptance of the app is evidenced by the high satisfaction ratings from participants and therapists alike. The high acceptance rate, especially among adolescents, may be partly due to their general openness to digital interventions [6]. In combination with the therapeutic relationship and the facilitated way of communicating directly with the therapist via the app, e-MICHI, or DHIs in general, seem to be an added value in psychiatric-psychotherapeutic treatment for adolescents.

The inclusion of a mood diary, as appreciated by the therapists, could enhance therapeutic engagement and outcomes due to its potential to track and reflect on emotional changes [36]. This consensus, together with the high level of satisfaction, supports the usefulness of the app as a complementary tool in psychotherapy and its potential to strengthen the therapeutic alliance, as also shown by Tremain et al [37].

In our study, additional contacts beyond the planned 3 sessions were rare, indicating efficient management of patient care with reduced direct contact frequency compared to our clinical experience, where patients in outpatient settings are typically seen every 1 to 2 weeks. Additionally, compared to treatment as usual, messenger contacts helped provide feedback on mood diary entries and suggested steps for e-MICHI daily task implementation.

This finding aligns with the objective of providing more accessible care [38] as it can reduce both waiting times to start psychotherapy and the need to travel long distances to receive care. It must be noted, that technical problems and difficulties with devices or limited network coverage or access to free Wi-Fi may be barriers to the use of therapeutic apps, especially for adolescents. In this study, 4/57 (7%) of originally included participants dropped out due to technical problems. Access to free Wi-Fi seems to be essential for adolescents to be able to use the app in their living environments. Ways to use the app offline would be desirable to make it accessible to even more patients.

Regarding trends of efficacy, there was a noticeable decrease in depressive symptoms during the intervention, as reported by both participants and external evaluators. This significant improvement highlights the potential of e-MICHI in the treatment of depression, warranting further validation through randomized controlled trials (RCTs) to more robustly establish its clinical efficacy.

Given the sample, which included participants with severe depression, drop out due to referral to other psychiatric treatment services was not surprising. The advantage of the blended approach is evident in the case of a patient who was admitted for crisis intervention and was able to resume the intervention after discharge providing continuity of care and additional therapist support, consequently increasing adherence.

Regarding safety, this study presents an optimistic perspective on the e-MICHI blended CBT intervention for adolescents with depression, as evidenced by a small number of unrelated AEs and no suicide attempts. Comparatively, Rasing et al [39] observed 2 suicide attempts in a similar blended CBT intervention, but found no significant differences in AEs between face-to-face and blended treatment conditions. A recent review indicated no increased risk of suicidality with blended treatment for adolescents with depression [40]. Concerning self-harm, Fleming et al [41] reported a comparable number of 2 incidents in a similar sample size, neither of which were related to the intervention. The absence of suicide attempts during the e-MICHI intervention, along with the findings from Rasing et al [39] and Fleming et al [41], suggests that blended interventions may not substantially increase the risk of severe AEs, such as suicide attempts and self-harm. Nonetheless, Domhardt et al [8] emphasize the need for continued vigilance and monitoring in blended mental health interventions, especially for high-risk populations.

Contrasting the results of e-MICHI with similar blended interventions, e-MICHI allowed participants to use their own devices, incorporating beneficial strategies such as text messaging for homework reminders and self-monitoring between sessions [42], and included participants with severe depressive symptoms, addressing the criticism regarding the exclusion of such cases in previous research [43]. As noted, we found no significant difference in adherence rates in participants with severe depressive symptoms. While this is consistent with previous findings [44], this result may also be due to the small sample size.

Comparing e-MICHI to unguided internet-based CBT [40] highlights the potential benefits of a blended approach. The higher adherence rates and satisfaction scores suggest that integrating therapist support into an e-health platform may

XSL•FO
RenderX

enhance the perceived usefulness of these interventions, as well as how well adolescents are able to apply the therapeutic strategies they learn to their daily lives.

Limitations

The non-RCT design of this study limits the generalizability of our findings on efficacy. As a feasibility study we wanted to detect trends of efficacy; the trial was not designed for efficacy testing. An RCT will be the next step in generating evidence for the e-MICHI intervention. Most of this study's population had a history of psychotherapy. Therefore, our participants might be more familiar with therapeutic contents than adolescents who use such a therapeutic app as a first-time psychotherapeutic intervention. The adherence scoring method faced challenges due to several therapy documentations missing or lacking sufficient information for rating. Given the small sample size, it is possible that these factors could have influenced the findings. For the development of therapeutic interventions according to precision medicine, it would be of interest to know about precise times of use of specific contents (eg, in the morning, evening, etc), direct effects on symptoms (eg, mood), etc. Unfortunately, due to technical and privacy restrictions, we were unable to evaluate this data.

Conclusion

Due to its easy application, the e-MICHI intervention seems to be a promising tool to enhance face-to-face treatment of depression in adolescents. This study identified areas for further development of the app based on user feedback and technical issues, to improve its effectiveness and user experience in the future. It contributes to filling the existing evidence gap on the feasibility and efficacy of e-therapy interventions via apps for adolescents with depression in the German health care system. According to the results of this study, the potential benefits of blended therapy for depression by intensifying therapy and facilitating patient-therapist communication are evident. As digital interventions continue to evolve, further research is imperative to address challenges, for example, accommodating a broad range of individuals, crisis intervention, and high dropout rates, ensuring that such platforms can effectively serve as viable therapeutic options for adolescents with depression.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

MK receives royalties from publishing companies as an author of books. He served as a scientific advisor for Janssen. He received grants from the German Ministry of Research and Education and the Ministry of Social Affairs, Health and Sports Mecklenburg-Vorpommern. This work is unrelated to the above grants and relationships. LL, TM-D, AD, AS, DAN, NS-K, SB, and OR declare no competing interests.

References

1. Kölch M, Reis O, Ulbrich L, Schepker R. Mental disorders in minors during the COVID-19 pandemic-an analysis of nationwide hospital treatment data. Dtsch Arztebl Int 2023 May 19;120(20):362-363. [doi: 10.3238/arztebl.m2023.0010] [Medline: 37493050]
2. Witte J, Zeitler A, Batram M, Diekmannshemke J, DAK-Kinder HL. Und jugendreport 2022: DAK im dialog. 2022 URL: https://caas.content.dak.de/caas/v1/media/45734/data/21b30080687f14a8c09d7a54370f8860/kinder-und-jugendreport-2022-sammelband.pdf [accessed 2025-03-27]
3. Dolle K, Schulte-Körne G. Behandlung von depressiven Störungen bei Kindern und Jugendlichen: Evidenz- und konsensbasierte Leitlinie. Prax Kinderpsychol Kinderpsychiatr 2013;63(3):237-263. [doi: 10.13109/prkk.2014.63.3.237]
4. Depression in children and young people: identification and management (NG134). National Institute for Health and Care Excellence. 2019. URL: https://www.nice.org.uk/guidance/ng134 [accessed 2025-03-26]
5. Allgaier K, Schmid J, Hollmann K, Reusch PA, Conzelmann A, Renner TJ. Times are changing: digitalisation in child and adolescent psychotherapy. Eur Child Adolesc Psychiatry 2021 Nov;30(11):1667-1670. [doi: 10.1007/s00787-020-01610-8] [Medline: 32737600]
6. Lincke L, Ulbrich L, Reis O, et al. Attitudes toward innovative mental health treatment approaches in Germany: e-mental health and home treatment. Front Psychiatry 2022;13:889555. [doi: 10.3389/fpsyt.2022.889555] [Medline: 35911231]
7. Furukawa TA, Suganuma A, Ostinelli EG, et al. Dismantling, optimising, and personalising internet cognitive behavioural therapy for depression: a systematic review and component network meta-analysis using individual participant data. Lancet Psychiatry 2021 Jun;8(6):500-511. [doi: 10.1016/S2215-0366(21)00077-8] [Medline: 33957075]
8. Domhardt M, Steubl L, Baumeister H. Internet- and mobile-based interventions for mental and somatic conditions in children and adolescents. Z Kinder Jugendpsychiatr Psychother 2020 Jan;48(1):33-46. [doi: 10.1024/1422-4917/a000625] [Medline: 30422059]
9. Achilles MR, Anderson M, Li SH, Subotic-Kerry M, Parker B, O'Dea B. Adherence to e-mental health among youth: considerations for intervention development and research design. Digit Health 2020;6:2055207620926064. [doi: 10.1177/2055207620926064] [Medline: 32547775]

XSL•FO
RenderX

10. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
11. Lehtimaki S, Martic J, Wahl B, Foster KT, Schwalbe N. Evidence on digital mental health interventions for adolescents and young people: systematic overview. JMIR Ment Health 2021 Apr 29;8(4):e25847. [doi: 10.2196/25847] [Medline: 33913817]
12. Kuosmanen T, Fleming TM, Newell J, Barry MM. A pilot evaluation of the SPARX-R gaming intervention for preventing depression and improving wellbeing among adolescents in alternative education. Internet Interv 2017 Jun;8(40-47):40-47. [doi: 10.1016/j.invent.2017.03.004] [Medline: 30135827]
13. Seiferth C, Vogel L, Aas B, et al. How to e-mental health: a guideline for researchers and practitioners using digital technology in the context of mental health. Nat Mental Health 2023;1(8):542-554. [doi: 10.1038/s44220-023-00085-1]
14. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176. [doi: 10.2196/jmir.7332] [Medline: 28546138]
15. Christ C, Schouten MJ, Blankers M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in adolescents and young adults: systematic review and meta-analysis. J Med Internet Res 2020 Sep 25;22(9):e17831. [doi: 10.2196/17831] [Medline: 32673212]
16. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. World Health Organization. 1992. URL: https://cdn.who.int/media/docs/default-source/classification/other-classifications/9241544228_eng.pdf [accessed 2025-03-27]
17. Beck AT, Steer RA, Brown G. Manual for Beck Depression Inventory-II: Psychological Corporation; 1996.
18. Keller F, Grieb J, Ernst M, Spröber N, Fegert JM, Kölch M. Children's Depression Rating Scale-Revised (CDRS-R): development of a German version and psychometric properties in a clinical sample. Z Kinder Jugendpsychiatr Psychother 2011 May;39(3):179-185. [doi: 10.1024/1422-4917/a000090] [Medline: 21563109]
19. Guy W. ECDEU Assessment Manual for Psychopharmacology: Government Printing Office; 1976.
20. Schneider S, Pflug V, Margraf J, In-Albon T. Kinder-DIPS: Diagnostic Interview for Mental Disorders in Children and Adolescents [Diagnostisches Interview Bei Psychischen Störungen Im Kindes- Und Jugendalter]: Mental Health Research and Treament Center, Ruhr-Universität Bochum; 2017.
21. Spröber N, Fegert JM, Kölch M, Straub J. Depressionen Im Jugendalter: MICHI ; Manual Für Die Gruppentherapie: Beltz; 2012.
22. Straub J, Plener PL, Keller F, Fegert JM, Spröber N, Kölch MG. MICHI–eine Gruppen-Kurzzeitpsychotherapie zur Behandlung von Depressionen bei Jugendlichen. Kindheit Entwickl 2015 Jul;24(3):189-198. [doi: 10.1026/0942-5403/a000175]
23. McCarty CA, Weisz JR. Effects of psychotherapy for depression in children and adolescents: what we can (and can't) learn from meta-analysis and component profiling. J Am Acad Child Adolesc Psychiatry 2007 Jul;46(7):879-886. [doi: 10.1097/chi.0b013e31805467b3] [Medline: 17581452]
24. Zone35 individual. innovative. digital. Zone35. 2020. URL: https://zone35.de/#home [accessed 2025-03-26]
25. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MFG. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. BMJ 2012 Apr 18;344:e2598. [doi: 10.1136/bmj.e2598] [Medline: 22517917]
26. Lincke L, Kilb H, Martin-Döring T, et al. Blended care for adolescents with depression from the perspective of treatment providers - feasibility study in an outpatient setting [Blended care für jugendliche mit depressiven erkrankungen aus der sicht von behandelnden - machbarkeitsstudie in einem ambulanten setting]. Kindheit Entwickl .
27. Hollis C, Falconer CJ, Martin JL, et al. Annual research review: digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
28. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
29. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
30. Schuster R, Leitner I, Carlbring P, Laireiter AR. Exploring blended group interventions for depression: randomised controlled feasibility study of a blended computer- and multimedia-supported psychoeducational group intervention for adults with depressive symptoms. Internet Interv 2017 Jun;8(63-71):63-71. [doi: 10.1016/j.invent.2017.04.001] [Medline: 30135830]
31. Melville KM, Casey LM, Kavanagh DJ. Dropout from Internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(Pt 4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]
32. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2009 May;39(5):705-712. [doi: 10.1017/S0033291708004224] [Medline: 18812006]
33. Denison-Day JL, Muir S, Newell C, Appleton KM. The role of aesthetics in intentions to use digital health interventions. PLOS Digit Health 2023 Jun;2(6):e0000274. [doi: 10.1371/journal.pdig.0000274] [Medline: 37347727]

XSL•FO
RenderX

34. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
35. Lindhiem O, Bennett CB, Rosen D, Silk J. Mobile technology boosts the effectiveness of psychotherapy and behavioral interventions: a meta-analysis. Behav Modif 2015 Nov;39(6):785-804. [doi: 10.1177/0145445515595198] [Medline: 26187164]
36. Matthews M, Doherty G, Coyle D, Sharry J. Designing mobile applications to support mental health interventions. In: Lumsden J, editor. Handbook of Research on User Interface Design and Evaluation For Mobile Technology 2008:635-656. [doi: 10.4018/978-1-59904-871-0.ch038]
37. Tremain H, McEnery C, Fletcher K, Murray G. The therapeutic alliance in digital mental health interventions for serious mental illnesses: narrative review. JMIR Ment Health 2020 Aug 7;7(8):e17204. [doi: 10.2196/17204] [Medline: 32763881]
38. Bundespsychotherapeutenkammer. Wartezeiten in Der Psychotherapie: BPtK Spezial; 2011, Vol. 6:1-2.
39. Rasing SPA, Stikkelbroek YAJ, den Hollander W, et al. Pragmatic quasi-experimental controlled trial evaluating the outcomes of blended CBT compared to face-to-face CBT and treatment as usual for adolescents with depressive disorders. Int J Environ Res Public Health 2021 Mar 17;18(6):3102. [doi: 10.3390/ijerph18063102] [Medline: 33802913]
40. Rasing SPA. Blended treatment for depressive disorders in youth: a narrative review. J Cogn Ther 2021 Mar;14(1):47-58. [doi: 10.1007/s41811-020-00088-1]
41. Fleming T, Dixon R, Frampton C, Merry S. A pragmatic randomized controlled trial of computerized CBT (SPARX) for symptoms of depression among adolescents excluded from mainstream education. Behav Cogn Psychother 2012 Oct;40(5):529-541. [doi: 10.1017/S1352465811000695] [Medline: 22137185]
42. Kobak KA, Mundt JC, Kennard B. Integrating technology into cognitive behavior therapy for adolescent depression: a pilot study. Ann Gen Psychiatry 2015;14(37):37. [doi: 10.1186/s12991-015-0077-8] [Medline: 26535048]
43. Sethi S, Campbell AJ, Ellis LA. The use of computerized self-help packages to treat adolescent depression and anxiety. J Technol Hum Serv 2010 Aug 31;28(3):144-160. [doi: 10.1080/15228835.2010.508317]
44. De Vos R, Tromp N, Bodden D, Stikkelbroek Y. Therapeut onmisbaar bij 'blended therapie' tegen depressie. Kind Adolesc Prakt 2017 Jun;16(2):6-13. [doi: 10.1007/s12454-017-0014-x]

Abbreviations

AE: adverse event
BDI-II: Beck Depression Inventory-II
CBT: cognitive behavioral therapy
CDRS-R: Children's Depression Rating Scale-Revised
DHI: digital health intervention
***ICD-10*:** *International Statistical Classification of Diseases and Related Health Problems, 10th Revision*
MDD: major depressive disorder
MedDRA: Medical Dictionary for Regulatory Activities
MICHI: Manualized Intervention to Cope With Depressive Symptoms, Help Strengthen Resources, and Improve Emotion Regulation
RCT: randomized controlled trial
SAE: serious adverse event

Edited by A Mavragani; submitted 18.03.24; peer-reviewed by K Sprange; revised version received 14.01.25; accepted 15.01.25; published 01.05.25.

Please cite as:
Lincke L, Martin-Döring T, Daunke A, Sadkowiak A, Nolkemper DA, Sproeber-Kolb N, Bienioschek S, Reis O, Kölch M
Integration of a Mental Health App (e-MICHI) Into a Blended Treatment of Depression in Adolescents: Single-Group, Naturalistic Feasibility Trial
JMIR Form Res 2025;9:e58427
URL: https://formative.jmir.org/2025/1/e58427
doi:10.2196/58427

© Lena Lincke, Tim Martin-Döring, Andrea Daunke, Antonia Sadkowiak, Daria Alexandra Nolkemper, Nina Sproeber-Kolb, Stefanie Bienioschek, Olaf Reis, Michael Kölch. Originally published in JMIR Formative Research (https://formative.jmir.org), 1.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,

XSL•FO
RenderX

provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Identification of Major Bleeding Events in Postoperative Patients With Malignant Tumors in Chinese Electronic Medical Records: Algorithm Development and Validation

Hui Li[1], PhD; Haiyang Yao[2], MD; Yuxiang Gao[2], BA; Hang Luo[2], MD; Changbin Cai[3], MD; Zhou Zhou[3], MD; Muhan Yuan[3], MD; Wei Jiang[4], MD

[1]Department of Thoracic Surgery, Beijing Chao-Yang Hospital, Capital Medical University, No.8 South Road of Workers' Stadium, Chaoyang District, Beijing, China

[2]Department of Technology, Shanghai Palan DataRx Co., Ltd, Shanghai, China

[3]Sanofi China Medical Affairs, Sanofi China, Shanghai, China

[4]Pharmaceutical Business Division, Basebit (Shanghai) Information Technology Co., Ltd, Shanghai, China

Corresponding Author:
Hui Li, PhD
Department of Thoracic Surgery, Beijing Chao-Yang Hospital, Capital Medical University, No.8 South Road of Workers' Stadium, Chaoyang District, Beijing, China

Abstract

Background: Postoperative bleeding is a serious complication following abdominal tumor surgery, but it is often not clearly diagnosed and documented in clinical practice in China. Previous studies have relied on manual interpretation of medical records to determine the presence of postoperative bleeding in patients, which is time-consuming and laborious. More critically, this manual approach severely hinders the efficient analysis of large volumes of medical data, impeding in-depth research into the incidence patterns and risk factors of postoperative bleeding. It remains unclear whether machine learning can play a role in processing large volumes of medical text to identify postoperative bleeding effectively.

Objective: This study aimed to develop a machine learning model tool for identifying postoperative patients with major bleeding based on the electronic medical record system.

Methods: This study used data from the available information in the National Health and Medical Big Data (Eastern) Center in Jiangsu Province of China. We randomly selected the medical records of 2,000 patients who underwent in-hospital tumor resection surgery between January 2018 and December 2021 from the database. Physicians manually classified each note as present or absent for a major bleeding event during the postoperative hospital stay. Feature engineering involved bleeding expressions, high-frequency related expressions, and quantitative logical judgment, resulting in 270 features. Logistic regression (LR), K-nearest neighbor (KNN), and convolutional neural network (CNN) models were developed and trained using the 1600-note training set. The main outcomes were accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for each model.

Results: Major bleeding was present in 4.31% (69/1600) of the training set and 4.75% (19/400) of the test set. In the test set, the LR method achieved an accuracy of 0.8275, a sensitivity of 0.8947, a specificity of 0.8241, a PPV of 0.2024, an NPV of 0.9937, and an F_1-score of 0.3301. The CNN method demonstrated an accuracy of 0.8900, sensitivity of 0.8421, specificity of 0.8924, PPV of 0.2807, NPV of 0.9913, and an F_1-score of 0.4211. While the KNN method showed a high specificity of 0.9948 and an accuracy of 0.9575 in the test set, its sensitivity was notably low at 0.2105. The C-statistic for the LR method was 0.9018 and for the CNN method was 0.8830.

Conclusions: Both the LR and CNN methods demonstrate good performance in identifying major bleeding in patients with postoperative malignant tumors from electronic medical records, exhibiting high sensitivity and specificity. Given the higher sensitivity of the LR method (89.47%) and the higher specificity of the CNN method (89.24%) in the test set, both models hold promise for practical application, depending on specific clinical priorities.

(JMIR Form Res 2025;9:e66189) doi:10.2196/66189

KEYWORDS

machine learning; electronic medical record; postoperative patients with malignant tumors; postoperative bleeding; tumor surgery; abdominal

XSL•FO
RenderX

Introduction

Bleeding events are frequent complications encountered in postoperative clinical settings and can stem from the use of anticoagulant or antiplatelet drugs, invasive surgical treatment, or patient-related conditions and the existence of comorbidities, which are associated with increased morbidity, mortality, and health care costs [1-4]. Patients experiencing gastrointestinal bleeding while in the intensive care unit face a fourfold increase in mortality risk compared to those without bleeding issues, along with an additional eight-day stay within the unit [4]. Ample clinical evidence supports the correlation between intraoperative blood loss, especially excessive blood loss and adverse effects on the tumor prognosis. On the other hand, concern about postoperative bleeding may become the main reason why clinicians might be overly cautious in using medications to prevent venous thrombosis (VTE), despite VTE being the second leading cause of death among patients with tumors. While the mortality rate of patients with tumors and VTE is three times higher than that of other patients, timely detection of bleeding risks through progress notes and balanced selection of anticoagulants are crucial for postoperative patients with malignant tumors [5,6].

Currently, there is a growing number of retrospective studies on postoperative bleeding risk in cancer patients. It is also worth noting the lack of fully established risk prediction schemes or risk assessment tools for postoperative bleeding. However, a major challenge in conducting these studies is determining whether patients experienced bleeding events during their past treatment processes. In the clinical setting in China, current approaches to identifying bleeding episodes predominantly rely on diagnostic records, which often lack precision due to inconsistent descriptions and missing details. The task of pinpointing bleeding events from a patient's medical history can be particularly challenging, especially when dealing with extensive records [7,8]. Despite the transition to electronic medical records, these valuable sources of data are frequently underused. Within these records, details within course notes, including physical examination reports and discharge summaries, often contain firsthand accounts of bleeding incidents or clear indications of their absence. Nonetheless, manual identification of these events can be both time-consuming and arduous, raising concerns about accuracy and practicality. Thus, emphasis should be placed on the imperative need to develop methodologies geared toward effectively identifying bleeding events within existing medical records.

In the realm of clinical research, the accurate identification of bleeding events within large clinical datasets holds paramount importance. Regrettably, the current landscape lacks automated and scalable machine learning (ML) methodologies tailored for this objective. This is a significant unmet need, especially considering the clinical need for early detection of postoperative bleeding to improve patient outcomes. ML has been regarded as a method for developing models that depict intricate nonlinear systems and handling a vast array of potential variables found in contemporary electronic medical records. ML techniques have found application in various health care scenarios such as forecasting cancer susceptibility, automatically categorizing clinical images, and predicting post-transplant prognosis. This is enabled by the abundance of real-world longitudinal datasets derived from the extensive integration of electronic health record (EHR) [9-11]. For health care challenges involving large and complex datasets, especially those with unstructured data, ML methods have demonstrated advantages over traditional statistical regression methods [12,13]. At present, mainstream text recognition methods such as support vector machine (SVM) or random forest (RF) are mostly suitable for large sample size data with high positivity rates, and the recognition of bleeding events may be a process of searching for sporadic positive events in large sample sizes [14,15]. The aim of this study was to develop a machine learning model tool for identifying postoperative patients with major bleeding events based on the electronic medical record system.

Methods

Population

This study used a retrospective design. The data for this study were obtained from the National Health and Medical Big Data (Eastern) Center in Jiangsu Province, which is maintained by the Jiangsu Provincial Health Commission. This database includes clinical data from hospitals in Jiangsu province, with sensitive and identifiable information removed to protect privacy. Retrospective electronic medical records data, including clinical notes were used for identifying the study population. We specifically focused on patients with malignant tumor, who underwent surgical procedures between January 2018 and December 2021, reflecting actual clinical scenarios. To more accurately identify high-risk individuals for bleeding, we chose to focus on patients with chest, abdominal and gynecological malignant tumor instead of urinary system tumors. This decision was based on the fact that surgeries for chest, abdominal, and gynecological tumors involve greater surgical trauma, making postoperative bleeding more likely. Additionally, secondary tumors were excluded because they were typically palliatively resected, resulting in a lower risk of bleeding compared to primary tumors.

The primary objective was to develop a classifier capable of recognizing significant cases of bleeding with clinical relevance. To achieve this, we randomly allocated 1600 notes for training purposes and 400 notes for testing the model.

The inclusion criteria in the study population are as follows: (1) patients were diagnosed with primary chest, abdominal, or gynecological malignant tumors; (2) patients underwent surgical resection of the malignant tumor; and (3) patients were ≥18 years old. The exclusion criteria are as follows: (1) patients were diagnosed with urinary system tumors; (2) patients were diagnosed with a secondary tumor; (3) patients only underwent endoscopic surgery (ie, gastroscopy, enteroscopy, cystoscopy); (4) patients diagnosed with bleeding before surgery or who underwent surgery due to bleeding events; and (5) hospitalization course record for the given visit is missing.

Each note was classified as major bleeding present, indicating that clinically relevant bleeding was referenced in the note, or

XSL•FO
RenderX

major bleeding absent, indicating that clinically relevant bleeding was not referenced in the note. Major bleeding was defined as fatal or symptomatic bleeding in a critical area or organ or bleeding causing a fall in hemoglobin level of ≥2 g/dL or transfusion of ≥2 units of erythrocyte concentrate, according to the definition of the International Society of Thrombosis and Haemostasis [16].

We included medical history and physical examination notes, progress notes, and discharge summaries that were signed by an attending physician. All investigators with direct access to the data completed a data use agreement. An independent medical professional was selected to read the medical and surgical records and label whether there was bleeding, and the bleeding classification (ie, major bleeding or nonmajor bleeding). Then, another independent medical professional was designated to review the labeled cases and extract or highlight the text content in the medical records where bleeding events occurred for reference in developing bleeding-related regular expressions. Any disagreements or unclear content in the text reading between the two medical professionals were recorded and discussed in an external expert workshop. In the workshop, questionable clinical scenarios were confirmed by two experienced medical experts.

Feature Engineering

Given the low positivity rate and limited sample size of major bleeding events in the existing data, the methodology for identifying major bleeding events during manual annotation was carefully considered when crafting features. A significant portion of these features was extracted through the application of natural language processing (NLP) techniques like Jieba (for Chinese word splitting) and regular expressions (for text and value extraction). These features are categorized as follows:

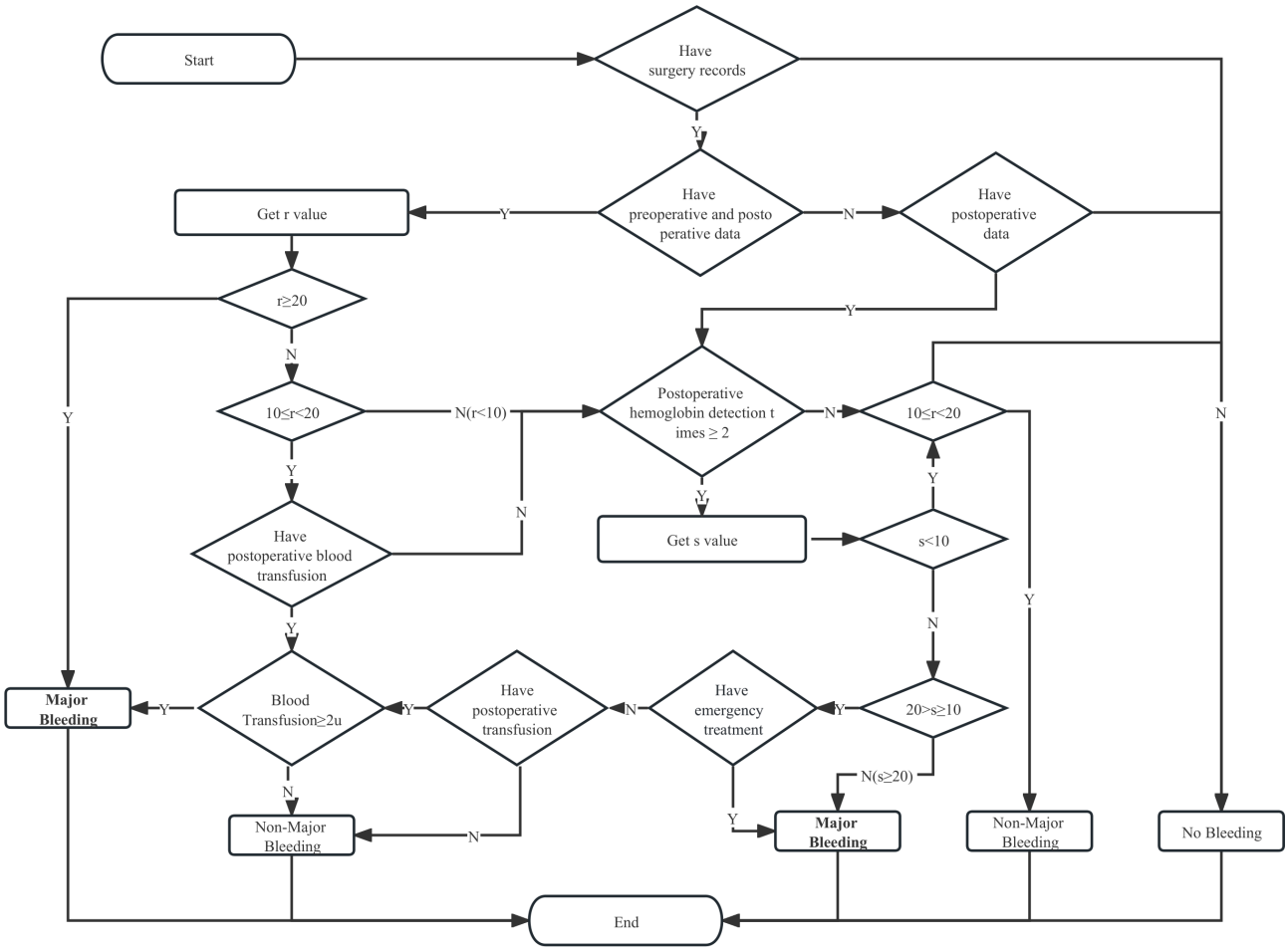
The first category: These features were generated from the content of manually annotated major bleeding events in the training set. This set includes 20 key features strongly linked to postoperative bleeding, such as terms like “postoperative,” “ hematochezia,” “ hemorrhagic fluid,” and “treatment with hemostasis surgery.” Furthermore, 241 additional features were compiled by segmenting factors associated with major bleeding events.

The second category: These features were derived by tokenizing the patient’s course texts from the training set using Jieba and arranging them based on their frequency of occurrence. These features represent words that exhibit some relevance to major bleeding events but were not present in the first part. Features in this category were chosen based on a minimum frequency threshold of 900, aiming to complement the initial features and mitigate any shortcomings due to manually crafted features.

The third category: This feature reflects insights from clinical experts and primarily revolves around indicators like preoperative, intraoperative, and postoperative hemoglobin levels, bleeding volume, and transfusion volume. They are instrumental in establishing the logic for identifying postoperative major bleeding for each patient. Features in this segment heavily rely on quantitative logical assessments to fill in any gaps present in the preceding two feature groups. For instance, transfusion volumes extracted from structured surgical records and transfusion documentation are analyzed in conjunction with the standardized decision-making flowchart to determine major bleeding status. A detailed, structured process identification diagram is presented in Figure 1.

XSL•FO
RenderX

Figure 1. Structured process identification diagram for feature selection.



Note:
r value: Preoperative last hemoglobin value (g/L) - Postoperative first hemoglobin value (g/L) - Intraoperative blood loss (ml)/33.3
s value: The difference between the first and second hemoglobin values (g/L) taken within 3 days after surgery.

Model Training

A supervised learning approach was adopted, where a dataset comprising 2000 patients was divided into a training set of 1600 patients and a test set of 400 patients. Using the features generated through feature engineering and the corresponding labels assigned during manual review, three distinct ML models were successfully trained using different algorithms: logistic regression (LR), k-nearest neighbors (KNN), and convolutional neural networks (CNN).

Here is a breakdown of how each model was implemented:

1. LR:

- Model implementation: developed using Keras
- Architectural details: constructed with a single dense layer
- Activation function: sigmoid function
- Loss function: SigmoidFocalCrossEntropy
- Optimizer: Adam optimizer
- Validation split: 20% of the training dataset is set aside for validation

1. KNN:

- Model implementation: Used the KNeighborsClassifier class from Scikit-Learn.

1. CNN:

- Model implementation: implemented using Keras
- Architectural details: featured 3 hidden layers - a 1-dimensional convolutional layer, a max pooling layer, and a dense layer
- Activation functions: Rectified Linear Unit for the convolutional layer and sigmoid for the output layer
- Output layer: a single dense unit
- Validation split: Again, 20% of the training data is reserved for validation.

All three ML models were trained on the training set consisting of 1600 cases to learn the underlying patterns between the extracted features and the corresponding labels. The use of LR, KNN, and CNN allows for a diverse exploration of the dataset with different algorithms to harness their unique strengths in handling and learning from the data.

Model Evaluation

Following the training phase, the evaluation of the three ML models—LR, KNN, and CNN—was conducted on the test set containing 400 cases. Various performance metrics were calculated for each model to assess their effectiveness in making predictions. These metrics included accuracy, sensitivity,

XSL•FO
RenderX

positive predictive value (PPV), F_1-score, negative predictive value (NPV), and specificity of each model.

After calculating these metrics, the ML models with the highest sensitivity and specificity were chosen for further consideration due to their critical role in the diagnostic accuracy required by the specific use case. By prioritizing both sensitivity and specificity, you ensured a balanced assessment of the models' performance regarding true positive and true negative rates.

In addition to the confusion matrices, which provide a detailed overview of true positives, false negatives, false positives, and true negatives at the default prediction threshold, we used the receiver operating characteristic curve and the corresponding area under the curve for each model. This comprehensive analysis of the receiver operating characteristic curves and area under the curve values allowed for a deeper understanding of the trade-offs between true positive and false positive rates across different thresholds and enabled the selection of the best-performing models based on a more nuanced evaluation beyond conventional accuracy metrics.

Ethical Considerations

The study protocol was conducted in accordance with the Declaration of Helsinki and approved by the institutional review board at Shanghai Ethics Committee for Clinical Research (Approval number: SECCR/2023-119-01). The data used in this study were deidentified. Informed consent was waived by the Ethics Committee owing to the use of deidentified data.

Results

Baseline Characteristics

The training set represented 1600 patients, of whom 48.5% (776) were female, the mean age was 62.86 (SD 9.16) years, with a total of 33654 course records. The test set represented 400 patients, of whom 49.3% (n=197) were female; the mean age was 62.81 (SD 9.21) years, with a total of 8491 course records. The ratio of postoperative major bleeding was 4.31% (n=69) in the training set and 4.75% (n=19) in the test set (Table 1).

Table . Population characteristics of training and test sets.

Characteristics	Training set (n=1600)	Test set (n=400)
Gender (female) n (%)	776 48.5	197 49.3
Age (years), mean (SD)	62.86 (9.16)	62.81 (9.21)
Course records, n	33,654	8491
Proportion of major bleeding, n (%)	69 (4.31)	19 (4.75)

Note-Based Feature Selection

A total of 270 features were selected; first, 261 features related to postoperative bleeding were selected by regular manual disassembly. Second, Jieba was used to segment the disease course text and select 8 features according to word frequency classification. Furthermore, the last feature was created by a logical recognition graph generated from expert opinion. More details are provided in the feature engineering section. The frequency of the top 20 features is shown in Table 2.

XSL•FO
RenderX

Table . The frequency of the top 20 features selection.

Features (description or regularization in English)	Frequency
Postoperative	1592
(chest | abdomen | pelvis) cavity | rectum. {0,2} depression | posterior fornix | anastomosis | pancreas | digestive tract | stomach | vagina	1579
(out | lose) blood (stop | break into)?| hematoma | subcutaneous Ecchymosis | congestion	1033
Bleeding	1002
Bloody fluid | red.{0,2} drainage fluid	995
Hemostasis	528
Laparotomy	513
Dark red	347
Emergency	343
(To | Perform).{0,15} to stop bleeding	303
Swelling	273
(To | Perform).{0,5} to stop bleeding	253
(((chest | abdomen | pelvis) cavity | rectum.{0,2} depression | posterior fornix | anastomosis | pancreas | digestive tract | stomach).{0,10}bleeding | vagina.{0,2}large .{0,2}(out | lose)blood)	251
Blood loss	246
Pelvic cavity.{0,5}Drainage.{0,5}(bloody liquid | fluid dark red)	225
(?: inject | give).{0,20}(?:suspended less white (?:red blood cell | erythrocyte blood transfusion volume | red blood cell)|whole blood | red blood cell | less white suspended red blood cell | red suspension | suspended less white Red).{1,20}(?:U | ML | CC | ml | u | ml | mL | cc | infusion | unit) | (?: in | out | to).{0,5}(? :U |ML |CC | milliliter | u | ml | mL | cc | infusion | unit).{0,5}(?: suspended less white(?:cell red blood cell | red blood cell transfusion amount | red blood cell) | whole blood | Red blood cells | less white suspension red blood cells | red suspension | suspension less white red)	220
((out | loss) blood | introduction | bloody Effusion).{0,15}(U | ML | CC | ml | u | ml | mL | cc | unit)	220
Lose.{0,20}(red blood cells | whole blood | red suspension)	220
Structured recognition results	206
(exist | has).{0,5} bleed	178

Comparison and Verification of the Efficiency of Three Machine Learning Models

In the identification of major bleeding events within the test set, the LR method had an accuracy of 0.8275, sensitivity of 0.8947, specificity of 0.8241, PPV of 0.2024, NPV of 0.9937, and F_1-score of 0.3301. The accuracy of the CNN method in the testing set was 0.8900, the sensitivity was 0.8421, the specificity was 0.8924, the PPV was 0.2807, the NPV was 0.9913, and the F_1-score was 0.4211. The KNN method had an accuracy of 0.9575, sensitivity of 0.2105, specificity of 0.9948, PPV of 0.6667, NPV of 0.9619, and F_1-score of 0.3200 (Table 3). The C-statistic was higher in the LR method (C=0.9018), followed by the CNN method (C=0.8830) (Figure 2).

XSL•FO
RenderX

Table . The performance of each model in identifying major bleeding events.

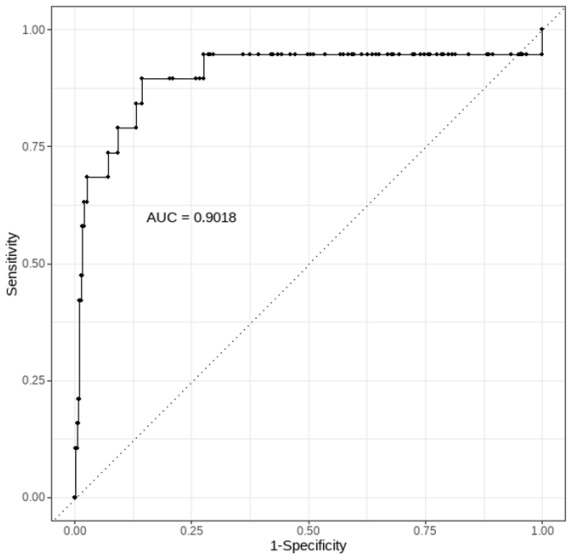
Sets and models	Accuracy	Sensitivity	Positive predictive value	F_1-score	Negative predictive value	Specificity
Training set (n=1600)						
LR[a]	0.8231	1.0000	0.1960	0.3278	1.000	0.8152
CNN[b]	0.9056	0.9710	0.3102	0.4702	0.9986	0.9027
KNN[c]	0.9643	0.1884	0.9286	0.3132	0.9647	0.9993
Test set (n=400)						
LR	0.8275	0.8947	0.2024	0.3301	0.9937	0.8241
CNN	0.8900	0.8421	0.2807	0.4211	0.9913	0.8924
KNN	0.9575	0.2105	0.6667	0.3200	0.9619	0.9948

[a]LR: logistic regression.

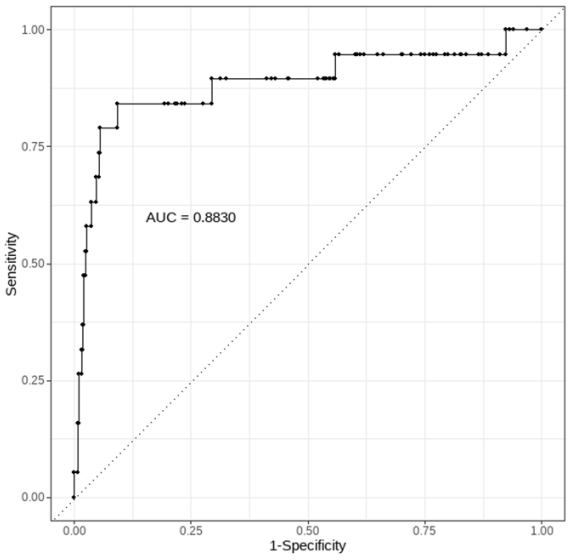
[b]CNN: convolutional neural network.

[c]KNN: K-nearest neighbor.

Figure 2. Receiver operating characteristic curves for identifying major bleeding from clinical notes using the LR and CNN methods. AUC: area under the curve; LR: logistic regression; CNN: convolution neural network.



LR



CNN

Discussion

Principal Findings

In this study, we implemented LR, CNN, and KNN algorithms for ML detection of bleeding events within electronic medical record systems. Notably, CNN was specifically applied to categorize radiology free-text reports and exhibited commendable accuracy [17]. The CNN approach demonstrated a sensitivity of 84% and a specificity of 89% in detecting recorded bleeding events in patients at a threshold of 0.50, while the LR method showcased a sensitivity of 89% and a specificity of 82% at the same threshold. The KNN algorithm can be negatively impacted by high dimensionality in feature matrices. These findings underscore the viability of employing algorithms to pinpoint patients with bleeding events within extensive textual disease records.

Evaluation Metrics and Application Context

F_1-scores are frequently used in the evaluation of classifiers. The F_1-score can be considered the harmonic mean between the precision and recall (sensitivity) [18]. Thus, it symmetrically represents both precision (how accurately the model identifies only true positive instances) and sensitivity (how well the classifier identifies all actual positive instances) in one metric. Although the three machine learning models in this investigation did not attain high F_1-scores on the test set, it must be clarified that the goal of the investigation was to detect bleeding events as accurately as possible in practical applications for subsequent manual assessment. Therefore, greater attention is paid to the sensitivity and specificity (ie, how well the classifier identifies all actual negative instances) of the model, the high values of which imply that the model can detect hemorrhagic events while avoiding false alarms, which is crucial for this investigation. Consequently, priority is placed on the models' sensitivity and the attainment of greater specificity. Hence, we maintain that

XSL•FO
RenderX

the models developed in this study remain valuable for identifying bleeding events within disease records and represent a promising avenue for future research endeavors.

Comparison to Prior Work

The integration of NLP and ML has proven successful across various domains. For instance, in the context of health care, predictive models leveraging ML and statistical methods have demonstrated the ability to forecast occurrences such as postpartum hemorrhage upon labor admission with reasonable discriminatory power. A prior study demonstrated the effectiveness of a Hybrid CNN-LSTM Autoencoder model for the detection of bleeding events within EHR data. This was accomplished through the integration of a supervised CNN with a pretrained, unsupervised Bidirectional Long Short-Term Memory autoencoder. The primary objective was to accurately predict the presence of a bleeding event within a given English sentence from an EHR record [19].

While prior research primarily focused on English EHR data and general bleeding event detection, such as the study by Li et al that used a Hybrid CNN-LSTM Autoencoder for sentence-level bleeding detection, and a more recent work that applied retrieval augmented generation with large language models for detecting nonsurgical major bleeding events in English EHRs [20], our investigation centers on the identification of major bleeding events within Chinese EHR. Chinese NLP presents inherent complexities, especially in word segmentation within EHR text, posing a significant hurdle. Furthermore, the shift in focus to major bleeding, which exhibits a considerably lower incidence rate compared to general bleeding events, substantially reduces the availability of positive samples, thereby intensifying the challenge for robust predictive modeling. Building upon these complexities, unlike the prior studies' sentence-level analysis, our research takes a holistic approach to identifying major bleeding events within the entire patient visit context. This broader, patient-centric perspective further complicates the analysis and requires a more comprehensive understanding of the clinical narrative. These critical distinctions—language specificity, event granularity, and analytical scope—collectively highlight the greater challenges of our study compared to existing work in the field.

Strengths and Limitations

A key strength of this study lies in the strategic approach to model selection and feature engineering. Initially, SVM and RF algorithms were explored. However, recognizing their reliance on effective word segmentation and the challenges posed by the heterogeneous nature of the text data, the focus was deliberately shifted to LR, KNN, and CNN. To further enhance model performance, term frequency-inverse document frequency (TF-IDF) scoring was effectively used to compute feature representations, addressing the complexities of free-text data. Moreover, to overcome the limitation of a relatively small sample size, particularly the scarcity of positive samples (fewer than 100), and the resulting dispersed TF-IDF features, a targeted feature engineering approach was strategically adopted. This involved selecting predefined features associated with bleeding based on expert medical domain knowledge. This approach allowed the ML models to concentrate on clinically relevant features, represented as binary values (0 and 1), rather than solely relying on traditional TF-IDF metrics applied to the entire text. This targeted feature engineering proved crucial for enhancing performance in a limited data setting. As a result of these strengths, including strategic model selection, targeted feature engineering, manual annotation, and training with limited samples, the LR and CNN models demonstrated superior performance compared to SVM and RF models based on TF-IDF features for detecting major bleeding events. This aligns with findings from similar research, as evidenced by a study that underscored the enhanced performance of CNNs when expert annotation of text data is incorporated [19], further validating the efficacy of CNNs in text classification tasks within the medical domain.

However, manual feature selection has certain limitations. First, engineers must dedicate time to writing code for feature extraction based on the definitions provided by the medical field, which is a more time-intensive process compared to the automatic feature computation performed by machines. To partially mitigate this, we prioritized high-impact features identified through clinical expert consultations, but future studies could integrate semiautomated pipelines to balance efficiency and domain specificity. Second, as previously discussed in related research [21], the identification of postoperative bleeding and feature selection depends on manual procedures, thereby increasing the likelihood of bias and oversight. While dual annotations by multiple clinicians were used to reduce subjectivity, discrepancies were resolved through consensus rather than quantitative metrics, potentially affecting reproducibility. Future work should adopt standardized annotation protocols with inter-rater reliability assessments. Third, the chosen features encapsulate domain-specific knowledge, sometimes being closely tied to specific hospital departments with varying requirements and documentation practices in electronic medical records. Broader applicability requires multicenter collaboration to harmonize feature definitions across institutions. Moreover, our study included a retrospective analysis with a relatively limited number of specimens, a factor to be considered for broader applicability in ML analyses. Similar challenges have been observed in other comparable studies [19]. Consequently, not only does this constrained universality impede application in diverse fields, but also when extending to other illnesses, hospitals, and departments, there is often a need to reimagine and recreate the features.

Conclusions

Based on our new text feature selection method, both the LR and CNN methods perform well in identifying major bleeding occurring in postoperative patients with malignant tumors from electronic medical records.

XSL•FO
RenderX

Acknowledgments

Medical writing support, under the direction of the authors, was provided by Hao Hu (Shanghai Palan DataRx Co.,Ltd) and funded by Sanofi. This study was sponsored by Sanofi.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to patient privacy concerns but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: H Li, HY, YG, H Luo, CC, ZZ, MY
Data curation: WJ
Formal analysis: H Luo, HY, YG
Methodology: H Li, HY, YG, H Luo, CC, ZZ, MY
Supervision: H Li
Writing – original draft: HY, YG, H Luo, WJ
Writing – review & editing: H Li, CC, ZZ, MY

Conflicts of Interest

The authors individually and collectively are responsible for all content and editorial decisions and received no payment from Sanofi related to the development/presentation of this publication. CC, ZZ are Sanofi employees and may hold shares and/or stock options in the company. MY was previously a Sanofi employee and is now employed at Beijing Baheal Chengchuang Pharmaceutical Investment Co., Ltd. The other authors declare that there is no conflict of interest.

References

1. Tasu JP, Vesselle G, Herpe G, et al. Postoperative abdominal bleeding. Diagn Interv Imaging 2015;96(7-8):823-831. [doi: 10.1016/j.diii.2015.03.013] [Medline: 26078019]
2. Maxwell MJ, Wilson MJA. Complications of blood transfusion. Continuing Education in Anaesthesia Critical Care & Pain 2006 Dec;6(6):225-229. [doi: 10.1093/bjaceaccp/mkl053]
3. Berger JS, Bhatt DL, Steg PG, et al. Bleeding, mortality, and antiplatelet therapy: results from the Clopidogrel for High Atherothrombotic Risk and Ischemic Stabilization, Management, and Avoidance (CHARISMA) trial. Am Heart J 2011 Jul;162(1):98-105. [doi: 10.1016/j.ahj.2011.04.015]
4. Cook DJ, Griffith LE, Walter SD, et al. The attributable mortality and length of intensive care unit stay of clinically important gastrointestinal bleeding in critically ill patients. Crit Care 2001 Dec;5(6):368-375. [doi: 10.1186/cc1071] [Medline: 11737927]
5. Khorana AA, Mackman N, Falanga A, et al. Cancer-associated venous thromboembolism. Nat Rev Dis Primers 2022 Feb 17;8(1):11. [doi: 10.1038/s41572-022-00336-y] [Medline: 35177631]
6. Anderson DR, Morgano GP, Bennett C, et al. American Society of Hematology 2019 guidelines for management of venous thromboembolism: prevention of venous thromboembolism in surgical hospitalized patients. Blood Adv 2019 Dec 10;3(23):3898-3944. [doi: 10.1182/bloodadvances.2019000975]
7. Taggart M, Chapman WW, Steinberg BA, et al. Comparison of 2 natural language processing methods for identification of bleeding among critically ill patients. JAMA Netw Open 2018 Oct 5;1(6):e183451. [doi: 10.1001/jamanetworkopen.2018.3451] [Medline: 30646240]
8. Corey KM, Kashyap S, Lorenzi E, et al. Development and validation of machine learning models to identify high-risk surgical patients using automatically curated electronic health record data (Pythia): a retrospective, single-site study. PLoS Med 2018;15(11):e1002701. [doi: 10.1371/journal.pmed.1002701]
9. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018;1:18. [doi: 10.1038/s41746-018-0029-1] [Medline: 31304302]
10. Weller GB, Lovely J, Larson DW, Earnshaw BA, Huebner M. Leveraging electronic health records for predictive modeling of post-surgical complications. Stat Methods Med Res 2018 Nov;27(11):3271-3285. [doi: 10.1177/0962280217696115] [Medline: 29298612]
11. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2015;13:8-17. [doi: 10.1016/j.csbj.2014.11.005]
12. Murdoch TB, Detsky AS. The inevitable application of big data to health care. JAMA 2013 Apr 3;309(13):1351-1352. [doi: 10.1001/jama.2013.393] [Medline: 23549579]
13. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature New Biol 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]

XSL•FO
RenderX

14. Delate T, Jones AE, Clark NP, Witt DM. Assessment of the coding accuracy of warfarin-related bleeding events. Thromb Res 2017 Nov;159:86-90. [doi: 10.1016/j.thromres.2017.10.004] [Medline: 29035718]
15. Wilchesky M, Tamblyn RM, Huang A. Validation of diagnostic codes within medical services claims. J Clin Epidemiol 2004 Feb;57(2):131-141. [doi: 10.1016/S0895-4356(03)00246-4] [Medline: 15125622]
16. Schulman S, Angerås U, Bergqvist D, et al. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in surgical patients. J Thromb Haemost 2010 Jan;8(1):202-204. [doi: 10.1111/j.1538-7836.2009.03678.x] [Medline: 19878532]
17. Chen MC, Ball RL, Yang L, et al. Deep learning to classify radiology free-text reports. Radiology 2018 Mar;286(3):845-852. [doi: 10.1148/radiol.2017171115] [Medline: 29135365]
18. Rai T, Morisi A, Bacci B, et al. Keeping pathologists in the loop and an adaptive F1-score threshold method for mitosis detection in canine perivascular wall tumours. Cancers (Basel) 2024 Feb 2;16(3):644. [doi: 10.3390/cancers16030644] [Medline: 38339394]
19. Li R, Hu B, Liu F, et al. Detection of bleeding events in electronic health record notes using convolutional neural network models enhanced with recurrent neural network autoencoders: deep learning approach. JMIR Med Inform ;7(1):e10788. [doi: 10.2196/10788]
20. Kaplinsky P, Singh R, Fusillo TF, Leader A, Zwicker JI, Mantha S. Retrieval augmented generation for the detection of major bleeding events in the electronic health record. Blood 2024 Nov 5;144(Supplement 1):2263-2263. [doi: 10.1182/blood-2024-203911]
21. Jing X, Wang X, Zhuang H, Fang X, Xu H. Multiple machine learning approaches based on postoperative prediction of pulmonary complications in patients with emergency cerebral hemorrhage surgery. Front Surg 2021;8:797872. [doi: 10.3389/fsurg.2021.797872] [Medline: 35127804]

Abbreviation

CNN: Convolutional neural network
KNN: K-nearest neighbor
LR: Logistic regression
ML: Machine learning
NLP: natural language processing
NPV: negative predictive value
PPV: positive predictive value
RF: random forest
SVM: support vector machine
VTE: venous thrombosis

Edited by A Mavragani; submitted 05.09.24; peer-reviewed by NH Le, Y Zhang; revised version received 25.03.25; accepted 07.04.25; published 01.05.25.

Please cite as:
Li H, Yao H, Gao Y, Luo H, Cai C, Zhou Z, Yuan M, Jiang W
Identification of Major Bleeding Events in Postoperative Patients With Malignant Tumors in Chinese Electronic Medical Records: Algorithm Development and Validation
JMIR Form Res 2025;9:e66189
URL: https://formative.jmir.org/2025/1/e66189
doi:10.2196/66189

© Hui Li, Haiyang Yao, Yuxiang Gao, Hang Luo, Changbin Cai, Zhou Zhou, Muhan Yuan, Wei Jiang. Originally published in JMIR Formative Research (https://formative.jmir.org), 1.5.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile Health App for Adolescent Asthma Self-Management: Development and Usability Study of the Pulmonary Education and Knowledge Mobile Asthma Action Plan

Xing He[1,2], PhD; Jiang Bian[1,2], PhD; Ariel Berlinski[3,4], MD; Yi Guo[5], PhD; A Larry Simmons[3,6], MD; S Alexandra Marshall[7], MPH, PhD; Carolyn J Greene[8], PhD; Rita Hudson Brown[3,4], BA, ACRP; Jessica Turner[4], MPH; Tamara T Perry[3,4], MD

[1]Department of Biostatistics and Health Data Science, Indiana University, Indianapolis, IN, United States

[2]Regenstrief Institute, Indianapolis, IN, United States

[3]Department of Pediatrics, University of Arkansas for Medical Sciences, Little Rock, AR, United States

[4]Arkansas Children's Research Institute, 13 Children's Way, Little Rock, AR, United States

[5]Department of Health Outcomes & Biomedical Informatics, University of Florida, Gainesville, FL, United States

[6]Arkansas Children's Hospital, Little Rock, AR, United States

[7]Department of Health Behavior & Health Education, University of Arkansas for Medical Sciences, Little Rock, AR, United States

[8]United States Department of Veterans Affairs, Mann-Grandstaff VA Medical Center, Spokane, WA, United States

Corresponding Author:
Tamara T Perry, MD
Department of Pediatrics, University of Arkansas for Medical Sciences, Little Rock, AR, United States

Abstract

Background: Adolescents with asthma are vulnerable to poor asthma outcomes due to inadequate self-management skills and nonadherence to medications. Mobile health (mHealth) apps have shown promise in improving asthma control, medication adherence, and self-efficacy. However, existing mHealth asthma apps lack personalization and real-time feedback and are not tailored for at-risk adolescents.

Objective: This study aimed to design, develop, and test a smartphone-based mHealth Asthma Action Plan for adolescents, called Pulmonary Education and Knowledge Mobile Asthma Action Plan (PEAK-mAAP), in preparation for a large-scale randomized controlled trial.

Methods: We employed user-centered design principles to develop our app, leveraging our previous work and following guidelines from the National Heart, Lung, and Blood Institute. The app consists of a patient-facing mobile app and a provider-facing portal. A convenience sample of 13 adolescents (aged 12 - 20 years) was recruited from the Arkansas Children's Research Institute database or direct health care provider referrals. Participants underwent a task-based usability assessment followed by the System Usability Scale assessment to measure user satisfaction, interface effectiveness, and overall system usability.

Results: PEAK-mAAP integrates 7 core modules supporting personalized asthma self-management, symptom monitoring, medication tracking, and real-time feedback. The mean System Usability Scale score was 83/100 (SD 5.54), indicating high user satisfaction and system usability. Notably, older adolescents (>17 years) reported higher usability scores (87.5) than younger users (77.5), suggesting potential age-related differences in app navigation and engagement.

Conclusions: The results demonstrate that PEAK-mAAP is a feasible and user-friendly mHealth intervention for adolescent asthma self-management. While the high usability score reflects a positive user experience, some participants encountered initial usability challenges, highlighting the need for minor refinements and user training materials. The integration of personalized self-management tools and real-time feedback distinguishes PEAK-mAAP from existing asthma apps, addressing key barriers to adherence and engagement. Moving forward, an ongoing randomized controlled trial will assess its clinical effectiveness, long-term engagement, and impact on asthma outcomes, providing further insights into its potential as a scalable solution for adolescent asthma care.

(JMIR Form Res 2025;9:e64212) doi:10.2196/64212

KEYWORDS

asthma self-management; mobile health; user-centered design; usability; adolescents; mobile phone

XSL•FO
RenderX

Introduction

According to the Global Asthma Report 2022 [1], asthma impacts 262 million individuals globally across all ages. The report highlighted that 1 in 10 children have asthma symptoms, and half have poorly controlled symptoms. Like most chronic diseases, a proactive way to mitigate the cost and adverse effects caused by asthma is to equip patients with effective self-management strategies [2]. Earlier investigations of mobile health (mHealth) asthma management apps show a positive effect on asthma control, medication adherence, and self-efficacy [3,4]. mHealth technology that combines personalized health management into patients' daily routines can effectively enhance patients' asthma self-management [4]. For instance, Farzandipour et al [5] designed a smartphone-based asthma self-management app with educational content, data recording, reminders, and physician communication features, which was rated as "good" in usability and positively impacted patients' self-management knowledge. Kosse et al [6] found that an mHealth intervention significantly improved medication adherence in adolescents with asthma who had poor adherence rates, suggesting that health care providers should consider a tailored mHealth approach in their treatment. Fedele et al [7] conducted a randomized controlled trial (RCT) to test the feasibility and efficacy of a mHealth intervention to enhance caregiver support and asthma self-management in early adolescents, revealing clinically significant improvements in asthma control. Nevertheless, existing mHealth asthma apps lack personalization and real-time feedback, and are not tailored for at-risk adolescents. Furthermore, recent review highlights a significant lack of large-scale RCTs evaluating the efficacy and sustainability of eHealth and mHealth interventions in adolescents and young adults with chronic diseases, including asthma [8].

Adolescents are at high risk of adverse asthma outcomes due to multiple factors, such as (1) their self-management skills are often insufficient, which may lead to nonadherence to medication; and (2) adolescence is typically when there is a transition from dependence on parental supervision to medical autonomy [9]. Current practices should ensure adolescents have the correct knowledge, beliefs, and positive attitudes to self-manage their illness [10]. National Heart, Lung, and Blood Institute (NHLBI) recommends using Asthma Action Plans (AAPs) to facilitate asthma self-management [11]. In our previous studies [12,13], we developed a smartphone-based mobile health Asthma Action Plan (mAAP) tailored for adolescents. Our findings demonstrated its feasibility as a self-management tool, significantly benefiting adolescents with uncontrolled asthma. Leveraging the insights from these feasibility studies, we adopted a user-centered design (UCD) [14] approach, collaborating with a diverse group of stakeholders (ie, adolescents with asthma, their parents, clinicians, nurses, and researchers) to iteratively redesign and re-engineer the mAAP and named it, Pulmonary Education and Knowledge Mobile Asthma Action Plan (PEAK-mAAP). Critically, we designed PEAK-mAAP to address specific barriers to adopting health behaviors based on the well-established health belief model (HBM) [15]. Built on the Apache Cordova (Adobe Systems) [16] mobile app framework, PEAK-mAAP has cross-platform compatibility to meet the needs of Android (Google), iOS (Apple Inc), and web users. Complementing PEAK-mAAP, we introduced a web portal for health providers and guardians, allowing them to oversee the adolescents' engagement with PEAK-mAAP. In this paper, we detail the design, development, and usability assessment of PEAK-mAAP.

Methods

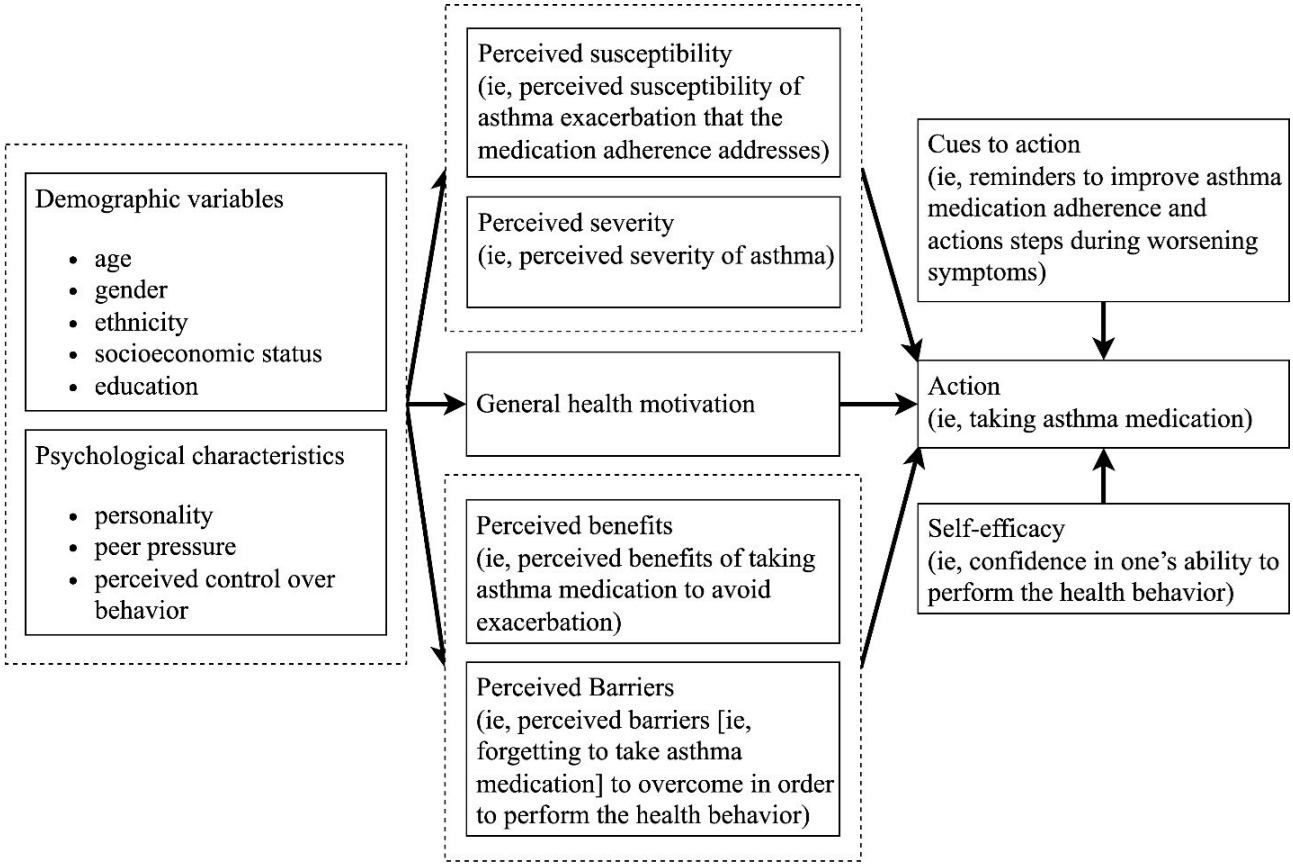
Design of PEAK-mAAP

We designed the functionalities of PEAK-mAAP based on the HBM to affect behavior change (ie, asthma self-management based on prescribed AAP). Figure 1 depicts how the design of PEAK-mAAP addresses the different constructs of the HBM. Notably, the HBM integrates considerations of individual demographics and psychological traits—termed modifying variables. These variables indirectly influence health-related actions by impacting factors like health motivation, perceived risks, severity, benefits, and obstacles.

After the preliminary design of PEAK-mAAP, we engaged in iterative enhancements by involving all stakeholders through a UCD approach. We convened a community advisory board with our intended end users, that is, adolescent (ie, teens and emerging adults aged between 12 and 20) patients with asthma, their caregivers, asthma care providers, UCD specialists, and researchers. community advisory board members provided feedback on the features of mAAP and iterative feedback and beta testing of the final app.

XSL•FO
RenderX

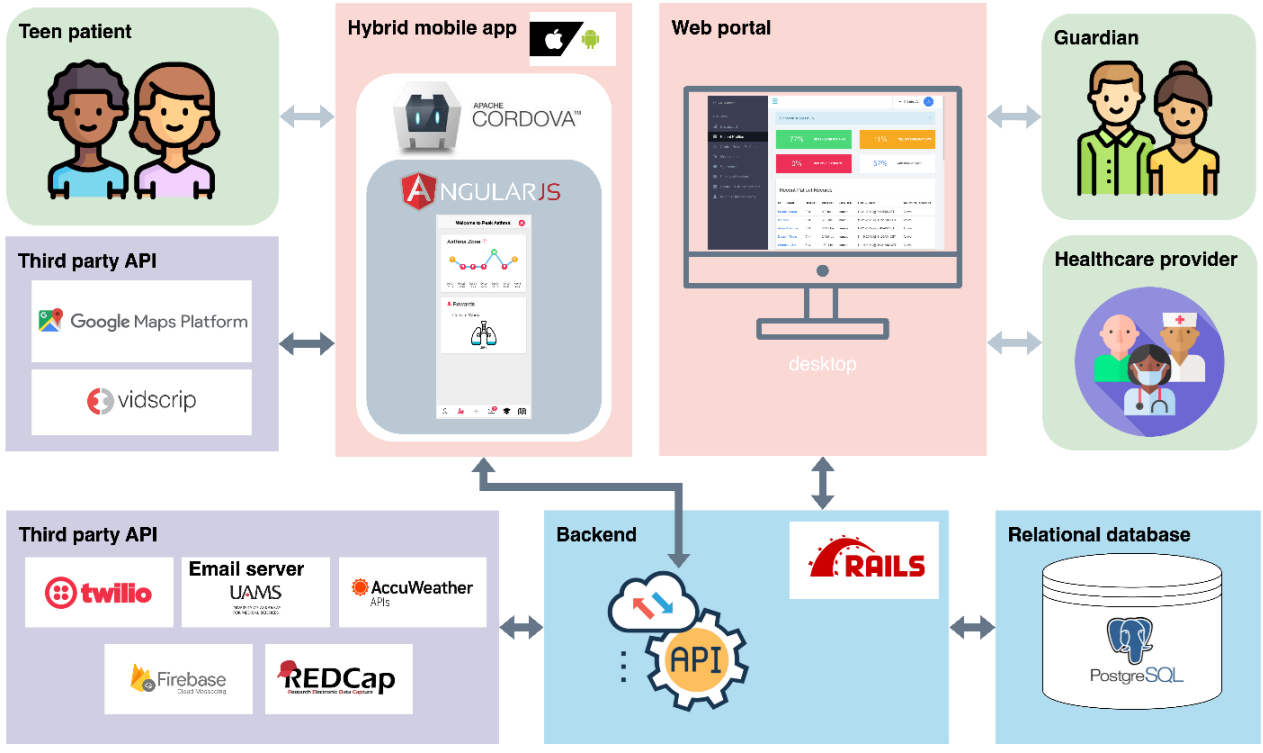
Figure 1. Design of Pulmonary Education and Knowledge Mobile Asthma Action Plan according to the health belief model.



Implementation of PEAK-mAAP

PEAK-mAAP is a hybrid mobile app developed using standard web technologies and housed within a native container—specifically Apache Cordova—for adolescents. The web portal, designed for health care providers and guardians of adolescents, is built with Ruby on Rails [17]. Figure 2 illustrates the 5-component system architecture that supports the functionality of the app and the web portal.

Figure 2. The 5-component system architecture which supports Pulmonary Education and Knowledge Mobile Asthma Action Plan and the web portal. API: application programming interface; REDCap: Research Electronic Data Capture.



XSL•FO
RenderX

First, a relational database, PostgreSQL [18], an advanced open-source object-relational database system, is employed to manage and categorize a range of data, including patient demographic details, guardian and provider profiles, medication records, and specific activity logs such as peak flow recordings. In addition, the database houses preset content, like educational tips and instructional videos. To safeguard against any potential data loss, regular backups are performed.

Second, our backend, built on the robust Ruby on Rails framework, offers RESTful application programming interfaces (APIs) to bridge our hybrid mobile app with its backend database. It also integrates with 5 external APIs, like weather, to enrich the mobile app's features. Besides, adhering to web development best practices, this backend employs the model-view-controller design pattern to support the web portal. The web portal is tailored for adolescent health care providers such as doctors, nurses, and researchers, as well as guardians like parents and grandparents, enabling them to monitor PEAK-mAAP usage effectively.

Third, we implemented PEAK-mAAP as a hybrid mobile app to aid adolescents in managing their asthma. This approach involved the development of a cross-platform web app, encapsulated within a native app container. The hybrid mobile app was crafted using Apache Cordova and AngularJS (Google) [19] (ie, an open-source JavaScript framework primarily used for building dynamic single-page web apps). This allows programmers to use standard web technologies such as HTML5, CSS3, and JavaScript for cross-platform development.

Fourth, we engineered a model-view-controller–based web app to facilitate monitoring adolescents' mobile app usage by both health care providers and guardians. Health care providers can access a comprehensive view of all patients' information, mobile app usage, and asthma-related activities through a specialized web portal. Concurrently, guardians are provided a tailored view allowing them to observe their children's information and interactions with the PEAK-mAAP, reinforcing transparency and personalized care.

Fifth, to enhance our mobile app for adolescents' asthma self-management, we integrated 7 third-party APIs. Directly, we incorporated the Google Maps [20] API for locating hospitals and pharmacies and Vidscrip (Clear.md Inc) [21] for educational asthma videos. Indirectly, through our Ruby on Rails backend, we added the Twilio [22] API for alert calls and texts during severe asthma exacerbation; an email server hosted by the University of Arkansas for Medical Sciences to notify health care providers and guardians; the AccuWeather [23] API for real-time weather and pollen levels; the Firebase Cloud Messaging [24] API for cross-platform notifications; and the REDCap (Research Electronic Data Capture; Vanderbilt University) [25,26] API for automated monthly survey distributions. These integrations ensure a holistic approach to asthma management for adolescents.

Usability Assessment

The PEAK-mAAP usability assessment was conducted over two months, spanning January and February 2019, before the formal enrollment of research participants in the subsequent RCT. Usability assessment is a widely applied method across various industries, including aviation, automotive, and software development, aimed at enhancing system functionality by assessing an app's effectiveness and efficiency. Effectiveness refers to the extent to which users can successfully accomplish tasks and achieve their goals, while efficiency evaluates the time and effort required to complete a given task.

Participants were selected using a convenience sampling method and a total of 13 individuals, ranging from 12 to 20 years old in age, were recruited. All participants had a confirmed clinical diagnosis of asthma, as documented in their medical records. Recruitment efforts targeted potential candidates through 2 primary channels: the Arkansas Children's Research Institute database, which provided access to a broad pool of pediatric patients with existing asthma diagnoses, and direct referrals from health care providers familiar with eligible patients. To qualify for the usability assessment, participants were required to meet specific eligibility criteria designed to ensure they could effectively engage with the PEAK-mAAP system. Inclusion criteria consisted of (1) a current diagnosis of asthma, verified by a health care professional; (2) previous exposure to or familiarity with AAPs, to facilitate comparative feedback on the mobile app; (3) proficiency in English to ensure clear communication during the assessment and accurate comprehension of the app's content; and (4) the absence of any visual or auditory impairments that could hinder the ability to interact with the app's interface. Individuals were excluded if they had significant respiratory conditions other than asthma, such as cystic fibrosis or chronic obstructive pulmonary disease, which could introduce variability in the management strategies assessed. In addition, participants with developmental delays or cognitive impairments that could interfere with their ability to comprehend instructions, navigate the app, or provide reliable feedback were not eligible for the study. The sample size of 13 was deemed sufficient, adhering to the standard usability principle that even 5 participants can identify the most severe usability issues and 80% of all usability issues [27].

To ensure consistency, all participants were provided with the same iPad (Apple Inc) device for testing and watched a prerecorded demonstration video of the PEAK-mAAP app. Following the video, participants were instructed to explore the app using the study iPad to familiarize themselves with its features. Subsequently, they were asked to complete eight tasks using the app: (1) logging Peak Flow, (2) logging Symptoms, (3) setting a reminder for logging their Peak Flow or symptoms, (4) reading daily tips and viewing educational videos, (5) reading notifications, (6) describing directions or locations of the hospital or pharmacy displayed on the map, (7) describing the weather for patient's location, and (8) reporting number of days they logged and rewards within the app. A stopwatch was used to time each user while performing each task or scenario. Any extra steps performed during the scenarios were recorded to determine if there were common mistakes or additional steps taken for any of the app's features. During this process, users were asked to adopt the think-aloud [28] process by verbalizing any trouble they had with the app, features they found exciting or beneficial, as well as features that were confusing or not useful. This approach provided evaluators with deeper insights

XSL•FO
RenderX

beyond those obtainable through standard cognitive walkthroughs. There were at least 2 evaluators in the room during each usability test session. While observing the user, evaluators tracked issues and comments, and then aggregated the information from all usability tests. Observed errors were categorized using Nielsen-Shneiderman heuristics [29] (Table 1), a set of principles that provide guidelines and best practices for user interface design based on established usability heuristics.

During the study, evaluators assigned a severity rating to each identified issue, categorizing them into (1) cosmetic, (2) minor, (3) major, and (4) catastrophic. A cosmetic issue refers to a superficial concern that does not affect users' ability to complete a task, such as formatting inconsistencies or color preferences; these issues are typically of lower priority for resolution. A minor issue represents a problem with an available workaround, such as restricting numerical input to whole numbers rather than allowing decimals. Major issues, however, can significantly impact data quality, user satisfaction, or system functionality. Examples include the inability to access a function by clicking an icon, necessitating the use of a menu bar instead, or encountering difficulty when attempting to exit a screen. These issues require prioritization. Finally, a catastrophic severity rating indicates a critical system failure that demands immediate remediation. For instance, a user's inability to locate or input essential information constitutes a catastrophic issue.

XSL•FO
RenderX

Table . Nielsen-Shneiderman heuristics.

Error	Heuristic	Description
1	Consistency	Consistency and standards: Users should not have to wonder whether different words, situations, or actions mean the same thing. Standards and conventions in product design should be followed (*eg, font and layout and position*).
2	Visibility	Visibility of system state: Users should be informed about what is going on with the system through appropriate feedback and display of information (eg, what is current state of the system and what changes are made after an action).
3	Match	Match between system and the world: The image of the system perceived by users should match the model the users have about the system (eg, user model matches system image, and objects on the system should match objects of the task).
4	Minimalist	Any extraneous information is a distraction and a slow-down (eg, less is more and simple is efficient).
5	Memory	Minimize memory load: Users should not be required to memorize a lot of information to carry out tasks. Memory load reduces users' capacity to carry out the main tasks (eg, recognition vs concrete examples such as DD/MM/YY, eg, 10/20/15).
6	Feedback	Informative feedback: Users should be given prompt and informative feedback about their actions (eg, response time, concrete and specific, and not abstract and general).
7	Flexibility	Flexibility and efficiency: Users always learn and users are always different. Give users the flexibility of creating customization and shortcuts to accelerate their performance (eg, shortcuts for experienced users, templates, and bookmarks).
8	Message	Good error messages: The messages should be informative enough such that users can understand the nature of errors, learn from errors, and recover from errors (eg, precise not vague or general and polite messages).
9	Error	Prevent errors: It is always better to design interfaces that prevent errors from happening in the first place (eg, avoid text wraps, avoid slips or mistakes, and interfaces that make errors impossible).
10	Closure	Clear closure: Every task has a beginning and an end. Users should be clearly notified about the completion of a task (eg, clear beginning, middle and end, and clear feedback that goal was achieved).
11	Undo	Reversible actions: Users should be allowed to recover from errors. Reversible actions also encourage exploratory learning (eg, reverse actions in a single action and prevent serious errors).
12	Language	Use users' language: The language should be always presented in a form understandable by the intended users (eg, use standard meaning, language for specialized group, and users' perspective).

XSL•FO
RenderX

Error	Heuristic	Description
13	Control	Users in control: Do not give users that impression that they are controlled by the systems (eg, users are initiators not responders to actions, and avoid surprising actions and tedious sequences).
14	Documentation	Help and documentation: Always provide help when needed.

User Satisfaction Assessment

Satisfaction refers to the user's contentment and attitude toward the system. After completion of the think-aloud process, users were asked to complete a System Usability Scale (SUS) [30] survey. The SUS is a short, 10-question survey tool with high reliability and validity that measures user satisfaction. The survey employs a Likert scale ranging from 1 (strongly disagree) to 4 (strongly agree) to generate a single composite score representing the app's overall usability. Participants recorded their responses on paper, which were later digitized by the study team. At the end of the assessment, users were encouraged to share any concerns or recommendations.

Ethical Considerations

This study was approved by the University of Arkansas for Medical Sciences Institutional Review Board (protocol #206110). Informed consent (Multimedia Appendix 1) was obtained from all participants after explaining the study's purpose and procedures. Participant data were deidentified to ensure privacy and confidentiality. Each participant received US $50 as compensation for their time and involvement.

Results

Design of PEAK-mAAP

Using the HBM to explain current behaviors within our population, we hypothesize that the "cues to action" provided by PEAK-mAAP will be an essential factor in triggering behavior change. The app provides asthma educational video modules to address the "perceived susceptibility and severity" of asthma. The personalized mobile app features also provide additional "cues to action" during worsening symptoms to address "perceived benefits and barriers" (ie, give a clear action plan for the users to manage current asthma symptoms). Based on the education provided combined with self-management assistance from PEAK-mAAP, we hypothesize that self-efficacy and adherence will improve and result in improved asthma control for PEAK-mAAP users. To prove our hypotheses, we are conducting an RCT. The results of the RCT will be published upon completion of the trial.

Final Implementation of PEAK-mAAP

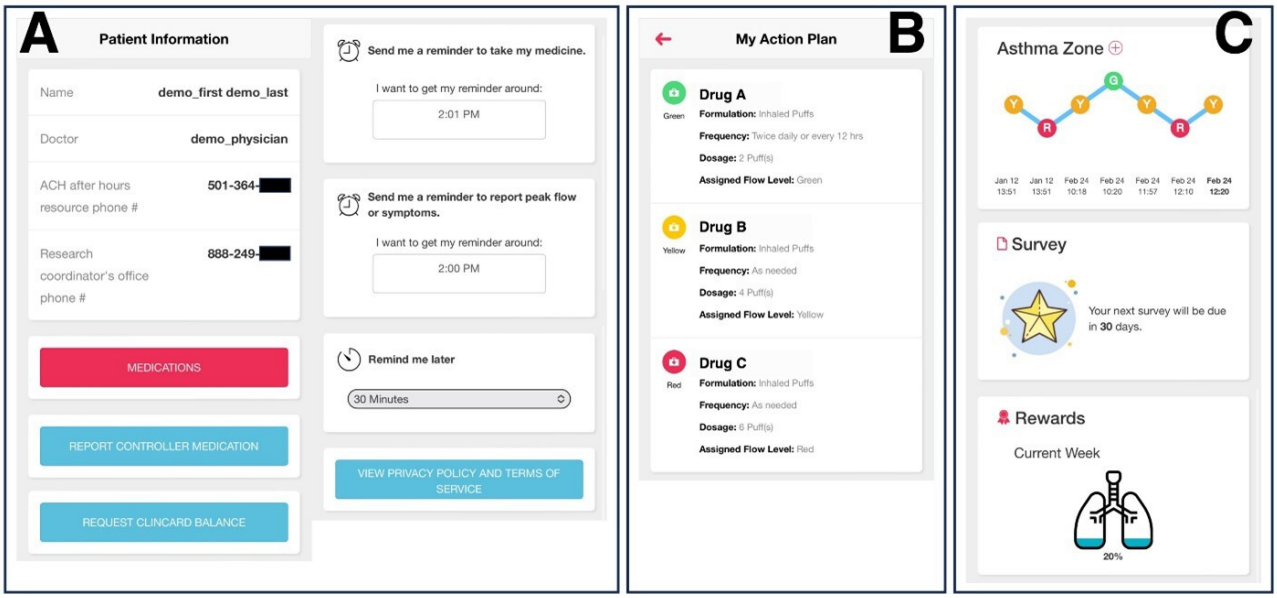
PEAK-mAAP is structured around 7 key modules.

First, as illustrated in Figure 3A, the patient information module encompasses a table presenting patient and health care provider details; a red "MEDICATIONS" button linking to the patient's asthma medication plan; a blue "REPORT CONTROLLER MEDICATION" button for logging medications; a "REQUEST CLINCARD BALANCE" button for users to request their ClinCard's balance, a reloadable debit card for research participant incentive; customizable reminder settings; and a "VIEW PRIVACY POLICY AND TERMS OF SERVICE" button to access the PEAK-mAAP's privacy terms. Figure 3B highlights a patient's personalized AAP, guiding actions including controller medication and rescue medication use based on their asthma zone, whether green, yellow, or red.

Second, a dashboard of patient activities, as seen in Figure 3C; the "Asthma Zone" allows visualization of the patient's recent asthma zones through a line chart, where each zone is depicted by uniquely colored circles. For color-blind users' ease, each circle also carries a letter mark indicating the logged zone. This intuitive design lets patients swiftly gauge their recent asthma status. The "Survey" card indicates the days leading up to the RCT's monthly survey due date. The "Rewards" card tracks weekly progress toward participant logging goals. As part of the PEAK-mAAP RCT trial, users earn nominal monetary incentives for active app engagement each week. The first daily record of a Peak Flow value or symptom will begin to fill the lung icon, signifying task completion and eligibility for the day's incentive. Completion of tasks each day will increase filling the lungs until the week's goals (eg, logging daily or completing the RCT's monthly survey) are met and the lungs fill completely. Such tangible rewards can bolster users' adherence to the app.

XSL•FO
RenderX

Figure 3. (**A**) The patient information tab; (**B**) the patient medication page with the patient's personalized Asthma Action Plan; and (**C**) the asthma zone record line chart, the survey countdown module, and the reward progress tracker.

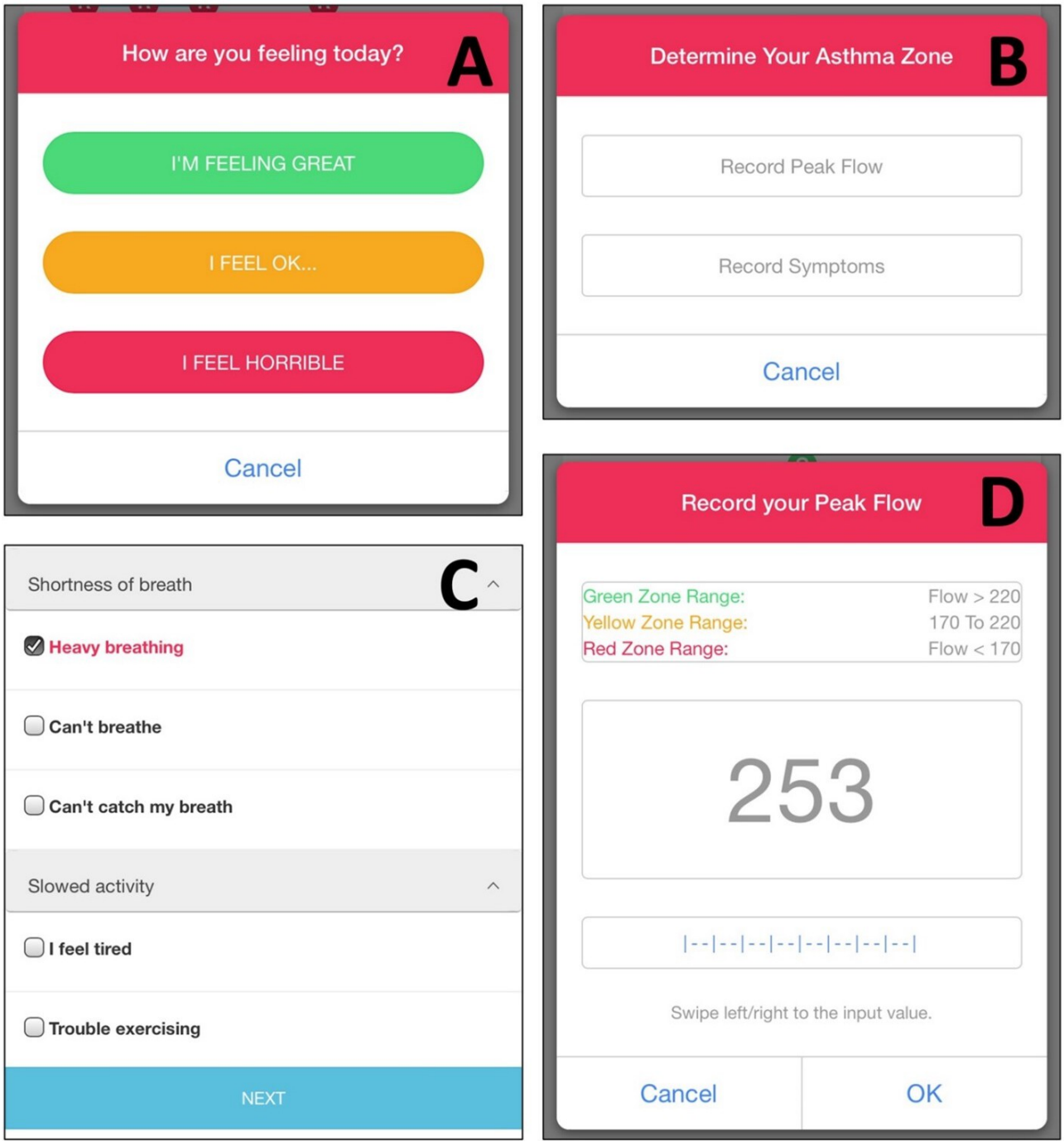


Third, a Peak Flow or asthma symptom record module: based on self-recorded Peak Flow value or asthma symptoms, PEAK-mAAP will classify the patient into their personalized asthma zones (ie, green, yellow, and red zones) and give zone-specific instructions (ie, action plan) such as taking emergency medications for acute symptoms. As shown in Figure 4A, before recording asthma-related symptoms or Peak Flow value, the system will first ask the participant "How are you feeling today?" The options are (1) "I'm feeling Great," (2) "I feel OK," and (3) "I feel horrible." Afterward, as shown in Figure 4B, the patient is asked to select how they prefer to determine their asthma zone, either Peak Flow or asthma symptoms. Figure 4C depicts the "Record your Peak Flow" screen, in which the patient can swipe on the blue bar to change the Peak Flow value. When the patient wants to report symptoms, a screen shown in Figure 4D will pop up, and the patient selects from a panel of 6 groups of symptoms to report.

Fourth, a notification module, as seen in Figure 5A, which shows a list of example notifications. The most recent messages are on the top of the list, and the unread and read notifications are separated by an "OLD MESSAGE" line separator. Figure 5B is an example of the in-app notification and Figure 5C is an example of the push notification to mobile phone native notification system.

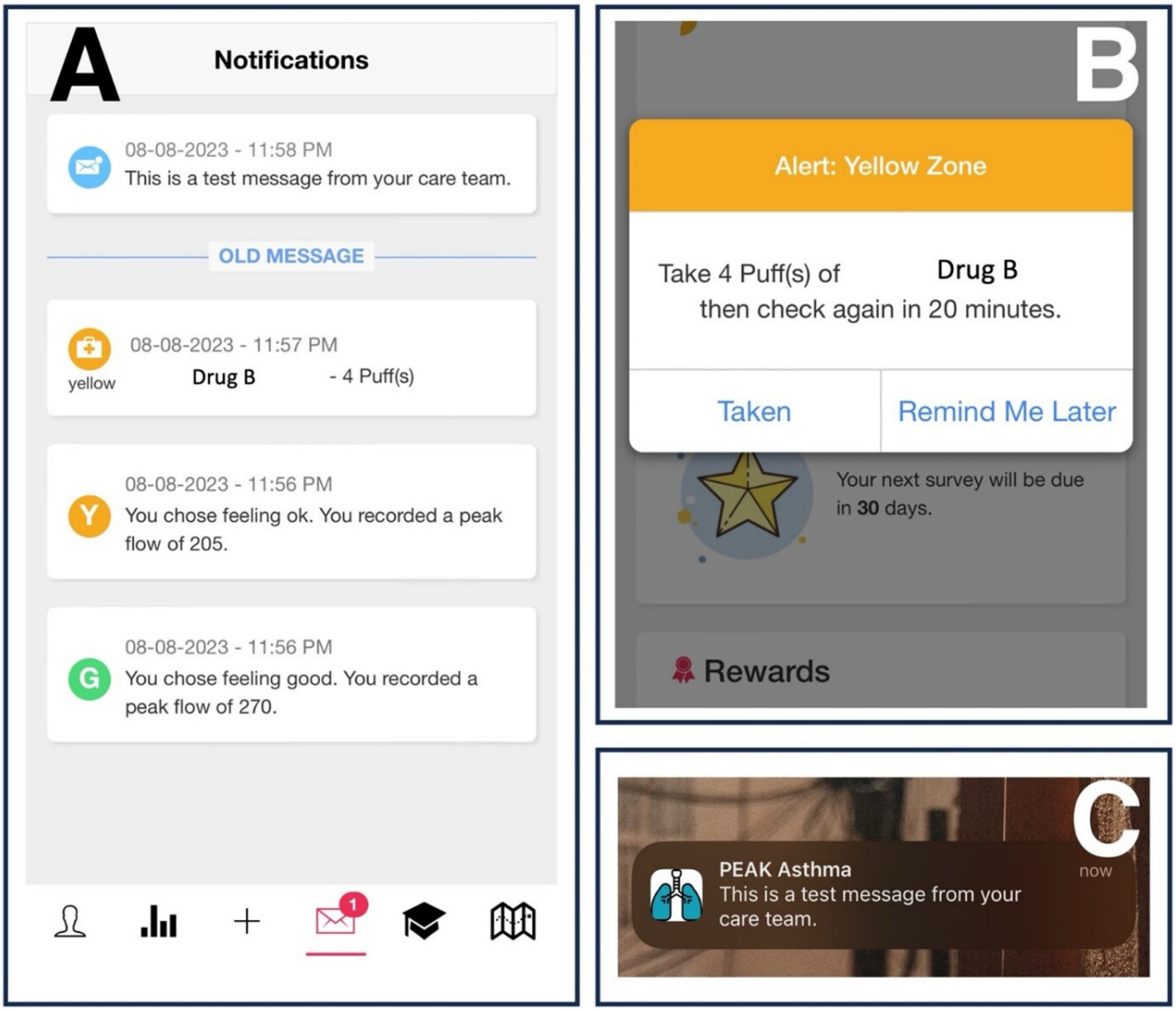
XSL•FO
RenderX

Figure 4. Pulmonary Education and Knowledge Mobile Asthma Action Plan navigates the patient through a series of simple steps to log daily asthma status and during any episodes of increased symptoms to determine their asthma zone: (**A**) "How are you feeling today?"; (**B**) "Determine your Asthma Zone"; (**C**) "Record your Peak Flow"; and (**D**) symptoms recording page. Patients can elect to enter either Peak Flow or symptoms at their preference.



XSL•FO
RenderX

Figure 5. (**A**) The history of notifications that were sent to the patient; (**B**) the in-app notification screen; and (**C**) the mobile system's native notifications.



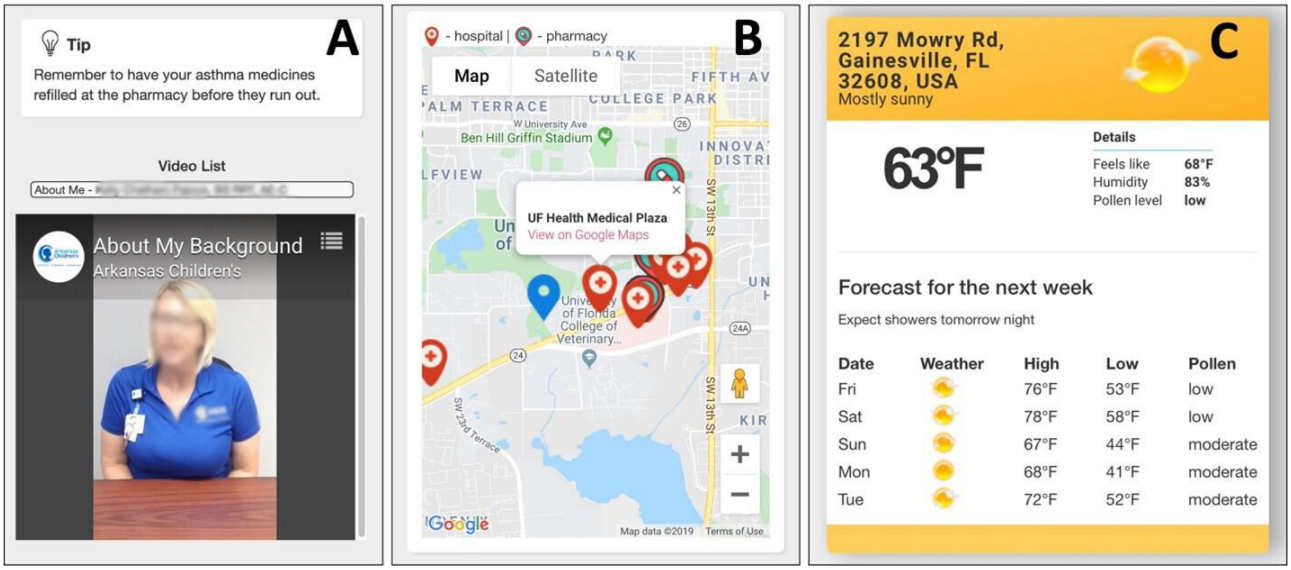
Fifth, an education module, as shown in Figure 6A, which consists of (1) an asthma educational tip card and (2) a list of asthma educational videos. The educational tip card will be refreshed weekly and the tips will also be pushed to the patients through the notification function. The asthma educational videos were created by the research team including asthma clinicians, asthma educators, and respiratory therapists using the Vidscrip platform. These resources aim to heighten patient awareness and empower them to manage their asthma more effectively.

Sixth, a nearby medical help map, as seen in Figure 6B, displaying a Google Map pinpointing nearby medical facilities, including hospitals and pharmacies. Users can click on markers for more detailed information about a facility and also receive directions via the Google Maps mobile app.

Seventh, a weather information module, as shown in Figure 6C, consisting of a weather information card, which shows real-time weather data and pollen level information that may affect patients' asthma outcomes.

XSL•FO
RenderX

Figure 6. (**A**) The education module; (**B**) the nearby medical help map; and (**C**) the weather information module.



Web Portal

We developed a web portal for the health care providers and guardians of the patients to enable monitoring of the patient's usage of PEAK-mAAP and keep track of patients' asthma status. The health care providers are granted full access, which means they can view all the patients' information and see the statistics related to the use of PEAK-mAAP. Furthermore, health care providers can create patient accounts, edit the information of the patients, and send notifications to all patients or a specific one. On the other hand, guardians are only allowed to access limited information specifically related to the patient they are associated with.

Effectiveness and Efficiency of PEAK-mAAP

During the think-aloud process, evaluators recorded a total of 39 instances in which participants encountered difficulties in correctly performing a task, referred to as errors. These errors corresponded to violations of 9 out of the 14 heuristics under evaluation. The most frequently violated heuristic principles were match (n=12) and visibility (n=8). These were followed by memory (n=5), control (n=4), and language (n=3). Additional violations were observed in the closure, minimalist, document, and error heuristics. Examples of errors related to the match principle included confusion between the notification tab and the reminder tab, as well as difficulties in navigating the map to locate hospitals and pharmacies. The visibility principle was the second most frequently violated, with errors such as difficulty distinguishing between old and new messages within the app and challenges in locating and logging symptoms that could trigger relevant notifications. A memory error example was that a participant could not remember how to log their activity for the week to receive a reward. An example of a control error was that the participant had trouble recording their peak flow using the dial to select the correct number. Not understanding the icons used on the map was an example of a language error. There were 5 principles that were not violated, that are, undo, message, consistency, feedback, and flexibility. Of the 39 total errors, 26% (n=10) were rated as minor, 69% (n=27) as major, and 5% (n=2) as catastrophic.

Satisfaction Assessment of PEAK-mAAP

Participant satisfaction with the PEAK-mAAP mobile app was assessed using the SUS, yielding an overall mean score of 83 (SD 5.54), indicating high usability satisfaction. Older participants (≥17 years old) reported higher usability scores, with a mean of 87.5 (SD 1.71), compared with younger participants (12 - 16 years old), who had a mean score of 77.5 (SD 2.36). These findings suggest that older users found the app more user-friendly. The highest agreeability was found in survey question 7 stating, "*I would imagine that most people would learn to use the app very quickly.*" In total, 12 participants reported that they strongly agreed with this statement, and the remaining responses indicated that they agreed with this statement. The greatest variability was seen in survey question 10 stating, "*I needed to learn a lot of things before I could get going with the app*" (strongly disagree=6, disagree=2, neither or neutral=2, agree=1, strongly agree=2).

Discussion

Principal Findings

The primary objective of this study was to develop and evaluate PEAK-mAAP, an innovative mHealth app designed to improve asthma self-management among adolescents. Findings indicate that PEAK-mAAP, rooted in the HBM and developed using a UCD approach, enhances adolescents' engagement in asthma management. The usability assessment provided positive feedback on the app's ease of use and its potential for supporting behavior change. These results align with our initial hypotheses that a well-integrated mHealth platform—combining educational content, symptom tracking, and behavior modification strategies—can effectively promote self-management practices.

UCD and Usability Assessment

A key factor in PEAK-mAAP's potential success is its UCD approach, which ensures the app is intuitive and user-friendly. By incorporating user feedback early in development, the app was designed to meet the needs and expectations of adolescents, thereby encouraging consistent use. When users feel that an app

XSL•FO
RenderX

is tailored to their needs, they are more likely to engage with it over time. The usability assessment results reinforced this; participants rated the app highly in usability, achieving an impressive usability score. Overall user satisfaction with PEAK-mAAP was high. The SUS scores averaged 83 (SD 5.54), with older adolescents (>17 years) rating it higher (87.5, SD 1.71) than younger participants (77.5, SD 2.36). This trend may indicate that older adolescents are generally more comfortable using mobile apps, or that certain app features are more intuitive for them. Regardless of age, participants agreed that most users would be able to learn how to use the app quickly.

PEAK-mAAP performed well in many heuristic categories. Users encountered no issues with consistency, feedback, flexibility, message, or undo. However, errors in "match" (12/39, 31%) and "visibility" (8/39, 21%) posed significant challenges. The "match" errors stemmed from inconsistencies between user expectations and actual system functions, particularly regarding reminders and navigation tools. Meanwhile, "visibility" issues were especially problematic, as some users struggled to distinguish new messages from old ones. Notably, 2 visibility errors were classified as catastrophic, as they significantly impacted user interaction with notifications.

However, despite strong usability performance, testing identified critical issues that could impact the user experience if left unaddressed. For example, users had difficulty identifying the correct icon to set Peak Flow logging reminders—many mistakenly selected the notification icon instead. In addition, the dial interface for recording Peak Flow was not user-friendly, making it challenging for users to input accurate data. This issue was particularly significant because the peak flow value determines the personalized asthma management guidance provided by the app. Younger participants also faced challenges with reading the map, likely due to their limited experience with navigation tools.

Identified Usability Issues and Modifications

To enhance usability and ensure a seamless experience, several modifications were made before the RCT. The reminder icon was redesigned to align with users' expectations, and the notification icon was changed from a bell to an envelope to prevent confusion. Similarly, the dial's sensitivity was adjusted to allow for more precise Peak Flow input, reducing the risk of inaccurate data entries. To address visibility concerns, new messages on the notification screen were color-coded (blue for unread messages, gray for read ones), and a dividing line was added to separate them. In addition, on the login screen, a password reset option was implemented to improve accessibility. Another significant modification was made to the educational content delivery. Initially, weekly asthma education videos under the education tab played consecutively, which could be overwhelming. This feature was adjusted to allow users to control playback, preventing automatic continuous play. For motivation and engagement, the app incorporated an age-appropriate reward system, represented by a progressively filling blue lung icon that visually tracked the user's daily logging progress. Finally, usability feedback also informed the development of training materials to better introduce participants to the app's features, ensuring that users fully understood its capabilities.

Advantages of the Hybrid Mobile App Approach

The decision to develop PEAK-mAAP as a hybrid mobile app brings several benefits. Hybrid apps work across different platforms, meaning more people can use the app without compatibility issues. This approach is also cost-effective since developers do not need to create separate versions for different platforms. Plus, hybrid apps can easily integrate with a phone's built-in features, enhancing the user experience. Compared with a native implementation, where the app must be developed separately for Android and iOS platforms, the hybrid approach significantly shortened the development time frame. It facilitated a rapid reach to a broader user base and proved to be more cost-effective in construction and maintenance. Furthermore, the hybrid mobile app can run on browsers, ensuring that adolescents remain engaged with PEAK-mAAP even if access to the mobile platform is unavailable, such as in the case of a lost phone.

Limitations

Despite these promising outcomes, several limitations must be acknowledged. Some users experienced an initial learning curve, which could impact short-term engagement. Longitudinal studies are needed to assess sustained usage and long-term effectiveness. Furthermore, as the efficacy of PEAK-mAAP has not yet been tested through an RCT, conclusions regarding its impact on clinical outcomes remain speculative at this stage.

Conclusion

This study presents the design, development, and usability evaluation of PEAK-mAAP, a hybrid mHealth intervention designed to support asthma self-management in adolescents. Rooted in the HBM and developed through a UCD approach, PEAK-mAAP integrates educational content, symptom tracking, medication adherence tools, and real-time feedback to enhance self-management behaviors. Usability testing revealed high user satisfaction, with strong SUS scores and positive engagement. While some challenges—particularly in navigation, reminders, and visibility—were identified, subsequent modifications improved the user experience before launching the RCT. The hybrid mobile app approach ensures cross-platform compatibility, cost efficiency, and accessibility, positioning PEAK-mAAP as a scalable and adaptable solution for adolescent asthma management. Although usability results are promising, further research should assess long-term effectiveness and sustainability in improving asthma control through a rigorous RCT. Addressing usability concerns and incorporating ongoing user feedback will be critical to fostering widespread adoption and adherence among adolescents. Ultimately, PEAK-mAAP represents a significant advancement in mHealth technology, bridging gaps in adolescent asthma care with a personalized and evidence-based approach to self-management.

XSL•FO
RenderX

Acknowledgments

This work was supported by the National Institutes of Health (NIH; grant 1R01NR015988-01A1). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We did not use generative artificial intelligence in any portion of the manuscript writing.

Data Availability

The data generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

TTP and JB designed methods and experiments. All coauthors contributed to design the app. XH developed and maintained the app. XH, JB, and TTP drafted the main manuscript. All coauthors provided critical edits of the early draft and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Usability testing agreement.
[PDF File, 87 KB - formative_v9i1e64212_app1.pdf]

References

1. The Global Asthma Report 2022. int J tuberc lung dis 2022 Nov 25;26(1):1-104. [doi: 10.5588/ijtld.22.1010]
2. Farzandipour M, Nabovati E, Sharif R, Arani MH, Anvari S. Patient self-management of asthma using mobile health applications: a systematic review of the functionalities and effects. Appl Clin Inform 2017 Oct;8(4):1068-1081. [doi: 10.4338/ACI-2017-07-R-0116] [Medline: 29241254]
3. Alquran A, Lambert KA, Farouque A, et al. Smartphone applications for encouraging asthma self-management in adolescents: a systematic review. Int J Environ Res Public Health 2018 Oct 29;15(11):2403. [doi: 10.3390/ijerph15112403] [Medline: 30380692]
4. Song T, Yu P, Zhang Z. Design features and health outcomes of mhealth applications for patient self-management of asthma: a systematic review: mhealth apps for asthma self-management. Presented at: ACSW '22: Proceedings of the 2022 Australasian Computer Science Week; Feb 14-18, 2022; Brisbane, Australia. [doi: 10.1145/3511616.3513110]
5. Farzandipour M, Nabovati E, Heidarzadeh Arani M, Akbari H, Sharif R, Anvari S. Enhancing asthma patients' self-management through smartphone-based application: design, usability evaluation, and educational intervention. Appl Clin Inform 2019 Oct;10(5):870-878. [doi: 10.1055/s-0039-1700866] [Medline: 31724144]
6. Kosse RC, Bouvy ML, de Vries TW, Koster ES. Effect of a mHealth intervention on adherence in adolescents with asthma: a randomized controlled trial. Respir Med 2019 Mar;149:45-51. [doi: 10.1016/j.rmed.2019.02.009] [Medline: 30803885]
7. Fedele DA, Thomas JG, McConville A, et al. Using mobile health to improve asthma self-management in early adolescence. Journal of Adolescent Health 2021;69(6):1032-1040. [doi: 10.1016/j.jadohealth.2021.06.011]
8. Li Z, Lu F, Wu J, et al. Usability and effectiveness of eHealth and mHealth interventions that support self-management and health care transition in adolescents and young adults with chronic disease: systematic review. J Med Internet Res 2024 Nov 26;26(1):e56556. [doi: 10.2196/56556] [Medline: 39589770]
9. Nanzer AM, Lawton A, D'Ancona G, Gupta A. Transitioning asthma care from adolescents to adults: severe asthma series. Chest 2021 Oct;160(4):1192-1199. [doi: 10.1016/j.chest.2021.05.019] [Medline: 34019863]
10. Holley S, Morris R, Knibb R, et al. Barriers and facilitators to asthma self-management in adolescents: a systematic review of qualitative and quantitative studies. Pediatr Pulmonol 2017 Apr;52(4):430-442. [doi: 10.1002/ppul.23556] [Medline: 27717193]
11. Asthma NAEAPP Third Expert Panel on the Diagnosis and Management of Asthma. Expert panel report 3: guidelines for the diagnosis and management of asthma. : National Heart, Lung, and Blood Institute (US); 2007.
12. Burbank AJ, Lewis SD, Hewes M, et al. Mobile-based asthma action plans for adolescents. J Asthma 2015;52(6):583-586. [doi: 10.3109/02770903.2014.995307] [Medline: 25494553]
13. Perry TT, Marshall A, Berlinski A, et al. Smartphone-based vs paper-based asthma action plans for adolescents. Ann Allergy Asthma Immunol 2017 Mar;118(3):298-303. [doi: 10.1016/j.anai.2016.11.028] [Medline: 28111110]
14. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Bainbridge W, editor. Encyclopedia of Human-Computer Interaction: Sage Publications; 2004:445-456.
15. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988;15(2):175-183. [doi: 10.1177/109019818801500203] [Medline: 3378902]
16. Apache Cordova. URL: https://cordova.apache.org/ [accessed 2023-08-09]

XSL•FO
RenderX

17. Ruby on Rails. URL: https://rubyonrails.org/ [accessed 2023-08-09]
18. PostgreSQL. URL: https://www.postgresql.org/ [accessed 2023-08-09]
19. AngularJS. URL: https://angularjs.org/ [accessed 2023-08-09]
20. Google Maps platform. Google for Developers. URL: https://developers.google.com/maps [accessed 2023-08-09]
21. Vidscrip. URL: https://www.vidscrip.com/ [accessed 2023-08-09]
22. Twilio. URL: https://www.twilio.com/en-us [accessed 2023-08-09]
23. Accuweather. URL: https://www.accuweather.com/ [accessed 2023-08-09]
24. Moroney L. Firebase cloud messaging. In: Moroney L, editor. The Definitive Guide to Firebase: Build Android Apps on Google's Mobile Platform: Apress; 2017:163-188. [doi: 10.1007/978-1-4842-2943-9_9]
25. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
26. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
27. Nielsen J. Why you only need to test with 5 users. Nielsen Norman Group. URL: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ [accessed 2023-08-09]
28. Charters E. The use of think-aloud methods in qualitative research an introduction to think-aloud methods. Brock Education Journal 2003;12(2). [doi: 10.26522/brocked.v12i2.38]
29. Zhang J, Johnson TR, Patel VL, Paige DL, Kubose T. Using usability heuristics to evaluate patient safety of medical devices. J Biomed Inform 2003;36(1-2):23-30. [doi: 10.1016/s1532-0464(03)00060-1] [Medline: 14552844]
30. Brooke J. SUS: a "quick and dirty" usability. In: Usability Evaluation in Industry 1996:189-194 URL: https://books.google.com/books?hl=en&lr=&id=ujFRDwAAQBAJ&oi=fnd&pg=PA189&dq=SUS+A+quick+and+dirty+usability+scale+&ots=ZaTJvZ-TAy&sig=wuNAJAYIO9uNOkFTJwcMMBBL0M [accessed 2025-04-15]

Abbreviations

AAP: Asthma Action Plan
API: application programming interface
HBM: health belief model
mAAP: mobile health Asthma Action Plan
mHealth: mobile health
PEAK-mAAP: Pulmonary Education and Knowledge Mobile Asthma Action Plan
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
SUS: System Usability Scale
UCD: user-centered design

Edited by A Mavragani; submitted 11.07.24; peer-reviewed by E Nabovati, M Praena-Crespo; revised version received 04.03.25; accepted 14.03.25; published 24.04.25.

Please cite as:
He X, Bian J, Berlinski A, Guo Y, Simmons AL, Marshall SA, Greene CJ, Brown RH, Turner J, Perry TT
Mobile Health App for Adolescent Asthma Self-Management: Development and Usability Study of the Pulmonary Education and Knowledge Mobile Asthma Action Plan
JMIR Form Res 2025;9:e64212
URL: https://formative.jmir.org/2025/1/e64212
doi:10.2196/64212

© Xing He, Jiang Bian, Ariel Berlinski, Yi Guo, A Larry Simmons, S Alexandra Marshall, Carolyn J Greene, Rita Hudson Brown, Jessica Turner, Tamara T Perry. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Tailoring Rheumatoid Arthritis Visit Timing Based on mHealth App Data: Mixed Methods Assessment of Implementation and Usability

Robert S Rudin[1], PhD; Leah M Santacroce[2], MS; Ishani Ganguli[3,4], MD, MPH; Daniel H Solomon[2,4,5], MD, MPH

[1]RAND, 20 Park Plaza, Boston, MA, United States

[2]Division of Rheumatology, Inflammation, and Immunity, Brigham and Women's Hospital, Boston, MA, United States

[3]Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, MA, United States

[4]Harvard Medical School, Boston, MA, United States

[5]Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Boston, MA, United States

Corresponding Author:
Robert S Rudin, PhD
RAND, 20 Park Plaza, Boston, MA, United States

Abstract

Background: Visits to medical subspecialists are common, with follow-up timing often based on heuristics rather than evidence. Unnecessary visits contribute to long wait times for new patients. Specialists could enhance visit timing and reduce frequency by systematically monitoring patients' symptoms between visits, especially for symptom-driven conditions like rheumatoid arthritis (RA). We previously designed an intervention using a mobile health (mHealth) app to collect patient-reported outcomes (PRO). One of several aims of the app was to assist rheumatologists in determining visit timing for patients with RA. The intervention did not reduce visit frequency.

Objective: To explore possible reasons for the lack of association between the intervention and visit frequency, we describe app usage, assess usability, and identify barriers and facilitators for using between-visit PRO data to reduce visits when patients' symptoms are stable.

Methods: We analyzed patients' use of the app by reporting adherence (percent of PRO questionnaires completed during the 12-month study) and retention (use in the last month of the study). To examine rheumatologists' experiences, we summarized views of the electronic health record (EHR)–embedded PRO dashboard and EHR inbox messages suggesting early or deferred visits. We assessed app usability using the interactive mHealth App Usability Questionnaire for Ease of Use and Usefulness for patients and the System Usability Scale for rheumatologists. We assessed rheumatologist-level effects of intervention usage using Kruskal-Wallis rank sum and equality of proportion tests. We identified barriers and facilitators through interviews and surveys.

Results: The analysis included 150 patients with RA and their 11 rheumatologists. Patients answered a median of 53.3% (IQR 34.1%-69.2%) of PRO questionnaires; this proportion varied by rheumatologist (range 40.7% - 67%). Over half of the patients used the app during the final month of the study (56%, range 51% - 65%, by rheumatologists); the median number of months of use was 12 (IQR 9-12). Rheumatologists viewed the dashboard 78 times (17.6% of 443 visits) with significant differences in viewing rates by rheumatologist (range 10% - 66%; $P<.01$). There were 108 generated messages sent to rheumatologists suggesting a deferred visit (24.4% of 443 visits) with significant differences in message counts received per visit by rheumatologist (range 10.8% - 22.6%; $P=.03$). Rheumatologists' reported barriers to offering visit deferrals included already scheduling as far out as they were comfortable and rescheduling complexities for staff. Based on 39 patient interviews and 44 surveys, patients reported 2 main barriers to app usage: questionnaire frequency not being tailored to them and reduced motivation after not discussing PRO data with their rheumatologist. A total of 5 interviewed patients received the option to defer their visits, of which 3 elected to defer the appointment and 2 chose to keep it.

Conclusions: While an mHealth app for reporting RA PROs was used frequently by patients, using these data to reduce the frequency of unneeded visits was not straightforward. Better engagement of clinicians may improve the use of PRO data.

(JMIR Form Res 2025;9:e60854) doi:10.2196/60854

KEYWORDS

mobile health; patient-reported outcomes; visit timing; visit frequency; rheumatoid arthritis; rheumatology

XSL•FO
RenderX

Introduction

Visits to medical specialists represent a substantial component of health care services. About half of patient visits to physicians in the United States are to specialists, totaling roughly 500 million visits per year at a cost of billions of dollars [1,2]. The current approach to care delivery inefficiently allocates specialty care resources: visits are scheduled largely in preset intervals (eg, 3, 6, or 12 months) that are not evidence-based and have a weak relationship to clinical need [3,4]. Sometimes these schedules are required for insurance purposes, that is, to have medications covered. Evidence for optimal visit timing is scant [3,5]. As a result, specialists' schedules fill with potentially unneeded visits, creating inefficiencies, wasteful spending, and patient burdens [6], and contribute to longer wait times for other patients who need specialist care; these access issues are especially pronounced for rural, minority, and other vulnerable populations [7].

We previously designed a mobile health (mHealth) intervention to improve patient care in rheumatology [8]. It was well accepted by patients, and we became interested in whether one additional value of the app could be to reduce visit follow-up frequency when clinically appropriate [8,9]. The rationale was that a reasonable proportion of visits were for stable patients where remote symptom reporting could help clinicians defer visits. We focused on rheumatology because this specialty has a great need for optimizing visit timing due to projected workforce shortages in the next decade [10]. Furthermore, prior works suggest a potential for reducing visit frequency, in part, because a substantial number of visits, for rheumatoid arthritis (RA) in particular, do not result in a change in treatment, suggesting limited clinical use [8,11-13]. The mHealth intervention involved patients using a smartphone app to record their symptoms using patient-reported outcome (PRO) questionnaires; based on these data, their rheumatologists received notifications that either suggested the need for an earlier visit if symptoms were acutely worsening or the potential to defer a scheduled visit if symptoms were stable (see Methods for details).

Our initial clinical trial of this intervention found no association with reduced visit frequency. To explore possible reasons for this finding and inform further intervention development, we examined quantitative and qualitative data on patient and rheumatologist engagement and data from surveys and interviews. We describe the app intervention usage, assess usability, and identify possible barriers and facilitators to improve visit follow-up frequency using between-visit PRO data.

Methods

Setting and Study Population

The prior study was conducted at 2 rheumatology practices affiliated with Brigham and Women's Hospital, which is part of Mass General Brigham, a large academic medical center in Boston, MA, United States. All clinics within this health system share an electronic health record (EHR) system (Epic Systems, Inc). The intervention was implemented by engaging 11 clinically focused rheumatologists who previously provided input on the intervention design [8,9]. As described previously, we identified eligible patients by searching upcoming patient visit schedules for the 11 rheumatologists [8,9]. We used prior diagnosis codes to identify patients with RA. Only established patients were included, and therefore all prescheduled visits were follow-ups. We recruited these patients via patient portal messages, phone calls, and in-person prior to scheduled visits.

mHealth Intervention

The intervention was designed with extensive input from patients and rheumatologists through several iterations as part of a user-centered design process [8,9,14,15]. Briefly, the intervention included a patient-facing mHealth app that used push notifications to prompt patients to complete one of four types of validated PRO questionnaires regarding pain interference, function, fatigue, and RA disease activity [8]. Each questionnaire was repeated every 8 days in a staggered fashion so that patients completed one questionnaire every 2 days. Rheumatologists were able to view all results within the patient's EHR chart. In addition, the PRO data could trigger 3 types of EHR inbox messages to the rheumatologists. First, all rheumatologists received a message that data were available for viewing 48 hours before a visit with a patient who had entered PRO data. Second, if PRO data suggested several weeks of worsening symptoms, the rheumatologist received a message with a suggestion to offer an earlier visit. Finally, if PRO data suggested stable symptoms during the visit interval from the prior visit until 2 weeks prior to a scheduled visit, rheumatologists received a message suggesting that the patient be offered a deferred visit. The specific logic that determined when these messages are sent has been described previously [8].

Study Design and Data Sources

As noted, this initial trial of the mHealth app found no change in visit frequency attributable to the PRO data and EHR inbox messages [9]. We sought to explore potential reasons for these findings and inform further intervention development by (1) describing use of the mHealth app by patients and use of the PRO data by rheumatologists, (2) assessing mHealth app and EHR-embedded PRO data usability, and (3) identifying barriers and facilitators. Patient and rheumatologist perspectives were assessed through surveys during and after the study period. Data sources included the EHR (for demographic and clinical characteristics of study patients and visit timing), mHealth app and dashboard usage logs (for patient adherence and retention, rheumatologist views of the PRO data, and messages sent suggesting early or deferred visits), interviews with patients (for barriers and facilitators), and surveys of patients and rheumatologists (for usability and barriers and facilitators).

For analyses of intervention usage and usability, rheumatologists with 15 or more patients in the study were analyzed as individuals, and the rest were aggregated. Because only one of the instances in which an earlier visit was suggested resulted in an earlier visit, we focused the analysis primarily on aspects related to deferred visits [9].

XSL•FO
RenderX

Intervention Usage Analysis

We used descriptive statistics to summarize patient adherence and retention rates with the mHealth app. We used the Kruskal-Wallis rank sum test to assess differences in median patient PRO completion percentage across rheumatologists. We used descriptive statistics to summarize volumes and timing of EHR inbox messages sent to the rheumatologists suggesting the potential for an early or deferred visit and view of the EHR-embedded PRO data. For both measures, we tested for equality of proportions across rheumatologists.

Usability Analysis

We summarized results from mHealth App Usability Questionnaire (MAUQ) items, which were administered to patients after they completed the study [16]. The MAUQ is a validated instrument measuring mHealth app usability. Specifically, we used the interactive versions of the ease of use and satisfaction (MAUQ_E) and usefulness (MAUQ_U) subscales of the questionnaire because they were most relevant to this intervention. The MAUQ scales range from 1 (worst, ie, strongly disagree with all questions) to 7 (best, ie, strongly agree with all questions), with 4 meaning neutral.

For rheumatologists, we summarized the results of the System Usability Scale (SUS) administered after all the rheumatologists' patients had completed the study [17,18]. The SUS is a widely used usability measure with a range from 0 (worst) to 100 (best), with an average score across evaluations of 68‐70.

Barriers and Facilitators Analysis

We invited patients who had received a suggested early or deferred visit to participate in an interview. We conducted those interviews virtually and transcribed the recordings. Using conventional content analysis, one research team member (RSR) created a codebook and summarized the results using the framework method, and a second research team member (DHS) reviewed a portion of the results to confirm findings [19,20]. We used similar methods to analyze free-text responses from patient survey results administered at study completion. When presenting qualitative results from interviews, we used "some" to indicate a finding was present in fewer than 5 responses, "many" to denote greater than 5 responses, and "most" to denote greater than half of the responses.

For rheumatologists, we summarized data from a survey administered after study completion about how often they offered patients the option to defer a visit and the reasons for not offering that option.

All quantitative analyses were conducted in R (version 4.2.2; R Foundation for Statistical Computing).

Ethical Considerations

This study was reviewed and approved by the institutional review board at Mass General Brigham (number 2021P000790). All written and audiovisual data (which were not anonymized) were stored on secure servers accessible only to the research team. Written consent was obtained for all patients to participate in the trial and for the use of their medical data in analyses, verbal consent was obtained for interviews, and consent information was presented before all electronic survey questions. Participants were compensated with a gift card for the trial and a US $25 check for the interviews.

Results

Participants

Characteristics of the 150 intervention patients included in the study are shown in Table 1. Patients were mostly female and White.

XSL•FO
RenderX

Table . Characteristics of patients using the mHealth app for rheumatoid arthritis (N=150).

Characteristics	Values
Sex, n (%)	
Female	125 (83.3)
Male	25 (16.7)
Age (years), median (IQR)	62 (52-69)
Race, n (%)	
Asian	4 (2.7)
Black	5 (3.3)
White	136 (90.7)
Declined or others	5 (3.3)
Seropositivity status, n (%)	
Missing	36 (24)
0	39 (26)
1	75 (50)
CRP[a,b], median (IQR)	1.4 (0.7 - 4.5)
DMARD[c,d] use, n (%)	
csDMARD[e]	93 (62)
btsDMARD[f]	90 (60)
Other medications, n (%)	
NSAIDs[g]	30 (20)
Corticosteroid	55 (36.7)
Opioids	25 (16.7)
Comorbidities, n (%)	
Hypertension	26 (17.3)
Obesity	15 (10)
Depression	7 (4.7)
Osteoporosis	19 (12.7)

[a]CRP: C-reactive protein.

[b]Two values were missing for CRP.

[c]DMARD: disease-modifying antirheumatic drug.

[d]It is common for patients to use more than 1 DMARD.

[e]csDMARD: conventional synthetic disease-modifying antirheumatic drug.

[f]btsDMARD: biologic or targeted synthetic disease-modifying antirheumatic drug.

[g]NSAID: nonsteroidal anti-inflammatory drug.

Intervention Usage

Four rheumatologists each had at least 15 patients in the intervention and were analyzed as individual observations; the other 7 rheumatologists were aggregated and treated as one observation. The median percentage of PRO questionnaires completed was 53.3% across all patients (Table 2). This completion rate varied by rheumatologists, ranging from 40.7% (the rheumatologist with the most patients) to 67% (the rheumatologist with the third most patients). Differences in completion rates between all rheumatologists or groups were not statistically significant based on the Kruskal-Wallis rank sum test (P=.50). The median number of months patients answered questionnaires was 12 (IQR 9-12), with little variation by rheumatologist; 56% (84/150) of patients completed at least one questionnaire in the final month of the study. See app usage by rheumatologists in Table 2 and over time in Figure 1.

XSL•FO
RenderX

Table . App usage by rheumatoid arthritis patients grouped by their rheumatologist.

Rheumatologist[a]	Patients, n	Patient visits, n	Percent PRO[b] questionnaires completed,[c] median (IQR)	Months with answered questionnaires,[d] median (IQR)	Patients who completed at least 1 PRO in the final month of the study, n (%)
All rheumatologists	150	443	53.3 (34.1-69.2)	12 (9-12)	84 (56)
Four rheumatologists with the greatest volume of study patients					
1st highest	51	148	40.7 (25.8-63.7)	11 (8-12)	26 (51)
2nd highest	25	68	48.9 (34.1-64.8)	12 (9-12)	13 (52)
3rd highest	20	53	67 (49.5-78.6)	12 (11-12)	13 (65)
4th highest	15	49	46.7 (44-61)	12 (9-12)	8 (53)
All other rheumatologists					
5th-11th highest	39	125	58.8 (47.3-72)	12 (11-12)	24 (62)

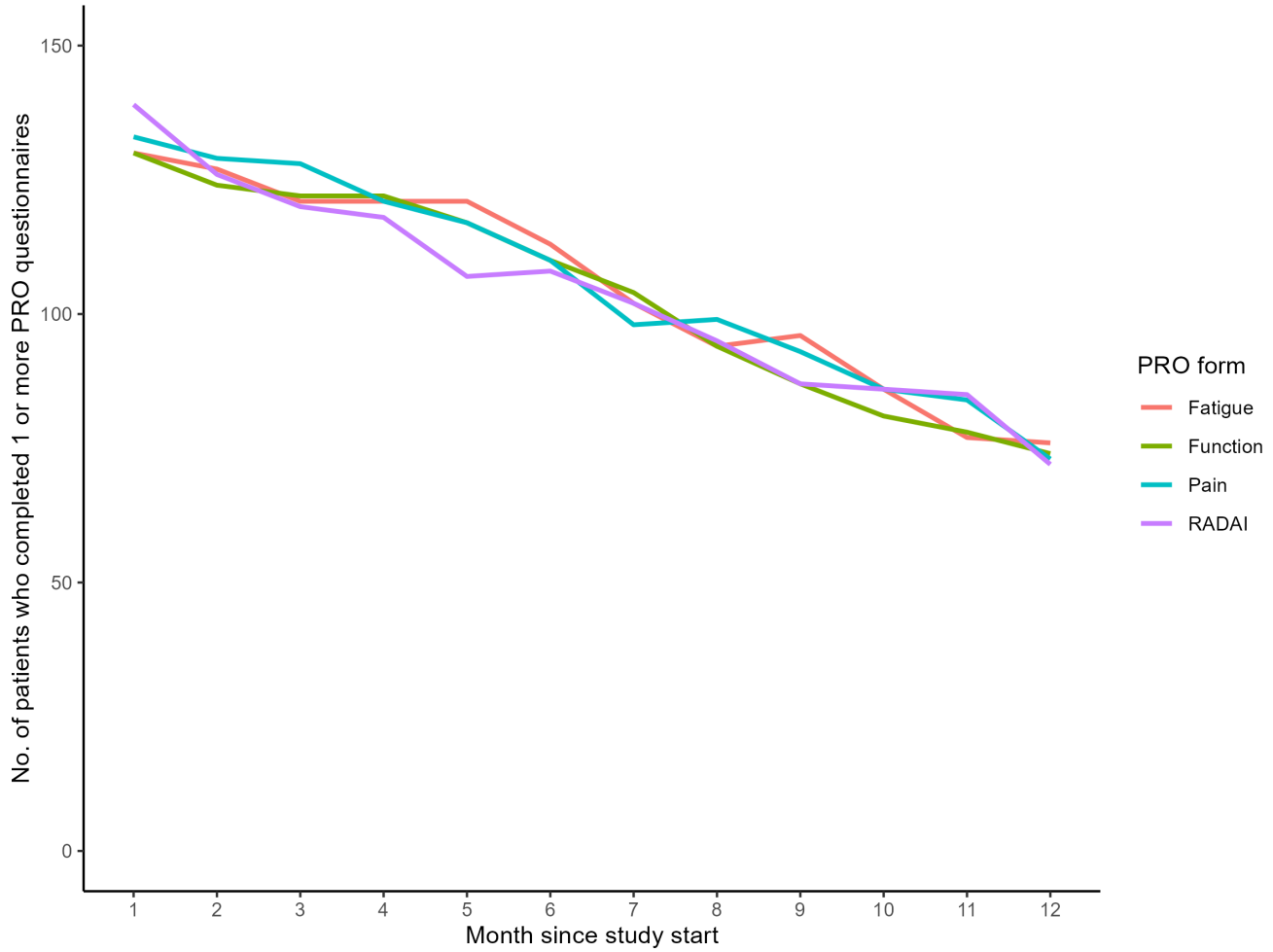
[a]Ordered by volume of study patients who receive care from that rheumatologist.

[b]PRO: patient-reported outcome.

[c]The total possible PRO questionnaires offered were 182.

[d]Number of months within the 12-month study in which patients answered at least 1 PRO questionnaire.

Figure 1. Patient questionnaire completion rates within the rheumatoid arthritis app over the 12-month study period. PRO: patient-reported outcome; RADAI: Rheumatoid Arthritis Disease Activity Index.



There was variation by rheumatologist in the percentage of their received messages that suggested a deferred visit, ranging from 18.9% of patient visits (the rheumatologist with the most patients) to 37.7% (the rheumatologist with the third most patients) (Table 3). In a 6-sample test for equality of proportions, there was a statistically significant difference in the percent of deferred visit suggestions received by rheumatologists (P=.03). There were no significant differences among early visits. For rheumatologist usage of the PRO data as measured by the number of views, there was variation by rheumatologist ranging

XSL•FO
RenderX

from 10.1% of patient visits (rheumatologist with the most patients) to 66% (rheumatologist with the third most patients). In a 6-sample test for equality of proportions, the difference was statistically significant ($P<.01$), which means that at least one value (number of views of PRO data per patient visit) differed significantly from the others.

Table . Usage of rheumatologist-facing intervention information generated by the rheumatoid arthritis app, available within the electronic health record.

Rheumatologist[a]	Patients, n	Patient visits, n	Messages sent to rheumatologist suggesting visit deferral, n (%)	Messages sent to rheumatologist suggesting early visit, n (%)	Rheumatologist views of PRO[b] data in EHR[c], n (%)
All rheumatologists	150	443	108 (24.4)	31 (6.9)	78 (17.6)
Four rheumatologists with the greatest volume of study patients					
1st highest	51	148	28 (18.9)	8 (5.4)	15 (10.1)
2nd highest	25	68	13 (19.1)	5 (7.4)	11 (16.2)
3rd highest	20	53	20 (37.7)	5 (9.4)	35 (66)
4th highest	15	49	18 (36.7)	3 (6.1)	7 (14.3)
All other rheumatologists					
5th-11th highest	39	125	29 (23.2)	10 (8)	10 (8)

[a]Ordered by volume of study patients who receive care from that provider.

[b]PRO: patient-reported outcome.

[c]EHR: electronic health record.

Usability

Of the 150 intervention patients, 88 completed the MAUQ surveys with a median MAUQ_E of 6.3 (IQR 5.6-6.9) and median MAUQ_U of 4.8 (IQR 4.1-5.6) (Table 4 and Figure 2). Based on Figure 2, most patients found the app easy to use across all questions. However, the usefulness of the app was less clear for patients. Most patients agreed that the app "would be useful for my health and well-being," but about half of patients were neutral about its usefulness on the other questions. There was a small range across rheumatologists of the aggregated MAUQ_E (6.2 - 6.9) and a more substantial range for the MAUQ_U (4.6 - 6.3).

Of 11 participating rheumatologists, 10 completed the SUS to assess the EHR-embedded dashboard and reported a mean of 68.25 (range 45 - 82.5). (The average score for the SUS across a wide range of apps is 68 (SD 12.5) [21].)

Table . Patient usability scores of the rheumatoid arthritis app at the completion of the 12-month study.

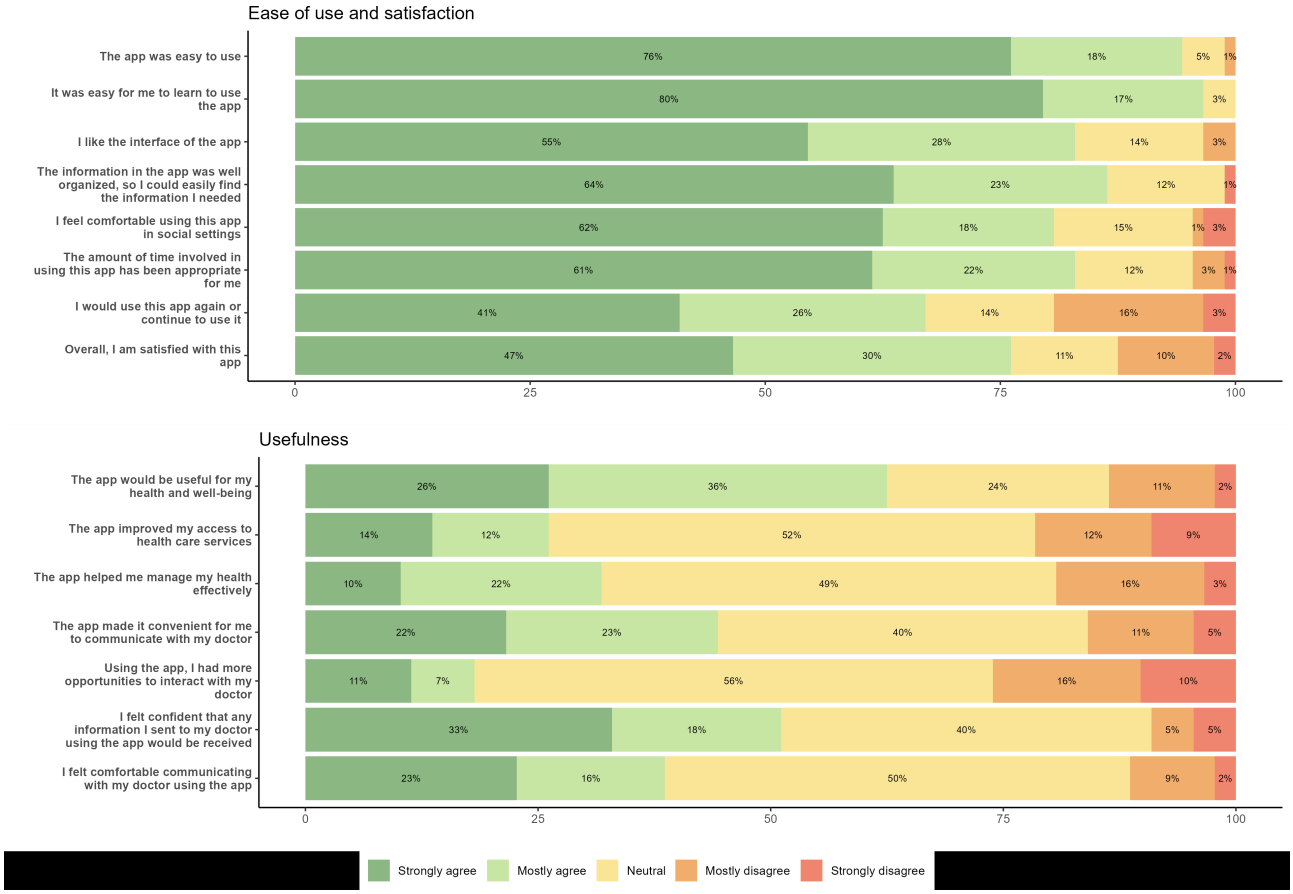
Rheumatologist[a]	Patients, n	MAUQ_E[b], median (IQR)	MAUQ_U[b], median (IQR)
All rheumatologists	88	6.3 (5.6-6.9)	4.8 (4.1-5.6)
Four rheumatologists with the greatest volume of study patients			
1st highest	31	6.2 (5.5-6.6)	4.6 (3.4-5.3)
2nd highest	17	6.3 (6.2-7.0)	4.6 (4.2-5.0)
3rd highest	12	6.7 (6.4-7.0)	5.2 (4.3-6.3)
4th highest	4	6.9 (6.2-7.0)	6.3 (5.7-6.6)
All other rheumatologists			
5th-11th highest	24	6.3 (5.6-6.9)	4.8 (3.8-5.3)

[a]Ordered by volume of study patients who receive care from that provider.

[b]The mHealth App Usability Questionnaire for Ease of Use (MAUQ_E) and Usefulness (MAUQ_U) were rescaled from 1 - 5 to 1 - 7 (worst to best) to be consistent with their standard use, with 1 indicating strongest disagreement with all questions and 7 indicating strongest agreement with all questions.

XSL•FO
RenderX

Figure 2. Rheumatoid arthritis app usability as measured by the mHealth App Usability Questionnaire for Ease of Use (MAUQ_E) and Usefulness (MAUQ_U) for clinically integrated apps, administered at the completion of the 12-month study.



Barriers and Facilitators

We interviewed 39 patients shortly after their responses triggered a message to their rheumatologist suggesting a possible early or deferred visit. We also assessed free text comments from 44 patients who completed the final study survey. Facilitators and barriers are summarized in Table 5. The most cited benefit of the intervention mentioned by patients was that the app improved awareness of their symptoms: “I really enjoyed using this app and felt like it gave me a great way to check in with myself on my symptoms and take a pulse on how my RA was” [Patient 78]. Patients described benefits both from being aware of when their symptoms were worse and from recognizing when they were feeling well, such as when a medication change was working (“I’m in the green every time, it’s amazing!” [Patient 141]).

XSL•FO
RenderX

Table . Facilitators and barriers to rheumatoid arthritis patient engagement by intervention component based on interviews with users and open-ended survey responses.

Intervention component	Facilitators	Barriers
PRO[a] questionnaires	Increased awareness of symptoms was perceived as a benefit; graph of responses helped assess the impact of medications	• Frequency: Too often for those with well-controlled or stable symptoms, and too little for those with changing symptoms, reduced motivation to complete questionnaires • Type: Questions did not always capture the nature of symptoms; many wanted to record free text notes • Timing: Some had trouble remembering symptoms during the prior 8 days
Discussing PRO scores with rheumatologists	Discussion motivated patients to continue using the app and be aware of their symptoms	• Lack of discussion reduced motivation to continue using the app
Offers for deferred visits	Perceived time saved from avoiding unneeded visits was viewed positively	• Lack of knowledge that the app would inform visit timing reduced motivation to use the app; perception that laboratory testing needs would dictate visit timing impeded the use of deferrals; receiving care from multiple rheumatologists complicated visit deferral decisions; living in different locations during the year reduced opportunities for visit deferrals; concern about the ability to find a visit slot in a suitable time frame reduced interest in visit deferrals
Offers for early visits	Potential for early visits helped make patients feel looked after; patients trust in rheumatologist's judgment made them likely to accept an earlier visit	• Lack of clarity over whether symptoms were RA-related[b] caused confusion over whether an early visit was needed
User experience with technology	App was viewed as easy to use	• Technical issues, the need for questionnaires to be available earlier in the morning, and the lack of SMS notifications were challenges to using app

[a]PRO: patient-reported outcome.

[b]RA: rheumatoid arthritis.

Although most of the patients we spoke with did not discuss the app with their rheumatologists, many did, and most of those patients were appreciative and motivated to use the app because of it: "She said 'well I can see on this graph'...it was definitely comforting because I'm in the good part...so it was good she brought it up" [Patient 28]; "I did discuss the app data with my doctor...It was very helpful to review them with her" [Patient 180]. In a few cases, patients described examples in which decisions regarding medications were made based on discussion of the app data. In 2 cases, a patient described how the app data helped inform the decision during a visit to further increase visit intervals. Some of the patients who discussed the app data with their rheumatologists said the app added little to their care: "I was cognizant that [my doctor] looked at the app and has referred to it in discussing my condition, but she is a sufficiently good and attentive communicator that I didn't notice much in the way of changed communications with her (which I considered excellent to begin with)" [Patient 145].

However, many patients were not aware that their rheumatologist has access to or viewed the data: "My doctor and I have never discussed the use of this app. I'm not even sure she knows about the app" [Patient 144]. This reduced motivation to use the app: "I was disappointed that the app was not used at all, really, in my care. I felt that the information just went into cyberspace and was never heard from again" [Patient 119]; "I don't see any connection between the app and the care I receive with my particular rheumatologist...I found it interesting that she didn't bring it up so that gave me the message that she doesn't find it helpful in her decision making" [Patient 30]. One patient noted a disconnect between her impression of the primary purpose of the app (for her rheumatologist to oversee her PRO responses and reach out if needed) and her rheumatologist's view of the app's main purpose (for patient self-management).

Of the 5 interviewed patients who reported receiving a call offering them the option to defer their visit because of stable symptoms they reported in the app, 3 accepted the offer to defer the visit: "Not having to take the trip down to Boston from Maine, as long as [my doctor] thought was fine, we said okay and we rescheduled it [from September until November]" [Patient 200]. One of those patients had a symptom flare shortly after electing to defer the visit: "the following week I had a bad rheumatoid attack and I thought 'oh shoot I should have kept that appointment' but I think it was okay" [Patient 190]. Two

XSL•FO
RenderX

decided to keep their appointment: “when she called me I was in the middle of...about 5 different joints that were inflamed and so I said to her that I’d like to keep the visit” [Patient 145]. When asked about the potential to receive such a call, almost all patients interviewed had a positive view: “If I’m feeling well and it’s just kind of a check-in, then I wouldn’t need to drive into Boston for it” [Patient 70]; “I’d do a cartwheel...it is such a hassle [to drive into Boston]...traffic can be brutal” [Patient 141].

The only patient we interviewed who received a call from the clinic offering them the option for an earlier visit based on their PRO data believed that they did not need to come in early for their RA. When asked about the potential to receive such a call, almost all patients interviewed had a positive view: “when you have to call in...you’re hesitant to do that because you know you’re always wondering ‘oh is this something that’s just going to pass in a little while’...whereas if you guys could actually see the graph changing and call...it is comforting” [Patient 28]; “I think that would be great!...I’d feel like I’m not just spinning the wheel for nothing, answering the questions” [Patient 70]. However, a few believed they were already in good communication with their rheumatologist and would have contacted them if they needed help: “Very likely I would’ve already been there...I would’ve already thought to give her a call and let her know I’m not feeling great” [Patient 30].

About half of patients interviewed or who provided free text comments on the survey mentioned challenges related to the app questionnaire: some found them too frequent (“I did not see the value and found the number of surveys excessive and as I am in remission” [Patient 84]); some found them not frequent enough (“Every day would be better than every other day because you know if you do it every day then you know when you miss a day” [Patient 103]); and some found them inadequate for capturing their relevant symptoms (“This app did NOT capture the total wellbeing of the patient. It is strictly about joints. My RA involves my lungs” [Patient 3]).

Of the 11 rheumatologists, 10 completed the final survey (Table 6). Responses showed that the major barrier for rheumatologists not offering deferred visits was that they were already scheduling patients out as far as they were comfortable with. The complexity of rescheduling visits and the need for patients to come in for labs anyway were also notable barriers.

XSL•FO
RenderX

Table . Rheumatologist survey of reasons for not offering patients the option to defer a visit, administered after completion of the 12-month study.[a]

Reasons for not offering patients with stable PROs[b] the option to defer a visit[c]	Top ranked, n[d]	In top three, n[d]	Indicated at least once
I am too busy to spend time deciding if the visit could be delayed	0	1	Yes
Most of my patients should come in even if their PROs are stable	0	1	Yes
Most of my patients need to come in for labs anyway	1	5	Yes
Rescheduling visits makes scheduling more complicated for me and my staff	2	6	Yes
Rescheduling visits increases my workload by removing easier visits	0	0	No
I am already scheduling patients as far out as I am comfortable with	3	8	Yes
If the open slots from a delayed visit are not filled, there will be lost revenue	0	0	Yes
I do not trust the PRO data to inform the decision to delay a visit	0	0	No
Delaying a visit could complicate prior authorizations	0	2	Yes
My time needed to determine if the visit could be delayed and not reimbursed	0	1	Yes
Other[e]	0	0	Yes

[a]N=10 responses; 1 rheumatologist had left the clinic, and all others responded.

[b]PRO: patient-reported outcome.

[c]Exact wording for questions: You received in-basket messages suggesting delayed visits during the study about patients with stable PROs who had visits scheduled in 2 weeks. The messages suggested that you consider asking your assistant to call the patient and give them the option to delay the visit. (1) Of these in-basket messages, please estimate how many for which you gave the patient the option to delay the visit (choose 1 response): none, some, all. (2) For patients for whom you did not offer a delayed visit, what do you think were the primary reasons? Rank all factors that influenced you (1=most influential and so on). (3) Please add any other thoughts you have about offering patients the option to delay a visit based on their PROs. (Write as much or as little as you'd like.)

[d]"Top ranked" means the respondent listed the item as their first choice. "In top three" means the respondent listed the item as their first, second, or third choice. Some respondents did not rank their responses numerically but rather selected options using an 'x'—we did not count those in the "top ranked" and only included them in the "top three" if they selected three or fewer.

[e]One rheumatologist wrote, "some patients seem to want to come in anyway, even if they don't have to come."

Discussion

Principal Results

Our examination of intervention usage, usability, and barriers and facilitators provides some potential reasons—from both rheumatologist and patient perspectives—for why our prior study did not find an effect of an intervention designed to improve visit follow-up frequency.

From the rheumatologist perspective, although usability of the EHR-embedded dashboard was satisfactory, use of the dashboard varied substantially across rheumatologists, with most using it in a small percentage of visits. Many participating rheumatologists did not offer patients deferred visits because they were not comfortable increasing visit intervals any further, believed rescheduling visits was complicated for them and their staff, or allowed the laboratory testing schedule to determine visit timing.

From the patient perspective, most continued to use the app in the final month of the study and found the app easy to use as measured by the MAUQ_E. The MAUQ_U scores were less favorable, suggesting that patients found the app less useful than easy to use. Relatively few disagreed that it was useful, but many were neutral on its usefulness. While some interviewed patients found the app useful to improve self-awareness of their symptoms, many found the questionnaires too frequent, and one patient found that the lack of questionnaires about nonarticular symptoms was problematic. Some patients were not aware that their rheumatologist would use their reported data to determine visit timing (which was correct in some cases) or that their rheumatologist even had access to these data, which

XSL•FO
RenderX

reduced their motivation to complete the questions. However, patients were largely enthusiastic about the idea of reducing the need for visits because of the time it could save them. Although some patients liked the idea of being called for earlier visits, many were skeptical that the data they reported in the app would be useful to their rheumatologists to help inform that decision.

Implications and Comparison With Prior Work

Although we identified several rheumatologist- and patient-facing barriers, the participants who engaged in the intervention provided a proof of concept that this intervention has the potential to improve the efficiency of face-to-face visits by facilitating visit deferrals when patients have stable symptoms. Between-visit symptom monitoring has other benefits, such as helping patients be more aware of their symptoms so they can get help sooner and facilitating discussions of symptoms during visits [14,15,22-26]. This work provides early evidence that improving visit efficiency can be added to that list of potential benefits. Qualitative data and the wide range of dashboard usage by rheumatologists suggest that clinician engagement is likely the critical factor that determines the extent to which visit deferrals will be offered and patients find the app useful. Other clinically integrated digital interventions have also encountered this challenge [27-29].

Clinician-facing enhancements may include generating evidence of the benefits of further spacing out visits to improve buy-in, simplifying the patient rescheduling process, and making patients preferences for longer visit intervals available to them. Patient-facing enhancements may include personalizing the timing and topics of the questionnaires based on patient responses or specified preferences, allowing the patients to add notes to elaborate on their symptoms when the questionnaires do not adequately capture them, assessing patient interest in changing visits, and informing patients within the user interface of how their rheumatologists will use the data. These enhancements require additional design and user testing. This approach may also prove promising in other medical specialties such as pulmonology and cardiology, which will also require additional design effort.

To our knowledge, this is the first attempt to use data collected between visits to inform visit timing in the United States. Prior work in rheumatology has attempted to free up appointments using scheduled nurse phone calls but did not base decisions on longitudinally collected patient data [13]. A recent rheumatology trial in the Netherlands found that app-supported patient-initiated care achieved comparable clinical results with fewer rheumatology consults for patients with stable low disease activity [12].

Limitations

This work has several limitations. First, the intervention was implemented in one academic health center with 11 rheumatologists who lacked direct incentives to be more efficient with visit timing. Other settings with different incentives may have had different experiences with the intervention. Second, the clinical logic for when an EHR inbox message was sent to the rheumatologists suggesting early or deferred visits was determined based on the expert opinion of one rheumatologist (DHS) with review and agreement by the participating study rheumatologists rather than through empirical evaluation. Therefore, the logic may have suggested changes to visit timing that were not appropriate or missed opportunities to change visits that were appropriate. Third, we were unable to comprehensively attribute an actual early or deferred visit to an app-generated message because we did not have data for when a provider initiated a patient outreach to offer an early or deferred visit in response to a study-generated message (that data is not recorded reliably or systematically in the EHR). Finally, responses to the rheumatologist survey may be influenced by social desirability bias [30]. Specifically, it is notable that none indicated that removing easier visits would be a barrier, and few selected barriers related to revenue and reimbursement, which may nonetheless be important factors to some clinicians.

Conclusions

Informing visit timing systematically based on data collected from patients between specialty visits would, among other benefits, represent a fundamental advance in health care delivery—a move from one-size-fits-all visit timing to data-informed determinations that personalize care to individual patient needs. We examined how a PRO-focused mHealth app for RA might contribute to deferring unneeded follow-up visits. The app suggested many opportunities to defer visits, some of which resulted in rescheduling visits, though many did not. Based on our quantitative and qualitative results, clinicians had inconsistent buy-in, which patients perceived. The experience of the rheumatologists who did engage in the intervention and the patients who received the option to defer a visit provide a proof of concept that this intervention has the potential to succeed in reducing visit frequency if further enhanced based on user-centered design to improve clinician engagement and patient-perceived usefulness. It is also possible that in different health care settings, that is, nonacademic practices, or when used by clinicians practicing under different incentives (eg, capitation and performance measures relevant to access), such an app that allows for rescheduling patients may achieve greater uptake and produce significant improvements in access. Directly addressing clinician buy-in is likely the key factor and requires experimentation with different kinds of implementation strategies, such as stronger leadership support, audit feedback, and direct incentives. If successfully adopted, this type of intervention may reduce per-patient visit frequency, improve patient satisfaction, and improve access.

Acknowledgments

This study was supported by the Rheumatology Research Foundation and VERITY (P30-AR072577). RSR carried out this work independently, outside of his employment at RAND.

XSL•FO
RenderX

Data Availability

The datasets generated and analyzed during this study are not publicly available to protect the privacy of the research participants.

Conflicts of Interest

RSR has received consulting fees from the Assistant Secretary for Technology Policy at the US Department of Health and Human Services unrelated to this work. DHS receives separate research funding from CorEvitas, Janssen Pharmaceuticals, and Novartis. He also receives royalty payments from UpToDate on unrelated content. IG receives consulting fees from FPrime unrelated to this work.

References

1. Santo L, Kang K. National hospital ambulatory medical care survey: 2019 national summary tables. : Centers for Disease Control and Prevention; 2023 Jan 20. [doi: 10.15620/cdc:123251]
2. Martin AB, Hartman M, Lassman D, Catlin A, National Health Expenditure Accounts Team. National health care spending in 2019: steady growth for the fourth consecutive year. Health Aff (Millwood) 2021 Jan;40(1):14-24. [doi: 10.1377/hlthaff.2020.02022] [Medline: 33326300]
3. Javorsky E, Robinson A, Boer Kimball A. Evidence-based guidelines to determine follow-up intervals: a call for action. Am J Manag Care 2014;20(1):17-19. [Medline: 24512163]
4. Ganguli I, Wasfy JH, Ferris TG. What is the right number of clinic appointments? Visit frequency and the accountable care organization. JAMA 2015 May 19;313(19):1905-1906. [doi: 10.1001/jama.2015.3356] [Medline: 25844726]
5. Jiang Y, Rudin RS, Solomon DH. Clinical visit frequencies in rheumatology: a systematic literature review. Arthritis Care Res (Hoboken) 2023 Oct;75(10):2054-2062. [doi: 10.1002/acr.25106] [Medline: 36807719]
6. Ganguli I, Chant ED, Orav EJ, Mehrotra A, Ritchie CS. Health care contact days among older adults in traditional Medicare: a cross-sectional study. Ann Intern Med 2024 Feb;177(2):125-133. [doi: 10.7326/M23-2331] [Medline: 38252944]
4. Sequist TD. Ensuring equal access to specialty care. N Engl J Med 2011 Jun 9;364(23):2258-2259. [doi: 10.1056/NEJMe1103390]
8. Solomon DH, Dalal AK, Landman AB, et al. Development and testing of an electronic health record-integrated patient-reported outcome application and intervention to improve efficiency of rheumatoid arthritis care. ACR Open Rheumatol 2022 Nov;4(11):964-973. [doi: 10.1002/acr2.11498] [Medline: 36099161]
9. Solomon DH, Altwies H, Santacroce L, et al. A mobile health application integrated in the electronic health record for rheumatoid arthritis patient–reported outcomes: a controlled interrupted time - series analysis of impact on visit efficiency. Arthritis Rheumatol 2024 May;76(5):677-683. [doi: 10.1002/art.42774]
10. Battafarano DF, Ditmyer M, Bolster MB, et al. 2015 American College of Rheumatology Workforce Study: supply and demand projections of adult rheumatology workforce, 2015–2030. Arthritis Care Res (Hoboken) 2018 Apr;70(4):617-626. [doi: 10.1002/acr.23518]
11. Hewlett S, Kirwan J, Pollock J, et al. Patient initiated outpatient follow up in rheumatoid arthritis: six year randomised controlled trial. BMJ 2005 Jan 22;330(7484):171. [doi: 10.1136/bmj.38265.493773.8F] [Medline: 15546895]
12. Seppen B, Wiegel J, Ter Wee MM, et al. Smartphone-assisted patient-initiated care versus usual care in patients with rheumatoid arthritis and low disease activity: a randomized controlled trial. Arthritis Rheumatol 2022 Nov;74(11):1737-1745. [doi: 10.1002/art.42292] [Medline: 35818342]
13. Butt S, Newman E, Smith N. Nurse scheduled telephone visit: the right rheumatology care for the right patient at the right time. Arthritis Rheumatol 2016;68(Suppl 10) [FREE Full text]
14. Colls J, Lee YC, Xu C, et al. Patient adherence with a smartphone app for patient-reported outcomes in rheumatoid arthritis. Rheumatology (Sunnyvale) 2021 Jan 5;60(1):108-112. [doi: 10.1093/rheumatology/keaa202]
15. Lee YC, Lu F, Colls J, et al. Outcomes of a mobile app to monitor patient-reported outcomes in rheumatoid arthritis: a randomized controlled trial. Arthritis Rheumatol 2021 Aug;73(8):1421-1429. [doi: 10.1002/art.41686] [Medline: 33559338]
16. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500. [doi: 10.2196/11500] [Medline: 30973342]
17. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
18. Lewis JR. The System Usability Scale: past, present, and future. Int J Hum Comput Interact 2018 Jul 3;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
19. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
20. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13(1):117. [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
21. Hyzy M, Bond R, Mulvenna M, et al. System Usability Scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290. [doi: 10.2196/37290] [Medline: 35980732]

XSL•FO
RenderX

22. Rudin RS, Fanta CH, Qureshi N, et al. A clinically integrated mHealth app and practice model for collecting patient-reported outcomes between visits for asthma patients: implementation and feasibility. Appl Clin Inform 2019 Oct;10(5):783-793. [doi: 10.1055/s-0039-1697597] [Medline: 31618782]
23. Rudin RS, Fanta CH, Predmore Z, et al. Core components for a clinically integrated mHealth app for asthma symptom monitoring. Appl Clin Inform 2017 Oct;8(4):1031-1043. [doi: 10.4338/ACI-2017-06-RA-0096] [Medline: 29241243]
24. Sulca Flores JA, Dalal AK, Sousa J, et al. Evaluation of a primary care-integrated mobile health intervention to monitor between-visit asthma symptoms. Appl Clin Inform 2024 Aug;15(4):785-797. [doi: 10.1055/s-0044-1788978] [Medline: 39357877]
25. Meirte J, Hellemans N, Anthonissen M, et al. Benefits and disadvantages of electronic patient-reported outcome measures: systematic review. JMIR Perioper Med 2020 Apr 3;3(1):e15588. [doi: 10.2196/15588] [Medline: 33393920]
26. Bartlett SJ, De Leon E, Orbai AM, et al. Patient-reported outcomes in RA care improve patient communication, decision-making, satisfaction and confidence: qualitative results. Rheumatology (Oxford) 2020 Jul 1;59(7):1662-1670. [doi: 10.1093/rheumatology/kez506] [Medline: 31665477]
27. Rudin RS, Thakore N, Mulligan KL, Ganguli I. Addressing the drivers of medical test overuse and cascades: user-centered design to improve patient-doctor communication. Jt Comm J Qual Patient Saf 2022 Apr;48(4):233-240. [doi: 10.1016/j.jcjq.2022.01.005] [Medline: 35177360]
28. Ganguli I, Mulligan KL, Chant ED, et al. Effect of a peer comparison and educational intervention on medical test conversation quality: a randomized clinical trial. JAMA Netw Open 2023 Nov 1;6(11):e2342464. [doi: 10.1001/jamanetworkopen.2023.42464] [Medline: 37943557]
29. Greenhalgh T, Wherton J, Papoutsi C, et al. Analysing the role of complexity in explaining the fortunes of technology programmes: empirical application of the NASSS framework. BMC Med 2018 May 14;16(1):66. [doi: 10.1186/s12916-018-1050-6] [Medline: 29754584]
30. Kreuter F, Presser S, Tourangeau R. Social desirability bias in CATI, IVR, and web surveys: the effects of mode and question sensitivity. Public Opin Q 2008 Dec 1;72(5):847-865. [doi: 10.1093/poq/nfn063]

Abbreviations

EHR: electronic health record
MAUQ: mHealth App Usability Questionnaire
MAUQ_E: mHealth App Usability Questionnaire for Ease of Use
MAUQ_U: mHealth App Usability Questionnaire for Usefulness
mHealth: mobile health
PRO: patient-reported outcome
RA: rheumatoid arthritis
SUS: System Usability Scale

Edited by A Mavragani; submitted 02.12.24; peer-reviewed by J Knitza, M Bass; revised version received 28.01.25; accepted 28.01.25; published 21.04.25.

Please cite as:
Rudin RS, Santacroce LM, Ganguli I, Solomon DH
Tailoring Rheumatoid Arthritis Visit Timing Based on mHealth App Data: Mixed Methods Assessment of Implementation and Usability
JMIR Form Res 2025;9:e60854
URL: https://formative.jmir.org/2025/1/e60854
doi:10.2196/60854

© Robert S Rudin, Leah M Santacroce, Ishani Ganguli, Daniel H Solomon. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Developing an Online Community Advisory Board (CAB) of Parents From Social Media to Co-Design an Human Papillomavirus Vaccine Intervention: Participatory Research Study

Regan M Murray[1], MPH; Shawn C Chiang[2], MPH, PhD; Ann C Klassen[3], PhD; Jennifer A Manganello[4], MPH, PhD; Amy E Leader[5], MPH, DrPH; Wen-Juo Lo[6], MA, PhD; Philip M Massey[7], MPH, PhD

[1]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[2]Department of Health Behavior, Texas A&M University School of Public Health, College Station, TX, United States

[3]Department of Community Health and Prevention, Drexel University, Philadelphia, PA, United States

[4]Department of Health Policy, Management and Behavior, University at Albany, State University of New York, Albany, NY, United States

[5]Division of Population Science, Department of Medical Oncology, Thomas Jefferson University, Philadelphia, PA, United States

[6]Department of Counseling, Leadership, and Research Methods, University of Arkansas, Fayetteville, AR, United States

[7]Department of Community Health Sciences, University of California, Los Angeles, 650 Charles E Young Drive S, 4th Fl, Los Angeles, CA, United States

Corresponding Author:
Philip M Massey, MPH, PhD
Department of Community Health Sciences, University of California, Los Angeles, 650 Charles E Young Drive S, 4th Fl, Los Angeles, CA, United States

Abstract

Background: Social media health interventions have grown significantly in recent years. However, researchers are still developing innovative methods to meaningfully engage online communities to inform research activities. Little has been documented describing this approach of using online community advisory boards (CABs) to co-create health communication interventions on social media.

Objective: This study describes the formation, engagement, and maintenance of an online CAB focused on co-creating a health education intervention for parents regarding the human papillomavirus (HPV) vaccine. The study provides guiding principles for public health researchers implementing such CABs in future digital health interventions.

Methods: In May 2020, Twitter was used to recruit parents of children aged 9‐14 years, who were active users of the platform and were interested in serving on a CAB focused on child health and online programs. The recruitment campaign included Twitter (rebranded as X in 2023) advertising tools (eg, “interests” and “audience look-a-likes”). A total of 17 parents completed a screening survey and 6 completed a follow-up phone interview. Following phone interviews, 6 parents were invited to join the CAB, where they committed to a 1-year involvement. The CAB participated in eleven 1-hour online meetings in the first year, contributing to monthly feedback through participatory workbooks. Long-term engagement was sustained through icebreakers and casual online interactions, as well as providing real-time updates to demonstrate CAB feedback integration. An anonymous midterm evaluation was conducted at the end of the project’s first year to assess processes and identify future growth opportunities.

Results: A total of 6 parents (5 females and 1 male) with children aged 9-14 years from diverse racial and ethnic backgrounds (African American, South Asian American, and White) across 6 states in the United States, representing urban, suburban, and rural areas, agreed to serve as CAB members. All 6 CAB members committed to 1 year of service beginning in July 2020 with 4 extending their participation into a second year (August 2021-August 2022). The CAB provided expert insights and feedback to co-develop the intervention, including character development, narrative content creation, study recruitment, survey development, and intervention delivery. The midterm evaluation showed 100% (6/6) satisfaction among CAB members, who valued the connections with other parents and their contribution to research. While all members felt confident discussing HPV, 83% (5/6) suggested diversifying the group and increasing informal bonding to enhance engagement and inclusivity, especially for differing vaccination views.

Conclusions: This study demonstrates that online CABs are a highly effective model for co-creating and informing online health communication interventions. The engagement of parents from diverse backgrounds and the structured use of online tools (eg, interactive workbooks) creates a constructive and thoughtful environment for incorporating parent contributions to research. This study highlights guiding principles to forming, engaging, and maintaining an online CAB to enhance health research and practice.

(JMIR Form Res 2025;9:e65986) doi:10.2196/65986

XSL•FO
RenderX

KEYWORDS

online community advisory boards; community engagement; social media; digital health; digital health intervention; HPV vaccine; human papillomavirus; HPV; parent health; child health

Introduction

Community advisory boards (CABs) are a widely integrated methodology in participatory health research to understand community needs, advise research, and ultimately improve health behaviors or health outcomes [1]. Comprising community members, leaders, stakeholders, and others, CABs are united by a common health interest, identity, status, or experience [2]. In the field of public health, CABs are integral to research programs to provide unique perspectives and guidance related to proposed research questions and interventions. Researchers often implement participatory strategies and inspire innovative discussions that lead to novel approaches to address health-related issues that are aligned with community needs [3]. Over time, as community needs evolve and change, CABs can also shape the development and adaptation of interventions to better align with those identified needs. Engaging community members early and often in the research process is an effective strategy to increase efficacy of the designed intervention, enhance transparency, and reduce population-level disparities [4].

In the context of child health, advisory boards consisting of parents or caregivers are essential for ensuring that public health research and interventions are aligned with family values and needs [5-7]. CABs comprising parents and caregivers have addressed a wide range of public health topics including but not limited to asthma [8], Early Head Start programs [9], childhood obesity [10], and childhood vaccination [11]. For parents, opportunities to engage in research focused on child health have been shown to be driven by altruistic motivations, previous experience in research or health care, and desire to share personal perspectives [12]. Parents are often the ultimate decision makers when it comes to their child's health; incorporating their input and understanding their perspectives provides important insights into facilitating the adoption of health behaviors among children.

As public health research and practice are incorporated more into the social media and online environment, the need for online CABs is essential. For parents, online and internet-delivered health interventions targeting child health continue to grow [13-15] and CABs will be an important approach to engage and work alongside diverse communities. Already, online opportunities to engage parents in research have allowed participation of varied perspectives across diverse geographies and facilitated ease of involvement by working around demanding parent schedules [12]. By shifting to the online environment, research teams can engage diverse populations through flexible and accessible means. Heightened flexibility allows researchers to swiftly adapt to the evolving needs of the community, as well as the ever-evolving nature of social media platforms and the online environment.

Engaging with diverse populations with varied perspectives is crucial to fostering trust, developing culturally relevant and accepted interventions, and encouraging participation in health behavior interventions [16]. This is further bolstered by expanding into the sphere of community-based participatory research through online platforms. Furthermore, as elements of everyday life increasingly transition to online platforms, the need for robust online CABs becomes paramount. This aligns with the broader trend of hybrid and remote work environments following the pandemic [17], offering flexibility that accommodates the diverse schedules and locations of members, similar to how many jobs and careers have adapted to remote work.

While the need to integrate online CABs into public health research is apparent, particularly as social media becomes an increasingly influential health intervention tool, there is a gap in evidence regarding the process of developing and maintaining online CABs. Addressing this gap is essential to promote thoughtful and systematically replicable online CABs in future public health efforts. Guiding frameworks play an important role in the development of effective CABs by providing researchers with structure and strategies for engaging CAB members from creation to evaluation [18]. Scholars interested in evaluating processes related to community-engaged advisory strategies have adapted earlier frameworks to meet the needs of systematically evaluating CABs [2]. The framework by Newman et al [2] identifies 3 primary domains that support the necessary functions of community advisory boards: formation, operation, and maintenance. Other researchers have found utility in following this guidance as well [19]. This study uses the framework developed by Newman et al [2] with some slight modifications in terminology; we refer to the "operation" domain labeled by Newman as "engagement."

This case study contributes to limited evidence on approaches and procedures for developing online CABs. We provide a case study, focusing on the development of human papillomavirus (HPV) vaccine messages for parents who use Twitter (rebranded as X in 2023), a microblogging social media platform, to share evidence and findings to support the development of future online CAB creation. Briefly, the HPV vaccine is an important topic for child and adolescent health, and parents are critical stakeholders in deciding whether their child receives the HPV vaccine. The HPV vaccine is routinely recommended for males and females between the ages of 9 and 12 years and can also be administered up to the age of 26 years and older in some cases [20]. The vaccine protects against certain strains of HPV that are associated with various cancers, including cervical, anal, throat, and penile cancers, as well as genital warts [20]. Parent involvement in online CABs is supported by previous studies showing that the inclusion of parents in HPV vaccine communication interventions resulted in improved vaccination rates [21], decreased vaccine hesitancy [22], as well as increased education and knowledge [23].

In this study, we describe how we assembled an online CAB consisting of parents from across the United States to serve in

XSL•FO
RenderX

an advisory role in the development of an online intervention. The goal of the CAB was to co-create a social media intervention on Twitter to serve as an information source for parents when deciding to get the HPV vaccine for their child. We used the framework by Newman et al [2] of formation, engagement, and maintenance to organize our Methods, Results, and Discussion sections. The purpose of this study is to describe promising strategies to be used and adapted for the creation of online CABs, ultimately to continue to support the development of participatory online health research and practice.

Methods

Formation

To initiate the creation of the online CAB, the study team first clearly defined the purpose, roles, expected functions, and membership and recruitment strategies tailored to an online format [2]. It was determined the CAB would assume an advisory role to provide support, feedback, and review of digital materials based on the project's needs. Given the study's focus on developing an online digital health intervention delivered through Twitter, recruitment was purposefully conducted on the same social media platform to align with the target audience. To recruit potential CAB members, the study team leveraged Twitter Ads Manager's Promoted Ad feature to create and disseminate targeted advertisements aimed at parents with children aged 9 - 14 years in the United States. At the time this study took place, the platform allowed extensive targeting capabilities and enabled the team to specify demographic filters and audience interests for purposive recruitment. Figure 1 shows an example of these recruitment ads.

In May 2020, parents were recruited using Twitter's "Promoted Ad" feature, which allowed the team to target broad user groups by specifying specific demographics, interests, and online behaviors and connections [24], making it a useful tool for online CAB recruitment as well as study participant recruitment more broadly. All advertisements were tailored based on age, gender, location, and language. To maximize outreach and engagement within the online environment, the study team optimized targeting by filtering for specific "keywords," "interests," "conversation topics," and "follower look-alikes" to reach parents with children aged 9 - 14 most likely to engage with the intervention platform. Table 1 provides a detailed breakdown of ad parameters, including follower look-alikes targeting, where the ads targeted parents with similar interests to followers of popular parent-related accounts on Twitter.

Interested parents were directed to an online screening survey linked within the promoted ads. The survey collected demographic information, Twitter and other social media usage patterns, and HPV vaccine awareness. Multimedia Appendix 1 provides the full list of screening questions. Parents who met the selection criteria (ie, used Twitter at least once a week, had a child aged 9 - 14 years, and agreed to be contacted by the study team) were then invited to online interviews conducted through videoconferencing by a member of the study team.

The interview guide, provided in Multimedia Appendix 2, collected information about the individual's background, interest in the project topic (ie, HPV vaccination), and assessed their fit as potential CAB members. The interviews also served as an opportunity for the study team to provide more detailed information about the project, the online nature of the CAB, member responsibilities, compensation, and next steps. CAB responsibilities described included the anticipated time commitment (ie, 10 online meetings over one year beginning in August 2020).

Figure 1. Recruitment advertisement graphic used on Twitter to recruit parents to the online community advisory board in May 2020.

XSL•FO
RenderX

Table . Twitter ad parameters used for recruiting parents to the community advisory board in May 2020.

Parameter	Definition	Ads for mothers	Ads for fathers
Sex	—[a]	Female	Male
Age range (years)	—	25-54	25-54
Location	—	United States	United States
Keywords	Target or exclude people who searched for, tweeted with, or engaged with keywords.	• Mom • Parent • Family • Health • Parenting	• Dad • Parent • Health • Parenting • Family
Interests	Choose broad or narrow interests to reach the right people	• Parenting teens • Health news and general info	• Parenting teens • Health news and general info
Conversation topics	Target people who tweet about or engage with specific conversation topics.	• Parenting teens	• Parenting teens
Follower look-alikes	Target people with interests similar to an account's followers. For example, enter @TwitterAds to target people likely to be interested in advertising on Twitter.	• @ScaryMommy • @AmerAcadPeds • @MomCentral • @RookieMoms	• @FatherlyHQ • @LifeofDadShow • @AmerAcadPeds

[a]Not available.

Engagement

CAB members collaborated with the study team to establish key activities, leadership structure, and operating principles. The group collectively determined meeting frequency, refined member roles and responsibilities, and reviewed the online research processes, including obligations to the University's Institutional Review Board, as well as the study design and procedures. The remote nature of the CAB allowed for us to adopt a flattened hierarchy to ensure equal responsibility and input among members without a formal leadership structure.

Recognizing the importance of capturing meaningful, systematic feedback in an online setting, the study team developed interactive online workbooks that included questions, examples, and open-ended prompts to elicit parent input on project materials. These workbooks were shared digitally with CAB members one week before each scheduled monthly meeting, as members indicated this to be the optimal time frame for preparation. Completed workbooks were analyzed by the study team to identify themes for discussion during meetings. The workbooks served as a structured and accessible tool for gathering organized feedback on topics such as parenting experiences and HPV vaccine perceptions. Multimedia Appendix 3 provides an example workbook used to solicit input and feedback on developing non-narrative messages for the intervention.

For virtual meeting logistics, each session began with a short icebreaker to build rapport among members, workbooks were then discussed at length, giving CAB members an opportunity to provide real-time feedback on workbook topics. All meetings were recorded, and meeting minutes were distributed digitally to both the study team and CAB members, ensuring those unable to attend (which was rare) could stay updated. This online approach enabled flexibility and sustained engagement, accommodating busy schedules and parental responsibilities.

Feedback from the CAB was incorporated iteratively into the study. The study team maintained transparency through regular online updates, presenting how CAB suggestions influenced the intervention design and delivery. This iterative process strengthened trust and collaboration in a fully online setting.

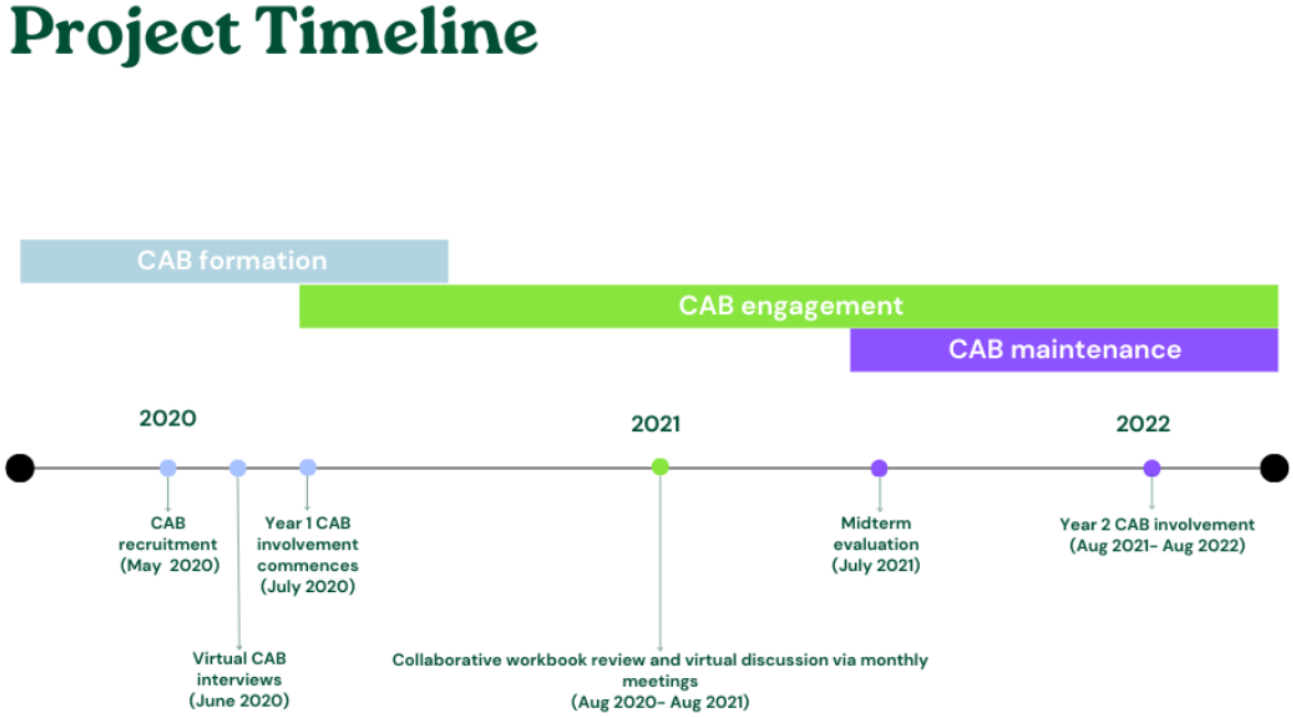
Maintenance

To evaluate the success and identify challenges encountered by CAB members during the first year, an anonymous midterm was conducted in July 2021, using an online survey and videoconference discussions at the end of the first year. The evaluation included measures examining the overall member satisfaction, meeting logistics, quality of activities and involvement, and what they envisioned for the future of the CAB. Multimedia Appendix 4 provides the full list of midterm evaluation questions.

To encourage sustained engagement, multiple digital strategies were used, including compensating members for their time and participation (ie, US $40 for workbook preparation and meeting attendance), conducting regular online meetings (ie, monthly during the first year) through videoconferencing, and sharing annual progress presentations to highlight study progress and CAB contributions to the study. To further celebrate their involvement, CAB members were provided a certificate of membership at the conclusion of the 1-year commitment. CAB members were invited to continue their participation for another year, extending their advisory role until August 2022. After completing the midterm evaluation, CAB activities were adjusted based on feedback collected and the remaining activities in the project lifecycle. Figure 2 provides an overview of the CAB activity timeline.

XSL•FO
RenderX

Figure 2. Project timeline of community advisory board activities (May 2020-August 2022). CAB: community advisory board.



Ethical Considerations

Evaluation activities were deemed exempt by the University of Arkansas Institutional Review Board (protocol 2108351667). All results presented in this manuscript are anonymous or deidentified, with no identification of individual CAB members included in the paper or supplementary materials. In addition, CAB members were compensated for their time and participation in the project, receiving US $40 in e-gift cards per meeting (US $20 for completing a workbook and US $20 for attending the meeting).

Results

Formation

In May 2020, as a part of a targeted social media recruitment campaign on Twitter, 17 parents completed the CAB screening survey. Of the 17, 6 responded to requests for an interview through videoconferencing. Following these interviews, all 6 parents were invited to join the parent advisory board, with all 6 accepting the offer. Table 2 details the CAB membership consisting of 5 mothers and 1 father, reflecting diversity across several key dimensions. The CAB represented diversity in terms of geography (Texas, Illinois, Pennsylvania, Indiana, Georgia, and Washington DC), community-type (urban, suburban, and rural), vaccination experience (half have vaccinated their child, though none were against getting the HPV vaccine for their child), and race and ethnicity (Black or African American, South Asian, Hispanic, and White).

Parents committed to 1-year involvement, which included participation in monthly online meetings. At the conclusion of the first year, 2 members stepped down due to personal commitments, while 4 members elected to remain active advisors to the project for an additional year, until the project's completion.

Table . Community advisory board demographic composition (deidentified), July 2020.

Parental role	Race and ethnicity	Location	Community type	Vaccination perspective
Mother	Black	Washington, DC	Urban	Vaccine advocate
Mother	White	Indiana	Rural	Vaccine advocate
Mother	White	Georgia	Suburban	Vaccine hesitant
Mother	Hispanic	Pennsylvania	Urban	Vaccine advocate
Mother	South Asian	Maryland	Suburban	Vaccine advocate
Father	Black	Texas	Urban	Vaccine advocate

Engagement

During the first year of the study, the CAB actively participated in 11 1-hour meetings conducted through Zoom (Zoom Communications). Meetings were held in the evening, typically 7 PM Eastern Time, to best accommodate parents' schedules across time zones. As the study transitioned into its second year and moved into the implementation phase, meeting frequency was reduced to six 1-hour meetings over one year. This adjustment reflected the reduced need for iterative feedback.

XSL•FO
RenderX

The CAB provided valuable feedback on a range of HPV vaccine topics informed by their lived experiences as parents. Key contributions included identifying scientifically relevant messaging for the HPV vaccine, developing personas and refining the storylines used in the narrative communication intervention [25], co-creating Twitter messages, and reviewing study recruitment ads. Table 3 describes how CAB input informed various aspects of the intervention, as well as survey refinement and pilot testing of materials.

CAB involvement ultimately led to the development of a 2-week HPV vaccine intervention delivered on Twitter. The intervention included eight topics related to the HPV vaccine: (1) character introductions, (2) vaccine normative beliefs, (3) vaccine knowledge and awareness, (4) vaccine accessibility, (5) vaccine safety, (6) HPV vaccine for boys, (7) vaccine disparities and equity, and (8) cancer prevention. Multimedia Appendix 5 provides examples of the first two topics, or "chapters," of the program, entitled "Meet the parents!" and "Parents are getting their kids the HPV vaccine."

XSL•FO
RenderX

Table . Examples of community advisory board's impact on quality, reach, and outcomes of the intervention (July 2020-August 2022).

Topic	Timeframe	CAB[a] activity goal	CAB feedback and integration
CAB formation and project integration	July 2020-December 2021	• Create CAB operating procedures • Inform study activities with CAB input and feedback.	• "Having the full study team present is not a problem-it is actually an advantage to be able to put names and faces together" • "I really enjoy the breakout rooms and feel we can speak about the topic but also get to know the other members a bit better." • "I like the mix of face to face zoom meetings and pdf workbooks."
Persona development	January 2021-May 2021	• Alter personas to be more relevant and relatable.	• "Many elements of this [persona] description remind me of myself or some in my closest circle." • "As I was reading [the persona descriptions], I could name specific people in my life who fit the characteristics of each character."
Intervention content creation	April 2021-August 2021	• Identify key themes for presenting information in the intervention. • Review narrative intervention content.	• "Knowledge and awareness has to come first to introduce the HPV vaccine into the general conversation. With information about accessibility, the audience can learn that [HPV is] not something that only certain people can get…" • "Love the images and think they will be effective for communities of color in particular and all parents in general." • "The use of gifs and memes is great."
Survey development	April 2021-December 2021	• Ensure questions are relevant to parents and their vaccine decision making.	• "People have very different relations with their providers and that plays a role in their decision making, making it important to include."
Study recruitment	November 2021-August 2022	• Develop recruitment ads with relevant language and images.	• "I'd like more graphics to share for recruitment"
Intervention delivery	December 2021-August 2022	• Advise on the various ways parents use Twitter. • Advise on the logistics of implementing the intervention on Twitter.	• "The order of tweets showing up in the feed is confusing and out of order." • "Using [Twitter] threads [to deliver the intervention], makes the most sense."

[a]CAB: community advisory board.

Maintenance

Table 4 shows the summarized themes from the midterm evaluation, and the full list of evaluation questions is provided in Multimedia Appendix 4.

XSL•FO
RenderX

Table . Qualitative themes of the anonymous community advisory board (CAB) midterm evaluation (July 2021).

Theme	Subtheme	Illustrative finding
Function of the CAB[a]	• Frequency • Convenience	• Online meetings after normal business hours work best. • Parents are happy with the meeting time and meeting frequency, and some would like to meet more often.
CAB research engagement	• Confidence • Knowledge building	• All parents expressed confidence in describing the project and purpose of the advisory board. • Everyone feels a great deal more knowledgeable about HPV[b] and the HPV vaccine since joining the advisory board. • All parents thought their input was well incorporated into the project.
Quality of the CAB	• Comfort • Collaboration	• Parents largely feel safe and comfortable sharing ideas and opinions. • Some discomfort sharing unpopular opinions (eg, vaccine hesitancy). • Parents really enjoy working with one another and appreciate the connection they have with other board members.
Next steps for the CAB	• Increased diversity	• All parents feel that it would be beneficial to recruit additional, more diverse people to join the advisory board. • Parents have an interest in participating in additional research activities

[a]CAB: community advisory board.

[b]HPV: human papilloma virus.

The midterm evaluation highlighted a high level of satisfaction among CAB members regarding their experience as advisory board members. Members particularly valued the connections they formed with other parents and their ability to contribute meaningfully to research in real time. All 6 CAB members reported that the monthly meeting frequency, timing, and format were well-suited to their needs as parents. However, 2 members suggested that additional meetings could further strengthen their connections with other members.

> *[Meeting logistics are] just right. It would be nice to maybe have a meet-up that isn't HPV-related so we can bond more. I really enjoy getting to know everyone.* [CAB member 4]
>
> *I don't know that anything can be done to improve the format. I think the team has done a very good job of facilitating the meetings.*
>
> *[CAB member 1]*
>
> *I love the communication! Everyone is so nice and fun to speak with. Even when we disagree it is never an issue. Maybe we should have a group chat.* [CAB member 5]

CAB members unanimously expressed confidence in describing the project to other parents and indicated that they gained more knowledge related to HPV and the HPV vaccine since joining the advisory board.

> *I personally felt like I was knowledgeable about HPV and the vaccine but this experience helps me understand it on a more deeper and informative level.* [CAB member 4]
>
> *The presentation we had regarding the HPV vaccine helped give me a clear understanding of the history, efficacy and uses for the vaccine. I have used some of the information that I learned and have used it while talking to friends and family about the HPV vaccine.* [CAB member 3]
>
> *I think I just took for granted that I knew enough about HPV and the HPV vaccine to understand it as being one more required shot kids need. I assumed it to be a simple STD, but now appreciate the wider repercussions of not getting vaccinated against it.* [CAB member 5]

The most enjoyed aspects or activities of the CAB members were expressed as: being part of a group that is improving the lives of their children, learning about HPV vaccination, learning about other parent perspectives, and seeing the project grow, being a part of a parent-led group, contributing to scientific research, and learning about HPV more broadly.

XSL•FO
RenderX

> *That I am actually a part of something that will help my son and generations to come.* [CAB member 5]
>
> *It has been really enjoyable to watch the project grow from the beginning to where it is now. I've also enjoyed hearing everyone's different perspectives.* [CAB member 3]
>
> *Our group was able to create an authentic appreciation of the importance for work like this. It's not just another social media survey, but real scientific research.* [CAB member 4]

Despite the positive feedback, areas of improvement were identified. While all 6 members expressed comfort with sharing their input with the full project team and CAB members, 5 felt more comfortable doing so during the smaller breakout sessions. This underscores the need to create a safe and inclusive space for parents to voice diverse or unpopular opinions, including vaccine hesitancy. In addition, members emphasized the importance of recruiting more parents with diverse backgrounds and perspectives to enrich group discussion, particularly those with differing views on vaccination.

> *I do think we need new members to diversify the group and add new voices and ideas.*
>
> *[CAB member 3]*
>
> *I feel that most everyone on the team is in complete support of the vaccination and is not really open to my questioning of the vaccinations.*
>
> *[CAB member 2]*
>
> *It always takes me a little while to warm up to new groups, but a tone of openness and support was established early on that made our group a safe space!* [CAB member 6]

Discussion

Principal Findings

This study describes strategies used for forming, engaging, and maintaining an online CAB for a parent-focused HPV vaccine intervention developed for social media. We strategically leveraged social media tools to recruit parents from the same social media platform where the intervention would be delivered (ie, Twitter). To engage with the CAB, we implemented interactive workbooks to ensure that every member had opportunities to provide input and feedback on intervention content and study direction. These digital tools streamlined the collection of data and facilitated meaningful discussions despite geographic barriers. An anonymous midterm evaluation was conducted to assess the first year's outcomes and build upon successes, with 4 of the 6 CAB members continuing to advise on the project during its final second-year phase.

By using social media–based strategies to establish the CAB, we successfully increased the relevance of health materials for parents, improved the interventions reach, and enhanced its overall quality. The recommendations from the CAB significantly influenced the development and delivery of the intervention, while the online format enabled greater flexibility and accessibility for participants. To guide this process, we applied and adapted the Newman et al (2011) framework for community advisory boards, synthesizing best practices of community-based participatory research to focus on three key domains: formation, engagement, and maintenance. These domains provided an essential structure for successful operations, ensuring that the virtual CAB we formed effectively represented community interests and contributed effectively to the research process [2].

Online CAB Formation

We found the need to clearly define the unique purpose, roles, functions, and responsibilities of an online advisory board to be paramount and supported by previous work [26]. There was great utility in applying Newman's framework, which was originally designed for in-person CABs, to the development of online CAB [2]. While it can be challenging to build trust, transfer accurate knowledge, and maintain communication in in-person interactions, these challenges can be further amplified in the online environment as other researchers have also noted [19,27,28]. To address this, our approach emphasized establishing clear digital communication strategies, fostering a sense of community that grew beyond the project, and ensuring that all members felt equally involved and respected. Our study builds on these principles and details the promising approaches we encountered so that future research can replicate our processes of building a supportive environment and collaborative spirit, as online interventions and online advisory boards continue to expand.

Social media played a critical role in the recruitment and formation process of the CAB and this finding adds to the growing body of evidence supporting this important and effective tool [27]. By using Twitter's targeted advertising tools, in addition to our recruitment survey and interviews, we reached a population of parents with diverse thoughts, backgrounds, and parental experiences. We prioritized the recruitment of vaccine naïve parents who were interested in learning about the HPV vaccine. Notably, intention to vaccinate their child was not a factor in determining who was interviewed or selected for membership, allowing us to include a wide range of perspectives. This decision draws from research demonstrating the need to be intentional in building an advisory board with diverse experiences and perspectives [29]. By embracing this diverse, flexible, and forward-thinking approach, we were able to create an online CAB that was resilient, adaptable, and invested in the project's success.

Recruiting advisory board members in May 2020, at the onset of the COVID-19 pandemic, presented unique challenges to research also experienced in other studies [28]. While most of society was just beginning to adapt to changes in work and communication structures, this project had already committed to forming an online CAB to align with the online nature of the intervention. This early commitment to the online CAB and the parents' intention to participate in a child-health focused research topic, likely contributed to strong engagement with the project tasks despite the quickly changing nature of the pandemic, especially in early 2020. In addition, the digital nature of the CAB and intervention allowed the project to move forward seamlessly during periods of mandatory lockdowns

XSL•FO
RenderX

and rapidly changing guidance. This decision to prioritize an online framework also highlighted the usability and effectiveness of online CAB models in health research.

Online CAB Engagement

Flexibility was vital for effective online CAB engagement. For parents, various factors such as work, holidays, long weekends, school events, personal obligations, or religious observations can affect their ability to respond timely, attend meetings, and complete tasks. By anticipating these factors, researchers can enhance participation and sustain engagement. To accommodate these needs, we used strategies such as asynchronous activities, including digital workbooks, to collect feedback systematically and reliably. This approach builds from the successes of other studies, demonstrating the effectiveness of flexibility and asynchronous activities [30]. These workbooks allowed CAB members to provide input at their convenience and served as a reliable tool for integrating their feedback into the intervention.

During online meetings, the use of videoconferencing tools facilitated dynamic discussions and collaborative decision-making. This approach reinforced expectations of ongoing participation and capitalized on the momentum of active collaboration. Meetings were recorded and made available to those who could not attend in real time (though this was rare), further enhancing flexibility and engagement.

Engaging parents through Twitter also provided an important avenue for buy-in and oversight of research activities. Parents' familiarity with the platform facilitated a sense of ownership and connection to the intervention, which was ultimately delivered on the same social media platform. Their breadth of experiences and depth of understandings provided important grounding in the study's health messaging element and enriched its development. This approach demonstrated the ability to use social media not only as a recruitment tool but also as a means for ongoing engagement with community stakeholders.

Online CAB Maintenance

Maintaining and sustaining an online CAB required a strong emphasis on diversity, transparency, and active engagement to ensure the board's long-term effectiveness and participation. The small membership size (n=6), allowed for personalized interactions and the cultivation of strong connections among members. "Icebreakers" and informal conversations during meetings were particularly effective for building camaraderie, uncovering common interests, and strengthening the sense of community within the CAB, an approach that has demonstrated effectiveness across studies and populations [2,26,28].

Frequent meetings were instrumental in explaining project activities and tasks, the purpose tasks (ie, workbooks), and demonstrating how CAB contributions were integrated into the intervention. This "give-back" approach, fostered a strong sense of ownership and commitment among members, as they could see the tangible impact of their participation. To sustain engagement between meetings, members were encouraged to stay connected through online platforms, such as LinkedIn, or private group cats, creating additional opportunities for collaboration and relationship building.

The digital-first format also allowed us to celebrate member contributions in meaningful ways. For example, certificates of membership were shared electronically, and members received regular updates on how their feedback had influenced the intervention's development. These findings, demonstrating the effectiveness of real-time updates, adds to existing evidence to support the approach [19,26]. These practices emphasized the value of their input and strengthened their motivation to continue participating with the project.

Limitations

Our approach had several limitations worth noting. First, we began recruiting CAB members during the height of the COVID-19 pandemic, a period marked by widespread skepticism of online information and heightened vaccine-related distress. These factors may have limited the number of people interested in the project, particularly for those hesitant about engaging in a vaccine-related initiative. Furthermore, as the pandemic continued, fatigue and continued experienced by parents and the broader public may have contributed to a slight drop-off in CAB membership between years 1 and 2.

Second, while social media platforms provided effective tools for recruitment, reliance on Twitter alone may have excluded individuals or limited diversity. Expanding recruitment efforts to include additional platforms could diversify participant pools and enhance inclusivity in future studies. Furthermore, increasing the diversity of the CAB, whether in terms of age, race, gender, disability status or viewpoints, would further enrich discussions, better represent the community, and ultimately better inform the research. For instance, incorporating parents who are not digitally native or who face barriers to social media engagement could offer valuable perspectives often overlooked in online-only recruitment strategies.

We also recognized the underrepresentation of fathers on the CAB (n=1). Although Twitter's targeting tools allowed for gender-based recruitment, leveraging specific social media groups or communities tailored to fathers, such as online parenting forms or father-focused groups, may help address this imbalance. Leveraging connections to recruit more fathers could be a valuable strategy for future studies that seek to engage parents, as increased representation and diversity only broadens the range of perspectives and ensures the CAB truly reflects the community it serves. Importantly, while underrepresentation of fathers may be viewed as a limitation, our formative research found that fathers were more likely to defer to the mother regarding child health decisions, perhaps explaining their underrepresentation in this study.

While the online format offers flexibility and accessibility, some CAB members noted challenges related to ensuring that all voices, particularly those with differing opinions, were fully incorporated into the project. Online group discussions may sometimes discourage dissenting opinions due to the absence of in-person cues or the fast-paced nature of digital conversations. Using structured feedback mechanisms, such as anonymous surveys or one-on-one digital interviews, may help future online CABs better capture a broad range of perspectives and ensure that all voices are equally heard. The online model reflected how the study was developed and was essential for

XSL•FO
RenderX

navigating the constraints of the pandemic; however, it inherently posed limitations for fostering the depth of connection and collaboration that may arise in face-to-face interactions. Future studies could consider incorporating hybrid models, blending online meetings with occasional in-person events, to balance flexibility of online formats with the relational benefits of in-person collaboration.

Despite these limitations, we believe that an online CAB offers unique opportunities to engage diverse populations, facilitate inclusive participation, and provide meaningful contributions to health programs delivered online. By iterating on these approaches and addressing the highlighted limitations, future research can further refine online CAB methodologies and amplify their impact in public health initiatives.

Conclusion

This study demonstrates the feasibility and effectiveness of using online and social media–based CABs in health research. Textbox 1 shows a list of guiding principles for developing online CABs based on findings from this study and supported by previous research [26-30]. Future research should explore additional methods for integrating digital tools to further enhance engagement and maintain long-term participation in online CABs.

Online and digital health interventions have become influential tools for public health, providing innovative means to address health disparities and influence health behaviors on a broad scale. As online interventions continue to expand, so must our efforts to strategically and effectively engage online advisors within the community. Social media and online platforms serve as powerful avenues for recruiting diverse and representative CAB members, fostering active engagement in research processes, and generating quality feedback that can significantly enhance research programs. These platforms not only enable the inclusion of voices from geographically dispersed and underserved populations, but also create opportunities for dynamic, real-time collaboration that aligns with the evolving digital landscape. By leveraging these approaches, researchers can maximize the potential of online CABs as flexible, scalable, and inclusive mechanisms for participatory research. This model holds promise for driving meaningful health outcomes, reducing inequities, and fostering a deeper connection between research initiatives and the communities they aim to serve.

Textbox 1. Guiding principles for developing online community advisory boards.

Forming the online community advisory board:

- Clearly define the purpose, roles, functions, and responsibilities expected of the online community advisory board members, ensuring responsibilities align with the format and digital engagement.
- Leverage social media platforms and tools, such as targeted ads and other adjustable parameters to recruit diverse members from specific demographics.
- Prioritize diversity in geography, online behaviors, and parental perspectives by using digital methods to reach underrepresented groups and those with varied social media experience.

Engaging the online community advisory board:

- Use digital tools like interactive workbooks and online surveys to collect feedback systematically, ensuring ease of access and usability in an online environment.
- Build a sense of community and connection by facilitating collaborative discussions during video calls and leverage shared online spaces to connect members meaningfully, as whole people, despite geographic distance.
- Ensure flexibility in scheduling and participation, leveraging asynchronous tools such as recorded meetings, digital materials, and follow-up communication to accommodate members' busy online and offline schedules.

Maintaining the online community advisory board:

- Enhance inclusivity by using social media and other online platforms to recruit members from diverse online communities, ensuring broad representation across digital spaces and perspectives.
- Foster a sense of belonging by incorporating icebreakers and encouraging casual online interactions, such as LinkedIn connections or private text or chat groups to sustain engagement between meetings and outside of the project.
- Continuously demonstrate the impact of members' input by providing real-time updates through online platforms and integrating feedback into the digital intervention to maintain motivation and accountability.

Acknowledgments

This study was supported by the National Cancer Institute of the National Institutes of Health (award R01CA229324). The content was solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

XSL•FO
RenderX

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screening questions for potential community advisory board members, May 2020.
[DOCX File, 15 KB - formative_v9i1e65986_app1.docx]

Multimedia Appendix 2
Community advisory board interview guide.
[DOCX File, 16 KB - formative_v9i1e65986_app2.docx]

Multimedia Appendix 3
Sample workbook for community advisory board (CAB) members.
[PDF File, 2357 KB - formative_v9i1e65986_app3.pdf]

Multimedia Appendix 4
Parent advisory board year 1 midterm evaluation.
[DOCX File, 16 KB - formative_v9i1e65986_app4.docx]

Multimedia Appendix 5
Human papillomavirus (HPV) vaccine content for Twitter co-created by community advisory board (CAB) members and study team.
[PDF File, 20554 KB - formative_v9i1e65986_app5.pdf]

References

1. Barkin S, Schlundt D, Smith P. Community-engaged research perspectives: then and now. Acad Pediatr 2013;13(2):93-97. [doi: 10.1016/j.acap.2012.12.006] [Medline: 23498079]
2. Newman SD, Andrews JO, Magwood GS, Jenkins C, Cox MJ, Williamson DC. Community advisory boards in community-based participatory research: a synthesis of best processes. Prev Chronic Dis 2011 May;8(3):A70. [Medline: 21477510]
3. Concannon TW, Fuster M, Saunders T, et al. A systematic review of stakeholder engagement in comparative effectiveness and patient-centered outcomes research. J Gen Intern Med 2014 Dec;29(12):1692-1701. [doi: 10.1007/s11606-014-2878-x] [Medline: 24893581]
4. Connors E, Selove R, Canedo J, et al. Improving community advisory board engagement in precision medicine research to reduce health disparities. J Health Dispar Res Pract 2019;12(6):80-94. [Medline: 32832256]
5. Alexander R, Estabrooks P, Brock DJP, Hill JL, Whitt-Glover MC, Zoellner J. Capacity development and evaluation of a parent advisory team engaged in childhood obesity research. Health Promot Pract 2021 Jan;22(1):102-111. [doi: 10.1177/1524839919862251] [Medline: 31409144]
6. Hartling L, Elliott SA, Buckreus K, Leung J, Scott SD. Development and evaluation of a parent advisory group to inform a research program for knowledge translation in child health. Res Involv Engagem 2021 Jun 7;7(1):38. [doi: 10.1186/s40900-021-00280-3] [Medline: 34099069]
7. Curran JA, Bishop A, Chorney J, MacEachern L, Mackay R. Partnering with parents to advance child health research. Healthc Manage Forum 2018 Mar;31(2):45-50. [doi: 10.1177/0840470417744568] [Medline: 29400092]
8. Halladay JR, Donahue KE, Sleath B, et al. Community advisory boards guiding engaged research efforts within a clinical translational sciences award: key contextual factors explored. Prog Community Health Partnersh 2017;11(4):367-377. [doi: 10.1353/cpr.2017.0044] [Medline: 29332850]
9. McAllister CL, Green BL, Terry MA, Herman V, Mulvey L. Parents, practitioners, and researchers: community-based participatory research with early head start. Am J Public Health 2003 Oct;93(10):1672-1679. [doi: 10.2105/ajph.93.10.1672] [Medline: 14534219]
10. Brock DJP, Estabrooks PA, Hill JL, et al. Building and sustaining community capacity to address childhood obesity: a 3-year mixed-methods case study of a community-academic advisory board. Fam Community Health 2019;42(1):62-79. [doi: 10.1097/FCH.0000000000000212] [Medline: 30431470]

XSL•FO
RenderX

11. Schoeppe J, Cheadle A, Melton M, et al. The immunity community: a community engagement strategy for reducing vaccine hesitancy. Health Promot Pract 2017 Sep;18(5):654-661. [doi: 10.1177/1524839917697303] [Medline: 28398837]
12. Engster SA, Fascetti C, Mykita A, Pompa K, Reis EC. Bringing parent voices into a pediatric research network through a virtual parent panel. J Am Board Fam Med 2020;33(5):665-674. [doi: 10.3122/jabfm.2020.05.200050] [Medline: 32989061]
13. Liverpool S, Mota CP, Sales CMD, et al. Engaging children and young people in digital mental health interventions: systematic review of modes of delivery, facilitators, and barriers. J Med Internet Res 2020 Jun 23;22(6):e16317. [doi: 10.2196/16317] [Medline: 32442160]
14. Oh C, Carducci B, Vaivada T, Bhutta ZA. Digital interventions for universal health promotion in children and adolescents: a systematic review. Pediatrics 2022 May 1;149(Suppl 5):e2021053852H. [doi: 10.1542/peds.2021-053852H] [Medline: 35503335]
15. Brigden A, Anderson E, Linney C, et al. Digital behavior change interventions for younger children with chronic health conditions: systematic review. J Med Internet Res 2020 Jul 31;22(7):e16924. [doi: 10.2196/16924] [Medline: 32735227]
16. Shea L, Pesa J, Geonnotti G, Powell V, Kahn C, Peters W. Improving diversity in study participation: patient perspectives on barriers, racial differences and the role of communities. Health Expect 2022 Aug;25(4):1979-1987. [doi: 10.1111/hex.13554] [Medline: 35765232]
17. Iogansen X, Malik JK, Lee Y, Circella G. Change in work arrangement during the COVID-19 pandemic: a large shift to remote and hybrid work. Transp Res Interdiscip Persp 2024 May;25:100969. [doi: 10.1016/j.trip.2023.100969]
18. Cargo M, Mercer SL. The value and challenges of participatory research: strengthening its practice. Annu Rev Public Health 2008;29:325-350. [doi: 10.1146/annurev.publhealth.29.091307.083824] [Medline: 18173388]
19. Cramer ME, Lazoritz S, Shaffer K, Palm D, Ford AL. Community advisory board members' perspectives regarding opportunities and challenges of research collaboration. West J Nurs Res 2018 Jul;40(7):1032-1048. [doi: 10.1177/0193945917697229] [Medline: 28367677]
20. Guideline for human papillomavirus (HPV) vaccine use. American Cancer Society. 2024. URL: https://www.cancer.org/health-care-professionals/american-cancer-society-prevention-early-detection-guidelines/hpv-guidelines.html [accessed 2024-08-30]
21. Lockhart S, Dempsey AF, Pyrzanowski J, O'Leary ST, Barnard JG. Provider and parent perspectives on enhanced communication tools for human papillomavirus vaccine-hesitant parents. Acad Pediatr 2018;18(7):776-782. [doi: 10.1016/j.acap.2018.05.012] [Medline: 30177028]
22. Beavis AL, Meek K, Moran MB, Fleszar L, Adler S, Rositch AF. Exploring HPV vaccine hesitant parents' perspectives on decision-making and motivators for vaccination. Vaccine X 2022 Dec;12:100231. [doi: 10.1016/j.jvacx.2022.100231] [Medline: 36337836]
23. Petagna CN, Perez S, Hsu E, et al. Facilitators and barriers of HPV vaccination: a qualitative study in rural Georgia. Res Sq 2024 Mar 8:rs.3.rs-3979079. [doi: 10.21203/rs.3.rs-3979079/v1] [Medline: 38496559]
24. What are promoted ads? Twitter. URL: https://business.twitter.com/en/help/overview/what-are-promoted-ads.html [accessed 2023-05-31]
25. Massey PM, Chiang SC, Rose M, et al. Development of personas to communicate narrative-based information about the HPV vaccine on twitter. Front Digit Health 2021;3:682639. [doi: 10.3389/fdgth.2021.682639] [Medline: 34713151]
26. Matthews AK, Anderson EE, Willis M, Castillo A, Choure W. A community engagement advisory board as a strategy to improve research engagement and build institutional capacity for community-engaged research. J Clin Transl Sci 2018 Apr;2(2):66-72. [doi: 10.1017/cts.2018.14] [Medline: 31660220]
27. Gelinas L, Pierce R, Winkler S, Cohen IG, Lynch HF, Bierer BE. Using social media as a research recruitment tool: ethical issues and recommendations. Am J Bioeth 2017 Mar;17(3):3-14. [doi: 10.1080/15265161.2016.1276644] [Medline: 28207365]
28. Bosak J, Drainoni ML, Christopher M, et al. Community advisory board members' perspectives on their contributions to a large multistate cluster RCT: a mixed methods study. J Clin Transl Sci 2024;8(1):e1. [doi: 10.1017/cts.2023.673] [Medline: 38384918]
29. Greenwood DA, Litchman ML, Ng AH, et al. Development of the intercultural diabetes online community research council: codesign and social media processes. J Diabetes Sci Technol 2019 Mar;13(2):176-186. [doi: 10.1177/1932296818818455] [Medline: 30614252]
30. Greenwood MJ, Heninger JR. Structured communication for patient safety in emergency medical services: a legal case report. Prehosp Emerg Care 2010;14(3):345-348. [doi: 10.3109/10903121003760788] [Medline: 20438344]

Abbreviations

CAB: community advisory board
HPV: human papillomavirus

XSL•FO
RenderX

Edited by A Mavragani; submitted 31.08.24; peer-reviewed by D Carvalho, N Godwin; revised version received 10.03.25; accepted 11.03.25; published 16.04.25.

Please cite as:
Murray RM, Chiang SC, Klassen AC, Manganello JA, Leader AE, Lo WJ, Massey PM
Developing an Online Community Advisory Board (CAB) of Parents From Social Media to Co-Design an Human Papillomavirus Vaccine Intervention: Participatory Research Study
JMIR Form Res 2025;9:e65986
URL: https://formative.jmir.org/2025/1/e65986
doi:10.2196/65986

© Regan M Murray, Shawn C Chiang, Ann C Klassen, Jennifer A Manganello, Amy E Leader, Wen-Juo Lo, Philip M Massey. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Machine Learning Approach to Identifying Empathy Using the Vocals of Mental Health Helpline Counselors: Algorithm Development and Validation

Ruvini Sanjeewa[1], BSc; Ravi Iyer[1], PhD; Pragalathan Apputhurai[1], PhD; Nilmini Wickramasinghe[2], PhD; Denny Meyer[1], PhD

[1]School of Health Sciences, Swinburne University of Technology, Hawthorn, PO Box 218, John Street, Melbourne, Australia

[2]School of Computing, Engineering & Mathematical Sciences, La Trobe University, Melbourne, Australia

Corresponding Author:
Ruvini Sanjeewa, BSc
School of Health Sciences, Swinburne University of Technology, Hawthorn, PO Box 218, John Street, Melbourne, Australia

Abstract

Background: This research study aimed to detect the vocal features immersed in empathic counselor speech using samples of calls to a mental health helpline service.

Objective: This study aimed to produce an algorithm for the identification of empathy from these features, which could act as a training guide for counselors and conversational agents who need to transmit empathy in their vocals.

Methods: Two annotators with a psychology background and English heritage provided empathy ratings for 57 calls involving female counselors, as well as multiple short call segments within each of these calls. These ratings were found to be well-correlated between the 2 raters in a sample of 6 common calls. Using vocal feature extraction from call segments and statistical variable selection methods, such as L1 penalized LASSO (Least Absolute Shrinkage and Selection Operator) and forward selection, a total of 14 significant vocal features were associated with empathic speech. Generalized additive mixed models (GAMM), binary logistics regression with splines, and random forest models were used to obtain an algorithm that differentiated between high- and low-empathy call segments.

Results: The binary logistics regression model reported higher predictive accuracies of empathy (area under the curve [AUC]=0.617, 95% CI 0.613 - 0.622) compared to the GAMM (AUC=0.605, 95% CI 0.601 - 0.609) and the random forest model (AUC=0.600, 95% CI 0.595 - 0.604). This difference was statistically significant, as evidenced by the nonoverlapping 95% CIs obtained for AUC. The DeLong test further validated these results, showing a significant difference in the binary logistic model compared to the random forest (D=6.443, *df*=186283, *P*<.001) and GAMM (Z=5.846, *P*<.001). These findings confirm that the binary logistic regression model outperforms the other 2 models concerning predictive accuracy for empathy classification.

Conclusions: This study suggests that the identification of empathy from vocal features alone is challenging, and further research involving multimodal models (eg, models incorporating facial expression, words used, and vocal features) are encouraged for detecting empathy in the future. This study has several limitations, including a relatively small sample of calls and only 2 empathy raters. Future research should focus on accommodating multiple raters with varied backgrounds to explore these effects on perceptions of empathy. Additionally, considering counselor vocals from larger, more heterogeneous populations, including mixed-gender samples, will allow an exploration of the factors influencing the level of empathy projected in counselor voices more generally.

(JMIR Form Res 2025;9:e67835) doi:10.2196/67835

KEYWORDS

vocal features; voice characteristics; empathy; mental health care; crisis helpline service

Introduction

Empathy is defined as experiencing the emotions (emotional empathy) and cognitions (cognitive empathy) of others and responding to them appropriately [1]. Empathy is especially important for patient care, where the lived experience of the patient must be understood by responding health care professionals, while also conveying this understanding in conjunction with a desire to help the patient [2]. The effectiveness of physician empathy has been shown to improve patient satisfaction and commitment to recovery while reducing anxiety and distress levels, leading to better clinical results [3]. Furthermore, empathic behavior by mental health (MH) care providers reduces their own risk of burnout [4].

XSL•FO
RenderX

Telephone helpline services offer an effective means of supporting those who need immediate MH care [5]. The demand for such services has increased dramatically since the outbreak of the COVID-19 pandemic [6], increasing the expectations of counseling staff to provide support for people with complex MH concerns [7]. As a basic counseling skill, empathy is key to successful engagement with patients in the context of complex psychosocial needs.

Besides emotional and cognitive empathy in understanding the status of a patient, contextual awareness is equally important for therapeutic engagement [8]. This means that empathic responses need to be contextually appropriate by considering environmental cues, culture, demographic factors, and the specific circumstances of the patient to understand the broader context of their MH status [9,10]. This allows counselors to tailor responses, based on context, to engage in effective communication with distressed patients, thereby delivering better outcomes [3].

Verbal cues and tone of voice are crucial when communicating empathy [11]. For example, reduced speech rate and lower pitch are perceived as more empathic by patients when receiving bad news from health care providers in an oncology setting [12] and while actively listening to telephone callers, nurses have been found to express empathy through their choice of words, voice and intonation, projection of compassion and warmth, as well as “tuning in” to the caller’s story and identifying with the caller’s emotions [13].

Unfortunately, the global demand for MH support is not being met by the existing workforce [14]. This service gap is leading to growing interest in alternative digital technological solutions. Technological innovations, such as the design of conversational agents, have demonstrated potential in facilitating effective and immediate patient care. However, to optimize upon end user acceptance, conversational agents need to display empathy [15,16]. However, we have yet to identify the precise vocal features most associated with an empathic human response, and it is also not known if it is possible to categorize empathy levels using such vocal features. Thus, the aim of this paper is to (1) identify the vocal features significantly associated with empathy in a large collection of telephone helpline counseling call recordings; and to (2) evaluate the accuracy of a machine learning algorithm to correctly designate short segments of each recording to categories of low and high empathy.

This study has been reported in accordance with the TRIPOD+AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis+Artificial Intelligence) checklist as shown in Checklist 1.

Methods

Data Collection

Recordings of telephone helpline calls (n=57) were obtained from On The Line, Australia, a suicide helpline counseling service. Participants were counselors for a Suicide Call-Back Service (a national helpline service coordinated by On The Line, Australia). Calls were randomly sampled from July 1, 2019, to June 30, 2021, stratified by organizationally determined suicide risk level (high or low). The level of suicide risk of each caller had been previously assessed by counselors using the Columbia Suicide Severity Rating Scale (C-SSRS) to differentiate between calls featuring high suicide risk (with C-SSRS ratings of 6‐7) and calls with low risk of suicide (with C-SSRS ratings of 1‐2; please refer to Iyer et al [17] for further details [18]). Only the counselors’ voice recordings were used in this study. No information was provided that could be used to identify the callers or counselors for any of these calls.

Ethical Considerations

This study was approved by the Swinburne Human Research Ethics Committee (Ref:20226835‐11907). The application was given a waiver of consent from the ethics committee for the use of nonidentifiable secondary research data. However, the research team also signed a confidentiality agreement that restricted the discussion of the contents of the recordings only within the research team. The secondary data and annotation results were saved on OneDrive for Business (Microsoft Corp) and were only accessible to the research team. No compensation was paid to participants in this study because they could not be identified by the research team.

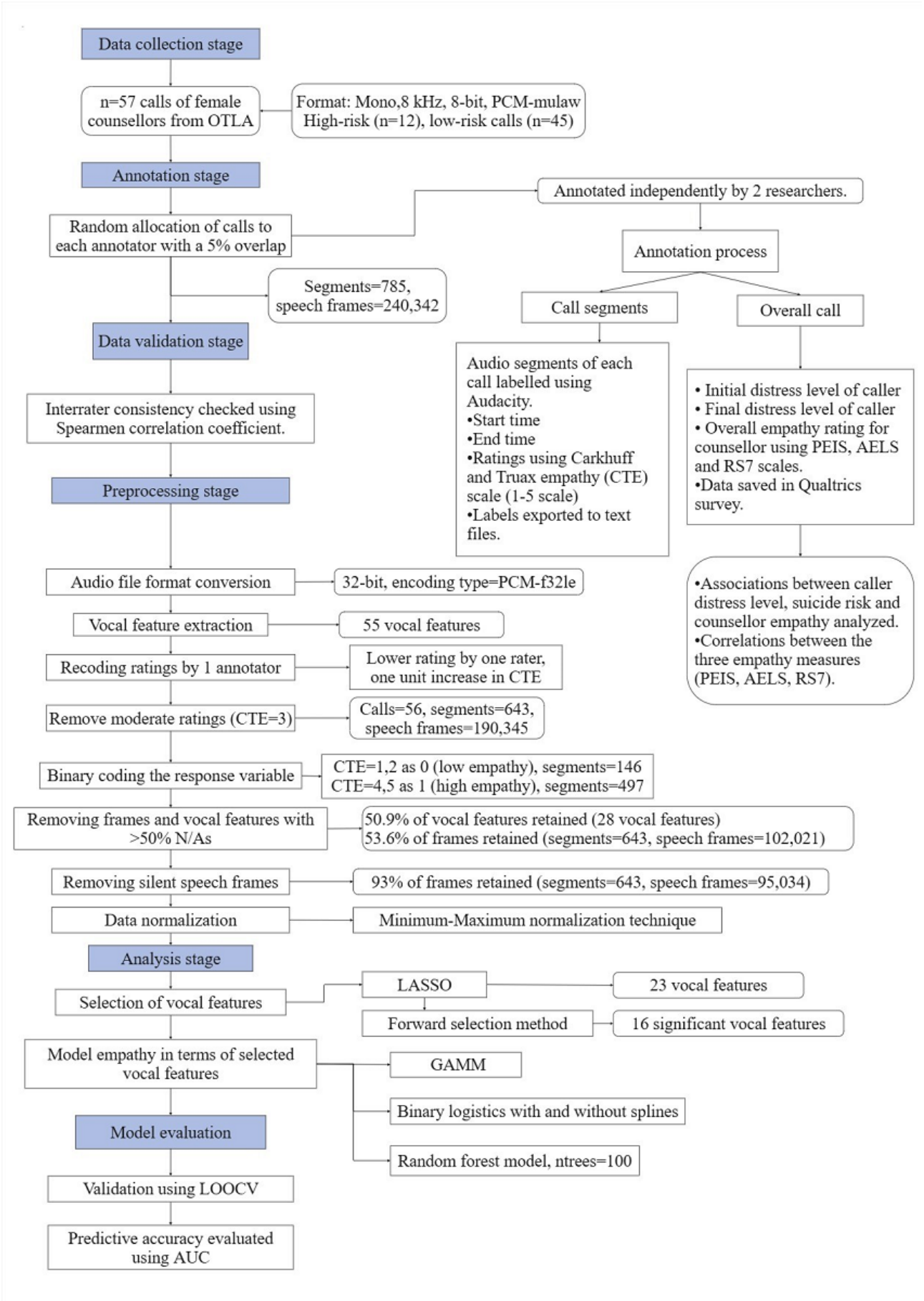
Annotation of Call Segments and Overall Call Empathy for Counselors

Annotations of counselor empathy were conducted by 2 independent researchers (Inge Gnatt and Sarah Dunning) using RStudio (version 2024.04.2, build 764; R Foundation) [19]. The call annotators Inge Gnatt and Sarah Dunning were recruited via research team networks. Both had extensive experience working as counselors for a MH helpline service. Inge Gnatt had experience working as an annotator on a similar project. Segments of the counselor voices were selected from each call using Audacity (version 3.5.1, CMake Release Build; Muse Group & contributors) [20], ensuring that overlaps between the caller and responding counselor voices were minimized. Empathy displayed by the counselor within each call segment was rated using the Carkhuff and Truax Empathy (CTE) scale [21]. A weekly project team meeting, attended by a clinical psychologist (Maja Nedeljkovic), was used to reconcile any disparities in ratings. A Qualtrics web-based questionnaire was used to also collect data on the overall level of empathy displayed by the responding counselor during each call and to evaluate caller distress at the commencement and conclusion of each call. Figure 1 shows the flow of the voice analysis process during this study.

XSL•FO
RenderX

Figure 1. Overview of the voice analysis process for classifying empathy in counselor voices. AELS: Active-Empathic Listening Scale; AUC: area under the curve; CTE: Carkhuff and Truax Empathy; LASSO: Least Absolute Shrinkage and Selection Operator; LOOCV: Leave One Out Cross Validation; GAMM: generalized additive mixed model; N/A: not available; OTLA: On the Line Australia; PCM-f321e: pulse code modulation float; PEIS: Perceived Emotional Intelligence Scale; RS7: Rating Scale 7 (7-item).

XSL•FO
RenderX



XSL•FO
RenderX

Measures of Counselor Empathy and Caller Distress

The CTE scale was used to rate the audio segments selected from each call on a 5-point Likert-style scale (1="low empathy" to 5="high empathy") [21]. Additionally, 3 measures were used to assess overall counselor empathy for each call. These 3 scales were modified to suit counselor-caller conversations through an iterative process in which members of the research team provided independent feedback to achieve the final questionnaire. Examples were developed by the annotators for each item included in these scales, ensuring clarity and consistency of ratings. The details for these scales are provided in Multimedia Appendix 1. The measurement scales included:

1. The Perceived Emotional Intelligence (PEI) Scale [22] identified variations in PEI in the counselor's vocals. The PEI is a 20-item scale with each item scored with 1="never or almost never true" and 7="almost or almost always true."
2. The Active-Empathic Listening (AEL) Scale [23] was modified appropriately to produce 10 items measuring empathic listening using a 7-point Likert-style scale with 1="never or almost never true" and 7="almost or almost always true."
3. Rating Scale 7 (RS7): A single item 7-point Likert Scale [24] was used to rate overall empathy with 1="low empathy" and 7="high empathy."

Finally, at the start and end of each call the annotators assessed the level of caller distress using the distress thermometer [25], a visual analogue 11-point scale (0="no distress" to 10="extreme distress").

Data Validation Through Interrater Consistency Check

Six (10%) calls were chosen at random to measure interrater reliability. The empathy ratings of the more experienced rater (Inge Gnatt) were used as the reference, against which the SD ratings were compared. Spearman correlation [26] was calculated for each of the 3 scales used to rate overall counselor empathy for each call. The Mann-Whitney *U* test was then used to check for the significance of differences between the ratings provided by the annotators.

Relationships Between Perceptions of Empathy and Call Context

The associations between perceived counselor empathy and call context were explored using a combination of empathy ratings, caller distress at the beginning and end of the call, and caller suicide risk. Caller distress and suicide risk were correlated with perceived counselor empathy to evaluate the relationship between level of empathy and caller disposition.

Preprocessing Stage: Audio File Format Conversion and Vocal Features Extraction

The input call recordings were obtained as 8 kHz sample rate, 8-bit depth .wav files. The encoding type of the files were transformed to PCM float format with 32-bit depth to ensure compatibility with RStudio for analysis. Vocal features (n=55) were extracted per 30-millisecond speech frames (50% overlap; Blackman windows) within each annotated segment using RStudio (version 2.7.0, *Soundgen* package [27]).

Removing Moderate Ratings and Binary Coding Empathy Level

Overview

The vocal segments that scored a rating of 3 out of 5 on the CTE scale were removed (n=142, 18%) from further analysis because of their neutral empathic character. A binary response variable was then created for each of the 643 remaining segments (190,345 speech frames of 30 ms) with an empathy rating of 4 - 5 coded as high empathy (n=146 segments) and a rating of 1 - 2 coded as low empathy (n=497 segments).

Removing Missing Values in Speech Frames and Vocal Features

Vocal features and 30-millisecond speech frames with more than 50% missing values were removed to maintain the quality of data and improve the overall accuracy of the results. The resulting data retained 50.9% (n=28) of the original vocal features and 53.6% (n=102,021) of the original speech frames.

Removing Silent Speech Frames and Normalization

The silent speech frames were also removed from each of the 643 segments leaving 95,034 speech frames in the final analysis sample. Finally, the minimum and maximum normalizing technique [28] was applied to reduce the influence of background noise.

Analysis Stage

Selection of Vocal Features

Variable selection was performed to identify vocal features that were strongly associated with empathy. L1 penalized LASSO (Least Absolute Shrinkage and Selection Operator) regression was used to select the most relevant variables by shrinking the coefficients of the least relevant variables to 0 [29]. Tenfold cross-validation was used to optimize the tuning parameter, lambda. Further refinement per the selected vocal features was then conducted using a forward stepwise regression model.

Models for Identifying Empathy Level With Selected Vocal Features

Three methods were used to classify low and high empathy segments based on this final set of selected vocal features. A generalized additive mixed model (GAMM) included vocal features as fixed effects, with each call treated as a random effect. Spline functions for the selected vocal features were used to account for nonlinearity [30]. The GAM function [31] of package *mgcv* [32] in RStudio was used for the analysis.

Random forest classification also allowed for nonlinear relationships using step functions while more efficiently processing large datasets [33]. This has been a prominent classification model used in studies involving vocal analysis [34,35]. The binary logistic regression model was the third model considered, again accounting for nonlinearity using splines, and was used in our study as the baseline model [36].

Model Evaluation

Probabilistic predictions for high versus low empathy levels were obtained for each segment using Leave One Caller Out

XSL•FO
RenderX

Cross Validation. Based on these probabilities, receiver operating characteristic curves were created, and areas under the curves (AUCs) [37,38] were used to compare the reliability of these models. The Youden index [39] was used to decide the optimal cut point for classifying segments based on their estimated high versus low empathy probabilities.

Results

Overview

The following results were obtained from the 57 calls for female counselors. This sample of calls included 12 (12/57, 21%) calls at high risk of suicide and 45 (45/57,79%) calls at low risk of suicide.

Annotation of Call Segments

Using the CTE scale, 146 (18.6%) segments showed low empathy, 142 (18.1%) showed medium empathy, and 497 (63.3%) segments demonstrated high empathy.

Overall Call Empathy Ratings

Multimedia Appendix 2 provides descriptive statistics for the overall empathy ratings for the 57 calls using the 3 scales. Excellent reliability is observed from both the raters, Inge Gnatt and Sarah Dunning, for the PEI and AEL scales, with Cronbach α values above 0.9. The descriptive statistics and the Spearman correlation statistics between the annotators are shown in Multimedia Appendix 3. A strong agreement between raters was observed with the PEI measure and the RS7 empathy measure approaching statistical significance.

Relationships Between Perceptions of Empathy and Call Context

The relationship of counselor empathy ratings with caller's distress at the start and end of the call and suicide risk are shown in Table 1. While the correlations between the initial distress and the 3 empathy measures were not significant, a moderate, statistically significant negative correlation between the final distress of the caller and the empathy of the counselor was observed for both the PEI and RS7 (P<.01 and P<.001, respectively). The suicide risk of callers as measured using the C-SSRS, had a statistically significant but weak positive correlation with counselor empathy across the PEI and AEL measures. Strong statistically significant correlations among the 3 empathy measures were found, validating the empathy measurement process.

Table . Spearman correlation coefficients for caller distress and suicide risk with counselor empathy ratings[a].

	Caller context			Ratings for counselor empathy		
Empathy rating	Initial distress	Final distress	Suicide risk	PEIS[b]	AELS[c]	RS7[d]
PEIS	–0.011	–0.414**	0.310*	1		
AELS	0.102	–0.233	0.308*	.822***	1	
RS7	–0.024	–0.443***	0.055	.847***	.851***	1

[a]***P<.001 (2-tailed), **P<.01 (2-tailed), and *P<.05 (2-tailed).

[b]PEIS: Perceived Emotional Intelligence Scale.

[c]AELS: Active-Empathic Listening Scale.

[d]RS7: Rating Scale 7 (7-item).

Analysis Stage

The selection of vocal features using LASSO to predict high versus low empathy used a cross-validation of the training dataset to reveal an optimum Log (Lambda) parameter=–8.542. The relationship between this parameter and the binomial deviance is shown in Multimedia Appendix 4. Using this lambda value, 23 vocal features were retained. These 23 vocal features were then passed on to the forward selection binary logistic regression model, which identified 16 significant vocal features as shown in Table 2.

XSL•FO
RenderX

Table . Results of the forward binary logistic regression model for vocal feature selection for identifying high versus low empathy in counselor voices.

Vocal features	Coefficient	*Z* value	*P* value
Depth of amplitude	–1.468	–15.994	<.001
Frequency of amplitude (Hz)	0.068	2.524	.01
Frequency of amplitude (Hz) via MS[a]	–0.462	–12.350	<.001
Purity of amplitude via MS	–0.834	–9.596	<.001
Amplitude (dB)	–3.391	–51.688	<.001
Dominant frequency (Hz)	0.985	3.136	.002
Entropy	3.47	20.843	<.001
Shannon entropy	–6.016	–25.27	<.001
Epoch	0.289	5.04	<.001
First formant frequency (Hz)	1.244	5.08	<.001
First formant width (Hz)	0.194	4.181	<.001
Second formant frequency (Hz)	–0.191	–2.808	.005
Second formant width (Hz)	–0.063	–1.802	.07
Third formant frequency (Hz)	–0.099	–1.521	.13
Third formant width (Hz)	0.004	0.118	.91
Spectral flux	0.11	1.435	.15
HNR[b] (dB)	–0.37	2.869	.004
Spectral novelty	0.035	0.715	.47
Peak frequency (Hz)	–0.293	–1.827	.06
25th percentile frequency (Hz)	–1.311	–4.720	<.001
50th percentile frequency (Hz)	–0.262	–1.586	.11
Spectral centroid (Hz)	5.782	15.717	<.001
Spectral slope (Hz)	–3.944	–17.881	<.001

[a]MS: modulation spectrum.

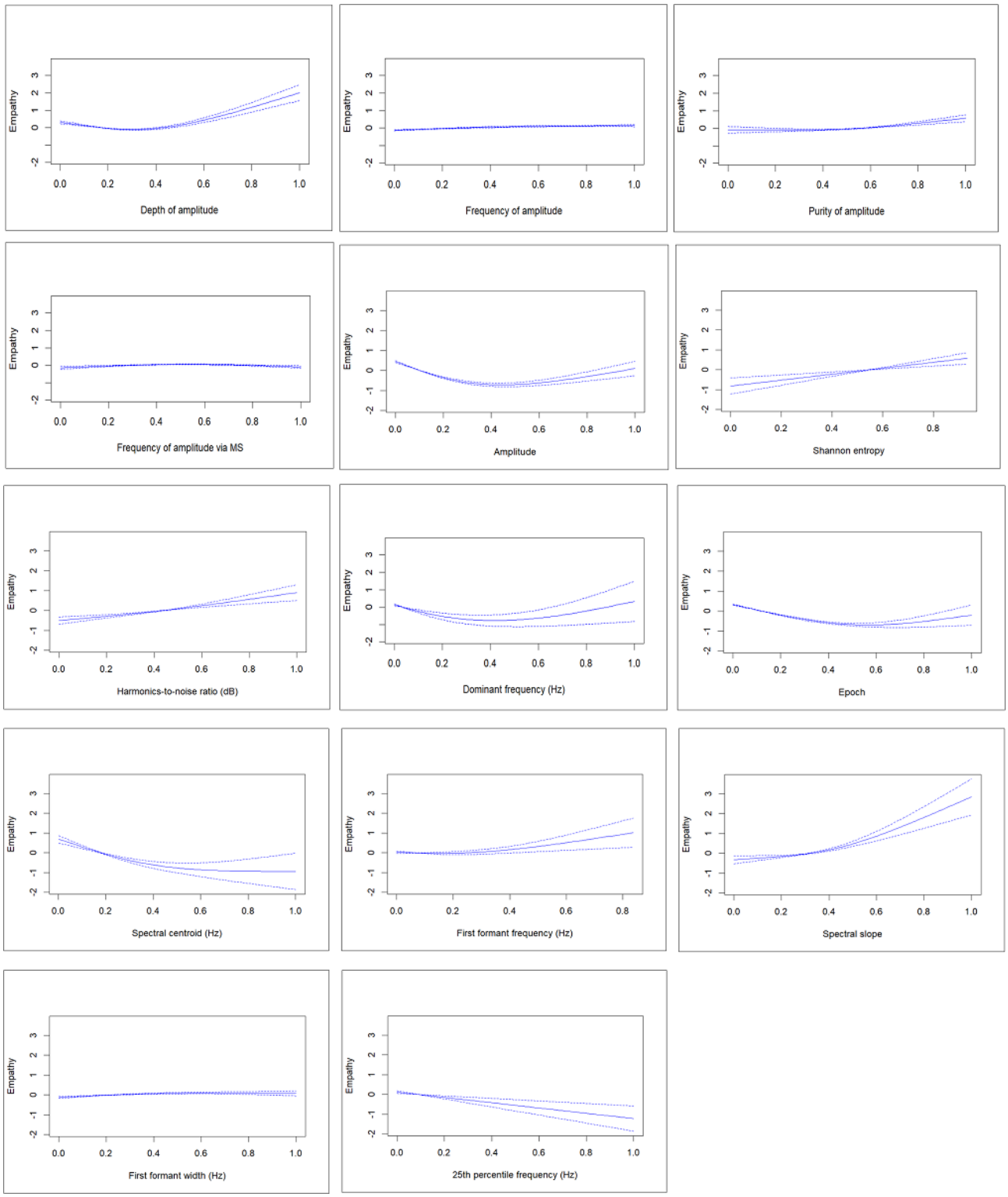
[b]HNR: harmonics-to-noise ratio.

Feature Extraction and Classification

The results of the GAMM are shown in Multimedia Appendix 5. Of the 16 selected vocal features, the GAMM used only 14. Based on the effective *df* values, 2 of the vocal features (Shannon entropy and the 25th percentile frequency) show a linear relationship in the GAMM. Figure 2 illustrates the nonlinear nature of all other relationships. An AUC value of 0.605 was obtained for the GAMM [40].

XSL•FO
RenderX

Figure 2. The smoothed relationship between the selected vocal features and the standardized predicted level of empathy displayed in counselor speech with 95% CIs. MS: modulation spectrum.



The binary logistic classification model showed higher AUC values and nonoverlapping CIs at 95%. The DeLong test confirmed that the binary logistic model outperformed both the random forest (D=6.443, *df*=186283, *P*<.001) and GAMM (Z=5.846, *P*<.001) models [41]. However, this model achieved a lower classification accuracy than the other 2 methods, as shown in Table 3, suggesting that the probability cut point used for classification purposes was not ideal.

XSL•FO
RenderX

Table . Performance comparison of the 3 classification models for identifying high versus low counselor empathy.

Classification model	Accuracy (%)	AUC[a] value	95% CIs
GAMM[b]	75	0.605	0.601 - 0.609
Random forest	74	0.6	0.595 - 0.604
Binary logistic regression	69	0.617	0.613 - 0.622

[a]AUC: area under the curve.

[b]GAMM: generalized additive mixed model.

As illustrated in Figure 2, higher empathy was associated with higher values for the first formant frequency, dominant frequency, Shannon entropy, spectral slope, and harmonic-to-noise ratio. In contrast, higher empathy was associated with lower values for the 25th percentile frequency and spectral centroid. Finally, lower empathy was associated with intermediate values for depth of amplitude, amplitude, dominant frequency, epoch, and first formant frequency of speech.

The GAMM was able to differentiate between low and high 30-millisecond segments of speech to a classification accuracy of 75%. This was superior to both random forest and binary logistic regression models (74% and 69%, respectively).

Epoch (39%), amplitude (22%), and depth of amplitude (7%) were the top 3 vocal features contributing to empathic speech in the GAMM. The vocal features that contribute the most toward the identification of empathic speech vary across the 3 methods used, as shown in Multimedia Appendix 5.

For purposes of validation, the voice algorithm was also evaluated on a synthesized dataset of female voices created from a text-to-speech application. This approach yielded significant differences between high and low empathic voices in the validation dataset for the GAMM, random forest, and binary logistic regression classification models.

Discussion

Principal Findings

This study was undertaken to identify a range of vocal features that can predict the level of empathy exhibited in the recordings of female counselors and to accurately classify short segments of each recording according to low or high empathy ratings. We were successful in identifying 14 unique vocal features that significantly distinguished between low and high empathy ratings in the GAMM. Furthermore, we were able to successfully classify short segments of speech using these vocal features to an accuracy level of 75% using this model. Although this study considered only 57 calls, each call featured multiple segments annotated for the level of empathy, and each of these segments was further divided into 30-millisecond speech frames that were the observational points considered during the modeling process. Reducing the number of vocal features in the model led to lower AUC values, suggesting that the original model featuring 14 vocal features was not overfitted. Furthermore, the 95% AUC CIs for the binary logistic regression did not overlap those for the other 2 methods, confirming that the binary logistic regression model produced superior results.

Empathy is an important component of human interactions and social connections that promotes general well-being [42]. It is an essential component of MH care support, required to enhance therapeutic alliance and rapport building. Empathy embodies the ability to understand and compassionately reflect the range of feelings and experiences communicated by others. Empathic communication relies upon verbal (words and vocal features) and nonverbal means such as body language and facial expression [43]. However, it is only verbal expressions of empathy through vocal features that are the subject of this study.

The detection of empathy is traditionally based upon subjective human perception captured using standardized empathy scales or questionnaires [44]. Research on detecting empathic speech through voice feature extraction has been the focus of at least 2 recent studies, both showing similar accuracies to what we have found. For example, the first study by Chen et al [45] analyzed the acoustic prosodic features of speech recorded in YouTube (Google LLC) videos used for empathy training purposes. The use of these features resulted in classification accuracies of 59% (F_1-score) when differentiating between empathy and neutral categories. A second study by Alam et al [46] explored acoustic and lexical features of empathic speech using annotations sourced from Italian call center conversations. In this study, classification using support vector machines yielded accuracies of 68.1%.

However, our study is unique in identifying a range of vocal features that identify the level of empathy using an ecologically valid dataset of counselor-caller conversations obtained from On the Line Australia, an Australian helpline service. Machine learning techniques are involved in both vocal feature selection and empathy classification. The 2 annotators chosen for empathy labeling purposes had a psychological background to strengthen the validity of this process. This led to the development of an algorithm that detects the human vocal features that are associated with empathic speech. This algorithm has the potential to enhance the training of counselors in the use of empathic speech and offers valuable insights into effective human communication in the MH care domain.

Call Context and Counselor Empathy in a Crisis Helpline Setting

The analysis revealed that there is a strong negative correlation between the final distress level of the caller and the counselor's level of empathy, suggesting that empathic communication with a caller can lower their level of distress by the end of the call. Higher levels of empathy allow the counselors to build rapport and trust with their patients, allowing for effective emotional support during a crisis. These findings align with the existing

XSL•FO
RenderX

literature about the benefits of empathic interactions. The suicide risk of the caller was positively related to counselor empathy. This is an indication that counselors effectively recognize crisis situations, exhibiting higher levels of empathy when speaking to callers with high risks of suicide. It also confirms that the level of empathy displayed by counselors is adapted to the situation of the caller. These relationships of counselor empathy with caller distress and suicide risk confirm the importance of counselor empathy in the context of crisis helpline services.

Characteristics of Empathic Vocals

This study has identified several vocal features associated with empathy. The depth of amplitude in speech reflects varying levels of loudness, emphasizing the expressiveness and dynamic nature of the human vocals apparent in empathic speech. A stable, more consistent emotional delivery during speech (purity of amplitude) also helps to convey empathy. Quieter vocals (amplitude and lower tonal frequency or dominant frequency) are also associated with higher empathy.

Higher first formant frequencies are associated with “a” vowel sound, which corresponds with high ratings of empathy. Additionally, a higher harmonic-to-noise ratio, indicating greater clarity and more pleasant-sounding vocals, is also associated with greater empathy. The spectral slope has a strong positive relationship to empathy, while the spectral centroid shows the opposite relationship. This indicates that a lower spectral centroid, with more low-frequency components, makes a speaker sound more empathic.

Based on these findings, it is evident that empathy in vocals is provided by a combination of multiple human vocal features, and variations in each of the features exert a different impact on empathy. In particular, the way that a specific threshold of loudness in the vocals decides the delivery of perceived empathy in the context of effective counseling provides compelling evidence. This further suggests that the right balance of each of these vocal features is needed, where stability, energy, and clarity play a pivotal role.

However, the study of empathy in vocals is a complex topic and has challenges. Especially the subjective nature of empathy perception is an area that requires further study. This study relied on empathy ratings provided by 2 psychology-trained female raters of English heritage. Different results may have been obtained if raters with different cultural, social, and educational backgrounds had been included [47].

Vocal Feature Extraction and Empathy Classification

Three methods were used to study the association between empathy and relevant vocal features. A GAMM, a random forest, and a logistic regression approach with splines were fitted and compared using AUC values and using the Leave One Caller Out Cross Validation method for evaluating these classifiers. The accuracy of the empathy level classifications achieved was similar (75%, 74%, and 69%, respectively) when Youden index was used to choose the probability cut point, as were their AUC values (0.605, 0.600, and 0.617, respectively). The GAMM and binary logistic regression with splines required significantly more computational time compared to the random forest, which used 100 trees. Despite the significant improvement in AUC value with the logistic regression with splines in comparison to the random forest model, the AUC results do not seem to be dissimilar. However, the AUC 95% CIs for the binary logistic regression do not overlap with those for the other 2 methods, confirming that the logistic regression model provides a better fit to the data. This was further validated by the DeLong test results showing significant differences in AUC values of the binary logistic model compared to the other 2 methods. The partial overlap between the AUC CIs of the random forest and binary GAMMs indicates a significant difference in model fit. Therefore, the binary logistic classification model outperforms the other 2 methods in its ability to distinguish high-empathic speech from low-empathic speech. However, the slightly lower classification accuracy (69%) for this model suggests that the Youden method used to determine the probability cut point used for classification purposes may not be optimal.

These relatively low AUC values, also seen in other related research projects, can perhaps be partly attributed to the difficulties encountered in providing accurate ratings of empathy. The algorithms developed for detecting empathy from vocal features were reliant on the quality of the input data provided for the empathy ratings. A larger sample of raters and a larger sample of calls might have produced more reliable data, and this is recommended for future research in this area.

However, the multimodal approaches commonly used for empathy recognition through the use of words, vocals, visual signs, and psychological signals reflect the multifaceted nature of empathy [48]. The importance of facial expressions for recognizing empathy is particularly emphasized in this literature, where factors such as observation time and the type of emotion expressed significantly influence the accuracy of identification [49]. These multimodal approaches suggest that a higher accuracy in detecting empathy can be achieved when all these factors are collectively considered, rather than vocal approaches on their own.

Limitations

The inherent differences of empathy perception among the annotators of this study were a concern in this study. An additional analysis was conducted to explore this further, incorporating a third annotator without psychological expertise and from a different cultural background (Multimedia Appendix 6). The findings from this analysis produced an even lower model performance, confirming that perceptions of empathy do vary between individuals and cultures. This analysis included both female and nonfemale counselor vocals, which may also have contributed to this poorer performance. While this finding underscores the complexity of recognizing empathy, it also highlights how cultural differences, personal experience, and psychological knowledge of individuals contribute to subjective perceptions. Therefore, future research would benefit from accommodating these differences in empathy perception within the model by including multiple annotators with varied backgrounds.

Another limitation of this study was due to the very small number of available male counselor recordings (n=13). The main analysis could therefore only be conducted on female

XSL•FO
RenderX

vocals, limiting the generalization of our findings. Ideally it would have been possible to develop separate models for empathy in male and female counselors and to test whether there were significant differences between these models. Larger samples of counselor vocals would also have been preferable, providing more diversity in the data. Unfortunately, only a very small sample of male counselor calls was obtained. A recent study that focused on the vocal characteristics of distressed adults using machine learning techniques identified a significant difference between male and female vocal behaviors. This suggested that it would not be appropriate to use a single model to describe the voices of both male and female counselors [50]. For this reason, only female counselor voices were considered in this study. Recent research on machine learning applications has explored synthesized data as an avenue to address class imbalance, and this was considered as a possible remedy to address the lower representation of male voices in this study [51]. However, using synthetic speech, such as through text-to-speech applications, faces the challenge of reflecting the diversity inherent in natural human voices [52]. So, this option was not considered for boosting the number of male counselor calls, so as to make an analysis of these data possible.

However, this option was used when the algorithm developed in this study was validated on an external synthesized dataset consisting of female recordings generated using a text-to-speech application. These data also showed successful outcomes for the algorithm in differentiating between high and low empathic voices in female voices. These results highlight synthetic voice augmentation as a promising future research direction in machine learning applications. Alternatively, a balanced representation of gender, as well as cultural background, empathy levels, and individual characteristics, are necessary considerations for the counselor recordings used in future research of this nature.

A further limitation is the source of the calls used for this study. The context of a suicide helpline service is very specific, and it may be that more algorithmic success would have been possible in a less stressed environment. Additionally, for most of the calls, the level of empathy was assessed by a single rater. As mentioned above, it would have been preferable if a greater duplication of ratings could have been used to provide the dependent variable for the models that have been used to identify the level of empathy in counselor vocals. However, this pilot study, using machine learning techniques to identify vocal features related to empathy in female counselor voices, has shown promise. Therefore, further explorations of this approach with increases in the call sample size and a more balanced gender representation, and with more annotators to allow more duplication of ratings and more variation in annotator background while also including calls from more than 1 helpline service, will benefit future research on this topic.

Implications of This Study

The importance of empathy in reducing the distress of callers confirms the need for the incorporation of empathic communication skills in training programs for counselors. Additionally, a statistically significant positive correlation was found between suicide risk and counselor empathy ($P<.05$). This suggests that counselors tend to be more empathic toward high-risk callers. This perhaps highlights the need for counselors to adhere to a more caller-centered approach, ensuring that empathy is consistently exhibited for both low-risk and high-risk callers. Resources should perhaps be allocated equally for all callers rather than having a crisis intervention strategy that is tailored to prioritize callers needing emergency support.

To the best of the authors' knowledge, this is the first study that has identified the unique features of human vocals that are associated with the communication of empathy in a MH care setting. The results of this study have implications for the training of counselors and psychologists working for MH-related telephone helpline services. Additionally, these findings can serve as a training resource for MH professionals more broadly, enhancing the quality of care provided. The engineering of empathic chatbots, especially within a triage capacity, is another significant area of research that would benefit from the findings of this study [53].

However, collecting the vocal data of individuals for research purposes raises important ethical concerns. This research needs to prioritize user consent and caller privacy. It is recommended that people with lived experience in telephone counseling and MH be asked to assist with the co-design and coproduction of such research to ensure that any resulting training programs or monitoring systems are acceptable to users and meet consumer needs.

Acknowledgments

This research was funded by Swinburne University of Technology. The data call recordings were provided by On The Line Australia who was the main research partner of this study. We acknowledge Dr Jakqui Barnfield for her support from On The Line, Australia. We also thank Inge Gnatt, BPsych (Hons), and Sarah Dunning, BPsych, for their contributions as research annotators and Dr Maja Nedeljkovic for her supervision of the annotation process.

Data Availability

All data analyzed in this study are included in summary form under Multimedia Appendices. Our ethics approval does not allow the reporting of individual call data.

XSL•FO
RenderX

Authors' Contributions

RS developed the first draft of this study's manuscript and conducted a voice analysis, including the creation of the algorithm. RI contributed to developing the algorithm for detecting voice features that relate to empathy. RS contributed to the call annotation process, where variations in counselor empathy within the calls were identified. DM and RI contributed their supervision during the entire annotation process, confirming the well-being of the researchers involved. DM, RI, PA, and NW were involved in this study's manuscript revisions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Online questionnaires used for data collection.
[DOCX File, 30 KB - formative_v9i1e67835_app1.docx]

Multimedia Appendix 2
Descriptive statistics of the analysis.
[DOCX File, 16 KB - formative_v9i1e67835_app2.docx]

Multimedia Appendix 3
Results of consistency between the raters.
[DOCX File, 16 KB - formative_v9i1e67835_app3.docx]

Multimedia Appendix 4
The results of LASSO regression. LASSO: Least Absolute Shrinkage and Selection Operator.
[DOCX File, 23 KB - formative_v9i1e67835_app4.docx]

Multimedia Appendix 5
Detailed results of the significance of the vocal features through GAMM, random forest, and binary logistics regression models. GAMM: Generalized Additive Mixed Model.
[DOCX File, 28 KB - formative_v9i1e67835_app5.docx]

Multimedia Appendix 6
Impact of expanded dataset and rater diversity on voice analysis results.
[DOCX File, 29 KB - formative_v9i1e67835_app6.docx]

Checklist 1
TRIPOD+AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis+Artificial Intelligence) statement for reporting the modeling process.
[PDF File, 620 KB - formative_v9i1e67835_app7.pdf]

References

1. Smith A. Cognitive empathy and emotional empathy in human behavior and evolution. Psychol Rec 2006 Jan;56(1):3-21. [doi: 10.1007/BF03395534]
2. Hojat M. Empathy in Health Professions Education and Patient Care: Springer Cham; 2016.
3. Derksen F, Bensing J, Lagro-Janssen A. Effectiveness of empathy in general practice: a systematic review. Br J Gen Pract 2013 Jan;63(606):e76-e84. [doi: 10.3399/bjgp13X660814] [Medline: 23336477]
4. Sturzu L, Lala A, Bisch M, Guitter M, Dobre D, Schwan R. Empathy and burnout - a cross-sectional study among mental healthcare providers in France. J Med Life 2019;12(1):21-29. [doi: 10.25122/jml-2018-0050] [Medline: 31123521]
5. Peppou LE, Economou M, Skali T, Papageorgiou C. From economic crisis to the COVID-19 pandemic crisis: evidence from a mental health helpline in Greece. Eur Arch Psychiatry Clin Neurosci 2021 Mar;271(2):407-409. [doi: 10.1007/s00406-020-01165-4] [Medline: 32666279]
6. Betancourt JA, Rosenberg MA, Zevallos A, Brown JR, Mileski M. The impact of COVID-19 on telemedicine utilization across multiple service lines in the United States. Healthcare (Basel) 2020 Oct 1;8(4):380. [doi: 10.3390/healthcare8040380] [Medline: 33019667]

XSL•FO
RenderX

7. Pavlova A, Scarth B, Witt K, Hetrick S, Fortune S. COVID-19 related innovation in Aotearoa/New Zealand mental health helplines and telehealth providers - mapping solutions and discussing sustainability from the perspective of service providers. Front Psychiatry 2022;13:973261. [doi: 10.3389/fpsyt.2022.973261] [Medline: 36111308]
8. Gibbons SB. Understanding empathy as a complex construct: a review of the literature. Clin Soc Work J 2011 Sep;39(3):243-252. [doi: 10.1007/s10615-010-0305-2]
9. Angus L, Kagan F. Empathic relational bonds and personal agency in psychotherapy: implications for psychotherapy supervision, practice, and research. Psychotherapy (Chic) 2007 Dec;44(4):371-377. [doi: 10.1037/0033-3204.44.4.371] [Medline: 22122315]
10. Pugh MA, Vetere A. Lost in translation: an interpretative phenomenological analysis of mental health professionals' experiences of empathy in clinical work with an interpreter. Psychol Psychother 2009 Sep;82(Pt 3):305-321. [doi: 10.1348/147608308X397059] [Medline: 19208292]
11. Tan L, Le MK, Yu CC, et al. Defining clinical empathy: a grounded theory approach from the perspective of healthcare workers and patients in a multicultural setting. BMJ Open 2021 Sep 14;11(9):e045224. [doi: 10.1136/bmjopen-2020-045224] [Medline: 34521657]
12. McHenry M, Parker PA, Baile WF, Lenzi R. Voice analysis during bad news discussion in oncology: reduced pitch, decreased speaking rate, and nonverbal communication of empathy. Support Care Cancer 2012 May;20(5):1073-1078. [doi: 10.1007/s00520-011-1187-8] [Medline: 21573770]
13. Gustafsson SR, Wahlberg AC. The telephone nursing dialogue process: an integrative review. BMC Nurs 2023;22(1):345. [doi: 10.1186/s12912-023-01509-0]
14. Mental health atlas 2020: review of the Eastern Mediterranean region. World Health Organization. 2022. URL: https://applications.emro.who.int/docs/9789292740443-eng.pdf?ua=1 [accessed 2025-03-27]
15. Callejas Z, Griol D. Conversational agents for mental health and wellbeing. dialog systems: a perspective from language, logic and computation. In: Dialog Systems: Springer, Cham; 2021:219-244. [doi: 10.1007/978-3-030-61438-6_11]
16. Pelau C, Dabija DC, Ene I. What makes an AI device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry. Comput Human Behav 2021 Sep;122:106855. [doi: 10.1016/j.chb.2021.106855]
17. Iyer R, Nedeljkovic M, Meyer D. Using voice biomarkers to classify suicide risk in adult telehealth callers: retrospective observational study. JMIR Ment Health 2022 Aug 15;9(8):e39807. [doi: 10.2196/39807] [Medline: 35969444]
18. Posner K, Brown GK, Stanley B, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry 2011 Dec;168(12):1266-1277. [doi: 10.1176/appi.ajp.2011.10111704] [Medline: 22193671]
19. RStudio. RStudio IDE (Version 2024.04.2+764). 2024. URL: https://posit.co/products/open-source/rstudio/ [accessed 2025-04-09]
20. Audacity 3.5.1. Audacity Support. 2024. URL: https://support.audacityteam.org/additional-resources/changelog/older-versions/audacity-3.5/audacity-3.5.1 [accessed 2025-03-27]
21. Heck EJ, Davis CS. Differential expression of empathy in a counseling analogue. J Couns Psychol 1973;20(2):101-104. [doi: 10.1037/h0034171]
22. Mayer JD, Salovey P, Caruso DR, Sitarenios G. Measuring emotional intelligence with the MSCEIT V2.0. Emotion 2003 Mar;3(1):97-105. [doi: 10.1037/1528-3542.3.1.97] [Medline: 12899321]
23. Bodie GD. The Active-Empathic Listening Scale (AELS): conceptualization and evidence of validity within the interpersonal domain. Commun Q 2011 Jul;59(3):277-295. [doi: 10.1080/01463373.2011.583495]
24. Joshi A, Kale S, Chandel S, Pal DK. Likert Scale: explored and explained. BJAST 2015;7(4):396-403. [doi: 10.9734/BJAST/2015/14975]
25. Ownby KK. Use of the distress thermometer in clinical practice. J Adv Pract Oncol 2019 Mar;10(2):175-179. [Medline: 31538028]
26. Bikker AP, Fitzpatrick B, Murphy D, Mercer SW. Measuring empathic, person-centred communication in primary CARE nurses: validity and reliability of the Consultation and Relational Empathy (CARE) measure. BMC Fam Pract 2015 Oct 23;16:149. [doi: 10.1186/s12875-015-0374-y] [Medline: 26493072]
27. Anikin A. Soundgen: an open-source tool for synthesizing nonverbal vocalizations. Behav Res Methods 2019 Apr;51(2):778-792. [doi: 10.3758/s13428-018-1095-7] [Medline: 30054898]
28. Casale S, Russo A, Scebba G, Serrano S. Speech emotion classification using machine learning algorithms. Presented at: Proceedings of the 2th IEEE International Conference on Semantic Computing (ICSC 2008); Aug 4-7, 2008; Santa Monica, CA. [doi: 10.1109/ICSC.2008.43]
29. Rasmussen MA, Bro R. A tutorial on the LASSO approach to sparse modeling. Chemometr Intell Lab Syst 2012 Oct;119:21-31. [doi: 10.1016/j.chemolab.2012.10.003]
30. Sóskuthy M. Generalised additive mixed models for dynamic analysis in linguistics: a practical introduction. arXiv. Preprint posted online on Mar 15, 2017. [doi: 10.48550/arXiv.1703.05339]
31. Wood SN, Pya N, Säfken B. Smoothing parameter and model selection for general smooth models. J Am Stat Assoc 2016 Oct 1;111(516):1548-1563. [doi: 10.1080/01621459.2016.1180986]

XSL•FO
RenderX

32. Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J R Stat Soc Series B Stat Methodol 2011 Jan 1;73(1):3-36. [doi: 10.1111/j.1467-9868.2010.00749.x]
33. Noroozi F, Sapiński T, Kamińska D, Anbarjafari G. Vocal-based emotion recognition using random forests and decision tree. Int J Speech Technol 2017 Jun;20(2):239-246. [doi: 10.1007/s10772-017-9396-2]
34. Guo CY, Chen F, Chang YJ, Yan JT. Applying random forest classification to diagnose autism using acoustical voice-quality parameters during lexical tone production. Biomed Signal Process Control 2022 Aug;77:103811. [doi: 10.1016/j.bspc.2022.103811]
35. Raahul A, Sapthagiri R, Pankaj K, Vijayarajan V. Voice based gender classification using machine learning. IOP Conf Ser Mater Sci Eng 2017 Nov 1;263:042083. [doi: 10.1088/1757-899X/263/4/042083]
36. Jacob A. Modelling speech emotion recognition using logistic regression and decision trees. Int J Speech Technol 2017 Dec;20(4):897-905. [doi: 10.1007/s10772-017-9457-6]
37. Tafiadis D, Kosma EI, Chronopoulos SK, et al. Voice handicap index and interpretation of the cutoff points using receiver operating characteristic curve as screening for young adult female smokers. J Voice 2018 Jan;32(1):64-69. [doi: 10.1016/j.jvoice.2017.03.009] [Medline: 28392085]
38. Rice ME, Harris GT. What does it mean when age is related to recidivism among sex offenders? Law Hum Behav 2014 Apr;38(2):151-161. [doi: 10.1037/lhb0000052] [Medline: 23876093]
39. Uloza V, Latoszek BBV, Ulozaite-Staniene N, Petrauskas T, Maryn Y. A comparison of Dysphonia Severity Index and Acoustic Voice Quality Index measures in differentiating normal and dysphonic voices. Eur Arch Otorhinolaryngol 2018 Apr;275(4):949-958. [doi: 10.1007/s00405-018-4903-x] [Medline: 29442165]
40. Carter JV, Pan J, Rai SN, Galandiuk S. ROC-ing along: evaluation and interpretation of receiver operating characteristic curves. Surgery 2016 Jun;159(6):1638-1645. [doi: 10.1016/j.surg.2015.12.029] [Medline: 26962006]
41. Zou L, Choi YH, Guizzetti L, Shu D, Zou J, Zou G. Extending the DeLong algorithm for comparing areas under correlated receiver operating characteristic curves with missing data. Stat Med 2024 Sep 20;43(21):4148-4162. [doi: 10.1002/sim.10172] [Medline: 39013403]
42. Gerdes KE, Segal E. Importance of empathy for social work practice: integrating new science. Soc Work 2011 Apr;56(2):141-148. [doi: 10.1093/sw/56.2.141] [Medline: 21553577]
43. Regenbogen C, Schneider DA, Finkelmeyer A, et al. The differential contribution of facial expressions, prosody, and speech content to empathy. Cogn Emot 2012 Sep;26(6):995-1014. [doi: 10.1080/02699931.2011.631296]
44. Concannon S, Tomalin M. Measuring perceived empathy in dialogue systems. AI Soc 2024 Oct;39(5):2233-2247. [doi: 10.1007/s00146-023-01715-z]
45. Chen R, Chen H, Kulkarni A, et al. Detecting empathy in speech. Presented at: Interspeech 2024; Sep 1-5, 2024; Kos, Greece URL: https://www.isca-archive.org/interspeech_2024 [accessed 2025-04-09] [doi: 10.21437/Interspeech.2024-347]
46. Alam F, Danieli M, Riccardi G. Annotating and modeling empathy in spoken conversations. Comput SPEECH Lang 2018 Jul;50:40-61. [doi: 10.1016/j.csl.2017.12.003]
47. Zhao Q, Neumann DL, Cao Y, et al. Culture-sex interaction and the self-report empathy in Australians and mainland Chinese. Front Psychol 2019;10:396. [doi: 10.3389/fpsyg.2019.00396] [Medline: 30914986]
48. Hasan MR, Hossain MZ, Ghosh S, Krishna A, Gedeon T. Empathy detection from text, audiovisual, audio or physiological signals: task formulations and machine learning methods. arXiv. Preprint posted online on Dec 17, 2024. [doi: 10.48550/arXiv.2311.00721]
49. Besel LDS, Yuille JC. Individual differences in empathy: the role of facial expression recognition. Pers Individ Dif 2010 Jul;49(2):107-112. [doi: 10.1016/j.paid.2010.03.013]
50. Iyer R, Nedeljkovic M, Meyer D. Using vocal characteristics to classify psychological distress in adult helpline callers: retrospective observational study. JMIR Form Res 2022 Dec 19;6(12):e42249. [doi: 10.2196/42249] [Medline: 36534456]
51. Thölke P, Mantilla-Ramos YJ, Abdelhedi H, et al. Class imbalance should not throw you off balance: choosing the right classifiers and performance metrics for brain decoding with imbalanced data. Neuroimage 2023 Aug 15;277:120253. [doi: 10.1016/j.neuroimage.2023.120253] [Medline: 37385392]
52. Minixhofer C, Klejch O, Bell P. Evaluating and reducing the distance between synthetic and real speech distributions. 2023 Presented at: Interspeech 2023; Aug 20-24, 2023; Dublin, Ireland URL: https://www.isca-archive.org/interspeech_2023/ [accessed 2025-03-27] [doi: 10.21437/Interspeech.2023-1978]
53. Sanjeewa R, Iyer R, Apputhurai P, Wickramasinghe N, Meyer D. Empathic conversational agent platform designs and their evaluation in the context of mental health: systematic review. JMIR Ment Health 2024 Sep 9;11:e58974. [doi: 10.2196/58974] [Medline: 39250799]

Abbreviations

AEL: Active-Empathic Listening
AUC: area under the curve
C-SSRS: Columbia Suicide Severity Rating Scale
CTE: Carkhuff and Truax Empathy

XSL•FO
RenderX

GAMM: generalized additive mixed model
LASSO: Least Absolute Shrinkage and Selection Operator
MH: mental health
PEI: Perceived Emotional Intelligence
RS7: Rating Scale 7
TRIPOD+AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis+Artificial Intelligence

Edited by A Mavragani; submitted 22.10.24; peer-reviewed by R Gore, S Chowdhury; revised version received 16.02.25; accepted 18.02.25; published 16.04.25.

Please cite as:
Sanjeewa R, Iyer R, Apputhurai P, Wickramasinghe N, Meyer D
Machine Learning Approach to Identifying Empathy Using the Vocals of Mental Health Helpline Counselors: Algorithm Development and Validation
JMIR Form Res 2025;9:e67835
URL: https://formative.jmir.org/2025/1/e67835
doi:10.2196/67835

© Ruvini Sanjeewa, Ravi Iyer, Pragalathan Apputhurai, Nilmini Wickramasinghe, Denny Meyer. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Acoustic and Natural Language Markers for Bipolar Disorder: A Pilot, mHealth Cross-Sectional Study

Cristina Crocamo[1], PhD; Riccardo Matteo Cioni[1], MD; Aurelia Canestro[1], MD; Christian Nasti[1], MD; Dario Palpella[1], MD; Susanna Piacenti[1], MD; Alessandra Bartoccetti[1], MD; Martina Re[1], MD; Valentina Simonetti[2], MSc; Chiara Barattieri di San Pietro[2,3], PhD; Maria Bulgheroni[2], MSc; Francesco Bartoli[1], MD, PhD; Giuseppe Carrà[1], MD, MSc, PhD

[1]School of Medicine and Surgery, University of Milano-Bicocca, via Cadore 48, Monza, Italy

[2]Ab.Acus, Milan, Italy

[3]Laboratory of Neurolinguistics and Experimental Pragmatics (NEP), University School for Advanced Studies IUSS, Pavia, Italy

Corresponding Author:
Cristina Crocamo, PhD
School of Medicine and Surgery, University of Milano-Bicocca, via Cadore 48, Monza, Italy

Abstract

Background: Monitoring symptoms of bipolar disorder (BD) is a challenge faced by mental health services. Speech patterns are crucial in assessing the current experiences, emotions, and thought patterns of people with BD. Natural language processing (NLP) and acoustic signal processing may support ongoing BD assessment within a mobile health (mHealth) framework.

Objective: Using both acoustic and NLP-based features from the speech of people with BD, we built an app-based tool and tested its feasibility and performance to remotely assess the individual clinical status.

Methods: We carried out a pilot, observational study, sampling adults diagnosed with BD from the caseload of the Nord Milano Mental Health Trust (Italy) to explore the relationship between selected speech features and symptom severity and to test their potential to remotely assess mental health status. Symptom severity assessment was based on clinician ratings, using the Young Mania Rating Scale (YMRS) and Montgomery-Åsberg Depression Rating Scale (MADRS) for manic and depressive symptoms, respectively. Leveraging a digital health tool embedded in a mobile app, which records and processes speech, participants self-administered verbal performance tasks. Both NLP-based and acoustic features were extracted, testing associations with mood states and exploiting machine learning approaches based on random forest models.

Results: We included 32 subjects (mean [SD] age 49.6 [14.3] years; 50% [16/32] females) with a MADRS median (IQR) score of 13 (21) and a YMRS median (IQR) score of 5 (16). Participants freely managed the digital environment of the app, without perceiving it as intrusive and reporting an acceptable system usability level (average score 73.5, SD 19.7). Small-to-moderate correlations between speech features and symptom severity were uncovered, with sex-based differences in predictive capability. Higher latency time (ρ=0.152), increased silences (ρ=0.416), and vocal perturbations correlated with depressive symptomatology. Pressure of speech based on the mean intraword time (ρ=–0.343) and lower voice instability based on jitter-related parameters (ρ ranging from –0.19 to –0.27) were detected for manic symptoms. However, a higher contribution of NLP-based and conversational features, rather than acoustic features, was uncovered, especially for predictive models for depressive symptom severity (NLP-based: R^2=0.25, mean squared error [MSE]=110.07, mean absolute error [MAE]=8.17; acoustics: R^2=0.11, MSE=133.75, MAE=8.86; combined: R^2=0.16; MSE=118.53, MAE=8.68).

Conclusions: Remotely collected speech patterns, including both linguistic and acoustic features, are associated with symptom severity levels and may help differentiate clinical conditions in individuals with BD during their mood state assessments. In the future, multimodal, smartphone-integrated digital ecological momentary assessments could serve as a powerful tool for clinical purposes, remotely complementing standard, in-person mental health evaluations.

(JMIR Form Res 2025;9:e65555) doi:10.2196/65555

KEYWORDS

digital mental health; remote assessment; mHealth; speech; NLP; natural language processing; acoustic; symptom severity; machine learning; markers; mental health; bipolar disorders; app; applications; multimodal; mobile health; voice; vocal; bipolar; verbal; emotion; emotional; psychiatry; psychiatric; mental illness

XSL•FO
RenderX

Introduction

Bipolar disorder (BD) is a lifelong, episodic illness characterized by mood recurrences, including manic or hypomanic, depressive, and mixed episodes [1-3]. The burden associated with BD, affecting families, carers, and mental health care systems, is heavy [4]. Community services often struggle in delivering regular monitoring of BD treatment needs, resulting in relapses that seem difficult to predict [4-6].

Language disturbances are among the core symptoms of acute episodes in BD, since speech patterns are modulated by the emotional and neurophysiological status [7,8]. Therefore, language may play a key role in the assessment of an individual's current experiences, emotions, thought patterns, and symptoms. While content analysis may reveal grandiosity associated with elevated mood, impulsivity, or changes in goal-directed activities, natural language may provide insights into mood fluctuations, cognitive processes, and behavioral patterns [9]. In particular, changes in the rate of speech are likely to indicate mood oscillations, including pressure of speech and increased verbosity during manic episodes [10] and poverty of speech and increased pause times during depressive episodes [11-13]. Clinicians are trained to recognize variations in language and voice, along with gestures and facial expressions, implicitly assessing both coherence and organization of speech and natural language features. However, this process is inevitably vulnerable to inconsistencies and biases.

Recent research in mental health and computer science has put forward computational approaches for speech analysis across a variety of mental disorders, proposing automated methods to assess and monitor the individual's mental state through speech patterns [14-18]. Promising techniques in speech acoustic signal processing [10,11,17,19-21], using mobile health (mHealth) technology, can bridge subjective and objective components across various stages, such as prediction of illness onset, diagnostic processes, assessment of severity, and forecast of treatment outcomes [22-25]. Indeed, natural language processing (NLP) techniques, exploring language resources (eg, lexical choices, syntax, and semantics) both qualitatively and quantitatively (eg, topic modeling, clustering, and classification), may produce deeper insights across different clinical conditions [9,26]. For example, observable linguistic traits (eg, increased use of both first-person pronouns and negative emotion expressions) can be identified among people with BD [23]. However, although linguistic features are informative, they are context-dependent and inferred according to word transcriptions [27]. Thus, speech analyses combining acoustic-dependent features (eg, speech prosody and voice quality) with NLP-based measures appear more promising in terms of model predictions, possibly providing a more accurate mental health assessment [23,27,28].

Indeed, research has shown that acoustic features are markers of emotional states in BD [29], and that quantifiable speech differences can predict the scores of scales such as the Young Mania Rating Scale (YMRS) and the Montgomery-Åsberg Depression Rating Scale (MADRS) [13,27]. On the other hand, recent evidence has shown how smartphone-based voice data [30] can enhance BD monitoring in real time, detecting possible mood changes [31,32]. Thus, speech-based systems embedded in smartphones might be useful tools for complementary, continuous assessments of BD clinical states. We therefore built an app-based tool, jointly using acoustic and NLP-based features from the speech of people with BD who delivered a narrative, and carried out a pilot study aimed at testing its feasibility and performance to remotely assess the individual clinical status. Continuous, uninterrupted spoken accounts, as supplied by individuals, provided the unique opportunity to combine communication style information from an in-depth set of acoustic features and NLP-based scores as potential digital markers of symptom severity in speech. We rigorously chose to test the tool's performance against standard psychometric assessments of mania and depression in order to explore its potential for remote, complementary assessments.

Methods

The report of this study adheres to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement (checklist presented in Multimedia Appendix 1) [33].

Study Design and Sampling Strategies

We conducted a pilot, cross-sectional study involving adult participants (aged 18 years or older) from the caseload of the Nord Milano Mental Health Trust (Italy). The Trust includes 2 psychiatric intensive care units, with a total of 27 beds, and also provides community mental health care for the same 280,000 inhabitants of the northern area of the Metropolitan City of Milan through 4 community mental health teams with multidisciplinary staff. The relevant catchment area comprises highly urbanized, both deprived and affluent, districts.

Inclusion criteria comprised a diagnosis of BD and the willingness to participate in the study. People with physical impairments affecting their acoustic capabilities were excluded. Based on inclusion and exclusion criteria, eligible individuals were identified among individuals consecutively admitted to the Trust. Then, they were approached by the research team, explaining the purpose of the study and, if any, potential risks.

Ethical Considerations

Recruitment efforts were carried out in accordance with ethical guidelines to ensure the well-being and safety of all participants. Study participants signed a written informed consent and were not compensated for their involvement. The study received ethical approval (protocol number 172 - 17032023) from the local ethical committee. To maintain participant privacy and confidentiality, all study data were pseudonymized prior to analysis. No individual participants are identifiable in any images included in this manuscript or Multimedia Appendices.

Procedures

Acoustic data were retrieved by asking participants to self-administer verbal performance tasks through a mobile app on their smartphones (SPEAKapp; [34]). Clinical testing and app usage took place on the same day in the study setting (inpatient and outpatient services). Then, the System Usability Scale (SUS), a short 10-item questionnaire based on a 5-point

XSL•FO
RenderX

Likert scale, was administered to assess the usability [35] of the app.

Verbal performance in terms of prose recall was based on the Babcock test [36], for which participants were asked to listen to a short story characterized by graphic and intense contents (eg, a death in a car crash) and then to repeat what she or he remembered from this narrative. This enabled to capture speech timing patterns based on sustained speech samples.

The app gathered participants' verbal production by using the smartphone-integrated microphone, recording and processing participants' speech by leveraging Google Speech-To-Text APIs [37] and Python libraries (eg, Parselmouth for the Praat software [38]). Recordings involved the use of one audio channel based on the participant's voice in a controlled environment with minimal acoustic conditions. Both the raw audio data and the transcribed text content were processed to extract acoustic and NLP-based features from speech outputs. NLP and acoustic signal models were embedded in the backend part of the mobile app.

Measures

Consistent with recent evidence, we assumed speech as verbal behavior, the spoken output of the mental system underlying the language [39]. Through speech recognition, acoustic and linguistic features were extracted. Then, based on both NLP and acoustic features, we considered a multidimensional framework in order to generate appropriate discriminative information for the potential use of speech patterns as digital markers in BD [27,31]. A full description of selected features is provided in Table S1 in Multimedia Appendix 2.

NLP-Based, Semantic, and Conversational Indices

NLP-based scores were computed according to distributional semantic models, encompassing vectorial representations for the meaning of words in a multi-dimensional space.

Standard linguistic scores included both the number of words, indicative of poverty of speech, and the number of words produced that matched the story text. On the other hand, novel NLP-based scores integrated mean intraword time, estimating the average time taken to articulate or pronounce subsequent words, as an indicator of processing speed, as well as word mover's distance (WMD), capturing both lexical overlap and semantic similarity. In particular, WMD was estimated as the minimum cumulative distance between words required to exactly match the point cloud of the text of the full correct story (ie, the content distance between the full correct story and the story narrative produced by the participant), thus incorporating the semantic similarity between individual word pairs into the word distance metric [40]. In addition, latency time was calculated as a novel NLP-based score, taking into account the delay between the initiation of a spoken utterance or action and the production of its intended outcome or response when starting the task (ie, the first word).

Additional objective information was extracted from speech data. These quantitative measures included (1) speech duration, (2) speaking time (ie, phonation), (3) silence, (4) ratios of speaking time to speech duration as well as of silence to speaking time, and (5) speech rate.

Acoustic Indices From Vocal Signals (Prosodic Cues Indices)

Measures for prosodic cues (acoustic indices quantifying how people talk during conversations) were based on the signal's frequency and energy or amplitude. These were assumed to contribute to conveying paralinguistic meaning [41]. Based on nontextual data, acoustic components of speech were defined as the key phonetic elements, that is, objectively and reproducibly quantified speech sounds [27,42]. Fundamental frequency (F0) was measured by the frequency of phonation [43]. The short-term instability of the vibration of the vocal cords during phonation (ie, jitter-related indices) was also extracted (Table S1 in Multimedia Appendix 2). Higher jitter values indicated speech patterns likely characterized by irregularities or hesitations, thus mirroring potential underlying psychological distress or emotional instability. Furthermore, microperturbations of the ampleness of the signal (ie, how variable acoustic peaks refer to the period-to-period variability of the signal peak-to-peak amplitude) were identified as small fluctuations in the intensity of vocal sound waves by shimmer-related measures, with higher values indicating greater variability or instability, while lower ones suggesting more stable vocal intensity (ie, smoother and more regular speech production).

Since both periodic and nonperiodic sound waves may characterize the voice, the mean harmonics-to-noise ratio was used to measure the relationship between harmonic and nonharmonic voice elements. Noisier, more raucous voices (ie, not smooth or clear) were expected to show lower harmonics-to-noise ratios, indicating vocal cord tension or irritation, possibly suggesting emotional distress.

Psychometric Measures

Diagnosis of BD was confirmed by the Structured Clinical Interview for *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*; SCID-5). Based on clinician-rated assessments, depressive symptom severity was measured by the MADRS [44], while YMRS was used to assess manic symptoms [45]. Scores ranged from 0 to 60 for both MADRS [44] and YMRS [45]. In addition, cutoffs for severe mood symptoms were either a YMRS score ≥20 [46,47] or a MADRS score ≥19 [48].

Statistical Analyses

First, we summarized participants' characteristics, providing standard statistics for continuous and categorical variables. For both MADRS and YMRS, continuous scores were used. However, a supplementary analysis was performed based on clinically meaningful thresholds for symptom severity. A bivariate analysis was then carried out to measure the strength of the potential association between speech indices and psychometric measures. Features' summary statistics were plotted, and correlation coefficients (Pearson and Spearman, according to assumptions on data distribution, eg, normality) were estimated. Color gradient heat plots were also generated for data visualization. Taking into account potential sex

XSL•FO
RenderX

differences in speech acoustic indices [49-51], subgroup analyses were performed. Statistical significance was set at $P<.05$.

Second, based on state-of-the-art algorithms, NLP and acoustic features extracted from natural language and audio streams (Table S1 in Multimedia Appendix 2) were used to train machine-learning models to detect depressive and manic states by means of scores from MADRS and YMRS. Data were randomly split using a 5-fold nested cross-validation approach for training and testing in order to provide an unbiased evaluation of the model's performance. In particular, random forest (RF) models, with the potential to handle both linear and nonlinear relationships between features and the target variable, were implemented. The supervised learning algorithm, with no assumptions about the distribution of the target variable, was based on the ensemble learning method of different decision trees, whose predictions were aggregated using the scikit-learn library in Python. Exploiting the bagging techniques, building multiple decision trees, RF contributed to minimizing overfitting issues by randomizing the feature selection during each tree split. This was assumed to reduce sensitivity to noise and to make decision trees less correlated through the use of a unique subset of the initial data for every base model. Moreover, we deemed features scaling unnecessary due to both the properties of the RF model and the performance metrics of comparisons. Relevant models were trained to test final performance by metrics (ie, mean squared error [MSE], mean absolute error [MAE], and *R*-squared [R^2]). These tested overall performance, even controlling for sex. Shapley Additive Explanations values, showing features' impact, were plotted. Data were analyzed using Stata release 18 and Python (version 3.10.9).

Results

Sample Characteristics

We included 32 subjects with BD (mean age 49.6, SD 14.3 years; 50% [16/32] females). The mean (SD) age at onset was 24.4 (10) years. As a whole, participants experienced more manic (median 4, IQR 8) than depressive episodes (median 2, IQR 5). About 40% (12/32) of participants reported a previous mood episode within 1 year before study enrollment. The MADRS median (IQR) score was 13 (21), while the YMRS median (IQR) score was 5 (16). Considering the app usage, participants reported high SUS scores on average (mean 73.5, SD 19.7). Demographic and clinical details are fully provided in Table 1.

XSL•FO
RenderX

Table . Sample[a] characteristics.

Characteristics	BD[b] (N=32)
Sex, n (%)	
Female	16 (50)
Male	16 (50)
Age (years), mean (SD)	49.6 (14.3)
Marital status, n (%)	
In a relationship	12 (37)
Family situation, n (%)	
Living alone	11 (34)
Education, n (%)[c]	
Elementary	1 (3)
Middle	12 (37)
High	13 (41)
University or superior	5 (16)
Employment, n (%)[c]	
Employed	13 (41)
Setting, n (%)	
Outpatient	13 (41)
Inpatient	19 (59)
Polarity of first episode, n (%)[c]	
Depressive	12 (38)
Hypomaniac or maniac	13 (41)
Unknown	1 (3)
Age of onset[c] (years), mean (SD)	24.4 (10)
Family history[c]	11 (34%)
Hospitalizations, median (IQR)	
Lifetime	3 (7.5)
12 months	1 (2)
Suicide attempts (lifetime), n (%)	10 (31)
Alcohol use disorder (lifetime), n (%)	3 (9)
Substance use disorder (lifetime), n (%)	6 (19)
Medication, n (%)	
FGA[d]	6 (19)
SGA[e]	28 (87)
Mood stabilizer	26 (81)
Antidepressant	8 (25)
Benzodiazepine	16 (50)
Psychometric assessment	
Depressive symptoms (MADRS[f]), median (IQR)	13 (21)
MADRS <19, n (%)	17 (53)
MADRS ≥19, n (%)	15 (47)

XSL•FO
RenderX

Characteristics	BD[b] (N=32)
Manic symptoms (YMRS[g]), median (SD)	5 (16)
YMRS <20, n (%)	26 (81)
YMRS ≥20, n (%)	6 (19)
SUS[h] score, mean (SD)	73.5 (19.7)

[a]The sample is for a pilot, cross-sectional study in Italy.

[b]BD: bipolar disorder.

[c]Missing values: education (1), employment (2), age of onset (10), polarity of first episode (6), family history (10), alcohol use disorder (2), substance use disorder (1), FGA (2), SGA (1), mood stabilizer (2), antidepressant (3), benzodiazepine (4).

[d]FGA: first-generation antipsychotics.

[e]SGA: second-generation antipsychotics.

[f]MADRS: Montgomery-Åsberg Depression Rating Scale.

[g]YMRS: Young Mania Rating Scale.

[h]SUS: System Usability Scale (range 0 - 100).

Associations Between Symptom Severity and Speech Features

For descriptive purposes, NLP-based, conversational, and acoustic features are summarized in Figures S1A-S1D and S2A-S2D in Multimedia Appendix 3 by depressive and manic symptom severity, respectively.

In particular, grouping data into 2 categories (Multimedia Appendix 3), statistically significant differences by depressive symptoms' severity were found for many NLP-based and conversational-like measures, including word number, phonation (also as percentage over the speech duration), and mean intraword time. Correlation analyses, based on Spearman nonparametric analysis of symptom severity continuous scores, are displayed in Figures 1A-C and 2A-C. These showed that both the total number of words and the length of phonation, as well as the related percentage out of segment duration, were negatively correlated (coefficients=−0.35, −0.32, and −0.42) to depressive symptoms (Figure 1A). Consistent results were observed for the ratio between silence and phonation (coefficient=0.42), as well as for mean intraword time, which was positively correlated to depressive (coefficient=0.53) and negatively to manic (coefficient=−0.34) symptoms. Among items for depressive symptoms assessment, this correlation was particularly clear between acoustic features and suicidal thoughts (coefficients ranging from 0.18 to 0.51). In addition, latency time also showed a moderate, though obviously opposite, correlation with manic and depressive symptoms, respectively (coefficients=−0.28 and 0.15).

Subgroup analyses for NLP-based and conversational features revealed more pronounced relationships in females (Figure 1C) as compared with males (Figure 1B), showing a high correlation between depressive symptoms and mean intraword time (coefficient=0.75), phonation percentage (coefficient=−0.56), and, consequently, the silence-phonation ratio (coefficient=0.56). Similarly, latency time was negatively correlated to manic symptoms among females (coefficient=−0.60).

XSL•FO
RenderX

Figure 1. Correlation heatmap of NLP-based, semantic and conversational features in people with bipolar disorder. (A) Overall sample; (B) Male subgroup; (C) Female subgroup. MADRS: Montgomery-Åsberg Depression Rating Scale, YMRS: Young Mania Rating Scale.

XSL•FO
RenderX

On the other hand, a small positive correlation was uncovered between depressive symptoms and higher values of instability in speech patterns (jitter-related indices, with coefficients ranging from 0.10 to 0.16). In contrast, small-to-moderate negative correlations were observed between manic symptoms and lower values of instability (jitter-related indices, with coefficients ranging from −0.19 to −0.27). Small estimates were found for F0, respectively (coefficient=0.16 and −0.18; Figure 2A). Except for shimmer_apq11 (manic symptoms coefficient=−0.22), we did not find any substantial relationship between shimmer-related indices (describing stable and unstable vocal intensity and speech production) and symptomatology.

Subgroup analyses suggested a role for sex also in influencing acoustic features. In particular, we found deeper connections in males as compared with females, especially in terms of F0 and jitter-related indices (Figure 2B and C).

XSL•FO
RenderX

Figure 2. Correlation heatmap of acoustic features in people with bipolar disorder. (A) Overall sample; (B) Male subgroup; (C) Female subgroup. MADRS: Montgomery-Åsberg Depression Rating Scale, YMRS: Young Mania Rating Scale.

XSL•FO
RenderX

Predictive Models From Speech Features

Considering depressive symptoms, performance metrics showed a contribution of NLP-based and conversational features higher than what was attributable to acoustic ones (Table 2). In particular, mean intraword time, silence-phonation ratio, ppq5 jitter (ie, perturbations in F0), WMD, and percentage of phonation over duration all ranked high in terms of relative importance.

Including sex into the analysis, a differential contribution of various features (NLP-based and conversational vs acoustics) to the predictive models for depressive (Figure 3A) and manic (Figure 3B) symptoms can be found. However, as for manic symptoms, although a relative contribution of different NLP-based and acoustic (eg, F0 SD) features was recorded, we could not find any reliable estimates for the relevant model, even including sex. Table 2 shows detailed estimated performance metrics for testing for the trained RF regressors, even controlling for sex.

XSL•FO
RenderX

Table . Performance estimates for random forest regression models in people with bipolar disorder.

Performance[a]	Depressive symptoms		Manic symptoms	
	Unadjusted	Adjusted[b]	Unadjusted	Adjusted[b]
NLP[c]				
R^2 average	0.26	0.25	—[d]	—
Fold 1	0.10	−0.55	−0.54	0.18
Fold 2	0.48	0.53	−0.13	0.02
Fold 3	0.06	0.37	0.25	0.12
Fold 4	0.54	0.64	0.23	−0.42
Fold 5	0.13	0.26	0.01	147.98
Mean squared error average	105.46	110.07	153.78	147.98
Fold 1	136.73	259.26	223.06	92.05
Fold 2	33.02	46.49	121.79	156.74
Fold 3	137.64	104.25	135.66	121.60
Fold 4	79.32	33.00	134.85	167.15
Fold 5	140.61	107.35	153.32	202.35
Mean absolute error average	8.08	8.17	10.58	10.13
Fold 1	9.58	13.64	12.47	7.79
Fold 2	3.36	5.57	9.28	10.90
Fold 3	10.31	8.56	10.40	9.26
Fold 4	7.59	4.34	9.82	9.71
Fold 5	9.26	8.74	10.96	13.00
Acoustics				
R^2 average	—	0.11	—	—
Fold 1	0.29	–0.22	0.002	–0.22
Fold 2	–0.83	–0.10	–0.02	–0.15
Fold 3	–0.59	0.03	–0.14	–0.14
Fold 4	0.23	0.18	–0.38	–0.44
Fold 5	0.36	0.64	–0.28	0.01
Mean squared error average	161.64	133.75	162.86	163.51
Fold 1	47.97	222.18	68.9	125.14
Fold 2	333.00	200.40	160.47	122.80
Fold 3	202.17	85.04	185.63	175.34
Fold 4	128.54	148.52	272.25	225.23
Fold 5	96.49	12.62	127.06	170.30
Mean absolute error average	10.02	8.86	10.35	10.73
Fold 1	5.27	11.76	7.09	9.77
Fold 2	16.43	13.8	10.94	8.82
Fold 3	11.77	6.00	12.48	11.70
Fold 4	9.69	10.13	14.05	12.34

XSL•FO
RenderX

Performance[a]	Depressive symptoms		Manic symptoms	
	Unadjusted	Adjusted[b]	Unadjusted	Adjusted[b]
Fold 5	6.95	2.62	7.21	11.01
Combined				
R^2 average	0.05	0.16	—	—
Fold 1	0.32	0.60	–0.56	–0.13
Fold 2	–0.09	0.11	0.24	0.07
Fold 3	–0.29	0.07	0.08	–0.54
Fold 4	0.10	0.04	0.06	0.18
Fold 5	0.22	0.22	–0.41	0.14
Mean squared error average	120.90	118.53	135.94	140.03
Fold 1	87.51	34.13	158.54	183.39
Fold 2	183.67	111.11	60.67	122.32
Fold 3	184.71	164.45	192.81	126.43
Fold 4	47.71	148.83	178.84	112.36
Fold 5	100.91	134.10	88.83	155.68
Mean absolute error average	8.65	8.68	9.61	10.00
Fold 5	6.69	4.37	11.21	11.50
Fold 5	11.49	7.33	6.95	8.86
Fold 5	11.04	11.29	11.40	10.30
Fold 5	5.57	10.15	11.19	8.26
Fold 5	8.46	10.26	7.28	11.08

[a]Metrics for testing based on a nested cross-validation approach (pilot, cross-sectional study, N=32). Range for symptom scores: 0 - 60.

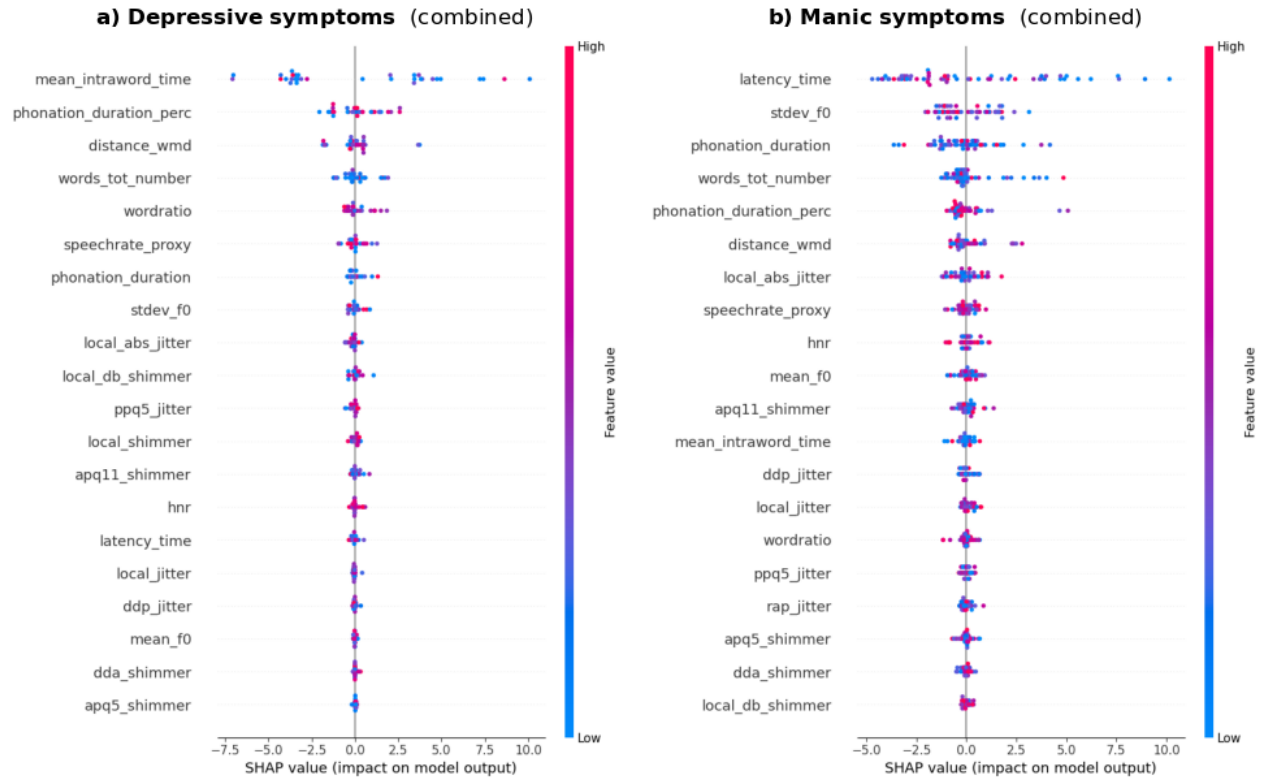
[b]Including sex.

[c]NLP: natural language processing.

[d]Not available.

XSL•FO
RenderX

Figure 3. Individual features contribution to depressive and manic symptoms predictions in sex-adjusted models among people with bipolar disorder.



Discussion

Main Findings

This study aimed at piloting the simultaneous use of speech acoustics, as well as natural language features, to glean insights into BD depressive and manic symptoms. Our findings corroborate evidence on the relationships between symptom severity and speech features, supporting the potential predictive role for clinical purposes of digital mental health applications, embedded in a mHealth integrated system.

First, the speech of participants with BD showed that vocal perturbations (eg, higher instability and hesitations considering voice quality), latency time, and increased silences and pauses over time speaking all correlated to depressive symptoms. Consistently, increased depressive symptoms resulted in NLP-based features such as a smaller number of words and longer mean intraword time, with lower pressure of speech. In our exploratory study, this relationship was particularly clear among females. This effect was corroborated by the predictive model, showing a contribution of NLP-based and conversational features higher than for acoustic ones. This finding aligns with prior evidence, advocating that text-based features contribute more to model accuracy than audio parameters [18]. However, also the latter component (ie, fundamental frequency, jitter- and shimmer-related indices) deserves a careful assessment, since our findings show that these indices might have an impact at least among males to predict future episodes. Indeed, recent evidence from healthy populations sheds light on sex differences in speech markers (eg, prosodic features) with different acoustic cues conveying various emotions [50]. A combination of inherent biological dissimilarities, socialization processes, influences of the social environment, and cultural expectations might contribute to these differences in both expression and perception of related emotional prosody [52,53]. Moreover, individuals may modulate their speech to align with the dominant pitch range within a specific linguistic community [54], and similar modulation may occur in conversational dialogues versus monologues and in spontaneous versus elicited speech. Thus, this criterion should be taken into account when designing apps with speech recognition and processing tasks for people with BD [31].

Second, voice instability and hesitations, as well as mean intraword time, were negatively correlated to manic symptoms. However, the interpretation of the relationship between manic features and vocal abnormalities is not straightforward. Mixed findings emerged on the relationships between speech features and manic symptoms, preventing us from supporting our original hypothesis. One plausible explanation may stem from the sample characteristics. Indeed, our participants were more likely to report depressive symptoms, and just a few had severe manic features.

However, the overall moderate correlations between speech markers and symptom severity were consistent with previous work that used speech smartphone data to discriminate between different mood states [20,21]. It has been argued that speech features may be useful to detect a trait [55] rather than a state [56] in BD. However, alterations in voice perturbations have been observed when assessing vocal markers of suicidal ideation [57], and this makes further research for vocal features reasonable, at least for depressive conditions.

XSL•FO
RenderX

Smartphone-Based Applications

Consistent with previous research on smartphone-based applications designed to record and analyze speech patterns in real time, our findings emphasize the feasibility of a simple, yet clinically useful, application of digital technology [13]. In particular, we developed the frontend of the app as a basic digital environment, freely managed by participants on their own smartphones. Participants reported a high level of engagement with the tool, showing an acceptable system usability level as assessed by SUS [35], without perceiving intrusiveness of the recording of both elicited and spontaneous conversations.

Comparisons of the vocal performance of people with BD with unaffected relatives and healthy controls have shown a clear speech "fingerprint" of the clinical condition [58], suggesting the utility of multilevel inputs [59]. However, there is also the need for a wider understanding of fluctuations in symptom severity and mood states in this population [60]. The major strength of our study consists in the usefulness of different speech data (eg, linguistic, conversational, acoustics) to differentially identify symptoms of BD. Thus, for relapse prevention purposes, future research should possibly explore systems combining smartphone-based generated objective acoustics data with additional information, such as from facial expressions and gestures [61]. This would ultimately improve BD state prediction, even considering classification tasks [21,62-64].

Clinical Implications: Interdisciplinary Perspective

This pilot study represents a step forward in the identification and utilization of digital biomarkers for BD from natural language and audio streams, with implications for personalized mental health care and early intervention strategies. Our approach holds promise for complementary, remote assessments enhancing depressive and partly manic states prediction by exploiting participants' speech. This would have significant implications, especially considering BD fluctuating symptomatology. Nonetheless, leveraging live speech recordings as a predictive tool, repeated assessments are needed to identify individuals at risk of transitioning to depressive and manic states.

Despite promising findings from automated assessments, mental health care heavily relies on participant interviews, yet with often subjective reports, cognitive limitations, and stigma [18]. Integrated systems, aiming at taking advantage of candidate digital markers from speech recognition, would possibly boost a care approach in which digital technology enhances, but does not replace, existing models from clinical assessment [30]. Indeed, automated assessment does not inherently lead to adherence and engagement of individuals with BD [65].

Finally, clinical, hypothesis-driven research on BD should not be dismissed, since algorithms may not be considered a black-box replacement for traditional data modeling, but they rather integrate with other systems, embedding a substantial clinical validation [66,67].

Limitations and Future Directions

We should acknowledge some limitations of this study. Analyzing speech and natural language in individuals with BD implies a challenge due to the nature of the disorder and to ethical considerations.

First, properties of chosen machine-learning models may hamper identification of unknown patterns based on values that fall outside the training set. Effective NLP and supervised learning models may require high-quality, annotated datasets. While exploratory in nature, the study's limited sample size may have constrained the model's statistical power and the ability to capture the full complexity of the underlying data distribution, thereby hindering meaningful subgroup comparisons. Our preliminary findings should be replicated and extended in a larger, more diverse sample of people with BD to mitigate the risks associated with overfitting. Furthermore, future research should address classification approaches based on severity thresholds for both MADRS and YMRS. Accordingly, there is potential for alternative modeling approaches for regression tasks (eg, splines) that might be implemented in the future. While still considering the number of predictors, these may possibly enable a better understanding of the nature of the existing relationships and nonlinear patterns.

Consistently, the lack of standardized (linguistic and acoustic) markers represents a barrier when studying relationships with mood states. Indeed, the model may still learn to overfit to irrelevant or noisy features the data may contain, especially if they are informative in the training set by chance.

Furthermore, the speaker's identity may show a possible confounding role in a between-subject design. Therefore, studies with a longitudinal design (ie, within-subjects) should be recommended, deploying Ecological Momentary Assessment approaches [24,68]. In addition, speech patterns may generate misinterpretations if individual cultural and linguistic factors are not accounted for [69]. Similarly, speech during manic episodes may exhibit circumstantiality or tangentiality, where individuals provide excessive details or veer off-topic. Rapid speech, tangential thinking, or unconventional language use pose challenges for automatic speech recognition systems. Analyzing such complex speech patterns requires a deep evaluation of language and context, achieving appropriate understanding of an individual's usual way of communicating in order to distinguish changes associated with BD episodes.

Furthermore, in our study, speech features were averaged over relevant duration, thus constraining the role of temporal variations across related measures in predicting symptom severity. Future research should endeavor to integrate dynamic aspects of speech on mood states transitioning.

Finally, other clinical variables, not investigated in our sample, are likely to influence the individual's speech. For instance, it should be noted that anxiety and anxious distress, often co-occurring with bipolar depression [70], may significantly influence speech features [71], as well as medication prescribed [72-74] and drug or alcohol comorbid conditions [75].

XSL•FO
RenderX

Conclusions

Speech patterns, underlying both linguistic and acoustic features, are able to yield quantifiable differences, thus embodying digital markers of symptom severity. Multimodal, smartphone-integrated digital assessments could serve as powerful tools for clinical purposes to remotely complement standard mental health evaluations, potentially contributing to distinguish clinical conditions in people with BD. Feasibility of similar systems seems promising, though issues related to privacy, intrusiveness, and clinical therapeutic relationships should be carefully considered.

Acknowledgments

This research was supported by the FSE REACT-EU Competitive Research Grant Axis-IV DM 1062/2021: “Natural Language Processing in Digital Mental Health.” The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

The datasets including audio data streams supporting the conclusions of this article are not publicly available as the original source has not granted permission to share that information but may be available from the corresponding author on reasonable request.

Authors' Contributions

CC, FB, and GC handled conceptualization. RMC, AC, CN, DP, SP, AB, and MR performed investigation. CC, FB, VS, CB, and MB contributed to methodology. VS and MB assisted with software. CC conducted formal analysis. GC performed supervision. CC contributed to writing – original draft. CC, RMC, AC, CN, DP, SP, AB, MR, VS, CB, MB, FB, and GC contributed to writing – review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Checklist.
[DOC File, 87 KB - formative_v9i1e65555_app1.doc]

Multimedia Appendix 2
Features.
[DOCX File, 16 KB - formative_v9i1e65555_app2.docx]

Multimedia Appendix 3
Supplementary analyses.
[DOCX File, 329 KB - formative_v9i1e65555_app3.docx]

References

1. Bartoli F, Crocamo C, Carrà G. Clinical correlates of DSM-5 mixed features in bipolar disorder: a meta-analysis. J Affect Disord 2020 Nov 1;276:234-240. [doi: 10.1016/j.jad.2020.07.035] [Medline: 32697704]
2. Grande I, Berk M, Birmaher B, Vieta E. Bipolar disorder. Lancet 2016 Apr 9;387(10027):1561-1572. [doi: 10.1016/S0140-6736(15)00241-X] [Medline: 26388529]
3. McIntyre RS, Berk M, Brietzke E, et al. Bipolar disorders. Lancet 2020 Dec 5;396(10265):1841-1856. [doi: 10.1016/S0140-6736(20)31544-0] [Medline: 33278937]
4. Karambelas GJ, Filia K, Byrne LK, Allott KA, Jayasinghe A, Cotton SM. A systematic review comparing caregiver burden and psychological functioning in caregivers of individuals with schizophrenia spectrum disorders and bipolar disorders. BMC Psychiatry 2022 Jun 23;22(1):422. [doi: 10.1186/s12888-022-04069-w] [Medline: 35733174]
5. Fajutrao L, Locklear J, Priaulx J, Heyes A. A systematic review of the evidence of the burden of bipolar disorder in Europe. Clin Pract Epidemiol Ment Health 2009 Jan 23;5:3. [doi: 10.1186/1745-0179-5-3] [Medline: 19166608]
6. Ogilvie AD, Morant N, Goodwin GM. The burden on informal caregivers of people with bipolar disorder. Bipolar Disord 2005;7 Suppl 1(Suppl 1):25-32. [doi: 10.1111/j.1399-5618.2005.00191.x] [Medline: 15762866]
7. Goodwin FK, Jamison KR. Manic-Depressive Illness: Bipolar Disorders and Recurrent Depression: Oxford University Press, USA; 2007.

XSL•FO
RenderX

8. Karam ZN, Provost EM, Singh S, et al. Ecologically valid long-term mood monitoring of individuals with bipolar disorder using speech. Proc IEEE Int Conf Acoust Speech Signal Process 2014 May;2014:4858-4862. [doi: 10.1109/ICASSP.2014.6854525] [Medline: 27630535]
9. Harvey D, Lobban F, Rayson P, Warner A, Jones S. Natural language processing methods and bipolar disorder: scoping review. JMIR Ment Health 2022 Apr 22;9(4):e35928. [doi: 10.2196/35928] [Medline: 35451984]
10. Birnbaum ML, Abrami A, Heisig S, et al. Acoustic and facial features from clinical interviews for machine learning-based psychiatric diagnosis: algorithm development. JMIR Ment Health 2022 Jan 24;9(1):e24699. [doi: 10.2196/24699] [Medline: 35072648]
11. Gideon J, Provost EM, McInnis M. Mood state prediction from speech of varying acoustic quality for individuals with bipolar disorder. Proc IEEE Int Conf Acoust Speech Signal Process 2016 Mar;2016:2359-2363. [doi: 10.1109/ICASSP.2016.7472099] [Medline: 27570493]
12. Guidi A, Schoentgen J, Bertschy G, Gentili C, Scilingo EP, Vanello N. Features of vocal frequency contour and speech rhythm in bipolar disorder. Biomed Signal Process Control 2017 Aug;37:23-31. [doi: 10.1016/j.bspc.2017.01.017]
13. Maxhuni A, Muñoz-Meléndez A, Osmani V, Perez H, Mayora O, Morales EF. Classification of bipolar disorder episodes based on analysis of voice and motor activity of patients. Pervasive Mob Comput 2016 Sep;31:50-66. [doi: 10.1016/j.pmcj.2016.01.008]
14. Matton K, McInnis MG, Provost EM. Into the wild: transitioning from recognizing mood in clinical interactions to personal conversations for individuals with bipolar disorder. 2019 Presented at: Interspeech 2019; Sep 15-19, 2019; Graz, Austria p. 1438-1442 URL: https://www.isca-archive.org/interspeech_2019 [doi: 10.21437/Interspeech.2019-2698]
15. Arevian AC, Bone D, Malandrakis N, et al. Clinical state tracking in serious mental illness through computational analysis of speech. PLoS ONE 2020;15(1):e0225695. [doi: 10.1371/journal.pone.0225695] [Medline: 31940347]
16. Girard JM, Vail AK, Liebenthal E, et al. Computational analysis of spoken language in acute psychosis and mania. Schizophr Res 2022 Jul;245:97-115. [doi: 10.1016/j.schres.2021.06.040] [Medline: 34456131]
17. Low DM, Bentley KH, Ghosh SS. Automated assessment of psychiatric disorders using speech: a systematic review. Laryngoscope Investig Otolaryngol 2020 Feb;5(1):96-116. [doi: 10.1002/lio2.354] [Medline: 32128436]
18. Malgaroli M, Hull TD, Zech JM, Althoff T. Natural language processing for mental health interventions: a systematic review and research framework. Transl Psychiatry 2023 Oct 6;13(1):309. [doi: 10.1038/s41398-023-02592-2] [Medline: 37798296]
19. Cummins N, Baird A, Schuller BW. Speech analysis for health: current state-of-the-art and the increasing impact of deep learning. Methods 2018 Dec 1;151:41-54. [doi: 10.1016/j.ymeth.2018.07.007] [Medline: 30099083]
20. Faurholt-Jepsen M, Busk J, Frost M, et al. Voice analysis as an objective state marker in bipolar disorder. Transl Psychiatry 2016 Jul 19;6(7):e856. [doi: 10.1038/tp.2016.123] [Medline: 27434490]
21. Faurholt-Jepsen M, Rohani DA, Busk J, et al. Discriminating between patients with unipolar disorder, bipolar disorder, and healthy control individuals based on voice features collected from naturalistic smartphone calls. Acta Psychiatr Scand 2022 Mar;145(3):255-267. [doi: 10.1111/acps.13391] [Medline: 34923626]
22. Daus H, Bloecher T, Egeler R, De Klerk R, Stork W, Backenstrass M. Development of an emotion-sensitive mHealth approach for mood-state recognition in bipolar disorder. JMIR Ment Health 2020 Jul 3;7(7):e14267. [doi: 10.2196/14267] [Medline: 32618577]
23. Dikaios K, Rempel S, Dumpala SH, Oore S, Kiefte M, Uher R. Applications of speech analysis in psychiatry. Harv Rev Psychiatry 2023;31(1):1-13. [doi: 10.1097/HRP.0000000000000356] [Medline: 36608078]
24. Dunster GP, Swendsen J, Merikangas KR. Real-time mobile monitoring of bipolar disorder: a review of evidence and future directions. Neuropsychopharmacology 2021 Jan;46(1):197-208. [doi: 10.1038/s41386-020-00830-5] [Medline: 32919408]
25. Marzano L, Bardill A, Fields B, et al. The application of mHealth to mental health: opportunities and challenges. Lancet Psychiatry 2015 Oct;2(10):942-948. [doi: 10.1016/S2215-0366(15)00268-0] [Medline: 26462228]
26. Le Glaz A, Haralambous Y, Kim-Dufor DH, et al. Machine learning and natural language processing in mental health: systematic review. J Med Internet Res 2021 May 4;23(5):e15708. [doi: 10.2196/15708] [Medline: 33944788]
27. Farrús M, Codina-Filbà J, Escudero J. Acoustic and prosodic information for home monitoring of bipolar disorder. Health Informatics J 2021;27(1):1460458220972755. [doi: 10.1177/1460458220972755] [Medline: 33438502]
28. Gong Y, Poellabauer C. Topic modeling based multi-modal depression detection. 2017 Oct 23 Presented at: MM '17; Oct 23-27, 2017; Mountain View California USA p. 2017 URL: https://dl.acm.org/doi/proceedings/10.1145/3133944 [doi: 10.1145/3133944.3133945]
29. Muaremi A, Gravenhorst F, Grünerbl A, Arnrich B, Tröster G. Assessing bipolar episodes using speech cues derived from phone calls. In: Lect Notes Inst Comput Sci Soc Inform Telecommun Eng 2014:103-114. [doi: 10.1007/978-3-319-11564-1_11]
30. Bond RR, Mulvenna MD, Potts C, O'Neill S, Ennis E, Torous J. Digital transformation of mental health services. Npj Ment Health Res 2023 Aug 22;2(1):13. [doi: 10.1038/s44184-023-00033-y] [Medline: 38609479]
31. Flanagan O, Chan A, Roop P, Sundram F. Using acoustic speech patterns from smartphones to investigate mood disorders: scoping review. JMIR Mhealth Uhealth 2021 Sep 17;9(9):e24352. [doi: 10.2196/24352] [Medline: 34533465]

XSL•FO
RenderX

32. de Oliveira L, Portugal LCL, Pereira M, et al. Predicting bipolar disorder risk factors in distressed young adults from patterns of brain activation to reward: a machine learning approach. Biol Psychiatry Cogn Neurosci Neuroimaging 2019 Aug;4(8):726-733. [doi: 10.1016/j.bpsc.2019.04.005] [Medline: 31201147]
33. von Elm E, Altman DG, Egger M, et al. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007 Oct 20;335(7624):806-808. [doi: 10.1136/bmj.39335.541782.AD] [Medline: 17947786]
34. Ab.acus Srl. URL: https://www.ab-acus.eu/index.php/portfolio-items/speakapp/ [accessed 2024-03-19]
35. Hyzy M, Bond R, Mulvenna M, et al. System Usability Scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290. [doi: 10.2196/37290] [Medline: 35980732]
36. Italian standardization and classification of neuropsychological tests. The Italian Group on the Neuropsychological Study of Aging. Ital J Neurol Sci 1987 Dec;Suppl 8(1-120):1-120. [Medline: 3330072]
37. Google speech-to-text apis. Google Cloud. URL: https://cloud.google.com/speech-to-text [accessed 2025-04-09]
38. Jadoul Y, Thompson B, de Boer B. Introducing Parselmouth: a Python interface to Praat. J Phon 2018 Nov;71:1-15. [doi: 10.1016/j.wocn.2018.07.001]
39. de Boer JN, Brederoo SG, Voppel AE, Sommer IEC. Anomalies in language as a biomarker for schizophrenia. Curr Opin Psychiatry 2020 May;33(3):212-218. [doi: 10.1097/YCO.0000000000000595] [Medline: 32049766]
40. Kusner M, Sun Y, Kolkin N, Weinberger K. From word embeddings to document distances. 2015 Presented at: Proceedings of the 32nd International Conference on Machine Learning, PMLR; Jul 5-7, 2015; Lille, France p. 957-966.
41. Nadeu M, Prieto P. Pitch range, gestural information, and perceived politeness in Catalan. J Pragmat 2011 Feb;43(3):841-854. [doi: 10.1016/j.pragma.2010.09.015]
42. Smith SW. Digital Signal Processing: A Practical Guide for Engineers and Scientists: California Technical Publishing; 2002:978. [doi: 10.1016/B978-0-7506-7444-7.X5036-5]
43. Ververidis D, Kotropoulos C. Emotional speech recognition: resources, features, and methods. Speech Commun 2006 Sep;48(9):1162-1181. [doi: 10.1016/j.specom.2006.04.003]
44. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979 Apr;134:382-389. [doi: 10.1192/bjp.134.4.382] [Medline: 444788]
45. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry 1978 Nov;133:429-435. [doi: 10.1192/bjp.133.5.429] [Medline: 728692]
46. Lukasiewicz M, Gerard S, Besnard A, et al. Young Mania Rating Scale: how to interpret the numbers? Determination of a severity threshold and of the minimal clinically significant difference in the EMBLEM cohort. Int J Methods Psychiatr Res 2013 Mar;22(1):46-58. [doi: 10.1002/mpr.1379] [Medline: 23526724]
47. Samara MT, Levine SZ, Leucht S. Linkage of Young Mania Rating Scale to Clinical Global Impression Scale to enhance utility in clinical practice and research trials. Pharmacopsychiatry 2023 Jan;56(1):18-24. [doi: 10.1055/a-1841-6672] [Medline: 35896419]
48. Thase ME, Harrington A, Calabrese J, Montgomery S, Niu X, Patel MD. Evaluation of MADRS severity thresholds in patients with bipolar depression. J Affect Disord 2021 May 1;286:58-63. [doi: 10.1016/j.jad.2021.02.043] [Medline: 33677183]
49. Eichhorn JT, Kent RD, Austin D, Vorperian HK. Effects of aging on vocal fundamental frequency and vowel formants in men and women. J Voice 2018 Sep;32(5):644. [doi: 10.1016/j.jvoice.2017.08.003] [Medline: 28864082]
50. Ertürk A, Gürses E, Kulak Kayıkcı ME. Sex related differences in the perception and production of emotional prosody in adults. Psychol Res 2024 Mar;88(2):449-457. [doi: 10.1007/s00426-023-01865-1] [Medline: 37542581]
51. Mendoza E, Valencia N, Muñoz J, Trujillo H. Differences in voice quality between men and women: use of the long-term average spectrum (LTAS). J Voice 1996 Mar;10(1):59-66. [doi: 10.1016/s0892-1997(96)80019-1] [Medline: 8653179]
52. Chaplin TM. Gender and emotion expression: a developmental contextual perspective. Emot Rev 2015 Jan;7(1):14-21. [doi: 10.1177/1754073914544408] [Medline: 26089983]
53. Lin Y, Ding H, Zhang Y. Gender differences in identifying facial, prosodic, and semantic emotions show category- and channel-specific effects mediated by encoder's gender. J Speech Lang Hear Res 2021 Aug 9;64(8):2941-2955. [doi: 10.1044/2021_JSLHR-20-00553] [Medline: 34310173]
54. Aung T, Puts D. Voice pitch: a window into the communication of social power. Curr Opin Psychol 2020 Jun;33:154-161. [doi: 10.1016/j.copsyc.2019.07.028] [Medline: 31445439]
55. Zhang J, Pan Z, Gui C, et al. Analysis on speech signal features of manic patients. J Psychiatr Res 2018 Mar;98:59-63. [doi: 10.1016/j.jpsychires.2017.12.012] [Medline: 29291581]
56. Guidi A, Salvi S, Ottaviano M, et al. Smartphone application for the analysis of prosodic features in running speech with a focus on bipolar disorders: system performance evaluation and case study. Sensors (Basel) 2015 Nov 6;15(11):28070-28087. [doi: 10.3390/s151128070] [Medline: 26561811]
57. Ozdas A, Shiavi RG, Silverman SE, Silverman MK, Wilkes DM. Investigation of vocal jitter and glottal flow spectrum as possible cues for depression and near-term suicidal risk. IEEE Trans Biomed Eng 2004 Sep;51(9):1530-1540. [doi: 10.1109/TBME.2004.827544] [Medline: 15376501]

XSL•FO
RenderX

58. Faurholt-Jepsen M, Rohani DA, Busk J, Vinberg M, Bardram JE, Kessing LV. Voice analyses using smartphone-based data in patients with bipolar disorder, unaffected relatives and healthy control individuals, and during different affective states. Int J Bipolar Disord 2021 Dec 1;9(1):38. [doi: 10.1186/s40345-021-00243-3] [Medline: 34850296]
59. Torous J, Bucci S, Bell IH, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct;20(3):318-335. [doi: 10.1002/wps.20883] [Medline: 34505369]
60. Balcombe L, De Leo D. Digital mental health challenges and the horizon ahead for solutions. JMIR Ment Health 2021 Mar 29;8(3):e26811. [doi: 10.2196/26811] [Medline: 33779570]
61. Soenksen LR, Ma Y, Zeng C, et al. Integrated multimodal artificial intelligence framework for healthcare applications. NPJ Digit Med 2022 Sep 20;5(1):149. [doi: 10.1038/s41746-022-00689-4] [Medline: 36127417]
62. Grünerbl A, Muaremi A, Osmani V, et al. Smartphone-based recognition of states and state changes in bipolar disorder patients. IEEE J Biomed Health Inform 2015 Jan;19(1):140-148. [doi: 10.1109/JBHI.2014.2343154] [Medline: 25073181]
63. Osmani V. Smartphones in mental health: detecting depressive and manic episodes. IEEE Pervasive Comput 2015;14(3):10-13. [doi: 10.1109/MPRV.2015.54]
64. Cohen J, Richter V, Neumann M, et al. A multimodal dialog approach to mental state characterization in clinically depressed, anxious, and suicidal populations. Front Psychol 2023;14:1135469. [doi: 10.3389/fpsyg.2023.1135469] [Medline: 37767217]
65. Or F, Torous J, Onnela JP. High potential but limited evidence: using voice data from smartphones to monitor and diagnose mood disorders. Psychiatr Rehabil J 2017 Sep;40(3):320-324. [doi: 10.1037/prj0000279] [Medline: 28891659]
66. Garcia-Ceja E, Riegler M, Nordgreen T, Jakobsen P, Oedegaard KJ, Tørresen J. Mental health monitoring with multimodal sensing and machine learning: a survey. Pervasive Mob Comput 2018 Dec;51:1-26. [doi: 10.1016/j.pmcj.2018.09.003]
67. McCoy LG, Brenna CTA, Chen SS, Vold K, Das S. Believing in black boxes: machine learning for healthcare does not need explainability to be evidence-based. J Clin Epidemiol 2022 Feb;142:252-257. [doi: 10.1016/j.jclinepi.2021.11.001] [Medline: 34748907]
68. Yerushalmi M, Sixsmith A, Pollock Star A, King DB, O'Rourke N. Ecological momentary assessment of bipolar disorder symptoms and partner affect: longitudinal pilot study. JMIR Form Res 2021 Sep 2;5(9):e30472. [doi: 10.2196/30472] [Medline: 34473069]
69. Clark EL, Easton C, Verdon S. The impact of linguistic bias upon speech-language pathologists' attitudes towards non-standard dialects of English. Clin Linguist Phon 2021 Jun 3;35(6):542-559. [doi: 10.1080/02699206.2020.1803405] [Medline: 32781853]
70. Bartoli F, Bachi B, Callovini T, et al. Anxious distress in people with major depressive episodes: a cross-sectional analysis of clinical correlates. CNS Spectr 2024 Feb;29(1):49-53. [doi: 10.1017/S1092852923002377] [Medline: 37489522]
71. Malgaroli M, Hull TD, Calderon A, Simon NM. Linguistic markers of anxiety and depression in somatic symptom and related disorders: observational study of a digital intervention. J Affect Disord 2024 May 1;352:133-137. [doi: 10.1016/j.jad.2024.02.012] [Medline: 38336165]
72. Bartoli F, Crocamo C, Clerici M, Carrà G. Allopurinol as add-on treatment for mania symptoms in bipolar disorder: systematic review and meta-analysis of randomised controlled trials. Br J Psychiatry 2017 Jan;210(1):10-15. [doi: 10.1192/bjp.bp.115.180281] [Medline: 27856422]
73. Bartoli F, Cavaleri D, Bachi B, et al. Repurposed drugs as adjunctive treatments for mania and bipolar depression: a meta-review and critical appraisal of meta-analyses of randomized placebo-controlled trials. J Psychiatr Res 2021 Nov;143:230-238. [doi: 10.1016/j.jpsychires.2021.09.018] [Medline: 34509090]
74. Bartoli F, Cavaleri D, Nasti C, et al. Long-acting injectable antipsychotics for the treatment of bipolar disorder: evidence from mirror-image studies. Ther Adv Psychopharmacol 2023;13:20451253231163682. [doi: 10.1177/20451253231163682] [Medline: 36994116]
75. Carrà G, Scioli R, Monti MC, Marinoni A. Severity profiles of substance-abusing patients in Italian community addiction facilities: influence of psychiatric concurrent disorders. Eur Addict Res 2006;12(2):96-101. [doi: 10.1159/000090429] [Medline: 16543745]

Abbreviations

BD: bipolar disorder
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
MADRS: Montgomery-Åsberg Depression Rating Scale
MAE: mean absolute error
mHealth: mobile health
MSE: mean squared error
NLP: natural language processing
RF: random forest
SCID-5: Structured Clinical Interview for DSM-5
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

XSL•FO
RenderX

WMD: word mover's distance
YMRS: Young Mania Rating Scale

Edited by A Mavragani; submitted 20.08.24; peer-reviewed by D Meyer, V Martin; revised version received 29.01.25; accepted 12.02.25; published 16.04.25.

Please cite as:
Crocamo C, Cioni RM, Canestro A, Nasti C, Palpella D, Piacenti S, Bartoccetti A, Re M, Simonetti V, Barattieri di San Pietro C, Bulgheroni M, Bartoli F, Carrà G
Acoustic and Natural Language Markers for Bipolar Disorder: A Pilot, mHealth Cross-Sectional Study
JMIR Form Res 2025;9:e65555
URL: https://formative.jmir.org/2025/1/e65555
doi:10.2196/65555

© Cristina Crocamo, Riccardo Matteo Cioni, Aurelia Canestro, Christian Nasti, Dario Palpella, Susanna Piacenti, Alessandra Bartoccetti, Martina Re, Valentina Simonetti, Chiara Barattieri di San Pietro, Maria Bulgheroni, Francesco Bartoli, Giuseppe Carrà. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Using the Healthy Native Youth Implementation Toolbox to Provide Web-Based Adolescent Health Promotion Decision Support to American Indian and Alaska Native Communities: Implementation Study

Amrita Sidhu[1*], MPH, MDS; Ross Shegog[1*], PhD; Stephanie Craig-Rushing[2*], PhD; Nicole Trevino[2*], MA; Michelle Singer[2*], BSc; Cornelia Jessen[3*], MA; Gwenda Gorman[4*], BSc; Sean Simpson[5*]; Melissa Peskin[1], PhD; Belinda Hernandez[1], PhD; Christine Markham[1*], PhD

[1]Department of Health Promotion and Behavioral Sciences, School of Public Health, UTHealth Houston, 7000 Fannin St, Houston, TX, United States

[2]Northwest Portland Area Indian Health Board, Portland, OR, United States

[3]Alaska Native Tribal Health Consortium, Anchorage, AK, United States

[4]Inter Tribal Council of Arizona, Inc, Phoenix, AZ, United States

[5]Good Medicine Tribal Public Health Consulting, Madison, WI, United States

[*]these authors contributed equally

Corresponding Author:
Amrita Sidhu, MPH, MDS
Department of Health Promotion and Behavioral Sciences, School of Public Health, UTHealth Houston, 7000 Fannin St, Houston, TX, United States

Abstract

Background: American Indian and Alaska Native (AI/AN) youth experience numerous health inequities, including those in sexual, reproductive, and mental health. Implementation of culturally relevant, age-appropriate evidence-based programs may mitigate these inequities. However, numerous barriers limit the adoption and implementation of evidence-based adolescent health promotion programs in AI/AN communities.

Objective: This study examines user reach and engagement from 2022 to 2024 of web-based decision support (the Healthy Native Youth [HNY] website and the embedded HNY Implementation Toolbox), designed to increase the implementation of evidence-based adolescent health promotion programming in AI/AN communities.

Methods: Promotional strategies were designed for optimal geographic reach to Tribal organizations, opinion leaders, federal decision makers, and funders. Promotional channels included grassroots, community, and professional networks. We used Google Analytics to examine the uptake of the HNY website and HNY Implementation Toolbox from January 2022 to January 2024. The Toolbox provides culturally relevant tools and templates to help users navigate through 5 phases of program adoption and implementation: Gather, Choose, Prepare, Implement, and Grow. User reach was estimated by demographic characteristics and geographic location; user engagement was estimated by visit frequency and duration, bounce rates, and frequency of page and tool access.

Results: Over the study period, page views of the HNY website and HNY Toolbox increased 10-fold and 27-fold, respectively. Over the 2-year evaluation period since the Toolbox "go live" date, approximately 1 in 8 users of the HNY website visited the Toolbox. The majority of HNY website users were located in Washington (n=1515), California (n=1290), and Oregon (n=1019) and were aged between 18 and 24 (n=1559, 21.7%) and 25 - 34 (n=1676, 23.29%) years. Toolbox users were primarily located in California (n=1238), Washington (n=1142), and Oregon (n=986), mostly aged between 35 and 44 years (n=444, 35%). Both website and Toolbox users were primarily female, who accessed the site and Toolbox via desktop computers. The most frequently accessed phase pages within the Implementation Toolbox were Gather, Choose, Implement, and Prepare, as supported by bounce rates and average time on page. The most viewed phase was the "Gather" phase, with 3278 views. The most frequently downloaded tools within the Toolbox were Gather: Community Needs and Resource Assessment, with 136 downloads. The phases and tools accessed may have differed based on the user's goal or stage of implementation.

Conclusions: Findings indicate positive initial reach and engagement of the HNY website and HNY Implementation Toolbox among AI/AN educators that has consistently increased over the 2 years. The provision of web-based decision support that guides AI/AN users through the adoption, implementation, and maintenance of culturally relevant, age-appropriate, evidence-based

XSL•FO
RenderX

adolescent health promotion programs in their communities may help increase the implementation of effective adolescent health promotion programs to ultimately increase health equity among AI/AN youth.

(JMIR Form Res 2025;9:e67885) doi:10.2196/67885

KEYWORDS

implementation; culturally relevant program; evidence-based health promotion; user engagement; reach; decision support system; American Indian; Alaska; native communities; youth; adolescent; decision support; Alaska native; health inequities; sexual; reproductive; mental health; AI/AN; Tribal organization; Google Analytics; toolbox

Introduction

In the United States, as of 2022, the federal government recognizes 574 distinct American Indian and Alaska Native (AI/AN) Tribes across 37 states [1]. Together, the combined AI/AN population comprises 9.7 million people, representing 2.9% of the US population in 2020 [1]. The AI/AN population is young, with 30% being up to 18 years old, compared with 24% of the US total population [2], making adolescent health an important topic in native communities. However, AI/AN youth face many health inequities compared with other racial/ethnic groups, including a disproportionate burden of adverse mental, sexual, and reproductive health, and violence-related outcomes [3-5].

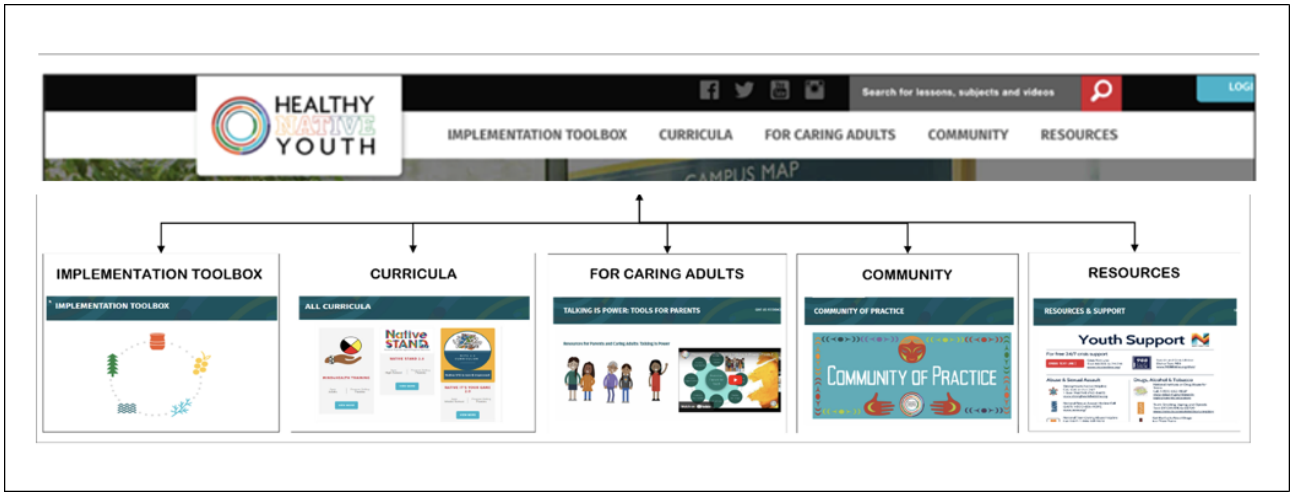
Evidence-based programs (EBPs) that promote adolescent health may help mitigate these inequities among AI/AN teens [6]. An increasing number of EBPs incorporating the strengths and cultural teachings of native communities have been developed or adapted for AI/AN youth [7-13]. Despite efforts to increase access to culturally relevant EBPs, barriers exist that limit their adoption and implementation by AI/AN educators [14-16]. Native communities may lack implementation readiness to discuss and approach adolescent health, experience difficulty navigating Tribal review and school board approval processes [16], have inadequate funding or program integration that compromises program sustainability, or be geographically isolated with limited access to programs [15,17]. Lack of administrative and parental support, lack of trained personnel, and lack of program integration with cultural values represent additional barriers to the implementation of EBPs in this population [15].

The Healthy Native Youth (HNY) website was developed in 2016 in collaboration with AI/AN community advisors to mitigate barriers to access and to enhance the adoption and implementation of culturally relevant adolescent health EBPs in native communities to promote health equity [6,14,18,19] (Figure 1). Digital platforms are an accepted delivery channel to enable greater reach of evidence-based interventions to underserved populations at lower costs [20-23]. Digital health interventions for AI/AN youth have been reported to contribute to enhanced health metrics such as mental health, substance use, and sexual risk reduction [10,24]. Prior to the HNY Website, such programs need to be located and accessed individually, without structured dissemination and implementation decision support. The HNY website provides a "one-stop-shop" for Tribal health advocates to access engaging, culturally relevant, age-appropriate health curricula and programs for AI/AN youth within a single website structured to provide implementation decision support [14]. The website comprises 5 sections: "Implementation Toolbox" (guidance for adoption and implementation of EBPs), "Curricula" (information about culturally informed EBPs), "For Caring Adults" (empowerment for nurturing adult-youth communication), "Community" (networking for advocates of adolescent health), and "Resources" (links and materials for youth advocates). The Curricula section of the HNY website contains comparative information on the content, specifications, and access to 19 culturally relevant, age-appropriate evidence-based for AI/AN youth focused on multiple topics, including mental health, sexual health, substance use, and suicide prevention. The user can filter and compare curricula on multiple dimensions to determine "best-fit" for their community and includes implementation materials at no cost.

The HNY Implementation Toolbox, housed within the HNY website is maintained and owned by the Northwest Portland Area Indian Health Board (NPAIHB) and provides stepped web-based decision support designed to facilitate the adoption and implementation of adolescent health EBPs in native communities (Figure 1). The Toolbox was developed on WordPress, an open-source content management system using an iterative design process informed by native practitioners and academicians to adapt an existing web-based decision support system (iCHAMPSS), which is described elsewhere [17,25]. Digital decision support can provide web-based expert guidance to facilitate complex tasks [26]. Findings from a 6-month user experience study with 25 AI/AN educators rated the Toolbox as credible, easy to use, and more helpful than current resources [17]. In the 6-month postsurvey, participants indicated increased knowledge, advocacy skills, and self-efficacy to adopt, implement, and maintain culturally relevant adolescent health promotion programs in their community [17]. The Toolbox launched in January 2022 and the sustainability of the program is through ongoing grant funding through the aforementioned NPAIHB. National Institutes of Health R21 funds (approximating US $440,000 total costs; US $164,000 indirect costs) were sufficient to develop and formatively assess the HNY Website Toolbox.

XSL•FO
RenderX

Figure 1. Roadmap of the Healthy Native Youth website landing page (top) and landing pages for each section (bottom) for access by American Indian and Alaska Native (AI/AN) communities.

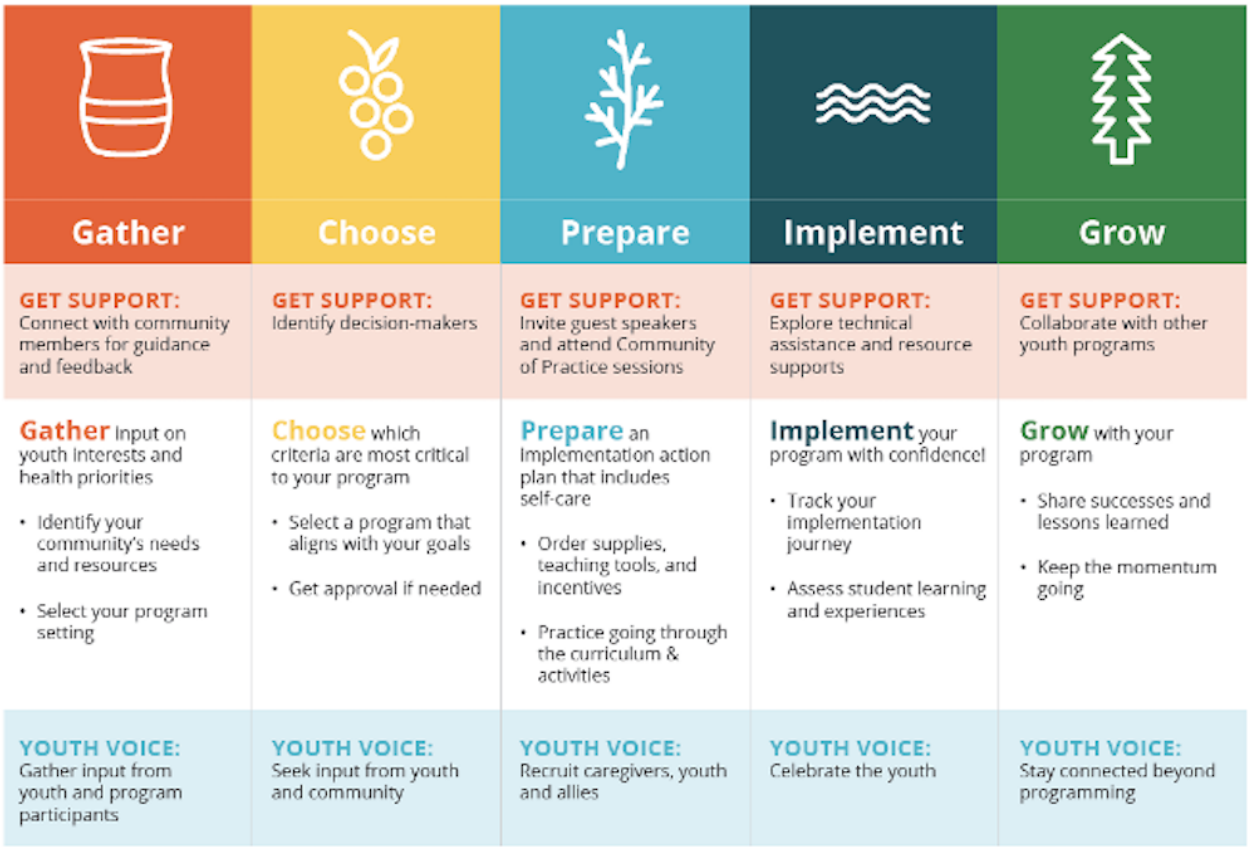


The Toolbox is designed to guide users through five culturally adapted dissemination and implementation phases: (1) Gather, (2) Choose, (3) Prepare, (4) Implement, and (5) Grow. (Figure 2) [17]. The "Gather" phase guides users to identify the health priorities and interests of youth in their community. "Choose" guides users to select a culturally relevant and age-appropriate evidence-based health program for their community. "Prepare" guides users to implement the selected program, including facilitating support from key decision makers. "Implement" guides users to deliver their selected program while emphasizing listening, reflection, and feedback. Finally, "Grow" guides users to sustain their selected program [17]. Each phase includes strategies to generate support for community members, elders, and youth throughout the planning process. The Toolbox comprises a "resource tools library" with over 20 phase-specific support tools, templates, helpful tips, links to web-based resources, and stories from the field (testimonials from experienced native health educators) [17]. All rights, including copyright, for this website and content on this Site are owned by or licensed to the NPAIHB.

The purpose of this study was to examine user reach and engagement with the HNY website and the embedded HNY Implementation Toolbox during an initial 2-year roll-out of the Toolbox from January 2022 to January 2024. Study findings of the reach of these web-based decision support strategies can inform further strategies to increase dissemination and implementation and provide a foundation for future hybrid effectiveness-implementation studies.

XSL•FO
RenderX

Figure 2. The 5 culturally adapted dissemination and implementation phases and associated tasks in the Healthy Native Youth Implementation Toolbox.



Methods

HNY Website and Implementation Toolbox Dissemination

Outreach dissemination strategies over the 2-year period encompassed channels through grassroots, community, and professional networks. These information channels comprised monthly AI/AN Community of Practice Network meetings highlighting the HNY Website and Toolkit for professionals who plan, manage, and facilitate healthy youth programs in Indian country, social media posts through existing networks in the AI/AN community devoted to health messages and event promotion, field studies with AI/AN practitioners and community members and related manuscript publication, presentations at professional conferences (eg, American Public Health Association, Department of Health and Human Services Office of Population Affairs Teen Pregnancy Prevention) and Tribal meetings (eg, The Tribal Public Health Conference) reporting formative, outcome, and process data from the HNY Toolbox, and presentations on varied health topics that promoted the HNY website as a resource. Rollouts were planned to ensure optimal breadth (geographically) and depth of exposure (including local Tribal organizations, opinion leaders, national Federal decision makers, and funders).

Data Extraction and Analysis

Data from the HNY website and the embedded HNY Implementation Toolbox were assessed using Google Analytics collected from January 2022 to January 2024. Google Analytics is a web analytics service that tracks and reports website traffic and user behavior [27,28]. This tool is used extensively by digital marketers and businesses to test and improve digital marketing performance, assess customer behavior, and identify strategies to optimize and improve traffic to a website by providing insight into user behavior.

Data were collected to assess reach, user demographics, and user engagement through a real-time, interactive dashboard accessible through a registered account on the Google Analytics website [29].

Reach is the extent to which an intervention or program is delivered to the priority population and was assessed as the number of users from each US state [29]. It is a necessary antecedent for community exposure to intervention content and any opportunity to establish effectiveness [29,30]. A user was defined as a visitor to the sites who had an assigned unique identifier with one or more associated sessions [29]. User demographics were assessed by geographic location, user gender and age, and the type of device used to access the sites. User engagement was assessed by the number of page views, number of pages viewed per session, number of "tools" downloaded, average duration of sessions, bounce rate, and exit rate. Page views were assessed as the number of separate views each page received. A session was defined as a user visit to, and interaction with, the webpage (eg, button clicks, downloads, and page navigation). Pages per session were assessed as the number of web pages of the site that the user viewed in a single session [29]. Duration of a session was assessed as the average length of time in minutes that users spent on a website/webpage during

XSL•FO
RenderX

a session [29]. Bounce rate, or nonengagement rate, was assessed as the percentage of single-page sessions a specific webpage received. If a user loaded a page but did not interact a second time (by viewing another page) within 30 minutes, this was considered a bounce [29]. For example, a bounce rate of 50% means that half the users engaged with the website on the visited page only. Exit rates were assessed as the percentage of users that left the website from a webpage (ie, the webpage was the last page visited by them before they exited the website) [29]. For example, an exit rate of 20% means that 20% of users left the website from this page. Reporting bounce rates and exit rates assess if the website/tool met the expectations of a user. If these are lower than industry standards then the site is considered to be engaging [29]. Downloads were assessed by user selections of a download option for a given resource template.

Ethical Considerations

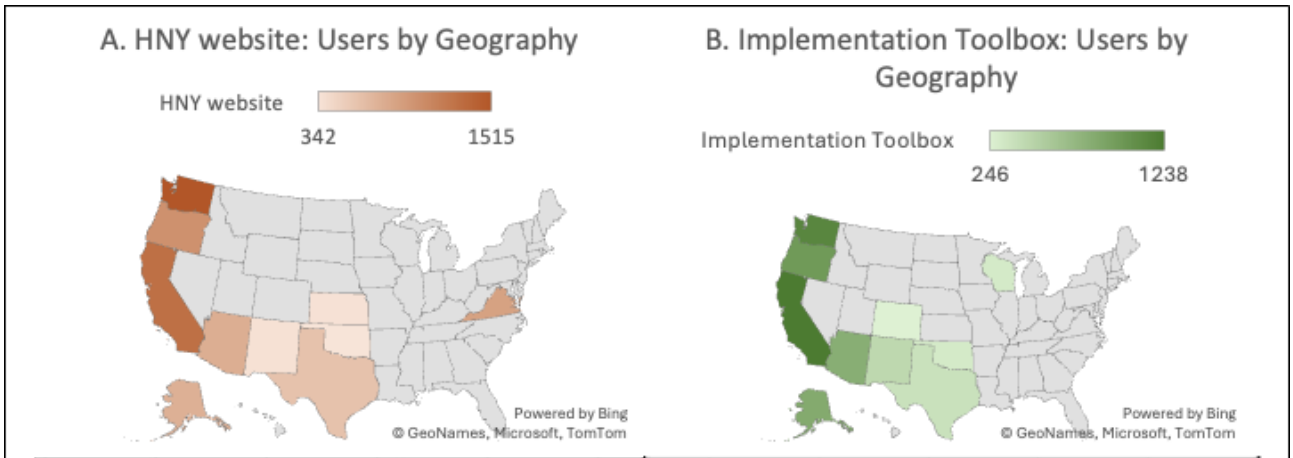
This study was reviewed and approved by the Committee for the Protection of Human Subjects at the University of Texas Health Science Center at Houston (institutional review board number: HSC-SPH-18 - 0958). The HNY website also outlines its privacy policy in the footer section of the website as a means of informed consent from users and describes how user data are collected and may be used for research purposes. All data used are deidentified and anonymous. This manuscript, including related images and supplementary material, does not allow the identification of individual participants or users. No compensation was offered to participants as this paper reports analytics of a web-based resource that is accessible free of cost.

Results

HNY Website Analytics

From January 2022 to January 2024 the HNY website had 10,237 total users. Most users were located in Washington (n=1515), California (n=1290), and Oregon (n=1019) (Figure 3A). The majority of users identified as female (n=5217, 70.0%), were mainly 25 - 34 years (n=1676, 23.0%) or 18 - 24 (n=1559, 22.0%) years of age, and accessed the website primarily via desktop (80.0%) and mobile (19.0%) devices (Table 1). The most frequently accessed sections (in decreasing order) were Curricula (9902 views), Resources (6397 views), HNY Implementation Toolbox home (4364 views), Community (4054 views), and For Caring Adults (3766 views) (Table 2). Page duration ranged from 1.7 minutes (Implementation Toolbox) to 3.06 minutes (Curricula), bounce rate from 50.2% (Curricula) to 65.5% (Community), and exit rates from 15.9% (Curricula) to 60.0% (For Caring Adults). Access to the HNY website increased from 2479 page views in the first quarter of 2022 to a cumulative total of 33,004 page views by the fourth quarter of 2023 (Figure 4).

Figure 3. Geographic distribution (top 10 US states) and number of users by state who visited the HNY Website (left column) and HNY Implementation Toolbox (right column) from January 1, 2022, to January 1, 2024.



State of origin	User (n)
Washington	1515
California	1290
Oregon	1019
Virginia	884
Arizona	779
Alaska	764
Texas	617
New Mexico	366
Kansas	362
Oklahoma	342

State of origin	User (n)
California	1238
Washington	1142
Oregon	986
Alaska	847
Arizona	815
New Mexico	456
Texas	382
Oklahoma	307
Wisconsin	307
Colorado	246

XSL•FO
RenderX

Table . User demographics (gender, age, and device used) for the Healthy Native Youth (HNY) website (left column) and HNY Implementation Toolbox (right column) from January 2022 to January 2024.

	HNY website	HNY Implementation Toolbox
Users by gender, n (%)		
Male	2213 (30.0)	244 (34.0)
Female	5217 (70.0)	472 (66.0)
Users by age (years), n (%)		
18 - 24	1559 (22.0)	253 (20.0)
25 - 34	1676 (23.0)	319 (25.0)
35 - 44	1638 (23.0)	444 (35.0)
45 - 54	1216 (17.0)	135 (11.0)
55 - 64	690 (9.0)	82 (6.0)
65 and older	416 (6.0)	39 (3.0)
Users by device (%)[a]		
Mobile	19.0	19.0
Tablet	1.0	0
Desktop	80.0	81.0

[a] During the time period of this study, the number of users by device was not a part of the analytic dashboard. Therefore, only percentage values have been reported.

Table . Most frequently accessed sections in the Healthy Native Youth (HNY) website from January 2022 to January 2024 (Google Analytic data from January 2022 to January 2024).

Section[a]	Total users, n	Page views, n	Page views/session[b]	Page duration[b] (min)	Bounce rate[c], %	Exit rate, %
HNY website home	10,237	38,009	2.4	2.75	40.7	19.6
Curricula	3494	9902	3.8	3.06	50.2	15.9
Resources	2849	6397	2.51	2.83	64.6	40.3
Implementation Toolbox home	2566	4364	1.8	1.71	59.8	35.4
Community	2486	4054	1.8	1.3	57.4	60.0
For caring adults	1727	3766	3.64	2.0	65.5	36.6

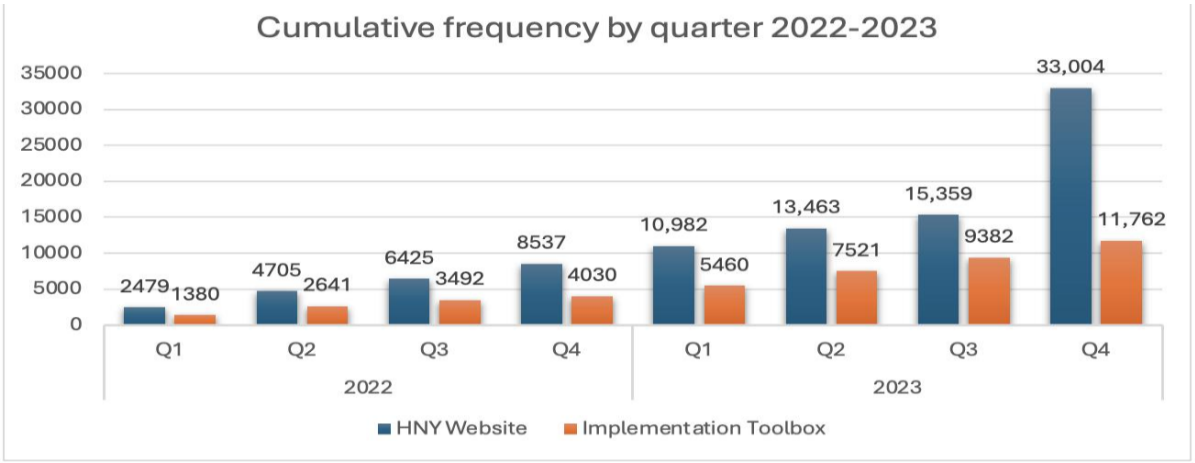
[a]Corresponds to top-level sections in the HNY website (Figure 1).

[b]Average.

[c]Nonengagement rate.

XSL•FO
RenderX

Figure 4. Cumulative user frequency over time by quarter from Jan 2022 to Jan 2024 and associated dissemination strategies (outreach events). AI/AN: American Indian and Alaska Native; HNY: Healthy Native Youth; NA: not available; Q: quarter.



Year of evaluation	2022				2023			
Quarter	1	2	3	4	1	2	3	4
Months	Jan - Mar	Apr - Jun	Jul - Sept	Oct - Dec	Jan - Mar	Apr - Jun	Jul - Sept	Oct - Dec
Website page views (n)	2479	2226	1918	2112	2395	2,531	1,896	17,645
Website page views cumulative	2479	4705	6425	8537	10,932	13,463	15,359	33,004
Toolbox page views	1380	1261	851	538	1430	2061	1861	2380
Toolbox page views cumulative	1380	2641	3492	4030	5460	7521	9382	11,762
Dissemination strategy								
AI/AN Community of Practice Network Meetings (Date, Attendees)	NA	NA	NA	NA	NA	NA	Sept. (n=5)	Nov. (n=30) Oct. (n=13)
Social media (Instagram) posts.	14	4	5	3	6	6	3	6
HNY presentations promoting the HNY Website and HNY Toolbox at professional meetings and Tribal gatherings	1	0	4	4	5	3	0	1
Field study - Formative and usability studies.	x		x	x				
Publication				x				

HNY Implementation Toolbox Analytics

From January 2022 to January 2024 the Implementation Toolbox had a total of 2566 total users (Table 3) and 11,762 page views (Figure 4). Most users were located in California (n=1238), Washington (n=1142), and Oregon (n=986) (Figure 3B). The majority of users identified as female (n=472, 66%), were mainly 35 - 44 years (n=444, 35%) or 25 - 34 (n=319, 25%) years of age and accessed the Toolbox mainly via desktop (81%) and mobile (19%) devices (Table 1). Phase page views (in decreasing order) were Gather (3278 views), Choose (1275 views), Prepare (907 views), Implement (843), and Grow (618 views) (Table 3). Page duration ranged from 1.67 minutes (Grow) to 2.08 minutes (Choose), bounce rate from 41.8% (Gather) to 72.9% (Prepare), and exit rates from 22.1% (Prepare) to 34.5% (Grow). The most frequently downloaded tools within the Toolbox were from the Gather phase and included the "Community needs and resource assessment" factsheet (136 views), the "Identify youth advocates and community partners" factsheet (111 views), and a "Bingo data collection" activity guide (83 views) (Table 4). Access to the HNY Implementation Toolbox increased from 1380 page views in the first quarter of 2022 to a cumulative total of 11,762 page views by the fourth quarter of 2023 in conjunction with dissemination activities of scientific presentations, community of practice sessions, social media posts, and webinars (Figure 4).

XSL•FO
RenderX

Table . Phases of the Healthy Native Youth (HNY) Implementation Toolbox accessed from January 2022 to January 2024 across the United States by user metrics.

Implementation phase	Total users, n	Page views, n	Page views/session[a]	Page duration[a] (min)	Bounce rate, %	Exit rate, %
Implementation Toolbox landing page	2566	4364	1.8	1.71	60.3	35.5
Gather	1863	3278	16.7	2.07	41.8	33.9
Choose	876	1275	12.3	2.08	49.5	22.3
Prepare	575	907	27.1	1.40	72.9	22.1
Implement	488	843	7.8	1.7	64.0	30.8
Grow	353	618	5.7	1.66	57.8	34.5

[a]Average.

Table . Top 10 most frequently downloaded tools from the Healthy Native Youth (HNY) Implementation Toolbox by phase, type, and name from January 2022 to January 2024 across the United States.

#	Implementation phase	Tool type	Tool name	Downloads, n
1	Gather	Factsheet	Community needs and resource assessment	136
2	Gather	Factsheet	Identify youth advocates and community partners	111
3	Gather	Activity sheet	Bingo data collection activity	83
4	Prepare	Template	Curriculum implementation plan template	76
5	Gather	Checklist	Site selection	72
6	Implement	Log	Session reflection log template	44
7	Prepare	Inventory list	Example list of materials to order for native STAND	40
8	Implement	Log	Class attendance sheet template	30
9	Choose	Factsheet	Identify youth advocates and community partners	29
10	Gather	Factsheet	Alaska Tribal Health Consortium adolescent health action plan	28

Discussion

Principal Findings

The HNY website and embedded HNY Implementation Toolbox provide web-based resources to support the adoption, implementation, and maintenance of culturally relevant, age-appropriate evidence-based adolescent health promotion programs in AI/AN communities. The HNY website is a "one-stop-shop" for Tribal health advocates to access culturally relevant and age-appropriate health curricula, programs, and related resources for AI/AN youth [14]. The embedded HNY Implementation Toolbox provides guidance to enable them to successfully implement these curricula in their communities. Over the 2-year evaluation period since the Toolbox's "go live" date, approximately 1 in 8 users of the HNY website visited the Toolbox. This, and expanded geographic reach, corroborate the previously reported utility of the Toolbox [17]. The differential traffic and reach between the HNY website and the Toolbox may reflect the comparative maturity of the HNY website compared with the relatively newer embedded Toolbox. The greater reach in the western states is expected because the HNY website was developed and is housed in the NPAIHB located in Washington State, and early promotion was among affiliated Tribes in Washington State, Oregon, and Idaho. Demographically, the Toolbox users were older, with 35% (n=444) being 35 - 44 years of age. This is likely a reflection of the demographics of community health advocates who are evaluating community needs, facilitating decision-making on the curricula best addressing these needs, and actively engaging community support to adopt and implement these curricula in their communities. This, and whether users had attended Tribal

XSL•FO
RenderX

health conferences or Community of Practice gatherings in which the Toolbox was promoted could not be verified from the Google data.

Over the evaluation period, page views of the HNY Website and the HNY Toolbox increased 10-fold and 9-fold, respectively. Page views to the Toolbox witnessed a steady increase from 2022 to 2023. The reason for this is uncertain but is possibly attributable to the Toolbox being showcased in monthly Community of Practice meetings that gather dozens of advocates for Tribal youth in Indian country. This suggests the potential for embedded practitioner networks as a particularly utilitarian strategy to disseminate youth programs [27,28]. Given that the HNY website and Implementation Toolbox coexist, the promotion of one likely promotes traffic to the other. This is the strength of a collaborative approach that provides multiple resources that address a particular theme, whether it be a priority population, setting, or health domain. The cumulative page views over time are also suggestive of the need to provide a persistent and longitudinal promotional strategy. We will continue to build awareness nationally via presentations, workshops, and round table discussions at Tribal, education, and public health conferences.

The Toolbox's average bounce rate (57.7%) was lower than that of the HNY website (62.1%) and also the industry average (62.3%); the average number of Toolbox pages visited per session was almost 5 times than that of the industry average (11.9 vs 2 pages) [14,31,32]. Further, the duration of use on both sites (2.4 and 1.8 respectively) are close to averages on comparable websites (2.58) [33]. These findings suggest that the Toolbox is "sticky," and that users are finding material commensurate with their needs. This is also supportive of previously reported positive usability ratings for the site from AI/AN community decision makers.

User preferences within the Toolbox were for the Gather Phase and tools related to this phase that enable needs assessment and identification of community advocates. This may be indicative of Tribal communities being in this early phase of adopting and implementing adolescent health programs, that these tools are particularly useful for the work of health educators, or simply that community members are likely to start at the first step in their review and exploration of a novel support tool like the Toolbox. Future studies can investigate if progression through stages occurs over time, the association of phase use with "in-situ" community implementation, as well as challenges to implementation that can be a function of differing state or local policies, regional internet accessibility, funding, community readiness to adopt programs, and logistics of obtaining school board approvals [16].

Study findings need to be interpreted in the context of several methodologic limitations and biases due to data collection. Google Analytics was originally intended for marketing and sales purposes and is limited largely to descriptive data [34,35]. Guidelines for the use and interpretation of Google Analytics do not exist. Discrepancies in the data are possible due to a new client ID being assigned every time a user deletes cookies or signs in from a different device or browser [36]. This may produce an inaccurate count of "new" users because different IDs could be allocated to the same user [34]. Additionally, deletion of cookies may degrade data captured via Google Analytics which relies on user cooperation for successful data collection [20]. Google Analytics assigns a value of 0 seconds if a user visits a webpage until they navigate to a second page, which can lead to inaccurate page duration assessments [34]. Google Analytics does not collect demographic and contextual data on ethnicity, Tribe, or the context of the application of the HNY website and Toolbox. The gender of users captured by Google Analytics may only reflect a subset of users that log into their Google accounts when visiting the page and it is difficult to determine how the demographics align with the target population [34]. Page views and resource downloads do not necessarily equate to the actual adoption and implementation of programs. There is a risk of selection bias wherein the analytics reported may only reflect a subset of users that log into their Google accounts The purpose of use, whether in the context of seeking guidance on program implementation, review and feedback for research, or simply to assuage curiosity, was not ascertained in this study and will need to be if the utility of the HNY Website and Toolbox is to be fully understood. To address the above biases and limitations, reporting using Google Analytics may be applied in combination with other field data collection (qualitative or quantitative) to triangulate data sources. This study reported on an array of promotional strategies to achieve increased dissemination. Future studies, perhaps using a mixed methods design, are recommended to investigate their comparative effectiveness.

Despite limitations, the use of Google Analytics to assess and gather feedback on web-based decision support and to track user behavior on multiple devices is well accepted [20]. Advantages are that it can be used for self-comparison over time and provide a comparison to industry standards. This can provide a positive feedback loop wherein higher website traffic builds positive normative perceptions that generate further user traffic through "word of mouth." An assessment of reach represents an early phase in understanding program dissemination and implementation that provides the foundation for future hybrid effectiveness-implementation evaluation, including network analysis to assess the dynamics of community diffusion.

Conclusions

The HNY website and HNY Implementation Toolbox provide AI/AN educators and youth advocates with step-by-step guidance to support the adoption, implementation, and maintenance of culturally relevant, age-appropriate evidence-based adolescent health promotion programs in their communities. Google Analytics provides insight into the use and reach of the HNY website and Implementation Toolbox among AI/AN youth advocates, which will help inform ongoing dissemination efforts. Future studies are needed to assess the Toolbox's actual impact on the uptake and implementation of adolescent health promotion EBPs in native communities.

XSL•FO
RenderX

Acknowledgments

Funding for this study was received from the National Institutes of Health Office of Minority Health (1R21MD01396001A1) and the Indian Health Service HIV Program and the Minority AIDS Initiative Fund (contract number 248-96-0011).

Data Availability

The datasets generated or analyzed during this study are not publicly available because the data could contain identifiable information linked to the users, but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Federally recognized American Indian tribes and Alaska Native entities. USA.gov. URL: https://www.usa.gov/indian-tribes-alaska-native [accessed 2025-04-02]
2. Facts for features: American Indian and Alaska Native heritage month: November 2021. US Census Bureau. 2022. URL: https://www.census.gov/newsroom/facts-for-features/2021/aian-month.html [accessed 2022-02-27]
3. Trends in teen pregnancy and childbearing. Office of Population Affairs. URL: https://opa.hhs.gov/adolescent-health/reproductive-health-and-teen-pregnancy/trends-teen-pregnancy-and-childbearing [accessed 2024-03-17]
4. Allen J, Wexler L, Rasmus S. Protective factors as a unifying framework for strength-based intervention and culturally responsive American Indian and Alaska Native suicide prevention. Prev Sci 2022 Jan;23(1):59-72. [doi: 10.1007/s11121-021-01265-0] [Medline: 34169406]
5. Clayton HB, Kilmer G, DeGue S, et al. Dating violence, sexual violence, and bullying victimization among high school students -youth risk behavior survey, United States, 2021. MMWR Suppl 2023 Apr 28;72(1):66-74. [doi: 10.15585/mmwr.su7201a8] [Medline: 37104527]
6. Morrison-Beedy D, Mazurek Melnyk B. Making a case for integrating evidence-based sexual risk reduction and mental health interventions for adolescent girls. Issues Ment Health Nurs 2019 Nov 2;40(11):932-941. [doi: 10.1080/01612840.2019.1639087]
7. Spencer LM, Schooley MW, Anderson LA, et al. Seeking best practices: a conceptual framework for planning and improving evidence-based practices. Prev Chronic Dis 2013 Dec 12;10:E207. [doi: 10.5888/pcd10.130186] [Medline: 24331280]
8. Craig Rushing S, Gaston A, Kaufman C, Markham C, Jessen C, Gorman G, et al. Using technology to promote health and wellbeing among American Indian and Alaska Native teens and young adults. In: Dyson LE, Grant S, Hendricks M, editors. Indigenous People and Mobile Technologies: Taylor & Francis Group; 2016:163-178.
9. Tingey L, Chambers R, Patel H, et al. Prevention of sexually transmitted diseases and pregnancy prevention among Native American youths: a randomized controlled trial, 2016–2018. Am J Public Health 2021 Oct;111(10):1874-1884. [doi: 10.2105/AJPH.2021.306447]
10. Shegog R, Craig Rushing S PhD, MPH, Jessen C MA, et al. Native IYG: improving psychosocial protective factors for HIV/STI and teen pregnancy prevention among youth in American Indian/Alaska Native communities. J Appl Res Child 2017;8(1). [doi: 10.58464/2155-5834.1320]
11. Hafner SP, Craig Rushing S. Sexual health, STI and HIV risk, and risk perceptions among American Indian and Alaska Native emerging adults. Prev Sci 2019 Apr;20(3):331-341. [doi: 10.1007/s11121-018-0920-7] [Medline: 30006906]
12. Skye M, McCoy T, Kelley A, et al. Effectiveness of Native STAND: a five-year study of a culturally relevant sexual health intervention. Adolescents 2021;1(3):321-334. [doi: 10.3390/adolescents1030024]
13. Kaufman CE, Whitesell NR, Keane EM, et al. Effectiveness of circle of life, an HIV-preventive intervention for American Indian middle school youths: a group randomized trial in a Northern Plains tribe. Am J Public Health 2014 Jun;104(6):e106-e112. [doi: 10.2105/AJPH.2013.301822]
14. Craig Rushing S, Stephens D, Shegog R, et al. Healthy Native Youth: improving access to effective, culturally-relevant sexual health curricula. Front Public Health 2018;6:225. [doi: 10.3389/fpubh.2018.00225]
15. Sacca L, Shegog R, Hernandez B, et al. Barriers, frameworks, and mitigating strategies influencing the dissemination and implementation of health promotion interventions in indigenous communities: a scoping review. Implement Sci 2022 Feb 21;17(1):18. [doi: 10.1186/s13012-022-01190-y] [Medline: 35189904]
16. Markham C, Torres J, Rushing SC, Gorman G, Jessen C, Gaston A, et al. Usability and psychosocial impact of decision support to increase sexual health education in American Indian and Alaska Native communities. J Health Dispar Res Pract 2018;11(1) [FREE Full text]
17. Markham CM, Rushing SC, Manthei J, et al. The Healthy Native Youth Implementation Toolbox: using implementation mapping to adapt an online decision support system to promote culturally-relevant sexual health education for American Indian and Alaska Native youth. Front Public Health 2022;10:889924. [doi: 10.3389/fpubh.2022.889924] [Medline: 36388328]

XSL•FO
RenderX

18. Peskin MF, Hernandez BF, Markham C, et al. Sexual health education from the perspective of school staff: implications for adoption and implementation of effective programs in middle school. J Appl Res Child 2011;2(2). [doi: 10.58464/2155-5834.1046]
19. Hernandez BF, Peskin M, Shegog R, et al. Choosing and maintaining programs for sex education in schools: the CHAMPSS model. J Appl Res Child 2011;2(2). [doi: 10.58464/2155-5834.1047]
20. Marcu G, Ondersma SJ, Spiller AN, Broderick BM, Kadri R, Buis LR. Barriers and considerations in the design and implementation of digital behavioral interventions: qualitative analysis. J Med Internet Res 2022 Mar 30;24(3):e34301. [doi: 10.2196/34301] [Medline: 35353043]
21. Marcu G, Ondersma SJ, Spiller AN, Broderick BM, Kadri R, Buis LR. The perceived benefits of digital interventions for behavioral health: qualitative interview study. J Med Internet Res 2022 Mar 30;24(3):e34300. [doi: 10.2196/34300] [Medline: 35353045]
22. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336. [doi: 10.1002/wps.20673] [Medline: 31496095]
23. Nelson LA, Spieker AJ, Mayberry LS, McNaughton C, Greevy RA. Estimating the impact of engagement with digital health interventions on patient outcomes in randomized trials. J Am Med Inform Assoc 2021 Dec 28;29(1):128-136. [doi: 10.1093/jamia/ocab254] [Medline: 34963143]
24. Toombs E, Kowatch KR, Dalicandro L, McConkey S, Hopkins C, Mushquash CJ. A systematic review of electronic mental health interventions for Indigenous youth: results and recommendations. J Telemed Telecare 2021 Oct;27(9):539-552. [doi: 10.1177/1357633X19899231] [Medline: 31937199]
25. Peskin MF, Hernandez BF, Gabay EK, et al. Using intervention mapping for program design and production of iCHAMPSS: an online decision support system to increase adoption, implementation, and maintenance of evidence-based sexual health programs. Front Public Health 2017;5:203. [doi: 10.3389/fpubh.2017.00203] [Medline: 28848729]
26. Bhargava HK, Power DJ, Sun D. Progress in web-based decision support technologies. Decis Support Syst 2007 Aug;43(4):1083-1095. [doi: 10.1016/j.dss.2005.07.002]
27. Noar AP, Jeffery HE, Subbiah Ponniah H, Jaffer U. The aims and effectiveness of communities of practice in healthcare: a systematic review. PLoS One 2023;18(10):e0292343. [doi: 10.1371/journal.pone.0292343] [Medline: 37815986]
28. Bodison SC, Sankaré I, Anaya H, et al. Engaging the community in the dissemination, implementation, and improvement of health-related research. Clin Transl Sci 2015 Dec;8(6):814-819. [doi: 10.1111/cts.12342] [Medline: 26546337]
29. Balasubramanian BA, Fernald D, Dickinson LM, et al. REACH of interventions integrating primary care and behavioral health. J Am Board Fam Med 2015;28 Suppl 1(Suppl 1):S73-S85. [doi: 10.3122/jabfm.2015.S1.150055] [Medline: 26359475]
30. Reilly KL, Kennedy S, Porter G, Estabrooks P. Comparing, contrasting, and integrating dissemination and implementation outcomes included in the RE-AIM and implementation outcomes frameworks. Front Public Health 2020;8:430. [doi: 10.3389/fpubh.2020.00430] [Medline: 32984239]
31. Average bounce rate by industry 2024 [benchmark]. Capturly. URL: https://capturly.com/blog/average-bounce-rate-by-industry-2023-benchmark [accessed 2024-07-22]
32. Pages per session: a measure of engagement. Analyzify. URL: https://analyzify.com/hub/pages-per-session-a-measure-of-engagement [accessed 2024-07-22]
33. Average session duration: industry benchmarks. First Page Sage. URL: https://firstpagesage.com/reports/average-session-duration-by-industry [accessed 2024-07-22]
34. Analytics tools & solutions for your business. Google Marketing Platform. URL: https://marketingplatform.google.com/about/analytics [accessed 2024-07-22]
35. Henström M, Duncanson K, Collins CE, Ashton LM, Davidson E, Ball R. Online reach and engagement of a child nutrition peer-education program (PICNIC): insights from social media and web analytics. BMC Public Health 2022 Apr 26;22(1):836. [doi: 10.1186/s12889-022-13252-3] [Medline: 35473610]
36. Client ID vs user ID. Analytics Mania. URL: https://www.analyticsmania.com/post/google-analytics-client-id-vs-user-id [accessed 2024-07-22]

Abbreviations

AI/AN: American Indian and Alaska Native
EBP: evidence-based program
HNY: Healthy Native Youth
NPAIHB: Northwest Portland Area Indian Health Board

XSL•FO
RenderX

Edited by A Mavragani; submitted 23.10.24; peer-reviewed by TF Ojo; revised version received 08.01.25; accepted 09.01.25; published 16.04.25.

Please cite as:
Sidhu A, Shegog R, Craig-Rushing S, Trevino N, Singer M, Jessen C, Gorman G, Simpson S, Peskin M, Hernandez B, Markham C Using the Healthy Native Youth Implementation Toolbox to Provide Web-Based Adolescent Health Promotion Decision Support to American Indian and Alaska Native Communities: Implementation Study
JMIR Form Res 2025;9:e67885
URL: https://formative.jmir.org/2025/1/e67885
doi:10.2196/67885

© Amrita Sidhu, Ross Shegog, Stephanie Craig-Rushing, Nicole Trevino, Michelle Singer, Cornelia Jessen, Gwenda Gorman, Sean Simpson, Melissa Peskin, Belinda Hernandez, Christine Markham. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluation of a Curriculum-Based Nutrition Education Intervention Protocol in Elementary Schools: Nonrandomized Feasibility Study

Jacqueline Marie Brown[1], MHSc; Nicholas Rita[2], MEd; Beatriz Franco-Arellano[1], PhD; Ann LeSage[2], PhD; Joanne Arcand[1], PhD, RD

[1]Faculty of Health Sciences, Ontario Tech University, 2000 Simcoe Street North, Oshawa, ON, Canada

[2]Mitch and Leslie Frazer Faculty of Education, Ontario Tech University, Oshawa, ON, Canada

Corresponding Author:
Joanne Arcand, PhD, RD
Faculty of Health Sciences, Ontario Tech University, 2000 Simcoe Street North, Oshawa, ON, Canada

Abstract

Background: Improving children's food literacy through school-based interventions can support developing healthy eating habits. However, teachers lack appropriate resources, time, and training to provide nutrition education in schools. Serious games, which are games designed for a purpose other than entertainment, have been demonstrated to improve children's food literacy and dietary intake and can address the barriers teachers face in providing nutrition education. Foodbot Factory (Arcand Lab) is a nutrition education intervention that is aligned with curricula and uses a serious game to provide nutrition education to students. Further evidence is needed to understand how serious games, including Foodbot Factory, can be researched in schools to support nutrition education.

Objective: The objective of this study was to evaluate the feasibility of a research study protocol that implements the curriculum-based nutrition education intervention Foodbot Factory into a real-world classroom setting. The evaluation of the protocol included study processes, resources, and management feasibility outcomes, as well as a preliminary assessment of scientific outcomes relevant to the intervention.

Methods: A nonrandomized study determined the feasibility of intervention implementation. Grade 4 and 4/5 classrooms were assigned to have nutrition education lessons for 5 days with either the Foodbot Factory or a control intervention. Outcomes were assessed in 4 feasibility domains of study processes (eg, recruitment and attrition rates), resources (eg, time taken to deliver the intervention), and management (eg, challenges with intervention delivery), and a preliminary assessment of scientific outcomes pertaining to the acceptability and impacts of the interventions. These outcomes were captured in semistructured field notes completed by study staff and a Nutrition Attitudes and Knowledge questionnaire and acceptability questionnaire completed by participants. Data were analyzed descriptively and using a paired *t* test to assess within-group changes in nutrition knowledge.

Results: In total, 4 classrooms participated in the feasibility study, with varying recruitment rates for schools (3/20, 15%), classrooms (4/4, 100%), parents (54/102, 53%), and children (49/54, 91%). The time required to implement the research protocol, including data collection and lesson plans, was sufficient and management of the intervention implementation was overall successful. Some challenges were experienced with classroom management during data collection, specifically with electronic data collection. After the intervention, participants reported a positive affective experience (26/41, 63%) and learning something new about healthy eating (31/41, 76%). Participants in both study groups improved their nutrition knowledge, but the changes were not statistically significant. The Foodbot Factory group had a statistically significant improvement in their knowledge of vegetables and fruit (P=.04) and protein foods (P=.03).

Conclusions: These findings indicate that the study protocol is feasible to implement and evaluate Foodbot Factory in a representative sample with select modifications to improve recruitment and data collection procedures.

(JMIR Form Res 2025;9:e69242) doi:10.2196/69242

KEYWORDS

nutrition education; serious games; children; food literacy; school nutrition intervention; feasibility

Introduction

One way to support children in acquiring and maintaining healthy eating patterns is by developing their food literacy. Food literacy describes the set of interrelated food and nutrition knowledge (ie, understanding food groups and the nutrients in foods), food skills (ie, ability to prepare foods and read recipes), and attributes (eg, self-efficacy and confidence) that interact with our broader socioecological environment to shape our dietary behaviors [1]. Research has demonstrated that children

XSL•FO
RenderX

and adolescents with higher levels of nutrition knowledge, a core component of food literacy, are more likely to have a higher quality dietary pattern [2,3]. Improving food literacy is especially relevant in Canada where the average child exceeds recommended intakes for saturated fat, sodium and free sugars, and consumes 21% - 25% of their daily caloric intake from foods that are not recommended by dietary guidelines [4,5]. This dietary pattern can increase the future risk of noncommunicable diseases such as cardiovascular disease and type 2 diabetes [6].

School curriculum, policies, and programming are an important way to support the development of child food literacy and healthy eating behaviors. All jurisdictions in Canada have food and nutrition as a core component of elementary health curricula [7], however, teachers face several barriers to implementing nutrition education in their classrooms due to a lack of training and dedicated time and resources for nutrition [8-10]. Published literature shows that well-designed curriculum-based interventions can effectively improve nutrition knowledge and behaviors [11]. A meta-analysis of school- and curriculum-based nutrition education interventions found that experiential learning approaches, such as school gardens and cooking classes, had the greatest impact on child nutrition knowledge and dietary intake [12]. While these experiential teaching approaches enhance the student learning experience, they are resource- and time-intensive and fail to address the needs of teachers. Thus, alternative experiential learning approaches are needed that can effectively support children's food literacy development and address the teacher-reported barriers to providing nutrition education.

Technology-based nutrition interventions can address some of these teacher-reported barriers, due to high access and ease of use in the classroom [13]. Serious games, which are games designed for a primary reason other than entertainment, have emerged as a leading technology-based educational platform as they use an experiential learning approach [14]. Research on nutrition-focused serious games has found that they can improve vegetable and fruit intake and nutrition knowledge among children [15-17]. Unfortunately, these nutrition-focused games are often not available to the public [18], do not consistently align with relevant nutrition curriculum, and are underinvestigated in classrooms as a resource to support curriculum implementation. Research and resources are required to support teachers in providing nutrition education and create opportunities for children to develop their nutrition knowledge and food literacy in the classroom.

The Foodbot Factory intervention was developed by an interdisciplinary team of dietitians, game developers, teachers, and researchers, to support elementary teachers with curriculum-based nutrition education and to improve children's nutrition knowledge [19]. The intervention includes a serious game for students played on a mobile app, with lesson plans for teachers. The content is designed for children ages 8 - 12 years (Grades 4 and 5) and aligns with the 2019 Ontario Health and Physical Education curriculum and Canada's Food Guide [20,21]. Our previous proof-of-concept research demonstrated that the Foodbot Factory serious game significantly improved children's nutrition knowledge compared with a control food-themed game [22]. However, this study was conducted in a controlled setting, not in the intended classroom environment. The lesson plans developed have also not yet been tested in classrooms, which is critical to understand their implementation [23]. Before embarking on a larger randomized trial to evaluate the impact of Foodbot Factory in classrooms, a setting that comes with unique implementation challenges, a feasibility assessment of an intended research protocol is warranted. Therefore, the objective of this research was to determine the feasibility of a research protocol to evaluate the Foodbot Factory intervention and a control intervention with children in the classroom setting. Study processes, resources, management, and scientific outcomes were the feasibility elements assessed. Such data will inform future research protocols to evaluate the efficacy of Foodbot Factory as part of a randomized trial.

Methods

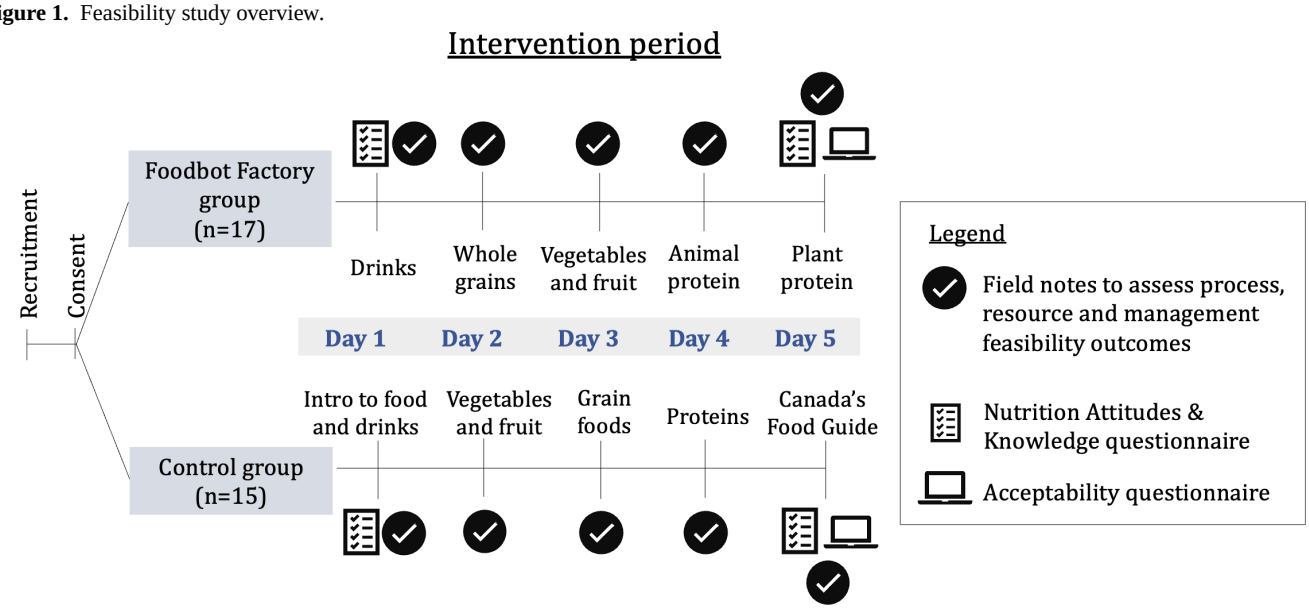
Study Design

This was a nonrandomized study to assess the feasibility of implementing a trial protocol for a nutrition education intervention (Foodbot Factory) and a control intervention among Grade 4, 4/5, and 5 elementary school classrooms, over a 5-day period. Classrooms were assigned in a 1:1 ratio to one of two groups: (1) the Foodbot Factory group receiving nutrition education using the Foodbot Factory serious game and lesson plans or (2) the control group receiving nutrition education using nontechnology based learning activities. This study protocol was cocreated with several school board partners. This approach increased the feasibility of the study protocol from the school board's perspective, particularly in relation to the length of time required for the study and the use of a study teacher to provide the intervention (described in the Interventions: Overview section).

The primary objective was to assess the feasibility, or ability to be successful in implementing the study protocol as planned using feasibility outcomes that are evaluated across 4 domains, that are study processes, resources, management, and a preliminary assessment of scientific outcomes [24]. The specific outcomes within each domain were based on established guidelines for conducting feasibility research [24]. Study process, resource, and management outcomes were documented throughout the study period. The preliminary assessment of scientific outcomes was assessed by having participants complete questionnaires on their nutrition knowledge (day 1 and day 5) and intervention acceptability (day 5; Figure 1).

XSL•FO
RenderX

Figure 1. Feasibility study overview.



Ethical Considerations

This study was approved by the Ontario Tech University Research Ethics Board (#16930) and by the participating school board. Informed consent (Multimedia Appendix 1) was obtained from participants' parents or guardians, along with informed assent from child participants. All study data are in a deidentified format. Classroom teachers received a $50 CAD (US $37.50) gift card as an acknowledgment for having their classroom participate in the research. All children in the classroom (regardless of participation in the study) received an Ontario Tech University water bottle as a token of appreciation.

Participants

Grade 4, 4/5, and 5 classrooms in a Greater Toronto Area school board were eligible to participate if they had not yet covered the healthy eating component of the curriculum. Furthermore, 4 classrooms from 2 elementary schools participated in the study. The research team randomly selected schools and contacted them via email to gauge interest in participation. Principals who expressed interest subsequently invited classroom teachers to participate in the study. In each school we recruited 2 classrooms, assigning 1 to the Foodbot Factory group and the other to the control group. Classrooms, rather than individuals, were assigned to the group as the intervention is designed for classrooms and it is not logistically feasible to assign individuals within the same classroom to different study groups. In the first school, 1 classroom was randomly assigned to their group and the second classroom was assigned to the opposite group. In the second school, classrooms were assigned to the study groups based on enrollment to ensure a comparable number of individual participants in each group. Before the study, classroom teachers emailed parents the web-based consent form. Child assent was obtained on day 1 of the study (Monday) from those with parental consent. For participants who were absent on day 1, assent was collected on day 5 for the purposes of collecting data on intervention acceptability only. All children in a classroom participated in the intervention, but only those with parental consent participated in data collection and analysis.

Interventions

Overview

Both interventions were 5 consecutive days in duration and classrooms received nutrition education for 35 - 40 minutes per day [23]. The interventions were provided by a hired certified teacher, who was part of the research team ("study teacher"). Due to the nature of their role, the study teacher was not blinded to the intervention group. The study teacher was trained to follow standardized operating procedures for the delivery of the interventions to reduce sources of bias and ensure a high level of fidelity. The study teacher was accompanied by an observer who was trained in the same procedures. The research team originally estimated that data collection would take a maximum of 20 minutes, and that lesson plan delivery would take 40 minutes. Classroom teachers were not involved in teaching the lessons, but they were present in the classroom during the study and supported classroom management.

Foodbot Factory Intervention

Classrooms in the Foodbot Factory group received nutrition education using the Foodbot Factory serious game and corresponding curriculum-based lesson plans [19,25]. The serious game is based on experiential learning theory [26], while the lesson plans are rooted in constructivist learning theory, to help children connect knowledge learned in the game to their lived experiences [27]. The Foodbot Factory intervention consists of 5 nutrition education lessons, with additional instructions for teachers on the importance of using food-neutral language, suggestions to incorporate cultural foods into each lesson and modifications for different learning needs. Each lesson follows the 3-phase lesson structure, which is an effective format for structuring a lesson [28]. The first phase of the lesson, "Minds On", introduces children to the lesson topic and establishes expectations (5 - 10 min). In the second phase, "Exploration", children play through 1 module of the Foodbot Factory serious game on a tablet provided by the research team (10 - 15 min). For the final phase, "Consolidation", discussion and teacher-led activities allow children to connect their new

XSL•FO
RenderX

knowledge from the serious game to their previous knowledge and lived experiences (10 min).

Control Intervention

Classrooms in the control group received nutrition education that covers the same topics as the Foodbot Factory intervention but used nontechnology based learning materials (eg, activity sheets). The control intervention also consists of 5 nutrition education lessons, following the same 3-phase lesson structure (Table 1). However, in place of the Foodbot Factory serious game, the lesson plans incorporated pre-existing resources that were sourced from a popular website repository of educational materials [29]. These resources were carefully reviewed and selected by the research team to closely match the Foodbot Factory intervention learning goals and ensure their alignment with curriculum and quality. Unlike clinical practice, where there are often existing guidelines for usual care that may serve as a control group in research, there is great variation in the strategies teachers implement for nutrition education. However, a consistent control intervention is required for comparative research purposes. The approach for the control intervention in this study was informed by qualitative interviews and focus groups conducted with Canadian elementary school teachers who frequently reported using pre-existing resources found online for nutrition [10].

XSL•FO
RenderX

Table . Foodbot Factory and control intervention learning topics and activities.

Daily learning topic and study group	Phase 1: "Minds On" introductory activities and set expectations (approximately 10 min)	Phase 2: "Action" main lesson activity (approximately 10-15 min)	Phase 3: "Consolidation" summarize and review lesson (approximately 10 min)
Day 1—drinks (Foodbot Factory) and introduction to food and drinks (control)			
Foodbot Factory	• Introductory slide show on food and drinks • Teacher-led class activity • Class discussion	• Play "Drinks" module in Foodbot Factory serious game	• Complete phase 1 class activity, adding to it based on what was learned • Class discussion
Control	• Introductory slide show • Complete activity sheet on food groups and drinks independently • Class discussion	• Complete activity sheet on different food groups and drinks independently	• Think-pair-share activity sheet • Take up answers from phase 2 activity • Class discussion
Day 2/3[a]—whole grain foods (Foodbot Factory day 2 and control group day 3)			
Foodbot Factory	• Introductory slide show • Class discussion	• Play "Whole Grain Foods" module in Foodbot Factory serious game	• Small-group activity and take up answers • Class discussion
Control	• Introductory slide show • Class discussion	• Complete activity sheets on whole grain foods independently or in pairs	• Take up answers from phase 2 activity • Class discussion
Day 2/3—vegetables and fruit (Foodbot Factory day 3 and control group day 2)			
Foodbot Factory	• Introductory slide show • Complete activity sheet on vegetables and fruit independently • Class discussion	• Play "Vegetables & Fruit" module in Foodbot Factory serious game	• Complete phase 1 activity sheet, adding to it based on what was learned • Class discussion
Control	• Introductory slide show • Class discussion	• Complete activity sheets on vegetables and fruit in pairs or small groups	• Take up answers from phase 2 activity • Complete activity sheet in pairs or small groups
Day 4—animal protein (Foodbot Factory) and protein foods (control)			
Foodbot Factory	• Introductory slide show • Class discussion	• Play "Animal Protein" module in Foodbot Factory serious game	• Teacher-led class activity • Class discussion
Control	• Introductory slide show • Class discussion	• Complete activity sheets on animal and plant protein foods independently or in pairs	• Take up answers from phase 2 activity • Small group activity sheet • Class discussion
Day 5—plant protein (Foodbot Factory) and Canada's Food Guide (control)			
Foodbot Factory	• Introductory slide show • Class discussion	• Play "Plant Protein" module in Foodbot Factory serious game	• Teacher-led class activity • Class discussion
Control	• Introductory slide show • Class discussion	• Complete activity sheet on Canada's Food Guide independently	• Take up answers from phase 2 activity • Class discussion

[a]Study groups cover both the topics of "Whole Grain Foods" and "Vegetables and Fruit" on different days.

XSL•FO
RenderX

Outcomes

We holistically determined if each study outcome was feasible for a research protocol by considering the context of each individual outcome, and if the outcome would reasonably facilitate a fully powered study.

Study Processes

Feasibility of study processes refers to the main elements of a study that are necessary for its success, including recruitment and retention of participants [24]. We assessed 7 variables related to study processes including school recruitment, classroom recruitment, parent recruitment, child recruitment, attrition rate, data collection, and instrument collection completion rates. Recruitment rates were calculated as the number of recruited school or participants over the total number of contacted or eligible participants (eg, school recruitment rate=schools successfully recruited/schools contacted). The attrition rate was calculated as the number of children recruited to the study with data available from 2 data collection questionnaires, a baseline Nutrition Attitudes and Knowledge (NAK) questionnaire and postintervention NAK, over the total number of child participants. Furthermore, 2 data collection completeness rates were calculated. The first was calculated as the number of child participants with 2 fully complete NAK questionnaires over the total number of child participants who completed the study. The second data collection completion rate was determined by dividing the number of structured field notes completed by study staff by the total number of expected field notes. The instrument collection completeness rate was calculated as the number of child participants with all study questionnaires completed (2 NAKs and the acceptability questionnaire) over the total number of child participants.

Study Resources

Feasibility of study resources refers to the time and resources required to run the study [24]. Furthermore, 3 variables related to study resources were assessed including time taken to deliver the intervention, time taken to collect data, and impacts on study personnel who were the study teacher and outcome assessor (ie, what was the qualitative experience for those conducting the study). Variables related to time were taken with a stopwatch and impacts on study personnel were assessed qualitatively. This study did not assess outcomes related to material resources as all necessary resources were brought into classroom by the study team (eg, tablets to play the Foodbot Factory game, activity sheets, and writing supplies).

Study Management

Feasibility of study management assesses possible issues with study implementation [24]. In this feasibility study, we collected data on 2 variables related to study management, which included challenges with data collection and challenges with intervention delivery. These outcomes were assessed qualitatively.

Preliminary Assessment of Scientific Outcomes

A preliminary assessment of the scientific outcomes of the study included an evaluation of the impacts of the intervention on nutrition knowledge and its safety. These data are intended to inform a fully powered future study, not to determine the effectiveness of an intervention [24]. In total, 3 variables were assessed in this category including acceptability of the intervention, adverse events, and impacts of the intervention on nutrition knowledge. An acceptability questionnaire assessed the perceived acceptability of the intervention by participants. The acceptability questionnaire was based on the theoretical framework of acceptability [30]. The questionnaire was adapted from the generic theoretical framework of acceptability questionnaire to be specific to the nutrition education interventions and ensure readability for children [31]. It consisted of 11 items on a numerical 5-point Likert scale questions defined by level of agreement (1=strongly disagree to 5=strongly agree). Adverse events were documented in study field notes and a self-report question on the acceptability questionnaire asking participants about feelings of stress due to the intervention. Nutrition knowledge was measured using the validated NAK questionnaire, which was developed and validated by our research team and is sensitive to detecting changes in nutrition knowledge [22,32]. The NAK questionnaire consists of 20 questions to assess a child's overall nutrition knowledge, and 4 subscores of nutrition knowledge (5 questions per subscore) to assess knowledge of a specific food group (eg, drinks, whole grain foods, vegetables and fruit, and protein foods). NAK questions are guided by the recommendations in Canada's Food Guide [20], which also are part of the Foodbot Factory intervention.

Outcome Assessment

Feasibility outcomes were collected throughout the study from (1) structured daily field notes that were created and completed by the research team to document the aforementioned study process, resource, and management outcomes; (2) assessments of nutrition knowledge using the NAK questionnaire [32]; and (3) an acceptability questionnaire. The 2 questionnaires were used to provide a preliminary assessment of the scientific impacts of the intervention.

Field notes were completed each day of the study by the study teacher and observer (days 1 through 5). The NAK questionnaire was administered to participants by an outcome assessor on day 1, before the first nutrition education lesson, and again on day 5, after the final lesson. The acceptability questionnaire was also administered by the outcome assessor and completed on day 5, after participants completed the NAK questionnaire. The questionnaires were completed independently by participants at their desks in the classroom with the outcome assessor circulating to monitor progress. The outcome assessor was present in each classroom throughout the intervention period as an observer, thus they were not blinded to the intervention group. In the first and second classrooms, students completed electronic forms using Qualtrics (Silver Lake); however, there were several issues with electronic data collection (described in Results) and the remaining 2 classrooms completed paper-based forms.

Sample Size

Our research team previously conducted a pilot study of the Foodbot Factory serious game, which provided us with an understanding of the distribution and effects sizes for the outcome of nutrition knowledge [22]. Thus, our objective with

XSL•FO
RenderX

this feasibility study was pragmatic in that we needed to understand the logistics of recruitment, data collection, and intervention delivery in the classroom setting [33]. For this type of feasibility study, 12 to 30 participants per group is suggested by other researchers [34,35]. Therefore, we aimed to recruit 4 classrooms as the feasibility challenges in this study were likely to arise related to both the classroom environment and individual participants. This resulted in 102 children eligible to participate and allowed us to assess the feasibility of implementing the interventions in different neighborhoods and understand the pragmatic elements required to scale the intervention up in a fully powered study.

Data Analysis

Outcomes related to study processes were analyzed using descriptive statistics, frequencies, and percentages. Resource outcomes were assessed using descriptive statistics and impacts on study personnel were assessed narratively. Outcomes related to management were assessed narratively. The scientific outcome of acceptability was analyzed using descriptive statistics. On all Likert scale questions, a response of 1 or 2 was considered as disagreement, a response of 3 was considered neutral, and a response of 4 or 5 was considered as agreement. Intervention safety was analyzed using both descriptive statistics and narratively. The scientific outcome of intervention impacts was assessed using descriptive statistics but was not considered in the final determination of protocol feasibility as the sample size would not have sufficient power. Within-group analysis for changes in overall nutrition knowledge and nutrition knowledge subscores was completed using a paired *t* test, after confirming the data were normally distributed through the Shapiro-Wilk test. Participants were excluded from the analysis of the overall knowledge score if there were any missing data. However, participants were included in nutrition knowledge subscore analyses if they had complete data for a particular subscore. All statistical analyses were completed using R Statistical Software version 4.2.0 (R Core Team) [36].

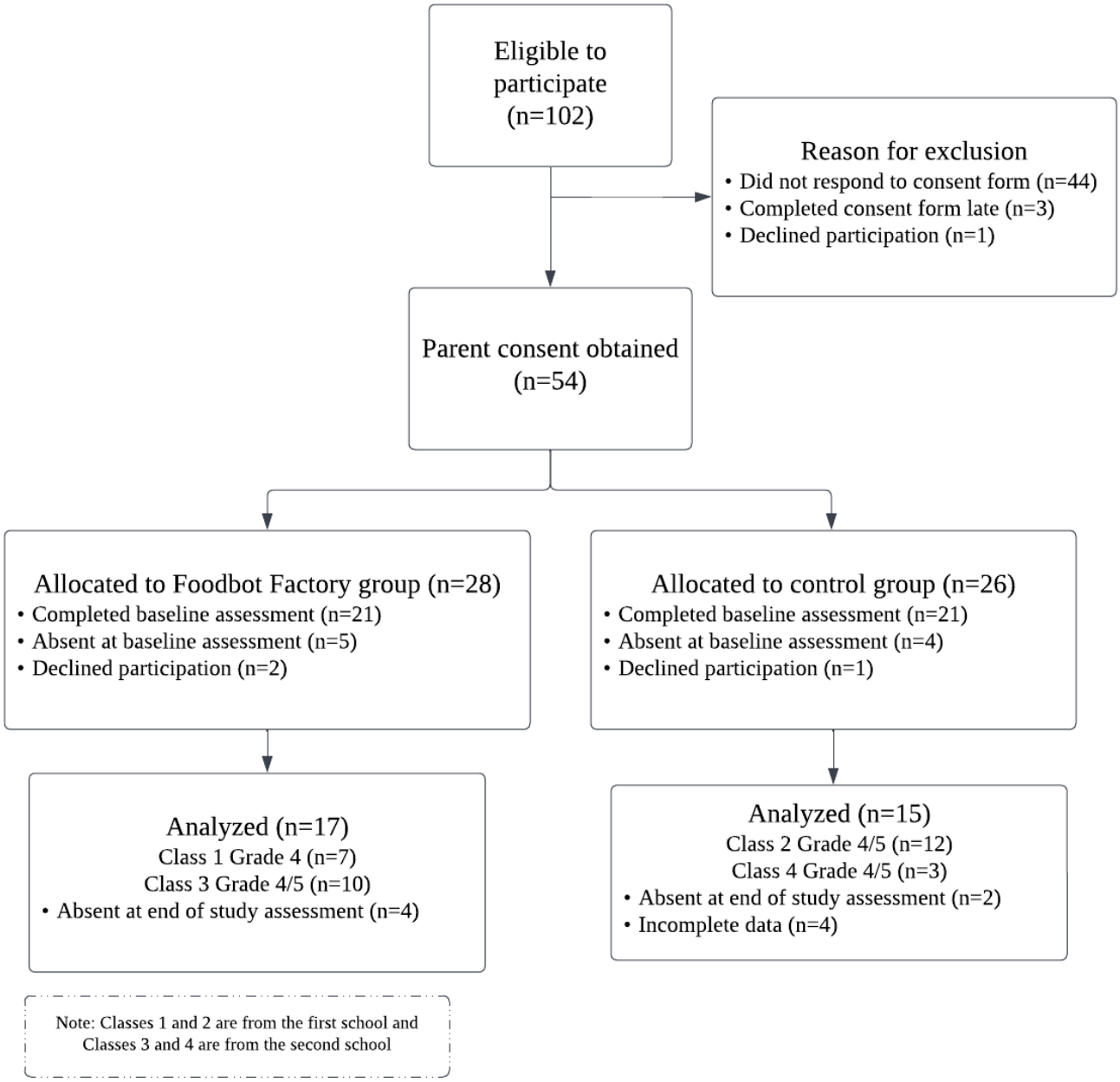
Results

Overview

In total, 4 classrooms participated from 2 schools between February and May 2023 (Figure 2). Both schools were in urban areas in culturally and linguistically diverse neighborhoods. The first school was in a higher socioeconomic-status neighborhood while the second school was in a neighborhood with a higher proportion of newcomers to Canada and government-subsidized housing. Participating classrooms included one Grade 4 and three Grade 4/5 split classes.

XSL•FO
RenderX

Figure 2. Flowchart of participant progress through the study.



Study Process Outcomes

The school recruitment rate was low (3/20, 15%). Of the 20 principals who were contacted by email, 70% (n=13) did not respond, 10% (n=2) declined participation, and 5% (n=1) replied with a question about the study but lost to follow-up. To counter the low email response rate, the research used phone calls for recruitment and was able to recruit 1 school after 3 phone calls. Once a principal agreed to have their school participate, 100% of classroom teachers in those schools agreed to have their classroom participate. In total, 4 classrooms from 2 of the 3 recruited schools were scheduled for participation, with a total of 102 children. While a third school was recruited to participate, they ultimately were not scheduled to be in the study, due to a staffing shortage among the research team. Parent recruitment rate was low (54/102, 53%); 43% (44/102) did not respond to the consent form, 3% (3/102) consented after the study start date, and 1% (1/102) declined participation. To improve parent recruitment, the research team requested that classroom teachers send a reminder email. Reminders were sent by teachers in 2 classrooms, resulting in a 77% parent recruitment rate compared with 40% in classrooms with no reminders. Anecdotally, 1 classroom with lower parent recruitment had a high number of parents who did not speak English as a first language, and the classroom teacher noted this was likely a barrier to participating in the study. Most children with parental consent (49/54, 91%) agreed to participate in the study, with 24 in the intervention group and 25 in the control group. The average age of participants was 9.6 (SD 0.68) years. Participants self-reported their gender as a boy (24/49, 49%), girl (18/49, 37%), or a self-defined identity (2/49, 4%), with the remaining participants preferring not to answer (5/49, 8%). In the intervention group, 17 participants completed the study; however, only 10 had a complete set of data. In the control group, 15 participants completed the study, with 9 having a complete set of data. The attrition rate was 19%. For participants, the instrument collection completion rate was moderate (73%) and the data collection completion rate for the NAK questionnaire was low (59%). The

XSL•FO
RenderX

data collection completion rate of the structured field notes by study staff was 100%.

Study Resource Outcomes

The average time required for child assent was 6.5 minutes, lesson plan delivery was 34 minutes (8 min for minds on, 14 min for exploration, and 12 min for consolidation), acceptability questionnaire completion was 5.5 minutes, and NAK questionnaire completion was 13 minutes; all consistent with original estimates. Study personnel, including the study teacher and outcome assessor, reported an overall positive experience as they enjoyed interacting with the students. In one classroom, classroom management was a significant challenge due to a handful of students who frequently interrupted the lessons and outcome assessments. This was a minor psychological stressor for the study teacher. There were challenges with retaining the hired study teachers on the project due to career changes and an unanticipated emergency, requiring the research team to reschedule 2 classrooms to a later date.

Study Management Outcomes

Data collection procedures and delivery of the intervention were executed as planned, although some challenges were experienced. Tablets were originally used for participants to complete data collection forms, but this posed a few unanticipated challenges. For example, participants intentionally or unintentionally "refreshed" or navigated away from the web-based NAK questionnaire, leading to the loss of completed responses. In addition, once a NAK questionnaire was submitted, study personnel were unable to verify if all questions had been completed. This resulted in incomplete data for some participants. Several participants struggled to focus on completing the data collection forms due to distractions in the classroom. The most significant distraction occurred when other children who were not participating in the study were allowed by the classroom teacher to engage in other activities, namely the use of electronic devices. Furthermore, as participants completed the questionnaire independently, those who finished early then moved on to a different activity. Participants who needed more time to complete the questionnaire became increasingly distracted as others around them finished, which may have contributed to them rushing to finish and skipping questions.

Based on the implementation of the interventions in the first 2 classrooms, some changes were made to the intervention delivery that were implemented in the subsequent 2 classrooms. The changes clarified instructions, modified select discussion questions to improve wording and added slideshows with the daily learning objectives instead of writing them on the board. Overall, these changes were relatively minor and were intended to improve the clarity and accessibility of the lesson delivery. In the study, classroom management was particularly challenging in 1 classroom due to some children repeatedly interrupting the study teacher. Study personnel found that having the classroom teacher and educational assistants support classroom management was helpful. In 2 classrooms, the intervention schedule was modified to occur over 4 days instead of 5, due to a conflicting school-wide track and field event. On one of the 4 days, 2 intervention lessons were provided consecutively instead of having 1 lesson each day. While it was logistically feasible to provide 2 lessons back-to-back in 1 day, student engagement was lower during the second lesson. Another challenge occurred in the Foodbot Factory intervention group, where some children struggled with the augmented reality components in the Foodbot Factory serious game. The research team responded to this by improving instructions about how the augmented reality features work. In addition, some children struggled to transition away from using the tablet to play the Foodbot Factory serious game to the next learning activity as they found the game engaging and wanted to keep playing.

Preliminary Assessment of Scientific Outcomes

Based on observations, participants enjoyed participating in both the Foodbot Factory and control intervention lessons and were engaged in learning about nutrition through the activities. Overall, intervention acceptability among participants was moderate to high, with most participants reporting that they had a positive experience (26/41, 63%), understood the goals of the nutrition education lessons (37/41, 90%), and that they learned something new (31/41, 76%; Figure 3). No participants reported feeling sad or stressed about food and nutrition from the lessons on the acceptability questionnaire nor did study personnel observe any verbal or nonverbal signs of concern from children in the classroom.

Scores on the NAK questionnaire were normally distributed as determined by the Shapiro-Wilk test. Both groups demonstrated improvements in their overall nutrition knowledge (Foodbot Factory group: mean 10.8, SD 2.1 to mean 12.5, SD 3.5; control group: mean 11.8, SD 1.9 to mean 12.7, SD 1.8), although these improvements were not statistically significant. For subscores of nutrition knowledge, participants in the intervention group demonstrated statistically significant improvements in knowledge of vegetables and fruit and protein foods. Participants in the control group demonstrated statistically significant improvements in knowledge of grain foods (Table 2).

XSL•FO
RenderX

Figure 3. Acceptability of the nutrition education interventions as rated by participants (n=41).

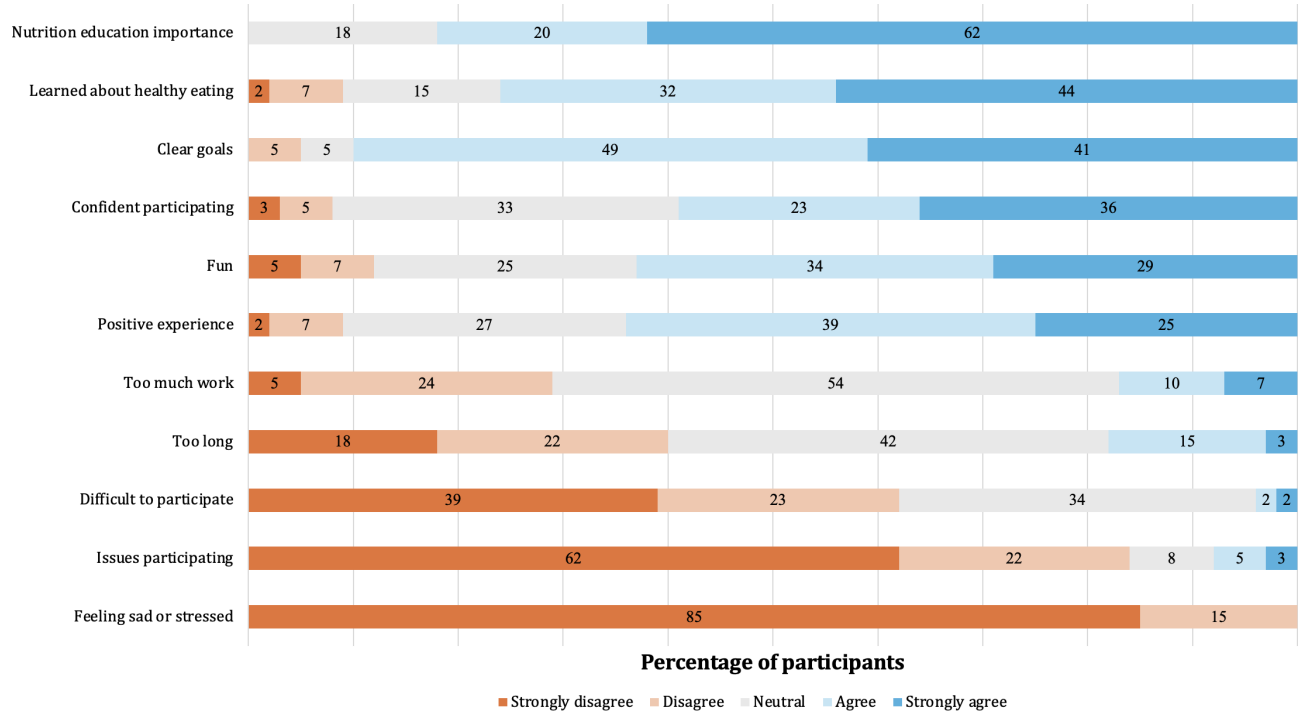


Table . Changes in overall and subscores of nutrition knowledge within the Foodbot Factory and control groups[a].

Nutrition knowledge	Foodbot Factory group				Control group			
	Score, n	Baseline, mean (SD)	End of study, mean (SD)	*P* value	Score, n	Baseline, mean (SD)	End of study, mean (SD)	*P* value
Overall nutrition knowledge	10	10.8 (2.1)	12.5 (3.5)	.13	9	11.8 (1.9)	12.7 (1.8)	.09
Drinks	17	3.94 (0.66)	3.76 (0.56)	.33	15	3.93 (0.70)	3.93 (0.59)	≥.99
Grain foods	16	2.31 (1.2)	2.75 (1.1)	.17	15	2.87 (1.1)	3.40 (1.1)	.03
Vegetables and fruit	15	2.27 (0.88)	3.00 (0.93)	.04	14	2.86 (0.86)	3.07 (1.0)	.61
Protein foods	12	1.92 (0.90)	2.75 (1.6)	.03	10	2.60 (1.1)	1.70 (0.95)	.05

[a]Data presented as means and SDs with paired *t* tests used to assess changes in knowledge from baseline to the end-of-study within groups. Sample sizes vary as only participants with a complete set of responses were included in the analysis.

Discussion

Principal Findings

Our research study determined the feasibility of implementing the Foodbot Factory nutrition education intervention as part of a research protocol in Grade 4 and 4/5 classrooms, identifying 9 facilitators to support a larger study and 4 risks that will require mitigation. Elements of the study that were deemed feasible included classroom recruitment, child recruitment, the attrition rate, time taken to deliver the intervention and collect data, the impacts of the study on personnel, management of delivering the intervention, lack of adverse events, and acceptability of the intervention. However, strategies will be needed to improve school recruitment, parent recruitment, data collection completeness, and the management of data collection procedures. In this feasibility study, we made minor modifications to the protocol on implementation to improve recruitment and intervention implementation (eg, conducting recruitment phone calls and modifying intervention instructions for the classroom). With further modifications to mitigate the aforementioned risks, our findings largely support the feasibility of this research protocol to evaluate the Foodbot Factory nutrition education intervention as part of a fully powered randomized controlled trial.

The primary strength of this study was our ability to better understand how to collaborate with schools, teachers, parents, and children in our future research. Key takeaways learned from this study that will be valuable to other researchers, include the importance of consulting with school boards before conducting research, ensuring recruitment materials and methods are relevant for the target audience, engaging classroom teachers and testing data collection methods before their implementation. We discussed our methodology and objectives with staff from several school boards before seeking school board approval to

XSL•FO
RenderX

conduct the research. From the outset, this approach resulted in a protocol that would be more feasible and acceptable to implement in classrooms and ensure value to our participants and the school board. For example, the initial version of our study protocol requested 5 hours of total class time, which was perceived as too much time taken away from other curriculum. Not only did we modify the protocol to reduce the total time by 1 hour, we also updated our recruitment materials to highlight alignment of the study with curriculum. This was a critical change as teachers have very limited time to cover content that is outside of the curriculum. Ultimately, the resulting protocol is not only acceptable to school boards, but the duration of the intervention now better aligns with the time teachers typically allocate to a lesson and clearly communicates the value of the study to possible participants. These features increase real-world acceptability of the intervention. Researchers should consult with their local school boards and teachers to establish a collaborative relationship from the beginning to develop interventions and research protocols that are suitable and valuable.

When recruiting individual schools, classroom teachers, and parents, recruitment materials should be succinct and highlight the value of study participation. In this study, phone calls were effective at recruiting schools as they enable direct consultation on the practical aspects of the study and allow the school to quickly evaluate if the study will work for them. Low recruitment of schools was still seen in this study and has been reported as a challenge in other school-based health studies due to low interest, poor timing, and other commitments [37,38]. During this study, engagement of the classroom teacher was critical for parent recruitment, as we were not permitted to contact parents directly. We found low parent recruitment was driven by parents not returning consent forms, rather than them explicitly declining participation, which has been observed in other studies [37]. We also observed significant improvements in recruitment when the classroom teacher sent reminders to parents about the study, highlighting the important role of classroom teachers in engaging parents in study participation. In this study, classroom teachers were asked to contact all parents and guardians to share the consent form and reminders to complete it. However, classroom teachers may introduce selection bias if they selectively reach out to parents and guardians. Instructions for sending out parent consent forms to classroom teachers should be clear and concise to avoid this bias.

In this study, we show the importance of testing data collection methods before implementation in a classroom setting. First, an important takeaway from this study is that electronic data collection presented several challenges. Although other studies have used electronic data collection methods, it may present more challenges if the technology has not been used before in that context [39]. In this study, training was not provided to participants on how to complete the electronic data collection form before data collection as we did not have sufficient time allotted for such training. Training on data collection methods could minimize issues like participants accidentally refreshing the page and skipping questions [39]. In this study, questions on the electronic data collections forms were not made mandatory for submission as study participation was voluntary. Second, participants completed data collection independently, requiring them to use their literacy skills to read and respond to each question. This data collection approach reduced accessibility and led to participants finishing at varying time points and the classroom environment becoming increasingly distracting for participants who needed more time. Missing data from both missed responses and missing surveys in other school-based studies is a common phenomenon ranging from 11% - 51% at a given outcome assessment time point [40,41]. In a future protocol, data collection could be modeled on the procedures classrooms use when administering standardized tests. This would align research data collection with procedures that are already familiar to children, provide clear instructions, increase accessibility, allow for verification of data collection completeness to encourage complete responses and directly guide the classroom as a group through the questions so all participants finish at the same time.

This study also examined the impacts of the developed nutrition education interventions on children's nutrition knowledge. Statistically insignificant improvements in overall nutrition knowledge were observed in both groups, and participants in the Foodbot Factory group significantly improved their knowledge of vegetables and fruit and protein foods. The results within-groups at the individual level indicate improvements in nutrition knowledge but do not have the same effect size as a larger proof-of-concept study of the Foodbot Factory serious game, where significant improvements in children's overall nutrition knowledge were observed both within and between groups [22]. This indicates that the findings in this feasibility study for nutrition knowledge are inconclusive, due to the insufficiently powered sample size, and should be used primarily for descriptive purposes and informing improvements to the intervention [42]. Some changes were made to the lesson plans during the study to improve clarity and accessibility (eg, using a slideshow to share learning objectives instead of writing them on the board) as the process of conducting feasibility studies and developing research protocols is recognized to be iterative and adaptive [43]. Since these changes were relatively minor and were focused on delivery, rather than intervention content, we do not anticipate they influenced the outcomes of interest.

This study had several strengths and limitations that can inform future research. This study included a small sample size, which is primarily driven by low parent recruitment. Recruitment may be a primary feasibility challenge moving forward to a fully powered study. Including more classrooms may have revealed additional feasibility issues; however, the classrooms were intentionally sampled from schools in neighborhoods with known differences in their sociodemographic profiles. While the challenges in this study are similar to those observed in international research [37,38,40,41], the results may be most applicable for those conducting school-based research in the Greater Toronto Area, or other urban environments in Ontario, Canada. In this study, we chose not to randomize classrooms as we were unsure of how successful our recruitment would be and this approach would allow us to ensure balance between study groups. Nonrandomized feasibility studies are common for research in earlier phases of preparation for a trial and when

XSL•FO
RenderX

there are significant unknowns [44]. To address the lack of randomization and balance school- and neighborhood-level characteristics, the research team assigned 1 classroom at each school to a different group; however, we acknowledge the potential of selection bias and an uneven balance of characteristics between groups that may have been introduced. In future research, we will recruit 2 classrooms per school and use pair-matching randomization to both reduce bias and ensure groups are balanced on school and neighborhood level variables that can impact our outcomes of interests. Due to the study design, where classrooms rather than individuals were assigned to each treatment arm, we were unable to assess differences between groups as the data should be analyzed at the classroom cluster level, which this study was not powered to do. In future research, an adequately powered randomized controlled trial that randomizes classrooms as clusters would be more appropriate to assess the impact of the interventions in a pragmatic fashion. In this study, we provided all material resources necessary, including tablets for students to use the Foodbot Factory serious game. This can be considered as a strength, as it increased the feasibility of conducting the research, but also a limitation as we did not evaluate possible material resource limitations (and the knowledge and processes required to implement them) for conducting technology-based research in schools. Researchers should carefully consult with school boards to understand the availability of resources to ensure success or invest in the required technology to avoid material resource barriers as we have done in this study. We also chose to use a certified study teacher who was trained to implement the lesson plans as it increases feasibility from the school board. We will use this approach in a future efficacy study. Future research should aim to also evaluate the effectiveness of the research when the lesson plans are implemented by classroom teachers. At this stage, our research team did not assess retention of knowledge or examine other variables that may be impacted by a nutrition education intervention, including dietary intake and behaviors. We also did not assess acceptability of the study protocol from the perspective of classroom teachers, which was not the focus of this study. Retention of knowledge and dietary intake will be assessed in future research and qualitative interviews will also be conducted with classroom teachers to inform implementation needs and strategies for the Foodbot Factory intervention. A current limitation of the Foodbot Factory intervention is that it predominantly covers nutrition knowledge and does not address all elements of food literacy. The serious game is also only available for use on mobile devices. Our future work aims to make a web-accessible version of the Foodbot Factory serious game and expand the content to different age groups and food literacy concepts.

Conclusions

In conclusion, the majority of feasibility outcomes assessing study processes, resources, management, and intervention acceptability suggest that the study protocol was feasible [45]. The data offer several insights to inform future studies and, when the protocol is implemented into a fully powered trial, the successes identified can be leveraged, and the risks associated with school recruitment, parent recruitment, and data collection can be averted with targeted mitigation strategies. These lessons learned have been applied to our protocol for a fully powered cluster randomized controlled trial evaluating the efficacy of Foodbot Factory in classrooms [46]. The development and evaluation of the protocol in this study was an important step in understanding how to best evaluate technology-based nutrition education interventions in Canadian classrooms. The lessons learned in this study will also support the rising number of researchers embarking on school-based nutrition and health interventions in school and classroom settings both in Canada and abroad.

Acknowledgments

Funding for this research was provided by Ontario Tech University's Dean's Discretionary Fund Small Research Grants Program (JA), a Social Sciences and Humanities Research Council (SSHRC) Explore Grant (BFA), an Ontario Graduate Scholarship (JMB), and a Canadian Institutes of Health Research (CIHR) Canada Graduate Doctoral Scholarship (JMB). The authors would like to thank the principals, teachers, parents, and children for being part of this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Informed consent form.
[DOCX File, 54 KB - formative_v9i1e69242_app1.docx]

References

1. Thomas H, Azevedo Perry E, Slack J, et al. Complexities in conceptualizing and measuring food literacy. J Acad Nutr Diet 2019 Apr;119(4):563-573. [doi: 10.1016/j.jand.2018.10.015] [Medline: 30670348]
2. Thakur S, Mathur P. Nutrition knowledge and its relation with dietary behaviour in children and adolescents: a systematic review. Int J Adolesc Med Health 2022 Dec 1;34(6):381-392. [doi: 10.1515/ijamh-2020-0192] [Medline: 33594848]
3. Vaitkeviciute R, Ball LE, Harris N. The relationship between food literacy and dietary intake in adolescents: a systematic review. Public Health Nutr 2015 Mar;18(4):649-658. [doi: 10.1017/S1368980014000962] [Medline: 24844778]

XSL•FO
RenderX

4. Hack S, Jessri M, L'Abbé MR. Nutritional quality of the food choices of Canadian children. BMC Nutr 2021 May 29;7(1):16. [doi: 10.1186/s40795-021-00422-6] [Medline: 34049592]
5. Ahmed M, Praneet Ng A, L'Abbe MR. Nutrient intakes of Canadian adults: results from the Canadian Community Health Survey (CCHS)-2015 Public Use Microdata File. Am J Clin Nutr 2021 Sep 1;114(3):1131-1140. [doi: 10.1093/ajcn/nqab143] [Medline: 34020449]
6. Afshin A, Sur PJ, Fay KA, et al. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet 2019 May;393(10184):1958-1972. [doi: 10.1016/S0140-6736(19)30041-8]
7. School nutrition: support for providing healthy food and beverage choices in schools. Canadian Pediatric Society. 2020. URL: https://cps.ca/documents/position/school-nutrition-support#:~:text=Schools%20provide%20important%20opportunities%20to,healthier%20dietary%20choices%20and%20behaviours [accessed 2022-02-01]
8. Jones AM, Zidenberg-Cherr S. Exploring nutrition education resources and barriers, and nutrition knowledge in teachers in California. J Nutr Educ Behav 2015;47(2):162-169. [doi: 10.1016/j.jneb.2014.06.011] [Medline: 25257845]
9. Perera T, Frei S, Frei B, et al. Improving nutrition education in U.S. elementary schools: challenges and opportunities. JEP 2015;6(30):41-50 [FREE Full text]
10. Franco-Arellano B, Vogel E, Stein M, et al. A qualitative study exploring elementary school teachers' experiences teaching nutrition in Canada. Presented at: International Society for Behavioral Nutrition and Physical Activity Annual Meeting; Jun 14-17, 2023; Uppsala, Sweden p. 354 URL: https://annualmeeting.isbnpa.org/wp-content/uploads/2023/07/ISBNPA-2023-Abstract-Book-web-final.pdf [accessed 2025-04-02]
11. Graziose MM, Koch PA, Wang YC, Lee Gray H, Contento IR. Cost-effectiveness of a nutrition education curriculum intervention in elementary schools. J Nutr Educ Behav 2017 Sep;49(8):684-691. [doi: 10.1016/j.jneb.2016.10.006] [Medline: 27843129]
12. Cotton W, Dudley D, Peralta L, Werkhoven T. The effect of teacher-delivered nutrition education programs on elementary-aged students: an updated systematic review and meta-analysis. Prev Med Rep 2020 Dec;20:101178. [doi: 10.1016/j.pmedr.2020.101178] [Medline: 32944494]
13. Education Indicators in Canada: An International Perspective: Statistics Canada; 2020.
14. Laamarti F, Eid M, El Saddik A. An overview of serious games. International Journal of Computer Games Technology 2014;2014:1-15. [doi: 10.1155/2014/358152]
15. Baranowski T, Ryan C, Hoyos-Cespedes A, Lu AS. Nutrition education and dietary behavior change games: a scoping review. Games Health J 2019 Jun;8(3):153-176. [doi: 10.1089/g4h.2018.0070] [Medline: 30339086]
16. Thompson D, Bhatt R, Vazquez I, et al. Creating action plans in a serious video game increases and maintains child fruit-vegetable intake: a randomized controlled trial. Int J Behav Nutr Phys Act 2015 Mar 18;12(1):39. [doi: 10.1186/s12966-015-0199-z] [Medline: 25890060]
17. Hermans RCJ, van den Broek N, Nederkoorn C, Otten R, Ruiter ELM, Johnson-Glenberg MC. Feed the Alien! The effects of a nutrition instruction game on children's nutritional knowledge and food intake. Games Health J 2018 Jun;7(3):164-174. [doi: 10.1089/g4h.2017.0055] [Medline: 29634366]
18. Brown JM, Franco-Arellano B, Froome H, Siddiqi A, Mahmood A, Arcand J. The content, quality, and behavior change techniques in nutrition-themed mobile apps for children in Canada: app review and evaluation study. JMIR Mhealth Uhealth 2022 Feb 16;10(2):e31537. [doi: 10.2196/31537] [Medline: 35171100]
19. Brown JM, Savaglio R, Watson G, et al. Optimizing child nutrition education with the Foodbot Factory mobile health app: formative evaluation and analysis. JMIR Form Res 2020 Apr 17;4(4):e15534. [doi: 10.2196/15534] [Medline: 32301743]
20. Canada's food guide. Government of Canada. 2019. URL: https://food-guide.canada.ca/en [accessed 2024-10-24]
21. Ontario Ministry of Education. Health and Physical Education Grades 1-8: Government of Ontario: Toronto; 2019.
22. Froome HM, Townson C, Rhodes S, et al. The effectiveness of the Foodbot Factory mobile serious game on increasing nutrition knowledge in children. Nutrients 2020 Nov 6;12(11):3413. [doi: 10.3390/nu12113413] [Medline: 33172094]
23. Brown JM, Rita N, Franco-Arellano B, LeSage A, Arcand J. Co-design and refinement of curriculum-based Foodbot Factory intervention to support elementary school nutrition education. Nutrients 2024 Nov 2;16(21):3769. [doi: 10.3390/nu16213769] [Medline: 39519601]
24. Thabane L, Ma J, Chu R, et al. A tutorial on pilot studies: the what, why and how. BMC Med Res Methodol 2010 Jan 6;10(1):1. [doi: 10.1186/1471-2288-10-1] [Medline: 20053272]
25. Franco-Arellano B, Brown JM, Daggett Q, et al. Updating the Foodbot Factory serious game with new interactive engaging features and enhanced educational content. Appl Physiol Nutr Metab 2024 Jan 1;49(1):52-63. [doi: 10.1139/apnm-2023-0214] [Medline: 37905542]
26. Kiili K. Digital game-based learning: towards an experiential gaming model. Internet High Educ 2005 Jan;8(1):13-24. [doi: 10.1016/j.iheduc.2004.12.001]
27. Gordon M. Toward a pragmatic discourse of constructivism: reflections on lessons from practice. Educ Stud 2009 Feb 3;45(1):39-58. [doi: 10.1080/00131940802546894]
28. Walle JA, Bay-Williams JM, Karp KS. Elementary and Middle School Mathematics: Teaching Developmentally, 9th edition: Pearson; 2016.

XSL•FO
RenderX

29. About us. Teachers Pay Teachers. 2022. URL: https://www.teacherspayteachers.com/About-Us [accessed 2022-03-06]
30. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88. [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
31. Sekhon M, Cartwright M, Francis JJ. Development of a theory-informed questionnaire to assess the acceptability of healthcare interventions. BMC Health Serv Res 2022 Mar 1;22(1):279. [doi: 10.1186/s12913-022-07577-3] [Medline: 35232455]
32. Franco-Arellano B, Brown JM, Froome HM, LeSage A, Arcand J. Development and pilot testing of the Nutrition Attitudes and Knowledge Questionnaire to measure changes of child nutrition knowledge related to the Canada's Food Guide. Appl Physiol Nutr Metab 2021 Dec;46(12):1495-1501. [doi: 10.1139/apnm-2021-0170] [Medline: 34289315]
33. Chan CL, Taljaard M, Lancaster GA, Brehaut JC, Eldridge SM. Pilot and feasibility studies for pragmatic trials have unique considerations and areas of uncertainty. J Clin Epidemiol 2021 Oct;138:102-114. [doi: 10.1016/j.jclinepi.2021.06.029] [Medline: 34229091]
34. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharm Stat 2005 Oct;4(4):287-291. [doi: 10.1002/pst.185]
35. Johanson GA, Brooks GP. Initial scale development: sample size for pilot studies. Educ Psychol Meas 2010 Jun;70(3):394-400. [doi: 10.1177/0013164409355692]
36. R Core Team. R: a language and environment for statistical computing. : R Foundation for Statistical Computing; 2022.
37. Braga-Pontes C, Simões-Dias S, Lages M, Guarino MP, Graça P. Nutrition education strategies to promote vegetable consumption in preschool children: the Veggies4myHeart project. Public Health Nutr 2022 Apr;25(4):1061-1070. [doi: 10.1017/S1368980021004456] [Medline: 34702425]
38. Epstein DS, Enticott JC, Larson HJ, Barton C. Recruiting for research on sensitive topics in schools: an experience with Vaxcards, a collectable vaccine card game. Trials 2021 May 3;22(1):320. [doi: 10.1186/s13063-021-05288-4] [Medline: 33941247]
39. Hodges A, Cordier R, Joosten A, Bourke-Taylor H, Chen YW. Evaluating the feasibility, fidelity, and preliminary effectiveness of a school-based intervention to improve the school participation and feelings of connectedness of elementary school students on the autism spectrum. PLOS ONE 2022;17(6):e0269098. [doi: 10.1371/journal.pone.0269098] [Medline: 35648779]
40. Mabungane S, Ramroop S, Mwambi H. Analysis of missing data in progressed learners: the use of multiple imputation methods. African Journal of Research in Mathematics, Science and Technology Education 2023 May 4;27(2):112-122. [doi: 10.1080/18117295.2023.2193496]
41. van de Weijer-Bergsma E, Langenberg G, Brandsma R, Oort FJ, Bögels SM. The effectiveness of a school-based mindfulness training as a program to prevent stress in elementary school children. Mindfulness (N Y) 2012. [doi: 10.1007/s12671-012-0171-9]
42. Tickle-Degnen L. Nuts and bolts of conducting feasibility studies. Am J Occup Ther 2013;67(2):171-176. [doi: 10.5014/ajot.2013.006270] [Medline: 23433271]
43. Eldridge SM, Lancaster GA, Campbell MJ, et al. Defining feasibility and pilot studies in preparation for randomised controlled trials: development of a conceptual framework. PLoS ONE 2016;11(3):e0150205. [doi: 10.1371/journal.pone.0150205] [Medline: 26978655]
44. Lancaster GA, Thabane L. Guidelines for reporting non-randomised pilot and feasibility studies. Pilot Feasibility Stud 2019;5(1):114. [doi: 10.1186/s40814-019-0499-1] [Medline: 31608150]
45. Teresi JA, Yu X, Stewart AL, Hays RD. Guidelines for designing and evaluating feasibility pilot studies. Med Care 2022 Jan 1;60(1):95-103. [doi: 10.1097/MLR.0000000000001664] [Medline: 34812790]
46. Brown JM, Tahir S, Franco-Arellano B, et al. Efficacy of the Foodbot Factory digital curriculum-based nutrition education intervention in improving children's nutrition knowledge, attitudes and behaviours in elementary school classrooms: protocol for a cluster randomised controlled trial. BMJ Open 2025 Jan 28;15(1):e092426. [doi: 10.1136/bmjopen-2024-092426] [Medline: 39880422]

Abbreviations

NAK: Nutrition Attitudes and Knowledge

XSL•FO
RenderX

Edited by A Mavragani; submitted 25.11.24; peer-reviewed by L Aljerf, MCE Battjes-Fries; revised version received 05.03.25; accepted 18.03.25; published 16.04.25.

Please cite as:
Brown JM, Rita N, Franco-Arellano B, LeSage A, Arcand J
Evaluation of a Curriculum-Based Nutrition Education Intervention Protocol in Elementary Schools: Nonrandomized Feasibility Study
JMIR Form Res 2025;9:e69242
URL: https://formative.jmir.org/2025/1/e69242
doi:10.2196/69242

© Jacqueline Marie Brown, Nicholas Rita, Beatriz Franco-Arellano, Ann LeSage, Joanne Arcand. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluating the Acceptability of a Brief Web-Based Alcohol Misuse Prevention Program Among US Military Cadets: Mixed Methods Formative Evaluation

Emily Schmied[1,2], MPH, PhD; Lauren Hurtado[3,4], BS; W Ken Robinson[5], DMin; Cynthia M Simon-Arndt[4], MBA, MA; Richard Moyer III[6], PsyD; Leslie Wilson[7], PhD; Mark Reed[1], PhD; Shannon M Blakey[8], PhD; Marni Kan[8], PhD

[1]School of Public Health, San Diego State University, San Diego, CA, United States

[2]Institute for Behavioral and Community Health, San Diego, CA, United States

[3]Leidos Inc, San Diego, CA, United States

[4]Psychological Health and Readiness Department, Military Population Health Directorate, Naval Health Research Center, San Diego, CA, United States

[5]United States Air Force Academy, CO, United States

[6]San Diego State University Research Foundation, San Diego, CA, United States

[7]Counseling and Psychological Services, San Diego State University, San Diego, CA, United States

[8]RTI International, Research Triangle Park, NC, United States

Corresponding Author:
Emily Schmied, MPH, PhD
School of Public Health, San Diego State University, San Diego, CA, United States

Abstract

Background: As alcohol misuse remains pervasive within the military, evidence-based prevention programs that are feasible to implement and appropriately tailored to meet the needs and norms of military personnel are critically needed. Further, programs that target future military leaders, such as trainees, recruits, and cadets, may be especially impactful. eCHECKUP TO GO is a web-based, evidence-based brief alcohol intervention designed to reduce alcohol misuse through education and personalized feedback that may be suitable for military trainees. However, because it was developed for civilian students, efforts to adapt the content for military settings are needed.

Objective: This study aimed to evaluate the acceptability of a military version of eCHECKUP TO GO, tailored to include military-specific terminology and alcohol use statistics.

Methods: US Air Force Academy cadets were recruited to participate in a single-arm, mixed methods study. Following the completion of eCHECKUP TO GO, participants completed a survey that assessed satisfaction with specific aspects of the user experience, including ease of use, design, and relevance of the information and personalized feedback (range: 1=strongly disagree to 7=strongly agree). A subset of cadets also participated in a focus group to expound on the survey responses.

Results: Survey participants included 22 cadets (n=12, 55% male; mean age 19.6, SD 1.8 years). In addition, 6 (27%) cadets participated in the focus group. Participants were satisfied with the program overall (mean 5.8, SD 0.9) and gave the highest ratings to ease of use (mean 6.6, SD 0.7), site design (mean 6.5, SD 0.6), and site interactivity (mean 6.4, SD 1.0). Items pertaining to tailoring, relevance, and amount of content specific to cadets scored lowest (mean 5.8, SD 1.4; mean 5.6, SD 1.4; and mean 5.5, SD 1.5, respectively). Most (n=15, 68%) participants said they would act upon the information they were provided. Focus group participants made suggestions for improved tailoring, such as increasing content on social aspects of drinking and military-specific risks of alcohol misuse (eg, Uniform Code of Military Justice violations).

Conclusions: Although the acceptability of eCHECKUP TO GO was high, continued efforts are needed to ensure the content accurately reflects the experiences of cadets. Researchers who design military health promotion interventions need to consider the varied contexts within the force and rigorously evaluate the acceptability of all content before implementation.

(JMIR Form Res 2025;9:e67637) doi:10.2196/67637

KEYWORDS

alcohol misuse prevention; digital interventions; formative research; military health; acceptability; alcohol prevention; US military; military; United States; formative evaluation; alcohol use; evidence-based prevention; alcohol intervention; mixed methods study; survey; alcohol use disorder; alcohol misuse; heavy drinker; educational web-based intervention; web-based intervention

XSL•FO
RenderX

Introduction

Alcohol misuse is pervasive within the US military [1-4]. Data from the 2018 Department of Defense (DoD) Health Related Behaviors Survey indicated that binge drinking and heavy drinking [5] are highly prevalent (34% and 9.8%, respectively). For many service members, alcohol misuse may lead to serious adverse health, career, and social outcomes [6-8]. For instance, among all active duty service members, alcohol use disorder was among the top 10 causes of outpatient medical visits for both men and women in 2022 [2,3]. During the same period, alcohol-related issues were also a leading cause of inpatient hospitalization [3]. It has also been associated with earlier separation from service and disciplinary action [6,7]. Further, it has been noted that alcohol misuse during service may persist following separation [9,10], potentially contributing to the widely documented health disparities between veterans and civilians [11,12].

Alcohol misuse can pose serious threats to the health and performance of service members, thus interventions designed to prevent it before it causes lasting harm are urgently needed. Prior research has shown that the implementation of alcohol misuse prevention programs in the military is complicated by numerous factors, including the size, diversity, and wide geographic distribution of the force [11]. One possible approach is to target individuals in the earliest stages of their careers who likely already consume alcohol but have yet to develop problematic alcohol habits [13-15], namely those enrolled in military service academies or training settings. Another promising strategy is to use brief alcohol interventions, or BAIs, which often consist of singular sessions and thus require fewer resources to implement [16]. BAIs are endorsed for use among service members who screen positive for unhealthy alcohol use in the current Department of Veterans Affairs and DoD joint clinical practice guidelines for the Management of Substance Use Disorders [17].

BAIs have been used in military settings with varying results, though most reports are from studies conducted in veteran or general active duty populations [18]. The available evidence from training settings shows promise. For example, Klesges et al [19] evaluated the Alcohol Misconduct Prevention Program (AMPP) in Air Force technical training [20,21]. AMPP, which consisted of a combination of a single-session, group-administered BAI, and random breathalyzer tests, was associated with a significant reduction in alcohol-related risk incidents (eg, alcohol-related disciplinary infractions). While AMPP was found to be cost-effective [22], the continued implementation of in-person group trainings may not be sustainable in all settings. Further, because alcohol use is prohibited in military training settings [11], participant concerns about confidentiality may limit the impact of such methodologies [23]. Web-based BAIs that give personalized feedback to individual program participants may provide a solution to these challenges.

Prior web-based trainings (including BAIs) evaluated in active duty personnel and veterans have been shown to be acceptable and efficacious, but few have been evaluated in service academies or training settings, leaving a critical knowledge gap [24-27]. This study addresses this gap by assessing the acceptability of a web-based, individualized BAI among military cadets. eCHECKUP TO GO (alcohol-specific version; San Diego State University) is an evidence-based, theoretically driven web-based intervention originally developed for civilian adolescents and young adults. It is designed to reduce alcohol misuse by targeting cognitive risk factors (eg, peer normative behaviors) and increasing positive behavioral strategies (eg, increasing time between drinks) to minimize drinking-related risks by providing personalized feedback to participants. The efficacy of eCHECKUP TO GO has been repeatedly demonstrated in young adults [28-32]. Given the purpose, format, and prior success of eCHECKUP TO GO, it may be suitable for decreasing alcohol misuse in military academies. However, before implementing eCHECKUP TO GO, it is important to determine whether the format and existing content are acceptable among potential program participants. Specifically, while there are demographic similarities between civilian young adults and military cadets and midshipmen, the program may need to be tailored to account for cultural and contextual differences between the communities. The results of this study can be used to determine whether eCHECKUP TO GO may be appropriate for use in military academies or training settings.

Methods

Study Design and Procedures

Participants in this study were a convenience sample of 22 cadets enrolled at the US Air Force Academy (USAFA). Prospective participants were recruited through third-party announcements distributed by local study partners by email and word-of-mouth. Interested cadets attended a recruitment event where the study purpose and procedures were explained by members of the research team and those who provided written consent to participate immediately began their participation. A mixed methods design was used in which all consenting participants completed the eCHECKUP TO GO program (alcohol version) and an electronically administered, anonymous feedback survey. Immediately after completing the feedback survey, a subset of participants volunteered to participate in a 20-minute focus group discussion designed to expound on the content of the survey. The focus group was facilitated by two members (ES and LH) of the research team and was audio-recorded and transcribed for analysis. Data collection took place in January 2023.

Description of Intervention

eCHECKUP TO GO is a brief, educational, web-based intervention designed to minimize alcohol misuse by providing personalized feedback. The program includes two components. First, participants complete a self-assessment, in which they are prompted to report their demographic characteristics (eg, age, gender, height, and weight), current alcohol consumption and behaviors (eg, frequency and quantity of drinking and driving while drinking), and alcohol-related cognitions (eg, reasons for drinking, such as “alcohol helps me reduce stress”). Participants also describe their goals and priorities, such as the importance

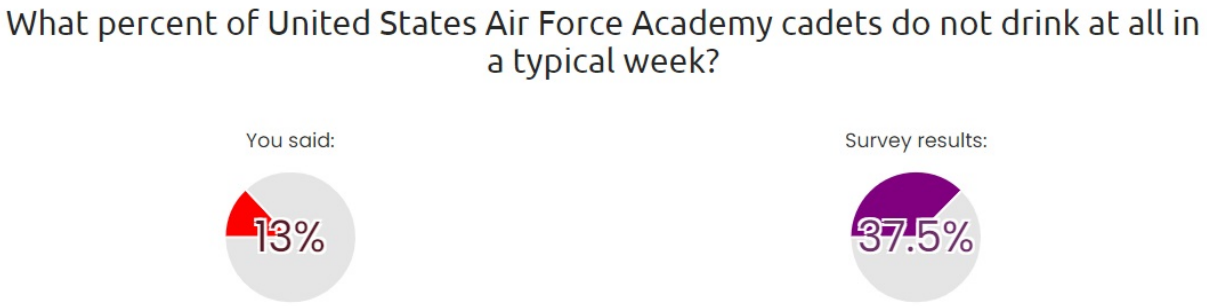
XSL•FO
RenderX

of their fitness levels and relationships with others. The assessment also prompts participants to estimate the alcohol consumption behaviors of individuals in their peer groups (ie, people of the same age, gender, and occupation). Once the assessment is completed, participants are directed to a series of modules that provide personalized feedback based on the responses provided during the assessment. The feedback modules include a range of information. For instance, one module addresses perceived peer norms by showing how the participants' alcohol consumption compares with that of their peers. Other modules show participants how much money they may spend on alcohol, their risk of becoming a problem drinker, and the potentially negative consequences of continued drinking based on their current use. Additional modules provide suggestions for how participants might change their habits and cognitions to reduce their risk for alcohol misuse and better achieve their self-reported goals (eg, alternate alcoholic drinks with water and set a drinking limit before going out). The version of eCHECKUP TO GO used in this study was adapted for use in military populations by incorporating service-specific terminology and imagery. Additionally, the peer normative feedback provided to participants was based on alcohol use data reported by Airmen in the most recent DoD Health Related Behaviors Survey [1], a trusted source used to inform military programming and resource allocation decisions. The program takes 20‐30 minutes to complete. Figures 1 and 2 show examples of program feedback.

Figure 1. eCHECKUP TO GO example alcohol misuse risk feedback.

XSL•FO
RenderX

Figure 2. eCHECKUP TO GO example peer normative feedback.



Ethical Considerations

All study procedures and measurement tools were reviewed and approved by the Institutional Review Board at San Diego State University (protocol HS-2022 - 0009) and the USAFA Human Research Protection Program and Cadet Wing. Participants provided written consent to participate, and only those who did so were invited to complete the study activities. Survey data were collected without personal identifiers and only anonymized data were used for the study; all data were stored in a secured drive with access limited to approved study personnel. Participants received no compensation for participating.

Measures

Survey

The self-report survey assessed participant demographic characteristics and satisfaction with various aspects of the eCHECKUP TO GO program, including the intended utility of the information. Sociodemographic characteristics included gender, ethnoracial identity, and age.

Satisfaction With eCHECKUP TO GO Content and User Experience

An abbreviated 11-item version of the general Web trust questionnaire [33-35] assessed perceptions of aspects of the program known to be important to individuals' reviews of web-based health information: credibility of information (ie, "The advice seemed credible" and "I trusted the information on the site"), quality of the information and content (ie, "The site had a professional design"), and personalization of information (ie, "It felt like the advice was tailored to me personally"). Response options ranged from 1=strongly disagree to 7=strongly agree. Three original items designed by the researchers were added to obtain more specific information about participants' perceptions of the program content. Two were included to assess satisfaction with the alcohol-related content (eg, "the feedback I received about alcohol use"), and another to assess satisfaction with the military content (ie, "amount of information about cadets or military service members"). Participants rated their degree of satisfaction on a scale from 1=completely dissatisfied to 7=completely satisfied. In addition to the rating scales, participants were asked to respond to two open-ended questions: "What did you like about the program?" and "How do you think the program could be improved to best help others?"

Use of Information

The intended outcome of eCHECKUP TO GO is to inspire behavioral changes related to alcohol, so participants were also asked at the conclusion of the program whether they intended to act on the advice they received related to alcohol consumption (yes or no). If yes, they were asked to explain in a free-response field which of the actions or strategies they intended to use, thereby providing an additional indicator of program satisfaction and the perceived relevance of the information [33].

Focus Group Guide

The focus group guide included nine questions that assessed participants' perceptions of the eCHECKUP TO GO program. Participants were prompted to provide their opinions on the aesthetics, functional features, and content of the program (eg, "Tell me what you thought about the look and feel of the program" and "How relevant did you find the recommendations were to you?"). Participants were also asked for general feedback (ie, "What else can we do to improve the eCHECKUP TO GO program?").

Analysis

Quantitative survey data and qualitative data were first analyzed separately. Descriptive statistics, including means and frequencies, were computed for all close-ended survey items to assess the distribution and completeness of responses. Open-ended survey questions and focus group data were combined and analyzed using the rigorous and accelerated data reduction (RADaR) technique [36]. The RADaR technique was selected to provide expedited feedback about necessary program adaptations to be implemented before making eCHECKUP TO GO more widely available to cadets [37]. Using RADaR, two coauthors (ES and LH) engaged in an iterative process of data reduction and coding to identify pertinent themes related to the following study research questions: (1) What aspects of eCHECKUP TO GO were participants most satisfied with? and (2) How could eCHECKUP TO GO be improved for use among cadets? Once a codebook was finalized, two analysts (ES and LH) independently reviewed all study data and compared results to identify potential discrepancies. Few discrepancies were identified and all were resolved through discussion. Study analyses were completed by integrating quantitative and qualitative data to examine their convergence and complementarity, with results shown in a joint display [38,39]. Data collection and analyses were led by researchers without

XSL•FO
RenderX

firsthand military service experience. The lead analyst (ES) is a former military beneficiary and the second analyst is a civilian (LH); both have experience conducting military health research.

Results

Overview

The survey participants were 22 cadets. A slight majority (n=12, 54%) were male. The mean age was 19.64 (SD 1.86; range 18 - 26) years, with 6 (27%) participants aged 21 years or older. Approximately, 40% (n=9) of participants identified as non-Hispanic White, approximately one-fifth (n=4, 18%) of participants identified as Hispanic or Latinx or multiracial, 14% (n=3) of participants identified as Black or African American, and 4% (n=1) of participants identified as Asian Pacific Islander or Asian. When compared to the cadet population at large, this sample includes proportionally more female and individual participants from minoritized ethnoracial groups [40]. About one-quarter (n=6, 27%) also participated in the focus group. Demographic characteristics were not recorded for focus group participants to preserve their anonymity; thus, it is not possible to determine whether they significantly differed from the larger group. Further, participants were not asked to report their current or prior alcohol use behaviors because alcohol use is not permitted in this population or setting.

Results of the survey and focus group are jointly displayed in Table 1. Through analysis of the qualitative data, the research team identified three overarching themes addressed by participants: design and functionality, credibility of content, and tailoring and relevance of content. Further, it was observed that participant satisfaction with these aspects of eCHECKUP TO GO differed, with more positive responses provided about the design and functionality than the other themes. A review of the quantitative survey responses revealed a parallel pattern, such that items pertaining to design and functionality were rated highest, while items pertaining to tailoring and relevance of content were rated lowest. A summary of the feedback provided by participants is presented here by theme.

XSL•FO
RenderX

Table . Joint display of quantitative and qualitative responses demonstrating participant perception of the eCHECKUP TO GO program.

Theme	Survey responses and exemplifying quotations from open-ended survey items and focus groups
Design and functionality	 • Exemplifying quotations: satisfaction • "The program was very accessible and easy to understand in terms of direction." [SP[a] #19] • "I liked how interactive the overall site was and how it gave me information based on my inputs." [SP #1] • Exemplifying quotations: suggestions for improvement • "I think the program is overall very good. Some of the questions I misinterpreted so having like a small description below with a better or more explanation of the question could probably help." [SP #7]
Credibility of content	 • Exemplifying quotations: satisfaction • "I appreciated the stats." [SP #21] • "I like that it gave a lot of helpful information on drinking and how to be a safer drinker." [SP #16] • "Yeah, I think it was very feasible. I guess, having the discipline to do it." [FGP[b] #3] • Exemplifying quotations: suggestions for improvement • "Provide more information and stats on alcohol use in the military, specifically cadets." [SP #21]
Tailoring and relevance of content	 • Exemplifying quotations: satisfaction • "Yeah, I liked how if you didn't put yourself down as drinking. It would, it was like, here's things you can do to recommend to others instead of you need to do these things." [FGP #2] • "I think by asking us to input what we thought the statistics were and what was important to us made it much more personal." [SP #12] • Exemplifying quotations: Suggestions for improvement • "You had all this stuff like with the airmen and cadets, but when you went to the consequences, it was largely just like social and physical consequences, maybe like help emphasize like the military aspect of it ... Just to reinforce that like drinking can ruin your career, which is what we're all here to set up for." [FGP #2] • "Provide more military-tailored advice ([it's a] different drinking culture here than civilian schools)." [SP #10]

[a]SP: survey participant.

[b]FGP: focus group participant.

Design and Functionality

On the survey, participants were most satisfied with the ease of use (on the 1 - 7 scale: mean 6.64, SD 0.73), how easy it was to understand the information (mean 6.45, SD 0.91), and the design and interactivity (mean 6.55, SD 0.60; and mean 6.45, SD 1.01, respectively). The qualitative data additionally elucidated the design features of the program most liked by participants. Several survey respondents commented on the features of the program that made it easy to use. For example, one participant wrote, "I thought it was clean and easy to follow and did not have unnecessary filler" [survey participant (SP) #17]. In the focus group, a participant who appreciated the embedded skip logic that allowed participants to advance throughout the assessment based on their responses said, "I liked in the parts where it was like kind of the yes or no that would just automatically move down to the next question instead of like having to scroll after answering each one" [focus group participant (FGP) #4]. Others commented on the interactivity of the program ("I liked how interactive the overall site was and how it gave me information based on my inputs" [SP #3]).

When it came to suggesting improvements in the design and functionality, most participants commented on the

XSL•FO
RenderX

self-assessment component of the program. Several participants described difficulty completing the drinking history questions, which are structured in the format of a calendar where respondents enter the number and type of drinks they consumed on different days of the week and during certain hours. In the focus group, two participants described difficulty with the calendar, stating that "the calendar was too stressful" [FGP #2] and "I don't pay attention to, like, how many drinks I had an hour or in the specific hours, which like, how long I've been drinking" [FGP #3].

Credibility of Content

Survey responses revealed a high degree of trust in the information and suggestions provided in eCHECKUP TO GO, as indicated by the high scores for credibility of information (mean 6.36, SD 0.66) and feasibility of alcohol misuse prevention strategies provided (mean 6.41, SD 0.79). The qualitative results helped to explain the highest ratings, as the participant responses revealed an appreciation for data, which are frequently presented to eCHECKUP TO GO users. One respondent noted, "I appreciated the stats" [SP #21], while others remarked, "I liked how it showed graphs and other ways for people to see how those products affect themselves and their peers" [SP #4], and "[I liked the] statistics at the end, as well as the estimation questions to display knowledge" [SP #10]. Others noted the specific information they recalled, such as "I really liked the part where it told you about like the calories and like how it can impact muscle mass, especially because physicality is such a big thing here that if, like, if people are drinking, they just don't realize the calories and how it can hurt you in your muscles" [FGP #2]. Further showing the importance of data to participants, the suggestions provided related to the credibility of content were to add even more data ("Maybe more data and statistics" [SP #12]).

Tailoring and Relevance of Content

As shown in Table 1, the scores for tailoring and relevance of eCHECKUP TO GO content were relatively lower (ie, amount of information about cadets: mean 5.48, SD 1.57). Qualitative responses provided context for these lower scores, showing that while many participants appreciated the personalized feedback they received, they felt that the content did not include enough information tailored to the unique experiences of military cadets. In terms of the personalized feedback, one participant noted that what they liked best about the program was it was "very personally [sic] to me, showed me the statistics, how much I am at risk" [SP #22], while another said, "I think personally it made me more aware of how much I was drinking or how much others are drinking ... it can make us more aware of our habits and our spending" [FGP #3].

When describing opportunities to improve the program tailoring, focus group participants discussed a need for more information on the social aspect of drinking. One participant explained:

> *There could be ... a lengthy section for the social aspect of drinking ... a lot of cadets just go out and drink together. And I don't think the survey emphasized that really enough. I think it was individually tailored, which I think is good. But also, it didn't go into the, too much, in my opinion, to the causes of like why people go out and drink, maybe just because of like the social pressure, because again, especially being cadets, there's always that implied social pressure.* [FGP #2]

Another participant also suggested adding more content on peer pressure and drinking culture, saying, "I feel like there should be a section on ... how much of the overall culture, overall climate of drinking contribute to your own and like do others pressure you to drink and stuff like that" [FGP #4]. Relatedly, another cadet requested content about "how to help friends and notice if they have an alcohol problem" [SP #22].

Participants recommended improving content tailoring to address other aspects of cadet or military life as well. Some felt that the existing content of eCHECKUP TO GO did not consider the unique schedules of cadets. They explained that there are periods of time when they are forced to completely abstain from drinking and do not have access to alcohol and then periods where it is almost expected that they would indulge, making it hard to answer the self-assessment items that asked about behaviors in the past week or a "typical" month. According to one participant, "We don't get off base much. So, our drinking habits may look very irrational, making it hard to say I drink x on a typical week" [SP #16]. Other suggestions pertained to the personalized feedback provided about the consequences of drinking. It was suggested that the military-specific consequences be included, which previously were not in the program:

> *When you went to the consequences, it was largely just like social and physical consequences. Maybe, like help emphasize, like, the military aspect of it. [Such as] like this many people a year get put on alcohol probation here [or] like in the actual Air Force, like this many people get in trouble for drinking, and like drinking can lead to these Uniform Code of Military Justice violations and has been linked to them or something like that. Just to reinforce that like drinking can ruin your career, which is what we're all here to set up for.* [FGP #2]

Another noted that the eCHECKUP TO GO program provided feedback about how much money they spent on alcohol compared with a rent payment, which cadets do not pay, adding, "So, we don't pay rent or at least we don't see it ... So, I guess if it's specific to cadets, maybe just don't add that" [FGP #3]. Finally, some requested "more military-tailored advice (different drinking culture here than civilian schools)" [SP #21].

Future Use of Alcohol Misuse Prevention Strategies

Over two-thirds of participants (n=15, 68%) said they intended to use at least one alcohol misuse prevention strategy they learned from eCHECKUP TO GO. Those who answered affirmatively were asked to describe which, if any, of the alcohol misuse prevention strategies learned from eCHECKUP TO GO they would use in the future. Through qualitative analysis, it was determined that responses could be categorized into two salient themes: strategies to reduce the amount consumed and strategies to protect others. Several participants planned to reduce their intake by spacing out their drinks, saying, "I intend

XSL•FO
RenderX

to use the 'spacing drinks over time' strategy" [SP #3] and "When reaching the legal drinking age, to practice safe drinking habits like spacing out drinks and setting limits for myself" [SP #20]. Additionally, numerous participants planned to use strategies to protect their peers, such as one participant who explained, "Even if I am not drinking and don't have to worry about putting myself in danger, I can look out for my friends when they are drinking and make sure they are making good decisions regarding alcohol" [SP #1]. Others indicated that they would act as the designated driver for their friends, stating, "I will likely be the DD for my group on more occasions" [SP #17].

Discussion

Principal Findings

Addressing the longstanding issue of alcohol misuse in the military requires creative solutions designed to mitigate the challenges inherent to promoting health in large, diverse communities [1,2,4,11,41]. This study sought to determine whether a military-tailored version of eCHECKUP TO GO is an acceptable option for use among military cadets, recruits, or trainees. Study results identified aspects of the program that were well received by participants, including the web-based format and interface, and the personalized feedback about alcohol risk and protective strategies. Importantly, no concerns were expressed by participants about confidentiality. These findings suggest that this internet-based, individually completed BAI may be acceptable in this community, providing a possible solution to the logistical issues and privacy concerns that often impede the implementation and efficacy of health promotion programs in military settings.

Although eCHECKUP TO GO was largely endorsed by participants, opportunities for improvement were identified. Though participants liked the self-assessment and personalized feedback, many felt that these modules did not sufficiently acknowledge the broader social context impacting alcohol use, particularly cultural norms and peer influence. Heavy alcohol use is widely known to be a norm within the military and among young adults, and research has shown that consuming alcohol with peers may be viewed as an opportunity to build cohesion [41-46]. Participants touched on these influences when they described situations in which they felt expected to drink or chose to drink to avoid feeling left out. These findings align with a study by Meadows et al [43] that found positive associations between perceived alcohol culture within respondents' command or unit and binge and heavy drinking. While alcohol use is an individual behavior, it cannot be disentangled from the social context in which it occurs when working with military cadets or trainees [43]. Thus, future alcohol misuse prevention interventions may be more acceptable and efficacious if they directly address the impact of peer influence and provide concrete strategies for navigating social situations in which alcohol may be present.

Study results also highlight another cultural norm that should be incorporated into eCHECKUP TO GO and other alcohol misuse prevention programming: service before self. "Service before self" is a core value of the Air Force [47], and the other service branches uphold similar maxims. Feedback from study participants indicates that their ingrained commitment to service should be used as a cornerstone of alcohol misuse prevention content. For example, participants who intended to act upon the advice received as part of the program largely intended to use strategies to protect their comrades, and some asked for more content on how to help recognize alcohol misuse in others. Further, participants felt the program could be more impactful if it included information about career consequences of alcohol misuse, such as the Uniform Code of Military Justice violations. These findings parallel results from a study by Aycock et al [48], which also noted that alcohol misuse prevention messaging for Airmen should emphasize adverse career impacts. Framing alcohol misuse interventions as a means of protecting comrades and fulfilling a commitment to serve could bolster buy-in and work to dismantle longstanding alcohol-related cultural norms [41-43].

Limitations

Methodological limitations must be considered when interpreting the results of this study. Due to the relatively small sample size and unique population, the results may not generalize to the USAFA community or other military service academies. Additional formative research may be needed to finalize intervention refinement before subsequent efficacy trials. In particular, more data may be needed to ensure the experiences and preferences of participants from minoritized communities who have historically been underrepresented in research are considered. Additionally, while no concerns about anonymity were expressed by participants, it remains possible that participants did not feel comfortable discussing their opinions about alcohol use out of fear of accidental disclosure of their identity. However, the responses received likely reflect participants' opinions and behaviors because participants were given the option of skipping questions they did not wish to answer. Finally, as this was a formative evaluation, the results cannot be used to estimate the efficacy of eCHECKUP TO GO completion on alcohol misuse among cadets. The next steps in this line of research should include an efficacy trial powered sufficiently to determine the impact of eCHECKUP TO GO completion on attitudes toward alcohol misuse, alcohol-related protective behavioral strategies, and alcohol consumption among military cadets.

Conclusions

The results of this study indicate that a military version of eCHECKUP TO GO may be suitable for use in military training settings, though additional tailoring is needed to account for the unique norms and values of this community. Study findings have important implications for the development and implementation of military alcohol misuse prevention programs, but the lessons learned can also be applied to programs that address other health behaviors. In communities that value service above self, program content should emphasize the ways in which improving individual health behaviors can benefit the community at large.

XSL•FO
RenderX

Acknowledgments

This work was sponsored by the US Government under Other Transactions number W81XWH-21-9-0011.012, Award W81XWH-21-9-0011.012. The US Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the US Government. WKR and CMSA are employees of the US Government. This work was prepared as part of their official duties. Title 17, U.S.C. §105 provides that copyright protection under this title is not available for any work of the US Government. Title 17, U.S.C. §101 defines US Government work as work prepared by a military service member or employee of the US Government as part of that person's official duties. This work was supported by the Defense Health Agency Medical Technology Enterprise Consortium under work unit 61135. The views expressed in this work are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, or the US Government.

Data Availability

Data are not available for public use because participants did not provide consent for the data to be used for additional research purposes.

Authors' Contributions

MK, SMB, ES, WKR, MR, LW, and RM conceptualized the project. ES, LH, and CMSA collected all project data. ES and LH analyzed project data. ES led the development of the manuscript; all other authors contributed to editing.

Conflicts of Interest

LW and RM may earn royalties for the use of the eCHECKUP TO GO Program.

References

1. Meadows SO, Engel CC, Collins RL, et al. 2015 department of defense health related behaviors survey (HRBS). Rand Health Q 2018;8:5. [Medline: 30323988]
2. Armed Forces Health Surveillance Division. Ambulatory visits among active component members, U.S. Armed Forces, 2022. MSMR 2023 Jun 20;30(6):19-25. [Medline: 37540151]
3. Armed Forces Health Surveillance Division. Hospitalizations among active component members, U.S. Armed Forces, 2022. MSMR 2023 Jun 20;30(6):12-18. [Medline: 37540152]
4. Armed Forces Health Surveillance Division. Absolute and relative morbidity burdens attributable to various illnesses and injuries, active component, U.S. Armed Forces, 2021. MSMR 2022 Jun 1;29(6):2-9. [Medline: 36250737]
5. Drinking levels defined. National Insitute on Alcohol Abuse and Alcoholism. 2023. URL: https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking#:~:text=NIAAA%20defines%20binge%20drinking%20as,)%2C%20in%20about%202%20hours [accessed 2024-06-01]
6. Mattiko MJ, Olmsted KLR, Brown JM, Bray RM. Alcohol use and negative consequences among active duty military personnel. Addict Behav 2011 Jun;36(6):608-614. [doi: 10.1016/j.addbeh.2011.01.023] [Medline: 21376475]
7. Porter B, Rodriguez LM, Woodall KA, Pflieger JC, Stander VA. Alcohol misuse and separation from military service: a dyadic perspective. Addict Behav 2020 Nov;110:106512. [doi: 10.1016/j.addbeh.2020.106512] [Medline: 32623237]
8. Waller M, McGuire ACL, Dobson AJ. Alcohol use in the military: associations with health and wellbeing. Subst Abuse Treat Prev Policy 2015 Jul 28;10:27. [doi: 10.1186/s13011-015-0023-4] [Medline: 26216215]
9. Norman SB, Schmied E, Larson GE. Predictors of continued problem drinking and substance use following military discharge. J Stud Alcohol Drugs 2014 Jul;75(4):557-566. [doi: 10.15288/jsad.2014.75.557] [Medline: 24988254]
10. Derefinko KJ, Hallsell TA, Isaacs MB, et al. Substance use and psychological distress before and after the military to civilian transition. Mil Med 2018 May 1;183(5-6):e258-e265. [doi: 10.1093/milmed/usx082] [Medline: 29420791]
11. Haibach JP, Haibach MA, Hall KS, et al. Military and veteran health behavior research and practice: challenges and opportunities. J Behav Med 2017 Feb;40(1):175-193. [doi: 10.1007/s10865-016-9794-y] [Medline: 27678001]
12. Hoerster KD, Lehavot K, Simpson T, McFall M, Reiber G, Nelson KM. Health and health behavior differences: U.S. military, veteran, and civilian men. Am J Prev Med 2012 Nov;43(5):483-489. [doi: 10.1016/j.amepre.2012.07.029] [Medline: 23079170]
13. Shirvani M, Reed MB, Clingan S. The relationship between emerging adult alcohol consumption and military enlistment. Mil Med 2017 Sep;182(9):e1731-e1737. [doi: 10.7205/MILMED-D-16-00312] [Medline: 28885929]
14. Derefinko KJ, Klesges RC, Bursac Z, Little MA, Hryshko-Mullen A, Talcott GW. Alcohol issues prior to training in the United States Air Force. Addict Behav 2016 Jul;58:142-148. [doi: 10.1016/j.addbeh.2016.02.028] [Medline: 26945450]

XSL•FO
RenderX

15. Taylor JE, Haddock K, Poston WSC, Talcott WG. Relationship between patterns of alcohol use and negative alcohol-related outcomes among U.S. Air Force recruits. Mil Med 2007 Apr;172(4):379-382. [doi: 10.7205/milmed.172.4.379] [Medline: 17484307]
16. Brief alcohol interventions for unhealthy alcohol use. : Department of Defense Psychological Health Center of Excellence; 2023 URL: https://health.mil/Reference-Center/Publications/2021/04/26/PHCoE-Evidence-Brief-Alcohol-Interventions-for-Unhealthy-Alcohol-Use-508 [accessed 2025-03-28]
17. Va/dod clinical practice guidelines for the management of substance use disorder (2021). : U.S. Department of Veterans Affairs; 2021.
18. Doherty AM, Mason C, Fear NT, Rona R, Greenberg N, Goodwin L. Are brief alcohol interventions targeting alcohol use efficacious in military and veteran populations? A meta-analysis. Drug Alcohol Depend 2017 Sep 1;178:571-578. [doi: 10.1016/j.drugalcdep.2017.05.029] [Medline: 28750345]
19. Klesges RC, Talcott W, Ebbert JO, et al. Effect of the Alcohol Misconduct Prevention Program (AMPP) in air force technical training. Mil Med 2013 Apr;178(4):445-451. [doi: 10.7205/MILMED-D-12-00400] [Medline: 23707832]
20. Little MA, Klesges RC, Mallawaarachchi I, et al. Prevention of alcohol-related incidents in the U.S. Air Force: results from a cluster randomized trial. Mil Med 2023 Jul 22;188(7-8):e1874-e1881. [doi: 10.1093/milmed/usab489]
21. Derefinko KJ, Linde BD, Klesges RC, et al. Dissemination of the brief alcohol intervention in the United States Air Force: study rationale, design, and methods. Mil Behav Health 2018 Jan 2;6(1):108-117. [doi: 10.1080/21635781.2017.1397569]
22. Li T, Waters TM, Kaplan EK, et al. Economic analyses of an alcohol misconduct prevention program in a military setting. Mil Med 2017 Jan;182(1):e1562-e1567. [doi: 10.7205/MILMED-D-16-00098] [Medline: 28051974]
23. Griffin KW, Williams C, Travis W, Tharp A. Prevention of unwanted sexual contact among cadets at the United States Air Force academy: a brief small-group intervention. Am J Public Health 2021 Apr;111(4):672-674. [doi: 10.2105/AJPH.2020.306050] [Medline: 33476231]
24. Pedersen ER, Parast L, Marshall GN, Schell TL, Neighbors C. A randomized controlled trial of A web-based, personalized normative feedback alcohol intervention for young-adult veterans. J Consult Clin Psychol 2017 May;85(5):459-470. [doi: 10.1037/ccp0000187] [Medline: 28287799]
25. Hurtado S, Shaffer RA, Schuckit MA, et al. Evaluation of an alcohol misuse prevention program in a military population. : Naval Health Research Center; 2003 URL: https://apps.dtic.mil/sti/tr/pdf/ADA421218.pdf [accessed 2025-03-28]
26. Pemberton MR, Williams J, Herman-Stahl M, et al. Evaluation of two web-based alcohol interventions in the U.S. military. J Stud Alcohol Drugs 2011 May;72(3):480-489. [doi: 10.15288/jsad.2011.72.480] [Medline: 21513685]
27. Hester RK, Squires DD, Delaney HD. The Drinker's Check-up: 12-month outcomes of a controlled clinical trial of a stand-alone software program for problem drinkers. J Subst Abuse Treat 2005 Mar;28(2):159-169. [doi: 10.1016/j.jsat.2004.12.002] [Medline: 15780546]
28. Doumas DM, Esp S, Turrisi R, Hausheer R, Cuffee C. A test of the efficacy of a brief, web-based personalized feedback intervention to reduce drinking among 9th grade students. Addict Behav 2014 Jan;39(1):231-238. [doi: 10.1016/j.addbeh.2013.10.011] [Medline: 24148137]
29. Doumas DM, Haustveit T, Coll KM. Reducing heavy drinking among first year intercollegiate athletes: a randomized controlled trial of web-based normative feedback. J Appl Sport Psychol 2010 Jul 23;22(3):247-261. [doi: 10.1080/10413201003666454]
30. Doumas DM, Mastroleo NR. Reducing alcohol use among high school athletes: a randomized controlled trial of a brief, web-based personalized normative feedback intervention. Sport Psychol 2021 Mar;35(1):55-63. [doi: 10.1123/tsp.2020-0022] [Medline: 38962423]
31. Ganz T, Braun M, Laging M, Schermelleh-Engel K, Michalak J, Heidenreich T. Effects of a stand-alone web-based electronic screening and brief intervention targeting alcohol use in university students of legal drinking age: a randomized controlled trial. Addict Behav 2018 Feb;77:81-88. [doi: 10.1016/j.addbeh.2017.09.017] [Medline: 28985586]
32. Doumas DM, Esp S, Turrisi R, Bond L. A randomized controlled trial of the eCHECKUP to GO for high school seniors across the academic year. Subst Use Misuse 2021;56(13):1923-1932. [doi: 10.1080/10826084.2021.1958862] [Medline: 34347564]
33. Briggs P, Burford B, De Angeli A, Lynch P. Trust in online advice. Soc Sci Comput Rev 2002;20(3):321-332. [doi: 10.1177/08939302020003009]
34. Harris PR, Sillence E, Briggs P. Perceived threat and corroboration: key factors that improve a predictive model of trust in internet-based health information and advice. J Med Internet Res 2011 Jul 27;13(3):e51. [doi: 10.2196/jmir.1821] [Medline: 21795237]
35. Sillence E, Blythe JM, Briggs P, Moss M. A revised model of trust in internet-based health information and advice: cross-sectional questionnaire study. J Med Internet Res 2019 Nov 11;21(11):e11125. [doi: 10.2196/11125] [Medline: 31710297]
36. Watkins DC. Rapid and rigorous qualitative data analysis: the "RADaR" technique for applied research. Int J Qual Methods 2017;16(1). [doi: 10.1177/1609406917712131]

XSL•FO
RenderX

37. Gale RC, Wu J, Erhardt T, et al. Comparison of rapid vs in-depth qualitative analytic methods from a process evaluation of academic detailing in the Veterans Health Administration. Implement Sci 2019 Feb 1;14(1):11. [doi: 10.1186/s13012-019-0853-y] [Medline: 30709368]
38. Palinkas LA, Aarons GA, Horwitz S, Chamberlain P, Hurlburt M, Landsverk J. Mixed method designs in implementation research. Adm Policy Ment Health 2011 Jan;38(1):44-53. [doi: 10.1007/s10488-010-0314-z] [Medline: 20967495]
39. Guetterman TC, Fàbregues S, Sakakibara R. Visuals in joint displays to represent integration in mixed methods research: a methodological review. Methods Psychol 2021 Dec;5:100080. [doi: 10.1016/j.metip.2021.100080]
40. Demographic profile of the incomine USAFA class of 2025. Department of Defense, United States Air Force Academy. URL: https://www.usafa.edu/app/uploads/CL2025-Class-Profile.pdf [accessed 2025-02-07]
41. Osborne AK, Wilson-Menzfeld G, McGill G, Kiernan MD. Military service and alcohol use: a systematic narrative review. Occup Med (Lond) 2022 Jul 11;72(5):313-323. [doi: 10.1093/occmed/kqac045] [Medline: 35674143]
42. Teachman J, Anderson C, Tedrow LM. Military service and alcohol use in the United States. Armed Forces Soc 2015 Jul;41(3):460-476. [doi: 10.1177/0095327X14543848]
43. Meadows SO, Beckman R, Engel CC, Jeffery DD. The culture of alcohol in the U.S. military: correlations with problematic drinking behaviors and negative consequences of alcohol use. Armed Forces Soc 2023 Apr;49(2):531-555. [doi: 10.1177/0095327X211069162]
44. Knox J, Schneider J, Greene E, Nicholson J, Hasin D, Sandfort T. Using social network analysis to examine alcohol use among adults: a systematic review. PLoS ONE 2019;14(8):e0221360. [doi: 10.1371/journal.pone.0221360] [Medline: 31437257]
45. Ames GM, Duke MR, Moore RS, Cunradi CB. The impact of occupational culture on drinking behavior of young adults in the U.S. Navy. J Mixed Methods Res 2009 Apr;3(2):129-150. [doi: 10.1177/1558689808328534]
46. Besse K, et al. How soldiers perceive the drinking environment in communities near military installations. J Alcohol Drug Educ 2018;62(1):71-90 [FREE Full text]
47. Vision: it's our promise to protect. US Air Force. URL: https://www.airforce.com/vision [accessed 2024-06-15]
48. Aycock CA, Mallawaarachchi I, Wang XQ, et al. Developing a text messaging intervention to prevent binge and heavy drinking in a military population: mixed methods development study. JMIR Form Res 2024 Mar 19;8:e55041. [doi: 10.2196/55041] [Medline: 38502165]

Abbreviations

AMPP: Alcohol Misconduct Prevention Program
BAI: brief alcohol intervention
DoD: Department of Defense
FGP: focus group participant
RADaR: rigorous and accelerated data reduction
SP: survey participant
USAFA: US Air Force Academy

Edited by A Mavragani; submitted 16.10.24; peer-reviewed by M Choi, T Bakas; revised version received 10.02.25; accepted 11.02.25; published 16.04.25.

Please cite as:
Schmied E, Hurtado L, Robinson WK, Simon-Arndt CM, Moyer III R, Wilson L, Reed M, Blakey SM, Kan M
Evaluating the Acceptability of a Brief Web-Based Alcohol Misuse Prevention Program Among US Military Cadets: Mixed Methods Formative Evaluation
JMIR Form Res 2025;9:e67637
URL: https://formative.jmir.org/2025/1/e67637
doi:10.2196/67637

© Emily Schmied, Lauren Hurtado, W Ken Robinson, Cynthia M Simon-Arndt, Richard Moyer III, Leslie Wilson, Mark Reed, Shannon M Blakey, Marni Kan. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impact of a Virtual Reality Video ("A Walk-Through Dementia") on YouTube Users: Topic Modeling Analysis

Xiaoli Li[1], PhD; Xiaoyu Liu[2], PhD; Cheng Yin[3], PhD; Sandra Collins[1], PhD; Eman Alanazi[4], PhD

[1]School of Health Sciences, Southern Illinois University Carbondale, 1365 Douglas Drive, ASA Building - Mail Code 6615, Carbondale, IL, United States

[2]College for Public Health & Social Justice, Saint Louis University, Saint Louis, MO, United States

[3]Department of Rehabilitation and Health Services, University of North Texas, Denton, TX, United States

[4]Health Informatics Department, College of Health Sciences, Saudi Electronic University, Riyadh, Saudi Arabia

Corresponding Author:
Xiaoli Li, PhD
School of Health Sciences, Southern Illinois University Carbondale, 1365 Douglas Drive, ASA Building - Mail Code 6615, Carbondale, IL, United States

Abstract

Background: Emerging research has highlighted the potential of virtual reality (VR) as a tool for training health care students and professionals in care skills for individuals with Alzheimer disease and related dementias (ADRD). However, there is limited research on the use of VR to engage the general public in raising awareness about ADRD.

Objective: This research aimed to examine the impact of the VR video “A Walk-Through Dementia” on YouTube users by analyzing their posts.

Methods: We collected 12,754 comments from the VR video series “A Walk-Through Dementia,” which simulates the everyday challenges faced by individuals with ADRD, providing viewers with an immersive experience of the condition. Topic modeling was conducted to gauge viewer opinions and reactions to the videos. A pretrained Bidirectional Encoder Representations from Transformers (BERT) model was used to transform the YouTube comments into high-dimensional vector embeddings, allowing for systematic identification and detailed analysis of the principal topics and their thematic structures within the dataset.

Results: We identified the top 300 most frequent words in the dataset and categorized them into nouns, verbs, and adjectives or adverbs using a part-of-speech tagging model, fine-tuned for accurate tagging tasks. The topic modeling process identified eight8 initial topics based on the most frequent words. After manually reviewing the 8 topics and the content of the comments, we synthesized them into 5 themes. The predominant theme, represented in 2917 comments, centered on users’ personal experiences with the impact of ADRD on patients and caregivers. The remaining themes were categorized into 4 main areas: positive reactions to the VR videos, challenges faced by individuals with ADRD, the role of caregivers, and learning from the VR videos.

Conclusions: Using topic modeling, this study demonstrated that VR applications serve as engaging and experiential learning tools, offering the public a deeper understanding of life with ADRD. Future research should explore additional VR applications on social media, as they hold the potential to reach wider audiences and effectively disseminate knowledge about ADRD.

(JMIR Form Res 2025;9:e67755) doi:10.2196/67755

KEYWORDS

Alzheimer disease; Alzheimer disease and related dementias; ADRD; dementia; public awareness; text mining; older adult; health care student; training; health care professional; simulation; digital world; virtual environment; virtual tour; walk-through dementia; virtual reality; VR; VR video; VR application; topic modeling; YouTube; Bidirectional Encoder Representations from Transformers; BERT; social media comments; experiential learning tool

Introduction

With a rapidly aging population, the number of older adults living with chronic conditions is on the rise. Among these conditions, Alzheimer disease and related dementias (ADRD) have become leading causes of disability and dependency worldwide, characterized by memory loss and other cognitive impairments [1]. In the later stages, individuals with ADRD may lose the ability to interact with their surroundings, making care management increasingly complex both within health care settings and at home [2]. This condition not only affects the quality of life for those diagnosed but also places a significant burden on health care systems and families. While most older adults prefer to stay in their homes as long as possible, many eventually require higher levels of care, often transitioning to nursing homes or memory care facilities [3]. Caregivers—whether family members or professionals—must

XSL•FO
RenderX

have a thorough understanding of ADRD to provide the best care possible. In addition, various policy initiatives, such as those promoting the creation of "dementia-friendly communities" [4], underscore the importance of raising public awareness about dementia. These communities aim to foster environments where people understand ADRD and where individuals with dementia feel included, involved, and empowered to maintain control over their daily lives [5]. Such initiatives emphasize the need for widespread dementia awareness and education within communities to improve the support and care provided to those affected by ADRD.

However, there is a lack of public awareness and understanding of what ADRD is and its impact on individuals' abilities, well-being, and capacity to live meaningful lives [6]. As a result, misunderstanding and discrimination among people living with ADRD are common. For instance, there are false beliefs that people with ADRD cannot experience quality of life, are entirely dependent on others, and have lost all autonomy and dignity [6]. Such stigma and misunderstandings can have negative effects on the psychosocial well-being, self-esteem, and overall quality of life for both those living with ADRD and their caregivers. In addition, stigma can contribute to social exclusion, isolation, and even self-stigmatization, further exacerbating the challenges faced by individuals with ADRD and their families [7].

Disseminating reliable information about ADRD is especially crucial because addressing modifiable factors can help reduce misunderstandings, provide support to caregivers, and raise awareness about the condition [8,9]. Social media platforms, particularly YouTube, have become powerful tools for disseminating health-related information to audiences that are not easily accessible [10]. With 73% of middle-aged adults and 45% of older adults in the United States using social media such as Facebook (Meta) and YouTube, these platforms are vital tools for raising awareness, educating communities, and deepening public understanding of ADRD [11]. They not only facilitate the sharing of personal stories but also provide educational content, making them well-positioned to increase knowledge and combat the stigma associated with ADRD [12].

Previous studies have explored various topics involving YouTube and ADRD, such as assessing the quality of ADRD caregiving information on the platform [13], describing social media activities for aging and ADRD education [14], developing internet-based ADRD training modules [15], examining YouTube's effectiveness in dementia education to a target community [16], and exploring the role of personal stories in raising awareness about dementia [17]. In addition, research has examined the role of personal stories in raising awareness about dementia and how social media mobilizes knowledge about pain management in dementia [18]. However, these studies focused on traditional videos as educational tools, and no research has examined the use of virtual reality (VR) videos for public ADRD education.

VR videos provide immersive, 360-degree environments that allow users to engage with simulated worlds using devices like enclosed headsets or body suits [19,20]. In the field of ADRD education, unlike traditional media, VR offers a unique opportunity to simulate the experiences of individuals living with dementia, providing a more visceral understanding of their challenges. While VR has primarily been used to educate health care students and the caregiving workforce about ADRD [21-23], its potential for raising public awareness has not been fully explored. This gap is particularly important, as VR can deliver an engaging and experiential form of learning that may influence attitudes towards ADRD.

This study aims to explore the impact of a free YouTube VR video series titled "A Walk-through Dementia" (AWTD), which simulates the experiences of individuals living with dementia [20]. AWTD consists of 3 videos, each lasting 3 to 5 minutes, and was developed by Alzheimer's Research United Kingdom to enhance understanding of the challenges faced by people with dementia. This research seeks to understand the general public's experience with VR videos on ADRD. By analyzing YouTube comments, the study will explore whether these VR videos can enhance public awareness of ADRD and serve as an effective tool for addressing misunderstandings and stigma surrounding ADRD.

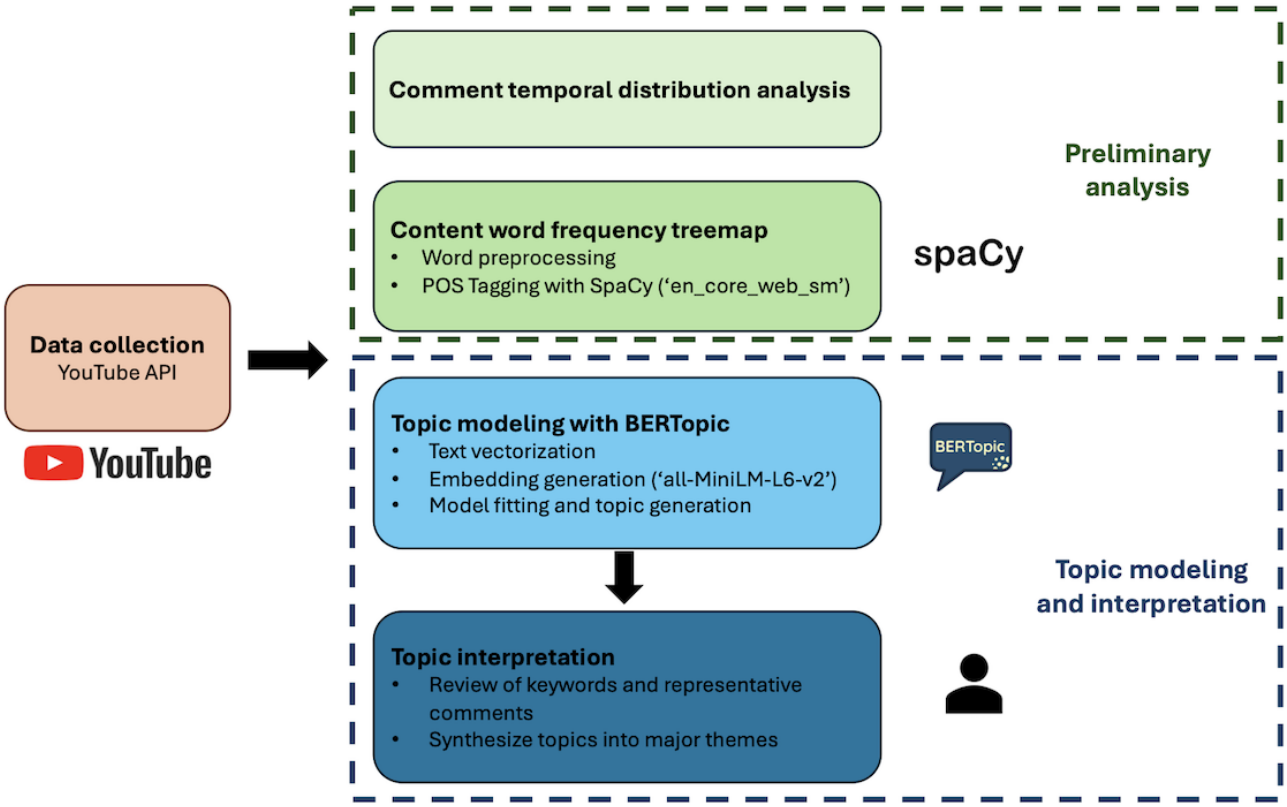
Methods

Overview

We followed a 3-step exploratory research approach to conduct the study. First, we collected comprehensive data by gathering the most viewed VR video series on YouTube. Second, we performed topic modeling in the comments section to identify the main themes discussed. Third, we used Thematic analysis to conduct these themes into concise, easily digestible summaries. The workflow of the methods is outlined in Figure 1 below.

XSL•FO
RenderX

Figure 1. The workflow of the methods. It includes data collection via the YouTube API, preliminary analysis, and topic modeling using BERTopic. The final stage involves interpreting topics by reviewing keywords and synthesizing them into major themes. API: application programming interface; BERT: Bidirectional Encoder Representations from Transformers.



Datasets

The analyzed video comments are from a series of 360° videos titled "A Walk-Through Dementia," the most viewed videos under the search terms "dementia" and "Alzheimer disease." Developed by Alzheimer's Research United Kingdom, this series aims to raise awareness about the impact of dementia on individuals' lives by engaging the public, health care professionals, and caregivers. The series comprises 3 main scenarios depicted in separate videos (Textbox 1) .

Textbox 1. Three main scenarios in "A Walk-Through Dementia."

- At the supermarket: the video demonstrates how dementia affects tasks such as reading shopping lists, dealing with short-term memory loss, changes in food preferences, and managing social anxiety at checkout counters.
- Walking home: the video highlights the difficulties in navigating busy streets, recognizing familiar places and people, and managing visual and spatial misperceptions.
- At home: the video illustrates challenges within a familiar environment, such as difficulty remembering instructions, issues with hand-eye coordination, and managing multi-step tasks like making a cup of tea.

Compared to other VR videos retrieved using the same search terms, the AWTD series has garnered the highest view counts (walking home [WH]: 3.2 million views; at the supermarket [AS]: 2.7 million views; at home [AH]: 0.12 million views) relative to the most viewed videos by other creators (eg, "Coping with Alzheimer's, Together and Apart" with 67 thousand views). In addition, this series has attracted the most comments. To maintain content consistency and consider the low number of comments in other VR videos (eg, "Coping with Alzheimer's, Together and Apart" with 66 comments), we decided to focus solely on the comment sections of the 3 videos in this series.

Preliminary Analysis

To collect the comments for this study, we used the official YouTube application programming interface, retrieving both primary comments and their corresponding replies. This resulted in a total of 12,754 comments. After removing duplicate and irrelevant entries, the final dataset included 11,543 unique comments.

Before conducting topic modeling, we performed a series of preliminary analyses to gain an initial understanding of the content across the 3 videos. These analyses involved examining the temporal distribution of comment postings and calculating word frequencies. The purpose of this step was to identify key content words essential for interpreting the overarching themes in the comments.

To ensure focus on the most meaningful words, we restricted our analysis to content words, specifically nouns, main verbs, adjectives, and most adverbs. We identified the top 300 most frequent words in the dataset and categorized them into nouns, verbs, and adjectives or adverbs using a part-of-speech tagging

XSL•FO
RenderX

model, fine-tuned for accurate tagging tasks. In addition, we manually corrected any miscategorized words during the automated tagging process. The frequencies of these words were calculated and used to create a treemap visualization, highlighting the distribution and prominence of these content words across the comments.

Topic Modeling With BERTopic

Topic modeling is widely used in health informatics and related disciplines for text mining large datasets, such as comments or tweets, to extract main topics from extensive texts. For our study, we used BERTopic (Bidirectional Encoder Representations from Transformers), a state-of-the-art topic modeling technique that leverages transformer-based deep learning models to identify topics in large text collections. Its powerful capabilities have been demonstrated in various studies [24-26]. It consistently outperforms traditional topic modeling methods, such as latent Dirichlet allocation (LDA) and nonnegative matrix factorization (NMF), with a particularly pronounced advantage in analyzing short and unstructured texts, where embedding-based models show greater potential for generating meaningful insights [27]. To validate this choice, we conducted empirical comparisons between BERTopic, LDA, and NMF using topic diversity scores, which measure the uniqueness of generated topics. It is calculated as the ratio of the number of unique words to the product of the number of top words per topic and the total number of topics [28]. BERTopic achieved a significantly higher score (0.8750) compared to LDA (0.2375) and NMF (0.25), indicating its superior ability to produce distinct and less redundant topics. These results align with previous research, which highlights BERTopic's effectiveness in capturing contextual meaning and semantic nuances often missed by traditional models [27,29]. All analyses and visualizations were conducted using Python in Colab, using the BERTopic package (version 0.16.0).

We began by preprocessing the text data and removing special characters, symbols, and punctuation. Then, we used the vectorization technique CountVectorizer to convert the preprocessed text into numerical representations. These numerical representations capture the frequency or importance of words in each document and can be directly input into topic modeling algorithms.

Subsequently, we conducted topic modeling with BERTopic. Based on a preliminary review of the comments, we estimated the number of topics to range from 5 to 10. After training models with various predefined topic numbers, we selected 8 as the final number because it provided a balance between maintaining model coherence and capturing meaningful thematic distinctions within the comments and retained only topics that included more than 50 comments. As part of the topic modeling process, we analyzed both unigrams and bigrams to generate topics. While unigrams offer a broad sense of the topics by identifying the most frequent individual terms, bigrams provide a more fine-grained perspective by highlighting word pairs that are often used together [30]. Analyzing both unigrams and bigrams helped create a more comprehensive understanding of the topics discussed.

Topic modeling is a common approach to analyzing social media content, such as Facebook (Meta) posts, tweets, and Reddit communities, related to public health. It identifies key themes, including awareness campaigns, personal caregiving experiences, and scientific advancements [31]. This approach provides a framework for understanding how structured health campaigns influence public discourse on social media [32]. Therefore, topic modeling is suitable for this study to examine comments on VR videos on YouTube, contributing to a more comprehensive view of ADRD in social media contexts.

Ethical Considerations

This study consisted of observations of public behaviors using secondary research data. The study data are anonymous and deidentified. Therefore, ethical approval was not required.

Results

Overview of the Video Comments

We extracted all comments from the 3 VR videos in the AWTD series. Both WH and AH were posted on June 2, 2016, while AS was posted the following year on March 21, 2017. As of September 2024, when the data was collected, AS had the highest number of comments, with 6572 (including replies), WH had 5954 comments, and AH had 122 comments.

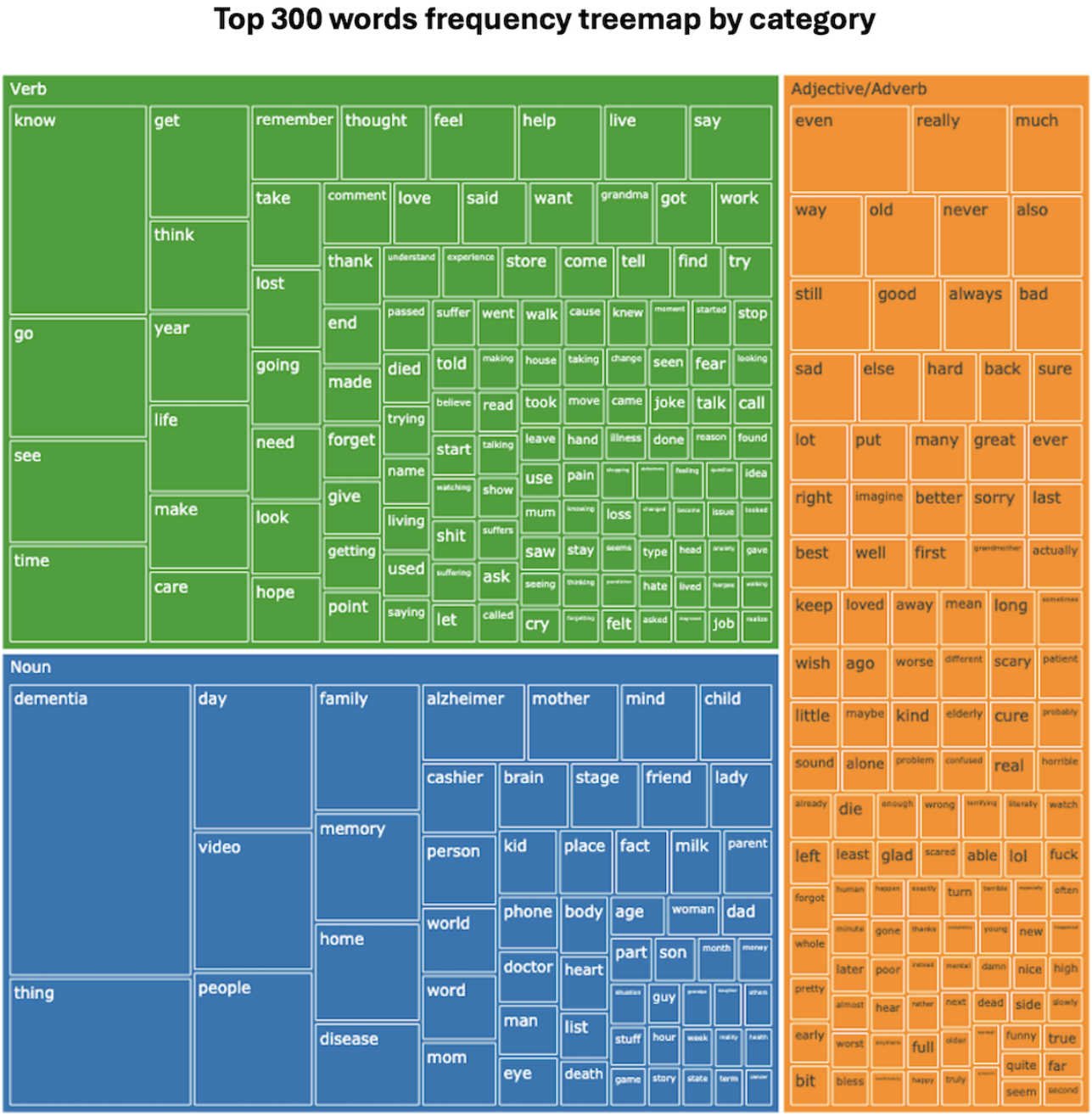
The treemap visualization illustrates the most frequently used words in the comments across the 3 videos. Each rectangle represents a word, with its size and color intensity corresponding to the word's frequency in the comments for each video. The word "dementia" is the most prominent in the noun category, indicating that it is a central topic in the discussions. Other significant nouns include "family," "home," "memory," "people," "video," "mom," and "disease." These nouns suggest that the comments often focus on personal experiences, caregiving, videos as a tool, and the impact of dementia on individuals and families.

Among the verbs, words including "know," "remember," "think," and "lost" are frequent. This shows that the comments are action-oriented, reflecting emotions, actions taken, or care experiences of the commenters. The presence of verbs like "forgot" and "care" suggests a focus on ADRD symptom and caregiving actions.

In the adjectives or adverbs category, "even," "really," "much," "old," and "still" are prominent. These words typically modify other words to express degree, quality, or timing, indicating that commenters are emphasizing the severity, persistence, or specific qualities of the experiences or events they are discussing. In addition, words like "sorry," "bad," "good," "best," "sad," and "hard" reflect emotional responses and value judgments, indicating that many comments convey personal sentiments or evaluations of the situations being described (Figure 2).

XSL•FO
RenderX

Figure 2. Top 300 words frequency treemap categorized by part-of-speech. The visualization groups the most frequently used words from YouTube comments into verbs (green), nouns (blue), and adjectives or adverbs (orange). The size of each word reflects its frequency.



Topics of the Video Comments

The topic modeling process identified 8 initial topics. The 2D dense clusters of these topics, visualized in Figure 3, depict the text embeddings. Each cluster, represented by a different color, corresponds to a unique topic derived from the dataset. Each point signifies a document, with spatial proximity indicating similarity between comments. The topic size decreases as the topic number increases, with Topic 0 being the largest, comprising 2040 comments. This topic is labeled "dementia_like_just" by BERTopic. The most representative words characterizing each topic are shown in Figure 4. For instance, the top 5 words in Topic 7 are "walls," "god," "live," "jesus," and "christ," indicating that this topic revolves around religious discussions.

XSL•FO
RenderX

Figure 3. Topic distribution and clustering of "A Walk-Through Dementia" comments. Comments are distributed into 8 distinct topics, each represented by a different color. The topics are ranked from 0 to 7 based on their size, with Topic 0 ("dementia_just_like") being the largest and Topic 7 ("walls_god_live") the smallest.

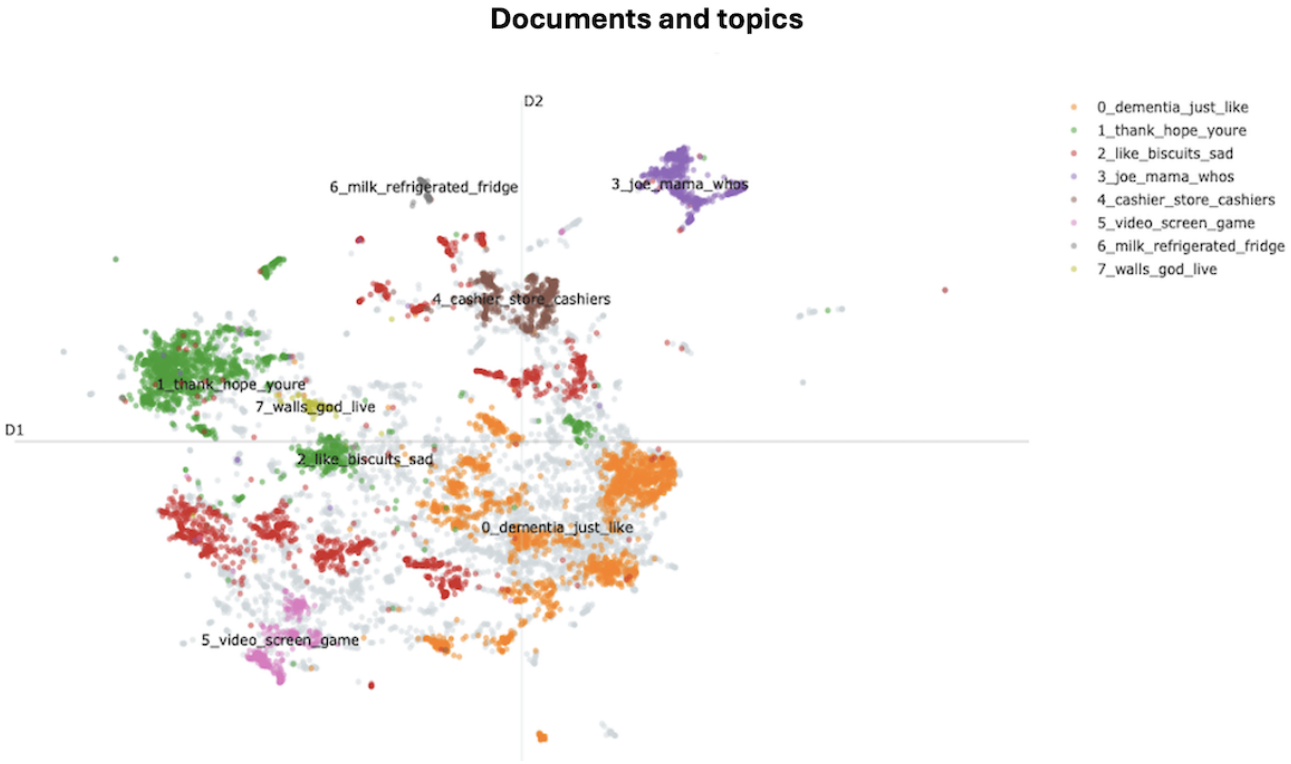
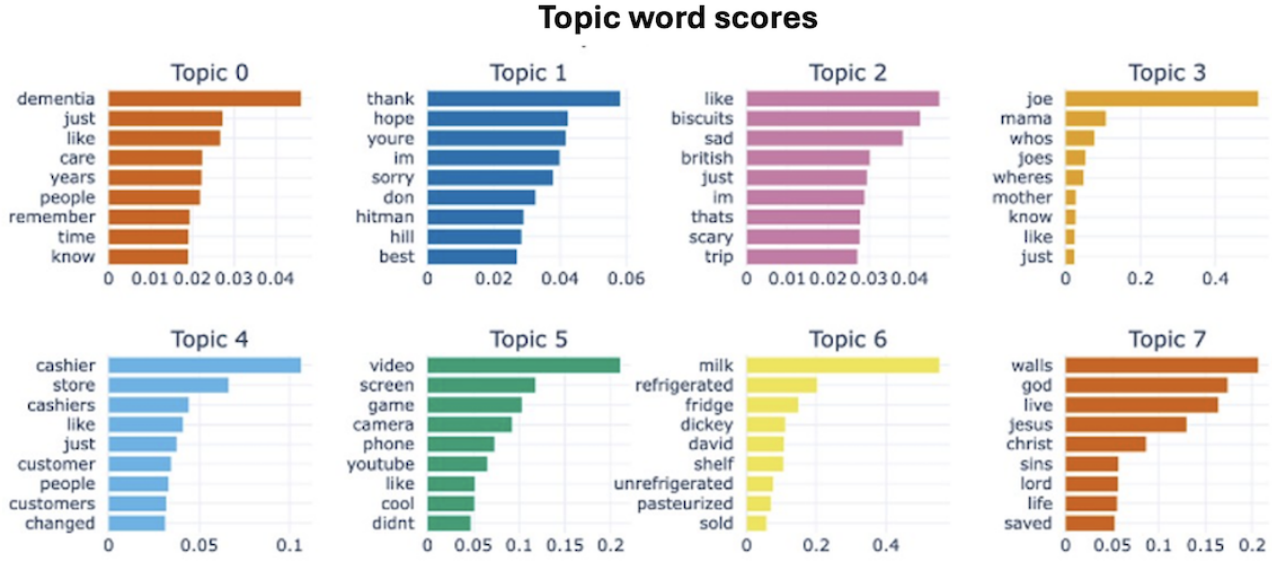


Figure 4. Top words by topic from "A Walk-Through Dementia" (AWTD) comments. It presents the most relevant words for each of the 8 identified topics in AWTD YouTube comments. The length of the bars after each word indicates the word's importance within that topic. For example, the word "dementia" is the most significant word in Topic 0.



After manually reviewing the 8 topics and content of comments, we synthesized them into five cohesive themes reflecting viewers' thoughts and needs: (1) the personal experience of ADRD care and expressions of empathy and support, (2) positive reactions to the VR videos, (3) challenges faced by individuals with ADRD, (4) caregivers' role, and (5) learning from VR videos. Notably, the majority of comments reflected a deep emotional connection to the content, underscoring the significant impact of personal experiences related to ADRD. Many viewers expressed a newfound understanding and empathy toward those living with dementia and the challenges they encounter in daily life, highlighting the educational value of the immersive experience. The reviewers also discussed the roles and responsibilities of a caregiver of a person with ADRD and what the members of the general public learned from AWTD. Table 1 illustrates our topic analysis and theme development process, listing the initial topics, keywords, representative comments, and summarized themes.

XSL•FO
RenderX

Table . Overview of topic analysis and theme development. The table presents 8 topics identified by BERTopic, along with associated keywords and representative comments.

Topic ID	Keywords	Representative comment	Summarized theme
0	dementia, just, like, care, years, people, home, time, family, know	• "My mom died of Alzheimer's. It is hard when someone you love forgets where they live, who you are, what's going on." • "That would break my heart. I guess some people get used to it and live with this sickness of one of their family members, but for an outsider, like me, reading your comment was just a tough read."	Personal experiences of ADRD[a] care, empathy and support
1	thank, your, I'm, hope, sorry, don, like, don't	• "I hope you are doing well, your Strength and ability to push on every day is so admirable and inspiring. Thank you for being such an amazing human being…" • "I am a member of a search team, and we are called upon to search urban areas, forests, swamps and such all the time for elderly people with dementia who 'wandered off'"	Personal experiences of ADRD care, empathy and support
2	biscuits, like, sad, British, trip, tea, I'm, scary, sugar, that's	• "Like a bad trip you have no control over" • "I love the subtle details, like how there was an orange car parked by her house, but not the same type she saw earlier." • "When a person mistakes a puddle for a hole in the ground, don't just accept that as normal behavior, look into it"	Challenges faced by individuals with ADRD
3	joe, mama, who's, son, joes, where's, just, like, voice, know	• "who's Joe?" "Old lady's son" "where's Joe?" • "For me, the most terrifying part of the whole video was when Joe closed the door, leaving his mom all alone. She's now isolated in a big home with tall stairs and couldn`t easily walk out." • "Children, you should make a cup of tea for your mum" • "Dementia is often overlooked and extremely underestimated."	Caregivers' role

XSL•FO
RenderX

Topic ID	Keywords	Representative comment	Summarized theme
4	cashier, store, cashiers, like, just, customer, people, customers, changed, rude	• "The cashier clearly has no understanding. She is rushing the poor women to pay. So rude. It's super devastating to see this because we know they can't stand up for themselves because of how much they struggle." • "When we put ourselves in someone's shoes, we learn so much about others and conditions we never really paid much thought to" • "I DEFINITELY need to be a more patient person with my aging mother! Thanks for posting this." • "When I was working in the nursing home, I always loved to just hold their hand, now I understand why they never refuse."	Learning from A Walk-Through Dementia
5	video, screen, camera, game, phone, YouTube, cool, didn't, like, realize	• "This is a great 360 video that puts us right in the middle of someone's illness" • "What a powerful video. Did anyone realize that if you move your phone, it pans the camera? Very cool!" • "This is so helpful, thank you! People without dementia are challenged to empathize because they don't have the insight this video provides." • "This is so much more informative than any government add or script" • "I could feel the anxiety realizing she was home alone. What a powerful message." • "I sit here reflecting on this incredible powerful and educational tool. I (My Eyes, My Heart, My Soul, My Feelings, My Spirit, My Brain) have just experience." • "I find the 360 aspect is really cool, the video is of great help to the general public in understanding the life difficulties of dementia cases."	Positive reactions to the VR[b] videos
6	milk, refrigerated, fridge, dickey, David, shelf, unrefrigerated, pasteurized, sold, isle		Challenges faced by individuals with ADRD

XSL•FO
RenderX

Topic ID	Keywords	Representative comment	Summarized theme
		• "Milk in all, 2spoon, 1s, 1/2s.; the milk wasn't refrigerated; Why is the milk not in the fridge" • "I would need to write it down for sure" • "Guests can serve themselves. Or else make a list for the mother, and tape it on a cupboard door for the mother to make it easier" • "They should really get her a place with no stairs, easy access to beds and rooms, cushions and pads in case she slips. Music and things to keep her distracted."	
7	god, Jesus, Christ, sins, life, saved, lord, John, repent, loves	• I'm taking care of my demented grandmother… alone, 24/7. Jesus Christ it is difficult, I don't know how long I can keep on doing this" • "God bless them. I take care of dementia patients all the time at the hospital, but this sheds new light."	Personal experiences of ADRD care, empathy, and support

[a]ADRD: Alzheimer disease and related dementias.

[b]VR: virtual reality.

Discussion

Principal Findings

The study evaluated AWTD, a series of free YouTube VR videos designed to raise public awareness of ADRD. At the time of data collection, AWTD was the most viewed video for the search terms "dementia" and "Alzheimer's disease" on YouTube, boasting over 6 million views. This series also garnered the highest number of comments, totaling 12,648, compared to a similar VR video, "Coping with Alzheimer's: Together and Apart," which had 67,000 views and only 66 comments. The comments from viewers predominantly came from individuals with a personal connection to someone with ADRD, professionals working with affected individuals, or general audience members. This suggests that the YouTube VR videos were particularly effective in disseminating information about ADRD to a diverse and broader audience.

Our analysis identified five key themes from the comments: (1) positive reactions to the VR videos, (2) personal experience of ADRD along with expressions of empathy and support, (3) the role of caregivers, (4) challenges faced by individuals with ADRD, and (5) learning from AWTD. These findings suggest that VR has significant potential to raise public awareness of ADRD and deepen understanding of the condition, particularly through immersive and emotionally engaging experiences.

Positive Reactions to the VR Experience and Immersive Learning

The immersive nature of VR proved highly effective in engaging viewers, allowing them to experience the challenges of ADRD firsthand, significantly enhancing their understanding of the condition. Many commenters described it as a convincing and immersive experience, both insightful and evocative, providing a powerful educational tool for ADRD care. Its appeal stemmed in part from its novelty in the field of dementia education, particularly for YouTube users, including caregivers, family members, and professionals with extensive experience in dementia care. Commenters noted that they had never encountered anything like this in their personal or professional lives. This is unsurprising, as academic literature recognizes VR as an emerging area of interest in dementia education [23,33], highlighting the need for further exploration of its applications and potential impact.

Our findings suggest that the success of the VR video lies in its ability to immerse viewers in the perspective of someone living with dementia. By wearing the VR headset, users are fully absorbed in the experience, allowing them to focus entirely on each scenario and draw connections to their personal lives or work experiences. Commenters noted how the VR video placed them directly in the midst of the illness, enabling them to feel the anxiety of a person with ADRD who is home alone. One commenter remarked, "My eyes, my heart, my soul, my feelings, my spirit, my brain have just experienced this," emphasizing how immersion creates a profound sense of presence, making

XSL•FO
RenderX

viewers feel as if they were in the shoes of someone with ADRD, inspiring empathy and support.

These findings are consistent with research from other scholars who emphasize the benefits of VR dementia education for medical students and professionals [34-36], as well as broader VR literature [37], which highlights the critical role of immersion and presence in virtual scenarios. In addition, AWTD stands out as a free VR video available on YouTube, unlike most VR content, which is typically developed and hosted on developers' websites, requiring purchase or offering limited access [38,39]. By being widely accessible, AWTD effectively disseminates information and knowledge about ADRD to the general public, offering significant benefits for public education on the condition. This aligns with Lupton's findings, which state that social media platforms like YouTube provide opportunities for researchers to step beyond traditional academic approaches and use innovative, free methods to share knowledge with new and diverse audiences [40].

Role of Caregivers and Observed Gaps

The VR experience prompted viewers to reflect on the crucial role of caregivers, with many commenters noting gaps in the videos, which demonstrates how VR can stimulate critical thinking and awareness of caregiving responsibilities in ADRD. Some commenters discussed the noticeable absence of caregivers in the 3 VR videos. They expressed concern that caregivers, particularly family members, should have been more present and attentive in the depicted scenarios. Comments like "Who is Joe?" (son of an individual with ADRD), "Where is Joe?", and "The daughter should make tea for her mother (individual with ADRD)," reflected the viewers' belief that caregivers were missing in moments when they were needed most. In these scenarios, family members, who play a crucial role in caregiving, appeared to overlook the early signs of ADRD, such as confusion, memory lapses, and difficulty completing daily tasks. By neglecting these key symptoms, caregivers failed to provide necessary support, which may have contributed to negative outcomes for individuals with ADRD, including diminished psychosocial well-being, lowered self-esteem, and a reduced overall quality of life [41]. The VR videos can help family caregivers become more aware of their roles and responsibilities in caring for individuals with ADRD, potentially encouraging further learning and training on dementia care. This heightened awareness can lead to more proactive approaches to caregiving, allowing family members to better recognize and respond to the needs of their loved ones, ultimately improving the quality of care provided to individuals living with ADRD.

Personal Connections and Empathy

By simulating the experiences of individuals with ADRD, VR video fostered a deeper emotional connection, encouraging viewers to express empathy and reflect on how the condition affects both patients and their families. Scenarios depicted individuals struggling to live with ADRD, allowing viewers to face challenges that they may have been previously unaware of, which were vividly portrayed through the VR experience. Commenters specifically noted moments like navigating around a puddle, which the person with ADRD believes to be a hole in the ground (street scenario), highlighting their disorientation. Other scenes—such as struggling to retrieve items from shelves and feeling rushed by a cashier (supermarket scenario) or becoming confused when asked to make tea for her daughter's guests (AH scenario)—underscored the daily difficulties faced by individuals with ADRD. Experiencing these challenges from the perspective of someone with ADRD seemed to provide viewers with a deeper insight and better understanding of the condition's impact.

These experiences prompted viewers to recognize how negative attitudes, impatience, or misunderstanding from others exacerbate the challenges faced by individuals with ADRD. Building a dementia-friendly or ADRD-aware community is essential to support people with ASRD to live well and carry out their daily activities. One potential approach to achieving this is to incorporate the VR video into a broader public training program, where the scenarios are reflected upon and discussed by the learners. As some commenters stated:

> *The cashier clearly has no understanding of ADRD. It's super devastating to see this because we know they can't stand up for themselves because of how much they struggle.*
>
> *When we put ourselves in someone's shoes, we learn so much about others and conditions we never really paid much thought to.*
>
> *I definitely need to be a more patient person with my aging mother! Thanks for posting this.*
>
> *When I was working in the nursing home, I always loved to just hold their hand, now I understand why they never refuse.*

VR video is a powerful way to raise awareness of ADRD, helping the public shift their perception of the condition and learn essential skills for caring for individuals with ADRD. Similar findings regarding VR as a learning tool have been noted for medical students and professional caregivers [42,43]. The VR video therefore has a fundamental role to play in helping to create equal, inclusive communities for all, improving the actions people undertake and the language they use when interacting with those with ADRD [21].

The research findings underscore the importance of immersive learning in fostering empathy and understanding for individuals with ADRD. To build on this success, there is a clear need for more VR tools that can be used in conjunction with follow-up discussions featuring educational takeaway messages tailored for diverse audiences with varying levels of knowledge and experience regarding ADRD [44]. Creating VR programs that cater to different linguistic and cultural contexts could also address gaps in accessibility. The existing scenarios might oversimplify the condition and fail to capture the wide variability in ADRD experiences. A more diverse set of scenarios covering different stages and severities of the condition would likely provide a more nuanced understanding. The insights gained from this research can guide these discussions and aid in the development of the format and learning materials for a future VR educational program based on AWTD.

XSL•FO
RenderX

Social Media Influence and VR in Health Communication

To contextualize the impact of VR as an educational tool, we ground our study within social media influence theory [45], which posits that social media platforms and their users shape public discourse, influence attitudes, and drive behavioral change through content creation, sharing, and interaction. Social media platforms serve as powerful conduits for disseminating health-related content, leveraging both algorithmic curation and social interactions to amplify messages. VR, as an immersive medium, engages user through both central (cognitive processing of health messages) and peripheral (emotional and experiential engagement) routes. This dual-pathway influence enhances message retention and can drive behavioral change, making VR an effective medium for public health communication.

The interactive nature of social media ensures that VR content does not exist in isolation but is actively discussed, shared, and reframed within online discourse. Algorithms prioritize highly engaging and emotionally resonant content, increasing its visibility and potential impact. User-generated discussions, as seen in the YouTube comments analyzed in this study, serve as mechanisms through which social media influence is both reinforced and coconstructed. When users share personal stories, express emotions, or react to the VR experience, they contribute to a collective conversation that reinforces common beliefs and introduces new perspectives on ADRD. This aligns with Social Media Influence theory's assertion that online interactions shape public discourse and personal identities, ultimately contributing to the normalization or transformation of health-related attitudes and behaviors.

These theoretical perspectives allow us to interpret the thematic findings within a broader framework of digital intervention effectiveness and audience engagement. By integrating topic modeling and BERT-based analysis with media influence theory, our study not only examines the immediate impact of VR-based health communication but also contributes to a deeper understanding of how digital interventions shape public awareness and health behaviors. The methodological framework established here can be adapted and extended to different populations and varying digital health content, enabling broader applications of VR in public health communication and education.

Limitations

One limitation of this study is the absence of demographic information about the commenters, particularly regarding their ethnicity, socioeconomic status, and professional roles. This information could provide insights into how social factors affect the experience of VR videos. In addition, as all VR video content was delivered in English, non-English-speaking populations may have been excluded, limiting the representativeness of the research sample in relation to the general public. Furthermore, the comments analyzed in this study constitute only a small portion of the total posts made in the forum, potentially failing to capture the full range of discussions surrounding ADRD. Further research could focus on gathering a more comprehensive dataset across platforms and including different languages to capture a broader perspective. Finally, this study focused exclusively on one social media platform, YouTube, and one VR video. Future research could broaden its scope by incorporating data from multiple social media platforms and a wider array of VR videos to enhance our understanding of the topic.

Conclusion

This study used topic modeling to highlight the potential of VR videos as a powerful and immersive educational tool, offering the public a more profound understanding of the experiences of individuals living with ADRD. By fostering emotional and cognitive engagement, this approach has the potential to cultivate greater empathy and community support for those affected by ADRD. Beyond its immediate findings, this research underscores the transformative role of VR in reshaping how complex health conditions are communicated and understood. Future studies should expand on this foundation by exploring the broader applications of VR on social media platforms, such as Dementia VR Film, Beatriz Lab VR Module, and virtual reality scenarios. These tools not only have the capacity to reach wider audiences but also to drive meaningful societal change by enhancing awareness, reducing stigma, and promoting inclusive support systems for individuals with ADRD and their caregivers.

Acknowledgments

We would like to express our gratitude to the creators and uploaders of the publicly available content on YouTube, which served as the foundation for this secondary data analysis. Their contributions enabled us to conduct this research and explore the potential of virtual reality videos as an educational tool for Alzheimer disease and related dementias.

Data Availability

All data generated or analyzed during this study are included in this published article.

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. Dementia. World Health Organization. 2023 Mar 15. URL: https://www.who.int/news-room/fact-sheets/detail/dementia [accessed 2024-05-15]
2. Shu S, Woo BK. Use of technology and social media in dementia care: current and future directions. World J Psychiatry 2021 Apr 19;11(4):109-123. [doi: 10.5498/wjp.v11.i4.109] [Medline: 33889536]
3. Prince M, Wimo A, Guerchet M, Ali GC, Wu YT, Prina M. The global impact of dementia: an analysis of prevalence, incidence, cost and trends. : Alzheimer's Disease International; 2015 URL: https://www.alzint.org/u/WorldAlzheimerReport2015.pdf [accessed 2024-05-15]
4. Dementia friendly communities: global developments. Alzheimer's Disease International. 2017. URL: https://www.alzint.org/u/dfc-developments.pdf [accessed 2024-05-15]
5. Green G, Lakey L. Building dementia-friendly communities: a priority for everyone. Act on Alzheimer's. 2013. URL: https://actonalz.org/sites/default/files/2023-01/Dementia_friendly_communities_full_report.pdf [accessed 2024-05-15]
6. Milne A. The "D" word: reflections on the relationship between stigma, discrimination and dementia. J Ment Health 2010 Jun;19(3):227-233. [doi: 10.3109/09638231003728166] [Medline: 20441486]
7. Mukadam N, Livingston G. Reducing the stigma associated with dementia: approaches and goals. Aging health 2012 Aug;8(4):377-386. [doi: 10.2217/ahe.12.42]
8. Cheng ST, Lam LCW, Chan LCK, et al. The effects of exposure to scenarios about dementia on stigma and attitudes toward dementia care in a Chinese community. Int Psychogeriatr 2011 Nov;23(9):1433-1441. [doi: 10.1017/S1041610211000834] [Medline: 21729424]
9. Cheston R, Hancock J, White P. A cross-sectional investigation of public attitudes toward dementia in Bristol and South Gloucestershire using the approaches to dementia questionnaire. Int Psychogeriatr 2016 Oct;28(10):1717-1724. [doi: 10.1017/S1041610216000843] [Medline: 27353013]
10. Chen J, Wang Y. Social media use for health purposes: systematic review. J Med Internet Res 2021 May 12;23(5):e17917. [doi: 10.2196/17917] [Medline: 33978589]
11. Auxier B, Anderson M. Social media use in 2021. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2024-05-15]
12. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85. [doi: 10.2196/jmir.1933] [Medline: 23615206]
13. Wu S, Foster K, Ramirez M, Jin H, Joe E. How well do videos on YouTube meet information needs of ADRD family caregivers? A qualitative analysis. Innovation in Aging 2020 Dec 16;4(Supplement_1):273-274. [doi: 10.1093/geroni/igaa057.875]
14. Teano AL, Scott A, Gipson C, Albert M, Pettigrew C. Social media programs for outreach and recruitment supporting aging and alzheimer disease and related dementias research: longitudinal descriptive study. JMIR Aging 2024 Jul 9;7:e51520. [doi: 10.2196/51520] [Medline: 38981112]
15. Dassel K, Butler J, Telonidis J, Edelman L. Development and evaluation of alzheimer's disease and related dementias (ADRD) best care practices in long-term care online training program. Educ Gerontol 2020 Mar 3;46(3):150-157. [doi: 10.1080/03601277.2020.1717079]
16. Lam NHT, Woo BKP. YouTube as a new medium for dementia education among Chinese Americans. Community Ment Health J 2020 Apr;56(3):435-439. [doi: 10.1007/s10597-019-00493-7] [Medline: 31641910]
17. Heward M, Palfreman-Kay J, Innes A. In their words: how television and visual media can raise awareness of dementia and other health conditions that carry stigma, including disabilities. Journal of Popular Television, The 2015 Oct 1;3(2):229-242. [doi: 10.1386/jptv.3.2.229_1]
18. Castillo LIR, Hadjistavropoulos T, Brachaniec M. The effectiveness of social media in the dissemination of knowledge about pain in dementia. Pain Med 2021 Nov 26;22(11):2584-2596. [doi: 10.1093/pm/pnab157] [Medline: 34009395]
19. Jiang H, Vimalesvaran S, Wang JK, Lim KB, Mogali SR, Car LT. Virtual reality in medical students' education: scoping review. JMIR Med Educ 2022;8(1):e34860. [doi: 10.2196/34860]
20. Kamińska D, Sapiński T, Wiak S, et al. Virtual reality and its applications in education: survey. Information 2019;10(10):318. [doi: 10.3390/info10100318]
21. Hicks B, Konovalova I, Myers K, Falconer L, Board M. Taking 'A walk through dementia': exploring care home practitioners' experiences of using a virtual reality tool to support dementia awareness. Ageing Soc 2023 May;43(5):1042-1067. [doi: 10.1017/S0144686X21000994]
22. A walk through dementia. Alzheimer's Research UK. 2016. URL: https://www.alzheimersresearchuk.org/campaigns/awtd/ [accessed 2024-05-15]
23. Li X, Liu X, Collins S, Mutthina R, Chiu C. The virtual reality intervention programs and dementia care education: a systematic review. Educ Gerontol 2025:1-17. [doi: 10.1080/03601277.2025.2466131]
24. Lee H, Hong Lee S, Re Lee K, Hyun Kim J. ESG discourse analysis through BERTopic: comparing news articles and academic papers. Computers, Materials & Continua 2023;75(3):6023-6037. [doi: 10.32604/cmc.2023.039104]
25. Wang Z, Chen J, Chen J, Chen H. Identifying interdisciplinary topics and their evolution based on BERTopic. Scientometrics 2024 Nov;129(11):7359-7384. [doi: 10.1007/s11192-023-04776-5]

XSL•FO
RenderX

26. Hristova G, Netov N. Media coverage and public perception of distance learning during the COVID-19 pandemic: a topic modeling approach based on bertopic. Presented at: 2022 IEEE International Conference on Big Data (Big Data); Dec 17-20, 2022; Osaka, Japan. [doi: 10.1109/BigData55660.2022.10020466]
27. Egger R, Yu J. A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify twitter posts. Front Sociol 2022;7:886498. [doi: 10.3389/fsoc.2022.886498] [Medline: 35602001]
28. Asnawi MH, Pravitasari AA, Herawan T, Hendrawati T. The combination of contextualized topic model and MPNet for user feedback topic modeling. IEEE Access 2023;11:130272-130286. [doi: 10.1109/ACCESS.2023.3332644]
29. Chong M, Chen H. Racist framing through stigmatized naming: a topical and geo‐locational analysis of #Chinavirus and #Chinesevirus on Twitter. Proceedings of the Association for Information Science and Technology 2021 Oct;58(1):70-79. [doi: 10.1002/pra2.437]
30. Nokel M, Loukachevitch N. A method of accounting bigrams in topic models. In: Proceedings of the 11th Workshop on Multiword Expressions: Association for Computational Linguistics; 2015:1-9. [doi: 10.3115/v1/W15-0901]
31. Paul MJ, Dredze M. Discovering health topics in social media using topic models. PLoS ONE 2014;9(8):e103408. [doi: 10.1371/journal.pone.0103408] [Medline: 25084530]
32. Kong D, Chen A, Zhang J, et al. Public discourse and sentiment toward dementia on Chinese social media: machine learning analysis of weibo posts. J Med Internet Res 2022 Sep 2;24(9):e39805. [doi: 10.2196/39805] [Medline: 36053565]
33. Hirt J, Beer T. Use and impact of virtual reality simulation in dementia care education: a scoping review. Nurse Educ Today 2020 Jan;84:104207. [doi: 10.1016/j.nedt.2019.104207]
34. Adefila A, Graham S, Clouder L, Bluteau P, Ball S. myShoes – the future of experiential dementia training? The Journal of Mental Health Training, Education and Practice 2016 May 9;11(2):91-101. [doi: 10.1108/JMHTEP-10-2015-0048]
35. Morganti F, Palena N, Savoldelli P, Greco A. Take the first-person perspective to become dementia-friendly: the use of 360° video for experiencing everyday-life challenges with cognitive decline. Front Psychol 2020;11:1117. [doi: 10.3389/fpsyg.2020.01117]
36. Wijma EM, Veerbeek MA, Prins M, Pot AM, Willemse BM. A virtual reality intervention to improve the understanding and empathy for people with dementia in informal caregivers: results of a pilot study. Aging Ment Health 2018 Sep;22(9):1115-1123. [doi: 10.1080/13607863.2017.1348470] [Medline: 28691861]
37. Falconer L. Situated learning in virtual simulations: researching the authentic dimension in virtual worlds. J Interact Learn Res 2013;24(3):285-300.
38. Bard JT, Chung HK, Shaia JK, Wellman LL, Elzie CA. Increased medical student understanding of dementia through virtual embodiment. Gerontol Geriatr Educ 2023;44(2):211-222. [doi: 10.1080/02701960.2022.2067850] [Medline: 35451921]
39. Sung HC, Su HF, Lee WL, Yamakawa M, Wang HM. Effects of a dementia virtual reality-based training with peer support for home care workers: a cluster randomized controlled trial. Int J Geriatr Psychiatry 2022 Sep;37(9). [doi: 10.1002/gps.5799] [Medline: 35996760]
40. Lupton D. 'Better understanding about what's going on': young Australians' use of digital technologies for health and fitness. Sport Educ Soc 2020 Jan 2;25(1):1-13. [doi: 10.1080/13573322.2018.1555661]
41. Li X, Dorstyn D, Mpofu E, et al. Nursing assistants and resident satisfaction in long-term care: a systematic review. Geriatr Nurs (Lond) 2021 Nov;42(6):1323-1331. [doi: 10.1016/j.gerinurse.2021.08.006]
42. Yamaguchi Y, Ryuno H, Fukuda A, et al. Effects of a virtual reality intervention on dementia care education among acute care nurses in Japan: a non-randomised controlled trial. Geriatr Nurs 2022;48:269-273. [doi: 10.1016/j.gerinurse.2022.10.013] [Medline: 36332442]
43. Stargatt J, Bhar S, Petrovich T, Bhowmik J, Sykes D, Burns K. The effects of virtual reality-based education on empathy and understanding of the physical environment for dementia care workers in Australia: a controlled study. J Alzheimers Dis 2021;84(3):1247-1257. [doi: 10.3233/JAD-210723] [Medline: 34633323]
44. Board M, Phipps L, Mitchell R, Murphy J. Evaluating the impact of the virtual reality app 'A walk through dementia' on year one health care students clinical practice. Presented at: The Alzheimer's Show; Jun 7-8, 2019; Olympia, London URL: https://eprints.bournemouth.ac.uk/33667/ [accessed 2024-05-15]
45. Rachmad Y. Social Media Influence Theory: Phoenix Cactus Book Publishing; 2023.

Abbreviations

ADRD: Alzheimer disease and related dementias
AH: at home
AS: at the supermarket
AWTD: A Walk-Through Dementia
BERT: Bidirectional Encoder Representations from Transformers
LDA: latent Dirichlet allocation
NMF: nonnegative matrix factorization
VR: virtual reality

XSL•FO
RenderX

WH: walking home

Edited by A Mavragani; submitted 20.10.24; peer-reviewed by Y Sutherland, Z Wang; revised version received 17.02.25; accepted 18.02.25; published 16.04.25.

Please cite as:
Li X, Liu X, Yin C, Collins S, Alanazi E
Impact of a Virtual Reality Video ("A Walk-Through Dementia") on YouTube Users: Topic Modeling Analysis
JMIR Form Res 2025;9:e67755
URL: https://formative.jmir.org/2025/1/e67755
doi:10.2196/67755

© Xiaoli Li, Xiaoyu Liu, Cheng Yin, Sandra Collins, Eman Alanazi. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Feasibility, Acceptability, and Effectiveness of a Smartphone App to Increase Pretransplant Vaccine Rates: Usability Study

Amy G Feldman[1,2,3], MD, PhD; Brenda L Beaty[1,3], MSPH; Susan L Moore[1,3,4], MSPH, PhD; Sheana Bull[4], MPH, PhD; Kumanan Wilson[5,6,7], MSc, MD, FRCP; Katherine M Atkinson[8], PhD; Cameron Bell[7], BEng; Kathryn M Denize[7], MSc; Allison Kempe[1,3,9], MPH, MD

[1]Section of Pediatric Gastroenterology, Hepatology, and Nutrition, University of Colorado, 13123 E 16th Ave, Aurora, CO, United States

[2]Digestive Health Institute, Children's Hospital Colorado, Aurora, CO, United States

[3]Adult and Child Center for Outcomes Research and Delivery Science, Aurora, CO, United States

[4]Colorado School of Public Health, Aurora, CO, United States

[5]Department of Medicine, University of Ottawa, Ottawa, ON, Canada

[6]Bruyère Health Research Institute, Ottawa, ON, Canada

[7]CANImmunize Inc, Ottawa, ON, Canada

[8]Department of Global Public Health, Karolinska Institutet, Stockholm, Sweden

[9]Children's Hospital Colorado, Aurora, CO, United States

Corresponding Author:
Amy G Feldman, MD, PhD
Section of Pediatric Gastroenterology, Hepatology, and Nutrition, University of Colorado, 13123 E 16th Ave, Aurora, CO, United States

Abstract

Background: Vaccine-preventable infections result in significant morbidity, mortality, and costs in pediatric transplant recipients. Despite intensive medical care in the pretransplant period, less than 20% of children are up to date for age-appropriate vaccines at the time of transplant. Mobile health apps have the potential to improve pretransplant vaccine rates.

Objective: This paper aimed to perform phase 2 beta testing of the smartphone app, Immunize PediatricTransplant, to determine (1) if it was effective in achieving up-to-date vaccine status by the time of transplant in a cohort of children awaiting transplants and (2) if the app was feasible and acceptable to parent and transplant provider users.

Methods: We recruited 25 dyads of parents and providers of a child awaiting a liver, kidney, or heart transplant at Children's Hospital Colorado, Ann and Robert H. Lurie Children's Hospital, and the Children's Hospital of Philadelphia. Parents and providers filled out an entry questionnaire before app use to gather baseline information. A research team member entered the child's vaccine records into the app. The parent and provider downloaded and used the app until the transplant to view vaccine records, read vaccine education, communicate with team members, and receive overdue vaccine reminders. After the transplant (or on April 1, 2024, the conclusion of the study), the parent and provider filled out an exit questionnaire to explore feasibility and acceptability of the app. The child's vaccine records were reviewed to determine if the child was up to date on vaccines at the time of transplant.

Results: Twenty-five parent and provider dyads were enrolled; 56% (14/25) had a child awaiting a liver transplant, 28% (7/25) had a child awaiting a kidney transplant, and 16% (4/25) had a child awaiting a heart transplant. At the conclusion of the study, 96% (24/25) of the children were up to date on vaccines. Of the 36 parents and providers who filled out an exit questionnaire, 97% (n=35) agreed or strongly agreed that they felt knowledgeable about pretransplant vaccine use and 86% (n=31) agreed or strongly agreed that communication around vaccines was good after using the app. Further, 91% (20/22) of parents and 79% (11/14) of providers recommended the app to future parents and providers of transplant candidates. Parents and providers suggested that in the future the app should connect directly to the electronic medical record or state vaccine registries to obtain vaccine data.

Conclusions: The overwhelming majority of children whose parents and providers used the Immunize PediatricTransplant app were up to date on vaccines at the time of transplant. The majority of app users felt the app was feasible and acceptable. In future iterations of the app and subsequent clinical trials, we will explore whether application programming interfaces might be used to extract vaccine data from the electronic medical record. If implemented broadly, this app has the potential to improve pretransplant vaccine rates, resulting in fewer posttransplant infections and improved posttransplant outcomes.

(JMIR Form Res 2025;9:e68855) doi:10.2196/68855

XSL•FO
RenderX

KEYWORDS

mobile app; vaccines; immunizations; transplantation; children; beta test; mobile health; mHealth

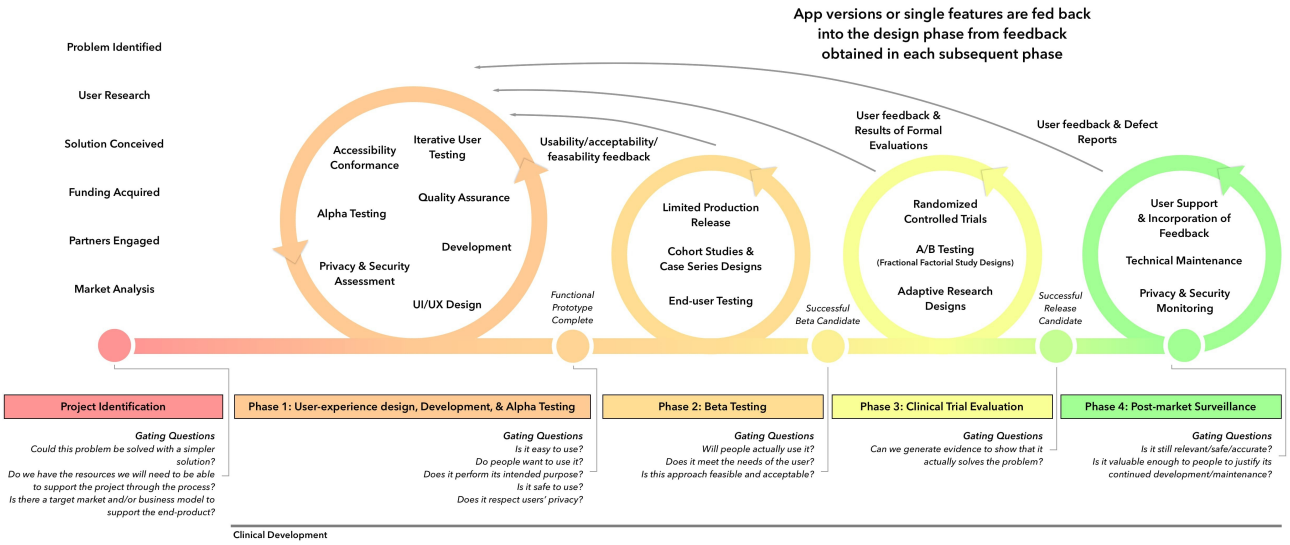
Introduction

Transplant recipients are on life-long immunosuppressive medications to prevent graft rejection. As a result of these medications, they are at increased risk for life-threatening infections including vaccine-preventable infections (VPIs) [1-3]. In the first 5 years posttransplant, VPIs occur in pediatric solid organ transplant recipients at rates up to 87 times higher than in the general pediatric population [4,5]. VPIs result in significant morbidity, mortality, and increased hospitalization costs [4,5]. Pretransplant vaccines are an important strategy to decrease the risk of VPIs posttransplant. Unfortunately, despite intensive medical management in the pretransplant period, less than 20% of pediatric liver transplant recipients have received all age-appropriate vaccines by the time of transplant [6]. Barriers to pretransplant vaccination include gaps in knowledge about pretransplant vaccines, lack of communication between team members regarding vaccines, difficulty remembering when vaccines are due amidst other acute medical problems, and lack of an easily accessible centralized vaccine record [7].

Mobile health (mHealth) apps can assist with medical education, team communication, and patient adherence [8-12]. Specifically, mHealth apps have demonstrated utility in facilitating vaccine delivery in the general population [13-20]. In a previous study, our team developed a novel mHealth app, Immunize PediatricTransplant, to help improve vaccine rates in the pretransplant period [21]. Specifically, the content of the app (1) provides information about vaccines in the pretransplant period; (2) houses a cloud-based, easily accessible vaccine record; (3) includes a chat and communication feature to facilitate communication between parents and transplant providers about vaccines; and (4) provides text or email reminders for parents and providers when vaccines are overdue [21].

According to the mHealth Agile Development and Evaluation Lifecycle (Figure 1) [22], after a functional mobile tool is developed, phase 2 beta testing is needed to test the mHealth product amongst external users to understand if the prototype meets the needs of users in a way that will lead to sustained adoption. Therefore, we designed this study to (1) evaluate the effectiveness of the Immunize PediatricTransplant app in achieving up-to-date vaccine status by the day of transplant (or conclusion of the study on April 1, 2024) and (2) explore app feasibility and acceptability amongst parent and provider users.

Figure 1. mHealth Agile Development and Evaluation Lifecycle (adapted from Wilson et al [22]). mHealth: mobile health.



Methods

Study Design, Participants, and Recruitment

This prospective cohort study was conducted with enrollment between November 1, 2021, and July 31, 2023. Dyads of parents and transplant providers caring for a child awaiting a heart, liver, or kidney transplant were recruited from the Children's Hospital Colorado, Ann and Robert H. Lurie Children's Hospital, and the Children's Hospital of Philadelphia via an in-clinic, email, or electronic medical record (EMR) invitation. Parents of a child being transplanted for acute organ failure or retransplantation were excluded. Parents who refused vaccines for religious or ethical reasons, who were not fluent in English, or who did not have access to a mobile phone were excluded. Certain transplant providers had more than 1 patient enrolled in the study and used the app for all enrolled patients. All participants received a US $100 gift card for participation.

The primary outcome of this study was up-to-date vaccine status of the child on the day of transplant (or end of the study) according to standard Centers for Disease Control (CDC) vaccine recommendations [23]. Secondary outcomes included feasibility and acceptability of the app amongst parent and provider app users.

XSL•FO
RenderX

Study Procedures and Data Collection

Getting to Know You Entry Questionnaire

Each parent and transplant provider were given a baseline survey to gather information about them.

Immunize PediatricTransplant App

After the Getting to Know You questionnaire was completed, the parent and provider both downloaded the app. A research team member entered the child's vaccine records (using state registries, the EMR, and physician and parental records) to initiate the app to send text and email reminders to the parents and providers when a vaccine was due. Parents and providers were instructed to log into the app as often as they wished to view their child's vaccine records, read information about vaccines, or communicate with team members about vaccines. Vaccine reminders were sent to parents and transplant providers (via text, email, and app notification) when vaccines were due starting at enrollment and going until the time of transplant or end of the study.

Exit Questionnaire

After the transplant, the parent and provider filled out a survey with questions assessing app feasibility (How many months did you use the app for? Did you experience technical difficulties with the app?) and acceptability (Would you recommend the app to others? Did you feel communication between families and providers around vaccines was good after using the app? Did you feel knowledgeable about pretransplant vaccines after app use?). The exit questionnaire included an open-ended question asking for suggestions for future app improvement. The exit questionnaire was sent via email after the transplant was completed. Weekly reminders were sent until the questionnaire was complete.

Vaccine Rates at the Time of Transplant

A research team member collected vaccine records at the time of transplant to evaluate whether the child was up to date on age-appropriate vaccines at the time of transplant.

Statistical Analysis

Effectiveness

We calculated the percentage of transplant recipients who were completely up to date on age-appropriate vaccines at the time of transplant according to standard CDC vaccine recommendations [23].

Feasibility

Data from feasibility questions in the exit questionnaire were summarized through the proportion of the feasibility outcome.

Acceptability

Data from acceptability questions in the exit questionnaire were summarized through the proportion of the acceptability outcome.

Ethical Considerations

The study was approved by the Colorado Multiple Institutional Review Board (CO-18‐0045). All participants provided consent for study participation. Participation was voluntary. An email and password were required to log in to the study app. Participants could access all study content through a mobile app available on both the iOS and Android platforms. All study data were encrypted at rest and in transit. Only authorized researchers were provided access to the participant data. The application's data was stored exclusively on Amazon Web Services in the United States. Industry standard information security practices were leveraged in the development of the app. All data were deidentified to ensure participant privacy and confidentiality. Participants who completed the full study received US $100 compensation. No artificial intelligence was used in any portion of the manuscript writing.

Results

Participants

We enrolled 25 dyads of parents and guardians of heart, liver, and kidney transplant recipients and their transplant provider. The majority (15/25, 60%) of children were <5 years old (Tables 1 and 2).

XSL•FO
RenderX

Table . Parent and guardian demographic characteristics (n=25).

Characteristic	Value, n (%)
Gender identity	
Female	20 (80)
Male	5 (20)
Ethnicity	
Caucasian	18 (72)
African American	2 (8)
Asian	2 (8)
Hispanic-Latino	1 (4)
Other or prefer not to answer	2 (8)
Center of care	
Children’s Hospital Colorado	19 (76)
Children’s Hospital of Philadelphia	3 (12)
Ann and Robert H. Lurie Children’s Hospital	3 (12)
How far do you live from your transplant center?	
<1 hour drive	8 (32)
1 - to 3-hour drive	10 (40)
>3-hour drive	3 (12)
We are in a different state than our center and need to fly to our center	4 (16)
How would you describe the town you live in?	
Rural	7 (28)
Suburban	13 (52)
Urban	5 (20)
How old is your child?	
0 - 11 months	4 (16)
12 months to 5 years	11 (44)
6 - 11 years	4 (16)
12 - 18 years	6 (24)
What organ is your child listed for?	
Heart	4 (16)
Liver	14 (56)
Kidney	7 (28)

XSL•FO
RenderX

Table . Provider demographic characteristics (n=15; there were providers who had multiple patients in the study).

Characteristic	Value, n (%)
Role on transplant team	
Cardiologist, hepatologist, or nephrologist	4 (27)
Transplant fellow	1 (7)
Transplant nurse coordinator	10 (67)
What transplant center do you work with?	
Children's Hospital Colorado	10 (67)
Children's Hospital of Philadelphia	2 (13)
Ann and Robert H. Lurie, Children's Hospital	3 (20)
How long have you worked as an attending transplant physician, transplant fellow, or transplant nurse coordinator?	
<1 year	3 (20)
1 - 5 years	5 (36)
6 - 10 years	3 (20)
>10 years	3 (20)
Gender identity	
Female	14 (93)
Male	1 (7)
Ethnicity	
Caucasian	11 (73)
African American	0 (0)
Asian	1 (7)
Hispanic-Latino	1 (7)
Other or prefer not to answer	2 (13)

Effectiveness of the App

Vaccine records were obtained on the day of transplant; 96% (24/25) of children were up to date on all age-appropriate vaccines. The 1 patient who was not up to date was a 15-year-old who was missing a second human papillomavirus (HPV) shot. The patient's team had specifically chosen to delay the second HPV vaccine until after transplant because the patient was on high-dose steroids for autoimmune hepatitis and was unlikely to mount a strong immune response to a pretransplant vaccine.

Feasibility of the App

App Usage

Parents and providers were instructed to use the Immunize PediatricTransplant app until the day of their child's transplant (or until the end of the study on April 1, 2024, whichever came first). No children were transplanted in less than a month, 27% (6/22) were transplanted in 1 - 3 months, 23% (5/22) in 4 - 6 months, and 50% (11/22) in 7 or more months.

Technical Difficulties While Using the App

Only 23% (5/22) of parents and 21% (3/14) of providers reported technical difficulties with the app. Difficulties described included initial log on (n=5), entry of historical vaccines or vaccines given while using the app (n=2), and communication with team members through the portal (n=1).

Acceptability of the App

Of the 25 parents and guardians and 15 providers, 88% (22/25) and 93% (14/15), respectively, completed an exit questionnaire. Of these, 97% (35/36) agreed or strongly agreed that they felt knowledgeable about vaccine use in the pretransplant period after using the app and 86% (31/36) agreed or strongly agreed that communication around vaccines between families and providers was good after using the app. Users rated the following features as extremely or moderately useful: the centralized vaccine record (n=30, 83%), the automated vaccine reminders (n=24, 67%), and the written educational information (n=21, 58%; Table 3). Further, 91% (20/22) of parents and 79% (11/14) of providers recommended the app to future families and transplant teams with children awaiting a transplant.

XSL•FO
RenderX

Table . Acceptability of the various features of the app (n=36; parents, guardians, and providers combined).

	Extremely or moderately useful, n (%)	Somewhat useful, n (%)	Not useful, n (%)
Written educational information in the app	21 (58)	12 (33)	3 (8)
Communication tool (chat feature)	18 (50)	9 (25)	9 (25)
Automated vaccine reminders	24 (67)	8 (22)	4 (11)
Centralized vaccine record	30 (83)	2 (6)	4 (11)

Suggestions for App Improvement

The open-ended responses to the question of "Do you have any suggestions for app improvement?" are listed in Textbox 1. The most common suggestion (n=6) was to populate the vaccine history from the EMR or other vaccine databases such as immunization information systems (IISs).

Textbox 1. Open-ended responses in the exit questionnaire to "Do you have any suggestions for app improvement?"

Responses

- Pull vaccine records from electronic medical record
- Automatically populate the vaccination history
- When entering vaccines, it would be nice to enter more than one at a time
- Scan vaccine documents or have a professional health care provider enter in the vaccine data, not the family, there's too much room for error
- Allow downloads of vaccine records from state immunization information systems (IIS)
- Connect the app to a vaccine data base
- Space for additional vaccines (extra COVID doses or repeat series)
- Make the live chat a more noticeable feature
- Improve ease of use for vaccine entry
- Connect primary care provider with the app
- Have a way to submit option of family refusal to avoid getting reminders about a vaccine your family is refusing
- Tech support for users
- Create a way for medical providers to silence alerts for vaccines that family is refusing
- App appearance
- If we could have the primary care pediatrician participate, that would be great
- Encourage primary care pediatricians to use it for communication
- Integration with vaccine databases such as the Colorado Department of Public Health and Environment

Discussion

Principal Findings

We sought to beta test the Immunize PediatricTransplant mHealth app to understand whether (1) it could help address the problem of underimmunization of pediatric solid organ transplant candidates and (2) it would be feasible and acceptable to parent and transplant provider users. The most significant finding of the study is the success of the app in getting nearly all children up to date on age-appropriate vaccines by the time of transplant. Historically, the majority of pediatric transplant candidates have not been up to date on age-appropriate vaccines at the time of transplant. In a study of 281 children who underwent a pediatric liver transplant across the United States between August 2017 and August 2018, only 19% were up to date on age-appropriate vaccines on the day of transplant and amongst those who were not up to date, 51% were missing 4 or more immunizations [6]. Increasing immunization rates pretransplant could significantly improve posttransplant outcomes by decreasing posttransplant VPIs and the associated morbidity, mortality, and increased medical costs [4,5].

While the app was overall feasible and acceptable to users (31/36, 86% of users recommended the app to future parents and providers), in the open-ended question in the exit questionnaire asking for suggestions for app improvement we learned that users did not want to have to enter vaccine records themselves into the app (in real-world use, there would be no research coordinator to enter vaccines). Multiple users recommended that the app connect directly with the EMR or IISs to retrieve accurate vaccine data.

There are several apps currently available in the App Store to promote childhood vaccination [24]. To our knowledge,

XSL•FO
RenderX

Immunize PediatricTransplant is the first app designed specifically for a group of children with complex medical needs at a higher risk for acquiring VPIs. In a 2020 systemic review of 28 studies evaluating 25 immunization apps, 4 studies described significant benefits of implementation of the app and 4 studies reported a significant impact on knowledge and vaccine decision-making [25]. Our study adds additional evidence that an mHealth app can improve vaccine rates in a high-risk population of children. In addition to providing vaccine education, record keeping, and reminder systems, Immunize PediatricTransplant also includes a communication portal for families and providers to directly communicate in real time about vaccine questions.

Future Directions

Based on user feedback, in the future we will explore utilizing application programming interfaces (APIs) like Health Level 7 FHIR (Fast Healthcare Interoperability Resources) [26] and SMART on FHIR [27] to allow the app to directly connect and communicate with the EMR and IISs [28]. While APIs are legal according to the 21st Century Cures Act [29], it would be critical to ensure compliance with Health Insurance Portability and Accountability Act (HIPAA) regulations protecting patient privacy. We will then need to evaluate the tool in a clinical trial (phase 3 of the mHealth Agile Development and Evaluation Lifecycle) [22] to see if it results in up-to-date immunization status at the time of transplant for a diverse population of pediatric transplant candidates across the country. As this is a "low-risk app" that does not pose any risk to user safety if it does not function (providers should still be reviewing vaccines as part of standard medical care), an alternative to a formal clinical trial would be ongoing A/B testing where a group of users trial a new version of the app that contains minor variations to a specific component of the app [22]. Based on user feedback, ongoing iterations of the app could be made. In future studies, we will modify the primary end point to not classify patients with deliberately missed vaccines (as instructed by the medical team) as not up to date.

After clinical trial or A/B testing, postmarket surveillance will be important to gather ongoing feedback about if the app remains useful and effective and whether further improvements are needed to keep it relevant as new vaccines come into the market [22]. In future research, it will be important to expand the qualitative sections of the exit questionnaire or perform postintervention interviews or focus groups to gain a deeper understanding of the users' experiences, likes, and dislikes about the app and recommendations for future app users including primary care physicians.

Limitations

Multiple limitations exist in this study beta testing the Immunize PediatricTransplant app. Our sample size was small (only 25 dyads of parents and transplant providers, which equated to 36 app users) and the majority of participants came from 1 large academic pediatric transplant center. Additionally, the app was only trialed by English-speaking parents and providers, limiting generalizability to all transplant parents and providers. Participation in the study was voluntary, which could introduce selection bias where people who were more interested in or comfortable with mHealth technology were more likely to participate. Similarly, there could be social desirability bias, where people were more likely to report finding the app to be beneficial in an attempt to please researchers. Finally, it is possible that simply by having parents and providers participate in an intervention-based research study, it increased attention and awareness to vaccines, in turn increasing the amount of transplant recipients with up-to-date vaccine statuses. As this was not a randomized trial, there was no control group to compare with. These limitations are all inherent to beta testing, and the app will need to go through future phase 3 clinical trial evaluation.

Conclusions

Each year, the United States invests $1.2 billion dollars in pediatric solid organ transplants [30]; however, we fail to protect our investment by not ensuring that these children are fully immunized. A novel mobile app, Immunize PediatricTransplant, has now been developed to help overcome barriers to pretransplant vaccines. The app was feasible and acceptable to parent and provider users and successfully resulted in nearly all children being up to date on age-appropriate vaccines by the time of transplant. However, the app was reliant on a research team member entering historical vaccine information, which is not practical for real-world use. In future iterations, we will need to explore whether APIs could be leveraged to extract vaccine information from the EMR or IISs and directly input these data into the app.

Acknowledgments

The authors would like to acknowledge all transplant stakeholders who participated in this research. AGF is supported by the Agency for Healthcare Research and Quality (grant K08 HS026510). Contents are the authors' sole responsibility and do not necessarily represent official National Institutes of Health (NIH) views.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

KW and CB are cofounders of CANImmunize Inc. KW is a paid member of the independent data safety board for the Medicago vaccine trial.

XSL•FO
RenderX

References

1. Burroughs M, Moscona A. Immunization of pediatric solid organ transplant candidates and recipients. Clin Infect Dis 2000 Jun;30(6):857-869. [doi: 10.1086/313823] [Medline: 10852737]
2. Kaplan LJ, Daum RS, Smaron M, McCarthy CA. Severe measles in immunocompromised patients. JAMA 1992 Mar 4;267(9):1237-1241. [doi: 10.1001/jama.1992.03480090085032] [Medline: 1538561]
3. McGregor RS, Zitelli BJ, Urbach AH, Malatack JJ, Gartner JC Jr. Varicella in pediatric orthotopic liver transplant recipients. Pediatrics 1989 Feb;83(2):256-261. [Medline: 2536479]
4. Feldman AG, Sundaram SS, Beaty BL, Kempe A. Hospitalizations for respiratory syncytial virus and vaccine-preventable infections in the first 2 years after pediatric liver transplant. J Pediatr 2017 Mar;182:232-238. [doi: 10.1016/j.jpeds.2016.12.021] [Medline: 28088400]
5. Feldman AG, Beaty BL, Curtis D, Juarez-Colunga E, Kempe A. Incidence of hospitalization for vaccine-preventable infections in children following solid organ transplant and associated morbidity, mortality, and costs. JAMA Pediatr 2019 Mar 1;173(3):260-268. [doi: 10.1001/jamapediatrics.2018.4954] [Medline: 30640369]
6. Feldman AG, Sundaram SS, Beaty BL, Torres R, Curtis DJ, Kempe A. Immunization status at the time of liver transplant in children and adolescents. JAMA 2019 Nov 12;322(18):1822-1824. [doi: 10.1001/jama.2019.14386] [Medline: 31714979]
7. Feldman AG, Marsh R, Kempe A, Morris MA. Barriers to pretransplant immunization: a qualitative interview study of pediatric solid organ transplant stakeholders. J Pediatr 2020 Dec;227:60-68. [doi: 10.1016/j.jpeds.2020.07.038] [Medline: 32681988]
8. Krishna S, Boren SA. Diabetes self-management care via cell phone: a systematic review. J Diabetes Sci Technol 2008 May;2(3):509-517. [doi: 10.1177/193229680800200324] [Medline: 19885219]
9. Scherr D, Kastner P, Kollmann A, et al. Effect of home-based telemonitoring using mobile phone technology on the outcome of heart failure patients after an episode of acute decompensation: randomized controlled trial. J Med Internet Res 2009 Aug 17;11(3):e34. [doi: 10.2196/jmir.1252] [Medline: 19687005]
10. Kao CK, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. PM R 2017 May;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
11. Domin A, Spruijt-Metz D, Theisen D, Ouzzahra Y, Vögele C. Smartphone-based interventions for physical activity promotion: scoping review of the evidence over the last 10 years. JMIR Mhealth Uhealth 2021 Jul 21;9(7):e24308. [doi: 10.2196/24308] [Medline: 34287209]
12. Tabi K, Randhawa AS, Choi F, et al. Mobile apps for medication management: review and analysis. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e13608. [doi: 10.2196/13608] [Medline: 31512580]
13. Wilson K, Atkinson KM, Penney G. Development and release of a national immunization app for Canada (ImmunizeCA). Vaccine 2015 Mar;33(14):1629-1632. [doi: 10.1016/j.vaccine.2015.02.022] [Medline: 25704801]
14. Wilson K, Atkinson KM, Westeinde J. Apps for immunization: leveraging mobile devices to place the individual at the center of care. Hum Vaccin Immunother 2015;11(10):2395-2399. [doi: 10.1080/21645515.2015.1057362] [Medline: 26110351]
15. Wilson K, Atkinson KM, Deeks SL, Crowcroft NS. Improving vaccine registries through mobile technologies: a vision for mobile enhanced Immunization information systems. J Am Med Inform Assoc 2016 Jan;23(1):207-211. [doi: 10.1093/jamia/ocv055] [Medline: 26078414]
16. CANImmunize. Apple App Store. URL: https://itunes.apple.com/ca/app/immunize-canada/id574951915?mt=8 [accessed 2025-04-02]
17. Feldman AG, Atkinson K, Wilson K, Kumar D. Underimmunization of the solid organ transplant population: an urgent problem with potential digital health solutions. Am J Transplant 2020 Jan;20(1):34-39. [doi: 10.1111/ajt.15605] [Medline: 31553135]
18. Zaidi S, Shaikh SA, Sayani S, et al. Operability, acceptability, and usefulness of a mobile app to track routine immunization performance in rural Pakistan: interview study among vaccinators and key informants. JMIR Mhealth Uhealth 2020 Feb 13;8(2):e16081. [doi: 10.2196/16081] [Medline: 32053115]
19. Mekonnen ZA, Gelaye KA, Were M, Tilahun B. Effect of mobile phone text message reminders on the completion and timely receipt of routine childhood vaccinations: superiority randomized controlled trial in northwest Ethiopia. JMIR Mhealth Uhealth 2021 Jun 15;9(6):e27603. [doi: 10.2196/27603] [Medline: 34128813]
20. Mekonnen ZA, Gelaye KA, Were MC, Gashu KD, Tilahun BC. Effect of mobile text message reminders on routine childhood vaccination: a systematic review and meta-analysis. Syst Rev 2019 Jun 28;8(1):154. [doi: 10.1186/s13643-019-1054-0] [Medline: 31253186]
21. Feldman AG, Moore S, Bull S, et al. A smartphone app to increase immunizations in the pediatric solid organ transplant population: development and initial usability study. JMIR Form Res 2022 Jan 13;6(1):e32273. [doi: 10.2196/32273] [Medline: 35023840]
22. Wilson K, Bell C, Wilson L, Witteman H. Agile research to complement agile development: a proposal for an mHealth research lifecycle. NPJ Digit Med 2018;1:46. [doi: 10.1038/s41746-018-0053-1] [Medline: 31304326]

XSL•FO
RenderX

23. Wodi AP, Murthy N, McNally V, Cineas S, Ault K. Advisory Committee on Immunization Practices Recommended Immunization Schedule for Children and Adolescents Aged 18 Years or Younger - United States, 2023. MMWR Morb Mortal Wkly Rep 2023 Feb 10;72(6):137-140. [doi: 10.15585/mmwr.mm7206a1] [Medline: 36757872]
24. Vaccination. Google Play Store. URL: https://play.google.com/store/search?q=vaccination&c=apps&gl=GB [accessed 2024-12-10]
25. de Cock C, van Velthoven M, Milne-Ives M, Mooney M, Meinert E. Use of apps to promote childhood vaccination: systematic review. JMIR Mhealth Uhealth 2020 May 18;8(5):e17371. [doi: 10.2196/17371] [Medline: 32421684]
26. What is HL7 FHIR? Office of the National Coordinator for Health Information Technology. URL: https://www.healthit.gov/sites/default/files/page/2021-04/What%20Is%20FHIR%20Fact%20Sheet.pdf [accessed 2024-12-10]
27. SMART on FHIR. SMART Health IT. URL: https://docs.smarthealthit.org [accessed 2024-12-10]
28. What is an API? Amazon Web Services. URL: https://aws.amazon.com/what-is/api/ [accessed 2024-01-11]
29. 21st Century Cures Act, HR 34, 114th Cong (2015). URL: https://www.congress.gov/bill/114th-congress/house-bill/34/text [accessed 2025-02-18]
30. Bentley TS, Phillips SJ. 2017 U.S. organ and tissue transplant cost estimates and discussion. : Milliman; 2017 URL: https://www.milliman.com/en/insight/2017-us-organ-and-tissue-transplant-cost-estimates-and-discussion [accessed 2025-04-02]

Abbreviations:

API: application programing interface
CDC: Centers for Disease Control
EMR: electronic medical record
FHIR: Fast Healthcare Interoperability Resources
HIPAA: Health Insurance Portability and Accountability Act
HPV: human papillomavirus
IIS: immunization information system
mHealth: mobile health
VPI: vaccine-preventable infection

Edited by A Mavragani; submitted 16.12.24; peer-reviewed by IA Eide; revised version received 18.02.25; accepted 20.02.25; published 15.04.25.

Please cite as:
Feldman AG, Beaty BL, Moore SL, Bull S, Wilson K, Atkinson KM, Bell C, Denize KM, Kempe A
Feasibility, Acceptability, and Effectiveness of a Smartphone App to Increase Pretransplant Vaccine Rates: Usability Study
JMIR Form Res 2025;9:e68855
URL: https://formative.jmir.org/2025/1/e68855
doi:10.2196/68855

© Amy G Feldman, Brenda L Beaty, Susan L Moore, Sheana Bull, Kumanan Wilson, Katherine M Atkinson, Cameron Bell, Kathryn M Denize, Allison Kempe. Originally published in JMIR Formative Research (https://formative.jmir.org), 15.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Development of an eHealth Mindfulness-Based Music Therapy Intervention for Adults Undergoing Allogeneic Hematopoietic Stem Cell Transplantation: Qualitative Study

Sara E Fleszar-Pavlovic[1], PhD; Blanca Noriega Esquives[1], MD, PhD; Padideh Lovan[1,2], PhD, RD, LDN; Arianna E Brito[1], BS; Ann Marie Sia[3]; Mary Adelyn Kauffman[1], DMA, MT-BC, NMT; Maria Lopes[4], BA; Patricia I Moreno[1,5], PhD; Tulay Koru-Sengul[1,5], PhD, MHS; Rui Gong[1,6], PhD; Trent Wang[1,6], DO, MPH; Eric D Wieder[1,6], PhD; Maria Rueda-Lara[1,7], MD; Michael Antoni[1,8], PhD; Krishna Komanduri[9], MD; Teresa Lesiuk[1,10], PhD, MT-BC; Frank J Penedo[1,8], PhD

[1]Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, FL, United States

[10]Frost School of Music, University of Miami, Coral Gables, FL, United States

[2]School of Nursing and Health Studies, University of Miami Miller School of Medicine, Miami, FL, United States

[3]Department of Undergraduate Research, University of Miami, Coral Gables, FL, United States

[4]Department of Psychology, College of Arts and Sciences, University of Miami, Coral Gables, FL, United States

[5]Department of Public Health Sciences, University of Miami Miller School of Medicine, Miami, FL, United States

[6]Department of Medicine, University of Miami Miller School of Medicine, Miami, FL, United States

[7]Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, Miami, FL, United States

[8]Departments of Psychology and Medicine, University of Miami Miller School of Medicine, 1120 NW 14th Street, Miami, FL, United States

[9]Department of Medicine, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Frank J Penedo, PhD
Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, FL, United States

Abstract

Background: Allogeneic hematopoietic stem cell transplantation (allo-SCT) is an effective treatment for various hematologic cancers, though it often results in severe side effects and psychological distress, which can negatively impact health outcomes. Integrative therapies like mindfulness-based stress reduction (MBSR), mindfulness meditation (MM), and music therapy (MT) yield promising results in enhancing both psychosocial outcomes (eg, reducing anxiety and depression) and physiological adaptation (eg, decreasing inflammation) in cancer patients.

Objective: We developed and refined, using focus groups and environmental and field testing, an eHealth-delivered mindfulness-based music therapy (eMBMT) intervention aimed at improving health-related quality of life, symptom burden (ie, pain, fatigue, and sleep), disease activity (ie, chronic graft-versus-host disease, cytomegalovirus activation, and infections) and psychosocial (ie, depression, anxiety, and cancer-specific distress) and physiological adaptation (ie, inflammation and immune reconstitution) tailored to adults receiving allo-SCT.

Methods: eMBMT intervention content is grounded in MT, MM, and MBSR, developed by a multidisciplinary team, and adapted for adults undergoing allo-SCT. eMBMT content was refined through focus groups and usability and field testing. Focus groups used a semistructured interview guide, while field testing used the “think aloud” method. Usability was evaluated using the 30-item Usefulness, Satisfaction, and Ease of Use (USE) questionnaire. Descriptive statistics analyzed the USE questionnaire and participant characteristics, while rapid qualitative analysis was applied to focus groups and field-testing sessions. Survivors eligible to participate in the focus groups and usability and field testing were adults (>18 years old) who received an allo-SCT (<36 months) for myelodysplastic syndrome, acute myeloid leukemia, or chronic myeloid leukemia, and were in remission for greater than 3 months.

Results: During the focus groups, participants (n=11; mean age 43.6, SD 17.8 years) provided qualitative feedback highlighting the shock of diagnosis, challenges during hospitalization, and coping strategies posttreatment. The eMBMT platform received positive evaluations for usefulness (mean 6.47, SD 0.29), ease of use (mean 6.92, SD 0.60), and satisfaction (mean 6.16, SD 0.82). Key themes from field testing highlighted the significance of social support, hope, and maintaining an active lifestyle. Suggestions for improvement included incorporating more representative content, reducing text, enhancing guidance, offering diverse music options, and streamlining blood sample collection.

XSL•FO
RenderX

Conclusions: The eMBMT intervention is a comprehensive, user-friendly eHealth tool tailored to the unique needs of allo-SCT patients. The positive feedback and identified areas for improvement underscore its potential to enhance well-being, symptom management, and overall quality of life for cancer survivors. A future pilot randomized controlled trial will further evaluate the feasibility, acceptability, and preliminary efficacy of the eMBMT intervention in improving health-related quality of life, symptom burden, disease activity, and psychosocial and physiological adaptation.

Trial Registration: ClinicalTrials.gov NCT05968963; https://clinicaltrials.gov/study/NCT05968963

(JMIR Form Res 2025;9:e65188) doi:10.2196/65188

KEYWORDS

allogeneic stem cell transplantation; hematologic malignancy; bone marrow transplant; mindfulness-based music therapy; mindfulness; music therapy; eHealth; music therapy intervention; adult; adolescence; allogeneic; stem cell transplantation; stem cell; transplantation; qualitative study; treatment; hematologic cancers; psychological distress; side effects; mindfulness-based stress reduction; stress reduction; anxiety; depression; diagnosis; blood sample collection; eHealth tool; quality of life; cancer survivors

Introduction

In the United States, hematologic malignancies account for approximately 10% of all new cancer diagnoses and 9.4% of cancer-related deaths annually [1]. It is projected that over 180,000 new cases of malignant hematologic cancers will be diagnosed in 2024 [2]. Allogeneic hematopoietic stem cell transplantation (allo-SCT) is an effective therapy for various hematologic cancers, including acute myeloid leukemia, chronic myeloid leukemia, and myelodysplastic syndrome [3]. Allo-SCT is a multiphase therapy starting with a pretransplant evaluation and finding a compatible donor. This is followed by intensive conditioning (chemotherapy, radiation, or both) to destroy cancer cells, suppress the immune system, and prepare for stem cell engraftment. The patient then receives the donor's stem cells intravenously and enters the posttransplant phase, focusing on engraftment and monitoring for complications like graft-versus-host disease (GVHD) and infections [3].

Despite its curative potential, allo-SCT may be accompanied by negative side effects (eg, nausea, vomiting, pain, fatigue, mood disturbance, and risk of infections) [4]. In addition, poor psychosocial functioning and emotional distress (eg, anxiety, depression, and social isolation) can persist well beyond active treatment [5-7]. These negative psychological factors, common among allo-SCT recipients [5,6], are associated with higher mortality rates and less favorable health outcomes in the first-year posttreatment [8-10], including greater rates of GVHD and infection, longer hospitalization stays, and higher readmission rates [11]. Furthermore, research indicates that psychological distress can affect physiological responses to treatment such as increased inflammation [12,13]. For example, in patients receiving allo-SCT, pretransplant anxiety, depression, and stress have been associated with delayed white blood cell recovery, slower immune function recovery, and shorter survival. In contrast, pretransplant optimism, positive affect, and social support correlate with a shorter time to engraftment and reduced incidence of GVHD [14-17]. Therefore, interventions targeting psychological distress and physical burden in allo-SCT recipients may enhance health-related quality of life (HRQoL), improve patient-reported outcomes, and support immune recovery and overall health in cancer survivors.

A growing body of research suggests that mindfulness-based interventions, including mindfulness-based stress reduction (MBSR) and mindfulness meditation (MM) yield promising results in enhancing both psychosocial outcomes (eg, reducing anxiety and depression) and physiological adaptation (eg, decreasing inflammation) in cancer patients [18-21]. The core principle of mindfulness involves maintaining nonjudgmental awareness of the present moment, which includes recognizing and accepting one's thoughts, emotions, and bodily sensations as they arise. Practicing mindfulness fosters a greater sense of self-awareness and emotional regulation, which can contribute to improved mental and physical well-being [22]. Although highly underused, mindfulness-based interventions have proven effective in alleviating cancer symptoms, cancer-related distress, and depression in various patient populations. Mindfulness-based interventions also enhance self-efficacy, coping skills, and HRQoL in cancer patients [12,20,23-27] and individuals with other chronic diseases [28-31].

In addition to mindfulness, music therapy (MT) is a well-recognized approach for addressing various psychosocial and physiological needs in clinical populations [32-44]. It has demonstrated effectiveness in reducing stress and enhancing emotional well-being among cancer patients [35,45-49]. MT offers numerous benefits for cancer patients by using both interactive (eg, singing) and receptive techniques (eg, listening to music and music imagery) to improve mood, reduce stress, pain, and anxiety, as well as promote relaxation [50]. Although most findings suggest psychological improvement associated with MT, a recent meta-analysis of controlled trials provided evidence that MT can significantly improve physical symptoms (eg, fatigue) and reduce pain in cancer patients [51,52]. MT can also serve as a foundation for planning effective rehabilitation programs and improving overall quality of life [50]. Using MT is known to be a low-risk, minimally invasive, and cost-effective approach to managing patients' well-being particularly in the often-stressful environment of a cancer clinic [53].

Although previous research has demonstrated that practices related to mindfulness and MT can positively affect fatigue, pain, anxiety, and HRQoL in cancer patients [54,55], there is a lack of well-powered randomized controlled trials and reproducible interventions specific for allo-SCT patients. In addition, significant gaps persist in our understanding of how

XSL•FO
RenderX

specific components of mindfulness and MT impact psychosocial and physiological adaptation, HRQoL, and overall health outcomes in this population. Therefore, we designed an eHealth-delivered mindfulness-based music therapy (eMBMT) aimed at reducing symptom burden and improving HRQoL in patients undergoing allo-SCT. We selected an eHealth intervention based on its proven advantages, including patient accessibility and convenience, improved communication, and reduced burden on health care systems. Overall eHealth interventions offer a valuable approach to complement traditional cancer care [56]. This study aims to refine and finalize, using focus groups and environmental and field testing, the newly developed eMBMT intervention.

Methods

Intervention Content Development

The eMBMT content was developed by a multidisciplinary team (eg, board-certified music therapists [MT-BC], psychologists, oncologists, and psychiatrists) and is based on MT, MM, MBSR [57], and Lesiuk's [58,59] MBMT pilot trial with breast cancer survivors. eMBMT was further adapted for patients undergoing allo-SCT and structured to be delivered through BrightOutcome's SmartManage: Tools for Health Living [60,61] platform. SmartManage is a patient-centered web application that is easy to use, Health Insurance Portability and Accountability Act (HIPAA) compliant, and accessible through multiple platforms (ie, tablets and smartphones). SmartManage serves as the primary platform for hosting the MBMT content, organized into 8 modules corresponding to sessions led by the MT-BC. SmartManage enables participants to engage in real time during sessions and supports independent practice by providing access to session materials, guided exercises, and additional resources for extended learning. Participants were encouraged to engage with SmartManage outside of scheduled sessions, using its resources to review session materials and independently practice intervention components at their own pace, thereby reinforcing the therapeutic process. The eMBMT program consists of eight 60-minute sessions led by a MT-BC. The sessions integrate MT and mindfulness attitudes (ie, nonjudging, patience, letting go, acceptance, trust, beginners mind, and nonstriving) developed by Kabat-Zinn [57]. MT-BCs guide patients in engaging with video and audio materials, practicing active and receptive music therapy exercises, incorporating music and journaling, and participating in video conferencing, all aimed at fostering mindfulness attitudes through music therapy. eMBMT's 8 tailored sessions span the allo-SCT treatment and recovery journey beginning upon hospital admission (approximately 7 days before transplant) and ending postdischarge (approximately 75 - 115 days posttransplant).

eMBMT begins with an orientation session to provide an overview of the program and its aims. The following 7 sessions are aligned with the patient's treatment trajectory, each emphasizing a specific mindfulness attitude to be practiced throughout their allo-SCT treatment: (1) nonjudging encourages evaluating stimuli without habitual reactions, whether from internal dialogue, sensations, or external events; (2) patience helps manage worry by allowing events to unfold over time; (3) letting go involves accepting things as they are and being open to change; (4) acceptance focuses on viewing one's experience without comparisons and allowing recovery to unfold at its own pace; (5) trust emphasizes believing in oneself and empowering self-reliance over thoughts; (6) beginner's mind encourages seeing newness in familiar environments, fostering a sense of wonder, and shedding negative perceptions; and (7) nonstriving promotes the willingness to try new experiences and suspend critical judgments, particularly through new music experiences. MT-BCs focus on guiding patients through each session to first learn the attitude through music exercises, then how to apply the attitude to potentially improve their treatment journey.

Intervention Refinement

Following the initial content development, we used focus groups and usability and field-testing sessions to refine and finalize the eMBMT content and intervention.

Sampling and Recruitment

Participants were recruited by (a) physician referral or (b) the University of Miami's consent-to-contact database to participate in a focus group and a subsequent one-on-one usability and field-testing session. Patients at the University of Miami Health System can voluntarily enroll and consent to be contacted about clinical study opportunities for which they qualify through the Consent to Contact Registry managed by the University of Miami Clinical and Translational Science Institute. Potential participants were identified using the Clinical and Translational Science Institute Consent-to-Contact Registry through the generation authorized patient lists based on International Classification of Diseases (ICD-10) codes. Allo-SCT survivors were eligible to participate if they were (1) aged >18 years; (2) received an allo-SCT for myelodysplastic syndrome, acute myeloid leukemia, or chronic myeloid leukemia <36 months; and (3) were in remission >3 months. Survivors who (1) had a history of severe psychiatric illness (eg, psychosis, active suicidality, or inpatient treatment) in the past 12-months, (2) severe cognitive impairment (per the Short Portable Mental Status Questionnaire) [62], or (3) hearing impairment were excluded from the study. Patients who were referred by their physician or were identified by the consent-to-contact list were contacted by telephone by study staff to assess interest in participating and eligibility. Of the survivors (n=64) screened for eligibility, 34 were ineligible, 11 declined screening, and 7 did not consent to participate. A total of 12 survivors agreed to participate and provided informed consent; however, 1 participant was ill and unable to attend the focus group. The study team conducted 2 focus groups (focus group 1: n=6 and focus group 2: n=5) and 9 one-on-one usability and field-testing sessions.

Procedures

Before attending a focus group, participants completed a survey assessing sociodemographic (eg, age, gender, race, and income) and medical characteristics (eg, cancer type and time in remission). Focus groups were conducted through Zoom (Zoom Communications) and facilitated by 2 MT-BCs and study staff

XSL•FO
RenderX

trained in qualitative methods. The aim of the focus groups was to understand the needs and burdens of allo-SCT patients and gather qualitative feedback to inform eMBMT content and platform design. Facilitators followed a semistructured interview guide including open-ended questions assessing: (1) challenges associated with receiving an allo-SCT (eg, uncertainty of the disease course, treatment-related side effects, and management of emotional and physical symptoms), (2) use of eHealth-based delivery for mindfulness, music therapy, and mindfulness-based music exercises, and (3) preferences for eMBMT length and detail of content. Facilitators also presented 2 modules from the eMBMT SmartManage platform and participants were asked about design and layout preferences (eg, theme, style, colors, and fonts), content detail and type (eg, text, images, and videos), and functionality features (eg, navigation, accessibility, and user interaction). Each focus group was facilitated by 2 experienced moderators (ie, "facilitators") with training in qualitative research methods (SEF-P, BNE, TL, and MAK). To ensure balanced participation and that all voices were heard, the facilitators used strategies such as directly inviting quieter participants to share their thoughts, using prompts to encourage diverse perspectives, and managing dominant voices to maintain an inclusive environment [63,64]. Focus groups were used to capture a diverse range of perspectives and facilitate interactive discussions among participants, allowing for the exploration of shared experiences and unique viewpoints related to the intervention. This approach was chosen because participants could build on each other's ideas, providing deeper insights into coping strategies and preferences. By leveraging the group dynamic, focus groups enabled the identification of common challenges and individual needs, ensuring the intervention was informed by both collective and personal experiences. Focus groups lasted approximately 3 hours and were audio recorded and detailed notes were taken to supplement the recordings.

Participants were contacted 1 week post–focus group participation to complete a usability and field-testing session. The usability assessment evaluated the ease of learning and identified platform design issues. Field testing evaluated both technical (eg, technical problems) and functional (eg, application use and integrating eMBMT into daily life) reliability. Sessions were conducted in-person or through Zoom and facilitated by study staff trained in usability and field testing of eHealth platforms. For usability testing, participants completed the 30-item USE questionnaire [65] which contains 4 subscales assessing usefulness (eg, "It helps me be more effective"), ease of use (eg, "It is easy to use"), ease of learning (eg, "I learned to use it quickly"), and satisfaction (eg, "I am satisfied with it") on an 8-point Likert scale (1=strongly disagree to 8=strongly agree). For field testing, we used the "think aloud" method [66], which encourages participants to vocalize any thoughts or comments while performing or immediately after performing a task, such as reading or watching a video. Participants were asked to share their thoughts and reactions as they explored the eMBMT and platform. Finally, participants were prompted with questions addressing technical reliability (eg, technology problems) and functional reliability (eg, application use and integrating eMBMT into daily life). The usability and field-testing sessions lasted approximately 3 hours and were audio-recorded.

Data Analysis

The Usefulness, Satisfaction, and Ease of Use (USE) questionnaire [65] and sociodemographic and medical characteristics were analyzed with descriptive statistics. We applied rapid qualitative analysis (RQA) to analyze the focus groups and usability and field-testing sessions following Watkins' guidelines [67]. Relative to traditional qualitative analysis methods, such as thematic analysis and grounded theory, which demand extensive time for tasks like multiple transcript readings and detailed coding, RQA is a more efficient and less time-demanding approach. RQA methods involve team members individually summarizing key points from qualitative data, and combining individual data by categorizing related key points to produce relevant themes [68]. RQA is a pragmatic, valid, and reliable method, which is crucial for guiding the refinement of digital interventions [68]. Study staff were trained on RQA, developed a coding matrix, individually coded the transcripts, and finally, met to complete a comprehensive final matrix summarizing data to facilitate the identification of themes. Themes were developed through iterative coding of the interview transcripts, with the team working to identify patterns and nuances in participant responses. While some themes aligned with the interview prompts, further analysis captured additional insights into coping mechanisms and patient preferences, which were critical for tailoring the intervention.

Ethical Considerations

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. We complied with the following ethical considerations. First, we recruited and obtained informed consent from all participants before the focus groups and usability and field-testing sessions, as approved by the University of Miami's Institutional Review Board (eProst 20230726). Participants were informed that their participation was voluntary, the focus groups and usability and field-testing sessions would be recorded, the collected information would be deidentified and securely stored, and they would receive compensation of US $50 for completing the focus group and US $50 for completing the usability and field-testing session, for a total of US $100. Second, we ensured that all transcript data were deidentified by replacing participant names with a unique identifier. All collected data are securely stored with access restricted to authorized researchers. Third, all researchers completed research compliance training through the Collaborative Institutional Training Initiative, ensuring adherence to best practices and ethical considerations in working with participants. This project is funded by the National Cancer Institute (1R61CA263335-01A1) and registered on ClinicalTrials.gov (NCT05968963).

Results

Participant Characteristics

Table 1 presents participant demographics and medical characteristics. Participants (N=11; mean age 43.6, SD 17.8; range 19 - 76 years) were female (55%, n=6), White (91%,

XSL•FO
RenderX

n=10), Hispanic (73%, n=8), with a history of acute myeloid leukemia (100%, n=11), and a mean time of 1.64 (SD 0.81) years post-SCT at the time of the focus groups.

Table . Demographics and medical characteristics of allogeneic stem cell transplantation (allo-SCT) survivors who participated in focus groups as part of the refinement of the eHealth-based mindfulness-based music therapy (eMBMT) intervention.

Demographics & medical characteristics	Total sample (N=11)	Focus group 1 (n=6)	Focus group 2 (n=5)
Age (years), mean (SD)	43.6 (17.8)	42.1 (15.9)	45.9 (18.4)
Gender, n (%)			
Female	6 (54.5)	4 (66.7)	2 (40)
Male	5 (45.5)	2 (33.3)	3 (60)
Race, n (%)			
White or Caucasian (including White Hispanic)	10 (91)	5 (83.3)	5 (100)
Black, African American, or Afro-Caribbean Black	1 (9)	1 (16.7)	0 (0)
Ethnicity, n (%)			
Hispanic	8 (72.7)	4 (66.7)	4 (80)
Non-Hispanic	3 (27.3)	2 (33.3)	1 (20)
Employment status, n (%)			
Full-time	5 (45.5)	3 (50)	2 (40)
Part-time	3 (27.3)	1 (16.7)	2 (40)
Unemployed with disability	2 (18.2)	2 (33.3)	0 (0)
Unemployed without disability	1 (9)	0 (0)	1 (20)
Education, n (%)			
High school or GED[a]	2 (18.2)	2 (33.3)	0 (0)
Some college	1 (9)	1 (16.7)	0 (0)
Graduated from a 2-year college	1 (9)	1 (16.7)	0 (0)
Graduated from a 4-year college	4 (36.5)	1(16.7)	3 (60)
Master's degree	3 (27.3)	1 (16.7)	2 (40)
Cancer type, n (%)			
Acute myeloid leukemia	11 (100)	6 (100)	5 (100)
Treatment type, n (%)			
Allogeneic stem cell transplant	11(100)	6 (100)	5 (100)
Time since diagnosis (months), n	26	26.7	25.3

[a]General Education Development diploma.

Focus Group Findings

Focus groups were conducted between April and May 2023. Qualitative analyses identified 4 themes.

The Initial Shock of a Cancer Diagnosis and Distress Over Unfavorable Outcomes

Most participants described the diagnosis of cancer as a profoundly shocking experience because they were diagnosed when they were feeling well and without any noticeable symptoms ("I didn't know I was sick. I had regular blood work done April of 2021 and everything was normal. And then June 1st I had a lump like a lymph node in my neck that was a little enlarged, but I wasn't feeling sick."). In addition, some participants hesitated to seek care because they were worried about exposure to COVID-19, resulting in delayed diagnosis and treatment ("I had previous history of asthma. So, my concern was: I go to the hospital, I'm gonna die… it's gonna hit people with respiratory problems, so I delayed it."). Once diagnosed, they grappled with uncertainty, the fear of missing out on meaningful activities, and the pressure to make tough treatment decisions (eg, fertility-related choices). After not responding to the initial treatment and given the option to receive an allo-SCT, participants faced frustration and anxiety related to finding a stem cell match, coping with chemotherapy's

XSL•FO
RenderX

physical side effects, and age-related issues (eg, hearing and vision loss, reduced mobility).

Challenges Associated With Hospitalization, Loss of Normalcy, and Adjusting to Life After Treatment

During hospitalization, participants experienced psychological distress due to prolonged social isolation and lack of personal interactions with their family and friends (eg, "...So, my isolation turned into 60, 70, 90 days, so that made it even worse"). They had feelings of loneliness, anxiety, boredom, and frustration surrounding missing major life events such as graduation, prom activities, or their children's events (eg, sport or school events). The COVID-19 pandemic restrictions negatively impacted the process of connecting patients to support services. Most participants reported a lack of access to support services during their hospital stay, which exacerbated their feeling of loneliness (eg, "I couldn't see anyone because it was COVID. So, for 2, 3 months there was nothing. Nobody offered me anything"). Following the conditioning phase and the stem cell infusion, participants experienced multiple physical side effects including weight loss, hair loss, nausea, vomiting, diarrhea, mouth sores, and fatigue. These symptoms negatively impacted their daily activity levels during their hospital stay and after being discharged. They also indicated several psychosocial side effects including, fear of recurrence, anxiety, depression, feeling very different *(*eg, "I'm not even close to who I was before this happened to me"), as well as a substantial impact on body image and its associated stigma (eg, "I don't want to be called the cancer kid at school"). They also noted that their families and friends were impacted by their disease (eg, "Another thing that I think it's forgotten is about the children. Most of us, we have children, and they have seen us in the darkest moments of our lives… and you think that they wouldn't remember, but they remember").

Coping Strategies for Living With a Hematologic Cancer

Participants developed various coping strategies for managing their disease. Participants vocalized several perceived useful strategies, such as (1) seeking social support from family, friends, health care staff, and other patients; (2) denial (eg, "I don't have cancer, I'm not going to admit it"); (3) spirituality and faith; (4) positive thinking and having an optimistic outlook; (5) goal setting; and (6) staying active and busy (eg, exercising, journaling, or watching television). In contrast, receiving adult coloring books and health communication materials was perceived as unhelpful or burdensome because they experienced changes in their ability to concentrate or memory lapses *(*"We don't read the green pamphlet because it's too much, too overwhelming, and you don't have the time or the mind to think about that, you're just literally surviving"). Support services were generally limited during their hospital stay due to COVID-19. However, when patients received supportive services, they valued the interaction with the therapist, provider, or both during those difficult times (eg, art therapy, music therapy, massage therapy, and physical therapy). In addition, although they mentioned that telehealth and videoconferencing were very convenient for maintaining continuity of care, they also experienced "screen fatigue." Participants also noted the lack of follow-up after being discharged and that they had to request support services rather than being automatically referred to them. A participant mentioned that "there's a waiting period of a year to see a sexologist in the gynecology department."

Usefulness of Mindfulness and Music Therapy to Cope With Cancer

Participants perceived mindfulness as useful for staying positive, strengthening their spirituality, and coping with uncertainty. Although participants were open to practicing mindfulness through online means, they preferred in-person sessions and the ability to share their experiences with other patients. Participants perceived music therapy as helpful to change their mood, and bring joy, confidence, and faith. In contrast, a participant described herself as "not a music person" and that she "could go in life without music and would be fine." Most participants were open to participating in online music therapy sessions; however, they underscored the need to account for patients' physical and psychological symptoms, which could potentially impact their participation.

Feedback on MBMT Prototype

Most participants agreed with the proposed number of sessions (n=8) and duration of the intervention (approximately 60 minutes/session). However, participants recommended implementing more sessions pretransplant to better help them be emotionally and psychologically prepared for the allo-SCT. They also suggested more sessions postdischarge to help them reintegrate into their day-to-day activities. Participants agreed with providing blood samples, but also suggested streamlining the process and obtaining samples along with already scheduled blood work. Participants described the eMBMT prototype within the SmartManage platform as "easy to read," "easy to navigate," "very straightforward," and "very appealing." Participants endorsed the order in which the mindfulness attitudes were presented (eg, "there is definitely a pattern, I think…There's sort of a journey to it"). They believed that the in-person interaction with the music therapist would be very helpful. Participants recommendations included (1) incorporating diverse music genres(eg, "disco when you need to move" or "hip hop when you need to walk"), (2) allowing music selection based on patient's mood, (3) recording audio or video clips to reduce the amount of text, (4) adding a sleep sounds feature, (5) reflecting the patient population within the content (eg, images specific allo-SCT concerns), (6) involving caregivers, (7) exploring a partnership with meditation and mindfulness apps, and (8) providing the necessary tools (eg, iPads or headphones).

Usability and Field-Testing Session Findings

Overview

Usability and field-testing sessions were completed between June and July 2023. The 2 participants who completed the focus groups were unable to participate in usability and field-testing sessions. Participants (n=9) agreed that the eMBMT platform was useful (mean 6.47 SD 0.29), easy to use (mean 6.92, SD 0.60) and learn (mean 7.39, SD 0.75), and were satisfied with the platform (mean 6.16, SD 0.82).

Usability and field-testing sessions were completed between June and July 2023. Qualitative analysis identified 2 themes.

XSL•FO
RenderX

Participants' Opinions About the MBMT Intervention

Overall, participants considered the program "helpful," "easy to read," and easy to navigate. They indicated that if the program had been provided during their hospitalization, they would have used it. Participants also liked the "concept" of the program (eg, holistic approach, accessibility, mindfulness meditation, and music to calm and distract from treatment). Regarding the aesthetics, participants perceived that the program was pleasant and appreciated personalized features (eg, their name on the front page). They also liked the colors, text size, layout, and images; although, 2 participants noted that the pictures did not represent the patient population accurately (eg, no photos of cancer patients without hair or racially diverse). Furthermore, some participants mentioned that the amount of text was overwhelming, section titles were confusing because they were very similar, and the platform lacked navigation guidance. They also mentioned that the exercises were repetitive with a lack of music options and that they could not have done some of the proposed activities during their hospital stay because of the physical symptoms (eg, mouth sores or extreme fatigue). Participants also experienced a few technical issues, such as some of the features not working (ie, audio), poor audio quality (ie, "white noise"), and slow internet connection.

Suggestions for Improving the eMBMT Intervention

Participants provided specific suggestions related to the content, including (1) reducing the amount of text by recording video and audio clips, (2) stating a specific goal at the beginning of each session (eg, this session will help you reduce your anxiety), (3) adding more relatable content (eg, "missing being outside," "you may be feeling fatigued today. Here is what you can do") and more practical knowledge (eg, "How will this help me?" and "Is there any section that helps you control pain by breathing more?") rather than definitions (eg, "what is mindfulness?"), (4) connecting the mindfulness attitudes to patients' experiences, (5) adding more guidance throughout the modules, (6) embedding YouTube videos with an introduction slide rather than links, (7) adding different music genres and the ability to choose, and (8) including more representative images. Participants also suggested new features such as (1) implementing a checkmark to indicate sections that have been reviewed, (2) allowing them to interact with other participants (eg, enabling a chat feature and support groups), (3) a journaling or writing section and a feedback box, (4) reducing the amount of clicking and scrolling, and (5) creating a Spotify playlist with the music from the program.

Discussion

Principal Findings

Overview

This study refined and finalized, using focus groups and environmental and field testing, an eMBMT intervention aimed at reducing symptom burden and improving HRQoL in patients undergoing allo-SCT. The focus group findings of this study underscore the significant emotional and physical challenges faced by allo-SCT patients, including the distress of diagnosis, hospitalization, and posttransplant recovery. These challenges align with previous research indicating that allo-SCT patients frequently experience psychological distress, including anxiety, depression, and feelings of isolation, which can negatively impact recovery and long-term survival [5,6,9]. Focus group findings also highlight that allo-SCT survivors had an appreciation for the potential positive impacts of the mindfulness and music therapy intervention components. In addition, the iterative refinement process, informed by usability and field-testing feedback, led to key modifications including enhanced content representativeness, reduced text with audio options, personalized music selection, and adjustments to session timing to better support patients pre- and posttransplant.

The development and refinement of the novel eMBMT intervention represents a significant step towards providing integrative interventions addressing HRQoL, symptom burden, disease activity, and psychosocial and physiological adaptation of adults undergoing allo-SCT. Grounded in theory and well-established therapeutic techniques (ie, MT, MM, and MBSR), the eMBMT intervention provides a structured delivery of in-person and eHealth sessions facilitated by an MT-BC, and is designed to be accessible and user-friendly. An eHealth intervention may be especially well suited for this population as shifting treatment phases and unexpected periods of isolation make it difficult to coordinate in-person support services.

The iterative process of refining the eMBMT content through focus groups and usability and field-testing sessions was crucial for its refinement. Engaging allo-SCT survivors in the development process allowed for the collection of valuable data that directly informed the content, design, and functionality of the intervention. Overall, participants reported positive experiences with the platform and described the content as useful and most participants expressed that they would have participated in such an intervention had it been available at the time of their treatment.

The qualitative analysis identified key themes that highlight the complex emotional and physical challenges faced by allo-SCT patients. These include the initial shock of a cancer diagnosis, the distress associated with hospitalization and treatment, and the ongoing struggle to regain normalcy posttransplant. Participants also shared coping strategies, underscoring the importance of social support, positive thinking, and maintaining an active lifestyle. The perceived usefulness of mindfulness and music therapy further validates the inclusion of these components in the eMBMT intervention.

The use of rapid qualitative analysis enabled the timely synthesis of participant feedback, which was essential for adapting the intervention to the unique needs of patients undergoing allo-SCT. Rapid qualitative analysis has been increasingly recognized as a valid and efficient method for evaluating eHealth interventions and has been successfully used in previous studies to inform digital health tools [67,68]. While a traditional qualitative approach might have provided additional granularity, the chosen method ensured that the findings were actionable within the project's timeline. We acknowledge that rapid qualitative analysis may result in themes that closely reflect the structure of the interview prompts. To address this, the research team used iterative coding and thematic refinement to ensure

XSL•FO
RenderX

the themes captured meaningful nuances and extended beyond the initial prompts, thereby providing deeper insights into participants' experiences and needs.

The feedback from usability and field-testing sessions was overwhelmingly positive, with participants rating the platform highly in terms of usefulness, ease of use, and satisfaction. However, several areas for improvement were identified, which were invaluable for further refining the intervention to better meet the needs of patients. Below we discuss the key areas of improvement and subsequent modifications made to the eMBMT content and intervention.

Enhance the Representativeness of the Content

More representative images of the patient population were included within the eMBMT platform content (eg, an image of a cancer patient in hospital robe listening to music, patients without hair, more racially and ethnically diverse).

Reduce the Amount of Text and Reading

We reduced the amount of text and provided an option to listen to content. All intervention content was professionally audio-recorded and integrated into the respective section of the platform.

Provide Headphones

All participants will now receive a pair of personal headphones for ease of listening to content.

Add More Guidance Throughout the Modules

Additional instructions guiding participants through the modules were added.

Modify the eMBMT Session Delivery Timeline

The eMBMT session delivery timeline was modified such that more sessions are provided pretransplant to aid with the emotional and psychological preparation for the allo-SCT. In addition, more sessions are provided postdischarge to help with reintegration into their day-to-day activities. Table 2 presents the sessions, content topics, and the delivery timeline.

Table . Overview of the 8-session eHealth-based mindfulness-based music therapy (eMBMT) intervention for allogeneic stem cell transplantation (allo-SCT) patients, detailing session content, mindfulness principles, and time relative to the transplant.

Session	SCT[a] treatment phase	Mindfulness attitude	MT[b] component	Attitude carry over after eMBMT
1	Admission	MBMT[c] Orientation	Breathing to music	Focus on the present
2	Conditioning	Nonjudging	Music listening and journaling	Unpleasant and pleasant experiences and writing
3	Early post transplant	Patience	Focused music and environmental sound listening, simple body movements to music	Awareness and experience of patience
4	Early post transplant	Letting go	Music-assisted relaxation and scripts	Discussion and journaling of letting go
5	Engraftment to discharge	Acceptance	Music-assisted relaxation and body scan	Nonjudgmental awareness and acceptance of physical sensations, thoughts, and feelings
6	Postdischarge	Trust	Vocal improvisation	Trusting oneself
7	Postdischarge	Beginner's mind	Novel instrument sounds	Discussion and journaling of beginner's mind
8	Postdischarge	Nonstriving	Music improvisation	Focus on the process, letting go of preconceived goals with what "should be"

[a]SCT: stem cell transplantation.

[b]MT: music therapy.

[c]MBMT: mindfulness-based music therapy.

Include Different Genres of Music and Preferred Music

Participants communicated the desire for a more varied genre of music and the ability to choose their preferred music. Thus, more diverse samples of music were included as well as allowing participants to choose their preferred music to complete eMBMT activities.

Provide a Journaling Section

Modules now provide sections within each module where participants can journal, write their thoughts, or take notes.

Streamline Blood Sample Collection

The blood sample collection was streamlined such that at baseline, blood is collected only after the patient has received their port and when the patient is scheduled for a follow-up appointment post-discharge.

XSL•FO
RenderX

Address Technical Issues

All technical issues (eg, audio was not functioning properly) were resolved. Furthermore, audio that was deemed of poor quality (ie, “white noise”) was professionally re-recorded.

Feedback from the future pilot randomized controlled trial will shed light on potential further modifications to the eMBMT content and intervention as well as provide examples of how participants successfully implement mindfulness attitudes in their own lives throughout their allo-SCT treatment.

Limitations

This study has several limitations. First, the sample size was small and primarily consisted White and Hispanic individuals, which may limit the generalizability of the findings to more diverse populations undergoing allo-SCT. Second, while the focus groups and usability and field testing provided valuable qualitative insights, the results may be influenced by selection bias, as participants who opted into the study may have been more motivated or receptive to mindfulness and music therapy interventions. Third, the rapid qualitative analysis approach, while efficient, may have limited the depth of thematic exploration compared to more traditional qualitative methodologies. Finally, while usability testing identified technical issues and design preferences, future iterations should consider additional real-world implementation challenges, including potential barriers to patient engagement with digital health interventions.

Conclusion

The development and refinement of the eMBMT intervention through focus groups and usability and field testing has resulted in an eHealth intervention for use with allo-SCT patients. The study findings underscore the importance of patient involvement in intervention design, ensuring that content is both relevant and engaging. As digital interventions become increasingly prevalent in health care, eMBMT can serve as a model for leveraging technology to deliver complementary, evidence-based psychosocial support. Future research should explore its long-term impact, scalability, and integration into standard transplant care to optimize patient outcomes and reduce health care burdens. A future pilot randomized controlled trial will evaluate the feasibility, acceptability, and preliminary efficacy of the eMBMT on health-related quality of life, symptom burden, disease activity, and psychosocial and physiological adaptation [69].

Acknowledgments

This study was supported by National Cancer Institute (NCI) 5R61CA263335-02 (FJP and TL were the principal investigators). SEFP and BNE are funded by the Ruth L Kirschstein NRSA Institution Research Training Grant (T32; 5T32CA251064-03) in Cancer Training in Disparities and Equity (C-TIDE). PIM is supported by an NCI career development award (K01CA258955).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to them containing information that could compromise patient privacy but are available from the corresponding author on reasonable request.

Authors' Contributions

SEFP, FJP, TL, KK, TW, MA, MRL, EDW, TKS, MAK, and BNE contributed to the study’s conception and design. TL, MAK, BNE, SEFP, and AEB conducted the interviews. BNE, SEFP, and AEB conducted the data analysis. SEFP, BNE, PL, and AEB wrote the first draft of the manuscript. All authors read, provided suggestions, and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Cancer facts & figures 2024. American Cancer Society. 2024. URL: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/2024-cancer-facts-figures.html [accessed 2025-03-26]
2. Zhang C. Blood cancer. National Foundation for Cancer Research. 2024. URL: https://www.nfcr.org/cancer-types/cancer-types-blood-cancer/ [accessed 2025-03-26]
3. Gyurkocza B, Rezvani A, Storb RF. Allogeneic hematopoietic cell transplantation: the state of the art. Expert Rev Hematol 2010 Jun;3(3):285-299. [doi: 10.1586/ehm.10.21] [Medline: 20871781]
4. Bevans MF, Mitchell SA, Marden S. The symptom experience in the first 100 days following allogeneic hematopoietic stem cell transplantation (HSCT). Support Care Cancer 2008 Nov;16(11):1243-1254. [doi: 10.1007/s00520-008-0420-6] [Medline: 18322708]
5. Mosher CE, Redd WH, Rini CM, Burkhalter JE, DuHamel KN. Physical, psychological, and social sequelae following hematopoietic stem cell transplantation: a review of the literature. Psychooncology 2009 Feb;18(2):113-127. [doi: 10.1002/pon.1399] [Medline: 18677717]
6. Chang G, Orav EJ, McNamara TK, Tong MY, Antin JH. Psychosocial function after hematopoietic stem cell transplantation. Psychosomatics 2005;46(1):34-40. [doi: 10.1176/appi.psy.46.1.34] [Medline: 15765819]

XSL•FO
RenderX

7. Malard F, Holler E, Sandmaier BM, Huang H, Mohty M. Acute graft-versus-host disease. Nat Rev Dis Primers 2023 Jun 8;9(1):27. [doi: 10.1038/s41572-023-00438-1] [Medline: 37291149]
8. Prieto JM, Atala J, Blanch J, et al. Patient-rated emotional and physical functioning among hematologic cancer patients during hospitalization for stem-cell transplantation. Bone Marrow Transplant 2005 Feb;35(3):307-314. [doi: 10.1038/sj.bmt.1704788] [Medline: 15580279]
9. Prieto JM, Atala J, Blanch J, et al. Role of depression as a predictor of mortality among cancer patients after stem-cell transplantation. JCO 2005 Sep 1;23(25):6063-6071. [doi: 10.1200/JCO.2005.05.751]
10. Park JE, Kim KI, Yoon SS, et al. Psychological distress as a negative survival factor for patients with hematologic malignancies who underwent allogeneic hematopoietic stem cell transplantation. Pharmacotherapy 2010 Dec;30(12):1239-1246. [doi: 10.1592/phco.30.12.1239] [Medline: 21114391]
11. Majhail NS, Omondi NA, Denzen E, Murphy EA, Rizzo JD. Access to hematopoietic cell transplantation in the United States. Biol Blood Marrow Transplant 2010 Aug;16(8):1070-1075. [doi: 10.1016/j.bbmt.2009.12.529] [Medline: 20036337]
12. Antoni MH, Dhabhar FS. The impact of psychosocial stress and stress management on immune responses in patients with cancer. Cancer 2019 May 1;125(9):1417-1431. [doi: 10.1002/cncr.31943] [Medline: 30768779]
13. Vogelzangs N, de Jonge P, Smit JH, Bahn S, Penninx BW. Cytokine production capacity in depression and anxiety. Transl Psychiatry 2016 May 31;6(5):e825. [doi: 10.1038/tp.2016.92] [Medline: 27244234]
14. Knight JM, Lyness JM, Sahler OJZ, Liesveld JL, Moynihan JA. Psychosocial factors and hematopoietic stem cell transplantation: potential biobehavioral pathways. Psychoneuroendocrinology 2013 Nov;38(11):2383-2393. [doi: 10.1016/j.psyneuen.2013.06.016] [Medline: 23845514]
15. Amonoo HL, Brown LA, Scheu CF, et al. Positive psychological experiences in allogeneic hematopoietic stem cell transplantation. Psychooncology 2019 Aug;28(8):1633-1639. [doi: 10.1002/pon.5128] [Medline: 31128072]
16. McGregor BA, Syrjala KL, Dolan ED, Langer SL, Redman M. The effect of pre-transplant distress on immune reconstitution among adult autologous hematopoietic cell transplantation patients. Brain Behav Immun 2013 Mar;30 Suppl(Suppl):S142-S148. [doi: 10.1016/j.bbi.2012.07.020] [Medline: 22910186]
17. Lee SJ, Loberiza FR, Rizzo JD, Soiffer RJ, Antin JH, Weeks JC. Optimistic expectations and survival after hematopoietic stem cell transplantation. Biol Blood Marrow Transplant 2003 Jun;9(6):389-396. [doi: 10.1016/s1083-8791(03)00103-4] [Medline: 12813447]
18. Duncan M, Moschopoulou E, Herrington E, et al. Review of systematic reviews of non-pharmacological interventions to improve quality of life in cancer survivors. BMJ Open 2017 Nov 28;7(11):e015860. [doi: 10.1136/bmjopen-2017-015860] [Medline: 29187408]
19. Cramer H, Lauche R, Klose P, Lange S, Langhorst J, Dobos GJ. Yoga for improving health-related quality of life, mental health and cancer-related symptoms in women diagnosed with breast cancer. Cochrane Database Syst Rev 2017 Jan 3;1(1):CD010802. [doi: 10.1002/14651858.CD010802.pub2] [Medline: 28045199]
20. Oberoi S, Yang J, Woodgate RL, et al. Association of mindfulness-based interventions with anxiety severity in adults with cancer: a systematic review and meta-analysis. JAMA Netw Open 2020 Aug 3;3(8):e2012598. [doi: 10.1001/jamanetworkopen.2020.12598] [Medline: 32766801]
21. Antoni MH, Moreno PI, Penedo FJ. Stress management interventions to facilitate psychological and physiological adaptation and optimal health outcomes in cancer patients and survivors. Annu Rev Psychol 2023 Jan 18;74:423-455. [doi: 10.1146/annurev-psych-030122-124119] [Medline: 35961041]
22. Crane RS, Brewer J, Feldman C, et al. What defines mindfulness-based programs? The warp and the weft. Psychol Med 2017 Apr;47(6):990-999. [doi: 10.1017/S0033291716003317] [Medline: 28031068]
23. Elkins G, Fisher W, Johnson A. Mind-body therapies in integrative oncology. Curr Treat Options Oncol 2010 Dec;11(3-4):128-140. [doi: 10.1007/s11864-010-0129-x] [Medline: 21116746]
24. Johns SA, Brown LF, Beck-Coon K, Monahan PO, Tong Y, Kroenke K. Randomized controlled pilot study of mindfulness-based stress reduction for persistently fatigued cancer survivors. Psychooncology 2015 Aug;24(8):885-893. [doi: 10.1002/pon.3648] [Medline: 25132206]
25. Lin LY, Lin LH, Tzeng GL, et al. Effects of mindfulness-based therapy for cancer patients: a systematic review and meta-analysis. J Clin Psychol Med Settings 2022 Jun;29(2):432-445. [doi: 10.1007/s10880-022-09862-z] [Medline: 35249176]
26. Liu H, Gao X, Hou Y. Effects of mindfulness-based stress reduction combined with music therapy on pain, anxiety, and sleep quality in patients with osteosarcoma. Braz J Psychiatry 2019;41(6):540-545. [doi: 10.1590/1516-4446-2018-0346] [Medline: 31116262]
27. Zhang Q, Zhao H, Zheng Y. Effectiveness of mindfulness-based stress reduction (MBSR) on symptom variables and health-related quality of life in breast cancer patients-a systematic review and meta-analysis. Support Care Cancer 2019 Mar;27(3):771-781. [doi: 10.1007/s00520-018-4570-x] [Medline: 30488223]
28. Cash E, Salmon P, Weissbecker I, et al. Mindfulness meditation alleviates fibromyalgia symptoms in women: results of a randomized clinical trial. Ann Behav Med 2015 Jun;49(3):319-330. [doi: 10.1007/s12160-014-9665-0] [Medline: 25425224]
29. Hilton L, Hempel S, Ewing BA, et al. Mindfulness meditation for chronic pain: systematic review and meta-analysis. Ann Behav Med 2017 Apr;51(2):199-213. [doi: 10.1007/s12160-016-9844-2] [Medline: 27658913]

XSL•FO
RenderX

30. Vieten C, Laraia BA, Kristeller J, et al. The mindful moms training: development of a mindfulness-based intervention to reduce stress and overeating during pregnancy. BMC Pregnancy Childbirth 2018 Jun 1;18(1):201. [doi: 10.1186/s12884-018-1757-6] [Medline: 29859038]
31. Victorson D, Kentor M, Maletich C, et al. Mindfulness meditation to promote wellness and manage chronic disease: a systematic review and meta-analysis of mindfulness-based randomized controlled trials relevant to lifestyle medicine. Am J Lifestyle Med 2014;9(3):185-211. [doi: 10.1177/1559827614537789]
32. van der Steen JT, van Soest-Poortvliet MC, van der Wouden JC, Bruinsma MS, Scholten RJ, Vink AC. Music-based therapeutic interventions for people with dementia. Cochrane Database Syst Rev 2017 May 2;5(5):CD003477. [doi: 10.1002/14651858.CD003477.pub3] [Medline: 28462986]
33. Lemieux AF, Fisher JD, Pratto F. A music-based HIV prevention intervention for urban adolescents. Health Psychol 2008 May;27(3):349-357. [doi: 10.1037/0278-6133.27.3.349] [Medline: 18624599]
34. Wachi M, Koyama M, Utsuyama M, Bittman BB, Kitagawa M, Hirokawa K. Recreational music-making modulates natural killer cell activity, cytokines, and mood states in corporate employees. Med Sci Monit 2007 Feb;13(2):CR57-CR70. [Medline: 17261984]
35. Archie P, Bruera E, Cohen L. Music-based interventions in palliative cancer care: a review of quantitative studies and neurobiological literature. Support Care Cancer 2013 Sep;21(9):2609-2624. [doi: 10.1007/s00520-013-1841-4] [Medline: 23715815]
36. Fancourt D, Ockelford A, Belai A. The psychoneuroimmunological effects of music: a systematic review and a new model. Brain Behav Immun 2014 Feb;36(15-26):15-26. [doi: 10.1016/j.bbi.2013.10.014] [Medline: 24157429]
37. Bittman B, Poornima I, Smith MA, Heidel RE. Gospel music: a catalyst for retention, engagement, and positive health outcomes for African Americans in a cardiovascular prevention and treatment program. Adv Mind Body Med 2020;34(1):8-16. [Medline: 32277749]
38. Magee WL, Clark I, Tamplin J, Bradt J. Music interventions for acquired brain injury. Cochrane Database Syst Rev 2017 Jan 20;1(1):CD006787. [doi: 10.1002/14651858.CD006787.pub3] [Medline: 28103638]
39. Lee JH. The effects of music on pain: a meta-analysis. J Music Ther 2016;53(4):430-477. [doi: 10.1093/jmt/thw012] [Medline: 27760797]
40. Dileo C. Effects of music and music therapy on medical patients: a meta-analysis of the research and implications for the future. J Soc Integr Oncol 2006;4(2):67-70. [doi: 10.2310/7200.2006.002] [Medline: 19442338]
41. Leggieri M, Thaut MH, Fornazzari L, et al. Music intervention approaches for Alzheimer's disease: a review of the literature. Front Neurosci 2019;13:132. [doi: 10.3389/fnins.2019.00132] [Medline: 30930728]
42. Haase JE, Robb SL, Burns DS, et al. Adolescent/young adult perspectives of a therapeutic music video intervention to improve resilience during hematopoietic stem cell transplant for cancer. J Music Ther 2020 Feb 25;57(1):3-33. [doi: 10.1093/jmt/thz014] [Medline: 31802124]
43. Robb SL, Burns DS, Stegenga KA, et al. Randomized clinical trial of therapeutic music video intervention for resilience outcomes in adolescents/young adults undergoing hematopoietic stem cell transplant: a report from the Children's Oncology Group. Cancer 2014 Mar 15;120(6):909-917. [doi: 10.1002/cncr.28355] [Medline: 24469862]
44. Russ KA, Holochwost SJ, Perkins SM, et al. Cortisol as an acute stress biomarker in young hematopoietic cell transplant patients/caregivers: active music engagement protocol. J Altern Complement Med 2020 May;26(5):424-434. [doi: 10.1089/acm.2019.0413] [Medline: 32073877]
45. Bradt J, Dileo C, Myers-Coffman K, Biondo J. Music interventions for improving psychological and physical outcomes in people with cancer. Cochrane Database Syst Rev 2021 Oct 12;10(10):CD006911. [doi: 10.1002/14651858.CD006911.pub4] [Medline: 34637527]
46. Gramaglia C, Gambaro E, Vecchi C, et al. Outcomes of music therapy interventions in cancer patients-a review of the literature. Crit Rev Oncol Hematol 2019 Jun;138(241-254):241-254. [doi: 10.1016/j.critrevonc.2019.04.004] [Medline: 31121392]
47. Dóro CA, Neto JZ, Cunha R, Dóro MP. Music therapy improves the mood of patients undergoing hematopoietic stem cells transplantation (controlled randomized study). Support Care Cancer 2017 Mar;25(3):1013-1018. [doi: 10.1007/s00520-016-3529-z] [Medline: 27942868]
48. Bates D, Bolwell B, Majhail NS, et al. Music therapy for symptom management after autologous stem cell transplantation: results from a randomized study. Biol Blood Marrow Transplant 2017 Sep;23(9):1567-1572. [doi: 10.1016/j.bbmt.2017.05.015] [Medline: 28533058]
49. Cassileth BR, Vickers AJ, Magill LA. Music therapy for mood disturbance during hospitalization for autologous stem cell transplantation: a randomized controlled trial. Cancer 2003 Dec 15;98(12):2723-2729. [doi: 10.1002/cncr.11842] [Medline: 14669295]
50. Stanczyk MM. Music therapy in supportive cancer care. Rep Pract Oncol Radiother 2011 Jun 8;16(5):170-172. [doi: 10.1016/j.rpor.2011.04.005] [Medline: 24376975]
51. Gutgsell KJ, Schluchter M, Margevicius S, et al. Music therapy reduces pain in palliative care patients: a randomized controlled trial. J Pain Symptom Manage 2013 May;45(5):822-831. [doi: 10.1016/j.jpainsymman.2012.05.008] [Medline: 23017609]

XSL•FO
RenderX

52. Köhler F, Martin ZS, Hertrampf RS, et al. Music therapy in the psychosocial treatment of adult cancer patients: a systematic review and meta-analysis. Front Psychol 2020;11(2095):651. [doi: 10.3389/fpsyg.2020.00651] [Medline: 32373019]
53. Harper FWK, Heath AS, Moore TF, Kim S, Heath EI. Using music as a tool for distress reduction during cancer chemotherapy treatment. JCO Oncol Pract 2023 Dec;19(12):1133-1142. [doi: 10.1200/OP.22.00814] [Medline: 37433094]
54. Tsai HF, Chen YR, Chung MH, et al. Effectiveness of music intervention in ameliorating cancer patients' anxiety, depression, pain, and fatigue: a meta-analysis. Cancer Nurs 2014;37(6):E35-E50. [doi: 10.1097/NCC.0000000000000116] [Medline: 24662723]
55. Lee JH. Erratum to: the effects of music on pain: a meta-analysis. J Music Ther 2021 Aug 24;58(3):34286339. [doi: 10.1093/jmt/thab012] [Medline: 34286339]
56. Penedo FJ, Oswald LB, Kronenfeld JP, Garcia SF, Cella D, Yanez B. The increasing value of eHealth in the delivery of patient-centred cancer care. Lancet Oncol 2020 May;21(5):e240-e251. [doi: 10.1016/S1470-2045(20)30021-8] [Medline: 32359500]
57. Kabat-Zinn J. Full Catastrophe Living: Using the Wisdom of Your Body and Mind to Face Stress, Pain, and Illness, 15th edition: Bantam Dell; 2005.
58. Lesiuk T. The effect of mindfulness-based music therapy on attention and mood in women receiving adjuvant chemotherapy for breast cancer: a pilot study. Oncol Nurs Forum 2015 May;42(3):276-282. [doi: 10.1188/15.ONF.276-282] [Medline: 25901379]
59. Lesiuk T. The development of a mindfulness-based music therapy (MBMT) program for women receiving adjuvant chemotherapy for breast cancer. Healthcare (Basel) 2016 Aug 9;4(3):27517966. [doi: 10.3390/healthcare4030053] [Medline: 27517966]
60. Puccinelli M, Seay J, Otto A, et al. An adapted cognitive behavioral stress and self-management intervention for sexual minority men living with HIV and cancer using the SmartManage ehealth platform: protocol and study design. JMIR Res Protoc 2022 Jul 18;11(7):e37822. [doi: 10.2196/37822] [Medline: 35849435]
61. BrightOutcome. URL: https://www.brightoutcome.com/index.php [accessed 2023-11-09]
62. Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. J Am Geriatr Soc 1975 Oct;23(10):433-441. [doi: 10.1111/j.1532-5415.1975.tb00927.x] [Medline: 1159263]
63. Krueger RA, Casey MA. Focus Groups: A Practical Guide for Applied Research, 5th edition: Sage Publications; 2014.
64. McQuarrie EF, Krueger RA. Focus groups: a practical guide for applied research. J Mark Res 1989 Aug;26(3):371. [doi: 10.2307/3172912]
65. Lund AM. Measuring usability with the USE questionnaire. Usability Interface 2001;8(2):3-6 [FREE Full text]
66. Fonteyn ME, Kuipers B, Grobe SJ. A description of think aloud method and protocol analysis. Qual Health Res 1993 Nov;3(4):430-441. [doi: 10.1177/104973239300300403]
67. Watkins DC. Rapid and rigorous qualitative data analysis: the "RADaR" technique for applied research. Int J Qual Methods 2017;16:1. [doi: 10.1177/1609406917712131]
68. St George SM, Harkness AR, Rodriguez-Diaz CE, Weinstein ER, Pavia V, Hamilton AB. Applying rapid qualitative analysis for health equity: lessons learned using "EARS" with Latino communities. Int J Qual Methods 2023;22:38463016. [doi: 10.1177/16094069231164938] [Medline: 38463016]
69. Fleszar-Pavlovic SE, Esquives BN, Brito AE, et al. eHealth mindfulness-based music therapy for patients undergoing allogeneic hematopoietic stem cell transplantation: A pilot randomized controlled trial protocol. Contemp Clin Trials 2024 Jul;142(107577). [doi: 10.1016/j.cct.2024.107577] [Medline: 38763308]

Abbreviations

allo-SCT: allogeneic stem cell transplantation
eMBMT: eHealth-based mindfulness-based music therapy
GVHD: graft-versus-host disease
HIPAA: Health Insurance Portability and Accountability Act
HRQoL: health-related quality of life
***ICD*:** *International Classification of Diseases*
MBMT: mindfulness-based music therapy
MBSR: mindfulness-based stress reduction
MM: mindful meditation
MT: music therapy
MT-BC: board-certified music therapist
RQA: rapid qualitative analysis
USE: Usefulness, Satisfaction, and Ease of Use

XSL•FO
RenderX

Edited by A Mavragani; submitted 08.08.24; peer-reviewed by MT Knobf, R Knoerl; revised version received 12.02.25; accepted 13.02.25; published 11.04.25.

Please cite as:
Fleszar-Pavlovic SE, Noriega Esquives B, Lovan P, Brito AE, Sia AM, Kauffman MA, Lopes M, Moreno PI, Koru-Sengul T, Gong R, Wang T, Wieder ED, Rueda-Lara M, Antoni M, Komanduri K, Lesiuk T, Penedo FJ
Development of an eHealth Mindfulness-Based Music Therapy Intervention for Adults Undergoing Allogeneic Hematopoietic Stem Cell Transplantation: Qualitative Study
JMIR Form Res 2025;9:e65188
URL: https://formative.jmir.org/2025/1/e65188
doi:10.2196/65188

© Sara E Fleszar-Pavlovic, Blanca Noriega Esquives, Padideh Lovan, Arianna E Brito, Ann Marie Sia, Mary Adelyn Kauffman, Maria Lopes, Patricia I Moreno, Tulay Koru-Sengul, Rui Gong, Trent Wang, Eric D Wieder, Maria Rueda-Lara, Michael Antoni, Krishna Komanduri, Teresa Lesiuk, Frank J Penedo. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluating User Experience and Satisfaction in a Concussion Rehabilitation App: Usability Study

Michael G Hutchison[1,2,3], PhD; Alex P Di Battista[1,4], PhD; Kyla L Pyndiura[1,3], MSc

[1]Faculty of Kinesiology and Physical Education, University of Toronto, 100 Devonshire Pl, Toronto, ON, Canada

[2]David L. MacIntosh Sport Medicine Clinic, Faculty of Kinesiology & Physical Education, University of Toronto, Toronto, ON, Canada

[3]Centre for Sport-Related Concussion Research, Innovation, and Knowledge, University of Toronto, Toronto, ON, Canada

[4]Defence Research and Development Canada, Toronto Research Centre, Toronto, ON, Canada

Corresponding Author:
Michael G Hutchison, PhD
Faculty of Kinesiology and Physical Education, University of Toronto, 100 Devonshire Pl, Toronto, ON, Canada

Abstract

Background: Evidence-based guidelines support the use of structured exercise to facilitate concussion recovery. Despite the growing number of mobile health (mHealth) apps aimed at managing concussions, few focus on delivering exercise rehabilitation protocols. Therefore, a mobile app was developed to provide personalized rehabilitation programs based on evidence-based exercise principles designed to cater to individuals recovering from concussions.

Objective: This study aimed to evaluate the usability and user experience of a mobile app designed to deliver an evidence-based rehabilitation program to individuals recovering from concussions.

Methods: A two-week prospective single-arm pilot study was conducted among adults with a physician-diagnosed concussion. Participants engaged in home-based rehabilitation exercises through a mobile app. Usability was assessed using a combination of the mHealth App Usability Questionnaire (MAUQ) and five custom questions evaluating confidence in recommendations, exercise flow, clarity of voice commands, and usability of the exercise report feature. Following the two-week period, participants rated each question on a 7-point Likert scale ranging from “strongly agree” to “strongly disagree”.

Results: Twenty-six participants consented and were enrolled in the study, with 23 participants (82%) completing all study components at the end of two weeks. The majority of participants were women, aged 26-38 years, and on average, approximately at three months postconcussion. Responses to both the MAUQ and custom questions were overwhelmingly positive. Overall, seven MAUQ questions received 100% positive responses, with no single question scoring below 83% positive responses. In the “ease of use and satisfaction” category, 100% of users responded positively to questions on ease of learning, usability, interface likeness, and comfort in social settings, while 83%-96% (19-22) of users responded positively to the remaining four questions. In the “system information arrangement” category, 100% (N=23) of users rated screen navigation, function usability, and health care service acceptability positively, with 87%-96% of users approving action acknowledgment, error recovery, and expected functions. Under the “usefulness” category, 96% (n=22) of users found the app beneficial for health and well-being, and 91% (n=21) users felt it effectively managed their health. For the five custom questions, 100% (N=23) users responded positively to voice command clarity, exercise awareness, ease of following exercises, and report understandability, with a single unfavorable response noted for confidence in app recommendations.

Conclusions: The findings of this study indicate that the mobile app is a user-friendly platform for delivering evidence-based exercise rehabilitation to individuals recovering from concussions. Positive user feedback, particularly in the areas of recommendation confidence, ease of exercise flow, and clarity of voice commands, highlights the app’s potential to support concussion recovery. Future iterations of the app will aim to improve time efficiency and streamline error recovery processes to further enhance the user experience.

(JMIR Form Res 2025;9:e67275) doi:10.2196/67275

KEYWORDS

mild traumatic brain injury; recovery; mHealth; app; digital health; smartphone; eHealth; digital; technology; usability; concussion rehabilitation; brain injury; rehabilitation protocols; evidence-based exercise; single-arm pilot study; home-based rehabilitation; user-friendly; questionnaire; telehealth; telemedicine

XSL•FO
RenderX

Introduction

Concussion is a common form of traumatic brain injury (TBI) that can occur in a variety of settings such as sports, motor vehicle accidents, workplaces, and daily life. Symptoms range from physical and cognitive disruptions to emotional and sleep disturbances [1], underscoring the highly individualized nature of the injury. Following an initial 24‐ to 48-hour period of cognitive and physical rest, research has demonstrated the benefits of aerobic exercise in enhancing concussion recovery. This approach not only alleviates symptom severity but can also expedite the recovery process [2-4].

The recommendation of aerobic exercise as an intervention for concussion often involves establishing a personalized heart rate (HR) threshold set by trained professionals [5,6]. More recently, the determination of exercise intensity has evolved to include the use of age-predicted maximum HR formulas [3], offering a more accessible and individualized approach to concussion management. However, to enhance the impact of these advancements, it is crucial to integrate exercise protocols into clinical practices beyond sports and exercise medicine. This can ensure that a wider patient population benefits from the most effective, evidence-based strategies for concussion management, thus achieving a more comprehensive and equitable standard of care.

Mobile health (mHealth) has emerged as a widely accessible medium for individuals to obtain medical care. This technology is particularly beneficial for individuals in remote areas or those who face challenges in accessing health care facilities. Consequently, there has been a surge in the development of health care–focused mobile apps in recent years, demonstrating significant benefits. Notably, mHealth interventions in mental health have shown promise in reducing symptoms of depression and anxiety across diverse population groups [7,8]. Additionally, mHealth apps have been effectively used in various health care domains, including cardiovascular disease prevention and diabetes management, showcasing their versatility and impact on modern health care [9-12].

The application of mHealth is increasingly recognized for its potential benefits in managing TBI. Various mobile apps aimed at providing education about TBI, symptom tracking, daily reminders for patients with TBI, and aiding in the screening or assessment of TBI are now available [13-15]. Despite these advancements, a notable gap exists in apps specifically tailored for the management of TBIs [13]. Moreover, many of these apps have yet to achieve widespread user adoption; a significant concern is that many available apps lack evidence-based interventions, making it difficult for health care providers to recommend them and for patients to use them with confidence [13]. In concussion care, the primary focus of mobile apps has been on symptom reporting, monitoring activity levels, and concussion recognition targeted at parents, coaches, athletic trainers, and athletes [16-18]. However, current apps do not focus on interventions involving exercise or physical activity, indicating an area for further development in mHealth solutions for TBI and concussion management.

Hutchison et al [19] conducted a study evaluating the feasibility of a mobile app designed to deliver resistance exercises using minimal equipment for individuals recovering from concussions. The app provides an aerobic stimulus by guiding users to achieve a percentage of their age-predicted maximal HR through tailored exercise routines. This innovative tool includes instructional videos to assist users in their rehabilitation process. Initial results demonstrated that most participants were able to reach the target HR zones by following the app-based exercise videos [19]. Building on these preliminary findings, which focused on the app's ability to elicit the desired physiological responses [19], the present study aims to expand on this research by assessing the app's overall usability. Specifically, we sought to identify potential areas for improvement, gauge user satisfaction, and evaluate the app's effectiveness in supporting concussion recovery.

Methods

Study Design

The usability study consisted of a two-week, community-oriented, prospective single-arm observational pilot study involving adults aged 18 years and older with a physician-diagnosed concussion. Participants were engaged in an active, home-based rehabilitation program accessed through the mobile app. Individuals were recruited through word-of-mouth, flyers, and social media platforms (eg, X, formerly known as Twitter, Facebook) and were provided detailed information regarding study requirements and procedures. Participant enrollment was conducted remotely via Zoom, and individuals were screened by a research assistant to confirm inclusion criteria consisting of a concussion diagnosis by a physician, aged at least 18 years, and access to an iPhone. Furthermore, individuals with coexisting conditions such as musculoskeletal or soft-tissue injuries, vestibular disorders, or neurological conditions that could impede physical activity, were excluded from the study. Twenty-six individuals consented and enrolled in the study. Upon enrollment, participants were mailed an Apple Watch Series 6 to use with their iPhones and access to the Rhea Health Inc. mobile app. Ultimately, 23 participants completed the usability questionnaire after using the mobile app for two weeks.

Ethical Considerations

This study received ethical approval from the Health Sciences Research Ethics Board at the University of Toronto (protocol #39143). To ensure comprehensive understanding and informed decision-making, prospective participants were given an information sheet by a member of the research team. This sheet detailed the study's eligibility criteria, purpose, methodologies, and any potential risks involved. Before participants gave their consent, their eligibility was confirmed via email. A Zoom videoconference was then conducted to clarify any queries regarding the study and reinforce the voluntary nature of participation and the freedom to withdraw at any point without adverse consequences. Informed consent was secured through an electronic process in which participants reviewed and signed a digital document that outlined the study's purpose, procedures, and potential risks. This consent form was then submitted using

XSL•FO
RenderX

the REDCap (Research Electronic Data Capture) platform, hosted by the University of Toronto.

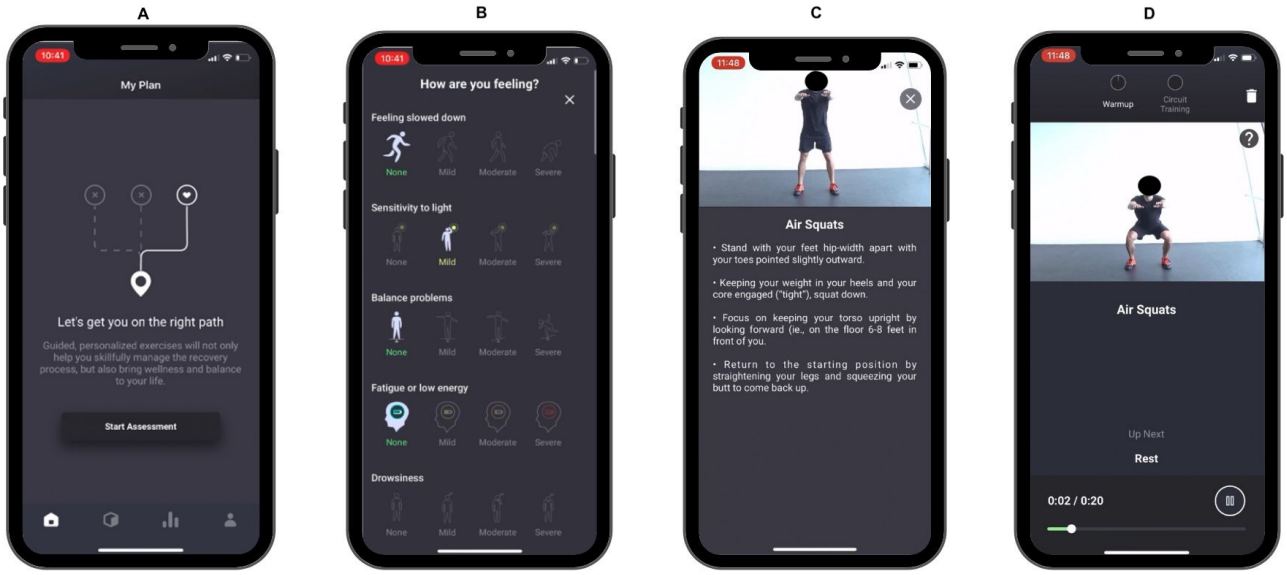
Data collected via the mobile app and Apple Watch were securely stored on the Amazon Web Services' DynamoDB database, located in Montreal, Quebec, Canada. This storage solution employed Amazon Web Services' proprietary encryption and restricted access system to ensure data security. Although participants were required to use their emails and passwords for accessing the mobile app and syncing data from their Apple Watch to Apple Health, personally identifiable information was stored in a separate database table and was not required for data analysis.

Upon the study's completion, participants were compensated CAD $50 (equivalent to approximately US $37.16) for their participation and were entered into a lottery for a chance to win an Apple Watch Series 6. This gesture not only acknowledged their contribution but also encouraged engagement and retention in the study.

The Mobile App

The rehabilitation protocol in this study was delivered through the Rhea mobile app (Rhea Health Inc.) (Figure 1), a rehabilitation platform that devises personalized programs based on an individual's symptoms and time elapsed since injury. The app integrates symptom data, including 22 items from the Sports Concussion Assessment Tool (SCAT-5) [20], along with seven additional questions related to vision and balance. Participants assessed their symptom severity using a 4-point Likert Scale, ranging from 0 (none) to 3 (severe). Following a staged approach endorsed by consensus statements on concussion management, the Rhea rehabilitation protocol employed a six-stage model that progressively increased in intensity and complexity. These stages extend from Initial Rest (Stage 1) to Return (Stage 6). Participants commenced their rehabilitation at Stage 2, completing three sessions at this level before progressing. One of the Rhea app's unique features is its dynamic adjustment of the rehabilitation plan that is based on participant feedback, which categorizes symptomology as "same," "better," or "worse." This feedback-driven mechanism ensures that the participants either stay at their current stage, advance to a more challenging stage, or revert to a less intensive stage, as required. The primary objective of this adaptive approach is to fine-tune the intensity and complexity of exercises to avoid exacerbating symptoms, thus promoting a safe and effective journey through the recovery stages. In addition to the six core stages, Rhea's program provided participants with exercises specifically tailored to neck, vision, and balance rehabilitation. These exercises are designed to address each participant's unique symptom profile, further personalizing the recovery experience. This approach not only ensures adherence to the rehabilitation protocol but also maximizes the potential for effective recovery tailored to individual needs. To monitor adverse events, participants were instructed to inform a member of the research team of any adverse events (eg, injuries, medical conditions) that were experienced while using the mobile app.

Figure 1. Screenshots of the Rhea mobile app from left to right display (A) the home page of a participant's plan, (B) a symptom checklist during onboarding that assists with exercise prescription, (C) an exercise preview screen providing description and cues for correct form, and (D) an exercise session screen with a timer and details of the next exercise for the user.



Usability Questionnaire

The mHealth App Usability Questionnaire (MAUQ) [21] contains 21 questions rated on a 7-point Likert scale, ranging from strongly agree to strongly disagree, under three categories: ease of use and satisfaction, system information arrangement, and usefulness. Studies have identified the MAUQ to be a reliable and valid tool for assessing mHealth app usability [21]. In this study, we used 16 questions from the MAUQ and omitted the 5 questions from the usefulness category, as they assess aspects related to sharing information with health care providers and clinical utility. Due to the preliminary nature of the research and the use of a research-grade tool that lacked standards and approval for clinical use, we did not ask or expect participants to engage with health care professionals using the tool. Additionally, we included five custom questions to evaluate the app's critical features related to the delivery of exercise rehabilitation, as the app was based on evidence-based exercise

XSL•FO
RenderX

rehab principles. These custom questions were designed to assess user confidence in the app's recommendations, clarity of exercise instructions, the flow of the exercise content, and the usability of the app's voice commands. More specifically, given the central role of accurate exercise prescription in rehabilitation, it was important to gauge whether users trusted the app's recommendations; therefore the question, "I had confidence in the recommendations I received from the app" was included. Since effective rehabilitation relies on the clarity and ease of exercise guidance, we included two questions targeting the flow and clarity of the exercise sessions: "The flow of the exercise sessions on the app is easy to follow" and "I knew exactly which exercises I had to do during an exercise session." These questions allowed us to determine whether participants found the session progression intuitive and well-structured. We also added a question about voice commands: "The in-app voice commands are clear and understandable." This was a critical feature for users who might benefit from auditory guidance, particularly those experiencing cognitive fatigue or visual sensitivity due to their concussion. Clear voice commands would support a hands-free interaction and reduce the cognitive load associated with reading instructions on a screen. Lastly, we evaluated the app's reporting functions through the question, "The exercise report section is easy to follow and understand." A clear and accessible exercise report is vital for tracking progress and ensuring participants understand their rehabilitation. These custom questions were specifically tailored to capture the nuances of delivering exercise rehabilitation through a digital platform, ensuring that the app was not only usable but also effective in guiding users through their recovery process.

Data Analysis

The purpose of the current study was to summarize and report participant feedback following a two-week period of using the mobile app. Therefore, the Likert-scale responses from the MAUQ and additional customized app usability questions were summarized and reported as raw counts and percentages. For additional continuous variables, the median (IQR) were calculated.

Results

Participant Characteristics

Study participant characteristics can be seen in Table 1. Briefly, the sample comprised predominantly of female participants (n=18, 78%), aged 26 - 38 years. On average, the participants were nearly three months post-concussion, with 48% (n=11) sustaining injury during sports or physical activity.

XSL•FO
RenderX

Table . Demographic and clinical characteristics of participants, including age, sex, mechanism of injury, and time since concussion.

Characteristics	Participants (N=23)
Sex, n (%)	
Female	18 (78)
Male	5 (22)
Age, median (IQR)	28.4 (25.1-37.3)
Days from injury, median (IQR)	82 (25-461)
Mechanism of injury, n (%)	
Sports/PA[a]	11 (48)
MVC[b]	1 (4.3)
Work	1 (4.3)
Fall	2 (8.7)
Other	8 (35)
History of concussion, n (%)	13 (59)
High performance athlete, n (%)	5 (22)
Exercise program, n (%)	8 (35)
Physical activity (times/week)	
0	3 (13)
1	2 (8.7)
2	4 (17)
3	3 (13)
4	2 (8.7)
5	4 (17)
6	3 (13)
≥7	2 (8.7)

[a]PA: physical activity.

[b]MVC: Motor vehicle collision.

Concussion Characteristics

Participants entered the study with an average of 14 symptoms and a symptom severity score of 29. On average, they reported feeling at approximately 70% of their usual health, and 43% (n=10) of the participants stated that they were getting better following their injury at first, but later experienced symptom worsening. Please see Table 2 for the complete symptom profiles of the participants.

Table . Symptom profile at onboarding, including symptom severity, perceived normalcy, and symptom progression trends.

Characteristics	Participants (N=23)
Symptom number, median (IQR)	14 (11-18)
Symptom severity, median (IQR)	29 (19-44)
Perceived percent normalcy, median (IQR)	70 (50-75)
Trends in symptoms since injury, n (%)	
No change	1 (4.3)
Steadily better	8 (35)
Steadily worse	1 (4.3)
Worse, then better	3 (13)
Better, then worse	10 (43)

XSL•FO
RenderX

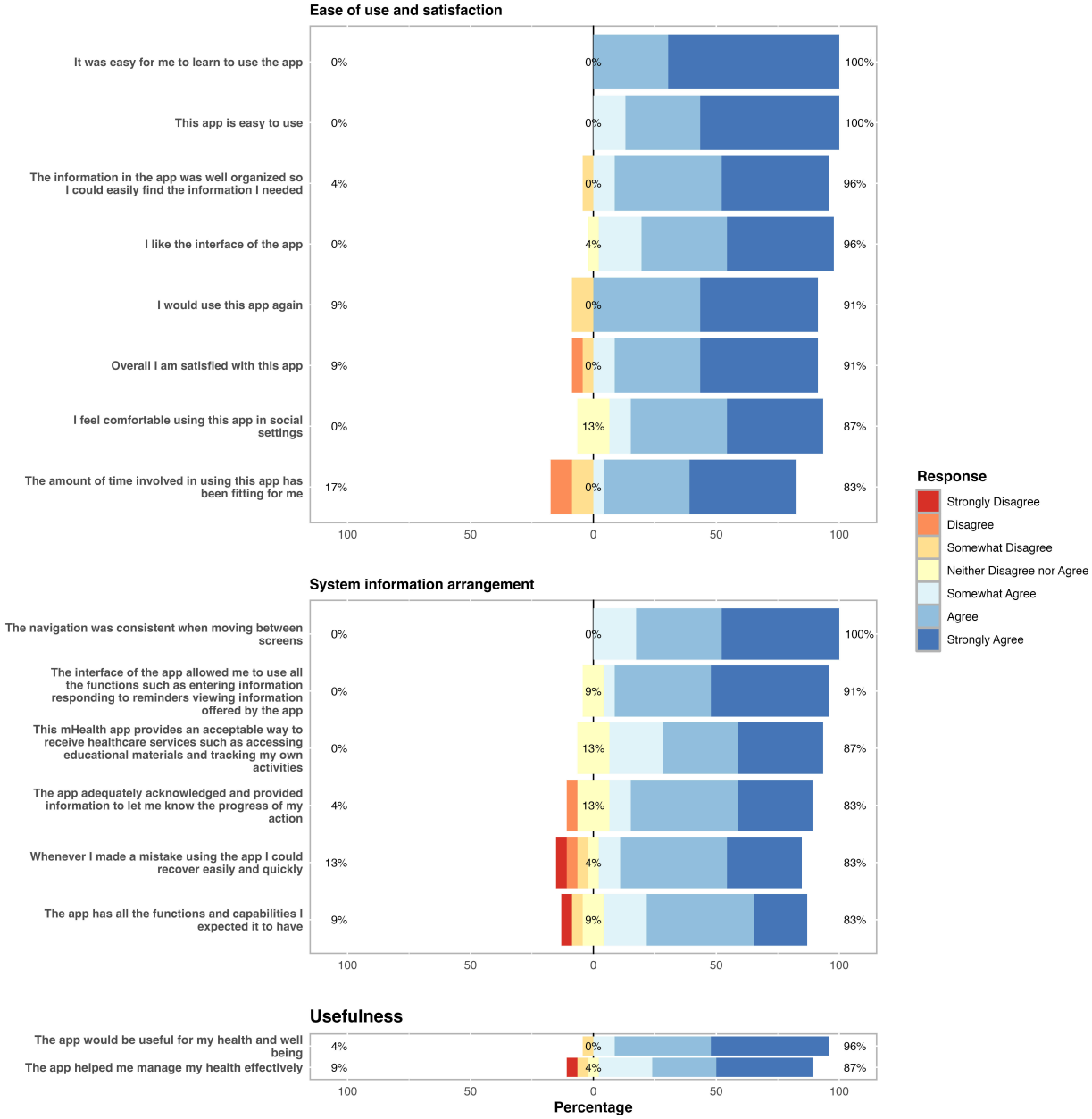
MAUQ Findings

Participants' responses to the MAUQ can be seen in Figure 2. Across the 16 MAUQ questions, "agree" or "strongly agree" were the most common responses. Under the "ease of use and satisfaction" category, all 23 participants selected "agree" or "strongly agree" for the question "easy to learn." Twenty participants (87%) selected "agree" or "strongly agree" for both "easy to use" and "well organized" questions. The question with the least favorable response in this category was "appropriate time involved," where 4 participants (17%) selected "disagree" or "somewhat disagree."

Under the category "system information arrangement," 20 (87%) participants selected "agree" or "strongly agree" for the question "all functions usable," and 19 (87%) for the question "screen navigation consistent." The most unfavorably responded question for this category was "ease of recovery following mistake," where 3 participants (13%) chose either "strongly disagree," "somewhat disagree," or "disagree."

In the "usefulness" category, 20 (87%) participants responded with "agree" or "strongly agree" for "health and well-being usefulness," and 21 participants (87%) reported similarly for "helped manage health." Please see Multimedia Appendix 1 for a summary of participant responses to each question of the MAUQ.

Figure 2. Responses from participants to each of the 16 mHealth app usability questionnaire (MAUQ). Participants ranked each question using a 7-point Likert scale ranging from strongly agree (dark blue) to strongly disagree (red). Each horizontal bar indicates the breakdown of each question's responses as percentages.

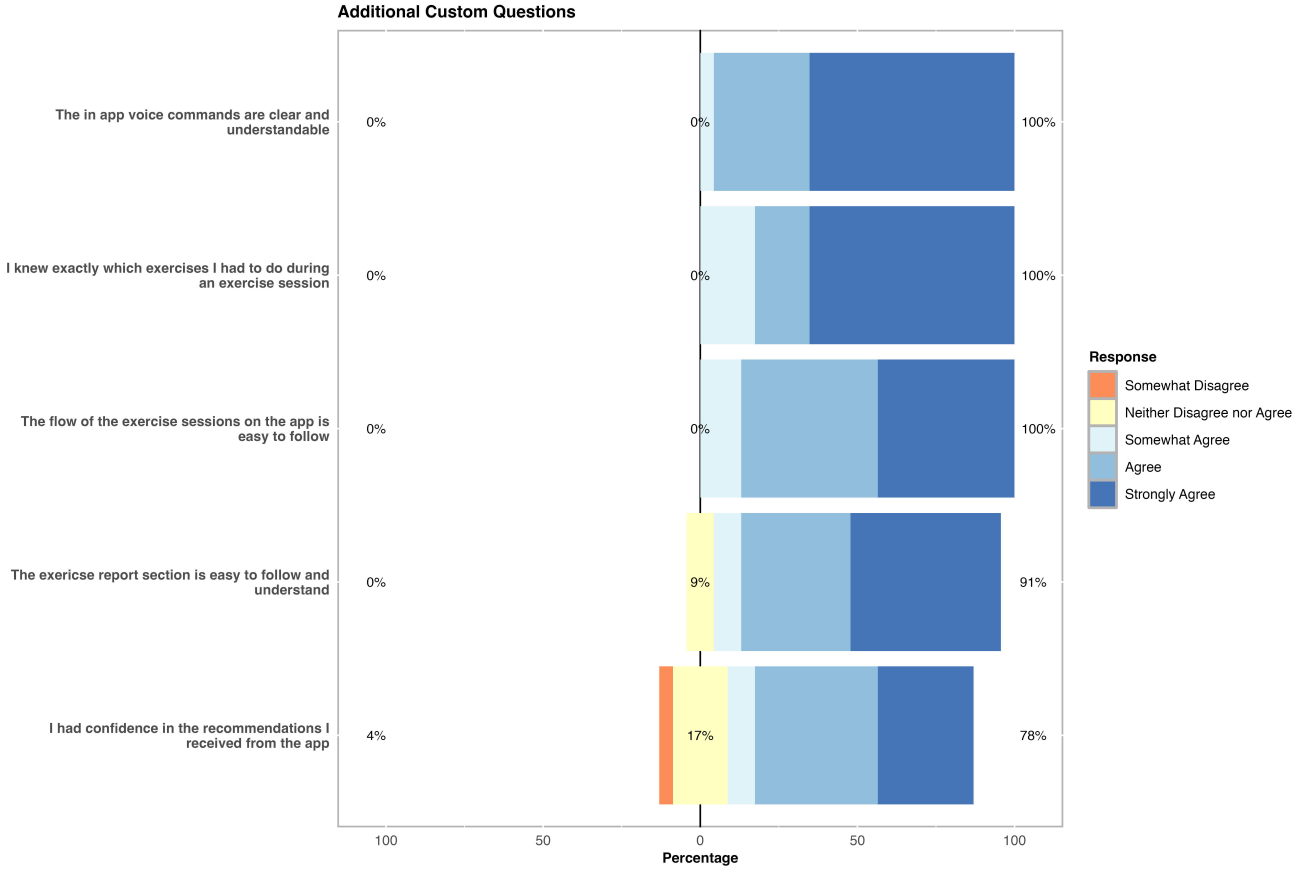


XSL•FO
RenderX

Additional Customized Questions

The additional questions specific to the app yielded similar responses to the MAUQ, with "agree" or "strongly agree" being the most common response for all five questions (Figure 3); only a single negative response was reported across all questions. The most favorable response was for "voice command clarity" where all but one participant (n=22, 95%) reported either "agree" or "strongly agree"; the other participant selected "somewhat agree." Please see Multimedia Appendix 2 for a summary of the participant responses for the additional questions.

Figure 3. Participants' responses to five additional custom questions. Participants ranked each question using a 7-point Likert scale ranging from strongly agree to strongly disagree; however, only 5 responses were chosen by participants, ranging from strongly agree (dark blue) to somewhat disagree (orange). Each horizontal bar indicates the breakdown of each question's responses as percentages.



Discussion

This study aimed to assess the usability of a mobile app that provides exercise rehabilitation for individuals with persistent concussion symptoms. Usability was assessed using a combination of the MAUQ and additional custom questions. Overall, the responses were strongly positive, indicating that participants found the app both useful and easy to navigate.

The app's core features were well received by users, as indicated by the absence of disagreements for seven of the MAUQ questions. Responses to these questions were either neutral or positive among participants, which strongly supports the overall user satisfaction with the app. The most favorable responses were primarily in the "ease of use and satisfaction" category, followed closely by the "system information arrangement" category. This positive feedback is particularly significant for individuals recovering from concussions, as symptoms such as cognitive complaints (eg, difficulty concentrating) and sensitivity to screens and light are commonly reported [22-24]. Providing a mobile app that is easy to navigate and well organized is essential for enhancing adoption and encouraging consistent use throughout the recovery process.

The strongly positive feedback based on the custom questions provides evidence that the mobile app effectively meets the needs of individuals recovering from concussions. Participants reported high confidence in the app's recommendations, which is particularly significant given that the app is grounded in evidence-based exercise rehabilitation principles. Confidence in the accuracy and safety of the recommendations is critical in concussion recovery, as missteps in physical activity intensity can lead to symptom exacerbation or worsening of neurocognitive performance [25]. The intuitive flow of exercise sessions and the clarity in guiding participants on which exercises to perform suggest that the app successfully translates complex rehabilitation protocols into an easy-to-follow format. Additionally, including clear voice commands helps reduce cognitive load, making the app accessible even to those with cognitive fatigue or visual sensitivity. Furthermore, the ease of understanding the exercise report section implies that users could track their progress effectively, promoting engagement and adherence to the rehabilitation protocol. These findings underscore the importance of mHealth solutions that provide

XSL•FO
RenderX

evidence-based care and inspire user confidence, ultimately enhancing the recovery experience for concussion patients.

While participants generally found the app easy to use, our study revealed concerns regarding the time commitment and error recovery processes, both of which require further attention. Specifically, some participants found the app's required time commitment to be excessive, and others experienced difficulty with the app's response to user errors. To address these concerns, future developments of the app will aim to find an optimal balance between screen time and cognitive load. This balance is crucial for supporting individuals with concussions, who may be particularly sensitive to long periods of screen interaction and cognitive stress. Enhancements will include more efficient navigation and task completion processes, as well as integrated educational features that prompt users to take necessary breaks and provide guidance on managing screen time effectively. Moreover, improving how the app handles mistakes—making error recovery more intuitive and less stressful—is essential for fostering a supportive and rehabilitative environment. These adjustments will not only enhance the user experience but also play a pivotal role in the app's functionality, ensuring it meets the unique needs of its users during their recovery.

There are several limitations to this study. First, the sample comprised predominantly female participants, which may limit the generalizability of the findings to male participants. Future studies should aim for more balanced demographics to validate the app's usability across diverse populations. Additionally, our sample consisted of individuals with persistent symptoms, leaving the usability of the app in the acute and subacute phases of concussion recovery still unexplored. Larger-scale studies with a more diverse population are necessary to further confirm the app's usability and effectiveness across different recovery stages. It is also important to note that we intentionally excluded five questions from the original MAUQ due to the early iteration of this tool. Specifically, we chose not to assess aspects related to sharing information with health care providers and clinical utility, as practitioners were not expected to engage with a research-grade tool at this stage. As the app evolves, future research should include these additional questions to better understand how the app complements health care practitioners and facilitates integration into clinical practice.

In conclusion, the Rhea app represents a promising mHealth solution for concussion rehabilitation. The overwhelmingly positive usability feedback supports the potential for broader adoption. However, continued refinement based on user feedback and further research with larger, more diverse populations is essential to ensure the app's long-term success in facilitating concussion recovery.

Acknowledgments

This research received funding from the Connaught Fund Innovation Award, the University of Toronto and the XSeed Funding Program, Faculty of Applied Science and Engineering, University of Toronto.

Data Availability

The datasets generated or analyzed during this study are available in the public GitHub repository [26].

Conflicts of Interest

Rhea Health Inc is a mobile health tech start-up developed by the University of Toronto, and MGH and the University of Toronto own equity shares in Rhea Health Inc. APDB is an unpaid data analytic lead for Rhea Health Inc.

Multimedia Appendix 1
Summary of participants' responses to MAUQ questions.
[DOCX File, 18 KB - formative_v9i1e67275_app1.docx]

Multimedia Appendix 2
Summary of participants' responses to additional custom questions.
[DOCX File, 16 KB - formative_v9i1e67275_app2.docx]

References

1. Patricios JS, Schneider KJ, Dvorak J, et al. Consensus statement on concussion in sport: the 6th International Conference on Concussion in Sport-Amsterdam, October 2022. Br J Sports Med 2023 Jun;57(11):695-711. [doi: 10.1136/bjsports-2023-106898] [Medline: 37316210]
2. Lawrence DW, Richards D, Comper P, Hutchison MG. Earlier time to aerobic exercise is associated with faster recovery following acute sport concussion. PLoS ONE 2018;13(4):e0196062. [doi: 10.1371/journal.pone.0196062] [Medline: 29668716]
3. Hutchison MG, Di Battista AP, Lawrence DW, Pyndiura K, Corallo D, Richards D. Randomized controlled trial of early aerobic exercise following sport-related concussion: progressive percentage of age-predicted maximal heart rate versus usual care. PLoS ONE 2022;17(12):e0276336. [doi: 10.1371/journal.pone.0276336] [Medline: 36548338]

XSL•FO
RenderX

4. Micay R, Richards D, Hutchison MG. Feasibility of a postacute structured aerobic exercise intervention following sport concussion in symptomatic adolescents: a randomised controlled study. BMJ Open Sport Exerc Med 2018;4(1):e000404. [doi: 10.1136/bmjsem-2018-000404] [Medline: 30018795]
5. Leddy JJ, Hinds AL, Miecznikowski J, et al. Safety and prognostic utility of provocative exercise testing in acutely concussed adolescents: a randomized trial. Clin J Sport Med 2018 Jan;28(1):13-20. [doi: 10.1097/JSM.0000000000000431] [Medline: 29257777]
6. Leddy J, Baker JG, Haider MN, Hinds A, Willer B. A physiological approach to prolonged recovery from sport-related concussion. J Athl Train 2017 Mar;52(3):299-308. [doi: 10.4085/1062-6050-51.11.08] [Medline: 28387557]
7. Mohr DC, Kwasny MJ, Meyerhoff J, Graham AK, Lattie EG. The effect of depression and anxiety symptom severity on clinical outcomes and app use in digital mental health treatments: meta-regression of three trials. Behav Res Ther 2021 Dec;147:103972. [doi: 10.1016/j.brat.2021.103972] [Medline: 34600398]
8. Bendtsen M, Müssener U, Linderoth C, Thomas K. A mobile health intervention for mental health promotion among university students: randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 20;8(3):e17208. [doi: 10.2196/17208] [Medline: 32196462]
9. Burke LE, Ma J, Azar KMJ, et al. Current science on consumer use of mobile health for cardiovascular disease prevention: a scientific statement from the American Heart Association. Circulation 2015 Sep 22;132(12):1157-1213. [doi: 10.1161/CIR.0000000000000232] [Medline: 26271892]
10. Klimis H, Thakkar J, Chow CK. Breaking barriers: Mobile health interventions for cardiovascular disease. Can J Cardiol 2018 Jul;34(7):905-913. [doi: 10.1016/j.cjca.2018.02.012] [Medline: 29887218]
11. Gandhi S, Chen S, Hong L, et al. Effect of mobile health interventions on the secondary prevention of cardiovascular disease: systematic review and meta-analysis. Can J Cardiol 2017 Feb;33(2):219-231. [doi: 10.1016/j.cjca.2016.08.017] [Medline: 27956043]
12. de Ridder M, Kim J, Jing Y, Khadra M, Nanan R. A systematic review on incentive-driven mobile health technology: as used in diabetes management. J Telemed Telecare 2017 Jan;23(1):26-35. [doi: 10.1177/1357633X15625539] [Medline: 26888421]
13. Christopher E, Alsaffarini KW, Jamjoom AA. Mobile health for traumatic brain injury: a systematic review of the literature and mobile application market. Cureus 2019 Jul 10;11(7):e5120. [doi: 10.7759/cureus.5120] [Medline: 31523551]
14. Juengst SB, Hart T, Sander AM, Nalder EJ, Pappadis MR. Mobile health interventions for traumatic brain injuries. Curr Phys Med Rehabil Rep 2019 Dec;7(4):341-356. [doi: 10.1007/s40141-019-00240-9]
15. Schmidt M, Babcock L, Kurowski BG, Cassedy A, Sidol C, Wade SL. Usage patterns of an mHealth symptom monitoring app among adolescents with acute mild traumatic brain injuries. J Head Trauma Rehabil 2022;37(3):134-143. [doi: 10.1097/HTR.0000000000000768] [Medline: 35125434]
16. Huber DL, Thomas DG, Danduran M, Meier TB, McCrea MA, Nelson LD. Quantifying activity levels after sport-related concussion using actigraph and mobile (mHealth) technologies. J Athl Train 2019 Sep;54(9):929-938. [doi: 10.4085/1062-6050-93-18] [Medline: 31411898]
17. Lee H, Sullivan SJ, Schneiders AG, et al. Smartphone and tablet apps for concussion road warriors (team clinicians): a systematic review for practical users. Br J Sports Med 2015 Apr;49(8):499-505. [doi: 10.1136/bjsports-2013-092930] [Medline: 24668048]
18. Worthen-Chaudhari L, McGonigal J, Logan K, Bockbrader MA, Yeates KO, Mysiw WJ. Reducing concussion symptoms among teenage youth: evaluation of a mobile health app. Brain Inj 2017;31(10):1279-1286. [doi: 10.1080/02699052.2017.1332388] [Medline: 28665690]
19. Hutchison MG, Di Battista AP, Loenhart MM. A continuous aerobic resistance exercise protocol for concussion rehabilitation delivered remotely via a mobile app: feasibility study. JMIR Form Res 2023 Jun 19;7:e45321. [doi: 10.2196/45321] [Medline: 37335605]
20. Echemendia RJ, Meeuwisse W, McCrory P, et al. The sport concussion assessment tool 5th Edition (SCAT5): background and rationale. Br J Sports Med 2017 Jun;51(11):848-850. [doi: 10.1136/bjsports-2017-097506] [Medline: 28446453]
21. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth app usability questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500. [doi: 10.2196/11500] [Medline: 30973342]
22. Stillman AM, Madigan N, Torres K, Swan N, Alexander MP. Subjective cognitive complaints in concussion. J Neurotrauma 2020 Jan 15;37(2):305-311. [doi: 10.1089/neu.2018.5925] [Medline: 31407632]
23. Mansur A, Hauer TM, Hussain MW, et al. A nonliquid crystal display screen computer for treatment of photosensitivity and computer screen intolerance in post-concussion syndrome. J Neurotrauma 2018 Aug 15;35(16):1886-1894. [doi: 10.1089/neu.2017.5539] [Medline: 30074876]
24. Abusamak M, Alrawashdeh HM. Post-concussion syndrome light sensitivity: a case report and review of the literature. Neuroophthalmology 2022;46(2):85-90. [doi: 10.1080/01658107.2021.1983612] [Medline: 35273410]
25. Majerske CW, Mihalik JP, Ren D, et al. Concussion in sports: postconcussive activity levels, symptoms, and neurocognitive performance. J Athl Train 2008;43(3):265-274. [doi: 10.4085/1062-6050-43.3.265] [Medline: 18523563]
26. Hutchison-et-al-2025-RHEA-usability. GitHub. 2025. URL: https://github.com/dibatti5/Hutchison-et-al-2025-RHEA-Usability [accessed 2025-04-03]

XSL•FO
RenderX

Abbreviations

HR: heart rate
MAUQ: mHealth app usability questionnaire
mHealth: mobile health
SCAT-5: sports concussion assessment tool
TBI: traumatic brain injury

Edited by A Mavragani; submitted 07.10.24; peer-reviewed by J Greenberg; revised version received 10.02.25; accepted 20.02.25; published 11.04.25.

Please cite as:
Hutchison MG, Di Battista AP, Pyndiura KL
Evaluating User Experience and Satisfaction in a Concussion Rehabilitation App: Usability Study
JMIR Form Res 2025;9:e67275
URL: https://formative.jmir.org/2025/1/e67275
doi:10.2196/67275

© Michael G Hutchison, Alex P Di Battista, Kyla L Pyndiura. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Oxidative Stress Markers and Prediction of Severity With a Machine Learning Approach in Hospitalized Patients With COVID-19 and Severe Lung Disease: Observational, Retrospective, Single-Center Feasibility Study

Olivier Raspado[1], Dr Med; Michel Brack[2], Dr Med; Olivier Brack[3]; Mélanie Vivancos[4]; Aurélie Esparcieux[1], Dr Med; Emmanuelle Cart-Tanneur[5], Dr Bio; Abdellah Aouifi[1], Dr Med

[1]Infirmerie Protestante, 1 Chemin du Penthod, Caluire-et-Cuire, France

[2]Oxidative Stress College, La Garenne-Colombes, France

[3]Statistique Industrielle Khi² Consulting (KSIC), Bayet, France

[4]Clinical Research and Innovation Department, Infirmerie Protestante, Caluire-et-Cuire, France

[5]Eurofins Biomnis Laboratory, Lyon, France

Corresponding Author:
Olivier Raspado, Dr Med
Infirmerie Protestante, 1 Chemin du Penthod, Caluire-et-Cuire, France

Abstract

Background: Serious pulmonary pathologies of infectious, viral, or bacterial origin are accompanied by inflammation and an increase in oxidative stress (OS). In these situations, biological measurements of OS are technically difficult to obtain, and their results are difficult to interpret. OS assays that do not require complex preanalytical methods, as well as machine learning methods for improving interpretation of the results, would be very useful tools for medical and care teams.

Objective: We aimed to identify relevant OS biomarkers associated with the severity of hospitalized patients' condition and identify possible correlations between OS biomarkers and the clinical status of hospitalized patients with COVID-19 and severe lung disease at the time of hospital admission.

Methods: All adult patients hospitalized with COVID-19 at the Infirmerie Protestante (Lyon, France) from February 9, 2022, to May 18, 2022, were included, regardless of the care service they used, during the respiratory infectious COVID-19 epidemic. We collected serous biomarkers from the patients (zinc [Zn], copper [Cu], Cu/Zn ratio, selenium, uric acid, high-sensitivity C-reactive protein [hs-CRP], oxidized low-density lipoprotein, glutathione peroxidase, glutathione reductase, and thiols), as well as demographic variables and comorbidities. A support vector machine (SVM) model was used to predict the severity of the patients' condition based on the collected data as a training set.

Results: A total of 28 patients were included: 8 were asymptomatic at admission (grade 0), 14 had mild to moderate symptoms (grade 1) and 6 had severe to critical symptoms (grade 3). As the first outcome, we found that 3 biomarkers of OS were associated with severity (Zn, Cu/Zn ratio, and thiols), especially between grades 0 and 1 and between grades 0 and 2. As a second outcome, we found that the SVM model could predict the level of severity based on a biological analysis of the level of OS, with only 7% misclassification on the training dataset. As an illustrative example, we simulated 3 different biological profiles (named A, B, and C) and submitted them to the SVM model. Profile B had significantly high Zn, low hs-CRP, a low Cu/Zn ratio, and high thiols, corresponding to grade 0. Profile C had low Zn, low selenium, high oxidized low-density lipoprotein, high glutathione peroxidase, a low Cu/Zn ratio, and low glutathione reductase, corresponding to grade 2.

Conclusions: The level of severity of pulmonary damage in patients hospitalized with COVID-19 was predicted using an SVM model; moderate to severe symptoms in patients were associated with low Zn, low plasma thiol, increased hs-CRP, and an increased Cu/Zn ratio among a panel of 10 biomarkers of OS. Since this panel does not require a complex preanalytical method, it can be used and studied in other pathologies associated with OS, such as infectious pathologies or chronic diseases.

(JMIR Form Res 2025;9:e66509) doi:10.2196/66509

KEYWORDS

oxidative stress; COVID-19; SARS-CoV-2; coronavirus; respiratory; infectious; pulmonary; respiration disorders; hospitalization; machine learning; ML; biomarker; lung; severity; prediction

XSL•FO
RenderX

Introduction

Oxidative stress (OS) is an imbalance between the production of free radicals and the body's ability to neutralize them, leading to damage to cells, proteins, and DNA [1-3]. This is a natural physiological process that is not harmful if it remains balanced, but it can be involved in the onset or worsening of chronic diseases such as cancer [4], metabolic disorders [5], atherosclerosis, [6] and cardiovascular diseases [7,8]. OS can also contribute to nerve cell death and deterioration of brain function, leading to neurodegenerative diseases [9,10].

Healthy lifestyle habits can help reduce OS and its harmful effects, for example, learning to manage emotional stress [11]; performing regular physical activity to promote the production of endogenous antioxidant enzymes [12]; and adopting a balanced diet that is able to provide antioxidants such as vitamins C and E, beta-carotene, and polyphenols and is rich in fruits, vegetables, whole grains, and lean protein sources [13]. Avoiding or reducing exposure to environmental toxins, such as toxic chemicals, cigarette smoke, and air pollutants, can also reduce OS [14].

Furthermore, many respiratory viral infections, including COVID-19, cause death of the infected cells, activation of the innate immune response, and secretion of inflammatory cytokines. All these processes are associated with the development of OS, which makes an important contribution to the pathogenesis of viral infections. COVID-19 is a complex disease in which interaction of the virus with target cells, action of the immune system, and the body's systemic response to these events are closely intertwined [15].

COVID-19 infection has various levels of pulmonary involvement, with asymptomatic, benign, serious, or even fatal forms, especially when the infected patient has comorbidities such as a history of cancer or malignant hematological diseases; concurrent chemotherapy treatment; severe chronic kidney disease, with or without dialysis; a history of solid organ transplantation or allogeneic hematopoietic stem cell transplantation; chronic polypathologies such as diabetes, high blood pressure, and obesity; or 2 or more organ failures [16]. All of these associated and chronic pathologies will influence the response of the immune system to COVID-19 infection and inflammation, as well as the level of OS present before and during the infection. Recent observational studies have reported that increasing COVID-19 severity may be responsible for a worse prognosis [17] among diabetes patients [18] and increased severity of lung disease [19]. Two meta-analyses showed that supplementation with antioxidants (in the form of vitamins and trace elements) in critically ill patients was associated with decreased mortality and counteracted OS damage [20,21].

The identification of OS biomarkers is essential for early diagnosis of chronic disease, for evaluating treatment efficacy, for monitoring lifestyle interventions, and for understanding the mechanisms involved. The decrease in the blood concentration of thiol proteins is one of the most relevant markers of OS [22]. However, clusters of evidence are emerging from the recent literature showing that levels of biomarkers of oxidative damage in biological fluids can be used for the prediction of measured concentrations of a limited number of exogenous and endogenous antioxidants [23].

Since March 2020, our health care and medical team began to hypothesize that antioxidant supplementation for patients in the intensive care unit (ICU) could improve their prognosis. We were heavily involved in treatment on a daily basis and tried to understand this serious and complex respiratory pathology and treat it as effectively as possible. Our patients benefit from a precise clinical assessment and comprehensive blood biological assays during their initial care and during their hospital stay. Other medical teams have been able to establish scores or panels composed of different blood biological assays to evaluate certain markers of OS and the inflammatory response [16,18], but the measurement and analysis of these biomarkers require complex preanalytical and analytical processes that cannot be carried out during hospitalization in the acute phase, that is, at the patient's bedside. To identify relevant OS biomarkers, we designed the OXYCOVID study as a feasibility study with the primary objective of testing whether the methodological protocol would be appropriate and feasible in the context of the ICU and ICU admission. The secondary objectives of this study were to identify relevant OS biomarkers associated with the severity of hospitalized patients and to identify whether there is a correlation between the clinical status of hospitalized patients with COVID-19 and severe lung disease at admission using artificial intelligence tools, such as a support vector machines (SVMs). SVMs are used to solve classification problems (eg, category assignment), discrimination problems, and regression problems. They require only a few parameters. The performance of the SVM family of machine learning algorithms is generally as good as, or even better than, neural networks or Gaussian models. These biological assays, if they can be used to determine an OS level, could become a routine element of care, allowing a more precise assessment of the severity of lung lesions in the event of COVID-19 infection.

Methods

Study Design

The OXYCOVID study is an observational, retrospective, single-center study based on medical data collected during hospitalization. All adult patients (ie, aged >18 years) hospitalized with COVID-19 at the Infirmerie Protestante (Lyon, France) were included, regardless of the care service they used. Patients were not included if they did not consent to the collection and processing of their personal data, if they were pregnant or breastfeeding, or if they had a protected status (ie, curatorship or other legal protection; deprivation of liberty by judicial or administrative decision). The data were collected from the period from February 9, 2022, to May 18, 2022. The final sample for analysis included 28 patients meeting the inclusion and exclusion criteria. All biological data were recorded in each patient's digital medical record.

Data Collection

The data were first collected and then sorted specifically for each patient after checking for the absence of errors and for missing data. After verification, no discrepancies were found and the clinical and biological data were judged to be usable.

XSL•FO
RenderX

The data are fully anonymized. The data included sociodemographics (sex and age), comorbidities, and risk factors (weight, height, smoking status, hypertension, diabetes, chronic renal disease, severe chronic respiratory disease, autoimmune disease, immune deficiency, and cancer). COVID-19 vaccines were also recorded, as were data related to the current COVID-19 infection (date of the latest polymerase chain reaction test from a nasopharyngeal swab, variant, and date of first symptoms) and the reasons for hospitalization.

At admission, clinical symptoms were recorded, including the presence of fever (temperature >38°C), cough (yes or no), dyspnea (felt by the patient during the initial questioning), myalgia (felt by the patient during the initial questioning), fatigue (felt by the patient during the initial questioning), diarrhea, oxyge saturation (SpO_2), oxygenation system used (nasal oxygen therapy, mask oxygen therapy, or Optiflow system), and pulmonary embolism (determined by a thoracic computed tomography angiogram). Biological data were taken from venous blood samples during hospitalization and included complete blood count; levels of urea, creatinine, C-reactive protein (CRP), troponin, D-dimer, ferritin, lactase dehydrogenase, and prothrombin; activated partial thromboplastin time; levels of total bilirubin, gamma-GT, transaminase, albumin, and interleukin-6; and an OxyCheck panel including the following 10 markers: zinc (Zn), copper (Cu), Cu/Zn ratio, selenium, uric acid, high sensitivity CRP (hs-CRP), oxidized low-density lipoprotein, glutathione peroxidase, glutathione reductase, and thiols.

Patients were triaged at admission into 3 mutually exclusive grades. Grade 0 represented asymptomatic patients. Grade 1 represented patients with mild to moderate clinical signs (SpO_2 >94% in ambient air or with oxygen [maximum 6 L/min], fever, cough, dyspnea, and respiratory rate between 20 and 30 respirations/min). Grade 2 represented the most severe patients, that is, those with severe pneumonia (respiratory rate >30 respirations/min; arterial oxygen saturation <90% despite fraction of inspired oxygen >40% [>6 L/min]) or a rapid increase in oxygen needs (tachycardia [heart rate >120 beats/min]), sweating, agitation, a need for noninvasive ventilatory support with Optiflow (flow rate <60 L/min at 60% fraction of inspired oxygen, ROX index [oxygen saturation/respiration rate index] >3). Grade 2 also included critical patients with severe pneumonia or acute respiratory distress syndrome (index ROX <3 [if the ROX index was <2.5, the patient was intubated]; Optiflow flow rate >60 L/min at 60% fraction of inspired oxygen [if the Optiflow flow rate was >80 L/min at 100% of fraction of inspired oxygen, the patient was intubated]; partial pressure of oxygen in the arterial blood/fraction of inspired oxygen ≤300 mm Hg) and patients with septic shock; acute, life-threatening organ dysfunction related to a known or suspected infection; an altered mental state (delirium or confusion); oliguria; skin mottling; laboratory evidence of coagulopathy, thrombocytopenia, acidosis, or elevated lactic acid; or hyperbilirubinemia.

Statistical Analysis

Baseline characteristics are reported as counts and percentages for categorical variables and medians and IQR for continuous variables. Each patient was assigned only one grade, so the observations are considered independent. We tested for significant outliers among all the biomarkers, determined if distributions followed a normal distribution using the Shapiro-Wilk test, and used the Levene test for homogeneity of variances. Due to the small sample size, the criteria for outlier values were not strictly applied in the ANOVA. Tukey post-hoc tests were used for multiple pairwise comparisons between severity groups when significant differences were observed in the ANOVA. Finally, an SVM was used to predict assignment for each grade, taking into account the entire analysis profile. Statistical analysis was performed with R (version 4.1.1; R Project for Statistical Computing) and the *tidyverse*, *ggpubr*, and *rstatix* libraries. The level of significance was set at $P<.05$.

Considering that collinearities within our dataset induced by obvious links between biological parameters disrupted traditional methods (eg, logistic regression or the classification method), we used an SVM to predict assignment to each grade for each patient, considering the entire analysis profile. The SVM algorithm is a supervised learning algorithm that uses training data to build a model to predict or classify new observations. SVMs are a machine learning method for nonlinear regression classification [24-26]. Our approach consisted in using the biomarker database as a training dataset for the SVM model (eg, the SVM training dataset). This resulted in a model capable of predicting the probability that a patient would be assigned a specific grade based on the results of a biological analysis, provided that the values did not exceed those used in the initial dataset; that is, the sum of the 3 probabilities was equal to 1. This method resulted in a profile in the form of a graph establishing the link between a combination of biomarkers and the probability of being assigned to the different severity grades. The SVM analysis was performed using JMP Pro (version 17.2; JMP Statistical Discovery).

Ethical Considerations

Participation in this study was voluntary, and all participants received an individual notice at admission to the ICU, which included informed consent (the individual notice is included in Multimedia Appendix 1). This study was based on data available in the medical record and did not require any additional medical visits or examinations; it also did not lead to any changes in the management or treatment of patients. The data were irreversibly deidentified for the statistical analysis. Participants did not receive any compensation. This study was approved by the ethics committee of Infirmerie Protestante of Lyon (CE-23‐01).

Results

Baseline Characteristics of Patients

A total of 28 patients met the inclusion criteria. The median age was 75 (IQR 68.75-80) years, and the patients were mostly male (n=17, 60.71%; Table 1). At admission, 8 patients were asymptomatic (grade 0), 14 had mild- to moderate-severity illness (grade 1) and 6 had severe to critical symptoms (grade 3). The most common reported symptoms were cough and fever.

XSL•FO
RenderX

Table . Baseline characteristics of the 28 patients hospitalized for COVID-19, including hospitalized patients; socio-demographics, vaccination status, clinical symptoms at admission, and comorbidities.

	All patients (N=28)	Asymptomatic patients (grade 0; n=8)	Mild to moderate patients (grade 1; n=14)	Severe to critical patients (grade 2; n=6)
Sociodemographics				
Sex ratio	1.54	0.6	1.8	5
Male, n (%)	17 (61)	3 (38)	9 (64)	5 (83)
Age (years), median (IQR)	75 (68.75-80)	71.5 (65.5-75.25)	75.5 (68.75-80.25)	71.0 (65.5-78.75)
COVID-19 vaccination, n (%)				
Vaccinated (primary series)	5 (18)	1 (12)	3 (21)	1 (17)
Unvaccinated	3 (11)	0 (0)	2 (14)	0 (0)
Unknown	20 (71)	7 (87)	9 (64)	5 (83)
Clinical symptoms at admission, n (%)				
Alteration of the general state	12 (43)	0 (0)	9 (64)	3 (50)
Fever (temperature >38°C)	12 (43)	0 (0)	8 (57)	4 (67)
Cough	10 (36)	0 (0)	9 (63)	1 (17)
Dyspnea (>20 respirations/min)	10 (36)	0 (0)	8 (57)	2 (33)
Myalgia	2 (7)	2 (25)	0 (0)	0 (0)
Diarrhea	3 (11)	0 (0)	3 (21)	0 (0)
Pulmonary embolism	1 (4)	0 (0)	1 (7)	0 (0)
Comorbidities, n (%)				
Diabetes	11 (39)	3 (37)	5 (36)	3 (50)
Hypertension	14 (50)	3 (37)	8 (57)	3 (50)
Cancer	13 (46)	1 (12)	9 (64)	3 (50)

Descriptive Analysis of the OS Biomarkers

Results for the detection of outliers, the normality test, and the homogeneity of variance are provided in Table 2. The OS biomarkers that met the criteria for the ANOVA were Zn (P=.002), Cu/Zn ratio (P=.009), hs-CRP (P=.02), and thiols (P=.004).

XSL•FO
RenderX

Table . Verification of the conditions for the ANOVA (outlier detection, verification of the normality of values between groups, and verification of homogeneity of variance between groups) and results of a 1-way ANOVA.

	Outlier detection, n			Shapiro-Wilk test [a]			Levene test[b]	ANOVA
	Grade 0	Grade 1	Grade 2	Grade 0	Grade 1	Grade 2		
Zinc	1	0	0	.97	.97	.66	.35	0.002
Copper	0	0	0	.98	.31	.80	.78	0.343
Copper/zinc ratio	0	1	0	.68	.008	.24	.52	0.009
Selenium	0	0	0	.50	.46	.045	.63	0.303
Uric acid	1	1	0	.02	.19	.60	.89	0.528
High-sensitivity C-reactive protein	0	0	0	<.001	.04	.72	.08	0.017
Oxidized low-density lipoprotein	1	0	0	.18	.70	.33	.55	0.872
Glutathione peroxydase	0	0	1	.18	.51	.03	.99	0.380
Glutathione reductase	0	0	0	.23	.88	.19	.03	0.194
Thiols	1	0	0	.34	.76	.07	.63	0.004

[a]If the data are normally distributed, the *P* value is greater than .05.

[b]If the variances are homogeneous, the *P* value is greater than .05.

Statistical differences were observed between the grade 0 and grade 1 patients and between the grade 0 and grade 2 patients for the 3 OS biomarkers, but the difference was nonsignificant between the grade 1 and 2 patients (Table 3).

Table . Comparison between the severity grades depending on the 3 OS biomarkers using the Tukey post-hoc test.

	Grade 0 versus grade 1	Grade 0 versus grade 2	Grade 1 versus grade 2
Zinc	0.005	0.003	0.609
Copper/zinc ratio	0.012	0.029	0.981
Thiols	0.004	0.019	1

Predictive Model Using SVM

Table 4 shows the results obtained from the SVM (Multimedia Appendix 2 provides detailed results). The clinical classification and the SVM prediction differed in only 2 of the 28 cases. One of these cases was a patient who was clinically classified as grade 1; he was symptomatic but did not meet the criteria for intubation. The biological analysis profile was abnormal (low thiols [4.1], low zinc, very high Cu/Zn ratio [2.2]). The SVM model classified this patient as grade 2 with a probability of 70.7%. He remained hospitalized for 1 month with active therapy, made good progress, and was discharged. The other patient was initially classified as grade 2; he was strongly symptomatic with high dependence on high-flow oxygen therapy. The biological analysis profile was abnormal (Cu/Zn ratio=3.57). The SVM model classified this case as grade 1 with a probability of 65.8%, probably due to slightly decreased plasma thiol and zinc levels. The clinical course of this patient quickly became favorable.

With a misclassification rate of only 7%, we estimated that the training of the SVM model was good enough and that it was able to predict the grade of patients with COVID-19 infection based on the results of the initial specific biological assays (upon arrival).

Table . Comparison of the clinically observed grade and the most likely grade (eg, a higher probability of belonging to the grade) predicted by the support vector machine (SVM).

Clinically observed grade	Most likely grade predicted by SVM model (patients), n		
	Grade 0	Grade 1	Grade 2
Grade 0	8	0	0
Grade 1	0	13	1
Grade 2	0	1	5

XSL•FO
RenderX

Profiles of Biological Results as Illustrative Examples

As an illustrative example, we created 3 biological result profiles (named A, B, and C). These profiles were submitted to the SVM model to determine the probability that they belonged to grade 0, 1, or 2 (Table 5). Profile B had a significantly high Zn level, low hs-CRP, a low Cu/Zn ratio, and high thiols and was predicted to be grade 0. At the opposite end, profile C had low Zn, low selenium, high oxidized LDL, high glutathione peroxidase, a low Cu/Zn ratio, and low glutathione reductase, corresponding to grade 2. Figure 1, Figure 2, and Figure 3 illustrate the predicted probabilities per biomarker and per profile for profiles A to C.

Table . Examples of 3 different biological profiles and the probability that they belonged to each grade of severity, as determined by the trained support vector machine (Multimedia Appendix 2 provides complete results).

Profile	Zinc (µg/L)	Copper (µg/L)	Selenium (µg/L)	Uric acid (µmol/L)	High-sensitivity C-reactive protein (mg/L)	Oxidized low-density lipoprotein (U/L)	Glutathione peroxidase (U/L)	Copper/zinc ratio	Glutathione reductase (U/g of hemoglobin)	Thiols (µmol/g of protein)	P[a] grade 0	P grade 1	P grade 2
A	584.68	1300.82	55.96	329.82	134.34	58.75	403.54	2.35	9.80	4.58	12.8%	73.9%	13.3%
B	850.00	850.00	70.00	350.00	2.00	30.00	400.00	1.00	9.00	7.00	99%	0.7%	0.3%
C	595.00	1313.00	75.00	266.00	206.00	91.00	600.00	2.20	9.70	4.10	5.9%	23.9%	70.2%

[a]P: probabililty.

Figure 1. Predicted probabilities of belonging to grade 0, 1, or 2 per biomarker for profile A. Values shown in red indicate the specific values for the profile. P: probability.

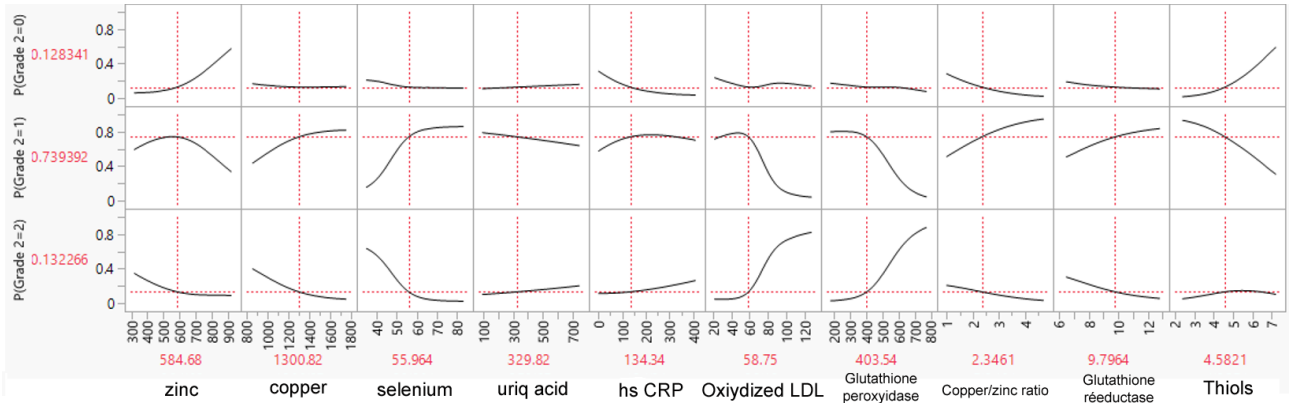
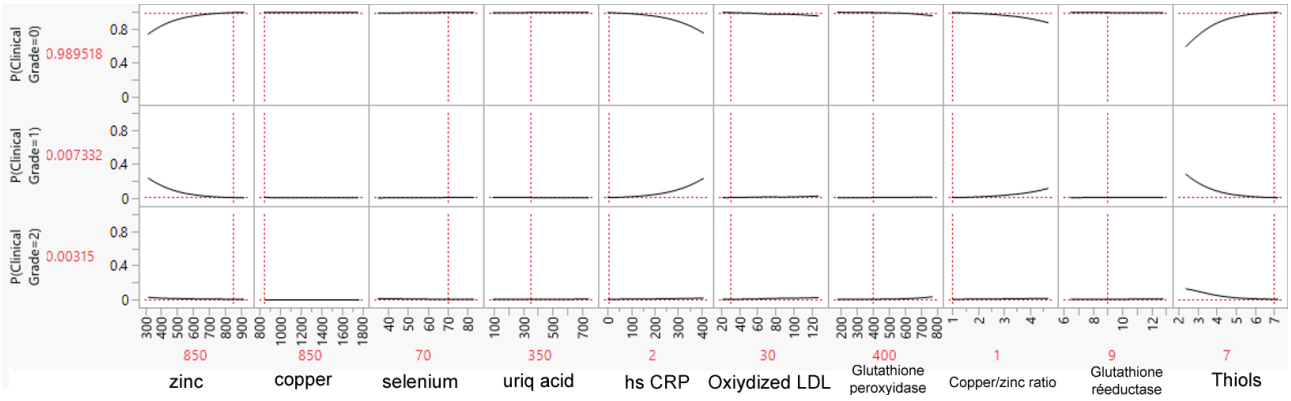
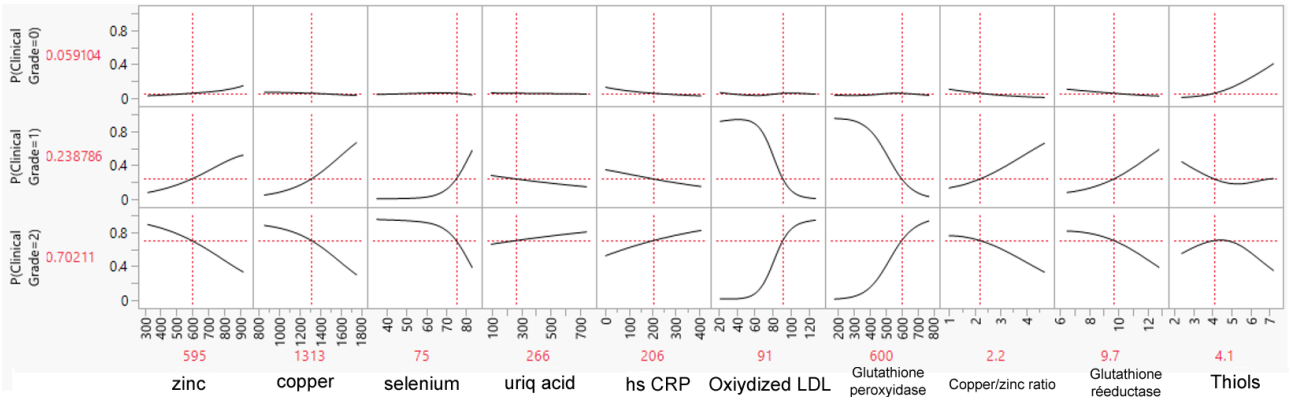


Figure 2. Predicted probabilities of belonging to grade 0, 1, or 2 per biomarker for profile B. Values shown in red indicate the specific values for the profile. P: probability.



XSL•FO
RenderX

Figure 3. Predicted probabilities of belonging to grade 0, 1, or 2 per biomarker for profile C. Values shown in red indicate the specific values for the profile P: probability.



Discussion

Main Findings

Initially, we assumed that antioxidant supplementation for patients in the ICU would reduce the severity of clinical signs and improve their prognosis. The main challenge of this feasibility study was performing a preanalytical and analytical process for measuring OS biomarkers. As a first observational finding, this study identified 3 biomarkers of OS (Zn, Cu/Zn ratio, and thiols) associated with the severity of patients, especially patients with a severity grade of 0 (ie, asymptomatic) or 1 (ie, mild to moderate severity). As a second analytical finding, we found that the SVM model could provide a prediction of the level of severity based on a biological analysis of the level of OS in a relatively limited cohort and during an epidemic of respiratory infectious disease, with only 7% misclassified cases in the training dataset.

The originality of this work is the determination of the grade of clinical severity through the analysis of specific OS biomarkers using a machine learning model. What distinguishes an SVM algorithm from other classical approaches is that when the response is continuous, the fitted models are called a support vector regression (SVR) model. In a typical regression problem, the goal is to fit a model that minimizes the error between a predicted response and the actual response. In an SVR problem, the goal is to fit a model such that the error between a predicted response and the actual response lies in a range of $-\varepsilon$ to ε. This allows for more flexible fitting. In JMP Pro, ε is equal to 0.1. The SVR algorithm doubles the data by creating 2 classes, $Y + \varepsilon$ and $Y - \varepsilon$. Then, the same algorithm used for the classification problem is also used for the prediction problem (ie, SVR).

Maximization in SVM algorithms is done by solving a quadratic programming problem. In JMP Pro, the algorithm used by the SVM platform is based on the sequential minimal optimization (SMO) algorithm introduced by John Platt in 1998 [27]. Typically, the SVM quadratic programming problem is very large. The SMO algorithm divides the overall quadratic programming problem into a series of smaller problems. The smaller problems are solved analytically rather than numerically, which means that they produce closed-form solutions. Therefore, the SMO algorithm is more efficient than solving the overall quadratic programming problem. While this univariate analysis did not find any differences between grade 1 and grade 2, the trained SVM model estimated the severity grade from the same OS biomarkers; this could have a significant impact in the management of patients at risk of sudden respiratory decompensation. Although the power of the model would benefit from being improved with clinical and biological data from a larger cohort, this approach could be applied to other types of epidemics and other biological assays, as well as to the investigation of other diseases.

Comparison With Prior Work

We know that moderate and severe forms of COVID-19 are associated with high OS and a low total antioxidant level, which is measured with complex preanalytical and analytical methods [28,29] that are not generally used in all laboratories, have long delays in receiving results, and are not compatible with management in ICUs [30]. Moderate and severe forms of COVID-19 are accompanied by an increase in inflammatory biomarkers, such as CRP, procalcitonin, the neutrophil to lymphocyte ratio, and hemostasis parameters (activated partial thromboplastin time, prothrombin time, D-dimers, and fibrinogens) [29]. The natural correction of low levels of endogenous antioxidants is slow, with a tendency to recover 3 months after hospital discharge [28].

High OS is not only specific to severe disease or infection, but positively influences their severity. A basic infectiology study with a porcine animal model of infection with the avian-type H1N1 European swine influenza virus has shown that prior infection of animals with mycoplasma generates a high level of OS, increasing the severity of influenza and reducing animal performance; host responses could be influenced by diet [31].

The systemic OS status, determined using many biomarkers, was still significantly increased in recovered COVID-19 patients during the recovery phase. Nonhospitalized individuals with COVID-19 presented signs of systemic OS, which is longitudinally associated with the development of post–COVID-19 condition. Regular assessment of plasma thiol levels as a monitoring biomarker and supplementation in cases of deficiency might be potential therapeutic targets at the onset of post–COVID-19 condition [32].

From the beginning of the epidemic, plasma thiol assays at admission were a promising tool to predict ICU admission in patients with COVID-19 [19]. The principles of nutrition for patients in intensive care follow validated recommendations

XSL•FO
RenderX

with a high level of evidence [33]. Faced with inflammatory syndrome and significant catabolism, caloric and protein intake must be sufficient to combat muscle wasting and the risk of infection. In products intended for parenteral or enteral nutrition, specific micronutrients or macronutrients intended to modulate inflammatory processes and optimize the body's immunological or metabolic responses are added, such as the antioxidants β-carotene, vitamin A, vitamin C, or vitamin E, in order to combat OS [34], which plays a central role in the development of major inflammatory states and visceral failure through the production of reactive oxygen species (free radicals), which influence the consumption, distribution, and level of antioxidants [35]. Lowered levels of antioxidants are common and are associated with an increase in morbidity and mortality [36]. Two meta-analyses showed that supplementation with antioxidants (eg, vitamins and trace elements) among ICU patients is associated with a reduction in mortality [20,21] and a reduction in mechanical ventilation time [20], but two more recent meta-analyses showed that the evidence for these benefits remained questionable because they were measured in small study populations and with highly variable cocktails of antioxidant nutrients [23,37]. Currently, it is not recommended to provide massive doses of antioxidants to patients treated in intensive care.

Strengths and Limitations

The main limitation of this study is the sample size, with less than 10 patients in the grade 0 and grade 2 groups. Despite the small number of patients, we observed a statistical difference between grade 0 and grade 1 patients in terms of biomarkers associated with severity. The small sample size was caused by a short study inclusion period due to the hospital context and the ongoing crisis, which was not suited to the inclusion of more patients or to the inclusion of other hospitals. The medical and nursing team has continued to collect relevant clinical and biological data concerning patients treated for COVID-19 infection since the end of the study. These new data could be aggregated with those collected during this study and analyzed with an SVM in order to improve the performance of the severity stage prediction. As a beneficial side effect, this study allowed us to test the ability of the clinical research team to study OS biomarkers and identify covariates and risk factors for further investigation after using propensity scoring to adjust for comorbidities and sociodemographic variables of patients.

As early as 2020, applying the SVM method to COVID-19 was reported in the scientific literature, but it was mainly used for diagnosis and prediction of severity in a radiological manner and as a scanner [26]. The prediction of severity stage by SVM has been extended to the analysis of biomarkers classically used in inflammatory and infectious pathologies [27,28] but has not provided a consensus or an infallible score to predict the level of severity. To our knowledge, this study is the first to show the SVM method being used for predicting the severity stage of COVID-19 with pulmonary involvement.

An important point to consider before generalizing the use of prediction scores based on this panel of blood biological OS analyses is the dosage method and standardization of reference values for each analysis and between each laboratory.

Conclusions

In patients with COVID-19 infection, moderate to severe symptoms were correlated with a lowered Zn level, a lowered plasma thiol level, an increased hs-CRP level, and an increased Cu/Zn ratio in a panel of 10 OS biomarkers not requiring complex preanalysis (Zn, Cu, Cu/Zn ratio, selenium, uric acid, hs-CRP, oxidized LDL, glutathione peroxidase, glutathione reductase, and thiols). Applying a machine learning method (an SVM) to the results of this same panel to evaluate the OS level in a retrospective cohort allowed prediction of the level of severity of pulmonary damage in hospitalized patients with COVID-19. Since this panel does not require a complex preanalytical method, it can be used and studied in other pathologies associated with OS, such as infectious pathologies or chronic diseases. Intensive care teams are accustomed to monitoring and managing major inflammatory symptoms, and although it is likely that they need to augment their practices to manage OS and the supply of antioxidants, the biological exploration of OS requires highly complex preanalytical and analytical processes that, to date, are highly limiting factors in the development and dissemination of relevant OS assessments. The fragility and thermolability of certain biomarkers means that any redox reactions that might take place in the sample must be halted; thus, the mandatory use of −80 °C conditions throughout the preanalytical phase makes analysis complex and in some cases even technically unfeasible. We have therefore developed an OS panel with 10 biomarkers that overcomes all these constraints without compromising scientific relevance by performing an assessment of OS based on a series of biomarkers that do not require complex preanalysis (simple freezing at −20 °C); these biomarkers were selected with a machine learning approach in 2009 [23]. The results from this panel showed real significance in estimating OS with robust and reproducible results. Iterative training of this SVM model with a biomarker panel in a larger cohort of patients with OS would give it more power and precision. By determining the ideal weighting of each of the biomarkers with a scientifically validated algorithm, we could obtain a score to estimate the level of OS. A Belgian intensive care team assessed total antioxidant capacity with an electrochemical method (eg, the Total Antioxidant Power technology) rather than by the analysis of biological samples (blood, saliva, or urine), finding that it had potential as a less expensive and much faster alternative to the individual analysis of biomarkers linked to pro-oxidants [30]. This would allow intensive care teams to monitor and respond in real time to systemic OS in their patients.

The World Health Organization has already announced the very probable risk of the appearance of a new epidemic caused by new types of severe acute respiratory syndrome or by the "humanization" of the avian influenza A (H5N1) virus in the years to come [38]. Given the contagiousness and the risk of decompensation, particularly respiratory decompensation, or even death from these viral infections, considering systemic OS in addition to the inflammatory syndrome could be a determining factor in the effectiveness of care in the future. The early and systematic use of machine learning models such as SVMs in clinical, radiological, and biological assessments at the start of an epidemic could provide a better understanding of the

XSL•FO
RenderX

functioning and mechanisms of novel pathogens and reduce human loss of life. Sharing of results between different care centers with dedicated databases would accelerate learning and improve the accuracy of these statistical tools.

From a public health perspective, individuals are subject to a range of exposures through the environment (air pollution, solar radiation, and temperature variations) and lifestyle (lack of sleep, psychological stress, tobacco, and poor diet); this is summarized in the term “exposome.” This directly impacts the OS level and therefore the body’s ability to defend itself in the event of infection. Taking action to maintain a healthy environment and lifestyle would help mitigate the severity of diseases such as COVID-19 [39].

Acknowledgments

This study was funded by the Infirmerie Protestante of Lyon. Preanalytical and analytical processes were performed by Biomnis. EpiComet SAS provided scientific writing assistance in the preparation and development of the manuscript in accordance with the Good Publication Practice guidelines.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Individual notice information.
[DOCX File, 51 KB - formative_v9i1e66509_app1.docx]

Multimedia Appendix 2
Support vector machine results from 28 patients, including observed oxidative stress biomarkers and the probability of patients being assigned severity grades 0, 1, or 2.
[DOCX File, 23 KB - formative_v9i1e66509_app2.docx]

References

1. Forman HJ, Zhang H. Targeting oxidative stress in disease: promise and limitations of antioxidant therapy. Nat Rev Drug Discov 2021 Sep;20(9):689-709. [doi: 10.1038/s41573-021-00233-1] [Medline: 34194012]
2. Pizzino G, Irrera N, Cucinotta M, et al. Oxidative stress: harms and benefits for human health. Oxid Med Cell Longev 2017;2017:8416763. [doi: 10.1155/2017/8416763] [Medline: 28819546]
3. Ji LL, Yeo D. Oxidative stress: an evolving definition. Fac Rev 2021;10:13. [doi: 10.12703/r/10-13] [Medline: 33659931]
4. Jelic MD, Mandic AD, Maricic SM, Srdjenovic BU. Oxidative stress and its role in cancer. J Cancer Res Ther 2021;17(1):22-28. [doi: 10.4103/jcrt.JCRT_862_16] [Medline: 33723127]
5. Rezzani R, Franco C. Liver, oxidative stress and metabolic syndromes. Nutrients 2021 Jan 21;13(2):301. [doi: 10.3390/nu13020301] [Medline: 33494242]
6. Li M, Xin S, Gu R, et al. Novel diagnostic biomarkers related to oxidative stress and macrophage ferroptosis in atherosclerosis. Oxid Med Cell Longev 2022;2022:8917947. [doi: 10.1155/2022/8917947] [Medline: 36035208]
7. Lorenzon Dos Santos J, Quadros AD, Weschenfelder C, Garofallo SB, Marcadenti A. Oxidative stress biomarkers, nut-related antioxidants, and cardiovascular disease. Nutrients 2020 Mar 3;12(3):682. [doi: 10.3390/nu12030682] [Medline: 32138220]
8. Gtif I, Abdelhedi R, Ouarda W, et al. Oxidative stress markers-driven prognostic model to predict post-discharge mortality in heart failure with reduced ejection fraction. Front Cardiovasc Med 2022;9:1017673. [doi: 10.3389/fcvm.2022.1017673] [Medline: 36419488]
9. Ton AMM, Campagnaro BP, Alves GA, et al. Oxidative stress and dementia in Alzheimer’s patients: effects of synbiotic supplementation. Oxid Med Cell Longev 2020;2020:2638703. [doi: 10.1155/2020/2638703] [Medline: 32411323]
10. Ionescu-Tucker A, Cotman CW. Emerging roles of oxidative stress in brain aging and Alzheimer’s disease. Neurobiol Aging 2021 Nov;107:86-95. [doi: 10.1016/j.neurobiolaging.2021.07.014] [Medline: 34416493]
11. Salim S. Oxidative stress: a potential link between emotional wellbeing and immune response. Curr Opin Pharmacol 2016 Aug;29:70-76. [doi: 10.1016/j.coph.2016.06.006] [Medline: 27400336]
12. El Assar M, Álvarez-Bustos A, Sosa P, Angulo J, Rodríguez-Mañas L. Effect of physical activity/exercise on oxidative stress and inflammation in muscle and vascular aging. Int J Mol Sci 2022 Aug 5;23(15):8713. [doi: 10.3390/ijms23158713] [Medline: 35955849]

XSL•FO
RenderX

13. Ávila-Escalante ML, Coop-Gamas F, Cervantes-Rodríguez M, Méndez-Iturbide D, Aranda-González II. The effect of diet on oxidative stress and metabolic diseases-clinically controlled trials. J Food Biochem 2020 May;44(5):e13191. [doi: 10.1111/jfbc.13191] [Medline: 32160647]
14. Gangwar RS, Bevan GH, Palanivel R, Das L, Rajagopalan S. Oxidative stress pathways of air pollution mediated toxicity: recent insights. Redox Biol 2020 Jul;34:101545. [doi: 10.1016/j.redox.2020.101545] [Medline: 32505541]
15. Chernyak BV, Popova EN, Prikhodko AS, Grebenchikov OA, Zinovkina LA, Zinovkin RA. COVID-19 and oxidative stress. Biochemistry (Mosc) 2020 Dec;85(12):1543-1553. [doi: 10.1134/S0006297920120068] [Medline: 33705292]
16. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506. [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
17. Suhail S, Zajac J, Fossum C, et al. Role of oxidative stress on SARS-CoV (SARS) and SARS-CoV-2 (COVID-19) infection: a review. Protein J 2020 Dec;39(6):644-656. [doi: 10.1007/s10930-020-09935-8] [Medline: 33106987]
18. Paul AK, Hossain MK, Mahboob T, et al. Does oxidative stress management help alleviation of COVID-19 symptoms in patients experiencing diabetes? Nutrients 2022 Jan 13;14(2):321. [doi: 10.3390/nu14020321] [Medline: 35057501]
19. Ducastel M, Chenevier-Gobeaux C, Ballaa Y, et al. Oxidative stress and inflammatory biomarkers for the prediction of severity and ICU admission in unselected patients hospitalized with COVID-19. Int J Mol Sci 2021 Jul 12;22(14):7462. [doi: 10.3390/ijms22147462] [Medline: 34299080]
20. Manzanares W, Dhaliwal R, Jiang X, Murch L, Heyland DK. Antioxidant micronutrients in the critically ill: a systematic review and meta-analysis. Crit Care 2012 Dec 12;16(2):R66. [doi: 10.1186/cc11316] [Medline: 22534505]
21. Heyland DK, Dhaliwal R, Suchner U, Berger MM. Antioxidant nutrients: a systematic review of trace elements and vitamins in the critically ill patient. Intensive Care Med 2005 Mar;31(3):327-337. [doi: 10.1007/s00134-004-2522-z] [Medline: 15605227]
22. Kalem AK, Kayaaslan B, Neselioglu S, et al. A useful and sensitive marker in the prediction of COVID-19 and disease severity: thiol. Free Radic Biol Med 2021 Apr;166:11-17. [doi: 10.1016/j.freeradbiomed.2021.02.009] [Medline: 33588050]
23. de la Villehuchet AM, Brack M, Dreyfus G, et al. A machine-learning approach to the prediction of oxidative stress in chronic inflammatory disease. Redox Rep 2009;14(1):23-33. [doi: 10.1179/135100009X392449] [Medline: 19161675]
24. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995 Sep;20(3):273-297. [doi: 10.1007/BF00994018]
25. Vapnik VN. The support vector method. In: Gerstner W, Germond A, Hasler M, Nicoud JD, editors. Artificial Neural Networks — ICANN'97 ICANN 1997 Lecture Notes in Computer Science, Vol 1327: Springer; 1997. [doi: 10.1007/BFb0020166]
26. Hsu CW, Chang CC, Lin CJ. A practical guide to support vector classification. National Taiwan University. URL: https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf [accessed 2025-03-27]
27. Platt JC. Fast training of support vector machines using sequential minimal optimization. In: Burges CJC, Schölkopf B, Smola AJ, editors. Advances in Kernel Methods: Support Vector Learning: MIT Press; 1998:41-65. [doi: 10.7551/mitpress/1130.003.0016]
28. Jindal M, Lehl SS, Jaswal S, Bhardwaj N, Gupta M. Total antioxidant status and oxidative stress in patients with COVID-19 infection. J Assoc Physicians India 2024 Aug;72(8):36-39. [doi: 10.59556/japi.72.0575] [Medline: 39163060]
29. Djordjevic J, Ignjatovic V, Vukomanovic V, et al. Laboratory puzzle of oxidative stress, parameters of hemostasis and inflammation in hospitalized patients with COVID-19. Biomedicines 2024 Mar 13;12(3):636. [doi: 10.3390/biomedicines12030636] [Medline: 38540248]
30. Pincemail J, Rousseau AF, Kaux JF, et al. A pilot study on oxidative stress during the recovery phase in critical COVID-19 patients in A rehabilitation facility: potential utility of the PAOT® technology for assessing total anti-oxidative capacity. Biomedicines 2023 Apr 28;11(5):1308. [doi: 10.3390/biomedicines11051308] [Medline: 37238982]
31. Deblanc C, Robert F, Pinard T, et al. Pre-infection of pigs with Mycoplasma hyopneumoniae induces oxidative stress that influences outcomes of a subsequent infection with a swine influenza virus of H1N1 subtype. Vet Microbiol 2013 Mar 23;162(2-4):643-651. [doi: 10.1016/j.vetmic.2012.11.028] [Medline: 23266108]
32. Vlaming-van Eijk LE, Bulthuis MLC, van der Gun BTF, et al. Systemic oxidative stress associates with the development of post-COVID-19 syndrome in non-hospitalized individuals. Redox Biol 2024 Oct;76:103310. [doi: 10.1016/j.redox.2024.103310] [Medline: 39163767]
33. Principes de nutrition en reanimation. Société française d'anesthésie et de réanimation. URL: https://sfar.org/wp-content/uploads/2018/10/14-Principes-de-nutrition-en-reanimation.pdf [accessed 2025-04-09]
34. Jones NE, Heyland DK. Pharmaconutrition: a new emerging paradigm. Curr Opin Gastroenterol 2008 Mar;24(2):215-222. [doi: 10.1097/MOG.0b013e3282f4cdd8] [Medline: 18301274]
35. Fuhrman MP. Antioxidant supplementation in critical illness: what do we know? Nutrition 2000 Jun;16(6):470-471. [doi: 10.1016/s0899-9007(00)00360-9] [Medline: 10869911]
36. Forceville X, Vitoux D, Gauzit R, Combes A, Lahilaire P, Chappuis P. Selenium, systemic immune response syndrome, sepsis, and outcome in critically ill patients. Crit Care Med 1998 Sep;26(9):1536-1544. [doi: 10.1097/00003246-199809000-00021] [Medline: 9751590]

XSL•FO
RenderX

37. Gudivada KK, Kumar A, Shariff M, et al. Antioxidant micronutrient supplementation in critically ill adults: a systematic review with meta-analysis and trial sequential analysis. Clin Nutr 2021 Mar;40(3):740-750. [doi: 10.1016/j.clnu.2020.06.033] [Medline: 32723509]
38. Avian influenza A(H5N1) - Cambodia. World Health Organization. URL: https://www.who.int/emergencies/disease-outbreak-news/item/2024-DON533 [accessed 2025-03-27]
39. Morales JS, Valenzuela PL, Castillo-García A, et al. The exposome and immune health in times of the COVID-19 pandemic. Nutrients 2021 Dec 22;14(1):24. [doi: 10.3390/nu14010024] [Medline: 35010900]

Abbreviations

CRP: C-reactive protein
Cu: copper
hs-CRP: high-sensitivity C-reactive protein
ICU: intensive care unit
OS: oxidative stress
ROX index: oxygen saturation/respiration rate index
SMO: sequential minimal optimization
SpO_2: oxygen saturation
SVM: support vector machine
SVR: support vector regression
Zn: zinc

Edited by A Mavragani; submitted 15.09.24; peer-reviewed by A Akomolafe, N Idowu; revised version received 19.02.25; accepted 20.02.25; published 11.04.25.

Please cite as:
Raspado O, Brack M, Brack O, Vivancos M, Esparcieux A, Cart-Tanneur E, Aouifi A
Oxidative Stress Markers and Prediction of Severity With a Machine Learning Approach in Hospitalized Patients With COVID-19 and Severe Lung Disease: Observational, Retrospective, Single-Center Feasibility Study
JMIR Form Res 2025;9:e66509
URL: https://formative.jmir.org/2025/1/e66509
doi:10.2196/66509

© Olivier Raspado, Michel Brack, Olivier Brack, Mélanie VivancosI, Aurélie Esparcieux, Emmanuelle Cart-Tanneur, Abdellah Aouifi. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Association of Social Media Recruitment and Depression Among Racially and Ethnically Diverse Metabolic and Bariatric Surgery Candidates: Prospective Cohort Study

Jackson M Francis[1], MPH; Sitapriya S Neti[2,3], MPH; Dhatri Polavarapu[2,3], MPH; Folefac Atem[2,3], PhD; Luyu Xie[1], PharmD, PhD; Olivia Kapera[4], MPH; Matthew S Mathew[1], MS; Elisa Marroquin[5], MS, PhD; Carrie McAdams[6], MD, PhD; Jeffrey Schellinger[7], MCH, RD, CSOWM, LD; Sophia Ngenge[1], MPH; Sachin Kukreja[1], MD; Benjamin E Schneider[8], MD; Jaime P Almandoz[7], MBA, MD; Sarah E Messiah[1,9], MPH, PhD

[1]Peter O'Donnell Jr. School of Public Health, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX, United States

[2]School of Public Health, University of Texas Health Science Center at Houston (UTHealth), Dallas, TX, United States

[3]Department of Biostatistics & Data Science, School of Public Health, University of Texas Health Science Center at Houston, Dallas, TX, United States

[4]School of Public Health, The University of Texas at Austin, Austin, TX, United States

[5]Department of Nutritional Sciences, Texas Christian University, Fort Worth, TX, United States

[6]Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, TX, United States

[7]Division of Endocrinology, Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX, United States

[8]Department of Surgery, University of Texas Southwestern Medical Center, Dallas, TX, United States

[9]Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX, United States

Corresponding Author:
Jackson M Francis, MPH
Peter O'Donnell Jr. School of Public Health, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX, United States

Abstract

Background: Due to the widespread use of social media and the internet in today's connected world, obesity and depression rates are increasing concurrently on a global scale. This study investigated the complex dynamics involving social media recruitment for scientific research, race, ethnicity, and depression among metabolic and bariatric surgery (MBS) candidates.

Objective: This study aimed to determine (1) the association between social media recruitment and depression among MBS candidates and (2) racial and ethnic differences in social media recruitment engagement.

Methods: The analysis included data from 380 adult MBS candidates enrolled in a prospective cohort study from July 2019 to December 2022. Race and ethnicity, recruitment method (social media: yes or no), and depression status were evaluated using χ^2 tests and logistic regression models. Age, sex, and ethnicity were adjusted in multivariable logistic regression models.

Results: The mean age of the candidates was 47.35 (SD 11.6) years, ranging from 18 to 78 years. Participants recruited through social media (n=41, 38.32%) were more likely to report past or current episodes of depression compared to nonsocial media–recruited participants (n=74, 27.11%; *P*=.03), with a 67% increased likelihood of depression (odds ratio [OR] 1.67, 95% CI 1.04 - 2.68, *P*=.03). Further analysis showed that participants with a history of depression who were below the mean sample age were 2.26 times more likely to be recruited via social media (adjusted OR [aOR] 2.26, 95% CI 1.03 - 4.95; *P*=.04) compared to those above the mean age. Hispanic (n=26, 38.81%) and non-Hispanic White (n=53, 35.10%) participants were significantly more likely to be recruited via social media than non-Hispanic Black (n=27, 18.37%) participants (*P*<.001). After adjusting for covariates, non-Hispanic Black participants were 60% less likely than non-Hispanic White participants to be recruited via social media (aOR 0.40, 95% CI 0.22 - 0.71; *P*=.002).

Conclusions: We found that individuals recruited through social media channels, especially younger participants, were more likely to report past or current episodes of depression compared to those recruited through nonsocial media. The study also showed that non-Hispanic Black individuals are less likely to engage in social media recruitment for scientific research versus other racial and ethnic groups. Future mental health–related studies should consider strategies to mitigate potential biases introduced by recruitment methods to ensure the validity and generalizability of research findings.

(JMIR Form Res 2025;9:e58916) doi:10.2196/58916

XSL•FO
RenderX

KEYWORDS

social media; social media recruitment; depression; depressed; major depressive disorder; MDD; depressive disorder; depressive; race; racial; racial differences; ethnic; ethnic differences; ethnicity; metabolic surgery; bariatric surgery; obesity; obese; online health; ethics; mental health; eHealth; internet; digital health; digital mental health; online interventions; public engagement; public health

Introduction

Over the past 2 decades, obesity (ie, BMI ≥30 kg/m^2) prevalence among adults in the United States has risen from 30.5% to 42.4% [1], and has nearly tripled worldwide [2]. Further, the prevalence of severe obesity (BMI ≥40 kg/m^2) has doubled from 4.7% to 9.2% [3]. Non-Hispanic Black (NHB) adults are disproportionately impacted, with an age-adjusted obesity prevalence of 49.9%, followed by Hispanic (45.6%), non-Hispanic White (NHW) (41.4%), and non-Hispanic Asian (16.1%) individuals [3].

Various obesity treatment options, including diet control, exercise, cognitive and family-based behavioral therapy, pharmacological therapy, and metabolic and bariatric surgery (MBS) have been explored to address this critical issue. Of these, MBS is a safe and effective treatment option, particularly for individuals with severe obesity who are resistant to other interventions [4]. A 2014 meta-analysis of 164 studies involving 160,000 patients found that BMI decreased by at least 11.8 kg/m² one-year post MBS, with sustained weight loss observed five years post surgery [5]. Most individuals who undergo MBS experience a reduction or complete elimination of obesity-related health issues, including high blood pressure, heart disease, diabetes, asthma, and osteoarthritis [6,7]. Additionally, long-term outcomes of MBS in adolescents aged 16 - 21 years demonstrated sustained weight reduction and improvement in comorbidities, including reduced depression prevalence over a decade post surgery [7]. Given the increasing prevalence of obesity and the proven safety, effectiveness, and long-term success of MBS, there has been an increase in the number of MBS procedures being performed, resulting in a parallel increase in MBS-related research [8,9]. However, the COVID-19 pandemic led to a temporary decrease in MBS surgeries due to delays in elective procedures [10]. Research studies conducted during the pandemic faced unique challenges, prompting a shift toward innovative recruitment methods. Researchers have increasingly used social media for participant recruitment, survey administration, and trend analysis. Although literature reviews underscore the efficacy of social media as a recruitment methodology, its use comes with challenges and drawbacks [11,12].

This study acknowledges the dual nature of social media, which, while a potent recruitment tool, has also been implicated in mental health and body image concerns. Celebrities and influencers on social media platforms have contributed to rising negative body image perceptions [13], which have been associated with higher depressive symptoms, loss of confidence, loneliness, and risk of developing eating disorders [13,14]. Studies have shown that viewing images that have been heavily edited can negatively impact an individual's body image and that editing selfies—regardless of whether they are posted—was linked to self-objectification and dissatisfaction with facial features [15]. Additionally, overall time spent on image-based social media platforms has been associated with weight-bias internalization [16]. A meta-analysis examining the relationship between body image and eating disorders found that both body checking and body image avoidance were associated with eating disorder pathology [17].

Acknowledging the gaps in our understanding of social media's impact on mental health among individuals with severe obesity and who are considering MBS, this study aimed to determine (1) the association between social media recruitment and depression among MBS candidates and (2) racial and ethnic differences in social media recruitment engagement. It was hypothesized that depression prevalence would be higher in individuals recruited through social media and that no significant racial and ethnic differences would exist between social media–and nonsocial–based media recruitment.

Methods

Study Design and Eligibility Criteria

The Bariatric Health Study (NIH/NIMHD 5R01MD011686) is an ongoing prospective cohort research study enrolling participants who qualify and have been referred for MBS based on established medical criteria. The American Society for Metabolic and Bariatric Surgery recommends MBS for individuals with a BMI >35 kg/m^2, regardless of the presence or absence of comorbidities; for individuals with type 2 diabetes and a BMI >30 kg/m^2; or for individuals with a BMI between 30 - 34.9 kg/m^2 who have not achieved adequate weight loss or improvement in comorbidities with nonsurgical obesity treatment methods [18]. Eligibility criteria mandated that participants be at least 18 years old, medically eligible for MBS, and referred for surgery by a clinical provider. To manage and collect study data securely and in compliance with HIPAA (Health Insurance Portability and Accountability Act of 1996) regulations, we used REDCap (Research Electronic Data Capture; Vanderbilt University), a web-based database management platform [19,20].

Ethical Considerations

The study was approved by the University of Texas Health System Institutional Review Board (IRB# HSC-SPH-18 - 0850) and the Committee for the Protection of Human Subjects. Informed consent was obtained from all study participants. Participant data are stored in a HIPAA-compliant, password-protected, encrypted database with limited access. The data used for this analysis were deidentified. Study participants were compensated with a US $50 gift card upon completion of the initial Mini International Neuropsychiatric Interview (MINI) and surveys. Upon surgery completion, participants were compensated at 6, 12, and 24 months with US

XSL•FO
RenderX

$100, US $125, and US $150 gift cards for completing the additional MINI interviews and surveys.

Participant Recruitment

Recruitment for this study was open from August 2019 to October 2022. Recruitment sites for this prospective cohort study consisted of MBS specialty clinics in the North Texas region. These included Minimally Invasive Surgical Associates, Lee Bariatrics Clinic, City Hospital at White Rock, Dallas, TX, and the UT Southwestern Weight Wellness Subspecialties Clinic, Dallas, TX. Before the COVID-19 pandemic, eligible MBS patients were recruited by research coordinators during mandated educational seminars held by MBS team members before surgery. Research coordinators provided interested candidates with information and a sign-up sheet at these seminars. In addition, the institutional review board–approved brochures and flyers were placed in the MBS clinic waiting rooms. Interested participants completed a web-based eligibility screener and were contacted by research coordinators to complete informed consent and baseline measurements. However, these educational seminars transitioned to virtual platforms due to the pandemic. Additionally, social media channels were added as a recruitment source to meet our recruitment targets, as elective surgeries at our academic center were paused on several occasions due to surges in COVID-19. Specifically, web-based social media recruitment was conducted through Facebook and Instagram and was managed by the study's research coordinators. Targeted advertisements on these platforms directed outreach toward individuals meeting our inclusion criteria. Interested individuals completed the study's online eligibility screener to ensure that the inclusion criteria were met. Following this, informed consent was completed, and research coordinators performed baseline measurements. To sustain study engagement, biweekly informative videos and posts addressing topics related to MBS were created.

Main Outcome

The primary outcome measure was self-reported depression, measured by the MINI [21], a validated tool for evaluating major psychiatric disorders, comprising dichotomous questions with follow-up questions for positive responses [21]. Previous studies have demonstrated that the MINI accurately diagnoses psychiatric conditions [22]. The interview primarily features "yes" or "no" questions, with additional clarifications provided to participants when necessary. Following the completion of questions for each psychiatric disorder, research coordinators (interviewers) tallied the "yes" and "no" responses to identify if the interviewee screened positively for a variety of potential psychiatric disorders. In this study, we assessed current, past, and recurrent major depressive episodes, posttraumatic stress disorder, agoraphobia, social anxiety disorder, general anxiety disorder, and panic disorder. Following the initial interview, research coordinators determined the presence or absence of depression based on participants' responses. Participants who screened positive for a current, past, or recurrent major depressive episode during the MINI interview were classified as having a history of depression. Conversely, if none of these conditions were selected (current, past, or recurrent), participants were categorized as not having a history of depression.

Main Exposure

We assessed each participant's recruitment method through the enrollment survey. Participants responded to the question, "How did you learn about this study?" They could choose from options such as "Lee Bariatrics, Texas Presbyterian Hospital, UT Southwestern Medical Center, Minimally Invasive Surgical Associates, UTSW Weight Wellness Program, City Hospital at White Rock, Flyer, Brochure or Pamphlet, Social media (ie, Facebook, Instagram), or Other." We classified participants as positive for social media recruitment if they chose "Social media (ie, Facebook, Instagram)" or mentioned social media recruitment in the "Other" section. We categorized participants as negative for social media recruitment if they selected other options.

Covariates

Covariates including age, sex, race and ethnicity, BMI, education level, and perceived financial well-being were obtained through self-report. Age was treated as a continuous variable, representing the participants' age in years. Sex was a categorical variable, categorized as male, female, or other. Race and ethnicity were also categorical variables, encompassing diverse categories such as NHW, NHB, Hispanic, and Other. BMI was calculated via self-reported height and weight and verified via medical records provided by participants when available. Education was a categorical variable, categorized as none, elementary school, high school, college, and graduate or professional degree. In our survey, we did not have a direct measure of income. However, as a proxy measure, we asked the question, "In the last two weeks, have you had enough money to meet your needs?" This variable was categorical, with answers being "not at all, a little, moderately, mostly, and completely." These covariates served as essential demographic factors for our analysis, providing a comprehensive understanding of their potential influence on the investigated variables.

Statistical Analysis

We analyzed descriptive variables, including sex, age, race and ethnicity (ie, NHB, NHW, Hispanic), BMI, and history of depression, comprising both categorical and continuous variables. To explore the associations between recruitment methods, social media or nonsocial media, and depression history (yes/no), we conducted Pearson χ^2 tests. Crude odds ratios were calculated using univariable logistic regression models to assess the relationship between recruitment method and depression history. Furthermore, we performed multivariable logistic regression by adjusting for age, race and ethnicity, and sex to calculate the adjusted odds ratios for depression in relation to the recruitment method used. There were less than 3% missing data for demographic variables (ie, age, sex, and ethnicity); no data was missing for the outcome variable. We performed all analyses using STATA (v.17.1, Stata Corp LP), with statistical significance set at a *P* value below 5%.

XSL•FO
RenderX

Results

Participant Characteristics

Participant characteristics (N=380) stratified by recruitment method, social media (n=107), and nonsocial media (n=273) are presented in Table 1. Participants recruited through social media had a mean age of 47.27 (SD 1.02) years, whereas those in the nonsocial media group had a mean age of 47.38 (SD 0.73) years (*P*=.93). Sex distribution in the social media group was 9 (8.41%) males and 98 (91.59%) females, while the nonsocial media group included 41 (15.02%) males and 232 (84.98%) females, *P*=.09. The mean BMI for the social media group was 44.92 (SD 1.05) kg/m^2 and the nonsocial media group was 45.68 (SD 0.64) kg/m^2 (*P*=.54). Regarding racial and ethnic backgrounds, the social media group had 49.53% (n=53) of participants identifying as NHW, 25.23% (n=27) as NHB, 24.3% (n=26) as Hispanic, and 0.93% (n=1) as other. In the nonsocial media group, 35.9% (n=98) were NHW, 43.96% (n=120) were NHB, 15.02% (n=41) were Hispanic, and 5.13% identified as Other (*P*<.001).

Table . Baseline characteristics of metabolic and bariatric surgery candidates recruited via social media and traditional methods in a prospective cohort study (N=380, 2019 - 2022).

Variables	Participants (N=380)		*P* value
	Social media (n=107)	Nonsocial media (n=273)	
Age (years), mean (SD)	47.27 (1.02)	47.38 (0.73)	.93
Sex, n (%)			.09
Males (n=50, 13.16%)	9 (8.41)	41 (15.02)	
Female (n=330, 86.84%)	98 (91.59)	232 (84.98)	
BMI (kg/m^2), mean (SD)	44.92 (1.04)	45.68 (0.64)	.54
Race/ethnicity, n (%)			<.001[a]
Non-Hispanic White (n=151)	53 (49.53)	98 (35.9)	
Non-Hispanic Black (n=147)	27 (25.23)	120 (43.96)	
Hispanic (n=67)	26 (24.3)	41 (15.02)	
Other (n=15)	1 (0.93)	14 (5.13)	

[a]Significant.

Depression Prevalence by Recruitment Method

Table 2 presents the prevalence of depression categorized by recruitment method and race and ethnicity among all participants (N=380). Of the participants, 115 (30.26%) had a history of depression, while 265 (69.74%) had no history of depression. Our findings showed a difference between the prevalence of depression and the recruitment methods (*P*=.03).

In the nonsocial media group (n=273, 71.84%), the prevalence rates of depression (n=74, 27.11%) varied by race and ethnicity. In the NHW subgroup (n=98), 26 (26.53%) participants had a history of depression, while 72 (73.47%) did not. In the NHB subgroup (n=120), 28 (23.33%) participants had a history of depression, whereas 92 (76.67%) had no history. Among the Hispanic subgroup (n=41), 15 (36.59%) participants had a history of depression, while 26 (63.41%) had no history. In the “Other” subgroup (n=14), 5 (35.71%) participants had a history of depression, while 9 (64.29%) had no history.

Within the social media group (n=107, 28.16%), subgroups based on race and ethnicity exhibited varied depression prevalence rates (n=41, 38.32%). In the NHW subgroup (n=53), 22 (41.51%) had a history of depression, while 31 (58.49%) had no history. Among the NHB subgroup (n=27), 7 (25.93%) had a history of depression, while 20 (74.07%) had no history. In the Hispanic subgroup (n=26), 11 (42.31%) individuals had a history of depression, whereas 15 (57.69%) had no history. A single participant in the “Other” subgroup had a history of depression.

XSL•FO
RenderX

Table . Prevalence of depression among metabolic and bariatric surgery candidates by recruitment method and race and ethnicity in a prospective cohort study (2019 - 2022).

Variable and group	Participants (N=380), n (%)		*P* value
	History of depression	No history of depression	
Overall prevalence	115 (30.26)	265 (69.74)	.03[a]
Social media group (n=107)	41 (38.32)	66 (61.68)	—[b]
NHW[c] (n=53)	22 (41.51)	31 (58.49)	
NHB[d] (n=27)	7 (25.93)	20 (74.07)	
Hispanic (n=26)	11 (42.31)	15 (57.69)	
Other (n=1)	1 (100)	0 (0)	
Nonsocial media group (n=273)	74 (27.11)	199 (72.89)	—
NHW (n=98)	26 (26.53)	72 (73.47)	
NHB (n=120)	28 (23.33)	92 (76.67)	
Hispanic (n=41)	15 (36.59)	26 (63.41)	
Other (n=14)	5 (35.71)	9 (64.29)	

[a]There was a significant difference between depression prevalence in the social media group versus nonsocial media group.

[b]Not applicable.

[c]NHW: non-Hispanic White.

[d]NHB: non-Hispanic Black.

Recruitment Method by Race and Ethnicity

Table 3 presents the odds of recruitment via social media, adjusting for age, sex, and BMI, stratified by race and ethnicity. In the unadjusted model, representing the crude odds, Hispanic participants had odds of 1.17 (95% CI 0.24 - 0.71, *P*=.60) for being recruited via social media, while NHB participants had odds of 0.42 (95% CI 0.65 - 2.12, *P*=.001), with NHW participants serving as the reference group.

In the adjusted model, adjusted for age, sex, BMI, education, and financial status, Hispanic participants had odds of 1.22 (95% CI 0.63 - 2.37, *P*=.55) for being recruited via social media, while NHB participants had odds of 0.40 (95% CI 0.22 - 0.72, *P*=.002). These findings indicate that after adjusting for demographic variables, NHB participants demonstrated significantly lower odds of recruitment via social media compared to NHW participants.

Table . Odds of being recruited via social media by race and ethnicity among metabolic and bariatric surgery candidates, adjusting for age, sex, BMI, education, and financial status.

Variables	Odds ratio (95% CI)	*P* value
Race/ethnicity		
Model 1[a] (crude odds ratio)		
Hispanic	1.17 (0.24 - 0.71)	.60
NHB[b]	0.42 (0.65 - 2.12)	.001[c]
NHW[d]	1.0 (ref)	—[e]
Model 2[f] (adjusted odds ratio)		
Hispanic	1.22 (0.63 - 2.37)	.55
NHB	0.40 (0.22 - 0.72)	.002[c]
NHW	1.0 (ref)	—

[a]Model 1: crude (social media as primary outcome).

[b]NHB: non-Hispanic Black.

[c]Significant.

[d]NHW: non-Hispanic White.

[e]Not applicable.

[f]Model 2: adjusted for age, sex, BMI, education, and financial status (social media as primary outcome).

XSL•FO
RenderX

History of Depression and Social Media Recruitment

Univariate logistic models showed that individuals with a history of depression were 67% more likely (OR 1.67, 95% CI 1.04 - 2.68, *P*=.03) to be recruited via social media than those with no history of depression (Table 4). However, after adjustment for age, sex, race and ethnicity, BMI, education, and financial status, the association between depression history and recruitment method was no longer significant. This indicated that one or more of these covariates influenced the relationship between our outcome variables, recruitment method, and depression history, warranting further subanalysis.

Table . Odds of depression for social media recruits, adjusting for age among metabolic and bariatric surgery candidates, race and ethnicity, sex, BMI, education, and financial status.

Variables	Odds ratio (95% CI)	*P* value
Depression history		
Model 1[a] (crude odds ratio)		
Depression	1.67 (1.04 - 2.68)	.03[b]
No depression	1.0 (ref)	—[c]
Model 2[d] (adjusted odds ratio)		
Depression	1.51 (0.88 - 2.62)	.13
No depression	1.0 (ref)	—

[a]Model 1: crude (social media as primary outcome).

[b]Significant.

[c]Not applicable.

[d]Model 2: adjusted for age, race/ethnicity, sex, BMI, education, and financial status (social media as primary outcome).

Age, Depression, and Recruitment Method

After adjusting for sex, race and ethnicity, BMI, education, and financial status, participants with a history of depression who were below the mean age were over twice as likely (OR 2.26, 95%CI, 1.03 - 4.95) to be recruited via social media compared to those with no history of depression (Table 5). However, with the same adjustments, no significant association was found among participants above the mean sample age.

Table . Relationship between age and recruitment method on odds of depression among metabolic and bariatric surgery candidates, after adjusting for sex, race and ethnicity, BMI, education, and financial status.

Variables	Adjusted odds ratio (95% CI)	*P* value
Depression history		
Model 1[a]		
Depression	2.26 (1.03 - 4.95)	.04[b]
No depression	1.0 (ref)	—[c]
Model 2[d]		
Depression	1.13 (0.47 - 2.73)	.79
No depression	1.0 (ref)	—

[a]Model 1: age below mean, adjusted for sex, race/ethnicity, BMI, education, and financial status (social media as primary outcome).

[b]Significant.

[c]Not applicable.

[d]Model 2: age above mean, adjusted for sex, race/ethnicity, BMI, education, and financial status (social media as primary outcome).

Discussion

Principal Findings

This study aimed to investigate the relationship between social media recruitment, depression, race, and ethnicity among individuals seeking MBS. Our analysis of 380 MBS candidates showed that participants recruited via social media exhibited a 67% higher likelihood of depression history compared to nonsocial media recruits. Furthermore, participants with a history of depression below the mean sample age were over twice as likely to be recruited via social media. Our findings also revealed significant disparities in social media recruitment among racial and ethnic groups. Hispanic and NHW participants were more likely to be recruited through social media than NHB. Specifically, adjusted models demonstrated a 60% lower likelihood of social media recruitment among NHB participants compared to NHW participants. These findings underscore the

XSL•FO
RenderX

significance of exploring the interplay between social media, mental health disorders, and demographic factors in this population.

Our first key finding reveals a significant association between social media recruitment and a history of depression, particularly among younger individuals below the sample mean age, who were twice as likely to report such a history. Extensive studies have explored the relationship between social media usage and depression, emphasizing the heightened risk of depression, anxiety, and psychological distress associated with addictive or problematic platform use [23]. A systematic examination of 11 studies focusing on young individuals corroborates this, revealing a modest yet statistically significant association between social media usage and symptoms of depression [24]. The observation that younger individuals with a history of depression are more likely to be recruited via social media underscores the growing concern about the impact of social media on mental health, particularly in younger populations.

Other contributing factors between social media and depression include sleep impairment and sedentary behavior [25]. These factors, combined with the other potentially negative effects of problematic social media use, defined as a "nonsubstance–related disorder by which detrimental effects occur as a result of preoccupation and compulsion to excessively engage in social media platforms despite negative consequences" [26], may contribute to the higher prevalence of depression observed in our sample of MBS candidates. Addressing these issues through increased physical activity could offer protective benefits, as regular exercise has been shown to reduce depressive symptoms, support healthy weight management, and improve sleep quality [25,27-32]. While both depression and obesity are multifaceted conditions influenced by various biological, psychological, and social determinants, promoting physical activity represents a practical starting point for improving overall mental and physical health.

Our second key finding was that NHB participants were 60% less likely to be recruited via social media. This finding is particularly concerning given that NHB individuals bear a disproportionate burden of obesity and its health-related consequences. According to the National Center for Health Statistics, NHB women represent the demographic group with the highest risk for overweight, obesity, and severe obesity, followed closely by NHB men [33]. However, this underrepresentation is not surprising, given the significant history of distrust due to unethical practices that have occurred when enrolling this population in previous scientific studies. Prior research has indicated that NHB individuals are less inclined to participate in scientific studies overall [34]. Our results corroborate this disparity and underscore the challenges of using social media for recruitment in scientific research.

The efficacy of social media as a tool for recruiting diverse samples remains a topic of debate [11]. The success of such methods often depends on a comprehensive understanding of the target audience's social media habits [35]. Recent research has shown promising results, with one study demonstrating a substantial increase in racial and ethnic diversity among participants recruited through social media for a clinical trial [36]. Other studies have raised concerns about potential biases, noting an overrepresentation of younger NHW individuals with higher education and income levels [37]. Our findings align with these observations, as we also reported a higher likelihood of recruiting Hispanic and NHW individuals through social media compared to NHB individuals. This result emphasizes the need for proactive measures to address potential biases in social media recruitment.

Additionally, obesity has been consistently associated with an increased risk of depression [38]. However, the significance of social support, whether received in person or through web-based platforms, cannot be overstated in optimizing positive outcomes for patients undergoing MBS. Previous research has underscored the positive correlation between heightened social support and decreased depression, along with improved weight loss among both pre- and post-MBS groups [39]. In one study, Facebook was used as a platform to deliver weight management program materials, resulting in documented weight loss among participants [40]. Furthermore, a 2016 study examining MBS-related Facebook support groups identified various recurring themes in online discussions, ranging from eating recommendations and postsurgery changes to anxiety, depression, and body image concerns [41]. The study concluded that online groups are valuable for fostering support, sharing experiences, and discussing challenges for pre- and post-MBS patients.

Overall, social media as a tool for research recruitment has several advantages, including a broad reach, cost-effectiveness, and the potential to foster social support related to weight management, eating disorders, and mental health. However, researchers must approach social media recruitment cautiously or risk recruiting homogenous samples and should be aware of social media's association with depression. Combining more traditional recruitment methods with online recruitment can help ensure diverse study samples and enhance the validity, generalizability, and equity of findings.

Future research should explore culturally sensitive recruitment methods that include social media and other innovative techniques while continuing to examine the relationship between mental health and social media usage. Improved recruitment methods should consist of partnerships with community organizations, tailored outreach campaigns, and evaluations of effectiveness. As our digital world continues to evolve, enhancing recruitment diversity is critical to reducing health inequities and advancing the quality and fairness of clinical and public health research.

Limitations and Strengths

Despite providing valuable insights, this study has limitations. One limitation of our study is that the sample was drawn exclusively from a single geographic region, which may impact the generalizability of the findings to a broader population. However, it is crucial to highlight that our sample includes a diverse participant population. Furthermore, our sample consisted specifically of MBS candidates, and we examined social media recruitment among this group and not social media use itself. All results should be interpreted with these key factors and limitations in mind.

XSL•FO
RenderX

Conclusion

In conclusion, our findings show that MBS candidates recruited through social media had a higher prevalence of depression, particularly among younger individuals, and were more likely to be NHW or Hispanic. Therefore, our findings underscore the significance of using various recruitment methods to prevent bias and optimize and ensure sample diversity in research.

Acknowledgments

This work was funded by the National Institutes of Health, National Institute on Minority Health and Health Disparities (Grant R01MD011686). The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

Data Availability

The data supporting this study's findings are not openly available due to reasons of sensitivity and are available from the corresponding author upon reasonable request. Data are in controlled access data storage at UTHealth School of Public Health.

Authors' Contributions

Formal analysis: FA, JMF
Investigation: BES, CMA, JS, JPA, SK
Supervision: SEM
Writing – original draft: DP, JMF, SSN
Writing – review & editing: DP, EM, JMF, LX, MSM, OK, SEM, SN, SSN

Conflicts of Interest

None declared.

References

1. Boutari C, Mantzoros CS. A 2022 update on the epidemiology of obesity and a call to action: as its twin COVID-19 pandemic appears to be receding, the obesity and dysmetabolism pandemic continues to rage on. Metab Clin Exp 2022 Aug;133:155217. [doi: 10.1016/j.metabol.2022.155217]
2. Obesity and overweight. World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2023-12-01]
3. Stierman B, Afful J, Carroll MD, et al. National health and nutrition examination survey 2017-March 2020 prepandemic data files-development of files and prevalence estimates for selected health outcomes. Natl Health Stat Report 2021 Jun 14(158). [doi: 10.15620/cdc:106273] [Medline: 39380201]
4. Valerio G, Maffeis C, Saggese G, et al. Diagnosis, treatment and prevention of pediatric obesity: consensus position statement of the Italian Society for Pediatric Endocrinology and Diabetology and the Italian Society of Pediatrics. Ital J Pediatr 2018 Jul 31;44(1):88. [doi: 10.1186/s13052-018-0525-6] [Medline: 30064525]
5. Chang SH, Stoll CRT, Song J, Varela JE, Eagon CJ, Colditz GA. The effectiveness and risks of bariatric surgery: an updated systematic review and meta-analysis, 2003-2012. JAMA Surg 2014 Mar;149(3):275-287. [doi: 10.1001/jamasurg.2013.3654] [Medline: 24352617]
6. Haase CL, Lopes S, Olsen AH, Satylganova A, Schnecke V, McEwan P. Weight loss and risk reduction of obesity-related outcomes in 0.5 million people: evidence from a UK primary care database. Int J Obes 2021 Jun;45(6):1249-1258. [doi: 10.1038/s41366-021-00788-4]
7. de la Cruz-Muñoz N, Xie L, Quiroz HJ, et al. Long-term outcomes after adolescent bariatric surgery. J Am Coll Surg 2022 Oct 1;235(4):592-602. [doi: 10.1097/XCS.0000000000000325] [Medline: 36102560]
8. Messiah SE, Xie L, de la Cruz-Muñoz N, Lipshultz SE. Use of metabolic and bariatric surgery among US youth. JAMA Pediatr 2023 Aug 1;177(8):856-857. [doi: 10.1001/jamapediatrics.2023.0803] [Medline: 37252718]
9. Alsuhibani A, Thompson JR, Wigle PR, et al. Metabolic and bariatric surgery utilization trends in the United States: evidence from 2012 to 2021 National Electronic Medical Records Network. Ann Surg Open 2023 Dec;4(4):e317. [doi: 10.1097/AS9.0000000000000317] [Medline: 38144499]
10. Messiah SE, Xie L, Mathew MS, et al. Impact of the COVID-19 pandemic on metabolic and bariatric surgery utilization and safety in the United States. Obesity Surg 2022 Jul;32(7):2289-2298. [doi: 10.1007/s11695-022-06077-x] [Medline: 35499637]
11. Topolovec-Vranic J, Natarajan K. The use of social media in recruitment for medical research studies: a scoping review. J Med Internet Res 2016 Nov 7;18(11):e286. [doi: 10.2196/jmir.5698] [Medline: 27821383]

XSL•FO
RenderX

12. Darko EM, Kleib M, Olson J. Social media use for research participant recruitment: integrative literature review. J Med Internet Res 2022 Aug 4;24(8):e38015. [doi: 10.2196/38015] [Medline: 35925655]
13. Pedalino F, Camerini AL. Instagram use and body dissatisfaction: the mediating role of upward social comparison with peers and influencers among young females. Int J Environ Res Public Health 2022 Jan 29;19(3):1543. [doi: 10.3390/ijerph19031543] [Medline: 35162562]
14. Yazdani N, Siedlecki KL. Mediators of the relationship between cognition and subjective well-being. J Happiness Stud 2021 Oct;22(7):3091-3109. [doi: 10.1007/s10902-021-00357-6]
15. Vandenbosch L, Fardouly J, Tiggemann M. Social media and body image: recent trends and future directions. Curr Opin Psychol 2022 Jun;45:101289. [doi: 10.1016/j.copsyc.2021.12.002] [Medline: 35030460]
16. Moufawad M, Hoque A, Kells M, Sonneville KR, Hahn SL. Social media use and weight bias internalization: association moderated by age and weight perception. J Eat Disord 2024 Jun 18;12(1):84. [doi: 10.1186/s40337-024-01043-7] [Medline: 38890765]
17. Walker DC, White EK, Srinivasan VJ. A meta-analysis of the relationships between body checking, body image avoidance, body image dissatisfaction, mood, and disordered eating. Int J Eat Disord 2018 Aug;51(8):745-770. [doi: 10.1002/eat.22867] [Medline: 29659039]
18. Is metabolic and bariatric surgery right for you? American Society for Metabolic and Bariatric Surgery. URL: https://asmbs.org/patients/is-metabolic-and-bariatric-surgery-right-for-you [accessed 2024-03-19]
19. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
20. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
21. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59(20):22-33. [Medline: 9881538]
22. Pettersson A, Modin S, Wahlström R, Af Winklerfelt Hammarberg S, Krakau I. The Mini-International Neuropsychiatric Interview is useful and well accepted as part of the clinical assessment for depression and anxiety in primary care: a mixed-methods study. BMC Fam Pract 2018 Jan 24;19(1):19. [doi: 10.1186/s12875-017-0674-5] [Medline: 29368585]
23. Keles B, McCrae N, Grealish A. A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. Int J Adolesc Youth 2020 Dec 31;25(1):79-93. [doi: 10.1080/02673843.2019.1590851]
24. McCrae N, Gettings S, Purssell E. Social media and depressive symptoms in childhood and adolescence: a systematic review. Adolescent Res Rev 2017 Dec;2(4):315-330. [doi: 10.1007/s40894-017-0053-4]
25. Lederman O, Ward PB, Firth J, et al. Does exercise improve sleep quality in individuals with mental illness? A systematic review and meta-analysis. J Psychiatr Res 2019 Feb;109:96-106. [doi: 10.1016/j.jpsychires.2018.11.004] [Medline: 30513490]
26. Shannon H, Bush K, Villeneuve PJ, Hellemans KG, Guimond S. Problematic social media use in adolescents and young adults: systematic review and meta-analysis. JMIR Ment Health 2022 Apr 14;9(4):e33450. [doi: 10.2196/33450] [Medline: 35436240]
27. Krittanawong C, Maitra NS, Qadeer YK, et al. Association of depression and cardiovascular disease. Am J Med 2023 Sep;136(9):881-895. [doi: 10.1016/j.amjmed.2023.04.036] [Medline: 37247751]
28. Popovic D, Bjelobrk M, Tesic M, et al. Defining the importance of stress reduction in managing cardiovascular disease - the role of exercise. Prog Cardiovasc Dis 2022;70:84-93. [doi: 10.1016/j.pcad.2022.01.008] [Medline: 35131232]
29. Ozemek C, Lavie CJ, Kaminsky LA. Review of recent cardiac rehabilitation research related to enrollment/adherence, mental health, and other populations. J Cardiopulm Rehabil Prev 2021;41(5):302-307. [doi: 10.1097/HCR.0000000000000649]
30. Boidin M, Grégoire CA, Gagnon C, et al. Effects of variation in exercise training load on cognitive performances and neurotrophic biomarkers in patients with coronary artery disease. J Appl Physiol (1985) 2024 Nov 1;137(5):1158-1167. [doi: 10.1152/japplphysiol.00636.2023] [Medline: 38961824]
31. Lavie CJ, Zhang I, Yang D, Liu M. Improving fitness through exercise will improve our heart and mind. Heart and Mind 2023;7(1):49-51. [doi: 10.4103/hm.hm_59_22]
32. Banno M, Harada Y, Taniguchi M, et al. Exercise can improve sleep quality: a systematic review and meta-analysis. PeerJ 2018;6:e5172. [doi: 10.7717/peerj.5172] [Medline: 30018855]
33. Welsh LK, Luhrs AR, Davalos G, et al. Racial disparities in bariatric surgery complications and mortality using the MBSAQIP data registry. Obesity Surg 2020 Aug;30(8):3099-3110. [doi: 10.1007/s11695-020-04657-3] [Medline: 32388704]
34. Milani SA, Swain M, Otufowora A, Cottler LB, Striley CW. Willingness to participate in health research among community-dwelling middle-aged and older adults: does race/ethnicity matter? J Racial Ethn Health Disparities 2021 Jun;8(3):773-782. [doi: 10.1007/s40615-020-00839-y] [Medline: 32808194]
35. Khatri C, Chapman SJ, Glasbey J, et al. Social media and internet driven study recruitment: evaluating a new model for promoting collaborator engagement and participation. PLoS One 2015;10(3):e0118899. [doi: 10.1371/journal.pone.0118899] [Medline: 25775005]

XSL•FO
RenderX

36. Strauch J, Agnew A, Meenaghan E, Miller KK, Haines MS. Recruitment strategies to increase racial and ethnic diversity in anorexia nervosa clinical research. J Eat Disord 2023 Jul 15;11(1):118. [doi: 10.1186/s40337-023-00844-6] [Medline: 37454157]
37. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res 2017 Aug 28;19(8):e290. [doi: 10.2196/jmir.7071] [Medline: 28851679]
38. Luppino FS, de Wit LM, Bouvy PF, et al. Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. Arch Gen Psychiatry 2010 Mar;67(3):220-229. [doi: 10.1001/archgenpsychiatry.2010.2] [Medline: 20194822]
39. Conceição EM, Fernandes M, de Lourdes M, Pinto-Bastos A, Vaz AR, Ramalho S. Perceived social support before and after bariatric surgery: association with depression, problematic eating behaviors, and weight outcomes. Eat Weight Disord 2020 Jun;25(3):679-692. [doi: 10.1007/s40519-019-00671-2] [Medline: 30859467]
40. Napolitano MA, Hayes S, Bennett GG, Ives AK, Foster GD. Using Facebook and text messaging to deliver a weight loss program to college students. Obesity (Silver Spring) 2013 Jan;21(1):25-31. [doi: 10.1002/oby.20232] [Medline: 23505165]
41. Koball AM, Jester DJ, Domoff SE, Kallies KJ, Grothe KB, Kothari SN. Examination of bariatric surgery Facebook support groups: a content analysis. Surg Obes Relat Dis 2017 Aug;13(8):1369-1375. [doi: 10.1016/j.soard.2017.04.025] [Medline: 28600115]

Abbreviations

aOR: adjusted odds ratio
HIPAA: Health Insurance Portability and Accountability Act
MBS: metabolic and bariatric surgery
MINI: Mini International Neuropsychiatric Interview
NHB: non-Hispanic Black
NHW: non-Hispanic White
OR: odds ratio
REDCap: Research Electronic Data Capture

Edited by A Mavragani; submitted 12.04.24; peer-reviewed by C Lavie, H Hecht, K Phankitnirundorn, MB Ellur; revised version received 06.02.25; accepted 07.02.25; published 10.04.25.

Please cite as:
Francis JM, Neti SS, Polavarapu D, Atem F, Xie L, Kapera O, Mathew MS, Marroquin E, McAdams C, Schellinger J, Ngenge S, Kukreja S, Schneider BE, Almandoz JP, Messiah SE
Association of Social Media Recruitment and Depression Among Racially and Ethnically Diverse Metabolic and Bariatric Surgery Candidates: Prospective Cohort Study
JMIR Form Res 2025;9:e58916
URL: https://formative.jmir.org/2025/1/e58916
doi:10.2196/58916

© Jackson M Francis, Sitapriya S Neti, Dhatri Polavarapu, Folefac Atem, Luyu Xie, Olivia Kapera, Matthew S Mathew, Elisa Marroquin, Carrie McAdams, Jeffrey Schellinger, Sophia Ngenge, Sachin Kukreja, Benjamin E Schneider, Jaime P Almandoz, Sarah E Messiah. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Encouraging the Voluntary Mobilization of Mental Resources by Manipulating Task Design: Explorative Study

Lina-Estelle Louis[1], PhD; Saïd Moussaoui[1], Prof Dr; Sébastien Ravoux[2], ING; Isabelle Milleville-Pennel[1], Prof Dr

[1]École Centrale Nantes, Centre National de la Recherche Scientifique (CNRS), Laboratoire des Sciences du Numérique de Nantes (LS2N), UMR 6004, Nantes Université, Nantes, France

[2]Entreprise Onepoint, Nantes, France

Corresponding Author:
Isabelle Milleville-Pennel, Prof Dr
École Centrale Nantes, Centre National de la Recherche Scientifique (CNRS), Laboratoire des Sciences du Numérique de Nantes (LS2N), UMR 6004, Nantes Université, Nantes, France

Abstract

Background: Cognitive training is increasingly being considered and proposed as a solution for several pathologies, particularly those associated with aging. However, trainees need to be willing to invest enough mental effort to succeed and make progress.

Objective: In this study, we explore how gamification in a narrative context (ie, the addition of visual game-like elements [GLEs] embedded in real-world contexts) could contribute increase in perceived playfulness (PP) and voluntary mental effort allocated to a cognitive task. In such context, narrative elements and GLEs can be designed to align with a commonly relatable scenario (like simulating fishing or gardening activity) to ground the task in familiar, real-world contexts. We also consider if the supposed effect of GLEs on PP and voluntary mental effort could endure while manipulating an intrinsic variable of the task (ie, by increasing cognitive solicitation).

Methods: In total, 20 participants (average age 33.6, SD 8.6 y) took part in 3 cognitive tasks proposed in a numerical format: a classic version of the Corsi test (Classic Corsi, a spatial memory task), a playful version of the Classic Corsi test (Playful Corsi), with added visual GLEs in a narrative context, and a playful version of the Classic Corsi test with added cognitive solicitation, that is, mental motor inhibition (Playful Corsi Multi). We assessed the impact of visual GLEs and cognitive solicitation on PP (1 question) and mental workload (MWL) using NASA-Task Load Index (NASA-TLX) and workload profile (WP) questionnaires.

Results: Results showed that PP was not influenced by interface's playful characteristics (Classic Corsi [mean 62.4, SD 8.8] vs Playful Corsi [mean 66, SD 8.8]; W=77; *P*=.30) but decreased the time necessary to complete the task (Classic Corsi [mean 10.7, SD 2.1 s] vs Playful Corsi [mean 6.8, SD 1.6 s]; W=209; *P*<.001) as well as performance (Classic Corsi [mean 92.4, SD 9.1] vs Playful Corsi [mean 88.2, SD 11.3]; W=140.5; *P*=.02). So, possibly, visual GLEs could raise the stakes of the task slightly and implicitly encourage people to go a bit faster. Furthermore, visual GLEs increased MWL regarding attentional resources (assessed by WP: Classic Corsi [mean 52.4, SD 10.9] vs Playful Corsi [mean 65.8, SD 10.9]; W=27.5; *P*=.04), while manipulating cognitive solicitation impacted MWL when linked to task requirements (assessed by NASA-TLX: Playful Corsi [mean 54.2, SD 9.4] vs Playful Corsi Multi [mean 67.5, SD 9.4]; W=35.5; *P*=.01) without impacting the performance to the task (Playful Corsi [mean 83.8, SD 13.9] vs Playful Corsi Multi [mean 94, SD 5.5]; W=27; *P*=.007). Thus, working on the way cognitive functions are solicited would be wiser than adding visual GLEs to improve users' voluntary mental effort while preserving performance.

Conclusion: These results offer valuable insights to improve users' experience during gamified cognitive tasks and serious games.

(JMIR Form Res 2025;9:e63491) doi:10.2196/63491

KEYWORDS

visual game-like elements; gamification; multiplicity of cognitive functions; cognitive tasks; perceived playfulness; mental workload; performance; cognitive training; aging; mental effort; cognitive function; cognitive skills; cognitive tests

Introduction

Cognitive training is increasingly being considered and proposed as a solution for several pathologies, particularly those associated with aging. However, trainees must be willing to invest enough mental effort to succeed and make progress. During the execution of an activity, we have the possibility of voluntarily activating our mental resources to reach a level of performance set by the task or by ourselves. This voluntary modulation of the quantity of mental resources invested when performing a task corresponds to the notion of mental effort [1-3]. Voluntary mental effort is a dimension of mental workload (MWL). The latter is a broader concept that also encompasses

XSL•FO
RenderX

notions such as the perception of our own performance, or emotional states such as frustration. Voluntary mental effort should be distinguished from mental effort imposed by the task properties and undergone by the participant. Indeed, voluntary mental effort must be seen as a "positive" mental effort as it corresponds to a voluntary deployment of resources with positive consequences for the participant (high perceived performance and low frustration). Thus, in this context it should be considered regarding the corresponding perceived performance and frustration when performing the task. Positive mental effort should be accompanied by high perceived performance and low frustration. To make the notion of positive mental effort operational, we return to the definition of this concept, specifying that it is part of the MWL construct. In the scientific literature, NASA-Task Load Index (NASA-TLX) [2] and workload profile (WP) [4] questionnaires permit to measure MWL linked respectively to task requirements and the mobilization of attentional resources (since the WP employs the multiple resource theory of [5-7] as its foundation). NASA-TLX is particularly interesting for addressing the notion of positive mental effort, as it considers various dimensions of MWL, including mental effort, perceived performance and frustration, that we considered as crucial in the definition of positive mental effort.

Mental effort has been studied in various contexts since it is a key element for enhancing performance. However, few studies have focused on the notion of positive mental effort during cognitive training, which aims to enhance general cognitive skills such as memory, attention, or processing speed through repeated interventions using cognitive tasks or intellectually demanding activities [8-10]. Our aim is thus to study the factors that influence voluntary mobilization of mental effort when confronted to cognitive tasks used in a cognitive training process. To do that, we need to consider a task allowing to assess voluntary mobilization of mental effort but without the confounding effects associated with longitudinal cognitive training paradigms (reinforcement, error feedback, and repetition). Thus, in this study, our aim is not to design interventions for cognitive training, but rather to propose isolated cognitive tasks as controlled experimental tools to investigate the role of some factors. These factors could then be integrated in tasks developed for cognitive training. So, we wondered how we could encourage individuals to voluntarily mobilize greater mental effort to achieve better performance. To this end, we wanted to explore one possibility, that is, the gamification of a cognitive task.

Gamification is defined as the use of game-like elements (GLEs) in nongame contexts. GLEs are elements typically found in games such as visual elements (icons, colored patterns, evocative images, or playful animations), badges systems, narration, or leaderboards [9,11]. This definition of gamification differentiates it from serious games. Indeed, according to Vermeir et al [9], serious games use full-fledged games in nongame contexts, whereas gamification uses elements of a game integrated into real-world contexts (for instance, a cognitive task that uses colored patterns).

Furthermore, gamification can be linked to cognitive tasks, but its introduction may affect the validity of cognitive tests by potentially introducing new cognitive demands. Therefore, the gamification of cognitive tasks is most applied within the context of cognitive training [9,11-13]. The reason is because cognitive tasks are perceived as demotivating and frustrating given their repetitive and effortful aspects (Lumsden et al [14]) which may impact on perceived playfulness (PP). PP is the sense of being focused on an activity that is viewed as inherently enjoyable, fun, and interesting [15,16]. Higher PP results in higher user intention to perform a task [17], a stronger attitude toward using a particular technology [16], a better performance, and a higher effective response to computer training tasks [18]. Thus, we predict that the increasing of PP would impact the voluntary mobilization of mental resources and thus positive mental effort. There would therefore be a real interest in increasing the PP of a cognitive task during cognitive training. Integrating GLEs into cognitive tasks could be one solution, as their inclusion is supposed to increase PP [11,19].

Among all GLEs, performance feedback via a reward points system (the unit of measurement quantifying a user progression) is one of the most widely used [19] and therefore the most studied in the literature. Nevertheless, this system provides feedback on performance and thus contributes to its improvement. While it does indeed increase motivation and PP, it also introduces a bias in the sense that it induces a form of learning. In this study, we aimed to understand the factors that encourage the voluntary mobilization of mental resources in a cognitive task (positive mental effort), with the hypothesis that this mobilization will have a positive impact on performance. Introducing reinforcement would induce a sort of learning process that could mask the effect of other variables allowing to increase positive mental effort. Thus, we provide no reinforcement through performance feedback.

Consequently, only a selected few GLEs, specifically visual ones, which have no impact on performance feedback, should be incorporated. Visual GLEs can be defined as the deliberate arrangement of several visual variables, such as position and form [20], integrated into a particular context. Visual GLEs are contextualized with narrative elements (short story explaining the task and a visual surrounding corresponding to the story) to make them meaningful for the participant and add ecological aspect to the cognitive task. By ecological aspect, we mean a less arbitrary situation. The task has an objective that can be linked to a game that could be encountered in real life. For example, the narrative elements can be designed to align with commonly relatable scenarios, (like aiming at targets, arranging objects to solve a puzzle, navigating a maze, or simulating a fishing or gardening activity) to ground the task in familiar, real-world contexts. These elements serve two purposes, that is, (1) the short story provides a framework for participants to understand the task's purpose, making the activity less abstract and (2) the visual surroundings (eg, buoys, and water effects) enhance immersion, reinforcing the narrative and fostering a sense of engagement. Together, these components aim to reduce the perceived arbitrariness of the cognitive task, encouraging participants to connect the task with a recognizable and meaningful activity from their own experiences.

Nevertheless, despite showing that GLEs increase PP [19], no study has specifically investigated the contribution of visual

XSL•FO
RenderX

GLEs integrated with a narrative context within cognitive tasks and the impact on both PP and positive mental effort. Thus, our first aim (hypothesis 1) is to ascertain if visual GLEs integrated with a narrative context (which will be called a playful task) are enough to increase PP and then positive mental effort (characterized by higher perceived mental effort and performance but lower perceived frustration) when compared with a classic design. And therefore, we predicted that this increase in voluntary mental effort could also increase real performances.

Nevertheless, in everyday life, an individual must involve various amounts of cognitive functions to meet the requirements of an activity. The number of cognitive functions solicited is a variable intrinsic to task requirements which would have a direct impact on mental demands. Mental demands is the dimension of MWL directly related to the amount of mental activities required [2]. This could also impact mental effort imposed by the task itself. We can thus assume that this effect could also play a role in positive mental effort, notably via perceived performance and frustration. Thus, an increase in mental effort required by the task could lead to a decrease in positive mental effort. To our knowledge, no study has investigated the impact of increasing the solicitation of cognitive functions on PP, positive mental effort, or the resulting actual performances in a narrative ecological context. Thus, our second aim (hypothesis 2) is to determine if the multiplicity of cognitive solicitation of the playful task might decrease PP, positive mental effort (characterized by higher perceived effort and frustration but fewer perceived performance), and actual performances when compared with the simple playful task.

Hence, this study will explore how extrinsic features (visual GLEs within a narrative context) impact PP and positive mental effort, and whether this effect persists when manipulating intrinsic demands of the task (eg, the multiplicity of cognitive solicitation in a narrative context). To this end, we have set up a protocol containing various cognitive tasks. In the first step, we will consider the impact of adding GLEs to a standard cognitive task. Then we will investigate further the impact of adding cognitive solicitation to this task. The tasks and the logic behind their design are detailed in the Methods section.

Methods

Ethical Considerations

This study has received a positive decision from the Northwest III Ethics Commitee for the Protection of Persons, with the reference number n° 21.04612.000059, in accordance with the ethical guidelines for human participants research and the Declaration of Helsinki. Informed consent was obtained from all participants. All data were anonymized to ensure privacy and were securely stored on a protected cloud platform accessible only to the research team. No identifiable participant information is included in the manuscript or supplementary materials. At the end of the study, each participant received a €50 (US $54.69) gift voucher as compensation.

Participants

In total, 20 (9 women, 11 men) took part in this experiment. Their average age was 33.6 (SD 8.6) years. All of them had at least a bachelor's degree and 85% (17/20) of them had at least a bac+3 level. Furthermore, 70% (14/20) worked for Onepoint company, 20% (4/20) came from social network LinkedIn (Microsoft) and the others were acquaintances of the experimenters or from Nantes University Hospital.

Recruitment

Participants were recruited by mailing an information letter inviting interested people to contact the experimenter if they wanted to participate in the study. At the reception of the mail, the experimenter contacted the participant in order to confirm (by oral statement) that he or she were healthy French-speaking volunteers residing in Nantes region that met the inclusion criteria (fluent in French, right-handed, having normal or corrected-to-normal vision, and having normal or corrected-to-normal hearing). Then the participant was informed about the experimental duration and procedure. After acceptance by the participant, an appointment was made for the experimental tests. The consent document was signed at the participant's arrival, before the beginning of the experiment.

Tasks and Procedure

Overview

The experimentation took place at Nantes University Hospital. Participants performed 3 experimental tasks in a random order on a screen positioned in front of them. All tasks were devised using Unity3D development software (version 2019.4.30f1; Unity Technologies). A link to the videos of all tasks is available in Multimedia Appendix 1.

To determine the level of complexity of the tasks, we considered several factors. Primarily, the complexity of a simple task (only soliciting 1 cognitive function) should not be arduous. Indeed, a simple task should accommodate the integration of an additional cognitive function in a condition that involves multiple cognitive functions. On the other hand, we did not want the simple task to appear too simplistic. Thus, we relied on the intermediate complexity levels from the study conducted by Louis et al [21], which offers a compromise considering the aforementioned conditions. Indeed, Louis et al [21] compared several cognitive tasks with different complexity levels, and the 5-item Corsi test induced an intermediate MWL (based on NASA-TLX) while still allowing high performance. Thus, the 5-item Corsi test is a sufficiently cognitively demanding task without being difficult. Consequently, to verify hypothesis 1 (visual GLEs integrated with a narrative context increase PP, positive mental effort, and performances) and 2 (the multiplicity of cognitive solicitation might decrease PP, positive mental effort, and performances), 3 digitized versions of the Corsi test were proposed, that are Classic Corsi [22], Playful Corsi, and Playful Corsi Multi.

Classic Corsi

This task [22] is a visual-spatial memory span task involving remembering a sequence of cubes pointed by the computer and reproducing the sequence in the same order by clicking with

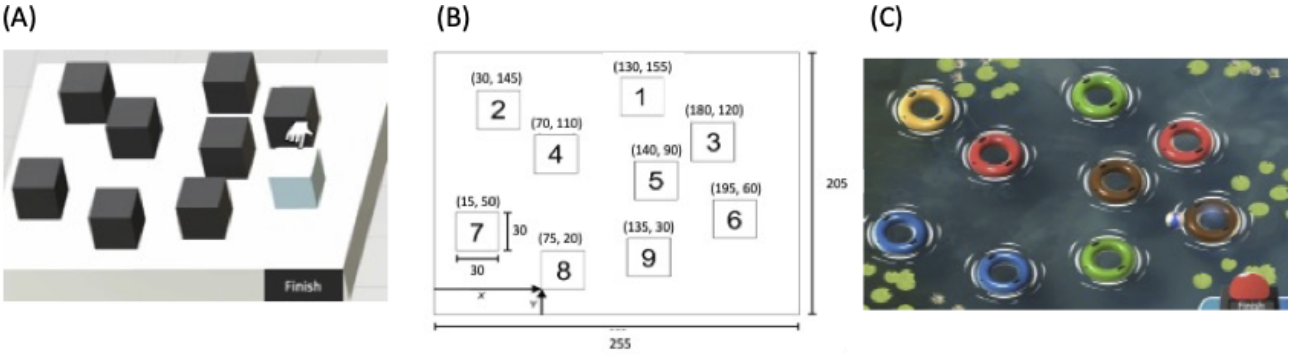
XSL•FO
RenderX

the mouse on screen elements. Thus, the cubes have been selected by a mouse click; in the following sections, we will use the term "clicked cubes" to designate this action. The material (Figure 1A) used is consistent with the study by Kessels et al [23] that proposed for each cube a number which is not visible by the participants (Figure 1B). Furthermore, 5 trials involving the memorization of 5 cubes (intermediate level as adopted from a study by Louis et al [21]) are proposed. The five sequences of cubes were (1) 5-2-1-8-6, (2) 4-2-7-3-1, (3) 3-9-2-4-8, (4) 7-8-2-9-4, and (5) 4-2-6-8-1 (Figure 1B).

The sequence of cubes for the training session was 2 - 6.

Performance dimensions include clicked cubes (number of cubes that are clicked by the participant), correct cubes (the cubes correctly recalled whatever the order), false cubes (the cubes incorrectly recalled), omissions (failure to press a cube), total time, and expected responses (common performance indicator for all tasks) corresponds to this calculation: 100×(25–(False cubes+Omissions)÷25) and exact sequences (number of cubes clicked in the expected order and position) which can be compared with the terminology of memory or block span, corresponding to the length of the last correctly repeated sequence, in a traditional Corsi test [23]. However, our task did not aim to determine a memory span, as is typically the case in Corsi tests. In a memory span assessment, the goal is to identify the maximum number of items a participant can correctly retain and reproduce. In contrast, our task focused on exact sequences, and the instructions provided to participants did not explicitly emphasize memory span as a performance objective. Therefore, we have chosen to retain the term exact sequences to better reflect the specific nature and goals of our task.

Figure 1. (A) Visual environment of the digitized version of the Classic Corsi test. (B) Positions of cubes in the Classic Corsi test from study by Kessels et al [23]. (C) Playful version of the digitized version of the Classic Corsi test: cubes are replaced by 9 buoys with different colors integrated in a narrative context (throwing balls towards buoys floating in a pond).



Playful Corsi

This task is identical to the classic version, but cubes are replaced by 9 buoys with different colors integrated in a narrative context (throwing balls toward buoys floating in a pond; Figure 1C). This task uses the same performance metrics as Classic Corsi.

To verify hypothesis 2 (the multiplicity of cognitive solicitation might decrease PP, positive mental effort, and performances), we needed a task that solicited another cognitive function. We called it "Playful Corsi Multi."

Playful Corsi Multi

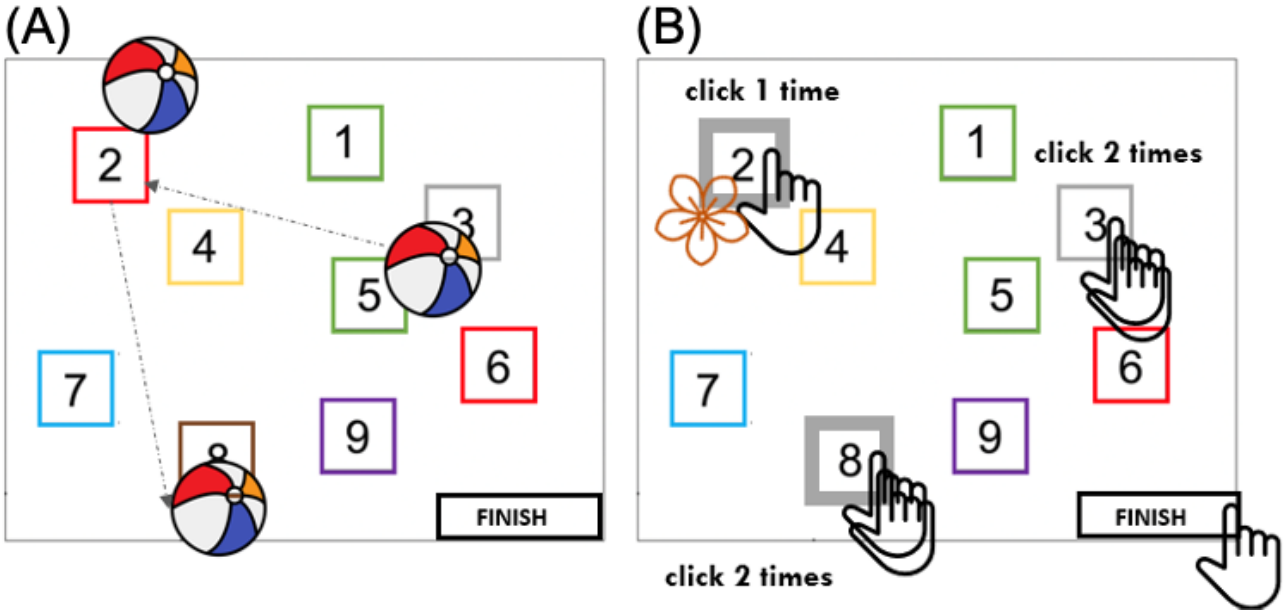
Following several pretests in which various cognitive functions (such as mental flexibility or working memory updating) were compared, it was observed that mental motor inhibition, as required in tasks such as go no-go [24], was the most suitable cognitive function to integrate with the Corsi test. This choice minimized the impact on the extrinsic characteristics of the task while maintaining its core objectives.

The Playful Corsi Multi uses the same material as Playful Corsi (Figure 1). Participants memorize a sequence of 5 ball throws, reproducing it by clicking twice on buoys. During sequence reproduction, flowers may appear on buoys that must be clicked. In this case, participants do not have to make the second click on the gray buoy with a flower (no-go), and the response is validated with a single click (Figure 2). Furthermore, 4 pointing sequences with 2 no-go targets (gray buoy with flower) are proposed. In all sequences, a "trap" buoy with a flower on a nongray buoy is present and must be clicked twice, as similarly required for a buoy (of any color) without a flower. The sequence of 5 pointed buoys for the training session was with 1 no-go and 1 ''trap'' buoy with a flower on a nongray buoy.

Performance dimensions encompass false buoys, false alarms (2 clicks instead of 1 upon the appearance of a no-go target), omissions (failure to perform a second click on a buoy), clicked buoys, exact buoys (number of buoys clicked in the expected order and position), expected responses (correspond to the accuracy and are calculated as 100×[25−(False alarms+omissions)÷25], *exact sequences* (number of buoys clicked in the expected order and position), reaction time for correct responses (the buoys correctly recalled), reaction time for all items combined, and total time (between the appearance of the red buzzer [finish button] and pressing it).

XSL•FO
RenderX

Figure 2. Schematic figures of the Playful Corsi Multi. (A) Ball displacement corresponding to memorization instructions given to participants; (B) actions required by the participants to respond to the instructions (2 clicks if there is no flower on the buoys and 1 click if there is a flower on the buoys).



Subjective Measures

Studies by Rubio et al [25] and Paxion [26] indicate that WP and NASA-TLX questionnaires complement each other. Furthermore, our study aimed to assess MWL related to attentional resources (WP questionnaire) and cognitive functions, emphasizing mental demands and mental effort (NASA-TLX questionnaire).

NASA-TLX Questionnaire

NASA-TLX [2] evaluates perceived workload across 6 distinct subscales. In total, 3 dimensions relate to activity, that are mental demands (requirements for mental and perceptual activity), physical demands (requirements for physical activity), and temporal demands (sense of time pressure), 2 dimensions relate to strategies, that are performance (feeling of success in achieving the objectives set by the experimenter or oneself) and effort (the mental or physical effort required to reach the performance level targeted), and 1 dimension relate to emotional state, that is frustration (irritation, stress, insecurity, and discouragement). Following each task level, participant score each dimension from 0 (no demand) to 100 (maximum demand). In our study, we used the unweighted version of the questionnaire by calculating an average of the 6 dimensions to derive a raw task load index validated by Byers et al [27].

WP Questionnaire

The WP questionnaire [4] asked the participants to provide the proportion of attentional resources across 8 subscales. Indeed, the dimensions of workload in this method were aligned with those in the multiple resource theory (described in Introduction). According to Tsang and Velazquez [4], 2 dimensions of WP questionnaire relate to “stages of processing”: perceptual or central processing (attentional resources required for activities like perceiving, remembering, problem-solving, and decision making) and response selection and execution (attentional resources required for response selection and execution; eg, the selection of the appropriate pedal to stop an automobile). Furthermore, 2 other dimensions are related to “processing codes”: spatial processing (according to Tsang and Velazquez [4], for tasks that are spatial in nature) and verbal processing (for tasks that are verbal in nature). In addition, 2 other dimensions are related to “input modality”: visual processing (for tasks that are performed based on the visual information received) and auditory processing (for tasks that are performed based on the auditory information received). And finally, 2 dimensions are related to “output modalities”: manual output (for tasks that required considerable attention for producing manual response), and speech output (for other tasks requiring considerable attention for producing speech responses).

Participants score each dimension from 0 (no demand) to 100 (maximum demand) representing the proportion of allocated attentional resources.

To reduce the time spent on questionnaires and given the absence of problem solving but also considering that the NASA-TLX’s mental demand dimension also questioned this source of MWL, the dimension perceptual or central processing was not considered. For the same reasons, because of the absence of actioners’ selection, auditory and speech tasks in our study, the dimensions response selection and execution, auditory processing, and speech output were not considered. Thus, only 4 of the 8 dimensions were evaluated. Indeed, as demonstrated by Tsang and Velazquez [4], the dimensions can be considered independently and rated in a 1D way. Therefore, we decided to focus on the spatial processing (WP3) and visual processing (WP5) as Corsi test is a visuo-spatial task, verbal processing (WP4) as the participants mentally verbalize the go no-go items (eg, the buoys colors), and manual output (WP7) dimensions, in relation to the participant’s movement of the mouse to click on the various items.

PP Measure

Regarding PP, we assessed it using a question presented on a Likert scale ranging from 0 to 100: “To what extent did you find the task you have just completed playful and fun?”

XSL•FO
RenderX

Experimental Procedure

After receiving approval from the Northwest III Ethics Committee for the Protection of Persons, the recruitment started. Potential participants received a study summary via email, followed by individual meetings held 7 days before the Nantes University Hospital study. Then, participants signed a consent letter. The experimental phase proceeded as follows: Participants were welcomed at Nantes University Hospital. Their eligibility and nonopposition to participation in the study were verified and they completed a pretest questionnaire (age, gender, and level of education). Then, they began the experiment. Each task had a training session (refer to Task and Procedure section) and after each task level, participants received NASA-TLX and WP questionnaires, and PP measure. Including questionnaires, each Corsi test (classic and playful one) has a total duration of 11 min and Playful Corsi Multi has a total duration of 16 min.

Data Analysis

Raw data is available in Multimedia Appendix 2. Subjective measures (NASA-TLX, WP, and PP) were normalized through Cousineau correction [28]. This normalization allows for more accurate estimation of within - participants CIs for subjective data in the case of experimental designs with several conditions. Cousineau correction was described as follows. Let y_{ij} be the ith participant's score in the jth condition (i=1, ..., N; j=1, …, M). Then, the normalized observations z_{ij} were defined as follows:

$$zij=yij-1N\sum j=1Myij+1NM\sum i=1N\sum j=1Myij$$

Furthermore, some distributions of our data were not normal, and variances were not homogeneous. Thus, we used nonparametric statistics (Wilcoxon signed-rank tests and Conover post hoc comparisons) with the JASP (Jeffrey's Amazing Statistics Program) software (version JASP 0.16; University of Amsterdam).

Finally, for the P value, the lower error threshold of .05 was considered significant.

Results

Comparisons Between Classic Corsi and Playful Corsi

Performance Measures

For the Corsi tests, 2 relevant performance dimensions appeared: expected responses and total time. The other Corsi performances (clicked cubes, correct cubes, false cubes, omissions, and exact sequences) were included within expected responses dimension (refer to Methods section).

Between the 2 tasks, expected responses and total time were significantly higher for the Classic Corsi than the Playful Corsi (Table 1; Figure 3).

Table . Statistical comparisons between Classic Corsi and Playful Corsi concerning performances data and mental workload questionnaires.

	Classic Corsi, mean (SD)	Playful Corsi, mean (SD)	W	*P* value
Expected responses	92.4 (9.1)	88.2 (11.3)	140.5	.02
Total time	10.7 (2.1)	6.8 (1.6)	209	<.001
Perceived playfulness	62.4 (8.8)	66 (8.8)	77	.30
Overall	187 (52.9)	186.3 (52.8)	65	.23
Mental demands	46.6 (13.8)	54.2 (13.9)	53	.16
Physical demands	13.6 (5.9)	13.4 (5.2)	36	.82
Temporal demands	35.6 (11.3)	38.9 (15.4)	93	.95
Effort	39.5 (11.7)	43.8 (13.6)	71	.54
Perceived performance	23.7 (10.4)	30.4 (9.9)	55.5	.19
Frustration	27.8 (16.8)	35.8 (16.8)	64.5	.23
WP3[a]	52.4 (12.1)	59.1 (12.1)	85.5	.48
WP5[b]	52.4 (10.9)	65.8 (10.9)	27.5	.04
WP7[c]	15.3 (12.8)	23.6 (12.8)	42.5	.30

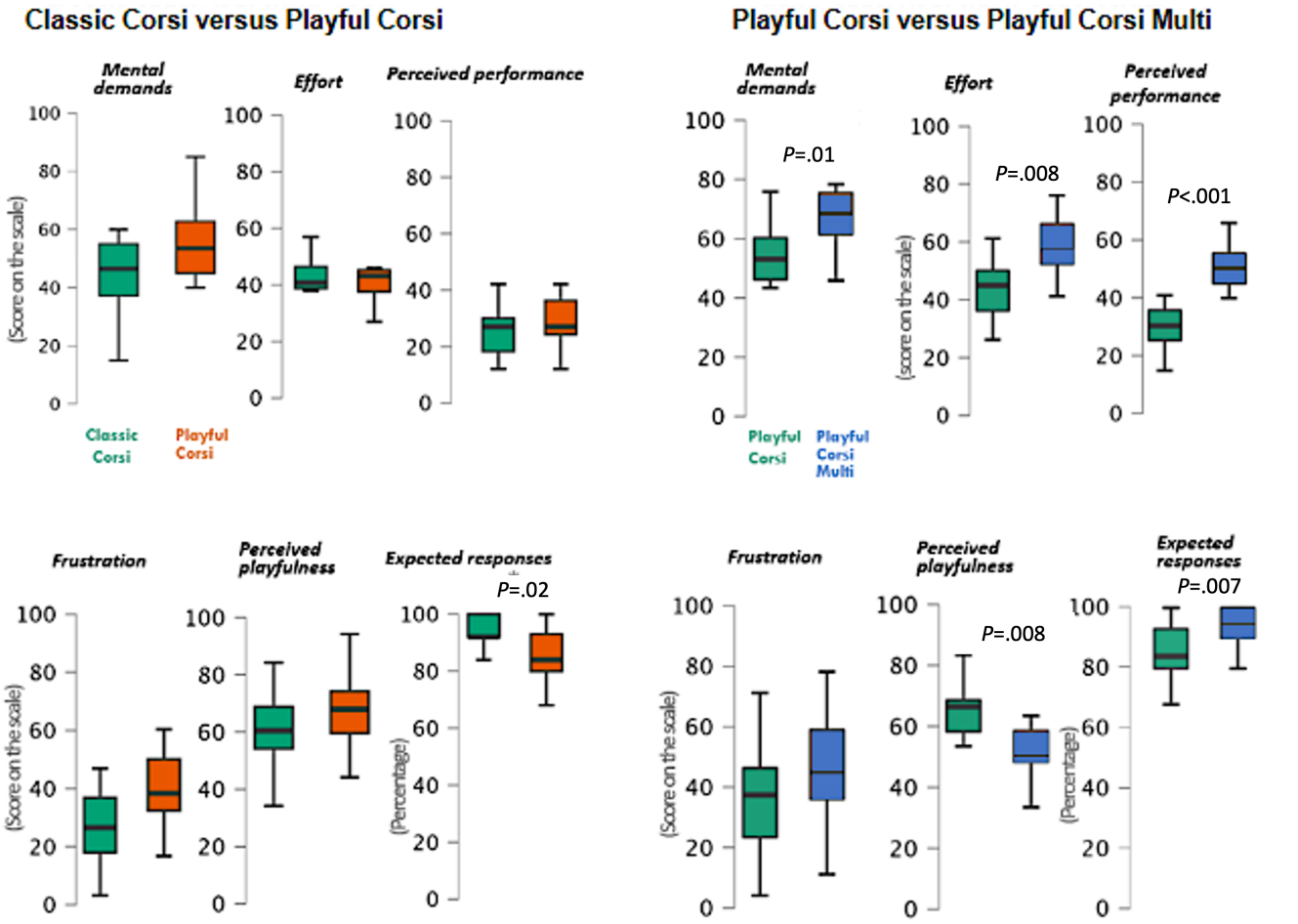
[a]WP3: workload profile on spatial processing.

[b]WP5: workload profile on visual processing.

[c]WP7: workload profile on manual output.

XSL•FO
RenderX

Figure 3. Comparisons of mental demands, effort, perceived performance, frustration, perceived playfulness, and expected responses between Classic Corsi and Playful Corsi and between Playful Corsi and Playful Corsi Multi.



NASA-TLX Questionnaire

Comparing Classic Corsi and Playful Corsi, the overall NASA-TLX score (corresponding to the sum of the 6 NASA-TLX dimensions which are on a scale of 600) showed no significant difference (Table 1), also for mental demands, physical demands, temporal demands, effort, performance, and frustration.

WP Questionnaire

We did not consider the WP4-verbal dimension since it was not a solicited dimension by the Corsi test, a visuo-spatial task (but it was solicited by Multi Playful Corsi test that contained mental inhibition based on the item colors; described in more detail in the Multicolored Playful Corsi vs Playful Corsi Multi section). Between the Classic Corsi and the Playful Corsi, only the WP5-visual dimension was significantly higher for the Playful Corsi than for the Classic Corsi (Table 1). But, between the 2 conditions, WP3-spatial and WP7-manual showed no significant difference (Table 1).

PP Measure

Between the 2 tasks, PP showed no significant difference (Table 1).

Multicolored Playful Corsi vs Playful Corsi Multi

Performance Measures

Total time was not considered, as in the Playful Corsi Multi, there are 2 clicks per response, instead of 1 for the Multicolored Playful Corsi. Therefore, a trial-by-trial comparison was not possible. Concerning expected responses between the 2 tasks, they were significantly better for Playful Corsi Multi than Playful Corsi (Table 2; Figure 3).

XSL•FO
RenderX

Table . Statistical comparisons between Playful Corsi and Playful Corsi Multi concerning performances data and mental workload questionnaires.

	Playful Corsi, mean (SD)	Playful Corsi Multi, mean (SD)	W	*P* value
Expected responses	83.8 (13.9)	94 (5.5)	27	.007
Perceived playfulness	66 (11.9)	50.9 (11.9)	161	.008
Overall	216.6 (40.5)	284.4 (40.5)	25.5	.003
Mental demands	54.2 (9.4)	67.5 (9.4)	35.5	.01
Physical demands	13.4 (8.9)	15.4 (8.9)	33	.40
Temporal demands	38.9 (12.4)	46.6 (12.4)	56	.20
Effort	43.8 (9.9)	58.1 (9.9)	29	.008
Perceived performance	30.4 (7)	50.2 (7)	1	<.001
Frustration	35.8 (17.7)	46.5 (17.7)	63	.20
WP3[a]	59.1 (12.6)	64.4 (12.6)	81.5	.40
WP4[b]	10.2 (11.7)	16.3 (11.7)	14	.30
WP5[c]	65.8 (7.4)	68.3 (7.4)	71	.50
WP7[d]	23.6 (14.7)	21.1 (14.7)	93	.80

[a]WP3: workload profile on spatial processing.

[b]WP4: workload profile on verbal processing.

[c]WP5: workload profile on visual processing.

[d]WP7: workload profile on manual output.

NASA-TLX Questionnaire

The overall MWL was significantly higher for Playful Corsi Multi than Playful Corsi (Table 2; Figure 3), as were the mental demands, effort, and performance. However, physical demands, temporal demands, and frustration showed no significant difference between the 2 tasks.

WP Questionnaire

Between the tasks, the WP3, WP4, WP5, and WP7 showed no significant difference (Table 2; Figure 3).

PP Measure

PP was significantly higher for the Playful Corsi compared to the Playful Corsi Multi (Table 2; Figure 3).

Discussion

Principal Objectives

In this study we wanted to explore (1) how extrinsic variable (visual GLEs within a narrative context) impacts PP, positive mental effort, and performances at the very first stage of facing with a task devoted to cognitive training, and (2) if this effect endures while manipulating intrinsic variable (multiplicity of cognitive solicitation in a narrative context).

Impact of Visual GLEs on PP and Positive Mental Effort

Our first hypothesis that visual GLEs integrated with a narrative context increase PP and positive mental effort is invalidated for various reasons. First, we observed no difference in terms of PP between Classic Corsi and its playful version. In the classic version of the Corsi, the PP is not poor but falls short of optimal. Consequently, we could have potentially improved the PP by manipulating the visual GLEs, but this was not the case. In this way, the PP did not seem to depend on the visual aspect of the task.

Concerning the supposed action of visual GLEs on positive mental effort, it appears that mental effort, perceived performance, and frustration did not differ between the 2 tasks. Thus, visual GLEs are not enough increasing positive mental effort.

On the contrary, MWL related to the mobilization of visual resources increased when we manipulated the extrinsic variables to task requiring (presence of visual GLEs). This is not surprising as, in the Playful Corsi, there are additional visual elements, such as waves or water lilies, which increase the MWL based on visual processing. Nevertheless, no other difference was noted about spatial processing and manual output, indicating that the tasks are relatively similar on the other dimensions.

Concerning the impact on effective performances, visual GLEs decreased the total time compared with a classic design, potentially attributed to the inherent challenge of game-like environments. There are no speed instructions for this task. Speed is therefore neither a performance criterion nor an objective to be achieved. The increase in speed observed in the playful condition may be due to an implicit incentive to accelerate, given the game-like design of the task, where speed is often an inherent goal. Concerning the expected responses (corresponded to the accuracy), our result contradicted the study by Redlinger et al [11] which specified that visual GLEs did not affect performance. Indeed, in the study by Redlinger et al

XSL•FO
RenderX

[11], the presence of backgrounds led to slight reductions in accuracy but not significantly. In our study, this result is to consider in light to the observed impact on the response time. Indeed, we can assume that going faster leads to more errors. This can be linked to the well-known speed-accuracy trade-off [29], that is, the accuracy of a response varies with the time taken to produce it. Furthermore, given that in our study, the visual GLEs included changes to the background, additionally to the object on which participants had to interact (cubes or buoys), this could have induced greater distraction for participants and therefore significant differences in performance in favor of the Classic Corsi.

In conclusion, our results were in line with the meta-analysis of Vermeir et al [9], indicating that gamified tasks are perceived as significantly more demanding, at least considering the perceived demand. Furthermore, this gamification is not accompanied with a voluntary mental effort improvement nor a better PP. A deterioration in performance has even been observed. So, adding visual GLEs with narrative context is not always beneficial for the task.

Impact of Increasing Cognitive Demanding

Our second hypothesis that the multiplicity of cognitive demanding of the playful task might decrease PP and positive mental effort is partially confirmed. PP is negatively influenced by intrinsic variables (solicitation of multiple cognitive functions) but there is every reason to believe that this manipulation has increased the positive mental effort. This is supported by a higher perceived mental effort in conjunction with a higher perceived performance (confirmed by better effective performances) but no increase in the perceived frustration.

Concerning the decrease in PP, increasing the solicitation of cognitive processing increases mental demand, which could have a negative impact on PP. Indeed, Fang et al [17] and Donovan et al [30] showed that PP had a significant and negative correlation with task complexity. In our study, this could also stem from the way we question the notion of PP: “To what extent did you find the task you have just completed playful and fun?” As the task becomes more cognitively demanding, it can be perceived as more “serious,” which could reduce the feeling of “playful.” Also, we acknowledge the limitation of combining 2 descriptors (“playful” and “fun”) in a single question, as they may not hold the same meaning for all participants. “Playful” could refer to the task’s design or its gamified elements, while “fun” might reflect the participant’s subjective enjoyment or emotional response. This overlap introduces potential ambiguity, as participants may weigh one descriptor more heavily than the other. For the next studies, we will consider separating these descriptors into distinct questions to capture more nuanced insights into PP and enjoyment, thereby reducing interpretive variability.

Nevertheless, even though the augmentation of the complexity of a cognitive task decreases the PP, this manipulation increases mental effort and perceived performance without increasing frustration. This is the opposite of what we expected to see. This can stem from the fact that “serious” is not antagonistic to “pleasant,” “enjoyable,” or “motivating.” It just means that what would help induce greater voluntary mental effort, would correspond more to stronger cognitive stimulation (agreed) than just the “fun” characteristics of an activity. This means that, to encourage the voluntary mobilization of mental resources, one possibility is to manipulate the cognitive demanding of a task rather than its design only. Here, by “cognitive demanding,” we are talking about cognitive challenge, not difficulty. In fact, the actual performance, which is better in this condition, shows us that the task is not more complex to perform. But everything seems to indicate that the participant is more determined to perform. Such a performance boost, resulting from cognitive function multiplicity, could be a drawback when performing cognitive assessment but is likely to be desirable in a cognitive training scenario [14]. Indeed, in cognitive training scenarios, the goal is to enhance overall cognitive functioning by engaging multiple cognitive processes simultaneously since it mimics the complexity and demands of real-life situations where multiple cognitive functions are often required at once.

Therefore, to encourage the voluntary mobilization of mental resources, based on our results, a possible approach could be to manipulate the cognitive demanding of a cognitive task (increasing mental effort and performance without increasing frustration) rather than its design but allowing it to maintain a high level of performance.

Limitations

Finally, it is important to specify that this study should be extended to include a larger number of participants to strengthen statistical significance. Given that the experiments were carried out at the Nantes Hospital, which is far from the city center and has few medical staff available, we were constrained by logistical considerations. In addition, the choice of terms used for measuring PP was not without ambiguity and may require refinement in future studies.

Conclusion and Implications for Research

This study can provide cognitive testing practitioners or even social game practitioners with valuable information on specific game-like guidelines to impact users’ PP, while taking into consideration the impact on MWL. We showed that the PP is a skillful balance between design and the solicitation of cognitive functions. Possibly, visual GLEs could raise the stakes of the task slightly and implicitly encourage people to go a bit faster (as observed with the Corsi test), but working on the way cognitive functions are solicited would be wiser to improve users’ voluntary mental effort than adding GLEs. Future research could explore further this last hypothesis as well as the influence of additional GLEs, such as sounds, variations in narrative context, or performance feedback.

XSL•FO
RenderX

Acknowledgments

The authors want to thank all participants for voluntarily taking part in this study and Sébastien Maire who implemented the task battery. The authors also thank Nantes Hospital, where the experiments were conducted, and more particularly, Aurélien Van Langhenhove and Vincent Roualdes for their hospitality, expertise, and discussions.

Data Availability

The data that support the findings of this study are available from the corresponding author (IMP), upon reasonable request.

Authors' Contributions

LEL contributed to the conceptualization, methodology, investigation, data curation, formal analysis, writing – original draft, visualization, and project administration. IMP contributed to the conceptualization, methodology, formal analysis, writing – review and editing, project administration, and supervision. SM handled the resources and supervision. SR handled the supervision and funding acquisition.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Video demonstration of the tasks used in the study.
[MP4 File, 107500 KB - formative_v9i1e63491_app1.mp4]

Multimedia Appendix 2
Raw data on how to encourage the voluntary mobilization of mental resources in a cognitive task.
[XLSX File, 36 KB - formative_v9i1e63491_app2.xlsx]

References

1. Kahneman D. Attention and Effort: Prentice-Hall; 1973.
2. Hart SG, Staveland LE. Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock PA, Meshkati N, editors. Human Mental Workload: North-Holland Press:139-183. [doi: 10.1016/S0166-4115(08)62386-9]
3. Delignières D, Deschamps T. L'effort mental [Book in French]. In: Delignières D, editor. L'effort: Revue EPS; 2000:25-40.
4. Tsang PS, Velazquez VL. Diagnosticity and multidimensional subjective workload ratings. Ergonomics 1996 Mar;39(3):358-381. [doi: 10.1080/00140139608964470] [Medline: 8849491]
5. Wickens CD. Processing resources in attention. In: Parasuraman R, Davis DR, editors. Varieties of Attention: Academic Press; 1984:63-101.
6. Wickens CD. Multiple resources and performance prediction. Theor Issues Ergon Sci 2002 Jan;3(2):159-177. [doi: 10.1080/14639220210123806]
7. Wickens CD. Multiple resources and mental workload. Hum Factors 2008 Jun;50(3):449-455. [doi: 10.1518/001872008X288394] [Medline: 18689052]
8. Rabipour S, Raz A. Training the brain: fact and fad in cognitive and behavioral remediation. Brain Cogn 2012 Jul;79(2):159-179. [doi: 10.1016/j.bandc.2012.02.006] [Medline: 22463872]
9. Vermeir JF, White MJ, Johnson D, Crombez G, Van Ryckeghem DML. The effects of gamification on computerized cognitive training: systematic review and meta-analysis. JMIR Serious Games 2020 Aug 10;8(3):e18644. [doi: 10.2196/18644] [Medline: 32773374]
10. Gobet F, Sala G. Cognitive training: a field in search of a phenomenon. Perspect Psychol Sci 2023 Jan;18(1):125-141. [doi: 10.1177/17456916221091830] [Medline: 35939827]
11. Redlinger E, Glas B, Rong Y. Impact of visual game-like features on cognitive performance in a virtual reality working memory task: within-subjects experiment. JMIR Serious Games 2022 Apr 28;10(2):e35295. [doi: 10.2196/35295] [Medline: 35482373]
12. Ninaus M, Pereira G, Stefitz R, et al. Game elements improve performance in a working memory training task. IJSG 2015;2(1). [doi: 10.17083/ijsg.v2i1.60]
13. Mohammed S, Flores L, Deveau J, et al. The benefits and challenges of implementing motivational features to boost cognitive training outcome. J Cogn Enhanc 2017 Dec;1(4):491-507. [doi: 10.1007/s41465-017-0047-y] [Medline: 30221244]
14. Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR. Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. JMIR Serious Games 2016 Jul 15;4(2):e11. [doi: 10.2196/games.5888] [Medline: 27421244]

XSL•FO
RenderX

15. Moon JW, Kim YG. Extending the TAM for a World-Wide-Web context. Information & Management 2001 Feb;38(4):217-230. [doi: 10.1016/S0378-7206(00)00061-6]
16. Hung SY, Tsai JCA, Chou ST. Decomposing perceived playfulness: a contextual examination of two social networking sites. Information & Management 2016 Sep;53(6):698-716. [doi: 10.1016/j.im.2016.02.005]
17. Fang X, Chan S, Brzezinski J, Xu S. A study of task characteristics and user intention to use handheld devices for mobile commerce. Presented at: The 2nd Annual Workshop on HCI Research in MIS; Dec 11-13, 2003; Seattle, WA, United States.
18. Lin CS, Wu S, Tsai RJ. Integrating perceived playfulness into expectation-confirmation model for web portal context. Information & Management 2005 Jul;42(5):683-693. [doi: 10.1016/j.im.2004.04.003]
19. Codish D, Ravid G. Academic course gamification: the art of perceived playfulness. IJELL 2014;10(1):131-151. [doi: 10.28945/2066]
20. Bertin J. Semiology of Graphics: Diagrams, Networks, Maps: University of Wisconsin Press:983.
21. Louis LEL, Moussaoui S, Van Langhenhove A, et al. Cognitive tasks and combined statistical methods to evaluate, model, and predict mental workload. Front Psychol 2023;14:1122793. [doi: 10.3389/fpsyg.2023.1122793] [Medline: 37251030]
22. Corsi PM. Human memory and the medial temporal region of the brain [PhD thesis]. : McGill University; 1972.
23. Kessels RP, van Zandvoort MJ, Postma A, Kappelle LJ, de Haan EH. The Corsi Block-Tapping Task: standardization and normative data. Appl Neuropsychol 2000;7(4):252-258. [doi: 10.1207/S15324826AN0704_8] [Medline: 11296689]
24. Vidal J, Mills T, Pang EW, Taylor MJ. Response inhibition in adults and teenagers: spatiotemporal differences in the prefrontal cortex. Brain Cogn 2012 Jun;79(1):49-59. [doi: 10.1016/j.bandc.2011.12.011] [Medline: 22325785]
25. Rubio S, Díaz E, Martín J, Puente JM. Evaluation of subjective mental workload: a comparison of SWAT, NASA - TLX, and workload profile methods. Applied Psychology 2004 Jan;53(1):61-86. [doi: 10.1111/j.1464-0597.2004.00161.x]
26. Paxion J. Complexite´ des situations, expe´rience, tension et vigilance: quels impacts sur la charge de travail et les performances de jeunes conducteurs. sciences cognitives [Thesis in French]. : Aix-Marseille Université; 2014.
27. Byers JC, Bittner AC, Hill SG. Traditional and raw task load index (TLX) correlations: are paired comparisons necessary? In: Mital A, editor. Advances in Industrial Ergonomics and Safety: Taylor and Francis; 1989:481-485.
28. Morey RD. Confidence intervals from normalized data: a correction to Cousineau (2005). TQMP 2008;4(2):61-64. [doi: 10.20982/tqmp.04.2.p061]
29. Heitz RP. The speed-accuracy tradeoff: history, physiology, methodology, and behavior. Front Neurosci 2014;8:150. [doi: 10.3389/fnins.2014.00150] [Medline: 24966810]
30. Donovan JJ, Hafsteinsson LG, Lorenzet SJ. The interactive effects of achievement goals and task complexity on enjoyment, mental focus, and effort. J Applied Social Pyschol 2018 Mar;48(3):136-149. [doi: 10.1111/jasp.12498]

Abbreviations

GLE: game-like element
JASP: Jeffrey's Amazing Statistics Program
MWL: mental workload
NASA-TLX: NASA-Task Load Index
PP: perceived playfulness
WP: workload profile
WP3: workload profile on spatial processing
WP4: workload profile on verbal processing
WP5: workload profile on visual processing
WP7: workload profile on manual output

Edited by A Mavragani; submitted 24.06.24; peer-reviewed by LK Chua, P Harvey; revised version received 07.02.25; accepted 08.02.25; published 10.04.25.

Please cite as:
Louis LE, Moussaoui S, Ravoux S, Milleville-Pennel I
Encouraging the Voluntary Mobilization of Mental Resources by Manipulating Task Design: Explorative Study
JMIR Form Res 2025;9:e63491
URL: https://formative.jmir.org/2025/1/e63491
doi:10.2196/63491

© Lina-Estelle Louis, Saïd Moussaoui, Sébastien Ravoux, Isabelle Milleville-Pennel. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction

XSL•FO
RenderX

in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Support of Home-Based Structured Walking Training and Prediction of the 6-Minute Walk Test Distance in Patients With Peripheral Arterial Disease Based on Telehealth Data: Prospective Cohort Study

Fabian Wiesmüller[1,2,3], MSc; Andreas Prenner[4], Dr; Andreas Ziegl[5], PhD; Gihan El-Moazen[3], BSc; Robert Modre-Osprian[5], PhD; Martin Baumgartner[1,3], PhD; Marianne Brodmann[4], MD; Gerald Seinost[4], MD; Günther Silbernagel[4], MD; Günter Schreier[1], PhD; Dieter Hayn[1,2], PhD

[1]Center for Health & Bioresources, AIT Austrian Institute of Technology GmbH, Graz, Austria

[2]Ludwig Boltzmann Institute for Digital Health and Prevention, Ludwig Boltzmann Gesellschaft, Salzburg, Austria

[3]Institute of Neural Engineering, Graz University of Technology, Graz, Austria

[4]Division of Angiology, Department of Internal Medicine, Medical University of Graz, Graz, Austria

[5]telbiomed Medizintechnik und IT Service GmbH, Graz, Austria

Corresponding Author:
Fabian Wiesmüller, MSc
Center for Health & Bioresources, AIT Austrian Institute of Technology GmbH, Graz, Austria

Abstract

Background: Telehealth has been effective in managing cardiovascular diseases like stroke and heart failure and has shown promising results in managing patients with peripheral arterial disease. However, more work is needed to fully understand the effect of telehealth-based predictive modeling on the physical fitness of patients with peripheral arterial disease.

Objective: For this work, data from the Keep Pace study were analyzed in depth to gain insights on temporal developments of patients' conditions and to develop models to predict the patients' total walking distance at the study end. This could help to determine patients who are likely to benefit from the telehealth program and to continuously provide estimations to the patients as a motivating factor.

Methods: This work analyzes continuous patient-reported telehealth data, in combination with in-clinic data from 19 Fontaine stage II patients with peripheral arterial disease who underwent a 12-week telehealth-based walking program. This analysis granted insights into the increase of the total walking distance of the 6-minute walk tests (6MWT) as a measure for physical fitness, the steady decrease in the patients' pain, and the positive correlation between well-being and the total walking distance measured by the 6MWT.

Results: This work analyzed trends of and correlations between continuous patient-generated data. Findings of this study include a significant decrease of the patients' pain sensation over time (P=.006), a low but highly significant correlation between pain sensation and steps taken on the same day (r=−0.11; P<.001) and the walking distance of the independently performed 6MWTs (r=−0.39; P<.001). Despite the reported pain, adherence to the 6MWT measurement protocol was high (85.53%). Additionally, patients significantly improved their timed-up-and-go test times during the study (P=.002). Predicting the total walking distance at the study end measured by the 6MWT worked well at study baseline (root mean squared error of 30 meters; 7.04% of the mean total walking distance at the study end of 425 meters) and continuously improved by adding further telehealth data. Future work should validate these findings in a larger cohort and in a prospective setting based on a clinical outcome.

Conclusions: We conclude that the prototypical trend estimation has great potential for an integration in the telehealth system to be used in future work to provide tailored patient-specific advice based on these predictions. Continuous data from the telehealth system grant a deeper insight and a better understanding of the patients' status concerning well-being and level of pain as well as their current physical fitness level and the progress toward reaching set goals.

Trial Registration: ClinicalTrials.gov Identifier: NCT05619835; https://tinyurl.com/mrxt7y9u

(JMIR Form Res 2025;9:e65721) doi:10.2196/65721

XSL•FO
RenderX

KEYWORDS

mHealth; telehealth; peripheral arterial disease; home-based structured walking training; trend estimation; predictive modeling; continuous data; walking; walking training; prediction; prediction model; cardiovascular disease; stroke; heart failure; physical fitness; telehealth system

Introduction

Background

Peripheral arterial disease is a significant global health concern with an estimated 202 million people affected worldwide [1]. Prevalence in Europe is reported at 5.3%, increasing to 15% - 20% among individuals over the age of 70 years [2]. Peripheral arterial disease is characterized by insufficient oxygen supply due to arterial vessel alterations in the extremities. This leads to pain when walking, forcing patients to stop walking and thus affecting mobility and quality of life. Pain may occur even at rest. Furthermore, the healing of wounds is slowed down, and peripheral arterial disease may ultimately lead to necrosis. Peripheral arterial disease is classified into distinct grades according to the Rutherford system, which includes 6 categories, and the Fontaine classifications, which include 4 stages. Evidence-based guidelines for the management of peripheral arterial disease are provided by organizations such as the European Society of Cardiology, European Society of Vascular Medicine, and American Heart Association [3-5].

Treatment for Fontaine stage I is typically limited to risk factor management (eg, nicotine abstention and blood pressure management), which remains a relevant factor throughout all following stages. In Fontaine stage II, medication regimen and structured walking training (SWT) become additional treatment options. While medication is also recommended throughout stages III and IV, SWT is only prescribed in stages II and III. Stages III and IV have surgical interventions as an additional treatment option [3,4].

SWT can more than double the pain-free walking distance of patients with peripheral arterial disease within a 12-week program [6,7], and achieve long-term outcomes comparable to surgical intervention (revascularization) while being less expensive and non-invasive [4,8]. To further foster SWT in patients with peripheral arterial disease, the Trans-Atlantic Inter-Society Consensus-II group has defined a training protocol for patients. This recommendation suggests performing walking exercises at an intensity that induces pain for 3 - 5 minutes in total within a 30 - 60-minute session. Such a session should be conducted 3 times per week for 3 months [9]. To evaluate the progress of SWT and to assess the physical fitness of patients with peripheral arterial disease, standardized tests have been developed.

The state-of-the-art method for measuring the physical fitness of patients with peripheral arterial disease is the 6-minute walk test (6MWT), during which a patient walks as far as possible within 6 minutes on an incline-free surface [10]. Usually, the 6MWT is performed under the supervision of a health care professional (HCP) who measures the total walking distance after the completion of the test [11]. Another test to measure physical fitness is the timed-up-and-go (TUG) test. The TUG test starts with the participant sitting on a chair. The participant is then instructed to stand up, walk for 3 meters, turn around, walk back, and sit back down. The time the participant needs to complete this sequence of tasks is measured [12]. Even though TUG tests are not primarily used for patients with peripheral arterial disease, they provide information about the patients' general mobility and frailness. Additionally, a previous publication identified a significant correlation between the 6MWTs and TUG tests in patients with peripheral arterial disease [13].

Telehealth is defined as the delivery of health care or rehabilitation services to patients remotely using digital technologies [14], which, over the last years, has proven to be effective in the field of cardiovascular diseases, particularly heart failure and stroke [15,16]. Telehealth and telerehabilitation solutions have recently also shown promising results in the field of peripheral arterial disease [17,18]. Telehealth can support early detection of (post-operative) complications, with the potential to improve functional capacity, claudication onset time, and quality of life for patients with peripheral arterial disease [19].

Despite these studies, the benefits and the potential of telehealth in the setting of peripheral arterial disease are not yet fully explored. Aspects like the influence of continuously measured parameters and digital tools on the physical fitness of patients with peripheral arterial disease must be further investigated.

Keep Pace Study Protocol

During the Keep Pace study, 19 patients with peripheral arterial disease underwent a 12-week program with digital assistance to evaluate the impact of telehealth solutions on patients with peripheral arterial disease with a focus on the decision of undergoing surgery after the program [20]. The participants for the Keep Pace study were recruited at the outpatient clinics of the Department of Angiology of the University Hospital Graz, Austria. Patients who were shortlisted for getting an angioplasty were contacted and asked to participate in the Keep Pace study. Patients who gave informed consent and fit the inclusion criteria were included in this study. The inclusion criteria were as follows: age between 18 and 80 years, peripheral arterial disease in Fontaine stage IIa or IIb, ownership of an Android mobile device, and capability to use telehealth services. Patients lacking compliance to perform SWT were excluded. At the onset of the study, an initial in-clinic visit was performed, during which the patient's medical history and subjective limitations in daily activities, including pain-free and total walking distance, were assessed. All patients were diagnosed with peripheral arterial disease based on imaging diagnostics before study entrance. Because the location and grade of the occlusion were not part of the analysis, all diagnostic imaging options, such as duplex ultrasound, computer tomography, and magnetic resonance angiography, were accepted, which did not provide information about the location and grade of the occlusion for all patients. In addition to the imaging peripheral arterial disease criteria,

XSL•FO
RenderX

claudication had to be present for patients to be eligible for this study. In the following 8 weeks participants were telemedically supervised by a nurse. In the subsequent remaining 4 weeks, patients performed their SWT therapy independently without regular telehealth consultations. Thereafter, an in-clinic study end visit was performed [20] .

During the Keep Pace study, the collected telehealth data have been used to continuously monitor the patients' status. They might also be used for predictive modeling applications, eg, to detect already at early stages, which patients will benefit from the training and which patients might require additional attention. It is expected that such a prediction could help to achieve a more patient-specific therapeutic regime, tailored exercise support, and expectation management. However, currently, there is no evidence to support this hypothesis. This work takes an in-depth look at the telehealth data recorded during the Keep Pace study [20] by focusing on analyzing trends in patient-reported timeseries data.

Objective

The aim of this investigation was to test the feasibility of predicting the total 6-minute walk test distance at the study end in patients with peripheral arterial disease, which can be used to determine patients who will likely benefit more from the telehealth program than others and additionally provide patient-specific feedback and training regime adaptions throughout the patients' time in the telehealth system.

Methods

Dataset

For this work, data of 19 patients with peripheral arterial disease in the Fontaine stage II were analyzed, which were collected over a 12-week study period in the Keep Pace study [20]. While the study protocol only specified that participants had to be patients with peripheral arterial disease in Fontaine stage II, the actual patient cohort was unanimously in Fontaine stage IIb. The characteristics of the patients are shown in Table 1.

Data from 2 sources were collected during Keep Pace and further processed in this analysis: (1) data assessed in-clinic by HCPs at the beginning and at the end of the study (Textbox 1) and (2) patient-reported data collected with a telehealth system during the 12-week study period (Textbox 2).

Table . Patient characteristics by gender and in total (SD) [20].

Characteristics	Female (n=7)	Male (n=13)	Total (n=20)
Age, years, mean (SD)	64.14 (8.68)	62.46 (5.39)	63.05 (6.55)
Weight, kg, mMean (SD)	70.14 (9.5)	81.67 (11.95)	77.64 (12.27)
Height, m, mean (SD)	1.65 (0.03)	1.76 (0.05)	1.72 (0.07)
BMI, kg/m²	25.79 (3.44)	26.33 (4.51)	26.15 (4.08)
Subjective pain-free walking distance, m, mean (SD)	73.57 (50.22)	81.54 (37.16)	78.75 (41.04)
Fontaine stage IIa, n	0 (0%)	0 (0%)	0 (0%)
Fontaine stage IIb, n	7 (100%)	13 (100%)	20 (100%)
Diabetics, n (%)	1 (14%)	2 (15%)	3 (15%)
Hypercholesterolemia, n (%)	7 (100%)	13 (100%)	20 (100%)
Smokers, n (%)	2 (28%)	10 (77%)	12 (60%)
Coronary artery disease, n (%)	3 (43%)	4 (31%)	7 (35%)

Textbox 1. In-clinic data.

Patient characteristics: sex, height, age, weight, disease-related data like the date of diagnosis, cardiovascular risk factors, current medication, diabetes and smoking assessment and year of diagnosis

6MWTs: manual measurement by an HCP and automated measurement with a dedicated 6MWT app [21]

TUG test data: automated measurement with a TUG test device as described in [21,22], including the total TUG time and sub-task times (stand-up, walk-forward, turnaround, walk-back, sit-down, total walk time, and mobility time)

Textbox 2. Patient-reported telehealth data.

6MWTs: automated measurement with a 6MWT app [21] — recommended once per week

Steps per day: automatically measured by the mobile phone

Subjective well-being: recommended once per day

Pain: recommended once per day

XSL•FO
RenderX

The telehealth data described in (Textbox 2) were gathered by the patients using a dedicated mobile application. Patients did not receive an incentive for completing the 6MWTs and the SWT apart from the inherent incentive of improving their physical fitness and quality of life by adhering to the measurement and training regime. Subjective well-being was assessed on a 3-item scale (good, medium, and bad) and pain was reported on a scale (0=no pain to 10=maximum pain). The steps per day were recorded with the smartphones' native step counter, and patients could perform unsupervised 6MWTs using an integrated previously developed GPS-based 6MWT-app [21].

Predictive Modeling

We investigated different models for predicting the total walking distance at the study end right at the onset of the program (baseline) and at week *n* of the program, only based on data recorded until week *n*. A leave-one-out setting was applied, which excluded 1 patient from the training process and used this patient as a validator for the developed algorithm, resulting in 19 different data splits per tested model. This ensures that each patient is used as the validator once per model. The quality of the prediction as compared to the measured data at the study end was calculated using the root mean squared error (RMSE). The following predictive models were applied.

Distance at Baseline

This approach was mainly used as a reference for other models. The 6MWT distance at the study end was estimated as equal to the 6MWT distance at baseline (ie, this model assumed no improvement during the study).

Average Improvement Based on All Patients

This model assumed that all patients had similar improvements. Therefore, for estimating the 6MWT distance at the study end of patient *i* in week *n* ($d_{av,n,i}$), the mean change of all other patients from week *n* to the study end was calculated and added to the respective patient's 6MWT distance $d_{n,i}$ in week *n* (equation 1).

(1)dav,n,i= dn,i+ 118∑j=1j≠i19(dend, j−dn,j)

Logarithmic Interpolation Based on an Individual Patient

Logarithmic interpolation was used to fit a logarithmic function as defined by the parameters $a_{n,i}$ and $b_{n,i}$ into all measured 6MWT distances $d_{k,j}$ performed at week $x_{k,j}$ until week *n* by the respective patient *i*. The estimated distance at the study end as derived in week *n* for patient *i* ($d_{log\_ind,n,i}$) was calculated according to equation 2.

(2)dlog_ind,n,i=(an,i*loge x+bn,i)|x=12

with[an,i,bn,i]=logfit(xk,j,dk,j) |∀k≤n, j=i

Logarithmic Interpolation Based on All Patients

Assuming that all patients improved similarly, logarithmic interpolation was also applied to data from all patients when defining parameters a_n and b_n in week *n*. Therefore, all measurements $d_{k,j}$ were first corrected for the mean of each patient (dk,j−) (see equation 3).

(3)dlog_all,n,i=(an,i*loge x+bn,i)|x=12

with[an,bn]=logfit(xk,j,dk,j−) |∀k≤n,∀j

Linear Regression

All models described earlier are based on 6MWT distances only. With the linear regression approach, also other data assessed during the program were considered. Feature candidates applied for predictions in week *n* were derived from all data recorded during the in-clinic visit at baseline (see In-clinic data in the Methods section) and from telehealth data recorded until week *n*. Telehealth data contained the feature candidates, as shown in Textbox 3:

All possible combinations of up to 8 features were evaluated. Therefore, each feature was first tested on its own, and subsequently features were added iteratively to test all possible combinations. However, due to the small sample size, a maximum number of 4 features was the goal.

Textbox 3. Feature candidates from telehealth data.

Mean 6MWT distance up to week *n*: this parameter represents all total walking distance measurements from the 6MWT app from the baseline up to the week of the prediction.
Mean 6MWT distance in week *n*: this parameter considers only measurements of 6MWTs performed in the week of the prediction.
Logarithmic prediction in week *n*: this feature is derived from the trend-based analysis, where a predicted outcome is available for each week.

Statistical Analyses

All statistical analyses were implemented in MATLAB (The MathWorks) and all predictive models were either implemented in MATLAB or Python. Wherever reasonable, pre- and post-study comparisons were performed on data assessed at the in-clinic visits at baseline and the study end, using paired tests. Continuously measured parameters were tested for correlation with one another. Unless stated otherwise, correlations were tested using the Pearson's correlation coefficient with a previous Anderson-Darling test to test for a normal distribution of the data. Trend analyses were performed for all the continuous measurements mentioned earlier for which the data were aggregated on a weekly basis by calculating mean values per parameter. Predictive models were evaluated by calculating the root mean square error between the estimated 6MWT total walking distance at the study end and the actually measured distance. For this work, a *P* value <.05 was considered to be significant.

Ethical Considerations

Ethical approval for the study was obtained from the Ethics Committee of the Medical University of Graz (34 - 127 ex 21/22 1566 - 2021, ClinicalTrials.gov Identifier: NCT05619835). All analyses performed during this work are covered by this ethics approval. All participants received oral

XSL•FO
RenderX

and written information prior to the study entry and provided written informed consent to participate in this study. This informed consent also covers the analysis of secondary study outcomes performed in this paper. All data were recorded in a pseudonymized way and anonymity was ensured for all participants during the presentation of the findings. Patients did not receive any financial incentives for participating in the study.

Results

Data Characteristics and Trend Analysis

The 6-Minute Walk Test

The study protocol recommended 12 6MWTs. 57.89% (11 out of 19) patients performed 12 or more tests and were 100% adherent to the measurement protocol. The 8 remaining patients performed, on average, 7.88 (SD 2.36) 6MWTs and hence achieved a mean adherence of 65.62% (SD 19.64%) of the 12 recommended tests. On average, patients performed 13.37 (SD 8.90) 6MWTs and overall adherence to the measurement routine was 85.53% (SD 21.31%).

Figure 1 illustrates the 6MWT total walking distance aggregated on a weekly basis for all patients. The total walking distance increased over time, following a logarithmic trend.

As can be seen in Table 2, a significant correlation was found between the 6MWT total walking distance at baseline versus the distance at the study end. However, no significant correlation between either the number of tests and the distance at baseline, or the number of tests and the distance at the study end, or the number of tests and the change in walking distance during the program, or between the distance at baseline and the change in walking distance during the program has been identified.

Figure 1. Course over time of weekly aggregated 6-minute walk test (6MWT) total walking distances for all patients, illustrated as lines per patient and boxplots. Data are normalized for the patient-individual mean. Dashed lines illustrated interpolations over weeks with missing values. In-clinic measurements of the first and last visit are shown in dark blue; tests measured with the 6MWT app are shown in azure. The number n of patients who performed at least 1 6MWT in the respective week is shown on the bottom.

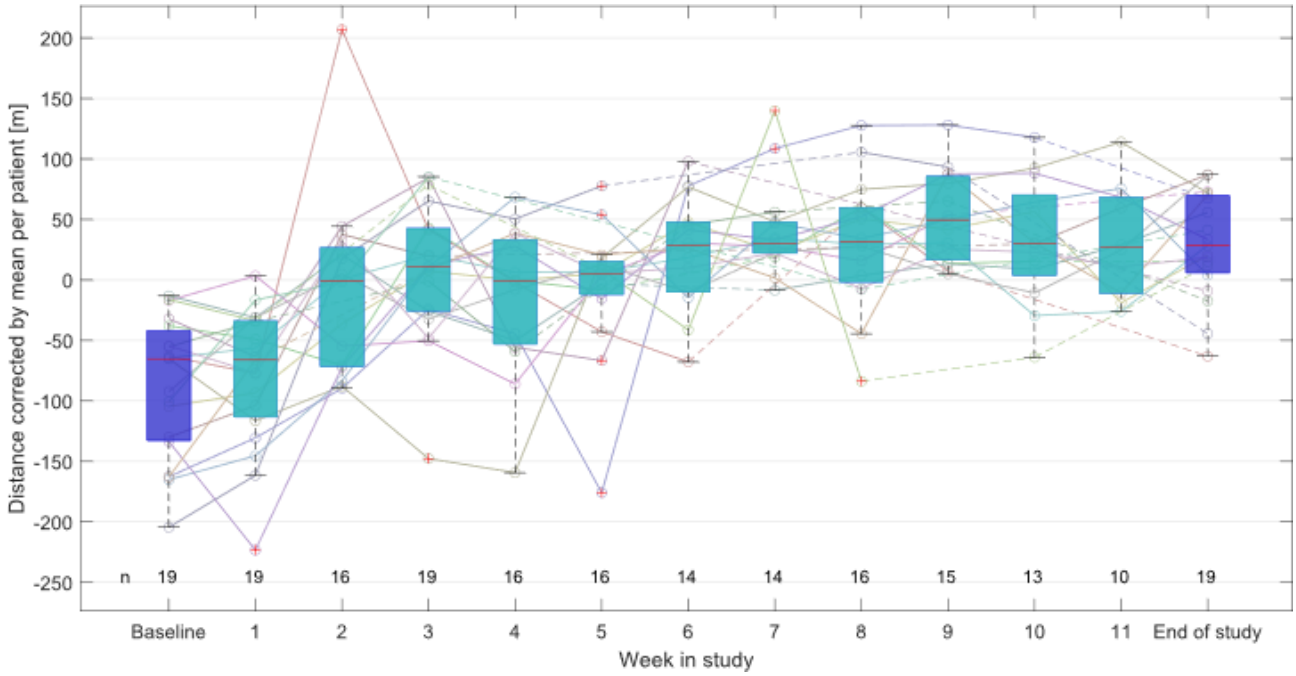


Table . Correlation between the final walking distance and the number of tests and the change in the walking distance between baseline and the study end (R: correlation coefficient; PC: Pearson correlation coefficient; S: Spearman correlation coefficient).

	Distance at the study end	Number of tests	Change in walking distance
Distance at baseline	*r*=0.78, *P*<.001, S	*R*=–0.14, *P*=.56, S	*r*=0.02, *P*=.93, S
Distance at the study end	N/A	*r*=−0.17, *P*=.48, S	*r*=0.71, *P*<.001, S
Number of tests	*r*=−0.13, *P*=.60, PC	N/A	*r*=0.02, *P*=.93, S

Timed-Up-and-Go Test

Table 3 summarizes all TUG subtask measures at baseline and the study end. Table 3 shows that the sit-down time significantly decreased from 2.21 to 1.89 seconds (*P*=.038), and the total walk time significantly decreased from 6.96 to 6.57 seconds (*P*=.003). Additionally, the overall TUG time significantly decreased from 10.6 seconds at baseline to 9.89 seconds at the study end (*P*=.002). Also, all other parameters decreased but did not reach statistical significance.

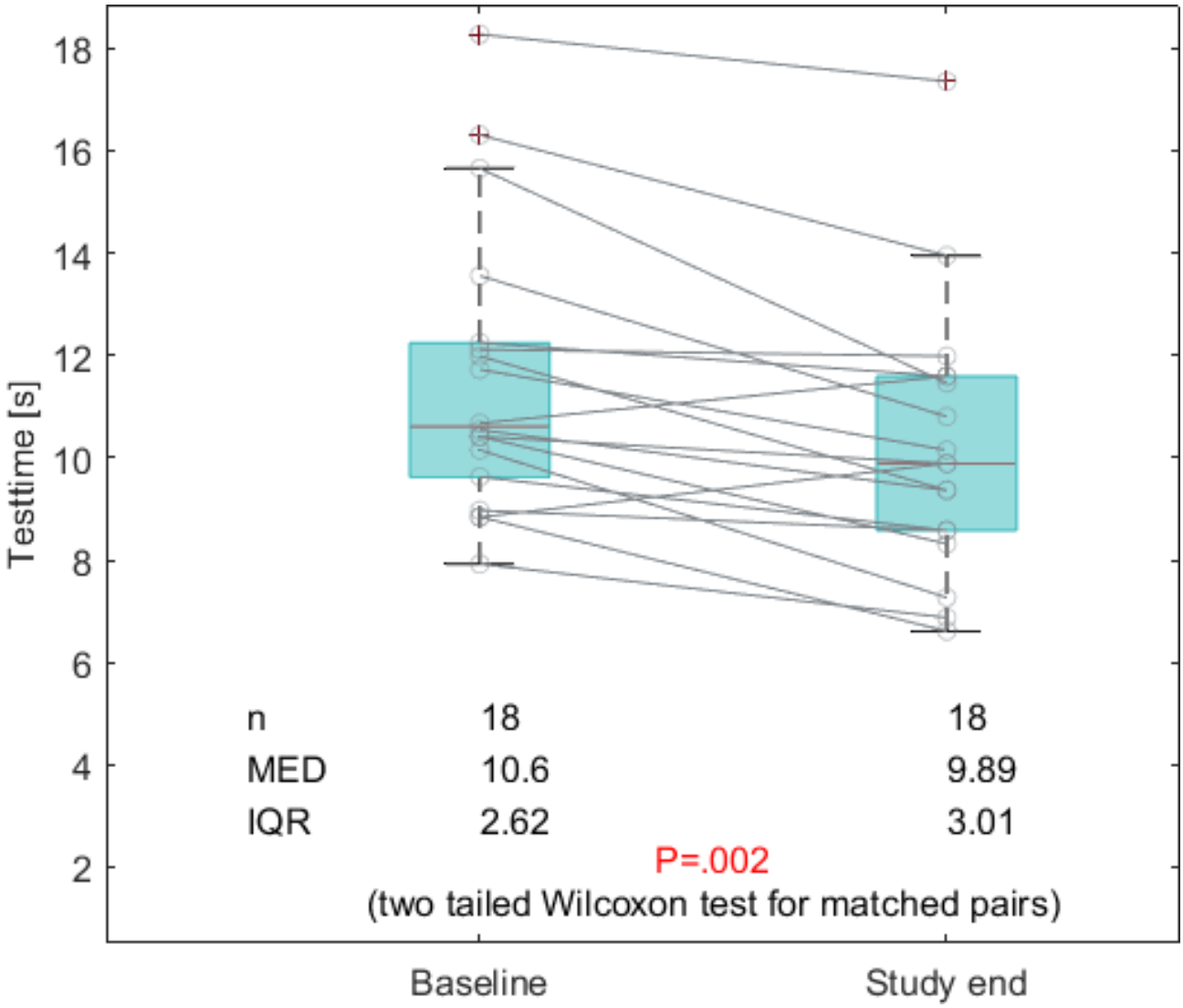
Figure 2 shows the decrease of the total TUG time.

XSL•FO
RenderX

Table . TUG subtask times (in seconds; median (IQR)) at baseline and the study end with the *P* values calculated with the 2-tailed Wilcoxon test for matched pairs.

	Baseline, median (IQR)	Study end, median (IQR)	*P* value
Stand-up	2.67 (2.34-4.44)	2.54 (2.10-3.00)	.09
Walk-forward	1.95 (1.70-2.21)	1.76 (1.69-2.10)	.06
Turnaround	1.82 (1.43-2.34)	1.56 (1.15-2.08)	.11
Walk-back	2.09 (1.95-2.35)	1.95 (1.82-2.21)	.06
Sit-down	2.21 (2.10-2.87)	1.89 (1.56-2.60)	.04
Total walk time	6.96 (5.98-9.78)	6.57 (5.33-7.42)	.003
Mobility time	4.10 (3.77-4.43)	3.64 (3.51-4.04)	.08
Total TUG time	10.6 (9.63-12.25)	9.89 (8.58-11.59)	.002

Figure 2. Improvement of the timed-up-and-go test time at baseline versus the study end (n: number of observations; MED: median; IQR: interquartile range).



Steps Per Day

The mean number of steps per day as measured by the smartphone's native app was 4176.58 (SD 2742.73). No significant change in steps per day was found when aggregating over all patients. However, 3 patients showed a significant reduction of steps, and 4 patients significantly improved. No relation between the change of recorded steps and the decision of undergoing surgery after the program was observed.

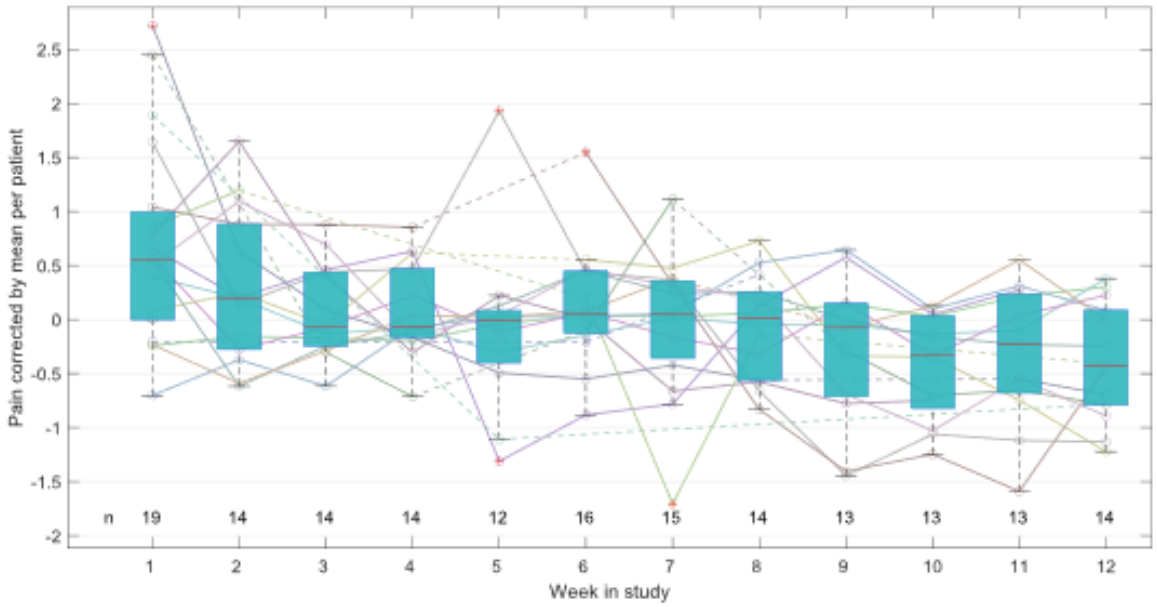
Pain

Overall, patient-reported pain was significantly reduced from a mean of 4.67 (SD 1.90) at baseline to 3.60 (SD 2.39) at the study end ($P<.05$). Additionally, pain showed a very low but

XSL•FO
RenderX

highly significant correlation with steps per day (r=−0.11; P<.001) and the distance of the 6MWTs (r=−0.39; P<.001) on days when patients submitted both measurements. Figure 3 illustrates the course over time of weekly pain values corrected for the mean per patient.

Figure 3. Course over time of weekly aggregated pain measurements for all patients, illustrated as lines per patient and boxplots. The level of pain is normalized for the patient-individual mean. Dashed lines illustrated interpolations over weeks with missing values. The number of patients who reported at least 1 pain level in the respective week is shown on the bottom.



Subjective Well-Being

Out of 1597 subjective well-being data points, 41 (2.63%) were reported as bad, 686 (43.0%) as medium, and 870 (54.5%) as good. No trends or patterns in the development of subjective well-being and no significant change from baseline to the study end were identified. On 5 days with bad, 160 with medium, and 106 with good well-being, 6MWTs were performed by the respective patient. On days with a submitted subjective well-being and a 6MWT, the number of bad well-being was 1.88%, which is in the same magnitude as the overall bad well-being submissions. Subjective well-being showed a significant correlation with the 6MWT total walking distance (r=0.26, P<.001, Spearman). On days with subjective well-being and step count data available (40 bad, 657 medium, and 835 good), a weak but highly significant correlation between subjective well-being and steps per day was found (r=0.19; P<.001).

Trend Prediction and Predictive Modeling

Table 4 shows the RMSEs between estimated and measured 6MWT total walking distance at the study end (week 12), depending on the week of estimation and the respective model. The distance at the baseline model, which estimated the distance at the study end equal to the distance at baseline, revealed an error of 136 meters. The average improvement in the week n method achieved an error of 109 meters at baseline and a continuous decrease of the error up to week 7 with an error of 43 meters. The logarithmic interpolation with a patient individual fit was not able to outperform the improvement in the week n method. However, the logarithmic fit on all patients was able to outperform the improvement in the week n method from week 1 to week 6. The linear regression model used data from the baseline of the study in combination with the time-dependent features, described in the Methods ection under the subsection Predictive modeling. The linear regression model continuously decreased the prediction error from week to week since the time-dependent features became more proximate to the outcome. For this analysis, the number of features was set to be between 1 and 4, whichever combination performed the best according to Table 4. The linear regression was able to outperform the logarithmic interpolation in every week. The error of 30 meters achieved by the linear regression model at baseline was 7.04% the mean total walking distance at the study end of 425 meters. The progress of the estimation errors over time is shown in Figure 4.

XSL•FO
RenderX

Table . Root mean square error (m) between the estimated 6-minute walk test (6MWT) total walking distance at the study end and measured distance, as achieved with different models based on the available data in the corresponding week.

Model	Baseline	Week										
		1	2	3	4	5	6	7	8	9	10	11
Distance at baseline	136	—[a]	—	—	—	—	—	—	—	—	—	—
6MWT data only												
Average improvement in week n	109	112	87	69	59	73	52	43	54	63	51	9
Logarithmic interpolation for individual fit	—	233	141	91	87	82	67	62	61	65	58	60
Logarithmic interpolation for all patient fit	—	50	59	50	50	48	48	48	49	50	50	49
6MWT + other data												
Linear regression	30	30	29	29	29	29	25	25	22	22	14	14

[a]Not applicable.

XSL•FO
RenderX

Figure 4. Course over time of the root mean square error (RMSE) between estimated and measured 6-minute walk test total walking distance, depending on the week of estimation, per model.

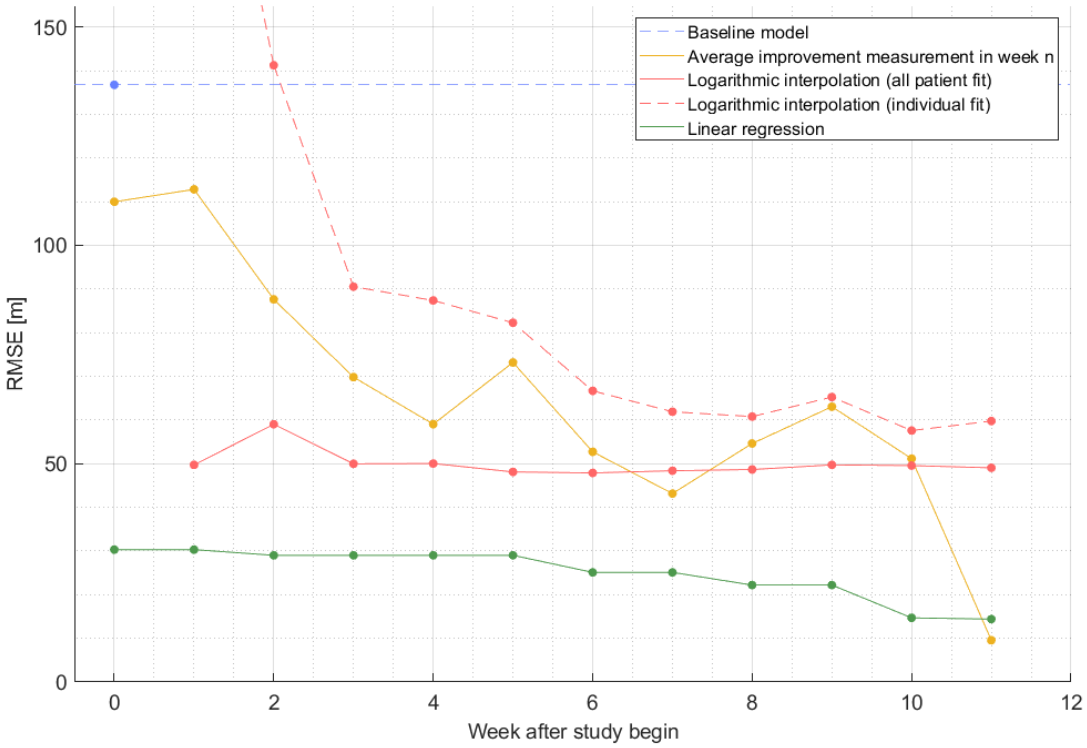


Table 5 shows the RMSE of models using data assessed in-clinic at the baseline of the study. Only the RMSE of the best-performing feature or feature combination is shown in the table. The ideal number of features for the linear regression was 4 (distance at the onset of the study, the patients' sex, the difference between subjectively estimated distance until the onset of pain and the actual distance, and the ratio between the 2 TUG tests, both performed at the onset).

Table . Root mean square errors (m) of the estimated 6-minute walk test total distance as compared to the actually measured distance in week 12, based on the number of features at the baseline.

Model	Number of features							
	1	2	3	4	5	6	7	8
Linear regression	52.08	38.89	31.10	30.31	30.63	30.59	31.90	33.23

Discussion

Principal Results

In the current work, an in-depth analysis of previously not analyzed data recorded during the Keep Pace Study was performed, in which 19 patients with peripheral arterial disease performed a 12-week walking program with telemonitoring assistance. In addition to the improved pain-free walking distance as previously published in [20], we identified significant improvements in physical activity scores during the study, based on both the TUG test and the 6MWT. The regular recording of 6MWTs by the patients at home in combination with the telehealth service allowed HCPs to monitor the individual success of the SWT in real time. Based on these data (Figure 1), we identified that the majority of the 6MWT improvement happened in the first 3 weeks of the program, while saturating thereafter, which is in good accordance with the training effect in other studies, even though this stagnation of improvement happens later in healthy individuals [23]. We also identified a significant reduction of pain during the program. In contrast to the training effect, this improvement did not show any saturation but spread until the end of the program.

Despite the pain that was reported by the patients, the overall adherence to the 6MWT measurement protocol was high (85.53%). Unfortunately, we were only able to measure the adherence to the test protocol and not the adherence to the actual SWT, which was not recorded during the Keep Pace study. Since we had no insight into whether the patients continuously wore their smartphones with the step count, we could not use the daily steps to draw conclusions about the adherence to the WT. However, based on the high adherence to the measurement protocol and the overall improvement in walking distance, we assume that most patients were adherent to the training program, too. Neither adherence nor improvement during the program

XSL•FO
RenderX

was dependent on the initial 6MWT distance, which indicates the program's potential across all baseline fitness levels.

While both the 6MWT distance and the TUG time improved during the program, no significant increase in the number of steps per day, as another measurement of the daily physical activity, was observed. However, since steps were tracked by the smartphones' native apps only, and since smartphones are not expected to be carried continuously, the number of steps might not be representative.

The TUG test is currently not a primary assessment tool for physical fitness in patients with peripheral arterial disease. However, TUG times and 6MWT distances are highly correlated in patients with peripheral arterial disease. Since the TUG test takes only approximately 20 - 30 seconds (as compared to the 6 minutes of the 6MWT), it is much less burdensome and time-consuming. Further studies are required to analyze whether parts of the 6MWT can be replaced by TUG tests or if TUG data can provide additional insights in combination with 6MWTs to optimize peripheral arterial disease management.

The data showed that the physical activity and fitness levels decreased with an increase in the pain sensation, as to be expected since pain is a limiting factor for physical activity. Similarly, the general well-being correlated with the outcome of the 6MWTs. However, due to the highly unequal distribution of subjective well-being on days on which a 6MWT was performed, the correlation between well-being and 6MWT total distance must be interpreted with care. Additionally, using a scale from 1 to 3 for measuring the well-being might result in inaccurate depictions due to a lack of granularity in the reporting compared to a continuous scale from 0 to 10 which was used for the measurement of pain. This lack of granularity might explain why no significant change in the well-being during the study has been identified.

The developed predictive models highlighted that—based on telehealth data—the outcome can be predicted with an error of 30 meters at baseline, which corresponds to 7.04% of the mean total walking distance at the study end. Whether this prediction is accurate enough to help adjust the training to the individual patient or to apply patient-specific measures, such as additional phone calls, must be evaluated separately in a prospective setting. With the modeling approaches, we have shown 2 potential use cases, where (1) a prediction at study baseline can be done accurately to identify patients with great potential for the system and (2) continuous, weekly predictions to give patient-specific feedback. A prediction of the final walking distance as in case (1) could be used to not only identify patients with high potential to benefit from the system but also to motivate them to join the system by providing them with a probable result if they were adherent to the program. The predictions of (2) could be used to provide patient-specific feedback and training regime adaptions.

Limitations

The results of this study must be interpreted with care since our dataset contained only 19 patients. Additionally, the study population was very homogeneous with only patients with peripheral arterial disease of Fontaine stage IIb since this stage benefits the most from SWT. While this work showed that predictive modeling and trend estimation work well for patients with Fontaine stage IIb, future work must validate these findings for other stages of peripheral arterial disease. Additionally, with a rather short follow-up time of 90 days, the effects of the telehealth solution on the long-term adherence to the walking program could not be evaluated.

These limitations apply especially for the predictive models. Therefore, we applied a leave-one-out train-test split to maximize the available training data and limited the number of features for the models. In general, the trend estimation worked well compared to the distance-at-baseline method but showed great differences in the RMSE between the methods.

As expected, telehealth data can improve the accuracy of models that predict 6MWT distances at the study end. Simple models showed high instability when applied on data of the first weeks only, especially when based on individual patients' data only. However, even these models improved when telehealth data from further weeks were included. The lowest error of all models was achieved with the average improvement model applied in week 11. However, this approach performed worse in previous weeks, and predictions in week 11 might be of limited interest in a real-world application. Additionally, we assume that this extraordinarily good result might have been achieved by chance, due to the small sample size. The algorithms used for predicting the total walking distance at the study end were low in complexity due to the small study size of only 19 patients. A random forest was also tested as a possible approach but was outperformed by linear regression in every instance. More complex models like the random forest or support vector machines, or even (deep) neuronal networks, might perform even better. However, much larger datasets will be required to train these models, even if transfer learning is applied.

Future Work

Future work should validate the effect of predictive modeling in supporting home-based structured walking training in patients with peripheral arterial disease with (a) a larger sample size, (b) prospectively, and (c) based on clinical outcomes compared to a control group only using the telehealth system without continuous predictive modeling support. Finally, our results need to be integrated into real-world telehealth settings, ideally in a prospective study to evaluate the impact of informed decision-making by the HCP based on recorded telehealth data and on the results of the predicted outcomes. This approach might not only help HCPs provide patient-specific care but could also help the patients with their motivation by seeing the progress and the prediction of what they could achieve if they stayed adherent to the walking program.

Conclusions

This paper shows that the prototypical trend estimation has great potential for an integration in the telehealth system to be used in future work to provide tailored patient-specific advice based on these predictions. Additionally, the continuous data from the telehealth system grant a deeper insight and a better understanding of the patients' status concerning well-being and

XSL•FO
RenderX

level of pain as well as their current physical fitness level and the progress towards reaching set goals.

Acknowledgments

Supported by TU Graz Open Access Publishing Fund.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to privacy concerns and based our ethics approval but are available from the corresponding author on reasonable request.

Authors' Contributions

FW had the main responsibility for preparing the final manuscript and for the statistical analyses. AP was responsible for preparing and conducting the KEEP PACE study and interpretation of the results. AZ contributed to the study design, study preparation, statistical analyses, and interpreting the results. GEM contributed to conducting the study preparation and interpretation of the results. RMO contributed to the study preparations and interpretation of the results. MB contributed to the interpretation of the results. MB contributed to the study design. GS contributed to the study preparation and helped conducting the study. GS contributed to the study preparation and helped conducting the study. GS contributed to preparing the final manuscript, interpreting the results, and the statistical analyses. DH contributed to preparing the final manuscript and the statistical analyses. All authors discussed the results, supervised the findings, and contributed to the article and approved the submitted version. No funding was requested for this project. It was a self-financed project of Meduni Graz and AIT within the framework of a dissertation. Generative artificial intelligence has been used to improve grammar and language during the writing of this paper.

Conflicts of Interest

None declared.

References

1. Song P, Rudan D, Wang M, Chang X, Rudan I. National and subnational estimation of the prevalence of peripheral artery disease (PAD) in China: a systematic review and meta-analysis. J Glob Health 2019 Jun;9(1):010601. [doi: 10.7189/jogh.09.010601] [Medline: 30873278]
2. Olinic DM, Spinu M, Olinic M, et al. Epidemiology of peripheral artery disease in Europe: VAS Educational Paper. Int Angiol 2018 Aug;37(4):327-334. [doi: 10.23736/S0392-9590.18.03996-2] [Medline: 29936722]
3. Aboyans V, Ricco JB, Bartelink M, et al. 2017 ESC Guidelines on the Diagnosis and Treatment of Peripheral Arterial Diseases, in collaboration with the European Society for Vascular Surgery (ESVS): document covering atherosclerotic disease of extracranial carotid and vertebral, mesenteric, renal, upper and lower extremity arteries. Endorsed by: the European Stroke Organization (ESO)The Task Force for the Diagnosis and Treatment of Peripheral Arterial Diseases of the European Society of Cardiology (ESC) and of the European Society for Vascular Surgery (ESVS). Eur Heart J 2018 Mar 1;39(9):763-816. [doi: 10.1093/eurheartj/ehx095] [Medline: 28886620]
4. Frank U, Nikol S, Belch J, et al. ESVM Guideline on peripheral arterial disease. Vasa 2019 Sep;48(Suppl 102):1-79. [doi: 10.1024/0301-1526/a000834] [Medline: 31789115]
5. Gerhard-Herman MD, Gornik HL, Barrett C, et al. 2016 AHA/ACC guideline on the management of patients with lower extremity peripheral artery disease: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol 2017 Mar 21;69(11):e71-e126. [doi: 10.1016/j.jacc.2016.11.007] [Medline: 27851992]
6. Murphy TP, Cutlip DE, Regensteiner JG, et al. Supervised exercise versus primary stenting for claudication resulting from aortoiliac peripheral artery disease: six-month outcomes from the claudication: exercise versus endoluminal revascularization (CLEVER) study. Circulation 2012 Jan 3;125(1):130-139. [doi: 10.1161/CIRCULATIONAHA.111.075770] [Medline: 22090168]
7. Murphy TP, Cutlip DE, Regensteiner JG, et al. Supervised exercise, stent revascularization, or medical therapy for claudication due to aortoiliac peripheral artery disease: the CLEVER study. J Am Coll Cardiol 2015 Mar 17;65(10):999-1009. [doi: 10.1016/j.jacc.2014.12.043] [Medline: 25766947]
8. Nicolaï SPA, Teijink JAW, Prins MH, Exercise Therapy in Peripheral Arterial Disease Study Group. Multicenter randomized clinical trial of supervised exercise therapy with or without feedback versus walking advice for intermittent claudication. J Vasc Surg 2010 Aug;52(2):348-355. [doi: 10.1016/j.jvs.2010.02.022] [Medline: 20478681]
9. Norgren L, Hiatt WR, Dormandy JA, et al. Inter-Society Consensus for the Management of Peripheral Arterial Disease (TASC II). Eur J Vasc Endovasc Surg 2007;33 Suppl 1:S1-75. [doi: 10.1016/j.ejvs.2006.09.024] [Medline: 17140820]

XSL•FO
RenderX

10. Nordanstig J, Behrendt CA, Baumgartner I, et al. Editor's Choice -- European Society for Vascular Surgery (ESVS) 2024 Clinical Practice Guidelines on the Management of Asymptomatic Lower Limb Peripheral Arterial Disease and Intermittent Claudication. Eur J Vasc Endovasc Surg 2024 Jan;67(1):9-96. [doi: 10.1016/j.ejvs.2023.08.067] [Medline: 37949800]
11. Shu J, Santulli G. Update on peripheral artery disease: epidemiology and evidence-based facts. Atherosclerosis 2018 Aug;275(379-81):379-381. [doi: 10.1016/j.atherosclerosis.2018.05.033] [Medline: 29843915]
12. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991 Feb;39(2):142-148. [doi: 10.1111/j.1532-5415.1991.tb01616.x] [Medline: 1991946]
13. Wiesmüller F, Prenner A, Ziegl A, El-Moazen G, Modre-Osprian R, Seinost G, et al. Correlation between timed-up-and-go test and 6-minute walk test in patients with peripheral artery disease. : IEEE; 2024 Presented at: 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).
14. Brennan DM, Mawson S, Brownsell S. Telerehabilitation: enabling the remote delivery of healthcare, rehabilitation, and self management. Stud Health Technol Inform 2009;145(231-48):231-248. [Medline: 19592797]
15. Poelzl G, Egelseer-Bruendl T, Pfeifer B, et al. Feasibility and effectiveness of a multidimensional post-discharge disease management programme for heart failure patients in clinical practice: the HerzMobil Tirol programme. Clin Res Cardiol 2022 Mar;111(3):294-307. [doi: 10.1007/s00392-021-01912-0] [Medline: 34269863]
16. Zerna C, Jeerakathil T, Hill MD. Telehealth for remote stroke management. Can J Cardiol 2018 Jul;34(7):889-896. [doi: 10.1016/j.cjca.2017.12.025] [Medline: 29459240]
17. Cucato G, Snowden C, McCone E, et al. Evaluating the feasibility and acceptability of an exercise and behaviour change intervention in socioeconomically deprived patients with peripheral arterial disease: the textpad study protocol. PLoS One 2022;17(6):e0269999. [doi: 10.1371/journal.pone.0269999] [Medline: 35749440]
18. Schwamm LH, Chumbler N, Brown E, et al. Recommendations for the implementation of telehealth in cardiovascular and stroke care: a policy statement from the American Heart Association. Circulation 2017 Feb 14;135(7):e24-e44. [doi: 10.1161/CIR.0000000000000475] [Medline: 27998940]
19. Haveman ME, Kleiss SF, Ma KF, et al. Telemedicine in patients with peripheral arterial disease: is it worth the effort? Expert Rev Med Devices 2019 Sep;16(9):777-786. [doi: 10.1080/17434440.2019.1649595] [Medline: 31379218]
20. Prenner A, Ziegl A, Wiesmüller F, et al. Usability of a telehealth-nurse supported home-based walking training for peripheral arterial disease - The Keep Pace! pilot study. Vasa 2024 Jul;53(4):246-254. [doi: 10.1024/0301-1526/a001127] [Medline: 38808475]
21. Ziegl A, Rzepka A, Kastner P, et al. mHealth 6-minute walk test - accuracy for detecting clinically relevant differences in heart failure patients. Annu Int Conf IEEE Eng Med Biol Soc 2021 Nov;2021:7095-7098. [doi: 10.1109/EMBC46164.2021.9630118] [Medline: 34892736]
22. Ziegl A, Kastner P, Modre-Osprian R, Schreier G. Automated measurement and subtask analysis of the timed up-and-go test in the field of geriatrics. Annu Int Conf IEEE Eng Med Biol Soc 2018 Jul;2018(1526-9):1526-1529. [doi: 10.1109/EMBC.2018.8512464] [Medline: 30440683]
23. Scharhag-Rosenberger F, Meyer T, Walitzek S, Kindermann W. Time course of changes in endurance capacity: a 1-yr training study. Med Sci Sports Exerc 2009 May;41(5):1130-1137. [doi: 10.1249/MSS.0b013e3181935a11] [Medline: 19346973]

Abbreviations

6MWT: 6-minute walk test
HCP: health care professional
RMSE: root mean squared error
SWT: structured walking training
TUG: timed-up-and-go

Edited by A Mavragani; submitted 23.08.24; peer-reviewed by EL Garcia; revised version received 04.12.24; accepted 05.12.24; published 10.04.25.

Please cite as:
Wiesmüller F, Prenner A, Ziegl A, El-Moazen G, Modre-Osprian R, Baumgartner M, Brodmann M, Seinost G, Silbernagel G, Schreier G, Hayn D
Support of Home-Based Structured Walking Training and Prediction of the 6-Minute Walk Test Distance in Patients With Peripheral Arterial Disease Based on Telehealth Data: Prospective Cohort Study
JMIR Form Res 2025;9:e65721
URL: https://formative.jmir.org/2025/1/e65721
doi:10.2196/65721

XSL•FO
RenderX

© Fabian Wiesmüller, Andreas Prenner, Andreas Ziegl, Gihan El-Moazen, Robert Modre-Osprian, Martin Baumgartner, Marianne Brodmann, Gerald Seinost, Günther Silbernagel, Günter Schreier, Dieter Hayn. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Web-Based Explainable Machine Learning-Based Drug Surveillance for Predicting Sunitinib- and Sorafenib-Associated Thyroid Dysfunction: Model Development and Validation Study

Fan-Ying Chan[1], MSc; Yi-En Ku[1], BSc; Wen-Nung Lie[2], PhD; Hsiang-Yin Chen[1,3], PharmD

[1]Department of Clinical Pharmacy, College of Pharmacy, Taipei Medical University, 250 Wuxing St, Xinyi Dist, Taipei, Taiwan

[2]Department of Electrical Engineering, National Chung Cheng University, Chiayi, Taiwan

[3]Department of Pharmacy, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

Corresponding Author:
Hsiang-Yin Chen, PharmD
Department of Clinical Pharmacy, College of Pharmacy, Taipei Medical University, 250 Wuxing St, Xinyi Dist, Taipei, Taiwan

Abstract

Background: Unlike one-snap data collection methods that only identify high-risk patients, machine learning models using time-series data can predict adverse events and aid in the timely management of cancer.

Objective: This study aimed to develop and validate machine learning models for sunitinib- and sorafenib-associated thyroid dysfunction using a time-series data collection approach.

Methods: Time series data of patients first prescribed sunitinib or sorafenib were collected from a deidentified clinical research database. Logistic regression, random forest, adaptive Boosting, Light Gradient-Boosting Machine, and Gradient Boosting Decision Tree were used to develop the models. Prediction performances were compared using the accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve, and area under the precision-recall curve. The optimal threshold for the best-performing model was selected based on the maximum F1-score. SHapley Additive exPlanations analysis was conducted to assess feature importance and contributions at both the cohort and patient levels.

Results: The training cohort included 609 patients, while the temporal validation cohort had 198 patients. The Gradient Boosting Decision Tree model without resampling outperformed other models, with area under the precision-recall curve of 0.600, area under the receiver operating characteristic curve of 0.876, and F1-score of 0.583 after adjusting the threshold. The SHapley Additive exPlanations analysis identified higher cholesterol levels, longer summed days of medication use, and clear cell adenocarcinoma histology as the most important features. The final model was further integrated into a web-based application.

Conclusions: This model can serve as an explainable adverse drug reaction surveillance system for predicting sunitinib- and sorafenib-associated thyroid dysfunction.

(JMIR Form Res 2025;9:e67767) doi:10.2196/67767

KEYWORDS

thyroid dysfunction; machine learning; cancer; sunitinib; sorafenib; TKI; tyrosine kinase inhibitor

Introduction

Background

Sunitinib- and sorafenib-associated thyroid dysfunction are time-varying and underreported adverse drug reactions (ADR). Despite the efficacy of multitargeted tyrosine kinase inhibitors TKI as first- or second-line therapies for solid and hematologic cancers, thyroid dysfunction—especially hypothyroidism—may complicate treatment regimens using sunitinib and sorafenib. Although not life-threatening, such adverse events can lead to a suboptimal quality of life, physical discomfort, or the need for thyroid dysfunction treatment [1,2]. Incidences of thyroid dysfunction vary from 10%- to 85% for sunitinib and from 6.3% to 27% for sorafenib [3-6]. Hypothyroidism develops late and is prolonged, with onset ranging from 5 to 20 months [1,6-8]. Due to its high variability and difficulty in predicting, close monitoring and proactive ADR surveillance with machine learning (ML) models may be warranted to manage sunitinib- and sorafenib-associated thyroid dysfunction.

Incorporating real-time laboratory data can enhance the performance of ML models in predicting thyroid adverse events. Previous studies have demonstrated the effectiveness of ML in predicting thyroid dysfunction, such as amiodarone-induced dysfunction using time-series data collection methods with a robust performance [9]. In contrast, predicting immune checkpoint inhibitor-induced adverse thyroid events without continuous laboratory testing resulted in an area under the precision-recall curve (AUPRC) of 0.510 [10]. Another study

XSL•FO
RenderX

developed a predictive model for immune checkpoint inhibitor-induced adverse thyroid events without detailed time-point data collection, achieving an area under the receiver operating characteristic curve (AUROC) of 0.885 with thyroid-related features [11]. However, no studies have focused on predicting thyroid adverse events caused by TKIs such as sunitinib and sorafenib. Adopting timely data collection with detailed clinical biochemical tests to develop thyroid dysfunction surveillance systems can help clinicians timely adjust TKI treatment by balancing risks and benefits.

Applying model interpretation and web-based applications can alleviate the black-box drawback of ML models and enhance understanding of ADRs. The SHapley Additive exPlanations (SHAP) analysis is a mathematical method that has become increasingly popular for explaining ML models [12-14]. It is based on game theory concepts that calculate feature importance and contributions to predicted outcomes at both population and individual levels [15]. Model deployment in web-based or smartphone applications were constructed in recent years to strengthen the scalability of predicting specific diseases or ADRs [13,16,17]. Through a user-friendly interface and explainable model, applications can provide clinicians with a greater understanding and actionable preventive remedies for their patients. These techniques successfully have increased the interpretability and accessibility of ML models to support clinical decision-making.

The landscape of drug surveillance has experienced notable changes, shifting from passive reporting systems or rules-based alerting systems to active identification of adverse drug events, driven by advancements in electronic health records and artificial intelligence. Traditional methods often rely on spontaneous reporting or rules-based alerting, which may lead to underreporting and delays in identifying ADRs. The ML models, particularly tree-based algorithms, can analyze large-scale clinical data in real time, proactively detecting ADRs and enhancing medication safety at the point of care [18]. By continuously collecting and analyzing patient data, ML-driven drug surveillance based on time-series data extraction enables dynamic risk assessment, facilitating early intervention and personalized treatment adjustments. Given the underreported and time-varying nature of thyroid dysfunction, this progressive approach can serve as a valuable tool for predicting ADRs in patients undergoing TKI therapy.

Objective

The objective of this study was to develop and validate progressive ML predictive models for sunitinib- and sorafenib-induced thyroid dysfunction. Five algorithms—logistic regression (LR), random forest (RF), Gradient Boosting Decision Tree (GBDT), Light Gradient Boosting Machine (LGBM), and Adaptive Boosting (AdaBoost)—were used to construct the models. The specific aims of this study were (1) to use time-series data collected at the baseline, and at 1, 2, 3, 4, 5, 6, 9, 12, 18, 24, 30, and 36 months after the index date to generate predictive models using the five algorithms; (2) to select the best model by comparing the accuracy, precision, recall, F1-score, AUROC, and AUPRC, after adjusting for the optimal threshold; (3) to interpret the best-performing model with a SHAP analysis to analyze feature importance levels and contributions, comparing findings with a previous amiodarone-induced thyroid dysfunction predictive model; and (4) to deploy the best-performing model by constructing a web-based application.

Methods

Ethical Considerations

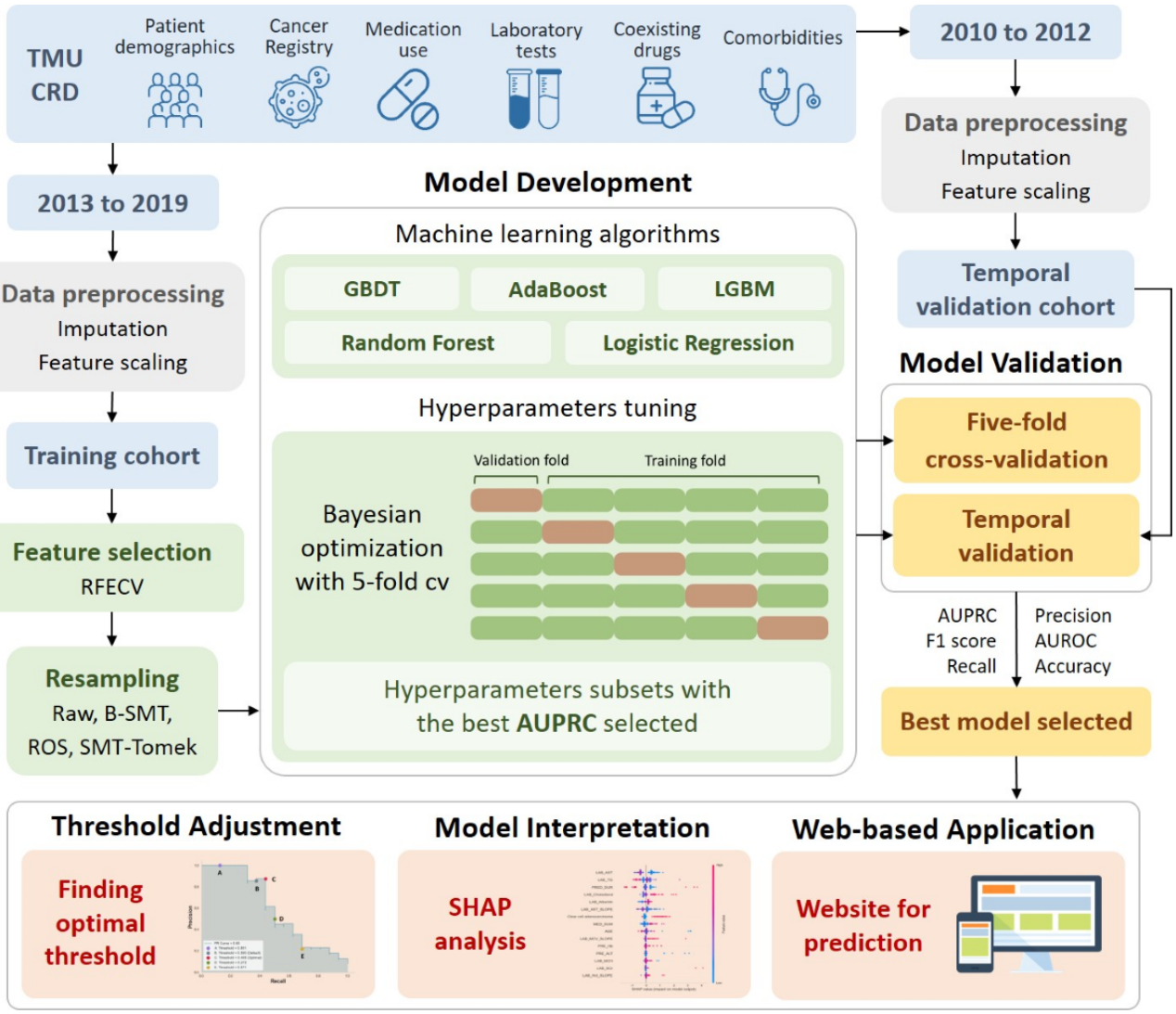
This retrospective study used the deidentified Clinical Research Database (CRD) which includes data from Taipei Medical University Hospital, Wan Fang Hospital, and Shuang Ho Hospital of the TMU health care system. Ethical review for this study was waived by the TMU-Joint Institutional Review Board (approval no.: N202202053). As the data had been deidentified, the requirement for informed consent was waived. No compensation was provided to the participants.

Study Design and Patient Cohort

Data from patients who were first prescribed sunitinib or sorafenib from 2013 to 2019 were collected as the derivation (training) cohort, whereas data from patients treated between 2010 and 2012 were collected as a temporal validation (testing) cohort. The data-splitting time point was chosen considering a common proportion of 7:3 for the training and testing cohorts and similar incidences of thyroid dysfunction in both cohorts for better model development and validation. More recent data were used for model training to better reflect current clinical practice, allowing models to be developed with up-to-date treatment patterns and patient characteristics. Patients were excluded if they were younger than 18 years, pregnant, with a history of hypothyroidism, had a diagnosis of thyroid carcinoma, had undergone thyroidectomy, used levothyroxine, or had abnormal thyroid-stimulating hormone (TSH) levels within 1 year before the index date, which was the first day of sunitinib or sorafenib use. Patients were also excluded if they were lost to follow up. Each individual was followed up for 3 years, and data were collected until the end of the study period, loss to follow-up, death, or occurrence of thyroid dysfunction. Figure 1 demonstrates the study design of the proposed models. This study followed the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis + Artificial Intelligence (TRIPOD + AI) reporting guideline [19]. Registration was not required, and no separate study protocol was prepared.

XSL•FO
RenderX

Figure 1. Study design of the proposed models. The study process included data collection, data splitting, data preprocessing, feature selection, resampling, model development, model validation, threshold adjustment, model interpretation, and development of a web-based application. AdaBoost: Adaptive Boosting; AUPRC: area under the precision recall curve; AUROC: area under the receiver operating characteristic curve; B-SMT: Oversampling with the Borderline Synthetic Minority Oversampling Technique; CRD: clinical research database; GBDT: Gradient Boosting Decision Tree; LGBM: Light Gradient Boosting Machine; RFECV: Recursive Feature Elimination with Cross-Validation; SMT: Synthetic Minority Oversampling Technique; TMU: Taipei Medical University.



Time-Series Data Collection and Preprocessing

Variables including patient demographics, cancer-related information, medication use, laboratory tests, coexisting drugs, and comorbidities were collected from the CRD. Patient demographics including age, gender, and cancer-related information were collected once at the baseline. Laboratory tests, medication use, comorbidities, coexisting drugs, weight, and body mass index (BMI) were collected at the baseline, and at 1, 2, 3, 4, 5, 6, 9, 12, 18, 24, 30, and 36 months after the index date (Multimedia Appendix 1). Reengineering features included calculating the cumulative dose, summed days of medication, duration of sunitinib or sorafenib, and the slope of laboratory tests at recent and previous data collection time points. The number of coexisting thyroid-related drugs was calculated to determine whether the aggregated use of these medications increased the risk of thyroid dysfunction.

Data preprocessing including missing value imputation and feature scaling was performed to ensure data quality. Features with over 80% missing values, except for the TSH level, were first excluded from the model. Missing values were imputed using multivariate imputation by chained equations [20]. If previous data were traceable, missing values, using time-series data collection, were precisely imputed using the last observation carried forward. Missing values in laboratory test slopes were imputed with zero. A feature-scaling method—the robust scaler—was used to reduce the effect of extreme values of continuous variables. Features were first selected through a literature review and clinical domain knowledge to identify thyroid-related variables. Recursive feature elimination, a feature selection technique that recursively removes the least important features with 5-fold cross-validation was then used, and the AUROC was considered to select optimal features for each model. The codebook and missing rates of each variable in the training and testing set are described in Multimedia Appendix 2.

XSL•FO
RenderX

Prediction Outcome

The predictive outcome of this study was the occurrence of sunitinib- or sorafenib-associated thyroid dysfunction. Patients were defined as a TSH level ≥4 mIU/L, a diagnosis of hypothyroidism with International Classification of Diseases codes (*ICD-9-CM*; *243* or *244* or *ICD-10*; *E02* or *E03)*, or use of levothyroxine (Anatomical Therapeutic Chemical code: H03AA02). These criteria were defined following previous studies that focused on drug-induced thyroid dysfunction [9,21,22].

Model Development and Validation

The study used three resampling strategies and raw data applied to the five algorithms to develop 20 machine learning models. The five algorithms included LR, RF, AdaBoost [23], LGBM [24], and GBDT [25]. Resampling strategies used were random over-sampling (ROS), borderline synthetic minority oversampling technique (B-SMT), and a combination of over- and under-sampling using SMT and Tomek links (STMK). Bayesian optimization with 5-fold cross-validation and the AUPRC were considered for hyperparameter tuning [26]. For boosting models, hyperparameters such as learning_rate and n_estimators were optimized to balance convergence speed and model complexity. Specifically, in LGBM, subsample and colsample_bytree were fine-tuned to mitigate overfitting. Key parameters including max_depth, min_samples_split, and min_samples_leaf were adjusted to improve generalization and prevent overfitting for RF. Table S3.1 in Multimedia Appendix 3 lists the ranges of hyperparameters used for tuning in each algorithm. Pseudocodes of the model development process are listed in Multimedia Appendix 4.

A stratified 5-fold cross-validation was first used for internal validation of the training dataset, and the performances of the 20 models were evaluated using the temporal validation cohort. The accuracy, precision, recall, F1-score, AUROC, and AUPRC were the metrics used for evaluating model performance. Model performance was compared and evaluated following the sequence of the highest AUPRC, AUROC value, and F1-score to select the best-performing model.

Threshold Adjustment, Model Interpretation, and Web-Based Application

To optimize the performance of the best-performing model, different thresholds were selected on the PRC based on different percentages of outcome predictions. Five cutoff points were chosen: (1) a high-risk threshold for identifying a lower number of patients potentially at risk; (2) a default threshold of 0.50; (3) an optimal threshold based on the maximum F1-score; (4) an equal threshold where precision, recall, and F1-score were equal, and (5) the low-risk threshold for identifying a higher number of patients potentially at risk. The F1-score, recall, precision, and accuracy were then compared for these 5 thresholds. The best model was further examined by SHAP analysis to explain feature importance and contribution at both the population and individual levels [27]. At the population level, SHAP summary plots were used to demonstrate feature importance and how the top-ranked features impacted outcome predictions. On the other hand, SHAP force plots were used to visualize how the features contributed to outcome predictions in specific patients.

The model was further integrated into a web-based application to increase the accessibility of the model. The Python Flask framework was used to develop the application programming interface. The value of each feature served as input data that were fed back into the model to generate a predictive probability of thyroid dysfunction. The consistency of feature scaling between input data and training data was ensured by applying the same scaling transformation. The input variables with missing values were imputed with the last observation or the same multivariate imputation by chained equations transformation applied during model development to ensure consistency in predictions.

Statistical Analysis

Patient characteristics were evaluated with independent *t* tests or the Wilcoxon rank-sum test for continuous variables, and the χ^2 test or Fisher exact test for categorical variables. A two-sided *P* value <.05 was considered statistically significant. Data were analyzed using SAS (version 9.4; SAS Institute), Python (version 3.9.5; Python Software Foundation), and R software (version 4.2.2; R Foundation for Statistical Computing). The statistical significance of the AUPRC was calculated using MedCalc software (version 22.001).

Results

Patient Characteristics and Multivariate Analysis

This study enlisted data from 900 patients prescribed sunitinib or sorafenib from initial screening from the CRD. A STROBE (Strengthening The Reporting of Observational Studies in Epidemiology) flowchart of patient selection is described in Multimedia Appendix 5. After applying the exclusion criteria, 807 patients remained. Patient characteristics and multivariate analysis of the training and temporal validation cohorts are listed in Multimedia Appendix 6. The training cohort contained 609 patients, with 52 (8.5%) patients developing thyroid dysfunction, while 16 (8.1%) patients experienced thyroid dysfunction in the temporal validation cohort of 198 patients. There were no significant differences in age or gender in the derivation or temporal validation cohort. Cancer type, histology, aspartate aminotransferase (AST), and bilirubin levels significantly differed (*P*<.001) between patients with and those without thyroid dysfunction in the two cohorts.

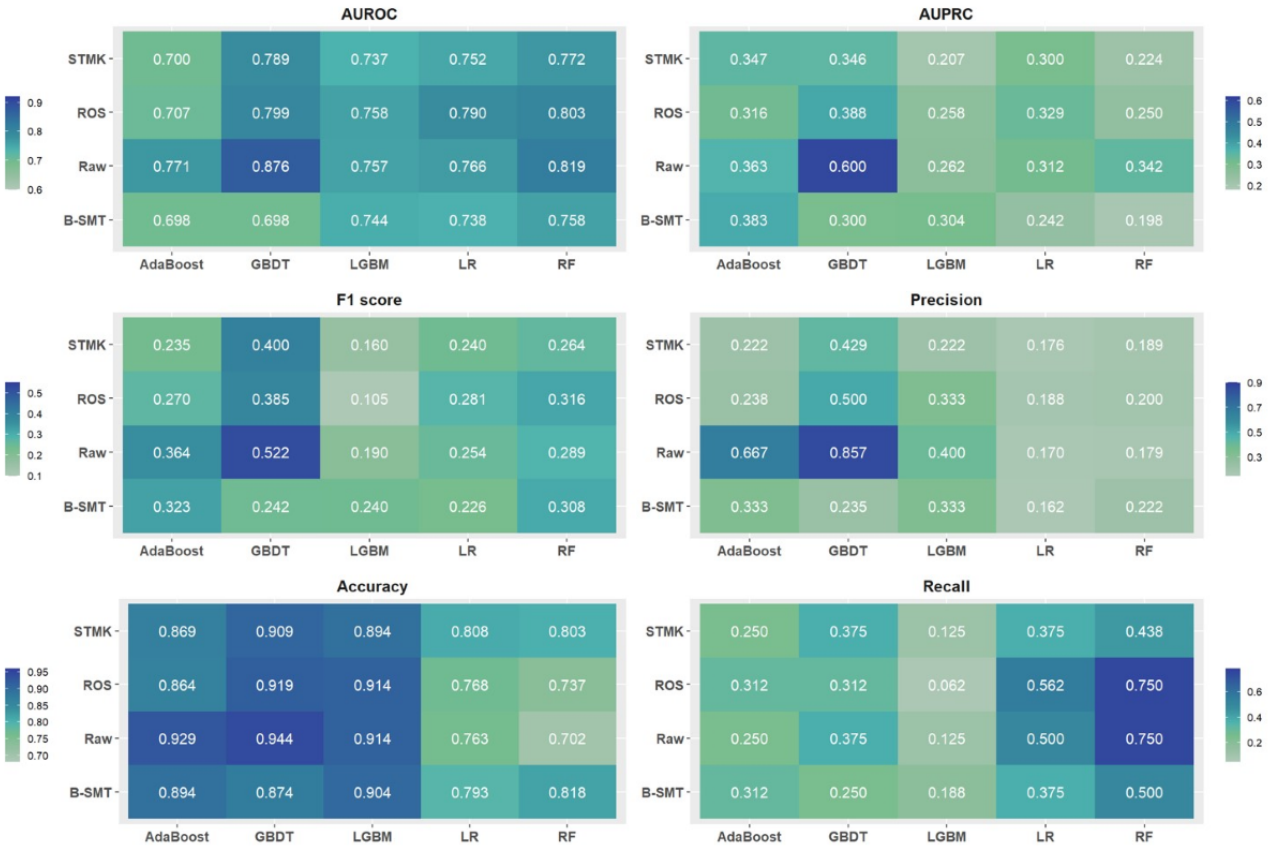
Model Development and Validation

The total number of features selected by recursive feature elimination were 20, 18, 15, 40, and 20 in GBDT, AdaBoost, LGBM, RF, and LR, respectively. The most frequently selected features were recent laboratory tests including AST, alanine transaminase (ALT), bilirubin, cholesterol, and triglyceride levels. Patient demographics such as age and BMI summed days of medication use, and follow-up duration were also among the top-ranked selected features (Multimedia Appendix 7). The optimal hyperparameter subsets selected by Bayesian optimization are listed in Table S3.2 in Multimedia Appendix 3.

XSL•FO
RenderX

Figure 2 shows the model performance based on the AUPRC, AUROC, F1-score, precision, recall, and accuracy in 20 ML models of the temporal validation cohort. Most of the GBDT and AdaBoost models had higher AUPRC values and F1-scores, while recall was higher in the RF and LR models. Of all the ML models, the GBDT without resampling (GBDT_RAW) outperformed the other models, with an AUPRC of 0.600, AUROC of 0.876, and an F1-score of 0.522. The AUPRC of GBDT_RAW (0.600, 95% CI 0.350-0.798) was significantly higher than those of the GBDT_ROS (0.388, 95% CI 0.175 - 0.618; P<.05), GBDT_BMST (0.300, 95% CI 0.124 - 0.549; P<.05), and GBDT_STMK (0.346, 95% CI 0.147 - 0.582; P<.05). Multimedia Appendix 8 lists model performances of 5-fold cross-validation with the training cohort and the statistical significance tests of the AUPRC of all the models.

Figure 2. Model performance on temporal validation. The model performance evaluated by the six metrics is demonstrated in a heatmap. Blue and green colors represent higher and lower values, respectively. ROS: random oversampling; BSMT: Oversampling with Borderline Synthetic Minority Oversampling Technique; STMK: Synthetic Minority Oversampling Technique-Tomek Links; GBDT: Gradient Boosting Decision Tree; AdaBoost: Adaptive Boosting; LGBM: Light Gradient Boosting Machine; RF: random forest; LR: logistic regression; AUPRC: area under the precision-recall curve; AUROC: area under the receiver operating characteristic curve.

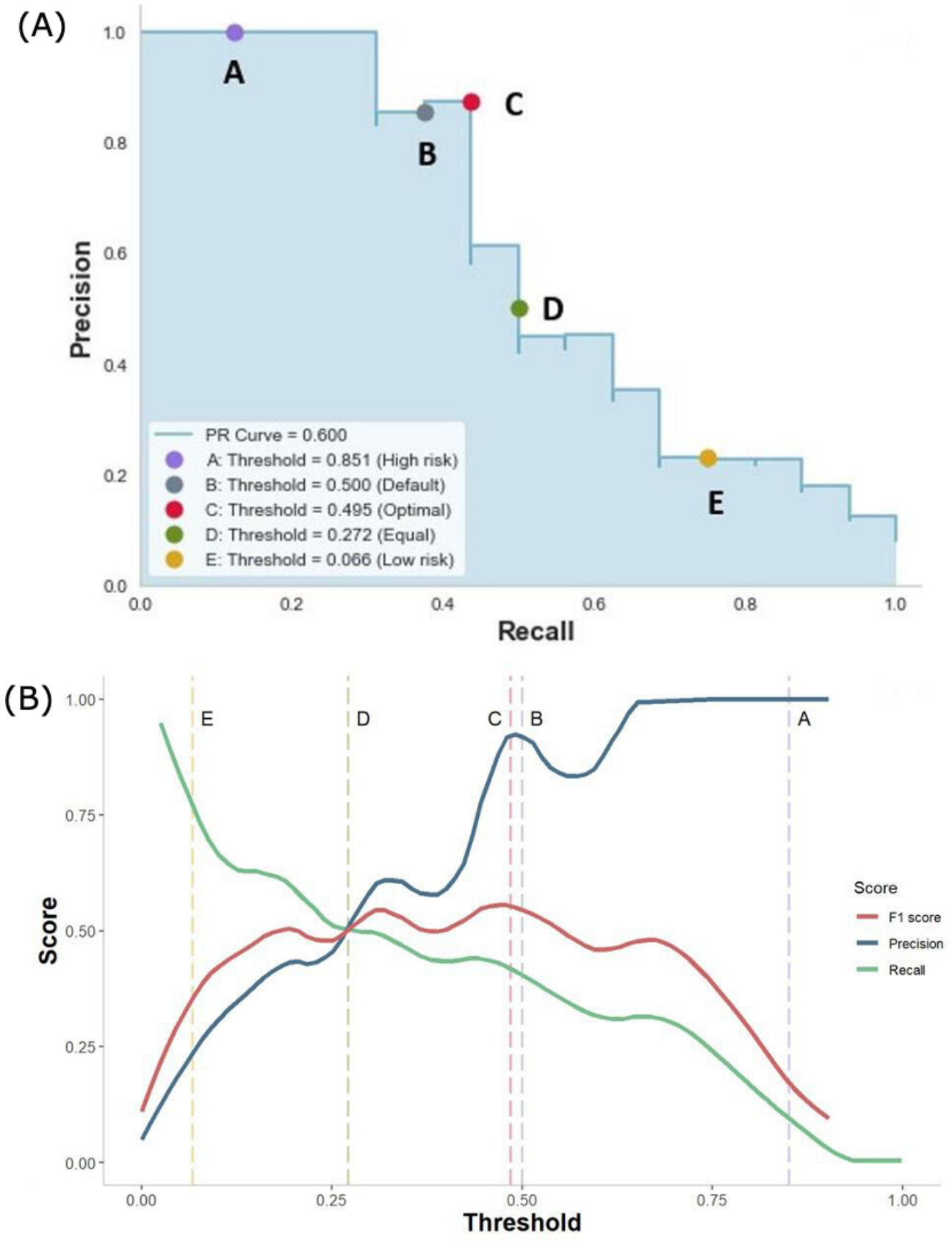


Threshold Adjustment

Figure 3 shows the five cutoff points on the PRC and the precision, recall, and F1-score at different thresholds of the GBDT_RAW model. Based on different percentages of outcome predictions, cutoff points were 0.851 for identifying the top 1% of patients (2 predictive positive cases, point A — high-risk threshold),0.500 for the top 3.5% (7 predictive positive cases, point B — default threshold), 0.495 for the top 4% (8 predictive positive cases, point C — optimal threshold based on the maximum F1-score), 0.272 for the top 8% (16 predictive positive cases, point D — equal threshold), and 0.071 for the top 26% of patients (51 predictive positive cases, point E — low-risk threshold). When moving the threshold from 0.500 to 0.851 (point A), the precision increased from 0.875 to 1.000, while recall significantly decreased to 0.125. In contrast, recall significantly increased to 0.750 and the precision decreased to 0.231 when changing the threshold to 0.066 (point E). The precision, recall, and F1-score reached 0.500 when the threshold was adjusted to 0.272 (point D). The optimal threshold of 0.495 (point C) for the GBDT_RAW model was selected based on the maximum F1-score, with a precision of 0.875, recall of 0.438, and F1-score of 0.583.

XSL•FO
RenderX

Figure 3. Threshold adjustment for the GBDT_RAW model. The five cutoff points on the precision-recall curve (A) represent different percentages of outcome predictions, precision, recall, and F1-scores based on different thresholds; (B) ,The X-axis represents the value of the threshold and the Y-axis shows values of the precision, recall, and F1-score.

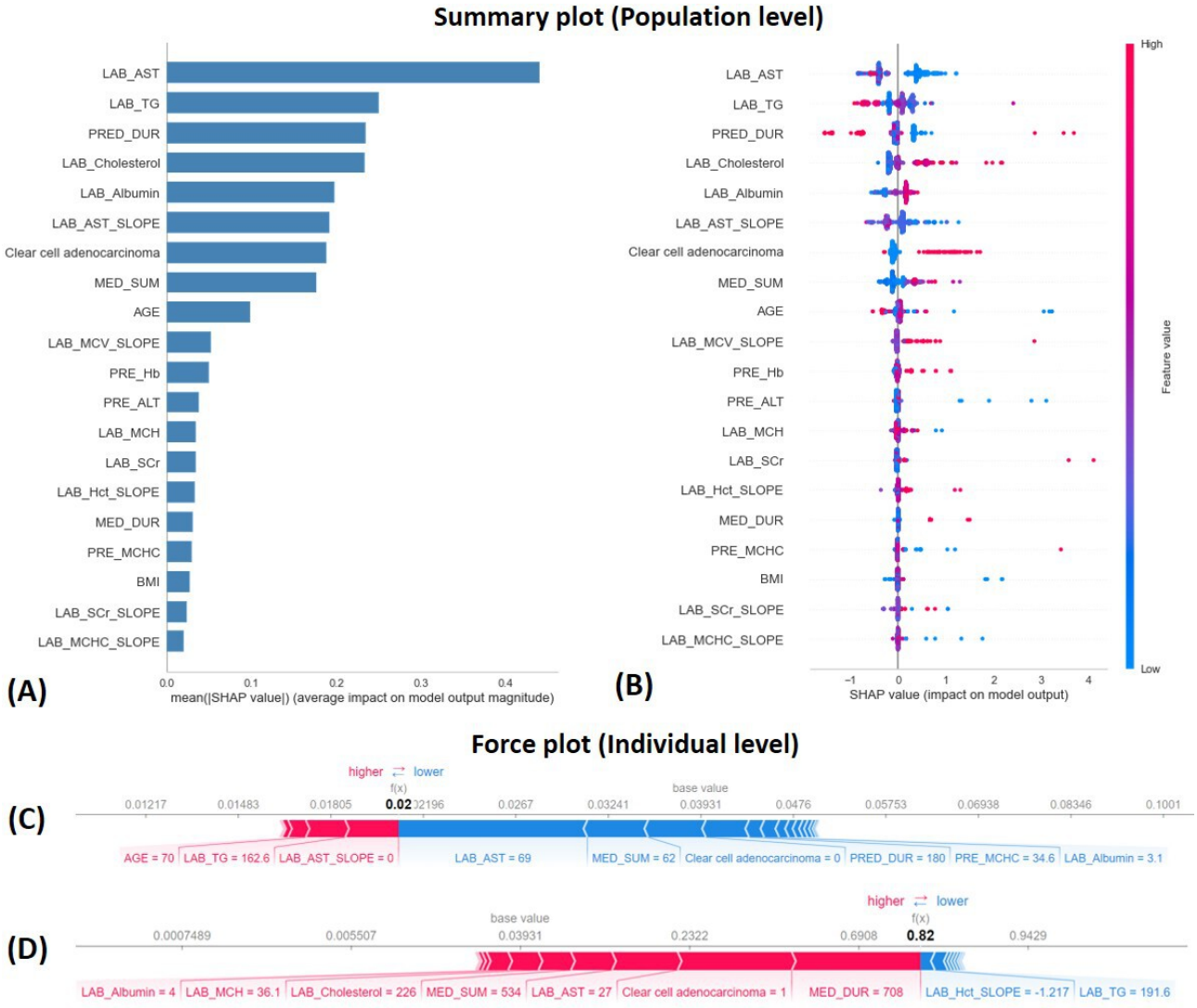


XSL•FO
RenderX

Model Interpretation

Figure 4 shows model interpretation implemented with the SHAP analysis on the best-performing GBDT_RAW model. The summary plot in Figure 4A displays that the top five important predictors were the AST level, TG level, follow-up duration, cholesterol level, and albumin level. Figure 4B shows that patients with lower AST levels, higher cholesterol levels, higher albumin levels, longer durations of medication use, and clear cell adenocarcinoma histology had higher SHAP values for predicting the occurrence of thyroid dysfunction. At the individual level, force plots showed the impacts of features on predicting thyroid dysfunction in two specific patients. Figure 4C shows a patient with a predictive probability of 0.02, resulting from his AST level of 69 IU/L and the summed days of medications of 62 days, which contributed most negatively to outcome predictions. For a patient with a predictive probability of 0.82 in Figure 4D, the medication duration of 708 days and clear cell adenocarcinoma served as the most important factors that positively impacted outcome predictions.

Figure 4. Model interpretation with a SHAP analysis for the GBDT_RAW model. At the population level, the summary bar plot (A) Mean SHAP values of all features, and a summary dot plot; (B) shows how each feature impacted the outcome prediction. The x-axis shows SHAP values of all features, and the colors represent the feature values, with red and blue respectively demonstrating higher and lower values. At the individual level, force plots in (C) and (D) show how features contributed to the model output value in each patient. Red and blue arrows represent positive and negative contributions of features, and the length of the arrow shows the magnitude of the impact on outcome predictions. LAB_AST: recent aspartate aminotransferase level; LAB_TG: recent triglyceride level; PRED_DUR: follow-up days; LAB_Cholesterol: recent cholesterol level; LAB_Albumin: recent albumin level; LAB_AST_SLOPE: slope of aspartate aminotransferase level; MED_SUM: sum days of medication; LAB_MCV_SLOPE: slope of mean corpuscular volume; PRE_Hb: previous hemoglobin level; PRE_ALT: previous alanine transaminase level; LAB_MCH: recent mean corpuscular haemoglobin level; LAB_SCr: recent serum creatinine level; LAB_Hct_SLOPE: slope of hematocrit level; MED_DUR: duration of medication; PRE_MCHC: previous mean corpuscular hemoglobin concentration level; LAB_SCr_SLOPE: slope of serum creatinine level; LAB_MCHC_SLOPE: slope of mean corpuscular hemoglobin concentration level.



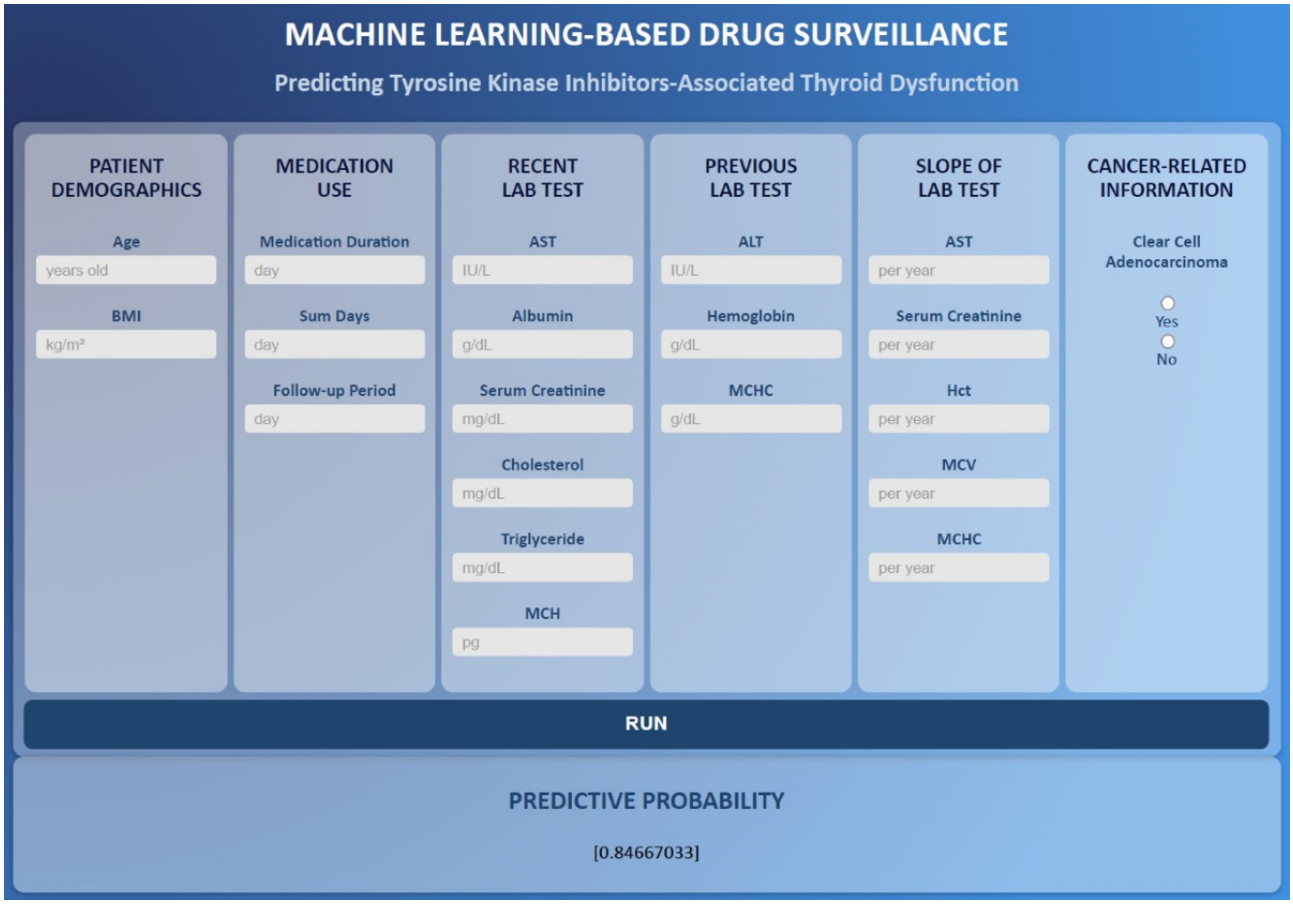
Web-Based Application

A web-based application for predicting sunitinib- and sorafenib-associated thyroid dysfunction was developed using the GBDT_RAW model (Figure 5). The application provides a user-friendly interface by showing 20 features selected by the GBDT model. Values of the features can be filled in with the appropriate units shown in each input box. The model then generates a predictive probability for the risk of sunitinib- or

XSL•FO
RenderX

sorafenib-associated thyroid dysfunction. Ultimate predictions for the occurrence of thyroid dysfunction were determined based on thresholds adopted by clinical users.

Figure 5. Web-based application with a user-friendly interface. A web-based application developed using the GBDT_RAW model showing 20 selected features selected and the predictive probability.



Discussion

Principal Findings

This study developed explainable ML models by collecting time-series data to predict sunitinib- and sorafenib-associated thyroid dysfunction. The present model allowed ongoing predictions according to the rapidly shifting status of the disease for patients undergoing long-term TKI treatment. The best-performing GBDT without resampling model was optimized by threshold moving strategies to achieve a maximum F1-score. This study further used a SHAP analysis that provided interpretability at both the population and individual levels, identifying key predictors such as AST, cholesterol, and albumin levels. The feature selection process revealed similarities with prior research on amiodarone-induced thyroid dysfunction, supporting the model's potential applicability to other drug-induced thyroid dysfunctions [9]. Integrating the model into a web-based application demonstrated its practical utility by allowing real-time risk estimation based on patient-specific data.

Time-Series Data Collection

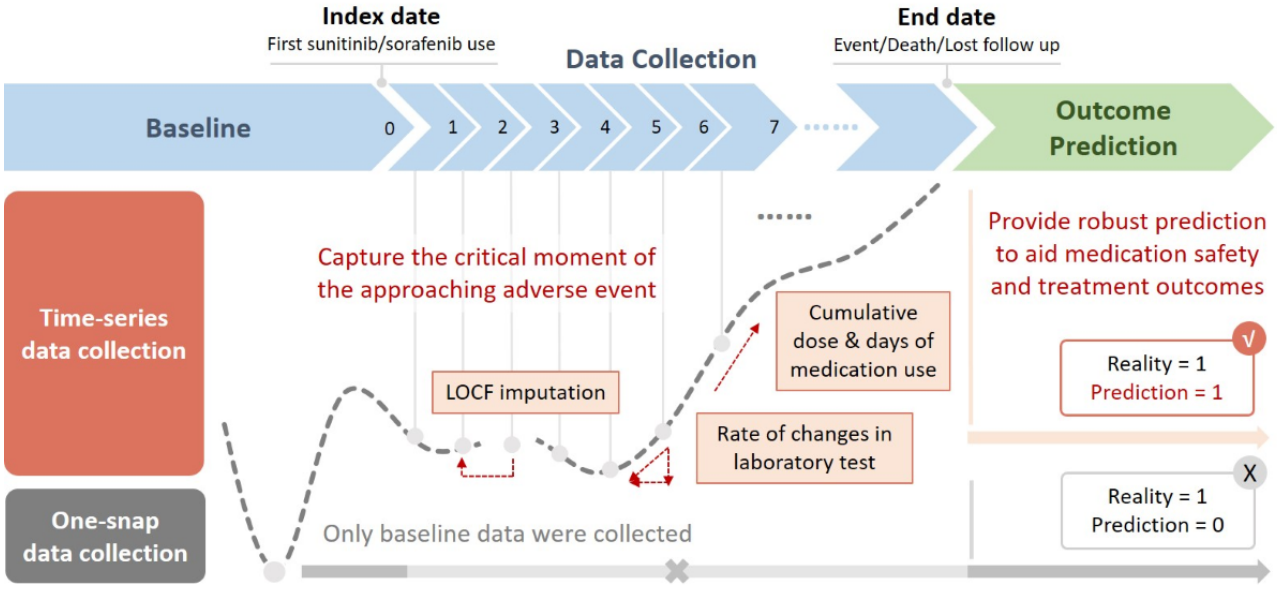
Time-series data collection in this study had a few advantages for thyroid dysfunction predictions on TKI users as illustrated in Figure 6. Clinical data at multiple time points in a long-term follow-up period were collected for model building. Time-series data allowed reengineering features by calculating rates of change of laboratory tests at recent and previous time points. The last observation carried forward method using the last value to replace missing data in subsequent time points ensured that missing values were imputed with plausible estimates [28]. This time-series data collection closely monitored patients newly prescribed TKI users and those on long-term TKI therapy. The time-series model captured the critical moment of an approaching adverse event for an individual patient, providing oncologists with invaluable information to treat patients on TKIs.

Identifying thyroid dysfunction was reported to potentially serve as a prognostic indicator for treatment outcomes. Sunitinib and sorafenib respectively demonstrated progression-free survival (PFS) periods of 11 and 5.5 months in patients with renal cell carcinoma [29,30]. Interestingly, studies indicated that patients with thyroid dysfunction had a longer PFS compared to patients with normal thyroid function among sunitinib (11.9 vs 8.8 months) and sorafenib users (19.3 vs 5.5 months) [4,31]. These findings suggest that thyroid dysfunction may not only represent an adverse event but could serve as a marker of enhanced therapeutic response. Leveraging a time-series data collection approach, this study can predict thyroid dysfunction risk while simultaneously identifying patients who may derive greater

XSL•FO
RenderX

clinical benefits from sustained TKI therapy. Unlike one-snap data collection which arbitrarily differentiates high-risk patients from a low-risk group, this time-series approach allows high-risk patients to use lifesaving TKIs for a relatively longer grace period which can significantly enhance their survival times. Models with time-series data collection balance the risks and benefits of survival and adverse drug events, bestowing precious survival periods for end-stage cancer patients with reasonable medication safety and improved quality of life.

Figure 6. Graphic comparison of one-snap and time-series data collection methods. Time-series and one-snap data collection methods were compared. The dashed line represents possible changes in variables during a long-term follow-up period. LOCF: Last Observation Carried Forward



Best-Performing Model With Threshold Adjustment

The GBDT_RAW model outperformed other machine learning models in this study, with thresholds adjusted to optimize model performance. The gradient boosting method analyzed nonlinear relationships and handled imbalanced data to predict diagnoses, hospital readmissions, and clinical outcomes [32-34]. During the hyperparameter tuning process, the AUPRC was considered, rather than the F1-score, without adjusting the threshold for imbalanced data [35,36]. The model without resampling outperformed resampling approaches by maintaining high precision, while the sparsity of the minority class in this study probably caused resampling to introduce noise or unrealistic synthetic data, failing to improve recall. The final model then used five cutoff points on the PRC to decide optimal thresholds and percentages of positive predictions. A relatively low threshold can identify all patients at risk but may increase the number of false positives, overwhelming clinicians with numerous warnings to review. On the contrary, choosing a relatively high threshold reduces the number of false positives but increases the chances of missing true positive cases. The selected threshold increased the precision, recall, and F1-score of the model's performance, while considering the relative costs of false positives and false negatives to meet clinical needs [37].

Model Interpretation and Potential for Model Expansion

Similar features for predicting drug-induced thyroid dysfunction were selected by the presence of TKIs and a previous amiodarone model built by our research team [9]. The feature selection process identified age, BMI, medication duration, summed days of medication, rate of change in serum creatinine levels, and recent mean corpuscular hemoglobin, AST, and cholesterol levels as key predictors of thyroid dysfunction in this study. Interestingly, these same features were also selected in our previous amiodarone model, suggesting their potential relevance across different drug-induced thyroid dysfunction models. Similar features highlight the potential for model expansion to increase generalizability. Top-ranked features identified in this study and their contributions calculated by the SHAP analysis were aligned with clinical domain knowledge. Higher cholesterol levels were found to positively impact predictions, which can be explained by the effect of thyroid function on lipoprotein metabolism [38]. The present study also found that longer summed days of medication increased the risk of thyroid dysfunction. This finding was confirmed by a long-term safety study of sunitinib, which showed the frequency of thyroid dysfunction, unlike other adverse events, gradually increased over time [39]. The AST level and histology of clear cell carcinoma served as important predictors, which may have resulted from a relatively lower incidence of thyroid dysfunction in patients with hepatocellular carcinoma and renal cell carcinoma.

Clinical Implications of Web-Based Drug Surveillance

ML-based drug surveillance provides a promising tool for predicting and managing thyroid dysfunction caused by TKIs. Traditional clinical decision support systems (CDSS) rely on rule-based alerts for drug interactions or contraindications but often fail to capture the multifactorial nature of drug-induced thyroid dysfunction, which depends on thyroid-related factors [40]. AI-driven CDSSs, leveraging ML algorithms, can overcome these limitations by identifying complex patterns in patient data, providing personalized risk assessments, and generating timely alerts for thyroid function monitoring. Deploying such ML models on cloud-based platforms alongside

XSL•FO
RenderX

rule-based CDSSs could enhance real-time risk stratification, enabling early detection, intervention, and potential adjustments in therapy to prevent endocrine complications. Further exploration of CDSSs incorporating ML predictions are warranted to improve clinical practice and medication safety [40].

Limitations

There are a few limitations of this study. Data were from a single healthcare system and included limited numbers of sunitinib- or sorafenib-treated patients. There were missing data in the diagnosis of comorbidities and medication use. The nature of a retrospective study introduces underestimations of the incidence of thyroid dysfunction. As a result, different features selected by the present model and our previous amiodarone model mainly resulted from features with higher missing rates. Multicenter and multicountry studies for improving model extrapolation are needed before clinical application. Future qualitative research and prospective studies with the involvement of physicians should also be conducted to assess the usability and accessibility of the model for real-world evidence.

Conclusions

This study applied time-series data collection to capture the critical moment of sunitinib- and sorafenib-associated thyroid dysfunction for ADR surveillance. The optimal threshold can balance the precision and recall meeting clinical needs. Feature importance was explained at the population and individual levels. The web-based application increased the model accessibility, allowing clinical users to receive real-time predictions. The comparison of features with amiodarone-induced thyroid dysfunction highlighted the potential for future model expansion.

Acknowledgments

This work was supported by the National Science and Technology Council (NSTC) [grant no. NSTC-113-2320-B-038-057].

Data Availability

The data were obtained from the Taipei Medical University clinical research database; this was approved by the Institutional Review Board of Taipei Medical University (TMU-JIRB-N202202053). The data and web-based application are available from the corresponding author upon reasonable requests, subject to appropriate request mechanisms and additional Institutional Review Board approval for personal information protection.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Time-series data collection.
[DOCX File, 364 KB - formative_v9i1e67767_app1.docx]

Multimedia Appendix 2
Codebook and missing rates of variables.
[DOCX File, 24 KB - formative_v9i1e67767_app2.docx]

Multimedia Appendix 3
Hyperparameters tuning.
[DOCX File, 21 KB - formative_v9i1e67767_app3.docx]

Multimedia Appendix 4
Pseudocodes of the model developing process.
[DOCX File, 18 KB - formative_v9i1e67767_app4.docx]

Multimedia Appendix 5
Patient selection flowchart.
[DOCX File, 141 KB - formative_v9i1e67767_app5.docx]

Multimedia Appendix 6
Patient characteristics and multivariate analysis.
[DOCX File, 23 KB - formative_v9i1e67767_app6.docx]

XSL•FO
RenderX

Multimedia Appendix 7
Feature selection with recursive feature elimination.
[DOCX File, 21 KB - formative_v9i1e67767_app7.docx]

Multimedia Appendix 8
Model performance of five-fold cross-validation.
[DOCX File, 24 KB - formative_v9i1e67767_app8.docx]

References

1. Torino F, Corsello SM, Longo R, Barnabei A, Gasparini G. Hypothyroidism related to tyrosine kinase inhibitors: an emerging toxic effect of targeted therapy. Nat Rev Clin Oncol 2009 Apr;6(4):219-228. [doi: 10.1038/nrclinonc.2009.4] [Medline: 19333228]
2. Abdel-Rahman O, Fouad M. Risk of thyroid dysfunction in patients with solid tumors treated with VEGF receptor tyrosine kinase inhibitors: a critical literature review and meta analysis. Expert Rev Anticancer Ther 2014 Sep;14(9):1063-1073. [doi: 10.1586/14737140.2014.929501] [Medline: 24927771]
3. Walko CM, Aubert RE, La-Beck NM, et al. Pharmacoepidemiology of clinically relevant hypothyroidism and hypertension from sunitinib and sorafenib. Oncologist 2017 Feb;22(2):208-212. [doi: 10.1634/theoncologist.2016-0233] [Medline: 28167571]
4. Vasileiadis T, Chrisofos M, Safioleas M, Kontzoglou K, Papazisis K, Sdrolia A. Impact of sunitinib-induced hypothyroidism on survival of patients with metastatic renal cancer. BMC Cancer 2019 Apr 30;19(1):407. [doi: 10.1186/s12885-019-5610-8] [Medline: 31039771]
5. Wu S, Chen JJ, Kudelka A, Lu J, Zhu X. Incidence and risk of hypertension with sorafenib in patients with cancer: a systematic review and meta-analysis. Lancet Oncol 2008 Feb;9(2):117-123. [doi: 10.1016/S1470-2045(08)70003-2] [Medline: 18221915]
6. Sabatier R, Eymard JC, Walz J, et al. Could thyroid dysfunction influence outcome in sunitinib-treated metastatic renal cell carcinoma? Ann Oncol 2012 Mar;23(3):714-721. [doi: 10.1093/annonc/mdr275] [Medline: 21653681]
7. Clemons J, Gao D, Naam M, Breaker K, Garfield D, Flaig TW. Thyroid dysfunction in patients treated with sunitinib or sorafenib. Clin Genitourin Cancer 2012 Dec;10(4):225-231. [doi: 10.1016/j.clgc.2012.08.002] [Medline: 23017335]
8. Jannin A, Penel N, Ladsous M, Vantyghem MC, Do Cao C. Tyrosine kinase inhibitors and immune checkpoint inhibitors-induced thyroid disorders. Crit Rev Oncol Hematol 2019 Sep;141:23-35. [doi: 10.1016/j.critrevonc.2019.05.015] [Medline: 31202955]
9. Lu YT, Chao HJ, Chiang YC, Chen HY. Explainable machine learning techniques to predict amiodarone-induced thyroid dysfunction risk: multicenter, retrospective study with external validation. J Med Internet Res 2023 Feb 7;25:e43734. [doi: 10.2196/43734] [Medline: 36749620]
10. Kim W, Cho YA, Kim DC, Jo AR, Min KH, Lee KE. Factors associated with thyroid-related adverse events in patients receiving PD-1 or PD-L1 inhibitors using machine learning models. Cancers (Basel) 2021 Oct 30;13(21):21. [doi: 10.3390/cancers13215465] [Medline: 34771631]
11. Wang Q, Wu T, Zhao R, et al. Development and validation of a prediction model for thyroid dysfunction in patients during immunotherapy. Endocr Pract 2024 Oct;30(10):943-950. [doi: 10.1016/j.eprac.2024.07.006] [Medline: 39004306]
12. Deng H, Eftekhari Z, Carlin C, et al. Development and validation of an explainable machine learning model for major complications after cytoreductive surgery. JAMA Netw Open 2022 May 2;5(5):e2212930. [doi: 10.1001/jamanetworkopen.2022.12930] [Medline: 35612856]
13. Hu J, Xu J, Li M, et al. Identification and validation of an explainable prediction model of acute kidney injury with prognostic implications in critically ill children: a prospective multicenter cohort study. EClinicalMedicine 2024 Feb;68:102409. [doi: 10.1016/j.eclinm.2023.102409] [Medline: 38273888]
14. Jiang Z, Bo L, Wang L, et al. Interpretable machine-learning model for real-time, clustered risk factor analysis of sepsis and septic death in critical care. Comput Methods Programs Biomed 2023 Nov;241:107772. [doi: 10.1016/j.cmpb.2023.107772] [Medline: 37657148]
15. Slack D, Hilgard S, Jia E, Singh S, Lakkaraju H. Fooling LIME and SHAP: adversarial attacks on post hoc explanation methods. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society 2020:180-186. [doi: 10.1145/3375627.3375830]
16. Ma M, Wan X, Chen Y, et al. A novel explainable online calculator for contrast-induced AKI in diabetics: A multi-centre validation and prospective evaluation study. J Transl Med 2023 Jul 31;21(1):517. [doi: 10.1186/s12967-023-04387-x] [Medline: 37525240]
17. Mosquera-Lopez C, Ramsey KL, Roquemen-Echeverri V, Jacobs PG. Modeling risk of hypoglycemia during and following physical activity in people with type 1 diabetes using explainable mixed-effects machine learning. Comput Biol Med 2023 Mar;155:106670. [doi: 10.1016/j.compbiomed.2023.106670] [Medline: 36803791]

XSL•FO
RenderX

18. Syrowatka A, Song W, Amato MG, et al. Key use cases for artificial intelligence to reduce the frequency of adverse drug events: a scoping review. Lancet Digit Health 2022 Feb;4(2):e137-e148. [doi: 10.1016/S2589-7500(21)00229-6] [Medline: 34836823]
19. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024 Apr 16;385:e078378. [doi: 10.1136/bmj-2023-078378] [Medline: 38626948]
20. van Buuren S, Oudshoorn CG. Multivariate Imputation by Chained Equations: Leiden: TNO; 2000. URL: https://stefvanbuuren.name/publications/MICE%20V1.0%20Manual%20TNO00038%202000.pdf
21. Kornelius E, Chiou JY, Yang YS, Peng CH, Lai YR, Huang CN. Iodinated contrast media increased the risk of thyroid dysfunction: a 6-year retrospective cohort study. J Clin Endocrinol Metab 2015 Sep;100(9):3372-3379. [doi: 10.1210/JC.2015-2329] [Medline: 26168278]
22. Feldt S, Schüssel K, Quinzler R, et al. Incidence of thyroid hormone therapy in patients treated with sunitinib or sorafenib: a cohort study. Eur J Cancer 2012 May;48(7):974-981. [doi: 10.1016/j.ejca.2012.01.036] [Medline: 22382202]
23. Freund Y, Schapire RE. Experiments with a new boosting algorithm. ICML'96: Proceedings of the Thirteenth International Conference on International Conference on Machine Learning 1996 Jul:148-156 [FREE Full text]
24. Ke G, Meng Q, Finley T, et al. Lightgbm: A highly efficient gradient boosting decision tree [abstract]. 2017 Presented at: 31st Conference on Neural Information Processing Systems (NIPS 2017); Dec 4-9, 2017 URL: https://papers.nips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html [accessed 2025-03-31]
25. Friedman JH. Greedy function approximation: A gradient boosting machine. Ann Statist 2001;29(5):1189-1232. [doi: 10.1214/aos/1013203451]
26. Snoek J, Larochelle H, Adams RP. Practical bayesian optimization of machine learning algorithms [abstract]. 2012 Presented at: Advances in neural information processing systems; Dec 3-6, 2012; Lake Tahoe, Nevada URL: https://papers.nips.cc/paper_files/paper/2012/hash/05311655a15b75fab86956663e1819cd-Abstract.html
27. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. 2017 Presented at: 31st Conference on Neural Information Processing Systems (NIPS 2017); Dec 4-9, 2017; Long Beach, CA URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf [accessed 2025-03-31]
28. Lachin JM. Fallacies of last observation carried forward analyses. Clin Trials 2016 Apr;13(2):161-168. [doi: 10.1177/1740774515602688] [Medline: 26400875]
29. Escudier B, Eisen T, Stadler WM, et al. Sorafenib in advanced clear-cell renal-cell carcinoma. N Engl J Med 2007 Jan 11;356(2):125-134. [doi: 10.1056/NEJMoa060655] [Medline: 17215530]
30. Motzer RJ, Hutson TE, Tomczak P, et al. Overall survival and updated results for sunitinib compared with interferon alfa in patients with metastatic renal cell carcinoma. J Clin Oncol 2009 Aug 1;27(22):3584-3590. [doi: 10.1200/JCO.2008.20.1293] [Medline: 19487381]
31. Schmidinger M, Vogl UM, Bojic M, et al. Hypothyroidism in patients with renal cell carcinoma: blessing or curse? Cancer 2011 Feb 1;117(3):534-544. [doi: 10.1002/cncr.25422] [Medline: 20845482]
32. Silva NCD, Albertini MK, Backes AR, Pena GDG. Machine learning for hospital readmission prediction in pediatric population. Comput Methods Programs Biomed 2024 Feb;244:107980. [doi: 10.1016/j.cmpb.2023.107980] [Medline: 38134648]
33. Duan H, Zhang Y, Qiu H, et al. Machine learning-based prediction model for distant metastasis of breast cancer. Comput Biol Med 2024 Feb;169:107943. [doi: 10.1016/j.compbiomed.2024.107943] [Medline: 38211382]
34. May SB, Giordano TP, Gottlieb A. Generalizable pipeline for constructing HIV risk prediction models across electronic health record systems. J Am Med Inform Assoc 2024 Feb 16;31(3):666-673. [doi: 10.1093/jamia/ocad217]
35. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE 2015;10(3):e0118432. [doi: 10.1371/journal.pone.0118432] [Medline: 25738806]
36. Lipton ZC, Elkan C, Naryanaswamy B. Optimal thresholding of classifiers to maximize F1 measure. Mach Learn Knowl Discov Databases 2014;8725:225-239. [doi: 10.1007/978-3-662-44851-9_15] [Medline: 26023687]
37. Rizos CV, Elisaf MS, Liberopoulos EN. Effects of thyroid dysfunction on lipid profile. Open Cardiovasc Med J 2011;5:76-84. [doi: 10.2174/1874192401105010076] [Medline: 21660244]
38. Porta C, Gore ME, Rini BI, et al. Long-term safety of sunitinib in metastatic renal cell carcinoma. Eur Urol 2016 Feb;69(2):345-351. [doi: 10.1016/j.eururo.2015.07.006] [Medline: 26215605]
39. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17. [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
40. Flynn A. Using artificial intelligence in health-system pharmacy practice: finding new patterns that matter. Am J Health Syst Pharm 2019 Apr 17;76(9):622-627. [doi: 10.1093/ajhp/zxz018] [Medline: 31361834]

Abbreviations

AdaBoost: Adaptive Boosting

XSL•FO
RenderX

ADR: adverse drug reaction
ALT: alanine transaminase
AST: aspartate aminotransferase
AUPRC: area under the precision recall curve
AUROC: area under the receiver operating characteristic curve
CDSS: clinical decision support systems
CRD: clinical research database
GBDT : Gradient Boosting Decision Tree
LGBM: Light Gradient Boosting Machine
LR: logistic regression
ML: machine learning
RF: random forest
TKI: tyrosine kinase inhibitor
TMUH: Taipei Medical University Hospital
TRIPOD+AI: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis + artificial intelligence
TSH: thyroid-stimulating hormone

Edited by A Mavragani; submitted 23.10.24; peer-reviewed by AP Vit, L Guo; revised version received 03.03.25; accepted 04.03.25; published 10.04.25.

Please cite as:
Chan FY, Ku YE, Lie WN, Chen HY
Web-Based Explainable Machine Learning-Based Drug Surveillance for Predicting Sunitinib- and Sorafenib-Associated Thyroid Dysfunction: Model Development and Validation Study
JMIR Form Res 2025;9:e67767
URL: https://formative.jmir.org/2025/1/e67767
doi:10.2196/67767

© Fan-Ying Chan, Yi-En Ku, Wen-Nung Lie, Hsiang-Yin Chen. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Assessment of Gait Parameters Using Wearable Sensors and Their Association With Muscle Mass, Strength, and Physical Performance in Korean Older Adults: Cross-Sectional Study

Jinyoung Shin[1], MD, PhD; Hyuk Jung Kweon[2], MD, PhD; Jaekyung Choi[1], MD, PhD

[1]Department of Family Medicine, Konkuk University Medical Center, Konkuk University School of Medicine, Seoul, Republic of Korea

[2]Department of Family Medicine, Konkuk University Medical Center, Chungju Hospital, Konkuk University School of Medicine, 268 Chungwon-daero, Chungju, Republic of Korea

Corresponding Author:
Hyuk Jung Kweon, MD, PhD
Department of Family Medicine, Konkuk University Medical Center, Chungju Hospital, Konkuk University School of Medicine, 268 Chungwon-daero, Chungju, Republic of Korea

Abstract

Background: Gait speed indicates the onset or decline of physical performance in sarcopenia. However, real-time measurements of other gait parameters, such as step length, stride length, step width, and support time, are limited. The advent of wearable technology has facilitated the measurement of these parameters, necessitating further investigation into their potential applications.

Objective: This study aimed to investigate the relationship between gait parameters measured using wearable sensors and muscle mass, strength, and physical performance in community-dwelling older adults.

Methods: In a cross-sectional study of 91 participants aged ≥65 years, gait parameters, such as step count, step length, cadence, single and double support times, vertical oscillation, and instantaneous vertical loading rate (IVLR), measured using a wireless earbud device, were analyzed on the basis of the appendicular skeletal muscle mass index (SMI), calf circumference, handgrip strength, 5-time chair stand test, short physical performance battery (SPPB), and the SARC-F (strength, assistance with walking, rise from a chair, climb stairs and fall frequency) questionnaire. This study was conducted from July 10 to November 1, 2023, at an outpatient clinic of a university hospital in Seoul, Korea. Multiple regression analysis was performed to investigate independent associations after adjusting for age, sex, BMI, and comorbidities.

Results: Among 91 participants (45 men and 46 women; mean age 74.1 years for men and 73.6 years for women), gait speed and vertical oscillation showed negative associations with their performance in the 5-time chair stand test (P<.001) and SARC-F and positive associations with their performance in the SPPB (P<.001). Vertical oscillations were also associated with grip strength (P=.003). Single and double support times were associated with performance in the 5-time chair stand test and SPPB (P<.001). In addition, double support time was associated with SARC-F scores (P<.001). Gait speed, support time, vertical oscillation, and IVLR showed independent associations with performance in the 5-time chair stand test and SPPB (P<.001), both related to muscle strength or physical performance. Gait speed, double support time, and vertical oscillation all had significant associations with SARC-F scores.

Conclusions: This study demonstrated a significant association between gait monitoring using wearable sensors and quantitative assessments of muscle strength and physical performance in older people. Furthermore, this study substantiated the extensive applicability of diverse gait parameters in predicting sarcopenia.

(JMIR Form Res 2025;9:e63928) doi:10.2196/63928

KEYWORDS

gait analysis; sarcopenia; wearable electronic devices; muscle mass; physical performance; older adults; geriatric; cross-sectional study; outpatient clinic; Korea; mHealth; mobile health

Introduction

Sarcopenia is an age-related condition characterized by the progressive loss of skeletal muscle mass and strength, as well as physical performance [1], which increases the risk of disability, falls, cognitive decline, loss of independence, and mortality among community-dwelling older adults [2-5]. The International Clinical Practice Guidelines for Sarcopenia recommend annual clinical assessments or more frequent evaluations after major health events, such as hospitalization, to detect early signs of frailty [6,7]. According to the 2019 Asian Working Group for Sarcopenia (AWGS) consensus update, the diagnosis of sarcopenia should be based on changes in muscle strength, physical performance, and muscle mass, with case

XSL•FO
RenderX

findings leading to a comprehensive assessment and diagnosis in primary care or community preventive settings [8-10].

Gait monitoring, particularly gait speed, has been identified as a valid indicator of sarcopenia and is associated with health outcomes such as disability and mortality in older adults [11,12]. Although various gait parameters such as step length, stride length and width, and single support time during walking can be detected, their use in research has been limited owing to challenges in real-time measurement and result standardization [13]. Wearable sensors have made it easier to conduct real-time evaluations. A meta-analysis of 82 studies addressing over 100 gait parameters outcomes discovered high validity and reliability, particularly for spatiotemporal parameters such as step time and stride time [14].

Because smartphone apps enable the measurement of spatiotemporal outcomes while older adults walk, the clinical utility of real-time monitoring of changes in mobility among older adults using wearable sensors has been highlighted [15]. Daily walking count guidance using wearable sensors improves physical activity and reduces sarcopenia in older adults in Taiwan [16]. A study examining men older than 50 years in Korea using a wearable smart belt for 4 weeks discovered significant correlations between gait speed and sarcopenia, as defined by grip strength and dual-energy x-ray absorptiometry [17]. Several studies measured gait parameters using watches or footwear [18]. However, few studies have investigated the association between various gait parameters other than gait speed or step length and sarcopenia diagnosis based on muscle strength, muscle mass, and physical performance using an earbud device [19,20]. This study aimed to investigate the relationship between various gait parameters measured using wearable sensors (particularly earbud-type sensors) and sarcopenia, based on muscle mass, strength, and physical performance, among community-dwelling older adults.

Methods

Study Design and Participants

This is a cross-sectional study. The primary objective of this study was to investigate the association between gait parameters and various sarcopenia measurements such as muscle mass, strength, and physical performance. The secondary objective was to determine if newly usable gait parameters, which can be easily measured using wearable sensors, were significantly associated with sarcopenia. In the initial study, the target sample size for both men and women was 45 individuals in an initial study [21]. This analysis was performed as a secondary study, using data from a comparison study of a portable device for muscle mass and bioelectrical impedance analysis (BIA). At an older adult community center, we recruited 91 participants aged 65 years or older who could perform daily activities independently. We displayed a recruitment poster at the community center, inviting older adults interested in the study to apply to participate. This study was performed at the outpatient clinic of a university hospital in Seoul, Korea, between July 10 and November 1, 2023. The exclusion criterion was ongoing treatment for uncontrolled conditions such as malignancy, acute stroke, or dementia.

Muscle Mass, Strength, and Physical Performance Measurements

Muscle mass was determined using the appendicular skeletal muscle mass index (SMI) and calf circumference. The SMI is calculated by dividing the total lean mass of both arms and legs by the square of the height (kg/m^2) as measured using BIA (InBody 770; Cerritos). Calf circumference was measured at the thickest point on both sides, with moderate to high sensitivity and specificity for detecting low skeletal muscle mass [10,22]. Muscle strength was determined by the grip strength measured 3 times at 30-second intervals for each hand in a standing position, alternating between the dominant and nondominant sides, using a Smedley handheld dynamometer [23]. The 5-time chair stand test, which assesses both physical performance and muscle strength, requires participants to stand and sit 5 times as quickly as possible from a straight-backed chair with no arm support [24]. A short physical performance battery (SPPB) was used to assess balance, gait, and lower limb strength. This included standing with feet side-by-side, semitandem, and tandem positions; walking 8 feet; and rising from a chair to return to a seated position 5 times [25]. The SARC-F (strength, assistance with walking, rise from a chair, climb stairs and fall frequency) questionnaire was used to assess sarcopenia risk based on its 5 components, with fall frequency ranging 0-10. A total score of ≥4 indicates sarcopenia. The Korean version of the SARC-F was previously validated [26,27]. According to the AWGS consensus report in 2014, we used the cutoff values for SMI (less than 7.0 kg/m^2 for men and less than 5.7 kg/m^2 for women using BIA) and calf circumference (<34 cm for men and <33 cm for women) [1]. Low muscle strength was defined as grip strength of <28 kg for men and <18 kg for women. The cutoff value for the 5-time chair stand test was 12 seconds or more, while that of the SPPB was ≤9 (range 0 - 12) points [10].

Gait Parameters

Gait characteristics were measured using a wireless earbud (Beflex) device (Figure 1) equipped with a BiomechEngine chip for accurate accelerometer data analysis and gait parameter estimation. In young adults, the earbud device's validity was confirmed by measuring walking and running parameters [28]. Participants were instructed to insert the device into both ears and walk for a minimum of 30 seconds to record various gait metrics such as step count, step length, cadence, gait speed, times of single and double support, vertical oscillation, and instantaneous vertical loading rate (IVLR). Step count is defined as the number of steps analyzed within a 2-second interval, where a step from the left foot to the right foot is considered a single cycle of 2 steps. Step length was calculated as the distance between the heel of the front foot and heel of the back foot during the walking cycle. Gait speed and cadence were automatically calculated. Single support time was defined as the average duration of one foot supporting the body while walking. Double support time was defined as the average duration of both feet's support and vertical oscillation, which was defined as the average distance of the body's center moved vertically during walking [28]. The IVLR measures the steepest slope of the vertical ground reaction force at the points of left- and right-foot contact. The algorithms used to estimate these

XSL•FO
RenderX

gait parameters are not currently disclosed; however, they are commercially available. Gait parameters were measured a few minutes after walking.

Figure 1. Wearable sensors for gait monitoring (5.7 g, 17.1 × 20.2 × 22 mm^3).

Ethical Considerations

The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the institutional review board of Konkuk University Medical Center (KUMC 2022-11-033). Informed consent was obtained from all subjects enrolled in the study. Written informed consent was obtained from the patients for publication of this paper. The authors attest that there was no use of generative artificial intelligence technology in the generation of text, figures, or other informational content in this manuscript.

Statistical Analysis

Descriptive statistics were used to summarize participants' characteristics based on sex. Independent samples 2-tailed *t* tests were used to compare continuous variables, whereas chi-square tests were used for categorical variables. Pearson correlation analysis was used to determine the relationship between gait parameters and sarcopenia. We performed a logarithmic transformation to measure gait, muscle, and physical performance to adjust for a skewed distribution. After adjusting for age, sex, BMI, and comorbidities, multiple regression analysis was used to investigate the independent associations between sarcopenia and specific gait parameters for the secondary end point. Participants reported their age, sex, and comorbidities, as well as any daily life difficulties caused by low vision or hearing impairment ("yes" or "no"). Height and weight were measured to calculate the BMI (kg/m^2) and waist circumference at the narrowest point. Bonferroni correction for the evidential threshold was performed based on the number of exposures for the 6 sarcopenia indices; therefore, statistical significance was set at $P<.05$ divided by 6 (ie, a threshold of $P<.0083$ per exposure). All statistical analyses were conducted using IBM SPSS Statistics (version 27.0).

Results

This study included 91 participants (45 men and 46 women). The participants' characteristics are summarized in Table 1. The mean age of the study participants was 73.9 years, with more than half of them aged between 65 and 74 years. According to Asian criteria for overweight and obesity, the participants had a normal BMI and waist circumference [29].

While there were differences in grip strength, calf circumference, and SMI between men and women, there were no differences in their performance in the 5-time chair stand test, SPPB, or SARC-F (Table 2). The prevalence of sarcopenia varied according to the definition used for each measurement, ranging from 4.3% (SARC-F) to 82.6% (5-time chair stand test). The prevalence of sarcopenia according to grip strength was higher in women than in men. However, the prevalence of sarcopenia, defined by the SMI, was higher in men than in women. There were no in the other indicators between men and women.

Table 3 lists the association between various gait parameters and sarcopenia measurements. Gait speed, single and double support times, vertical oscillation, and IVLR demonstrated significant correlations with performance on the 5-time chair stand test and the SPPB. Additionally, gait speed, double support time, and vertical oscillation were associated with performance on the SARC-F. Vertical oscillation positively correlated with grip strength, and step count negatively correlated with performance on the 5-time chair stand test. However, after Bonferroni correction, step length and cadence did not show

XSL•FO
RenderX

any significant correlations with sarcopenia. Calf circumference and SMI were not associated with gait parameters.

After adjusting for confounding factors, gait speed, double support time, and vertical oscillation were significantly correlated with the 5-time chair stand test, SPPB, and SARC-F scores (Table 4). Single support time and IVLR were significantly correlated with performance in the 5-time chair stand test and SPPB.

Table . Characteristics of participants.

		Total (n=91)	Men (n=45)	Women (n=46)
Age (years), mean (SD)		73.9 (5.5)	74.1 (5.0)	73.6 (6.1)
Age groups (years), n (%)				
	65 - 74	51 (56.0)	24 (53.3)	27 (58.7)
	75 - 84	38 (41.8)	21 (46.7)	17 (37.0)
	≥85	2 (2.2)	0 (0)	2 (4.3)
BMI (kg/m^2), mean (SD)		24.4 (2.8)	24.6 (2.5)	24.2 (3.1)
Waist circumference (cm), mean (SD)		80.0 (8.9)	83.0 (7.4)	77.0 (9.3)
Number of comorbidities, mean (SD)		1.46 (1.07)	1.58 (0.97)	1.35 (1.16)
Comorbidities, n (%)				
	None	17 (18.7)	5 (11.1)	12 (26.1)
	Cardiovascular disease	69 (75.8)	40 (88.9)	29 (63.0)
	Musculoskeletal disease	14 (15.4)	2 (4.4)	12 (26.1)

Table . Diagnostic measurements of sarcopenia in study participants.

	Measurements			Sarcopenia		
	Men (n=45), mean (SD)	Women (n=46), mean (SD)	*P* value	Men (n=45), n (%)	Women (n=46), n (%)	*P* value
5-time chair stand test[a]	15.7 (5.5)	15.7 (4.6)	.96	33 (73.3)	38 (82.6)	.29
SPPB[b]	9.0 (1.5)	8.9 (1.5)	.73	29 (64.4)	29 (63.0)	.89
SARC-F[c]	0.66 (1.51)	0.87 (1.34)	.50	3 (6.7)	2 (4.3)	.63
Grip strength[d]	28.8 (5.6)	18.1 (4.8)	<.001	13 (28.9)	23 (50.0)	<.001
Calf circumference[e]	34.2 (2.8)	32.1 (2.5)	<.001	20 (44.4)	25 (54.3)	.35
Skeletal muscle index[f]	7.7 (0.71)	6.2 (0.6)	<.001	7 (15.6)	5 (10.9)	.02

[a]Cutoff was ≥12 seconds.

[b]SPPB: short physical performance battery; cutoff was ≤9 points.

[c]SARC-F: strength, assistance with walking, rise from a chair, climb stairs and fall frequency; cutoff was ≥4 points.

[d]Cutoff was <28 kg for men and <18 kg for women.

[e]Cutoff was <34 cm for men and <33 cm for women.

[f]Cutoff was <7.0 kg/m^2 for men and <5.7 kg/m^2 for women.

XSL•FO
RenderX

Table . Associations between gait parameters and sarcopenia measurements (n=91). All variables were subjected to correlation analysis after log transformation and Bonferroni correction was applied (*P*<.05/6 [.0083]).

		5-time chair stand test	SPPB[a]	SARC-F[b]	Grip strength	Calf circumference	SMI[c]
Step count							
	Coefficient	−0.322	0.163	−0.137	−0.071	−0.198	−0.204
	P value	.002	.12	.20	.50	.06	.053
Step length							
	Coefficient	−0.221	0.237	−0.209	−0.162	−0.176	−0.162
	P value	.04	.02	.047	.13	.10	.13
Cadence							
	Coefficient	−0.018	−0.036	−0.191	0.057	−0.041	0.040
	P value	.86	.75	.07	.59	.70	.71
Gait speed							
	Coefficient	−0.437	0.381	−0.281	−0.176	−0.196	−0.230
	P value	<.001	<.001	.007	.10	.06	.03
Single support time							
	Coefficient	0.489	−0.350	0.176	0.167	0.158	0.255
	P value	<.001	.001	.10	.11	.14	.02
Double support time							
	Coefficient	0.604	−0.541	0.371	−0.007	0.083	0.151
	P value	<.001	<.001	<.001	.95	.43	.15
Vertical oscillation							
	Coefficient	−0.387	0.518	−0.467	0.306	0.107	0.198
	P value	<.001	<.001	<.001	.003	.31	.06
IVLR[d]							
	Coefficient	−0.530	0.449	−0.270	−0.032	−0.128	−0.177
	P value	<.001	<.001	.01	.76	.23	.09

[a]SPPB: short physical performance battery.

[b]SARC-F: strength, assistance with walking, rise from a chair, climb stairs and fall frequency.

[c]SMI: appendicular skeletal muscle mass index.

[d]IVLR: instantaneous vertical loading rate.

Table . Multiple regression analysis for sarcopenia and gait indexes in older adults, having adjusted for age, sex, waist circumference, BMI, and comorbidities. Bonferroni correction was applied (*P*<.05/6 [.0083]).

	5-time chair stand test		SPPB[a]		SARC-F[b]	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
Gait speed	−0.393	<.001	0.331	.001	−0.328	.003
Single support time	0.545	<.001	−0.384	.001	0.222	.055
Double support time	0.558	<.001	−0.452	<.001	0.389	<.001
Vertical oscillation	−0.306	.008	0.399	<.001	−0.528	<.001
IVLR[c]	−0.496	<.001	0.392	<.001	−0.267	.02

[a]SPPB: short physical performance battery.

[b]SARC-F: strength, assistance with walking, rise from a chair, climb stairs and fall frequency.

[c]IVLR: Instantaneous vertical loading rate.

XSL•FO
RenderX

Discussion

Principal Findings

Gait measurements using wearable sensors show a strong correlation with physical performance and muscle strength. This study confirmed that the gait speed, double support time, and vertical oscillation measured using wearable sensors were reliable walking indicators of sarcopenia in community-dwelling older adults. Furthermore, using real-time gait monitoring measurements, we found that single support time and IVLR were good indicators of sarcopenia.

Age-related alterations in gait, such as shorter steps, increased double support time, lower cadence, or a wider step width, indicate a decline in gait speed [30]. These changes are attributed to a combination of sensorimotor factors, including motor unit reduction, impaired muscular activation, fiber-type substitution, decreased cutaneous sensation, and slower nerve and reaction speeds [12]. Furthermore, functional brain impairment, manifested as reduced brain volume, may be associated with deterioration in gait performance [31]. Accordingly, among the diagnostic indicators of sarcopenia, gait parameters may show a high correlation with the physical performance of participants.

However, these gait parameters were not associated with calf circumference or SMI, which are indicators of muscle mass. BIA, combined with the measurement of anthropometric metrics such as calf circumference, provides a practical approach for estimating muscle mass in nonclinical settings [11]. Individual factors, such as edema or diuretics may influence BIA [32], and muscle mass changes in older adults are not uniform [22], necessitating caution when interpretating their correlation with gait parameters. However, it is important to note that muscle mass is not a good predictor of adverse health–related outcomes such as mobility limitation, falls, and mortality in community-dwelling older adults [9]. Therefore, sarcopenia in older adults should be diagnosed holistically with muscle strength and physical performance taking precedence over muscle mass. Although this study did not identify a universal correlation between handgrip strength and gait parameters, quadriceps strength has been shown to influence gait speed, step length, and cadence [30]. This emphasizes the limitations of handgrip strength as an indicator of overall muscle strength in healthy older adults [33,34]. Further studies are required to determine the association between upper and lower limb muscle strength and gait parameters.

Among the gait parameters, gait speed, either alone or as part of a combined tool, such as the SPPB, has been confirmed to be associated with sarcopenia, which is the most comprehensive indicator of brain function and structural abnormalities [35]. Double support time correlates with balance in the cortico-subcortical white matter sensory and motor tracts [35]. Gait speed and double support time are direct indicators obtained while walking, whereas vertical oscillation and IVLR are known to be sensitive in reflecting changes in brain function as indirect data derived from direct indicators [36]. We confirmed the independent association between vertical oscillation and sarcopenia. Therefore, analyzing various gait parameters could improve the accuracy of sarcopenia diagnosis in older adults. Gait measurement using wearable sensors, which are easy to use and allow continuous and periodic monitoring of daily activities, offers significant promise for the early detection of sarcopenia in both the home and caregiving facilities. In general, older adults have difficulty with consistent use on consecutive days. Therefore, compliance with wearable sensors must be considered when evaluating the effectiveness of wearables in monitoring physical activity in older populations [37].

Limitations

However, several limitations of this study should be considered when applying gait measurements to older adults. First, the majority of participants were 65-74 years old, relatively healthy, and capable of independent daily living; hence, our results may not be applicable to all older populations. Additional research is required to apply this device to older adults living in nursing homes or other specialized care facilities. Second, we were unable to compare differences in wearable sensor techniques because our study used earbud-type devices rather than those more commonly placed on the foot or shank or used an insole pressure-sensing method [38]. We did not conduct repeated tests to assess test-retest reliability because it was difficult to revisit older participants. Third, because this study was conducted using a cross-sectional design, we could not establish a temporal relationship between the deterioration of gait parameters and sarcopenia. It is important to note that differences in the prevalence of sarcopenia based on the definition of muscle mass, strength, and physical performance may influence the correlation with gait parameters. Specifically, because our study did not compare gait parameters based on different rates of sarcopenia diagnosis, clinical interpretation should be cautious when predicting sarcopenia directly. Because race or lifestyle habits may have affected our results, it is important to consider the characteristics of the participants when applying the findings of this study. Finally, single and double support times, vertical oscillation, and IVLR still have limitations in practical applications, because diagnostic criteria have yet to be established. The algorithms for estimating these gait parameters are not currently disclosed; therefore, further studies are required to identify the clinical cutoff points to predict sarcopenia.

Conclusions

Using earbud-type wearable sensors, this study confirmed that gait parameters, particularly gait speed, double support time, and vertical oscillation, are correlated to muscle strength and physical performance in community-dwelling older adults. Given the limitations associated with the current diagnostic criteria and the potential unreliability of single measurements, further research is imperative to establish efficacy measures and broaden the applicability of diverse gait parameters in sarcopenia diagnosis.

XSL•FO
RenderX

Acknowledgments

We thank B-Lab and Beflex Inc for their cooperation. This research was funded by the Olive Healthcare (South Korea). The funders had no role in the design of the study, collection of data, analyses, interpretation of data, writing of the manuscript, or the decision to publish the results.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: JS, JC, and HJK
Formal analysis: JS and HJK
Funding acquisition: JS
Methodology: JS, JC, and HJK
Writing—original draft: JS
Writing—review and editing: JC and HJK

Conflicts of Interest

None declared.

References

1. Chen LK, Liu LK, Woo J, et al. Sarcopenia in Asia: consensus report of the Asian Working Group for Sarcopenia. J Am Med Dir Assoc 2014 Feb;15(2):95-101. [doi: 10.1016/j.jamda.2013.11.025] [Medline: 24461239]
2. Liang C, Shi L, Li B, He Z. The mediating role of sarcopenia in the association between physical activity and falls among Chinese older adults: a cross-sectional study. Healthcare (Basel) 2023 Dec 12;11(24):3146. [doi: 10.3390/healthcare11243146] [Medline: 38132036]
3. Chou MY, Nishita Y, Nakagawa T, et al. Role of gait speed and grip strength in predicting 10-year cognitive decline among community-dwelling older people. BMC Geriatr 2019 Jul 5;19(1):186. [doi: 10.1186/s12877-019-1199-7] [Medline: 31277579]
4. Wang DXM, Yao J, Zirek Y, Reijnierse EM, Maier AB. Muscle mass, strength, and physical performance predicting activities of daily living: a meta-analysis. J Cachexia Sarcopenia Muscle 2020 Feb;11(1):3-25. [doi: 10.1002/jcsm.12502] [Medline: 31788969]
5. Oh MH, Shin HE, Kim KS, Won CW, Kim M. Combinations of Sarcopenia Diagnostic Criteria by Asian Working Group of Sarcopenia (AWGS) 2019 guideline and incident adverse health outcomes in community-dwelling older adults. J Am Med Dir Assoc 2023 Aug;24(8):1185-1192. [doi: 10.1016/j.jamda.2023.04.005] [Medline: 37209713]
6. Dent E, Morley JE, Cruz-Jentoft AJ, et al. Physical frailty: ICFSR International Clinical Practice Guidelines for Identification and Management. J Nutr Health Aging 2019;23(9):771-787. [doi: 10.1007/s12603-019-1273-z] [Medline: 31641726]
7. Dent E, Morley JE, Cruz-Jentoft AJ, Arai H, Kritchevsky SB, Guralnik J, et al. International Clinical Practice Guidelines for Sarcopenia (ICFSR): screening, diagnosis and management. J Nutr Health Aging 2018;22(10):1148-1161. [doi: 10.1007/s12603-018-1139-9] [Medline: 30498820]
8. Voulgaridou G, Tyrovolas S, Detopoulou P, et al. Diagnostic criteria and measurement techniques of sarcopenia: a critical evaluation of the up-to-date evidence. Nutrients 2024 Feb 1;16(3):436. [doi: 10.3390/nu16030436] [Medline: 38337720]
9. Bhasin S, Travison TG, Manini TM, et al. Sarcopenia definition: the position statements of the Sarcopenia Definition and Outcomes Consortium. J Am Geriatr Soc 2020 Jul;68(7):1410-1418. [doi: 10.1111/jgs.16372] [Medline: 32150289]
10. Chen LK, Woo J, Assantachai P, et al. Asian Working Group for Sarcopenia: 2019 consensus update on sarcopenia diagnosis and treatment. J Am Med Dir Assoc 2020 Mar;21(3):300-307. [doi: 10.1016/j.jamda.2019.12.012] [Medline: 32033882]
11. Mijnarends DM, Meijers JMM, Halfens RJG, et al. Validity and reliability of tools to measure muscle mass, strength, and physical performance in community-dwelling older people: a systematic review. J Am Med Dir Assoc 2013 Mar;14(3):170-178. [doi: 10.1016/j.jamda.2012.10.009] [Medline: 23276432]
12. Abellan van Kan G, Rolland Y, Andrieu S, et al. Gait speed at usual pace as a predictor of adverse outcomes in community-dwelling older people an International Academy on Nutrition and Aging (IANA) Task Force. J Nutr Health Aging 2009 Dec;13(10):881-889. [doi: 10.1007/s12603-009-0246-z] [Medline: 19924348]
13. Lord S, Howe T, Greenland J, Simpson L, Rochester L. Gait variability in older adults: a structured review of testing protocol and clinimetric properties. Gait Posture 2011 Oct;34(4):443-450. [doi: 10.1016/j.gaitpost.2011.07.010] [Medline: 21920755]
14. Kobsar D, Charlton JM, Tse CTF, et al. Validity and reliability of wearable inertial sensors in healthy adult walking: a systematic review and meta-analysis. J Neuroeng Rehabil 2020 May 11;17(1):62. [doi: 10.1186/s12984-020-00685-3] [Medline: 32393301]

XSL•FO
RenderX

15. Contreras C, Stanley EC, Deschamps-Prescott C, Burnap S, Hopkins M, Browning B, et al. Evaluation of smartphone technology on spatiotemporal gait in older and diseased adult populations. Sensors (Basel) 2024 Sep 9;24(17):5839. [doi: 10.3390/s24175839] [Medline: 39275750]
16. Ho MH, Peng CY, Liao Y, Yen HY. Efficacy of a wearable activity tracker with step-by-step goal-setting on older adults' physical activity and sarcopenia indicators: clustered trial. J Med Internet Res 2024 Nov 1;26:e60183. [doi: 10.2196/60183] [Medline: 39486024]
17. Kang MG, Kang SJ, Roh HK, et al. Accuracy and diversity of wearable device-based gait speed measurement among older men: observational study. J Med Internet Res 2021 Oct 11;23(10):e29884. [doi: 10.2196/29884] [Medline: 34633293]
18. Biswas N, Chakrabarti S, Jones LD, Ashili S. Smart wearables addressing gait disorders: a review. Mater Today Commun 2023 Jun;35:106250. [doi: 10.1016/j.mtcomm.2023.106250]
19. Seifer AK, Kuderle A, Dorschky E, Moradi H, Hannemann R, Eskofier BM. Step length and gait speed estimation using a hearing aid integrated accelerometer: a comparison of different algorithms. IEEE J Biomed Health Inform 2024 Nov;28(11):6619-6628. [doi: 10.1109/JBHI.2024.3454824] [Medline: 39236137]
20. Seifer AK, Dorschky E, Küderle A, Moradi H, Hannemann R, Eskofier BM. EarGait: estimation of temporal gait parameters from hearing aid integrated inertial sensors. Sensors (Basel) 2023 Jul 20;23(14):6565. [doi: 10.3390/s23146565] [Medline: 37514858]
21. Shin J, Park E. Comparison between discrete multi-wavelength near-infrared spectroscopy and bioelectrical impedance analysis in the assessment of muscle mass for community-dwelling older people. J Clin Med 2024 Apr 18;13(8):2350. [doi: 10.3390/jcm13082350] [Medline: 38673621]
22. Hansen SS, Munk T, Knudsen AW, Beck AM. Concordance between changes in calf circumference and muscle mass exists: a narrative literature review. Clin Nutr ESPEN 2024 Feb;59:171-175. [doi: 10.1016/j.clnesp.2023.11.026] [Medline: 38220372]
23. do Amaral C, da Luz Goulart C, da Silva BM, et al. Low handgrip strength is associated with worse functional outcomes in long COVID. Sci Rep 2024 Jan 24;14(1):2049. [doi: 10.1038/s41598-024-52401-z] [Medline: 38267519]
24. Ryu JY, Kim M, Kim KS, Kim S, Won CW. Chair stand test as a proxy for physical performance and muscle strength in sarcopenia diagnosis: the Korean Frailty and Aging Cohort Study. Aging Clin Exp Res 2022 Oct;34(10):2449-2456. [doi: 10.1007/s40520-022-02172-2] [Medline: 35918606]
25. Guralnik JM, Simonsick EM, Ferrucci L, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol 1994 Mar;49(2):M85-M94. [doi: 10.1093/geronj/49.2.m85] [Medline: 8126356]
26. Kim M, Won CW. Sarcopenia in Korean community-dwelling adults aged 70 years and older: application of screening and diagnostic tools from the Asian Working Group for Sarcopenia 2019 Update. J Am Med Dir Assoc 2020 Jun;21(6):752-758. [doi: 10.1016/j.jamda.2020.03.018] [Medline: 32386844]
27. Kim S, Kim M, Won CW. Validation of the Korean Version of the SARC-F questionnaire to assess sarcopenia: Korean Frailty and Aging Cohort Study. J Am Med Dir Assoc 2018 Jan;19(1):40-45. [doi: 10.1016/j.jamda.2017.07.006] [Medline: 28843526]
28. Jung CK, Kim J, Rhim HC. Validation of an ear-worn wearable gait analysis device. Sensors (Basel) 2023 Jan 21;23(3):1244. [doi: 10.3390/s23031244] [Medline: 36772282]
29. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 2004 Jan 10;363(9403):157-163. [doi: 10.1016/S0140-6736(03)15268-3] [Medline: 14726171]
30. Callisaya ML, Blizzard L, Schmidt MD, McGinley JL, Lord SR, Srikanth VK. A population-based study of sensorimotor factors affecting gait in older people. Age Ageing 2009 May;38(3):290-295. [doi: 10.1093/ageing/afp017] [Medline: 19264860]
31. Rosano C, Aizenstein H, Brach J, Longenberger A, Studenski S, Newman AB. Special article: gait measures indicate underlying focal gray matter atrophy in the brain of older adults. J Gerontol A Biol Sci Med Sci 2008 Dec;63(12):1380-1388. [doi: 10.1093/gerona/63.12.1380] [Medline: 19126852]
32. Genton L, Karsegard VL, Kyle UG, Hans DB, Michel JP, Pichard C. Comparison of four bioelectrical impedance analysis formulas in healthy elderly subjects. Gerontology 2001;47(6):315-323. [doi: 10.1159/000052821] [Medline: 11721145]
33. Porto JM, Nakaishi APM, Cangussu-Oliveira LM, Freire Júnior RC, Spilla SB, Abreu DCC. Relationship between grip strength and global muscle strength in community-dwelling older people. Arch Gerontol Geriatr 2019;82:273-278. [doi: 10.1016/j.archger.2019.03.005] [Medline: 30889410]
34. Samuel D, Rowe P. An investigation of the association between grip strength and hip and knee joint moments in older adults. Arch Gerontol Geriatr 2012;54(2):357-360. [doi: 10.1016/j.archger.2011.03.009] [Medline: 21481952]
35. Rosano C, Brach J, Longstreth WT Jr, Newman AB. Quantitative measures of gait characteristics indicate prevalence of underlying subclinical structural brain abnormalities in high-functioning older adults. Neuroepidemiology 2006;26(1):52-60. [doi: 10.1159/000089240] [Medline: 16254454]
36. Wang Z, Ren K, Li D, et al. Assessment of brain function in patients with cognitive impairment based on fNIRS and gait analysis. Front Aging Neurosci 2022 May 24;14:799732. [doi: 10.3389/fnagi.2022.799732] [Medline: 35686022]

XSL•FO
RenderX

37. Ding H, Ho K, Searls E, et al. Assessment of wearable device adherence for monitoring physical activity in older adults: pilot cohort study. JMIR Aging 2024 Oct 25;7:e60209. [doi: 10.2196/60209] [Medline: 39454101]
38. Prasanth H, Caban M, Keller U, et al. Wearable sensor-based real-time gait detection: a systematic review. Sensors (Basel) 2021 Apr 13;21(8):2727. [doi: 10.3390/s21082727] [Medline: 33924403]

Abbreviations

AWGS: Asian Working Group for Sarcopenia
BIA: bioelectrical impedance analysis
IVLR: instantaneous vertical loading rate
SARC-F: strength, assistance with walking, rise from a chair, climb stairs and fall frequency
SMI: appendicular skeletal muscle mass index
SPPB: short physical performance battery

Edited by A Mavragani; submitted 03.07.24; peer-reviewed by G Welk, S Okita; revised version received 20.11.24; accepted 18.12.24; published 10.04.25.

Please cite as:
Shin J, Kweon HJ, Choi J
Assessment of Gait Parameters Using Wearable Sensors and Their Association With Muscle Mass, Strength, and Physical Performance in Korean Older Adults: Cross-Sectional Study
JMIR Form Res 2025;9:e63928
URL: https://formative.jmir.org/2025/1/e63928
doi:10.2196/63928

©Jinyoung Shin, Hyuk Jung Kweon, Jaekyung Choi. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Centering Youth Voice in the Adaptation of an mHealth Intervention for Young Adults With HIV in South Texas, United States: Human-Centered Design Approach

Nhat Minh Ho[1], BS; Catherine Johnson[2], MSc; Autumn Chidester[3], MD, MPH; Ruby Viera Corral[1], BS; Jacundo Ramos[4], BS; Miguel Garcia[4], BFA; Rishi Gonuguntla[5], MD, MPH; Cyrena Cote[4], BS; Divya Chandramohan[4], MD; Hueylie Lin[1], MD, MPH; Anna Taranova[2], MD, MS; Ank E Nijhawan[6], MD, MPH, MSCS; Susan Kools[7], PhD; Karen Ingersoll[8], PhD; Rebecca Dillingham[8], MD, MPH; Barbara S Taylor[4], MD, MS

[1]Joe R. and Teresa Lozano Long School of Medicine, The University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

[2]Department of Public Health, Innovation and Equity, University Health, San Antonio, TX, United States

[3]Department of Women's Health, Dell Medical School, The University of Texas at Austin, Austin, TX, United States

[4]Division of Infectious Diseases, Department of Medicine, The University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Dr., MSC 7881, San Antonio, TX, United States

[5]Department of Psychiatry, Washington University School of Medicine in St. Louis, St Louis, MO, United States

[6]Division of Infectious Diseases, Departament of Medicine, The University of Texas Southwestern Medical Center, Dallas, TX, United States

[7]School of Nursing, University of Virginia, Charlottesville, VA, United States

[8]School of Medicine, University of Virginia, Charlottesville, VA, United States

Corresponding Author:
Barbara S Taylor, MD, MS
Division of Infectious Diseases, Department of Medicine, The University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Dr., MSC 7881, San Antonio, TX, United States

Abstract

Background: Young adults living with HIV are less likely to engage in care and achieve viral suppression, compared to other age groups. Young adults living with HIV also have a high degree of self-efficacy and willingness to adopt novel care modalities, including mobile health (mHealth) interventions. Interventions to increase care engagement could aid young adults living with HIV in overcoming structural and social barriers and leveraging youth assets to improve their health outcomes.

Objective: The objective of the paper was to use an assets-based framework, positive youth development, and human-centered design principles to adapt an existing mHealth intervention, PositiveLinks (PL), to support care engagement for 18 - to 29-year-olds with HIV.

Methods: We conducted a formative evaluation including semistructured interviews with 14 young adults with HIV and focus groups with 26 stakeholders (providers, nurses, case managers, and clinic staff). Interviews covered barriers to care, provider communication, and concerns or suggestions about mHealth interventions. The research team used thematic analysis to review interview transcripts. In the second phase, human-centered design processes informed adaptation of the existing PL platform using data from real-time use suggestions of 3 young adults with HIV. Throughout the formative evaluation and adaptation, a Youth Advisory Board (YAB) provided input.

Results: Young adults with HIV and stakeholders identified common elements of an mHealth intervention that would support care engagement including: the convenience of addressing needs through the app, online support groups to support interconnection, short videos or live chats with other young adults with HIV or providers, appointment and medication reminders, and medical information from a trustworthy source. Stakeholders also mentioned the need for youth empowerment. Concerns included worries about confidentiality, unintentional disclosures of status, urgent content in an unmoderated forum, and the impersonality of online platforms. Design suggestions from young adults with HIV included suggestions on appearance, new formatting for usability of the online support group, and prioritization of local content. Based on the feedback received, iterative changes were made to transform PL into Positive Links for Youth (PL4Y). Final votes on adaptations were made by the YAB. The overall appearance of the platform was changed, including logo, color, and font. The online support group was divided into 3 channels which support hashtags and content searches. The "Resources" and "Frequently Asked Questions" sections were condensed and revised to prioritize South Texas–specific content.

Conclusions: Our assets-based framework supported young adults with HIV and stakeholder input in the transformation of an mHealth intervention to meet the needs of 18- to 29-year-olds in South Texas. The human-centered design approach allowed

XSL•FO
RenderX

young adults with HIV to suggest specific changes to the intervention's design to support usability and acceptability. This adapted version, PL4Y, is now ready for pilot testing in the final phase of this implementation science project.

(JMIR Form Res 2025;9:e60531) doi:10.2196/60531

KEYWORDS

HIV; implementation science; youth; mHealth; adherence; young; mobile health; mobile health intervention; AIDS; US; adult; self-efficacy; willingness; health outcomes; mHealth intervention; interview; human-centered design; acceptability; usability; mobile phone

Introduction

Background

Young adults living with HIV face difficulty with care engagement [1]. While existing studies of young adults with HIV include varying age ranges, spanning 13 to 29 years, most agree that young adults with HIV across the age spectrum face unique challenges in care engagement. These include limited opportunities for care, higher rates of missed appointments, and medication nonadherence, all of which contribute to difficulties in achieving viral suppression [2-4]. Interventions to increase care engagement could help young adults with HIV overcome structural and social barriers to improve health outcomes. While young adults with HIV often seek mobile health technologies to support care engagement, few were developed or adapted with their input or needs in mind.

Most young adults with HIV have a high degree of self-efficacy and willingness to adopt novel care modalities, particularly technological innovations [5]. In 2018, a national survey revealed 45% of adolescents report being on the web almost constantly, and 95% of teens own a smartphone [6]. Mobile health (mHealth) interventions are well suited to these digital natives. Data show that mobile phone apps are an acceptable and often preferred way for adolescents to receive sexual health information and successfully engage those susceptible to HIV [7-10]. Considering the developmentally-appropriate desire for self-determination found among young adults with HIV, they may appreciate mHealth interventions offering an expanded range of services, particularly functions that support self-management of HIV and digital connections [11,12]. However, the majority of current mHealth interventions offer a limited number of functions, such as medication or appointment reminders [7,8,10,13].

mHealth Interventions and Care Engagement for Young Adults With HIV

A few mHealth interventions focusing on young adults with HIV have shown promising results [9,10,13]. A mobile intervention that included peer navigation and a webcomic series increased entry to care, decreased rates of loss to follow up, and supported sustained viral load suppression for young adults with HIV [9]. A review of mHealth intervention modalities for 18- to 34-year-old people living with HIV found advantages such as improved accessibility to information, enhanced self-efficacy, and reduced negative feelings about HIV diagnoses [13]. Other forms of mHealth interventions previously used in promoting the health of young adults with HIV include remote coaching, short message services, phone calls, and social media outreach [14,15]. mHealth intervention implementation faces the challenge of needing continuous updates based on input from stakeholders and app users [10].

Few studies in the existing literature describe how mHealth interventions can be tailored to the needs of young adults with HIV or adaptation processes that ensure they play an active role in intervention design. The Positive Youth Development (PYD) framework suggests that youth involvement and empowerment can lead to better health and reduced risk-taking behaviors [1,16,17]. Aligned with PYD, peer empowerment, knowledge seeking, and taking responsibility for health outcomes improve engagement in health-promoting processes and resilience among gay and bisexual youth with HIV [18]. These data suggest that the PYD framework may be used to incorporate youth input into intervention design and, ideally, lead to more impactful interventions.

Young adults with HIV in South Texas experience significant barriers to care engagement for HIV, making the region an important location to study innovative interventions [19]. In 2016, the US Centers for Disease Control and Prevention (CDC) began using phylogenetic analysis to identify geographic areas where HIV transmission and new diagnoses were more likely. The first iteration of this analysis identified several clusters of new diagnoses in Texas, the largest of which was in South Texas, where 78% of the individuals newly diagnosed were under 30 years old. The majority were also uninsured, without access to HIV prevention services, and Hispanic or Latino [20,21]. Other literature suggests that young men who have sex with men of color experience structural racism, which may alienate them from health care services [4,22]. Despite these challenges, South Texas also has a strong history of collaborative work between community clinics, the county health system, and the local health authority to end the HIV epidemic [23,24]. These factors make the region an ideal site to create and implement a new mHealth intervention targeted specifically toward young adults with HIV.

Investigation Goals and Significance

The aim of this investigation is to collaborate with young adults with HIV and create a space for their voices and input to adapt an existing mHealth intervention, PositiveLinks (PL), to improve care engagement in 18- to 29-year-old people living with HIV in South Texas [25]. This age range was selected as they experience higher rates of new diagnoses and a lower prevalence of sustained viral suppression [4]. PL is a clinic-based multimodal HIV care intervention that increases care engagement and viral suppression [25]. The platform includes daily medication adherence tracking, mood and stress

XSL•FO
RenderX

self-assessment, appointment reminders, secure messaging with clinic providers, and an anonymous online peer support network [25]. We used human-centered design thinking processes in this adaptation to ensure young adults with HIV input, as interventions are more effective when impacted communities are involved in their design and implementation [5].

We conducted a formative evaluation using data from semistructured interviews with young adults with HIV and stakeholders, as well as focus groups with stakeholders to understand acceptable and desired intervention characteristics [19]. Subsequently, we used design thinking processes to aid in our adaptation of PL [16]. This paper describes findings from each phase to enhance understanding of the components of mHealth interventions that are accepted and sought after by young adults with HIV.

Methods

Overview

Adaptation of the existing mHealth intervention, PL, into its novel iteration, Positive Links for Youth (PL4Y), was informed by three data sources: (1) components of the formative evaluation that addressed acceptable mHealth interventions, (2) a human-centered design thinking process, and (3) A Youth Advisory Board (YAB) of young adults with HIV that provided input throughout the formative evaluation and adaptation phases (Figure 1). There was no overlap in membership between the young adults with HIV participating in interviews and the YAB during the formative evaluation. A total of 3 young adults with HIV interviewed during the formative evaluation joined the YAB for subsequent phases of the project, the human-centered design thinking process, and the pilot intervention with the adapted PL4Y app, currently underway. The project used the Consolidated Framework for Implementation Research (CFIR), gathering knowledge of individuals involved and the inner and outer settings surrounding the intervention from a formative evaluation, and then iteratively adapting the intervention to fit the context [26]. A PYD framework also informed study activities, which incorporated input of young adults with HIV in the form of interviews, active feedback during the human-centered design thinking process, and the YAB [11,12]. All 3 components were overseen by the design working group, which included the principal investigator (PI), study coordinator, research assistants, and the scientists, clinic staff, and software designers on the original PL team.

XSL•FO
RenderX

Figure 1. Triangulation of data sources for adaptation of the mobile health (mHealth) intervention to the needs of young adults with HIV in South Texas. PL: PositiveLinks; YAB: Youth Advisory Board; FAQ: frequently asked questions.

XSL•FO
RenderX

Recruitment

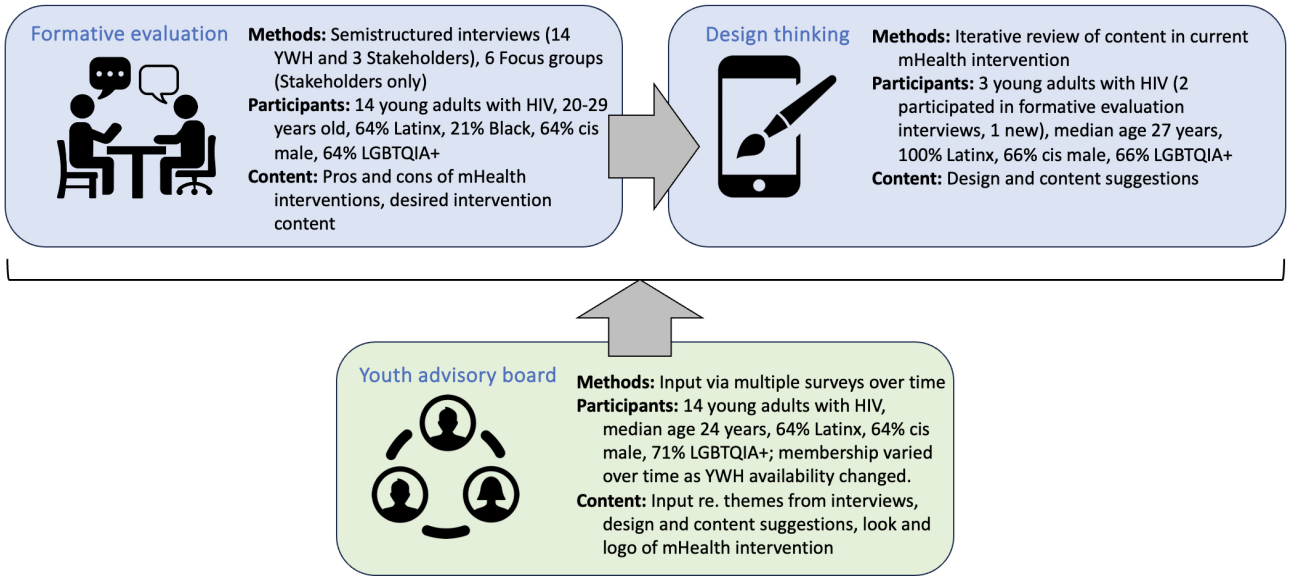
Members of the YAB were recruited through meetings with community-based organizations, flyer distribution in HIV service organizations and clinics, and collaboration with advocates for young adults with HIV. For the formative evaluation phase and Think-Aloud interviews, young adults with HIV were recruited from 6 HIV treatment centers in South Texas using flyers and direct outreach to clinic staff and providers. A total of 14 participants were enrolled in the formative evaluation phase. Of the 3 participants in the Think-Aloud session, 2 had previously participated in the formative evaluation phase. Inclusion criteria required members to be 18- to 29-year-old and living with HIV in San Antonio, Texas. Stakeholders from 6 regional HIV treatment centers were also recruited for the formative evaluation phase.

Youth Advisory Board Role and Input

PL4Y began with the creation of a YAB consisting of 18 - to 29-year-old people living with HIV in South Texas [1,19]. As described previously by Chidester et al [1], the YAB's structure and format evolved in response to the COVID-19 pandemic and feedback from YAB members. YAB engagements occurred quarterly, with input collected through Qualtrics surveys to preserve YAB member anonymity.

Topics covered in the surveys included desired elements of an mHealth intervention, topics for the community message board, preferred mobile app security options (password, fingerprint, face ID, etc), and feedback on the design and look of the app. In the final stage of design iteration, responses regarding the graphics and the app interface from previous YAB surveys and the Think-Aloud sessions were reviewed by the design working group. The software developers incorporated this feedback into the design of a new look and logo for PL4Y. A Qualtrics survey was distributed to the YAB, asking respondents to rate and rank potential logos and designs (N=11). The top 3 logos and fonts (determined by average ranking and average rating) were implemented into a second Qualtrics survey (N=9). The highest average ranked combination of logo and font was selected as the official logo of PL4Y (Figure 2).

Figure 2. Overview of 4 human-centered design thinking driven adaptations to the existing intervention to respond to expressed needs of young adults with HIV during Think-Aloud sessions and the Youth Advisory Board (YAB). mHealth: mobile health; LGBTQIA+: lesbian, gay, bisexual, transgender, queer, intersex, asexual, and other minoritized gender and sexual identities.



Formative Evaluation of Acceptable mHealth Intervention Components

The formative evaluation of care engagement for young adults with HIV in South Texas was conducted from October 2020 to December 2021. Eligibility criteria included being 18 - 29 years of age, residing in South Texas, fluency in either English or Spanish, and reporting current or past challenges with care engagement. Stakeholders included 26 office, nursing, and provider staff members from the regional HIV treatment centers, who provided input either through 6 focus groups or 3 individual interviews.

Interview and focus group guides were developed based on the theory of planned behavior (TPB) theoretical model, the PYD framework, and additional themes provided by the YAB. young adults with HIV were asked about first-hand experiences that affect their care engagement as young people residing in South Texas. Stakeholders were asked to describe their experiences and beliefs regarding serving young adults with HIV and to provide system-level insights on strategies for improving care engagement. To reduce bias and support generative reflection on desired components, neither young adults with HIV or stakeholders were provided with access to the original version of the PL app. A transcription service was used to transcribe individual interviews and focus groups. Transcripts were reviewed and edited by the interviewer (CJ) for accuracy. Transcripts were uploaded into NVivo 1.6.2 (Lumivero), which allows for tracking of responses by stakeholder or young adults with HIV characteristics and by clinic site.

The qualitative analysis team (QAT) created a master codebook of themes derived from TPB and PYD theoretical frameworks for thematic analysis [1,16,17]. The QAT included 4 research assistants with training in qualitative analysis (CJ, AC, HL, and

XSL•FO
RenderX

NMH), 5 faculty-level researchers with expertise in HIV care engagement (AT, AN, KI, RD, and BT), and 1 youth development expert (SK). A member of the primary coding team independently coded transcripts inductively (CJ, AC, and HL), with review by the QAT lead (BT). Secondary coders (CJ, AC, HL, and NMH) expanded on coding as needed. At this stage, the QAT also examined themes by participant type and noted differences between young adults with HIV and stakeholder perspectives. The QAT resolved differences in coding by consensus and confirmed revisions to codebooks. The full QAT participated in the final round of coding, where all themes were reviewed, representative quotes discussed, and themes were renamed, merged, or discarded. Overall findings from the formative evaluation regarding care engagement for young adults with HIV in this context were the topic of a separate analysis [19]. This analysis reviewed themes on advantages and disadvantages of mHealth interventions in general and themes regarding specific mHealth intervention components desired by young adults with HIV or stakeholders.

Human-Centered Design Adaptation of Positive Links by Young Adults with HIV

A total of 3 Think-Aloud interviews were conducted with young adults with HIV using human-centered design principles to ensure end-user input informed an iterative adaptation process [17,18]. A total of 2 Think-Aloud sessions were conducted in person at a local clinic, and 1 was conducted online through Zoom. In addition, 3 researchers (CJ, AC, and NMH) were present for interviews and a modified system usability scale guided conversation [27]. The Think-Aloud participants reviewed each aspect and function of a previous version of the app by walking through it on their phones in real time, providing suggestions for possible design changes and usability improvement. Adaptation recommendations were categorized by ease of implementation and frequency. The most common and feasible suggestions were conveyed to the University of Virginia app development team for final design and development.

Ethical Considerations

All young adults with HIV and stakeholders participating in the formative evaluation and human-centered design components reviewed an institutional review board (IRB)–approved information sheet about the study and provided verbal consent to participate. Participants in the focus groups and Think-Aloud sessions received a US $25 electronic gift card. Stakeholders did not receive monetary incentives for participation. YAB members also received a US $25 electronic gift card upon completion of each survey. This protocol was approved by the UT Health San Antonio and University Health IRBs (HSC20220752HU)

Results

Acceptability and Concerns Regarding mHealth Interventions From the Formative Evaluation

Qualitative data from young adults with HIV and stakeholders revealed common themes about benefits and challenges of an mHealth care engagement intervention. Overall, young adults with HIV and stakeholders felt ease of access to resources, labs, and clinic personnel would be beneficial, and that the anonymous online support group could help young adults with HIV feel less isolated. However, concerns were raised about confidentiality on an mHealth platform and potential pitfalls of the online support group, including the need for moderation and issues with web-based connections (Textbox 1).

Textbox 1. Summary of perceived benefits, challenges, and critiques of an mHealth care engagement intervention expressed by 18- to 29-year-olds living with HIV and stakeholders.

Benefits:

- Increased accessibility of care team.
- Increased access to accurate information about health issues and available resources.
- App-based medication and appointment reminders.
- Ability to track treatment progress through improvements in lab parameters.
- Improved access to a safe community through a online support group.
- Anonymity provided by online support group platform.

Challenges:

- Potential confidentiality concerns.
- Lack of streamlined features.
- Difficulty connecting to other members in the online setting.
- Need for moderation of online support group with potential for cyberbullying.
- Live chat could trigger negative or apprehensive feelings.

XSL•FO
RenderX

Perspectives on Benefits of the mHealth Intervention

Accessibility to Care and Knowledge

Youth and stakeholders believed the app would be highly effective in connecting patients with providers and case managers. Young adults with HIV noted that having an app on their smartphone could increase accessibility to their care team, making them more likely to participate in their care. They also emphasized the app could help them understand what programs they were eligible for and how to apply for support without visiting a clinical setting.

> *You'll be more likely to actually participate in that if it's, you know, easy. It's, like, virtual; so that's a big thing.* [Youth Voice YV 8]

> *If you have a phone or some type of smart device, you would be able to access it wherever you are.* [YV5]

Stakeholders further noted the online nature of the app provides a unique benefit, particularly in the COVID-19 era, by enabling patients to connect with their support network without meeting in person. They believed this benefit extended beyond the pandemic, as it also improves access for patients in rural settings.

> *We don't know how many people are at home alone.* [Stakeholder Voice SV 9]

Furthermore, young adults with HIV and stakeholders felt the app could centralize trustworthy information and serve as an assessable resource.

> *You may not want to ask the doctor, but just, like, you want a quick question answered.* [YV6]

Advantages of an Online Support Group

Youth recognized the benefits of an online support group, describing it as a safe community where people with similar life experiences could connect. They noted that such a group could reduce stigma by fostering a sense of camaraderie.

> *I think with a support group, it's, like, knowing that you're not alone and that, like, it's not a end-of-the-world thing; and you're not dying. You know, you have other people who are living with it and surviving; and it kind of decreases that stigma because you know that there are millions, you know, that are dealing with the same issue.* [YV6]

Both young adults with HIV and shareholders appreciated the anonymity provided by an app-based support group. Young adults with HIV emphasized the app's online platform offered privacy, as they would not need to visit a physical location to engage with a support group.

> *It's good because it's like I don't have to go and park there to go see somebody.* [YV13]

Stakeholders believed newly diagnosed patients would be able to ask about symptoms without feeling guilt. In addition, they highlighted that the anonymity of the app may enable more people to feel comfortable contributing to the community without fear of being judged.

> *Like if they want us to stay a little bit more anonymous but have some place where they can talk to some other people that are going through something similar, or to be like, hey, I'm having these weird symptoms, anybody else experienced that or whatever.* [SV25]

Reminders

Young adults with HIV noted app-based medication reminders would have been helpful when they were first diagnosed and had not already established a medication routine.

> *I think it would be good for other people that are barely starting, so they can remember to drink their meds.* [YV3]

Young adults with HIV and stakeholders reported they would welcome medication refill reminders and appointment reminders.

> *Oh, that would definitely be helpful—medication reminders for sure because I will go to my very last, like, tablet; and I will forget, and then it's a hassle.* [YV6]

Stakeholders identified that reminders could provide encouragement by tracking treatment progress and enabling users to share their milestones within a community. This, in turn, would empower young adults with HIV to continue engaging in care.

> *Hey, your viral load was this high and look how it is now. I think it would be helpful, you know, when they become undetectable or they can, you know, share with partners.* [SV18]

Young adults with HIV reported they wished to have "testimonials/videos of people living with HIV," and "facilitated conversations" on the app.

Disadvantages of an mHealth Intervention

Confidentiality and Anonymity

Although the app's presence on a smartphone increases accessibility, young adults with HIV and stakeholders expressed concerns about potential for unintentional disclosure if someone else sees the app on their device. Both groups identified notifications as a particular risk for exposure.

> *Somebody else gets access to your account, and I mean, that's biggest my fear… Maybe getting notifications when I didn't want to get a notification and it kind of exposing me.* [YV7]

Design Critiques

Some young adults with HIV and stakeholders mentioned having too many settings in the app may limit its usability. They wanted a streamlined design with access to educational content. In addition, youth expressed concerns about potential redundancy in notification, where the same message might appear in multiple locations on the smartphone.

> *It's just redundant when I'm using the same device over and over again.* [YV6]

XSL•FO
RenderX

Young adults living with HIV and stakeholders voiced concerns that online interactions could feel impersonal. A participant reported negative experiences with previous web-based communities of people living with HIV.

> *But I just never really, I never really identified—or like connected with any of them. They were … people from (other countries). But no Hispanic people.* [YV11]

Other young adults living with HIV reported that online resources did not feel genuine and preferred in-person interactions.

> *Going to an actual in-person group, having the actual in-person interaction with people for me would have more of an impact than just kind of talking to someone on my phone.* [YV5]

Despite the potential impersonal experience, stakeholders pointed out that people are more accustomed to virtual settings than in the past.

> *But a lot of things just nowadays is really virtual, just like what we're doing here today. You know, a lot of it is just really impersonal.* [SV3]

Concerns for Unmoderated Interactions

Young adults living with HIV were concerned that without moderation, there was the possibility of smaller communities developing within mHealth interventions and online communities, making the platform less approachable for new users. In addition, concerns were voiced about the lack of moderation allowing users to write irrelevant or offensive posts within the forum. A participant described their experience reading comments on a thread for a online discussion group:

> *Every time you scroll down and you think you're done, there's like 20 more in the bottom and they're just saying random stuff.* [YV9]

Young adults living with HIV and stakeholders had concerns regarding the potential for "trolls" or "cyberbullying." They were particularly concerned about those who would be able to download and join the app community following implementation.

> *Maybe if a certain type of group of people get ahold of it… They have easy access to basically come on and troll everybody.* [YV14]

Stakeholders additionally expressed concerns about delayed responses to urgent messages written in the live chats, given staff are not constantly monitoring the service. However, they believe this risk could be mitigated by educating patients that urgent messages should not be conveyed through the app and that messages will be reviewed within a designated time.

> *Patients may just send the chat like saying… I have fever and shortness of breath. And then Friday at 5:30 and nobody's there in the chat feature. So, I think that that's a little bit dangerous.* [SV18]

Young adults with HIV voiced apprehension about how video support groups or live chats would affect them. A user expressed they might experience a, "feeling of insignificance, just because maybe somebody is going through it worse than you are" (YV2). In addition, young adults living with HIV stated concerns about how listening to other people's testimonies may, "make it more real for them" (YV5).

and it is unpleasant if people minimize their experience on the app. They also speculated that users interested in building connections may migrate to other platforms, such as social media, for deeper conversations.

> *But I don't think that they're going to really utilize just this App to message back and forth. Because they'll be like, "Hey, you're cool." "Oh, you're cool, too. Here's my Instagram." And then they're going to use that.* [YV13]

Design Suggestions From Young Adults With HIV Derived From Think-Aloud Sessions

Young adults with HIV wanted to know who would have access to the information they put into the app, such as medication adherence, mood, or stress. App users can report daily medication adherence, stress levels, and mood "check ins" to log their emotions. Think-Aloud participants emphasized the importance of ensuring that no one views their data without consent.

Participants provided specific suggestions to improve the app's aesthetic appeal, requesting a more "modern look." They had strong opinions against the original color scheme of the app and the old logo, which featured two chain links. The original blue color scheme was deemed "too medical" by participants, who suggested either a "plain" or "vibrant" color scheme. Participants were quick to point out incorrect scaling on thumbnails and other visually unappealing color schemes that reduced the visual charm of the original PL app.

During Think-Aloud sessions, participants reported certain design aspects were nonintuitive, hindering the ease of use. Participants suggested reorganizing the community message board from a single thread to allow participants to self-organize threads, optimizing the search bar for resources and FAQs, having convenient "learn more" hyperlinks within lab results, and implementing "read receipts" to confirm whether messages and documents sent were seen by all users, including health care providers.

Participants expressed particular interest in finding local resources relevant to South Texas. They were eager to learn about the resources offered within the app but suggested making them more community specific. Participants also found the FAQ section useful for patients newly diagnosed with HIV and suggested including a glossary of acronyms to help understand frequently used terminology in HIV care. Participants also emphasized the need for customizability and improved integration within their devices. Specifically, they requested ways to customize the app's home screen icon and to modify notification settings to allow for privacy.

Several suggestions made by Think-Aloud participants were determined to be beyond the scope of the current adaptation, but important to consider in the future. These potential adaptations include saving posts, blocking specific users, and

XSL•FO
RenderX

prioritizing conversations within community message boards. Young adults living with HIV also requested the ability to add photos to posts, replies, and messages. Participants additionally suggested new features for access to care, such as online bus passes.

Adaptation Process

The design working group incorporated findings from the formative evaluation and Think-Aloud sessions into the app's redesign. The laboratory section was updated to be more understandable for young adults with HIV. The community message board was divided into 3 distinct channels, allowing members to "like" other posts, and implementing a search by word or hashtag function. The color scheme, font, and logo were updated based on feedback by the YAB, as described above (Figure 2). The "Resources and FAQ" section was condensed and revised to prioritize material relevant to South Texas. This adaptation process is ongoing as we continue to add requested content, such as videos explaining the steps to an HIV clinic visit.

Discussion

Principal Findings

We incorporated the voices of young adults with HIV throughout the process of adapting an mHealth care engagement intervention using an implementation science–informed approach. This process led to significant changes in the intervention. Young adults with HIV identified several advantages of mHealth interventions, including ease of access, community building, and medications and appointment reminders. However, they also voiced concerns about confidentiality, the impersonal nature of online interactions, and lack of content moderation. Based upon feedback from the formative evaluation, Think-Aloud sessions, and the YAB, the intervention was adapted to meet expressed needs. Modifications included changing the graphic design and logo, adding South Texas specific resources, updating visualization of laboratory data, and designing a more user-friendly and youth-appropriate community message board.

The medical literature suggests young adults with and those without HIV accept app-based mHealth interventions, which can be used to improve health behaviors for youth and emerging adults [9,10,13,28]. An app-based intervention was found to be successful in an adolescent population with asthma, as well as in an adolescent population in recovery following hematopoietic stem cell transplant [29,30]. There have been few studies investigating the usability of app-based interventions in patients with HIV. Aladin et al [9] enrolled 113 patients with HIV and found their app-based intervention to be associated with increased communication and engagement with care when combined with peer navigation and a comic series. However, the majority of mHealth interventions are text message–based between patient and clinic, and users lack the ability to form connections with each other [31]. Brooks et al [13] found that app-based interventions were most useful for young adults with HIV if they had medical information and reminders, leading to increased self-efficacy, supported feelings of connection to a community, and reduced stigma around HIV. These results indicate users of app-based interventions may benefit from "discussion board" components that allow users to communicate with one another and develop a community. Schnall et al [32] studied the functions of 15 publicly available app-based interventions targeting people living with HIV and found 4 enabled communication between peers and providers, 6 had medication reminders, 7 had medication logs, and 6 included a search function [32]. This analysis aligns with the existing literature as both users and stakeholders identified all of the above as critical functionalities of an app-based intervention, and these components were included in the final redesign.

Our findings add to this existing knowledge base by highlighting the importance young adults with HIV place on easy, asynchronous access to services. Young adults with HIV welcome the ability to message providers or other members of the care team, quick links to web-based resources for HIV care and social services, and automatic appointment and medication reminders. We discovered many young adults with HIV wished for connections to other young adults with HIV, something not currently featured in mHealth interventions for this demographic. They also appreciated the online, anonymous format of the community message board. Finally, young adults with HIV repeatedly emphasized that the graphic design and usability of the intervention should be influenced by what was most acceptable and appealing to them. These data influenced our own design adaptation of PL into PL4Y but are also relevant to framing other mHealth interventions for young adults with HIV or other youth communities.

Few studies investigate patient thoughts regarding current mHealth interventions or discuss the concerns of young adults with HIV with this intervention modality. A review of 7 mHealth interventions by Conserve et al [33] found cellular network restrictions were a significant barrier to the use of mobile technologies for young adults with HIV. Our findings echo this, as both young adults with HIV and the stakeholders anticipated access to a smartphone could be a barrier to use of our intervention.

We found additional young adults with HIV concerns with mHealth intervention that, to our knowledge, are not reflected in the existing literature. Many young adults with HIV expressed concerns about confidentiality and security of the data on their phone. Others worried that communicating by electronic messaging, rather than in person or by phone call, would feel impersonal. Young adults living with HIV were also concerned with exposure to harmful or disempowering narratives on the community message board. Though our intervention already has robust security and content monitoring protocols, these concerns will need to be addressed during implementation.

Limitations

Despite the meaningful findings of this investigation, there are limitations. Our adaptation process incorporated viewpoints of a limited number of young adults with HIV participants and YAB members. While their perspectives were valuable, we cannot assume they represent the perspectives of all young adults with HIV. Furthermore, our participants reflect the South Texas population, with the majority identifying as Latinx, which may limit the generalizability of our findings. Despite these

XSL•FO
RenderX

limitations, this study fills a crucial gap in the literature by identifying crucial barriers and facilitators to the adoption of an app-based mHealth intervention for young adults with HIV. It also increases our understanding of the understudied South Texas community's potential response to other mHealth interventions [34].

Conclusions

mHealth interventions targeting young adults with HIV can improve care engagement and health outcomes. However, few studies have identified the specific characteristics that promote or hinder patient adoption of app-based interventions [9,10,13,28]. Our adaptation process successfully incorporated input from a majority Hispanic or Latinx group of young adults with HIV in the design of a mHealth intervention. Our process demonstrated that young adults with HIV should play integral roles in the design of interventions for this population. We identified novel concerns not previously reflected in the literature including both barriers and facilitators to app-based mHealth interventions for young adults with HIV. We also describe a rigorous and iterative adaptation process that centers youth voice to ensure that the new intervention meets the needs of the community, successfully transforming PL into PL4Y. Currently, our team is testing the feasibility and acceptability of the adapted PL4Y mHealth intervention in a randomized control trial among young adults with HIV in South Texas. Our findings can also be used in the design of future app-based mHealth interventions to promote their adoption by young adults with HIV.

Acknowledgments

We are deeply grateful to all the HIV treatment centers and youth-focused community-based organizations in South Texas who partnered with us in this effort, and to all study participants and Youth Advisory Board members for their time and commitment to this process. The funding source for this project was a grant from the National Institute of Mental Health (NIH/NIMH R34 MH122332). The sponsor played no role in the development, review, or approval of this manuscript for publication.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to the fact that individual participant stories of their experiences with HIV could be identifiable in the transcripts, but are available from the corresponding author on reasonable request.

Conflicts of Interest

RD and KI provide consulting services for Warm Health Technology, LLC, which distributes the PositiveLinks app outside Virginia. None declared by the other authors.

References

1. Chidester AB, Johnson CJ, Lin H, et al. Nothing about us without us: involving youth living with HIV in a virtual advisory board. J Adolesc Health 2023 Dec;73(6):1158-1161. [doi: 10.1016/j.jadohealth.2023.06.028] [Medline: 37665305]
2. Ryscavage PA, Anderson EJ, Sutton SH, Reddy S, Taiwo B. Clinical outcomes of adolescents and young adults in adult HIV care. J Acquir Immune Defic Syndr 2011;58(2):193-197. [doi: 10.1097/QAI.0B013E31822D7564]
3. Kapogiannis BG, Koenig LJ, Xu J, et al. The HIV continuum of care for adolescents and young adults attending 13 urban US HIV care centers of the NICHD-ATN-CDC-HRSA SMILE collaborative. J Acquir Immune Defic Syndr 2020 May 1;84(1):92-100. [doi: 10.1097/QAI.0000000000002308] [Medline: 32267659]
4. Behavioral and clinical characteristics of persons with diagnosed HIV infection-medical monitoring project, HIV surveillance special report united states. : Centers for Disease Control and Prevention; 2020.
5. Hightow-Weidman LB, Muessig KE, Bauermeister J, Zhang C, LeGrand S. Youth, technology, and HIV: recent advances and future directions. Curr HIV/AIDS Rep 2015 Dec;12(4):500-515. [doi: 10.1007/s11904-015-0280-x] [Medline: 26385582]
6. Anderson M. Teens, social media & technology 2018. Pew Research Center. 2018. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2023-10-05]
7. Cooper V, Clatworthy J, Whetham J, Consortium E. mHealth interventions to support self-management In HIV: a systematic review. Open AIDS J 2017;11:119-132. [doi: 10.2174/1874613601711010119] [Medline: 29290888]
8. Hightow-Weidman LB, Muessig K, Rosenberg E, et al. University of North Carolina/Emory Center for Innovative Technology (iTech) for addressing the HIV epidemic among adolescents and young adults in the United States: protocol and rationale for center development. JMIR Res Protoc 2018 Aug 3;7(8):e10365. [doi: 10.2196/10365] [Medline: 30076126]
9. Aladin B, Thompson M, Addison D, et al. The YGetIt? program: a mobile application, PEEP, and digital comic intervention to improve HIV care outcomes for young adults. Health Promot Pract 2023 Jul;24(4):658-668. [doi: 10.1177/15248399221150789] [Medline: 36924286]
10. Paul ME, Castillo M, Emmanuel P, et al. Scale up mHealth HIV interventions: site and public health perspectives and lessons learned from P3. Mhealth 2021;7:38. [doi: 10.21037/mhealth-20-65] [Medline: 33898607]

XSL•FO
RenderX

11. Benson PL, Scales PC, Sesma Jr A. Positive youth development: theory, research, and applications. In: Handbook of Child Psychology: John Wiley & Sons; 2007. [doi: 10.1002/9780470147658]
12. Tamis-LeMonda LB, Catherine S, editors. Toward a New Vision and Vocabulary about Adolescence: Theoretical, Empirical, and Applied Bases of a "Positive Youth Development" Perspective: Child Psychol 2nd ed Psychology Press; 2005.
13. Brooks RA, Nieto O, Swendeman D, et al. Qualitative evaluation of social media and mobile technology interventions designed to improve HIV health outcomes for youth and young adults living with HIV: a HRSA SPNS initiative. Health Promot Pract 2020 Sep;21(5):693-704. [doi: 10.1177/1524839920938704] [Medline: 32757839]
14. Amico KR, Lindsey JC, Hudgens M, et al. Randomized Controlled Trial of a Remote Coaching mHealth Adherence Intervention in Youth Living with HIV: Springer; 2022, Vol. 26:3897-3913. [doi: 10.1007/s10461-022-03717-2]
15. Goldstein M, Archary M, Adong J, et al. Systematic review of mhealth interventions for adolescent and young adult HIV prevention and the adolescent HIV continuum of care in low to middle income countries. AIDS Behav 2023 May;27(Suppl 1):94-115. [doi: 10.1007/s10461-022-03840-0] [Medline: 36322217]
16. Harte R, Glynn L, Rodríguez-Molinero A, et al. A human-centered design methodology to enhance the usability, human factors, and user experience of connected health systems: A three-phase methodology. JMIR Hum Factors 2017 Mar 16;4(1):e8. [doi: 10.2196/humanfactors.5443] [Medline: 28302594]
17. Liedtka J, King A, Bennett K. Solving Problems with Design Thinking: Ten Stories of What Works: Columbia University Press; 2011.
18. Liedtka J, Ogilvie T, Brozenske R. Designing for Growth: A Design Thinking Tool Kit for Managers: Columbia University Press; 2011.
19. Johnson C, Chidester A, Chandramohan D, et al. A call for youth voice to support engagement in care for 18- to 29-year olds living with HIV in the US South. AIDS Patient Care STDS 2024 May;38(5):238-248. [doi: 10.1089/apc.2024.0006] [Medline: 38662471]
20. Monterosso A, Minnerly S, Goings S, et al. Identifying and investigating a rapidly growing HIV transmission cluster in Texas. 2017 Presented at: Conference on Retroviruses and Opportunistic Infections; Feb 19-22, 2023; Seattle, Washington.
21. Oster AM, France AM, Panneer N, et al. Identifying clusters of recent and rapid HIV transmission through analysis of molecular surveillance data. J Acquir Immune Defic Syndr 2018 Dec 15;79(5):543-550. [doi: 10.1097/QAI.0000000000001856] [Medline: 30222659]
22. Allan-Blitz LT, Mena LA, Mayer KH. The ongoing HIV epidemic in American youth: challenges and opportunities. Mhealth 2021;7:33. [doi: 10.21037/mhealth-20-42] [Medline: 33898602]
23. Taylor B, Sapien H, Courtois F, et al. San antonio's end stigma end HIV alliance/fast-track cities initiative: community-wide collaboration and peer advocate driven strategies to end the HIV epidemic. : Fast Track Cities Conference; 2019 Presented at: .
24. Taylor BS, Sapién H. Determining the direction of human immunodeficiency virus transmission: benefits and potential harms of taking phylogenetic analysis one step further. Clin Infect Dis 2021 Jan 23;72(1):38-40. [doi: 10.1093/cid/ciz1248] [Medline: 31922539]
25. Dillingham R, Ingersoll K, Flickinger TE, et al. PositiveLinks: A mobile health intervention for retention in HIV care and clinical outcomes with 12-month follow-up. AIDS Patient Care STDS 2018 Jun;32(6):241-250. [doi: 10.1089/apc.2017.0303] [Medline: 29851504]
26. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implementation Sci 2009 Dec;4(1):50. [doi: 10.1186/1748-5908-4-50]
27. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
28. Biello KB, Marrow E, Mimiaga MJ, Sullivan P, Hightow-Weidman L, Mayer KH. A mobile-based app (MyChoices) to increase uptake of HIV testing and pre-exposure prophylaxis by young men who have sex with men: protocol for a pilot randomized controlled trial. JMIR Res Protoc 2019 Jan 7;8(1):e10694. [doi: 10.2196/10694] [Medline: 30617042]
29. Rodgers CC, Krance R, Street RL, Hockenberry MJ. Symptom prevalence and physiologic biomarkers among adolescents using a mobile phone intervention following hematopoietic stem cell transplantation. Oncol Nurs Forum 2014 May;41(3):229-236. [doi: 10.1188/14.ONF.229-236] [Medline: 24769589]
30. Burbank AJ, Lewis SD, Hewes M, et al. Mobile-based asthma action plans for adolescents. J Asthma 2015;52(6):583-586. [doi: 10.3109/02770903.2014.995307] [Medline: 25494553]
31. Fedele DA, Cushing CC, Fritz A, Amaro CM, Ortega A. Mobile health interventions for improving health outcomes in youth: a meta-analysis. JAMA Pediatr 2017 May 1;171(5):461-469. [doi: 10.1001/jamapediatrics.2017.0042] [Medline: 28319239]
32. Schnall R, Mosley JP, Iribarren SJ, Bakken S, Carballo-Diéguez A, Brown Iii W. Comparison of a user-centered design, self-management app to existing mHealth apps for persons living with HIV. JMIR Mhealth Uhealth 2015 Sep 18;3(3):e91. [doi: 10.2196/mhealth.4882] [Medline: 26385783]

XSL•FO
RenderX

33. Conserve DF, Jennings L, Aguiar C, Shin G, Handler L, Maman S. Systematic review of mobile health behavioural interventions to improve uptake of HIV testing for vulnerable and key populations. J Telemed Telecare 2017 Feb;23(2):347-359. [doi: 10.1177/1357633X16639186] [Medline: 27056905]
34. Ramirez AG, Thompson IM, Vela L. South texas population characteristics. In: Ramirez AG, Thompson IM, Vela L, editors. South Texas Health Status Review: A Health Disparities Roadmap: Springer; 2013:7-9. [doi: 10.1007/978-3-319-00233-0_2]

Abbreviations

CDC: Centers for Disease Control and Prevention
CFIR: Consolidated Framework for Implementation Research
IRB: institutional review board
mHealth: mobile health
PI: principal investigator
PL: PositiveLinks
PL4Y: Positive Links for Youth
PYD: Positive Youth Development
QAT: qualitative analysis team
SV: Stakeholder Voice
TPB: theory of planned behavior
YAB: Youth Advisory Board
YV: Youth Voice

Edited by A Mavragani; submitted 14.05.24; peer-reviewed by D Cain, EDWARD NSUBUGA; revised version received 04.02.25; accepted 05.02.25; published 08.04.25.

Please cite as:
Ho NM, Johnson C, Chidester A, Viera Corral R, Ramos J, Garcia M, Gonuguntla R, Cote C, Chandramohan D, Lin H, Taranova A, Nijhawan AE, Kools S, Ingersoll K, Dillingham R, Taylor BS
Centering Youth Voice in the Adaptation of an mHealth Intervention for Young Adults With HIV in South Texas, United States: Human-Centered Design Approach
JMIR Form Res 2025;9:e60531
URL: https://formative.jmir.org/2025/1/e60531
doi:10.2196/60531

© Nhat Minh Ho, Catherine Johnson, Autumn Chidester, Ruby Viera Corral, Jacundo Ramos, Miguel Garcia, Rishi Gonuguntla, Cyrena Cote, Divya Chandramohan, Hueylie Lin, Anna Taranova, Ank E Nijhawan, Susan Kools, Karen Ingersoll, Rebecca Dillingham, Barbara S Taylor. Originally published in JMIR Formative Research (https://formative.jmir.org), 8.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Chatbot to Support the Mental Health Needs of Pregnant and Postpartum Women (Moment for Parents): Design and Pilot Study

Kelsey McAlister[1], PhD; Lara Baez[1], PhD; Jennifer Huberty[1], PhD; Marianna Kerppola[2], MS, MBA

[1]Fit Minded Inc, 14848 N 46th Place, Phoenix, AZ, United States

[2]Poisera Inc (dba Moment for Parents), Ann Arbor, MI, United States

Corresponding Author:
Kelsey McAlister, PhD
Fit Minded Inc, 14848 N 46th Place, Phoenix, AZ, United States

Abstract

Background: Maternal mental health disorders are prevalent, yet many individuals do not receive adequate support due to stigma, financial constraints, and limited access to care. Digital interventions, particularly chatbots, have the potential to provide scalable, low-cost support, but few are tailored specifically to the needs of perinatal individuals.

Objective: This study aimed to (1) design and develop Moment for Parents, a tailored chatbot for perinatal mental health education and support, and (2) assess usability through engagement, usage patterns, and user experience.

Methods: This study used a human-centered design to develop Moment for Parents, a rules-based chatbot to support pregnant and postpartum individuals. In phase 1, ethnographic interviews (n=43) explored user needs to inform chatbot development. In phase 2, a total of 108 pregnant and postpartum individuals were recruited to participate in a pilot test and had unrestricted access to the chatbot. Engagement was tracked over 8 months to assess usage patterns and re-engagement rates. After 1 month, participants completed a usability, relevance, and satisfaction survey, providing key insights for refining the chatbot.

Results: Key themes that came from the ethnographic interviews in phase 1 included the need for trusted resources, emotional support, and better mental health guidance. These insights informed chatbot content, including mood-based exercises and coping strategies. Re-engagement was high (69/108, 63.9%), meaning users who had stopped interacting for at least 1 week returned to the chatbot at least once. A large proportion (28/69, 40.6%) re-engaged 3 or more times. Overall, 28/30 (93.3%) found the chatbot relevant for them, though some noted repetitive content and limited response options.

Conclusions: The Moment for Parents chatbot successfully engaged pregnant and postpartum individuals with higher-than-typical retention and re-engagement patterns. The findings underscore the importance of flexible, mood-based digital support tailored to perinatal needs. Future research should examine how intermittent chatbot use influences mental health outcomes and refine content delivery to enhance long-term engagement and effectiveness.

(JMIR Form Res 2025;9:e72469) doi:10.2196/72469

KEYWORDS

perinatal support; human-centered design; digital health; maternal health; chatbot; digital tool

Introduction

Maternal mental health disorders are a critical public health concern due to their high and rising prevalence. The World Health Organization (WHO) estimates that 10% of pregnant women and 13% of postpartum women develop a mental health disorder [1]. In the United States, 1 in 5 women experience a mental health disorder during pregnancy or within the first year after birth [2]. While postpartum depression is the most prevalent, affecting approximately 1 in 8 women [3], other maternal mental health disorders affect women, including anxiety, psychosis, obsessive-compulsive disorder, bipolar disorder, posttraumatic stress disorder, and schizophrenia [4]. These challenges are even more pronounced among underrepresented populations, who experience higher rates of maternal mood disorders [5,6]. Although treatable, 75% of women with maternal mental health disorders go untreated [2], which poses serious health concerns for both the mother and the infant. Untreated maternal mental health conditions are linked to increased risks for preterm birth, low birth weight, developmental delays, and impaired maternal-infant bonding, and also raise the mother's risk of substance use, psychosis, and suicide [4,7]. The lack of ongoing education and support of mental health issues during and after pregnancy often leaves these women without timely support or interventions, exacerbating the problem.

There are several barriers to care contributing to untreated maternal mental health disorders, including stigma, lack of access to mental health care professionals, and high-cost challenges [8-11]. These challenges are especially prevalent in

XSL•FO
RenderX

underrepresented populations, such as Black women, who often face additional obstacles such as systemic racism, cultural insensitivity, and mistrust of health care providers [12-14]. Other barriers include reluctance to acknowledge symptoms and a lack of support from partners, family members, and health care professionals [11,15]. Social pressures and stigma often prevent women from discussing their feelings openly, leading to feelings of shame and fear of losing custody of their child [15]. Additionally, many mothers have limited knowledge of maternal mental health problems, hindering their ability to recognize symptoms and seek timely help [10,11]. In some cases, individuals may turn to online sources or personal networks for support [16,17], but these resources can be incomplete or inaccurate. Without adequate support, these barriers perpetuate high rates of untreated maternal mental health disorders, negatively affecting both mother and child health outcomes.

Digital tools are a promising solution to addressing barriers to maternal mental health care. These tools offer several advantages, including scalability, affordability, and accessibility, and provide real-time support without the stigma associated with in-person care. A wealth of digital tools for perinatal mental health have been tested in clinical trials, and a recent meta-analysis found that these tools are modestly effective in reducing symptoms of anxiety and depression [18]. However, challenges have emerged in engaging and retaining users of many digital tools due to the excess burden they place on users (eg, needing to download and learn to use a new mobile app). Chatbots have emerged recently as a way to engage perinatal users with digital interventions using a mode of communication that is ubiquitous and comfortable (eg, texting) [19]. Furthermore, chatbots simulate conversations with users, which may help them stay engaged [20]. Chatbots are promising because they offer cost-effective, real-time engagement [21,22]. Chatbots monitor mental health symptoms longitudinally, with each conversation functioning as an ecological momentary assessment, capturing real-time insights into emotional well-being and enabling timely interventions [23]. Importantly, chatbots promote honest disclosure for emotional well-being [24,25].

The success of these tools depends on adopting a human-centered approach, which ensures that the tool aligns with the specific needs and preferences of its users. A human-centered design (HCD) approach involves feedback from users during development, ensuring that interactions are empathetic, personalized, and accessible [26]. HCD integrates human factors, ergonomics, and usability techniques to enhance product design [26,27]. Additionally, it often involves ethnographic research and strategies that explore how individuals interact with systems, ensuring products are safer, easier to use, and aligned with the intended users' needs and abilities [26,27]. This approach helps create engaging experiences, fostering trust and consistent use, which are critical for maintaining long-term effectiveness in mental health support [26].

A growing body of research has demonstrated the utility of engaging in HCD products when developing tools for perinatal populations. Evidence shows that perinatal women widely use and value health-related apps, particularly for support, education, and self-monitoring. For example, many women find apps helpful for personalized health tracking, connecting with online communities, and addressing mental health concerns [28,29]. Furthermore, studies indicate that apps are most effective when designed with input from users and health care professionals, addressing gaps such as information reliability, security, and clinical validation [30,31]. HCD approaches have shown high feasibility and satisfaction, emphasizing the importance of tailoring interventions to the specific challenges and preferences of this population [32].

Despite the potential of digital tools, there remains a critical gap in solutions that offer both education and meaningful mental health support for pregnant and postpartum individuals. Many chatbots use evidence-based techniques such as cognitive behavioral therapy but are not designed for the unique needs of perinatal populations, often failing to address fluctuating emotions, stigma, and cultural sensitivities [33,34]. Additionally, these tools frequently lack longitudinal support and proactive interventions for perinatal mood and anxiety disorders [35]. Therefore, the purpose of this study was to develop and evaluate a chatbot called Moment for Parents. The goal of Moment for Parents is to address the current need for a digital mental health intervention to support the needs of pregnant and postpartum people. This study aimed to (1) design and develop a chatbot tailored specifically to educate and support pregnant and postpartum people on emotional well-being and mental health and (2) evaluate the usability of Moment for Parents in a pilot test by examining usage patterns, engagement levels, and user experience.

Methods

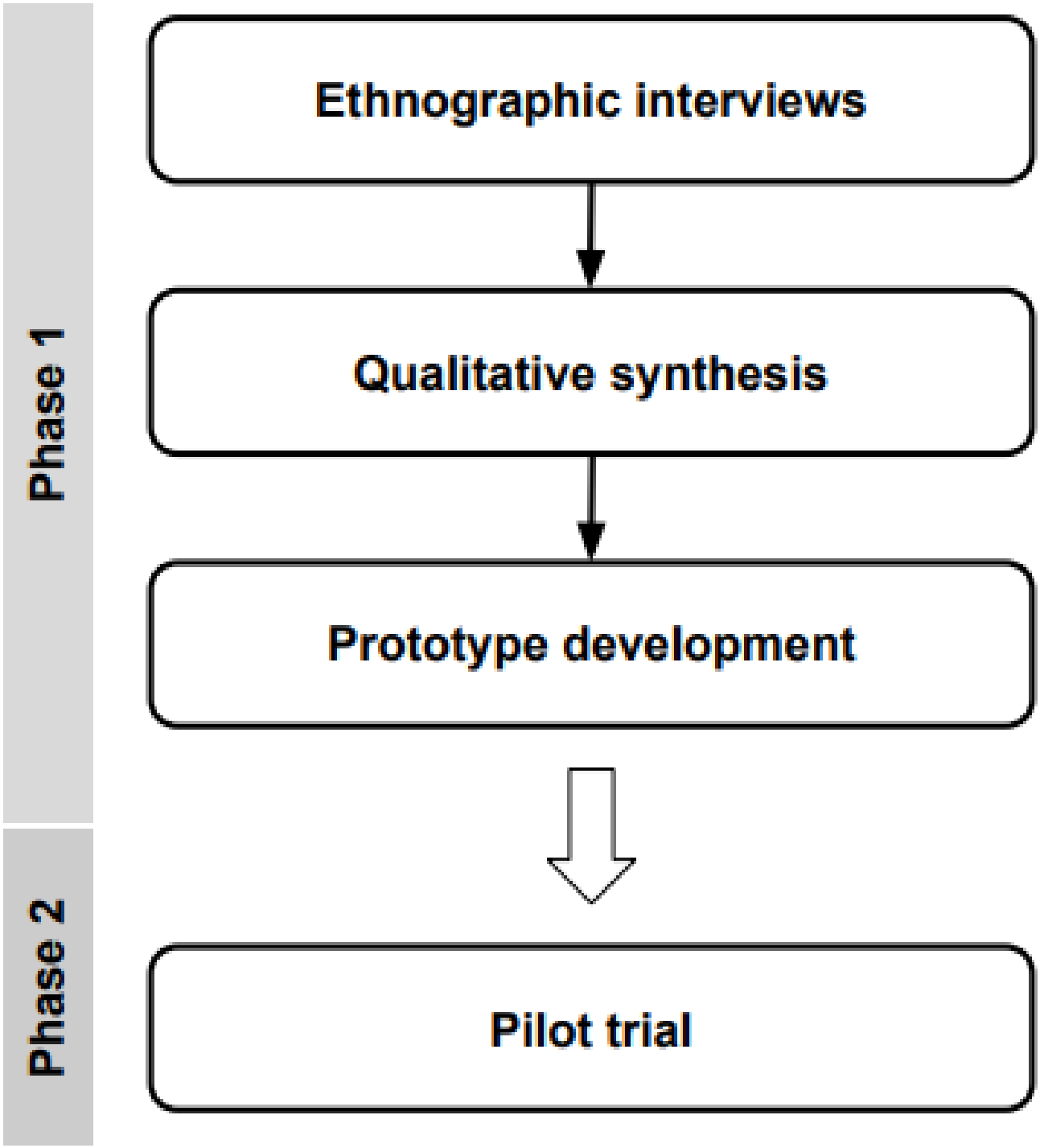
Design and Development

This study used a HCD approach to develop the Moment for Parents chatbot. HCD is an interactive system development approach that aims to create useful products by focusing on users, their needs, and their requirements [26,36]. HCD leads to inclusive health technology that improves mental health equity because the system engages users to improve adherence, adoption, and effectiveness [26,27,37]. By applying these principles, we aimed to develop a more inclusive and supportive chatbot for the mental health and well-being of pregnant and postpartum end users.

Figure 1 depicts the Moment for Parents study design. In phase 1, we engaged in ethnographic interviews with pregnant and postpartum end users to understand their needs and preferences. These findings informed the initial prototype of the Moment for Parents chatbot. In phase 2, the initial prototype was launched for pilot-testing to evaluate its usability, functionality, and relevance to users' needs. Throughout phase 2, the Moment for Parents chatbot experience was iteratively updated as participants provided feedback.

XSL•FO
RenderX

Figure 1. Overview of this study's design.



Ethical Considerations

This study is a retrospective analysis of secondary data, and therefore, it was exempt from additional consent requirements under human subjects regulations and received approval from the BRANY (Biomedical Research Alliance of New York) Institutional Review Board (study ID 24-338-1949). All study data were deidentified. Participants did not receive compensation for participation.

Phase 1 – Ethnographic Interviews

Participants

A total of 43 participants were recruited via snowball sampling, starting with existing networks and communities and expanding through participant referrals. Participants were eligible to participate in this study if they self-reported being pregnant or up to 1 year post partum.

Procedure

Data collection took place between March and November 2019. Ethnographic interviews were conducted via the videoconferencing service Zoom (Zoom Communications, Inc).

XSL•FO
RenderX

The same interviewer conducted all interviews. Interviews were recorded with participant permission, and the interviewer took reflexive notes throughout the interviews.

Measures

Of ethnographic interview, participants were asked to provide their gestational age if pregnant or the age of their youngest child if post partum. Participants were also asked how many additional children they had and their ages. No other demographic information was collected from participants. The ethnographic interviews were designed to explore participants' mental health needs, challenges during pregnancy or post partum, and expectations for digital support tools. Questions addressed a range of topics, including emotional well-being, resources they have found helpful, and desired features for a mental health chatbot. Sample questions included:

> *How has your pregnancy (if pregnant) been going so far? What have been your go-to resources to stay sane during pregnancy? What tools and/or resources were the most helpful?? After you had your baby (if postpartum), what tools or resources were the most helpful postpartum? Is there anything else you wish you had to help prepare for birth/motherhood? As you reflect on your pregnancy and postpartum experience, how did you think about self-care? What were the biggest hurdles/challenges in taking care of yourself?*

In addition to ethnographic interviews, additional input was collected from a multidisciplinary team of experts, including yoga and meditation instructors, midwives, doulas, public health professionals, and therapists specializing in maternal mental health. Insights for the Moment for Parents chatbot were also drawn from relevant maternal mental health literature, literature reviews on chatbots, and the team's lived experience.

Data Analysis

Audio recordings of interviews were transcribed, and reflexive interview notes were reviewed. Data were analyzed using a qualitative description approach [38], wherein the coder identified themes and subthemes relevant to the Moment for Parents prototype development.

Phase 2 – Pilot Test

Participants

A new sample of participants was recruited for the pilot test of the Moment for Parents prototype. Participants were recruited to the pilot test through personal networks and social media groups targeting pregnant and postpartum people who were interested in accessing mental health support. By recruiting these networks and online groups, we sought to reach a diverse range of users to ensure the prototype would be tested among people with different experiences in pregnancy and postpartum care.

Procedure

Data collection took place between November 2019 and June 2020. Upon enrollment, participants were provided with a no-cost, indefinite version of the Moment for Parents prototype, allowing participants to freely engage with the chatbot's features without time constraints. After 1 month of using Moment for Parents, participants were invited to complete a survey that asked about their perceptions of the Web app (ie, prototype feedback survey).

Measures

Chatbot Engagement

Participants' interactions with the Moment for Parents chatbot were tracked to assess usage patterns and calculate engagement metrics.

Prototype Feedback Survey

After 1 month of using Moment for Parents, participants completed a structured survey evaluating their experience with the chatbot. Participants received a text message inviting them to complete the survey, which included a link directing them to the questions in the Web app. Survey items assessed usability, perceived utility, relevance of content, and overall satisfaction. These responses allowed us to quantify participant satisfaction and identify trends in user experience.

Data Analysis

Descriptive statistics were used to summarize participant characteristics, engagement patterns, and feedback survey responses. Means and SDs were calculated for continuous variables, while categorical variables were reported as frequencies and percentages. Engagement over time was analyzed by calculating the proportion of participants who remained active each month. Re-engagement patterns were assessed by identifying users who returned to the chatbot after at least 1 week of inactivity. Stata (version 15; StataCorp LLC) was used for the analyses.

Results

Phase 1 – Ethnographic Interviews

Participants

A total of 43 individual interviews were conducted. Demographic characteristics of participants included 56% (n=24) who were first-time pregnant mothers, 23% (n=10) who were mothers expecting their second or third child, and 21% (n=9) who were post partum with at least 1 child under the age of 1 year.

Qualitative Description Results

A summary of the key themes and subthemes derived from the qualitative description can be found in Table 1. The three main themes identified were (1) resources, (2) mental health journey, and (3) gaps in support. When asked what types of support users were looking for during the perinatal period, they reported seeking guidance from friends, family, peers, mental health professionals, and medical providers. Users also described the emotional trajectory they experienced, highlighting unique challenges present during the various phases of pregnancy and postpartum. Finally, users identified a variety of gaps in support that they felt could be addressed with a chatbot. These included a lack of reliable evidence-based psychoeducation on maternal

XSL•FO
RenderX

mental health, social isolation from peers, and insufficient emotional support from medical professionals.

XSL•FO
RenderX

Table . Qualitative synthesis of findings from ethnographic interviews.

Theme		Description	Quotes
Resources[a]			
	Social support	Friends, family, and online communities eased anxiety through shared experiences and validation.	• "I feel like I leaned a lot on my friends." [P09[b], postpartum, aged 10 mo] • "It's good to just hear opinions of other pregnant people instead of like professional opinions or thoughts." [P22, pregnant, third trimester]
	Mental health tools	Apps and therapists provided emotional support, tracking, and coping strategies.	• "I've been in touch with my therapist since late last year... It's been really helpful just having an actual therapist to talk to and talk my anxieties out." [P41, pregnant, second trimester]
	Clinical guidance	Health care providers and doulas were essential for reassurance, though often limited in time for deeper support.	• "I listen to my doctor above all else. For my first birth, I didn't want a C-Section, and they let me go three days and respected my wishes. I had such a positive experience with my health care team, which I know not everyone has, so I trust my doctor above all else." [P07, pregnant, second trimester, aged 3 y]
Mental health journey[c]			
	Emotional trajectory	Anxiety, isolation, and the unpredictability of pregnancy and parenthood shaped mental health challenges. The pregnancy journey was marked by fluctuating emotions, anxiety, and adjustments in mental health.	• "Mentally a very isolating time anyways – feel like unless you know someone close to you who is pregnant around the same time, even in normal world standards, it can be very isolating." [P16, pregnant, third trimester] • "I'm not as happy as I was in the second trimester, I feel always tired. A little bit more emotionally trying, especially since I can't hang out with my friends." [P02, pregnant, third trimester]
	Early pregnancy stress	Intense physical symptoms and fear of complications caused heightened anxiety in the first trimester.	• "Trying to think that you know I don't want the baby to feel this. I want the baby to know that she's loved and don't want her to feel affected by the situation." [P40, postpartum, aged 15 mo]
	Late pregnancy anxiety	Worries about labor and delivery escalated in the third trimester, along with physical discomfort.	• "You're supposed to be happy about this, it shouldn't be terrible, not a walk in the park, but unknowns surrounding childbirth have been the biggest worry in back of my mind" [P30, pregnant, third trimester]

XSL•FO
RenderX

Theme		Description	Quotes
	Postpartum adjustment	Burnout, hormonal changes, and lack of support created emotional strain during the postpartum period.	• "Honestly, I was burned out. I was giving and giving, and it felt like I had nothing left for myself." [P15, postpartum, aged 10 mo] • "I think the hardest part for me postpartum was not taking care of the baby but how do I take care of myself." [P40, postpartum, aged 15 mo]
Gaps in support[d]			

XSL•FO
RenderX

Theme		Description	Quotes
	Maternal mental health	Overemphasis on baby care left mothers feeling unsupported in managing their own emotional well-being.	• "... would have been nice to have more information on raising kids developmentally along with doing things that help me cope with boiling of stress." [P07, pregnant, second trimester, aged 3 y] • "All the apps that I have are pretty much focused on the baby. Besides the workout one, I would say there's not an app that's more focused on the mother." [P11, postpartum, aged 5 mo]
	Community isolation	Young or first-time moms felt isolated and lacked connections to peers for emotional support.	• "I've never been around people that are pregnant and haven't spent much time around little babies so I feel ignorant to most things regarding pregnancy, giving birth and having a child. Starting from complete scratch…" [P01, pregnant, second trimester]
	Limited time and support from clinicians	Frustration with short, impersonal doctor visits that prioritized medical information over emotional support	• "[Appointments are so short] that I didn't want to bother my doctor all the time with little things I was feelings." [P16, pregnant, third trimester] • "Yeah, the first few months were hard and I felt like that was when I needed the most like reassurance that like 'This is temporary.'" [P26, postpartum, aged 1 y, aged 4 y]
	Lack of understandable and credible sources	Frequently encountered misinformation online and desired reliable, easy-to-understand sources.	• "Message boards mostly don't work for me -- joined some of the What to Expect forums, and 75% of the time they lead you down such a rabbit hole, people have such different opinions and its not helpful and maybe scares you more. 25% of the time I like them because knowing a bunch of people who are going through the same things is helpful." [P35, pregnant, second trimester, aged 3 y]

[a]Integration into prototype: Moment for Parents was designed to offer validation, reassurance, and coping strategies for mothers. Developed by emotional well-being experts—including a doula, perinatal therapist, and yoga and meditation instructor—it incorporated real-life examples from mothers to highlight shared experiences.

[b]P: participant.

[c]Integration into prototype: Moment for Parents included acceptance and commitment therapy exercises to boost self-acceptance and psychological flexibility, helping moms process emotions such as isolation and anxiety. It also used cognitive behavioral therapy techniques to address worries about physical symptoms and mindfulness practices for coping with discomfort. During each check-in, the chatbot encouraged users to create a self-care plan and suggested activities lasting 5, 15, 30, and 60 minutes.

[d]Integration into prototype: Moment for Parents focused entirely on mothers' emotional and mental well-being, reinforcing the importance of self-care in supporting their baby. The chatbot guided moms in building a support system through exercises on accepting and asking for help, challenging the "do it all" myth, and promoting a mindset of community care.

XSL•FO
RenderX

Prototype Development

The results from the analysis of the ethnographic interviews directly influenced the content in the first Moment for Parents prototype. First, the "resources" theme resulted in exercises aiming to provide validation, reassurance, and coping strategies for users. Given specific user needs around social support, Moment for Parents also included real examples from parents with mental health struggles to illustrate shared experiences. Second, results from the "mental health journey" theme inspired acceptance and commitment therapy based exercises around self-acceptance and psychological flexibility. The goal for these exercises was to help users process their negative emotions, including anxiety and isolation, that users reported experiencing throughout the perinatal period. The "mental health journey" theme also led to the development of exercises tailored to the different phases of the perinatal period. For example, cognitive behavioral therapy exercises to increase coping with worries about physical symptoms in early pregnancy, and exercises focused on increasing self-care exercises during the postpartum period. Finally, the "gaps in support" theme led to the inclusion of exercises designed to help users increase help-seeking and build their support systems. The multidisciplinary team of experts vetted and fleshed out the content based on the user interview themes. Other aspects of the prototype, such as the structure (eg, frequency of messaging), were also informed by insights from the team of experts.

The first prototype consisted of a rules-based chatbot called Moment for Parents embedded within a Web app. Upon logging in, participants completed a mood check-in. Based on their mood at the check-in, the chatbot offered an exercise tailored to their current mood. The Moment for Parents prototype had a library of 35 exercises that the chatbot would select based on the users' reported mood. A second version of the prototype was adopted after 4 weeks based on participant feedback. In this version, participants were offered the option to select their own journey through the app. Options included the following: (1) an exercise based on mood, (2) an exercise on a predetermined topic (ie, "topic of the day"), and (3) a self-care check-in. The following day, the chatbot followed up with participants to encourage accountability and engagement in the exercises provided by the chatbot.

Phase 2 – Pilot Test

Participant Characteristics

A total of 108 participants, 29 pregnant and 79 post partum, participated in the pilot test of the Moment for Parents prototype. Among pregnant participants, 69% (n=20) were experiencing their first pregnancy, while 31% (n=9) were pregnant for their second time or more. Additionally, 17% (n=5) of pregnant participants were in their first trimester, 31% (n=9) in their second trimester, and 52% (n=15) in their third trimester. For postpartum participants, 82% (n=65) had 1 child, and 18% (n=14) had multiple children. Additionally, 25% (n=20) participants were 0 - 13 weeks post partum, 25% (n=20) participants were 14 - 26 weeks postpartum, 17% (n=13) participants were 27 - 40 weeks postpartum, and 33% (n=26) participants were 41 - 52 weeks postpartum.

Overall Engagement

Participants completed a total of 1036 check-ins and exercises between November 2019 and June 2020, with an average of 9.6 (SD 14.5) check-ins per participant and a mean of 4.83 weeks (SD 5.17) of engagement. A total of 31.9% (n=331) of check-ins occurred during the first 2 weeks, and 46.1% (n=478) occurred within the first month. Over the second month, check-ins dropped to 17.5% (n=181), followed by 9.6% (n=99) in the third month. During the fourth and fifth months, 6.5% (n=67) of check-ins occurred each month, followed by a slight rise to 6.8% (n=68) in the sixth month. Check-ins then decreased to 2.9% (n=30) in the seventh month and 1.0% (n=10) in the eighth month.

Table 2 presents the number of participants who engaged with the Web app at least once within each given month for all participants and separately for pregnant and postpartum groups. At 1 month, nearly all participants (98.1%) remained active, with higher retention among postpartum users (100%) compared to pregnant users (93.1%). Retention remained high at 2 months, with almost half (47.2%) of users still engaged. While engagement declined over time, postpartum users showed sustained interest, with 15.2% active at 7 months and 7.6% at 8 months.

XSL•FO
RenderX

Table . Moment for parents monthly active users over the 8-month evaluation (n=108).[a]

	All (N=108), n (%)	Pregnant (n=29), n (%)	Postpartum (n=79), n (%)
1 month	106 (98.1)	27 (93.1)	79 (100)
2 months	51 (47.2)	9 (31.0)	42 (53.2)
3 months	22 (20.4)	3 (10.3)	19 (24.1)
4 months	17 (15.7)	3 (10.3)	14 (17.7)
5 months	22 (20.4)	4 (13.8)	18 (22.8)
6 months	14 (13)	1 (3.4)	13 (16.5)
7 months	12 (11.1)	—[b]	12 (15.2)
8 months	6 (5.6)	—	6 (7.6)

[a]Most users (n=90, 83.3%) were active during their first week of use; 47.2% (51/108; 9/29, 31% pregnant, 42/79, 53.2% postpartum) used the app during their second week.

[b]Not available.

Re-Engagement Patterns

A total of 25 (5 pregnant and 20 postpartum) participants used the app for a median of 1 week (mean 1.8, SD 3.33 wk) before stopping and never returning. In contrast, 69 (14 pregnant and 55 postpartum) participants eventually re-engaged with the app after at least one week of inactivity. Among these participants, 44.9% (31/69) returned only once (8/14, 57% of pregnant participants; 23/55, 42% of postpartum participants). Re-engagement was operationalized as the number of times participants returned to the app after stopping for at least 1 week of inactivity. Of participants who re-engaged, a large proportion (28/69, 41%) did so 3 or more times after at least 1 week of inactivity. Postpartum participants re-engaged more often and more quickly than pregnant participants. Most re-engagements took place between 1 and 2 weeks after initial re-engagement, but a significant percentage of re-engagements occurred after 4 weeks (58/402, 14%; Table 3).

Table . Re-engagement patterns among those who re-engaged with moment for parents after stopping (n=69 participants; n=402 check-ins).

		All	Pregnant	Postpartum
Time to re-engagement (week), mean (SD)[a]		2.8 (2.6)	3.2 (3.2)	2.6 (2.5)
Number of re-engagements, n (%)[a]				
	1 re-engagement	31 (44.9)	8 (57.1)	23 (41.8)
	2 re-engagements	10 (14.5)	2 (14.3)	8 (14.5)
	3+ re-engagements	28 (40.6)	4 (28.6)	24 (43.6)
Number of check-ins after re-engagement, n (%)[b]				
	After 1 week	160 (39.8)	28 (38.9)	132 (40)
	After 2 weeks	106 (26.4)	19 (26.4)	87 (26.4)
	After 3 weeks	78 (19.4)	15 (20.8)	63 (19.1)
	After 4+ weeks	58 (14.4)	10 (13.9)	48 (14.5)

[a]All (n=69), pregnant (n=14), and postpartum (n=55).

[b]All (n=402), pregnant (n=72), and postpartum (n=330).

Prototype Feedback

Feedback from users (n=30) who completed the 1-month survey is in Table 4. Most users signed up for Moment for Parents because a friend recommended it (n=15, 55%). Other top reasons for signing up included an interest in becoming more resilient (n=14, 47%), wanting support for a self-care routine (n=13, 43%), and looking for advice to take care of themselves (n=12, 40%). The vast majority (n=28, 93%) felt that Moment for Parents was created for people such as them. Most users thought that Moment for Parents was a good fit for them because it kept them accountable for self-care (n=21, 70%), it reminded them to reflect (n=19, 63%), and it helped them take a mental health break (n=17, 58%). A large proportion of users also felt that Moment for Parents validated what they were going through (n=13, 43%), gave helpful advice (n=13, 43%), and helped them navigate the ups and downs of motherhood (n=12, 40%). While users generally had a positive experience with Moment for Parents, some users reported several “pet peeves” or negative experiences. Half of the users felt that the check-in topics were repetitive (n=15), and 37% (n=11) of users reported a pet peeve when the check-ins had a glitch. A third of users felt that the check-ins did not have the response options they wanted (n=10). Less frequently reported pet peeves included the daily text

XSL•FO
RenderX

reminders (n=8, 27%), the lack of access via a mobile app (n=6, 20%), and feeling as if the check-in topic was not relevant (n=6, 20%). The vast majority of users spent less than 5 minutes completing their check-ins (n=28, 93%), and most (n=24, 80%) felt that the check-in length was “just right.”

XSL•FO
RenderX

Table . Feedback from users who completed the 1-month survey (n=30).

Feedback		Value, n (%)
Why did you sign up for Moment for Parents?		
	Friend recommended	15 (55.3)
	Interested in becoming more resilient	14 (46.7)
	Interested in trying a new product for moms	9 (30)
	Looking for advice to take care of baby	4 (13.3)
	Looking for advice to take care of myself	12 (40)
	Help coping with emotions	10 (33.3)
	Support for self-care routine	13 (43.3)
	Other	1 (3.3)
Was Moment for Parents created for moms like you?		
	Yes	28 (99.3)
Why is Moment for Parents a good fit for you?		
	Feels like a friend	11 (36.7)
	Gives me confidence as a mom	5 (16.7)
	Gives me helpful advice	13 (43.3)
	Helps me become resilient	7 (23.3)
	Helps me navigate ups and downs of motherhood	12 (40)
	Helps me take a mental health break	17 (56.7)
	I know it’s there in case I need it	1 (3.3)
	It keeps me accountable for self-care	21 (70)
	It offers helpful mindfulness exercises	9 (30)
	It reminds me to reflect	19 (63.3)
	It surprises me with new perspectives	6 (20)
	It understands what I’m going through	7 (23.3)
	It validates what I’m going through	13 (43.3)
What are your pet peeves about Moment for Parents?		
	Can’t access via an app	6 (20)
	Daily text reminders	8 (26.7)
	None–it’s great as-is	3 (10)
	Not enough support for sleep	4 (13.3)
	Tone is too empathetic	1 (3.3)
	Something else	5 (16.7)
	Too much emphasis on self-care	3 (10)
	When check-in has a glitch	11 (36.7)
	When check-in questions don’t have the response option I want	10 (33.3)
	When texts come too quickly	3 (10)
	When texts come too slowly	1 (3.3)
	When check-in topic doesn’t feel relevant to me	6 (20)
	When check-in topics are repetitive	15 (50)
	When Moment for Parents doesn’t respond to me appropriately	5 (16.7)
Time to complete check-in		

XSL•FO
RenderX

Feedback		Value, n (%)
	Less than 1 min	3 (10)
	1 - 3 min	20 (66.7)
	3 - 5 min	5 (16.7)
	More than 5 min	2 (6.7)
Check-in length		
	Just right	24 (80)
	Too long	5 (16.7)
	Too short	1 (3.3)

Discussion

Summary of Findings

The purpose of this study was to (1) design and develop a chatbot tailored specifically to educate and support pregnant and postpartum people on emotional well-being and mental health and (2) evaluate the usability of Moment for Parents in a pilot test by examining usage patterns, engagement levels, and user experience. In the design phase, we found that user needs included social support, coping strategies, more attention to mental health and self-care, and reliable medical and mental health information. Users also emphasized a need for support that is tailored to changes over the perinatal period. The end result of phase 1 was an initial prototype of the Moment for Parents chatbot that was used in the phase 2 pilot test. The results from the pilot test showed that participants engaged with the app early on and that sustained engagement decreased over time. However, many participants re-engaged with the chatbot after inactivity, suggesting that users may have turned to Moment for Parents when they needed support. The results from the feedback survey showed that participants were primarily interested in using Moment for Parents to build resilience, support their self-care routines, and seek emotional support during pregnancy and postpartum.

Themes from the qualitative analysis highlighted the key needs and preferences of perinatal users. Participants consistently emphasized the importance of evidence-based strategies to support their mental health, particularly as a complement to in-person care, where provider time for emotional support may be limited. Additionally, users described the fast-paced and unpredictable nature of the perinatal period, reinforcing the need for flexible and adaptive digital interventions that align with their evolving emotional and psychological needs. These findings underscore the importance of delivering personalized, flexible, and easily accessible digital support that aligns with the evolving needs of perinatal individuals. Such alignment may also play a critical role in sustaining user engagement over time.

In the pilot trial, user engagement was greatest in the first 2 months, then decreased over the next 6 months, both in terms of check-ins with the app and monthly active users. Importantly, 47% of users were still active after 2 months—this level of usage is much greater than metrics described in other publicly available mental health applications where users drop to only 4% of the initial sample after 2 weeks [39]. There are several possible explanations for this finding. First, Moment for Parents was developed using HCD, which is thought to enhance engagement with digital mental health interventions [40]. Feedback showed that most participants felt that the information and exercises in Moment for Parents were highly personalized and relevant, which may have led to more sustained engagement. Another possible explanation is that, given the time-consuming demands faced by perinatal individuals such as attending medical appointments, managing physical symptoms, breastfeeding, and childcare, engaging with Moment for Parents required relatively little time and energy from users, as most users reported spending less than 5 minutes with Moment for Parents daily. This is in contrast to other apps and chatbots that are more time and resource intensive [41,42].

While most participants disengaged after 2 months, many re-engaged with Moment for Parents at least once during the pilot. Among those who returned, a large proportion did so 3 or more times, suggesting they found the app useful or relevant to their needs. Re-engagement was highest in the first week (40%), but a notable portion (14%) continued to return after 4 weeks, indicating that users may have turned to the app specifically when they needed support or information. A recent review of 16 trials of digital mental health interventions for perinatal populations found that disengagement varied widely across trials (13% - 90% retention rates) and highlights the challenge of engaging perinatal users [43]. Moment for Parents was designed to deliver targeted support based on mood check-ins rather than continuous interaction, highlighting the importance of minimal contact and flexible mental health tools. Future research should prioritize understanding the timing and reasons behind user re-engagement, rather than solely aiming to increase engagement frequency, to enhance the mental health chatbots' ability to deliver timely, relevant support tailored to users' needs. This could include exploring the specific triggers for re-engagement, the impact of intermittent use on mental health outcomes, and how to refine content delivery to maximize effectiveness when users do return.

Interestingly, both engagement and re-engagement were greater among postpartum participants compared to pregnant participants. Postpartum users may have been experiencing greater mental health challenges than their pregnant counterparts in this sample and, therefore, turned to the app more frequently for exercises to improve their mood. Relatedly, postpartum users may have sought more information from the app as they learned to care for their infants, managed their postpartum

XSL•FO
RenderX

mental health, and adjusted to new parenthood. Additionally, the postpartum period is marked by significant hormonal changes, sleep deprivation, and increased caregiving demands, all of which contribute to low mood [44-46]. Given these challenges, postpartum individuals may be more open to digital mental health support as a convenient and accessible resource for coping strategies, reassurance, and parenting guidance [47]. In contrast, pregnant users may have relied more on in-person medical support or other existing resources, leading to lower engagement with the app. Alternatively, some users indicated in informal feedback that the content presented in Moment for Parents was geared toward postpartum more than pregnant user needs. Overall, the limited number of pregnant individuals in our sample prevents definitive conclusions, highlighting the need for further research on tailoring interventions to pregnant user needs.

The findings from the 1-month feedback survey reinforced the value of Moment for Parents in supporting perinatal mental health. The users reported that the tool provided helpful advice, supported self-care accountability, and facilitated emotional reflection, highlighting its role in bridging gaps in traditional care [48,49]. Many found the content relevant and beneficial, though some noted frustration with repetitive or irrelevant check-ins, emphasizing the need for customizable, phase-specific content. These findings underscore the potential of personalized digital interventions to enhance accessibility and engagement in perinatal mental health support.

Limitations

This study was not without limitations. First, the samples used in both phases of this study were relatively small and prone to bias because of the snowball sampling method used. Certain groups or participant characteristics may have been overrepresented, leading to skewed results. Furthermore, only 25% of the initial sample completed the 1-month survey, leading to potentially biased survey results if, for example, only users who enjoyed the app completed the survey. Relatedly, another limitation was that no demographic data was collected about study participants in either phase, limiting our ability to speak to the generalizability or biases in the samples studied. Second, the sample primarily consisted of postpartum users, limiting our ability to make inferences about the feasibility and utility of Moment for Parents across the perinatal period. Future work should focus on recruiting a larger sample of pregnant users at various stages of pregnancy, as well as postpartum users at various postpartum stages (eg, early postpartum vs late postpartum). Third, the way the feedback questions were asked may have led participants to provide overly positive feedback (eg, “why was Moment for Parents a good fit” vs “to what extent do you agree that Moment for Parents was a good fit?”). Future evaluations of Moment for Parents should include a more balanced survey to capture a variety of perspectives. Fourth, systematic data were not collected around the nature and timing of the iterative updates, which limits our ability to understand how these updates impacted engagement or satisfaction with Moment for Parents. Finally, data were not collected around user enactment of the Moment for Parents exercises outside of the check-ins. Future research should focus on understanding how users integrate the strategies they learn from Moment for Parents into their daily lives.

Conclusions

This study provides early evidence that Moment for Parents is a practical and scalable tool for perinatal mental health support. HCD informed the chatbot’s development, ensuring it aligned with the needs of perinatal users before being tested in a pilot trial. Findings showed that participants engaged with the chatbot far longer than is typical for digital interventions and often returned for support even weeks after initially stopping use—highlighting its usability and value. However, more research is needed to fully understand the chatbot’s efficacy in mitigating adverse perinatal mental health outcomes. These results contribute to the growing body of research on digital mental health interventions in the perinatal period, reinforcing the potential of chatbots to deliver accessible, on-demand support.

Data Availability

The dataset used for this retrospective analysis is not publicly available. However, aggregated and anonymized data may be shared with interested parties upon reasonable request.

Conflicts of Interest

MK is employed by Poisera, Inc (dba Moment for Parents), and Moment for Parents was used in this study. JH is a paid consultant of Moment for Parents, and KM and LB are employed by JH. However, authors’ employment status or salary are not dependent upon the results of their research.

References

1. Maternal mental health. World Health Organization. URL: https://www.who.int/teams/mental-health-and-substance-use/promotion-prevention/maternal-mental-health [accessed 2025-03-28]
2. Perinatal mental health and substance use disorders [white paper]. Americal Psychiatric Association. 2023. URL: https://www.psychiatry.org/getmedia/344c26e2-cdf5-47df-a5d7-a2d444fc1923/APA-CDC-Perinatal-Mental-and-Substance-Use-Disorders-Whitepaper.pdf [accessed 2025-03-28]
3. Reproductive health: symptoms of depression among women. CDC. URL: https://www.cdc.gov/reproductive-health/depression/index.html [accessed 2025-03-28]

XSL•FO
RenderX

4. Howard LM, Khalifeh H. Perinatal mental health: a review of progress and challenges. World Psychiatry 2020 Oct;19(3):313-327. [doi: 10.1002/wps.20769] [Medline: 32931106]
5. Petersen EE, Davis NL, Goodman D, et al. Racial/ethnic disparities in pregnancy-related deaths - United States, 2007-2016. MMWR Morb Mortal Wkly Rep 2019 Sep 6;68(35):762-765. [doi: 10.15585/mmwr.mm6835a3] [Medline: 31487273]
6. Ceballos M, Wallace G, Goodwin G. Postpartum depression among African-American and Latina mothers living in small cities, towns, and rural communities. J Racial Ethn Health Disparities 2016 Oct 19;4:916-927. [doi: 10.1007/s40615-016-0295-z] [Medline: 27761728]
7. Slomian J, Honvo G, Emonts P, Reginster JY, Bruyère O. Consequences of maternal postpartum depression: a systematic review of maternal and infant outcomes. Womens Health (Lond) 2019;15:1745506519844044. [doi: 10.1177/1745506519844044] [Medline: 31035856]
8. Biaggi A, Conroy S, Pawlby S, Pariante CM. Identifying the women at risk of antenatal anxiety and depression: a systematic review. J Affect Disord 2016 Feb;191:62-77. [doi: 10.1016/j.jad.2015.11.014] [Medline: 26650969]
9. Kingston D, Austin MP, Heaman M, et al. Barriers and facilitators of mental health screening in pregnancy. J Affect Disord 2015 Nov 1;186:350-357. [doi: 10.1016/j.jad.2015.06.029] [Medline: 26281038]
10. Ford E, Roomi H, Hugh H, van Marwijk H. Understanding barriers to women seeking and receiving help for perinatal mental health problems in UK general practice: development of a questionnaire. Prim Health Care Res Dev 2019 Dec 12;20:e156. [doi: 10.1017/S1463423619000902] [Medline: 31826794]
11. Button S, Thornton A, Lee S, Shakespeare J, Ayers S. Seeking help for perinatal psychological distress: a meta-synthesis of women's experiences. Br J Gen Pract 2017 Oct;67(663):e692-e699. [doi: 10.3399/bjgp17X692549] [Medline: 28847773]
12. Brailey C, Slatton BC. Centering Black women's voices: illuminating systemic racism in maternal healthcare experiences. Societies (Basel) 2024 May;14(5):70. [doi: 10.3390/soc14050070]
13. Estriplet T, Morgan I, Davis K, Crear Perry J, Matthews K. Black perinatal mental health: prioritizing maternal mental health to optimize infant health and wellness. Front Psychiatry 2022;13:807235. [doi: 10.3389/fpsyt.2022.807235] [Medline: 35573337]
14. Moore D, Ayers S, Drey N. A thematic analysis of stigma and disclosure for perinatal depression on an online forum. JMIR Ment Health 2016 May 19;3(2):e18. [doi: 10.2196/mental.5611] [Medline: 27197516]
15. Megnin-Viggars O, Symington I, Howard LM, Pilling S. Experience of care for mental health problems in the antenatal or postnatal period for women in the UK: a systematic review and meta-synthesis of qualitative research. Arch Womens Ment Health 2015 Dec;18(6):745-759. [doi: 10.1007/s00737-015-0548-6] [Medline: 26184835]
16. Peahl AF, Moniz MH, Heisler M, et al. Experiences with prenatal care delivery reported by black patients with low income and by health care workers in the US: a qualitative study. JAMA Netw Open 2022 Oct 3;5(10):e2238161. [doi: 10.1001/jamanetworkopen.2022.38161] [Medline: 36279136]
17. Song H, Cramer EM, McRoy S, May A. Information needs, seeking behaviors, and support among low-income expectant women. Women Health 2013;53(8):824-842. [doi: 10.1080/03630242.2013.831019] [Medline: 24215275]
18. Lewkowitz AK, Whelan AR, Ayala NK, et al. The effect of digital health interventions on postpartum depression or anxiety: a systematic review and meta-analysis of randomized controlled trials. Am J Obstet Gynecol 2024 Jan;230(1):12-43. [doi: 10.1016/j.ajog.2023.06.028]
19. Suharwardy S, Ramachandran M, Leonard SA, et al. Feasibility and impact of a mental health chatbot on postpartum mental health: a randomized controlled trial. AJOG Glob Rep 2023 Aug;3(3):100165. [doi: 10.1016/j.xagr.2023.100165] [Medline: 37560011]
20. Gaffney H, Mansell W, Tai S. Conversational agents in the treatment of mental health problems: mixed-method systematic review. JMIR Ment Health 2019 Oct 18;6(10):e14166. [doi: 10.2196/14166] [Medline: 31628789]
21. Car LT, Dhinagaran DA, Kyaw BM, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug 7;22(8):e17158. [doi: 10.2196/17158] [Medline: 32763886]
22. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464. [doi: 10.1177/0706743719828977] [Medline: 30897957]
23. Booth F, Potts C, Bond R, et al. A mental health and well-being chatbot: user event log analysis. JMIR mHealth uHealth 2023 Jul 6;11:e43052. [doi: 10.2196/43052] [Medline: 37410539]
24. Lucas GM, Gratch J, King A, Morency LP. It's only a computer: virtual humans increase willingness to disclose. Comput Human Behav 2014 Aug;37:94-100. [doi: 10.1016/j.chb.2014.04.043]
25. Lucas GM, Rizzo A, Gratch J, et al. Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. Front Robot AI 2017;4:51 [FREE Full text] [doi: 10.3389/frobt.2017.00051]
26. Stiles-Shields C, Cummings C, Montague E, Plevinsky JM, Psihogios AM, Williams KDA. A call to action: using and extending human-centered design methodologies to improve mental and behavioral health equity. Front Digit Health 2022;4:848052. [doi: 10.3389/fdgth.2022.848052] [Medline: 35547091]
27. Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018 Aug 1;25(8):1080-1088. [doi: 10.1093/jamia/ocy052] [Medline: 29788380]

XSL•FO
RenderX

28. Are pregnant and postpartum women interested in health-related apps? Implications for the prevention of perinatal depression. Consensus. URL: https://consensus.app/papers/are-pregnant-and-postpartum-women-interested-in-osma-barrera/43c0340c7c765fca9ee27233f11516f4/?utm_source=chatgpt [accessed 2025-03-28]
29. Connor K, Wambach K, Baird MB. Descriptive, qualitative study of women who use mobile health applications to obtain perinatal health information. J Obstet Gynecol Neonatal Nurs 2018 Nov;47(6):728-737. [doi: 10.1016/j.jogn.2018.04.138] [Medline: 29856963]
30. Tsai Z, Kiss A, Nadeem S, et al. Evaluating the effectiveness and quality of mobile applications for perinatal depression and anxiety: a systematic review and meta-analysis. J Affect Disord 2022 Jan 1;296:443-453. [doi: 10.1016/j.jad.2021.09.106] [Medline: 34624434]
31. Mackintosh N, Agarwal S, Adcock K, et al. Online resources and apps to aid self-diagnosis and help seeking in the perinatal period: a descriptive survey of women's experiences. Midwifery 2020 Nov;90:102803. [doi: 10.1016/j.midw.2020.102803] [Medline: 32717660]
32. Gordon M, Henderson R, Holmes JH, Wolters MK, Bennett IM, SPIRIT (Stress in Pregnancy: Improving Results with Interactive Technology) Group. Participatory design of ehealth solutions for women from vulnerable populations with perinatal depression. J Am Med Inform Assoc 2016 Jan;23(1):105-109. [doi: 10.1093/jamia/ocv109] [Medline: 26342219]
33. Rathnayaka P, Mills N, Burnett D, De Silva D, Alahakoon D, Gray R. A mental health chatbot with cognitive skills for personalised behavioural activation and remote health monitoring. Sensors (Basel) 2022 May 11;22(10):3653. [doi: 10.3390/s22103653] [Medline: 35632061]
34. Abd-Alrazaq AA, Alajlani M, Alalwan AA, Bewick BM, Gardner P, Househ M. An overview of the features of chatbots in mental health: a scoping review. Int J Med Inform 2019 Dec;132:103978. [doi: 10.1016/j.ijmedinf.2019.103978] [Medline: 31622850]
35. Mancinelli E, Magnolini S, Gabrielli S, Salcuni S. A chatbot (Juno) prototype to deploy a behavioral activation intervention to pregnant women: qualitative evaluation using a multiple case study. JMIR Form Res 2024 Aug 14;8:e58653. [doi: 10.2196/58653] [Medline: 39140593]
36. ISO 9241-210:2019. ISO. URL: https://www.iso.org/standard/77520.html [accessed 2025-03-28]
37. López L, Green AR, Tan-McGrory A, King R, Betancourt JR. Bridging the digital divide in health care: the role of health information technology in addressing racial and ethnic disparities. Jt Comm J Qual Patient Saf 2011 Oct;37(10):437-445. [doi: 10.1016/s1553-7250(11)37055-9] [Medline: 22013816]
38. Bradshaw C, Atkinson S, Doody O. Employing a qualitative description approach in health care research. Glob Qual Nurs Res 2017;4:2333393617742282. [doi: 10.1177/2333393617742282] [Medline: 29204457]
39. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567. [doi: 10.2196/14567] [Medline: 31573916]
40. Mohr DC, Riper H, Schueller SM. A solution-focused research approach to achieve an implementable revolution in digital mental health. JAMA Psychiatry 2018 Feb 1;75(2):113-114. [doi: 10.1001/jamapsychiatry.2017.3838] [Medline: 29238805]
41. Darcy A, Beaudette A, Chiauzzi E, et al. Anatomy of a Woebot® (WB001): agent guided CBT for women with postpartum depression. Expert Rev Med Devices 2023;20(12):1035-1049. [doi: 10.1080/17434440.2023.2280686] [Medline: 37938145]
42. Hussain-Shamsy N, Shah A, Vigod SN, Zaheer J, Seto E. Mobile health for perinatal depression and anxiety: scoping review. J Med Internet Res 2020 Apr 13;22(4):e17011. [doi: 10.2196/17011] [Medline: 32281939]
43. Davis JA, Ohan JL, Gibson LY, Prescott SL, Finlay-Jones AL. Understanding engagement in digital mental health and well-being programs for women in the perinatal period: systematic review without meta-analysis. J Med Internet Res 2022 Aug 9;24(8):e36620. [doi: 10.2196/36620] [Medline: 35943773]
44. Schiller CE, Meltzer-Brody S, Rubinow DR. The role of reproductive hormones in postpartum depression. CNS Spectr 2015 Feb;20(1):48-59. [doi: 10.1017/S1092852914000480] [Medline: 25263255]
45. Iranpour S, Kheirabadi GR, Esmaillzadeh A, Heidari-Beni M, Maracy MR. Association between sleep quality and postpartum depression. J Res Med Sci 2016;21(1):110. [doi: 10.4103/1735-1995.193500]
46. Saharoy R, Potdukhe A, Wanjari M, Taksande AB. Postpartum depression and maternal care: exploring the complex effects on mothers and infants. Cureus 2023 Jul;15(7):e41381. [doi: 10.7759/cureus.41381] [Medline: 37546054]
47. Feldman N, Perret S. Digital mental health for postpartum women: perils, pitfalls, and promise. NPJ Digit Med 2023 Jan 26;6(1):11. [doi: 10.1038/s41746-023-00756-4] [Medline: 36702915]
48. Zeng Z, Peng J, Liu L, Gong W. Translating research evidence into marketplace application: cohort study of internet-based intervention platforms for perinatal depression. J Med Internet Res 2023 Apr 17;25(1):e42777. [doi: 10.2196/42777] [Medline: 37067855]
49. Feldman N, Back D, Boland R, Torous J. A systematic review of mHealth application interventions for peripartum mood disorders: trends and evidence in academia and industry. Arch Womens Ment Health 2021 Dec;24(6):881-892. [doi: 10.1007/s00737-021-01138-z] [Medline: 33929636]

Abbreviations

BRANY: Biomedical Research Alliance of New York

XSL•FO
RenderX

HCD: human-centered design

Edited by A Mavragani; submitted 10.02.25; peer-reviewed by A Lewkowitz, R Joly; revised version received 10.03.25; accepted 11.03.25; published 08.04.25.

Please cite as:
McAlister K, Baez L, Huberty J, Kerppola M
Chatbot to Support the Mental Health Needs of Pregnant and Postpartum Women (Moment for Parents): Design and Pilot Study
JMIR Form Res 2025;9:e72469
URL: https://formative.jmir.org/2025/1/e72469
doi:10.2196/72469

© Kelsey McAlister, Lara Baez, Jennifer Huberty, Marianna Kerppola. Originally published in JMIR Formative Research (https://formative.jmir.org), 8.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Anticipated Acceptability of Blended Learning Among Lay Health Care Workers in Malawi: Qualitative Analysis Guided by the Technology Acceptance Model

Tiwonge E Mbeya-Munkhondya[1], PhD; Caroline J Meek[2], MPH; Mtisunge Mphande[3], BSc; Tapiwa A Tembo[3], MSc; Mike J Chitani[3], BA; Milenka Jean-Baptiste[4], MPH; Caroline Kumbuyo[3], BSoc; Dhrutika Vansia[3], MPH; Katherine R Simon[3,5], MD; Sarah E Rutstein[6], MD, PhD; Victor Mwapasa[7], PhD; Vivian Go[4], PhD; Maria H Kim[3,5], MD; Nora E Rosenberg[4], PhD

[1]School of Nursing, Kamuzu University of Health Sciences, Lilongwe, Malawi

[2]RTI International, Research Triangle Park, NC, United States

[3]Baylor College of Medicine Children's Foundation, Golden Peacock, 5 Floor, Lilongwe, Malawi

[4]Gillings School of Public Health, Chapel Hill, NC, United States

[5]Department of Pediatrics, Baylor College of Medicine, Houston, TX, United States

[6]Department of Medicine, University or North Carolina, Chapel Hill, NC, United States

[7]Department of Community and Environmental Health, Kamuzu University of Health Sciences, Blantyre, Malawi

Corresponding Author:
Dhrutika Vansia, MPH
Baylor College of Medicine Children's Foundation, Golden Peacock, 5 Floor, Lilongwe, Malawi

Abstract

Background: HIV index case testing (ICT) aims to identify people living with HIV and their contacts, engage them in HIV testing services, and link them to care. ICT implementation has faced challenges in Malawi due to limited counseling capacity among lay health care workers (HCWs). Enhancing capacity through centralized face-to-face training is logistically complex and expensive. A decentralized blended learning approach to HCW capacity-building, combining synchronous face-to-face and asynchronous digital modalities, may be an acceptable way to address this challenge.

Objective: The objective of this analysis is to describe factors influencing HCW anticipated acceptability of blended learning using the Technology Acceptance Model (TAM).

Methods: This formative qualitative study involved conducting 26 in-depth interviews with HCWs involved in the ICT program across 14 facilities in Machinga and Balaka, Malawi (November-December 2021). Results were analyzed thematically using TAM. Themes were grouped into factors affecting the 2 sets of TAM constructs: perceived usefulness and perceived ease of use.

Results: A total of 2 factors influenced perceived usefulness. First, HCWs found the idea of self-guided digital learning appealing, as they believed it would allow for reinforcement, which would facilitate competence. They also articulated the need for opportunities to practice and receive feedback through face-to-face interactions in order to apply the digital components. In total, 5 factors influenced perceived ease of use. First, HCWs expressed a need for orientation to the digital technology given limited digital literacy. Second, they requested accessibility of devices provided by their employer, as many lacked personal devices. Third, they wished for adequate communication surrounding their training schedules, especially if they were going to be asynchronous. Fourth, they wished for support for logistical arrangements to avoid work interruptions. Finally, they wanted monetary compensation to motivate learning, a practice comparable with offsite trainings.

Conclusions: A decentralized blended learning approach may be an acceptable method of enhancing ICT knowledge and skills among lay HCWs in Malawi, although a broad range of external factors need to be considered. Our next step is to integrate these findings into a blended learning package and examine perceived acceptability of the package in the context of a cluster randomized controlled trial.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2023-077706

(JMIR Form Res 2025;9:e62741) doi:10.2196/62741

KEYWORDS

blended learning; technology acceptance model; index case testing; acceptability; partner notification; contact tracing

XSL•FO
RenderX

Introduction

While Malawi is making substantial progress in achieving the Joint United Nations Programme on HIV and AIDS (UNAIDS) 95-95-95 targets by 2030, the first target (HIV status awareness by persons living with HIV) has not been reached [1]. To find persons living with HIV unaware of their HIV status, the World Health Organization (WHO) promotes HIV index case testing (ICT), in which contacts of individuals living with HIV are identified, counselled, offered HIV testing, and linked to care [2]. Specifically, WHO recommends an option for voluntary assisted partner notification, in which health care workers help index clients invite their contacts for HIV testing, an approach that has been incorporated into Malawian standard of care [3-5].

Despite these recommendations, health care workers (HCWs) have limited counseling capacity that hinders full realization of voluntary assisted partner notification. For instance, formative work in Malawi identified limited counselling capacity among HCWs as a major challenge to enhancing the implementation of ICT [5]. ICT is offered by lay HCWs, cadres who lack professional training but are expected to perform select tasks within the health care system based on brief in-service training. Standard ICT training for lay HCWs is centralized, in-person, and delivered didactically. However, such approaches are logistically challenging and costly. A blended learning implementation package, which can be offered in a decentralized manner, asynchronously, and with nondidactic components is a promising alternative training approach for HCWs.

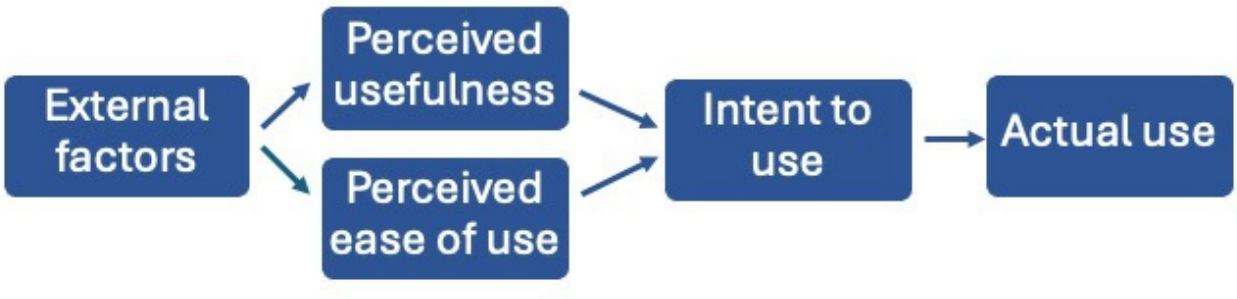
Blended learning, which combines face-to-face with digital modalities [6-8], is often more effective than digital or face-to-face training alone [9-11]. In recent years, the use of digital and blended educational approaches have increased, especially with COVID-19–driven pivots to remote learning [12-17]. Blended learning has promoted student motivation, skill-building, and achievement in several contexts [18-20]. However, blended learning has not been fully evaluated for implementation outcomes, especially among lay HCWs in low- and middle-income countries (LMICs) with limited technological literacy [21,22]. Thus, there is limited evidence as to whether lay cadres would find blended learning to be acceptable. We thus developed a cluster randomized controlled trial to assess the impact of a blended learning implementation package compared with the standard of care training on implementation, effectiveness, and cost-effectiveness outcomes [23]. Before developing this package, we conducted formative research to understand factors that may influence the acceptability of blended learning among lay HCWs in LMICs with limited technological literacy [24,25]. This formative work uses in-depth interviews (IDIs) to assess lay HCWs' perceptions of a blended learning approach to ICT training. Our analysis was guided by the Technology Acceptance Model (TAM).

Methods

Conceptual Framework

To explore the anticipated acceptability of a proposed blended learning implementation package, we used the TAM to guide data analysis. TAM is a theoretical model that examines constructs influencing the acceptance or rejection of technologies [26]. The model proposes that acceptability of any technology can be predicted by 2 factors: perceived usefulness and perceived ease of use (Figure 1). Perceived usefulness refers to the belief that a certain technology would increase an individual's job performance. Perceived ease of use is a belief that the use of the new technology would be effortless [27]. TAM proposes that a populations' perceptions of usefulness and ease of use of a technology impact the population's intention to use the technology, which in turn influences the population's actual usage of the technology. Furthermore, the key constructs of perceived usefulness and perceived ease of use are influenced by external factors, such as pedagogy, digital literacy, infrastructure, and the learning environment. Understanding these external factors is essential for designing a such a package. We used the TAM in this formative sub-study to guide the analysis of the relevant external factors influencing perceived usefulness and perceived ease of use surrounding blended learning in the context of ICT.

Figure 1. The Technology Acceptance Model.



Setting

This study was conducted at health centers in Balaka and Machinga, two districts in Southeastern Malawi. Malawi is a country in Southeastern Africa with an adult HIV prevalence of 8.9% [1]. Although Malawi has met the second two UNAIDS 95-95-95 targets, it has not yet reached the first 95% HIV testing target, with only 88.3% of adults living with HIV aware of their HIV status at the time of the study [1]. Simultaneously, Malawi has one of the lowest rates of qualified HCWs globally with less than 1 medical doctor per 10,000 people [28], resulting in task shifting of HIV services to lay cadres with less formal training [29]. Particularly relevant to the proposed blended learning strategy, is the fact that Malawi has one of the lowest digital literacy levels (having the necessary skills and knowledge

XSL•FO
RenderX

to confidently and effectively use technology) globally ranking 131 out of 134 countries by Network Readiness Index [30].

Study Design and Population

This formative work was conducted to inform a cluster randomized controlled trial designed to examine the implementation, effectiveness, and cost-effectiveness of a blended learning ICT package [23]. Data for this qualitative substudy were generated through IDIs conducted with lay HCWs between November and December 2021. This formative study purposively selected facilities from both districts (Machinga and Balaka) and different levels of care (dispensaries, health centers, and hospitals) for purposes of generalizability. To be eligible for participation, HCWs had to be at least 18 years old, work full-time at one of the health facilities, and be involved in counseling index clients or tracing contact clients. With engagement from facility supervisors, the research team identified up to 2 eligible participants in each facility. These individuals were actively involved in ICT at the time of the study. A research team member visited all the eligible participants in the 14 facilities to give them more details about the study and to seek their consent. Contact details of all consenting participants were shared with one of the research team members who made arrangements with the participants for the date and time of the interviews.

Data Collection

IDIs were conducted by 3 trained Malawian qualitative interviewers in Chichewa, the local language. All interviews were conducted by 3 female external interviewers. Among them, 1 interviewer had a doctoral degree in social sciences and 2 had bachelor's degrees in social sciences. The interviewers were trained on techniques of normalization and probing to reduce the risk of socially desirable response bias and had previous training and experience in qualitative data collection and research positions within the organization (postdoctoral fellow, senior research officer, and research officer). Furthermore, interviews were conducted in a private setting at the facilities, with only the interviewers and participants present. Interviews were conducted using a semistructured guide developed for the study through a series of rigorous, iterative discussions among the research team. The questions for this analysis focused on feasibility and acceptability of a blended learning implementation package.

After consenting and before conducting the interview, the research staff introduced themselves and their role, explained the purpose of the interview, and reiterated that they could refuse to answer questions, that all data would be anonymized, and that responses would have no consequences on their employment. Interviewers first asked questions about participants' views of the standard training. Interviewers then provided participants with a brief description of the planned blended learning approach, explaining that it involved tablet-based asynchronous lessons modeling counseling approaches, as well as interactive synchronous in-person group sessions to practice the skills learned. Finally, they asked about perceived feasibility, probing on potential barriers. Before the start of the IDIs, the interview guides were translated, back translated, and pilot tested with iterative refinement. Interviews were conducted over the phone or face-to-face and lasted approximately an hour long. Repeat interviews were not carried out. Interviewers summarized each IDI immediately after it was conducted. Audio recordings were transcribed verbatim in both Chichewa and translated to English by professional transcribers. Transcriptions were not returned to the participants.

Data Analysis

The team took a thematic approach to coding [31]. First, the team read the summaries to identify key themes and observed that they were framed by TAM. The research team then developed a qualitative codebook based on a focused examination of the transcripts, debrief reports, and interview summaries. The codebook consisted of both deductive codes from the interview guides and inductive codes that mapped onto TAM constructs. In addition, 2 trained research analysts (TEM and CJM) applied codes to transcript text using the Dedoose (SocioCultural Research Consultants) web application. As a first step, each analyst independently coded the same transcript and compared code applications to ensure agreements in the coding process. The analysts discussed and resolved any coding disparities, updating the codebook as needed to facilitate a shared understanding of all codes. The analysts then met weekly throughout the coding process, which provided space to review code applications and address any challenges. After coding was completed, analysts reviewed the coded text to extract key themes through the identification of patterns, naming and defining themes and refining the identified themes. Analysis was organized by factors that influence the TAM's constructs: perceived usefulness (what participants perceived as gained or lost by the blended learning implementation package) and perceived ease of use (what participants perceived as facilitators or barriers to acceptance of the blended learning implementation package). The findings were not shared with the participants for their feedback.

Ethical Considerations

Ethical clearance was provided by the University of North Carolina at Chapel Hill institutional review board (#20 - 1810), the Malawi National Health Sciences Research Committee (#20/06/2566) and the Baylor College of Medicine institutional review board (H-48800). Interviewers obtained written informed consent from all participants before starting the IDIs, reminding participants that their participation was voluntary and could be withdrawn at any time. Participants were compensated the equivalent of US $10 per interview. All data were deidentified to protect participant confidentiality.

Results

Demographic Summary

A total of 14 facilities participated in the IDIs. Among those, 2 of the facilities were district hospitals, 9 were health centers, and 3 were dispensaries. Furthermore, 9 facilities were in Machinga and 5 were in Balaka. In total, 26 HCWs participated in the IDIs. A total of 7 were female and 19 were male. Furthermore, 16 participants were from Machinga and 10 were from Balaka. There were no participants that refused to participate. All the participants were involved in both screening

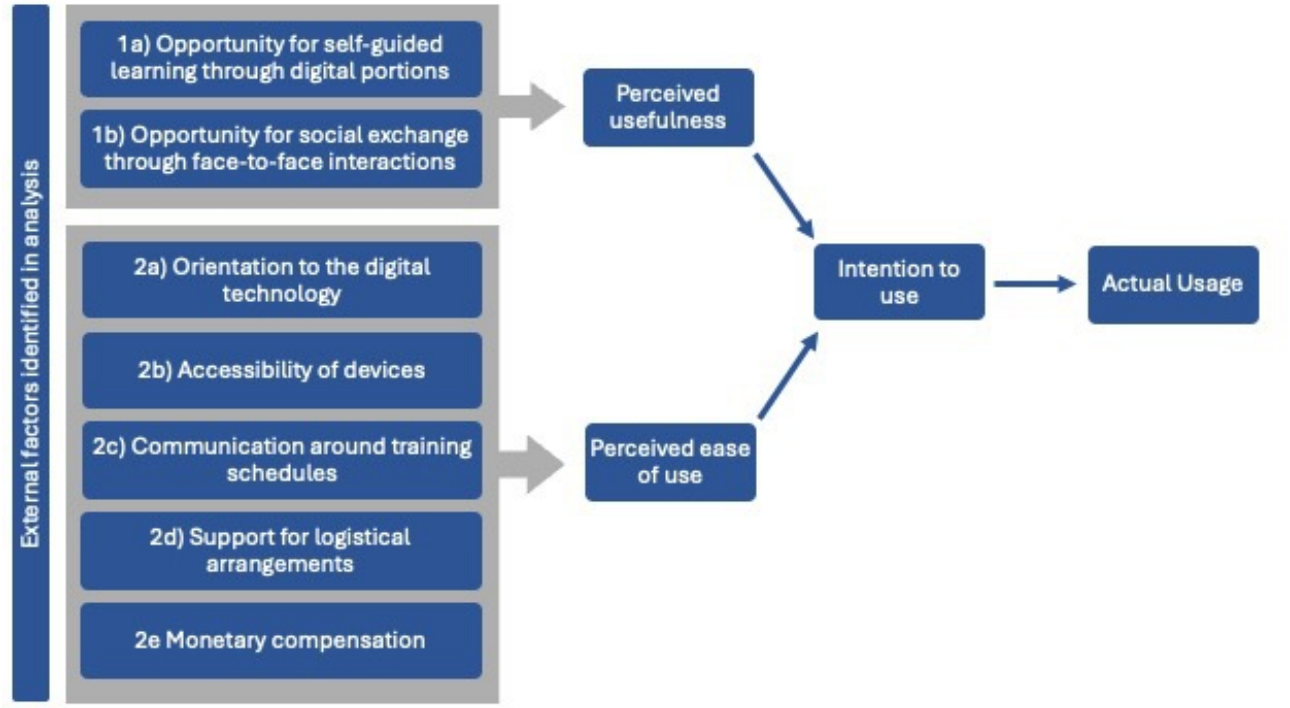
XSL•FO
RenderX

and counselling index clients and conducting contact tracing and had previous exposure to standard ICT training.

Themes

Our analysis identified two sets of factors influencing perceived usefulness: (1a) opportunity for self-guided learning through digital portions and (1b) opportunity for social exchange through face-to-face interactions. We identified 5 factors influencing perceived ease of use. In addition, 2 factors; (2a) need for orientation to the digital technology and (2b) accessibility of devices were related to perceived ease of use to the technology itself, while 3 factors (2c) communication around training schedules, (2d) support for logistical arrangements, and (2e) monetary compensation were related to perceived ease of use of the training approach (Figure 2). The following sections develop these themes, and Table 1 shows the illustrative quotes.

Figure 2. External factors that contribute to perceived usefulness and perceived ease of use.



XSL•FO
RenderX

Table . Detailed explanations and illustrative quotations for each external factor.

External factors		Quote
1. Perceived usefulness		
a. Opportunity for self-guided learning		
	Able to learn better.	• …. because I will be doing the things myself, I will be able to understand unlike having someone read to me or stand in front and say that this is this. [P 8021]
	Able to review the training content at their own pace	• You can repeat at your own time where you do not understand. [P 8007]
b. Opportunity for interaction		
	Ability to ask questions and seek clarifications.	• ….the tablet cannot respond to questions that one may have as you do when you have a facilitator. [P 8003] • …. with the face-to-face portion if you have other questions maybe of which you failed to ask the tablet, … then you will ask the facilitator and he will. [P 8024]
	Ability to practice and receive feedback.	• … face to face, gives us … a chance to be practicing and see which areas we are doing well and areas in which we are lacking and help one another while we are together. [P 8013]
	Interacting with others	• … it means we will be able to meet people from different facilities and we will be able to share experiences. [P 8006]
2. Perceived ease of use		
a. Need for orientation to the digital technology.		
	Need for orientation on use of tablet.	• There can be some people who have never used a tablet and before the training, they should be asked; “have you ever used a tablet?” If the answer is “no,” they should be taught how to use the tablet. [P 8011]
b. Accessibility to the digital services		
	Questions and concerns regarding tablet availability	• ….if there would be a number of people with few tablets that will be a ‘challenge’ since you will be waiting for your friends to finish then wait for another one. [P 8023] • …if the networks are malfunctioning, it will definitely affect your work. [P 8012]

XSL•FO
RenderX

External factors			Quote
		Technical issues affecting usage and safety that could affect tablet access.	• …okay maybe at the time when you are using it [tablet] (…) maybe has run out of power or maybe that thing has a technical fault, [P 8007] • ….the tablet can get lost or maybe it fell down and it has broken. [P 8009]
	c. Communication around training schedules		
		Engaging in open communication about why certain individuals have been selected to attend tablet training.	• People can have thoughts if they can be seeing you working on a tablet and if the supervisor did not say that [name of participant] will be working on a tablet on this day. Your friends can think that you are just playing but if the supervisor can tell them they can know what is going on. [P 8004]
	d. Logistical arrangements		
		Location—Preference for the training location	• I would love if the training can be conducted in my catchment area and the other people would also want the training to be conducted in their catchment area. So, my opinion is that the training should be conducted at a convenient venue for most of the participants. [P 8006] • ….Just because' maybe 'you are far from' your home, 'you feel' that you are at a training… [P 8007]
		Time – perception of having the training over the weekend.	• So, if the training has happened on a weekend, there's no problem, it is your free time. [P 8012] • …we like using the weekend to…focus on our own things for our lives yeah so maybe many people would be affected if the training is done during the weekend. [P 8007]
	e. Monetary compensation		

XSL•FO
RenderX

External factors		Quote
	Need for allowance.	• On the allowances I can say that it would be good if we are given a little something so that we can be assisted. [P 8019] • …an allowance is an allowance. It needs to be there. [P 8022]
	Allowances motivate participants to concentrate.	• …this motivates people and even the concentration during the training is not compromised. [P 8006] • The benefit of receiving allowances is that the money helps you to care for yourself while you are doing the training, it can also help your family at the time when you are away from the family. It will help you to focus on the trainings and your home will not be disturbed due to lack of money. [P 8003]
	Need for equity when giving allowances.	• … So, the one working in NGO, because the training is hosted by the NGO, they only receive dinner allowance while those from government they receive dinner and daily subsistence allowance (DSA). [P 8005]
	Compensation is not the sole motivation	• I am not very concerned about the allowances because they just motivate us with the allowances…. what is paramount is the new knowledge that that we gain from the training. [P 8002]

Perceived Usefulness

Opportunity for Self-Guided Learning Through Digital Portions

Participants viewed the digital portion of blended learning as an opportunity for self-guided instruction. Some participants expected that a tablet-led training would allow them to learn ICT content better than facilitator-led training. This stemmed from a desire to review the training content at their own time and pace with no disturbances and to repeat difficult sections as needed. Participants expected they would keep the tablets at their facilities for future reference, an appealing feature.

Opportunity for Social Exchange Through Face-to-Face Interactions

Participants viewed opportunities to interact with facilitators and colleagues in-person as a central part of learning. Although participants were interested in individual learning, almost all also wished for social interaction during the learning process. Most participants were worried that working on the tablet would be like “talking to oneself” with no feedback mechanisms. They expected they would not get immediate feedback and clarifications. Participants viewed face-to-face sessions as essential for seeking clarification. In addition, participants felt face-to-face learning would allow them to practice and receive feedback. Practicing the skills with a facilitator and colleagues was considered useful because it would help them reinforce content and ultimately facilitate acquisition of knowledge and skills. It was also a time in which participants could receive support from their peers and facilitators on areas they needed to strengthen. In addition, participants considered the sharing of experiences with colleagues as integral.

Perceived Ease of Use

Need for Orientation to the Digital Technology

Orientation to the technology was highlighted as a key factor that would facilitate the ease of use of the blended learning implementation package. Conversely, some participants expressed fear about using a tablet for training without proper orientation due to not being well acquainted with the technology. To this end, participants suggested that they needed orientation on the use of the tablet as well as the applications that would

XSL•FO
RenderX

be used during the training since technological skills varied across participants.

Accessibility of Devices

Equipment access was a significant factor influencing perceived ease of use. Participants typically did not have personal tablets and assumed a tablet would be provided. Participants had several questions regarding the availability of the tablets within their facility, such as the duration when they would be available and whether they would be assigned to individuals or the facility. In addition, participants identified tablet storage, tablets running out of power, and poor network connectivity as key challenges that would affect tablet access. Participants feared that that the tablets were at risk of being lost at the facility, thus having appropriate tablet storage measures was suggested. Furthermore, participants were concerned that in seasons with frequent power blackouts, they would have challenges to charge the tablets if they ran out of power. Some participants stated that poor mobile network connectivity in some facilities could affect their ability to complete a tablet-based training.

Communication Around Training Schedules

Participants emphasized the need for open communication regarding training arrangements. They noted that training attendance can evoke envy among colleagues. To address this, participants suggested ensuring prior communication to all health care workers about the selection process and purpose of training, especially for tablet training, to minimize work disruption and misunderstandings.

Support for Logistical Arrangements

Time and location for both portions of the training were highlighted as key logistical considerations that would drive ease of use. Most participants considered having digital training at their place of work or a nearby location acceptable. They also were open to having portions of the training over the weekend or during workdays, with pros and cons for each of these different locations and times. Weekend training was considered desirable by some as it would not disrupt work schedules, but others noted it infringed on their personal commitments. Some participants explained that having the training on-site involved less effort for travel and would not disrupt the work schedule at the facility. However, to some individuals, having the training offsite was important because it created a mindset of higher excitement, motivation, and a sense of purpose for participants. They also noted that participants in offsite training would experience fewer distractions than in a work environment, allowing participants to fully focus on the training content.

Monetary Compensation

Provision of a training allowance (ie, monetary compensation) was noted as a contributor to ease of use of a blended training. The norm in the Malawian context is to receive an allowance to cover accommodation and meals when attending a centralized face-to-face training. Participants were asked to share their views regarding such allowances under a decentralized blended learning approach. The general perception was that they would still expect to receive an allowance. In fact, some participants indicated that they preferred offsite trainings, specifically because of the certainty of an allowance. They described how allowances act as a motivating factor that allows people to focus on training content and apply what they learn. In addition, allowances were perceived as important because they support the basic expenses for their household needs without detracting from their family's financial commitments while they were in the training. Participants preferred to receive a monetary allowance, rather than meals and accommodation, because it provided them with the possibility to save some of the money for other purposes. Participants voiced a desire for allowances to be disbursed equally across organization type, distance travelled, and rank. Although most participants noted allowances as important, many emphasized that this was not their sole ore even primary motivation for training.

Discussion

This study explored the anticipated acceptability of blended learning among lay HCWs in Malawi. Interest in self-guided learning was a key factor underlying perceived usefulness in digital learning. The desire for interaction was an important consideration for in-person learning. With respect to ease of use, blended learning could be facilitated through orientation, accessibility of digital devices, communication, and consideration of training logistics and compensation.

This study is among the first to examine perceived acceptability of a blended learning implementation package within the context of in-service training among lay HCWs in LMICs. Our study suggests that the use of technology could be acceptable among HCWs in this setting despite low digital literacy, as long as there is sufficient orientation and device access. The study revealed positive perspectives from HCWs on the use of the digital component as it would allow participants to learn the training content better at their own time, place, and pace. Consistently, studies that have used digital technology in tertiary education throughout health professions have reported positive perspectives [32-35], due largely to the ability to review content [16]. We posit that if blended learning proves acceptable in a country with one of the lowest digital literacy levels and many technology barriers, it can also be implemented in areas with higher digital literacy and fewer barriers.

Our findings support previous findings that suggest learners do not want to completely forego human interactions, which they perceive enhance learning [16,17,36-38]. While most participants welcomed the idea of digital ICT training, the lack of interaction was highlighted as a major hindrance to the use of technology alone. Hence, participants preferred to also have a face-to-face portion to resolve their unanswered questions and allow them to practice new skills. The desire for face-to-face encounters has also been highlighted in clinical training programs [32,39]. These findings around interest in both modalities reinforce the value of blended approaches [40] in which participants experience the flexibility of digital self-training and engagement and interactivity in face-to-face sessions [41]. We incorporated these findings into our package by having a set of digital modules focused on reviewing content, followed by face-to-face modules focused on practicing and receiving feedback.

XSL•FO
RenderX

Orientation was perceived as a key factor to facilitate the ease of use of digital technologies, such as tablets in this setting. Participants came from a setting with low technological literacy. According to the World Bank, mobile phone penetration in Malawi was at only at 60% in 2021, suggesting that digital skills are not as ubiquitous as they are in other settings. Therefore, for a blended learning implementation package to be successful, participants need basic navigation skills before initiation [14], as well as technical support to improve ease of use [42]. We incorporated these findings into our package training site supervisors on how to use the blended learning package. The site supervisors, in turn, were empowered to orient staff at their site and support those who experienced technology challenges.

Participants had different preferences regarding the time and location of the training, but they unanimously agreed on the desire for allowances, even with this new training modality. In addition, there was consensus around the importance of equity when distributing allowances. Allowances were felt to serve as motivating factors that allow participants to focus on training content. Such allowances are often provided to motivate participants [43-45]. However, an appeal of the blended learning approach is its potential to expand the skill set of HCWs in a more economical and less disruptive manner. By offering tablet-based training at or near their workplace, HCWs spend less time away from their other duties, and the overall costs could be reduced if no lodging or transportation expenses were incurred. Our study team took these findings under consideration in the development of a blended learning package that is now being tested in a randomized trial. In the trial, to address equity considerations, we offered the training to all lay health workers at each facility and provided a small allowance to recognize participants' time.

Our study has several limitations and next steps. The study was conducted before HCWs were exposed to blended learning and evaluated anticipated perceptions; actual perceptions may differ. This is something we will evaluate when we implement our trial. In addition, the study was conducted in a few facilities in 2 districts and perceptions may differ in other settings. However, we sampled from a broad range of districts and facility types and achieved saturation. These facilities are not meaningfully different from other facilities in these districts or the country overall, suggesting that our results may generalize broadly in Malawi.

In summary, our formative study showed that a blended learning implementation package may be an acceptable method of training HCWs in the Malawian setting if program developers attend to contextual factors that affect perceived usefulness and ease of use. Although this work offered immediate insights for our planned trial, we recommend additional research to assess acceptability of blended learning approaches in a range of LMIC settings among lay health workers.

Authors' Contributions

NER, TEM-M, TAT, MHK, VG, VM, SER, and KRS contributed to conceptualization. CK, TAT, MJC, and MM assisted with data collection. TEM-M, CJM, and MM assisted with the analysis. TEM-M, CJM, NER, TAT, MJC, MP, MJ-B, and DV contributed to the drafting of the paper and figures. All authors critically reviewed, edited, and approved the final paper.

Conflicts of Interest

None declared.

References

1. Malawi Ministry of Health. Malawi final report 2020-2021. 2022 Dec 1 URL: https://phia.icap.columbia.edu/malawi-final-report-2020-2021/ [accessed 2025-03-12]
2. World Health Organization. Guidelines on HIV Self-Testing and Partner Notification: Supplement to Consolidated Guidelines on HIV Testing Services: World Health Organization; 2016.
3. Rosenberg NE, Mtande TK, Saidi F, et al. Recruiting male partners for couple HIV testing and counselling in Malawi's option B+ programme: an unblinded randomised controlled trial. Lancet HIV 2015 Nov;2(11):e483-e491. [doi: 10.1016/S2352-3018(15)00182-4] [Medline: 26520928]
4. Kahabuka C, Plotkin M, Christensen A, et al. Addressing the First 90: a highly effective partner notification approach reaches previously undiagnosed sexual partners in Tanzania. AIDS Behav 2017 Aug;21(8):2551-2560. [doi: 10.1007/s10461-017-1750-5] [Medline: 28299518]
5. Tembo TA, Kim MH, Simon KR, et al. Enhancing an HIV index case testing passive referral model through a behavioural skills-building training for healthcare providers: a pre-/post-assessment in Mangochi District, Malawi. J Int AIDS Soc 2019 Jul;22 Suppl 3(Suppl Suppl 3):e25292. [doi: 10.1002/jia2.25292] [Medline: 31321917]
6. Picciano AG. Blended learning: implications for growth and access. OLJ 2006;10(3):95-102. [doi: 10.24059/olj.v10i3.1758]
7. Graham CR, Woodfield W, Harrison JB. A framework for institutional adoption and implementation of blended learning in higher education. Internet High Educ 2013 Jul;18:4-14. [doi: 10.1016/j.iheduc.2012.09.003]
8. Sakina R, Kulsum EM, Uyun AS. Integrating technologies in the new normal: a study of blended learning. ijqrm 2020;1(4):181-193 [FREE Full text] [doi: 10.46336/ijqrm.v1i4.81]
9. Means B, Toyama Y, Murphy R, Baki M. The effectiveness of online and blended learning: a meta-analysis of the empirical literature. Teachers College Record: The Voice of Scholarship in Education 2013 Mar;115(3):1-47. [doi: 10.1177/016146811311500307]

XSL•FO
RenderX

10. Karamizadeh Z, Zarifsanayei N, Faghihi AA, Mohammadi H, Habibi M. The study of effectiveness of blended learning approach for medical training courses. Iran Red Crescent Med J 2012 Jan;14(1):41-44. [Medline: 22737553]
11. Singh H. Building effective blended learning programs. In: Challenges and Opportunities for the Global Implementation of E-Learning Frameworks: IGI Global; 2021:15-23. [doi: 10.4018/978-1-7998-7607-6.ch002]
12. Skulmowski A, Rey GD. COVID-19 as an accelerator for digitalization at a German university: Establishing hybrid campuses in times of crisis. Hum Behav Emerg Technol 2020 Jul;2(3):212-216. [doi: 10.1002/hbe2.201] [Medline: 32838228]
13. Finlay MJ, Tinnion DJ, Simpson T. A virtual versus blended learning approach to higher education during the COVID-19 pandemic: The experiences of A sport and exercise science student cohort. J Hosp Leis Sport Tour Educ 2022 Jun;30:100363. [doi: 10.1016/j.jhlste.2021.100363] [Medline: 34867086]
14. Batubara HS, Riyanda AR, Rahmawati R, Ambiyar A, Samala AD. Implementasi model pembelajaran blended learning di masa pandemi Covid-19: meta-analisis [Article in Indonesian]. basicedu 2022;6(3):4629-4637 [FREE Full text] [doi: 10.31004/basicedu.v6i3.2816]
15. Mali D, Lim H. How do students perceive face-to-face/blended learning as a result of the Covid-19 pandemic? The International Journal of Management Education 2021 Nov;19(3):100552. [doi: 10.1016/j.ijme.2021.100552]
16. Baber H. Social interaction and effectiveness of the online learning – A moderating role of maintaining social distance during the pandemic COVID-19. AEDS 2022 Jan 3;11(1):159-171. [doi: 10.1108/AEDS-09-2020-0209]
17. Nacaroğlu O, Kızkapan O, Demir H. Middle school students' motivations and learning competencies in science: mediating role of digital literacy. Psychol Sch 2025. [doi: 10.1002/pits.23400]
18. Islam S, Baharun H, Muali C, et al. To boost students' motivation and achievement through blended learning. J Phys: Conf Ser 2018 Nov 1;1114:012046. [doi: 10.1088/1742-6596/1114/1/012046]
19. Ramachandran R, Kaur R, Roy A, Goel P, Deorari AK. Perceived usefulness of a blended learning approach for skills training among medical interns: a pilot study. BMC Med Educ 2024 Nov 25;24(1):1357. [doi: 10.1186/s12909-024-06379-x] [Medline: 39587590]
20. World Health Organization. WHO Guideline: Recommendations on Digital Interventions for Health System Strengthening: Web Supplement 2: Summary of Findings and GRADE Tables: World Health Organization; 2019.
21. Kenney J, Newcombe E. Adopting a blended learning approach: challenges encountered and lessons learned in an action research study. OLJ 2011;15(1):45-57. [doi: 10.24059/olj.v15i1.182]
22. Tembo TA, Mollan K, Simon K, et al. Does A blended learning implementation package enhance HIV index case testing in Malawi? A protocol for A cluster randomised controlled trial. BMJ Open 2024 Jan 22;14(1):e077706. [doi: 10.1136/bmjopen-2023-077706] [Medline: 38253452]
23. Hamilton AB, Finley EP. Qualitative methods in implementation research: An introduction. Psychiatry Res 2019 Oct;280:112516. [doi: 10.1016/j.psychres.2019.112516] [Medline: 31437661]
24. Elsey H, Khanal S, Manandhar S, et al. Understanding implementation and feasibility of tobacco cessation in routine primary care in Nepal: a mixed methods study. Implementation Sci 2015 Dec;11(1):1-12. [doi: 10.1186/s13012-016-0466-7]
25. Marangunić N, Granić A. Technology acceptance model: a literature review from 1986 to 2013. Univ Access Inf Soc 2015 Mar;14(1):81-95. [doi: 10.1007/s10209-014-0348-1]
26. Wallace LG, Sheetz SD. The adoption of software measures: A technology acceptance model (TAM) perspective. Information & Management 2014 Mar;51(2):249-259. [doi: 10.1016/j.im.2013.12.003]
27. Medical doctors (per 10,000).: World Health Organization; 2020. URL: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/medical-doctors-(per-10-000-population) [accessed 2023-04-12]
28. Flick RJ, Simon KR, Nyirenda R, et al. The HIV diagnostic assistant: early findings from a novel HIV testing cadre in Malawi. AIDS 2019 Jun 1;33(7):1215-1224. [doi: 10.1097/QAD.0000000000002159] [Medline: 31045942]
29. Network readiness index 2023-malawi. : Portulans Institute; 2023.
30. Clarke V, Braun V, Hayfield N. Thematic analysis. In: Qualitative Psychology: A Practical Guide to Research Methods: Sage Publications; 2015, Vol. 3:222-248.
31. Kelly RF, Mihm-Carmichael M, Hammond JA. Students' engagement in and perceptions of blended learning in a clinical module in a veterinary degree program. J Vet Med Educ 2021 Apr;48(2):181-195. [doi: 10.3138/jvme.2019-0018] [Medline: 31873057]
32. Bhat S, Madi˙ M. Blended learning in undergraduate dental education. Cumhuriyet Dent J 2022;25(1):91-93. [doi: 10.7126/cumudj.968711]
33. Nagaraj C, Yadurappa SB, Anantharaman LT, Ravindranath Y, Shankar N. Effectiveness of blended learning in radiological anatomy for first year undergraduate medical students. Surg Radiol Anat 2021 Apr;43(4):489-496. [doi: 10.1007/s00276-020-02572-x] [Medline: 32964270]
34. Bouarar AC, Mouloudj S, Umar TP, Mouloudj K. Antecedents of physicians' intentions to engage in digital volunteering work: an extended technology acceptance model (TAM) approach. JICA 2023 Dec 6;31(4):285-299. [doi: 10.1108/JICA-03-2023-0017]
35. Jiang W. Interdependence of roles, role rotation, and sense of community in an online course. Distance Education 2017 Jan 2;38(1):84-105. [doi: 10.1080/01587919.2017.1299564]

XSL•FO
RenderX

36. Wu JH, Tennyson RD, Hsia TL. A study of student satisfaction in A blended e-learning system environment. Comput Educ 2010 Aug;55(1):155-164. [doi: 10.1016/j.compedu.2009.12.012]
37. Jusoff K, Khodabandelou R. Preliminary study on the role of social presence in blended learning environment in higher education. IES 2009;2(4):79-83. [doi: 10.5539/ies.v2n4p79]
38. Koch J, Ramjan LM, Everett B, Maceri A, Bell K, Salamonson Y. "Sage on the stage or guide on the side"-Undergraduate nursing students' experiences and expectations of bioscience tutors in A blended learning curriculum: A qualitative study. J Clin Nurs 2020 Mar;29(5-6):863-871. [doi: 10.1111/jocn.15140] [Medline: 31855301]
39. Suwannaphisit S, Anusitviwat C, Tuntarattanapong P, Chuaychoosakoon C. Comparing the effectiveness of blended learning and traditional learning in an orthopedics course. Ann Med Surg (Lond) 2021 Dec;72:103037. [doi: 10.1016/j.amsu.2021.103037] [Medline: 34815858]
40. Kemp N, Grieve R. Face-to-face or face-to-screen? Undergraduates' opinions and test performance in classroom vs. online learning. Front Psychol 2014;5:1278. [doi: 10.3389/fpsyg.2014.01278] [Medline: 25429276]
41. Kassably C, Ashaal A, Tahan S, Bouzakhem N, Youssef S, Ghosn F. Identification of factors influencing educators perception of blended learning and triggering adoption decisions. Migrat Lett 2024;21(4):126-144. [doi: 10.59670/ml.v21iS6.7217]
42. Nkamleu GB, Kamgnia BD. Uses and abuses of per-diems in africa: a political economy of travel allowances. : African Development Bank Group; 2014.
43. Vian T, Miller C, Themba Z, Bukuluki P. Perceptions of per diems in the health sector: evidence and implications. Health Policy Plan 2013 May;28(3):237-246. [doi: 10.1093/heapol/czs056] [Medline: 22684639]
44. Samb OM, Essombe C, Ridde V. Meeting the challenges posed by per diem in development projects in southern countries: a scoping review. Global Health 2020 May 28;16(1):48. [doi: 10.1186/s12992-020-00571-6] [Medline: 32466774]
45. Derntl M, Motschnig-Pitrik R. Patterns for blended, person-centered learning: strategy, concepts, experiences, and evaluation. Presented at: Proceedings of the 2004 ACM Symposium on Applied Computing; Mar 14-17, 2004; Nicosia, Cyprus. [doi: 10.1145/967900.968087]

Abbreviations

HCW: health care worker
ICT: index case testing
IDI: in-depth interview
LMIC: low- and middle-income country
TAM: Technology Acceptance Model
UNAIDS: Joint United Nations Programme on HIV and AIDS
WHO: World Health Organization

Edited by A Mavragani; submitted 30.05.24; peer-reviewed by K Mouloudj; revised version received 20.02.25; accepted 24.02.25; published 07.04.25.

Please cite as:
Mbeya-Munkhondya TE, Meek CJ, Mphande M, Tembo TA, Chitani MJ, Jean-Baptiste M, Kumbuyo C, Vansia D, Simon KR, Rutstein SE, Mwapasa V, Go V, Kim MH, Rosenberg NE
Anticipated Acceptability of Blended Learning Among Lay Health Care Workers in Malawi: Qualitative Analysis Guided by the Technology Acceptance Model
JMIR Form Res 2025;9:e62741
URL: https://formative.jmir.org/2025/1/e62741
doi:10.2196/62741

© Tiwonge E Mbeya-Munkhondya, Caroline J Meek, Mtisunge Mphande, Tapiwa A Tembo, Mike J Chitani, Milenka Jean-Baptiste, Caroline Kumbayo, Dhrutika Vansia, Katherine R Simon, Sarah E Rutstein, Victor Mwapasa, Vivian Go, Maria H Kim, Nora E Rosenberg. Originally published in JMIR Formative Research (https://formative.jmir.org), 7.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Impact of a Human Papillomavirus Facebook-Based Intervention (#HPVVaxTalks) Among Young Black (African American and Sub-Saharan African Immigrants) Adults: Pilot Pre- and Poststudy

Adebola Adegboyega[1], PhD; Amanda Wiggins[1], PhD; Abubakari Wuni[1], MS; Melinda Ickes[2], PhD

[1]College of Nursing, University of Kentucky, Lexington, KY, United States

[2]Department of Kinesiology and Health Promotion, College of Education, University of Kentucky, Lexington, KY, United States

Corresponding Author:
Adebola Adegboyega, PhD
College of Nursing, University of Kentucky, Lexington, KY, United States

Abstract

Background: Despite the availability of prophylactic human papillomavirus (HPV) vaccines, uptake remains suboptimal among young Black adults. Social media is a platform for the dissemination of health information and can be used to promote HPV vaccination among young Black adults.

Objective: This study aimed to assess the impact of a Facebook-based intervention (#HPVVaxTalks), which consisted of 40 posts over 8 weeks in improving cognitive outcomes, reducing vaccine hesitancy, and increasing vaccine intention, and uptake among young Black adults aged 18-26 years.

Methods: A pilot 1-group pre- and poststudy was conducted among 43 young Black adults who engaged in an 8-week Facebook intervention (#HPVVaxTalks). #HPVVaxTalks was developed in collaboration with a youth community advisory committee. Participants were actively recruited by research staff from community settings using flyers, and flyers were posted in public places in communities. Eligible participants were screened for eligibility and consented prior to study participation. Participants completed baseline surveys and were added to a Facebook page created for the study to receive intervention posts. Participants completed pre- and postdata on HPV knowledge, HPV vaccine knowledge, vaccine hesitancy, and vaccine uptake via REDCap (Research Electronic Data Capture) surveys distributed by email. Participants' satisfaction with the intervention was collected via individual interviews. Data were analyzed using 2-tailed paired *t* tests and repeated measures analysis

Results: Overall, 32 of the 43 (74%) participants completed the follow-up survey, and of the 23 participants who reported not having ever received the vaccine at baseline, 7 (30%) reported receiving the vaccine at follow-up. Participants demonstrated significant improvements in HPV knowledge and receiving the vaccine at follow-up. Participants demonstrated significant improvements in HPV knowledge (pre: mean 7.3, SD 4.2 and post: mean 11.1, SD 4.3; *P*=.004) and HPV vaccine knowledge (pre: mean 2.8, SD 2.5 and post: mean 4.7, SD 2.2; *P*=.003) and reduction in vaccine hesitancy (pre: mean 28.3, SD 4.2 and post: mean 29.9, SD 3.6; *P*=.007) after the intervention. However, there were no significant changes in other outcomes. Feedback from open-ended questions and qualitative interviews highlighted participants' satisfaction with the intervention and its role in increasing HPV and HPV vaccine awareness.

Conclusions: The findings from this study underscore the potential of social media platforms for health promotion among underrepresented populations and the importance of advocating for culturally appropriate interventions to improve HPV vaccination rates and reduce disparities.

(JMIR Form Res 2025;9:e69609) doi:10.2196/69609

KEYWORDS

human papillomavirus vaccination; HPV vaccination; Facebook-based intervention; young Black adults; social media

Introduction

Background

Human papillomavirus (HPV) vaccination provides primary prevention against HPV-related diseases [1,2]. However, HPV vaccine uptake remains low for female participants and lower for male participants. Within 12 years of HPV vaccine introduction, infections with the 4 HPV types prevented by Gardasil decreased to 88% among 14‐ to 19-year-old female participants and to 81% among 20‐ to 24-year-old female participants in the United States [3]. According to the Centers

XSL•FO
RenderX

for Disease Control and Prevention, only 21.5% of adults aged 18 - 26 years received the recommended number of doses of the HPV vaccine [4]. Completion rates for the HPV vaccine are low in Black female participants (44.7% vs 57.9% White female participants) and lower among Black male participants (29.4% vs 26.6% among White male participants) aged 18 - 26 years [5]. HPV vaccination is recommended for everyone through age 26 years if not adequately vaccinated when younger. Hence, Black young adults are a priority population to improve HPV vaccine uptake, given the low HPV vaccine uptake among this group [6,7].

In general, one of the most common barriers to HPV vaccination uptake is the lack of awareness of HPV and HPV infections. Research shows that sub-Saharan African immigrants have even less knowledge of HPV and HPV vaccination compared to White and Black American–born individuals [8-10]. The lack of knowledge results in misconceptions and the thinking that the HPV vaccine is unnecessary because individuals do not perceive themselves at risk of HPV infection [11]. In addition, the potential cost of the HPV vaccine series completion, perceptions of vaccine safety, and worry regarding anticipated side effects are barriers to the uptake of vaccine uptake for young adults [12]. In addition, system-level barriers such as limited access to vaccines, vaccine documentation requirements, and lack of strong provider recommendations may lead to missed vaccination opportunities [13-15]. Additionally, the African American population has been participants of historically unethical research, which may contribute to low vaccine uptake, medical mistrust, and misconceptions about vaccines [8,16,17].

Despite the high prevalence and complications that can occur from unresolved HPV infections, gaps in HPV-related cognition and HPV-related disparities exist for Black individuals. Catch-up HPV vaccination is recommended for young adults (up through age 26 years) who have not been previously vaccinated. However, few interventions have focused on Black young adults (18 - 26 years), much of the research has explored parental decision-making and system-level interventions for HPV vaccination among adolescents [18-22]. Transition to young adulthood is a critical period to institute health-promoting behaviors [23]. As young adults transition to adulthood, they begin to explore their sexuality, including initiating sexual activity for the first time or engaging in riskier sexual behaviors than in the past. These behaviors increase the risk of acquiring sexually transmitted infections including HPV, providing a strong justification for health-promoting behavior such as catch-up HPV vaccination [24]. As Black young adults assume health care decision-making responsibility and independence, they may benefit from interventions to promote cognitive outcomes (HPV and HPV vaccine-related knowledge, benefits, severity, self-efficacy, risk perception), vaccine intention, and catch-up HPV vaccination.

Social media is increasingly becoming the preferred platform for the dissemination of health information and can be used to raise awareness about HPV and HPV vaccinations, therefore increasing HPV vaccination rates [25]. Social media platforms are multifaceted, and their features can be used to serve multiple purposes and reach a wide audience. Social media platforms include video and photo posting, blogs, polls, and other functions [26,27]. In the United States, Facebook is the single most popular social network with over 2 billion daily users, and 73% of Black adults have ever used Facebook [28,29]. In addition, 83% of adult social media users report visiting Facebook in a typical week, and on average, young adults aged 18 to 24 years spend about 11 minutes every day on Facebook [30].

Recent studies have used Facebook-based interventions to promote HPV vaccination among young adults [31-33]. For example, a technology-mediated intervention including private Facebook posts improved knowledge among undergraduate college students in the HPV vaccination intervention relative to those in the control condition [32]. Similarly, a sponsored advertisement campaign on Facebook targeted at young female participants aged 18 - 26 years found that participants engaged with and were receptive to watching narrative-based health information videos on Facebook. Higher video engagement was associated with stronger intentions to talk with a health care professional, talk with friends or family, and vaccinate against HPV [31]. Despite the growing popularity of Facebook-based interventions, few have targeted young Black adults including sub-Saharan African immigrants. To fill this gap, we sought to leverage Facebook for an HPV vaccination intervention among young Black adults—a priority catch-up population with the highest burden of HPV. Thus, the objective of this study was to examine the impact of a Facebook-based intervention (#HPVVaxTalks) on HPV vaccination uptake and secondary outcomes among Black (African American and sub-Saharan African immigrants) young adults (18 - 26 years) of an 8-week Facebook-based intervention. We hypothesize that young Black adults who complete the intervention will have improvements in HPV knowledge and HPV vaccine knowledge and a reduction in vaccine hesitancy after the intervention.

Theoretical Framework

The theory of planned behavior states that the intention to perform a behavior is the strongest predictor of whether an individual engages in certain behaviors [34]. The theory of planned behavior has been used to understand and predict HPV vaccination rates by examining what constructs most correlate with people's intentions to receive HPV vaccine. For example, in a sample of college students, more positive attitudes and increased subjective norms were strongly correlated with increased intention to receive complete doses of the HPV vaccine [35]. Knowledge and health beliefs are antecedents of intention and subsequent behavior. Hence, the theory of planned behavior provides a framework to understand individual actions by identifying beliefs relevant to individuals and groups and to develop interventions to change beliefs [34].

Methods

Study Designs and Participants

This was a 1-group pre- and post-Facebook–based intervention pilot study conducted in a southeastern state. We used a mixed methods approach by administering a pre- and postintervention survey to the participants as well as conducting follow-up interviews with a subset of participants.

XSL•FO
RenderX

Participants for the study were young Black (African American and sub-Saharan African immigrants) adults between 18 and 26 years, given the age for HPV vaccination catch-up through age 26 years. We used the STROBE (Strengthening the Reporting of Observation Studies in Epidemiology) checklist to guide the reporting of the study (Checklist 1).

Intervention

A culturally appropriate Facebook-based intervention (#HPVVaxTalks) was developed in collaboration with a youth community advisory committee [36,37]. The research team created a library of evidenced-based Facebook posts, and a youth community advisory board reviewed the preliminary version of #HPVVaxTalks intervention materials and provided critiques and suggestions for refinement. The advisory board provided recommendations and input on different iterations of the posts for content preferences, priorities, user-friendliness, relevance, and cultural appropriateness [36]. To ensure posts were culturally appropriate, we followed strategies from Kreuter et al [38] including evidential, peripheral, linguistic, constituent-involving, and sociocultural strategies. Engagement with a youth community advisory board fostered the opportunity to discuss priorities reflecting the target group's interest, ensure the appeal and relevance of the intervention, and draw on the experience of the target populations. In addition, we conducted a content validation of the intervention by obtaining experts' feedback and participants' satisfaction on the components of the #HPVVaxTalks for content usability, acceptability, relevance, and visual appeal to inform further refinements [37]. The intervention addressed general information on HPV, the importance of HPV vaccination, vaccine hesitancy, self-efficacy, HPV vaccination as a cancer prevention strategy, risk perception, perceived behavioral control, and subjective norm. Intervention contents were mapped to constructs of the theory of planned behavior, and important concepts informed by the literature and youth advisory board. The Facebook posts contained relevant images, graphics, memes, short texts, and information on where to receive the HPV vaccine in the community (see Figures 1 and 2 for sample posts). We included embedded links to information from lead organizations (eg, American Cancer Society, American Sexual Health Association, and Centers for Disease Control and Prevention). The intervention consisted of 40 messages posted over 8 weeks, with 4 to 5 messages posted each week. One "secret" and "closed" Facebook group page solely for this study was created. Facebook was chosen as the platform for delivering HPV vaccination messages due to its features that allow for the creation of private groups. These groups enable only invited participants to see posts and engage in discussions, ensuring a level of privacy and confidentiality that can encourage open discussion [33]. Eligible participants were screened and consented, they received a link to complete the baseline survey and were invited to join #HPVVaxTalks Facebook group. Trained research staff posted intervention posts as scheduled (4 to 5 messages posted each week for 8 weeks) to promote engagement but not burden the participants with too much information. Enrolled participants were encouraged to turn on Facebook notification to know when posts are posted, view posts regularly, comment, like, and ask questions. The Facebook page was monitored throughout the intervention by the research team.

Figure 1. This is a sample of Facebook graphic posted as one of the messages for HPV vaccination promotion among young Black adults. Courtesy of National Foundation for Infectious Diseases [39].

XSL•FO
RenderX

Figure 2. This is a sample of Facebook graphic posted as one of the messages for HPV vaccination promotion among young Black adults [39]. HPV: human papillomavirus.

Recruitments

Black young adults aged 18 - 26 years were recruited throughout a southeastern state between April 2023 and June 2024 using approved study flyers, snowballing, and word of mouth. In addition, research staff carried out active recruitment throughout community organizations that have a high client base of young adults. Study flyers with research team contact information were posted in public areas and community settings throughout the state. Flyers described the project and asked interested participants to contact the study team to learn more about the study. Interested participants were provided with additional information about the study and screened for eligibility. Study eligibility included the following: self-identification as African American and sub-Saharan African immigrants, age 18 - 26 years, fluency in the English language, access to the internet, a Facebook account, and willingness to participate.

Ethical Considerations

Prior to the start of the study, all study procedures and materials were reviewed and approved by the University of Kentucky Institutional Review Board (protocol 76589). Participation was voluntary, and all study participants provided written informed consent prior to participation in any study activities. Confidentiality of participant's information was maintained at all times. Patient's data were coded with a unique subject identifier, and all data deidentified prior to analysis. Participants were compensated with US $30 gift card from a retail store for completing the baseline survey and US $20 for postintervention surveys.

Quantitative Data Collection

All research activities were completed remotely. Participants received REDCap (Research Electronic Data Capture; Vanderbilt University) links via emails to complete baseline and postintervention surveys. Quantitative data were collected using self-administered REDCap-based surveys at baseline and after the intervention (after 8 weeks). After the intervention, we repeated the items from the baseline survey and added open-ended questions to assess what participants liked about the intervention and ways for improvement. In addition, we assessed intervention satisfaction and asked if participants had received a dose of the HPV vaccination.

Measures

Overview

Participants completed a self-administered survey including sociodemographic characteristics, HPV vaccination history, HPV knowledge items, HPV vaccine knowledge items, vaccine hesitancy measures, risk perception items, HPV severity, HPV vaccination benefits, and HPV vaccination self-efficacy. Valid and reliable measures were adapted from the literature [40-43].

XSL•FO
RenderX

Sociodemographic Characteristics

Sociodemographic information collected via self-report included age, sex, country of origin, education level, income, and health insurance coverage.

HPV Vaccination Uptake

HPV vaccination uptake was assessed among those who reported not being vaccinated at baseline. In the follow-up survey, participants were asked if they had received at least 1 dose of the HPV vaccine since enrolling in the study (yes or no).

HPV Knowledge

HPV knowledge items included 16 items assessing knowledge of HPV (including transmission, consequences, and risk factors). Response options were true or false or do not know, with "do not know" scored as incorrect [41,42]. A summative score was calculated with a potential range of 0 - 16. The scale has high reliability (Cronbach α=0.84) among diverse populations [42], and Cronbach α for this study was 0.91.

HPV Vaccine Knowledge

HPV vaccine knowledge was assessed with an 8-item scale regarding the protection offered by HPV vaccines [41]. A summative score was calculated with a potential range of 0 - 8. Cronbach α for this sample was 0.84.

Vaccine Hesitancy

Vaccine hesitancy was measured with a 9-item scale that captures general attitudes to vaccination [43], sample items include "the HPV vaccine is effective" and "the HPV vaccine is important for my health." Responses options followed a 4-point Likert scale ranging from 1=strongly disagree to 4=strongly agree [43]. A summative score was calculated with a potential range of 9 - 36, with lower scores indicating greater hesitancy. Cronbach α for this sample was 0.80.

HPV-Related Risk Perception

HPV-related risk perception was assessed with the item, "If I don't get vaccinated for HPV, I think my chances of getting [HPV infection; HPV-related cancer; genital warts] sometime in the future would be," with responses of 1=almost zero, 2=small, 3=moderate, 4=large, and 5=almost certain [40].

HPV Severity, Benefits, and Self-Efficacy

These constructs were assessed with items adapted from the literature [40]. HPV severity was assessed with 4 items (eg, being infected with HPV would have major consequences on my life), HPV benefits were assessed with 3 items (eg, getting vaccinated for HPV will help protect me from HPV infection), and self-efficacy was assessed with 3 items (eg, I feel confident in my ability to get vaccinated for HPV, even if it is expensive). Responses options for each item followed a 5-point Likert scale ranging from 1=strongly disagree to 5=strongly agree. Summative scores were calculated for each construct, with potential ranges of 4 - 20, 3 - 15, and 3 - 15 for severity, benefits, and self-efficacy, respectively. Cronbach α for all scales exceeded 0.77.

HPV Vaccine Intention

Vaccine intention was assessed with an item, "I intend to get the HPV vaccine in the next 3 months?" Responses followed a 5-point Likert scale ranging from 1=strongly disagree to 5=strongly agree.

Intervention Satisfaction

Intervention satisfaction was assessed with items to measure usability (6 items; Cronbach α=0.93), acceptability (5 items; Cronbach α=0.92), and relevance of HPV posts or messages on Facebook (5 items; Cronbach α=0.92). Responses to each item followed a 5-point Likert scale ranging from 1=strongly disagree to 5=strongly agree, and summative scores were calculated for each construct separately [44,45].

Data Analysis: Quantitative Data Analysis

Sociodemographic characteristics among those who completed the follow-up survey (completers) and those who did not (noncompleters) were examined using the 2-tailed *t* test, chi-square test of association, or Mann-Whitney *U* test, as appropriate.

Descriptive statistics were used to analyze vaccination uptake. Among those who had not received the HPV vaccine at baseline, unadjusted changes in outcomes (HPV-related knowledge, HPV vaccine knowledge, vaccine hesitancy, risk perception, severity, vaccination benefits, self-efficacy, and intention and vaccination uptake) were evaluated using the paired 2-tailed *t* test. These changes were also evaluated using repeated measures analysis, adjusting for age, sex, and ethnicity. The repeated measures analysis is exploratory due to the small size. Intervention satisfaction, specifically usability, acceptability, and relevance of the messages, was summarized descriptively using means and SDs. All analysis was conducted using SAS (version 9.4; SAS Institute Inc), with an α level of .05 throughout.

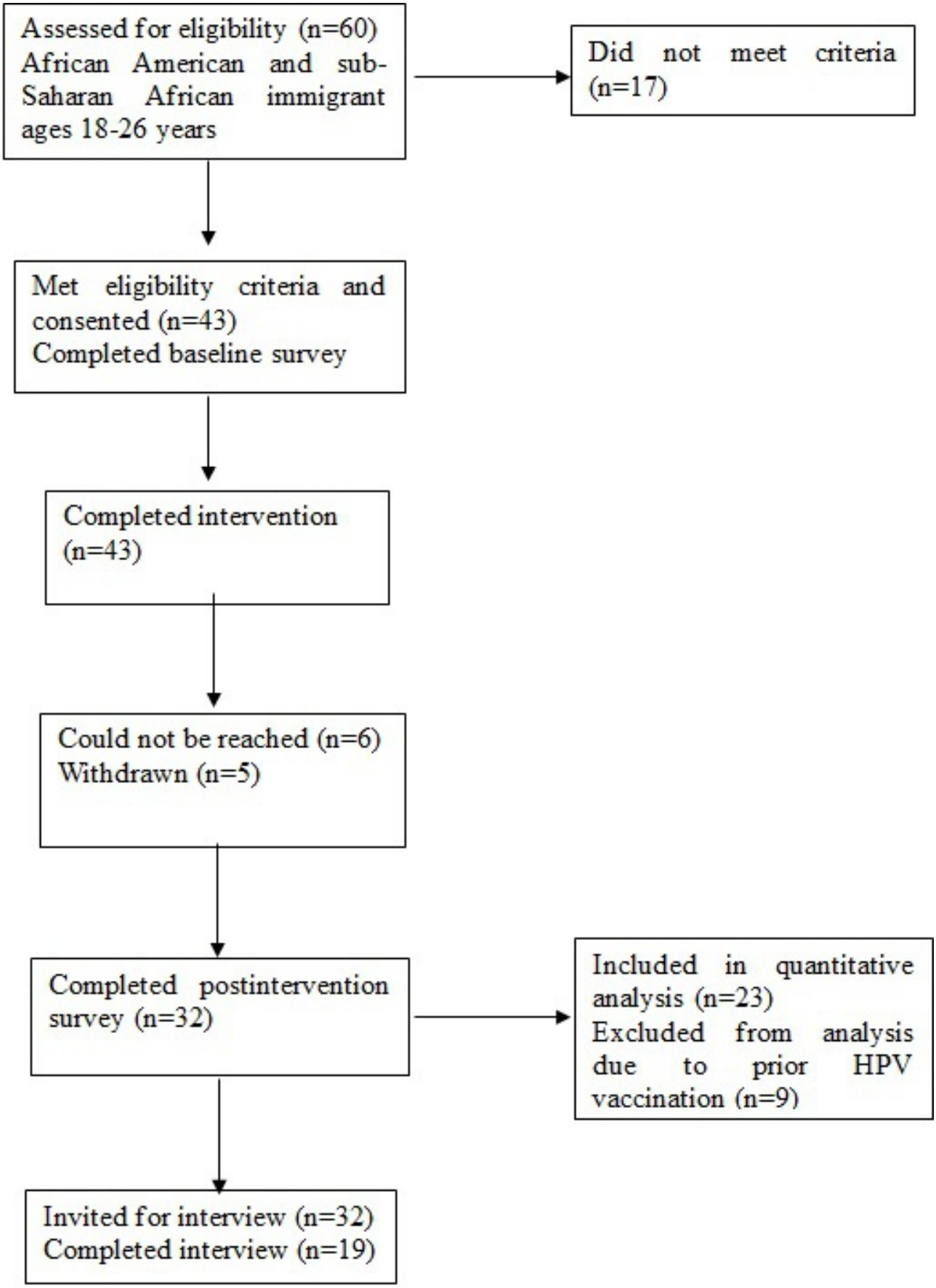
Results

Quantitative Findings: Sample Characteristics

In total, 43 young Black adults enrolled in the study, 32 of the 43 (74%) participants completed the follow-up survey, and 9 participants were ineligible for HPV vaccination because of prior HPV vaccination or other reasons not identified at the time of recruitment. Flow participant diagram through pre- and postintervention is shown in Figure 3. In the baseline sample of 43 participants, the average age was 21.7 (SD 2.7 years; Table 1). The majority were female (32/43, 74%), and more than half were African-born (25/43, 58%). Participants were well-educated, with nearly half having a college degree or beyond (20/33, 46%), and over three-quarters of participants had health insurance coverage (33/43, 77%). In comparing the completers and noncompleters of the follow-up survey, a higher proportion of completers were African-born (*P*=.002), had higher educational attainment (*P*=.02), and had higher income levels (*P*=.04). There was no difference in age, sex, health insurance coverage, or receipt of the HPV vaccine (at baseline) between completers and noncompleters.

XSL•FO
RenderX

Figure 3. Flow participant diagram of the progress through the phases of a pre and post 1-group intervention (enrollment, baseline, completed intervention, follow-up, and data analysis). HPV: human papillomavirus.



XSL•FO
RenderX

Table . Descriptive summary of sociodemographic characteristics of young adults (18 - 26 years).

	Full sample (N=43)	Completers (n=32)	Noncompleters (n=11)	*P* value
Age (years), mean (SD)	21.7 (2.7)	22.1 (2.9)	20.5 (1.6)	.09[a]
Sex, n (%)				.88[b]
Male	11 (26)	8 (25)	3 (27)	
Female	32 (74)	24 (75)	8 (73)	
Country of origin, n (%)				002[b]
US-born	18 (42)	9 (28)	9 (82)	
African-born	25 (58)	23 (72)	2 (18)	
Education, n (%)				.02[c]
High school or below	4 (9)	2 (6)	2 (18)	
Some college	19 (44)	12 (38)	7 (64)	
College graduate	13 (30)	11 (34)	2 (18)	
Master or higher	7 (16)	7 (22)	0 (0)	
Household income (US $), n (%)				.04[c]
Less than $25,000	14 (33)	11 (34)	3 (27)	
$25,000-$49,999	12 (28)	12 (38)	0 (0)	
$50,000-$79,999	7 (16)	5 (16)	2 (18)	
$80,000 or more	10 (23)	4 (13)	6 (55)	
Health insurance coverage, n (%)				.23[b]
Yes	33 (77)	26 (81)	7 (64)	
No	10 (23)	6 (18)	4 (36)	
Ever recieved the human papillomavirus vaccine, n (%)				.96[b]
Yes	12 (28)	9 (28)	3 (27)	
No or do not know	31 (72)	23 (72)	8 (73)	

[a]Two-tailed *t* test.

[b]Chi-square test of association.

[c]Mann-Whitney *U* test.

Among the 23 participants who completed the follow-up survey that reported not having ever received a dose of the HPV vaccine at baseline, 7 (30%) reported receipt of at least 1 dose of the HPV vaccine at follow-up. Among participants (n=23), there was a significant increase in HPV knowledge (*P*=.01) and HPV vaccine knowledge (*P*=.004; Table 2) in the unadjusted analysis. Based on a potential range of 0 - 16, average HPV knowledge scores increased from 7.3 (SD 4.2) to 11.1 (SD 4.3) from baseline to follow-up. For HPV vaccine knowledge, based on a potential range of 0 - 8, average scores increased from 2.8 (SD 2.5) on the baseline survey to 4.7 (SD 2.2) on the follow-up. Both of these changes remained statistically significant in the models adjusting for age, sex, and ethnicity (*P*=.004 and *P*=.003, respectively). There was also a significant change in HPV vaccine hesitancy. Scores on this scale increased from baseline (mean 28.3, SD 4.2) to follow-up (mean 29.9, SD 3.6; *P*=.02), which indicates a decrease in vaccine hesitancy. This change remained significant in the adjusted analysis (*P*=.007).

XSL•FO
RenderX

Table . Pre- and postchanges from baseline to follow-up among young adults (18 - 26 years) who have not received the human papillomavirus (HPV) vaccine at baseline (n=23).

	Potential range	Baseline, mean (SD)	Postintervention, mean (SD)	Unadjusted *P* value	Adjusted *P* value[a]
HPV knowledge	0 - 16	7.3 (4.2)	11.1 (4.3)	.01	.004
HPV vaccine knowledge	0 - 8	2.8 (2.5)	4.7 (2.2)	.004	.003
HPV vaccine hesitancy	9 - 36	28.3 (4.2)	29.9 (3.6)	.02	.007
HPV-related risk perception	1 - 5	2.4 (1.1)	2.7 (1.1)	.44	.58
HPV severity	4 - 20	16.5 (2.1)	16.5 (3.7)	.96	.93
HPV vaccination benefits	3 - 15	12.3 (1.8)	13.1 (2.2)	.19	.12
HPV vaccination self-efficacy	3 - 15	10.7 (2.1)	11.4 (2.6)	.34	.13
HPV vaccine intention	1 - 5	2.7 (0.8)	2.6 (1.1)	.88	.90

[a]Models adjusted for age, ethnicity, and sex.

There were no significant changes in HPV severity, vaccination benefits, vaccination self-efficacy, or intention among those who had not received the HPV vaccine at baseline. Among the completers of the intervention, baseline vaccine hesitancy and risk perception scores were near the midpoint of the scales, while HPV severity, vaccination benefits, and vaccination self-efficacy were relatively high.

For intervention satisfaction, participants rated the usability high (eg, information was important, credible, helpful, etc) with an average score of 24.8 (SD 3.4) of a potential maximum of 30. Similarly, based on potential ranges of 0 - 25, overall intervention acceptability (mean 20.0, SD 3.3; potential range 5 - 25) and relevance of the posts (mean 20.1, SD 2.7) were high. The usability, acceptability, and relevance of the posts or messages were supported further by one-on-one interviews (n=19) and open-ended questions (n=32).

Qualitative Data Collection

All participants who completed the postintervention assessments were invited for a follow-up interview via email. Follow-up one-on-one interviews were conducted with 19 participants to assess intervention satisfaction at 8 weeks (postintervention). The interviews were conducted on Zoom (Zoom Video Communications) by the first author (AA), a PhD-level qualitatively trained Black female researcher. Each interview began with an introduction, and the interview was guided by a semistructured interview guide developed by the research team. Each interview lasted 15 to 25 minutes. The interview questions were centered on how well the posts resonated with participants, how the intervention could be improved, and how the information learned from the intervention would be used. In addition, we asked if participants had received at least 1 dose of the HPV vaccination or scheduled an appointment. Participants were compensated with US $10 for interviews. Interviews were audio-recorded, transcribed verbatim, and deidentified for analysis.

Qualitative Data Analysis

Data from the one-on-one interviews and open-ended questions were analyzed by 2 qualitative trained research team members (Adegboyega Adebola and Wuni Abubakari). The researchers (AA and AW) read the transcripts and independently carried out a content analysis of the data. We did not use any qualitative software. Transcripts were coded line-by-line by the 2 researchers. Coding involved aggregating the data text into small categories of information and assigning a label to the code [46]. We used an iterative procedure to refine the codes and identify emergent themes until reaching saturation. Any discrepancies between results were discussed and resolved with the entire research team.

Qualitative Findings

In total, 19 participants completed the interviews, 12 (63%) were female participants. Two main themes emerged from the data: (1) feedback and recommendation on the intervention and (2) perceived impact of the intervention on vaccination decision.

Theme 1: Feedback and Recommendation on the Intervention

Participants expressed strong appreciation for the visual content used in the intervention, particularly memes and graphics. Participants found these visuals to be effective in delivering clear and engaging messages about HPV and vaccination. One participant stated, “the use of memes and graphics made the information clear and easy to understand.” Another participant noted, “visuals really caught my attention and made me more interested in learning about HPV.” The participants positively noted the notification system for new posts and the use of incentives to encourage engagement. One participant noted that “I like the fact that I get notified of when new information is posted.” Opportunities for improvements suggested by participants included the promotion of more participant engagement, the use of additional platforms, and shorter intervention duration. One participant suggested, “You should make it into a group that everybody is active, maybe like text

XSL•FO
RenderX

messaging, SMS, WhatsApp or something like a smaller group." Another participant noted, "I think I will like it shorter for half of the time it was done, like four weeks."

Theme 2: Perceived Impact of the Intervention on Vaccination Decision

Some participants who took the vaccine after the intervention reported that it significantly influenced their decision to vaccinate against HPV. Some participants indicated that the information provided through the intervention helped clarify misconceptions and increased their confidence in the vaccine's necessity. One participant noted, "After learning more from the intervention, I decided to get vaccinated." Another noted, "The intervention gave me the information I needed to make an informed decision about HPV vaccination."

Discussion

Principal Findings

This study was a pilot 1-group pre- and post-Facebook–based intervention (#HPVVaxTalks) study among 43 young Black adults in a southeastern state. Participants in the 8-week intervention showed significant improvements in HPV knowledge and HPV vaccine knowledge and a reduction in vaccine hesitancy after the intervention. Participants were satisfied with the intervention, and it was well-received. Participants did not show significant changes in HPV severity, vaccination benefits, vaccination self-efficacy, or intention among those who had not received the HPV vaccine at baseline.

The intervention led to HPV vaccination uptake, with 7 of 23 (30%) reporting that they have received at least 1 dose of the HPV vaccine at 8 weeks follow-up. This is a noteworthy result, given that the intervention targeted a high-risk group. This finding is in line with a study where 30% of Korean-American female participants aged 21 to 29 years who participated in a mobile health intervention received the first dose of the HPV vaccine after the intervention of a text messaging intervention to promote HPV vaccination [47]. Our finding highlights the impact of culturally appropriate intervention in promoting HPV vaccination among previously unvaccinated individuals. However, the low vaccination rates and no significant changes in HPV vaccine intention reported from our study show that this intervention had limited influence on HPV vaccination outcomes. According to the theory of planned behavior, the intention to perform a behavior is the strongest predictor of whether an individual engages in certain behaviors. The intention to engage in vaccination is determined by a person's beliefs, attitudes, and subjective norms within a specific context [34]. Hence, future interventions should use multicomponent interventions to address young adults' HPV vaccine–related beliefs, attitudes, and subjective norms. Multicomponent interventions should include effective patient-provider communication and HPV vaccine recommendations as a cancer prevention tool. In addition, there is a need for comprehensive social media campaigns to address myths and misinformation related to HPV targeted at young adults, given their wide reach and influence among young adults.

This study found a significant increase in HPV knowledge and HPV vaccine knowledge in the pre- and postintervention. The findings from this study are congruent with a social media intervention study where there was a statistically significant increase in knowledge after the intervention [48]. Furthermore, qualitative postintervention interviews with study participants supported the findings that our social media intervention increased their awareness and knowledge of HPV, particularly regarding transmission risks and preventive measures. This is in line with a similar study on mobile health interventions among unvaccinated youths, which showed improvement in HPV-related knowledge, vaccination intention, or vaccine initiation in a postintervention [49]. However, the findings from our study are inconsistent with a Twitter campaign that reported no significant changes in HPV knowledge after the campaign [50]. However, knowledge alone is insufficient for increasing HPV vaccination rates, as other factors beyond knowledge deficiency influence HPV vaccination decisions.

Our study found a significant reduction in HPV vaccine hesitancy. This finding supports the use of a Facebook-based intervention to decrease HPV vaccine hesitancy among young Black adults. A large-scale study across 67 countries showed that vaccine hesitancy can partly be attributed to a lack of confidence in vaccine safety, perceptions that vaccines do not work, distrust of information, perceived low risks of vaccine-preventable diseases, as well as a lack of trust in health care providers, authorities, and pharmaceutical companies [51]. Research has shown that interventions that support risk communication, community engagement, use of trained vaccine champions, tailored communication campaigns, and increased confidence in vaccines tend to reduce HPV vaccine hesitancy [52]. For this pilot study, we used a community-engaged approach by engaging with a youth advisory board for the development of #HPVVaxTalks intervention, which improved its cultural appropriateness for young Black adults. Future studies could use young adults as vaccine champions to deliver training and advocacy tailored to their communities to positively impact social norms [52].

Our study found no significant differences in HPV self-efficacy, intention to vaccinate, perceived severity, and perceived vaccination benefits. This may be because of the small sample size and relatively high baseline values for these variables. Future research with large sample size and longitudinal approach is warranted to test whether a Facebook-based intervention is effective in promoting significant change in HPV self-efficacy, intention to vaccinate, perceived severity, perceived vaccination benefits, and HPV vaccination uptake.

The intervention piloted in this study was well-received, with participants rating its usability, acceptability, and relevance highly. This study reiterates the importance of including the target population in formative research to promote usability, acceptability, and relevance to the target population. These findings are supported by research that reports that young adults are open to using social media to seek out health information and learn more about preventative health behaviors [32,48,50]. Using social media platforms could be an avenue and platform for delivering crucial information to reach young Black adults on a larger scale.

XSL•FO
RenderX

Participants reported that the intervention positively influenced their perception of the HPV vaccine's necessity and encouraged them to consider vaccination. These findings are supported by research that reports that young Black adults can benefit from social media interventions and seek out correct health information from social media platforms. Leader et al [31] showed that young adults were receptive to watching narrative-based health information videos on social media. Based on suggested opportunities for improvement, future research may explore shorter intervention duration to prevent attrition and use other social media avenues that may be relevant for health promotion and engagement for African American and sub-Saharan African immigrant Black young adults.

Implications for Practice or Research

The intervention's effect on HPV-related knowledge, vaccine knowledge, and vaccine hesitancy demonstrates its efficacy. Researchers can extend these findings by undertaking larger-scale research with varied groups and using other social media platforms to demonstrate the intervention's reproducibility and generalizability. Follow-up studies could determine whether knowledge increases, and behavioral improvements are sustainable over time. To determine if intervention effects persist beyond study completion and if participants complete the HPV vaccination series, longitudinal research is needed to monitor participants' HPV vaccination series completion. In addition, further research can determine which parts of the intervention (eg, specific message kinds and frequency of exposure) are most important to its effectiveness.

Tailored and community-driven social media interventions can be implemented to increase overall community uptake and support of the intervention and may improve health outcomes among minoritized and underserved populations. Health care professionals can use social media platforms to share evidence-based information regarding HPV and HPV vaccination. The intervention's success in increasing knowledge and vaccination shows that such initiatives could be beneficial on a larger scale. Messages can be tailored to young Black adults who are active on social media while taking into consideration cultural contexts and incorporating strategies to increase engagement and the effectiveness of future interventions. The usability, acceptability, and relevance of the intervention point to its usefulness in informing future large-scale Facebook-based studies.

Strengths and Limitations

Strengths of this study include the use of theory-driven intervention developed by a youth community advisory board, the use of a social media intervention, and a mixed methods approach. There are limitations to this study to consider. The sample is small and limited to those who had access to Facebook, and results are not generalizable to all young Black adults. The use of convenience sampling could have led to a homogenous participant pool and limited the diversity of perspectives. We collected self-reported measures, which could have social desirability bias. We also experienced a high attrition rate in the study. Future studies should oversample and send multiple routine reminders about follow-ups to minimize attrition. Another limitation of this study is the lack of a comparable control group. Similar to many pilot studies, the study was not powered, but provides a foundation for future studies, to be able to pilot-test the intervention and posts.

Conclusions

This study highlights significant improvements in HPV knowledge and HPV vaccine knowledge and a reduction in HPV vaccine hesitancy following a Facebook-based intervention among young Black adults. The findings from this study contribute to available research on the use of Facebook intervention among young Black adults. Intervention addressing knowledge gaps and improving HPV hesitancy among potential vaccine recipients is pivotal to promoting vaccine uptake. More research is needed to assess the type of messages and intervention dose that can most effectively facilitate changes in HPV perceived severity, vaccination benefits, self-efficacy, and intention to vaccinate to promote HPV vaccination behavior among young Black adults.

Acknowledgments

The authors thank all the volunteers who participated in the study. The study was supported by the American Cancer Society Institutional Grant (#IRG 19-140-31-IRG). The content is solely the responsibility of the authors and does not necessarily represent the official views of the sponsors.

Data Availability

The datasets analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Checklist 1
STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) statement—checklist of items that should be included in reports of cohort studies.
[DOCX File, 33 KB - formative_v9i1e69609_app1.docx]

XSL•FO
RenderX

References

1. McNamara M, Batur P, Walsh JME, Johnson KM. HPV update: vaccination, screening, and associated disease. J Gen Intern Med 2016 Nov;31(11):1360-1366. [doi: 10.1007/s11606-016-3725-z] [Medline: 27184752]
2. Han JJ, Beltran TH, Song JW, Klaric J, Choi YS. Prevalence of genital human papillomavirus infection and human papillomavirus vaccination rates among US adult men: National Health and Nutrition Examination Survey (NHANES) 2013-2014. JAMA Oncol 2017 Jun 1;3(6):810-816. [doi: 10.1001/jamaoncol.2016.6192] [Medline: 28114440]
3. Rosenblum HG, Lewis RM, Gargano JW, Querec TD, Unger ER, Markowitz LE. Declines in prevalence of human papillomavirus vaccine-type infection among females after introduction of vaccine—United States, 2003–2018. MMWR Morb Mortal Wkly Rep 2021 Mar 26;70(12):415-420. [doi: 10.15585/mmwr.mm7012a2] [Medline: 33764964]
4. Chiou L, Tucker C. Social distancing, internet access and inequality. Public Goods & Regulation in Health Economics eJournal. 2020. URL: https://www.oxy.edu/sites/default/files/assets/Economics/Chiou/covid19_inequality_social_distancing.pdf [accessed 2020-03-10]
5. Boersma P, Black LI. Human papillomavirus vaccination among adults aged 18-26, 2013-2018. NCHS Data Brief 2020 Jan(354):1-8. [Medline: 32487295]
6. De P, Budhwani H. Human papillomavirus (HPV) vaccine initiation in minority Americans. Public Health (Fairfax) 2017 Mar;144:86-91. [doi: 10.1016/j.puhe.2016.11.005] [Medline: 28274389]
7. Pérez AE, Agénor M, Gamarel KE, Operario D. Nativity disparities in human papillomavirus vaccination among U.S. adults. Am J Prev Med 2018 Feb;54(2):248-258. [doi: 10.1016/j.amepre.2017.10.019] [Medline: 29241719]
8. Amboree TL, Darkoh C. Barriers to human papillomavirus vaccine uptake among racial/ethnic minorities: a systematic review. J Racial Ethn Health Disparities 2021 Oct;8(5):1192-1207. [doi: 10.1007/s40615-020-00877-6]
9. Adegboyega A, Adeyimika D, Omoadoni O, Mark D. HPV vaccination and cervical cancer screening promotion among Black individuals: social ecological perspectives from key informants interviews. Ethn Health 2023 Oct;28(7):1026-1040. [doi: 10.1080/13557858.2023.2193360] [Medline: 36973897]
10. Adegboyega A, Obielodan O, Wiggins AT, Dignan M, Williams LB. Beliefs and knowledge related to human papillomavirus (HPV) vaccine among African Americans and African immigrants young adults. Cancer Causes Control 2023 May;34(5):479-489. [doi: 10.1007/s10552-023-01678-y] [Medline: 36854988]
11. Zheng L, Wu J, Zheng M. Barriers to and facilitators of human papillomavirus vaccination among people aged 9 to 26 years: a systematic review. Sex Transm Dis 2021 Dec 1;48(12):e255-e262. [doi: 10.1097/OLQ.0000000000001407] [Medline: 33783412]
12. Kutz JM, Rausche P, Gheit T, Puradiredja DI, Fusco D. Barriers and facilitators of HPV vaccination in sub-Saharan Africa: a systematic review. BMC Public Health 2023 May 26;23(1):974. [doi: 10.1186/s12889-023-15842-1] [Medline: 37237329]
13. Holman DM, Benard V, Roland KB, Watson M, Liddon N, Stokley S. Barriers to human papillomavirus vaccination among US adolescents: a systematic review of the literature. JAMA Pediatr 2014 Jan;168(1):76-82. [doi: 10.1001/jamapediatrics.2013.2752] [Medline: 24276343]
14. Rosen BL, Shepard A, Kahn JA. US health care clinicians' knowledge, attitudes, and practices regarding human papillomavirus vaccination: a qualitative systematic review. Acad Pediatr 2018 Mar;18(2S):S53-S65. [doi: 10.1016/j.acap.2017.10.007] [Medline: 29502639]
15. Zhu X, Jacobson RM, MacLaughlin KL, Sauver JS, Griffin JM, Finney Rutten LJ. Parent-reported barriers and parental beliefs associated with intentions to obtain HPV vaccination for children in a primary care patient population in Minnesota, USA. J Community Health 2023 Aug;48(4):678-686. [doi: 10.1007/s10900-023-01205-9] [Medline: 36920709]
16. Galbraith-Gyan KV, Lechuga J, Jenerette CM, Palmer MH, Moore AD, Hamilton JB. HPV vaccine acceptance among African-American mothers and their daughters: an inquiry grounded in culture. Ethn Health 2019 Apr;24(3):323-340. [doi: 10.1080/13557858.2017.1332758] [Medline: 28553758]
17. Washington A, Chabaan J, Fakih A, et al. "Should I give it to my kids?": Factors that influence HPV vaccine hesitancy among African American parents. Cancer Epidemiol Biomarkers Prev 2022 Jul 1;31(7):1512-1512. [doi: 10.1158/1055-9965.EPI-22-0478]
18. Shapiro GK, Tatar O, Amsel R, et al. Using an integrated conceptual framework to investigate parents' HPV vaccine decision for their daughters and sons. Prev Med 2018 Nov;116:203-210. [doi: 10.1016/j.ypmed.2018.09.017]
19. Cox DS, Cox AD, Sturm L, Zimet G. Behavioral interventions to increase HPV vaccination acceptability among mothers of young girls. Health Psychol 2010 Jan;29(1):29-39. [doi: 10.1037/a0016942] [Medline: 20063933]
20. Perkins RB, Zisblatt L, Legler A, Trucks E, Hanchate A, Gorin SS. Effectiveness of a provider-focused intervention to improve HPV vaccination rates in boys and girls. Vaccine (Auckl) 2015 Feb;33(9):1223-1229. [doi: 10.1016/j.vaccine.2014.11.021]
21. McFadden SM, Ko LK, Shankar M, et al. Development and evaluation of an online continuing education course to increase healthcare provider self-efficacy to make strong HPV vaccine recommendations to East African immigrant families. Tumour Virus Res 2021 Jun;11:200214. [doi: 10.1016/j.tvr.2021.200214] [Medline: 33647533]
22. Chu H, Ko LK, Ibrahim A, et al. The impact of an educational forum intervention on East African mothers' HPV vaccine-related knowledge, attitudes, and intentions to vaccinate their adolescent children. Vaccine (Auckl) 2021 Jun 23;39(28):3767-3776. [doi: 10.1016/j.vaccine.2021.05.029] [Medline: 34053792]

XSL•FO
RenderX

23. Baranoski AS, Stier EA. Factors associated with time to colposcopy after abnormal Pap testing in HIV-infected women. J Womens Health (Larchmt) 2012 Apr;21(4):418-424. [doi: 10.1089/jwh.2011.3046]
24. Baranowski T, Cullen KW, Basen-Engquist K, et al. Transitions out of high school: time of increased cancer risk? Prev Med 1997;26(5 Pt 1):694-703. [doi: 10.1006/pmed.1997.0193] [Medline: 9327479]
25. Korda H, Itani Z. Harnessing social media for health promotion and behavior change. Health Promot Pract 2013 Jan;14(1):15-23. [doi: 10.1177/1524839911405850] [Medline: 21558472]
26. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. P T 2014 Jul;39(7):491-520. [Medline: 25083128]
27. Mendoza-Herrera K, Valero-Morales I, Ocampo-Granados ME, Reyes-Morales H, Arce-Amaré F, Barquera S. An overview of social media use in the field of public health nutrition: benefits, scope, limitations, and a Latin American experience. Prev Chronic Dis 2020 Aug 6;17:E76. [doi: 10.5888/pcd17.200047] [Medline: 32762809]
28. Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. Pew Research Center. 2019. URL: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018 [accessed 2025-03-03]
29. Lebow S. Facebook is by far the most popular US social media platform. EMARKETER Inc. URL: https://www.emarketer.com/content/facebook-by-far-most-popular-us-social-media-platform [accessed 2025-03-10]
30. Average time spent on Facebook per day. Oberlo. 2024. URL: https://www.oberlo.com/statistics/average-time-spent-on-facebook [accessed 2025-03-03]
31. Leader AE, Miller-Day M, Rey RT, Selvan P, Pezalla AE, Hecht ML. The impact of HPV vaccine narratives on social media: testing narrative engagement theory with a diverse sample of young adults. Prev Med Rep 2022 Oct;29:101920. [doi: 10.1016/j.pmedr.2022.101920] [Medline: 36161120]
32. Brandt HM, Sundstrom B, Monroe CM, et al. Evaluating a technology-mediated HPV vaccination awareness intervention: a controlled, quasi-experimental, mixed methods study. Vaccines (Basel) 2020 Dec 10;8(4):749. [doi: 10.3390/vaccines8040749] [Medline: 33321975]
33. Wysota CN, Abroms LC, DeVarona H, Koban D, Napolitano M, Broniatowski DA. User experiences with a moderated Facebook group to promote vaccination. Am J Health Promot 2025 Jan;39(1):89-102. [doi: 10.1177/08901171241272061] [Medline: 39110567]
34. Montaño DE, Kasprzyk D. Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. In: Glanz K, Rimer BK, Viswanath KV, editors. Health Behavior: Theory, Research and Practice, 5th edition: Jossey-Bass/Wiley; 2015:95-124.
35. Catalano H, Richards K, Hawkins K. Theory of planned behavior-based correlates of HPV vaccination intentions and series completion among university students in the Southeastern United States. Health Educ (Muncie) 2017;49(2):35-44 [FREE Full text]
36. Adegboyega A, Francis DB, Ebikwo C, Ntego T, Ickes M. Engagement with a youth community advisory board to develop and refine a Facebook HPV vaccination promotion intervention (#HPVvaxtalks) for young black adults (18-26 years old). Health Promot Pract 2023 Dec 5:15248399231216731. [doi: 10.1177/15248399231216731] [Medline: 38050905]
37. Adebola A, Charis E, Tristan N, Melinda I. Content validation of a Facebook HPV vaccination promotion intervention (#HPVvaxtalks) created for young Black adults (18–26 years old). Public Health Nurs 2024 Sep;41(5):1178-1187. [doi: 10.1111/phn.13356]
38. Kreuter MW, Lukwago SN, Bucholtz R, Clark EM, Sanders-Thompson V. Achieving cultural appropriateness in health promotion programs: targeted and tailored approaches. Health Educ Behav 2003 Apr;30(2):133-146. [doi: 10.1177/1090198102251021] [Medline: 12693519]
39. HPV social media graphics. National Foundation for Infectious Diseases. URL: https://www.nfid.org/resource/hpv-social-media-graphics/ [accessed 2025-03-10]
40. Gerend MA, Shepherd JE. Predicting human papillomavirus vaccine uptake in young adult women: comparing the health belief model and theory of planned behavior. Ann Behav Med 2012 Oct;44(2):171-180. [doi: 10.1007/s12160-012-9366-5] [Medline: 22547155]
41. Marlow LAV, Zimet GD, McCaffery KJ, Ostini R, Waller J. Knowledge of human papillomavirus (HPV) and HPV vaccination: an international comparison. Vaccine (Auckl) 2013 Jan;31(5):763-769. [doi: 10.1016/j.vaccine.2012.11.083]
42. Waller J, Ostini R, Marlow LAV, McCaffery K, Zimet G. Validation of a measure of knowledge about human papillomavirus (HPV) using item response theory and classical test theory. Prev Med 2013 Jan;56(1):35-40. [doi: 10.1016/j.ypmed.2012.10.028]
43. Luyten J, Bruyneel L, van Hoek AJ. Assessing vaccine hesitancy in the UK population using a generalized vaccine hesitancy survey instrument. Vaccine (Auckl) 2019 Apr;37(18):2494-2501. [doi: 10.1016/j.vaccine.2019.03.041]
44. Jones HJ, Kovacic MB, Bakas T. Establishing validity of the midlife Black women's stress-reduction wellness program materials using a mixed methods approach. J Cardiovasc Nurs 2022;37(5):446-455. [doi: 10.1097/JCN.0000000000000876] [Medline: 34935740]

XSL•FO
RenderX

45. Horvath KJ, Oakes JM, Rosser BRS, et al. Feasibility, acceptability and preliminary efficacy of an online peer-to-peer social support ART adherence intervention. AIDS Behav 2013 Jul;17(6):2031-2044. [doi: 10.1007/s10461-013-0469-1] [Medline: 23553347]
46. Creswell JW. Qualitative Inquiry and Research Design: Choosing Among Five Approaches: Sage Publications; 2012.
47. Lee HY, Koopmeiners JS, McHugh J, Raveis VH, Ahluwalia JS. mHealth pilot study: text messaging intervention to promote HPV vaccination. Am J Health Behav 2016 Jan;40(1):67-76. [doi: 10.5993/AJHB.40.1.8] [Medline: 26685815]
48. Reiter PL, Gower AL, Kiss DE, et al. Effects of a web-based HPV vaccination intervention on cognitive outcomes among young gay, bisexual, and other men who have sex with men. Hum Vaccin Immunother 2022 Nov 30;18(6):2114261. [doi: 10.1080/21645515.2022.2114261] [Medline: 36069662]
49. Ou L, Chen ACC, Amresh A. The effectiveness of mHealth interventions targeting parents and youth in human papillomavirus vaccination: systematic review. JMIR Pediatr Parent 2023 Nov 21;6:e47334. [doi: 10.2196/47334] [Medline: 37988155]
50. Allen JD, Hollander J, Gualtieri L, Alarcon Falconi TM, Savir S, Agénor M. Feasibility of a twitter campaign to promote HPV vaccine uptake among racially/ethnically diverse young adult women living in public housing. BMC Public Health 2020 Jun 1;20(1):830. [doi: 10.1186/s12889-020-08824-0] [Medline: 32487045]
51. Larson HJ, de Figueiredo A, Xiahong Z, et al. The state of vaccine confidence 2016: global insights through a 67-country survey. EBioMedicine 2016 Oct;12:295-301. [doi: 10.1016/j.ebiom.2016.08.042] [Medline: 27658738]
52. Tuckerman J, Kaufman J, Danchin M. Effective approaches to combat vaccine hesitancy. Pediatr Infect Dis J 2022 May 1;41(5):e243-e245. [doi: 10.1097/INF.0000000000003499] [Medline: 35213864]

Abbreviations

HPV: human papillomavirus
REDCap: Research Electronic Data Capture
STROBE: Strengthening the Reporting of Observation Studies in Epidemiology

Edited by A Mavragani; submitted 03.12.24; peer-reviewed by J Cunningham-Erves, W Tang; revised version received 07.02.25; accepted 08.02.25; published 02.04.25.

Please cite as:
Adegboyega A, Wiggins A, Wuni A, Ickes M
The Impact of a Human Papillomavirus Facebook-Based Intervention (#HPVVaxTalks) Among Young Black (African American and Sub-Saharan African Immigrants) Adults: Pilot Pre- and Poststudy
JMIR Form Res 2025;9:e69609
URL: https://formative.jmir.org/2025/1/e69609
doi:10.2196/69609

© Adebola Adegboyega, Amanda Wiggins, Abubakari Wuni, Melinda Ickes. Originally published in JMIR Formative Research (https://formative.jmir.org), 2.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Medical Device Based on a Virtual Reality–Based Upper Limb Rehabilitation Software: Usability Evaluation Through Cognitive Walkthrough

Seojin Hong[1], MS; Hyun Choi[2], PhD; Hyosun Kweon[1], PhD

[1]Department of Clinical Research Rehabilitation, National Rehabilitation Center, 58 Samgaksan-ro, Gangbuk-gu, Seoul, Republic of Korea

[2]Department of Healthcare and Public Health, National Rehabilitation Center, Seoul, Republic of Korea

Corresponding Author:
Hyosun Kweon, PhD
Department of Clinical Research Rehabilitation, National Rehabilitation Center, 58 Samgaksan-ro, Gangbuk-gu, Seoul, Republic of Korea

Abstract

Background: The use of virtual reality (VR) technology in rehabilitation therapy has been growing, leading to the development of VR-based upper-limb rehabilitation softwares. To ensure the effective use of such software, usability evaluations are critical to enhance user satisfaction and identify potential usability issues.

Objective: This study aims to evaluate the usability of a VR-based upper-limb rehabilitation software from the perspective of occupational therapists. Specifically, the study seeks to identify usability challenges and provide insights to improve user satisfaction.

Methods: The VR-based upper-limb rehabilitation software was tailored for therapists to operate while delivering therapy to patients. Usability testing was conducted with occupational therapists from the Korean National Rehabilitation Center using cognitive walkthroughs and surveys. Participants performed tasks that simulated real clinical scenarios, including turning the device on, assisting patients with wearing the device, and shutting it down. Observers recorded user reactions during task performance, and participants completed surveys to assess the ease of use of the user interface. This mixed-methods approach provided qualitative insights into user difficulties and their root causes.

Results: Usability evaluations were conducted with 6 participants. Cognitive walkthroughs revealed potential areas for improvement in the software, including (1) enhancements to the graphical user interface for ease of use, (2) refinements in the natural user interface, and (3) better user manuals for clearer product instructions. The ease-of-use score for the user interface averaged 1.58 on a 5-point scale (1=very easy to 5=very difficult).

Conclusions: This study provides valuable insights into improving user satisfaction by focusing on the needs of occupational therapists who operate a VR-based rehabilitation software. Future research should explore software refinement and clinical efficacy to maximize the therapeutic potential of such technologies.

(JMIR Form Res 2025;9:e68149) doi:10.2196/68149

KEYWORDS

usability; cognitive walkthrough; virtual reality-based upper limb rehabilitation software; upper limb; limb rehabilitation; rehabilitation; therapist; virtual reality; VR; medical device; formative evaluation; quantitative; qualitative; occupational therapy; user safety; usability testing; software; risk factor

Introduction

Background

Upper limb rehabilitation plays a critical role in restoring functional abilities in individuals with stroke or neurological injuries, enhancing their independence and autonomy [1,2]. Virtual reality (VR) technology has emerged as a promising tool in rehabilitation therapy, offering immersive and interactive features that can improve patient engagement and therapeutic outcomes [3-5]. Through tailored feedback and adjustable difficulty levels, VR-based rehabilitation programs can address diverse patient needs and therapeutic goals [6]. To achieve optimal clinical use, VR-based rehabilitation software must be designed to meet the needs of end users, including health care professionals such as occupational therapists, within their specific work environments. Usability evaluation is a critical step in this process, as it helps identify design flaws, reduce user errors, and enhance operational efficiency and satisfaction [7]. While existing studies on VR-based rehabilitation have largely focused on technical development and clinical efficacy [4], research on usability issues experienced by health care

XSL•FO
RenderX

professionals during actual software use remains limited [3] . Specifically, there is a lack of comprehensive analysis regarding the intuitive interaction, learning curve, and practical applicability of such software in real-world clinical settings [7].

Systematic usability evaluation targeting occupational therapists, as key stakeholders in VR-based rehabilitation, is essential to identify and address usability issues. Formative evaluation methods, such as cognitive walkthroughs, provide valuable insights during the early stages of software development, enabling iterative improvements and better user experience [3].

Usability Evaluation

Errors encountered during the use of medical devices are often linked to poorly designed interfaces rather than user mistakes. Although not all user errors result from design flaws, usability evaluations conducted before commercialization can effectively mitigate design-related issues [8,9]. Usability engineering processes play a vital role in identifying and addressing predictable user errors in clinical environments, ensuring the safety and efficiency of medical devices [10,11].

Usability evaluations can be categorized as formative or summative, depending on the timing and purpose of the evaluation [12]. Formative evaluations aim to identify usability issues and guide design improvements during the product development phase, while summative evaluations focus on validating outcomes at the final stages of development through objective data analysis [13]. Formative evaluations commonly use methods such as cognitive walkthroughs, task analysis, usability testing, and heuristic evaluation to explore and address usability challenges in depth [14].

Cognitive walkthroughs, in particular, are highly effective for evaluating software usability. They focus on assessing the system from the perspective of new users, enabling the identification of potential issues such as inconsistent interface behavior or a lack of clarity [15]. By providing insights into users' cognitive processes, cognitive walkthroughs empower developers to design more user-friendly, efficient, and satisfying interfaces.

Objective

This study aims to conduct a formative evaluation of a prototype VR-based upper limb rehabilitation software, focusing on its usability for occupational therapists. By identifying usability challenges and gathering feedback from intended users, this research seeks to provide actionable insights to guide iterative improvements in the software design. Ultimately, this study aims to contribute foundational data for user-centered design and optimize the software for effective clinical implementation.

Methods

Study Design

This study conducted a usability evaluation of a prototype VR-based upper limb rehabilitation software targeted for use by occupational therapists. The evaluation used cognitive walkthroughs and surveys, covering the entire user workflow, including activating the device, setting it up, and completing tasks. The software was assessed as an embedded system, including its hardware components.

Formative Evaluation Procedures

The formative evaluation was conducted in the clinical rehabilitation testbed of the Korean National Rehabilitation Center. Each session lasted approximately 90 minutes per participant (Table 1). Before the evaluation, the facilitator introduced the usability testing process, including its objectives, methods, and relevant product information.

XSL•FO
RenderX

Table . Formative evaluation procedures. The formative evaluation involved practitioners and evaluators, and was conducted for 90 minutes under the guidance of a facilitator.

Composition	Details	Time (minutes)
Orientation	• The facilitator explains the target product, purpose, overview, methodology, and other relevant aspects of the usability evaluation to the participant. • The facilitator introduces the concept of usability evaluation to participants who are unfamiliar with it.	10
Guidance on consent	• The facilitator provides the participant with detailed instructions regarding recording and transcription procedures. • The facilitator ensures that the participant fully comprehends the evaluation content and voluntarily provides written consent to participate in the usability evaluation.	10
Perform cognitive walkthrough	• The facilitator conducts a cognitive walk-through following a predefined script and product usage instructions. • The participant operates the medical device in accordance with the facilitator's guidance. • The observer documents and analyzes the participant's responses during task performance.	50
Survey	• The facilitator administers a survey to the participant to assess the usability of the user interface.	20

Participants were asked about their previous experience with similar devices. Subsequently, they followed a structured script guiding them through a series of tasks using the software. Participants operated the software as instructed, while observers recorded their behaviors and responses throughout the process. Post-testing, participants were asked to evaluate their satisfaction with the system based on the VR-based upper limb rehabilitation software.

The usability testing was conducted in a simulated clinical environment resembling the intended use setting for the device (Figure 1). Environmental conditions, including lighting, temperature, and noise, were measured before the evaluation: lighting at 550 (SD 100) lx, temperature at 24°C (SD 2°C), relative humidity at 60% (SD 10%), and noise level at 50 (SD 5) dBA.

Figure 1. Test environment (left) and scene (right). The simulation environment is configured to resemble a rehabilitation treatment room where evaluators are used.

Cognitive Walkthrough

The cognitive walkthrough method evaluates how well users can navigate an interface without previous training or information about the device [16]. In this study, 6 occupational therapists conducted the cognitive walkthrough, assessing the ease of navigation for specific scenarios outlined in Table 2. Observers recorded and analyzed participants' reactions based on predefined observation sheets. Observation sheets are designed to record problems discovered during each task, user reactions and actions, and system feedback [17,18].

XSL•FO
RenderX

Table . Use scenarios for cognitive walkthrough. The facilitator verbally provided detailed subtask instructions to the occupational therapist, who performed the tasks in accordance with the usage scenario, while the simulated patient participated in the process under the occupational therapist's guidance.

Task	Subtask
1. Reviewing the user manual	• 1.1. Reviewing the user manual.
2. Checking components and power supply	• 2.1. Checking the HMD[a], controller, disposable face mask, PC set, and software in the PC. • 2.2. Checking the product's power supply.
3. Launching the program	• 3.1. Launching the "RehabwareVR" (Tech Village) icon and entering the password.
4. Creating patient information	• 4.1. Entering patient information. • 4.2. Providing instructions on patient data consent.
5. Donning the equipment	• 5.1. Instructing the patient to wear the disposable face mask, HMD, and controller.
6. Running the upper rehabilitation content	• 6.1. Selecting one of the upper limb rehabilitation exercises, explaining it to the patient, and running.
7. Running the content	• 7.1. Selecting the patient and configuring the catch ball activity. • 7.2. Saving these settings as User Setting 3, then starting the catch ball activity. • 7.3. After completing at least five repetitions, stopping the activity, modifying the environment settings, and rerunning the activity. • 7.4. Switching to the fruit transfer activity and running it. • 7.5. After completing the fruit transfer activity, proceeding with the bubble activity.
8. Updating patient information	• 8.1. Updating the patient information.
9. Running contents after updating patient information	• 9.1. Selecting the patient and starting the hammering activity. • 9.2. After completing the hammering activity, switching to the meteor avoidance activity. • 9.3. Ending the session.
10. Doffing the product	• 10.1. Removing the HMD, controllers, and disposable face mask from the patient.
11. Reviewing statistical analysis results	• 11.1. Reviewing the data graph of the patient. • 11.2. Reviewing the fruit transfer activity data for patient. • 11.3 Deleting the hammering activity data for patient.
12. Ending the program	• 12.1. Ending the program.

[a]HMD: head-mounted display.

Survey

Before the cognitive walkthrough, participants provided demographic information, clinical experience, and previous experience with similar devices. After completing the cognitive walkthrough, participants completed a survey to further assess the usability of the user interface based on the scenarios. The survey consisted of 23 items rated on a 5-point Likert scale (1=very easy to 5=very difficult). Results were analyzed using means and SDs.

VR-Based Upper Limb Rehabilitation Software

The prototype VR-based upper limb rehabilitation software used in this study was designed to improve upper limb function, including shoulder mobility, in patients with stroke or neurological injuries (Figure 2). Patients wear a head-mounted display (HMD; HTC VIVE Pro Full-Kit including VIBE base stations and controllers) and use handheld controllers to perform voluntary movements in a VR environment for rehabilitation (Figure 3). Occupational therapists, as intended users, set up the hardware, operate the software, and assist patients in improving their upper limb functions.

XSL•FO
RenderX

Figure 2. User interface used by occupational therapists in the virtual reality (VR)–based upper limb rehabilitation software: home screen (left), activity selection screen (middle), and activity options manipulation screen (right).

Figure 3. Hardware device of the virtual reality (VR)–based upper limb rehabilitation software. To use VR content, the patient wears the head-mounted device (HMD; left) on their head and holds the controller (right).

The software includes eight rehabilitation activities (Figure 4): (1) throwing a catching balls, (2) hammering nails, (3) popping bubbles, (4) moving fruits, (5) dodging meteors, (6) throwing balls, (7) playing the xylophone, and (8) stacking blocks.

Figure 4. Examples of virtual reality within the patient-worn head-mounted device (HMD), such as throwing a catching balls (left) and hammering nails (right).

The minimum computer requirements to run the software are Microsoft Windows 7 SP, Windows 8.1, or Windows 10, with a display resolution of 1280×72 0r higher. Recommended specifications include Microsoft Windows 10 Professional with the same resolution. The hardware used was the VIVE Pro Full-Kit.

Participants

The usability evaluation involved 3 usability practitioners (a facilitator and 2 observers), 6 occupational therapists as participants, and 3 mock patients. The facilitator conducted the evaluation, while observers recorded and analyzed participants' responses. The sample size of 6 participants was determined to balance cost and effectiveness, as this number is sufficient to identify most usability issues [19]. Mock patients were recruited to simulate the role of patients during testing, prioritizing safety by selecting healthy individuals. In the early stages of development, formative evaluations typically involve the use of simulated patients before including real patients. This approach ensures safety and control while being effective in

XSL•FO
RenderX

identifying functional issues in devices [20,21]. Furthermore, unlike real patients, simulated patients are trained to provide consistent conditions and responses, thereby maintaining the reliability and consistency of the evaluation process [22,23]. Accordingly, this study employed simulated patients instead of real patients to conduct the evaluation. All participants were informed of the study's purpose, methods, and procedures before testing and provided signed consent.

Recruitment Process

Recruitment was conducted through posted announcements at the Korean National Rehabilitation Center. The inclusion criteria encompassed possession of an occupational therapist license and previous experience with rehabilitation medicine devices, such as computerized cognitive rehabilitation programs, whereas the exclusion criteria comprised individuals who did not provide consent to participate in the study. Through eligibility screening, 6 occupational therapists were selected. The mock patients were 3 healthy adults aged 19 years and older and had no experience using the evaluation product. They were provided detailed explanations of the study and coordinated testing schedules. Written informed consent was obtained from all participants.

Ethical Considerations

This study was approved by the Institutional Review Board of the Korean National Rehabilitation Center (NRC-2023-06-041). All participants provided informed consent, and no financial compensation was offered. Data were anonymized to ensure participant privacy and confidentiality.

Results

Participant Characteristics

The clinical experience of the occupational therapists ranged from 6 months to 6 years, with an average of 3 years and 9 months. In addition, they had previous experience using similar medical devices, such as Comcog and RAPAEL (Neofect) (computerized cognitive rehabilitation programs), with their experience ranging from a minimum of 2 months to a maximum of 5 years (Table 3).

Table . General characteristics of participants. The participant in the cognitive walkthrough was an occupational therapist who had experience using similar medical devices.

Participant Number	Sex	Age (years)	Experience	Occupation	Experience using similar medical devices	Name of similar medical device
1	Female	28	6 years	Occupational therapist	6 months, twice a week	Comcog (Neofect), RAPAEL Smart Glove (Neofect), and RAPAEL Smart pegboard
2	Female	24	6 months	Occupational therapist	2 months, twice a week	RAPAEL
3	Female	27	5 years	Occupational therapist	5 years, twice a week	Comcog, RAPAEL, and RehaCom (Hasomed)
4	Female	25	3 years	Occupational therapist	3 years, once a week	RAPAEL and Xbox (Microsoft)
5	Male	27	3 years, 9 months	Occupational therapist	2 years, once a day	Comcog
6	Female	29	5 years, 8 months	Occupational therapist	1 year, 3 times a month	Comcog and RehaCom

Cognitive Walkthrough Results

A total of 12 usability errors were identified using cognitive walkthrough. In task 4, "Creating Patient Information," 4 errors were identified. In subtask 4.1, "Locating the Add Patient Icon," participants were confused as they could not find the icon leading to the screen for adding patient information. In subtask 4.2, "Registering Patient Information," participants hesitated when they had to enter such details as the diagnosis and dominant hand manually because moving to the next input field required a click on the mouse, as the tab key shortcut did not work. In subtask 4.3, "Guiding Patients' Consent for Using Personal Information," participants skipped the data consent step and almost automatically pressed the confirm button without explaining the process of consent for using personal information to the mock patient.

In task 7, "Performing the Activity," 5 usability issues were identified. In subtask 7.1, "Selecting Activity," participants attempted to select activities without first selecting the patient, which triggered a popup message indicating "No patient selected." In subtask 7.2, "Configuring the Activity Settings," participants committed multiple errors, including failure to estimate the perceived distance between the hand and the object within the virtual reality, and identifying which hand was being used for interaction. Furthermore, although the users clicked on the "User Settings" button, they failed to save the configurations. There were also challenges in adjusting precise decimal values using mouse clicks and dragging, and this time-consuming process delayed the therapeutic session. In subtask 7.4, "Changing Activities," participants switched activities without saving current data, which led to the loss of unsaved records. In addition, some therapists had difficulty

XSL•FO
RenderX

adjusting the settings after completing an activity, as they were unsure whether they needed to press the confirm button before proceeding to the next activity.

In task 11, "Reviewing Statistical Analysis Results," 3 errors were identified. In subtask 11.1, "Reviewing Data Graphs," no issues were observed. In subtask 11.2, "Reviewing Activity Data," participants were unable to review the data because activity records were not saved, and there was inconsistency between the activity names on the playscreen and the statistical analysis screen, leading to further confusion. In subtask 11.3, "Deleting Data" participants clicked on the pie chart and searched for a delete button on the detailed graph screen.

Survey Results

The survey assessing the usability of user interface showed that most participants rated usability as a score of 2 or lower, indicating ease of use on a 5-point Likert scale (1=very easy to 5=very difficult; Table 4).

XSL•FO
RenderX

Table . Survey results of usability of user interface (n=5). Usability of user interface showed that most participants rated usability as a score of 2 (easy) or lower (1=very easy; 5=very difficult).

Usability of UI[a]	Mean (SD)
Reviewing user manual	
Identifying information	1.67 (SD 0.82)
Comprehensibility	1.83 (SD 0.98)
Checking the components	
Differentiating between components	1.50 (SD 0.55)
Verifying connection and power	1.17 (SD 0.41)
Launching the program	
Launching the software	1.00 (SD 0.00)
Entering the password	1.00 (SD 0.00)
Creating patient information	
Creating patient information	2.17 (SD 0.75)
Consenting to the disclosure of personal information	1.67 (SD 1.03)
Donning the product	
Donning the product	2.00 (SD 1.10)
Instructions regarding the product	1.83 (SD 0.98)
Running the upper rehabilitation content	
Engaging in the content	1.50 (SD 0.55)
Explaining the contents	1.50 (SD 0.55)
Performing the activity	
Pausing the content	1.67 (SD 1.03)
Configuring detailed settings	2.17 (SD 1.33)
Changing detailed settings	2.17 (SD 1.33)
Ending the content	1.67 (SD 1.03)
Updating patient information	
Updating patient information	1.33 (SD 0.82)
Running contents after updating patient information	
Changing the patient	1.67 (SD 1.03)
Doffing the product	
Doffing the product	1.50 (SD 0.84)
Reviewing statistical analysis results	
Checking the graph	1.33 (SD 0.82)
Checking data	1.33 (SD 0.82)
Deleting data	1.50 (SD 1.22)
Ending the program	
Ending the software	1.17 (SD 0.41)

[a]UI: user interface.

Discussion

Principal Findings

The following recommendations were suggested from the results of the cognitive walkthrough to enhance usability and user satisfaction with the interface of the VR-based upper limb rehabilitation software. First, enhancements to the graphical user interface (GUI) are essential. A GUI facilitates more intuitive and efficient information presentation using graphics instead of text [24]. This study identified usability issues stemming from the lack of clarity in the “Add Patient” icon,

XSL•FO
RenderX

underscoring the need for a more intuitive icon that clearly conveys its function. In addition, improvements in typography, such as optimizing font size and color contrast, are necessary to enhance the readability of activity-setting screens.

Furthermore, to streamline data entry, manual input for all patient information fields should be replaced with input controls such as pickers, which allow users to select from predefined options, or steppers, which enable users to increment or decrement numbers or values using "+" and "-" buttons. In addition, field navigation should be optimized by enabling users to switch between fields using the tab key, thereby facilitating user input. Finally, the functionality for controlling buttons within the software should be visibly displayed on the interface. When developing GUI-based medical software devices, the interface design should aim to integrate clear control mechanisms and simplify tasks, thereby minimizing users' cognitive load and memory demands [25].

Second, integrating natural user interface (NUI) elements could further enhance usability. NUI refers to an interface that relies on gestures rather than traditional input devices, such as a mouse or keyboard, to control content [26]. While the VR-based upper limb rehabilitation software uses a gesture-based interface using an HMD and controllers, the system does not differentiate between left and right sides. This limitation can hinder therapists from properly administering treatments for patients with unilateral paralysis. To address this issue, the software should include features to either automatically detect the controller's orientation or provide manual configuration options to distinguish the affected side. Furthermore, therapists reported difficulties in adjusting the perceived distance between the hand and objects displayed on their monitors, highlighting the need for improved distance control mechanisms in the VR environment. Therefore, a VR software interface should be designed to ensure ease of use for both therapists and patients.

Third, the user interface (UI) design requires improvements to enhance the interaction between the medical device and its users. In this study, we observed that therapists used inconsistent methods for adjusting settings and some failed to save their configurations before proceeding with the tasks. Although most medical device users are well-trained and experienced, errors can still occur when they rely heavily on their own previous knowledge. If the system can provide warning alerts when settings are unsaved or incorrect actions are detected, and provide confirmation prompts when actions are correct, unintended errors can be mitigated [27]. Therefore, warning alerts, confirmation popups, detailed instructions, and help guides within the software interface are recommended to prevent usage errors.

Formative evaluation is a type of usability test conducted during the medical device development phase to improve product design. In this study, we conducted a cognitive walkthrough and surveys with occupational therapists as participants to explore potential areas for improvement in the VR-based upper limb rehabilitation software.

Cognitive walkthrough is a usability test method used to systematically identify interactions between users and a system without previous knowledge of the system. It is particularly suitable for evaluating screen-based systems where users must actively navigate the interface [16]. Test tasks should be selected from the user's perspective, and the sequence of potential actions should first be outlined, followed by a step-by-step analysis of continuous actions to assess whether appropriate actions are taken at the correct moments. Successful interaction is determined by whether the system feedback helps users achieve their intended goals [28].

In this study, we identified usability issues that need to be addressed in the UI of the VR-based upper limb rehabilitation software as a medical device. However, there are some limitations to note. Although both health care providers delivering the treatment and patients receiving it were considered users, actual patients were not included in the participant group. In addition, the therapists involved in the study did not represent a diverse range of clinical experience levels. Despite these limitations, the rehabilitation specialists who participated in this study provided valuable insights, including perspectives from the patients' point of view. Future research should expand the scope of such work to include a broader range of users and conduct a more comprehensive analysis of the usability across various features and tasks.

Conclusion

VR-based rehabilitation represents a rapidly emerging field in health care. In this study, we conducted a usability evaluation of a prototype VR-based upper limb rehabilitation software. By ensuring a user-centered interface and addressing the identified UI issues and areas for improvement, enhancing the usability of the software as a medical device will contribute to the development of a solution that fully meets the needs of its users. These efforts will not only prioritize user satisfaction and ease of use but also accelerate the commercialization of the VR-based upper limb rehabilitation software. Furthermore, improving user-device interaction will enhance the usability and satisfaction associated with the device, ultimately contributing to the development of a medical device that delivers error-free performance, as intended by the system.

Acknowledgments

This study was supported by the National Rehabilitation Center Rehabilitation Research & Development Support Program, funded by the Ministry of Health and Welfare (NRCRSP-25TB01).

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

The research design and data analysis were primarily conducted by SH, while data collection and experimental execution were contributed by HC. The interpretation of results and drafting of the manuscript were carried out jointly. The final review and approval of the manuscript, as well as research direction adjustments, were overseen by HK.

Conflicts of Interest

None declared.

References

1. Carson RG, Hayward KS. Using mechanistic knowledge to appraise contemporary approaches to the rehabilitation of upper limb function following stroke. J Physiol 2025 Feb;603(3):635-650. [doi: 10.1113/JP285559] [Medline: 39129269]
2. Zhang T, Yao Z, Chen F, et al. Enhancing upper limb function and motor skills post-stroke through an upper limb rehabilitation robot. J Vis Exp 2024 Sep 6(211). [doi: 10.3791/66938] [Medline: 39311613]
3. Zhang T, Booth R, Jean-Louis R, et al. A primer on usability assessment approaches for health-related applications of virtual reality. JMIR Serious Games 2020 Oct 28;8(4):e18153. [doi: 10.2196/18153] [Medline: 33112243]
4. Seo NJ, Arun Kumar J, Hur P, Crocher V, Motawar B, Lakshminarayanan K. Usability evaluation of low-cost virtual reality hand and arm rehabilitation games. J Rehabil Res Dev 2016;53(3):321-334. [doi: 10.1682/JRRD.2015.03.0045] [Medline: 27271199]
5. Marques B, Oliveira S, Ferreira H, Amorim P, Dias P, Santos BS. The role of collaborative virtual reality engagement in stroke survivors' rehabilitation. In: Accessibility, Assistive Technology and Digital Environments: AHFE International; 2024, Vol. 121. URL: https://openaccess.cms-conferences.org/publications/book/978-1-958651-97-1/article/978-1-958651-97-1_3 [accessed 2025-01-26]
6. Oliveira S, Marques B, Amorim P, Dias P, Sousa Santos B. Stepping into recovery with an immersive virtual reality serious game for upper limb rehabilitation: a supermarket experience for stroke survivors. In: Chen JYC, Fragomeni G, editors. Virtual, Augmented and Mixed Reality Lecture Notes in Computer Science: Springer Nature Switzerland; 2024, Vol. 14708:142-153. [doi: 10.1007/978-3-031-61047-9_9]
7. Kamińska D, Zwoliński G, Laska-Leśniewicz A. Usability testing of virtual reality applications-the pilot study. Sensors (Basel) 2022 Feb 10;22(4):1342. [doi: 10.3390/s22041342] [Medline: 35214246]
8. Johnson TR, Tang X, Graham MJ, et al. Attitudes toward medical device use errors and the prevention of adverse events. Jt Comm J Qual Patient Saf 2007 Nov;33(11):689-694. [doi: 10.1016/s1553-7250(07)33079-1] [Medline: 18074717]
9. Koppel R, Metlay JP, Cohen A, et al. Role of computerized physician order entry systems in facilitating medication errors. JAMA 2005 Mar 9;293(10):1197-1203. [doi: 10.1001/jama.293.10.1197] [Medline: 15755942]
10. Bond RR, Finlay DD, Nugent CD, Moore G, Guldenring D. A usability evaluation of medical software at an expert conference setting. Comput Methods Programs Biomed 2014;113(1):383-395. [doi: 10.1016/j.cmpb.2013.10.006] [Medline: 24210868]
11. IEC I 62366-1: 2015 medical devices—part 1: application of usability engineering to medical devices. : International Electrotechnical Commission (IEC); 2015 URL: https://www.iso.org/standard/63179.html#:~:text=IEC%2062366%2D1%3A2015%20specifies,as%20it%20relates%20to%20SAFETY [accessed 2025-03-25]
12. Lewis JR. Usability: lessons learned … and yet to be learned. Int J Hum Comput Interact 2014 Sep 2;30(9):663-684. [doi: 10.1080/10447318.2014.930311]
13. Sauro J. A Practical Guide to Measuring Usability: 72 Answers to the Most Common Questions about Quantifying the Usability of Websites and Software: Measuring Usability LLC; 2010. URL: https://measuringu.com/wp-content/uploads/2017/05/QuantitativeUsabilityTestOnline.pdf [accessed 2025-03-25]
14. Choi CJ, Park IH, Kim SM, Kwon JY. A study on the usability evaluation for medical devices design using heuristic evaluation. Presented at: 2019 Spring Conference of ESK; May 31 to Jun 1, 2019; Yeosu, Korea.
15. Rieman J, Franzke M, Redmiles D. Usability evaluation with the cognitive walkthrough. Presented at: CHI '95: Conference Companion on Human Factors in Computing Systems; May 7-11, 1995; Denver, Colorado, United States. [doi: 10.1145/223355.223735]
16. Liu Y, Osvalder AL, Dahlman S. Exploring user background settings in cognitive walkthrough evaluation of medical prototype interfaces: a case study. Int J Ind Ergon 2005 Apr;35(4):379-390. [doi: 10.1016/j.ergon.2004.10.004]
17. Lavery D, Cockton G. Cognitive walkthrough: usability evaluation materials. Northumbria University Library. 1997. URL: https://nrl.northumbria.ac.uk/id/eprint/26027/ [accessed 2025-01-26]
18. Wharton C, Rieman J, Lewis C, Polson P. The cognitive walkthrough: a practitioner's guide. In: Usability Inspection Methods: John Wiley; 1994.
19. Nielsen J. Usability Engineering: Academic Press; 1993.
20. Clanchy K, Mitchell J, Mulholland K, et al. Towards co-design of rehabilitation technologies: a collaborative approach to prioritize usability issues. Front Rehabil Sci 2024;5:1302179. [doi: 10.3389/fresc.2024.1302179] [Medline: 38450206]

XSL•FO
RenderX

21. Alhasani R, George N, Radman D, Auger C, Ahmed S. Methodologies for evaluating the usability of rehabilitation technologies aimed at supporting shared decision-making: scoping review. JMIR Rehabil Assist Technol 2023 Aug 15;10:e41359. [doi: 10.2196/41359] [Medline: 37581911]
22. Park HK. Medical education using standardized patients. Hanyang Med Rev 2012;32(1):35. [doi: 10.7599/hmr.2012.32.1.35]
23. Gorski S, Prokop-Dorner A, Pers M, et al. The use of simulated patients is more effective than student role playing in fostering patient-centred attitudes during communication skills training: a mixed method study. Biomed Res Int 2022;2022:1498692. [doi: 10.1155/2022/1498692] [Medline: 36573197]
24. Study on the course of smart phones GUI UX and experiences design elements. journalofdigitaldesign 2014 Apr;14(2):545-554. [doi: 10.17280/jdd.2014.14.2.051]
25. Oh YH. A study on user interface design for improving usability of the medical device. In: KSDS Conference Proceeding: Korean Society of Design Science; 2009:126-127.
26. Kim B. A study on user experience of gesture-base interface in virtual reality [Master's thesis]. : Aju University; 2019.
27. Chae SZ, Kim AJ. A study on process of medical device development and guideline of ergonomics for safety improvement. ksdc 2017 Jun 30;23(2):701-710. [doi: 10.18208/ksdc.2017.23.2.701]
28. Hanington B, Martin B. The Pocket Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas and Design Effective Solutions: Rockport; 2017.

Abbreviations

GUI: graphical user interface
HMD: head-mounted display
NUI: natural user interface
UI: user interface
VR: virtual reality

Edited by A Mavragani; submitted 03.11.24; peer-reviewed by B Marques, G Guerrero-Contreras, J Fielder, JH Lee; revised version received 25.01.25; accepted 13.02.25; published 01.04.25.

Please cite as:
Hong S, Choi H, Kweon H
Medical Device Based on a Virtual Reality–Based Upper Limb Rehabilitation Software: Usability Evaluation Through Cognitive Walkthrough
JMIR Form Res 2025;9:e68149
URL: https://formative.jmir.org/2025/1/e68149
doi:10.2196/68149

© Seojin Hong, Hyun Choi, Hyosun Kweon. Originally published in JMIR Formative Research (https://formative.jmir.org), 1.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Development of a Comprehensive Decision Support Tool for Chemotherapy-Cycle Prescribing: Initial Usability Study

Sanna Iivanainen[1], PhD, MD; Reetta Arokoski[2], MSc; Santeri Mentu[2], MSc; Laura Lang[2], MSci; Jussi Ekström[2], PhD; Henri Virtanen[2], BDSc; Vesa Kataja[2], MD, PhD; Jussi Pekka Koivunen[1], MD, PhD

[1]Department of Oncology and Radiotherapy, Oulu University Hospital, Kajaanintie 50, Oulu, Finland

[2]Kaiku Health Oy, Helsinki, Finland

Corresponding Author:
Sanna Iivanainen, PhD, MD
Department of Oncology and Radiotherapy, Oulu University Hospital, Kajaanintie 50, Oulu, Finland

Abstract

Background: Chemotherapy cycle prescription is generally carried out through a multistep manual process that is prone to human error. Clinical decision support tools can provide patient-specific assessments that support clinical decisions, improve prescribing practices, and reduce medication errors.

Objective: We hypothesized that a knowledge-based, patient-derived, evidence-directed decision support tool consisting of multiple modules focusing on the core duties preceding chemotherapy-cycle prescription could result in a more cost-effective and error-free approach and streamline the workflow.

Methods: A 1-arm, multicenter, prospective clinical trial ("Follow-up of Cancer Patients Receiving Chemotherapy or Targeted Therapy by Electronic Patient Reported Outcomes-tool" [ECHO] 7/2019-1/2021; NCT04081558) was initiated to investigate the tool. The most important inclusion criteria were the presence of colorectal cancer (CRC) treated with oxaliplatin-based chemotherapy, age ≥18 years, Eastern Cooperative Oncology Group [ECOG] performance score of 0 to 2, and internet access. A decision support tool that included digital symptom monitoring, a laboratory value interface, and treatment schedule integration for semiautomated chemotherapy cycle prescribing was integrated into the care pathway. Performance was assessed by the percentage of chemotherapy cycles with sent and completed symptom questionnaires, while perceptions of health care professionals (HCPs) on the feasibility of the approach were collected through a 1-time semistructured interview.

Results: The ECHO trial included 43 patients with CRC treated with doublet or triplet chemotherapy in an adjuvant or metastatic setting. Altogether, 843 electronic patient-reported outcome (ePRO) symptom questionnaires were completed. Of the 15 recorded symptoms, fatigue (n=446, 52.9%) and peripheral neuropathy (n=429, 50.9%) were reported most often, while 137 grade 3 to 4 symptoms were recorded, of which diarrhea (n=5, 4%) and peripheral neuropathy (n=4, 3%) were the most common. During the study, 339 chemotherapy cycles were prescribed, and for the 77% (n=262) of new chemotherapy cycles, ePRO questionnaire data were available within preset limits (completed within 3 days prior to chemotherapy scheduling) while 65% of the cycles (n=221) had symptom questionnaire grading at ≤1%, and 67% of the cycles (n=228) had laboratory values in a preset range. The recommendations by the tool for a new chemotherapy cycle were tier 1 (green; meaning "go") in 145 (42.8%) of the cycles, tier 2 (yellow; "evaluate") in 83 (25%), and tier 3 (red; "hold") in 111 (32.7%). HCPs (n=3) were interviewed with a questionnaire (comprising 8 questions), revealing that they most valued the improved workflow, faster patient evaluation, and direct messaging option.

Conclusions: In this study, we investigated the feasibility of a decision support system for chemotherapy-cycle pre-evaluation and prescription that was developed for the prospective ECHO trial. The study showed that the functionalities of the investigated tool were feasible and that an automated approach to chemotherapy-cycle prescription was possible for nearly half of the cycles.

Trial Registration: ClinicalTrials.gov NCT04081558; https://clinicaltrials.gov/study/NCT04081558

(JMIR Form Res 2025;9:e62749) doi:10.2196/62749

KEYWORDS

cancer; chemotherapy; ePRO; electronic patient-reported outcome; decision support system

Introduction

In health care, digital approaches can potentially improve accessibility, increase the comprehensiveness of care, and streamline processes in a cost-effective manner. Digital tools are especially convenient in facilitating communication independently of time and place, as well as facilitating duties requiring repeated numerical comparison [1-3]. In oncology,

XSL•FO
RenderX

electronic patient-reported outcome (ePRO)–based monitoring has gained a great deal of interest in recent years. ePROs have been shown to improve quality of life (QoL) and survival and reduce the number of unscheduled visits among patients receiving chemotherapy for advanced cancer [4-6]. Reasons for the observed beneficial effects of ePROs are not well known but might be related to urgency algorithms, long-term symptom data, better management of side effects, symptom control, patient empowerment, and facilitated communication between the patient and care team [7-11].

For cytotoxic chemotherapy, the most limiting factors for dose intensity are generally bone marrow function, neuropathy, fatigue, and nausea. Dose intensity typically correlates to treatment outcomes [12], and methods to maintain the intensity level with a personalized approach could improve treatment. For monitoring of dose-intensity limiting factors, both laboratory values and symptom profiles are required.

In oncology units, prescription of a new chemotherapy cycle is based on general laboratory values such as neutrophil counts and transaminases being in the selected reference range and the absence of severe symptoms indicating treatment side effects or cancer progression. In most units, precycle assessment is done by phone-based symptom evaluation and manual comparison of laboratory values to a safety reference. ePRO-collected symptoms can be converted to numerical data, which enables automated comparison to a safety reference scale; the quality of automated comparison surpasses manual review.

Delivering evidence-based, personalized care that involves patients requires profound changes in the traditional, hospital organization–based structure, process, and organization of care, along with clinically validated incentives to support such changes. Health care providers must engage and apply a vast amount of scientific information to provide high-quality patient care. Decision support tools can provide patient-specific assessments that support clinical decisions, improve prescribing practices, and reduce medication errors, and these tools could be part of a solution to the increasing cognitive burden currently placed on clinicians [13].

Clinical decision support systems (CDSSs) have been classified and subdivided into various categories, frequently as knowledge based or non–knowledge based. In knowledge-based systems, outputs are based on rules (IF-THEN statements) relying on literature-based, practice-based, or patient-directed evidence. Similarly, CDSSs that are non–knowledge based require a data source, but these systems use artificial intelligence (AI) methodology in work-up rather than being programmed to follow expert medical knowledge. Non–knowledge-based CDSSs, although a rapidly growing use case for AI in medicine, come with challenges including problems understanding the logic that the AI uses in creating recommendations (the so-called black box issue) and problems with validated data availability [14,15].

Taking into consideration the present hurdles in the use of non–knowledge-based CDSSs in clinical practice, we chose to develop a practice-based, ePRO-directed multidimensional platform to personalize as well as automate the chemotherapy-prescribing process. We hypothesized that generation of a decision support tool consisting of multiple modules focusing on the core duties preceding a chemotherapy-cycle prescription could be a more cost-effective and error-free approach. The functionalities of the tool would include chemotherapy cycle–based scheduling for data collection and analysis, an ePRO symptom collection–linked urgency algorithm, personalized symptom management, laboratory value interface, and decision support.

Methods

Study Design

A 1-arm, multicenter, prospective clinical trial ("Follow-up of Cancer Patients Receiving Chemotherapy or Targeted Therapy by Electronic Patient Reported Outcomes-tool" [ECHO] 7/2019-1/2021; NCT04081558) investigated the use of a novel ePRO tool, Kaiku Health, in cancer care. The most important patient inclusion criteria were colorectal cancer (CRC), a plan to receive oxaliplatin-based chemotherapy as an adjuvant therapy or in the first- or second-line setting of advanced disease, age ≥18 years, an Eastern Cooperative Oncology Group [ECOG] performance score of 0 to 2, and internet access. A decision support tool consisting of a digital symptom-monitoring tool with an urgency algorithm, laboratory values, and integration of a treatment schedule to automate chemotherapy-cycle prescribing was created for the trial.

The ePRO Tool

The Kaiku Health ePRO follow-up module consists of a questionnaire that assess both the presence and severity of symptoms related to typical side effects of chemotherapy (blood in urine, dysuria, eye symptoms [decreased vision or other symptoms], peripheral sensory neuropathy, pain, constipation, cough, decreased appetite, diarrhea, fatigue, fever, mouth sores, nausea, rash or skin changes, shortness of breath, and vomiting). The tool also allows recording the presence of a symptom and a severity algorithm that grades the symptom according to National Cancer Institute Common Terminology Criteria for Adverse Events (NCI-CTCAE). The severity algorithm triggered an email alert to the study physician in the care unit within preset limits (the presence of a grade 3 or higher symptom or a rise in symptom severity from grade 0 to 2). The patients were informed that the ePRO follow-up was intended only for nonurgent communication, and, in critical matters, patients were advised to contact emergency care. In addition, patients received tailored, evidence-based, personalized self-care advice according to electronically reported symptoms and their grade. The tool also included a messaging option through which patients could communicate with their care team directly.

Laboratory Values

Prior to chemotherapy infusion, peripheral blood samples were taken to evaluate bone marrow, kidney, and liver function. The laboratory values include counts for red and white cells, thrombocytes, alanine and aspartate aminotransferases, bilirubin, and creatinine. Based on the summary of product characteristics of the chemotherapeutic agents used by the patient, prespecified laboratory values were set to assess the fitness of the patient for a new course of cancer medical therapy.

XSL•FO
RenderX

The Decision Support Tool

We developed and added a novel element to the Kaiku Health remote symptom-monitoring tool: in addition to collecting and grading ePROs, it automatically collected laboratory values via a novel interface (MyLab). The timing of the ePRO symptom questionnaires and laboratory value collection was automated according to chemotherapy cycles through another interface with a treatment-appointment reservation module (Oberon). The decision support tool analyzed the completed symptom questionnaires and preceding laboratory values and compared the results to a set of institutional reference values. The decision support tool provided a 3-tier recommendation to the care team on the initiation of the preplanned chemotherapy cycle. The 3-tier recommendations included green (meaning “go”), yellow (“evaluate and possibly postpone”), or red (“postpone”). A recommendation of go versus no-go and potential treatment visit details were sent to the patient through the Kaiku Health tool. With a no-go decision, an automated ePRO questionnaire and laboratory value collection for the rescheduled chemotherapy cycle were initiated based on the reservation module interface.

Ethical Considerations

The study was approved by the ethics committee of the Northern Ostrobothnia Hospital District (273/2017) and registered in an international clinical study registry (ClinicalTrials.gov NCT04081558). The study was carried out in accordance with the principles of the Declaration of Helsinki and Good Clinical Practice. All patients provided written informed consent. Study participants were not compensated for their participation in the trial.

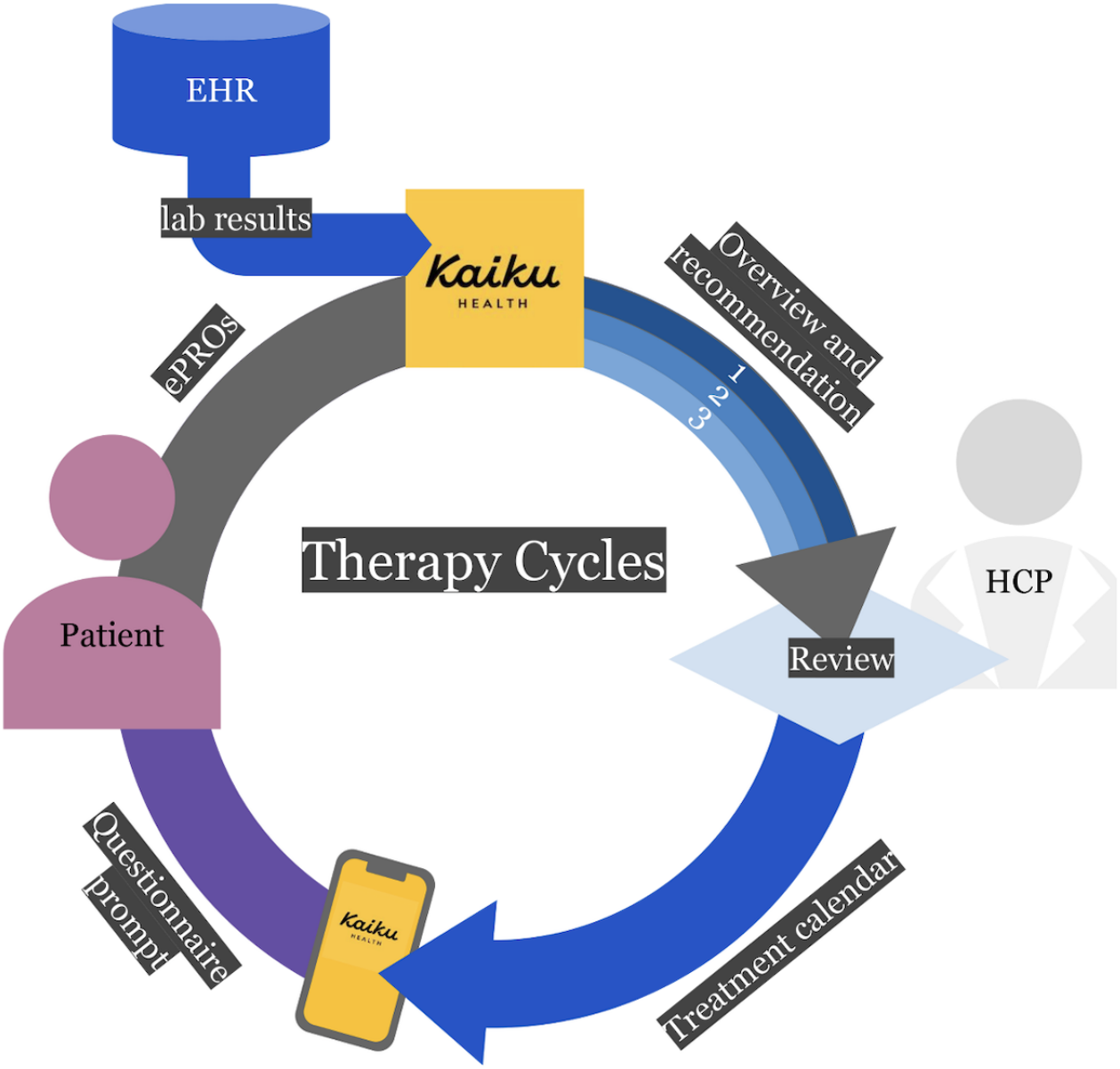
Results

Development of the Decision Support Tool

Before a new chemotherapy cycle, a patient’s laboratory values and symptoms were evaluated, for example, for indications of treatable chemotherapy side effects, excess chemotherapy toxicity, and cancer progression. To streamline this work process, increase the comprehensiveness and repeatability of the precycle evaluation, and automate duties, we developed a decision support tool. The designed tool had 6 different elements that aimed to increase the efficiency and quality of cancer care (Figure 1; Multimedia Appendix 1).

XSL•FO
RenderX

Figure 1. Flowchart of the methodological process of the integration of electronic patient-reported outcomes (ePROs) and laboratory values. The tool included a scheduling module that synchronized the sending of symptom questionnaires to patients via Kaiku Health and laboratory value collection from the hospital's electronic health record (EHR) via an interface. The tool also included a symptom module with an ePRO-based questionnaire for 15 core symptoms. An evaluation module compared the symptoms and laboratory values to values in a set range and assigned a recommendation in 3 tiers (1=go; 2=hold; 3=postpone) for the chemotherapy cycle. HCP: health care professional.

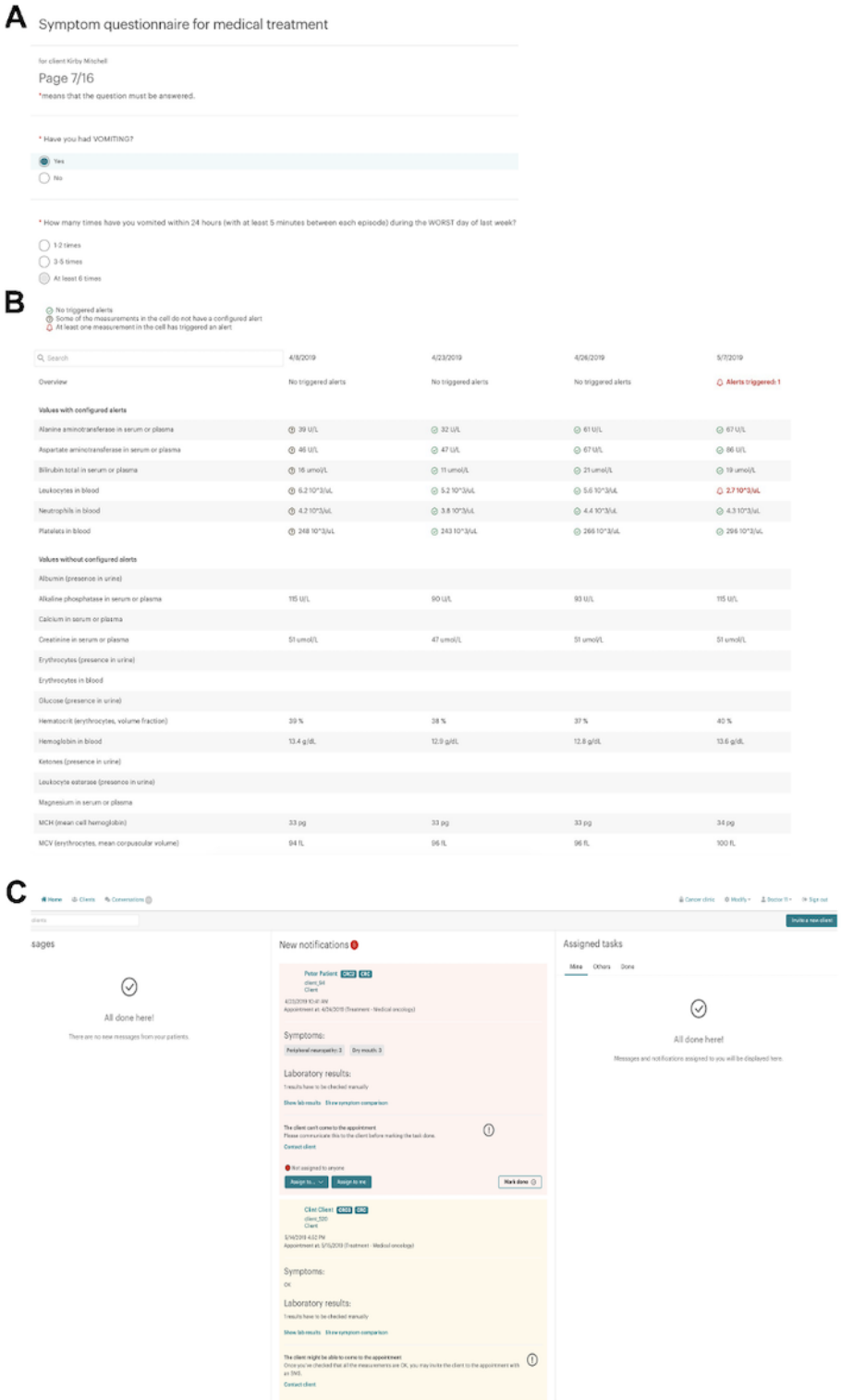


The first element was a scheduling module that synchronized the sending of symptom questionnaires to patients via Kaiku Health and laboratory value collection from hospital registries via an interface. The symptom module used ePRO-based (Kaiku Health) grading (NCI-CTCAE) for 15 core symptoms. The laboratory-value module collected specific central values via a software interface (Figure 2A). The evaluation module compared the symptoms and laboratory values to values in a set range and assigned a recommendation for the chemotherapy cycle in 1 of 3 tiers (Figure 2B). For the tier-1 (green/go) recommendation, laboratory values needed to be within a set range and all the symptoms needed to be at grade ≤1; thus, the nurse made the planned chemotherapy order to the hospital pharmacy, and an SMS text message with information on the timetable of drug infusion was sent through Kaiku Health. For the tier-2 (yellow/evaluate) recommendation, laboratory values had to be within a set range and ePROs had to include ≥1 symptom that was grade ≥2 (if they were not already at baseline); the nurse individually assessed the symptoms. If there was a need for cancer drug modification, typically dose reduction or supportive medications, the study physician was consulted as to whether the chemotherapy cycle could be initiated or modified, or if supportive medication should be initiated or modified. In the case of a "go" decision, confirmation of the treatment time was sent via Kaiku Health. In the tier-3 (red/hold) scenario, laboratory values were out of the set range, and the nurse postponed the chemotherapy cycle and sent the new infusion date through Kaiku Health. In the case of a postponed chemotherapy cycle, the scheduling module programmed a new

XSL•FO
RenderX

collection of symptoms and laboratory values according to the new infusion date.

Figure 2. Overview of the core functionalities of the decision support tool. (A) Electronic patient-reported outcome symptom questionnaire. (B) Laboratory integration module and its output. (C) Decision support tool recommendation.



The care team could always override the recommendation of the decision support system. In the case of missing ePRO or laboratory data (or if the data were >3 days old) before the chemotherapy cycle, the decision support tool gave a tier-2 (yellow/evaluate) recommendation; thus, manual evaluation had to be done prior to prescribing the treatment. The SMS messaging module was used to inform the patient about the recommendation and potential treatment decision. Furthermore, the messaging module enabled secure messaging between the patient and care team. The symptom support module also provided the patient with personalized guidance for self-care based on the presence of specific symptoms and their grade (Figure 2C).

XSL•FO
RenderX

Performance of Decision Support Tool–Directed Chemotherapy Scheduling

Next, we analyzed the performance of the decision support tool based on data collected in a prospective, multicenter, single-arm clinical trial (ECHO; NCT04081558). Patient recruitment for the prospective cohort took place from July 2019 to January 2021 at the Oulu University Hospital and Vaasa Central Hospital oncology units. The study dataset included patients with CRC (n=43) in the ECHO trial who gave informed consent and were treated with doublet or triplet chemotherapy in an adjuvant or metastatic setting. During the study, a total of 339 chemotherapy cycles were prescribed. Altogether, 843 ePRO symptom questionnaires were completed during the ECHO study. Of the recorded 15 symptoms, fatigue (n=446, 52.9%) and peripheral neuropathy (n=429, 50.9%) were reported most often, while 137 grade 3 or 4 symptoms were recorded, of which diarrhea (n=5, 4%) and peripheral neuropathy (n=4, 3%) were the most common. Since the functioning of the Kaiku Health symptom tool has been evaluated in multiple previous trials [16-18], we focused here on the performance of the scheduling and evaluation modules.

The function of the scheduling module was assessed by evaluating the percentage of chemotherapy cycles with sent and completed symptom questionnaires. There were no false-positive or false-negative decision support tool recommendations. For over two-thirds (n=262, 77%) of the cycles, ePRO questionnaire data were available. Preceding a new chemotherapy cycle, 65% (n=221) of cycles had symptom questionnaires graded at ≤1, while 67% (n=228) of the cycles had laboratory values in the preset range. Evaluation module recommendations were tier 1 (green/go) in 145 (42.8%) of the cycles, tier 2 (yellow/evaluate) in 83 (25%), and tier 3 (red/hold) in 111 (32.7%), while 61% of the patients (n=27) needed a phone call before any of the planned chemotherapy cycles (Table 1).

Table . Performance metrics of the developed decision support tool for automated chemotherapy-cycle prescription. The feasibility of the scheduling module was assessed by evaluating the percentage of chemotherapy cycles with sent and completed symptom questionnaires and the availability of laboratory work. The 3-tier recommendation system operated as follows: tier 1 indicated “go” (the nurse proceeded with the initial treatment plan); tier 2 indicated “hold” (the laboratory values had to be within a set range, while the reported symptoms [at least one] were grade ≥2; the study physician was consulted); and tier 3 indicated “postpone” (electronic patient-reported outcomes were acceptable, but laboratory values were outside the preset limits; the nurse scheduled the laboratory work and postponed the infusion).

Criteria	Cycles (n=339), n (%)
Was an automated symptom questionnaire sent and completed prior to the chemotherapy cycle?	
Yes	262 (77.3)
No	77 (23)
Was the symptom questionnaire grading in the acceptable range prior to the planned chemotherapy cycle?	
Yes	221 (65.2)
No	118 (34.8)
Was the laboratory work in the acceptable range prior to the planned chemotherapy cycle?	
Yes	228 (67.3)
No	111 (32.7)
Evaluation module guidance (tier)	
1	145 (42.8)
2	83 (25)
3	111 (32.7)

The experience of the health care professionals (HCPs) with the decision support system integration into the care pathway was evaluated via interviews (n=4) that explored 5 topics: improved workflow, improved care, use of resources, challenges, and points for improvement. The HCPs appreciated the improved workflow, faster patient evaluation, and direct messaging option. The laboratory module originally collected data only daily, which was considered too infrequent by the HCPs, and data collection was reprogrammed to occur multiple times a day, which improved the usability of the tool. Some inconveniences were experienced with the scheduling module for patients with Baxter pump–based treatments, for whom 2 subsequent treatment reservations were made at once (treatment and pump removal); this led to incorrect scheduling of the symptom questionnaires. The HCPs felt that the messaging option should include read notifications, which would enable verification that the patient was aware of the chemotherapy schedule.

Discussion

Principal Findings

We present the results of a feasibility study investigating an integrated approach with an automated decision support system for chemotherapy-cycle prescription. We show that the investigated multidimensional integrated system is feasible and

XSL•FO
RenderX

can provide clinically meaningful decision support. The approach could decrease the HCP workload and increase the quality and safety of cancer care.

Comparison to Prior Work

In recent years, the use of ePRO symptom data has gained a great deal of interest in the oncology field. Previous studies have shown that ePRO monitoring can reduce emergency room visits, improve QoL, and, more importantly, increase survival [4-6]. Integration of an ePRO system into electronic patient records is considered essential for large-scale adoption of the approach. Timely integration of data from multiple electronic health care systems into ePROs could further enhance the clinical value of the captured data and facilitate decision support.

AI-based approaches can be superior for automatic detection, classification, and interpretation using big data. However, AIs can support improper decisions if they are presented with data beyond their original training set data. One could speculate that AI-based decision support systems could surpass our approach with a simple comparison to reference values. However, chemotherapy has a narrow therapeutic window and relying on prescription decisions made by AI alone may not satisfy ethical requirements without wide-scale training and validation [19-21].

A fast track for AI integration in clinical practice could lie in the use of large language models (LLMs). Methodological developments in LLMs such as in-context learning and retrieval-augmented generation could improve LLM performance and reduce hallucinations, consequently making the use of LLMs possible in clinical practice. It has been suggested that the recent development of advanced LLMs could improve patient care across several areas, such as clinical decision support or helping to answer patients' questions, and could even improve patient outcomes by identifying social determinants of health and reducing health disparities [22-24].

Strengths and Limitations

In this study, we investigated a next-generation approach aiming to streamline and automate the prescription of chemotherapy cycles. We identified symptom profiles and laboratory values as central decision mediators for safe chemotherapy prescription. We hypothesized that symptom profiles could be collected via ePROs in a more comprehensive manner and used as input for a decision support tool that would also collect laboratory data and schedule collections based on the expected cycle date in an automated fashion. Furthermore, our approach included an expanded ePRO module with an integrated urgency algorithm, symptom-based personalized feedback and self-care advice, and a safe messaging platform.

We were able to show that the investigated platform is technically feasible, with good user experiences among patients and HCPs. More importantly, we found that automated "go" decisions for new chemotherapy cycles were made in almost half of the chemotherapy cycles. It is likely that this proportion could be increased by focusing only on core symptoms, instead of the total of 17 symptoms that were recorded in this study. We speculate that the investigated tool would decrease the workload for HCPs and enable allocation of human resources to more value-creating duties. We strongly believe that the computerized comparison of data to a set reference is more precise and error-free than human assessment, even though our study did not provide any metrics on the specific matter.

Nevertheless, a few modifications to the platform were suggested by end users. First, in the testing phase, the frequency of laboratory data collection was already considered too low; thus, the laboratory module was reprogrammed to search for data multiple times a day, which greatly improved the usability of the subtool. Second, the HCPs felt that the messaging option regarding the schedule of the planned chemotherapy cycle should include read notifications, which would enable verification that the patient was aware of the coming cycle. Furthermore, due to the low number of study participants and the observational nature of the study, the results are to be interpreted with caution, and the generalizability of the feasibility of the studied platform is unclear.

Future Directions

Due to the feasibility design of the study, the results are generally hypothesis generating. Since the study investigated a complex technical tool, initial testing with a feasibility approach is therefore warranted before progressing to a larger clinical study. The integration of an LLM-based module for interactivity, such as a chat box, could enhance the supportive role of the developed tool by providing educational elements to empower patients as well as give guidance on milder-grade chemotherapy toxicity management on a more individual level. Furthermore, the rather large proportion of symptom questionnaires with grade 2 symptoms might be decreased by fine-tuning the symptom evaluation, possibly by using AI-based models to mimic diagnostic pattern recognition to identify red-flag symptoms in parallel with laboratory value assessment. To conclude, the results of our trial are promising and lay the groundwork for further studies as well as development of the platform. A randomized study design would be required to fully validate the clinical utility and cost-effectiveness of the system.

Acknowledgments

Generative artificial intelligence tools were not used in preparation of this manuscript. The study was funded by Roche Diagnostics. Kaiku Health employees were involved in data acquisition and analysis.

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

SI, JPK, and RA are the guarantors of integrity of the entire study. SI, JPK, and RA contributed to study conceptualization and design. SI and JPK performed the literature search. SI, JPK, LL, JE, HV, and VK contributed to the clinical studies. SI, JPK, RA, SM, LL, and JE contributed to the experimental studies and data analysis. SI, JPK, SM, and JE performed the statistical analysis. SI, JPK, RA, SM, JE, LL, HV, and VK prepared the manuscript. SI, JPK, RA, SM, JE, LL, HV, and VK edited the manuscript. All authors read and approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Kaiku Bridge combines the Kaiku Health symptom data from the electronic patient-reported outcome platform and laboratory value data from the hospital records (Effica) according to the scheduled care events (Oberon). After evaluating both symptoms and laboratory values within predefined limits, the Kaiku interface provides guidance in 3 tiers on new chemotherapy-cycle prescription.
[PNG File, 89 KB - formative_v9i1e62749_app1.png]

References

1. Holch P, Warrington L, Bamforth LCA, et al. Development of an integrated electronic platform for patient self-report and management of adverse events during cancer treatment. Ann Oncol 2017 Sep 1;28(9):2305-2311. [doi: 10.1093/annonc/mdx317] [Medline: 28911065]
2. Kotronoulas G, Kearney N, Maguire R, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. J Clin Oncol 2014 May 10;32(14):1480-1501. [doi: 10.1200/JCO.2013.53.5948] [Medline: 24711559]
3. Lizée T, Basch E, Trémolières P, et al. Cost-effectiveness of web-based patient-reported outcome surveillance in patients with lung cancer. J Thorac Oncol 2019 Jun;14(6):1012-1020. [doi: 10.1016/j.jtho.2019.02.005] [Medline: 30776447]
4. Basch E, Deal AM, Dueck AC, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. JAMA 2017 Jul 11;318(2):197-198. [doi: 10.1001/jama.2017.7156] [Medline: 28586821]
5. Denis F, Lethrosne C, Pourel N, et al. Randomized trial comparing a web-mediated follow-up with routine surveillance in lung cancer patients. J Natl Cancer Inst 2017 Sep 1;109(9):9. [doi: 10.1093/jnci/djx029] [Medline: 28423407]
6. Denis F, Yossi S, Septans AL, et al. Improving survival in patients treated for a lung cancer using self-evaluated symptoms reported through a web application. Am J Clin Oncol 2017 Oct;40(5):464-469. [doi: 10.1097/COC.0000000000000189] [Medline: 25811297]
7. Osborn J, Ajakaiye A, Cooksley T, Subbe CP. Do mHealth applications improve clinical outcomes of patients with cancer? A critical appraisal of the peer-reviewed literature. Support Care Cancer 2020 Mar;28(3):1469-1479. [doi: 10.1007/s00520-019-04945-4] [Medline: 31273501]
8. Moradian S, Voelker N, Brown C, Liu G, Howell D. Effectiveness of internet-based interventions in managing chemotherapy-related symptoms in patients with cancer: a systematic literature review. Support Care Cancer 2018 Feb;26(2):361-374. [doi: 10.1007/s00520-017-3900-8] [Medline: 28948360]
9. Trajkovic-Vidakovic M, de Graeff A, Voest EE, Teunissen S. Symptoms tell it all: a systematic review of the value of symptom assessment to predict survival in advanced cancer patients. Crit Rev Oncol Hematol 2012 Oct;84(1):130-148. [doi: 10.1016/j.critrevonc.2012.02.011] [Medline: 22465016]
10. Bennett AV, Jensen RE, Basch E. Electronic patient-reported outcome systems in oncology clinical practice. CA Cancer J Clin 2012;62(5):337-347. [doi: 10.3322/caac.21150] [Medline: 22811342]
11. Benze G, Nauck F, Alt-Epping B, et al. PROutine: a feasibility study assessing surveillance of electronic patient reported outcomes and adherence via smartphone app in advanced cancer. Ann Palliat Med 2019 Apr;8(2):104-111. [doi: 10.21037/apm.2017.07.05] [Medline: 29156896]
12. Lyman GH. Impact of chemotherapy dose intensity on cancer patient outcomes. J Natl Compr Canc Netw 2009 Jan;7(1):99-108. [doi: 10.6004/jnccn.2009.0009] [Medline: 19176210]
13. Walsh S, de Jong EEC, van Timmeren JE, et al. Decision support systems in oncology. JCO Clin Cancer Inform 2019 Feb;3(3):1-9. [doi: 10.1200/CCI.18.00001] [Medline: 30730766]
14. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17. [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
15. Nafees A, Khan M, Chow R, et al. Evaluation of clinical decision support systems in oncology: an updated systematic review. Crit Rev Oncol Hematol 2023 Dec;192:104143. [doi: 10.1016/j.critrevonc.2023.104143] [Medline: 37742884]

XSL•FO
RenderX

16. Iivanainen S, Alanko T, Peltola K, et al. ePROs in the follow-up of cancer patients treated with immune checkpoint inhibitors: a retrospective study. J Cancer Res Clin Oncol 2019 Mar;145(3):765-774. [doi: 10.1007/s00432-018-02835-6] [Medline: 30666409]
17. Iivanainen S, Alanko T, Vihinen P, et al. Follow-up of cancer patients receiving anti-PD-(L)1 therapy using an electronic patient-reported outcomes tool (KISS): prospective feasibility cohort study. JMIR Form Res 2020 Oct 28;4(10):e17898. [doi: 10.2196/17898] [Medline: 33112242]
18. Iivanainen SME, Ravi R, Jekunen A, et al. ePRO follow-up of colorectal cancer patients receiving oxaliplatin-based adjuvant chemotherapy is feasible and enhances the quality of patient care. Ann Oncol 2021 Sep;32:S1184-S1185. [doi: 10.1016/j.annonc.2021.08.1667] [Medline: 36809503]
19. Castaneda C, Nalley K, Mannion C, et al. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. J Clin Bioinforma 2015;5(1):4. [doi: 10.1186/s13336-015-0019-3] [Medline: 25834725]
20. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. JAMA 2018 Dec 4;320(21):2199-2200. [doi: 10.1001/jama.2018.17163] [Medline: 30398550]
21. Simon G, DiNardo CD, Takahashi K, et al. Applying artificial intelligence to address the knowledge gaps in cancer care. Oncologist 2019 Jun;24(6):772-782. [doi: 10.1634/theoncologist.2018-0257] [Medline: 30446581]
22. Iivanainen S, Lagus J, Viertolahti H, Sippola L, Koivunen J. Investigating large language model (LLM) performance using in-context learning (ICL) for interpretation of ESMO and NCCN guidelines for lung cancer. J Clin Oncol 2024 Jun 1;42(16_suppl):e13637-e13637. [doi: 10.1200/JCO.2024.42.16_suppl.e13637]
23. Chen D, Huang RS, Jomy J, et al. Performance of multimodal artificial intelligence chatbots evaluated on clinical oncology cases. JAMA Netw Open 2024 Oct 1;7(10):e2437711. [doi: 10.1001/jamanetworkopen.2024.37711] [Medline: 39441598]
24. Guevara M, Chen S, Thomas S, et al. Large language models to identify social determinants of health in electronic health records. NPJ Digit Med 2024 Jan 11;7(1):6. [doi: 10.1038/s41746-023-00970-0] [Medline: 38200151]

Abbreviations

AI: artificial intelligence
CDSS: clinical decision support system
CRC: colorectal cancer
ECHO: Follow-up of Cancer Patients Receiving Chemotherapy or Targeted Therapy by Electronic Patient Reported Outcomes-tool
ECOG: Eastern Cooperative Oncology Group
ePRO: electronic patient-reported outcome
HCP: health care professional
LLM: large language model
NCI-CTCAE: National Cancer Institute Common Terminology Criteria for Adverse Events
QoL: quality of life

Edited by A Mavragani; submitted 30.05.24; peer-reviewed by JCL Chow, SS Goswami; revised version received 12.02.25; accepted 13.02.25; published 31.03.25.

Please cite as:
Iivanainen S, Arokoski R, Mentu S, Lang L, Ekström J, Virtanen H, Kataja V, Koivunen JP
Development of a Comprehensive Decision Support Tool for Chemotherapy-Cycle Prescribing: Initial Usability Study
JMIR Form Res 2025;9:e62749
URL: https://formative.jmir.org/2025/1/e62749
doi:10.2196/62749

© Sanna Iivanainen, Reetta Arokoski, Santeri Mentu, Laura Lang, Jussi Ekström, Henri Virtanen, Vesa Kataja, Jussi Pekka Koivunen. Originally published in JMIR Formative Research (https://formative.jmir.org), 31.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Barriers and Determinants of Referral Adherence in AI-Enabled Diabetic Retinopathy Screening for Older Adults in Northern India During the COVID-19 Pandemic: Mixed Methods Pilot Study

Anshul Chauhan[1], PhD; Anju Goyal[2], PhD; Ritika Masih[1], MPH; Gagandeep Kaur[1], B Optometry; Lakshay Kumar[1], B Optometry; Neha[1], B Optometry; Harsh Rastogi[1], MSc; Sonam Kumar[1], MPH; Bidhi Lord Singh[1], PhD; Preeti Syal[2], MS; Vishali Gupta[1], MS; Luke Vale[3], PhD; Mona Duggal[1,4], MD

[1]Advanced Eye Centre, Post Graduate Institute of Medical Education and Research, Number 139, Kairon Block, Chandigarh, India

[2]Department of Statistics, Panjab University, Chandigarh, India

[3]Newcastle University, Newcastle, United Kingdom

[4]National Institute for Research in Digital Health and Data Sciences, Indian Council of Medical Research, New Delhi, India

Corresponding Author:
Mona Duggal, MD
Advanced Eye Centre, Post Graduate Institute of Medical Education and Research, Number 139, Kairon Block, Chandigarh, India

Abstract

Background: Diabetic retinopathy (DR) is a leading cause of blindness globally. DR has increasingly affected both individuals and health care systems as the population ages.

Objective: This study aims to explore factors and identify barriers associated with nonadherence to referral recommendations among older adult participants after DR screening (DRS) during the COVID-19 pandemic.

Method: This paper presents findings from a pilot study on artificial intelligence–enabled DRS conducted in two districts in Punjab, India (Moga and Mohali) during the COVID-19 pandemic. The screenings were conducted from March to June 2022 at community health center Badhani Kalan in Moga and from March to June 2021 in community settings (homes) in Block Boothgarh, Mohali. Participants were referred to the district hospital for an ophthalmological review based on artificial intelligence–enabled screening. After 1 month, the participants were contacted by telephone to assess adherence to the referral recommendations. Participants who did not adhere to the referral were then interviewed alongside health care providers to understand the barriers explaining their nonadherence.

Results: We aimed to recruit 346 and 600 older adult participants from 2 sites but enrolled 390. Key challenges included health facility closures due to COVID-19, low motivation among health personnel for recruitment, incomplete nonparticipation data, and high participant workloads. Approximately 45% of the participants were male and 55% female. Most participants (62.6%) were between 60 and 69 years old, while 37.4% were 70 or older, with a mean age of 67.2 (SD 6.2) years. In total, 159 participants (40.8%) were referred, while 231 participants (59.2%) were not. Only 23 (14.5%) of those referred followed through and visited a health facility for ophthalmological review, while 136 (85.5%) did not pursue further evaluation. Our analysis revealed no significant differences in the characteristics between adherent and nonadherent participants, suggesting that demographic and health factors alone do not predict adherence behavior in patients with DR. Interviews identified limited knowledge about DR, logistical challenges, financial constraints, and attitudinal barriers as the primary challenges.

Conclusions: This study, conducted during the COVID-19 pandemic, showed suboptimal adherence to referral recommendations among older adult patients due to knowledge gaps, logistical challenges, and health system issues. Quantifying and understanding adherence factors are crucial for targeted interventions addressing barriers to referral recommendations after DRS. Integrating teleophthalmology into and strengthening infrastructure for artificial intelligence–enabled diabetic retinopathy screening to enhance access and outcomes.

(JMIR Form Res 2025;9:e67047) doi:10.2196/67047

KEYWORDS

diabetic retinopathy; diabetes; gerontology; geriatric; old; aging; aged; artificial intelligence; retinopathy; retinal; referral; screening; optometry; ophthalmology; adherence; barriers

XSL•FO
RenderX

Introduction

Diabetic retinopathy (DR) is a prevalent and severe microvascular complication of diabetes mellitus (DM) [1]. In India, 10.9% (7.2% - 16.3%) of individuals aged 65 years and above with diabetes have DR, with 2.3% (1.2% - 4.4%) suffering from vision-threatening diabetic retinopathy (VTDR) [1], characterized by severe retinopathy or macular edema [2]. DR is typically asymptomatic in its early stages, and it can lead to visual impairment or blindness if left untreated [3]. The rate of blindness due to VTDR is expected to rise proportionately to the exponential increase in DM prevalence [4,5]. Visual impairment due to DR increases by 9.6% in individuals over 50 [6], emphasizing advancing age as a significant risk factor [7]. Individuals with VTDR face greater difficulties in vision-related tasks, such as reading, watching television, and driving [8]. Additionally, people with DR have a higher risk of falls, resulting in reduced mobility, social isolation, and increased physical dependence [9].

India is set to become home to the world's second-largest older adult population, with the number of people over 60 years of age projected to rise from 100 million in 2013 to 198 million by 2030 [10]. The National Programme for Health Care of the Elderly in India advocates for accessible, affordable, high-quality, long-term, comprehensive care services tailored to the needs of an aging population [10]. Significant scientific evidence shows that early screening and timely treatment referral can prevent most visual loss caused by DR [11]. Conventionally, DR screening (DRS) includes fundus (retina) examination by ophthalmologists or color fundus photography using conventional cameras (mydriatic or nonmydriatic) conducted by trained eye technicians or optometrists [12]. However, the sharp rise in diabetes cases, coupled with a shortage of trained retinal specialists and ophthalmologists [13], makes DRS services accessibility challenging, particularly as 80% of India's older adult population resides in rural areas [10]. The COVID-19 pandemic posed additional significant challenges to health care systems worldwide, inevitably leading to the curtailment of health services accessibility, including DRS [14,15].

Automated computer-based analysis of fundus images could ease the strain on health care systems by streamlining DRS and offering a more efficient management solution [13]. Artificial intelligence (AI) systems have reduced costs, enhanced diagnostic precision, and expanded patient accessibility to DRS services [16] and immediate AI-supported feedback on referral status was linked to higher referral adherence [17].

Previous studies have demonstrated telemedicine-based DRS programs' effectiveness, scalability, and sustainability in improving accessibility and reducing disparities [18]. The COVID-19 pandemic further emphasized the need for AI-enabled DRS and teleconsultations to standardize workflows and reduce human interactions to mitigate transmission risks [19,20].

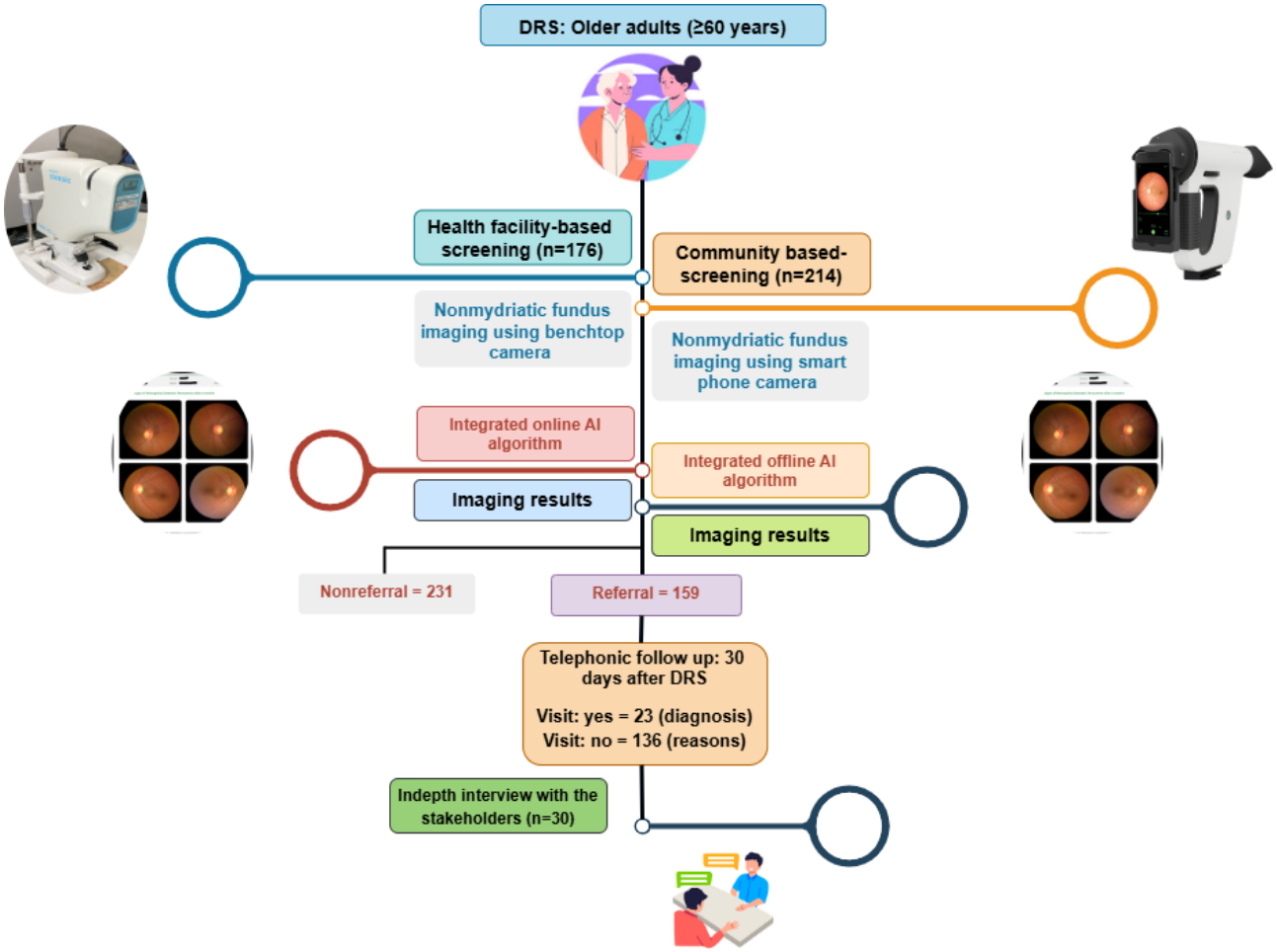
These studies are particularly relevant for designing and implementing effective AI-enabled DRS in India, where challenges of health care accessibility and resource constraints exist [20]. Suboptimal adherence to follow-up recommendations impacts treatment outcomes and can undermine even the most effective treatment options [21]. A multicenter analysis showed a decline in laser procedures and vitrectomies for retinal detachment repair due to reduced clinic visits during the pandemic [22]. Patients faced barriers to follow-up eye exams, including transportation issues, high costs, and long wait times [23]. Programs such as a point-of-care DR examination program highlight the potential of telemedicine in reducing barriers to DRS, though adherence to follow-up recommendations after screening remains a major challenge [24].

However, some factors (individual, health) influencing adherence and nonadherence to referral services remain unknown [25]. During the COVID-19 pandemic, patients were prompted to limit hospital visits, minimize time spent there, and reduce direct contact with health care providers (HCPs). Limited research exists in India on older adults' adherence to DRS referral recommendations and the barriers influencing their decisions.

This study aims to (1) evaluate the referral adherence rates among older adult participants after AI-enabled DRS and identify the key factors affecting adherence during the COVID-19 pandemic and (2) identify and analyze the barriers to adherence to referral recommendations during the COVID-19 pandemic. The overall study design is described in Figure 1.

XSL•FO
RenderX

Figure 1. Overall study design. AI: artificial intelligence; DRS: diabetic retinopathy screening.



Methods

Study Setting and Participants

Quantitative

This paper presents findings from 2 districts in Punjab, India: Moga and Mohali. Screenings in Moga occurred at the community health center (CHC) Badhani Kalan from March to June 2022, while in Mohali, they were conducted in community settings (home) in Block Boothgarh from March to June 2021. The study included older adults [10] aged 60 years and above with a history of DM at both sites. Participants without a history of DM, those unwilling to provide informed consent, and those with a history of intraocular surgery, conjunctivitis, red eye, injury, or any eye inflammation were excluded.

A nurse at the medicine clinic at the CHC, Badhani Kalan, in district Moga, identified the elderlyolder adult participants with DM for DRS. In district Mohali, the list of diabeticparticipants with diabetes was obtained from thePa primary Hhealth Ccentere (PHC), in Boothgarh, and approached with the help of the village's Accredited Social and Health Activists (ASHA) worker. The participants agreeing to undergo DRS received an appointment card, and a reminder telephone call was placed a day before the screening to confirm their attendance at home.

The repeated lockdowns during the COVID-19 pandemic disrupted health care services, significantly increasing the burden on health systems. Consequently, recruitment efforts were hindered, making it difficult to reach the target sample sizes of 348 at CHC Badhani Kalan and 600 at Block Boothgarh. Since the populations shared similar health systems and implementation challenges, the data from both sites were pooled and presented. Consequently, data pooling enhanced sample size and statistical power and effectively addressed research questions [26].

Qualitative

The in-depth interviews were conducted with 2 distinct categories of participants: older adults and HCPs. Older adult participants who did not adhere to the recommended referral instructions were purposely sampled and approached for in-depth interviews to understand their barriers to referral adherence after DRS. The nurse at the ophthalmology clinic invited older adult participants to the CHC for in-depth interviews, while ASHA workers contacted older adult participants in the community to conduct interviews at homes.

The second group of participants, HCPs, were purposely selected according to their roles in diabetes and DR diagnosis, referral, and treatment (Table S1 in Multimedia Appendix 1). The inclusion criteria were having at least 3 years of experience working in public health care settings and a willingness to describe patient experiences and barriers to referral adherence (Table S2 in Multimedia Appendix 1). The HCPs included ASHAs from the community settings [27], Community Health

XSL•FO
RenderX

Officers (CHOs) from health and wellness centers [28], optometrists from a CHC [29] , ophthalmologists from the district hospital [30], and retina specialists from the retina unit of a tertiary eye care hospital.

Semistructured interviews were conducted to explore participants perspectives on the barriers to adhering to referral recommendations following AI-enabled DRS. The in-depth interview guides were developed in English, informed by existing literature on barriers to DR screening and referral [31-34], and later translated into Hindi and Punjabi (Table S3 in Multimedia Appendix 1). These guides, featuring open-ended questions and probes, explored participants' perspectives on referral adherence after DRS. Pilot interviews with 2 people with DM, an ASHA, and a CHO led to revisions for clarity, replacing ambiguous terms with simpler language for better participant understanding.

Ethical Considerations

Written informed consent was obtained from all participants, with both studies approved by the Institutional Ethics Committee of the Post Graduate Institute of Medical Education and Research, Chandigarh, India (PGI/IEC/2020/000741 and PGI/IEC/2020/001342). To ensure patient privacy and confidentiality, all images and associated metadata were de-identified prior to analysis. The studies adhered to the Declaration of Helsinki guidelines, and the study protocol is registered with the Clinical Trials Registry, India (CTRI 2022/10/046283 and 2022/10/046185).

Hardware and Software for DRS

Hardware

The data were collected from 2 locations across districts in Punjab, each employing a different DRS methodology. Health facility–based screening at CHCs requires desktop nonmydriatic fundus cameras, while community screenings (home) rely on portable cameras for enhanced accessibility.

Fundus Benchtop Camera

The screening at the CHC Badhani Kalan utilized the Forus 3 Nethra classic (benchtop) camera, supported by a laptop running MS Windows 10 or higher, featuring a 64-bit operating system and an i3 10th-generation processor [35,36].

Smartphone-Based Fundus Camera

The lightweight, portable Fundus on Phone Non Mydriatic 10 (Remidio) handheld device, with a power backup, was used for screening in community settings in Block Boothgarh [37].

Software

Both sites used validated AI screening algorithms for DRS. At the CHC, an online DRS algorithm (Revelo) was integrated into the benchtop fundus camera [38]. A camera-integrated offline AI algorithm (Medios) was used for community-based screening [37,39].

Data Collection

Fundus Photography Protocol

Training Protocol

Before patient recruitment, 2 optometrists underwent 15 days of 1-on-1 training at the Advanced Eye Centre of Post Graduate Institute of Medical Education and Research, Chandigarh, and a primary health care center in Block Boothgarh of district Mohali. The training covered recruitment and consenting procedures, history recording, retinal image acquisition, troubleshooting poor-quality images, automated grading with the AI system, and delivering grading reports to patients. Training and supervision continued until staff were comfortable independently collecting data and performing imaging.

Image Acquisition and Grading

Adequate ventilation was maintained for patient consultations and fundus photography. Optometrists wore N95 masks and followed strict hand hygiene. Patients without masks were required to wear one during interactions and in waiting areas. Optometrists obtained nonmydriatic macula and disc-centered image color fundus photographs at 45° field of view using low-cost cameras (Figure 1) for participants at both sites. The screening at the CHC was carried out in an ophthalmology clinic. On the day of screening, the optometrists were assisted by ASHA workers at the homes of the eligible participants. The AI algorithms provided results at both sites. Two human graders, a certified optometrist and an ophthalmologist, graded all the fundus images. A senior retina specialist resolved any disagreements between the graders. All the classifications of DR were based on the International Classification for Diabetic Retinopathy [40].

Referral and Telephonic Follow-Up

Participants received an AI-generated report detailing their DR status, severity, and referral recommendations. Patients diagnosed with moderate non-proliferative DR and above were referred to as referable DR [41] and, with ungradable images, were advised to visit the district hospital for an ophthalmological review. One month after the screening, participants were followed up with telephone calls to confirm adherence and obtain diagnosis details. Each participant was contacted thrice, on 3 consecutive days, before being classified as nonresponsive. A structured questionnaire was administered to participants via telephone. The questionnaire assessed factors influencing appointment attendance and treatment adherence, with responses recorded as yes, no, or do not know for each question [42,43]. The questionnaire assessed sociodemographic characteristics and perceived barriers to referral (knowledge, attitude, logistics) and included an open-ended question for unlisted reasons behind incomplete referrals.

In-Depth Interviews

Two research fellows (AC and HR) with master's degrees and over 5 years of qualitative research experience conducted the in-depth interviews in the participant's preferred languages (Hindi, Punjabi, and English). Participants with DM were interviewed in quiet hospital locations at the CHC or at home in a properly ventilated space while HCPs were interviewed

XSL•FO
RenderX

face-to-face or via Zoom, based on availability. Study procedures were explained, and informed consent was obtained, either written for in-person interviews or verbal for Zoom sessions. Each interview was audio recorded and lasted for 35 - 40 minutes. The final sample size for the qualitative study was determined based on the principle of data saturation [44,45].

Data Analysis

Using a mixed methods approach, the results from both the quantitative and qualitative components were compared, synthesized, and discussed.

Quantitative

All participant data were entered into Research Electronic Data Capture (REDCap) [46], exported in Excel format, cleaned, and subsequently analyzed using Stata/IC (version 14.2; StataCorp LLC). The analysis involved reporting the descriptive statistics. Histograms were used to assess the normality of continuous variables. Summary statistics include frequencies (percentages) for categorical variables and means (standard deviations) for normally distributed continuous variables; otherwise, medians and interquartile ranges were reported. A chi-square (X^2) test was conducted to compare the distribution of categorical variables (eg, gender, education, marital status) between the adherent and nonadherent participant groups.

Univariate logistic regression was used to estimate odds ratios (ORs) and 95% CIs for the association between each independent variable (eg, gender, age, education, occupation, marital status, income, health insurance, duration of diabetes, hypertension status, and previous eye facility visits) and the dependent variable (referral adherence status: adherent versus nonadherent). The multivariate logistic regression model included independent variables with a P value <.05 in univariate analysis to adjust for potential confounding factors and identify independent predictors of referral nonadherence, using a significance level of P<.05 for the final model.

Qualitative

The qualitative researchers proficient in Hindi and Punjabi transcribed the recordings verbatim, translated them into English, and removed all personal identifiers before analysis. We analyzed the interview transcripts using thematic analysis, following Braun and Clarke's 6-phase approach [47]. Although the phases progressed logically, the process was not linear but recursive and iterative, involving movement back and forth between phases as needed [47]. We systematically organized the data, coded it, and grouped similar codes into themes to identify barriers to referral adherence. The research team members (AC: research scholar; RM: research associate; HR: research scholar) performed independent open transcript coding. MD, an experienced public health specialist, reviewed and confirmed the initial coding framework. Coding and categorization under themes were done using Atlas.Ti 23 software.

We then began extracting relevant portions of text from each interview related to the categories, sorting and selecting quotes and placing them under the appropriate themes. Quotes were edited for readability to retain their original meaning; ellipses (...) show removed text. The participants are denoted in italics with a unique ID number and their group designator: P (people with DM), Opt (Ophthalmologists), ASHA (ASHA workers), CHO (Community Health Officer), Optom (Optometrists), and RS (Retina Specialist). Given that both study sites are in Punjab and share similar health system building blocks [48], the in-depth interviews were analyzed and presented collectively.

Results

Quantitative

Participant Characteristics

A total of 390 (41%) of 948 older adult participants were enrolled across both sites, with 176 (45%) males and 214 (55%) females, primarily due to COVID-19 restrictions and participant refusals. Participants aged 60 - 69 years comprised 62.6244% of the sample, while those aged ≥70 years constituted 37.4%, with a mean age of 67.2 (SD 6.2) years. Around 48.5% had no formal education, 45.9% had education up to the 10th standard, and 5.6% had education at the 12th standard or higher. Occupationally, 69.7% were unemployed or engaged in home duties, 13.8% were retired, and 16.5% were in other categories. Marital status indicated that 74% were married. Additionally, 69% had a monthly household income below 30,000 Indian rupees (INR; approximately US $350), 52.3% had a history of hypertension, and the mean DM duration was 8.03 (SD 6.9) years. Site-wise demographic details are available in Table S4 in Multimedia Appendix 1.

Referral Recommendations and Adherence Rates

Out of 390 screened participants, 231 (59.2%) were not referred, while 159 (40.8%) were referred for further review and management. Among those referred, only 23 (14.5%) followed the advice and visited a health facility (details in Table S5 in Multimedia Appendix 1), while 136 (85.5%) did not adhere to the referral recommendations or seek further medical attention.

Differences Between Referral Categories (Adherent and Nonadherent)

Table 1 shows the demographic and medical characteristic differences between adherent and nonadherent participants in the referral category. The comparison of 159 referral participants showed no statistically significant differences in demographics and health characteristics between adherent (n=23) and nonadherent (n=136) groups (all P>.05). This includes gender distribution (P=.23), age groups (P=.48), educational levels (P=.27), occupational status (P=.33), marital status (P=.97), household income (P=.71), health insurance coverage (P=.93), duration of DM (P=.51), presence of hypertension (P=.16), and previous eye facility visits (P=.45).

XSL•FO
RenderX

Table . Difference in demographics and health characteristics between adherent and nonadherent participants in the referral category.

Variable	Referral category (n=159)		
	Adherent (n=23)	Nonadherent (n=136)	*P* value
Gender, n (%)			.23
Male	12 (52.7)	53 (39)	
Female	11 (47.8)	83 (61)	
Age group (years), n (%)			.48
60 - 69	14 (60.8)	72 (53)	
≥70	9 (39.1)	64 (47)	
Education, n (%)			.27
No formal education	8 (34.8)	72 (53)	
Up to 10th standard	14 (60.8)	59 (43.4)	
12th standard and above	1 (4.4)	5 (3.6)	
Occupation, n (%)			.33
Retired from service	4 (17.4)	15 (11)	
Unemployed/home duties	14 (60.9)	103 (75.8)	
Other	5 (21.7)	18 (13.2)	.97
Marital status, n (%)			
Married	16 (69.5)	94 (69)	
Unmarried, divorced, separated, widow, widower	7 (30.4)	42 (31)	
Monthly household income (INR)[a], n (%)			.71
<30,000	15 (65.2)	94 (69)	
>30,000	8 (34.8)	42 (31)	
Health insurance, n (%)			.93
Yes	8 (34.8)	46 (33.8)	
No	15 (65.2)	90 (66.2)	
Duration of DM[b] (years), n (%)			.51
0 - 10	17 (73.9)	91 (67)	
≥10	6 (26.1)	45 (33)	
Hypertension (years), n (%)			.16
Yes	15 (65.2)	67 (49.2)	
No	8 (34.8)	69 (50.8)	
Previous visit to the eye facility, n (%)			.45
Once	19 (82.6)	120 (88.2)	
Never	4 (17.4)	16 (11.8)	

[a]INR: Indian rupee. 30,000 INR=US $350.

[b]DM: diabetes mellitus.

Analysis of Determinants for Nonadherence to Referral Recommendations

Table 2 compares determinant factors between 23 adherent and 136 nonadherent participants using univariate and multivariate logistic regression analyses. None of the variables showed statistically significant differences (all *P*>.05), indicating no evidence of associations with adherence status across the examined factors. Gender (female) showed nonsignificantly higher odds of nonadherence, for univariate (OR 1.7, 95% CI 0.7-4.2*; P*=.23) and multivariate (OR 1.07, 95% CI 0.31-3.72*; P*=.91). Age ≥70 years had univariate and multivariate ORs of 1.4, 95% CI 0.56-3.4 *(P*=.48) and 1.4, 95% CI 0.5-4.09 *(P*=.47), respectively, indicating no significant association with adherence. Similarly, no significant differences were found in occupation categories, marital status, household income, health

XSL•FO
RenderX

insurance status, duration of DM, hypertension, or previous eye facility visits (all $P>.05$).

Table . Univariate and multivariate analysis of the determinants for nonadherence to referral recommendations. Reference: Base category used for comparison.

Variable	Adherent (n=23)		Nonadherent (n=136)	
	Univariate OR[a] (95% CI)	*P* value	Multivariate OR (95% CI)	*P* value
Gender, n (%)				
Male	Reference		Reference	
Female	1.7 (0.7-4.2)	.23	1.07 (0.31-3.72)	.91
Age group (years)				
60 - 69	Reference		Reference	
≥70	1.4 (0.56-3.4)	.48	1.4 (0.5-4.09)	.47
Education, n (%)				
No formal education	Reference		Reference	
Up to 10th standard	0.47 (0.18-1.19)	.11	0.49 (0.16-1.45)	.20
12th and above	0.5 (0.06-5.36)	.61	0.57 (0.04-7.40)	.67
Occupation, n (%)				
Retired	Reference		Reference	
Unemployed/home	1.9 (0.56-6.7)	.28	1.98 (0.38-10.14)	.41
Other	0.9 (0.21-4.2)	.95	0.98 (0.17-5.6)	.98
Marital status, n (%)				
Married	Reference		Reference	
Unmarried, divorced, separated, widow, widower	1.02 (0.39-2.67)	.96	0.68 (0.23-1.97)	.48
Monthly household income[b], n (%)				
<30,000	Reference		Reference	
>30,000	0.84 (0.33-2.12)	.71	0.94 (0.32-2.77)	.92
Health insurance				
Yes	Reference		Reference	
No	1.3 (0.55-3.4)	.50	0.71 (0.24-2.03)	.52
Duration of diabetes mellitus (years)				
0 - 10	Reference		Reference	
≥10	1.4 (0.52-3.8)	.51	1.56 (0.52-4.68)	.42
Hypertension (years)				
Yes	Reference		Reference	
No	1.9 (0.76-4.8)	.10	2.09 (0.79-5.48)	.13
Previous visit to the eye facility				
Once	Reference		Reference	
Never	0.63 (0.19-2.09)	.46	0.54 (0.14-2.04)	.37

[a]OR: odds ratio.

[b]INR: Indian rupee. 30,000 INR=US $350.

Telephone Follow-Up: Assessing Referral Adherence Rates and Reasons

Participants with ungradable images or diagnosed with referable DR (n=136) were contacted by phone 1 month after screening. The most commonly reported barriers were categorized as other factors (n=70, 51.5%), logistical challenges (n=31, 22.8%), and attitudinal issues (n=27, 19.9%).

XSL•FO
RenderX

Regarding attitude, 17 (12.5%) participants believed their eyes were fine, while logistical obstacles included lack of family support (n=21, 15.4%) and family problems (n=8, 5.8%). Financial constraints, particularly high treatment costs, affected 4 (2.9%) participants. Additionally, 4 (2.9%) were unaware of the importance of DR treatment. Seasonal harvesting (n=14, 10.3%) and other health issues (n=9, 6.6%) were also contributing factors (Table 3).

Table . Reasons for nonadherence to referral recommendations.

Categories and barriers to referral adherence		Participants (n=136), n (%)
Attitude		
	Eyes are fine	17 (12.5)
	Not willing to treatment	3 (2.2)
	Lack of concern about diabetic retinopathy	3 (2.2)
	Not interested in treatment	4 (2.9)
Financial		
	High treatment cost	4 (2.9)
Awareness		
	Not aware that DR treatment is important	4 (2.9)
Logistical support		
	Lack of family support	21 (15.4)
	Travel distance to a health facility	2 (1.6)
	Family problems	8 (5.8)
Other		
	Harvesting seasons	14 (10.3)
	Work commitments	4 (2.9)
	Out of station	6 (4.4)
	Phone with family members	3 (2.2)
	Planning to visit a health facility soon	2 (2.2)
	Other health issues	9 (6.6)
	Weather conditions	1 (0.7)
	Technical error (wrong contact numbers, switch-offs, unavailability)	31 (22.8)

Qualitative

Characteristics of the Participants

In total, 28 in-depth interviews were conducted with 9 older adult people with DM and 19 HCPs. Of the 9 people with DM, 4 (44%) were male and 5 (56%) were female. Most (n=7, 78%) were aged 60 - 69 years, and 8 (89%) were married. Two-thirds (n=6) lacked formal education, and 6 (67%) were unemployed or engaged in home duties (Tables S6 and S7 in Multimedia Appendix 1). The HCP group included 1 retina specialist (5%), 4 ophthalmologists (21%), 2 optometrists (10%), 3 CHOs (16%), and 9 ASHA workers (47%). Among 19 health care providers, 13 (68%) were female and 6 (32%) were male. Most (n=12, 63%) were aged 31 - 40 years.

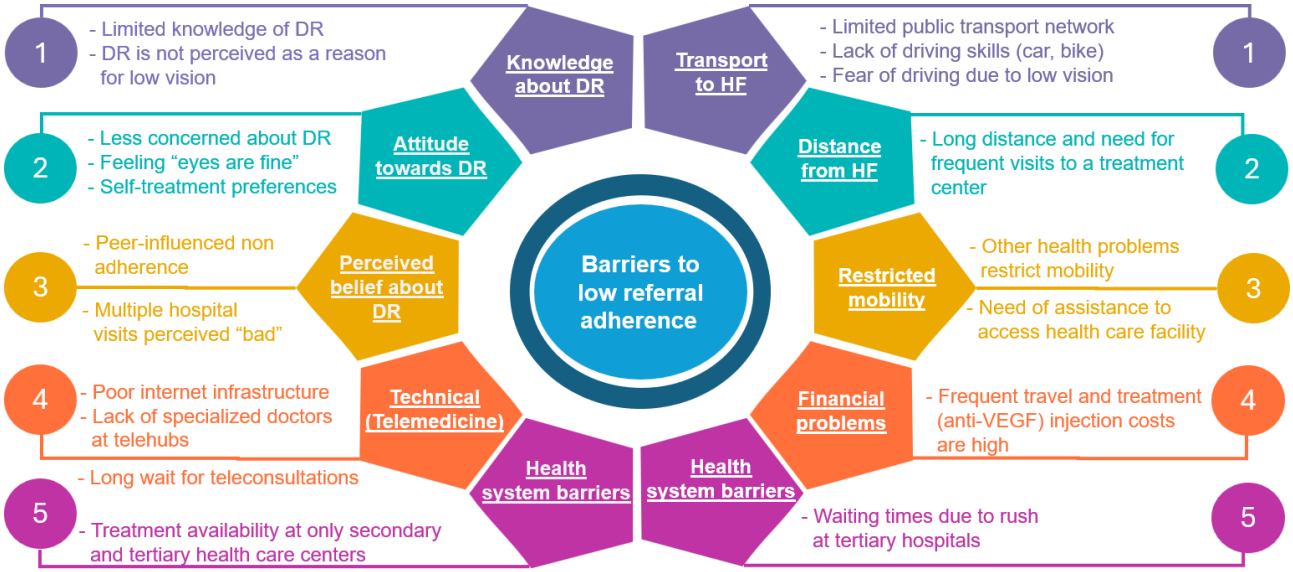
Barriers Associated With Low Referral Adherence

Overview

The data revealed three key themes that best explained the barriers to referral adherence: (1) awareness and knowledge-related obstacles, (2) logistical support challenges, and (3) health care system limitations. The results of the in-depth interviews with the stakeholders complemented the quantitative findings of nonadherent participants and shed further light on their findings (Figure 2). The thematic analysis suggested a range of factors that impact adherence to referral instructions and put low adherence or nonadherence into perspective. The quotes under the categories are available in Table S8 in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 2. Barriers to referral nonadherence. DR: diabetic retinopathy; HF: health facility; VEGF: vascular endothelial growth factor.



Awareness and Knowledge-Related Barriers

The HCPs observed that patients with DR often delayed treatment due to a lack of awareness about the condition and its complications, frequently mistaking vision changes for aging, cataracts, or simply needing new glasses. These misconceptions often lead patients to request remedies such as eye drops or new glasses, mistakenly believing these will resolve their vision problems. One patient recounted an interaction with fellow patients who had undergone extensive treatments and incurred substantial expenses without experiencing any improvement in vision. This perception highlights a common gap in patient knowledge. Individuals often rely excessively on peer anecdotal evidence, hindering adherence to evidence-based medical treatments. This issue is further complicated by a limited awareness of the crucial role regular follow-up appointments play in managing DR, resulting in poor compliance with treatment protocols. Additionally, some patients tend to avoid health care facilities under the false assumption that their ocular condition has stabilized while engaging in self-treatment, which introduces significant risks and compromises effective disease management.

Logistical Support Barriers

Mobility and Transportation

The HCPs said that the lack of public transport restricts participants from accessing local transport options for treatment, making it hard for patients to reach the facility for their appointments. One patient with a 2-wheeler reported difficulty driving alone because of their low vision. Longer distances to treatment facilities also decreased the frequency of visits, resulting in higher dropout rates despite initial engagement in treatment. Sometimes, multiple health issues can limit mobility and hinder access to eye care services, leading individuals to prioritize managing other health conditions over maintaining regular DR treatment appointments. This preference reflects a practical decision based on immediate health needs and logistical challenges posed by mobility restrictions rather than a lack of recognition of the importance of DR care.

Financial

Cost presents a formidable barrier, restricting many from affording regular travel to treatment centers. Initially, patients may adhere to treatment plans diligently, but as time progresses, dropout rates escalate. A patient articulated their predicament: "I was referred to another hospital for injections, which are financially out of reach due to the distance from home." The ophthalmologist and retina specialist underscored the substantial financial burden of DR treatment, specifically anti-VEGF (vascular endothelial growth factor) injections, which significantly impacts adherence. Financially disadvantaged patients frequently prioritize other expenses over necessary visits to eye care facilities, thereby exacerbating the challenge of managing DR effectively and consistently. This underscores the critical need for accessible and affordable eye care solutions to mitigate the economic barriers hindering patient care and treatment adherence.

Health Care System Barriers

Burdened Health System

The tertiary eye care hospital experiences high demand, resulting in delayed appointments that led many patients to discontinue treatment after 1-2 visits. Long queues also discourage patients from seeking care at larger eye care centers, underscoring the systemic challenges that affect access to DR treatment services. The large number of patients and extended waiting times pose significant challenges for individuals trying to attend appointments and receive timely treatment.

Technical Challenges

Telemedicine facilities are beneficial for extending eye care access [49], but they face implementation challenges. The limited availability of specialized doctors and unreliable internet in peripheral areas hinder teleconsultations. Village residents encountered prolonged wait times at telemedicine centres and expressed reluctance to wait for consultations with doctors at these facilities.

XSL•FO
RenderX

Discussion

Principal Findings

This study investigated barriers to referral adherence after AI-enabled DRS at primary (community) and secondary (CHC) sites among those who did not follow referrals to district hospitals for further management. We implemented AI-enabled DRS programs for older adults in public health settings in Punjab, India, during the COVID-19 pandemic. AI has been promoted as a transformative tool in health care to help attain health and health-related Sustainable Development Goals among older adults [50,51]. However, the impact of implementing AI on patient outcomes depends on many factors beyond diagnostic accuracy, including care access, referral adherence, and the implementation of treatment and management recommendations [52]. In our study, only 23 of 159 (14.5%) patients screened adhered to referral recommendations. Of these, only 9 visited the district hospital, while the others chose different facilities, indicating various factors influencing their decision-making regarding referral locations [53]. This finding contrasts with a study conducted in South India examining the factors influencing the utilization of referral services from secondary to tertiary eye care, which reported noncompliance rates of 31.65% and 24.2%, respectively [33].

Telephone interviews identified key barriers in our study, including low awareness about DR (n=4, 2.9%), attitudinal issues (n=27, 19.9%), and a lack of family support (n=21, 15.4%). Economic 27 (16.4%) and attitudinal barriers (44.2%) to referral service access were consistent with findings by Padhy et al [33] Our analysis showed no differences in the characteristics of adherent and nonadherent participants. However, nonstatistically significant differences between adherent and nonadherent groups do not imply that these factors are irrelevant or that there are no relationships. This indicates that we did not detect significant associations with the current sample size and variability. However, the wide confidence intervals suggest additional quantitative data are needed to improve estimate precision.

Nevertheless, we have no evidence that demographic and health variables alone are sufficient to understand or predict adherence behavior in patients with DR. Technical errors (Table 3) prevented us from contacting 31 of 136 (22.8%) participants, potentially impacting our quantitative analysis, including the univariate and multivariate analysis results. The unprecedented COVID-19 pandemic may have also negatively affected health-seeking behavior and treatment adherence [19,22].

We prioritized incorporating insights from key stakeholders through qualitative, in-depth interviews to guide informational power. Our study found that older adult individuals lack adequate awareness about DR complications, resulting in misconceptions about the causes of vision changes. Lack of awareness discourages service use, enabling unchecked harmful conditions and potentially leading to adverse outcomes due to delayed care-seeking. Further, the participants encountered difficulties affording transportation and treatment expenses and taking time off work for medical appointments, driven by concerns about their daily income [54]. Additionally, attitudes such as believing their eyes are healthy and logistical challenges like lack of family support impeded referral adherence [23]. Reliance on anecdotal evidence (do not go for eye treatment) from peers discouraged adherence to referral recommendations. Local and regional factors, such as the peak harvesting season (Table 3), hindered older adult adherence to referral recommendations (n=14, 10.3%). In-depth stakeholder interviews complemented the telephone follow-up findings, revealing barriers to referral recommendations. Similar programs, like the point-of-care DR examination program initiative, reported suboptimal follow-up rates despite targeted interventions, highlighting the need for more personalized strategies to enhance adherence [24]. Community-based programs have proven effective in enhancing access to eye care and improving patient adherence to continuity of care recommendations [24,55]. Bonilla-Escobar et al [24] showed higher adherence to referral recommendations when personalized approaches, such as phone calls, voicemails, and result letters with appointment details, were used. Transportation interventions, including bus passes, taxi vouchers, reimbursed travel, and tailored services, enhance health outcomes in chronic diseases, particularly for older adults and women [56].

Teleophthalmology, for instance, has accelerated health care access in rural and remote areas [57], overcoming significant travel constraints, time constraints, and economic barriers through community-based program initiatives [58,59]. However, the HCPs noted that technical challenges, such as underdeveloped internet infrastructure in rural areas, pose challenges to implementing teleophthalmology programs [60]. Hence, it is critical to identify methods for establishing DRS programs tailored to available resources and health care infrastructure [24,61]. In our study, older adult participants highlighted financial constraints as barriers to accessing referral services, a sentiment corroborated by the HCPs' in-depth interviews. The findings align with existing literature on economic challenges among individuals with visual impairment in developing countries [62]. The high costs of DR treatment often compel financially disadvantaged patients to prioritize other essential expenses over necessary eye care. Patients and HCPs should be informed about the comprehensive coverage offered by the Ayushman Bharat Pradhan Mantri Jan Arogya Yojana public health insurance program in India, including DRS and treatment. This ensures adequate access to necessary services through insurance benefits [63].

Our study findings revealed issues in the care process in rural areas, such as patients self-referring and bypassing the referral system hierarchy [64]. Systemic challenges such as delayed appointments and long queues discourage patients from continuing treatment beyond 1 or 2 visits due to access issues [23]. The World Report on Vision has recommended an integrated care model incorporating primary, secondary, and tertiary care, ensuring clear roles and appropriate referrals at each health care system level [65].

This study's strength lies in triangulating quantitative findings with qualitative insights to address research questions regarding referral adherence. Additionally, conducting the study across primary (community) and secondary (CHC) health care systems

XSL•FO
RenderX

provides comprehensive information on referral adherence within a network of public health systems.

A limitation of this study is its focus on a public health network in North India within one state, potentially limiting the generalizability of its findings to broader settings. Potential biases from self-reported data during telephone follow-ups may lead to underreporting or overreporting adherence based on perceived expectations and limited generalizability due to the study's regional focus. The study was conducted during the COVID-19 pandemic, marked by significant disruptions to health care systems and changes in patient behavior. These unique circumstances, including the impact of COVID-19 on decision-making regarding nonemergency medical care, may have introduced factors related to health care access that are specific to this time. Furthermore, nearly 22.8% of screened patients (Table 3) were not successfully contacted despite multiple attempts, which would increase the imprecision of estimates yielded by the regression analysis. For example, only 23 adherent participants contributed data to the multivariate model and therefore the analyses may be hindered due to issues related to statistical power and imbalanced group sizes [66]. This study did not extensively collect detailed information on health care system factors affecting adherence, such as hospital waiting times, hospital accessibility, and staff attitudes [33]. This highlights the necessity of exploring additional factors that influence adherence, including psychological, social, and systemic barriers [23,67].

In this pilot study, approximately 390 of 948 (41%) participants were recruited due to COVID-19 pandemic restrictions. Additionally, the limited sample size in the nonadherent group and high variability in the collected data resulted in insufficient statistical power to accurately identify determinants of nonadherence, leading to large confidence intervals and reduced precision in the estimates.

Conclusion and Policy Implications

This study, conducted during the COVID-19 pandemic, has demonstrated suboptimal adherence among older adult patients to referral recommendations. Participants considered barriers such as knowledge gaps, logistical challenges, and health system–related issues to have hindered referral adherence. Given the unique context of the pandemic, the results should be generalized with caution. Future studies should consider pragmatic trial study designs incorporating implementation science approaches using mixed methods to examine barriers to referral and adherence across diverse contexts. This could include exploring community beliefs and other social determinants influencing service uptake. Policy-level changes are essential to integrating teleophthalmology into public health programs and improving infrastructure for AI-enabled DRS, enhancing access and outcomes. An effective, well-coordinated referral system with defined protocols, streamlined communication, reliable transportation, trained personnel, integrated systems, and collaboration may address the challenges in the referral systems [68,69].

Acknowledgments

The authors sincerely thank all participants for their invaluable contributions to this research and the Department of Health and Family Welfare, Punjab, India, for granting permission to conduct the study. All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[PDF File, 248 KB - formative_v9i1e67047_app1.pdf]

References

1. Raman R, Vasconcelos JC, Rajalakshmi R, et al. Prevalence of diabetic retinopathy in India stratified by known and undiagnosed diabetes, urban-rural locations, and socioeconomic indices: results from the SMART India population-based cross-sectional screening study. Lancet Glob Health 2022 Dec;10(12):e1764-e1773. [doi: 10.1016/S2214-109X(22)00411-9] [Medline: 36327997]
2. Ting DSW, Cheung GCM, Wong TY. Diabetic retinopathy: global prevalence, major risk factors, screening practices and public health challenges: a review. Clin Exp Ophthalmol 2016 May;44(4):260-277. [doi: 10.1111/ceo.12696] [Medline: 26716602]
3. Vujosevic S, Aldington SJ, Silva P, et al. Screening for diabetic retinopathy: new perspectives and challenges. Lancet Diabetes Endocrinol 2020 Apr;8(4):337-347. [doi: 10.1016/S2213-8587(19)30411-5] [Medline: 32113513]
4. IDF Diabetes Atlas. URL: www.diabetesatlas.org [accessed 2025-03-26]
5. Teo ZL, Tham YC, Yu M, et al. Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. Ophthalmology 2021 Nov;128(11):1580-1591. [doi: 10.1016/j.ophtha.2021.04.027] [Medline: 33940045]

XSL•FO
RenderX

6. Flaxman SR, Bourne RRA, Resnikoff S, et al. Global causes of blindness and distance vision impairment 1990-2020: a systematic review and meta-analysis. Lancet Glob Health 2017 Dec;5(12):e1221-e1234. [doi: 10.1016/S2214-109X(17)30393-5] [Medline: 29032195]
7. Leley SP, Ciulla TA, Bhatwadekar AD. Diabetic retinopathy in the aging population: a perspective of pathogenesis and treatment. Clin Interv Aging 2021;16:1367-1378. [doi: 10.2147/CIA.S297494] [Medline: 34290499]
8. Lamoureux EL, Tai ES, Thumboo J, et al. Impact of diabetic retinopathy on vision-specific function. Ophthalmology 2010 Apr;117(4):757-765. [doi: 10.1016/j.ophtha.2009.09.035] [Medline: 20122736]
9. Fenwick E, Rees G, Pesudovs K, et al. Social and emotional impact of diabetic retinopathy: a review. Clin Exp Ophthalmol 2012;40(1):27-38. [doi: 10.1111/j.1442-9071.2011.02599.x] [Medline: 21575125]
10. National Programme for Health Care of the Elderly. Ministry of Health and Family Welfare. URL: https://mohfw.gov.in/?q=major-programmes/Non-Communicable-Diseases/Non-Communicable-Diseases-1 [accessed 2025-03-26]
11. Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness certifications in England and Wales in working age adults (16-64 years), 1999-2000 with 2009-2010. BMJ Open 2014 Feb 12;4(2):e004015. [doi: 10.1136/bmjopen-2013-004015] [Medline: 24525390]
12. Ryan ME, Rajalakshmi R, Prathiba V, et al. Comparison Among Methods of Retinopathy Assessment (CAMRA) study: smartphone, nonmydriatic, and mydriatic photography. Ophthalmology 2015 Oct;122(10):2038-2043. [doi: 10.1016/j.ophtha.2015.06.011] [Medline: 26189190]
13. Padhy SK, Takkar B, Chawla R, Kumar A. Artificial intelligence in diabetic retinopathy: a natural step to the future. Indian J Ophthalmol 2019 Jul;67(7):1004-1009. [doi: 10.4103/ijo.IJO_1989_18] [Medline: 31238395]
14. Abbasi K. The scandals of covid-19. BMJ 2020 Apr(369):m1434. [doi: 10.1136/bmj.m1434]
15. Shih KC, Kwong ASK, Wang JHL, et al. Diabetic retinopathy screening during the coronavirus disease 2019 pandemic. Eye (Lond) 2020 Jul;34(7):1246-1247. [doi: 10.1038/s41433-020-0928-7] [Medline: 32367007]
16. Grzybowski A, Brona P, Lim G, et al. Artificial intelligence for diabetic retinopathy screening: a review. Eye (Lond) 2020 Mar;34(3):451-460. [doi: 10.1038/s41433-019-0566-0] [Medline: 31488886]
17. Mathenge W, Whitestone N, Nkurikiye J, et al. Impact of artificial intelligence assessment of diabetic retinopathy on referral service uptake in a low-resource setting: the RAIDERS randomized trial. Ophthalmol Sci 2022 Dec;2(4):100168. [doi: 10.1016/j.xops.2022.100168] [Medline: 36531575]
18. Bonilla-Escobar FJ, Ghobrial AI, Gallagher DS, Eller A, Waxman EL. Comprehensive insights into a decade-long journey: the evolution, impact, and human factors of an asynchronous telemedicine program for diabetic retinopathy screening in Pennsylvania, United States. PLoS One 2024;19(7):e0305586. [doi: 10.1371/journal.pone.0305586] [Medline: 38995899]
19. Ahmed I, Liu TYA. The impact of COVID-19 on diabetic retinopathy monitoring and treatment. Curr Diab Rep 2021 Sep 8;21(10):40. [doi: 10.1007/s11892-021-01411-6] [Medline: 34495377]
20. Kumar A. The transformation of the Indian healthcare system. Cureus 2023 May;15(5):e39079. [doi: 10.7759/cureus.39079] [Medline: 37378105]
21. Burkhart PV, Sabaté E. Adherence to long-term therapies: evidence for action. J Nurs Scholarsh 2003;35(3):207. [Medline: 14562485]
22. Breazzano MP, Nair AA, Arevalo JF, et al. Frequency of urgent or emergent vitreoretinal surgical procedures in the United States during the COVID-19 pandemic. JAMA Ophthalmol 2021 Apr 1;139(4):456-463. [doi: 10.1001/jamaophthalmol.2021.0036] [Medline: 33662093]
23. Piyasena M, Murthy GVS, Yip JLY, et al. Systematic review on barriers and enablers for access to diabetic retinopathy screening services in different income settings. PLoS One 2019;14(4):e0198979. [doi: 10.1371/journal.pone.0198979] [Medline: 31013274]
24. Bonilla-Escobar FJ, Eibel MR, Le L, Gallagher DS, Waxman EL. Follow-up in a point-of-care diabetic retinopathy program in Pittsburgh: a non-concurrent retrospective cohort study. BMC Ophthalmol 2024 Aug 20;24(1):356. [doi: 10.1186/s12886-024-03581-9] [Medline: 39164678]
25. Khanna RC, Kim S, Giridhar P, Mettla AL, Marmamula S, Rao GN. Barriers to uptake of referral services from secondary care to tertiary care and its associated factors in L V Prasad Eye Institute network in Southern India: a cross-sectional study. BMJ Open 2018 Jul 7;8(7):e020687. [doi: 10.1136/bmjopen-2017-020687] [Medline: 29982204]
26. Adhikari K, Patten SB, Patel AB, et al. Data harmonization and data pooling from cohort studies: a practical approach for data management. Int J Popul Data Sci 2021;6(1):1680. [doi: 10.23889/ijpds.v6i1.1680] [Medline: 34888420]
27. ASHA training modules. Ministry of Health and Family Welfare. URL: https://nhm.gov.in/index1.php?lang=1&level=3&sublinkid=184&lid=257 [accessed 2025-03-26]
28. Ayushman bharat comprehensive primary health care through health and wellness centres operational guidelines. Ministry of Health and Family Welfare. URL: http://www.ab-hwc.mohfw.gov.in/ [accessed 2023-10-13]
29. Indian public health standards. Community health centre 2022. National Health Systems Resource Centre. URL: https://nhsrcindia.org/sites/default/files/CHC%20IPHS%202022%20Guidelines%20pdf.pdf [accessed 2023-10-08]
30. Indian public health standards, sub district and district health hospital. Ministry of Health & Family Welfare. URL: https://nhsrcindia.org [accessed 2025-03-26]

XSL•FO
RenderX

31. Mukamel DB, Bresnick GH, Wang Q, Dickey CF. Barriers to compliance with screening guidelines for diabetic retinopathy. Ophthalmic Epidemiol 1999 Mar;6(1):61-72. [doi: 10.1076/opep.6.1.61.1563] [Medline: 10384685]
32. Watson MJG, McCluskey PJ, Grigg JR, Kanagasingam Y, Daire J, Estai M. Barriers and facilitators to diabetic retinopathy screening within Australian primary care. BMC Fam Pract 2021 Nov 30;22(1):239. [doi: 10.1186/s12875-021-01586-7] [Medline: 34847874]
33. Padhy D, Pyda G, Marmamula S, Khanna RC. Barriers to uptake of referral services from secondary eye care to tertiary eye care and its associated determinants in L V Prasad Eye Institute network in Southern India: a cross-sectional study-report II. PLoS ONE 2024;19(5):e0303401. [doi: 10.1371/journal.pone.0303401] [Medline: 38743737]
34. Kumar S, Kumar G, Velu S, et al. Patient and provider perspectives on barriers to screening for diabetic retinopathy: an exploratory study from southern India. BMJ Open 2020 Dec 10;10(12):e037277. [doi: 10.1136/bmjopen-2020-037277] [Medline: 33303431]
35. Shah P, Mishra DK, Shanmugam MP, Doshi B, Jayaraj H, Ramanjulu R. Validation of deep convolutional neural network-based algorithm for detection of diabetic retinopathy - artificial intelligence versus clinician for screening. Indian J Ophthalmol 2020 Feb;68(2):398-405. [doi: 10.4103/ijo.IJO_966_19] [Medline: 31957737]
36. Darwish DY, Patel SN, Gao Y, et al. Diagnostic accuracy and reliability of retinal pathology using the Forus 3nethra fundus camera compared to ultra wide-field imaging. Eye (Lond) 2019 May;33(5):856-857. [doi: 10.1038/s41433-019-0339-9] [Medline: 30679873]
37. Sosale B, Sosale AR, Murthy H, Sengupta S, Naveenam M. Medios- an offline, smartphone-based artificial intelligence algorithm for the diagnosis of diabetic retinopathy. Indian J Ophthalmol 2020 Feb;68(2):391-395. [doi: 10.4103/ijo.IJO_1203_19] [Medline: 31957735]
38. Duggal M, et al. Responsible adoption of cloud-based artificial intelligence in health care: A validation case study of multiple artificial intelligence algorithms for diabetic retinopathy screening in public health settings (forthcoming). .
39. Rajalakshmi R, Subashini R, Anjana RM, Mohan V. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. Eye (Lond) 2018 Jun;32(6):1138-1144. [doi: 10.1038/s41433-018-0064-9] [Medline: 29520050]
40. Wong TY, Sun J, Kawasaki R, et al. Guidelines on diabetic eye care: the International Council of Ophthalmology recommendations for screening, follow-up, referral, and treatment based on resource settings. Ophthalmology 2018 Oct;125(10):1608-1622. [doi: 10.1016/j.ophtha.2018.04.007] [Medline: 29776671]
41. Raman R, Ramasamy K, Rajalakshmi R, Sivaprasad S, Natarajan S. Diabetic retinopathy screening guidelines in India: All India Ophthalmological Society diabetic retinopathy task force and Vitreoretinal Society of India Consensus Statement. Indian J Ophthalmol 2021 Mar;69(3):678-688. [doi: 10.4103/ijo.IJO_667_20] [Medline: 33269742]
42. Lee BW, Sathyan P, John RK, Singh K, Robin AL. Predictors of and barriers associated with poor follow-up in patients with glaucoma in South India. Arch Ophthalmol 2008 Oct;126(10):1448-1454. [doi: 10.1001/archopht.126.10.1448] [Medline: 18852425]
43. Vengadesan N, Ahmad M, Sindal MD, Sengupta S. Delayed follow-up in patients with diabetic retinopathy in South India: social factors and impact on disease progression. Indian J Ophthalmol 2017 May;65(5):376-384. [doi: 10.4103/ijo.IJO_620_16] [Medline: 28573993]
44. Mason M. Sample size and saturation in PhD studies using qualitative interviews. Forum Qual Soc Res 2010;11(3). [doi: 10.17169/fqs-11.3.1428]
45. Ritchie J, Lewis J. Qualitative Research Practice: A Guide for Social Science Students and Researchers: Sage Publications; 2003. URL: https://mthoyibi.files.wordpress.com/2011/10/qualitative-research-practice_a-guide-for-social-science-students-and-researchers_jane-ritchie-and-jane-lewis-eds_20031.pdf [accessed 2025-03-27]
46. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
47. Byrne D. A worked example of Braun and Clarke's approach to reflexive thematic analysis. Qual Quant 2022 Jun;56(3):1391-1412. [doi: 10.1007/s11135-021-01182-y]
48. Monitoring the building blocks of health systems: a handbook of indicators and their measurement strategies. World Health Organization. URL: https://apps.who.int/iris/bitstream/handle/10665/258734/9789241564052-eng.pdf [accessed 2025-03-26]
49. Kalogeropoulos D, Kalogeropoulos C, Stefaniotou M, Neofytou M. The role of tele-ophthalmology in diabetic retinopathy screening. J Optom 2020;13(4):262-268. [doi: 10.1016/j.optom.2019.12.004] [Medline: 31948924]
50. Sajwani RA. Artificial intelligence for sustainability development in healthcare. In: BUiD Doctoral Research Conference 2023 2024:264-272. [doi: 10.1007/978-3-031-56121-4_26]
51. Ageing, older persons and the 2030 agenda for sustainable development acknowledgements. : UNDP; 2017 URL: https://www.undp.org/publications/ageing-older-persons-and-2030-agenda-sustainable-development [accessed 2025-03-27]
52. Haynes B. Can it work? Does it work? Is it worth it? The testing of healthcareinterventions is evolving. BMJ 1999 Sep 11;319(7211):652-653. [doi: 10.1136/bmj.319.7211.652] [Medline: 10480802]

XSL•FO
RenderX

53. Piyasena M, Murthy GVS, Yip JLY, et al. Systematic review and meta-analysis of diagnostic accuracy of detection of any level of diabetic retinopathy using digital retinal imaging. Syst Rev 2018 Nov 7;7(1):182. [doi: 10.1186/s13643-018-0846-y] [Medline: 30404665]
54. Taber JM, Leyva B, Persoskie A. Why do people avoid medical care? A qualitative study using national data. J Gen Intern Med 2015 Mar;30(3):290-297. [doi: 10.1007/s11606-014-3089-1] [Medline: 25387439]
55. Solomon SD, Shoge RY, Ervin AM, et al. Improving access to eye care: a systematic review of the literature. Ophthalmology 2022 Oct;129(10):e114-e126. [doi: 10.1016/j.ophtha.2022.07.012] [Medline: 36058739]
56. Starbird LE, DiMaina C, Sun CA, Han HR. A systematic review of interventions to minimize transportation barriers among people with chronic diseases. J Community Health 2019 Apr;44(2):400-411. [doi: 10.1007/s10900-018-0572-3] [Medline: 30206755]
57. Chabouis A, Berdugo M, Meas T, et al. Benefits of Ophdiat, a telemedical network to screen for diabetic retinopathy: a retrospective study in five reference hospital centres. Diabetes Metab 2009 Jun;35(3):228-232. [doi: 10.1016/j.diabet.2008.12.001] [Medline: 19349202]
58. Hark LA, Katz LJ, Myers JS, et al. Philadelphia telemedicine glaucoma detection and follow-up study: methods and screening results. Am J Ophthalmol 2017 Sep;181:114-124. [doi: 10.1016/j.ajo.2017.06.024] [Medline: 28673747]
59. Wang YZ, He YG, Mitzel G, Zhang S, Bartlett M. Handheld shape discrimination hyperacuity test on a mobile device for remote monitoring of visual function in maculopathy. Invest Ophthalmol Vis Sci 2013 Aug 13;54(8):5497-5505. [doi: 10.1167/iovs.13-12037] [Medline: 23860761]
60. Yim D, Chandra S, Sondh R, Thottarath S, Sivaprasad S. Barriers in establishing systematic diabetic retinopathy screening through telemedicine in low- and middle-income countries. Indian J Ophthalmol 2021 Nov;69(11):2987-2992. [doi: 10.4103/ijo.IJO_1411_21] [Medline: 34708733]
61. Wong TY, Sabanayagam C. Strategies to tackle the global burden of diabetic retinopathy: from epidemiology to artificial intelligence. Ophthalmologica 2020;243(1):9-20. [doi: 10.1159/000502387] [Medline: 31408872]
62. Sengo DB, Marraca NA, Muaprato AM, García-Sanjuan S, Caballero P, López-Izquierdo I. Barriers to accessing eye health services in suburban communities in Nampula, Mozambique. Int J Environ Res Public Health 2022 Mar 25;19(7):7. [doi: 10.3390/ijerph19073916] [Medline: 35409600]
63. Government of india ministry of health and family welfare, AB-PMJAY health benefit package master 2022. National Health Authority. URL: https://pmjay.gov.in/sites/default/files/2022-04/HBP%202022%20..pdf [accessed 2024-07-14]
64. Eskandari M, Abbaszadeh A, Borhani F. Barriers of referral system to health care provision in rural societies in Iran. J Caring Sci 2013 Sep;2(3):229-236. [doi: 10.5681/jcs.2013.028] [Medline: 25276731]
65. Rao GN, Khanna RC, Athota SM, Rajshekar V, Rani PK. Integrated model of primary and secondary eye care for underserved rural areas: the L V Prasad Eye Institute experience. Indian J Ophthalmol 2012;60(5):396-400. [doi: 10.4103/0301-4738.100533] [Medline: 22944748]
66. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996 Dec;49(12):1373-1379. [doi: 10.1016/s0895-4356(96)00236-3] [Medline: 8970487]
67. Piyasena M, Murthy GVS, Yip JLY, et al. A qualitative study on barriers and enablers to uptake of diabetic retinopathy screening by people with diabetes in the Western Province of Sri Lanka. Trop Med Health 2019;47:34. [doi: 10.1186/s41182-019-0160-y] [Medline: 31139011]
68. Kruk ME, Gage AD, Arsenault C, et al. High-quality health systems in the Sustainable Development Goals era: time for a revolution. Lancet Glob Health 2018 Nov;6(11):e1196-e1252. [doi: 10.1016/S2214-109X(18)30386-3] [Medline: 30196093]
69. Kamau KJ, Osuga BO, Njuguna S. Challenges facing implementation of referral system for quality health care services in Kiambu County, Kenya. Health Syst Policy Res 2017;04(1):01. [doi: 10.21767/2254-9137.100067]

Abbreviations

ASHA: Accredited Social and Health Activists
CHC: community health center
CHO: community health officer
DM: diabetes mellitus
DR: diabetic retinopathy
DRS: DR screening
HCP: health care provider
VTDR: vision-threatening diabetic retinopathy

XSL•FO
RenderX

Edited by A Mavragani; submitted 30.09.24; peer-reviewed by F Bonilla-Escobar, T Intharah; revised version received 22.02.25; accepted 25.02.25; published 31.03.25.

Please cite as:
Chauhan A, Goyal A, Masih R, Kaur G, Kumar L, Neha , Rastogi H, Kumar S, Singh BL, Syal P, Gupta V, Vale L, Duggal M
Barriers and Determinants of Referral Adherence in AI-Enabled Diabetic Retinopathy Screening for Older Adults in Northern India During the COVID-19 Pandemic: Mixed Methods Pilot Study
JMIR Form Res 2025;9:e67047
URL: https://formative.jmir.org/2025/1/e67047
doi:10.2196/67047

© Anshul Chauhan, Anju Goyal, Ritika Masih, Gagandeep Kaur, Lakshay Kumar, Neha, Harsh Rastogi, Sonam Kumar, Bidhi Lord Singh, Preeti Syal, Vishali Gupta, Luke Vale, Mona Duggal. Originally published in JMIR Formative Research (https://formative.jmir.org), 31.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Personalized mHealth Intervention (StepAdd) for Increasing Physical Activity in Japanese Patients With Type 2 Diabetes: Secondary Analysis of Social Cognitive Theory Measurements of a Single-Arm Pilot Study

Kayo Waki[1,2,3], MPH, MD, PhD; Syunpei Enomoto[3], MAS; Toshimasa Yamauchi[2], MD, PhD; Masaomi Nangaku[4], MD, PhD; Kazuhiko Ohe[1,3], MD, PhD

[1]Department of Biomedical Informatics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan

[2]Department of Diabetes and Metabolic Diseases, The University of Tokyo, Tokyo, Japan

[3]Department of Planning, Information and Management, University of Tokyo Hospital, Tokyo, Japan

[4]Division of Nephrology and Endocrinology, The University of Tokyo, Tokyo, Japan

Corresponding Author:
Kayo Waki, MPH, MD, PhD
Department of Biomedical Informatics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan

Abstract

Background: A 12-week pilot of the StepAdd mobile health (mHealth) behavior change intervention based on social cognitive theory (SCT) saw an 86.7% increase in mean daily step counts among patients with type 2 diabetes. Due to the lack of exploration of theoretical implications in mHealth intervention studies, there is a need to understand the mechanism underlying the behavioral change to inform the future design of digital therapeutics.

Objective: This study aimed to examine the SCT drivers underlying the mean increase in exercise among Japanese patients with type 2 diabetes who participated in the StepAdd intervention.

Methods: This is a post hoc analysis of data collected in the single-arm pilot study of the 32 patients who completed the StepAdd intervention. The StepAdd app uses self-mastery and coping strategies to increase self-efficacy and thus increase walking. Self-mastery was measured by the goal completion (GC) rate, which is the percentage of days in which patients met these adapting goals. The use of coping strategies was measured by the strategy implementation (SI) rate, which is the percentage of days in which patients applied their selected coping strategies. We assessed correlations between GC, SI, and self-efficacy to increase walking via linear regression and analyzed relationships via structural equation modeling.

Results: We found statistically significant support for the SCT approach, including a correlation coefficient (ρ) of 0.649 between step increase and GC rate (P<.001); a ρ of 0.497 between the coping SI rate and self-efficacy increase (P=.004); a ρ of 0.446 between GC rate and self-mastery increase (P=.01); and a ρ of 0.355 between self-regulation increase and step increase (P=.046), giving us insight into why the behavior intervention succeeded. We also found significant correlations between self-efficacy for barriers and self-efficacy for task-specific behavior (ρ=0.358; P=.04), as well as self-regulation and self-efficacy for task-specific behavior (ρ=0.583; P<.001). However, a cross-lagged panel modeling analysis found no significant evidence that changes in self-efficacy preceded behavior changes in line with SCT.

Conclusions: Self-mastery and coping strategies contributed to the walking behavior change in StepAdd, supporting the SCT model of behavior change. Future research is needed to better understand the causal pathways proposed by SCT.

(JMIR Form Res 2025;9:e60221) doi:10.2196/60221

KEYWORDS

social cognitive theory; mHealth; mobile health; behavior change; diabetes; diabetic; type 2 diabetes; walk; step; exercise; physical activity; walking; digital therapeutics; secondary analysis; personalization; coping; self-efficacy

Introduction

Digital therapeutics (DTx), defined as evidence-based therapeutic interventions that are driven by high-quality software programs to treat, manage, or prevent a disease or disorder, can be effective in treating diabetes by inducing behavior changes [1]. However, studies of DTx generally do not explore the implications of the underlying behavior change theory. For example, we were unable to explain the mechanism behind behavior change with a randomized controlled trial (RCT) that

XSL•FO
RenderX

showed significant improvement in hemoglobin A_{1C} (HbA_{1c}) and albuminuria among patients with type 2 diabetes via DTx [2]. However, given the progress in DTx, it is not enough to simply show outcomes. Lifestyle change involves complex patient behavior comprising numerous decisions and significant effort. We need to explain the mechanism behind behavior change, to show the characteristics of the app in question that differ from those of other apps, and to explain the rationale for choosing the app to treat patients.

We analyzed results from a 12-week pilot study of 33 Japanese patients with diabetes who participated in the StepAdd mobile health (mHealth) behavior change intervention to explain the mechanism behind the behavior change [3]. This was a single-arm pilot study from August 2021 to December 2021 using pre-post evaluation to assess the feasibility and preliminary efficacy of a personalized mHealth intervention using the StepAdd app to improve physical activity among patients with type 2 diabetes, conducted in collaboration with Mitsui Memorial Hospital; Nihon Chouzai Co, Ltd (a community pharmacy); and Mitsui & Co, Ltd. Patients receiving type 2 diabetes treatment at Mitsui Memorial Hospital, with an HbA_{1c} of 7.5% or higher, were recruited by their physicians to participate in the pilot study evaluating the preliminary efficacy of StepAdd in increasing physical activity. StepAdd applies behavior change methods based on social cognitive theory (SCT) [4,5], and the pilot collected SCT-related measures throughout the intervention. The study included 6 visits spanning an initial step baseline phase and a 12-week continuous measurement phase. The transtheoretical model was used to categorize the participants who were currently at the contemplation stage (willing to change health behavior within the next 6 months), preparation stage (willing to change health behavior within the next month), or action stage (has made modifications to health behavior) to achieve the target goal of 10,000 steps a day. A community pharmacist briefed participants on this research study and on how to use the smartphone and the pedometer. The pilot saw an 86.7% increase in mean daily step counts, from 5436 (SD 2231) per day to 10,150 (SD 3908) per day, and a reduction of mean HbA_{1c} from 8.58% (SD 1.02%) to 7.79% (SD 1.11%) (−0.79%, SD 1.04%) [4]. A 24-week StepAdd RCT, with 80 participants per arm, is currently underway [6]. Analysis of RCTs has proven that increased walking causes clinically significant improvements to glycemic control among patients with diabetes [7,8]. The StepAdd pilot results show overall causality—the intervention caused the desired behavior and health improvement.

This secondary analysis seeks to use SCT measurements to understand the factors underlying the behavioral change, both to inform future analyses of the StepAdd RCT results and to improve the design of future interventions. We want to understand what features caused the change and whether the causal relationships defined by SCT can be proved.

SCT posits a central role to self-efficacy, the individual's belief that they can perform the behavior, with increases in self-efficacy causing increases in the behavior. StepAdd uses self-mastery and coping strategies to increase self-efficacy and thus increase walking. For self-mastery, StepAdd implements achievable but challenging goals, which are adapted weekly, with self-mastery being measured by the goal completion (GC) rate. For coping strategies, StepAdd has participants choose a barrier and an associated coping strategy to implement, increasing self-efficacy, with coping strategy use being measured by the strategy implementation (SI) rate. During the StepAdd pilot, we collected 3 measures of the underlying latent construct of self-efficacy at weeks 0, 4, 8, and 12: self-efficacy in achieving the targeted behavior, self-efficacy in dealing with barriers, and self-regulation (SR).

Methods

Overview

This is a post hoc analysis of data collected in the StepAdd pilot for the 32 patients who completed the intervention (the pilot analyzed 33 patients, but 1 dropped out after week 4 because he expected the study to be just 4 weeks). Data came from pedometers, the StepAdd app, and questionnaire instruments, with data collected at weeks 0, 4, 8, and 12 of the intervention.

The step data came from daily counts from wireless pedometers (OMRON HJA-405 T-W or Yamasa AW-001), averaged for the 2 weeks prior to the week 0, 4, 8, and 12 time points.

The intervention used 2 SCT-based methods. First, it targeted improving self-mastery by using a series of attainable yet growing goals. The GC rate measured the percentage of days in which patients met these adapting goals [3]. Second, it targeted enhancing SR and improving sociocultural factors in the SCT framework by allowing patients to evaluate possible barriers to achieving their step goals and choose to implement associated coping strategies. The SI rate measured the percentage of days in which patients applied their selected coping strategies [3].

The intervention is based in particular on increasing self-efficacy, the belief in one's own efficacy, applying it in this case to daily walking. We used 3 measures (self-efficacy for barriers [SE-B], self-efficacy for task-specific behavior [SE-T], and SR) of self-efficacy. SE-B is a self-efficacy scale to deal with barriers in achieving the targeted daily steps [9]. It is the sum of 4 questions, each rated 1 - 5, delivered via a questionnaire, amounting to a total score range of 4 - 20. SE-T is a self-efficacy scale for exercise in achieving the targeted daily steps [10]. It is the average of a participant's confidence (in percentages) in their ability to walk 4000; 6000; 8000; 10,000; or 12,000 steps daily, as measured via a questionnaire. SR was measured using the Japanese version of the 12-item Physical Activity Self-Regulation Scale (PASR-12) [11]. It is the sum of 12 questions, each rated 1 - 5, amounting to a total score range of 12 - 60.

We analyzed measurements using means and standard deviations. We assessed correlation via linear regression and analyzed relationships via structural equation modeling. All statistical analyses used SAS 9.4, with the significance level for all the statistical tests at $P<.05$ (2-tailed). We calculated 95% CIs where applicable.

XSL•FO
RenderX

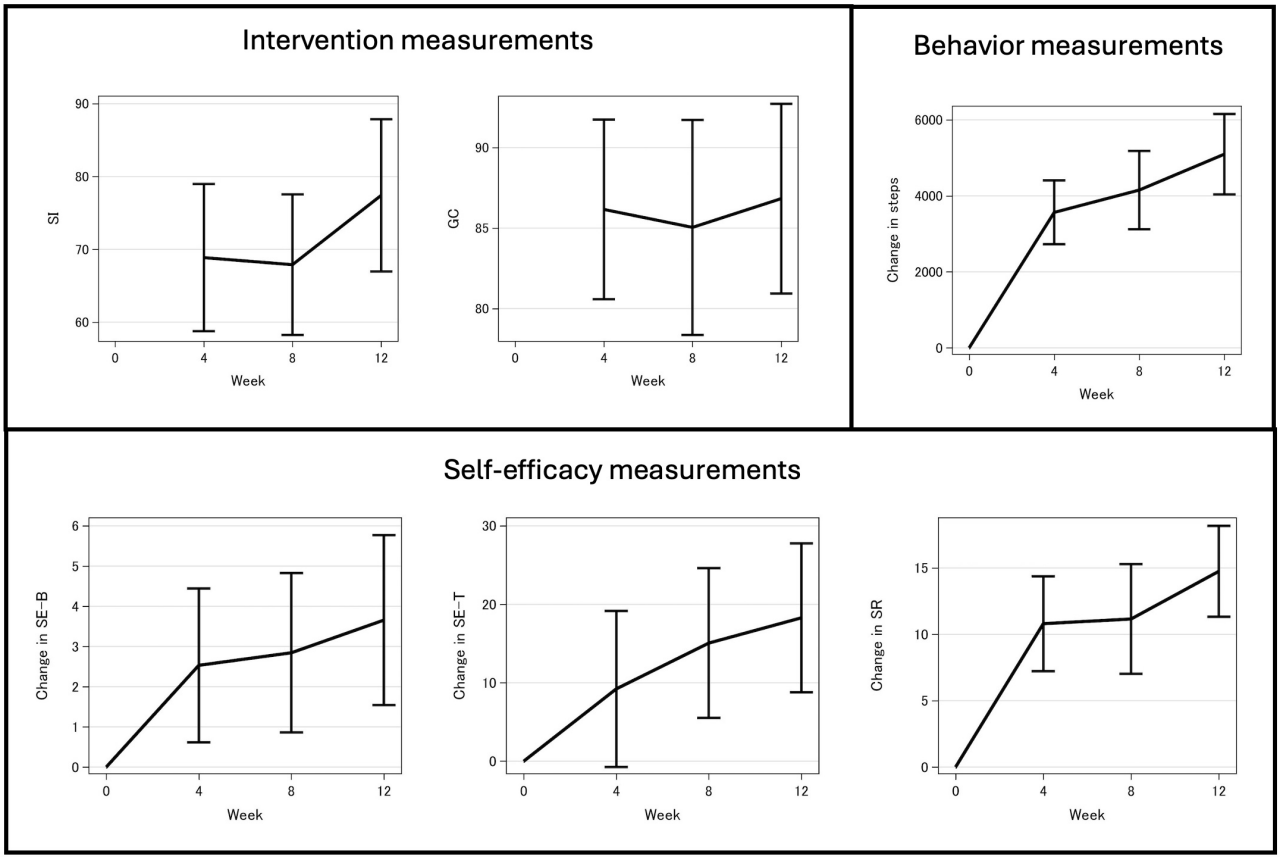
Ethical Considerations

This study was approved by the Institutional Review Board of the University of Tokyo School of Medicine, under approval number 2021084NI-(2). This approval covers secondary analysis without additional consent. Study data were deidentified to ensure confidentiality and to protect all research participants.

Results

The behavior of interest, walking, increased steadily over the 12-week intervention (Figure 1). Both intervention method measurements (GC and SI) were consistently high (68% - 87%). The measures of self-efficacy (SE-B, SE-T, and SR) grew throughout the intervention.

Figure 1. Means and 95% CIs of change in 6 measured items over 12 weeks in the StepAdd intervention for Japanese patients with type 2 diabetes. GC: goal completion; SE-B: self-efficacy for barriers; SE-T: self-efficacy for task-specific behavior; SI: strategy implementation; SR: self-regulation.



Analysis of the correlation of changes from baseline to week 12 among the 6 measurements revealed 6 relationships with statistical significance (Table 1 and Figure 2). When mapped to a simplified representation of SCT applied to this study, these relationships support but do not prove SCT's central mechanism of changes in self-efficacy causing changes in behavior. The correlation coefficient (ρ) of 0.497 between SI and change in SE-B suggests that the coping strategies target of the intervention was effective. Similarly, the ρ of 0.649 between change in steps and GC and the ρ of 0.446 between GC and change in SE-T suggest that the self-mastery target of the intervention was effective. The ρ of 0.355 between change in SR and change in steps supports the centrality of SR in the changes to exercise behavior [3].

XSL•FO
RenderX

Table . Correlations of social cognitive theory constructs at week 12 of the StepAdd intervention for Japanese patients with type 2 diabetes.

	SI[a]	GC[b]	ΔSteps	ΔSE-B[c]	ΔSE-T[d]	ΔSR[e]
SI						
ρ	1					
P value	—[f]					
GC						
ρ	0.267	1				
P value	.140	—				
ΔSteps						
ρ	0.028	0.649	1			
P value	.879	<.001	—			
ΔSE-B						
ρ	0.497	0.254	0.134	1		
P value	.004	.161	.465	—		
ΔSE-T						
ρ	0.328	0.446	0.279	0.358	1	
P value	.067	.01	.122	.04	—	
ΔSR						
ρ	−0.073	0.297	0.355	0.348	0.583	1
P value	.691	.098	.046	.051	<.001	—

[a]SI: strategy implementation.

[b]GC: goal completion.

[c]SE-B: self-efficacy for barriers.

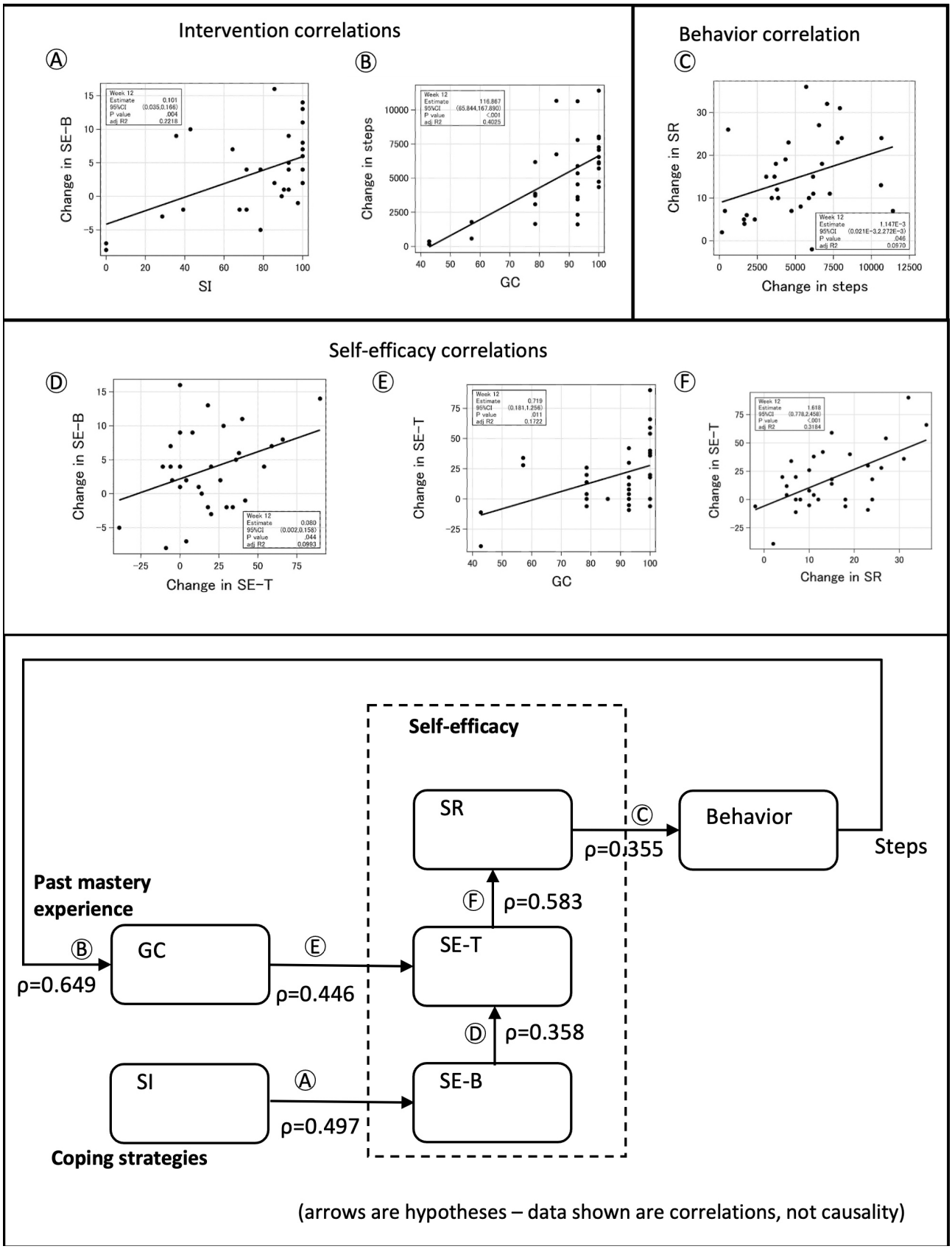
[d]SE-T: self-efficacy for task-specific behavior.

[e]SR: self-regulation.

[f]Not applicable.

XSL•FO
RenderX

Figure 2. Social cognitive theory correlations at week 12 of the StepAdd intervention for Japanese patients with type 2 diabetes. (A) Intervention correlation between change in SE-B and SI. (B) Intervention correlation between change in steps and GC. (C) Behavior correlation between change in SR and change in steps. (D) Self-efficacy correlation between change in SE-B and change in SE-T. (E) Self-efficacy correlation between change in SE-T and GC. (F) Self-efficacy correlation between change in SE-T and change in SR. GC: goal completion; SE-B: self-efficacy for barriers; SE-T: self-efficacy for task-specific behavior; SI: strategy implementation; SR: self-regulation.



XSL•FO
RenderX

We found significant correlations among changes in the 3 measures of self-efficacy (SE-T to SE-B and SR to SE-T). We used structural equation modeling to conduct a confirmatory factor analysis to explore how the 3 self-efficacy–related measures relate to a latent variable of self-efficacy. The analysis with 3 variables and the 32-participant dataset did not support an assessment of goodness-of-fit. The StepAdd RCT currently underway [6] targets 160 patients, and this larger dataset may support a more effective confirmatory factor analysis.

We used cross-lagged panel modeling to explore whether changes in the self-efficacy measures preceded changes in the behavior, a necessary condition for the causality postulated in SCT. The analysis with our 32 participants and 4-week data spacing did not reveal any significant cross-lagged correlations.

Discussion

This study aimed to examine the SCT drivers underlying StepAdd's mean increase in exercise. Overall, our analyses found statistically significant support for the SCT approach in the StepAdd intervention. Significant correlation was found for the following parameters: step increase and GC rate; SR increase and step increase; GC rate and self-mastery increase; and coping SI rate and self-efficacy increase.

Our study findings are consistent with a study that reported that interventions that focus on removing obstacles to intended planned fitness activities are effective in increasing self-efficacy to engage in exercise [12]. Effective use of coping strategies such as problem-solving has been shown to enhance self-efficacy, which in turn promotes engagement in physical activity [13]. Self-mastery, which has been shown to be one of the most important predictors of self-efficacy for physical activity in community-dwelling older adults, is also reflected in our study findings [14]. Achieving goals allows individuals to experience a sense of mastery with authentic evidence of their capabilities, which in turn boosts self-efficacy to increase physical activity [15]. Further, SR strategies such as self-monitoring and goal setting have been associated with exercise adherence in older adults [16].

Despite the significant correlations, there was a lack of support for directional associations from the cross-lagged panel modeling results. A key limitation of this study is the difference between the timescale of behavioral decisions and the measurement intervals. Participants made physical activity decisions multiple times per day, often in response to real-time feedback from the intervention. However, the study's 4-week measurement intervals may not have fully captured these rapid behavior-feedback loops, making it difficult to assess causal relationships between intervention components and behavior change. Assessing causality is an area that requires further investigation. The pilot was relatively small, and further research is needed.

Our results provide insight into designing an effective DTx solution. Although some view DTx as centered on software implementation [17], we emphasize how DTx can remove time and space barriers, not just supporting patients but accompanying them 24/7 where they live, as a cost-effective low-labor augmentation to infrequent scheduled visits with health care professionals in a clinic.

Applying digital health to increase exercise levels is an active area of research [18], and more research is warranted. Interventions that are based on behavior change theory have been shown to be more effective than those not based on theory [19]. In our experience, building an intervention around a theory helps focus the effort on the key aspects, prioritizing them over less critical features. Our group has conducted numerous interventions not specifically based on a theory of human behavior change [2,20-23] that have produced results of interest, but we have found theory-based interventions [3,6,24] to be more effective. For example, the StepAdd study, which employs a theory-based approach, demonstrated a consistent increase in the average step count from week 1 to week 12 of the intervention [3]. In contrast, in the DialBetesPlus study, which is not theory-based, the average step count remained steady throughout the 12-month intervention period [2].

By understanding how the intervention relates to complex human behavior, researchers can put more emphasis on treatment elements that are effective over those that are not. Our results suggest that both of our intervention mechanisms, self-mastery and coping strategies, contributed to the change in behavior, supporting the SCT model of behavior change.

Acknowledgments

We thank Daniel Lane for his support in manuscript editing and scientific discussions. This work was supported by Mitsui & Co, Ltd. The funder played no role in the study design, data collection, analysis and interpretation of data, or writing of this manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request for noncommercial purposes.

Authors' Contributions

KW drafted the manuscript, which was revised and approved by all authors, who also assume responsibility for its content. KW, TY, MN, and KO initiated the study. SE performed the analyses, and all authors discussed the analysis plan and results and provided input to the manuscript. All authors had access to the final study results. KW is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

XSL•FO
RenderX

Conflicts of Interest

A patent resulting from this research work was issued to Raxi Co, Ltd, and the University of Tokyo.

References

1. Digital Therapeutics Alliance. URL: https://dtxalliance.org/ [accessed 2025-03-04]
2. Waki K, Nara M, Enomoto S, et al. Effectiveness of DialBetesPlus, a self-management support system for diabetic kidney disease: randomized controlled trial. NPJ Digit Med 2024 Apr 27;7(1):104. [doi: 10.1038/s41746-024-01114-8] [Medline: 38678094]
3. Sze WT, Waki K, Enomoto S, et al. StepAdd: a personalized mHealth intervention based on social cognitive theory to increase physical activity among type 2 diabetes patients. J Biomed Inform 2023 Sep;145:104481. [doi: 10.1016/j.jbi.2023.104481] [Medline: 37648101]
4. Bandura A. Human agency in social cognitive theory. Am Psychol 1989 Sep;44(9):1175-1184. [doi: 10.1037/0003-066x.44.9.1175] [Medline: 2782727]
5. Beauchamp MR, Crawford KL, Jackson B. Social cognitive theory and physical activity: mechanisms of behavior change, critique, and legacy. Psychol Sport Exerc 2019 May;42:110-117. [doi: 10.1016/j.psychsport.2018.11.009]
6. Waki K, Tsurutani Y, Waki H, et al. Efficacy of StepAdd, a personalized mHealth intervention based on social cognitive theory to increase physical activity among patients with type 2 diabetes mellitus: protocol for a randomized controlled trial. JMIR Res Protoc 2024 Feb 23;13:e53514. [doi: 10.2196/53514] [Medline: 38393770]
7. Qiu S, Cai X, Schumann U, Velders M, Sun Z, Steinacker JM. Impact of walking on glycemic control and other cardiovascular risk factors in type 2 diabetes: a meta-analysis. PLoS One 2014;9(10):e109767. [doi: 10.1371/journal.pone.0109767] [Medline: 25329391]
8. Moghetti P, Balducci S, Guidetti L, et al. Walking for subjects with type 2 diabetes: a systematic review and joint AMD/SID/SISMES evidence-based practical guideline. Nutr Metab Cardiovasc Dis 2020 Oct 30;30(11):1882-1898. [doi: 10.1016/j.numecd.2020.08.021] [Medline: 32998820]
9. Yamawaki K, Harada K, Lee E, Oka K, Nakamura Y. Development of the stages of change and self-efficacy scales for walking behavior [Article in Japanese]. Nikken Kyogaku (Jpn J Health Educ Promot) 2009;17(2):87-96. [doi: 10.11260/kenkokyoiku.17.87]
10. Fukushima H. Development of a physical activity self-efficacy scale for young people [Article in Japanese]. Bulletin of the Faculty of Human Development Sciences 2011;6(1):99-106 [FREE Full text]
11. Watanabe K, Kawakami N, Adachi H, Inoue S, Meyer MRU. Internal consistency, convergent validity, and structural validity of the Japanese version of the Physical Activity Self-Regulation scale (PASR-12) among Japanese workers: a validation study. J Occup Health 2017 Jan 24;59(1):24-32. [doi: 10.1539/joh.16-0143-OA] [Medline: 27885244]
12. Artistico D, Pinto AM, Douek J, Black J, Pezzuti L. The value of removing daily obstacles via everyday problem-solving theory: developing an applied novel procedure to increase self-efficacy for exercise. Front Psychol 2013;4:20. [doi: 10.3389/fpsyg.2013.00020] [Medline: 23372560]
13. Pekmezi D, Jennings E, Marcus BH. Evaluating and enhancing self-efficacy for physical activity. ACSMs Health Fit J 2009;13(2):16-21. [doi: 10.1249/FIT.0b013e3181996571] [Medline: 29910597]
14. Warner LM, Schüz B, Wolff JK, Parschau L, Wurm S, Schwarzer R. Sources of self-efficacy for physical activity. Health Psychol 2014 Nov;33(11):1298-1308. [doi: 10.1037/hea0000085] [Medline: 24707842]
15. Bandura A. Self-Efficacy: The Exercise of Control: W H Freeman & Co; 1997.
16. McAuley E, Mullen SP, Szabo AN, et al. Self-regulatory processes and exercise adherence in older adults: executive function and self-efficacy effects. Am J Prev Med 2011 Sep;41(3):284-290. [doi: 10.1016/j.amepre.2011.04.014] [Medline: 21855742]
17. Wang C, Lee C, Shin H. Digital therapeutics from bench to bedside. NPJ Digit Med 2023 Mar 10;6(1):38. [doi: 10.1038/s41746-023-00777-z] [Medline: 36899073]
18. Mönninghoff A, Kramer JN, Hess AJ, et al. Long-term effectiveness of mHealth physical activity interventions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2021 Apr 30;23(4):e26699. [doi: 10.2196/26699] [Medline: 33811021]
19. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010;31:399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
20. Waki K, Fujita H, Uchimura Y, et al. DialBetics: smartphone-based self-management for type 2 diabetes patients. J Diabetes Sci Technol 2012 Jul 1;6(4):983-985. [doi: 10.1177/193229681200600433] [Medline: 22920828]
21. Waki K, Fujita H, Uchimura Y, et al. DialBetics: a novel smartphone-based self-management support system for type 2 diabetes patients. J Diabetes Sci Technol 2014 Mar;8(2):209-215. [doi: 10.1177/1932296814526495] [Medline: 24876569]
22. Kondo M, Okitsu T, Waki K, Yamauchi T, Nangaku M, Ohe K. Effect of information and communication technology-based self-management system dialbeticslite on treating abdominal obesity in the specific health guidance in Japan: randomized controlled trial. JMIR Form Res 2022 Mar 24;6(3):e33852. [doi: 10.2196/33852] [Medline: 35323122]

XSL•FO
RenderX

23. Yamaguchi S, Waki K, Nannya Y, Nangaku M, Kadowaki T, Ohe K. Usage patterns of GlucoNote, a self-management smartphone app, based on ResearchKit for patients with type 2 diabetes and prediabetes. JMIR Mhealth Uhealth 2019 Apr 24;7(4):e13204. [doi: 10.2196/13204] [Medline: 31017586]
24. Shibuta T, Waki K, Miyake K, et al. Preliminary efficacy, feasibility, and perceived usefulness of a smartphone-based self-management system with personalized goal setting and feedback to increase step count among workers with high blood pressure: before-and-after study. JMIR Cardio 2023 Jul 21;7:e43940. [doi: 10.2196/43940] [Medline: 37477976]

Abbreviations

DTx: digital therapeutics
GC: goal completion
HbA_{1c}: hemoglobin A_{1c}
mHealth: mobile health
PASR-12: 12-item Physical Activity Self-Regulation Scale
RCT: randomized controlled trial
SCT: social cognitive theory
SE-B: self-efficacy for barriers
SE-T: self-efficacy for task-specific behavior
SI: strategy implementation
SR: self-regulation

Edited by A Mavragani; submitted 05.05.24; peer-reviewed by A Woodward; revised version received 15.02.25; accepted 16.02.25; published 28.03.25.

Please cite as:
Waki K, Enomoto S, Yamauchi T, Nangaku M, Ohe K
Personalized mHealth Intervention (StepAdd) for Increasing Physical Activity in Japanese Patients With Type 2 Diabetes: Secondary Analysis of Social Cognitive Theory Measurements of a Single-Arm Pilot Study
JMIR Form Res 2025;9:e60221
URL: https://formative.jmir.org/2025/1/e60221
doi:10.2196/60221

© Kayo Waki, Syunpei Enomoto, Toshimasa Yamauchi, Masaomi Nangaku, Kazuhiko Ohe. Originally published in JMIR Formative Research (https://formative.jmir.org), 28.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Accuracy of Smartphone-Mediated Snore Detection in a Simulated Real-World Setting: Algorithm Development and Validation

Jeffrey Brown, MD; Zachary Mitchell, BS; Yu Albert Jiang, MS; Ryan Archdeacon, BS

Bodymatter, Inc, 4343 Von Karman Ave, Suite 150J, Newport Beach, CA, United States

Corresponding Author:
Jeffrey Brown, MD
Bodymatter, Inc, 4343 Von Karman Ave, Suite 150J, Newport Beach, CA, United States

Abstract

Background: High-quality sleep is essential for both physical and mental well-being. Insufficient or poor-quality sleep is linked to numerous health issues, including cardiometabolic diseases, mental health disorders, and increased mortality. Snoring—a prevalent condition—can disrupt sleep and is associated with disease states, including coronary artery disease and obstructive sleep apnea.

Objective: The SleepWatch smartphone app (Bodymatter, Inc) aims to monitor and improve sleep quality and has snore detection capabilities that were built through a machine-learning process trained on over 60,000 acoustic events. This study evaluated the accuracy of the SleepWatch snore detection algorithm in a simulated real-world setting.

Methods: The snore detection algorithm was tested by using 36 simulated snoring audio files derived from 18 participants. Each file simulated a snoring index between 30 and 600 snores per hour. Additionally, 9 files with nonsnoring sounds were tested to evaluate the algorithm's capacity to avoid false positives. Sensitivity, specificity, and accuracy were calculated for each test, and results were compared by using Bland-Altman plots and Spearman correlation to assess the statistical association between detected and actual snores.

Results: The SleepWatch algorithm showed an average sensitivity of 86.3% (SD 16.6%), an average specificity of 99.5% (SD 10.8%), and an average accuracy of 95.2% (SD 5.6%) across the snoring tests. The positive predictive value and negative predictive value were 98.9% (SD 2.6%) and 93.8% (SD 14.4%) respectively. The algorithm performed exceptionally well in avoiding false positives, with a specificity of 97.1% (SD 3.5%) for nonsnoring files. Inclusive of all snoring and nonsnore tests, the aggregated accuracy for all trials in this bench study was 95.6% (SD 5.3%). The Bland-Altman analysis indicated a mean bias of −29.8 (SD 41.7) snores per hour, and the Spearman correlation analysis revealed a strong positive correlation (r_s=0.974; P<.001) between detected and actual snore rates.

Conclusions: The SleepWatch snore detection algorithm demonstrates high accuracy and compares favorably with other snore detection apps. Aside from its broader use in sleep monitoring, SleepWatch demonstrates potential as a tool for identifying individuals at risk for sleep-disordered breathing, including obstructive sleep apnea, on the basis of the snoring index.

(JMIR Form Res 2025;9:e67861) doi:10.2196/67861

KEYWORDS

snore detection; snore tracking; machine learning; SleepWatch; Bodymatter; neural net; mobile device; smartphone; smartphone application; mobile health; sleep monitoring; sleep tracking; sleep apnea

Introduction

The central importance of high-quality sleep in the maintenance of physical and mental health cannot be overstated, as well as the critical role of sleep in avoiding disease states and other physiologic impairments. Insufficient or poor-quality sleep has extensively documented associations with an array of negative consequences that affect both personal and public health. These issues include cardiometabolic disease states, such as hypertension, diabetes, and obesity, as well as increased frequency of heart attacks and strokes [1-3]. Sleep disruption profoundly impacts mental health and emotional well-being, increasing the frequency of depression, anxiety, inattentiveness, learning and memory impairment, and mood disorders [4-6]. From a public health perspective, poor-quality sleep correlates with workplace and vehicular accidents, and health issues arising from sleep impairment impart a massive economic burden to the health care system, as well as an increase in all-cause mortality [2,7-9].

Although the health risks of insufficient or poor-quality sleep are well known among the scientific community and society at large, most individuals struggle to obtain sleep of appropriate quality and quantity. The frequency of insufficient sleep varies by study, but as a conservative estimate, at least one-third of

XSL•FO
RenderX

the US adult population reports regularly sleeping less than the recommended 7 hours each night [10-12]. Moreover, there is evidence that sleep problems are worsening over time globally; the percentage of US adults reporting short sleep durations (defined as less than 6 h) has consistently increased over the last several decades [13,14]. The precise cause of this declining trend in sleep health is multifactorial but is influenced in large part by poor sleep hygiene related to the ongoing technological evolution of the modern world [15,16]. These concerns emphasize the need for interventions to address the profound personal and societal issue of sleep impairment.

Among other factors related to poor-quality sleep, a noteworthy and underappreciated cause of sleep impairment is the physiological phenomenon of snoring. Snoring is an auditory process that occurs during sleep as a result of air passing over relaxed tissues of the oropharynx, which produces vibrations and the sounds that are classically associated with snoring. Snoring is extremely common, especially among older adults, male individuals, and individuals with obesity; the reported frequency of habitual snoring among adults ranges from 20% to 60%, depending on risk factors [17-20]. Snoring is often benign but may be a source of personal embarrassment and a cause of relationship distress between sleep partners [19,21]. On a more concerning note, snoring may be associated with serious clinical conditions, including coronary artery disease, obstructive sleep apnea, and depressive disorders [17,22,23].

As with other areas of health optimization, smartphone-based apps have been developed to help address the need for improved sleep quality and sleep duration, and several such apps include snore detection capabilities. Because snoring occurs during sleep, individuals often lack insight into how frequently or how intensely they snore, and obtaining this information through formal polysomnography can be inconvenient, time-consuming, and cost-prohibitive. As such, there is an unmet need for novel, personalized methods of monitoring sleep quality, including evaluation of snoring and sleep-disordered breathing. There are several smartphone-based snoring detection apps in existence, and while these apps are promising in terms of their cost and scalability, most lack rigorous validation, and questions remain with regard to their real-world accuracy for individual users. This paper highlights the results of a bench study that evaluated a proprietary snore detection algorithm in a simulated real-world setting.

Methods

Study Design

The smartphone-based, acoustical snore detection algorithm used in the SleepWatch app was developed by the Bodymatter team in-house, using a deep neural net model trained on over 60,000 individually validated, real-world snore and nonsnore sounds. The SleepWatch snore detection capabilities were then tested against an array of snoring and sleep audio files in a controlled acoustic setting. Sleep audio files were derived from volunteers in an anonymized fashion, as outlined in the *Ethical Considerations* subsection. No raw or identifying human participant data were used in this study. Additionally, all training data were derived entirely from audio sources that were separate from the validation and test audio sets used in this bench study and prior trials.

In total, 36 trials were conducted based on test audio files that were 10 minutes in duration and were derived from 18 individual SleepWatch participants (male: n=9; female: n=9). Audio quality fidelity of the test files was ensured during their editing, exporting, and uploading. There were 2 audio files made from each participant, simulating a snoring index of 30, 60, 120, 240, 360, or 600 snores per hour. For each file with a snoring index of 120, 240, 360, or 600 snores per hour, 10 unique snores were repeated at regular intervals. For each file with a snoring index of 30 or 60 snores per hour, 5 unique snores were repeated at regular intervals. The duration of snore events was limited to a maximum of 5 seconds, in accordance with the scoring methodology described further in this paper. The intervening audio between inserted acoustic events was the ambient noise present at baseline during a given participant's sleep recording. Additionally, 9 trials were conducted based on test audio files that exclusively contained confounding, nonsnoring audio events (including percussive sounds, coughing, speech sounds, movement of bedding, etc). These confounding audio events reflected common ambient noise that was expected during typical sleeping conditions and had the potential to be erroneously detected as a snore. Again, no confounding acoustic event lasted longer than 5 seconds, in accordance with the scoring methodology. Of note, any audio samples that were determined to have corruption at any level of the editing or exporting process and resulted in distorted playback were discarded.

For each trial, a SleepWatch sleep recording session was started on an iPhone 13 smartphone (Apple Inc) that ran the SleepWatch app (v8.2.3.0-pkf33, public release v8.3.0) shortly before simultaneously starting the test audio playback. Test audio files were played from a dedicated JBL 305P MKII (Harman International Industries, Inc) studio speaker in close proximity to the detection smartphone to simulate product implementation in a real-world setting, where users would place their smartphone within 5 feet of their sleeping position. Trials were conducted under acoustically controlled ambient conditions. For each test, audio playback intensity was confirmed via an independent decibel meter and limited to a 125-millisecond (fast response) peak of 65 A-weighted decibels (dB[A]) and a floor of 50 dB(A). Trials were performed for 10 minutes, after which the SleepWatch app performance was recorded.

After each trial, the SleepWatch-detected snore count was compared against the known number of snoring events present in each audio file. Additionally, app performance was assessed by reporting sensitivity (true-positive rate), specificity (true-negative rate), and accuracy (ratio of true events to total events). As a scoring method, each audio file was discretized into 5-second intervals (N=120). An appropriately identified snoring event was considered a true-positive event for a 5-second interval, and conversely, a sound that was inaccurately reported as snoring was considered a false-positive event for the corresponding 5-second interval. False-negative events were missed snores, and true-negative events were 5-second regions of silence or ambient noise that were not reported as snoring. Sensitivity was calculated as "(true positive)/(true positive +

XSL•FO
RenderX

false negative)," and specificity was calculated as "(true negative)/(true negative + false positive)." Positive and negative predictive values were also determined, calculated as "(true positive)/(true positive + false positive)" and "(true negative)/(false negative + true negative)," respectively. Accuracy was determined as "(true positive + true negative)/all events" (all events: N=120). In the case of the 9 trials that used confounding nonsnore sounds, only specificity and accuracy could be described due to the absence of positive events, and this was factored into the aggregate accuracy. This scoring methodology is summarized in Table 1.

After the initial analysis, a Bland-Altman plot, which compared the average snoring index (snores/h) and difference in the snoring index between the app and test audio files, was produced as a visual method of analyzing bias associated with the app's performance. A Spearman correlation was performed between known and detected snoring index values, and the correlation coefficient (r_s) and statistical significance were determined. All statistical analyses were completed by using IBM SPSS Statistics (version 29.0.2.0; IBM Corp).

Table . Analytical validation reference table showing the schematic and equations[a,b,c,d,e] used to report snore detection performance of the SleepWatch app in this study.

		Constructed sleep audio (gold standard)	
		Snoring occurred	Snoring did not occur
Smartphone app			
	Snoring detected	True positive	False positive
	Snoring not detected	False negative	True negative

[a]Accuracy = (true positive + true negative)/(true positive + true negative + false positive + false negative).

[b]Sensitivity = true positive/(true positive + false negative).

[c]Specificity = true negative/(false positive + true negative).

[d]Positive predictive value = true positive/(true positive + false positive).

[e]Negative predictive value = true negative/(false negative + true negative).

Ethical Considerations

A principal aim of this study was to closely reproduce a real-world environment for the validation of Bodymatter software. As such, raw audio files were voluntarily obtained from individuals who were willing to participate in this study. Written consent was obtained for the use of their sleep audio for internal testing purposes. No compensation was provided to participants for their participation in this study. All audio files were anonymized at both the level of audio editing and the level of audio playback to ensure participant privacy. As no direct intervention was performed on study participants and this study posed no risk to participants, this study was granted an exemption from formal institutional review board approval for human subjects research.

Results

Across the 18 participants in this study, 9 were male and 9 were female. The age distribution skewed toward a younger adult demographic, with 28% (5/18) of participants aged 25 to 34 years, 44% (8/18) aged 35 to 44 years, 22% (4/18) aged 45 to 54 years, and 6% (1/18) aged 55 to 64 years. The average BMI was 31.21 (SD 5.03) kg/m^2 for male participants and 33.93 (SD 8.07) kg/m^2 for female participants. With respect to the presence of sleep partners, 56% (10/18) of participants reported sleeping alone, and 44% (8/18) reported having a consistent sleep partner.

The snore detection performance was evaluated for the 36 snoring tests and 9 confounding noise tests. During the development and internal validation of the SleepWatch snore detection algorithm, several trends were observed. The SleepWatch snore detection algorithm has a strong capacity to appropriately designate true-positive events (sensitivity) and an extremely strong capacity to appropriately define true-negative events (specificity); that is, the algorithm almost never reports a snore when one did not occur. Table 2 reports the sensitivity, specificity, and accuracy for both the snoring tests and the confounding audio (nonsnore) tests. The cumulative sensitivity for the snoring tests was 86.3% (SD 16.6%), the cumulative specificity was 99.5% (SD 10.8%), and the cumulative accuracy was 95.2% (SD 5.6%). The positive and negative predictive values were 98.9% (SD 2.6%) and 93.8% (SD 14.4%), respectively.

Consistent with the performance observed for the dedicated snore audio files, the SleepWatch algorithm continued to perform exceptionally well for the dedicated nonsnore audio files, with an overall specificity (ie, the true-negative capacity) of 97.1% (SD 3.5%). The sensitivity, positive predictive value, and negative predictive value cannot be reported, as there were no snoring or positive events in these audio files, but a false-positive rate of 2.9% was observed across all the noise audio files—defined as the ratio of erroneously detected snores to the total number of nonsnoring time intervals. Inclusive of the nonsnore audio files, the aggregated accuracy for all trials in this bench study was 95.6% (SD 5.3%).

XSL•FO
RenderX

Table . Detection performance of the SleepWatch app for cumulative, snoring, and nonsnore noise audio tests. The results were produced from 36 snoring trials and 9 nonsnoring noise trials.

Test condition	Sensitivity (%), mean (SD)	Specificity (%), mean (SD)	Positive predictive value (%), mean (SD)	Negative predictive value (%), mean (SD)	Accuracy (%), mean (SD)
Snore audio	86.3 (16.6)	99.5 (10.8)	98.9 (2.6)	93.8 (14.4)	95.2 (5.6)
Nonsnore audio	—[a]	97.1 (3.5)	—	—	97.1 (3.5)
Cumulative	86.3 (16.6)	98.0 (9.8)	98.9 (2.6)	93.8 (14.4)	95.6 (5.3)

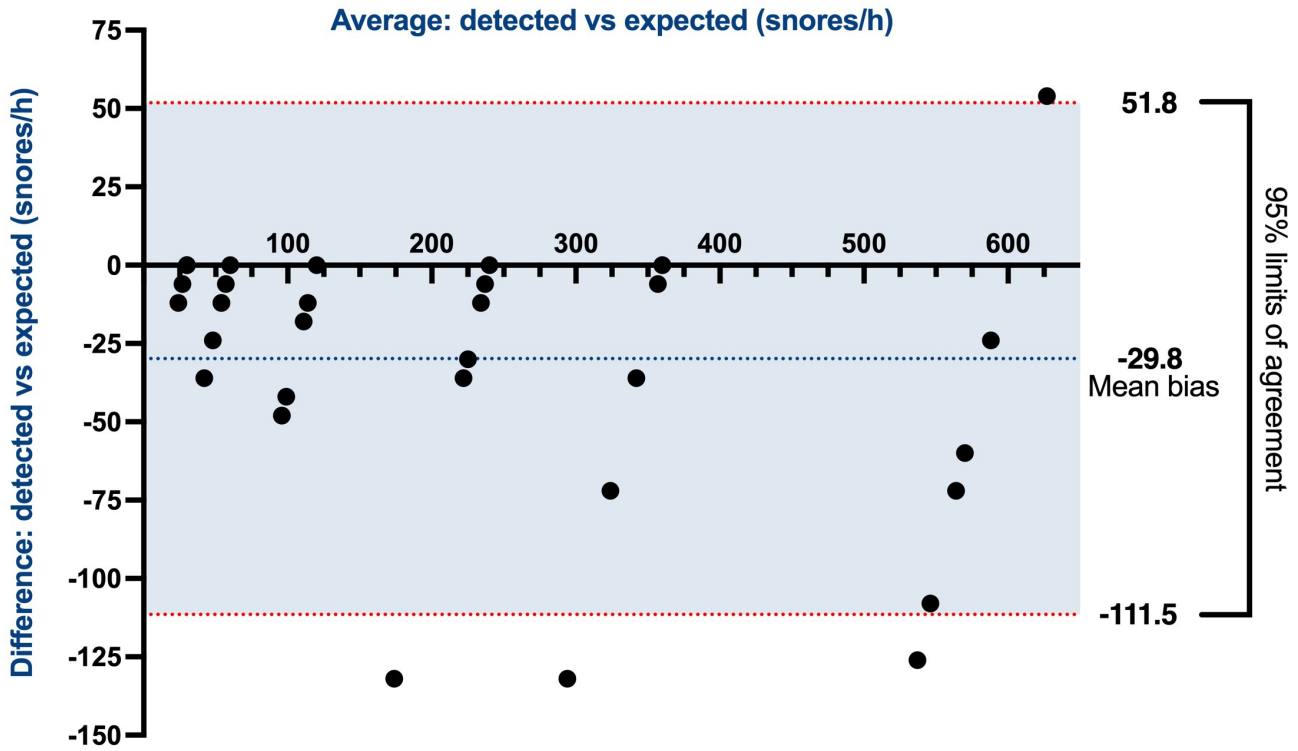
[a]Not applicable.

Of note, the detected false-positive events were not evenly distributed across files. Algorithm performance was strong for coughing, percussive sounds, and loud speaking, for which no false-positive events were observed. Intermediate performance was observed for movement sounds, such as sounds resulting from pacing around a room, climbing stairs, or rustling pillows or bedding. Algorithm performance was poor for a specific file that included whispering and sleep talking, for which a false-positive rate of 10.8% was observed.

The results of the 36 snoring tests were compiled and displayed in a Bland-Altman plot (Figure 1). The average of the known versus detected snoring index (snores/h) was plotted on the x-axis, and the difference in the known versus detected snoring index was plotted on the y-axis. As expected, there were roughly 6 vertical data clusters around the x values 30, 60, 120, 240, 360, and 600 snores per hour, corresponding to the planned variations in snoring rate for each audio file. If the average of the known versus detected snores per hour aligned closely with these six x values, then excellent performance was interpreted. In contrast, if the difference in the known versus detected snores per hour was close to 0 on the y-axis across all tests, irrespective of the snoring rate, then ideal performance was also indicated. The mean bias of the SleepWatch app for this study was −29.8 (SD 41.7) snores per hour, indicating an average difference of approximately 30 snores per hour between the detected and expected snores. Most values were contained within the 95% limits of agreement (51.8 and −111.5 snores/h).

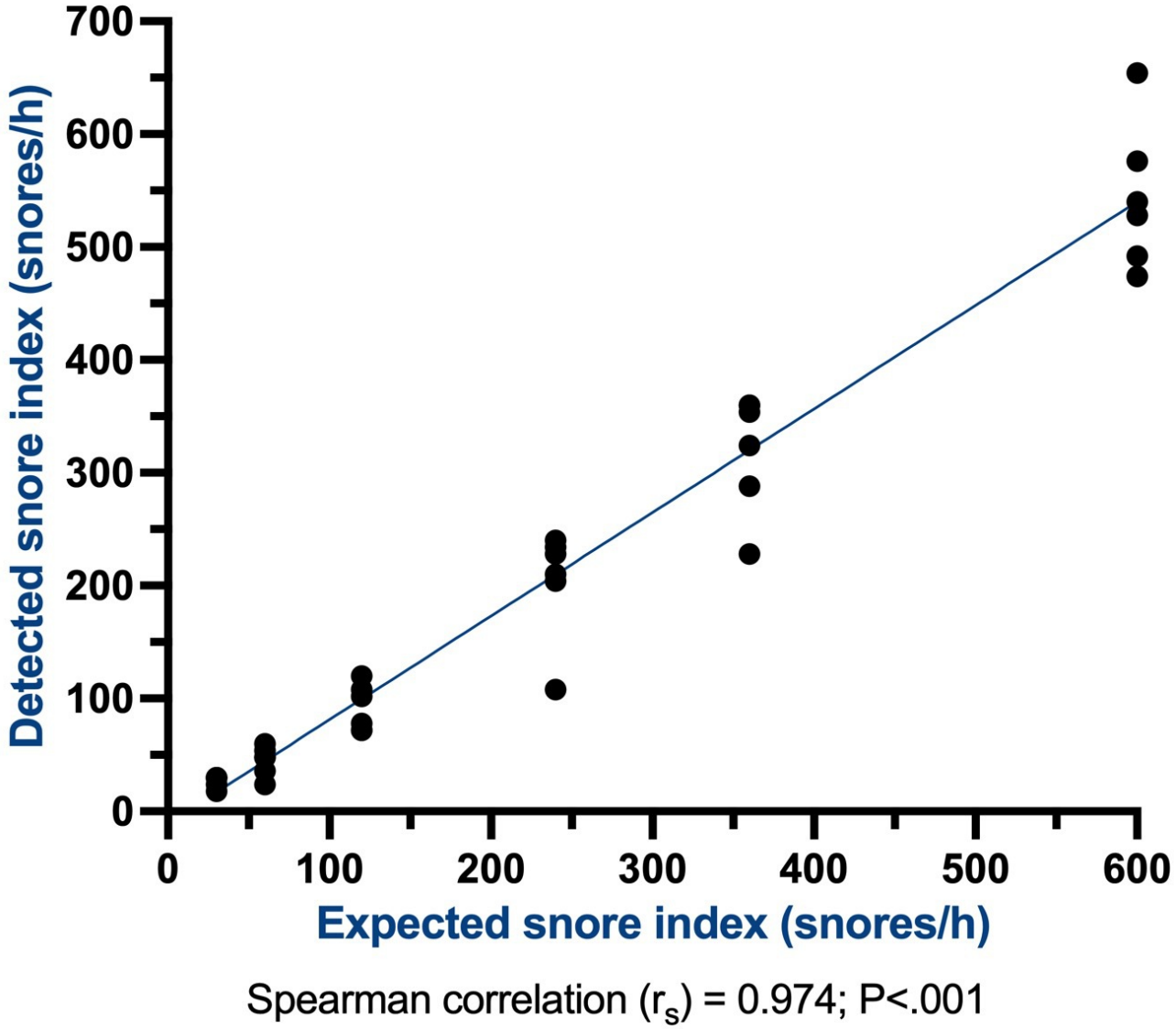
A Spearman correlation coefficient was calculated to statistically evaluate the association between the detected and known snoring index (snores/h) in each test audio file (Figure 2). The Spearman coefficient (r_s) for this cohort was 0.974, denoting an extremely strong, positive correlation between detected snores and expected snores (P<.001). These findings are highly encouraging and demonstrate a robust capacity for identifying true snores in a simulated real-world setting.

Figure 1. Bland-Altman plot of the difference in and average of the detected versus expected snores/h. The Bland-Altman plot demonstrates a mean bias of –29.8 (95% limits of agreement –111.5 to 51.8) snores/h for the average of the detected versus expected snores/h graphed against the difference in the detected versus expected snores/h. This demonstrates a relatively tight relationship between detected and expected snores.



XSL•FO
RenderX

Figure 2. The expected snore index graphed against the snore index detected by the SleepWatch app. This demonstrates a tight grouping and a Spearman correlation coefficient of 0.974, indicating a strong, positive correlation between expected and detected snores.



Discussion

SleepWatch Snore Detection Accuracy Compares Favorably to Similar Snore Detection Modalities

The SleepWatch app provides users with consistent, valuable insight into their sleep quality, and the snore detection feature constitutes an accuracy benchmark for acoustic snore detection software. This study reports the performance of the SleepWatch snore detection algorithm in detail; the algorithm showed an average sensitivity of 86.3% (SD 16.6%), an average specificity of 99.5% (SD 10.8%), and an average accuracy of 95.2% (SD 5.6%). Furthermore, the accuracy of the SleepWatch snore detection feature reflects performance that is superior to that of similar apps, and this feature is solely dependent on existing smartphone hardware; thus, it has powerful implications for improving sleep quality and screening for sleep-disordered breathing on an individualized basis.

There are several snore detection modalities in existence that have published results with respect to the accuracy of snore detection. The accuracy, sensitivity, specificity, positive predictive values, and negative predictive values of such modalities are summarized in Table 3, though these results must be compared with caution due to the differing methodologies between studies. Shin et al [24] integrated and analyzed a series of sleep and noise datasets for 10 individuals, attempting to distinguish verified snoring from ambient noise (playing music, running a fan, talking, etc). They reported a 95.1% overall accuracy in detecting snores and a sensitivity and specificity of 98.6% and 94.6%, respectively, as well as a positive and negative predictive value of 70.4% and 99.8%, respectively. Chiang et al [25] evaluated the sleep records of 11 patients using the Snore Clock app. The authors reported an overall mean snore detection accuracy of 95% and a sensitivity and specificity of 78.2% and 97%, respectively, along with a positive and negative predictive value of 65.3% and 97%, respectively. They also commented that the false-negative rate (type II error) was higher in patients with greater snoring frequencies. In a cohort of 19 patients, Klaus et al [26] compared the performance of the SnoreLab app with that of formal polysomnography, reporting an accuracy of 94.7% and a sensitivity and specificity of 100% and 94.1%, respectively, as well as a positive and

XSL•FO
RenderX

negative predictive value of 66.6% and 100%, respectively. These findings leave room for misinterpretation however, as these statistics were observed only in a group of participants that snored for >50% of the time recorded, constituting approximately 10% of the cohort. Inclusive of users who snored for >5% of the time recorded—a far less restrictive threshold comprising 42% of the participants—the accuracy, sensitivity, and specificity were 73.6%, 100%, and 54.5%, respectively, with a positive and negative predictive value of 61.5% and 100%. Klaus et al [26] also noted that the app had a strong tendency to overestimate the ratio of snoring time to quiet sleep time, relative to the comparison polysomnogram.

Table . Cross-study comparison of our study's snore detection app and other snore detection modalities reported in available literature. Our study's modality compares favorably against similar snore detection modalities relying on acoustic input.

Study	Accuracy, %	Sensitivity, %	Specificity, %	Positive predictive value, %	Negative predictive value, %
Our study	95.6	86.3	99.5	98.9	93.8
Shin et al [24][a]	95.1	98.6	94.6	70.4	99.8
Klaus et al [26][b]	73.6	100	54.5	61.5	100
Chiang et al [25][c]	95	78.2	97	65.3	97

[a]Shin et al [24] only used a smartphone for audio recording; snore detection was processed through a separate modality.

[b]Statistics from the Klaus et al [26] study are reported for participants with a snoring ratio of >5%, inclusive of the largest subset of participants evaluated.

[c]Chiang et al [25] selected participants from patients with known sleep apnea who were using an oral device that was intended for reducing snoring.

Collectively, the previously mentioned studies suggest that there is reasonable accuracy for smartphone-based snoring detection software, but there are several limitations with comparing our study to those snore detection studies. Notably, only 2 of those studies used working smartphone apps that recorded and processed snoring. Shin et al [24] only used a smartphone as a recording device and processed sleep audio and snore detection through a separate modality. Chiang et al [25] and Klaus et al [26] used smartphone apps with a methodology that was consistent with our study. However, Chiang et al [25] recruited participants with known sleep-disordered breathing or obstructive sleep apnea who had received a mandibular advancement device that was intended for reducing airway obstruction and snoring, which could potentially introduce a degree of selection bias and thereby limit cross-study comparison. Further, as previously stated, Klaus et al [26] reported results for an arbitrarily defined cohort that was considered to have spent >50% of the sleep session snoring, relative to quiet sleeping. App performance was far worse when the snoring threshold was >5%, with an accuracy of 73.6% in that subcohort. This discrepancy in methodology imparts limitations in the comparison of snore detection statistics between the study by Klaus et al [26] and our study, making cross-study comparison largely irrelevant.

The SleepWatch app by Bodymatter stands out for several reasons. Given that only 2 of the previously referenced studies involved a working smartphone app running snore detection software, many of the processes and protocols in the referenced studies do not seem practical to implement on an individualized, large-scale basis, whether due to concerns regarding privacy or concerns regarding convenience. Furthermore, the SleepWatch snore detection feature far outcompetes the referenced studies' modalities in positive predictivity and specificity, which, in practical language, denotes that (1) there is an extremely high likelihood that a true snore actually occurred when a snore was identified by the algorithm, and (2) the app has an extremely strong capacity to appropriately discriminate silence and nonsnoring noise events, with the latter representing one of the most significant challenges in snore detection. This reflects an engineering priority of avoiding false-positive reported snores in the development of the Bodymatter SleepWatch algorithm. Finally, this study carries the advantages of reporting performance on a snore-by-snore basis and including ambient noise and confounding sound events in the testing of the snore detection software, thereby evaluating performance in conditions that more closely reflect real-world settings and expected use.

Screening for Sleep-Disordered Breathing

As previously described, snoring is independently associated with an array of disease states, and snore detection smartphone apps provide individual users with insight into the presence and frequency of their snoring. Loud snoring is a cardinal symptom of obstructive sleep apnea—a disease characterized by airway collapse during sleep, which results in intermittent deoxygenation, apneic events, and arousal at night [27]. Furthermore, obstructive sleep apnea results in daytime sleepiness and an associated increase in accidents, as well as an increased risk of hypertension, coronary artery disease, depression, and death [28,29].

Diagnostic criteria for obstructive sleep apnea center around the apnea-hypopnea index (AHI), which describes the number of episodes where an individual stops breathing due to airway obstruction (apnea) or experiences decreased oxygenation due to inadequate breathing (hypopnea). A diagnosis of moderate-severity sleep apnea is made when >15 apneic or hypopneic events occur in 1 hour (AHI>15) [30]. Previous studies have described a positive correlation between the snoring index and AHI and, thereby, a positive correlation with the diagnosis of obstructive sleep apnea [31,32]. Moreover, it has been estimated that as many as 85% of individuals with obstructive sleep apnea are undiagnosed, highlighting the importance of novel, affordable, and scalable tools for disease screening [33].

XSL•FO
RenderX

These findings suggest that there is utility in exploring the SleepWatch app as a screening tool for stratifying users along a risk spectrum for sleep-disordered breathing, including obstructive sleep apnea. As previously mentioned, the snoring index alone is a useful metric for assessing user risk, aside from formal diagnostic criteria, such as apneic and hypopneic events. Ultimately, there is both significant need and potential for at-home technological innovations to assist with screening for underdiagnosed health conditions, including obstructive sleep apnea, and SleepWatch may have potential uses in a clinical context, aside from providing individual users with insights into their sleep health.

Limitations

Although the reported results of the SleepWatch snore detection feature are promising, this study must be evaluated with several limitations in mind. As mentioned previously, while false-positive snore events were infrequent, it is difficult to entirely eradicate such events. This is a priority for the future development of the snore detection algorithm. Further, this initial study involved a relatively small cohort, and a study with a larger sample size might more comprehensively evaluate the performance of the snore detection algorithm, due to a greater degree of acoustic heterogeneity. Additionally, there are intrinsic limitations in participant-derived audio collection, as individuals may place their phones at varying distances from the sleeping location, and the ambient acoustic environment naturally differs from person to person. From another perspective however, this acoustic diversity may represent a strength of this study, as it provides evidence that app performance is robust, irrespective of ambient sound conditions.

Conclusion

The SleepWatch app by Bodymatter has a strong track record of providing users with individualized insights into their sleep hygiene and represents a powerful and scalable modality for improving overall sleep quality. This study evaluated the performance of the deep neural net algorithm powering the SleepWatch snore detection feature. Over a series of trials involving snoring and nonsnoring sounds from 18 participants, a strong, positive correlation (r_s=0.974; P<.001) was found between detected and reference snoring indices that ranged from 30 to 600 snores per hour. The snore detection feature exhibited an overall accuracy (95.6%) that was competitive with those of similar modalities, as well as a remarkable ability to discriminate true snores from nonsnores, as reflected by the superior specificity and positive predictive value. This positions SleepWatch as a powerful tool for monitoring sleep quality and snoring. Although financial, geographic, and time constraints may prevent individuals from accessing traditional health and wellness resources, SleepWatch represents an accessible, scalable, and intuitive instrument for monitoring an individual's snoring over time and potentially identifying individuals at risk for sleep-disordered breathing, thus helping reduce the growing disease burden of impaired sleep.

Acknowledgments

No use of generative artificial intelligence was involved in the background research, writing, or editing of this study, and the writing in this paper represents original work and research by its contributors.

Data Availability

Consistent with the confidentiality agreement outlined for participants in this study, the raw and edited audio recordings used in this study are unavailable for public review. However, a summary table reporting the app performance for each participant is available in Multimedia Appendix 1. Readers may contact the corresponding author for further requests or information regarding access to data.

Conflicts of Interest

JB is a paid employee of Bodymatter, Inc. ZM and AJ are paid employees of and have stock incentives in Bodymatter, Inc. RA is a paid employee and board member of Bodymatter, Inc, and has stock incentives in Bodymatter, Inc.

Multimedia Appendix 1
Snore detection results broken down by participant, with app performance metrics for each individual.
[PDF File, 44 KB - formative_v9i1e67861_app1.pdf]

References

1. Institute of Medicine (US) Committee on Sleep Medicine and Research. Extent and health consequences of chronic sleep loss and sleep disorders. In: Colten HR, Altevogt BM, editors. Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem: National Academies Press (US); 2006. URL: https://www.ncbi.nlm.nih.gov/books/NBK19961/ [accessed 2024-10-07]
2. Medic G, Wille M, Hemels ME. Short- and long-term health consequences of sleep disruption. Nat Sci Sleep 2017 May 19;9:151-161. [doi: 10.2147/NSS.S134864] [Medline: 28579842]
3. Hanson JA, Huecker MR. Sleep Deprivation: StatPearls Publishing; 2024. URL: http://www.ncbi.nlm.nih.gov/books/NBK547676 [accessed 2024-10-07]

XSL•FO
RenderX

4. Chattu VK, Manzar MD, Kumary S, Burman D, Spence DW, Pandi-Perumal SR. The global problem of insufficient sleep and its serious public health implications. Healthcare (Basel) 2018 Dec 20;7(1):1. [doi: 10.3390/healthcare7010001] [Medline: 30577441]
5. Scott AJ, Webb TL, Martyn-St James M, Rowse G, Weich S. Improving sleep quality leads to better mental health: a meta-analysis of randomised controlled trials. Sleep Med Rev 2021 Dec;60:101556. [doi: 10.1016/j.smrv.2021.101556] [Medline: 34607184]
6. Telzer EH, Fuligni AJ, Lieberman MD, Galván A. The effects of poor quality sleep on brain function and risk taking in adolescence. Neuroimage 2013 May 1;71:275-283. [doi: 10.1016/j.neuroimage.2013.01.025] [Medline: 23376698]
7. Lim DC, Najafi A, Afifi L, et al. The need to promote sleep health in public health agendas across the globe. Lancet Public Health 2023 Oct;8(10):e820-e826. [doi: 10.1016/S2468-2667(23)00182-2] [Medline: 37777291]
8. Grandner MA. The cost of sleep lost: implications for health, performance, and the bottom line. Am J Health Promot 2018 Sep;32(7):1629-1634. [doi: 10.1177/0890117118790621a] [Medline: 30099900]
9. The Lancet Diabetes Endocrinology. Sleep: a neglected public health issue. Lancet Diabetes Endocrinol 2024 Jun;12(6):365. [doi: 10.1016/S2213-8587(24)00132-3] [Medline: 38740028]
10. Liu Y, Wheaton AG, Chapman DP, Cunningham TJ, Lu H, Croft JB. Prevalence of healthy sleep duration among adults--United States, 2014. MMWR Morb Mortal Wkly Rep 2016 Feb 19;65(6):137-141. [doi: 10.15585/mmwr.mm6506a1] [Medline: 26890214]
11. Centers for Disease Control and Prevention (CDC). Effect of short sleep duration on daily activities--United States, 2005-2008. MMWR Morb Mortal Wkly Rep 2011 Mar 4;60(8):239-242. [Medline: 21368739]
12. Grandner MA, Seixas A, Shetty S, Shenoy S. Sleep duration and diabetes risk: population trends and potential mechanisms. Curr Diab Rep 2016 Nov;16(11):106. [doi: 10.1007/s11892-016-0805-8] [Medline: 27664039]
13. Wang S, Rossheim ME, Nandy RR. Trends in prevalence of short sleep duration and trouble sleeping among US adults, 2005-2018. Sleep 2023 Jan 11;46(1):zsac231. [doi: 10.1093/sleep/zsac231] [Medline: 36152031]
14. Ford ES, Cunningham TJ, Croft JB. Trends in self-reported sleep duration among US Adults from 1985 to 2012. Sleep 2015 May 1;38(5):829-832. [doi: 10.5665/sleep.4684] [Medline: 25669182]
15. Fuller C, Lehman E, Hicks S, Novick MB. Bedtime use of technology and associated sleep problems in children. Glob Pediatr Health 2017 Oct 27;4:2333794X17736972. [doi: 10.1177/2333794X17736972] [Medline: 29119131]
16. Hysing M, Pallesen S, Stormark KM, Jakobsen R, Lundervold AJ, Sivertsen B. Sleep and use of electronic devices in adolescence: results from a large population-based study. BMJ Open 2015 Feb 2;5(1):e006748. [doi: 10.1136/bmjopen-2014-006748] [Medline: 25643702]
17. Bhattacharyya N. Sleep and health implications of snoring: a populational analysis. Laryngoscope 2015 Oct;125(10):2413-2416. [doi: 10.1002/lary.25346] [Medline: 25946644]
18. Pevernagie D, Aarts RM, De Meyer M. The acoustics of snoring. Sleep Med Rev 2010 Apr;14(2):131-144. [doi: 10.1016/j.smrv.2009.06.002] [Medline: 19665907]
19. Huang Z, Aarab G, Ravesloot MJL, de Vries N, Hilgevoord AAJ, Lobbezoo F. The annoying nature of snoring sounds is not only about intensity: a pilot study on exposed test subjects. J Clin Med 2023 Mar 31;12(7):2630. [doi: 10.3390/jcm12072630] [Medline: 37048713]
20. Knuiman M, James A, Divitini M, Bartholomew H. Longitudinal study of risk factors for habitual snoring in a general adult population: the Busselton Health Study. Chest 2006 Dec;130(6):1779-1783. [doi: 10.1378/chest.130.6.1779] [Medline: 17166996]
21. Virkkula P, Bachour A, Hytönen M, Malmberg H, Salmi T, Maasilta P. Patient- and bed partner-reported symptoms, smoking, and nasal resistance in sleep-disordered breathing. Chest 2005 Oct;128(4):2176-2182. [doi: 10.1378/chest.128.4.2176] [Medline: 16236871]
22. Hägg SA, Ilieva E, Ljunggren M, et al. The negative health effects of having a combination of snoring and insomnia. J Clin Sleep Med 2022 Apr 1;18(4):973-981. [doi: 10.5664/jcsm.9764] [Medline: 34753555]
23. Sowho M, Sgambati F, Guzman M, Schneider H, Schwartz A. Snoring: a source of noise pollution and sleep apnea predictor. Sleep 2020 Jun 15;43(6):zsz305. [doi: 10.1093/sleep/zsz305] [Medline: 31837267]
24. Shin H, Cho J. Unconstrained snoring detection using a smartphone during ordinary sleep. Biomed Eng Online 2014 Aug 15;13:116. [doi: 10.1186/1475-925X-13-116] [Medline: 25128409]
25. Chiang JK, Lin YC, Lin CW, Ting CS, Chiang YY, Kao YH. Validation of snoring detection using a smartphone app. Sleep Breath 2022 Mar;26(1):81-87. [doi: 10.1007/s11325-021-02359-3] [Medline: 33811634]
26. Klaus K, Stummer AL, Ruf S. Accuracy of a smartphone application measuring snoring in adults-how smart is it actually? Int J Environ Res Public Health 2021 Jul 8;18(14):7326. [doi: 10.3390/ijerph18147326] [Medline: 34299777]
27. Slowik JM, Sankari A, Collen JF. Obstructive Sleep Apnea: StatPearls Publishing; 2024. URL: http://www.ncbi.nlm.nih.gov/books/NBK459252 [accessed 2024-10-08]
28. Tregear S, Reston J, Schoelles K, Phillips B. Obstructive sleep apnea and risk of motor vehicle crash: systematic review and meta-analysis. J Clin Sleep Med 2009 Dec 15;5(6):573-581. [doi: 10.5664/jcsm.27662] [Medline: 20465027]
29. Harding SM. Complications and consequences of obstructive sleep apnea. Curr Opin Pulm Med 2000 Nov;6(6):485-489. [doi: 10.1097/00063198-200011000-00004] [Medline: 11100957]

XSL•FO
RenderX

30. Adult Obstructive Sleep Apnea Task Force of the American Academy of Sleep Medicine. Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults. J Clin Sleep Med 2009 Jun 15;05(3):263-276. [doi: 10.5664/jcsm.27497]
31. Chiang JK, Lin YC, Lu CM, Kao YH. Snoring index and neck circumference as predictors of adult obstructive sleep apnea. Healthcare (Basel) 2022 Dec 15;10(12):2543. [doi: 10.3390/healthcare10122543] [Medline: 36554066]
32. Chiang JK, Lin YC, Lu CM, Kao YH. Correlation between snoring sounds and obstructive sleep apnea in adults: a meta-regression analysis. Sleep Sci 2022;15(4):463-470. [doi: 10.5935/1984-0063.20220068] [Medline: 36419807]
33. Motamedi KK, McClary AC, Amedee RG. Obstructive sleep apnea: a growing problem. Ochsner J 2009;9(3):149-153. [Medline: 21603432]

Abbreviations

AHI: apnea-hypopnea index
dB(A): A-weighted decibels

Edited by A Mavragani; submitted 22.10.24; peer-reviewed by M Agbede, X Gao; revised version received 24.02.25; accepted 25.02.25; published 28.03.25.

Please cite as:
Brown J, Mitchell Z, Jiang YA, Archdeacon R
Accuracy of Smartphone-Mediated Snore Detection in a Simulated Real-World Setting: Algorithm Development and Validation
JMIR Form Res 2025;9:e67861
URL: https://formative.jmir.org/2025/1/e67861
doi:10.2196/67861

© Jeffrey Brown, Zachary Mitchell, Yu Albert Jiang, Ryan Archdeacon. Originally published in JMIR Formative Research (https://formative.jmir.org), 28.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile Health Tool to Capture Social Determinants of Health and Their Impact on HIV Treatment Outcomes Among People Who Use Drugs: Pilot Feasibility Study

Rachel E Gicquelais[1], PhD, MPH; Caitlin Conway[2], BSN; Olivia Vjorn[3], MS; Andrew Genz[4], MPH; Gregory Kirk[4,5], PhD, MD; Ryan Westergaard[1], MD, PhD, MPH

[1]School of Medicine and Public Health, University of Wisconsin-Madison, 603 WARF Office Building, 610 Walnut Street, Madison, WI, United States

[2]University of Wisconsin-Madison School of Nursing, Madison, WI, United States

[3]Center for Health Enhancement Systems Studies, University of Wisconsin-Madison College of Engineering, Madison, WI, United States

[4]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[5]Division of Infectious Diseases, Johns Hopkins University School of Medicine, Baltimore, MD, United States

Corresponding Author:
Rachel E Gicquelais, PhD, MPH
School of Medicine and Public Health, University of Wisconsin-Madison, 603 WARF Office Building, 610 Walnut Street, Madison, WI, United States

Abstract

Background: Active substance use, food or housing insecurity, and criminal legal system involvement can disrupt HIV care for people living with HIV and opioid use disorder (OUD). These social determinants of health are not routinely captured in clinical settings.

Objective: We evaluated whether real-time reports of social and behavioral factors using a smartphone app could predict viral nonsuppression and missed care visits to inform future mobile health interventions.

Methods: We enrolled 59 participants from the AIDS Linked to the Intravenous Experience (ALIVE) Study in Baltimore, Maryland, into a 12-month substudy between February 2017 and October 2018. Participants were eligible if they had OUD and had either a measured HIV RNA ≥1000 copies/mL or a ≥1-month lapse in antiretroviral therapy in the preceding 2 years. Participants received a smartphone and reported HIV medication adherence, drug use or injection, and several disruptive life events, including not having a place to sleep at night, skipping a meal due to lack of income, being stopped by police, being arrested, or experiencing violence on a weekly basis, through a survey on a mobile health app. We described weekly survey completion and investigated which factors were associated with viral nonsuppression (HIV RNA ≥200 copies/mL) or a missed care visit using logistic regression with generalized estimating equations adjusted for age, gender, smartphone comfort, and drug use.

Results: Participants were predominantly male (36/59, 61%), Black (53/59, 90%), and had a median of 53 years old. At baseline, 16% (6/38) were virally unsuppressed. Participants completed an average of 23.3 (SD 16.3) total surveys and reported missing a dose of antiretroviral therapy, using or injecting drugs, or experiencing any disruptive life events on an average of 13.1 (SD 9.8) weekly surveys over 1 year. Reporting use of any drugs (adjusted odds ratio [aOR] 2.3, 95% CI 1.4 - 3.7), injecting drugs (aOR 2.3, 95% CI 1.3 - 3.9), and noncompletion of all surveys (aOR 1.6, 95% CI 1.1 - 2.2) were associated with missing a scheduled care visit over the subsequent 30 days. Missing ≥2 antiretroviral medication doses within 1 week was associated with HIV viral nonsuppression (aOR 3.7, 95% CI: 1.2 - 11.1) in the subsequent 30 days.

Conclusions: Mobile health apps can capture risk factors that predict viral nonsuppression and missed care visits among people living with HIV who have OUD. Using mobile health tools to detect sociobehavioral factors that occur prior to treatment disengagement may facilitate early intervention by health care teams.

(JMIR Form Res 2025;9:e59953) doi:10.2196/59953

KEYWORDS

HIV; drug use; social determinants of health; mobile health; mHealth; smartphone

XSL•FO
RenderX

Introduction

In 2019, the US Department of Health and Human Services launched the Ending the HIV Epidemic initiative, establishing a goal to reduce new infections of HIV in the United States by 90% by the year 2030 [1]. Achieving this ambitious goal hinges on rapid linkage of newly diagnosed individuals to antiretroviral therapy (ART) and sustained HIV viral suppression to prevent transmission. Active injection drug use and use of other substances, such as alcohol, are associated with inconsistent adherence to ART and viral nonsuppression [2-5]. Accordingly, people who use drugs and are living with HIV are among the most vulnerable to disruptions in care that threaten sustained viral suppression.

The link between drug and alcohol use and viral nonsuppression often relates to unmet needs broadly defined as social determinants of health [6]. For example, housing instability among people who use drugs has been linked to an increased risk of both acquiring HIV and viral nonsuppression [6]. Injecting drugs raises the likelihood of unstable housing and hinders improvements in housing situations among people living with HIV [7,8], thereby threatening viral suppression. Further, inadequate access to food is highly prevalent among people who use drugs [9,10] and has been associated with viral nonsuppression [11], mortality [12], and engaging in behaviors that raise HIV transmission risk, such as receptive syringe sharing [13] and unprotected sexual intercourse [14]. The criminalization of drug use and social stigma of HIV further contribute to difficulties in engaging people who use drugs in both HIV and substance use disorder treatment [15] as well as to disruptions in care due to arrest or incarceration [16,17]. The link between social determinants of health and patients' capacity to engage and sustain in HIV care illustrates a need for innovative strategies that provide timely support to address the root causes of care disruption and viral nonsuppression.

Mobile health (mHealth) is defined broadly by the World Health Organization as "mobile and wireless technologies to support the achievement of health objectives" [18]. mHealth tools can provide a platform that facilitates real-time detection and the provision of timely assistance when events that could disrupt care engagement occur. Further, information captured through mHealth apps could detect early warning signs of care disengagement, such as missed ART doses or increases in drug use. These technologies may be particularly well-suited to augment existing patient-centered, comprehensive models of HIV care that are enabled by federal Ryan White funding in the United States [19], as these care models already feature medical case management, food pantries, support to obtain housing, and other services to address social determinants of health. In the present pilot feasibility study, we explored the potential of a smartphone app to detect disruptive life events and early warning signs of HIV care disengagement over a 1-year period among 59 people who use drugs living with HIV. Using a weekly survey within a smartphone app designed to support substance use disorder recovery, we assessed how often people who use drugs used the smartphone app and reported experiencing disruptive life events. Through subsequent linkage with electronic health record data, we assessed whether these disruptive life events were associated with 2 outcomes representing care disengagement: viral nonsuppression and missed care visits.

Methods

Participants, Recruitment, and Follow-Up

This manuscript reports on a pilot feasibility study of an mHealth app designed to detect disruptive events that could threaten sustained engagement in HIV care and examine the preliminary associations of these disruptive events with HIV care engagement. Participants for this study came from the AIDS Linked to the IntraVenous Experience (ALIVE) study, a prospective cohort study of people who inject drugs in Baltimore, Maryland, that has been described previously [20]. A total of 5506 individuals have been enrolled since the study began in 1988, and approximately 524 were living and had HIV at the time enrollment for the present substudy, which began in 2017 [21]. To enroll in the ALIVE study, participants must be aged 18 years or older and have a history of injection drug use. Participants are tested for HIV at enrollment. Follow-up is ongoing and involves twice-yearly visits to the ALIVE clinic where participants complete surveys about health history in the past 6 months and repeat HIV testing. They also provide consent for the researchers to access medical records to obtain health care utilization data.

The present analysis includes 59 ALIVE study participants living with HIV who were recruited through the HIV care clinic affiliated with Johns Hopkins Hospital in Baltimore, Maryland, United States. Medical records from ALIVE participants who had received HIV care at Johns Hopkins were reviewed to determine eligibility. If eligible, they were invited to meet with a study coordinator to complete a brief screening survey and enroll in the study. Recruitment flyers were also posted in the ALIVE study office and the Johns Hopkins clinic, inviting patients to self-refer for eligibility screening. The inclusion criteria were (1) ever had opioid use disorder (OUD), (2) used a nonprescribed opioid in the past year, (3) not treated with ART for ≥1 month in the past 2 years or had an HIV RNA viral load of ≥1000 copies/mL in the past 2 years, (4) a patient of the local HIV clinic from which participants were recruited, and (5) a participant of the ALIVE study.

All enrolled substudy participants received a smartphone with an unlimited voice and data plan and underwent a brief training with a study coordinator on using the study mHealth app. Participants completed an interviewer-administered baseline survey and follow-up surveys every 3 months for a 1-year period. The study team received HIV viral load testing results from the electronic health record at the local HIV clinic where recruitment was conducted and from routine semiannual testing performed as part of the ALIVE study follow-up. The date of all scheduled care visits at the local HIV clinic and whether visits were missed or attended was also collected. Enrollment and follow-up occurred during 2017 and 2018.

Ethical Considerations

This study was approved by the institutional review boards at the University of Wisconsin-Madison (protocol ID:

XSL•FO
RenderX

2016‐1190) and Johns Hopkins University (protocol ID: IRB00007523). Results were reported using the CONSORT (Consolidated Standards of Reporting Trials) guidelines for pilot studies. Written informed consent was obtained from all individual participants included in the study. All data collected were coded using a unique study identification number to ensure that individual participants could not be identified. Participants were compensated US $20 cash for completing a baseline survey and another US $20 for each completed follow-up survey. They also received a bonus of US $10 remuneration for each completed survey at the end of the 1-year follow-up period (up to US $50).

The A-CHESS mHealth App

The mHealth system used in this study, the Addiction-Comprehensive Health Enhancement Support System (A-CHESS), was designed to support recovery from alcohol use disorder and was found to reduce drinking in a prior study of people with alcohol use disorder [22]. A-CHESS has since been adapted for several other risk groups, including those living with OUD [23,24]. The theoretical framework supporting A-CHESS is the self-determination theory [25], which emphasizes that adaptive functioning to manage health conditions is improved by meeting 3 needs: competence (ie, the ability to cope with challenges), social relatedness (ie, having meaningful relationships and a feeling of connection), and autonomy (ie, a sense of control over behavior and goals). A-CHESS is therefore organized around these 3 pillars. Competence is supported by an extensive resource library containing podcasts, information, and other resources to support recovery. An in-app discussion board and private messaging function support relatedness by allowing app users to connect with one another and with case managers or providers who also have accounts on the app. Weekly self-monitoring surveys and a goal-setting feature within the app provide opportunities to build autonomy. Participants receive notifications to complete the self-monitoring survey on a weekly basis. For this study, the weekly self-monitoring survey was adapted to investigate potentially disruptive events that could threaten sustained engagement in HIV care, such as being arrested or not having a place to live, as well as events that may signal or raise risk of HIV care disengagement, such as missed ART doses and drug use (further described in Measures). Participants were invited to use A-CHESS on their study-provided cell phone for 12 months.

Measures

At the study enrollment visit, participants provided informed consent and completed a survey administered by study staff. They reported their HIV diagnosis date and whether they were currently using ART. Regarding drug use, participants reported whether they had used any illegal or street drugs (including marijuana) or used any medications not as prescribed in the past 30 days, what drugs they used, and whether they were currently taking any medications for OUD (including methadone, buprenorphine, or naltrexone). Participants also reported whether they had ever been diagnosed with any mental health conditions, including anxiety, depression, bipolar disorder, attention deficit hyperactivity disorder, obsessive-compulsive disorder, panic disorder, or post-traumatic stress disorder. Participants reported their comfort level using a smartphone on a numeric scale (1=not at all comfortable) to (5=very comfortable). Finally, participants reported sociodemographic characteristics, including age, gender, race, ethnicity, education, and employment.

The 2 primary outcomes of the study were abstracted through a review of electronic health records and ALIVE semiannual study visit data available for each participant during the 1-year follow-up period. These included viral nonsuppression, defined by the Centers for Disease Control and Prevention as an HIV viral load of ≥200 copies/mL (vs viral load of <200 copies/mL) [26] and missing a scheduled care visit, defined as missing any scheduled visit to the local HIV care clinic affiliated with the study (vs attending the visit). The HIV viral load at baseline was considered the closest HIV measurement to a participant's baseline survey within the period of 1 year before to 7 days after study enrollment. Electronic health records were also used to characterize baseline CD4+ count defined as the closest CD4 measurement to a participant's baseline survey within the 1 year before to 60 days after the baseline date. For the missed visit outcome, a total of 841 care visits were included in the analysis. These 841 visits include scheduled visits (regardless of visit attendance) with a medical provider, nurse, case manager, or mental health provider to reflect the breadth of care that patients living with HIV may receive through the United States Health Resources & Services Administration Ryan White HIV/AIDS Program. The primary outcome was whether a scheduled visit was missed versus attended. A total of 147 scheduled visits that were canceled before the appointment occurred were excluded from analyses.

The weekly self-monitoring survey included questions about the number of days ART was not taken, any drug use (defined by combining separate questions on use of heroin, any other opioids (including oxycodone, morphine, fentanyl, etc), cocaine, methamphetamine, or sedatives (including Valium, Ativan, or Xanax)), any injection drug use, and several potentially disruptive life events over the past 7 days. Potentially disruptive life events included not having a place to sleep, skipping a meal because of not having enough income, being stopped by police, being put in jail, and being robbed or beat up. Responses to questions on the weekly self-monitoring survey were used to form several binary indicator variables summarizing whether events of interest were reported on any weekly survey completed in the 30 days prior to having a viral load measurement and separately, in the 30 days prior to a scheduled care visit. Binary variables for noncompletion of all weekly surveys in the 30 days prior to each outcome and for reporting any significant event (ie, missed ART dose, any drug use, injecting drugs, or any disruptive life event) in the 30 days prior to each outcome were also created. These indicators were analyzed as the primary independent variables for the study. A-CHESS system use was summarized as the average number of days A-CHESS was used and the average number of weekly surveys completed per participant over the entire 1-year study period and per study month.

XSL•FO
RenderX

Analysis

Descriptive characteristics of the sample at baseline were calculated and compared between those who had at least 1 HIV viral load ≥200 copies/mL (vs those for whom all HIV test results across the 1-year study period indicated viral suppression). These characteristics were also compared for those who missed ≥50% of their care visits (vs <50%). Engagement with the A-CHESS smartphone app was described in the total sample and by the primary outcomes. To characterize whether reporting missed ART doses, any drug use, injecting drugs, or experiencing a disruptive life event were associated with viral nonsuppression or a missed care visit, logistic regression with generalized estimating equations accounting for repeated outcome measurements for each participant were used to estimate odds ratios with 95% CIs, as participants could have multiple HIV viral loads taken or care visits scheduled over the 1-year follow-up period. Unadjusted models and models adjusted for several confounders selected a priori were computed. Confounders included in adjusted models were sociodemographic characteristics (age, race, ethnicity, and gender), comfort using a smartphone (treated as a continuous variable with a range of 1-5), and whether the participant had used any drugs in the past 30 days, all of which were reported at study enrollment. In a sensitivity analysis, models were run with an alternate definition of HIV viral nonsuppression, ≥1000 copies/mL, as this matched one of the eligibility criteria for the study and was used to define virological failure in a prior, foundational study of HIV transmission [27].

Results

Participant Characteristics

A total of 59 participants enrolled in the 1-year study (Table 1). At baseline, participants were a mean age of 53 years and predominantly male (n=36, 61%), Black race (n=53, 90%), and non-Hispanic ethnicity (n=54, 92%). Approximately 52% (n=31) had less than a high school education, and only 5% (n=3) were employed at the time of the study. In the month before enrollment, 83% (n=49) had used any drugs, 71% (n=42) used opioids, 48% (n=28) used stimulants, and 27% (n=16) had injected drugs. While nearly all (n=55, 93%) reported currently taking ART, 16% (n=6) were virally nonsuppressed (defined as having a viral load of ≥1000 copies/mL at baseline), and 20% (n=10) had a CD4+ count <200 cells/mm^3. Participants had been living with HIV for a mean of 19 years prior to study enrollment, with 8% (n=5) having been diagnosed in the last 5 years.

XSL•FO
RenderX

Table . Descriptive characteristics at baseline among a sample of 59 people living with HIV and opioid use disorder enrolled in a pilot feasibility study of the A-CHESS (Addiction-Comprehensive Health Enhancement Support System) mobile health app during 2017 and 2018.

Characteristic	Total (n=59)	HIV viral suppression		Missed care visits	
		≥1 test ≥200 copies/mL (n=25)	All tests <200 copies/mL (n=27)	Missed ≥50% visits (n=23)	Missed <50% visits (n=33)
Age (years), median (IQR)	53 (11)	53 (9.0)	53 (12.5)	51 (10.0)	53 (11.1)
Female (vs male), n (%)	23 (39)	11 (44)	10 (37)	7 (30)	15 (46)
Black or African American race (vs another race), n (%)	53 (90)	25 (100)	24 (89)	21 (91)	29 (88)
Hispanic or Latinx (vs non-Hispanic), n (%)	5 (8)	0 (0)	4 (15)	2 (9)	3 (9)
Less than high school education (vs high school or more), n (%)	31 (52)	12 (48)	14 (52)	12 (52)	17 (52)
Employed (vs unemployed), n (%)	3 (5)	1 (4)	2 (7)	3 (13)	0 (0)
Comfort using a smartphone, mean (SD)	3.8 (1.3)	3.8 (1.4)	3.7 (1.3)	3.8 (1.3)	3.9 (1.3)
Any mental health concerns (vs none), n (%)	40 (68)	15 (60)	20 (74)	15 (65)	25 (76)
Used drugs in past 30 days (vs none), n (%)	49 (83)	19 (76)	25 (93)	20 (87)	27 (82)
Used opioids in past 30 days (vs none), n (%)	42 (71)	16 (64)	22 (82)	16 (70)	25 (76)
Used stimulants in past 30 days (vs none), n (%)	28 (48)	11 (44)	15 (56)	13 (56)	15 (46)
Injected drugs in past month (vs not), n (%)	16 (27)	6 (24)	7 (26)	7 (30)	8 (24)
Years living with HIV, mean (SD)	19.4 (8.7)	19.1 (7.5)	20.0 (9.2)	17.8 (7.2)	20.4 (9.1)
Currently taking antiretroviral therapy (vs not), n (%)	55 (93)	26 (96)	23 (92)	22 (96)	30 (91)
HIV RNA ≥1,000 copies/mL[a] (vs<1,000), n (%)	6 (16)	5 (29)	1 (6)	2 (14)	4 (19)
CD4+ count < 200cells/mm^3[b] (vs ≤200), n (%)	10 (20)	8 (36)	1 (4)	8 (40)	1 (4)

[a]37 participants had an available HIV viral load result at baseline.

[b]50 participants had an available CD4 count at baseline.

Of 52 participants who had at least 1 HIV viral load test during follow-up (comprising 237 total viral load tests), nearly half of the participants (n=25, 48%) had at least 1 detectable viral load of ≥200 copies/mL during the follow-up period. Regarding missed care visits, of 56 participants who had a visit scheduled during the follow-up period (comprising 841 total scheduled visits), 41% (n=23) of participants missed at least half of their scheduled care visits during the study period. The modest sample size limited our ability to test for differences in baseline characteristics between those who were virally nonsuppressed at least once over the follow-up period or who missed over half of their care visits, as shown in Table 1.

A-CHESS Engagement and Weekly Reports of Missed ART Doses, Drug Use, and Disruptive Life Events

Over the 1-year follow-up period, participants used A-CHESS an average of 119.4 (SD 82.8) total days and on an average of 11.5 (SD 4.4) days per month. The weekly survey was

XSL•FO
RenderX

completed an average of 23.3 (SD 16.3) times per participant across the follow-up period, leaving a total of 1353 weekly surveys available for analysis (Table 2). In total, 46.3% (627/1353) of weekly surveys included a report of missed ART doses, drug use, or a disruptive life event. Missing a dose of ART was reported on 25.2% (341/1353) of surveys, an average of 8.5 (SD 7.1) times per participant across the study period. Using any drugs was reported an average of 9.5 (SD 9.3) times per participant across the study period, and injecting drugs was reported 5.8 (SD 8.2) times per participant. Skipping a meal was reported an average of 6.6 (SD 6.5) times per participant. Not having a place to sleep was reported an average of 2.7 (SD 3.4) times per participant. Other disruptive life events (ie, being put in jail and being robbed or beaten up) were reported <2 times, on average, per participant.

Table . Missed antiretroviral therapy (ART) doses, drug use, and disruptive life events reported on weekly surveys completed by 59 participants living with HIV and opioid use disorder enrolled in a pilot feasibility study of the A-CHESS (Addiction-Comprehensive Health Enhancement Support System) mobile health app during 2017 and 2018.

Survey item	Number of weekly surveys, n (%)	Number of weekly surveys completed per participant, mean (SD)
Total number of weekly surveys completed	1353 (100)	23.3 (16.3)
Reported any missed ART, drug use, or disruptive life event	627 (46.3)	13.1 (9.9)
1 or more missed ART doses	341 (25.2)	8.5 (7.1)
2 or more missed ART doses	190 (14)	5.9 (6.1)
Any drug use	417 (30.8)	9.5 (9.3)
Injected drugs	116 (8.6)	5.8 (8.2)
Did not have a place to sleep at night	27 (2)	2.7 (3.4)
Skipped a meal because I didn't have enough money	131 (9.7)	6.6 (6.5)
Was stopped by police	22 (1.6)	1.8 (1.2)
Was put in jail	9 (0.7)	1.5 (0.6)
Was robbed or beat up	13 (1)	1.6 (0.9)

Table S1 in Multimedia Appendix 1 shows differences in A-CHESS utilization frequency and weekly survey completion during follow-up between participants who were virally nonsuppressed at least once over the follow-up period (vs suppressed throughout follow-up) as well as between those who missed 50% or more of their scheduled care visits (vs who missed fewer than 50% of their care visits). These comparisons indicate that more weekly surveys were completed by those who did not have any viral loads ≥200 copies/mL (P=.01) and who did not miss any care visits (P=.005). Those who did not miss any care visits during follow-up also tended to use A-CHESS more often (P=.03).

Association of Missed ART, Drug Use, and Disruptive Life Events Reported Through A-CHESS With Missed Care Visits and Viral Nonsuppression

In bivariable and adjusted analyses, injecting drugs, any drug use, and not completing any weekly surveys in the prior month were associated with missing a scheduled care visit in the subsequent 30 days (Table 3). Specifically, injecting drugs was associated with 2.3-fold higher adjusted odds (95% CI 1.3 - 3.9) and using any drugs was associated at a similar magnitude with missing a scheduled care visit after adjustment for age, comfort using a smartphone, gender, and any drug use at baseline. Not completing the weekly survey (vs completing at least 1 weekly survey) was associated with 57% higher odds of missing a care visit in the following 30 days (adjusted odds ratio [aOR] 1.6, 95% CI 1.1 - 2.2).

XSL•FO
RenderX

Table . Association of reporting missed antiretroviral therapy (ART) doses, drug use, and disruptive life events in the prior 30 days with missing a care visit among 59 participants living with HIV and opioid use disorder enrolled in a pilot feasibility study of the A-CHESS (Addiction-Comprehensive Health Enhancement Support System) mobile health app during 2017 and 2018.

Events reported in the month prior to scheduled care visit[a]	Unadjusted OR[b] (95% CI)	Adjusted OR (95% CI)[c]
Any missed ART, drug use, or disruptive life event	1.20 (0.80 - 1.81)	1.25 (0.81 - 1.93)
1 or more missed ART doses	0.91 (0.61 - 1.35)	0.89 (0.57 - 1.39)
2 or more missed ART doses	1.22 (0.70 - 2.15)	1.30 (0.69 - 2.43)
Any drug use	2.07 (1.28 - 3.35)	2.27 (1.41 - 3.66)
Injected drugs	2.11 (1.24 - 3.59)	2.26 (1.30 - 3.91)
Did not have a place to sleep at night	1.87 (0.38 - 9.27)	1.76 (0.28 - 11.14)
Skipped a meal due to not enough money	0.76 (0.36 - 1.60)	0.81 (0.37 - 1.74)
Stopped by police	1.07 (0.45 - 2.54)	1.11 (0.42 - 2.97)
No weekly surveys completed in past 30 days	1.42 (1.00 - 2.02)	1.57 (1.10 - 2.22)

[a]Odds ratios are reported relative to a referent group that completed at least 1 weekly survey in the 30 days prior to a scheduled care visit but did not indicate experiencing the disruptive life event.

[b]OR: odds ratio.

[c]Adjusted for age, gender, comfort using a smartphone (score: 1=not at all comfortable to 5=very comfortable), and any drug use in the past 30 days, all of which were reported at study enrollment.

In both unadjusted and adjusted analyses, missing 2 or more doses of ART was associated with having a nonsuppressed viral load in the subsequent 30 days (aOR 3.7, 95% CI 1.2 - 11.1; Table 4). These associations were similar in a sensitivity analysis in which the outcome was defined as a viral load of load ≥1000 (Table S2 in Multimedia Appendix 1). Notably, we were unable to adjust for race or ethnicity given the predominantly Black and non-Hispanic racial or ethnic makeup of the sample (90% Black and 8% Hispanic). Further, because being put in jail and robbed or beaten up were each reported on only a few surveys (<15 surveys each), we were unable to analyze their relationship with the primary outcomes of interest.

Table . Association of reporting missed antiretroviral therapy (ART) doses, drug use, and disruptive life events in the prior 30 days with HIV viral nonsuppression (viral load ≥200 copies/mL) among 59 participants living with HIV and opioid use disorder enrolled in a pilot feasibility study of the A-CHESS (Addiction-Comprehensive Health Enhancement Support System) mobile health app during 2017 and 2018.

Events reported in the month prior to viral load measurement[a]	Unadjusted OR[b] (95% CI)	Adjusted OR (95% CI)[c]
Any missed ART, drug use, or disruptive life event	1.30 (0.72 - 2.33)	1.28 (0.68 - 2.44)
1 or more missed ART doses	1.96 (0.88 - 4.35)	1.92 (0.76 - 4.76)
2 or more missed ART doses	4.00 (1.61 - 10.00)	3.70 (1.23 - 11.11)
Any drug use	0.93 (0.46 - 1.85)	0.92 (0.44 - 1.92)
Injected drugs	0.79 (0.22 - 2.86)	0.74 (0.19 - 2.78)
Did not have a place to sleep at night	2.38 (0.56 - 10.00)	2.44 (0.54 - 11.11)
Skipped a meal due to not enough money	0.64 (0.25 - 1.67)	0.6 (0.22 - 1.64)
Stopped by police	1.72 (0.34 - 8.33)	1.67 (0.32 - 9.09)
No weekly surveys completed in past 30 days	0.94 (0.53 - 1.69)	0.87 (0.46 - 1.64)

[a]Odds ratios are reported relative to a referent group that completed at least 1 weekly survey in the 30 days prior to a viral load test but did not indicate experiencing the disruptive life event.

[b]OR: odds ratio.

[c]Adjusted for age, gender, comfort using a smartphone (score: 1=not at all comfortable to 5=very comfortable), and any drug use in the past 30 days, all of which were reported at study enrollment.

Discussion

This study is a proof-of-concept demonstrating the utility and feasibility of using an mHealth system to capture time-varying threats to HIV viral suppression and care engagement among people living with HIV and OUD. Our system focused on detecting early warning signs of HIV care disengagement and disruptive events that fall within the realm of social determinants

XSL•FO
RenderX

of health known to impact the health of people who use drugs [6]. Over a 1-year period, we captured >600 reports of events that could be disruptive to HIV care and be early warning signs of care disengagement. Using or injecting drugs and reporting missed doses of ART were associated with an elevated likelihood to miss a care visit or be virally nonsuppressed during follow-up, respectively. Not completing any of the weekly check-in surveys was also associated with missing a care visit. While our study lacked the statistical power needed to assess whether disruptive life events reported less frequently, including housing insecurity, experiencing violence, and being arrested, were associated with poor outcomes, our results support the need to conduct larger studies of time-varying threats to care engagement in the future. An mHealth system like ours could be used in HIV care settings to assist case managers and social workers in proactively detecting and addressing these threats.

Our sample of people who use drugs answered an average of 23 of 52 total weekly surveys available to them during the 1-year follow-up period, a response rate of 44%. Despite that participants were not remunerated for completing weekly surveys, as they were operationalized as a self-assessment tool to enhance self-determined motivation as part of the A-CHESS app, this response rate is similar to several studies among people living with HIV [28]. Nonetheless, the utility of the weekly survey as an informative tool for clinical action may require a higher response rate. Thus, future studies may consider strategies such as incentivizing weekly surveys or sending surveys at a time most likely to get a response [29].

At the same time, our findings suggest that missing up to 4 consecutive surveys may be a helpful threshold to prompt a case manager or social worker to proactively reach out to a patient about any difficulties they may be experiencing and ensure any needs are met through appropriate referrals or direct provision of support. This conclusion is supported by considering multiple of our findings together, including our modest response rate, our finding that not answering at least 1 survey over a 4-week period was associated with missing a care visit, and our finding that those who did not experience viral nonsuppression or missed visits tended to engage with the app more frequently. A focus on monthly, rather than weekly, engagement with A-CHESS may also help to reduce alert fatigue, a phenomenon wherein people using technology become desensitized after receiving many alerts or notifications [30-33] and balance the extra staff time required to access A-CHESS and follow up with disengaged participants. Indeed, another ongoing study of A-CHESS in the substance use treatment setting suggests that staff time is a major barrier to using a clinician dashboard feature added to A-CHESS [34], which provides an easy-to-interpret interface to monitor patient survey results and app engagement. Prior studies of A-CHESS also suggest that providing dedicated support to implement A-CHESS within a health care organization using strategies such as coaching and process improvement models to ensure clinicians have dedicated time, technical assistance, and resources to integrate this new clinical tool into existing workflows [35,36]. Our finding that using and injecting drugs were associated with subsequently missing care visits further suggests that understanding the context of drug use in weekly surveys may assist with identifying supports available to app users at the time when drug use or craving occurs [37].

The need for novel and effective mHealth interventions is underscored by findings from several demonstration projects that were part of the Health Resources and Services Administration's Special Projects of National Significance focused on social media and mHealth for supporting retention in HIV care [38]. This work identified 3 promising foci for mHealth tools, including (1) managing HIV care; (2) fostering feelings of support and personal connectedness; and (3) helping alleviate negative feelings about status and mitigating HIV-related stigma [39]. A-CHESS can meet all 3 of these needs. Through its theoretical foundations in self-determination theory [25], A-CHESS facilitates patient-provider connection and support through messaging functions and a clinician dashboard that can highlight weekly self-monitoring survey results. Peer connections are augmented by the use of a discussion board, targeted resources, and other key features.

Limitations

The modest sample size of the study limits its internal and external validity and our ability to adjust for a broad set of confounders. Thus, we presented both unadjusted and adjusted results that could account for key confounders we identified a priori, including sociodemographic characteristics, comfort level with using a smartphone, and recent drug use at baseline. The overall conclusions drawn from the results were consistent in adjusted and unadjusted models. Additionally, several of the disruptive life events of interest (ie, housing instability, food insecurity, and law enforcement interaction) were reported infrequently over the follow-up period, which did not allow us to examine their association with missing a care visit or viral nonsuppression. As these are known contributors to disengagement with HIV care, they will be important to study in future, larger studies. Our results suggest that, when experienced, our smartphone app platform was able to capture at least some of these events on a weekly basis. Future research should also use qualitative methods to assess user experience, acceptability, and desired supports when disruptive life events occur. Finally, we were unable to include canceled care visits, and it was impossible to determine if or when the visit was rescheduled. Further analysis of patients who cancel a visit and do not return is warranted.

Conclusions

A-CHESS is a promising tool for capturing time-varying threats to engagement in HIV care among people who use drugs. Future research should examine whether receiving tailored support based on disruptive life events reported in A-CHESS is effective in reducing viral nonsuppression and HIV care disengagement and evaluate how to integrate A-CHESS within the HIV medical home setting.

XSL•FO
RenderX

Acknowledgments

The authors acknowledge the National Institute on Drug Abuse (NIDA) Avenir Award DP2DA042424 and NIDA U01DA036297.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the small sample size but are available from the study principal investigator, RW, on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental tables.
[DOCX File, 358 KB - formative_v9i1e59953_app1.docx]

Checklist 1
CONSORT (Consolidated Standards of Reporting Trials) checklist.
[DOCX File, 3600 KB - formative_v9i1e59953_app2.docx]

References

1. Ending the HIV epidemic HIV.gov 2022. US Department of Health and Human Services. 2022. URL: https://www.hiv.gov/federal-response/ending-the-hiv-epidemic/overview [accessed 2023-04-26]
2. Jones TPW, Lampe FC, Arenas-Pinto A, et al. Alcohol, smoking, recreational drug use and association with virological outcomes among people living with HIV: cross-sectional and longitudinal analyses. HIV Med 2022 Mar;23(3):209-226. [doi: 10.1111/hiv.13156] [Medline: 34634176]
3. Ladak F, Socias E, Nolan S, et al. Substance use patterns and HIV-1 RNA viral load rebound among HIV-positive illicit drug users in a Canadian setting. Antivir Ther 2019;24(1):19-25. [doi: 10.3851/IMP3265] [Medline: 30230474]
4. Langebeek N, Gisolf EH, Reiss P, et al. Predictors and correlates of adherence to combination antiretroviral therapy (ART) for chronic HIV infection: a meta-analysis. BMC Med 2014 Aug 21;12:142. [doi: 10.1186/PREACCEPT-1453408941291432] [Medline: 25145556]
5. Westergaard RP, Hess T, Astemborski J, Mehta SH, Kirk GD. Longitudinal changes in engagement in care and viral suppression for HIV-infected injection drug users. AIDS 2013 Oct 23;27(16):2559-2566. [doi: 10.1097/QAD.0b013e328363bff2] [Medline: 23770493]
6. Galea S, Vlahov D. Social determinants and the health of drug users: socioeconomic status, homelessness, and incarceration. Public Health Rep 2002;117(Suppl 1):S135-S145. [Medline: 12435837]
7. Kidder DP, Wolitski RJ, Campsmith ML, Nakamura GV. Health status, health care use, medication use, and medication adherence among homeless and housed people living with HIV/AIDS. Am J Public Health 2007 Dec;97(12):2238-2245. [doi: 10.2105/AJPH.2006.090209] [Medline: 17971562]
8. Marcus R, de Groot A, Bachman S, et al. Longitudinal determinants of housing stability among people living with HIV/AIDS experiencing homelessness. Am J Public Health 2018 Dec;108(S7):S552-S560. [doi: 10.2105/AJPH.2018.304772] [Medline: 32941778]
9. Strike C, Rudzinski K, Patterson J, Millson M. Frequent food insecurity among injection drug users: correlates and concerns. BMC Public Health 2012 Dec 8;12(1058):1058. [doi: 10.1186/1471-2458-12-1058] [Medline: 23216869]
10. Schmitz J, Kral AH, Chu D, Wenger LD, Bluthenthal RN. Food insecurity among people who inject drugs in Los Angeles and San Francisco. Public Health Nutr 2016 Aug;19(12):2204-2212. [doi: 10.1017/S1368980016000306] [Medline: 26956477]
11. Singer AW, Weiser SD, McCoy SI. Does food insecurity undermine adherence to antiretroviral therapy? A systematic review. AIDS Behav 2015 Aug;19(8):1510-1526. [doi: 10.1007/s10461-014-0873-1] [Medline: 25096896]
12. Anema A, Chan K, Chen Y, Weiser S, Montaner JSG, Hogg RS. Relationship between food insecurity and mortality among HIV-positive injection drug users receiving antiretroviral therapy in British Columbia, Canada. PLoS ONE 2013;8(5):e61277. [doi: 10.1371/journal.pone.0061277] [Medline: 23723968]
13. Rouhani S, Allen ST, Whaley S, et al. Food access among people who inject drugs in West Virginia. Harm Reduct J 2021 Aug 21;18(1):90. [doi: 10.1186/s12954-021-00536-x] [Medline: 34419045]
14. Shannon K, Kerr T, Milloy MJ, et al. Severe food insecurity is associated with elevated unprotected sex among HIV-seropositive injection drug users independent of HAART use. AIDS 2011 Oct 23;25(16):2037-2042. [doi: 10.1097/QAD.0b013e32834b35c9] [Medline: 21811140]
15. Wood E, Kerr T, Tyndall MW, Montaner JSG. A review of barriers and facilitators of HIV treatment among injection drug users. AIDS 2008 Jul 11;22(11):1247-1256. [doi: 10.1097/QAD.0b013e3282fbd1ed] [Medline: 18580603]

XSL•FO
RenderX

16. Westergaard RP, Kirk GD, Richesson DR, Galai N, Mehta SH. Incarceration predicts virologic failure for HIV-infected injection drug users receiving antiretroviral therapy. Clin Infect Dis 2011 Oct;53(7):725-731. [doi: 10.1093/cid/cir491] [Medline: 21890777]
17. Westergaard RP, Hochstatter KR, Andrews PN, et al. Effect of patient navigation on transitions of HIV care after release from prison: a retrospective cohort study. AIDS Behav 2019 Sep;23(9):2549-2557. [doi: 10.1007/s10461-019-02437-4] [Medline: 30790170]
18. mHealth: New Horizons for Health through Mobile Technologies: World Health Organization; 2011.
19. Beane SN, Culyba RJ, DeMayo M, Armstrong W. Exploring the medical home in Ryan White HIV care settings: a pilot study. J Assoc Nurses AIDS Care 2014;25(3):191-202. [doi: 10.1016/j.jana.2013.10.007] [Medline: 24560357]
20. Vlahov D, Anthony JC, Muñoz A, et al. The ALIVE Study: a longitudinal study of HIV-1 infection in intravenous drug users: description of methods. J Drug Issues 1991 Oct;21(4):759-776. [doi: 10.1177/002204269102100406]
21. Sun J, Mehta SH, Astemborski J, et al. Mortality among people who inject drugs: a prospective cohort followed over three decades in Baltimore, MD, USA. Addiction 2022 Mar;117(3):646-655. [doi: 10.1111/add.15659] [Medline: 34338374]
22. Gustafson DH, McTavish FM, Chih MY, et al. A smartphone application to support recovery from alcoholism: A randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572. [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
23. Gustafson DH Sr, Landucci G, McTavish F, et al. The effect of bundling medication-assisted treatment for opioid addiction with mHealth: study protocol for a randomized clinical trial. Trials 2016 Dec 12;17(1):592. [doi: 10.1186/s13063-016-1726-1] [Medline: 27955689]
24. Gustafson DH, Landucci G, Vjorn OJ, et al. Effects of bundling medication for opioid use disorder with an mHealth intervention targeting addiction: a randomized clinical trial. Am J Psychiatry 2024 Feb 1;181(2):115-124. [doi: 10.1176/appi.ajp.20230055] [Medline: 37789744]
25. Deci EL, Ryan RM. Intrinsic Motivation and Self-Determination in Human Behavior: Plenum Press; 1985.
26. HIV treatment as prevention? HIV.gov. 2023. URL: https://www.hiv.gov/tasp [accessed 2025-03-18]
27. Cohen MS, Chen YQ, McCauley M, et al. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med 2011 Aug 11;365(6):493-505. [doi: 10.1056/NEJMoa1105243] [Medline: 21767103]
28. Smiley SL, Milburn NG, Nyhan K, Taggart T. A systematic review of recent methodological approaches for using ecological momentary assessment to examine outcomes in U.S. based HIV research. Curr HIV/AIDS Rep 2020 Aug;17(4):333-342. [doi: 10.1007/s11904-020-00507-0] [Medline: 32594365]
29. Aminikhanghahi S, Schmitter-Edgecombe M, Cook DJ. Context-aware delivery of ecological momentary assessment. IEEE J Biomed Health Inform 2020 Apr;24(4):1206-1214. [doi: 10.1109/JBHI.2019.2937116] [Medline: 31443058]
30. Ash JS, Sittig DF, Campbell EM, Guappone KP, Dykstra RH. Some unintended consequences of clinical decision support systems. AMIA Annu Symp Proc 2007 Oct 11;2007(26–30):26-30. [Medline: 18693791]
31. Baseman JG, Revere D, Painter I, Toyoji M, Thiede H, Duchin J. Public health communications and alert fatigue. BMC Health Serv Res 2013 Aug 5;13(1):295. [doi: 10.1186/1472-6963-13-295] [Medline: 23915324]
32. Gephart S, Carrington JM, Finley B. A systematic review of nurses' experiences with unintended consequences when using the electronic health record. Nurs Adm Q 2015;39(4):345-356. [doi: 10.1097/NAQ.0000000000000119] [Medline: 26340247]
33. Muench F, Baumel A. More than a text message: dismantling digital triggers to curate behavior change in patient-centered health interventions. J Med Internet Res 2017 May 26;19(5):e147. [doi: 10.2196/jmir.7463] [Medline: 28550001]
34. Fleddermann K, Molfenter T, Jacobson N, et al. Clinician perspectives on barriers and facilitators to implementing e-health technology in substance use disorder (SUD) treatment facilities. Subst Abuse 2021;15:11782218211053360. [doi: 10.1177/11782218211053360] [Medline: 34720585]
35. Gustafson D Sr, Horst J, Boss D, et al. Implementation of smartphone systems to improve quality of life for people with substance use disorder: interim report on a randomized controlled trial. JMIR Hum Factors 2022 Jul 14;9(3):e35125. [doi: 10.2196/35125] [Medline: 35834315]
36. Quanbeck A, Gustafson DH, Marsch LA, et al. Implementing a mobile health system to integrate the treatment of addiction into primary care: a hybrid implementation-effectiveness study. J Med Internet Res 2018 Jan 30;20(1):e37. [doi: 10.2196/jmir.8928] [Medline: 29382624]
37. Linas BS, Latkin C, Westergaard RP, et al. Capturing illicit drug use where and when it happens: an ecological momentary assessment of the social, physical and activity environment of using versus craving illicit drugs. Addiction 2015 Feb;110(2):315-325. [doi: 10.1111/add.12768] [Medline: 25311241]
38. Medich M, Swendeman DT, Comulada WS, et al. Promising approaches for engaging youth and young adults living with HIV in HIV primary care using social media and mobile technology interventions: protocol for the SPNS social media initiative. JMIR Res Protoc 2019 Jan 31;8(1):e10681. [doi: 10.2196/10681] [Medline: 30702434]
39. Brooks RA, Nieto O, Swendeman D, et al. Qualitative evaluation of social media and mobile technology interventions designed to improve HIV health outcomes for youth and young adults living with HIV: a HRSA SPNS initiative. Health Promot Pract 2020 Sep;21(5):693-704. [doi: 10.1177/1524839920938704] [Medline: 32757839]

XSL•FO
RenderX

Abbreviations

A-CHESS: Addiction Comprehensive Health Enhancement Support System
ALIVE: AIDS Linked to the Intravenous Experience
aOR: adjusted odds ratio
ART: antiretroviral therapy
CONSORT: Consolidated Standards of Reporting Trials
mHealth: mobile health
OUD: opioid use disorder

Edited by A Mavragani; submitted 26.04.24; peer-reviewed by S Ramsey, T Finlayson; revised version received 28.02.25; accepted 02.03.25; published 26.03.25.

Please cite as:
Gicquelais RE, Conway C, Vjorn O, Genz A, Kirk G, Westergaard R
Mobile Health Tool to Capture Social Determinants of Health and Their Impact on HIV Treatment Outcomes Among People Who Use Drugs: Pilot Feasibility Study
JMIR Form Res 2025;9:e59953
URL: https://formative.jmir.org/2025/1/e59953
doi:10.2196/59953

© Rachel E Gicquelais, Caitlin Conway, Olivia Vjorn, Andrew Genz, Gregory Kirk, Ryan Westergaard. Originally published in JMIR Formative Research (https://formative.jmir.org), 26.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile App-Based Interactive Care Plan for Migraine: Survey Study of Usability and Improvement Opportunities

Nathan P Young[1], DO; Jennifer I Stern[1], MD; Stephanie J Steel[1], MD; Jon O Ebbert[2], MD

[1]Department of Neurology, Mayo Clinic, 200 First St. SW Eighth Floor, Rochester, MN, United States

[2]Department of Internal Medicine, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Nathan P Young, DO
Department of Neurology, Mayo Clinic, 200 First St. SW Eighth Floor, Rochester, MN, United States

Abstract

Background: We implemented a novel mobile app-based Migraine Interactive Care Plan (MICP) integrated with our electronic health records (EHRs). The MICP facilitates remote assessment of adult patients with migraine, educational content delivery, and care team communication. Feasibility of the MICP was demonstrated in a pilot implementation study.

Objective: We aimed to assess the preferences and satisfaction of patients with migraine users of a mobile app-based care plan integrated with the EHR.

Methods: An electronic survey was administered to a single cohort of MICP users between December 6, 2021, and December 30, 2021. The survey assessed patient preferences for which data to track, frequency of tracking, and satisfaction with the MICP. Survey responses were compared between subsets determined by patient-reported headache frequency and treatment with and without botulinum toxin and calcitonin gene-related peptide (CGRP) antagonist therapy. The Wilcoxon rank-sum test was used for continuous variables and the χ^2 test or Fisher exact test for categorical variables.

Results: The total sample size was 184 and the survey response rate was 30.4% (56/184). No significant differences in age (P=.26) or sex (P=.19) between respondents and nonrespondents were observed. Respondent median age was 42 (range 20 - 72) years and 94.6% (53/56) were female. Headache frequency was (1) 0 to 8 days (26/56, 46.4%), (2) 9 to 14 days (12/56, 21.4%), and (3) 15 or more days (18/56, 32.1%). No difference was observed in any survey responses based on headache frequency or treatment. The majority of respondents preferred to track headache days weekly (30/56, 53.6%) or daily (15/56, 26.8%) and preferred to change the frequency of headache tracking reminders (42/56, 75%). Respondents were somewhat or very interested in daily tracking personal observations in free text (41/52, 78.8%), medication treatment (43/52, 82.7%) and treatment response (39/56, 69.6%), class of medication treatment (36/52, 69.2%), severity of functional impairment (39/56, 69.6%), type of functional impairment (35/53, 66%), headache day (40/54, 74.1%), and headache pain level on a scale of 1 to 10 (38/53, 71.7%). Respondents agreed or strongly agreed that the education content was useful (31/51, 60.8%) but lacked personalization (25/51, 49%). Most respondents agreed or strongly agreed that they were satisfied with the MICP (38/50, 76%) and that it helped them communicate with their care team (38/53, 71.7%).

Conclusions: Most MICP users were motivated to track headache frequency, medication treatment with response, functional impairment, and pain intensity. Opportunities to improve the MICP include (1) allowing patients to change the frequency of assessments and notifications; (2) recording personal observations or comments through free text, which may include headache triggers; (3) assessment of headache severity using a 1 to 10 pain scale; and (4) tailoring headache education based on frequency and severity (episodic vs chronic migraine). These observations may be useful to improve the usability of the MICP and similar EHR-integrated migraine care platforms that others may develop.

(JMIR Form Res 2025;9:e66763) doi:10.2196/66763

KEYWORDS

migraine; remote monitoring; mobile app; mHealth; patient-reported outcomes; care plan; digital health; app; smartphone; eHealth; technology; survey study; headache; electronic health record; remote assessment; older adult; adult; electronic survey; pain; mobile phone; telehealth; telemedicine

XSL•FO
RenderX

Introduction

Background

Patients with migraine often face significant barriers in access to evidence-based care, with approximately one-quarter of those with episodic migraine and less than 5% of those with chronic migraine receiving appropriate treatment [1-4]. Access to care is especially challenging for women, racial and ethnic minority groups, and patients living in rural areas [5]. Novel health care delivery models may improve care access when resources are limited and projected to worsen [2,6]. Studies have demonstrated that telemedicine improves access to migraine care, and that it is associated with favorable outcomes including patient and provider satisfaction [7,8]. Internet or smartphone app-based remote assessment and monitoring of migraine [9,10] may increase the efficiency of care delivery and facilitate telemedicine, electronic [11], and face-to-face visits while delivering migraine educational content. Headache clinicians are comfortable treating patients through telemedicine and most have indicated a high level of interest in prescribing migraine apps [12] while citing the importance of integrating remote monitoring data into the electronic health record (EHR) [13]. Smartphone-based apps may also improve diagnosis [14], deliver educational content [15], guide behavioral treatment [16-18], and advance clinical trials [19]. There are improvement opportunities with respect to accessibility of apps for headache [20].

Previous Work

We implemented a novel smartphone app-based Migraine Interactive Care Plan (MICP) integrated with the EHR (Epic, Epic Systems) in our community neurology practice, and potentially available to other institutions using Epic EHR. Development and testing of the MICP has been previously published [15]. Feasibility of the MICP was demonstrated in a pilot implementation study, but usability did not reach a predetermined threshold of 75% of users completing at least one electronically assigned task 127/171 (74.3%) [15].

Access to the MICP was ordered by a provider at the time of a face-to-face neurology consult and integrated into the existing EHR app (ie, Epic MyChart platform). This integration allowed patient access to their medical record, care team messaging, and appointment management. The MICP was designed to facilitate remote monitoring of adult patients with migraine, deliver educational content, and streamline care delivery. Key features of the MICP included (1) weekly assessments of headache days, treatment days, reduced function, and absenteeism; (2) monthly assessments of patient satisfaction (Likert scale) with their migraine treatment plan; (3) Migraine Disability Assessment (MIDAS) survey [21] assessment every 3 months; (4) scheduled delivery of migraine education content; and (5) weekly assessment of medication compliance and concerns and electronic messaging with the care team if needed.

Goal of This Study

In this study, we aimed to identify opportunities to improve MICP usability by surveying MICP users to assess their satisfaction and preferences and whether they differ based on migraine frequency or type of preventive treatment.

Methods

Ethical Considerations

This quality improvement project was granted an exemption waiver for written consent from the Mayo Clinic (IRB 20 - 000606). All respondents electronically signed the Health Insurance Portability and Accountability Act (HIPAA) agreement before starting the survey. No compensation was offered to complete the survey. All survey responses were deidentified before analysis, and the data were on a secure server.

Overview

A novel electronic survey was developed and formatted (on Qualtrics). Between December 6, 2021, and December 30, 2021, all MICP users (n=184) who completed MICP enrollment and completed at least one task received an email invitation to a secure web-based survey with a personalized letter from a physician on the migraine care team (NPY) explaining the goals of the survey and encouragement to voluntarily participate. The survey cohort had all completed a neurological headache consultation within a general community neurology practice before enrollment in the MICP. Weekly email reminders were sent by a study coordinator to nonrespondents for a total of 4 weeks. Only 1 survey completion was allowed. Demographic variables, including age and sex, were retrieved from the EHR for respondents and nonrespondents and were compared using the *t* test. Categorical variables were summarized. The overall survey response rate was defined by the percentage of surveys returned. The analysis of the results from each individual question included only the completed response to the question and nonresponses to individual questions were excluded. Survey responses were compared between subsets determined by patient reported headache frequency (0 to 8 days, 9 to 14 days, or 15 or more days), survey self-reported treatment with and without botulinum toxin and calcitonin gene-related peptide (CGRP) antagonist therapy. The Wilcoxon rank-sum test was used for continuous variables and χ^2 test or Fisher exact test for categorical variables using a 2-tailed α level of 5%. Missing survey response data were excluded from analysis. Analysis was performed using SAS software (version 9.4; SAS Inc).

Results

The survey response rate was 30.4% (56/184). No significant differences in age (*P*=.26) or sex (*P*=.19) between respondents and nonrespondents were observed. Table 1 summarizes the survey responses of all respondents, and a comprehensive table including the subgroup analysis and *P* values are summarized in Multimedia Appendix 1 and Multimedia Appendix 2.

XSL•FO
RenderX

Table . Survey responses of migraine users of the Migraine Interactive Care Plan.

Survey question	Total respondents (N=56), n (%)
How often would you prefer to track your headache days on a smartphone app?	
Daily	15 (26.8)
Weekly	30 (53.6)
Monthly	8 (14.3)
Every 3 months	3 (5.4)
How often would you like to be reminded to record a headache day on a smartphone app?	
Daily	14 (25)
Weekly	33 (58.9)
Monthly	7 (12.5)
Every 3 months	2 (3.6)
Would you prefer to control and change the frequency of headache tracking reminders?	
Yes	42 (75)
No	14 (25)
In addition to headache days, please tell us what else you are interested in tracking with your headache days?	
Migraine/headache triggers	41 (73.2)
Stress level	38 (67.8)
Sleep	37 (66.1)
Step count	9 (16.1)
Heart rate	4 (7.1)
Diet	16 (28.6)
Exercise	15 (26.8)
Diet calories	9 (16.1)
Please rate your level of interest and motivation to track, on a daily basis, the following factors in the Mayo Clinic Migraine Care Plan	
Headache days (yes/no)	
Not at all interested	3 (5.6)
Not very interested	3 (5.6)
Neutral	8 (14.8)
Somewhat interested	15 (27.8)
Very Interested	25 (46.3)
If yes, then pain level on scale 1 - 10	
Not at all interested	3 (5.7)
Not very interested	1 (1.9)
Neutral	11 (20.8)
Somewhat interested	17 (32.1)
Very Interested	21 (39.6)
Functional impairment (mild, moderate, or severe)	
Not at all interested	3 (5.7)
Not very interested	2 (3.8)
Neutral	9 (17)

XSL•FO
RenderX

Survey question	Total respondents (N=56), n (%)
Somewhat interested	17 (32.1)
Very Interested	22 (41.5)
What type of functions was impaired (work, school, family, or personal)	
Not at all interested	3 (5.7)
Not very interested	4 (7.5)
Neutral	11 (20.8)
Somewhat interested	16 (30.2)
Very Interested	19 (35.8)
Did you take medication (yes or no)	
Not at all interested	3 (5.8)
Not very interested	0 (0)
Neutral	6 (11.5)
Somewhat interested	14 (26.9)
Very Interested	29 (55.8)
If yes, then which type of medication (pick from multiple in list)	
Not at all interested	5 (9.6)
Not very interested	1 (1.9)
Neutral	10 (19.2)
Somewhat interested	19 (36.5)
Very interested	17 (32.7)
Response to medication	
Not at all interested	3 (5.8)
Not very interested	1 (1.9)
Neutral	9 (17.3)
Somewhat interested	15 (28.8)
Very interested	24 (46.2)
Your own personal observations or comments (free text)	
Not at all interested	3 (5.8)
Not very interested	1 (1.9)
Neutral	7 (13.5)
Somewhat interested	20 (38.5)
Very interested	21 (40.4)
Please rate your level of agreement with the following statements about the Mayo Clinic Migraine Care Plan.	
I felt confident using the Mayo Clinic Care Plan	
Strongly disagree	1 (1.9)
Disagree	2 (3.8)
Neither agree nor disagree	8 (15.1)
Agree	24 (45.3)
Strongly agree	18 (34)
The Mayo Clinic Care Plan app was easy to use	
Strongly disagree	2 (3.8)
Disagree	4 (7.5)

XSL•FO
RenderX

Survey question	Total respondents (N=56), n (%)
Neither agree nor disagree	11 (20.8)
Agree	21 (39.6)
Strongly agree	15 (28.3)
The equipment helped in my care at home	
Strongly disagree	1 (1.9)
Disagree	8 (15.1)
Neither agree nor disagree	23 (43.4)
Agree	11 (20.8)
Strongly agree	10 (18.9)
I felt comfortable interacting with my care team through the Mayo Clinic Care Plan	
Strongly disagree	0 (0)
Disagree	3 (5.8)
Neither agree nor disagree	7 (13.5)
Agree	24 (46.2)
Strongly agree	18 (34.6)
It helped me better understand my condition	
Strongly disagree	1 (1.9)
Disagree	7 (13.2)
Neither agree nor disagree	23 (43.4)
Agree	12 (22.6)
Strongly agree	10 (18.9)
It helped me understand how to care for myself	
Strongly disagree	1 (1.9)
Disagree	9 (17)
Neither agree nor disagree	24 (45.3)
Agree	9 (17)
Strongly agree	10 (18.9)
It helped me understand what I should be tracking throughout my care	
Strongly disagree	1 (1.9)
Disagree	5 (9.4)
Neither agree nor disagree	24 (45.3)
Agree	14 (26.4)
Strongly agree	9 (17)
It helped me understand what steps I could take to improve my health	
Strongly disagree	1 (1.9)
Disagree	8 (15.1)
Neither agree nor disagree	25 (47.2)
Agree	11 (20.8)
Strongly agree	8 (15.1)
It helped me communicate with my care team	
Strongly disagree	2 (3.8)
Disagree	5 (9.4)

XSL•FO
RenderX

Survey question	Total respondents (N=56), n (%)
Neither agree nor disagree	8 (15.1)
Agree	27 (50.9)
Strongly agree	11 (20.8)
It helped to inform me when to contact my care team about concerning symptoms	
Strongly disagree	2 (3.8)
Disagree	9 (17)
Neither agree nor disagree	15 (28.3)
Agree	17 (32.1)
Strongly agree	10 (18.9)
Please rate your level of agreement with the following statements about the education provided while using the Mayo Clinic Care Plan.	
The educational materials were useful to me	
Strongly disagree	2 (3.9)
Disagree	6 (11.8)
Neither agree nor disagree	12 (23.5)
Agree	22 (43.1)
Strongly agree	9 (17.6)
The information was easy to understand	
Strongly disagree	1 (2)
Disagree	0 (0)
Neither agree nor disagree	11 (21.6)
Agree	26 (51)
Strongly agree	13 (25.5)
I was comfortable with how often I received educational materials	
Strongly disagree	2 (3.9)
Disagree	3 (5.9)
Neither agree nor disagree	11 (21.6)
Agree	26 (51)
Strongly agree	9 (17.6)
I was able to find the educational materials when I needed them	
Strongly disagree	3 (5.9)
Disagree	2 (3.9)
Neither agree nor disagree	13 (25.5)
Agree	25 (49)
Strongly agree	8 (15.7)
The educational materials matched my personal needs	
Strongly disagree	1 (2)
Disagree	5 (9.8)
Neither agree nor disagree	20 (39.2)
Agree	19 (37.3)
Strongly agree	6 (11.8)
The educational information from the Mayo Clinic Care Plan matched the information received from my Mayo Clinic Care Team	

XSL•FO
RenderX

Survey question	Total respondents (N=56), n (%)
Strongly disagree	1 (2)
Disagree	0 (0)
Neither agree nor disagree	18 (35.3)
Agree	21 (41.2)
Strongly agree	11 (21.6)
Please rate your level of agreement with the following statements.	
I would recommend the Mayo Clinic Care Plan to others with similar health condition(s)	
Strongly disagree	1 (2)
Disagree	2 (4)
Neither agree nor disagree	8 (16)
Agree	24 (48)
Strongly agree	15 (30)
Overall, I am satisfied with the Mayo Clinic Care Plan	
Strongly disagree	1 (2)
Disagree	4 (8)
Neither agree nor disagree	7 (14)
Agree	25 (50)
Strongly agree	13 (26)

The median age of respondents was 42 (range 20 - 72) years and 94.6% (53/56) were female. Respondents reported treatment with an injectable CGRP antagonist (17/56, 30.4%) or botulinum toxin (22/56, 39.3%). Patient-reported headache frequency at the time of the survey was (1) 0 to 8 days (26/56, 46.4%), (2) 9 to 14 days (12/56, 21.4%), and (3) 15 or more days (18/56, 32.1%). No differences in survey responses were observed comparing patients reporting CGRP antagonist treatment versus no CGRP treatment, botulinum toxin treatment versus no botulinum toxin treatment, or between headache frequency groups at the time of the survey.

In terms of frequency, respondents had the highest preference for tracking headache days weekly (30/56, 53.6%) followed by daily (15/56, 26.8%) and monthly (8/56, 14.3%), with a tracking frequency of every 3 months being the least preferred (3/56, 5.4%). Respondents indicated a similar preference for how frequently they would like to be reminded to record a headache day. Most respondents preferred to have control over the frequency of headache tracking reminders (42/56, 75%). In addition to tracking headache days, most respondents indicated a preference for tracking migraine and headache triggers (42/56, 73.2%), stress level (38/56, 67.8%), and sleep (37/56, 66.1%). Fewer participants preferred tracking diet (16/56, 28.6%), exercise (15/56, 26.8%), caloric intake (9/56, 16.1%), step count (9/56, 16.1%), and heart rate (4/56, 7.1%).

When asked about the level of interest and motivation to track on a daily basis, respondents were somewhat or very interested in tracking personal observations or comments in free text (41/52, 78.8%), medication treatment (43/52, 82.7%), response to medication (39/56, 69.6%), class of medication treatment chosen from a list (36/52, 69.2%), functional impairment graded mild, moderate, or severe (39/56, 69.6%), type of function impaired including work, school, family, personal (35/53, 66%), headache day (40/54, 74.1%), and headache pain level on scale 1 - 10 (38/53, 71.7%).

The majority of respondents agreed or strongly agreed with statements about the Migraine Care Plan that they “felt confident using” (42/53, 79.2%), “was easy to use” (36/53, 68%), “felt comfortable interacting with my care team” (42/52, 80.8%), “helped to the inform me when to contact my care team about concerning symptoms” (27/53, 50.9%), and “helped me communicate with my care team” (38/53, 71.7%). A minority of respondents agreed or strongly agreed with statements that “the equipment helped in my care at home” (21/53, 40%), “helped me better understand my condition” (22/53, 41.5%), “understand how to care for myself” (19/53, 35.8%), “understand what I should be tracking throughout my care” (23/53, 43.4%), and “understand steps I can take to improve my health” (19/53, 35.8%).

Respondents agreed or strongly agreed that the education materials were useful (31/51, 60.8%) and easy to understand (39/51, 76.5%), but fewer agreed that the education content “matched my personal needs” (25/51, 49%). Most respondents agreed or strongly agreed that they were satisfied with the Migraine Care Plan (38/50, 76%).

XSL•FO
RenderX

Discussion

Principal Findings

We assessed the satisfaction and preferences of MICP and did not observe differences based on migraine frequency or type of preventive treatment. However, analysis of all respondents highlight that the most notable MICP improvement opportunities include allowing patients to enter free text observational data and adjust the frequency of monitoring.

In addition, we observed that most respondents prefer to track headache outcomes that are of interest to clinicians engaged in migraine management, including headache days, treatment days and response, and functional impairment [22]. Most respondents in this study preferred weekly tracking as well as the option to change the frequency of tracking reminders along with the ability to enter free text data and observations and rate the severity of pain on a 1 to 10 pain scale, all of which were not included in the MICP. Most respondents were not interested in monitoring non-headache data such as diet, exercise, caloric intake, or heart rate. Most respondents agreed that the educational content delivered was useful but that the content lacked personalization. Most respondents were satisfied with the MICP and felt that it helped them communicate with their care team, supporting ongoing use in our practice.

Key MICP Improvement Opportunities and Comparison With Previous Work

Free Text Capability

We observed that 78.8% (41/52) of respondents were somewhat or very interested in recording their own free text observations. This capability was not included in the MICP because of concern that free text data might be difficult for clinicians to interpret or might not be clinically actionable or easy to summarize in the EHR. Minen et al [23] also reported a similar preference of a patient with migraine for recording free text data. Although such data may be difficult for clinicians to efficiently review within the EHR and contain clinical irrelevant observations [22], it may be important for patient engagement and improve usability. As we enter the era of artificial intelligence [24], such free text data [25] may be a powerful data set for artificial intelligence natural language processing tools that may be able to quickly summary such data or use in algorithm development to facilitate patient assessment and management [26-29].

Frequency of Monitoring

Only 15/56 (26.8%) respondents in our study indicated a preference for daily monitoring, and only 5/171 (2.9%) users in the pilot study of the MICP responded to nearly all assigned assessment tasks [15]. Raffaeli et al described a cohort of highly engaged users of a commercial app that engaged with the app daily for more than 7 - 13 months. This cohort consisted of only 1.8% of 85,000 active app users who used the app daily for 7 months [10]. These observations suggest that sustained engagement on a daily or weekly basis may not be a realistic goal for most patients. Whether giving patients the option to change the frequency of assessments would improve overall engagement is not known.

Our MICP allowed weekly tracking of the number of headache days, treatment days, and days with functional impairment. However, most of our patients indicated an interest in changing this tracking frequency (42/56, 75%). The range of different preferences in tracking frequency supports allowing patients to choose the frequency of monitoring even though the accuracy of the data may be less precise the longer a patient may be asked to recall. It is also possible that the preference for tracking frequency might depend on migraine frequency, for example, patients with less frequent episodic migraine may prefer less frequent monitoring than a patient with uncontrolled chronic migraine. We suspect that as patients are effectively treated with appropriate acute and preventative migraine therapies, headache frequency and disability decrease such that daily and even weekly monitoring would no longer be necessary. A study in a larger population of patients is needed to determine if flexibility would increase overall engagement.

Patient preferences for tracking frequency may not always align with clinician preference or insurance requirements. In addition, different tracking frequencies serve different purposes. For example, the diagnosis of episodic versus chronic migraine depends on 3 months of data, whereas a shorter period may be ideal for making changes in medications or other aspects of care.

The survey suggests that patients would prefer to control their tracking and tracking reminder frequency. Individualization of tracking and reminder frequency may enhance patient engagement and understanding, especially for those with a new diagnosis of migraine. Requiring monthly assessments of headache days, treatment days, function, and satisfaction will allow the clinician access to data that will inform best migraine care. For example, monthly tracking may quickly uncover medication overuse behavior. It may also allow the clinician early detection of the transition between episodic and chronic migraine, which in turn may allow new treatment options such as onabotulinumtoxin A. The MIDAS assessment every 3 months may not be needed except for research purposes to allow comparison with other studies or migraine care model intervention. Monthly data trends would be simple to summarize and easy to interpret for busy clinicians.

Educational Materials

Respondents agreed or strongly agreed that the education materials were useful (31/51, 60.8%) but the education did not always match personal needs (25/51, 49%). The migraine educational content was generally applicable and may be useful to a patient with a new migraine diagnosis. Most MICP users in this study had frequent migraine that led to neurology consultation, and as most were treated with botox or CGRP, it can be assumed majority failed to respond to multiple first-line preventive options. Education specific to chronic migraine was minimal. The MICP may be improved by delivering educational content that is tailored to the migraine diagnosis or current treatment. Delivering content that is already known or simply not of interest may disengage patients.

XSL•FO
RenderX

Strengths

We present survey findings from a unique group of patients that used a novel digital tool directly interfacing with the EHR within a community neurology practice. We assessed preferences for a broad range of features that may both inform migraine care and promote patient engagement. We compared respondents with nonrespondents and between subsets of patients. The low response rate was similar to other survey studies of the same population [6,30,31].

Limitations

Limitations include the small sample size, and the overall survey response rate was low, reducing the reproducibility and generalizability of our findings. In addition, the study was underpowered to detect differences in user preferences based on headache frequency or treatment type. Most patients in the survey likely had chronic migraine or frequent episodic migraine as is typical of our specialty practice. The majority of patients were treated with botulinum toxin injections or CGRP antagonist therapy at the time of the study, and the majority of patients reported less than 15 headache days per month, likely reflecting the effectiveness of their current treatment plan. We did not retrospectively review records to confirm the migraine diagnosis documented in the medication record. In addition, most of the population we assessed were white and female, with high migraine frequency and access to a neurology specialty clinic and for these reasons our observations are not generalizable to other patient populations or health care systems. Finally, we did not reassess the frequency or level of engagement of MICP users who completed the survey, increasing potential response bias.

Conclusions

Most MICP users were motivated to track clinically actionable data, including headache frequency, medication treatment with response, functional impairment, and pain intensity; features that should be maintained in the MICP and similar tools that others may develop. Opportunities to improve the MICP that may be useful for other similar remote monitoring tools include allowing patients to (1) change the frequency of assessments and notifications; (2) record personal observations or comments through free text, which may include headache triggers, accompanying symptoms, and prodrome and postdrome symptoms; (3) report headache severity using a 1 to 10 pain scale; and (4) tailor headache education based on frequency and severity (episodic vs chronic migraine). To balance the needs for patient engagement with the changes above, the MICP may be simplified by requiring a monthly assessment of headache days, treatment days, function, and satisfaction with treatment plan without MIDAS every 3 months. Remote assessment and monitoring of patients with migraine is feasible, and continued work to refine remote monitoring tools informed by patient preferences and reassessment of user preferences within diverse and larger populations, including both episodic and chronic patients with migraine, is needed.

Acknowledgments

We acknowledge Jay Mandrekar, PhD for his assistance with summarizing the survey data and statistical support. The study and manuscript were not created with the use of generative artificial intelligence. This study was funded by a grant from the Noaber Foundation and the Mayo Clinic Center for Digital Health.

Conflicts of Interest

Mayo Clinic and Epic are partners in the development and commercialization of Care Plans. No individuals involved in this study have a direct financial interest in the Migraine Interactive Care Plan. JOE has consulting agreements with K Health, Exact Sciences, Alnylam Pharmaceuticals, and MedinCell; and serves on the Scientific Advisory Board for Applied Aerosol Technologies. None declared by the other authors.

Multimedia Appendix 1
Survey responses in subsets based on headache frequency.
[DOCX File, 51 KB - formative_v9i1e66763_app1.docx]

Multimedia Appendix 2
Survey responses in subsets based on migraine treatment type.
[DOCX File, 40 KB - formative_v9i1e66763_app2.docx]

References

1. Ashina M, Katsarava Z, Do TP, et al. Migraine: epidemiology and systems of care. Lancet 2021 Apr 17;397(10283):1485-1495. [doi: 10.1016/S0140-6736(20)32160-7] [Medline: 33773613]
2. Raffaelli B, Rubio-Beltrán E, Cho SJ, et al. Health equity, care access and quality in headache - part 2. J Headache Pain 2023 Dec 13;24(1):167. [doi: 10.1186/s10194-023-01699-7] [Medline: 38087219]
3. Lipton RB, Silberstein SD. Episodic and chronic migraine headache: breaking down barriers to optimal treatment and prevention. Headache 2015 Mar;55 Suppl 2(S2):103-122. [doi: 10.1111/head.12505_2] [Medline: 25662743]

XSL•FO
RenderX

4. Dodick DW, Loder EW, Manack Adams A, et al. Assessing barriers to chronic migraine consultation, diagnosis, and treatment: results from the chronic migraine epidemiology and outcomes (CaMEO) study. Headache 2016 May;56(5):821-834. [doi: 10.1111/head.12774] [Medline: 27143127]
5. Buse DC, Armand CE, Charleston L 4th, et al. Barriers to care in episodic and chronic migraine: Results from the Chronic Migraine Epidemiology and Outcomes Study. Headache 2021 Apr;61(4):628-641. [doi: 10.1111/head.14103] [Medline: 33797078]
6. Young NP, Philpot LM, Vierkant RA, et al. Episodic and chronic migraine in primary care. Headache 2019 Jul;59(7):1042-1051. [doi: 10.1111/head.13543] [Medline: 31032913]
7. Robblee J. Telemedicine in headache medicine: a narrative review. Curr Pain Headache Rep 2023 Sep;27(9):371-377. [doi: 10.1007/s11916-023-01141-2] [Medline: 37515743]
8. Berk T, Silberstein S, McAllister P. A novel virtual-based comprehensive clinical approach to headache care. J Clin Med 2023 Aug 17;12(16):37629391. [doi: 10.3390/jcm12165349] [Medline: 37629391]
9. Mosadeghi-Nik M, Askari MS, Fatehi F. Mobile health (mHealth) for headache disorders: A review of the evidence base. J Telemed Telecare 2016 Dec;22(8):472-477. [doi: 10.1177/1357633X16673275] [Medline: 27799450]
10. Raffaelli B, Mecklenburg J, Overeem LH, et al. Determining the evolution of headache among regular users of a daily electronic diary via a smartphone app: observational study. JMIR Mhealth Uhealth 2021 Jul 7;9(7):e26401. [doi: 10.2196/26401] [Medline: 34255716]
11. Robblee J, Starling AJ. E-consultation in headache medicine: a quality improvement pilot study. Headache 2020 Nov;60(10):2192-2201. [doi: 10.1111/head.13981] [Medline: 33063855]
12. Minen MT, Szperka CL, Kaplan K, et al. Telehealth as a new care delivery model: The headache provider experience. Headache 2021 Jul;61(7):1123-1131. [doi: 10.1111/head.14150] [Medline: 34309828]
13. Minen MT, George A, Katara A, et al. Headache clinicians' perspectives on the remote monitoring of patients' electronic diary data: A qualitative study. Headache 2023;63(7):917-925. [doi: 10.1111/head.14519] [Medline: 37313636]
14. Roesch A, Dahlem MA, Neeb L, Kurth T. Validation of an algorithm for automated classification of migraine and tension-type headache attacks in an electronic headache diary. J Headache Pain 2020 Jun 12;21(1):75. [doi: 10.1186/s10194-020-01139-w] [Medline: 32532222]
15. Young NP, Ridgeway JL, Haddad TC, et al. Feasibility and usability of a mobile app-based interactive care plan for migraine in a community neurology practice: development and pilot implementation study. JMIR Form Res 2023 Oct 5;7:e48372. [doi: 10.2196/48372] [Medline: 37796560]
16. Minen MT, Adhikari S, Padikkala J, et al. Smartphone-delivered progressive muscle relaxation for the treatment of migraine in primary care: a randomized controlled trial. Headache 2020 Nov;60(10):2232-2246. [doi: 10.1111/head.14010] [Medline: 33200413]
17. Minen MT, Friedman BW, Adhikari S, et al. Introduction of a smartphone based behavioral intervention for migraine in the emergency department. Gen Hosp Psychiatry 2021;69(12-9):12-19. [doi: 10.1016/j.genhosppsych.2020.12.009] [Medline: 33485090]
18. Spreadbury JH, Young A, Kipps CM. A comprehensive literature search of digital health technology use in neurological conditions: review of digital tools to promote self-management and support. J Med Internet Res 2022 Jul 28;24(7):e31929. [doi: 10.2196/31929] [Medline: 35900822]
19. Santoro A, Delussi M, Leone M, et al. Effects of botulinum toxin on migraine attack features in chronic migraine: a six-month open-label observation study through electronic diary smartphone application. Toxins (Basel) 2019 Nov 15;11(11):668. [doi: 10.3390/toxins11110668] [Medline: 31731628]
20. Minen MT, George A, Camacho E, et al. Assessment of smartphone apps for common neurologic conditions (headache, insomnia, and pain): cross-sectional study. JMIR Mhealth Uhealth 2022 Jun 21;10(6):e36761. [doi: 10.2196/36761] [Medline: 35727625]
21. Lipton RB, Stewart WF, Sawyer J, Edmeads JG. Clinical utility of an instrument assessing migraine disability: the Migraine Disability Assessment (MIDAS) questionnaire. Headache 2001 Oct;41(9):854-861. [doi: 10.1111/j.1526-4610.2001.01156.x] [Medline: 11703471]
22. Dodick DW, Tepper SJ, Lipton RB, et al. Improving medical communication in migraine management: a modified delphi study to develop a digital migraine tracker. Headache 2018 Oct;58(9):1358-1372. [doi: 10.1111/head.13426] [Medline: 30362524]
23. Minen MT, Jaran J, Boyers T, Corner S. Understanding what people with migraine consider to be important features of migraine tracking: an analysis of the utilization of smartphone-based migraine tracking with a free-text feature. Headache 2020 Jul;60(7):1402-1414. [doi: 10.1111/head.13851] [Medline: 33300599]
24. Gutman B, Shmilovitch AH, Aran D, Shelly S. Twenty-five years of AI in neurology: the journey of predictive medicine and biological breakthroughs. JMIR Neurotech 2024;3:e59556. [doi: 10.2196/59556]
25. Lefkovitz I, Walsh S, Blank LJ, Jetté N, Kummer BR. Direct clinical applications of natural language processing in common neurological disorders: scoping review. JMIR Neurotech 2024;3:e51822. [doi: 10.2196/51822]

XSL•FO
RenderX

26. Chiang CC, Luo M, Dumkrieger G, et al. A large language model-based generative natural language processing framework finetuned on clinical notes accurately extracts headache frequency from electronic health records. medRxiv. 2023 Oct 3. [doi: 10.1101/2023.10.02.23296403] [Medline: 37873417]
27. Kreimeyer K, Foster M, Pandey A, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. J Biomed Inform 2017 Sep;73(14-29):14-29. [doi: 10.1016/j.jbi.2017.07.012] [Medline: 28729030]
28. Jensen K, Soguero-Ruiz C, Oyvind Mikalsen K, et al. Analysis of free text in electronic health records for identification of cancer patient trajectories. Sci Rep 2017 Apr 7;7(46226):46226. [doi: 10.1038/srep46226] [Medline: 28387314]
29. Messinger AI, Luo G, Deterding RR. The doctor will see you now: How machine learning and artificial intelligence can extend our understanding and treatment of asthma. J Allergy Clin Immunol 2020 Feb;145(2):476-478. [doi: 10.1016/j.jaci.2019.12.898] [Medline: 31883444]
30. Stern JI, Ebbert JO, Vierkant RA, Young NP. Migraine patient care model preferences in a community practice. Neurol Clin Pract 2023 Apr;13(2):e200131. [doi: 10.1212/CPJ.0000000000200131] [Medline: 37064592]
31. Jonker L, Fitzgerald L, Vanderpol J, Fisher S. Digital diary app use for migraine in primary care: prospective cohort study. Clin Neurol Neurosurg 2022 May;216(107225):107225. [doi: 10.1016/j.clineuro.2022.107225] [Medline: 35364371]

Abbreviations

CGRP: calcitonin gene-related peptide
EHR: electronic health record
MICP: Migraine Interactive Care Plan
MIDAS: Migraine Disability Assessment

Edited by A Mavragani; submitted 08.10.24; peer-reviewed by G Peral, M Straburzynski; revised version received 02.03.25; accepted 04.03.25; published 26.03.25.

Please cite as:
Young NP, Stern JI, Steel SJ, Ebbert JO
Mobile App-Based Interactive Care Plan for Migraine: Survey Study of Usability and Improvement Opportunities
JMIR Form Res 2025;9:e66763
URL: https://formative.jmir.org/2025/1/e66763
doi:10.2196/66763

© Nathan P Young, Jennifer I Stern, Stephanie J Steel, Jon O Ebbert. Originally published in JMIR Formative Research (https://formative.jmir.org), 26.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Changes in Health Education Literacy After Structured Web-Based Education Versus Self-Directed Online Information Seeking in Patients Undergoing Carpal Tunnel Release Surgery: Nonrandomized, Controlled Study

Mariella Seel[1], BA, MSc, Dr. sc. med.; Julian Alexander Mihalic[1,2], MD; Stefan Mathias Froschauer[1,2], MD, PhD; Bernhard Holzner[3], MD, PhD; Jens Meier[1,4], MD, PhD; Tobias Gotterbarm[1,2], MD, PhD; Matthias Holzbauer[1,2], MD, LLM

[1]Faculty of Medicine, Johannes Kepler University of Linz, Linz, Austria

[2]Department for Orthopedics and Traumatology, Kepler University Hospital GmbH, Krankenhausstrasse 9, Linz, Austria

[3]Department of Psychiatry, Psychotherapy, Psychosomatics and Medical Psychology, University Hospital of Psychiatry II, Innsbruck Medical University, Innsbruck, Austria

[4]Department of Anesthesiology and Intensive Care Medicine, Kepler University Hospital GmbH, Johannes Kepler University of Linz, Linz, Austria

Corresponding Author:
Matthias Holzbauer, MD, LLM
Faculty of Medicine, Johannes Kepler University of Linz, Linz, Austria

Abstract

Background: With advancements in anesthesiologic and surgical techniques, many surgeries are now performed as day-surgery procedures, requiring greater responsibilities for self-management from patients during the perioperative process. Online health information often lacks reliability and comprehensibility, posing risks for patients with low health literacy. Carpal tunnel release (CTR) surgery, a common day-surgery procedure, necessitates effective patient education for optimal recovery and self-management.

Objective: This study introduces the CTS Academy, a web-based education program designed for patients undergoing CTR day surgery. The study aimed to evaluate the CTS Academy's impact on patients' health education literacy (HEL) compared with self-directed online information seeking.

Methods: A scoping review on education programs focusing on the perioperative process of CTR was conducted before this study. In a nonrandomized controlled study, 60 patients scheduled for CTR were assigned to 2 groups based on the patients' preferences; the test group used the CTS Academy, while the control group performed self-directed online searches. HEL was assessed using the Health Education Literacy of Patients with chronic musculoskeletal diseases (HELP) questionnaire, focusing on patients's comprehension of medical information (COMPR), patients's ability to apply health-related information in an everyday life (APPLY), and patient's ability to communicate with health care professional (COMM). Secondary outcomes included content comprehensibility, patient preferences, platform usability, and clinical carpal tunnel syndrome (CTS)–related parameters.

Results: In the scoping review, 17 studies could be identified and included for full-text analysis. Eighteen patients each were included in the test group (13 women and 5 men) and in the control group (11 women and 7 men). The average time spent in the study was 167 and 176 days for the test and control groups, respectively. The test group showed significant improvements in APPLY (mean 28, SD 7.99 vs mean 24, SD 5.14; $P<.05$) and COMM (mean 30, SD 10.52 vs mean 25, SD 6.01; $P=.02$) after using the CTS Academy in a longitudinal analysis. No significant changes were observed in the control group. In a comparison between groups, the test group had significantly higher APPLY scores at follow-up (mean 24, SD 5.14 vs mean 33, SD 14.78; $P=.044$) and fewer comprehension issues at baseline (mean 38, SD 16.60 vs mean 50, SD 19.00; $P=.03$). The CTS-related knowledge assessment yielded 92% (66/72) versus 90% (65/72) correct answers in the test and control groups, respectively. The test group rated the CTS Academy highly in usability (6.22 of 7.00 points) and utility (6.13 of 7.00 points). Preferences leaned toward using CTS Academy alongside doctor consultations (16/18, 89%) and over self-directed searches (15/18, 84%). No significant differences were found in CTS-related symptoms between groups.

Conclusions: The CTS Academy effectively enhanced patients' HEL, especially in applying and communicating medical information. The platform's usability and utility were rated favorably, and patients preferred it over independent online information seeking. This suggests that structured, web-based education enhances patient self-management during the day surgery process.

(JMIR Form Res 2025;9:e65114) doi:10.2196/65114

XSL•FO
RenderX

KEYWORDS

carpal tunnel release surgery; patient education; structured web-based education; online information; health education literacy; web-based; health education; information seeking; carpal tunnel release; carpal tunnel surgery; non-randomized; controlled study; self-management; perioperative; online health information; health literacy; day surgery; online search; carpal tunnel; carpal

Introduction

Background

Due to continuous advancements in surgical and anesthesiologic techniques, an increasing number of surgical interventions are being performed in a day surgery setting. Consequently, patients spend less time in the hospital and are often overwhelmed by a large amount of information regarding their condition and the perioperative process provided during face-to-face interactions with medical professionals [1]. Thus, patients are required to assume greater responsibilities for self-management and self-care at home [1]. However, many patients lack the necessary knowledge and ability to effectively manage the day surgery situation, leading to an increased demand for comprehensive patient education [2]. Patients seeking online health information frequently encounter the risk of unreliable sources and incorrect information due to insufficient quality control or issues with comprehensibility [3]. Considering the low health literacy levels among some patients, this poses a substantial threat to effective health communication and patient safety [4].

Despite the low-risk profile of many ambulatory interventions, the recovery phase poses challenges, emphasizing the need for patient self-management and comprehensive education programs addressing both physical and psychological aspects [1,5] to empower patients. Patient empowerment can be defined as "a process through which people gain greater control over decisions and actions affecting their health" [6]. It can be achieved through patient education, a "process of assisting consumers of health care to learn how to incorporate health related behaviors (knowledge, skill, attitude) into everyday life with the purpose of achieving the goal of optimal health" [7]. This approach should provide patients with the health-related knowledge and skills to manage their medical conditions effectively [8]. Closely related is the concept of health education literacy (HEL), a patient's "cognitive and social skills […] to understand and use the information as conveyed in health education programmes" [9]. HEL is based on the concept of health literacy and consists of three dimensions, that are (1) the patient's ability to apply health-related information in an everyday life (APPLY), (2) the patient's ability to communicate with health care professionals (COMM), and (3) the patient's comprehension of medical information (COMPR) [9].

Goal of This Study

Open carpal tunnel release (CTR) surgery is a common day-surgery procedure for carpal tunnel syndrome (CTS) [10]. CTS affects approximately 3% of the general adult population [11] and is caused by compression of the median nerve and results in symptoms such as weakness, tingling, and pain [12]. In a recent meta-analysis, CTR as gold standard treatment results in reliable improvements in hand function and symptom severity on the long-term follow-up, while manual therapy only provides short term effects regarding pain relief [13]. To address time constraints and information overload experienced by patients during clinical visits and to enhance their comprehension of their medical condition and treatment, we aim to introduce the CTS Academy, a patient-centered web-based education program, in this study. Since the implementation of patient-centered web-based education programs focusing on the perioperative process of CTR is rare, a scoping review was conducted before this study to highlight recent gaps in existing research. Our program is tailored to the specific needs of patients undergoing day surgery for CTR and aims to provide valid and comprehensible information in a familiar environment, thereby facilitating patient self-management.

In a nonrandomized controlled study, we aimed to evaluate the CTS Academy for its effect on patients' HEL compared with self-directed online information seeking, 3 weeks after CTR (in comparison with the baseline survey).

We hypothesize that HEL is increased by participating in a web-based, structured educational programme (CTS Academy) for patients undergoing CTR. Furthermore, we examine the comprehensibility of educational content, patients' preferences in education, the usability of the CTS Academy platform, and clinical changes in CTS-related parameters. Our research addresses the scarcity of comparable implementations and aims to acquire further insights into the existing limited quantitative research in this domain.

Methods

Scoping Review of the Literature

Before the conduction of this study, we performed a scoping review of literature discussing the quality of online CTS information. From September to December 2022, we searched Epistemonikos, the Cochrane Library, PubMed, Connected Papers, and Google Scholar. Search prompts were for example "perioperative education," "web-based," "carpal tunnel release," and "structured patient education," and the complete search strategy can be requested from the authors. The search aimed to identify systematic reviews, original research, and related studies on web-based patient education and online information for patients undergoing open CTR surgery. No filters were applied for publication date or status, but only English and German full texts were analyzed. Publications with low evidence levels or lacking full study descriptions were excluded, while a broad search strategy ensured relevant materials not explicitly titled as web-based education were included. Reporting was performed based on the Joanna Briggs Institute Scoping Review Reporting Template [14] and the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) Checklist [15].

XSL•FO
RenderX

Design of the CTS Academy

We implemented the CTS Academy platform prototype using Microsoft Sway [16], a freely accessible web-based software for creating single-page websites.

The target group for this program consists of patients scheduled for open CTR day surgery who seek relevant information related to physical and mental preparation for CTR surgery.

The goal of CTS Academy is to assist patients preparing for surgery and managing the postoperative phase by providing textual information, multimedia content, self-check quizzes, downloadable or printable checklists for patients' self-directed online learning.

In the CTS Academy, we primarily focus on biophysiological, functional, experiential, and social aspects, excluding financial information due to the Austrian public health insurance system. The ethical and cognitive perspectives, emphasizing patient recognition, appreciation, and knowledge form the foundation of the CTS Academy. Patient education programs that prioritize patient-centeredness and engagement rely on evidence-based strategies, including plain language and multimodal information delivery [17]. Keulers et al [18] propose practical implications for interactive patient education, advocating for varied information presentation through text, graphics, and videos, with structured text using headings and subheadings. Summaries of lengthy texts determine the relevance of content, and the inclusion of graphs, pictures, and patient role models, such as animated characters, aids comprehension [18]. Incorporating entertaining elements, such as storytelling, enhances patients' self-efficacy, with a preference for visual materials like videos and images over plain text [19,20]. Modular approaches, including basic modules for essential knowledge and additional modules for in-depth information on living with a specific disease, are also recommended [20].

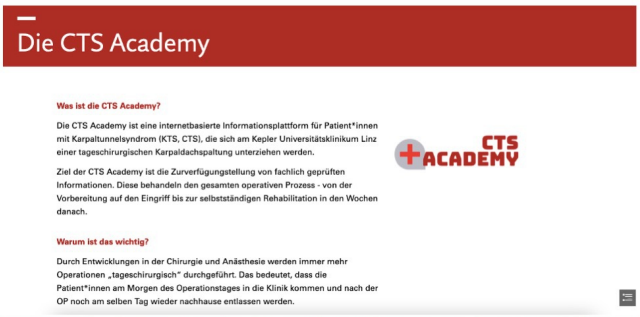
The CTS Academy is based on the information provided for informed surgery consent [21], but adds further structured and comprehensive content for various stages of day surgery. Furthermore, we included aspects of mental preparation for surgery [22-26]and instructions for the postoperative period, as well as training exercises [27]. Patients can use a printable checklist for perioperative preparation and access external links, for example, the Austrian pharmacy search to locate nearby pharmacies after hospital discharge. An animated character named Karla serves as a patient role model, guiding patients through the day-surgery process and sharing important advice. To improve accessibility, the spoken text is also displayed as written text throughout the entire video.

While the platform lacks structured patient assessments, it encourages self-assessment based on chapter summaries and offers knowledge checks. In Multimedia Appendix 1, we outline the content and media types presented on the CTS Academy platform. In Figure 1, screenshots of the CTS Academy platform are presented.

XSL•FO
RenderX

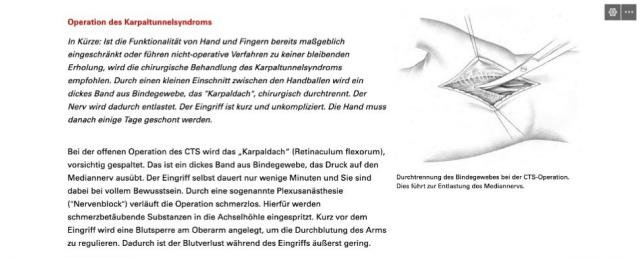
Figure 1. Screenshots of CTS academy including textual explanations of its content. It aims to support patients in preparing for surgery and managing postoperative recovery through self-directed online learning by offering textual information, multimedia content, self-check quizzes, and downloadable or printable checklists.

XSL•FO
RenderX



- Explanation of the project, the platform, project goals, relevance for the patients etc.
  - Why should I – as a patient – use this?

- Short summary at the beginning of a new chapter
  - Most important information as overview
  - Quick decision for continue reading or not
- Information on preoperative situation

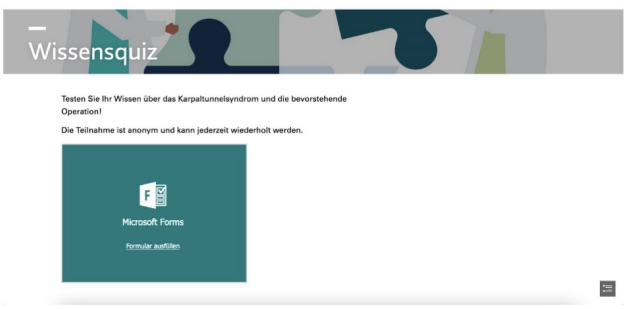


- Short summary
- Followed by longer and more detailed explanation
- Images for better comprehensibility

Unterwegs mit Karla (YouTube-Video)

- Video with spoken and written text
- Animated character “Karla” as a patient role model, sharing advice for undergoing CTS surgery
- Information on perioperative situation

- Mental preparation for targeting stress and anxiety before surgery
  - Talks, exercise, music, breathing exercises, essential oils etc.



- Short quiz for checking understanding of important information
  - What is CTS? recurrence after surgery? etc.

XSL•FO
RenderX

Design of the Evaluation Study

Between April 2022 and April 2023, all patients presenting with CTS symptoms at the Department of Orthopedics and Traumatology and scheduled for CTR surgery after medical assessment were assessed for eligibility for inclusion in this nonrandomized, controlled, confirmatory study.

During consultations, patients received detailed information about their condition and perioperative procedures. After obtaining written informed consent for CTR, we introduced the CTS Academy study and explained the test or control groups. Allocation to the groups was based on the patient's preference—either participation in CTS Academy or self-directed online information seeking.

Inclusion criteria were a scheduled appointment for CTR surgery, voluntariness and age of majority (18 y of age or older), physical and cognitive aptitude to participate, sufficient language skills of written and spoken German, and access to web-connected devices.

Furthermore, patients needed sufficient skills for internet use and an active pursuit of online health information.

Patients in the test group were granted unlimited access to the CTS Academy (from informed consent to surgery until 3 w postoperatively), while the control group conducted independent online searches. Throughout the self-study period of both groups, reminders were sent via SMS text messages or email to encourage information access.

Participants completed an online questionnaire after inclusion for the study, and 3 weeks postoperatively (follow-up appointment). As the CTS Academy study was conducted with German-speaking patients, all questionnaires were used in validated German versions.

The primary outcome was assessed using the Health Education Literacy of Patients with chronic musculoskeletal diseases (HELP) questionnaire by Farin, Ullrich, and Nagl [9]. It comprises 3 subtopics, that are comprehension of medical information (COMPR), applying medical information (APPLY), and communicative competence in provider interactions (COMM). Patients rated their perceived difficulty in health-related situations on a 5-point scale, with lower values indicating fewer difficulties.

Secondary outcomes included assessing the comprehensibility of content provided in the CTS Academy using the Comprehensibility Of Health Education Programmes (COHEP) questionnaire by Farin-Glattacker et al [28] and Farin, Nagl, and Ullrich [29]. This questionnaire comprises 4 subscales: transferability scale, medical information scale, amount scale, and the trainer scale, while the latter was dismissed because we did not offer lessons by a trainer [29]. Patients had to rate the amount of information provided or the perceived usefulness of information on a 6-point ranging from 1 (strongly agree) to 6 (strongly disagree).

Knowledge checks involved answering 4 statements about CTS and its treatment [30] to check their knowledge at baseline after inclusion to the study, and 3 weeks postoperatively.

Patient preferences of information seeking in the test group were assessed using the following statements on a 5-point rating scale (1=strongly disagree, 2=disagree, 3=no preference, 4=agree, and 5=strongly agree) in the follow-up questionnaire:

- I prefer the additional use of the CTS Academy (in addition to talking with a doctor) over the sole standard patient information provided in the medical informed consent conversation.
- I prefer the use of the CTS Academy over independent online information-seeking.

Usability and utility were evaluated using selected items of meCUE 2.0 [31], rated in the follow-up questionnaire on a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree).

Clinical CTS-related parameters were examined using the validated German-language version [32] of the Boston Carpal Tunnel Questionnaire (BCTQ) [33] using symptom severity scale, functional status scale, as well as the total BCTQ score at baseline and follow-up.

Before patient testing, we conducted a pre-, pilot-, and prototype testing of the CTS Academy platform in August 2021 to examine the usability as well as to detect functionality and usability issues. This testing involved 16 voluntary participants: 11 laypersons (some with previous CTS surgery experience) and 5 experts in hand surgery, psychology, sociology, e-learning, and online didactics. Laypersons were asked to imagine themselves in the role of CTS patients undergoing surgery. Experts reviewed the platform professionally. The feedback and suggestions were discussed with the other project team members and led to updates in pre- and postoperative recommendations, as well as revisions for clarity and gender-appropriate language (shortening of the sentence length, simplifying certain terms, and standardizing gender-inclusive language). Recommendations were added regarding preoperative and postoperative smoking abstinence, as well as the ergonomic use of computer mice and keyboards after surgery. A chapter about the author of the CTS Academy was added.

Statistical Analysis

Sample Size Estimation and Enrollment

We performed sample size estimation in G*Power (version 3.1.9.7) [34] using the following parameters: Wilcoxon-Mann-Whitney test (2 groups), 1-tailed hypothesis, Cohen *d* of 0.8, α error probability of 0.05, and power of 0.8 [35]. This yielded a required sample size of 21 patients per group, totaling 42 patients. Anticipating a 20% dropout rate, we aimed to recruit at least 50 patients. However, due to the COVID-19 pandemic's impact on surgical resourced and increased dropouts, we ultimately enrolled 60 patients. Between April 2022 and April 2023, 60 patients scheduled for open CTR at our hospital participated in the CTS Academy study. Patients presenting with CTS symptoms at the hand surgery outpatient clinic underwent diagnostic evaluation, and those requiring surgery were scheduled accordingly. During preoperative consultations, they received information about the procedure, risks, and postoperative care before providing written informed consent. Eligible patients—adults with sufficient German

XSL•FO
RenderX

language skills, cognitive and physical ability, regular internet use as well as willingness to seek online health information—were invited to join the study. No restrictions were placed on gender or education to ensure a representative sample. Interested patients could assign themselves to either the test group (structured web-based education) or control group (self-directed online information seeking) and received study details, including access to an online baseline questionnaire. Those in the test group were provided instructions for using the CTS Academy platform. Throughout the study, patients received reminders via email and SMS to complete study-related tasks, including the baseline and follow-up questionnaires. They were also notified before surgery and received postoperative messages encouraging engagement. Nonresponders were contacted within 2 days. Throughout the study, participants could reach the study team via phone, SMS, or email.

Data Exclusion and Dropouts

In our online questionnaires, all fields were mandatory except for the BCTQ functional scale, as some patients may not perform tasks specified in this BCTQ scale. Patients could review and change their answers before submitting the questionnaire. In the test group, 1% (2/144) of values were missing at baseline and at follow-up. In the control group, 6% (8/144) and 1% (2/144) were missing at baseline and follow-up, respectively. Following BCTQ guidelines, we excluded participants with more than 1 missing value, resulting in the exclusion of 1 participant from each group [36]. Since BCTQ activities are not directly comparable (eg, holding a book, doing chores, and getting dressed), missing values were not imputed.

Due to the per-protocol analysis of this study, 12 patients in each group dropped out of the study due to various reasons such as nonsubmission of questionnaires, withdrawal of consent, or canceled surgery appointments.

Data Preparation and Analysis

Microsoft Excel 16.72 for Mac, and IBM SPSS (version 29) were used for statistical testing.

Descriptive statistics included means and SD as well as the IQR to identify potential outliers. Normal distribution was tested using the Shapiro-Wilk-Test. The HELP questionnaire results were analyzed by summing each subscale (APPLY, COMM, and COMPR) for individual patients and transforming these sums to a 0‐100 scale. Delta values between pre- and postoperative scores (DIFF_APLLY, DIFF_COMM, and DIFF_COMPR) were calculated.

Depending on the parameters' normal distribution, either the exact Wilcoxon test or the *t* test for dependent variables and the Mann-Whitney *U* test or the *t* test for independent variables were used for comparisons. The total HELP score (a sum of the subtypes APPLY, COMM, and COMPR) was tested at baseline, follow-up, and as difference between the former using the multivariate analysis of variances (MANOVA).

Nominal data were tested using the chi-square test or the Fisher exact test if the requirements of the former test were not met. Effect sizes were determined using Psychometrica [37].

Demographic data were examined for group comparability, focusing on age and educational level. Statistical tests were primarily 1-sided, with significance set at $P<.05$. We converted results from default 2-sided tests to 1-sided outcomes and did not perform subgroup analyses.

The COHEP questionnaire results were analyzed by summing subscales and transforming them to a 0‐100 scale. Knowledge check results were reported as the frequency of correct answers within each group. Usability and utility of the CTS Academy platform were assessed using the meCue 2.0 data template [38]. For the assessment of CTS-related symptom changes, symptom severity, functional status, and total BCTQ, of which a delta value (DIFF_BCTQ) was calculated, were compared between cohorts as well as longitudinally.

Ethical Considerations

This study was approved and registered by the institutional review board (EK1212-2021). Written informed consent (Multimedia Appendix 2) based on the hospital's regulative for clinical studies was obtained from every patient willing to participate in the study. To maintain data security, patients were pseudonymized by assigning each one an alphanumeric ID at the time of study enrollment. This ID was used exclusively for all surveys. A list of IDs alongside patient names was kept in the study office and was accessible to the project management team, for example, to use the reminder system (eg, reminders to complete the questionnaire). During the analysis of the results, it was not possible at any time to link the identity of the test participants to the available data. No identification of individual participants in any of the figures or additional material is possible. Patients received no compensation for study participation.

The study was reported using the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist (Multimedia Appendix 3) [39].

Results

Scoping Review

In total, 17 reports were included for full-text analysis, finding that patients searching online for CTS information risk encounter incorrect or confusing content [3], which underlines the need for reliable education programs. An overview of the results is presented in Table 1.

XSL•FO
RenderX

Table . Outcomes of the 17 references included in the scoping review on patient-centered web-based education programs focusing on the perioperative process of carpal tunnel release conducted in Epistemonikos, the Cochrane Library, PubMed, Connected Papers, and Google Scholar up to December 2022, discussing the quality of online carpal tunnel syndrome information.

Reference	Topic	Relevant results
Cook et al [40]	Complexity of online information on hand surgery	The recommended 6th-grade reading level was not met in any of the analyzed CTS[a] material, which decreases understandability for many patients.
Goyal et al [41]	Analysis of misleading CTS information in YouTube videos	78% (47/60) of the analyzed videos contained at least 1 potentially misconception-reinforcing statement, more than 335% (21/60) contained more than 3 statements, mostly on CTS causes and treatment options.
Lutsky et al [42]	Re-evaluation of CTS information on the internet (updated study from 2000)	Significant improvement of online CTS information quality (P<.001), further improvement is needed.
Steimle et al [43]	Comparison of online CTS information with the Clinical Practice Guidelines of the American Academy of Orthopaedic Surgeons	Incorrect information might provoke misunderstandings and potential risks for patients.
Özbek et al [44]	Quality and reliability of CTS videos on YouTube	Significant differences in reliability and informational scores (P<.05) depending on the source of the video, and generally low reliability and educational quality.
Fang et al [45]	Content and design of online CTS education handouts	Handouts contained misleading information, lacked references or citations, and were complicated to understand.
Frueh et al [46]	Quality of online CTS information	Significant difference in information quality in older compared to newer CTS websites (P<.001). Significant difference in explaining the handling of complications (P=.03).
Frické et al [47]	Indicators of accuracy for online CTS information	Indicators like high Google ratings and inlinks were correlated with accuracy (P<.05).
Beredjiklian et al [48]	Source and content quality of online CTS information	Several websites had misleading or unconventional about CTS treatment. The mean information value score of all websites was low and differed based on authorship.
Kwak et al [49]	Quality and reliability of CTS YouTube videos	Low reliability and low educational quality of the analyzed videos.
Johnson et al [50]	Comprehensibility and applicability, cultural sensitivity, and readability of English and Spanish online CTS information material	Low comprehensibility, applicability, and readability scores, and no statistically significant differences in the overall educational quality between English and Spanish material.
Eberlin et al [51]	Readability of online CTS material	The recommended 6th grade reading level was not met in any of the analyzed CTS material, which decreases understandability for many patients.
Mert and Bozgeyik [52]	Quality and content of YouTube videos about CTS	Generally insufficient quality of content, only few videos reached good scores for global quality and accuracy.
Sproule et al [53]	Quality of online information about carpal tunnel release, Dupuytren's fasciectomy, and trigger finger release	The majority of material provided accurate but incomplete information, and a lack of information on prognosis and complications of surgical treatment.
Ozdemir et al [54]	Quality and reliability of CTS videos on YouTube	Overall low quality of videos concerning accuracy and transparency. Lack of information on treatment risks and benefits. Significantly better scores when authored by medical centers (P=.02).

XSL•FO
RenderX

Reference	Topic	Relevant results
Park et al [55]	Accuracy of online CTS images	Only 30% (6/20) of the analyzed images had incorrect anatomical information, for example, the position of the median nerve.
Akinleye et al [56]	Readability of online hand pathology material including CTS	The recommended 6th grade reading level was not met in any of the analyzed CTS material, which decreases understandability for many patients.

[a]CTS: carpal tunnel syndrome.

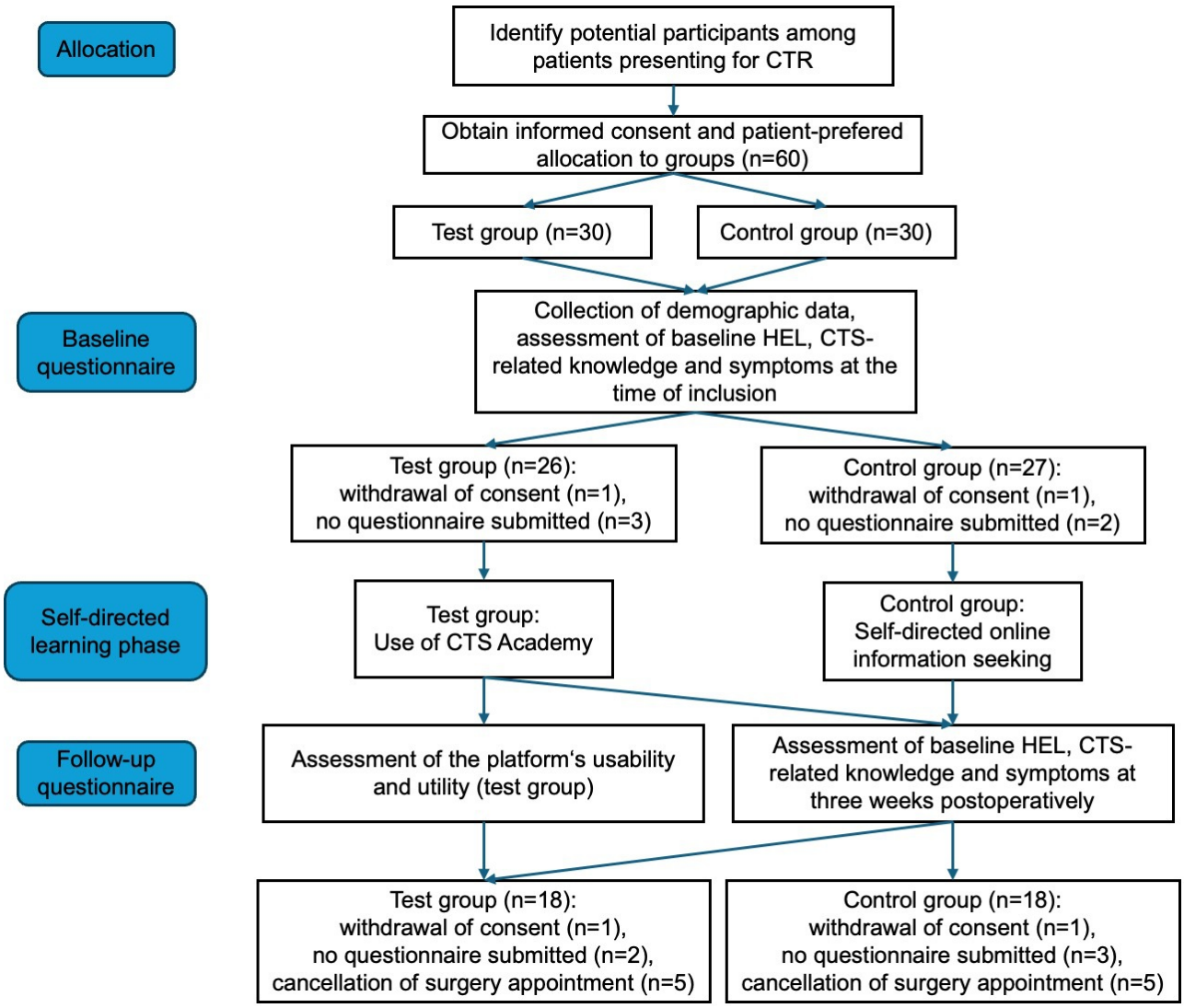
Patient Characteristics

Between April 2022 and April 2023, 30 patients were enrolled in each group. Figure 2 displays a flowchart of patients including the reasons for dropout.

Ultimately, both the test and the control group comprised of 18 patients each, whose characteristics are summarized in Table 2.

The average educational phase duration (CTS Academy platform or self-directed online information seeking) was 167 (range 43-267) days, from inclusion to 176 (range 43-253) days postoperatively.

Figure 2. Consort flowchart of patients' enrollment, follow-up, and analysis in the test group (CTS Academy) and control group (Self-directed online information seeking). CTR: carpal tunnel syndrome; CTS: carpal tunnel sydrome; HEL: health education literacy.



XSL•FO
RenderX

Table . Patient characteristics and demographics of the final test group (CTS Academy platform) and control group (self-directed online information seeking).

Categories and subcategories		Test group (n=18), n	Control group (n=18), n
Sex			
	Male	5	7
	Female	13	11
Age group			
	<30 years	0	0
	30 - 40 years	4	3
	41 - 50 years	3	3
	51 - 60 years	5	2
	61 - 70 years	3	6
	71 - 80 years	1	4
	>80 years	2	0
Highest education			
	Compulsory schooling or less	2	5
	High school (without A-levels) or apprenticeship	11	9
	High school with A-levels, diploma, apprenticeship on "Master" level	4	4
	Undergraduate degree (bachelor)	1	0
	Graduate degree (master, PhD, etc)	0	0
CTS[a] surgery			
	First time	14	11
	Previous CTS surgery (other hand)	4	6
	Previous CTS surgery (same hand)	0	0
	Previous CTS surgery (both hands)	0	1
Internet use			
	Several times a month	5	7
	Several times a week	5	3
	Daily	8	8
Average duration of CTS Academy use (days)		167 (range 43-267)	176 (range 43-253)

[a]CTS: carpal tunnel syndrome.

Primary Outcome

Outcomes of the primary parameter and its subtypes are presented in Table 3 and Table 4.

XSL•FO
RenderX

Table . Outcomes (pre-post intervention analysis) of health education literacy of patients with chronic musculoskeletal diseases questionnaire (health education literacy) via subitems comprehension of medical information, applying medical information, and communicative competence in provider interactions of the final test group (CTS Academy) and control group (self-directed online information seeking) including statistical results (all *P* values are reported as 1-sided *P* values).

	APPLY[a]			COMM[b]			COMPR[c]		
	Baseline, mean (SD)	Follow-up, mean (SD)	*P* value	Baseline, mean (SD)	Follow-up, mean (SD)	*P* value	Baseline, mean (SD)	Follow-up, mean (SD)	*P* value
Test group (n=18)	28 (7.99)	24 (5.14)	.006[d]	30 (10.52)	25 (6.01)	.017[d]2	38 (16.60)	35 (9.90)	.24
Control group (n=18)	33 (14.17)	33 (14.78)	.5	35 (17.12)	30 (15.19)	.0596	50 (19.00)	46 (23.75)	.18

[a]APPLY: applying medical information.

[b]COMM: communicative competence in provider interactions.

[c]COMPR: comprehension of medical information.

[d]Statistically significant results (α<.05).

Table . Outcomes (comparison between groups [*P* values]) of health education literacy of patients with chronic musculoskeletal diseases questionnaire (health education literacy) via subitems comprehension of medical information, applying medical information, and communicative competence in provider interactions of the final test group (CTS Academy) and control group (self-directed online information seeking) including statistical results (all *P* values are reported as 1-sided *P* values).

	APPLY[a]	COMM[b]	COMPR[c]	HEL[d]
Baseline	.2	.3	.03[e]	.13
Follow-up	.04[e]	.27	.06	.05[e]
Difference	.09	.38	.37	—[f]

[a]APPLY: applying medical information.

[b]COMM: communicative competence in provider interactions.

[c]COMPR: comprehension of medical information.

[d]HEL: health education literacy.

[e]Statistically significant results (α<.05).

[f]Not applicable.

In the test group, we found a statistically significant decrease in the group means for APPLY (*P*=.006) and COMM (*P*=.02) after using the CTS Academy, compared with baseline. However, we found no statistically significant difference regarding the COMPR subparameter. An overview of statistically significant postoperative changes on item-level is provided in Multimedia Appendix 4. In the control group analysis, we found no statistically significant difference.

In the comparison between the test and control group, significantly higher APPLY values were found in the test group during the follow-up assessment (*P*=.04). For COMPR, we found the test group had significantly fewer problems with comprehension of medical information at baseline (*P*=.03).

We found a statistically significant difference between the groups on the combined dependent variables at follow-up (*P*=.05). In the subsequent post hoc univariate ANOVA for each dependent variable (between-participant effects), we found a statistically significant difference between the groups for APPLY in the follow-up examination (*P*=.007).

Secondary Outcomes

As lower means indicate better results, we found the CTS Academy yielded good results on the transferability (mean 27, SD 9.74) and the medical information scale (mean 28, SD 10.12) and scored average results for the amount of medical information provided (mean 58, SD 24.61).

Both groups achieved high accuracy in describing CTS and the surgical procedure at baseline (63/72, 88% vs 62/72, 86% in test group and control group). At follow-up, the test group showed an average accuracy of 92%, with incorrect answers for only 1 statement. The control group yielded an average accuracy of 90% with incorrect answers for all 4 statements. No control group patient answered all statements correctly.

Regarding patients' preference for using the CTS Academy, 10 out of 18 (56%) patients totally agreed and 6 out of 18 (33%) agreed that they preferred the additional use of the CTS Academy (in addition to the consultation of a medical doctor). Furthermore, 2 patients had no preferences. When asked whether they preferred the CTS Academy over independent online information-seeking, 10 out of 18 (56%) patients totally agreed and 5 out of 18 (28%) agreed, while 3 patients had no preferences.

Patients rated usability with a mean 6.22 (SD 0.82) points (4.67 to 7.00 points), and utility with a mean 6.13 (SD 0.89) points (4.00 to 7.00 points).

XSL•FO
RenderX

The CTS-related symptoms assessed via the BCTQ are displayed in Table 5.

We found no statistically relevant differences between the groups (comparison of groups at baseline: P=.34 [2-sided], follow-up: P=.47 [2-sided], and for DIFF_BCTQ: P=.86 [2-sided]).

Table . Carpal tunnel syndrome–related symptoms of test group and control group at baseline and follow-up, assessed with the Boston carpal tunnel questionnaire.

	Symptom severity, mean (SD)		Functional status, mean (SD)	
	Baseline	Follow-up	Baseline	Follow-up
Test group (n=18)	3.03 (0.68)	1.99 (0.71)	2.24 (0.73)	2.24 (0.87)
Control group (n=18)	3.36 (0.92)	2.19 (0.98)	2.40 (1.04)	2.48 (1.04)

Discussion

Principal Results

This study introduced the CTS Academy to improve patients' HEL compared with self-directed online information seeking. Significant improvements were observed in the HELP subscales APPLY and COMM within the test group, indicating that the CTS Academy enhances patients' ability to apply medical information and communicate with health care professionals. Patients in the study group rated both utility (usefulness) and usability (ease of use) of the CTS Academy favorably.

CTS Academy content was evaluated favorably for transferability and medical information, with average ratings for the amount of medical information provided. Development relied on existing research in clinical patient education due to a lack of comprehensive guidelines, potentially affecting perceived content quality [17]. Future improvements could involve personalized pathways based on patient knowledge levels.

Both groups showed strong understanding of CTS etiology before and after the intervention. The test group's understanding of CTR surgery improved notably, with all participants (N=18) having all answers correct at follow-up. Misconceptions about complete symptom removal persisted in both groups, which might be due to a focus on postoperative recovery over potential complications.

Patients preferred the CTS Academy over self-directed online searches, indicating its value as a reliable information source. We conclude that integrating the CTS Academy into patient education could effectively enhance access to information and understanding of the perioperative situation, complementing discussions with medical professionals during informed consent.

High ratings for utility and usability suggest the platform is user-friendly and effective for CTR day surgery.

No significant differences were found between groups regarding perceived symptom severity or functional status, reflecting consistent surgical intervention.

The CTS Academy is the first web-based patient education program for CTS. No direct comparisons in medical literature are available, especially because no comparable programs have health education as primary outcome. Future updates to the CTS Academy will focus on restructuring content and enhancing readability and comprehensibility. The platform has the potential to evolve into an educational program with user accounts, enabling patients to follow structured, personalized learning paths. The integration of technologies such as artificial intelligence could further personalize and enhance the learning experience. Another future goal of this project is expanding the platform to include multiple languages relevant to clinical practice at our hospital.

Comparison With Previous Work

Patient education is crucial for day surgery, as it helps prevent adverse events and rehospitalization [57]. Pekonen et al [58] conducted a systematic review on tools for measuring patient empowerment after patient education interventions. While the Health Education Impact Questionnaire is well-established [59], the HELP questionnaire specifically addresses HEL. Because the latter not just assesses knowledge acquisition but also patients' ability to apply and understand educational interventions, it was deemed more appropriate for the CTS Academy study.

A publication by Farin-Glattacker et al [28] assessed HEL using the HELP questionnaire as an outcome measure in patients undergoing rehabilitation for chronic back pain or osteoarthritis [60]. The study group stated a high level of self-reported HEL, with the average patient experiencing slight difficulty on the APPLY and COMM scales, and slight to moderate difficulty on the COMPR scale [9]. It aligns with the findings from the CTS Academy that difficulty in understanding medical terminology appears to be a common challenge [61-64].

Regarding computer-based education on CTS, a randomized study compared knowledge levels and satisfaction scores between patients receiving the same information either using computer-based education or using face-to-face education with a medical doctor. Ultimately, the computer-based education group scored significantly higher on the knowledge questionnaire [35].

In total, 2 previous publications focused on the topic of structured web-based patient education versus self-directed online information seeking. Finkelstein and Bedra [65] conducted this comparison on hypertension knowledge in 60 patients with hypertension. The test group used a structured web-based patient education program on hypertension, while the control group searched the internet for information. The test group showed significantly higher hypertension knowledge scores compared with the baseline and compared with the

XSL•FO
RenderX

control group. Semistructured interviews indicated a better learning experience in the test group, particularly for those with limited computer and internet experience. Castro et al [66] found comparable results in a similar cohort study. Thus, both concluded that web-based patient education was effective for hypertension learning, even for participants with limited internet experience [65,66].

Although health literacy is often seen as empowering, it does not automatically lead to empowerment [67]. Health education is crucial for promoting health literacy [67], with health knowledge being a key dimension [68,69] according to the conceptual model of health literacy by Sørensen et al [70]. Low health literacy levels are associated with poorer health outcomes, increased morbidity and mortality [71], highlighting the importance of patient education. Several studies have proven that multimedia education programs can improve health knowledge, such as in patients with type 2 diabetes [72]. Despite existing literature on digital health interventions for improving health literacy, only few studies specifically measure health literacy outcomes, often focusing on knowledge-related outcomes instead [73]. This distinction underscores the importance of CTS Academy's focus on HEL.

Various publications on web-based patient education address the quality and readability of online educational materials, but few compare structured web-based patient education and self-directed online information seeking. Many examined online resources presenting orthopedic, trauma surgery, pediatrics, or cardiology content [74-76] exceed the recommended sixth-grade reading level [77], posing comprehension challenges. The quality and accuracy of online educational resources varies, with those authored by medical professionals only being slightly better than those authored by commercial organizations or laypersons [44,48,54]. Due to a lack of structured analyses of German educational online resources, quality and accuracy of online information for German-speaking CTS patients cannot be assessed. However, studies indicate that more than one-third of German-speaking orthopedic patients used the internet before their hospital appointment, and 80% (63/79) of these patients intended to gain additional knowledge of their condition and treatment options. The majority of patients performing self-directed online information seeking would recommend it to other orthopedic patients [78]. Hertling et al [79] reports that 66% (832/1262) of orthopedic patients used the internet as a source of disease-related information, and the majority found the information trustworthy and useful. However, it is unclear whether patients can assess the reliability and correctness of online information and identify incorrect information [79]. While both studies indicated widespread use and acceptance of online sources among German orthopedic patients, the quality of the information sources remains unexamined.

Comparing the comprehensibility of CTS Academy content with other programs using the COHEP questionnaire is challenging due to differing contexts. Existing studies by Farin-Glattacker et al [28], Nagl et al [60], and Finkelstein et al [65] focused on different patient populations and did not provide detailed comparative data.

Patients provided positive feedback on CTS Academy's usability and utility, appreciating its design and functionality, facilitating effective access and utilization of educational materials. Overall, these findings suggest that the CTS Academy is a valuable tool for educating and supporting patients undergoing day surgery for CTR.

In our study, we collected postoperative clinical parameters 3 weeks after CTR. Our clinical results align with the ones published by Mack et al [80] at 10 - 14 days postoperatively and Kim et al [81] and Aloi et al [82] at 14 days postoperatively, despite slight differences in patient characteristics and surgical approaches.

Limitations

This study also involves several limitations. The study was affected by COVID-19, especially by prolonged surgery waiting times leading patients to cancel their surgery appointments. This led to a smaller sample size than intended, potentially biasing our results. Furthermore, the patient-performed allocation process may have introduced selection bias. Moreover, group sizes were relatively small, potentially affecting the generalizability of the results. We did not ask patients to monitor the duration of information-seeking or the frequency of use of the CTS Academy. In addition, the quality of information accessed by the control group was not controlled, which could have influenced their outcomes and skewed comparisons with the intervention group. However, both groups completed identical baseline questions on CTS before accessing the platform content or other online resources, ensuring a basic understanding of the scope of information. These questions required basic knowledge, independent of the depth of research or exclusive platform content. This highlights a common challenge in independent online research, that is, patients struggle to identify essential information, and available content can be difficult to interpret. While the test group benefited from a patient-centered platform design, this advantage did not extend to the fundamental information itself. To preserve the natural research and learning process, patients were not required to document their sources, as this might disrupt their learning experience. For privacy reasons and to prevent artificially inflated research activities, browser histories were not collected to avoid potential identification or bias. In future research, randomization allocation to either the test or control group should be conducted to further investigate effects of the CTS Academy platform on HEL, ensuring that the results are not affected by a selection bias. In terms of usability, common usability testing guidelines report that 5 participants already detect almost all of the possible errors in a product [83]. Therefore, the usability conclusions can be seen as valid despite of the generally decreased study's power and generalizability.

Conclusions

Knowledge is commonly measured as a primary outcome in patient education programs for its ease of pre and post comparisons. However, the ability to apply knowledge in daily life is crucial for effective self-management. Health literacy models emphasize both knowledge and its practical application. Hence, we recommend a comprehensive evaluation for patient education programs, beyond knowledge assessments. Our

XSL•FO
RenderX

findings highlight the effectiveness of our educational intervention for patients undergoing CTR day-surgery. Despite challenges in comparing outcomes with other studies, statistical analysis favored the CTS Academy in improving overall HEL. Therefore, we rejected the null hypothesis, confirming the platform's efficacy over self-directed online information seeking. Future research could explore the quality of German-language online educational resources by having patients document and analyze their information sources. In addition, the effectiveness of the CTS Academy following its implementation at the Department of Orthopedics and Traumatology could be assessed. Investigating patient preferences and the utility of each specific chapter could help identify the most relevant educational needs and could be analyzed on item-level of each HELP or COHEP subscale for a granulated evaluation of patients' learning experience.

Acknowledgments

This study is supported by Johannes Kepler University Open Access Publishing Fund and the Federal State Upper Austria.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization, methodology, software, data curation, and writing—original draft preparation were contributed and handled by MS. Validation was contributed by TG, JM, and BH. Formal analysis and investigation was contributed by MS, JAM, and SMF. Resources was handled by TG. Writing—review and editing was contributed by MH. Visualization was contributed by MS and MH. Supervision was managed by BH, JM, and TG. Project administration was handled by SMF.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of content and media types used in the CTS Academy (in order of appearance on the platform).
[DOCX File, 74 KB - formative_v9i1e65114_app1.docx]

Multimedia Appendix 2
Informed consent.
[PDF File, 239 KB - formative_v9i1e65114_app2.pdf]

Multimedia Appendix 3
Cherries (Checklist for Reporting Results of Internet E-Surveys) checklist.
[PDF File, 670 KB - formative_v9i1e65114_app3.pdf]

Multimedia Appendix 4
Overview of pre- and postoperative health education literacy on item-level for the test group.
[DOCX File, 24 KB - formative_v9i1e65114_app4.docx]

References

1. Hälleberg Nyman M, Nilsson U, Dahlberg K, Jaensson M. Association between functional health literacy and postoperative recovery, health care contacts, and health-related quality of life among patients undergoing day surgery: secondary analysis of a randomized clinical trial. JAMA Surg 2018 Aug 1;153(8):738-745. [doi: 10.1001/jamasurg.2018.0672] [Medline: 29710226]
2. Berg K, Arestedt K, Kjellgren K. Postoperative recovery from the perspective of day surgery patients: a phenomenographic study. Int J Nurs Stud 2013 Dec;50(12):1630-1638. [doi: 10.1016/j.ijnurstu.2013.05.002] [Medline: 23726224]
3. Tonsaker T, Bartlett G, Trpkov C. Health information on the internet: gold mine or minefield? Can Fam Physician 2014 May;60(5):407-408. [Medline: 24828994]
4. Connelly RA, Speer ME. Health literacy and health communication. In: Connelly RA, Turner T, editors. Health Literacy and Child Health Outcomes: Springer International Publishing; 2017:1-13. [doi: 10.1007/978-3-319-50799-6_1]
5. Charalambous A, Papastavrou E, Valkeapää K, et al. Content of orthopedic patient education provided by nurses in seven European countries. Clin Nurs Res 2018 Sep;27(7):770-789. [doi: 10.1177/1054773817713178] [Medline: 28691509]
6. World Health Organization. Empowerment. In: Health Promotion Glossary of Terms: World Health Organization; 2021. URL: https://www.who.int/publications/i/item/9789240038349 [accessed 2025-03-21]

XSL•FO
RenderX

7. Bastable SB, editor. Essentials of Patient Education, 2nd edition: Jones & Bartlett Learning; 2017.
8. Wittink H, Oosterhaven J. Patient education and health literacy. Musculoskelet Sci Pract 2018 Dec;38:120-127. [doi: 10.1016/j.msksp.2018.06.004] [Medline: 30017902]
9. Farin E, Ullrich A, Nagl M. Health education literacy in patients with chronic musculoskeletal diseases: development of a new questionnaire and sociodemographic predictors. Health Educ Res 2013 Dec;28(6):1080-1091. [doi: 10.1093/her/cyt095] [Medline: 24113084]
10. Towfigh H, Hierner R, Langer M, Friedel R, editors. Handchirurgie Bd: Springer; 2011, Vol. 2.
11. Atroshi I, Gummesson C, Johnsson R, Ornstein E, Ranstam J, Rosén I. Prevalence of carpal tunnel syndrome in a general population. JAMA 1999 Jul 14;282(2):153-158. [doi: 10.1001/jama.282.2.153] [Medline: 10411196]
12. American Academy of Orthopaedic Surgeons. Carpal tunnel syndrome. : OrthoInfo; 2016. URL: https://orthoinfo.aaos.org/en/diseases--conditions/carpal-tunnel-syndrome/ [accessed 2021-01-26]
13. Donati D, Boccolari P, Tedeschi R. Manual therapy vs. surgery: which is best for carpal tunnel syndrome relief? Life (Basel) 2024 Oct 11;14(10):1286. [doi: 10.3390/life14101286] [Medline: 39459587]
14. The Joanna Briggs Institute. Templates for scoping reviews. Downloadable templates in Word guide and assist reviewers in developing a scoping review protocol and scoping review. 2022. URL: https://jbi.global/scoping-review-network/resources [accessed 2022-12-30]
15. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
16. Microsoft Corporation. Microsoft sway. : 2021 URL: https://sway.office.com/ [accessed 2025-03-21]
17. Health Care Education Association. Patient education practice guidelines for health care professionals. 2021. URL: https://www.hcea-info.org/assets/hcea%20guidelines%20color%201-25-2021.pdf [accessed 2025-03-21]
18. Keulers BJ, Keulers MJ, Scheltinga MRM, Spauwen PHM. Evidence-based development of interactive patient education programs: a guide for achieving optimal computer-based patient education. Eur J Plast Surg 2006 Nov 13;29(4):169-175. [doi: 10.1007/s00238-006-0085-7]
19. Raunsbæk Knudsen L, Lomborg K, Ndosi M, Hauge EM, de Thurah A. The effectiveness of e-learning in patient education delivered to patients with rheumatoid arthritis: The WebRA study-protocol for a pragmatic randomised controlled trial. BMC Rheumatol 2021 Dec 20;5(1):57. [doi: 10.1186/s41927-021-00226-y] [Medline: 34924034]
20. Knudsen LR, Lomborg K, de Thurah A. Design and development of an e-learning patient education program for self-management support in patients with rheumatoid arthritis. PEC Innov 2022 Dec;1:100004. [doi: 10.1016/j.pecinn.2021.100004] [Medline: 37364010]
21. Sauerbier M, Allert S. Operation des karpaltunnelsyndroms (KTS) offen/endoskopisch [report in german]. 2017 URL: https://thieme-compliance.de/de/shop/Artikel/Aufkl%C3%A4rungsb%C3%B6gen/Karpaltunnelsyndrom-%28KTS%29%2C-Operation/p/DE61748501 [accessed 2025-03-25]
22. Norelli SK, Long A, Krepps JM. Relaxation techniques. In: StatPearls: StatPearls Publishing; 2021. URL: http://www.ncbi.nlm.nih.gov/books/NBK513238 [accessed 2023-07-19]
23. Anxiety and Depression Association of America (ADAA). Exercise for stress and anxiety. : Anxiety and Depression Association of America (ADAA); 2021. URL: https://adaa.org/living-with-anxiety/managing-anxiety/exercise-stress-and-anxiety [accessed 2021-05-06]
24. Pardede J, Simanjuntak GV, Manalu N. Effectiveness of deep breath relaxation and lavender aromatherapy against preoperative patent anxiety. Divers Equal Health Care 2020;17(4):168-173. [doi: 10.36648/206]
25. Uğraş GA, Yıldırım G, Yüksel S, Öztürkçü Y, Kuzdere M, Öztekin SD. The effect of different types of music on patients' preoperative anxiety: a randomized controlled trial. Complement Ther Clin Pract 2018 May;31:158-163. [doi: 10.1016/j.ctcp.2018.02.012] [Medline: 29705448]
26. Braden R, Reichow S, Halm MA. The use of the essential oil lavandin to reduce preoperative anxiety in surgical patients. J Perianesth Nurs 2009 Dec;24(6):348-355. [doi: 10.1016/j.jopan.2009.10.002] [Medline: 19962101]
27. American Academy of Orthopaedic Surgeons, OrthoInfo. Therapeutic exercise program for carpal tunnel syndrome. 2018. URL: https://orthoinfo.aaos.org/globalassets/pdfs/a00789_therapeutic-exercise-program-for-carpal-tunnel_final.pdf [accessed 2025-03-21]
28. Farin-Glattacker E, Schöpf A, Ullrich A, Nagl M. Abschlussbericht zum projekt „entwicklung und evaluation einer patientenschulung zur Förderung der gesundheitskompetenz Für chronisch kranke" (geko) [article in german[. Universitätsklinikum Freiburg. 2014. URL: https://docplayer.org/14332968-Abschlussbericht-zum-projekt.html [accessed 2025-03-21]
29. Farin E, Nagl M, Ullrich A. The comprehensibility of health education programs: questionnaire development and results in patients with chronic musculoskeletal diseases. Patient Educ Couns 2013 Feb;90(2):239-246. [doi: 10.1016/j.pec.2012.10.004] [Medline: 23127897]
30. Aung KK, Wu WK, Tokumi A, Kuo P, Day CS. Does a directive to an internet site enhance the doctor-patient interaction? A prospective randomized study for patients with carpal tunnel syndrome. J Bone Joint Surg Am 2015 Jul 1;97(13):1112-1118. [doi: 10.2106/JBJS.N.00741] [Medline: 26135078]

XSL•FO
RenderX

31. Minge M, Thüring M, Wagner I, Kuhr CV. The mecue questionnaire: A modular tool for measuring user experience. In: Soares M, Falcão C, Ahram TZ, editors. Advances in Ergonomics Modeling, Usability & Special Populations: Springer International Publishing; 2017, Vol. 486:115-128. [doi: 10.1007/978-3-319-41685-4_11]
32. Keilani M, Pernicka E, Paternostro-Sluga T, et al. Übersetzung und Validierung des „Boston Carpal Tunnel Syndrome Questionnaire" zum Einsatz bei deutschsprachigen Patienten. Phys Rehab Kur Med 2008 Jun;18(3):136-144. [doi: 10.1055/s-2007-1004550]
33. Levine DW, Simmons BP, Koris MJ, et al. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. J Bone Jt Surg 1993 Nov;75(11):1585-1592. [doi: 10.2106/00004623-199311000-00002]
34. Buchner A, Erdfelder E, Faul F, Lang AG. G*Power. URL: https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower [accessed 2025-03-21]
35. Keulers BJ, Welters CFM, Spauwen PHM, Houpt P. Can face-to-face patient education be replaced by computer-based patient education? A randomised trial. Patient Educ Couns 2007 Jul;67(1-2):176-182. [doi: 10.1016/j.pec.2007.03.012] [Medline: 17448621]
36. Multanen J, Ylinen J, Karjalainen T, Kautiainen H, Repo JP, Häkkinen A. Reliability and validity of the Finnish version of the Boston carpal tunnel questionnaire among surgically treated carpal tunnel syndrome patients. Scand J Surg 2020 Dec;109(4):343-350. [doi: 10.1177/1457496919851607] [Medline: 31132964]
37. Lenhard W, Lenhard A. Computation of effect sizes. 2016. [doi: 10.13140/RG.2.2.17823.92329]
38. Minge. meCue Data Template. 2013. URL: http://mecue.de/Homepage%20Content/english/meCUE_data_template.xlsx [accessed 2023-04-12]
39. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34. [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
40. Cook JA, Sasor SE, Tholpady SS, Momeni A, Chu MW. Hand surgery resources exceed American health literacy. Hand (N Y) 2018 Sep;13(5):547-551. [doi: 10.1177/1558944717725384] [Medline: 28825342]
41. Goyal R, Mercado AE, Ring D, Crijns TJ. Most YouTube videos about carpal tunnel syndrome have the potential to reinforce misconceptions. Clin Orthop Relat Res 2021 Oct 1;479(10):2296-2302. [doi: 10.1097/CORR.0000000000001773] [Medline: 33847604]
42. Lutsky K, Bernstein J, Beredjiklian P. Quality of information on the internet about carpal tunnel syndrome: an update. Orthopedics 2013 Aug;36(8):e1038-e1041. [doi: 10.3928/01477447-20130724-20] [Medline: 23937750]
43. Steimle J, Gabriel S, Tarr R, Kohrs B, Johnston P, Martineau D. Comparing diagnostic and treatment recommendations of carpal tunnel syndrome available on the internet with AAOS clinical practice guidelines. Hand (N Y) 2020 Jul;15(4):514-520. [doi: 10.1177/1558944718821417] [Medline: 30654676]
44. Özbek EA, Armangil M, Karaca MO, Merter A, Dursun M, Kocaoğlu H. Evaluation of the reliability and quality of information in carpal tunnel syndrome shared on YouTube. J Wrist Surg 2022 Aug;11(4):295-301. [doi: 10.1055/s-0041-1735231] [Medline: 35971474]
45. Fang Y, Baker NA, Dole J, Roll SC. Quality of carpal tunnel syndrome patient education handouts available on the internet: a systematic analysis of content and design. Arch Phys Med Rehabil 2022 Feb;103(2):297-304. [doi: 10.1016/j.apmr.2021.08.010] [Medline: 34547274]
46. Frueh FS, Palma AF, Raptis DA, Graf CP, Giovanoli P, Calcagni M. Carpal tunnel syndrome: analysis of online patient information with the EQIP tool. Chir Main 2015 Jun;34(3):113-121. [doi: 10.1016/j.main.2015.04.003] [Medline: 26022522]
47. Frické M, Fallis D, Jones M, Luszko GM. Consumer health information on the internet about carpal tunnel syndrome: indicators of accuracy. Am J Med 2005 Feb;118(2):168-174. [doi: 10.1016/j.amjmed.2004.04.032] [Medline: 15694903]
48. Beredjiklian PK, Bozentka DJ, Steinberg DR, Bernstein J. Evaluating the source and content of orthopaedic information on the internet. The case of carpal tunnel syndrome. J Bone Joint Surg Am 2000 Nov;82(11):1540-1543. [doi: 10.2106/00004623-200011000-00004] [Medline: 11097441]
49. Kwak D, Park JW, Won Y, Kwon Y, Lee JI. Quality and reliability evaluation of online videos on carpal tunnel syndrome: a YouTube video-based study. BMJ Open 2022 Apr 15;12(4):e059239. [doi: 10.1136/bmjopen-2021-059239] [Medline: 35428646]
50. Johnson AR, Doval AF, Granoff MD, et al. A comparative multimetric assessment of English and Spanish carpal tunnel syndrome materials. J Surg Res 2019 Jun;238:64-71. [doi: 10.1016/j.jss.2019.01.032] [Medline: 30739070]
51. Eberlin KR, Vargas CR, Chuang DJ, Lee BT. Patient education for carpal tunnel syndrome: analysis of readability. Hand (N Y) 2015 Sep;10(3):374-380. [doi: 10.1007/s11552-014-9718-7] [Medline: 26330766]
52. Mert A, Bozgeyik B. Quality and content analysis of carpal tunnel videos on YouTube. Indian J Orthop 2022 Jan;56(1):73-78. [doi: 10.1007/s43465-021-00430-5] [Medline: 35070145]
53. Sproule JA, Tansey C, Burns B, Fenelon G. The web: friend or foe of the hand surgeon? Hand Surg 2003 Dec;8(2):181-185. [doi: 10.1142/s0218810403001716] [Medline: 15002095]
54. Ozdemir O, Diren F, Boyali O, Civelek E, Kabatas S. Metric evaluation of reliability and transparency of the videos about carpal tunnel syndrome surgery in the online platforms: assessment of YouTube videos' content. Neurospine 2021 Jun;18(2):363-368. [doi: 10.14245/ns.2142030.015] [Medline: 34218617]

XSL•FO
RenderX

55. Park KE, Allam O, Kim S, Prsic A. Accuracy of online pictorial resources for common hand surgery procedures. Hand (N Y) 2022 Mar;17(2):361-365. [doi: 10.1177/1558944720937360] [Medline: 32713204]
56. Akinleye SD, Garofolo-Gonzalez G, Montuori M, Culbertson MD, Hashem J, Edelstein DM. Readability of the most commonly accessed online patient education materials pertaining to pathology of the hand. Hand (N Y) 2018 Nov;13(6):705-714. [doi: 10.1177/1558944717726138] [Medline: 28836453]
57. Strouse S, Rosenberg S. Health literacy and discharge education. MSJ 2022;31(6):402 [FREE Full text] [doi: 10.62116/MSJ.2022.31.6.402]
58. Pekonen A, Eloranta S, Stolt M, Virolainen P, Leino-Kilpi H. Measuring patient empowerment - a systematic review. Patient Educ Couns 2020 Apr;103(4):777-787. [doi: 10.1016/j.pec.2019.10.019] [Medline: 31767243]
59. Osborne RH, Elsworth GR, Whitfield K. The Health Education Impact Questionnaire (heiQ): an outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient Educ Couns 2007 May;66(2):192-201. [doi: 10.1016/j.pec.2006.12.002] [Medline: 17320338]
60. Nagl M, Ullrich A, Farin E. Verständlichkeit von Patientenschulungen in der orthopädischen Rehabilitation: Qualitative Erhebung bei Rehabilitanden und Schulungsleitern. Rehabilitation (Bonn) 2012;52(1):34-39. [doi: 10.1055/s-0032-1312664]
61. Akbolat M, Amarat M, Ünal Ö, Şantaş G. A survey of health information seeking by cancer patients indicates some problems over medical explanations and terminology. Health Info Libraries J 2023 Mar;40(1):29-41. [doi: 10.1111/hir.12387]
62. Cosic F, Kimmel L, Edwards E. Patient comprehension of common orthopedic terminology. Health Lit Res Pract 2019 Jul;3(3):e187-e193. [doi: 10.3928/24748307-20190624-04] [Medline: 31428736]
63. Watermeyer J, Thwala Z, Beukes J. Medical terminology in intercultural health interactions. Health Commun 2021 Aug;36(9):1115-1124. [doi: 10.1080/10410236.2020.1735700] [Medline: 32202159]
64. Derevianchenko N, Lytovska O, Diurba D, Leshchyna I. Impact of medical terminology on patients' comprehension of healthcare. Georgian Med News 2018 Nov(284):159-163. [Medline: 30618411]
65. Finkelstein J, Bedra M. Is Internet search better than structured instruction for web-based health education? Stud Health Technol Inform 2013;190:65-67. [Medline: 23823377]
66. Castro H, Hise M, Finkelstein J. A comparison of two models of web-based education in older adults. AMIA Annu Symp Proc 2005;2005:914. [Medline: 16779201]
67. Crondahl K, Eklund Karlsson L. The nexus between health literacy and empowerment. SAGE Open 2016 Apr 1;6(2):215824401664641. [doi: 10.1177/2158244016646410]
68. Stars I. Health literacy as a challenge for health education. SHS Web Conf 2018;40:02004. [doi: 10.1051/shsconf/20184002004]
69. Gellert P, Tille F. What do we know so far? The role of health knowledge within theo-ries of health literacy. Eur Health Psychol 2015;16:266-274 [FREE Full text]
70. Sørensen K, Van den Broucke S, Fullam J, et al. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12(1):80. [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
71. Yim CK, Shumate L, Barnett SH, Leitman IM. Health literacy assessment and patient satisfaction in surgical practice. Ann Med Surg (Lond) 2018 Nov;35:25-28. [doi: 10.1016/j.amsu.2018.08.022] [Medline: 30263114]
72. Kandula NR, Nsiah-Kumi PA, Makoul G, et al. The relationship between health literacy and knowledge improvement after a multimedia type 2 diabetes education program. Patient Educ Couns 2009 Jun;75(3):321-327. [doi: 10.1016/j.pec.2009.04.001] [Medline: 19395223]
73. Davaris MT, Bunzli S, Trieu J, Dowsey MM, Choong PF. The role of digital health interventions to improve health literacy in surgical patients: a narrative review in arthroplasty. ANZ J Surg 2022 Oct;92(10):2474-2486. [doi: 10.1111/ans.17931] [Medline: 35924880]
74. Eltorai AEM, Ghanian S, Adams CA Jr, Born CT, Daniels AH. Readability of patient education materials on the american association for surgery of trauma website. Arch Trauma Res 2014 Jun;3(2):e18161. [doi: 10.5812/atr.18161] [Medline: 25147778]
75. Man A, van Ballegooie C. Assessment of the readability of web-based patient education material from major Canadian pediatric associations: cross-sectional study. JMIR Pediatr Parent 2022 Mar 16;5(1):e31820. [doi: 10.2196/31820] [Medline: 35293875]
76. Wasir AS, Volgman AS, Jolly M. Assessing readability and comprehension of web-based patient education materials by American Heart Association (AHA) and CardioSmart online platform by American College of Cardiology (ACC): How useful are these websites for patient understanding? Am Heart J Plus 2023 Aug;32:100308. [doi: 10.1016/j.ahjo.2023.100308] [Medline: 38510202]
77. Bühn S, Huppertz E, Weise A, et al. The effects of modifying elements of written informed consent forms for elective surgical or invasive procedures: a systematic review. Patient Educ Couns 2023 Feb;107:107576. [doi: 10.1016/j.pec.2022.107576] [Medline: 36455317]
78. Pennekamp P, Diedrich O, Schmitt O, Kraft C. Prävalenz und Stellenwert der Internetnutzung orthopädischer Patienten. Z Orthop Ihre Grenzgeb 2006 Jan;144(05):459-463. [doi: 10.1055/s-2006-942237]
79. Hertling S, Matziolis G, Graul I. Die Rolle des Internets als medizinische Informationsquelle für orthopädische Patienten [Article in German]. Orthopädie 2022;51(7):521-530. [doi: 10.1007/s00132-022-04238-5]

XSL•FO
RenderX

80. Mack EM, Callinan NJ, Reams M, Bohn DC, Chmielewski TL. Patient-reported outcomes after open carpal tunnel release using a standard protocol with 1 hand therapy visit. J Hand Ther 2017 Jan;30(1):58-64. [doi: 10.1016/j.jht.2016.03.007]
81. Kim JK, Koh YD, Kim JO, Choi SW. Changes in clinical symptoms, functions, and the median nerve cross-sectional area at the carpal tunnel inlet after open carpal tunnel release. Clin Orthop Surg 2016 Sep;8(3):298-302. [doi: 10.4055/cios.2016.8.3.298] [Medline: 27583113]
82. Aloi NF, Rahman H, Fowler JR. Changes in cross-sectional area of the median nerve and Boston carpal tunnel questionnaire scores after carpal tunnel release. Hand (N Y) 2024 Mar;19(2):212-216. [doi: 10.1177/15589447221127336] [Medline: 36189893]
83. Nielsen J. How Many Test Users in a Usability Study? NN/g Nielsen Norman Group. 2012. URL: https://www.nngroup.com/articles/how-many-test-users/ [accessed 2020-04-12]

Abbreviations

APPLY: patient's ability to apply health-related information in an everyday life
BCTQ: Boston carpal tunnel questionnaire
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
COHEP: Comprehensibility Of Health Education Programmes
COMM: patient's ability to communicate with health care professionals
COMPR: patient's comprehension of medical information
CTR: carpal tunnel release
CTS: carpal tunnel syndrome
HEL: health education literacy
HELP: Health Education Literacy of Patients with chronic musculoskeletal diseases
MANOVA: multivariate analysis of variances
PRISMA-ScR: Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews

Edited by A Mavragani; submitted 05.08.24; peer-reviewed by R Tedeschi, X Liu; revised version received 03.02.25; accepted 04.02.25; published 25.03.25.

Please cite as:
Seel M, Mihalic JA, Froschauer SM, Holzner B, Meier J, Gotterbarm T, Holzbauer M
Changes in Health Education Literacy After Structured Web-Based Education Versus Self-Directed Online Information Seeking in Patients Undergoing Carpal Tunnel Release Surgery: Nonrandomized, Controlled Study
JMIR Form Res 2025;9:e65114
URL: https://formative.jmir.org/2025/1/e65114
doi:10.2196/65114

© Mariella Seel, Julian Alexander Mihalic, Stefan Mathias Froschauer, Bernhard Holzner, Jens Meier, Tobias Gotterbarm, Matthias Holzbauer. Originally published in JMIR Formative Research (https://formative.jmir.org), 25.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Impact of ChatGPT Exposure on User Interactions With a Motivational Interviewing Chatbot: Quasi-Experimental Study

Jiading Zhu[1], BASc; Alec Dong[2*]; Cindy Wang[1*]; Scott Veldhuizen[3,4,5], PhD; Mohamed Abdelwahab[1], MSc; Andrew Brown[1], MASc; Peter Selby[3,5,6], MBBS, MHSc; Jonathan Rose[1,3], PhD

[1]The Edward S. Rogers Sr. Department of Electrical & Computer Engineering, University of Toronto, Toronto, ON, Canada

[2]Department of Mechanical & Industrial Engineering, University of Toronto, Toronto, ON, Canada

[3]INTREPID Lab, Centre for Addiction and Mental Health, Toronto, ON, Canada

[4]Institute for Mental Health Policy Research, Centre for Addiction and Mental Health, Toronto, ON, Canada

[5]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada

[6]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Jonathan Rose, PhD
The Edward S. Rogers Sr. Department of Electrical & Computer Engineering, University of Toronto, Toronto, ON, Canada

Abstract

Background: The worldwide introduction of ChatGPT in November 2022 may have changed how its users perceive and interact with other chatbots. This possibility may confound the comparison of responses to pre-ChatGPT and post-ChatGPT iterations of pre-existing chatbots, in turn affecting the direction of their evolution. Before the release of ChatGPT, we created a therapeutic chatbot, MIBot, whose goal is to use motivational interviewing to guide smokers toward making the decision to quit smoking. We were concerned that measurements going forward would not be comparable to those in the past, impacting the evaluation of future changes to the chatbot.

Objective: The aim of the study is to explore changes in how users interact with MIBot after the release of ChatGPT and examine the relationship between these changes and users' familiarity with ChatGPT.

Methods: We compared user interactions with MIBot prior to ChatGPT's release and 6 months after the release. Participants (N=143) were recruited through a web-based platform in November of 2022, prior to the release of ChatGPT, to converse with MIBot, in an experiment we refer to as MIBot (version 5.2). In May 2023, a set of (n=129) different participants were recruited to interact with the same version of MIBot and asked additional questions about their familiarity with ChatGPT, in the experiment called MIBot (version 5.2A). We used the Mann-Whitney *U* test to compare metrics between cohorts and Spearman rank correlation to assess relationships between familiarity with ChatGPT and other metrics within the MIBot (version 5.2A) cohort.

Results: In total, 83(64.3%) participants in the MIBot (version 5.2A) cohort had used ChatGPT, with 66 (51.2%) using it on a regular basis. Satisfaction with MIBot was significantly lower in the post-ChatGPT cohort (U=11,331.0; P=.001), driven by a decrease in perceived empathy as measured by the Average Consultation and Relational Empathy Measure (U=10,838.0; P=.01). Familiarity with ChatGPT was positively correlated with average response length (ρ=0.181; P=.04) and change in perceived importance of quitting smoking (ρ=0.296; P<.001).

Conclusions: The widespread reach of ChatGPT has changed how users interact with MIBot. Post-ChatGPT users are less satisfied with MIBot overall, particularly in terms of perceived empathy. However, users with greater familiarity with ChatGPT provide longer responses and demonstrated a greater increase in their perceived importance of quitting smoking after a session with MIBot. These findings suggest the need for chatbot developers to adapt to evolving user expectations in the era of advanced generative artificial intelligence.

(JMIR Form Res 2025;9:e56973) doi:10.2196/56973

KEYWORDS

chatbot; digital health; motivational interviewing; natural language processing; ChatGPT; large language models; artificial intelligence; experimental; smoking cessation; conversational agent

XSL•FO
RenderX

Introduction

Background

Generative chatbots are conversational systems that provide synthesized replies using deep learning techniques [1]. In recent years, generative chatbots based on large language models (LLMs) have made major advancements in their ability to engage in natural and human-like conversations [2]. ChatGPT, a popular LLM-based generative chatbot developed by OpenAI [3], has demonstrated significant potential to be applied in public health and medicine for a variety of purposes such as providing health information, supporting patient education, summarizing clinical notes, and assisting in administrative tasks [4,5]. However, these applications have not typically focused on therapeutic interactions, such as those using motivational interviewing (MI) [6] techniques.

Prior to ChatGPT's worldwide release in November 2022, our team developed MIBot [7], a therapeutic chatbot using MI techniques to guide smokers toward the decision to quit smoking. MIBot has a structured conversation flow, using scripted questions and LLM-generated MI-style reflections [6]. However, the popularity of ChatGPT and its advanced conversational capabilities raise questions about its influence on user interactions and expectations with simpler chatbots such as MIBot. These potential influences may in turn affect our future experiments with newer versions of MIBot, as the measurements used in future versions of MIBot may not be comparable to those in past versions, due to ChatGPT's potential influences on participants.

Related Work

Several recent studies have delved into the realm of generative chatbots, used in a medical or therapeutic context, and explored their effect on users. Perski et al [8] quantified the effect of the addition of a supportive chatbot to their smoking cessation application and found that it has increased user engagement and resulted in higher rates of quit success. Boucher et al [9] provided a comprehensive review of artificially intelligent chatbots in digital mental health interventions, discussing their acceptability, effects on user engagement and mental health outcomes, as well as their weaknesses and risks at the time, such as language proficiency and understanding. Moilanen et al [10] examined the effect of personality traits of a mental health chatbot on user engagement, finding that chatbots with a conscientious personality elicit the most user engagement and that users prefer their chatbot to be informative and confident rather than monotonic. Chow et al [11] framed ChatGPT as a disruptive technology and explored its impact on medical chatbots, claiming that it has the potential to improve access to health care services while expressing concerns over factors of risks such as reliability, transparency, privacy, and bias.

Together, these findings illustrate the growing influence of generative chatbots on user engagement and expectations in health-related contexts, particularly as advanced chatbots like ChatGPT redefine conversational norms. However, the impact of ChatGPT's widespread use on users' perceptions and behaviors toward other chatbots remains underexplored, raising questions about how exposure to such an advanced technology shifts user expectations and affects user behaviors. These findings served as important inspirations and motivations for this study, where we specifically investigate how users of one chatbot (MIBot) would change after being exposed to another chatbot (ChatGPT).

Study Objective

This study aims to investigate how user interactions with MIBot have changed following the release of ChatGPT and examines the relationship between familiarity with ChatGPT and these changes. By comparing dialogue sessions with MIBot conducted before and after ChatGPT's release, this study analyzes metrics such as average response length, satisfaction ratings, and effectiveness of the therapy. Additionally, we assess the relationship between participants' familiarity with ChatGPT and these metrics to better understand the impact of ChatGPT on user behaviors and expectations. Findings from this study are expected to provide insights into how user expectations have evolved since the release of ChatGPT and inform future development strategies for chatbots in health-related applications.

Methods

Experiments

Overview

An experiment was conducted on November 18, 2022, about 2 weeks before the worldwide introduction of ChatGPT by OpenAI [3]. We then repeated that experiment on May 16, 2023, about half a year after ChatGPT was released. The 2 experiments used the exact same version of MIBot [7] but on different groups of independently recruited participants. From this part on, we will refer to the first experiment as MIBot (version 5.2) and the second experiment as MIBot (version 5.2A).

Overall Experiment Flow

The overall experiment flow strictly followed what was detailed in the original MIBot paper [7]. Here is a brief overview:

1. Recruitment: Participants were recruited through the Prolific [12] paid web-based recruitment system after providing consent.
2. Preconversation surveys: Participants filled out surveys about their smoking habits on a custom website, including metrics used in the filtering process.
3. Conversation initiation: MIBot initiated with a text chat about smoking, continuing only with participant consent.
4. Core conversation: The main chatbot conversation about smoking cessation, which contained prescripted questions, responses from the user, and then generated reflections from the chatbot, described in more detail in the Design of MIBot section.
5. Postconversation: Participants completed another readiness-to-quit survey, the Consultation and Relational Empathy (CARE) Measure [13], and additional qualitative questions.
6. Reporting: Completion of tasks was recorded in the Prolific system [12].

XSL•FO
RenderX

7. One-week-later survey: A week later, participants answered a follow-up survey, also including readiness-to-quit, the completion of which, along with passing a manual data quality review, was required for the participants to receive their payments.

Design of MIBot

MIBot asked 2 types of questions. We call the first type "main questions" (Textbox 1) and the second type "yes or no questions" (Textbox 2). The main questions were open-ended questions that promoted self-reflection within users about their smoking habits, while yes or no questions were close-ended questions posed after each generated reflection to find out whether or not the generated reflection made sense, was on topic, and was used as a transition to the next main question.

Textbox 1. Main questions in MIBot conversations.

1. To start, what is the thing you like most about smoking?
2. What else do you like about smoking?
3. Now, what is the thing you like least about smoking?
4. What else do you dislike about smoking?
5. Now, what is one thing about your smoking that you would like to change?
6. What will it look like when you have made this change in your smoking addiction?
7. Finally, what are the steps you need to take to make this change?

Textbox 2. Yes or no questions in MIBot conversations.

1. Did that make sense?
2. Did what I said make sense to you?
3. Does this make sense to you?

The overall conversation structure was fixed as follows:

1. Introductory statement about MIBot and asking for permission to talk about the participant's smoking habits
2. Five to seven repetitions of
   a. Main question
   b. User response
   c. Generated reflection
   d. Yes or no question
   e. User response
3. Conclusion and thanking the participant for their time

Smoking Status

The context of our study requires us to keep track of the smoking status of the participants, measured by the following three metrics:

1. User status: A label that denotes if a participant would have passed the screening from their preconversation survey imported from our previous MIBot studies [7]. If they have passed the screening, they are denoted as low confidence or discordant; if not, they are instead denoted as high confidence and not discordant.
2. Heaviness of Smoking Index [14]: A validated survey metric calculated from cigarettes per day and time to the first cigarette of the day. Heaviness of Smoking Index is an integer, where a higher value indicates a heavier smoking habit.
3. Quit attempts made: A binary value that denotes if a participant has made at least 1 quit attempt in the week leading up to the MIBot conversation.

MIBot Surveys

There were 3 surveys delivered to the participants from the original MIBot experiment flow: preconversation survey, postconversation survey, and 1-week-later survey. These surveys were used to determine how effective MIBot was in motivating smokers to make the decision to quit smoking as well as how participants felt about the conversation with MIBot.

The relevant parts of the survey for this study are the following:

1. The CARE Measure [13] consists of questions regarding the participant's satisfaction with the conversation. Each question asks the participant to rate how well they think MIBot did on a scale from 1 to 5 or 0 if it does not apply. For example, participants were asked to evaluate MIBot's ability to "make you feel at ease" or "letting you tell your 'story'" back when they were interacting with MIBot. The CARE Measure is only included in the postconversation survey.
2. The Readiness Ruler [15] measures how ready the participant is to quit smoking. It consists of 3 ratings on a 0 - 10 scale: how confident they are about quitting smoking now, how important it is for them to quit smoking, and how ready they feel to quit smoking now. The Readiness Ruler is included in all 3 surveys.
3. Two additional feedback questions, included only in the postconversation survey:
   a. "What are 3 words that you would use to describe the chatbot?"
   b. "What would you change about the conversation?"

XSL•FO
RenderX

ChatGPT Survey

To determine the extent of exposure to ChatGPT, for each participant in MIBot (version 5.2A), we included an additional short survey in the 1-week-later survey referred to as the ChatGPT survey. It contained 8 new questions designed to evaluate the participant's knowledge and use of ChatGPT prior to engaging in MIBot (version 5.2A). The full ChatGPT survey can be found in Multimedia Appendix 1.

Recruitment and Data Inclusion

The steps taken for recruiting the participants for both MIBot (version 5.2) and MIBot (version 5.2A) were the exact same as detailed in the original MIBot paper [7]. Participants were recruited through the Prolific [12] web-based recruitment system, where they were informed by a recruitment description that they would engage in a text-based conversation with a chatbot designed to promote readiness to quit smoking, provide feedback on their experience, and complete the 1-week-later survey. The study was framed as an opportunity to contribute to research on chatbot-assisted smoking cessation, emphasized the confidentiality of participant data, and clearly stated that participation was voluntary, with the option to opt out if they did not agree to the terms. The entire study was delivered to the participants remotely, and they could participate in the study using their personal computers or mobile devices.

Notably, participants were screened using Prolific's filters based on the following inclusion criteria: participants could be located in any country, were at least 18 years of age, were fluent in English, had a smoking status defined as either a current smoker (smoking at least 5 cigarettes a day for at least 1 year) or a recent smoker (smoking at least 5 cigarettes a day for less than 1 year), and had a minimum approval rate of 90% on their prior Prolific studies. Additionally, Prolific was set to recruit an equal number of male and female participants. However, due to additional screening conducted after recruitment, the final sample was not perfectly balanced by sex.

The data collected were manually reviewed for data inclusion following the same criteria as listed in the MIBot paper [7] except for 1 major difference. In the original MIBot study, we only included participants with either of the following qualities in their Readiness Ruler scores:

- Low confidence: confidence level less than or equal to 5
- Discordant: importance level more than 5 points below the confidence level

As the focus of this study is instead about ChatGPT's influence on participants, we decided to include as many valid data entries as possible. Therefore, this filtering was not conducted in this study. As a result, the number of participants we included in this study increased from 100 to 143.

Evaluation Metrics

To evaluate MIBot conversations and user behaviors from various aspects, we designed several evaluation metrics, for which detailed definitions are provided below.

Response Length

In this study, the key aspect of user behavior we focused on was the length of their responses to questions prompted by MIBot. Response length is an important indicator of user engagement and willingness to interact with the chatbot. Longer responses suggest that participants are more actively reflecting on the conversation, which aligns with the therapeutic goals of MI sessions.

We defined response length by its word count, excluding any punctuation. We also categorized the questions MIBot asks into 2 categories and looked at the user responses separately to gain more insight. As a result, there are three metrics we applied to measure the lengths of participant responses:

1. Average response length: average length of responses to all questions.
2. Average response length (main): average length of responses to main questions, which are the scripted questions MIBot asks to provoke contemplation of the users' smoking habits.
3. Average response length (yes or no): average length of responses to yes or no questions, which are shown in Textbox 2.

Satisfaction With MIBot

Satisfaction with MIBot is a metric designed to measure user satisfaction after participating in a conversation with MIBot, a value between 0 and 1, calculated as the mean of the following three quantities:

1. Average CARE Measure (integers ranging from 1 to 5, in order of increasing satisfaction with MIBot), excluding answers where does not apply was selected. The mean is then rescaled to a number between 0 and 1 by subtracting 1 from the original value and dividing the result by 4.
2. Feedback sentiment score, based on answers to the first feedback question, "What are 3 words that you would use to describe the chatbot?" (integers ranging from 1 to 5, in order of increasing positivity). These scores were generated by automatically measuring the sentiment of the 3 words given by the user as feedback, described in the MIBot Survey section. The sentiment is computed using the neural network bert-base-multilingual-uncased-sentiment [16], which provides a score between 0 and 1.
3. MIBot improvement indicator, a binary value (0 or 1) if the participant indicated (in written feedback) that they would like MIBot improvement in response to the second feedback question, "What would you change about the conversation?" Responses are individually checked and manually labeled by a human collaborator. Since participants who suggested improvements are assumed to be less satisfied with MIBot, the value (1−MIBot improvement indicator) is used in the averaging calculation to get satisfaction with MIBot.

Effectiveness of MIBot

The goal of MIBot is to guide participants toward the decision to quit smoking by provoking contemplation. The more effective the conversation was, the more change we should observe in a participant. The effectiveness of MIBot is measured using 3

XSL•FO
RenderX

metrics related to the Readiness Ruler, as described in the MIBot Surveys section. The first of which, confidence change, is calculated as the difference between the confidence level before the conversation and the confidence level measured 1 week later, as recorded on the corresponding Readiness Rulers filled by the participants. Similarly, the importance change and readiness change were determined. As noted in the original MIBot paper [7], the primary outcome is the confidence change, as this is the most predictive of smoking cessation success [17-19].

Familiarity With ChatGPT

Familiarity with ChatGPT was given as a score between 0 and 126 (inclusive) to each participant based on their answers to the ChatGPT survey. Participants were first given a starting score of either 0 or 1: those who had not heard of ChatGPT were given a score of 0, and those who had were given a score of 1. Among those who had heard of ChatGPT, the score of those who had never used ChatGPT stayed at 1; for those who had used ChatGPT on top of knowing about it, they received an additional score according to the answers they had provided for 3 questions in the ChatGPT survey regarding their past use of ChatGPT. The three questions are as follows: (1) How long ago did you start using ChatGPT? (2) How frequently do you use ChatGPT? (3) On average, how long do your ChatGPT sessions last?

These 3 questions were presented as multiple-choice questions, the answers to which are presented in an ordered way and assigned a corresponding integer score from 1 to 5, as presented in Table 1. The final familiarity score they received was the product of the 3 integer scores, in addition to the 1 they already received. As a result, familiarity scores ranged from 0 to 126, with 0 being having never heard of ChatGPT, 1 being having heard of but never used ChatGPT, 2 being having used ChatGPT once a month or less for less than 5 minutes each session starting from less than a week ago, and 126 being having used ChatGPT multiple times per day with each session lasting longer than an hour for the past 5 months.

Table . Corresponding description for familiarity score assignment.

Corresponding score	Duration of use	Frequency of use	Length of each session
1	Less than a week	Once a month or less	Less than 5 minutes
2	Less than a month	Every 2 weeks	Less than 10 minutes
3	1 - 3 months	Every week	10 to 30 minutes
4	3 - 5 months	Every day	30 minutes to 1 hour
5	Over 5 months	Multiple times per day	More than 1 hour

Statistical Analysis

Comparison of Pre- and Post-ChatGPT Cohorts

To capture the differences in behavior between participants in the 2 cohorts—November 2022 (MI version 5.2) and May 2023 (MI version 5.2A)—we used the Mann-Whitney *U* test to compare various metrics, such as average response length, satisfaction with MIBot, and changes in Readiness Ruler scores. This nonparametric test provided test statistics and *P* values for each metric, which are used to determine significant differences between the 2 groups. The test was done using the Python programming language (Python Software Foundation) with the pandas [20] and SciPy [21] libraries.

Relationships Between Familiarity With ChatGPT and Other Variables

To examine the impact of familiarity with ChatGPT within the MIBot (version 5.2A) cohort, we performed Spearman rank correlation analysis. This test assessed the relationships between participants' familiarity scores (ranging from 0 to 126) and key metrics, such as average response length, satisfaction with MIBot, and changes in Readiness Ruler scores. Spearman rank correlation provided correlation coefficients and *P* values, indicating the strength and significance of each relationship. The test was also done using Python with the pandas [20] and SciPy [21] libraries.

Ethical Considerations

This research was approved by the University of Toronto Research Ethics Board (protocol # 35567), as amended on June 29, 2022. Participants voluntarily provided informed consent by agreeing to all the terms stated in a consent form presented digitally to all participants during recruitment, fully outlining the study's goals, procedures, potential risks, and privacy guarantees. It specified that no personally identifiable information would be collected, and any inadvertent identifiers would be removed prior to analysis or publication. Participants received a total of US $6.25 for MIBot (version 5.2) and US $6.18 for MIBot (version 5.2A) as compensation for completing all tasks in the study, and partial compensation was not provided to those who did not complete all components of the experiment. No identifiable participant information is included in any images or supplementary materials in the manuscript.

Results

Participant Demographic

Participant demographic data of both MIBot (version 5.2) and MIBot (version 5.2A) are listed in Tables 2 and 3, including categorical and continuous variables, respectively. Additional data of the demographics of the participants in both MIBot (version 5.2) and MIBot (version 5.2A) are listed in Multimedia Appendix 2.

XSL•FO
RenderX

Table . Categorical demographic data of participants in cohort MIBot (version 5.2; November 2022, pre-ChatGPT) and MIBot (version 5.2A; May 2023, post-ChatGPT).

Characteristic		MIBot (version 5.2) (n=143), n (%)	MIBot (version 5.2A) (n=129), n (%)
Sex			
	Male	72 (50.3)	63 (48.8)
	Female	71 (49.7)	66 (51.2)
User status			
	Low confidence or discordant	100 (69.9)	91 (70.5)
	High confidence and not discordant	43 (30.1)	38 (29.5)
Smoking status			
	I am a current smoker (smoke at least 5 cigarettes a day and have smoked this amount for at least 1 year)	121 (84.6)	107 (82.9)
	I am a recent smoker (smoke at least 5 cigarettes a day and have smoked this amount for less than 1 year)	22 (15.4)	22 (17.1)
Quit attempts made			
	Yes	69 (48.3)	46 (35.7)
	No	74 (51.7)	83 (64.3)

Table . Continuous demographic data of participants in cohort MIBot (version 5.2; November 2022, pre-ChatGPT) and MIBot (version 5.2A; May 2023, post-ChatGPT).

Characteristic	MIBot (version 5.2)		MIBot (version 5.2A)	
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)
Age (years)	29.22 (9.69)	26.00 (23.00 - 33.00)	32.75 (11.36)	29.00 (24.00 - 39.00)
HSI[a]	1.45 (1.38)	1.00 (0.00 - 2.00)	1.67 (1.45)	1.00 (0.00 - 3.00)
Readiness Rulers, preconversation				
Confidence	4.52 (2.75)	4.00 (2.00 - 7.00)	4.12 (2.60)	4.00 (2.00 - 6.00)
Importance	5.84 (2.84)	6.00 (3.00 - 8.00)	5.60 (2.65)	6.00 (3.00 - 8.00)
Readiness	5.36 (2.83)	5.00 (3.00 - 8.00)	4.79 (2.58)	5.00 (3.00 - 7.00)

[a]HSI: Heaviness of Smoking Index.

Reach of ChatGPT

The familiarity score, described in the Familiarity With ChatGPT section, provides an estimate of the reach of ChatGPT to the public. Table 4 provides a more detailed description of the distribution of the familiarity score, where we bin the participants based on familiarity score ranges and assign labels to each score range. “Unexposed” is the group of participants who have not heard of ChatGPT prior to this study. “Aware” denotes that the participants have heard of ChatGPT but did not use it in any capacity. People who have used ChatGPT are separated into 3 groups, which are “casual user,” “consistent user,” and “dedicated user,” based on their familiarity score.

Table . Distribution of ChatGPT familiarity levels among participants in the MIBot (version 5.2A) cohort.

ChatGPT familiarity level	Count, n (%)	Score
Unexposed	12 (9.3)	0
Aware	34 (26.4)	1
Casual user	17 (13.2)	2 to 8
Consistent user	37 (28.7)	9 to 27
Dedicated user	29 (22.5)	28 or higher

XSL•FO
RenderX

Comparison of Pre- and Post-ChatGPT Cohorts

Table 5 shows the Mann-Whitney *U* test results that compare between participants from the November 2022 (MI version 5.2) and May 2023 (MI version 5.2A) cohorts. In terms of average response length, there was no significant difference between the 2 cohorts, whether looking at all responses together or breaking them down into responses to main questions and responses to yes or no questions. Satisfaction with MIBot was significantly higher in MIBot (version 5.2) than in MIBot (version 5.2A; *U*=11,331.0; *P*=.001). Specifically, of the 3 metrics satisfaction with MIBot was aggregated of, Average CARE Measure was the only one showing any significant difference, with it being significantly higher in MIBot (version 5.2) compared to MIBot (version 5.2A; *U*=10,838; *P*=.01). As for changes in the Readiness Ruler, no significant differences were found between the cohorts in terms of confidence change, importance change, or readiness change.

Table . Results of the Mann-Whitney *U* test comparing key metrics between cohort MIBot (version 5.2; pre-ChatGPT) and MIBot (version 5.2A; post-ChatGPT).

Variable		MIBot (version 5.2)		MIBot (version 5.2A)		Test statistic (*U*)	*P* value
		Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)		
Response length							
	Average response length	4.98 (2.49)	4.46 (3.31 - 5.89)	5.36 (3.54)	4.29 (3.13 - 7.00)	9248.0	.97
	Average response length (main)	8.18 (4.00)	7.60 (5.50 - 9.69)	8.04 (4.71)	6.62 (4.86 - 10.43)	9936.5	.27
	Average response length (yes or no)	2.02 (1.91)	1.00 (1.00 - 2.07)	2.44 (3.04)	1.17 (1.00 - 2.80)	8301.5	.13
Satisfaction scores							
	Satisfaction with MIBot	0.74 (0.22)	0.85 (0.52 - 0.91)	0.66 (0.21)	0.61 (0.49 - 0.85)	11,331.0	.001
	Average CARE[a] Measure	0.77 (0.16)	0.78 (0.66 - 0.90)	0.72 (0.17)	0.74 (0.58 - 0.84)	10,838.0	.01
	MIBot improvement indicator	0.57 (1.01)	0.00 (0.00 - 1.00)	0.76 (1.12)	0.00 (0.00 - 1.00)	8192.5	.06
	Feedback sentiment score	0.79 (0.15)	0.80 (0.80 - 0.80)	0.76 (0.18)	0.80 (0.80 - 0.80)	10,028.0	.15
Readiness Ruler							
	Confidence change	0.86 (2.32)	1.00 (0.00 - 2.00)	0.60 (2.08)	0.00 (0.00 - 2.00)	9893.0	.29
	Importance change	0.69 (1.85)	0.00 (0.00 - 2.00)	0.40 (1.98)	0.00 (0.00 - 1.00)	10,003.5	.21
	Readiness change	0.35 (1.82)	0.00 (−1.00 to 1.00)	0.33 (2.07)	0.00 (−1.00 to 1.00)	9260.0	.95

[a]CARE: Consultation and Relational Empathy.

Relationships Between Familiarity With ChatGPT and Other Variables

Table 6 shows the Spearman rank correlation results evaluating relationships between familiarity with ChatGPT and various other variables within the MIBot (version 5.2A) cohort. In terms of average response length, there was a significant positive correlation between familiarity with ChatGPT and overall average response length (ρ=0.181; *P*=.04) as well as with average response length for main questions (ρ=0.180; *P*=.04) and average response length for yes or no questions (ρ=0.197; *P*=.03). Satisfaction with MIBot had a weak negative correlation with familiarity score, but it was not statistically significant (ρ=−0.171; *P*=.05). Of the 3 metrics aggregated into satisfaction with MIBot, only the MIBot improvement indicator showed a significant positive correlation with familiarity with ChatGPT (ρ=0.188; *P*=.03). Among the Readiness Ruler changes, a significant positive correlation was observed between familiarity with ChatGPT and importance change (ρ=0.296; *P*<.001) but not for confidence change or readiness change.

XSL•FO
RenderX

Table . Spearman rank correlation analysis for the MIBot (version 5.2A) cohort.

Variable		Familiarity with ChatGPT	
		ρ	*P* value
Response length			
	Average response length	0.181	.04
	Average response length (main)	0.180	.04
	Average response length (yes or no)	0.197	.03
Satisfaction scores			
	Satisfaction with MIBot	−0.171	.05
	Average CARE[a] Measure	−0.112	.21
	MIBot improvement indicator	0.188	.03
	Feedback sentiment score	−0.023	.80
Readiness Ruler			
	Confidence change	0.165	.06
	Importance change	0.296	<.001
	Readiness change	0.078	.38

[a]CARE: Consultation and Relational Empathy.

Discussion

Principal Findings

This study aimed to investigate how user interactions with MIBot changed following the release of ChatGPT as well as the relationship between ChatGPT familiarity and these changes. From the results, we found that users interacting with MIBot after the release of ChatGPT were less satisfied with the chatbot, particularly in terms of perceived empathy, but users more familiar with ChatGPT provided longer responses and showed greater increase in their perceived importance of quitting smoking.

The premise of this study is that people have been exposed to ChatGPT due to its high popularity, which is supported by the distribution of familiarity scores in the MIBot (version 5.2A) cohort. From Table 4, ChatGPT is observed to have a widespread reach, with only 12 (9.3%) participants of MIBot (version 5.2A) having never heard of it, and 34 (26.4%) having only learned about it but never used it. Collectively, 83 (64.3%) participants have used ChatGPT at least to some capacity, most of whom are consistent or even dedicated users, indicating that the public has received massive exposure to the popular chatbot.

Using the Mann-Whitney *U* test (Table 5) to compare participants from the MIBot (version 5.2) and MIBot (version 5.2A) cohorts, we observed that most variables did not show significant differences between the groups. However, satisfaction with MIBot was significantly lower in the MIBot (version 5.2A) cohort (*U*=11,331.0; *P*=.001), a finding primarily driven by differences in the Average CARE Measure. This specific component of the satisfaction score, which reflects perceived empathy in interactions, was significantly higher in the MIBot (version 5.2) cohort (*U*=10,838.0; *P*=.01). These findings suggest that participants in the post-ChatGPT cohort, many of whom were familiar with or at least aware of ChatGPT's advanced conversational capabilities, may have developed higher expectations for chatbot empathy, which MIBot may not fully meet.

Within the MIBot (version 5.2A) cohort, Spearman rank correlation analysis revealed that familiarity with ChatGPT was positively correlated with average response length across all response types, indicating that more experienced users were likely to engage in longer interactions. However, while users who are more familiar with ChatGPT are more likely to suggest improvements for MIBot (ρ=0.188; *P*=.03), satisfaction with MIBot as a whole had only a weak, nonsignificant negative correlation with familiarity (ρ=−0.171; *P*=.05), even though the *U* test results showed that satisfaction was significantly lower in the MIBot (version 5.2A) cohort. This discrepancy suggests that while ChatGPT exposure may influence satisfaction overall, familiarity alone does not predict satisfaction levels. It may be that other factors unique to the MIBot (version 5.2A) cohort, such as general exposure to more advanced conversational artificial intelligence or broader trends in technology expectations, are impacting satisfaction with MIBot in a way that the specific familiarity score we designed failed to capture. In other words, the drop in satisfaction with MIBot might reflect a general shift in expectations for chatbot performance rather than an effect specific to individual familiarity with ChatGPT.

Interestingly, although familiarity with ChatGPT appears to make users more critical of MIBot's capabilities, it is also associated with an increase in participants' perceived importance of quitting smoking (ρ=0.296; *P*<.001), a key measure of MIBot's effectiveness. This seemingly paradoxical finding could suggest that while exposure to advanced generative chatbots raises expectations for technical quality, it might also enhance users' receptiveness to the broader goals of MIBot, such as helping users resolve ambivalence and guiding them

XSL•FO
RenderX

toward change, and as a result, making a less advanced chatbot more effective.

Practical Implications

The findings from this study suggest several practical implications for future chatbot development, especially in health-related applications. The decrease in satisfaction with MIBot from post-ChatGPT users highlights a need for researchers and developers to adapt to higher expectations of chatbots, especially for empathy. Additionally, the tendency of users familiar with ChatGPT to provide longer, more detailed responses suggests that chatbots should be capable of handling more nuanced interactions. Finally, the positive correlation between ChatGPT familiarity and MIBot's effectiveness in increasing the perceived importance of quitting smoking implies that users more experienced with advanced generative chatbots may have more trust in the technology and be more receptive to interactions guiding themselves toward change.

Comparison With Prior Works

Unlike prior studies that examined and reviewed various aspects of chatbots that may affect user engagement and effectiveness in different contexts [8-10], this study specifically and uniquely investigates how exposure to and familiarity with a more advanced chatbot (ChatGPT) affect user interactions with a simpler chatbot (MIBot). Building on to our previous work, MIBot [7], our new experiments suggest that MIBot continues to increase users' readiness to quit even after the introduction of ChatGPT. Nevertheless, overall user satisfaction has decreased, possibly affected by higher user expectations among those exposed to ChatGPT. This could be a disruptive effect from ChatGPT as predicted by Chow et al [11]. Moreover, our findings also provide unique insights not found in related prior works, such as that familiarity with ChatGPT positively correlates with core metrics for user engagement with and effectiveness of MIBot, a simpler chatbot than ChatGPT. It also presented a measurement, at a specific point in time, of the exposure level to ChatGPT among our recruited participants, which demonstrates the extent of ChatGPT's reach at that time.

Limitations

There are several limitations regarding this study. Since we followed the same experiment flow as the original MIBot paper [7] and used mostly the same evaluation metrics, the limitations of the MIBot experiments documented in the original MIBot paper [7] also apply to this study. Moreover, there exists an additional type of sampling bias specific to this study. Since for both of our experiments, we recruited participants through Prolific [12], a web-based recruitment system, it is possible that these participants are more fluent with technology than the general public and therefore more likely to have been exposed to ChatGPT or have the effects of ChatGPT manifest in a way that does not fairly represent a more general population. Furthermore, our 2 experiments used different groups of people with some demographic differences, notably with a difference in their average age, which may result in different smoking behaviors and success rates of smoking cessation [22,23]. Finally, it is important to acknowledge the possibility of confounding factors not accounted for in our study, which may have influenced how participants interacted with MIBot or perceived its effectiveness, along with familiarity with ChatGPT.

Conclusions

Given MIBot, a generative chatbot for smoking cessation less advanced than ChatGPT, this study aimed to find any potential changes to how users interact with MIBot by comparing user interactions before and after ChatGPT's release. We found that post-ChatGPT users are less satisfied with MIBot overall, particularly in terms of perceived empathy. Moreover, as users gain more familiarity with ChatGPT, they provide longer responses and show a greater increase in their perceived importance of quitting smoking. These findings suggest the need for continuous innovation in chatbot technology, particularly in digital health, to meet the evolving expectations of users accustomed to more advanced chatbot interactions.

Acknowledgments

The authors acknowledge their long-standing collaboration with the Intrepid Lab of Dr Peter Selby at the Centre for Addiction and Mental Health in Toronto and Professor Matt Ratto in the Faculty of Information at the University of Toronto. This research was funded by a Natural Sciences and Engineering Research Council of Canada Discovery grant (RGPIN-2019-04395) and a New Frontiers in Research Fund grant (NFRFE-2021-00449).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

JZ and JR conceptualized this research work and oversaw project administration in terms of managing, planning, and coordinating the experiments. JZ, AD, CW, MA, and AB contributed to data curation and to developing the chatbot and the scripts. JZ, AD, CW, and SV contributed to conducting all the formal analyses. PS and JR secured funding for the project and provided overall supervision of all the research activities. JZ, AD, CW, and MA did all the investigation work related to the study. JZ, SV, AB, and JR contributed to developing the methodology. JR provided the computing and experimental resources for the study. JZ, AD, and CW contributed to the validation and visualization of the data collected. JZ, AD, CW, and JR were the writers of the original

XSL•FO
RenderX

draft, while JZ and JR critically reviewed and edited the manuscript. All authors made significant contributions to this work and approved it for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questions related to ChatGPT on the week-later survey.
[DOCX File, 511 KB - formative_v9i1e56973_app1.docx]

Multimedia Appendix 2
Additional demographics data.
[DOCX File, 20 KB - formative_v9i1e56973_app2.docx]

References

1. Adamopoulou E, Moussiades L. Chatbots: history, technology, and applications. Mach Learn Appl 2020 Dec;2:100006. [doi: 10.1016/j.mlwa.2020.100006]
2. Shanahan M, McDonell K, Reynolds L. Role play with large language models. Nature New Biol 2023 Nov;623(7987):493-498. [doi: 10.1038/s41586-023-06647-8] [Medline: 37938776]
3. Introducing ChatGPT. OpenAI. 2022. URL: https://openai.com/blog/chatgpt/ [accessed 2023-12-19]
4. Biswas SS. Role of Chat GPT in public health. Ann Biomed Eng 2023 May;51(5):868-869. [doi: 10.1007/s10439-023-03172-7] [Medline: 36920578]
5. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med 2023 Aug;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
6. Miller W, Rollnick S. Motivational Interviewing: Helping People Change and Grow: The Guilford Press; 2023.
7. Brown A, Kumar AT, Melamed O, et al. A motivational interviewing chatbot with generative reflections for increasing readiness to quit smoking: iterative development study. JMIR Ment Health 2023 Oct 17;10:e49132. [doi: 10.2196/49132] [Medline: 37847539]
8. Perski O, Crane D, Beard E, Brown J. Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study [published correction appears in Digit Health. 2020 May 22;6:2055207620930958]. Digit Health 2019 Jan;5:2055207619880676. [doi: 10.1177/2055207619880676]
9. Boucher EM, Harake NR, Ward HE, et al. Artificially intelligent chatbots in digital mental health interventions: a review. Expert Rev Med Devices 2021 Dec;18(sup1):37-49. [doi: 10.1080/17434440.2021.2013200] [Medline: 34872429]
10. Moilanen J, Visuri A, Suryanarayana SA, Alorwu A, Yatani K, Hosio S. Measuring the effect of mental health chatbot personality on user engagement. Presented at: MUM 2022: 21th International Conference on Mobile and Ubiquitous Multimedia; Nov 27-30, 2022; Lisbon, Portugal. [doi: 10.1145/3568444.3568464]
11. Chow JCL, Sanders L, Li K. Impact of ChatGPT on medical chatbots as a disruptive technology. Front Artif Intell 2023;6. [doi: 10.3389/frai.2023.1166014]
12. Palan S, Schitter C. Prolific.ac—a subject pool for online experiments. J Behav Exp Finance 2018 Mar;17:22-27. [doi: 10.1016/j.jbef.2017.12.004]
13. Mercer SW, Maxwell M, Heaney D, Watt GC. The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. Fam Pract 2004 Dec;21(6):699-705. [doi: 10.1093/fampra/cmh621.cmh621]
14. Heatherton TF, Kozlowski LT, Frecker RC, Rickert W, Robinson J. Measuring the heaviness of smoking: using self-reported time to the first cigarette of the day and number of cigarettes smoked per day. Br J Addict 1989 Jul;84(7):791-799. [doi: 10.1111/j.1360-0443.1989.tb03059.x] [Medline: 2758152]
15. Readiness Ruler. Center for Evidence-Based Practices | Case Western Reserve University. 2021. URL: https://case.edu/socialwork/centerforebp/resources/readiness-ruler [accessed 2023-12-19]
16. Peirsman Y. bert-base-multilingual-uncased-sentiment. Hugging Face. 2020. URL: https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment [accessed 2020-12-19]
17. Smit ES, Hoving C, Schelleman-Offermans K, West R, de Vries H. Predictors of successful and unsuccessful quit attempts among smokers motivated to quit. Addict Behav 2014 Sep;39(9):1318-1324. [doi: 10.1016/j.addbeh.2014.04.017] [Medline: 24837754]
18. Gwaltney CJ, Metrik J, Kahler CW, Shiffman S. Self-efficacy and smoking cessation: a meta-analysis. Psychol Addict Behav 2009 Mar;23(1):56-66. [doi: 10.1037/a0013529] [Medline: 19290690]
19. Von Ah D, Ebert S, Ngamvitroj A, Park N, Kang DH. Factors related to cigarette smoking initiation and use among college students. Tob Induc Dis 2005 Dec 15;3(1):27-40. [doi: 10.1186/1617-9625-3-1-27] [Medline: 19570288]

XSL•FO
RenderX

20. pandas—Python Data Analysis Library homepage pandas—Python Data Analysis Library. URL: https://pandas.pydata.org/ [accessed 2024-11-17]
21. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020 Mar 2;17(3):261-272. [doi: 10.1038/s41592-019-0686-2]
22. Townsend J, Roderick P, Cooper J. Cigarette smoking by socioeconomic group, sex, and age: effects of price, income, and health publicity. BMJ 1994 Oct 8;309(6959):923-927. [doi: 10.1136/bmj.309.6959.923] [Medline: 7950662]
23. Fidler J, Ferguson SG, Brown J, Stapleton J, West R. How does rate of smoking cessation vary by age, gender and social grade? Findings from a population survey in England. Addiction 2013 Sep;108(9):1680-1685. [doi: 10.1111/add.12241] [Medline: 23668684]

Abbreviations

CARE: Consultation and Relational Empathy
LLM: large language model
MI: motivational interviewing

Edited by A Mavragani; submitted 02.02.24; peer-reviewed by A Carcone, I Balan, J Wilson; revised version received 20.02.25; accepted 21.02.25; published 21.03.25.

Please cite as:
Zhu J, Dong A, Wang C, Veldhuizen S, Abdelwahab M, Brown A, Selby P, Rose J
The Impact of ChatGPT Exposure on User Interactions With a Motivational Interviewing Chatbot: Quasi-Experimental Study
JMIR Form Res 2025;9:e56973
URL: https://formative.jmir.org/2025/1/e56973
doi:10.2196/56973

© Jiading Zhu, Alec Dong, Cindy Wang, Scott Veldhuizen, Mohamed Abdelwahab, Andrew Brown, Peter Selby, Jonathan Rose. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Brief Cognitive Behavioral Therapy–Based Digital Intervention for Reducing Hazardous Alcohol Use in South Korea: Development and Prospective Pilot Study

Manjae Kwon[1,2*], MD; Daa Un Moon[3*], MD; Minjae Kang[1,2], MD; Young-Chul Jung[1,2], MD, PhD

[1]Department of Psychiatry, Yonsei University College of Medicine, Seoul, Republic of Korea

[2]Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

[3]Department of Psychiatry and Neurosciences, Charité Campus Mitte, Charité Universitätsmedizin Berlin, Corporate Member of Freie Universität and Humboldt-Universität zu Berlin, Berlin, Germany

[*]these authors contributed equally

Corresponding Author:
Young-Chul Jung, MD, PhD
Department of Psychiatry, Yonsei University College of Medicine, Seoul, Republic of Korea

Abstract

Background: Alcohol consumption is a leading cause of death and disability worldwide, associated with numerous acute and chronic medical conditions. Digital health interventions offer a promising solution to overcome barriers associated with traditional treatment methods, providing accessible, scalable, and cost-effective means to support individuals in reducing hazardous drinking.

Objective: This pilot study aims to evaluate the feasibility, acceptability, and preliminary efficacy of the Sober smartphone app in individuals with hazardous alcohol use.

Methods: This single-group, pre- and postpilot study included 20 participants with risky alcohol use, identified using the Alcohol Use Disorder Identification Test. Participants used the Sober app for 4 weeks, incorporating cognitive behavioral therapy–based interventions. Feasibility was assessed by study and session completion rates, acceptability by participant satisfaction and perceived usefulness, and preliminary efficacy by changes in alcohol consumption and psychiatric symptoms. Semistructured interviews with participants and clinicians provided qualitative perspectives on the app's usability, efficacy, and areas for improvement.

Results: Of the 20 enrolled participants, 17 completed the study. The app demonstrated high feasibility with an 85% (17/20) study completion rate, and 59% (10/17) completed all cognitive behavioral therapy sessions. Participants reported positive acceptability, with average satisfaction and usefulness ratings of 3.8 and 3.7 of 5, respectively. Preliminary efficacy outcomes showed significant improvements: abstinence days increased from 67% to 85% (z=−3.17; P=.002), heavy drinking episodes decreased from 3.3 to 1.9 (t_{16}=−2.97; P=.003), and total alcohol consumption reduced from 456.8 to 195.9 mL (t_{16}=3.16; P=.002). Alcohol Use Disorder Identification Test scores dropped from 17.5 to 10.7 (t_{16}=4.51; P<.001). Additionally, depression (Patient Health Questionnaire-9) scores decreased from 5.8 to 4.4 (t_{16}=2.91; P=.01), and anxiety (Generalized Anxiety Disorder-7) scores from 3.4 to 2.1 (z=−2.80; P=.005). No adverse events were reported. Qualitative analysis found participants valued daily logging but noted usability issues, while clinicians called for tailored goals, enhanced communication features, and age-specific content.

Conclusions: The mobile app Sober shows promise as an effective tool for reducing hazardous alcohol consumption and improving related psychiatric symptoms. The study demonstrated high feasibility and positive acceptability, with significant preliminary efficacy in reducing alcohol use. Qualitative findings provided actionable evidence for refining the app's usability and clinical integration. Further research through a randomized controlled trial is warranted to confirm these findings and optimize the app's features and content.

Trial Registration: ClinicalTrials.gov NCT06502756; https://clinicaltrials.gov/study/NCT06502756

(JMIR Form Res 2025;9:e64459) doi:10.2196/64459

KEYWORDS

alcohol; hazardous alcohol use; digital intervention; cognitive behavioral therapy; mobile apps; prevention; therapy-based; cognitive behavioral; alcohol use; South Korea; prospective pilot study; pilot study; alcohol consumption; death; disability; chronic medical condition; digital health interventions; traditional treatment methods; Korean; hazardous drinking; acceptability; feasibility; smartphone app; alcohol use disorder; psychiatric symptoms; mobile phone

XSL•FO
RenderX

Introduction

Alcohol is the most widely used psychoactive substance worldwide and poses a significant public health problem with a massive socioeconomic burden [1,2]. It is a leading cause of death and disability globally, contributing to a wide range of acute and chronic medical conditions. Acute medical consequences of high alcohol intake include injuries, car accidents, and violence [3]. Chronic diseases associated with alcohol consumption include liver cirrhosis, cardiovascular diseases, and various cancers [4-8]. Additionally, alcohol use disorders are linked to numerous psychiatric comorbidities, such as mood disorders, anxiety disorders, personality disorders, and suicide [9,10].

Hazardous drinking, defined as a drinking behavior that significantly increases the risk of adverse alcohol-related outcomes [11], is prevalent among those who consume alcohol, with a global prevalence of 17% and 15.1% in South Korea [2,12]. Effective prevention typically includes screening and brief interventions to identify and address hazardous drinking early [13-15]. These interventions involve trained staff using a brief screening tool combined with motivational interviewing techniques to provide personalized feedback and clinical options tailored to the individual's level of risk, including a referral to treatment component, if necessary [11,16-18]. However, several barriers hinder the implementation of these interventions, including concerns about stigmatization, lack of resources, insufficient training, high clinician workload, and logistical and financial obstacles for patients seeking help [19-22].

Digital health interventions offer a promising solution to overcome this barrier, especially in countries like South Korea, which has the highest smartphone ownership rate in the world at approximately 95%, with minimal generational gaps in ownership [23]. With near-ubiquitous smartphone ownership, these interventions can improve health outcomes by providing inexpensive, easily accessible, and scalable methods to deliver education, support, and monitoring through personal devices [24]. Currently, many apps in commercial app stores aim to help users reduce their alcohol consumption, but few have been rigorously evaluated, and only a minority have published efficacy information [25]. While the overall evidence for alcohol reduction through digital interventions is promising, it remains inconclusive [25-31]. Moreover, reporting theoretical foundations in these apps is often limited and unclear, indicating a need for more rigorous evaluation and transparency in their development [31].

Hazardous alcohol use frequently co-occurs with psychological distress, including symptoms of depression and anxiety [32]. These symptoms often create a cyclical relationship with problematic alcohol use, where they exacerbate drinking patterns and complicate treatment outcomes [33,34]. Research shows that brief interventions targeting hazardous alcohol use can reduce consumption and alleviate psychological distress, depression, and anxiety, addressing both alcohol use and co-occurring mental health symptoms [35,36]. Studies also highlight the potential of digital interventions to achieve similar dual benefits, providing scalable solutions to this complex challenge [37,38].

This pilot study aims to estimate the potential of a digital intervention using the Sober smartphone app (WELT Corp Ltd) for secondary prevention in individuals with hazardous alcohol use. The primary objective is to evaluate the app's feasibility, acceptability, and preliminary efficacy in reducing hazardous drinking behaviors. In addition, the study examines secondary outcomes, including the impact of the intervention on symptoms such as depression, anxiety, and stress. The findings from this pilot study will be instrumental in refining the app's features and content and optimizing the study design for a planned randomized controlled trial (RCT).

Methods

Design and Setting

This single-group, pre- and postpilot study assessed the preliminary efficacy, feasibility, and acceptability of the mobile app Sober for reducing hazardous alcohol consumption. The single-group design was chosen to provide initial insights into the intervention's effects and to inform the design of future RCTs.

Ethical Considerations

The study was approved by the institutional review board of Yonsei Severance Hospital (approval 4-2023-0338) and adhered to the ethical standards of the Declaration of Helsinki. Prior to study participation, all participants provided written informed consent, including consent for the collection and analysis of their data. Participants were informed of their right to withdraw from the study at any time. No modifications to the content or methodology were made after the study commenced. We anonymized the participants using unique ID numbers. No identifying or potentially identifying information is included in this paper or Multimedia Appendices 1 and 2. All identifying or potentially identifying data are securely maintained by Yonsei Severance Hospital and accessible only to the authorized research team. The participants received approximately US $34 for each clinic visit for their time and travel expenses.

Recruitment

Participants were recruited through outpatient clinic bulletin boards, web-based advertisements (eg, university websites), flyers, and word of mouth. Recruitment targeted individuals who self-identified as having concerns about their drinking habits and sought mental health support. The study was conducted at Yonsei Severance Hospital, a tertiary care teaching hospital in Seoul, South Korea. Interested individuals received detailed information about the study and underwent an in-person screening process after providing written informed consent. Board-certified psychiatrists (MK, DUM, and YCJ) reviewed each participant's medical and medication history to determine eligibility following a psychiatric assessment [39], including a review of medical and medication history.

Eligibility Criteria

Participants aged 19 - 65 years who scored 8 or higher on the Alcohol Use Disorder Identification Test (AUDIT) [40,41] were

XSL•FO
RenderX

included, with risky alcohol use encompassing hazardous use, harmful use, and probable dependence. Additional inclusion criteria required participants to possess sufficient Korean language skills, own an Android smartphone, and have no difficulty using mobile apps independently. They also needed to understand the study information and provide written informed consent. Exclusion criteria included major psychiatric disorders such as schizophrenia spectrum disorders, bipolar disorder, and major depressive disorder. Participants with substance use disorders involving substances other than alcohol or nicotine (eg, illicit drugs or prescription medications) were excluded. Participants with active and progressive physical illnesses, unstable medical conditions, or a life expectancy of less than 6 months were also excluded. Pregnant individuals or those planning to become pregnant during the study period, as well as individuals who participated in another clinical trial within 4 weeks before screening, were not eligible.

Procedure

Enrollment and follow-up for the study occurred between May and August 2023 (Table 1). Participants who consented to join the pilot study downloaded the Sober app and were given usernames to sign in. All participants were encouraged to use the app for self-monitoring and to log in regularly over the 4 weeks. Additionally, participants were provided with a Samsung Galaxy Smart Watch 5 (model: SM-R900NZAAKOO) to collect biometric data throughout the study. Participants were required to visit the clinic 3 times: at baseline (visit 1), 2 weeks after the baseline visit (visit 2), and 4 weeks after the baseline visit (visit 3) for assessments. Each visit included a consultation with a board-certified psychiatrist who provided standard psychiatric consultation regarding alcohol consumption based on the data entered into the mobile app.

XSL•FO
RenderX

Table . Schedule of enrollment, interventions, and assessments for a single-group, pre-post pilot study evaluating the feasibility, acceptability, and preliminary efficacy of the Sober mobile app[a].

	Study period and time points			
	Enrollment	Visit 1	Visit 2	Visit 3 and closeout
	0 days	0 days	14±4 days	28±4 days
Enrollment				
Eligibility screen and informed consent	✓			
Intervention				
Mobile app Sober and consultation with a psychiatrist at each visit		✓	✓	✓
Assessments				
Sociodemographic and clinical characteristics	✓			
Alcohol consumption history	✓			
Primary objectives				
Feasibility of the Sober app			✓	✓
Acceptability of the Sober app				✓
Days abstinent in the past 2 weeks (%)		✓	✓	✓
Secondary objectives				
AUDIT[b]	✓		✓	✓
PHQ-9[c]		✓	✓	✓
GAD-7[d]		✓	✓	✓
PSS[e]		✓	✓	✓
CIWA-Ar[f]		✓	✓	✓
Amount of alcohol consumed (mL) in the past 2 weeks		✓	✓	✓
Adverse events		✓	✓	✓

[a]The study was conducted at Yonsei Severance Hospital, Seoul, South Korea, from May to August 2023.

[b]AUDIT: Alcohol Use Disorders Identification Test.

[c]PHQ-9: Patient Health Questionnaire-9.

[d]GAD-7: Generalized Anxiety Disorder-7.

[e]PSS: Perceived Stress Scale.

[f]CIWA-Ar: Clinical Institute Withdrawal Assessment of Alcohol Scale—Revised.

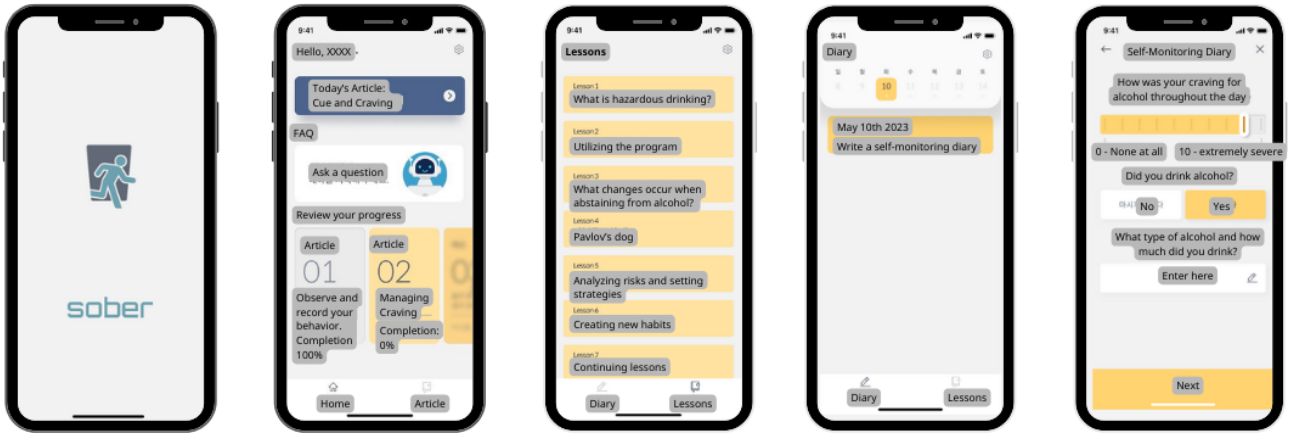
Sober App

The Sober app was developed as a digital intervention tool to reduce and prevent hazardous drinking (Figure 1). The development process involved a multidisciplinary team of experienced clinicians, software developers, product designers, and behavioral scientists to ensure clinical effectiveness and user-friendliness. Before developing the program, the team reviewed relevant materials, such as treatment manuals, intervention descriptions, guidelines, patient reports, and trial results. The program's development involved iterative evaluations of prototypes, with the findings used to refine and enhance the final app design. Designed for Android smartphones, the app supports clinicians, especially those with minimal addiction training, in providing structured, evidence-based interventions to hazardous drinkers in the general population. The cognitive behavioral therapy (CBT)–based app incorporates interventions that include psychoeducation, motivational interviewing, cognitive behavioral skills training (eg, problem-solving and behavior

XSL•FO
RenderX

substitution), and personalized normative feedback. It provides content on the effects of alcohol on the brain and behavior, risk awareness, and strategies for reducing risky drinking habits while encouraging users to consider alternatives to heavy drinking.

Figure 1. Screenshots of the Sober app's interface. Illustrative screenshots of the mobile app used in this study, showcasing features for daily logging, educational modules, and progress tracking. Original content in Korean has been translated into English for presentation.



The modularly structured app features educational modules, self-monitoring tools, and interactive elements. The CBT-based modules are divided into 4 chapters, which include 8 sessions, each addressing different aspects of hazardous drinking and providing skills training and risk awareness (Table 2). The content is organized sequentially. The first chapter, "What is hazardous drinking?," introduces participants to the concept and helps them set initial goals, such as abstaining from alcohol for 4 weeks, entering self-monitoring reports, and engaging in the CBT program. Participants also receive personalized normative feedback on their drinking habits. The second chapter, "What changes occur when abstaining from alcohol?," covers psychoeducation about changes in sleep, appetite, and emotions resulting from abstinence, with participants recording these changes in their self-monitoring logs. The third chapter, "Pavlov's dog," explains the concepts of cues, cravings, behaviors, and outcomes, helping participants identify high-risk situations and understand how cravings lead to drinking behavior. This chapter guides participants in exploring and selecting alternative behaviors to replace drinking. The fourth chapter, "Forming a healthy drinking habit," reviews the previous 4 weeks and reinforces the skills learned. To facilitate self-monitoring and promote self-awareness, the app included a daily log function, where participants recorded their alcohol consumption, mood, appetite, and sleep quality.

XSL•FO
RenderX

Table . Description of cognitive behavioral therapy–based modules in the Sober mobile app intervention.

Week and chapter		Session	Key principles and contents
Week 1			
	What is hazardous drinking?	Do I belong to high-risk drinking group?	Introduction to the High-Risk Drinking Prevention Program: this includes a description of high-risk drinking behaviors and a self-assessment for participants to evaluate their own drinking habits. Normative feedback is provided for the participants' level of current drinking habit. The program also offers psychoeducation about the long-term risks associated with continuing these drinking habits.
	What is hazardous drinking?	Promise for 4 weeks abstinence	We set up the action plan with the initial goal of abstaining from alcohol for the next 4 weeks while participating in the program. Evaluate individuals' expectations regarding the outcomes of alcohol abstinence. Provide instructions for recording sleep, emotions, appetite, and alcohol consumption in the app.
Week 2			
	What changes occur when abstaining from alcohol?	Alcohol and sleep	Psychoeducation regarding the link between alcohol and sleep.
	What changes occur when abstaining from alcohol?	Alcohol and appetite	Psychoeducation regarding the link between alcohol and appetite.
	What changes occur when abstaining from alcohol?	Alcohol and emotion	Psychoeducation regarding the link between alcohol and emotion.
Week 3			
	Pavlov's dog	Cue	Description of the cue-craving-behavior-result cycle in alcohol consumption. This session focuses on identifying the cues that lead to this cycle.
	Pavlov's dog	Craving	Psychoeducation on the various ways "craving" may be identified in participants. Participants will set up alternative behaviors to resort to when they experience alcohol cravings.
Week 4			
	Forming a healthy drinking habit	Review and relapse prevention	Review of the intervention and establishment of relapse prevention strategies.

Assessments

At baseline (visit 1), participants provided sociodemographic information, lifestyle habits, past disease history, previous medication, and detailed information on their alcohol consumption habits. At each of the 3 visits, participants were evaluated using several standardized instruments: The AUDIT was used to assess hazardous and harmful patterns of alcohol consumption [40,41]. Symptoms of depression were measured using the Patient Health Questionnaire-9 (PHQ-9) [42,43], while symptoms of anxiety were evaluated using the Generalized Anxiety Disorder-7 (GAD-7) scale [44,45]. The Perceived Stress Scale (PSS) was used to assess perceived levels of stress [46,47], and the Clinical Institute Withdrawal Assessment of Alcohol Scale—Revised (CIWA-Ar) was used to evaluate alcohol withdrawal symptoms [48,49]. At the final visit (visit 3), the completion rate of the app-based sessions was assessed along with participants' satisfaction ratings, perceived usefulness of the app, and ease of mastering the app program, all assessed through a 5-point Likert scale. Adverse events were monitored and recorded at each visit to ensure participant safety.

XSL•FO
RenderX

Outcomes

The study's primary outcomes focused on feasibility, acceptability, and preliminary efficacy. Feasibility was assessed at visit 3 based on the study completion rate and the completion of all 8 CBT-based intervention sessions. Acceptability was evaluated at visit 3 using 3 measures: participant satisfaction, perceived usefulness of the app, and ease of mastering the app.

Preliminary efficacy was analyzed by comparing outcomes between baseline (visit 1) and postintervention (visit 3). The primary outcome was the change in the number of days participants abstained from alcohol in the past 14 days. Secondary outcomes included changes in total AUDIT scores, the total amount of alcohol consumed in the past 14 days, and psychiatric symptom scores (PHQ-9, GAD-7, PSS, and CIWA-Ar), all assessed at both baseline (visit 1) and postintervention (visit 3).

Statistical Analysis

All statistical analyses were conducted using SPSS (version 26 for Windows; IBM Corp). Frequencies and percentages described categorical variables, while means and SDs were used for continuous variables. Depending on the data distribution's normality (Shapiro-Wilk test), paired *t* tests (2-tailed) or Wilcoxon signed rank tests were used to compare pre- and postintervention outcomes. Statistical significance was set at $P<.05$. Results were reported with corresponding *P* values.

Participant and Clinician Interviews

Qualitative data were collected through semistructured, open-ended interviews conducted at the final visit (visit 3). All 17 participants who completed the study participated in individual interviews to explore their experiences with the app and the clinical trial. Additionally, 2 clinicians involved in the trial were interviewed separately following the study's conclusion to provide insights from a clinical perspective.

The interviews were conducted face-to-face (in private rooms at the study site) or via videoconferencing, depending on participant preferences, and lasted approximately 20‐30 minutes. An interview guide was used to ensure consistency across interviews. For participants, questions focused on app usability, content relevance, and perceived impact on drinking behaviors. For clinicians, the interviews explored the app's integration into clinical workflows, feasibility, and suggestions for tailoring the intervention to diverse patient needs. All interviews were conducted by members of the research team involved in the study. Interviews were audio-recorded with participant consent, transcribed verbatim, and anonymized to maintain confidentiality.

Thematic analysis was used to analyze the qualitative data, following Braun and Clarke's 6-phase framework [50]: 2 independent researchers (DUM and Yujin Lee) coded the transcripts to identify key patterns and themes. Codes were grouped into overarching themes and refined iteratively to ensure they accurately represented the data. Representative verbatim quotes were selected to illustrate key findings and provide depth to the analysis. Discrepancies in coding were resolved through discussion. The findings are reported following the COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines [51].

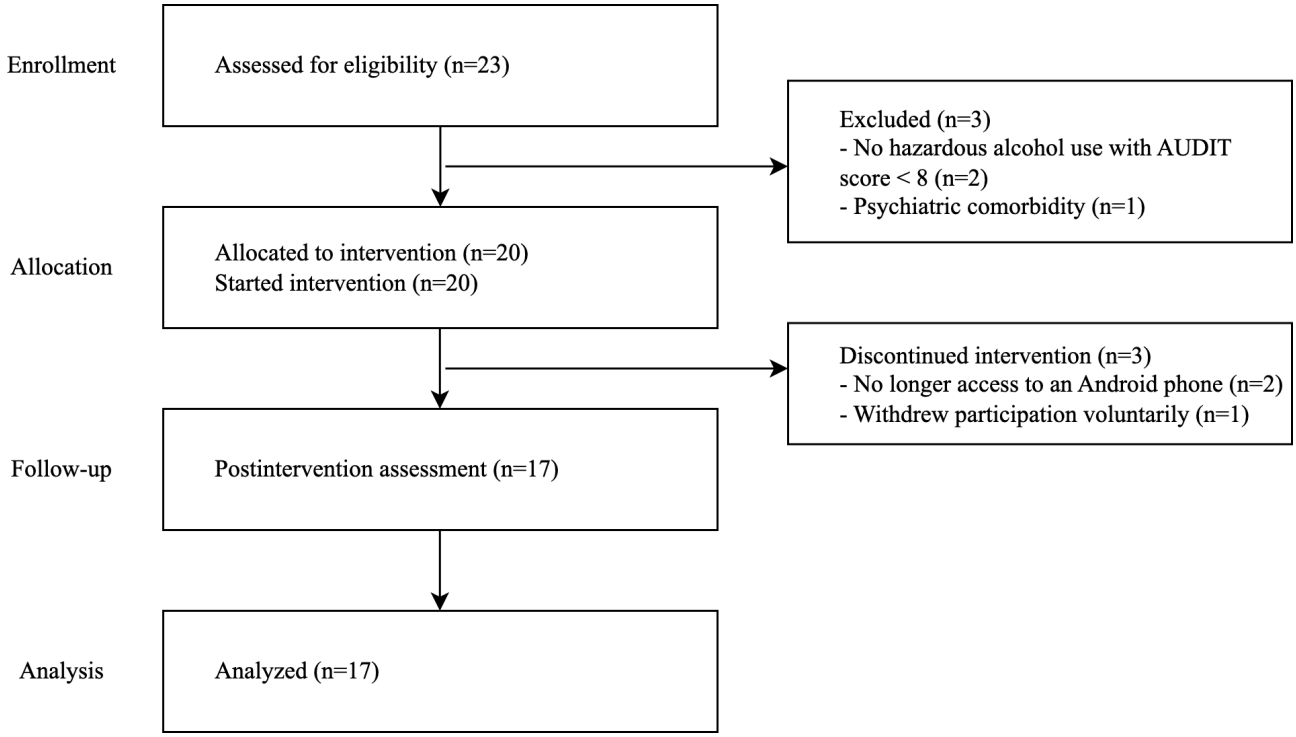
Results

Participant Eligibility and Baseline Characteristics

A total of 20 eligible participants with hazardous alcohol use, identified based on their AUDIT scores, were enrolled in the study. Of these, 17 completed the 4-week clinical trial. In total, 2 participants dropped out because they no longer had access to Android smartphones, and 1 withdrew voluntarily (Figure 2). All participants were male, with a mean age of 22.29 (SD 8.3) years. Most were students and single. Regarding smoking status, 3 participants were current smokers, 3 were ex-smokers, and 11 were never smokers. The mean age at first alcohol drinking was 18.9 (SD 1.2) years, and the mean age at first heavy drinking episode was 19.7 (SD 1.0) years. In total, 7 participants had a family history of alcohol-related problems, and 3 showed insights into their alcohol problems (Table 3).

XSL•FO
RenderX

Figure 2. Flowchart summarizing the participant recruitment, screening, enrollment, intervention, and follow-up processes. AUDIT: Alcohol Use Disorders Identification Test.



XSL•FO
RenderX

Table . Participants' baseline characteristics (n=17).

Characteristics	Values
Sex (male), n (%)	17 (100)
Age (years), mean (SD)	22.29 (8.3)
Height (cm), mean (SD)	174.9 (3.9)
Weight (kg), mean (SD)	74.4 (12.8)
Education, n (%)	
Middle school	1 (6)
High school	12 (71)
Bachelor's degree	3 (18)
Master's degree or postgraduate	1 (6)
Occupation, n (%)	
Student	14 (82)
Office worker	1 (6)
Labor worker	1 (6)
Unemployed	1 (6)
Religion, n (%)	
Christian	1 (6)
Others	0 (0)
None	16 (94)
Marital status, n (%)	
Single	16 (94)
Married	0 (0)
Divorced	1 (6)
Living arrangement, n (%)	
Living alone	8 (47)
Living with others	9 (53)
Smoking status, n (%)	
Current smokers	3 (18)
Ex-smoker	3 (18)
Never smoker	11 (65)
Age at first alcohol drinking (years), mean (SD)	18.9 (1.2)
Age at first binge drinking (years), mean (SD)	19.7 (1.0)
Presence of alcohol-related accident, n (%)	
No	10 (59)
Yes (once)	5 (29)
Yes (multiple)	2 (12)
Presence of alcohol-related legal problems, n (%)	
No	16 (94)
Yes (once)	0 (0)
Yes (multiple)	1 (56)
Presence of drunken driving, n (%)	
No	17 (100)
Yes	0 (0)

XSL•FO
RenderX

Characteristics	Values
Family history of alcohol-related problems, n (%)	
No	10 (59)
Yes	7 (41)
Insight of alcohol problems, n (%)	
No	14 (82)
Yes	3 (18)

Feasibility and Acceptability

The feasibility of the intervention was assessed based on the completion rates of the study and the CBT-based educational sessions within the app. Of the 20 participants initially enrolled, 17 (85%) completed the study. Among these 17 participants, 10 (59%) completed all 8 CBT-based sessions. Acceptability was evaluated in the 17 participants who completed the final visit (visit 3). Participants were asked to rate their satisfaction with the app on a scale from 0=not satisfied at all to 5=very satisfied. The average satisfaction score was 3.8 (SD 0.9). When asked to rate the usefulness of the app from 0=not useful at all to 5=very useful, the average response was 3.7 (SD 0.9). Additionally, participants rated their ease of mastering the app on a scale from 0=could not use without help to 5=used naturally without difficulty, with an average score of 3.5 (SD 0.6).

Preliminary Efficacy Outcomes

Preliminary efficacy outcomes showed significant improvements in several key measures (Table 4). The percentage of days participants remained abstinent increased from 67% to 85% (z=−3.17; P=.002). The frequency of heavy drinking episodes decreased from 3.3 to 1.9 (t_{16}=−2.97; P=.003). The amount of alcohol consumed reduced significantly from 456.8 to 195.9 mL (t_{16}=3.16; P=.002). Total AUDIT scores dropped from 17.5 to 10.7 (t_{16}=4.51; P<.001). Regarding psychological assessment outcomes, PHQ-9 scores decreased from 5.8 to 4.4 (t_{16}=2.91; P=.01), and GAD-7 scores from 3.4 to 2.1 (z=−2.80; P=.005). There were nonsignificant changes in PSS scores (17.2 to 16.1; P=.19) and CIWA-Ar scores (0.3 to 0.8; P=.74). No adverse events were reported by the study participants.

Table . Changes in alcohol consumption, psychological symptoms, and related measures before and after a 4-week intervention using the mobile app "Sober."

	Baseline (version 1), mean (SD)	Postintervention (version 3), mean (SD)	Difference, mean (SD)	t test (df=16)[a]	P value
Days abstinent in the past 2 weeks (%)	67 (25)	85 (16)	18.0 (4.2)	z=−3.17[b]	.002
Frequency of heavy drinking (AUDIT[c], question 3)	3.3 (0.7)	1.9 (1.4)	−1.4 (0.3)	−2.97	.003
Amount of alcohol consumed (mL)	456.8 (441.9)	195.9 (320.2)	−260.9 (82.6)	3.16	.002
AUDIT	17.5 (6.2)	10.7 (9.0)	−6.7 (1.5)	4.51	<.001
PHQ-9[d]	5.8 (5.4)	4.4 (5.6)	−1.4 (0.5)	2.91	.01
GAD-7[e]	3.4 (3.4)	2.1 (3.5)	−1.3 (0.5)	z=−2.80[b]	.005
PSS[f]	17.2 (2.9)	16.1 (2.6)	−1.1 (0.8)	1.38	.19
CIWA-Ar[g]	0.3 (0.6)	0.8 (2.7)	0.5 (0.7)	z=0.01[b]	.74

[a]Paired t test (2-tailed).

[b]Wilcoxon signed rank test.

[c]AUDIT: Alcohol Use Disorder Identification Test.

[d]PHQ-9: Patient Health Questionnaire-9.

[e]GAD-7: Generalized Anxiety Disorder-7.

[f]PSS: Perceived Stress Scale.

[g]CIWA-Ar: Clinical Institute Withdrawal Assessment of Alcohol Scale—Revised.

XSL•FO
RenderX

Participant and Clinician Perspectives on the Intervention

Qualitative feedback from participants and clinicians provided insights into the app's usability, effectiveness, and areas for improvement. For a detailed summary, see Multimedia Appendices 1 and 2.

Participant Feedback

Interviews with the 17 participants highlighted 3 primary themes: behavioral impact, usability challenges, and content suggestions.

Participants consistently reported that using the app increased their awareness of drinking behaviors and encouraged self-reflection, which helped them reduce alcohol consumption:

> *When I logged my drinking every day, it made me stop and think about how much I was drinking. It naturally helped me reduce my alcohol consumption.* [Participant 1]
>
> *Tracking how often and how much I drink made me more aware of my habits.* [Participant 3]

Participants also noted the app's role in helping them observe connections between their drinking and other aspects of their health, such as sleep and appetite:

> *It was helpful to see how my drinking affected things like my appetite and sleep. It made me think about how drinking impacts my health.* [Participant 13]

However, a participant reported unintended effects, such as increased cravings when focusing on abstinence:

> *Sometimes, thinking about not drinking actually made me want to drink more. It felt like a trigger at times.* [Participant 7]

Usability issues were commonly mentioned, including app crashes, lack of notifications, and vague criteria. Participants suggested improvements to enhance functionality:

> *The app would crash a lot when I tried to log something, which was really frustrating.* [Participant 6]
>
> *I wish the app had reminders, like notifications in the morning to remind me to log.* [Participant 4]
>
> *It felt like some of the criteria in the app were too vague. I wasn't always sure what I was supposed to do.* [Participant 3]
>
> *It would be nice to see trends over time, like a graph showing my progress.* [Participant 9]

Participants also recommended expanding the app's content with more educational materials, such as strategies for managing cravings and understanding the risks of alcohol consumption:

> *The app could include more about the long-term effects of drinking.* [Participant 17]
>
> *It would be great if I could log more details, like my emotions or cravings. Just using a simple scale didn't capture everything.* [Participant 12]

Clinician Feedback

Interviews with 2 clinicians involved in the trial provided additional perspectives on the app's clinical utility and suggestions for improvement.

Clinicians observed that participants who recorded their mood and sleep alongside abstinence became more aware of positive changes in their mental and physical health, reinforcing their motivation to reduce alcohol consumption:

> *By recording their mood and sleep daily while abstaining from alcohol, participants seemed to independently recognize the positive changes in their mind and body, which, in turn, reinforced their commitment to staying alcohol-free.* [Clinician 1]

They emphasized the importance of flexibility in accommodating participants' diverse goals, such as achieving abstinence or reducing binge episodes, suggesting features like weekly goal setting and tracking achievements:

> *Allowing for participants to set personalized goals every week and adding features that allow participants to record their achievement would be beneficial.* [Clinician 1]

To enhance clinical integration, clinicians recommended features that facilitate communication and monitoring, such as messaging options and tools to assess participants' motivation:

> *Features that connect clinicians and patients, like messaging or feedback options, would make follow-ups easier and more effective.* [Clinician 2]
>
> *If patients' motivation levels for change can be assessed through the app, it will help clinicians effectively tailor their approach to patient interventions in real-life settings.* [Clinician 2]

Additionally, clinicians underscored the importance of tailoring app content to different age groups and suggested incorporating features that allow participants to reflect on their motivations for change:

> *Younger and older users have very different needs and expectations. The app should include content that's tailored to different age groups.* [Clinician 2]

Discussion

Principal Findings

This pilot study aimed to assess the feasibility, acceptability, and preliminary efficacy of the Sober mobile app for reducing hazardous alcohol consumption. The results demonstrated significant improvements in key measures of alcohol use, including increased days of abstinence, reduction in heavy drinking episodes, and decreased overall alcohol consumption. Additionally, notable reductions in AUDIT scores and improvements in psychological assessments (PHQ-9 and GAD-7 scores) were observed.

Qualitative feedback from participants and clinicians provided valuable insights into the app's strengths and areas for improvement. Participants emphasized the app's ability to increase awareness of drinking behaviors and promote

XSL•FO
RenderX

self-reflection, which supported behavior change. However, usability challenges, such as technical issues and limited options for personalization, were highlighted. Clinicians stressed the importance of tailoring the app to diverse user goals and incorporating features to facilitate clinical integration. These findings will inform the refinement of the app and future iterations of the study design.

Comparison With Prior Work

Our findings align with previous research indicating the effectiveness of digital interventions in reducing alcohol consumption, particularly among the young male population [52], which mirrors the demographic of our study. Other studies have also shown positive effects in the general population, suggesting that such interventions could have broad applicability [29,31,53-55]. The promising use of CBT in digital formats is consistent with existing evidence that CBT delivered in clinical settings can reduce alcohol consumption [11,56,57]. Other digital interventions using CBT-based approaches have also demonstrated effectiveness in reducing alcohol consumption [58-60]. However, in the CBT framework, various interventions, including normative feedback, self-monitoring, psychoeducation, action planning, goal setting, problem-solving skills, and identifying or managing triggers and cravings, aim to strengthen "reflective" cognitive processes used to control behavior [25]. This approach diversity highlights the need for a more standardized framework for developing and evaluating digital interventions [31]. This study also supports findings that digital interventions can positively impact co-occurring psychological symptoms, such as depression and anxiety, further emphasizing their dual utility [61,62].

The app's high feasibility, demonstrated by an 85% (17/20) study completion rate, suggests that it is well-received and manageable for participants. However, the fact that only 59% (10/17) completed all 8 CBT-based educational sessions highlights the challenge of maintaining engagement with the full content. Strategies to enhance user engagement, such as incorporating gamification elements, personalized feedback, and real-time adjustments using predictive modeling [63-65], could address this gap.

Acceptability ratings were positive, with satisfaction and perceived usefulness ratings averaging 3.8 and 3.7 of 5, respectively. Feedback indicated that daily logging and self-monitoring features were valuable in promoting self-reflection and awareness of drinking habits. However, several usability improvements were suggested, such as showing averages of past drinking, fixing app crashes, adding notification settings, and integrating with other health apps. Clinician feedback emphasized the need for customizable goals, treatment periods, and content tailored to different age groups. These suggestions will be considered in future iterations of the app to ensure that it meets the diverse needs of its users. Furthermore, incorporating interactive elements and practical synthesis in workflows could enhance the overall intervention.

Strengths and Limitations

This study demonstrated the potential of the Sober app for reducing risky alcohol consumption; yet, several limitations must be acknowledged. The single-arm, pre- and postdesign limits causal inferences, and the study was not sufficiently powered. The short duration of the study and lack of long-term follow-up limit the ability to assess the sustained effects of the intervention. The inclusion of face-to-face sessions with clinicians may have influenced the outcomes, as the reinforcement and availability of clinicians could have impacted the results [66]. Furthermore, the homogeneous participant demographic, consisting solely of young male participants, limits the generalizability of the findings.

Despite these limitations, the study had several notable strengths. To our knowledge, this is the first study to evaluate a brief digital intervention for hazardous alcohol use in South Korea. The significant preliminary efficacy demonstrated suggests potential for real-life clinical implementation. The high overall completion rate and positive acceptability ratings indicate that the app is well-received and safe to use. Importantly, the inclusion of qualitative feedback provides a nuanced understanding of user experiences, directly informing app improvements and study design.

At a population level, addressing interventions for individuals who exhibit hazardous drinking can have the greatest impact on reducing alcohol-related problems [67]. By providing an easily accessible, scalable, and cost-effective intervention, the Sober app could help reduce the overall burden of hazardous drinking on health care systems. While numerous studies have shown that screening and brief intervention can lead to significant decreases in alcohol use in primary care populations [13,68], digital intervention lessens the burden on primary care providers, requires no extensive training, and mitigates the fear of stigma, as interventions can be anonymous [69]. This is particularly relevant in South Korea, where hazardous drinking rates are significant [12], and smartphone ownership is nearly ubiquitous [23].

Future Research

Future research should include an RCT to assess the app's efficacy compared to standard care without app use, with larger samples and follow-up assessments to determine long-term effects. Expanding the study to include a more diverse population will enhance the generalizability of the findings [70]. Optimizing the app's tailoring, structure, and content based on participant and clinician feedback is essential. Personalized content, interactive elements [71], gamification [72,73], and predictive modeling using data from daily logs and wearables could improve user engagement and intervention effectiveness by providing personalized feedback and real-time adjustments [63-65]. Integrating the app with other health monitoring tools and clinician workflow could give a more holistic approach to managing alcohol consumption [71,74].

Conclusions

The mobile app Sober shows promise as an effective tool for reducing hazardous alcohol consumption and improving related psychiatric symptoms. This pilot study demonstrated significant preliminary efficacy, high overall feasibility, and positive acceptability. The app's features, such as daily logging and CBT-based content, helped users develop greater self-awareness

XSL•FO
RenderX

and control over their drinking habits, showing potential for integration into clinical practice as a supportive tool. Qualitative feedback from participants and clinicians identified critical areas for improvement, such as usability, goal customization, and age-specific content, which will guide future app refinement. With further research and refinement, the Sober app has the potential to make an impact on public health by providing a scalable and accessible intervention for hazardous drinking.

Acknowledgments

This work was supported by a Seoul Business Agency grant funded by the Seoul Metropolitan Government (BT210029). The funding organization had no role in the study design, data collection, analysis, interpretation of results, or the decision to submit the manuscript for publication. The authors would like to acknowledge Yujin Lee (College of Pharmacy, Yonsei Institute of Pharmaceutical Sciences, Yonsei University, Incheon, Republic of Korea) for her contribution as one of the 2 independent researchers responsible for coding and analyzing the qualitative data. Generative artificial intelligence was not used in any portion of this manuscript.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

DUM was employed by WELT Corp Ltd. The other authors declare no competing interests.

Multimedia Appendix 1
Participant codes with acceptability measures.
[DOCX File, 20 KB - formative_v9i1e64459_app1.docx]

Multimedia Appendix 2
Thematic quotes from participant and clinician interviews.
[DOCX File, 21 KB - formative_v9i1e64459_app2.docx]

References

1. Health at a glance 2023: OECD indicators. : OECD Publishing; 2023. [doi: 10.1787/7a7afb35-en]
2. Global status report on alcohol and health and treatment of substance use disorders. World Health Organization. 2024. URL: https://www.who.int/publications/i/item/9789240096745 [accessed 2025-01-21]
3. Global status report on alcohol and health 2018. World Health Organization. 2018. URL: https://www.who.int/publications/i/item/9789241565639 [accessed 2025-01-21]
4. Rehm J, Gmel GE Sr, Gmel G, et al. The relationship between different dimensions of alcohol use and the burden of disease—an update. Addiction 2017 Jun;112(6):968-1001. [doi: 10.1111/add.13757] [Medline: 28220587]
5. Rehm J, Patra J, Brennan A, et al. The role of alcohol use in the aetiology and progression of liver disease: a narrative review and a quantification. Drug Alcohol Rev 2021 Nov;40(7):1377-1386. [doi: 10.1111/dar.13286] [Medline: 33783063]
6. Wild C, Weiderpass E, Stewart BW. World Cancer Report: Cancer Research for Cancer Prevention. : International Agency for Research on Cancer; 2020.
7. White IR, Altmann DR, Nanchahal K. Alcohol consumption and mortality: modelling risks for men and women at different ages. Bmj 2002 Jul 27;325(7357):191-191. [doi: 10.1136/bmj.325.7357.191]
8. Zahr NM, Kaufman KL, Harper CG. Clinical and pathological features of alcohol-related brain damage. Nat Rev Neurol 2011 May;7(5):284-294. [doi: 10.1038/nrneurol.2011.42] [Medline: 21487421]
9. Grant BF, Goldstein RB, Saha TD, et al. Epidemiology of DSM-5 alcohol use disorder: results from the national epidemiologic survey on alcohol and related conditions III. JAMA Psychiatry 2015 Aug;72(8):757-766. [doi: 10.1001/jamapsychiatry.2015.0584] [Medline: 26039070]
10. Puddephatt JA, Irizar P, Jones A, Gage SH, Goodwin L. Associations of common mental disorder with alcohol use in the adult general population: a systematic review and meta-analysis. Addiction 2022 Jun;117(6):1543-1572. [doi: 10.1111/add.15735] [Medline: 34729837]
11. MacKillop J, Agabio R, Feldstein Ewing SW, et al. Hazardous drinking and alcohol use disorders. Nat Rev Dis Primers 2022 Dec 22;8(1):80. [doi: 10.1038/s41572-022-00406-1] [Medline: 36550121]
12. Hong JW, Noh JH, Kim DJ. The prevalence of and factors associated with high-risk alcohol consumption in Korean adults: the 2009–2011 Korea National Health and Nutrition Examination Survey. PLoS ONE 2017;12(4):e0175299. [doi: 10.1371/journal.pone.0175299]

XSL•FO
RenderX

13. Kaner EF, Beyer FR, Muirhead C, et al. Effectiveness of brief alcohol interventions in primary care populations. Cochrane Database Syst Rev 2018 Feb 24;2(2):CD004148. [doi: 10.1002/14651858.CD004148.pub4] [Medline: 29476653]
14. Solberg LI, Maciosek MV, Edwards NM. Primary care intervention to reduce alcohol misuse ranking its health impact and cost effectiveness. Am J Prev Med 2008 Feb;34(2):143-152. [doi: 10.1016/j.amepre.2007.09.035] [Medline: 18201645]
15. The SAFER technical package: five areas of intervention at national and subnational levels. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/the-safer-technical-package [accessed 2025-01-21]
16. US Preventive Services Task Force, Curry SJ, Krist AH, et al. Screening and behavioral counseling interventions to reduce unhealthy alcohol use in adolescents and adults: US Preventive Services Task Force recommendation statement. JAMA 2018 Nov 13;320(18):1899-1909. [doi: 10.1001/jama.2018.16789] [Medline: 30422199]
17. Miller WR, Rollnick S. Motivational Interviewing: Helping People Change: Guilford Press; 2012.
18. Vaca FE, Winn D. The basics of alcohol screening, brief intervention and referral to treatment in the emergency department. West J Emerg Med 2007 Aug;8(3):88-92. [Medline: 19561690]
19. Johnson M, Jackson R, Guillaume L, Meier P, Goyder E. Barriers and facilitators to implementing screening and brief intervention for alcohol misuse: a systematic review of qualitative evidence. J Public Health (Oxf) 2011 Sep 1;33(3):412-421. [doi: 10.1093/pubmed/fdq095]
20. Nygaard P, Aasland OG. Barriers to implementing screening and brief interventions in general practice: findings from a qualitative study in Norway. Alcohol Alcohol 2011 Jan 1;46(1):52-60. [doi: 10.1093/alcalc/agq073]
21. Wallhed Finn S, Bakshi AS, Andréasson S. Alcohol consumption, dependence, and treatment barriers: perceptions among nontreatment seekers with alcohol dependence. Subst Use Misuse 2014 May 12;49(6):762-769. [doi: 10.3109/10826084.2014.891616]
22. Kilian C, Manthey J, Carr S, et al. Stigmatization of people with alcohol use disorders: an updated systematic review of population studies. Alcoholism Clin Exp Res 2021 May;45(5):899-911. [doi: 10.1111/acer.14598]
23. Taylor K, Silver L. Smartphone ownership is growing rapidly around the world, but not always equally. Pew Research Center. 2019. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2023-05-01]
24. Chong SOK, Pedron S, Abdelmalak N, Laxy M, Stephan AJ. An umbrella review of effectiveness and efficacy trials for app-based health interventions. NPJ Digit Med 2023 Dec 16;6(1):233. [doi: 10.1038/s41746-023-00981-x] [Medline: 38104213]
25. Colbert S, Thornton L, Richmond R. Smartphone apps for managing alcohol consumption: a literature review. Addict Sci Clin Pract 2020 Dec;15(1):17. [doi: 10.1186/s13722-020-00190-x]
26. Manning V, Piercy H, Garfield JBB, Clark SG, Andrabi MN, Lubman DI. A personalized approach bias modification smartphone app ("SWiPE") to reduce alcohol use: open-label feasibility, acceptability, and preliminary effectiveness study. JMIR Mhealth Uhealth 2021 Dec 10;9(12):e31353. [doi: 10.2196/31353] [Medline: 34890355]
27. Witkiewitz K, Desai SA, Bowen S, Leigh BC, Kirouac M, Larimer ME. Development and evaluation of a mobile intervention for heavy drinking and smoking among college students. Psychol Addict Behav 2014 Sep;28(3):639-650. [doi: 10.1037/a0034747] [Medline: 25000269]
28. Boyle SC, LaBrie JW, Trager BM, Costine LD. A gamified personalized normative feedback app to reduce drinking among sexual minority women: randomized controlled trial and feasibility study. J Med Internet Res 2022 May 13;24(5):e34853. [doi: 10.2196/34853] [Medline: 35559854]
29. Chiauzzi E, Green TC, Lord S, Thum C, Goldstein M. My student body: a high-risk drinking prevention web site for college students. J Am Coll Health 2005 May 1;53(6):263-274. [doi: 10.3200/JACH.53.6.263-274]
30. McDermott KT, Noake C, Wolff R, et al. Digital interventions to moderate alcohol consumption in young people: a Cancer Prevention Europe overview of systematic reviews. Front Digit Health 2023;5:1178407. [doi: 10.3389/fdgth.2023.1178407] [Medline: 37288171]
31. Kaner EF, Beyer FR, Garnett C, et al. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. Cochrane Database Syst Rev 2017 Sep 25;9(9):CD011479. [doi: 10.1002/14651858.CD011479.pub2] [Medline: 28944453]
32. Haynes JC, Farrell M, Singleton N, et al. Alcohol consumption as a risk factor for anxiety and depression: results from the longitudinal follow-up of the National Psychiatric Morbidity Survey. Br J Psychiatry 2005 Dec;187(6):544-551. [doi: 10.1192/bjp.187.6.544] [Medline: 16319407]
33. Strid C, Andersson C, Öjehagen A. The influence of hazardous drinking on psychological functioning, stress and sleep during and after treatment in patients with mental health problems: a secondary analysis of a randomised controlled intervention study. BMJ Open 2018 Mar 6;8(3):e019128. [doi: 10.1136/bmjopen-2017-019128] [Medline: 29511011]
34. Smith JP, Randall CL. Anxiety and alcohol use disorders: comorbidity and treatment considerations. Alcohol Res 2012;34(4):414-431. [Medline: 23584108]
35. Baker AL, Thornton LK, Hiles S, Hides L, Lubman DI. Psychological interventions for alcohol misuse among people with co-occurring depression or anxiety disorders: a systematic review. J Affect Disord 2012 Aug;139(3):217-229. [doi: 10.1016/j.jad.2011.08.004]

XSL•FO
RenderX

36. Magee KE, Connell AM. The role of substance use coping in linking depression and alcohol use from late adolescence through early adulthood. Exp Clin Psychopharmacol 2021 Dec;29(6):659-669. [doi: 10.1037/pha0000436] [Medline: 33539119]
37. Schouten MJE, Christ C, Dekker JJM, Riper H, Goudriaan AE, Blankers M. Digital interventions for people with co-occurring depression and problematic alcohol use: a systematic review and meta-analysis. Alcohol Alcohol 2022 Jan 8;57(1):113-124. [doi: 10.1093/alcalc/agaa147] [Medline: 33588432]
38. Stapinski LA, Prior K, Newton NC, et al. Are we making Inroads? A randomized controlled trial of a psychologist-supported, web-based, cognitive behavioral therapy intervention to reduce anxiety and hazardous alcohol use among emerging adults. eClinMed 2021 Sep;39:101048. [doi: 10.1016/j.eclinm.2021.101048] [Medline: 34622183]
39. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33. [Medline: 9881538]
40. Babor TF, Higgins-Biddle JC, Saunders JB, Monteiro MG. AUDIT: the Alcohol Use Disorders Identification Test: guidelines for use in primary health care. World Health Organization. 2001. URL: https://www.who.int/publications/i/item/WHO-MSD-MSB-01.6a [accessed 2025-01-21]
41. Kim JS, Oh MK, Park BK, Lee MK, Kim GJ. Screening criteria of alcoholism by Alcohol Use Disorders Identification Test (AUDIT) in Korea. J Korean Acad Fam Med 1999;20(9):1152-1159.
42. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
43. Shin C, Lee SH, Han KM, Yoon HK, Han C. Comparison of the usefulness of the PHQ-8 and PHQ-9 for screening for major depressive disorder: analysis of psychiatric outpatient data. Psychiatry Investig 2019 Apr;16(4):300-305. [doi: 10.30773/pi.2019.02.01]
44. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
45. Ahn JK, Kim Y, Choi KH. The psychometric properties and clinical utility of the Korean version of GAD-7 and GAD-2. Front Psychiatry 2019;10:127. [doi: 10.3389/fpsyt.2019.00127]
46. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
47. Lee EH. Review of the psychometric evidence of the Perceived Stress Scale. Asian Nurs Res (Korean Soc Nurs Sci) 2012 Dec;6(4):121-127. [doi: 10.1016/j.anr.2012.08.004]
48. Choi HY, Seo JS, Lee SK. Assessment and treatment of alcohol withdrawal syndrome. Korean J Gastroenterol 2020 Aug 25;76(2):71-77. [doi: 10.4166/kjg.2020.76.2.71]
49. Sullivan JT, Sykora K, Schneiderman J, Naranjo CA, Sellers EM. Assessment of alcohol withdrawal: the revised Clinical Institute Withdrawal Assessment for Alcohol scale (CIWA-Ar). Br J Addict 1989 Nov;84(11):1353-1357. [doi: 10.1111/j.1360-0443.1989.tb00737.x] [Medline: 2597811]
50. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
51. Booth A, Hannes K, Harden A, Noyes J, Harris J, Tong A. COREQ (Consolidated Criteria for Reporting Qualitative Studies). In: Guidelines for Reporting Health Research: A User's Manual 2014:214-226. [doi: 10.1002/9781118715598]
52. Bertholet N, Cunningham JA, Faouzi M, et al. Internet‐based brief intervention for young men with unhealthy alcohol use: a randomized controlled trial in a general population sample. Addiction 2015 Nov;110(11):1735-1743. [doi: 10.1111/add.13051]
53. Kraepelien M, Sundström C, Johansson M, Ivanova E. Digital psychological self-care for problematic alcohol use: feasibility of a new clinical concept. BJPsych Open 2023 May 24;9(3):e91. [doi: 10.1192/bjo.2023.73] [Medline: 37222099]
54. Bendtsen P, Bendtsen M, Karlsson N, White IR, McCambridge J. Online alcohol assessment and feedback for hazardous and harmful drinkers: findings from the AMADEUS-2 randomized controlled trial of routine practice in Swedish universities. J Med Internet Res 2015 Jul 9;17(7):e170. [doi: 10.2196/jmir.4020] [Medline: 26159179]
55. Blankers M, Koeter MWJ, Schippers GM. Internet therapy versus internet self-help versus no treatment for problematic alcohol use: a randomized controlled trial. J Consult Clin Psychol 2011;79(3):330-341. [doi: 10.1037/a0023498]
56. Magill M, Ray L, Kiluk B, et al. A meta-analysis of cognitive-behavioral therapy for alcohol or other drug use disorders: treatment efficacy by contrast condition. J Consult Clin Psychol 2019 Dec;87(12):1093-1105. [doi: 10.1037/ccp0000447] [Medline: 31599606]
57. Ray LA, Meredith LR, Kiluk BD, Walthers J, Carroll KM, Magill M. Combined pharmacotherapy and cognitive behavioral therapy for adults with alcohol or substance use disorders. JAMA Netw Open 2020 Jun 1;3(6):e208279. [doi: 10.1001/jamanetworkopen.2020.8279]
58. Augsburger M, Kaal E, Ülesoo T, et al. Effects of a minimal-guided on-line intervention for alcohol misuse in Estonia: a randomized controlled trial. Addiction 2022 Jan;117(1):108-117. [doi: 10.1111/add.15633] [Medline: 34184795]
59. Baumgartner C, Schaub MP, Wenger A, et al. "Take Care of You"—efficacy of integrated, minimal-guidance, internet-based self-help for reducing co-occurring alcohol misuse and depression symptoms in adults: results of a three-arm randomized

XSL•FO
RenderX

controlled trial. Drug Alcohol Depend 2021 Aug 1;225:108806. [doi: 10.1016/j.drugalcdep.2021.108806] [Medline: 34171823]
60. Kiluk BD, Ray LA, Walthers J, Bernstein M, Tonigan JS, Magill M. Technology-delivered cognitive-behavioral interventions for alcohol use: a meta-analysis. Alcohol Clin Exp Res 2019 Nov;43(11):2285-2295. [doi: 10.1111/acer.14189] [Medline: 31566787]
61. Schuckit MA, Hesselbrock V. Alcohol dependence and anxiety disorders: what is the relationship. Focus Tech Coop 2004;151(3):1723-1453. [doi: 10.1176/foc.2.3.440]
62. Boden JM, Fergusson DM. Alcohol and depression. Addiction 2011 May;106(5):906-914. [doi: 10.1111/j.1360-0443.2010.03351.x] [Medline: 21382111]
63. Chih MY, Patton T, McTavish FM, et al. Predictive modeling of addiction lapses in a mobile health application. J Subst Abuse Treat 2014 Jan;46(1):29-35. [doi: 10.1016/j.jsat.2013.08.004] [Medline: 24035143]
64. Bae S, Chung T, Ferreira D, Dey AK, Suffoletto B. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: implications for just-in-time adaptive interventions. Addict Behav 2018 Aug;83:42-47. [doi: 10.1016/j.addbeh.2017.11.039] [Medline: 29217132]
65. Lindner P, Johansson M, Gajecki M, Berman AH. Using alcohol consumption diary data from an internet intervention for outcome and predictive modeling: a validation and machine learning study. BMC Med Res Methodol 2020 May 11;20(1):111. [doi: 10.1186/s12874-020-00995-z] [Medline: 32393245]
66. Kay-Lambkin F, Baker A, Lewin T, Carr V. Acceptability of a clinician-assisted computerized psychological intervention for comorbid mental health and substance use problems: treatment adherence data from a randomized controlled trial. J Med Internet Res 2011 Jan 27;13(1):e11. [doi: 10.2196/jmir.1522] [Medline: 21273184]
67. McGovern R, Kaner E. Alcohol prevention and treatment: interventions for hazardous, harmful, and dependent drinkers. In: The Handbook of Health Behavior Change, 4th edition: Springer Publishing Company; 2014:211-227.
68. Bertholet N, Daeppen JB, Wietlisbach V, Fleming M, Burnand B. Reduction of alcohol consumption by brief alcohol intervention in primary care. Arch Intern Med 2005 May 9;165(9):986. [doi: 10.1001/archinte.165.9.986]
69. Lapham GT, Hawkins EJ, Chavez LJ, et al. Feedback from recently returned veterans on an anonymous web-based brief alcohol intervention. Addict Sci Clin Pract 2012 Aug 28;7(1):17. [doi: 10.1186/1940-0640-7-17] [Medline: 23186354]
70. Saunders JB, Kypri K, Walters ST, Laforge RG, Larimer ME. Approaches to brief intervention for hazardous drinking in young people. Alcoholism Clin Exp Res 2004 Feb;28(2):322-329. [doi: 10.1097/01.ALC.0000113418.12889.08]
71. Fischer LC, Kölligan V, Wieland N, Klein M. Development and evaluation of a digital health intervention for substance use reduction in young refugees with problematic use of alcohol and/or cannabis-study protocol for a single-armed feasibility trial. Front Public Health 2021;9:557431. [doi: 10.3389/fpubh.2021.557431] [Medline: 33869122]
72. Boendermaker WJ, Boffo M, Wiers RW. Exploring elements of fun to motivate youth to do cognitive bias modification. Games Health J 2015 Dec;4(6):434-443. [doi: 10.1089/g4h.2015.0053] [Medline: 26421349]
73. Prins PJM, Brink ET, Dovis S, et al. "Braingame Brian": toward an executive function training program with game elements for children with ADHD and cognitive control problems. Games Health J 2013 Feb;2(1):44-49. [doi: 10.1089/g4h.2013.0004]
74. Glass JE, Tiffany B, Matson TE, et al. Approaches for implementing digital interventions for alcohol use disorders in primary care: a qualitative, user-centered design study. Implement Res Pract 2022;3:26334895221135264. [doi: 10.1177/26334895221135264] [Medline: 37091101]

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
CBT: cognitive behavioral therapy
CIWA-Ar: Clinical Institute Withdrawal Assessment of Alcohol Scale—Revised
COREQ: Consolidated Criteria for Reporting Qualitative Research
GAD-7: Generalized Anxiety Disorder-7
PHQ-9: Patient Health Questionnaire-9
PSS: Perceived Stress Scale
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by A Mavragani; submitted 22.07.24; peer-reviewed by S Nkosi; revised version received 02.12.24; accepted 02.12.24; published 19.03.25.

Please cite as:
Kwon M, Moon DU, Kang M, Jung YC
A Brief Cognitive Behavioral Therapy–Based Digital Intervention for Reducing Hazardous Alcohol Use in South Korea: Development and Prospective Pilot Study
JMIR Form Res 2025;9:e64459
URL: https://formative.jmir.org/2024/1/e64459
doi:10.2196/64459

© Manjae Kwon, Daa Un Moon, Minjae Kang, Young-Chul Jung. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring the Use of an Augmented Reality Device Learning Tool for Multidisciplinary Staff Training on Domestic Abuse and Sexual Violence: Postintervention Qualitative Evaluation

Dilroshini Karunaratne[1], MBChB; Jessica Whittock[2], BSc, MA; Amber Moore[1], MBBS, BSc; Krishna Dasigan[1], MBBS, MRCP; Jasmine Chevolleau[1], MBChB; Brent Bartholomew[3], MBChB, BSc, MRCP; Nikki Kelly[4]; Charlotte E Cohen[5], MBBS, BSc, MRCP

[1]Department of Medical Education, Chelsea and Westminster Hospital NHS Foundation Trust, London, United Kingdom

[2]Department of Safeguarding Adults and Children, Chelsea and Westminster NHS Foundation Trust, Standing Together Against Domestic Abuse, London, United Kingdom

[3]Department of Acute Medicine and Emergency Services, Faculty of Medicine, Chelsea and Westminster Hospital NHS Foundation Trust, London, United Kingdom

[4]The SafeLives Practice Team, SafeLives, London, United Kingdom

[5]Department of Genitourinary Medicine and HIV Services, Faculty of Medicine, Chelsea and Westminster Hospital NHS Foundation Trust, London, United Kingdom

Corresponding Author:
Dilroshini Karunaratne, MBChB
Department of Medical Education, Chelsea and Westminster Hospital NHS Foundation Trust, London, United Kingdom

Abstract

Background: Legislative policies published by National Health Service, England and the UK Government focus on prioritizing the creation of a stronger system. These frameworks emphasize on the improvement of health care staff's ability to identify and refer domestic abuse (DA) survivors as key areas for supporting workforce development. Health care staff are often the first professional contact of survivors of DA, and insufficient staff training is a key barrier to survivors being identified and directed to support. The Microsoft HoloLens2 is a mixed-reality headset that allows virtual objects (holograms) to be integrated into the real world. Mixed-reality headsets are being increasingly used within medical education and have the advantage of independent operation, reducing the staffing requirements for teaching. The HoloLens2 can be used to project HoloPatients (HPs), which resemble clinically unwell patients, into the classroom. Two of these HPs have been specifically designed to portray survivors of DA and sexual violence (SV).

Objective: This study explored potential uses of the HP in DA and SV training as a potential survivor-centered educational initiative that could be used as an adjunct to existing training for health care professionals and community sector workers.

Methods: Frontline staff and community stakeholders from the national health service, DA, and law enforcement sectors were invited on 3 separate occasions (n=14, 12, 22) to a HoloLens2 demonstration that displayed 9 HPs. The patient voice was to be outlined by personalized scripts, co-created alongside sector charities, ensuring survivor engagement and participation. Participants were given the opportunity to wear the headset and familiarize themselves with the technology during the sessions. A post-intervention evaluation research model was used to explore the feasibility and functionality of the HP as an educational tool.

Results: Thematic analysis described the HP as a "realistic," "adjustable" tool that "creates a safe learning environment." Participants suggested it could be useful in "pre-exposure preparation" by "improving communication" and allowing different approaches to be trialed in a safe environment. The use of survivor scripts was described as a useful tool to "bring the survivor into the learning space" in a safe way. Participants identified the HP as a suitable tool for workers inside and outside health care, including social sectors such as law enforcement (32%).

Conclusions: The HP acts as a low-risk, adaptable tool for trainees to develop skills in a safe environment. This study demonstrates that professionals perceived the HoloLens as an innovative means to amplify the lived experience voice. Further research will evaluate this additional impact on trainees' confidence and responses to survivors disclosing DA and SV within different disciplines to drive improved outcomes.

Trial Registration: Permission to conduct this study granted by the Research Innovation and Quality Improvement Team at Chelsea and Westminster NHS Foundation Trust

(JMIR Form Res 2025;9:e60075) doi:10.2196/60075

XSL•FO
RenderX

KEYWORDS

augmented reality; virtual reality; medical technology; domestic abuse; sexual violence; medical education; training; domestic violence; violence; assault; victim; survivor-centered; staff; community stakeholders; social care; innovation

Introduction

In the UK, it is estimated that nearly 16.6% of adults aged 16 years and older have experienced sexual assault [1]. Furthermore, it is estimated that 2.1 million people aged 16 years and older experienced domestic abuse (DA) in the year ending March 2023 [2]. These statistics are comparable to the number of patients currently living with type 2 diabetes mellitus [3]. However, unlike diabetes and other medical conditions, there is little evidence discussing the pathogenesis behind DA and sexual assault, thereby making prevention programs, both difficult to design and sustainably fund. With one of the highest rates of repeat victimization than any other crime [4], DA affects 1 in 5 adults during their lifetime [5]. In addition, the detrimental effects of DA are wide-ranging, encompassing physical, mental, behavioral, and reproductive consequences [6]. Survivors of DA seek health care more frequently than individuals who have not experienced abuse [7], varying from presentations related to the physical consequences of abuse to psychological health problems, including depression, anxiety, and post-traumatic stress disorder [8]. Therefore, health care providers have a key role in identifying and supporting survivors. A systematic review of barriers and facilitators to disclosure of abuse to health care professionals (HCPs) found that perception of the HCP's ability to respond to the disclosure was impactful. Any notion that the HCPs did not possess the skills or knowledge to appropriately respond deterred survivors from disclosing their experiences [8]. HCPs are often the first point of contact for survivors; hence, it is vital that they are equipped with the skills and tools to competently identify and support survivors [9].

According to the UK Government, all behaviors, inflicted on anyone perceived to be of a controlling, coercive, harmful, or sexual nature can be defined as DA or sexual violence (SV) [10]. Whilst the majority of survivors of DA are women [2], it is important that we do not neglect the experiences of those from the LGBTQ+ (lesbian, gay, bisexual, transgender, and queer) community, who are also more vulnerable to familial abuse and honor-based violence [11]. There are many barriers to LGBTQ+ individuals seeking support, including discriminative attitudes from the police leading to mistrust, as well as a lack of training for professionals from the criminal justice system and public sector on how to recognize and respond to DA in the LGBTQ+ community [12]. Although the data for England and Wales for the year ending March 2024 showed no statistically significant differences in the prevalence of domestic violence between ethnic groups [2], the literature shows that Black, Asian, and minority ethnic individuals experience multiple barriers in seeking help following abuse, including structural racism, cultural exclusion, and victim-blaming from both formal services and communities [13]. Furthermore, the 2024 statistics for England and Wales showed that a higher proportion of people with a disability experienced DA than those without [2]. These issues should not be considered as separate entities, but rather an intersectional approach must be undertaken, so we can understand how inequalities due to race, sex, disability, class, and sexual identity may compound each other, further preventing survivors from seeking help [11]. These statistics outline the need for services to adopt a survivor-centered approach. Strategic partnerships consulting survivors of all protected characteristics are needed to enable the development of pathways specifically designed to address their psychosocial needs. This is outlined in The Strategic Direction for Sexual Assault and Abuse Services [14], which is an initiative from National Health Service (NHS) England, with the goal of radically improving access to services and support for survivors of sexual abuse to help them heal and rebuild their lives. This requires collaboration between authorities, health services and strategic partners in the voluntary and community sectors as part of a multiagency care model [15].

Existing literature suggests a gap in the knowledge of HCPs on how to identify and respond to survivors of domestic violence and sexual assault, largely due to a lack of training [16,17]. A 2020 metasynthesis of 47 qualitative studies examined health practitioners' readiness to address domestic violence and abuse. Support from the health system was the largest theme, with HCPs suggesting further support was needed through upskilling staff in how to address domestic violence, as well as making enquiries into domestic violence routine and allocating additional time for sensitive conversations [16]. HCPs spoke about the importance of training in how to identify and respond to DA in their own specific clinical setting [16]. Evaluations of existing domestic violence support training programs for HCPs have demonstrated significant improvements in knowledge and confidence in how to identify and respond to survivors [17] and hence, we set out to design a training program adjunct, using a mixed reality device.

The emerging technologies of virtual, augmented, and mixed reality devices create an expansive new realm for education. Virtual reality is an immersive experience where a virtual world replaces the physical world, whereas augmented reality uses technology to superimpose virtual data (visual information) onto the physical world (real-life objects) [18]. Mixed reality is an in-between medium which allows virtual objects to be integrated and anchored onto the physical world [18]. The use of mixed reality headsets, such as the Microsoft Hololens2, is being increasingly used within medical education. There are many advantages to using these devices, including the option of repeated practice of learning, skills, and communication, without adverse effects on the patient [19]. These devices also can be used independently, reducing the time burden of teaching sessions on clinical educators. So far, mixed reality devices have been primarily used for teaching anatomy (through understanding of 3D object layers and components), procedures, and clinical skills [19-21]. We wanted to explore the benefits of using them for DA and SV training. In order to effectively teach HCPs about domestic violence and sexual assault, it is important to use examples based on real-life experiences of

XSL•FO
RenderX

survivors. To ask a survivor to facilitate a training session could be considered unethical, as it may be triggering to the survivor, with potential negative effects on their mental health and well-being; hence, the use of mixed reality devices creates an opportunity to recreate these learning opportunities without adverse effects on survivors.

The Domestic Abuse team at Chelsea & Westminster Hospital NHS Foundation Trust adopts a "whole health" approach by pioneering a longstanding partnership between health care and community stakeholders [22] widely accepted as a best response in acute health. Their model centers on 3 key features: a co-located Domestic Abuse Co-ordinator (DAC) responsible for training workers in DA; the co-location of Independent Domestic Violence Advisors (IDVAs) to provide holistic and specialist support to survivors that are identified; and a DA Lead. Data suggest that survivors 'felt safe' and were more likely to report to an on-site IDVA while in the hospital [23]. Early identification of potential survivors by frontline workers is essential and stresses the importance of high-quality in-house training on the identification, management and referral of survivors of DA and SV.

The aim of this study was to improve the quality of staff training on the identification and management of DA and SV survivors. In collaboration with the Undergraduate Education Team at Chelsea and Westminster Hospital NHS Foundation Trust, the DA Team proposed the use of the Microsoft HoloLens2 as a tool to channel the survivor voice. This evaluation explores both health care and community stakeholder views of the augmented reality headset as an additional learning tool to promote survivor-centered training in a safe setting. We hypothesized that the following experience would demonstrate the Microsoft HoloLens' potential as a teaching adjunct when educating staff on the identification and management of DA and SV survivors.

Methods

We conducted a formative evaluation of the Microsoft HoloLens2, specifically the GiGXR HoloPatient (HP) application, as a training tool within the DA and SV landscape.

Intervention

The Microsoft HoloLens2 is a mixed reality headset which has had a substantial impact in various industries since its release in 2019, including health care, education, and engineering [24]. Dissimilar to virtual reality, the headset is only part-immersive as it displays augmented content superimposed on the real surrounding environment. Organizations like GiGXR have created a series of augmented holograms, known as HPs, that resemble clinically unwell patients. The HP menu listed 17 holograms, each representing a clinical emergency, 2 of which were specifically designed to portray a survivor of DA (Lydia Johnson) and SV (Jenny Li). Each case consists of a series of video clips that depict the patient at different stages of their hospital journey (Figure 1). Aside from Jenny Li who vocalizes her experience via a prerecorded audio clip, all other holograms are nonverbal. As a result, the user has the flexibility to adapt the HP by overlaying a different patient script to suit their educational needs. This application, if used effectively, could expose students to unique training opportunities that are ethically challenging and difficult to recreate in a real clinical environment.

XSL•FO
RenderX

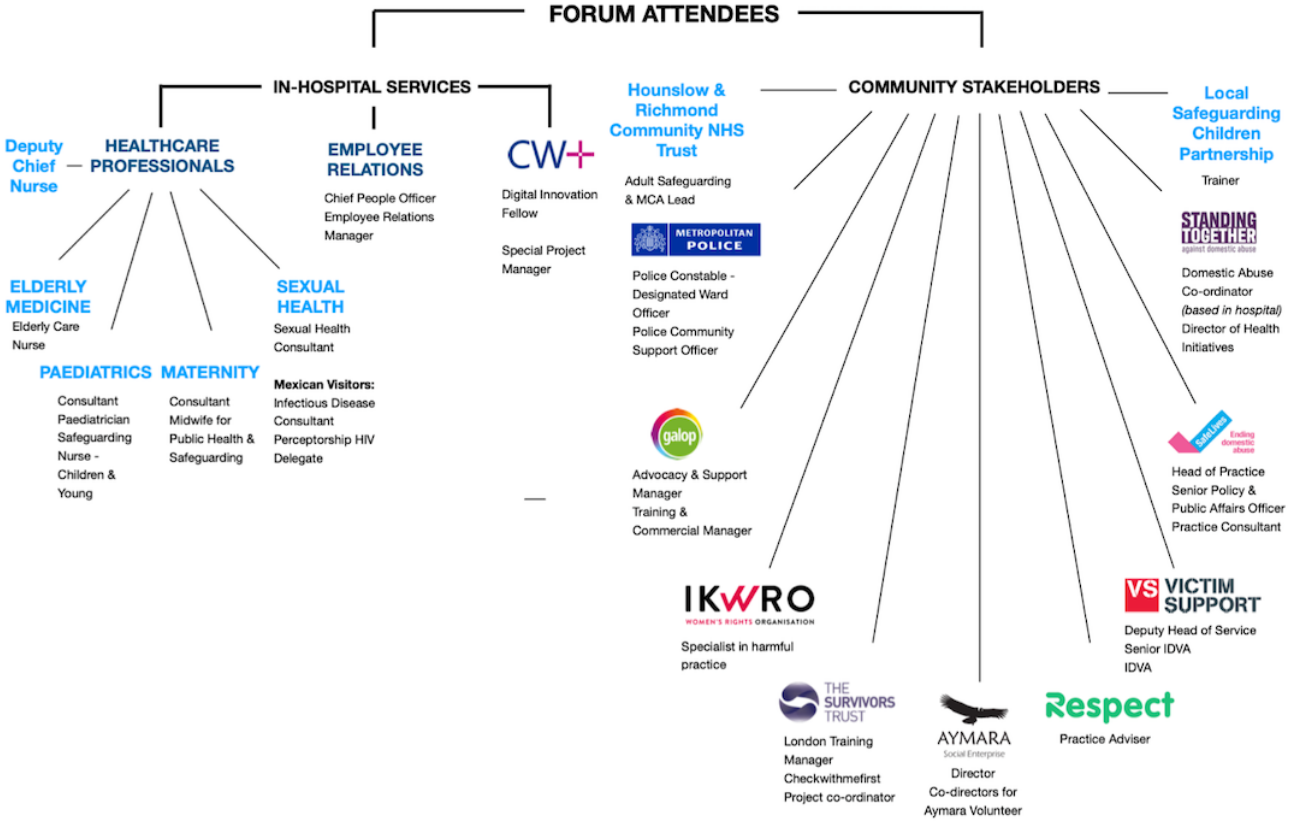
Figure 1. Lydia Johnson, a HoloPatient survivor of domestic abuse and sexual violence. Image A demonstrates Lydia on arrival to the emergency department. Image B displays her as a patient on a ward.

Recruitment

The evaluation was conducted in the form of live demonstrations. These demonstrations were held within a lecture theater at Chelsea and Westminster Hospital. Frontline staff and community stakeholders from the NHS, DA, social care, and law enforcement sectors within the North West London area were invited to 3 HP events, held between February and July 2023 (see spider diagram displaying the range of attendees; Figure 2). Participants from these sectors were invited due to their increased likelihood of encountering survivors in a professional setting and their shared use of North West London Safeguarding guidelines. All participants were invited by email, which described the aim and method of the study in detail (see recruitment email in Multimedia Appendix 1). Participants were reminded that participation is voluntary and that all data were to be anonymized on publication. Written consent was obtained at the stakeholder meeting after participants were given the opportunity to read further information on the HoloLens and its purpose in this study (see consent form in Multimedia Appendix 2).

XSL•FO
RenderX

Figure 2. Spider diagram displaying all the organizations within North West London which attended the stakeholder forums.

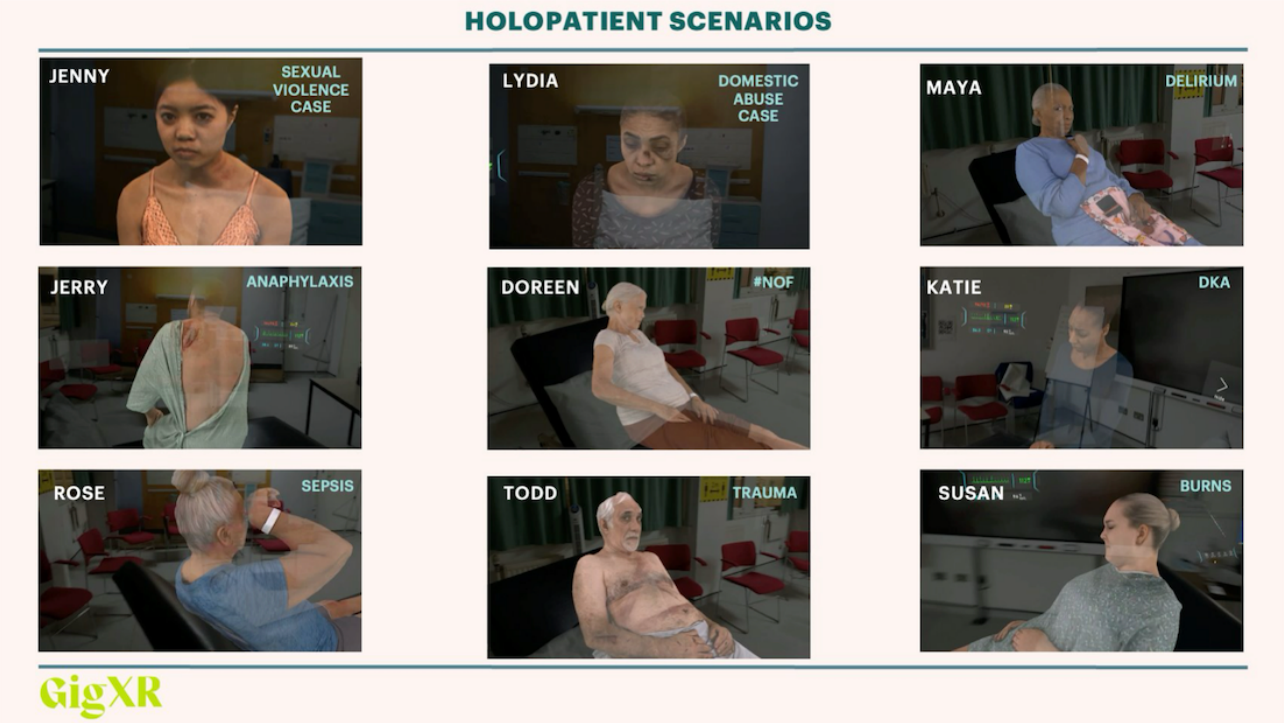


The Demonstrations

Following a brief introduction into augmented reality, attendees were shown data that evidenced the HP as an immersive tool when trialed in an undergraduate setting [25]. The HoloLens was ultimately proposed as a teaching adjunct to be used whilst training frontline workers on the management of DA and SV cases (see full presentation in Multimedia Appendix 3). It was envisaged to be a visual aid that amplified the lived experience, to be used after their initial training with the DAC. The patient voice was to be outlined by personalized scripts, co-created alongside sector charities, ensuring survivor engagement and participation. In total, 9 HPs, different in gender, age, and ethnicity, were projected into the room (Figure 3). Participants were given the opportunity to wear the headset and familiarize themselves with the technology during the sessions.

XSL•FO
RenderX

Figure 3. Nine HoloPatients selected to be displayed during the demonstrations. Their original clinical presentation demarcated in blue writing within each box.



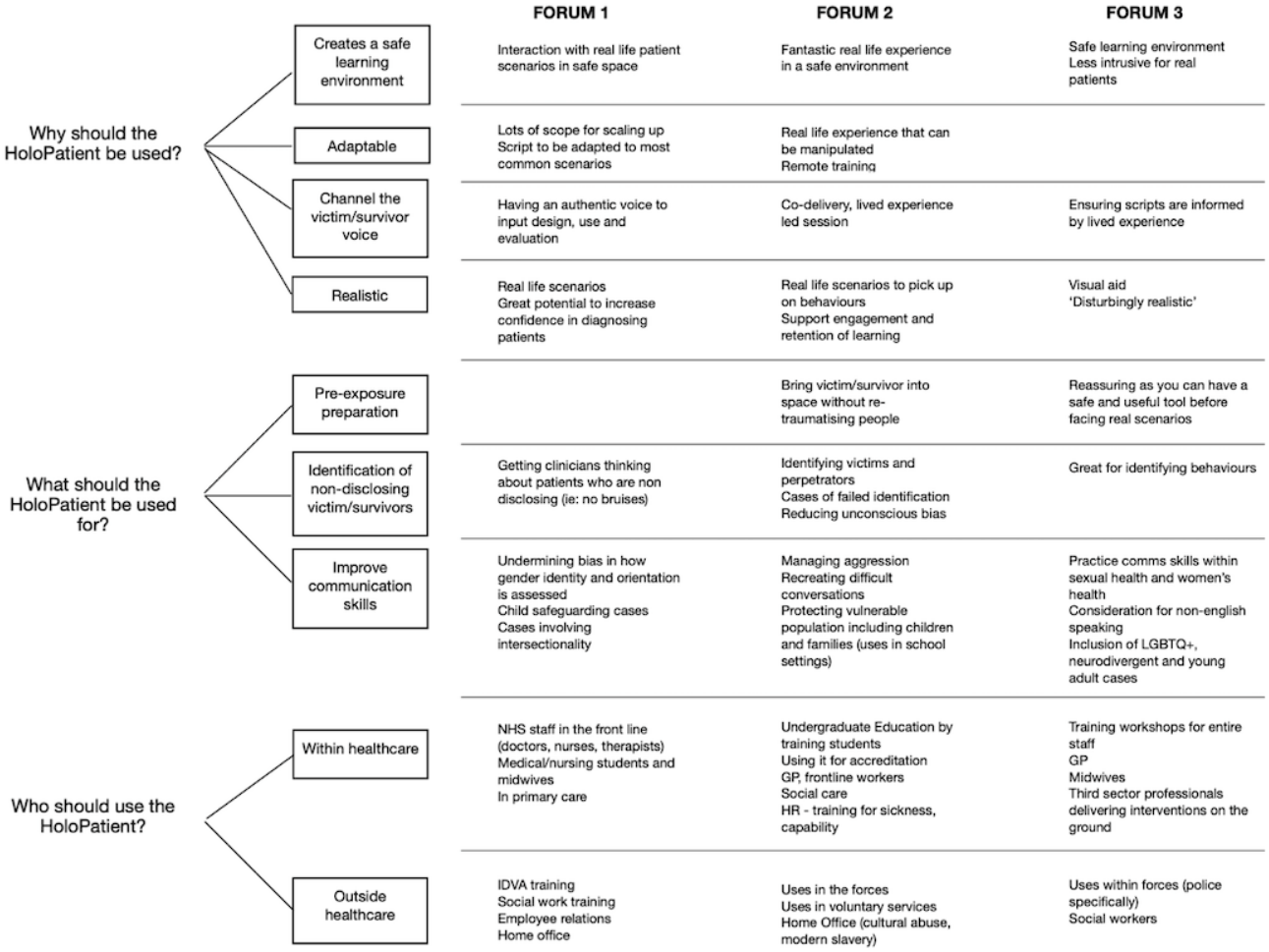
Data Collection and Analysis

A postintervention evaluation research model was used to explore the feasibility and functionality of the HP as an educational tool within DA and SV training. Furthermore, it evaluated the potential for a centralized multiagency augmented reality hub for professionals to access educational opportunities in the future. In order to achieve this, participants were prompted to answer 3 overarching questions: why should the HP be used, what should the HP be used for, and who should use the HP?

Data were collected in the form of a standardized questionnaire to improve reliability (see in Multimedia Appendix 4). It consisted of Likert scaling [26] and open-ended questions to collect quantitative and qualitative data, respectively. Likert scales were used for participants to rate how strongly they agreed or disagreed with the following statements: “HoloLens is an effective tool used in DA and SV,” and “Technological limitations prevent me from using this tool.” Open-ended questions gave participants the opportunity to provide feedback on how they felt wearing the device, if they thought it would be useful in health care or other fields, and if they felt there was value in having a centralized multiagency HoloLens training hub. A basic analysis conducted on the Likert scale questions displayed statistical means and modes. The qualitative data were thematically analyzed using Braun and Clark’s 6-step method [27] (see data sets in Multimedia Appendices 5, 6, and 7). One researcher (medically qualified) read each survey repeatedly to gain familiarity. Initial interpretations were discussed with the project supervisor, and together they extracted evidence from the surveys to refine each theme. Finally, a list of themes was presented with illustrative examples on multiple occasions to the wider team (see coding tree in Figure 4). As a final checking step, draft findings were discussed with a few participants (n=3) prior to publication. To ensure that this study was conducted in accordance with qualitative reporting guidelines, a COREQ checklist was completed (Checklist 1) [28].

XSL•FO
RenderX

Figure 4. Coding tree across all 3 stakeholder forums.



Ethical Considerations

This evaluation was approved by the Research Innovation and Quality Improvement board of Chelsea and Westminster Hospital NHS Foundation Trust under reference number 6038. Written informed consent was obtained from all participants. All study data are anonymous, and HP scripts, although informed by DA and sexual assault survivors, did not contain any identifiable information. Involvement in this study was optional and voluntary, and there was no financial compensation for participants or researchers.

Results

Thematic Analysis

A total of 48 questionnaires were collected across all 3 demonstrations (n=14, 12, 22). Approximately 29% (n=14) of the sample were local HCPs. The remainder consisted of representatives from community charities, local councils, and law enforcement (see spider diagram displaying range of attendees in Figure 2). No participants refused to participate or dropped out.

Analysis of participants' perspectives of the HP as a training tool in DA and SV sectors was divided into 3 overarching themes, based on our research questions: Why should the HP be used? What should the HP be used for? Who should use the HP?

A summary of this section is displayed in Table 1.

XSL•FO
RenderX

Table . Realistic table summarizing qualitative data following thematic analysis

Theme	Reason
Why should the HoloPatient be used?	
Realistic	• "disturbingly realistic" • "impactful" • "increased their sense of reality" • improving "retention of learning" • "more realistic than role-play"
Creates a safe learning environment	• brought "real-life learning experiences into a safe environment" • trainees can repeatedly interact, make errors, stumble • openly discuss case with peers without risk of re-traumatization
Adaptable	• allows "real-life experiences to be manipulated" • "scenarios adapted to each trainee's requirement" • "helpful for all professionals' • remote feature allows "scope to scale up"
Channel the survivor voice	Use of survivor script • "essential" feature that could "bring the survivor into the learning space" • to "incorporate language and cultural specific aspects" • "inclusion of survivors when creating the stories" • cover "common scenarios in different clinical settings and diverse cases"
What should the HoloPatient be used for?	
Pre-exposure preparation	• "safe and useful tool for practice before facing real-life scenarios" "important to understand how to deal with the situation before interacting with survivors" • law enforcement felt this could "support junior officers responding to their first call outs in a safer space."
Identification of nondisclosing survivors	• prepare trainees to identify survivors presenting without visible injury • "masking patients" • those "with a hidden agenda" • "great for identifying behaviors or visible indicators"
Improving communication skills with and when	
Vulnerable patients	• elderly • mental illness • neurodiversity • women • young people • NHS staff or other staff
LGBTQ + survivors	• "look at undermining bias in how people are assessed in regards to their their gender identity and sexual orientation" • support exploration of "intimate partner violence" and "hate crime" • have an "LGBTQ + lived experience voice" in the scenarios
Engaging with perpertrators	• "look at undermining bias in how people are assessed in regards to their their gender identity and sexual orientation"
Scenario could be adapted to	• where the survivor is "aggressive and protective of the perpetrator" • "on how to approach a survivor and respond appropriately with a perpetrator present"
Recreating difficult conversations	• "practicing mental capacity assessments" • "decision-making, end-of-life care planning" • "dealing with confrontation and bullying and harassment cases" • demonstrate "an example of an incorrect way to deal with a situation" • encourage "open discussion"

XSL•FO
RenderX

Theme	Reason
Who should use the HoloPatient?	
Within health care	
Undergraduate education	• "medical and nursing students" • training "first with the HoloPatient and then a certificate before the real practice"
Maternity	• "the risk of domestic abuse during pregnancy" • "trust building that happens between patient and midwife"
Other frontline staff	• 'all health professionals who may come into contact with a survivor' • 'no them and us' • targeting 'third sector professionals who intend to deliver interventions on the ground' • could be used in IDVA[a] training to 'significantly reduce the time taken to train'
Outside of health care	
Social care sector	• social workers • paid carers • other care home staff
Public sector	• within law enforcement • Home Office to recreate cases involving "minority ethnic survivors," "honour-based abuse" and "modern slavery.'

[a]IDVA: Independent Domestic Violence Advisor

Why Should the HoloPatient Be Used?

Realistic

Many attendees found the HP "disturbingly realistic" (Participant 1 (P1)) and were impressed with how "impactful" it was (P2, P3, and P4) (participant references taken from data sets—see Multimedia Appendices 5, 6, and 7). Attendees found that the HP "increased their sense of reality" (P5 and P6) hence improving engagement and "retention of learning" (P7). One participant found the HP "more realistic than role-play" (P2), as dynamics could be affected if a trainee is familiar with their actor or mentor.

Creates a Safe Learning Environment

Participants felt the HP brought "real-life learning experiences into a safe environment" (P8), where both trainees and survivors are better protected. In contrast to role-play, trainees can repeatedly interact, make errors, stumble, and openly discuss the HP's case with peers without risk of re-traumatization.

Adaptable

Attendees acknowledged that the HP allows "real-life experiences to be manipulated" (P9) by using "scenarios adapted to each trainee's requirement" (P1). As a result, a majority of participants saw it as a tool helpful for all professionals. HP demonstrations were live-streamed via Microsoft Teams to remote participants. One participant called this feature a "key benefit" (P10), which added "scope to scale up."

Channel the Survivor Voice

The concept of the survivor script was presented to all participants at the HoloLens demonstrations. The scripts received positive feedback, with stakeholders seeing it as an "essential" feature that could "bring the survivor into the learning space" (P11). The idea was to overlay the HP scenario with a voice-over, either prerecorded or delivered by a survivor facilitator on the day. Participants particularly encouraged the use of the facilitator's voice "to ensure that nuance isn't lost" (P9). The potential to create scripts that cover "common scenarios in different clinical settings (eg, Emergency Department, sexual health)" (P12) and "diverse cases" with different protected characteristics were discussed. The idea to "incorporate language and cultural specific aspects" (P7) into the scripts was strongly considered. Attendees agreed with the "inclusion of survivors when creating the stories" to ensure that the scripts are "informed by a lived experience" (P3), with agreement that "specialist domestic abuse" (P2) input was necessary to maintain the use of trauma-informed language. The creation of "an authentic voice advisory group" (P13) for their overall evaluation was suggested.

Overall, all stakeholders strongly agreed (n=45) or somewhat agreed (n=3) that the Microsoft HoloLens 2 could be an effective tool to deliver training on the identification and management of DA and SV survivors.

What Should the HoloPatient Be Used For?

Pre-Exposure Preparation

Stakeholders saw the HoloLens as a "safe and useful tool for practice before facing real-life scenarios" (P1). One participant described Jenny, the hologram of a SV survivor, as "incredibly uncomfortable to watch" (P14) explaining that it was "important to understand how to deal with the situation before interacting with survivors" (P14). Themes of realism and the HP's ability

XSL•FO
RenderX

to create a safe learning environment were commonly mentioned in conjunction with this theme. This is something that respondents cited clinicians often do not have access to, following any training. The HoloLens offers a training tool to embed the teaching from the DAC, before practicing in real life with patients. Law enforcement similarly felt this could support junior officers responding to their first call-outs in a safer space.

Identification of Nondisclosing Survivors

The HP's potential to prepare trainees to identify survivors presenting without visible injury became a reoccurring theme across all demonstrations. One participant described the device as "great for identifying behaviors or visible indicators" (P2).

Improve Communication Skills

Communicating With Vulnerable patients

Once all the HPs were displayed, participants were prompted to suggest potential scenarios that could aid the identification of vulnerable survivors of DA or SV. Public Health England defines vulnerability as "in need of special care, support, or protection because of age, disability, risk of abuse, or neglect" [29]. Categorized into these subgroups, the suggested scenarios are listed below in ascending order of prevalence based on the questionnaire data: Elderly patients (n=2): "delirium," "care-home scenarios"; women (n=4): "pregnant women," "female genital mutilation"; patients with mental illness (n=7): "suicidal tendencies," "eating disorders," and "psychosis"; patients with visible and nonvisible disability (n=8): "scenarios that include adults with learning disability," and "involving paid carers"; children and adolescents (n=16): "whole family scenarios," "involving teachers," "preparing families for court proceedings," "forced marriage," and "substance abuse".

Communicating With LGBTQ+ Survivors

A participant stated that "generalist staff have not always got the skills to deal with LGBTQ+ survivors" (P15). In response, they suggested the potential use of the HP to "look at undermining bias in how people are assessed in regards to their gender identity and sexual orientation" (P15). They saw how this device could support exploration of "intimate partner violence" and "hate crime" by editing the scenarios to have an "LGBTQ+ lived experience voice" (P15). A few participants highlighted the absence of any transgender HPs (P1, P15, P16, and P17). When asked to suggest designs for new HPs, 22% (n=11) of participants said LGBTQ+ HPs.

Dealing With Perpetrators

In the first stakeholder meeting, a scenario was adapted to include a perpetrator by getting a facilitator to interact with the HP by standing in the background. They are visible to the HoloLens user and all participants watching the live stream. It was suggested that the scenario could be adapted to "challenge the stereotype of a passive survivor" (P9) where the survivor is "aggressive and protective of the perpetrator" (P9). Stakeholders saw its training potential within the police force (n=19), "specifically on how to approach a survivor and respond appropriately with a perpetrator present" (P3). However, concerns with safety were highlighted. One participant stated that "we need to be clear on the expectation around engaging with a perpetrator and how to do it safely if at all" (P9). This was seen as a key advantage of the HoloLens tool, as this could be explored in a safe space.

Recreating Difficult Conversations

Stakeholders listed "practicing mental capacity assessments, decision-making, end-of-life care planning, dealing with confrontation, and bullying and harassment cases" (P18) as challenging scenarios that could be practiced with the HoloLens. One participant suggested to display "an example of an incorrect way to deal with a situation" (P8) as a teaching tool to encourage "open discussion" (P18).

Used in a Central Multiagency Training Hub

During the final demonstration, which took place in July 2023, attendees (n=22) were asked whether they supported the creation of a central multiagency training hub, potentially at Chelsea and Westminster NHS Foundation Trusts, which incorporated the use of augmented and virtual reality headsets. The intention of the hub is to provide community and health care services within the North West London region, shared access to the technology, and other training materials. This allows for a standardized training approach with greater opportunity for communication between teams. Overall creating more collaborative DA and SV multi-agency services in the future. A majority of participants either strongly (n=17) or somewhat agreed (n=2) in the creation of a central hub. It was stated that "sculpting multiagency responses into this (central hub) would be incredible and will change the health care landscape. The HoloLens can be used to continuously improve practice by dropping into an immersion center" (P4). One participant neither agreed nor disagreed with the hub, and 2 participants did not vote.

Who Should Use the HoloPatient?

Within Health Care

Undergraduate Education

When asked where this technology could be most useful, approximately 14% (n=7) of participants suggested its use in undergraduate education. In particular, "medical and nursing students" (P5, P18, P19, P20, P21, and P22). One attendee suggested using the HoloLens for accreditation by training "first with the HP and then a certificate before the real practice" (P20).

Maternity

At least one participant per demonstration (total n=3) acknowledged its need in midwifery due to "the risk of DA during pregnancy and the trust building that happens between patient and midwife. Additional attendees (n=5) saw potential for the HP to be used in "IDVA training" (P4), indicating it could significantly reduce the time taken to train if this were used in routine practice. IDVAs are trained specialists that work with clients "from the point of crisis to assess, discuss, and develop safety plans"[30].

Other Frontline Staff

Approximately 52% (n=25) of attendees felt as though the technology was suitable for "all health professionals who may come into contact with a survivor" (P9, P23). "No them and

XSL•FO
RenderX

us" (P24) was stated by one participant. Particular attention was given to primary care (n=6), emergency, and sexual health service staff. Other scoping included targeting "third sector professionals who intend to deliver interventions on the ground" (P7). One participant stated that training "could be in real-time with the holograms" (P5), providing clinicians with "authentic role-play" (P5) that was "less intrusive than real patients" (P5). This is a unique concept when compared to existing training programs, which mainly consist of digital modules at a beginner level [31].

Outside of Health Care

Social Care Sector

At all 3 demonstrations, over one-third of attendees (n=16) named social workers, paid carers, and other care home staff as members of the public who would benefit from DA and SV training with the HoloLens.

Public Sector

Similarly, around 39% (n=19) of participants described the HP's potential to train staff within law enforcement. Particularly those who "attend a domestic abuse emergency" (P12).

The use of the technology in the Home Office was discussed, especially to recreate cases involving "minority ethnic survivors" (P18), "honor-based abuse" (P2, P24, and P25), and "modern slavery" (P17 and P18).

Discussion

Study Findings and Comparison With Previous Works

The findings of this stakeholder evaluation indicate that despite minor technological limitations, all stakeholders saw the Microsoft HoloLens2 as a novel and effective tool to deliver training on DA and SV, within health, social care, law enforcement, and specialist settings. The data encourage the use of co-created survivor scripts alongside the HPs, to create a training program that amplifies their voice and lived experience. This enables the training to remain survivor-centered without needing a survivor present in the room as a facilitator. This minimizes the potential for adverse effects on the survivor.

A common theme across the groups was the ability to identify behaviors that signaled abuse in patients who did not have obvious physical signs of abuse such as bruising or other physical injury. The literature highlights how survivors often do not freely disclose abuse and often actively try to hide it [7], and hence being able to identify the more subtle signs is key to having good and comprehensive training on how to identify survivors. Positive feedback on the ease of use of the HoloLens device was received from all 3 groups, spanning health care workers, law enforcement, and the charity sector, demonstrating that this is a widely acceptable medium to deliver training.

Stakeholders suggested that the HPs were useful for pre-exposure preparation, which aligns with how we envisage it being used in the future. There was a strong theme of realism, with participants describing how the holograms had a profound emotional impact, which we hope will mirror some of the thoughts and feelings experienced when encountering survivors in real life. Enabling this first encounter to occur in a safe and simulated environment empowers participants to trial different approaches to managing the scenario without the risk of harm to a survivor so that they are better equipped to appropriately handle these situations in real life.

The existing literature clearly shows that one of the key barriers to recognizing and supporting survivors is a lack of training [12,16,17]. Ideally, this training needs to be embedded into the undergraduate curriculum of HCPs, so that doctors and nurses graduate with the skills to appropriately manage these encounters. We have demonstrated the potential of using the HoloLens2 in a professional setting and hope this could also be used as a tool for educating undergraduate health care students.

Limitations

We discovered numerous limitations within our project, both related to the Hololens2 device itself and to the application. Wi-Fi connectivity poses a significant issue. One attendee stated that the HoloLens is a "useful tool when in a Wi-Fi range." To open the HP application on the Microsoft HoloLens headset, the device must be connected to a high-speed wireless network. This is a key limitation, as internet connectivity is not reliable, especially in the community. To overcome this, the Undergraduate Education Team at Chelsea and Westminster Hospital has used the smartphone "Wi-Fi hotspot" feature to supply a mobile connection. If the HPs are to be rolled out as a training program more widely, then considerations will need to be made into the creation of a downloadable version that can be operated without an internet connection.

We also found the limited battery life of the headset was a barrier to producing longer or back-to-back training sessions. Microsoft estimate the HoloLens' battery life to have 2 - 3 hours of active use [32]. As both the Microsoft Teams and GiGXR HP applications were open during the demonstrations, it caused the device to malfunction at approximately 1.5 hours. This was highlighted as "an issue" by participants. One user felt that it hindered their ability to "maximize opportunity when using it in the training room." Microsoft has stated that the devices are fully functional while charging [32]. Therefore, the Undergraduate Education Team inserted a portable battery pack into the USB-C port to prolong battery life during the demonstrations. Despite these technological limitations, 64% of participants felt as though it did not affect their opinion of the HoloLens. Overall statistics demonstrated that 14% stated otherwise. while 23% remained impartial. Another limitation of the device is that as it uses more power, it heats significantly. This poses potential safety risks and makes the device less comfortable to wear. To combat this, training sessions will need to be limited to shorter periods of time. We would suggest not using the device for more than an hour at a time.

A further limitation of the HP application was the lack of diversity of the characters. Most participants (roughly 57%, n=27) stated that they would have liked a wider variety of "visible clues" to represent intersectional characters [33]. For example, "different ethnicities, cultural headwear, visible disabilities, walking aids, transgender patients, pregnant women, and children." Unfortunately, this would require the HP menu to be redesigned, at significant cost, which remains a limiting

XSL•FO
RenderX

factor when considering scalability. However, augmented reality allows the hologram's background to be adjusted to the user's preference therefore walking aids and other visual accessories can be added, in an alternative manner or with a real-life person in the scenario, either as a friend, a family member or a perpetrator.

We must also consider limitations related to the study design. Data were collected as postintervention questionnaires. Although this provided insights into how the HPs could be used, the questionnaire was fairly brief, consisting of 13 questions. A focus group of a random sample of participants may have provided more detailed insights into the utility and acceptability of the Hololens2, and the HPs as a training tool for DA and SV. Furthermore, a larger sample size would allow for more reliable results.

Conclusions

This technology has the potential to educate professionals on the use of trauma-informed language while challenging preconceived assumptions around protected characteristics in safe settings. Our evaluation has demonstrated that the Hololens2 is a widely acceptable tool for DA and SV training, that can be used as an adjunct to existing training programs. There is currently very little research on how virtual, mixed, and augmented reality devices can be used for training in DA and SV; hence, we hope the positive results from our evaluation will make way for further large-scale studies [34]. As the HoloLens2 can be operated independently, it gives scope for training sessions to be delivered with minimal staffing requirements.

Virtual reality companies like SimX have developed DA-specific virtual reality patient simulations. In the future we would want to evaluate the efficacy of these modalities as a teaching adjunct in the DA and SV space with the aim to create a core framework applicable to all future mixed and virtual reality technologies [35,36]. It could be of additional direct benefit to employee relation team training around sexual misconduct following the release of NHS England's sexual safety organizational charter [37]. The goal will be to roll out augmented reality Holosurvivors as a standard teaching tool, in addition to DACs, supporting the best multiagency team public health response to DA and SV in the UK.

Acknowledgments

We would like to thank GiGXR, who own the rights to the application and holograms, for granting us permission to evaluate their product throughout this study. We wish to acknowledge our hospital charity CW+, who enabled the purchasing of the technology to assist with teaching and training in the Trust, supported the stakeholder forums, and will be collaborating on future research with our domestic abuse team. Additionally, we would like to thank our stakeholders; Standing Together Against Domestic Abuse, Victim Support, Galop, SafeLives, Respect, IKWRO Women's Rights Organisation, Aymara, Metropolitan Police, Adult Community Safeguarding, and Children's Social Care for their attendance and much valued feedback during the stakeholder forums. Finally, we would like to praise our Undergraduate Education Team and our Domestic Abuse team at Chelsea and Westminster NHS Foundation Trust for their hard work and dedication on this project.

The authors attest that there was no use of generative artificial intelligence (AI) technology in the generation of text, figures, or other informational content of this manuscript.

Data Availability

The authors confirm that the data supporting the findings of this study are available within the accompanying multimedia appendix (see Multimedia Appendix 5-Multimedia Appendix 7).

Authors' Contributions

DK, JW, AM, KD, BB, NK, and CC conceptualized the study. DK, AM, KD, BB, and CC contributed to methodology; validation was done by DK, BB, and CC. Formal analysis was done by DK and CC. Investigation was done by DK, JW, AM, KD, BB, NK, and CC. DK, JW, AM, KD, BB, NK, and CC contributed to resources. DK and CC involved in writing – original draft. DK, JC, and CC involved in writing – review & editing. Visualization was done by DK and CC. CC involved in supervision. DK and CC involved in project administration.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sample recruitment email sent to stakeholders.
[PDF File, 102 KB - formative_v9i1e60075_app1.pdf]

Multimedia Appendix 2
Sample participant information and consent form.

XSL•FO
RenderX

[PDF File, 104 KB - formative_v9i1e60075_app2.pdf]

Multimedia Appendix 3
Powerpoint presentation delivered by Undergraduate Education Team during HoloLens demonstrations.
[PDF File, 3454 KB - formative_v9i1e60075_app3.pdf]

Multimedia Appendix 4
Sample questionnaire given to participants during the HoloLens demonstrations.
[PDF File, 106 KB - formative_v9i1e60075_app4.pdf]

Multimedia Appendix 5
Questionnaire data from the first HoloLens demonstration (23/2/23).
[XLSX File, 12 KB - formative_v9i1e60075_app5.xlsx]

Multimedia Appendix 6
Questionnaire data from the second HoloLens demonstration (26/4/23).
[XLSX File, 13 KB - formative_v9i1e60075_app6.xlsx]

Multimedia Appendix 7
Questionnaire data from the third HoloLens demonstration (13/7/23).
[XLSX File, 17 KB - formative_v9i1e60075_app7.xlsx]

Checklist 1
COREQ checklist
[PDF File, 396 KB - formative_v9i1e60075_app8.pdf]

References

1. Office for National Statistics. Sexual offences in England and Wales overview - Office for National Statistics. 2023. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/sexualoffencesinenglandandwalesoverview/march2022 [accessed 2025-03-03]
2. Office for National Statistics. Domestic Abuse in England and Wales Overview. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/domesticabuseinenglandandwalesoverview/november2023 [accessed 2024-12-09]
3. Diabetes UK. Diabetes is Serious. 2022. URL: https://diabetes-resources-production.s3.eu-west-1.amazonaws.com/resources-s3/public/2022-04/Diabetes%20is%20Serious%20Report%20Digital_0.pdf [accessed 2025-03-03]
4. Government Equalities Office. Strengthening women's voices in government: a response to the public consultation. 2011. URL: https://assets.publishing.service.gov.uk/media/5a78ae7b40f0b62b22cbbf11/strengthening-womens-voices.pdf [accessed 2024-02-28]
5. Domestic abuse in england and wales overview: November 2022 (2022). Census 2021. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/domesticabuseinenglandandwalesoverview/november2022 [accessed 2024-04-17]
6. World Health Organization. Understanding and addressing violence against women: Intimate partner violence: World Health Organization. 2012. URL: https://apps.who.int/iris/handle/10665/77432 [accessed 2025-03-03]
7. Tarzia L, Bohren MA, Cameron J, et al. Women's experiences and expectations after disclosure of intimate partner abuse to a healthcare provider: a qualitative meta-synthesis. BMJ Open 2020 Nov 27;10(11):e041339. [doi: 10.1136/bmjopen-2020-041339] [Medline: 33247027]
8. Heron RL, Eisma MC. Barriers and facilitators of disclosing domestic violence to the healthcare service: a systematic review of qualitative research. Health Soc Care Community 2021 May;29(3):612-630. [doi: 10.1111/hsc.13282] [Medline: 33440034]
9. Dheensa S, Halliwell G, Daw J, Jones SK, Feder G. "From taboo to routine": a qualitative evaluation of a hospital-based advocacy intervention for domestic violence and abuse. BMC Health Serv Res 2020 Feb 21;20(1):129. [doi: 10.1186/s12913-020-4924-1] [Medline: 32085771]
10. Department of Justice. Domestic and Sexual Violence and Abuse. 2016. URL: https://www.justice-ni.gov.uk/articles/domestic-and-sexual-violence-and-abuse [accessed 2025-03-03]
11. McGregor K. Domestic violence and abuse in LGBTQ+ communities. In: Ali P, Rogers MM, editors. Gender-Based Violence: A Comprehensive Guide: Springer; 2023. [doi: 10.1007/978-3-031-05640-6_34]

XSL•FO
RenderX

12. Wellock S, Tarpey E. LGBTQ + Community perceptions of domestic abuse, responsivity and inclusion within the criminal justice system. Sexual and Gender Diversity in Social Services 2024 Jul 2;36(3):354-378. [doi: 10.1080/10538720.2023.2267480]
13. Hulley J, Bailey L, Kirkman G, et al. Intimate partner violence and barriers to help-seeking among Black, Asian, minority ethnic and immigrant women: a qualitative metasynthesis of global research. Trauma Violence Abuse 2023 Apr;24(2):1001-1015. [doi: 10.1177/15248380211050590] [Medline: 35107333]
14. NHS England. Strategic direction for sexual assault and abuse services: lifelong care for victims and survivors: 2018-2023. 2018 URL: https://www.england.nhs.uk/wp-content/uploads/2018/04/strategic-direction-sexual-assault-and-abuse-services.pdf [accessed 2025-03-03]
15. Liu Y, Dong H, Zhang L, El Saddik A. Technical evaluation of HoloLens for multimedia: a first look. IEEE MultiMedia 2018;25(4):8-18. [doi: 10.1109/MMUL.2018.2873473]
16. Hegarty K, McKibbin G, Hameed M, et al. Health practitioners' readiness to address domestic violence and abuse: a qualitative meta-synthesis. PLoS ONE 2020;15(6):e0234067. [doi: 10.1371/journal.pone.0234067] [Medline: 32544160]
17. Cleak H, Hunt G, Hardy F, Davies B, Bell J. Health staff responses to domestic and family violence: the case for training to build confidence and skills. Australian Social Work 2021 Jan 2;74(1):42-54. [doi: 10.1080/0312407X.2020.1808029]
18. Lungu AJ, Swinkels W, Claesen L, Tu P, Egger J, Chen X. A review on the applications of virtual reality, augmented reality and mixed reality in surgical simulation: an extension to different kinds of surgery. Expert Rev Med Devices 2021 Jan;18(1):47-62. [doi: 10.1080/17434440.2021.1860750] [Medline: 33283563]
19. Barteit S, Lanfermann L, Bärnighausen T, Neuhann F, Beiersmann C. Augmented, mixed, and virtual reality-based head-mounted devices for medical education: systematic review. JMIR Serious Games 2021 Jul 8;9(3):e29080. [doi: 10.2196/29080] [Medline: 34255668]
20. Banjar A, Xu X, Iqbal MZ, Campbell A. A systematic review of the experimental studies on the effectiveness of mixed reality in higher education between 2017 and 2021. Computers & Education: X Reality 2023 Dec;3:100034. [doi: 10.1016/j.cexr.2023.100034]
21. Kolecki R, Pręgowska A, Dąbrowa J, et al. Assessment of the utility of mixed reality in medical education. Translational Research in Anatomy 2022 Sep;28:100214. [doi: 10.1016/j.tria.2022.100214]
22. Chelsea and Westminster Hospital NHS Foundation Trust. The Queen Consort visits our domestic abuse service. 2022. URL: https://www.chelwest.nhs.uk/about-us/news/news-archive/2022/the-queen-consort-visits-our-domestic-abuse-service [accessed 2024-02-28]
23. Halliwell G, Dheensa S, Fenu E, et al. Cry for health: a quantitative evaluation of a hospital-based advocacy intervention for domestic violence and abuse. BMC Health Serv Res 2019 Oct 21;19(1):718. [doi: 10.1186/s12913-019-4621-0] [Medline: 31638998]
24. Park S, Bokijonov S, Choi Y. Review of Microsoft HoloLens applications over the past five years. Appl Sci (Basel) 2021;11(16):7259. [doi: 10.3390/app11167259]
25. Karunaratne D, Moore A, Dasigan K, Bartholomew B. Exploring the use of the hololens as an innovative means to facilitate early clinical exposure in medical school. : MedEd Publish; 2023. [doi: 10.21955/mep.1115444.1]
26. Likert R. A technique for the measurement of attitudes. Arch Psychol 1932;22(140):55.
27. Braun V, Clarke V. Thematic analysis: APA handbooks in psychology®. In: Handbook of Research Methods in Psychology Ed: American Psychological Association; 2012:57-71. [doi: 10.1037/13620-004]
28. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
29. Office for Health Improvement & Disparities. Vulnerabilities: applying All Our Health. 2022. URL: https://www.gov.uk/government/publications/vulnerabilities-applying-all-our-health/vulnerabilities-applying-all-our-health [accessed 2025-03-03]
30. SafeLives. National definition of IDVA work. 2014. URL: https://safelives.org.uk/sites/default/files/resources/National%20definition%20of%20IDVA%20work%20FINAL.pdf [accessed 2025-03-03]
31. e-lfh. E-learning for Domestic Violence and Abuse. 2020. URL: https://www.e-lfh.org.uk/programmes/domestic-violence-and-abuse-elearning-for-health-visitors-and-nurses [accessed 2025-03-03]
32. Microsoft. About HoloLens 2. 2023. URL: https://learn.microsoft.com/en-us/hololens/hololens2-hardware [accessed 2025-03-03]
33. Evans E. Diversity matters: intersectionality and women's representation in the USA and UK: Table 1. Parliam Aff 2016 Jul;69(3):569-585. [doi: 10.1093/pa/gsv063]
34. Xue J, Hu R, Zhang W, et al. Virtual reality or augmented reality as a tool for studying bystander behaviors in interpersonal violence: scoping review. J Med Internet Res 2021 Feb 15;23(2):e25322 [FREE Full text] [doi: 10.2196/25322] [Medline: 33587044]
35. SimX. Domestic Violence Virtual Reality Medical Simulation. 2024 Sep 6. URL: https://www.simxvr.com/marketplace/domestic-violence-2/ [accessed 2024-12-17]
36. Hobson L. Virtual reality used to help domestic abuse in north east courts. BBC News. 2024 Sep 16 URL: https://www.bbc.co.uk/news/articles/cp8n67wxgjro [accessed 2025-03-03]

XSL•FO
RenderX

37. Russell S. Sexual safety in healthcare – organisational charter.: NHS England; 2023 Sep 4 URL: https://www.england.nhs.uk/publication/sexual-safety-in-healthcare-organisational-charter [accessed 2025-03-03]

Abbreviations

DA: domestic abuse
HCP: health care professional
HP: HoloPatient
IDVA: Independent Domestic Violence Advisor
NHS: National Health Service
SV: sexual violence

Edited by A Mavragani; submitted 30.04.24; peer-reviewed by P Vemavarapu, V Baker; revised version received 26.01.25; accepted 28.01.25; published 19.03.25.

Please cite as:
Karunaratne D, Whittock J, Moore A, Dasigan K, Chevolleau J, Bartholomew B, Kelly N, Cohen CE
Exploring the Use of an Augmented Reality Device Learning Tool for Multidisciplinary Staff Training on Domestic Abuse and Sexual Violence: Postintervention Qualitative Evaluation
JMIR Form Res 2025;9:e60075
URL: https://formative.jmir.org/2025/1/e60075
doi:10.2196/60075

© Dilroshini Karunaratne, Jessica Whittock, Amber Moore, Krishna Dasigan, Jasmine Chevolleau, Brent Bartholomew, Nikki Kelly, Charlotte E Cohen. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Optimizing Testimonials for Behavior Change in a Digital Intervention for Binge Eating: Human-Centered Design Study

Isabel R Rooper[1,2], BA; Adrian Ortega[1,3], PhD; Thomas A Massion[1,2], MA; Tanvi Lakhtakia[1,3], BA; Macarena Kruger[1,3], BS; Leah M Parsons[1,2], BS; Lindsay D Lipman[1,2], BA; Chidiebere Azubuike[1,2], MS; Emily Tack[1,3], BA; Katrina T Obleada[4,5], PhD; Andrea K Graham[1,2], PhD

[1]Center for Behavioral Intervention Technologies, Feinberg School of Medicine, Northwestern University, 750 N Lake Shore Dr, 10th Floor, Chicago, IL, United States

[2]Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[3]Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[4]Department of Psychiatry and Behavioral Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[5]Potocsnak Family Division of Adolescent and Young Adult Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Lurie Children's Hospital, Chicago, IL, United States

Corresponding Author:
Andrea K Graham, PhD
Center for Behavioral Intervention Technologies, Feinberg School of Medicine, Northwestern University, 750 N Lake Shore Dr, 10th Floor, Chicago, IL, United States

Abstract

Background: Testimonials from credible sources are an evidence-based strategy for behavior change. Behavioral health interventions have used testimonials to promote health behaviors (eg, physical activity and healthy eating). Integrating testimonials into eating disorder (ED) interventions poses a nuanced challenge because ED testimonials can promote ED behaviors. Testimonials in ED interventions must therefore be designed carefully. Some optimal design elements of testimonials are known, but questions remain about testimonial speakers, messaging, and delivery, especially for ED interventions.

Objective: We sought to learn how to design and deliver testimonials focused on positive behavior change strategies within our multisession digital binge eating intervention.

Methods: We applied human-centered design methods to learn users' preferences for testimonial speakers, messaging, and delivery (modalities, over time, and as "nudges" for selecting positive behavior change strategies they could practice). We recruited target users of our multisession intervention to complete design sessions. Adults (N=22, 64% self-identified as female; 32% as non-Hispanic Black, 41% as non-Hispanic White, and 27% as Hispanic) with recurrent binge eating and obesity completed individual interviews. Data were analyzed using methods from thematic analysis.

Results: Most participants preferred designs with testimonials (vs without) for their motivation and validation of the intervention's efficacy. A few distrusted testimonials for appearing too "commercial" or personally irrelevant. For speakers, participants preferred sociodemographically tailored testimonials and were willing to report personal data in the intervention to facilitate tailoring. For messaging, some preferred testimonials with "how-to" advice, whereas others preferred "big picture" success stories. For delivery interface, participants were interested in text, video, and multimedia testimonials. For delivery over time, participants preferred testimonials from new speakers to promote engagement. When the intervention allowed users to choose between actions (eg, behavioral strategies), participants preferred testimonials to be available across all actions but said that selectively delivering a testimonial with one action could "nudge" them to select it.

Conclusions: Results indicated that intervention users were interested in testimonials. While participants preferred sociodemographically tailored testimonials, they said different characteristics mattered to them, indicating that interventions should assess users' most pertinent identities and tailor testimonials accordingly. Likewise, users' divided preferences for testimonial messaging (ie, "big picture" vs "how-to") suggest that optimal messaging may differ by user. To improve the credibility of testimonials, which some participants distrusted, interventions could invite current users to submit testimonials for future integration in the intervention. Aligned with nudge theory, our findings indicate testimonials could be used as "nudges" within interventions—a ripe area for further inquiry—though future work should test if delivering a testimonial only with the nudged choice improves its uptake. Further research is needed to validate these design ideas in practice, including evaluating their impact on behavior change toward improving ED behaviors.

(JMIR Form Res 2025;9:e59691) doi:10.2196/59691

XSL•FO
RenderX

KEYWORDS

health behavior; health narratives; binge eating; user engagement; personalization; behavior change; digital health; intervention; human-centered design; behavioral health; preferences

Introduction

Testimonials are an evidence-based behavior change strategy with beneficial impacts on health behaviors. Often characterized by the first-person voice of a typical person [1], testimonials can explain an individual's experience with a health condition or intervention and can be integrated as part of an intervention's didactic approach. Testimonial speakers are often credible *sources* (eg, someone who has used an intervention), which is useful because highlighting credible sources is an evidence-based behavior change technique [2]. A meta-analysis by Xu [3] demonstrated that health narratives like testimonials have comparable effectiveness to statistical evidence for supporting behavior change. Specifically, testimonials produce effects on behavioral intentions and are persuasive for health beliefs and attitudes [4]. There is support for using testimonials to encourage health behavior change in a range of health-promoting interventions, including for physical activity [5,6], healthy eating [7], cancer screening [8], smoking cessation [9], and diabetes self-management [10].

In the literature about eating disorders (EDs), the behavioral impacts of testimonials are more nuanced. A small systematic review on mental health recovery narratives identified that in the context of anorexia nervosa, exposure to ED recovery testimonials can increase ED behaviors in testimonial viewers, indicating that particular care should be taken to avoid potentially triggering content in testimonials [11]. Specific to binge eating behaviors, in a qualitative study that exposed young women to video blogs depicting binge eating, participants reported that exposure to testimonials produced both positive (eg, reductions in their own binge eating) and negative effects (eg, desire to replicate depicted binge eating behaviors) [12]. Indeed, personal stories from credible peer sources on so-called "pro-Ana" (anorexia nervosa) and "pro-Mia" (bulimia nervosa) web-based forums can undermine ED recovery and validate ED behaviors [13]. Yet the popularity of these sites among people with EDs [14]—as well as the prominence of online communities [15] and social media influencers [16,17] focused on ED recovery—indicates that people often seek out and are exposed to ED-related testimonials online, such that they may be interested in testimonials from peer sources within interventions as well. To deliver testimonials to users of ED interventions, it is therefore critical to avoid potential safety risks by carefully designing testimonials.

Extant literature provides some guidance for designing safe, effective, and engaging testimonials. De Graaf et al [4] conducted a meta-review of 153 experimental studies of health narratives and identified features of effective testimonials. For testimonial messaging, they found that depicting a positive health behavior in a testimonial (eg, physical activity) is more effective for changing behavioral intentions than depicting the consequences of an unhealthy behavior (eg, health risks associated with a sedentary lifestyle), though this finding was mixed in physical activity studies [18,19]. The review also found that first-person testimonials are most effective, but that users do not find testimonials from speakers with similar characteristics to them (eg, shared gender, race, and age) more persuasive than testimonials from dissimilar speakers. However, others have found that testimonials from speakers with similar characteristics to users are more persuasive for changing health beliefs [20], improving engagement with the health message [21], and increasing behavioral intentions [22]. Indeed, a more recent meta-analysis by Chen et al [23] identified that sharing characteristics with a testimonial speaker improves testimonial persuasiveness. However, to our knowledge, the design of testimonials for ED interventions has not been explored, such that much remains unknown about designing testimonials for populations with EDs.

The purpose of this paper is to inform the design of testimonials for a digital binge eating intervention. This study was conducted as part of our ongoing efforts applying human-centered design methods to create a 16-week digital health intervention (FoodSteps) that targets binge eating and weight-related behaviors. We used human-centered design methodologies because they centralize the perspectives of users [24] and are useful for learning how to carefully design and package psychosocial interventions and their constituent parts [25]. Indeed, including users in human-centered design processes can yield engaging digital tools with high clinical impact [26]. We thus conducted virtual "design sessions" with users to learn their preferences related to testimonials.

We examined several specific components. First, given the behavioral benefits of testimonials and the popularity of ED testimonials in nonintervention contexts (eg, online ED discussion forums) [14,15], we sought to learn if users were interested in viewing testimonials as part of a digital ED intervention. Second, we investigated users' preferences for testimonial messaging, speakers, and delivery. Within testimonial delivery, we were interested in delivery modalities, as well as learning how to deliver testimonials (1) over time and (2) as "nudges" toward optimal choices within an intervention. Although many digital health interventions that use testimonials are longitudinal (ie, delivered over weeks or months) [27,28], to our knowledge, how to deliver testimonials over time has been unexplored and represents an important design question for motivating ongoing behavior change in a longitudinal intervention. Relatedly, we wanted to learn users' perspectives on using testimonials to "nudge" a health-oriented behavior, since nudge theory in behavioral economics posits that attaching a "nudge" to an optimal choice is a useful method to influence users' behavior [29]. Testimonials have been used for nudging in nonhealth contexts [30,31] and could be used in health interventions to nudge users toward a particular choice. We therefore sought to learn whether intervention users are open to testimonials as "nudges" and whether they perceive testimonials delivered as nudges would influence their choices. Collectively, results of this human-centered design study can inform the design of testimonials within health interventions,

XSL•FO
RenderX

with particular relevance for interventions that target binge eating and weight-related behaviors.

Methods

We conducted a qualitative study of FoodSteps target users' preferences for testimonial design and delivery. In reporting our data, we adhered to the Standards for Reporting Qualitative Research (see supplementary files) [32]. During codebook formation, the research team consisted of 3 bachelor's level researchers, 1 master's level researcher, and 3 clinical psychologist researchers.

Participants and Procedures

From October to December 2022, we recruited potential participants via flyers in Chicago, Illinois, and online via Craigslist, social media, and ResearchMatch. Participants were recruited to participate in design sessions to help design a mobile app for managing binge eating and weight-related behaviors. Through tailored recruitment materials, methods, and locations, we sought to recruit a sociodemographically diverse sample based on gender, race, ethnicity, age, and socioeconomic status in an effort to increase generalizability and relevance.

Interested individuals completed an online screener, and potential participants were invited to complete a baseline assessment to confirm eligibility. Participants were eligible if they were non-pregnant, English-speaking adults (≥18 years old) with regular access to a smartphone. Eligible participants also had obesity (BMI ≥30 kg/m^2 based on self-reported height and weight) and recurrent binge eating (≥12 objectively large binge eating episodes in the past 3 months based on the diagnostic portion of the Eating Disorder Examination semi-structured interview) [33]. To be eligible, participants had to endorse struggling with losing weight and be willing to use an app to reduce their binge eating. We established this inclusion criteria to obtain a sample of participants that mirrors target users of the FoodSteps intervention. We invited a representative group of 25 participants to our design sessions. Of these, 22 participants completed the interviews; 3 were lost to follow-up.

Ethical Considerations

This study was approved by Northwestern University's Institutional Review Board (STU00216998). All participants gave verbal informed consent. Participants received US$25 financial compensation for completing study procedures.

Design Sessions

The design sessions aimed to learn how to design and deliver testimonials in a digital binge eating intervention. The principal investigator (AKG) and a research assistant (TAM) iteratively developed a 1-hour interview guide for the design sessions. Individual virtual interviews, held via Zoom (Zoom Communications, Inc.), were led by TAM and frequently co-led by AKG.

During the semistructured design sessions, the research team member leading the interview began by briefly explaining the FoodSteps intervention to the participant. They were then shown nonfunctional mock-ups of intervention features, including designs that featured a testimonial (see Figure 1). Participants were asked which versions of the testimonial designs they preferred and their preferences for testimonial speakers (eg, sociodemographic characteristics), messaging (eg, how-to advice), delivery interface (eg, video), and delivery over time and as a "nudge" to a particular action. Specifically, because FoodSteps involves users choosing behavioral strategies each week, participants were shown designs in which a testimonial is used to nudge users to select a particular strategy.

In Table 1, we present the interview guide questions on testimonials. Because the interviews were semi-structured, we followed-up on participants' unique responses with additional questions based on the feedback that participants shared.

As saturation emerged (ie, when enough participants provided the same feedback such that the research team felt consensus had been reached), the designs and interview guide were iteratively updated, consistent with the iterative human-centered design process [26]. For example, as saturation developed regarding the set of identity characteristics participants wanted to share with the testimonial speaker, we asked more about participants' preferences for the longitudinal delivery of testimonials. As the study progressed, we also iteratively updated the visualizations we showed during design sessions based on participants' feedback. As such, participants were shown slightly different designs, and interviews were focused on different questions depending on the stage of the study at which they completed their design session.

XSL•FO
RenderX

Figure 1. Example slide of testimonial design options shown to participants. Designs iterations include (1) no testimonial, (2) Shanice's "big picture" testimonial, and (3) Shanice's "how-to" testimonial. In the latter options, a testimonial is used to nudge a user toward a specific action (in this case, the behavioral strategy "improve self-image").

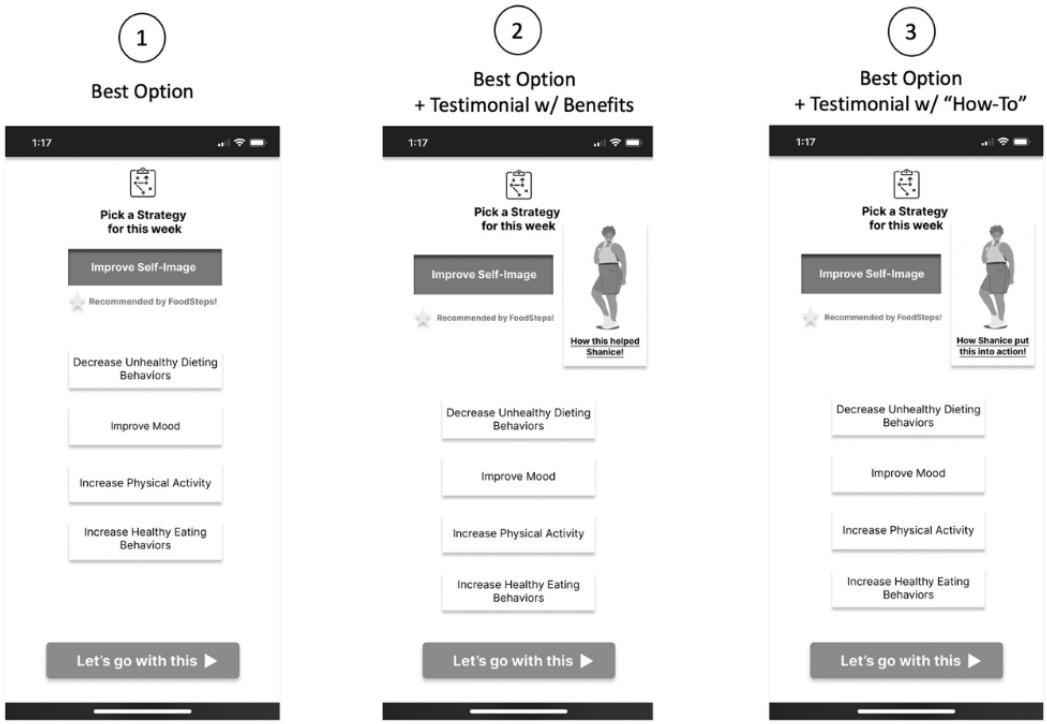


Table . Guiding questions from our interview guide to learn target users' preferences for testimonial design and delivery.

#1	What would you want to see in a testimonial?
#2	How would you want that testimonial presented to you (eg, video, article, pictures)?
#3	Is it important to you for the testimonials to be paired with your own identity? (eg, if you are a mother or father, seeing testimonials from other mothers or fathers)
#3a	If yes, ask: In order to make the testimonials related to you, would you be comfortable answering personal questions at the start to inform this personalization?
#4	Would you expect or want to see the same testimonial throughout the program, or different ones?
#5	Which of these designs do you prefer or like best, if any?
#5a	If there is one you prefer, why is it your favorite or what stands out to you about it?
#6	Which design(s) do you like least? What about them do you not like, find confusing, or feel is unhelpful?

Measures

We assessed participants' sociodemographic characteristics at screening (gender, race, ethnicity, age, and BMI) and through a questionnaire administered as part of the baseline eligibility assessment (household income and education). Objective binge eating was assessed at baseline by a trained assessor who administered the Eating Disorder Examination interview [33]; this assessor did not conduct the design sessions.

Analyses

We qualitatively analyzed interview text for themes using methods from thematic analysis [34,35]. Interviews were automatically transcribed by Zoom and edited for accuracy by a research assistant. A research team member (IRR) reviewed the interview data in full, then iteratively generated and applied an initial set of codes using Dedoose (SocioCultural Research Consultants), a qualitative data analysis software. After refining the codebook with another coder (AKG), IRR reviewed and recoded the data and began identifying themes. To strengthen the rigor of the coding process, the full study team reviewed the codebook and discussed the themes. IRR, AO, and AKG

XSL•FO
RenderX

refined and named the themes and met often to resolve coding ambiguities.

Results

Sample Characteristics

Of the 22 participants, 64% self-identified as female; 27% as Hispanic, 32% as non-Hispanic Black, and 41% as non-Hispanic White. On average, participants were 42 years (range 23-69 years; median 40 years; SD 14.10 years); had a BMI of 39.42 kg/m^2 (range 30.03-57.56 kg/m^2; median 37.48 kg/m^2; SD 7.15 kg/m^2); and experienced 49 binge episodes over the past 3 months (range 15-111; median 41; SD 30.59). Just over half (54%) had at least a bachelor's degree. Excluding 4 missing data points, participants' mean household income was US $79,616 (range US $10,000-$250,000; median US $53,000; SD US $64,991). See Table 2 for full sample characteristics.

Table . Sample characteristics of the target intervention users who participated in this study's design sessions.

ID	Gender	Race	Hispanic	Age (y)	Household income (USD)	Education	BMI (kg/m^2)	Number of binge eating episodes in the past 3 months
1	Female	Black	No	40	—	Master's degree	44.62	46
2	Female	White	No	36	$205,000	Master's degree	31.17	41
3	Female	White	No	43	$250,000	Master's degree	36.94	20
4	Female	White	No	56	$10,000	2-year college	38.16	36
5	Female	White	No	65	—	4-year college	30.99	15
6	Female	Black	No	29	$50,000	High school/GED	46.00	34
7	Female	Black	No	54	$56,000	Master's degree	47.42	111
8	Female	Black	No	28	—	Some college	57.56	22
9	Female	Black	No	40	$113,104	Professional degree	35.87	64
10	Male	White	No	69	$110,000	Professional degree	36.33	60
11	Female	White	Yes	28	—	4-year college	32.87	18
12	Female	More than one race	Yes	27	$30,000	Some college	35.66	111
13	Male	White	No	68	$60,000	Master's degree	41.55	84
14	Male	White	Yes	34	$65,000	Master's degree	34.72	48
15	Female	White	Yes	52	$99,000	Master's degree	40.85	108
16	Male	White	No	43	$150,000	4-year college	38.01	41
17	Male	White	No	48	$30,000	Some college	54.88	24
18	Female	Black	No	23	$50,000	Some college	38.84	67
19	Male	More than one race	Yes	30	$40,000	Some college	36.59	15
20	Male	Black	No	25	$50,000	Some college	36.18	32
21	Other	White	No	34	$25,000	Some college	42.06	32
22	Female	American Indian or Alaskan Native	Yes	45	$40,000	Some college	30.04	51

XSL•FO
RenderX

Preferences for Including Testimonials

Most users preferred designs featuring testimonials. They said testimonials personalized the intervention ("*that makes it a little more personalized for you*," P7) and validated its efficacy ("*look at all these people from various lots of life who have used this and are successful*," P2). Participants also found testimonials motivating ("*it will encourage someone*," P6) and engaging:

> *The testimonial person engages you. It gives you hope. It motivates you.* [P19]

Many said testimonials fostered connection ("*it's sort of a camaraderie in a way*," P13). They said binge eating is an illness for which social support is limited, so testimonials can fill this gap:

> *It's good to know that other people have the same issue.* [P11]
>
> *[It] feels more like you're doing it as a team effort, and that's […] important with such an isolating illness.* [P21]

While they generally preferred designs with testimonials, a subset distrusted testimonials for reasons of relevance, authenticity, and utility. A few recommended excluding testimonials altogether. Some participants found testimonials irrelevant because eating and weight are subjective topics:

> *The reason I don't like the testimonial option is because it's a very personal thing that you're doing.* [P3]
>
> *When it comes to my own weight loss journey, I wouldn't find how it helped somebody else [useful] […] that's just not something I would care to look into.* [P12]

Some found testimonials too "commercial" or distrusted the speaker:

> *I relate it mentally to too many, not scams, but like trying to sell a product [...] I might think it's just a paid actor. I don't trust it enough.* [P20]
>
> *It's like looking at an infomercial on TV […] Here's Suzy Cream Cheese, and she lost four hundred pounds in one week.* [P10]

A few participants also said that they would prefer advice come from the intervention, not a former user:

> *The 'how-to' is more important to me than the testimonial.* [P16]
>
> *Having some instruction as to how I can put it into action might be more beneficial for me personally than hearing how someone else put it into action.* [P9]

Participants also cited unique reasons for disliking testimonials. This participant said the speaker's success did not guarantee their own, so an unhelpful self-comparison could ensue:

> *If it doesn't help you and it did help someone else, then how are you going to internalize that? That's why I didn't like that one.* [P3]

Other critiques were of the illustration's larger body size, which raised concerns about the intervention's efficacy ("*if this helps Shanice, why does Shanice still look like this?*", P9) and prompted size comparisons ("*am I fatter than this person?*", P15). One participant worried the speaker's racial identity could alienate users ("*I would not use the image and the name, just because our country is so* [...] *ugly about race, and that could turn some people off*," P4).

Design Preferences

Participants described their preferences for testimonial speakers, interface, and messaging.

Speaker

Participants preferred speakers with shared identity characteristics such as age, race, gender, body size, lifestyle (ie, being busy), parenthood, and socioeconomic status.

> *I would immediately look at it and say I'm not a middle-aged African American woman.* [P10]
>
> *If it was someone smaller, I would be a bit more skeptical. But if it was someone that looks like me, I could say okay, she or he goes through the same struggle as I do.* [P18]
>
> *She's a busy lady like you are, and […] she still found time to squeeze it in.* [P1]
>
> *There's different challenges if you're a working mother compared to a stay-at-home mother, or someone who's not a mother.* [P2]
>
> *When Oprah lost her weight it was like, you have a personal chef, it's easy for you.* [P7]

Participants said sharing characteristics with speakers made those testimonials more comfortable ("*it's just something I feel comfortable with*," P22), relevant ("*I don't want to see how somebody was successful who I can't relate to*," P10), and inspiring ("*if it's someone I can relate to* […] *it makes it more real*," P7). They said these testimonials would make them more likely to listen:

> *If it shows some 6-foot tall, muscular dude […] I'm not going to listen to him.* [P8]

Yet, one user cautioned that sharing characteristics does not assure identification ("*just because this person is a female, and I'm a female, does not mean that I identify with this person*," P1).

Understanding which intersecting identities matter most to which users is important. We asked a subset of participants if they would be willing to report personal data in the intervention to facilitate identity-based matching and personalization; all of them said yes.

> *When you introduce that [the demographic questions] you should say: in order to really focus in on you, we're going to ask you these questions […] I want to know you're focusing in on me, because then it's worth my while to continue.* [P10]

These findings indicate that it will be important for interventions to deliver tailored testimonials to users, which can be facilitated by assessing their identity characteristics within the intervention.

XSL•FO
RenderX

Messaging

Generally, participants were shown 2 types of testimonials (see Figure 1). The first ("how this helped Shanice") conveyed the "big picture" impact of the behavioral strategy; the second ("how Shanice put this into action") conveyed "how-to" advice. Participants reported varying preferences between these options. Those who preferred "big picture" testimonials found the speaker's end results more motivating than the speaker's process ("*I'm more excited about the final results of losing weight, so this is more engaging*," P19). Others thought the big picture impact was obvious ("*it's not rocket science […] they know the benefits*," P8).

Some participants found "how-to" advice unhelpful ("*how one put it into action may not be how someone else puts it into action*," P9). Others favored "how-to" testimonials for their actionable insights ("*it's concrete. Easily implementable*," P14) that operationalize behavior change:

> *Sometimes when things are broken down […] it makes it a lot easier, especially if it's something that's new to you.* [P18]

These varying preferences indicate that both types of testimonial messaging ("big picture" and "how-to") are motivating to different users. To that end, this participant recommended assessing users' preferences within the intervention to discern the optimal messaging to motivate a user:

> *I could see there being an option in the setup for the app to choose […] Do you want to know testimonials? Do you want how-to? Which ones do you want?* [P17]

Understanding what testimonial messaging to deliver and to whom to best facilitate motivation is critical to optimize testimonials' behavioral impact.

Interface

Participants wanted testimonials with a simple interface and varying multimedia, such as videos, to encourage engagement and connection ("*it's more a personal connection when you have a video*," P2). A participant said they would prefer an image of the speaker, instead of an illustration ("*to the extent it could be a real person, that would be helpful*," P9). Others preferred testimonials delivered multimodally (eg, text and video) to give options for viewing the content:

> *Maybe having a video, and then the transcript below […] so I can kind of listen in my head where I may not have the audio, but I can still read and see them speaking.* [P2]
>
> *Every week, if there's something different; sometimes there might be audio of somebody talking about their experience, then there might be a write-up, and then there might be a video […] that might be motivating […] something that changes as weeks go by.* [P1]

One participant wanted multiple testimonials available so they could choose between speakers:

> *If you had a choice, like if you wanted to see Shanice, or Bob, or whomever. Like you could watch multiple testimonials from different people.* [P2]

Taken together, these preferences for multimodal content and choice between speakers can inform how intervention designers optimize testimonials' user interface.

Delivery Preferences

Testimonials could be a strategy to nudge users toward a particular action within the intervention when choice is involved (eg, in our intervention, users choose between behavioral strategies). We therefore asked participants about their preferences for seeing testimonials when there is an element of user choice. If multiple actions are available, this participant said a testimonial should only be delivered with the nudged action:

> *If I clicked on 'mood' this week, because 'mood' is not my recommended, there would be no testimonial […] I think I would be more motivated [to select the recommendation] if there was a testimonial.* [P2]

Other participants preferred to see testimonials across all available actions:

> *I thought it was going to show a visual next to [each behavioral strategy]. Like, on 'physical activity,' I thought it would have been somebody with weights, and on 'behaviors,' it would have somebody eating.* [P8)]
>
> *It would be great if even if it wasn't a recommendation, that they would have a testimonial there, too […] I still would like that option added to it.* [P7)]

Additionally, because most behavioral interventions are multi-session, we sought to understand users' preferences for the longitudinal delivery of testimonials. In our design sessions, we assessed this by asking users about their preferences after envisioning a week had passed in the program. Many participants wanted testimonial speakers to vary over time because new testimonials continue validating the intervention's efficacy:

> *It validates [it] if you can see real life people and every day it's a brand-new person.* [P19)]
>
> *I think different people each week. So that way you can see a variety of people, so you don't get into, like, oh my gosh, only one person's used this.* [P2)]

Some preferred testimonials from varied speakers each week to reduce potential boredom from hearing the same person's story repeatedly ("*especially if they're going through this week after week […] I really don't need to see that again*," P17). Delivering varied testimonials over time also increases the odds of identifying with the speaker:

> *If I click on it one day, and I see a Tom or something […] I don't really connect if it were Tom, but maybe next week I get somebody else.* [P1)]

However, one participant wanted to see the same speaker over time ("*I'd feel a lot more impressed if you showed me Bob all the time […] It's just like having a personal trainer*," P10). These findings indicate that when choice is involved, selectively delivering testimonials could nudge users toward the optimal action, though users preferred testimonials be available across

XSL•FO
RenderX

all actions. Understanding how to deliver testimonials over time will be crucial for intervention designers to facilitate motivation and engagement throughout a multi-session intervention.

Discussion

Principal Findings

We conducted design sessions to learn how to design and deliver testimonials within a multisession digital binge eating intervention. We found that participants generally preferred designs with testimonials because they promoted engagement, motivation, and connection. This finding indicates that target users of a digital intervention for binge eating and weight-related behaviors are interested in viewing testimonials as part of the intervention and aligns with other literature indicating interest in testimonials from peer sources among individuals with EDs [14,15]. We also described participants' preferences for testimonial design and delivery over time. From these results, we collated design recommendations, summarized in Table 3.

Table . Recommendations for optimizing the design of testimonials in digital health interventions.

General	Include testimonials from former users as part of digital interventions for eating and weight-related behaviors.
	Assess users' testimonial preferences in-app and incorporate choice around whether and how to receive testimonials (eg, speaker, messaging, or delivery).
Speaker	Design testimonials that match the user sociodemographically to improve relevance, understanding, and listening.
	Highlight the authenticity of the speaker to establish trust and avoid appearing too "commercial".
	Avoid showing users irrelevant testimonials that can disengage them.
	Facilitate sociodemographic matching of testimonials by leveraging users' willingness to input personal information in the app for matching purposes.
	Recognize that some users may prefer following the same person over time; as feasible, create and disseminate this content.
Messaging	Highlight actionable, nonprescriptive content that recognizes the personal nature of health journeys to transfer information without alienating users. eg,
	"See how Shanice made this strategy work for her. Her unique approach might motivate you to find your own way to increase your daily activity."
	Highlight the "big picture" impact of behavior changes to inspire users. eg,
	"Here's how increasing her daily physical activity made Shanice feel stronger and helped her connect with her kids."
	Include language to help users avoid fear of failure or unproductive self-comparison to the testimonial speaker. eg,
	"The key is finding an approach that fit in my life. For me, finding what worked was a personal journey, and every small step has added up to big changes."
Interface	Deliver testimonials through images, video, and audio to foster connection between the viewer and speaker, and capture motivating effects.
	Vary delivery method of testimonials to engage users.
Delivery	Offer testimonials from new speakers weekly (or on an intervention-appropriate cadence).
	When choice between actions is involved, consider leveraging testimonials to nudge users toward the optimal action.

Participants preferred testimonial speakers who shared their sociodemographic characteristics. This finding contrasts with de Graaf's [4] meta-analysis, but aligns with more recent research affirming the effectiveness of culturally tailored testimonials [36] and the persuasive influence of sharing race, age, and gender with a testimonial speaker [37]. When asked why tailored testimonials mattered to them, some participants said that nonrelatable testimonials (eg, from speakers with higher socioeconomic status or different racial identities) could alienate users, which underscores the importance of tailoring testimonials to users' personal characteristics. Notably, tailoring based on body size poses a nuanced challenge for weight-related interventions. While most participants preferred speakers with relatable body sizes, one said that this would make them think the intervention was ineffective. Indeed, because sharing certain characteristics was important to some participants, but not others, our results indicate that intervention designers should consider ways to assess users' preferences and characteristics

XSL•FO
RenderX

at the onset of the intervention to identify users' most pertinent identities (eg, is it more important for testimonials to match a user's gender or socioeconomic status?) and tailor testimonials accordingly. This can be facilitated by users' willingness to report personal information in the intervention, though continued design work is needed to understand user tolerance for testimonial personalization over time. Indeed, Robbins et al [38] observed that some Black participants were uncomfortable with a video testimonial intervention that featured overwhelmingly Black speakers. In multisession interventions, testimonial tailoring based on characteristics like race could feel increasingly targeted over time. It is thus necessary to learn users' preferences for tailored testimonials throughout an intervention during its actual delivery. Further research is also needed to empirically test if tailoring testimonials improves their effectiveness in the context of eating and weight-related interventions.

For testimonial messaging, we found users were relatively split between receiving "how-to" information versus "big picture" information, in that some preferred testimonials with actionable advice, whereas others preferred testimonials that highlighted the end result of the intervention. This divide also appears in the literature on testimonials for deterring maladaptive health behaviors. While Keer et al [39] found that testimonials describing the positive consequences of abstinence from binge drinking were most effective, Jawed and Hogan [40] found that highlighting actionable strategies for smoking cessation improved testimonial effectiveness and engagement. Our findings affirm that both forms of testimonial messaging may be effective for different subsets of users, depending on their preferred style of messaging. To this end, participants suggested incorporating choice around whether and how to receive testimonials, depending on the elements they find most motivating. User-driven personalization is highly feasible in digital interventions, given a sufficient bank of testimonial content, such that this is a promising area for further inquiry. Future work also would benefit from identifying factors that predict which design elements (eg, forms of messaging) are preferred by which subsets of users, toward effectively personalizing testimonials to users. Importantly, we did not explore users' preferences for testimonial messaging that included descriptions of binge eating or dangerous compensatory behaviors, given the behavioral consequences that can arise when viewers of ED testimonials are exposed to this kind of messaging [11,12]. Future work could explore opportunities to safely design testimonials that mention such behaviors, though our results indicate that potential users of our binge eating intervention were interested in and motivated by testimonials about positive health behaviors, such that testimonials describing ED behaviors could be avoided in interventions to reduce potential risks.

For delivery, we found that users were interested in written testimonials as well as video, audio, and multimedia testimonials. Because de Graaf [4] did not identify any modality as uniquely effective, and given participants' varied preferences, assessing users' preferred delivery modality at the onset of intervention is another promising opportunity for tailoring. For delivery over time, users preferred testimonials from varied speakers because they were more engaging and validated the intervention's efficacy. We did not learn how their messaging preferences may change over time, which is important because these preferences could meaningfully evolve (eg, a new user initially prefers "big picture" testimonials but later prefers "how tos"). Future work should thus examine methods to longitudinally assess users' testimonial preferences. When the intervention allowed users to choose between actions (in our case, behavioral strategies), users preferred testimonials to be delivered across all choices. While intervention designers could do this, our results indicate that selectively delivering a testimonial with the recommended action could encourage users to select it. This finding aligns with nudge theory in behavioral economics [29] and may be especially relevant for just-in-time adaptive interventions, which have deployed nudges as part of their behavior change strategies [41]. Our finding that users positively perceived testimonials indicates that testimonials are a promising design element to explore for deployment as nudges in digital health interventions, though future work should test if exclusively delivering a testimonial for the optimal action improves the action's uptake.

Despite the generally positive perceptions of testimonials, some users said testimonials would not motivate behavior change. This finding needs to be validated in practice through actual intervention delivery. Although much research supports the behavioral effects of testimonials [3], including for promoting abstinence from maladaptive health behaviors [9], Shen et al [42] found that cessation testimonials did not have significant effects, as opposed to prevention testimonials (eg, using condoms) and detection testimonials (eg, cancer screening), which did produce significant effects. This indicates that intervention designers—especially those who are designing interventions to dissuade maladaptive health behaviors—must continue identifying other elements to influence behavior and design to preempt testimonial pitfalls. Indeed, the subset of users who said testimonials would not influence their health behavior change universally critiqued the testimonials' trustworthiness. This objection can be understood through the lens of persuasive intent (ie, a message's implicit goal to persuade), which has been shown to undermine the credibility and persuasiveness of health communications [43,44]. Designing testimonials to highlight source credibility may help improve trust in the testimonial message [45]. Another approach to improve the credibility of testimonials could involve inviting users to submit their own testimonials for future integration into the intervention. This strategy could also help create a bank of testimonials from a diversity of users, which could then be deployed in the intervention.

Further research is needed to validate these findings with FoodSteps users in practice. While we believe these findings are relevant to interventions for other health topics, our study is limited by its focus on an intervention for binge eating and weight-related behaviors. Designers of other health interventions may wish to follow similar human-centered design methods to identify intervention-specific design preferences.

XSL•FO
RenderX

Conclusions

This study sought to learn how to optimize the design and delivery of testimonials in a multisession digital health intervention. Our findings confirm that target users of a digital intervention for binge eating and weight-related behaviors wanted testimonials delivered as part of the intervention. We generated novel insights into users' preferences for testimonial messaging, speakers, and delivery (through varied modalities, over time, and as a "nudge" toward a particular action). These results indicate opportunities for digital health intervention designers to tailor testimonials to users based on their preferences and identities. This study extends existing research by applying human-centered design methods to identify design preferences that can now be implemented and tested in practice. Our work offers a foundation for further inquiry into areas such as user preferences for sociodemographic tailoring over time and testimonials as "nudges" toward optimal actions in health interventions. Future research should continue applying human-centered design methods to address these and other questions, toward the goal of optimizing the behavioral impact of health testimonials.

Acknowledgments

This work was funded by the National Institute of Diabetes and Digestive and Kidney Diseases under grant number R01 DK133300 and the National Institute of Mental Health under grant number T32 MH115882.

Authors' Contributions

AKG and TAM designed the study. Data were analyzed by IRR, AO, and AKG. IRR, AO, and AKG drafted the manuscript. TL, TAM, MK, LMP, LDL, CA, ET, and KO critically reviewed and improved the manuscript with intellectual content. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Checklist 1
SRQR checklist.
[PDF File, 947 KB - formative_v9i1e59691_app1.pdf]

References

1. Bilandzic H, Busselle R. Narrative persuasion. In: The SAGE Handbook of Persuasion: Developments in Theory and Practice, 2nd edition: Sage Publications, Inc; 2013:200-219.
2. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
3. Xu J. A meta-analysis comparing the effectiveness of narrative vs. statistical evidence: health vs. non-health contexts. Health Commun 2023 Dec;38(14):3113-3123. [doi: 10.1080/10410236.2022.2137750] [Medline: 36278821]
4. de Graaf A, Sanders J, Hoeken H. Characteristics of narrative interventions and health effects: a review of the content, form, and context of narratives in health-related narrative persuasion research. RCR 2016;4:88-131. [doi: 10.12840/issn.2255-4165.2016.04.01.011]
5. Perrier MJ, Martin Ginis KA. Changing health-promoting behaviours through narrative interventions: a systematic review. J Health Psychol 2018 Sep;23(11):1499-1517. [doi: 10.1177/1359105316656243] [Medline: 27387514]
6. Falzon C, Radel R, Cantor A, d'Arripe-Longueville F. Understanding narrative effects in physical activity promotion: the influence of breast cancer survivor testimony on exercise beliefs, self-efficacy, and intention in breast cancer patients. Support Care Cancer 2015 Mar;23(3):761-768. [doi: 10.1007/s00520-014-2422-x] [Medline: 25186211]
7. Oh HJ, Larose R. Tell me a story about healthy snacking and I will follow: comparing the effectiveness of self-generated versus message-aided implementation intentions on promoting healthy snacking habits among college students. Health Commun 2015;30(10):962-974. [doi: 10.1080/10410236.2014.910289] [Medline: 25256795]
8. Perrier MJ, Martin Ginis KA. Narrative interventions for health screening behaviours: a systematic review. J Health Psychol 2017 Mar;22(3):375-393. [doi: 10.1177/1359105315603463] [Medline: 26359288]
9. Igartua JJ, Rodríguez-Contreras L. Narrative voice matters! Improving smoking prevention with testimonial messages through identification and cognitive processes. Int J Environ Res Public Health 2020 Oct 5;17(19):7281. [doi: 10.3390/ijerph17197281] [Medline: 33028030]
10. Gardner L, Leshner G. The role of narrative and other-referencing in attenuating psychological reactance to diabetes self-care messages. Health Commun 2016;31(6):738-751. [doi: 10.1080/10410236.2014.993498] [Medline: 26528578]

XSL•FO
RenderX

11. Rennick-Egglestone S, Morgan K, Llewellyn-Beardsley J, et al. Mental health recovery narratives and their impact on recipients: systematic review and narrative synthesis. Can J Psychiatry 2019 Oct;64(10):669-679. [doi: 10.1177/0706743719846108] [Medline: 31046432]
12. Sung B, Lee JY. The influence of YouTube binge eating vlog viewing on perception and behavior among South Korean women, aged 20s to 30s, with binge eating tendencies: a qualitative analysis. Int J Ment Health 2024:1-27. [doi: 10.1080/00207411.2024.2377810]
13. Firkins A, Twist J, Solomons W, Keville S. Cutting ties with Pro-Ana: a narrative inquiry concerning the experiences of Pro-Ana disengagement from six former site users. Qual Health Res 2019 Aug;29(10):1461-1473. [doi: 10.1177/1049732319830425] [Medline: 30810096]
14. Sharpe H, Musiat P, Knapton O, Schmidt U. Pro‐eating disorder websites: facts, fictions and fixes. J Public Ment Health 2011 Mar 15;10(1):34-44. [doi: 10.1108/17465721111134538]
15. Nutley SK, Falise AM, Henderson R, Apostolou V, Mathews CA, Striley CW. Impact of the COVID-19 pandemic on disordered eating behavior: qualitative analysis of social media posts. JMIR Ment Health 2021 Jan 27;8(1):e26011. [doi: 10.2196/26011] [Medline: 33465035]
16. Nikolova I, LaMarre A. “If I Unfollow Them, It’s Not A Dig at Them”: A narrative analysis of Instagram use in eating disorder recovery. Psychol Women Q 2023 Sep;47(3):387-401. [doi: 10.1177/03616843231166378]
17. Herrick SSC, Hallward L, Duncan LR. “This is just how I cope”: an inductive thematic analysis of eating disorder recovery content created and shared on TikTok using #EDrecovery. Int J Eat Disord 2021 Apr;54(4):516-526. [doi: 10.1002/eat.23463] [Medline: 33382136]
18. Gray JB, Harrington NG. Narrative and framing: a test of an integrated message strategy in the exercise context. J Health Commun 2011 Mar;16(3):264-281. [doi: 10.1080/10810730.2010.529490] [Medline: 21058142]
19. Wirtz JG, Kulpavaropas S. The effects of narrative and message framing on engagement and eating intention among a sample of adult Hispanics. J Nutr Educ Behav 2014;46(5):396-400. [doi: 10.1016/j.jneb.2013.12.005] [Medline: 24517968]
20. de Graaf A. The effectiveness of adaptation of the protagonist in narrative impact: similarity influences health beliefs through self-referencing. Hum Commun Res 2014 Jan;40(1):73-90 [FREE Full text] [doi: 10.1111/hcre.12015]
21. Kim M, Shi R, Cappella JN. Effect of character-audience similarity on the perceived effectiveness of antismoking PSAs via engagement. Health Commun 2016 Oct;31(10):1193-1204. [doi: 10.1080/10410236.2015.1048421] [Medline: 26891148]
22. Lu AS. An experimental test of the persuasive effect of source similarity in narrative and nonnarrative health blogs. J Med Internet Res 2013 Jul 25;15(7):e142. [doi: 10.2196/jmir.2386] [Medline: 23887135]
23. Chen M, Dong Y, Wang J. A meta-analysis examining the role of character-recipient similarity in narrative persuasion. Communic Res 2024 Feb;51(1):56-82. [doi: 10.1177/00936502231204834]
24. Maguire M. Methods to support human-centred design. Int J Hum Comput Stud 2001 Oct;55(4):587-634. [doi: 10.1006/ijhc.2001.0503]
25. Lyon AR, Koerner K. User-centered design for psychosocial intervention development and implementation. Clin Psychol (New York) 2016 Jun;23(2):180-200. [doi: 10.1111/cpsp.12154] [Medline: 29456295]
26. Graham AK, Wildes JE, Reddy M, Munson SA, Barr Taylor C, Mohr DC. User-centered design for technology-enabled services for eating disorders. Int J Eat Disord 2019 Oct;52(10):1095-1107. [doi: 10.1002/eat.23130] [Medline: 31313370]
27. Brewer LC, Kumbamu A, Smith C, et al. A cardiovascular health and wellness mobile health intervention among Church-going African Americans: formative evaluation of the FAITH! app. JMIR Form Res 2020 Nov 17;4(11):e21450. [doi: 10.2196/21450] [Medline: 33200999]
28. Bruehlman-Senecal E, Hook CJ, Pfeifer JH, et al. Smartphone app to address loneliness among college students: pilot randomized controlled trial. JMIR Ment Health 2020 Oct 20;7(10):e21496. [doi: 10.2196/21496] [Medline: 33079071]
29. Patel MS, Volpp KG, Asch DA. Nudge units to improve the delivery of health care. N Engl J Med 2018 Jan 18;378(3):214-216. [doi: 10.1056/NEJMp1712984] [Medline: 29342387]
30. Gottschewski-Meyer PO, Lüddemann N, der LM, von VC. Exploring digital social norms nudges in E-grocery: typical consumer testimonials with a warm glow. 2023 Presented at: Internationale Tagung Wirtschaftsinformatik URL: https://aisel.aisnet.org/wi2023/59
31. Ouvrard B, Préget R, Reynaud A, Tuffery L. Nudging and subsidising farmers to foster smart water meter adoption. Eur Rev Agric Econ 2023 Jun 13;50(3):1178-1226. [doi: 10.1093/erae/jbad013]
32. O’Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med J Assoc Am Med Coll 2014;89(9):1245-1251. [doi: 10.1097/ACM.0000000000000388]
33. Fairburn CG, Cooper Z, O’Connor ME. The Eating Disorder Examination (EDE) 17.0D. 2014. URL: https://www.credo-oxford.com/7.2.html [accessed 2024-04-07]
34. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
35. Braun V, Clarke V. Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern‐based qualitative analytic approaches. Couns Psychother Res 2021 Mar;21(1):37-47. [doi: 10.1002/capr.12360]

XSL•FO
RenderX

36. Nugraha YR, Aditjondro E, Svenson D, et al. Harnessing the power of testimonial videos: a deep dive into the #1Minute2Quit campaign for spreading awareness on the health hazards of tobacco. JKMJC 2024;40(4):450-465 [FREE Full text] [doi: 10.17576/JKMJC-2024-4004-25]
37. Kim M. When similarity strikes back: conditional persuasive effects of character-audience similarity in anti-smoking campaign. Hum Commun Res 2019 Jan;45(1):52-77. [doi: 10.1093/hcr/hqy013] [Medline: 30631219]
38. Robbins R, Senathirajah Y, Williams NJ, et al. Developing a tailored website for promoting awareness about obstructive sleep apnea (OSA) among Blacks in community-based settings. Health Commun 2019 May;34(5):567-575. [doi: 10.1080/10410236.2018.1423865] [Medline: 29338353]
39. Keer M, van den Putte B, de Wit J, Neijens P. The effects of integrating instrumental and affective arguments in rhetorical and testimonial health messages. J Health Commun 2013;18(9):1148-1161. [doi: 10.1080/10810730.2013.768730] [Medline: 23659225]
40. Jawed A, Hogan A. Lessons learned from telltale testimonies: a descriptive study assessing coverage of the tips from former smokers campaign on YouTube. Am J Health Educ 2024 Mar 3;55(2):108-118. [doi: 10.1080/19325037.2023.2296950]
41. Hsu TCC, Whelan P, Gandrup J, Armitage CJ, Cordingley L, McBeth J. Personalized interventions for behaviour change: a scoping review of just-in-time adaptive interventions. Br J Health Psychol 2025 Feb;30(1):e12766. [doi: 10.1111/bjhp.12766] [Medline: 39542743]
42. Shen F, Sheer VC, Li R. Impact of narratives on persuasion in health communication: a meta-analysis. J Advert 2015 Apr 3;44(2):105-113. [doi: 10.1080/00913367.2015.1018467]
43. Lee EJ, Kim HS, Joo MH. Social media vs. mass media: mitigating the suspicion of ulterior motives in public health communication. Health Commun 2023 Oct;38(11):2450-2460. [doi: 10.1080/10410236.2022.2074781] [Medline: 35610554]
44. Wang W, Shen F. The effects of health narratives: examining the moderating role of persuasive intent. Health Mark Q 2019;36(2):120-135. [doi: 10.1080/07359683.2019.1575061] [Medline: 30907268]
45. Xu Y, Margolin D, Niederdeppe J. Testing strategies to increase source credibility through strategic message design in the context of vaccination and vaccine hesitancy. Health Commun 2021 Oct;36(11):1354-1367. [doi: 10.1080/10410236.2020.1751400] [Medline: 32308037]

Abbreviations

ED: eating disorder

Edited by A Mavragani; submitted 19.04.24; peer-reviewed by K Cooper; revised version received 15.01.25; accepted 10.02.25; published 21.03.25.

Please cite as:
Rooper IR, Ortega A, Massion TA, Lakhtakia T, Kruger M, Parsons LM, Lipman LD, Azubuike C, Tack E, Obleada KT, Graham AK
Optimizing Testimonials for Behavior Change in a Digital Intervention for Binge Eating: Human-Centered Design Study
JMIR Form Res 2025;9:e59691
URL: https://formative.jmir.org/2025/1/e59691
doi:10.2196/59691

© Isabel R Rooper, Adrian Ortega, Thomas A Massion, Tanvi Lakhtakia, Macarena Kruger, Leah M Parsons, Lindsay D Lipman, Chidiebere Azubuike, Emily Tack, Katrina T Obleada, Andrea K Graham. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Interactive Care Coordination and Navigation mHealth Intervention for People Experiencing Homelessness: Cost Analysis, Exploratory Financial Cost-Benefit Analysis, and Budget Impact Analysis

Hannah P McCullough[1], PharmD, MPH; Leticia R Moczygemba[1], PharmD, PhD; Anton L V Avanceña[1,2], PhD; James O Baffoe[1], PharmD

[1]Health Outcomes Division, College of Pharmacy, The University of Texas at Austin, 2409 University Avenue, Austin, TX, United States

[2]Department of Internal Medicine, Dell Medical School, The University of Texas at Austin, Austin, TX, United States

Corresponding Author:
Leticia R Moczygemba, PharmD, PhD
Health Outcomes Division, College of Pharmacy, The University of Texas at Austin, 2409 University Avenue, Austin, TX, United States

Abstract

Background: The Interactive Care Coordination and Navigation (iCAN) mobile health intervention aims to improve care coordination and reduce hospital and emergency department visits among people experiencing homelessness.

Objective: This study aimed to conduct a three-part economic evaluation of iCAN, including a (1) cost analysis, (2) exploratory financial cost-benefit analysis, and (3) budget impact analysis (BIA).

Methods: We collected cost and expenditure data from a randomized controlled trial of iCAN to conduct a cost analysis and exploratory financial cost-benefit analysis. Costs were classified as startup and recurring costs for participants and the program. Startup costs included participant supplies for each participant and SMS implementation costs. Recurring costs included the cost of recurring services, SMS text messaging platform maintenance, health information access fees, and personnel salaries. Using the per participant per year (PPPY) costs of iCAN, the minimum savings reduction in the average health care costs among people experiencing homelessness that would lead to a benefit-cost ratio >1 for iCAN was calculated. This savings threshold was calculated by dividing the PPPY cost of iCAN by the average health care costs among people experiencing homelessness multiplied by 100%. The benefit-cost ratio of iCAN was calculated under different savings thresholds from 0% (no savings) to 50%. Costs were calculated PPPY under different scenarios, and the results were used as inputs in a BIA. A probabilistic sensitivity analysis was conducted to incorporate uncertainty around cost estimates. Costs are in 2022 US $.

Results: The total cost of iCAN was US $2865 PPPY, which was made up of US $265 in startup (9%) and US $2600 (91%) in recurring costs PPPY. The minimum savings threshold that would cause iCAN to have a positive return on investment is 7.8%. This means that if average health care costs (US $36,917) among people experiencing homelessness were reduced by more than 7.8% through iCAN, the financial benefits would outweigh the costs of the intervention. When health care costs are reduced by 25% ($9229/$36,917; equal to 56% [$9229/$16,609] of the average cost of an inpatient visit), the benefit-cost ratio is 3.22, which means that iCAN produces US $2.22 in health care savings per US $1 spent. The BIA estimated that implementing iCAN for 10,250 people experiencing homelessness over 5 years would have a financial cost of US $28.7 million, which could be reduced to US $2.2 million if at least 8% ($2880/$36,917) of average health care costs among people experiencing homelessness are reduced through the intervention.

Conclusions: If average costs of emergency department and hospital visits among people experiencing homelessness were reduced by more than 7.8% ($2880/$36,917) through iCAN, the financial benefits would outweigh the costs of the intervention. As the savings threshold increases, it results in a higher benefit-cost ratio.

Trial Registration: ClinicalTrials NCT05365867; https://clinicaltrials.gov/study/NCT05365867

(JMIR Form Res 2025;9:e64973) doi:10.2196/64973

KEYWORDS

people experiencing homelessness; budget impact analysis; financial cost-benefit analysis; mHealth; care coordination; care; mobile health; smartphones; homeless; hospitalization; cost analysis; health care cost; economic; emergency department; United States; cost-benefit; digital health

XSL•FO
RenderX

Introduction

People experiencing homelessness have higher rates of emergency department (ED) and hospital visits than housed groups. A systematic review reported that, on average, across 10 studies, the number of ED visits among people experiencing homelessness ranged from 0.7 to 5.8 visits annually. Furthermore, 5 of these studies compared ED use between people experiencing homelessness and housed groups and found that the rates were higher in people experiencing homelessness [1]. It has also been reported that people experiencing homelessness have 1.9 times the odds of hospital readmissions compared with housed groups [2]. Furthermore, a study of the hospital use patterns of people experiencing homelessness using electronic health data in an integrated health system in the United States found that people experiencing homelessness have higher rates of 30-day readmissions and longer lengths of stay in the hospital compared with those that are not homeless [3]. Some of the reasons for high ED and hospital utilization in people experiencing homelessness includes inadequate access to primary care and challenges taking medications due to medications being lost or stolen and competing priorities for basic needs such as shelter and food [4,5].

Mobile health (mHealth) is a type of digital technology that involves the use of mobile devices such as mobile phones to deliver health services [6]. The ubiquity and declining costs of mobile phones in the United States and other countries has generated interest in the use of mHealth to improve population and public health [7,8]. Studies have shown that mHealth can be effective in improving outcomes such as increasing patient adherence and healthy behaviors in chronic disease management [9], and reducing ED use [10-12]. According to 2 systematic reviews of cost effectiveness of digital interventions (eg, telephone, SMS text messaging, and mHealth apps), for those studies that included some type of economic evaluation, there is some evidence that these interventions may be cost-beneficial [9,13]. Even in the mHealth interventions where the outcomes were not statistically significant or cost-savings were not addressed, a reduction in preventable and nonemergent ED visits suggest likely significant cost-savings for health care payers in the community [13].

Given that mobile phones can reach traditionally hard to reach groups, including people experiencing homelessness [14], mHealth interventions should be considered when designing interventions to decrease ED and hospital visits in people experiencing homelessness. In a small pre-post study, people experiencing homelessness indicated that it was easy to use a mobile phone to participate in an mHealth intervention, and a qualitative study of people experiencing homelessness in the pre-post study found that mobile phones empowered people experiencing homelessness to self-manage their health and social needs [15,16]. A scoping review of 12 studies found that using technology on mobile phones such as SMS text messaging and apps could help improve access to mental health services for youth experiencing homelessness, with evidence of feasibility and acceptability by users [17]. The scalability of mHealth interventions for people experiencing homelessness will depend, in part, on whether its benefits outweigh the costs of the intervention. Thus, in this study, we will estimate the costs of an mHealth intervention, which will be described in the Methods section, that was implemented to decrease ED and hospital use among people experiencing homelessness and conduct an exploratory financial cost-benefit analysis and preliminary budget impact analysis to evaluate the potential cost-savings from this intervention for people experiencing homelessness.

Methods

Overview

Interactive Care Coordination and Navigation (iCAN) is a community-based mHealth intervention in the United States that was designed using findings from previous studies related to acceptability and benefits of mHealth interventions in people experiencing homelessness [15,16] and with feedback from homeless care providers and people with the lived experience of homelessness. The purpose of iCAN, which is implemented at community-based sites that provide homeless services, is to improve care coordination between health and social services for people experiencing homelessness with a goal of reducing hospitalizations and ED visits. It is a multicomponent intervention comprised of SMS text messages, preinstalled apps and resources, including a bus pass, and a case manager who conducted telephone encounters and used bidirectional SMS texting to communicate with participants. Participants received 3 - 5 messages daily regarding medication adherence and appointment reminders, general health messages, motivational messages such as encouragement to keep goals, and as needed messages for local information (eg, weather updates and resource fair information). Participants were also able to text the case manager or study staff. Within 72 hours of enrollment, the case manager conducted a telephone assessment to identify relevant health and social needs whereby they could connect participants to medical and social services in the community. Case managers received weekly reports twice about ED and hospital use from the local health information exchange (HIE) and contacted participants within 72 hours of an ED or hospital visit to help coordinate care needs for managing discharge instructions. People experiencing homelessness who were 18 years or older, owned a cell phone with service, had at least 2 chronic health conditions, had at least 2 hospitalizations or ED visits 6 months before enrollment, scored 4 or higher on the Rapid Estimate of Adult Literacy in Medicine-Short Form, and scored greater than 17 on the Mini-Mental State Exam were eligible to participate in iCAN. There was a total of 60 participants in the iCAN intervention. The mean age of participants in the iCAN intervention was 50.1 (SD 10.2) years. Most (42/60, 70%) were men; over half (34/60, 57%) reported their race as White and 32% (19/60) reported their race as Black. The randomized controlled trial (RCT) of iCAN completed recruitment in August 2023 and data analysis is ongoing.

Informed by previous studies of mHealth interventions [18-21], we conducted a three-part economic evaluation of iCAN, including a (1) cost analysis, (2) exploratory financial cost-benefit analysis, and (3) budget impact analysis. For the cost analysis, we used expenditure data from the iCAN RCT to determine the total cost and cost per participant per year (PPPY).

XSL•FO
RenderX

Using these costs, we conducted an exploratory cost-benefit analysis to understand the minimum health care savings that would cause iCAN to have a positive return on investment. Finally, we conducted a budget impact analysis to determine the financial requirements of scaling up the iCAN intervention in a metropolitan area. For the cost and cost-benefit analyses, we used scenario and probabilistic sensitivity analyses to determine the effect of different assumptions and uncertainty in existing data on our findings.

We followed the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) [22], the International Training and Education Center for Health (I-TECH) Applied Economic Evaluation of Digital Health Interventions Report [23], the The International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Budget Impact Analysis Report [24], and other guidance [25] in conducting this study. We used a public-payer perspective, assuming that the majority of people experiencing homelessness do not have private health insurance and cannot pay for out-of-pocket health care costs [26,27]. Thus, the costs of iCAN and health care services used by people experiencing homelessness are borne by the public sector through public health care payers (eg, Medicaid), health care systems, and local governments.

Data Collection

Intervention Costs

We collected expenditure records from the ongoing iCAN RCT to determine intervention costs. Intervention costs were separated into startup and recurring costs using a microcosting approach [28].

Startup or capital costs were separated into participant-related and program-related costs. Participant-related startup costs included a cell phone, phone case, armband, charging cable and block, rechargeable power bank, and drawstring backpack. Program-related startup costs included the initial payment for the SMS text messaging platform and printing costs for training materials.

Recurring costs were separated into participant-related, program-related, and personnel costs. Participant-related recurring costs included a monthly bus pass and a SIM card with a phone plan, which are given to each participant. Program-related recurring costs included monthly SMS text messaging platform maintenance fees and HIE data access fees. Access to HIE data are required to monitor ED and inpatient health care use of participants during the trial, however, this expense may be excluded during the scaling up of iCAN in the community. We explored this scenario in sensitivity analysis.

Personnel-related recurring costs included the salaries of a case manager and program coordinator. The case manager spent approximately 20 hours per week communicating with 24 enrolled participants; for 60 participants, this was equivalent to 1.25 case managers. The program coordinator, a role assumed by graduate research assistants during the iCAN trial, is responsible for several operational and administrative tasks, including setting up participants' cell phones (1 h per phone), training participants on the intervention (1 h per participant), and troubleshooting any technology-related issues (2 h per week). The case manager's annual salary was estimated based on the average hourly wage of social workers in Texas from the Bureau of Labor Statistics multiplied by the total work hours (2500) per year [29]. The salary of the program coordinator was based on the average hourly pay of graduate research assistants, which is available on the human resources website of The University of Texas at Austin [30]. The hours spent by the program coordinator on specific tasks were determined through interviews with trial staff and a review of staff records, and the total time spent was used to calculate the program coordinator's annual salary.

Health Care Costs

We used existing data to estimate the annual health care costs related to inpatient and ED care among people experiencing homelessness. We relied on previous studies on health care utilization among people experiencing homelessness in Texas to determine the average number and range of ED [31-34] and inpatient visits [31,35] (Multimedia Appendix 1). We took the weighted average and median of ED and inpatient visits across studies, which resulted in equivalent values and were comparable with the estimates from other jurisdictions and the United States [36-40]. We then multiplied these figures by the average cost of each service from the 2022 Medical Expenditure Panel Survey-Household Component (MEPS-HC) database [41]. With the average cost of inpatient (US $16,609, [SE $860]) and ED visits (US $1233 [SE $66]) from MEPS-HC, we used a total health care cost of US $36,916 PPPY. This average cost is comparable with what previous health care utilization studies among people experiencing homelessness who use health care services have reported in Texas and other locations [31-40].

Analysis

Cost Analysis

We calculated the total costs of iCAN by adding up all the startup and recurring intervention costs. We also calculated the PPPY costs by dividing the total costs by 60, which is the number of enrolled participants in the iCAN RCT. In scenario analyses, we varied assumptions around the cost inputs (Multimedia Appendix 2) and calculated total and PPPY costs under each scenario.

Cost-Benefit Analysis

Using the PPPY costs of iCAN, we calculated the minimum savings or reduction in the average health care costs among people experiencing homelessness that would lead to a benefit-cost ratio >1 for iCAN, where the financial benefits associated with the intervention outweigh its costs. We calculated this "savings threshold" by dividing the PPPY cost of iCAN by the average health care costs among people experiencing homelessness multiplied by 100%. For example, if the PPPY cost of iCAN is 10% of the average health care costs among people experiencing homelessness, a health care cost reduction of >10% (the savings threshold) would lead to a benefit-cost ratio >1 for iCAN. These savings can be accrued by reducing the number or severity of ED and inpatient visits. We then calculated the benefit-cost ratio of iCAN under different savings thresholds from 0% (no savings) to 50%.

XSL•FO
RenderX

To account for uncertainty in various cost inputs (ie, cost of the iCAN, number of ED, and inpatient visits), we conducted a probabilistic sensitivity analysis. A common approach in economic modeling, probabilistic sensitivity analysis involves generating probability distributions for each input using the best available parameter values and then drawing 10,000 random values of each input to produce a distribution of results. From these 10,000 estimates, we determined the mean, median, and 95% probable interval of the savings threshold for iCAN. We assigned gamma distributions for iCAN and health care costs and Poisson distributions for the annual number of inpatient and ED visits among people experiencing homelessness.

Budget Impact Analysis

We conducted a budget impact analysis to determine the financial requirements of implementing iCAN for people experiencing homelessness in a metropolitan area. We used a 5-year time horizon and selected Travis County, Texas, which includes the city of Austin where the iCAN trial was conducted, as the setting of the analysis. The number of unsheltered people experiencing homelessness in the first year was assumed to be 3990 based on estimates of the Homelessness Response System [42]. We assumed that the people experiencing homelessness population would grow annually by 3.4%, which is the estimated population growth rate for people experiencing homelessness by the US Department of Housing and Urban Development in 2023 [43]. We further assumed that only 60% of people experiencing homelessness would be eligible to participate in iCAN based on previous use of ED and inpatient services based on trial data and previously published studies. Finally, we assumed that only 80% of eligible people experiencing homelessness would participate in iCAN based on engagement data from the trial.

We included fixed and variable startup costs and variable recurring costs in the budget impact analysis. Fixed start-up costs, which were only paid in the first year of program implementation regardless of the number of participants, included the costs of the SMS text messaging platform and iCAN training materials. Variable startup costs included the 1-time cost of supplies (eg, cell phone) per participant. Recurring costs were the cost of services and salaries that varied based on the number of participants; these costs included SMS text messaging platform maintenance fees, HIE data access, personnel salaries, and annual bus passes and phone plans for each participant. We used a medical inflation rate of 2.3% in estimating future costs.

Ethical Considerations

The trial described in the study was approved by the University of Texas Institutional Board (STUDY00002666). The consent form indicated that research data would be kept for use in secondary analyses.

Results

Cost Analysis

The total cost of implementing iCAN to 60 participants for 1 year was US $170,921. The startup costs of iCAN (US $15,925) were estimated to be 9.3% ($15,925/$171,921) of the total cost (Table 1) and the remainder ($155,996/$171,921, 90.7%) were attributable to recurring costs (US $155,996), particularly personnel costs (US $84,908; Table 2). The startup and recurring costs PPPY were US $265 and US $2600, respectively, leading to a total PPPY cost of US $2865 PPPY (Table 3).

Table . Startup costs of Interactive Care Coordination and Navigation, a mobile health intervention for people experiencing homelessness in the United States, for a cost analysis. Total costs are calculated based on 60 participants. Costs are in 2022 US $.

Item		Unit cost per participant (US $)	Total cost (US $)
Participant-related			
	Cell phone	149.99	8999.40
	Phone case	12.99	779.40
	Arm band	8.99	539.40
	Charging block	3.50	210.00
	USB-C charging cable	3.50	210.00
	Power bank	14.99	899.40
	Drawstring backpack	1.34	80.40
	Subtotal	195.30	11,718.00
Program-related			
	Text messaging platform implementation	59.80	3588.00
	Printing costs for training binder supplies	10.31	618.59
	Subtotal	70.11	4206.59
Total		265.41	15,924.59

XSL•FO
RenderX

Table . Recurring costs of Interactive Care Coordination and Navigation, an mobile health intervention for people experiencing homelessness in the United States, for a cost analysis. Total costs were calculated for 60 participants. Costs are in 2022 US $.

Item		Unit cost (US $)	Units	Total annual cost (US $)
Participant-related				
	Bus pass	41.25 per month	12 months per participant	29,700
	Sim card or phone plan	210 per 6 months	12 months per participant	25,200
	Subtotal	—[a]	—	54,900
Program-related				
	SMS text messaging platform maintenance	99 per month	12 months for all participants	1188
	HIE[b] data access	200 per hour	750 hours for all participants	15,000
	Subtotal	—	—	16,188
Personnel-related				
	Program coordinator[c]	17 per hour	224 hours for all participants	3808
	Case worker	32.44 per hour	2500 hours for all participants	81,100
	Subtotal	—	—	84,908
Total		—	—	155,996

[a]Not applicable.

[b]HIE: health information exchange.

[c]Includes time spent on setting up (1 h/participant) and training (1 h/participant) participants and maintenance (2 h/wk).

Table . Per participant per year costs of Interactive Care Coordination and Navigation, a mobile health intervention for people experiencing homelessness in the United States, for a cost analysis. Costs are in 2022 US $.

Category		Cost PPPY[a] (US $)
Startup cost		
	Participant-related	195.30
	Program-related	70.11
	Subtotal	265.41
Recurring cost		
	Participant-related	915.00
	Program-related	269.80
	Personnel-related	1415.13
	Subtotal	2599.93
Total cost		2865.34

[a]PPPY: per participant per year.

The total PPPY cost of iCAN intervention varied slightly under different scenarios (Table 4). For example, when startup costs were excluded (Scenario A), recurring costs PPPY was US $2600. Excluding the cost of HIE data access (Scenario C) led to a similar PPPY cost of US $2615. The removal of bus passes and HIE data access (Scenario D) produced the lowest cost of US $2120 PPPY.

XSL•FO
RenderX

Table . Results of scenario analyses for a cost analysis of Interactive Care Coordination and Navigation, an mobile health intervention for people experiencing homelessness in the United States. All costs are in PPPY[a] 2022 US $.

Type of cost	Base case	Scenario A	Scenario B	Scenario C	Scenario D
Startup costs (US $)	265	0	265	265	265
Recurring costs (US $)	2600	2600	2105	2350	1855
Total iCAN[b] costs (US $)	2865	2600	2370	2615	2120
Savings threshold	7.8	7.0	6.4	7.1	5.7

[a]PPPY: per participant per year.

[b]iCAN: Interactive Care Coordination and Navigation.

In Table 4, the savings threshold refers to the percent (%) reduction in average health care costs for ED and inpatient care among people experiencing homelessness. The base case includes both startup and recurring costs. Scenario A includes recurring costs only. Scenario B includes startup costs and recurring costs without the bus pass. Scenario C include startup costs and recurring costs without HIE data access. Scenario D includes startup costs and recurring costs without the bus pass and HIE data access.

Cost-Benefit Analysis

When the cost of iCAN PPPY equals the base-case estimate (US $2865), the savings threshold is 7.8% (Table 2). This means that if average health care costs (US $36,917) among people experiencing homelessness were reduced by more than 7.8% through iCAN, the financial benefits would outweigh the costs of the intervention. A higher savings threshold is associated with a higher benefit-cost ratio for iCAN (Figure 1). When 10% ($3692/$N36,917) of average health care costs among people experiencing homelessness are averted (equal to the average cost of 3 ED visits), iCAN has a benefit-cost ratio of 1.29. When health care costs are reduced by 25% ($9229/N$36,917; equal to 56% [$9229/N$16,609] of the average cost of an inpatient visit), the benefit-cost ratio is 3.22, which means that iCAN produces US $2.22 in health care savings per US $1 spent.

Figure 1. Benefit-cost ratio of Interactive Care Coordination and Navigation, an mobile health intervention for people experiencing homelessness in the United States, under varying savings thresholds.

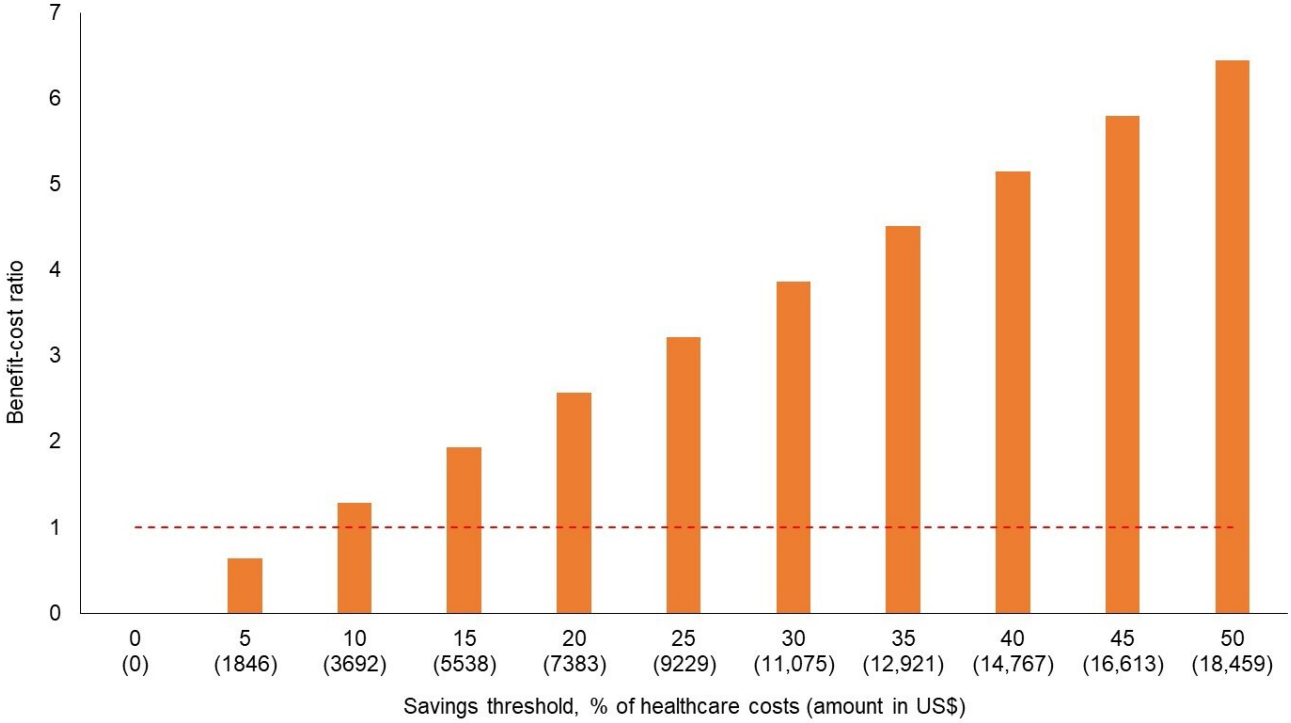


Figure 1 shows the benefit-cost ratio of iCAN under different savings thresholds, which is the percent (%) reduction in average health care costs for ED and inpatient care among people experiencing homelessness. Average health care costs were calculated using the weighted average of ED and inpatient visits among people experiencing homelessness from previous studies, whose value was equivalent to the median. The dotted red line marks a benefit-cost ratio of 1, which occurs when the benefits or financial savings of iCAN outweigh its costs. The probabilistic sensitivity analyses, which account for uncertainty in the costs of iCAN and the average health care costs of people experiencing homelessness, found a median savings threshold of 8.2% (95% probable interval 3.2 - 94.0). The average savings threshold was higher at 21%.

XSL•FO
RenderX

Budget Impact Analysis

The 5-year financial cost of implementing the iCAN intervention in Travis County for 10,250 people experiencing homelessness was around US $28.7 million (Table 5). If at least 5% of average health care costs among people experiencing homelessness are averted, the net budget impact is US $9.7 million. If average health care savings increased to 7%, the net budget impact of iCAN is further reduced to US $2.2 million.

Table . A 5-year budget impact analysis of Interactive Care Coordination and Navigation, an mobile health intervention for people experiencing homelessness in the United States (the 5-year total costs were calculated for the total number of people experiencing homelessness participating in Interactive Care Coordination and Navigation and health care savings were calculated based on the average health care costs for emergency department and inpatient care among people experiencing homelessness per year [US $36,917] among those who use these services).

	Year 1	Year 2	Year 3	Year 4	Year 5	5-year total
Total number of participants, n	1915	1980	2048	2117	2189	10,250
Intervention costs (US $)						
Fixed costs	4207	0	0	0	0	4207
Startup costs, PPPY[a]	195	195	195	195	195	2,001,764
Recurring costs, PPPY	2600	2600	2600	2600	2600	26,648,502
Total financial cost (US $)	5,357,637	5,535,448	5,723,653	5,918,257	6,119,478	28,654,472
Net budget impact (US $)						
5% health care savings	1,822,466	1,880,080	1,944,002	2,010,099	2,078,442	9,735,088
7% health care savings	408,397	417,933	432,142	446,835	462,028	2,167,335

[a]PPPY: per participant per year.

Discussion

Principal Results

In this study, we found, using a base-case estimate of US $2865, that a reduction in ED and hospital costs (estimated to be US $36,916 PPPY) of about 8%, which equals the cost of 2 ED visits PPPY, through iCAN would outweigh the costs of the intervention. Higher reductions in ED and hospital costs would result in a higher benefit-cost ratio for iCAN. While the economic value of mHealth interventions have been previously explored, none have focused on interventions for people experiencing homelessness. In total, 29 out of 39 studies evaluated in a systematic review conducted by Iribarren et al [44] in 2017 concluded that mHealth interventions show cost-effectiveness, economic benefit, or cost savings. Just over half of the mHealth interventions included SMS text messaging as the mHealth-related function in the intervention, with 77% (17/22) of these interventions reporting positive cost outcomes [44]. However, many limitations were identified in this systematic review. Of the 39 studies included in the review, only 12 were considered high-quality economic evaluations, which was defined as meeting at least 90% (21/24) of the items in the CHEERS checklist. In 2022, Gentili et al [13] conducted a systematic review of digital health interventions, which includes mHealth and other technologies such as telehealth, evaluating studies published after the time period evaluated by Iribarren et al [44]. Furthermore, 35 studies were included, and the findings were relatively consistent with that of Iribarren et al [44]. Only 12 of the 35 studies met at least 90% (21/24) of the items on the CHEERS checklist [13]. While the quality of economic evaluation was the focus of these systematic reviews, there is little evaluation of the budgetary impact and scalability of the included studies. Budget impact analyses were not conducted for these studies, and any evaluations on the economic scalability and sustainability of the interventions were preliminary, suggesting the need for formal budget impact analyses on mHealth interventions.

Another mHealth intervention used SMS to increase participation in their own care and decrease the number of no-show appointments in homeless veterans at a metropolitan Veterans Health Administration hospital. There was a significant decrease in the number of appointments cancelled by the patient and no-show appointments, and there was a significant reduction in ED visits. This study also showed that there were significant cost-savings when comparing the costs of implementing and maintaining SMS appointment reminders with the cost of unutilized services due to cancelled and no-show appointments [45]. A systematic review conducted in 2014 found that studies that evaluated cost-effectiveness showed cost savings associated with SMS services compared with standard of care while maintaining equal efficacy of interventions [46]. Even in the mHealth interventions where the outcomes were not statistically significant or cost-savings were not addressed, a reduction in preventable and nonemergent ED visits suggest likely significant cost-savings for health care payers in the community [13,44,47].

Our study adds to this literature by estimating the minimum health care savings needed for an mHealth intervention like iCAN to have positive net financial benefits. However, our analysis excludes other benefits that may be associated with iCAN, such as improvements in health outcomes and health-related quality of life. These health-related benefits

XSL•FO
RenderX

should be quantified, valued, and added to the cost-benefit analysis or cost-effectiveness analysis to gain a more comprehensive picture of iCAN's overall value.

Research has shown that people experiencing homelessness use more health care services than individuals with stable housing. In our analysis, we relied on previously published estimates of ED and inpatient health care use among people experiencing homelessness in Texas, which all found that use of these services is common (>1) among people experiencing homelessness who have a history of using these services. For example, single-center retrospective studies in Harris County, Texas, and Fort Worth, Texas found that people experiencing homelessness who used the ED had an average of 3.82 and 3.29 ED visits, respectively [33,34]. These numbers align with recent estimates from the National Hospital Ambulatory Medical Care Survey that found an ED visit rate of 310 per 100 persons per year in 2020 - 2021[37]. Estimates of inpatient visits among people experiencing homelessness in Texas are more limited and appear higher than numbers from other jurisdictions. A study of veterans in Pittsburgh, Pennsylvania, estimated the average number of inpatient visits at 0.7 (SD 1.6) among those who were hospitalized in 2012 - 2013, while single-center studies in Boston, Massachusetts, and Seattle, Washington State, found slightly higher numbers (>2) [40,48], which align with research showing a high risk of readmission among people experiencing homelessness [2,3]. Using these health care utilization numbers, we calculated an average health care cost comparable with previously published estimates of the cost of care among people experiencing homelessness [35,49,50]. Additional research is needed to understand the health care use and costs among people experiencing homelessness.

This study not only provides insight to the cost-benefit of the iCAN mHealth intervention, but also highlights the scalability and sustainability of an mHealth intervention like iCAN in a metropolitan area. Including a budget impact analysis of iCAN demonstrates that expanding the services provided by iCAN over a 5-year period is both possible and cost-saving. While existing literature on the economic implications of mHealth interventions exists, the majority of these studies did not conduct formal budget impact analyses to determine the sustainability of implementing an intervention on a larger scale [13,44]. Our economic evaluation and budget impact analysis balances both real-world data with necessary assumptions on a realistic scale, and many of the assumptions used in this study can be replaced with more specific real-world data after the iCAN RCT has concluded.

This economic evaluation and budget impact analysis provides insightful information in an emerging field where concrete data are lacking in several aspects. In published studies conducting budget impact analyses, the interventions being studied are typically medications or treatment protocols, not multicomponent behavioral health interventions [21,51-54]. While there are dozens of economic evaluations of mHealth interventions, there are very few formal budget impact analyses associated with these economic evaluations [13,44]. Our economic evaluation and budget impact analysis aims to fill this gap about mHealth interventions and the economic scalability and sustainability associated with expanding the services provided.

Limitations

The primary limitations of this study involve the assumptions made in the economic evaluation. The costs associated with the iCAN intervention were estimated from expenditure data from the RCT, which likely overestimate the actual costs of the intervention if implemented in real-world setting due to costs required to ensure accurate trial design and adherence to protocol [20]. We addressed this issue by adjusting the fixed costs included in our budget impact analysis, however, other economies of scale are potentially still excluded from our analysis. Health care costs related to ED and inpatient visits were estimated using national health care expenditure data (MEPS-HC) because such data were not available through the trial or for Travis County specifically. With possible variations in health care cost and utilization among people experiencing homelessness, our probabilistic sensitivity analyses explored the impact of uncertainty on our findings and indeed found a wide range of savings thresholds for iCAN. This suggests iCAN's economic value will be highly influenced by the change in health care costs and utilization of people experiencing homelessness who receive the intervention. Finally, we also had to make simplifying assumptions in the budget impact analysis, including the current size and growth of the people experiencing homelessness population in Travis County over the next 5 years. Our budget impact analysis can be easily updated with more precise inputs to understand the financial aspects of expanded access to iCAN.

Conclusions

This exploratory economic evaluation gives estimates of the cost of the iCAN mHealth intervention while exploring the cost-savings and budget impact of implementing the iCAN mHealth intervention on a larger scale. By using data from the iCAN RCT and other available resources, this economic evaluation determined that implementing the iCAN mHealth intervention for people experiencing homelessness in a metropolitan area provides financial cost-benefit if 1 hospitalization or 2 ED visits can be avoided. Future studies should explore the feasibility of implementing the iCAN mHealth intervention using data from the completed iCAN RCT to limit uncertainty in the cost of scaling up the intervention. Health care payers should examine the benefits of implementing mHealth interventions like this one to prevent unnecessary health care utilization and the associated costs among people experiencing homelessness.

Acknowledgments

Portions of the content (ie, the randomized controlled trial described in the methods) of this manuscript are funded under grant 1R18 HS027750-01 from the Agency for Healthcare Research and Quality (AHRQ), US Department of Health and Human

XSL•FO
RenderX

Services (HHS). The authors are solely responsible for the contents, findings, and conclusions, which do not necessarily represent the views of AHRQ. Readers should not interpret any statement in this report as an official position of AHRQ or of HHS. None of the authors have any affiliation or financial involvement that conflict with the material presented in this report. Portions of this manuscript were previously presented as abstracts and posters at the St. David's Center for Health Promotion and Disease Prevention Research in Underserved Populations Conference at the University of Texas School of Nursing in March 2023 and the AcademyHealth Annual Research Meeting in Seattle, WA, in June 2023.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Annual emergency department and inpatient health care utilization among people experiencing homelessness.
[DOCX File, 16 KB - formative_v9i1e64973_app1.docx]

Multimedia Appendix 2
Scenarios explored in the cost analysis of Interactive Care Coordination and Navigation, an mHealthmobile health intervention for people experiencing homelessness in the United States.
[DOCX File, 15 KB - formative_v9i1e64973_app2.docx]

References

1. Vohra N, Paudyal V, Price MJ. Homelessness and the use of emergency department as a source of healthcare: a systematic review. Int J Emerg Med 2022 Jul 28;15(1):32. [doi: 10.1186/s12245-022-00435-3] [Medline: 35902803]
2. Miyawaki A, Hasegawa K, Figueroa JF, Tsugawa Y. Hospital readmission and emergency department revisits of homeless patients treated at homeless-serving hospitals in the USA: observational study. J Gen Intern Med 2020 Sep;35(9):2560-2568. [doi: 10.1007/s11606-020-06029-0] [Medline: 32666492]
3. Banerjee S, Yassin M, Dyer WT, Thomas TW, Rodriguez LA, Schmittdiel J. Hospital readmissions among patients experiencing homelessness: an electronic health record data study. Perm J 2024 Mar 15;28(1):55-61. [doi: 10.7812/TPP/23.075] [Medline: 38108331]
4. Liu M, Hwang SW. Health care for homeless people. Nat Rev Dis Primers 2021;7(1):5. [doi: 10.1038/s41572-020-00241-2] [Medline: 33446661]
5. Thurman W, Moczygemba LR, Baffoe JO. "Without my medication, I'm a wreck": photo-elicitation to explore medication use among people experiencing homelessness. Res Social Adm Pharm 2022 Jul;18(7):3149-3157. [doi: 10.1016/j.sapharm.2021.08.015] [Medline: 34479800]
6. What is digital health? US Food and Drug Administration. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health [accessed 2025-01-13]
7. Velmovitsky PE, Kirolos M, Alencar P, Leatherdale S, Cowan D, Morita PP. Leveraging mHealth technologies for public health. JMIR Public Health Surveill 2024 Sep 12;10:e49719. [doi: 10.2196/49719] [Medline: 39265164]
8. mHealth: use of appropriate digital technologies for public health: report by the director-general. : World Health Organization; 2017 URL: https://iris.who.int/bitstream/handle/10665/274134/B142_20-en.pdf?sequence=1&isAllowed=y [accessed 2025-03-06]
9. Kyaw TL, Ng N, Theocharaki M, Wennberg P, Sahlen KG. Cost-effectiveness of digital tools for behavior change interventions among people with chronic diseases: systematic review. Interact J Med Res 2023 Feb 16;12:e42396. [doi: 10.2196/42396] [Medline: 36795470]
10. Arora S, Peters AL, Burner E, Lam CN, Menchine M. Trial to examine text message-based mHealth in emergency department patients with diabetes (TExT-MED): a randomized controlled trial. Ann Emerg Med 2014;63(6):745-754. [doi: 10.1016/j.annemergmed.2013.10.012] [Medline: 24225332]
11. Eustache JH, Latimer EA, Liberman AS, et al. A mobile phone app improves patient-physician communication and reduces emergency department visits after colorectal surgery. Dis Colon Rectum 2023 Jan 1;66(1):130-137. [doi: 10.1097/DCR.0000000000002187] [Medline: 34933314]
12. Ladley A, Hieger AW, Arthur J, Broom M. Educational text messages decreased emergency department utilization among infant caregivers: a randomized trial. Acad Pediatr 2018 Aug;18(6):636-641. [doi: 10.1016/j.acap.2018.02.003] [Medline: 29432907]
13. Gentili A, Failla G, Melnyk A, et al. The cost-effectiveness of digital health interventions: a systematic review of the literature. Front Public Health 2022;10:787135. [doi: 10.3389/fpubh.2022.787135] [Medline: 36033812]

XSL•FO
RenderX

14. Moczygemba LR, Cox LS, Marks SA, Robinson MA, Goode JVR, Jafari N. Homeless patients' perceptions about using cell phones to manage medications and attend appointments. Int J Pharm Pract 2017 Jun;25(3):220-230. [doi: 10.1111/ijpp.12321] [Medline: 27896909]
15. Moczygemba LR, Thurman W, Tormey K, Hudzik A, Welton-Arndt L, Kim E. GPS mobile health intervention among people experiencing homelessness: pre-post study. JMIR Mhealth Uhealth 2021 Nov 3;9(11):e25553. [doi: 10.2196/25553] [Medline: 34730550]
16. Thurman W, Semwal M, Moczygemba LR, Hilbelink M. Smartphone technology to empower people experiencing homelessness: secondary analysis. J Med Internet Res 2021 Sep 29;23(9):e27787. [doi: 10.2196/27787] [Medline: 34586073]
17. Lal S, Elias S, Sieu V, Peredo R. The use of technology to provide mental health services to youth experiencing homelessness: scoping review. J Med Internet Res 2023 Jan 16;25:e41939. [doi: 10.2196/41939] [Medline: 36645703]
18. Prinja S, Gupta A, Bahuguna P, Nimesh R. Cost analysis of implementing mHealth intervention for maternal, newborn & child health care through community health workers: assessment of ReMIND program in Uttar Pradesh, India. BMC Pregnancy Childbirth 2018;18(1):390. [doi: 10.1186/s12884-018-2019-3] [Medline: 30285669]
19. Cheng Q, Dang T, Nguyen TA, et al. mHealth application for improving treatment outcomes for patients with multidrug-resistant tuberculosis in Vietnam: an economic evaluation protocol for the V-SMART trial. BMJ Open 2023;13(12):e076778. [doi: 10.1136/bmjopen-2023-076778] [Medline: 38081668]
20. Heapy AA, Driscoll MA, Buta E, et al. Co-operative pain education and self-management (copes) expanding treatment for real-world access (extra): pragmatic trial protocol. Pain Med 2020;21(Suppl 2):S21-S28. [doi: 10.1093/pm/pnaa365] [Medline: 33313733]
21. Avanceña ALV, Brody C, Chhoun P, Tuot S, Yi S. Connecting female entertainment workers in Cambodia to health care services using mHealth: economic evaluation of mobile link. JMIR Form Res 2024 Jul 25;8:e52734. [doi: 10.2196/52734] [Medline: 39052328]
22. Husereau D, Drummond M, Augustovski F, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. BMJ 2022;376:e067975. [doi: 10.1136/bmj-2021-067975] [Medline: 35017145]
23. Babigumira JB, Dolan S, Shade S, et al. Applied economic evaluation of digital health interventions. I-TECH. URL: https://www.go2itech.org/wp-content/uploads/2021/02/I-TECH_HIS_Economic_Evaluation.pdf [accessed 2024-07-30]
24. Sullivan SD, Mauskopf JA, Augustovski F, et al. Budget impact analysis-principles of good practice: report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. Value Health 2014;17(1):5-14. [doi: 10.1016/j.jval.2013.08.2291] [Medline: 24438712]
25. Rascati K. Essentials of Pharmacoeconomics: Lippincott Williams & Wilkins; 2020.
26. Baggett TP, O'Connell JJ, Singer DE, Rigotti NA. The unmet health care needs of homeless adults: a national study. Am J Public Health 2010;100(7):1326-1333. [doi: 10.2105/AJPH.2009.180109] [Medline: 20466953]
27. Fryling LR, Mazanec P, Rodriguez RM. Barriers to homeless persons acquiring health insurance through the Affordable Care Act. J Emerg Med 2015 Nov;49(5):755-762. [doi: 10.1016/j.jemermed.2015.06.005] [Medline: 26281811]
28. Špacírová Z, Epstein D, García-Mochón L, Rovira J, Olry de Labry Lima A, Espín J. A general framework for classifying costing methods for economic evaluation of health care. Eur J Health Econ 2020;21(4):529-542. [doi: 10.1007/s10198-019-01157-9] [Medline: 31960181]
29. US Bureau of Labor Statistics. URL: https://data.bls.gov/timeseries/CUUR0000SAM?output_view=pct_12mths [accessed 2022-12-14]
30. Student employee compensation. The University of Texas at Austin Human Resources. URL: https://hr.utexas.edu/student/student-employee-compensation [accessed 2024-07-27]
31. Petrovich JC, Hunt JJ, North CS, Pollio DE, Roark Murphy E. Comparing unsheltered and sheltered homeless: demographics, health services use and predictors of health services use. Community Ment Health J 2020 Feb;56(2):271-279. [doi: 10.1007/s10597-019-00470-0] [Medline: 31552539]
32. Schiefelbein EL, Olson JA, Moxham JD. Patterns of health care utilization among vulnerable populations in Central Texas using data from a regional health information exchange. J Health Care Poor Underserved 2014;25(1):37-51. [doi: 10.1353/hpu.2014.0020] [Medline: 24509011]
33. Amato S, Nobay F, Amato DP, Abar B, Adler D. Sick and unsheltered: homelessness as a major risk factor for emergency care utilization. Am J Emerg Med 2019 Mar;37(3):415-420. [doi: 10.1016/j.ajem.2018.06.001] [Medline: 29891125]
34. Holmes CT, Holmes KA, MacDonald A, et al. Dedicated homeless clinics reduce inappropriate emergency department utilization. J Am Coll Emerg Physicians Open 2020;1(5):829-836. [doi: 10.1002/emp2.12054] [Medline: 33145527]
35. Buck DS, Brown CA, Mortensen K, Riggs JW, Franzini L. Comparing homeless and domiciled patients' utilization of the Harris County, Texas public hospital system. J Health Care Poor Underserved 2012;23(4):1660-1670. [doi: 10.1353/hpu.2012.0171] [Medline: 23698680]
36. Coe AB, Moczygemba LR, Harpe SE, Gatewood SBS. Homeless patients' use of urban emergency departments in the United States. J Ambul Care Manage 2015;38(1):48-58. [doi: 10.1097/JAC.0000000000000034] [Medline: 25469578]
37. Schappert SM, Santo L. Emergency department visits by homeless status and sex: United States, 2016-2021. Natl Health Stat Report 2024 Jun(204):1-21. [Medline: 38912857]

XSL•FO
RenderX

38. Ball MAZ, Sack DE, Druffner SA, et al. Characteristics and health care utilization of patients with housing insecurity in the ED. JAMA Netw Open 2024 Apr 1;7(4):e248565. [doi: 10.1001/jamanetworkopen.2024.8565] [Medline: 38669017]
39. Gabrielian S, Yuan AH, Andersen RM, Rubenstein LV, Gelberg L. VA health service utilization for homeless and low-income Veterans: a spotlight on the VA Supportive Housing (VASH) program in greater Los Angeles. Med Care 2014;52(5):454-461. [doi: 10.1097/MLR.0000000000000112] [Medline: 24714583]
40. Lin WC, Bharel M, Zhang J, O'Connell E, Clark RE. Frequent emergency department visits and hospitalizations among homeless people with Medicaid: implications for Medicaid expansion. Am J Public Health 2015 Nov;105 Suppl 5(Suppl 5):S716-S722. [doi: 10.2105/AJPH.2015.302693] [Medline: 26447915]
41. MEPS-HC data tools. AHRQ. URL: https://datatools.ahrq.gov/meps-hc [accessed 2024-11-08]
42. Austin count of homeless people returns this weekend. Here's what we know. Austin American-Statesman. URL: https://www.statesman.com/story/news/2023/01/27/austin-texas-annual-count-homeless-people-ending-community-homlessness-coalition/69839931007/ [accessed 2024-03-07]
43. Department of Housing and Urban Development (HUD). URL: https://www.hud.gov/ [accessed 2024-03-07]
44. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. Plos one 2017;12(2):e0170581. [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
45. McInnes DK, Petrakis BA, Gifford AL, et al. Retaining homeless veterans in outpatient care: a pilot study of mobile phone text message appointment reminders. Am J Public Health 2014;104(Suppl 4):S588-S594. [doi: 10.2105/AJPH.2014.302061] [Medline: 25100425]
46. Beratarrechea A, Lee AG, Willner JM, Jahangir E, Ciapponi A, Rubinstein A. The impact of mobile health interventions on chronic disease outcomes in developing countries: a systematic review. Telemed J E Health 2014;20(1):75-82. [doi: 10.1089/tmj.2012.0328] [Medline: 24205809]
47. Katz R, Mesfin T, Barr K. Lessons from a community-based mHealth diabetes self-management program: it's not just about the cell phone. J Health Commun 2012;17(Suppl 1):67-72. [doi: 10.1080/10810730.2012.650613] [Medline: 22548601]
48. Srebnik D, Connor T, Sylla L. A pilot study of the impact of housing first-supported housing for intensive users of medical hospitalization and sobering services. Am J Public Health 2013;103(2):316-321. [doi: 10.2105/AJPH.2012.300867] [Medline: 23237150]
49. O'Toole TP, Johnson EE, Borgia M, et al. Population-tailored care for homeless veterans and acute care use, cost, and satisfaction: a prospective quasi-experimental trial. Prev Chronic Dis 2018 Feb 15;15:E23. [doi: 10.5888/pcd15.170311] [Medline: 29451116]
50. Mitchell MS, León CLK, Byrne TH, Lin WC, Bharel M. Cost of health care utilization among homeless frequent emergency department users. Psychol Serv 2017;14(2):193-202. [doi: 10.1037/ser0000113] [Medline: 28481604]
51. Faleiros DR, Álvares J, Almeida AM, et al. Budget impact analysis of medicines: updated systematic review and implications. Expert Rev Pharmacoecon Outcomes Res 2016;16(2):257-266. [doi: 10.1586/14737167.2016.1159958] [Medline: 26923561]
52. Jahn B, Todorovic J, Bundo M, et al. Budget impact analysis of cancer screening: a methodological review. Appl Health Econ Health Policy 2019 Aug;17(4):493-511. [doi: 10.1007/s40258-019-00475-6] [Medline: 31016686]
53. Mohammadnezhad G, Sattarpour M, Moradi N. Budget impact analysis of breast cancer medications: a systematic review. J Pharm Policy Pract 2022;15(1):105. [doi: 10.1186/s40545-022-00493-1] [Medline: 36581921]
54. Luo Z, Ruan Z, Yao D, Ung COL, Lai Y, Hu H. Budget impact analysis of diabetes drugs: a systematic literature review. Front Public Health 2021;9:765999. [doi: 10.3389/fpubh.2021.765999] [Medline: 34869180]

Abbreviations

CHEERS: Consolidated Health Economic Evaluation Reporting Standards
ED: emergency department
HIE: health information exchange
I-TECH: International Training and Education Center for Health
iCAN: Interactive Care Coordination and Navigation
ISPOR: The International Society for Pharmacoeconomics and Outcomes Research
MEPS-HC: Medical Expenditure Panel Survey-Household Component
mHealth: mobile health
PPPY: per participant per year
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by A Mavragani; submitted 01.08.24; peer-reviewed by L Pirhonen, MP Montgomery; revised version received 13.01.25; accepted 16.01.25; published 18.03.25.

Please cite as:
McCullough HP, Moczygemba LR, Avanceña ALV, Baffoe JO
The Interactive Care Coordination and Navigation mHealth Intervention for People Experiencing Homelessness: Cost Analysis, Exploratory Financial Cost-Benefit Analysis, and Budget Impact Analysis
JMIR Form Res 2025;9:e64973
URL: https://formative.jmir.org/2025/1/e64973
doi:10.2196/64973

© Hannah P McCullough, Leticia R Moczygemba, Anton L V Avanceña, James O Baffoe. Originally published in JMIR Formative Research (https://formative.jmir.org), 18.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Digital Health Literacy and Attitudes Toward eHealth Technologies Among Patients With Cardiovascular Disease and Their Implications for Secondary Prevention: Survey Study

Greta Ullrich[1], Dr Med; Alexander Bäuerle[2,3], PhD; Hannah Vogt[1]; Amir Abbas Mahabadi[1], Prof Dr Med; Katrin Paldán[4], Dr Phil; Daniel Messiha[1], Dr Med; Lisa Maria Jahre[2,3], MSc; Christos Rammos[1], Prof Dr Med; Tienush Rassaf[1], Prof Dr Med; Julia Lortz[1], Prof Dr Med

[1]Department of Cardiology and Vascular Medicine, West-German Heart and Vascular Center Essen, University of Duisburg-Essen, Essen, Germany

[2]Clinic for Psychosomatic Medicine and Psychotherapy, LVR-University Hospital Essen, Essen, Germany

[3]Center for Translational Neuro- and Behavioral Sciences (C-TNBS), University of Duisburg-Essen, Essen, Germany

[4]HCT Research, Vorarlberg University of Applied Science, Dornbirn, Austria

Corresponding Author:
Julia Lortz, Prof Dr Med
Department of Cardiology and Vascular Medicine, West-German Heart and Vascular Center Essen, University of Duisburg-Essen, Essen, Germany

Abstract

Background: Cardiovascular disease is the major cause of death worldwide, leading to a significant socioeconomic burden. Although secondary prevention is a cornerstone in chronic disease management, adherence to guideline recommendations in this regard often falters, leading to suboptimal outcomes. While eHealth technologies are promising for improving treatment adherence, they also represent a new approach to secondary prevention. However, a common critique is that extensive digitalization may not adequately address the needs of older adults with chronic medical conditions.

Objective: This study aims to analyze eHealth literacy, digital use patterns, and general attitudes toward digital technologies in a collective of patients with cardiovascular disease to identify potential obstacles in implementing mobile health technologies in secondary preventive therapy.

Methods: This survey-based study was a part of the baseline examination of the PreventiPlaque trial. It involved 240 participants with known coronary artery disease. The assessment evaluated their current understanding of the general use of digital devices. The questionnaire covered aspects such as the duration of daily use, personal attitudes, and the perceived burden associated with digital media. eHealth literacy was assessed within the target population and general demographic data were gathered, focusing on cardiovascular comorbidities and risk factors.

Results: The analysis revealed an average age of 61.9 (SD 8.9) years, with 59.9% (n=144) of the participants being male. Overall, 37.3% (n=90) of the participants had previous knowledge of digital health interventions, while only 17.8% (n=41) had used them. Despite the generally low practical application within this study population, there was a high level of confidence in handling digital devices, with 61.9% (n=149) expressing themselves as either rather confident or very confident. Regarding the levels of eHealth literacy among the participants, 71.2% (n=170) claimed to be familiar with locating health information on the internet, and 64% (n=153) of participants felt capable of critically evaluating its quality. These levels of digital confidence were consistent across all age groups. Moreover, internet use rates remained high even among the older participants, with 80% (n=192) of those participants older than 75 years using the internet for 1-3 hours a day.

Conclusions: The study unveiled a notable confidence level among participants regarding the use of digital devices, coupled with a favorable attitude toward digital media evident across all age brackets. Remarkably, internet use rates remained high, even among older participants. The actual utilization of digital health interventions was relatively low, potentially stemming from challenges in locating reliable sources. These findings emphasize the prospect of future eHealth interventions customized to the distinct needs and preferences of patients in cardiovascular disease management. Recognizing the incongruity between confidence in device use and the restricted adoption of digital health tools can guide the development of focused interventions to narrow this divide.

Trial Registration: ClinicalTrials.gov NCT05096637; https://clinicaltrials.gov/study/NCT05096637

(JMIR Form Res 2025;9:e63057) doi:10.2196/63057

XSL•FO
RenderX

KEYWORDS

cardiovascular diseases; telemedicine; eHealth; patient-centered approach; digital health literacy; digital health; cardiovascular disease; mortality; artery disease; ischemic; heart disease; diabetes mellitus; obesity; patient education; eHealth literacy; mobile phone

Introduction

Cardiovascular disease remains the leading cause of global mortality, with coronary artery disease and ischemic heart disease alone contributing to 16% of total annual deaths worldwide [1]. Over the past 2 decades, ischemic heart disease has demonstrated the most rapid growth in overall death rates globally, resulting in nearly 9 million annual deaths [1]. Cardiovascular disease is intricately linked to various comorbidities and a severe mental health burden [2]. While advancements in the understanding and treatment of cardiovascular disease persist, there is an increasing emphasis on secondary prevention. Following guideline recommendations, secondary prevention aims to address modifiable risk factors such as hyperlipidemia, nicotine consumption, arterial hypertension, diabetes mellitus, obesity, chronic stress, and lack of physical activity [3]. Despite the preventive potential of lifestyle modifications supported by guidelines, their implementation remains inadequate. Primary obstacles include the challenge of ensuring adherence to long-term behavioral changes and the scarcity of medical resources and time needed for comprehensive patient education in chronic disease management [4,5]. Digital health interventions emerge as a promising avenue to educate and empower patients, encouraging an active role in disease management while optimizing the control and monitoring of modifiable risk factors [6]. They offer a potential solution to the current deficiencies in medical infrastructure, addressing the escalating demand associated with the increasing prevalence of cardiovascular disease [7]. However, a frequently cited challenge in the expanding digitization of health care is the consideration of special needs among older patients concerning the development and structure of digital health interventions [8]. Given that cardiovascular disease predominantly affects older individuals, it is crucial to not disadvantage them to fully use the modern treatment options.

The objective of this study was to delve into the digital use patterns of patients with cardiovascular disease expressing an interest in digital health interventions. The aim was to assess and identify potential obstacles and challenges in the implementation of mobile health (mHealth) or eHealth technologies for secondary prevention of cardiovascular disease. The study sought to evaluate patients' existing knowledge and opinions regarding smartphones, internet, and digital health interventions while focusing on uncovering possible age-dependent differences. The overarching goal was to gain insights that would inform the design of digital health interventions tailored to meet the specific needs and requirements of individuals managing cardiovascular conditions.

Methods

Study Design and Participants

We conducted a survey-based assessment at the University Hospital Essen, West German Heart and Vascular Center, Department of Cardiology and Vascular Medicine. The questionnaires used were part of the baseline examination of the PreventiPlaque trial. PreventiPlaque is a registered randomized clinical trial (NCT05096637) testing the effects of a smartphone app that included atherosclerotic plaque visualization on adherence to secondary preventive therapy [9]. The recruitment of participants for the PreventiPlaque trial occurred in 2022. Patients with atherosclerotic cardiovascular disease were eligible for participation in this trial. This included patients with documented ischemic heart disease, acute coronary syndrome, and patients with proven peripheral artery disease [10]. Another requirement was adult age (18 years and older). Participants had to own a smartphone that was suitable for potential app use. Finally, patients had to give written informed consent to comply with the study protocol and had to be willing to participate in the study. Patients with insufficient knowledge of the German language, or unwillingness to use a smartphone app, were excluded from the study.

Ethical Considerations

The conduct of the trial was approved by the ethics committee of the Medical Faculty of the University of Duisburg-Essen (20 - 9157-BO). Written informed consent was provided by each participant before inclusion in the study. The study was conducted in accordance to the Declaration of Helsinki. Data were collected and analyzed using a pseudonymous form. The participants did not receive any form of financial compensation for participating in this trial.

Measurements

Sociodemographic and Medical Data

Basic sociodemographic data and medical data were collected, using a questionnaire with 15 items, including the patient's marital status, level of education, and current profession. Moreover, patients were asked to self-assess their level of physical activity, as well as the quality of their diet. We extracted medical information from the hospital's electronic medical records regarding the presence of major cardiovascular risk factors such as diabetes mellitus, arterial hypertension, and hypercholesterolemia, as well as the prevalence of cardiovascular comorbidities including coronary artery disease, peripheral artery disease, stroke, aortic syndrome, or chronic heart failure.

eHealth Literacy

The revised German eHealth Literacy Scale (GR-eHEALS) was used to assess the participants' skills to find and critically evaluate health information on the internet [11]. It is a validated

XSL•FO
RenderX

tool to measure eHealth literacy in patients with cardiovascular disease and the modified version contains 8 items and 2 subscales [12]. Responses could be given on a 5-point Likert scale (eg, "I know how to use the internet to find answers to my health-related questions," 1=does not apply to me, 5=does apply to me).

eHealth-Related Data

The self-generated eHealth data questionnaire started with 3 items examining the participant's general confidence in handling digital media, digital devices, and web-based platforms on the internet. They could respond on a 5-point Likert scale (1=not confident at all, 5=very confident). Patients were asked to self-assess their daily internet utilization period for private or work-related purposes. Moreover, internet anxiety was measured by 6 items with possible responses on a 5-point Likert scale (eg "I have concerns about using the internet," 1=does not apply to me, 5=does apply to me). Finally, it was determined whether the participants were already experienced in using digital health interventions, had heard about them, or knew where to find them. The self-generated eHealth items had already been used in previous studies [13-16] and have proven good reliability.

Statistical Analysis

We performed descriptive statistical analysis using SPSS (version 23; IBM Corp). Variables were presented as frequencies and percentages, as well as means and SDs. The participant's BMI was calculated using the patient's weight and height according to our hospital's electronic medical records. Since it is often brought forward that older patients may not be a suitable target group for digital health interventions, we analyzed internet use frequency and digital confidence in relation to the participant's age. To serve this purpose, we divided the participants into 4 age groups. To reach comparable sample sizes, we divided the age groups into <55, 55 - 64, 65 - 75, and >75-year-olds. Using these age groups, we performed an age-adjusted analysis of the variables "Internet-use," as well as "confidence in handling internet platforms/digital media/mobile technologies." This included an age-adjusted means comparison, using the Kruskal-Wallis test. We also conducted bivariate correlations between the participant's age and the 4 items "Internet-use" and "confidence in handling Internet platforms/digital media/mobile technologies," using Spearman's Rho correlation for ordinally scaled variables. Finally, we analyzed these items mentioned above, describing the participants' digital confidence adjusted to gender, and comparing the 2 formed subgroups "male" and "female."

Results

Socioeconomic Characteristics

In total, 240 patients completed the assessment. With a mean age of 61.9 (SD 8.9) years and 17.4% (n=42) of participants being older than 70 years, the study showed an older people collective with a total of 59.9% (n=144) of all participants being male. With 60% (n=144) in total, most participants were married, 51.5% (n=124) of participants were retired, and 22.5% (n=54) of participants were working full-time.

Comorbidities and Cardiovascular Risk Factors

An analysis of the prevalence of major cardiovascular risk factors showed that 79.6% (n=191) and 80.4% (n=193) of the study participants had hypercholesterinemia and arterial hypertension, respectively. One-fourth of the participants were active smokers (n=60, 25%) and 17% (n=41) had type-2 diabetes while 36.5% (n=88) of participants were overweight with a BMI of 25 - 30 kg/m² and another 34.9% (n=84) of participants were obese with a BMI of >30 kg/m² (Table S1 in Multimedia Appendix 1). While all patients were diagnosed with coronary artery disease, a total of 36.7% (n=88) of patients had also been diagnosed with peripheral artery disease. Moreover, 26.3% (n=63) of participants had known congestive heart failure and 5.8% (n=14) of participants had once experienced a stroke (Table S1 in Multimedia Appendix 1).

eHealth Data

Overview

Analyzing the frequency and duration of internet use per day, only 5.9% (n=14) of the study population did not use the internet daily and 53.4% (n=128) of participants stated an internet use of 1 - 3 hours per day on average (Table 1). In general, levels of confidence when handling digital media, digital devices, and internet platforms were high in this study population (Table 2). Regarding internet-based programs to promote health or provide health information, only 34.4% (n=83) of participants knew how these programs worked and 47.5% (n=114) of participants knew where to find them (Table 3). When asked about having doubts about using the internet, a total of 75.7% (n=181) of the study population stated that this did not apply at all or rather did not apply. Furthermore, almost 80% (n=192) of the study population did not feel negatively affected by carrying a mobile phone with them (Table 4).

XSL•FO
RenderX

Table . Cardiovascular risk factors and comorbidities (N=240).

Cardiovascular risk factors	Prevalence, n (%)
BMI>30	84 (34.9)
BMI>25	88 (36.5)
Type 2 diabetes	41 (17)
Nicotine dependency	60 (25)
Hypercholesterolemia	191 (79.6)
Arterial hypertension	193 (80.4)
Congestive heart failure	63 (26.3)
Peripheral artery disease	88 (36.7)
Stroke	14 (5.8)

Table . Daily internet use (n=236).

	How long are you using the internet for private purposes per day?, n (%)
Not at all	14 (5.9)
Less than 1 hour	69 (29.2)
1 - 3 hours	126 (53.4)
3 - 5 hours	20 (8.5)
More than 5 hours	7 (3)

Table . Digital confidence in handling digital media (n=236).

	Very insecure, n (%)	Rather insecure, n (%)	Partly insecure, n (%)	Rather confident, n (%)	Very confident, n (%)
How confident are you when handling digital media?	16 (6.8)	17 (7.2)	56 (23.7)	92 (39)	55 (23.3)
How confident are you when handling internet platforms?	20 (8.5)	21 (8.9)	61 (25.8)	81 (34.3)	53 (22.5)
How confident are you when handling digital devices?	20 (8.5)	16 (6.8)	54 (22.9)	92 (39)	54 (22.9)

Table . Knowledge regarding internet-based programs to promote health (n=236).

	I do not agree at all, n (%)	I rather do not agree, n (%)	Neither, n (%)	I rather agree, n (%)	I fully agree, n (%)
I can imagine what that might be	27 (11.4)	45 (19.1)	35 (14.8)	97 (41.1)	32 (13.6)
I know how those programs work	36 (15.3)	66 (28)	53 (22.5)	67 (28.4)	14 (5.9)
I know where to find these programs	35 (14.8)	50 (21.2)	39 (16.5)	84 (35.6)	28 (11.9)

eHealth Literacy

As presented in Table S2 in Multimedia Appendix 1, the statistical analysis of the eHealth literacy questionnaire results showed that the majority (n=152, 63.5%) of the study participants rather or fully agreed to know how to find internet platforms with helpful information regarding their overall health. Moreover, a total of 71.2% (n=171) of participants stated to knew how to use the internet as a useful tool to get answers to their questions. When asked, whether the participants thought they were able to critically evaluate the credibility of health information on the internet, a total of 64% (n=152) of participants stated that they were able to do so. However, only 40% (n=96) of the participants stated to knew digital sources available for health information, and 47.9% (n=115) of

XSL•FO
RenderX

participants felt secure enough to use information from the internet in order to make health-related decisions (Table 5).

Table . eHealth literacy (n=236).

	I do not agree at all, n (%)	I rather not agree, n (%)	Neither, n (%)	I rather agree, n (%)	I fully agree, n (%)
I know how to find internet platforms with helpful health information (n=236)	17 (7.2)	30 (12.7)	39 (16.5)	89 (37.7)	61 (25.8)
I know how to use the internet to get answers to my questions (n=236)	16 (6.8)	28 (11.9)	24 (10.2)	108 (45.8)	60 (25.4)
I know which sources for health information are available (n=235)	18 (7.7)	45 (19.1)	45 (19.1)	90 (38.3)	37 (1.7)
I know where I can find helpful health information on the internet (n=236)	18 (7.6)	32 (19.1)	39 (16.5)	103 (43.6)	44 (18.6)
I know how to use health information from the internet to help me (n=235)	24 (10.2)	48 (20.4)	50 (21.3)	88 (37.4)	25 (10.6)
I can critically evaluate health information on the internet (n=236)	17 (7.2)	23 (9.7)	45 (19.1)	101 (42.8)	50 (21.2)
I can distinguish between questionable and trustworthy internet resources with health information (n=236)	19 (8.1)	34 (14.4)	41 (17.4)	105 (44.5)	37 (15.7)
I feel secure to use information from the internet to make decisions regarding my health (n=236)	27 (11.4)	48 (20.3)	48 (20.3)	78 (33.1)	35 (14.8)

Age-Related Differences in Internet Use and Digital Confidence

As Figure 1 shows, using the internet for more than one hour a day, particularly for 1 - 3 hours per day, was the most common answer when asked about the duration of daily internet use. This result can be seen throughout all age groups. A longer duration of daily internet use of at least 3 hours was more common in participants younger than 55 years than in older groups. When analyzing confidence in handling digital media, it becomes clear that with older age, starting from the age group of 65 - 74-year-olds, the share of "partly confident," "rather not confident," and "not confident" individuals slightly increases, while still more than 50% (n=120) of the population remains "rather confident" or "very confident" (Figure 2). The share of at least "partly confident" participants is even higher in terms of confidence in handling digital devices such as smartphones and computers (Figure 3). This is again apparent throughout all age groups, even the oldest group of aged older than 75 years, contains less than 20% (n=48) of participants who are "not confident at all" in handling digital devices (Figure 3). Finally, confidence in handling internet platforms was generally lower than confidence in handling digital devices, and the share of participants who were "very confident," as well as "rather confident," decreased with older age. In the oldest age group of aged older than 75 years, 42% (n=101) of participants were "very confident" or "rather confident" in using internet platforms (Figure 4). Using the Kruskal Wallis test to compare internet use rates and digital confidence between the 4 age groups, there was no statistically significant difference for any of the examined items (>.05). Testing for correlations between the participants age and digital confidence, as well as daily internet use, the Spearman's-Rho correlation shows the only statistically significant result regarding the confidence in using internet platforms (ρ [rho]) of –.128; *P*=.048). In addition, there was no age-related statistically significant difference in the levels of confidence in using mobile technologies (ρ of –0.70; *P*=.28), digital media (ρ of –0.96*; P*=.14), or daily internet use (ρ of 0.36*; P*=.58).

XSL•FO
RenderX

Figure 1. Daily internet use in relation to age groups (indicated in years along the x-axis).

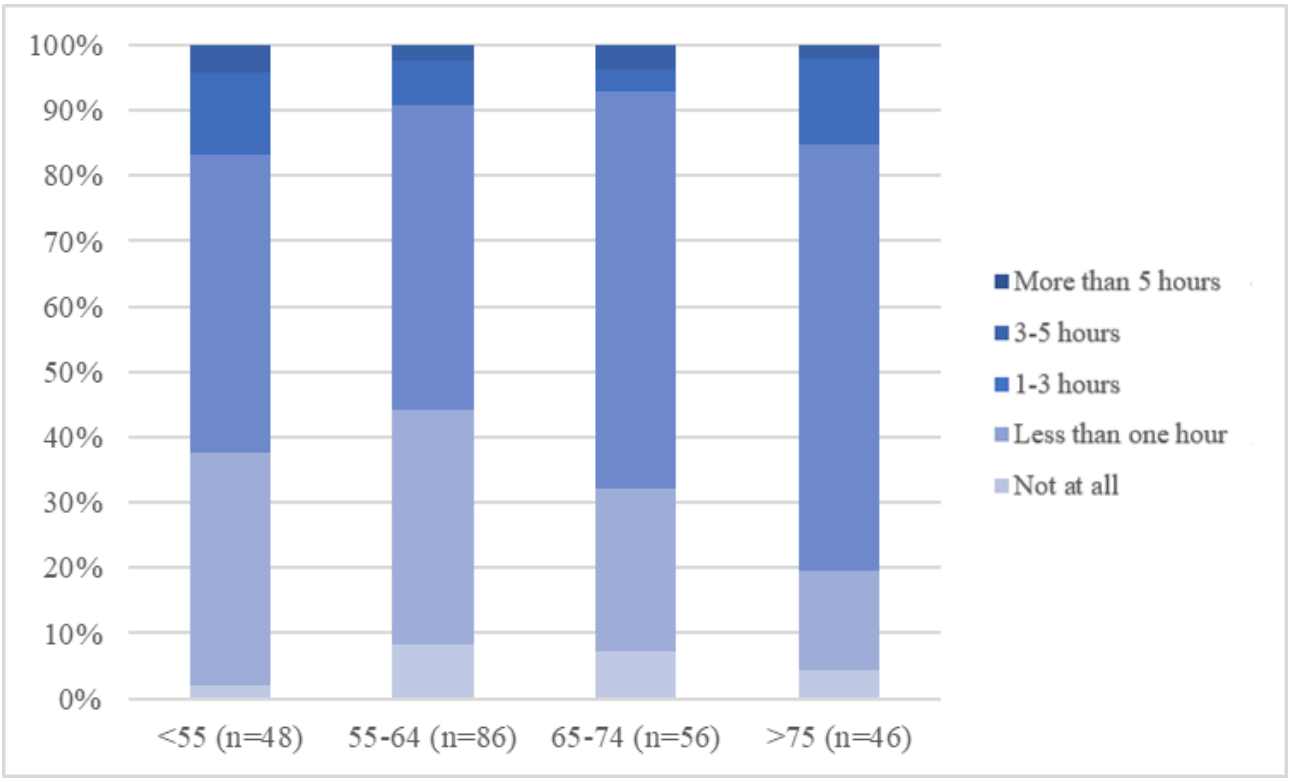
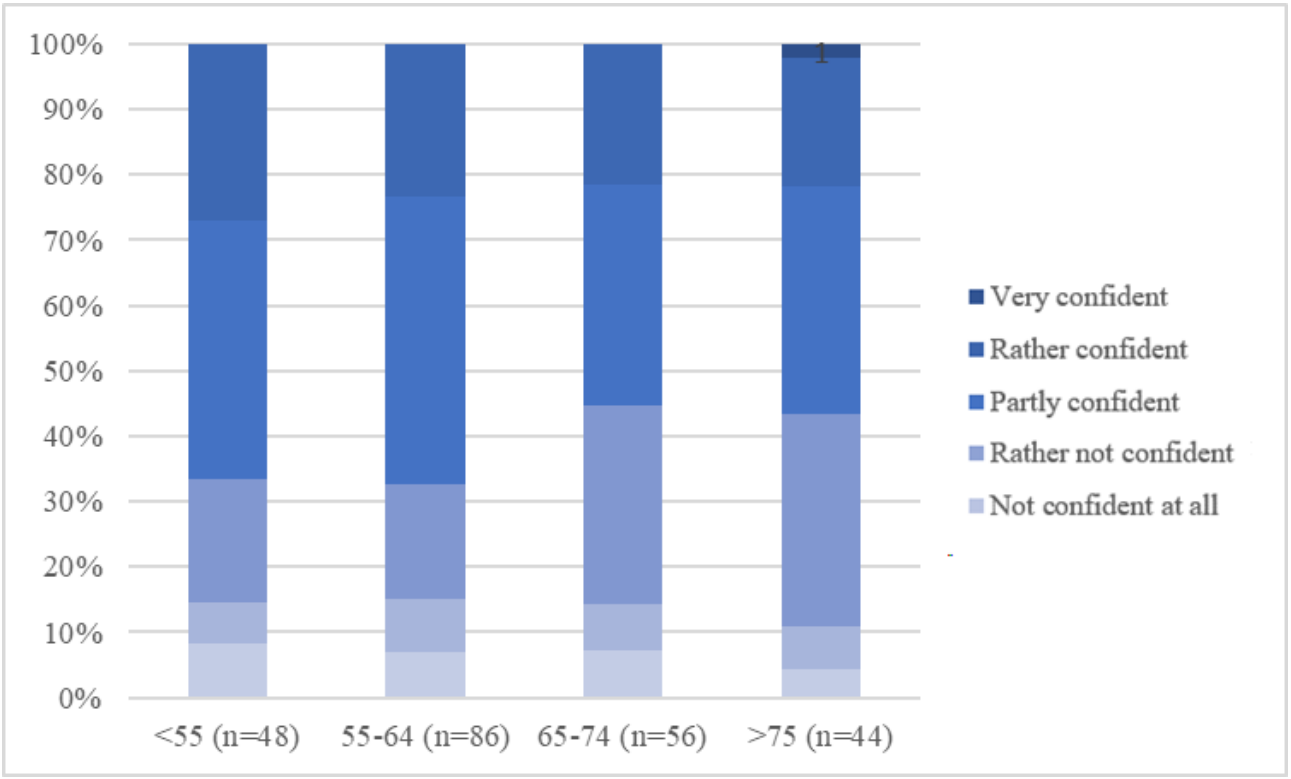


Figure 2. Confidence when handling digital media in relation to age groups (indicated in years along the x-axis).



XSL•FO
RenderX

Figure 3. Confidence when handling digital devices in relation to age groups (indicated in years along the x-axis).

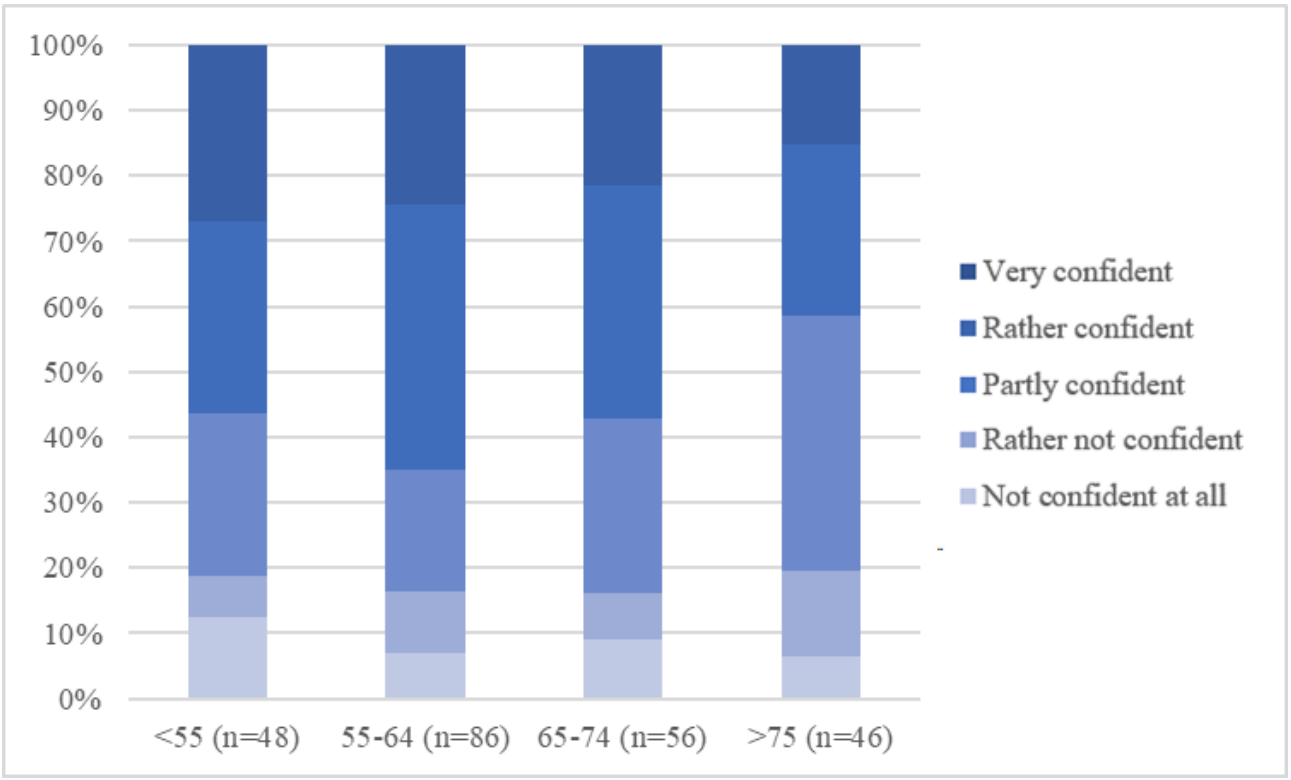
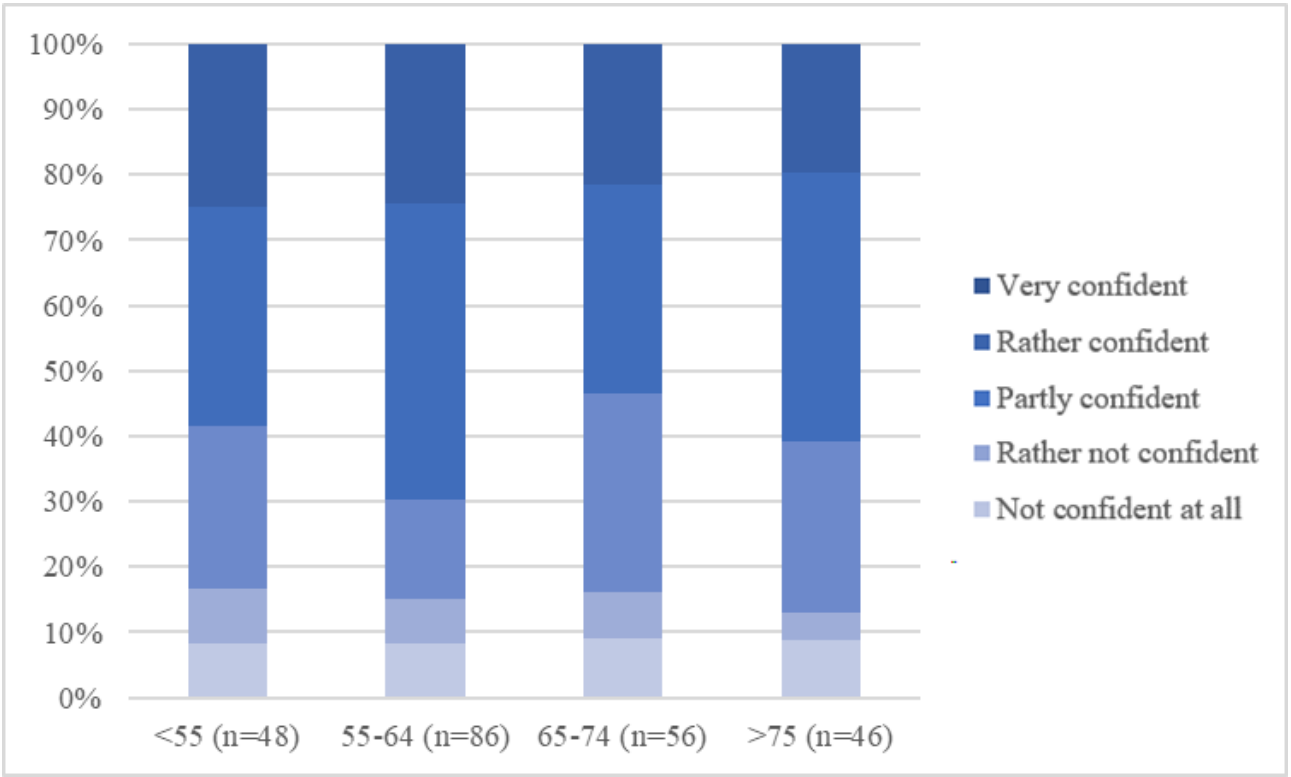


Figure 4. Confidence when handling internet platforms in relation to age groups (indicated in years along the x-axis).

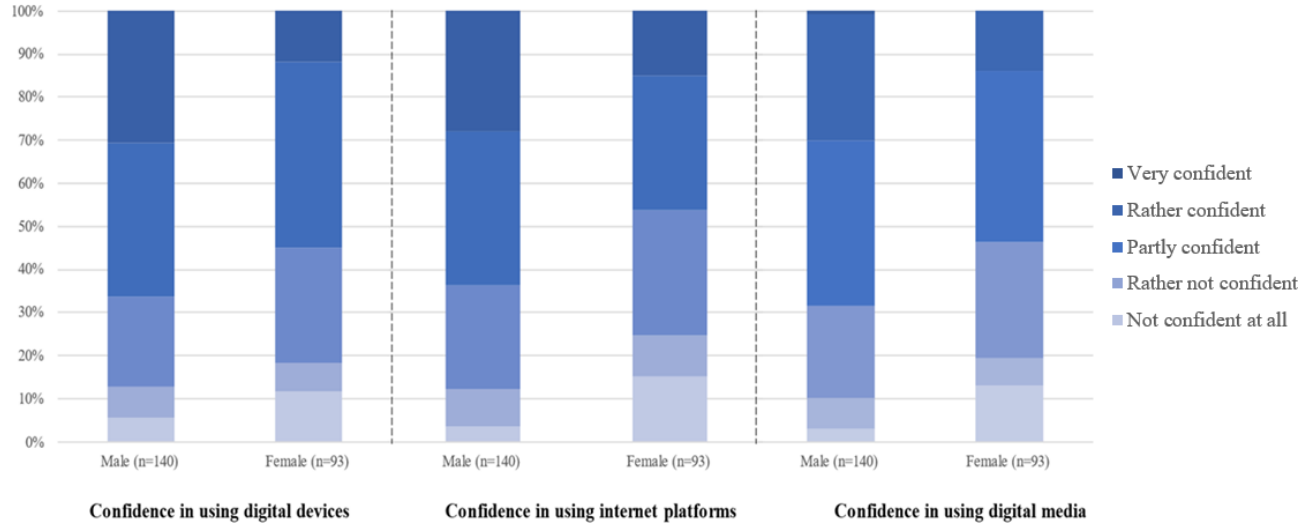


Gender-Specific Differences in the Level of Digital Confidence

As shown in Figure 5, we compared the differences in the levels of digital confidence in male and female participants. It is striking, that in all 3 categories, the levels of digital confidence seem to be at least slightly higher within the male subgroup. This difference becomes the most apparent comparing the levels of confidence when using internet platforms, with only about 10% (n=24) of the male participants feeling “rather not” or “not confident,” while about 25% (n=60) of the female participants feel that way. Moreover, the share of participants who stated to be “very confident” in either of the 3 subcategories is relevantly higher in the male patients.

XSL•FO
RenderX

Figure 5. Gender-specific digital confidence.



Discussion

Internet Use and Digital Knowledge

We conducted an analysis of current internet use rates among individuals affected by cardiovascular disease, examining their attitudes toward the internet and mobile technology, with a specific focus on digital health interventions. Additionally, we assessed participants' confidence in using digital technology and their ability to find credible health information on the internet.

Recent years have witnessed a notable increase in internet and smartphone use, particularly among the older population. From 2009 to 2019, the percentage of German citizens older than 65 years using the internet increased significantly, from approximately 30% to over 67% [17]. This trend was reflected in our trial, where only less than 6% (n=14) of the participants reported not using the internet for private purposes at all. Furthermore, the majority of the study collective (n=156, 64.9%) showed substantial daily internet use, spending at least one hour a day on the internet. Despite seemingly high internet use rates, knowledge about digital health interventions and internet-based health programs appeared to be limited. Only 34.3% (n=82) of participants indicated to understood how these programs worked, and merely 47.5% (n=114) of participants knew where to find them. This knowledge gap might be attributed to the potential lack of accessibility of digital interventions and a scarcity of programs that align with the specific needs of this patient demographic [18,19]. In the future, further research on patients' specific concerns and barriers regarding implementing digital health interventions will be necessary to gain a more in-depth understanding of the current low utilization rates.

Digital Confidence

Confidence levels in using digital media, internet platforms, and digital devices were notably high within this patient population, with over 60% (n=144) of participants expressing confidence in these 3 domains. Participants demonstrated particular assurance in handling digital media and devices. This heightened digital confidence is a crucial factor when contemplating the implementation of digital health interventions in this specific group, as a lack of confidence in digital technology is often cited as a significant barrier to integrating eHealth technologies into patients' chronic disease management [20,21]. Notably, these results deviate from prior research, which has frequently indicated that older patients, in particular, harbored doubts about eHealth technologies and felt insecure about using them [8]. Given the mean age of 61.9 (SD 8.9) years of the study participants, the observed high levels of digital confidence may signify the rapidly increasing rates of smartphone ownership and internet use among the older population [17]. Supporting this hypothesis, even in age-adjusted analyses, participants displayed sustained high levels of digital confidence, with most individuals feeling assured in handling digital media and devices. Only a minority expressed a lack of confidence in using these technologies. However, to ensure the inclusion of these individuals and provide them with an opportunity to benefit from digital health interventions, additional education on the matter will be imperative.

eHealth Literacy

To derive benefits from digital health solutions and effectively integrate them into daily life, possessing eHealth literacy is another crucial prerequisite [22]. Especially, in this study collective of patients with cardiovascular disease with a high prevalence of cardiovascular comorbidities and major risk factors, eHealth literacy is indispensable in order to use the rising possibilities that come with digitalization in chronic disease management.

eHealth literacy is defined as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem" [23]. It also involves the capability to discern between more or less credible sources of information. A common barrier to the adoption of digital health interventions is often the identification of trustworthy sources, leading to concerns about the security and privacy of these interventions [18,24]. In this trial, it was evident that most participants knew how to use the internet to find valuable health information and answers to health-related queries. However, it is noteworthy that only 40% (n=96) of the participants agreed that they knew about the various kinds of sources available for health

XSL•FO
RenderX

information on the internet, with only 1.7% (n=4) fully agreeing. This lack of awareness could be a contributing factor to the low rates of prior engagement with digital health interventions. Previous studies have highlighted that the lack of knowledge about eHealth interventions remains a pertinent barrier, especially among older patients who may not be adequately informed about the diverse options available to receive health support [25]. The trial's results indicated moderate levels of eHealth literacy. Most of the participants (n=154, 64%) felt capable of critically evaluating health information on the internet and distinguishing between questionable and trustworthy web-based resources. Nevertheless, only 47.9% (n=115) of the participants felt secure or reasonably secure in incorporating information from the internet into their health-related decision-making. Previous studies have indicated that even if health information is deemed credible, its implementation often falters due to the digital presentation, which is frequently described as "not user-friendly," "not meaningful," or generally challenging to comprehend [24]. One reason for this may be the sense of depersonalization of health information when presented digitally rather than in a face-to-face interaction, making it harder for individuals to connect with and apply relevant information [20].

Outlook—Challenges and Opportunities of Digital Health Interventions

To address patients' concerns regarding the security and reliability of health information on the internet, Germany took a significant step toward the digitalization of the health care system in 2019 by introducing digital health applications (DiGA) into standard care and supporting their use with statutory health insurance funding [26]. The approval of a mHealth or eHealth technology as a DiGA is strictly regulated and quality assurance for both prescribing physicians and patients using the intervention has to be ensured on a high level [27]. As highlighted in this study, the challenge of finding credible sources for health information on the internet remains a major concern. Having clear criteria to assess the quality and validity of health information that can be found on the internet is crucial to use the full potential of today's digital technology [28]. Confidence in using internet platforms was generally lower in this patient population compared to confidence in using digital devices (Figures 3 and 4). Therefore, officially approved digital health interventions with proven health care benefits that can be prescribed by treating physicians present an opportunity to make them available to more patients and ensure the high quality and credibility of the provided health information. Although this approach appears promising, the level of acceptance by both patients and physicians still needs to be thoroughly evaluated [29].

Limitations

It should be noted that the eHealth literacy score was self-assessed, and this study did not try to match the perceived eHealth literacy score to the actual abilities to use the internet to find health information and implement it in daily life. It has been shown that self-assessed eHealth literacy scores and actual eHealth literacy can differ when put into practice [30,31]. Digital use patterns and internet use were also assessed using self-reported data, which makes them subject to potential bias. For future research, it could be beneficial to include tracking individuals' daily use patterns in order to avoid potential bias and inaccuracies.

Moreover, since the study population was actively interested in participating in a trial revolving around mHealth technologies, a possible sampling bias should be considered when transferring the results to the general population. In addition to this, this trial particularly included patients diagnosed with coronary artery disease, which leads to a rather homogenous collective. This may limit the generalizability of these findings to a collective of patients with different cardiovascular diseases, such as heart failure, which was only present in one-fourth of the participants of this trial.

Conclusions

This study revealed that individuals with atherosclerotic cardiovascular disease were generally well-acquainted with using the internet and handling digital devices in all age groups. The internet was an integral part of their daily lives, leading to a high level of confidence in digital technologies. However, the adoption of digital health interventions remained relatively low, possibly reflecting reservations about incorporating internet-derived information into daily routines. This hesitancy might stem from uncertainties in identifying credible and trustworthy health sources on the internet, as well as the perceived lack of personalization in digital health interventions.

Despite their older age, the results suggest that patients with cardiovascular disease can be a suitable target group for digital health interventions, given their high internet use rates and digital confidence. The patient population appears representative in terms of cardiovascular comorbidities and the prevalence of cardiovascular risk factors. To address persistent skepticism toward digital technologies, it becomes crucial to incorporate personalized health information and adopt a patient-centered approach. Therefore, we need to delve into the specific needs and requirements of our patient collective regarding digital health interventions. This approach is essential to enhance the acceptance and effectiveness of digital health interventions within this patient population.

Acknowledgments

This work was financially supported by the Welcome-Back Grant of the University Duisburg-Essen Medical Faculty to Julia Lortz (grant D/401-00004). All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

XSL•FO
RenderX

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

AB conceived and designed the experiments, and analyzed and interpreted the data. AAM analyzed and interpreted the data. CR contributed reagents, materials, analysis tools, and data. DM performed the experiments. GU conceived and designed the experiments and performed the experiments. HV performed the experiments. JL conceived and designed the experiments, analyzed and interpreted the data, and contributed reagents, materials, analysis tools, or data. KP contributed reagents, materials, analysis tools, and data. LMJ analyzed and interpreted the data. TR conceived and designed the experiments and analyzed and interpreted the data. All authors wrote the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File, 12 KB - formative_v9i1e63057_app1.docx]

References

1. The top 10 causes of death. WHO. 2020. URL: https://www.who.int/news-room/facts-sheets/details/the-top-10-causes-of-death [accessed 2024-02-16]
2. Hare DL, Toukhsati SR, Johansson P, Jaarsma T. Depression and cardiovascular disease: a clinical review. Eur Heart J 2014 Jun 1;35(21):1365-1372. [doi: 10.1093/eurheartj/eht462] [Medline: 24282187]
3. Visseren FLJ, Mach F, Smulders YM, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice. Eur Heart J 2021 Sep 7;42(34):3227-3337. [doi: 10.1093/eurheartj/ehab484] [Medline: 34458905]
4. Chowdhury R, Khan H, Heydon E, et al. Adherence to cardiovascular therapy: a meta-analysis of prevalence and clinical consequences. Eur Heart J 2013 Oct;34(38):2940-2948. [doi: 10.1093/eurheartj/eht295] [Medline: 23907142]
5. Dall TM, Gallo PD, Chakrabarti R, West T, Semilla AP, Storm MV. An aging population and growing disease burden will require a large and specialized health care workforce by 2025. Health Aff (Millwood) 2013 Nov;32(11):2013-2020. [doi: 10.1377/hlthaff.2013.0714] [Medline: 24191094]
6. Gandapur Y, Kianoush S, Kelli HM, et al. The role of mHealth for improving medication adherence in patients with cardiovascular disease: a systematic review. Eur Heart J Qual Care Clin Outcomes 2016 Oct 1;2(4):237-244. [doi: 10.1093/ehjqcco/qcw018] [Medline: 29474713]
7. Schorr EN, Gepner AD, Dolansky MA, et al. Harnessing mobile health technology for secondary cardiovascular disease prevention in older adults: a scientific statement from the American Heart Association. Circ Cardiovasc Qual Outcomes 2021 May;14(5):e000103. [doi: 10.1161/HCQ.0000000000000103] [Medline: 33793309]
8. Nymberg VM, Bolmsjö BB, Wolff M, Calling S, Gerward S, Sandberg M. "Having to learn this so late in our lives…" Swedish elderly patients' beliefs, experiences, attitudes and expectations of e-health in primary health care. Scand J Prim Health Care 2019 Mar;37(1):41-52. [doi: 10.1080/02813432.2019.1570612] [Medline: 30732519]
9. Ullrich G, Dönmez A, Mahabadi AA, et al. Effect of visual presentation of atherosclerotic carotid plaque on adherence to secondary preventive therapy using mHealth technologies (PreventiPlaque app): study protocol for a randomized controlled trial. Heliyon 2023 Mar;9(3):e14052. [doi: 10.1016/j.heliyon.2023.e14052] [Medline: 36915490]
10. Mach F, Baigent C, Catapano AL, et al. 2019 ESC/EAS guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk: The Task Force for the management of dyslipidaemias of the European Society of Cardiology (ESC) and European Atherosclerosis Society (EAS). Eur Heart J 2019;41(1):111-188. [doi: 10.1093/eurheartj/ehz826]
11. Marsall M, Engelmann G, Skoda EM, Teufel M, Bäuerle A. Measuring electronic health literacy: development, validation, and test of measurement invariance of a revised German version of the eHealth literacy scale. J Med Internet Res 2022 Feb 2;24(2):e28252. [doi: 10.2196/28252] [Medline: 35107437]
12. Bäuerle A, Marsall M, Jahre LM, et al. Psychometric properties of the German revised version of the eHealth literacy scale in individuals with cardiac diseases: validation and test of measurement invariance. Digital Health 2023;9:20552076231194915. [doi: 10.1177/20552076231194915] [Medline: 37588160]
13. Bäuerle A, Mallien C, Rassaf T, et al. Determining the acceptance of digital cardiac rehabilitation and its influencing factors among patients affected by cardiac diseases. J Cardiovasc Dev Dis 2023 Apr 17;10(4):174. [doi: 10.3390/jcdd10040174] [Medline: 37103053]
14. Nurtsch A, Teufel M, Jahre LM, et al. Drivers and barriers of patients' acceptance of video consultation in cancer care. Digital Health 2024;10:20552076231222108. [doi: 10.1177/20552076231222108] [Medline: 38188860]

XSL•FO
RenderX

15. Schröder J, Bäuerle A, Jahre LM, et al. Acceptance, drivers, and barriers to use eHealth interventions in patients with post-COVID-19 syndrome for management of post-COVID-19 symptoms: a cross-sectional study. Ther Adv Neurol Disord 2023;16:17562864231175730. [doi: 10.1177/17562864231175730] [Medline: 37255668]
16. Rentrop V, Damerau M, Schweda A, et al. Predicting acceptance of e-Mental health interventions in patients with obesity by using an extended unified theory of acceptance model: cross-sectional study. JMIR Form Res 2022 Mar 17;6(3):e31229. [doi: 10.2196/31229] [Medline: 35297769]
17. Ziebach M. Datenreport 2021—internetnutzung [Report in German]. 2021 URL: https://www.bpb.de/kurz-knapp/zahlen-und-fakten/sozialbericht-2024/553201/internetnutzung/ [accessed 2024-01-29]
18. Frederix I, Caiani EG, Dendale P, et al. ESC e-Cardiology Working Group Position Paper: overcoming challenges in digital health implementation in cardiovascular medicine. Eur J Prev Cardiol 2019 Jul;26(11):1166-1177. [doi: 10.1177/2047487319832394] [Medline: 30917695]
19. Torrent-Sellens J, Díaz-Chao Á, Soler-Ramos I, Saigí-Rubió F. Modelling and predicting eHealth usage in Europe: a multidimensional approach from an online survey of 13,000 European Union internet users. J Med Internet Res 2016 Jul 22;18(7):e188. [doi: 10.2196/jmir.5605] [Medline: 27450189]
20. Lie SS, Karlsen B, Oord ER, Graue M, Oftedal B. Dropout from an eHealth intervention for adults with type 2 diabetes: a qualitative study. J Med Internet Res 2017 May 30;19(5):e187. [doi: 10.2196/jmir.7479] [Medline: 28559223]
21. Herrera S, Salazar A, Nazar G. Barriers and supports in eHealth implementation among people with chronic cardiovascular ailments: integrative review. Int J Environ Res Public Health 2022 Jul 7;19(14):14. [doi: 10.3390/ijerph19148296] [Medline: 35886149]
22. Gee PM, Greenwood DA, Paterniti DA, Ward D, Miller LMS. The eHealth enhanced chronic care model: a theory derivation approach. J Med Internet Res 2015 Apr 1;17(4):e86. [doi: 10.2196/jmir.4067] [Medline: 25842005]
23. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9. [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
24. Ware P, Bartlett SJ, Paré G, et al. Using eHealth technologies: interests, preferences, and concerns of older adults. Interact J Med Res 2017 Mar 23;6(1):e3. [doi: 10.2196/ijmr.4447] [Medline: 28336506]
25. Pywell J, Vijaykumar S, Dodd A, Coventry L. Barriers to older adults' uptake of mobile-based mental health interventions. Digital Health 2020;6:2055207620905422. [doi: 10.1177/2055207620905422] [Medline: 32110429]
26. Ludewig G, Klose C, Hunze L, Matenaar S. Digital health applications: statutory introduction of patient-centred digital innovations into healthcare [Article in German]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021 Oct;64(10):1198-1206. [doi: 10.1007/s00103-021-03407-9] [Medline: 34529096]
27. Sauermann S, Herzberg J, Burkert S, Habetha S. DiGA—a chance for the German healthcare system. J Eur CME 2022;11(1):2014047. [doi: 10.1080/21614083.2021.2014047] [Medline: 34992948]
28. Duncker D, Billig H, Frey N, et al. Professioneller Umgang mit sozialen Medien in der Kardiologie [Article in German]. Kardiologie 2024 Mar;18(2):176-186. [doi: 10.1007/s12181-024-00665-0]
29. Uncovska M, Freitag B, Meister S, Fehring L. Patient acceptance of prescribed and fully reimbursed mHealth apps in Germany: an UTAUT2-based online survey study. J Med Syst 2023 Jan 27;47(1):14. [doi: 10.1007/s10916-023-01910-x] [Medline: 36705853]
30. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 9;13(4):e86. [doi: 10.2196/jmir.1840] [Medline: 22071338]
31. Richtering SS, Morris R, Soh SE, et al. Examination of an eHealth literacy scale and a health literacy scale in a population with moderate to high cardiovascular risk: Rasch analyses. PLoS ONE 2017;12(4):e0175372. [doi: 10.1371/journal.pone.0175372] [Medline: 28448497]

Abbreviations

mHealth: mobile health

Edited by A Mavragani; submitted 08.06.24; peer-reviewed by J Kaswija, U Sinha; revised version received 21.10.24; accepted 07.11.24; published 19.03.25.

Please cite as:
Ullrich G, Bäuerle A, Vogt H, Mahabadi AA, Paldán K, Messiha D, Jahre LM, Rammos C, Rassaf T, Lortz J
Digital Health Literacy and Attitudes Toward eHealth Technologies Among Patients With Cardiovascular Disease and Their Implications for Secondary Prevention: Survey Study
JMIR Form Res 2025;9:e63057
URL: https://formative.jmir.org/2025/1/e63057
doi:10.2196/63057

XSL•FO
RenderX

© Greta Ullrich, Alexander Bäuerle, Hannah Vogt, Amir Abbas Mahabadi, Katrin Paldán, Daniel Messiha, Lisa Maria Jahre, Christos Rammos, Tienush Rassaf, Julia Lortz. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Interprofessional Discussion for Knowledge Transfer in a Digital "Community of Practice" for Managing Pneumoconiosis: Mixed Methods Study

Varinn Avi Sood[1]; Heidi Rishel Brakey[2], MA; Orrin Myers[2], PhD; Xin Shore[2], PhD; Akshay Sood[2], MPH, MD

[1]Central New Mexico Community College, La Cueva High School, Albuquerque, NM, United States

[2]University of New Mexico Health Sciences Center, 1 University of New Mexico MSC 10 5550, Albuquerque, NM, United States

Corresponding Author:
Akshay Sood, MPH, MD
University of New Mexico Health Sciences Center, 1 University of New Mexico MSC 10 5550, Albuquerque, NM, United States

Abstract

Background: Pneumoconiosis prevalence is increasing in the United States, especially among coal miners. Contemporaneously with an increased need for specialized multidisciplinary care for miners, there is a shortage of experts to fulfill this need. Miners' Wellness ECHO (Extension for Community Health Outcomes) is a digital community of practice based on interprofessional discussion for knowledge transfer. The program has been demonstrated to increase participants' self-efficacy for clinical, medicolegal, and "soft" skills related to miners' health.

Objective: We aimed to examine characteristics associated with interprofessional discussions and suggest ways to strengthen knowledge transfer.

Methods: This mixed methods study used an exploratory sequential design. We video-recorded and transcribed ECHO sessions over 14 months from July 2018 to September 2019 and analyzed content to examine participant discussions. We focused on participants' statements of expertise followed by other participants' acceptance or eschewal of these statements (utterances). We conducted quantitative analyses to examine the associations of active participation in discussion (primary outcome variable, defined as any utterance). We analyzed the association of the outcome on the following predictors: (1) participant group status, (2) study time frame, (3) participant ECHO experience status, (4) concordance of participant group identity between presenter and participant, (5) video usage, and (6) attendance frequency. We used the generalized estimating equations approach for longitudinal data, logit link function for binary outcomes, and *LSMEANS* to examine least squares means of fixed effects.

Results: We studied 23 sessions with 158 unique participants and 539 total participants, averaging 23.4 (SD 5.6) participants per session. Clinical providers, the largest participant group, constituting 36.7% (n=58) of unique participants, were the most vocal group (mean 21.74, SD 2.11 average utterances per person-session). Benefits counselors were the least vocal group, with an average utterance rate of 0.57 (SD 0.29) per person-session and constituting 8.2% (n=13) of unique participants. Thus, various participant groups exhibited different utterance rates across sessions (P=.003). Experienced participants may have dominated active participation in discussion compared to those with less or intermediate experience, but this difference was not statistically significant (P=.11). When the didactic presenter and participant were from the same participant group, active participation by the silent group participants was greater than when both were from different groups. This association was not seen in vocal group participants (interaction P=.003). Compared to those participating by audio, those participating on video tended to have higher rates of active participation, but this difference was not statistically significant (P=.11).

Conclusions: Our findings provide insight into the mechanics of interprofessional discussion in a digital community of practice managing pneumoconiosis. Our results underscore the capacity of the novel ECHO model to leverage technology and workforce diversity to facilitate interprofessional discussions on the multidisciplinary care of miners. Future research will evaluate whether this translates into improved patient outcomes.

(JMIR Form Res 2025;9:e67999) doi:10.2196/67999

KEYWORDS

digital community of practice; knowledge transfer; pneumoconiosis; telementoring; rural health care; rural professionals; multidisciplinary management; interprofessional discussion; miner health; health equity; mixed methods; digital health; public health; digital community; self-efficacy; quantitative analyses; quantitative; technology; multidisciplinary care; patient outcome

XSL•FO
RenderX

Introduction

Mining dust-related diseases (eg, pneumoconiosis) are increasing in rural United States, especially among coal miners [1-4]. The 2017 prevalence of radiographic pneumoconiosis for coal miners with over 25 years of underground mining experience was greater than 10%, double the prevalence from the late 1990s [1]. Similarly, the 2014 rate of complicated pneumoconiosis (a particularly deadly form) among long-tenured underground coal miners was 1.1%, compared to 0.3% at its lowest point in the late 1990s [5].

Miners are medically vulnerable, underserved, and often underinsured. Many miners live in rural, remote, and mountainous locations in the Mountain West and Appalachia, constituting the "hot spot" regions of pneumoconiosis prevalence and mortality [6]. The prevalence of radiographic and complicated pneumoconiosis is the highest among all US miners in rural central Appalachia [7,8].

The re-emergence of pneumoconiosis presents unique challenges for rural communities. Compared with urban residents, those in rural areas have less access to outpatient pulmonary rehabilitation [9], primary care physicians [10], and pulmonologists [11]. Rural practitioners also face unique challenges, including professional isolation and complex patient profiles [12], and describe multiple barriers to knowledge acquisition, such as resources, personal costs, physical distance, and time [13].

Contemporaneously with an increase in the need for specialized multidisciplinary care for US miners, there is a shortage of experts to fulfill this need [14]. This critical gap is being addressed in New Mexico by innovative technology-based interventions such as mobile clinics equipped for telemedicine [15] and digital communities of practice for knowledge transfer. These interventions are recognized as rural and training innovations by the Rural Health Information Hub and the American Thoracic Society, respectively. The Miners' Wellness ECHO (Extension for Community Health Outcomes) Program is a digital community of practice, defined as a group of people who "share a concern or a passion for something they do and learn how to do better as they interact regularly" [16].

Project ECHO began in 2003 to connect primary care providers in rural New Mexico at "spoke" sites to University-based experts at a "hub" site so the former could effectively treat their patients with hepatitis C locally, instead of referring them to the University of New Mexico [17]. Since then, the ECHO movement has grown exponentially, spanned many health topics, and reached people around the globe. The Miners' Wellness ECHO Program provides structured longitudinal interprofessional telementoring to professionals caring for US miners in pneumoconiosis mortality hot spots, including respiratory therapists, home health professionals, benefits counselors, attorneys, clinical providers, and others.

Previous research of the Miners' Wellness ECHO established that participants in the pneumoconiosis mortality hot spots in the United States valued telementoring sessions for delivering relevant, evidence-based, balanced, and objective content [18]. Participants also showed a significant increase in self-efficacy scores overall and for domains related to clinical, medicolegal, and "soft" skills pertaining to miners' health, with "soft" skills including interpersonal and communication skills needed to navigate highly collaborative work in the care of miners [14]. The fact that existing participants showed greater improvement than fresh (newer) participants indicated that there might be a dose effect of telementoring on change in self-efficacy. In addition, greater improvement in outcomes among participants belonging to the clinical versus nonclinical professional groups might reflect the disproportionate emphasis on clinical didactics and case discussions in the program. Participants rated the collective efficacy of the virtual community of practice highly, stating feelings of being "closely knit," trustworthiness, and willingness to help each other [19].

The foundation of knowledge transfer in the program is the "all teach, all learn" model, which implies that every participant has unique knowledge to contribute and share during interprofessional discussions. Our previous research showed that most Miners' Wellness ECHO participants used knowledge ties from outside their participant group, emphasizing the interprofessional nature of knowledge transfer in miners' care [20]. The knowledge transfer was more efficient for participants with greater ties to rural-based miners than those with lesser ties [20]. Given the compensation challenges faced by US miners, benefits counselors and attorneys played an outsized role in knowledge transfer [18]. The findings from the preliminary evaluation establish the usefulness of this unique program and provide the study rationale for further examination of its learning processes, communication, and participation, an area of limited scholarship [21].

The objective of our study was to examine characteristics associated with interprofessional discussions and suggest ways to strengthen knowledge transfer. The long-term goal of this study is to reduce health inequity through greater investment in interprofessional telementoring efforts that promote collaborative health care in medically underserved mining communities by fusing technology with specialized multidisciplinary expertise. This approach may help rural communities counter the re-emergence of the pneumoconiosis epidemic by ameliorating their shortage of skilled expertise in mining-related diseases.

Methods

ECHO Model

The ECHO approach differs from traditional telemedicine, where clinical providers (eg, physicians, physician assistants, and advanced practice providers) assume short-term direct care of individual patients. Further, unlike webinars or conventional didactic lectures, the ECHO model provides a digital discussion of cases with expert panels in real time that is highly contextualized and thereby fulfills key learning theories, such as deliberate practice [22], social cognitive theory [23], and situated learning and communities of practice [24]. As detailed in a previous publication [18], the ECHO model is based on the following key principles: (1) use of internet-based technology for multipoint videoconferencing, which helps leverage scarce

XSL•FO
RenderX

resources; (2) use of an established evidence-based disease-management model that has been demonstrated to improve outcomes by reducing variation in processes of care and sharing best practices [17,25-27]; (3) use of case-based learning, based on discussion with experts and peers; (4) use of a digital community of practice, which emphasizes reciprocity in the sharing of skills and information; and (5) use of an internet-based database (ie, iECHO software [Hangzhou IECHO Science & Technology Co, Ltd]) to monitor outcomes.

Miners' Wellness ECHO Structure

Professionals involved in miners' health are invited to attend the program using an emailed web link connected to a REDCap (Research Electronic Data Capture; Vanderbilt University) survey [28]. As published previously [18], the ECHO sessions are scheduled at the same time twice a month for 75 minutes, adhering to a standard format. Following a facilitator's initial 10 minutes of introduction and announcements, an invited expert delivers a 15-minute didactic followed by a 20-minute facilitated question-answer session and a 30-minute digital facilitated case discussion. The medical director of the Miners' Wellness ECHO program is trained to facilitate group discussions during a 3-day immersion program and is periodically retrained [29]. The ECHO format, employing adult learning principles, focuses more on active learning through discussion than on didactic training. Upon completing a session-specific survey, participants can receive continuing education credits without charge. A multidisciplinary committee of experts follows a structured curriculum that is continually adapted based on the needs of the virtual community of practice, which are identified through a review of the continuing evaluation reports.

Attendance at ECHO sessions is open and voluntary. Through session attendance, participants can participate in the didactic and case discussions, provide information and insight from their own experience, and receive mentoring from their peers and an expert panel. Outside the program sessions, participants continue to have access to peers and experts for urgent consultation requests via telephone or email. Over time, with iterative practice and feedback, participants gain additional expertise, and become more confident in their skill sets related to caring for miners in their respective fields [14,18,20]. Recorded didactic sessions are made available through a web-based archive. The didactic and case discussion topics were not repeated during this study's time frame. All cases presented during this study's time frame were resolved using the collective expertise of the multidisciplinary community of practice, which included experts. The University of New Mexico has a sustainable business model for its ECHO Programs, with the cost of personnel shared between university budgets and grants and contracts.

Study Design and Recruitment

This 14-month exploratory sequential mixed methods study involved participants in the Miners' Wellness ECHO Program from July 2018 to September 2019. Participants are invited from across the nation by email, flyers, and word of mouth referrals to join the ECHO sessions. Their session participation is voluntary. At the start of the session, participants are informed that the session is video recorded. Together, the University of New Mexico School of Medicine and its rural partner Miners' Colfax Medical Center constitute the hub site of experts. The spoke partner sites are located across the pneumoconiosis mortality hot spot regions of the United States [18]. Although participants were allowed to attend multiple sessions, our analyses emphasize cross-sectional comparisons based on participant characteristics.

We asked the following 5 research questions on common ECHO processes based on the experience of this study's team, which are also related to the engagement theory model that focuses on participation, engagement, and learning within communities of practice [30]:

1. Do some participant groups make more utterances than others (eg, do clinical providers make more utterances than benefit counselors)?
2. Do participants who are more experienced with ECHO have greater rates of active participation in discussion than those with lower levels of experience?
3. Does the participant group identity of the didactic presenter predict active participation in discussion from the same participant group during that session (eg, if a didactic presenter is an attorney, are other attorneys more likely to participate in discussion in that session)?
4. Does video usage predict active participation in discussion, as compared to audio usage?
5. Does session attendance frequency predict active participation in discussion?

Video Recording and Transcription

Each ECHO session during this study time frame was video recorded and professionally transcribed. Participants were informed of the recording at the start of each session.

Summative Qualitative Content Analysis

Using an exploratory approach from a naturalistic point of view, we conducted summative qualitative content analysis to code interprofessional discussions among ECHO participants in each video. Specifically, we examined participants' verbal utterances, focusing on statements of expertise followed by acceptance or eschewal of the statements by other participants, using latent analysis for interpretation of content. We did not use a guiding theory, as this was exploratory. See Table 1 for the qualitative codebook with definitions of each. See Multimedia Appendix 1 for examples of conversations demonstrating statements of expertise and responses of acceptance and eschewal.

XSL•FO
RenderX

Table . Qualitative and quantitative content analysis codebook, from the mixed methods study of Miners' Wellness ECHO[a] participants, July 2018 to September 2019.

Variable, variable label, and variable code value	Definition or example of code value	Level of measurement
Qualitative codebook		
Utterance type[b]		
What type of utterance did the participant make?		
1=Statement of expertise	Defining solutions to problems or answering questions. Anyone who speaks as an expert is an expert.	Nominal
2=Acceptance of a statement of expertise	Others' acceptance of a statement of expertise. Includes adding to statement (expanding is acceptance) or verbalizations such as "That's great!" "Thank you so much for that." "That's really helpful!"	Nominal
3=Eschewal of a statement of expertise	Others' eschewal in response to a statement of expertise. Eschewal can include providing new "expertise" that contradicts statements in disagreement with previously stated opinions, information, evidence, or statements. Can include partial disagreement or eschewal.	Nominal
4=Neutral or no response to a statement of expertise	Neutral or no response. For example, if a speaker moves to next question or topic.	Nominal
Quantitative codebook		
Number of utterances		
How many utterances did the participant make during the session?	Total number of expertise, acceptance, eschewal, or neutral statements during a session.	Ordinal
Participation		
Did this person participate in a session?		
0=No	At least 1 utterance during a session	Nominal
1=Yes	At least 1 utterance during a session	Nominal
Attendance		
Did this person attend a session?		
0=No	Logged into a session	Nominal
1=Yes	Logged into a session	Nominal
Session		

XSL•FO
RenderX

Variable, variable label, and variable code value		Definition or example of code value	Level of measurement
	What is the session number?	23 sessions from July 2018 to September 2019	Ordinal
Video			
	Did the participant use their video camera during the session?		
	1=Audio only	Use of video camera during a session	Nominal
	2=Video	Use of video camera during a session	Nominal
	3=Unknown	Use of video camera during a session	Nominal
Presenter			
	Was the participant a didactic presenter during the session?		
	0=No	—[c]	Nominal
	1=Yes	—	Nominal
Participant group			
	What participant group was the person a part of?		
	1=Attorney	Self-defined participant group status	Nominal
	2=Clinical provider	Self-defined participant group status	Nominal
	3=Home health professional	Self-defined participant group status	Nominal
	4=Respiratory therapist	Self-defined participant group status	Nominal
	5=Benefits counselor	Self-defined participant group status	Nominal
	6=Other	Self-defined participant group status	Nominal
Participant ECHO experience status			
	What joining group was the person a part of?		
	1=Experienced	Participant joined before the start of this study, May 9, 2018.	Nominal
	2=Intermediate	Participant joined at the start of this study, May 9, 2018.	Nominal
	3=Less experienced	Joined after September 12, 2018.	Nominal

[a]ECHO: Extension for Community Health Outcomes.

[b]Utterances include statements of expertise, acceptance, and eschewal. Additional details are provided in Tables S1 and S2 in Multimedia Appendix 1.

[c]Not applicable.

We realize that evaluating others' acceptance or eschewal can be subjective and difficult to determine by reading transcripts. Additionally, body language and tone of voice can convey meaning that written words cannot. To minimize these risks, our qualitative team consisted of 4 people with varied and complementary areas of expertise: (1) the principal investigator (AS) who is an expert pulmonologist and medical director/session facilitator for the Miners' Wellness Tele-ECHO Clinic; (2) an expert qualitative methodologist (HRB); and (3) two experts of ECHO methods and processes (director of program operations, Rachelle Rochelle, MPA, and senior program manager, Stephen Murillo, MBA). This team of 4 analysts met frequently during the project to interpret session content and discuss coding, potential biases, and reflexivity. We concurrently coded video recordings of participant behavior alongside transcripts using NVivo (version 12; QSR

XSL•FO
RenderX

International). The senior analyst created the preliminary codebook and independently coded the presence of each code across all ECHO sessions. She met with the principal investigator and 2 other team members to review the trial coding of the first 2 sessions as a group, modifying coding and definitions as needed. The principal investigator independently reviewed 2 additional sessions, and the other analysts independently reviewed 5 sessions; therefore, 7 sessions were trial-coded by 3 (5 sessions) or 4 (2 sessions) team members. The team regularly met to discuss coding until they were 100% in agreement, at which point the senior analyst coded the remaining 16 sessions. She continued to meet with the team periodically to discuss her coding generally and anything that was coded to be "unclear" (as defined in Table 1), at which point they jointly decided on whether it was better suited for another code. We removed all statements of expertise, acceptance, and eschewal of the session facilitator (AS), so we only evaluated participants' utterances. This approach minimized the bias resulting from the effect of the role and relationships of the session facilitator (AS) in the data analysis. There were no unintended consequences from the program participation or analysis. Since thematic saturation was not the goal of this study, the sampling was guided by the number of sessions rather than sampling saturation.

Quantitative Data Analysis

We worked with our senior statistician (OM) to convert our qualitative coding from NVivo into quantitative coding in Excel (Microsoft Corp). We coded additional variables to be analyzed with SAS (version 9.4; SAS Institute Inc). See Table 1 for the quantitative codebook.

Outcome and Predictor Variables

We examined the primary outcome variable, "active participation in discussion," defined by making any utterances (which included statements of expertise, acceptance, and eschewals). We analyzed the association of the outcome on the following predictor variables: (1) participant group status, (2) study time frame, (3) participant ECHO experience status, (4) concordance of participant group identity between presenter and participant, (5) video usage, and (6) attendance frequency.

Participant groups included attorneys, benefits counselors, clinical providers, home health professionals, respiratory therapists, and others. The participants' ECHO experience status was determined by the participant's date of joining the ECHO Program. Participants' ECHO experience status was defined by (1) the experienced group: those who joined the Miners' Wellness ECHO before the start of this study, May 9, 2018; (2) the intermediate group: those who joined at the start of this study, May 9, 2018; and (3) the less experienced group: those who joined after September 12, 2018. The participant groups were also classified as "silent" and "vocal" based on their preliminary analysis of their rates of utterances during the ECHO sessions. Participant group identity was considered concordant when the didactic presenter and participant belonged to the same participant group. Video status was determined if the video camera was turned on for any duration during a session by the participant.

Analytic Strategy

The number of utterances by type (statements of expertise, acceptance, and eschewal) and overall for each participant and ECHO session were collected. An average number of utterances by participant group, session, and type were computed. Binary variables for any active participation in discussion (1=yes for ≥1 utterance in the session, 0=no for 0 utterances in the session) were made for each utterance type and overall. Frequencies and percentages for active participation variables and other categorical variables were compiled by utterance type, participant group, and session.

We used the generalized estimating equations approach to extend the generalized linear model to handle longitudinal data, including predictor variables. The logit link function was applied for binary outcomes. Repeated statements with the subject option identified each session as a cluster. *LSMEANS* were used to compute and compare least squares means of fixed effects. A 2-tailed *P* value less than .05 was considered statistically significant.

Ethical Considerations

This study was an arm of a larger study. This arm of this study included participants who consented and enrolled in the larger study and those who were not enrolled in the larger study. This arm was approved as exempt by the institutional review board, University of New Mexico Human Research Protections Office (HRRC#18 - 386). Informed consent was waived by the institutional review board for this arm, separate from the larger study. The study data were deidentified for analysis. No compensation was offered for this arm, separate from the larger study.

Results

Descriptive Characteristics

We studied 23 ECHO sessions during this study's time frame, with 158 unique participants and 539 total participants. Further, 78 (49.4%) participants attended a single session, 28 (17.7%) attended 2 sessions, 11 (7%) attended 3 sessions, 13 (8.2%) attended 4 - 5 sessions, 13 (8.2%) attended 6 - 10 sessions, and 15 (9.5%) attended 11 - 20 sessions (overall mean 3.4, SD 4.1 sessions). Individual session attendance averaged 23.4 (SD 5.6; median 24, IQR 8; range=13 - 36) participants per session. During these sessions, 23 participants presented 23 patient cases, and 23 invited experts presented 23 didactics.

Do Some Participant Groups Make More Utterances Than Others?

Clinical providers, the largest participant group constituting 36.7% (58/158) of unique participants, were also the most vocal group (mean 21.74, SD 2.11, average utterances per person-session; Table 2). Attorneys, while the smallest group in number (11/158, 6.9%), were the second most vocal group (mean 6.73, SD 4.29, average utterances per person-session). Benefits counselors and home health professionals were the least vocal groups, with average utterance rates of 0.57 (SD 0.29) and 2.86 (SD 2.24) per person-session, respectively, constituting 8.2% (13/158) and 22.8% (36/158) of unique

XSL•FO
RenderX

participants, respectively. Thus, the various participant groups exhibited significantly different rates of utterances across sessions (P=.003).

Table . Summary of participant groups and their participation in discussion, with utterance rates per person-session across 23 Miners' Wellness ECHO[a] sessions during this study's time frame from July 2018 to September 2019.[b,c]

Participant group	Total PAS[d], n	Total PRAS[e], n/N (%)	Unique PAS, n	Unique PRAS, n/N (%)	Total statements of expertise, n	Average expertise per person-session	Total statements of acceptance, n	Average acceptance per person-session	Total statements of eschewal, n	Average eschew per person-session	Total utterances across sessions, n	Average utterances per person-session
Attorney	54	54/539 (10)	11	11/158 (6.9)	77	3.5	58	2.64	13	0.59	148	6.73
Benefits counselor	87	87/539 (16.1)	13	13/158 (8.2)	6	0.26	7	0.3	0	0	13	0.57
Clinical provider	186	186/539 (34.5)	58	58/158 (36.7)	280	12.17	182	7.91	38	1.65	500	21.74
Home health professional	89	89/539 (16.5)	36	36/158 (22.8)	36	1.71	21	1	3	0.14	60	2.86
Other or unknown	63	63/539 (11.7)	27	27/158 (17.1)	56	2.43	48	2.09	7	0.3	111	4.83
Respiratory therapist	60	60/539 (11.1)	13	13/158 (8.2)	53	2.3	31	1.35	12	0.52	96	4.17
Total	539	539/539 (100)	158	158/158 (100)	508	22.09	347	15.09	73	3.17	928	40.35

[a]ECHO: Extension for Community Health Outcome.

[b]Utterances include statements of expertise, acceptance, and eschewal. Additional detail of utterances is provided in Table S1 in Multimedia Appendix 1.

[c]Details of individual ECHO session content are shown in Table S2 in Multimedia Appendix 1 to provide more context to data shown in Figure 1.

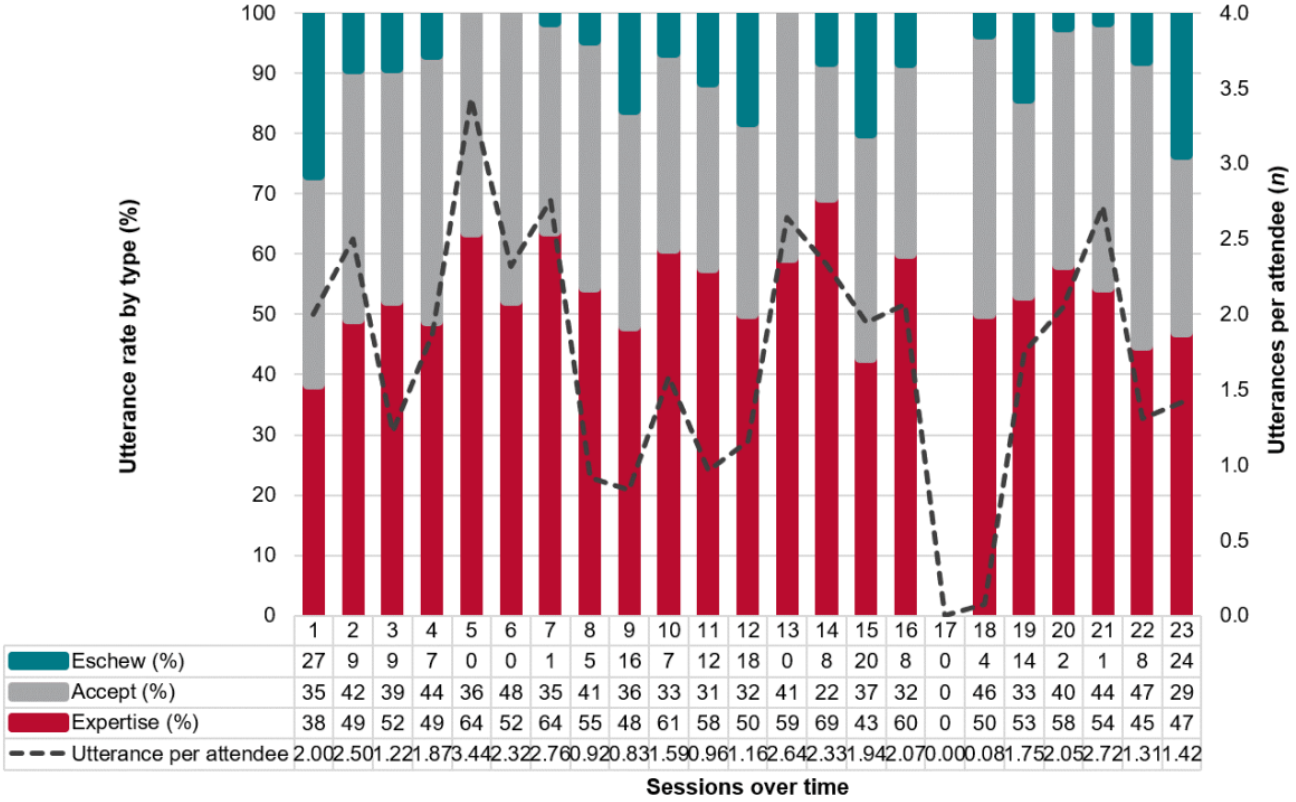
[d]PAS: participants across sessions.

[e]PRAS: participant representation in all sessions.

Some topics generated a lot of discussion (ie, high rates of utterances), while others did not (Figure 1). Overall, utterances per person-session did not change over time (P=.50, Figure 1). It should be noted that in session 17, there were no utterances—this is the only session facilitated by a different person who was an experienced ECHO participant and expert clinical provider but not a trained facilitator. As shown in Table 2, most utterances are statements of expertise, and most statements are accepted. As shown in Figure 1, some topics generate a lot of controversy during discussion (ie, high rates of eschewal), while others do not.

XSL•FO
RenderX

Figure 1. Rates of utterances across the 23 Miners' Wellness ECHO sessions during this study's time frame, July 2018 to September 2019. This figure shows the average number of utterances per person-session (in gray line) and the percentage breakdown of each type of utterance (teal for eschewal, gray for acceptance, and red for statements of expertise). ECHO: Extension for Community Health Outcomes.



Do Participants Who Are More Experienced With ECHO Have Greater Rates of Active Participation in Discussion Than Those With Lower Levels of Experience?

Experienced and intermediate experienced groups accounted for 36.8% (n=104) and 26.9% (n=29) of active participation in discussion, respectively, while the active participation from the less experienced group accounted for 28.6% (n=40). Experienced participants may have, thus, tended to dominate active participation in discussion compared to participants with less or intermediate levels of experience, but this difference was not statistically significant (*P*=.11, Table 3).

Table . Association of participants' ECHO[a] experience with active participation in discussion, as measured by any utterance during an individual session, during this study's time frame from July 2018 to September 2019.

Group	Nonactive (n)	Active, n (%)	OR[b] (95% CI)	*P* value	Overall *P* value
Experienced	179	104 (36.8)	1.45 (0.88-2.39)	.14	.11
Intermediate experience	79	29 (26.9)	0.92 (0.52-1.64)	.77	Referent
Less experienced	100	40 (28.6)	Referent	—[c]	Referent

[a]ECHO: Extension for Community Health Outcomes.

[b]OR: odds ratio.

[c]Not applicable.

Does the Participant Group Identity of the Didactic Presenter Predict Active Participation in Discussion From the Same Participant Group During That Session?

Participant group identity was considered concordant when the didactic presenter and participant belonged to the same participant group. For this analysis, the participant groups were categorized as silent versus vocal, in which the benefits counselors and home health professionals constituted the former group and the clinical providers, attorneys and respiratory therapists in the latter group (with the other or unknown group excluded). When the didactic presenter and participant were from the same participant group, active participation in discussion (measured by any utterances within an individual ECHO session) by the silent group participants was 26.9% (n=7) compared to 9.5% (n=14) when the didactic presenter was from a different group (odds ratio 3.5, 95% CI 1.47-8.54, Table 4). Among vocal group participants, active participation was 46.4% (n=96) when presenters were in the same group as the

XSL•FO
RenderX

participant, compared to 46.2% (n=43) when presenters and participants were from different groups (odds ratio 1.01, 95% CI 0.58-1.75), with an interaction *P*=.003.

Table . Association of active participation in Miners' Wellness ECHO[a] session discussion by silent and vocal groups on concordance of participant group identity between didactic presenter and participant, during this study's time frame from July 2018 to September 2019.

Participant group concordance between presenter or participant[b]	Vocal group (n=300)			Silent group (n=173)			Interaction *P* value
	Nonactive (n)	Active, n (%)	OR[c] (95% CI)	Nonactive (n)	Active, n (%)	OR (95% CI)	
Concordant	111	96 (46.4)	1.01 (0.58-1.75)	19	7 (26.9)	3.5 (1.47-8.54)	.003
Discordant	50	43 (46.2)	Referent	133	14 (9.5)	Referent	Referent

[a]ECHO: Extension for Community Health Outcomes.

[b]For this analysis, the participant groups were categorized as silent versus vocal, in which the benefits counselors and home health professionals constituted the former group and the clinical providers, attorneys and respiratory therapists in the latter group (with the other or unknown group excluded).

[c]OR: odds ratio.

Does Video Usage Predict Active Participation in Discussion, as Compared to Audio Usage?

Compared to those joining a session by audio, those joining by video tended to have higher rates of active participation in discussion, as measured by any utterance during an individual session. However, this comparison did not reach statistical significance (Table 5, *P*=.11).

Table . Association of active participation in Miners' Wellness ECHO[a] session discussion on video camera use, during this study's time frame from July 2018 to September 2019.

Group	Nonactive participation (n)	Active participation, n (%)	OR[b] (95% CI)	*P* value
Audio[c]	110	54 (32.9)	Referent	.11
Video[c]	175	119 (40.5)	1.39 (0.93-2.07)	Referent

[a]ECHO: Extension for Community Health Outcomes.

[b]OR: odds ratio.

[c]Those whose audio and video status could not be determined (n=81) were excluded from the analysis.

Does Session Attendance Frequency Predict Active Participation in Discussion?

Unique participants' attendance ranged from 1 to 20 sessions and was categorized into 3 groups (Table 6). Overall, active participation in the discussion was significantly associated with attendance (*P*=.004); however, no dose response was noted.

Table . Association of active participation in Miners Wellness ECHO[a] session discussion on frequency of session attendance, during this study's time frame from July 2018 to September 2019.

Session attendance frequency	Nonactive participation (n)	Active participation, n (%)	OR[b] (95% CI)	*P* value	Overall *P* value
≤5 sessions	76	56 (42.4)	Referent	—[c]	.004
6 - 10 sessions	1	12 (92.3)	16.3 (2.1-128.9)	.008	Referent
11 - 20 sessions	4	11 (73.3)	3.7 (1.1-12.3)	.03	Referent

[a]ECHO: Extension for Community Health Outcomes.

[b]OR: odds ratio.

[c]Not applicable.

XSL•FO
RenderX

Discussion

Principal Findings

In this digital community of practice for managing pneumoconiosis with previously published efficacy outcomes [14,18,20], our study findings show that some participant groups are vocal while others are silent, based on utterance rates during interprofessional discussion. Some topics generate a lot of discussion (ie, relatively high rates of utterances), and some topics generate a lot of controversy during discussion (ie, relatively high rates of eschewal of statements of expertise). Inadequately trained facilitators may impede discussion during a session. Experienced participants may dominate the discussion and inhibit participation by less experienced participants, but this association did not reach statistical significance (P=.11, Table 3). When the didactic presenter and participant are from the same participant group, participation in discussion by the silent group participants rises significantly. Compared to those participating by audio, those participating by video camera may tend to have higher rates of participation in discussion, but this association did not reach statistical significance. Overall, active participation in the discussion was significantly associated with attendance frequency. Our study findings lend themselves to several best practice recommendations, as discussed below.

The primary objective of the Project ECHO movement is to decentralize knowledge for the care of patients through exchanging insights and information using the all teach, all learn principle. All participants have unique knowledge sets, and discussion inside and outside the ECHO session within and across participant groups facilitates the transfer of knowledge that would otherwise remain siloed within individual participant groups. Further, interprofessional discussion in the ECHO model may allow greater access to new and thought-provoking ideas and perspectives that foster learning and other growth-enhancing actions. Our study findings confirm the existence of silent participant groups that, by not fully engaging in discussions, may not be as effective in exchanging knowledge (including providing and acquiring knowledge) within the ECHO session. These groups need to be actively supported by a trained facilitator and by the judicious use of didactic presenters sharing participant group identity concordant with the silent group participants. However, our previous objective measurement of knowledge transfer using social network principles indicates that benefits counselors are among the groups most effective in knowledge transfer [18]. This prior finding would suggest that either benefits counselors convey relevant knowledge succinctly within the ECHO sessions or participate in knowledge transfer outside the strict bounds of the ECHO program via follow-up phone calls, emails, and virtual and physical meetings. This finding warrants further research as it is a possible indication that the informal “social network” system, perhaps provoked by or aided by the ECHO sessions, may be more important than and reinforce the formal sessions.

Project ECHO formally trains its facilitators on the best practices for managing discussions in a digital community of practice. Substituting a trained facilitator by another untrained expert can affect participation in the discussion. Experienced participants usually offer the greatest rates of statements of expertise, acceptance, and eschewal, reflecting their expert status in this complex field. However, the dominance of experienced groups can be avoided by the facilitator actively encouraging less experienced groups to make their minority opinions heard during discussion. Using video cameras during participation in this digital community of practice helps build community and promotes accountability and engagement at the cost of rural internet bandwidth. This practice should be encouraged, even though the findings did not reach statistical significance.

Our study has multiple strengths. To the best of our knowledge, our approach of studying interprofessional discussion in a digital “community of practice” for managing pneumoconiosis has never been used previously by another group of investigators. This study involves the innovative ECHO model intervention that addresses barriers to the care of miners by providing a multidisciplinary community of practice approach, which has been well studied in other diseases [17,25-27]. This study is topical and significant because it addresses a critical gap related to the emerging pneumoconiosis epidemic in rural United States. Since no one refused participation, there was likely no potential participation bias in this study. Since the ECHO model has been adopted nationwide and worldwide to care for patients with numerous chronic diseases, infrastructure already exists to rapidly scale the Miners’ Wellness ECHO Program nationwide and worldwide. Other strengths include the detailed contemporaneous interpretation of videos and transcripts of discussions; strict quality control in qualitative analysis; a mixed methods study design; and the simultaneous use of qualitative, quantitative, and subject content experts as coinvestigators.

Although our study focused on the ECHO program, which emphasizes chronic disease management with a strong focus on mentorship and collaborative learning, our findings have broader implications for other digital communities of practice that may be structured differently and operate differently. Examples include Elpha (Women in Tech Network), a community of practice for women in technology; Stack Overflow (Stack Exchange Inc) for programmers and developers; Digital Nomad World for nonprofessional digital nomads [31]; and various leadership communities of practice that emphasize succession planning and leadership skills development [32]. Such unrelated communities of practice can also benefit from our evidence-based findings and recommendations to reach out to silent groups, break down silos, provide facilitator training, and encourage active participation. On the other hand, the ECHO program could learn from other communities of practice on how to reward active participation and knowledge sharing (Stack Overflow) or create asynchronous platforms to supplement synchronous engagement (Digital Nomad World).

There are also limitations to this study. We cannot correlate interprofessional discussion for knowledge transfer to patient outcomes or changes in provider behavior. However, we have previously published a listing of qualitative changes our ECHO participants reported they would make in their practice, obtained as part of a continuing medical education survey requested at the end of each ECHO session [20]. Participants whose group identity was other or unknown were 11.7% (63/539) of the total

XSL•FO
RenderX

participant representation in all sessions. When analyzing participant group identity (as shown in Table 4), the other or unknown group was excluded. Although a small sample size raises the possibility of a type I error, individual professionals and teams of professionals trained in the ECHO model can reach a large number of miners, with the potential for creating exponential change. Miners impacted with pneumoconiosis do not participate in the ECHO model. This model characteristic is, however, not a limitation for the following reasons. The model ensures that knowledge is disseminated widely and effectively by focusing on provider education. By reaching out to more miner professionals, the ECHO model indirectly benefits more miners and is more scalable than traditional telemedicine models focusing on individual patients. Case-based learning, where providers present anonymized patient cases to a panel of experts, allows for in-depth discussion and learning without compromising patient confidentiality while maintaining the traditional provider-patient relationship [25,33].

Conclusions

Our findings provide insight into the mechanics of interprofessional discussion for knowledge transfer in a digital community of practice managing pneumoconiosis and potential recommendations to enhance the same. Our results underscore the capacity of the Project ECHO model to leverage technology and workforce diversity to facilitate interprofessional discussions on the multidisciplinary and complex care of miners and ultimately promote health equity among rural and medically underserved mining communities. Although this approach addresses a critical gap related to the emerging pneumoconiosis epidemic, future research will evaluate whether this translates into improved patient outcomes in mining communities, a priority need in rural United States.

Acknowledgments

Generative artificial intelligence was not used in any portion of this paper. This study was funded by the Alpha Foundation (grant AFC719-14 of US $250,000) and partly supported by the National Center for Advancing Translational Sciences, National Institutes of Health (UL1TR001449). The sponsors played no role in conducting, analyzing, and disseminating this study or reviewing and approving this paper for publication. We want to thank Rachelle Rochelle, MPA, and Stephen Murillo, MBA, who were integral to the collection of this data.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

All authors were involved with the following: (1) substantial contributions to the conception or design of this work (HRB, VAS, XS, OM, and AS) and the acquisition (HRB, XS, OM, and AS), analysis (HRB, XS, and OM), or interpretation of data for this work (HRB, VAS, XS, OM, and AS); (2) drafting this work or reviewing it critically for important intellectual content (HRB, VAS, XS, OM, and AS); (3) final approval of the version to be published (HRB, VAS, XS, OM, and AS); and (4) agreement to be accountable for all aspects of this work in ensuring that questions related to the accuracy or integrity of any part of this work are appropriately investigated and resolved (HRB, VAS, XS, OM, and AS).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example conversations and details of individual ECHO (Extension for Community Health Outcomes) session content.
[DOCX File, 35 KB - formative_v9i1e67999_app1.docx]

References

1. Blackley DJ, Halldin CN, Laney AS. Continued increase in prevalence of coal workers' pneumoconiosis in the United States, 1970-2017. Am J Public Health 2018 Sep;108(9):1220-1222. [doi: 10.2105/AJPH.2018.304517] [Medline: 30024799]
2. Hall NB, Blackley DJ, Halldin CN, Laney AS. Continued increase in prevalence of r-type opacities among underground coal miners in the USA. Occup Environ Med 2019 Jul;76(7):479-481. [doi: 10.1136/oemed-2019-105691] [Medline: 31023786]
3. Laney AS, Attfield MD. Coal workers' pneumoconiosis and progressive massive fibrosis are increasingly more prevalent among workers in small underground coal mines in the United States. Occup Environ Med 2010 Jun;67(6):428-431. [doi: 10.1136/oem.2009.050757] [Medline: 20522823]
4. Laney AS, Blackley DJ, Halldin CN. Radiographic disease progression in contemporary US coal miners with progressive massive fibrosis. Occup Environ Med 2017 Jul;74(7):517-520. [doi: 10.1136/oemed-2016-104249] [Medline: 28408654]

XSL•FO
RenderX

5. Blackley DJ, Halldin CN, Laney AS. Resurgence of a debilitating and entirely preventable respiratory disease among working coal miners. Am J Respir Crit Care Med 2014 Sep 15;190(6):708-709. [doi: 10.1164/rccm.201407-1286LE] [Medline: 25221884]
6. Dwyer-Lindgren L, Bertozzi-Villa A, Stubbs RW, et al. Trends and patterns of differences in chronic respiratory disease mortality among US counties, 1980-2014. JAMA 2017 Sep 26;318(12):1136-1149. [doi: 10.1001/jama.2017.11747] [Medline: 28973621]
7. Almberg KS, Halldin CN, Blackley DJ, et al. Progressive massive fibrosis resurgence identified in U.S. coal miners filing for black lung benefits, 1970-2016. Ann Am Thorac Soc 2018 Dec;15(12):1420-1426. [doi: 10.1513/AnnalsATS.201804-261OC] [Medline: 30114941]
8. Arnold C. A scourge returns: black lung in Appalachia. Environ Health Perspect 2016 Jan;124(1):A13-A18. [doi: 10.1289/ehp.124-A13] [Medline: 26720594]
9. Fan VS, Giardino ND, Blough DK, Kaplan RM, Ramsey SD, Nett Research Group. Costs of pulmonary rehabilitation and predictors of adherence in the National Emphysema Treatment Trial. COPD 2008 Apr;5(2):105-116. [doi: 10.1080/15412550801941190] [Medline: 18415809]
10. Kirby JB, Yabroff KR. Rural-urban differences in access to primary care: beyond the usual source of care provider. Am J Prev Med 2020 Jan;58(1):89-96. [doi: 10.1016/j.amepre.2019.08.026] [Medline: 31862103]
11. Croft JB, Lu H, Zhang X, Holt JB. Geographic accessibility of pulmonologists for adults with COPD: United States, 2013. Chest 2016 Sep;150(3):544-553. [doi: 10.1016/j.chest.2016.05.014] [Medline: 27221645]
12. Miedema B, Hamilton R, Fortin P, Easley J, Tatemichi S. The challenges and rewards of rural family practice in New Brunswick, Canada: lessons for retention. Rural Remote Health 2009;9(2):1141. [Medline: 19496642]
13. Crockett LK, Leggett C, Curran JA, et al. Knowledge sharing between general and pediatric emergency departments: connections, barriers, and opportunities. CJEM 2018 Feb 22:1-9. [doi: 10.1017/cem.2018.7] [Medline: 29467040]
14. Sood A, Pollard C, Kalishman S, et al. Telementoring of healthcare teams in the care of miners. ATS Sch 2020 Dec 22;2(1):66-83. [doi: 10.34197/ats-scholar.2020-0073OC] [Medline: 33870324]
15. Evans K, Lerch S, Boyce TW, et al. An innovative approach to enhancing access to medical screening for miners using a mobile clinic with telemedicine capability. J Health Care Poor Underserved 2016;27(4A):62-72. [doi: 10.1353/hpu.2016.0182] [Medline: 27818414]
16. Wenger E. Communities of Practice: Learning, Meaning and Identity: Cambridge University Press; 1998.
17. Arora S, Thornton K, Murata G, et al. Outcomes of treatment for hepatitis C virus infection by primary care providers. N Engl J Med 2011 Jun 9;364(23):2199-2207. [doi: 10.1056/NEJMoa1009370] [Medline: 21631316]
18. Sood A, Assad N, Jarrell W, et al. A virtual community-of-practice approach by rural stakeholders in managing pneumoconiosis in the USA: A cross-sectional analysis. Rural Remote Health 2020 Aug;20(3):5784. [doi: 10.22605/RRH5784] [Medline: 32847365]
19. Files JA, Blair JE, Mayer AP, Ko MG. Facilitated peer mentorship: a pilot program for academic advancement of female medical faculty. J Womens Health (Larchmt) 2008;17(6):1009-1015. [doi: 10.1089/jwh.2007.0647] [Medline: 18681821]
20. Soller B, Myers O, Sood A. Transfer of knowledge on pneumoconiosis care among rural-based members of a digital community of practice: cross-sectional study. JMIR Form Res 2024 Jan 24;8:e52414. [doi: 10.2196/52414] [Medline: 38265861]
21. Ranmuthugala G, Plumb JJ, Cunningham FC, Georgiou A, Westbrook JI, Braithwaite J. How and why are communities of practice established in the healthcare sector? A systematic review of the literature. BMC Health Serv Res 2011 Oct 14;11:273. [doi: 10.1186/1472-6963-11-273] [Medline: 21999305]
22. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. Acad Emerg Med 2008 Nov;15(11):988-994. [doi: 10.1111/j.1553-2712.2008.00227.x] [Medline: 18778378]
23. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
24. Wenger E. How we learn. Communities of practice. The social fabric of a learning organization. Healthc Forum J 1996;39(4):20-26. [Medline: 10158755]
25. Arora S, Kalishman S, Thornton K, et al. Expanding access to hepatitis C virus treatment--Extension for Community Healthcare Outcomes (ECHO) project: disruptive innovation in specialty care. Hepatology 2010 Sep;52(3):1124-1133. [doi: 10.1002/hep.23802] [Medline: 20607688]
26. Bouchonville MF, Hager BW, Kirk JB, Qualls CR, Arora S. Endo echo improves primary care provider and community health worker self-efficacy in complex diabetes management in medically underserved communities. Endocr Pract 2018 Jan;24(1):40-46. [doi: 10.4158/EP-2017-0079] [Medline: 29368967]
27. Deming P, Arora S. Taribavirin in the treatment of hepatitis C. Expert Opin Investig Drugs 2011 Oct;20(10):1435-1443. [doi: 10.1517/13543784.2011.606214] [Medline: 21854301]
28. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]

XSL•FO
RenderX

29. Moss P, Hartley N, Russell T. Project ECHO®: a global cross-sectional examination of implementation success. BMC Health Serv Res 2024 May 3;24(1):583. [doi: 10.1186/s12913-024-10920-5] [Medline: 38702685]
30. Kearsley G, Shneiderman B. Engagement theory: a framework for technology-based teaching and learning. Educ Technol 1998;38(5):20-23.
31. Burton C. 6 community of practice examples (+ different types). Thinkific. 2022 Aug 16. URL: https://www.thinkific.com/blog/community-of-practice-examples/#non-professional [accessed 2025-02-22]
32. DePodesta M. The development of leadership communities of practice: utilizing the ADKAR framework. Nurs Adm Q 2024;48(3):225-233. [doi: 10.1097/NAQ.0000000000000648] [Medline: 38848484]
33. Zhou C, Crawford A, Serhal E, Kurdyak P, Sockalingam S. The impact of Project ECHO on participant and patient outcomes: a systematic review. Acad Med 2016 Oct;91(10):1439-1461. [doi: 10.1097/ACM.0000000000001328] [Medline: 27489018]

Abbreviations

ECHO: Extension for Community Health Outcomes
REDCap: Research Electronic Data Capture

Edited by A Mavragani; submitted 25.10.24; peer-reviewed by G Kernohan, W LaMendola; revised version received 23.01.25; accepted 24.01.25; published 13.03.25.

Please cite as:
Sood VA, Rishel Brakey H, Myers O, Shore X, Sood A
Interprofessional Discussion for Knowledge Transfer in a Digital "Community of Practice" for Managing Pneumoconiosis: Mixed Methods Study
JMIR Form Res 2025;9:e67999
URL: https://formative.jmir.org/2025/1/e67999
doi:10.2196/67999

© Varinn Avi Sood, Heidi Rishel Brakey, Orrin Myers, Xin Shore, Akshay Sood. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Telemedicine Prescribing by US Mental Health Care Providers: National Cross-Sectional Survey

Mollie R Cummins[1,2], BSN, MSN, PhD; Julia Ivanova[2], PhD; Hiral Soni[2], MS, MBA, PhD; Zoe Robbins[1], DNP; Brian E Bunnell[2,3], PhD; Esteban López[2], MD, MBA; Brandon M Welch[2,4], MS, PhD

[1]College of Nursing, University of Utah, 10 S 2000 East, Salt Lake City, UT, United States

[2]Doxy.me, Inc, Charleston, SC, United States

[3]Department of Psychiatry and Behavioral Neurosciences, Morsani College of Medicine, University of South Florida, Tampa, FL, United States

[4]Biomedical Informatics Center, Public Health and Sciences, Medical University of South Carolina, Charleston, SC, United States

Corresponding Author:
Mollie R Cummins, BSN, MSN, PhD
College of Nursing, University of Utah, 10 S 2000 East, Salt Lake City, UT, United States

Abstract

Background: In the postpandemic era, telemedicine continues to enable mental health care access for many people, especially persons living in areas with mental health care provider shortages. However, as lawmakers consider long-term telemedicine policy decisions, some question the safety and appropriateness of prescribing via telemedicine, and whether there should be requirements for in-person evaluation, especially for controlled substances.

Objective: Our objective was to assess US telemental health care provider perceptions of comfort and perceived safety in prescribing medications, including controlled substances, via telemedicine.

Methods: We conducted a web-based, cross-sectional survey of US telemental health care providers who prescribe via telemedicine, using nonprobability, availability sampling of a national telehealth research panel from February 13 to April 28, 2024. We used descriptive statistics, visualization, and thematic analysis to analyze results. We assessed differences in response distribution by health care provider licensure type (physician vs nonphysician) and specialty (psychiatry vs nonpsychiatry) using the Mann-Whitney *U* test.

Results: A total of 115 screened and eligible panelists completed the survey. Overall, participants indicated high levels of comfort with prescribing via telemedicine, with 84% (102/115) of health care providers indicating they strongly agree with the statement indicating comfort in prescribing medications via telemedicine. However, participants indicated less comfort in prescribing if they have never seen a patient in person, or if the patient is located out-of-state. Most participants indicated they can safely prescribe controlled substances via telemedicine, without having previously provided care to a patient in person. However, 14.8% (17/115) to 19.1% (30/115) of health care providers (by schedule) felt that they could rarely or never safely prescribe controlled substances. There were some differences in perception of comfort and safety by licensure and specialty. Among controlled substance schedules, participants indicated the least perceived safety with schedule IV medications, and the most safety with schedule II and III medications.

Conclusions: These health care providers were highly comfortable prescribing both scheduled and unscheduled medications via telemedicine. Comfort and perceived safety with telemedicine prescribing varied somewhat by licensure type (physician vs nonphysician) and specialty (psychiatry vs nonpsychiatry). Perceived safety varied moderately for scheduled medications (controlled substances), especially for schedule IV and V medications. Participants indicated use of adaptive strategies to prescribe safely depending upon the clinical context. In ongoing efforts, we are analyzing additional survey results and conducting qualitative research related to telemedicine prescribing. A strong understanding of prescriber perspectives and experience with telemedicine prescribing is needed to support excellent clinical practice and effective policy making in the United States.

(JMIR Form Res 2025;9:e63251) doi:10.2196/63251

KEYWORDS

telemedicine; telehealth; telemental; provider; professional; experience; attitude; opinion; perception; perspective; prescribing; prescription; drug; pharmacology; pharmacotherapy; pharmaceutic; pharmaceutical; medication; mental health; digital health; informatics; buprenorphine; ketamine; cross sectional; survey; questionnaire

XSL•FO
RenderX

Introduction

During the COVID-19 public health emergency, the need for mental health services increased and telemedicine enabled critical access to care [1,2]. In the postpandemic era, telemedicine continues to enable mental health care access for many people, especially persons living in areas with mental health care provider shortages [3]. However, as lawmakers consider long-term policy decisions related to telemedicine, some have raised questions about the safety and appropriateness of prescribing via telemedicine, and whether there should be requirements for in-person evaluation, especially for controlled substances.

Before the US COVID-19 public health emergency declaration, most health care providers lacked telemedicine prescribing experience due to limited telemedicine adoption. Additionally, telemedicine-based prescribing of controlled substances was restricted during the prepandemic era, in compliance with the 2008 Ryan Haight Online Pharmacy Consumer Protection Act (Ryan Haight Act). Pursuant to the Ryan Haight Act, the US Drug Enforcement Agency (DEA) required that health care providers conduct an in-person evaluation of patients, before prescribing controlled substances via telemedicine. This measure was intended to prevent health care providers from prescribing potentially harmful medications with only minimal and inadequate online interaction with a patient. During the US public health emergency declaration, temporary policy flexibilities enabled controlled substance prescribing without the requirement of an in-person evaluation. In part, this enabled crucial access to buprenorphine for opioid use disorder treatment [4,5]. Now, consumers and stakeholders are calling for new policy that enables continued access to mental health care and substance use treatment, including medication-based treatment, via telemedicine [6]. The challenge is to develop evidence-based policy that supports safety while enabling critical health care access.

Current evidence is inconclusive but suggests that prescribing patterns can differ when care is provided via telemedicine versus in-person care. A 2023 study at a single institution showed that orthopedic patients are prescribed higher doses (milligram morphine equivalent) of opioids via video telemedicine than during in-person visits. McCabe et al [7] found that telemedicine is used less frequently than in-person visits for prescribing antibiotics. However, other studies have shown little or no difference in the prescribing patterns associated with telemedicine use [8,9]. These varied findings likely relate to differences in the clinical context for prescribing, including patient and health care provider characteristics, system-level factors, or the specific medications and their unique requirements for appropriate initiation and monitoring (and whether those requirements can be met via telemedicine). Evidence supporting safety and quality of telemedicine prescribing in mental health care is scarce, and also likely to be context dependent. However, a recent, large cohort study of health care claims data found no difference in safety or quality for telemedicine-based treatment versus in-person treatment of opioid use disorder, including medication-based opioid use disorder treatment [10].

Current health care provider perspectives and practices related to prescribing via telemedicine for mental or behavioral health care are not well-characterized. However, a better understanding of telemental health care provider's perspectives and experiences related to prescribing is needed to inform appropriate telemedicine program design, identify key research questions, pursue clinical practice guidelines, and develop curricula for professional educational programs. The perspectives and experience of practicing telemental health care providers should also inform policy decisions, such as the currently pending US DEA rules governing telemedicine prescribing of buprenorphine for opioid use disorder. The purpose of this study was to assess the perspectives of US telemental health care providers related to their comfort and perceived safety in prescribing medications, including controlled substances, via telemedicine.

Methods

Study Design

We conducted a web-based, cross-sectional survey of telemental health providers who prescribe via telemedicine, using nonprobability, availability sampling of a telehealth research panel from February 13 to April 28, 2024.

Survey

The web-based survey, administered using Qualtrics, was created by the research team and designed to elicit perspectives and practices of US mental or behavioral health care providers related to prescribing via telemedicine. The survey consisted of 8 sections: informed consent (1 question), verification of eligibility (5 questions), demographic and practice information (13 questions), comfort and safety (6 questions), health history (7 questions), physical assessment and diagnostic testing (7 questions), issuing a prescription (8 questions), and legal and regulatory environment (8 questions). A copy of the survey is provided in Multimedia Appendix 1.

This study is focused on a subset of the survey items, including Likert-scale items that measured agreement with statements indicating comfort with prescribing in varied scenarios, including in person (no telemedicine), via telemedicine, and via telemedicine with and without previous in-person care. We also analyzed Likert-scale items that measured agreement with statements of ability to safely prescribe medications, by US DEA scheduling. Recognizing that safe prescribing can depend upon individual patient characteristics and contextual factors aside from telemedicine use or pharmacotherapy, we designed the Likert scale to measure how often medications of a given schedule can be prescribed safely via telemedicine. Optional open-ended items prompted participants to elaborate upon their responses.

Survey Development

The interdisciplinary research team, which includes expertise in telemedicine, clinical pharmacology, and advanced practice nursing in primary care, developed the initial survey based on a literature review. A small group of telemedicine practitioners reviewed an early draft of the survey and provided input on its content, format, and relevance for clinical practice. We then refined the survey according to their input. Before initiating

XSL•FO
RenderX

data collection, we informally pretested and modified the survey within the research team, then formally pretested the survey with the target audience of 5 prescribing telemental health providers located in Utah and North Carolina who met the inclusion criteria. All 5 individuals who pretested the survey indicated that they strongly agree with the following statements: (1) the content and wording of the survey is appropriate; (2) the survey is easy to understand; (3) the survey is free of errors; (4) the time required to complete the survey is reasonable; and (5) the time required to complete the survey would be reasonable if the participants are compensated. Pretesters completed the survey in mean 11.9 (SD 4.24) minutes and recommended US $25 - $50 as compensation for completing the survey. The survey was configured to assign completion codes, so that the process of participant compensation, which requires collecting additional personal information, could be managed separately from data collection.

Sampling and Recruitment

We implemented nonprobability, availability sampling of a telehealth research panel. We calculated a target sample size of 382 for 95% CI. We currently lack precise data describing the number of US health care providers who (1) provide mental health care and (2) prescribe via telemedicine. Therefore, we based the target sample size Health Resources and Services Administration estimates of the total number of US mental health care providers in prescribing roles [11].

We recruited survey participants from the TelehealthEngage Research Panel, a panel of 7134 telemedicine users who have consented to be contacted about opportunities to participate in research. TelehealthEngage includes individuals from 49 states who are active users of Doxy.me, a commercial telemedicine platform. We invited all TelehealthEngage panelists who identified as a health professional in mental health, psychiatry, family practice, internal medicine, general practice, neurology, integrative medicine, or unknown fields or who identified as a physician, hospitalist, physician assistant (PA), nurse practitioner, or nurse. Recruitment was initiated on February 13, 2024, and closed on April 28, 2024. We divided our 2 recruitment waves over the course of 3 days each: the first invite was sent to 2343 unique panelists and the second to 3490. We sent a reminder email to any panelist who had not opened the email invitation (1491 in the first wave and 3200 in the second wave) after 3 weeks. The survey was closed out 2 weeks after the last invitation reminder was sent out. We required responses to all categorical or numeric survey items to complete the survey and be compensated for participation with a US $50 gift card.

Ethical Considerations

This study was reviewed and approved by the BRANY (Biomedical Research Alliance of New York) Institutional Review Board (IRB00010793). To protect the anonymity of participants, we configured Qualtrics to assign a random completion code to each completed survey. Upon survey completion, a code was displayed to participants along with instructions to visit a separate Qualtrics survey to submit the code and contact information for compensation. In this way, we separated participants' responses from their contact information. We stored the data in a secure environment with access limited to essential study personnel. Participants were compensated for participation with a US $50 gift card.

Analysis

We analyzed all completed surveys. We calculated descriptive statistics and frequencies, visualized response distributions, and aggregated free-text responses for qualitative analysis. Further, we assessed group differences in response distributions by licensure type (physician vs nonphysician) and specialty type (psychiatric vs nonpsychiatric) for items using Mann-Whitney *U* tests. We performed all statistical analysis using SPSS (version 29; IBM Corp) and performed a qualitative, thematic analysis of open-ended items using MAXQDA (MAXQDA - Distribution by VERBI GmbH). The results of the quantitative analyses informed the way we organized codes from the thematic analysis [12,13]. Further, 1 author coded the qualitative responses from surveys using the entire participant response as the unit of analysis. After 3 iterations of coding, a codebook was developed and honed with the help of all authors. We assigned each response one or more codes, and so the total number of codes exceeds the number of responses.

Results

Participation and Response

In total, 350 TelehealthEngage panelists accessed the survey. Of those, 336 (96%) completed all eligibility screening questions, and 121 (36%) of those screened were eligible. Further, 115/121 (95%) of screened and eligible panelists completed the survey. There was no missing data in completed surveys given the requirement that all items be completed, except optional free text fields. Only 4 surveys were initiated but not completed, so we did not analyze the data or individual items for nonresponse bias. See Table 1 illustrating study participation results.

Table . Participant count by stage.

Stage	Value, n
Accessed survey	350
Completed consent and screening items	336
Eligible	121
Initiated survey	119
Completed survey	115

XSL•FO
RenderX

Participant Characteristics

Demographic characteristics of the participants are summarized in Table 2 and Figure 1. The participants were largely White, non-Hispanic, and female, with a mean age of 51.2 (SD 12.4) years and primarily physicians and advanced practice nurses in smaller-sized practices. The most common specialties were psychiatry or psychiatry specialties. However, 28 (24.3%) indicated nonpsychiatry specialties. Geographically, the participants were distributed across 26 American states (Figure 1).

XSL•FO
RenderX

Table . Demographic and practice characteristics of participants (N=115).

Characteristic	Values, n (%)
Race	
American Indian or Alaska Native	1 (0.9)
Asian	9 (7.8)
Black or African American	8 (7)
Native Hawaiian or Pacific Islander	0 (0)
White	83 (72.2)
More than 1 race	5 (4.3)
Other race	2 (1.7)
Prefer not to answer	7 (6.1)
Unknown	0 (0)
Hispanic or Latino	
Hispanic or Latino	8 (7)
Not Hispanic or Latino	97 (84.3)
Prefer not to answer	6 (5.2)
Unknown	4 (3.5)
Gender	
Male	46 (40)
Female	66 (57.4)
Nonbinary or third gender	0 (0)
Prefer not to say	3 (2.6)
Licensure	
Doctor of medicine or doctor of osteopathic medicine	69 (60)
Advanced practice registered nurse or physician assistant	40 (34.8)
Doctor of philosophy clinical psychologist	4 (3.)
Pharmacist	0 (0)
Other	2 (1.74)
Percent of clients seen via telehealth	
None (0%)	0 (0)
Few (1% - 24%)	10 (8.7)
Some (25% - 49%)	23 (20)
Most (50% - 74%)	33 (28.7)
Almost all (75% - 99%)	29 (25.2)
All (100%)	20 (17.4)
Practice size	
Independent	53 (46.1)
Small group (2 - 5 health care providers)	29 (25.2)
Mid-size (6 - 15 health care providers)	20 (17.4)
Large group practice (16+ health care providers)	13 (11.3)
Setting[a]	
Academic	6 (5.2)
Community	47 (40.9)
Hospital	4 (3.5)

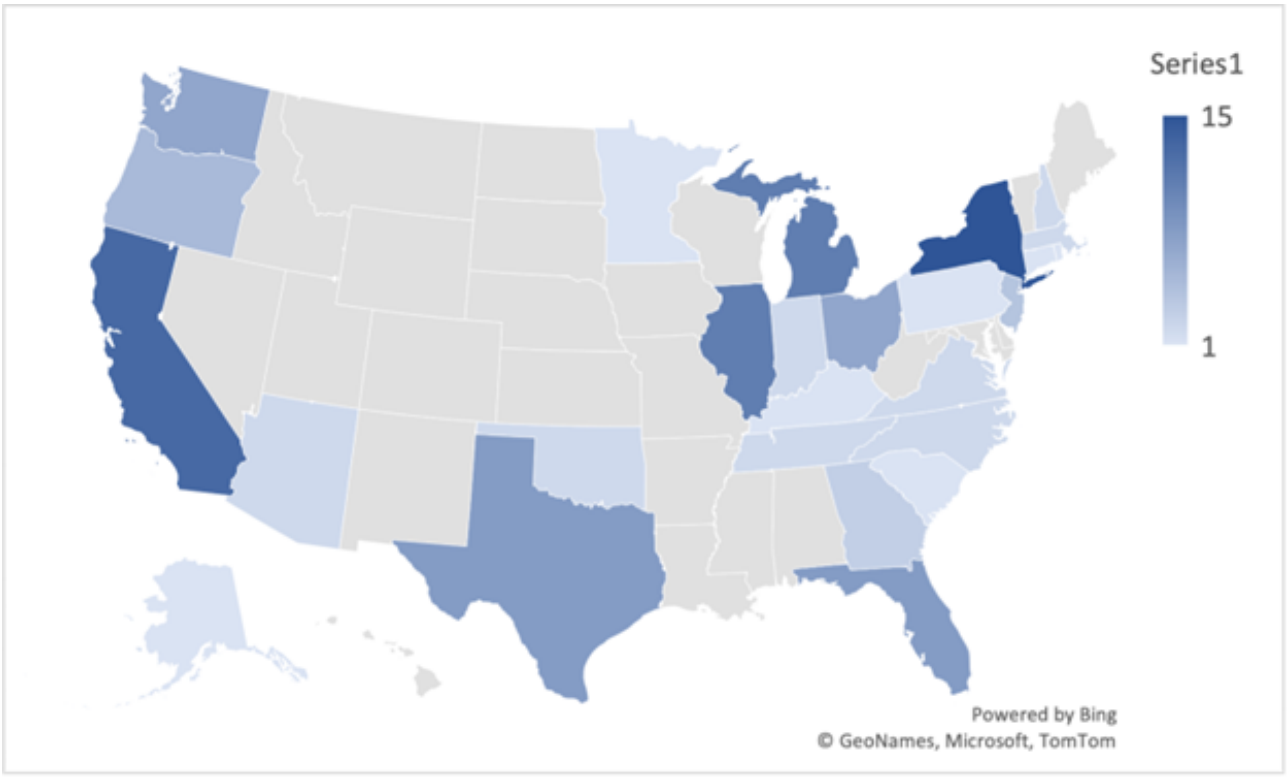
XSL•FO
RenderX

Characteristic	Values, n (%)
Clinic	58 (50.4)
School	0 (0)
Corrections	0 (0)
Federally qualified health center	4 (3.5)
Digital health care	17 (4.8)
Other	6 (5.2)
Specialty	
Psychiatry	63 (54.8)
Addiction medicine or psychiatry	6 (5.2)
Child and adolescent psychiatry	14 (12.2)
Geriatric psychiatry	4 (3.5)
Forensic psychiatry	0 (0)
Consultation liaison psychiatry	0 (0)
Family practice	11 (9.6)
Internal medicine	6 (5.2)
Pediatrics	3 (2.6)
Other	8 (7)
Experience in specialty (years)	
0 - 5	24 (20.9)
6 - 10	20 (17.4)
11 - 15	23 (20)
16+	48 (41.7)
Experience in telehealth (years)	
0 - 3.7 (since onset of COVID-19 pandemic)	70 (60.9)
3.7 - 10	39 (33.9)
11 - 15	3 (2.6)
16+	3 (2.6)

[a]Participants could select more than 1 descriptor.

XSL•FO
RenderX

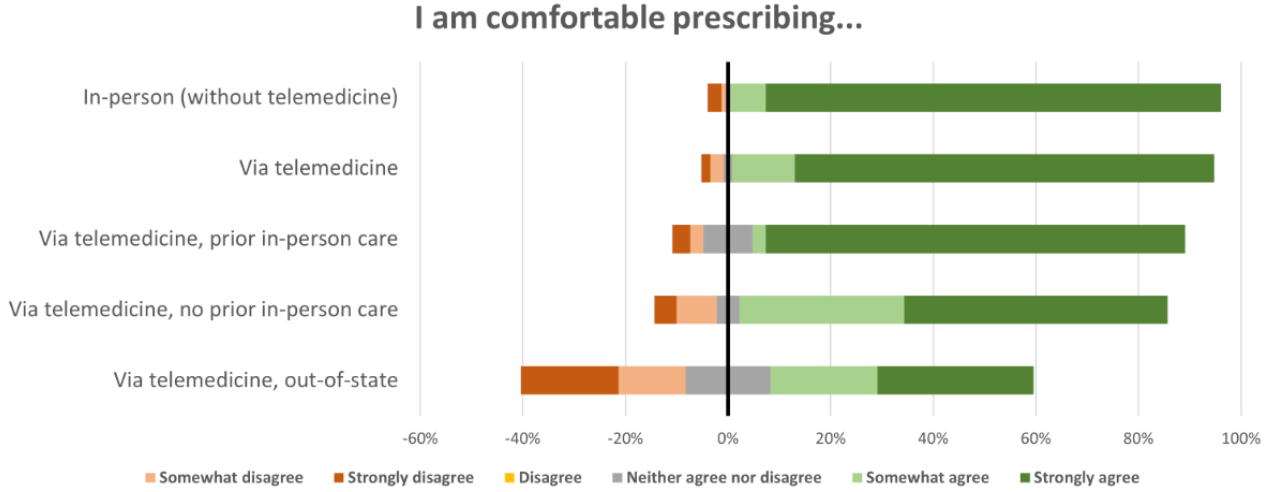
Figure 1. Geographic distribution of participants (N=115). Created using Bing Maps [14] and published under the platform's limited license [15].



Main Findings

Overall, participant health care providers indicated high levels of comfort with prescribing via telemedicine, with 84% (102/115) of participants indicating they strongly agree with the statement indicating comfort in prescribing medications via telemedicine. Only 5 (4.3%) participants somewhat or strongly disagreed. However, participants indicated less comfort in prescribing if they have never seen a patient in person, or if the patient is located out-of-state. Figure 2 shows a visualization of the response distribution; detailed descriptive statistics are provided in Multimedia Appendix 2.

Figure 2. Participant agreement with statements of comfort in prescribing via telemedicine (N=115).



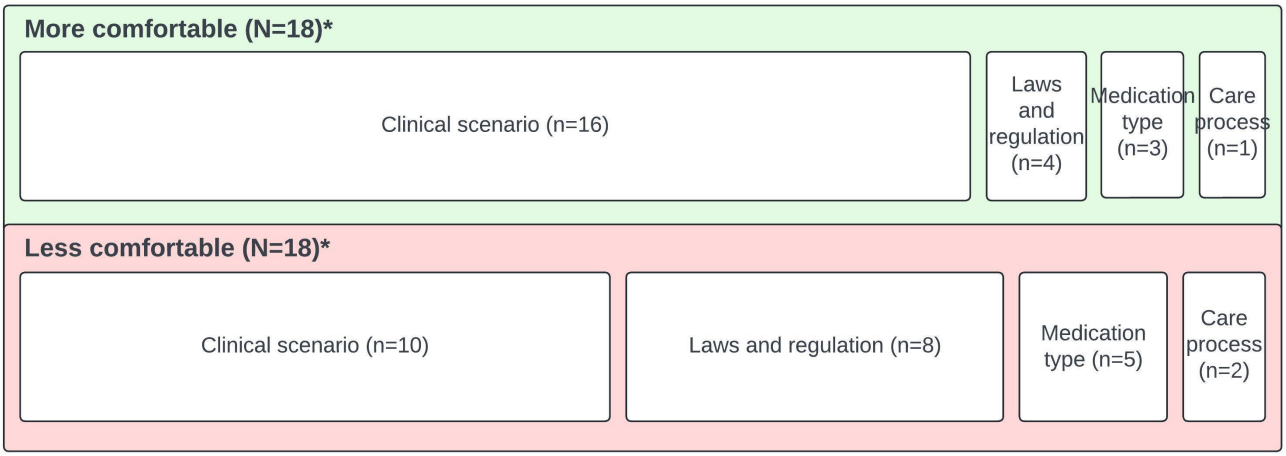
Further, 50% (58/115) of participants responded to the optional question “Please tell us more about the situations in which you feel comfortable or not comfortable prescribing.” We found 3 general categories of comments related to comfort: comments related to feeling uncomfortable (1 instance), comfortable (17 instances), or more or less comfortable depending on the specific situation (conditionally comfortable, 36 instances).

The singular response that a participant health care provider unequivocally did not feel comfortable prescribing over telemedicine was in regard to controlled substances: “Will not prescribe classified drugs or anything for ADHD online.” Health care providers who noted they unequivocally felt comfortable prescribing over telemedicine (17 instances) also included some explanation of prescribing situations regarding laboratories and assessments, laws and regulations, and types of medication:

XSL•FO
RenderX

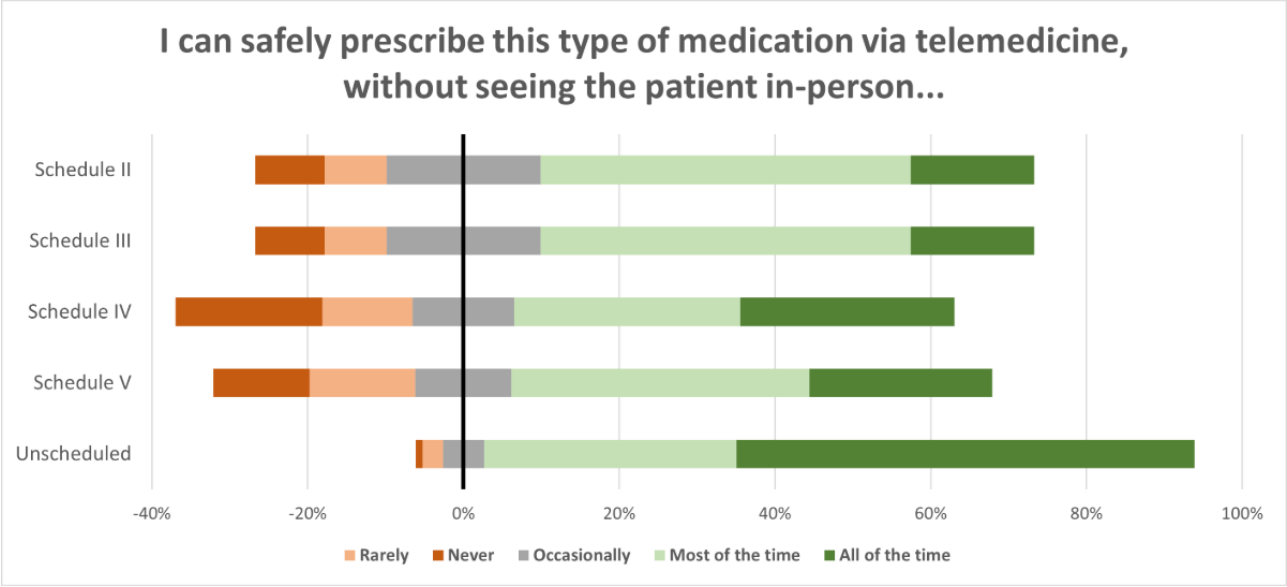
"All but one of my current patients I've met in person, but my practice is now solely telemedicine. I do my best to assess the patient's condition, personality, and response. I think I am as comfortable prescribing online as in person." The majority (36 instances) of health care providers reported their comfort levels depended upon certain conditions. The approximate distribution of condition types is depicted in Figure 3. Health care providers were more (18 codes) or less comfortable (18 codes) due to reasons relating to the clinical scenario (including patient characteristics or behavior); laws and regulations; types of medications; and the care process (characteristics of the patient's care plan or care delivery including specific types of visits, assessments, and laboratories). Numerous health care providers noted how an initial in-person intake (9 codes) or working with established patients (9 codes) affected their experiences and comfort: "I felt comfortable after having an in person visits [sic] with the patient before telemedicine. I do a physical exam on all patients during the first visit snd sm [sic] not sure I feel comfortable prescribing without that initial physical encounter." Complete results of thematic analysis for this item are provided in Multimedia Appendix 3.

Figure 3. Condition types where health care providers feel conditionally more or less comfortable prescribing via telemedicine. *N refers to codes of more or less comfortable, depending upon conditions. n refers to code counts for the condition type. The total number of codes for condition type exceeds that of codes for comfort.



For telemedicine-based prescribing of controlled substances, the results are visualized in Figure 4 (detailed descriptive statistics in Multimedia Appendix 2). Most participants indicated that they can safely prescribe controlled substances via telemedicine, without having previously provided care to a patient in person. However, 14.8% - 19.1% of health care providers (by schedule) felt that they could rarely or never safely prescribe controlled substances. Among controlled substance schedules, participants indicated the least perceived safety with schedule IV medications, and the most safety with schedule II and III medications.

Figure 4. Perceived safety of telemedicine-based prescribing for medication types, by US Drug Enforcement Agency drug scheduling (N=115).



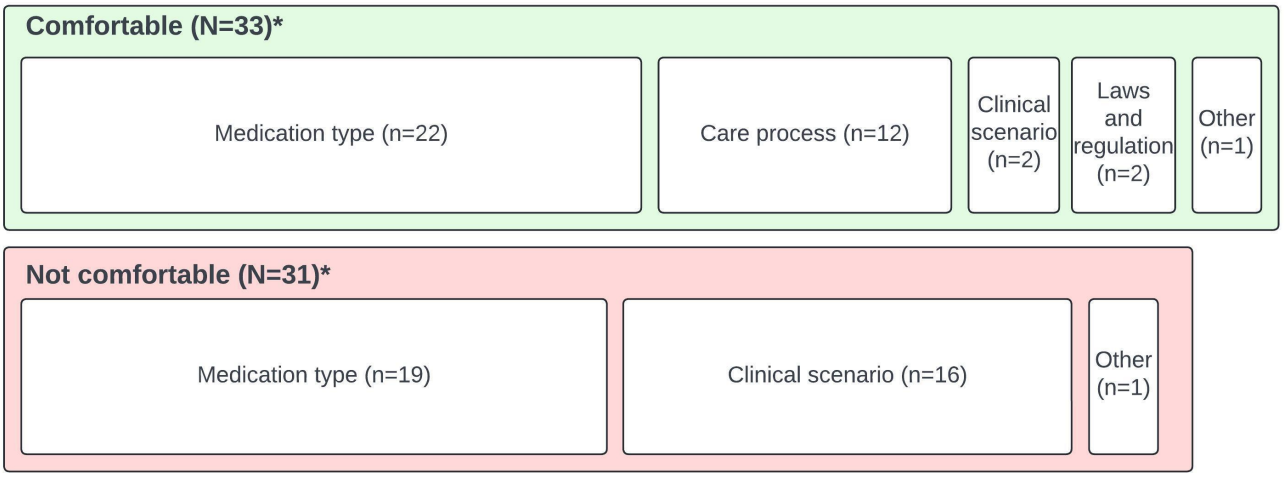
Further, 44% (51/115) of participants responded to the optional question "Tell us more about the types of medications you feel comfortable prescribing and/or monitoring via telemedicine, and the circumstances in which you feel an in-person assessment is appropriate…" We found 2 general categories of comments

XSL•FO
RenderX

related to comfort: comfortable (33 instances) and not comfortable (33 instances).

As depicted in (Figure 5), health care providers reported they felt comfortable (33 instances) prescribing certain types of medications (22 codes), prescribing with given certain care activities or processes (12 codes), in relation to laws and regulations (2 codes), in certain clinical scenarios (2 codes), and other (1 code). Further, 1 health care provider explained their requirements regarding medication types for prescribing over telemedicine: "No analgesics prescribed (no opiates, etc.) but I do prescribe stimulants for psychiatric reasons (ADHD, post-concussive pathology, as an adjunct for depression). My use of benzos is very conservative. Otherwise, I rarely prescribe anything but 'standard' psych meds–anti-depressants, mood stabilizers, non-benzo anxiolytics, at times anti-psychotics." Health care providers reported not feeling comfortable (31 instances) due to prescribing certain types of medications (19 codes), certain clinical scenarios (16 codes), and preferring in-person prescribing (1 code). Further, 1 health care provider noted how certain clinical scenarios and medication types affect their comfort: "I would only feel uncomfortable prescribing controlled substances virtually if I was [sic] concerned about other illicit substance use or elevated blood pressure/heart rate. I would want to see [the] patient in person for vital signs and UDS [urine drug screen]. The [state controlled substance database] allows us to monitor controlled prescriptions our patients have obtained." Complete results of thematic analyis for this item are provided in Multimedia Appendix 3.

Figure 5. Comments on comfort with specific types of medications and need for in-person appointments. *N refers to incidents and n refers to code counts. The total number of codes exceeds that of incidents.



Findings by Licensure and Specialty

Table 3 summarizes the results of the Mann-Whitney *U* tests to determine whether response differed according to licensure type (physician vs nonphysician) or specialty (psychiatry vs nonpsychiatry). We found that comfort in prescribing differed by licensure type for in-person prescribing (no use of telemedicine, U=1857.5; P=.005; common language effect size [CL]=0.59), with physicians indicating stronger baseline agreement that they are comfortable with in-person prescribing (mean rank of 61.92 vs a mean rank of 52.12 for nonphysicians). We also found that physicians differed from nonphysicians in perception of their ability to safely prescribe schedule V medications (U=2018.5; P=.01; CL=0.64), with visualization showing that physicians indicated greater perceived safety (mean rank 64.25 vs 48.62). There were no other statistically significant differences in the distribution of responses to statements of comfort in prescribing by licensure type.

XSL•FO
RenderX

Table . Difference in distribution of responses by licensure type and specialty (N=115).

Item	Licensure (physician or nonphysician)			Specialty (psychiatric or nonpsychiatric)		
	U(SD)	*P* value	CL[a]	*U*	*P* value	CL[a]
I am comfortable prescribing medications…						
In person (no telemedicine)	1857.5	.005[a]	0.59	1216	.98	0.50
Telemedicine	1796.5	.08	0.57	1292.5	.47	0.53
Telemedicine, prior in person	1666	.50	0.52	1215.5	.98	0.50
Telemedicine, no prior in person	1435	.34	0.45	1571	.01	0.64
Telemedicine, out-of-state	1651.5	.71	0.52	1039	.23	0.43
I can safely prescribe this type of medication via telemedicine, without seeing the patient in person:						
Schedule II	1701.5	.50	0.54	1409.5	.19	0.58
Schedule III	1562	.88	0.49	880	.02	0.36
Schedule IV	1749	.34	0.55	1487.5	.07	0.61
Schedule V	2018.5	.01	0.64	1063	.30	0.44
Unscheduled medications	1774.5	.22	0.56	1419.5	.13	0.58

[a]CL: common language effect size.

By specialty, we found that the distribution of responses for prescribers from psychiatric specialties differed from that of nonpsychiatric specialty prescribers for comfort in prescribing via telemedicine when there has been no previous in-person care (*U*=1571; *P*=.01; CL=0.64), with prescribers from psychiatric specialties indicating more comfort (mean rank 62.06 vs 45.39). Prescribers from psychiatric specialties also differed for the perceived safety of prescribing schedule III medications (*U*=880; *P*=.02; CL=0.36), with visualization showing that prescribers from psychiatric specialties indicated lower perceived safety for this type of medication (mean rank 54.11 vs 70.07). Every participant with a specialty of addiction medicine or psychiatry (n=6) indicated that they can safely prescribe schedule III medications either “always” or “almost always.” However, the number of addiction medicine or psychiatry specialists was not sufficient to support testing of group differences.

Discussion

Principal Findings

In this survey of US telemental health care providers, we examined prescribing telemental health care provider perceptions of comfort and safety in prescribing medications via telemedicine. Largely White, non-Hispanic, female, and middle-aged, the participants’ demographic characteristics reflect the known demographics of the US mental health care workforce [16]. Licensure type and medical specialty varied. Participant practice settings were primarily individual or small group, and nonacademic. Most participants were physicians, advanced practice registered nurses, or PAs, and psychiatry was the most common specialty. Geographically, participants were fairly distributed within the United States. However, there was no representation of the mountain west.

Overall, we found that telemental health care providers were very comfortable prescribing via telemedicine. However, they were slightly less comfortable when prescribing to a patient not previously seen in person, and even less comfortable prescribing to an out-of-state patient. Psychiatry specialists expressed more comfort than nonpsychiatry specialists with providing care to patients not previously seen in person. The effect size was moderate for these differences. Qualitative analysis showed that individual prescriber comfort varied depending on the context, including the clinical scenario, types of medications being prescribed, laws and regulations, and care process. This finding is consistent with prior research indicating that a health care provider’s discomfort in prescribing over telemedicine appears to be tempered by access to additional information regarding the patient and situation such as access to laboratories, in-person follow-up, and coordination of care [17,18]

We further examined the participants’ perspectives on their ability to safely prescribe unscheduled medications as well as scheduled medications, which are controlled substances. Participants indicated that they are highly comfortable prescribing unscheduled medications. However, they varied in their comfort prescribing controlled substances. While participants in this study indicated they can usually prescribe schedule II-V medications safely, a minority indicated that they can rarely or never prescribe these medications safely. Variation in comfort by DEA scheduling may relate more to prescriber familiarity and experience with those medications, than with safety issues. For example, we found that physicians felt more comfortable than nonphysicians in prescribing schedule V

XSL•FO
RenderX

medications, with a moderate effect size. Examples of schedule V medications are pregabalin (Lyrica) and diphenoxylate or atropine (Lomotil). Given most psychiatric medications fall under schedules II-IV, this could reflect a broader medical practice of physicians in comparison to other licensure types. Psychiatry specialists were less comfortable than nonpsychiatry specialists in prescribing schedule III medications with no prior in-person care. Examples of schedule III medications include buprenorphine (Suboxone), testosterone, and ketamine. It is plausible that nonpsychiatry specialists, unlikely to prescribe buprenorphine or ketamine in their practices, may have based their response on other medications within schedule III, medications with which they are more familiar.

All addiction psychiatry or medicine specialists indicated a high level of comfort in prescribing schedule III medications. It is plausible that addiction specialists have cultivated practices that enhance safety and reduce harm, increasing their sense of safety with schedule III medications. A typical addiction specialist may enhance safety through a policy of checking controlled substance databases, requiring random drug toxicology screens and securing permission to speak with patients' families. Prescribers who routinely treat addiction also often have the advantage of intersecting with integrative partners such as substance use disorder treatment centers, court appointed case managers, etc, which may enhance their sense of comfort and safety in prescribing. Addiction specialists may be better prepared to learn that a patient has diverted medications, sought additional prescribers, or knowingly taken drugs or prescriptions outside of their prescribed regimen, and could be less likely to fear these circumstances in the context of treatment.

In open-ended responses, participants described substantial tailoring of their individual prescribing practices and decision-making according to the clinical context and their comfort level in prescribing. This finding is indicative of individual adaptation of practice to perceived risks and uncertainties. Currently, best practices for prescribing via telemedicine are general in nature, and health care providers rely on local practice settings' rules and processes for guidance [19]. Individual prescribers may differ in how they approach virtual and hybrid (virtual and in-person) care, but likely adapt their practice within their comfort level through measures that include coordination of care, in-person follow-up appointments, requirements for assessments and laboratories, and compliance with laws and regulations [20]. Ultimately, our findings show that telemental health care providers prescribe over telemedicine at their discretion, in the interest of their patients' health, and according to their individual sense of comfort. Our findings also initiate that prescribers implement care activities and processes that enhance safety.

Limitations

This study used a nonprobability sampling approach; we recruited a convenience sample from a research panel of telemental health care providers. The sample size and study design do not support the generalizability of findings to all US mental health care providers. However, these findings reflect the perspectives of a large, national sample of these health care providers. TelehealthEngage panelists are users of the Doxy.me telemedicine platform which is heavily used by solo and small clinic practices, and this is reflected in the participant demographics. Perspectives may differ for health care providers who are part of large enterprise settings or academic settings. Additionally, we may not have captured nuanced considerations of prescribing safety or the conditions under which prescribers feel comfortable prescribing via telemedicine. Medications within a single controlled substance schedule are used for different purposes and have different safety considerations. For example, the use of telemedicine to prescribe ketamine is currently highly controversial, as the mental health community continues to debate the appropriateness of using a medication that induces a dissociative state without the support of an in-person therapist or guide.

We also recognize that telemedicine implementation models vary and likely influence perspectives on prescribing comfort and safety. For example, a prescriber who provides telemental health care in the context of a hub-and-spoke model, where a registered nurse or medical assistant augments virtual care with an in-person assessment, may have perceptions of greater comfort and safety in prescribing than a prescriber practicing without such a model. In ongoing work, we are conducting qualitative research that entails interviewing prescribers to more fully elucidate this information. However, this survey study allowed us to first characterize the overarching perspectives of a larger number of prescribers.

Conclusion

We conducted a national, cross-sectional survey of US telemental health providers to assess their comfort and perceived safety in prescribing medications, including controlled substances, via telemedicine. The participants included physicians, advanced practice nurses, PAs, and clinical psychologists. These health care providers were highly comfortable prescribing both scheduled and unscheduled medications via telemedicine. Comfort and perceived safety with telemedicine prescribing varied somewhat by licensure type (physician vs nonphysician) and specialty (psychiatry vs nonpsychiatry). Perceived safety varied moderately for scheduled medications (controlled substances), especially for schedule IV and V medications. Participants indicated use of adaptive strategies to prescribe safely depending upon the clinical context. In ongoing efforts, we are analyzing additional survey results and conducting qualitative research related to telemedicine prescribing. A strong understanding of prescriber perspectives and experience with telemedicine prescribing is needed to support excellent clinical practice and effective policy making in the United States.

Acknowledgments

We gratefully acknowledge Beju Shah, PharmD, MBA, for his assistance in identifying individuals to pretest the survey.

XSL•FO
RenderX

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

MC and JI conceived and designed this study, obtained ethical approval, and conducted the statistical analysis. MC developed the survey instrument, conducted the statistical analysis, and coordinated pretesting. HS and JI implemented sampling, recruitment, and participant compensation. MC, JI, BEB, EL, ZR, and BMW contributed to the literature search and interpretation of findings. MC wrote the first draft of this paper. All authors reviewed and edited this paper and approved the final version of this paper.

Conflicts of Interest

BMW is a shareholder, and MC, JI, HS, EL and BEB are employees of Doxy.me Inc, a commercial telemedicine company. The authors declare no other competing interests.

Multimedia Appendix 1
Survey of telemental health providers and informed consent form.
[PDF File, 563 KB - formative_v9i1e63251_app1.pdf]

Multimedia Appendix 2
Descriptive statistics.
[DOCX File, 18 KB - formative_v9i1e63251_app2.docx]

Multimedia Appendix 3
Results of thematic analysis of open-ended questions.
[DOCX File, 19 KB - formative_v9i1e63251_app3.docx]

References

1. Penninx B, Benros ME, Klein RS, Vinkers CH. How COVID-19 shaped mental health: from infection to pandemic effects. Nat Med 2022 Oct;28(10):2027-2037. [doi: 10.1038/s41591-022-02028-2] [Medline: 36192553]
2. Baumgart JG, Kane H, El-Hage W, et al. The early impacts of the COVID-19 pandemic on mental health facilities and psychiatric professionals. Int J Environ Res Public Health 2021 Jul 29;18(15):8034. [doi: 10.3390/ijerph18158034] [Medline: 34360334]
3. Larson EH, Patterson DG, Garberson LA, Andrilla CHA. Supply and distribution of the behavioral health workforce in rural America. WWAMI Rural Health Research Center, University of Washington. 2016. URL: https://depts.washington.edu/fammed/rhrc/wp-content/uploads/sites/4/2016/09/RHRC_DB160_Larson.pdf [accessed 2025-01-24]
4. Huskamp HA, Riedel L, Uscher-Pines L, et al. Initiating opioid use disorder medication via telemedicine during COVID-19: implications for proposed reforms to the Ryan Haight Act. J Gen Intern Med 2022 Jan;37(1):162-167. [doi: 10.1007/s11606-021-07174-w] [Medline: 34713386]
5. Lynch MJ, Vargas D, Winger ME, et al. A telemedicine bridge clinic improves access and reduces cost for opioid use disorder care. Drug Alcohol Depend Rep 2024 Jun;11:100227. [doi: 10.1016/j.dadr.2024.100227] [Medline: 38550513]
6. Zionts A. The DEA relaxed online prescribing rules during covid. now it wants to rein them in. KFF Health News. URL: https://kffhealthnews.org/news/article/dea-online-prescribing-rules-telehealth-changes/ [accessed 2025-01-24]
7. McCabe BK, Linder JA, Doctor JN, et al. The protocol of improving safe antibiotic prescribing in telehealth: a randomized trial. Contemp Clin Trials 2022 Aug;119:106834. [doi: 10.1016/j.cct.2022.106834] [Medline: 35724841]
8. Fathy R, Briker S, Rodriguez O, Barbieri JS. Comparing antibiotic prescription rates between in-person and telemedicine visits. J Am Acad Dermatol 2022 Aug;87(2):438-440. [doi: 10.1016/j.jaad.2021.08.064] [Medline: 34499992]
9. Patel R, Irving J, Brinn A, et al. Impact of the COVID-19 pandemic on remote mental healthcare and prescribing in psychiatry: an electronic health record study. BMJ Open 2021 Mar;11(3):e046365. [doi: 10.1136/bmjopen-2020-046365]
10. Hailu R, Mehrotra A, Huskamp HA, Busch AB, Barnett ML. Telemedicine use and quality of opioid use disorder treatment in the US during the COVID-19 pandemic. JAMA Netw Open 2023 Jan 3;6(1):e2252381. [doi: 10.1001/jamanetworkopen.2022.52381] [Medline: 36692880]
11. Mapping supply of the US psychiatric workforce. Health Workforce Technical Assistance Center. 2018 Oct. URL: https://www.healthworkforceta.org/research-alerts/mapping-supply-of-the-us-psychiatric-workforce/ [accessed 2025-02-6]
12. Clarke V, Braun V. Thematic analysis. J Posit Psychol 2017 May 4;12(3):297-298. [doi: 10.1080/17439760.2016.1262613]
13. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
14. Bing Maps. 2025. URL: https://www.bingmapsportal.com/ [accessed 2025-05-05]

XSL•FO
RenderX

15. Microsoft bing maps platform apis terms of use. Microsoft. 2025. URL: https://www.microsoft.com/en-us/maps/bing-maps/product [accessed 2025-05-05]
16. Data tool: demographics of the U.S psychology workforce. American Psychological Association. 2024. URL: https://www.apa.org/workforce/data-tools/demographics [accessed 2025-01-24]
17. Uscher-Pines L, Sousa J, Raja P, Mehrotra A, Barnett M, Huskamp HA. Treatment of opioid use disorder during COVID-19: experiences of clinicians transitioning to telemedicine. J Subst Abuse Treat 2020 Nov;118:108124. [doi: 10.1016/j.jsat.2020.108124] [Medline: 32893047]
18. Uscher-Pines L, Parks AM, Sousa J, et al. Appropriateness of telemedicine versus in-person care: a qualitative exploration of psychiatrists' decision making. Psychiatr Serv 2022 Aug 1;73(8):849-855. [doi: 10.1176/appi.ps.202100519] [Medline: 35080419]
19. Smith K, Ostinelli E, Macdonald O, Cipriani A. COVID-19 and telepsychiatry: development of evidence-based guidance for clinicians. JMIR Ment Health 2020 Aug 28;7(8):e21108. [doi: 10.2196/21108] [Medline: 32658857]
20. Gough F, Budhrani S, Cohn E, et al. ATA practice guidelines for live, on-demand primary and urgent care. Telemed J E Health 2015 Mar;21(3):233-241. [doi: 10.1089/tmj.2015.0008] [Medline: 25658882]

ABBREVIATIONS

BRANY: Biomedical Research Alliance of New York
CL: common language effect size
DEA: Drug Enforcement Agency
PA: physician assistant

Edited by A Mavragani; submitted 14.06.24; peer-reviewed by MH Pourasad, V Chavarria; revised version received 20.12.24; accepted 21.12.24; published 11.03.25.

Please cite as:
Cummins MR, Ivanova J, Soni H, Robbins Z, Bunnell BE, López E, Welch BM
Telemedicine Prescribing by US Mental Health Care Providers: National Cross-Sectional Survey
JMIR Form Res 2025;9:e63251
URL: https://formative.jmir.org/2025/1/e63251
doi:10.2196/63251

© Mollie R Cummins, Julia Ivanova, Hiral Soni, Zoe Robbins, Brian E Bunnell, Esteban López, Brandon M Welch. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Supplementing Consent for a Prospective Longitudinal Cohort Study of Infants With Antenatal Opioid Exposure: Development and Assessment of a Digital Tool

Jamie E Newman[1], MPH, PhD; Leslie Clarke[2], MS, MBA, RN; Pranav Athimuthu[1], BA; Megan Dhawan[3], MSN, CRNP; Sharon Owen[4], RN; Traci Beiersdorfer[5], RN; Lindsay M Parlberg[1], MPH; Ananta Bangdiwala[1], MS; Taya McMillan[1], MPH; Sara B DeMauro[3], MSCE, MD; Scott Lorch[3], MSCE, MD; Myriam Peralta-Carcelen[4], MPH, MD; Deanne Wilson-Costello[2], MD; Namasivayam Ambalavanan[4], MD; Stephanie L Merhar[5], MS, MD; Brenda Poindexter[6], MD; Catherine Limperopoulos[7], PhD; Jonathan M Davis[8], MD; Michele Walsh[9], MD; Carla M Bann[1], PhD

[1]RTI International, 3040 Cornwallis Road, Research Triangle Park, NC, United States

[2]Case Western Reserve University, Cleveland, OH, United States

[3]Children's Hospital of Philadelphia, Philadelphia, PA, United States

[4]University of Alabama at Birmingham, Birmingham, AL, United States

[5]University of Cincinnati College of Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

[6]Emory University, Atlanta, GA, United States

[7]Children's National Hospital, Washington, DC, United States

[8]Tufts Medical Center, Boston, MA, United States

[9]National Institute of Child Health and Human Development, Rockville, MD, United States

Corresponding Author:
Jamie E Newman, MPH, PhD
RTI International, 3040 Cornwallis Road, Research Triangle Park, NC, United States

Abstract

Background: The Outcomes of Babies With Opioid Exposure (OBOE) study is an observational cohort study examining the impact of antenatal opioid exposure on outcomes from birth to 2 years of age. COVID-19 social distancing measures presented challenges to research coordinators discussing the study at length with potential participants during the birth hospitalization, which impacted recruitment, particularly among caregivers of unexposed (control) infants. In response, the OBOE study developed a digital tool (consenter video) to supplement the informed consent process, make it more engaging, and foster greater identification with the research procedures among potential participants.

Objective: We aim to examine knowledge of the study, experiences with the consent process, and perceptions of the consenter video among potential participants of the OBOE study.

Methods: Analyses included 129 caregivers who were given the option to view the consenter video as a supplement to the consent process. Participants selected from 3 racially and ethnically diverse avatars to guide them through the 11-minute video with recorded voice-overs. After viewing the consenter video, participants completed a short survey to assess their knowledge of the study, experiences with the consent process, and perceptions of the tool, regardless of their decision to enroll in the main study. Chi-square tests were used to assess differences between caregivers of opioid-exposed and unexposed infants in survey responses and whether caregivers who selected avatars consistent with their racial or ethnic background were more likely to enroll in the study than those who selected avatars that were not consistent with their background.

Results: Participants demonstrated good understanding of the information presented, with 95% (n=123) correctly identifying the study purpose and 88% (n=112) correctly indicating that their infant would not be exposed to radiation during the magnetic resonance imaging. Nearly all indicated they were provided “just the right amount of information” (n=123, 98%) and that they understood the consent information well enough to decide whether to enroll (n=125, 97%). Survey responses were similar between caregivers of opioid-exposed infants and unexposed infants on all items except the decision to enroll. Those in the opioid-exposed group were more likely to enroll in the main study compared to the unexposed group (n=49, 89% vs n=38, 51%; $P<.001$). Of 81 caregivers with known race or ethnicity, 35 (43%) chose avatars to guide them through the video that matched their background. Caregivers selecting avatars consistent with their racial or ethnic background were more likely to enroll in the main study (n=29, 83% vs n=43, 57%; $P=.01$).

XSL•FO
RenderX

Conclusions: This interactive digital tool was helpful in informing prospective participants about the study. The consenter tool enhanced the informed consent process, reinforced why caregivers of unexposed infants were being approached, and was particularly helpful as a resource for families to understand magnetic resonance imaging procedures.

Trial Registration: ClinicalTrials.gov NCT04149509; https://clinicaltrials.gov/study/NCT04149509

(JMIR Form Res 2025;9:e59954) doi:10.2196/59954

KEYWORDS

informed consent digital tool; avatars; video-assisted consent; MRI; antenatal opioid exposure; infant; antenatal; opioid exposure; caregiver; survey; magnetic resonance imaging; Outcomes of Babies With Opioid Exposure

Introduction

The importance of the informed consent process in today's clinical trials stems from ethical standards detailed in landmark documents such as the Nuremberg Code, the Declaration of Helsinki, and the Belmont Report. Recent research suggests that there is room for improvement in the consent processes to better inform prospective participants of study aims and potential risks. Numerous studies have suggested that minority populations, those with low levels of education, and those with intellectual disabilities are among the groups in which improvements could be made to heighten participants' understanding of the informed consent process (see Tam et al [1] for a review). Some studies have observed that fewer than half of the participants were able to correctly answer questions about the study's purpose after reviewing an informed consent form [2,3]. In the review by Tam et al [1], only 55% could name a risk associated with a study, with even fewer being able to define key methodological concepts such as the use of placebo or randomization.

Informed consent materials should be easily comprehended by the average research participant. Grade-level reading standards are the most common metric for measuring the readability of consent documents. It is generally agreed that material should target 6th- to 8th-grade level standards to be considered widely interpretable. Yet, a 2003 study exploring informed consent materials furnished by research institutions across the United States found that only 8% of their sample met their own institution's standards for comprehension, with the average readability score of a consent document exceeding stated standards by almost 3 grade levels [4]. If prospective participants cannot understand the structure and content of a consent form, they may decline to participate. If a study using English language materials involves nonnative English speakers, minority populations, or foreign immigrants, the reading level must be further simplified [5].

In response to criticisms of written consent forms, recent studies have harnessed technology to create consent procedures that simplify the information being conveyed while better engaging prospective participants. While some approaches have used electronic forms, others have developed videos, audio files, and multimedia tutorials to supplement the consent process.

Electronic consent tools found their earliest uses in medical procedures, and most evaluations of their effectiveness continue to reside in that domain [6,7]. Studies evaluating electronic consent have focused on comprehension [8], engagement [8-10], satisfaction and procedure-related anxiety reduction [11], recall of risks and benefits [4], and the potential to streamline the consent process by reducing time spent by staff explaining study procedures [7]. There are mixed findings on the effectiveness of video-assisted consent compared with written consent forms [9]. This may be due, in part, to technology advances over time. Videos created 20 years ago would have been limited in the graphics capacity of procedures they could display, whereas modern tools have the capacity to include 3D animation and guided walk-throughs. Multiple studies that have successfully demonstrated the effectiveness of video-assisted consent involve minority populations, study participants from rural areas, populations with English as a second language, older adults, and those with intellectual disabilities [12-15].

Certain capacities show particular promise at recruiting people traditionally underrepresented in medical research and vulnerable populations, warranting consideration in future studies. For example, Lawrence et al [16] describe an electronic consent framework using REDCap (Research Electronic Data Capture; Vanderbilt University) and customizable avatars. Videos provided participants with a choice of virtual assistants voiced by community members that guided them through the consent process. Participants had the ability to select their avatar from a variety of backgrounds, which were developed in consultation with targeted communities. Participants "emphasized the importance of representing minority populations [and] depicting supportive and helpful interactions with medical staff," as well as their strong preference for seeing procedures depicted using avatars [16]. Participants reported being inspired by the avatars and being able to see themselves in the avatars, which have translated into stronger intervention outcomes for the treatment group [17].

Digitally assisted consent also offers potential advantages in other research domains for informing participants of study procedures and potential risks such as neuroimaging methods (eg, magnetic resonance imaging [MRI]). Although MRI is largely a low-risk, noninvasive procedure that substitutes magnetic fields for high-energy radiation, misconceptions about its use abound, including false beliefs that MRI exposes participants to radiation [18]. To our knowledge, no studies of electronic consent tools have assessed the potential of video-assisted consent to help communicate risks associated with MRI procedures. The few studies involving neuroimaging have focused on computed tomography and other radiation methods that do not carry as many misconceptions [13]. Thus, our research fills a gap in this literature by assessing whether using a video supplement to visually depict MRI scanning

XSL•FO
RenderX

procedures and associated risks can adequately inform potential participants and dispel common MRI misperceptions in the Outcomes of Babies With Opioid Exposure (OBOE) study. It also builds on past literature on consent procedures, by including ethnically diverse avatars to describe MRI procedures.

Social distancing measures to mitigate the spread of COVID-19 presented new challenges to research coordinators talking at length with caregivers during the birth hospitalization about the OBOE study and building rapport through these face-to-face interactions. To mitigate these challenges, we developed the OBOE study consenter video to supplement the informed consent process and to make it more engaging for potential study participants, especially those from underrepresented populations. We developed this interactive digital tool to inform prospective participants about the study, reinforce why caregivers of unexposed (control) infants were being approached, provide a resource for families to understand the MRI procedures, and describe the longitudinal nature of the study while detailing visits across the 2-year study period. To enhance participant diversity, we developed 3 racially and ethnically diverse avatars, which participants could choose to guide them through the video. If potential participants were able to see themselves in the avatars, this may foster greater identification with the research procedures. This paper provides an overview of the digital tool and examines knowledge of the study, experiences with the consent process, and perceptions of the video among potential participants in the OBOE study.

Methods

OBOE Study Overview

The Advancing Clinical Trials in Neonatal Opioid Withdrawal OBOE study is a multisite prospective longitudinal cohort study of infants with antenatal opioid exposure and unexposed infants from birth to 2 years of age (ClinicalTrials.gov NCT04149509). The OBOE study protocol is described elsewhere in detail [19]. Briefly, the observational study aims to determine the impact of antenatal opioid exposure on brain development and neurodevelopmental outcomes over the first 2 years of life and to explore whether family, home, and community factors modify developmental trajectories. The study consists of 3 MRI visits at 0 - 1 month, 6 months, and 24 months; 1 home visit at 12 months; 1 telephone interview at 18 months; and developmental testing at 12 and 24 months. The OBOE study addresses a growing health concern—opioid use and misuse—among a particularly vulnerable population: pregnant individuals and their infants. Overdose has become one of the leading causes of maternal mortality [20], mirroring the rise in misuse among pregnant individuals and the greater population throughout the opioid epidemic.

Digital Tool Development: The Consenter Video

The first infant was enrolled into the OBOE study in August 2020, and the OBOE study began developing the consenter video a year later to assist with the slower-than-anticipated enrollment and to enhance participant diversity. The consenter tool is a patient-centered, interactive digital tool for informing prospective participants about clinical research studies and enhancing informed consent processes [21,22]. It transforms the informed consent process from the passive receipt of complex medical and scientific information to an active, engaging visual experience that can improve individuals' understanding of clinical research processes and result in a more informed, dedicated research participant [23]. Five main steps were involved in developing the OBOE study consenter tool. First, a detailed storyboard was developed, which included verbatim audio text and a written description of proposed visuals. The audio text was primarily pulled from the single institutional review board (IRB)–approved consent form. Next, the team created illustrations and visuals to describe the study and assessments. Third, a group of digital designers developed customized avatars. Fourth, the team selected 3 actors to record voice-overs that corresponded to the race and ethnicity of each avatar. Finally, the video production team combined the multimedia elements into a single video file. The development process was guided by a communication scientist to ensure that evidence-based communication practices were incorporated. Throughout each step in the process, staff at the clinical sites, including research coordinators, nurses, physicians, and MRI staff, reviewed the products, and their feedback was incorporated. All study team members attended a virtual consenter orientation held before implementation. The OBOE study launched the IRB-approved consenter tool in June 2022, at which time approximately half of the target sample had been enrolled.

Ethical Considerations

Human Subject Ethics Review Approvals

Through a single IRB at Cincinnati Children's Hospital Medical Center, all 4 clinical sites (Cincinnati Children's Hospital Medical Center, University of Alabama at Birmingham, Children's Hospital of Pennsylvania, and Case Western Reserve University), the Neuroimaging Core at Children's National Hospital, and the Data Coordinating Center at the RTI (Research Triangle Institute) received approval for human subjects research activities for the OBOE study, and informed consent was obtained for all participants. The STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) reporting guidelines for observational studies were followed [24].

Informed Consent

Caregivers of opioid-exposed infants and unexposed infants were asked to view the consenter video in addition to talking to coordinators about the study as well as reading and signing the informed consent form. Caregivers of infants born at or after 37 weeks gestation were recruited at the 4 clinical sites primarily during hospital stays following birth and, to a lesser extent, prenatal clinics. Caregivers of opioid-exposed infants were approached if their infant was exposed to opioids in the second or third trimester of pregnancy. Caregivers of unexposed (control) infants were approached if there was no antenatal drug exposure as determined by maternal history or maternal urine toxicology screen at delivery. OBOE study exclusion criteria for both groups were confirmed at delivery and are detailed elsewhere [19].

XSL•FO
RenderX

Privacy and Confidentiality

Caregivers approached for the OBOE study were assigned a participant ID number, and they chose from one of three racially and ethnically diverse avatars to guide them through the 11-minute video with recorded voiceovers and illustrations of study procedures (Figure 1). After viewing the video, caregivers were asked to complete a brief survey using their participant ID, regardless of their decision to participate in the main study. The survey assessed their knowledge of the study, experiences with the consent process, and perception of the video.

Figure 1. Consenter avatars that caregivers choose to guide them through the video as a supplement to the informed consent form.

Compensation Details

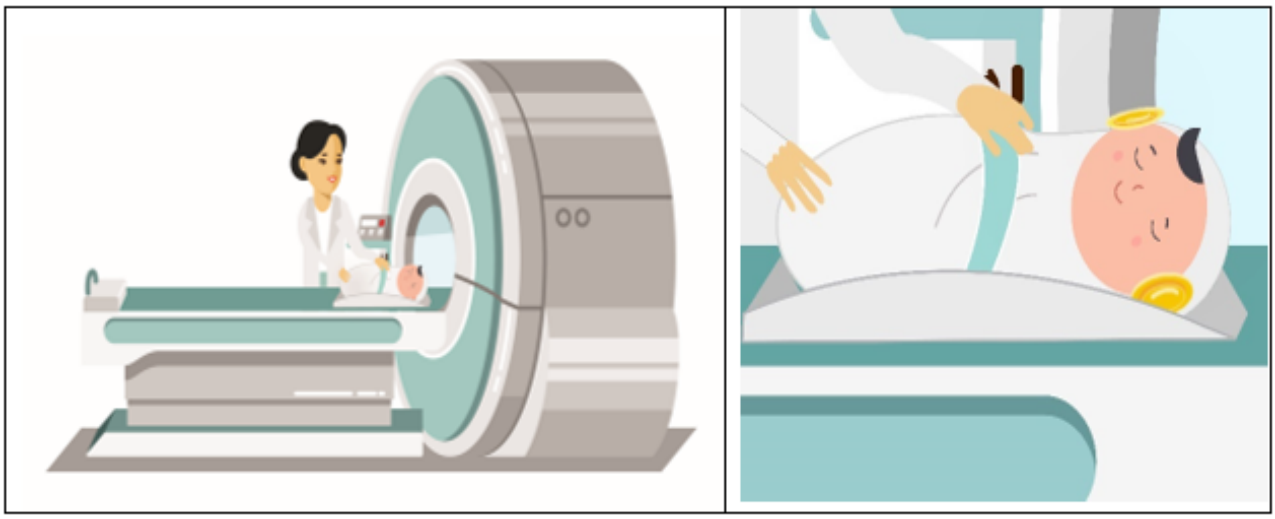
Participants were given a US $10 gift card for completing the survey.

Digital Tool Overview

To ensure ease of access to the consenter tool, participants were provided 3 options for viewing the video. These included: (1) an iPad with the video preloaded onto an app, which the participants used alongside a coordinator; (2) a coordinator lanyard card with a QR code that participants could scan to watch the consenter video on their personal mobile device; and (3) business cards with a QR code and web link directing to the video so that participants could access the video asynchronously away from the study site.

The interactive video described the study purpose, detailed the risks and benefits of joining the study, explained that participation is voluntary, and depicted MRI procedures via a verbal description and illustrations of a staff member preparing a swaddled, sleeping baby with protective earmuffs to enter the MRI scanner (Figure 2). The avatar explained that MRI scans are safe and do not expose the infant to any radiation, dispelling a common misperception of MRI.

Figure 2. Consenter video screenshots illustrating MRI procedures at the 0 - 1 month OBOE study visit. MRI: magnetic resonance imaging; OBOE: Outcomes of Babies With Opioid Exposure.



The consenter tool detailed the longitudinal nature of the study and described assessments at each visit time point. Figure 3 illustrates developmental testing at the 24-month visit.

XSL•FO
RenderX

Figure 3. Consenter video screenshot illustrating developmental testing at the 24-month OBOE study visit. OBOE: Outcomes of Babies With Opioid Exposure.

Statistical Analysis

All analyses were performed using SAS (version 9.4; SAS Institute). Chi-square tests were used to examine (1) whether there were differences between opioid-exposed and unexposed groups in survey responses and (2) whether participants who selected avatars consistent with their racial or ethnic background were more likely to report that they decided to enroll in the study than those who selected avatars that were not consistent with their background.

Results

Overview

Analyses included 129 caregivers (55 caregivers of opioid-exposed infants and 74 caregivers of unexposed infants) who were given the option to view the consenter video as a supplement to the informed consent process and take the survey.

Survey Responses

Participants demonstrated a good understanding of the information presented in the video (Table 1). Overall, 95% (n=123) correctly identified the purpose of the OBOE study. The majority also correctly indicated that the research team would collect the baby's umbilical cord (n=120, 93%), the infant would not be exposed to radiation during the MRI scan (n=112, 88%), and the study would not include a blood test (n=105, 83%). Nearly all indicated they were provided "just the right amount of information" (n=123, 98%) and understood the consent information well enough to decide whether to enroll (n=125, 97%).

XSL•FO
RenderX

Table . Survey responses of caregivers of opioid-exposed infants and unexposed infants after watching consenter video (N=129).

Item	All (n=129), n (%)	Opioid exposed (n=55), n (%)	Unexposed (n=74), n (%)
Knowledge of study			
1. Correctly identified the purpose of the OBOE[a] study	123 (95)	53 (96)	70 (95)
2. Correctly indicated, "If I enroll in the study, the research team will collect a piece of my baby's umbilical cord at delivery that would otherwise be thrown away."	120 (93)	50 (91)	70 (95)
3. Correctly indicated that infant would not be exposed to radiation during the MRI[b] scan	112 (88)	48 (87)	64 (88)
4. Which of the following is not one of the study activities:			
a. Three brain MRIs over the first 2 years of my baby's life	4 (3)	3 (5)	1 (1)
b. A 12-month home visit	9 (7)	6 (11)	3 (4)
c. A blood test (correct response)	105 (83)	44 (80)	61 (86)
d. Parent or caregiver surveys	8 (6)	2 (4)	6 (8)
Experience with consent process			
5. During the consent process, do you feel you were provided:			
a. Too much information	0 (0)	0 (0)	0 (0)
b. Just the right amount of information	123 (98)	53 (98)	70 (97)
c. Too little information	3 (2)	1 (2)	2 (3)
6. Felt you understood the consent information well enough to decide whether you and your baby should be in the study	125 (97)	55 (100)	70 (95)
Perceptions of consenter video			
7. Did you watch the video describing the study?			
a. Yes	124 (96)	51 (93)	73 (99)
b. No	5 (4)	4 (7)	1 (1)
7a. If yes, what did you like most about the video[c]?			
a. The explanation of why the study is important	62 (51)	27 (53)	35 (49)
b. The explanation of the study procedures	28 (23)	12 (24)	16 (23)
c. The explanation of the MRI	19 (16)	6 (12)	13 (18)
d. The explanation of how my personal information would be protected	7 (6)	4 (8)	3 (4)
e. I did not like watching the video	2 (2)	1 (2)	1 (1)
f. Other	4 (3)	1 (2)	3 (4)
7b. If yes, what did you not like about the video?			
a. It was too long	5 (4)	1 (2)	4 (6)
b. It was hard to understand	0 (0)	0 (0)	0 (0)

XSL•FO
RenderX

Item	All (n=129), n (%)	Opioid exposed (n=55), n (%)	Unexposed (n=74), n (%)
c. It did not cover everything I needed to know about the study	0 (0)	0 (0)	0 (0)
d. I had a hard time getting the video to play	10 (8)	3 (6)	7 (10)
e. I liked watching the video	93 (79)	41 (82)	52 (76)
f. Other	10 (8)	5 (10)	5 (7)
Study enrollment intentions			
8. Did you decide to enroll in the OBOE study[d]?			
a. Yes	87 (67)	49 (89)	38 (51)
b. No	25 (19)	1 (2)	24 (32)
c. Not yet decided	17 (13)	5 (9)	12 (16)
8a. If you decided not to enroll or have not yet decided, what was the primary reason?			
a. Too much time or effort to complete the study visits or surveys	3 (8)	1 (17)	2 (6)
b. I do not trust that my information will be kept safe	2 (5)	0 (0)	2 (6)
c. I do not want to disclose any information, even if my name is not attached to it	0 (0)	0 (0)	0 (0)
d. I am not interested in participating in any research	5 (13)	0 (0)	5 (15)
e. I live too far away from the hospital or follow-up clinic	5 (13)	0 (0)	5 (15)
f. Financial burden to come back for the visits	1 (3)	0 (0)	1 (3)
g. I do not want my baby to have a brain MRI	6 (15)	0 (0)	6 (18)
h. Other	18 (45)	5 (83)	13 (38)

[a]OBOE: Outcomes of Babies With Opioid Exposure.

[b]MRI: magnetic resonance imaging.

[c]Percentages do not add to 100% because 3 respondents selected multiple responses.

[d]$P<.001$.

Survey responses did not differ significantly between caregivers of opioid-exposed infants and those of unexposed infants on all items except decision to enroll. Those in the opioid-exposed group were significantly more likely to report that they decided to enroll in the study compared to the unexposed group (n=49, 89% vs n=38, 51%; $P<.001$).

Avatar Selection

Of 129 respondents who completed the consenter video survey, 1 (1%) was Asian, 1 (1%) was Black and Hispanic, 29 (22%) were Black and non-Hispanic, 69 (53%) White and non-Hispanic, and 29 (22%) were missing race. Among 81 respondents of known race who watched the video, 35 (43%) chose avatars to guide them through the video that matched their background. Participants who selected avatars consistent with their racial or ethnic background were more likely to report that they decided to enroll in the study (n= 29, 83% vs n=43, 57%; $P=.01$).

Other Findings

An unanticipated benefit of the consenter video was its usefulness in informing new caregivers about the study when there were custody changes. As the longitudinal study progressed, coordinators indicated that the consenter video was helpful in providing information to new caregivers (eg, foster parents) about the study. We only asked caregivers to take the consenter survey at their infant's enrollment into the study, so we do not have data to assess what new caregivers thought about the video in cases of custody changes.

Discussion

Principal Findings

Social distancing measures to mitigate the spread of COVID-19 placed limitations on face-to-face interactions between study staff and participants. The OBOE study addressed this challenge

XSL•FO
RenderX

by developing an avatar-based consenter tool to supplement the consent process by creating an interactive video to explain study procedures while engaging a diverse cohort of study participants, particularly those from underrepresented populations who are most likely to have difficulty with traditional consent materials. Of the 129 caregivers who viewed the consenter video and took the survey, almost all (n=123, 95%) were able to correctly identify the purpose of the study, and nearly all felt that the video provided "just the right amount" of information and understood the consent information well enough to decide whether to enroll.

Participants' most frequently cited strengths of the video were its ability to convey the importance of the study and its ability to explain the MRI and study procedures. This is an important finding given the lack of existing research assessing video-assisted consent for neuroimaging studies. In addition, when asked what they did not like about the video, 79% (n=93) of participants indicated they in fact liked watching the video, suggesting they were engaged in viewing the video-assisted consent. Compared with written consent forms, animation and visuals and such as those used in the consenter tool can better engage participants, potentially resulting in more informed research participants [22-24].

Participants who indicated that they did not like the consenter video cited technical issues (n=10, 8%) and length (n=5, 4%) as the primary reasons. Duration is a key element that affects engagement if the video length does not correlate with the complexity of the information conveyed. Researchers must try to avoid including excessive details, because doing so may result in informational overload thus minimizing the benefits of using a video [25,26]. Additionally, study staff must be adequately equipped to handle the technological needs associated with using digital tools. The OBOE study addressed the need for flexible viewing options by providing 3 ways for participants to view the video.

Given the potential for race-matched avatars to facilitate connections between participants and the study, we anticipated that participants would tend to choose the avatar that best matched their group identification. However, only 43% of participants whose race was known and who completed the consenter survey chose an avatar that aligned with their racial background. It is possible that participants chose avatars for reasons aside from self-representation, such as perceptions of avatars' attractiveness or how authentic they felt an avatar looked. Lee [27] found that Black and female participants of a virtual video game were more likely to choose Black and female avatars (respectively) when assured that they were entering a high-diversity virtual world with other minority participants, but without this assurance, many Black participants chose White avatars. Lee [27] interpreted this finding using a social identity theory framework [28], explaining that virtual avatars can be used to facilitate self-expression and feelings of belonging, or they can create "escapes" for minority participants by allowing them to camouflage themselves as a member of the majority group. When virtual worlds are framed as high in diversity, participants tend toward the former, projecting their identity onto their avatar and customizing it accordingly. However, without this framing, participants may default to using ingrained cultural lenses and may "pass" under White avatars. Zimmermann et al [29], using the theories of transformed social interaction [30] and impression management [31], explain this "passing" behavior as people's attempt to present an idealized representation of themselves in a given context.

Researchers must make active efforts to assure participants, through study framing and multimedia tool design, that the world they will enter as study participants is one that accepts them, includes them, and actively recognizes their autonomy to choose and consent, regardless of their background.

Limitations

Specific activities that supported the launch of the consenter tool included receiving robust feedback from a diverse set of clinical site staff on the video illustrations, avatar images, and voice-overs to ensure the video was culturally relevant and met the needs of the caregivers approached for the study. However, a limitation is that we did not explicitly obtain feedback during video development from those with lived experience, such as caregivers of opioid-exposed infants. For example, Naeim et al [32] describe the implementation of electronic video consent in their precision health research after the formation of a community advisory board consisting of respected individuals from the participants' home community who could speak of the communities' perspectives. They solicited regular feedback from the community advisory board, ensuring that the video language was suitable for a lay population and spoke effectively to the study's racially diverse sample.

Conclusions

This interactive digital tool was helpful in informing prospective participants about the study and reinforcing why caregivers of unexposed infants were being approached. The video was a helpful resource for families to understand the study's MRI procedures, because it provided a verbal description and included illustrations of a staff member preparing a swaddled, sleeping baby with protective earmuffs to enter the MRI scanner.

Acknowledgments

The authors acknowledge the caregivers who agreed to take part in this study; the medical and nursing colleagues at Case Western Reserve University, Cincinnati Children's Hospital Medical Center, University of Alabama at Birmingham, and Children's Hospital of Philadelphia; the Neuroimaging Core colleagues at Children's National Hospital; the data coordinating center staff at RTI (Research Triangle Institute) International; and National Institute of Child Health and Human Development colleagues Dr Andrew Bremer, MAS, MD, PhD, Dr Nahida Chakhtoura, MD, and Ms Stephanie Archer, MA. This research was supported by the National Institutes of Health (NIH) through the NIH Helping to End Addiction Long-Term (HEAL) Initiative (awards

XSL•FO
RenderX

1PL1HD101059-01, 3PL1HD101059-01S2). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or its NIH HEAL Initiative.

Data Availability

The dataset analyzed during this study is not currently publicly available because the OBOE study is ongoing. Following OBOE study completion, a public use dataset will be deposited in the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) Data and Specimen Hub (DASH).

Conflicts of Interest

None declared.

References

1. Tam NT, Huy NT, Thoa LTB, et al. Participants' understanding of informed consent in clinical trials over three decades: systematic review and meta-analysis. Bull World Health Organ 2015 Mar 1;93(3):186-98H. [doi: 10.2471/BLT.14.141390] [Medline: 25883410]
2. Eiser C, Davies H, Jenney M, Glaser A. Mothers' attitudes to the randomized controlled trial (RCT): the case of acute lymphoblastic leukaemia (ALL) in children. Child Care Health Dev 2005 Sep;31(5):517-523. [doi: 10.1111/j.1365-2214.2005.00538.x] [Medline: 16101646]
3. Manafa O, Lindegger G, Ijsselmuiden C. Informed consent in an antiretroviral trial in Nigeria. Indian J Med Ethics 2007;4(1):26-30. [doi: 10.20529/IJME.2007.009] [Medline: 18630217]
4. Paasche-Orlow MK, Taylor HA, Brancati FL. Readability standards for informed-consent forms as compared with actual readability. N Engl J Med 2003 Feb 20;348(8):721-726. [doi: 10.1056/NEJMsa021212] [Medline: 12594317]
5. Palazzani L, Macioce F, Daverio M, Ferro V, Persampieri L. New strategies for increasing participation of patients from diverse cultural and religious backgrounds in clinical trials. BioLaw Journal - Rivista di BioDiritto 2019;1(15):101-122. [doi: 10.15168/2284-4503-403]
6. Rossi M, McClellan R, Chou L, Davis K. Informed consent for ankle fracture surgery: patient comprehension of verbal and videotaped information. Foot Ankle Int 2004 Oct;25(10):756-762. [doi: 10.1177/107110070402501011] [Medline: 15566709]
7. Shqaidef AJ, Saleh MYN, Ismail FK, Abu-Awad M, Khambay BS. A comparative assessment of information recall and comprehension between conventional leaflets and an animated video in adolescent patients undergoing fixed orthodontic treatment: a single-center, randomized controlled trial. Am J Orthod Dentofacial Orthop 2021 Jul;160(1):11-18. [doi: 10.1016/j.ajodo.2020.03.028] [Medline: 33902979]
8. Hall EW, Sanchez TH, Stein AD, et al. Use of videos improves informed consent comprehension in web-based surveys among internet-using men who have sex with men: a randomized controlled trial. J Med Internet Res 2017 Mar 6;19(3):e64. [doi: 10.2196/jmir.6710] [Medline: 28264794]
9. Glaser J, Nouri S, Fernandez A, et al. Interventions to improve patient comprehension in informed consent for medical and surgical procedures: an updated systematic review. Med Decis Making 2020 Feb;40(2):119-143. [doi: 10.1177/0272989X19896348] [Medline: 31948345]
10. Schenker Y, Fernandez A, Sudore R, Schillinger D. Interventions to improve patient comprehension in informed consent for medical and surgical procedures: a systematic review. Med Decis Making 2011;31(1):151-173. [doi: 10.1177/0272989X10364247] [Medline: 20357225]
11. Gesualdo F, Daverio M, Palazzani L, et al. Digital tools in the informed consent process: a systematic review. BMC Med Ethics 2021 Feb 27;22(1):18. [doi: 10.1186/s12910-021-00585-8] [Medline: 33639926]
12. Biesecker B, Raspa M, Rupert D, Moultrie R, Furberg R, McCormack LA. Making Clinical Trials More Patient-Centered Using Digital Interactive E-Consent Tools: RTI Press; 2019. [doi: 10.3768/rtipress.2019.op.0063.1910]
13. Cowan EA, Calderon Y, Gennis P, Macklin R, Ortiz C, Wall SP. Spanish and English video-assisted informed consent for intravenous contrast administration in the emergency department: a randomized controlled trial. Ann Emerg Med 2007 Feb;49(2):221-230. [doi: 10.1016/j.annemergmed.2006.07.934] [Medline: 17011074]
14. Fanaroff AC, Li S, Webb LE, et al. An observational study of the association of video- versus text-based informed consent with multicenter trial enrollment: lessons from the PALM study (patient and provider assessment of lipid management). Circ Cardiovasc Qual Outcomes 2018 Apr;11(4):e004675. [doi: 10.1161/CIRCOUTCOMES.118.004675] [Medline: 29625993]
15. McCormack LA, Wylie A, Moultrie R, et al. Supporting informed clinical trial decisions: results from a randomized controlled trial evaluating a digital decision support tool for those with intellectual disability. PLoS ONE 2019;14(10):e0223801. [doi: 10.1371/journal.pone.0223801] [Medline: 31644588]
16. Lawrence CE, Dunkel L, McEver M, et al. A REDCap-based model for electronic consent (eConsent): moving toward a more personalized consent. J Clin Transl Sci 2020 Apr 3;4(4):345-353. [doi: 10.1017/cts.2020.30] [Medline: 33244416]

XSL•FO
RenderX

17. Heffner JL, Watson NL, Serfozo E, et al. An avatar-led digital smoking cessation program for sexual and gender minority young adults: intervention development and results of a single-arm pilot trial. JMIR Form Res 2021 Jul 30;5(7):e30241. [doi: 10.2196/30241] [Medline: 34328430]
18. Hadskis M, Kenny N, Downie J, Schmidt M, D'Arcy R. The therapeutic misconception: a threat to valid parental consent for pediatric neuroimaging research. Account Res 2008;15(3):133-151. [doi: 10.1080/08989620801946917] [Medline: 18792535]
19. Bann CM, Newman JE, Poindexter B, et al. Outcomes of Babies with Opioid Exposure (OBOE): protocol of a prospective longitudinal cohort study. Pediatr Res 2023 May;93(6):1772-1779. [doi: 10.1038/s41390-022-02279-2] [Medline: 36042329]
20. Campbell J, Matoff-Stepp S, Velez ML, Cox HH, Laughon K. Pregnancy-associated deaths from homicide, suicide, and drug overdose: review of research and the intersection with intimate partner violence. J Womens Health (Larchmt) 2021 Feb;30(2):236-244. [doi: 10.1089/jwh.2020.8875] [Medline: 33295844]
21. Furberg RD, Ortiz AM, Moultrie RR, et al. A digital decision support tool to enhance decisional capacity for clinical trial consent: design and development. JMIR Res Protoc 2018 Jun 6;7(6):e10525. [doi: 10.2196/10525] [Medline: 29875084]
22. Furberg RD, Raspa M, Wheeler AC, McCormack LA, Bailey DB. A digital health app to assess decisional capacity to provide informed consent: protocol for a randomized controlled trial. JMIR Res Protoc 2018 Nov 19;7(11):e10360. [doi: 10.2196/10360] [Medline: 30455171]
23. Raspa M, Fitzgerald T, Furberg RD, et al. Mobile technology use and skills among individuals with fragile X syndrome: implications for healthcare decision making. J Intellect Disabil Res 2018 Oct;62(10):821-832. [doi: 10.1111/jir.12537] [Medline: 30105880]
24. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008 Apr;61(4):344-349. [doi: 10.1016/j.jclinepi.2007.11.008]
25. Bester J, Cole CM, Kodish E. The limits of informed consent for an overwhelmed patient: clinicians' role in protecting patients and preventing overwhelm. AMA J Ethics 2016 Sep 1;18(9):869-886. [doi: 10.1001/journalofethics.2016.18.9.peer2-1609] [Medline: 27669132]
26. Mohammed-Ali AI, Gebremeskel EI, Yenshu E, et al. Informed consent in a tuberculosis genetic study in Cameroon: information overload, situational vulnerability and diagnostic misconception. Res Ethics 2022 Oct;18(4):265-280. [doi: 10.1177/17470161221106674]
27. Lee JER. Does virtual diversity matter?: Effects of avatar-based diversity representation on willingness to express offline racial identity and avatar customization. Comput Human Behav 2014 Jul;36:190-197. [doi: 10.1016/j.chb.2014.03.040]
28. Tajfel H, Turner JC. An integrative theory of intergroup conflict. In: Hogg MA, Abrams D, editors. Intergroup Relations: Essential Readings: Psychology Press; 1979:1-4 URL: https://alnap.org/help-library/resources/an-integrative-theory-of-intergroup-conflict/ [accessed 2025-02-19]
29. Zimmermann D, Wehler A, Kaspar K. Self-representation through avatars in digital environments. Curr Psychol 2023 Sep;42(25):21775-21789. [doi: 10.1007/s12144-022-03232-6]
30. Bailenson JN, Yee N, Blascovich J, Beall AC, Lundblad N, Jin M. The use of immersive virtual reality in the learning sciences: digital transformations of teachers, students, and social context. J Learn Sci 2008 Feb 15;17(1):102-141. [doi: 10.1080/10508400701793141]
31. Goffman E. The Presentation of Self in Everyday Life: Bantam; 1959.
32. Naeim A, Dry S, Elashoff D, et al. Electronic video consent to power precision health research: a pilot cohort study. JMIR Form Res 2021 Sep 8;5(9):e29123. [doi: 10.2196/29123] [Medline: 34313247]

Abbreviations

IRB: institutional review board
MRI: magnetic resonance imaging
OBOE: Outcomes of Babies With Opioid Exposure
REDCap: Research Electronic Data Capture
RTI: Research Triangle Institute
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

XSL•FO
RenderX

Edited by A Mavragani; submitted 26.04.24; peer-reviewed by L Narbey; revised version received 22.01.25; accepted 22.01.25; published 04.03.25.

Please cite as:
Newman JE, Clarke L, Athimuthu P, Dhawan M, Owen S, Beiersdorfer T, Parlberg LM, Bangdiwala A, McMillan T, DeMauro SB, Lorch S, Peralta-Carcelen M, Wilson-Costello D, Ambalavanan N, Merhar SL, Poindexter B, Limperopoulos C, Davis JM, Walsh M, Bann CM
Supplementing Consent for a Prospective Longitudinal Cohort Study of Infants With Antenatal Opioid Exposure: Development and Assessment of a Digital Tool
JMIR Form Res 2025;9:e59954
URL: https://formative.jmir.org/2025/1/e59954
doi:10.2196/59954

© Jamie E Newman, Leslie Clarke, Pranav Athimuthu, Megan Dhawan, Sharon Owen, Traci Beiersdorfer, Lindsay M Parlberg, Ananta Bangdiwala, Taya McMillan, Sara B DeMauro, Scott Lorch, Myriam Peralta-Carcelen, Deanne Wilson-Costello, Namasivayam Ambalavanan, Stephanie L Merhar, Brenda Poindexter, Catherine Limperopoulos, Jonathan M Davis, Michele Walsh, Carla M Bann. Originally published in JMIR Formative Research (https://formative.jmir.org), 4.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Digital Pornography Education Prototype Co-Designed With Young People: Formative Evaluation

Jake Turvey[1], MPH; Michelle Raggatt[1,2], BBiomedSci (Hons); Cassandra J C Wright[1,3], PhD; Angela C Davis[1,2], PhD; Meredith J Temple-Smith[4], DHSc; Megan S C Lim[1,2,5], PhD

[1]Young People's Health Research, Burnet Institute, 85 Commercial Rd, Melbourne, Australia
[2]School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia
[3]Alcohol and other Drugs team, Menzies School of Health Research, Casuarina, Australia
[4]Department of General Practice and Primary Care, University of Melbourne, Melbourne, Australia
[5]Melbourne School of Population and Global Health, University of Melbourne, Parkville, Australia

Corresponding Author:
Megan S C Lim, PhD
Young People's Health Research, Burnet Institute, 85 Commercial Rd, Melbourne, Australia

Abstract

Background: Interventions to help young people make sense of sex and relationships in the context of widely available pornography are becoming increasingly supported in school settings. However, young people who experience disruptions to their education often have less access to such programs. Digital platforms may offer a more accessible method to deliver tailored sexual health and pornography literacy to young people who are disengaged from mainstream schooling, or who experience other types of structural disadvantage.

Objective: This study aimed to describe the formative evaluation of "The Gist" a co-designed online sexual health education and pornography literacy prototype designed to meet the sexual health information needs of structurally marginalized young people in Australia.

Methods: We conducted iterative workshops with 33 young people aged between 15 and 24 years recruited from an alternative education school in Melbourne, Australia. Through interactive activities, participants evaluated the overall prototype design, including its usability, desirability, inclusiveness, and potential for impact.

Results: Participants reported The Gist to be easy to use (17/20, 85%) and safe (19/23, 83%), with "hot" branding (25/30, 83%). However, perceived content relevance was dependent on the participants' existing level of sexual health knowledge and experience, with only 31% (7/23) agreeing that "The Gist feels like it was made for me." The interactive learning activities such as the debunked (myth-busting) and quiz features were among the most used and well-liked on The Gist platform. Low unprompted engagement with the prototype outside of facilitated workshop settings also confirmed previous researcher postulations that The Gist as a standalone digital platform is unlikely to meet the needs of this population group. Further design refinements are needed to improve user experience, including more interactive activities and visual information in place of heavily text-based features.

Conclusions: This study provides important insights into the design and sexual health information needs of structurally marginalized young people. Further research is needed to assess the overall efficacy of The Gist prototype, as well as its ability to positively influence young people's sexual attitudes, beliefs, and behaviors. Future iterations should consider hybrid or face-to-face delivery models to better capture student engagement.

(JMIR Form Res 2025;9:e65859) doi:10.2196/65859

KEYWORDS

pornography; education; website; prototyping; evaluation; sexual health; sexual wellbeing; pornography literacy; young people; youth; adolescents; The Gist; sexual education; Australia; efficacy; digital health; co-design

Introduction

Pornography and Pornography Literacy

Pornography is easily accessible to, and commonly viewed by young people. In a 2017 cross-sectional survey of 941 Australians aged 15 - 29 years, 84% of males and 23% of females viewed pornography at least weekly [1]. Some young people report that pornography is an important component of their sexual development [2], helping to define understandings of sexual pleasure, how to have sex, and diverse sexual experiences and sexualities [3]. However, pornographic content also commonly includes behavior which portrays problematic themes, including the promotion of heteronormative and

XSL•FO
RenderX

misogynistic gender norms, violence-supportive sexual objectification of women, lack of consent, and the poor representation of diverse people and body types [4]. Research has further shown that pornography can shape young people's sexual development in line with these problematic themes, and negatively influence their developing attitudes, beliefs, and sexual health behaviors [2,5,6].

Pornography literacy education is a technique which aims to reduce the adverse impacts of pornography use. Originating from media literacy theory, pornography literacy programs seek to increase young people's knowledge and understanding of healthy relationships and sexualities, whilst developing their ability to critique, evaluate, and challenge the various themes commonly present within pornographic content [7]. However, there remains a scarcity of formally evaluated pornography literacy interventions [8-10]. In a 2018 pilot study conducted in the United States, it was observed that students demonstrated improvements in various pornography-related knowledge items after participating in a college-based pornography literacy class [11]. Similarly, a longitudinal study of Dutch teenagers found that the relationship between pornography use and sexist attitudes became weaker the more a participant engaged in school-based pornography literacy classes [12]. These findings therefore suggest that there is potential for pornography literacy programs to shape the way in which young people perceive common underlying themes and messages within online pornographic content [12]. While pornography literacy programs show promise in mitigating the adverse impacts of pornography on young people, determining how best to integrate and deliver these programs within existing educational frameworks is crucial for their success.

Relationships and Sexuality Education

School-based sex education initiatives have traditionally been the preferred approach to improving sexual health outcomes amongst young people and mitigating risks of sexual health related harms such as sexually transmitted infections (STIs) and unplanned pregnancy [13]. In Australia, the delivery of relationships and sexuality education (RSE) has evolved in recent decades to include a broader range of contemporary sexual health topics and themes [14]. However, creating programs that adequately cater to the complex sexual health needs of diverse young people has posed an ongoing challenge for Australian educators and policymakers alike [15]. Furthermore, the actual education received by students is highly varied nationwide, as the development and implementation of RSE curricula remains the responsibility of each individual state or territory, while content delivery often differs between individual schools [14]. In a 2018 national survey of 6327 Australian high school students, more than 60% of respondents described their RSE experience as either irrelevant or marginally relevant to their sexual health needs, with many advocating for a greater focus on sexual communication, relationships, and pleasure [16]. The Australian government introduced mandatory consent education from 2023, covering areas of sexual consent, gender stereotypes, and power and relationship dynamics. However, there have been growing calls for the further addition other contemporary topics (such as pornography literacy) that take a more sex positive approach, and are more inclusive of alternative sexualities, gender identities, and other intersectionalities [17-19].

Furthermore, school-based education may not meet the needs of all young people. Young people have many different intersecting experiences that mean they have more or less opportunities for healthy sexual education and development. Young people who are disengaged from, or experience disruptions to their mainstream education or for example, have less access to school-based sex education programs [20]. In addition, school-based sex education is often not tailored to meet the specific needs of young people who are lesbian, gay, bisexual, transgender, queer, questioning, intersex, asexual (LGBTQIA+), young people from diverse cultural backgrounds, or young people with experiences of poverty [21-23]. International research also indicates that young people with experiences of trauma, family conflict, and other adverse childhood experiences have poorer access to sexual health information compared with other mainstream cohorts, and have an increased likelihood of engaging in risky sexual health behaviors [24]. Interventions to support young people's sexual development in the context of pornography therefore, need to account for this substantial variation in sexual health development among young people, and must consider the specific needs of individuals who have been excluded from current relationships and sexuality education.

Digital platforms have shown increasing promise as an acceptable way to reach and deliver tailored sexual health information to young people [25]. Recent systematic reviews have highlighted their effectiveness at shaping young people's knowledge of healthy sexual development, and reducing risky sexual health behaviors [25,26]. However, many of these interventions remain highly focused on singular or complementary behavioural outcomes in adolescents, such as STI prevention and contraception use [26]. As such, there is a distinct need for an online sexual health platform that is not only informed by, and tailored to the specific needs of structurally marginalized young people, but also provides them with a more comprehensive education on a wide range of sexual health topics, including pornography literacy.

Aims

The Gist is an online relationships- and sexuality-education platform, incorporating elements of pornography literacy, that was co-designed by young people, sexual health researchers, educators, and web developers [4]. Here, we describe a formative evaluation of The Gist prototype, an evaluation designed to pretest a program with the target group before implementation [27]. This evaluation aims to assess The Gist's usability, desirability, inclusiveness, and potential for impact, as well as identifying areas for future improvement.

Methods

The Intervention

Development of The Gist used a human-centred and co-design methodology [4]. Several months before this study, young people who identified as LGBTQIA+, were culturally and linguistically diverse, had experiences of family conflict or

XSL•FO
RenderX

educational disengagement participated in a series of iterative workshops to provide in-depth insights and design ideas for a sexual health education resource. Based on their insights and design recommendations, a minimally functioning digital prototype was developed. This is a working website that demonstrated most of the planned features of The Gist with some of the planned content. This study aimed to test and further refine the prototype design. To give context to this evaluation, the key design considerations and features of The Gist are briefly described below. More in-depth information about the methodology and findings from the design and development phase is available in a separate publication [4].

To give context to this evaluation, the key design considerations and features of The Gist are briefly described below. More in-depth information about the methodology and findings from the design and development phase is available in a separate publication [4].

The Gist prototype was designed as an online website application, with mobile-friendly capacity. Topics covered included pornography, body diversity, sex, consent, pleasure, healthy relationships, and sexual preferences. Language and imagery remained gender- and sexuality-neutral. Critical pornography literacy was built into this content by contrasting the healthy and ethical information about sex, relationships, and bodies with problematic messages often espoused in pornographic materials. The visual design of the digital prototype had an alternative, “retro 90s” and comic-book style as preferred by participants from the earlier co-design stage.

The website prototype presented short facts about various sexual health topics to engage, surprise and challenge the user. Users were then linked to more in-depth information about the topic, and encouraged to engage with interactive activities that further challenged their knowledge and built critical thinking skills around the respective topic. Key platform features included:

1. Get Facts: a series of rolling screens presented a short fact with an image, designed to reproduce the “stories” function used on social media apps such as Instagram and Snapchat.
2. Debunked*:* presented users with a short statement on a series of sexual health topics and prompted users to decide whether the statement was true or false. This feature was designed to challenge commonly held misconceptions and beliefs about certain sexual health topics.
3. Quizzes: brief interactive quizzes presented through multiple-choice questions designed to test the user’s knowledge of key sexual health themes.
4. Sexual Health Articles: provided in-depth written information on various sexual health topics broken down into digestible subsections.

Evaluation Questions

This paper reports on the testing and evaluation phase of the human-centered design process, which is part of an iterative cycle of prototyping, testing, and refining. The insights from this evaluation phase will be used to further refine and develop The Gist prototype for improved application. The key evaluation questions were: (1) is The Gist usable and accessible for young people?, (2) is The Gist desirable and relevant to young people?, (3) is The Gist inclusive and safe for young people?, and (4) do young people believe that The Gist has potential to positively impact other young people’s knowledge, attitudes, and confidence around respectful and healthy relationships, sexual health and wellbeing, and pornography literacy?

Recruitment and Setting

Evaluation participants were recruited from across 3 campuses of an alternative education school in Melbourne, Australia. Students at this school have experiences of fragmented school attendance or disengagement from mainstream school settings for a range of reasons including family conflict or breakdown, bullying, being gender and sexually diverse, living with a disability, or being culturally and linguistically diverse. Researchers first spoke to school staff, providing them with an overview of the study and demonstration of The Gist prototype so that teachers were informed of the content that would be shown to students, and could voice any concerns about certain topics that some students may find distressing. After feedback and strong support from staff was received, the research team then spoke to students, and provided a brief overview of the study, workshop activity and content themes, before distributing plain language participant information and consent forms to those interested.

Participants

A total of 33 young people participated in at least 1 of 7 workshops held across a 3-week period. Of the 33 participants; 15 identified as female, 13 identified as male and 1 identified as nonbinary (4 participants elected not to provide demographic information). Participants’ ages ranged from 15 - 24 years with a median age of 19 years. A total of 18 participants identified as heterosexual, 11 identified as LGBTQIA+, and 8 spoke a language other than English or in addition to English at home.

Workshop Methodology

In total, 3 groups of students across 2 campuses attended a series of 2 workshops each. Only 1 workshop was carried out at the third campus as these students had previously participated in the initial co-design research process for The Gist earlier in the year and were therefore already familiar with The Gist prototype.

All workshops were facilitated by 3 members of the research team, including 1 digital designer and 2 public health researchers. Two of these facilitators had been involved in the development of the prototype, all identified as young women (age<35 years old). School staff were present in 1 workshop (teacher supervision was required due to an incident earlier in the day unrelated to the research project). The duration of the workshops ranged from 70 - 135 minutes. An overview of the workshop aims is available in Multimedia Appendix 1. Workshop activities were specifically designed to be engaging, interactive, and trauma-informed, due to the unique learning and classroom needs of the participants. A total of 10 different activities were used across the various workshops, with a descriptive overview of each workshop activity available in Multimedia Appendix 2.

XSL•FO
RenderX

Data Collection

A range of qualitative and quantitative data was collected from across the 7 workshops.

Qualitative Data

Qualitative data included audio recordings of workshop activities and classroom discussions and physical material produced by students during workshop activities (ie, images of whiteboard brainstorming and completed worksheet activities). One researcher recorded written observations on classroom activities, levels of student engagement, and overheard classroom discussions. They used a semistructured note-taking guide developed by the lead researcher, to supplement audio recording with observations of nonverbal interactions. In addition, while students explored the prototype, 1 participant in each group (n=3) volunteered to be individually observed to collect more detailed information on website navigation and interactions. The facilitator recorded their actions in detail including which buttons were pressed and which features were accessed.

One researcher recorded written observations on classroom activities, levels of student engagement, and overheard classroom discussions. They used a semi-structured note-taking guide developed by the lead researcher, to supplement audio recording with observations of non-verbal interactions.

In addition, while students explored the prototype, one participant in each group (n=3) volunteered to be individually observed to collect more detailed information on website navigation and interactions. The facilitator recorded their actions in detail including which buttons were pressed and which features were accessed.

Quantitative Data

Quantitative data included surveys conducted during workshops explored students' receptiveness to The Gist prototype, including measures on usability and accessibility, desirability and relevance, safety and inclusivity, and potential to impact (Multimedia Appendix 3). A five-point Likert scale was used to assess survey responses, with response options ranging from 1= "Nah Shit," 2="Pfft," 3="Meeh," 4="It's OK," 5="Love it!!!." Students used stickers to indicate their responses on a large poster displaying the questions and website analytical data including unique page visits.

Website analytical data including unique page visits.

Data Analysis

This research was embedded in the constructivist paradigm. Following initial data cleaning and familiarisation, author [MR] developed a thematic framework guided by the evaluation questions, as well as emerging themes from initial data familiarisation processes. This framework was then applied to the data by authors [MTS and MR] by coding the observer notes, audio recordings, and images directly into NVivo 12 software (Lumivero). Researchers then summarized emerging themes into charts and tables to visualize the data as a whole, before mapping and synthesising differences across different participant groups and workshops. Throughout each phase, the thematic framework was refined as new themes or subthemes emerged.

Simultaneously, the quantitative data sourced from workshop activities and the workshop surveys were analysed by author [MR] using descriptive statistics. Then, a narrative integration approach was used to weave both qualitative and quantitative findings together on a theme-by-theme basis by authors [JT, MR, CW, and ML]. To achieve this, the quantitative findings were mapped to the thematic framework, and used to support or expand upon the emerging study themes, providing a stronger and more comprehensive understanding of the themes. Findings and key themes are arranged according to the evaluation questions. All supporting quotes are presented anonymously.

Ethical Considerations

Ethics approval for the research was granted by the Alfred Health Human Research Ethics Committee (373/18). All participants provided written informed consent. For participants under the aged of 18 years or younger, written or verbal parental consent confirmed via school staff was sought before workshop commencement. All data have been deidentified. Participants received a AUD20 (USD 13.90) voucher per workshop, plus a AUD10 (USD6.95) voucher if attending both workshops. Participants who were already familiar with the prototype and were only invited to a single workshop received a AUD25 (USD17.40) voucher.

Results

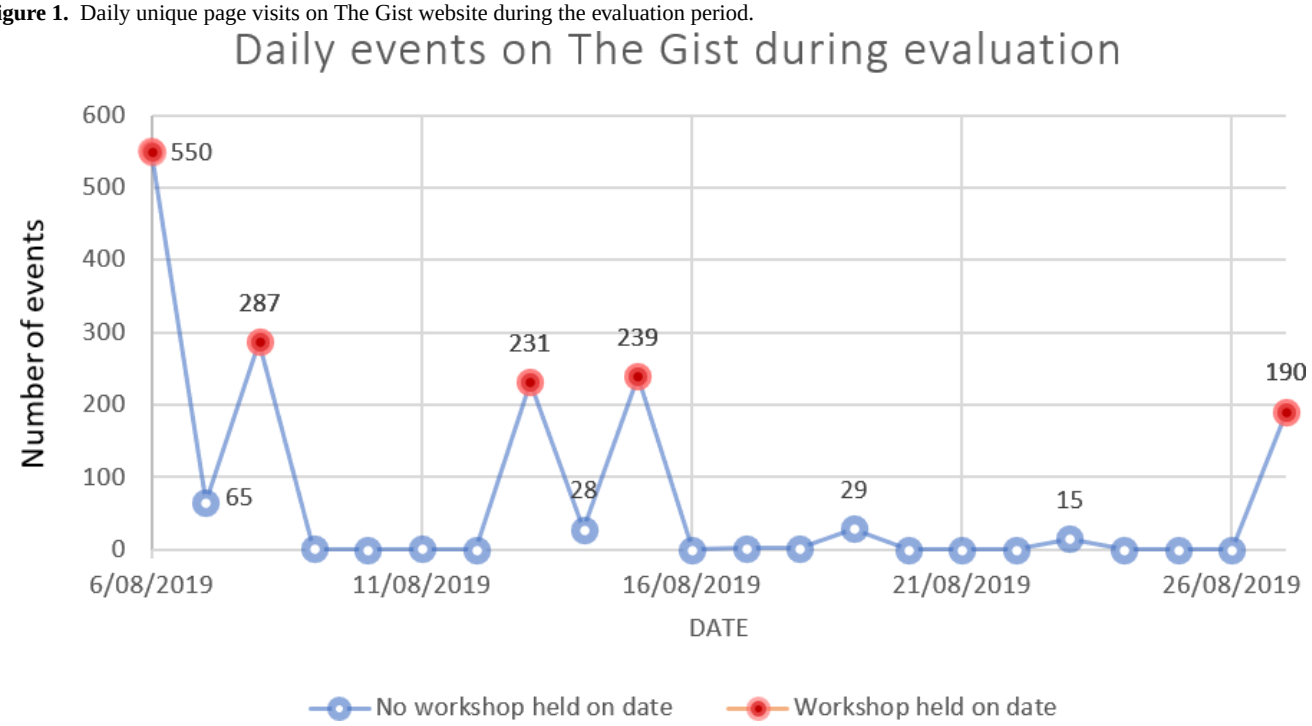
Usable and Accessible

Use and Usability

In general, participants were able to access and engage with The Gist website effectively, with 85% of workshop participants (17/20) responding with either "Love it!!!" or "It's OK" when asked if "The Gist is easy to use." However, website analytical data suggests that actual engagement with The Gist outside of the facilitated workshops was extremely low. Of the 1641 unique page visits recorded across the evaluation period, 1497 (91%) were recorded during workshop sessions (Figure 1), suggesting few participants independently engaged with the website.

XSL•FO
RenderX

Figure 1. Daily unique page visits on The Gist website during the evaluation period.



Researcher observations also noted that students primarily engaged with The Gist in small groups during workshop sessions, with many participants preferring to explore the website features alongside their peers, rather than as an individual endeavor. Students seemed to value the shared discussions which were stimulated by The Gist content and engaged each other in a collective learning process.

Non-Intuitive Design Features

Participants also reported design issues that meant several of The Gist features were not as intuitive or easy to use as participants would have liked. The Get Facts feature for example, was not clearly identifiable as content, with several participants initially attempting to use it as a navigation function. The ability to filter content within the search function, was not obvious to many participants who would either scroll past this option or requested clarification from facilitators.

> *It's all organised but I wouldn't really know how to go around using it such as what the app offers. But after a while at using it I would get used to it*

Further Information and Learning

Participants were receptive to the different interactive activities available. Over the evaluation period, the most popular and well used features were "Debunked" (302 events), "Quizzes" (127 events), and "Get Facts" (100 events). The home page received 240 visits and "Articles" were viewed 84 times.

> *I first checked "Get Facts" tab section. Informative, liked the flicking facts.*
>
> *When you click something and then you learn something, it's not complicated. The Get Facts [feature] works*

Articles were generally the least used learning feature, owing largely to their length and heavy use of text over interactive content and images, despite our attempts to prioritise brevity. Several participants commented that they did not enjoy reading, and quickly abandoned the feature or skipped over the articles altogether, preferring to access information through more interactive means. This is despite feedback from participants suggesting that the content did not go far enough with the information presented, with participants wanting the opportunity to learn more about specific topics that interested them. However, amongst the participants who reviewed the article feature, the majority (17/21, 81%) enjoyed (ie, "it's OK" or "Love it!!!") the format of the articles which used subheadings to break down relevant information, while 95% (20/21) thought the articles were easy to read. While a dot point summary of each article was provided, this was placed at the end of the article, with facilitators once again observing many participants either abandoning the article before reaching the summary, or simply scrolling past.

> *For me, because I don't like reading, I would say the articles weren't very important to me. But that's because I hate reading.*
>
> *I really love the [article] layout and the expandable boxes – the word I'd use is nifty.*
>
> *The articles were easy to read, [and] I did learn some cool things about consent.*

Desirable and Relevant

Content Relevance

Overall, participants held generally positive opinions of The Gist and its content, with 68% (15/22) of workshop participants agreeing with the statement, "The Gist is definitely something I would use." More than 73% (16/22) of participants also agreed that they "learnt some cool stuff from the Gist." Many participants reflected that The Gist was both helpful and informative, with the features of the website serving as effective

XSL•FO
RenderX

learning strategies to deliver new and engaging sexual health information.

> *Yes [relevant] because some of these things happen in my life*

Participants particularly enjoyed how the content was specifically catered toward young people, and sought to openly discuss topics that were frequently absent from within other mainstream sexual health education programs.

> *I really liked how The Gist is better than any sex education you get in school*
>
> *It's more appropriate than other sites because it's made for young people that don't understand sex ed at school or just want to learn.*
>
> *…get info about things that are not typically spoken about*
>
> *I like that it was open and uncensored*

However, while most participants felt that The Gist was relevant to young people in general, several participants reflected that the prototype was not overly relevant to them personally. Two opposing explanations were offered for this, with some participants suggesting that The Gist's content was more appropriate for a younger audience, with quizzes being "too easy," and that they already knew much of the information presented. In contrast, several other participants reflected that much of the content was too advanced for them, especially as they had little previous knowledge or sexual experience. They explained that there was not enough information for "beginners" who were not comfortable learning about sex-based topics. Similar apprehensions were expressed by a small number of students who identified as asexual, or those who had never engaged in watching pornographic material.

> *I probably won't use it that much because the info is stuff I mostly know.*
>
> *More information for younger/less experienced.*

Survey results supported this mixed response. When asked if "The Gist feels like it was made for me," 6 participants (26%) responded negatively ("Nah shit" or "Pfft"), 7 participants (31%) responded positively ("It's OK" or "Love it!!!"), and the largest proportion of students (10/22, 43%) remained neutral ("Meeh"). While many students did not think the content was personally relevant, most remained interested in participating and providing feedback on the prototype itself.

Aesthetically Appealing

Participants also gave feedback on the design and branding of The Gist, with the overall look and feel of the application appealing well to young people who took part in the evaluation. In total, 83% of participants (25/30) responded either "It's ok" or "I love it!!!" to the statement "The Gist branding is hot!" signifying significant approval of the website aesthetic and imagery.

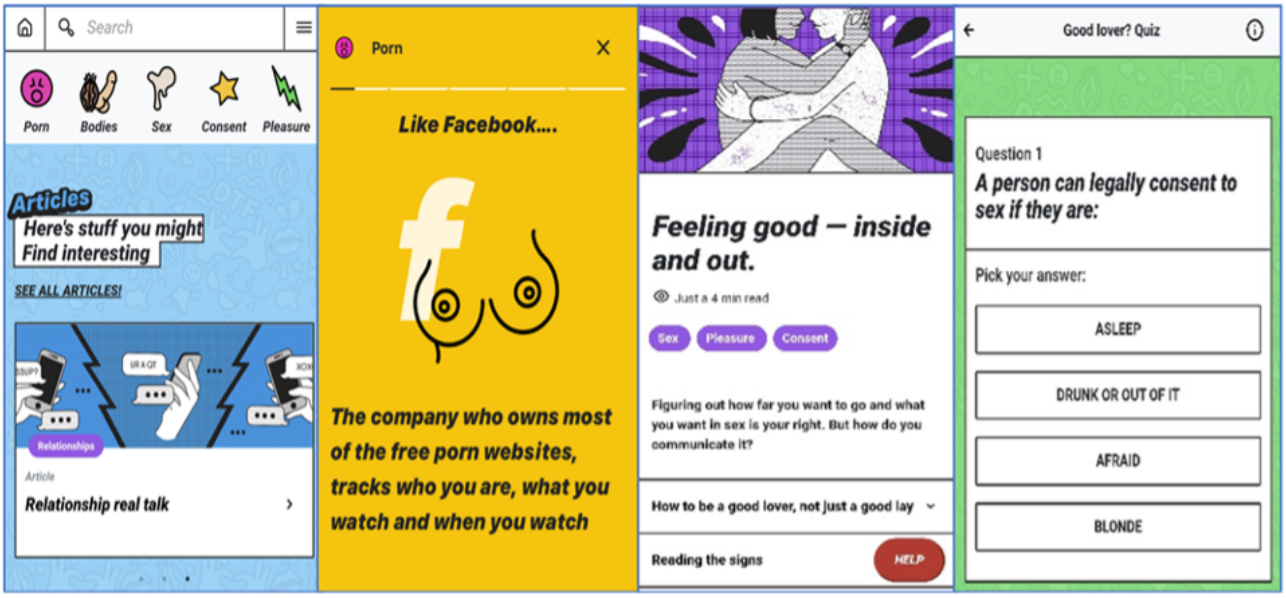
Participants also indicated that they enjoyed the use of bright colours and specifically approved of the retro-90s styling which simultaneously appealed to a wide range of young people and created a nostalgic and safe feeling whilst using the site.

> *Comic book aesthetic is something that everyone can enjoy*
>
> *I really enjoy the colours and art style: it makes me feel safe and nostalgic somehow*

For some students, the visual depiction of genitalia, sex, and pornography (Figure 2) was seen as unnecessary and excessive at times, and may have worked to dissuade apprehensive students from engaging with the website.

> *Some of it is too much, this one that indicates cum in the Get Facts section, it's not necessary*

Figure 2. Features of The Gist, a digital prototype for pornography education. Panels show screenshots of the home page, Get Facts, article,and quiz.



XSL•FO
RenderX

Safe and Inclusive

User Safety

When asked about how safe they felt using The Gist website, the majority (19/23, 83%) of participants responded either "It's ok" or "I love it!!!." Participants particularly enjoyed the ability to browse The Gist anonymously, especially due to the highly sensitive nature of The Gist's content and associated sexual health topics.

> *Safe place to get information that you might not want to ask someone about*
>
> *I like that its anonymous, you don't need an account to log in*

Participants were also asked if they would feel comfortable accessing The Gist website in public spaces such as on a train or tram, with a slight majority (12/22, 54%) responding with either "It's OK" or "I love it!!!." Those who felt less safe using The Gist expressed that these sentiments were largely driven by feelings of being unsafe on the internet more generally due to concerns of data monitoring and internet privacy rather than factors specifically related to The Gist website.

Inclusion of LGBTQIA+ Themes

In general, the content and language used throughout The Gist was seen as inclusive, particularly by those participants who identified as LGBTQIA+. Participants appreciated the nongendered wording used across the site, and the inclusive approach to sex and pornography use amongst gender and sexually diverse people.

> *It's not gender specific which is good*

Several participants, however, did express a desire to see greater inclusion of LGBTQIA+ themes, and for content to appear more obviously inclusive of sexual and gender diverse young people. This included having a dedicated LGBTQIA+ filter or tab section with relevant content rather than just having inclusive language and content spread throughout the website.

> *Make it more trans/non-binary (like a language filter). Have LGBTQIA+ filter e.g.: I click trans + all the content is trans-based*
>
> *fully inclusive – could be more obvious. Have LGBTQIA+ section*

Differences in Cultural Perspectives Toward Sex and Pornography

Discussions with participants who were recent migrants or from culturally and linguistically diverse backgrounds suggested that cultural understandings played a significant role in how they experienced The Gist website and its related content. Some young people indicated that in their culture or religion, acts of intimacy and sex were typically seen as only appropriate within marriage, and therefore influenced their view on the necessity of a sexual health education and pornographic literacy platform.

Several other participants were willing to discuss issues surrounding healthy relationships and consent, but were less comfortable discussing sexual behaviors and pornography. Even though some of these participants felt overwhelmed at times during the first workshop, all participants returned to the second workshop and appeared to be comfortable accessing The Gist and fully participating in the remaining activities. This highlights the importance of giving participants' time to develop rapport and to "warm up" to discussing sensitive topics during co-design processes.

> *Feel uncomfortable when quiz answers mention sex in relationships*
>
> *Confronting images*

Potential of The Gist to Positively Impact Other Young People

Participants were asked whether they believed The Gist had the potential to impact other young people's knowledge, attitudes, and confidence around healthy relationships, sexual health issues, and pornographic literacy. Most participants perceived that it would be an effective sexual health resource for other young students.

> *I like it and think it would be helpful and informative for teens and young adults.*
>
> *…it is a web site that helps young people no matter the gender/ethnicity to understand more about erotic matters.*

However, once again participants reflected that the ability of The Gist to impact other young people's sexual health and pornographic literacy was dependent on its relevance to the user, with greater efforts needed to provide appropriate information to different ages and knowledge levels.

> *It depends on what kinda age group that you want, I think more detail would be good if you want older people.*
>
> *I wish that it appealed to an older audience as well as a younger one*

When asked if "The Gist will make people want to watch porn," 63% of participants (14/22) disagreed with the statement, and only 1/22 agreed ("probs"). Participants stated that The Gist was more likely to reduce the stigma associated with watching pornography, and therefore makes young people feel less "scared" about watching pornographic material.

> *Using the website was easy to read & understand the gist will not make you want to watch porn it will instead give you an insight what porn really is!*
>
> *I think it might make people feel comfortable not scared to watch porn*

Discussion

Principal Findings

The Gist prototype was developed through a multistage co-design process, which includes this evaluation as part of the iteration and refinement stage. A key criticism made by participants was that much of the Gist's content and messaging was too text-heavy, particularly when providing users with in-depth or supplementary information on key sexual health topics. Feedback from participants emphasised this as a barrier to engagement, with many expressing their dislike of the articles

XSL•FO
RenderX

feature and other text-based elements of the website. This is despite this feature being intentionally designed based on previous research of young people's user experiences and online interaction patterns, as well as feedback from the initial prototype co-design stage, in which young people supported the progressive disclosure of supplementary text information [4]. This highlights the importance of a multistage co-design and prototype refinement process, with further work needed to optimize the Gist's articles for enhanced user engagement and acceptability. In contrast, the interactive learning activities such as the "Debunked" (myth-busting) and "quiz" features were amongst the most used and well-liked on The Gist platform. This affinity for highly interactive, image-based activities is unsurprising, given the well documented preference for kinaesthetic and visual learning styles amongst this cohort [28]. Additional efforts are therefore needed to further adapt key supplementary health information on The Gist website to more appropriate styles, including interactive activities, image-based content, and audio-visual materials. However, consideration must also be given to the resources associated with producing and maintaining such a large volume of supplementary content, including the capacity to update this information as young people's information needs change.

Despite enthusiastic engagement from young people during the evaluation workshops, website analytical data indicated that few participants accessed The Gist outside of the classroom setting. It was also noted that many participants preferred to explore The Gist collectively with fellow classmates, rather than as an individual learning exercise. This suggests that young people are potentially more likely to engage with The Gist as a sexual health and pornography literacy resource within a facilitated and supported learning environment. Consequently, this poses important questions for future iterations of The Gist, including how to strengthen independent engagement with the website, and whether The Gist as a standalone digital platform is a sufficient intervention to meet the needs of this population group. As noted by the research team in earlier co-design stages [4], previous international research has also shown that while young people appreciate the availability of sexual health materials online, they often seek out the guidance of trusted adults and authority figures to help them make sense of this information, and to determine its validity [29,30]. Thus, consideration should be given to the overall structure of The Gist as an educational intervention, and whether alternative delivery methods, (such as a potential hybrid model involving face-to-face sessions at schools, community sites, or services) are more effective approaches to supporting the sexual development and wellbeing of structurally marginalised young people.

This study adds to limited research into the development and evaluation of pornography education. A recent systematic review identified only 2 education programs that addressed pornography as a major focus [10]. Neither have described in depth the process of designing and developing the program.

Wider Implications

As so called "digital natives," there is a widespread perception of young people as highly proficient consumers and navigators of digital information. However, digital literacy is associated with socio-economic status, with online health information less accessible and interpretable to young Australians of lower socioeconomic standing [31,32]. The principal findings of this evaluation, therefore, provide valuable insights into how structurally marginalized young people perceive The Gist as a digital health resource, including key user interface and design preferences, information needs, and online learning requirements.

As pornography literacy programs emerge within mainstream education settings throughout Australia [7], it is important to recognize that young people are not homogenous, and their information needs are likely to differ despite similarities in demographic characteristics. Even amongst our small cohort of structurally marginalized young people, there was a wide variation in experience levels, previous knowledge, and attitudes toward sexual and pornographic themes. These differences reflect the fact that adolescent sexual development is a highly individualised process, shaped by a range of biological, psychological, and social factors [33]. For our participants, this also meant that the perceived relevance and desirability of the Gist prototype differed depending upon their own set of unique personal circumstances. Participants from culturally and linguistically diverse backgrounds also shared how their differing cultural perspectives and understandings of sex and pornography shaped their experiences with the prototype and receptiveness to the key health messages within. Future pornography literacy programs, therefore, need to ensure that the information and content presented is suited to, and tailorable for a wide range of experience and knowledge levels, even within similar demographic cohorts, and is offered in a way that is inclusive of all sexuality, gender, and culturally diverse people.

This evaluation has developed recommendations for further prototype refinements. It is vital that additional testing be carried out following these modifications to ensure The Gist remains highly relevant to structurally marginalized young people seeking accessible sexual health and pornography information. Future research is also needed to assess the impact of The Gist on young people's knowledge, attitudes, and behaviors, and to assess its effectiveness as a standalone sexual health and pornography literacy intervention.

Limitations

There are several limitations in which the findings described above should be considered. Study participants were recruited through a single service provider, and were therefore already familiar with other participants and supporting staff members. While this familiarity may have helped some participants to feel more comfortable during workshop sessions, it also introduces the potential for social and group biases. In the presence of their peers and teachers, participants may have sought to conform to the dominant group narrative by offering opinions that felt more socially acceptable. The sensitive nature of the research content may have further exacerbated this, with participants potentially hesitant to share their own personal experiences with sex and pornography. Similarly, while facilitators emphasized the need for critical and honest feedback,

XSL•FO
RenderX

participants may have felt inclined to provide positive assessments of the website as an unintended courtesy to the research team, particularly as 2 of the facilitators had been involved in development of the prototype. The design of the evaluation and the analysis of data was conducted by 3 researchers who were not directly involved in development of the prototype; however, they were close colleagues of those who did develop the prototype.

Young people were also free to engage with The Gist prototype as much or as little as they liked, however, not all participants attended both workshops, meaning that exposure varied. During the workshops, facilitators provided additional prompts and guidance only as needed, and there was no prerequisite that all elements of The Gist be explored and evaluated by participants. While this flexibility is an integral part of co-design research with structurally marginalized cohorts, it also means that again, participants may have had variable experiences with The Gist website, leading to differing assessments of the prototype. Similarly, as The Gist was an early prototype, access was only available through a password protected URL link, further limiting potential student engagement with the site. Finally, the evaluation did not use any validated instruments or apply techniques to enhance the validity or credibility of data analysis.

Conclusion

Amid growing concerns over the potential impact of pornography on the sexual development of young people, digital resources such as The Gist are emerging as potential alternatives to contemporary interventions seeking to combat these concerns. Research is still needed to determine the efficacy and impact of digital pornography literacy resources on young people's sexual attitudes, beliefs, and behaviors. The findings from this formative evaluation highlight the potential for creating engaging and highly pertinent online platforms for pornographic literacy, specifically tailored to the distinctive requirements of structurally marginalized young people. However, low engagement with the prototype outside of facilitated settings suggests that a hybrid delivery model that incorporates face-to-face elements may be preferable. This study emphasizes that intervention development is a long-term process requiring multiple iterations of prototype evaluation and refinement.

Acknowledgments

The project was funded by the Victorian Health Promotion Foundation (VicHealth), Melbourne, Australia. Angela Davis, Cassandra Wright, and Megan Lim were supported by National Health and Medical Research Council fellowships. Michelle Raggatt was supported by a Monash University Australian Postgraduate Award. The project was delivered in partnership with Hester Hornbrook Academy. The Gist prototype's content, illustration, design, and production were created with contributions from Mike Ebinum, Stacey Murphy, Jenny Walsh, Cat Wall, and Bonnie Eichelberger. The authors gratefully acknowledge the contribution to this work of the Victorian Operational Infrastructure Support Program received by Burnet Institute.

Data Availability

The data generated in this study are not publicly available due to participant confidentiality. Selected extracts of data could be made available upon reasonable request to the corresponding author, subject to approval from the relevant research ethics committee.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Workshop aims.
[DOCX File, 19 KB - formative_v9i1e65859_app1.docx]

Multimedia Appendix 2
Workshop activities.
[DOCX File, 19 KB - formative_v9i1e65859_app2.docx]

Multimedia Appendix 3
Workshop survey results.
[DOCX File, 19 KB - formative_v9i1e65859_app3.docx]

References

1. Lim MSC, Agius PA, Carrotte ER, Vella AM, Hellard ME. Young Australians' use of pornography and associations with sexual risk behaviours. Aust N Z J Public Health 2017 Aug;41(4):438-443. [doi: 10.1111/1753-6405.12678] [Medline: 28664609]

XSL•FO
RenderX

2. Quadara A, El-Murr A, Latham J. The Effects of Pornography on Children and Young People: An Evidence Scan: Australian Institute of Family Studies; 2017.
3. Waling A, Bellamy R, Ezer P, Kerr L, Lucke J, Fisher C. “It’s kinda bad, honestly”: Australian students’ experiences of relationships and sexuality education. Health Educ Res 2020 Dec 23;35(6):538-552. [doi: 10.1093/her/cyaa032] [Medline: 32929480]
4. Davis AC, Wright CJ, Murphy S, et al. A digital pornography literacy resource co-designed with vulnerable young people: development of “The Gist”. J Med Internet Res 2020 Jun 1;22(6):e15964. [doi: 10.2196/15964] [Medline: 32348268]
5. Marshall EA, Miller HA. The role of sexual scripts in the relationship between pornography use and sexual coercion. J Interpers Violence 2023 Apr;38(7-8):5519-5541. [doi: 10.1177/08862605221123291] [Medline: 36181240]
6. Massey K, Burns J, Franz A. Young people, sexuality and the age of pornography. Sex Cult 2021 Feb;25(1):318-336. [doi: 10.1007/s12119-020-09771-z]
7. Albury K. Porn and sex education, porn as sex education. Porn Studies 2014;1(1-2):172-181. [doi: 10.1080/23268743.2013.863654]
8. A presentation for parents and carersis now available!. It’s Time We Talked. URL: https://itstimewetalked.com [accessed 2025-02-14]
9. A teaching resource for years 7 to 10. The Practical Guide to Love, Sex and Relationships. URL: https://www.lovesexrelationships.edu.au [accessed 2025-02-12]
10. Aznar-Martínez B, Lorente-De-Sanz J, López-i-Martín X, Castillo-Garayoa JA. Pornography and gender-based violence: two neglected topics in sexuality education. A systematic review. Sex Educ :1-17. [doi: 10.1080/14681811.2024.2316154]
11. Rothman EF, Adhia A, Christensen TT, Paruk J, Alder J, Daley N. A pornography literacy class for youth: results of a feasibility and efficacy pilot study. Am J Sex Educ 2018 Jan 2;13(1):1-17. [doi: 10.1080/15546128.2018.1437100]
12. Vandenbosch L, van Oosten JMF. The relationship between online pornography and the sexual objectification of women: the attenuating role of porn literacy education. J Commun 2017 Dec;67(6):1015-1036 [FREE Full text] [doi: 10.1111/jcom.12341]
13. Crocker BCS, Pit SW, Hansen V, John-Leader F, Wright ML. A positive approach to adolescent sexual health promotion: a qualitative evaluation of key stakeholder perceptions of the Australian Positive Adolescent Sexual Health (PASH) Conference. BMC Public Health 2019 Jun 3;19(1):681. [doi: 10.1186/s12889-019-6993-9] [Medline: 31159767]
14. Pfitzner N, Ollis D, Stewart R, Allen KA, Fitz-Gibbon K, Flynn A. Respectful Relationships Education in Australian: National Stocktake and Gap Analysis of Respectful Relationships Education Material and Resources Final Report: Monash university Gender and Family Violence Prevention Centre; 2022.
15. Meehan C. “They’re much too young”: the entanglement of porn, pleasure and age in sex education. J Sex Res 2024 Apr 8:1-9. [doi: 10.1080/00224499.2024.2326960] [Medline: 38588605]
16. Fisher CM, Waling A, Kerr L, Bellamy R, Ezer P, Mikolajczak G, et al. 6th National Survey of Australian Secondary Students and Sexual Health 2018: Australian Research Centre in Sex, Health & Society, La Trobe University; 2019. [doi: 10.3389/fpubh.2019.00217]
17. Power J, Kauer S, Fisher C, Chapman-Bellamy R, Bourne A. The 7th National Survey of Australian Secondary Students and Sexual Health 2021: The Australian Research Centre in Sex, Health and Society, La Trobe University; 2022.
18. McKibbin G, Hamilton B, Humphreys C. Hold pornography to account – not education programs – for children’s harmful sexual behaviour. The Conversation 2016.
19. Crabbe M, Flood M. School-based education to address pornography’s influence on young people: a proposed practice framework. Am J Sex Educ 2021 Jan 2;16(1):1-37. [doi: 10.1080/15546128.2020.1856744]
20. Brown G, Sorenson A, Hildebrand J. How they got it and how they wanted it: marginalised young people’s perspective on their experiences of sexual health education. Sex Educ 2012 Nov;12(5):599-612. [doi: 10.1080/14681811.2011.634141]
21. O’Farrell M, Corcoran P, Davoren MP. Examining LGBTI+ inclusive sexual health education from the perspective of both youth and facilitators: a systematic review. BMJ Open 2021 Sep 2;11(9):e047856. [doi: 10.1136/bmjopen-2020-047856] [Medline: 34475162]
22. Lirios A, Mullens AB, Daken K, et al. Sexual and reproductive health literacy of culturally and linguistically diverse young people in Australia: a systematic review. Cult Health Sex 2024 Jun 2;26(6):790-807. [doi: 10.1080/13691058.2023.2256376]
23. Atkins R, Sulik MJ, Hart D, Ayres C, Read N. The effects of school poverty on adolescents’ sexual health knowledge. Res Nurs Health 2012 Jun;35(3):231-241. [doi: 10.1002/nur.21471] [Medline: 22431188]
24. Wood SK, Ford K, Madden HCE, Sharp CA, Hughes KE, Bellis MA. Adverse childhood experiences and their relationship with poor sexual health outcomes: results from four cross-sectional surveys. Int J Environ Res Public Health 2022 Jul 21;19(14):8869. [doi: 10.3390/ijerph19148869] [Medline: 35886718]
25. Wadham E, Green C, Debattista J, Somerset S, Sav A. New digital media interventions for sexual health promotion among young people: a systematic review. Sex Health 2019 Apr;16(2):101-123. [doi: 10.1071/SH18127] [Medline: 30819326]
26. Sewak A, Yousef M, Deshpande S, Seydel T, Hashemi N. The effectiveness of digital sexual health interventions for young adults: a systematic literature review (2010-2020). Health Promot Int 2023 Feb 1;38(1):1-14. [doi: 10.1093/heapro/daac104] [Medline: 36757346]

XSL•FO
RenderX

27. Bauman AE, Nutbeam D. In: North R, editor. Evaluation in A Nutshell: A Practical Guide to the Evaluation of Health Promotion Programs, 2nd edition: McGraw-Hill Education Medical; 2014.
28. Olivares-Cuhat G. Learner factors in a high-poverty urban middle school. Urban Educ 2011;9(1):1-11.
29. Albury K. 'Recognition of competition' versus will to app: rethinking digital engagement in Australian youth sexual health promotion policy and practice. Media International Australia 2019 May;171(1):38-50. [doi: 10.1177/1329878X19850224]
30. Lupton D. Young people's use of digital health technologies in the global north: narrative review. J Med Internet Res 2021 Jan 11;23(1):e18286. [doi: 10.2196/18286] [Medline: 33427684]
31. Thomson S, Bortoli L. Preparing Australian students for the digital world: results from the PISA 2009 digital reading literacy assessment. Australian Council for Educational Research 2012.
32. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19. [doi: 10.2196/jmir.1619] [Medline: 22357448]
33. Kar SK, Choudhury A, Singh AP. Understanding normal development of adolescent sexuality: a bumpy ride. J Hum Reprod Sci 2015;8(2):70-74. [doi: 10.4103/0974-1208.158594] [Medline: 26157296]

Abbreviations

LGBTQIA+: lesbian, gay, bisexual, transgender, queer, questioning, intersex, asexual
STI: sexually transmitted infection
VicHealth: Victorian Health Promotion Foundation

Edited by A Mavragani; submitted 27.08.24; peer-reviewed by H Pratomo, M Lindroth; revised version received 26.01.25; accepted 28.01.25; published 04.03.25.

Please cite as:
Turvey J, Raggatt M, Wright CJC, Davis AC, Temple-Smith MJ, Lim MSC
A Digital Pornography Education Prototype Co-Designed With Young People: Formative Evaluation
JMIR Form Res 2025;9:e65859
URL: https://formative.jmir.org/2025/1/e65859
doi:10.2196/65859

© Jake Turvey, Michelle Raggatt, Cassandra J C Wright, Angela C Davis, Meredith J Temple-Smith, Megan S C Lim. Originally published in JMIR Formative Research (https://formative.jmir.org), 4.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Purposes of Smartphone Health Care Apps and the Practicality of Their Functions in Disaster Situations: Qualitative Function Assessment Study

Sayuri Nonaka[1*], PhD; Susumu Fujii[1*], PhD; Kosuke Chris Yamada[1,2*], PhD

[1]Disaster Medical Informatics Lab, International Research Institute of Disaster Science, Tohoku University, Sendai, Japan

[2]Research Institute of Sport Medical Science, Tokai University, Hiratsuka, Japan

[*]all authors contributed equally

Corresponding Author:
Sayuri Nonaka, PhD
Disaster Medical Informatics Lab, International Research Institute of Disaster Science, Tohoku University, Sendai, Japan

Abstract

Background: Japan has experienced various natural disasters, including the Great East Japan Earthquake in 2011. It becomes crucial to focus on strengthening self-help measures through health care apps that are used in normal times to help people during disasters. However, little is known about what health care apps would be useful in times of disaster.

Objective: This study aimed to investigate the prevalent functionalities and purposes of using health care apps during normalcy, explore their potential utility, and propose strategies for disaster response through their utilization.

Methods: We focus on highly ranked health care apps (within the top 100 in the health care category for iPhones by Apple, Inc for a certain period) and reclassify their purpose of use, such as sleep, relaxation, and exercise, in detail. We also investigate the functions within each health care app (measurement, recording, advice, content provision, and guidance to actual services), based on which we determine their potential utilization during disasters and anticipate potential solutions to address disaster-related challenges. We also consider the ideal framework of health care apps in disaster response, exploring possibilities such as the necessity of new disaster-specific apps or the adaptation of existing health care apps for disaster scenarios.

Results: Among the 70 free apps, the predominant functions included "recording" (n=60 cases, 86%) and "measurement" (n=47 cases, 67%), primarily encompassing the mechanical functions of wearable devices and smartphones. A similar trend was seen in the 77 paid apps, but "content provision" (n=54 cases, 70%) was the most prevalent. Furthermore, the "content provision" function was particularly common in the "purposes of use" categories "sleep" (χ^2_{11}=29; P<.001), "relaxation" (χ^2_{11}=14.6; P<.001), and "exercise" (χ^2_{11}=9.3; P=.002). This suggested the possibility of using the content provision function in existing health care apps to support mental and physical health even during a disaster.

Conclusions: The widespread use of apps during normal times could minimize hesitation in adopting them during disasters. The findings emphasize the potential for augmenting disaster-specific content within existing apps rather than developing new ones. This approach aligns with the likelihood of preinstalled app use during emergencies, indicating a pragmatic strategy for enhancing disaster response content within prevalent apps.

(JMIR Form Res 2025;9:e56862) doi:10.2196/56862

KEYWORDS

health care; disaster medicine; mobile app; survey; disaster situations; self-reliance; Japan; disaster response; mobile phone

Introduction

Japan has experienced various natural disasters, including the Great East Japan Earthquake (GEJE) in 2011 [1-3], the Kumamoto earthquake in 2016, and wind and flood damage induced by Typhoons 15 and 19 in 2019 [4-6]. Furthermore, there is a 70% probability of a Nankai Trough earthquake and an earthquake directly under the Tokyo metropolitan area by 2050, necessitating urgent attention to disaster prevention, mitigation, and response [7].

The issues in drug prescribing that became apparent after the GEJE have been partially resolved using the Medication Record Book held by those affected by the disaster [8]. Electronic standards such as E-Kusu links have been established, and many drugstores and local governments are transitioning from paper-based to electronic medication handbooks [9]. Furthermore, various issues, such as managing chronic diseases, including diabetes and mental health during disasters, have been identified [10-17]. To reduce the burden on public and mutual aid support systems in times of disaster, we thought that familiar

XSL•FO
RenderX

digital technology, health care apps for smartphones, could be expected to strengthen self-help [18-21]. In this study, we focused on strengthening self-help measures during disasters using health care apps that are used in everyday life.

The best way to determine whether an app can be used effectively during a disaster is to investigate whether it can be used during a disaster. However, it is difficult to conduct such a survey during a disaster. In addition, it is time-consuming to apply the findings to the next disaster. Therefore, we assume that apps commonly used in the health care field related to human life will be immediately available at the time of a disaster and divide them into subgenres related to the issues at the time of a disaster. The potential for use in times of disaster was considered by surveying the implemented IT functions. This is because we thought that it would be difficult to use during a disaster if the function is mainly just measurement and record keeping. If the function can provide detailed methods such as content, there is a high possibility that it can be used during a disaster. In other words, rather than creating an app that provides content useful only for disasters, we thought it would be more effective to add only content for disaster apps within an app already installed and tied to the issues at the time of a disaster.

This study aimed to investigate the prevalent functionalities and purposes of using health care apps during normalcy, explore their potential utility, and propose strategies for disaster response through their utilization.

Specifically, we focus on highly ranked health care apps (within the top 100 in the health care category for iPhones by Apple, Inc for a certain period) and reclassify their purpose of use, such as sleep, relaxation, and exercise, in detail. We also investigate the functions within each health care app (measurement, recording, advice, content provision, and guidance to actual services), based on which we determine their potential utilization during disasters and anticipate potential solutions to address disaster-related challenges. We also consider the ideal framework of health care apps in disaster response, exploring possibilities such as the necessity of new disaster-specific apps or the adaptation of existing health care apps for disaster scenarios.

Methods

Survey Procedure

Overview

In this study, we conducted a comprehensive survey of current health care–related smartphone apps to determine their functionalities and intended purposes. Subsequently, we investigated the functions implemented for each specific purpose. Finally, based on the findings, we evaluated the potential utilization of each purpose during disaster scenarios, identifying their utility and challenges.

The procedure that we followed for the surveys is elaborated below.

Initial Selection (Survey 1)

For approximately 3 months, we targeted commonly used daily life health care or fitness apps by many Japanese people that were ranked within the top 100 on app download sites. The download site is Apple's App Store.

Functional Survey (Survey 2)

We focused on technical functionalities such as measurement (including wearable device integration), data recording, advice, and video provision. We classified functionalities into 5 categories: measurement function, recording function, advice provision function, content provision function, and actual service function.

Purpose Classification (Survey 3)

We surveyed and categorized the intended purposes for each app into eating habits, self-management, sleep, relaxation, walking, running, exercise, cycle management (menstruation or pregnancy), habit formation, childcare or parenting, teaching materials or practice, and actual service provision.

Analysis of Characteristics of Apps (Survey 4)

We examined the characteristics of functional implementation unique to each purpose identified in surveys 2 and 3. We also considered how to use them in the event of a disaster.

Survey Description

Overview

Using the cumulative results from surveys 1 to 4, we assessed the utility and challenges of each purpose during disaster situations. We evaluated whether the existing system could be directly used or required customization to address disaster-specific issues. Detailed descriptions of each survey method follow below.

Survey 1: Targeted Apps (Commonly Used Daily-Life Health Care Apps)

We included iPhone apps available on the App Store of Apple, Inc. There was no category for emergencies or disasters in the App Store, so the category for medical or health care–related options would be the best choice for dealing with disasters. According to the App Store category descriptions, the "Medical" category is described as "Apps that are focused on medical education, information management, or health reference for patients or health care professionals" [22]. The "Health & Fitness" category comprises "Apps related to healthy living, including stress management, fitness, and recreational activities" [22]. We targeted apps in the "Health & Fitness" category because they are accessible for general daily use rather than exclusively for health care professionals. The survey period was from October 2021 to April 2022, and the data were collected 4 times. The target apps were chargeable and free in the "Health & Fitness" category displayed exclusively on iPhones, excluding rankings on iPads and Macs. In addition, all apps had to be ranked within the top 100 during all 4 data collection periods. The App Store has a primary category for the app. If the category is inappropriate for an app, it violates the guidelines [22]. Whether an app is classified in the "Health & Fitness" category is determined by Apple, Inc.

XSL•FO
RenderX

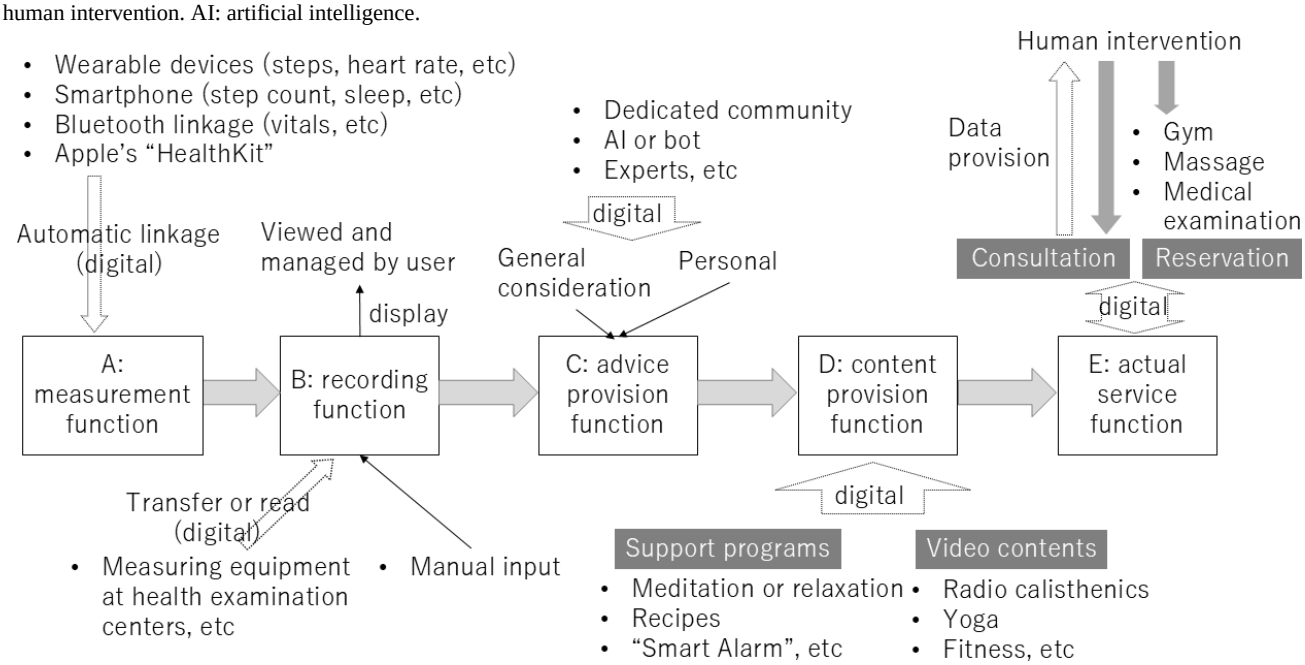
Survey 2: Classification of "Implemented Function"

This survey aimed to identify 5 distinct functions (A: measurement function, B: recording function, C: advice provision function, D: content provision function, and E: actual service function) within the health care apps targeted in survey 1. Assessment of function implementation relied on scrutinizing app descriptions and screenshot images available in the apps. We did not verify these functions' actual performance or effectiveness or whether they precisely aligned with the description provided. As there is no existing classification system for these 5 functions, researchers (SF and KCY) who are an expert in app development and evaluation defined them, and another expert (SN) carried out the classification work. Multiple researchers discussed any doubt about the decision. The elucidation of each function is as follows.

- "A: measurement function" was examined to determine whether the function displays numerical values (measured values) by linking with smartphone sensors or external devices worn by individuals (eg, wearable devices and home-use measuring devices). Information on whether the values were stored was not sought. If the description mentioned support for the Apple health care app or Health Kit, it was deemed linked to a wearable device and was considered to possess a "measurement" function. If the value could only be entered manually, no "measurement function" was provided.
- "B: recording function" was defined as the ability to link, read, and save (record) values acquired from manual inputs and view the data. Information on the storage duration (history period) or storage frequency (history count) was sought. The recorded values could be manually entered, read from external data, or include sensor values obtained from "A: measurement function."
- "C: advice provision function" focused on investigating whether the app offered advice to users from various sources, including dedicated communities, experts, or artificial intelligence functions. The advice content was not limited to advice optimized based on personal data, advice from nonexperts in the community, or generalized advice. However, supportive advice, such as encouraging users, was not considered. Additionally, advice that only explained how to use the app or explained the provided content was not considered.
- "D: content provision function" was defined as providing specific health care methods (content), such as videos and recipes, within the app.
- "E: actual service function" focused on scrutinizing whether health care services were provided through human intervention. Services were assumed to be provided through digital services, such as scheduling appointments at physical stores, phone calls, and chats.

Figure 1 illustrates an app's function through a series of flow diagrams, depicted as a sequence from A to E.

Figure 1. Classification of functions. The function of an app through a series of flow diagrams, depicted as a sequence from A to E. Data measured by the user is recorded, advice is offered based on the recordings, content aligned with the advice is provided, and the actual service is offered in the form of human intervention. AI: artificial intelligence.



Survey 3: Categorization of "Purpose of Use"

Survey 1 results, categorized broadly under the health care classification by Apple, were reclassified into more detailed purposes of use because of the initial classification's broadness. The reclassification aimed to offer a more granular breakdown of the apps' intended purposes and was based on the service purpose outlined by the provider without considering any unintended or user-applied purposes of use. Determining the purpose of use relied on the information provided within the app description and screenshot images. Initially, the apps were classified into 15 categories (Multimedia Appendix 1), including "eating habits," "sleep," "relaxation," "walking," "running," "exercise," "cycle management (menstruation or pregnancy)," and "childcare or parenting." If an app did not fit any of these categories, they were further classified, and if an app served

XSL•FO
RenderX

multiple purposes, it was assigned multiple purposes of use. As there is no existing classification system for this category, researchers (SF and KCY) who are an expert in app development and evaluation defined them, and another expert (SN) carried out the classification work. Multiple researchers discussed any doubt about the decision.

Survey 4: Evaluation of Cross-Tabulation Results Between "Implemented Function" and "Purpose of Use"

Survey 4 involved a cross-tabulation evaluation between the purpose of use (survey 3) and the implemented functions (survey 2) classifications. The primary objective was to explore any notable characteristics in the implemented functions (A to E) concerning each purpose of use. For instance, the assessment aimed to discern commonalities, such as whether "A: measurement function" and "B: recording function" were prevalent in apps designed for sleep-related use or whether "A: measurement function" was rare in apps offering actual services. We also analyzed which "implemented function" was prominently implemented for each "purpose of use." Multiple response test was used for the analysis. A commercially available software program JMP (version 17.1.0; SAS Institute) was used for the analyses. This survey was compiled and analyzed by an expert health care app specialist (SN) and verified by some experts (SF and KCY) in disaster medicine. This evaluation, encompassing surveys 1‐4, facilitated discussions on the utility and challenges posed during disasters for each purpose of use. Additionally, it provided insights into precautions and suggestions for customization to enhance utility during disaster scenarios.

Ethical Considerations

This study does not involve the collection of data or interventions with human participants but is based on existing published app rankings and app descriptions. Therefore, we confirmed that this study did not meet the ethical review requirements at our institution. More specifically, this study satisfied the following conditions: it did not involve experiments, observations, or interventions in the human body. It did not collect or reuse personal information from participants and used only publicly available information, so the identification of individuals is impossible.

Results

Result 1: Targeted Survey of Commonly Used Daily-Life Health Care Apps

Apps in Apple's "Health & Fitness" category included those with "in-app purchases" (Apple Inc, 2023) for both free and paid downloads. "In-app purchases" included "consumable," "nonconsumable," "auto-renewable subscriptions," and "nonrenewing subscriptions" [23]. The presence of "in-app purchases" was not used to distinguish between free and paid download apps and those that required a fee to commence use were classified as paid download apps; Apple's classification at the time of the survey was used. The evaluation of free apps occurred on 4 occasions: October 22, 2021; October 29, 2021; November 22, 2021; and January 11, 2022. Paid apps were similarly scrutinized on 4 different dates: February 2, 2022; February 22, 2022; March 22, 2022; and April 12, 2022. In the second survey, the proportion of free apps was 86%, while that of paid apps was 84%. In the third survey, these values changed to 77% and 81%, respectively. By the fourth survey, the proportions further declined to 70% for free apps and 77% for paid apps. The results of survey 1 indicated that 147 (70 free and 77 paid) apps ranked in the top 100 on all 4 occasions.

Result 2: Classification of "Implemented Function"

Figure 2 provides an overview of the survey procedure. The results regarding the implemented functions within the apps are summarized in Table 1. It includes the 147 apps from Result 1, which ranked in the top 100 across all 4 survey points. Among the 70 free apps, the predominant function was "B: recording function" (n=60 cases, 86%), followed by "A: measurement function" (n=47 cases, 67%). These functionalities primarily leveraged the mechanical capabilities inherent in wearable devices and smartphones. The 77 paid apps also displayed a high implementation rate for "B: recording function" (n=48 cases, 62%) and "A: measurement function" (n=39 cases, 51%); however, the most notable difference lay in the implementation of "D: content provision function" (n=54 cases, 70 %) within paid apps, showcasing a higher rate than in free apps (n=28 cases, 40%). The app vendor specifically curated this content, marking a distinct trend from the mechanically driven performance witnessed in free apps.

XSL•FO
RenderX

Figure 2. Survey procedure overview. After survey 1 narrowed the target apps to 147, we examined the functions implemented in these apps (survey 2) and the purpose of using these apps (survey 3). Survey 4 analyzed the characteristics of apps based on their implemented functions and purpose.

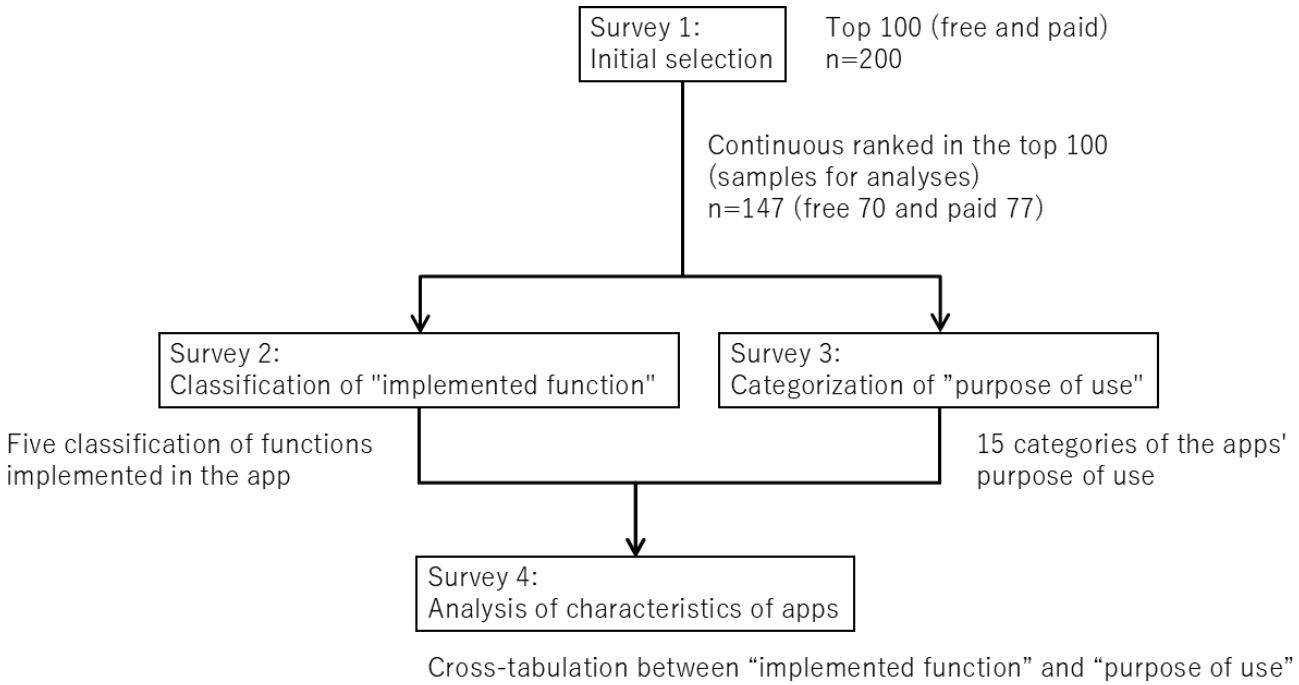


Table . Results of classification of functions implemented in the app.

	A: measurement	B: recording	C: advice provision	D: content provision	E: actual service
Free apps[a] (n=70), n (%)	47 (67)	60 (86)	32 (46)	28 (40)	10 (14)
Paid apps[b] (n=77), n (%)	39 (51)	48 (62)	18 (23)	54 (70)	0 (0)
Total downloaded apps (n=147), n (%)	86 (58.5)	108 (73.5)	50 (34)	82 (55.8)	10 (6.8)

[a]"Free" means free download apps.

[b]"Paid" means paid download apps.

Result 3: Categorization of "Purpose of Use"

The results of the "purpose of use" categorization are summarized in Table 2. Apple's categorization had a coarse granularity regarding the apps' "purpose of use"; therefore, the recategorization was performed. No additions other than the 15 "purpose of use" categorizations set during planning were performed. Among the 70 free apps, "walking" was the most common (n=16 cases, 23%), followed by "exercise" and "self-management" (n=15 cases each, 21%). Among the 77 paid apps, "sleep" (n=21, 27%) and "relaxation" (n=20, 26%) were the most common, followed by "exercise" (n=15, 20%). Overall (147 combined free and paid apps), the most popular activities were "exercise" and "sleep" (n=30 cases each, 20.4%), followed by "relaxation" (n=28, 19%) and "self-management" (n=25, 17%). Note that only paid apps were available for "childcare or parenting" and "teaching materials or practice," whereas only free apps were available for "actual service provision." The designation "not applicable" refers to apps lacking health care functionality. For the categories of "self-management," "habit formation," "teaching materials or practice," and "tools for Apple Health," a single purpose of use was assigned if it could not be determined to be the primary purpose of use, even if multiple purposes of use were plausible. However, if an app implied multiple purposes, such as "eating habits" and "exercise," the apps were tallied separately within each applicable category, leading to a cumulative count across various purposes.

XSL•FO
RenderX

Table . Results of the categorization of the apps' purpose of use.

Purpose of use	Free (n=70), n (%)	Paid (n=77), n (%)	Total (n=147), n (%)
Eating habits	10 (14)	3 (4)	13 (8.8)
Sleep	9 (13)	21 (27)	30 (20.4)
Relaxation	8 (11)	20 (26)	28 (19)
Walking	16 (23)	2 (3)	18 (12.2)
Running	5 (7)	2 (3)	7 (4.8)
Exercise	15 (21)	15 (20)	30 (20.4)
Cycle management (menstruation or pregnancy)	8 (11)	4 (5)	12 (8.2)
Childcare or parenting	0 (0)	3 (4)	3 (2)
Actual service provision	10 (14)	0 (0)	10 (6.8)
Self-management	15 (21)	10 (13)	25 (17)
Habit formation	2 (3)	4 (5)	6 (4.1)
Teaching materials or practice	0 (0)	4 (5)	4 (2.7)
Tools for Apple Health	0 (0)	3 (4)	3 (2)
Not applicable	9 (13)	0 (0)	9 (6.1)
Pack of apps	0 (0)	2 (3)	2 (1.4)

Result 4: Evaluation of Cross-Tabulation Results Between "Implemented Function" and "Purpose of Use"

The survey was conducted to determine the characteristics of the provided functions according to the purpose of use. "Tools for Apple Health," "pack of apps," and "not applicable" were excluded from survey 4 because they were not directly relevant to health-related purposes of use. The results of combining the "implemented function" categorized in survey 2 for each "purpose of use" classified in survey 3 are presented in Table 3.

Table . Cross-tabulation results of "implemented function" classifications by "purpose of use" of apps.

Purpose of use	A: measurement, n (%)	B: recording, n (%)	C: advice provision, n (%)	D: content provision, n (%)	E: actual service, n (%)
Eating habits (n=13)	12 (92)	13 (100)	10 (77)	11 (85)	1 (8)
Sleep (n=30)	17 (57)	16 (53)	11 (37)	30 (100)	1 (3)
Relaxation (n=28)	12 (43)	14 (50)	7 (25)	26 (93)	1 (4)
Walking (n=18)	18 (100)	18 (100)	8 (44)	8 (44)	2 (11)
Running (n=7)	7 (100)	7 (100)	5 (71)	6 (86)	0 (0)
Exercise (n=30)	21 (70)	27 (90)	13 (43)	26 (87)	2 (7)
Cycle management (menstruation or pregnancy; n=12)	8 (67)	12 (100)	11 (92)	7 (58)	3 (25)
Childcare or parenting (n=3)	1 (33)	1 (33)	1 (33)	3 (100)	0 (0)
Actual service provision (n=10)	4 (40)	9 (90)	6 (60)	6 (60)	10 (100)
Self-management (n=25)	22 (88)	25 (100)	9 (36)	3 (12)	0 (0)
Habit formation (n=6)	1 (17)	6 (100)	0 (0)	0 (0)	0 (0)
Teaching materials or practice (n=4)	0 (0)	0 (0)	0 (0)	4 (100)	0 (0)

Table 4 examines trends in the "purpose of use" for each "implemented function." For "A: measuring function," a significantly greater number of respondents reported "eating habits," "walking," "running," and "self-management," while

XSL•FO
RenderX

fewer respondents reported “relaxation,” “habit formation,” and “teaching materials or practice.” In the case of “B: recording function,” “eating habits,” “walking,” “cycle management,” and “self-management” were significantly more common, whereas “sleep,” “relaxation,” and “teaching materials or practice” were significantly less common. Regarding “C: advice provision function,” significantly more instances were found in “eating habits” and “cycle management” whereas “habit formation” reported significantly fewer occurrences. For “D: content provision function,” significantly more instances were observed in “sleep,” “relaxation,” “exercise,” and “teaching materials or practice,” whereas “self-management” and “habit formation” reported significantly fewer occurrences. In the case of “E: actual service function,” significantly more instances were noted in “actual service provision,” whereas “self-management” reported significantly fewer occurrences.

XSL•FO
RenderX

Table . Results of multiple response test: “purpose of use” categorized by “implemented function” of apps.

	χ^2 (*df*)	*P* value
A: measurement function		
Eating habits	5.7 (11)	.02
Sleep	0.7 (11)	.41
Relaxation	5.1 (11)	.02
Walking	16.1 (11)	<.001
Running	6.3 (11)	.01
Exercise	0.5 (11)	.48
Cycle management (menstruation or pregnancy)	0.0 (11)	.84
Childcare or parenting	1.2 (11)	.28
Actual service provision	2.4 (11)	.13
Self-management	7.5 (11)	.006
Habit formation	5.7 (11)	.02
Teaching materials or practice	8.2 (11)	.004
B: recording function		
Eating habits	5.4 (11)	.02
Sleep	12.0 (11)	<.001
Relaxation	13.8 (11)	<.001
Walking	7.5 (11)	.006
Running	2.9 (11)	.09
Exercise	1.8 (11)	.18
Cycle management (menstruation or pregnancy)	5.0 (11)	.03
Childcare or parenting	3.3 (11)	.07
Actual service provision	0.6 (11)	.44
Self-management	10.4 (11)	.001
Habit formation	2.5 (11)	.11
Teaching materials or practice	13.4 (11)	<.001
C: advice provision function		
Eating habits	8.4 (11)	.004
Sleep	0.0 (11)	.92
Relaxation	2.0 (11)	.16
Walking	0.4 (11)	.55
Running	3.3 (11)	.07
Exercise	0.4 (11)	.52
Cycle management (menstruation or pregnancy)	15.6 (11)	<.001
Childcare or parenting	0.0 (11)	.88
Actual service provision	2.1 (11)	.15
Self-management	0.0 (11)	.87
Habit formation	5.7 (11)	.02
Teaching materials or practice	3.8 (11)	.05
D: content provision function		
Eating habits	3.3 (11)	.07
Sleep	29.0 (11)	<.001

XSL•FO
RenderX

	χ^2 (*df*)	*P* value
Relaxation	14.6 (11)	<.001
Walking	2.2 (11)	.14
Running	2.0 (11)	.16
Exercise	9.3 (11)	.002
Cycle management (menstruation or pregnancy)	0.1 (11)	.81
Childcare or parenting	2.9 (11)	.09
Actual service provision	0.0 (11)	.91
Self-management	26.7 (11)	<.001
Habit formation	11.5 (11)	<.001
Teaching materials or practice	3.9 (11)	.049
E: actual service function		
Eating habits	0.0 (11)	.98
Sleep	0.9 (11)	.33
Relaxation	0.8 (11)	.38
Walking	0.3 (11)	.59
Running	1.1 (11)	.30
Exercise	0.0 (11)	.86
Cycle management (menstruation or pregnancy)	3.4 (11)	.06
Childcare or parenting	0.5 (11)	.49
Actual service provision	51.8 (11)	<.001
Self-management	3.9 (11)	.048
Habit formation	0.9 (11)	.33
Teaching materials or practice	0.6 (11)	.43

Discussion

Principal Results

The results from surveys 1 - 4 provided insights into the potential usefulness of health care apps in the event of a disaster and highlighted specific concerns that need to be addressed for each purpose of use. We also discussed precautions and customization suggestions for use during times of disaster.

Survey 1 indicated that apps with enduring high rankings are more likely to be used during regular times. Considering the challenges of discovering and adopting new apps during disasters, using established apps beforehand could offer substantial advantages.

Survey 2 revealed that free apps primarily focus on measurement and recording, raising doubts about their direct applicability in times of disaster. By contrast, many paid apps offer content that can provide concrete advice, suggesting they could be used during disasters.

Survey 3 suggested a possibility of comprehensively addressing care demands during disasters by categorizing “exercise” as “physical” to maintain physical health and “sleep” and “relaxation” as “mental” (wellness). The survey also suggested using the app to provide comprehensive care during disasters.

Survey 4 explored the relationship between implemented functions and purpose of use, suggesting that health care apps require implementing functions such as recording and content provision. The implementation rate of content-providing functions was high for “physical” (eg, “exercise”) and “mental wellness” (eg, “sleep” and “relaxation”). In other words, apps frequently used during normal times may provide comprehensive (physical and mental) and specific content (response methods) during disasters. Using them during disaster scenarios is highly beneficial as they are already installed and configured.

In the following subsections, we discuss the anticipated effects of the apps’ specific “purpose of use” during disasters and suggest ways to improve their effectiveness in such scenarios, considering the apps’ “implemented functions.” The timing of using health care apps after a disaster is thought to vary from person to person. Therefore, the disaster phase we will examine here is the period after evacuation and after securing personal safety.

“Eating Habits” Apps

Eating habits during disasters are biased toward carbohydrates such as rice, bread, and noodles, rendering maintaining a nutritionally balanced diet difficult [24]. Moreover, this can lead to life-threatening problems for those with allergies or chronic illnesses because they cannot prepare their meals [25].

XSL•FO
RenderX

A study reported that, following the GEJE, patients with type 1 diabetes required 6 months for their HbA_{1c} values to return to predisaster levels [26].

The "eating habits" apps are characterized by their "C: advice provision function"; however, solely relying on this function in situations such as evacuation shelters can be challenging. In such times, obtaining suitable food items and meals customized to individual situations becomes crucial but can also be difficult to procure. For instance, devising a meal tailored to one's circumstances, especially considering evacuation rations, poses a significant challenge. Future endeavors should implement the "D: content provision function" to provide specific content. Moreover, the "C: advice provision function" includes a feature that allows information to be exchanged within a dedicated community. This feature enables the people impacted by the disaster to mutually exchange information and receive advice from medical specialists and nutrition managers.

"Sleep" and "Relaxation" Apps

Daily sleep is an indicator of mental health [27]. Based on sleep status, even in the event of a disaster, the early detection of mental health problems may be possible [28,29].

Many of the "sleep" and "relaxation" apps were equipped with "D: content provision function." Several users used paid apps, indicating that they have routine difficulties with "sleep" and "relaxation" and seek content that addresses these concerns. Although whether the content has a scientific basis is debatable, sleep and relaxation are particularly important for maintaining mental health even in times of disaster, and the possibility of providing a concrete method (content) can render the content useful in disaster situations. In the future, providing content tailored to the special environment of a disaster, such as living in an evacuation center, will be crucial.

"Walking," "Running," and "Exercise" Apps

In evacuation centers, space limitations restrict physical activity, leading to a sedentary lifestyle and potential health issues such as poor circulation and muscle weakness due to a lack of exercise. Physical exercises, particularly for older adults, play a crucial role in preventing frailty [30]. However, challenges were observed during the GEJE, where exercise acceptance within evacuation centers was limited [31]. To address these issues, strategies have been proposed to include exercise instructors in disaster medical assistance teams and form a disaster-exercise support network, emphasizing collaboration with other organizations and exercise initiatives that include local governments [32].

While "walking" and "running" apps showed no significant implemented functions, the "exercise" apps tended to have more "D: content provision function." Notably, "A: measurement function" was significantly implemented in "walking" and "running" apps. However, measuring the number of steps or distances might offer limited utility during disasters. Unless a specific means of "exercise" is provided, effectively using the system in the event of a disaster will be difficult. The challenge will be to provide an "exercise" content that can be implemented within limited spaces, including evacuation centers.

"Cycle Management (Menstruation or Pregnancy)" Apps

In the aftermath of the GEJE, female evacuees encountered challenges, including inadequate sanitary products, underwear, and essentials for infants and young children [33,34]. Women (especially pregnant women and caregivers) are at higher risk of facing mental health problems during disasters [35,36], highlighting the stressful environment for several women in such situations. Frameworks such as The Sendai Framework for Disaster Risk Reduction, which describes women's perspectives and participation in disaster reduction and building women's capacities [37], and guidelines from the Cabinet Office ("Women's Perspective for Strengthening Disaster Response Capabilities: Disaster Prevention and Recovery Guidelines from a Gender Equality Perspective") [38,39] focus on enhancing women's roles in disaster reduction and recovery.

Several women manage their menstrual and pregnancy cycles during normal times, suggesting that incorporating health care apps into their daily lives might be straightforward. For instance, as of 2022, the "LunaLuna" app, one of the "cycle management" apps, has been downloaded more than 19 million times [40].

By predicting these things, we can predict the demand for sanitary products, determine what is in short supply, and know when there is a problem with our bodies, such as when the cycle is different from normal. We can use this app effectively.

"Childcare or Parenting" Apps

Only 3 apps were identified, all of which were paid apps used by parents to support their children's development and growth. Therefore, their use during disasters would be challenging.

"Actual Service Provision" Apps

The "actual service" apps provided convenience to existing members, such as digital reservations and consultations. Currently, these apps are only intended for use during normal times. However, they could prove useful during evacuations if they offer real services related to supplies and medical care.

"Self-Management" Apps

Previous studies have demonstrated the potential benefits of using wearable devices to track activity, physiological functions, and lifestyle data. Having undergone rigorous testing for data accuracy, they can have a positive physical and mental impact [41-44].

The "self-management" apps tended to implement "A: measurement function" and "B: recording function," but lacked "C: advice provision function" and "D: content provision function." Currently, these apps only measure and record data from wearable devices, with limited use of the collected data.

However, ongoing research on wearable device technology suggests that it may become an essential tool in future health care apps by facilitating behavioral change and offering benefits to health care professionals and patients [45]. To date, most "self-management" apps focus on measurement and recording, rendering them less effective during a disaster scenario.

XSL•FO
RenderX

"Habit Formation" Apps

These apps primarily encouraged behavioral change through reminders and consistent recording. However, its utility during disaster situations seems improbable at present.

"Teaching Materials or Practice" Apps

These apps served only "D: content provision function." They could prove useful if tailored precisely to specific needs. Creating content based on accurate knowledge of the situation would be safer and more precise during disasters than a general internet search. They might serve as a straightforward contingency preparedness tool not typically used in normal circumstances.

Overall Perspective

Disaster situations require health care apps that not only measure and record day-to-day activities but also provide specific methods and actual services addressing unique disaster-related issues. These apps should encompass functionalities that can provide concrete methods specially tailored to disaster situations. We believe this is feasible using the content provision function in this study. For example, in terms of food, the blood sugar levels of diabetics will likely be difficult to control when eating emergency food in the event of a disaster. Emergency food generally consists mainly of carbohydrates, so there is a risk of sudden fluctuations in blood sugar levels and high blood sugar. In addition, there is a possibility that dehydration symptoms will worsen diabetes symptoms due to a lack of hydration due to concerns about using the toilet [46]. Therefore, if the app content could encourage people to drink more water and provide information on the fact that the rice balls and sweet buns provided immediately after a disaster contain sodium and the recommended amount of food to eat, it could be very useful. However, avoiding hypoglycemia and replenishing sugar levels would still be necessary.

Similarly, there is a trend for apps that deal with issues such as sleep, exercise, and stress management during disasters to have a significant implementation of the "D: content provision function," which also suggests the possibility of managing health in terms of both "mental" and "physical" aspects even during disasters. In other words, the key is not to develop new apps but to prepare content specific to disasters. Because these apps are widely used normally, they are highly reliable and have low initial implementation costs. Therefore, enhancing existing content is more reasonable than developing new apps as disaster countermeasures.

Limitations

There are some limitations to this study. First, this study focused only on Apple's App Store apps. Although we assumed that the top 100 apps were used among all the users of smartphones, they are not necessarily apps that are used frequently. It could be needed to expand the investigation to Android apps as well. Second, this study could not clearly define the "purpose of use" classification. The "purpose of use" classification assumed identical functionality for apps serving multiple purposes. For example, if an app served both "eating habits" and "exercise," the assumption was that identical functionality existed for both "purposes of use." However, while the "D: content provision function" might be implemented for "eating habits," it might not extend to "exercise." Despite this, the "D: content provision function" was also assumed to be implemented in the "exercise" category for discussion purposes. This was because of its existence in the app itself. In categories such as "self-management," "habit formation," and "teaching materials or practice," where the app providers did not specify distinct health care purposes, leaving them to the users, overlap between multiple "purposes of use" was assumed to be absent. Third, there are other aspects of the app's characteristics. In this study, we investigated the characteristics of apps that could be used during a disaster in ordinary time. It is also possible that apps from genres (categories) uncovered in this study could be adapted to the health field. Finally, we believe that the results of this study are not relevant in times of immediate disaster or response to disaster but are effective in the aftermath and the recovery phase.

Conclusions

In this study, we underscored the high utility of apps used in everyday contexts, highlighting their potential usefulness even in disaster situations owing to their reliability and reduced initial implementation.

The "D: content provision function" emerged as crucial, particularly during disasters when specific solutions are essential. Since the "D: content provision function" tended to be significantly implemented in the "sleep," "relaxation," and "exercise" categories, we observed the potential for health management in both the "mental" and "physical" realms even amid disasters. We believe that providing tailored content that can be used effectively during such crises is an important issue for the future.

In the future, apps in subgenres where the content provision function had a significantly high implementation rate will be studied individually to verify their effectiveness.

Acknowledgments

Part of this research was conducted as part of a joint research project with NTT Space Environment and Energy Laboratories entitled "Study of an intervention model for the construction of a next-generation evacuation system, the improvement of life after evacuation and the promotion of actions toward disaster readiness by local residents" and Research on "Digital Hospitals in Metaverse Spaces that Protect Local Lives" Realized through the Integration of Disaster Science and IOWN, and JST-OPERA: Joint Research Promotion Program for Industry-Academia Co-Creation Platform (Consortium for Co-Creation of Electric Power and Telecommunications Fusion Networks). Generative artificial intelligence was not used in any portion of the manuscript writing.

XSL•FO
RenderX

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

SN and SF conceived the study. SN compiled the data, developed the statistical analysis plan, and conducted statistical analyses. KCY contributed to the interpretation of the results. SN drafted the original manuscript. SF supervised the conduct of this study. All authors reviewed the manuscript draft and revised it critically on intellectual content. All authors approved the publication of the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Categorization of the app's purpose of use.
[DOCX File, 16 KB - formative_v9i1e56862_app1.docx]

References

1. 2011 Tohoku-Oacific Ocean earthquake (Great East Japan Earthquake) [Web page in Japanese]. Cabinet Office. 2022. URL: https://www.bousai.go.jp/2011daishinsai/pdf/torimatome20220308.pdf [accessed 2023-11-17]
2. Egawa S. Progress of disaster medicine during ten years after the 2011 Great East Japan Earthquake. Tohoku J Exp Med 2021 Mar;253(3):159-170. [doi: 10.1620/tjem.253.159] [Medline: 33762509]
3. Ishigaki A, Higashi H, Sakamoto T, Shibahara S. The Great East-Japan Earthquake and devastating tsunami: an update and lessons from the past Great Earthquakes in Japan since 1923. Tohoku J Exp Med 2013 Apr;229(4):287-299. [doi: 10.1620/tjem.229.287] [Medline: 23583960]
4. Damage caused by Typhoon No 15 in 2019 [Web page in Japanese]. Cabinet Office. 2019. URL: https://www.bousai.go.jp/updates/r1typhoon15/pdf/r1typhoon15_30.pdf [accessed 2023-11-17]
5. Damage from the 2016 Kumamoto Earthquake with the epicenter in Kumamoto Prefecture, Japan [Web page in Japanese]. Cabinet Office. 2019. URL: https://www.bousai.go.jp/updates/h280414jishin/pdf/h280414jishin_55.pdf [accessed 2023-11-17]
6. Damage caused by Typhoon no 19 in 2019 [Web page in Japanese]. Cabinet Office. 2020. URL: https://www.bousai.go.jp/updates/r1typhoon19/pdf/r1typhoon19_45.pdf [accessed 2023-11-17]
7. Summary of the white paper on land, infrastructure, transport and tourism in Japan. MLIT. 2021. URL: https://www.mlit.go.jp/statistics/content/001442571.pdf [accessed 2023-11-17]
8. Examples of the use of the medication handbook during the Great East Japan Earthquake [Web page in Japanese]. Japan Pharmaceutical Association. 2012. URL: https://www.nichiyaku.or.jp/files/co/activities/shinsai_techo.pdf [accessed 2025-02-26]
9. E-kusu link [Web page in Japanese]. Japan Pharmaceutical Association. 2016. URL: https://www.nichiyaku.or.jp/e_kusulink/ [accessed 2023-11-17]
10. Abe M, Akaishi T, Ishizawa K, et al. A massive natural disaster, the Great East Japan Earthquake, and the incidence of dialysis due to end-stage kidney disease. J Nephrol 2022 Apr;35(3):719-724. [doi: 10.1007/s40620-021-01140-9] [Medline: 34642850]
11. Hirai H, Okazaki K, Ohira T, et al. Suboptimal diabetic control and psychological burden after the triple disaster in Japan: the Fukushima Health Management Survey. BMJ Open Diabetes Res Care 2021 Sep;9(1):e002007. [doi: 10.1136/bmjdrc-2020-002007] [Medline: 34497045]
12. Ishikuro M, Obara T, Murakami K, et al. Relation of disaster exposure with maternal characteristics and obstetric outcomes: the Tohoku medical megabank project birth and three-generation cohort study. J Epidemiol 2023 Mar 5;33(3):127-135. [doi: 10.2188/jea.JE20210052] [Medline: 34219121]
13. Kishimoto M, Noda M. Diabetes care: after the Great East Japan Earthquake. J Diabetes Investig 2013 Jan 29;4(1):97-102. [doi: 10.1111/jdi.12025] [Medline: 24843638]
14. Kotozaki Y, Tanno K, Sakata K, et al. Association between social isolation and total mortality after the Great East Japan Earthquake in Iwate Prefecture: findings from the TMM CommCohort study. Int J Environ Res Public Health 2022 Apr 5;19(7):4343. [doi: 10.3390/ijerph19074343] [Medline: 35410023]
15. Nakaya N, Nakamura T, Tsuchiya N, Tsuji I, Hozawa A, Tomita H. The association between medical treatment of physical diseases and psychological distress after the Great East Japan Earthquake: the Shichigahama Health Promotion Project. Disaster Med Public Health Prep 2015 Aug;9(4):374-381. [doi: 10.1017/dmp.2015.52] [Medline: 25912962]
16. Ohira T, Nakano H, Nagai M, et al. Changes in cardiovascular risk factors after the Great East Japan Earthquake. Asia Pac J Public Health 2017 Mar;29(2_suppl):47S-55S. [doi: 10.1177/1010539517695436] [Medline: 28330394]

XSL•FO
RenderX

17. Orui M, Sato Y, Tazaki K, Kawamura I, Harada S, Hayashi M. Delayed increase in male suicide rates in tsunami disaster-stricken areas following the Great East Japan Earthquake: a three-year follow-up study in Miyagi Prefecture. Tohoku J Exp Med 2015 Mar;235(3):215-222. [doi: 10.1620/tjem.235.215] [Medline: 25765170]
18. Bucci S, Schwannauer M, Berry N. The digital revolution and its impact on mental health care. Psychol Psychother 2019 Jun;92(2):277-297. [doi: 10.1111/papt.12222] [Medline: 30924316]
19. Desveaux L, Shaw J, Saragosa M, et al. A mobile app to improve self-management of individuals with type 2 diabetes: qualitative realist evaluation. J Med Internet Res 2018 Mar 16;20(3):e81. [doi: 10.2196/jmir.8712] [Medline: 29549070]
20. Dugas M, Wang W, Crowley K, et al. Engagement and outcomes associated with contextual annotation features of a digital health solution. J Diabetes Sci Technol 2022 Jul;16(4):804-811. [doi: 10.1177/1932296820976409] [Medline: 33355003]
21. Fujii S, Nonaka S, Nakayama M. Use of medical information and digital services for self-empowerment before, during, and after a major disaster. Tohoku J Exp Med 2021 Nov;255(3):183-194. [doi: 10.1620/tjem.255.183] [Medline: 34853210]
22. Choosing a category. Apple Developer. 2023. URL: https://developer.apple.com/app-store/categories/ [accessed 2023-11-17]
23. In-app purchase. Apple Developer. 2023. URL: https://developer.apple.com/in-app-purchase/ [accessed 2023-11-17]
24. Tokumaru O, Fujita M, Nagai S, Minamikawa Y, Kumatani J. Medical problems and concerns with temporary evacuation shelters after great earthquake disasters in Japan: a systematic review. Disaster Med Public Health Prep 2022 Aug;16(4):1645-1652. [doi: 10.1017/dmp.2021.99] [Medline: 34103106]
25. Sudo N, Shimada I, Tsuboyama-Kasaoka N, Sato K. Revising “nutritional reference values for feeding at evacuation shelters” according to nutrition assistance by public health dietitians based on past major natural disasters in Japan: a qualitative study. Int J Environ Res Public Health 2021 Sep 24;18(19):10063. [doi: 10.3390/ijerph181910063] [Medline: 34639364]
26. Kamimura M, Hakoda A, Kanno J, Nishii A, Igarashi Y, Fujiwara I. Glycemic control in type 1 diabetic patients following the Great East Japan Earthquake and tsunami. J Jpn Diabetes Soc 2014;57:16-21. [doi: 10.11213/tonyobyo.57.16]
27. Recognize early signs of mental illness [Web page in Japanese]. National Center of Neurology and Psychiatry. 2023. URL: https://kokoro.ncnp.go.jp/health_understanding.php#02 [accessed 2023-11-17]
28. Li X, Buxton OM, Hikichi H, et al. Predictors of persistent sleep problems among older disaster survivors: a natural experiment from the 2011 Great East Japan earthquake and tsunami. Sleep 2018 Jul 1;41(7):zsy084. [doi: 10.1093/sleep/zsy084] [Medline: 29726979]
29. Nakamura Y, Suda T, Murakami A, et al. Sleep disturbance of evacuees in Minamisanriku Town after Great East Japan Earthquake: risk factors and treatment. Tohoku J Exp Med 2020 Jul;251(3):207-216. [doi: 10.1620/tjem.251.207] [Medline: 32669518]
30. Maeda K, Shamoto H, Furuya S. Feeding support team for frail, disabled, or elderly people during the early phase of a disaster. Tohoku J Exp Med 2017 Aug;242(4):259-261. [doi: 10.1620/tjem.242.259] [Medline: 28747593]
31. Yamashita M, Seino S, Nofuji Y, et al. The Kesennuma study in Miyagi, Japan: study design and baseline profiles of participants. J Epidemiol 2022 Dec 5;32(12):559-566 [FREE Full text] [doi: 10.2188/jea.JE20200599] [Medline: 33840651]
32. Survey and research project on physical and mental support and utilization of exercise and sports in the areas affected by the Great East Japan Earthquake [Web page in Japanese]. Japan Health Promotion & Fitness Foundation. 2012. URL: https://www.health-net.or.jp/tyousa/houkoku/pdf/h23_shinsaihoukokusyo.pdf [accessed 2023-11-17]
33. Disaster prevention and reconstruction from a gender equal society perspective—lessons from the Great East Japan Earthquake—(from the “White Paper on Gender Equality 2012”). Gender Equality Bureau Cabinet Office. 2023. URL: https://www.gender.go.jp/english_contents/about_danjo/whitepaper/pdf/2012-01.pdf [accessed 2025-02-26]
34. Survey on the status of response to the earthquake from the perspective of gender equality (July 2012) [Web page in Japanese]. Gender Equality Bureau Cabinet Office. 2012. URL: https://www.gender.go.jp/policy/saigai/pdf/hokokusyo_P17-P22.pdf [accessed 2023-11-17]
35. Orui M, Tanaka E, Maeda M, et al. Mental health and suicide prevention during disasters. Nihon Koshu Eisei Zasshi 2020;67(2):101-110. [doi: 10.11236/jph.67.2_101] [Medline: 32092726]
36. Research group for "Research on cooperation methods between disaster psychiatric assistance team (DPAT) and community mental health systems. Manual for Mental Health and Medical Welfare Activities in Disasters in Local Governments. 2021. URL: https://www.mhlw.go.jp/content/000772550.pdf [accessed 2023-11-17]
37. Sendai framework for disaster risk reduction 2015-2030. PreventionWeb. 2016. URL: https://www.preventionweb.net/files/43291_sendaiframeworkfordrren.pdf [accessed 2023-11-17]
38. Women’s perspective for strengthening disaster response capabilities—disaster prevention and recovery guidelines from a gender equality perspective [Web page in Japanese]. Gender Equality Bureau Cabinet Offic. 2020. URL: https://www.gender.go.jp/policy/saigai/fukkou/pdf/guidelene_01.pdf [accessed 2023-11-17]
39. Women’s perspective for strengthening disaster response capabilities. Gender Equality Bureau Cabinet Office. URL: https://www.gender.go.jp/policy/saigai/fukkou/pdf/pamphlet_en.pdf [accessed 2023-11-17]
40. LunaLuna. URL: https://sp.lnln.jp/brand/services/lnln_services [accessed 2023-11-17]
41. Bliudzius A, Svaikeviciene K, Puronaite R, Kasiulevicius V. Physical activity evaluation using activity trackers for type 2 diabetes prevention in patients with prediabetes. Int J Environ Res Public Health 2022 Jul 6;19(14):8251. [doi: 10.3390/ijerph19148251] [Medline: 35886100]

XSL•FO
RenderX

42. Lima FV, Kadiyala V, Huang A, et al. At the crossroads! time to start taking smartwatches seriously. Am J Cardiol 2022 Sep 15;179:96-101. [doi: 10.1016/j.amjcard.2022.06.020] [Medline: 35842279]
43. Magal N, Rab SL, Goldstein P, Simon L, Jiryis T, Admon R. Predicting chronic stress among healthy females using daily-life physiological and lifestyle features from wearable sensors. Chronic Stress (Thousand Oaks) 2022;6:24705470221100987. [doi: 10.1177/24705470221100987] [Medline: 35911618]
44. Schmidt LI, Jansen CP, Depenbusch J, Gabrian M, Sieverding M, Wahl HW. Using wearables to promote physical activity in old age: feasibility, benefits, and user friendliness. Z Gerontol Geriatr 2022 Aug;55(5):388-393. [doi: 10.1007/s00391-022-02083-x] [Medline: 35849158]
45. Kang HS, Exworthy M. Wearing the future-wearables to empower users to take greater responsibility for their health and care: scoping review. JMIR Mhealth Uhealth 2022 Jul 13;10(7):e35684. [doi: 10.2196/35684] [Medline: 35830222]
46. Hoshide S, Nishizawa M, Okawara Y, et al. Salt intake and risk of disaster hypertension among evacuees in a shelter after the Great East Japan Earthquake. Hypertension 2019 Sep;74(3):564-571. [doi: 10.1161/HYPERTENSIONAHA.119.12943] [Medline: 31280649]

Abbreviations

GEJE: Great East Japan Earthquake

Edited by A Mavragani; submitted 28.01.24; peer-reviewed by A Bodini, J Sands; revised version received 06.01.25; accepted 16.01.25; published 05.03.25.

Please cite as:
Nonaka S, Fujii S, Yamada KC
Purposes of Smartphone Health Care Apps and the Practicality of Their Functions in Disaster Situations: Qualitative Function Assessment Study
JMIR Form Res 2025;9:e56862
URL: https://formative.jmir.org/2025/1/e56862
doi:10.2196/56862

© Sayuri Nonaka, Susumu Fujii, Kosuke Chris Yamada. Originally published in JMIR Formative Research (https://formative.jmir.org), 5.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Assessment of the Sensitivity of a Smartphone App to Assist Patients in the Identification of Stroke and Myocardial Infarction: Cross-Sectional Study

Amar Dhand[1,2], MD, DPhil; Rama Mangipudi[3], BPharm, MSc; Anubodh S Varshney[4], MD; Jonathan R Crowe[5,6], MSc, MD, MPH; Andria L Ford[7], MD; Nancy K Sweitzer[8], MD, PhD; Min Shin[9], PhD; Samuel Tate[9], MS; Haissam Haddad[3], MD; Michael E Kelly[10], MD, PhD; James Muller[11], MD; Jay S Shavadia[3], MBChB

[1]Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, 65 Landsdowne Street, Cambridge, MA, United States

[10]Division of Neurosurgery, Department of Surgery, University of Saskatchewan, Saskatoon, SK, Canada

[11]Division of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

[2]Network Science Institute, Northeastern University, Boston, MA, United States

[3]Division of Cardiology, Department of Medicine, Unversity of Saskatchewan, Saskatoon, SK, Canada

[4]Division of Cardiovascular Medicine, Department of Medicine, Stanford University, Palo Alto, CA, United States

[5]Department of Neurology, University of Virgina, Charlottesville, VA, United States

[6]Department of Public Health Sciences, University of Virgina, Charlottesville, VA, United States

[7]Department of Neurology, Washington University School of Medicine, St. Louis, MO, United States

[8]Division of Cardiology, Department of Medicine, Washington University School of Medicine, St. Louis, MO, United States

[9]Department of Computer Science, University of North Carolina at Charlotte, Charlotte, NC, United States

Corresponding Author:
Amar Dhand, MD, DPhil
Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, 65 Landsdowne Street, Cambridge, MA, United States

Abstract

Background: Most people do not recognize symptoms of neurological and cardiac emergencies in a timely manner. This leads to delays in hospital arrival and reduced access to therapies that can open arteries. We created a smartphone app to help patients and families evaluate if symptoms may be high risk for stroke or heart attack (myocardial infarction, MI). The ECHAS (Emergency Call for Heart Attack and Stroke) app guides users to assess their risk through evidence-based questions and a test of weakness in one arm by evaluating finger-tapping on the smartphone.

Objective: This study is an initial step in the accuracy evaluation of the app focused on sensitivity. We evaluated whether the app provides appropriate triage advice for patients with known stroke or MI symptoms in the Emergency Department. We designed this study to evaluate the sensitivity of the app, since the most dangerous output of the app would be failure to recognize the need for emergency evaluation. Specificity is also important, but the consequences of low specificity are less dangerous than those of low sensitivity.

Methods: In this single-center cross-sectional study, we enrolled patients presenting with symptoms of possible stroke or MI. The ECHAS app assessment consisted of a series of evidence-based questions regarding symptoms and a test of finger-tapping speed and accuracy on the phone's screen to detect unilateral arm weakness. The primary outcome was the sensitivity of the ECHAS app in detecting the need for ED evaluation. The secondary outcome was the sensitivity of the ECHAS app in detecting the need for hospital admission. Two independent and blinded board-certified physicians reviewed the medical record and adjudicated the appropriateness of the ED visit based on a 5-point score (ground truth). Finally, we asked patients semistructured questions about the app's ease of use, drawbacks, and benefits.

Results: We enrolled 202 patients (57 with stroke and 145 with MI). The ECHAS score was strongly correlated with the ground truth appropriateness score (Spearman correlation 0.41, $P<.001$). The ECHAS app had a sensitivity of 0.98 for identifying patients in whom ED evaluation was appropriate. The app had a sensitivity of 1.0 for identifying patients who were admitted to the hospital because of their ED evaluation. Patients completed an app session in an average of 111 (SD 60) seconds for the stroke pathway and 60 (SD 33) seconds for the MI pathway. Patients reported that the app was easy to use and valuable for personal emergency situations at home.

XSL•FO
RenderX

Conclusions: The ECHAS app demonstrated a high sensitivity for the detection of patients who required emergency evaluation for symptoms of stroke or MI. This study supports the need for a study of specificity of the app, and then a prospective trial of the app in patients at increased risk of MI and stroke.

(JMIR Form Res 2025;9:e60465) doi:10.2196/60465

KEYWORDS

time-to-treatment; digital health; patient-centered care; eHealth apps; smartphones; apps; cross-sectional study; stroke; myocardial infarction; heart attack; neurology; neurological assessment; accuracy; emergency department; finger-tapping test; hospital admission; physicians; medical records; mobile health; mhealth; symptom recognition; mobile phone

Introduction

Timely recognition and intervention in neurological and cardiac emergencies can prevent severe complications and improve survival rates. Prehospital care is emerging as a critical window for deploying innovative strategies to enhance early recognition, triage, and transport in these emergencies [1]. Prehospital care is particularly important in ischemic stroke and myocardial infarction (MI), where the time from symptom onset to treatment profoundly affects prognosis [2,3]. This is particularly true in marginalized populations, where prehospital delay is a key driver of disparities in outcomes [4,5]. Thus, addressing these delays offers broad benefits for public health, economic savings, and equity in health care [6].

The urgency in addressing delays in care for patients with stroke and MI is underscored by a wealth of literature emphasizing the “golden hour”—the initial period when medical intervention is most effective [7,8]. For stroke, thrombolytic or thrombectomy therapy within this critical window reduces the extent of brain damage and improves functional outcomes [9,10]. Similarly, for MI, early reperfusion therapy is crucial in reducing infarct size and mortality [11-13]. However, despite large-scale public health campaigns and education efforts, the majority of patients fail to recognize symptoms as life-threatening, leading to delayed presentations to the emergency department (ED) and missed opportunities for early intervention [2,14,15].

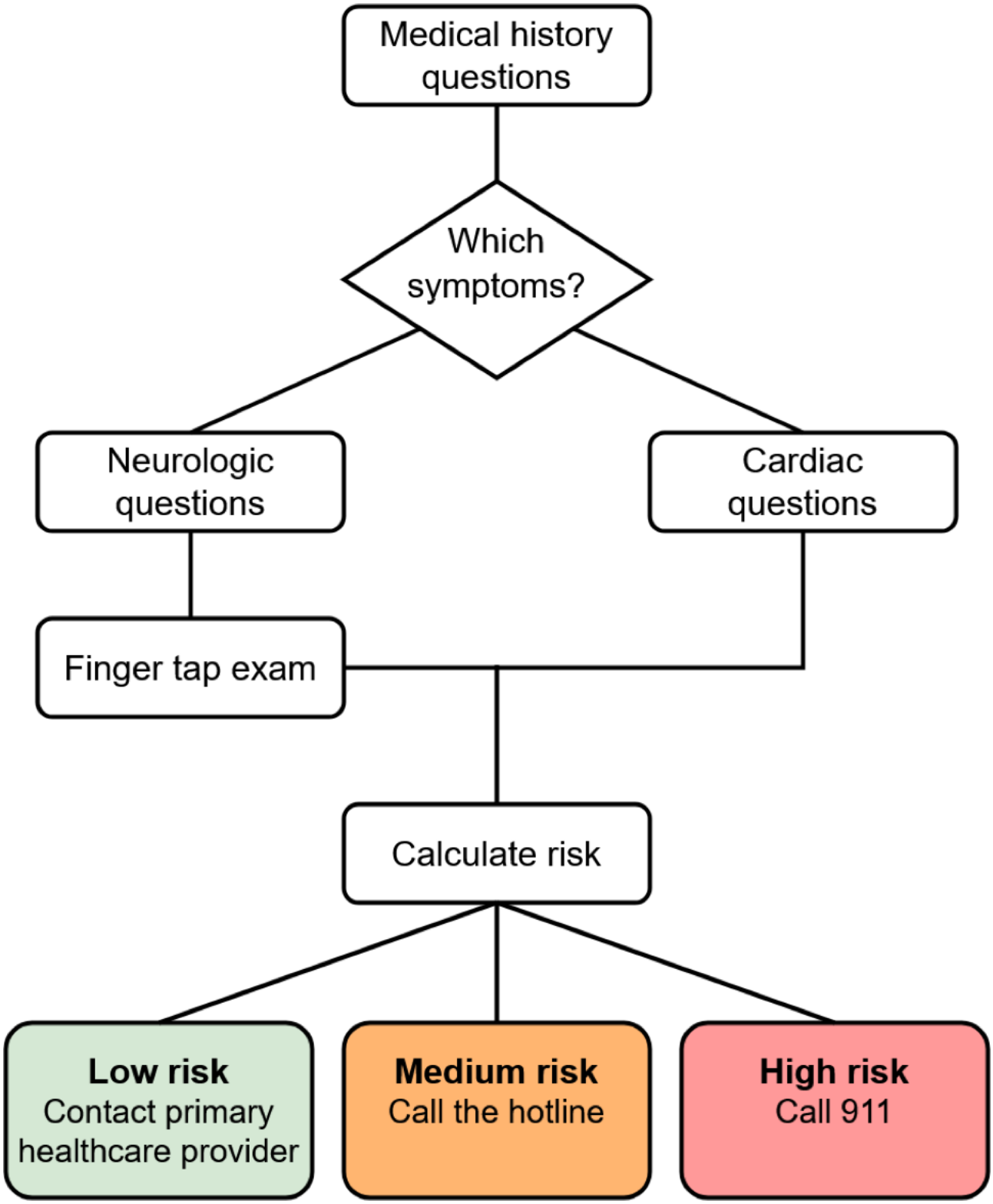
In addition to patient delay, patients often make mistakes in triage decisions. Some never seek the needed emergency evaluation, while others visit an ED for symptoms that do not require emergency evaluation [16]. The initial problem is one of sensitivity and the latter is one of specificity. For screening tests, it is important to first prioritize sensitivity before specificity. This is because not recognizing the need for emergency evaluation could lead to dangerous outcomes (eg, death or morbidity) which is more serious than overdiagnosis. We acknowledge the latter is also important for any decision support tool to be useful before final deployment [17].

The ECHAS (Emergency Call for Heart Attack and Stroke) smartphone app is an innovative approach to bridge this gap by assisting patients in identifying symptoms of neurological and cardiac emergencies and facilitating accurate triage and a quicker intervention (Figure 1). Leveraging the ubiquity of smartphones, ECHAS aims to be a scientifically validated and regulatory-approved digital medicine technology. The app is modeled on the “history and examination” of a neurologist or cardiologist by asking a series of evidence-based questions about the user’s medical history and symptoms, as well as a finger-tapping test designed to detect unilateral weakness. Based on the user or family responses, the app calculates a risk score and recommends one of the following three actions: (1) call 911, (2) call a medical hotline, or (3) call a primary care physician now.

In the first of a series of validation steps, this study evaluated the sensitivity and usability of ECHAS in appropriately triaging patients who have already sought care in the ED for symptoms of a possible stroke, MI, or related conditions.

XSL•FO
RenderX

Figure 1. ECHAS (Emergency Call for Heart Attack and Stroke) flowchart of "history and examination" functions for symptoms concerning cardiac or stroke emergencies.



Methods

Study Design and Participants

This was a cross-sectional study at the Royal University Hospital in Saskatoon, Canada between June 2022 and July 2023. We recruited eligible patients in the ED once their initial evaluation was complete, and they were in stable condition.

The inclusion criteria for the study were patients who presented to the ED for evaluation of a possible stroke or MI. These patients were identified for exhibiting symptoms that warranted an "Stroke alert" or "MI alert" as judged by standard clinical protocol. The symptoms indicative of a stroke alert included slurred speech, asymmetric weakness or numbness, balance problems, vision changes, and headache. The symptoms that triggered an MI alert included chest discomfort, chest pressure

XSL•FO
RenderX

or pain, palpitations, shortness of breath, lightheadedness or presyncope, and syncope. In addition, eligible participants were required to be 18 years or older and capable of providing informed consent.

Patients were excluded from eligibility if they were non-English speaking, had a documented history of dementia, had severe visual impairment, were unable to use a smartphone (due to physical, cognitive, or other reasons), or were unable to provide consent. A single trained research staff member enrolled all patients.

Regarding sample size, we determined the sample size a priori based on statistical and practical considerations. We estimated that a sample size of 200 patients was sufficient to provide power to detect small to medium effect sizes, ensuring robust statistical analysis of sensitivity of the app. In addition, this number was deemed feasible for a single-center study within the constraints of the available budget and study period.

The sampling method was consecutive sampling. It involved including all patients who met inclusion criteria sequentially, as they presented within a defined period, until the desired sample size was reached. This minimized selection bias. For recruitment, the research team screened eligible participants through daily chart review. We then collaborated with nurse clinicians in the ED, cardiology, and neurology units to identify potential participants. We also displayed a poster in these nursing units to assist nurses and patients to understand the study. The research team requested the bedside nurse to gauge patient interest. If patient was interested, the research team approached them at the bedside. The research team member introduced the study, obtained informed written consent, and then began collecting data.

Ethical Considerations

We followed a protocol approved by the University of Saskatchewan Research Ethics Board (REB Bio #3380). All patients provided written informed consent (Multimedia Appendix 1) to the study. Regarding privacy, all data collected in the app was deidentified. Specifically, the responses to the questions and the finger-tapping amounts were not connected to the user. For compensation, participants did not receive any financial payment for their involvement in the study. We ensured that no individual participants may be identified in this paper.

ECHAS Description

All eligible patients completed a single session of ECHAS on a study-standardized Apple iPhone 13. Figure 2 displays screenshots of ECHAS. A single session included questions of symptoms and phone sensor–based examination of unilateral weakness for the stroke pathway. Participants were instructed to answer questions about symptoms based on what they experienced before coming to the ED. They answered yes or no questions about these symptoms and their previous history aligned with standard of clinical care (refer to Multimedia Appendix 2 for list of questions). Patients were asked about symptoms occurring during their decision to seek emergency evaluation that had occurred hours before engaging with the app. In many instances, the symptoms that had brought them to the ED were still present at the time of data entry in the app.

XSL•FO
RenderX

Figure 2. Screenshots of ECHAS (Emergency Call for Heart Attack and Stroke) including stroke question, myocardial infarction question, finger-tapping, and final suggestion. We obtained explicit consent from these individuals to use their images for publication.

Patients in the stroke pathway additionally completed a phone sensor–based evaluation evaluating finger-tapping on the left and right hands. The app quantified the three following finger-tapping characteristics: (1) speed of finger-tapping for 10 seconds against literature-based standards, (2) comparison of finger-tapping speed on the left versus right hand, and (3) the accuracy of tapping in the middle of the target provided. Participants' responses to questions and their finger-tapping statistics were aggregated into an ECHAS score. The ECHAS score guided one of the following three triage decisions: (1) call 911, (2) call the hot line, or (3) call your primary care physician now.

Clinical End Points

The ground truth for the triage decision assessment was the appropriateness of the patient's visit to the ED as judged by a

XSL•FO
RenderX

retrospective review of the ED record performed by a board-certified neurologist and a board-certified cardiologist using pre-established criteria (Table 1). Each physician was blinded to the ECHAS score or triage decision and to the other adjudicator's decision. A positive (ie, considered reasonably appropriate) ED visit was defined as a patient whose Appropriateness Scale Score was ≥3. A negative (ie, considered inappropriate) ED visit was defined as a patient whose Appropriateness Scale Score was ≤2.

XSL•FO
RenderX

Table . The Appropriateness Scale for emergency department evaluation for potential neurological and cardiac emergency.

Appropriateness Scale score	Definition
Neurological	
1	No neurological imaging or neurology consultation
2	CT[a] head performed
3	Neurology consultation completed
4	Any of the following: • CTA[b] or CTV[c] head and neck performed • MRI[d] brain performed • MRA[e] or MRV[f] head and neck performed • Invasive angiogram performed • Lumbar puncture performed • Admitted to observation unit or hospital admission for noncerebrovascular diagnosis
5	Any of the following: • Diagnosed with stroke, TIA[g], ICH[h], epidural hematoma, SDH[i], SAH[j] • Admitted to hospital for a cerebrovascular diagnosis
Cardiac	
1	No ECG[k] taken nor troponin value measured
2	ECG taken or single troponin value measured
3	ECG taken, ≥2 troponin values measured
4	• ECG taken, ≥2 troponin values measured PLUS: • Coronary CTA performed • TTE[l] performed • Stress test performed • Invasive coronary angiography performed OR (no ECG or troponin required) • Arrhythmia requiring cardioversion • Admission to observation unit or hospital admission for a noncardiovascular diagnosis
5	Any of the following: • Diagnosed with ST-segment elevation MI[m] (STEMI), Non-ST elevation-acute coronary syndrome (NSTE-ACS), acute PE[n], acute aortic dissection, cardiac arrest, or cardiogenic shock • Admitted to hospital for a cardiovascular diagnosis

[a]CT: computed tomography.
[b]CTA: computed tomography angiography.
[c]CTV: computed tomography venography.
[d]MRI: magnetic resonance imaging.
[e]MRA: magnetic resonance angiography.
[f]MRV: magnetic resonance venography.
[g]TIA: transitory ischemic attack.
[h]ICH: intracerebral hemorrhage.
[i]SDH: subdural hematoma.
[j]SAH: subarachnoid hemorrhage.

XSL•FO
RenderX

[k]ECG: electrocardiogram.

[l]TTE: transthoracic echocardiography.

[m]MI: myocardial infarction.

[n]PE: pulmonary embolism.

The primary outcome was the sensitivity of the ECHAS app in detecting patients who needed ED evaluation. We compared the 2 highest acuity triage decisions of the app (call 911 or call the hotline) versus a positive classification (Appropriateness Scale Score ≥3) by the blinded physician adjudicators. We also calculated the sensitivity of ECHAS app to predict admission to the hospital for workup (Appropriateness Scale Score cutoff of 5 as positive and ≤4 as negative).

In the usability study, we determined the time to complete the ECHAS assessment and recorded the feedback of patients on the usability of the app. We recorded patient's feedback both in the form of 2-choice or 5-choice Likert scales to measure opinions and semistructured interview responses.

Statistical Analysis

We first completed descriptive analyses of the ECHAS score and the Appropriateness Scale Score. Then, we assessed their relationship using scatterplots, Spearman's correlation, and unadjusted linear regression. Finally, we created 2×2 tables of ECHAS versus the ground truth and calculated the sensitivity. The study was focused on sensitivity since failure to identify emergency cases is the most dangerous potential outcome of use of the app. The study population in this ED-based context had a high prevalence of serious neurological or cardiac conditions warranting ED evaluation. Specificity will be evaluated in a subsequent study in patients who did not seek ED evaluation.

For secondary outcomes, we calculated the mean and median time that patients took to complete the ECHAS session. We created bar plots of patients' responses on usability. Finally, we completed a qualitative analysis of the participants' interview responses. As the qualitative analysis was ancillary to the quantitative work, we followed a "small q qualitative research" approach [18]. Using single-participant interviews with fully structured, open-ended questions, we elicited brief responses, often single sentences. We coded for patterns to describe and summarize topics and categories [19]. Throughout the process, we engaged in reflexivity by discussing patterns as a group before, during, and after coding. These discussions ensured the confirmability, dependability, credibility, and transferability of our findings.

We used the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) Statement Guideline in reporting results (Checklist 1).

Results

Cohort Characteristics

The cohort examined in this study consisted of 202 enrolled participants, of which 200 had ECHAS scores available. In 2 patients, the scores were not recorded due to technical error. The distribution of patients by clinical presentation included 57 patients with stroke and 145 patients with MI. The overall sex distribution was 76 women and 126 men. The median age of the participants was 62 (IQR 51-67) years. Racial composition of the cohort comprised White (n=179), Asian (n=10), Black or African American (n=1), and Indigenous (n=12). In Table 2, we present the clinical characteristics stratified by illness including baseline medical, ED presentation, stroke, and MI data.

XSL•FO
RenderX

Table . Clinical characteristics of 202 study participants from Royal University Hospital emergency department in 2022 - 2023.

Variables		Presented to ED[a] for symptoms	
		Stroke (n=57)	MI[b] (n=145)
Baseline medical data			
Sex, n (%)			
	Male	25 (44)	102 (70)
	Female	32 (56)	44 (30)
Weight (kg), median (IQR)		86.0 (75.0-97.5)	90.0 (76.0-102.0)
Height (cm), median (IQR)		168.0 (163.0-175.0)	173.5 (165.0-180.3)
Race, n (%)			
	American Indian or Alaska Native	0 (0)	0 (0)
	Asian	0 (0)	10 (7)
	Black or African American	1 (1)	0 (0)
	Native Hawaiian or Other Pacific Islander	0 (0)	0 (0)
	White	53 (93)	127 (87)
	Other	3 (5)	9 (6)
Cigarette smoking, n (%)			
	Current (within last 30 days)	19 (33)	41 (28)
	Never	34 (60)	72 (50)
	Quit ≥1 month	4 (7)	32 (22)
Diabetes mellitus		9 (16)	30 (21)
Diabetes type, n (%)			
	Type 1	1 (11)	2 (7)
	Type 2	8 (89)	28 (93)
Medications taken to control blood sugars		9 (100)	23 (77)
Diabetes drug type, n (%)			
	Oral medication	7 (12)	20 (14)
	GLP-1[c] analogues	1 (2)	2 (1)
	Insulin	3 (5)	10 (7)
Dyslipidemia requiring treatment, n (%)		16 (28)	43 (29)
Hypertension requiring treatment, n (%)		19 (33)	65 (45)
Previous MIs, n (%)		2 (4)	26 (18)
Previous intervention within 30 days, n (%)		0 (0)	2 (1)
Type of intervention, n (%)			
	CABG[d]	0 (0)	0 (0)
	PCI[e] or stenting	0 (0)	2 (1)
	Hear ablation	0 (0)	0 (0)
	Heart valve surgery	0 (0)	0 (0)
Acute coronary syndrome, n (%)		3 (5)	28 (19)
Previous stroke, n (%)		6 (11)	7 (5)
Previous transient ischemic attack (TIA), n (%)		7 (12)	3 (2)

XSL•FO
RenderX

Variables		Presented to ED[a] for symptoms	
		Stroke (n=57)	MI[b] (n=145)
ED presentation data			
Hours from symptom onset to ED arrival, n (%)			
	Less than 3 hours	24 (43)	58 (40)
	3 to 6 hours	5 (9)	18 (12)
	More than 6 hours	27 (48)	69 (46)
Time from symptom onset to arrival, median (IQR)		340 (113-1140)	276 (85-2880)
Mode of transportation to ED, n (%)			
	EMS[f] from home or residence	13 (23)	29 (20)
	EMS from rehab or other health care facility	0 (0)	1 (0.7)
	Personal auto	39 (68)	110 (76)
	Taxi	0 (0)	0 (0)
	Transfer from other hospital	4 (7)	5 (3)
	Other	1 (2)	0 (0)
Stroke data			
Stroke severity by NIH[g] Stroke Scale, n (%)			
	Mild (0 - 4)	45 (79)	—[h]
	Moderate (5-14)	11 (19)	—
	Severe (15-42)	1 (2)	—
Finger-tapping assessment, n (%)			
	Both sides are normal speed and equal	39 (75)	—
	Right side is slow compared with left	6 (12)	—
	Left side is slow compared with right	7 (14)	—
	Both sides are slow	0 (0)	—
CT[i] results, n (%)			
	CT not done	10 (18)	—
	Normal	30 (53)	—
	Abnormal	17 (30)	—
MRI[j] results, n (%)			
	MRI not done	36 (63)	—
	Normal	9 (16)	—
	Abnormal	12 (21)	—
Source document to be submitted, n (%)			
	CT Report	46 (81)	—
	MRI Report	20 (35)	—
TPA given, n (%)		2 (4)	—
Mechanical thrombectomy completed, n (%)		1 (2)	—
MI data			

XSL•FO
RenderX

Variables		Presented to ED[a] for symptoms	
		Stroke (n=57)	MI[b] (n=145)
Cardiac biomarkers, n (%)			
	Troponin T	—	13 (9)
	Troponin I	—	17 (12)
	High sensitive troponin T	—	130 (89)
	Creatine kinase	—	121 (83)
	Not performed	—	11 (8)
Troponins test 1 result (ng/mL), median (IQR)		—	0.016 (0.007-0.057)
Troponins test 1 upper reference range, median (IQR)		—	0.014 (0.014-0.050)
Troponins test 2 result (ng/mL), median (IQR)		—	0.063 (0.016-0.711)
Troponins test 2 upper reference range, median (IQR)		—	0.014 (0.014-0.050)
Troponins test 3result (ng/mL), median (IQR)		—	0.083 (0.033-1.114)
Troponins test 3 upper reference range, median (IQR)		—	0.014 (0.014-0.050)
Creatine kinase result (U/L), median (IQR)		—	159.0 (94.0-1010.0)
ECG[k] changes, n (%)			
	Normal ECG during the event	—	49 (34)
	Abnormal	—	35 (24)
	ECG showed signs of new MI	—	34 (24)
	ECG showed signs of possible new MI	—	19 (13)
	ECG showed non-specific signs	—	7 (5)
Cardiac catheterization performed, n (%)		—	77 (53)
Coronary imaging performed, n (%)		—	12 (8)
Stress echo ordered, n (%)		—	3 (2)
New cardiac medications started, n (%)		—	80 (55)
Class of new cardiac medicines prescribed, n (%)			
	Antiplatelet	—	76 (52)
	Anticoagulation	—	74 (51)
	Lipid-lowering	—	30 (21)
	Antihypertensive	—	24 (16)

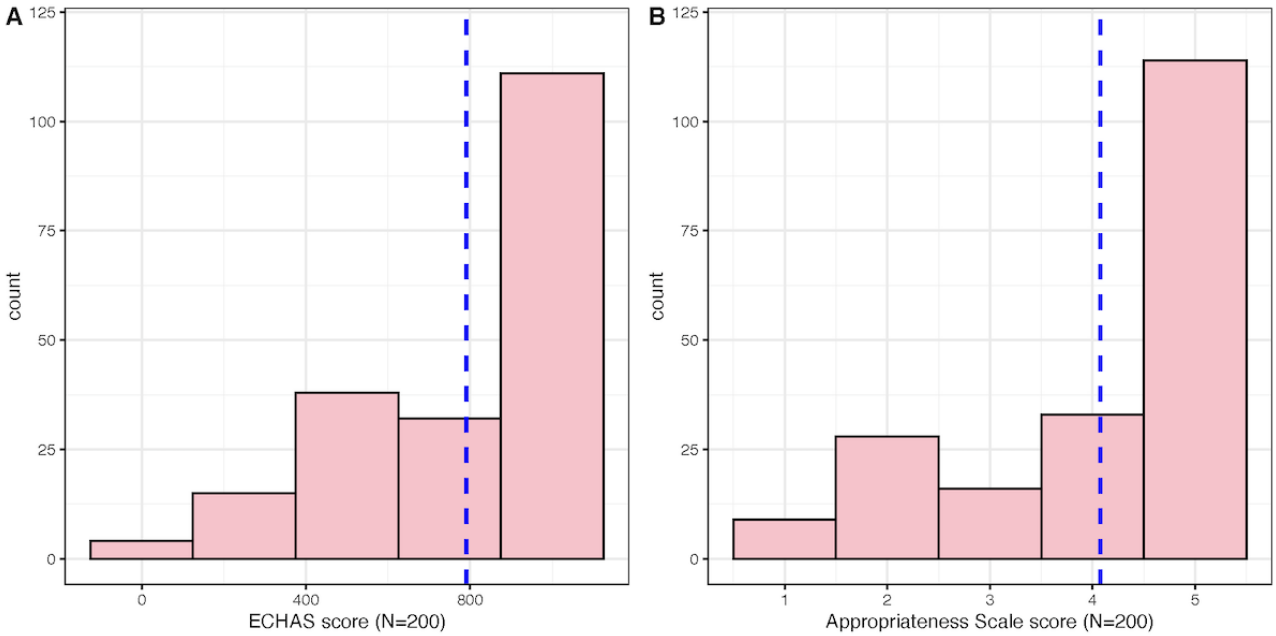
[a]ED: emergency department.
[b]MI: myocardial infarction.
[c]GLP-1: Glucagon-like peptide-1.
[d]CABG: coronary artery bypass graft.
[e]PCI: percutaneous coronary intervention.
[f]EMS: emergency medical services.
[g]NIH: National Institutes of Health.
[h]Not applicable.
[i]CT: computed tomography.
[j]MRI: magnetic resonance imaging.
[k]ECG: electrocardiogram.

XSL•FO
RenderX

ECHAS Score Versus Appropriateness Scale Scores

The distribution of scores for the ECHAS score and the Appropriateness Scale were similar, as illustrated in Figure 3. This similarity suggests alignment in the measurement of clinical outcomes by these 2 scales.

Figure 3. Distribution of the ECHAS (Emergency Call for Heart Attack and Stroke) score and appropriateness scale score (ground truth) with the median indicated by the blue line in 202 study participants.



We observed a significant correlation between the ECHAS score and the Appropriateness Scale Score, with a Spearman correlation coefficient 0.41 (P<.001). This pattern of correlation was consistent when stratified for patients in both the MI and stroke pathways. This association was confirmed in a simple linear regression, where the ECHAS β-coefficient was determined to be 0.002 (SE=0.003, P<.01), with a constant of 2.557.

The agreement of ECHAS scores with the ground truth, as adjudicated by 2 independent physicians, is presented in Table 3. This 2×2 table shows that the physicians adjudicated the cases had 97% agreement, suggesting low interobserver variability.

Table . Two-by-two tables of the ECHAS (Emergency Call for Heart Attack and Stroke) app triage advice versus adjudicators' determination of appropriateness of emergency department visit.

ECHAS app		Positive	Negative	Sensitivity	Specificity
Adjudicator 1				0.98	0.08
	Positive	159	34		
	Negative	4	3		
Adjudicator 2				0.97	0.07
	Positive	153	40		
	Negative	4	3		

The average sensitivity of ECHAS for appropriate triage was 0.98. The sensitivity for identifying patients admitted to the hospital for possible stroke or MI was 1.0, indicating that ECHAS detected all patients who were highest acuity. The average specificity of ECHAS for appropriate triage was 0.08.

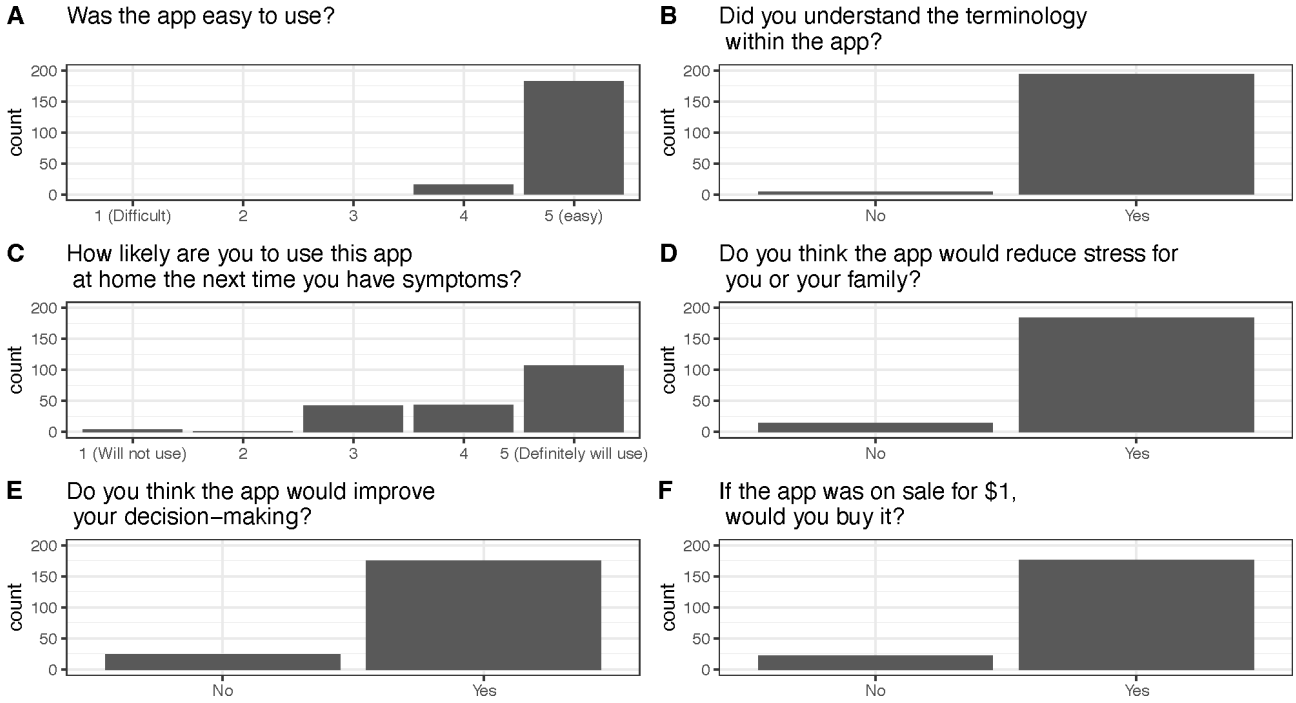
Secondary Outcomes

The average time to complete the assessment using ECHAS varied by 111 (SD 60) seconds for stroke and 60 (SD 33) seconds for MI.

Patients reported finding ECHAS usable, understandable, and worthwhile in acute emergencies at home, as shown in Figure 4. Specific comments from patients, which provide insights into their experiences and perceptions of the app, are detailed in Textbox 1.

XSL•FO
RenderX

Figure 4. Usability of the ECHAS (Emergency Call for Heart Attack and Stroke) app according to study participants.



Textbox 1. Patients' experiences and perceptions of the ECHAS (Emergency Call for Heart Attack and Stroke) after use in the Royal University Hospital emergency department.

App is helpful, useful, and positive

- *Good app. Very useful for delayers. Could save lives.*
- *I wish I had this app in my mobile 2 days back.*
- *The next time I experience any symptoms that I don't understand, I would use the app to help me make the correct decision.*
- *I would inform my colleagues that this app is available. I would use it myself if I was having chest pain again.*
- *Consider if vitals measurement could be done by the smart phone.*
- *If I have symptoms again I would use again.*

App as important within family networks

- *I would definitely buy the app for my family members.*
- *If the app could get the summary of all the questions answered at the end (like a PDF copy)- may be beneficial for EMS/ER to look at the responses to reduce the time for reevaluation or asking same questions.*
- *I would use the app whenever I thinks my father is having a possible heart attack or stroke.* [The patient's daughter]

Suggestions for improvement:

- *Don't know what to select. I have both severe headache and chest pain. Thought could be either of stroke or heart attack.*
- *Chest pain felt like a prior heart attack- so called 911. I would call 911 irrespective of the app.*
- *Don't remember my blood pressure.*
- *Not sure if I am having heart attack or stroke which one to select to begin.*

Discussion

Principal Results

In this initial validation study, we evaluated the sensitivity and usability of the ECHAS smartphone app in 202 patients presenting to an ED for symptoms of a possible stroke or MI. We found a significant correlation between ECHAS scores and the Appropriateness Scale scores, indicating the app was successful in identifying patients who should seek ED evaluation. We found that ECHAS demonstrated high sensitivity of 0.98, especially in identifying high-acuity stroke or MI patients, with a unanimous detection rate for those patients admitted. In addition, the app proved to be efficient, requiring on average 111 (SD 60) seconds for stroke, and 60 (SD 33) seconds for MI. Patient feedback highlighted the potential

XSL•FO
RenderX

usability and effectiveness of ECHAS in emergency situations at home.

These findings support continued development of the app which requires study of specificity and a prospective study. The specificity in this study of patients who were already on the MI or Stroke Alert pathway in the emergency room was low (0.08). This means that ECHAS did not detect patients who did not need more comprehensive cardiac workup (ECG taken, ≥2 troponin values measured or more advanced studies) or neurological workup (neurology consultation and advanced imaging). The low specificity could mean that the app would over-recommend ED visits leading to false positives. However, it is unknown how ECHAS would perform in otherwise healthy persons at home. This needs further prospective study. In this study, we deliberately prioritized sensitivity because the consequences of low sensitivity are more dangerous than those of low specificity.

Comparison With Previous Work

The development of ECHAS is informed by a communication and health behavior theory suggesting that mobile health technologies address constraints of space and time that hinder more traditional forms of mass media communication [20]. Specifically, mobile health tools have shown promise to enhance self-management, adherence to treatment, and health literacy among patients with chronic conditions [21]. Furthermore, certain mobile health solutions provide an opportunity for real-time monitoring and assessment, which is particularly valuable in emergency scenarios where every second counts [22]. The ECHAS study contributes to these trends in the literature by demonstrating initial accuracy of a potentially ubiquitous tool that combines a physician-informed assessment and sensor-based monitoring to empower patients and families to determine actions during a worrisome moment.

The ECHAS study adds to the evolving landscape of digital medicine technologies aimed at enhancing acute symptom detection and emergency medical system alerting [23,24]. For example, Wasselius et al [25] showed the potential of wearable accelerometers to detect unilateral arm paresis. In the Apple Heart Study, Perez et al [26] showed that a wearable algorithm could correctly identify atrial fibrillation in users who were notified of irregular pulses. Garcia et al [27] introduced a mobile app that detects key symptoms of stroke, such as smile asymmetry, speech difficulties, and arm weakness, showing effective early detection capabilities. Similarly Stroke Cognitive Medical Assistant by Khriyenko et al [28] used IBM Watson tools to help users detect face asymmetry, speech difficulties, and arm weakness. These advancements collectively underscore a significant shift toward integrating technology in the fight against acute medical emergencies.

Study Strengths

This study's strengths included a large sample size of patients in acute settings, the use of 2 board-certified and blinded adjudicators for end point determination, and the evaluation of ECHAS against the appropriateness of ED visits rather than just confirmed stroke or MI diagnoses. In addition, this collaborative effort between neurologists and cardiologists addresses the needs of patients with vascular risk factors who are at risk for both neurological and cardiac emergencies. Finally, these positive results were achieved with a sensor that only detects finger-tapping weakness. Future versions of ECHAS, being tested at the time of this study, include sensors for facial droop detection and slurred or incomprehensible speech detection. This should improve specificity. Together with finger-tapping weakness, ECHAS future versions would allow objective detection of face, arm, and speech abnormalities (FAST) on a smartphone.

Limitations

As an initial investigation of ECHAS, this study had limitations. Its cross-sectional design may introduce recall bias, affecting the accuracy of reported outcomes. In addition, the performance of the app, tested in a controlled clinical environment on a standardized phone, might not mirror its usability or accuracy in real-world, at-home scenarios during worrisome moments. The study was conducted in a single center with limited diversity in terms of demographics and disabilities. The version of ECHAS evaluated in this study included sensors limited to detecting unilateral weakness through finger tapping. As stated above, ECHAS versions now being tested also include sensors for detecting facial asymmetry and speech difficulties. The study's cohort composition limited the ability to assess specificity, which is the app's ability to correctly identify those without the condition. This is an important metric for avoiding false positives that requires further study. We are planning future prospective studies involving at-home app deployment to address these limitations and provide a clearer understanding of ECHAS' accuracy, specificity, and usability in real-life settings with lower risk cohorts. We also aim to study ECHAS across diverse populations to test accessibility and usability in a wide range of users.

Conclusions

In this study of 202 patients, we found that ECHAS was highly sensitive and easily usable in guiding patients to identify medical emergencies without input by health care personnel. The findings support further research to determine specificity and a study of the app's prospective use in home settings. Overall, this "history and examination" app adds to the burgeoning array of digital medicine tools designed to improve prehospital care, highlighting its potential to empower patients and enhance early intervention strategies.

Acknowledgments

This study was funded by ECHAS, LLC. The cofounders of ECHAS, LLC played an active role in the design of the study, analysis, interpretation of data, and writing of the manuscript. We acknowledge the support of Seed Investors, particularly JanaLee Cherneski and Ian Desai, who enabled this study.

XSL•FO
RenderX

Data Availability

The data sets generated and/or analyzed during this study are not publicly available due to their sensitive medical nature protected by HIPAA, but are available from the corresponding author on reasonable request.

The underlying code for ECHAS is not publicly available for proprietary reasons. We understand that this restricts replication and transparency. Therefore, we will make the code for ECHAS used in this study available to qualified researcher who provide a reasonable request to the corresponding author.

Authors' Contributions

AD analyzed and interpreted the data and wrote the first draft of the manuscript. RM and JSS collected the data and were major contributors to the manuscript writing. ASV and JC adjudicated all data, adjudicated all cases to determine the Appropriateness Score, and were major contributors to the manuscript writing. ASV, JC, AF, NKS, MS, ST, HH, MEK, JM interpreted the data and provided substantial revisions to the data. ST, MS, and AD developed the smartphone app. All authors read and approved the final manuscript.

Conflicts of Interest

Authors AD and JM are co-founders of ECHAS, LLC. AD, JM, and MS hold shares in ECHAS, LLC and declare a financial interest. Authors ASV, JC, and ST have acted as paid consultants for ECHAS, LLC and declare a financial interest. All other authors declare no financial or non-financial competing interests.

Multimedia Appendix 1
Informed consent form.
[PDF File, 209 KB - formative_v9i1e60465_app1.pdf]

Multimedia Appendix 2
List of ECHAS questions.
[PDF File, 72 KB - formative_v9i1e60465_app2.pdf]

Checklist 1
STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist.
[PDF File, 154 KB - formative_v9i1e60465_app3.pdf]

References

1. Fladt J, Ospel JM, Singh N, Saver JL, Fisher M, Goyal M. Optimizing patient-centered stroke care and research in the prehospital setting. Stroke 2023 Sep;54(9):2453-2460. [doi: 10.1161/STROKEAHA.123.044169] [Medline: 37548010]
2. Hoschar S, Albarqouni L, Ladwig KH. A systematic review of educational interventions aiming to reduce prehospital delay in patients with acute coronary syndrome. Open Heart 2020;7(1):e001175. [doi: 10.1136/openhrt-2019-001175] [Medline: 32201586]
3. Matsuo R, Yamaguchi Y, Matsushita T, et al. Association between onset-to-door time and clinical outcomes after ischemic stroke. Stroke 2017 Nov;48(11):3049-3056. [doi: 10.1161/STROKEAHA.117.018132] [Medline: 28974631]
4. Powell-Wiley TM, Baumer Y, Baah FO, et al. Social determinants of cardiovascular disease. Circ Res 2022 Mar 4;130(5):782-799. [doi: 10.1161/CIRCRESAHA.121.319811] [Medline: 35239404]
5. Cruz-Flores S, Rabinstein A, Biller J, et al. Racial-ethnic disparities in stroke care: the American experience: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2011 Jul;42(7):2091-2116. [doi: 10.1161/STR.0b013e3182213e24] [Medline: 21617147]
6. National Institutes of Health, National Institute of Neurological Disorders and Stroke. Inequities in Access and Delivery of Acute Stroke Care.: National Institutes of Health; 2022. URL: https://www.ninds.nih.gov/sites/default/files/documents/Inequities-in-Access-and-Delivery-of-Acute-Stroke-Care-A-Brain-Attack-Coalition-Report_508C.pdf [accessed 2025-01-23]
7. Hand M, Brown C, Horan M, Simons-Morton D. Access to timely and optimal care of patients with acute coronary syndromes — community planning considerations: a report by the National Heart Attack Alert Program. J Thromb Thrombolysis 1998;6(1):19-46. [doi: 10.1023/A:1008820104852]
8. Randhawa AS, Pariona-Vargas F, Starkman S, et al. Beyond the golden hour: treating acute stroke in the Platinum 30 minutes. Stroke 2022 Aug;53(8):2426-2434. [doi: 10.1161/STROKEAHA.121.036993] [Medline: 35545939]
9. Saver JL, Fonarow GC, Smith EE, et al. Time to treatment with intravenous tissue plasminogen activator and outcome from acute ischemic stroke. JAMA 2013 Jun 19;309(23):2480-2488. [doi: 10.1001/jama.2013.6959] [Medline: 23780461]
10. Saver JL, Goyal M, van der Lugt A, et al. Time to treatment with endovascular thrombectomy and outcomes from ischemic stroke: a meta-analysis. JAMA 2016 Sep 27;316(12):1279-1288. [doi: 10.1001/jama.2016.13647] [Medline: 27673305]

XSL•FO
RenderX

11. Moser DK, Kimble LP, Alberts MJ, et al. Reducing delay in seeking treatment by patients with acute coronary syndrome and stroke: a scientific statement from the American Heart Association Council on cardiovascular nursing and stroke council. Circulation 2006 Jul 11;114(2):168-182. [doi: 10.1161/CIRCULATIONAHA.106.176040] [Medline: 16801458]
12. Nallamothu BK, Normand SL, Wang Y, et al. Relation between door-to-balloon times and mortality after primary percutaneous coronary intervention over time: a retrospective study. Lancet 2015 Mar;385(9973):1114-1122. [doi: 10.1016/S0140-6736(14)61932-2]
13. Jollis JG, Granger CB, Zègre-Hemsey JK, et al. Treatment time and in-hospital mortality among patients with ST-segment elevation myocardial infarction, 2018-2021. JAMA 2022 Nov 22;328(20):2033-2040. [doi: 10.1001/jama.2022.20149] [Medline: 36335474]
14. Yoon CW, Oh H, Lee J, et al. Comparisons of prehospital delay and related factors between acute ischemic stroke and acute myocardial infarction. J Am Heart Assoc 2022 May 3;11(9):e023214. [doi: 10.1161/JAHA.121.023214] [Medline: 35491981]
15. Saczynski JS, Yarzebski J, Lessard D, et al. Trends in prehospital delay in patients with acute myocardial infarction (from the Worcester Heart Attack Study). Am J Cardiol 2008 Dec 15;102(12):1589-1594. [doi: 10.1016/j.amjcard.2008.07.056] [Medline: 19064010]
16. Newton EH. Addressing overuse in emergency medicine: evidence of a role for greater patient engagement. Clin Exp Emerg Med 2017 Dec;4(4):189-200. [doi: 10.15441/ceem.17.233] [Medline: 29306268]
17. Power M, Fell G, Wright M. Principles for high-quality, high-value testing. Evid Based Med 2013 Feb;18(1):5-10. [doi: 10.1136/eb-2012-100645] [Medline: 22740357]
18. Braun V, Clarke V. Toward good practice in thematic analysis: avoiding common problems and be(com)ing a knowing researcher. Int J Transgend Health 2023;24(1):1-6. [doi: 10.1080/26895269.2022.2129597] [Medline: 36713144]
19. Saldana J. The Coding Manual for Qualitative Researchers, 4th edition: SAGE Publications; 2021.
20. Glanz K, Rimer BK. Health Behavior: Theory, Research, and Practice, 5th edition: Jossey-Bass; 2015.
21. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52. [doi: 10.2196/jmir.3951] [Medline: 25803266]
22. Albahri OS, Albahri AS, Mohammed KI, et al. Systematic review of real-time remote health monitoring system in triage and priority-based sensor technology: taxonomy, open challenges, motivation and recommendations. J Med Syst 2018 Mar 22;42(5):80. [doi: 10.1007/s10916-018-0943-4] [Medline: 29564649]
23. Bat-Erdene BO, Saver JL. Automatic acute stroke symptom detection and emergency medical systems alerting by mobile health technologies: a review. J Stroke Cerebrovasc Dis 2021 Jul;30(7):105826. [doi: 10.1016/j.jstrokecerebrovasdis.2021.105826] [Medline: 33932749]
24. Varshney AS, Madias C, Kakkar R, Martin DT. Watching for disease: the changing paradigm of disease screening in the age of consumer health devices. J Gen Intern Med 2020 Jul;35(7):2173-2175. [doi: 10.1007/s11606-019-05626-y] [Medline: 31942673]
25. Wasselius J, Lyckegård Finn E, Persson E, et al. Detection of unilateral arm paresis after stroke by wearable accelerometers and machine learning. Sensors (Basel) 2021 Nov 23;21(23):7784. [doi: 10.3390/s21237784] [Medline: 34883800]
26. Perez MV, Mahaffey KW, Hedlin H, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917. [doi: 10.1056/NEJMoa1901183] [Medline: 31722151]
27. García L, Tomás J, Parra L, Lloret J. An m-health application for cerebral stroke detection and monitoring using cloud services. Int J Inf Manage 2019 Apr;45:319-327. [doi: 10.1016/j.ijinfomgt.2018.06.004]
28. Khriyenko O, Rönkkö K, Tsybulko V, Piik K. Stroke Cognitive Medical Assistant (StrokeCMA). GSTF J Comput 2018;6(1). [doi: 10.5176/2251-3043_6.1.112]

Abbreviations

ECHAS: Emergency Call for Heart Attack and Stroke
ED: emergency department
MI: myocardial infarction
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

XSL•FO
RenderX

Edited by A Mavragani; submitted 13.05.24; peer-reviewed by D Patel; revised version received 22.01.25; accepted 23.01.25; published 03.03.25.

Please cite as:
Dhand A, Mangipudi R, Varshney AS, Crowe JR, Ford AL, Sweitzer NK, Shin M, Tate S, Haddad H, Kelly ME, Muller J, Shavadia JS
Assessment of the Sensitivity of a Smartphone App to Assist Patients in the Identification of Stroke and Myocardial Infarction: Cross-Sectional Study
JMIR Form Res 2025;9:e60465
URL: https://formative.jmir.org/2025/1/e60465
doi:10.2196/60465

© Amar Dhand, Rama Mangipudi, Anubodh S Varshney, Jonathan R Crowe, Andria L Ford, Nancy K Sweitzer, Min Shin, Samuel Tate, Haissam Haddad, Michael E Kelly, James Muller, Jay S Shavadia. Originally published in JMIR Formative Research (https://formative.jmir.org), 3.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Text Messaging Interventions for Unhealthy Alcohol Use in Emergency Departments: Mixed Methods Assessment of Implementation Barriers and Facilitators

Megan O'Grady[1], PhD; Laura Harrison[2], MPH; Adekemi Suleiman[1], MBBS, MPH; Morica Hutchison[1], PhD, LMFT; Nancy Kwon[2,3], MD; Frederick Muench[4], PhD; Sandeep Kapoor[2,3], MD

[1]Department of Public Health Sciences, University of Connecticut School of Medicine, 263 Farmington Avenue, Farmington, CT, United States

[2]Northwell Health, New Hyde Park, NY, United States

[3]Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, United States

[4]Partnership to End Addiction, New York, NY, United States

Corresponding Author:
Megan O'Grady, PhD
Department of Public Health Sciences, University of Connecticut School of Medicine, 263 Farmington Avenue, Farmington, CT, United States

Abstract

Background: Many patients with unhealthy alcohol use (UAU) access health care in emergency departments (EDs). Scalable supports, such as SMS text messaging interventions, are acceptable and feasible to enhance care delivery for many health issues, including substance use. Further, SMS text messaging interventions have been shown to improve patient outcomes related to alcohol consumption (eg, reduced consumption compared to no intervention, basic health information, or drink tracking), but they are rarely offered in clinical settings.

Objective: This paper describes a mixed methods study using the Integrated Promoting Action on Research Implementation in Health Services (i-PARIHS) framework. The goal of this study was to use a stakeholder-engaged mixed methods design to assess barriers and facilitators to the implementation of SMS text messaging interventions for UAU in EDs with a focus on the recipient's characteristics, the innovation's degree of fit within the existing practice, and the unique nature of the inner and outer context.

Methods: This study was conducted in a large health system in the northeastern United States. We examined electronic health record data on alcohol screening in 17 EDs; surveyed 26 ED physician chairpersons on implementation feasibility, acceptability, and appropriateness; and interviewed 18 ED staff and 21 patients to understand barriers and facilitators to implementation. Interviews were analyzed according to the i-PARIHS framework to assess recipient characteristics, innovation degree of fit, and inner and outer context.

Results: Electronic health record data revealed high variability in alcohol screening completion (mean 73%, range 35% - 93%), indicating potential issues in identifying patients eligible to offer the intervention. The 26 ED chair surveys revealed a relatively high level of implementation confidence (mean 4, SD 0.81), acceptability (mean 4, SD 0.71), and appropriateness (mean 3.75, SD 0.69) regarding the UAU SMS text messaging intervention; feasibility (mean 3.5, SD 0.55) had the lowest mean, indicating concerns about integrating the text intervention in the busy ED workflow. Staff were concerned about staff buy-in and adding additional discussion points to already overwhelmed patients during their ED visit but saw the need for additional low-threshold services for UAU. Patients were interested in the intervention to address drinking and health-related goals.

Conclusions: ED visits involving UAU have increased in the United States. The results of this formative study on barriers and facilitators to the implementation of UAU SMS text messaging interventions in EDs indicate both promise and caution. In general, we found that staff viewed offering such interventions as appropriate and acceptable; however, there were concerns with feasibility (eg, low alcohol risk screening rates). Patients also generally viewed the SMS text messaging intervention positively, with limited drawbacks (eg, slight concerns about having time to read messages). The results provide information that can be used to develop implementation strategies that can be tested in future studies.

Trial Registration: ClinicalTrials.gov NCT05350878; http://clinicaltrials.gov/ct2/show/NCT05350878

(JMIR Form Res 2025;9:e65187) doi:10.2196/65187

KEYWORDS

unhealthy alcohol use; text messaging intervention; emergency department; barriers and facilitators; implementation; alcohol use; unhealthy; mixed methods assessment; mixed methods; assessment; facilitators; alcohol; health care; scalable supports; text

XSL•FO
RenderX

messaging; patient outcomes; health system; electronic health record; EHR; alcohol screening; acceptability; feasibility; survey; health-related goals

Introduction

Unhealthy alcohol use (UAU) costs the United States nearly US $250 billion per year [1], and alcohol-related deaths among those aged 16 years and older recently increased by 25% [2]. Emergency departments (EDs), where alcohol-related visits are rising [3-5], are sometimes the only health care touchpoint for patients with UAU, making it a promising point of intervention [6,7]. SMS text messages are one of the most ubiquitous and salient modes of mobile health interventions. They use brief, supportive messages to promote behavior change without significant user burden [8-10] and are acceptable and feasible to enhance care delivery for many health issues, ranging from high blood pressure to substance use [9,11-16]. A meta-analysis and study indicated that alcohol-targeted texting interventions reduced alcohol consumption compared with no intervention, basic health information, or drink tracking [17,18]. Notably, UAU text interventions for ED and trauma patients are acceptable and effective in reducing drinking [19-22]. However, there is limited uptake of text-messaging interventions in ED settings.

Reviews of substance use technology interventions have noted minimal focus on implementation strategies or outcomes; rather, most studies focus only on patient clinical outcomes [23,24]. Thus, prioritizing the examination of barriers and facilitators to understand ways to increase uptake is an important next step [25]. This paper describes a mixed methods study using the Integrated Promoting Action on Research Implementation in Health Services (i-PARIHS) framework, which posits optimal implementation occurs when interactive problem-solving promotes acceptance and use of a new innovation based on the recipient's characteristics (eg, motivation), the innovation's degree of fit within the existing practice (eg, usability), and the unique nature of the inner (eg, culture) and outer context (eg, mandates) [26,27]. The examination of barriers and facilitators to implementing UAU text interventions in EDs [26,28] was done by examining electronic health records (EHRs), surveys, and interview data from a large health system using i-PARIHS elements known to influence the uptake of new innovations.

Methods

Study Population

The study was conducted in 17 adult-serving EDs in the northeastern US urban and suburban locations; it is part of a larger project examining SMS text messaging interventions for UAU [29].

Ethical Considerations

This study was approved by the institutional review board at the Feinstein Institutes for Medical Research (approval number 21-0756). All participants consented to study procedures using verbal consent processes and consent information sheets. All qualitative data were deidentified prior to analysis; survey data were anonymous. Patients received a US $15 gift card incentive; staff participants were not compensated. Generative artificial intelligence was not used in any part of the manuscript writing.

Procedures and Materials

We used an explanatory, sequential mixed methods approach (quantitative → qualitative) [30]. Data were integrated at several points. Quantitative EHR and survey data classified sites on low or high implementation potential; classifications were used for data collection and site selection for qualitative interviews and analysis. Data were integrated during result interpretation. Reporting is in accordance with the COREQ (Consolidated Criteria for Reporting Qualitative Research) and CHERRIES (Checklist for Reporting Results of Internet e-Surveys) checklists for qualitative health research involving interviews and internet e-surveys.

EHR Data

The 3-item AUDIT-C (Alcohol Use Disorders Identification Test-Consumption) [31] detects UAU (≥3 for women and ≥4 for men) and was built into all participating ED EHRs; it provided information on each ED's capacity to address UAU (eg, low screening completion rates may indicate a lack of time). Data, extracted from 2021, included percentage of patients screened (number of AUDIT-C completions divided by the total patient census).

Staff Surveys

ED physician chairpersons from the 17 participating EDs received an email from the study team inviting them to complete a 13-item web-based survey of three validated 4-item scales assessing conceptually distinct, but interrelated constructs often used in formative research to examine implementation barriers: feasibility (eg, "A text messaging intervention for UAU seems possible in my ED"), acceptability (eg, "I would welcome our clinical team offering a text messaging intervention for UAU to patients during their ED visit"), and appropriateness (eg, "A text messaging intervention for UAU seems fitting for patients in my ED") [32]. An additional confidence item ("I am confident that our clinical ED team would offer a text messaging intervention for UAU to patients during their ED visit") was included. All items were rated on a 5-point scale (1=completely disagree to 5=completely agree). To complete site classification, we calculated each site's survey totals (range 4 - 65; higher values indicate positive implementation outcomes). All invited EDs participated; in cases where both primary and secondary chairs completed the survey, scores were averaged.

Implementation Potential Classification

We classified low versus high implementation potential based on AUDIT-C completion rates and survey scores (low: ≤median on both measures; high: >median on both measures), resulting in 5 low sites and 4 high. Sites with mixed classifications were not recruited. We considered several factors (eg, census and location) to select the 4 final EDs for interviews (2 low and 2 high) using stratified purposeful sampling [33,34].

XSL•FO
RenderX

Staff Interviews

In the 4 final EDs, chairpersons were interviewed and then each identified staff. They were invited via emails from study staff, resulting in 18 (82%) out of 22 invited interviewies. Interviews lasted ~30 minutes, were conducted over Zoom, recorded, and cleaned for analysis. Interview guides covered views on alcohol screening, SMS text messaging interventions and how to best offer them, barriers and facilitators to intervention implementation, and general practice change.

Patient Interviews

Patients were recruited from 6 EDs; sites were selected based on location and census to gain perspectives from different populations and ED sizes (eg, race or ethnicity and urban or suburban). Patients were approached during their ED visit if they screened positive on the AUDIT-C. Of the 30 patients invited, 21 (70%) were interviewed. Interviews lasted ~20 minutes and were recorded and transcribed. The interview guide was adapted from previous ED studies (eg, mobile device use and interest in receiving supportive messages on UAU reduction) [35,36].

Qualitative Analysis

We used rapid qualitative analysis [37] to analyze interview data following procedures previously outlined [37,38]. Two researchers (MO and AS) summarized 3 transcripts from patient and staff interviews to ensure consistent summaries. One researcher (AS) completed the remaining summaries with the second researcher (MO) reviewing for accuracy and alignment; discrepancies were resolved via discussion. For patients, we created comparison groups between AUDIT-C scores indicative of moderate to high risk (score 4 - 7) versus severe risk (score 8 - 12) to examine response differences; for staff, we created comparison groups by low versus high implementation potential sites. See Multimedia Appendix 1 for demographic information for chair surveys (n=26), staff (n=18), and patient interviews (n=21).

Results

EHR and Chair Survey

Average AUDIT-C completion rates (73%) were relatively high but ranged greatly (35% - 93%). Survey implementation scores indicated a relatively high level of confidence, acceptability, and appropriateness regarding the UAU texting intervention; feasibility had the lowest mean, indicating logistical implementation concerns. See Multimedia Appendix 2 for absolute values on AUDIT-C completion rates and implementation scores.

Staff Interviews

Table 1 illustrates implementation barriers and facilitators. In comparing staff in low (n=7) versus high (n=11) implementation sites, low sites more frequently indicated addressing alcohol use was not a high priority. All staff discussed challenges to include information about the intervention in discharge papers, noting the paperwork is bulky and patients rarely read it. Facilitators were similar across high and low sites. Low sites discussed the difficulty of making changes, and that change takes time and requires strong justification. When asked about recent practice changes, low sites had fewer examples of what worked well and more frequently mentioned low staff morale, high turnover, and burnout.

XSL•FO
RenderX

Table . ED[a] staff qualitative interview barrier and facilitator themes from rapid analysis in accordance with i-PARIHS[b] domains.

Overall theme	Subtheme	Exemplar quote	i-PARIHS domain
Implementation barriers			
Screening views	There is not 100% buy-in from staff to do alcohol risk screening (eg, feeling that patients are not truthful on screenings, that the threshold is too low for "positive," that there is screening fatigue, or that patients may be surprised by being asked).	"...We try to have as good of a completion as possible...there isn't a one hundred percent buy in from the staff when it comes to it...a lack of an education, because the staff is in our population as well...when you work in an emergency room, and you see a lot of crazy things, and you deal with a lot of crazy health issues, somebody comes in for one thing, you're not necessarily too concerned about doing a screening kind of unrelated to the main diagnosis. But I think that is part is the main reason why completion rates aren't one hundred percent..."	Context/inner (culture/climate)
Screening process	Physicians or other providers generally do not view alcohol screening results.	"It's kind of cumbersome, unless it's actually pertinent to the case...We're juggling so much. It's a little too much for us to handle."	Context/inner (culture/climate)
Need for services	There was a focus mostly on the need for services for those with more severe problems who are frequent visitors to the ED.	"I think that it could be very helpful for a certain set of patients...the patients that keep coming back that we've tried to help...And yet, a week later they back."	Recipients (motivation/beliefs)
ED as an appropriate setting for addressing alcohol use	The ED is viewed as a difficult environment to address problems (lack of privacy and acute health problems) and have conversations beyond screening, doing more falls to low mixed priority due to staff workload/burden unless the patient is there for an alcohol issue.	"...I think it really depends on if the patient is there, specifically for the alcohol issue...In addition to the medical aspect, we try to make it a priority. But if a patient is identified in triage,...needing an SBIRT consult and they're not there for then it's not a priority it just becomes another thing in the chart that you know we just don't have time to address, unfortunately."	Context/inner (culture/climate)
SMS text messaging intervention usefulness for ED patients	The text intervention was viewed as potentially less useful for patients with severe alcohol problems, less motivated to change their drinking, or certain populations (eg, older adults and low mobile phone users).	"To be honest, I don't think it would be that effective to the majority, the patients that we see like the so-called frequent flyers, or the one-time binge drinker kind of deal. They're not going to be interested in that..."	Innovation (degree of fit)
Reasons patients may or may not accept the text intervention	Some patients may not have mobile phones/texting plans or may have lower motivation to change, which may deter from accepting a text intervention.	"...Some of them don't have cell phones or some of them have like prepaid cell phones...I honestly don't think it would be helpful."	Innovation (degree of fit)

XSL•FO
RenderX

Overall theme		Subtheme	Exemplar quote	i-PARIHS domain
	Offering the text intervention to patients	Patients are overwhelmed during the ED visit and receive a lot of discharge instructions, so text intervention information could get lost.	"...whereas the doctor might put it on the discharge paper...who reads them? They're twelve pages. It's going in the garbage. No one is really you know paying attention to it."	Innovation (degree of fit)
	Barriers to implementing a text intervention in the ED	It was expressed that physicians and nurses do not have the time to discuss alcohol use with patients, and that there are too many screens already being done, and the EHR[c] is already full of things to click.	"...It would be tough to operationalize that in a busy...any ED but...to have that as another initiative for our PAs, attendings for cases that are not necessarily for that, may be difficult to operationalize. It's not impossible I just don't know if it would be feasible in every single case."	Recipients (time, resources, support)
	Staff training needs	Staff are already overwhelmed with training; it would be difficult to add more training modules for a new intervention.	"...We get so many emails and have so many required modules that we need to watch that I think that those would probably not be the most effective way..."	Recipients (time, resources, support)
	General ease of making practice changes in ED	For something new to be implemented, staff must see value for the patient, and it must fit into the workflow without adding a burden.	"...when the staff buy into it and understand why we're doing it for the patient. I think that helps. They need to understand why we're implementing change."	Context/inner (innovation/change)
	Reflections on recent practice changes	It was perceived that change takes a long time based on recent experiences in implementing new practices.	"...I think it could be challenging to make changes. Especially if it requires like extra work. I think for people, it takes time or some of them just will never do it that they're just not complying with the changes. But I think it does take time and people to get used to the idea."	Context/inner (innovation/change)
	Culture of innovation	Staff morale and burnout are still recovering from COVID and some long-time staff are not as open to change; this creates some ambivalence about making changes.	"...burnt out, probably from the most part, I think most staff is burnt still from Covid, and is still in a recovery phase in general, because volume is kind of coming back, everybody is a little deconditioned in terms of volume, because we had very low levels, and we had high levels. So I think there's a level of burnout out there that we're constantly battling that's probably the biggest challenge that we're having at the moment."	Context/inner outer (culture/climate)
Implementation facilitators				

XSL•FO
RenderX

Overall theme		Subtheme	Exemplar quote	i-PARIHS domain
	Views on alcohol screening	The ED is a good place for screening because alcohol-related health issues in the ED are common, and it is beneficial information for the clinical team because the problem may not be identified or visible otherwise; it helps to normalize and destigmatize the issue of substance use.	"...I think it's a pretty good place. Um, Ah to identify um people who have alcohol abuse because, like yeah, like, I said, we do have like a nice, decent amount of people that do come in..."	Context/inner (culture/climate)
	Views on the alcohol screening process	Most patients are being screened. AUDIT[d] screening tool is built into EHR, and most patients are initially screened by triage/primary nursing and followed up via social work or screening, brief intervention, and referral to treatment (SBIRT[e]) health coaches.	"AUDIT screening tool is built into EMR, and most patients are initially screened by triage/primary nursing and followed up via social work or SBIRT."	Context/inner (culture/climate)
	The need for services for people with alcohol problems	It was viewed that the prevalence of those with alcohol problems in the ED is high; they see people across the continuum of use.	"...I see people on two ends of the spectrum right. Either they're completely intoxicated, and they're over utilizing the system for a bed to sleep, or they're coming in from like a bar or a nightclub overly intoxicated. Or I see the other end of this spectrum, which is somebody that's in severe withdrawal, delirium tremens. I've seen people that need to be admitted to the ICU setting or a hospital monitored bed because that's the normal vital signs risk procedure...Then you also see the other complications of alcohol abuse, such as like pancreatitis, for example, That's...the majority of the cases that I interact with."	Recipients (motivation/beliefs)
	ED as an appropriate setting for addressing alcohol use	The ED is a good place to identify problems and get the conversation started. The ED is a key access point that sees a high volume of people.	"I think it's a good place to identify...I think there are resources around that we can pull from. But again, if there's not a coach or somebody here, we struggle with where to refer people to."	Context/inner (culture/climate)

XSL•FO
RenderX

Overall theme		Subtheme	Exemplar quote	i-PARIHS domain
	SMS text messaging intervention usefulness for ED patients	It was viewed that there was no downside to offering the text intervention, even with a potentially small/incremental benefit, and patients may be more inclined because it is something they can engage with after an ED visit and because it is automated.	"I think it would be helpful for a patient that would be willing and wanting to need services...it really just offers another resource. I think nowadays one of the many issues that are occurring is that patients don't have access to resources. So I think any step that we can do to help provide other resources to patients that are in need, would be helpful. Something in real time or something where they don't have to wait on the phone don't have to dial number can just kind of use their phone and text would be definitely helpful. I think it would be good."	Innovation (degree of fit)
	Reasons patients may or may not accept the text intervention	The text intervention could be introduced to patients in a way that describes how it could be beneficial while ensuring privacy/confidentiality.	"...just letting them know that it's confidential, that it's helpful for them. They might, maybe someone of a clinical nature could tell them that they think it's important that they address this with someone, with a professional."	Innovation (degree of fit)
	Offering the text intervention to patients	Ways to facilitate offering the text intervention could be pagers/flyers with strong marketing, including it in discharge instructions, and using a team approach to offering. Materials plus a human touch would be beneficial.	"...The other way to do this is a passive way in terms of flyers...somebody could just opt in reading material while they are there...then in terms of explaining it is probably just a script so you give them a flyer and you can train probably a lot of support staff to say I know this is not the reason for your visit today we are working on this initiative, We give this to everybody..."	Innovation (degree of fit)
	Barriers to implementing a text intervention in the ED	Tapping into already existing workforce members (eg, social work, SBIRT, and referral coordinators) could be useful for offering the text intervention.	"...I find that nurses have more time to talk to patients, but I think it would have to be a team approach coming from the attending and working with the nurses...I think it would have to really be a team effort on everybody's part."	Recipients (time, resources, support)

XSL•FO
RenderX

Overall theme		Subtheme	Exemplar quote	i-PARIHS domain
	Barriers to implementing a text intervention in the ED	It could be useful to plug into already operating workflows or offer passive ways of offering the text intervention (eg, QR code).	"...like QR Codes for people to open up YouTube videos as distraction videos...But I think that places like the bathroom walls and provide to our codes for them to scan phone numbers and then champions. And as for health coaches and our social workers, and our case managers are often utilized to help with, uh disposition of those patients, right?"	Context/inner (innovation/change)
	Staff training needs	Brief training using familiar modalities would help with educating team members about the text intervention (eg, learning management system modules, morning huddles, Microsoft PowerPoint slides with all relevant information about the text program, and morning brief).	"...PowerPoint that you could always like reference or something that they could go back on. Um. I know a lot of the times like providers kind of say like, Oh, social work could do this...Some type of education um would be good, I think, including maybe, examples of what the text messages are, and then examples of like, If they did reach out to the program, what are they able to actually help with..."	Recipients (time, resources, support)
	General ease of making practice changes in ED	The ED is a setting used to change and is very adaptable. It values flexibility and change and is nimble.	"...I hate to say it depends on the change right so, yeah, but I think I think to answer your question a little easier our department is open to change pretty regularly we change things, most emergency departments do but we're pretty aggressive about being flexible, we pride ourselves sort of on that..."	Context/inner (innovation/change)
	Reflections on recent practice changes	EDs have a lot of recent experiences with change and methods to disseminate information and gain feedback and to have consistency in communication.	"...the changes with COVID testing and requirements and admission process...but I think what went well, people appreciate consistency, I think in communication, so I think what I appreciate now is that everyone knows, where to look for updates and things like that..."	Context/inner (innovation/change)

XSL•FO
RenderX

Overall theme	Subtheme	Exemplar quote	i-PARIHS domain
Culture of innovation	EDs are overall welcoming of innovation and embrace a culture of change; leadership and a cohesive team are important in change/innovation.	"There has been a lot of change here. It's always changing. They've received multiple certifications. (redacted). Um, They've made new rooms. They've gone through a renovation. They've gone through new leadership..."	Context/inner outer (culture/climate)

[a]ED: emergency department.

[b]i-PARIHS: Integrated Promoting Action on Research Implementation in Health Services.

[c]EHR: electronic health record.

[d]AUDIT: Alcohol Use Disorders Identification Test.

[e]SBIRT: Screening, Brief Intervention, and Referral to Treatment.

Patient Interviews

Patient results indicated strong interest and potential benefit from a UAU text intervention with limited concerns (Table 2). When comparing patients who screened moderate to high (n=6) versus severe (n=15) on the AUDIT-C, only patients at moderate to high risk said they had a goal to use the intervention to prevent future heavy drinking, while both groups mentioned abstinence and moderation. Perceived benefits also varied. Patients at severe risk thought the program could help stop the progression of alcohol use disorder symptoms, change their lifestyle, and be a distraction from drinking; patients at moderate to high risk thought the program could improve their job roles and family relationships, and provide information about drinking and change their thoughts about drinking. Both groups thought the intervention could provide support and improve their health and had similar content preferences. Patients at severe risk additionally mentioned using the text intervention to track drinking, set and complete goals, engage family support, avoid people who trigger their use, and reduce heavy drinking during drinking episodes.

XSL•FO
RenderX

Table . ED[a] patient qualitative interview themes from rapid analysis in accordance with i-PARIHS[b] domains.

Overall theme	Subtheme	Exemplar quote	i-PARIHS domain
Patient phone use patterns	• Most patients had mobile phones and used them for a variety of purposes, including frequent texting • It was rare that patients used their mobile phones for health or to manage their alcohol use	N/A	Recipients • Time, resources, and support • Motivations, values, and beliefs
Interest in a text intervention for alcohol use	• Most patients were interested in receiving the text intervention • Patients were interested in different goals (eg, to prevent future heavy drinking, reduce current drinking, or stop drinking completely) • Patients were interested in a variety of reasons (eg, improve health and relationships and change drinking cognitions)	• "Yeah. Maybe for myself if, if, you know, in the future, I decide to drink – I start drinkin' more. When I see myself drinkin' more, I wanna know information. Maybe I could prevent myself from, you know, gettin' into alcoholism or somethin' like that." [Participant 27] • "Personally, maybe it might change my mind on why I'm drinking at that moment. Or if I'm feeling low, maybe it'd inspire me not to drink as much. Maybe those texts'll be useful." [Participant 17]	Innovation (degree of fit)
Content	• Patients were interested in a variety of content (eg, quotes, uplifting, inspirational, or humorous messages, memes, images, resources, videos, and links with information) • Patient preference on the desired frequency of texts ranged from 3 times per week to several times per day	• "I like resources. Resources as in like what you should do besides picking up a drink." [Participant 13]	Innovation (degree of fit)
Offer preferences	• Most patients had no preference of who (eg, nurse or doctor) should offer the intervention, but it was more important that it be a personal conversation showing concern and empathy • Most patients had no preference for whether they would prefer the intervention to be offered through the health system versus a third party	• "Um, maybe towards the end and I'm, um, gettin' discharged or somethin,' like, when I feel better." [Participant 27]	Innovation (degree of fit)
Concerns	• Patients did not generally cite concerns related to privacy or cost; concern was that the program does not send too many messages, that they would not be able to read messages immediately, or that messages would be too long	• "So if I can't get to the message right away, I won't. But usually when I open the messages up, I do take concern in what's being said or written." [Participant 17]	Innovation (degree of fit)

XSL•FO
RenderX

Overall theme	Subtheme	Exemplar quote	i-PARIHS domain
Motivation to change	• Most participants had at least some motivation to change their drinking (eg, quit or cut down) and mentioned several reasons (eg, family and job)	• "I'm at a 10 where I wanna change, Um, I have tried. I cut back, um, 'cause I used to drink like every day and I cut back a lot, so…looking to cut it back even further to, and just stop drinking alcohol." [Participant 25]	Recipients (motivations, values, beliefs)
Benefits to enrolling	• Patients saw a number of benefits to enrolling (eg, help with mental and physical health, receiving information about negative effects of alcohol, having help in limiting alcohol use, or prompting in the moment to think differently about drinking)	• "It just healthy to stop. Make your body feel better, more healthy." [Participant 26]	Innovation

[a]ED: emergency department.

[b]i-PARIHS: Integrated Promoting Action on Research Implementation in Health Services.

Discussion

The results of this formative study on barriers and facilitators to implementing UAU texting interventions in EDs indicate both promise and caution. Quantitative data suggested concerns with feasibility (eg, low alcohol risk screening rates and feasibility scores). Qualitative results confirmed and extended quantitative results; providers viewed offering the intervention as appropriate and acceptable, especially to address the high volume of alcohol-related visits and support patients longer term; patients viewed the intervention as a positive way to meet drinking goals and saw several health benefits; they saw limited drawbacks (eg, slight concerns about time to read messages).

Providers felt there was no downside to offering the intervention even if there was incremental patient benefit. However, differences were noted among low and high implementation potential sites: variable screening rates and lower ratings on feasibility (eg, competing priorities, staff burnout, and turnover), indicating potential additional support strategies needed to create staff buy-in and sustain intervention engagement in low-potential sites. Larger EDs with higher patient census' may have unique implementation barriers, as they tended to have lower alcohol risk screening completion rates that would impact their ability to identify those who would benefit from text interventions. Other studies examining provider acceptance, barriers, and facilitators to implementing technology-based interventions in their EDs (not focused on alcohol use, however) revealed similar findings: high ratings of usage and value, but lower rates of actual use due to insufficient training, lack of awareness of the program, and technical factors (eg, compatibility) [39,40]. Future studies should examine ED characteristics that may impact implementation and design and test specific strategies to address these characteristics.

Patients of varying alcohol use risk severity indicated potential differences in drinking goals: those with higher severity may be less likely to have moderation goals and more likely to have abstinence goals. Results may reflect differences in how drinking has differently impacted health, relationships, functioning, and experience of alcohol use disorder symptoms among those with lower or higher severity. Patients' reasons for using a text program for UAU in our study were similar to that of another study examining patients' reasons for attending a residential program for alcohol use, including health, lifestyle, and sobriety goals [41]. Future studies of SMS text messaging programs for UAU should take into consideration different patient goals and experiences and how this could impact implementation strategies to engage patients. Designing programs that can address both moderate and more severe patients' needs and uses is important; the ability of the program to tailor to the needs and preferences of individual patients should be inherent in the design.

This study had limitations, including being conducted in one health system, which may affect generalizability, and classifying implementation potential using the 5-point Likert scale, which was meaningful in identifying potential implementation issues, but may not indicate practically meaningful differences.

In conclusion, SMS text messaging interventions for alcohol use have the potential to offer ED staff another tool to assist patients with meeting alcohol-related health goals. Patients welcome these interventions as a way to receive support and information. However, EDs will have varying levels of motivation and readiness to implement these driven by barriers related to feasibility. Findings may be useful for the development and testing of implementation strategies in larger cluster-randomized trials, an important area for mobile behavioral health [42]. Automating enrollment (eg, QR codes and text invites), capitalizing on existing workflows, and engaging ED leadership may be fruitful implementation strategies to test in future ED implementation research.

XSL•FO
RenderX

Acknowledgments

We are thankful to the sites and staff for their participation in the study. This study is supported by the National Institute on Alcohol Abuse and Alcoholism of the National Institutes of Health under award number R21AA029734 (to MO). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Demographics for chair surveys (n=26), staff interviews (n=18), and patient interviews (n=21).
[DOCX File, 16 KB - formative_v9i1e65187_app1.docx]

Multimedia Appendix 2
ED EHR alcohol screening data and feasibility, appropriateness, acceptability, and confidence of implementing alcohol SMS text messaging intervention in the ED in 17 sites. ED: emergency department; EHR: electronic health record.
[DOCX File, 17 KB - formative_v9i1e65187_app2.docx]

Checklist 1
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[DOCX File, 17 KB - formative_v9i1e65187_app3.docx]

Checklist 2
CHERRIES (Checklist 2 for Reporting Results of Internet e-Surveys) checklist.
[DOCX File, 18 KB - formative_v9i1e65187_app4.docx]

References

1. Sacks JJ, Gonzales KR, Bouchery EE, Tomedi LE, Brewer RD. 2010 national and state costs of excessive alcohol consumption. Am J Prev Med 2015 Nov;49(5):e73-e79. [doi: 10.1016/j.amepre.2015.05.031] [Medline: 26477807]
2. White AM, Castle IJP, Powell PA, Hingson RW, Koob GF. Alcohol-related deaths during the COVID-19 pandemic. JAMA 2022 May 3;327(17):1704-1706. [doi: 10.1001/jama.2022.4308] [Medline: 35302593]
3. Theriault KM, Rosenheck RA, Rhee TG. Increasing emergency department visits for mental health conditions in the United States. J Clin Psychiatry 2020 Jul 28;81(5):20m13241. [doi: 10.4088/JCP.20m13241] [Medline: 32726001]
4. White AM, Slater ME, Ng G, Hingson R, Breslow R. Trends in alcohol-related emergency department visits in the United States: results from the nationwide emergency department sample, 2006 to 2014. Alcohol Clin Exp Res 2018 Feb;42(2):352-359. [doi: 10.1111/acer.13559] [Medline: 29293274]
5. Suen LW, Makam AN, Snyder HR, et al. National prevalence of alcohol and other substance use disorders among emergency department visits and hospitalizations: NHAMCS 2014-2018. J Gen Intern Med 2022 Aug;37(10):2420-2428. [doi: 10.1007/s11606-021-07069-w] [Medline: 34518978]
6. Strayer RJ, Friedman BW, Haroz R, et al. Emergency department management of patients with alcohol intoxication, alcohol withdrawal, and alcohol use disorder: a white paper prepared for the American Academy of Emergency Medicine. J Emerg Med 2023 Apr;64(4):517-540. [doi: 10.1016/j.jemermed.2023.01.010] [Medline: 36997435]
7. Hawk KF, D'Onofrio G. Time to treat alcohol use disorder in the emergency department. Ann Emerg Med 2023 Apr;81(4):450-452. [doi: 10.1016/j.annemergmed.2022.11.013] [Medline: 36775724]
8. Berman AH, Gajecki M, Sinadinovic K, Andersson C. Mobile interventions targeting risky drinking among university students: a review. Curr Addict Rep 2016;3(2):166-174. [doi: 10.1007/s40429-016-0099-6] [Medline: 27226948]
9. Tofighi B, Nicholson JM, McNeely J, Muench F, Lee JD. Mobile phone messaging for illicit drug and alcohol dependence: a systematic review of the literature. Drug Alcohol Rev 2017 Jul;36(4):477-491. [doi: 10.1111/dar.12535] [Medline: 28474374]
10. Ranney ML, Pisani AR, Chernick LS. The role of texting in addressing mental health. In: Moreno MA, Radovic A, editors. Technology and Adolescent Mental Health: Springer International Publishing; 2018:207-215. [doi: 10.1007/978-3-319-69638-6_15]
11. Keoleian V, Polcin D, Galloway GP. Text messaging for addiction: a review. J Psychoactive Drugs 2015;47(2):158-176. [doi: 10.1080/02791072.2015.1009200] [Medline: 25950596]

XSL•FO
RenderX

12. Schwebel FJ, Larimer ME. Text message reminders as an adjunct to a substance use intervention for adolescents and young adults: pilot feasibility and acceptability findings. Digit Health 2020;6. [doi: 10.1177/2055207620965052] [Medline: 33110614]
13. Hutton A, Prichard I, Whitehead D, et al. mHealth interventions to reduce alcohol use in young people: a systematic review of the literature. Compr Child Adolesc Nurs 2020 Sep;43(3):171-202. [doi: 10.1080/24694193.2019.1616008] [Medline: 31192698]
14. Mason M, Ola B, Zaharakis N, Zhang J. Text messaging interventions for adolescent and young adult substance use: a meta-analysis. Prev Sci 2015 Feb;16(2):181-188. [doi: 10.1007/s11121-014-0498-7] [Medline: 24930386]
15. Sibley AL, Noar SM, Muessig KE, et al. An automated text messaging intervention to reduce substance use self-stigma (Project RESTART): protocol for a feasibility and acceptability pilot study. JMIR Res Protoc 2024 Aug 9;13:e59224. [doi: 10.2196/59224] [Medline: 39121478]
16. Spieker AJ, Nelson LA, Rothman RL, et al. Feasibility and short-term effects of a multi-component emergency department blood pressure intervention: a pilot randomized trial. J Am Heart Assoc 2022 Mar;11(5):e024339. [doi: 10.1161/JAHA.121.024339] [Medline: 35195015]
17. Bendtsen M, McCambridge J, Åsberg K, Bendtsen P. Text messaging interventions for reducing alcohol consumption among risky drinkers: systematic review and meta-analysis. Addiction 2021 May;116(5):1021-1033. [doi: 10.1111/add.15294] [Medline: 33047865]
18. Muench F, Madden SP, Oommen S, et al. Automated, tailored adaptive mobile messaging to reduce alcohol consumption in help-seeking adults: a randomized controlled trial. Addiction 2024 Mar;119(3):530-543. [doi: 10.1111/add.16391] [Medline: 38009576]
19. Suffoletto B, Chung T, Muench F, Monti P, Clark DB. A text message intervention with adaptive goal support to reduce alcohol consumption among non-treatment-seeking young adults: non-randomized clinical trial with voluntary length of enrollment. JMIR Mhealth Uhealth 2018 Feb 16;6(2):e35. [doi: 10.2196/mhealth.8530] [Medline: 29453191]
20. Suffoletto B, Kristan J, Callaway C, et al. A text message alcohol intervention for young adult emergency department patients: a randomized clinical trial. Ann Emerg Med 2014 Dec;64(6):664-672.e4. [doi: 10.1016/j.annemergmed.2014.06.010] [Medline: 25017822]
21. Sharpe S, Kool B, Whittaker R, et al. Effect of a text message intervention to reduce hazardous drinking among injured patients discharged from a trauma ward: a randomized controlled trial. NPJ Digit Med 2018;1:13. [doi: 10.1038/s41746-018-0019-3] [Medline: 31304298]
22. Burner E, Zhang M, Terp S, et al. Feasibility and acceptability of a text message-based intervention to reduce overuse of alcohol in emergency department patients: controlled proof-of-concept trial. JMIR Mhealth Uhealth 2020 Jun 4;8(6):e17557. [doi: 10.2196/17557] [Medline: 32496203]
23. Nesvåg S, McKay JR. Feasibility and effects of digital interventions to support people in recovery from substance use disorders: systematic review. J Med Internet Res 2018 Aug 23;20(8):e255. [doi: 10.2196/jmir.9873] [Medline: 30139724]
24. Ramsey AT, Satterfield JM, Gerke DR, Proctor EK. Technology-based alcohol interventions in primary care: systematic review. J Med Internet Res 2019 Apr 8;21(4):e10859. [doi: 10.2196/10859] [Medline: 30958270]
25. Kmiec J, Suffoletto B. Implementations of a text-message intervention to increase linkage from the emergency department to outpatient treatment for substance use disorders. J Subst Abuse Treat 2019 May;100:39-44. [doi: 10.1016/j.jsat.2019.02.005] [Medline: 30898326]
26. Harvey G, Kitson A. PARIHS revisited: from heuristic to integrated framework for the successful implementation of knowledge into practice. Implement Sci 2016 Mar 10;11:33. [doi: 10.1186/s13012-016-0398-2] [Medline: 27013464]
27. Swindle T, Johnson SL, Whiteside-Mansell L, Curran GM. A mixed methods protocol for developing and testing implementation strategies for evidence-based obesity prevention in childcare: a cluster randomized hybrid type III trial. Implement Sci 2017 Jul 18;12(1):90. [doi: 10.1186/s13012-017-0624-6] [Medline: 28720140]
28. Kitson A, Harvey G. Facilitating an evidence-based innovation into practice: the novice facilitator's role. In: Implementing Evidence-Based Practice in Healthcare: Routledge; 2015:85-104. [doi: 10.4324/9780203557334-6]
29. O'Grady MA, Kapoor S, Harrison L, Kwon N, Suleiman AO, Muench FJ. Implementing a text-messaging intervention for unhealthy alcohol use in emergency departments: protocol for implementation strategy development and a pilot cluster randomized implementation trial. Implement Sci Commun 2022 Aug 6;3(1):86. [doi: 10.1186/s43058-022-00333-y] [Medline: 35933560]
30. Palinkas LA, Aarons GA, Horwitz S, Chamberlain P, Hurlburt M, Landsverk J. Mixed method designs in implementation research. Adm Policy Ment Health 2011 Jan;38(1):44-53. [doi: 10.1007/s10488-010-0314-z] [Medline: 20967495]
31. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Arch Intern Med 1998;158(16):1789-1795. [doi: 10.1001/archinte.158.16.1789] [Medline: 9738608]
32. Weiner BJ, Lewis CC, Stanick C, et al. Psychometric assessment of three newly developed implementation outcome measures. Implement Sci 2017 Aug 29;12(1):108. [doi: 10.1186/s13012-017-0635-3] [Medline: 28851459]

XSL•FO
RenderX

33. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015 Sep;42(5):533-544. [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
34. Benoot C, Hannes K, Bilsen J. The use of purposeful sampling in a qualitative evidence synthesis: a worked example on sexual adjustment to a cancer trajectory. BMC Med Res Methodol 2016 Feb 18;16:21. [doi: 10.1186/s12874-016-0114-6] [Medline: 26891718]
35. Ranney ML, Choo EK, Wang Y, Baum A, Clark MA, Mello MJ. Emergency department patients' preferences for technology-based behavioral interventions. Ann Emerg Med 2012 Aug;60(2):218-227. [doi: 10.1016/j.annemergmed.2012.02.026] [Medline: 22542311]
36. Kool B, Smith E, Raerino K, Ameratunga S. Perceptions of adult trauma patients on the acceptability of text messaging as an aid to reduce harmful drinking behaviours. BMC Res Notes 2014 Jan 4;7:4. [doi: 10.1186/1756-0500-7-4] [Medline: 24387293]
37. Gale RC, Wu J, Erhardt T, et al. Comparison of rapid vs in-depth qualitative analytic methods from a process evaluation of academic detailing in the Veterans Health Administration. Implement Sci 2019 Feb 1;14(1):11. [doi: 10.1186/s13012-019-0853-y] [Medline: 30709368]
38. Hamilton AB. Qualitative methods in rapid turn-around health services research. Presented at: 2013 VA HSR&D Cyberseminar Spotlight on Women's Health; Dec 11, 2013.
39. Fujimori R, Liu K, Soeno S, et al. Acceptance, barriers, and facilitators to implementing artificial intelligence-based decision support systems in emergency departments: quantitative and qualitative evaluation. JMIR Form Res 2022 Jun 13;6(6):e36501. [doi: 10.2196/36501] [Medline: 35699995]
40. Jun S, Plint AC, Campbell SM, Curtis S, Sabir K, Newton AS. Point-of-care cognitive support technology in emergency departments: a scoping review of technology acceptance by clinicians. Acad Emerg Med 2018 May;25(5):494-507. [doi: 10.1111/acem.13325] [Medline: 28960689]
41. Marsh TN, Eshakakogan C, Spence M, et al. Patient experience of an abstinence-based indigenous residential treatment program in Northern Ontario: a descriptive qualitative study. Front Health Serv 2024;4:1387184. [doi: 10.3389/frhs.2024.1387184] [Medline: 39777229]
42. Graham AK, Lattie EG, Powell BJ, et al. Implementation strategies for digital mental health interventions in health care settings. Am Psychol 2020 Nov;75(8):1080-1092. [doi: 10.1037/amp0000686] [Medline: 33252946]

Abbreviations

AUDIT-C: Alcohol Use Disorders Identification Test-Consumption
CHERRIES: Checklist for Reporting Results of Internet e-Surveys
COREQ: Consolidated Criteria for Reporting Qualitative Research
ED: emergency department
EHR: electronic health record
i-PARIHS: Integrated Promoting Action on Research Implementation in Health Services
UAU: unhealthy alcohol use

Edited by A Mavragani; submitted 08.08.24; peer-reviewed by H Hutton; revised version received 30.01.25; accepted 30.01.25; published 03.03.25.

Please cite as:
O'Grady M, Harrison L, Suleiman A, Hutchison M, Kwon N, Muench F, Kapoor S
Text Messaging Interventions for Unhealthy Alcohol Use in Emergency Departments: Mixed Methods Assessment of Implementation Barriers and Facilitators
JMIR Form Res 2025;9:e65187
URL: https://formative.jmir.org/2025/1/e65187
doi:10.2196/65187

© Megan O'Grady, Laura Harrison, Adekemi Suleiman, Morica Hutchison, Nancy Kwon, Frederick Muench, Sandeep Kapoor. Originally published in JMIR Formative Research (https://formative.jmir.org), 3.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Patients' Experience With Evaluation by Both a Musculoskeletal Physician and Physical Therapist in the Same Digital Visit: Survey Study

Mary I O'Connor, MD; Carolyn Chudy, DO; Kaitlyn C Peters, NP; Megan Ribaudo; Carrie McCulloch, MD; Jared Aguilar, DPT; Trista Taylor, MPA; Ryan A Grant, MS, MBA, MD

Vori Health, Inc, Nashville, TN, United States

Corresponding Author:
Mary I O'Connor, MD
Vori Health, Inc, Nashville, TN, United States

Abstract

Background: Patients undergoing evaluation for musculoskeletal concerns are often seen by a physician and physical therapist in the in-person setting in a sequential manner. This process typically delays the onset of nonoperative care, inclusive of physical therapy, and creates the risk of inadequate clinical collaboration between physician and physical therapist. To address these issues, we designed a novel initial patient evaluation to a group visit in which both a specialty-trained musculoskeletal physician and physical therapist simultaneously evaluate a patient together in the digital encounter.

Objective: The aim of the study is to gain insights from patients on their experience with this innovative digital simultaneous musculoskeletal medical doctor and physical therapist (MD+PT) visit format for the initial evaluation of musculoskeletal concerns.

Methods: An electronic 7-question survey was sent to 750 patients who completed an MD+PT visit asking them to comment on prior musculoskeletal evaluations and their experience with the MD+PT format.

Results: In total, 195 (26%) patients responded to the survey with the frequent body regions of diagnosis being lumbar spine (n=65), knee (n=32), shoulder (n=21), cervical spine (n=20), hip (n=14), and hand (n=11). Most patients had prior musculoskeletal experience with a physician or nurse practitioner (171/195, 87.7%) or physical therapist (148/195, 75.9%) with nearly all such encounters in the in-person setting (161/171,94.2% for physician or nurse practitioner and 144/148, 97.3% for physical therapy). Only 3.1% (6/193) of patients reported seeing both a physician and physical therapist during the same in-person visit. Patients rated the simultaneous MD+PT visit very favorably: this type of digital evaluation saved them time (179/192, 93.2%) and permitted them to promptly start their treatment plan (174/192, 90.6%). Overall, 87.5% (168/192) rated the MD+PT visit as enjoyable, and 92.2% (177/192) responded that it increased their confidence with understanding their medical condition and how to start treating it.

Conclusions: Our early experience with the evaluation of patients with musculoskeletal conditions by both a specialty-trained musculoskeletal physician and physical therapist simultaneously in the same digital visit resulted in patients reporting a very positive experience with high satisfaction, engagement, and confidence in understanding their diagnosis and how to start treating it.

(JMIR Form Res 2025;9:e66744) doi:10.2196/66744

KEYWORDS

telemedicine; musculoskeletal care; patient satisfaction; multidisciplinary care; digital visit

Introduction

Musculoskeletal disorders are a leading cause of disability and health care expenditure worldwide [1]. Inappropriate care exacerbates the cost of musculoskeletal care as seen, for example, with high-cost spine [2] and total knee replacement [3] surgeries. Efficient and effective evidence-based evaluation of musculoskeletal disorders is critical to ensuring patient-centric care, delivering quality patient outcomes, and lowering health care costs. The type of care a patient with a musculoskeletal concern seeks and the way that care is delivered—especially at the onset of a care journey—can impact the risk of inappropriate and expensive imaging [4-6], procedures [7,8], and surgery [2,3,6]. In other words, what type of clinician a patient initially starts with influences the care path a patient will receive.

Most patients with musculoskeletal concerns first seek care from their primary care clinician [9]; however, many primary care clinicians lack fundamental competency in the musculoskeletal examination and do not feel comfortable evaluating, diagnosing, and treating musculoskeletal conditions [10]. In many states, patients may now directly see a physical

XSL•FO
RenderX

therapist without a physician referral, via direct access physical therapy, due to the expansion of the scope of practice for physical therapists [11]. From either primary care or physical therapy, many patients are subsequently referred to an orthopedic or spine surgeon, or in the minority of cases a specialty nonoperative musculoskeletal physician (eg, physical medicine and rehabilitation or sports medicine physician), collectively referred to as specialty musculoskeletal physician. Creation of a comprehensive treatment plan for these patients will be dependent upon the completion of a specialty musculoskeletal physician's clinical assessment, which often is delayed given the time required for separate physician and physical therapist appointments, and potentially by discordance in evaluations and advice by these 2 clinicians.

For many patients, musculoskeletal care can be effectively delivered through telemedicine. Digital delivery of medical services can be very convenient for patients, particularly for those with challenges to access in-person clinicians [12]. Physical examination of the patient with a musculoskeletal condition can be readily performed in the digital space [13]. The adequacy of that digital physical examination is supported by research showing good to very good concordance of the diagnosis of the patient with a musculoskeletal condition between the digital and in-person clinician [13]. Moreover, the concordance of treatment plans for patients with musculoskeletal conditions created in these 2 practice settings was "probably good to excellent" [14]. Finally, a meta-analysis of 35 systematic reviews assessing the use of telemedicine versus in-person care for musculoskeletal conditions showed telemedicine to be beneficial to or equal in patient-reported outcomes measures and objective measures, including physical function, and to have notably lower costs. Furthermore, in the small number of studies in the review reporting on patient experience measures, telemedicine was equivalent to in-person care [15].

We developed an initial clinical digital visit structure to address these delays in treatment created by the traditional sequential clinical visits between physician and physical therapist, streamline the patient experience, and advance evidence-based patient care. In this model, patients see a nonoperative musculoskeletal specialty physician (or a musculoskeletal nurse practitioner) and a physical therapist together—literally at the same time—during a simultaneous digital medical visit via our telehealth platform [16]. To our knowledge, such a simultaneous evaluation, which we will abbreviate as an "MD+PT" (musculoskeletal medical doctor and physical therapist) digital evaluation, has not been previously systematically implemented in either the in-person or digital care environment. To gain insights into the patient experience with this simultaneous patient-centric model, we created a descriptive survey to better understand the experience from the patients' point of view.

Methods

Ethical Considerations

Our study, including the consent language and questionnaire, was reviewed by The Institute for Evaluation and Research Institutional Review Board and approved as exempt. Patients who completed an MD+PT digital evaluation were sent an email asking them to voluntarily consent to the research study and complete a 7-question survey about their appointment experience with the MD+PT evaluation (Textbox 1). All patients were 18 years or older of age. The email informed patients that the purpose of the study was to examine the impact of enabling patients to see a medical doctor or nurse practitioner and physical therapist together in the same digital appointment. Patients were provided no compensation for completing the survey. Those who elected to participate were informed that they could withdraw from the study at any time and skip questions if they chose to do so. The consent informed patients that their survey results would be confidential, shared with our researchers and not their medical team. Participants could answer the questionnaire in an anonymous manner and complete it only once. The questions could not be answered until participants consented to participate in the study. Data integrity was protected with access only provided to the researchers conducting the data analysis.

XSL•FO
RenderX

Textbox 1. The 7-question survey used in this study of 195 patients who completed an evaluation with both a nonoperative musculoskeletal physician and physical therapist during the same digital encounter.

1. Before coming to Vori Health, have you ever seen a medical doctor (or nurse practitioner or physician assistant) for back, muscle, or joint pain?

- Yes, no

If yes, were any of these visits telehealth appointments? Yes/no

2. Before coming to Vori Health, have you ever seen a physical therapist for back, muscle, or joint pain?

- Yes, no

If yes, were any of these visits telehealth appointments? Yes/no

3. Before coming to Vori Health, have you ever been seen by both a medical doctor/nurse practitioner and physical therapist together in the same appointment for back, muscle, or joint pain?

- Yes, no

If yes, were any of these visits telehealth appointments? Yes/no

The next few questions will ask you about your experience at Vori Health seeing a medical doctor or nurse practitioner and physical therapist together in the same appointment.

4. Seeing a medical doctor/nurse practitioner and physical therapist together in the same appointment saved me time.

- Strongly agree, agree, neutral, disagree, strongly disagree

5. Seeing a medical doctor/nurse practitioner and physical therapist together in the same appointment increased my confidence in understanding my medical condition and how to start treating it.

- Strongly agree, agree, neutral, disagree, strongly disagree

6. Seeing a medical doctor/nurse practitioner and physical therapist together in the same appointment helped me to start my treatment plan faster compared to if I had to see the doctor and physical therapist in different appointments.

- Strongly agree, agree, neutral, disagree, strongly disagree

7. Seeing a medical doctor/nurse practitioner and physical therapist together in the same appointment made my care experience more enjoyable.

- Strongly agree, agree, neutral, disagree, strongly disagree

If disagree, or strongly disagree, show question below:

Please indicate why the appointment was not enjoyable for you (choose all that apply):

o The experience felt too complicated.

o I did not feel I had enough time with each clinician.

o I did not feel comfortable sharing confidential medical information with anyone other than my medical doctor.

o Other (free text):

Questionnaire

The questionnaire sought patients' prior experience with musculoskeletal care, both in the in-person and telemedicine settings. Patients were asked if they had ever been seen by both a medical doctor (or nurse practitioner or physician associate) and physical therapist in the same appointment. Additional questions were asked related to patients' experiences with their MD+PT digital evaluation. To recruit patients to the study, an email survey was sent to all patients who completed the MD+PT evaluation until we had 200 patients who consented to participate in the study. There were no other inclusion or exclusion criteria.

The MD+PT Clinical Evaluation and Care Plan

The MD+PT digital evaluation was performed by a specialty musculoskeletal physician (typically, physical medicine and rehabilitation), or for a smaller number of patients by an experienced musculoskeletal nurse practitioner, and a physical therapist who had a minimum of 5 years of musculoskeletal experience. Each integrated visit was 40 minutes in duration. The physician or nurse practitioner, physical therapist, and patient were each in separate physical locations and logged into the same videoconferencing link to be together in the digital space for the evaluation. The history and digital physical examination were performed by a combination of physician or nurse practitioner and physical therapist. Each clinician could ask additional questions to the patient to provide more details to the medical history. The physical examination was completely digital and based on published guidelines [13]. Both the physician or nurse practitioner and physical therapist could instruct the patient to perform various movements and tests to complete the physical examination to the satisfaction of both clinicians. Based on the available evidence, the physician or nurse practitioner then rendered a medical diagnosis, and the physical therapist provided a functional diagnosis. If outside records, inclusive of previous imaging, were not available to the clinicians at the time of this initial visit, then the physician or nurse practitioner would request the outside records and imaging for subsequent review.

XSL•FO
RenderX

At the completion of this visit, patients were immediately started on an evidence-based biopsychosocial personalized care plan that may include evidence-based prescriptions, imaging, laboratory studies, and a personalized home exercise program facilitated by computer vision motion-tracking technology. Patients were subsequently scheduled for follow-up digital physical therapy care and, if applicable, dedicated health coaching and registered dietitian care, given the overlap of modifiable risk factors with musculoskeletal conditions (eg, obesity). Follow-up physician or nurse practitioner visits were also scheduled as clinically appropriate.

Results

Overview

Surveys were sent to a total of 750 patients, with 200 patients consenting to the study. In total, 5 patients consented but did not answer any questions on the survey resulting in 195 responses, a 26% survey response rate. Surveys were fully answered by 192 patients and incompletely answered by 3 patients who each did not complete 1 - 2 questions. Patients who responded to the survey had a variety of diagnoses involving the following body regions: lumbar spine (n=65), knee (n=32), shoulder (n=21), cervical spine (n=20), hip (n=14), hand (n=11), ankle (n=8), elbow or wrist (n=6), thoracic spine (n=5), and other and nonspecific body regions (n=13).

Prior Musculoskeletal Clinical Experience

Patients were asked about their prior experience with the evaluation of back, muscle, or joint pain. In total, 171 of 195 (87.7%) patients had previously seen a physician or nurse practitioner, with 161 of 171 (94.2%) of these encounters occurring in the in-person setting. A total of 148 of 195 (75.9%) patients had previously seen a physical therapist, with 144 of 148 (97.3%) of these encounters occurring in the in-person setting.

We asked patients about prior experience of being evaluated by both the physician (or nurse practitioner or physician assistant) and physical therapist in the same clinical visit for back, muscle, or joint pain. Only 3.1% (6/193) reported a prior visit that combined a physician and physical therapist with 1 patient commenting that this occurred during a hospital stay, with no patient reporting the encounter as digital. We concluded that our cohort had essentially no experience with simultaneous musculoskeletal clinical visits, particularly in the digital setting.

MD+PT Digital Experience

The remainder of our survey questions focused on the experience of the patient with the MD+PT digital evaluation. Table 1 shows that 93.2% (179/192) agreed that this appointment format saved them time (strongly agreed: 162/192, 84.4% and somewhat agreed: 17/192, 8.9%).

Table . Survey responses (n=192) to question asking patients whether they felt the MD+PT digital evaluation saved them time.

Response	Participant count, n
Strongly agree	162
Somewhat agree	17
Neither agree nor disagree	11
Somewhat disagree	0
Strongly disagree	2

When patients were asked if the MD+PT digital evaluation allowed them to start their treatment plan faster than if they had seen a physician and physical therapist in separate appointments, 90.6% (174/192) of patients responded affirmatively (strongly agreed: 155/192, 80.7% and somewhat agreed: 19/192, 9.9%; Table 2).

Table . Survey responses (n=192) to question asking patient whether the MD+PT digital evaluation helped them start a treatment plan faster compared to seeing a doctor and physical therapist in separate appointments.

Response	Participant count, n
Strongly agree	155
Somewhat agree	19
Neither agree nor disagree	16
Somewhat disagree	1
Strongly disagree	1

Given our focus on the patient experience, the 2 most significant survey questions asked about patient enjoyment of their visit experience and confidence in medical decision-making. Overall, Table 3 shows that 87.5% (168/192) rated the visit as enjoyable (strongly agree: 135/192, 70.3% and somewhat agree: 33/192, 17.2%).

XSL•FO
RenderX

Table . Survey responses (n=192) to question asking patients whether the MD + PT digital evaluation made their patient care experience more enjoyable.

Response	Participant count, n
Strongly agree	135
Somewhat agree	33
Neither agree nor disagree	21
Somewhat disagree	1
Strongly disagree	2

The patients who responded as “somewhat disagreed” (n=1) or “strongly disagreed” (n=2) each provided a comment: (1) the practitioner and therapist were not very engaging, and I could not connect with them; (2) digital appointment diminished the transfer of information between the patient and provider; and (3) in-person health care is so very important, especially the first visit.

When asked if MD+PT digital evaluation increased their confidence with understanding their medical condition and how to start treating it, Table 4 illustrates that 92.2% (177/192) positively responded (strongly agree: 137/192, 71.4% and somewhat agree: 40/192, 20.8%).

Table . Survey responses (n=192) to question asking patients whether the MD+PT digital evaluation increased their confidence in understanding their medical condition and how to start treating it.

Response	Participant count, n
Strongly agree	137
Somewhat agree	40
Neither agree nor disagree	12
Somewhat disagree	1
Strongly disagree	2

Discussion

Overview

With 1 in 2 US adults estimated to experience a musculoskeletal disorder [9], improving musculoskeletal care must be a national, if not global, priority. We innovated our digital initial clinical evaluation to bring both the physician or nurse practitioner and physical therapist together simultaneously with the patient for the duration of the telemedicine evaluation. More specifically, both clinicians were together with the patient for the entirety of the first digital visit. This allows the patient to provide 1 history and the clinicians to collaboratively perform 1 digital physical examination. The physician or nurse practitioner renders a medical diagnosis, and the physical therapist provides a functional diagnosis, with the opportunity for collaboration between both clinicians during the visit to drive clinician concordance. The resultant treatment plan is aligned with the patient’s values and goals (ie, “What matters to me”) [17], and the patient can ask clarifying questions to both clinicians. Our clinicians subjectively commented that this integrated, combined visit was a more satisfying and collaborative experience for them.

Principal Findings

Our survey results showed a very high level of positive responses from patients. Patients felt that this visit format saved them time (179/192, 93.2%), allowed them to start their treatment plan faster (174/192, 90.6%), and was enjoyable (168/192, 87.5%). Moreover, 92.2% (177/192) of patients responded that the visit increased their confidence with understanding their medical condition and how to start treating it. Improved patient understanding of their medical condition and associated treatment positively affects health outcomes [18]. Effective patient education supports an atmosphere of trust between patient and clinician, empowers patients to participate in their own health care, and supports informed decision-making [19].

We postulate that the high patient confidence in understanding their medical condition and how to start treating it was influenced by our MD+PT visit format. Our model was structured to create diagnostic and therapeutic agreements between clinicians. The model also provided patients with the benefit, real or perceived, of having both a musculoskeletal physician or nurse practitioner and physical therapist together with them in the clinical encounter. While direct access physical therapy has been shown to be a safe, less expensive, and effective model for triage of patients with musculoskeletal conditions compared to a required referral from a physician [20], physical therapists cannot order imaging, prescribe medications, or render a medical diagnosis in the United States. Patients may have a perception of a more complete evaluation when the physician (or nurse practitioner) is evaluating them, even if imaging or prescription medications are not ordered. Patients may also have a higher level of trust when a physician or nurse practitioner is creating the treatment plan with the physical therapist, particularly for patients with complex conditions or comorbidities. Future research comparing the experience of patients being evaluated by a physical therapist versus evaluation by both a physician or nurse practitioner and

XSL•FO
RenderX

physical therapist may provide data on the potential value (or not) of the MD+PT visit.

Research on concordance [21] in diagnosis between physicians and physical therapists shows varied results. In a systemic review of 19 studies involving 1745 patients and a total of 35 advanced practice physical therapists (APPTs), Lafrance et al [22] showed concordance between APPTs and physicians for musculoskeletal disorders to be "probably good to very good for diagnosis" (κ coefficient [21] 0.76) and "good to very good for surgical triage" (κ coefficient 0.71). All studies in this systemic review were conducted outside the United States, and in 5 studies, APPTs had the authority to order diagnostic imaging studies, including magnetic resonance imaging [22]. The ability of some APPTs in this study to order diagnostic imaging limits the application of these results to concordance between a physical therapist and a physician in the United States.

In other research, Madsen et al [23] reported on 69 patients independently evaluated twice on the same day by an orthopedic surgeon and an extended scope physical therapist. Agreement between the surgeon and physical therapist was 62% on the primary diagnosis and 79% on a combination of diagnoses [23]. In 2 small studies, the κ coefficient between APPTs and musculoskeletal surgeons for referral for surgical treatment was 0.46 (shoulder disorders) [24] and 0.69 (lumbar spine pain) [25], suggesting an opportunity for improvement. The experience of the physical therapist is a variable that may influence these outcomes: in a study of 105 patients with carpal tunnel syndrome, physical therapists with advanced experience had 100% agreement with surgeons, whereas early-career physical therapists only had fair agreement [26]. Depressive symptoms were also shown in one study of knee disorders to decrease diagnostic concordance between physical therapists and medical musculoskeletal physicians [27]. While all our clinicians have a least 5 years of musculoskeletal clinical experience, our MD+PT model may mitigate diagnostic discordance when the clinical experience of one of the clinicians is more limited.

We believe that concordance in the treatment plan in our MD+PT model was facilitated by direct and immediate communication between clinicians. In a survey of interprofessional communication between Canadian orthopedic surgeons and physical therapists, 65.6% of clinicians in stand-alone practices report a negative view of communication compared to 48.4% of those in collaborative practice. This same study also found that orthopedic surgeons and physical therapists did not agree on the clinical information that should be shared regarding postoperative rehabilitation [28]. In a study of 600 survey responses of physical therapists treating patients who underwent rotator cuff repair, 33% of physical therapists reported receiving copies of the operative report in <25% of patients, and 16% reported not receiving an operative note >50% of the time. With postoperative therapy protocols based on the size of the rotator cuff tear and the complexity of the repair, this lack of communication could negatively impact clinical care [29].

We also highlight the importance of patients being able to promptly begin their treatment plan. In the traditional model, the time to diagnosis and treatment of a patient often depends upon the sequencing of physician and physical therapist evaluations as well as the potential need for insurance authorization for care, with frequent delays in care. While in many states, patients can access physical therapy (direct PT) for initial treatment with no delay in beginning the physical therapy plan of care, one small study found that 25.6% of patients who began physical therapy and were subsequently referred to a physician had their physician visit at a range of 0.5 to 10 weeks following the initial physical therapy evaluation [30]. We were unable to find data on the length of time for a patient to see the physical therapist after referral from the physician, which is likely dependent on many factors including patient insurance and the need for prior authorization. Delaying the onset of treatment may lengthen the recovery time and negatively impact clinical outcomes [31,32]. Finally, 2 separate clinical visits can result in additional expenses to patients related to travel costs, time lost from work, and disruption of normal activities.

Finally, a very high percentage (168/192, 87.5%) of patients found the MD+PT visit enjoyable. We chose to ask this question as an attempt to gauge the emotional experience of the patient with 1 simple question. We believe that taking evidence-based care to the next level of effectiveness also requires delivering a better patient experience. We are not aware of others using this question to gain insights into the patient experience. The current Consumer Assessment of Healthcare Providers and Systems (CAPHS) survey, a standard instrument for measuring the patient experience, asks patients about various aspects of their experience including whether their visit started on time and if the clinician showed them respect, listened carefully, and gave them enough time. CAPHS asks patients to rate their visit on a scale of 0 (worse visit possible) to 10 (best visit possible) [33]. In our opinion, CAPHS is a very valuable tool but does not query the emotional response of the patient to the encounter. While we recognize "enjoyable" may never be a realistic experience for patients with very serious medical issues, we challenge the health care system to explore this, and similar, questions that assess the emotional response of the patient. We believe that the more enjoyable the clinical experience is for patients, the higher their engagement will be, which will translate into improved clinical outcomes.

Limitations

Our research has several limitations. The main limitation was our survey response rate of 26% (N=195) raising the potential for nonresponder selection bias. Consensus as to an acceptable survey response rate in medical research is lacking [34,35]. Studies show a range of patient response rates to email surveys. In a global overview paper of surgical patients, email studies gave an average response rate of 68%, although the United States was noted to have a lower overall response rate of 64.2% [33]. In a primary care practice, the response rate to a postvisit email survey was 60.9%. However, only 45.15% of patients who initially consented to completing a paper waiting room survey agreed to receive an email survey [36].

While these response rates are higher than ours, they may not be directly comparable. Patients who have undergone a surgical

XSL•FO
RenderX

procedure may be more willing to participate in a research study compared to those who had a telemedicine visit. In the primary care practice study cited in the preceding paragraph, the response rate to an email survey was 27.7% (60.9% of the 45.15% of patients who consented to an email survey), which is comparable to our response rate of 26% (N=195). In our study, we did not ask patients for consent to be surveyed, rather we sent the email survey to all eligible patients until 200 patients consented to the study.

In studies with methodology more aligned with ours, our survey response rate is favorable. A 2016 primary care study showed a 19.8% response rate to a 6-question email survey to patients on access to care and patient centeredness [37]. Another publication in 2018 reported 28% of patients opening an email containing a patient experience survey, with 80% completing the survey [38]. In a 2024 poster on patient satisfaction and outcomes, an email survey of patients having a genetic information session for kidney disease via telephone had a response rate of 12.1%, although the survey was much longer at 47 questions [39].

We did not conduct a corresponding survey of our clinicians using this model to potentially identify episodes of initial discordance in clinical assessment and treatment plan (and how these were resolved). All patients received a personalized, evidence-based, biopsychosocial treatment plan supported by the physician or nurse practitioner and physical therapist at the completion of the initial visit. Health coaches and registered dietitians were brought into the clinical team to support the patient as appropriate. We also did not survey our clinicians on their perception of whether communication between clinicians was facilitated with this model or whether the clinicians found this model to be enjoyable for them. We did not correlate the results of this survey with clinical outcomes or health care use; these are areas we plan to study in the future.

Patient's perception of the quality of care provided by each clinician, or the quality of care provided by the MD+PT visit, was not studied and is an area for potential future research. However, given that patients reported high confidence in their medical plan with the MD+PT visit, we postulate that this is a surrogate for patients' perceptions of a higher quality of care. In the literature, one study found that patients with osteoarthritis reported that physical therapists provided better information about how to take care of their condition, while surgeons provided patients with greater involvement in the decision-making process and facilitated a greater degree of expectations being met [40]. When compared to primary care clinicians, another study found that patient perception of physical therapy–led care was superior in all quality aspects evaluated [41].

Our model requires the resources of both a physician or nurse practitioner and physical therapist in the initial evaluation. Direct access physical therapy has been shown to be clinically effective and resource-sensitive, as many patients with musculoskeletal conditions will not require a physician or nurse practitioner assessment [11]. The scope of our study did not support any type of cost analysis of our model versus direct access physical therapy. We believe that patients with low-complexity musculoskeletal issues can be effectively evaluated in a physical therapy first model and those with high complexity are better evaluated in our MD+PT model.

Finally, our study did not directly address whether health literacy was improved in our patients. Health literacy is defined as "the degree to which individuals have the ability to find, understand, and use information and services to inform health-related decisions and actions for themselves and others" [42]. We speculate that patients having increased confidence in their understanding of their condition and how to start treating it may have greater health literacy. As annual health care costs of individuals with low health literacy are 4 times higher than those of higher health literacy individuals, and patients with higher health literacy are more likely to adhere to treatment recommendations [43], we will continue our research to identify the potential impact of our model on health literacy.

Conclusions

We found our innovative model of a musculoskeletal physician or nurse practitioner and physical therapist evaluating the patient with a musculoskeletal condition together (at the same time during the initial telemedicine visit) resulted in a high degree of positive patient feedback. Overwhelmingly, patients reported that this visit model saved them time, allowed them to start their treatment plan faster, was enjoyable, and increased their confidence in understanding their medical condition and how to start treating it. We will continue to seek additional ways to innovate patient-centric care that enhances the patient experience, clinical outcomes, and appropriate health care use.

Acknowledgments

There was no use of generative artificial intelligence in the generation of text, figures, or other informational content of this manuscript.

Data Availability

The datasets generated and analyzed during this study are not publicly available because the public sharing of respondent data was not included in the consent for study participation permission.

XSL•FO
RenderX

Authors' Contributions

MIO wrote and edited the manuscript and contributed to the study conception, methodology, and data analysis. MR curated the data, provided tables for the manuscript, and contributed to the study conception, methodology, and data analysis. CC and JA contributed to the study conception and methodology. KCP contributed to study conception, methodology, and data analysis. CM contributed to methodology and writing the manuscript. TT determined the software used in the study and supported data curation. RAG contributed with writing and editing the manuscript. All authors reviewed the final manuscript.

Conflicts of Interest

All authors are employees of Vori Health Medical Group, PLLC or Vori Health, Inc, and have a financial interest in the company.

References

1. Gill TK, Mittinty MM, March LM. Global, regional, and national burden of other musculoskeletal disorders, 1990-2020, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. Lancet Rheumatol 2023 Nov;5(11):e670-e682. [doi: 10.1016/S2665-9913(23)00232-1] [Medline: 37927903]
2. Unnecessary back surgery: older Americans put at risk while billions in Medicare funds wasted. Lown Institute. URL: https://lownhospitalsindex.org/unnecessary-back-surgery [accessed 2024-12-23]
3. Riddle DL, Jiranek WA, Hayes CW. Use of a validated algorithm to judge the appropriateness of total knee arthroplasty in the United States: a multicenter longitudinal cohort study. Arthritis Rheumatol 2014 Aug;66(8):2134-2143. [doi: 10.1002/art.38685] [Medline: 24974958]
4. Jacobs JC, Jarvik JG, Chou R, et al. Observational study of the downstream consequences of inappropriate MRI of the lumbar spine. J Gen Intern Med 2020 Dec;35(12):3605-3612. [doi: 10.1007/s11606-020-06181-7] [Medline: 32989711]
5. Khan MM, Pincher B, Pacheco R. Unnecessary magnetic resonance imaging of the knee: how much is it really costing the NHS? Ann Med Surg (Lond) 2021 Oct;70:102736. [doi: 10.1016/j.amsu.2021.102736]
6. Sogbein OA, Chen AG, McClure JA, et al. Unnecessary interventions for the management of hip osteoarthritis: a population-based cohort study. Can J Surg 2024;67(4):E300-E305. [doi: 10.1503/cjs.001624] [Medline: 39089818]
7. Shelton S, Yochum A DO, Walker E, Waters N. Effectiveness of epidural steroid injections for low back pain. Am Fam Physician 2022 Jul;106(1):89-90. [Medline: 35839356]
8. Hastings CS, Schilthuis AJ, Jackson CD. Physical therapy is better than glucocorticoid injections for osteoarthritis of the knee. J Gen Intern Med 2021 Oct;36(10):3301-3303. [doi: 10.1007/s11606-021-06870-x] [Medline: 34322835]
9. National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Board on Health Care Services; Committee on Identifying Disabling Medical Conditions Likely to Improve with Treatment. Selected Health Conditions and Likelihood of Improvement with Treatment: National Academies Press; 2020.
10. Wu V, Goto K, Carek S, et al. Family medicine musculoskeletal medicine education: a CERA study. Fam Med 2022 May;54(5):369-375. [doi: 10.22454/FamMed.2022.975755]
11. APTA report: use of direct access among PTs is widespread, but barriers need to be addressed. American Physical Therapy Association. 2017. URL: https://www.apta.org/article/2017/10/20/apta-report-use-of-direct-access-among-pts-is-widespread-but-barriers-need-to-be-addressed [accessed 2025-02-12]
12. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2021;2:100117. [doi: 10.1016/j.sintl.2021.100117] [Medline: 34806053]
13. Laskowski ER, Johnson SE, Shelerud RA, et al. The telemedicine musculoskeletal examination. Mayo Clin Proc 2020 Aug;95(8):1715-1731. [doi: 10.1016/j.mayocp.2020.05.026] [Medline: 32753146]
14. Vincent R, Charron M, Lafrance S, Cormier AA, Kairy D, Desmeules F. Investigating the use of telemedicine by health care providers to diagnose and manage patients with musculoskeletal disorders: systematic review and meta-analysis. J Med Internet Res 2024 Sep 23;26:e52964. [doi: 10.2196/52964] [Medline: 39312765]
15. Bargeri S, Castellini G, Vitale JA, et al. Effectiveness of telemedicine for musculoskeletal disorders: umbrella review. J Med Internet Res 2024 Feb 2;26:e50090. [doi: 10.2196/50090] [Medline: 38306156]
16. Vori Health. URL: https://www.vorihealth.com [accessed 2024-12-23]
17. Barry MJ, Edgman-Levitan S. Shared decision making—pinnacle of patient-centered care. N Engl J Med 2012 Mar 1;366(9):780-781. [doi: 10.1056/NEJMp1109283] [Medline: 22375967]
18. Simonsmeier BA, Flaig M, Simacek T, Schneider M. What sixty years of research says about the effectiveness of patient education on health: a second order meta-analysis. Health Psychol Rev 2022 Sep;16(3):450-474. [doi: 10.1080/17437199.2021.1967184] [Medline: 34384337]
19. AAFP core educational guidelines: recommended core educational guidelines for family practice residents. American Academy of Family Physicians. URL: https://www.aafp.org/pubs/afp/issues/2000/1001/p1712.html [accessed 2024-12-23]
20. Gallotti M, Campagnola B, Cocchieri A, Mourad F, Heick JD, Maselli F. Effectiveness and consequences of direct access in physiotherapy: a systematic review. J Clin Med 2023 Sep 7;12(18):5832. [doi: 10.3390/jcm12185832] [Medline: 37762773]
21. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282. [Medline: 23092060]

XSL•FO
RenderX

22. Lafrance S, Vincent R, Demont A, Charron M, Desmeules F. Advanced practice physiotherapists can diagnose and triage patients with musculoskeletal disorders while providing effective care: a systematic review. J Physiother 2023 Oct;69(4):220-231. [doi: 10.1016/j.jphys.2023.08.005] [Medline: 37714771]
23. Madsen MN, Kirkegaard ML, Klebe TM, et al. Inter-professional agreement and collaboration between extended scope physiotherapists and orthopaedic surgeons in an orthopaedic outpatient shoulder clinic—a mixed methods study. BMC Musculoskelet Disord 2021 Jan 4;22(1):4. [doi: 10.1186/s12891-020-03831-z] [Medline: 33397350]
24. Lowry V, Bass A, Lavigne P, et al. Physiotherapists' ability to diagnose and manage shoulder disorders in an outpatient orthopedic clinic: results from a concordance study. J Shoulder Elbow Surg 2020 Aug;29(8):1564-1572. [doi: 10.1016/j.jse.2019.11.030] [Medline: 32199757]
25. Robarts S, Stratford P, Kennedy D, Malcolm B, Finkelstein J. Evaluation of an advanced-practice physiotherapist in triaging patients with lumbar spine pain: surgeon-physiotherapist level of agreement and patient satisfaction. Can J Surg 2017 Aug;60(4):266-272. [doi: 10.1503/cjs.013416] [Medline: 28730987]
26. Lewis KJ, Coppieters MW, Vicenzino B, Hughes I, Ross L, Schmid AB. Occupational therapists, physiotherapists and orthopaedic surgeons agree on the decision for carpal tunnel surgery. Int J Health Policy Manag 2022 Jul 1;11(7):1001-1008. [doi: 10.34172/ijhpm.2020.227] [Medline: 33590739]
27. Lowry V, Bass A, Vukobrat T, et al. Higher psychological distress in patients seeking care for a knee disorder is associated with diagnostic discordance between health care providers: a secondary analysis of a diagnostic concordance study. BMC Musculoskelet Disord 2021 Jul 30;22(1):650. [doi: 10.1186/s12891-021-04534-9] [Medline: 34330250]
28. Longstaffe R, Slade Shantz J, Leiter J, Peeler J. Surgeon-therapist communication: do all members see eye-to-eye? Phys Sportsmed 2015 Nov;43(4):381-387. [doi: 10.1080/00913847.2015.1077096] [Medline: 26309155]
29. Schultzel M, Scheidt KB, McNeill B, Klein CM, Blout C, Itamura JM. Surgeon-therapist communication must be improved in rotator cuff repair rehabilitation: an electronic survey of physical therapists on postoperative rehabilitation protocols and communication with treating surgeons. Perm J 2021 May;25:20.088. [doi: 10.7812/TPP/20.088] [Medline: 33970072]
30. Boissonnault WG, Ross MD. Physical therapists referring patients to physicians: a review of case reports and series. J Orthop Sports Phys Ther 2012 May;42(5):446-454. [doi: 10.2519/jospt.2012.3890] [Medline: 22282166]
31. Ojha HA, Wyrsta NJ, Davenport TE, Egan WE, Gellhorn AC. Timing of physical therapy initiation for nonsurgical management of musculoskeletal disorders and effects on patient outcomes: a systematic review. J Orthop Sports Phys Ther 2016 Feb;46(2):56-70. [doi: 10.2519/jospt.2016.6138] [Medline: 26755406]
32. Deslauriers S, Déry J, Proulx K, et al. Effects of waiting for outpatient physiotherapy services in persons with musculoskeletal disorders: a systematic review. Disabil Rehabil 2021 Mar;43(5):611-620. [doi: 10.1080/09638288.2019.1639222] [Medline: 31304824]
33. CAHPS Clinician & Group Survey. URL: https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/cg/adult-eng-cg40-3351a.pdf [accessed 2024-12-23]
34. Meyer VM, Benjamens S, Moumni ME, Lange JFM, Pol RA. Global overview of response rates in patient and health care professional surveys in surgery: a systematic review. Ann Surg 2022 Jan 1;275(1):e75-e81. [doi: 10.1097/SLA.0000000000004078] [Medline: 32649458]
35. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004;6(3):e34. [doi: 10.2196/jmir.6.3.e34]
36. Johnston S, Hogg W, Wong ST, Burge F, Peterson S. Differences in mode preferences, response rates, and mode effect between automated email and phone survey systems for patients of primary care practices: cross-sectional study. J Med Internet Res 2021 Jan 11;23(1):e21240. [doi: 10.2196/21240] [Medline: 33427675]
37. Slater M, Kiran T. Measuring the patient experience in primary care: comparing e-mail and waiting room survey delivery in a family health team. Can Fam Physician 2016 Dec;62(12):e740-e748. [Medline: 27965350]
38. Evaluation of the primary care patient experience survey tool. Te Tāhū Hauora Health Quality & Safety Commission. URL: https://www.hqsc.govt.nz/assets/Our-data/Publications-resources/MoH-PES-report-18April2018_2.pdf [accessed 2024-12-23]
39. Hager MM, Stein QP, Hendricks E, et al. Patient satisfaction and outcomes following a session with a genetic counselor for kidney diseases. J Am Soc Nephrol 2024;35(10S). [doi: 10.1681/ASN.2024w7s5gcnj]
40. Gustavsson L, Mohaddes M, Samsson K, Beischer S. No major difference in perceived quality of care in patients with hip or knee osteoarthritis assessed in a physical therapy-led triage compared with standard care: a randomized controlled trial. BMC Musculoskelet Disord 2023 Jun 29;24(1):530. [doi: 10.1186/s12891-023-06659-5] [Medline: 37386368]
41. Samsson KS, Bernhardsson S, Larsson MEH. Perceived quality of physiotherapist-led orthopaedic triage compared with standard practice in primary care: a randomised controlled trial. BMC Musculoskelet Disord 2016 Jun 10;17(17):257. [doi: 10.1186/s12891-016-1112-x] [Medline: 27286829]
42. Health literacy fact sheets. Center for Health Care Strategies. URL: https://www.chcs.org/resource/health-literacy-fact-sheets [accessed 2024-12-22]
43. Health literacy: statistics at-a-glance. AAACEUs. URL: https://www.aaaceus.com/courses/nl0610/article1.pdf [accessed 2024-12-22]

XSL•FO
RenderX

Abbreviations

APPT: advanced practice physical therapist
CAPHS: Consumer Assessment of Healthcare Providers and Systems
MD+PT: musculoskeletal medical doctor and physical therapist

Edited by A Mavragani; submitted 22.09.24; peer-reviewed by M Heng, R Vincent; revised version received 22.01.25; accepted 23.01.25; published 03.03.25.

Please cite as:
O'Connor MI, Chudy C, Peters KC, Ribaudo M, McCulloch C, Aguilar J, Taylor T, Grant RA
Patients' Experience With Evaluation by Both a Musculoskeletal Physician and Physical Therapist in the Same Digital Visit: Survey Study
JMIR Form Res 2025;9:e66744
URL: https://formative.jmir.org/2025/1/e66744
doi:10.2196/66744

© Mary I O'Connor, Carolyn Chudy, Kaitlyn C Peters, Megan Ribaudo, Carrie McCulloch, Jared Aguilar, Trista Taylor, Ryan A Grant. Originally published in JMIR Formative Research (https://formative.jmir.org), 3.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

App- and Wearable-Based Remote Monitoring for Patients With Myasthenia Gravis and Its Specialists: Feasibility and Usability Study

Maike Stein[1,2,3,4], MD; Regina Stegherr[5], PhD, Dr rer nat; Pushpa Narayanaswami[3], MD; David Legg[1], MSc; Meret Herdick[1,4], MD; Andreas Meisel[1,4,6], MD; Lea Gerischer[1,2,4*], MD; Sophie Lehnerer[1,2,4,6*], MD

[1]Department of Neurology with Experimental Neurology, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Charitéplatz 1, Berlin, Germany

[2]Digital Health Center, Berlin Institute of Health at Charité, Charité Universitätsmedizin Berlin, Berlin, Germany

[3]Department of Neurology, Beth Israel Deaconess Medical Center/Harvard Medical School, Boston, MA, United States

[4]Neuroscience Clinical Research Center, Department of Neurology with Experimental Neurology, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[5]Institute of Biometry and Clinical Epidemiology, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[6]Center for Stroke Research Berlin, Charité Universitätsmedizin Berlin, Berlin, Germany

[*]these authors contributed equally

Corresponding Author:
Maike Stein, MD
Department of Neurology with Experimental Neurology, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Charitéplatz 1, Berlin, Germany

Abstract

Background: Myasthenia gravis (MG) is rare, chronic autoimmune disorder of the neuromuscular junction that requires specialized care and often lifelong treatment, facing challenges due to its rarity and the limited availability of specialists. Telemedical solutions in specialized centers hold considerable promise in bridging this gap by increasing access to this care to a broader patient population in a timely manner. However, there is no research regarding interventional remote care solutions in the field of MG to date.

Objective: This study aimed to assess the feasibility and usability among patients with MG and specialists of a telemedicine platform, tailored to patients with MG and designed to facilitate remote monitoring, treated in a specialized center.

Methods: The telemedicine platform consisted of an app for patients and a web-based portal for physicians. Over a period of 3 months, 30 patients continuously monitored their vital parameters through external devices, including a digital spirometer and a wearable (activity tracker). Furthermore, patients completed 7 different patient-reported outcome measures (PROMs) through the app at predefined intervals. Specialists could review this monitoring data and adjust therapy as necessary. In addition, communication between patients and physicians was facilitated through a chat module. Feasibility was evaluated by total adherence rates for completing PROMs within the app, alongside the collection of spirometry and wearable data. Furthermore, user satisfaction was assessed among both patients with MG and physicians at the end of study.

Results: Total adherence rates ranged from 74.3% (1830/2464) to 97.9% (327/334) across all data types, with the highest adherence observed for PROMs (1139/1179, 96.6%), followed by spirometry (293/334, 87.7%) and wearables (1830/2261, 80.9%). Notably, patients wore the wearable longer than required by protocol and conducted a higher number of spirometry measurements during the study than required per protocol (median 20 h/d [IQR 15-24] vs 14 h/d and median 49 [IQR 15-59] measurements vs 11 measurements, respectively). Technical issues and discomfort with wearables were factors affecting lower adherence in some patients. The System Usability Scale yielded a median score of 85 indicating “excellent usability.” In addition, results from a more detailed user evaluation questionnaire showed high levels of user satisfaction among both patients and health care professionals across diverse categories, including their experience of the care program, communication, and evaluation of the program.

Conclusions: Remote monitoring of patients with MG through the telemedical platform demonstrated good feasibility and acceptability, as evidenced by above-average adherence rates and user satisfaction for both patients and physicians. The majority of patients wanted to continue using the app. These findings highlight the potential for user-friendly digital tools to enhance care for patients with MG, although addressing technical challenges and ensuring patient comfort with wearables are essential for

XSL•FO
RenderX

optimal implementation. Further research involving larger cohorts and longer study duration is warranted to validate these findings.

Trial Registration: Deutsches Register Klinischer Studien DRKS00029907; https://drks.de/search/en/trial/DRKS00029907

(JMIR Form Res 2025;9:e58266) doi:10.2196/58266

KEYWORDS

myasthenia gravis; myasthenia; remote monitoring; PROMs; digital platform; wearables; telemedicine; spirometry; app; usability; feasibility; autoimmune disorder; web-based portal; activity tracker; communication; wearable data; digital tool; mobile phone

Introduction

Myasthenia gravis (MG) is a rare autoimmune disease characterized by specific autoantibodies targeting the postsynaptic membrane of the neuromuscular junction, resulting in fluctuating fatigability and weakness of ocular, bulbar, and skeletal muscles, with the potential for life-threatening crises [1]. MG has a bimodal incidence pattern, with younger women and older men being most commonly affected [2]. Given its chronic nature, the majority of patients with MG require long-term and often lifelong, highly specialized care and therapy [3]. Furthermore, studies have shown that patients with MG experience a significantly reduced quality of life [4] and have a high demand for individualized counseling on various topics [5] and supportive interactions. In addition, characteristic symptom fluctuations are not captured effectively by conventional clinical assessments that rely on infrequent in-person assessments that are further challenged due to documentation overload for physicians, limited consultation time, and geographical barriers that limit access to specialists [6] . The characteristic symptom fluctuations in MG, but also the long intervals for therapies to effectively influence the disease including potential side effects, give rise to questions and challenges for timely therapy adjustments and individual counseling. However, in practice, these adjustments often do not occur in a timely manner or are conducted in ways that pose data privacy problems, are resource intensive and lack standardization. These challenges and the growing demand for personalized and on-demand available health care have spurred the development of telemedical and wearable devices for remote patient monitoring [7]. Wearable devices such as activity trackers facilitate continuous monitoring of various health parameters and have shown promise in reducing hospitalization costs, improving personalized health management [8-10] and health‐related quality of life [11]. Those digital health interventions hold potential for bridging the care gap by offering individualized support based on continuous health data derived from both subjective patient-reported outcome measures (PROMs) and objective wearable and sensor-based data. Although previous remote digital studies have shown the feasibility and usability of leveraging smartphones and wearable technology for assessing clinical changes in different diagnosis and naturalistic settings [12-14], there is to date no data available on telemedical interventions for patients with MG. Generally, successful integration of these technologies into health care relies on patients' active involvement and acceptance, underscoring the importance of digital literacy [15,16]. In the context of remote monitoring technologies, adherence serves as a crucial measure of patient compliance with study protocols (eg, the completion rate of providing data inputs at specified frequencies outlined in the protocol) and is dependent on their willingness and capacity to adopt new behaviors and practices. We hypothesize that the integration of digital tools into clinical care presents a viable approach for remotely monitoring the condition and symptoms of patients with MG. The aim of this study was to evaluate the feasibility and user satisfaction levels among patients with MG and health care professionals with a telemedicine platform specifically tailored to monitor patients with MG and enabling patient-physician interaction.

Methods

Telemedical Platform for Patients with MG

The telemedical platform for patients with MG, "MyaLink" used in the study was developed at Charité Universitätsmedizin Berlin together with software partner Qurasoft GmbH (based on the SaniQ platform). The platform is a Conformité Européenne (CE)-marked class-I medical device (Medical Device Directive) and is GDPR (General Data Protection Regulation) and BlnDSG (Berlin Data Protection Act) compliant. The platform consists of an app for patients (iOS and Android) and a web-based portal for physicians facilitating to review patient data, manage and set up individual monitoring plans and medication plans. It offers a secure interactive platform for exchange of information and health data between physicians and patients and informed treatment decisions (Figure 1). The MyaLink app contains the following modules: (1) transmission and integration of sensor-based data from wearables including heart rate, oxygen saturation, and step count (device: Garmin Vivosmart 5); (2) assessment of the forced vital capacity from a digital spirometer (device: MIR Spirobank Smart One); (3) assessment of questionnaires and PROMs (see details under Section 3.5); and a (4) communication module for individual requests and patient-physician interaction. Furthermore, the app contains a (5) medication plan with reminder function and (6) a module including data storage of important documents (eg, clinical reports), which can also be shared with the physician through the web-based portal. The web-based system for physicians enables real time access to this data and allows clinicians to obtain an accurate overview of the patients' condition and visualization of health data. If necessary, the patient can be contacted through the communication tool and adjustments to medication and the remote monitoring plans can be made.

XSL•FO
RenderX

Figure 1. Schematic overview of telemedical platform and its features (MyaLink): transmission and simultaneous integration of sensor-based data into the app is implemented by bluetooth from coupled external devices (wearable or digital spirometer). In addition, the platform facilitates the collection of patient-reported outcome measures (PROMs) through the app and a communication module enables interaction between patients and physicians. Health data can be shared with specialists who can review monitoring parameters through the web-based system and make therapy adjustments as needed. The data is transmitted between patient and physicians with double encryption (AES256, Advanced Encryption Standard). PROM: patient-reported outcome measure.

Subjects' Eligibility

Inclusion and exclusion criteria have been previously presented in detail in our methods paper [17]. Inclusion criteria included patients with a diagnosis of MG for at least 6 months before study enrollment, aged 18 years and older, and classified according to the Myasthenia Gravis Foundation of America (MGFA) classification status I-IV [18]. Exclusion criteria included patients with somatic, psychiatric, or neurological disorders who were undergoing treatments or were not stabilized under drug therapy that could adversely affect cognitive function. Patients were recruited through the specialized outpatient myasthenia center at Charité Universitätsmedizin Berlin.

Study Design

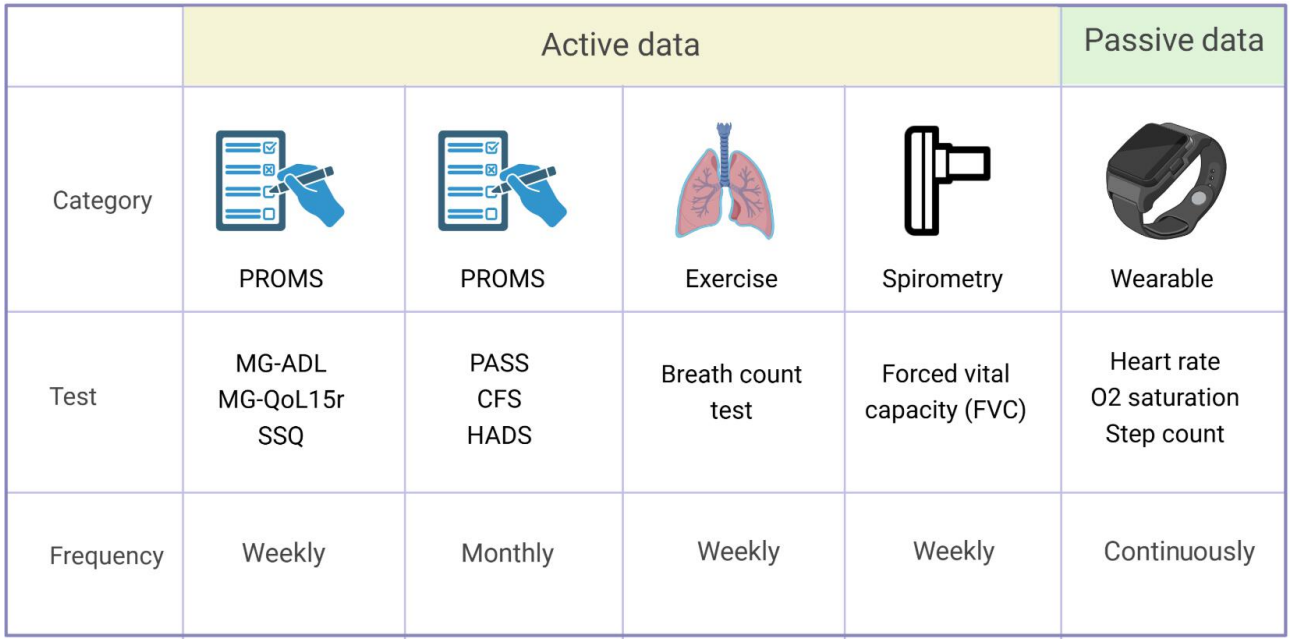
This analysis focused on data obtained from the telemedical intervention group from a randomized controlled trial (RCT) conducted at our specialized center at Charité Universitätsmedizin Berlin in Germany. The RCT included 45 patients with MG over a 12-week observation period between April and September 2023. Patients were randomized in a 2:1 ratio into the telemedical intervention group (n=30) or the control group (n=15), who received standard of care without telemedical intervention. Randomization was performed according to gender and clinical class stratified by MGFA score at baseline (using the randomizer tool [19]). The primary objective of the study was to investigate the feasibility of the remote monitoring platform in care of patients with MG. Feasibility was measured through adherence rates based on completion of PROMs in the app and data collection from external devices (including a digital spirometer and activity tracker). Furthermore, user satisfaction with the telemedical platform was assessed among patients and health care professionals comprised of the study team (n=5) at the end of the study. We, therefore, exclusively analyzed data from the 30 patients from the intervention group. Secondary objectives of the study (but not part of this analysis) included comparing standard of care with and without additional telemedical intervention. This comparison focused on evaluating clinical endpoints between the control and intervention group. Comprehensive baseline and end-of-study visits were conducted at the study center for both groups and included an in-person clinical assessment of medical history, medication, hospitalizations for MG, care-related questions (eg, accessibility of a specialist or areas of improvement regarding MG treatment), various PROMs and MG-specific clinical assessment through the quantitative myasthenia gravis score (QMGS) [20]. At baseline, the intervention group received a wearable (activity tracker) and a digital spirometer, along with the MyaLink app, which were set up and connected on the patients' smartphone. Instructions on app and external devices usage were provided by the study team. This enabled remote health monitoring throughout the study by active (PROMs and spirometry) and passive (wearable) continuous data collection. Patients were instructed to fill out PROMs and perform spirometry at predefined intervals. They were instructed to wear the wearable during waking hours (at least 14 h per day) throughout the study period and optionally at night. Telemedical check-ups occurred in weeks 4 and 8, comprising the reviewing of remote monitoring data by study physicians and contacting the patients if necessary (eg, changing medication). Patients could also initiate communication through the chat module at any time, with physicians monitoring messages on weekdays during working hours and responding within 24 hours. Patients completed usability questionnaires through the app a week before the end of the study, while health care professionals conducted their questionnaires after study completion. Furthermore, oral feedback from patients was collected and documented during the end-of-study visit. All data from the study visits and all data from the telemedical check-up was documented in REDCap (Research Electronic Data Capture) software (version 13.7.31; Vanderbilt University).

XSL•FO
RenderX

Patient-Reported Outcome Measures and Data From App

In total, 6 questionnaires and 1 test were assigned to patients through the app, administered weekly or monthly, depending on their specific type (Figure 2). Weekly questionnaires included the myasthenia gravis activities of daily living (MG-ADL) profile [21], myasthenia gravis quality of life, revised (MG-QoL15r) [22], and single simple question (SSQ) [23], which are commonly used in MG studies including large approval studies to assess functional status and quality of life. Patients were also instructed to perform a weekly breath count test [24]. Monthly questionnaires included Chalder Fatigue Scale (CFS 11) [25] and Hospital Anxiety and Depression Scale (HADS) [26] to assess common comorbidities in MG. In addition, Patient Acceptable Symptom State (PASS) was assessed monthly [27]. Furthermore, patients completed 2 user satisfaction questionnaires at the end of the study. For broad comparability, patients completed the System Usability Scale (SUS) [28], where patients ranked each of the 10 statements from 1 to 5 based on how much they agreed with the statement. The second questionnaire was adapted from Braun et al [29] and specifically tailored to the MyaLink telemedical platform. It features 17 items categorized in 3 sections and uses a 4-point Likert scale (strongly agree to strongly disagree). Originally formulated in German for publication purposes, it underwent consensus translation into English by 4 medical health care professionals, including 1 native speaker. After the study, the study team completed a modified version of the questionnaire adapted from Braun et al for health care professionals, respectively.

Figure 2. Monitoring plan and frequencies of assessed parameters: Active (PROMs, respiratory function assessment) and passive data (wearable data) were collected during the study period. PROMs data were collected weekly or monthly, depending on the questionnaire. Respiratory function assessment, assessed through breath count test and spirometry, was conducted weekly by the protocol. The collection of wearable data was continuous, with steps aggregated by the hour. PROMs: patient-reported outcome measures; MG-ADL: myasthenia gravis activities of daily living; MG-QoL15r: myasthenia gravis quality of life, revised; SSQ: single simple question; PASS: Patient Acceptable Symptom State; CFS: Chalder Fatigue Scale; HADS: Hospital Anxiety and Depression Scale.



Data From External Device or Wearable

Patients were instructed to perform spirometry measurements weekly, with reminders sent through the app through push notifications. The wearable continuously collected oxygen saturation level (every minute), heart rate (every minute), and steps (hourly aggregated). Data from both the digital spirometer and the wearable were transmitted directly to the app by bluetooth. The study team regularly checked the web-based system to ensure receipt of data from external devices and questionnaires, sending reminder messages through the chat module when necessary.

Analysis

The statistical calculations were performed using R software (version 4.2.2; R Foundation for Statistical Computing) [30]. Descriptive statistics (median [IQR] and absolute and relative frequencies) are presented depending on the scale of the variables.

Feasibility was assessed by adherence rates. For PROMs, assignments of the questionnaire started a week or month after the baseline visit respectively, depending on the questionnaire. Questionnaires were assigned together with a push notification on the patient's phone and disappeared from the app upon completion or were "overwritten" upon receiving a new assignment. Questionnaires could only be submitted by patients through the app if they were fully completed. Each questionnaire was required to be completed within 7 days to be counted as "completed," including the monthly assigned questionnaires, with missing entries calculated accordingly. The final day for data acquisition from the last assigned weekly questionnaires ended the day before the end-of-study visit. Total adherence rates were calculated as the total number of completed questionnaires divided by the total number of assigned

XSL•FO
RenderX

questionnaires. Additional questionnaire assignments were allowed in cases of clinical deterioration but were excluded from adherence calculations. The study was considered successful if at least 60% of assigned PROMs through the app were answered by patients of the intervention group. Across the 30 patients of the intervention group, a mean adherence rate and a 2-sided 95% CI are determined. If the observed average adherence is at least 76%, with data from 30 patients, the 95% CI would range from 60.7% to 91.3%, thus exceeding the 60% threshold.

Spirometry adherence rates were calculated accordingly (ratio of the total number of actual measurements performed during the study vs total number of measurements required per the study protocol).

Adherence calculations for wearable data included measurements starting the day after the baseline visit and ending the day before the end-of-study visit. Adherence to wearable data from the activity tracker was evaluated by aggregating data over hourly windows, with 100% adherence defined as wearing the device for 14 hours per day per patient (based on the estimated waking hours). This was then extrapolated over the study duration. Each hour necessitated at least one heart rate measurement to validate wear time, considering heart rate a more reliable marker than oxygen saturation. Overall adherence was determined by the sum of the hours the patients had with at least one heart rate measurement divided by sum of the hours per day corresponding to 100% individual adherence.

The user satisfaction questionnaire was descriptively evaluated using the Likert scale. The SUS ranked each of the 10 statements from 1 to 5 based on user agreement, from which a SUS score was calculated: Each patient responses undergo conversion into a new numerical value (0 - 4), which is then aggregated and multiplied by 2.5 to transform the original scoring range of 0 - 40 into a 0 - 100 scale. These scores do not represent percentages but a percentile ranking [31]. Research suggests that a SUS score above 68 is considered acceptable, with scores below 68 classified as not acceptable. Therefore, a score above 68 on the SUS is interpreted as a positive result further supporting the platform's usability. A more detailed classification considers SUS scores between 73 and 84 as good, between 85 and 99 as excellent, and 100 as the best.

This analysis focuses on the primary study endpoints assessing feasibility and usability. Further exploratory analyses of secondary endpoints, eg, comparative clinical endpoint analyses (between control group and intervention group), evaluations of communication patterns from the MyaLink chat and telemedical interventions are currently prepared and results will be presented in further studies.

Ethical Considerations

This study received ethics approval by the ethics committee at Charité Universitätsmedizin Berlin (EA2/157/22). Written informed consent was obtained from the study participants before the start of the study. The study was conducted in accordance with the Declaration of Helsinki (2013) and its later amendments. Data storage and handling responsibilities were approved by the ethics committee of Charité Universitätsmedizin Berlin. The data collected through MyaLink is stored on certified data servers, which adhere to a management system according to ISO/IEC 27001:2013.

Results

Patient Characteristics

Patients were between 23 and 82 years (median 47.5 years; IQR 40.5-57.5) at time of enrollment. The majority of patients were female (21/30, 70%). Most patients in the study had a mild to moderate disease severity, with 36.7% (11/30) categorized in MGFA II (a or b) status and 43.3% (13/30) in MGFA III (a or b) status at the time of enrolment. Median scores for MG-ADL and QMG were 10 (IQR 5.25-11) and 15.5 points (IQR 12.25-20.75), respectively. On average, patients were diagnosed 5.5 years (IQR 2-9.75) before entry in the study. One patient withdrew from the study due to health reasons unrelated to MG. All results presented include data from the dropout, including measurements taken before the time of dropout (with the patient's consent). Detailed patient characteristics are presented in Table 1.

XSL•FO
RenderX

Table . Patient characteristics: age, clinical class, time since diagnosis, study duration, and operating system for intervention group as well as subgroups (female and male).

Characteristics	Overall (n=30)	Female (n=21)	Male (n=9)
Age, median (IQR)	47.5 (40.5 - 57.5)	47 (42 - 53)	56 (40 - 60)
MGFA[a] score, n (%)			
I	1 (3.3)	0 (0)	1 (11.1)
IIa	5 (16.7)	4 (19)	1 (11.1)
IIb	6 (20)	3 (14.3)	3 (33.3)
IIIa	3 (10)	2 (9.5)	1 (11.1)
IIIb	10 (33.3)	8 (38.1)	2 (22.2)
IVa	0 (0)	0 (0)	0 (0)
IVb	5 (16.7)	4 (19)	1 (11.1)
MG-ADL[b] score, median (IQR)	10 (5.25 - 11)	10 (5 - 11)	9 (6-12)
MG-QoL15r[c] score, median (IQR)	15.5 (12.25 - 20.75)	17 (12 - 21)	13 (13 - 19)
QMG[d] score, median (IQR)	15 (12.25 - 17)	15 (13 - 18)	14 (10 - 15)
Years since MG[e] diagnosis, median (IQR)	5.5 (2 - 9.75)	5 (2-10)	7 (4-9)
System, n (%)			
Android	19 (63.3)	15 (71.4)	4 (44.4)
iOS	11 (36.7)	6 (28.6)	5 (55.6)
Days between baseline and end-of-study visit, median (IQR)	84 (84 - 85)	84 (83 - 85)	84 (84 - 86)
Dropout, n (%)	1 (3.3)	1 (4.8)	0 (0)

[a]MGFA: Myasthenia Gravis Foundation of America.

[b]MG-ADL: myasthenia gravis activities of daily living.

[c]MG-QoL15r: myasthenia gravis quality of life, revised.

[d]QMG: quantitative myasthenia gravis.

[e]MG: myasthenia gravis.

Adherence

The total adherence rates for PROMs and external devices data ranged from 74.3%- 97.9% (1830/2464 - 327/334) across all data types and are detailed in Table 2. When examining adherence by data category, adherence to PROMs was the highest, followed by spirometry, and then wearable data from the activity tracker.

XSL•FO
RenderX

Table . Total adherence rates for different data groups.

Data group	Adherence (total completed/total assigned), n/N (%)
Overall PROMs[a]	1139/1179 (96.6)
MG-ADL[b]	321/334 (96.1)
MG-QoL15r[c]	325/334 (97.3)
SSQ[d]	327/334 (97.9)
CFS[e]	55/59 (93.2)
HADS[f]	5/59 (93.2)
PASS[g]	56/59 (94.9)
Breath count test	326/334 (97.6)
Spirometry	
FVC[h]	293/334 (87.7)
Wearables (at least 14 h per day)	
Pulse	1830/2464 (74.3)
Pulse (excluding days without any wearable measurement from the analysis)	1830/2261 (80.9)

[a]PROMS: patient-reported outcome measures.

[b]MG-ADL: myasthenia gravis activities of daily living.

[c]MG-QoL15r: myasthenia gravis quality of life, revised.

[d]SSQ: single simple question.

[e]CFS: Chalder Fatigue Scale.

[f]HADS: Hospital Anxiety and Depression Scale.

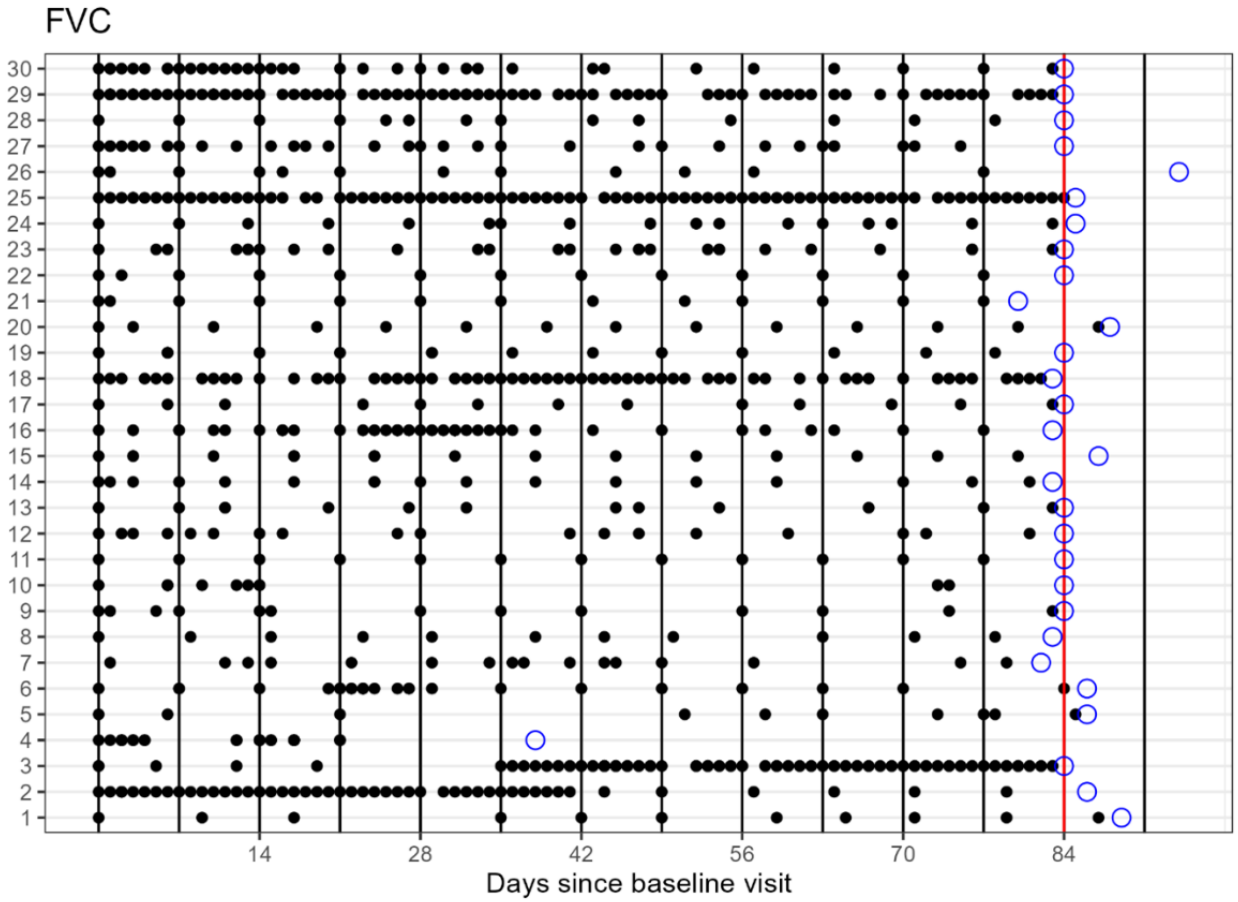
[g]PASS: Patient Acceptable Symptom State.

[h]FVC: forced vital capacity.

Adherence to PROMs ranged between 93.2% (55/59) and 97.9% (327/334), depending on the specific questionnaire. Patients were assigned weekly questionnaires and spirometry measurements between 11 and 13 times (drop out: 5 times) during the study, depending on individual study duration of a participant. Monthly questionnaires were assigned twice to all patients (drop out: 1). There was no difference in adherence regarding the questionnaire frequency. For the breath count test, adherence was 97.6% (326/334). Adherence to spirometry was slightly lower than PROMs (293/334, 87.7%), likely due to the higher task burden associated with these active measurements (Figure 3). Patients conducted a median of 49 spirometry measurements (IQR 15-59).

XSL•FO
RenderX

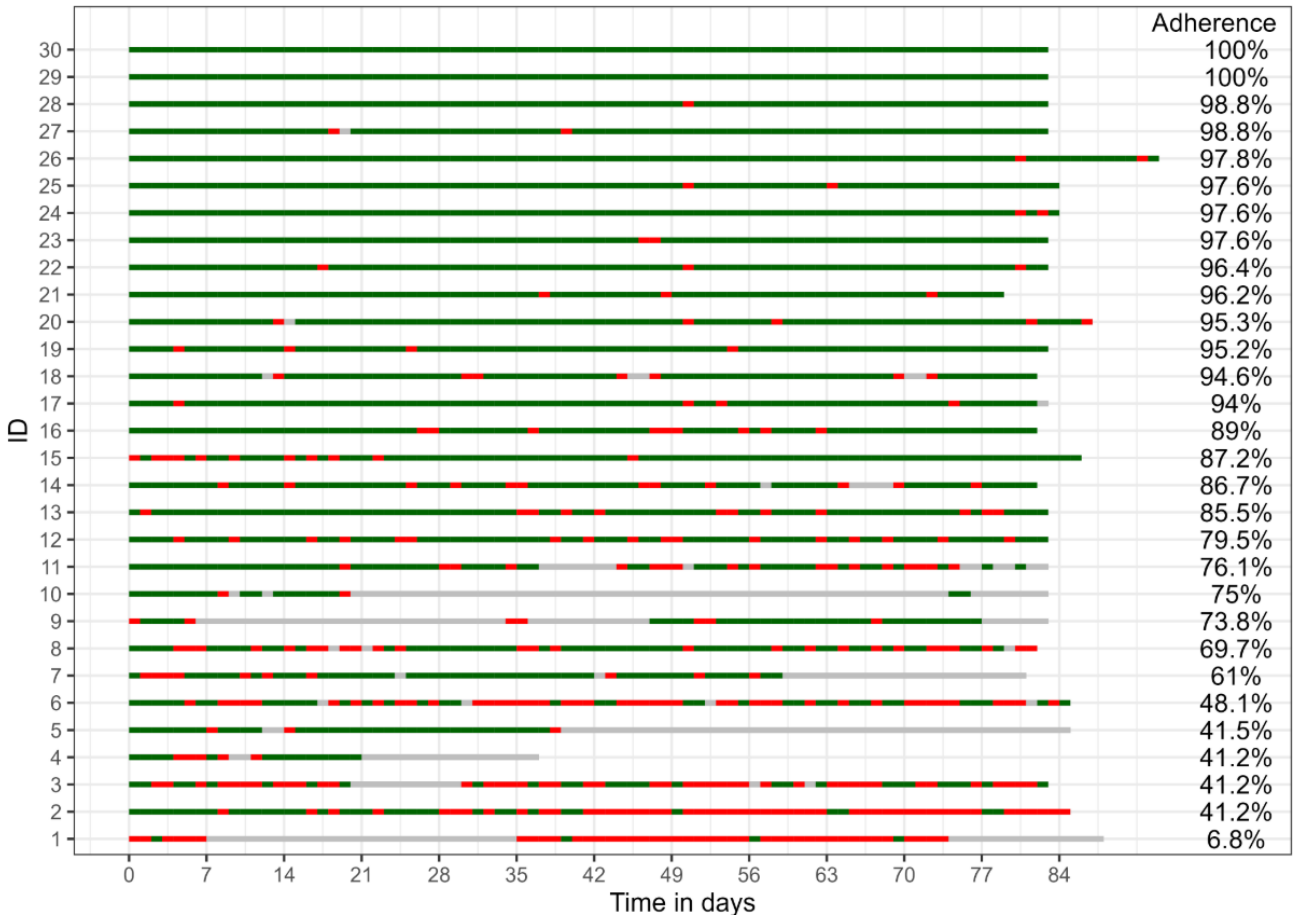
Figure 3. Adherence of spirometry measurements: The blue circles represent the end-of-study visit for each patient, while the red line indicates week 12, marking the duration of the study. Each black vertical line indicates when a spirometry measurement was assigned according to the study protocol (weekly intervals). Patients were given a 7-day window until the next weekly assignment for measurement to qualify as "completed." The black dots mark the times when the patient actually performed the spirometry measurement. Each measurement from the digital spirometer was timestamped and subsequently automatically transmitted to the app through bluetooth, with the values being integrated accordingly. FVC: forced vital capacity.



On average, patients wore the wearable for a median duration of 20 hours per day (IQR 15 - 24). Wearable adherence rates exhibited considerable variability among patients, ranging from 6.8% (4/59) to 100% (83/83) (Figure 4), despite an overall strong adherence of 80.9% (1830/2261) for all patients during the study when excluding days without any data input (no recorded heart rate measurements) from the calculation (Table 2). Including days without wearable data input to the calculation decreased adherence to 74.3% (1830/2464). Clinical insights from chat interactions, as evidenced by the timestamps within the chat messages, unveiled several factors impacting patients with decreased device use. For example, out of the 30 patients, 2 patients (6.7%) experienced a myasthenic crisis necessitating prolonged hospitalization and reported having forgotten the devices at home. Furthermore, 1 patient (3.3%) had difficulty with using the app (older patient who was overwhelmed by the technical operation and conducting measurements with the digital spirometer) with numerous interactions with the software provider's support team. Some patients experienced discomfort with the rubber strap of the activity tracker. Some patients had device-related issues such as (recurring) pairing problems and 1 out of the 30 patients (3.3%) received a defective activity tracker, necessitating a device replacement. However, even after replacement the patient reported ongoing connectivity issues between the wearable and his smartphone.

XSL•FO
RenderX

Figure 4. Adherence rates for wearable data from the activity tracker: Days since baseline visit on the X-axis, with patients represented on the Y-axis and the adherence rate per patient in the column on the right side. Days with wearable usage of at least 14 hours per day were classified as compliant (marked in green). A wearable hour was counted if at least one heart rate measurement value was recorded per hour. Days with less than 14 hours of wearable usage per day were marked as red. Days without any recorded heart rate measurements were marked in gray and excluded from the analysis.

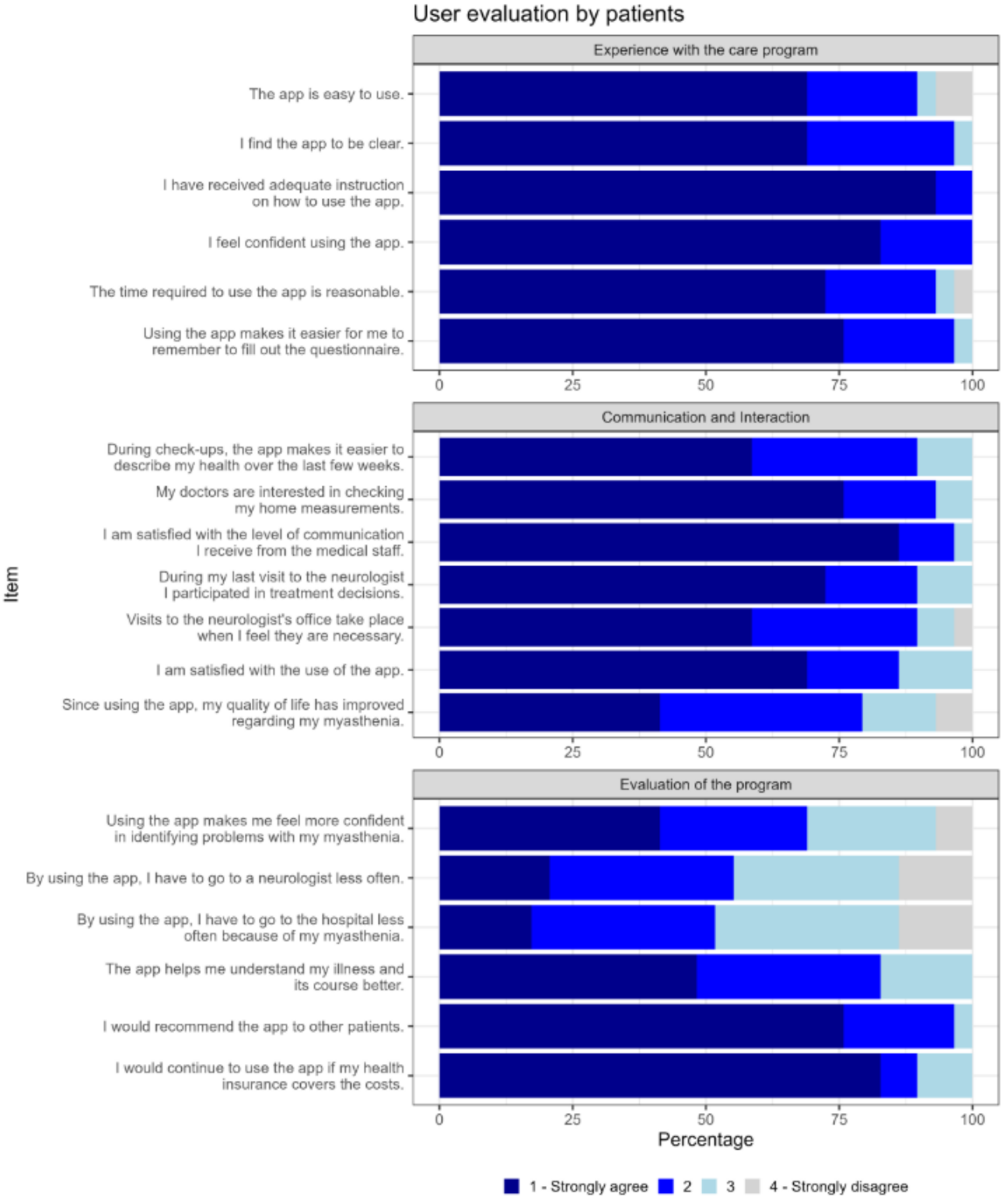


User Satisfaction

The highest levels of agreement were observed in the following order: experience with the care program, communication, and evaluation of the program. Regarding the care program evaluation, 89.7% (26/29) agreed that the app was easy to use while 93.1% (27/29) agreed that the time investment required for app usage was appropriate. In total, 89.6% (26/29) of patients agreed that describing their health status during follow-up appointments was easier. A total of 93.1% (27/29) indicated that treating physicians were interested in reviewing the collected monitoring parameters. In addition, 79.3% (23/29) of patients agreed that their quality of life had significantly improved since app usage. Futhermore, 96.6% (28/29) expressed willingness to recommend the app to other patients. In addition, 69% (20/29) agreed that they felt more confident in identifying MG-specific problems using the app. Views on the reduction of visits to neurologists or clinics were mixed; 55.2% (16/29) agreed to needing fewer visits to neurologists, while 51.7% (15/29) agreed to reduced visits to clinics, with 13.8% (4/29) strongly disagreeing with both categories (Figure 5). Out of the 29 patients, 14 patients (46.7%) responded that they did not agree that they required less frequent hospital visits; of those 14 patients, 8 patients (57.1%) were hospitalized during the study period due to MG (two of whom were in intensive care). In addition, 7 of these 14 patients (50%) experienced at least one exacerbation. Results of user evaluation from the health care professionals showed similar high acceptance rates regarding the telemedical platform, ease of use, and overall positive evaluation on the effect on patient care (Figure 6).

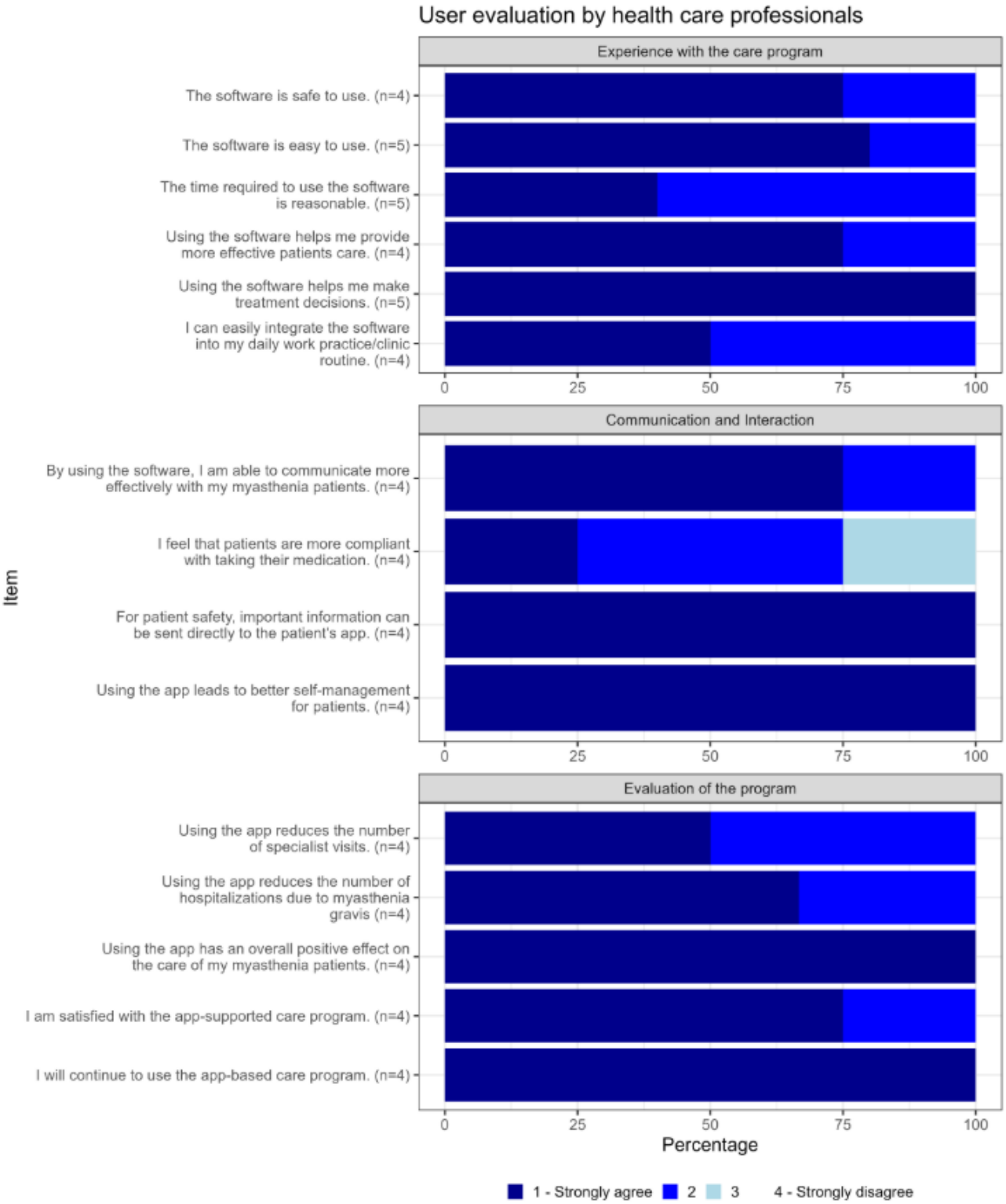
XSL•FO
RenderX

Figure 5. User evaluation for patients: The questionnaire was divided into 3 categories: “Experience with the care program,” “Communication and Interaction” and “Evaluation of the program.” All 17 questions could be answered using a 4-point Likert scale ranging from 1 (strongly agree) to 4 (strongly disagree).



XSL•FO
RenderX

Figure 6. User evaluation for health care professionals: The questionnaire was divided into 3 categories: "Experience with the care program," "Communication and Interaction," and "Evaluation of the program." All 17 questions could be answered using a 4-point Likert scale ranging from 1 (strongly agree) to 4 (strongly disagree). Depending on the role within the study team (eg, physician and study nurse), only certain questions in the questionnaire could be answered.

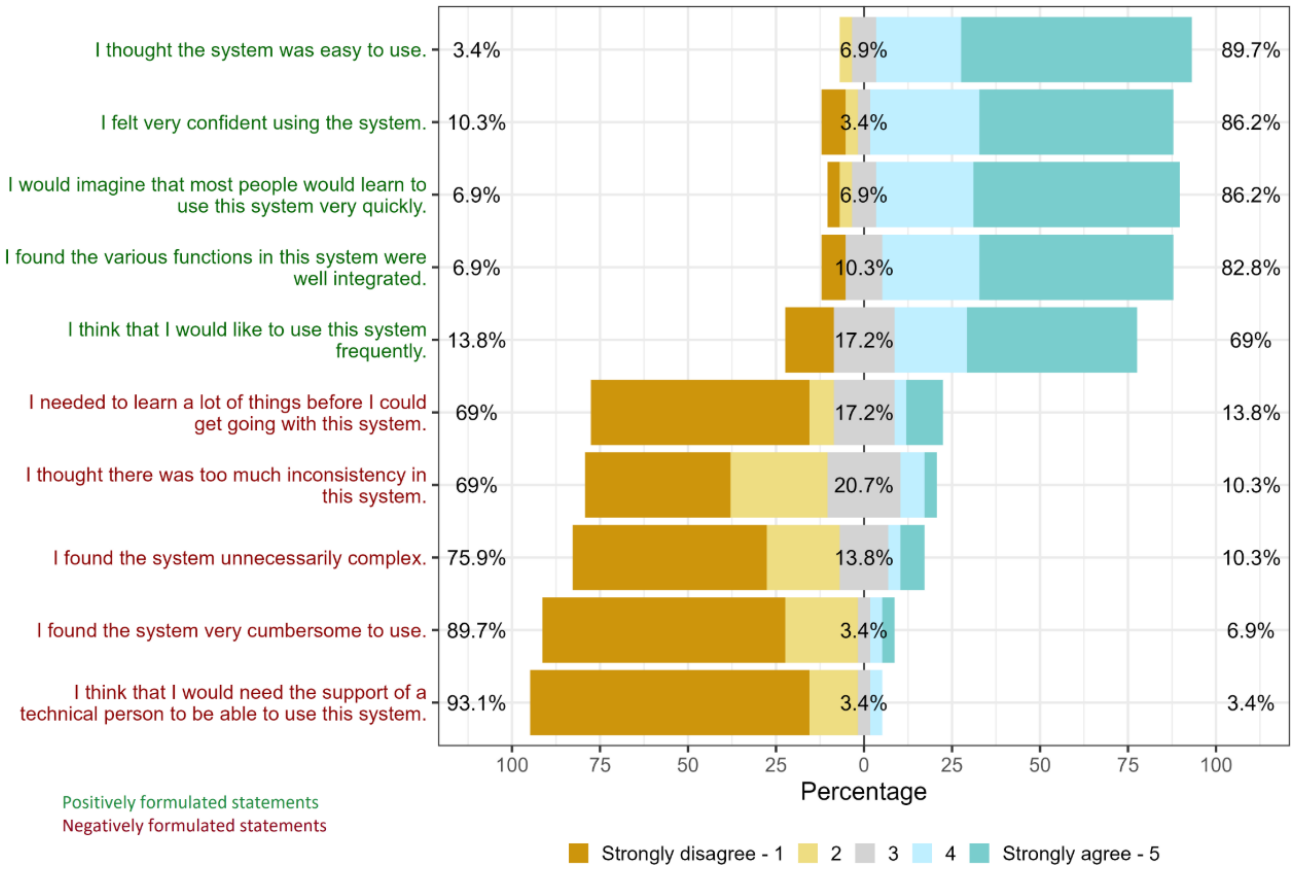


The overall median SUS score was 85 (IQR 75-90). Figure 7 shows that there was a high level of agreement regarding the system's ease of use (26/29, 90%), patients' confidence in operating it (25/29, 86%), and patients' belief that others would quickly learn to use the app (25/29, 86%). Notably, 21% (6/29) remained neutral regarding perceived inconsistencies in the app, although the majority (20/29, 69%) tended to disagree. Among patients (N=22) who responded at least once with a neutral (Likert-scale 3) or negative response to any question (agreeing or strongly agreeing with negatively formulated questions or disagreeing or strongly disagreeing with positively formulated questions), there were 14 females (63.6%) with a median age of 47.5 (IQR 39.25-61.25; mean 49.7, SD 15.70), and the MG-ADL score at baseline was median 9.5 (IQR 5-11; mean

XSL•FO
RenderX

8.5, SD 3.60). Of all 29 patients, 69% (20/29) expressed willingness to use the app regularly. Among the patients who responded negatively or neutral to this statement, the documented oral feedback collected at the end-of-study visit mainly encompassed technical feedback, such as suggestions for adjusting the size of the chat window, implementing filter options for monitoring parameters in the health diary, and presenting preferences for the parameter dashboard. Furthermore, patients also frequently noted necessary improvements for a more user-friendly medication plan. In addition, there were critiques regarding the omission of specific symptoms in the PROMs (eg, nonmotor symptoms, pain, and sleep) and requests from patients for a free-form diary function (eg, for documenting infections or daily fluctuations).

Figure 7. The System Usability Scale (SUS) uses a 10-item questionnaire with 5 response options for respondents, ranging from strongly agree to strongly disagree. Due to rounding, the percentage for Question 2 ("I feel confident using the app") is 99% instead of 100%.



Discussion

Our study evaluated the feasibility and usability of a telemedicine platform for patient-physician interaction and monitoring for MG, incorporating wearable data collection. Feasibility, measured by adherence, ranged from 74.3%- 97.9% (1830/2464 - 327/334) across all data types, with the highest adherence observed for completion of PROMs (1139/1179, 96.6%) within the app, followed by adherence to spirometry measurements (293/334, 87.7%), and then wearable adherence (1830/2261, 80.9%). Next to adherence, user satisfaction among patients and health care professionals was assessed and generally rated high to very high across different categories (care program, communication, and evaluation of the program) and considered "excellent" according to the SUS score. The majority of patients stated their willingness to continue using the app.

Our results demonstrate that the introduction of such a digital tool in a clinical setting is feasible and facilitates clinical monitoring regardless of age or gender. These findings represent the first results in the field of MG for telemedical intervention. There was mutual acceptance from both health care professionals and patients, as evidenced by the high adherence rates and good to very good user satisfaction assessed at the end of the study.

Overall, patients demonstrated greater engagement with the app questionnaires compared with data collection through wearables and spirometry. Passive data collection by wearables displayed higher adherence rates than active spirometry, interestingly diverging from previous research [32]. Consistent with previous clinical studies [33-35], there was a variability particularly in wearable adherence. Our additional clinical data from the chat interactions revealed that commonly technical issues (eg, pairing problems and device malfunction) and discomfort (eg, wristband intolerance) were factors contributing to lower adherence rates for some patients. Only 1 patient consistently experienced difficulties (and ultimately discontinued wearable use). However, patients wore the device significantly longer than required by the study protocol, including during nighttime (median 20 h per day (IQR 15-24)). Similarly, spirometry was performed more frequently than required by protocol (49 measurements (median; IQR 15-59)). This indicates a high need and willingness of patients to monitor their own symptoms. Our

XSL•FO
RenderX

study results align with broader research on telemedicine for chronic diseases, demonstrating high feasibility and user acceptance, consistent with findings in other disorders. Similar to studies in palliative care in cancer patients [36], depression monitoring [37] and heart failure [38], our results confirm the feasibility of continuous monitoring through mobile technologies for chronic conditions. With an adherence rate of 80.9% (1830/2261) for wearables, our study falls at the higher end of reported ranges (53% - 94%) [36-43]. However, the higher adherence rates for active data (PROMs) over passive data (wearables) in our results contrasts with other studies finding [40,42], indicating potential disease-specific differences requiring further investigation. In addition, patients' sense of control during active data input or technical challenges with wearables could have affected adherence to passive monitoring. Still, comparison with other studies for different diseases remains difficult, as literature on wearable adherence expose inconsistency in reporting and study design.

Overall user satisfaction was high, particularly patients agreed on the ease of use, satisfaction, and improvement in understanding their symptoms. This finding aligns with a study in multiple sclerosis that reported sustained satisfaction with remote monitoring using smartphone and wearable technology over 6 months [40]. Further qualitative analyses, such as focus groups with patients, could aid in gaining a deeper understanding of the reasons behind these perceptions and to identify areas for improvements. First approaches have been made by the research team within the framework of additional small-scale pilot phases involving patients testing the platform use and workshops for feedback collection and development requests. Furthermore, mixed opinions were indicated by patients regarding whether the use of the telemedicine platform could reduce visits to neurologists or clinics, indicating that the platform may primarily serve as a complement to existing care structures rather than a replacement of personal interactions.

For successful implementation in clinical care, the efficient usage and integration of such data into clinical decision-making processes will be essential. This includes logistical efforts, but also necessitates a careful assessment of data quality and interpretation, and the "appropriate" monitoring intensity for different patient groups (eg, those with a short disease duration, highly active disease, and fast-acting treatment regimens may require more frequent assessments). Furthermore, addressing regulatory challenges, particularly regarding reimbursement structures for health care professionals will be crucial for long-term integration within health care systems.

Limitations

Our study encountered several challenges associated with remote monitoring technology: These included missing passive wearable data primarily due to technical issues and synchronization problems and patients' discomfort with the wristband from the wearable. Addressing these challenges and providing solutions, such as training sessions and available technical support, is crucial to ensure optimal patient experience and adherence in future studies using these technologies. Furthermore, the study cohort is relatively small, and the study duration is short. Patients with higher digital literacy may have been more likely to participate, potentially impacting the representativeness of the cohort. Regarding user evaluation of health care professionals, it should be noted that the study team size was small and partially comprised members of the development team for the platform, which could add to bias in answering the questionnaire.

Conclusion

In conclusion, our results demonstrate the feasibility of remote monitoring for collecting both subjective (active) and objective (passive) data from multiple sources among patients with MG, as evidenced by above-average adherence rates and acceptability. This underscores the potential for user-friendly digital tools for long-term monitoring of patients with MG and their willingness to track disease progression. Larger and prolonged studies are needed to benchmark our findings and to further explore the effective implementation of such tools and their clinical implications in practice.

Acknowledgments

We thank the Neuroscience Clinical Research Center at Charité Universitätsmedizin Berlin for their support in administration and throughout the formal processes essential for the successful execution of this study. The authors especially wish to thank Maja Olzewska, Josefina Kreusler, and Gabriela Jiron-Lozano for their support in establishing the REDCap database. Furthermore, special acknowledgments are extended to Jens Brestrich and Dike Remstedt for their support in conducting the study. In addition, the authors thank Qurasoft GmbH, namely Lale Zils, for the project assistance during the study and Sebastian Thunert, Lucas Keune, and Ragnar Hirschbrich for their technical support throughout the study. We wish to thank all the patients who participated in the study. Figure 2 was created with BioRender.com. Argenx, UCB, and Hormosan provided a grant to support the study but were not involved in the design of the study. The project was partly funded and supported by the Berlin Institute for Health/Charité Universitätsmedizin Berlin.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

MS wrote the paper. MS, SL, and LG developed the study design. AM critically reviewed the study. MS and SL conducted the study. DL helped with observation of the data. RS performed statistical analyses. MS, SL, and LG interpreted the data. RS, PN, DL, MH, AM, LG, and SL critically reviewed and edited the paper.

Conflicts of Interest

MS has received speaker's honoraria and honoraria for attendance at advisory boards from Argenx and Alexion. MH has received speaker´s honoraria from Argenx and speaker's honoraria and honoraria for attendance at advisory boards from Alexion. PN has received grant support from PCORI, Alexion, Dianthus, and Janssen. PN has received honoraria for consulting services for Alexion, Argenx, CVS Caremark, Dianthus, GSK, Immune Abs, Janssen, Novartis, and UCB. PN has served on DSMB for Sanofi. AM received speaker's honoraria from Alexion, Grifols, and Hormosan, and honoraria from Argenx, Alexion, MorphoSys, and UCB for consulting services, and financial research support from Alexion and Octapharma. He is chairperson of the medical advisory board of the DMG. SL has received speaker's honoraria for attendance at patient events and advisory boards from Argenx, Alexion, Biogen, Hormosan, Roche, and UCB. None declared by other authors.

References

1. Gilhus NE, Tzartos S, Evoli A, Palace J, Burns TM, Verschuuren J. Myasthenia gravis. Nat Rev Dis Primers 2019 May 2;5(1):1-19. [doi: 10.1038/s41572-019-0079-y]
2. Vinciguerra C, Iacono S, Bevilacqua L, et al. Sex differences in neuromuscular disorders. Mech Ageing Dev 2023 Apr;211:111793. [doi: 10.1016/j.mad.2023.111793] [Medline: 36806604]
3. Narayanaswami P, Sanders DB, Wolfe G, et al. International consensus guidance for management of myasthenia gravis: 2020 Update. Neurology (ECronicon) 2021 Jan 19;96(3):114-122. [doi: 10.1212/WNL.0000000000011124] [Medline: 33144515]
4. Lehnerer S, Jacobi J, Schilling R, et al. Burden of disease in myasthenia gravis: taking the patient's perspective. J Neurol 2022 Jun;269(6):3050-3063. [doi: 10.1007/s00415-021-10891-1] [Medline: 34800167]
5. Stein M, Hoffmann S, Gerischer L, et al. Myasthenia gravis - a retrospective analysis of e-mail inquiries made to a patient organisation and specialized center to uncover unmet needs from patients and caregivers. BMC Neurol 2022 Dec 7;22(1):455. [doi: 10.1186/s12883-022-02981-y] [Medline: 36476357]
6. Masanneck L, Räuber S, Schroeter CB, et al. Driving time-based identification of gaps in specialised care coverage: An example of neuroinflammatory diseases in Germany. Digit Health 2023;9:20552076231152989. [doi: 10.1177/20552076231152989] [Medline: 36762020]
7. Holko M, Litwin TR, Munoz F, et al. Wearable fitness tracker use in federally qualified health center patients: strategies to improve the health of all of us using digital health devices. NPJ Digit Med 2022 Apr 25;5(1):53. [doi: 10.1038/s41746-022-00593-x] [Medline: 35469045]
8. Ramezani R, Cao M, Earthperson A, Naeim A. Developing a smartwatch-based healthcare application: notes to consider. Sensors (Basel) 2023 Jul 25;23(15):6652. [doi: 10.3390/s23156652] [Medline: 37571436]
9. Dierckx R, Pellicori P, Cleland JGF, Clark AL. Telemonitoring in heart failure: big brother watching over you. Heart Fail Rev 2015 Jan;20(1):107-116. [doi: 10.1007/s10741-014-9449-4] [Medline: 24972644]
10. Koehler F, Koehler K, Deckwart O, et al. Efficacy of telemedical interventional management in patients with heart failure (TIM-HF2): a randomised, controlled, parallel-group, unmasked trial. Lancet 2018 Sep 22;392(10152):1047-1057. [doi: 10.1016/S0140-6736(18)31880-4] [Medline: 30153985]
11. Smolis-Bąk E, Dąbrowski R, Piotrowicz E, et al. Hospital-based and telemonitoring guided home-based training programs: Effects on exercise tolerance and quality of life in patients with heart failure (NYHA class III) and cardiac resynchronization therapy. A randomized, prospective observation. Int J Cardiol 2015 Nov;199:442-447. [doi: 10.1016/j.ijcard.2015.07.041]
12. Zhang Y, Pratap A, Folarin AA, et al. Long-term participant retention and engagement patterns in an app and wearable-based multinational remote digital depression study. NPJ Digit Med 2023 Feb 17;6(1):25. [doi: 10.1038/s41746-023-00749-3] [Medline: 36806317]
13. Druce KL, Dixon WG, McBeth J. Maximizing engagement in mobile health studies: lessons learned and future directions. Rheum Dis Clin North Am 2019 May;45(2):159-172. [doi: 10.1016/j.rdc.2019.01.004] [Medline: 30952390]
14. De Angel V, Lewis S, White K, et al. Digital health tools for the passive monitoring of depression: a systematic review of methods. NPJ Digit Med 2022 Jan 11;5(1):3. [doi: 10.1038/s41746-021-00548-8] [Medline: 35017634]
15. Svennberg E, Tjong F, Goette A, et al. How to use digital devices to detect and manage arrhythmias: an EHRA practical guide. Europace 2022 Jul 15;24(6):979-1005. [doi: 10.1093/europace/euac038] [Medline: 35368065]
16. Hillmann HAK, Soltani S, Mueller-Leisse J, Hohmann S, Duncker D. Cardiac rhythm monitoring using wearables for clinical guidance before and after catheter ablation. J Clin Med 2022 Apr 26;11(9):2428. [doi: 10.3390/jcm11092428] [Medline: 35566556]

XSL•FO
RenderX

17. Stein M, Meisel A, Moench M, et al. The telemedical platform myalink for remote monitoring in myasthenia gravis: rationale and protocol for a proof of concept study. MedRxiv. Preprint posted online on Jul 11, 2024. [doi: 10.1101/2024.07.10.24310229]
18. Jaretzki A 3rd, Barohn RJ, Ernstoff RM, et al. Myasthenia gravis: recommendations for clinical research standards. Task Force of the Medical Scientific Advisory Board of the Myasthenia Gravis Foundation of America. Neurology (ECronicon) 2000 Jul 12;55(1):16-23. [doi: 10.1212/wnl.55.1.16] [Medline: 10891897]
19. Randomization service for multicenter clinical trials. Randomizer. URL: https://www.randomizer.at/ [accessed 2025-02-27]
20. Barohn RJ, McIntire D, Herbelin L, Wolfe GI, Nations S, Bryan WW. Reliability testing of the quantitative myasthenia gravis score. Ann N Y Acad Sci 1998 May 13;841:769-772. [doi: 10.1111/j.1749-6632.1998.tb11015.x] [Medline: 9668327]
21. Wolfe GI, Herbelin L, Nations SP, Foster B, Bryan WW, Barohn RJ. Myasthenia gravis activities of daily living profile. Neurology (ECronicon) 1999 Apr 22;52(7):1487-1489. [doi: 10.1212/wnl.52.7.1487] [Medline: 10227640]
22. Burns TM, Sadjadi R, Utsugisawa K, et al. International clinimetric evaluation of the MG-QOL15, resulting in slight revision and subsequent validation of the MG-QOL15r. Muscle Nerve 2016 Dec;54(6):1015-1022. [doi: 10.1002/mus.25198] [Medline: 27220659]
23. Abraham A, Breiner A, Barnett C, Katzberg HD, Bril V. The utility of a single simple question in the evaluation of patients with myasthenia gravis. Muscle Nerve 2018 Feb;57(2):240-244. [doi: 10.1002/mus.25720] [Medline: 28590545]
24. Kukulka K, Gummi RR, Govindarajan R. A telephonic single breath count test for screening of exacerbations of myasthenia gravis: A pilot study. Muscle Nerve 2020 Aug;62(2):258-261. [doi: 10.1002/mus.26987] [Medline: 32447763]
25. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary care evaluation of mental disorders. Patient health questionnaire. JAMA 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
26. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
27. Mendoza M, Tran C, Bril V, Katzberg HD, Barnett C. Patient-acceptable symptom states in myasthenia gravis. Neurology (ECronicon) 2020 Sep 22;95(12):e1617-e1628. [doi: 10.1212/WNL.0000000000010574] [Medline: 32759200]
28. Brooke J. SUS: A quick and dirty usability scale. In: Usability Evaluation in Industry: Taylor & Francis; 1995:189-194.
29. Braun M, Schmidt O, Schultz T, Woehrle H, Sundrup MG, Schöbel C. Erfahrungen mit der digitalen Versorgung von Patienten mit chronischen und akuten Lungenerkrankungen während der SARS-CoV-2-Pandemie [Article in German]. Internist 2022 Mar;63(3):255-265. [doi: 10.1007/s00108-022-01266-3]
30. R. The R Project for Statistical Computing. URL: https://www.r-project.org/ [accessed 2020-09-11]
31. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
32. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile sensing and support for people With depression: a pilot trial in the wild. JMIR Mhealth Uhealth 2016 Sep 21;4(3):e111. [doi: 10.2196/mhealth.5960] [Medline: 27655245]
33. Jacoby SF, Robinson AJ, Webster JL, Morrison CN, Richmond TS. The feasibility and acceptability of mobile health monitoring for real-time assessment of traumatic injury outcomes. Mhealth 2021;7:5. [doi: 10.21037/mhealth-19-200] [Medline: 33634188]
34. Lee S, Kim H, Park MJ, Jeon HJ. Current advances in wearable devices and their sensors in patients with depression. Front Psychiatry 2021;12:672347. [doi: 10.3389/fpsyt.2021.672347] [Medline: 34220580]
35. Chan YFY, Wang P, Rogers L, et al. The Asthma Mobile Health Study, a large-scale clinical observational study using ResearchKit. Nat Biotechnol 2017 Apr;35(4):354-362. [doi: 10.1038/nbt.3826] [Medline: 28288104]
36. Pavic M, Klaas V, Theile G, et al. Mobile health technologies for continuous monitoring of cancer patients in palliative care aiming to predict health status deterioration: a feasibility study. J Palliat Med 2020 May;23(5):678-685. [doi: 10.1089/jpm.2019.0342] [Medline: 31873052]
37. Pedrelli P, Fedor S, Ghandeharioun A, et al. Monitoring changes in depression severity using wearable and mobiles ensors. Front Psychiatry 2020;11:584711. [doi: 10.3389/fpsyt.2020.584711] [Medline: 33391050]
38. Werhahn SM, Dathe H, Rottmann T, et al. Designing meaningful outcome parameters using mobile technology: a new mobile application for telemonitoring of patients with heart failure. ESC Heart Fail 2019 Jun;6(3):516-525. [doi: 10.1002/ehf2.12425] [Medline: 30868756]
39. Wu JM, Ho TW, Chang YT, et al. Wearable-based mobile health app in gastric cancer patients for postoperative physical activity monitoring: focus group study. JMIR Mhealth Uhealth 2019 Apr 23;7(4):e11989. [doi: 10.2196/11989] [Medline: 31012858]
40. Midaglia L, Mulero P, Montalban X, et al. Adherence and satisfaction of smartphone- and smartwatch-based remote active testing and passive monitoring in people with multiple sclerosis: nonrandomized interventional feasibility study. J Med Internet Res 2019 Aug 30;21(8):e14863. [doi: 10.2196/14863] [Medline: 31471961]
41. Matcham F, Leightley D, Siddi S, et al. Remote Assessment of Disease and Relapse in Major Depressive Disorder (RADAR-MDD): recruitment, retention, and data availability in a longitudinal remote measurement study. BMC Psychiatry 2022 Feb 21;22(1):136. [doi: 10.1186/s12888-022-03753-1] [Medline: 35189842]

XSL•FO
RenderX

42. Sohn A, Speier W, Lan E, et al. Integrating remote monitoring into heart failure patients' care regimen: A pilot study. PLoS One 2020;15(11):e0242210. [doi: 10.1371/journal.pone.0242210] [Medline: 33211733]
43. Naslund JA, Aschbrenner KA, Bartels SJ. Wearable devices and smartphones for activity tracking among people with serious mental illness. Ment Health Phys Act 2016 Mar;10:10-17. [doi: 10.1016/j.mhpa.2016.02.001] [Medline: 27134654]

Abbreviations

BlnDSG: Berlin Data Protection Act
CFS: Chalder Fatigue Scale
GPDR: General Data Protection Regulation
HADS: Hospital Anxiety and Depression Scale
MG: myasthenia gravis
MG-ADL: myasthenia gravis activities of daily living
MG-QoL15r: myasthenia gravis quality of life, revised
MGFA: Myasthenia Gravis Foundation of America
PASS: Patient Acceptable Symptom State
PROMs: patient-reported outcome measures
QMGS: quantitative myasthenia gravis score
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
SSQ: single simple question
SUS: System Usability Scale

Edited by A Mavragani; submitted 03.04.24; peer-reviewed by C Vinciguerra, L Masanneck; revised version received 03.12.24; accepted 27.12.24; published 03.03.25.

Please cite as:
Stein M, Stegherr R, Narayanaswami P, Legg D, Herdick M, Meisel A, Gerischer L, Lehnerer S
App- and Wearable-Based Remote Monitoring for Patients With Myasthenia Gravis and Its Specialists: Feasibility and Usability Study
JMIR Form Res 2025;9:e58266
URL: https://formative.jmir.org/2025/1/e58266
doi:10.2196/58266

© Maike Stein, Regina Stegherr, Pushpa Narayanaswami, David Legg, Meret Herdick, Andreas Meisel, Lea Gerischer, Sophie Lehnerer. Originally published in JMIR Formative Research (https://formative.jmir.org), 3.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Enhancing Distress Tolerance Skills in Adolescents With Anorexia Nervosa Through the BALANCE Mobile App: Feasibility and Acceptability Study

Christina Miranda[1], BA; Brittany Matheson[2], PhD; Nandini Datta[2], PhD; Aileen Whyte[2], PhD; Hyun-Joon Yang[2], BS; Paul Schmiedmayer[3], PhD; Vishnu Ravi[3], MD; Oliver Aalami[3], MD; James Lock[2], MD, PhD

[1]School of Medicine, Stanford University, 291 Campus Dr, Stanford, CA, United States

[2]Department of Psychiatry and Behavioral Sciences, Stanford Medicine, Stanford, CA, United States

[3]Stanford Mussallem Center for Biodesign, Stanford University, Stanford, CA, United States

Corresponding Author:
Christina Miranda, BA
School of Medicine, Stanford University, 291 Campus Dr, Stanford, CA, United States

Abstract

Background: Anorexia nervosa is a severe psychiatric disorder with high morbidity and mortality, particularly among adolescents. Family-based treatment (FBT) is the leading evidence-based intervention for adolescent anorexia nervosa, involving parents in renourishment and behavior interruption. Despite its effectiveness, challenges in distress tolerance and emotion regulation during high-stress situations, such as mealtimes, contribute to suboptimal treatment outcomes, with only 35% to 50% of adolescents achieving full recovery. Enhancing distress tolerance skills during FBT may improve treatment responses and recovery rates. The BALANCE mobile app was developed to address this need, offering real-time, dialectical behavior therapy (DBT)–based distress tolerance skills to support adolescents and families during mealtimes.

Objective: Our aim was to explore the feasibility and acceptability of a mobile app designed to deliver distress tolerance skills to adolescents with and adolescents without anorexia nervosa. When fully programmed and optimized, we plan to use the mobile app to improve distress tolerance during mealtimes for adolescents with anorexia nervosa undergoing FBT.

Methods: BALANCE was developed collaboratively with Stanford University's Center for Biodesign, leveraging the expertise of clinical psychologists and using biodesign student input and the Stanford Spezi ecosystem. The app underwent an iterative development process, with feedback from adolescent users. The initial feasibility and acceptability of the app were assessed through self-reported questionnaires and structured interviews with 24 adolescents aged 12 to 18 years, including 4 diagnosed with anorexia nervosa and 20 healthy controls. Adolescents with anorexia nervosa specifically used the app during mealtimes, and healthy controls used it as needed. Participants assessed the app's usability, perceived effectiveness, and its impact on their distress tolerance.

Results: The app demonstrated high usability and acceptability. Of 24 participants, 83% (n=20) reported enjoying the app, 88% (n=21) would recommend it to peers, and 100% (n=24) found it user-friendly. Adolescents with anorexia nervosa reported that BALANCE helped them manage stressful mealtimes more effectively, highlighting features such as guided meditation, breathing exercises, and gamification elements as particularly effective. Healthy controls provided additional feedback, confirming the app's broad appeal to the target audience and potential scalability. Preliminary findings suggest that BALANCE may enhance distress tolerance in adolescents with and adolescents without anorexia nervosa.

Conclusions: BALANCE shows promise as an innovative mobile health intervention for enhancing distress tolerance in adolescents with anorexia nervosa. Its user-friendly design and tailored DBT-based skills make it a feasible tool for integration into FBT. Future research should explore its integration into clinical practice and its impact on treatment outcomes. As distress tolerance skills are relevant to a range of mental health conditions, future research may also expand BALANCE's application to broader adolescent populations.

(JMIR Form Res 2025;9:e70278) doi:10.2196/70278

KEYWORDS

mHealth; mobile health; mobile application; emotion regulation; eating disorders; family-based treatment; distress tolerance; mealtimes

XSL•FO
RenderX

Introduction

Anorexia nervosa is a severe psychiatric disorder that poses significant morbidity and mortality risks, particularly among adolescents. Family-based treatment (FBT) has been shown to be effective in addressing adolescent anorexia nervosa by involving parents to promote renourishment and disrupt eating disorder behaviors [1]. However, despite its success, only 35% to 50% of adolescents fully recover through FBT, highlighting the need for innovative strategies to enhance treatment outcomes [2,3].

Recent research highlights the role of distress tolerance in eating disorder recovery, with studies demonstrating that lower distress tolerance is associated with poorer treatment outcomes [4]. Despite the increasing adoption of mobile health (mHealth) interventions in psychiatric care, few apps specifically target distress tolerance during mealtimes for adolescents with anorexia nervosa. The BALANCE app was developed to address this gap, providing real-time, dialectical behavior therapy (DBT)–based distress tolerance strategies that can be integrated into FBT.

Several studies in the past 2 to 3 years have examined digital interventions for eating disorders, reinforcing the importance of personalized and interactive tools in improving treatment adherence [5-7]. Digital interventions have been found to enhance engagement by providing real-time coping strategies, facilitating adherence to treatment protocols, and supporting long-term behavior change [7]. Additionally, studies emphasize the importance of incorporating structured distress tolerance strategies into existing treatment models to optimize clinical outcomes [4]. By leveraging DBT principles, BALANCE aims to enhance distress tolerance skills, thereby ultimately improving treatment efficacy for adolescents with anorexia nervosa.

This study evaluated the feasibility and acceptability of the BALANCE mobile app, which was designed to improve distress tolerance in adolescents, with the aim of ultimately having the app be used during mealtimes for adolescents with anorexia nervosa undergoing FBT.

Methods

Recruitment Process and Procedures

To ensure a systematic and transparent recruitment process, we implemented the following step-by-step procedure: (1) identifying eligible participants, (2) screening for eligibility, (3) obtaining parental and participant consent, and (4) conducting an orientation session.

Identification of Eligible Participants

A total of 24 adolescents aged 12 to 18 years were recruited for this study. For the initial iteration of the app, we focused on acceptability and feasibility among the adolescent age group. In total, 20 healthy adolescent controls were recruited from the community through flyers on social media platforms. We also recruited 4 adolescents with anorexia nervosa from the Stanford University Comprehensive Care Program—a 15-bed inpatient unit within Lucile Packard Children's Hospital that provides multidisciplinary care for medically compromised adolescents with eating disorders—to preliminarily assess acceptability and feasibility with this clinical group.

Screening for Eligibility

Interested participants completed an initial phone call with study staff to determine eligibility based on diagnosis, age, and access to a smart device that could download the BALANCE app.

Parental and Participant Consent

For participants younger than 18 years, parental consent and adolescent assent were obtained. Participants aged 18 years provided informed consent independently. All participants received detailed study explanations, including explanations about potential risks and benefits.

Orientation Session

Eligible participants attended an orientation session where they learned how to use the app, including its features and intended usage during mealtimes.

Ethical Considerations

This study was approved by Stanford University's institutional review board, under the protocol titled "Refinement of Emotion Regulation App for Adolescents with Anorexia Nervosa at Mealtimes During Family-Based Treatment" (eProtocol #69550), adhering to ethical guidelines for human subject research. Informed consent was obtained from all participants aged 18 years, and for participants younger than 18 years, consent was secured from their parents or legal guardians alongside adolescent assent. To ensure privacy and confidentiality, all data were deidentified prior to analysis and securely stored on encrypted servers that were accessible only to the research team. Findings were reported in aggregate to prevent individual identification. Participants were not monetarily compensated but were provided early access to the BALANCE app and allowed continued use after this study's completion.

App Development and Collaboration

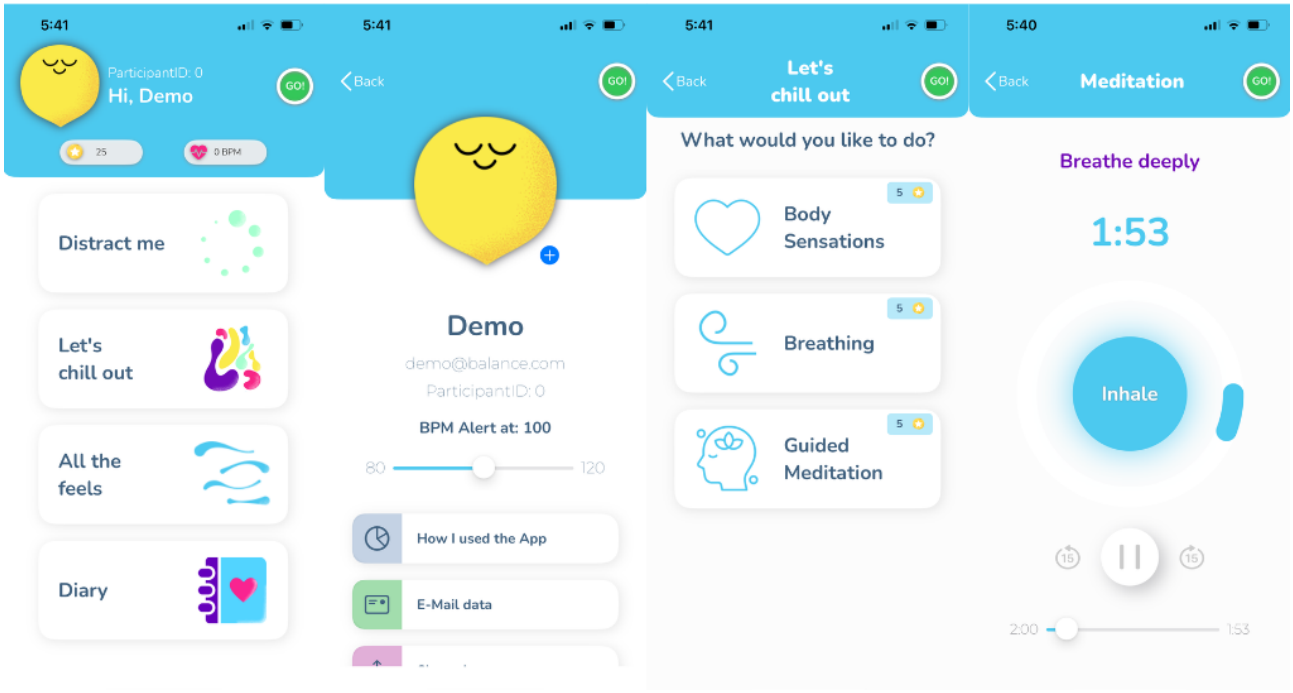
BALANCE was developed in collaboration with Stanford University's Eating Disorder Research Team and a Stanford biodesign course on mHealth app development [8]. The course enabled computer science undergraduate and graduate students to provide input on app features, interface design, and technical functionality, leveraging the Stanford Spezi open-source digital health framework [9] and student expertise in mHealth technology. Spezi was designed for the rapid development of modern, interoperable mHealth apps based on an ecosystem of modules that encapsulate common functionality and build upon international standards for health data. Prior research on app-based interventions for eating disorders demonstrated the importance of incorporating lived experiences of potential users into the development process [10]. Similarly, the BALANCE app development followed an iterative, user-centered design process, which incorporated 2 waves of feedback from adolescent users (N=10), ensuring the app's relevance and usability for the target adolescent age group. Licensed clinical psychologists with expertise in DBT-based distress tolerance skills guided the creation of the app's core content, while

XSL•FO
RenderX

students helped refine the user experience through feedback loops and usability testing [11].

BALANCE incorporates several features designed to improve distress tolerance in adolescents. Users can select a distress tolerance skill, such as distraction, relaxation, or emotion tracking, through the app's interface (Figure 1). Data on app use (eg, frequency, duration, and skill type) are collected. The app also includes a gamification feature, rewarding users with coins (ie, for customizing an avatar) based on their engagement. The app can be opened directly when prompted by the adolescents or their parents. Additionally, when the app is integrated with a user's personal smartwatch, notifications are sent to the user when their heart rate increases enough to indicate the possibility of emotional distress, prompting them to open the app. Therapists working with the adolescents in FBT have access to a dashboard that provides information about adolescents' app use in visual tables and figures. Therapists can integrate this information (eg, how often the app was used, what time of day app use occurred, and what app features were accessed) into FBT when working with the patients and their family.

Figure 1. BALANCE app interface and features. This figure showcases the design and core functionalities of the BALANCE app—a mobile health intervention developed to deliver distress tolerance skills to adolescents aged 12 to 18 years. The app incorporates guided meditation, emotion tracking, and gamification features, and it was developed through a collaborative process involving clinical psychologists and biodesign students. Screenshots depict the user interface and interactive elements.



Study Design

This study assessed the BALANCE app's feasibility and acceptability, using self-reported questionnaires and interviews with adolescent users. A secondary aim was to explore app use among adolescents with anorexia nervosa. Healthy adolescents used the app for at least 30 minutes, while adolescents with anorexia nervosa used the app during mealtimes. Measures included usability, perceived effectiveness, and distress tolerance outcomes. Structured interviews allowed participants to elaborate on survey responses to suggest specific modifications to the app, comment on the design, alert us about any usability issues, and share any other feedback that was not captured by survey questions.

Measures

Participants rated their experience with the app by using Likert-scale surveys and open-ended questions about the app's functionality, design, and impact on emotion regulation.

Results

Participant Demographics

Participants consisted of healthy adolescents (n=20) and adolescents with anorexia nervosa (n=4). Participants were racially and ethnically diverse; of the 24 participants, 16 (67%) identified as female, 8 (33%) identified as male, 13 (54%) identified as White, 8 (33%) identified as Asian, and 3 (13%) identified as Hispanic or Latino.

Survey Results

The majority of participants found the app user-friendly (n=24, 100%) and helpful (n=22, 92%). Of the 24 participants, 20 (83%) reported enjoying the app, and 21 (88%) would recommend it to a friend. Notably, the guided meditation and breathing exercises were highlighted as particularly helpful during stressful situations. Participant ratings of the app were generally consistent between adolescents with and adolescents without anorexia nervosa (Table 1). Preliminary findings suggest that the BALANCE app improves adolescents' self-reported ability to manage emotional states. Adolescents with anorexia

XSL•FO
RenderX

nervosa reported a decrease in distress and increased ability to tolerate discomfort associated with eating (Tables 1-3).

Table . Survey responses from the 24 adolescents (4 with anorexia nervosa and 20 healthy controls) on the BALANCE app. Responses (collected from July 2023 to January 2024) highlight user satisfaction, ease of use, and distress tolerance skills.[a]

Survey statement	Strongly disagree, n (%)	Disagree, n (%)	Neutral, n (%)	Agree, n (%)	Strongly agree, n (%)
I enjoyed using this application	0 (0)	0 (0)	0 (0)	20 (83)	4 (17)
This app has functionality that is easy to understand	0 (0)	0 (0)	2 (8)	17 (71)	5 (21)
This app has a user-friendly interface design	0 (0)	0 (0)	0 (0)	14 (58)	10 (42)
This app is exhausting or difficult to use	11 (46)	10 (42)	2 (8)	1 (4)	0 (0)
I would recommend this app to a friend	0 (0)	0 (0)	3 (13)	21 (88)	0 (0)
This app helped me to regulate my emotions	0 (0)	1 (4)	8 (33)	15 (63)	0 (0)
I found this app to be triggering or unhelpful to my emotional state	16 (67)	6 (25)	0 (0)	2 (8)	0 (0)

[a]Percentages may not add up to 100% due to rounding.

Table . Survey responses from the 20 healthy adolescent controls on the BALANCE app. Responses (collected from July 2023 to January 2024) highlight user satisfaction, ease of use, and distress tolerance skills.[a]

Survey statement	Strongly disagree, n (%)	Disagree, n (%)	Neutral, n (%)	Agree, n (%)	Strongly agree, n (%)
I enjoyed using this application	0 (0)	0 (0)	0 (0)	16 (80)	4 (20)
This app has functionality that is easy to understand	0 (0)	0 (0)	1 (5)	14 (70)	5 (25)
This app has a user-friendly interface design	0 (0)	0 (0)	0 (0)	12 (60)	8 (40)
This app is exhausting or difficult to use	9 (45)	10 (50)	0 (0)	1 (5)	0 (0)
I would recommend this app to a friend	0 (0)	0 (0)	3 (15)	17 (85)	0 (0)
This app helped me to regulate my emotions	0 (0)	1 (5)	6 (30)	13 (65)	0 (0)
I found this app to be triggering or unhelpful to my emotional state	13 (65)	5 (25)	0 (0)	2 (10)	0 (0)

[a]Percentages may not add up to 100% due to rounding.

XSL•FO
RenderX

Table . Survey responses from the 4 adolescents with anorexia nervosa on the BALANCE app. Responses (collected from July 2023 to January 2024) highlight user satisfaction, ease of use, and distress tolerance skills.[a]

Survey statement	Strongly disagree, n (%)	Disagree, n (%)	Neutral, n (%)	Agree, n (%)	Strongly agree, n (%)
I enjoyed using this application	0 (0)	0 (0)	0 (0)	4 (100)	0 (0)
This app has functionality that is easy to understand	0 (0)	0 (0)	1 (25)	3 (75)	0 (0)
This app has a user-friendly interface design	0 (0)	0 (0)	0 (0)	2 (50)	2 (50)
This app is exhausting or difficult to use	2 (50)	0 (0)	2 (50)	0 (0)	0 (0)
I would recommend this app to a friend	0 (0)	0 (0)	0 (0)	4 (100)	0 (0)
This app helped me to regulate my emotions	0 (0)	0 (0)	2 (50)	2 (50)	0 (0)
I found this app to be triggering or unhelpful to my emotional state	3 (75)	1 (25)	0 (0)	0 (0)	0 (0)

[a]Percentages may not add up to 100% due to rounding.

Discussion

Summary of Findings

This study demonstrated the feasibility and acceptability of the BALANCE mobile app for delivering distress tolerance skills to adolescents, including those with anorexia nervosa. The app was well received, with 83% (20/24) of participants reporting enjoyment, 88% (21/24) reporting that they would recommend it to peers, and 100% (24/24) finding it user-friendly. Adolescents with anorexia nervosa reported improvements in their ability to tolerate mealtime distress, suggesting that the app may serve as a valuable adjunct to FBT. These findings align with this study's primary objective to assess the feasibility and acceptability of the app, paving the way for its integration into clinical practice. Additionally, this study provides initial evidence that BALANCE may help address distress-related barriers to effective meal completion—a crucial factor in anorexia nervosa recovery.

Interpretations

The BALANCE app provides a structured and scalable solution to a critical gap in FBT by delivering real-time distress tolerance skills for high-stress mealtimes [12,13]. Distress tolerance deficits contribute to poor treatment adherence and increased dropout rates among adolescents undergoing FBT [12,13]. The high usability ratings suggest that BALANCE may effectively supplement traditional therapeutic interventions, thereby improving mealtime coping mechanisms.

Compared to other mHealth apps for eating disorders, BALANCE is unique in its focus on real-time distress tolerance rather than self-monitoring or general emotional regulation [5,7]. Its DBT-based approach aligns with research showing the efficacy of distress tolerance training in improving treatment engagement and emotional resilience in adolescents with anorexia nervosa [14,15]. Gamification and interactive features further support engagement—a crucial factor in digital intervention success [5,6].

As digital tools become more integrated into clinical practice, BALANCE could be incorporated into standard FBT protocols to improve accessibility and engagement. By providing just-in-time intervention during mealtimes, it may help adolescents manage high-anxiety eating situations and support better treatment adherence and long-term recovery.

Limitations

This study's limitations include a small sample size, particularly the limited number of adolescents with anorexia nervosa. Additionally, the reliance on self-reported data introduces potential biases, and the short duration of app use prevents long-term effectiveness assessment. Future research should address these limitations by conducting larger-scale trials, exploring longitudinal impacts, and assessing BALANCE's integration into standard FBT protocols. Investigating potential barriers to app engagement in diverse clinical settings is also necessary for optimizing its implementation.

Conclusions and Broader Implications

The BALANCE app is a novel, acceptable, and feasible tool for delivering distress tolerance skills to adolescents with anorexia nervosa. It shows promise for use during high-stress mealtimes, addressing a critical barrier to the effectiveness of FBT. The pilot testing of the app's integration into FBT is underway, assessing its impact on treatment adherence and distress tolerance in real-world clinical settings. If successfully implemented, BALANCE could enhance treatment outcomes, reduce barriers to recovery, and improve FBT effectiveness. Future research should explore its broader applicability to other mental health conditions wherein distress tolerance plays a key

XSL•FO
RenderX

role, to ensure its scalability and long-term benefits for adolescent mental health interventions.

Acknowledgments

This project was funded by the Packard Foundation and Stanford University's undergraduate health application development course participants. The authors would like to thank the students and faculty from the Stanford Mussallem Center for Biodesign for their collaboration in developing the BALANCE app [8,9].

Data Availability

The datasets generated or analyzed during this study are not publicly available due to participant privacy and confidentiality considerations but are available from the corresponding author on reasonable request.

Conflicts of Interest

JL receives royalties from Guilford Press and Routledge. He is also the codirector of the Training Institute for Child and Adolescent Eating Disorders, LLC. ND receives royalties from Routledge. BM receives royalties from Routledge.

References

1. Lock J, Le Grange D, Agras WS, Moye A, Bryson SW, Jo B. Randomized clinical trial comparing family-based treatment with adolescent-focused individual therapy for adolescents with anorexia nervosa. Arch Gen Psychiatry 2010 Oct;67(10):1025-1032. [doi: 10.1001/archgenpsychiatry.2010.128] [Medline: 20921118]
2. Le Grange D, Lock J, Agras WS, Bryson SW, Jo B. Randomized clinical trial of family-based treatment and cognitive-behavioral therapy for adolescent bulimia nervosa. J Am Acad Child Adolesc Psychiatry 2015 Nov;54(11):886-894.e2. [doi: 10.1016/j.jaac.2015.08.008] [Medline: 26506579]
3. Lock JD, Le Grange D, Bohon C, Matheson B, Jo B. Who responds to an adaptive intervention for adolescents with anorexia nervosa being treated with family-based treatment? Outcomes from a randomized clinical trial. J Am Acad Child Adolesc Psychiatry 2024 Jun;63(6):605-614. [doi: 10.1016/j.jaac.2023.10.012] [Medline: 38142046]
4. Datta N, Matheson BE, Citron K, Van Wye EM, Lock JD. Evidence based update on psychosocial treatments for eating disorders in children and adolescents. J Clin Child Adolesc Psychol 2023;52(2):159-170. [doi: 10.1080/15374416.2022.2109650] [Medline: 35950931]
5. Wilks CR, Gurtovenko K, Rebmann K, Williamson J, Lovell J, Wasil AR. A systematic review of dialectical behavior therapy mobile apps for content and usability. Borderline Personal Disord Emot Dysregul 2021 Dec 3;8(1):29. [doi: 10.1186/s40479-021-00167-5] [Medline: 34857035]
6. Guala MM, Bul K, Skårderud F, Søgaard Nielsen A. A serious game for patients with eating disorders (Maze Out): pilot user experience and acceptance study. JMIR Form Res 2023 Jan 27;7:e40594. [doi: 10.2196/40594] [Medline: 36705956]
7. Lindgreen P, Lomborg K, Clausen L. Patient experiences using a self-monitoring app in eating disorder treatment: qualitative study. JMIR Mhealth Uhealth 2018 Jun 22;6(6):e10253. [doi: 10.2196/10253] [Medline: 29934285]
8. Stanford University. CS342 / MED253 - building for digital health: winter quarter – academic year 2023/2024. Notion. 2023. URL: https://stanfordbdhg.notion.site/CS342-MED253-Building-for-Digital-Health-115008f9653881e3b42ddf7dfbf5a156 [accessed 2025-02-18]
9. Schmiedmayer P, Bauer A, Zagar P, Ravi V, Aalami O. Stanford Spezi. Zenodo. 2024 Oct 29. URL: https://zenodo.org/records/14006855 [accessed 2025-02-18]
10. Jarman HK, McLean SA, Rodgers R, et al. Informing mHealth and web-based eating disorder interventions: combining lived experience perspectives with design thinking approaches. JMIR Form Res 2022 Oct 31;6(10):e38387. [doi: 10.2196/38387] [Medline: 36315225]
11. Safer DL, Robinson AH, Jo B. Outcome from a randomized controlled trial of group therapy for binge eating disorder: comparing dialectical behavior therapy adapted for binge eating to an active comparison group therapy. Behav Ther 2010 Mar;41(1):106-120. [doi: 10.1016/j.beth.2009.01.006] [Medline: 20171332]
12. Rowsell M, MacDonald DE, Carter JC. Emotion regulation difficulties in anorexia nervosa: associations with improvements in eating psychopathology. J Eat Disord 2016 May 18;4:17. [doi: 10.1186/s40337-016-0108-0] [Medline: 27195121]
13. Harrison A, Sullivan S, Tchanturia K, Treasure J. Emotional functioning in eating disorders: attentional bias, emotion recognition and emotion regulation. Psychol Med 2010 Nov;40(11):1887-1897. [doi: 10.1017/S0033291710000036] [Medline: 20102669]
14. Vogel EN, Singh S, Accurso EC. A systematic review of cognitive behavior therapy and dialectical behavior therapy for adolescent eating disorders. J Eat Disord 2021 Oct 18;9(1):131. [doi: 10.1186/s40337-021-00461-1] [Medline: 34663452]
15. Rozakou-Soumalia N, Dârvariu Ş, Sjögren JM. Dialectical behaviour therapy improves emotion dysregulation mainly in binge eating disorder and bulimia nervosa: a systematic review and meta-analysis. J Pers Med 2021 Sep 18;11(9):931. [doi: 10.3390/jpm11090931] [Medline: 34575707]

XSL•FO
RenderX

Abbreviations

DBT: dialectical behavior therapy
FBT: family-based treatment
mHealth: mobile health

Edited by A Mavragani; submitted 18.12.24; peer-reviewed by HJ Watson, NJ Goel; revised version received 03.02.25; accepted 04.02.25; published 28.02.25.

Please cite as:
Miranda C, Matheson B, Datta N, Whyte A, Yang HJ, Schmiedmayer P, Ravi V, Aalami O, Lock J
Enhancing Distress Tolerance Skills in Adolescents With Anorexia Nervosa Through the BALANCE Mobile App: Feasibility and Acceptability Study
JMIR Form Res 2025;9:e70278
URL: https://formative.jmir.org/2025/1/e70278
doi:10.2196/70278

© Christina Miranda, Brittany Matheson, Nandini Datta, Aileen Whyte, Hyun-Joon Yang, Paul Schmiedmayer, Vishnu Ravi, Oliver Aalami, James Lock. Originally published in JMIR Formative Research (https://formative.jmir.org), 28.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Probing Public Perceptions of Antidepressants on Social Media: Mixed Methods Study

Jianfeng Zhu[1], PhD; Xinyu Zhang[2], MA; Ruoming Jin[1], Prof Dr; Hailong Jiang[1], PhD; Deric R Kenne[3], PhD

[1]Department of Computer Science, Kent State University, 800 E Summit St, Kent, OH, United States

[2]Department of Communication, Michigan State University, East Lansing, MI, United States

[3]Center for Public Policy and Health, College of Public Health, Kent State University, Kent, OH, United States

Corresponding Author:
Jianfeng Zhu, PhD
Department of Computer Science, Kent State University, 800 E Summit St, Kent, OH, United States

Abstract

Background: Antidepressants are crucial for managing major depressive disorders; however, nonadherence remains a widespread challenge, driven by concerns over side effects, fear of dependency, and doubts about efficacy. Understanding patients' experiences is essential for improving patient-centered care and enhancing adherence**,** which prioritizes individual needs in treatment.

Objective: This study aims to gain a deeper understanding of patient experiences with antidepressants, providing insights that health care providers, families, and communities can develop into personalized treatment strategies. By integrating patient-centered care, these processes may improve satisfaction and adherence with antidepressants.

Methods: Data were collected from AskaPatient and Reddit, analyzed using natural language processing and large language models. Analytical techniques included sentiment analysis, emotion detection, personality profiling, and topic modeling. Furthermore, demographic variations in patient experiences were also examined to offer a comprehensive understanding of discussions around antidepressants.

Results: Sentiment and emotion analysis revealed that the majority of discussions (21,499/36,253, 59.3%) expressed neutral sentiments, with negative sentiments following closely (13,922/36,253, 38.4%). The most common emotions were fear (16,196/36,253, 44.66%) and sadness (12,507/36,253, 34.49%). The largest topic, "Mental Health and Relationships," accounted for 11.69% (3755/36,253) of the discussions, which indicated a significant focus on managing mental health conditions. Discussions around nonadherence were marked by fear, followed by sadness, while self-care discussions showed a notable trend of sadness.

Conclusions: These psychological insights into public perceptions of antidepressants provide a foundation for developing tailored, patient-centered treatment approaches that align with individual needs, enhancing both effectiveness and empathy of care.

(JMIR Form Res 2025;9:e62680) doi:10.2196/62680

KEYWORDS

antidepressant; AskaPatient; natural language processing; BERTopic; large language models; Reddit

Introduction

Background

Antidepressants are a cornerstone in the treatment of major depressive disorder worldwide, improving the quality of life for millions of individuals. However, nonadherence to antidepressant therapy remains a significant barrier to achieving successful outcomes. According to the National Center for Health Statistics, 13.2% of adults in the United States used antidepressants within the past 30 days, with higher rates among women (17.7%) than among men (8.4%) between 2015 and 2018 [1]. This disparity highlights potential differences in how men and women experience depression and respond to antidepressant therapy, which may influence their adherence behaviors [2].

The factors driving nonadherence are multifaceted, encompassing both patient- and clinician-related challenges. Patients may avoid medicines due to concerns about side effects, fear of dependency, or doubts about their effectiveness [3]. Furthermore, depression's symptoms, such as low motivation and emotion dysregulated, often prevent patients from filling prescriptions or attending follow-up appointments [2]. Clinician-related factors, such as insufficient patient education and inadequate follow-up, also contribute to poor adherence. These challenges underscore the need for a more patient-centered approach in managing depression.

Patient-centered care (PCC) is designed to optimize treatment adherence by ensuring that clinical interventions are closely aligned with each patient's preferences, needs, and values. This approach enhances communication between caregivers and

XSL•FO
RenderX

patients, supports collaborative decision-making, and increase overall patient satisfaction, all of which are essential to improving adherence rates [4,5]. However, long-term treatment with selective serotonin reuptake inhibitors raises concerns about emotion blunting, a condition in which patients report feeling emotionally numb, impacting both their positive and negative emotions, which potentially alter personality [6-8]. This highlights the need to examine the emotional and psychological dimensions of antidepressant use in real-world setting.

Traditional methods such as surveys and interviews are limited by recall bias and cannot capture the day-to-day emotional fluctuations that significantly influence medication adherence [9,10]. Real-time insights are crucial to understanding how emotions, which change rapidly in response treatment, impact medication adherence. Given that medication adherence is closely linked to patient emotions during treatment, consequently there is a need for innovative data collection that can offer information on patients' real-time experience.

Social media offers a unique opportunity to capture real-time, unfiltered opinions from users [11-13]. Unlike structured clinical trials or surveys, social media discussions reflect spontaneous, in the moment experiences, offering deeper insights into how patients perceive their medications and mental health [14-16]. Furthermore, social media provides large-scale, real-time data on patient experiences, which can be analyzed to better understand the emotional and psychological factors influencing antidepressant adherence.

Advancements in artificial intelligence, particularly with natural language processing (NLP) and large language models (LLMs), such as GPT-4, allow for more nuanced analysis of patient language [17-19]. These tools provide deeper insights by not only detecting sentimental analysis but also profiling personality traits, allowing researchers to explore psychological dimensions of antidepressant users. Through NLP and LLMs, it is possible to assess emotional states, sentiment, and personality traits from large volumes of patient-generated content.

This study aims to bridge the gap between clinical assessment and patient-reported experiences by using real-time social media data. This advanced approach offers a more comprehensive understanding of patient experiences, capturing emotional and psychological factors often missed by traditional methods. The findings support decision-making strategies for caregivers and enhance PCC strategies leading to increase treatment adherence.

Literature Review

Nonadherence to Antidepressants

Research consistently highlights significant variability in antidepressant adherence across different populations. A review of studies over the past decade reveals that approximately 50% of patients discontinue antidepressant therapy prematurely, which underscore the pervasive nature of nonadherence [3]. Pompili et al [20] identified several key predictors of nonadherence in patients with mood disorders, including those of younger age (younger than 40 years), comorbid substance use or personality disorders, negative beliefs, poor insight, treatment-related side effects, and a weak therapeutic alliance. The study on 332 patients by Kamaradova et al [21] found a negative correlation between self-stigma and adherence, with greater illness severity associated with lower adherence across all diagnostic groups. Similarly, De las Cuevas et al [22] identified that adherence was closely linked to positive attitudes toward treatment, while nonadherence was associated with severe depression and concerns about side effects. These findings underscore the multifactorial nature of antidepressant nonadherence, influenced by patient beliefs, illness severity, and the quality of patient-provider relationships [23].

Patient-Centered Care

PCC, which prioritizes the patient's preferences and values, plays a crucial role in addressing nonadherence. It improves patient satisfaction and outcomes by considering the psychological challenges, such as fear and uncertainty, faced by patients and their families [24]. Studies show that shared decision-making and effective communication between health care providers and patients foster trust, enabling patients to express their emotional concerns, ultimately strengthening the therapeutic alliance, and improving adherence [25,26].

Integration sentiment analysis into PCC allows health care providers to gain real-time insights into patient emotions and concerns, enabling timely interventions that enhance adherence and promote self-care strategies [27]. These tools, combined with personality profiling, offer a more nuanced understanding of patient experiences than traditional methods [28], allowing for personalized care that improves mental health outcomes.

Use of Social Media for Health Research

Social media provides real-time, unfiltered insights into patient experiences with medication and mental health challenges. Platforms such as X (Twitter) and Reddit are rich sources of data that help researchers understand patient interactions with their mental health conditions and the health care system. Social media enables real-time health monitoring and predictive analysis, allowing for early detection of changes in patient mood or attitudes toward medications, which can trigger timely interventions for nonadherence [29-31]. Sentiment analysis of social media posts has proven effective in identifying concerns about side effects, medication efficacy, and treatment challenges, complementing traditional health care systems by offering new avenues for improving adherence to antidepressants [30].

Integration of NLP and LLMs in Mental Health Research

Recent NLP models have shown promise in identifying and classifying even rare neuropsychiatric diseases. For instance, the language model RoBERTa has been used to identify clusters of individuals with primary progressive aphasia through speech analysis, providing health care providers with valuable insights into patient experiences [32,33]. LLMs, which increasingly power conversational agents, can simulate synthetic personalities, allowing for fine-tuned, personalized patient interactions. These models amplify mental health support by offering tailored perspectives that improve adherence to therapy and foster more effective patient care [34].

XSL•FO
RenderX

Emotion-based models, such as those using emotional transition fingerprints, have demonstrated greater generalizability than content-based models, helping researchers understand the social dynamics of antidepressant use [35]. Furthermore, BERT has been used to categorize antidepressant-related discussions on social media, such as mood, sleep, and somatic symptoms, showcasing NLP's ability to provide targeted analysis in specific domains [36]. These tools have also been applied to predict drug interactions and optimize treatment plans using techniques such as Adaptive Fuzzy Logic Neural Networks, improving accuracy in predicting side effects [37].

Despite substantial research on antidepressant nonadherence, existing studies often neglect the dynamic interplay of emotions, personality traits, and social media discourse in shaping patient behaviors. Traditional approaches have provided foundational insights, yet they frequently fail to capture the real-time and nuanced patient experiences accessible through modern social media platforms. Furthermore, while PCC is increasingly recognized for its potential to improve adherence, there is a critical gap on how to leverage advanced NLP and LLMs to investigate patients' present psychological reaction with antidepressants using social media datasets.

This research aims to bridge the gap, which provides a more granular understanding of the psychological factors by analyzing sentiment, emotion, and personality profiles related to antidepressant use across diverse demographics. Furthermore, we explore topics, especially topics associated with 2 themes: nonadherence and self-care across sentiment and personality profile, which enhances our understanding of antidepressant users' real-time psychological patterns. The following 4 research questions (RQs) align to the goal of this study.

RQ1: How do sentiment analyses reveal differences in antidepressant experiences across demographics?

RQ2: What predominant emotions (eg, fear, sadness, and joy) are associated with antidepressant use, and how do these vary by demographics?

RQ3: What personality traits (eg, extraversion and neuroticism) are prevalent among users discussing antidepressants, and how do these traits differ across demographics?

RQ4: What are the prevailing topics in web-based antidepressant discussions, and how do sentiment, emotion, and personality insights inform our understanding of nonadherence and self-care themes?

Through these questions, this study addresses key limitations identified in the existing literature and harnesses advanced NLP techniques on social media data. The findings offer deeper understanding of patients' experience, improving antidepressant treatments.

Methods

Data Collection and Preprocessing for Antidepressant Analysis

To explore public perceptions of antidepressants, data were collected from 2 major platforms: AskaPatient [38] and Reddit [39], both of which offer diverse, patient-reported experiences. AskaPatient provides structured reviews and ratings, while Reddit features spontaneous, unfiltered discussions, offering a broader range of insights.

From AskaPatient, data were collected for the top 10 antidepressants, based on 1571 reviews posted between 2015 and 2023. This dataset includes demographic information (eg, gender and age), satisfaction ratings, side effects, and free-text comments (see Table 1 for sample data). The analyzed antidepressants were Cymbalta, Effexor XR, Lexapro, Zoloft, Wellbutrin XL, Wellbutrin, Celexa, Paxil, Wellbutrin SR, and Prozac.

We filtered posts from Reddit between 2018 and 2023 using the same top 10 antidepressants analyzed in the AskaPatient dataset. Due to the absence of explicit demographic details in Reddit posts, we used regular expressions (regex) to deduce demographic information such as gender and age. For instance, in the post: "Hi, I am a 19 y/o male. 6'2" and approximately 210 lbs. I currently take venlafaxine (Effexor) to treat depression and generalized anxiety disorder. I do not smoke. For the last several months, I have been experiencing hiccups…," we inferred the demographic information as "19 y/o male." (For a sample dataset, refer to Multimedia Appendix 1.)

This process constructed 36,252 posts with both inferred gender and age information. Preprocessing steps included text standardization such as lowercasing and stopwords removal. This dataset included 72.33% (25,972/36,252) of females and 27.67% (9936/36,252) of males, with 78% (28,277/36,252) of those aged 18 - 35 years, reflecting a predominantly younger demographic.

XSL•FO
RenderX

Table . Sample AskaPatient dataset.

Name	Reason	Satisfaction	Side effects	Comments	Sex	Age (years)
Celexa	Anxiety and depression	4	Loss of appetite, headache, insomnia, and feeling numb	Definitely changing my life. I feel like a new person and ready to live every day like I never did. The side effects are not that bad, my body is slowly feeling normal again. Hoping not having any side effects soon!	Female	21
Celexa	Autism and anxiety	3	Increased aggression with higher doses (5 mg)	Took edge off my son's anxiety but didn't do much else. He was more aggressive when the dose was raised.	Male	9
Celexa	Fibromyalgia and depression	3	Weight gain and fatigue not helping with the pain from fibromyalgia	Quit taking feel like there is something that will work better. Having anger issues and uncontrollable outburst.	Female	56
Celexa	Anxiety	1	Hair loss, vision haze, heart palps, and laziness	This was the start of my downfall I lost all my hair from it and lost my life slowly avoid this crap as its led me down a path of destruction.	Male	24

Analytical Techniques

To gain insights into patient experiences with antidepressants, we applied several advanced analytical methods, offering multiple perspectives on sentiment, emotions, personality traits, and latent themes.

Sentimental Analysis

We used the *twitter-roberta-base-sentiment-latest* model to classify text into negative, neutral, or positive sentiment. This model was trained on 124 million tweets from 2018 to 2021 and fine-tuned using the TweetEval benchmark [40]. For example, the comment "The side effects are not that bad, my body is slowly feeling normal again" yielded probabilities of 0.67 for positive, 0.28 for neutral, and 0.04 for negative sentiment. A 0.7 threshold was used for overall sentiment classification, but the full probability distributions were also analyzed for more nuanced insights.

Emotional Analysis

We used the English DistilRoBERTa-base model [41], optimized for detecting 6 basic emotions—anger, disgust, fear, joy, sadness, and surprise—along with neutral. This model trained on diverse datasets and enabled the identification of emotional states influencing antidepressant use.

Personality Analysis

LLMs, such as GPT-4 [42], were used for personality predictions, focusing on the Big Five traits—openness, conscientiousness, extraversion, agreeableness, and neuroticism. Each comment was assigned personality scores (1-5). Role prompting [43], a powerful artificial intelligence technique, was applied to guide the LLMs' response in interpreting personality traits based on patient comments. (Detailed prompts can be found in Multimedia Appendix 1.)

Topic Modeling Insights

We used BERTopic [44], a topic modeling approach leveraging transformers and class-based term frequency-inverse document frequency to generate coherent topics. To enhance interpretability, we used GPT-4 to refine topic labels by analyzing keywords and representative documents. The following prompt was used (Textbox 1).

To further explore the themes of nonadherence and PCC within the generated topics, we selected the 2 predominate topics related to nonadherence (side effects, withdrawal symptoms, etc) and PCC (personalized treatment, communication with providers, etc). These topics were most populated, which allowed the study to focus on the exploration of how sentiment, emotion, and personality traits associated with the 2 themes.

XSL•FO
RenderX

Textbox 1. Large language model prompt for topic label generation.

```
Prompt = """
I have topic that contains the following documents: \n[DOCUMENTS]
The topic is described by the following keywords: [KEYWORDS]
Based on the above information, can you give a short label of the topic?
"""
```

Ethical Considerations

The authors confirm that this research adhered to ethical guidelines and legal requirements in the United States. For the data obtained from AskaPatient, a data use agreement was signed to ensure compliance with ethical standards. Reddit data were sourced from publicly available platforms [45] and did not contain any personally identifiable information. The dataset was fully anonymized, and no private identifiers were included.

The data used in this research were obtained from 2 publicly accessible sources: AskaPatient and Reddit. Access to AskaPatient data was granted under a signed terms of use agreement between Consumer Health Resource Group, LLC, and authors, which provides a nonexclusive, revocable, and limited license for academic purposes. This agreement ensures compliance with ethical and legal requirements, including the stipulation that no more than 10 individual records or postings are included in any publication, with proper attribution to the AskaPatient website. Data from Reddit were publicly available and did not include any private, sensitive, or identifiable information. This study was partially supported by the National Science Foundation project (IIS-2041065) and adhered to the ethical guidelines and legal requirements outlined in the institutional review board regulations of Kent State University (reference no. KSU IRB20-182) and was classified as “exempt” under 45 CFR §46.104 (d), the federal policy for the protection of human subjects in the United States.

Results

In our analysis, traditional statistical methods (eg, *t* tests) were used to validate the observed trends in sentiment and emotion across demographic groups, with detailed results provided in Multimedia Appendix 1.

Sentiment Analysis

Significant gender differences were found in sentiment scores (Figures S4-S6 in Multimedia Appendix 1). Negative sentiment was higher in males ($t_{35,905}$=−6.79; P<.001), while females exhibited higher neutral ($t_{35,905}$=4.47; P<.001) and positive sentiment ($t_{35,905}$=5.60; P<.001). The majority of posts were classified as neutral (21,499/36,253, 59.3%), followed by negative (13,922/36,253, 38.4%), and a smaller proportion of positive posts (798/36,253, 2.2%).

The gender difference in overall sentiment indicates that females exhibited 38.8% (10,181/26,245) negative, 59.1% (15,515/26,245) neutral, and 2.1% (551/26,245) positive sentiments, while males showed 37.6% (3764/10,008) negative, 59.8% (5986/10,008) neutral, and 2.6% (260/10,008) positive sentiments.

Across age groups, negative sentiment was notably higher in younger age groups (0 - 18 years), while neutral sentiment was lower in the 0 - 18 years age groups than in the older age groups, particularly 31 - 45 years age group ($t_{35,903}$=−15.77; P<.001). Positive sentiment was also significantly lower in younger age groups (0 - 18 years) than in older age groups.

Sentiment by age range (Figure S7 in Multimedia Appendix 1) indicated that the negative sentiment was highest in the 0 - 18 years age group (median score ~0.62) and gradually decreases with age, stabilizing around 0.52 for the 46 - 60 and 60+ years age groups. Neutral sentiment was relatively stable across age, with the highest scores in the 31 - 45 and 60+ years age groups (~0.37) and the lowest scores in the 0 - 18 years age group (~0.31). Positive sentiment remained consistently low across all ages, peaking in the 46 - 60 years age group (~0.14).

Emotional Analysis

Significant gender differences were found for anger, disgust, joy, neutral, and sadness (P<.05). No significant differences were observed for fear and surprise between genders (Figures S10-S12 in Multimedia Appendix 1). The emotion analysis shows that fear was the most prevalent emotion across the dataset, present in 44.66% (16,196/36,253) of posts, followed by sadness (12,507/36,253, 34.49%). Surprise (2455/36,253, 6.77%) and anger (2132/36,253, 5.88%) were less common, while positive emotions such as joy appeared in only 4.88% (1769/36,253) of posts. Neutral emotions accounted for 2.67% (968/36,253), with disgust being the least frequent emotion (232/36,253, 0.64%).

A comparison of the mean emotion scores between females (**F**) and males (**M**) showed that fear was dominant for both genders (~0.40), with sadness slightly higher in females and joy more expressed in males, although both emotions had low overall score. Anger, disgust, and surprise were reported at similarly low levels for both genders, with females showing slightly higher levels of sadness and anger.

Across different age groups (0 - 18, 19 - 30, 31 - 45, 46 - 60, and 60+ years), fear remained the most prevalent emotion, peaking in the younger users (0 - 18 years) and gradually declining with age. Sadness followed a similar pattern, with higher scores among younger users before decreasing. Joy and neutral emotions remain low across all age groups, with slight increases in users aged 46 years and older. Anger, disgust, and surprise showed no significant variation across age ranges.

XSL•FO
RenderX

Personality Analysis

A gender difference was observed in agreeableness ($t_{35,903}$=3.78; P<.001). No significant gender differences were found for openness, conscientiousness, extraversion, or neuroticism (Figures S1-S17 in Multimedia Appendix 1). The proportion of users who scored the highest level (5) across the 5 personality traits, separated by gender, showed that agreeableness was highest in males (429/10,008, 4.29%) compared with females (1034/26,24, 53.94%). Conscientiousness and extraversion showed similar proportions, with 4% (1039/25,972) of females and 4.06% (403/9936) of males scoring 5. Neuroticism exhibited minimal differences, with females at 3.97% (1031/25,972) and males at 4.05% (402/9936). Openness scores were also closely matched, with females scoring 4.07% (1056/25,972) and males scoring 3.93% (391/9,936).

Pairwise *t* tests across age groups for openness, conscientiousness, extraversion, agreeableness, and neuroticism showed no significant differences after adjustment (all P>.05). Personality trait scores varied across age group. Agreeableness showed little variation across age groups, peaking in the 60+ years age group (1511/36,253, 4.17%) and reaching its lowest in the younger than 30 years age group (1450/36,253, 4%). Conscientiousness remained stable, with slightly higher scores in the 19 - 30 and 46 - 60 years age groups (1462/36,253, 4.03%) and lower scores in the 60+ years age group (1355/36,253, 3.74%). Extraversion was highest in the 60+ years age group (1555/36,253, 4.29%) and lowest in the 0 - 18+ years age group (1348/36,253, 3.72%). Neuroticism peaked in the 60+ years age group (1627/36,253, 4.49%) and was lowest in the 46 - 60 years age group (1360/36,253, 3.75%). Openness followed a similar trend, with the 60+ years age group scoring the highest (1704/36,253, 4.70%) and the 46 - 60 years age group scoring the lowest (1414/36,253, 3.90%).

Topic Analysis

Table 2 shows the results of BERTopic analysis within dataset, identifying key topics related to antidepressants user experiences. The largest topic, “Mental Health and Relationships,” accounted for 11.69% (3755/32,121) of the discussions, indicating a significant focus on managing these mental health conditions. Other notable topics included “Lexapro Experiences and Alternatives” (2292/32,121, 7.14%) and “Zoloft Withdrawal and Anxiety” (2184/32,121, 6.80%), reflecting common concerns about antidepressant experience and desire to switch to alternative medicines to manage their anxiety and depression.

Several discussions centered around side effects, including “Medication and Sexual Dysfunction” (2146/32,121, 6.68%) , “Chronic Abdominal Pain Issues (1949/32,121, 6.07%), “Skin Rashes and Sensitivities” (1700/32,121, 5.70%), “Heart Symptoms and Anxiety” (1638/32,121, 5.10%), and “Chronic Pain and Health Issues” (1300/32,121, 4.05%), alongside concerns related to weight gain and sleep disorders, which highlight frequent side effects users encounter during treatment.

Two primary themes—nonadherence and self-care—were extracted from the dataset based on the highest documented counts from relevant topics identified by the BERTopic model. In the nonadherence theme, “Zoloft Withdrawal and Anxiety,” ” Effexor Experiences and Withdrawals,” “Cymbalta Withdrawal Experiences and Effects,” and “Psychiatric Drug Withdrawal Effects” were assigned. In the self-care theme, the topics “mental health and relationships,” “relationship struggles and mental health,” and “toxic relationships and mental health” were chosen. Furthermore, we investigate sentiment, emotion, and personality traits across the 2 themes.

Sentiment distribution varied between the themes of nonadherence and self-care (Figure S19 in Multimedia Appendix 1). Neutral sentiment was most prevalent in both, accounting for 57.70% (3679/6376) in nonadherence and 47.65% (2929/6147) in self-care. Negative sentiment followed closely, with 36.68% (2339/6376) in nonadherence and 50.69% (3116/6147) in self-care, while positive sentiment remained low (358/6376, 5.61% and 102/6147, 1.66%, respectively).

The emotion distribution has been compared across the themes of nonadherence and self-care (Figure S20 in Multimedia Appendix 1). Fear was dominant emotion in both themes, making up 55.83% (3560/6376) of emotions in nonadherence and 35.90% (2207/6147) in self-care. Sadness was more prominent in self-care (2733/6147, 44.46%) than in nonadherence (1566/6376, 24.56%). Anger was also more frequent in self-care (592/6147, 9.63%) compared with nonadherence (265/6376, 4.16%).

Personality traits (scoring 5) also differed across themes (Figure S21 in Multimedia Appendix 1). In the nonadherence theme, agreeableness was the most common trait (2023/6376, 31.73%), followed by extraversion (1983/6376, 31.10%) and neuroticism (1892/6376, 29.68%). In the self-care theme, neuroticism was highest (1891/6147, 30.76%), followed by openness (1887/6147, 30.69%), conscientiousness (1804/6147, 29.34%), and agreeableness (1789/6,147, 29.10%).

XSL•FO
RenderX

Table . BERTopic result.

Topic	OpenAI label	Document count	Proportion
0	Mental Health and Relationships	3755	11.69
1	Lexapro Experiences and Alternatives	2292	7.14
2	Zoloft Withdrawal and Anxiety	2184	6.80
3	Medication and Sexual Dysfunction	2146	6.68
4	ADHD[a] Medication Challenges and Insights	2090	6.51
5	Chronic Abdominal Pain Issues	1949	6.07
6	Skin Rashes and Sensitivities	1700	5.70
7	Chronic Pain and Fatigue Issues	1651	5.14
8	Heart Symptoms and Anxiety	1638	5.10
9	Abnormal Blood Test Concerns	1446	4.50
10	Chronic Pain and Health Issues	1300	4.05
11	Weight Gain and Medication Struggles	1272	3.96
12	Relationship Struggles and Mental Health	1260	3.92
13	Experiences With Prozac for Anxiety	1176	3.66
14	Irregular Periods and Birth Control	1146	3.57
15	Toxic Relationships and Mental Health	1132	3.53
16	Wellbutrin Experiences and Challenges	1006	3.13
17	Chronic Sleep Disorders and Fatigue	990	3.08
18	Dizziness and Anxiety Symptoms	788	2.45
19	Chronic Headache and Migraine	773	2.40
20	Psychiatric Drug Withdrawal Effects	766	2.38
21	Anxiety and Substance Reactions	734	2.28
22	Effexor Experiences and Withdrawals	681	2.12
23	Chronic UTI[b] Symptoms and Challenges	517	1.61
24	Cymbalta Withdrawal Experiences and Effects	453	1.41
25	Alcohol, Anxiety, and Treatment Concerns	402	1.25
27	Eye Health and Vision Concerns	385	1.20
28	Chronic Ear Infections and Tinnitus	274	0.85
29	Managing Anxiety and Depression Tips	2	0.01

[a]ADHD: attention-deficit/hyperactivity disorder.

[b]UTI: urinary tract infection.

XSL•FO
RenderX

Discussion

Principal Findings

The finding reveals that sentiment of most of the antidepressant users is neutral within the dataset, while there remains substantial negative sentiment, and few show positive. In the meantime, the emotional analysis exhibits that fear and sadness command patients' emotional response. Those findings may consider the experiences when the patients use those antidepressants during the daily life. The community and health care can provide support to reduce the fear of side effects or failure of antidepressants, and family can facilitate comfort toward the sadness and negative emotions. Consequently, antidepressant users can practice self-care and enhance their treatment adherence.

Personality has not indicated a significant difference within the dataset. Topic analysis expresses that discussions mainly emphasize managing anxiety and depression, the efficacy of medications, and their side effects. Furthermore, in the nonadherence theme, the neutral sentiment was more prominent, followed by the significant fear emotion. In the self-care theme, the negative sentiment was more distinct, with the sadness emotion most dominant. Those findings suggest that the most beneficial route for antidepressant users is to exercise focused self-care for their mental health issues, while continuing their antidepressants treatment.

RQ1: How Do Sentiment Analyses Reveal Differences in Antidepressant Experiences Across Demographics?

The sentiment analysis indicates that public perception of antidepressants is predominantly neutral, with a substantial portion reflecting negative sentiment and very few positive expressions. The neutral sentiment suggests ambivalence or cautious optimism toward antidepressants, while the limited positive sentiment points to restricted satisfaction with treatment outcomes. These trends reflect broader society uncertainty or skepticism about antidepressant effectiveness and side effects, consistent with the literature identifying nonadherence as a pervasive issue [46].

In terms of gender, the results show minor differences in sentiment between males and females, with females exhibiting slightly higher negative sentiment and lower positive sentiment than males. This may reflect the differences in antidepressant experiences, societal pressures, or varying expectations regarding treatment outcomes across genders. Addressing these gender-specific concerns may enhance treatment satisfaction and adherence of antidepressants.

Notably, younger individuals expressed the highest levels of negative sentiment, which gradually decreases with age. This could be attributed to heightened sensitivity to side effects or unaccomplished expectations from treatment among younger users [47]. Older age groups demonstrated more stable neutral sentiment and slightly higher positive sentiment, particularly in the 46 - 60 years age group. These findings underscore the importance of patient's sentiment during antidepressant treatment. Mental health may benefit from tailoring interventions based on these emotional responses. Therefore, addressing negative bias early may be a key mechanism of antidepressant drug action and a potentially useful predictor of therapeutic response [48].

RQ2: What Predominant Emotions (eg, Fear, Sadness, and Joy) Are Associated With Antidepressant Use, and How Do These Vary by Demographics?

The emotional analysis revealed that fear and sadness dominate patients' emotional responses, with fear accounting for nearly half (44.66%) and sadness is the second most prominent emotion (34.49%). These findings suggest that patients' experiences with antidepressants are characterized by deep concerns and dissatisfaction. The prominence of fear, especially among younger users, may indicate anxiety related to side effects, potential long-term dependence, or uncertainties regarding the treatment's efficacy. This aligns with research showing that emotional blunting, commonly associated with antidepressants use, is also a symptom of depression and is linked to poorer remission outcomes [49].

The prominence of fear across both genders highlights widespread concerns about side effects, dependence, and treatment efficacy. However, the slightly higher level of sadness and anger in females suggests gender-specific emotional responses to antidepressants. Tailored communication strategies that address these gender-specific emotional concerns could improve treatment satisfaction and adherence.

As patients age, fear tends to decrease, potentially due to greater familiarity with antidepressants or prolonged exposure to treatment. However, the emotional burden remains strong among younger users, emphasizing the need for enhanced communication regarding treatment plans and potential side effects. The high prevalence of sadness**,** particularly among younger users, may reflect frustration or disappointment with treatment outcomes, potentially linked to unmet expectations, a lack of education over antidepressant, or emotional blunting. The intent of providing time-phased educational materials to patients is to maximize the relevance of such information by synchronizing it with typical recovery processes and issues. Additional efforts at engaging patients earlier after the initiation of treatment might be of most benefit [50].

RQ3: What Personality Traits (eg, Extraversion and Neuroticism) Are Prevalent Among Users Discussing Antidepressants, and How Do These Traits Differ Across Demographics?

Personality traits exhibited minor yet meaningful variations across gender and age groups. Agreeableness was higher in males, suggesting that empathy and social cooperation may positively influence adherence to mental health treatment. Concerning the age difference, one study found that the lower the extroversion score, the better the response to antidepressant treatment for adolescents [51]. From our finding, extraversion was highest in the 60+ years age group (4.29%) and lowest in the 0 - 18 years age group (3.72%). Compared with the higher negative sentiment, fear, and sadness emotions among the 0 - 18 years age group, the health care providers may tailor treatment to maximize the outcome of antidepressant medication.

XSL•FO
RenderX

Agreeableness is dominant in the nonadherence theme; neuroticism is the largest personality in the self-care theme. These are not notable findings that align with existing literatures for personality impact to antidepressant adherence and self-care; the result may provide a new perspective in the future research.

RQ4: What Are the Prevailing Topics in Web-Based Antidepressant Discussions, and How Do Sentiment, Emotion, and Personality Insights Inform Our Understanding of Nonadherence and Self-Care Themes?

The topic analysis identified that discussions primarily focus on mental health and relationships, medication efficacy, and side effects such as sexual dysfunction and withdrawal symptoms. These findings highlight clear concerns among users regarding the mental health issues and overall effectiveness of antidepressants. The dominance of neutral and negative sentiment in these discussions, combined with the high presence of fear and sadness, indicates that many users remain uncertain about their treatment outcomes and dissatisfy their daily life regarding their experience with the antidepressants.

Discussions around nonadherence are marked by fear, followed by sadness emotion, with patients expressing concerns about withdrawal symptoms and medication efficacy. In contrast, self-care discussions show a notable trend at sadness, accompanied by an increase in expressions of anger. A study found that stigma and fear of adverse events inhibited initiation of therapy, while adverse events and ineffectiveness of antidepressants contributed to discontinuation. Patients with strong perceptions of the necessity and few concerns about antidepressants were more likely to adhere to treatment at all phases of adherence [52].

Limitations

This study has several limitations. First, it relies on self-reported data from AskaPatient and Reddit, which may introduce bias as individuals who share their experiences on the web may not represent the broader population of antidepressant users. These users may disproportionately report extreme experiences, potentially skewing the findings. Furthermore, the focus on social media users limits the generalizability of the results to the wider population.

Although sentiment and emotion analysis tools are sophisticated, they may not fully capture the complexity of human emotions, particularly mixed or nuanced sentiments. Personality traits inferred from text via LLMs might not always align with real-world behaviors, possibly leading to misinterpretations. Furthermore, the study provides a snapshot of discussions at a particular time, without tracking changes over time. Furthermore, while age and gender are considered in the analysis, other demographic factors, such as socioeconomic status, education, and geographic location, were not controlled for and may influence user experiences with antidepressants. These factors should be considered in future studies to provide a more comprehensive understanding of patients' antidepressant use.

Conclusions

This study offers valuable insights into public perceptions of antidepressants, revealing that neutral and negative sentiments dominate discussions, with fear and sadness as the most common emotions, particularly among younger users. Personality traits exhibited minor yet meaningful variations across gender and age groups. Agreeableness was higher in males, suggesting that empathy and social cooperation may positively influence adherence to mental health treatment.

The topic analysis revealed that discussions primarily revolve around mental health and relationships, medication efficacy, and side effects, such as sexual dysfunction and withdrawal symptoms. Discussions about nonadherence are characterized by fear as the dominant emotion, followed by sadness, with patients voicing concerns about withdrawal symptoms and the effectiveness of medications. In contrast, self-care discussions convey a predominant sense of sadness, accompanied by heightened expression of anger.

The findings underscore the importance of self-care that addresses psychological dimensions of antidepressant use. Tailored interventions based on individual negative sentiment, fear, and sadness emotional challenges could improve overall treatment outcomes. By leveraging real-time social media data and advanced analytical tools, this research contributes to a more patient-centered understanding of antidepressant use, offering pathways for improving patient care and enhancing their nonadherence for antidepressants.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed supplementary data on public perceptions of antidepressants discussed on social media, including topic analysis results, sentiment breakdowns, and associated statistical summaries.
[DOCX File, 1836 KB - formative_v9i1e62680_app1.docx]

References

1. Brody DJ, Gu Q. Antidepressant use among adults: United States, 2015-2018. NCHS Data Brief 2020 Sep(377):1-8. [Medline: 33054926]
2. Marasine NR, Sankhi S. Factors associated with antidepressant medication non-adherence. Turk J Pharm Sci 2021 Apr 20;18(2):242-249. [doi: 10.4274/tjps.galenos.2020.49799] [Medline: 33902271]

XSL•FO
RenderX

3. Sansone RA, Sansone LA. Antidepressant adherence: are patients taking their medications? Innov Clin Neurosci 2012 May;9(5-6):41-46. [Medline: 22808448]
4. Deen TL, Fortney JC, Pyne JM. Relationship between satisfaction, patient-centered care, adherence and outcomes among patients in a collaborative care trial for depression. Adm Policy Ment Health 2011 Sep;38(5):345-355. [doi: 10.1007/s10488-010-0322-z] [Medline: 20978932]
5. Rathert C, Wyrwich MD, Boren SA. Patient-centered care and outcomes: a systematic review of the literature. Med Care Res Rev 2013 Aug;70(4):351-379. [doi: 10.1177/1077558712465774] [Medline: 23169897]
6. Kramer PD. Listening to Prozac: The Landmark Book about Antidepressants and the Remaking of the Self: Penguin; 1997.
7. Opbroek A, Delgado PL, Laukes C, et al. Emotional blunting associated with SSRI-induced sexual dysfunction. Do SSRIs inhibit emotional responses? Int J Neuropsychopharmacol 2002 Jun;5(2):147-151. [doi: 10.1017/S1461145702002870] [Medline: 12135539]
8. Sansone RA, Sansone LA. SSRI-induced indifference. Psychiatry (Edgmont) 2010 Oct;7(10):14-18. [Medline: 21103140]
9. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient non-compliance with paper diaries. BMJ 2002 May 18;324(7347):1193-1194. [doi: 10.1136/bmj.324.7347.1193] [Medline: 12016186]
10. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
11. Saha K, Sugar B, Torous J, Abrahao B, Kıcıman E, De Choudhury M. A social media study on the effects of psychiatric medication use. Proc Int AAAI Conf Weblogs Soc Media 2019;13:440-451. [doi: 10.1609/icwsm.v13i01.3242]
12. Saha K, Torous J, Kiciman E, De Choudhury M. Understanding side effects of antidepressants: large-scale longitudinal study on social media data. JMIR Ment Health 2021 Mar 19;8(3):e26589. [doi: 10.2196/26589] [Medline: 33739296]
13. Zhu J, Jin R, Kenne DR, Phan N, Ku WS. User dynamics and thematic exploration in r/depression during the COVID-19 pandemic: insights from overlapping r/SuicideWatch users. J Med Internet Res 2024 May 20;26:e53968. [doi: 10.2196/53968] [Medline: 38767953]
14. Carabot F, Fraile-Martínez O, Donat-Vargas C, et al. Understanding public perceptions and discussions on opioids through Twitter: cross-sectional infodemiology study. J Med Internet Res 2023 Oct 31;25:e50013. [doi: 10.2196/50013] [Medline: 37906234]
15. Park A, Conway M. Tracking health related discussions on Reddit for public health applications. AMIA Annu Symp Proc 2017;2017:1362-1371. [Medline: 29854205]
16. Zhu J, Yalamanchi N, Jin R, Kenne DR, Phan N. Investigating COVID-19's impact on mental health: trend and thematic analysis of reddit users' discourse. J Med Internet Res 2023 Jul 12;25:e46867. [doi: 10.2196/46867] [Medline: 37436793]
17. Stade EC, Stirman SW, Ungar LH, et al. Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. NPJ Ment Health Res 2024 Apr 2;3(1):12. [doi: 10.1038/s44184-024-00056-z] [Medline: 38609507]
18. Yang K, Zhang T, Kuang Z, Xie Q, Huang J, Ananiadou S. MentaLLaMA: interpretable mental health analysis on social media with large language models. Presented at: WWW '24; May 13-17, 2024; Singapore, Singapore. [doi: 10.1145/3589334.3648137]
19. Chowdhury AK, Sujon SR, Shafi MSS, et al. Harnessing large language models over transformer models for detecting Bengali depressive social media text: a comprehensive study. Nat Lang Process J 2024 Jun;7:100075. [doi: 10.1016/j.nlp.2024.100075]
20. Pompili M, Venturini P, Palermo M, et al. Mood disorders medications: predictors of nonadherence—review of the current literature. Expert Rev Neurother 2013 Jul;13(7):809-825. [doi: 10.1586/14737175.2013.811976] [Medline: 23898852]
21. Kamaradova D, Latalova K, Prasko J, et al. Connection between self-stigma, adherence to treatment, and discontinuation of medication. Pat Prefer Adherence 2016;10:1289-1298. [doi: 10.2147/PPA.S99136] [Medline: 27524884]
22. De las Cuevas C, Peñate W, Sanz EJ. Risk factors for non-adherence to antidepressant treatment in patients with mood disorders. Eur J Clin Pharmacol 2014 Jan;70(1):89-98. [doi: 10.1007/s00228-013-1582-9] [Medline: 24013851]
23. Sawada N, Uchida H, Suzuki T, et al. Persistence and compliance to antidepressant treatment in patients with depression: a chart review. BMC Psychiatry 2009 Jun 16;9:38. [doi: 10.1186/1471-244X-9-38] [Medline: 19531229]
24. Epstein RM, Street RL. The values and value of patient-centered care. Ann Fam Med 2011;9(2):100-103. [doi: 10.1370/afm.1239] [Medline: 21403134]
25. 2023 webinars planning information. UNC Lineberger Cancer Network. URL: https://unclineberger.org/unclcn/wp-content/uploads/sites/871/2024/03/2023-Telehealth-Webinar-Planning-Guide-1.pdf [accessed 2024-11-24]
26. Levinson W. Patient-centred communication: a sophisticated procedure. BMJ Qual Saf 2011 Oct;20(10):823-825. [doi: 10.1136/bmjqs-2011-000323]
27. McPherson S, Wicks C, Tercelli I. Patient experiences of psychological therapy for depression: a qualitative metasynthesis. BMC Psychiatry 2020 Jun 18;20(1):313. [doi: 10.1186/s12888-020-02682-1] [Medline: 32552748]
28. Ell G. Encouraging self-care and positive lifestyle changes in patients with depression. Pharm J 2020 [FREE Full text]
29. Sidana S, Mishra S, Amer-Yahia S, Clausel M, Amini MR. Health monitoring on social media over time. Presented at: SIGIR '16; Jul 7-21, 2016; Pisa Italy. [doi: 10.1145/2911451.2914697]

XSL•FO
RenderX

30. Naslund JA, Bondre A, Torous J, Aschbrenner KA. Social media and mental health: benefits, risks, and opportunities for research and practice. J Technol Behav Sci 2020 Sep;5(3):245-257. [doi: 10.1007/s41347-020-00134-x] [Medline: 33415185]
31. Roeder A. Social Media Use Can Be Positive for Mental Health and Well-Being: Harvard TH Chan School of Public Health; 2020.
32. Rezaii N, Wolff P, Price BH. Natural language processing in psychiatry: the promises and perils of a transformative approach. Br J Psychiatry 2022 Jan 7;220(5):1-3. [doi: 10.1192/bjp.2021.188] [Medline: 35048814]
33. Stewart R, Velupillai S. Applied natural language processing in mental health big data. Neuropsychopharmacology 2021 Jan;46(1):252-253. [doi: 10.1038/s41386-020-00842-1] [Medline: 32895453]
34. Serapio-García G, Safdari M, Crepy C, et al. Personality traits in large language models. Research Square. Preprint posted online on Aug 8, 2023. [doi: 10.21203/rs.3.rs-3296728/v1]
35. Guo X, Sun Y, Vosoughi S. Emotion-based modeling of mental disorders on social media. Presented at: WI-IAT '21; Dec 14-17, 2021; Melbourne, Australia. [doi: 10.1145/3486622.3493916]
36. de Anta L, Alvarez-Mon MA, Ortega MA, et al. Areas of interest and social consideration of antidepressants on English tweets: a natural language processing classification study. J Pers Med 2022 Jan 25;12(2):155. [doi: 10.3390/jpm12020155] [Medline: 35207644]
37. Duraisamy P, Natarajan Y, Preethaa KRS, Mouthami K. Sentiment analysis on drug reviews using diverse classification techniques. Presented at: 2022 3rd International Conference on Communication, Computing and Industry 40 (C2I4); Dec 15-16, 2022; Bangalore, India. [doi: 10.1109/C2I456876.2022.10051399]
38. AskaPatient. URL: https://www.askapatient.com/ [accessed 2025-02-13]
39. Reddit. URL: https://www.reddit.com/ [accessed 2025-02-13]
40. Barbieri F, Camacho-Collados J, Espinosa Anke L, Neves L. TweetEval: unified benchmark and comparative evaluation for tweet classification. Presented at: Findings of the Association for Computational Linguistics; Nov 16-20, 2020. [doi: 10.18653/v1/2020.findings-emnlp.148]
41. Sanh V, Debut L, Chaumond J, Wolf T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv. Preprint posted online on Oct 2, 2019. [doi: 10.48550/arXiv.1910.01108]
42. OpenAI. GPT-4 technical report. 2023. URL: https://openai.com/research/gpt-4 [accessed 2025-02-13]
43. Leonard L. Role prompting: the secret to effective AI interactions. The Software Forum. URL: https://www.thesoftwareforum.com/role-prompting [accessed 2025-02-13]
44. Grootendorst M. BERTopic: neural topic modeling with a class-based TF-IDF procedure. arXiv. Preprint posted online on Mar 11, 2022. [doi: 10.48550/arXiv.2203.05794]
45. Reddit comments/submissions 2005-06 to 2023-09. Academic Torrents. URL: https://academictorrents.com/details/89d24ff9d5fbc1efcdaf9d7689d72b7548f699fc [accessed 2025-02-13]
46. Dell'Osso B, Albert U, Carrà G, et al. How to improve adherence to antidepressant treatments in patients with major depression: a psychoeducational consensus checklist. Ann Gen Psychiatry 2020;19:61. [doi: 10.1186/s12991-020-00306-2] [Medline: 33062034]
47. Luft MJ, Lamy M, DelBello MP, McNamara RK, Strawn JR. Antidepressant-induced activation in children and adolescents: risk, recognition and management. Curr Probl Pediatr Adolesc Health Care 2018 Feb;48(2):50-62. [doi: 10.1016/j.cppeds.2017.12.001] [Medline: 29358037]
48. Godlewska BR, Browning M, Norbury R, Cowen PJ, Harmer CJ. Early changes in emotional processing as a marker of clinical response to SSRI treatment in depression. Transl Psychiatry 2016 Nov 22;6(11):e957. [doi: 10.1038/tp.2016.130] [Medline: 27874847]
49. Goodwin GM, Price J, De Bodinat C, Laredo J. Emotional blunting with antidepressant treatments: a survey among depressed patients. J Affect Disord 2017 Oct 15;221:31-35. [doi: 10.1016/j.jad.2017.05.048] [Medline: 28628765]
50. Mundt JC, Clarke GN, Burroughs D, Brenneman DO, Griest JH. Effectiveness of antidepressant pharmacotherapy: the impact of medication compliance and patient education. Depress Anxiety 2001;13(1):1-10. [doi: 10.1002/1520-6394(2001)13:1<1::aid-da1>3.0.co;2-s] [Medline: 11233454]
51. Ran LY, Liu XY, Wang W, et al. Personality traits predict treatment outcome of an antidepressant in untreated adolescents with depression: an 8-week, open-label, flexible-dose study. J Affect Disord 2024 Apr 1;350:102-109. [doi: 10.1016/j.jad.2024.01.015] [Medline: 38199422]
52. Gupta R, Ariefdjohan M. Mental illness on Instagram: a mixed method study to characterize public content, sentiments, and trends of antidepressant use. J Ment Health 2021 Aug;30(4):518-525. [doi: 10.1080/09638237.2020.1755021] [Medline: 32325006]

Abbreviations

LLM: large language model
NLP: natural language processing
PCC: patient-centered care
RQ: research question

XSL•FO
RenderX

Edited by A Mavragani; submitted 28.05.24; peer-reviewed by E Chan, H Rafi, R Norel; revised version received 08.11.24; accepted 13.11.24; published 26.02.25.

Please cite as:
Zhu J, Zhang X, Jin R, Jiang H, Kenne DR
Probing Public Perceptions of Antidepressants on Social Media: Mixed Methods Study
JMIR Form Res 2025;9:e62680
URL: https://formative.jmir.org/2025/1/e62680
doi:10.2196/62680

© Jianfeng Zhu, Xinyu Zhang, Ruoming Jin, Hailong Jiang, Deric R Kenne. Originally published in JMIR Formative Research (https://formative.jmir.org), 26.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Novel Evaluation Metric and Quantified Performance of ChatGPT-4 Patient Management Simulations for Early Clinical Education: Experimental Study

Riley Scherr[1], BS; Aidin Spina[1], BS; Allen Dao[2], BS; Saman Andalib[1*], BS; Faris F Halaseh[1*], BS; Sarah Blair[3], BS; Warren Wiechmann[4], MD, MBA, MSED; Ronald Rivera[4], MD

[1]School of Medicine, University of California, Irvine School of Medicine, 1001 Health Sciences Road, Irvine, CA, United States

[2]Department of Medicine, Stanford Medicine, Stanford, CA, United States

[3]School of Medicine, Tufts University, Boston, MA, United States

[4]Department of Emergency Medicine, University of California Irvine, Orange, CA, United States

[*]these authors contributed equally

Corresponding Author:
Riley Scherr, BS
School of Medicine, University of California, Irvine School of Medicine, 1001 Health Sciences Road, Irvine, CA, United States

Abstract

Background: Case studies have shown ChatGPT can run clinical simulations at the medical student level. However, no data have assessed ChatGPT's reliability in meeting desired simulation criteria such as medical accuracy, simulation formatting, and robust feedback mechanisms.

Objective: This study aims to quantify ChatGPT's ability to consistently follow formatting instructions and create simulations for preclinical medical student learners according to principles of medical simulation and multimedia educational technology.

Methods: Using ChatGPT-4 and a prevalidated starting prompt, the authors ran 360 separate simulations of an acute asthma exacerbation. A total of 180 simulations were given correct answers and 180 simulations were given incorrect answers. ChatGPT was evaluated for its ability to adhere to basic simulation parameters (stepwise progression, free response, interactivity), advanced simulation parameters (autonomous conclusion, delayed feedback, comprehensive feedback), and medical accuracy (vignette, treatment updates, feedback). Significance was determined with χ^2 analyses using 95% CIs for odds ratios.

Results: In total, 100% (n=360) of simulations met basic simulation parameters and were medically accurate. For advanced parameters, 55% (200/360) of all simulations delayed feedback, while the Correct arm (157/180, 87%) delayed feedback was significantly more than the Incorrect arm (43/180, 24%; $P<.001$). A total of 79% (285/360) of simulations concluded autonomously, and there was no difference between the Correct and Incorrect arms in autonomous conclusion (146/180, 81% and 139/180, 77%; $P=.36$). Overall, 78% (282/360) of simulations gave comprehensive feedback, and there was no difference between the Correct and Incorrect arms in comprehensive feedback (137/180, 76% and 145/180, 81%; $P=.31$). ChatGPT-4 was not significantly more likely to conclude simulations autonomously ($P=.34$) and provide comprehensive feedback ($P=.27$) when feedback was delayed compared to when feedback was not delayed.

Conclusions: These simulations have the potential to be a reliable educational tool for simple simulations and can be evaluated by a novel 9-part metric. Per this metric, ChatGPT simulations performed perfectly on medical accuracy and basic simulation parameters. It performed well on comprehensive feedback and autonomous conclusion. Delayed feedback depended on the accuracy of user inputs. A simulation meeting one advanced parameter was not more likely to meet all advanced parameters. Further work must be done to ensure consistent performance across a broader range of simulation scenarios.

(JMIR Form Res 2025;9:e66478) doi:10.2196/66478

KEYWORDS

medical school simulations; AI in medical education; preclinical curriculum; ChatGPT; ChatGPT-4; medical simulation; simulation; multimedia; feedback; medical education; medical student; clinical education; pilot study; patient management

XSL•FO
RenderX

Introduction

Background

With the rise of generative artificial intelligence (AI) applications such as OpenAI's ChatGPT, research into its medical application has expanded. Early investigations assessed whether large language models (LLMs) could pass medical trainee licensing examinations [1-5]. Such studies indicated that LLMs can pass medical exams and possess solid foundations in medical reasoning [3,6-8].

An expanding body of literature has focused on LLMs in medical education, including the perspectives of both students and seasoned clinicians [9-13]. Medical students have shown interest in LLMs and frequently use or intend to use them educationally [14]. Studies have also suggested that ChatGPT can be an effective tool for students when entering the clinical wards [15]. Moving from theoretical, classroom-based instruction to hands-on patient care introduces new challenges [16]. The steep learning curve heightens the need for reliable training tools, and generative AI technologies may satisfy this need.

Current literature suggests a need for rigorous validation of student use of generative AI. Frameworks such as RISE (role, input, steps, expectations; ie, inputting the LLM's role, anticipated input, required steps, and desired expectations) exist for prompt engineering but have been infrequently applied to medical student LLM use [17]. Initial work established prompts for medical trainees to practice common clinical scenarios in platforms such as ChatGPT [9,18]. These studies established that, given the precise wording of prompts, LLMs can act as an effective simulator of basic clinical scenarios. Despite this promising result, these technologies' unknown accuracy and precision in displaying desirable parameters limits their broader applicability [19-21].

In this study, we reviewed evidence-based resources for medical simulation and multimedia educational design. Using The Society for Simulation in Healthcare guidelines along with Richard Mayer's multimedia design principles, we created an evaluation system with 3 main evaluation categories: basic simulation parameters, advanced simulation parameters, and medical accuracy parameters. Each category was further divided into 3 subparameters. Basic parameters were composed of stepwise progression, free response, and interactivity. Advanced parameters were composed of autonomous conclusion, delayed feedback, and comprehensive feedback. Medical accuracy parameters were composed of clinical vignette, updates based on treatment, and feedback. See Figure 1 for definitions.

XSL•FO
RenderX

Figure 1. Simulation parameters. Desirable simulation parameters of ChatGPT clinical simulations were based on multimedia educational technology principles and simulation in health care recommendations. Three major parameters were each divided into 3 corresponding subparameters with clear definitions. Subparameters constitute the total 9-part metric by which ChatGPT simulations were evaluated. “P” or green indicates a pass, and “F” or red indicates a fail for whether a simulation exhibits that characteristic.

Basic Parameters

Stepwise progression in simulation is supported by the segmenting principle of multimedia learning, meaning that learning is maximized when broken into smaller units [22]. At breaks between units, the learner provides the next steps in the simulated patient’s management with reasoning. This is supported by Mayer’s self-explanation principle, which shows strong learning when previous knowledge is integrated into current learning by explaining one’s reasoning [23]. An user-responsive design further facilitates this and creates a generative learning activity poised to improve learning acquisition and retention [24]. These are both in-line with real medical practice where every step of a treatment or diagnosis must be chosen. The free-response format for responses similarly mimics real medical practice, as it allows space for treatment justifications. It is also supported by the active processing principle, which asserts that learning is improved when people actively organize information into cognitive models and integrate prior knowledge to address the new task, rather than passively absorbing teaching points. Just like in real life, where physicians create treatment plans de novo, rather than choosing from a selection of options [25]. Finally, interactivity promotes cognitive engagement and is beneficial to learning outcomes, especially when planned and augmented by peer-to-peer and peer-to-teacher learning [26,27].

Advanced Parameters

Advanced parameters were selected to improve authenticity and decrease the cognitive load devoted to simulation mechanics, allowing greater focus on learning engagement [28]. An autonomous conclusion is valuable because it provides a specific end point and prevents learners from following extraneous threads when learning goals have been achieved. Learners do not have to decide when they have learned enough, rather they must meet the simulation goals to finish. Delayed feedback

XSL•FO
RenderX

keeps novice learners from overrelying on immediate feedback cues to make decisions and promotes active learning in simulations [29]. However, Mayer suggests immediate feedback may be better for some tasks such as solo novice learning or guided reasoning with an expert, but the Society for Simulation in Healthcare guidelines suggest that delayed debrief is most appropriate, and we accordingly chose delayed feedback for our metric [30]. This also mimics the real-life practice of medicine, where the only indication of a treatment's accuracy is a change in patient stability. Delaying feedback also keeps these simulations in-line with the active-processing principle that active learning is more effective than passive learning from regular performance feedback cues [29]. Finally, comprehensive feedback falls in-line with Mayer's signaling principle. Signaling, or indicating what information is vital, highlights specific learning points rather than generalizing about a user's performance [31]. Comprehensive feedback thus ensures meaningful performance assessments with specific takeaways.

Medical Accuracy Parameters

Medical accuracy is perhaps the most important. This is a common-sense principle that undergirds all the other parameters' efficacy. Medical accuracy categories were defined whenever there were physiologic, pharmacologic, or pathologic assertions.

A 9-part model for evaluating AI medical simulations is therefore proposed. Using a previously engineered and verified starting prompt, this model helps quantify the consistency of an LLM (ChatGPT-4) in following prompt commands to create simulations exhibiting these 9 features of effective simulations. The provided results aim to better inform recommendations to students and educators who use such resources in their clinical education and establish a method for future evaluation of simulation reliability.

Methods

Study Design

The Society for Simulation in Healthcare recommends simulation sessions be high or low fidelity with low physical realism, making AI simulation sessions acceptable when high-fidelity sessions are unavailable [30]. This is also congruent with the authentic learning environments principle which suggests that learning can happen equivalently in any environment as long as the design adheres to effective learning principles [32]. The Society also recommends training should be spaced, frequent, short, and skill-oriented sessions. Therefore, we built our prompt around asthma exacerbation management, a short lesson with evidence-based best practices to be done in 5 to 10 minutes with specific learning takeaways. This is further supported by the Society's recommendation that each session end with focused feedback via structured debrief, so our prompt accordingly addressed this. The only criteria we were unable to meet were recommendations for in-situ practice and the use of interprofessional teams. However, the authentic learning environments principle again suggests that these simulations can have educational benefits even though they are not in situ, given our use of educational design principles. The following prompt, developed with features of RISE prompt engineering, was subsequently used [9]:

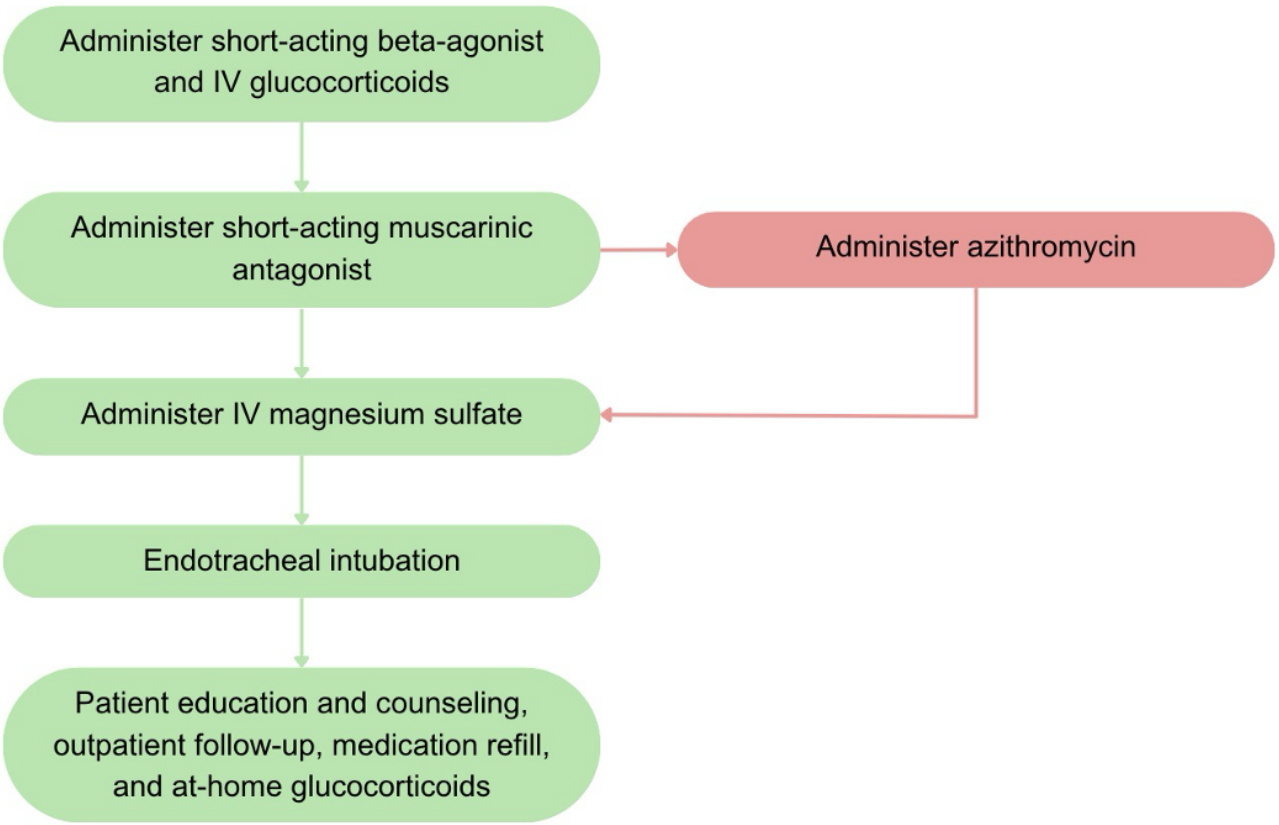
"Please create a clinical scenario on a patient presenting to the hospital with an acute asthma exacerbation and quiz me on what the proper next step of management is. Please make it free response and interactive, meaning you ask me what the next step is one question at a time, and then I write out what I would do, and then you ask me another question based on how my answer would affect the patient. Please update/change the patient's condition based on my actions, and do not tell me the right answers until the end of the scenario."

Simulations were run on 1 of 3 ChatGPT-4 accounts due to ChatGPT-4's token limits and split across Safari, Google Chrome, and Firefox browsers from March 29, 2024, to April 24, 2024. Each simulation was run on its own browser tab session of ChatGPT-4 to eliminate session-dependent memory. All cases simulated an acute asthma exacerbation and focused on treatment steps.

An appropriate treatment algorithm for acute asthma exacerbation was developed based on the review of recommendations from the American Thoracic Society (a leading pulmonology society) and the American Academy of Family Physicians, and confirmed by author RR, a board-certified emergency medicine physician at a high-volume academic medical center [33]. Movement to the next step of the algorithm depended on the simulated patient's clinical condition, and the tester could proceed to the final step (discharge) if the patient was appropriately stabilized in a stepwise manner. For example, if a patient was no longer in respiratory distress and had normal oxygen saturation after administration of a short-acting muscarinic antagonist, the tester would skip magnesium sulfate and endotracheal intubation and proceed to patient education and counseling. See Figure 2 for treatment workflow.

XSL•FO
RenderX

Figure 2. Acute asthma simulation treatment algorithms. Treatment algorithms for acute asthma exacerbation simulations were developed on societal guidelines and were used as a template for ChatGPT simulation inputs. The green pathway represents "correct treatment" and was followed for 180 simulations in the correct subgroup. The red path represents "incorrect treatment," a variation of the correct treatment algorithm, and was followed for 180 simulations in the incorrect subgroup.



Simulations were divided into 2 subgroups based on adherence to the treatment workflow. The "Correct Treatment" subgroup followed the treatment workflow. The "Incorrect Treatment" subgroup added administration of a macrolide antibiotic (azithromycin) as the third step (Figure 2). Progression through the algorithm for each simulation depended on the patient's response to treatment, but all incorrect simulations included steps 1, 2, 3, and 6 at minimum. In total, 180 simulations each were run for the Correct and Incorrect treatment arms, leaving a total of 360 simulations for analysis. Simulations were evaluated according to our 9-part model using the definitions in Figure 1.

Author RS resolved unclear data points. Simulations were reviewed by all authors for medical accuracy and confirmed with author RR, a licensed, board-certified emergency medicine physician with 8 years of clinical experience.

Descriptive statistics were run on GraphPad Prism (version 10.2.3; Graphstats Technologies) and Excel (version 16.85; Microsoft Corp). Statistical significance was determined by χ^2 test, with an α level set at .05.

Ethical Considerations

Study data were collected by an author (AD) and did not involve personal information or observation of any person's private or public information. As such data were collected from ChatGPT, all study data were inherently anonymous. This study was reviewed under the University of California, Irvine's Institutional Review Board Protocol #3213 and deemed not human subjects research and not requiring institutional review board review.

Results

A summary of the results is listed in Table 1 and represented in Figure 3.

XSL•FO
RenderX

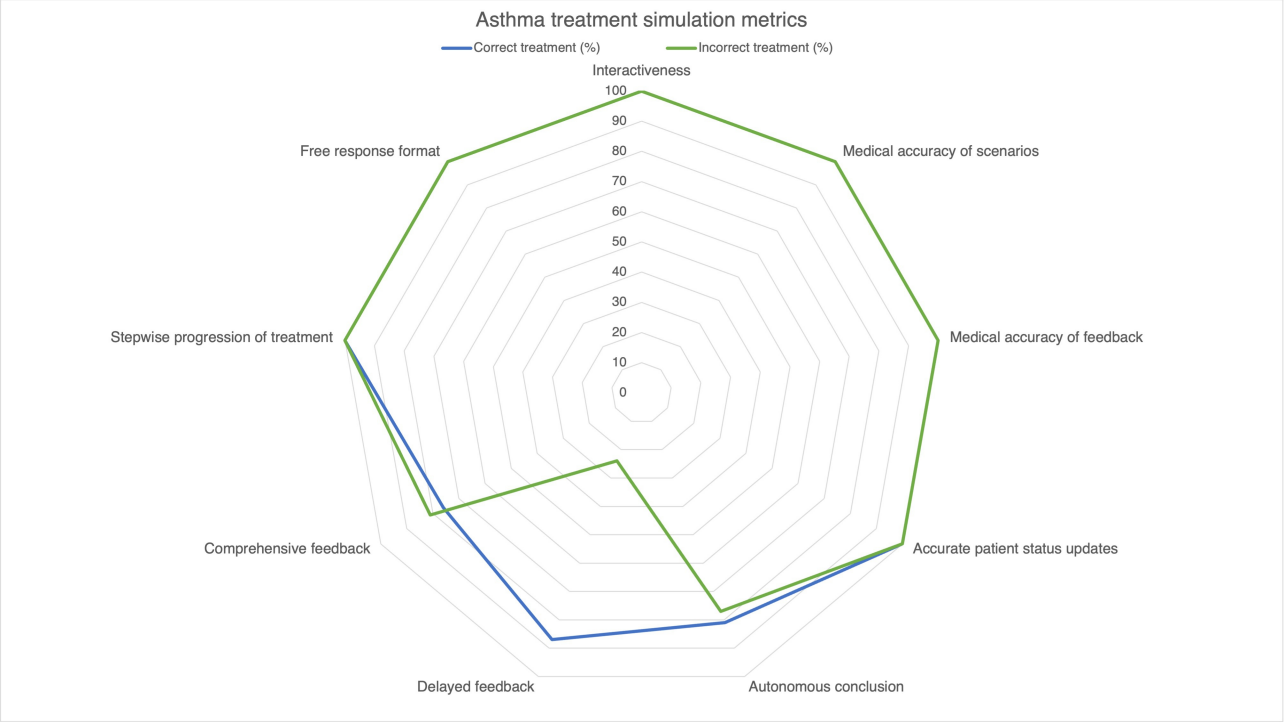
Table . Simulation characteristics[a].

	Combined (n=360), n (%)	Correct (n=180), n (%)	Incorrect (n=180), n (%)	χ^2 (*df*)	*P* value[b]
Basic parameters	360 (100)	180 (100)	180 (100)	0 (1)	≥.99
Medical accuracy	360 (100)	180 (100)	180 (100)	0 (1)	≥.99
Comprehensive feedback	282 (78)	137 (76)	145 (81)	1.05 (1)	.31
Autonomous conclusion	285 (79)	146 (81)	139 (77)	0.825 (1)	.36
Delayed feedback	200 (55)	157 (87)	43 (24)	146.2 (1)	<.001

[a]Simulations were given the correct or incorrect treatment algorithm inputs and evaluated on exhibition of characteristics against the 9-part metric. Successful exhibition of a parameter was calculated as a percent of all simulations in a subgroup. The table provides a summary of simulation outcomes.

[b]*P* values were calculated from χ^2 values

Figure 3. ChatGPT performance on simulation parameters by correct and incorrect inputs. Radar plot showing correct (blue) and incorrect (green) ChatGPT-4 performance on simulation parameters. Each point represents a simulation parameter. Lines indicate the percentage of simulations exhibiting that parameter, with the circumference of the shape indicating 100%. Simulations were given the correct or incorrect acute asthma exacerbation treatment algorithm inputs and evaluated by the 9-part metric developed from multimedia education and simulation principles. Successful exhibition of a parameter was calculated as a percent of all simulations in a subgroup and plotted on the radar plot. Correct and Incorrect subgroups performed similarly on all metrics except “delayed feedback.”



In 360 completed simulations, ChatGPT adhered to basic parameters of stepwise progression of treatment, appropriate patient status adjustments based on treatment inputs, and free response format in 100% (n=360) of simulations. It was also found to be 100% medically accurate in its vignettes, treatment updates, and feedback. There was no difference in medical accuracy between ChatGPT-4 outputs after correct versus incorrect tester treatment responses.

Results for advanced parameters were mixed. Among the correct treatment subgroup (n=180), 87% (157/180) of simulations delayed feedback until the end of the scenario. Only 24% (43/180) of the simulations with incorrect treatment (n=180) delayed feedback (*P*<.001). In all but 2 of these simulations with immediate (nondelayed) feedback, the feedback came after the administration of azithromycin. When combining correct and incorrect simulations (n=360), ChatGPT-4 demonstrated a 55% success rate in delaying feedback.

Simulations with correct treatments concluded the scenario autonomously 81% (146/180) of the time, while those with incorrect treatment concluded autonomously 77% (139/180; *P*=.36). Overall, 79% (285/360) of all simulations, regardless of treatment accuracy, concluded autonomously without tester input.

Correct simulations provided comprehensive feedback 76% (137/180) of the time while incorrect simulations provided comprehensive feedback 81% (145/180) of the time (*P*=.31).

XSL•FO
RenderX

Combined analysis showed ChatGPT-4 provided comprehensive feedback on 76% (137/180) of outputs.

Further analysis revealed that ChatGPT-4 was not more likely to conclude the simulation autonomously (P=.34) and provide comprehensive feedback (P=.27) when feedback was delayed compared to when feedback was not delayed (Table 2).

Table . Impact of delayed feedback on simulation autonomous conclusion and comprehensive feedback.

	Delayed feedback	Early feedback
Autonomous conclusion[a]	162	123
No autonomous conclusion	38	37
Comprehensive feedback[b]	161	121
No comprehensive feedback	39	39

[a]Number of all simulations regardless of subgroup that reached an autonomous conclusion versus those that did not based on delayed or early feedback timing.

[b]Number of all simulations regardless of subgroup that provided comprehensive feedback versus those that did not based on delayed or early feedback timing.

Discussion

Basic Parameters

ChatGPT-4 performed perfectly on basic simulation parameters, which is a requirement for it to be a viable educational tool. Its performance agrees with other research on prompt engineering showing proper prompt generation can produce high-quality results [34]. However, even given the use of proper prompt engineering techniques, perfect performance is surprising given ChatGPT's known inconsistencies [4,34,35]. This may reflect its algorithmic refinement and increased LLM training from widespread use. Given this known feature of LLMs, we should expect it to continue improving over time.

Medical Accuracy

ChatGPT was also medically accurate for both patient progression within the simulation and in its feedback; simulated patients had reasonable clinical presentations, responded appropriately to therapies, and feedback was in accordance with basic pharmacology, physiology, antibiotic stewardship, and clinical practice. This agrees with other findings on ChatGPT's clinical reasoning abilities, but the consistently excellent performance is again impressive. However, acute asthma exacerbation is not an exceedingly complex problem and has well-established management guidelines that likely exist in ChatGPT's training data, so its medical accuracy cannot be generalized to more complex clinical simulations like intensive care management [2,4]. Further research needs to be done into LLM performance differences between complex and simple simulations.

Advanced Parameters

ChatGPT's performance on feedback was variable. The observed significant difference in delaying feedback is noteworthy. Delayed feedback best mirrors real-life clinical scenarios where immediate feedback is not always available to redirect incorrect thought processes [36,37]. However, immediate feedback to incorrect answers mirrors the role a teacher might take with a student in simulation, where the teacher does not give a correct answer but instead explains why a learner's answer was incorrect in an attempt to give a better chance at the correct answer. It is also in-line with the expertise reversal effect that suggests novice learners will do better with more intense educational guidance, while advanced learners do better with less [38]. One possibility is that ChatGPT may be intentionally providing more structured guidance to learners who make errors, recognizing the need for additional support.

Delayed feedback best mirrors real-life clinical scenarios where immediate feedback is not always available to redirect incorrect thought processes. This may be particularly important for learners intent on practicing in smaller community settings where there are fewer opportunities for collaboration. However, this independence is a major developmental process in residency training, and thus, may be above the training level for which these simulations were targeted. Additionally, health care systems are interdisciplinary and collaborative. Expecting completely delayed feedback on patient care may be unrealistic considering shift hand-offs, pharmacists, nurses, medical record alerts, and other checks. Furthermore, the difference between delayed feedback for correct and incorrect treatments may offer different learning experiences for learners of different skills; it is unclear whether these different learning experiences would be inequitable or if they would offer new opportunities for learning (ie, learners who know the proper treatment are ready to practice independently, whereas learners still mastering treatment protocols are not forced to practice independently when they are not ready).

Another consideration is that different prompt engineering could resolve this variability. A prompt could include the instructions to "not give feedback until the end of the scenario, and instead of giving feedback on incorrect answers, make the patient more unstable as a sign that the choice was incorrect, then explain this incorrect choice in the summary feedback at the end of the simulation." Alternatively, because ChatGPT's algorithm learns from user inputs, learners running simulations in the same chat window could give ChatGPT feedback as they run simulations to hone in on their desired parameters, including desired feedback mechanisms. Future studies will need to evaluate differential prompt engineering, learner experiences, and learner preferences.

XSL•FO
RenderX

Of note, ChatGPT also gave hints to guide the tester after incorrect inputs. For example, ChatGPT could change from asking about the "next step in management" to asking about the "next step in controlling bronchoconstriction," thus nudging the tester toward administering a bronchodilator. This feature, as well as not delaying feedback after incorrect inputs, might be ChatGPT exhibiting flow, the educational and gaming principle that tasks should be made difficult enough to optimally challenge the learner, but not so difficult that the learner quits [39]. Again, further studies will need to evaluate learner experiences and preferences.

Comprehensive feedback was not significantly different between correct and incorrect simulations, with an overall rate of 78% (282/360). This high performance is encouraging, as comprehensive feedback offers more detailed learning opportunities and touchpoints for self-study. While this study did not explore the possibility of learners requesting expanded feedback, doing so could potentially increase the rate of comprehensive feedback to close to 100%, thereby enhancing the educational value of the simulations. ChatGPT concluded autonomously at similar rates to comprehensive feedback, which is also encouraging. Autonomous conclusion prevents learners from becoming "trapped" in the simulation, primed by the continuing simulation to believe there is more treatment needed even when the patient is stable and treatment should conclude. While autonomous conclusion at 100% would be best for efficiency and ease of use, forcing a learner to decide when to end the simulation can also mimic the testing style of oral examinations used by some medical specialties for board certification. This is a skill expected of advanced learners who have completed residency, but earlier learners could be easily prompted to end the simulation when they feel they have completed their treatments. Proper instructions and framing could address this problem from the prompt generator's perspective.

ChatGPT-4 was not significantly more likely to conclude simulations autonomously and provide comprehensive feedback when feedback was delayed. This suggests that ChatGPT-4 does not necessarily perform desirable actions in clusters, meaning that learners who make mistakes are not at risk of losing some of the simulation's desirable characteristics.

Future Work

This study focused on a single mistake in the treatment workflow, but future studies should look at performance with multiple mistakes. ChatGPT-4's ability to display desirable simulation parameters in the presence of multiple mistakes would be crucial for it to be a viable educational tool; real learners may make multiple mistakes of varying severity, and ChatGPT-4 would need to handle this. As is, the feasibility of using this tool given ChatGPT's varied performance on advanced parameters is still unclear. Medical educators may be hesitant to use a tool that is not completely standardized and risks inequitable learning outcomes, while others may welcome a tool that seemingly adapts to the learner's level. Quantified student learning gains after simulation use are thus needed for educators to make informed decisions in their classrooms and simulation centers and are the subject of ongoing work. However, with the aforementioned testing, we foresee the following potential implementation: educators would use the most advanced version of ChatGPT available and tailor the prompt provided here to their students' learning goals. Simulations would focus on simple diagnostic or therapeutic problems with straightforward algorithms. Simulations would be paired with robust pre- or postsimulation didactics to reinforce simulation concepts and not be the sole method of instruction. Ideally, simulations would be overseen by an experienced educator capable of monitoring simulations for accuracy.

Future work should focus on measuring student performance after using simulations to assess educational utility, trialing simulations on different LLMs to assess if the best LLM exists for this purpose, and trialing LLM performance on increasingly complex, nonalgorithmic simulations with multiple wrong answers. All of these are important for this tool's external validity.

Limitations

This study has several limitations which should be addressed. First, the simulations were limited to a single clinical scenario—acute asthma exacerbation—which may not be representative of ChatGPT-4's performance across a wider range of medical conditions. More complex scenarios reflecting the breadth of clinical practice and training level cannot be assumed to function as the simulations provided here. Our simulations were also run on only 1 prompt to limit confounders, but different prompts may produce varied results. Additionally, the study only compared correct treatment and 1 type of incorrect treatment, excluding a broader spectrum of clinical inaccuracies that may occur in real-world educational settings. Furthermore, simulations were run on 3 different ChatGPT-4 accounts and across different browsers, which, although necessary due to token limits, might introduce variability in the model's responses. Finally, the results of this study are contingent on the current application program interface settings established by ChatGPT and each new update of the LLM can potentially impact reproducibility. However, this does not detract from this study's conclusions given its intent to establish the methodology for evaluating simulations, rather than being a definitive recommendation for implementing LLM into preclinical medical education.

Conclusions

The 9-part model described here can be useful in evaluating ChatGPT simulations as a learning tool in accordance with established principles of medical simulation and multimedia design in educational technology. The use of this model can help standardize the evaluation of LLM simulation research. As an example use of this model, ChatGPT performed well in creating practice clinical scenarios that adhered to 3 simulation parameter categories. It performed excellently on 2 primary features, medical accuracy and basic simulation parameters, and reasonably well on advanced parameters. Variation in simulation characteristics based on the accuracy of learner inputs is a point of concern. This tool's impact on student learning is an important next step to explore, but these simulations

XSL•FO
RenderX

demonstrate ChatGPT's potential to be a reliable educational tool for the appropriate preclinical-to-clinical learning level.

Data Availability

All data generated or analyzed during this study are included in this published article and in Multimedia Appendix 1.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary files.
[XLSX File, 62 KB - formative_v9i1e66478_app1.xlsx]

References

1. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 8;9:e45312. [doi: 10.2196/45312] [Medline: 36753318]
2. Nakao T, Miki S, Nakamura Y, et al. Capability of GPT-4V(ision) in the Japanese National Medical Licensing Examination: evaluation study. JMIR Med Educ 2024 Mar 12;10:e54393. [doi: 10.2196/54393] [Medline: 38470459]
3. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digital Health 2023 Feb;2(2):e0000198. [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
4. Ali R, Tang OY, Connolly ID, et al. Performance of ChatGPT and GPT-4 on Neurosurgery written board examinations. Neurosurgery 2023 Dec 1;93(6):1353-1365. [doi: 10.1227/neu.0000000000002632] [Medline: 37581444]
5. Brin D, Sorin V, Vaid A, et al. Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments. Sci Rep 2023 Oct 1;13(1):16492. [doi: 10.1038/s41598-023-43436-9] [Medline: 37779171]
6. Mohammad B, Supti T, Alzubaidi M, et al. The pros and cons of using ChatGPT in medical education: a scoping review. Stud Health Technol Inform 2023 Jun 29;305:644-647. [doi: 10.3233/SHTI230580] [Medline: 37387114]
7. Boscardin CK, Gin B, Golde PB, Hauer KE. ChatGPT and generative artificial intelligence for medical education: potential impact and opportunity. Acad Med 2024 Jan 1;99(1):22-27. [doi: 10.1097/ACM.0000000000005439] [Medline: 37651677]
8. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT—reshaping medical education and clinical management. Pak J Med Sci 2023;39(2):605-607. [doi: 10.12669/pjms.39.2.7653] [Medline: 36950398]
9. Scherr R, Halaseh FF, Spina A, Andalib S, Rivera R. ChatGPT interactive medical simulations for early clinical education: case study. JMIR Med Educ 2023 Nov 10;9:e49877. [doi: 10.2196/49877] [Medline: 37948112]
10. Halaseh FF, Yang JS, Danza CN, Halaseh R, Spiegelman L. ChatGPT's role in improving education among patients seeking emergency medical treatment. West J Emerg Med 2024 Sep;25(5):845-855. [doi: 10.5811/westjem.18650] [Medline: 39319818]
11. Wu Y, Zheng Y, Feng B, Yang Y, Kang K, Zhao A. Embracing ChatGPT for medical education: exploring its impact on doctors and medical students. JMIR Med Educ 2024 Apr 10;10:e52483. [doi: 10.2196/52483] [Medline: 38598263]
12. Araujo SM, Cruz-Correia R. Incorporating ChatGPT in medical informatics education: mixed methods study on student perceptions and experiential integration proposals. JMIR Med Educ 2024 Mar 20;10:e51151. [doi: 10.2196/51151] [Medline: 38506920]
13. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023 Mar 19;11(6):887. [doi: 10.3390/healthcare11060887] [Medline: 36981544]
14. Alkhaaldi SMI, Kassab CH, Dimassi Z, et al. Medical student experiences and perceptions of chatgpt and artificial intelligence: cross-sectional study. JMIR Med Educ 2023 Dec 22;9:e51302. [doi: 10.2196/51302] [Medline: 38133911]
15. Skryd A, Lawrence K. ChatGPT as a tool for medical education and clinical decision-making on the wards: case study. JMIR Form Res 2024 May 8;8:e51346. [doi: 10.2196/51346] [Medline: 38717811]
16. Kim HW, Hong JW, Nam EJ, Kim KY, Kim JH, Kang JI. Medical students' perceived stress and perceptions regarding clinical clerkship during the COVID-19 pandemic. PLoS ONE 2022;17(10):e0277059. [doi: 10.1371/journal.pone.0277059] [Medline: 36315569]
17. Heston TF, Khun C. Prompt engineering in medical education. IME 2023;2(3):198-205. [doi: 10.3390/ime2030019]
18. Musallam E, Alhaj Ali A, Alkhafaji M. OpenAI's ChatGPT clinical simulation: an innovative teaching strategy for clinical nursing education. Nurse Educ 2024;49(6):E361-E362. [doi: 10.1097/NNE.0000000000001657] [Medline: 38728116]
19. Lucas HC, Upperman JS, Robinson JR. A systematic review of large language models and their implications in medical education. Med Educ 2024 Nov;58(11):1276-1285. [doi: 10.1111/medu.15402] [Medline: 38639098]

XSL•FO
RenderX

20. Jeyaraman M, K SP, Jeyaraman N, Nallakumarasamy A, Yadav S, Bondili SK. ChatGPT in medical education and research: a boon or a bane? Cureus 2023 Aug;15(8):e44316. [doi: 10.7759/cureus.44316] [Medline: 37779749]
21. van de Ridder JMM, Shoja MM, Rajput V. Finding the place of ChatGPT in medical education. Acad Med 2023 Aug 1;98(8):867. [doi: 10.1097/ACM.0000000000005254] [Medline: 37162206]
22. Clark RC. Multimedia learning in e-courses. In: Mayer RE, Fiorella L, editors. The Cambridge Handbook of Multimedia Learning, 2nd edition: Cambridge University Press; 2014:842-882. [doi: 10.1017/CBO9781139547369.040]
23. Wylie R, Chi MTH. The self-explanation principle in multimedia learning. In: Mayer RE, Fiorella L, editors. The Cambridge Handbook of Multimedia Learning, 2nd edition: Cambridge University Press; 2014:413-432. [doi: 10.1017/CBO9781139547369.021]
24. Luetner D, Schmeck A. The generative activity principle in multimedia learning. In: Mayer RE, Fiorella L, editors. The Cambridge Handbook of Multimedia Learning, 2nd edition: Cambridge University Press; 2014:433-448. [doi: 10.1017/CBO9781139547369.022]
25. Mayer RE. Cognitive theory of multimedia learning. In: Mayer RE, Fiorella L, editors. The Cambridge Handbook of Multimedia Learning, 2nd edition: Cambridge University Press; 2014:43-71. [doi: 10.1017/CBO9781139547369.005]
26. Low R, Sweller J. The modality principle in multimedia learning. In: Mayer RE, Fiorella L, editors. The Cambridge Handbook of Multimedia Learning, 2nd edition: Cambridge University Press; 2014:227-246. [doi: 10.1017/CBO9781139547369.012]
27. Plass JL, Schwartz RN. Multimedia learning with simulations. In: Mayer RE, Fiorella L, editors. The Cambridge Handbook of Multimedia Learning, 2nd edition: Cambridge University Press; 2014:729-761. [doi: 10.1017/CBO9781139547369.036]
28. Mayer RE, Fiorella L. Principles for reducing extraneous processing in multimedia learning. In: The Cambridge Handbook of Multimedia Learning, 2nd edition: Cambridge University Press; 2014:279-315. [doi: 10.1017/CBO9781139547369.015]
29. Johnson CI, Priest HA. The feedback principle in multimedia learning. In: Mayer RE, Fiorella L, editors. The Cambridge Handbook of Multimedia Learning, 2nd edition: Cambridge University Press; 2014:449-463. [doi: 10.1017/CBO9781139547369.023]
30. Stefanidis D, Cook D, Kalantar-Motamedi SM, et al. Society for simulation in healthcare guidelines for simulation training. Simul Healthc 2024 Jan 1;19(1S):S4-S22. [doi: 10.1097/SIH.0000000000000776] [Medline: 38240614]
31. van Gog T. The signaling (or cueing) principle in multimedia learning. In: Mayer RE, Fiorella L, editors. The Cambridge Handbook of Multimedia Learning, 2nd edition: Cambridge University Press; 2014:263-278. [doi: 10.1017/CBO9781139547369.014]
32. Feldon DF, Jeong S, Clark RE. Fifteen common but questionable principles of multimedia learning. In: Mayer RE, Fiorella L, editors. The Cambridge Handbook of Multimedia Learning, 3rd edition: Cambridge University Press; 2021:25-40. [doi: 10.1017/9781108894333.005]
33. Dabbs W, Bradley MH, Chamberlin SM. Acute asthma exacerbations: management strategies. Am Fam Physician 2024 Jan;109(1):43-50. [Medline: 38227870]
34. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med 2023 Aug;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
35. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: implications in scientific writing. Cureus 2023 Feb;15(2):e35179. [doi: 10.7759/cureus.35179] [Medline: 36811129]
36. Hattie J, Timperley H. The power of feedback. Rev Educ Res 2007 Mar;77(1):81-112. [doi: 10.3102/003465430298487]
37. Chamberland M, Setrakian J, St-Onge C, Bergeron L, Mamede S, Schmidt HG. Does providing the correct diagnosis as feedback after self-explanation improve medical students diagnostic performance? BMC Med Educ 2019 Jun 11;19(1):194. [doi: 10.1186/s12909-019-1638-3] [Medline: 31185971]
38. Kalyuga S. The expertise reversal principle in multimedia learning. In: Mayer RE, Fiorella L, editors. The Cambridge Handbook of Multimedia Learning, 2nd edition: Cambridge University Press; 2014:576-597. [doi: 10.1017/CBO9781139547369.028]
39. Csikszentmihalyi M, Abuhamdeh S, Nakamura J. In: Csikszentmihalyi M, editor. Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihaly: Springer; 2014:227-238. [doi: 10.1007/978-94-017-9088-8_15]

Abbreviations

AI: artificial intelligence
LLM: large language model
RISE: role, input, steps, expectations

XSL•FO
RenderX

Edited by A Mavragani; submitted 14.09.24; peer-reviewed by C Zickler, M Barakat, N Chien, S Patel, W Li; revised version received 30.01.25; accepted 31.01.25; published 27.02.25.

Please cite as:
Scherr R, Spina A, Dao A, Andalib S, Halaseh FF, Blair S, Wiechmann W, Rivera R
Novel Evaluation Metric and Quantified Performance of ChatGPT-4 Patient Management Simulations for Early Clinical Education: Experimental Study
JMIR Form Res 2025;9:e66478
URL: https://formative.jmir.org/2025/1/e66478
doi:10.2196/66478

© Riley Scherr, Aidin Spina, Allen Dao, Saman Andalib, Faris F Halaseh, Sarah Blair, Warren Wiechmann, Ronald Rivera. Originally published in JMIR Formative Research (https://formative.jmir.org), 27.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Capturing Everyday Parental Feeding Practices and Eating Behaviors of 3- to 5-Year-Old Children With Avid Eating Behavior: Ecological Momentary Assessment Feasibility and Acceptability Study

Abigail Pickard[1,2], PhD; Katie Edwards[1,3], PhD; Claire Farrow[1], PhD; Emma Haycraft[4], PhD; Jacqueline Blissett[1], PhD

[1]School of Psychology and Institute of Health and Neurodevelopment, Aston University, Birmingham, United Kingdom

[2]Department of Clinical Psychology, School of Health in Social Science, University of Edinburgh, Wilkie Building, Edinburgh, United Kingdom

[3]School of Psychology, University of Birmingham, Birmingham, United Kingdom

[4]School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, United Kingdom

Corresponding Author:
Abigail Pickard, PhD
School of Psychology and Institute of Health and Neurodevelopment, Aston University, Birmingham, United Kingdom

Abstract

Background: The wide use of smartphones offers large-scale opportunities for real-time data collection methods such as ecological momentary assessment (EMA) to assess how fluctuations in contextual and psychosocial factors influence parents' feeding practices and feeding goals, particularly when feeding children with high food approaches.

Objective: The main objectives of this study were to (1) assess parents/caregivers' compliance with EMA procedures administered through a smartphone app and (2) estimate the criterion validity of the EMA to capture children's eating occasions and parents' feeding practices. Participant adherence, technological challenges, and data quality were used to provide an overview of the real-time dynamics of parental mood, feeding goals, and contextual factors during eating occasions.

Methods: Parents in the United Kingdom with a child aged 3 to 5 years who exhibit avid eating behavior were invited to participate in a 10-day EMA study using a smartphone app. Of the 312 invited participants, 122 (39%) parents initiated the EMA study, of which 118 (96.7%) completed the full EMA period and the follow-up feasibility and acceptability survey.

Results: Of those parents who completed the EMA study, 104 (87.4%) parents provided at least 7 "full" days of data (2 signal surveys and 1 event survey), despite 51 parents (43.2%) experiencing technical difficulties. The parents received notifications for morning surveys (69.9% response rate), 3 daily mood surveys (78.7% response rate), and an end-of-day survey (84.6% response rate) on each of the 10 days. Over the EMA period, a total of 2524 child eating/food request surveys were self-initiated by the participants on their smartphones, an average of 2.1 times per day per parent (SD 0.18; min=1.7, max=2.3). The majority of parents felt that the surveys made them more aware of their feelings (105/118, 89%) and activities (93/118, 79%). The frequency of daily food requests estimated by parents at baseline was significantly correlated with the frequency of food requests reported daily during the EMA period (r=0.483, P<.001). However, the number of daily food requests per day estimated at baseline (mean 4.5, SD 1.5) was significantly higher than the number of food requests reported per day during the EMA period (mean 3.7, SD 1.1), (t_{116}=18.8, P<.001).

Conclusions: This paper demonstrates the feasibility of employing EMA to investigate the intricate interplay between parental mood, feeding goals, contextual factors, and feeding practices with children exhibiting an avid eating behavior profile. However, the use of EMA needs to be carefully developed and tested with parents' involvement to ensure successful data collection.

International Registered Report Identifier (IRRID): RR2-10.2196/55193

(JMIR Form Res 2025;9:e66807) doi:10.2196/66807

KEYWORDS

pediatric; paediatric; child; child eating; parent feeding; parent; ecological momentary assessment; mHealth; mobile health; mobile app; application; smartphone; digital; digital health; digital technology; digital intervention

XSL•FO
RenderX

Introduction

Children with an avid eating behavior profile pose unique challenges for parents, as this group of children has been shown to make greater requests for food, often in response to emotions and food cues in their environment rather than due to internal hunger signals [1,2]. Avid eating behavior in children is characterized by an intense and persistent interest in food, often leading to distinct challenges in managing feeding interactions [1,2]. Understanding the real-time dynamics of these interactions is crucial for developing targeted interventions aimed at promoting healthy eating behaviors in children.

However, despite evidence that both parental feeding practices and children's eating behaviors are dynamic, much research has investigated the association between the 2 constructs cross-sectionally, using questionnaires, interviews, and one-off mealtime observations [3]. Such research only allows between-subject investigation, neglecting the within-subject variation in both feeding practices and eating behavior. These limitations have spurred more recent research using ecological momentary assessment (EMA; also known as experience sampling methods).

EMA is a form of mobile health technology that allows the collection of comprehensive data on daily life, as it is directly experienced by an individual, by asking them to provide responses about their emotions, context, and behavior at several random time points each day over several days [4]. Commonly, EMA employs both signal-contingent surveys, which are scheduled by the researcher to be push-notified at specific times, and event-contingent surveys, which are to be initiated by the participant following a specific event or action [5]. Due to advancements in technology and the widespread use of smartphones, EMA methods have become more popular in eating behavior research [4,6]. Berge et al [7] conducted an EMA study examining feeding practices in a sample of parents (N=150) of a 5 - 7-year-old child with at least one other sibling between 2 and 12 years old. Their EMA study had very high compliance; 100% of participants (N=118) completed all 8 days of EMA data collection and met the minimum requirement per day (ie, 2 signal contingent, 1 event contingent, and 1 end-of-day contingent), with an average of 7.4 surveys per day [7].

A more recent study conducted by Loth et al [8] used EMA among a sample of 120 young adults with at least 1 child aged 2 - 5 years. Of the 118 participants who registered for their EMA study, 116 (98.3%) participants completed at least one eating occasion survey, 100 participants completed one "full" day (84.7%), and 80 (67.8%) participants completed "full" days for the complete 10-day EMA period (a "full" day was defined as providing data for 2 eating occasions, 2 random prompts, and 1 end of day prompt). Finally, a study examining parents' feeding practices as a predictor of children's physical activity and fruit and vegetable intake used an EMA design with 22 mothers, fathers, and 8 - 12-year-old-child triads completing a 7-day assessment period [9]. The response rates for the study period were high with 85%, 80%, and 85.9% responses to the survey notifications completed by the mothers, fathers, and children, respectively.

While these studies have provided important evidence for the use of EMA to investigate parental feeding interactions with children, there remains caution and concern within the field about the feasibility of using such methods with samples in which feeding and eating interactions may be more challenging. For example, it is unknown whether such methods are adequate to capture real-time information for parents of children with avid eating tendencies, who have previously been shown to have a higher risk of food insecurity [1] and where children are more likely to show an increased frequency of eating and food requests [2]. Therefore, this study assessed the feasibility, acceptability, and criterion validity of using EMA to explore the participation and everyday experiences of parenting children with avid eating behavior aged 3 to 5 years old across a 10-day sampling period.

Assessing the feasibility of such methods is not only important to guide future EMA research but also to ensure that the findings of this study are reliable. The study also aimed to determine the usefulness of using a commercially licensed software application on Android and iOS devices as a tool for EMA.

Methods

Overview

This EMA study was part of a larger program of research, the APPETItE project, which examines feeding and eating in preschool children with avid eating behavior to inform future intervention design and efficacy. An avid eating behavior profile in preschool children was identified and defined in an earlier APPETItE study using latent profile analysis (LPA) [1]. The LPA used the 8 subscales of the Children's Eating Behavior Questionnaire (CEBQ) [10] as index variables to create holistic yet distinct eating profiles for UK children aged 3 to 5 years.

Ethical Considerations

Recruitment, enrolment, and data collection for the EMA study took place between October 2023 and April 2024. Ethical approval was granted by the Aston University Health and Life Sciences Research Ethics Committee (HLS21079). All participants provided informed consent and were made aware that they had the right to opt out of the study at any time. For a comprehensive overview of the EMA study protocol and the measures included in each data collection wave, refer to Edwards et al [11]. The findings of the full study will be published at a later date.

Participants

Only parents of children aged 3 to 5 years who were assigned to the avid eating profile based on the screening questionnaire were eligible to participate in this EMA study. Eligibility criteria also included English-speaking primary caregivers from the United Kingdom who were responsible for feeding their child for more than half the time when their child was at home. Parents whose child was autistic, or had severe learning disabilities, or a chronic illness that directly influenced their dietary requirements and eating habits were not eligible to participate. A total of 312 parents met the eligibility criteria for the study and were invited to participate via email.

XSL•FO
RenderX

Recruitment

From the original LPA defining the avid eating profile, 101 parents with a child in the eligible age range (3 - 5 y) and assigned to the avid eating profile were invited to participate. The profile solution identified in the LPA study [1] was used to classify additional children whose parents expressed interest in the EMA study and submitted CEBQ [10] data through the recruitment platform Prolific. Using this profiling method, 168 children were identified as belonging to the avid eating profile. A further 43 parents who had completed the CEBQ screening questionnaire for a separate lab-based study as part of the APPETItE project were also identified as having eligible children and were invited to take part in the EMA study. If the parents completed the initial interest survey, they were then contacted via email with further instructions and a demo video [12] explaining how to complete the EMA study. If parents experienced any technical difficulties, had any questions, or wanted to go through the study procedures, they were able to contact the research team by email.

Data Collection Procedures

Participation in this study was remote; surveys were administered through a smartphone app, ExpiWell, which was downloaded directly to parents' smartphones. ExpiWell is a cloud-based platform designed for researchers to conduct experience sampling methods and EMA studies. It enables researchers to create, schedule, and distribute surveys to participants via free mobile applications available on iOS and Android devices. Parents who did not own a smartphone were informed that they could request one from the research team to use for the study period, but all parents used their own devices.

Parents completed a baseline questionnaire, 10 days of EMA, and an end-of-study questionnaire. The 10-day EMA period included both signal surveys (push notifications sent to the participant's smartphone) and event surveys (initiated by the participant following an eating occasion or request for food by the child). Parents were notified to complete 1 start-of-day survey, 3 signal-contingent examining mood and emotions, 1 end-of-day survey, and event surveys activated by the parent when their child ate or asked for food. Participants were able to schedule the morning and evening surveys to suit their typical routine. The full protocol and procedure are reported elsewhere [11].

Parents received a £100 (approximately US $126) shopping voucher if at least 8 "full" days of surveys were complete. Based on previous research, the criteria for one full day of EMA data included the completion of 2 signal-contingent surveys (morning, mood, or end-of-day) and one event-contingent (food) survey [8]. All parents who completed 10 days of EMA were entered into a prize draw to win an additional £100 (approximately US $126) shopping voucher. If participants withdrew prematurely from this study, their time was reimbursed on a pro-rated basis of £10 (approximately US $12.60) per complete day.

Measures

To evaluate the feasibility and criterion validity of using EMA to capture the eating and feeding experiences of parents with a child displaying avid eating, several questions were incorporated into the survey period. The baseline survey asked parents to report on average how many times their child requested food, and ate food, both on weekdays and weekends. In the end-of-day surveys, parents were also asked to report how many times the child had eaten or requested food on that specific day. These data were collected to determine how many of the average eating and food request occasions were being reported in the EMA period.

Feedback Questionnaire

A self-constructed feedback questionnaire, based on the items used by de Vries et al [13] and "The Open Handbook of Experience Sampling Methodology" [14], was used to gather participants' perspectives on the EMA procedure. Participants responded to 16 items about their experience of completing the study, the influence the EMA study had on their daily lives, the awareness the EMA study gave them of their actions and feelings, and the burden of completing the EMA study. Response options were based on a 5-point scale where 1="not at all," 2="a little," 3="moderately," 4="quite a bit," and 5="extremely." The final item asked about whether the participant had experienced any technical problems during the EMA period with response options of 1="yes," and 0="no." The full study, including a list of all the items used, is available on the Open Science Framework [15].

Data Analysis

Descriptives of the survey completion rates, forgotten reporting, withdrawal rates, nonresponse rates, survey duration, and any reported technical issues were examined to determine EMA feasibility. Criterion validity of EMA-reported frequency of eating occasions and food requests was examined by comparison with the initially reported frequencies and end-of-day frequencies using paired *t* tests and Pearson correlations. Study acceptability was assessed with descriptive statistics of the feedback responses at the end of the study.

Results

Demographics

Parents had a mean age of 34.97 years (SD 5.3; min=23.6, max=55.3) and children had a mean age of 52.57 months (SD 10.5; min=36.8, max=71.4). An overview of the parents' demographic background and child sex is provided in Table 1.

XSL•FO
RenderX

Table . Demographic details of the sample (N=118).

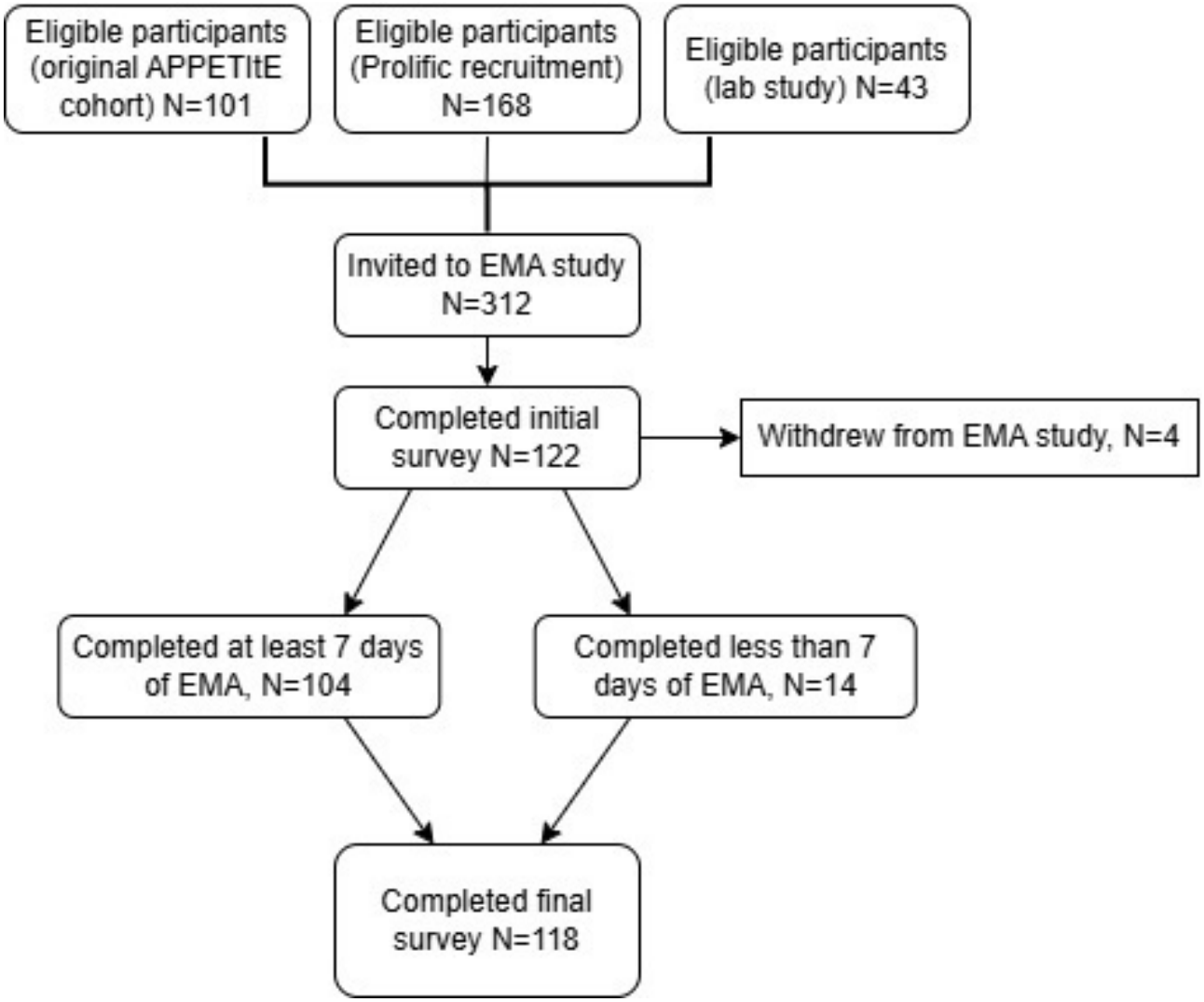
Demographics	Values, n (%)
Child sex	
Male	55 (46.6)
Female	63 (53.4)
Parent sex	
Male	26 (22)
Female	92 (78)
Parent ethnicity	
Asian	13 (11)
Black	10 (8.5)
White	93 (78.8)
Other	2 (1.7)
Education	
Degree	70 (59.3)
No degree	48 (40.7)
Parent or family structure	
Single parent	12 (10.2)
Dual parent	96 (81.4)
Cohabiting (not in relationship)	6 (5.1)
Extended family	4 (3.4)
Working status	
Unemployed	25 (21.2)
Working part-time (between 8 and 29 hours per week)	36 (30.5)
Working full-time (30 hours or more per week)	57 (48.3)
Adequacy of income	
Living comfortably	54 (45.8)
Managing	47 (39.8)
Finding it difficult	17 (14.4)
Household food security	
High or marginal food security	78 (66.1)
Low food security	20 (16.9)
Very low food security	20 (16.9)

EMA Feasibility

Of the 312 parents invited to participate, 122 (39.1%) parents downloaded the mobile application and completed the initial survey. Of those 122 participants, 118 (96.7%) participants completed the EMA period. Four parents withdrew from the EMA study due to either technical difficulties or time constraints. Of the remaining participants, 104 (87.4%) participants completed at least 7 days of the EMA period with an average of 58 completed surveys for the full 10 days (SD 18; min=9, max=103). Figure 1 provides an overview of the eligible sample sizes as well as completion and attrition rates at each phase of the EMA study.

XSL•FO
RenderX

Figure 1. Recruitment and retention figures for each EMA phase. EMA: ecological momentary assessment.



Over the 10-day EMA period, a total of 2524 child eating/food request surveys were self-initiated by the 118 participants on their smartphones, with 30 surveys partially completed. Parents self-initiated the request for food and eating surveys on average 2.1 times per day (SD 0.18; min=1.7, max=2.3). Of the 2524 eating and food request surveys completed at the moment, 2353 reported on an event in which the child ate food, and 171 reported on an instance in which the child requested food but did not eat.

At the end of all signal-contingent surveys and the food survey, parents were asked whether they had forgotten to report an eating occasion or a request for food from the child. On 870 occasions a parent retrospectively reported on an eating/food request that they had forgotten to report (Table 2 details the number of eating events that were reported on when the parent was reminded at the following survey).

XSL•FO
RenderX

Table . Number of retrospective eating occasions and food request reports.

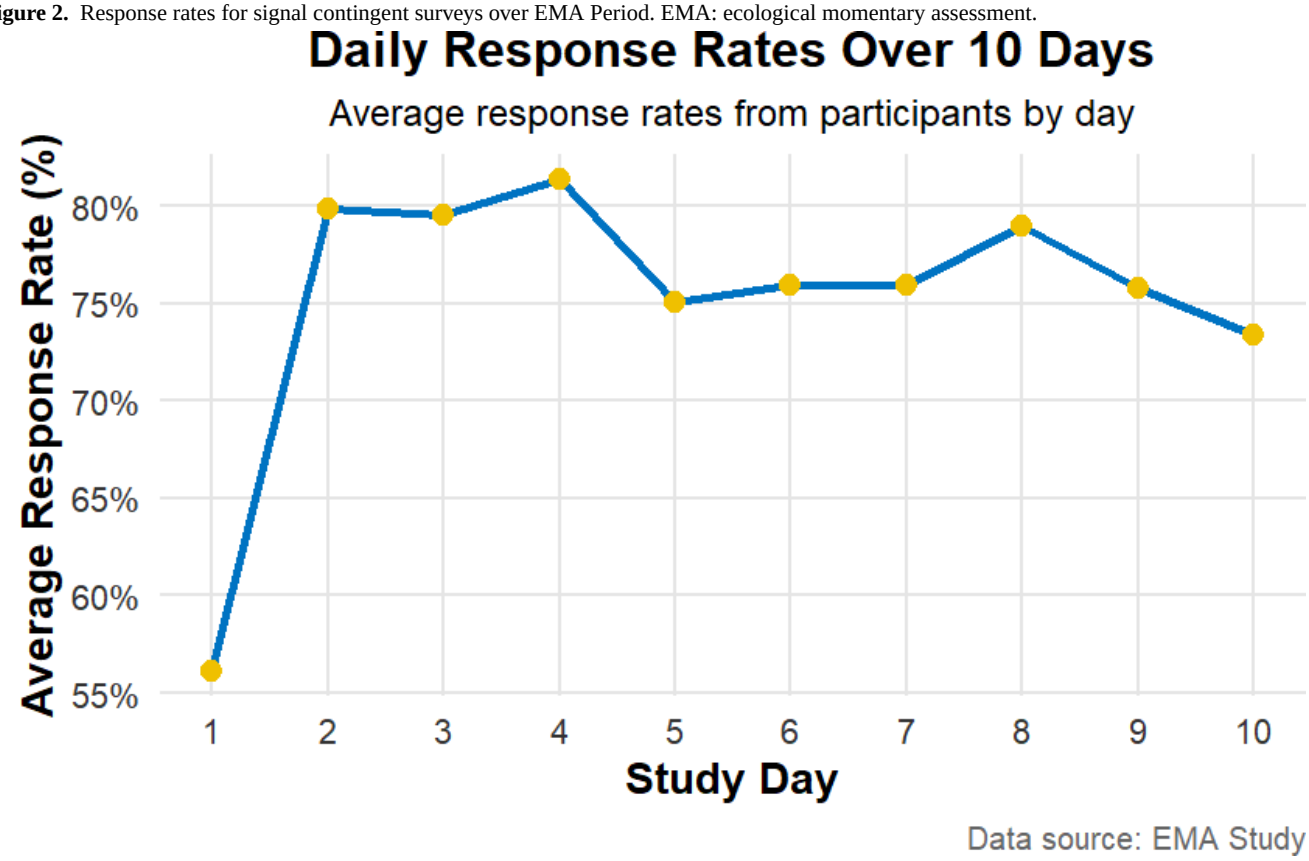
Survey	Number of responses
Morning surveys	
Forgotten eating event	111
Forgotten food request event	4
Mood surveys	
Forgotten eating event	413
Forgotten food request event	28
End-of-day surveys	
Forgotten eating event	137
Forgotten food request event	13
Food surveys	
Forgotten eating event	141
Forgotten food request event	23
Total number of forgotten events reported	870

Response rates were calculated for the signal-contingent surveys, which included 1 survey in the morning, 3 surveys during the day, and 1 survey in the evening. From the first notification, parents had 60 minutes to complete the surveys before the link expired. A total of 1180 push notifications were received by all the participants (10 per participant) to complete the morning surveys, 825 morning surveys were initiated by participants following the notification, with 18 incomplete responses (response rate=69.9%). The parents also received notifications for 3 signal-contingent surveys (mood surveys) at semirandom times within one of three 120-minute blocks. A possible 3540 signal-contingent mood surveys were available to the 118 participants throughout the 10-day duration (30 per participant) with a total completion of 2778 surveys and 8 incomplete responses (78.7% response rate). The signal contingent end-of-day survey was sent to each participant every evening with a potential total of 1180 responses, 998 end-of-day surveys were initiated when prompted with 4 incomplete responses (84.6% response rate).

To test whether more participants responded to the 5 prompted surveys at the beginning (eg, 2-4 d) compared with the end (eg, 8-10 d) a paired *t* test was conducted on the daily compliance rates. Day 1 was treated as a practice day due to significantly lower compliance rates as participants became accustomed to the EMA protocol (Figure 2). There was no significant difference in compliance rate for the start of the study period compared with the end of the study period (t_{118}=0.127, *P*=.90).

XSL•FO
RenderX

Figure 2. Response rates for signal contingent surveys over EMA Period. EMA: ecological momentary assessment.



A 1-way ANOVA was conducted to determine whether the working status of the parent was associated with the number of surveys completed. There was a significant difference in the number of reported eating and food request occasions ($F_{2,115}$=12.41, P=.011), with parents working part-time reporting significantly more food occasions (mean=2.5 food surveys per day) than parents working full-time (mean=1.75 food surveys per day) but not significantly more than unemployed parents (mean=2.13). There were no significant differences in survey completion for all signal-contingent surveys between parents of differing working statuses.

Table 3 displays the descriptives of survey completion duration. The average duration to complete the morning surveys was 2.76 minutes (SD 6.83 minutes). The average time taken for the signaled mood surveys was 2.09 minutes (SD 6.16 minutes). The average duration to complete the food surveys was 3.13 minutes (SD 7.87 minutes) and 2524 food surveys were self-initiated and completed.

Table . Average survey completion duration by survey type.

	Duration (MM:SS)		
	Mean (SD)	Minimum	Maximum
Signaled surveys			
Morning surveys	02:29:00 (04:00:00)	00:38:00	58:32:00
Mood surveys	01:52:00 (02:21:00)	00:27:00	39:17:00
End-of-day surveys	02:20:00 (02:15:00)	00:47:00	31:44:00
Event surveys			
Food surveys	02:47:00 (02:29:00)	00:48:00	52:47:00

Almost half of the participants experienced technical issues with the application (51/118, 43.2%). Pearson chi-square test with Yates' continuity correction revealed that Android users were more likely to experience technical issues (30.16% of iOS users experienced technical issues compared with 62.26% of Android users).

Over half of the participants (71/118, 60%) reported that they did not make any mistakes when completing the surveys.

Criterion Validity

To determine the criterion validity of using EMA to accurately capture eating occasions and food requests of parents from children with avid eating, several additional measures of the frequency of children's eating and requests for food were taken.

Parents were asked at baseline to estimate how frequently their child ate on a typical day, during the EMA period they were also asked to report how many times their child had eaten that

XSL•FO
RenderX

day in the evening survey. A paired-samples *t* test and paired-samples correlations were then conducted to determine whether the scores were representative of what was reported by parents. Although the parents' baseline reports of eating frequency were correlated with daily reported eating frequency (r=0.365, P<.001), eating frequency was significantly higher in the baseline reports compared with the daily reports (t_{116}=5.78, P<.001). Parents reported greater eating frequency in the baseline survey (mean 4.5, SD 1.5) compared with the frequency that they reported in the end-of-day surveys (mean 3.7, SD 1.1).

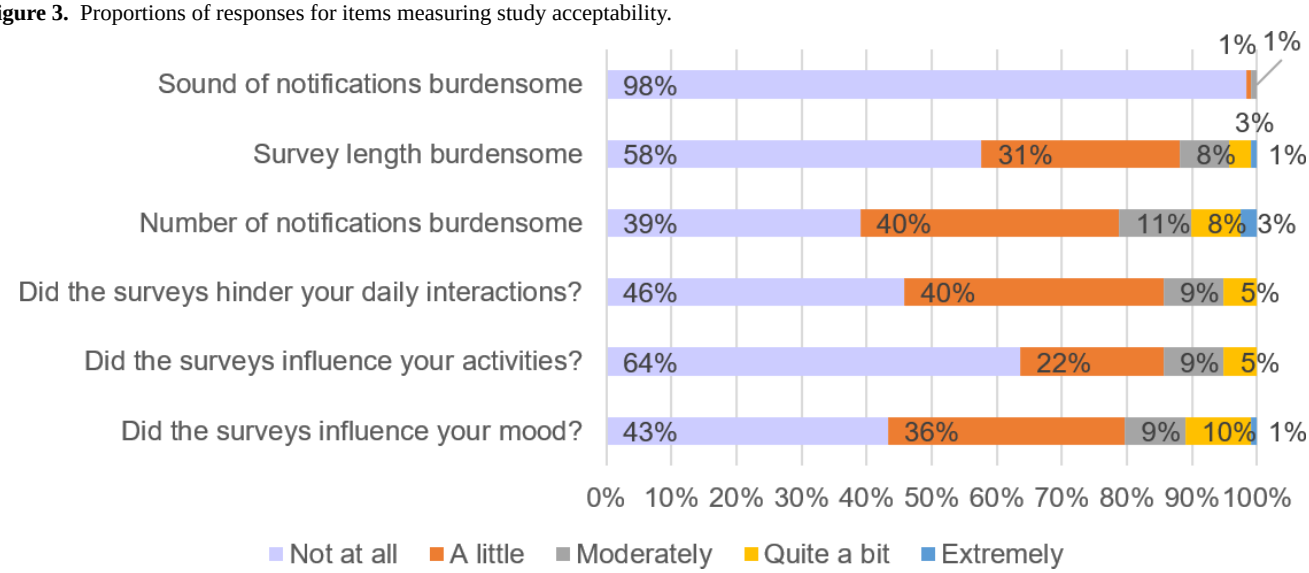
Pairwise *t* testing showed that the approximate frequency of food requests per day reported at baseline was significantly correlated with the frequency of food requests reported daily during the EMA period (r=0.483, P<.001). However, pairwise *t* testing showed that the number of food requests per day at baseline was significantly higher (mean 5.6, SD 2.4) than the number of food requests reported per day during the EMA period (mean 1.9, SD 1.1; t_{116}=18.8; and P<.001). The number of daily food requests and eating occasions reported by parents in the baseline survey was not significantly correlated with the number of daily food requests and eating occasions reported by parents in the EMA period (r=0.121, P=.097). Furthermore, the mean number of daily eating/request events at baseline (mean 10.1, SD 3.5) was significantly higher than the number of daily eating/request events during the EMA period (mean 2.04, SD 1.2; t_{116}=24.9, and P=.001). The frequency of the daily eating/request events reported in the end-of-day survey was significantly correlated with the number of reports for in-the-moment eating/request events (r=0.244, P=.004). However, reports of average eating and food requests at the end of the day (mean 5.6, SD 1.8) were significantly higher than the average frequency of in-the-moment reporting of eating/food requests (mean 2.06, SD 1.2; t_{116}=20.2; and P<.001).

Study Acceptability

To determine whether the questions included in the EMA study were acceptable for measuring eating and feeding interactions with children with avid eating behavior, the parents were asked to rate the degree to which the questions included in the survey accurately reflected their experiences. The majority of the participants felt that the EMA surveys and questions accurately reflected their experience of a parent feeding a 3 - 5-year-old child with avid eating behavior, 20% (n=24) said extremely well, 44% (n=52) said quite a bit, and 30% (n=35) said moderately well.

Figure 3 presents participants' feedback on the EMA period. Nearly all participants reported that they were not burdened by the sound of the study notifications (116/118, 98%). Although parents were asked to keep their devices on loud, it was at their discretion whether they muted the notifications. Over half of the participants were not burdened by the length of the surveys (68/118, 58%). Over half of the participants considered the number of notifications at least a "little" burdensome (73/118, 62%), but 39% (n=46) did not consider the frequency of notifications a burden. Over half of parents (64/118, 54%) found that the surveys hindered their daily interactions a little to quite a bit. However, (76/118, 64%) reported that the surveys did not influence their activities at all and (51/118, 43%) stated that the surveys did not influence their mood.

Figure 3. Proportions of responses for items measuring study acceptability.



The majority of parents (105/118, 89%) reported that filling in the surveys made them more aware of their feelings and this was experienced as pleasant by 55% (58/105), neutral by 37% (39/105), and unpleasant by 8% (8/105) of parents. A total of 79% of parents (93/118) felt they were made more aware of their activities through completing the EMA study, of which 58 parents experienced this as pleasant, 34 parents experienced this as neutral, and 1 parent experienced this as unpleasant.

Parents were also asked in the signal-contingent end-of-day survey to rate how difficult it was to fill out the surveys that day, with 1="not at all" and 5="extremely". After the scores had been aggregated by participants the mean level of difficulty was 2 (equivalent to "a little") (SD 0.8; min=1, max=5).

XSL•FO
RenderX

Discussion

Principal Findings

This study aimed to assess the feasibility, acceptability, and criterion validity of using EMA to investigate parental feeding practices and their dynamic interactions with children exhibiting avid eating behaviors. The results demonstrate that EMA is a feasible and acceptable method for capturing real-time data on child eating occasions and food requests, but it also highlights some discrepancies in the data collected and usability issues specific to parents with young children.

Feasibility of, and Compliance With, EMA Methodology

Despite the monetary incentive for the study, the recruitment rate to the study was lower than expected with only 122 (39.1%) of contacted parents initiating the study. However, the study found a high completion rate of 96.7% (n=118) for the EMA period, suggesting that the method was generally feasible for the participating parents. Most parents (104/118, 87.4%) completed at least 7 days of the 10-day EMA period, with an average of 58 surveys completed per participant over the 10 days, which is above the minimum requirement of 2 signal-contingent surveys and 1 event-contingent survey per day. The overall response rates for the signal-contingent surveys were relatively high, ranging from 69.9% to 84.6% depending on the type of survey. These findings suggest that it was generally feasible for the participants to engage with the EMA protocol, despite the demands of responding to notifications to complete signal-contingent surveys as well as self-initiating surveys multiple times per day. However, working status played a role in the frequency of reported eating and food requests, with part-time working parents reporting more food occasions than full-time working parents. This finding suggests that the employment status of parents may influence the ability to engage with EMA protocols, which could have implications for future studies in families.

The inclusion of pro-rated remuneration appeared to be valuable for retaining participants and appropriately rewarding ongoing completion of study components. A small percentage of parents withdrew from the study due to technical difficulties or time constraints, and nearly half of the participants (51/118, 43.2%) experienced technical issues with the application, which could have affected data quality and study adherence. The provision of an email address through which participants could access rapid technical support from our study team ensured the retention of most participants in the study despite these challenges. Such support must be built into any future studies using EMA methods to ensure participant retention. Careful testing and consideration of the most appropriate software are crucial in future EMA research. The feedback also indicated that while the majority of parents did not find the sound of notifications burdensome, some participants perceived the frequency and timing of surveys as a hindrance to daily activities. These findings highlight the importance of carefully considering the design and user experience of the EMA app to minimize the burden on participants and reduce barriers to accurate data collection.

Criterion Validity of EMA for Capturing Feeding Interactions

While the EMA method showed promise in capturing in-the-moment data on eating occasions and food requests, discrepancies were found between the EMA reports and baseline recall data, suggesting that EMA should be validated with observational studies, given the challenges with any self-report methods. There were notable discrepancies between baseline reports of perceived frequency and EMA data of reported eating episodes: parents reported significantly fewer eating occasions and food requests during the EMA period compared with their baseline estimates. This pattern was consistent across different measures of eating frequency and food requests, suggesting a possible overestimation in baseline self-reports, or, perhaps more likely, underreporting of events during the EMA period. The finding that participants retrospectively reported missed eating or food request occasions in subsequent surveys supports the latter interpretation, suggesting some level of underreporting in the initial EMA responses.

There were 870 instances where parents, when prompted, reported missed eating or food request events that they had forgotten to log in the moment, highlighting a significant gap in real-time reporting accuracy. This retrospective reporting underscores the challenges inherent in capturing accurate eating behaviors in real-time, particularly with parents of young children. These results are consistent with previous studies indicating that retrospective prompts can improve data accuracy in EMAs [16]. The high incidence of missed reports emphasizes the importance of considering both real-time and reflective reporting methods in future research to enhance the validity and comprehensiveness of monitoring children's eating behavior [16]. Thus, whilst EMA may provide a good assessment of some eating interactions demonstrated by the wealth of data we collected; it is unlikely to capture all such interactions in the period of study.

The observed discrepancies between baseline reports and EMA data reflect the challenges associated with self-reporting in real-time contexts, particularly for parents of young children who display avid eating. If parents are unable to complete surveys at the moment due to competing demands or distractions, the more chaotic or stressful feeding interactions that we aim to capture are likely to be missed. Additionally, the high frequency of child eating behaviors and food requests, particularly among children with avid eating behaviors, may lead to reluctance in parents to consistently report these events [2]. Despite these limitations, the general correlation between EMA-reported data and baseline data indicates that EMA remains a valuable tool for capturing dynamic changes in feeding practices and children's eating behaviors. However, to mitigate these challenges, future EMA studies could incorporate design modifications, such as reducing the EMA period, providing more frequent reminders, or simplifying survey interfaces to encourage immediate and consistent reporting.

Acceptability of EMA

The study's findings indicate that our EMA protocol was generally acceptable to parents. The majority of participants reported that they were not burdened by the notification sounds

XSL•FO
RenderX

(116/118, 98%) and were not significantly affected by the length of the surveys (68/118, 58%). However, the frequency of notifications was considered burdensome by some participants (73/118, 62%), and over half (64/118, 54%) reported that the surveys hindered their daily interactions to some extent. Despite this, most parents reported that completing the EMA surveys increased their awareness of their feelings and activities, which was perceived as pleasant or neutral by the majority. These findings suggest that while EMA can be intrusive for some participants, it is generally well-tolerated and may even have positive side effects by enhancing self-awareness.

Implications for Future Research

The findings of this study have important implications for future research using EMA to investigate feeding practices and eating behaviors in children. First, the study underscores the need to further refine the EMA methodology to ensure that it is user-friendly and minimizes the burden on participants. For example, the user interface could be enhanced to reduce technical difficulties, and the frequency and timing of surveys could be further optimized based on participant feedback.

Second, the study provides evidence supporting the use of EMA to capture in-the-moment mood and context data related to parental feeding practices and children's eating behaviors. This is particularly important given the dynamic and context-dependent nature of these behaviors, which cannot be adequately captured through traditional cross-sectional methods. By employing EMA, researchers can gain a more nuanced understanding of how contextual factors, such as parental mood, feeding goals, and external pressures, influence feeding practices on a moment-to-moment basis.

Finally, this study contributes to the growing body of literature supporting the use of mobile technology for real-time data collection in behavioral research. The high compliance rates observed in this study suggest that parents are willing and able to use smartphone apps to report on their feeding practices and their children's eating behaviors, even over extended periods. This opens up new opportunities for large-scale studies that leverage mobile technology to explore complex behavioral dynamics in diverse populations.

Limitations

It is important to recognize that EMA methods have certain limitations. Since assessments are conducted repeatedly and in close relation to the behaviors, thoughts, and emotions being studied, there is a chance that the methods themselves could influence the variables being examined. This indeed is evidenced in the high percentage of participants who said they were made more aware of their emotions and activities because of the study. As such, there is the potential that throughout the EMA period, an individual's behaviors and self-reports begin to shift and this should be accounted for in subsequent analysis. Future studies should incorporate features to address participant reactivity, such as randomizing the order of questions.

It must also be noted that there are several general limitations with EMA methodologies. In particular, the surveys are self-reported so this may result in the same social desirability reporting that is seen in other self-report methods such as unrepeated questionnaires. Additionally, in an attempt to keep the surveys brief and accessible, very basic responses (ie, yes or no) were provided, which limits capturing the detail and nuance of eating occasions. Ideally, examples of child-parent interactions should be observationally recorded to allow for a richer interpretation of feeding interactions in young children with avid eating as well as determining the accuracy of parent reports. The results of this study also highlight that relying on parents to trigger event-based surveys often results in underreporting of eating occasions which impacts the accurate reporting of the frequency of eating occasions or food requests in children with avid eating. The development of wearable devices and camera recordings could be used in future research to capture more precise data on children's eating patterns.

Additionally, while EMA studies can be designed to include a diverse sample with various racial, ethnic, and socioeconomic backgrounds, it is essential to consider how likely different groups are to complete an EMA survey and how well the results can be generalized to the broader population. Although participants were offered the use of a smartphone for the study, no participant requested this, but it may still be the case that access to a smartphone was a barrier to participation. Equally, participants with higher completion rates than others may have greater time and fewer pressures around feeding interactions to be able to complete the surveys. Furthermore, there is also the possibility that parents who choose to participate and complete an EMA study may have a stronger interest in promoting healthy eating habits for their child than those who choose not to participate, which could introduce bias into the findings. Finally, although the adherence in our study aligns closely with previously reported short-term EMA studies [17], the 10-day duration may limit the ability to generalize to contexts requiring sustained monitoring over longer periods.

Conclusion

In conclusion, this study demonstrates the feasibility, acceptability, and potential of using EMA to investigate the intricate interplay between parental mood, feeding goals, contextual factors, and feeding practices in children with avid eating behaviors. While the results are promising, further refinement of the EMA methodology is needed to address technical challenges and reduce participant burden. Future research should continue to explore the use of EMA to capture the dynamic nature of feeding interactions and to inform the development of targeted interventions to promote healthy eating behaviors in children.

Acknowledgments

We would like to thank all the families for participating in the study. We are also grateful to Croker H, Farrow C, Haycraft E, Herle M, Llewellyn, and Blissett J for conceptualizing the research project and acquiring funding. This work was supported by

XSL•FO
RenderX

an Economic and Social Research Council (ESRC) research grant (ES/V014153/1). The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of data; and preparation, review, or approval of the manuscript.

Data Availability

The data sets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Each named author has substantially contributed to conducting the underlying research and drafting this manuscript. JB, CF, and EH have contributed to the conception and funding acquisition of the study. AP and KE have led the study design and data collection. AP has led the data analysis and write-up of the manuscript. JB, CF, EH, and KE have been involved in the editing and proofing of the manuscript.

Conflicts of Interest

None declared.

References

1. Pickard A, Croker H, Edwards K, et al. Identifying an avid eating profile in childhood: associations with temperament, feeding practices and food insecurity. Appetite 2023 Dec 1;191:107050. [doi: 10.1016/j.appet.2023.107050] [Medline: 37793473]
2. Edwards KL, Blissett J, Croker H, et al. Examining parents' experiences and challenges of feeding preschool children with avid eating behaviour. Appetite 2024 Jul 1;198:107372. [doi: 10.1016/j.appet.2024.107372] [Medline: 38657683]
3. Loth KA, Ji Z, Kohli N, Fisher JO, Fulkerson JA. Parents of preschoolers use multiple strategies to feed their children: findings from an observational video pilot study. Appetite 2023 Aug 1;187:106615. [doi: 10.1016/j.appet.2023.106615] [Medline: 37236362]
4. Hektner JM, Schmidt JA, Csikszentmihalyi M. Experience Sampling Method: Measuring the Quality of Everyday Life: Sage; 2007. [doi: 10.4135/9781412984201]
5. Napa SC, Prieto CK, Diener E. Experience sampling: promises and pitfalls, strength and weaknesses. In: Diener E, editor. Assessing Well-Being: Springer; 2009, Vol. 39:157-180. [doi: 10.1007/978-90-481-2354-4_8]
6. Runyan JD, Steinke EG. Virtues, ecological momentary assessment/intervention and smartphone technology. Front Psychol 2015;6:481. [doi: 10.3389/fpsyg.2015.00481] [Medline: 25999869]
7. Berge JM, Tate A, Trofholz A, et al. Examining variability in parent feeding practices within a low-income, racially/ethnically diverse, and immigrant population using ecological momentary assessment. Appetite 2018 Aug 1;127:110-118. [doi: 10.1016/j.appet.2018.04.006] [Medline: 29684414]
8. Loth KA, Ji Z, Wolfson J, Neumark-Sztainer D, Berge JM, Fisher JO. A descriptive assessment of A broad range of food-related parenting practices in A diverse cohort of parents of preschoolers using the novel real-time parent feeding practices survey. Int J Behav Nutr Phys Act 2022 Mar 2;19(1):22. [doi: 10.1186/s12966-022-01250-y] [Medline: 35236392]
9. Lopez NV, Lai MH, Yang CH, Dunton GF, Belcher BR. Associations of maternal and paternal parenting practices with children's fruit and vegetable intake and physical activity: preliminary findings from an ecological momentary study. JMIR Form Res 2022 Aug 10;6(8):e38326. [doi: 10.2196/38326] [Medline: 35947425]
10. Wardle J, Guthrie CA, Sanderson S, Rapoport L. Development of the children's eating behaviour questionnaire. J Child Psychol Psychiatry 2001 Oct;42(7):963-970. [doi: 10.1111/1469-7610.00792] [Medline: 11693591]
11. Edwards K, Croker H, Farrow C, et al. Examining parent mood, feeding context, and feeding goals as predictors of feeding practices used by parents of preschool children with avid eating behavior: protocol for an ecological momentary assessment study. JMIR Res Protoc 2024 Mar 19;13:e55193. [doi: 10.2196/55193] [Medline: 38502178]
12. APPETItE smartphone study: information video. Youtube. URL: https://www.youtube.com/watch?v=Gp5Ydnhk3K4 [accessed 2025-01-03]
13. de Vries LP, Baselmans BML, Bartels M. Smartphone-based ecological momentary assessment of well-being: a systematic review and recommendations for future studies. J Happiness Stud 2021;22(5):2361-2408. [doi: 10.1007/s10902-020-00324-7] [Medline: 34720691]
14. Myin-Germeys I, Kuppens P. The open handbook of experience sampling methodology. In: A Step-by-Step Guide to Designing, Conducting, and Analyzing ESM Studies: Leuven: Center for Research on Experience Sampling and Ambulatory Methods Leuven; 2022.
15. Pickard A, Blissett J, Farrow C, Edwards K. APPETItE - ecological momentary analysis. Published online. 2023. [doi: 10.17605/OSF.IO/N48YV]
16. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]

XSL•FO
RenderX

17. Berge JM, Trofholz A, Tate AD, et al. Examining unanswered questions about the home environment and childhood obesity disparities using an incremental, mixed-methods, longitudinal study design: the family matters study. Contemp Clin Trials 2017 Nov;62:61-76. [doi: 10.1016/j.cct.2017.08.002] [Medline: 28800894]

Abbreviations

CEBQ: Children's Eating Behavior Questionnaire
EMA: ecological momentary assessment
LPA: latent profile analysis

Edited by A Mavragani; submitted 23.09.24; peer-reviewed by M Stach, N Lopez; revised version received 07.01.25; accepted 07.01.25; published 27.02.25.

Please cite as:
Pickard A, Edwards K, Farrow C, Haycraft E, Blissett J
Capturing Everyday Parental Feeding Practices and Eating Behaviors of 3- to 5-Year-Old Children With Avid Eating Behavior: Ecological Momentary Assessment Feasibility and Acceptability Study
JMIR Form Res 2025;9:e66807
URL: https://formative.jmir.org/2025/1/e66807
doi:10.2196/66807

© Abigail Pickard, Katie Edwards, Claire Farrow, Emma Haycraft, Jacqueline Blissett. Originally published in JMIR Formative Research (https://formative.jmir.org), 27.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobilizing Health Promotion Through Canada's Student Mental Health Network: Concurrent, Mixed Methods Process Evaluation

Amy Ecclestone[1], BHSc; Brooke Linden[2], PhD; Jessica Rose[3], BSc; Kiran Kullar[3], BSc

[1]School of Public Health Sciences, University of Waterloo, Waterloo, ON, Canada

[2]Health Services and Policy Research Institute, Queen's University, 21 Arch Street, Kingston, ON, Canada

[3]Queen's University, Kingston, ON, Canada

Corresponding Author:
Brooke Linden, PhD
Health Services and Policy Research Institute, Queen's University, 21 Arch Street, Kingston, ON, Canada

Abstract

Background: Mental health issues among Canadian postsecondary students are prevalent. In tandem, an increased acknowledgment of the need for upstream mental health support has been highlighted. While the majority of institutions offer some form of mental health promotion, research suggests students are failing to access support due to barriers including lack of awareness, geographical and financial barriers, and lack of relevance in offerings. Canada's Student Mental Health Network is a web-based knowledge mobilization initiative designed to fill these gaps. With content created and curated "for-students, by-students" and reviewed by subject matter experts, the Network serves as a one-stop shop for evidence-based, mental health support for postsecondary students.

Objective: The goal of this research was to conduct the first component of a comprehensive program evaluation of the Network. This paper details a formative, process evaluation after approximately 1 year of operations, with the goal of assessing acceptability and feasibility.

Methods: Using a concurrent mixed methods study design, quantitative and qualitative data were simultaneously collected from students in order to evaluate the acceptability and feasibility of the Network as a mental health promotion resource. Quantitative data were automatically collected through Google Analytics via the website over the course of the first year of operations. Data collected included the number of users accessing the website, user engagement, and user "stickiness." Quantitative data were used to evaluate both accessibility and feasibility. Qualitative data were collected via individual, digital interviews conducted with a modest sample of students (n=8) across areas and levels of study. Qualitative data derived more detailed insights into user experience and website attributes, as well as feedback on content delivery, providing evidence used to evaluate feasibility.

Results: A total of 1200 users globally accessed the Network within the first year of operations, with Canadian users accounting for nearly 90% of total website traffic. An overall 66% engagement rate was observed, with the average user visiting 7 pages per session. Further support for the acceptability of the Network is demonstrated in the Canada-wide reach of the content development and review team. Evidence for the feasibility of the Network was observed through website use statistics indicating the most frequently viewed pages aligned with our goals: providing mental health education and increasing awareness of available resources. Qualitative feedback provided additional context surrounding the feasibility of the space, including positive feedback on the esthetics, relevance, usability, inclusion, and accessibility. Areas for content expansion and improvements to accessibility were also identified.

Conclusions: The results of this study provide evidence in support of the feasibility and acceptability of the Network as a web-based knowledge mobilization initiative in support of postsecondary students' mental health and well-being. Future research will pursue a summative, impact assessment to evaluate utility.

(JMIR Form Res 2025;9:e58992) doi:10.2196/58992

KEYWORDS

mental health; health promotion; program evaluation; postsecondary; knowledge translation

Introduction

Prevalence estimates for mental health–related problems, including above-average stress, psychological distress, and symptoms of mental illnesses, have increased significantly among Canadian postsecondary students over the past several years [1]. Indeed, data collected from Canadian postsecondary institutions via the 2022 National College Health Assessment III survey (n=11,322) revealed large proportions of students reporting moderate (47%) to high (37%) stress levels within the past 30-day period, with over half (52%) reporting their

XSL•FO
RenderX

stress levels to have impacted their academic performance [2]. Additionally, approximately one-third (33%) of respondents screened positive for serious psychological distress, while many reported having received a diagnosis for anxiety (23%) or depression (25%) [2].

Importantly, excessive stress has been linked to a number of negative academic (eg, decreased performance and decreased motivation) and health (eg, anxiety and depression) outcomes for postsecondary students [3,4]. In recent years, a more in-depth exploration of the sources of student stress has been conducted in the development and deployment of the Postsecondary Student Stressors Index. This work has revealed that stressors experienced by students within the postsecondary setting reach beyond academics, also spanning the learning environment, campus culture, personal, and interpersonal domains [5,6]. Importantly, this suggests a need for holistic mental health promotion support, including providing resources that extend beyond the academic realm. Further complicating things, postsecondary-aged Canadians were also the age group identified as the most likely to report dramatically declining mental health throughout the COVID-19 pandemic [7]. The long-term mental health fallout of the pandemic continues to be unknown, further reinforcing the need to provide well-being support to students as they navigate the stressors associated with the postsecondary setting.

Strategies to address and support postsecondary mental health have emerged at the national, provincial, and institutional levels [8]. At the postsecondary level, many institutions have attempted to improve student mental health by offering both upstream and downstream mental health services [8,9]. Upstream services aim to intervene prior to the development of symptoms of mental illness (eg, mental health promotion and mental illness prevention), while downstream interventions (eg, counseling and pharmacological therapy) aim to treat individuals who have already reached a clinical threshold. These service levels align with Keyes' seminal conceptual framework, the dual continuum of mental health and mental illness, which conceptualizes mental health and mental illness as existing on interrelated, but separate continuums [10]. A student who is experiencing a period of poor mental health, but who has not yet reached the clinical threshold for a mental illness, may benefit greatly from the use of upstream services. Unfortunately, the majority of students experiencing a decline in their mental health seek downstream treatments as an initial solution. As a result, the demand for downstream campus mental health services (eg, counseling) over the past decade has far outpaced institutions' capacities to deliver timely care [11]. Placing additional focus on upstream mental health promotion and mental illness prevention resources may contribute toward alleviating the current bottleneck observed at the treatment level by encouraging students to implement appropriate forms of help-seeking that align with their level of need. Bolstering upstream supports, including mental health promotion, may also improve students' resilience and education regarding adaptive and effective coping mechanisms, empowering students to better manage stressful situations.

While some form of mental health promotion is offered on most postsecondary campuses, a national review of mental health and well-being services on Canadian postsecondary campuses found that only 70% of respondents felt students were well-informed about available resources on campus, and about mental health issues in general [12]. Notably, many respondents (including campus service providers and institution administrators; 84%) also indicated they felt there was room for improvement in the mental health promotion efforts currently offered on their campus [12]. Further, mental health promotion activities were highly varied, with some offering programs aimed to inform students about available campus mental health services, reduce stigma, and educate students about mental illness. Other institutions focused on outreach, aiming to encourage students to seek help when needed. In many cases, the mental health supports offered were dependent on the resources (eg, financial, logistic, and staffing) available to the institution, with smaller institutions unable to provide as much support as their larger counterparts [12].

Together, these gaps and inconsistencies suggest a need for a comprehensive and holistic approach to mental health promotion, where resources are universally accessible to postsecondary students across Canada. With the overall aim of filling these gaps, Canada's Student Mental Health Network ("the Network") was created—a web-based, "one-stop shop" for mental health education and evidence-based resources. Content is tailored specifically to postsecondary students and is provided in a centralized, universally accessible location, focusing on three primary pillars of action: improving mental health literacy, encouraging the development of strong social support networks, and improving understanding of appropriate help-seeking and awareness of available resources.

A comprehensive evaluation of the Network, including formative and summative components, has been detailed and published elsewhere [13]. The purpose of this paper is to present the results of the formative, process evaluation of the Network after approximately 1 year of operation. Process evaluations are intended to assess 2 key components of program function: (1) whether the program is reaching the intended targeted population (acceptability), and (2) whether its delivery and function are consistent with program design intentions (feasibility) [14]. The key objectives of this formative evaluation were to determine the feasibility and acceptability of the Network to date and to inform the development of the protocol for the subsequent summative impact assessment.

Methods

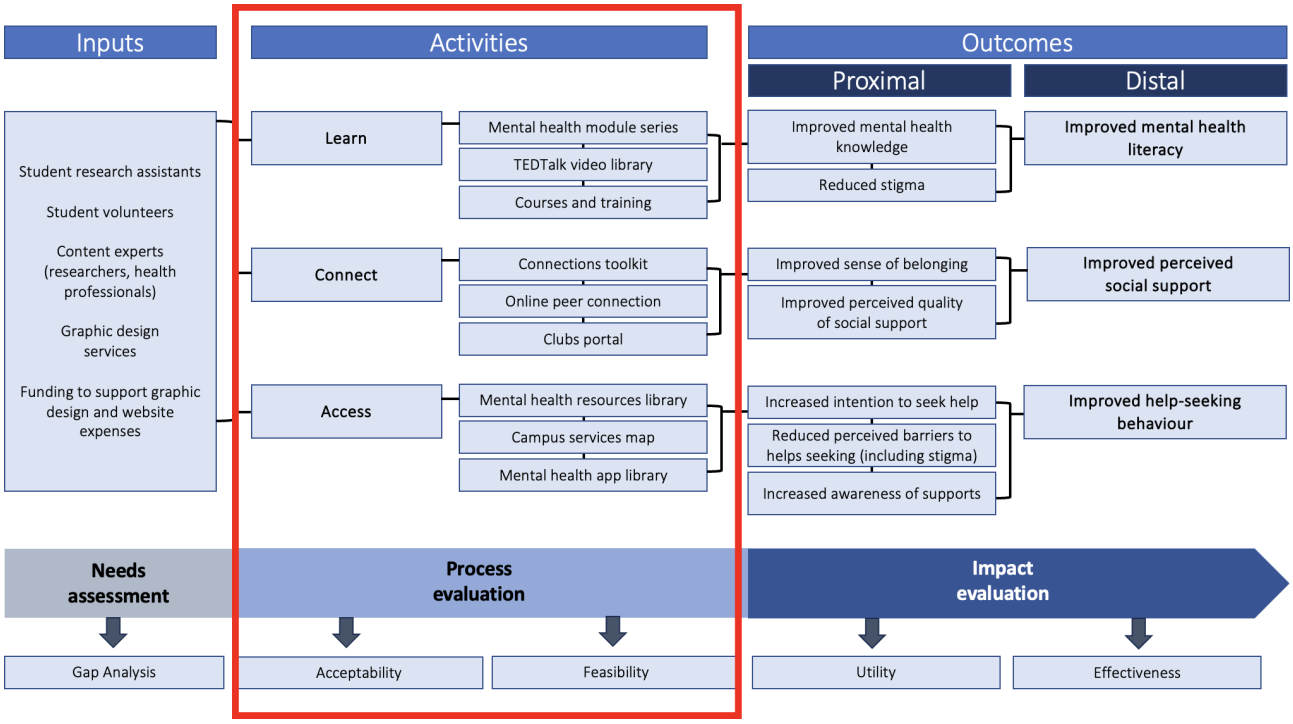
Study Design

We used a concurrent mixed methods evaluation study design, combining both quantitative and qualitative data to gain more in-depth insights into the acceptability and feasibility of the Network. Best practices in evaluation dictate that the chosen evaluation design should align with the maturity of the initiative. Given the age of the Network at the time of this study, a formative process evaluation design was chosen, with a focus laid on process indicators selected to provide information about feasibility and acceptability, rather than focusing on outcomes evaluating impact. A future summative impact evaluation is planned for when the initiative has matured appropriately [13].

XSL•FO
RenderX

Figure 1 displays the logic model used to visually depict the program theory supporting the process evaluation.

Figure 1. Logic model depicting program theory for the Student Mental Health Network, with components relevant to the formative process evaluation highlighted.



Initiative

The Student Mental Health Network is a website designed to be a web-based resource hub tailored toward postsecondary students, featuring mental health education and evidence-based tools and resources.

The target population for the Network itself is postsecondary students enrolled across Canada.

Centered around improving mental health literacy, encouraging the development of strong social support networks, and improving help-seeking behaviors, the content is intended to be relevant for all students, regardless of the type of institution attended (eg, university, college, and institute), years, areas, and levels of study. We have also aimed to develop content that is accessible, psychologically safe, and culturally competent for various subgroups of students who may be best supported by specific content areas, including medical students, members of the LGBTQ2S+ community, BIPOC students, Indigenous students, international students, and first-generation students. All content is created and curated by students and reviewed for validity by subject matter experts (eg, mental health researchers, clinicians, and medical professionals) using a collaborative, knowledge mobilization approach. Further information on the resources featured in each of the three content pillars (learn, connect, and access) has been published elsewhere [13].

Recruitment

To meet the objectives of this evaluation, data were collected using two methods: (1) quantitative website use statistics, collected through Google Analytics, and (2) qualitative data collected through individual, cognitive interviews.

To collect information regarding the reach and coverage of the Network in terms of website engagement and user experience, we used quantitative website metrics collected passively through Google Analytics. These metrics included: the total number of users, average number of users (weekly, monthly, and annually), average engagement time (time spent on each component of the website) and user "stickiness" (average length of visits), page views, event counts (clicks, file downloads), and user region. The definitions of these specific criteria per the Google Analytics glossary can be viewed in Multimedia Appendix 1.

Second, individual, cognitive interviews were conducted using a "think aloud" technique to further evaluate user experience and highlight gaps in the Network's delivery [15]. Participants were recruited using a convenience sampling method; a recruitment notice was posted on the Network website, as well as through the Network's social media channels. Participants had to be computer literate and currently enrolled at a postsecondary institution at the time of the interview. A total of 8 interviews were conducted before data saturation was reached (eg, the point at which additional interviews were not producing new information and we were beginning to hear repetitive answers to interview questions). Table 1 shows the demographic characteristics of the 8 interviewees. In terms of procedure, participants were asked to screen share over Zoom (Zoom Video Communications Inc) as they navigated the Network website and talked the interviewer through their exploration, identifying content-related issues or preferences, as well as any navigational challenges. The interviews included questions related to important website attributes (eg, appearance, content, interactivity, inclusivity or representation, and accessibility) and perceived acceptability and feasibility of the space. Interviews were audio recorded and automatically

XSL•FO
RenderX

transcribed, with field notes taken throughout the interview and immediately reviewed afterward.

Table . Descriptive statistics for demographics characteristics of cognitive interview sample.

Characteristic	Value (N=8), n (%)
Gender	
Woman	5 (60)
Man	3 (40)
Year of study	
Y1 undergraduate	1 (13)
Y2 undergraduate	2 (25)
Y3 undergraduate	2 (25)
Y4 undergraduate	2 (25)
Graduate	1 (13)
Area of study	
Health sciences	3 (38)
Science	3 (38)
Arts	1 (13)
Graduate studies	1 (13)
Institution	
Queen's University	5 (63)
Western University	1 (13)
University of Toronto	2 (25)

Data Analysis

Descriptive statistics provided through Google Analytics were displayed as frequency tables. Cognitive interview data were analyzed through the use of the website attribute codes displayed in Table 2. Field notes were taken electronically during the interviews and reviewed immediately afterward. Website attribute codes were applied to relevant transcript areas. In the event that a new issue arose that was not captured by existing problem codes, a new code was created.

Table . Website attribute codes used to categorize cognitive interview responses and appropriate corrective actions to be taken.

Website attribute	Description	Corrective action
Esthetics	Comments on the look and feel of the website, including whether the color scheme, images, and multimedia used are appropriate.	• Improve esthetics (additional graphics, multimedia, color)
Relevance	Comments on whether the content displayed on the website is not relevant to postsecondary student stress or mental health.	• Add additional relevant content • Remove irrelevant content
Accessibility	Comments on the accessibility of content is adequate (eg, ease of use for students with visual, audio, or other impairments).	• Improve accessibility features of the website
Inclusion	Comments on whether the content provided adequately caters to a wide variety of students and important subgroups.	• Change in wording • Change the order of items
Usability	Comments on ease of navigation of the website, whether the website renders on mobile and web browsers adequately, etc.	• Improve navigation • Improve optimization

Ethical Considerations

This research study was reviewed and approved for ethical compliance by the Queen's University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board (TRAQ #6041988). Google Analytics data were collected passively from website users and were not connected to any identifying information. Therefore, there was no consent component associated with this part of the study. Interview participants provided verbal, informed consent to participate (including

XSL•FO
RenderX

audio and video recording of the interview) after reviewing the letter of information for the study. All qualitative data (eg, transcripts and extracted quotations) were deidentified for analyses. Interview recordings were kept confidential and stored as password-protected electronic files by the principal investigator (BL) as the data custodian. Interview participants were compensated with a CAD $20 (US $14.04) e-gift card to Amazon as thanks for their participation.

Results

Acceptability

Recall that acceptability refers to whether or not the program or initiative is reaching its intended targeted population [14]. Website use statistics were extracted from Google Analytics for the dates of May 1, 2022, to June 1, 2023, inclusive. A total of 1200 users globally accessed the Network within this time frame, where Canadian users accounted for nearly 90% of total website traffic (Multimedia Appendix 1). During this time frame, about 1500 “engaged sessions” were recorded, defined as the number of sessions (or visits) on the website that exceeded 10 seconds in length or had ≥2 page views. Users also spent some time navigating the Network, with the average user visiting approximately 7 pages per session. This resulted in an overall 66% engagement rate during the first year of operations. This data suggests that users are not only visiting the website but are spending time exploring its resources. Similarly, during this time frame, a total of 21,532 “events” (a specific interaction or occurrence on a website, such as a page view, click, or download) were recorded. Although we were not able to capture the age of these users, the continued and increased presence of users from across regions of Canada over time provides preliminary evidence in support of acceptability.

The acceptability of the Network is also reflected in the Canada-wide reach demonstrated through our content development team. Our student volunteers, who develop content and promote the Network, attend postsecondary institutions across the country and are enrolled across various programs, levels, and areas of study, including medicine, health science, political science, computer science, public health, education, and more. In addition to contributing students, we have a team of subject matter experts who serve as expert reviewers of website content to ensure its validity and reliability. These experts include faculty specializing in related research, Student Wellness Services staff and clinicians, and family physicians. Figure 2 shows the Canada-wide reach of the Network, including each of these contributing groups after 1 year of operations.

XSL•FO
RenderX

Figure 2. Map depicting the cross-Canada reach of the Student Mental Health Network after approximately 1 year of operations via the postsecondary institutions at which both student contributors and subject matter experts reside.

Feasibility

Feasibility is determined by whether a program's delivery and function are consistent with program design intentions [14]. We first gained insight into the feasibility of the Network by examining the website use statistics to determine which pages were viewed most frequently (Multimedia Appendix 1). In line with our goals, resources from the learn section of the Network were some of the most sought after the first year of operations (eg, module series, toolkits, and online courses). This was consistent with our initial needs assessment, which suggested that students were eager to improve their mental health education. Notably, several resources within the access pillar of the website were also within the most frequently visited pages (eg, campus services map, mental health resources bank, stress management resources, and mental health app library). This is consistent with our goals to encourage student awareness of available mental health resources. Finally, resources within the connect pillar of the website (eg, clubs portal and peer connection resources) were also among the most frequently visited pages, consistent with our goal to encourage the development of strong social support networks.

We also derived evidence for feasibility by conducting cognitive interviews with 8 students over the course of the evaluation period. Interviewees were asked to navigate through the Network website and comment on the appearance, relevance, accessibility, and usability. Overall, feedback was positive across interviews, with participants sharing that the site was visually appealing, easy to access, and relevant to them or their peers. Table 3 depicts the results of the interviews, as well as assigned website attribute codes.

XSL•FO
RenderX

Table . Responses to website attributes as indicated by interview participants (N=8).

Interviewee	Esthetics	Relevance	Accessibility	Inclusion	Usability
001	+[a]	+	+, x[b]	+	+
002	+	+	+	+	+
003	+, x	+	+	+	+, x
004	+	+	+	+	x
005	+	+	+	+	+
006	+	+	+, x	+	+, x
007	+, x	+	+	+	+
008	+	+	+	+	+

[a]+ indicates a positive response to this attribute.

[b]x indicates that suggestions for improvement were recommended by the participant.

Esthetics

When asked about the appearance of the website, participants indicated liking the color scheme and its consistency. The use of multimedia (eg, images, graphics, and videos) was reported to increase the visual appeal of the website. Participants also found that the content was displayed in an engaging way. In particular, they reported enjoying the format in which the toolkits and modules were presented, especially the integration of different types of media (eg, links to web pages, TED Talks, and worksheets). Participants indicated that the resources displayed as interactive multimedia were most esthetically pleasing (as well as easiest to navigate), suggesting that resources currently provided as downloadable documents be adapted to be displayed in a more esthetic, interactive alterative format.

Usability

With respect to usability, participants noted that the website was easy to navigate, with clear pathways to associated resources. Participants appreciated the list of included items at the bottom of each page, clearly indicating each subsection of website pillars. Participants were also asked which type of device they used to access the online space, and whether the website and its resources rendered properly on their device. Despite the type of device and operating system used, the website was rendered properly. Many respondents suggested that a mobile app or version be developed specifically with tablets and mobile devices in mind.

Relevance

When asked about the relevance of the website content to the student target population, all participants agreed that the content was relevant and valuable to a postsecondary audience. Participants also shared that it was clear content was intended for students from all schools and provinces and did not appear to be preferential to a single region. When asked if they would recommend the site or any of its resources to a friend or family member, every participant said yes, with some indicating that they had already done so.

Accessibility and Inclusivity

Interview participants were asked to share their impressions of the accessibility and inclusivity of the website. Respondents shared positive comments regarding the accessibility of the site, stating that content appeared to be directed toward a wide range of student backgrounds and not targeted to any specific group. Participants also indicated that they were able to find everything they were looking for with ease, suggesting the website was user-friendly in terms of accessibility. Regarding inclusivity, participants thought the website content was generally applicable to all students, the more specific resources provided (eg, graduate students' module, Indigenous well-being module, and intersectionality module) were especially inclusive as they targeted specific struggles experienced by students belonging to those groups. Respondents encouraged the development of additional resources (eg, modules and toolkits) targeting other specialty populations to enhance the inclusivity of the content. Overall, positive feedback was provided regarding both the accessibility and inclusivity of the website. Participants' recommendations regarding additional resources and content that could be improved or enhanced are displayed in Table 4.

XSL•FO
RenderX

Table . Recommendations provided by interview participants (N=8) for additional or adapted content for the Student Mental Health Network.

Area of website	Suggested addition or update
Learn	• A library of recommended self-help books fostering positive mental health, resilience, and well-being • Development of an international student module, LGBTQ+[a] module, and expansion of resources for students of color, LGBTQ+ students, international students, and Indigenous students
Connect	• A library of social media accounts that aim to promote mental health
Access	• Reformatting of the Mental Health Resources bank to be displayed similarly to the clubs bank
General	• A newsletter updating interested students on new resources added to the site, and on general student mental health news (eg, recent impactful research) • A Q&A[b] page for the site to answer any common questions students may have while navigating our site • Option to increase text size • Option to select "read aloud" or prerecorded audio for some sections and resources • French translation of website content • Development of a mobile-optimized version of the Network

[a]LGBTQ+: lesbian, gay, bisexual, transgender, queer, and other sexual and gender minority.

[b]Q&A: question and answer.

Discussion

Principal Findings

As prevalence estimates for mental health issues have increased over the past decade [1], increasing attention has been paid to postsecondary mental health. The development of institutional frameworks and guiding documents such as the Okanagan Charter [16] and Canada's National Standard for the Psychological Safety and Wellbeing of Postsecondary Students [17] have emerged, alongside new forms of data collection such as the Canadian Campus Wellbeing Survey [18]. Across the board, the recommendation for improvements in the targeting of mental health support for students is evident [8]. With the demand for downstream treatment-based services (eg, counseling and medication) continually surpassing the campus' capacities to provide care to all, there is a glaring need to work upstream. Bolstering and improving the targeting of these upstream mental health supports is imperative to reducing the burden on downstream campus mental health services by empowering students to improve their resiliency and develop a toolkit of resources to turn to during times of increased stress where lower-intensity support may result in marked improvements.

The Student Mental Health Network offers 1 such solution as a web-based mental health promotion resource specifically targeted toward postsecondary students and universally accessible across Canada. The Network provides unlimited access to mental health education and resources, addressing many of the common barriers cited by students attempting to access mental health supports, including financial restrictions, lengthy wait times, stigma, and logistical access (eg, time, location, or travel-based constraints) [19]. The findings from this formative process evaluation have provided preliminary evidence for both the acceptability and feasibility of the Network. Results showed the imitative is reaching its intended target population (acceptability), and that its delivery and function are consistent with program design intentions (feasibility). In addition to positive feedback from the target population, opportunities for growth were also identified.

Results from Google Analytics provided evidence for accessibility (that the initiative is reaching its intended targeted population), with web use statistics indicating that a total of 1200 users globally accessed the Network within the first year of operations, the vast majority of whom resided in Canada (90%). This data suggests that the Network is reaching its intended target population. Providing additional support to this conclusion is the Canada-wide reach of the Network, evidenced through the regional distribution of our content contributors and expert reviewers, including postsecondary students, mental health researchers, clinicians, and physicians from across the country. Users also spent time navigating the Network, with the average user visiting approximately 7 pages per session. This finding provides evidence that users are not simply popping in and out of the space, but rather, are taking the time to view multiple pages and absorb some of the information provided. Overall, the Network reported a high engagement rate (66%) within its first year of operations. Overall, this data suggests that users are not only visiting the website but are spending time exploring its resources, as we hoped they would. Collecting passive user data via Google Analytics allowed us to cast a much wider net with respect to determining usership and avoid issues common to survey-based research, such as low response rates and selection bias.

Feasibility was supported by Google Analytics data in addition to the data derived from cognitive interviews. Recall that the

XSL•FO
RenderX

goal of the Network was to provide Canadian postsecondary students with a web-based, "one-stop shop" for mental health education and evidence-based resources focusing on mental health education, social support, and improving awareness of available mental health resources. After the first year of operations, results suggest that the Network's delivery and function have been consistent with our goals. Valid and reliable content has been created across all 3 areas of focus and is well-received by students. Data from Google Analytics showed that resources within all 3 sections fell within the top 20 most frequently visited pages of the Network, with resources in the learn and access sections being the most popular. Results from interviews further supported students' endorsement of the Network and its resources, with positive reviews regarding website esthetics, relevance of content, accessibility, inclusion, and usability. Interviewees identified some areas for improvement, including additional content to fill some existing gaps, as well as minor esthetic or display changes to existing resources, but no resource was identified for removal or irrelevance. Here, we see the added value of the mixed methods design of the evaluation, where valuable contextual information was provided as a result of the cognitive interviews compared to what would have been gleaned from the collection of quantitative data alone. These interviews allowed for a detailed walkthrough of the website, facilitating a rich conversation about numerous elements (eg, relevance, accessibility, and esthetics). Overall, the design elements highlighted by study participants echo those identified by Garret et al [20] as key to thoughtful website design and driving user engagement.

Limitations

Despite its strengths, there are several limitations to this research that should also be acknowledged. While the Network is a novel initiative aiming to fill important gaps, its novelty also presents evaluation challenges given the absence of a standardized process for evaluating this type of initiative. Despite this limitation, we did our best to thoughtfully design a comprehensive program theory and evaluation plan, including both the process evaluation outlined here, as well as a future summative impact evaluation. In terms of data collection, only 8 interviews were conducted with all students attending postsecondary institutions in Ontario, resulting in limited geographic variability in our qualitative sample. However, we did see good variability in terms of other characteristics, such as gender, area, and level of study. Furthermore, while our qualitative sample was small, we did find that we reached data saturation in our interviews. Additionally, data collected through Google Analytics were able to provide us with information on the number of people visiting the website over time and from which regions, but not the age groups to which those visitors belonged. As a result, we can conclude that website traffic increased over time, but we do not have quantitative evidence to confirm that those visitors were postsecondary-aged individuals. Additionally, while the Network has demonstrated a cross-Canada reach after the first year of operations, there is an uneven distribution of users and content contributors with the majority situated in Ontario. This may be due to the fact that the Network was initially created by a group of students at a single, Eastern Ontario university, or due to the vast majority of postsecondary institutions in Canada being located in this province. Moving forward, we plan to continue working on extending the reach of the Network across the country. Finally, website content is currently only available in English, which poses an accessibility challenge for French-speaking postsecondary students. Going forward with Network development, we aim to translate content into French to provide equitable access to mental health resources to Canadian students in both official languages.

Conclusions

Bolstering upstream supports like mental health promotion may help alleviate the current bottleneck observed at the downstream service level on postsecondary campuses. Canada's Student Mental Health Network aims to contribute toward alleviating this issue by delivering a novel, innovative method of providing universally accessible, evidence-based mental health promotion to postsecondary students across the country. The Network provides a centralized location for mental health education and resources, created for-students, by-students, and validated by subject matter experts, including physicians, clinicians, and mental health researchers. The space will contribute to breaking down common barriers to help-seeking, such as geographic restraints (eg, by providing universally accessible content that can serve online students, as well as those with travel or logistical obstacles), financial restraints (eg, by providing cost-free education and resources wherever possible), and lack of awareness (eg, by maintaining an active social media presence and partnering with Student Wellness Services at postsecondary institutions across Canada to bolster promotion). The participatory, for-students, by-students approach is a key strength of the Network and a characteristic that many students have highlighted as a unique and attractive quality of the space. The Network is a first-of-its-kind initiative that aims to fill gaps in current mental health promotion resources provided to students using an innovative, collaborative design. Through the provision of relevant and focused content and the facilitation of social connection, the Network will support and empower students to take ownership of their mental health and well-being by providing them with the toolbox they need to build and maintain resilience.

The formative process evaluation component for the Network has now been completed, with data collection for the impact evaluation currently underway. The next steps will include the ongoing creation, curation, and refreshing of the Network resources in order to meet the changing needs of the target population, as well as to fill any gaps identified throughout the evaluation process. Resources will continue to be updated to ensure a high level of accessibility and relevance. Preliminary results from this evaluation show the Network is meeting its preliminary objectives. The summative assessment of the Network will aim to further evaluate these objectives and provide additional evidence of overall impact.

XSL•FO
RenderX

Acknowledgments

The authors would like to acknowledge and thank all the student contributors involved in the development and curation of content for the Student Mental Health Network [21]. Dr. Brooke Linden's position is funded by the Bell Canada Mental Health and Anti-Stigma Research Chair.

Data Availability

The data generated or analyzed during this study are not publicly available due to Ethics Board restrictions but are available from the corresponding author for secondary analysis upon reasonable request and upon submission of an additional ethics request to the approving Ethics Board.

Authors' Contributions

AE and JR led data collection and analysis. AE, JR, and KK took lead roles in the drafting and writing of the manuscript. BL was the principal investigator on the research project and held a supervisory role throughout the data collection, analysis, reporting, writing, and editing of the manuscript. All authors approved the manuscript for final submission.

Conflicts of Interest

All authors have a nonfinancial competing interest as founding members of the Student Mental Health Network.

Multimedia Appendix 1
Google Analytics terms glossary and website usage statistics.
[DOCX File, 216 KB - formative_v9i1e58992_app1.docx]

References

1. Linden B, Boyes R, Stuart H. Cross-sectional trend analysis of the NCHA II survey data on Canadian post-secondary student mental health and wellbeing from 2013 to 2019. BMC Public Health 2021 Dec;21(1). [doi: 10.1186/s12889-021-10622-1]
2. Canadian reference group data report spring 2022. American College Health Association. 2022. URL: https://www.acha.org/wp-content/uploads/2024/07/NCHA-III_SPRING_2022_CANADIAN_REFERENCE_GROUP_DATA_REPORT.pdf [accessed 2025-02-07]
3. Pascoe MC, Hetrick SE, Parker AG. The impact of stress on students in secondary school and higher education. Int J Adolesc Youth 2020 Dec 31;25(1):104-112. [doi: 10.1080/02673843.2019.1596823]
4. Duffy A, Keown-Stoneman C, Goodday S, et al. Predictors of mental health and academic outcomes in first-year university students: identifying prevention and early-intervention targets. BJPsych Open 2020 May 8;6(3):e46. [doi: 10.1192/bjo.2020.24] [Medline: 32381150]
5. Linden B, Stuart H. Post-secondary stress and mental well-being: a scoping review of the academic literature. Can J Commun Ment Health 2020 May 1;39(1):1-32. [doi: 10.7870/cjcmh-2020-002]
6. Linden B, Stuart H, Ecclestone A. Trends in post-secondary student stress: a pan-Canadian study. Can J Psychiatry 2023 Jul;68(7):521-530. [doi: 10.1177/07067437221111365] [Medline: 35791667]
7. National polling initiative. Mental Health Research Canada. 2020. URL: https://www.mhrc.ca/national-polling-covid [accessed 2022-05-17]
8. Monaghan C, Linden B, Stuart H. Postsecondary mental health policy in Canada: a scoping review of the grey literature: Politique de santé mentale post-secondaire au Canada: un examen de la portée de la littérature grise. Can J Psychiatry 2021 Jul;66(7):603-615. [doi: 10.1177/0706743720961733] [Medline: 33016127]
9. DiPlacito-DeRango ML. Acknowledge the barriers to better the practices: support for student mental health in higher education. Can J Scholarsh Teach Learn 2016;7(2). [doi: 10.5206/cjsotl-rcacea.2016.2.2]
10. Keyes CLM. The mental health continuum: from languishing to flourishing in life. J Health Soc Behav 2002 Jun;43(2):207-222. [doi: 10.2307/3090197] [Medline: 12096700]
11. McGartland K. White paper: mental health on campus. Centre for Innovation in Campus Mental Health. 2018. URL: https://campusmentalhealth.ca/wp-content/uploads/2018/03/White-Paper-Mental-Health-on-Campus1.pdf [accessed 2022-05-17]
12. Jaworska N, De Somma E, Foneska B, Heck E, MacQueen G. Mental health services for students at postsecondary institutions: a national survey. Can J Psychiatry 2016;12(61). [doi: 10.1177/0706743716640752]
13. Ecclestone A, Linden B, Monaghan C, Zheng S. Canada's Student Mental Health Network: protocol for a comprehensive program evaluation. JMIR Res Protoc 2023 Jun 22;12:e41521. [doi: 10.2196/41521] [Medline: 37347511]
14. Rossi PH, Lipsey MW, Freeman HE. Evaluation: A Systematic Approach, 7th edition: SAGE Publications; 2003.
15. Wolcott MD, Lobczowski NG. Using cognitive interviews and think-aloud protocols to understand thought processes. Curr Pharm Teach Learn 2021 Feb;13(2):181-188. [doi: 10.1016/j.cptl.2020.09.005] [Medline: 33454077]

XSL•FO
RenderX

16. Okanagan charter: an international charter for health promoting universities and colleges. Best Practices in Canadian Higher Ed. 2015. URL: https://bp-net.ca/wp-content/uploads/2019/03/Okanagan-Charter.pdf [accessed 2025-01-20]
17. Mental health and well-being for post-secondary students (CSAZ 2003). : Canadian Standards Association; 2020.
18. Fagan MJ, Wunderlich KB, Faulkner G. Psychological distress among Canadian postsecondary students: a repeated cross-sectional analysis of the Canadian Campus Wellbeing Survey (CCWS) between spring 2020-2023. J Am Coll Health 2024 Jul 31;2024:1-6. [doi: 10.1080/07448481.2024.2382438] [Medline: 39083793]
19. Moghimi E, Stephenson C, Gutierrez G, et al. Mental health challenges, treatment experiences, and care needs of post-secondary students: a cross-sectional mixed-methods study. BMC Public Health 2023 Apr 6;23(1):655. [doi: 10.1186/s12889-023-15452-x] [Medline: 37020282]
20. Garett R, Chiu J, Zhang L, Young SD. A literature review: website design and user engagement. Online J Commun Media Technol 2016 Jul;6(3):1-14. [Medline: 27499833]
21. Student Mental Health Network. URL: www.studentmentalhealthnetwork.ca [accessed 2025-02-07]

Edited by A Mavragani; submitted 29.03.24; peer-reviewed by A AL-Asadi, KW Tay; revised version received 27.01.25; accepted 28.01.25; published 27.02.25.

Please cite as:
Ecclestone A, Linden B, Rose J, Kullar K
Mobilizing Health Promotion Through Canada's Student Mental Health Network: Concurrent, Mixed Methods Process Evaluation
JMIR Form Res 2025;9:e58992
URL: https://formative.jmir.org/2025/1/e58992
doi:10.2196/58992

© Amy Ecclestone, Brooke Linden, Jessica Rose, Kiran Kullar. Originally published in JMIR Formative Research (https://formative.jmir.org), 27.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluation of Financial Support Workshops for Patients Under State Pension Age With Degenerative Cervical Myelopathy: Survey Study

Tanzil Rujeedawa[1], MB BChir; Zahabiya Karimi[2], BA; Helen Wood[2], MA; Irina Sangeorzan[2], MSc; Roy Smith[2], BSc; Iwan Sadler[2]; Esther Martin-Moore[2]; Adrian Gardner[3], PhD; Andreas K Demetriades[4], BChir, MB MPhil; Rohitashwa Sinha[5], BMBS, PhD; Gordan Grahovac[6], MD, PhD; Antony Bateman[7], MSc; Naomi Deakin[1], PhD; Benjamin Davies[1], BSc, MB ChB, MPhil, MRCS

[1]Division of Neurosurgery, Department of Clinical Neurosciences, University of Cambridge, The Old Schools, Trinity Ln, Cambridge, United Kingdom
[2]Myelopathy.org, Cambridge, United Kingdom
[3]Aston University, Birmingham, United Kingdom
[4]Edinburgh Spinal Surgery Outcome Studies Group, Department of Neurosurgery, Royal Infirmary Edinburgh, Edinburgh, United Kingdom
[5]Department of Neurosurgery, Leeds General Infirmary, Leeds, United Kingdom
[6]King's College Hospital, London, United Kingdom
[7]Royal Derby Spinal Centre, Derby, United Kingdom

Corresponding Author:
Benjamin Davies, BSc, MB ChB, MPhil, MRCS
Division of Neurosurgery, Department of Clinical Neurosciences, University of Cambridge, The Old Schools, Trinity Ln, Cambridge, United Kingdom

Abstract

Background: Degenerative cervical myelopathy (DCM), a form of slow-motion and progressive spinal cord injury caused by spinal cord compression secondary to degenerative pathology, leads to high levels of disability and dependence, and may reduce quality of life. Myelopathy.org is the first global scientific and clinical charity for DCM, providing an accessible platform freely disseminating information relevant to the DCM diagnosis and its treatment. Significant transient and long-term change to earnings do occur and can thrust individuals into poverty. People with DCM face many challenges accessing state financial assistance. This can have a cumulative negative financial effect due to the association between DCM and low socioeconomic index. Financial support available to patients under pension age include Universal Credit (UC), a payment that helps with living costs, and Personal Independence Payment (PIP), which helps with extra living costs if someone has both a long-term health condition or disability and difficulty doing certain everyday tasks.

Objective: This study aimed to assess if delivering workshops centered around access to financial support could assist people with DCM living in the United Kingdom.

Methods: A series of 2 internet-based workshops was targeted at accessing financial support for English patients under the state pension age, with an anonymized survey delivered to participants after each session. The first session was on UC and the second on PIP. The survey consisted of a mixture of Likert scales, free text and yes or no answers. Survey responses were analyzed using descriptive statistics and free text answers underwent inductive thematic analysis.

Results: The average rating on the use of UC was 9.00/10. Presession self-rated confidence levels were 5.11/10 rising to 8.00/10. The mean score of wanting further similar sessions was 8.67/10 with 56% (5/9) of participants wanting one-to-one sessions. For PIP, the average session use rating was 10/10. Presession self-rated confidence levels were 4.43/10 rising to 9.57/10. The mean score of wanting further similar sessions was 8.71/10, with 43% (3/7) of participants wanting one-to-one sessions . Following inductive thematic analysis, themes regarding the usefulness of such sessions and the challenges to accessing financial support emerged. One participant gave negative feedback, which included the length of the session and perceived problems around confidentiality and data protection.

Conclusions: The pilot series was largely perceived as a success, with participants finding them useful and increasing their self-rated confidence in navigating the UK financial support system. Given the small sample size, it is hard to predict the success of future sessions. Finally, given that the hurdles in accessing financial support extend beyond DCM, such workshops may be relevant to other organizations.

(JMIR Form Res 2025;9:e59032) doi:10.2196/59032

XSL•FO
RenderX

KEYWORDS

myelopathy; degenerative; spine; spinal; benefits; aid; financial; money; income; poverty; disability; disabled; finance; workshop; education; service; access; accessibility; navigate; confidence; government

Introduction

Degenerative cervical myelopathy (DCM) is a form of slow-motion and progressive spinal cord injury caused by spinal cord compression secondary to degenerative pathology. This includes disease processes such as cervical spondylosis, ossification of the posterior longitudinal ligament, ossification of the ligamentum flavum, and degenerative disc disease [1,2]. Globally, it is estimated to affect 2% of adults, although less than 10% are formally diagnosed at this time [3,4]. DCM can cause sensorimotor dysfunction in the upper or lower limbs, gait disturbance, bladder or bowel dysfunction, and pain [2,5-9]. Surgery is the mainstay of treatment for DCM and aims to decompress the spinal cord [2,10-12]. Whilst this typically stops further deterioration and may afford meaningful symptomatic improvement, recovery is generally incomplete. People with DCM experience high levels of disability and dependence, and are among those with the lowest quality of life scores for any chronic disease [13]. In fact, almost half of diagnosed individuals may be dependent on the support of others for activities of daily living (41.9% (326/778) in a recent study [14]).

This landscape, in the study mentioned above, results in great socioeconomic impact, with over a third of participants unable to return to work (35.7%; 278/778 of people with DCM [14]) due to DCM. Furthermore, 25.8% (201/778) were employed full time and 11.3%(88/778) were employed part-time [14]. In addition, 35.7% (278/778) were unable to work due to their disability [14]. Surgery can be helpful, with Lønne et al [15] showing that by 12 months after surgery, 65% (589/906) had returned to work, increasing to 75% (680/906) by 36 months. Clearly, therefore, DCM impairs patients' ability to contribute economically before retirement and can lead people into poverty [16]. Davies et al [17] found that, based on the average age of cases at presentation in the United Kingdom and an average age of retirement of 65 years, there is a potentially 15.1 years of affected productivity in people with DCM [17]. This equates to an estimated lifetime inflated loss of income of up to £347,112 (US $435,150) using UK Office for National Statistics figures or £474,719 (US $595,121) using Office for Economic Co-Operation figures [17]. In fact, there is an estimated annual loss of productivity of £362.6m (US $453.3m), in addition to disability financial support totaling £280.2m (US $350.3m) and an overall cost to society for this cohort of £681.6m (US $ 853.9m) in the United Kingdom [17]. This has a cumulative negative financial effect due to the association of DCM with a low socioeconomic index, a delayed diagnosis, and poor outcomes [18].

Whilst this provides an estimate of the current "demand" and potential future "societal cost" of DCM, the practicalities of ensuring that people with DCM can access support is seldom straightforward, since social support systems differ by both jurisdiction and patient circumstance. A study from Norway, where financial support is available to all citizens suffering ill-health, found that 20% of patients received medical income compensation financial support 1 year before surgery, rising to 100% at the time of surgery. However, for most countries, this "gold standard" access is complex. First, the term "DCM" is relatively new, having been introduced in 2015 [1]. Furthermore, being focused on undergoing treatment and lacking a true understanding of the long-term implications of the disease, people with DCM rarely recognize (or accept) the need for financial support until needs are heightened [19-21]. Professionals may be aware of these challenges, as patients often turn to their medical team for support, guidance, and written opinions. However, many health care professionals are not trained to advise patients on available options and are likely unequipped to best navigate the claim and assessment process.

Myelopathy.org is the first global scientific and clinical charity for DCM. The charity provides an accessible internet-based platform freely disseminating information relevant to the DCM diagnosis and its treatment. As an organization, it hosts peer-to-peer support forums, and fields inquiries from those affected by DCM. Across this media, contacts requesting information or assistance for accessing financial support is a frequent topic.

Several active organizations exist that offer advice and guidance to people applying for financial support. These include Hastings Advice and Representation Centre (HARC) [22], Citizens Advice [23], and Welfare Benefits Unit [24]. In fact, certain charities such as the British Heart Foundation have dedicated web pages specific to applying for financial support [25]. Some, such as Age Scotland, offer workshops to their beneficiaries about accessing benefits [26]. However, there is no literature assessing the utility of these interventions. We hypothesized that delivering workshops centered around access to financial support could assist people with DCM.

Methods

Overview

The authors, in collaboration with Myelopathy.org, piloted a series of 2 workshops in May 2023 targeted at accessing financial support for English patients under the state pension age.

Participants

People with DCM were recruited to the sessions virtually, using an advertisement within the Facebook (Meta) support group run by Myelopathy.org, which had 2065 members at the time of the study. Interested participants contacted the charity directly by email to book sessions. This was the only form of recruitment. Inclusion criteria was any person with DCM under state pension age from England. Where a session was oversubscribed, a random number generator was used to select 10 participants.

XSL•FO
RenderX

Procedures

The sessions were delivered in a paid partnership with HARC, a charitable organization, which provides welfare financial support advice and information to the people of East Sussex and the South Coast of England [22]. It aims to address inequalities, relieve poverty, and improve the quality of life for vulnerable and disadvantaged individuals and families [22]. Funding was from the National Lottery community fund, a UK specific fund [27].

The overall aim was to equip participants with the skills to navigate the financial support system independently. The workshop program consisted of 2 sessions delivered as a set. All sessions were held virtually on Zoom (Zoom Video Communications) and each lasted up to 2 hours. The workshop format was chosen in partnership with HARC. HARC provided expertise from generic sessions delivered previously, with the chosen workshop format encouraging interactivity and the ability to ask questions, as well as being cost-effective.

The sessions were targeted to patients currently under state pension age, allowing content to be tailored to age demographics. This is dependent on when someone was born [28]. The first session was on Universal Credit (UC), a payment that helps with living costs [28] and the second on Personal Independence Payment (PIP), which helps with extra living costs if someone has both a long-term health condition or disability and difficulty doing certain everyday tasks or getting [29]. The sessions will be termed UC and PIP, respectively. Each session consisted of a comprehensive overview of either UC or PIP and their application process with a question time at the end.

Measures

Participants were evaluated using an anonymized survey (Google Forms), sent by email to the participants by the charity after attending each workshop.

Analysis

Survey responses were analyzed using descriptive statistics (mean, median, mode, and range) using R (version 4.2.1; R Foundation for Statistical Computing). Free text answers underwent inductive thematic analysis to identify key themes of feedback by TR and BD separately. Themes were identified manually with TR an BD having a discussion afterwards to ensure validity. Given the small quantity of data, no codebook software was used.

Ethical Considerations

This study was conducted with ethical approval from the Myelopathy.org Board (M2023001). At the start of the workshop, participants were provided with an overview of the workshop and that feedback would be asked at the end that may be used for publication. By continuing into the workshop, participants were providing informed consent to participate. All data collected were anonymous. No incentives were offered for completion of the surveys.

Results

A total of 24 people with DCM registered for the workshops. There were 9 attendees in UC and 7 in PIP, with 2 individuals dropping out from the first session to the second. Demographic information was not recorded.

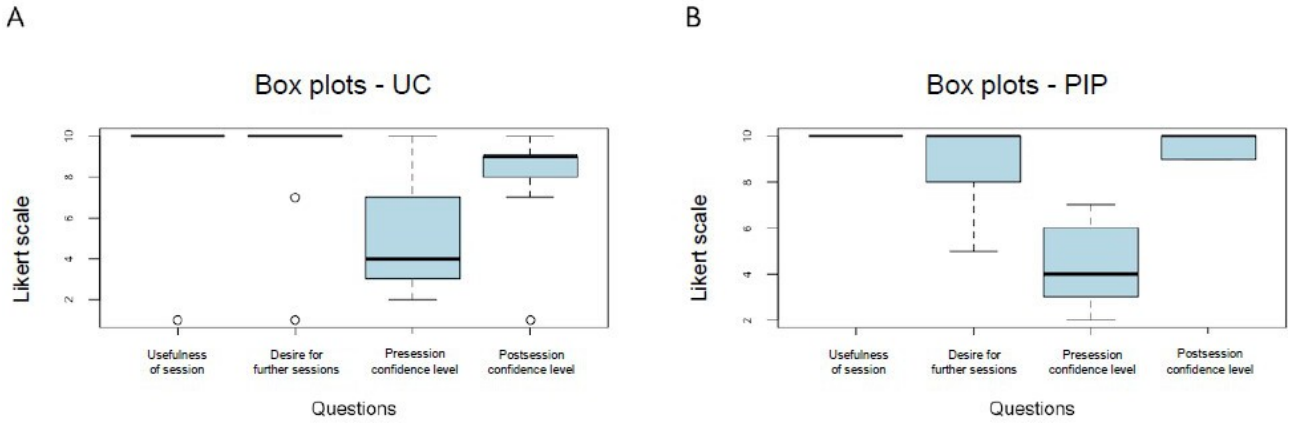
Quantitative Feedback

Participants were asked to rate the sessions across four questions, using a Likert scale (1 to 10) for each. Figure 1 summarizes the responses to the survey administered after each session, while Tables S1 and S2 (Multimedia Appendix 1) provide responses at the level of each participant.

The average rating on the usefulness of UC was 9.00/10 (SD 2.82). Presession self-rated confidence levels were 5.11/10 (SD 2.73) rising to 8.00/10 (SD 2.62) after the session. The mean score of wanting further similar sessions was 8.67/10 (SD 2.87), with 56% (5/9) of participants wanting one-to-one sessions.

For PIP, the average session usefulness rating was 10/10 (SD 0), with the desire for further sessions. Presession confidence levels were 4.43/10 (SD 1.76) rising to 9.57/10 (SD 0.49). The mean score of wanting further similar sessions was 8.71/10 (SD 1.75) with 43% (3/7) of participants wanting one-to-one sessions.

Figure 1. A. Box plots showing the response in the survey administered after the UC session. B. Box plots showing the response in the survey administered after the PIP session. UC: universal credit; PIP: personal independence payment.



XSL•FO
RenderX

Qualitative Feedback

Several themes emerged during the thematic analysis, notably with regards to the usefulness of the sessions and the challenges to accessing financial support.

An important theme that emerged was the usefulness of the sessions. Participants noted that the sessions were "interactive," "amazing," "informative," and that they had "enjoyed this very much." Both sessions were generally viewed as very helpful, with several participants noting the information given being particularly useful with a participant noting that the sessions contained "specific information that was tailored to participants' needs, rather than just a bullet point description of financial support available." The information on what the questions on the forms actually meant and how to answer them was noted as especially helpful. In fact, most participants noted that they could not suggest any improvements for the sessions. The desire for further sessions was also noted, with several participants keen for more as demonstrated by one of the participants saying that they "would like more of these types of sessions as they are so so beneficial and also make you feel like you are less alone in the journey," and another saying "just more sessions would be amazing." One participant even suggested group discussions that include Myelopathy.org members who had made successful claims.

The challenges to accessing financial support were also highlighted. The confusing system was the main challenge noted. The lack of knowledge about eligibility, lack of understanding and knowledge from professionals and fear of judgment were also mentioned. One participant "felt alone and like I was being judged" and another "never applied before because I never felt like I was entitled to any, but I feel more confident now." In addition, participants found it hard to understand how detailed answers to the forms needed to be. Finally, physical disability making it hard to fill in forms was another issue.

Certain negative themes also came across from only 1 of the participants. They did not find the session helpful due to what was perceived as "unnecessary personal introductions, and waffle," and some technical issues that prevented the screen from being shared properly at the start. In addition, they felt that details of the session were not properly explained in advance. They also perceived that the facilitator had strayed "into areas of confidentiality and data protection." This was due to an incident whereby another one of the participants had mentioned to the facilitator that they were going to the toilet and the facilitator accidentally mentioned it to the whole group. The participant involved had no issues with this.

Discussion

With the support of a charitable partner, Myelopathy.org delivered a series of workshops that were effective in increasing awareness and confidence in obtaining financial support in the United Kingdom. No other similar study was found through a search on PubMed for comparison, even in a non-myelopathy context, thereby illustrating the novel nature of this initiative. Workshops were considered highly useful (9-10/10) and increased confidence by >40%.

Following inductive thematic analysis, themes regarding the usefulness of the sessions and the challenges to accessing financial support emerged. The interactivity of the session was viewed as an important positive. The usefulness of the sessions was also noted as well as the desire for further sessions. The challenges of applying for financial support were also explored, notably the confusing system, physical disability, the lack of knowledge about eligibility, lack of understanding and knowledge from professionals, and fear of judgment. One participant had negative feedback, which included the length of the session and perceived problems around confidentiality and data protection. The feedback was taken seriously and given to the facilitator. An apology was presented to the participant by ZK, the director of Myelopathy.org, and the misunderstanding was resolved. This participant did not attend PIP and could represent a subgroup of the people with DCM population for whom another mode of engagement would be more suitable such as anonymized access of online web pages or documents.

While the intervention was considered successful, it is important to consider areas for improvement, particularly to ensure the intervention can be sustained and exported to other regions or countries. Of note, participants were split about wanting further one-to-one sessions despite thinking further sessions would be helpful. This could be potentially explained by the participants preferring to be in a group setting rather than alone.

Experience from other settings is limited and is an avenue that should be explored. It should also be acknowledged that individuals who sign up for a group workshop may be more likely to like such settings. This initial evaluation also does not collect objective data relevant to actual or successful financial support applications. Other limitations include the fact that the sessions were focused solely on financial support available in England and particular economic situations. How these findings relate to outside England and situations such as self-employed status is therefore unclear. Furthermore, delivery of sessions virtually, available to only a limited number of participants, may provide only a narrow snapshot of people with DCM. This situation is compounded by the fact that Facebook-only recruitment may skew participant demographics. Another limitation is that no demographic data were collected that could have helped in contextualizing findings and identifying possible trends across specific demographic groups.

Despite the limitations outlined above, the participants largely deemed this initial pilot to be a welcome success. Given this success, Myelopathy.org aims to provide further similar sessions both in England and other countries. In addition, this can serve as a template for small charities focusing on relatively unknown conditions that cause a financial burden. The socioeconomic impact of DCM is a global research priority [30]. The chronic disability experienced by most, even if transiently during the drawn-out diagnostic process and aftermath from surgery, can dramatically alter household incomes and thrust people into poverty [16]. There is a cumulative negative effect due to the association of DCM with low socioeconomic conditions.

XSL•FO
RenderX

Conclusion

Despite the presented limitations and a small amount of negative feedback from one individual, the sessions were largely perceived favorably, with participants finding them useful and increasing their self-rated confidence in navigating the UK financial support system. The small sample size limits any further extrapolation; it is indeed hard to predict how future sessions will be received. Nonetheless, Myelopathy.org intends to co-host future sessions with the lessons learned from this pilot. Finally, difficulties in accessing financial support extend beyond DCM, and hence, the lessons learned in this evaluation may be helpful to other patient groups and support organizations.

Acknowledgments

BD is supported by an NIHR Clinical Doctoral Fellowship. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

TR contributed to conceptualization, manuscript drafting, and preparation. ZK, HW, I Sangeorzan, RS, I Sadler, and EMM contributed to conceptualization. ND contributed to manuscript review. BD contributed to conceptualization and manuscript review.

Conflicts of Interest

BD, HW, I Sadler, and RS are Trustees of Myelopathy.org. ZK is the director of Myelopathy.org.

Multimedia Appendix 1
Supplementary tables.
[DOCX File, 16 KB - formative_v9i1e59032_app1.docx]

References

1. Nouri A, Tetreault L, Singh A, Karadimas SK, Fehlings MG. Degenerative cervical myelopathy: epidemiology, genetics, and pathogenesis. Spine (Phila Pa 1976) 2015 Jun 15;40(12):E675-E693. [doi: 10.1097/BRS.0000000000000913] [Medline: 25839387]
2. Badhiwala JH, Ahuja CS, Akbar MA, et al. Degenerative cervical myelopathy - update and future directions. Nat Rev Neurol 2020 Feb;16(2):108-124. [doi: 10.1038/s41582-019-0303-0] [Medline: 31974455]
3. Grodzinski B, Stubbs DJ, Davies BM. Most degenerative cervical myelopathy remains undiagnosed, particularly amongst the elderly: modelling the prevalence of degenerative cervical myelopathy in the United Kingdom. J Neurol 2023 Jan;270(1):311-319. [doi: 10.1007/s00415-022-11349-8] [Medline: 36056204]
4. Smith SS, Stewart ME, Davies BM, Kotter MRN. The prevalence of asymptomatic and symptomatic spinal cord compression on magnetic resonance imaging: a systematic review and meta-analysis. Global Spine J 2021 May;11(4):597-607. [doi: 10.1177/2192568220934496] [Medline: 32677521]
5. Tracy JA, Bartleson JD. Cervical spondylotic myelopathy. Neurologist 2010 May;16(3):176-187. [doi: 10.1097/NRL.0b013e3181da3a29] [Medline: 20445427]
6. Kalsi-Ryan S, Karadimas SK, Fehlings MG. Cervical spondylotic myelopathy: the clinical phenomenon and the current pathobiology of an increasingly prevalent and devastating disorder. Neuroscientist 2013 Aug;19(4):409-421. [doi: 10.1177/1073858412467377] [Medline: 23204243]
7. Harrop JS, Naroji S, Maltenfort M, et al. Cervical myelopathy: a clinical and radiographic evaluation and correlation to cervical spondylotic myelopathy. Spine (Phila Pa 1976) 2010 Mar 15;35(6):620-624. [doi: 10.1097/BRS.0b013e3181b723af] [Medline: 20150835]
8. Tetreault L, Goldstein CL, Arnold P, et al. Degenerative cervical myelopathy: a spectrum of related disorders affecting the aging spine. Neurosurgery 2015 Oct;77 Suppl 4:S51-S67. [doi: 10.1227/NEU.0000000000000951] [Medline: 26378358]
9. Boerger T, Alsouhibani A, Mowforth O, et al. Moving beyond the neck and arm: the pain experience of people with degenerative cervical myelopathy who have pain. Global Spine J 2022 Sep;12(7):1434-1442. [doi: 10.1177/2192568220986143] [Medline: 33626937]
10. Fehlings MG, Ibrahim A, Tetreault L, et al. A global perspective on the outcomes of surgical decompression in patients with cervical spondylotic myelopathy: Results from the prospective multicenter aospine international study on 479 patients. Spine (Phila Pa 1976) 2015;40(17):1322-1328. [doi: 10.1097/BRS.0000000000000988]

XSL•FO
RenderX

11. Fehlings MG, Wilson JR, Kopjar B, et al. Efficacy and safety of surgical decompression in patients with cervical spondylotic myelopathy. JBJS 2013;95(18):1651-1658. [doi: 10.2106/JBJS.L.00589]
12. Wilson JR, Tetreault LA, Kim J, et al. State of the art in degenerative cervical myelopathy: an update on current clinical evidence. Neurosurgery 2017 Mar 1;80(3S):S33-S45. [doi: 10.1093/neuros/nyw083] [Medline: 28350949]
13. Oh T, Lafage R, Lafage V, et al. Comparing quality of life in cervical spondylotic myelopathy with other chronic debilitating diseases using the short form survey 36-health survey. World Neurosurg 2017 Oct;106:699-706. [doi: 10.1016/j.wneu.2016.12.124] [Medline: 28065875]
14. Pope DH, Mowforth OD, Davies BM, Kotter MRN. Diagnostic delays lead to greater disability in degenerative cervical myelopathy and represent a health inequality. Spine (Phila Pa 1976) 2020 Mar 15;45(6):368-377. [doi: 10.1097/BRS.0000000000003305] [Medline: 31658234]
15. Lønne VV, Hara S, Gulati S, et al. Return to work after surgery for degenerative cervical myelopathy: a nationwide registry-based observational study. Acta Neurochir (Wien) 2023 Mar;165(3):779-787. [doi: 10.1007/s00701-023-05521-w] [Medline: 36795223]
16. Davies B, Matters SIM. Series 2 ep 1 - poverty & myelopathy. : Mylopathy matters URL: https://soundcloud.com/myelopathymatters/s2e1-poverty-myelopathy [accessed 2025-02-04]
17. Davies BM, Phillips R, Clarke D, et al. Establishing the socio-economic impact of degenerative cervical myelopathy Is fundamental to improving outcomes [AO Spine RECODE-DCM Research Priority Number 8]. Global Spine J 2022 Feb;12:122S-129S. [doi: 10.1177/21925682211039835] [Medline: 35174730]
18. Rethorn ZD, Cook CE, Park C, et al. Social risk factors predicting outcomes of cervical myelopathy surgery. J Neurosurg Spine 2022 Jul 1;37(1):41-48. [doi: 10.3171/2021.12.SPINE21874] [Medline: 35090132]
19. Sangeorzan I, Andriopoulou P, Davies BM, McNair A. The information needs of people with degenerative cervical myelopathy: A qualitative study to inform patient education in clinical practice. PLoS ONE 2023;18(5):e0285334. [doi: 10.1371/journal.pone.0285334] [Medline: 37205664]
20. Davies BM, Khan DZ, Barzangi K, et al. We choose to call it “Degenerative Cervical Myelopathy”: findings of AO Spine RECODE-DCM, an international and multi-stakeholder partnership to agree a standard unifying term and definition for a disease. Global Spine J 2024 Mar;14(2):503-512. [doi: 10.1177/21925682221111780] [Medline: 35769029]
21. Davies BM, Mowforth O, Wood H, et al. Improving awareness could transform outcomes in degenerative cervical myelopathy [AO Spine RECODE-DCM Research Priority Number 1]. Global Spine J 2022 Feb;12(1_suppl):28S-38S. [doi: 10.1177/21925682211050927] [Medline: 35174734]
22. HARC Hastings Advice and Representation Centre. URL: https://harcuk.com [accessed 2025-02-04]
23. Citizens advice. 2023. URL: https://www.citizensadvice.org.uk/benefits [accessed 2025-02-04]
24. Welfare benefits unit. 2023. URL: http://www.welfare-benefits-unit.org.uk [accessed 2025-02-04]
25. Can I get financial support or benefits if I have a heart condition? British Heart Foundation. URL: https://www.bhf.org.uk/informationsupport/support/practical-support/can-i-get-financial-support-or-benefits-if-i-have-a-heart-condition [accessed 2025-02-06]
26. Benefits workshops. Age Scotland. URL: https://www.agescotland.org.uk/information-advice/information-workshops/benefits-workshops [accessed 2025-02-06]
27. The national lottery community fund. URL: https://www.tnlcommunityfund.org.uk/ [accessed 2025-02-06]
28. Line A, insurance N, pension S, pension AS state pension. How and when you can claim. URL: https://www.gov.uk/universal-credit [accessed 2025-02-11]
29. Personal independence payment. Gov.uk. URL: https://www.gov.uk/pip [accessed 2025-02-04]
30. Davies BM, Phillips R, Clarke D, et al. Establishing the socio-economic impact of degenerative cervical myelopathy Is fundamental to improving outcomes [AO Spine RECODE-DCM Research Priority Number 8]. Global Spine J 2022 Feb;12(1_suppl):122S-129S. [doi: 10.1177/21925682211039835] [Medline: 35174730]

Abbreviations

DCM: degenerative cervical myelopathy
HARC: Hastings Advice and Representation Centre
PIP: personal independence payment
UC: universal credit

XSL•FO
RenderX

Edited by A Mavragani; submitted 31.03.24; peer-reviewed by F Khan, G Goodman; revised version received 26.11.24; accepted 15.12.24; published 24.02.25.

Please cite as:
Rujeedawa T, Karimi Z, Wood H, Sangeorzan I, Smith R, Sadler I, Martin-Moore E, Gardner A, K Demetriades A, Sinha R, Grahovac G, Bateman A, Deakin N, Davies B
Evaluation of Financial Support Workshops for Patients Under State Pension Age With Degenerative Cervical Myelopathy: Survey Study
JMIR Form Res 2025;9:e59032
URL: https://formative.jmir.org/2025/1/e59032
doi:10.2196/59032

©Tanzil Rujeedawa, Zahabiya Karimi, Helen Wood, Irina Sangeorzan, Roy Smith, Iwan Sadler, Esther Martin-Moore, Adrian Gardner, Andreas K Demetriades, Rohitashwa Sinha, Gordan Grahovac, Antony Bateman, Naomi Deakin, Benjamin Davies. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Snowball Group Usability Testing for Rapid and Iterative Multisite Tool Development: Method Development Study

Katherine L Dauber-Decker[1,2], PhD; David Feldstein[3], MD; Rachel Hess[4], MS, MD; Devin Mann[5,6], MD, MS; Eun Ji Kim[1,2], MS, MSc, MD; Pranisha Gautam-Goyal[7], MD; Jeffrey Solomon[1,2], BFA; Sundas Khan[8,9], MD; Fatima Malik[1,2], MHA; Lynn Xu[5], MPH; Ainsley Huffman[10], MS; Paul D Smith[3], MD; Wendy Halm[11], DNP, FNP-BC, APNP; Alice Yuroff[3], PhD; Safiya Richardson[5,6,12], MPH, MD

[1]Feinstein Institutes for Medical Research, Northwell Health, 600 Community Drive, Suite 403, Manhasset, NY, United States

[10]Utah Clinical and Translational Science Institute, University of Utah, Salt Lake City, UT, United States

[11]University of Wisconsin School of Nursing, Madison, WI, United States

[12]Institute for Excellence in Health Equity, NYU Langone Health, New York, NY, United States

[2]Northwell Health, 2000 Marcus Avenue, Suite 300, New Hyde Park, NY, United States

[3]University of Wisconsin School of Medicine and Public Health, Madison, WI, United States

[4]University of Utah Health, Salt Lake City, UT, United States

[5]Department of Population Health, NYU Grossman School of Medicine, New York, NY, United States

[6]Department of Health Informatics, Medical Center Information Technology, NYU Langone Health, New York, NY, United States

[7]Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Manhasset, NY, United States

[8]Department of Medicine, Baylor College of Medicine, Houston, TX, United States

[9]Center for Innovations in Quality, Effectiveness and Safety, Michael E. DeBakey VA Medical Center, Houston, TX, United States

Corresponding Author:
Katherine L Dauber-Decker, PhD
Feinstein Institutes for Medical Research, Northwell Health, 600 Community Drive, Suite 403, Manhasset, NY, United States

Abstract

Background: Usability testing is valuable for assessing a new tool or system's usefulness and ease-of-use. Several established methods of usability testing exist, including think-aloud testing. Although usability testing has been shown to be crucial for successful clinical decision support (CDS) tool development, it is often difficult to conduct across multisite development projects due to its time- and labor-intensiveness, cost, and the skills required to conduct the testing.

Objective: Our objective was to develop a new method of usability testing that would enable efficient acquisition and dissemination of results among multiple sites. We sought to address the existing barriers to successfully completing usability testing during CDS tool development.

Methods: We combined individual think-aloud testing and focus groups into one session and performed sessions serially across 4 sites (snowball group usability testing) to assess the usability of two CDS tools designed for use by nurses in primary and urgent care settings. We recorded each session and took notes in a standardized format. Each site shared feedback from their individual sessions with the other sites in the study so that they could incorporate that feedback into their tools prior to their own testing sessions.

Results: The group testing and snowballing components of our new usability testing method proved to be highly beneficial. We identified 3 main benefits of snowball group usability testing. First, by interviewing several participants in a single session rather than individuals over the course of weeks, each site was able to quickly obtain their usability feedback. Second, combining the individualized think-aloud component with a focus group component in the same session helped study teams to more easily notice similarities in feedback among participants and to discuss and act upon suggestions efficiently. Third, conducting usability testing in series across sites allowed study teams to incorporate feedback based on previous sites' sessions prior to conducting their own testing.

Conclusions: Snowball group usability testing provides an efficient method of obtaining multisite feedback on newly developed tools and systems, while addressing barriers typically associated with traditional usability testing methods. This method can be applied to test a wide variety of tools, including CDS tools, prior to launch so that they can be efficiently optimized.

Trial Registration: Clinicaltrials.gov NCT04255303; https://clinicaltrials.gov/study/NCT04255303

(JMIR Form Res 2025;9:e55316) doi:10.2196/55316

XSL•FO
RenderX

KEYWORDS

clinical decision support; CDS; decision aid; clinical aid; cough; sore throat; strep pharyngitis; snowball group usability testing; snowball group; usability testing

Introduction

Usability testing assesses a new tool or system's usefulness and ease-of-use. Usability testing is particularly important in clinical decision support (CDS) development. CDS tools are evidence-based tools that help clinicians make decisions to improve patient care. Usability testing during CDS development increases the likelihood of tool adoption and impact by creating a tool that is workflow-integrated, useful, and easy to use [1-5]. The current standard for usability testing of CDS tools is think-aloud testing in which approximately 3 - 5 end users are asked to think out loud while interacting with a prototype of the tool during individual sessions.

Although the benefits of usability testing are well established, it is often not conducted during typical CDS development due to tight project timelines. Think-aloud testing with 3 - 5 users may add weeks to a project timeline. This becomes a larger challenge in multisite development projects. Typically, usability testing is done on the same tool at multiple sites in parallel, where end-users at each site are likely to report similar issues, adding weeks to the project timeline at each site with minimal added value of testing at each site. An efficient, effective method for usability testing CDS across multiple sites is needed.

In addition to think-aloud testing done in individual sessions, focus groups are also used to obtain feedback on a tool. These two types of testing can provide different types of insight from future end-users [6,7]. Focus groups include interactions between different participants, which can lead to the development of consolidated ideas through conversation. However, a single vocal participant can influence the direction of the discussion and may not represent all the participants' views. One-on-one interviews, in contrast, enable the research team to discuss individual-level feedback with each participant [7]. Including both types of usability testing allows for the highest yield of testing [1,4].

We sought to develop a new method of usability testing, termed snowball group usability testing, that would enable efficient, high-yield multisite testing. We applied our method to a project aimed at decreasing antibiotic overprescribing in the outpatient setting. In the United States, antibiotics are frequently prescribed inappropriately, contributing to antibiotic resistance [8-13]. The goals of our testing were to (1) determine the feasibility and practicality of snowball group usability testing and (2) facilitate the development of two useful and usable CDS tools. Here we describe our method for snowball group usability testing.

Methods

Ethical Considerations

All study activities were approved by the New York University Langone Institutional Review Board (i19-01222). Verbal informed consent was obtained from all usability testing participants. Any identifiable data collected as part of the study were stored on secure drives and only accessible to members of the study team. Participants were not compensated for partaking in this research study. This manuscript does not include any identifiable study data.

Setting and Recruitment

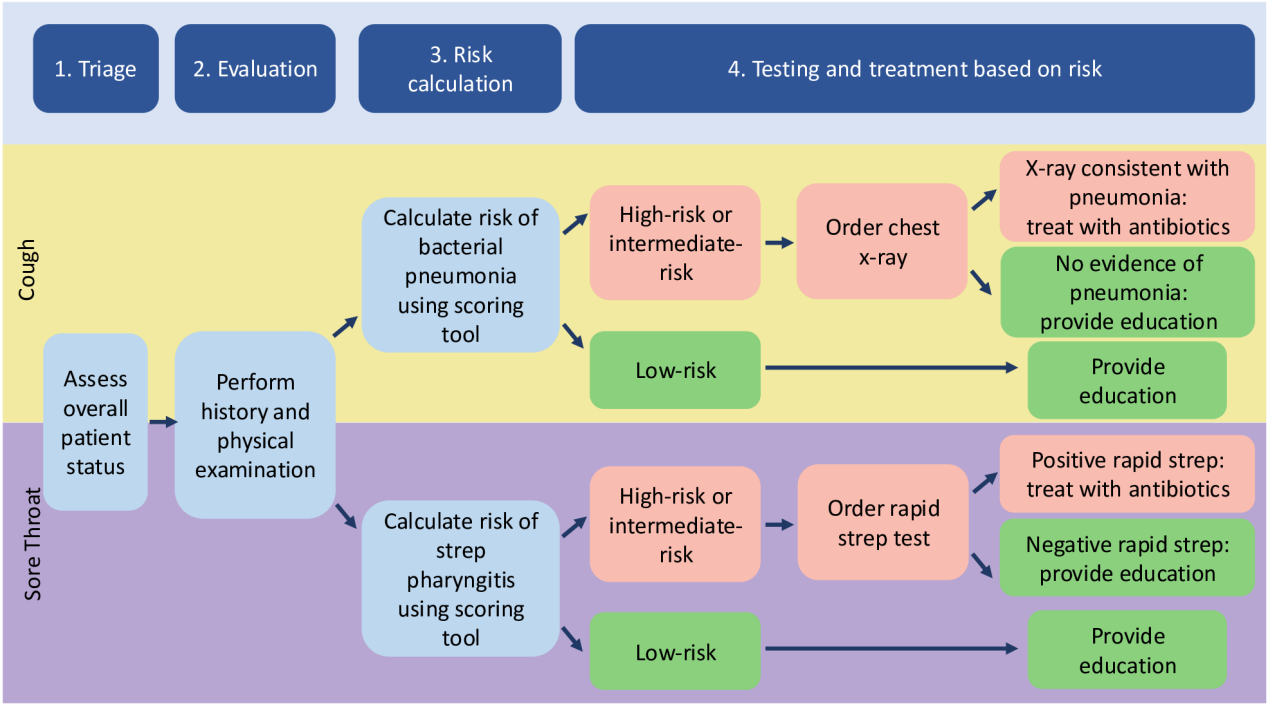
We conducted this study at 4 large academic sites in New York, Wisconsin, and Utah. The study teams at each site were experienced in CDS design, development, and evaluation, with a focus on using human-centered design strategies. Nurses who were eligible to participate in usability testing were identified based on their roles at study clinics and eligibility criteria. The study was presented to nurses by the study team. Eligible nurses were contacted by email for recruitment. Nurses who were interested in participating provided verbal consent and were given a key information sheet that outlined their research participation.

Description of CDS Tools

For this study, we used the Heckerling rule for cough [14] and the Centor score for sore throat [15-17] to develop CDS tools for use by nurses in an outpatient setting. We developed each of these tools for use in two different electronic health records (EHRs), Epic and Allscripts TouchWorks, due to site differences. Each tool included an in-person nurse visit section and a triage section that could be applied either over the phone or in-person. During triage, the nurse or medical assistant recorded the patient's symptoms using a tool that helped determine the risk and severity level for the patient. This was used to determine if the patient needed to be seen in person for those who were assessed over the phone. Patients who needed to be seen or were already in the office but did not have severe symptoms or other significant illnesses were deemed eligible for an in-person nurse visit. Alternatively, the patient could be sent to an emergency department or physician visit, or for those who were assessed over the phone, they could be advised that they do not need a visit and could self-treat with supportive care at home (Figure 1). The assessment questions were developed based on current site standards for patients with cough and sore throat [18].

XSL•FO
RenderX

Figure 1. Cough and sore throat clinical decision support workflow diagram. The workflow for both tools included 4 steps: (1) triage, (2) evaluation, (3) risk calculation, and (4) testing and treatment based on risk.



If a nurse visit was deemed appropriate for the patient, the nurse completed the visit section during an in-person encounter. The visit tools included a note for documenting history and the physical exam, a risk calculator to determine the risk of bacterial infection, and order sets based on the results of the risk calculator. Orders could be placed or pended by nurses to be signed by providers. The risk calculators were based on the Heckerling rule or Centor score for cough or sore throat, respectively, and recommendations were given based on the overall score (Table 1). Based on the number of criteria with which the patient presented, rule users could stratify patients into low-, intermediate-, or high-risk categories, indicating different levels of care.

Table . Heckerling rule and Centor score criteria.

Rule	Criteria	Low-risk	Intermediate- or high-risk
Heckerling rule (pneumonia)	• Temperature of >100°F (37.8 °C) • Heart rate of >100 beats/min • Crackles (rales) • Decreased breath sounds • No history of asthma	0 - 1: supportive care indicated	2 - 5: chest x-ray and antibiotics if x-ray was positive
Centor score (strep pharyngitis)	• History of fever: feels feverish, sudden onset of cold with shivering sweats, or oral temperature of ≥100.4°F (38.0 °C) • Absence of cough: new onset, frequent and/or persistent, different from baseline • Tonsillar exudates • Tender anterior cervical nodes	0 - 1: no rapid strep test indicated	2 - 3: rapid strep test indicated and antibiotics if the test is positive

Snowball Group Usability Testing Sessions

We combined individual think-aloud sessions and focus groups into one session. We conducted these sessions at each site serially, with the hope that each subsequent organization would reach saturation more quickly by benefitting from the previous site's insights. This form of testing will hereafter be referred to as snowball group usability testing, in which "snowballing" refers to conducting usability testing serially at each site, transferring feedback regarding the CDS tools from site to site as testing progresses, and "group" refers to conducting testing with multiple participants at once. To our knowledge, snowballing has not previously been used in the context of usability testing. We applied snowballing with the goal of increasing the efficiency at which results were obtained and passed along between study institutions. Each site performed a

XSL•FO
RenderX

minimum of one group think-aloud session for the cough tool and another for the sore throat tool with one exception. Site 4 did not complete a cough CDS testing session. Sessions were conducted with the EHR encoded tool within a sandbox environment (for the Epic EHR sites) or a Research Electronic Data Capture (REDCap) mock-up of the tool (for the Allscripts TouchWorks EHR site). Each session was structured to include an individual think-aloud component with 1 to 4 nurse participants (and in one session, a medical assistant) simultaneously, followed by a focus group component with all participants that included debriefing questions about the tool. Results were documented in a pretemplated format for rapid dissemination to the other study sites to enable the efficient integration of feedback prior to the next site's testing session. Sessions were completed remotely on Microsoft Teams, Cisco Webex, or Zoom depending on site requirements.

Study team moderators were different for each site due to Institutional Review Board regulations (ie, study team members at each site could only conduct testing with their own site's participants). The main moderator presented an overview of the session and the tools. This included instructions on how to moderate think-aloud sessions. To obtain individualized feedback, each nurse worked with an additional session moderator in an individual web-based breakout room. Here, the nurse used the think-aloud method while navigating through the tool or a prototype of the tool. Each nurse used the tool or prototype on mock clinical cases. The main moderator circulated among the breakout rooms throughout the session to answer any questions, resolve any issues, and ensure that the session ran smoothly.

Following the individual breakout room think-aloud sessions, all nurse participants were brought back together into the main meeting room for a focus group to answer debriefing questions about their opinions of the tool. All participants were encouraged to speak, and individual participants were asked for their thoughts if they were not frequently volunteering their opinions. Sites shared their results with each other between sessions so that any necessary tool modifications could be made prior to the next testing session, enabling rapid iteration of the tools. Sites requiring more feedback following their initial sessions completed an additional session.

Snowballing Approach to Method Development

Prior to initiating our usability testing sessions, we assembled a reporting form template in PowerPoint that we could quickly fill in with structured feedback from our sessions. As each site ran their testing session, and moderators filled out the template with session results and shared it with the other participating sites. This enabled the other sites to make changes as needed before conducting their own usability testing sessions. In addition, the PowerPoint template included space for sites to share information about how to iteratively adjust the methods to efficiently run the sessions based on their experiences. Each site highlighted any challenges they had with completing the sessions and suggestions for how to avoid such challenges in future sessions, allowing sites to implement changes as needed. With the goal of rapidly iterating on our tools among sites, we did not use traditional in-depth thematic analysis in favor of quick feedback summaries. Our informal thematic analysis enabled us to quickly resolve the most easily addressable issues between sessions.

Results

A summary of our session durations and participant breakdowns is provided in Table 2.

XSL•FO
RenderX

Table . Time and participant breakdown of snowball group usability testing sessions.

Variable	Average value, mean	Site 1	Site 2		Site 3	Site 4[a]
		Session 1	Session 1	Session 2[b]	Session 1	Session 1
Cough clinical decision support testing						
Participants, n	3.3	3	4	—	3	—
Facilitators, n	4.7	4	4	—	6	—
Think-aloud time (min)	35.8	30	30 - 45[c]	—	40	—
Focus group time (min)	18	21	—[d]	—	15	—
Sore throat clinical decision support testing						
Participants, n	3	4	2	1	3	5
Facilitators, n	4	4	4	3	6	3
Think-aloud time (min)	36.4	27	40	45	30	40
Focus group time (min)	14.1	24	—[d]	5	17.5	10

[a]Site 4 did not complete usability testing sessions for the cough tool.

[b]Site 2 completed 2 sessions to test the sore throat tool.

[c]This range represents differences in the amount of time participants were able to take part in the session.

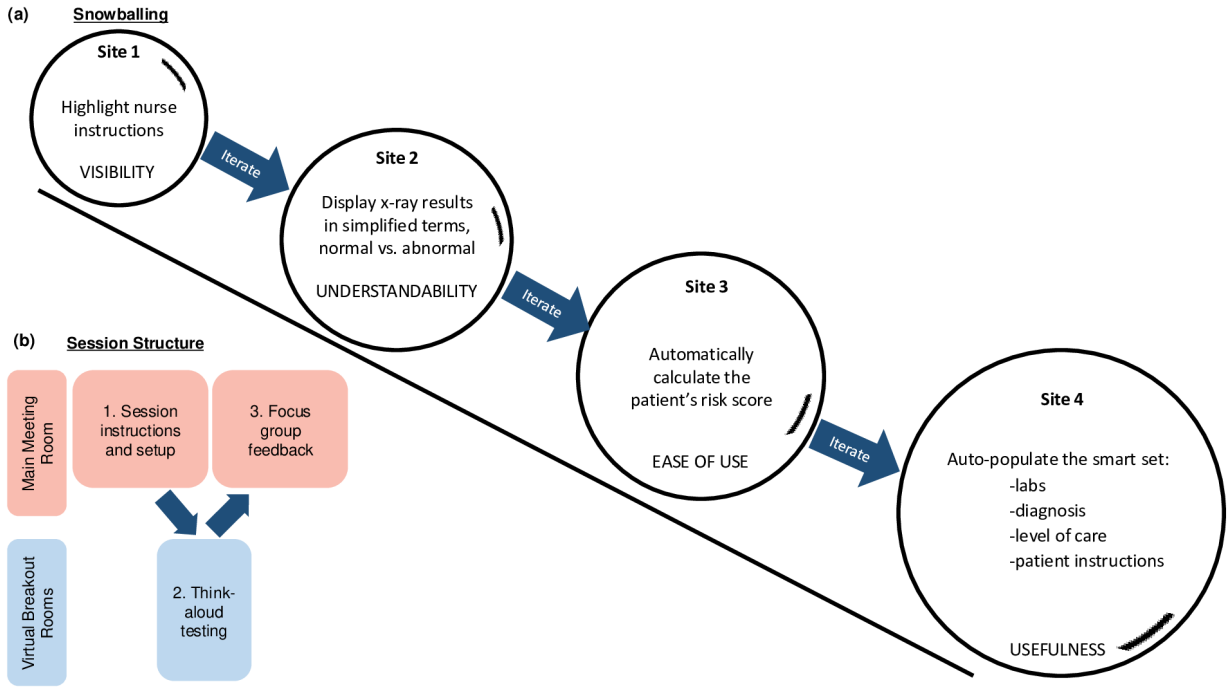
[d]This session did not include a focus group component.

Cough and Sore Throat CDS Tool Feedback

Snowball group usability testing enabled us to efficiently obtain and incorporate iterative feedback across our study sites (Figure 2). Overall, participants felt that the CDS tools were useful but identified some areas for potential improvement. Overarching themes from the sessions included suggestions for changes in the tool's wording and formatting, as well as changes to the overall workflow. For example, participants identified the wording of "recent travel" as unclear, as they were unsure what timeframe would constitute recent. Similarly, "disease exposure" contained ambiguity as to which diseases the exposure could encompass. In terms of workflow, participants at one site pointed out that it would not make sense for the tool to ask about COVID results during the visit, as this would have been a prerequisite to the patient coming in for a visit at the time this study was conducted. Each of these suggestions provided valuable feedback for modifications to be made prior to the next testing session.

XSL•FO
RenderX

Figure 2. Snowball group usability testing enabled efficient acquisition of iterative feedback across study sites. (**A**) Snowballing enabled iterative integration of feedback between testing sessions. Circles indicate feedback obtained at during each site's testing sessions. Snowballing is indicated by the increasing size of the circles from site to site. (**B**) Each session included both think-aloud testing and focus group components except where indicated in Table 2.



Focus Group Testing Component

The focus group testing component of our new method proved to be highly beneficial. We asked debriefing questions regarding participants' opinions of the tool in a focus group setting. One key feature was that, by collecting feedback from multiple participants at once rather than over a series of multiple separate sessions, we were more easily able to notice similarities in feedback among participants. In addition, each site's study team was able to meet soon after the single group session, rather than once after several sessions had been completed. This made it easier to discuss results, find similarities in participant feedback, and generalize, due to the recency of the information being obtained.

Snowballing Component

The snowballing component of our method also proved to be beneficial. A key benefit of snowballing was that study teams were able to incorporate changes based on previous sites' sessions. For example, nurses at one site did not properly follow instructions for the order of how to choose antibiotics based on patient allergies. The study team made changes to the cough and sore throat CDS tools at this site to highlight these instructions. These changes were also incorporated into the CDS tools under development at the 2 other sites that used Epic EHRs in the study. In addition, during the first site's testing session, it was determined that the location of the patient comorbidities, used to determine if a nurse visit was appropriate during triage, was not optimal. Comorbidity information was moved to below the nurse visit symptoms to improve ease of use. This feedback was incorporated into the second site's tool prior to their usability testing session.

While snowballing was useful for identifying and implementing certain changes to the CDS tools among sites, differences in the 2 EHRs and their respective limitations precluded some tool components from being transferable between sites. For example, while the Epic EHR versions of the tools were able to automatically calculate results, the TouchWorks EHR versions lacked this capability due to system limitations. In addition, a key feature of the Epic CDS tool was that it could link to order sets, enabling the user to seamlessly proceed from the CDS to the appropriate orders for a patient. TouchWorks did not support connection to order sets, and therefore, the transition from the CDS to ordering in the TouchWorks versions of the CDS tools was more manual. This, unfortunately, could not be changed. Finally, differences in primary care and urgent care workflows precluded uniformity in our approaches in these two environments.

Discussion

Principal Findings

Snowball group usability testing is a novel method that can be used to rapidly and iteratively test the usability of a new CDS tool or workflow. To our knowledge, this is the first time usability testing has been performed in this manner. We identified 3 main benefits of snowball group usability testing. First, by interviewing several participants in a single session rather than individuals over the course of weeks, each site was able to rapidly obtain a large amount of data at once. With traditional testing approaches, in which one participant is interviewed at a time and each must be scheduled separately, it may take weeks or even months to obtain the same amount of data. Importantly, snowball group usability testing helped to

XSL•FO
RenderX

expedite the overall usability testing timeline. By using our method sequentially (or in limited cases, concurrently) and at different sites, we were able to rapidly make changes, then retest our CDS tools with the next group of participants.

Second, combining the individualized think-aloud component with a focus group component in the same session helped us obtain a variety of feedback. In a focus group, one participant's feedback can become amplified, as a more talkative participant may dominate the conversations, and others may simply agree with their thoughts. Similar to any other focus group, session moderators encouraged all nurses in the group to contribute to the conversation, calling out individuals if they had not provided feedback after a while. By combining methodologies, we first obtained in-depth individual-level feedback with the think-aloud sessions and then elicited group-based feedback with the focus groups when participants built upon each other's opinions. Both types of feedback were useful; in the think-aloud sessions, the study team mostly observed, whereas in the focus group, the study team was more involved in asking for feedback on how to improve the tool. Therefore, combining approaches enabled us to get different types of feedback from the same set of participants all at once, which was invaluable when considering tight research and development timelines.

Third, our method enabled us to identify new and varied feedback from each site because we were able to edit key issues with our tools between sessions. Rather than hearing the same feedback in each session and making all of the changes at the end, our method resulted in a greater variety of feedback and a greater ability to optimize the tools.

Expediting the timeline and maximizing the amount of feedback gained are particularly valuable when developing EHR tools due to the inherent time constraints in the development process. Working with multiple sites and internal teams, such as those in charge of coding and implementing changes to the EHR, poses its own timing challenges as a result of the need to coordinate schedules and workloads. Thus, usability testing is often not considered a top priority; however, it remains an essential component of tool and system development. The use of snowball group usability testing effectively addresses this problem by reducing the time it takes to obtain and share feedback among participating sites. We also performed our testing remotely, which further addressed the issue of tight timelines and schedules, as participants did not have to travel from their clinics to be in the same location as other participants and the study team for the testing sessions. This enabled participants from different clinics to participate at the same time, without taking extra time out of their busy schedules for travel. Of note, remote testing can be used to mitigate scheduling issues with other forms of usability testing, including single-participant think-aloud sessions, interviews, and focus groups.

Limitations

There were challenges and limitations with snowball group usability testing. We experienced some incompatibility between different EHR systems, limiting transferability. Although the content of the CDS tools was transferable, the overall structure of the note and workflow were quite different. In addition, there were some components of the tools (such as the automatic calculator and linking to order sets) that were possible in one system but not the other, requiring site-specific modifications and testing.

In addition, timing, staffing, and availability issues presented a challenge. There were variations in each site's experience with usability testing. Finding 4 - 5 experienced usability testing moderators for a given session was challenging for some sites. As a result, staff needed to be trained in the methodology, but identifying times when staff were available was challenging. Another site experienced research staff turnover at the beginning of the usability testing phase of our study, leading to a need for extra training and reshuffling of resources. Additionally, nurses participating in the sessions had limited time available to complete their sessions and the study teams were unable to get through all of their planned cases and questions. For example, in one session, only think-aloud testing was completed due to time constraints, as the nurses had competing clinical demands. In the same session, one of the nurse participants was frequently interrupted by medical staff and phone calls. Two sites cut the think-aloud session short, setting a time limit to ensure that the focus group component was included. Additional challenges with time included accounting for the different speeds at which participants gave feedback and a lack of opportunity for the study team to iteratively improve the timing issues from session to session, as most sites only conducted one. However, learning from earlier sessions, we were able to change future sessions so that each nurse started off with a different mock case during their respective think-aloud sessions. This ensured that each case was tested, even though each nurse did not complete every case themselves.

Finally, an additional limitation was the fact that, due to the pragmatic nature of our study, we did not perform highly rigorous qualitative or quantitative evaluations of our usability testing results. Future testing of this method should include validated usability measures, such as the System Usability Scale [19], in all rounds of testing, as well as qualitative thematic analysis to rigorously assess the impact of snowball group usability testing on the usability of the CDS tools being developed. However, based on our collective years of experience performing usability testing, the suggestions and improvements derived of from snowball group usability testing were on par with what we would expect to obtain from established usability testing techniques, and the quality of our findings was higher than what we have seen using previous testing methods.

Conclusion

Snowball group usability testing provides a novel, efficient method of obtaining feedback on newly developed tools and systems, while addressing barriers typically associated with traditional usability testing methods. We successfully developed and used this method to test two CDS tools and rapidly iterated the process among 4 sites. Snowball group usability testing can be applied to test a wide variety of tools and workflows, including CDS tools, prior to launch so that they can be efficiently optimized, ultimately leading to higher adoption and end-user satisfaction.

XSL•FO
RenderX

Acknowledgments

This work was supported by grants 2R01AI108680-07A1 from the National Institute of Allergy and Infectious Diseases and K23HL145114 from the National Heart, Lung, and Blood Institute. Generative artificial intelligence was not used in any portion of the manuscript writing or for any part of this study.

Data Availability

Data for this study are limited to session recordings and notes. Deidentified usability testing data will be provided upon request.

Conflicts of Interest

KLDD's spouse works at the NYU IRB. The remaining authors declare no conflict of interest.

References

1. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004 Feb;37(1):56-76. [doi: 10.1016/j.jbi.2004.01.003] [Medline: 15016386]
2. Richardson S, Mishuris R, O'Connell A, et al. "Think aloud" and "near live" usability testing of two complex clinical decision support tools. Int J Med Inform 2017 Oct;106:1-8. [doi: 10.1016/j.ijmedinf.2017.06.003] [Medline: 28870378]
3. Press A, McCullagh L, Khan S, Schachter A, Pardo S, McGinn T. Usability testing of a complex clinical decision support tool in the emergency department: lessons learned. JMIR Hum Factors 2015 Sep 10;2(2):e14. [doi: 10.2196/humanfactors.4537] [Medline: 27025540]
4. Li AC, Kannry JL, Kushniruk A, et al. Integrating usability testing and think-aloud protocol analysis with "near-live" clinical simulations in evaluating clinical decision support. Int J Med Inform 2012 Nov;81(11):761-772. [doi: 10.1016/j.ijmedinf.2012.02.009] [Medline: 22456088]
5. Dauber-Decker KL, Basile M, King D, et al. Developing a decision aid to facilitate informed decision making about invasive mechanical ventilation and lung transplantation among adults with cystic fibrosis: usability testing. JMIR Hum Factors 2021 Apr 14;8(2):e21270. [doi: 10.2196/21270] [Medline: 33851921]
6. Gill P, Stewart K, Treasure E, Chadwick B. Methods of data collection in qualitative research: interviews and focus groups. Br Dent J 2008 Mar 22;204(6):291-295. [doi: 10.1038/bdj.2008.192] [Medline: 18356873]
7. Rosenthal M. Qualitative research methods: why, when, and how to conduct interviews and focus groups in pharmacy research. Curr Pharm Teach Learn 2016 Jul;8(4):509-516. [doi: 10.1016/j.cptl.2016.03.021]
8. Antibiotic resistance threats in the United States, 2019. : Centers for Disease Control and Prevention; 2019 Nov. [doi: 10.15620/cdc:82532]
9. Dadgostar P. Antimicrobial resistance: implications and costs. Infect Drug Resist 2019;12:3903-3910. [doi: 10.2147/IDR.S234610] [Medline: 31908502]
10. Bush K, Courvalin P, Dantas G, et al. Tackling antibiotic resistance. Nat Rev Microbiol 2011 Nov 2;9(12):894-896. [doi: 10.1038/nrmicro2693] [Medline: 22048738]
11. Nadimpalli ML, Chan CW, Doron S. Antibiotic resistance: a call to action to prevent the next epidemic of inequality. N Med 2021 Feb;27(2):187-188. [doi: 10.1038/s41591-020-01201-9] [Medline: 33462445]
12. Fleming-Dutra KE, Hersh AL, Shapiro DJ, et al. Prevalence of inappropriate antibiotic prescriptions among US ambulatory care visits, 2010-2011. JAMA 2016 May 3;315(17):1864-1873. [doi: 10.1001/jama.2016.4151] [Medline: 27139059]
13. Havers F, Thaker S, Clippard JR, et al. Use of influenza antiviral agents by ambulatory care clinicians during the 2012-2013 influenza season. Clin Infect Dis 2014 Sep 15;59(6):774-782. [doi: 10.1093/cid/ciu422] [Medline: 25034419]
14. Heckerling PS, Tape TG, Wigton RS, et al. Clinical prediction rule for pulmonary infiltrates. Ann Intern Med 1990 Nov 1;113(9):664-670. [doi: 10.7326/0003-4819-113-9-664] [Medline: 2221647]
15. Fine AM, Nizet V, Mandl KD. Large-scale validation of the Centor and McIsaac scores to predict group A streptococcal pharyngitis. Arch Intern Med 2012 Jun 11;172(11):847-852. [doi: 10.1001/archinternmed.2012.950] [Medline: 22566485]
16. McIsaac WJ, White D, Tannenbaum D, Low DE. A clinical score to reduce unnecessary antibiotic use in patients with sore throat. CMAJ 1998 Jan 13;158(1):75-83. [Medline: 9475915]
17. Centor RM, Witherspoon JM, Dalton HP, Brody CE, Link K. The diagnosis of strep throat in adults in the emergency room. Med Decis Making 1981;1(3):239-246. [doi: 10.1177/0272989X8100100304] [Medline: 6763125]
18. Briggs J. In: Magee SW, McAvey MM, Conville T, editors. Telephone Triage Protocols for Nursing, 5th edition: Wolters Kluwer Health; 2015.
19. Brooke J. SUS: A "quick and dirty" usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry: Taylor & Francis; 1996.

XSL•FO
RenderX

Abbreviations

CDS: clinical decision support
EHR: electronic health record
REDCap: Research Electronic Data Capture

Edited by A Mavragani; submitted 11.12.23; peer-reviewed by D Chrimes, J Brooke; revised version received 29.10.24; accepted 30.10.24; published 18.02.25.

Please cite as:
Dauber-Decker KL, Feldstein D, Hess R, Mann D, Kim EJ, Gautam-Goyal P, Solomon J, Khan S, Malik F, Xu L, Huffman A, Smith PD, Halm W, Yuroff A, Richardson S
Snowball Group Usability Testing for Rapid and Iterative Multisite Tool Development: Method Development Study
JMIR Form Res 2025;9:e55316
URL: https://formative.jmir.org/2025/1/e55316
doi:10.2196/55316

© Katherine L Dauber-Decker, David Feldstein, Rachel Hess, Devin Mann, Eun Ji Kim, Pranisha Gautam-Goyal, Jeffrey Solomon, Sundas Khan, Fatima Malik, Lynn Xu, Ainsley Huffman, Paul D Smith, Wendy Halm, Alice Yuroff, Safiya Richardson. Originally published in JMIR Formative Research (https://formative.jmir.org), 18.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring the Effect of an 8-Week AI-Composed Exercise Program on Pain Intensity and Well-Being in Patients With Spinal Pain: Retrospective Cohort Analysis

Annika Griefahn[1,2,3], MSc; Florian Avermann[3], MSc; Christoff Zalpour[2], Prof Dr Med; Robert Percy Marshall[4,5], MD; Inés Cordon Morillas, PhD; Kerstin Luedtke[1], Prof Dr

[1]Department of Physiotherapy, Institute of Health Sciences, University of Lübeck, Lübeck, Germany

[2]Faculty of Business Management and Social Sciences, Hochschule Osnabrück, Albrechtstraße 30, Osnabrück, Germany

[3]Evidence and Evaluation Department, medicalmotion GmbH, München, Germany

[4]Medical Department, RasenBallsport Leipzig GmbH, Leipzig, Germany

[5]Department of Orthopedic and Trauma Surgery, Martin Luther University Halle-Wittenberg, Halle, Germany

Corresponding Author:
Annika Griefahn, MSc
Department of Physiotherapy, Institute of Health Sciences, University of Lübeck, Lübeck, Germany

Abstract

Background: Spinal pain, one of the most common musculoskeletal disorders (MSDs), significantly impacts the quality of life due to chronic pain and disability. Physical activity has shown promise in managing spinal pain, although optimizing adherence to exercise remains a challenge. The digital development of artificial intelligence (AI)-driven applications offers a possibility for guiding and supporting patients with MSDs in their daily lives.

Objective: The trial aimed to investigate the effect of an 8-week AI-composed exercise program on pain intensity and well-being in patients with spinal pain. It also examined the relationship between exercise frequency, pain intensity, and well-being. In addition, app usage frequency was examined as a proxy for app engagement.

Methods: Data from users who met the inclusion criteria were collected retrospectively from the medicalmotion app between January 1, 2020, and June 30, 2023. The intervention involved the use of the medicalmotion app, which provides 3 - 5 personalized exercises for each session based on individual user data. The primary outcomes assessed pain intensity and well-being using the numeric rating scale (NRS) and the Likert scale. Data were collected at baseline (t0), 4 weeks (t1), and 8 weeks (t2). The correlation between exercise frequency, pain intensity, and well-being was analyzed as a secondary outcome. In addition, average session length and frequency were measured to determine app engagement. Statistical analysis included ANOVA and Spearman correlation analysis.

Results: The study included 379 participants with a mean age of 50.96 (SD 12.22) years. At t2, there was a significant reduction of 1.78 points on the NRS (P<.001). The score on the Likert scale for well-being improved by 3.11 points after 8 weeks. Pain intensity showed a negative correlation with the number of daily exercises performed at t1 and t2. Well-being had a small negative correlation with the average number of exercises performed per day. The average number of exercises performed per day was 3.58. The average session length was approximately 10 minutes, and the average interaction with the app was 49.2% (n=27.6 days) of the 56 available days.

Conclusions: Overall, the study demonstrates that an app-based intervention program can substantially reduce pain intensity and increase well-being in patients with spinal pain. This retrospective study showed that an app that digitizes multidisciplinary rehabilitation for the self-management of spinal pain significantly reduced user-reported pain intensity in a preselected population of app users.

Trial Registration: OSF Registries osf.io/kjhef; https://osf.io/kjhef

(JMIR Form Res 2025;9:e57826) doi:10.2196/57826

KEYWORDS

exercise; mHealth; app engagement; spinal pain; artificial intelligence; AI; intensity; well-being; mobile health; apps; applications; retrospective analysis; physical activity; adults; questionnaire

XSL•FO
RenderX

Introduction

Musculoskeletal disorders (MSDs) can affect muscles, tendons, cartilage, ligaments, and nerves [1]. MSDs are a major cause of chronic pain, physical disability, and loss of quality of life worldwide. They include more than 150 conditions and syndromes [2]. For example, one of the most common MSDs is spinal pain, and more specifically, low back pain (LBP) [2]. The lifetime prevalence of LBP is high, ranging from 75% to 85% [3]. Nonspecific LBP, characterized by the absence of a clear anatomical cause, accounts for approximately 90% of LBP cases [3]. Numerous risk factors for LBP have been identified, including age, obesity, sedentary lifestyle, poor posture, smoking, and psychosocial factors such as stress and depression [4]. These factors contribute to LBP's complex and multifactorial nature, including physical, psychological, occupational, and lifestyle influences [4].

Several studies have shown that regular physical activity is effective in reducing chronic spinal pain and is consistently recommended in clinical guidelines for the management of nonspecific LBP [5-8]. A systematic review by Jordan et al [9] assessed the effects of interventions that may improve adherence to exercise and physical activity. A total of 42 trials were included, mainly focusing on patients with knee osteoarthritis and spinal pain with relatively short follow-up measurements. Promising strategies to improve exercise adherence included supervised exercise, individualized exercise, refresher or follow-up sessions, the provision of supplementary materials such as audio- or videotapes with graded exercises, self-management programs, and cognitive behavioral techniques [9].

The Clinical Practice Guideline on Intervention for the Management of Acute and Chronic LBP suggests that specific complaints and causes should be considered when selecting exercises [10]. For example, for acute complaints involving leg pain, specific exercises to activate the trunk muscles and improve the strength as well as the endurance of the back muscles should be considered. For chronic spinal pain, exercises designed to improve movement control should be included [10].

The increasing digitalization of health care and therapeutic services allows patients to use various mobile and web applications to manage their conditions [11]. Digital solutions provide a scalable and widely accessible approach, enabling the management of spinal pain in rural areas where the availability of physiotherapists is limited [12]. Digital therapeutic care applications mostly include video-based treatment programs and educational materials [13]. Not surprisingly, the number of studies investigating digital interventions has increased in recent years, including studies on the effectiveness of digital exercises in patients with spinal pain [14-18].

Another aspect of digitalization is integrating artificial intelligence (AI) methods into digital therapy applications. AI can be used to better address the heterogeneity of patients with MSDs and to personalize therapy. The use of AI in health care, particularly to support diagnosis, has increased in recent years [19]. Few studies have investigated the effects of AI-based exercises [20-22]. In addition, maintaining an exercise routine is an important factor in reducing the risk of recurrence [7,23,24]. Nearly 70% of people who recover from spinal pain experience another episode within 12 months [25]. Most maintenance strategies involve supervised group exercise, with or without equipment. However, this can be costly for the individual and the health care system [7]. Therefore, an approach using digital health applications and AI offers a new perspective to increase individual adherence to exercise at a lower cost, which could reduce the risk of recurrent pain episodes in the future [26].

To increase knowledge and provide additional evidence on the use of digital applications, the following four research questions were defined: (1) What are the effects of an 8-week app-based AI-composed exercise program in patients with spinal pain regarding pain intensity? (2) What are the effects of an 8-week app-based AI-composed exercise program on patients with spinal pain regarding their well-being? (3) Is there a correlation between the frequency of exercise and pain intensity and well-being? (4) What are the parameters of app engagement, in terms of average session duration and frequency, for patients with spinal pain in an 8-week app-based AI-composed exercise program?

Methods

Ethical Considerations

The study was conducted in accordance with the ethical principles of the Declaration of Helsinki. All procedures and materials used in this study were approved by the ethics committee of the University of Applied Sciences Osnabrueck (HSOS/2022/1/2). The data provided by medicalmotion GmbH were anonymized and stored on a password-protected server at the University of Applied Sciences Osnabrueck. All users consented to the collection of data by agreeing to the terms and conditions of the use of medicalmotion.

Trial Design

The study is reported following the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guideline (Multimedia Appendix 1) [27]. The study was registered a priori via the Open Science Framework [28]. The study is a retrospective analysis of user data. Users were recruited to the medicalmotion app between January 1, 2020, and June 30, 2023, through various recruitment opportunities, such as health insurance companies or physiotherapists.

Participants

Inclusion criteria were as follows: (1) men and women who reported spinal pain (including upper and lower back) in the initial therapeutic questionnaire, (2) individuals aged between 18 and 65 years with access to the medicalmotion app, and (3) participants who had at least 8 interactions with the medicalmotion app during the 8-week evaluation period. Exclusion criteria were as follows: (1) persons diagnosed with a neurological or mental illness, (2) persons currently suffering from an infection or systemic disease, or (3) persons regularly taking medication for mental illness. The sample for this study consisted of all enrolled participants who met the inclusion criteria.

XSL•FO
RenderX

Intervention

The medicalmotion GmbH offers a comprehensive mobile and web application for various pain conditions. The mobile app not only offers AI-based exercises but also includes a wide range of additional features, including relaxation exercises, podcasts, a chat function, and a health cockpit, which tracks the user's condition throughout use. The AI-chosen exercises consider medical history data, including pain characteristics, lifestyle information, and well-being. In addition, user feedback from previous exercise sessions is incorporated into the composition process to ensure a personalized approach.

The system dynamically selects the most appropriate exercises based on real-time data such as the user's pain location and intensity, comfort level, and feedback from previous exercises. It creates an exercise needs profile and matches it with exercises from an extensive database. This adaptive approach allows adjustments for acute pain or changing circumstances. The user has the option to choose the number of daily exercises (3 to 5), all of which are delivered as real-time audio-based exercise videos. Each exercise has an average length of 167 seconds. The system also identifies potential causes of pain and adapts the exercise composition accordingly, ensuring a personalized exercise program with transparent traceability. Its training repertoire consists of a selection of 300 exercises. These exercises are categorized into 3 groups: release, mobility, and strength. Release exercises aim to induce muscle and tissue relaxation. Mobility exercises involve the full spectrum of body movement with complex movement patterns. Strength exercises strengthen muscles, tendons, and supporting tissues. An example of each type of exercise is included in Multimedia Appendix 2.

Outcomes

The primary outcome measure was pain intensity. An 11-point numeric rating scale (NRS) assessed participants' pain intensity. A change of 1 - 2 points was considered clinically relevant [29]. The secondary outcome was well-being, measured using an 11-point Likert scale. A score of zero indicates poor well-being, while a score of 10 indicates perfect well-being. Data were collected at baseline (t0), 4 weeks (t1), and 8 weeks (t2). In addition, the following parameters were collected at t0 via a therapeutic questionnaire as part of the onboarding within the medicalmotion app: age, work mode, sports frequency, and sex. The initial pain areas' anatomical location and duration were recorded using a body map. Pain duration was categorized into chronic, subacute, and acute. The anatomical location of pain was recorded for the buttocks, lower back, upper back, and neck areas. Multiple responses were allowed for anatomical location. Data collection included the weekly training volume, the number of training sessions with more than one exercise, the number of training sessions in which all exercises were completed, and the total number of exercises completed. The application evaluated the average length and frequency of sessions to assess engagement with the app. Data were only collected within the medicalmotion app.

Statistical Analysis

For the statistical analysis, an ANOVA with repeated measurement was conducted to examine the effects of the 8-week app usage on pain intensity and well-being. If the assumption of sphericity was not met, the Greenhouse-Geisser correction was used. For post hoc analysis, the Bonferroni correction was used.

In addition, Pearson correlation (r) between pain intensity, well-being, average number of exercises per day, total number of exercises, and skipped exercises was examined. If there was no linear correlation, Spearman correlation (ρ) was performed. The interpretation of the correlation coefficients was based on Cohen thresholds [30].

The average session length was calculated by multiplying the average number of exercises performed by the average exercise duration of 167 seconds. Session frequency was calculated by dividing the total number of active days by the total number of available days (56 days). Session frequency was expressed as a percentage. Session length was given in seconds and minutes.

The significance level was set at P=.05. The effect size was described by partial η^2 [30]. Results are reported as mean, SD, minimum, maximum, CIs, F value, and df. Statistical analysis of the data was performed using IBM SPSS Statistics version 29.

Results

Descriptive Results of the Participants

The total sample included 379 participants, of whom 138 (36.4%) were men and 241 (63.6%) were women. The mean age was 50.96 (SD 12.22) years. Data on mode of work, frequency of sport, initial pain areas, and completed exercises per day are shown in Table 1.

XSL•FO
RenderX

Table . Descriptive results of the cohort.

Variables			Values
Categorical variables			
	Participants, n (%)		
		Female	241 (63.6)
		Male	138 (36.4)
	Work mode, n (%)		
		Sitting and standing	165 (43.5)
		Sitting	177 (46.7)
		Hard work	17 (4.5)
		Standing	20 (5.3)
	Sport frequency, n (%)		
		Never	59 (15.6)
		1 - 3 times per week	248 (65.4)
		>3 times per week	72 (19)
	Initial pain areas (anatomical location), n (%)		
		Lower back	229 (60.4)
		Neck	201 (53)
		Buttocks	109 (28.8)
		Upper back	184 (48.5)
Continuous Variables			
	Age (years), mean (SD)		50.96 (12.22)
	Number of initial areas of pain based on the duration of the pain, mean (SD)		
		Total	4.86 (4.02)
		Chronic pain areas (>6 months)	2.43 (3)
		Subacute pain areas (>1 week and <6 months)	1.52 (2.18)
		Acute pain areas (<1 week)	0.91 (2.33)

Pain Intensity

At t0, the intervention group showed a mean pain intensity of 6.08 (SD 2.16) on the 0 - 10 NRS (Table 2). After 8 weeks (t2), the intervention group showed a mean reduction of 1.78 (SD 2.05) points on the 0 - 10 NRS (Figure 1). The factor “time” was significant ($F_{1.88,710.96}$=179.861, P<.001) and showed a large effect with η^2=0.32. Post hoc analysis with Bonferroni correction showed a significant difference at each time point with P<.001 (Table 2).

XSL•FO
RenderX

Table . Statistical analysis of pain intensity and well-being.

Measurement points	Mean (SD)	Mean difference (SD)		Percentage difference, %		95% CI		Min	Max	*P* value	Partial η^2[a]
		t0[b]	t1[c]	t0	t1	Lower	Upper				
Pain intensity (NRS[d])										<.001	0.32
t0	6.08 (2.16)	—[e]	—	—	—	5.85	6.3	1	10	—	—
t1	5.54 (2.23)	0.53 (1.9)	—	8.3	—	5.32	5.77	0	10	—	—
t2[f]	4.3 (2.28)	1.78 (2)	1.25 (1.6)	29.3	22.4	4.07	4.52	0	10	—	—
Well-being (Likert scale)										<.001	0.64
t0	4.95 (1.65)	—	—	—	—	4.8	5.1	2	10	—	—
t1	7.07 (1.43)	2.1 (1.8)	—	42.8	—	6.92	7.22	4	10	—	—
t2	8.06 (1.38)	3.1 (1.7)	1 (1.5)	62.8	14	7.9	8.2	4	10	—	—

[a]η^2 = effect size.

[b]t0: baseline.

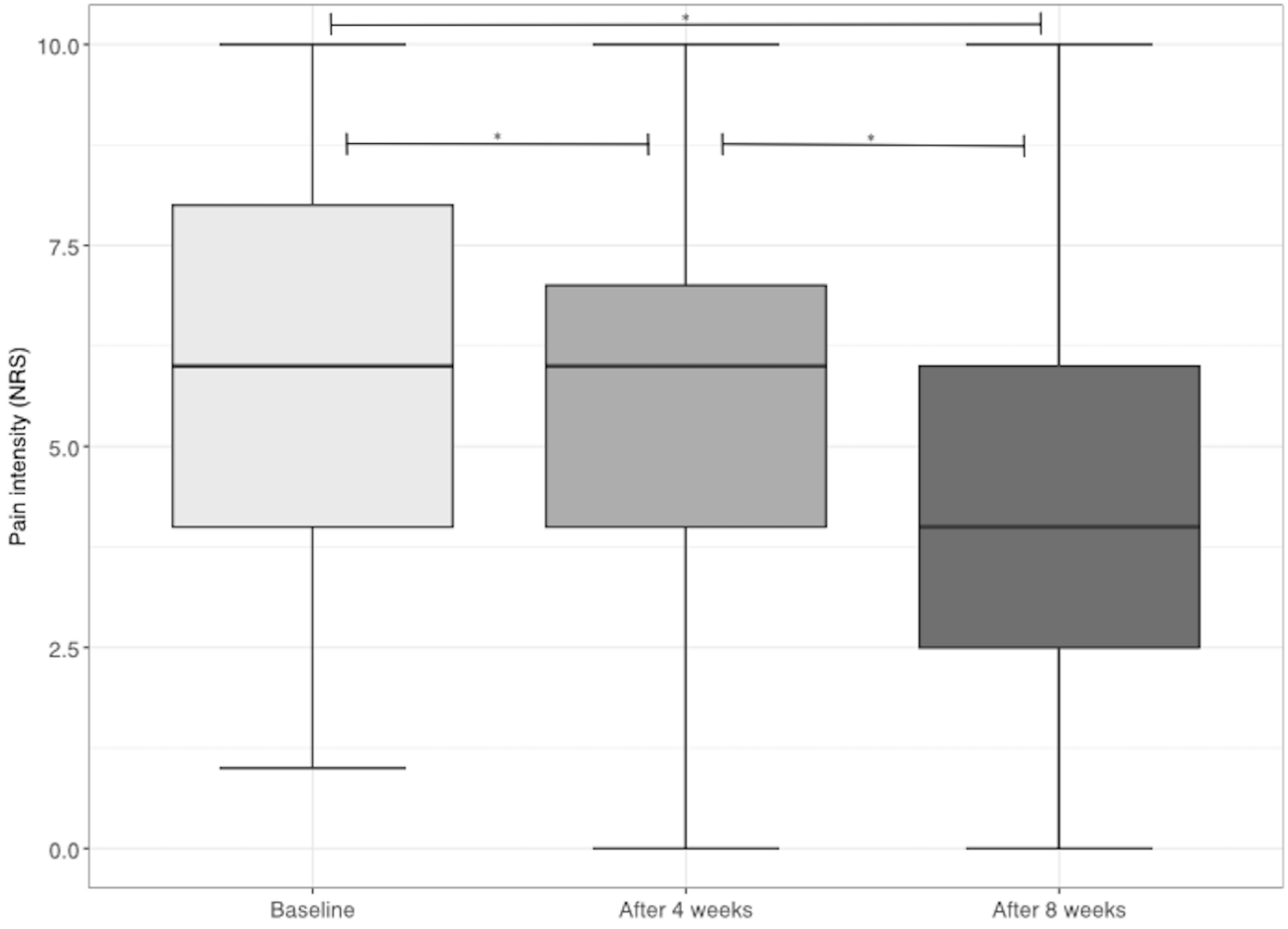
[c]t1: after 4 weeks.

[d]NRS: numeric rating scale.

[e]Not applicable.

[f]t2: after 8 weeks.

Figure 1. Changes in pain intensity over time. An asterisk (*) denotes a significant difference. NRS: numeric rating scale (from 0 to 10); lower whisker: minimum; upper whisker: maximum.

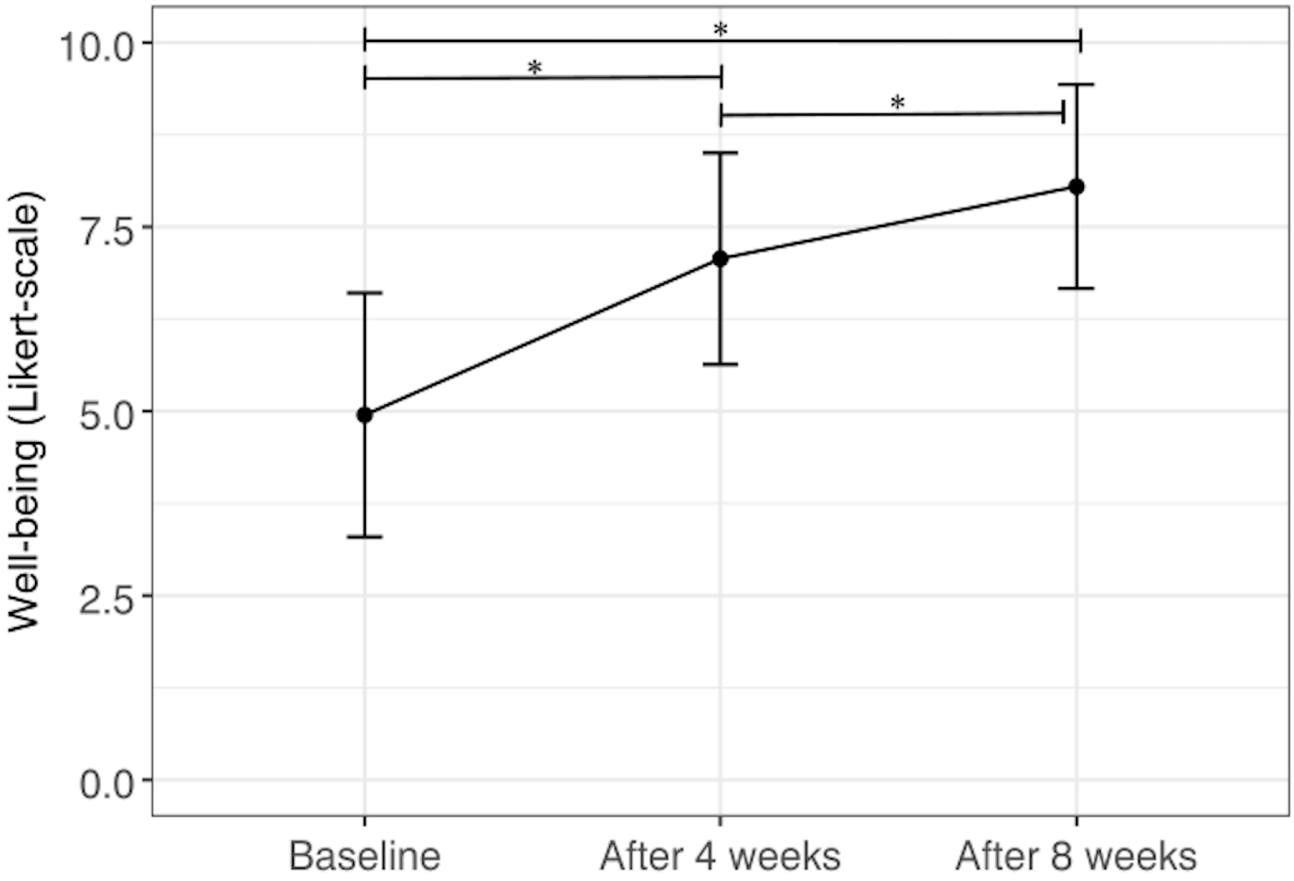


XSL•FO
RenderX

Well-Being

At t0, mean well-being was 4.95 (SD 1.65) on the 11-point Likert scale (Table 2). After 8 weeks (t2), the well-being improved by 3.11 points (Figure 2). The factor "time" was significant ($F_{1.94,732.57}$=671.97, P<.001) and showed a large effect with η^2=0.64. Post hoc analysis with Bonferroni correction showed a significant difference at each time point with P<.001 (Table 2).

Figure 2. Changes in well-being over time are shown with mean and SD. An asterisk (*) denotes a significant difference.



App Engagement

The average session length was 597.86 seconds (~10 minutes). The medicalmotion app was used on an average of 49.2% (27.6 days) of the 56 available days. Table 3 shows the variables: active days, total number, and average number of exercises performed.

Table . Overview of the average activity within the medicalmotion app.

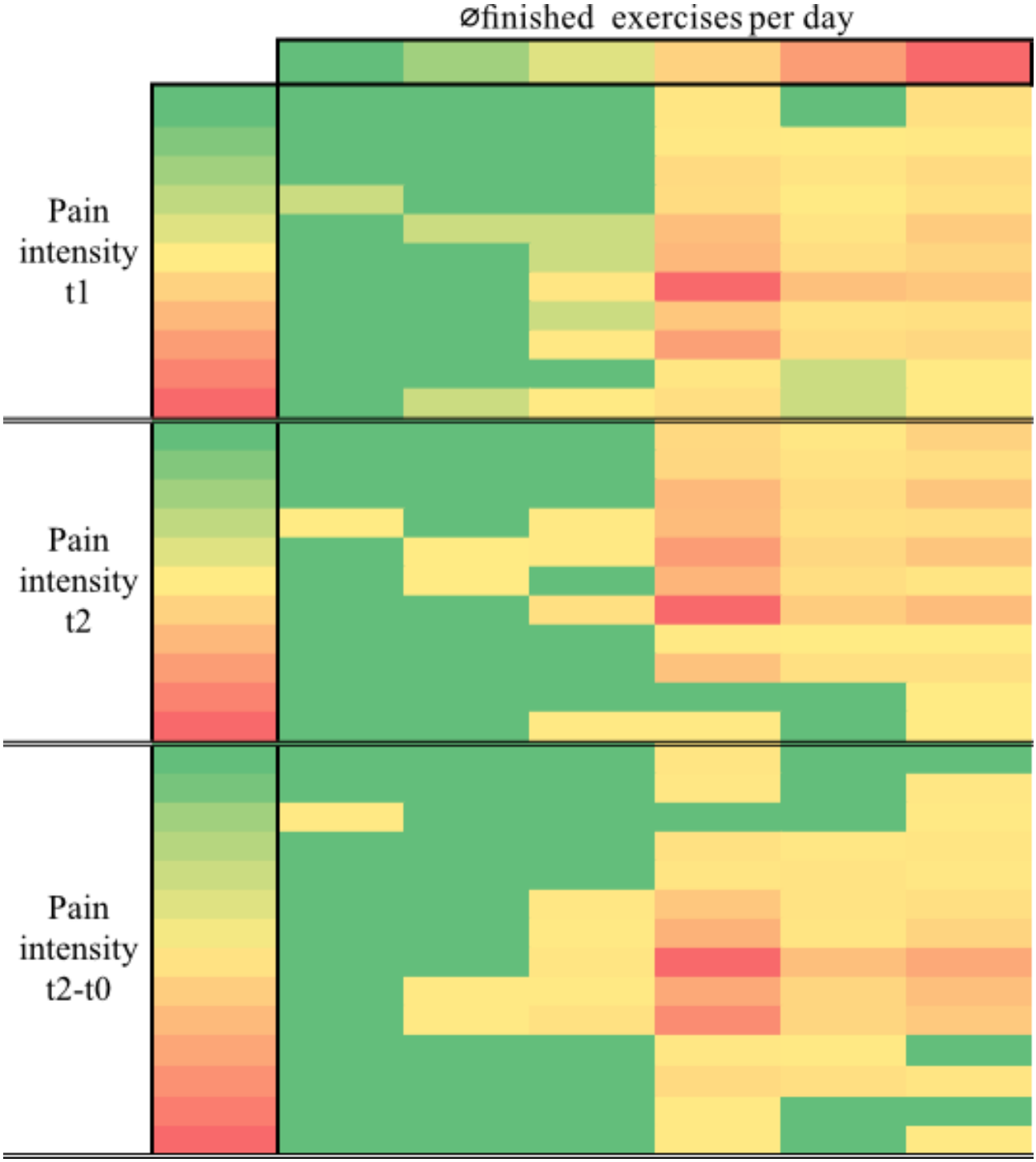
Variable	Mean (SD)	Min	Max
Active days within 8 weeks trial (days)	27.54 (10.98)	16	57
Finished exercises (number)	99.34 (49.9)	1	279
Skipped exercises (number)	1.36 (4.87)	0	67
finished exercises per day (number)	3.58 (0.9)	1	5

Correlation Between Pain Intensity and Exercise Completion

Pain intensity levels showed a statistically significant negative correlation with the average number of exercises performed per day at t1 (P<.001) and t2 (P=.004). This suggests that as pain intensity increased, the number of exercises performed per day tended to decrease. Conversely, as pain intensity decreased, the number of exercises performed per day tended to increase. A similar, although slightly weaker, negative correlation was observed between pain intensity at t2 and the total number of exercises completed (P=.04). Notably, no statistically significant correlation was found between missed exercises and pain intensity. Details of the correlation coefficients and P values can be found in Multimedia Appendix 3. Figure 3 shows the relationship between the pain intensity at t1 and t2, the difference between t2 and t0, and the average number of exercises per day using a heat map.

XSL•FO
RenderX

Figure 3. Heatmap of the Spearman correlation between pain intensity and average finished exercises per day. The x-axis represents the average number of exercises per day (0 - 5), and the y-axis represents pain intensity (0 - 10). The color scale from green to yellow to red represents the percentage distribution. Green means a low percentage, yellow a medium percentage, and red a high percentage.



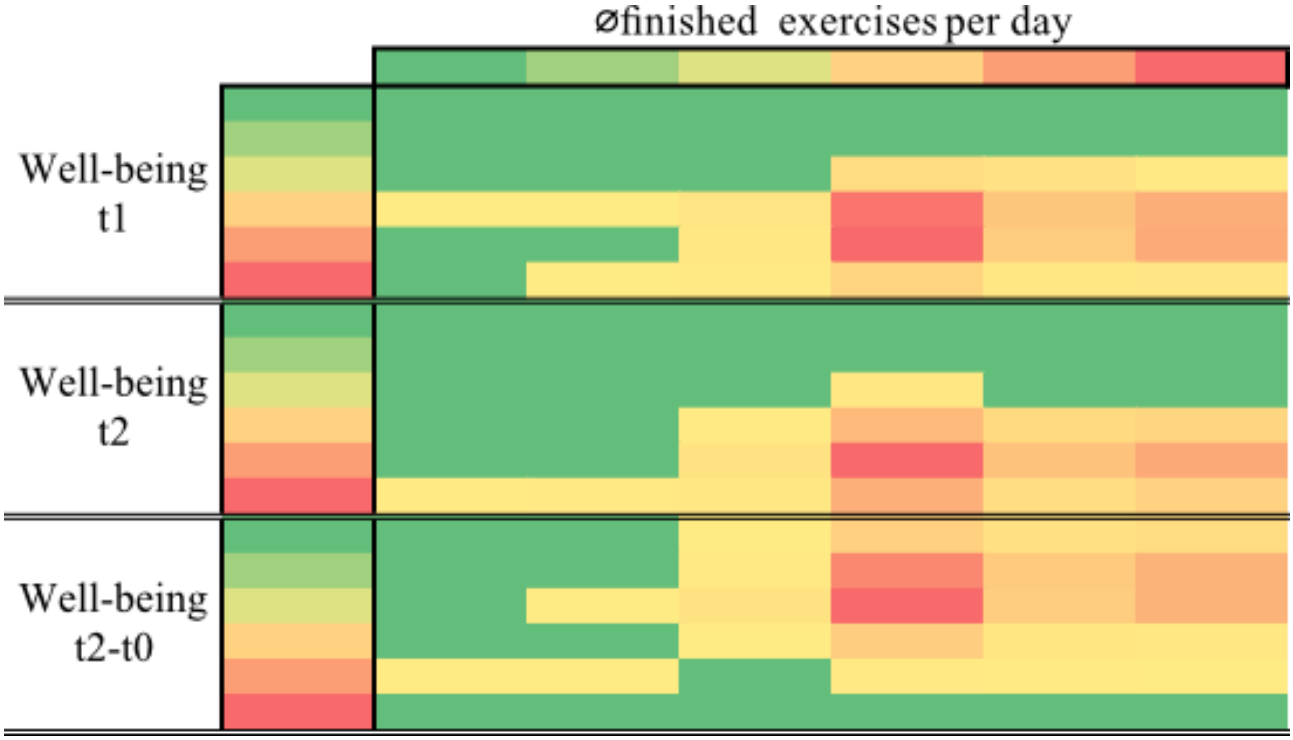
Correlation Between Well-Being and Exercise Completion

The average daily number of exercises completed showed a statistically significant, albeit small, negative correlation with the change in well-being over 8 weeks (P=.01). This means that as well-being increased, the number of exercises completed tended to decrease. Conversely, as well-being decreased, the average number of exercises completed tended to increase.

Figure 4 shows the relationship between the well-being at t1 and t2, the difference between t0 and t2, and the average number of exercises per day using a heat map.

XSL•FO
RenderX

Figure 4. Heatmap of the Spearman correlation between well-being and finished exercises per day. The x-axis represents the average number of exercises per day (0 - 5), and the y-axis represents well-being (0 - 10). The color scale from green to yellow to red represents the percentage distribution. Green means a low percentage, yellow a medium percentage, and red a high percentage.



Discussion

Principal Results

This study aimed to evaluate the effect of an app-based, AI-composed exercise program on pain and well-being in patients with spinal pain. A significant effect size (η^2=0.32) was achieved with a clinically relevant reduction in pain intensity of 1.78 points. A similar picture emerges with regard to well-being. The observed change in well-being of 3.11 points represents a substantial improvement. In addition, the large effect size (η^2=0.64) indicates that a significant proportion of the variation in well-being is attributable to the intervention. The study showed significant negative correlations between pain intensity levels at t1 (P<.001) and t2 (P=.004) and the average number of daily exercises, meaning that a higher pain intensity was associated with reduced daily exercise. A weaker negative correlation was observed between pain intensity at t2 and the total number of exercises performed (P=.04). The graph shows that both too much and too little exercise did not result in a significant change in pain intensity.

In addition, the average number of exercises completed per day showed a small but significant negative correlation with the change in well-being over 8 weeks (P=.01), suggesting that increased well-being was associated with a tendency to exercise less per day. In terms of session frequency, the medicalmotion app was used on 49.2% (27.6 days) of the 56 available days. With an average exercise duration of 167 seconds, this cohort's average use time was approximately 10 minutes per 8-week session.

Limitations

While this study provides valuable insights, its limitations require careful consideration. One limitation of the AI-based app is that it categorizes the exercises into strength, mobility, and release. However, it needs to be made clear whether these categories or the individualized plan contributed significantly to the observed effects. Future studies should focus on a more detailed evaluation of the specific effects of AI on outcomes.

Pain intensity (NRS) and well-being (Likert scale) were treated as metrically scaled in the statistical analysis, allowing for the calculation of repeated measures ANOVA. Therefore, the results are presented with the mean and SD. Based on the study by Nair and Diwan [31], parametric tests can be used for ordinal scaled data (pain intensity) if a normal distribution is present. This fact should be taken into account when interpreting the results.

Because of the broad inclusion criteria used, there is variability in the participants. This variability results from the wide age range and the broad categorization of spinal pain as an inclusion criterion. As a result, the population may have included individuals with a medical diagnosis of spinal pain as well as those with no diagnosis or an inconclusive diagnosis. This diversity in the study population could affect the influence of the app on the participants. In order to increase the specificity of the results and improve the assessment of effectiveness, it may be beneficial to introduce a differentiation based on diagnostic categories.

Another limitation is the data collection period, from 2020 to 2023. The COVID-19 pandemic that was ongoing during this period, with its peak years in 2020 and 2021, may have confounded the results. During this time, the amount of physical

XSL•FO
RenderX

activity and opportunities for public exercise in gyms were limited. In addition, many people were experiencing psycho-emotional stress, which may have affected their perception of pain. Finally, the long-term effects of COVID-19 infection can cause heterogeneous symptom complexes, which may further influence the physical complaints of the participants.

A further limitation of this study is the short duration of the intervention, which was limited to 8 weeks. Such a short period of time limits the conclusions that can be drawn about the long-term effects of the intervention and the sustainability of the results. In particular, aspects such as long-term compliance and the associated changes in exercise frequency could not be adequately assessed during this short observation period. Future studies with longer follow-ups are needed to make reliable statements about the long-term benefits and stability of the intervention.

In the context of these limitations, it is important to acknowledge that this study has inherent drawbacks, including potential selection bias, lack of a control group, and reliance on self-reported data. These limitations may introduce bias and affect the generalizability of the findings. The retrospective cohort study design also has its own set of drawbacks, including potential data quality issues and variability in user recruitment. In particular, the lack of control for concomitant pain-relieving interventions, such as physiotherapy, adds complexity and requires caution in drawing definitive conclusions. Consequently, the results of the study should be used primarily for hypothesis generation, with validation sought through prospective study designs to address these limitations and provide a more comprehensive understanding of the issue.

Comparison to Prior Work

These results on pain intensity reduction can be compared with home-based exercise therapy, as shown in a review by Quentin et al [32], who demonstrated a reduction in pain intensity in nonspecific LBP and a reduction in functional impairment with home-based exercise therapy and also reported a high effect size. A study by Weise et al [33] compared the effects of digital exercise therapy with physiotherapy. This study also showed a reduction in pain intensity of 2.92 points after 6 weeks in the intervention group. The results of this study showed similar effects. However, the added value of the AI-composed exercises cannot be deduced from these results. Therefore, it cannot be concluded whether the individualized exercises provided by the mobile app have added value in reducing pain or whether simply exercising induced the effect. However, a study by Cimarras-Otal et al [34] showed that exercises tailored to the needs of workers had a better outcome than general exercises.

The interaction between well-being and chronic pain has been reported previously [35]. For example, improvements in well-being have been shown to lead to reductions in pain catastrophizing and depression [35,36]. Based on the association between exercise and increased subjective well-being, the exercises provided in this study also influenced the psychological aspects of chronic pain.

Although a correlation only indicates an association and does not imply causation, the results of the correlation analysis can be interpreted in several ways. On the one hand, a reduction in pain could be the reason why users exercised more. In other words, more exercise or movement became possible with less pain. On the other hand, increased practice could also cause pain reduction; that is, users could reduce their pain through practice. Further studies are therefore needed to establish a causal relationship and the ideal frequency of exercise. The relationship between exercise frequency and changes in pain intensity was demonstrated in a large-scale study by Marshall et al [37]. Interestingly, the frequency of exercise was very similar between participants with significant improvement and those with deterioration. This fact may be due to the inclusion of different types of exercise, making it difficult to assess the effect of specific types of exercise. Therefore, an individualized training program for each person may be necessary to improve pain intensity successfully [37].

Another factor that may influence exercise frequency may be motivation. Increased pain and decreased well-being could both increase and decrease motivation to exercise. On the one hand, if you are in pain, you need to do something about it, or conversely, if you are in pain, there may be a fear or concern that exercise will only worsen the condition. Both scenarios are plausible and require further investigation in future studies. In this context, motivation to exercise should be assessed to identify possible relationships between pain intensity and motivation.

A 1985 review by Dishman et al [38] identified several factors that influence exercise frequency. Self-motivation, anticipated personal health benefits, perceived well-being, and enjoyment of physical activity positively contributed to the likelihood of participating in a supervised exercise program. Conversely, factors such as mood disturbance, health concerns, and knowledge about exercise or health negatively influenced the likelihood. This observation underscores the importance of motivation to exercise and highlights its significant influence on the frequency of engagement in physical activity.

A study by Meyer et al [39] showed a positive association between the frequency of physical activity and self-rated health. This finding contrasts somewhat with the current results, as a negative association was found between an average number of daily exercises and a change in self-rated health over 8 weeks. One possible explanation for this discrepancy could be the distinction between well-being and self-rated health. Another explanation could be that the lower the number of complaints, the lower the compliance to continue exercising. Users may have felt better and consequently exercised less because they no longer felt the need to exercise. One possible explanation may be the influence of self-motivation on training frequency, which is already a known factor for training frequency [38]. The correlation cannot answer these considerations; hence, further research is needed to establish a causal relationship [38]. In addition, the method of assessing the frequency of physical activity may also contribute to the difference. In the study by Meyer et al [39], participants provided a self-assessment of physical activity using a questionnaire, in contrast to the objective values obtained from the app in this study. Another factor to consider for the discrepancy between Meyer et al [39] and this study's results is the study population. The

XSL•FO
RenderX

aforementioned study examined healthy participants and not a pain-specific population [39].

To date, few studies have reported on the frequency of use of eHealth products. In a study by Labinsky et al [40], 9 out of 39 participants reported having only used the applications once. A study by Tian et al [41] also found that the average usage time of common digital products was between 2 and 10 minutes. No reliable values could be found specifically for eHealth or mobile health products.

Conclusions

In summary, the study investigated the multiple effects of an 8-week, app-based, AI-composed exercise program on patients with spinal pain. It addressed critical dimensions, including pain intensity and well-being, while exploring the potential relationship between exercise frequency and these outcomes. The results highlight the importance of exercise frequency in influencing pain intensity and well-being, revealing a compelling interplay between engagement in prescribed exercises and observed improvements in health outcomes. This finding strengthens the rationale for personalized exercise programs and highlights the importance of adherence and regularity in achieving optimal outcomes.

Hypotheses that were derived from this retrospective analysis for a randomized controlled trial: (1) participants in the intervention group following an 8-week AI-composed exercise program will experience a statistically significant reduction in pain intensity and a significant improvement in well-being compared with the control group (alternative hypothesis) and (2) the frequency of engagement in the prescribed exercises will be positively correlated with decreased pain intensity and increased well-being among participants with spinal pain (correlation hypothesis).

Acknowledgments

The authors would like to thank Philipp Schlüter for his assistance. We would also like to thank medicalmotion GmbH for providing the patient examples.

Author Ines Cordon Morillas is not affiliated with any institution but is an independent scholar/freelancer who mainly supported data analysis for this study.

No funding, grants, or other support has been received.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: AG, FA, KL
Formal analysis: AG, FA, ICM
Investigation: AG, FA, KL
Methodology: AG, CZ, KL
Project administration: KL, AG
Writing – original draft: AG, FA
Writing – review & editing: AG, FA, CZ, RPM, ICM, KL

Conflicts of Interest

AG and FA are employees of medicalmotion GmbH. They report personal fees and nonfinancial support provided by medicalmotion GmbH. The remaining authors declare no competing interests.

Multimedia Appendix 1
STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement.
[PDF File, 123 KB - formative_v9i1e57826_app1.pdf]

Multimedia Appendix 2
Exercise examples.
[PDF File, 68 KB - formative_v9i1e57826_app2.pdf]

Multimedia Appendix 3
Exercise-induced correlation of pain intensity and well-being.
[PDF File, 111 KB - formative_v9i1e57826_app3.pdf]

References

XSL•FO
RenderX

1. Musculoskeletal health. World Health Organization. 2022 Jul 14. URL: https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions [accessed 2023-03-24]
2. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392:1789-1858. [doi: 10.1016/S0140-6736(18)32279-7]
3. Wu A, March L, Zheng X, et al. Global low back pain prevalence and years lived with disability from 1990 to 2017: estimates from the Global Burden of Disease Study 2017. Ann Transl Med 2020 Mar;8(6):299. [doi: 10.21037/atm.2020.02.175] [Medline: 32355743]
4. Knezevic NN, Candido KD, Vlaeyen JWS, Van Zundert J, Cohen SP. Low back pain. Lancet 2021 Jul 3;398(10294):78-92. [doi: 10.1016/S0140-6736(21)00733-9] [Medline: 34115979]
5. Hayden JA, van Tulder MW, Malmivaara A, Koes BW. Exercise therapy for treatment of non-specific low back pain. Cochrane Database Syst Rev 2005 Jul 20;2005(3):CD000335. [doi: 10.1002/14651858.CD000335.pub2] [Medline: 16034851]
6. Geneen LJ, Moore RA, Clarke C, Martin D, Colvin LA, Smith BH. Physical activity and exercise for chronic pain in adults: an overview of Cochrane Reviews. Cochrane Database Syst Rev 2017 Jan 14;1(1):CD011279. [doi: 10.1002/14651858.CD011279.pub2] [Medline: 28087891]
7. Steffens D, Maher CG, Pereira LSM, et al. Prevention of low back pain: a systematic review and meta-analysis. JAMA Intern Med 2016 Feb;176(2):199-208. [doi: 10.1001/jamainternmed.2015.7431] [Medline: 26752509]
8. Oliveira CB, Maher CG, Pinto RZ, et al. Clinical practice guidelines for the management of non-specific low back pain in primary care: an updated overview. Eur Spine J 2018 Nov;27(11):2791-2803. [doi: 10.1007/s00586-018-5673-2] [Medline: 29971708]
9. Jordan JL, Holden MA, Mason EE, Foster NE. Interventions to improve adherence to exercise for chronic musculoskeletal pain in adults. Cochrane Database Syst Rev 2010 Jan 20;2010(1):CD005956. [doi: 10.1002/14651858.CD005956.pub2] [Medline: 20091582]
10. George SZ, Fritz JM, Silfies SP, et al. Interventions for the management of acute and chronic low back pain: revision 2021. J Orthop Sports Phys Ther 2021 Nov;51(11):CPG1-CPG60. [doi: 10.2519/jospt.2021.0304] [Medline: 34719942]
11. Hewitt S, Sephton R, Yeowell G. The effectiveness of digital health interventions in the management of musculoskeletal conditions: systematic literature review. J Med Internet Res 2020 Jun 5;22(6):e15617. [doi: 10.2196/15617] [Medline: 32501277]
12. Hou J, Yang R, Yang Y, et al. The effectiveness and safety of utilizing mobile phone-based programs for rehabilitation after lumbar spinal surgery: multicenter, prospective randomized controlled trial. JMIR Mhealth Uhealth 2019 Feb 20;7(2):e10201. [doi: 10.2196/10201] [Medline: 30785406]
13. Lewkowicz D, Slosarek T, Wernicke S, Winne A, Wohlbrandt AM, Bottinger E. Digital therapeutic care and decision support interventions for people with low back pain: systematic review. JMIR Rehabil Assist Technol 2021 Nov 19;8(4):e26612. [doi: 10.2196/26612] [Medline: 34807837]
14. Machado GC, Pinheiro MB, Lee H, et al. Smartphone apps for the self-management of low back pain: a systematic review. Best Pract Res Clin Rheumatol 2016 Dec;30(6):1098-1109. [doi: 10.1016/j.berh.2017.04.002] [Medline: 29103552]
15. Priebe JA, Haas KK, Moreno Sanchez LF, et al. Digital treatment of back pain versus standard of care: the cluster-randomized controlled trial, Rise-uP. J Pain Res 2020;13:1823-1838. [doi: 10.2147/JPR.S260761] [Medline: 32765057]
16. Bailey JF, Agarwal V, Zheng P, et al. Digital care for chronic musculoskeletal pain: 10,000 participant longitudinal cohort study. J Med Internet Res 2020 May 11;22(5):e18250. [doi: 10.2196/18250] [Medline: 32208358]
17. Lara-Palomo IC, Antequera-Soler E, Matarán-Peñarrocha GA, et al. Comparison of the effectiveness of an e-health program versus a home rehabilitation program in patients with chronic low back pain: a double blind randomized controlled trial. Digit Health 2022;8:20552076221074482. [doi: 10.1177/20552076221074482] [Medline: 35111332]
18. Özden F, Sarı Z, Karaman Ö, Aydoğmuş H. The effect of video exercise-based telerehabilitation on clinical outcomes, expectation, satisfaction, and motivation in patients with chronic low back pain. Ir J Med Sci 2022 Jun;191(3):1229-1239. [doi: 10.1007/s11845-021-02727-8] [Medline: 34357527]
19. Alowais SA, Alghamdi SS, Alsuhebany N, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ 2023 Sep 22;23(1):689. [doi: 10.1186/s12909-023-04698-z] [Medline: 37740191]
20. Rasmussen CDN, Sandal LF, Holtermann A, Stochkendahl MJ, Mork PJ, Søgaard K. Effect of a smartphone self-management digital support system for low-back pain (selfBACK) among workers with high physical work demands: secondary analysis of a randomized controlled trial. Scand J Work Environ Health 2024 Dec 1;50(8):613-621. [doi: 10.5271/sjweh.4186] [Medline: 39264347]
21. Marcuzzi A, Klevanger NE, Aasdahl L, et al. An artificial intelligence-based app for self-management of low back and neck pain in specialist care: process evaluation from a randomized clinical trial. JMIR Hum Factors 2024 Jul 9;11:e55716. [doi: 10.2196/55716] [Medline: 38980710]
22. Hartmann R, Avermann F, Zalpour C, Griefahn A. Impact of an AI app-based exercise program for people with low back pain compared to standard care: a longitudinal cohort-study. Health Sci Rep 2023 Jan;6(1):e1060. [doi: 10.1002/hsr2.1060] [Medline: 36660258]

XSL•FO
RenderX

23. de Campos TF, Maher CG, Fuller JT, Steffens D, Attwell S, Hancock MJ. Prevention strategies to reduce future impact of low back pain: a systematic review and meta-analysis. Br J Sports Med 2021 May;55(9):468-476. [doi: 10.1136/bjsports-2019-101436] [Medline: 32646887]
24. Huang R, Ning J, Chuter VH, et al. Exercise alone and exercise combined with education both prevent episodes of low back pain and related absenteeism: systematic review and network meta-analysis of randomised controlled trials (RCTs) aimed at preventing back pain. Br J Sports Med 2020 Jul;54(13):766-770. [doi: 10.1136/bjsports-2018-100035] [Medline: 31672696]
25. da Silva T, Mills K, Brown BT, et al. Recurrence of low back pain is common: a prospective inception cohort study. J Physiother 2019 Jul;65(3):159-165. [doi: 10.1016/j.jphys.2019.04.010] [Medline: 31208917]
26. Pocovi NC, Lin CWC, French SD, et al. Effectiveness and cost-effectiveness of an individualised, progressive walking and education intervention for the prevention of low back pain recurrence in Australia (WalkBack): a randomised controlled trial. Lancet 2024 Jul 13;404(10448):134-144. [doi: 10.1016/S0140-6736(24)00755-4] [Medline: 38908392]
27. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008 Apr;61(4):344-349. [doi: 10.1016/j.jclinepi.2007.11.008] [Medline: 18313558]
28. AI-based exercise recommendations for people with back pain: a retrospective cohort analysis. OSF Registries. URL: https://osf.io/kjhef [accessed 2025-02-03]
29. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. Eur J Pain 2004 Aug;8(4):283-291. [doi: 10.1016/j.ejpain.2003.09.004] [Medline: 15207508]
30. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Routledge; 2013.
31. Nair AS, Diwan S. Pain scores and statistical analysis—the conundrum. Ain-Shams J Anesthesiol 2020 Dec;12(1). [doi: 10.1186/s42077-020-00085-8]
32. Quentin C, Bagheri R, Ugbolue UC, et al. Effect of home exercise training in patients with nonspecific low-back pain: a systematic review and meta-analysis. Int J Environ Res Public Health 2021 Aug 10;18(16):8430. [doi: 10.3390/ijerph18168430] [Medline: 34444189]
33. Weise H, Zenner B, Schmiedchen B, et al. The effect of an app-based home exercise program on self-reported pain intensity in unspecific and degenerative back pain: pragmatic open-label randomized controlled trial. J Med Internet Res 2022 Oct 28;24(10):e41899. [doi: 10.2196/41899] [Medline: 36215327]
34. Cimarras-Otal C, Marcen-Cinca N, Rabal-Pelay J, et al. Adapted exercises versus general exercise recommendations on chronic low back pain in industrial workers: a randomized control pilot study. Work 2020;67(3):733-740. [doi: 10.3233/WOR-203322] [Medline: 33164978]
35. Yildizeli Topcu S. Relations among pain, pain beliefs, and psychological well-being in patients with chronic pain. Pain Manag Nurs 2018 Dec;19(6):637-644. [doi: 10.1016/j.pmn.2018.07.007] [Medline: 30181033]
36. Marshall PWM, Schabrun S, Knox MF. Physical activity and the mediating effect of fear, depression, anxiety, and catastrophizing on pain related disability in people with chronic low back pain. PLoS ONE 2017;12(7):e0180788. [doi: 10.1371/journal.pone.0180788] [Medline: 28686644]
37. Marshall PW, Morrison NMV, Gibbs M, Schabrun SM. The effect of exercise engagement on low back disability at 12-months is mediated by pain and catastrophizing in a community sample of people with chronic low back pain. Behav Res Ther 2022 Dec;159:104205. [doi: 10.1016/j.brat.2022.104205] [Medline: 36215852]
38. Dishman RK, Sallis JF, Orenstein DR. The determinants of physical activity and exercise. Public Health Rep 1985;100(2):158-171. [Medline: 3920714]
39. Meyer OL, Castro-Schilo L, Aguilar-Gaxiola S. Determinants of mental health and self-rated health: a model of socioeconomic status, neighborhood safety, and physical activity. Am J Public Health 2014 Sep;104(9):1734-1741. [doi: 10.2105/AJPH.2014.302003] [Medline: 25033151]
40. Labinsky H, Gupta L, Raimondo MG, Schett G, Knitza J. Real-world usage of digital health applications (DiGA) in rheumatology: results from a German patient survey. Rheumatol Int 2023 Apr;43(4):713-719. [doi: 10.1007/s00296-022-05261-7] [Medline: 36543961]
41. Tian Y, Zhou K, Pelleg D. What and how long: prediction of mobile app engagement. ACM Trans Inf Syst 2022 Jan 31;40(1):1-38. [doi: 10.1145/3464301]

Abbreviations

AI: artificial intelligence
LBP: low back pain
MSD: musculoskeletal disorder
NRS: numeric rating scale
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

XSL•FO
RenderX

Edited by A Mavragani; submitted 28.02.24; peer-reviewed by JB Lechauve, Z Rahemi; revised version received 11.12.24; accepted 12.12.24; published 18.02.25.

Please cite as:
Griefahn A, Avermann F, Zalpour C, Marshall RP, Cordon Morillas I, Luedtke K
Exploring the Effect of an 8-Week AI-Composed Exercise Program on Pain Intensity and Well-Being in Patients With Spinal Pain: Retrospective Cohort Analysis
JMIR Form Res 2025;9:e57826
URL: https://formative.jmir.org/2025/1/e57826
doi:10.2196/57826

© Annika Griefahn, Florian Avermann, Christoff Zalpour, Robert Percy Marshall, Inés Cordon Morillas, Kerstin Luedtke. Originally published in JMIR Formative Research (https://formative.jmir.org), 18.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Involving End Users in the Development and Usability Testing of a Smartphone App Designed for Individuals With Prediabetes: Mixed-Methods Focus Group Study

Natalie Grieve[1], BA, MSc; Kyra Braaten[1], BSc, MSc; Megan MacPherson[1], BA, MOT, PhD; Sam Liu[2], BPHE, MSc, PhD; Mary E Jung[1], HBKin, MSc, PhD

[1]Faculty of Health and Social Development, University of British Columbia, 1238 Discovery Way, Kelowna, BC, Canada

[2]Exercise Science, Physical and Health Education, University of Victoria, Victoria, BC, Canada

Corresponding Author:
Mary E Jung, HBKin, MSc, PhD
Faculty of Health and Social Development, University of British Columbia, 1238 Discovery Way, Kelowna, BC, Canada

Abstract

Background: Technology is more likely to be used when it is designed to meet the needs of end users. To supplement the Small Steps for Big Changes diabetes prevention program, a smartphone app was developed in partnership with past Small Steps for Big Changes clientele. Usability testing is critical for the ongoing use and adoption of mobile health apps by providing insight on where appropriate adjustments and improvements need to be made to ensure user satisfaction.

Objective: A focus group with 7 participants was conducted to examine the app's usability and collect feedback for future iterations.

Methods: Past Small Steps for Big Changes clientele participated in a cognitive walkthrough of 8 novel tasks and completed the System Usability Scale survey. Participants were then given the option to use the app for 3 weeks before completing the User-Mobile Application Rating Scale.

Results: Analysis of the cognitive walkthrough identified 26 usability problems; each was coded using a heuristic evaluation to describe usability errors. The most frequently coded errors included inappropriate progress feedback, information appearing in an illogical order, counterintuitive design, and issues with app aesthetics. A mean summary score of 66.8% (SD 18.91) was reported for the System Usability Scale, representing a marginal acceptability score and indicating that design issues needed to be resolved. A User-Mobile Application Rating Scale mean score of 3.59 (SD 0.33) was reported, implying an average acceptability rating.

Conclusions: These findings identified necessary improvements in the app, ranging from minor aesthetic problems to major functionality problems. Involving end users allows the app to be tailored to the client's preferences and increases the likelihood of usage. This app aligns with Small Steps for Big Changes' program components and behavior change techniques that can improve health outcomes for future clients and allow them to self-monitor their exercise, diet, and goals.

(JMIR Form Res 2025;9:e59386) doi:10.2196/59386

KEYWORDS

usability evaluation; mHealth; usability testing; app; end-user; focus group; participant; survey; diabetes; user-centered; cognitive walkthrough; cognitive walkthroughs; questionnaire; mobile phone; digital health; prediabetes

Introduction

Background

Involving end users in mobile health (mHealth) technology development increases the value and practical use of the product [1,2]. Recognizing a client's expertise through collaboration allows end users to effectively express their needs and expectations during the development phases of mHealth apps [3]. mHealth technologies have been used to increase accessibility to health care resources [4] and increase patient autonomy [5]. The uptake of mHealth technology has been shown to improve chronic disease management including improvements in weight management and hypertension, and decreased hospitalizations [6]. Within diabetes research, mHealth tools have assisted in self-management efforts, improved glycemic control [6-8], and are acceptable by users [9]. mHealth apps allow users to play an active role in their health, subsequently improving health outcomes. To capitalize on these advantages, an mHealth app was developed to supplement a diabetes prevention program run in British Columbia, Canada.

XSL•FO
RenderX

Small Steps for Big Changes is a community-based diabetes prevention program designed for individuals with prediabetes [10]. Small Steps for Big Changes is delivered in 6 sessions over 4 weeks using a client-centered counseling style known as motivational interviewing with follow-ups over 12 months. Clients work one-on-one with their coach to cover several topics including goal setting, healthy food choices, and exercise adherence. During the program's first session, clients are introduced to self-monitoring techniques, including tracking diet and exercise. Various forms of tracking are discussed, including commercial mHealth apps, journals, paper tracking, and exercise logs. While the most prominent method of tracking used in the program is commercially available mHealth apps, several challenges have been reported by Small Steps for Big Changes clients and coaches with such tools [11]. This program-specific mHealth app was developed to supplement the Small Steps for Big Changes program and to fill the gap in mHealth apps previously identified by Small Steps for Big Changes participants. Involving end users during development supported initiatives to ensure the app was relevant and useful. This mHealth app was developed to support the Small Steps for Big Changes program by integrating the frequently used behavior change techniques throughout the 6 sessions, emphasizing goals, planning, feedback, and monitoring [12].

Development of the Small Steps for Big Changes mHealth app was guided by the FASTER (Framework for Accelerated and Systematic Technology-Based Intervention Development and Evaluation Research), a 3-phase framework that recognizes the complexities of conducting research on mHealth technologies in a timely fashion while also consistently engaging end users throughout [13]. Phase 1 included a needs assessment, literature review, and preliminary evaluation; the feedback and results from phase 1 helped inform the app prototype used in this usability study [14]. This study focused on the second phase of FASTER, examining the progressive usability of the app [13]. In this phase, small-scale usability testing helped uncover usability errors, levels of acceptability, and improvements needed before large-scale implementation (phase 3 of FASTER). Usability testing is critical for sustained use and uptake of mHealth apps by identifying where changes can be made [13]. An app that is difficult to navigate with an interface that is nonintuitive limits usage. mHealth technologies provide opportunities to self-monitor and manage health behaviors for individuals with chronic disease [15]. Successful self-monitoring of behaviors involves consistency; if carried out on an app, this needs an interface that is easy to use and will encourage the likelihood that goals will be recorded and revisited. Understanding where usability errors exist will inform appropriate changes needed to improve app features and increase user satisfaction [13].

Objective

The aim of this study was the answer the following questions: where do existing usability errors exist in the Small Steps for Big Changes mHealth app and is the app deemed usable by past Small Steps for Big Changes participants? To answer this question, this study used a mixed-methods approach involving a cognitive walkthrough (CW) focus group, take-home testing period, and self-reported usability surveys to engage end users in the examination of usability of a Small Steps for Big Changes mHealth app.

Methods

Small Steps for Big Changes App Development

Pathverse, the app-building company used for this project that specializes in no-code app building for research and learning [16], worked with our research team to create a platform specific to the needs of the Small Steps for Big Changes program. Several meetings took place with Pathverse and the research team to organize content, create select features, and tailor the content to the Small Steps for Big Changes program. This app included basic self-monitoring features, including tracking for steps, minutes of daily exercise, weight, waist circumference, and hemoglobin A_{1c} (HbA_{1c}). Additional capabilities included logging exercise sessions, recording goals, diary entry opportunities, and access to additional frequently asked questions and resources including nutritional cooking videos and program content. Additional details on layout and design can be seen in the attached screenshots available in Multimedia Appendix 1.

Participants

A total of 7 participants were recruited to participate in this study. A sample size between 6 and 10 participants was determined a priori based on past literature exploring user-centered CW methods in mHealth chronic disease self-management [17,18]. Secondary measures used to collect additional usability data included the System Usability Scale (SUS) [19,20] and User-Mobile Application Rating Scale (uMARS) [21]. Past Small Steps for Big Changes clients who had completed the 4-week training phase of the program were invited to take part in this study. Eligible participants included past Small Steps for Big Changes clients who were adults aged 18 years or older, able to read and write in English, assessed for prediabetes by one of the following means: (1) physician-diagnosed prediabetes, (2) HbA_{1c} values between 5.7% and 6.4%, and (3) an American Diabetes Association risk questionnaire score indicating increased risk of diabetes (>5). Participants also needed to have access to a mobile device with internet access (such as a tablet or smartphone). In an attempt to represent the current demographic of the Small Steps for Big Changes program (75% of individuals identified as female and 25% as male), participants were purposively sampled. Participants were asked to bring an internet-enabled mobile device during their laboratory visit with capabilities to connect and access the device over the 3-week take-home period.

Ethical Considerations

Informed consent was obtained from all individual participants and all procedures performed were in accordance with the ethical standards of the University of British Columbia Behavioural Research Ethics Board (H22-01399).

CW: Phase 1

Participants were invited to engage in a CW focus group at the University of British Columbia Okanagan Campus. CWs involve a task-based method to explore the usability of digital health

XSL•FO
RenderX

technologies and when used in conjunction with questionnaires are part of a multistep process to evaluate usability [17,18,22]. As participants arrived, research assistants helped to provide assistance to connect to the internet, download the app, and create a user profile. A 10-minute demonstration of the Small Steps for Big Changes app on a large projector screen was presented before the CW began to account for the app's novelty. During this demonstration a researcher lead a familiarization walkthrough of each of the function landing pages (home page, tracking page, and resources page) as well as some of the button functions (go-back, save, and delete). Participants followed along on their devices and were given time to play with navigating the app before starting the novel tasks. After this period, each participant was provided with a task booklet outlining 8 novel tasks to be completed during the CW focus group. Each task had 3 corresponding questions used to identify usability problems about the task and assist with the facilitation of group discussion. The workbook was designed for additional observations to be recorded during each task. During the CW, participants were asked to rotate through leading the completion of a task using their instincts by informing the facilitator how to perform the task on a large screen for the group to follow. After each task, all participants were asked to record their observations and share their thoughts about any usability problems they encountered. Each participant lead at least 1 task during the CW. Once all tasks were completed in the CW, participants were granted access to the app for an additional 3 weeks to use it as they pleased, simulating a real-world setting. During this time, participants were encouraged to make note of any additional usability errors, feedback for future modifications, and general reactions or evaluations. This information was provided along with compensation following completion of this study.

Mobile Device Proficiency Questionnaire and Demographics Survey

The Mobile Device Proficiency Questionnaire (MDPQ-16) [23] was used to assess participants' mobile device proficiency, and demographics were collected for descriptive reporting. Participants self-reported their age, sex, gender, whether they were born in Canada, whether they identify as Indigenous or a visible minority, their highest level of education completed, occupational status, household income, and marital status. During the initial laboratory visit, participants were asked to complete a demographics survey and the MDPQ-16 on iPads (Apple Inc) before beginning the CW. The MDPQ-16 includes 16 questions organized into 8 subscales such as mobile device basics, communication, internet, troubleshooting, and are answered on a 5-point scale (1=never tried and 5=tried very easily). Each subscale score is averaged and then summed, with a maximum score of 40 (a higher score indicates greater proficiency [23]).

About SUS

After completing the CW tasks, participants were asked to complete the SUS [19]. The SUS was used to assess system usability. This brief 10-question survey asks general usability questions about ease of use and complexity on a 5 point-scale (1=strongly disagree and 5=strongly agree). The SUS is a widely used scale, a recent meta-analysis concluded that the SUS was a suitable measure for evaluating the usability of digital health apps [20]. Final SUS scores range from 0 - 100, with higher scores indicating better usability [19] with a benchmark acceptability score of 68 [20]. This survey was delivered immediately after the CW to each participant on an iPad.

User Version of the Mobile App Rating Scale: Phase 2

After completing an optional take home period, the uMARS [21] was used to assess the quality and usability of the mHealth app. The uMARS items are answered on a 5-point scale, representing the following 1-inadequate, 2-poor, 3-acceptable, 4-good, and 5-excellent [21]. The uMARS is a reliable tool to measure app quality and has been used to assess mHealth apps including self-management behaviors associated with type 2 diabetes [24]. This survey offered additional detail and depth compared to the SUS and was delivered to the participants after an optional 3-week at-home phase. The uMARS consists of 4 subscales focused on engagement, functionality, aesthetics and information. A mean score for each subscale is calculated, and a total app quality score is calculated by averaging the subscale means. The survey also includes additional questions about app subjective quality, perceived impact, and an option to leave further written comments. Participants were given information about the take-home phase of this study and were asked to use the app freely for 3 weeks. After this take-home period, participants were sent a survey link to the uMARS.

Analysis

All survey data from the demographics, MDPQ-16, SUS, and uMARS were descriptively analyzed using SPSS (version 27; IBM Corp) to examine mean scores, SD, and range. The CW was analyzed through an audio recording and transcript. The usability problems identified in the CW task workbook and transcript were coded using Nielsen's heuristic evaluation [25,26]. The 10 usability heuristics outline general principles for user interface design; if an identified usability problem violates any of the 10 heuristics, it is coded accordingly. Furthermore, 2 coders completed the qualitative analysis and assigned severity scores to each of the usability errors identified.

Results

Demographics and MDPQ-16

A total of 5 of the 7 participants were aged older than 70 years, and 2 were between 55 and 65 years of age. No individuals identified as Indigenous or a visible minority. Further, 2 participants identified as having a disability. There was variance in education level and family income; 5 participants were retired. Table 1 reports on more detailed demographic results. MDPQ-16 scores were reported to assess levels of mobile device proficiency. The mean (SD) MDPQ-16 total score was 36.36 and SD of 3.42, individual MDPQ-16 scores are reported in Multimedia Appendix 1. These MDPQ-16 scores represented high levels of mobile device proficiency for our participants compared to those identified in previous literature [23].

XSL•FO
RenderX

Table . Demographic characteristics of participants.

Sample characteristic	Participants who selected this response
Age (years), mean (SD)	70 (7.47)
Sex, n (%)	
Female	4 (57)
Male	3 (43)
Gender, n (%)	
Woman	4 (57)
Man	3 (43)
Born in Canada, n (%)	
Yes	4 (57)
No	3 (43)
No	7 (100)
No	7 (100)
Identifies as a person with a disability, n (%)	
Yes	2 (29)
No	4 (57)
Prefer not to answer	1 (14)
Level of education, n (%)	
High school	1 (14)
University certificate or diploma below the bachelor level	3 (43)
Postgraduate degree	2 (29)
College, CEGEP[a], or other nonuniversity certificate or diploma	1 (14)
Working status, n (%)	
Working full time	2 (29)
Retired	5 (71)
Annual income level (US $)	
$41,688 to $62,531	1 (14)
$62,532 to $82,681	1 (14)
$104,220 or more	3 (43)
Prefer not to answer	2 (29)
Marital status, n (%)	
Married	6 (86)
Widowed	1 (14)

[a]CEGEP, Collège d'enseignement general et professionnel, which means general and professional teaching college

About CW

The CW lasted 60 minutes in duration. Participants were asked to complete 8 novel tasks while engaging with the Small Steps for Big Changes app. A total of 26 usability problems were identified and assigned severity scores in the recording transcript and client workbooks. The highest number of usability problems identified for a specific task was 7 for task 1, and the lowest was 0 for task 8. All severity scores and usability problems identified for each task and corresponding Nielsen heuristics are reported in Table 2.

XSL•FO
RenderX

Table . Usability problems identified in associated workbook task completion.

Task	Number of usability problems identified	Mean severity score (SD)	Taps expected to complete task (n)	Actual taps to complete task (n)	Nielsen heuristics identified (*f*)
1-Record a weight value	7	2.1 (0.69)	5	8	1 (3), 2 (4), 3 (1), 4 (2), 6 (4), 8 (2)
2-Navigate to the FAQ[a] page	2	0.4 (0.53)	4	4	3 (1), 6 (1), 8 (1)
3-Log an exercise	3	1 (1.15)	17	18	1 (1), 2 (1), 4 (1), 6 (2), 8 (2)
4-Create a diary entry	1	0.5 (0.58)	7	7	10 (1)
5-Record a goal	5	1.9 (1.38)	12	15	1 (1), 2 (2), 3 (1), 4 (1), 5 (1), 6 (2), 9 (1)
6-Navigate to a module	4	1.6 (0.53)	6	9	2 (3), 3 (1), 8 (1), 10 (2)
7-Update the hemoglobin A_{1c} value within the tracking page	4	0 (0)	10	10	2 (1), 4 (2), 6 (2), 8 (1)
8-Navigate to a video resource	0	0 (0)	5	5	0

[a]FAQ: frequently asked questions.

The usability problems identified in the CW included issues with icon and text sizes, scroll sensitivity, system preferences, and logical location of features and information. Common system preference issues were challenges saving and editing goals and the exclusion of a 24-hour clock. Participants also had difficulty scrolling and reported favorability for a text box for selecting values. Usability problems were analyzed by coding each into 10 categories. The 26 usability problems highlighted errors associated with information appearing in an illogical order and lack of visibility of information. For example, the usability error named “tracker alignment” indicates that the weight tracking feature’s dates did not match what was seen on the graph displayed in the participant’s progress summary. This was coded as 1 (incorrect feedback provided by the system) and 2 (information appearing in an illogical order). All usability problems were coded accordingly; some were assigned only 1 Nielsen heuristic, while others were assigned multiple.

NG and KB double-coded the 26 identified usability errors according to heuristic evaluation. Nielsen heuristics of 2 (match between system and the real world) and 6 (recognition rather than recall) were the most frequently coded (11 times each), while heuristics of 5 (error prevention), 7 (flexibility and efficiency of use), and 9 (help users recognize, diagnose, and recover from errors) were rarely coded. Table 3 outlines each heuristic score, their coded frequency, and the corresponding percentage of the 49 total codes used.

Table . Nielsen heuristic frequency for usability problem codes from a total of 26 usability problems and 49 total codes.

Nielsen heuristic	Frequency (%[a])
Visibility of system status	5 (10)
Match between system and the real world	11 (22)
User control and freedom	4 (8)
Consistency and standards	6 (12)
Error prevention	1 (2)
Recognition rather than recall	11 (22)
Flexibility and efficiency of use	0 (0)
Aesthetic and minimalist design	7 (14)
Help users recognize, diagnose, and recover from errors	1 (2)
Help and documentation	3 (6)

[a]%: percentage of the total frequency of codes.

XSL•FO
RenderX

About SUS

The mean SUS score was 66.8% for the sample, with an SD of 18.91 and a large range from 50 to 97.5. Further, 4 participants rated the usability moderately low, between 50% and 60%, while the remaining 3 rated the usability highly, resulting in scores above 85%.

About uMARS

In total, 5 of the 7 clients who participated in the CW completed the full uMARS. The app quality total score was calculated to be a mean score of 3.59/5 with an SD of 0.33, with the highest-rated subcategory being aesthetics at a mean score of 3.8 (SD 0.51) and the lowest being engagement at 3.12 (SD 0.33). The highest perceived impact scores were behavior change and awareness, both receiving a mean score of 3.33 (SD 1.03), with knowledge being the lowest at a mean of 2.8 (SD 0.84). Additional feedback for improvements included adding analytics for the progress made on the app, adding synchronizing capabilities with smartwatches to avoid manually recording physical activity and including tracking standards for specific age groups and sex. A preference to the Fitbit app was also reported by 2 participants during discussion of additional feedback after the CW. Positive comments resulting from completion of the uMARs included comments about the potential for the app to be a very useful tool, the value of recording anthropometrics such as weight and HbA_{1c}, the emphasis on diabetes, and the use of the graphs as a visual representation of the client's progress. Survey scores for usability data are reported in Table 4, and additional results from the usability surveys are included in Multimedia Appendix 1.

Table . Usability scores for quantitative surveys (N=18).

Survey	Mean (SD)	Range
MDPQ-16[a]	36.36/40 (3.42)	31 - 40
SUS[b]	66.8/100 (18.91)	50 - 97.5
uMARS[c]	3.59/5 (0.33)	3.03 - 3.83

[a]MDPQ-16: Mobile Device Proficiency Questionnaire.

[b]SUS: System Usability Scale.

[c]uMARS: User-Mobile Application Rating Scale.

Follow-Up With Pathverse

Once the research team noted the usability problems and feedback, 2 follow-up meetings were conducted with Pathverse to address the usability concerns and make appropriate improvements. Updates made by the Pathverse team included correcting minor bugs with diary notification scheduling, fixing the scaling of images and icons, and bringing consistency across devices (Samsung vs Apple). Additions made to the app included 24-hour time when inputting goals and tracking, opportunity to input manual steps, and updates for some of the language and icons used (the menu symbol, save button, and trash icon).

Discussion

Findings

The 26 usability problems identified in the CW identified relevant usability errors seen during task-based discussion within the group of participants. Further, 9 of the 10 Nielsen heuristics were coded, indicating usability errors across multiple domains of the app; however, most usability problems were addressed for preference changes and minor updates in the most recent iteration of the app. Other suggestions, including a hover function to show a description of functions within the app, and the ability to link the app with a Fitbit have not yet been addressed due to cost, time, and practicality implications.

The usability score of 66.8 for the SUS represented a marginal usability score [19,20]. The variance could be due to a range in expectations of the app and its functions; a few users expressed they had been using different health monitoring apps, which they were quite satisfied with, which may have influenced their perceptions of the Small Steps for Big Changes app. After 3 weeks of optional use, a uMARS mean score of 3.59 was reported, describing an acceptable score [22]. MDPQ-16 scores were relatively high in our sample; we can have confidence that the feedback provided was independent of a lack of skills or ability to use mobile apps. Within our sample, moderate levels of usability may have been due to a low level of willingness to track in general. We did not purposively sample those who favored smartphone tracking over other methods such as tracking in a journal or workbook,; and some users may have had little interest in tracking at all.

The literature has identified perceived usability as an existing barrier to mHealth usage [27]; this aligns with the moderate levels of usability reported in this study. However, high levels of usability have been reported by select mHealth technologies designed for individuals with type 2 diabetes [9,28], which counter the moderate levels of usability seen in this study. Although the levels of usability have demonstrated mixed results, when comparing user feedback, various themes did align with our findings. Users agreed on noting similar improvements, including adding more knowledge and information available on the mHealth apps [28], suggesting that mHealth app content may be oversimplified. Interestingly, this study reported high levels of mobile device proficiency from adults with a mean age of 70 years, which may contradict previous research noting older age as a barrier to mHealth usage [27].

This study identified usability problems and necessary improvements for the Small Steps for Big Changes program

XSL•FO
RenderX

app. With only moderate levels of usability reported, these results emphasize the need for user feedback during the development process. This highlights the importance of the end user's involvement in app development and prototype stages. Several iterations of mHealth technology should be tested before large-scale implementation to ensure the end product is relevant and highly useful to users. This study demonstrates that early usability evaluations can help identify and address issues at the beginning of the development cycle, saving time and resources in the long run. The iterative nature of this study's approach, where usability problems were identified and subsequently addressed, underscores the importance of continuous improvement. Following the steps of the FASTER framework helped ensure that necessary steps were not overlooked; during this process, essential findings for decision-making can save on cost and the long-term success of an mHealth app [13]. This study highlights the importance of tailoring mobile apps, especially health-related ones, to meet the specific needs and preferences of older adults. Developers should prioritize user-centered design principles, considering factors such as font size, icon clarity, and ease of navigation to enhance usability for this demographic.

Past research has highlighted the importance of integrating multiple methods to assess usability in digital health evaluation [22],; this study incorporated various modes of feedback through a CW, real-world use for 3 weeks, validated questionnaires, and qualitative feedback. Diverse feedback sources provide a more comprehensive understanding of usability and user experiences. Each feedback method uncovers different aspects of usability. Combining various feedback sources allows for a holistic evaluation of the mobile app. This approach ensures that usability problems are not only identified but also validated across different dimensions. It helps prioritize issues that have the most significant impact on users. The real-world use over 3 weeks provides insights into how usability evolves over time. It helps identify whether initial issues persist or if users adapt to the app. The combined feedback informs tailored iterations. Specific issues can be addressed when identified through different sources in a targeted manner. For example, usability problems identified in CW may lead to design changes, while questionnaire data can guide overall satisfaction improvements.

Limitations

Limited diversity within our sample was seen, specifically for those identifying as Indigenous or a visible minority. While this is representative of the current demographics served by Small Steps for Big Changes, as Small Steps for Big Changes expands to more communities, the potential to target more diverse opinions should be highlighted, especially for those who may benefit the most from increased accessibility (eg, individuals living in rural communities). This study did not explore long-term user compliance and security issues; further research and considerations should be explored to identify solutions around these limiting factors [29]. Considerations including (1) a small sample size when reporting uMARS results and (2) an unrealistic simulation of app implementation (providing assistance during installation and a familiarization of the app) should be taken into account when interpreting the results.

Future Directions for Research

Future research may look toward phase 3 of FASTER through large-scale implementation and evaluation of the app as a supplement to the Small Steps for Big Changes diabetes prevention program to examine the app's effectiveness in improving health behaviors and clinical outcomes for clients. mHealth prompts such as texts and push notifications have been well-received by users in diabetes prevention [30]. Looking toward these capabilities within the Small Steps for Big Changes app is an additional function that may benefit client health outcomes by allowing them to self-monitor their exercise, diet, and goals. Future iterations of the app will call for additional cycles of usability testing and evaluation to ensure a satisfactory user experience is reached. Usability is an important aspect of user experience; however, additional factors including incentivizing, convenience, and practicality of mHealth apps may also influence decisions to integrate their use in practice. For example, the results of this study report that the lowest category in the uMARS scale was engagement. In the uMARS, user engagement refers to whether the app is entertaining, interesting, customizable and interactive; if users do not find the app engaging they may be less incentivized to use it.

Conclusions

Diabetes self-management efforts can be improved using mHealth technologies; it can help establish meaningful routines, allow clients to better understand their health conditions and status, and provide additional outlets for interactions with health care professionals and resources [31]. Providing tools to make behavior change efforts easier for clients will lead to better outcomes; however, these tools need to be well-accepted by users to increase the likelihood of usage. This study successfully assessed the usability of a self-monitoring app for individuals with prediabetes by identifying 26 usability errors and recognizing where improvements needed to be made. These activities are necessary for generating research evidence for technology-based interventions [13] to ensure usage is not a limiting factor. Moderate usability levels were reported by users, prompting a new iteration of the app, emphasizing the importance of user involvement. Involving end users in this process allows the mHealth app usage experts to make decisions during development. Overlooking usability testing during applicaton development can result in premature large-scale implementation before critical changes and errors can be addressed. Embracing multiple feedback modes reflects a user-centric approach to design and evaluation. It demonstrates a commitment to understanding the user experience from various angles, which is essential for creating user-friendly apps.

Acknowledgments

We respectfully acknowledge that the work and research conducted for this study was on the traditional, ancestral, and unceded territory of the Syilx Nation.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional usability data and screenshots of app platform.
[DOCX File, 2565 KB - formative_v9i1e59386_app1.docx]

References

1. de Beurs D, van Bruinessen I, Noordman J, Friele R, van Dulmen S. Active involvement of end users when developing web-based mental health interventions. Front Psychiatry 2017;8:72. [doi: 10.3389/fpsyt.2017.00072] [Medline: 28515699]
2. Timmers T, van der Weegen W, Janssen L, Kremer J, Kool RB. Importance of patient involvement in creating content for ehealth interventions: qualitative case report in orthopedics. JMIR Form Res 2022 Nov 3;6(11):e39637. [doi: 10.2196/39637] [Medline: 36326799]
3. Song T, Yu P, Bliokas V, et al. A clinician-led, experience-based co-design approach for developing mHealth services to support the patient self-management of chronic conditions: development study and design case. JMIR mHealth uHealth 2021 Jul 20;9(7):e20650. [doi: 10.2196/20650] [Medline: 34283030]
4. Price M, Yuen EK, Goetter EM, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436. [doi: 10.1002/cpp.1855] [Medline: 23918764]
5. Sahin C. Rules of engagement in mobile health: what does mobile health bring to research and theory? Contemp Nurse 2018;54(4-5):374-387. [doi: 10.1080/10376178.2018.1448290] [Medline: 29502472]
6. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR mHealth uHealth 2018 Jan 17;6(1):e23. [doi: 10.2196/mhealth.8873] [Medline: 29343463]
7. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97. [doi: 10.2196/jmir.4883] [Medline: 27185295]
8. Stevens S, Gallagher S, Andrews T, Ashall-Payne L, Humphreys L, Leigh S. The effectiveness of digital health technologies for patients with diabetes mellitus: a systematic review. Front Clin Diabetes Healthc 2022;3:936752. [doi: 10.3389/fcdhc.2022.936752] [Medline: 36992773]
9. Alanzi T, Alanazi NR, Istepanian R, Philip N. Evaluation of the effectiveness of mobile diabetes management system with social networking and cognitive behavioural therapy (CBT) for T2D. mHealth 2018;4:35. [doi: 10.21037/mhealth.2018.06.05] [Medline: 30221168]
10. Bean C, Dineen T, Locke SR, Bouvier B, Jung ME. An evaluation of the reach and effectiveness of a diabetes prevention behaviour change program situated in a community site. Can J Diabetes 2021 Jun;45(4):360-368. [doi: 10.1016/j.jcjd.2020.10.006] [Medline: 33323314]
11. Dineen TE, Bean C, Jung ME. Successes and challenges from a motivational interviewing-informed diabetes prevention program situated in the community. Health Promot Pract 2024 Mar;25(2):274-284. [doi: 10.1177/15248399221115066] [Medline: 36047459]
12. MacPherson MM, Dineen TE, Cranston KD, Jung ME. Identifying behaviour change techniques and motivational interviewing techniques in small steps for big changes: a community-based program for adults at risk for type 2 diabetes. Can J Diabetes 2020 Dec;44(8):719-726. [doi: 10.1016/j.jcjd.2020.09.011] [Medline: 33279096]
13. Wang RH, Kenyon LK, McGilton KS, et al. The time is now: a FASTER approach to generate research evidence for technology-based interventions in the field of disability and rehabilitation. Arch Phys Med Rehabil 2021 Sep;102(9):1848-1859. [doi: 10.1016/j.apmr.2021.04.009] [Medline: 33992634]
14. MacPherson M, Grieve N, Gala B, Braaten K, Jung M. Development of a small steps for big changes diabetes prevention app: application of the development phase of FASTER. J Technol Behav Sci 2024;9(4):681-702. [doi: 10.1007/s41347-023-00373-8]
15. Yi JY, Kim Y, Cho YM, Kim H. Self-management of chronic conditions using mhealth interventions in Korea: a systematic review. Healthc Inform Res 2018 Jul;24(3):187-197. [doi: 10.4258/hir.2018.24.3.187] [Medline: 30109152]
16. Liu S, La H, Willms A, Rhodes RE. A "no-code" app design platform for mobile health research: development and usability study. JMIR Form Res 2022 Aug 18;6(8):e38737. [doi: 10.2196/38737] [Medline: 35980740]
17. Georgsson M, Staggers N, Årsand E, Kushniruk A. Employing a user-centered cognitive walkthrough to evaluate a mHealth diabetes self-management application: a case study and beginning method validation. J Biomed Inform 2019 Mar;91:103110. [doi: 10.1016/j.jbi.2019.103110] [Medline: 30721757]
18. Beauchemin M, Gradilla M, Baik D, Cho H, Schnall R. A multi-step usability evaluation of a self-management app to support medication adherence in persons living with HIV. Int J Med Inform 2019 Feb;122:37-44. [doi: 10.1016/j.ijmedinf.2018.11.012] [Medline: 30623782]
19. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]

XSL•FO
RenderX

20. Hyzy M, Bond R, Mulvenna M, et al. System usability scale benchmarking for digital health apps: meta-analysis. JMIR mHealth uHealth 2022 Aug 18;10(8):e37290. [doi: 10.2196/37290] [Medline: 35980732]
21. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development validation of the user version of the mobile application rating scale and (uMARS). JMIR mHealth uHealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849] [Medline: 27287964]
22. Davis R, Gardner J, Schnall R. A review of usability evaluation methods and their use for testing ehealth hiv interventions. Curr HIV/AIDS Rep 2020 Jun;17(3):203-218. [doi: 10.1007/s11904-020-00493-3] [Medline: 32390078]
23. Roque NA, Boot WR. A new tool for assessing mobile device proficiency in older adults: the mobile device proficiency questionnaire. J Appl Gerontol 2018 Feb;37(2):131-156. [doi: 10.1177/0733464816642582] [Medline: 27255686]
24. Tripathi D, Vikram NK, Chaturvedi S, Bhatia N. Development of “DiabetesSutra” a mobile application for lifestyle management of type 2 diabetes in India. J Diabetes Metab Disord 2024 Jun;23(1):709-720. [doi: 10.1007/s40200-023-01339-6] [Medline: 38932794]
25. Nielsen J. Usability inspection methods. Presented at: Conference Companion on Human Factors in Computing Systems - CHI 94; Apr 24-28, 1994; Boston, Massachusetts, United States p. 413-414 URL: http://portal.acm.org/citation.cfm?doid=259963 [accessed 2024-01-31] [doi: 10.1145/259963.260531]
26. Kushniruk AW, Monkman H, Tuden D, Bellwood P, Borycki EM. Integrating heuristic evaluation with cognitive walkthrough: development of a hybrid usability inspection method. Stud Health Technol Inform 2015;208:221-225. [doi: 10.3233/978-1-61499-488-6-221]
27. Byambasuren O, Beller E, Hoffmann T, Glasziou P. Barriers to and facilitators of the prescription of mhealth apps in australian general practice: qualitative study. JMIR mHealth uHealth 2020 Jul 30;8(7):e17447. [doi: 10.2196/17447] [Medline: 32729839]
28. Goswami R, Garner SL, George CE. Usability of an mHealth application in India for type 2 diabetes mellitus. Health Educ J 2023 Feb;82(1):108-116. [doi: 10.1177/00178969221138483]
29. Istepanian RSH, Al-Anzi TM. m-Health interventions for diabetes remote monitoring and self management: clinical and compliance issues. mHealth 2018;4:4. [doi: 10.21037/mhealth.2018.01.02] [Medline: 29552566]
30. MacPherson MM, Merry KJ, Locke SR, Jung ME. mHealth prompts within diabetes prevention programs: a scoping review. mHealth 2022;8:20. [doi: 10.21037/mhealth-21-22] [Medline: 35449504]
31. Torbjørnsen A, Ribu L, Rønnevig M, Grøttland A, Helseth S. Users’ acceptability of a mobile application for persons with type 2 diabetes: a qualitative study. BMC Health Serv Res 2019 Sep 6;19(1):641. [doi: 10.1186/s12913-019-4486-2] [Medline: 31492176]

Abbreviations

CW: cognitive walkthrough
FASTER: Framework for Accelerated and Systematic Technology-Based Intervention Development and Evaluation Research
HbA_{1c}: hemoglobin A_{1c}
MDPQ-16: Mobile Device Proficiency Questionnaire
mHealth: mobile health
SUS: System Usability Scale
uMARS: User-Mobile Application Rating Scale

Edited by A Bucher; submitted 10.04.24; peer-reviewed by A Gandhi, K Rawlings, M Braun, Y Kim; revised version received 08.11.24; accepted 09.11.24; published 11.02.25.

Please cite as:
Grieve N, Braaten K, MacPherson M, Liu S, Jung ME
Involving End Users in the Development and Usability Testing of a Smartphone App Designed for Individuals With Prediabetes: Mixed-Methods Focus Group Study
JMIR Form Res 2025;9:e59386
URL: https://formative.jmir.org/2025/1/e59386
doi:10.2196/59386

© Natalie Grieve, Kyra Braaten, Megan MacPherson, Sam Liu, Mary E Jung. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete

XSL•FO
RenderX

bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Activity Trackers in Physical Therapy for People With Chronic Obstructive Pulmonary Disease in the Netherlands: Cross-Sectional Study on Current Use and Implementation Determinants

Darcy Ummels[1,2], PhD; Esther Bols[1], PhD; Roel Johannes Anna Frantzen[1], BSc; Tim Frantzen[1], BSc; Levy Robeerts[1], BSc; Emmylou Beekman[1,3], PhD

[1]Research Centre for Autonomy and Participation of Persons with a Chronic Illness, Academy for Speech and Language Therapy, Zuyd University of Applied Sciences, Heerlen, Netherlands

[2]Department of Rehabilitation Medicine, Care and Public Health Research Institute, Faculty of Health, Medicine & Life Sciences, Maastricht University, Maastricht, Netherlands

[3]Department of Family Medicine, Care and Public Health Research Institute, Faculty of Health, Medicine & Life Sciences, Maastricht University, Maastricht, Netherlands

Corresponding Author:
Darcy Ummels, PhD
Research Centre for Autonomy and Participation of Persons with a Chronic Illness, Academy for Speech and Language Therapy, Zuyd University of Applied Sciences, Heerlen, Netherlands

Abstract

Background: In the Netherlands, 545,900 people experienced chronic obstructive pulmonary disease (COPD) in 2022. Physical therapy following the Royal Dutch Society for Physiotherapy (Koninklijk Nederlands Genootschap voor Fysiotherapie) guidelines for COPD treatment is a proven effective treatment for people with COPD. The revised guidelines include a new recommendation: a patient's physical activity level should be assessed with an activity tracker (AT). Literature shows that the implementation of eHealth in clinical practice, in this case, ATs, is challenging.

Objective: This study aims (1) to assess how and why ATs are currently used in physical therapy in patients with COPD and (2) to determine which barriers and facilitators are of relevance for optimal implementation of ATs during the clinical reasoning process of physical therapists in patients with COPD.

Methods: A cross-sectional study was used to evaluate the implementation of ATs in physical therapy. Included participants were physical therapists who were affiliated with Chronisch ZorgNet and had a specialization in COPD treatment. The survey content was based on the Consolidated Framework for Implementation Research, the theory of planned behavior, the framework "experiences of patients with commercially available ATs," and the Koninklijk Nederlands Genootschap voor Fysiotherapie guidelines for COPD. Physical therapists were questioned via a digital survey.

Results: In total, 211 completed surveys were analyzed. Of the 211 participating physical therapists, 108 (51.2%) used ATs, whereas most of them (n=82, 75.9%) already used ATs before it was advised in the guidelines. Physical therapists indicated that the most important reason to use ATs is that they experience it as an added health care value. Both users and nonusers indicated that the most important reason why they do not use ATs is because their patients do not want to use an AT. The second reason was a lack of knowledge in the nonuser group. Moreover, both users and nonusers indicated that the implementation of ATs was not prepared and planned for within their center.

Conclusions: Overall, these results show that ATs are not yet fully implemented in the Dutch general physical therapy practice in patients with COPD, as recommended by current evidence-based guidelines. Physical therapists need guidance for the successful implementation of ATs. This could be accomplished by providing training for physical therapists, integrating ATs into the education of (future) physical therapists, and providing support during the implementation process of ATs for both the physical therapists and management.

(JMIR Form Res 2025;9:e59533) doi:10.2196/59533

KEYWORDS

physical therapy; activity tracker; implementation; technology; COPD; chronic obstructive pulmonary disease; eHealth; meaningful use; health measurement; active lifestyle; Netherlands; physical activity; barriers; therapy

XSL•FO
RenderX

Introduction

Chronic obstructive pulmonary disease (COPD) is a heterogeneous lung condition characterized by chronic respiratory symptoms (dyspnea, cough, sputum production, and/or exacerbations) due to abnormalities of the airways (bronchitis and bronchiolitis) and/or alveoli (emphysema) that cause persistent, often progressive, airflow obstruction [1,2]. In the Netherlands, 545,900 people experienced COPD in 2022. It is the fourth leading disease in relation to disease burden and the sixth leading disease that causes mortality in the Dutch population [3]. The health care expenses for people with COPD were estimated at US $775 million in 2019, totaling 1% of the total Dutch health care expenses and 23% of the total expenditure for all respiratory diseases [4]. Physical therapy according to the Royal Dutch Society for Physiotherapy (Koninklijk Nederlands Genootschap voor Fysiotherapie [KNGF]) guidelines for COPD treatment [5] is a proven effective treatment for people with COPD. In general, the treatment is a major component of pulmonary rehabilitation and aims to increase a patient's physical activity level and physical capacity and is accomplished with home exercises and education [5]. The KNGF guidelines describe the clinical reasoning process for physical therapists based on the latest evidence. In the revised guidelines, the new recommendation is that a patient's physical activity level (eg, number of steps) is measured with an activity tracker (AT). The patient's physical activity level is an important starting point, as it codetermines whether physical therapy is indicated or a lifestyle coach is sufficient. Therefore, if ATs are not or incorrectly used, the clinical reasoning process is compromised. This could result in misclassification of patients and hence the wrong allocation to physical therapy or lifestyle coaching.

Apart from assigning patients to treatment, ATs can also be used to monitor and evaluate the physical activity level during and after the treatment [6]. Using ATs as an intervention can also enhance the physical activity level of patients with COPD when combined with the guidance of a health care professional [7]. Moreover, they can improve patients' self-management and self-efficacy [8-10].

Literature shows that the implementation of eHealth in clinical practice, in this case, ATs, is challenging [6,11,12]. Barriers in the implementation of eHealth are among others, the lack of knowledge among health care professionals concerning eHealth possibilities, unclear benefits of eHealth, resistance to change, and clinicians' perception of impersonal care [13-16]. The degree of actual use of ATs and areas of application by physical therapists in patients with COPD are currently unclear. In addition, more insight into potential barriers and facilitators for successful future implementation of ATs is necessary to optimize the treatment of patients with COPD.

A theoretical framework to assess potential barriers and facilitators for implementing innovations is the Consolidated Framework for Implementation Research (CFIR) [17]. The CFIR consists of 5 domains associated with effective implementation: intervention characteristics, inner setting, characteristics of individuals, outer setting, and process. Besides assessing potential barriers and facilitators for successful implementation, physical therapists need to change their current behavior. The theory of planned behavior (TPB) supports predicting an individual's intention to engage in a (new) behavior. The TPB consists of 4 aspects: attitude, subjective norms, perceived behavioral control, and intention [18].

This study combines the aspects of the CFIR and the TPB as a background (1) to assess how and why ATs are currently used in physical therapy in patients with COPD and (2) to determine which barriers and facilitators are of relevance for optimal implementation of ATs during the clinical reasoning process of Dutch physical therapists in patients with COPD.

Methods

A cross-sectional study was performed, in which Dutch physical therapists were questioned via a voluntary digital closed survey.

Population

Included participants were physical therapists who were affiliated with Chronisch ZorgNet and had a specialization in COPD treatment. Chronisch ZorgNet is a nationwide network that provides training for treating patients with chronic diseases according to the latest evidence (guidelines). The network consists of 3500 physical therapists working in primary, secondary, and tertiary care, of which 1200 physical therapists are specialized in lung physical therapy [19]. This network provided access to this target population. The survey was distributed by emails from Chronisch ZorgNet. All specialized lung physical therapists who were Chronisch ZorgNet members were eligible to participate in this survey. Physical therapists were excluded from this study if they treated less than 1 unique patient with COPD per week.

Survey

The survey (Multimedia Appendix 1) was drafted from documents based on international scientific literature: (1) the checklist "the use of measurement instruments in clinical practice" [20] based on the CFIR [17] (Q1), (2) the questionnaire "meaningful use of patient-specific goal setting within the clinical reasoning of physical therapists" [21] based on the TPB [18] (Q2), (3) the framework "experiences of patients with commercially available ATs" [22] (Q3), and (4) the KNGF guidelines for COPD [5]. Questions of the checklist Q1 and the questionnaire Q2 were adapted for the application of ATs instead of general measurement instruments. To gain insight into relevant barriers and facilitators, the topics of the framework Q3 were used. The survey was supplemented with questions based on the recommendations for the use of ATs in the COPD guidelines [5]. In total, 9 questions determined the eligibility of the participants (eg, sex, age, and work experience). Based on the question "I use ATs in people with COPD" (yes or no), participants answered either 59 questions if they used ATs or 33 questions if they did not.

Of the 59 questions for the AT users, 26 were based on the checklist "the use of measurement instruments in clinical practice"' Q1 [20], 5 were based on questionnaire Q2 [21], 20 were based on the framework Q3 [22], and 8 were based on the COPD guidelines [5]. Of the 33 questions for the nonusers, 19

XSL•FO
RenderX

were based on Q1 [20], 5 were based on Q2 [21], and 9 were based on Q3 [22]. The questions were either answered on a 7-point Likert scale (0=totally disagree and 7=totally agree), on a multiple-answer scale, or were open comment questions. Only 4 questions were necessarily answered by yes or no (eg, ATs are provided to our patients: yes or no). Participants were able to review and change their answers before completing the survey.

To ensure the content validity of the combined questionnaires, the draft survey was sent to 5 experts for feedback on the usability and face validity of the survey: 2 experts in the field of measurement tools in health care, 1 expert in the field of ATs in health care, and 2 specialized physical therapists. The feedback required no changes to the survey and the final survey was constructed (Multimedia Appendix 1).

Data collection was performed in May 2021. The survey was completed via clickable link, and QR code in the email leading to the survey in QuestBack Essentials (2021). For each question, a digital page emerged; in total, there were between 33 and 59 pages excluding the start- and the end page and 7 questions regarding population characteristics. All questions were mandatory questions and needed to be completed before the survey automatically continued. One reminder was sent after 1 week. Participants were asked to answer the questions as accurate as possible, and it was explicitly addressed that the results were analyzed anonymously. Data were stored in a secured data repository of Zuyd University of Applied Sciences for 15 years, and only the participating researchers had access to these data.

Data Analyses

Data analyses were performed using SPSS Statistics (version 23.0; IBM Crop). Only fully completed surveys were analyzed. Descriptive statistics of the participants' characteristics were presented as a number (percentage) for the categorical variables sex, workplace setting, and whether they completed a recent COPD course. For the continuous variables, age, work experiences, and the number of treated unique patients with COPD per week were presented as a mean (SD). Answers to questions with a 7-point Likert scale were presented as mean (SD). Responses to open comment questions were treated as qualitative data and categorized into themes by using directed content analyses via inductive coding [23] and presented as a number (percentage) per theme. Open answers that did not fit any code were labeled as other. Responses were separately analyzed for users and nonusers of ATs. In case of any missing data, listwise deletion was applied.

Ethical Considerations

No ethics approval was required since the research was not subject to the Medical Research Involving Human Subjects Act (Wet medisch-wetenschappelijk onderzoek met mensen). The research was part of continuous quality monitoring and improvement of daily COPD care within a national quality network (Chronisch ZorgNet). This study did not require written informed consent. Participating physical therapists were informed about the purpose of this survey prior to commencing the survey. Physical therapists voluntarily participated in this study. Study results were processed anonymously. Physical therapists were always able to end the survey without any consequences. There was no compensation for the physical therapists. The CHERRIES (Checklist for Reporting Results of Internet E-Surveys) was used (Checklist 1).

Results

Population Characteristics

In total, 216 physical therapists completed the survey. Five physical therapists were excluded because they did not meet the inclusion criteria (treating less than 1 unique patient with COPD per week), leaving 211 responses for analysis. There were no missing data. The characteristics of the physical therapists are displayed in Table 1.

Table . Characteristics of the participating physical therapist (N=211).

Characteristic	Participants
Sex (male), n (%)	74 (35.1)
Age (years), mean (SD)	46.5 (11.7)
Work experience (years), mean (SD)	22.8 (11.1)
Work setting, n (%)	
Primary care	197 (93.5)
Secondary care	5 (2.5)
Primary and secondary care	8 (3.8)
Tertiary care	1 (0.5)
Number of unique patients with COPD[a] per week, mean (SD)	8.2 (6.5)
Completed a recent COPD course, n (%)	208 (98)

[a]COPD: chronic obstructive pulmonary disease.

Results of the survey are presented per research questions.

How ATs Are Used

Of the 211 participating physical therapists, 108 (51.2%) used an AT, whereas most of them (n=82, 75.9%) already used an

XSL•FO
RenderX

AT before it was advised in the guidelines. Most physical therapists measured the number of steps (n=64, 55.2%) and used a standard app on a smartphone (n=97, 89.8%). Other parameters named were, for example, activities of daily living, distance, physical capacity, and heart rate. The 2 most common goals for the use of an AT were to inventory the physical activity level (n=94, 87%) and to stimulate the physical activity level of the patient (n=88, 81.5%). All results on how ATs were used in physical therapy in patients with COPD are presented in Table 2.

XSL•FO
RenderX

Table . Results on how activity trackers (ATs) are currently used in physical therapy in patients with chronic obstructive pulmonary disease (COPD; N=211; based on digital survey).

Question	User (n=108)	Nonuser (n=103)
I use AT in people with COPD, n (%)	• Yes: 108 (100)[a]	• No: 103 (100)[a]
Before the release of the COPD guidelines, I already used AT in people with COPD, n (%)	• Yes: 82 (75.9) • No: 26 (24.1)	—[b]
It is clear what I want to measure with an AT[c], mean (SD)	• 5.8 (1.3)	—
Which parameter or concept do you measure with an AT? n (%)	• Number of steps: 64 (55.2) • Physical activity: 43 (37.1) • Other[d]: 9 (7.7)	—
It is clear to me with which AT I can measure these parameters[c], mean (SD)	• 5.3 (1.5)	—
There are agreements within our center about the choice of the used type of AT[c], mean (SD)	• 3.2 (1.9)	—
Within our center, we use the following type of AT, n (%)	• App on a smartphone: 97 (89.8) • Commercially available pedometer: 57 (52.7) • Commercially available accelerometer: 14 (12.9) • Already owned by patient: 2 (1.8)	—
Within our center, we use the following brand AT, n (%)	• Standard app on a smartphone: 26 (24.1) • Unknown brand: 23 (21.3) • Already owned by patient: 14 (12.9) • Fitbit: 12 (11.1) • Omron: 8 (7.4) • Other[d]: 7 (6.4) • Yamax: 3 (2.7) • McRobberts: 2 (1.8)	—
The AT is worn by people with COPD on the following locations, n (%)	• Trouser pocket: 88 (81.4) • Wrist: 52 (48.1) • Hip: 13 (12) • Bag: 13 (12) • Other[d]: 5 (4.6) • Chest pocket: 4 (3.7) • Lower back: 2 (1.8) • Ankle: 2 (1.8)	—
Which brand of AT would you recommend to your patients? n (%)	• Do not know: 49 (45.3) • Apps: 23 (21.2) • Fitbit: 21 (19.2) • I do not recommend a brand: 6 (5.5) • Garmin: 5 (4.6) • Other: 5 (4.6) • Apple: 4 (3.7)	—
I stimulate my patients to purchase an AT[c], mean (SD)	• 4.9 (1.4)	—
At this moment, in how many unique patients with COPD per week are you using an AT? mean (SD)	• 5.5 (4.2)	—
I use AT for the following goals in patients with COPD, n (%)	• Inventory: 94 (87) • Stimulation: 88 (81.5) • Allocating: 64 (59.2) • Evaluation: 60 (55.5)	—

XSL•FO
RenderX

Question	User (n=108)	Nonuser (n=103)
My patients with COPD wear an AT for the recommended number of days (7 days)[c], mean (SD)	• 4.4 (1.5)	—

[a]Of the 211 respondents, 108 (51.2%) answered "yes" and 103 (38.8) answered "no."

[b]Not available.

[c]Likert scale with a range of 0 - 7 (0=totally disagree and 7=totally agree).

[d]Multiple open answers which were categorized as other.

Why ATs Are Used

Physical therapists who used ATs because they believed that it is an added value to health care (n=46, 42.5%) and because the guidelines recommend to use ATs (n=44, 40.7%). Physical therapists considered the cutoff value of 5000 steps per day for patients with COPD to be a realistic measure of sufficient activity (mean 4.7, SD 1.5). The primary reason physical therapists did not use ATs both among users (n=58, 53.7%) and nonusers (n=35, 33.9%) was that patients themselves indicated that they did not want to use them. In the nonusers group, lack of knowledge (n=31, 30.9%) was the second important reason for nonuse. Other reasons mentioned were, for example, bad experiences, difficulty implementing ATs, and concerns about the clinometric properties of the ATs.

In most centers, there was no consensus on why they were going to use ATs (mean 3.4, SD 1.8). Users scored high on questions regarding whether ATs contributed to the assessment of physical activity levels (mean 5.8, SD 1.1), allocation of patients into profiles (mean 5.8, SD 1.1), evaluation of physical activity levels (mean 5.9, SD 0.9), and stimulation of physical activity levels (mean 5.9, SD 1.0). Results on why ATs were used in physical therapy in patients with COPD are presented in Table 3.

XSL•FO
RenderX

Table . Results on why activity trackers (ATs) are currently used in physical therapy in patients with chronic obstructive pulmonary disease (COPD; N=211; based on digital survey).

Question	User (n=108)	Nonuser (n=103)
Within our center, we have agreed on why we (are going to) use AT[a], mean (SD)	• 3.4 (2.0)	• 3.2 (1.9)
The choice for the concerned AT was made based on, n (%)	• Availability: 73 (67.6) • Costs: 49 (45.3) • Feasibility: 43 (39.8) • Own experience: 19 (17.5) • Validity: 15 (13.8) • Already owned by the center: 5 (4.6) • Other[b]: 6 (5.5%)	—[c]
I want to measure the following specific activities, n (%)	• Walking: 105 (97.2) • Bicycling: 52 (48.1) • Running: 11 (10.1) • Activities of daily living: 8 (7.4) • Swimming: 7 (6.4) • Other[b]: 5 (4.6) • Walking stairs: 2 (1.8)	—
I want to measure the following parameters, n (%)	• Number of steps: 104 (96.2) • Active minutes: 73 (67.5) • Walked distance: 53 (49) • Heart rate: 35 (32.4) • Passive minutes: 34 (31.4) • Calories: 13 (11.1) • Other[b]: 3 (2.7)	—
I think the cutoff value of 5000 steps per day to be sufficiently active is a realistic cutoff value[a], mean (SD)	• 4.7 (1.5)	—
The most important reason I am using an AT in patients with COPD is, n (%)	• Because it is an added health care value: 46 (42.5) • Because of the guidelines: 44 (40.7) • Own interest: 9 (8.3) • Other[b]: 5 (4.6) • Because it is mandatory in my network: 4 (3.7)	—
When I use an AT, it contributes to the inventory of the physical activity level and the physiotherapeutic diagnosis of a patient[a], mean (SD)	• 5.8 (1.1)	—
When I use an AT, it contributes to allocating patients into the profiles of the guidelines[a], mean (SD)	• 5.8 (1.1)	—
When I use an AT, it contributes to evaluating the physical activity during the treatment process in patients with COPD[a], mean (SD)	• 5.9 (0.9)	—
When I use an AT, it contributes to stimulating physical activity as part of my intervention[a], mean (SD)	• 5.9 (1.0)	—
I think AT should be used in patients with COPD[a], mean (SD)	• 5.2 (1.7)	• 4.3 (1.7)
My patients with COPD think it is important to (start) using AT[a], mean (SD)	• 4.1 (1.2)	• 3.9 (2.0)
My patients with COPD are more aware of their functioning by using an AT[a], mean (SD)	• 5.7 (1.1)	• 5.8 (1.4)

XSL•FO
RenderX

Question	User (n=108)	Nonuser (n=103)
Health care centers, comparable to my center are using AT[a], mean (SD)	• 4.2 (0.9)	• 3.3 (1.2)
External organizations are obligating the use of AT[a], mean (SD)	• 3.4 (1.8)	• 2.6 (1.6)
The most important reason I do not use AT in some patients with COPD is, n (%)	• Patients do not want to: 58 (53.7) • No added value: 13 (12) • No knowledge: 9 (8.3) • No AT available: 8 (7.4) • Costs: 7 (6.4) • Other[b]: 6 (5.5)	• Patients do not want to: 35 (33.9) • No knowledge: 31 (30.9) • No AT available: 9 (8.7) • Bad experience: 8 (7.7) • Other[b]: 8 (7.7) • No added value: 7 (6.7) • Costs: 7 (6.7)

[a]Likert scale with a range of 0 - 7 (0=totally disagree and 7=totally agree).

[b]Multiple open answers which were categorized as other.

[c]Not available.

Barriers and Facilitators for Implementation of ATs

No remarkable differences exist between users and nonusers regarding the experienced barriers and facilitators. For both groups, a relevant barrier to using ATs was the costs (users: n=75, 69.4% vs nonusers: n=77, 74.7%) and the cognitive or communication skills of the patient (users: n=75, 69.4% vs nonusers: n=50, 48.5%). Other reasons that were mentioned were prioritizing the patient over scientific pursuits, privacy concerns, and AT fails to measure parameters that therapists find meaningful. Important facilitators in both groups for using ATs were the motivation of the patient (users: n=90, 83.3% vs nonusers: n=70, 67.9%) and the user-friendliness of the AT (users: n=64, 59.2% vs nonusers: n=57, 55.3%). Both users (mean 3.6, SD 1.8) and nonusers (mean 2.5, SD 1.6) indicated that they lack the knowledge to provide a recommendation of commercially available ATs to their patients. Likewise, both groups indicated insufficient education or coaching possibilities for the health care professionals on ATs within their center (users: mean 3.7, SD 1.9 and nonusers: mean 2.7, SD 1.8). Further barriers and facilitators are presented in Table 4.

XSL•FO
RenderX

Table . Results on relevant barriers and facilitators for optimal implementation of activity trackers (ATs) during the clinical reasoning process of Dutch physical therapists in patients with chronic obstructive pulmonary disease (COPD; N=211; based on digital survey).

Question	User (n=108)	Nonuser (n=103)
I have sufficient knowledge to recommend commercially available AT to my patients regarding price, functions, and quality[a], mean (SD)	• 3.6 (1.8)	• 2.5 (1.6)
I am capable to draft personalized physical activity goals together with my patient[a], mean (SD)	• 5.8 (1.0)	• 5.8 (1.3)
I am capable to teach my patients about the functions, use, and interpretations of AT[a], mean (SD)	• 5.4 (1.3)	• 4.9 (1.8)
ATs are provided to our patients, n (%)	• No: 71 (65.7) • Yes: 37 (34.3)	• No: 99 (96.1) • Yes: 4 (3.9)
The following factors stimulate the use of AT during the clinical reasoning in patients with COPD, n (%)	• Motivation for the patient: 90 (83.3) • User-friendliness: 64 (59.2) • The added value: 59 (54.6) • Reliability and validity: 38 (35.1) • Costs: 28 (25.9) • Time investment: 25 (23.1)	• Motivation for the patient: 70 (67.9) • User-friendliness: 57 (55.3) • The added value: 54 (52.4) • Reliability and validity: 40 (38.8) • Costs: 25 (24.2) • Time investment: 18 (17.4) • Other[b]: 5 (4.8)
The following factors obstruct the use of AT during clinical reasoning in patients with COPD, n (%)	• Costs: 75 (69.4) • Cognitive or communication skills of the patient: 75 (69.4) • Time investment for the therapist: 37 (34.2) • User-friendliness: 36 (33.3) • Time investment for the patient: 34 (31.4) • Reliability and validity: 26 (24) • Other[b]: 3 (2.7)	• Costs: 77 (74.7) • Cognitive or communication skills of the patient: 50 (48.5) • Time investment for the therapist: 46 (44.6) • User-friendliness: 45 (43.6) • Reliability and validity: 40 (38.8) • Time investment for the patient: 40 (38.8) • Other[b]: 7 (6.7)
I know how I can measure physical activity with an AT[a], mean (SD)	• 5.5 (1.4)	—[c]
I know how I can interpret the results of the AT[a], mean (SD)	• 5.4 (1.2)	—
I know how I can discuss the results of the AT with my patients[a], mean (SD)	• 5.8 (1.0)	—
The use of AT is part of our mission and vision and is incorporated into our policy[a], mean (SD)	• 5.2 (2.0)	• 3.8 (2.3)
The use of AT fits in our care process[a], mean (SD)	• 5.6 (1.2)	• 4.6 (1.9)
There is sufficient support and involvement from the management when using AT[a], mean (SD)	• 5.2 (1.9)	• 4.1 (1.9)
In our center, there is someone who guides the implementation process of AT[a], mean (SD)	• 2.7 (2.0)	• 2.2 (1.7)
In our center, there is sufficient education or coaching for the health care professionals about AT[a], mean (SD)	• 3.7 (1.9)	• 2.7 (1.8)
In our center, there is sufficient time to try to use AT[a], mean (SD)	• 4.0 (1.9)	• 3.4 (1.7)
In our center, the use of AT is incorporated in our electronic patients' files[a], mean (SD)	• 4.5 (2.2)	• 3.4 (2.3)
In our center, there are sufficient AT available, n (%)	• No: 63 (58.3) • Yes: 24 (22.2) • Not applicable: 20 (18.5)	• No: 67 (65) • Yes: 5 (4.8) • Not applicable: 21 (20.3)

XSL•FO
RenderX

Question	User (n=108)	Nonuser (n=103)
I am open to (start to) using AT in patients with COPD[a], mean (SD)	• 6.5 (0.8)	• 5.8 (1.4)
I am planning to continue or start using AT in patients with COPD[a], mean (SD)	• 6.4 (0.8)	• 5.2 (1.5)
I have sufficient knowledge to (start to) use AT[a], mean (SD)	• 5.5 (1.3)	• 4.3 (1.7)
I trust I can use AT in patients with COPD[a], mean (SD)	• 5.5 (1.2)	• 4.8 (1.7)
My colleagues are open to (start) using an AT[a], mean (SD)	• 5.5 (1.5)	• 5.2 (2.0)
My colleagues have sufficient skills and knowledge to (start) using AT[a], mean (SD)	• 5.1 (1.7)	• 4.7 (2.2)
My patients with COPD are open to (start) using AT[a], mean (SD)	• 4.6 (1.1)	• 4.0 (1.8)
My patients with COPD have sufficient skills and knowledge to (start) using an AT[a], mean (SD)	• 4.0 (1.4)	• 3.6 (2.0)
My patients with COPD experience commercially available AT as technical complex[a], mean (SD)	• 5.1 (1.3)	• 5.7 (1.8)
In our center, the implementation of AT was prepared and planned[a], mean (SD)	• 3.5 (2.3)	• 2.7 (2.4)
In our center, we started using AT on a trial basis and started to adapt our care process[a], mean (SD)	• 4.2 (1.9)	• 2.2 (1.5)
In our center, relevant persons were involved in the implementation process: health care professionals, patients (representatives), and supporting professionals (policy)[a], mean (SD)	• 3.4 (2.0)	• 2.9 (2.5)
In our center, the use of AT are frequently evaluated and improved[a], mean (SD)	• 3.0 (1.9)	• 2.9 (2.8)

[a]Likert scale with a range of 0 - 7 (0=totally disagree and 7=totally agree).

[b]Multiple open answers which were categorized as other.

[c]Not available.

Discussion

Principal Findings

This research aimed to assess how and why ATs are currently used in physical therapy in patients with COPD and to determine which barriers and facilitators are of relevance for optimal implementation of ATs during the clinical reasoning process of Dutch physical therapists.

It is known that the implementation of an innovation is often challenging [11]. However, of the 211 participating physical therapists, 108 (51.2%) used an AT, and most of them (n=82, 75.9%) already used an AT before it was advised in the guidelines. This relatively high percentage could be explained by the fact that physical therapists indicated that the most important reason for using ATs is the experienced added health care value. Another explanation could be that the included physical therapists were affiliated with Chronisch ZorgNet and might therefore be more sensitized to provide evidence-based care. Furthermore, it was notable that regardless whether a physical therapist was a user or nonuser, the barriers and facilitators for using ATs were roughly the same. Both users and nonusers indicated that the most important reason for not using ATs is because their patients do not want to use an AT. However, it is debatable whether this is an assumption of the physical therapist or the actual opinion of the patient, as a recent review describes that patients with COPD in general enjoyed using ATs and found using the ATs easy [24]. The second (nonusers) and third (users) reason for physical therapists to not use ATs was a general lack of knowledge. Physical therapists indicated that they did not know which brand of commercially available AT they could recommend to their patients, which is related to the experienced lack of knowledge regarding price, functions, and quality. In contradiction, knowledge regarding how to measure with ATs, how to interpret the data of the AT, and how to discuss the data of the AT with a patient was scored

XSL•FO
RenderX

high by the physical therapists. A possible explanation could be that therapists lack different knowledge: the knowledge on how to measure with ATs, how to interpret the data of the AT, and how to discuss the data of the AT.

Comparison to Prior Work

The general adaption rate of eHealth in physical therapy seems to be similar to the adaption rate of 52% found in this study [25,26]. However, the use of the ATs in this study was self-reported; therefore, the presence of biases in the reported use and perception of ATs might be present. Most physical therapists used an app on a smartphone as an AT, which links to the fact that availability and costs were important factors when selecting an AT. Most smartphones automatically measure step count, and apps are mostly free of use. However, the validity of apps regarding measuring physical activity is low, especially in older adults during low walking speed, as is the case in activities of daily living [27]. Most physical therapists considered the cutoff value of 5000 steps per day to be sufficiently active, used in the new national guidelines, to be realistic. This is valuable information, considering that this cutoff value in the revised guidelines is based on previous studies. Several studies mention a minimum of 7500 - 10,000 steps per day in order to have a healthy active lifestyle [28-30]. However, it seems that a healthy person reaches only 5500 - 6000 steps per day on average, and people with a chronic disease 3500 - 5500 steps per day [29-31]. Consequently, in the national guidelines, a cutoff value of 5000 steps per day is used for people with COPD to be sufficiently active [32,33]. This study suggests that 5000 steps per day might be a minimal clinical important threshold in daily physical therapy practice.

Both users and nonusers indicated that the implementation of ATs was not prepared and planned for within their center. To successfully implement eHealth, internal and external facilitation, audit and feedback, management support, and training of clinicians are essential according to a recent systematic review [34] and are in line with the CFIR [17]. Wilde et al [35] similarly showed the physical therapists' need for guidance and information to support the integration of ATs within the practice. A solution could be to provide education and training to management and physical therapists. Management should be trained to guide an implementation trajectory; however, the use of implementation support practitioners could also be an option [36]. Implementation support practitioners are professionals who support organizations, leaders, and staff in their implementation of evidence-informed practices and policies [37]. Current physical therapists and physical therapy students should be educated and trained. To optimize the adoption of ATs in clinical settings, educational interventions should be developed based on the best available evidence, clinical expertise, and patient values and circumstances [38].

However, it is important to train physical therapists' and students' general eHealth competencies in order to use eHealth in a broader spectrum than ATs only. It is suggested that education consists of 5 competency themes: information communication technology attitudes and skills, interpretation and analysis of eHealth data, support and guidance, communication skills, and privacy and confidentiality [39]. Current health care professionals are insufficiently trained in all of these competencies and therefore struggle to perform these skills in their daily clinical practice [40]. A better understanding and more digital competency could also lead to better experiences and better outcomes. On the other hand, a previous study showed that patients with chronic diseases (including COPD) experience ATs as technically complicated and wish to receive support by their physical therapist during the use within a diagnostic and/or therapeutic process [22]. Currently, several (digital) tools and websites are available to support health care professionals in the goal-oriented selection of eHealth [41,42].

Strengths and Limitations

Several limitations of this study need to be mentioned. First, this study addressed technology acceptance on a limited basis. The level of technology acceptance might influence whether ATs are used or not. Studies showed that if health care professionals are already experienced eHealth users, they report fewer implementation barriers and experience more advantages (eg, a more positive attitude toward eHealth) [43]. Second, participating physical therapists were affiliated with Chronisch ZorgNet, an organization in which physical therapists treating patients with chronic diseases are trained to work evidence-based. This might have biased the results since affiliated physical therapists might be more prone to adhere to recent guidelines and be motivated to implement innovations. On the other hand, all participating physical therapists were members of Chronisch ZorgNet, which includes almost all physical therapists who are specialized in lung physical therapy. Therefore, the study sample constitutes a representative sample for primary care. A strength of this study is that the survey was based on relevant international scientific literature and models in relation to implementation and behavioral change (CFIR and TPB), which enables generalization beyond the Dutch context [5,13-18,22,44].

Conclusions

Overall, these results show that ATs are not yet fully implemented in the Dutch general physical therapy practice in patients with COPD, as recommended by current evidence-based guidelines. Both users and nonusers indicated that the most important reason for not using ATs is because their patients do not want to or at least the health care professionals experienced it that way. However, it is debatable whether this is an assumption of the physical therapist or the actual opinion of the patient. Physical therapists need guidance for the successful implementation of ATs. This could be accomplished by providing training for current physical therapists, integrating ATs into the education of (future) physical therapists, and providing support during the implementation process of ATs for both the physical therapists and management.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank Chronisch ZorgNet for their assistance and coordination in data collection and all participating physical therapists for their responses. The authors thank Anita Stevens for her expertise on adapting questionnaire Q2. The authors declare that no generative artificial intelligence was used in writing this manuscript.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to the restrictions of the quality network Chronisch ZorgNet but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire for both users and nonusers.
[DOCX File, 35 KB - formative_v9i1e59533_app1.docx]

Checklist 1
CHERRIES (Checklist for Reporting Results of Internet E-Surveys).
[DOCX File, 32 KB - formative_v9i1e59533_app2.docx]

References

1. Global strategy for the diagnosis management, and prevention of chronic obstructive pulmonary disease. Global Initiative for Chronic Obstructive Lung Disease. 2023. URL: https://goldcopd.org/2023-gold-report-2/ [accessed 2025-02-04]
2. Celli B, Fabbri L, Criner G, et al. Definition and nomenclature of chronic obstructive pulmonary disease: time for its revision. Am J Respir Crit Care Med 2022 Dec 1;206(11):1317-1325. [doi: 10.1164/rccm.202204-0671PP] [Medline: 35914087]
3. COPD [Web page in Dutch]. VZinfo. 2022. URL: https://www.vzinfo.nl/copd [accessed 2023-01-09]
4. COPD health care costs [Web page in Dutch]. VZinfo. 2019. URL: https://www.vzinfo.nl/copd/zorgtaven#:~:text=Zorgtaven%20voor%20COPD%20753%20miljoen%20euro%20in%20201&text=Dit%20maakt%20%2C8%25%20uitgaven%20voor%20alle%20luchtwegaandoeningen [accessed 2024-03-27]
5. Vreeken HL, Beekman E, Doornmaal M, Post MHL, Meerhoff GA, Spruit MA. COPD [Web page in Dutch]. KNGF. 2020. URL: https://www.kennisplatformfysiotherapie.nl/richtlijnen/copd/a-algemene-informatie/ [accessed 2025-02-04]
6. Ummels D, Beekman E, Braun SM, Beurskens AJ. Using an activity tracker in healthcare: experiences of healthcare professionals and patients. Int J Environ Res Public Health 2021 May 12;18(10):34066296. [doi: 10.3390/ijerph18105147] [Medline: 34066296]
7. Qiu S, Cai X, Wang X, et al. Using step counters to promote physical activity and exercise capacity in patients with chronic obstructive pulmonary disease: a meta-analysis. Ther Adv Respir Dis 2018;12:1753466618787386. [doi: 10.1177/1753466618787386] [Medline: 29993339]
8. Chiauzzi E, Rodarte C, DasMahapatra P. Patient-centered activity monitoring in the self-management of chronic health conditions. BMC Med 2015 Apr 9;13:77. [doi: 10.1186/s12916-015-0319-2] [Medline: 25889598]
9. Karapanos E, Gouveia R, Hassenzahl M, Forlizzi J. Wellbeing in the making: peoples' experiences with wearable activity trackers. Psychol Well Being 2016;6:4. [doi: 10.1186/s13612-016-0042-6] [Medline: 27376017]
10. Richard AA, Shea K. Delineation of self-care and associated concepts. J Nurs Scholarsh 2011 Sep;43(3):255-264. [doi: 10.1111/j.1547-5069.2011.01404.x] [Medline: 21884371]
11. Wouters M, Huygens M, Voogdt H, et al. E-health monitor: samen aan zet [Web page in Dutch]. Nivel. 2019. URL: https://www.nivel.nl/nl/publicatie/samen-aan-zet-ehealth-monitor-2019 [accessed 2025-02-04]
12. Taylor J, Coates E, Wessels B, Mountain G, Hawley MS. Implementing solutions to improve and expand telehealth adoption: participatory action research in four community healthcare settings. BMC Health Serv Res 2015 Dec 1;15:529. [doi: 10.1186/s12913-015-1195-3] [Medline: 26626564]
13. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and facilitators to the implementation of eHealth services: systematic literature analysis. J Med Internet Res 2019 Nov 22;21(11):e14197. [doi: 10.2196/14197] [Medline: 31755869]
14. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146. [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
15. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. J Telemed Telecare 2018 Jan;24(1):4-12. [doi: 10.1177/1357633X16674087]

XSL•FO
RenderX

16. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: systematic review of the literature. J Med Internet Res 2018 May 1;20(5):e10235. [doi: 10.2196/10235] [Medline: 29716883]
17. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implementation Sci 2009 Dec;4(1):50. [doi: 10.1186/1748-5908-4-50]
18. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
19. Chronish ZorgNet [Website in Dutch]. URL: https://www.chronischzorgnet.nl/ [accessed 2025-02-04]
20. Beurskens A, Swinkels R, Wittink H, Peppen R, Vermeulen H. Meten in de Praktijk: Bohn Stafleu van Loghum; 2020. [doi: 10.1007/978-90-368-2459-0]
21. Stevens JGA. Ready for Goal Setting?: Maastricht University; 2017.
22. Ummels D, Beekman E, Moser A, Braun SM, Beurskens AJ. Patients' experiences with commercially available activity trackers embedded in physiotherapy treatment: a qualitative study. Disabil Rehabil 2020 Nov;42(23):3284-3292. [doi: 10.1080/09638288.2019.1590470] [Medline: 30973026]
23. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
24. Wilde LJ, Sewell L, Percy C, Ward G, Clark C. What are the experiences of people with COPD using activity monitors?: A qualitative scoping review. COPD 2022;19(1):88-98. [doi: 10.1080/15412555.2022.2033192] [Medline: 35132933]
25. Rausch AK, Baur H, Reicherzer L, et al. Physiotherapists' use and perceptions of digital remote physiotherapy during COVID-19 lockdown in Switzerland: an online cross-sectional survey. Arch Physiother 2021 Jul 7;11(1):18. [doi: 10.1186/s40945-021-00112-3] [Medline: 34233763]
26. Miller MJ, Pak SS, Keller DR, Gustavson AM, Barnes DE. Physical therapist telehealth delivery at 1 year into COVID-19. Phys Ther 2022 Nov 6;102(11):36124699. [doi: 10.1093/ptj/pzac121] [Medline: 36124699]
27. Ummels D, Beekman E, Theunissen K, Braun S, Beurskens AJ. Counting steps in activities of daily living in people with a chronic disease using nine commercially available fitness trackers: cross-sectional validity study. JMIR Mhealth Uhealth 2018 Apr 2;6(4):e70. [doi: 10.2196/mhealth.8524] [Medline: 29610110]
28. Hancock C. The benefits of regular walking for health, well‐being and the environment. : C3 Collaborating for Health; 2012 URL: https://image.guardian.co.uk/sys-files/Guardian/documents/2012/11/02/C3-report-on-walking.pdf [accessed 2025-02-05]
29. Lee IM, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of step volume and intensity with all-cause mortality in older women. JAMA Intern Med 2019 Aug 1;179(8):1105-1112. [doi: 10.1001/jamainternmed.2019.0899] [Medline: 31141585]
30. Tudor-Locke C, Bassett DR. How many steps/day are enough? Sports Med 2004;34(1):1-8. [doi: 10.2165/00007256-200434010-00001]
31. Tudor-Locke CE, Myers AM. Methodological considerations for researchers and practitioners using pedometers to measure physical (ambulatory) activity. Res Q Exerc Sport 2001 Mar;72(1):1-12. [doi: 10.1080/02701367.2001.10608926] [Medline: 11253314]
32. Depew ZS, Novotny PJ, Benzo RP. How many steps are enough to avoid severe physical inactivity in patients with chronic obstructive pulmonary disease? Respirology 2012 Aug;17(6):1026-1027. [doi: 10.1111/j.1440-1843.2012.02207.x] [Medline: 22672739]
33. Tudor-Locke C, Craig CL, Thyfault JP, Spence JC. A step-defined sedentary lifestyle index: <5000 steps/day. Appl Physiol Nutr Metab 2013 Feb;38(2):100-114. [doi: 10.1139/apnm-2012-0235] [Medline: 23438219]
34. Varsi C, Solberg Nes L, Kristjansdottir OB, et al. Implementation strategies to enhance the implementation of eHealth programs for patients with chronic illnesses: realist systematic review. J Med Internet Res 2019 Sep 27;21(9):e14255. [doi: 10.2196/14255] [Medline: 31573934]
35. Wilde LJ, Percy C, Clark C, Ward G, Wark PA, Sewell L. Views and experiences of healthcare practitioners supporting people with COPD who have used activity monitors: "More than just steps". Respir Med 2023 Nov;218:107395. [doi: 10.1016/j.rmed.2023.107395] [Medline: 37633422]
36. Albers B, Metz A, Burke K. Implementation support practitioners—a proposal for consolidating a diverse evidence base. BMC Health Serv Res 2020 May 1;20(1):368. [doi: 10.1186/s12913-020-05145-1] [Medline: 32357877]
37. Burke K, Albers B, Ward C. Implementation support practitioner profile guiding principles and core competencies for implementation practice. : University of North Carolina at Chapel Hill; 2020.
38. Components of evidence-based practice. American Physical Therapy Association. 2020 Mar 23. URL: https://www.apta.org/patient-care/evidence-based-practice-resources/components-of-evidence-based-practice#:~:text=Evidence-based%20practice%20includes%20the,three%20elements%20are%20equally%20important [accessed 2024-10-16]
39. Barakat A, Woolrych RD, Sixsmith A, Kearns WD, Kort HS. eHealth technology competencies for health professionals working in home care to support older adults to age in place: outcomes of a two-day collaborative workshop. Med 20 2013;2(2):e10. [doi: 10.2196/med20.2711]

XSL•FO
RenderX

40. Wentink MM, Siemonsma PC, van Bodegom-Vos L, et al. Teachers' and students' perceptions on barriers and facilitators for eHealth education in the curriculum of functional exercise and physical therapy: a focus groups study. BMC Med Educ 2019 Sep 6;19(1):343. [doi: 10.1186/s12909-019-1778-5] [Medline: 31492129]
41. Meten met ehealth [Website in Dutch]. 2024. URL: https://metenmetehealth.nl [accessed 2025-02-05]
42. Beweegtech [Website in Dutch]. 2024. URL: https://beweegtech.nl/ [accessed 2025-02-05]
43. Ariens LF, Schussler-Raymakers FM, Frima C, et al. Barriers and facilitators to eHealth use in daily practice: perspectives of patients and professionals in dermatology. J Med Internet Res 2017 Sep 5;19(9):e300. [doi: 10.2196/jmir.7512] [Medline: 28874336]
44. Spruit MA, Van't Hul A, Vreeken HL, et al. Profiling of patients with COPD for adequate referral to exercise-based care: the Dutch model. Sports Med 2020 Aug;50(8):1421-1429. [doi: 10.1007/s40279-020-01286-9] [Medline: 32333261]

Abbreviations

AT: activity tracker
CFIR: Consolidated Framework for Implementation Research
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
COPD: chronic obstructive pulmonary disease
KNGF: Koninklijk Nederlands Genootschap voor Fysiotherapie
TPB: theory of planned behavior

Edited by A Mavragani; submitted 15.04.24; peer-reviewed by MV Ittersum, Y Chu; revised version received 19.12.24; accepted 19.12.24; published 12.02.25.

Please cite as:
Ummels D, Bols E, Frantzen RJA, Frantzen T, Robeerts L, Beekman E
Activity Trackers in Physical Therapy for People With Chronic Obstructive Pulmonary Disease in the Netherlands: Cross-Sectional Study on Current Use and Implementation Determinants
JMIR Form Res 2025;9:e59533
URL: https://formative.jmir.org/2025/1/e59533
doi:10.2196/59533

© Darcy Ummels, Esther Bols, Roel Johannes Anna Frantzen, Tim Frantzen, Levy Robeerts, Emmylou Beekman. Originally published in JMIR Formative Research (https://formative.jmir.org), 12.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Understanding Providers' Attitude Toward AI in India's Informal Health Care Sector: Survey Study

Sumeet Kumar, PhD; Snehil Rayal, BTech; Raghuram Bommaraju, PhD; Navya Pratyusha Varasala, MA; Sirisha Papineni, MD; Sarang Deo, PhD

Indian School of Business, Gachibowli, ISB Road, Hyderabad, India

Corresponding Author:
Sumeet Kumar, PhD
Indian School of Business, Gachibowli, ISB Road, Hyderabad, India

Abstract

Background: Tuberculosis (TB) is a major global health concern, causing 1.5 million deaths in 2020. Diagnostic tests for TB are often inaccurate, expensive, and inaccessible, making chest x-rays augmented with artificial intelligence (AI) a promising solution. However, whether providers are willing to adopt AI is not apparent.

Objective: The study seeks to understand the attitude of Ayurveda, Yoga and Naturopathy, Unani, Siddha, and Homoeopathy (AYUSH) and informal health care providers, who we jointly call AIPs, toward adopting AI for TB diagnosis. We chose to study these providers as they are the first point of contact for a majority of TB patients in India.

Methods: We conducted a cross-sectional survey of 406 AIPs across the states of Jharkhand (162 participants) and Gujarat (244 participants) in India. We designed the survey questionnaire to assess the AIPs' confidence in treating presumptive TB patients, their trust in local radiologists' reading of the chest x-ray images, their beliefs regarding the diagnostic capabilities of AI, and their willingness to adopt AI for TB diagnosis.

Results: We found that 93.7% (270/288) of AIPs believed that AI could improve the accuracy of TB diagnosis, and for those who believed in AI, 71.9% (194/270) were willing to try AI. Among all AIPs, 69.4% (200/288) were willing to try AI. However, we found significant differences in AIPs' willingness to try AI across the 2 states. Specifically, in Gujarat, a state with better and more accessible health care infrastructure, 73.4% (155/211) were willing to try AI, and in Jharkhand, 58.4% (45/77) were willing to try AI. Moreover, AIPs in Gujarat who showed higher trust in the local radiologists were less likely to try AI (odds ratio [OR] 0.15, 95% CI 0.03 - 0.69; *P*=.02). In contrast, in Jharkhand, those who showed higher trust in the local radiologists were more likely to try AI (OR 2.11, 95% CI 0.9 - 4.93; *P*=.09).

Conclusions: While most AIPs believed in the potential benefits of AI-based TB diagnoses, many did not intend to try AI, indicating that the expected benefits of AI measured in terms of technological superiority may not directly translate to impact on the ground. Improving beliefs among AIPs with poor access to radiology services or those who are less confident of diagnosing TB is likely to result in a greater impact of AI on the ground. Additionally, tailored interventions addressing regional and infrastructural differences may facilitate AI adoption in India's informal health care sector.

(JMIR Form Res 2025;9:e54156) doi:10.2196/54156

KEYWORDS

artificial intelligence; tuberculosis; health care providers; cross-sectional studies; trust; x-rays; India

Introduction

Tuberculosis (TB) remains a significant global health challenge, with over 80% of reported cases and deaths originating from low- and middle-income countries (LMICs) worldwide [1]. Among these countries, India shoulders a substantial burden, accounting for a quarter of all TB cases and resulting in approximately 89,000 deaths in the year 2019 alone [2]. The COVID-19 pandemic further worsened these global inequalities, particularly by disrupting TB diagnostic and treatment services [3,4]. Data from the World Health Organization (WHO) reveals a concerning trend in the incidence rate of TB. After experiencing a decline of around 2% per year over the past 2 decades, there has been a 3.6% increase in new TB cases per 100,000 population between 2020 and 2021, as indicated in the Global Tuberculosis Report [5]. Additionally, due to factors like underreporting, underdiagnosis, and limited access to health care for TB patients, a significant gap exists between the estimated and reported number of cases [6]. In India, 29% of TB cases were either undiagnosed or missed in 2021 [7,8]. This resurgence of TB cases [9] is particularly alarming for LMICs like India, which bear a disproportionate disease burden [10] and face numerous challenges [6] in their fragile health care supply chains.

XSL•FO
RenderX

To lessen this burden, recent advancements in diagnosis based on molecular testing and chest x-ray (CXR) scans offer promising avenues. However, the highly accurate molecular diagnostic tests for TB, such as Xpert MTB/RIF (*Mycobacterium tuberculosis*/rifampin), remain expensive, difficult to access, and challenging to maintain in LMICs [11]. Additionally, their roll-out has been slow [12]. Although WHO recommends molecular diagnostics as the preferred first-line testing method, only 38% of individuals diagnosed with TB in 2021 were tested with a WHO-recommended rapid molecular diagnostic at the initial diagnosis [12]. The other option of clinical diagnosis based on CXRs continues to be the mainstay among health care providers in India [13]. CXR is recommended by WHO for screening, triaging, and assisting in the diagnosis of TB. It is also featured in the national TB diagnostic guidelines of 19 out of 22 countries with the highest TB burden [14]. However, there is a shortage of adequately trained radiologists proficient in analyzing CXRs and producing high-quality reports [15]. This shortage has important implications, as the majority of TB care providers are not formally trained to interpret CXRs [16].

In the absence of qualified radiologists, the presence of automated artificial intelligence (AI) systems for interpreting CXRs could prove to be highly advantageous [17,18]. Employing AI-assisted interpretation of CXRs can improve the speed [19] and enhance the accuracy of TB diagnosis [20]. Addressing the delays in TB diagnosis through earlier TB detection is a key element of the WHO's TB control and elimination strategy [21]. Moreover, the more recent deep learning–based AI systems for CXR interpretation have demonstrated on par or higher specificity than radiologists [22-24], and their usage can significantly reduce the cost of TB screening [25]. These systems are also portable and can additionally be accessed remotely via the internet, making them a good choice for providers who often serve in remote locations.

However, the successful adoption of AI for the accurate diagnosis of TB depends on many factors [26]. In particular, adoption by the private sector is needed, as it plays a significant role in treating TB cases in India, diagnosing twice as many cases as compared to the public sector, amounting to an estimated 2.2 million TB cases [27]. The private sector is fragmented, consisting of a variety of health care providers, clinics, and practitioners that lack integration, especially in small towns and villages [28]. A study conducted [29] revealed that the average Indian village has 3.2 primary health care providers, with 68% lacking formal medical training and referred to as informal health care providers (IPs). In addition, India has Ayurveda, Yoga and Naturopathy, Unani, Siddha, and Homoeopathy (AYUSH) providers formally trained in one of the 6 traditional Indian systems of medicine, constituting 22.8% of formally trained medical practitioners in India [30]. From this point forward, we will use the abbreviation "AIPs" to collectively denote both informal health care providers and AYUSH practitioners. For faster and more accurate TB diagnosis, AIPs need to adopt AI systems. However, the adoption will only happen if AIPs perceive value in AI, which has not been examined in prior research.

In this paper, we conducted a survey study to understand the attitudes and perspectives of AIPs toward adopting AI for TB diagnosis. AIPs are the first point of contact in patient pathways for the majority of TB cases in India, so their willingness to incorporate AI-enhanced CXR systems into the decision-making is a key factor in driving AI adoption. Adopting AI could enhance AIPs' capability to interpret CXR scans and detect TB cases earlier and more accurately. In order to assess AIPs' beliefs in the diagnostic capabilities of AI and their willingness to adopt AI for TB diagnosis, we conducted a cross-sectional survey of 406 AIPs across the states of Jharkhand (162 participants) and Gujarat (244 participants).

Methods

Overview

We conducted a cross-sectional survey of AIPs in 2 Indian states (Gujarat and Jharkhand) to understand their current practices and potential barriers in adopting AI in their practice. The guidelines in the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) were followed (Checklist 1).

Jharkhand is one of the poorer states in India with lower technology penetration and lower access to radiologic facilities as well as radiologists, and overall a less developed private sector. In contrast, Gujarat is a more developed state with an active and accessible private health care infrastructure.

Participants Recruitment and Enrollment

We collaborated with World Health Partner (WHP), a non-profit organization that has partnered with x-ray labs and AIPs in Gujarat and Jharkhand. In Gujarat, WHP worked with 73 x-ray labs and 757 AIPs, while in Jharkhand, they collaborated with 113 x-ray labs and 406 AIPs.

Multiple methods were used to identify the AIPs. Field investigators from WHP mapped the AIPs in the districts of Ranchi and East Singhbhum in Jharkhand, as well as Surat and Gandhinagar in Gujarat. Investigators queried TB patient cohorts (those in their first year of treatment) about the health care providers they visited during their treatment. Allopathic providers were interviewed to identify nearby AIPs who frequently referred cases to them. Additionally, government officials and partner organizations at the community level helped identify local health care provider networks. Radiology units offering CXR services were also consulted to locate AIPs in their vicinity. The investigators used the mobile Commcare application to map these providers, capturing details such as contact information, qualifications, outpatient department load, referral linkages, and more, which were then used to generate a unique ID for each provider (form included in Multimedia Appendix 1). Each AIP clinic functioned as an independent study or clinical setting.

For our study, we asked all identified AIPs working with WHP for their participation and enrolled 406 AIPs based on their willingness to participate in the survey. In return, we provided a mobile recharge of 300 (US $3.47) for completing the survey. These participants were from 18 blocks across 2 districts in Gujarat and 22 blocks spanning 2 districts in Jharkhand. The survey was conducted between February 2022 and September 2022.

XSL•FO
RenderX

Data

We developed the initial questionnaire by leveraging questions from established scales and modifying them to suit our study's context. This process included adapting sections on AI in TB diagnosis from existing literature [31,32] and customizing questions to assess AIPs' comfort with technology and demographics [33]. Additionally, we adapted from the literature [34] the questions regarding AIPs' trust in radiologists, ensuring they were pertinent to our specific research objectives.

The questionnaire underwent 7 iterative revisions in collaboration with our field partner, WHP, known for its extensive experience with TB providers and patients. Subsequently, Outline India, the data collection agency, conducted pretests that led to further refinements based on feedback from pilot studies. This phase involved translating the questionnaire into Hindi and Gujarati and making 8 additional revisions to improve clarity and relevance. Our research team oversaw the translation process to ensure its accuracy. The final, refined version of the questionnaire, which incorporated comprehensive feedback and amendments, received validation by the strategy lead at WHP.

We structured the questionnaire into multiple subsections to evaluate the following aspects: AIPs' understanding of TB diagnosis; confidence in diagnosing TB patients; trust levels in local radiologists; beliefs and intentions regarding the adoption of AI in TB diagnosis; comfort levels with technology; and demographic details such as age, gender, state of residence, and years of experience. We conducted a survey of the registered AIPs by partnering with Outline India. This agency deployed field investigators proficient in the local languages (Hindi and Gujarati) and used specialized software for digital data collection.

We gathered data from 406 participating AIPs. The survey was conducted at the participants' locations and lasted 10-20 minutes. Out of 406 AIPs, 39.9% (162/406) participants were from Jharkhand and 60.1% (244/406) participants were from Gujarat. In total, 85.2% (346/406) of the respondents were men, while women comprised 14.8% (60/406). The age of respondents ranged from 22 to 86 (mean 41.8, SD 11.4) years.

Responses to most of the questions were on a 5-point Likert scale as shown in Table 1. To illustrate the use of this scale, a score of "4" indicates that the participant somewhat agrees with the statement under consideration, whereas a score of "5" indicates strong agreement with the statement. For a few questions, respondents' opinions were recorded as 1=Yes and 0=No using a dichotomous scale. The detailed questionnaire can be referenced in Multimedia Appendix 2.

Table . Likert Scale of 0 - 5 capturing responses from "strongly disagree" to "strongly agree".

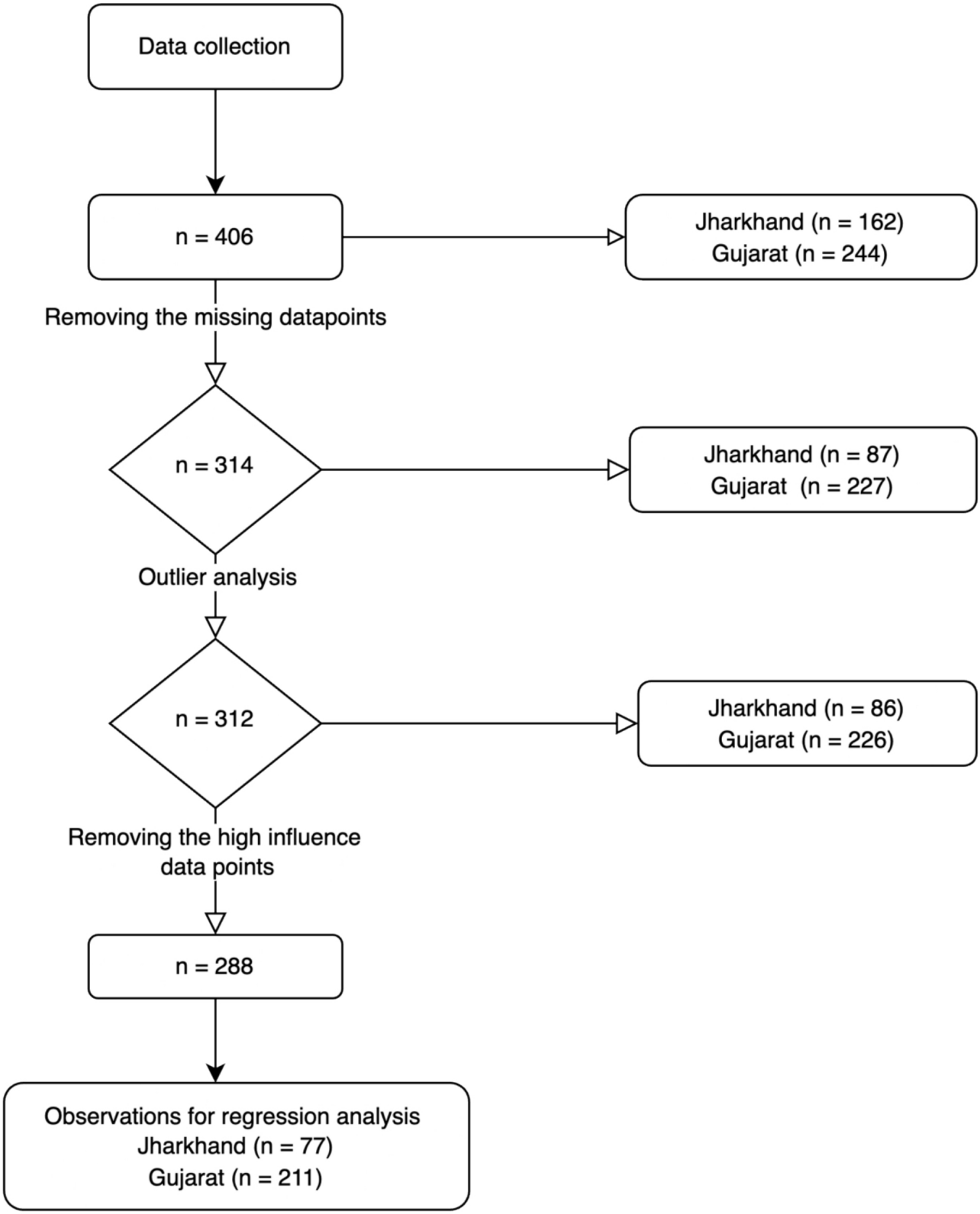
Scale	Respondent's opinion to question
5	Strongly agree
4	Somewhat agree
3	Neither agree nor disagree
2	Somewhat disagree
1	Strongly disagree

We encountered certain absent (not available) data points, stemming from incomplete surveys. These data points were eliminated before employing the data for regression, resulting in the removal of 92 observations. Out of these, 17 were from Gujarat, and 75 were from Jharkhand. The main reason for the reduction in the number of observations was the lack of responses for questions on "Trust in local radiologists" because many survey participants did not have access to radiologists.

In addition, the outliers were identified and removed using boxplot and Cook distance analysis to eliminate any highly influential data points that could distort the results. This process resulted in a final data set consisting of 288 observations, with 211 observations from Gujarat and 77 observations from Jharkhand. We show the entire workflow for data cleaning in Figure 1. Multimedia Appendix 3 contains further information on outlier analysis and on descriptor variables for both the initial sample collected and the final sample used in the regression analysis.

XSL•FO
RenderX

Figure 1. Data cleaning workflow reflecting multiple steps to arrive at the final dataset.



Variable Description and Analysis

In this study, we investigated two main constructs: (1) AI Intent and (2) AI Belief. Previous research on technology adoption has demonstrated that the actual utilization of a technology is preceded by factors such as "Intent to use," "Attitude toward technology," and considerations of "Ease of use" among others [35,36]. The ease of use of a product often evolves with the incorporation of new features and regular enhancements to the user experience, a common occurrence in the realm of internet technology products. Hence, we specifically focused on understanding individuals' "Intent to use" AI products, which we refer to as "AI Intent"; and their "Attitude toward AI," which we refer to as "AI Belief."

XSL•FO
RenderX

We coded the response to the survey question, "Are you willing to try an AI software that could generate an x-ray report from an uploaded x-ray?" as a categorical variable "AI Intent" with 2 options: "Yes" or "No." We used logistic regression to examine its association with the independent variables.

We created the "AI Belief" variable by combining the responses to 2 questions: "I believe AI algorithms can reliably detect imaging findings that suggest TB" and "I believe the use of AI could improve the accuracy of diagnosis of tuberculosis." We calculated the mean of the responses to these questions to construct the "AI Belief" variable. It represents participants' overall belief in the effectiveness of AI in improving TB diagnosis.

For independent variables in the regression analysis, we included "Confidence in diagnosing TB" and "Trust in local radiologists," and demographic variables such as age, gender, and state of residence.

We coded "Confidence in diagnosing TB" as the mean of the responses to 5 questions from the survey questionnaire: "I feel confident in my ability to read an x-rays report to detect a TB case," "I feel confident in my ability to interpret an x-rays film without an x-rays report to detect a TB case," "I feel confident in my ability to deal with TB patients (if allowed to treat or medicate)," "I feel confident in my ability to use relevant diagnostic tests to detect a TB case," and "I feel confident in my ability to collect relevant health-related information from patients."

Similarly, we coded "Trust in local radiologists" as the mean of the responses to 3 questions: "The language and style of radiology reports are mostly clear," "The radiologist who reads the x-rays and writes the report is competent," and "I am generally satisfied with x-rays reports I receive from radiologists." We calculated the Cronbach α using all the survey responses to assess the consistency of the scale. Gender and state of residence were treated as categorical.

We used Python (version 3.9.13) for data cleaning and data visualization, while data analysis was carried out using StataCorpSTATA SE (version 17.0) and Microsoft Excel.

Ethical Considerations

The study protocols underwent review and approval by the Institutional Review Board at the Indian School of Business (application number ISB-IRB 2022 - 04) and were approved on January 28, 2022. Informed consent was obtained from the participants before recording their responses. Information provided to participants included the purpose of the study, what participants will be asked to do if they participate, how much of the subjects' time participation may take, and what type of information they might be asked to provide. It was also made clear that participation is voluntary, and subjects may withdraw from the study at any time. The data were deidentified to protect participants' privacy. The consent form can be referenced in Multimedia Appendix 4.

Results

Descriptive Statistics

We found that over 90% of AIPs believed in the benefits of AI, with 93.8% (270/288) responding positively to "AI Improves Accuracy of TB Diagnosis," and 90.9% (262/288) responding positively to "AI Can Detect Image Findings." We also found that 69.4% (200/288) of AIPs were willing to try AI. For those who believed in AI, 71.9% (194/270) were willing to try AI.

Of the final 288 observations used for regression analysis, 73.3% (211/288) were from Gujarat, and 27.7% (77/288) were from Jharkhand. In the sample, 11.1% (32/288) were women, and 88.9% (256/288) were men with ages between 22 and 68 (mean 40.7, SD 9.7) years. See Table 2 for the summary statistics of the data.

Table . Summary statistics for each variable in regression analysis, for Jharkhand and Gujarat, February-September 2022.

	Jharkhand (n=77)[a]				Gujarat (n=211)[b]			
Variable	Mean	SD	Min	Max	Mean	SD	Min	Max
Age (years)	40.87	10.92	24	67	40.60	9.47	22	68
AI[c] Intent	0.58	0.50	0	1	0.73	0.44	0	1
AI Belief	4.39	0.71	2.5	5	4.59	0.55	3	5
Confidence in diagnosing TB[d]	4.41	0.46	3	5	4.50	0.50	2.4	5
Trust in local radiologists	4.56	0.62	1.33	5	4.85	0.34	3.33	5

[a]Of the 77 respondents, 61 (79.2%) were male and 16 (20.8) were female.

[b]Of the 211 respondents, 195 (92.4%) were male and 16 (7.6%) were female.

[c]AI: artificial intelligence.

[d]TB: tuberculosis.

The mean score for "Confidence in diagnosing TB" was similar in both states, 4.50 (SD 0.50) for Gujarat and 4.41 (SD 0.46)

XSL•FO
RenderX

for Jharkhand. The mean score for "Trust in local radiologists" was higher among AIPs in Gujarat (4.85, SD 0.34) than in Jharkhand (4.56, SD 0.62).

The Cronbach α value for the questions measuring "Confidence in diagnosing TB" was 0.65, while it was 0.82 for "Trust in the local radiologist" and 0.79 for "AI Belief."

AI Intent

The results of the logistic regression analysis of "AI Intent" on predictor variables are shown in Table 3. We found that "Confidence in diagnosing TB" was positively associated for both states but statistically significant only for Gujarat (odds ratio [OR] 1.943, 95% CI 1.038-3.637; *P*=.04) and not for Jharkhand (OR 2.186, 95% CI 0.695-6.875; *P*=.18).

Table . Regression results for dependent variable "AI Intent", for Jharkhand and Gujarat, February-September 2022.

	Jharkhand (n=77)				Gujarat (n=211)			
Predictor variables	Odds ratio	SE	*P* value	95% CI	Odds ratio	SE	*P* value	95% CI
Age (years)	0.964	0.025	.15	0.916-1.014	0.990	0.016	.52	0.959-1.021
Gender	2.573	1.874	.20	0.617-10.728	2.073	1.252	.23	0.635-6.771
Confidence in diagnosing TB[a]	2.186	1.278	.18	0.695-6.875	1.943	0.622	.04	1.038-3.637
Trust in local radiologists	2.107	0.914	.09	0.900-4.932	0.149	0.117	.02	0.032-0.691

[a]TB: tuberculosis.

"Trust in local radiologists" was positively associated with "AI Intent" for AIPs in Jharkhand (OR 2.107, 95% CI 0.9-4.932; *P*=.09) but negatively associated for AIPs in Gujarat (OR 0.149, 95% CI 0.032 - 0.691; *P*=.02).

AI Belief

The results of the multivariate linear regression (ordinary least squares) analysis are shown in Table 4. First, we observed that "Confidence in diagnosing TB" was positively associated with "AI Belief" in both Jharkhand (β=0.72, *P*<.001) and Gujarat (β=0.177, *P*=.008). Further, "Trust in local radiologists" was also positively associated with "AI Belief" in both Jharkhand (β=0.215, *P*=.02) and Gujarat (β=0.66, *P*<.001).

Table . Regression results for dependent variable "AI Belief", for Jharkhand and Gujarat, February-September 2022.

	Jharkhand (n=77)				Gujarat (n=211)			
Predictor variables	β	SE	*P* value	95% CI	β	SE	*P* value	95% CI
Age (years)	−0.007	0.006	.24	−0.020 to 0.005	0.002	0.004	.64	−0.005 to 0.009
Gender	0.166	0.196	.40	−0.225 to 0.557	0.021	0.105	.84	−0.186 to 0.227
Confidence in diagnosing TB[a]	0.720	0.137	<.001	0.447 to 0.993	0.177	0.066	.008	0.047 to 0.308
Trust in local radiologists	0.215	0.089	.02	0.037 to 0.392	0.660	0.105	<.001	0.453 to 0.866

[a]TB: tuberculosis.

We found that none of the demographic factors were significantly associated with "AI Belief."

Discussion

Principal Findings

The integration of AI with CXRs holds promise as a cost-effective and accurate solution for TB diagnosis. However, despite this potential, the adoption of AI tools among health care providers for TB diagnosis remains limited. In this study, we conducted a cross-sectional survey of AIPs in 2 Indian states (Gujarat and Jharkhand) to understand and assess their belief in AI; their willingness to adopt AI for improved diagnosis; and the association of these with the AIPs' confidence in diagnosing TB as well as trust in the local radiologists. We found that AIPs are mostly aware of AI and believe in AI, but the intent to adopt AI is lower and depends on AIPs' confidence in diagnosing TB and their trust in local radiologists. Previous studies that examined the perceptions of health care professionals about AI in the medical field have yielded mixed results [37,38]. For instance, Oh et al [37] found that physicians in Korea had a favorable attitude toward AI, considering it as a complementary

XSL•FO
RenderX

tool rather than a replacement for their roles. In contrast, Abdullah et al [38] discovered that health care employees in Saudi Arabia needed to be made aware of the benefits of AI and expressed concerns about being replaced by AI in their jobs.

While previous studies primarily focused on formal health care professionals [37-39], students [40], and the general public [41], our study contributes to understanding perceptions of AI among other medical professionals, specifically AIPs in India. Our findings indicate that most AIPs (over 90%) believed in AI's capabilities; however, a significant proportion (88/288, 30.6%) were unwilling to try AI. These findings align with previous research among physicians [42], where the intention to use AI was observed only when there was a belief in the role of AI. However, unlike in prior research [42], where the participants were medical college students who received AI education as part of their curriculum, AIPs are unlikely to have similar formal educational opportunities for exposure to AI. Therefore, AIPs would require additional support, for example, in terms of training programs to know about AI and subsequently try AI. Additionally, AI systems should strive to be more explainable to make it easier to believe in AI, for example, by providing underlying reasoning that leads to their output, which is also known as explainable AI. An explainable AI approach would assist AIPs in developing trust in the system's output and bolster their belief in AI.

Our study has uncovered a noteworthy association between AIPs' confidence in diagnosing TB and their willingness to adopt AI, which is somewhat surprising when considering prior research has indicated that highly confident clinicians are less likely to change their decision [43]. However, AIPs differ from general clinicians in a crucial aspect, as they lack formal training in reading x-rays films. Therefore, we posit that the confidence of AIPs is likely influenced by the availability of high-quality diagnostic inputs, such as access to microbiological tests and well-prepared radiology reports. While the contextual factors may be at play in our findings, this suggests that the overall impact of AI adoption on health care outcomes might be more restrained than earlier studies have suggested. This distinction arises from the fact that those choosing to adopt AI may already have access to a high-quality diagnostic infrastructure, and thus, their adoption of AI might not significantly boost health outcome metrics like case detection rates. In light of this insight, it is advisable for companies and health care organizations to not rely exclusively on mean performance differences between AI systems and radiologists when estimating the potential enhancements in health outcomes.

Moreover, we identified a substantial difference in the intention to adopt AI between the states of Gujarat and Jharkhand. A higher willingness to try AI was observed among respondents from Gujarat compared to their counterparts in Jharkhand. The observed disparity between the 2 states could be attributed to Gujarat being a more developed state in India, with higher levels of education and better health care infrastructure. The presence of a well-established, technology-enabled infrastructure in Gujarat may have provided first-hand experience of the benefits of technology in health care, leading to a stronger intent to try technological products like AI. This implies that, in the context of AIPs, states with better technology-based infrastructure are the ones likely to have higher adoption when AI-based systems are implemented on the ground.

Another factor that might have affected willingness to try AI was the unavailability of qualified radiologists in less developed states like Jharkhand. One would expect AIPs to be more willing to try AI systems in the absence of qualified radiologists. However, our findings indicate that AIPs in Jharkhand were less likely to try AI. The reasons for this attitude are not immediately apparent, but experiencing lower-quality radiology reports may bias AIPs into believing that CXR technology is unreliable. To overcome such barriers to adoption, implementing agencies will need to provide additional evidence of the benefits of AI and related technologies like CXR scans.

By integrating AI into the primary health care centers, which are already actively engaged in TB service provision, one can extend AI-powered TB diagnostics to grassroots levels [15]. While AIPs acknowledge the potential of AI to enhance TB diagnosis, concerns regarding the financial impact of such technology are significant. AI's integration into TB diagnosis through CXR screenings entails an additional cost, estimated at a modest 100 (approximately $1.3) per screening [44]. This cost, while relatively small, would be borne by AIPs and possibly transferred to patients. Considering India's per capita net income was 98,374 (US $1138.71) in 2022 - 23 [45], this expense may be manageable for a large population segment. However, affordability may pose a barrier in economically disadvantaged regions, potentially necessitating government intervention to facilitate access. The urgency of integrating AI in diagnostics must be contextualized within India's acute radiologist shortfall, with the current ratio standing at approximately one radiologist per 100,000 people, starkly below international norms [15]. This shortage underscores the broader necessity of AI deployment in augmenting health care capacity.

Our study has unveiled several important findings that are instrumental in understanding the landscape of AI in India's TB-related health care services. First and foremost, a significant fraction of AIPs have a positive attitude toward AI, with many expressing a readiness to integrate AI into their practice. However, we have identified a discernible gap between those who acknowledge AI's benefits and those who are prepared to embrace the technology. This discrepancy is further nuanced by a variation in attitudes based on providers' self-assessed diagnostic competencies, providers with greater confidence in their abilities tend to be more receptive to AI. Additionally, there is noticeable heterogeneity in the willingness to adopt AI across different geographic regions, which points to the influence of local health care infrastructure and socioeconomic factors on technology acceptance.

The implications of these findings are multifaceted and suggest several actionable strategies for enhancing AI adoption in health care settings. To translate the positive attitudes of AIPs into actual use of AI, tailored initiatives that address barriers to implementation are necessary. This might include providing more robust evidence of AI's effectiveness, integrating AI education into continuous professional development programs, and offering financial incentives. As diagnostic pathways in TB are majorly driven by the private sector [27], steps should be

XSL•FO
RenderX

taken to make the integration of AI in the TB care cascade profitable for providers to ensure its sustainability. Moreover, there is a compelling need to target AIPs with suboptimal diagnostic capabilities, as enhancements in this group's diagnostic services could positively impact overall health care outcomes.

To ensure these strategies are effective across India's diverse health care landscape, future research should replicate this study in multiple geographies. Such research should also examine the intricacies of trust and financial relationships between AIPs and radiologists to uncover deeper insights into the factors that influence AI adoption and to devise informed interventions that can bolster the technology's uptake.

Limitations

Like other survey-based studies, this study too has a few limitations. First, we had to remove many observations for the state of Jharkhand because many AIPs in Jharkhand did not respond to questions on "Trust in local radiologists." The lower response rate to these questions was because of AIPs' limited access to radiologists. To ascertain if the removal of such AIPs from the data impacted the findings, we compared the demographic variables of the dataset before and after the removal and did not find any statistically significant difference (more information in Multimedia Appendix 5).

Second, the responses in the questionnaire that measured opinions about AI may have been subjected to social desirability bias among AIPs. To mitigate this effect, we informed the participants about the anonymity of the survey and the separation of personal information from their responses.

Third, the AIPs' perception of AI in TB diagnosis may have been influenced by their overall opinion of AI and the extent of their knowledge of AI. To allow more informed responses, participants were apprised of the AI approach to recognize TB cases.

Conclusions

Our study provides insights into the attitudes of AYUSH and informal health care providers toward AI in TB diagnosis. It contributes to the understanding that adopting AI is not solely determined by a belief in AI but also relates to factors like confidence in diagnosing TB and other local factors like the quality of radiology services. Furthermore, regional variations between Gujarat and Jharkhand underscore the importance of tailored interventions to effectively integrate AI into diagnostic practices, with implications beyond TB diagnosis.

Acknowledgments

This research has been supported in part by the Max Institute of Healthcare Management (MIHM) at the Indian School of Business. Access to Ayurveda, Yoga and Naturopathy, Unani, Siddha, and Homoeopathy (AYUSH) practitioners and informal health care providers was facilitated by World Health Partners. We acknowledge a grant of 820,000 (US $9491.78) for the project titled "Effectiveness of AI for clinical diagnosis – Case of TB diagnosis in India." We also want to acknowledge the support of Dr Nisarg Desai for validating the survey questionnaire. We used the generative artificial intelligence tool ChatGPT by OpenAI in some parts of the paper for paraphrasing and checking grammar.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

SD, SP, SK, and RB designed the study. NPV and SP coordinated with the survey agency to conduct the study and collect the data. SR and SK conducted the analysis and prepared figures and tables. SK, SR, RB, and SD wrote the paper. All authors have reviewed the final manuscript and provided critical feedback.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Application form to map Ayurveda, Yoga and Naturopathy, Unani, Siddha, and Homoeopathy practitioners and informal health care providers.
[PDF File, 29 KB - formative_v9i1e54156_app1.pdf]

Multimedia Appendix 2
Survey Questionnaire.
[PDF File, 650 KB - formative_v9i1e54156_app2.pdf]

Multimedia Appendix 3
Outlier analysis.

XSL•FO
RenderX

[PDF File, 292 KB - formative_v9i1e54156_app3.pdf]

Multimedia Appendix 4
Institutional review board application and approval documents.
[PDF File, 2160 KB - formative_v9i1e54156_app4.pdf]

Multimedia Appendix 5
Descriptor variables.
[PDF File, 262 KB - formative_v9i1e54156_app5.pdf]

Checklist 1
STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist.
[PDF File, 117 KB - formative_v9i1e54156_app6.pdf]

References

1. Tuberculosis. WHO. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/tuberculosis [accessed 2025-01-27]
2. Varshney K, Patel H, Kamal S. Trends in tuberculosis mortality across India: improvements despite the COVID-19 pandemic. Cureus 2023 Apr;15(4):e38313. [doi: 10.7759/cureus.38313] [Medline: 37261163]
3. McQuaid CF, Vassall A, Cohen T, Fiekert K, Covid/Tb Modelling Working Group, White RG. The impact of COVID-19 on TB: a review of the data. Int J Tuberc Lung Dis 2021 Jun 1;25(6):436-446. [doi: 10.5588/ijtld.21.0148] [Medline: 34049605]
4. Migliori GB, Thong PM, Alffenaar JW, et al. Gauging the impact of the COVID-19 pandemic on tuberculosis services: a global study. Eur Respir J 2021 Nov;58(5):2101786. [doi: 10.1183/13993003.01786-2021] [Medline: 34446465]
5. Global tuberculosis report. WHO. 2022. URL: https://iris.who.int/bitstream/handle/10665/363752/9789240061729-eng.pdf?sequence=1 [accessed 2025-01-27]
6. Chakaya J, Khan M, Ntoumi F, et al. Global Tuberculosis Report 2020 - Reflections on the Global TB burden, treatment and prevention efforts. Int J Infect Dis 2021 Dec;113 Suppl 1(Suppl 1):S7-S12. [doi: 10.1016/j.ijid.2021.02.107] [Medline: 33716195]
7. Ricks S, Singh A, Sodhi R, Pal A, Arinaminpathy N. Operational priorities for engaging with India's private healthcare sector for the control of tuberculosis: a modelling study. BMJ Open 2024 Mar 19;14(3):e069304. [doi: 10.1136/bmjopen-2022-069304] [Medline: 38508628]
8. India TB report 2022, coming together to end TB altogether. : Ministry of Health and Family Welfare URL: https://tbcindia.mohfw.gov.in/wp-content/uploads/2023/05/TBAnnaulReport2022.pdf [accessed 2024-10-29]
9. Natarajan A, Beena PM, Devnikar AV, Mali S. A systemic review on tuberculosis. Indian J Tuberc 2020 Jul;67(3):295-311. [doi: 10.1016/j.ijtb.2020.02.005] [Medline: 32825856]
10. Suttels V, Guedes Da Costa S, Garcia E, et al. Barriers and facilitators to implementation of point-of-care lung ultrasonography in a tertiary centre in Benin: a qualitative study among general physicians and pneumologists. BMJ Open 2023 Jun 26;13(6):e070765. [doi: 10.1136/bmjopen-2022-070765] [Medline: 37369423]
11. Pai M, Dewan PK, Swaminathan S. Transforming tuberculosis diagnosis. Nat Microbiol 2023 May;8(5):756-759. [doi: 10.1038/s41564-023-01365-3] [Medline: 37127703]
12. Reid M, Agbassi YJP, Arinaminpathy N, et al. Scientific advances and the end of tuberculosis: a report from the Lancet Commission on Tuberculosis. Lancet 2023 Oct;402(10411):1473-1498. [doi: 10.1016/S0140-6736(23)01379-X]
13. Chest radiography in tuberculosis detection. WHO. 2016. URL: https://iris.who.int/bitstream/handle/10665/252424/9789241511506-eng.pdf [accessed 2025-01-27]
14. Pande T, Pai M, Khan FA, Denkinger CM. Use of chest radiography in the 22 highest tuberculosis burden countries. Eur Respir J 2015 Dec;46(6):1816-1819. [doi: 10.1183/13993003.01064-2015]
15. Thakur A. Shortage of radiologists a concern, fostering research in AI can augment capabilities of radiologists: experts. The Indian Express. 2023. URL: https://indianexpress.com/article/cities/bangalore/shortage-radiologists-research-ai-capabilities-experts-8884446/ [accessed 2025-01-27]
16. Guidance document on ntep_ayush collaborative activities to end TB. TBC India. URL: https://tbcindia.gov.in/showfile.php?lid=3631 [accessed 2023-09-19]
17. Hansun S, Argha A, Liaw ST, Celler BG, Marks GB. Machine and deep learning for tuberculosis detection on chest x-rays: systematic literature review. J Med Internet Res 2023 Jul 3;25:e43154. [doi: 10.2196/43154] [Medline: 37399055]
18. Kotei E, Thirunavukarasu R. A comprehensive review on advancement in deep learning techniques for automatic detection of tuberculosis from chest x-ray images. Arch Computat Methods Eng 2024 Jan;31(1):455-474. [doi: 10.1007/s11831-023-09987-w]
19. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology 2017 Aug;284(2):574-582. [doi: 10.1148/radiol.2017162326] [Medline: 28436741]

XSL•FO
RenderX

20. Rajpurkar P, O'Connell C, Schechter A, et al. CheXaid: deep learning assistance for physician diagnosis of tuberculosis using chest x-rays in patients with HIV. NPJ Digit Med 2020;3(1):115. [doi: 10.1038/s41746-020-00322-2] [Medline: 32964138]
21. Bhargava A, Bhargava M, Juneja A. Social determinants of tuberculosis: context, framework, and the way forward to ending TB in India. Expert Rev Respir Med 2021 Jul;15(7):867-883. [doi: 10.1080/17476348.2021.1832469] [Medline: 33016808]
22. Qin ZZ, Sander MS, Rai B, et al. Using artificial intelligence to read chest radiographs for tuberculosis detection: a multi-site evaluation of the diagnostic accuracy of three deep learning systems. Sci Rep 2019 Oct 18;9(1):15000. [doi: 10.1038/s41598-019-51503-3] [Medline: 31628424]
23. Codlin AJ, Dao TP, Vo LNQ, et al. Independent evaluation of 12 artificial intelligence solutions for the detection of tuberculosis. Sci Rep 2021 Dec 13;11(1):23895. [doi: 10.1038/s41598-021-03265-0] [Medline: 34903808]
24. Engle E, Gabrielian A, Long A, Hurt DE, Rosenthal A. Performance of Qure.ai automatic classifiers against a large annotated database of patients with diverse forms of tuberculosis. PLOS One 2020;15(1):e0224445. [doi: 10.1371/journal.pone.0224445] [Medline: 31978149]
25. Tzelios C, Nathavitharana RR. Can AI technologies close the diagnostic gap in tuberculosis? Lancet Dig Health 2021 Sep;3(9):e535-e536. [doi: 10.1016/S2589-7500(21)00142-4]
26. Kulkarni V, Gawali M, Kharat A. Key technology considerations in developing and deploying machine learning models in clinical radiology practice. JMIR Med Inform 2021 Sep 9;9(9):e28776. [doi: 10.2196/28776] [Medline: 34499049]
27. Arinaminpathy N, Batra D, Khaparde S, et al. The number of privately treated tuberculosis cases in India: an estimation from drug sales data. Lancet Infect Dis 2016 Nov;16(11):1255-1260. [doi: 10.1016/S1473-3099(16)30259-6]
28. Yellappa V, Lefèvre P, Battaglioli T, Devadasan N, Van der Stuyft P. Patients pathways to tuberculosis diagnosis and treatment in a fragmented health system: a qualitative study from a south Indian district. BMC Public Health 2017 Aug 4;17(1):635. [doi: 10.1186/s12889-017-4627-7] [Medline: 28778192]
29. Das J, Daniels B, Ashok M, Shim EY, Muralidharan K. Two Indias: the structure of primary health care markets in rural Indian villages with implications for policy. Soc Sci Med 2022 May;301:112799. [doi: 10.1016/j.socscimed.2020.112799] [Medline: 32553441]
30. The health workforce in india. WHO. 2016. URL: https://iris.who.int/bitstream/handle/10665/250369/9789241510523-eng.pdf?sequence=1 [accessed 2025-01-27]
31. Coppola F, Faggioni L, Regge D, et al. Artificial intelligence: radiologists' expectations and opinions gleaned from a nationwide online survey. Radiol Med 2021 Jan;126(1):63-71. [doi: 10.1007/s11547-020-01205-y]
32. Abuzaid MM, Elshami W, Tekin H, Issa B. Assessment of the willingness of radiologists and radiographers to accept the integration of artificial intelligence into radiology practice. Acad Radiol 2022 Jan;29(1):87-94. [doi: 10.1016/j.acra.2020.09.014]
33. Kelamane S, Satyanarayana S, Nagaraja SB, et al. Engaging informal private health care providers for TB case detection: experiences from RIPEND project in India. Tuberc Res Treat 2021;2021:9579167. [doi: 10.1155/2021/9579167] [Medline: 34239728]
34. Ignácio FDC, de Souza L, D'Ippolito G, Garcia MM. Radiology report: what is the opinion of the referring physician? Radiol Bras 2018;51(5):308-312. [doi: 10.1590/0100-3984.2017.0115] [Medline: 30369658]
35. Kemp A, Palmer E, Strelan P. A taxonomy of factors affecting attitudes towards educational technologies for use with technology acceptance models. Br J Educ Tech 2019 Sep;50(5):2394-2413. [doi: 10.1111/bjet.12833]
36. Rauniar R, Rawski G, Yang J, Johnson B. Technology acceptance model (TAM) and social media usage: an empirical study on Facebook. J Enterp Inf Manag 2014 Feb 4;27(1):6-30. [doi: 10.1108/JEIM-04-2012-0011]
37. Oh S, Kim JH, Choi SW, Lee HJ, Hong J, Kwon SH. Physician confidence in artificial intelligence: an online mobile survey. J Med Internet Res 2019 Mar 25;21(3):e12422. [doi: 10.2196/12422]
38. Abdullah R, Fakieh B. Health care employees' perceptions of the use of artificial intelligence applications: survey study. J Med Internet Res 2020 May 14;22(5):e17620. [doi: 10.2196/17620]
39. Nitiéma P. Artificial intelligence in medicine: text mining of health care workers' opinions. J Med Internet Res 2023 Jan 27;25:e41138. [doi: 10.2196/41138] [Medline: 36584303]
40. Teng M, Singla R, Yau O, et al. Health care students' perspectives on artificial intelligence: countrywide survey in Canada. JMIR Med Educ 2022 Jan 31;8(1):e33390. [doi: 10.2196/33390] [Medline: 35099397]
41. Beets B, Newman TP, Howell EL, Bao L, Yang S. Surveying public perceptions of artificial intelligence in health care in the United States: systematic review. J Med Internet Res 2023 Apr 4;25:e40337. [doi: 10.2196/40337] [Medline: 37014676]
42. Wagner G, Raymond L, Paré G. Understanding prospective physicians' intention to use artificial intelligence in their future medical practice: configurational analysis. JMIR Med Educ 2023 Mar 22;9:e45631. [doi: 10.2196/45631] [Medline: 36947121]
43. Dreiseitl S, Binder M. Do physicians value decision support? A look at the effect of decision support systems on physician opinion. Artif Intell Med 2005 Jan;33(1):25-30. [doi: 10.1016/j.artmed.2004.07.007]
44. qXR. India Health Fund. URL: https://www.indiahealthfund.org/our-portfolio/qxr/#:~:text=It%20is%20an%20affordable%20screening,100%20(~%241.3)%20per%20test [accessed 2025-01-27]

XSL•FO
RenderX

45. Income in india. Wikipedia. URL: https://en.wikipedia.org/wiki/Income_in_India [accessed 2025-01-27]

Abbreviations

AI: artificial intelligence
AIP: Ayurveda, yoga and naturopathy, Unani, Siddha, and homoeopathy practitioner and informal health care provider
AYUSH: Ayurveda, Yoga and Naturopathy, Unani, Siddha, and Homoeopathy
CXR: chest x-ray
LMIC: low- and middle-income country
MTB/RIF: *Mycobacterium tuberculosis*/rifampin
OR: odds ratio
TB: tuberculosis
WHO: World Health Organization
WHP: World Health Partner

Edited by A Mavragani; submitted 31.10.23; peer-reviewed by AK Madhukeshwar, K Varshney; revised version received 18.11.24; accepted 18.11.24; published 10.02.25.

Please cite as:
Kumar S, Rayal S, Bommaraju R, Varasala NP, Papineni S, Deo S
Understanding Providers' Attitude Toward AI in India's Informal Health Care Sector: Survey Study
JMIR Form Res 2025;9:e54156
URL: https://formative.jmir.org/2025/1/e54156
doi:10.2196/54156

© Sumeet Kumar, Snehil Rayal, Raghuram Bommaraju, Navya Pratyusha Varasala, Sirisha Papineni, Sarang Deo. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Developing an ICD-10 Coding Assistant: Pilot Study Using RoBERTa and GPT-4 for Term Extraction and Description-Based Code Selection

Sander Puts[1], PhD; Catharina M L Zegers[1], PhD; Andre Dekker[1], PhD; Iñigo Bermejo[1,2], PhD

[1]Department of Radiation Oncology (Maastro), GROW Research Institute for Oncology and Reproduction, Maastricht University Medical Centre+, P.O. Box 616, Maastricht, Netherlands

[2]Data Science Institute (DSI), Hasselt University, Belgium

Corresponding Author:
Sander Puts, PhD
Department of Radiation Oncology (Maastro), GROW Research Institute for Oncology and Reproduction, Maastricht University Medical Centre+, P.O. Box 616, Maastricht, Netherlands

Abstract

Background: The International Classification of Diseases (ICD), developed by the World Health Organization, standardizes health condition coding to support health care policy, research, and billing, but artificial intelligence automation, while promising, still underperforms compared with human accuracy and lacks the explainability needed for adoption in medical settings.

Objective: The potential of large language models for assisting medical coders in the ICD-10 coding was explored through the development of a computer-assisted coding system. This study aimed to augment human coding by initially identifying lead terms and using retrieval-augmented generation (RAG)–based methods for computer-assisted coding enhancement.

Methods: The explainability dataset from the CodiEsp challenge (CodiEsp-X) was used, featuring 1000 Spanish clinical cases annotated with ICD-10 codes. A new dataset, CodiEsp-X-lead, was generated using GPT-4 to replace full-textual evidence annotations with lead term annotations. A Robustly Optimized BERT (Bidirectional Encoder Representations from Transformers) Pretraining Approach transformer model was fine-tuned for named entity recognition to extract lead terms. GPT-4 was subsequently employed to generate code descriptions from the extracted textual evidence. Using a RAG approach, ICD codes were assigned to the lead terms by querying a vector database of ICD code descriptions with OpenAI's text-embedding-ada-002 model.

Results: The fine-tuned Robustly Optimized BERT Pretraining Approach achieved an overall F_1-score of 0.80 for ICD lead term extraction on the new CodiEsp-X-lead dataset. GPT-4-generated code descriptions reduced retrieval failures in the RAG approach by approximately 5% for both diagnoses and procedures. However, the overall explainability F_1-score for the CodiEsp-X task was limited to 0.305, significantly lower than the state-of-the-art F_1-score of 0.633. The diminished performance was partly due to the reliance on code descriptions, as some ICD codes lacked descriptions, and the approach did not fully align with the medical coder's workflow.

Conclusions: While lead term extraction showed promising results, the subsequent RAG-based code assignment using GPT-4 and code descriptions was less effective. Future research should focus on refining the approach to more closely mimic the medical coder's workflow, potentially integrating the alphabetic index and official coding guidelines, rather than relying solely on code descriptions. This alignment may enhance system accuracy and better support medical coders in practice.

(JMIR Form Res 2025;9:e60095) doi:10.2196/60095

KEYWORDS

International Classification of Diseases; ICD-10; computer-assisted-coding; GPT-4; coding; term extraction; code analysis; computer assisted coding; transformer model; artificial intelligence; AI automation; retrieval-augmented generation; RAG; large language model; LLM; Bidirectional Encoder Representations from Transformers; Robustly Optimized BERT Pretraining Approach; RoBERTa; named entity recognition; NER

Introduction

Background and Significance

The International Classification of Diseases (ICD), developed by the World Health Organization (WHO), serves as a universal standard for coding health-related conditions [1]. The ICD system steers health care policymaking, assists in billing and reimbursement, underpins research, ensures quality monitoring, and promotes the standardization of medical information exchange. ICD-10 is divided into diagnosis and procedure codes,

XSL•FO
RenderX

encompassing over 100,000 codes. The system is hierarchically organized and includes detailed coding guidelines.

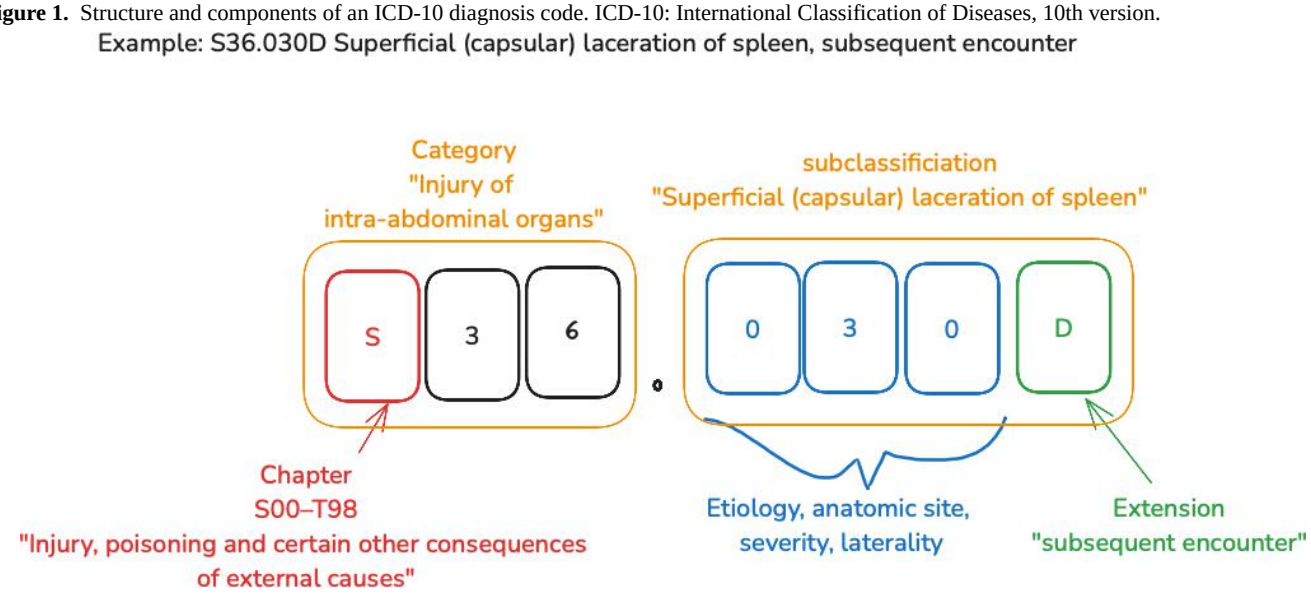
Medical coding entails the allocation of unique codes to medical records and is a standard procedure in most hospitals. Hospitals collect a list of ICD codes relevant to each patient's hospital admission. Medical coding is a time-consuming and error-prone task, which has led to interest in automating it, and in turn, to the emergence of the medical coding subfield within medical natural language processing (NLP) [1,2]. However, the task of effectively computerizing ICD coding remains a challenge. While the accuracy of codes assigned by medical coders may be debated [2-4], current state-of-the-art (SOTA) automated systems for ICD-10 coding still fall short of human performance [5-7]. In addition, many artificial intelligence (AI) solutions are "black boxes," not offering explanations for their predictions. However, model transparency is required to gain trust, which is crucial for its adoption in medical settings [8].

ICD Coding

Diagnosis codes in the ICD system are organized into an alphanumeric format across 21 chapters, capturing a range of categories including specific diseases, symptoms, and external causes, among others.

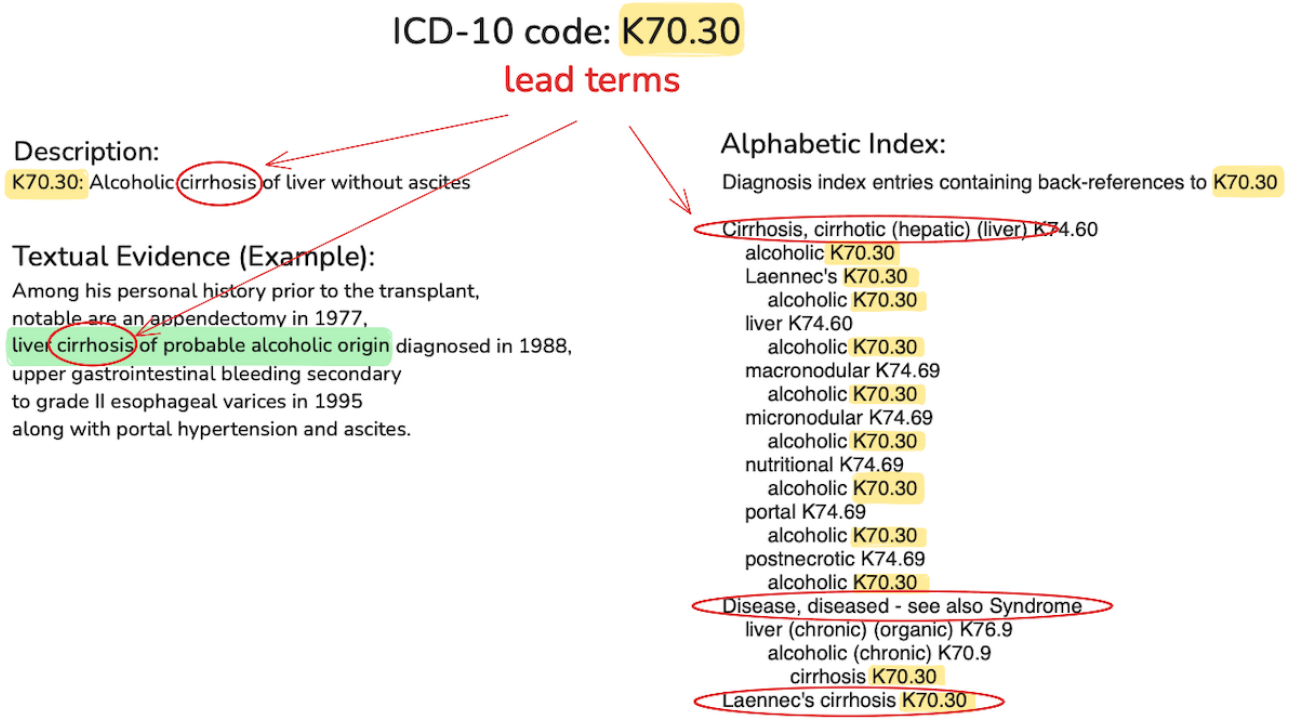
The hierarchical structure of ICD codes begins with a letter that corresponds to a chapter, together with the 2 subsequent digits that form the category of the code. Following the category code, the subclassification uses up to 5 additional alphanumeric characters. These detail specific aspects of the disease or condition, like laterality or severity, as illustrated in Figure 1. The procedure codes, separate from diagnostic codes, consist of 7 characters termed as "axis," with each axis representing a specific aspect of the medical procedure performed [9]. ICD-10 provides a comprehensive 3-volume coding manual that includes an alphabetical index. This index helps in finding medical codes using "lead terms," which are crucial keywords associated with medical conditions, listed in the alphabetical index. Medical coders first identify a diagnosis's lead term in the text (Figure 2). In this example, the lead term "cirrhosis" from the textual evidence corresponds to an alphabetic index entry. Under "Cirrhosis," the relevant subentry is "liver," "alcoholic," leading to the back-reference code K70.30. The next coding step involves consulting the tabular index for K70.30 for detailed instructions.

Figure 1. Structure and components of an ICD-10 diagnosis code. ICD-10: International Classification of Diseases, 10th version.



XSL•FO
RenderX

Figure 2. Illustration of concepts "code description," "textual evidence (green)," "lead terms (red)," and "back references" in the "alphabetical index". ICD: International Classification of Diseases.



For each diagnosis and procedure that needs an ICD code, the initial step is to identify the lead term, which is comparable to the named entity recognition (NER) task in NLP. NER involves extracting specific categories like names, locations, medications, and diseases from unstructured text. Only a few studies proposing lead term-based approaches for ICD coding have been published [5], possibly due to the absence of public datasets annotated with lead terms. Previous NER approaches focused on extracting full textual evidence, rather than lead terms [5].

Explainability

Recent studies have proposed adversarial robustness training strategies and feature attribution, such as the use of attention mechanisms, for the explainability of NLP [10-12]. A limitation of these approaches is the absence of efforts to validate whether the identified "explainable" terms align with official ICD guidelines. For instance, while "insulin" may be a predictive term for the disease "diabetes", it is not suitable as evidence for coding according to the WHO guidelines and educational textbooks [13] as insulin can be prescribed for various medical conditions other than diabetes.

Transformers

In the original transformer architecture, which encompasses both encoder and decoder components [14], it is crucial to recognize the distinct roles of these elements. Encoder-only models, such as the BERT (Bidirectional Encoder Representations from Transformers) architecture and its variant, RoBERTa (Robustly Optimized BERT Pretraining Approach) [15], excel at capturing relevant information from the input data and constructing meaningful representations [16-18] and can be found in current SOTA models in ICD coding [5,7]. Conversely, decoder models, exemplified by GPT, are specialized in generating coherent and contextually relevant text [19]. With the evolution of the GPT architecture, remarkable advancements have been achieved in NLP [20,21]. GPT has already been applied across various health care disciplines and use cases [22]. However, early studies indicate that while GPT-3.5, GPT-4, and ChatGPT show potential, their performance in ICD classification has been mixed, with out-of-the-box negative results [23]. Generative approaches may be especially useful for ICD coding of rare diseases [24]. OpenAI's GPT was, for a time, the leading large language model (LLM) in generative AI [20], but other models, such as Mixtral, Gemini, Grok, Llama, and Claude [25], are available and quickly catching up [26,27].

Objective

Our objective was to develop a GPT-4-based computer-assisted coding system to aid medical coders in ICD-10 coding. Our approach enhances the identification of diagnoses and procedures in texts by initially focusing on lead terms rather than extracting the full textual evidence for a code at once. The reason for initially extracting lead terms is that it aligns perfectly with the first step followed by medical coders in their procedure. This approach would therefore be a good fit for assisting with the process. For code normalization, our approach exclusively employs code descriptions, moving away from reliance on model-based training. Specifically, the ICD code for identified lead terms was assigned using GPT-4 and ICD descriptions, employing a retrieval-augmented generation (RAG) approach. This reliance on an external knowledge source is strategically chosen to improve generalizability, enhancing the system's ability to accurately classify codes unseen during training. Furthermore, this method offers increased adaptability to changes in the coding system. The research also investigates the optimal method for querying a database to retrieve ICD codes by description. It compares the effectiveness of 2

XSL•FO
RenderX

approaches: using direct textual evidence as the query and employing code descriptions generated (invented) by GPT-4 from that textual evidence. Additionally, the study aimed to assess GPT-4's ability to accurately select the best code by matching code descriptions with textual diagnoses. This objective is rooted in the understanding that ICD coding guidelines and descriptions are designed for humans, and it seeks to determine if GPT-4's capabilities in following instructions translate to robust performance in this specific context.

Methods

GPT-4

This study used GPT-4 (gpt-4 1106-preview) via OpenAI's application programming interface and the Langchain Python library [11]. To ensure reproducibility, the model's temperature was set to 0. GPT-4's parametric knowledge, or out-of-the-box performance, was evaluated in 2 ways. First, it was prompted to generate descriptions for 100 random ICD codes from the CodiEsp dataset. These descriptions were then compared with the ground truth by prompting GPT-4 in a new chat with, "Do both descriptions refer to the same ICD code?" Second, GPT-4 was tasked with matching the correct ICD-10 code to each of the 100 official descriptions.

Dataset CodiEsp

CodiEsp is a unique corpus of 1000 clinical cases in Spanish, annotated by professionals for explainability, offering token-level annotations that provide textual evidence for each assigned code [28]. The annotation agreement rates were 88.6% for diagnosis, 88.9% for procedures, and 80.5% for textual evidence [28]. The test set contains 1514 unique ICD codes (1143 diagnoses and 371 procedures), some of which are not present in the training set, specifically, 439 diagnosis codes (38.3%) and 231 procedure codes (62.3%). Among the procedure codes, 143 (38.5%) are 4-axis ICD-10 procedure codes, which officially should have 7-axis ICD-10 codes. These incomplete codes lack descriptions and account for 602 of the total of 1112 procedure codes (54%). Overall, 828 out of 4777 (17.3%) test annotations do not have an exact match in the training set. In the current SOTA on CodiEsp-X, the entity extraction and normalization are both transformer-based [5].

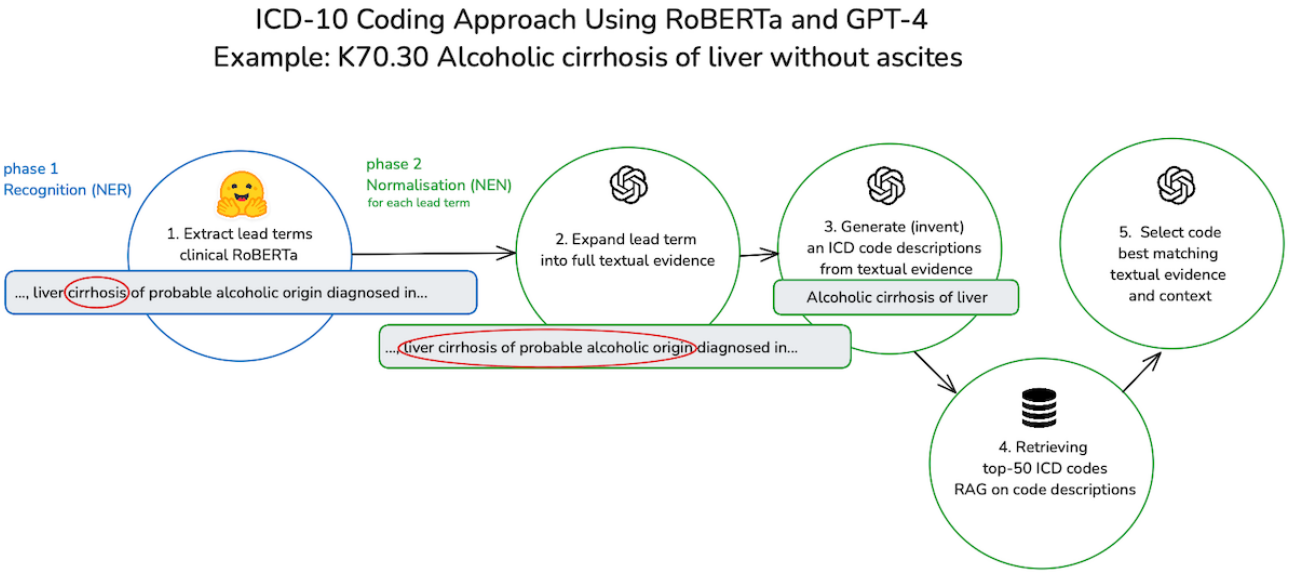
Dataset CodiEsp-X-Lead

This approach does not use NER to extract complete textual evidence; instead, it mimics human coders by first identifying lead terms. An annotated dataset was required to train and evaluate a NER model for this task. However, manual annotation of 18,435 items was deemed infeasible due to the time and effort involved. Instead, GPT-4 few-shot prompting with 20 manually curated examples was used to extract lead terms from CodiEsp-X textual evidence. The author (SP) reviewed and corrected the extraction. The resulting CodiEsp-X-lead dataset is identical to CodiEsp-X but includes additional columns for the GPT-extracted lead terms.

Approach

For this study, we followed the process described in 2 main phases and 5 steps in Figure 3, mimicking the process human medical coders follow.

Figure 3. Schematic overview of the process followed for ICD coding. ICD: International Classification of Diseases; RAG: retrieval-augmented generation; RoBERTa: Robustly Optimized BERT Pretraining Approach



Step 1: Extract Lead Terms

A RoBERTa-based biomedical-clinical language model for Spanish was fine-tuned for NER of lead terms using the CodiEsp-X-lead dataset and Hugging Face's transformer library (version 4.26.0) [19,20], using the official dataset splits. The fine-tuned model enabled precise identification of lead terms within medical documents. To evaluate its performance, the same approach was applied to the CodiEsp-X dataset to extract full-textual evidence, which is more commonly used in AI methods for code assignment.

Step 2: Expand Lead Terms Into Full Textual Evidence

Instead of using NER to extract textual evidence, GPT-4 was instructed to complete a template for each lead term which included a field to capture the full textual evidence for the code.

XSL•FO
RenderX

Alongside the textual evidence, the template includes a field designed to capture contextual information, which is described as the text snippet containing all the necessary information for coding the lead term. In the prompt, the entire text was supplied, using HTML tags to highlight the lead terms within the original text.

Step 3: Generate (Invent) an ICD Code Description From Textual Evidence

In this part, we tested the theory that GPT-4 can improve text embeddings-based database search accuracy by aligning queries with official ICD code descriptions. This assumption arose from the observation that terms in official ICD code descriptions often markedly differ from those used in clinical texts. GPT-4 was tasked to generate (invent) ICD code descriptions from extracted textual evidence, aiming to reconcile these terminological discrepancies. This process, known as query alignment, refines a query to mirror the language and structure of database entries more closely, with the goal of enhancing search accuracy. To verify this assumption, the study compared different types of input queries; raw clinical text versus GPT-4 refined descriptions, to assess their effectiveness in achieving precise database lookups.

Step 4: Retrieving ICD Codes Using Descriptions and RAG

The complete lists of code descriptions for diagnoses and procedures contain around 2.5 million tokens each, far exceeding the context window size of current generative models. While including an entire knowledge base in the prompt is feasible for smaller datasets, RAG is necessary here to retrieve relevant context efficiently. To address this, we converted the official ICD code descriptions into vectors using OpenAI's "text-embedding-ada-002" model via the Langchain library [21], and stored them in the Chroma vector database. This enables efficient searching by embedding-based similarity rather than exact matches. By retrieving only the most relevant information, RAG ensures that the results fit within the model's context window, allowing it to process the required details effectively.

Step 5: Select the Code Best Matching Textual Evidence

From the 50 results retrieved from the vector store query, GPT-4 was prompted to choose the best-matching ICD code. For the decision, GPT was provided with the lead term, textual evidence, and additional context extracted in Step 2.

Evaluation

The performance of the ICD coding pipeline and its components was evaluated using precision, recall, and F_1-score metrics calculated with the CodiEsp scripts [18,22]. To assess performance loss at each step, predictions were replaced with actual values in subsequent steps only if the predictions at the current step were accurate. Specifically, in lead term extraction, a prediction was deemed correct if the predicted text span overlapped with the textual evidence. For textual evidence extraction, if there was an overlap, all values except the span were replaced. In the code retrieval step, the predicted code was replaced with the actual code if the predicted code was among the top-k results. In the code assignment, the predictions were not altered. Additionally, the accuracy of lead term extraction was also measured using F_1-scores calculated for the actual spans, independently of the CodiEsp script.

Ethical Considerations

This study used only publicly available data from the CodiEsp challenge, a synthetic corpus of 1000 clinical case studies, that does not contain real patient information, ensuring anonymity and eliminating the need for ethical approval.

Results

GPT-4

GPT-4 generated descriptions for 100 randomly selected ICD diagnosis codes and in a new chat compared these with official ICD-10 descriptions. It found that only 52% matched the official descriptions, as detailed in Multimedia Appendix 1. When GPT-4 was tasked with associating the correct ICD-10 code to each of the 100 official descriptions, only 47% of the ICD codes were correct.

Dataset CodiEsp-X-Lead

A key outcome of this study is the creation of the CodiEsp-X-lead dataset using GPT-4. This dataset consists of 7748 unique textual evidence strings, with 6532 of these containing multiple terms. As instructed, GPT-4 typically extracted a single lead term from these multiterm strings. However, in 549 cases (8%), multiple lead terms were extracted. The researchers manually resolved 325 of these cases to a single lead term, leaving the remaining cases unmapped due to ambiguity in determining the appropriate lead term.

Step 1: Extracting Lead Terms

After fine-tuning the Biomedical-clinical language model for Spanish for NER to extract lead terms [29], it achieved an F_1-score of 0.80 (precision of 0.78 and a recall) of 0.83 on the CodiEsp-X-lead test set (Table 1). When this model was solely fine-tuned and tested for diagnoses, it exhibited a slightly better performance, with an F_1-score of 0.82 (Table 1).

XSL•FO
RenderX

Table . Named entity recognition (NER) performance on extracting lead terms and textual evidence for diagnoses and or procedures on CodiEsp-X and CodiEsp-X-lead.

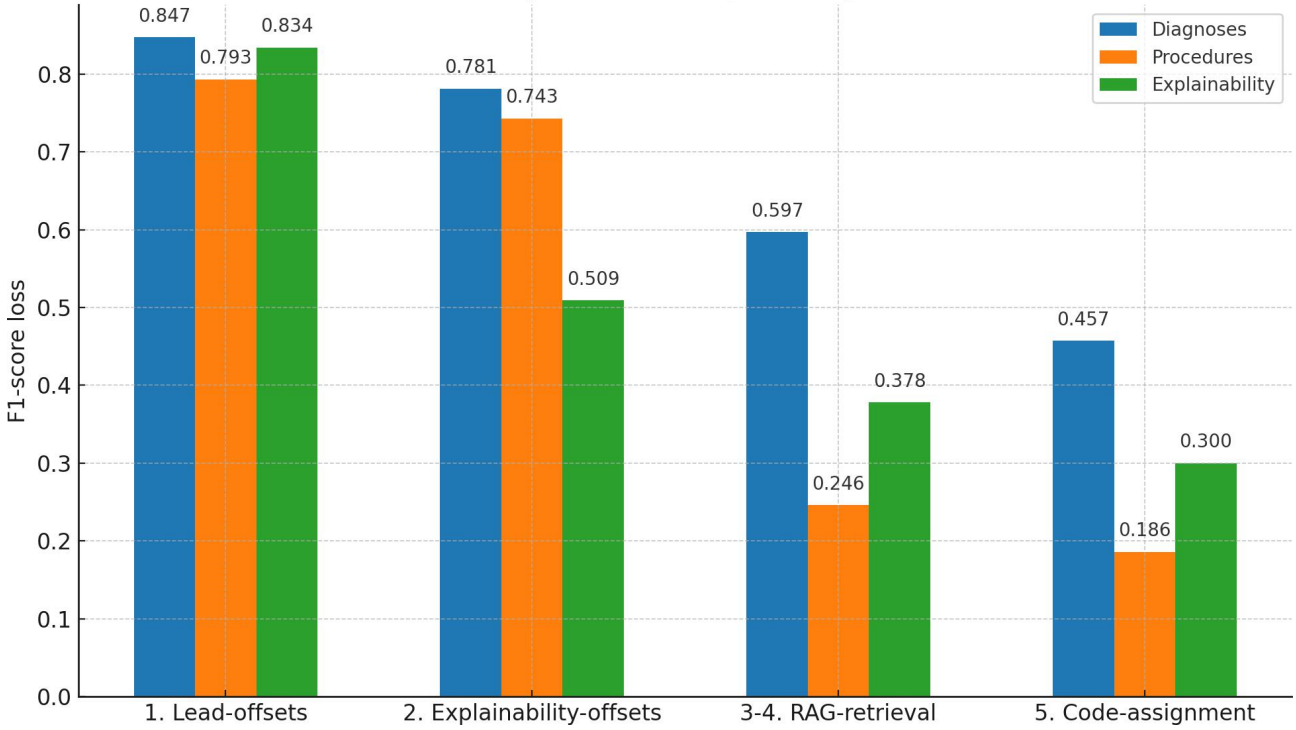
NER	F_1-score
Textual evidence diagnoses and procedures combined (ours-baseline) [29]	0.65
Textual evidence diagnoses (SOTA[a]) [5]	0.71
Textual evidence procedures (SOTA) [5]	0.61
Main-terms diagnoses (ours) [29]	0.82
Main-terms diagnoses and procedures combined (ours) [29]	0.80

[a]SOTA: state-of-the-art.

Step 2: Expand Lead Terms Into Full Textual Evidence

Using GPT-4, full textual evidence was extracted based on preannotated lead terms, which were identified in Step 1. GPT-4 completed a template that incorporated the full textual evidence for each provided lead term. This approach achieved an F_1-score of 0.509, as shown in Figure 4.

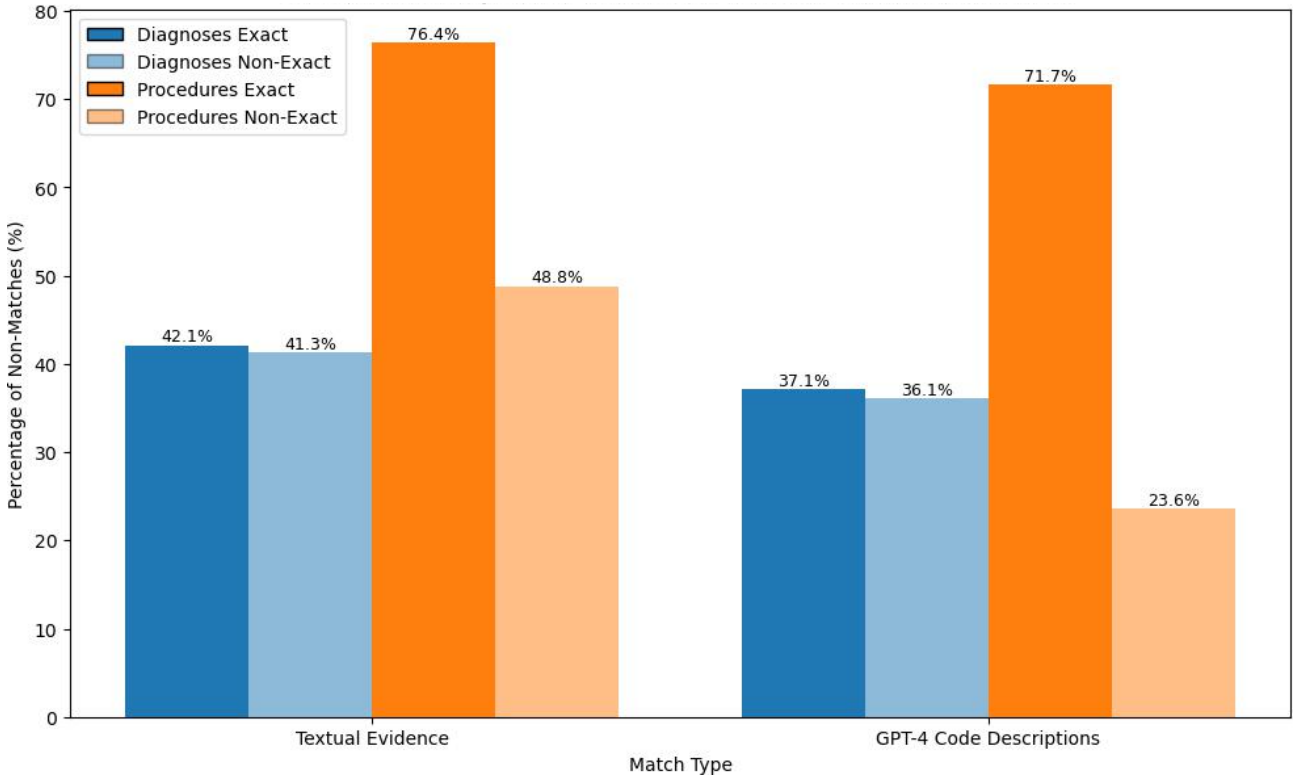
Figure 4. F1-Score performance loss for each individual step in the approach.



The base model from Step 1 was fine-tuned to extract textual evidence solely for benchmarking, achieving an F_1-score of 0.65 on the CodiEsp-X test set (Table 1). This model was not used in the actual approach, which aims to mimic medical coders by prioritizing the extraction of lead terms. Note that results from Table 1, assessing exact offsets, differ from those in Figure 5 where the official CodiEsp scripts allow for an error margin.

XSL•FO
RenderX

Figure 5. Comparison of failed retrievals @50 in a vector database containing all official ICD diagnosis and procedure codes. The chart compares results for 25 documents (507 codes), contrasting searching by textual evidence versus GPT-4 zero-shot generated code descriptions based on the textual evidence.



Step 3: Generate (Invent) an ICD Code Description From Textual Evidence

GPT-4 was used to generate ICD code descriptions from textual evidence, evaluated on 25 documents containing 507 codes in a RAG setting. When comparing GPT-4-generated descriptions to direct retrieval using textual evidence (Figure 5), retrieval failures decreased from 160 to 141 (42.1% to 37.1%) for the 380 diagnoses codes and from 97 to 91 (76.4% to 71.7%) for the 127 procedure codes.

For procedure codes, many were assigned as 4-axis codes, lacking the detail of the official 7-axis codes and without available descriptions in the resources. As a comparison, a nonexact matching criterion was applied, where a match was considered valid if the assigned code was a substring of the retrieved code. Under this nonexact approach, retrieval failures for procedures dropped significantly, from 62 to 30 out of 127 (48.8% to 23.6%). For diagnoses, which were mostly assigned in full detail, the decrease in failures was smaller, from 157 to 137 out of 380 (41.3% to 36.1%).

Step 4: Retrieving ICD Codes Using Descriptions and RAG

When RAG retrieval was evaluated within the full pipeline, a significant drop in explainability F_1-score was observed, falling from 0.509 to 0.378, as shown in steps 3 and 4 in Figure 4. Procedures had the largest contribution to this drop, with their F_1-score decreasing from 0.743 to 0.246. Diagnoses also experienced a decline, with their F_1-score dropping from 0.781 to 0.597.

Step 5: Select the Code Best Matching Textual Evidence

In the final step, GPT-4 had to select the correct ICD code from the top 50 retrieved results based on the previously predicted textual evidence. This selection process caused a further decline in performance, with the explainability F_1-score dropping from 0.378 to 0.3 (Table 2 and Figure 4).

XSL•FO
RenderX

Table . Comparison of performance of different models in the CodiEsp challenge.

Team name	Diagnoses (F_1-score)	Procedures (F_1-score)	Explainability (F_1-score)
FLE	0.679	0.514	0.611
IAM	0.687	0.522	0.611
BETO-Galén+mBERTGalén+XLM-R-Galén	—[a]	—	0.633
Ours	0.457	0.204	0.305

[a]Not available.

Discussion

Summary of Main Findings

In this study, we proposed methods to assist ICD-10 coding using LLMs. We successfully created a dataset of lead terms, "CodiEsp-X-lead," using few-shot prompting on GPT-4. Using this dataset, we effectively fine-tuned a transformer model (F_1-score of 0.80) that can assist medical coders in the initial step of medical coding: identifying ICD lead terms. To enhance RAG retrieval, we employed GPT-4 to generate code descriptions from textual evidence, aiming to improve ICD code identification within a code description database. Although the initial step, focusing on extracting lead terms, closely mimicked medical coders and performed well, subsequent RAG steps based on textual descriptions were less effective, yielding a poor F_1-score of 0.305 on the CodiEsp-X task. We will discuss the shortcomings of the RAG implementation and analyze each step's performance, highlighting areas of significant improvement.

Motivation of Choices

RoBERTa was selected over GPT-4 for NER of lead terms because encoder-based models are better suited for token-level tasks like NER, providing detailed context understanding. RoBERTa is lighter and more efficient to fine-tune, potentially outperforming generative, decoder-only models like GPT-4 [17,30]. Additionally, as of 2023, fine-tuning GPT-4 was not possible. Previous research indicates that GPT models perform poorly in traditional, example-rich NER tasks [31].

In contrast, GPT-4 excels in few-shot NER tasks and was chosen for other aspects of the study due to its SOTA performance in NLP. Medical coding involves following specific guidelines, and a well-configured generative model like GPT-4, trained for instruction-following, is likely to perform well in such roles. Therefore, GPT-4 was used to create the dataset by extracting lead terms through few-shot NER, as manually compiling the dataset was not feasible. Its natural language generation capabilities and instruction-following skills were also leveraged to generate code descriptions, align queries, and identify the correct ICD-10 codes from a given list.

Using code descriptions exclusively for normalizing lead terms to ICD codes offers independence from training data-based fine-tuning. Given the vastness of ICD-10, with over 100,000 codes, many are rarely encountered in practice. The analysis of the CodiEsp dataset revealed that 17.3% of specific test set codes were not included in the training data. If the proposed method, which primarily relies on official code descriptions, were to be successful, it likely would be very generalizable across different ICD datasets.

GPT-4 Parameteric Knowledge

The preliminary results of GPT-4's performance were disappointing. However, it is crucial to approach negative outcomes based on prompts with caution, as the prompts themselves may be suboptimal. Nevertheless, we can definitively state that GPT-4's out-of-the-box performance on 2 basic ICD-10 coding tasks—coding official descriptions and generating code descriptions for diagnoses—was mediocre. The approach of using GPT-4 generated descriptions for procedure codes was found to be completely ineffective in the study. These findings and those from the first available publications on GPT-3.5 and GPT-4's abilities with ICD billing codes underscore their subpar performance for downstream applications [23].

Creating CodiEsp-X-Lead With GPT-4

In this study, we introduced "CodiEsp-X-lead", a curated selection of lead terms from the CodiEsp-X dataset. Using GPT-4 as an annotator and employing few-shot prompting, we extracted these terms from the dataset's textual evidence. This method showcases its utility for other use cases where manual annotation is unfeasible, but reviewing the generated annotations is feasible. While the authors manually reviewed and made minor adjustments to this extraction, there is potential for further refinement. For improved lead term identification, lead terms derived from the codes assigned to that textual evidence could be suggested by using an inverse lookup on the official alphabetical index. This approach can benefit both the GPT-4 prompts given and our manual review process. Last, to ensure the highest accuracy, domain experts, such as medical coders, could annotate or validate a portion of the lead terms and can help improve the quality of the examples used in the (few-shot) prompt. Another suggestion to enhance the dataset's quality is to allow multiple lead terms for each assigned code. Limiting to a single lead term when several are applicable results in inconsistencies and reduces the quality of the dataset and the model's performance.

Step 1: Extracting Lead Terms

The results demonstrate that using NER for lead term identification holds distinct advantages over extracting the full textual evidence all at once.

XSL•FO
RenderX

First, since the task is more straightforward, significantly better performance can be obtained when extracting lead terms compared with extracting the full textual evidence. For instance, the SOTA approach for the CodiEsp-X task yielded an F_1-score of 0.71 for NER of full-textual evidence for diagnoses compared with our score of 0.82 for NER of lead terms (Table 1) [2]. High performance in extracting the correct lead term is crucial, as it sets the performance ceiling for the entire coding approach.

Second, starting with lead term extraction mirrors the initial step of human coders, facilitating seamless integration into their workflow and providing standalone value. A coding assistant that accurately identifies lead terms with high recall and precision would likely be favored over a system that offers full ICD assignments but has a higher error rate. The latter could compel human coders to frequently reread the text and make extensive corrections, frequently starting from scratch. In contrast, a fully autonomous approach that meets performance requirements without needing human intervention would likely favor extracting full-textual evidence at once, due to its potential for higher efficiency. However, in such a scenario, the relevance of explainability becomes questionable, particularly when the system operates satisfactorily with no human oversight.

The proposed NER model for extracting lead terms performs well, but further improvements can be achieved by enhancing the quality of the CodiEsp-X-lead dataset. It may be beneficial to fine-tune the model to prioritize recall over precision, as medical coders might favor removing incorrect lead term suggestions instead of revisiting a text to verify that all diagnoses are captured. Additionally, optimizing parameters might enhance the performance of the model.

Step 2: Expand Lead Terms Into Full Textual Evidence

From Step 2, our method deviated from ICD-10 coding guidelines by adopting a RAG approach, which requires a database lookup rather than following the traditional coding steps.

Using a zero-shot GPT-4 prompt to extract full textual evidence resulted in mediocre outcomes, with the F_1-score for explainability dropping from 0.824 to 0.509 (Figure 4). This highlights a significant limitation of this approach. In contrast, skipping lead term extraction and directly employing a fine-tuned BERT model yielded better performance, achieving F_1-scores of 0.71 and 0.61 for diagnoses and procedures respectively, which are SOTA, and a combined score of 0.65 for our model [5]. For future work, we propose to align more closely with medical coders' practices by continuing to use extracted lead terms and leveraging the alphabetic index and official guidelines for code lookups, allowing textual evidence to emerge naturally. This differs from the current RAG approach, which relies on a single lookup using predefined semantic queries.

Step 3: Generate (Invent) an ICD Code Description From Textual Evidence

The hypothesis we examined suggested that GPT-generated code descriptions, based on textual evidence and context, would semantically more closely align with a database of ICD code descriptions than the original textual evidence. This expected alignment is intended to enhance RAG retrieval, suggesting that rephrasing textual evidence into descriptions using AI could improve the accuracy of searching for ICD codes by their descriptions. From the experiments on a subset of the data (Figure 5), we confirm the hypothesis, as all cases show a reduction in failed retrievals. Although the improvements are subtle in most scenarios, with only a few percentage points fewer failed retrievals, there is a marked improvement for nonexact matching procedure codes. Specifically, failed retrievals dropped from 48.8% to 23.6%. This notable performance in nonexact matching is likely due to the majority of procedure codes lacking descriptions (59%), which is why nonexact matching was evaluated and resulted in a significant reduction in the number of classes.

Step 4: Retrieving ICD Codes Using Descriptions and RAG

In step 3, we tested nonexact matching for procedure codes, but the official Codiesp evaluation uses exact matching. Our approach, relying on NER extraction and descriptions for code lookup, fails when many assigned codes lack descriptions and are invalid.

In the RAG retrieval step, F_1-score performance for procedure codes drops sharply from 0.743 to 0.246. Diagnosis code performance is also mediocre at 0.597, but this would still be acceptable if GPT-4 could select the correct code in the next step without causing another drop.

Step 5: Select the Code Best Matching Textual Evidence

In step 5, GPT-4 selects the final answer from the retrieved codes, but another drop in performance is observed, resulting in a final F_1-score of 0.3 on CodiEsp-x, compared with the SOTA score of 0.633 (Table 2) [5]. It is likely that the proposed textual evidence (span) is incorrect in some cases, making it valuable to evaluate this step in isolation to assess its impact on overall performance.

Method Recap and Evaluation

While the initial NER step for extracting lead terms is promising and can directly assist medical coders, subsequent steps perform inadequately. Although using RAG with code descriptions seems intuitive for ICD-10 coding, the implementation underperforms. While lead term extraction works well, the following steps should better mimic the workflow of medical coders. If this is not done, step 2 could be replaced by extracting textual evidence via NER, which achieves a higher F_1-score (0.65 in Table 1) compared with the current method (0.509 in Figure 5).

Instead of relying on descriptions to find the correct code, it would be more effective to use the alphabetic index of lead terms. Textual evidence could then be derived from this lookup using iterative steps, rather than the current approach, which requires knowing the exact spans of evidence in advance before performing the database lookup, as we need to know exactly what we are searching for.

XSL•FO
RenderX

Additionally, terms like NEC (not elsewhere classified) and NOS (not otherwise specified), which are present in some code descriptions, could be useful for an agent-based approach, allowing an agent to interpret and follow these instructions in coding scenarios. This would enhance performance, as an agent can adaptively handle these cases, which are often challenging for a simple database lookup.

Incorporating more coding guidelines into the process would further align with an agent-based model, closely replicating the real-world workflow of medical coders.

Limitations

Our hypothesis was that GPT-4 can achieve effective medical ICD-10 code normalization (assignment) by solely using the official WHO-supplied ICD-10 code descriptions. The findings suggest that the combination of GPT-4 with solely code descriptions might not be sufficient.

Relying solely on the WHO descriptions for medical coding can be problematic. A prime example is the challenge in coding "lower limb weakness" when it must fit into a broad category description like "Other malaise", highlighting issues with semantic differences. Although the embedding methods used address some similarity issues, using an alphabetic index or leveraging training data might be more effective and accurate, but care must be taken to maintain generalizability.

A major challenge in using GPT-4 for research lies in its cost, processing speed, and privacy concerns. For each individual medical text, processing with GPT-4 using the proposed prompts typically takes several minutes at a minimum. Furthermore, when evaluating 50 code candidates, each accompanied by descriptions for the lead term extracted, the token use becomes significant, resulting in substantial costs. Additionally, GPT-4's closed-source nature presents further complications in research applications.

Another limitation is the exclusive use of the CodiEsp dataset, which restricts the generalizability of our findings. While the MIMIC dataset is a commonly used public resource for ICD coding, it lacks annotations for explainability. No other publicly available ICD dataset is currently coded for explainability.

Future Research

For NER of lead terms and their normalization to ICD codes, we suggest continuing with the NER approach for extracting lead terms, given its promising performance and its seamless integration into the workflow of medical coders. For future research, we propose implementing an agent-based LLM approach that mimics all steps of the medical coders' workflow in detail.

Another point for future research is to explore the use of locally runnable LLMs, ensuring that data can remain in-house for privacy and security purposes.

Conclusions

A dataset of ICD-10 lead terms, "CodiEsp-X-lead", was created using few-shot prompting on GPT-4. This method of model-assisted labeling showcases its utility for other use cases where manual annotation is unfeasible, but reviewing the generated annotations is feasible. As GPT models are not the SOTA in NER [31], a RoBERTa-based model was fine-tuned, achieving an F_1-score of 0.80. This model can aid medical coders in the primary step of medical coding by identifying ICD lead terms. GPT-4 was employed to generate code descriptions based on textual evidence, improving RAG retrieval for code lookups in a code description database. However, exclusive reliance on code descriptions, coupled with GPT-4 prompting for ICD-10 coding, led to only mediocre outcomes. Continuing with the NER approach for lead term extraction is advisable. Relying solely on code descriptions, along with GPT-4 prompts for ICD-10 coding, led to mediocre outcomes, partly because some codes in the dataset lacked descriptions. Additionally, this approach is likely less effective than mimicking the complete workflow of medical coders, which could yield more accurate results. A key limitation of using RAG for medical coding is its requirement for full-textual evidence upfront, making it less flexible than methods that extract evidence during the lookup process using official coding resources and guidelines.

Acknowledgments

The authors acknowledge the use of ChatGPT (GPT-4 and GPT-4o, OpenAI, 2023/2024) to assist in improving the language and readability of this manuscript. ChatGPT was specifically employed to rephrase sentences for clarity and enhance the overall articulation of the text. The authors carefully reviewed and refined the rephrased content to ensure that the final text accurately reflected their intended meaning and original ideas.

Data Availability

The datasets generated or analyzed during this study are available in the "Enhancing ICD coding with LLMs" repository [32].

The official CodiEsp dataset is openly available at [33].

Conflicts of Interest

None declared.

Multimedia Appendix 1
GPT-4 parametric ICD-10 knowledge.

XSL•FO
RenderX

[DOCX File, 17 KB - formative_v9i1e60095_app1.docx]

References

1. Alharbi MA, Isouard G, Tolchard B. Historical development of the statistical classification of causes of death and diseases. Cog Med 2021 Jan 1;8(1):1893422. [doi: 10.1080/2331205X.2021.1893422]
2. Kim BH, Ganapathi V. Read, attend, and code: pushing the limits of medical codes prediction from clinical notes by machines. 2021 Presented at: Proceedings of the 6th Machine Learning for Healthcare Conference; Aug 6-7, 2021; Proceedings of Machine Learning Research p. 196-208.
3. Joos C, Lawrence K, Jones AE, Johnson SA, Witt DM. Accuracy of ICD-10 codes for identifying hospitalizations for acute anticoagulation therapy-related bleeding events. Thromb Res 2019 Sep;181:71-76. [doi: 10.1016/j.thromres.2019.07.021] [Medline: 31357146]
4. Burns EM, Rigby E, Mamidanna R, Bottle A, Aylin P, Ziprin P, Faiz OD A, Aylin P, Ziprin P, et al. Systematic review of discharge coding accuracy. J Public Health (Oxf) 2012 Mar;34(1):138-148. [doi: 10.1093/pubmed/fdr054] [Medline: 21795302]
5. López-García G, Jerez JM, Ribelles N, Alba E, Veredas FJ. Explainable clinical coding with in-domain adapted transformers. J Biomed Inform 2023 Mar;139:104323. [doi: 10.1016/j.jbi.2023.104323] [Medline: 36813154]
6. Barros J, Rojas M, Dunstan J, Abeliuk A. Divide and conquer: an extreme multi-label classification approach for coding diseases and procedures in spanish. : Association for Computational Linguistics; 2022 Presented at: Proceedings of the 13th International Workshop on Health Text Mining and Information Analysis (LOUHI); Abu Dhabi, United Arab Emirates (Hybrid). [doi: 10.18653/v1/2022.louhi-1.16]
7. Huang CW, Tsai SC, Chen YN. PLM-ICD: automatic ICD coding with pretrained language models. arXiv. Preprint posted online on 2022 URL: http://arxiv.org/abs/2207.05289 [accessed 2023-02-22]
8. Hulsen T. Explainable artificial intelligence (XAI): concepts and challenges in healthcare. AI 2023;4(3):652-666. [doi: 10.3390/ai4030034]
9. Centers for Medicare & Medicaid Services. ICD-10-PCS official guidelines for coding and reporting 2023. URL: https://www.cms.gov/files/document/2023-official-icd-10-pcs-coding-guidelines.pdf [accessed 2024-11-20]
10. Xie X, Xiong Y, Yu PS, Zhu Y. EHR coding with multi-scale feature attention and structured knowledge graph propagation. : Association for Computing Machinery; 2019 Nov 3 Presented at: Proceedings of the 28th ACM International Conference on Information and Knowledge Management; Nov 3-7, 2019; Beijing China p. 649-658. [doi: 10.1145/3357384.3357897]
11. Amjad H, Ashraf M, Sherazi S, et al. Attention-based explainability approaches in healthcare natural language processing. : SCITEPRESS; 2023 Presented at: Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies; Lisbon, Portugal p. 689-696. [doi: 10.5220/0011927300003414]
12. Edin J, Maistro M, Maaløe L, Borgholt L, Havtorn JD, Ruotsalo T. An unsupervised approach to achieve supervised-level explainability in healthcare records. arXiv. Preprint posted online on 2024 URL: http://arxiv.org/abs/2406.08958 [accessed 2024-10-09]
13. Safian SC, Johnson MA. Let's Code It! 2019-2020 Code Edition, 2nd edition: McGraw-Hill Education; 2020.
14. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. arXiv. URL: http://arxiv.org/abs/1706.03762
15. Liu Y, Ott M, Goyal N, Du J J, Joshi M, Chen D, et al. RoBERTa: a robustly optimized BERT pretraining approach. arXiv. Preprint posted online on 2019 URL: http://arxiv.org/abs/1907.11692
16. Devlin J, Chang MW, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. arXiv. Preprint posted online on 2018 URL: http://arxiv.org/abs/1810.04805
17. Lu Q, Li R, Wen A, Wang J, Wang L, Liu H. Large language models struggle in token-level clinical named entity recognition. arXiv. Preprint posted online on 2024 URL: http://arxiv.org/abs/2407.00731
18. Clark K, Khandelwal U, Levy O, At M. What does BERT look at? an analysis of BERT's attention. arXiv. Preprint posted online on 2019 URL: http://arxiv.org/abs/1906.04341
19. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. OpenAI 2018 [FREE Full text]
20. OpenAI GPT-4 technical report. arXiv. Preprint posted online on 2023 URL: http://arxiv.org/abs/2303.08774
21. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. arXiv. Preprint posted online on 2020 URL: http://arxiv.org/abs/2005.14165 [accessed 2023-10-12]
22. Li J, Dada A, Puladi B, Kleesiek J, Egger J. ChatGPT in healthcare: a taxonomy and systematic review. Comput Methods Programs Biomed 2024 Mar;245:108013. [doi: 10.1016/j.cmpb.2024.108013] [Medline: 38262126]
23. Soroush A, Glicksberg BS, Zimlichman E, et al. Assessing GPT-3.5 and GPT-4 in generating international classification of diseases billing codes. medRxiv. Preprint posted online on Jul 7, 2023. [doi: 10.1101/2023.07.07.23292391]
24. Yang Z, Kwon S, Yao Z, Yu H. Multi-label few-shot ICD coding as autoregressive generation with prompt. Proc AAAI Conf Artif Intell 2023 Jun 26;37(4):5366-5374. [doi: 10.1609/aaai.v37i4.25668] [Medline: 37635946]
25. Safavi-Naini SAA, Ali S, Shahab O, Shahhoseini Z, Savage T, Rafiee ST, et al. Vision-language and large language model performance in gastroenterology. arXiv. Preprint posted online on 2024 URL: http://arxiv.org/abs/2409.00084

XSL•FO
RenderX

26. Gema AP, Leang JOJ, Hong G, Devoto A, Mancino ACM, Saxena R, et al. Are we done with MMLU. arXiv. Preprint posted online on 2024 URL: http://arxiv.org/abs/2406.04127
27. Chiang WL, Zheng L, Sheng Y, Angelopoulos AN, Li T, Li D, et al. Chatbot arena: an open platform for evaluating llms by human preference. arXiv. Preprint posted online on 2024 URL: http://arxiv.org/abs/2403.04132
28. Miranda-Escalada A, Gonzalez-Agirre A, Armengol-Estapé J, Krallinger M. Overview of automatic clinical coding: annotations, guidelines, and solutions for non-English clinical cases at codiesp track of CLEF eHealth 2020. 2020 Presented at: Working Notes of Conference and Labs of the Evaluation (CLEF) Forum CEUR Workshop Proceedings; Sep 22-25, 2020; Thessaloniki, Greece.
29. Carrino CP, Llop J, Pàmies M, et al. Pretrained biomedical language models for clinical NLP in spanish. : Association for Computational Linguistics; 2022 Presented at: Proceedings of the 21st Workshop on Biomedical Language Processing; Dublin, Ireland p. 193-199. [doi: 10.18653/v1/2022.bionlp-1.19]
30. Hu Y, Chen Q, Du J, Peng X, Keloth VK, Zuo X, et al. Improving large language models for clinical named entity recognition via prompt engineering. J Am Med Inform Assoc 2024 Sep 1;31(9):1812-1820. [doi: 10.1093/jamia/ocad259] [Medline: 38281112]
31. Wang S, Sun X, Li X, Ouyang R, Wu F, Zhang T, et al. GPT-NER: named entity recognition via large language models. arXiv. Preprint posted online on 2023 URL: http://arxiv.org/abs/2304.10428 [accessed 2023-07-20]
32. Enhancing ICD coding with LLMs. GitHub. 2024. URL: https://github.com/putssander/icdllmeval [accessed 2025-01-20]
33. Miranda EA, Gonzalez AA, Krallinger M. CodiEsp corpus: gold standard Spanish clinical cases coded in ICD10 (CIE10) - eHealth CLEF2020. Z 2020. [doi: 10.5281/zenodo.3625746]

Abbreviations

BERT: Bidirectional Encoder Representations from Transformer
ICD: International Classification of Diseases
LLM: large language model
NEC: not elsewhere classified
NER: named entity recognition
NLP: natural language processing
NOS: not otherwise specified
RAG: retrieval-augmented generation
RoBERTa: Robustly Optimized BERT Pretraining Approach
SOTA: state-of-the-art
WHO: World Health Organization

Edited by A Mavragani; submitted 01.05.24; peer-reviewed by SS Sohail, T Karen; revised version received 20.11.24; accepted 24.11.24; published 11.02.25.

Please cite as:
Puts S, Zegers CML, Dekker A, Bermejo I
Developing an ICD-10 Coding Assistant: Pilot Study Using RoBERTa and GPT-4 for Term Extraction and Description-Based Code Selection
JMIR Form Res 2025;9:e60095
URL: https://formative.jmir.org/2025/1/e60095
doi:10.2196/60095

© Sander Puts, Catharina M L Zegers, Andre Dekker, Iñigo Bermejo. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Web-Based, Algorithm-Guided Insulin Titration in Insulin-Treated Type 2 Diabetes: Pre-Post Intervention Study

Nishanth Thiagarajan, MBBS; Hong Chang Tan, MBBS; Suresh Rama Chandran, MBBS; Phong Ching Lee, MBChB; Yun Ann Chin, MBBS; Wanling Zeng, MD; Emily Tse Lin Ho, MBBS; David Carmody, MBChB; Su-Yen Goh, MBBS; Yong Mong Bee, MBBS

Department of Endocrinology, Singapore General Hospital, 20 College Road, Academia, Singapore, Singapore

Corresponding Author:
Nishanth Thiagarajan, MBBS
Department of Endocrinology, Singapore General Hospital, 20 College Road, Academia, Singapore, Singapore

Abstract

Background: Self-monitoring of blood glucose (SMBG) using web-based diabetes management platforms has demonstrated promise in managing type 2 diabetes (T2D). However, the effectiveness of such systems incorporating algorithm-guided insulin titration has not been extensively studied in Asian populations.

Objective: This study evaluates the efficacy and safety of the ALRT telehealth solution—a US Food and Drug Administration–cleared, web-based platform that integrates SMBG with algorithm-driven insulin dose adjustments—in improving glycemia in insulin-treated T2D.

Methods: This 24-week, pre-post intervention study enrolled 25 adults with T2D (mean age 58.9, SD 7.0 y; n=14, 56% male) on twice-daily premixed insulin. Inclusion criteria included a baseline hemoglobin A_{1c} (HbA_{1c}) level between 7.5% to 9.9% (58 - 86 mmol/mol), a BMI ≤40 kg/m², and experience with SMBG. Participants uploaded twice-daily SMBG data weekly via a mobile app, which generated insulin titration recommendations based on a predefined algorithm. Physicians reviewed and approved the recommendations, which were then communicated back to participants via the app. The primary outcome was the change in HbA_{1c} level from baseline to 24 weeks. Secondary outcomes included changes in fasting plasma glucose, insulin dose, hypoglycemia incidence, and SMBG adherence.

Results: Participants achieved significant reductions in HbA_{1c} level from 8.6% (70 mmol/mol) at baseline to 7.4% (57 mmol/mol) at 24 weeks (P<.001), with reductions of 0.8% and 0.4% in the first and second 12 weeks, respectively. Fasting plasma glucose decreased from 8.7 (SD 2.0) mmol/L to 7.1 (SD 1.4) mmol/L (P<.001). Mean total daily insulin dose increased modestly from 0.73 (SD 0.31) units/kg/day to 0.79 (SD 0.34) units/kg/day (P=.007). Participants demonstrated high adherence, completing 97.3% (327/336) of prescribed SMBG measurements. During the study, 48% (12/25) of participants experienced at least 1 hypoglycemia episode, predominantly mild hypoglycemia (85/96, 88.5%; glucose 3.0 - 3.9 mmol/L). Hypoglycemia episodes increased from 24 during weeks 0 - 12 to 72 during weeks 13 - 24. There were no episodes of severe hypoglycemia requiring external assistance. BMI increased slightly from 29.0 (SD 3.6) kg/m² to 29.5 (SD 3.6) kg/m² (P=.03), reflecting a modest weight gain associated with improved glycemia.

Conclusions: In conclusion, patients with insulin-treated T2D initiated on a web-based glucose monitoring system with algorithm-guided dosing recommendations showed significant improvement in glycemic control compared to baseline. High adherence rates underscore the feasibility of integrating algorithm-guided insulin titration into routine care. While hypoglycemia incidence rose slightly, episodes were predominantly mild, and no severe events occurred. This intervention shows promise for broader adoption in T2D management, particularly in resource-constrained settings.

(JMIR Form Res 2025;9:e68914) doi:10.2196/68914

KEYWORDS

diabetes; insulin; monitoring; technology; mobile; app; intervention

Introduction

The global prevalence of diabetes mellitus has risen dramatically, with projections estimating that over 1.3 billion people will be affected by 2050 [1]. This chronic disease profoundly impacts both quality of life [2] and health care expenditure [3], and has driven significant efforts toward early detection and treatment.

Effective glycemic control remains the corner stone of diabetes management, correlating with improved long-term outcomes [4-6]. Despite rapid advances in pharmacotherapy, up to

XSL•FO
RenderX

two-thirds of patients fail to achieve their glycemic targets [7]. Self-monitoring of blood glucose (SMBG) is a pivotal component of diabetes management [8], enhancing patient engagement and glycemic control [9]. The benefits of SMBG are most pronounced [10-12] when a structured approach is taken. Ideally, blood glucose readings should be collected at specified intervals and returned to health care providers for interpretation. In this regard, a safe and consistent dose adjustment regimen will be needed if any treatment modification is required. Finally, patients need to be promptly informed of the necessary changes. Thus, structured SMBG's effectiveness is contingent on regular, frequent physician reviews and adjustments of therapy [13,14].

However, having frequent in-person clinic visits to review SMBG and adjust insulin doses is impractical, given the massive patient load and long waiting time of many diabetes clinics. Furthermore, a vast amount of SMBG data needs to be reviewed and interpreted by the physician during the limited allocated clinic time. This process is laborious and time-consuming, yet decision-making for insulin dose adjustments needs to be rapid. Providers with less clinical experience will find this especially challenging. Consequently, the SMBG data are often not fully capitalized to improve a patient's diabetes management.

Several web-based programs and mobile apps have been developed to overcome some of the abovementioned challenges. Instead of manual recording of SMBG and then providing feedback to their provider in person, these programs allow patients to record and transmit SMBG data directly via the app, reducing the frequency of face-to-face visits [15-17]. In an earlier study, we demonstrated the ability of a web-based blood glucose monitoring system to optimize insulin dosages for patients receiving a basal-plus or basal-bolus insulin regimen. Importantly, glycemic control improved without increasing hypoglycemia rates [18]. However, although this approach reduced the need for face-to-face visits for insulin titration, physicians still need to review a large amount of SMBG data before recommending any insulin dose adjustment. In addition, insulin doses are adjusted based on the physician's discretion and experience. Such an approach requires extensive workforce resources and is unlikely to solve the real-world problem of delays in treatment optimization. One potential solution is to use an automated insulin dose adjustment system that could interpret SMBG data and provide dose adjustment recommendations based on a preprogrammed algorithm. Studies have demonstrated the superiority of such systems in improving glycemic control compared to physician-dependent adjustments alone [19]. However, such programs have not been studied in an Asian cohort.

We thus designed the Glucose Monitoring and Intervention in Insulin-treated Type 2 Diabetes Patients (GEMINI-T2D) study to examine the efficacy and safety of a web-based platform that incorporates monitoring and insulin titration recommendations based on a built-in algorithm in patients treated with insulin.

Methods

Study Design

The GEMINI-T2D study was a prospective, single-site, single-group, pre-post interventional study.

Study Participants

All participants were recruited from the Singapore General Hospital's Diabetes & Metabolism Centre from September 2020 to May 2022. Those deemed potentially eligible were contacted by phone, by email, or during scheduled visits to the Diabetes & Metabolism Centre. Those who expressed interest in participating were given the informed consent form and invited to participate. Final eligibility of those who consented to participate was based on satisfactory completion of the screening process, demonstrating full eligibility based on inclusion and exclusion criteria.

Inclusion Criteria

We recruited adults (aged ≥21 y) with type 2 diabetes (T2D) of more than 6-month duration on twice-daily premixed insulin regimen for ≥3 months. The total daily dose (TDD) of insulin must be <1 unit/kg, hemoglobin A_{1c} (HbA_{1c}) level must be between 7.5% (58 mmol/mol) to 9.9% (86 mmol/mol), and BMI must be ≤40 kg/m^2. We chose to study patients on premixed insulin as it represents a common and preferred insulin formulation in Asia [20,21]. Participants were required to have experience performing SMBG and own a compatible smartphone capable of weekly data uploads via the study app.

Exclusion Criteria

We excluded individuals with hypoglycemia unawareness (Gold score ≥4) to limit the risk of severe hypoglycemia during treatment. Pregnant or breastfeeding women, as well as those with severe renal impairment (estimated glomerular filtration rate <30 mL/min/1.73 m^2), hemoglobinopathies, systemic corticosteroid use, and medical disease with a life expectancy of less than 1 year, were excluded.

Intervention

Eligible participants were given a glucometer and instructed to perform SMBG twice daily. They were trained to use the ALRT system—a US Food and Drug Administration–cleared diabetes management system—and asked to upload their SMBG data to the ALRT app weekly for the next 24 weeks. Participants were also educated on hypoglycemia recognition and management, and instructed to perform SMBG when experiencing symptoms suggestive of hypoglycemia.

Insulin Dose Adjustment

The ALRT system analyzed uploaded SMBG data every 7 days. When capillary blood glucose readings consistently fell outside the predefined range of 4 - 8 mmol/L, a predefined algorithm adjusted the prebreakfast and/or predinner insulin doses, implementing an increment or decrement of up to 15% to 20% as necessary (details of the algorithm are provided in Multimedia Appendix 1). The treating physician was notified via a web-based user interface with the recommended insulin dose adjustments and was required to accept or decline the

XSL•FO
RenderX

recommendation. Subsequently, the adjusted insulin doses were communicated to the participants via the ALRT app every 7 days, or more frequently in the event of hypoglycemia. Participants were required to acknowledge the recommended dose adjustments via the app.

Baseline and Follow-Up Assessments

Medical history, vital signs, body measurements, and laboratory measurements were collected at baseline. HbA_{1c} level and fasting plasma glucose (FPG) were measured at baseline, week 12, and week 24.

Outcome Measures

The primary outcome measure was the change in HbA_{1c} level at the end of 24 weeks relative to baseline. Secondary outcome measures included the incidence of hypoglycemia, defined as the number of episodes of hypoglycemia per participant, and the adherence to the prescribed SMBG regimen. Episodes of hypoglycemia were classified as follows:

- Level 1: Glucose value of 3.0 - 3.9 mmol/L
- Level 2: Glucose value of <3.0 mmol/L
- Level 3: Severe hypoglycemia causing altered mentation requiring external assistance for recovery

Adherence to the SMBG regimen was defined as the proportion of the prescribed twice-daily SMBG that the participant successfully completed and uploaded to the ALRT platform. The incidence of hypoglycemia was defined as the number of episodes of hypoglycemia that occurred during each 12-week period of the study.

Statistical Analysis

Continuous data were presented as mean (SD) and categorical data were presented as numbers (percentages). We compared the variables at baseline, week 12, and week 24, using repeated-measures ANOVA for normally distributed variables with post hoc Tukey method, and Friedman ANOVA for nonnormally distributed variables with post hoc Wilcoxon signed-rank tests with Bonferroni correction. Statistical analyses were performed using R (version 4.1.1; R Foundation for Statistical Computing).

Ethical Considerations

The SingHealth Institutional Review Board reviewed and approved the study (reference 2019/2874). Informed consent was obtained from all participants, and this was inclusive of secondary analysis. The study data were deidentified. Participants were compensated with a nominal transport allowance of SGD $80 (US $59.14) for each of the clinic visits at weeks 0, 12, and 24.

Results

Participant Characteristics

A total of 25 participants were recruited and followed up over 24 weeks. The baseline demographic and clinical characteristics are summarized in Table 1. The study cohort had a mean age of 58.9 (SD 7.0) years, was 56% (n=14) male, and had a mean BMI of 29.0 (SD 3.6) kg/m^2.

XSL•FO
RenderX

Table . Baseline demographic and clinical characteristics of 25 adults with type 2 diabetes enrolled in the GEMINI-T2D[a] study, a 24-week pre-post intervention study conducted at the Singapore General Hospital from September 2020 to May 2022. All participants were on twice-daily premixed insulin and met inclusion criteria for HbA_{1c}[b] level (7.5% - 9.9%), BMI (≤40 kg/m²), and experience with self-monitoring of blood glucose.

Characteristics	Value (N=25)
Age (years), mean (SD)	58.9 (7.0)
Sex, n (%)	
Male	14 (56)
Female	11 (44)
Ethnicity, n (%)	
Chinese	13 (52)
Malay	4 (16)
Indian	6 (24)
Others	2 (8)
BMI (kg/m^2), mean (SD)	29.0 (3.6)
HbA_{1c} level (%), mean (SD)	8.6 (0.7)
HbA_{1c} level (mmol/mol), mean (SD)	70.4 (7.7)
Fasting plasma glucose (mmol/L), mean (SD)	8.7 (2.0)
Duration of diabetes (years), mean (SD)	18.9 (6.7)
Duration of insulin use (years), mean (SD)	6.3 (3.5)
Total daily dose of insulin (units/day), mean (SD)	57.3 (24.4)
Total daily dose of insulin (units/kg/day), mean (SD)	0.73 (0.31)
Hypertension, n (%)	22 (88)
Hyperlipidemia, n (%)	23 (92)
Ischemic heart disease, n (%)	3 (12)
Stroke, n (%)	1 (4)
Retinopathy, n (%)	12 (48)
Neuropathy, n (%)	4 (16)
Nephropathy, n (%)	6 (24)
Peripheral vascular disease, n (%)	2 (8)

[a]GEMINI-T2D: Glucose Monitoring and Intervention in Insulin-treated Type 2 Diabetes Patients.

[b]HbA_{1c}: hemoglobin A_{1c}.

HbA_{1c} Level and Other Glycemic Variables

Mean baseline HbA_{1c} level and FPG were 8.6% (SD 0.7%) and 8.7 (SD 2.0) mmol/L, respectively (Table 2). The mean baseline TDD of insulin was 57.3 (SD 24.4) units/day or 0.73 (SD 0.31) units/kg/day. All 25 participants experienced a reduction in HbA_{1c} level of ≥0.4% at 24 weeks. Mean HbA_{1c} level decreased by 1.2% over 24 weeks, from 8.6% (SD 0.7%) to 7.4% (SD 0.6%; *P*<.001; Figure 1). From weeks 0 - 12, the mean HbA_{1c} level decreased by 0.8% (*P*<.001), and from weeks 13 - 24, the mean HbA_{1c} level decreased by 0.4% (*P*<.001). Individuals with a higher baseline HbA_{1c} level experienced a greater absolute decrease in HbA_{1c} level over 24 weeks (*P*<.001; Figure 1).

Mean FPG also decreased by 1.6 mmol/L over 4 weeks, from 8.7 (SD 2.0) mmol/L to 7.1 (SD 1.4) mmol/L (*P*<.001; Figure 1). From weeks 0 - 12, mean FPG decreased by 1.3 mmol/L (*P*=.02), and from weeks 13 - 24, mean FPG decreased by 0.3 mmol/L (*P*>.99).

The mean TDD of insulin increased from 0.73 (SD 0.31) units/kg/day to 0.79 (SD 0.34) units/kg/day (*P*=.01; Figure 1). From weeks 0 - 12, the mean TDD of insulin increased by 0.05 units/kg/day (*P*=.02), and from weeks 13 - 24, the mean TDD of insulin increased by 0.01 units/kg/day (*P*>.99). BMI rose from 29.0 kg/m^2 to 29.5 kg/m^2 over 24 weeks (*P*=.03).

XSL•FO
RenderX

Table . Changes in key glycemic outcomes, insulin doses, and BMI from baseline to weeks 12 and 24 in the GEMINI-T2D[a] study, which evaluated the efficacy of a web-based, algorithm-guided insulin titration system in adults with insulin-treated type 2 diabetes at the Singapore General Hospital from September 2020 to May 2022. *P* values indicate trends over the study period.

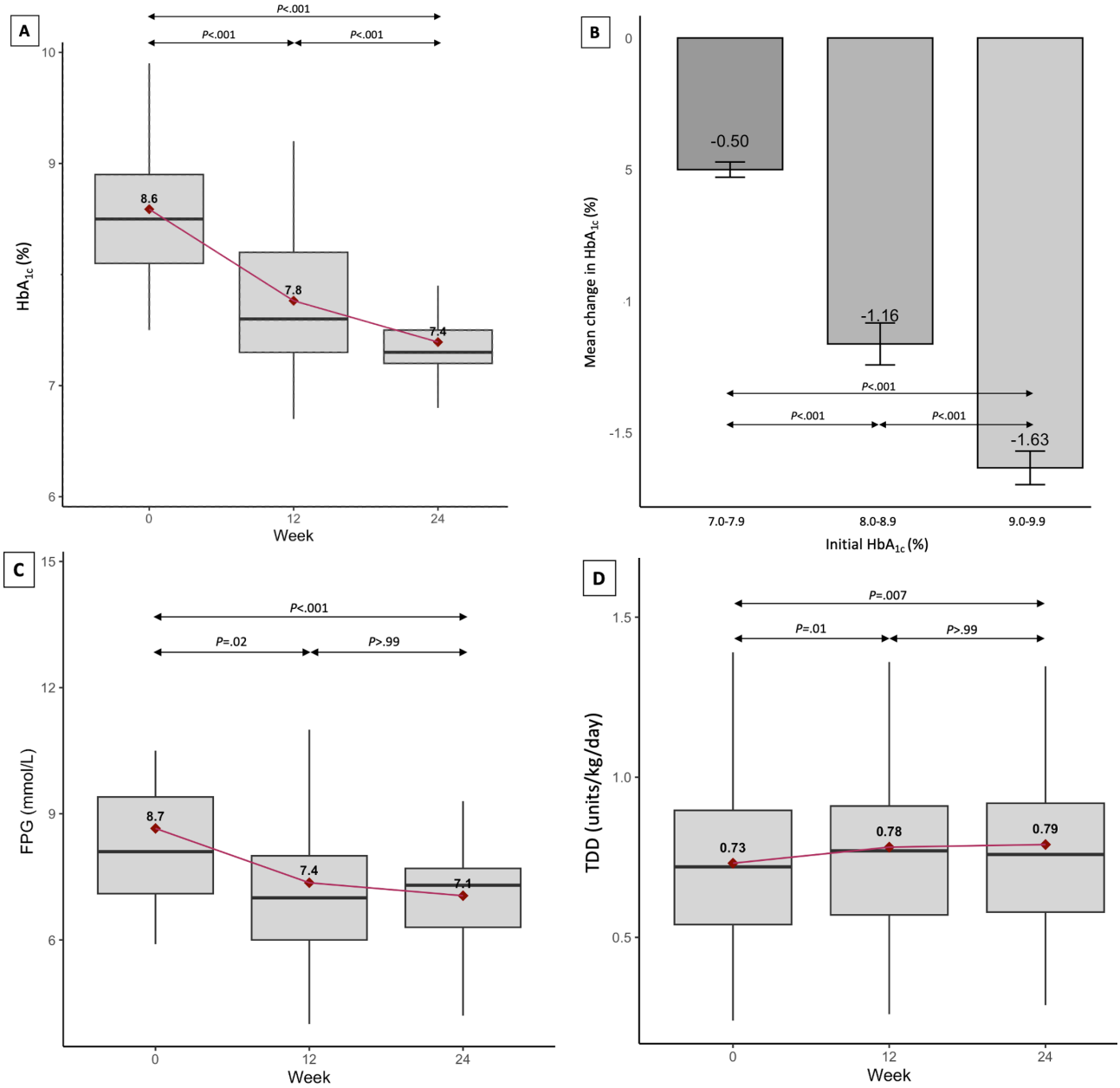
Variable	Baseline, mean (SD)	Week 12, mean (SD)	Week 24, mean (SD)	*P* value for trend
HbA_{1c}[b] level (%)	8.6 (0.7)	7.8 (0.6)	7.4 (0.6)	<.001
HbA_{1c} level (mmol/mol)	70.4 (7.7)	61.4 (7.0)	57.3 (6.7)	<.001
Fasting plasma glucose (mmol/L)	8.7 (2.0)	7.4 (1.9)	7.1 (1.4)	<.001
Total daily dose of insulin (units/day)	57.3 (24.4)	61.3 (25.7)	62.1 (25.4)	.01
Total daily dose of insulin (units/kg/day)	0.73 (0.31)	0.78 (0.33)	0.79 (0.34)	.007
BMI (kg/m^2)	29.0 (3.6)	29.3 (3.6)	29.5 (3.6)	.04

[a]GEMINI-T2D: Glucose Monitoring and Intervention in Insulin-treated Type 2 Diabetes Patients.

[b]HbA_{1c}: hemoglobin A_{1c}.

XSL•FO
RenderX

Figure 1. Trends in (A) hemoglobin A_{1c} (HbA_{1c}) level, (B) mean change in HbA_{1c} level stratified by baseline HbA_{1c} level, (C) fasting plasma glucose (FPG), and (D) total daily dose (TDD) of insulin (units/kg/day) during the 24-week GEMINI-T2D study. The study assessed the impact of algorithm-guided insulin titration delivered via a web-based telehealth platform in 25 adults with insulin-treated type 2 diabetes at the Singapore General Hospital from September 2020 to May 2022. GEMINI-T2D: Glucose Monitoring and Intervention in Insulin-treated Type 2 Diabetes Patients.



Hypoglycemia and Adherence to SMBG

At baseline, 44% (11/25) of participants reported at least 1 episode of hypoglycemia in the preceding month. During our study, 48% (12/25) of participants experienced at least 1 episode of hypoglycemia. In total, 24 episodes of hypoglycemia occurred in weeks 0 - 12, while 72 episodes occurred in weeks 13 - 24.

The episodes of hypoglycemia were predominantly confined to level 1 (mild; blood glucose 3.0 - 3.9 mmol/L), with 85 (89%) out of 96 episodes falling in this category (Table 3). There were no level-3 (severe) episodes of hypoglycemia reported during the study. Adherence to the prescribed SMBG frequency was 98.2% (165/168) for weeks 0 - 12 and 96.5% (162/168) for weeks 12 - 24.

XSL•FO
RenderX

Table . Incidence and severity of hypoglycemia episodes per participant during the 24-week GEMINI-T2D[a] study. Episodes were categorized by level: mild (3.0 - 3.9 mmol/L), moderate (<3.0 mmol/L), and severe (requiring external assistance). Data highlight an increase in mild hypoglycemia episodes from weeks 0 - 12 to weeks 13 - 24. The study was conducted in 25 adults with insulin-treated type 2 diabetes using a web-based telehealth platform at the Singapore General Hospital from September 2020 to May 2022.

Level of hypoglycemia	Incidence of hypoglycemia	
	Weeks 1 - 12	Weeks 13 - 24
1 (mild)	0.88	2.52
2 (moderate)	0.08	0.36
3 (severe)	0	0

[a]GEMINI-T2D: Glucose Monitoring and Intervention in Insulin-treated Type 2 Diabetes Patients.

Discussion

The GEMINI-T2D study demonstrated that remote monitoring of SMBG data via a glucose management platform, coupled with algorithm-guided insulin titration, may improve glycemic control in patients with insulin-treated T2D. To the best of our knowledge, this is the first study of an integrated, mobile phone–based solution in an Asian cohort with insulin-treated diabetes. The degree of HbA_{1c} improvement in our study was comparable to studies on smartphone-based diabetes management platforms [15,18,22] and a larger trial on an automated insulin titration system [19].

Several key factors likely contributed to the improvement in glycemic control in our cohort. The ALRT system empowered individuals to participate in structured SMBG, through a platform that allowed for ease of upload of blood glucose readings directly from a glucometer. It also provided a convenient way to store, visualize, and communicate the data with their physicians, receiving timely feedback. The weekly review of data followed by insulin dose adjustments also tightened the feedback loop between patients and their physicians, enabling more prompt insulin dose adjustments than usually possible in a clinic setting. Physicians are also likely to have been empowered by the algorithm-guided insulin dose recommendations, which provide a framework to base their therapeutic decisions upon.

In addition, we observed that the improvement in HbA_{1c} level during the first half of the study persisted into the second half, demonstrating sustained benefits from the intervention. Interestingly, although the average increase in TDD of insulin was only 0.05 units/kg/day or 3 units/day for a 60-kg person, HbA_{1c} level decreased by 1.7%. This modest increase in insulin dose alone is unlikely to account for such a substantial improvement in glycemia. We speculate that the “outsized” glycemic improvement may be attributed to a more optimal redistribution of insulin doses facilitated by structured SMBG.

Another contributing factor to the improved glycemic control could be enhanced patient empowerment, which may have led to positive changes in overall health, diet, and lifestyle. This is supported by the high adherence to the prescribed SMBG regimen observed in our cohort, exceeding rates typically reported in the literature [23]. Enhanced patient engagement, consistently recognized as a critical factor in managing chronic diseases such as diabetes [24-26], was a central focus of our intervention. This engagement was fostered through patient education, twice-daily SMBG, and frequent review with feedback from physicians.

With regard to the safety of our intervention, while an increase in level-1 (mild) episodes of hypoglycemia occurred with tighter glycemic control, we found a low rate of level-2 (moderate) hypoglycemia, with no episodes of level-3 (severe) hypoglycemia. This safety signal was likely contributed by a few measures, such as patient education on hypoglycemia recognition and management at the time of enrollment, along with automated alerts with algorithm-guided insulin dosing suggestions to the treating physician via the ALRT app in event of hypoglycemia, prompting timely adjustment of insulin doses ahead of the scheduled weekly SMBG reviews. The safety profile aligns with a US-based randomized controlled trial [19] that showed similar hypoglycemia rates with automated insulin titration guidance compared to standard care. Nevertheless, dynamic adjustments to the algorithm to minimize hypoglycemia rates will be warranted in future implementation of this intervention in other populations.

A web-based platform with algorithm-guided insulin dose titration holds promise for closing the titration gap for people on insulin therapy, offering dynamic and frequent adjustments based on data, while assisting busy physicians in delivering necessary care safely and conveniently. Moving forward, it is also fathomable that artificial intelligence may allow for yet larger scales of SMBG data management and analysis, and provide more personalized insulin dosing recommendations that take into account factors beyond glycemic trend such as diet, lifestyle, and general health.

The primary limitation of our study is a small sample size, which reflects its design as a pilot study of an intervention. While the pre-post intervention design of our study suggests the effectiveness of our intervention compared to prior standard care, a longer study with a larger sample size would be warranted to evaluate the long-term effectiveness and scalability of such a web-based platform in diabetes care. Incorporating a comparison arm would further clarify the intervention’s performance against standard care and help identify the factors contributing to the improved glycemic control observed in our cohort. Additionally, our strict inclusion criteria may limit the generalizability of our findings to a broader population with diabetes, warranting further exploration. Finally, understanding physicians’ experiences with this intervention would provide

XSL•FO
RenderX

valuable insights into the practical challenges and opportunities in its integration into routine clinical practice.

In conclusion, patients with insulin-treated T2D initiated on a web-based glucose monitoring system with algorithm-guided dosing recommendations showed significant improvement in glycemic control compared to baseline. High adherence rates underscore the feasibility of integrating algorithm-guided insulin titration into routine care. While hypoglycemia incidence rose slightly, episodes were predominantly mild, and no severe events occurred. This intervention shows promise for broader adoption in T2D management, particularly in resource-constrained settings.

Acknowledgments

No generative artificial intelligence (AI) or AI-assisted technologies were used in this study. This research was supported by A*STAR, Singapore, under its Industry Alignment Pre-Positioning Fund (grant H19/01/a0/023 – Diabetes Clinic of the Future), and ALR Technologies.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Insulin dose titration algorithm.
[DOCX File, 15 KB - formative_v9i1e68914_app1.docx]

References

1. GBD 2021 Diabetes Collaborators. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2023 Jul 15;402(10397):203-234. [doi: 10.1016/S0140-6736(23)01301-6] [Medline: 37356446]
2. Trikkalinou A, Papazafiropoulou AK, Melidonis A. Type 2 diabetes and quality of life. World J Diabetes 2017 Apr 15;8(4):120-129. [doi: 10.4239/wjd.v8.i4.120] [Medline: 28465788]
3. Magliano DJ, Boyko EJ, IDF Diabetes Atlas 10th Edition Scientific Committee. IDF Diabetes Atlas, 10th Edition: International Diabetes Federation; 2021. URL: https://www.ncbi.nlm.nih.gov/books/NBK581934/ [accessed 2025-01-20]
4. Diabetes Control and Complications Trial Research Group, Nathan DM, Genuth S, et al. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med 1993 Sep 30;329(14):977-986. [doi: 10.1056/NEJM199309303291401] [Medline: 8366922]
5. Ohkubo Y, Kishikawa H, Araki E, et al. Intensive insulin therapy prevents the progression of diabetic microvascular complications in Japanese patients with non-insulin-dependent diabetes mellitus: a randomized prospective 6-year study. Diabetes Res Clin Pract 1995 May;28(2):103-117. [doi: 10.1016/0168-8227(95)01064-k] [Medline: 7587918]
6. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HAW. 10-year follow-up of intensive glucose control in type 2 diabetes. N Engl J Med 2008 Oct 9;359(15):1577-1589. [doi: 10.1056/NEJMoa0806470] [Medline: 18784090]
7. Mannucci E, Monami M, Dicembrini I, Piselli A, Porta M. Achieving HbA1c targets in clinical trials and in the real world: a systematic review and meta-analysis. J Endocrinol Invest 2014 May;37(5):477-495. [doi: 10.1007/s40618-014-0069-6] [Medline: 24699860]
8. Benjamin EM. Self-monitoring of blood glucose: the basics. Clin Diabetes 2002 Jan 1;20(1):45-47. [doi: 10.2337/diaclin.20.1.45]
9. Karter AJ, Ackerson LM, Darbinian JA, et al. Self-monitoring of blood glucose levels and glycemic control: the Northern California Kaiser Permanente Diabetes registry. Am J Med 2001 Jul;111(1):1-9. [doi: 10.1016/s0002-9343(01)00742-2] [Medline: 11448654]
10. Polonsky WH, Fisher L, Schikman CH, et al. Structured self-monitoring of blood glucose significantly reduces A1C levels in poorly controlled, noninsulin-treated type 2 diabetes: results from the Structured Testing Program study. Diabetes Care 2011 Feb;34(2):262-267. [doi: 10.2337/dc10-1732] [Medline: 21270183]
11. Parkin CG, Buskirk A, Hinnen DA, Axel-Schweitzer M. Results that matter: structured vs. unstructured self-monitoring of blood glucose in type 2 diabetes. Diabetes Res Clin Pract 2012 Jul;97(1):6-15. [doi: 10.1016/j.diabres.2012.03.002] [Medline: 22464874]
12. Lalić NM, Lalić K, Jotić A, et al. The impact of structured self-monitoring of blood glucose combined with intensive education on HbA1c levels, hospitalizations, and quality-of-life parameters in insulin-treated patients with diabetes at

XSL•FO
RenderX

primary care in Serbia: the multicenter SPA-EDU study. J Diabetes Sci Technol 2017 Jul;11(4):746-752. [doi: 10.1177/1932296816681323] [Medline: 28625073]
13. Chow N, Shearer D, Aydin Plaa J, et al. Blood glucose self-monitoring and internet diabetes management on A1c outcomes in patients with type 2 diabetes. BMJ Open Diabetes Res Care 2016 Apr 29;4(1):e000134. [doi: 10.1136/bmjdrc-2015-000134] [Medline: 27158516]
14. Chircop J, Sheffield D, Kotera Y. Systematic review of self-monitoring of blood glucose in patients with type 2 diabetes. Nurs Res 2021;70(6):487-497. [doi: 10.1097/NNR.0000000000000542] [Medline: 34292228]
15. Yu Y, Yan Q, Li H, et al. Effects of mobile phone application combined with or without self-monitoring of blood glucose on glycemic control in patients with diabetes: a randomized controlled trial. J Diabetes Investig 2019 Sep;10(5):1365-1371. [doi: 10.1111/jdi.13031] [Medline: 30815973]
16. Cui M, Wu X, Mao J, Wang X, Nie M. T2DM self-management via smartphone applications: a systematic review and meta-analysis. PLoS One 2016 Nov 18;11(11):e0166718. [doi: 10.1371/journal.pone.0166718] [Medline: 27861583]
17. Katalenich B, Shi L, Liu S, et al. Evaluation of a remote monitoring system for diabetes control. Clin Ther 2015 Jun 1;37(6):1216-1225. [doi: 10.1016/j.clinthera.2015.03.022] [Medline: 25869625]
18. Rama Chandran S, Tan HC, Chen Q, et al. Telemonitoring with a connected glucose meter improves glycemia among people with insulin-treated type 2 diabetes. J Diabetes Sci Technol 2023 Jul;17(4):909-915. [doi: 10.1177/19322968231157387] [Medline: 36825611]
19. Bergenstal RM, Johnson M, Passi R, et al. Automated insulin dosing guidance to optimise insulin management in patients with type 2 diabetes: a multicentre, randomised controlled trial. Lancet 2019 Mar 16;393(10176):1138-1148. [doi: 10.1016/S0140-6736(19)30368-X] [Medline: 30808512]
20. Kalra S, Balhara YPS, Sahay BK, Ganapathy B, Das AK. Why is premixed insulin the preferred insulin? novel answers to a decade-old question. J Assoc Physicians India 2013 Jan;61(1 Suppl):9-11. [Medline: 24482980]
21. Chinese Elderly Type 2 Diabetes Prevention and Treatment of Clinical Guidelines Writing Group, Geriatric Endocrinology and Metabolism Branch of Chinese Geriatric Society, Geriatric Endocrinology and Metabolism Branch of Chinese Geriatric Health Care Society, Geriatric Professional Committee of Beijing Medical Award Foundation, National Clinical Medical Research Center for Geriatric Diseases (PLA General Hospital). Clinical guidelines for prevention and treatment of type 2 diabetes mellitus in the elderly in China [Article in Chinese]. Zhonghua Nei Ke Za Zhi 2022 Jan 1;61:12-50. [doi: 10.3760/cma.j.cn112138-20211027-00751] [Medline: 34979769]
22. Zhou W, Chen M, Yuan J, Sun Y. Welltang - a smart phone-based diabetes management application - Improves blood glucose control in Chinese people with diabetes. Diabetes Res Clin Pract 2016 Jun;116:105-110. [doi: 10.1016/j.diabres.2016.03.018] [Medline: 27321324]
23. Hu ZD, Zhang KP, Huang Y, Zhu S. Compliance to self-monitoring of blood glucose among patients with type 2 diabetes mellitus and its influential factors: a real-world cross-sectional study based on the Tencent TDF-I blood glucose monitoring platform. Mhealth 2017 Jun 27;3:25. [doi: 10.21037/mhealth.2017.06.01] [Medline: 28736734]
24. Barello S, Graffigna G, Vegni E. Patient engagement as an emerging challenge for healthcare services: mapping the literature. Nurs Res Pract 2012;2012:905934. [doi: 10.1155/2012/905934] [Medline: 23213497]
25. Forbat L, Cayless S, Knighting K, Cornwell J, Kearney N. Engaging patients in health care: an empirical study of the role of engagement on attitudes and action. Patient Educ Couns 2009 Jan;74(1):84-90. [doi: 10.1016/j.pec.2008.07.055] [Medline: 18790594]
26. Chang YT, Tu YZ, Chiou HY, Lai K, Yu NC. Real-world benefits of diabetes management app use and self-monitoring of blood glucose on glycemic control: retrospective analyses. JMIR Mhealth Uhealth 2022 Jun 15;10(6):e31764. [doi: 10.2196/31764] [Medline: 35704371]

Abbreviations

FPG: fasting plasma glucose
GEMINI-T2D: Glucose Monitoring and Intervention in Insulin-treated Type 2 Diabetes Patients
HbA_{1c}: hemoglobin A_{1c}
SMBG: self-monitoring of blood glucose
T2D: type 2 diabetes
TDD: total daily dose

XSL•FO
RenderX

Edited by A Mavragani; submitted 17.11.24; peer-reviewed by A Andersen, X Gao; revised version received 28.12.24; accepted 29.12.24; published 07.02.25.

Please cite as:
Thiagarajan N, Tan HC, Rama Chandran S, Lee PC, Chin YA, Zeng W, Ho ETL, Carmody D, Goh SY, Bee YM
Web-Based, Algorithm-Guided Insulin Titration in Insulin-Treated Type 2 Diabetes: Pre-Post Intervention Study
JMIR Form Res 2025;9:e68914
URL: https://formative.jmir.org/2025/1/e68914
doi:10.2196/68914

© Nishanth Thiagarajan, Hong Chang Tan, Suresh Rama Chandran, Phong Ching Lee, Yun Ann Chin, Wanling Zeng, Emily Tse Lin Ho, David Carmody, Su-Yen Goh, Yong Mong Bee. Originally published in JMIR Formative Research (https://formative.jmir.org), 7.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Assessment of Digital Capabilities by 9 Countries in the Alliance for Healthy Cities Using AI: Cross-Sectional Analysis

Hocheol Lee, PhD

Department of Health Administration, College of Software and Digital Healthcare Convergence, Yonsei University, PO Box 26493, Wonju, Repubic of Korea

Corresponding Author:
Hocheol Lee, PhD
Department of Health Administration, College of Software and Digital Healthcare Convergence, Yonsei University, PO Box 26493, Wonju, Repubic of Korea

Abstract

Background: The Alma-Ata Declaration of 1978 initiated a global focus on universal health, supported by the World Health Organization (WHO) through healthy cities policies. The concept emerged at the 1984 Toronto "Beyond Health Care" conference, leading to WHO's first pilot project in Lisbon in 1986. The WHO continues to support regional healthy city networks, emphasizing digital transformation and data-driven health management in the digital era.

Objective: This study explored the capabilities of digital healthy cities within the framework of digital transformation, focusing on member countries of the Asian Forum of Healthy Cities. It examined the cities' preparedness and policy needs for transitioning to digital health.

Methods: A cross-sectional survey was conducted of 9 countries—Australia, Cambodia, China, Japan, South Korea, Malaysia, Mongolia, the Philippines, and Vietnam—from August 1 to September 21, 2023. The 6-section SPIRIT (setting approach and sustainability; political commitment, policy, and community participation; information and innovation; resources and research; infrastructure and intersectoral; and training) checklist was modified to assess healthy cities' digital capabilities. With input from 3 healthy city experts, the checklist was revised for digital capabilities, renaming "healthy city" to "digital healthy city." The revised tool comprises 8 sections with 33 items. The survey leveraged ChatGPT (version 4.0; OpenAI, Microsoft), accessed via Python (Python Software Foundation) application programming interface. The *openai* library was installed, and an application programming interface key was entered to use ChatGPT (version 4.0). The "GPT-4 Turbo" model command was applied. A qualitative analysis of the collected data was conducted by 5 healthy city experts through group deep-discussions.

Results: The results indicate that these countries should establish networks and committees for sustainable digital healthy cities. Cambodia showed the lowest access to electricity (70%) and significant digital infrastructure disparities. Efforts to sustain digital health initiatives varied, with countries such as Korea focusing on telemedicine, while China aimed to build a comprehensive digital health database, highlighting the need for tailored strategies in promoting digital healthy cities. Life expectancy was the highest in the Republic of Korea and Japan (both 84 y). Access to electricity was the lowest in Cambodia (70%) with the remaining countries having had 95% or higher access. The internet use rate was the highest in Malaysia (97.4%), followed by the Republic of Korea (97.2%), Australia (96.2%), and Japan (82.9%).

Conclusions: This study highlights the importance of big data-driven policies and personal information protection systems. Collaborative efforts across sectors for effective implementation of digital healthy cities. The findings suggest that the effectiveness of digital healthy cities is diminished without adequate digital literacy among managers and users, suggesting the need for policies to improve digital literacy.

(JMIR Form Res 2025;9:e62935) doi:10.2196/62935

KEYWORDS

digital capabilities; digital health cities; digital transformation; Asian Forum of Healthy Cities; assessment; digital health; artificial intelligence; AI; World Health Organization; WHO; healthy city; data; health management; digital era; qualitative analysis; cross-sectional survey; database; digital health database; effectiveness; digital literacy

Introduction

The Alma-Ata Declaration of 1978 marked a global shift toward universal health, a movement endorsed by the World Health Organization (WHO) through its advocacy for healthy cities as a strategic policy to enhance global health [1]. The idea of a healthy city was first conceptualized at the 1984 Toronto "Beyond Health Care" conference, and the first pilot project was spearheaded by the WHO European Office in Lisbon in April 1986 [2]. Subsequently, the International Health Cities Committee was established, and over 3000 cities now have healthy cities policies. The WHO remains a steadfast supporter

XSL•FO
RenderX

of regional healthy city networks. In Europe, the WHO European Healthy Cities Network, founded in 1988, continuously promotes networking through assemblies every 5 years, involving over 1400 municipalities [3]. In North America, a network was established with Canada as the central player in 1984, which now includes over 200 North American cities. In the Western Pacific region, the Alliance for Healthy Cities (AFHC) was established in 2003. It certifies and manages healthy cities as alliance members, facilitating dynamic networking through biennial general assemblies [4]. In the Western Pacific region, 9 countries—Australia, Cambodia, China, Japan, the Republic of Korea, Malaysia, Mongolia, the Philippines, and Vietnam—are actively participating as full members of the AFHC. The AFHC is an international network focused on protecting and promoting the health of urban residents through collaborative efforts.

Healthy cities are dedicated to improving physical and social environments where all community members collaborate to enhance the health and quality of life of citizens [5]. The evaluation of healthy cities is categorized into distinct areas: personal, community, socioeconomic infrastructure, environment, and evaluation management. These domains all adopt a top-down approach, underscoring the pivotal role of policy makers in formulating and executing policies and initiatives for healthy cities [6]. Research has shown that the proactive stance of the city's leaders has a positive impact on community residents' health [7]. Additionally, healthy city projects must be developed considering the specific contemporary and environmental demands of the country. Worldwide, healthy cities are adopting unique characteristics, transitioning to models such as senior-friendly cities, smart healthy cities, digital healthy cities, and green healthy cities [8,9]. Following the coronavirus disease 2019 pandemic, there was an acceleration in digital transformation, a trend pivotal in the context of healthy cities. The WHO highlighted that healthy cities in the digital era should focus on data-driven monitoring and the implementation of digital platforms for managing the public's health [10]. With the advancement of digital technology, it has become crucial for healthy cities to prioritize the development of digital competencies.

The WHO's "Global Strategy on Digital Health 2020 - 2025" emphasizes the role of healthy cities in the digital age, including citizen-led health management; policy making based on big data; and the expansion of digital platforms, including telemedicine [11]. Particularly, the strategy underscores the importance of health management through information technology devices and the significance of digital health at the national level [11]. Additionally, the AFHC stresses the need to integrate aging, smart technology, and digital solutions in healthy city strategies for the Western Pacific region. It prioritizes digital-focused strategies for advancing healthy cities [12].

The world is preparing for the digital era through evident forms of digital transformation such as digital healthy cities and smart healthy cities. Previous studies have stressed the importance of equipping nations and cities with digital competency readiness to prepare for this transformation [13]. However, according to United Nations digital indicators, there are disparities in digital capabilities among countries in the Western Pacific region and the AFHC member countries [14].

Several methods are used to measure the capabilities of healthy cities, such as the WHO European Healthy City Evaluation Model and the AFHC SPIRIT (setting approach and sustainability political commitment, policy, and community participation; information and innovation; resources and research; infrastructure and intersectoral; and training) checklist [15]. These 2 models, conceptualized and evaluated by healthy city experts, provide a structured, logical framework for objectively assessing various aspects of a city, including its health status, capabilities, lifestyle, vision, strategy, and network. Healthy cities worldwide use these models, and they have recently been used to understand the capabilities of healthy cities in the digital era.

In the context of digital transformation, this study aims to investigate the capabilities of digital healthy cities in all AFHC member countries, exploring policy implications for areas that require further development and readiness to effectively promote the transition to digital healthy cities.

Methods

Study Design

This cross-sectional study aims to inspect the digital healthy city capabilities among AFHC member countries. Nine countries were included: Australia, Cambodia, China, Japan, the Republic of Korea, Malaysia, Mongolia, the Philippines, and Vietnam.

Instruments

We used a modified version of the SPIRIT checklist developed and used by AFHC to assess the capabilities of healthy cities [16,17]. The checklist consists of 6 sections: setting approach and sustainability, political commitment, policy, and community participation, information and innovation, resources and research, infrastructure and intersectoral, and training. We modified the tool with input from 3 healthy city experts to more specifically evaluate digital capabilities, particularly changing the term "healthy city" used in the original tool to "digital healthy city." The modified tool contained 8 sections with 33 items.

Data Collection

Data were collected using a database powered by a large-scale artificial intelligence (AI) application. ChatGPT (version 4.0), developed by OpenAI, was accessed via an application programming interface in Python and used to conduct our survey. We installed the *openai* library in Python and entered an application programming interface key issued for the use of ChatGPT (version 4.0). In the *openai* library, the "GPT-4 Turbo" model command was used; GPT-4 Turbo is a large AI language model that has acquired knowledge up to March 2023.

The query command was set up to allow the input of questions. Additionally, the "message" command was used to define the conversation's history for generation. The system option facilitated the setup of a chatbot helper for conversation options, and messages were relayed directly to the helper via the assistant. The user option was used to save previous

XSL•FO
RenderX

conversations and maintain dialogue continuity. The results were made available through the answer command.

The questions on the SPIRIT checklist were revised to reflect names of the 9 surveyed AFHC member countries. The survey was conducted from August 1 to September 21, 2023, and the collected data were stored in a report. No human participants were involved, and thus institutional review board (IRB) approval was not obtained.

Data Analysis

Five healthy city experts performed a qualitative analysis on the collected data through a group deep discussion. The experts were recruited from October 1, 2023, to November 3, 2023. They had experience conducting activities within the AFHC International Network, held PhDs, and were affiliated with healthy city networks in their respective countries. Specifically, the group comprised 3 public health PhDs from Yonsei University's Healthy City Research Center and two professors who are academic members of the Healthy City Society. A consultation meeting was held on November 13, 2023. Prior to the qualitative analysis meeting, written informed consent was obtained from the 5 experts; the consent covered this study's purpose and plan and the objectives of the qualitative analysis meeting. Through this process, we obtained the experts' advice on the appropriateness of the contents of the SPIRIT checklist.

First, the collected reports were summarized and organized based on key areas on the checklist. Second, a relational analysis was conducted of keywords from the summarized reports to visualize the results. Third, 3 digital healthy city experts conducted a final review and discussed the digital capabilities and directions for each country. The indicators for healthy cities in the 9 surveyed countries (Table 1) were analyzed using the official AFHC data, World Bank's DataBank, and International Telecommunication Union Data Hub [14,18,19].

XSL•FO
RenderX

Table . AFHC[a] member cities.

	AUS[b]	KHM[c]	CHN[d]	JPN[e]	KOR[f]	MYS[g]	MNG[h]	PHL[i]	VNM[j]
Number of healthy cities, n[k]	5	1	34	31	109	1	5	12	1
Population, million (year)	25.9 (2021)	16.7 (2022)	1410 (2022)	125.1 (2022)	51.6 (2022)	33.9 (2022)	3.3 (2022)	115.5 (2022)	98.1 (2022)
GDP[l] per capita, US $ (year)	64,491 (2022)	1786 (2022)	12,720 (2022)	33,815 (2022)	32,254 (2022)	11,971 (2022)	4946 (2022)	3498 (2022)	4163 (2022)
GDP growth, annual % (year)	3.6 (2022)	1.1 (2022)	3 (2022)	−0.4 (2022)	2.6 (2022)	8.7 (2022)	4.8 (2022)	7.6 (2022)	8 (2022)
Poverty headcount ratio, % (year)	0.5 (2018)	N/A[m]	0.1 (2019)	0.7 (2013)	0.2 (2016)	0 (2018)	0.7 (2018)	3 (2018)	0.7 (2018)
Life expectancy at birth, total years (year)	83 (2021)	70 (2021)	78 (2021)	84 (2021)	84 (2021)	75 (2021)	71 (2021)	69 (2021)	71 (2021)
Access to electricity, %	100	82	100	100	100	100	100	97	100
Safely managed sanitation services, %	74	N/A	70	81	100	77	56	61	N/A
Internet use rate, %	96.2	60.2	75.6	82.9	97.2	97.4	81.6	52.7	78.6

[a]AFHC: Alliance for Healthy Cities.

[b]AUS: Australia.

[c]KHM: Cambodia.

[d]CHN: China.

[e]JPN: Japan.

[f]KOR: the Republic of Korea.

[g]MYS: Malaysia.

[h]MNG: Mongolia.

[i]PHL: the Philippines.

[j]VNM: Vietnam.

[k]The number of cities enrolled in the Healthy Cities initiative in the respective country.

[l]GDP: gross domestic product.

[m]N/A: not available.

Ethical Considerations

All components of this survey were approved by the IRB of Yonsei University in Korea (IRB 1041849-202401-SB-018-01). This study does not involve human subjects research; therefore, no compensation procedures were required, and there are no issues related to privacy or confidentiality. These aspects were thoroughly reviewed and confirmed by the Yonsei University IRB.

Results

AFHC Member Countries

Of the 9 AFHC member countries included in this study, the Republic of Korea had the largest number of healthy cities (109), followed by China (34) and Japan (31). China had the largest population (1410 million), followed by Japan (125.1 million) and the Philippines (115.5 million). Gross domestic product per capita was the highest in Australia (64,491), followed by Japan (33,815) and Korea (32,254). Life expectancy was the highest in the Republic of Korea and Japan (both 84 y). Access to

XSL•FO
RenderX

electricity was the lowest in Cambodia (70%) with the remaining countries having 95% or higher access. The internet use rate was the highest in Malaysia (97.4%), followed by the Republic of Korea (97.2%), Australia (96.2%), and Japan (82.9%).

SPIRIT Approach for the Digital Era

Setting and Sustainability (S)

Digital healthy cities were evaluated in terms of setting and sustainability. The findings indicated that Australia, Cambodia, and the Republic of Korea must ensure sustainability through telemedicine, while China should establish a digital healthy city database by integrating the public's electronic medical records. Malaysia, Mongolia, and the Philippines need to sustain their digital healthy cities by establishing digital platform settings for communities (Table 2).

Table . Setting and sustainability (S) of the SPIRIT[a] checklist by country.

Country	Content
Australia	• Sustainability through telemedicine and digital health networking • Sustainable digital healthy city through digital medical infrastructure
Cambodia	• Provide high-quality medical service by addressing the digital medical disparity • Expand sustainability by providing stable high-speed internet
China	• Digital health database that integrates the public's electronic medical records • Health application, wearable, and digital health education
Japan	• Sustainable urban planning through exemplary cases of digital healthy cities • Policy decisions based on integration, education, and collaboration-focused data
The Republic of Korea	• Establish a telemedicine system by digitalizing the existing health care facilities • Establish systems to protect data and prevent leakage of personal information
Malaysia	• Digital healthy city approach through digital technology and public health collaboration • Secure data by strategically placing smart kiosks in the communities
Mongolia	• Establish a digital healthy city hub through a digital platform • Expand sustainability through digital health experts and digital coordinators
The Philippines	• Digital healthy city with community participation • Ensure sustainability by providing stable high-speed internet
Vietnam	• Establish a sustainable digital healthy city by installing data protection measures • Evidence-based digitally healthy city powered by research

[a]SPIRIT: setting approach and sustainability; political commitment, policy, and community participation; information and innovation; resources and research; infrastructure and intersectoral; and training.

Political Commitment, Policy, and Community Participation (P)

We examined political commitment, policy, and community participation in the context of digital healthy cities across 9 countries. The findings emphasized interdepartmental collaboration involving private organizations, schools, and civil institutions. Cambodia, China, and the Philippines referred to building digital healthy cities through community participation via digital platforms. Additionally, the Republic of Korea is implementing digital healthy cities by disclosing health-related public data (Table 3).

XSL•FO
RenderX

Table . Political commitment, policy, and participation (P) of the SPIRIT[a] checklist by country.

Country	Content
Australia	• Health in all policies (health care, environment, and education) • Interdepartmental collaboration • Competency skills training • Data protection • Monitoring
Cambodia	• Digital leadership • Interdepartmental collaboration • Strengthen competency for digital health • Community involvement (digital platform) • Partnership with the technology industry
China	• Community involvement (digital platform) • Intensive training of promising talents in each department • Centralized and comprehensive policy approach through digital healthy cities • Comprehensive and encompassing digital healthy city framework
Japan	• Health in All Policy approach • Digital healthy city policies centered on transportation, education, housing, and digital infrastructure • All departments demand to receive a health impact evaluation • Health in All Policy approach by forming an interdepartmental committee
The Republic of Korea	• Dedicated task force team for Health in All Policy approach • Open data policy to allow access to health-related data • Digital healthy city education for all personnel in charge of health-related matters in all departments • Accelerate the transition to a healthy city by installing digital devices in communities
Malaysia	• Establish digital health laboratories to make effective policies • Share policies among various departments and set common health goals • Devise digital healthy city policies to preserve the environment
Mongolia	• Integration of various areas in digital healthy city policies • Bolster responsibility of digitally healthy city policy makers • Build public-private partnership
The Philippines	• Propose digital healthy city policies through interspecialty collaboration, including health, transportation, education, environment, and information technology • Digital healthy city through digital health platforms
Vietnam	• Digital healthy city policy decisions based on data • Feedback from digital healthy city experts every 2 years

[a]SPIRIT: setting approach and sustainability; political commitment, policy, and community participation; information and innovation; resources and research; infrastructure and intersectoral; and training.

Information and Innovation (I)

We analyzed digital healthy cities in terms of information and innovation. Australia has approached this from a data-centralization perspective. Cambodia has provided telemedicine services through digital health kiosks. China is using digital mental health platforms such as lifestyle fitness apps, green technology, and AI diagnostic tools. Japan is enhancing the digital-use competencies of older adults in their aged society and is planning innovative digital healthy cities based on their needs, such as promoting digital detox weekend campaigns. The Republic of Korea is realizing digital healthy cities with AI platforms for mental health counseling and integrated electronic medical record systems. Malaysia is installing digital kiosks where community residents can obtain instant health information, training community health professionals for this purpose. Mongolia has recognized the need to build digital healthy cities for nomadic communities by facilitating mobile health and telemedicine participation. The Philippines provides telemedicine consultations through the digital health access drive and aims to develop digital healthy

XSL•FO
RenderX

cities based on health care big data. Vietnam is developing digital healthy cities based on data and the digital city profile (Table 4).

Table . Information and innovation (I) of the SPIRIT[a] checklist by country.

Country	Content
Australia	• Address urban noncommunicable diseases and mental health deterioration • Develop public parks and bike lanes • Require collaboration among various departments • Digital healthy city through centralization (eg, data centralization)
Cambodia	• Telemedicine service by implementing digital health kiosks • Require increasing awareness for a healthy lifestyle • Promote active participation of community members and analyze feedback • Improve urban living environment through public-private collaboration
China	• Digital mental health platform • Lifestyle fitness app • Green technology spaces • Data-based AI[b] diagnostic tool
Japan	• Strengthen older adults' digital competency • Innovative digital healthy city planning using digital technology for older adults • Digital detox weekends campaign
The Republic of Korea	• Digital platform for AI mental health counseling • Establish a real-time digital monitoring system for respiratory disorders • Comprehensive electronic medical record system • Digital program for walking through collaboration between Ministry of Health and Welfare and the Ministry of Land, Infrastructure, and Transport
Malaysia	• Immediately acquire health information by installing digital kiosks in public areas • Digital health education through community health care professionals • Expand digital health by integrating nonhealth sectors, such as education and entertainment
Mongolia	• Develop a digital healthy city system for nomadic communities • Reduce health disparity by promoting mobile health and telemedicine for nomads • Collaboration between the green healthy city and digital healthy city • Develop mobile health care facilities
The Philippines	• Provide telemedicine counseling and health education through the digital health access drive • Develop and analyze the digital city profile • Develop a digital healthy city based on health care big data
Vietnam	• Inspect walking facilities through Geographic Information System • Digital healthy city approach • Digital healthy city based on the digital city profile

[a]SPIRIT: setting approach and sustainability; political commitment, policy, and community participation; information and innovation; resources and research; infrastructure and intersectoral; and training.

[b]AI: artificial intelligence.

Resources and Research (R)

We examined the funding and research for digitally healthy cities. Australia is analyzing the digital health gap and planning research on digital healthy cities in cooperation with the Australian Research Council. Cambodia evaluated its digital infrastructure and defined key performance indicators by mapping health priorities. China monitors health burden

XSL•FO
RenderX

assessments and conducts research to enhance digital literacy while also studying the environment, policy, and regulations for digital healthy cities. Japan approaches digital healthy cities by analyzing digital health accessibility and identifying gaps. Additionally, they are conducting research on telemedicine platforms. The Republic of Korea aims to expand its digital healthy cities through the Korea Healthy City Partnership. The partnership seeks to develop key performance indicators and establish shared objectives for digital healthy cities. Mongolia is developing digital healthy cities with privacy-protected digital health regulations (Table 5).

Table . Resources and research (R) of the SPIRIT[a] checklist by country.

Country	Content
Australia	• Establish priority goals for a digital healthy city • Analyze digital divide • Health outcome metrics • Joint resources investment by public and private sectors • Research on digital healthy city through collaboration with the Australian Research Council
Cambodia	• Networks of major stakeholders (eg, leaders, health care professionals, urban planning, and companies) • Evaluate digital infrastructure (eg, internet, digital device use, and platform) • Health priority mapping • Define key performance indicators for a digitally healthy city
China	• Health burden evaluation monitoring • Research to improve digital literacy • Foster a digital healthy city environment through sociocultural evaluation • Research on policies and regulations for a digital healthy city
Japan	• Develop a framework to evaluate the need for a digital healthy city • Analyze digital health access among various demographic groups • Analyze health disparities for a digital healthy city • Research on artificial intelligence diagnosis and telemedicine platforms
The Republic of Korea	• Expand digital healthy city through the Korea Healthy Cities Partnership • Establish and share a digital healthy city database among cities • Establish plans based on the latest literature and information on digital healthy city
Malaysia	• Develop health initiatives to develop regions with poor digital infrastructure • Analyze digital health disparities and implement a digital healthy city focused on vulnerable regions and groups • Develop key performance indicators and establish common digital healthy city goals
Mongolia	• Understand Mongolia's legal digital health regulations and develop a digital healthy city that protects personal information
The Philippines	• Foster experts through focus group interviews with various stakeholders • Understand the digital health policy environment and innovate regulations • Develop key performance indicators and establish common digital healthy city goals
Vietnam	• Appropriate allocation of resources and funds through financial analysis, technical analysis, and risk assessment of digital healthy cities

[a]SPIRIT: setting approach and sustainability; political commitment, policy, and community participation; information and innovation; resources and research; infrastructure and intersectoral; and training.

XSL•FO
RenderX

Infrastructure and Intersectoral (I)

We examined the infrastructure and intersectoral measures for promoting digital healthy cities. Australia is strengthening the functions and capabilities of departments associated with digital healthy cities, requiring regular workshops for knowledge sharing. Cambodia and China are focusing on enhancing infrastructure through digital healthy city operation committees and collaboration with the Ministry of Health and Ministry of Urban Planning as well as other city administration departments. Japan is establishing a digital healthy city committee and uses local media and digital platforms for communication. The Republic of Korea is emphasizing building infrastructure for digital health companies and networking through digital healthy city conferences. The Philippines is developing infrastructure involving urban information and communication technology developers and coordinators to build networks of environmental experts (Table 6).

Table . Infrastructure and intersectoral (I) of the SPIRIT[a] checklist by country.

Country	Content
Australia	• Need to strengthen the functions and competencies of departments linked to digital healthy cities • Sharing of knowledge and skills through periodic workshops • Campaigning and promotions through online platforms and the internet
Cambodia	• Bolster infrastructure through the activities of the digital healthy city operation committee • Cambodia digital healthy city initiative operation committee needs to collaborate with the Ministry of Health, Ministry of Urban Planning, and city administrative bodies
China	• Bolster infrastructure through the activities of the digital healthy city operation committee • Collaboration among public agencies (Ministry of Health), private sector (information technology companies), hospital personnel (physicians and traditional Chinese medicine practitioners), academia (college faculty), and civil society (health nongovernmental organization)
Japan	• Establish a digital healthy city committee with the Ministry of Health and local government leaders • Communication through local media and digital platforms
The Republic of Korea	• Establish various digital health company infrastructure around artificial intelligence conglomerates • Expand the network around the Korea Healthy Cities Partnership
Malaysia	• Establish a department dedicated to technological advancements for digital healthy cities • Establish a digital healthy city network by hosting various digital healthy city conferences
Mongolia	• Launch Mongolia digital healthy city partnerships • Consider the expertise of the digital healthy city policy strategy team, operation, and research teams
The Philippines	• Develop infrastructure involving information and communication technology developers and development coordinators • Establish a network involving environmental experts to develop digital healthy cities
Vietnam	• Establish and execute sustainable digital healthy city plans by developing professional project teams for a digital healthy city

[a]SPIRIT: setting approach and sustainability; political commitment, policy, and community participation; information and innovation; resources and research; infrastructure and intersectoral; and training.

Training (T)

We analyzed training initiatives for realizing digital healthy cities in each country. Australia regularly engages in education and research through exchanges with academic institutions. Cambodia requires settings, such as a digital healthy city initiative conference and the Global Health Summit (Table 7).

XSL•FO
RenderX

Table . Training (T) of the SPIRIT[a] checklist by country.

Country	Content
Australia	• Educate digital healthy city-related personnel through periodic workshops • Education and research through exchanges with academic institutions
Cambodia	• Cambodia digital healthy city initiative conference • Need digital healthy city Asia conference • Global health summit setting
China	• Healthy China 2030 conference • Host a digital healthy city forum • Smart and healthy city exhibition
Japan	• Host conference with the involvement of Japan Smart Community Alliance • Education using digital kiosks • Education using community fitness apps
The Republic of Korea	• Provide public health promotion workshops focused on digital health • Expand digital healthy city through the Korea digital health association • Organize digital health master class
Malaysia	• Host Digital Health Malaysia to invite healthy cities in nearby countries for a joint education on digital healthy city planning and directions • Host smart city conference
Mongolia	• Education and training on digital innovation in health care • Organize Mongolian Smart City Association
The Philippines	• Host an international digital healthy city forum through the Digital Health Philippines Summit • Need collaboration through digital healthy city sister city agreements
Vietnam	• Actively promote digital healthy cities through social media • Research on health technologies and collaboration

[a]SPIRIT: setting approach and sustainability, political commitment, policy, and community participation, information and innovation, resources and research, infrastructure and intersectoral, and training.

Discussion

Principal Results

This study analyzed the digital capabilities of AFHC member countries amid the ongoing transition toward a digital era using the AFHC SPIRIT checklist with the *openai* library in Python and the ChatGPT (version 4.0) large language model. The following implications were obtained.

First, it is necessary to prepare for digital healthy cities by forming digital healthy city networks and committees to promote sustainability and infrastructure. Specifically, an interdepartmental committee should be established to facilitate collaboration among health city-related departments. Further, a holistic "Healthy in All Policy during digital era" approach should be adopted. The AFHC should enhance intercity networking in the digital era during its biennial general meetings. Additionally, it is important to create a framework based on an in-depth understanding of social dynamics, policy structures, governance networks, and urban infrastructure [20].

Second, as digital healthy cities increasingly rely on data-driven policies, the importance of data protection and personal information security has been highlighted. With the volume of daily generated data expanding worldwide, the era of big data is imminent. In such an era, personal information protection is a significant social issue [21]. This study found that data-based policy decisions are crucial for digital healthy cities and requires reviewing and analyzing various data types. The increasing use of community residents' data necessitates diverse protective measures, such as legal regulations and training for responsible personnel. For instance, some healthy cities have adopted AI care robots to provide home-based care services for older adults. As all information about the client is collected through these care robots, heightened precautions must be taken to protect personal information [22].

Third, there is a need for cooperation between various organizations and sectors, such as public agencies, industries, schools, hospitals, and private organizations, to build a digital healthy city network and implement intersectoral healthy city policies. Such cooperation is crucial to promote digital healthy

XSL•FO
RenderX

cities. It is particularly important to establish an infrastructure for community health management by linking digital information technology and AI-related industries with school research institutions and including public institutions, hospitals, and private organizations. Developing a shared platform across these sectors is essential to systematically implement digital healthy cities initiatives [7].

Fourth, there is a need for digital healthy city policies tailored to each country's unique digital capabilities, characteristics, policies, and cultures. The WHO reviewed the focus group analyses of healthy cities worldwide and emphasized the importance of developing appropriate policies for each country and city, defining this as the "glocal approach" [23]. With advancements in digital technology, there is a growing focus on personalized medicine and public health. Accordingly, digital healthy city policies must be customized, considering the diverse characteristics of each country.

Finally, this study highlights the importance of education for key groups in digital healthy cities—policy makers, civil servants, health care providers, and residents. As digital transformation gains traction worldwide, AFHC member countries are also witnessing accelerated digitalization. However, the digital divide between cities, genders, and socioeconomic classes in these countries is widening [24]. The WHO cautions that, without adequate digital literacy among managers and users, the efficacy of digital technologies remains limited even after commercialization [10]. Telemedicine, digital health kiosks, and data management systems are key components of digital healthy cities. To use these services as intended, competency education is essential. Most importantly, digital literacy education for older adults is crucial, given the rapid pace of population aging [25,26].

Limitations

This study has a few limitations. First, as ChatGPT (version 4.0) is trained only until April 2023, the latest developments were not reflected in our survey. Second, the number of healthy cities varied across countries. Third, the SPIRIT checklist was interpreted by only 3 experts. While objective assessments were performed based on their expertise, the number of experts may not have been sufficient for subjective results. Thus, future studies should involve at least 5 experts. Fourth, owing to the proprietary nature of OpenAI's ChatGPT large language model, its full workings and application in this study could not be completely explained. Additionally, the differing levels of political commitment and resources among countries complicate comparisons. Wealthier nations such as South Korea and Japan invest heavily in digital health, while countries such as Cambodia and Mongolia focus on basic health care needs. Tailored strategies are needed to account for these economic and health care disparities, and future studies should consider long-term progress in countries with less developed infrastructures.

Conclusions

This study analyzed the digital health city capabilities of 9 AFHC member countries in the digital era. The findings suggested that these countries must prepare to attain sustainable digital healthy cities by establishing networks and committees. Additionally, digital healthy cities develop big data-driven policies, and this requires systems to ensure personal information protection. This study also emphasized the need for collaborative initiatives across various sectors to systematically implement digital healthy cities. Finally, digital healthy cities are expected to be less effective even with good policies and environments if users have poor digital literacy; thus, policies for improving digital literacy among both managers and users will be required.

Acknowledgments

We would like to thank the stakeholders of Alliance for Healthy Cities countries for their contributions to digital healthy cities. This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1C1C2005464).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

References

1. World Health Organization. Revised list of indicators and procedure for monitoring progress towards health for all in the European region (1987–1988). : WHO European Bureau; 1989 URL: https://iris.who.int/bitstream/handle/10665/355931/39wd04e-HFATgtsUpdat-890007.pdf?sequence=1&isAllowed=y [accessed 2025-01-28]
2. Ashton J, Grey P, Barnard K. Healthy cities — WHO's new Public Health initiative. Health Promot Int 1986;1(3):319-324. [doi: 10.1093/heapro/1.3.319]
3. WHO European healthy cities network. World Health Organization. 2023. URL: https://www.who.int/europe/groups/who-european-healthy-cities-network [accessed 2024-05-03]

XSL•FO
RenderX

4. Nakamura K. A network of healthy cities in asia and the pacific: the alliance for healthy cities. In: Muto T, Nakahara T, Nam EW, editors. Asian Perspectives and Evidence on Health Promotion and Education: Springer Japan; 2011:155-161. [doi: 10.1007/978-4-431-53889-9_15]
5. Raea SM, Alsulami ANM, Aljuaid TH, et al. Designing healthy cities: linking infrastructure to quality of life. Int J of Health Sci 2020;4(S1):509-529. [doi: 10.53730/ijhs.v4nS1.15432]
6. Duhl LJ. The healthy city: its function and its future. Health Promot Int 1986;1(1):55-60. [doi: 10.1093/heapro/1.1.55]
7. de Leeuw E. Evidence for healthy cities: reflections on practice, method and theory. Health Promot Int 2009 Nov;24 Suppl 1:i19-i36. [doi: 10.1093/heapro/dap052] [Medline: 19914985]
8. Acuto M, Morissette M, Tsouros A. City diplomacy: towards more strategic networking? Learning with WHO Healthy Cities. Global Policy 2017 Feb;8(1):14-22 [FREE Full text] [doi: 10.1111/1758-5899.12382]
9. Cocchia A. Smart and digital city: a systematic literature review. In: Dameri RP, Rosenthal-Sabroux C, editors. Smart City: Springer International Publishing; 2014:13-43. [doi: 10.1007/978-3-319-06160-3_2]
10. Rudnicka E, Napierała P, Podfigurna A, Męczekalski B, Smolarczyk R, Grymowicz M. The World Health Organization (WHO) approach to healthy ageing. Maturitas 2020 Sep;139:6-11. [doi: 10.1016/j.maturitas.2020.05.018] [Medline: 32747042]
11. Lapão LV, Correia JC, Jevtic M. Public health framework for smart cities within the comprehensive approach to sustainability in Europe: case study of diabetes. Sustainability 2023;15(5):4269. [doi: 10.3390/su15054269]
12. Is the Eastern Mediterranean region ready for digitalizing health? implications from the global strategy on digital health. : World Health Organization; 2021 URL: https://www.slideshare.net/WHO-EMRO/is-the-eastern-mediterranean-region-ready-for-digitalizing-health [accessed 2025-01-28]
13. Lin Y, Shen Z. Evaluation of urban health efficiency in japanese cities using network data envelopment analysis. J Urban Plann Dev 2023 Jun;149(2):04023011. [doi: 10.1061/JUPDDM.UPENG-4069]
14. Mouton M, Ducey A, Green J, et al. Towards “smart cities” as “healthy cities”: health equity in a digital age. Can J Public Health 2019 Jun;110(3):331-334. [doi: 10.17269/s41997-019-00177-5] [Medline: 30701413]
15. World. ITU. URL: https://datahub.itu.int/data/ [accessed 2024-05-04]
16. Eriksson C. Learning and knowledge-production for public health: a review of approaches to evidence-based public health. Scand J Public Health 2000 Dec;28(4):298-308. [Medline: 11228118]
17. Nam EW, Moon JY, Lee A. Evaluation of Healthy City Project using SPIRIT checklist: Wonju city case. Korean J Health Educ Promot 2010;27:15-25 [FREE Full text] [doi: G704-000885.2010.27.5.001]
18. 2021 SPIRIT template. Alliance for Healthy Cities. 2021. URL: https://www.alliance-healthycities.com/docs/AFHCAward2021/AFHC%20Awards%202021_SPIRIT_TEMPLATE.docx [accessed 2024-05-12]
19. Members. Alliance for Healthy Cities. URL: https://www.alliance-healthycities.com/htmls/members/index_members.html [accessed 2024-03-03]
20. Data bank. World Bank Group. URL: https://databank.worldbank.org/ [accessed 2025-01-14]
21. Ramaswami A, Russell AG, Culligan PJ, Sharma KR, Kumar E. Meta-principles for developing smart, sustainable, and healthy cities. Science 2016 May 20;352(6288):940-943. [doi: 10.1126/science.aaf7160] [Medline: 27199418]
22. Zou H. Protection of personal information security in the age of big data. In: 12th International Conference on Computational Intelligence and Security (CIS): IEEE Publications; 2016:586-589. [doi: 10.1109/CIS.2016.0142]
23. Lee H, Chung MA, Kim H, Nam EW. The effect of cognitive function health care using artificial intelligence robots for older adults: systematic review and meta-analysis. JMIR Aging 2022 Jun 28;5(2):e38896. [doi: 10.2196/38896] [Medline: 35672268]
24. de Leeuw E. Global and local (glocal) health: the WHO Healthy Cities programme. Global Change and Human Health 2001;2(1):34-45. [doi: 10.1023/A:1011991014805]
25. Longoria IAI, Bustamante-Bello R, Ramírez-Montoya MS, Molina A. Systematic mapping of digital gap and gender, age, ethnicity, or disability. Sustainability 2022;14(3):1297. [doi: 10.3390/su14031297]
26. Lee H, Lim JA, Nam HK. Effect of a digital literacy program on older adults’ digital social behavior: a quasi-experimental study. Int J Environ Res Public Health 2022 Sep 29;19(19):12404. [doi: 10.3390/ijerph191912404] [Medline: 36231707]

Abbreviations

AFHC: Alliance for Healthy Cities
AI: artificial intelligence
IRB: institutional review board
SPIRIT: setting approach and sustainability; political commitment, policy, and community participation; information and innovation; resources and research; infrastructure and intersectoral; and training
WHO: World Health Organization

XSL•FO
RenderX

Edited by A Mavragani; submitted 05.06.24; this is a non–peer-reviewed article;accepted 05.12.24; published 07.02.25.

Please cite as:
Lee
Assessment of Digital Capabilities by 9 Countries in the Alliance for Healthy Cities Using AI: Cross-Sectional Analysis
JMIR Form Res 2025;9:e62935
URL: https://formative.jmir.org/2025/1/e62935
doi:10.2196/62935

© Hocheol Lee Originally published in JMIR Formative Research (https://formative.jmir.org), 7.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Web-Based Resource Informed by Cognitive Behavioral Therapy and Positive Psychology to Address Stress, Negative Affect, and Problematic Alcohol Use: A Usability and Descriptive Study

Ingrid Serck-Hanssen[1], MS, MBA; Marit Solheim-Witt[2], MA; Justin J Anker[3], PhD; Dawn E Sugarman[4,5], PhD

[1]Department of Clinical Counseling, Alpha Element Institute, LLC, 6160 Summit Dr N, Suite 200, Minneapolis, MN, United States

[2]Department of Holistic Health, Alpha Element Institute, LLC, Minneapolis, MN, United States

[3]Psychiatry and Behavioral Sciences, University of Minnesota, Minneapolis, MN, United States

[4]Division of Alcohol, Drugs, and Addiction, McLean Hospital, Belmont, MA, United States

[5]Department of Psychiatry, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Ingrid Serck-Hanssen, MS, MBA
Department of Clinical Counseling, Alpha Element Institute, LLC, 6160 Summit Dr N, Suite 200, Minneapolis, MN, United States

Abstract

Background: Research documents that drinking to cope behavior can be disrupted by enhancing emotion regulation and coping skills related to the experience of stress and negative affect. The Alpha Element Self-Coaching Plan incorporates principles of positive psychology and cognitive behavioral therapy to redirect negative thinking and emotions and, therefore, has the potential to benefit individuals who use alcohol to cope with stress.

Objective: This study aimed to evaluate satisfaction and usability of the web-based Alpha Element Self-Coaching Plan in order to inform the development of an expanded digital platform based on the Alpha Element framework.

Methods: Participants enrolled in the web-based program as part of their clinical care were eligible to participate. A total of 20 individuals (14 women and 6 men) between ages 30 and 79 (mean 54.5, SD 14.14) years completed web-based questionnaires to assess product performance in areas such as ease of technology use, quality of videos and handouts, and the value of the activities. Participants also completed the System Usability Scale (SUS) and provided background and demographic information, including alcohol use.

Results: Only 1 participant reported no alcohol use in the past year; 55% (11/20) of participants drank alcohol 2 - 4 times per month or less and 45% (9/20) reported drinking alcohol 2 - 3 times per week or more. The average SUS score of 76.38 (SD 17.85) was well above the commonly accepted threshold of 68, indicating high system usability. A majority of the sample (16/19, 84%) agreed or strongly agreed that the activities in the program inspired behavioral changes; and most agreed or strongly agreed that the program was engaging (16/20, 80%), well-organized (18/20, 90%), and easy to follow (17/20, 85%). Only 2 participants endorsed experiencing difficulty using the program on a smartphone. Suggestions for program improvements included expanding the platform, updating the web format, adding user interactivity, and enhancing navigation.

Conclusions: These data suggest that participants were generally satisfied with the web-based Alpha Element Self-Coaching Plan, and rated usability of the program as favorable. Importantly, a significant portion of participants reported that the program inspired behavioral changes. More research is needed with a larger sample to obtain specific data about alcohol consumption and investigate associations between alcohol use and program components, as well as examine gender differences. Data collected from this study will be used to expand the platform and improve user experience.

(JMIR Form Res 2025;9:e63819) doi:10.2196/63819

KEYWORDS

alcohol misuse; stress; drinking to cope; DTC; negative affect; positive psychology intervention; PPI; cognitive behavior therapy; CBT; alcohol use; drinking; usability; descriptive study; behavior; emotion; coping skill; positive psychology; psychology; online resource; mobile phone

XSL•FO
RenderX

Introduction

Stress is a well-documented risk factor associated with the initiation, maintenance and relapse of alcohol use disorder [1]. The self-medication model suggests some individuals drink to cope (DTC) with stress due to the belief that drinking will alleviate stress and negative affect [2,3]. However, although alcohol may provide temporary relief from emotional discomfort, in the long-term, drinking to cope with stress exacerbates negative affect and puts individuals at risk for alcohol-related consequences [4].

Positive psychology interventions (PPIs) may be beneficial in disrupting the DTC cycle of alcohol use, as emerging data suggest that the processes underlying PPI may facilitate more adaptive behavioral choices about substance use [5]. PPIs are strength-based approaches that apply individual traits along with positive subjective experiences to promote well-being [5,6]. PPIs aim to foster optimism, psychological well-being, and resilience through practices like cognitive flexibility, positive reframing, and the cultivation of strengths, representing a shift away from focusing solely on reducing symptoms and instead building emotional assets. Furthermore, research indicates that positivity can be learned and lead to increased well-being and decreased alcohol consumption [7]. A wide range of PPIs exist that vary in type, delivery mode, and intensity. PPIs have mostly been studied in nonclinical samples; however, preliminary evidence exists for their efficacy in clinical settings [8]. Research suggests that cognitive changes may explain the positive effect of PPIs on emotions [6]. Therefore, cognitive behavioral therapy (CBT) can be personalized with PPI components that target motivational and emotional factors. With a focus on increasing positive emotions, behaviors, and cognitions, PPI approaches align well with CBT principles and both can be leveraged with the objective to reduce stress with negative affect aimed at mitigating cravings for alcohol to self-medicate.

Critical to the success of CBT-based strategies in the clinical setting is engaging clients in homework exercises [9,10]. Technology-based tools such as smartphone apps and web-based self-management programs provide scalable solutions for individuals to engage with self-directed exercises. Emerging research indicates that web-based interventions can serve a protective function for mental health by mitigating the effects of stress [11]. Ongoing measurement of engagement with web-based homework exercises along with strategic approaches to tailoring intervention content delivery to individuals may help increase sustained practice [12]. The purpose of this quality improvement (QI) study was to evaluate usability and satisfaction of the web-based Alpha Element Self-Coaching Plan (hereafter referred to as the Self-Coaching Plan [13]) applied in a community clinical counseling setting. The Self-Coaching Plan incorporates a combination of CBT and PPI approaches and features content based on a novel positive psychological framework of Vitality Type patterns in behavior and cognition [14-17]. The goal of the Self-Coaching Plan is to empower individuals with tools to discover a new and positive perspective on themselves, along with psychoeducation and personalized skills to proactively manage stress and negative affect, as well as address problematic drinking behavior. The aim of this QI study was to examine satisfaction and usability of the web-based Self-Coaching Plan in order to facilitate ongoing improvement of this tool.

Methods

Study Setting

In this QI study the Self-Coaching Plan was offered as adjunct homework to clients in the clinical setting. Participation in the project was optional and did not affect the therapeutic relationship with their clinical provider. All participants in this study were receiving mental health counseling through telehealth in a community counseling clinic in Minneapolis, Minnesota.

Study Design

Data were gathered for administrative purposes within the context of routine efforts to examine and improve clinical care services and operations. This QI study was conceptualized as a formative evaluation to assess product performance of the Self-Coaching Plan. The evaluation surveys focused on system usability, technology ease of use, quality of videos, and downloadable handouts, as well as the value of the program content and activities.

Participants

Participants were recruited through a self-selection process. We applied an all-comers approach, meaning individuals enrolled in the Self-Coaching Plan as part of their standard care were invited to participate. Of the 35 clients invited to participate, 20 (14 women and 6 men) agreed. To recruit participants, a University of Minnesota REDCap (Research Electronic Data Capture; Vanderbilt University) survey link was included in weekly touch-base emails.

Intervention

A web-based tool for self-directed learning, the Self-Coaching Plan, was applied in this study as a homework resource in the clinical setting. The Self-Coaching Plan is based on the premise that introducing individuals to a framework of 4 Vitality Type patterns as a strategy to disrupt negative self-schemas can help provide new perspectives on existing behavior patterns as well as cognitive resources to reframe distorted thinking that is impairing the regulation of negative affect (Table 1).

XSL•FO
RenderX

Table . The premise of the Self-Coaching Plan is to apply 4 Vitality Type categories to self-identify empowered and disempowered patterns in behavior and cognition. Throughout the Self-Coaching Plan, individuals are reminded that words are cognitive tools that frame our perceptions and shape our reality.

Vitality Type patterns	Catalyst	Torchbearer	Pathfinder	Pragmatist
Empowered behavior patterns	• Discovery • Open-minded • Uplifting • Playful • Social • "Let's try it now."	• Action • Direct • Fast • Results-driven • Bold • "What can I do?"	• Development • Investigative • Process-oriented • Diplomatic • Consensus-building • "I'll think about it."	• Completion • Scrutinizing • Focused • Discerning • Structured • "It has to be done right."
Disempowered behavior patterns	• Self-indulgence • Erratic	• Irritability • Restless	• Avoidance • Overwhelmed	• Perfectionism • Judgmental

The 4 categories, Catalyst Vitality, Torchbearer Vitality, Pathfinder Vitality, and Pragmatist Vitality, represent distinct patterns in outlook and behavior (Table 1). The essence of a catalyst is to spark activity. Positive words associated with the Catalyst Vitality pattern include discovery, open-minded, uplifting, playful, and social. The essence of a torch bearer is to challenge status quo and pave the way to take action. Positive words associated with the Torchbearer Vitality pattern are action, direct, fast, results-driven, and bold. The essence of a pathfinder is to be methodical. Positive words associated with the Pathfinder Vitality pattern include development, investigative, process-oriented, diplomatic, and consensus-building. The essence of a pragmatist is to bring clarity and stability to their environment. Positive words associated with the Pragmatist Vitality pattern are completion, scrutinizing, focused, discerning, and structured. As part of the Self-Coaching Plan, participants study all 4 Vitality Type categories by exploring the words linked to each. Through a series of exercises, they apply these words to personalize their self-coaching strategies. By organizing positive words into categories (Table 1) the Vitality Type patterns combine elements from both CBT and PPI.

Exploring how to break unhealthy habits and cognitive distortions and replacing them with positive self-talk can create shifts in outlook that can lead to changes in behavior. The Vitality Type categories represent a range of possibilities in patterns of cognition and behavior which in the Self-Coaching Plan are applied as cognitive tools to help individuals self-identify and articulate aspects of their empowered and disempowered patterns. Exploring words linked to the Vitality Type patterns and applying these words as cognitive tools to change behavior and outlook, help individuals take back their power instead of leaning on alcohol and other negative coping strategies. It is posited that the framework of 4 Vitality Type categories offers resources to promote positive self-perception to inspire new self-awareness, self-discovery, and development of personalized self-management strategies. Vitality Type patterns can also be applied as a cognitive strategy for considering differences in relationships and interpersonal communication. Managing relationships with self and others are key aspects of managing stress and problematic drinking behavior.

It is proposed that learning about these patterns (schemas) can help individuals orient and learn new terminology to reframe negative schemas they have about themselves. By applying the constructs of 4 Vitality Types as cognitive resources to identify and articulate empowered and disempowered patterns, individuals are introduced to a fresh perspective on themselves along with terminology that can provide a bridge to describe potential movement away from pathology toward hope and well-being. This self-reflective process can lead to a deeper understanding of how behaviors and thoughts impact outlook and overall health. By using the Vitality Type categories as a framework of schemas and tools to foster cognitive flexibility, the Self-Coaching Plan adapts CBT elements and PPI principles as a therapeutic strategy to create synergy and inspire self-investigation. Therefore, the terminology of the 4 Vitality Type categories are applied to all activities to provide a map for organizing and personalizing CBT-PPI strategies.

In this QI study, the objective of the Self-Coaching Plan was for clients, between sessions, to explore and process patterns of disempowerment and negative outlook and learn skills to redirect and self-regulate. This web-based program offers psychoeducation and activities through videos and visual graphics; topics include how to apply the Vitality Type patterns to foster self-compassion, growth mindset, and preparing for positive change as well as facts about gender-specific vulnerabilities and substance use. During the Self-Coaching Plan, individuals are encouraged to use the Vitality Type patterns as tools to help achieve their goals and focus on what they are for—as a shift away from focusing on what they are against and working to change. Self-Coaching Plan activities are organized into sections with the objective of completing an action plan to stay motivated on their journey toward positive change (Table 2). It is recommended that the proposed order of activities is followed, however, individuals are free to choose the order they prefer. The goal is for individuals to apply the Vitality Type framework to gain a new perspective on predominant go-to behavior patterns and understand the importance of developing their less dominant patterns to self-advocate, set boundaries, and feel more empowered. The Self-Coaching Plan is designed to help individuals get prepared and stay motivated to achieve their goals over an 8-week period. They learn to apply words as cognitive tools to change their

XSL•FO
RenderX

behavior and take their power back instead of leaning on alcohol and other negative coping strategies. The Self-Coaching Plan can be applied as both a guided and unguided protocol.

XSL•FO
RenderX

Table . A web-based tool for self-directed learning, the Self-Coaching Plan, was applied in this study as a homework resource in the clinical setting. The goal was to empower individuals through a personalized process of activities and self-reflection to foster healthy habits, learn to set boundaries, manage emotions, and change unhealthy habits such as problematic drinking behavior.

Module name	Description	Components
Growth Mindset	Psychoeducation about how our abilities and frame of mind can be developed through dedication and hard work. A growth mindset is knowing that it is never too late to make a change.	• Welcome video. • Course intro. • Psychoeducation about growth mindset and self-compassion.
Get Ready for Positive Change	Psychoeducation to get clarity about priorities and steps to achieve positive lifestyle changes.	• Course objectives. • Gender specific vulnerabilities for substance use. • Substance use and drug facts.
Vitality Type Indicator	The objective of this self-assessment is to gain awareness of empowered and disempowered patterns in communication and relationships.	• Self-assessment objective. • Connection between stress management, outlook, and drinking behavior.
Vitality Type Patterns	Psychoeducation about how managing the relationship with self and others is key to managing the relationship with alcohol.	• Introducing the 4 Vitality Types as tools to self-manage and lay the foundation for developing a growth mindset. • Vitality Type matrix of empowered and disempowered patterns.
Top Priority	Identify the most important area of desired change. Individuals are encouraged to look at themselves as a priority.	• Exploring boundaries required for positive change. • Self-care as a means to manage stress. • Connection between mind, body and soul in healing.
Self-Coaching Guide	The Self-Coaching Guide is the Vitality Type of the highest score on the Vitality Type Indicator self-assessment.	• The approach of the Self-Coaching Guide is key to achieving our Top Priority. • Self-Coaching exercise applies the highest scoring Vitality Type pattern as a technique and practice to reflect on goals, challenges, and personal growth. • Exercise to create positive “I am” statements using words from the word-bank associated with their highest self-assessed Vitality Type score.
Hidden Resource	The Hidden Resource is the Vitality Type of the lowest score on the Vitality Type Indicator self-assessment.	• The approach of the Hidden Resource indicates untapped potential that can help on the path to sobriety. • Exercise to redirect thoughts by applying the lowest scoring Vitality Type as a strategy to address challenges and difficult situations in a new way. • Reflection exercise to discover how to leverage aspects of the Hidden Resource approach as support during the healing journey.
Shadow Aspects	Shadow Aspects represent disempowered behaviors taken on during times of stress or fear.	• Exercise to self-identify behaviors that are barriers in communication and relationships. • Reflection activity about self-compassion and growth mindset as strategies to help turn around a disempowered state.

XSL•FO
RenderX

Module name	Description	Components
Power Words	Power Words are positive characteristics and behaviors linked to each Vitality Type.	• Activities to choose Power Words, which are then applied to personalize strategies to help disrupt Shadow Aspects. • Exercise to identify 1 Power Word linked to the Self-Coaching Guide and another word linked to the Hidden Resource as strategy to stay proactive rather than reactive in stressful situations. • Reflection activity to become more aware of how to deal with joy and conflict can be key to understanding the relationship with alcohol.
Self-Coaching Plan Commitment Sheet	Course summary: maintaining sobriety and a sober-minded lifestyle is much more manageable when the focus is on empowerment, growth mindset and self-compassion.	• Resources to explore how to continue using the Vitality Types to stay committed to healing, health and wellness.

As part of the Self-Coaching Plan, individuals take a self-assessment (Vitality Type Indicator, Table 2) that sorts responses according to the categories of the 4 Vitality Type categories. The results of the self-assessment provide a map for individuals to discover a range of Vitality Type patterns in their behavior, cognition, and communication. Individuals' highest Vitality Type score indicates a predominant pattern and is identified as their Self-Coaching Guide. An individual's lowest Vitality Type score indicates untapped potential and is identified as their Hidden Resource. Attributes linked to an individual's Hidden Resource represent capabilities outside of the person's typical patterns of behavior. The concepts of Self-Coaching Guide and Hidden Resource are incorporated into the program as cognitive tools and a positive approach to develop personalized strategies for stress-management and self-regulation. Individuals are prompted to practice their Hidden Resource capabilities in order to prepare themselves to proactively handle challenges and difficult situations with a new perspective. This information is applied in the Self-Coaching Plan as CBT-PPI informed reflective learning activities and homework assignments.

Through a series of activities of using words as cognitive resources to make positive shifts, individuals practice using specific words that represent empowered behavior to formulate their desired change and develop healthy habits. A word bank is linked to each of the 4 Vitality Type categories as a tool to discover a range of possibilities. Individuals are asked to reflect on which words and aspects of the 4 patterns resonate with their view of themselves, and which do not. In a self-coaching module, individuals explore how to use their Self-Coaching Guide and their Hidden Resource (Table 2) to manage stress and move toward sobriety and a healthier, more balanced life. Table 1 presents an overview of the 4 categories and illustrates how each proactive (empowered) Vitality Type pattern has a reactive (disempowered) opposite. In the Self-Coaching Plan (Table 2), the disempowered opposite is referred to as a Shadow Aspect. Shadow Aspects represent disempowered behaviors taken on during times of stress or fear. To dig deeper into understanding their disempowered reactions, participants are asked to self-identify 2 primary Shadow Aspects. To redirect disempowerment, individuals then choose 2 words from the word bank they consider to be empowering. These words are referred to as Power Words. Power Words are positive characteristics and behaviors linked to each Vitality Type which are applied to develop their personalized strategies to redirect. During this CBT-PPI process of applying Vitality Type categories in self-discovery and reflective learning, individuals gain a newfound perspective on their go-to behavior patterns along with personalized strategies to use their less dominant patterns as a resource to set boundaries, self-advocate, and feel more empowered.

Procedures

A survey was offered through a REDCap link as part of touch-base emails frequently sent out to all registered users of the Self-Coaching Plan. We informed all users that the goal of the project was quality improvement and their answers would be used to improve user experience. Users were not able to proceed to the survey unless they had checked the box indicating they agreed to participate. Participants completed a custom questionnaire, the System Usability Scale (SUS), as well as background and demographic information, including alcohol use. The custom questionnaire was designed to evaluate (1) the relevance and ease-of use of the Self-Coaching Plan material and (2) satisfaction with the Self-Coaching Plan.

Measures

The SUS is a 10-item scale designed to provide a global view of subjective assessments of usability. The scale covers a variety of aspects of usability, including the need for support and training while using the system as well as its level of complexity. The SUS evaluation instrument yields a single number that represents a composite measure of the overall usability of the system being studied with current literature suggesting that a score of 68 as a useful benchmark (mean SUS score) [18].

The custom questionnaire was developed to gather feedback about program engagement along with dimensions of perceived interactivity and ease of use. The questions were developed by creators of the Self-Coaching Plan and other subject matter experts. The data were collected to establish a baseline along key dimensions against which change over time can be measured. Key dimensions measured included strengths and

XSL•FO
RenderX

limitations of the technology, videos, and exercises; need for tech support, notifications, and reminders; program navigation and logistics features; time commitment; aesthetics and visual elements; as well as suggestions for improvement. Specific questions were asked about engagement, self-perception and behavior change as well as knowledge about key terms and concepts, such as Vitality Types, Self-Coaching Guide, and Hidden Resource. Each item was rated on a 5-point Likert scale ranging from strongly disagree to strongly agree with a midpoint of neither agree nor disagree. Recommendations for enhancing program improvement were obtained through optional write-in responses to capture comments and suggestions in the language of participants. Alcohol consumption was measured with 1 question about frequency of alcohol use in the past year with the following scale: Never; Monthly or less; 2 - 4 times a month; 2 - 3 times a week; 4 or more times.

Data Collection

Participants completed the surveys through REDCap. REDCap is an encrypted, electronic database that is HIPAA (Health Insurance Portability and Accountability Act) compliant. It is approved by University of Minnesota Institutional Review Board for the administration and storage of human subject information.

Statistical Analysis

Analyses for this study focused on evaluating the Self-Coaching Plan using descriptive statistics and were performed using R statistical software (R Foundation for Statistical Computing). Mean and standard deviation were used for continuous variables such as age and SUS scores. For categorical variables, including demographic information (eg, sex, education level, and race) and survey responses (eg, engagement with self-care routine and frequency of alcohol consumption), frequencies were calculated.

Ethical Considerations

The project was determined not to qualify as human subject research by the University of Minnesota Institutional Review Board (STUDY00020268). Confidentiality was strictly protected. No identifying information was collected.

Results

Participant Characteristics

Participants (N=20) were between 30 and 79 years of age (mean 54.5, SD 14.14), predominantly female (n=14, 70%) White (n=15, 75%) and well educated with 35% (n=7) college graduates and 35% (n=7) reporting as graduate level or above. Only 1 participant reported no alcohol use in the past year; 55% (n=11) of the sample drank alcohol 2 - 4 times per month or less and 45% (n=9) reported drinking alcohol 2 - 3 times per week or more.

System Usability

Average SUS score from this study was 76.38 (SD 17.85). This result is well above the commonly accepted threshold of 68. The score indicates high usability of the system and its interfaces.

Custom Questionnaire

As shown in Table 3, a majority (16/19, 84%) of participants agreed or strongly agreed that the activities in the program inspired positive behavioral changes while 90% (18/20) agreed or strongly agreed that participation in the program positively influenced their self-perception. There was strong endorsement of knowledge and understanding about concepts unique to the program such as Vitality Types, Self-Coaching Guide and Hidden Resource. Of note, all participants (n=20, 100%) agreed or strongly agreed that the Vitality Type Indicator self-assessment was a helpful tool in their self-development; 90% (n=18) endorsed the purpose of the Self-Coaching Guide as clear and understandable while 85% (n=17) reported they could relate to their Self-Coaching Guide; 80% (n=16) of participants strongly endorsed understanding the concept of a Hidden Resource. A majority of participants (n=16, 80%) endorsed the program as engaging and 80% (n=16) found the videos were effective in motivating them to complete the homework activities and most participants agreed or strongly agreed that the visual design of the platform was appealing. With regards to technical interface, 80% (n=16) of participants did not experience significant difficulties using the Self-Coaching Plan on smartphones or computers. Generally, participants (n=18, 90%) found the program well-organized.

XSL•FO
RenderX

Table . Results from the quality improvement study (N=20).

Quality improvement survey questions	Strongly agreed or agreed, n (%)	Neither agreed or disagreed, n (%)	Disagreed or strongly disagreed, n (%)
The activities in this program inspired me to change my behavior.[a]	16 (84)	—[b]	3 (16)
Participating in this course positively impacted my self-perception.	18 (90)	2 (10)	—
The order of the activities in the program was logical and helped in my understanding.	18 (90)	1 (5)	1 (5)
The videos were effective in motivating me to complete the activities.	16 (80)	3 (15)	1 (5)
The Vitality Type categories were understandable.	18 (90)	2 (10)	—
The Vitality Type Indicator self-assessment was helpful in my self-development.	20 (100)	—	—
The purpose of the Self-Coaching Guide was clearly explained and understandable.	18 (90)	2 (10)	—
I related to my Self-Coaching Guide.	17 (85)	3 (15)	—
The concept of a Hidden Resource was clear to me.	16 (80)	4 (20)	—
I found the program to be engaging.	16 (80)	2 (10)	2 (10)
The content within the program was well organized.	18 (90)	2 (10)	—
The time required to complete this course felt appropriate to me.	17 (85)	2 (10)	1 (5)

[a]One participant missed this item.

[b]Not endorsed by any participants.

As part of the custom questionnaire, participants were asked to respond to 3 open-ended questions. The first question asked participants to share their thoughts about which program components they considered to be the most effective. Predominant themes that emerged from responses to this question were (1) empowerment, (2) personalization, and (3) engagement. The following comments illustrate a sense of empowerment: “Understanding better how I react to things” and “A clear view of my strengths and weaknesses really helped me choose where I wanted to focus my energy.” The theme personalization was illustrated by the following quotes: “I love that I get to use a personal Self-Coaching Guide based on my behaviors specifically designed for me!” and “I enjoyed exploring the Vitality Types that align with my personality and observing how they influence my daily behavior. After completing the test (Vitality Type self-assessment), I identified which trait was most prominent for me.” Third, the theme engagement, was illustrated by these comments: “I found the homework was very helpful for me to have. It encouraged me to continue my healing in between sessions and then to have a follow-up conversation with my therapist about the homework really helped tie everything together for me.” “The activities were most effective because they made you actually pause and reflect.”

The second open-ended question asked participants to share their views about program components they considered the least effective. Here the predominant theme was navigation, as illustrated by the following comments, “I did find it a little confusing in the beginning to find the correct files at the end of the program,” and “It was hard to find the Vitality Types breakdowns through the library.” The third open-ended question asked for recommendations to enhance the effectiveness of the program. Here feedback again focused on navigation, as the following comments illustrate. “I think the navigation could be cleaned up, but overall, it was an effective program.” “The online interface needs to be updated.” Other themes for program improvements included expanding the platform and adding user interactivity as illustrated by the following comments: “I’d be excited to witness user interaction on the platform. In addition, incorporating a gamification element, perhaps tracking our progress through the program, would be an engaging feature.” Another participant stated, “I hope for a platform where I can engage with others regarding Vitality Types, perhaps in the form of a community for individuals who have undergone this course.”

XSL•FO
RenderX

Discussion

Principal Findings

Findings from this study document high usability and satisfaction with the web-based Self-Coaching Plan in the clinical setting. A significant portion of participants reported that the program inspired behavioral changes as well as positive self-perception and found the program to be well-organized and easy to follow. Most participants (18/20, 90%) reported the Vitality Type categories were understandable, with 85% (17/20) stating they could relate to their Self-Coaching Guide, and 84% (16/19) endorsed that they were inspired by program activities to change their behavior. However, more research is needed to examine the relationship between these constructs, participant self-perception, and behavior change regarding alcohol use. Several participants expressed a desire for improved navigation.

Comparison With Previous Work

An increasing amount of research points to the double jeopardy of stress and negative affect as a detrimental set of vulnerabilities that lead to problematic drinking behavior and DTC [2,19]. This trend is of increasing concern and spotlights the need for research and development of innovative self-management approaches. Data from this QI study suggest that participants were generally satisfied with the Self-Coaching Plan, and rated usability of the program as favorable. The implications from this study support the application of the Vitality Type categories as a CBT-PPI approach for individuals to gain a fresh and more positive self-perception as well as personalized skills to self-regulate. All participants endorsed the Vitality Type Indicator as a helpful tool in their self-development. Through reflection activities, individuals were encouraged to discover a new view of themselves and apply the Vitality Type patterns as cognitive tools to disrupt barriers, reframe, and manage stress.

Findings from this QI study indicate the 4 Vitality Types patterns can be adapted to a web-based program for individuals to orient and reflect on these 4 categories of attributes and explore the possibility of using the terminology to describe aspects of themselves and develop strategies to redirect. Responses to the custom questionnaire document that 90% (18/20) of participants endorsed the purpose of the Self-Coaching Guide as clear and understandable. Furthermore, 85% (17/20) reported they could relate to their Self-Coaching Guide. The majority of participants 80% (16/20) strongly endorsed understanding the concept of a Hidden Resource. It is noteworthy that 90% (18/20) of participants thought the Self-Coaching Plan positively impacted their self-perception.

A focus of CBT approaches is to support clients in their efforts to examine their thoughts and behaviors and become scientific investigators of their own lives [9]. Results from the open-ended questions in the custom questionnaire document that participants applied the self-directed exercises in the Self-Coaching Plan as tools to self-reflect and personalize the content. This finding is underscored by the comment, “The activities were most effective because they made you actually pause and reflect.” Individuals who engage with homework activities are taking a hopeful step by demonstrating the belief that their own efforts will help them overcome their clinical problems [10]. It has been proposed that integrating PPIs into existing treatment approaches may be beneficial because they produce an immediate reward that is nonsubstance-related [20].

Historically, PPI research has focused on individual positive qualities and attributes such as gratitude, optimism, and kindness [21] and not development of novel conceptual frameworks of patterns in behavior and cognition to address the larger context of disempowerment created by negative self-schemas. The Vitality Type categories offer a map of words for individuals to reflect on a range of factors impacting their behavior and communication without focusing on pathology. The Self-Coaching Plan is distinguished from other web-based psychosocial programs by providing a novel conceptual framework of 4 Vitality Type categories as proposed schemas in behavior and cognition. The proposed schematic conceptualizations of distinct patterns of emotion, behavior and cognitions as 4 Vitality Type patterns represent a model that can be applied to expand upon the traditional positive psychology approaches. As a psychosocial approach to self-development, the Self-Coaching Plan represents a new context for self-reflection and an innovative way of looking at differences in relationships and interpersonal communication. The current lack of standardized PPI protocols may limit the implementation of PPIs in the clinical setting and it may be beneficial to develop new, positive-focused interventions for substance use and other conditions instead of simply including traditional PPIs into existing treatments [22]. One study found that addiction counselors were using positive psychology exercises in their sessions over half of the time without classifying their methods as explicitly PPI [23]. To our knowledge, current web-based PPI informed programs do not apply a map of 4 empowered and disempowered patterns in behavior and cognition as schemas to navigate stress and proactively manage negative affect. As documented by the findings in this study, 90% (18/20) of participants endorsed the Self-Coaching Plan as positively impacting their self-perception. These results indicate that awareness of the Vitality Type categories and terminology can be leveraged to help individuals gain a more optimistic view of themselves as well as develop a personalized self-coaching plan to work toward health and well-being.

Data collected from this study will be applied to improve and expand the web-based Self-Coaching Plan. More research is needed with a larger sample to obtain specific data about alcohol consumption and investigate associations between alcohol use and program components. Furthermore, web-based programs that incorporate gender-specific components have been shown to enhance outcomes for women with substance use disorders [24]. Therefore, it is important for future research to examine gender differences, as well as feedback on gender-specific elements of the program.

Limitations

The small number of participants is a limitation in this study. Participants were recruited through a self-selection process which could be a limitation if participants who chose to partake in the survey were more positively biased than nonparticipants.

XSL•FO
RenderX

Another limitation is that drinking to cope was not a criterion for eligibility, thus the intervention was not specifically tested on individuals who drink to cope. Individuals who drink to cope might have provided different data regarding acceptability and usefulness. Hence, future research should focus on testing the intervention in a target population with DTC. In addition, the sample was predominantly White, female, and college-educated. While this demographic distribution is representative of the community where the research was conducted, it limits the wider generalizability of the findings. Future research should seek to gain feedback on usability and satisfaction from a more diverse sample. Of note, recent research suggests alcohol use is increasing more among middle-aged and older adults than among younger drinkers [25], therefore the wide age range in the present sample could be an indication of the program's appeal to this age group.

Conclusions

In this QI study, we examined usability and satisfaction with a web-based Self-Coaching Plan that applies CBT and PPI principles featuring a novel framework of Vitality Type categories to engage clients in a structured self-development process to increase skills to manage stress and negative affect. Findings from this quality improvement study indicate that the Self-Coaching Plan was engaging as well as acceptable to participants, positively influencing their self-perception and behavior. Data from the custom questionnaire document that applying the framework of 4 Vitality Type patterns is an effective approach to empower individuals with information, strategies, and a plan to self-regulate. Taken together, results from this QI study warrant investigating if the Self-Coaching Plan's approach to personalize assets-based self-management strategies targeting stress and negative affect, could help empower individuals with healthy habits to cope with stress instead of turning to alcohol.

Acknowledgments

The study was internally funded by Alpha Element Institute, LLC. Effort for this paper was supported by the National Institute on Drug Abuse K23DA050780 (DES) and UG3DA059414 (JJA).

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

ISH and MSW cowrote the paper, assisted in study conception, development, and qualitative data analysis. JJA led data analysis and assisted in study conception and development. DES conceived of the study, oversaw the project, and edited the paper. All authors reviewed and approved the final paper.

Conflicts of Interest

ISH and MSW are cofounders and co-owners of Alpha Element Institute, LLC, and have equity in the company. DES and JJA report no conflicts of interest.

References

1. Peltier MR, Verplaetse TL, Mineur YS, et al. Sex differences in stress-related alcohol use. Neurobiol Stress 2019 Feb;10:100149. [doi: 10.1016/j.ynstr.2019.100149] [Medline: 30949562]
2. Frone MR. Work stress and alcohol use: developing and testing a biphasic self-medication model. Work Stress 2016;30(4):374-394. [doi: 10.1080/02678373.2016.1252971] [Medline: 28090129]
3. Hawn SE, Cusack SE, Amstadter AB. A systematic review of the self-medication hypothesis in the context of posttraumatic stress disorder and comorbid problematic alcohol use. J Trauma Stress 2020 Oct;33(5):699-708. [doi: 10.1002/jts.22521] [Medline: 32516487]
4. Armeli S, Sullivan TP, Tennen H. Drinking to cope motivation as a prospective predictor of negative affect. J Stud Alcohol Drugs 2015 Jul;76(4):578-584. [doi: 10.15288/jsad.2015.76.578] [Medline: 26098033]
5. Trinh CD, Girard R, Schick MR, Spillane NS. Positive psychological interventions on alcohol use and consequences: pilot randomized trial in a young adult cannabis-using sample. J Subst Use Addict Treat 2024 Mar;158:209241. [doi: 10.1016/j.josat.2023.209241] [Medline: 38056630]
6. Stone BM, Schmidt K. Positive psychological interventions and cognition. J Posit Psychol 2020 Sep 2;15(5):629-632. [doi: 10.1080/17439760.2020.1789714]
7. Akhtar M, Boniwell I. Applying positive psychology to alcohol-misusing adolescents. GPWK 2010;20(3):6-31. [doi: 10.1921/gpwk.v20i3.707]
8. Chakhssi F, Kraiss JT, Sommers-Spijkerman M, Bohlmeijer ET. The effect of positive psychology interventions on well-being and distress in clinical samples with psychiatric or somatic disorders: a systematic review and meta-analysis. BMC Psychiatry 2018 Jun 27;18(1):211. [doi: 10.1186/s12888-018-1739-2] [Medline: 29945603]

XSL•FO
RenderX

9. Padesky CA. Collaboration and guided discovery. Cogn Behav Pract 2022 Aug;29(3):545-548. [doi: 10.1016/j.cbpra.2022.02.003]
10. Strunk DR. Homework. Cogn Behav Pract 2022 Aug;29(3):560-563. [doi: 10.1016/j.cbpra.2022.02.009]
11. Enrique Roig A, Mooney O, Salamanca-Sanabria A, Lee CT, Farrell S, Richards D. Assessing the efficacy and acceptability of a web-based intervention for resilience among college students: pilot randomized controlled trial. JMIR Form Res 2020 Nov 11;4(11):e20167. [doi: 10.2196/20167] [Medline: 33174530]
12. Goldin PR, Lindholm R, Ranta K, Hilgert O, Helteenvuori T, Raevuori A. Feasibility of a therapist-supported, mobile phone-delivered online intervention for depression: longitudinal observational study. JMIR Form Res 2019 Jan 22;3(1):e11509. [doi: 10.2196/11509] [Medline: 30682726]
13. Transform your life with the Alpha Element Self-Coaching Plan. Alpha Element. 2023. URL: https://alphaelement.podia.com/alpha-element-academy [accessed 2025-01-28]
14. Anker JJ, Serck-Hanssen I, Solheim-Witt M, Sugarman DE. Quality improvement of an online cognitive behavioral therapy and positive psychology informed resource to manage alcohol use, stress, and negative affect. In: Alcohol, Clinical and Experimental Research: Volume 48, Issue S1 Supplement: Speaker and Poster Abstracts of the Research Society on Alcohol, 47th Annual Scientific Meeting, June, 2024: Wiley-Blackwell; 2024. URL: https://onlinelibrary.wiley.com/toc/29937175/2024/48/S1 [accessed 2025-01-28]
15. Solheim-Witt M, Serck-Hanssen I. Self-development and interpersonal communication. Presented at: Interprofessional Education Summit: High-Performance Innovation Teams: An Interprofessional Approach to Health; Apr 11, 2017; St Paul, MN, United States.
16. Solheim-Witt M, Serck-Hanssen I. Communication as a bridge to proactive behavior, resilience and well-being. Presented at: Interprofessional Education Summit: What if? The Impact of Innovation and Interprofessionalism on Healthcare; Apr 9, 2018; St Paul, MN, United States.
17. Solheim-Witt M, Serck-Hanssen I. Resilience, resourcefulness and self-leadership: evaluation of a self-management program and novel conceptual framework of vitality type patterns in behavior and communication. Presented at: National Institute on Drug Abuse Genetics and Epigenetics Cross Cutting Research Team Meeting; May 16-17, 2023; Bethesda, MD, United States.
18. Hyzy M, Bond R, Mulvenna M, et al. System usability scale benchmarking for digital health apps: meta-analysis. JMIR mHealth uHealth 2022 Aug 18;10(8):e37290. [doi: 10.2196/37290] [Medline: 35980732]
19. Anker JJ, Kushner MG, Thuras P, Menk J, Unruh AS. Drinking to cope with negative emotions moderates alcohol use disorder treatment response in patients with co-occurring anxiety disorder. Drug Alcohol Depend 2016 Feb 1;159:93-100. [doi: 10.1016/j.drugalcdep.2015.11.031] [Medline: 26718394]
20. Stone BM. Positive psychology for substance use disorders: a rationale and call to action. J Stud Alcohol Drugs 2022 Nov;83(6):959-961. [doi: 10.15288/jsad.22-00259] [Medline: 36484594]
21. Mahipalan M, S S. Workplace spirituality and subjective happiness among high school teachers: gratitude as a moderator. Explore (NY) 2019;15(2):107-114. [doi: 10.1016/j.explore.2018.07.002] [Medline: 30150147]
22. Stone BM. Integrating positive psychology into substance use treatments. Encyc 2023;3(3):1133-1144. [doi: 10.3390/encyclopedia3030082]
23. Krentzman AR, Barker SL. Counselors' perspectives of positive psychology for the treatment of addiction: a mixed methods pilot study. Alcohol Treat Q 2016;34(4):370-385. [doi: 10.1080/07347324.2016.1217705] [Medline: 27980355]
24. Sugarman DE, Meyer LE, Reilly ME, Greenfield SF. Feasibility and acceptability of a web-based, gender-specific intervention for women with substance use disorders. J Womens Health (Larchmt) 2020 May;29(5):636-646. [doi: 10.1089/jwh.2018.7519] [Medline: 31584313]
25. White AM, Orosz A, Powell PA, Koob GF. Alcohol and aging - an area of increasing concern. Alcohol 2023 Mar;107:19-27. [doi: 10.1016/j.alcohol.2022.07.005] [Medline: 35940508]

Abbreviations

CBT: cognitive behavioral therapy
DTC: drink to cope
HIPAA: Health Insurance Portability and Accountability Act
PPI: positive psychology intervention
QI: quality improvement
REDCap: Research Electronic Data Capture
SUS: System Usability Scale

XSL•FO
RenderX

Edited by A Mavragani; submitted 30.06.24; peer-reviewed by E Grekin; revised version received 05.12.24; accepted 06.12.24; published 10.02.25.

Please cite as:
Serck-Hanssen I, Solheim-Witt M, Anker JJ, Sugarman DE
A Web-Based Resource Informed by Cognitive Behavioral Therapy and Positive Psychology to Address Stress, Negative Affect, and Problematic Alcohol Use: A Usability and Descriptive Study
JMIR Form Res 2025;9:e63819
URL: https://formative.jmir.org/2025/1/e63819
doi:10.2196/63819

© Ingrid Serck-Hanssen, Marit Solheim-Witt, Justin J Anker, Dawn E Sugarman. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Step Count Accuracy of the Life Plus Connected Watch at Different Localizations and Speeds in Healthy Adults, Patients With Cardiovascular Disease, and Patients With Peripheral Artery Disease: Step Count Validation Study in Laboratory Settings

Anne-Noëlle Heizmann[1,2], PhD; Edouard Ollier[1,3], MCU-PH; Pierre Labeix[4], PhD; Ivan Goujon[4], MSc; Frédéric Roche[1,2,4], PUPH; Claire Le Hello[1,2,5], PUPH

[1]INSERM, SAINBIOSE U1059, Saint-Priest-en-Jarez, France

[2]University Jean Monnet, Saint-Etienne, France

[3]Innovation and Pharmacology Clinical Research Unit, University Hospital of Saint-Etienne, Saint-Etienne, France

[4]Department of Physiology, University Hospital of Saint-Etienne, Saint-Etienne, France

[5]Department of Therapeutics and Vascular Medicine, University Hospital of Saint-Etienne, Saint-Etienne, France

Corresponding Author:
Anne-Noëlle Heizmann, PhD
INSERM, SAINBIOSE U1059, Saint-Priest-en-Jarez, France

Abstract

Background: Smartwatches are increasingly used to monitor and motivate physical activity. Patients with cardiovascular disease (CVD) and peripheral artery disease (PAD) often do not meet national physical activity recommendations. They may, thus, benefit from a physical activity program using smartwatches. The Life Plus smartwatch is designed to facilitate activity monitoring by counting steps, but its validity needs to be determined, particularly in patients who may not have a normal gait, such as those with cardiovascular pathology.

Objective: This study evaluates the accuracy of the Life Plus smartwatch (versions 2 and 3) in healthy adults, patients with CVD, and patients with PAD at different walking speeds (1.8, 2.5, 3.2, and 4 km/h) and different localizations (wrists, hips, and ankles) to determine best accuracy.

Methods: In total, 34 participants, comprising healthy individuals (n=10), patients with CVD (n=14), and patients with PAD (n=10), wore 6 Life Plus watches simultaneously (3 of version 2 and 3 of version 3), located on wrists, hips, and ankles. Participants walked on a treadmill for 3-minute sessions at speeds of 1.8, 2.5, 3.2, and 4 km/h; they then performed a 10-minute free walking on the ground and again walked for 3-minute sessions on a treadmill at the same speeds. Actual step counts were recorded through video footage.

Results: When worn at the wrist, no significant difference between the actual number of steps and step count by version 2 watches was found in each group independently (healthy group: *P*=.25; CVD group: *P*=.50; and PAD group: *P*=.37). Significant differences were found with the version 2 watches at the wrist in the healthy group at 3.2 (−5.26%; *P*=.01) and 4 km/h (−6.13%; *P*=.008) and in the CVD group at 2.5 (−5.94%; *P*=.008), 3.2(−13.1%; *P*=.008), and 4 km/h (−13.96%*; P*=.004). When worn at the wrist, no significant difference between actual number of steps and step count by version 3 watches was found in the healthy group (*P*=.02) and the PAD group (*P*=.90). Significant differences were found with the version 3 watches at the wrist in the CVD group at 3.2 (−6.43%; *P*=.001) and 4 km/h (−7.3%; *P*=.01) and in the PAD group at 4 km/h (−5.77%; *P*=.04).

Conclusions: For optimized counting when worn at the wrist, healthy individuals should prefer version 2 watches for slow walking (1.8 and 2.5 km/h) and version 3 for higher speeds (3.2 and 4 km/h). Patients (CVD and PAD) should prefer version 2 watches at 1.8 km/h and version 3 watches for higher speeds (2.5, 3.2, and 4 km/h).

(JMIR Form Res 2025;9:e58964) doi:10.2196/58964

KEYWORDS

physical activity assessment; accuracy; validation; steps; smartwatches; wearables; cardiovascular disease; cardiovascular; cardiology; heart; artery; peripheral artery disease; wearable physical activity monitoring device; physical activity; exercise; walking; monitoring; mobile phone

XSL•FO
RenderX

Introduction

The benefits of regular physical activity on global health are well documented [1-4]. The key role of physical activity in cardiovascular diseases (CVDs) including peripheral artery disease (PAD) has been well reported within the scientific literature and has even been recognized as an effective nonmedical therapy and core component of cardiac rehabilitation [5,6]. Walking is the most natural physical activity that confers many benefits with minimal adverse effects [7-9]. In this context, daily step count seems to be an interesting target for achieving physical activity recommendations and setting physical activity goals [10]. Daily step count can easily be assessed and monitored by the patients themselves through wearable physical activity monitoring devices like a smartwatch. In healthy populations, wearable devices such as smartwatches are frequently used as motivational and self-monitoring tools for physical activity [11-14]. These wearable physical activity tracker devices offer considerable advantages in health care and personalized physical activity management in populations with chronic disease [15]. The gait of these patients can be different than those of healthy persons and may therefore influence the wearable device's accuracy [16-19]. Patients with PAD are known to have an altered gait cycle and a reduction in step lengths [20-23].

The Life Plus connected watch is a wearable physical activity monitoring device that provides information about daily steps and daily walking activity. It also offers personalized and progressive goals, feedback, encouragements, and personalized innovative challenges, taking into account the users' personal physical activity level and capacity. This watch was designed to be easy to use and not requiring a smartphone app, synchronization, or an internet connection. This watch seems to be an interesting tool to motivate patients to walk more. As this connected watch has not been evaluated, we wanted to estimate its accuracy in a healthy adult population as well as in populations with CVD or PAD. The aim of this study was to determine the accuracy of the Life Plus connected watch (versions 2 and 3) in counting steps under controlled treadmill walking conditions at different speeds in healthy adults and patients with CVD or PAD, with each participant wearing 6 watches at different locations (wrists, hips, and ankles). In view of the current literature, we hypothesize that the Life Plus watch is reliable for step counting with variations according to speeds, localizations, and the different groups (healthy adults, patients with CVD, and patients with PAD).

Methods

Recruitment

The inclusion criteria were adult healthy volunteers, adult persons with CVD or PAD (asymptomatic stage or stress ischemia stage) able to give written informed consent, to understand walking instructions to walk on a treadmill alone without technical help for several periods of 3 minutes at different speeds (1.8 to 4 km/h). Patients with CVD were recruited in the "Viellissement, Système Nerveux Autonome et Sommeil" (aging, autonomic nervous system, and sleep) center of the University Hospital of Saint-Etienne during their cardiac rehabilitation program. Patients with PAD were recruited in the Vascular Medicine Department of the University Hospital of Saint-Etienne during a scheduled ultrasound test for PAD. Investigators (CLH, ANH, FR, PL, and IG) explained the purpose and described the procedure to the patients who then decided whether or not they wanted to take part in the study. Healthy adults were volunteers recruited from medical and paramedical professionals at Saint-Etienne University Hospital, who gave their oral consent after a detailed presentation of the study. Exclusion criteria were volunteers and patients unable to walk independently, patients with PAD at chronic limb–threatening ischemia stage, and patients with walking limitations due to other pathologies. The ability to walk on a treadmill for 3 minutes at different paces was reported by all the patients recruited. None of the patients or healthy adults underwent any surgical procedures in the 3 months preceding the treadmill walking sessions.

Ethical Considerations

This study was approved by the ethics committee of the University Hospital of Saint-Etienne in June 2022 (reference CE 2022 - 36). The purpose of the study and the description of the procedure were explained to the participants before they signed an informed consent. Retrieved study data were deidentified and anonymized. No compensation was provided for participation.

Material and Data Collection

Each participant wore simultaneously 6 Life Plus connected watches: 3 version 2 and 3 version 3, version 2 positioned on one side with one at the wrist, one at the hip, and one at the ankle; and version 3 on the other side in the same localizations. This disposition was determined to evaluate versions 2 and 3 Life Plus connected watches and to identify the best placement for better accuracy since the position may influence its accuracy [24]. Treadmill walking provided a controlled and easily replicable physical activity for evaluating the accuracy of step counting at different speeds by watches [25].

The Life Plus connected watch uses a triaxial accelerometer. Versions 2 and 3 are embedded with different accelerometers and step counting algorithms. Life Plus watches integrate a microelectromechanical system technology that is a spring-levered step counter. Step count recording was collected directly through the Life Plus watch screen. Videos of the participant's lower limbs were recorded for a visual count of steps taken during all the walking sessions. Actual numbers of steps were counted a posteriori through visual inspection by the investigators with a camera of a Samsung Galaxy TAB S6 Lite under Android, which has a number of effective pixels of 12.8 megapixels (per million) with a full high-definition resolution of 1920×1080 (30 frames per screen).

The following data were retrieved: step counts by the watches, actual numbers of steps (videos), watch version (versions 2 and 3), group (healthy adults, patients with CVD, and patients with PAD), watch position (wrist, hip, and ankle), walking speed (1.8, 2.5, 3.2, and 4 km/h), help from the treadmill's bars (yes

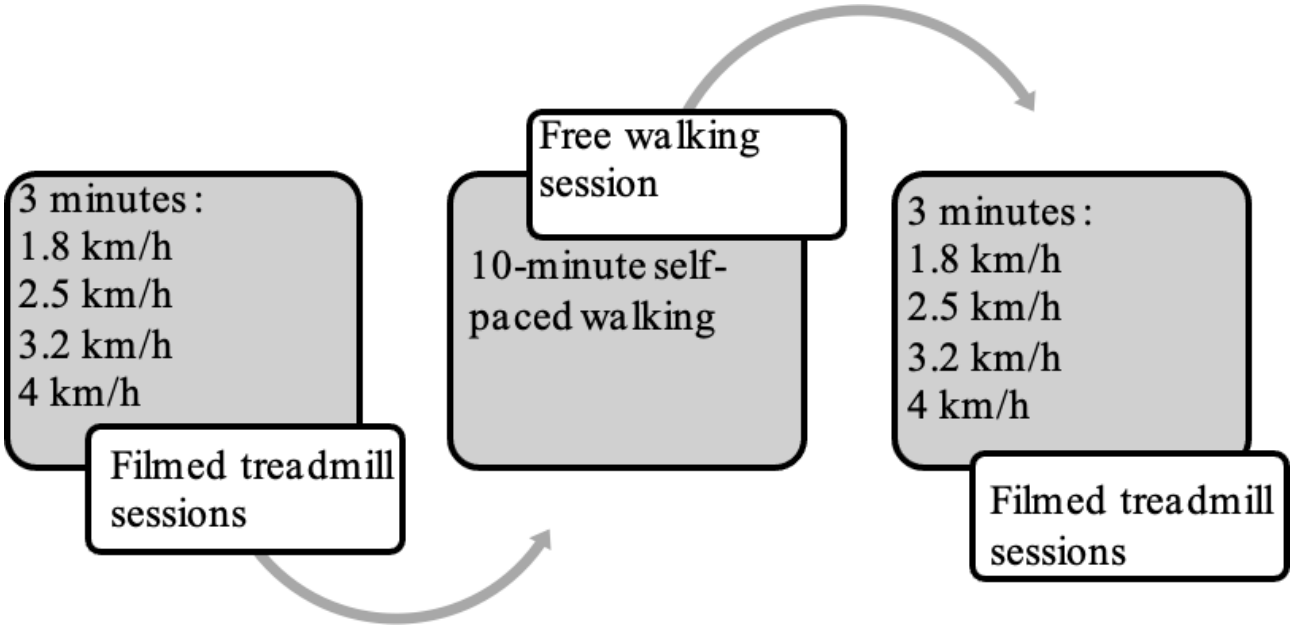
XSL•FO
RenderX

and no), patients characteristics (age, BMI, and ankle-brachial index for patients with PAD).

Study Design

After giving written informed consent, participants were invited to walk 4 sessions of 3 minutes on the treadmill at different speeds (1.8, 2.5, 3.2, and 4 km/h). As far as possible, participants were asked to let go of the treadmill bars as soon as they started their walking movement so that their arms could swing normally. Between each session, participants had to sit still on a chair for at least 2 minutes to allow step count data to be refreshed on the watch screen. After the first 4 sessions, the participants had a 10-minute free walking session at their own pace in the hospital corridor to simulate a fatigue pattern. Participants then again walked 4 sessions of 3 minutes on the treadmill at the same speeds as the first time. The treadmill walking sessions were filmed so that 2 investigators could count the actual number of steps taken during all the treadmill walking sessions (criterion measure). The study design is illustrated in Figure 1.

Figure 1. Study design.



Statistical Analysis

Each participant was assessed in 8 walking sessions at 1.8, 2.5, 3.2, and 4 km/h. In all the conditions, the difference between the step count recorded by the Life Plus connected watches and the actual number of steps counted by the investigators was computed. We calculated the quantiles Q1 and Q3 for the data on actual number of steps, then the IQR=Q3−Q1; then we calculated the lower and upper thresholds for the outliers with lower threshold=Q1−1.5*IQR and upper threshold=Q3+1.5*IQR. Outliers were removed from the data collected by the watches. For all analyses, the α significance threshold used was 5%. Concordance between the step count recorded by the connected watches and the actual number of steps was quantified by calculating the difference between the actual number of steps and the step count measured by the watches. These differences were analyzed using the Wilcoxon test as the Shapiro-Wilk test was performed and showed deviance from normality for actual number of steps and step count by watches.

ANOVA was computed to compare differences in age and BMI between groups, and the Tukey test was used to determine which differences were significant. Finally, Bland-Altman charts were constructed to evaluate the mean bias and the limits of agreement between step count by watches and actual numbers of steps. All analyses were conducted using the R Statistical language (version 4.0.4; R Foundation for Statistical Computing).

Results

Demographics

In total, 34 participants were included in the study: 10 in the healthy group, 14 in the CVD group, and 10 in the PAD group. Characteristics of the 3 groups are summarized in Table 1.

XSL•FO
RenderX

Table . Demographic characteristics of the 3 groups included in the Life Plus watch validation study (healthy adults, patients with cardiovascular disease [CVD], and patients with peripheral artery disease [PAD]).

	Healthy group (n=10)	CVD group (n=14)	PAD group (n=10)
Age (years), mean (SD)	32.5 (11)	59.7 (14.1)	65.1 (9.8)
Sex, n (%)			
Female	5 (50)	2 (14)	2 (20)
Male	5 (50)	12 (86)	8 (80)
BMI (kg/m^2), mean (SD)	22.18 (2.98)	24.22 (1.89)	26.85 (4.03)
Ankle-brachial index, mean (SD)			
Right	—[a]	—	0.84 (0.28)
Left	—	—	0.88 (0.18)
Toe pressure (mm Hg), mean (SD)			
Right	—	—	78.6 (20.1)
Left	—	—	77.6 (26.6)

[a]Not available.

No significant age difference between the CVD and PAD groups (P=.53) was found, but significant differences between the healthy group and PAD group and the healthy group and CVD group (P<.001) were found. For BMI, analyses showed no difference between the CVD and PAD groups (P=.14) and between the healthy group and CVD group (P=.29) but significant differences between the healthy group and PAD group (P=.005).

Evaluation Outcomes

Global statistical parameters are summarized in Table 2.

Table . Global statistical parameters for the number of steps counted by the investigators (number of actual steps) and counted by the Life Plus watches (versions 2 and 3).

	Number of actual steps	Number of steps measured by the watch (outliers removed)
Range	185-399	176-412
Median (IQR)	296.5 (264-324)	310 (277-343)
Mean (SD)	292.6 (43.73)	307.3 (49.71)

Independently from any condition, the Life Plus step watch overestimates the actual number of steps by mean+14.7 steps (P=.001; r=0.08; 95% CI 2-8.5).

Life Plus Watch Version 2

Independently from all conditions, the version 2 watch did not correctly trigger automatic step counting (no step counted or less than the lower threshold of outliers) in 34% (n=81) walking sessions of the healthy group, 39% (n=116) in the CVD group, and 40% (n=95) in the PAD group. When looking closer, this technical issue occurred more frequently at slow speeds (1.8 and 2.5 km/h), when participants were holding the treadmill bars, and when the watch was worn at the hip or ankle. In the healthy group at 3.2 and 4 km/h when worn at the wrist, the version 2 watch triggered step counting in each walking session. In the CVD and PAD groups, at 3.2 and 4 km/h speeds, the version 2 watch triggered step counting more efficiently. Detailed statistics are summarized in Table S1 in Multimedia Appendix 1.

When worn at the wrist, no significant difference between actual number of steps and step count by watches was found (P=.89; 95% CI −8.0 to 9.0). The same result was found in each group independently (healthy group: P=.25; 95% CI −5.99 to −21; CVD group: P=.50; 95% CI −19.9 to 8.9; and PAD group: P=.37; 95% CI −8.0 to 22).

Detailed mean differences and mean percentage differences between the actual number of steps and step count by the watches are summarized in Table 3.

In the healthy group, when the watches were worn at the wrist, without holding treadmill bars and with speeds of 3.2 and 4 km/h, they underestimated the actual number of steps by mean 5.4% (P=.008 at 3.2 and 4 km/h). Differences between actual number of steps and step count by the watches were the lowest and nonsignificant when the watch was worn at the wrist at 1.8 km/h (P=.50; 95% CI 8-22) and 2.5 km/h (P=.21; 95% CI −22 to 3.5). In the CVD group, the mean percentage difference between the actual number of steps and the step count by the watches ranged from −13.96% (n=) at 4 km/h (P=.004; 95% CI −15.5 to −2.5) to −0.83% at 1.8 km/h (P=.13; 95% CI −126 to 24) when the watch was worn at the wrist. In the PAD group, the mean percentage difference between actual number of steps and the number of steps counted by the watches ranged from −12.44% at 4 km/h (P=.65; 95% CI −46 to 13.5) to +0.70% at

XSL•FO
RenderX

1.8 km/h (*P*=.81; 95% CI −49.5 to 35) when the watch was worn at the wrist.

Table . Mean step differences and mean percentage differences between actual number of steps and number of steps counted by version 2 watch in the 3 groups (healthy adults, patients with cardiovascular disease [CVD], and patients with peripheral artery disease [PAD]) at different localizations (wrist, hip, and ankle) and different speeds (1.8, 2.5, 3.2, and 4 km/h).

Speed and position		Groups		
		Healthy, mean step difference (%)	CVD, mean step difference (%)	PAD, mean step difference (%)
1.8 km/h				
	Wrist	9.37 (7.01)	−11.93 (−0.83)	−6.64 (0.70)
	Hip	17.94 (10.17)	−3.37 (2.33)	1.92 (3.86)
	Ankle	55.87 (23.85)	34.57 (16.02)	39.86 (17.54)
2.5 km/h				
	Wrist	2.24 (1.89)	−19.06 (−5.94)	−13.78 (−4.42)
	Hip	10.80 (5.06)	−10.50 (2.78)	5.21 (−1.26)
	Ankle	48.73 (18.74)	27.43 (10.90)	32.72 (12.43)
3.2 km/h				
	Wrist	−16.92 (−5.26)	−38.00 (−13.10)	−32.93 (−11.57)
	Hip	−8.36 (−2.10)	−29.66 (−9.94)	24.37 (−8.41)
	Ankle	29.58 (11.58)	8.28 (3.75)	13.56 (5.27)
4 km/h				
	Wrist	−18.76 (−6.13)	−40.06 (−13.96)	−34.78 (−12.44)
	Hip	10.20 (−2.96)	−31.50 (−10.80)	−26.21 (−9.27)
	Ankle	27.73 (10.72)	6.43 (2.88)	11.72 (4.41)

The Bland-Altman plots revealed no systematic differences between the actual number of steps and the number of steps counted by the watches for the 3 groups when the watch was worn at the wrist (Figure S1 in Multimedia Appendix 1).

Life Plus Watch Version 3

Independently from all conditions, the version 3 watch did not correctly trigger automatic step counting (no step counted or less than the lower threshold of outliers) in 35% (n=83) of walking sessions in the healthy group, 43% (n=129) in the CVD group, and 43% (n=103) in the PAD group. When looking closer, this technical issue occurred more frequently at 1.8 and 2.5 km/h, when participants were holding the treadmill bars, and when the watch was worn at the hip or ankle. At 3.2 and 4 km/h and with the watch worn at the ankle, the watch version 3 triggered step counting more frequently. Detailed statistics are summarized in Table S2 in Multimedia Appendix 1.

When worn at the wrist, no significant difference between actual number of steps and number of steps counted by the watches (*P*=.52; 95% CI −15 to 7.0). The same result was found in the healthy group (*P*=.02; 95% CI 2.99-41.99) and the PAD group (*P*=.90; 95% CI −19 to +20). A significant difference was found in the CVD group (*P=.006; P*=.01; 95% CI −43 to +9) at higher speeds (3.2 and 4 km/h) but not at lower speeds (1.8 and 2.5 km/h). Mean differences and mean percentage differences between the actual number of steps and the number of steps counted by the watches are summarized in Table 4.

XSL•FO
RenderX

Table . Mean step differences and mean percentage differences between actual number of steps and number of steps counted by version 3 watch in the 3 groups (healthy adults, patients with cardiovascular disease [CVD], and patients with peripheral artery disease [PAD]) at different positions (wrist, hip, and ankle) and different speeds (1.8, 2.5, 3.2, and 4 km/h).

Speed and position		Groups		
		Healthy, mean step difference (%)	CVD, mean step difference (%)	PAD, mean step difference (%)
1.8 km/h				
	Wrist	28.28 (13.67)	6.98 (5.84)	12.27 (7.36)
	Hip	36.84 (16.83)	15.54 (8.99)	20.83 (10.52)
	Ankle	74.78 (30.51)	53.48 (22.68)	58.77 (24.20)
2.5 km/h				
	Wrist	21.15 (8.56)	−0.16 (0.72)	5.13 (2.25)
	Hip	29.71 (11.72)	8.41 (3.88)	13.70 (5.41)
	Ankle	67.64 (25.40)	46.34 (17.57)	51.63 (9.09)
3.2 km/h				
	Wrist	1.99 (1.40)	−14.03 (−6.43)	−19.31 (−4.91)
	Hip	10.55 (4.56)	−10.75 (−3.27)	−5.46 (−1.75)
	Ankle	48.49 (11.58)	27.18 (10.41)	32.47 (11.94)
4 km/h				
	Wrist	0.15 (0.54)	−21.16 (−7.30)	−15.87 (−5.77)
	Hip	8.71 (3.70)	−12.59 (−4.14)	−7.31 (−2.61)
	Ankle	46.64 (17.38)	25.34 (9.55)	30.63 (11.07)

Results showed that in the healthy group, differences between actual number of steps and steps counted by the watches were lower and nonsignificant when the watch was worn at the wrist at all speeds. In the healthy group, the version 3 watch tended to globally overestimate step counts in all conditions. In the CVD group, the mean percentage difference between actual number of steps and number of steps counted by the watch ranged from −7.29% (P=.01; 95% CI −32.5 to 108.5) at 4 km/h to +5.84% (P=.25; 95% CI −91 to 22) at 1.8 km/h when the watch was worn at the wrist. In the PAD group, the mean percentage of difference between the actual number of steps and the number of steps counted by the watches ranged from −5.77% (P=.04; 95% CI −59 to −4) at 4 km/h to +7.36% (P=1.0; 95% CI −156 to 81) at 1.8 km/h when the watch was worn at the wrist.

The Bland-Altman plots revealed no systematic differences between the actual number of steps and the step count obtained by the watches when the watch was worn at the wrist (Figure S2 in Multimedia Appendix 1).

Discussion

Principal Findings

Activity trackers were designed to help people to monitor their daily physical activity. A smartwatch able to accurately count the number of steps and provide motivating features, such as the Life Plus watch, would benefit patients' physical activity. Indeed, it is important to have a reliable activity tracker [26] to support and coach patients in home-based physical activity rehabilitation. Inaccurate information on physical activity may lead to frustration for the patient if his activity is underestimated or to false positive feedback if overestimated. As these tools often are tested in healthy populations, we wanted to test the accuracy of the Life Plus watch for counting steps in healthy volunteers, patients with CVD, and patients with PAD.

On the basis of our investigation, the Life Plus smartwatches seemed to be accurate for counting steps in healthy persons when walking on a treadmill without holding the lateral bars. The version 3 version was better in counting steps.

Our results indicated that in the healthy group, when worn at the wrist, the version 2 watches were more accurate at 1.8 and 2.5 km/h; whereas at 3.2 and 4 km/h, the version 3 watches were more accurate. Healthy adults should wear the watch, version 2 or 3, at the wrist for better accuracy. In the CVD and PAD groups, versions 2 and 3 watches have similar accuracy at 1.8 and 2.5 km/h when worn at the wrist or hips; at 3.2 and 4 km/h, the version 3 watches are more accurate than the version 2 when worn at the wrist or hips. In practice, the version 3 watch seems to be the better option in all groups with better accuracy at all speeds when worn at the wrist or hips.

Differences between the number of steps counted by watches according to their version (versions 2 and 3) are explained by the different accelerometers and corresponding step counting algorithm embedded in versions 2 and 3.

Comparison With Prior Work and Limitations

Globally, our investigation showed that step count accuracy decreased as walking speed decreased, which is consistent with

XSL•FO
RenderX

other validation studies [27,28]. In our study design, speeds were not counterbalanced; therefore, we could not ensure that there was no order effect on step counts.

The localization of the watch on the body also influences the step count accuracy, with overestimation when worn at the ankles, and underestimation when worn at the hips at higher speeds. This result follows other validation studies [29].

The watch had slightly lower accuracy in the CVD and PAD groups. This could be explained by the fact that the patients were more often holding the treadmill lateral bars to maintain stability, and wrist-worn wearable devices typically undercount the number of steps when hands are fixed on lateral hand bars [30]. Moreover, patients may have had some impaired arm swing and gait pattern as shown in several studies [17,18,20-23]. Furthermore, age, sex, and BMI also may influence the validity of wearable devices [31]. These could also have affected our results as the mean age of the healthy participant's group was much lower than that of the 2 disease groups. This makes the comparison between groups difficult. Results could also be influenced by the fact that the study design did not propose a familiarization session on the treadmill. Walking on a treadmill is not natural and requires a lot of attention at first. Several studies showed indeed higher cadences, shorter step lengths, shorter swing phases, longer stance phases, and higher walking variability when walking on a treadmill [32,33]. A 10-minute session to familiarize with walking on a treadmill could enable to be in conditions nearer to natural walking [34].

Participants only walked for 3 minutes for each speed. The accelerometer has a wake-up threshold, meaning it only starts measuring movement when a certain level of acceleration is detected. At slower speeds, the arm swing may not be sufficiently pronounced to reach the accelerometer's step count threshold. The algorithm has set specific thresholds that need to be met for a movement to be recognized as a step. If the arm swing is too subtle, it may not reach these thresholds, resulting in steps not being counted. This could explain the percentage of errors in step count.

Our research evaluated the accuracy of the Life Plus watch, versions 2 and 3, in laboratory conditions. Future research should investigate the accuracy of the watch during free-living conditions.

Conclusions

We investigated the accuracy of versions 2 and 3 of the Life Plus watch in healthy adults, patients with CVD, and patients with PAD. Accuracy of the version 3 was better for all walking speeds when worn at the wrist. Mean percentage differences between actual number of steps (videos) and number of steps counted by watches were less than 10% when version 3 was worn at the wrist by each group at all speeds except 1.8 km/h. Globally, independently from the watch version, accuracy slightly decreased when worn by patients with CVD or PAD walking at slower speeds. In the CVD and PAD groups, the accuracy of version 3 was better when worn at the wrist for slower walking speeds (1.8 and 2.5 km/h) and when worn at the hips for higher walking speeds. In practice, when worn at the wrist as it is intended, healthy individuals should prefer version 2 watches for slow walking (1.8 and 2.5 km/h) and version 3 for higher speeds (3.2 and 4 km/h). Patients (CVD and PAD) should prefer version 2 watches at 1.8 km/h and version 3 watches for higher speeds (2.5, 3.2, and 4 km/h).

Acknowledgments

The researchers wish to thank all the volunteers for their time and the staff of the Viellissement, Système Nerveux Autonome et Sommeil rehabilitation center of Saint Etienne University Hospital for the treadmill use and patient recruitment help.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

ANH and CLH conducted the experiment, EO helped with the statistical analyses, IG and PL helped in patients' recruitment, and FR allowed access to cardiac rehabilitation. All coauthors reviewed and approved the manuscript for publication.

Conflicts of Interest

ANH is a PhD student working in collaboration with the Life Plus company as part of her Convention Industrielle de Formation par la Recherche (industrial agreements for training through research) thesis (optimization and use of the Life Plus watch in rehabilitation of patients with cardiovascular disease). All other authors have no known conflict of interest to disclose.

Multimedia Appendix 1
Outliers table and Bland-Altman plots.
[DOCX File, 1096 KB - formative_v9i1e58964_app1.docx]

References

XSL•FO
RenderX

1. Ekelund U, Tarp J, Steene-Johannessen J, et al. Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. BMJ 2019 Aug 21;366:l4570. [doi: 10.1136/bmj.l4570] [Medline: 31434697]
2. Li J, Siegrist J. Physical activity and risk of cardiovascular disease—a meta-analysis of prospective cohort studies. Int J Environ Res Public Health 2012 Feb;9(2):391-407. [doi: 10.3390/ijerph9020391] [Medline: 22470299]
3. Warburton DER. Health benefits of physical activity: the evidence. Can Med Assoc J 2006 Mar 14;174(6):801-809. [doi: 10.1503/cmaj.051351]
4. Cardiovascular diseases (CVDs). World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) [accessed 2025-01-09]
5. Gerhard-Herman MD, Gornik HL, Barrett C, et al. 2016 AHA/ACC guideline on the management of patients with lower extremity peripheral artery disease: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation 2017 Mar 21;135(12):e686-e725. [doi: 10.1161/CIR.0000000000000470] [Medline: 27840332]
6. Casillas JM, Troisgros O, Hannequin A, et al. Rehabilitation in patients with peripheral arterial disease. Ann Phys Rehabil Med 2011 Oct;54(7):443-461. [doi: 10.1016/j.rehab.2011.07.001] [Medline: 21917541]
7. Hall KS, Hyde ET, Bassett DR, et al. Systematic review of the prospective association of daily step counts with risk of mortality, cardiovascular disease, and dysglycemia. Int J Behav Nutr Phys Act 2020 Jun 20;17(1):78. [doi: 10.1186/s12966-020-00978-9] [Medline: 32563261]
8. Morris JN, Hardman AE. Walking to health. Sports Med 1997 May;23(5):306-332. [doi: 10.2165/00007256-199723050-00004] [Medline: 9181668]
9. Oja P, Kelly P, Murtagh EM, Murphy MH, Foster C, Titze S. Effects of frequency, intensity, duration and volume of walking interventions on CVD risk factors: a systematic review and meta-regression analysis of randomised controlled trials among inactive healthy adults. Br J Sports Med 2018 Jun;52(12):769-775. [doi: 10.1136/bjsports-2017-098558] [Medline: 29858464]
10. Kraus WE, Janz KF, Powell KE, et al. Daily step counts for measuring physical activity exposure and its relation to health. Med Sci Sports Exerc 2019;51(6):1206-1212. [doi: 10.1249/MSS.0000000000001932]
11. Bravata DM, Smith-Spangler C, Sundaram V, et al. Using pedometers to increase physical activity and improve health: a systematic review. JAMA 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
12. Sanders JP, Loveday A, Pearson N, et al. Devices for self-monitoring sedentary time or physical activity: a scoping review. J Med Internet Res 2016 May 4;18(5):e90. [doi: 10.2196/jmir.5373] [Medline: 27145905]
13. Vetrovsky T, Siranec M, Marencakova J, et al. Validity of six consumer-level activity monitors for measuring steps in patients with chronic heart failure. PLOS ONE 2019;14(9):e0222569. [doi: 10.1371/journal.pone.0222569] [Medline: 31518367]
14. Brickwood KJ, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819. [doi: 10.2196/11819] [Medline: 30977740]
15. Park LG, Elnaggar A, Lee SJ, et al. Mobile health intervention promoting physical activity in adults post cardiac rehabilitation: pilot randomized controlled trial. JMIR Form Res 2021 Apr 16;5(4):e20468. [doi: 10.2196/20468] [Medline: 33861204]
16. Ata R, Gandhi N, Rasmussen H, et al. Clinical validation of smartphone-based activity tracking in peripheral artery disease patients. NPJ Digit Med 2018;1(1):66. [doi: 10.1038/s41746-018-0073-x] [Medline: 31304343]
17. Ozcan EB, Saglam M, Vardar-Yagli N, et al. Impaired balance and gait characteristics in patients with chronic heart failure. Heart Lung Circ 2022 Jun;31(6):832-840. [doi: 10.1016/j.hlc.2021.10.015] [Medline: 35177316]
18. Panizzolo FA, Maiorana AJ, Naylor LH, et al. Gait analysis in chronic heart failure: the calf as a locus of impaired walking capacity. J Biomech 2014 Nov 28;47(15):3719-3725. [doi: 10.1016/j.jbiomech.2014.09.015] [Medline: 25307437]
19. Thorup CB, Andreasen JJ, Sørensen EE, Grønkjær M, Dinesen BI, Hansen J. Accuracy of a step counter during treadmill and daily life walking by healthy adults and patients with cardiac disease. BMJ Open 2017 Mar 31;7(3):e011742. [doi: 10.1136/bmjopen-2016-011742] [Medline: 28363918]
20. Gardner AW, Forrester L, Smith GV. Altered gait profile in subjects with peripheral arterial disease. Vasc Med 2001;6(1):31-34. [doi: 10.1177/1358836X0100600106] [Medline: 11358158]
21. McDermott MM, Ohlmiller SM, Liu K, et al. Gait alterations associated with walking impairment in people with peripheral arterial disease with and without intermittent claudication. J Am Geriatr Soc 2001 Jun;49(6):747-754. [doi: 10.1046/j.1532-5415.2001.49151.x] [Medline: 11454113]
22. Myers SA, Johanning JM, Stergiou N, Celis RI, Robinson L, Pipinos II. Gait variability is altered in patients with peripheral arterial disease. J Vasc Surg 2009 Apr;49(4):924-931. [doi: 10.1016/j.jvs.2008.11.020] [Medline: 19217749]
23. Szymczak M, Krupa P, Oszkinis G, Majchrzycki M. Gait pattern in patients with peripheral artery disease. BMC Geriatr 2018 Feb 20;18(1):52. [doi: 10.1186/s12877-018-0727-1] [Medline: 29458330]
24. Tudor-Locke C, Barreira TV, Schuna JM. Comparison of step outputs for waist and wrist accelerometer attachment sites. Med Sci Sports Exerc 2015 Apr;47(4):839-842. [doi: 10.1249/MSS.0000000000000476] [Medline: 25121517]

XSL•FO
RenderX

25. Moore CC, McCullough AK, Aguiar EJ, Ducharme SW, Tudor-Locke C. Toward harmonized treadmill-based validation of step-counting wearable technologies: a scoping review. J Phys Act Health 2020 Aug 1;17(8):840-852. [doi: 10.1123/jpah.2019-0205] [Medline: 32652514]
26. Kang HS, Exworthy M. Wearing the future-wearables to empower users to take greater responsibility for their health and care: scoping review. JMIR Mhealth Uhealth 2022 Jul 13;10(7):e35684. [doi: 10.2196/35684] [Medline: 35830222]
27. Bassett DR, Toth LP, LaMunion SR, Crouter SE. Step counting: a review of measurement considerations and health-related applications. Sports Med 2017 Jul;47(7):1303-1315. [doi: 10.1007/s40279-016-0663-1]
28. Gould ZR, Mora-Gonzalez J, Aguiar EJ, et al. A catalog of validity indices for step counting wearable technologies during treadmill walking: the CADENCE-Kids study. Int J Behav Nutr Phys Act 2021 Dec;18(1):97. [doi: 10.1186/s12966-021-01167-y] [Medline: 36076265]
29. Kotler D, Wu C, Thompson K, et al. The accuracy of fitness trackers varies with their location on the body (P16-019-19). Curr Dev Nutr 2019 Jun;3:nzz050.P16-019-19. [doi: 10.1093/cdn/nzz050.P16-019-19]
30. Lynn R, Pfitzer R, Rogers RR, Ballmann CG, Williams TD, Marshall MR. Step-counting validity of wrist-worn activity monitors during activities with fixed upper extremities. J Meas Phys Behav 2020;3(3):197-203. [doi: 10.1123/jmpb.2019-0055]
31. Musa SB, Ellis R, Chafe B, et al. Wearable device validity in measuring steps, energy expenditure, and heart rate across age, gender, and body mass index: data analysis from a systematic review. J Phys Act Health 2023 Feb 1;20(2):100-105. [doi: 10.1123/jpah.2022-0160] [Medline: 36535270]
32. Murray MP, Spurr GB, Sepic SB, Gardner GM, Mollinger LA. Treadmill vs. floor walking: kinematics, electromyogram, and heart rate. J Appl Physiol (1985) 1985 Jul;59(1):87-91. [doi: 10.1152/jappl.1985.59.1.87] [Medline: 4030579]
33. Warabi T, Kato M, Kiriyama K, Yoshida T, Kobayashi N. Treadmill walking and overground walking of human subjects compared by recording sole-floor reaction force. Neurosci Res 2005 Nov;53(3):343-348. [doi: 10.1016/j.neures.2005.08.005] [Medline: 16182398]
34. Van de Putte M, Hagemeister N, St-Onge N, Parent G, de Guise JA. Habituation to treadmill walking. Biomed Mater Eng 2006;16(1):43-52. [Medline: 16410643]

Abbreviations

CVD: cardiovascular disease
PAD: peripheral artery disease

Edited by A Mavragani; submitted 29.03.24; peer-reviewed by G Piva, RJ Mays; revised version received 28.07.24; accepted 02.09.24; published 10.02.25.

Please cite as:
Heizmann AN, Ollier E, Labeix P, Goujon I, Roche F, Le Hello C
Step Count Accuracy of the Life Plus Connected Watch at Different Localizations and Speeds in Healthy Adults, Patients With Cardiovascular Disease, and Patients With Peripheral Artery Disease: Step Count Validation Study in Laboratory Settings
JMIR Form Res 2025;9:e58964
URL: https://formative.jmir.org/2025/1/e58964
doi:10.2196/58964

© Anne-Noëlle Heizmann, Edouard Ollier, Pierre Labeix, Ivan Goujon, Frédéric Roche, Claire Le Hello. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mediating Effect of Social Capital on the Association Between Digital Literacy and Life Satisfaction Among Older Adults in South Korea: Cross-Sectional Study

Hyein Jung[1], MA; Hocheol Lee[2], PhD; Eun Woo Nam[1,2], MPH, PhD

[1]Department of Health Administration, Yonsei University Graduate School, Wonju, Republic of Korea

[2]Division of Health Administration, College of Software and Digital Healthcare Convergence, Yonsei University, Unit 412, Chang-jo gwan 1 Yonseidae-gil, Wonju, Republic of Korea

Corresponding Author:
Eun Woo Nam, MPH, PhD
Department of Health Administration, Yonsei University Graduate School, Wonju, Republic of Korea

Abstract

Background: As Korea rapidly transforms into a super-aged society, research indicates that digital literacy among older adults enhances their life satisfaction. Digital literacy refers to the ability to efficiently use digital technologies, encompassing access, competency, and utilization. It reflects the capacity to navigate and benefit from digital environments effectively. Furthermore, social capital positively influences the quality of life, and digital literacy facilitates social capital formation. However, since most studies have only focused on the direct relationship between digital literacy and life satisfaction, research on the mediating role of social capital remains limited.

Objective: To analyze the effect of digital literacy on the life satisfaction of older adults in South Korea and to verify whether social capital acts as a mediating factor in this process.

Methods: This descriptive cross-sectional study used data from the 2023 Report on the Digital Divide—an annual survey conducted by the Korean Ministry of Science and Information and Communications Technology. The study targeted individuals aged 65 years or older. Descriptive statistics, the Pearson correlation analyses, and the 3-step multiple regression analysis proposed by Baron and Kenny were performed. The bootstrap method was employed, and all analyses were conducted using R, version 4.4.1.

Results: The study included 869 participants. Digital literacy had a significant positive effect on their life satisfaction (β=0.103; *P*=.008). Social capital was also positively associated with life satisfaction (β=0.337; *P*<.001). Mediation analysis showed that digital literacy influenced life satisfaction both directly (β=0.103; *P*=.006) and indirectly through social capital (β=0.037; *P*=.03). Bootstrapping confirmed the significance of the indirect effect (β=0.037, 95% CI 0.005 - 0.070; *P*=.03). The total effect of digital literacy on life satisfaction was also significant (β=0.140, 95% CI 0.058 - 0.230; *P*=.002).

Conclusions: This study analyzed the association between digital literacy, social capital, and life satisfaction among older adults in Korea. We identified that social capital mediates the association between digital literacy and life satisfaction among older adults. These findings indicate that tailored digital literacy programs and support policies that promote social capital formation could help bridge the digital divide and foster social inclusion. These measures would enable older adults to access essential services, reduce social isolation, and enhance health and well-being, ultimately improving the overall quality of life.

(JMIR Form Res 2025;9:e68163) doi:10.2196/68163

KEYWORDS

digital literacy; digital access; digital competency; digital utilization; life satisfaction; older adults; social capital; mediating effect; aging

Introduction

South Korea is rapidly approaching a super-aged society, where over 20% of the population is 65 years and older. As of 2024, approximately 9.9 million, or 19.2%, of the total population fall with in this age group. By 2025, this is projected to exceed 20%, officially marking the country's transition into a super-aged society [1]. This increase in the older adult population has led to a decrease in the proportion of the economically active population. Therefore, the financial and psychological burdens at the individual, family, and national levels are expected to increase due to aging-related diseases and care issues. In 2019, the medical expenses of older adults in Korea accounted for over 40% of the nation's total medical expenses, and the country recorded the highest elderly poverty rate among the Organisation for Economic Co-operation and Development (OECD) countries

XSL•FO
RenderX

[2]. Moreover, the life satisfaction of older adults aged 65 years and over decreased by 2.4% compared to the preceding year, reaching 31.9%, which is lower than the overall average. Life satisfaction also tends to decrease with age [1].

Enhancing the digital technology proficiency of older adults has been proposed as a potential solution to address these challenges [3,4]. Although internet use has grown significantly in recent years, disparities remain, especially among older adults. In South Korea, older adults aged 55 years and above are categorized as one of the 4 major digitally vulnerable groups, alongside people with disabilities, low-income households, farmers, and fishermen. Among these groups, older adults have the lowest level of digital literacy, highlighting a significant digital divide [5]. This digital gap limits access to essential resources and opportunities, hindering efforts to improve their quality of life.

As social systems become increasingly digitized, older adults experience considerable difficulties in managing various daily tasks, such as internet-based purchases, bill payments, and financial transactions. Furthermore, the rapid expansion of telemedicine and digital health care services following the COVID-19 pandemic has significantly increased reliance on digital platforms for health management. This shift has exacerbated challenges for older adults in effectively utilizing these technologies [6,7]. These difficulties are further intensified by resistance to new technologies, psychological barriers, and negative self-perceptions related to aging [4,8,9]. Consequently, these challenges isolate older adults socially and economically and hinder their ability to lead an independent and fulfilling life.

Digital literacy is the ability to efficiently use digital technologies, grounded in fundamental literacy skills such as reading, writing, and arithmetic. It encompasses digital access, digital competency, and digital utilization, enabling individuals to search, evaluate, utilize, share, and create content using digital tools such as the internet and smartphones [10,11].

Research has shown that digital literacy positively impacts the life satisfaction of older adults by enhancing personal development, social connections, and self-esteem [12]. Activities such as searching for financial and health information and engaging in internet-based communication boost self-efficacy, foster positive perceptions of aging, alleviate loneliness and isolation, and improve overall quality of life [13-15]. Domestic research has also demonstrated that the digital literacy level of older adults positively impacts their life satisfaction [16-18]. Moreover, it plays a crucial role in maintaining and expanding relationships with family, friends, and the community, thereby positively influencing the formation of social capital [19,20].

Social capital, which refers to the social networks, norms, and trust that arise from interactions among individuals and groups, serves as a resource that promotes a sense of community and enables individuals to lead fulfilling lives [21,22]. Previous studies have identified social capital as a vital resource that helps alleviate loneliness and enhances life satisfaction, thereby positively influencing the mental well-being of older adults [23-27]. When older adults effectively utilize digital technologies, their social capital is further strengthened, helping them overcome social isolation and enhance their quality of life [19]. Specifically, the enhancement of digital literacy contributes to the growth of social capital, thereby improving emotional stability and life satisfaction [15]. Digital literacy allows older adults to connect with broader social networks, which elevates their overall quality of life.

While several studies have explored the relationship between digital literacy and life satisfaction among older adults and the connection between social capital and quality of life, most have focused only on their direct relationships. Thus, research on how social capital mediates this connection remains limited. This highlights the need to better understand how digital literacy contributes to life satisfaction, considering the potentially mediating role of social capital. Therefore, this study aims to analyze the impact of digital literacy on the life satisfaction of Korean older adults and to verify whether social capital acts as a mediating factor in this process.

Methods

Study Design and Aim

This cross-sectional study is aimed at analyzing the mediating effect of social capital on the relationship between digital literacy and life satisfaction among older adults in Korea.

Data Collection

This study used raw data from the 2023 Report on the Digital Divide—an annual survey conducted by the Korean Ministry of Science and Information and Communications Technology (MSIT). Considering the OECD's definition of older adults and the age criteria commonly used for pension and social security benefits, this study defines older adults as those aged 65 years and above. Accordingly, response data from this demographic were extracted, processed, and analyzed. The survey employed a structured questionnaire administered through face-to-face interviews. Measures were taken to minimize response bias by standardizing the interview process and providing training for surveyors. The survey period lasted from October to December 2023. For analysis, the data were deidentified in compliance with the Personal Information Protection Act. The sample was selected using a proportional and stratified probability sampling method, accounting for inclusion probability based on region, gender, and age, and estimates of standard error for each stratum. Among the 1239 participants aged 65 years and older, data from 869 individuals were analyzed after excluding responses with missing values (n=370).

Study Variables

Life Satisfaction

The dependent variable in this study, life satisfaction, was measured across 8 domains of daily life: leisure and cultural activities, financial status, social activities, interpersonal relationships, family relationships, work, physical and mental health, and political engagement. These were assessed using a 4-point Likert scale, and the scores were converted to a 100-point scale. A Cronbach α of 0.806 was calculated to assess the internal consistency of the measurement items.

XSL•FO
RenderX

Digital Literacy

The independent variable in this study, digital literacy, was evaluated by integrating 3 components: digital access, digital competency, and digital utilization. Digital access was assessed using 4 items, examining whether participants owned digital devices such as desktops, laptops, and mobile phones and whether they had constant internet access (1 for yes, 0 for no). Digital competency was measured using 7 items on personal computer competency and 7 items on mobile device competency, each assessed using a 4-point Likert scale ranging from "Strongly disagree" (1) to "Strongly agree" (4). Digital utilization was assessed based on 50 items categorized by wired or mobile internet usage, the diversity of internet services used, and advanced internet utilization. The diversity of internet service usage included activities such as information and news searches, email, media content, educational content, social networking services, messengers, blogs, internet-based communities, cloud services, daily information, e-commerce, financial transactions, and public service utilization. Advanced internet utilization measured the extent of information production and sharing, networking, social participation, and economic activities. All items were rated on a 4-point Likert scale ranging from "Do not use at all" (1) to "Use frequently" (4). Scores for digital access, digital competency, and digital utilization were calculated on a weighted 100-point scale, as outlined in the 2023 Report on the Digital Divide. A Cronbach α of 0.967 confirmed the internal consistency of these items.

Social Capital

The mediating variable, social capital, was measured by assessing the presence of social support networks, such as family and friends, in the internet-based environment for older adults. The survey used a shortened version of the Internet Social Capital Scale (ISCS) developed by Williams [28], modified and adapted by the National Information Society Agency of Korea to evaluate social capital levels. The original ISCS consists of 20 items, with 10 items each for bonding social capital and bridging social capital. Bonding social capital assesses aspects such as emotional support, access to limited resources, and the ability to mobilize solidarity. Bridging social capital evaluates aspects such as an outward-looking perspective, contact with a wide range of people, a sense of belonging to a larger group, and interactions with broader communities. In this study, the shortened scale with 5 items each for bonding and bridging social capital, totaling 10 items, was used. All items were rated on a 4-point Likert scale from "Strongly disagree" (1) to "Strongly agree" (4) and converted to a 100-point scale, with a Cronbach α of 0.863 confirming reliability.

Control Variables

The control variables included sociodemographic characteristics such as gender, age, education level, disability status, household type, and monthly income. Age was categorized into 65 - 69, 70 - 74, 75 - 79, and 80 years or older. Education level was classified as lower than elementary school, middle school graduate, high school graduate, and college graduate or higher. Disability status was binary ("yes" or "no"), and household type was classified as either single-person households or households with 2 or more people. Monthly income was divided into 5 quintiles based on income levels.

Data Analysis

The data were analyzed using the R statistical program, version 4.4.1, and the specific analysis methods are as follows. First, descriptive statistical analyses, including frequency analysis, percentages, means, and SDs, were conducted to understand the participants' sociodemographic characteristics, digital literacy level, life satisfaction, and social capital. Second, the Pearson correlation analyses were performed to identify the associations between the main variables of the study model: digital literacy, life satisfaction, and social capital. Third, to verify the mediating effect of social capital in the association between digital literacy and life satisfaction, a 3-step multiple regression analysis based on Baron and Kenny's procedure [29] was conducted. The bootstrap method was applied to test the statistical significance of mediation effects, using a 95% CI.

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki and received review and approval from the Institutional Research Ethics Committee of the Yonsei University Institutional Review Board in Wonju, South Korea (Approval No.1041849 - 202410-SB-213-01). All procedures followed the relevant institutional guidelines and regulations, and informed consent was obtained from all participants or their legal guardians before participation. To ensure privacy and confidentiality, the data used in this study underwent anonymization and deidentification processes.

Results

Participants Characteristics

A total of 869 participants were included in this study (Table 1). There were marginally more women (n=450, 51.8%) than men (n=419, 48.2%). By age, the largest group included those aged 65 - 69 years old (n=367, 42.2%), followed by those aged 70 - 74 years old (n=296, 34.1%), 75 - 79 years old (n=164, 18.9%), and 80 years and older (n=42, 4.8%). Regarding education level, high school graduates accounted for the largest proportion (n=345, 39.7%), followed by middle school graduates (n=289, 33.3%), those with elementary school education or lower (n=182, 20.9%), and college graduates or higher (n=53, 6.1%). In terms of disability status, most respondents (n=855, 98.4%) did not have a disability, while 1.6% (n=14) reported having a disability. Household type showed that 18.2% (n=158) lived in single-person households, while 81.8% (n=711) lived in households with 2 or more people. Regarding monthly income, the largest group earned ₩2 - ₩2.99 million (US $1392.30-$2081.49) (n=233, 26.8%), followed by ₩1 - ₩1.99 million (US $696.15-$1385.34) (n=216, 24.9%), ₩4 million or more (US $2784.61) (n=188, 21.6%), ₩3 - ₩3.99 million (US $2088.45-$2777.64) (n=152, 17.5%), and less than ₩1 million (US $696.15) (n=80, 9.2%).

XSL•FO
RenderX

Table . Respondents' characteristics and degree of study variables.

Characteristic	Participants (n=869)
Gender, n (%)	
Male	419 (48.2)
Female	450 (51.8)
Age (years), n (%)	
65 - 69	367 (42.2)
70 - 74	296 (34.1)
75 - 79	164 (18.9)
≥80	42 (4.8)
Education level, n (%)	
Elementary school or lower	182 (20.9)
Middle school	289 (33.3)
High school	345 (39.7)
College or higher	53 (6.1)
Disability, n (%)	
No	855 (98.4)
Yes	14 (1.6)
Live with someone, n (%)	
No	158 (18.2)
Yes	711 (81.8)
Income, n (%)[a]	
<1 million	80 (9.2)
1 - 1.99 million	216 (24.9)
2 - 2.99 million	233 (26.8)
3 - 3.99 million	152 (17.5)
≥4 million	188 (21.6)
Digital literacy[b], mean (SD)[c]	56.2 (10.5)
Digital access	62.4 (16.6)
Digital competency	47.2 (16.5)
Digital utilization	62.2 (6.2)
Life satisfaction, mean (SD)[c]	64.9 (11.0)
Social capital, mean (SD)[c]	67.9 (12.0)

[a]Income reported in Korean won (₩1=US $0.00070) and categorized into ranges.

[b]Digital literacy: the sum of access, competency, and utilization, with respective weights of 0.2, 0.4, and 0.4.

[c]All scores presented on a 100-point scale.

The mean digital literacy score was 56.2 (SD 10.5), with the subcomponents showing the following mean scores: 62.4 (SD 16.6) for digital access; 47.2 (SD 16.5) for digital competency; and 62.2 (SD 6.2) for digital utilization. The mean life satisfaction score was 64.9 (SD 11.0), and the mean social capital score was 67.9 (SD 12.0).

Correlation Analysis Between Life Satisfaction, Digital Literacy, and Social Capital

The results of the Pearson correlation analyses, conducted to examine the correlation between digital literacy, social capital, and life satisfaction, are presented in Table 2. Life satisfaction showed a statistically significant positive correlation with both digital literacy (r=0.276, P<.001) and social capital (r=0.438, P<.001), with a stronger correlation observed between life satisfaction and social capital. This indicates that higher levels

XSL•FO
RenderX

of digital literacy and social capital are associated with greater life satisfaction. Additionally, digital literacy showed a significant positive correlation with social capital (r=0.216, P<.001), indicating that as digital literacy increases, social capital tends to increase as well.

Table . Correlation analysis of life satisfaction, digital literacy, and social capital.

Variables	Life satisfaction	Digital literacy	Social capital
Life satisfaction			
r	—[a]	0.276	0.438
P value	—	*P*<.001	*P*<.001
Digital literacy			
r	0.276	—	0.216
P value	*P*<.001	—	*P*<.001
Social capital			
r	0.438	0.216	—
P value	*P*<.001	*P*<.001	—

[a]Not applicable.

The Mediating Effect of Social Capital on the Association Between Digital Literacy and Life Satisfaction

Following the procedure proposed by Baron and Kenny [29], a 3-step multiple regression analysis was conducted to examine the effects of digital literacy on life satisfaction and the mediating role of social capital (Table 3), controlling for gender, age, education level, disability status, and income level.

XSL•FO
RenderX

Table . The mediating effect of social capital on the association between digital literacy and life satisfaction.[a]

Variables	Model 1[b]		Model 2[c]		Model 3[d]	
	β (SE)	*t* (*P*[e])	β (SE)	*t* (*P*[e])	β (SE)	*t* (*P*[e])
Gender						
Female (reference: male)	1.253 (0.818)	1.531 (.13)	−0.276 (0.736)	−0.375 (.71)	−0.699 (0.684)	−1.022 (.31)
Age (years)						
70 - 74 (reference: 65 - 69)	−1.324 (0.934)	−1.417 (.16)	−1.342 (0.840)	−1.597 (.11)	−0.896 (0.780)	1.148 (.25)
75 - 79 (reference: 65 - 69)	0.172 (1.171)	0.147 (.88)	0.138 (1.053)	0.131 (.90)	0.080 (0.977)	0.082 (.93)
≥80 (reference: 65 - 69)	−0.477 (1.940)	−0.246 (.81)	−0.287 (1.745)	−0.164 (.87)	−0.1255 (1.619)	−0.078 (.94)
Education level						
Middle school (reference: ≤ elementary school)	3.213 (1.173)	2.738 (.006)	0.745 (1.055)	0.706 (.48)	−0.339 (0.983)	−0.345 (.73)
High school (reference: ≤ elementary school)	4.244 (1.285)	3.302 (.001)	1.686 (1.156)	1.459 (.15)	0.255 (1.079)	0.236 (.81)
≥ College (reference: ≤ elementary school)	5.206 (2.084)	2.497 (.01)	5.399 (1.875)	2.879 (.004)	3.643 (1.745)	2.087 (.04)
Disability						
Yes (reference: no)	1.038 (3.182)	0.326 (.74)	3.403 (2.862)	1.189 (.24)	3.053 (2.655)	1.150 (.25)
Live with someone						
Yes (reference: no)	0.059 (1.224)	0.048 (.96)	−0.363 (1.101)	−0.330 (.74)	−0.383 (1.021)	−0.375 (.71)
Income[f]						
1 - 1.99 million (reference:<1 million)	1.643 (1.655)	0.992 (.32)	2.579 (1.489)	1.732 (.08)	2.025 (1.382)	1.465 (.14)
2 - 2.99 million (reference:<1 million)	3.676 (1.775)	2.071 (.04)	5.642 (1.596)	3.534 (<.001)	4.402 (1.484)	2.965 (.003)
3 - 3.99 million (reference:<1 million)	4.765 (1.939)	2.458 (.01)	6.524 (1.744)	3.741 (<.001)	4.916 (1.623)	3.029 (.003)
≥4 million (reference:<1 million)	5.502 (1.962)	2.805 (.005)	7.371 (1.765)	4.177 (<.001)	5.515 (1.644)	3.354 (<.001)
Digital literacy	0.110 (0.046)	2.364 (.02)	0.140 (0.042)	3.350 (<.001)	0.103 (0.039)	2.647 (.008)
Social capital	—[g]	—	—	—	0.337 (0.029)	11.81 (<.001)

[a]All models were adjusted for control variables, including gender, age, education level, disability status, household type, and income level.

[b]Model 1: effect of digital literacy on social capital (constant=54.969; R^2=0.096; adjusted R^2=0.082; F_{852}=6.51; P<.001).

[c]Model 2: effect of digital literacy on life satisfaction (constant=51.755; R^2=0.133; adjusted R^2=0.119; F_{852}=9.38; P<.001).

[d]Model 3: effect of digital literacy and social capital on life satisfaction (constant=33.214; R^2=0.255; adjusted R^2=0.242; F_{851}=19.48; P<.001).

[e]*P* values <.05 are considered significant.

[f]Income reported in Korean won (₩1=US $0.00070) and categorized into ranges.

[g]Not applicable.

XSL•FO
RenderX

Model 1 analyzed the effect of digital literacy on social capital. Digital literacy showed a statistically significant positive effect on social capital (β=0.110; *P*=.02), indicating that higher levels of digital literacy tend to increase social capital. Moreover, higher education and income levels were positively associated with social capital.

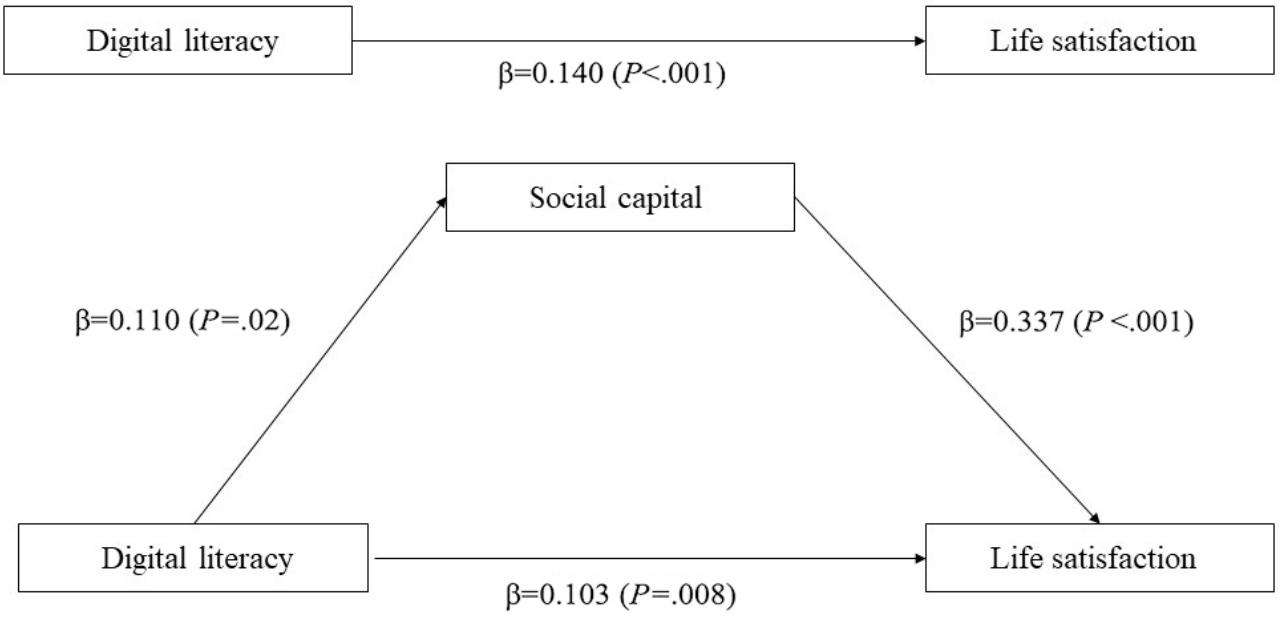
Model 2 examined the effect of digital literacy on life satisfaction. Digital literacy had a statistically significant positive effect on life satisfaction (β=0.140; *P*<.001). Higher life satisfaction was observed among those with a college degree or higher (β=5.399; *P*=.004) and those with a monthly income of ₩2 - ₩2.99 million (β=5.642; *P*<.001), ₩3 - ₩3.99 million (β=6.524; *P*<.001), and ₩4 million or more (β=7.371; *P*<.001).

Model 3 analyzed the effects of digital literacy and social capital on life satisfaction. Both digital literacy (β=0.103; *P*=.008) and social capital (β=0.337; *P*<.001) had statistically significant positive effects on life satisfaction, indicating that both factors contribute to enhancing life satisfaction. Additionally, higher life satisfaction was observed among those with a college degree or higher (β=3.643; *P*=.04) and those with a monthly income of ₩2 - ₩2.99 million (β=4.402; *P*=.003), ₩3 - ₩3.99 million (β=4.916; *P*=.003), and ₩4 million or more (β=5.515; *P*<.001).

The *F* and *P* values for each model (models 1, 2, and 3) were statistically significant (*P*<.001), indicating that the models fit the data well. Notably, model 3 demonstrated the highest explanatory power, with an adjusted R^2 of 0.242, when social capital was included.

The regression coefficient of the mediating variable, social capital, was statistically significant, and the direct effect of the independent variable, digital literacy, on the dependent variable, life satisfaction (β=0.140), decreased when social capital was included as a mediator (β=0.103). This confirms that social capital had a partial mediating effect on the relationship between digital literacy and life satisfaction (Figure 1). Thus, social capital played a key role in enhancing life satisfaction, and digital literacy positively influenced life satisfaction through social capital.

Figure 1. Association between digital literacy, social capital, and life satisfaction.



Verification of the Significance of the Mediating Effect of Social Capital

Table 4 presents the results of the bootstrapping analysis performed to verify the statistical significance of the mediating effect of social capital. First, the direct effect of digital literacy on life satisfaction, without mediation by social capital, was 0.103 (*P*=.006), confirming that digital literacy had a significant direct effect on life satisfaction. The indirect effect, representing the effect of digital literacy on life satisfaction through social capital, was 0.037 (*P*=.03). The total effect, which combines the direct and indirect effects, was 0.140 (*P*=.002), indicating that the overall impact of digital literacy on life satisfaction was statistically significant. Thus, digital literacy was an important factor that positively influenced life satisfaction, both directly and indirectly, through the mediating role of social capital.

XSL•FO
RenderX

Table . Verification of the significance of the mediating effect of social capital.[a]

Mediation pathway	β (95% CI)	*P* value[b]
Indirect effect		
Digital literacy → Social capital → Life satisfaction	0.037 (0.005-0.070)	.03
Direct effect		
Digital literacy → Life satisfaction	0.103 (0.030-0.180)	.006
Total effect	0.140 (0.058-0.230)	.002

[a]Results are based on bootstrapping simulations (n=5000).

[b]*P* values <.05 are considered significant.

Discussion

Principal Findings

This study analyzed the effect of digital literacy on life satisfaction among older adults in Korea and investigated the mediating role of social capital in this relationship. The results showed that digital literacy has a positive effect on life satisfaction. These findings align with prior research indicating that higher levels of digital access, competency, and utilization are associated with greater life satisfaction [16,18,30]. Studies on older adults have highlighted the internet's role as a valuable tool in daily life, supporting interpersonal communication, information searches, task completion, and leisure activities, all of which contribute positively to life satisfaction [31,32]. These findings underscore the internet as a crucial resource that empowers older adults to lead independent and active lives. This study reinforces these findings by demonstrating that a higher level of digital literacy positively influences life satisfaction among older adults in Korea.

The COVID-19 pandemic accelerated the digitalization of public services in Korea, significantly increasing reliance on digital platforms for essential services, including health care, administration, and finance. Although Korea is known for its advanced digital infrastructure, older adults face new challenges in accessing these essential services. This often leads to delays in medical care and limited access to health information, increasing the risk of adverse health outcomes [33]. Moreover, the rapid expansion of financial digitalization has exposed older adults with low digital literacy and limited decision-making abilities to economic risks, such as financial fraud [34]. According to a report comparing OECD countries, Korea's younger population (aged 25 - 34 years) demonstrates digital proficiency levels higher than the OECD average, whereas the older population aged 55 - 65 years exhibits proficiency levels below the OECD average. Consequently, Korea is identified as having one of the largest generational digital proficiency gaps among OECD countries [35].

Public and private organizations in Korea have provided digital literacy education for older adults, including training in computer and smartphone usage. However, these programs have predominantly taken the form of one-to-many classroom-style lectures focused on basic skills, which may not sufficiently address the diverse needs of older adults [36]. To better meet these needs, it is recommended to first assess digital literacy levels to determine older adults' specific digital proficiency levels and then provide tailored training based on these levels. For beginners, training on basic smartphone usage and sending text messages would be appropriate, while intermediate and advanced courses could cover digital health management, internet-based financial education, and the creation of digital photo albums.

Such tailored education can also be observed in international examples. The United Kingdom's Mi WiFi project (2017 - 2018) provided smart devices and digital education to vulnerable groups, including older adults and individuals with disabilities, helping them acquire basic digital skills. Similarly, the United States' Cyber Seniors project involved young mentors from volunteer organizations visiting older adults in their homes or facilities to provide one-to-one digital education, significantly enhancing their digital competency [37]. The older adult population differs widely in terms of income, education level, and health status. Therefore, digital education programs should be designed to reflect the diverse characteristics and needs of these groups.

In this study, social capital was found to have a positive impact on older adults' life satisfaction, consistent with previous studies' findings [38-41]. As social networks shrink and participation opportunities diminish with age, maintaining or acquiring social capital becomes increasingly challenging [42]. Social capital is often divided into 2 dimensions: bonding social capital and bridging social capital [43]. Bonding social capital refers to emotional support, access to resources, and a sense of solidarity formed within close relationships, such as family, friends, and neighbors. When older adults have networks from which they can obtain resources or receive emotional support, their life satisfaction significantly improves. Meanwhile, bridging social capital is formed through interactions with a wider community and diverse individuals, which helps reduce social isolation and contributes to higher life satisfaction among older adults.

These dimensions of social capital are shaped by components such as networks, social participation, norms, and social trust. Among these, social trust in family, friends, neighbors, and institutions was identified as the most important factor affecting life satisfaction in older adults [44]. By strengthening relationships with family and the local community, social trust contributes to psychological stability and enhances the quality of life for older adults. This highlights the need for

XSL•FO
RenderX

community-based support environments that consider the unique characteristics of older adults, enabling them to build and maintain social capital. Communities should provide programs and spaces that allow older adults to share their thoughts and experiences, fostering meaningful relationships. Such environments are essential to reinforcing social capital and improving the life satisfaction of older adults.

In this study, social capital showed a partial mediating effect on the relationship between digital literacy and life satisfaction among older adults. This implies that digital literacy not only directly enhances life satisfaction but also indirectly influences it through social capital. These findings are consistent with those of previous research showing that social capital mediates the relationship between digital utilization and life satisfaction among individuals aged 55 years and older [45]. Older adults with higher levels of digital literacy tend to build social relationships and participate in voluntary social activities through internet-based platforms, leading to improvements in their life satisfaction [46-48]. For example, Yoon et al [49] demonstrated that during the COVID-19 pandemic, a pilot program using digital platforms for social prescriptions effectively expanded the social networks of Korean older adults and improved their life satisfaction. South Korean society continues to emphasize family-centered values, with older adults particularly prioritizing their relationships with family members. Maintaining connections with family and friends via social networking services or remote communication positively impacts psychological well-being and overall quality of life [50,51].

Recently, some local governments in Korea have proposed tailored support measures for older adults, categorized based on their digital competency levels and specific needs. For instance, support programs have been planned for groups requiring digital tools for daily life, those seeking social interactions, and those pursuing self-development and cultural enrichment [52]. These programs aim to enable older adults to effectively utilize digital tools, foster social interactions, and mitigate social isolation. Ultimately, it is essential to implement practical support measures that expand social capital in a digital environment and enable older adults to lead more independent and fulfilling lives.

Limitations and Future Directions

This study has the following limitations. First, as a cross-sectional study, it cannot track changes in causal relationships between variables over time. Future studies should utilize longitudinal data to better understand how relationships between variables change over time. Second, this study focuses on older adults in Korea, which limits the generalizability of the results to other countries or groups with different cultural backgrounds. The relationship between digital literacy and social capital formation may differ across countries, highlighting the need for international comparative studies. Third, most of the study participants (855/869, 98.4%) were not disabled and belonged to the upper-middle-income group, which may restrict the generalizability of the findings. Therefore, the results should be interpreted with caution.

Nevertheless, this study is significant as it used nationwide raw data from the Korean MSIT, analyzing data representative of the older adult population in Korea. It emphasizes the importance of improving digital literacy and expanding social capital to enhance the quality of life of older adults in Korea.

Additionally, future research should explore the distinct impacts of the components of digital literacy—digital access, competency, and utilization—in greater detail. Particularly, examining specific aspects of digital literacy, such as media utilization skills, online safety awareness, and information-seeking behaviors, could help identify their relative effects on life satisfaction among older adults. This analysis would provide critical insights for effectively reducing the digital divide among older adults and designing tailored digital education and support policies to meet their specific needs.

Conclusion

This study analyzed the associations between digital literacy, social capital, and life satisfaction among older adults in Korea. The results showed significant correlations among digital literacy, social capital, and life satisfaction, with social capital mediating the relationship between digital literacy and life satisfaction. These findings underscore that digital literacy, when combined with social capital, contributes to enhancing the life satisfaction of older adults. Therefore, improving digital literacy and expanding social capital are crucial for promoting the psychological well-being and quality of life of older adults. Furthermore, strengthening digital literacy in the rapidly aging society of Korea can not only improve the quality of life for older adults but also help bridge generational gaps and foster greater social cohesion.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (2024S1A5A2A03038024).

Data Availability

The datasets generated and analyzed during this study were collected by the Ministry of Science and Information and Communication Technology (MSIT) and the Korean National Information Society Agency and are available on the agency's website [5].

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. 2024 elderly statistics press release [Web page in Korean]. Statistics Korea. 2024. URL: https://kostat.go.kr/board.es?mid=a10301010000&bid=10820&act=view&list_no=432917 [accessed 2025-01-22]
2. The 5th National Health Plan (Health Plan 2030, 2021-2030) [Web page in Korean]. Ministry of Health and Welfare. 2021. URL: https://www.mohw.go.kr/board.es?mid=a10401000000&bid=0008&tag=&act=view&list_no=365340 [accessed 2025-01-22]
3. Ageing in a digital world – from vulnerable to valuable. : ITU Publications; 2021 URL: https://www.unapcict.org/sites/default/files/2021-05/Ageing%20in%20a%20digital%20world%20%20from%20vulnerable%20to%20valuable.pdf [accessed 2025-01-22]
4. Hou F, Cao L, Wang Q. Call for the ageing-inclusive digital economy and related standards: based on a global survey. Presented at: ICEME 2022; Jul 16-18, 2022; Beijing, China. [doi: 10.1145/3556089.3556175]
5. The 2023 Report on the Digital Divide [Report in Korean]. : The Ministry of Science and ICT; 2024 URL: https://www.nia.or.kr/site/nia_kor/ex/bbs/List.do?cbIdx=81623 [accessed 2025-01-27]
6. Haimi M, Sergienko R. Adoption and use of telemedicine and digital health services among older adults in light of the COVID-19 pandemic: repeated cross-sectional analysis. JMIR Aging 2024 Apr 24;7:e52317. [doi: 10.2196/52317] [Medline: 38656768]
7. Mace RA, Mattos MK, Vranceanu AM. Older adults can use technology: why healthcare professionals must overcome ageism in digital health. Transl Behav Med 2022 Dec 30;12(12):1102-1105. [doi: 10.1093/tbm/ibac070] [Medline: 36073770]
8. Choi EY, Kim Y, Chipalo E, Lee HY. Does perceived ageism widen the digital divide? And does it vary by gender? Gerontologist 2020 Sep 15;60(7):1213-1223. [doi: 10.1093/geront/gnaa066] [Medline: 32539099]
9. Mubarak F, Suomi R. Elderly forgotten? Digital exclusion in the information age and the rising grey digital divide. Inquiry 2022;59:469580221096272. [doi: 10.1177/00469580221096272] [Medline: 35471138]
10. Bawden D. Origins and concepts of digital literacy. In: Lankshear C, Knobel M, editors. Digital Literacies: Concepts, Policies and Practices: Peter Lang Publishing; 2008:17-32.
11. van Dijk JAGM. The Deepening Divide: Inequality in the Information Society: Sage; 2005. [doi: 10.4135/9781452229812]
12. Hasan H, Linger H. Enhancing the wellbeing of the elderly: social use of digital technologies in aged care. Educ Gerontol 2016 Oct;42(11):749-757. [doi: 10.1080/03601277.2016.1205425]
13. Zheng R, Spears J, Luptak M, Wilby F. Understanding older adults' perceptions of internet use: an exploratory factor analysis. Educ Gerontol 2015 Mar 9;41(7):504-518. [doi: 10.1080/03601277.2014.1003495]
14. Fowler JG, Gentry JW, Reisenwitz TH. Analyzing Chinese older people's quality of life through their use of the internet. Int J Consumer Studies 2015 Apr;39(4):324-334. [doi: 10.1111/ijcs.12194]
15. Welch V, Ghogomu ET, Barbeau VI, et al. Digital interventions to reduce social isolation and loneliness in older adults: an evidence and gap map. Campbell Syst Rev 2023 Dec;19(4):e1369. [doi: 10.1002/cl2.1369] [Medline: 38024780]
16. Kim SK, Shin HR, Kim YS. The effect of mobile device capability of middle and older aged adults on life satisfaction: focusing on the mediating effect of mobile social participation [Article in Korean]. J Digit Convergence 2020;18(3):23-34. [doi: 10.14400/JDC.2020.18.3.023]
17. Lee SH, Han M, Lee J, Kim S. The effect of digital literacy in the elderly on life satisfaction: serial multiple mediating effects of social participation activities and depression [Article in Korean]. KAIS 2023;24(6):406-416. [doi: 10.5762/KAIS.2023.24.6.406]
18. Lee H. Analysis of the impact of digital literacy on life satisfaction (2019-2022) for older adults in South Korea: a national community-based panel study. Sci Rep 2024 Sep 2;14(1):20399. [doi: 10.1038/s41598-024-71397-0] [Medline: 39223233]
19. Morris ME, Adair B, Ozanne E, et al. Smart technologies to enhance social connectedness in older people who live at home. Australas J Ageing 2014 Sep;33(3):142-152. [doi: 10.1111/ajag.12154] [Medline: 24730370]
20. Khosravi P, Rezvani A, Wiewiora A. The impact of technology on older adults' social isolation. Comput Human Behav 2016 Oct;63:594-603. [doi: 10.1016/j.chb.2016.05.092]
21. Coleman JS. Social capital in the creation of human capital. American Journal of Sociology 1988 Jan;94:S95-S120. [doi: 10.1086/228943]
22. Putnam RD. The prosperous community: social capital and public life. Am Prospect 1993;13:35-42 [FREE Full text]
23. Pénard T, Poussing N. Internet use and social capital: the strength of virtual ties. J Econ Issues 2010 Sep;44(3):569-595. [doi: 10.2753/JEI0021-3624440301]
24. Mason M, Sinclair D, Berry C. Nudge or Compel? Can Behavioural Economics Tackle the Digital Exclusion of Older People?: International Longevity Centre; 2012.
25. Nyqvist F, Forsman AK, Giuntoli G, Cattan M. Social capital as a resource for mental well-being in older people: a systematic review. Aging Ment Health 2013;17(4):394-410. [doi: 10.1080/13607863.2012.742490] [Medline: 23186534]

XSL•FO
RenderX

26. Lu N, Wu B. Perceived neighborhood environment, social capital and life satisfaction among older adults in Shanghai, China. Sci Rep 2022 Apr 23;12(1):6686. [doi: 10.1038/s41598-022-10742-7] [Medline: 35461348]
27. Larnyo E, Tettegah S, Griffin B, et al. Effect of social capital, social support and social network formation on the quality of life of American adults during COVID-19. Sci Rep 2024 Feb 1;14(1):2647. [doi: 10.1038/s41598-024-52820-y] [Medline: 38302613]
28. Williams D. On and off the 'net: scales for social capital in an online era. J Comp Mediated Comm 2006 Jan;11(2):593-628. [doi: 10.1111/j.1083-6101.2006.00029.x]
29. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986;51(6):1173-1182. [doi: 10.1037/0022-3514.51.6.1173]
30. Hwang HJ, Hwang YS. Gaps and reasons of digital divide within the elderly in Korea: focusing on household composition [Article in Korean]. Journal of Social Sciences 2017 Sep 30;24:359-386. [doi: 10.46415/jss.2017.09.24.3.359]
31. Seifert A, Schelling HR. Seniors online: attitudes toward the internet and coping with everyday life. J Appl Gerontol 2018 Jan;37(1):99-109. [doi: 10.1177/0733464816669805] [Medline: 27655237]
32. Lifshitz R, Nimrod G, Bachner YG. Internet use and well-being in later life: a functional approach. Aging Ment Health 2018 Jan;22(1):85-91. [doi: 10.1080/13607863.2016.1232370] [Medline: 27657190]
33. Han MH. Effects of digital information competency on life satisfaction among the elderly in the era of COVID-19 [Article in Korean]. KAIS 2023;24(2):559-567. [doi: 10.5762/KAIS.2023.24.2.559]
34. Kim J. Digital divide among older consumers and measures to enhance their digital literacy [Article in Korean]. J Consum Policy Stud 2022;53:27-54. [doi: 10.15723/jcps.53.3.202212.27]
35. Skills Matter: Additional Results from the Survey of Adult Skills: OECD Publishing; 2019. [doi: 10.1787/1f029d8f-en]
36. Kim SW, Lee SS. A study of the current status of older adults' digital literacy programs in public libraries and improvement plans [Article in Korean]. Korean Soc Inform Manag 2022;39(4):49-74. [doi: 10.3743/KOSIM.2022.39.4.049]
37. A study on the human rights situation of the elderly due to the digital divide [Report in Korean]. : National Human Rights Commission of Korea; 2022 URL: https://www.humanrights.go.kr/base/board/read?boardManagementNo=17&boardNo=7608724&menuLevel=3&menuNo=115 [accessed 2025-01-27]
38. Chen Y, Li W, Zhang X, Cheng H, Tian Y, Yang H. Association between social capital and quality of life in older adults with subjective cognitive decline: a cross-sectional study. Appl Nurs Res 2024 Feb;75:151773. [doi: 10.1016/j.apnr.2024.151773] [Medline: 38490797]
39. Kim BJ, Linton KF, Lum W. Social capital and life satisfaction among Chinese and Korean elderly immigrants. J Soc Work (Lond) 2015 Jan;15(1):87-100. [doi: 10.1177/1468017313504699]
40. Kim J, Kim SE. Social capital and life satisfaction of elderly who lives alone: focusing on regional difference in depression and public pension [Article in Korean]. Journal of Social Science 2018 Jul 30;29(3):267-284. [doi: 10.16881/jss.2018.07.29.3.267]
41. Kim MY, Jeon HJ. The effects of smartphone use on life satisfaction in older adults: the mediating role of participation in social activities [Article in Korean]. Korean J Gerontol Soc Welfare 2017;72(3):343-370. [doi: 10.21194/kjgsw.72.3.201709.343]
42. Jeon SN. The effects of elderly's depression and social capital on successful aging [Article in Korean]. JKSSCHE 2019 Dec 31;20(3):29-42. [doi: 10.35133/kssche.20191231.03]
43. Chung S, Sung MH. Relationship between social capital and life satisfaction: a comparison of three different age groups [Article in Korean]. Health Soc Welf Rev 2012 Dec;32(4):249-272. [doi: 10.15709/hswr.2012.32.4.249]
44. Putnam RD. Bowling alone: the collapse and revival of American community. Presented at: CSCW00: Computer Supported Cooperative Work; Dec 2-6, 2000; Philadelphia, PA, United States. [doi: 10.1145/358916.361990]
45. Yoon HJ, Shin HR, Kim YS. A study on the relationship between level of digital information usage and life satisfaction: the mediating effect of the social capital [Article in Korean]. Info Policy 2020;27(4):85-100. [doi: 10.22693/NIAIP.2020.27.4.085]
46. Lee BJ, Myung SH. The impact of perception of informatization and internet use on the formation of social networks among the elderly [Article in Korean]. J Korean Assoc Reg Inf Soc 2010;13(4):151-179. [doi: 10.22896/karis.2010.13.4.007]
47. Delello JA, McWhorter RR. Reducing the digital divide: connecting older adults to iPad technology. J Appl Gerontol 2017 Jan;36(1):3-28. [doi: 10.1177/0733464815589985] [Medline: 26084479]
48. Francis J, Kadylak T, Cotten SR, Rikard RV. When it comes to depression, ICT use matters: a longitudinal analysis of the effect of ICT use and mattering on depression among older adults. In: Stephanidis C, editor. HCI International 2016 – Posters' Extended Abstracts: Springer; 2016:301-306. [doi: 10.1007/978-3-319-40542-1_49]
49. Yoon HJ, Kim HK, Lim JA, et al. Evaluation of the effectiveness of a social prescribing pilot project for the elderly using a digital platform in the COVID-19 pandemic [Article in Korean]. Korean Public Health Res 2022;48(2):1-17. [doi: 10.22900/kphr.2022.48.2.001]
50. Kim MY. The effects of smartphone use on life satisfaction, depression, social activity, and social support of older adults [Article in Korean]. J Korea Acad Ind Coop Soc 2018;19(11):264-277. [doi: 10.5762/KAIS.2018.19.11.264]
51. Kwon MO, Lee KT. The effects of self-disclosure on online social capital and subjective well-being in online network service [Article in Korean]. J Korean Manag Assoc 2015;28(5):1485-1503 [FREE Full text]

XSL•FO
RenderX

52. Kim CN, Jo JH, Lee HS, Lee HS, Yoon MH, Lee MY. A study on strategies to enhance digital literacy among the elderly in Gyeonggi province [Report in Korean]. : Gyonggi Welfare Foundation; 2022 2022-04 URL: https://ggwf.gg.go.kr [accessed 2025-01-22]

Abbreviations

OECD: Organisation for Economic Co-operation and Development
MSIT: Ministry of Science and Information and Communications Technology
ISCS: Internet Social Capital Scale

Edited by A Mavragani; submitted 02.11.24; peer-reviewed by C Nnajiofor, M Gasmi; revised version received 23.12.24; accepted 24.12.24; published 06.02.25.

Please cite as:
Jung H, Lee H, Nam EW
Mediating Effect of Social Capital on the Association Between Digital Literacy and Life Satisfaction Among Older Adults in South Korea: Cross-Sectional Study
JMIR Form Res 2025;9:e68163
URL: https://formative.jmir.org/2025/1/e68163
doi:10.2196/68163

© Hyein Jung, Hocheol Lee, Eun Woo Nam. Originally published in JMIR Formative Research (https://formative.jmir.org), 6.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Addressing the “Black Hole” of Low Back Pain Care With Clinical Decision Support: User-Centered Design and Initial Usability Study

Robert S Rudin[1], PhD; Patricia M Herman[2], ND, PhD; Robert Vining[3], DC, DHSc

[1]RAND, 20 Park Plaza, Suite 910, Boston, MA, United States

[2]RAND, Santa Monica, CA, United States

[3]Palmer Center for Chiropractic Research, Palmer College of Chiropractic, Davenport, IA, United States

Corresponding Author:
Robert S Rudin, PhD
RAND, 20 Park Plaza, Suite 910, Boston, MA, United States

Abstract

Background: Low back pain (LBP) is a highly prevalent problem causing substantial personal and societal burden. Although there are specific types of LBP, each with evidence-based treatment recommendations, most patients receive a nonspecific diagnosis that does not facilitate evidence-based and individualized care.

Objectives: We designed, developed, and initially tested the usability of a LBP diagnosis and treatment decision support tool based on the available evidence for use by clinicians who treat LBP, with an initial focus on chiropractic care.

Methods: Our 3-step user-centered design approach consisted of identifying clinical requirements through the analysis of evidence reviews, iteratively identifying task-based user requirements and developing a working web-based prototype, and evaluating usability through scenario-based interviews and the System Usability Scale.

Results: The 5 participating users had an average of 18.5 years of practicing chiropractic medicine. Clinical requirements included 44 patient interview and examination items. Of these, 13 interview items were enabled for all patients and 13 were enabled conditional on other input items. One examination item was enabled for all patients and 16 were enabled conditional on other items. One item was a synthesis of interview and examination items. These items provided evidence of 12 possible working diagnoses of which 3 were macrodiagnoses and 9 were microdiagnoses. Each diagnosis had relevant treatment recommendations and corresponding patient educational materials. User requirements focused on tasks related to inputting data, and reviewing and selecting working diagnoses, treatments, and patient education. User input led to key refinements in the design, such as organizing the input questions by microdiagnosis, adding a patient summary screen that persists during data input and when reviewing output, adding more information buttons and graphics to input questions, and providing traceability by highlighting the input items used by the clinical logic to suggest a working diagnosis. Users believed that it would be important to have the tool accessible from within an electronic health record for adoption within their workflows. The System Usability Scale score for the prototype was 84.75 (range: 67.5 - 95), considered as the top 10th percentile. Users believed that the tool was easy to use although it would require training and practice on the clinical content to use it effectively. With such training and practice, users believed that it would improve care and shed light on the “black hole” of LBP diagnosis and treatment.

Conclusions: Our systematic process of defining clinical requirements and eliciting user requirements to inform a clinician-facing decision support tool produced a prototype application that was viewed positively and with enthusiasm by clinical users. With further planned development, this tool has the potential to guide clinical evaluation, inform more specific diagnosis, and encourage patient education and individualized treatment planning for patients with LBP through the application of evidence at the point of care.

(JMIR Form Res 2025;9:e66666) doi:10.2196/66666

KEYWORDS

low back pain; clinical decision support; user-centered design; usability; back pain; low back pain care; pain; clinical decision; societal burden; substantial; burden; evidence-based; treatment; diagnosis; support tool; clinicians; chiropractic; chiropractor; reviews; scenario-based interviews; interviews

Introduction

Low back pain (LBP) is highly prevalent globally, causing personal and societal burden, chronic disability, and substantial health care expenditures, ranging from US $62 to US $124 billion annually in the United States [1-6]. Most LBP is caused by 1 or more benign conditions that are not associated with dangerous pathology [7]. Instead, LBP is influenced by injury, inflammation, and dysfunction of myofascial tissues, joints, and

XSL•FO
RenderX

nerve roots or peripheral nerves, with symptoms also influenced by nonphysical factors (eg, psychological, social, and environmental) and central sensitization processes [8]. However, despite extensive evidence in the literature informing diagnoses and treatments for specific types of LBP, nonpathological LBP is typically referred to as nonspecific [7]. The nonspecific LBP label lacks details to offer a working understanding of a patient's condition or inform personalized treatment plans. Furthermore, some diagnoses for common "specific" and neurologically related conditions, such as stenosis, neurogenic claudication, radiculopathy, and sciatica, are ill-defined, overlapping, and often used incorrectly, contributing to the challenge of addressing LBP with existing clinical evidence [9]. The uncertainty that pervades LBP diagnosis and management also causes patients substantial distress [10].

To help distill the evidence for LBP and make it more widely available in clinical practice, one of the authors (RV) conducted prior work that defines an evidence-based examination for LBP in a chiropractic research setting [11]; systematically reviewed the literature of LBP diagnostic studies, which recommends standardizing diagnostic terminology for common conditions contributing to LBP [9]; and developed a revised office-based clinical examination [12] and paper-based clinical decision aid that coordinates working diagnoses with evidence-based nonpharmacological treatments offered by doctors of chiropractic [12,13].

To further the goal of supporting clinicians in applying the latest evidence for LBP in busy clinical settings, we leveraged this prior work to develop an electronic clinical decision support (CDS) tool. Our goal was to produce a web-based application to support evidence-based in-office examination, diagnosis, treatment decisions, and patient education for nonpathological conditions contributing to LBP. We used an iterative, user-centered design process to identify clinical requirements from published evidence, elicit user requirements from iterative interviews, and develop and evaluate the usability of an electronic CDS tool that facilitates and encourages evidence-based diagnosis and treatment.

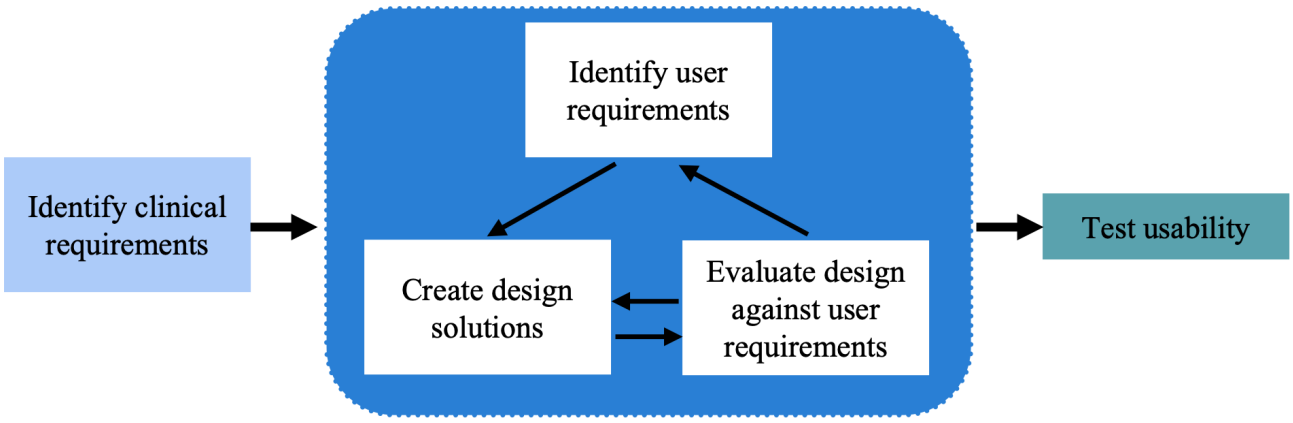
Methods

Overview

We conducted a three-step user-centered design process [14] to develop the CDS tool for LBP informed by best practices [15] in developing such tools: (1) identify clinical requirements based on our prior published reviews of the evidence for LBP diagnosis and treatment; (2) iteratively identify user requirements, create wireframes, and develop a working prototype; and (3) test usability as part of a preliminary assessment (Figure 1). Steps (1) and (2) occurred in June-August 2023, and Step (3) occurred in April and May 2024. We present the results of each step. User input was received during the second and third steps from 5 chiropractors serving clinical teaching faculty roles with Palmer College of Chiropractic. Participants were recruited from the College's main campus in Davenport, Iowa (n=4) and branch campus in Port Orange, Florida (n=1). Each campus maintains teaching clinics staffed by licensed chiropractic doctors who provide care and oversee senior student clerkships. Faculty were recruited via an institutional review board (IRB)-approved email that offered information about the study, including participation requirements, and a request for a reply if interested. Those interested were sent additional information and scheduled for an interview.

Participants were targeted for recruitment as part of a purposive sampling strategy with the intention of including variation in terms of users by sex and experience. We intentionally targeted clinicians with more extensive experience in clinical practice and teaching, rather than novice users, to gain expert insights into usability and workflow issues in real-world settings and educational contexts. This approach helps ensure that the tool meets the demands and complexities of their working environments and builds credibility and acceptance among the clinical community. Users were recruited via email invitation; all invited users agreed to participate. Users verbally consented before each session. All user sessions were audio- and video-recorded.

Figure 1. User-centered design process for low back pain decision support tool. After identifying relevant clinical findings from prior systematic literature reviews, we followed a standard, iterative process to identify and refine user requirements and design solutions before conducting a test of the usability using fictitious scenarios.



XSL•FO
RenderX

Clinical Requirements Identification

As described in the introduction, in prior work, we developed an evidence-based classification system for LBP, conducted a systematic review of LBP diagnostic studies, made recommendations for standardizing diagnostic terminology for common conditions contributing to LBP, and developed recommendations for LBP evaluation and treatment that were distilled in a series of publications including a paper-based clinical decision aid [9,11-13]. To identify clinical requirements for the CDS tool in this work, we reviewed our prior work to derive (1) clinical interview and examination inputs with adaptive characteristics that disable or enable items based on prior responses, (2) working diagnoses based on selected input items, and (3) evidence-based treatment recommendations corresponding to each working diagnosis. Furthermore, the clinician on the research team (RV) drafted patient education materials corresponding to each working diagnosis for user testing.

User Requirement Elicitation and CDS Tool Design and Development

We iteratively elicited user requirements and developed and refined our prototype through wireframing and then software programming. We developed a discussion guide covering topics that included current approaches used for diagnosing and treating LBP, workflows related to LBP, pain points, ideas for tool design, and reactions to the wireframes showing the input items and example diagnoses, treatment recommendations, and patient educational information. Two research team members with expertise in informatics and design (RSR) and LBP clinical practice and research (RV) interviewed each user. Prior to each first design session, users were sent prior published literature related to LBP diagnosis and treatment. Users were asked to review in advance as a reminder of current evidence related to LBP evaluation, diagnosis, and treatment, and to gain familiarity with key sources underlying the CDS tool. Users were encouraged to “think aloud” during the design sessions.

After each design session, 1 research team member (RSR) reviewed the recording and summarized key findings using conventional content analysis [16] and the framework method [17]. Categories of findings were related to user requirements, changes were needed to make the wireframes and prototypes address the emerging requirements, and new design ideas emerged during the sessions organized by user task. Between design sessions, we brainstormed or ideated new design ideas for the tool and worked with an experienced user interface designer to create additional wireframes. We then developed a fully functional prototype of the tool as a web-based application. A second research team member (RV) reviewed the recordings and confirmed or amended key findings. Results were finalized through consensus among research team members (RSR and RV).

Usability Test

To assess usability, 1 clinician-author (RV) developed 4 fictional clinical scenarios to help assess how the users would use the CDS tool in realistic situations. These involved a condition-specific history and examination findings for conditions matching criteria for acute nociceptive pain, chronic nociceptive pain, chronic nociceptive pain combined with symptoms of neurogenic claudication, and radicular pain. Each user was tested on 1 or 2 of these scenarios. At the start of each usability testing session, the user was given a brief overview of the tool and then asked to imagine that he or she was treating a patient with LBP. The user was then read a brief summary of a fictional patient’s LBP symptoms and was instructed to ask questions, as they would during a patient interview, to complete the input questions. The user was encouraged the “think aloud” as they used the CDS tool. After completing at least 1 scenario, the user was asked to complete the System Usability Scale (SUS), a 10-item measure that is commonly used to assess the usability [18]. SUS scores range from 0 to 100, with higher scores indicating greater usability. SUS scores >68 are considered above average, and those above 80 are in the top 10th percentile [19]. Users were encouraged to explain reasons for their answers and results were summarized [16].

Ethical Considerations

This study was reviewed and approved by the RAND Human Subjects Protection Committee (HSPC ID: 2023-N0175). The IRB at Palmer College of Chiropractic was the relying IRB through a formal reliance agreement. All written and audiovisual data (which were not anonymized) were stored on secure servers accessible only to the research team. Because the interviews did not address protected or sensitive topics, but instead focused on perceptions of a potential website and clinical workflows, no written consent was obtained. Instead, verbal consent was obtained prior to conducting interviews. Participants were compensated US $150 per interview, with a maximum of 2 interviews.

Results

User Characteristics

All users practiced chiropractic in teaching clinics at Palmer College of Chiropractic at either an Iowa or Florida campus location. See Table 1 for user characteristics. All encountered LBP on a daily basis when they treated patients. Users varied on how often they attempted to arrive at a specific diagnosis from rarely to every visit. One user recently switched to an administrative position and was no longer engaging in direct patient care.

XSL•FO
RenderX

Table . Characteristics of chiropractors at Palmer College of Chiropractic (N=5) who participated in user design sessions to assess the low back pain decision support tool.

Characteristic	Statistical values
Sex, n (%)	
Male	2 (40)
Female	3 (60)
Years practicing chiropractic, mean (range)	18.5 (7-34)
Location, n (%)	
Iowa	4 (80)
Florida	1 (20)
Patient visits per week, n	~100
Years using an EHR[a], mean (range)	10 (7 - 12)

[a]EHR: electronic health record.

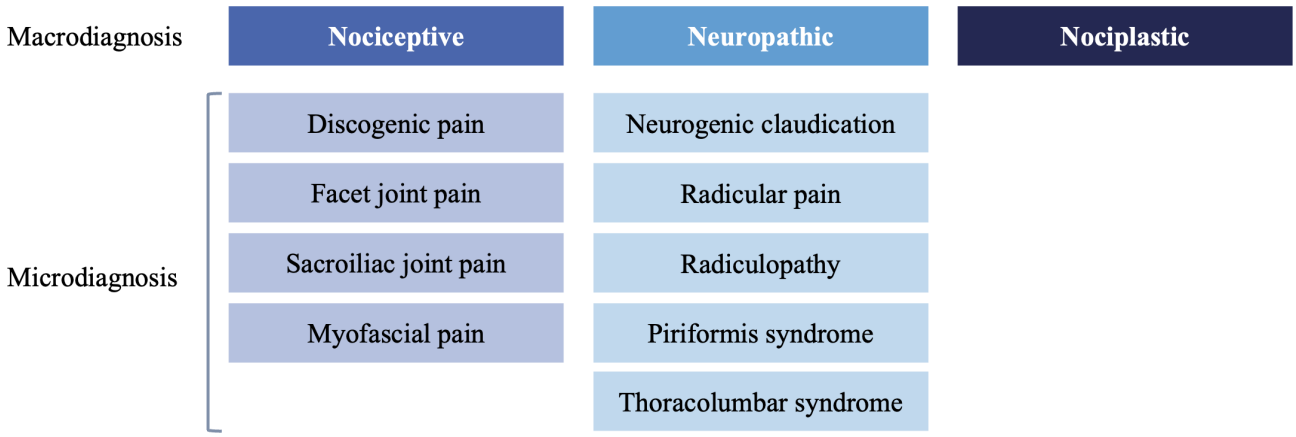
Clinical Requirements

We identified 14 possible working diagnoses divided into 3 macrocategories and 9 microcategories (Figure 2). For the input items and logic rules to enable them, we identified a total of 44 items: 13 interview items are enabled for all patients, 13 interview items are enabled conditional on other input items, 1 examination item is enabled for all patients, 16 examination items are enabled conditional on other input items, and 1 item is a synthesis of interview and examination findings and enabled for all patients. Items that were enabled for all patients helped identify macrodiagnoses; items that were conditionally enabled were related to microdiagnoses. For example, if the user selected "Sensory changes in a nerve root distribution," which offers some evidence for the macrodiagnostic category of neuropathic pain, 8 interview items and 6 examination items are enabled to clarify 1 or more potential microdiagnoses.

To specify the clinical logic rules for each working diagnosis, we used Boolean logic. For example, piriformis syndrome was included as a possible diagnosis based on the following logic combining the relevant input items: "radiating pain into an ipsilateral leg" OR "tenderness of the greater sciatic notch" OR "buttock pain" OR "positive straight leg raise test" OR "increased pain with prolonged sitting." The inputs and logic represent diagnostic information or criteria consistent with piriformis syndrome. If indicated by the literature, we included estimates of the strength of the evidence based on the quantity of inputs corresponding to a working diagnosis (eg, 1-5). Anchor terms of less likely (eg, 1) and more likely (eg, 5) are used because working diagnoses cannot be completely definitive.

Treatment recommendations were assigned to each working diagnosis based on published evidence. In many cases, 1 treatment recommendation applied to multiple diagnoses. For example, graded exposure and activity training are recommended for both neurogenic claudication and nociplastic pain. Some treatment recommendations apply to all working diagnoses (eg, education about condition and general exercise when safe and tolerated).

Figure 2. Types of low back pain. The literature identifies 3 macrodiagnoses and 9 microdiagnoses. These categories served as outputs of the decision support tool for low back pain.



User Requirements and CDS Tool Design

Example wireframes used in design sessions are shown in Figure S1 in Multimedia Appendix 1. Users were positive about the potential value of the CDS tool: "This tells you exactly where to focus your attention on. I like it" (User 3). Users said that the primary use case for using the tool would be when a new patient presents with LBP. Some might also use it if treatment was not working for reassessing their diagnoses and treatment plan and informing the potential for a referral. All users

XSL•FO
RenderX

suggested that it would be useful both for practicing chiropractors and as an education tool.

User requirements and design features are shown in Table 2 organized by user task. User input led to key refinements in the design, such as organizing the input questions by microdiagnosis, adding a patient summary screen that persists during data input and when reviewing the output, adding more information buttons and graphics to input questions, and providing traceability by highlighting the input items used by the clinical logic to suggest a working diagnosis.

The prototype CDS tool was built as a web-based application using the React javascript library and responsive design. For this prototype, all of the data entered by the user are stored within the user's web browser—no database or server-side storage of data were used (Figure 3). Screenshots of the tool are shown in Figure 4 and Figures S2 and S3 in Multimedia Appendix 1.

XSL•FO
RenderX

Table . User requirements and design features by task for the low back pain decision support tool.

Task	User requirement	Design feature
Input interview and examination findings	Make it easy to input the relevant data.	Order interviews items before examination items to parallel clinical workflows; disable input items for microdiagnoses until they are indicated based on items for more general diagnoses; and cluster input items based on microdiagnosis.
	Allow the user to learn more about data that need to be inputted if desired.	Information buttons and graphics.
	Show a summary of the interview/examination findings as user enters them.	Patient summary section.
Review potential working diagnoses	Allow the user to see all potential working diagnoses, which reflects the multifactorial nature of LBP[a], with relative strength of evidence based on interview/examination findings.	List diagnoses with likeliness scales to enable rapid visual confirmation of the extent of evidence supporting each condition, allowing users to confirm data accuracy and determine whether the interview and examination data align with the overall clinical presentation.
	For any specific working diagnosis, allow the user to view the specific interview/examination findings that constitute the evidence for it.	When user hovers over a diagnosis, highlight the related interview/examination findings in patient summary.
	Allow the user to go back and change/alter the interview/examination findings at any time to alter the potential working diagnoses.	Navigation arrow and button to return to the input screen.
Select potential working diagnoses	Allow the user to select/deselect from the list of working diagnoses based on the user's clinical assessment.	Toggles for each diagnosis.
	Allow the user to document his or her selected working diagnoses in his or her notes.	Allow copy-paste; automated input into EHR[b,c].
Save data for next visit	Allow the user to save his or her selected working diagnoses and interview/examination findings for review in future.	Save data tied to visit in EHR or database[b].
Review treatment recommendations	Allow the user to see all evidence-based treatment recommendations for selected working diagnoses.	List treatments based on evidence-based guidelines for diagnoses.
	For any specific treatment recommendation, allow the user to see which diagnoses he or she corresponds to.	List diagnoses below each treatment recommendation.
Determine treatment recommendations	Allow the user to select/deselect and customizing treatment recommendations based on the user's clinical assessment and experience with the various options.	Toggles for each treatment recommendation[b].
Select relevant patient educational information	Allow the user to easily print the patient educational information.	Make educational information available in PDF or other printable format.
	Allow the user to select/deselect and adapt patient educational information.	Select buttons to turn on/off specific contents[b].
	Allow the user to tailor the patient educational information based on patient and user preferences.	Let user edit content for each patient and configuration option for default content[b].
Global	Allow the user to access the tool within the EHR.	Implement CDS[d] tool within patient chart in EHR[b].

[a]LBP: low back pain.

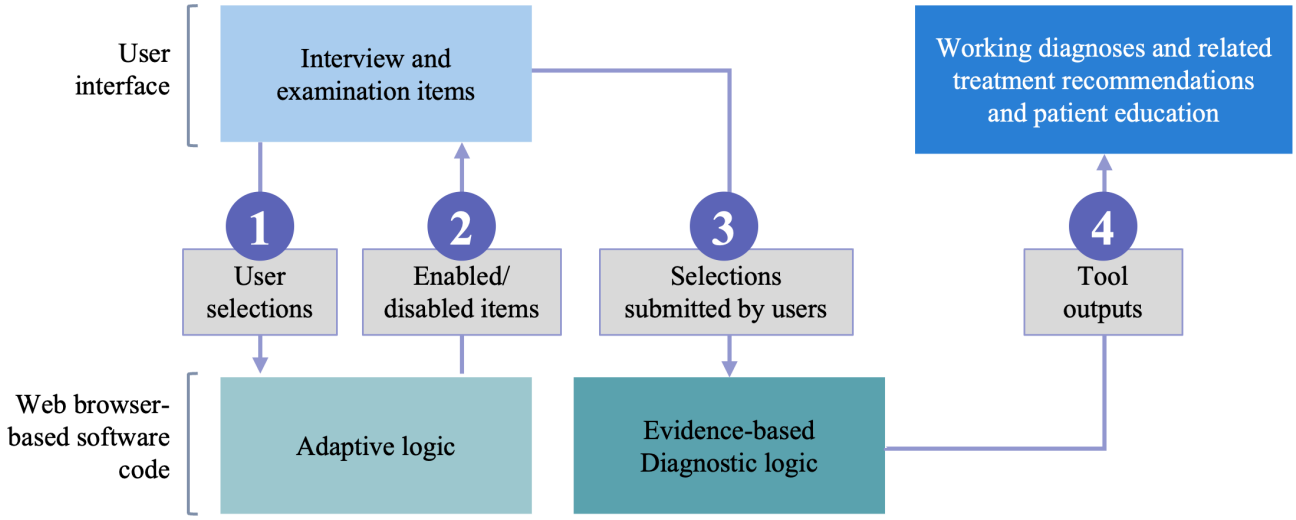
[b]Not yet implemented in prototype CDS tool.

[c]EHR: electronic health record.

[d]CDS: clinical decision support.

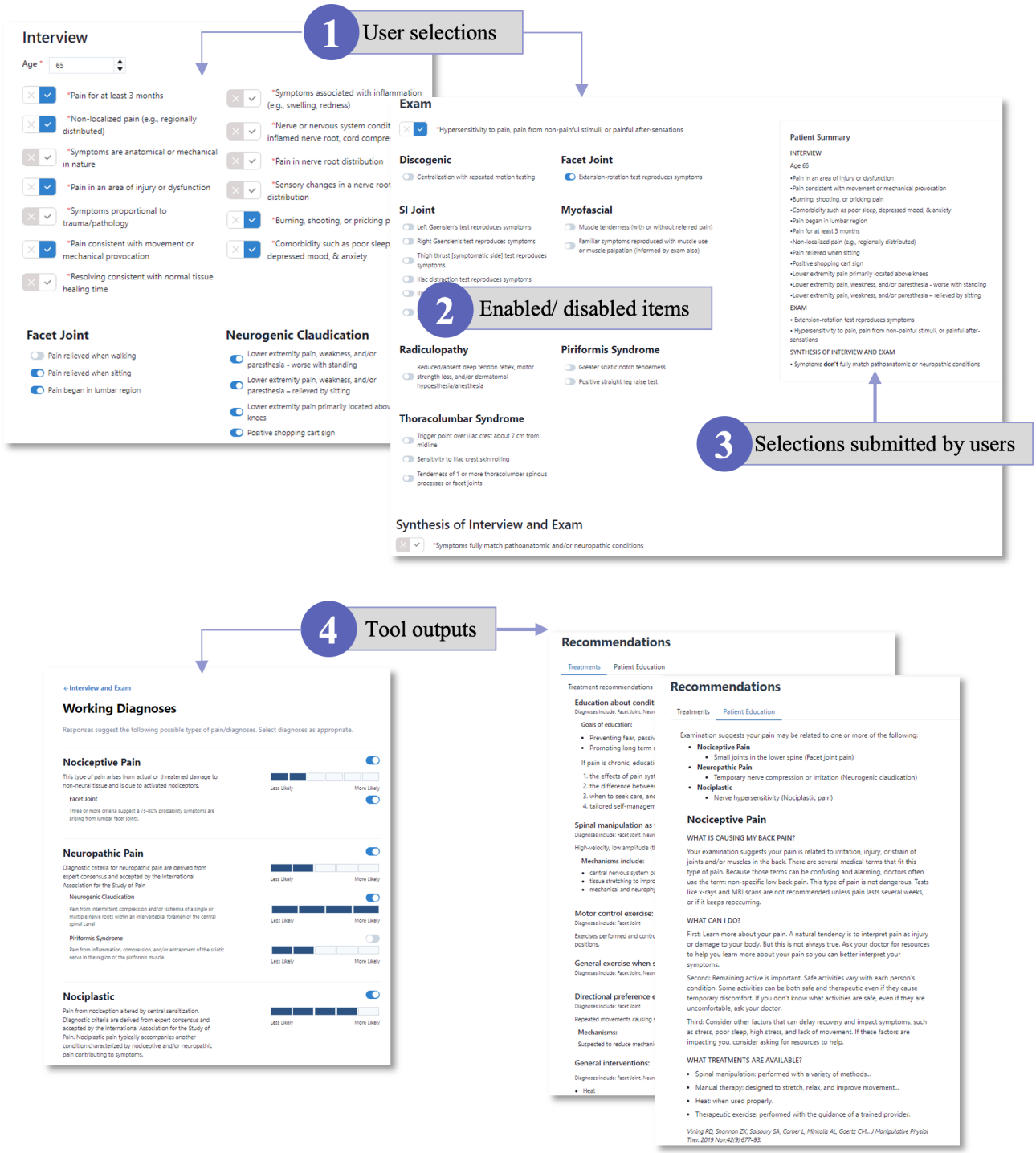
XSL•FO
RenderX

Figure 3. Technical architecture of web-based application for low back pain decision support. The sequence 1 - 4 shows each step that occurs as the user interacts with the tool.



XSL•FO
RenderX

Figure 4. Complex multifactorial diagnosis with evidence for 3 macrodiagnoses of low back pain (nociceptive, neuropathic, and nociplastic pain). Screenshots of the tool show (1) user selections based on the interview and examination, with (2) items disabled according to the adaptive logic in the model. The greatest evidence suggests neurogenic claudication with overlying nociplastic pain. Piriformis syndrome was rejected by the model because symptoms overlap with neurogenic claudication and because historical and examination evidence is stronger for neurogenic claudication. The (3) summary panel to the right, shown only in the top right screenshot, shows the selections as submitted by the user. The bottom screenshots show the tool outputs (4), with the result of the evidence-based diagnostic logic displaying the working diagnoses (left) and recommendations for treatments and patient education (right).



Usability Test

The mean score of the SUS was 84.75 (range 67.5 - 95). Qualitative results explained these findings (Table 3). Overall, users were enthusiastic about using the CDS tool because they perceived that it would help them focus on the most relevant information, especially for more complex patients. All users thought that the tool was easy to use from a technical perspective but required training and practice on the clinical content to use it effectively and efficiently. While using the tool during a study session, one user quickly transitioned from an initial impression that substantial effort would be needed to use it to expressing enthusiasm for the tool's potential use for improving back pain diagnosis and treatment: "This (tool) is like walking into a storm, but it wouldn't be like that the second time and certainly by the

XSL•FO
RenderX

50th time it would be fabulous" (User 4). One user said that the training needed to use the tool effectively was a good thing because that training would improve care for LBP—they strongly agreed with the statement from the SUS that said, "I needed to learn a lot of things before I could get going with this system" (User 3), but added that they meant that "in a positive connotation way."

Users also reacted to the usability of specific features. When using the input page, user reactions included the following: enabling examination items only when indicated could help eliminate tests that might be done routinely but are not clinically relevant; the patient summary was useful for guiding the input process; the wording or clinical intention behind some items was difficult to understand (eg, "symptoms fully match pathoanatomic and/or neuropathic condition;" and "resolving consistent with normal tissue healing time"); and it might be difficult to have time to go through all items during a visit.

When using the list of working diagnoses, reactions from users included positive views of the feature that highlighted relevant input items when hovering over a diagnosis to show the inputs (ie, evidence) supporting it; clarity of the scale used to indicate strength of evidence; and helpfulness of having a list of diagnoses to prevent anchoring on one and also for explaining the clinical reasoning process to patients. Reactions to the treatment recommendations were generally positive and believed to be helpful for decision-making. Two users noted that they had preferences for some treatment options more than others based on what they were familiar with.

For the patient education materials, 3 users emphasized the importance of being able to print out copies. Two believed that the materials would be acceptable to patients while 3 found them too complex especially for those with low health literacy levels. Two users preferred handwriting education information to make it more personal. One preferred to print out the materials and highlight key information.

Users also described how they would likely use this CDS tool in their workflows. All users believed that it was important to make the tool available within the EHR to allow for easy access to it and to minimize the need to switch between applications or devices. All users would use the CDS tool while with the patients but 2 users mentioned that they might leave the room to use the tool separately. One suggested that a freestanding web-based version of the tool maybe useful as well in some scenarios. Two users wanted to modify their patient intake questions to match the CDS tool's input items to help the clinician prepare for the visit and facilitate tool usage.

Users suggested several ideas to improve use and usability including use of colors on the input screen to break up the text more, autopopulating the CDS tool based on patient intake questions, providing a crosswalk between generated working diagnoses and *International Classification of Diseases, Tenth Revision* (*ICD-10*) codes to facilitate billing, indicating in the input screen if the combination input items selected did not make clinical sense and should be revisited, further emphasizing which treatments should be of higher or lower priority based on evidence (if it exists), and providing more visuals for patient education materials.

XSL•FO
RenderX

Table . Summary of qualitative usability evaluation results of the low back pain decision support tool.

Category	Selected finding	Illustrative quote
Overall impact of the CDS[a] tool on clinical care	Users believed that the tool would help users focus on the relevant information, improve diagnostic accuracy, and use of evidence-based treatments.	"I'm very likely to use this… you know why? It makes me better. It makes me smarter… This is an elegant way to approach the black hole of low back pain." (User 4)
		"It helps you focus your thoughts" (User 3)
Training required to use the CDS tool	Although the tool is simple to use technically, clinical knowledge is needed to use it effectively and will require training.	"It doesn't seem complex; it just seems like I need to familiarize myself with it." (User 1)
Input items	Enabling items adaptively can help guide information gathering.	"It is nice that it does close off those other options that maybe get in our way… I think this helps cater the treatment to the patient's diagnoses rather than allowing the provider to cater the treatment to their own preferences." (User 5)
	It might be challenging for users to go through all items during visits with a patient.	"I think it could be challenging even with familiarity… to answer it real-time [as] I'm having a conversation with somebody." (User 1)
Working diagnoses	Degree of evidence is helpful and clearly shown in bar format.	"Very visually clear." (User 4)
Treatment recommendations	The recommendations can help inform treatment for both clinician and student.	"Beyond nice, this is very educational… I love this, this is brilliant." (User 3)
Patient education	The content may be helpful to some patients but too complex for others.	"I think this is very useful especially if someone starts catastrophizing." (User 4)
		"I think it would be hugely beneficial at the report of findings to explain to the patient what the pain generator is and how that is affecting their presentation." (User 5)
		"[For] 90% of [my patient] this would be far too complex… it is far beyond a lot of our patients understanding." (User 3)

[a]CDS: clinical decision support.

Discussion

Principal Results

This study reports the initial development and evaluation of a CDS tool facilitating evidence-based diagnosis, clinical management, and patient education for LBP. The tool is web-based and can be embedded into the EHR for use during the clinical encounter as a support for interview, examination, care planning, and patient education. It can also be used independent of the patient encounter. The SUS score for the prototype was 84.75 (range: 67.5 - 95), which is in the top 10th percentile. Users believed that the CDS tool was easy to use, although training in clinical content and practice was considered necessary for effective and efficient use. With such training and practice, users believed that it would improve care and shed light on the "black hole" of LBP diagnoses and treatment.

The CDS tool inputs included 44 patient interview and examination items that provided evidence of 12 possible working diagnoses. Input items were ordered to begin with interview items followed by examination items—conditionally displayed if relevant—so as to parallel clinical workflows. The output user interface allows multiple potential working diagnoses, reflecting the multifactorial nature of LBP. Likeliness scales enable rapid visual assessment of the extent of evidence supporting each condition. This feature enables clinicians to confirm data accuracy and determine whether the interview and examination data align with the overall clinical presentation. The ability to remove or reject working diagnoses for which there is little evidence in favor of others supported by stronger evidence can assist in prioritizing care. The patient education materials written at an eighth-grade level were designed to help further explain LBP, offer reassurance, list possible treatment and self-care options, describe other factors influencing pain severity (eg, stress, poor sleep, and lack of movement), and encourage patients to communicate with clinicians about these topics. Some interviews suggested that these materials had potential to be useful, although some revision may be needed. Review by, and feedback from, patients with LBP would best inform further development of this feature.

Contribution and Comparison With Other Work

To our knowledge, this is the first provider-facing tool developed to assist clinicians in the diagnosis and treatment of LBP. A prior study describes the development a CDS tool to assist Australian pharmacists in advising patients about watchful waiting for acute LBP and recommending medications and provider types to seek for persistent symptoms [20]. Our work is complementary to that work in that it focuses on offering providers a CDS tool once patients seek care. The CDS tool offers a systematic process for a clinical evaluation which, in

XSL•FO
RenderX

turn, generates working diagnoses and links these outputs with guideline-based nonpharmacological treatments offered by chiropractors and other practitioners providing similar treatments. Our tool also shares some design similarities with currently available web-based risk-screening tools, such as the Lung Cancer Screening Risk Calculator and the Fracture Risk Assessment Tool (FRAX) [21,22]. These calculators are clinician-facing applications with input variables (eg, age, sex, smoking status, and bone mineral density) for estimating the risk for future lung cancer and osteoporotic fracture.

In addition to distilling the evidence for LBP diagnosis and treatment and making it available to clinicians at the point of care, this tool has the potential to help clinicians address patient needs identified in qualitative research for LBP. Costa et al [10] describe difficulties encountered by people with LBP due to uncertainty stemming from diagnoses and treatment. Our CDS tool has the potential to support clinicians in explaining the nature of the problem, treatment plan, and what patients can do on their own to address the problem [10]. Cox et al [23] make specific recommendations to clinicians about communicating LBP diagnoses, including sharing thought processes and diagnostic reasoning and providing written information about diagnosis. Our tool addresses these recommendations [23].

The CDS tool we developed may enable clinicians to avoid common errors in clinical reasoning leading to misdiagnosis, such as anchoring and premature closure [24]. Anchoring is an attachment to a diagnosis despite contradictory evidence; premature closure occurs when a diagnosis is accepted before verification. By offering objective, evidence-based, and repeatable provisional working diagnoses based on history and examination data, with visual displays of the strength of evidence supporting working diagnoses, this tool may help clinicians more systematically and consciously consider each potential working diagnosis in the context of the strength of evidence.

Limitations and Future Work

User input was gathered from 5 practicing chiropractic clinicians who participated in both the development and usability testing phases. This sample size is sufficient for formative research and for identifying major usability issues [25-27]. While there is a potential for bias due to their involvement in the design process, the likelihood of such bias is reduced, as the design sessions took place 7 months prior to the usability testing, limiting recall of specific design details. Informed by our findings and to address limitations in our scope, we plan to build on this work in 6 areas. First, we will implement the tool within a health system's EHR to enable easy access within the clinical workflow and enhance certain features such as allowing users to save information for each patient. Second, we will iteratively refine the patient-facing content with patients, clinicians, and with consideration of evidence in literature. Third, we will expand to other clinical specialties that treat LBP. Our initial focus was on chiropractic, a profession focused on nonpharmacological diagnosis and treatment of neuro-musculoskeletal conditions and consisting of approximately 77,000 licensed practitioners in the United States [28]. There is potential to further develop the tool for physical therapy, primary care, and other specialties. Fourth, the tool may be expanded beyond the biological perspective to include other factors (eg, psychological, spiritual, environmental, and social) relevant to LBP diagnosis and management [29]. Fifth, we will examine opportunities to incorporate artificial intelligence (AI). AI-based approaches are gaining traction in diagnosis [30,31] and other areas. However, AI models are limited by the data quality used to create them. By standardizing the definition of interview and examination inputs, our tool has the potential to improve the quality of data used by AI. Finally, the CDS tool developed in this study does not screen for signs or symptoms of pathology (eg, fracture, infection, and malignancy). There is no current consensus on such screening, and findings can be interpreted differently with varied clinical histories and among patients in different age groups. However, inputs for major red flags (eg, history of cancer and pulsatile pain with abdominal bruit suggesting aortic aneurysm) can be added to prompt and facilitate the clinical screening process.

Conclusions

The LBP tool we have developed has potential to help standardize the process of clinical evaluation leading to systematically derived working diagnoses that inform individually tailored treatment decisions. Our process of defining clinical requirements based on rigorous evidence reviews and applying user-centered design principles through eliciting user requirements produced a clinician-facing decision support tool that was viewed positively and with enthusiasm by clinical users. If implemented, this tool may shed light on the "black hole" of LBP care, facilitate dissemination of LBP evidence as it is produced, and catalyze innovation in LBP care.

Acknowledgments

Research reported in this publication was supported by the RAND Technology Incubator and the National Center for Complementary and Integrative Health of the National Institutes of Health under award number U24AT012549. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors thanks Andrew Sloan for writing the software code for the tool and Alyson Youngblood for help designing the user interface.

Data Availability

The datasets generated during and/or analyzed during this study are not publicly available to protect the privacy of the research subjects.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional examples of wireframes and screenshots.
[DOCX File, 2515 KB - formative_v9i1e66666_app1.docx]

References

1. Hartvigsen J, Hancock MJ, Kongsted A, et al. What low back pain is and why we need to pay attention. Lancet 2018 Jun 9;391(10137):2356-2367. [doi: 10.1016/S0140-6736(18)30480-X] [Medline: 29573870]
2. Pitcher MH, Von Korff M, Bushnell MC, Porter L. Prevalence and profile of high-impact chronic pain in the United States. J Pain 2019 Feb;20(2):146-160. [doi: 10.1016/j.jpain.2018.07.006] [Medline: 30096445]
3. Dutmer AL, Schiphorst Preuper HR, Soer R, et al. Personal and societal impact of low back pain: the Groningen spine cohort. Spine (Phila Pa 1976) 2019 Dec 15;44(24):E1443-E1451. [doi: 10.1097/BRS.0000000000003174] [Medline: 31369481]
4. Wu A, March L, Zheng X, et al. Global low back pain prevalence and years lived with disability from 1990 to 2017: estimates from the Global Burden of Disease Study 2017. Ann Transl Med 2020 Mar;8(6):299. [doi: 10.21037/atm.2020.02.175]
5. Lo J, Chan L, Flynn S. A systematic review of the incidence, prevalence, costs, and activity and work limitations of amputation, osteoarthritis, rheumatoid arthritis, back pain, multiple sclerosis, spinal cord injury, stroke, and traumatic brain injury in the United States: a 2019 update. Arch Phys Med Rehabil 2021 Jan;102(1):115-131. [doi: 10.1016/j.apmr.2020.04.001]
6. Dieleman JL, Cao J, Chapin A, et al. US health care spending by payer and health condition, 1996-2016. JAMA 2020 Mar 3;323(9):863-884. [doi: 10.1001/jama.2020.0734] [Medline: 32125402]
7. Maher C, Underwood M, Buchbinder R. Non-specific low back pain. Lancet 2017 Feb;389(10070):736-747. [doi: 10.1016/S0140-6736(16)30970-9]
8. Gräper PJ, Scafoglieri A, Clark JR, Hallegraeff JM. Sensory profiles predict symptoms of central sensitization in low back pain: a predictive model research study. J Clin Med 2024 Aug 1;13(16):4677. [doi: 10.3390/JCM13164677/S1]
9. Vining RD, Shannon ZK, Minkalis AL, Twist EJ. Current evidence for diagnosis of common conditions causing low back pain: systematic review and standardized terminology recommendations. J Manipulative Physiol Ther 2019 Nov;42(9):651-664. [doi: 10.1016/j.jmpt.2019.08.002]
10. Costa N, Butler P, Dillon M, et al. "I felt uncertain about my whole future"-a qualitative investigation of people's experiences of navigating uncertainty when seeking care for their low back pain. Pain 2023 Dec 1;164(12):2749-2758. [doi: 10.1097/j.pain.0000000000002975] [Medline: 37478013]
11. Vining R, Potocki E, Seidman M, Morgenthal AP. An evidence-based diagnostic classification system for low back pain. J Can Chiropr Assoc 2013 Sep;57(3):189-204. [Medline: 23997245]
12. Vining RD, Minkalis AL, Shannon ZK, Twist EJ. Development of an evidence-based practical diagnostic checklist and corresponding clinical exam for low back pain. J Manipulative Physiol Ther 2019 Nov;42(9):665-676. [doi: 10.1016/j.jmpt.2019.08.003]
13. Vining RD, Shannon ZK, Salsbury SA, Corber L, Minkalis AL, Goertz CM. Development of a clinical decision aid for chiropractic management of common conditions causing low back pain in veterans: results of a consensus process. J Manipulative Physiol Ther 2019 Nov;42(9):677-693. [doi: 10.1016/j.jmpt.2019.03.009] [Medline: 31864769]
14. ISO. ISO 9241-210:2019 —Ergonomics of human-system interaction — part 210: human-centred design for interactive systems. 2019. URL: https://www.iso.org/standard/77520.html [accessed 2024-06-25]
15. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005 Apr 2;330(7494):765. [doi: 10.1136/bmj.38398.500764.8F] [Medline: 15767266]
16. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
17. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013;13(1):117. [doi: 10.1186/1471-2288-13-117]
18. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
19. Measuring U. Measuring usability with the System Usability Scale (SUS). 2011. URL: https://measuringu.com/sus/ [accessed 2024-08-13]
20. Downie AS, Hancock M, Abdel Shaheed C, et al. An electronic clinical decision support system for the management of low back pain in community pharmacy: development and mixed methods feasibility study. JMIR Med Inform 2020 May 11;8(5):e17203. [doi: 10.2196/17203] [Medline: 32390593]

XSL•FO
RenderX

21. ScreenLC. Lung cancer screening risk calculator. URL: https://screenlc.com/dpp-vue/index.html [accessed 2024-09-03]
22. FRAX. Fracture risk assessment tool. URL: https://frax.shef.ac.uk/FRAX/tool.aspx [accessed 2024-09-03]
23. Cox C, Hatfield T, Willars J, Fritz Z. Identifying facilitators and inhibitors of shared understanding: an ethnography of diagnosis communication in acute medical settings. Health Expect 2024 Aug;27(4):e14180. [doi: 10.1111/hex.14180] [Medline: 39180375]
24. Rylander M, Guerrasio J. Heuristic errors in clinical reasoning. Clin Teach 2016 Aug;13(4):287-290. [doi: 10.1111/tct.12444] [Medline: 26395611]
25. Nielsen Norman Group. Why you only need to test with 5 users. 2000 Mar 18. URL: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ [accessed 2024-11-19]
26. Virzi RA. Refining the test phase of usability evaluation: how many subjects is enough? Hum Factors 1992 Aug;34(4):457-468. [doi: 10.1177/001872089203400407]
27. Shuldiner J, Kiran T, Agarwal P, et al. Developing an audit and feedback dashboard for family physicians: user-centered design process. JMIR Hum Factors 2023;10(1):e47718. [doi: 10.2196/47718]
28. NBCE. Practice analysis of chiropractic 2020—NBCE survey analysis. 2020. URL: https://www.nbce.org/practice-analysis-of-chiropractic-2020/ [accessed 2024-09-03]
29. King D. Diagnosis: Oxford University Press; 2015. [doi: 10.1093/ACREFORE/9780199381135.013.2138]
30. Fu Y, Zhou J, Li J. Diagnostic performance of ultrasound-based artificial intelligence for predicting key molecular markers in breast cancer: a systematic review and meta-analysis. PLoS ONE 2024;19(5):e0303669. [doi: 10.1371/journal.pone.0303669] [Medline: 38820391]
31. Nowroozi A, Salehi MA, Shobeiri P, et al. Artificial intelligence diagnostic accuracy in fracture detection from plain radiographs and comparing it with clinicians: a systematic review and meta-analysis. Clin Radiol 2024 Aug;79(8):579-588. [doi: 10.1016/j.crad.2024.04.009]

Abbreviations

AI: artificial intelligence
CDS: clinical decision support
FRAX: Fracture Risk Assessment Tool
LBP: low back pain
SUS: System Usability Scale

Edited by A Mavragani; submitted 19.09.24; peer-reviewed by S Kandaswamy; revised version received 02.12.24; accepted 03.12.24; published 04.02.25.

Please cite as:
Rudin RS, Herman PM, Vining R
Addressing the "Black Hole" of Low Back Pain Care With Clinical Decision Support: User-Centered Design and Initial Usability Study
JMIR Form Res 2025;9:e66666
URL: https://formative.jmir.org/2025/1/e66666
doi:10.2196/66666

© Robert S Rudin, Patricia M Herman, Robert Vining. Originally published in JMIR Formative Research (https://formative.jmir.org), 4.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Summer Research Internship Curriculum to Promote Self-Efficacy, Researcher Identity, and Peer-to-Peer Learning: Retrospective Cohort Study

Yulia A Levites Strekalova[1], MBA, PhD; Rachel Liu-Galvin[1], MBChB; Samuel Border[1], MS; Sara Midence[1], BS; Mishal Khan[1], MHA; Maya VanZanten[2], BS; John Tomaszewski[3], MD; Sanjay Jain[4], MD, PhD; Pinaki Sarder[1], PhD

[1]Department of Health Services Research, Management and Policy, College of Public Health and Health Professions, University of Florida, Gainesville, FL, United States

[2]Ann & Robert H Lurie Children's Hospital of Chicago, Chicago, IL, United States

[3]Department of Pathology and Anatomical Sciences, University at Buffalo, Buffalo, NY, United States

[4]John T Milliken Department of Medicine, Washington University, St. Louis, MO, United States

Corresponding Author:
Yulia A Levites Strekalova, MBA, PhD
Department of Health Services Research, Management and Policy, College of Public Health and Health Professions, University of Florida, Gainesville, FL, United States

Abstract

Background: Common barriers to students' persistence in research include experiencing feelings of exclusion and a lack of belonging, difficulties developing a robust researcher identity, perceptions of racial and social stigma directed toward them, and perceived gaps in research skills, which are particularly pronounced among trainees from groups traditionally underrepresented in research. To address these known barriers, summer research programs have been shown to increase the participation and retention of undergraduate students in research. However, previous programs have focused predominantly on technical knowledge and skills, without integrating an academic enrichment curriculum that promotes professional development by improving students' academic and research communication skills.

Objective: This retrospective pre-then-post study aimed to evaluate changes in self-reported ratings of research abilities among a cohort of undergraduate students who participated in a summer research program.

Methods: The Human BioMolecular Atlas Program (HuBMAP) piloted the implementation of a web-based academic enrichment curriculum for the Summer 2023 Research Internship cohort, which was comprised of students from groups underrepresented in biomedical artificial intelligence research. HuBMAP, a 400-member research consortium funded by the Common Fund at the National Institutes of Health, offered a 10-week summer research internship that included an academic enrichment curriculum delivered synchronously via the web to all students across multiple sites. The curriculum is intended to support intern self-efficacy, researcher identity development, and peer-to-peer learning. At the end of the internship, students were invited to participate in a web-based survey in which they were asked to rate their academic and research abilities before the internship and as a result of the internship using a modified Entering Research Learning Assessment instrument. Wilcoxon matched-pairs signed rank test was performed to assess the difference in the mean scores per respondent before and after participating in the internship.

Results: A total of 14 of the 22 undergraduate students who participated in the internship responded to the survey. The results of the retrospective pre-then-post survey indicated that there was a significant increase in students' self-rated research abilities, evidenced by a significant improvement in the mean scores of the respondents when comparing reported skills self-assessment before and after the internship (improvement: median 1.09, IQR 0.88-1.65; W=52.5, P<.001). After participating in the HuBMAP web-based academic enrichment curriculum, students' self-reported research abilities, including their confidence, their communication and collaboration skills, their self-efficacy in research, and their abilities to set research career goals, increased.

Conclusions: Summer internship programs can incorporate an academic enrichment curriculum with small-group peer learning in addition to a laboratory-based experience to facilitate increased student engagement, self-efficacy, and a sense of belonging in the research community. Future research should investigate the impact of academic enrichment curricula and peer mentoring on the long-term retention of students in biomedical research careers, particularly retention of students underrepresented in biomedical fields.

(JMIR Form Res 2025;9:e54167) doi:10.2196/54167

XSL•FO
RenderX

KEYWORDS

artificial intelligence; biomedical research; curriculum; training programs; workforce

Introduction

Several population groups continue to be underrepresented among trainees pursuing biomedical research disciplines [1-5]. According to 2023 census data, Hispanic, Black, and American Indian and Alaska Native individuals comprised 34.5% of the US population (19.5%, 13.7%, and 1.3%, respectively) [6]. Yet, only 16.0% of students awarded science and engineering doctorate degrees in 2020 were from these groups [7]. These disparities carry over into the workforce, with Hispanic individuals comprising only 15.0% of the STEM (science, technology, engineering, and math) workforce in 2021, 9.0% of Black individuals, and <1.0% of American Indians and Alaska Native individuals [7]. These figures illustrate a disparity between the racial and ethnic diversity of the US population and of individuals working in biomedical and STEM fields.

The underrepresentation of trainees from diverse backgrounds is particularly pronounced in the fields of biomedical artificial intelligence (AI) and machine learning [8,9]. Biomedical research has seen significant growth in AI applications with the development of tools for diagnosing, monitoring, and predicting prognoses [10,11]. With AI's increasing role in society, AI researchers' diversity must mirror the diverse populations they serve. A skilled research workforce in domain-specific AI applications is needed to maintain sustained progress in AI research. According to a 2019 report from the AI Now Institute, women comprised only 18% of authors at AI conferences, while 80% of AI professors were male [8]. A lack of diverse perspectives can have significant consequences on resulting research products, as lack of fairness (the ability of AI to be free from bias) and explainability (the ability to understand why an AI model makes certain predictions, which is crucial for identifying biases) have plagued the AI field for years [10,11].

Common barriers to students' participation and persistence in research include experiencing feelings of exclusion and a lack of belonging, difficulties developing a strong research identity, perceptions of racial and social stigma directed toward them, and perceived gaps in research skills [1-3,12,13]. One study, for example, observed that reasons for dropout from computer science degrees included unclear expectations about the chosen field, feeling as if they did not belong, and social isolation [12]. These challenges are particularly pronounced among trainees from groups traditionally underrepresented in research, as articulated by the National Institutes of Health (NIH) [14]. The NIH reports that several racial and ethnical groups are underrepresented in research: Black or African American individuals, Hispanic or Latinx individuals, American Indians or Alaska Natives, and Native Hawaiians and other Pacific Islanders, with women carrying a higher burden of underrepresentation [14]. To address these known barriers, research exposure programs for undergraduate students can include a curriculum of academic skill development components that promote student self-efficacy, identity, and a sense of belonging in research. Specifically, evidence shows that students encouraged to develop their "science identity" are more likely to integrate into the wider culture of scientific research [3]. Students can develop their self-confidence, self-efficacy, and identity as scientists through science support experiences and connections with peers and mentors, thus benefiting from exposure to diverse perspectives and experiences [5,15,16]. Additionally, experiences that foster a sense of belonging to the research community have been shown to lead to greater persistence in STEM fields, especially for students from groups underrepresented in research [16-19]. Furthermore, STEM undergraduates who engage in small-group and active learning demonstrate increased self-efficacy, achieve better grades, and are more likely to remain in STEM fields [16,20,21].

There is a critical need for nationwide collaborative training programs to fund, develop, and mentor students from underrepresented groups to ensure AI representation in biomedical research [22]. Summer research programs delivered via both in-person and web-based formats have been shown to increase the participation and retention of undergraduate students in research [23]. Evidence shows that involvement in both face-to-face and remote summer research experiences equips undergraduate students with knowledge and research skills and promotes enrollment in postgraduate training and entry into scientific- or health care–related careers [24,25]. In addition to being exposed to research and learning research skills, students can also gain confidence as scientists and develop a researcher identity and intention to pursue a career in STEM [19,26,27]. However, previous programs have primarily focused on equipping students with technical knowledge and skills, with varying degrees of emphasis on professional development, without integrating a universal academic enrichment curriculum that is delivered synchronously via the web to students across multiple sites and that focuses on active learning and improving students' academic and research skills [23-27]. The effective development of students' research skills and familiarity with scientific methods calls for supplementing laboratory research experience with academic enrichment sessions in order to equip students with the necessary academic and professional skills to thrive as researchers building their identities in the research space. These sessions should allow students to actively engage in discussions about their identities as researchers while offering a supportive environment that fosters a sense of belonging and facilitates peer-to-peer information exchange. This study aimed to report the findings of delivering academic enrichment sessions to undergraduate students from groups underrepresented in AI research and to test the hypothesis that a guided academic enrichment curriculum, as part of a summer research internship that also included a laboratory-based experience, promotes self-perceived research skills, researcher identity, and professional skills.

Methods

Program Background

The Human BioMolecular Atlas Program (HuBMAP) is a consortium of diverse research teams with expertise in biology,

XSL•FO
RenderX

imaging, computation, data management, and visualization. The program is funded by the NIH Common Fund and aims to create a multiscale spatial atlas of the human body at single-cell resolution by applying advanced technologies and disseminating resources to the community [28]. Extensive infrastructure has been established through a coordinated effort among various HuBMAP integration, visualization, and engagement teams; tissue-mapping centers; technology and tools development groups; and rapid technology implementation teams and working groups [29]. To ensure the pipeline of emerging research talent familiar with HuBMAP's tools, data, and processes, the program began offering summer research internships for undergraduate students in 2021. The internship included placement of the student within an active research laboratory for 8 - 10 weeks, during which time they were mentored by an established investigator, and at the end of which they were encouraged to deliver a presentation of their research project. The 2023 HuBMAP internship program hosted students at 19 different research laboratories across 12 different institutions. Participants were recruited nationally by the NIH coordinating center; they submitted applications that included a research interest and career goal statement and were selected for a paid internship based on matching their research interests to those of participating mentors. A complete list of participating laboratories and institutions is included in Multimedia Appendix 1.

In 2023, the HuBMAP internship program was enhanced with an academic enrichment curriculum. The curriculum was informed by the evidence-based Entering Research curriculum, developed by a faculty trained as an Entering Research facilitator [30], and designed to address the needs and challenges of a diverse group of undergraduate students from groups underrepresented in biomedical and AI research. The academic enrichment curriculum included a series of interactive workshops, peer discussions, and hands-on writing assignments aimed at supporting students to develop their research identities and enhance their research communication skills. Each session lasted 2 hours and was delivered synchronously via Zoom, enabling students participating in internships at 19 different research laboratories across 12 different institutions across the United States to attend. A complete list of the sessions constituting the academic enrichment curriculum can be found in Multimedia Appendix 2. Sessions were structured into 1 hour of didactics, small-group activities, and discussions and 1 hour of guided work on academic products. Examples of academic products included a research statement, an individual development plan, a web-based ORCID profile, and a research poster, among others. An institutional Google Drive account was used to capture interns' work systematically and securely, promote peer-to-peer learning, and facilitate continuous access to training materials and assignments.

Data Collection and Instruments

In the summer of 2023, a total of 22 students participated in the HuBMAP summer internship, and their self-reported demographic data were collected through the initial application form. Demographic variables collected during the application process included sex, race and ethnicity, class standing, degree major, and home institution type (categorized as per Carnegie classification).

At the end of the internship, all 22 students were invited to participate in a web-based survey in which they were asked to rate their research skills self-efficacy (operationalized on the survey instrument as academic and research abilities) before the internship (referred to as the before questions) and as a result of the internship (referred to as the after questions). It is important to note that the after questions specifically asked the students to rate their abilities as a result of the internship, and not as a result of the academic enrichment curriculum. The survey employed a 17-item scale and validated questions adapted from the Entering Research Learning Assessment instrument [31]. The Entering Research Learning Assessment instrument is designed to measure learning gains of undergraduate or graduate research trainees across 7 areas: research comprehension and communication skills, practical research skills, research ethics, researcher identity, equity and inclusion awareness and skills, and professional and career development skills. The survey was piloted with undergraduate students and presented at several meetings to internship leaders at the NIH and participating sites to gain expert feedback and ensure the content and external validity of the items. Examples of questions in the survey included those that asked about students' self-reported ability to communicate the context, methods, and results of their research, to tailor research communications to different audiences, to work comfortably in the research environment, to explore possible research career pathways, to set research career goals, and to develop a plan to pursue a research career. For each item, students were asked to indicate their abilities on a scale of 1 - 5, with 1 representing "no ability," 2 representing "a little ability," 3 representing "moderate ability," 4 representing "good ability," and 5 representing "great ability." The complete list of survey questions can be found in (Multimedia Appendix 3).

The retrospective pre-then-post survey was administered anonymously, and the participants were asked to rereport their sex, race and ethnicity, and class standing. Since the retrospective pre-post survey was anonymous, we were unable to tell which students had or had not responded to the survey; thus, we compared the demographic characteristics of the respondents (n=14) to that of the whole cohort (N=22).

Statistical Analyses

The Cronbach α was applied to measure the internal consistency of the items included to assess the participants' research skills and self-efficacy. Median, minimum, and maximum scores for each of the before and after questions were calculated. Mean scores per respondent for the before questions compared with the after questions were also calculated, and a Wilcoxon matched-pairs signed rank test was performed to assess the median difference in the mean scores per respondent for the before questions compared with the after questions. Although the survey results were anonymous, we were able to match responses to each before and after question to individual survey respondents based on anonymous identification numbers.

XSL•FO
RenderX

All statistical analyses were performed using SAS (version 9.4; SAS Institute Inc). All *P* values were 2-sided, with $P<.05$ considered statistically significant.

Ethical Considerations

The study was submitted for review to the University of Florida Institutional Review Board and determined to be exempt (#ET00041828). Informed consent was obtained from all survey respondents. No compensation was provided for participants in this study, and no identification of individual participants in any manuscript or supplementary material is possible.

Results

Participant Characteristics

Participants in the 2023 HuBMAP summer internship and curriculum included 22 students from undergraduate institutions across the United States. Students had declared majors in STEM fields such as biology, bioinformatics, biochemistry, and engineering. While class standing varied among the participants, half of the students were at a sophomore level before beginning their summer internship. Related to their incoming career goals, participants often reported interest in pursuing MD, PhD, or MD/PhD degrees and sought to use the internship to explore the various career opportunities available in biomedicine. Almost half of the participants (9/22, 40.9%) were from high (R2) and very high (R1) research activity academic institutions, with the rest of the participants (13/22, 59.1%) from master, baccalaureate, and associate colleges. Demographic information of the 22 students is presented in Table 1.

XSL•FO
RenderX

Table . Characteristics of Human BioMolecular Atlas Program summer 2023 interns (N=22) and survey respondents (N=14).

Demographic	Interns, n (%)	Survey respondents, n (%)
Sex		
Male	7 (31.8)	3 (21.4)
Female	14 (63.6)	10 (71.4)
Other	1 (4.5)	1 (7.1)
Race		
Black or African American	6 (27.3)	5 (35.7)
Asian or Pacific Islander	5 (22.7)	3 (21.4)
White	4 (18.2)	1 (7.1)
Other race	0 (0)	1 (7.1)
More than one race	2 (9.1)	1 (7.1)
N/A[a] or prefer not to answer	5 (22.7)	3 (21.4)
Ethnicity		
Hispanic or Latino	7 (31.8)	3 (21.4)
Not Hispanic or Latino	15 (68.2)	11 (78.6)
Class standing		
Freshman	5 (22.7)	2 (14.3)
Sophomore	11 (50.0)	7 (50.0)
Junior	3 (13.6)	2 (14.3)
Senior	3 (13.6)	3 (21.4)
Major		
Biology	8 (36.4)	—[b]
Biochemistry	5 (22.7)	—
Bioinformatics	2 (9.1)	—
Chemistry	1 (4.5)	—
Biomedical/electrical engineering	2 (9.1)	—
Genetics	1 (4.5)	—
Information technology	1 (4.5)	—
Nanomedicine	1 (4.5)	—
Neuroscience	1 (4.5)	—
Home institution type per Carnegie classification		
Doctoral universities: very high research activity (R1)	5 (22.7)	—
Doctoral universities: high research activity (R2)	4 (18.2)	—
Master colleges and universities: larger programs (M1)	4 (18.2)	—
Master colleges and universities: smaller programs (M3)	1 (4.5)	—
Baccalaureate colleges	7 (31.8)	—
Associate colleges	1 (4.5)	—

[a]N/A: not applicable.

[b]Not available.

XSL•FO
RenderX

Invitations to complete a web-based survey were sent out to all 22 students, and 14 survey responses were received and included in data analysis. The makeup of the survey respondents aligned with the overall sample, but the sample size was too small for statistical analyses. The demographic makeup of all interns who participated in the program and the subset who responded to the survey are reported in Table 1. The postprogram survey did not collect data on degree major or institutional affiliation; thus, this information is not reported for the Survey Respondents column in Table 1.

Results of Retrospective Pre-Post Survey

Cronbach α was calculated to measure the level of internal consistency among the survey questions, and its value of 0.96 for the preintervention questions and 0.92 for the postintervention questions indicated an excellent level of internal consistency, that is, all questions in the survey were assessing the primary underlying concepts being assessed: researcher skills, researcher identity, and self-efficacy in research.

Table 2 reports the median, minimum, and maximum values of the responses to each of the before questions compared with each of the after questions. Given the ordinal nature of the data, where the median was a decimal number due to the averaging of 2 middle values, the median was rounded up to the nearest whole number. Of all survey responses, 3 responses were missing (one to each of After Question 3, After Question 5, and After Question 9); missing responses therefore comprised <1% of all data.

XSL•FO
RenderX

Table . Median and range of responses to self-reported research skills assessment before versus after questions.

Skill assessed	Before, median (IQR; min, max)	After, median (IQR; min, max)
Research comprehension and communication skills		
Communicate the context, methods, and results of your research.	3 (2-4.25; 1, 5)	4 (4-5; 3, 5)
Tailor your research communications for different audiences (eg, general public and disciplinary conference)	3 (3-4; 1, 5)	5 (4-5; 4, 5)
Work in the research environment comfortably	3 (2-4.25; 1, 5)	5 (4-5; 4, 5)
Ask questions to clarify your understanding of your research project	4 (3-4; 1, 5)	5 (4-5; 4, 5)
Practice regular and open communication with your research mentor	3 (2.75-4; 1, 5)	5 (4-5; 4, 5)
Practice regular and open communication with your research team members	3 (2.75-4; 1, 5)	5 (4-5; 4, 5)
Practical research skills		
Use the tools, materials, and equipment needed to conduct research	3 (2-3.25; 1, 5)	5 (4-5; 4, 5)
Make a case for your research question based on the literature	3 (2-3; 1, 5)	4 (4-5; 3, 5)
Research ethics		
Identify forms of unethical practices or research misconduct	3 (2.75-4; 1, 5)	4 (4-5; 3, 5)
Researcher identity		
Think of yourself as a scientist/researcher	3 (2-3; 1, 5)	4 (3.75-5; 2, 5)
Feel like you belong in research	3 (2-3.25; 1, 5)	4 (3-5; 2, 5)
Researcher confidence and independence		
Work independently on your research project	2 (1.75-3.25; 1, 4)	5 (4-5; 2, 5)
Determine the next steps in your research project	2 (1.75-3; 1, 5)	4 (3.75-5; 3, 5)
Equity and inclusion awareness and skills		
Understand how others might experience research differently based on their identity (eg, race, socioeconomic status, first-generation status)	3 (1-4; 1, 5)	5 (4-5; 1, 5)
Professional and career development skills		
Explore possible research career pathways	3 (2-4; 1, 5)	5 (3.75-5; 3, 5)
Set research career goals	3 (2-3; 1, 5)	5 (4-5; 3, 5)
Develop a plan to pursue a research career (determine the next step in your training)	3 (2-3.25; 1, 4)	5 (4-5; 3, 5)

The results from the Wilcoxon signed rank matched-pairs test indicated that there was a significant improvement in the mean scores of the respondents when comparing the before questions to the after questions (improvement: median 1.09, IQR 0.88-1.65; W=52.5, P<.001). The mean scores of each respondent for the before questions compared with the after questions are presented in Multimedia Appendix 4. A review of individual items revealed an increase in all items, indicating that overall, students felt their research abilities had improved following the HuBMAP Summer Research Internship.

Discussion

This paper reports on the pilot implementation of an academic enrichment curriculum delivered as part of the summer, AI-focused, undergraduate research internship administered through the HuBMAP program. Following participation in the internship, students reported increases in self-assessed research skills, including their abilities to set research career goals, to think of themselves as researchers, and to work comfortably in the research environment, and increases in their feelings of belonging in research. Our curricular intervention included

XSL•FO
RenderX

academic enrichment content supplemented by peer learning and hands-on activities.

Earlier research has demonstrated that perceived gaps in research skills and feelings of isolation have detrimental effects on students' persistence in computer science and AI-focused training [12]. While most summer research internships focus on individual training and technical aspects of research with varying degrees of emphasis on professional development [23-27], the format and content of the supplemental academic enrichment curriculum we piloted showed promise in creating a nurturing, active, and collaborative learning environment for student interns across multiple institutions, with demonstrated self-reported improvements in abilities that went beyond technical research skills.

Based on our study's findings, we have formulated propositions and hypotheses about the mechanisms through which academic enrichment curricula promote research skills self-efficacy and researcher identity and result in sustained engagement in research training. First, delivering an academic enrichment curriculum enhances and amplifies the learning experiences of students participating in summer research programs and promotes the long-term retention of these students in biomedical research careers. Second, academic enrichment sessions help to build a research community and promote a sense of belonging among students. Third, providing students with opportunities to learn from one another encourages peer learning and enhances students' critical thinking skills, resourcefulness, and resilience within academia. Additional research questions that we would like to answer through future studies and longitudinal assessments include how students can learn from one another and whether peer learning is superior to learning from mentors. Future studies should also explore the effect of individual versus cohort-based research experiences. The academic enrichment curriculum we delivered to participants of the HuBMAP summer internship program would be considered a mix of cohort (delivered through the academic enrichment sessions) and hybrid (since students were based at separate laboratories and institutions with individual mentors) research experiences. In addition, further research is needed to determine the optimal structure of the academic enrichment curriculum, including the frequency of sessions (once weekly vs every other week, for example).

Our experience suggests that the integration of an academic enrichment curriculum into a laboratory-based AI-focused summer research internship can help students increase their confidence, communication skills, self-efficacy, and sense of belonging in the research community. In line with our findings, a growing body of evidence shows that a structured curriculum supports undergraduate students from traditionally underrepresented groups in developing their identity as scientists, and in their overall self-efficacy to use scientific thinking, through science-support experiences and connections formed with peers [5,15,16]. Thus, we recommend that future programs should incorporate small group peer learning, active learning strategies, and laboratory-based research experience as evidence-supported methods to facilitate increased self-efficacy and a sense of belonging in the research community, which may also have the potential to promote the retention of undergraduate students in STEM and biomedical disciplines [16,20,21].

Despite these encouraging results, it is important to recognize that the generalizability of the results of this study is limited by its relatively small sample size of 22 students, of which 14 students (63.6%) completed the pre- and postprogram surveys. As a result, this study's findings may not fully represent the broader experiences of other students across other programs. It is also important to note that it was not possible to separate the effect of our academic enrichment curriculum from other activities that students may have been engaged in during their HuBMAP Summer Research Internship. Multiple individual laboratories across different institutions hosted students participating in the 2023 HuBMAP summer internship program. Therefore, while we delivered this academic enrichment curriculum to all students via the web, some students may have been provided additional learning opportunities (remote or in-person) at the laboratories or institutions where they were based. The nature and quantity of such other learning opportunities may have varied across the different sites and should be explored in future studies. It is possible that a portion of the gains in students' self-reported research abilities were attributable to these supplementary activities and not because of the academic enrichment curriculum. Future research should therefore investigate the specific impact of academic enrichment curricula on the long-term retention of students in biomedical research careers, particularly retention of students underrepresented in biomedical fields. Furthermore, each student received one-to-one and group mentoring from lead investigators and senior trainees of the host labs. The effects of mentoring and the role of peer-to-peer interaction as contributing factors in student academic skill development can also be quantified and measured in future evaluations of summer research experiences. Additionally, this study used a pre-then-post evaluation design, which combined the questions in which students were asked to rate their research abilities "before" and "as a result of" the HuBMAP Summer Research Internship into a single survey administered post program. This methodological design allowed the students to assess knowledge gained from a perspective informed by a 10-week experience. However, the effects of recall and social desirability biases can be evaluated with sufficiently powered studies that use different evaluation formats.

Following their participation in the HuBMAP 2023 summer internship program, which included an academic enrichment curriculum delivered remotely, students' self-reported research abilities, including their confidence and communication skills, their self-efficacy in research, and their abilities to set research career goals, increased. Students also reported an increased ability to think of themselves as researchers and to work comfortably in the research environment, as well as increased feelings of belonging in research after participating in the program. We recommend that summer internship programs incorporate an academic enrichment curriculum with small-group peer learning in addition to a laboratory-based experience to facilitate increased student engagement, self-efficacy, and a sense of belonging in the research community. Strengthening the pipeline of skilled and diverse

XSL•FO
RenderX

professionals entering these fields will help in addressing both current and future workforce needs, reducing bias in AI fields, and cultivating the next generation of scientists and leaders in public health, with the goal of improving health outcomes for all populations.

Acknowledgments

This work was supported in part by the National Institutes of Health Common Fund (award number 3OT2OD033753).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

YALS contributed to the writing of the original draft, conceptualization, data curation, funding acquisition, investigation, methodology, resources, supervision, and validation. RLG was responsible for writing the original draft and conducting formal analysis. MV contributed to the writing of the original draft, data curation, and resources. SB reviewed and edited the manuscript and provided resources. SM managed project administration and provided resources. MK contributed to project administration, provided resources, and reviewed and edited the manuscript. JT reviewed and edited the manuscript and provided resources. SJ reviewed and edited the manuscript and contributed to funding acquisition. PS contributed to the review and editing of the manuscript, funding acquisition, resources, and supervision.

Conflicts of Interest

PS is an advisor of digipath.ai. The authors have no further interests to declare.

Multimedia Appendix 1
List of research laboratories/institutions that hosted students for the 2023 Human BioMolecular Atlas Program summer internship program.
[DOCX File, 15 KB - formative_v9i1e54167_app1.docx]

Multimedia Appendix 2
Schedule and objectives of academic enrichment sessions offered to summer 2023 Human BioMolecular Atlas Program interns.
[DOCX File, 17 KB - formative_v9i1e54167_app2.docx]

Multimedia Appendix 3
Academic enrichment evaluation instrument.
[DOCX File, 17 KB - formative_v9i1e54167_app3.docx]

Multimedia Appendix 4
Mean scores per respondent for before questions compared to after questions.
[DOCX File, 15 KB - formative_v9i1e54167_app4.docx]

References

1. Jantzer J, Kirkman T, Furniss KL. Understanding differences in underrepresented minorities and first-generation student perceptions in the introductory biology classroom. J Microbiol Biol Educ 2021 Dec;22(3):e00176-21. [doi: 10.1128/jmbe.00176-21] [Medline: 34804325]
2. Fisher AJ, Mendoza-Denton R, Patt C, et al. Structure and belonging: pathways to success for underrepresented minority and women PhD students in STEM fields. PLoS One 2019;14(1):e0209279. [doi: 10.1371/journal.pone.0209279] [Medline: 30625173]
3. Hurtado S, Cabrera NL, Lin MH, Arellano L, Espinosa LL. Diversifying science: underrepresented student experiences in structured research programs. Res High Educ 2009 Mar;50(2):189-214. [doi: 10.1007/s11162-008-9114-7] [Medline: 23503690]
4. Estrada M, Burnett M, Campbell AG, et al. Improving underrepresented minority student persistence in STEM. CBE Life Sci Educ 2016;15(3):es5. [doi: 10.1187/cbe.16-01-0038] [Medline: 27543633]
5. McGee R Jr, Saran S, Krulwich TA. Diversity in the biomedical research workforce: developing talent. Mt Sinai J Med 2012;79(3):397-411. [doi: 10.1002/msj.21310] [Medline: 22678863]
6. Population estimates, July 1, 2023, (V2023) quickfacts US. US Census Bureau. 2024. URL: https://www.census.gov/quickfacts/fact/table/US [accessed 2024-09-17]

XSL•FO
RenderX

7. Diversity and STEM: women, minorities, and persons with disabilities. National Science Foundation. 2023. URL: https://ncses.nsf.gov/pubs/nsf23315/report/ [accessed 2023-12-18]
8. West SM. Discriminating systems: gender, race, and power in AI - report. AI Now Institute. 2019. URL: https://ainowinstitute.org/publication/discriminating-systems-gender-race-and-power-in-ai-2 [accessed 2024-09-17]
9. Zhang D, Maslej N, Brynjolfsson E, et al. The AI index 2022 annual report. arXiv. Preprint posted online on 2022. [doi: 10.48550/arXiv.2205.03468]
10. Border SP, Sarder P. From what to why, the growing need for a focus shift toward explainability of AI in digital pathology. Front Physiol 2021;12:821217 [FREE Full text] [doi: 10.3389/fphys.2021.821217] [Medline: 35087427]
11. Wong C. AI "fairness" research held back by lack of diversity. Nature New Biol 2023 Mar 30. [doi: 10.1038/d41586-023-00935-z] [Medline: 36997714]
12. Schefer-Wenzl S, Miladinovic I, Bachinger-Raithofer S, Muckenhumer C. A study on reasons for student dropouts in a computer science bachelor's degree program. In: Auer ME, Cukierman UR, Vendrell Vidal E, Tovar Caro E, editors. Towards a Hybrid, Flexible and Socially Engaged Higher Education: Proceedings of the 26th International Conference on Interactive Collaborative Learning (ICL2023), Volume 4: Springer Nature; 2024:391-400. [doi: 10.1007/978-3-031-53382-2_38]
13. Estrada M, Woodcock A, Hernandez PR, Schultz PW. Toward a model of social influence that explains minority student integration into the scientific community. J Educ Psychol 2011;103(1):206-222. [doi: 10.1037/a0020743]
14. NOT-OD-20-031: notice of NIH's interest in diversity NIH.gov. National Institutes of Health. 2019. URL: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-20-031.html [accessed 2024-09-17]
15. Chemers MM, Zurbriggen EL, Syed M, Goza BK, Bearman S. The role of efficacy and identity in science career commitment among underrepresented minority students: efficacy and identity in science career commitment. J Soc Issues 2011 Sep;67(3):469-491. [doi: 10.1111/j.1540-4560.2011.01710.x]
16. Connors PK, Lanier HC, Erb LP, et al. Connected while distant: networking CUREs across classrooms to create community and empower students. Integr Comp Biol 2021 Oct 4;61(3):934-943. [doi: 10.1093/icb/icab146] [Medline: 34190987]
17. Ostrove JM, Long SM. Social class and belonging: implications for college adjustment. Rev High Educ 2007 Jun;30(4):363-389. [doi: 10.1353/rhe.2007.0028]
18. Rosenthal L, London B, Levy SR, Lobel M. The roles of perceived identity compatibility and social support for women in a single-sex STEM program at a co-educational university. Sex Roles 2011 Nov;65(9-10):725-736. [doi: 10.1007/s11199-011-9945-0]
19. Bruthers CB, Matyas ML. Undergraduates from underrepresented groups gain research skills and career aspirations through summer research fellowship. Adv Physiol Educ 2020 Dec 1;44(4):525-539. [doi: 10.1152/advan.00014.2020]
20. Springer L, Stanne ME, Donovan SS. Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: a meta-analysis. Rev Educ Res 1999 Mar;69(1):21-51. [doi: 10.3102/00346543069001021]
21. Freeman S, Eddy SL, McDonough M, et al. Active learning increases student performance in science, engineering, and mathematics. Proc Natl Acad Sci U S A 2014 Jun 10;111(23):8410-8415. [doi: 10.1073/pnas.1319030111] [Medline: 24821756]
22. Stathoulopoulos K, Mateos-Garcia JC. Gender diversity in AI research. SSRN J 2019. [doi: 10.2139/ssrn.3428240]
23. Ajayi TB, Mueller AL, Okwuosa IS, et al. Innovations in undergraduate research training through multisite collaborative programming: American Heart Association summer undergraduate research experience syndicate. J Am Heart Assoc 2022 Apr 19;11(8):e022380. [doi: 10.1161/JAHA.121.022380] [Medline: 35388707]
24. Wierzchowski A, Wink DJ, Zhang H, Kambanis K, Robles JOR, Rosenhouse-Dantsker A. CoLab: a workshop-based undergraduate research experience for entering college students. J Chem Educ 2022 Dec 13;99(12):4085-4093. [doi: 10.1021/acs.jchemed.1c01290] [Medline: 37519308]
25. Heggland SJ, Hovde CJ, Minnich SA, Liou LE, Daniels RL. Building a biomedical pipeline: the impact of the Idaho IDeA INBRE summer research experience at a primarily undergraduate institution. Adv Physiol Educ 2021 Dec 1;45(4):786-796. [doi: 10.1152/advan.00232.2020] [Medline: 34529538]
26. Johnson MDL, Knox CJ. National Summer Undergraduate Research Project (NSURP): a virtual research experience to deliver REAL (Retention, Equity, Access, and Life-Changing) outcomes for underrepresented minorities in STEM. J Microbiol Biol Educ 2022 Apr;23(1):e00335-21. [doi: 10.1128/jmbe.00335-21] [Medline: 35496677]
27. Hudson JK, Hudson BG. Aspirnaut: a rural high school pipeline to increase diversity in STEM. Nat Rev Nephrol 2019 Oct;15(10):591-592. [doi: 10.1038/s41581-019-0190-z]
28. The Human BioMolecular Atlas Program (HuBMAP). National Institutes of Health. 2024. URL: https://commonfund.nih.gov/HuBMAP [accessed 2024-09-17]
29. Jain S, Pei L, Spraggins JM, et al. Advances and prospects for the Human BioMolecular Atlas Program (HuBMAP). Nat Cell Biol 2023 Aug;25(8):1089-1100. [doi: 10.1038/s41556-023-01194-w] [Medline: 37468756]
30. Balster N, Pfund C, Rediske R, Branchaw J. Entering research: a course that creates community and structure for beginning undergraduate researchers in the STEM disciplines. CBE Life Sci Educ 2010;9(2):108-118. [doi: 10.1187/cbe.09-10-0073] [Medline: 20516356]

XSL•FO
RenderX

31. Butz AR, Branchaw JL. Entering Research Learning Assessment (ERLA): validity evidence for an instrument to measure undergraduate and graduate research trainee development. CBE Life Sci Educ 2020 Jun;19(2):ar18. [doi: 10.1187/cbe.19-07-0146] [Medline: 32412837]

Abbreviations

AI: artificial intelligence
HuBMAP: Human BioMolecular Atlas Program
NIH: National Institutes of Health
STEM: science, technology, engineering, and math

Edited by A Mavragani; submitted 14.03.24; peer-reviewed by CM Ambrosino, G White; revised version received 14.10.24; accepted 15.10.24; published 03.02.25.

Please cite as:
Levites Strekalova YA, Liu-Galvin R, Border S, Midence S, Khan M, VanZanten M, Tomaszewski J, Jain S, Sarder P
Summer Research Internship Curriculum to Promote Self-Efficacy, Researcher Identity, and Peer-to-Peer Learning: Retrospective Cohort Study
JMIR Form Res 2025;9:e54167
URL: https://formative.jmir.org/2025/1/e54167
doi:10.2196/54167

© Yulia A Levites Strekalova, Rachel Liu-Galvin, Samuel Border, Sara Midence, Mishal Khan, Maya VanZanten, John Tomaszewski, Sanjay Jain, Pinaki Sarder. Originally published in JMIR Formative Research (https://formative.jmir.org), 3.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Web-Based Exercise and Nutrition Intervention to Improve Leg Muscle Strength and Physical Functioning in Older Adults: Pre-Post Pilot Study

Berber Gijsbertha Dorhout[1,2], PhD; Nick Wezenbeek[3], MSc; Lisette C P G M de Groot[1], Prof Dr; Pol Grootswagers[1], PhD

[1]Division of Human Nutrition and Health, Wageningen University & Research, Stippeneng 4, PO Box 17, Wageningen, Netherlands

[2]Research Group Innovation of Human Movement Care, HU University of Applied Sciences Utrecht, Utrecht, Netherlands

[3]CleverMove, Wageningen, Netherlands

Corresponding Author:
Berber Gijsbertha Dorhout, PhD
Division of Human Nutrition and Health, Wageningen University & Research, Stippeneng 4, PO Box 17, Wageningen, Netherlands

Abstract

Background: The lifestyle intervention ProMuscle, which combines resistance exercise and an increased protein intake, was effective in improving muscle strength, muscle mass, and physical functioning in older adults. However, due to a growing shortage of health care professionals, the rapidly growing aging population cannot be personally guided in the future. Therefore, Uni2Move, a scalable web-based variant of ProMuscle, was designed to reach larger groups of older adults without putting additional burden on health care professionals.

Objective: The current study investigated the effects of a web-based lifestyle intervention on muscle strength, protein intake, and physical functioning in healthy older adults. In addition, we conducted a qualitative study to gather key insights of the participants involved, as little is known about older adults' perceptions of web-based lifestyle interventions.

Methods: A pre-post pilot study was conducted in the Netherlands. In the 24-week intervention, 19 healthy adults aged >55 years were included. They performed resistance training at home twice a week for 24 weeks via web-based workout videos. Videos (45 - 60 minutes) were recorded by the fitness trainer and mainly focused on training leg muscles. In addition, older adults were advised on increasing protein intake via two web-based consultations by a dietitian in the first 12 weeks and via an e-learning course in the second 12 weeks. Intervention adherence and acceptance was measured in week 25. The 1 repetition maximum knee extension strength, repeated chair rise test, and protein intake were measured at baseline, week 13, and week 25. Linear mixed models were used to test differences over time. Semistructured interviews were used to gather experiences of participants. Atlas.ti version 22 was used to analyze the interviews.

Results: The mean age of participants (n=19) at baseline was 69 (SD 7) years. The 1 repetition maximum knee extension strength and repeated chair rise test improved significantly during the 24-week intervention with a mean difference of 7.0 kg (95% CI 4.8-9.3; *P*<.001) and −1.2 seconds (95% CI −1.7 to −0.6; *P*<.001), respectively. Total protein intake per day did not change, whereas protein intake during breakfast had increased significantly after 13 weeks with a mean difference of 6.9 g (95% CI 1.1-12.7; *P*=.01). Qualitative research revealed that advantages of the program included no need to travel and exercising in their own environment. Disadvantages were the lack of physical interaction and no corrections by the trainer.

Conclusions: The results of the web-based exercise and nutrition intervention Uni2Move indicate potential improvements of muscle strength and physical functioning in healthy middle-aged and older adults. Providing such lifestyle interventions on the internet could reach an increased number of older adults, providing the opportunity to contribute to the health and independence of the rapidly growing aging population.

(JMIR Form Res 2025;9:e54392) doi:10.2196/54392

KEYWORDS

web-based lifestyle intervention; resistance exercise; protein intake; muscle strength; muscle mass; older adults; web-based exercise; nutrition; exercise; resistance training; sarcopenia

XSL•FO
RenderX

Introduction

The aging population is growing and life expectancy is increasing [1]. With advancing age, many people are experiencing sarcopenia. Sarcopenia can be defined as the age-related decline in skeletal muscle mass and function, highlighting the importance of advancing age as a factor of sarcopenia [2]. Sarcopenia is a natural part of the aging process, but its progression can be hastened by several factors, such as inflammation associated with aging (inflammaging), chronic diseases, lack of physical activity, inadequate nutrition, unintentional weight loss, and disuse events [3]. As the global population of older adults increases, the prevalence of sarcopenia is anticipated to grow accordingly [2,4].

As a consequence, the demand for care and support is increasing, whereas at the same time, the shortage of health care personnel is growing [5]. Governments worldwide have called for novel public health approaches concerning healthy aging, in response to the demographic shift and limited availability of health care professionals in the future [3,6]. Therefore, it is important to focus on the prevention of sarcopenia, with the goal of improving the health of the aging population and adding life to years [7].

Fortunately, lifestyle and behavioral interventions may reduce the prevalence of sarcopenia. The recommended lifestyle advice for treatment of sarcopenia consists of an increased dietary protein intake [8,9] and progressive resistance exercise [10,11]. The ProMuscle intervention, which combines resistance exercise with an increased protein intake, has been shown to be effective in improving muscle health of older adults in various settings [12-14]. In view of the rapidly growing aging population and the potential of digitalization in today's world, a web-based version of ProMuscle seems promising. For that purpose, ProMuscle was transformed to a web-based program: Uni2Move.

Several studies showed that web-based exercise training (ie, participants train at home) resulted in an improved physical performance, physical activity, balance, mobility, and muscle strength in older adults [15,16]. In addition, a recent study showed that a blended home-based exercise and dietary protein intervention resulted in improved gait speed, physical activity level, muscle mass, strength, and dietary protein intake in older adults [17]. However, it is unknown whether an intervention combining resistance exercise and an increased protein intake that is offered completely on the internet is effective in improving muscle health. In addition, little is known about the perceptions and preferences of middle-aged and older adults regarding web-based interventions. Therefore, we investigated the effects of the web-based lifestyle intervention Uni2Move on muscle strength, physical functioning, and protein intake in healthy middle-aged and older adults and conducted qualitative research to gather the key insights and experiences from participants involved.

Methods

Uni2Move Program

The Uni2Move program was a 24-week web-based lifestyle intervention, based on the ProMuscle program.

Web-based resistance exercise training

Participants followed full body progressive resistance exercise training at home twice a week via web-based videos. Workouts lasted 45 to 60 minutes. The main focus was training the leg muscles, as these muscles are important in performing daily activities. At least 50% of each training session consisted of leg exercises, while the main upper body muscle groups were alternately trained once per week (shoulders, chest, back, and abdominal muscles).

The resistance training sessions comprised of a warm-up (5 - 7 minutes), strength training (40 - 50 minutes) and a cooldown (3 - 5 minutes). The warm-up consisted of aerobic exercises, mobility exercises, and light strength exercises like air squats and push-ups against the wall. During the cooldown, light aerobic and dynamic stretching exercises were performed. The strength training session consisted of 4 - 5 blocks. Each block consisted of 2 - 4 unique exercises that were performed for 3 - 4 sets of 40 - 45 seconds. Each set of exercises was followed by a 20 second break. In total, 10 - 12 unique exercises were performed per training session. An example of a block is as follows: set 1 squats, 20 second break, set 2 push-ups, 20 second break, set 3 squats, 20 second break, set 4 push-ups, etc. Effective muscle resting time was 80 seconds. This set-up was chosen as we expected long resting breaks to be detrimental for motivation and adherence in a home-based situation with only digital interaction possibilities.

Participants used their body weight, resistance bands, and weights (eg, dumbbells or bottles filled with sand). Exercises performed with bodyweight or weights were multiple variations of squats, box squats, lunges, wall-sits, bent-over rows, hip bridges, calf raises, planks, shoulder presses, lateral shoulder raises, reverse flies, push-ups, pullovers, crunches, and dips. Exercises performed with resistance bands were leg curls, leg extensions, leg abductions, glute kickbacks, knee lifts, horizontal rows, and shoulder external rotations. The exercise regimen was designed in a progressive manner by increasing the difficulty of the exercises during the 24-week period and encouraging participants to use higher weights, stronger bands, or to choose a more difficult variation of an exercise, which was provided by the trainer. In addition, more unilateral exercise variations were included from week 7 onward.

Although the main focus was on training around 12 - 15 reps (2 - 3 seconds for the eccentric movement and 1 - 2 seconds for the concentric movement), a small number of exercises were performed isometrically (ie, wall-sit and plank) or with a focus on eccentric load (downward movement of a one-leg box squat). From week 13 onward, we introduced power training for the leg muscles in 50% of the trainings (ie, squats, lunges, and kettlebell swings) performed at high speed to stimulate improvements in physical functioning.

XSL•FO
RenderX

Training sessions were prerecorded by a fitness trainer and broadcasted at set times. A fitness trainer was present on the web at the start and end of the broadcast to welcome participants and answer questions. Participants could also perform a missed training at a time of their preference.

Increased protein intake

Before the start of the program, participants recorded their nutritional intake for 3 days in the Traqq app [18]. Based on their nutritional intake, participants received personal advice from the dietitian about increasing their protein intake up to 25 grams during breakfast, lunch, and dinner. A protein intake of 25 grams for each main meal was chosen because it has been shown to be the requirement for appropriate muscle protein synthesis [19]. A lower protein intake may attenuate the skeletal muscle protein synthesis response in older adults [20]. In addition, guidelines of the effective ProMuscle program, in which a dietary protein intake of 25 grams per main meal was advised, were used in the Uni2Move program [13]. Consultations took place via a web-based video call or by phone. During the evaluation consultation in week 6, the advice was adjusted if necessary. From week 13 onward, participants enrolled in a web-based nutrition module. The nutrition module consisted of 12 submodules, with a new submodule released every week. Various themes related to proteins were discussed in the form of videos, short texts, assignments, quizzes, and recipes. Each participant accessed the module with a personal account and performed it at their own pace.

The pilot study did not include a comparative control group of participants because of the previous ProMuscle studies performed. ProMuscle is an in-person program and has been proven to be effective in controlled and practice settings. This pilot study was an exploration to study the effects of ProMuscle when offered completely on the internet. In case of positive results, a larger study could be conducted including a larger study population and a control group.

Participants

Adults aged 55 years and older living in and around Wageningen (The Netherlands) were recruited through flyers and news items in local newspapers. The participant's eligibility for the web-based program was assessed through a physical screening; non-frail adults were allowed to participate. Participants' ability to use the internet was assessed using two questions: (1) are you able to use email? (yes/no); (2) are you able to use internet explorer? (yes/no). Their ability to adjust the performance of exercises based on instructions was assessed by visual inspection of the fitness trainer to ensure safe participation. Participants were excluded during screening if participation might not be safe, in cases of medical limitations (such as severe heart failure or recovering from hip or knee arthroplasty), or in cases of an uncontrolled disease (ie, diabetes type 1 or 2, chronic obstructive pulmonary disease, hypertension, and cancer).

In total, 21 middle-aged and older adults were eligible for inclusion. After screening, 2 participants were excluded. Reasons for exclusion were (1) computer technical problems, resulting in them being unable to follow the web-based intervention and (2) not specified by the participant.

Measurements and Analyses

Quantitative Measures

Questionnaires were used to collect baseline characteristics including age, sex, ethnicity, education level, care use, and sports. Body weight was measured to the nearest 0.1 kg using a digital scale and height was measured to the nearest 0.1 cm using a stadiometer. Measurements were performed at week 0, 13, and 25. Uni2Move was based on the effective ProMuscle program, which has a duration of 25 weeks, and effectiveness was measured at 0, 13, and 25 weeks [12]. In order to be able to compare Uni2Move (web-based program) to ProMuscle (in-person program) measurements of Uni2Move were also performed at the beginning, half way, and end of the program (0, 13, and 25 weeks). In addition, strength training needed to be performed for at least 6 months to assure maintenance of muscle strength and mass [21]. Muscle strength was measured by a fitness trainer with a 1-repetition maximum knee extension test using the leg extension machine. Physical performance was measured by the repeated chair rise test. The chair rise test was measured as the time to complete 5 chair rises with hands folded against the chest. The same chair was used during the measurements in week 0, 13, and 25. The measurements were performed according to a standardized protocol. In addition, fall frequency was measured using a questionnaire. Nutritional intake was measured for 3 days with the Traqq app in weeks 0, 13, and 25. The average protein intake over 3 days was calculated in gram per day, g/kg body weight/d and per meal moment (breakfast, lunch, or dinner).

Adherence

Training sessions were broadcasted using Webinargeek. The number of participants that watched the training session was tracked via Webinargeek. The percent adherence was calculated as the number of training sessions watched, divided by the number of training sessions that were provided, multiplied by 100. Adherence of the dietary consultations was registered by the dietitian for each participant separately. Engagement in the web-based nutrition module was registered for each submodule by the researchers by checking the percentage of participants that completed the submodule.

Acceptance

During the measurement in week 25, acceptance of the program was checked by asking the following question: "What grade do you give the Uni2Move program on a scale of 1 - 10?" The question was asked for the total program and separately for the training sessions, as well as for the consultations with the dietitian and the nutrition module.

Qualitative Measures

A trained research assistant conducted semistructured interviews with participants to gain insight into experiences with the web-based program. Interview questions were based on a pretested interview guide from the ProMuscle in Practice Study and the guide was supplemented with questions specifically focusing on the web-based program [22]. The goal of the interview was to answer the following research questions:

XSL•FO
RenderX

1. How did participants experience the web-based training sessions?
2. How did participants experience the web-based dietary consultations?
3. How did participants experience the web-based nutrition module?

Statistical Analyses

Univariate procedures were used to check for normal distribution of the data. Baseline data were expressed as means with SD or as percentages. Effects on muscle strength and physical performance were analyzed using linear mixed models in SPSS version 29 (IBM Corp). Fixed factors were age, sex, and measuring time. Statistical significance was indicated with a *P* value of <.05. Qualitative data were analyzed in Atlas.ti version 22 (Lumivero). Interviews were taped and transcribed verbatim, and transcripts from the interviews were coded using an inductive approach. First, the interviews were read to familiarize with the data. After that, parts of the transcript were identified and classified as open codes. Finally, repeating patterns were identified and sorted in themes.

Ethical Considerations

The research team of the study officially consulted the medical ethical committee, with the question if medical ethical approval was necessary. No medical ethical approval was needed as this research did not fall within the remit of the Dutch Medical Research Involving Human Subjects Act (in Dutch: WMO; File number METC East-Netherlands: 2021 - 13365). All steps in this study have been conducted ethically. Participants provided informed consent before participation [23]. Privacy was guaranteed by working via SmartPIA procedures. Participants were covered (for incidents related to the intervention) by the insurance of the organization. Study data were deidentified as every participant was assigned a participant number. Participants were not identifiable from participant numbers. Participants did not receive a compensation for the research. Since Uni2Move is a program that is supposed to be offered regularly, we asked participants to pay a set price of 157 euros (US $161.58) in total. We aimed to include older adults with a low socioeconomic position (SEP), by letting them participate for a reduced price of 30 euros (US $30.88) in total.

Results

Baseline Characteristics

A total of 21 community-dwelling older adults were eligible to participate, and 19 participants could be included. The mean age at baseline was 69 (SD 7) years and the majority were female (n=14, 74%; Table 1).

XSL•FO
RenderX

Table . Baseline characteristics of the intervention group (n=19).

Characteristic	Value
Age (years), mean (SD)	69 (7)
Sex, n (%)	
Female	14 (74)
Male	5 (26)
Height (cm), mean (SD)	168.4 (9.4)
Weight (kg), mean (SD)	71.4 (14.6)
Native Dutch ethnicity, n (%)	19 (100)
Education level, n (%)	
Low[a]	0 (0)
Intermediate[b]	2 (10)
High[c]	17 (90)
Receiving care[d], n (%)	
Yes	1[e] (5)
No	18 (95)
Doing sports weekly, n (%)	
Yes	15 (79)
No	4 (21)
Type of sports, n (%)	
Hiking	11 (58)
Cycling	9 (47)
Nordic walking	2 (10)
Running	1 (5)
Aqua gym	1 (5)
Swimming	1 (5)
Yoga	1 (5)
Tai chi	1 (5)
Dancing	1 (5)
Canoeing	1 (5)
Golf	1 (5)
Total exercise hours per week, mean (SD)	4 (2)

[a]Low: primary education, prevocational secondary education, secondary vocational education level 1, undergraduate senior general secondary education, or undergraduate preuniversity education.

[b]Intermediate: secondary vocational education level 2, 3, or 4; senior general secondary education; or preuniversity education.

[c]High: higher vocational education or university.

[d]Care was defined as informal care, housekeeping, personal care, nursing care, individual counseling, or group counseling.

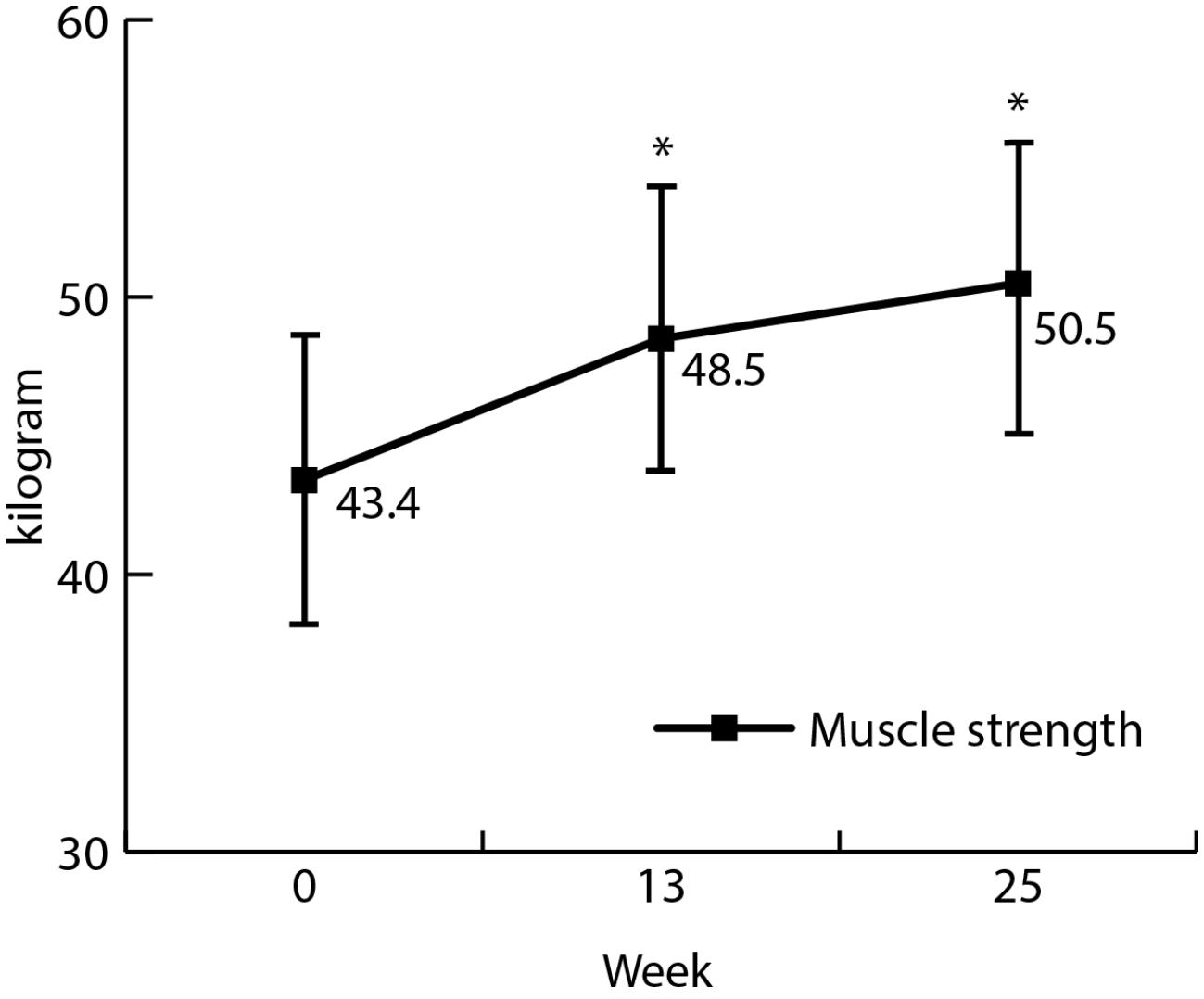
[e]Care received was housekeeping, once a week.

Quantitative Study

Results are presented in Table 2 and Figures 1 and 2. Participants significantly improved knee extension strength after 13 weeks by 5.1 kg (95% CI 2.8-7.3; *P*<.001) and by 7.0 kg (95% CI 4.8-9.3; *P*<.001) after 25 weeks, compared to baseline. Performance on the repeated chair test improved significantly after 13 weeks (−0.9 seconds, 95% CI −1.5 to −0.3; *P*=.001) and 25 weeks (−1.2 seconds, 95% CI −1.7 to −0.6; *P*<.001). Frequency of falls in the past 6 months decreased from 3 out of 17 (18%) at baseline to 1 out of 17 (6%) in weeks 13 and 25. Total protein intake did not change significantly between baseline and week 13 or week 25. Protein intake during breakfast increased significantly by 6.9 g (95% CI 1.1-12.7; *P*=.01) between week 0 and 13. Mean differences of protein intake during other meal moments did not change significantly.

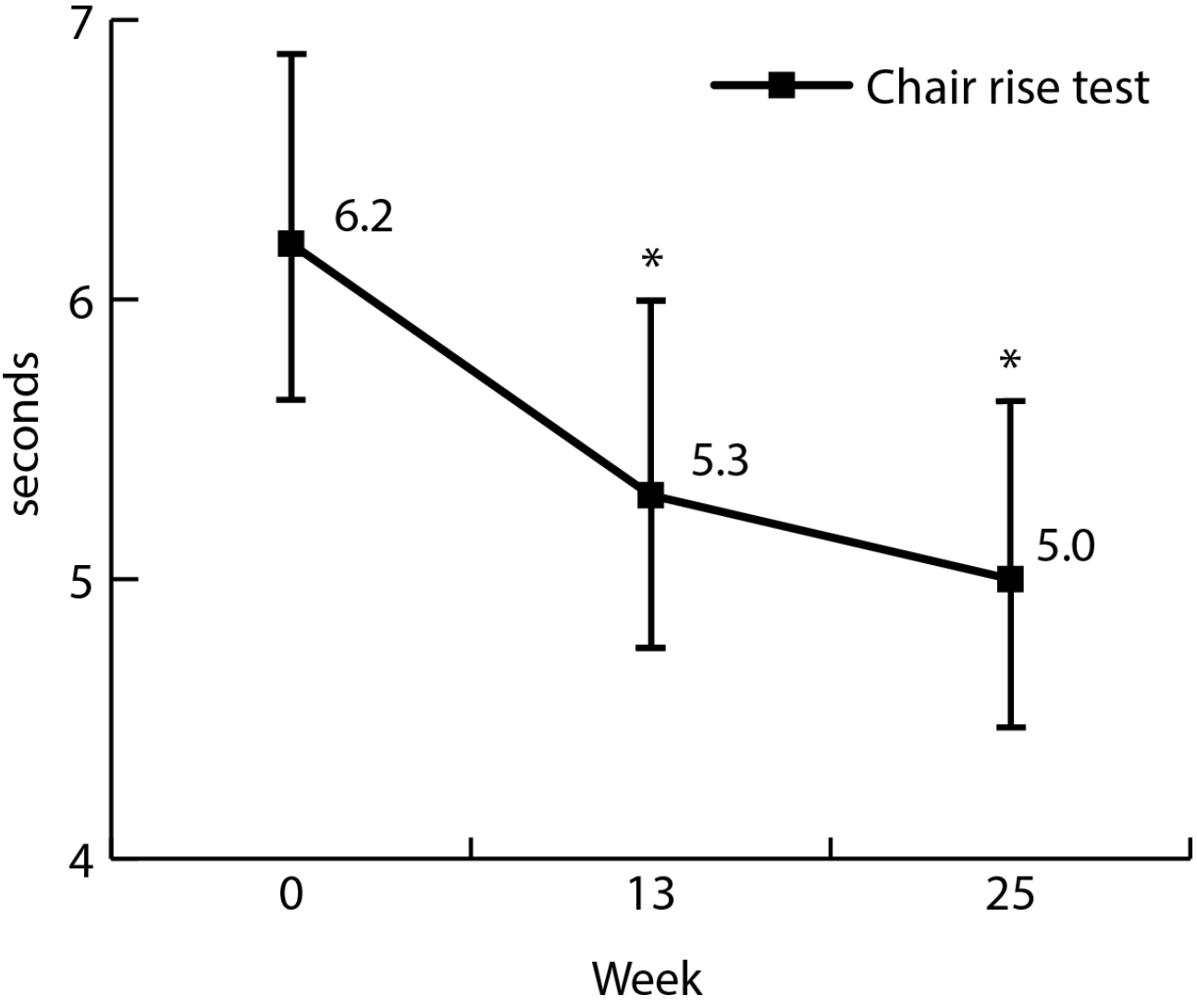
XSL•FO
RenderX

Figure 1. Effects of muscle strength in kilogram at week 0, 13, and 25. Error bars indicate SD. *Indicates a significant increase compared to baseline (*P*<.05).



XSL•FO
RenderX

Figure 2. Effects of repeated chair rise test in seconds at week 0, 13, and 25. Error bars indicate SD. *Indicates a significant increase compared to baseline (*P*<.05).



XSL•FO
RenderX

Table . Study outcomes at week 0, 13, and 25 and the time effects of weeks 0 - 25, 0 - 13, and 13 - 25.

Study outcome	Week 0, mean (95% CI)	Week 13, mean (95% CI)	Week 25, mean (95% CI)	Week 0 - 25		Week 0 - 13		Week 13 - 25	
				Mean difference (95% CI)	*P* value	Mean difference (95% CI)	*P* value	Mean difference (95% CI)	*P* value
Chair rise test (s)	6.2 (5.6-6.9)	5.3 (4.7-6.0)	5.0 (4.4-5.7)	−1.2 (−1.7 to −0.6)	<.001	−0.9 (−1.5 to −0.3)	.001	−0.3 (−0.9 to 0.3)	.65
1RM[a] knee extension (kg)	43.4 (38.5-48.3)	48.5 (43.6-53.4)	50.5 (45.5-55.4)	7.0 (4.8-9.3)	<.001	5.1 (2.8-7.3)	<.001	2.0 (−0.3 to 4.2)	.10
Protein intake (g/day)**	79.4 (65.6-93.2)	90.2 (75.9-104.4)	85.2 (70.3-100.2)	5.9 (−12.6 to 24.3)	>.99	10.8 (−6.9 to 28.5)	.39	−4.9 (−23.5 to 13.7)	>.99
Protein intake (g/kg body weight/day)	1.2 (0.9-1.5)	1.3 (1.0-1.6)	1.2 (0.9-1.5)	0.1 (−0.2 to 0.4)	>.99	0.2 (−0.1 to 0.4)	.54	−0.1 (−0.4 to 0.2)	>.99
Protein intake breakfast (g/day)	16.1 (10.9-21.3)	23.0 (17.6-28.4)	20.3 (14.6-26.1)	4.2 (−2.1 to 10.5)	.32	6.9 (1.1-2.7)	.01	−2.7 (−9.2 to 3.8)	.94
Protein intake lunch (g/day)	22.0 (17.1-27.0)	25.5 (20.0-31.0)	28.6 (22.9-34.4)	6.6 (−0.4 to 13.6)	.07	3.5 (−3.3 to 10.3)	.65	3.1 (−4.4 to 10.6)	.95
Protein intake dinner (g/day)	33.6 (25.9-41.4)	37.6 (29.2-45.9)	35.1 (26.3-44.0)	1.5 (−8.6 to 11.6)	>.99	3.9 (−5.6 to 13.5)	.96	−2.5 (−13.2 to 8.3)	>.99

[a]1RM: 1 repetition maximum.

Adherence

Participants followed on average 72% of the training sessions that were offered in the 24-week program. In the first 12 weeks, participants followed 77% of the training sessions, as compared to 67% in the second 12 weeks. The percentage of participants that participated in the web-based nutrition module, consisting of 12 modules, differed per week. On average, 80% of the participants performed the first 6 modules, and on average, 60% of the participants conducted the last 6 modules.

Qualitative Study

Participants' experiences were retrieved with interviews (n=7). The average length of the interviews was 27 minutes. For the interviewees, 6 out of 7 (86%) were woman with an average age of 66 years. The focus of the interviews was gaining insights in the experiences with the web-based training sessions, consultations, and nutrition module. Participants rated the total Uni2Move program 8.5 on average (scale 1 - 10).

Experiences With Web-Based Training Sessions

The training sessions were rated 8.6 on average (scale 1 - 10). Most of the participants did not follow the web-based training sessions live but caught up at a time of their preference. They indicated several advantages of a web-based program, such as no need to travel, exercising in their own environment, and the possibility to train at a self-determined time. Participants appreciated that the fitness trainer provided variations to exercises, for example receiving an alternative exercise if they had a prosthesis.

> *The benefit is that you have no travel time of course. All I have to do here is walk from one room to another.* [Participant, female]

Disadvantages mentioned by participants were the lack of physical interaction, social control, training in a group, and corrections by the trainer. Some participants also indicated that more discipline was needed.

> *You need a little more discipline to think: tomorrow morning I am going to do it.* [Participant, female]

> *A drawback is that there is a little bit less correction. You need to think by yourself; How do I place my back and how do I stand? There is nobody who says; Hello, a bit to the front, a bit to the back.* [Participant, female]

Experiences With Web-Based Dietary Consultations

Participants rated the consultations 7.3 on average (scale 1 - 10). Participants appreciated the fact that the consultations with the dietitian could be planned at any time of day. In addition, the form of the consultation was flexible and participants mentioned that the dietitian always checked whether the information they provided was sufficient.

> *She always ended the consultation with asking whether the advise she provided was clear and*

XSL•FO
RenderX

sufficient, and for me it was fulfilling. [Participant, female]

For several participants, the web-based consultations made the contact with the dietitian less personal, leaving limited room for discussion. In addition, the consultations were a challenge for a minority of the participants. Due to limited digital skills, video calling was challenging for them. As an alternative, these participants conducted phone consultations.

Experiences With the Web-Based Nutrition Module

The web-based nutrition module was rated 7.1 on average (scale 1 - 10). Participants mentioned that the level of difficulty for the modules was suitable. However, some participants thought it should be more challenging. Participants appreciated the schedule of the nutrition module, with a new module being presented on the web every week. In addition, the different elements of the nutrition module (knowledge clips, text, quizzes, and recipes) were mentioned as assets.

Participants missed the connection between the exercise and nutrition program, as they were offered separately and integration between the two was not present. According to the participants, a feature that should be added to the nutrition module is the possibility to have an web-based conversation in order to gather experiences from other users. On the content level, participants mentioned the need for information on skipping a meal, as well as vegetarian and vegan diets.

Discussion

Principal Findings

To our knowledge this is the first study evaluating the effects of an intervention combining resistance exercise and an increased protein intake that is completely web-based. Participants significantly improved knee extension strength after 13 weeks by 5.1 kg (95% CI 2.8-7.3; $P<.001$) and by 7.0 kg (95% CI 4.8-9.3; $P<.001$) after 25 weeks, compared to baseline. Performance on the repeated chair test improved significantly after 13 weeks by −0.9 seconds (95% CI −1.5 to −0.3; $P=.001$) and after 25 weeks by −1.2 seconds (95% CI −1.7 to −0.6; $P<.001$). Protein intake during breakfast increased significantly between week 0 and 13. Total protein intake per day did not change significantly. Participants positively valued Uni2Move, rating the total program an 8.5 (scale 1 - 10). Advantages of the web-based training sessions and dietary consultations were the flexibility to conduct the program in their own environment and at a self-determined time. Challenges were the lack of personal contact and corrections by the trainer.

Web-Based vs Offline Program

Results of the web-based Uni2Move program are in line with the offline program ProMuscle in Practice for the chair rise test after 13 weeks, which were −0.9 seconds (15%) versus −0.8 seconds (6%), respectively, and knee extension strength improved by 5.1 kg (12%) versus 9.6 kg (14%), respectively [13]. These findings show that the offline program is a promising strategy in improving muscle health of middle-aged and older adults. However, several challenges of the web-based program in comparison to the face-to-face program should be addressed.

The average protein intake decreased after week 13, and the adherence dropped slightly during the course of the web-based program. Several participants indicated in the interviews that discipline was needed to proceed with the adjusted diet, and personal contact was lacking in the nutrition module. In line with this, Aalbers et al [10] concluded in their systematic review that tailored web-based interventions were rated as more interesting and helped middle-aged participants more, compared to generic web-based interventions [24]. This emphasizes the importance of personalizing web-based interventions and corresponding to the needs of participants. As suggested by participants, the web-based nutrition module can be improved by adding an option to chat with other participants. Social support is known to be an effective technique for the maintenance of a behavior change [25]. In addition, the nutrition module can be extended with advanced modules for extra challenge and additional information.

Strengths and Limitations

First, a major strength of this study is the combination of quantitative data and qualitative data. In this way, insights on the effects of the program are combined with experiences of participants with the practicalities and feasibility of the program. Second, the limited costs of the intervention are a strength. Costs of preventive interventions are always challenging for professionals as well as participants. A web-based intervention limits costs by reducing travel motions, guiding large groups of participants during web-based training, reusing training videos, and reusing the nutrition module. This decreases a possible cost-related barrier for participation.

Several limitations should be mentioned. First, as participants perform the workout in their own environment, the performance of the exercises is less regulated. As a consequence, the intensity of the performed exercises could not be measured. On the other hand, the results show improvements for the chair rise test as well as the knee extension strength, and adherence levels and rating of the training sessions are satisfactory. Based on that, we may assume that participants performed the exercises with a sufficient intensity.

Second, the lack of physical attendance during the intervention limits the possibility of measuring adherence to the program in an accurate way. Attendance to training sessions was based on whether participants watched the video of the training session. In a face-to-face setting, the fitness trainer would be able to check who is present and actively participates in the training sessions. In addition, fitness trainers are not able to correct participants during at home training sessions. During the interviews, several participants also indicated the lack of guidance and corrections on the performance of exercises, whereas others did not miss the guidance. For older adults who live independently at home and without complex disorders, web-based guidance can be suitable. Physical guidance on site may be appropriate for frail older people who need more assistance. Further research is warranted to carefully assess eligibility of the program in different groups.

Third, the small study population is a limitation of this pilot study. In addition, this study population was highly educated for the most part. In general, adults with a higher education

XSL•FO
RenderX

level have better health compared to their less-educated peers [26]. The latter group should also be included in future research to study the effects of Uni2Move on their muscle health and to contribute to reducing health inequalities. During the recruitment phase, several strategies that were shown to be successful strategies for reaching adults with a low SEP were executed in order to include participants with a lower SEP [24]. Multilayered recruitment strategies were conducted to increase engagement. First, we adjusted the recruitment material by adopting the language level and a personal approach by recruiting at low budget shops with flyers suitable for adults with low literacy levels. Second, we deployed existing networks and community networks by recruiting at the local food bank and neighborhoods where many adults with a low SEP live. Third, we tackled the common barrier of having limited resources to join the program by lowering the price of participation from 157 euros (US $161.58) to 30 euros (US $30.88) for participants with a low SEP. These strategies led to the inclusion of 2 participants with an intermediate education level. SEP was not measured because of participants' privacy. Recently, Collombon et al [4] published a paper on recruitment strategies for reaching adults aged 50 years and older with low SEP for participation in web-based physical activity interventions [27]. Results show that recruitment via invitation letters administered by municipality was most promising; however, the costs were relatively high (>2000 euros; US $2058.34). This strategy could be deployed in future research to include a diverse group of participants.

Future Directions

The main challenges of the web-based intervention include recruiting participants with a low SEP and adherence to the program. Recruitment of more participants with a low SEP should be attempted in follow-up studies by sending invitations via municipalities, as well as by focusing on recruitment on social media and at gyms [27]. Adherence to the program can be increased in several ways. Promising strategies appear to be increasing personal contact and tailoring the intervention to the needs of participants. This can be executed by conducting a hybrid intervention, alternating web-based and physical training sessions. In this way, fitness trainers can tailor the level of the exercises to the needs of the participants and can correct participants if needed. Personal contact evolves during training sessions, not only between participants but also between participants and fitness trainers. The advantages of a hybrid intervention are lower participation costs, flexibility of training sessions at home, less health care professionals needed, increased safety due to face-to-face training sessions, and expectedly higher adherence on the long term due to tailoring and personal contact.

Conclusion

In summary, the web-based lifestyle intervention Uni2Move is a pilot study that resulted in an improvement of muscle strength and performance on the repeated chair rise test, comparable to the previous offline ProMuscle studies. The results should be interpreted with caution, as the study population is small. Advantages of the current intervention are flexibility for participants as well as professionals, as well as a lack of travel time and costs. Further research is needed to evaluate the intervention's effects in larger groups, and should include participants with a low SEP, be in a hybrid form, and focus on adherence in the long term.

Acknowledgments

Within Uni2Move, Wageningen University collaborated with CleverMove and dietitians from Zoeetje. CleverMove is specialized in guiding people over 55 with (web-based) strength training. Zoeetje is specialized in advising on nutrition in health and disease. We would like to thank the fitness trainer of CleverMove, the dietitians of Zoeetje, the participants of Uni2Move, Susanne Wildemast, and the students, Eline Jonkers and Carlijn de Vries, involved in conducting the Uni2Move program, for their contribution. The research described was funded by ZonMw (05460402110005). The development of the web-based nutrition module was financed by the Rijksinstituut voor Volksgezondheiden Milieu (RIVM). These organizations had no role in the design, analysis, or writing of this article. The study was funded by ZonMw.

Data Availability

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: BGD, NW, LCPGMG, PG
Data curation: BGD, PG
Formal analysis: BGD
Funding acquisition: BGD, NW, LCPGMG, PG
Investigation: BGD
Methodology: BGD, NW, LCPGMG, PG
Project administration: BGD, NW, PG
Resources: BGD, NW, PG
Software: BGD
Supervision: BGD, NW, LG, PG

XSL•FO
RenderX

Validation: BGD, NW, PG
Visualization: BGD
Writing – original draft: BGD
Writing – review & editing: NW, LG, PG

Conflicts of Interest

NW is owner of CleverMove, which is a commercial organization. NW conducted the training session of Uni2Move. NW had no role in the analyses or interpretation of the data.

References

1. Shen Y, Liu D, Li S, et al. Effects of exercise on patients important outcomes in older people with sarcopenia: an umbrella review of meta-analyses of randomized controlled trials. Front Med (Lausanne) 2022;9:811746. [doi: 10.3389/fmed.2022.811746] [Medline: 35186999]
2. Tieland M, Dirks ML, van der Zwaluw N, et al. Protein supplementation increases muscle mass gain during prolonged resistance-type exercise training in frail elderly people: a randomized, double-blind, placebo-controlled trial. J Am Med Dir Assoc 2012 Oct;13(8):713-719. [doi: 10.1016/j.jamda.2012.05.020] [Medline: 22770932]
3. Wet medisch-wetenschappelijk onderzoek met mensen. Overheid.nl. 2024. URL: https://wetten.overheid.nl/BWBR0009408/2025-01-01 [accessed 2025-01-23]
4. Collombon E, Bolman CAW, de Bruijn GJ, Peels DA, Lechner L. Recruitment strategies for reaching adults aged 50 years and older with low socioeconomic status for participation in online physical activity interventions. Front Digit Health 2024;6:1335713. [doi: 10.3389/fdgth.2024.1335713] [Medline: 38800097]
5. van Dongen EJI, Haveman-Nies A, Doets EL, Dorhout BG, de Groot L. Effectiveness of a diet and resistance exercise intervention on muscle health in older adults: ProMuscle in Practice. J Am Med Dir Assoc 2020 Aug;21(8):1065-1072. [doi: 10.1016/j.jamda.2019.11.026] [Medline: 31948853]
6. Duurzame zorg en preventie (SPR 2019-2022). : Rijksinstituut voor Volksgezondheiden Milieu; 2024 URL: https://www.rivm.nl/over-het-rivm/strategisch-programma-rivm/duurzame-zorg-en-preventie [accessed 2025-01-23]
7. Kwasnicka D, Dombrowski SU, White M, Sniehotta F. Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories. Health Psychol Rev 2016 Sep;10(3):277-296. [doi: 10.1080/17437199.2016.1151372] [Medline: 26854092]
8. Izquierdo M, Merchant RA, Morley JE, et al. International exercise recommendations in older adults (ICFSR): expert consensus guidelines. J Nutr Health Aging 2021;25(7):824-853. [doi: 10.1007/s12603-021-1665-8] [Medline: 34409961]
9. van Dongen EJI, Doets EL, de Groot L, Dorhout BG, Haveman-Nies A. Process evaluation of a combined lifestyle intervention for community-dwelling older adults: ProMuscle in Practice. Gerontologist 2020 Nov 23;60(8):1538-1554. [doi: 10.1093/geront/gnaa027] [Medline: 32259835]
10. Aalbers T, Baars MAE, Rikkert M. Characteristics of effective Internet-mediated interventions to change lifestyle in people aged 50 and older: a systematic review. Ageing Res Rev 2011 Sep;10(4):487-497. [doi: 10.1016/j.arr.2011.05.001] [Medline: 21628005]
11. Ouderen. Statistiek CB voor de. 2024. URL: https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/leeftijd/ouderen [accessed 2025-01-17]
12. Coletta G, Phillips SM. An elusive consensus definition of sarcopenia impedes research and clinical treatment: a narrative review. Ageing Res Rev 2023 Apr;86:101883. [doi: 10.1016/j.arr.2023.101883] [Medline: 36792012]
13. Paddon-Jones D, Rasmussen BB. Dietary protein recommendations and the prevention of sarcopenia. Curr Opin Clin Nutr Metab Care 2009 Jan;12(1):86-90. [doi: 10.1097/MCO.0b013e32831cef8b] [Medline: 19057193]
14. Lucassen DA, Brouwer-Brolsma EM, van de Wiel AM, Siebelink E, Feskens EJM. Iterative development of an innovative smartphone-based dietary assessment tool: Traqq. J Vis Exp 2021 Mar 19(169). [doi: 10.3791/62032] [Medline: 33818566]
15. Cruz-Jentoft AJ, Bahat G, Bauer J, et al. Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing 2019 Jul 1;48(4):601. [doi: 10.1093/ageing/afz046] [Medline: 31081853]
16. UN decade of healthy ageing plan of action 2021-2030. : World Health Organization; 2020 URL: https://cdn.who.int/media/docs/default-source/decade-of-healthy-ageing/decade-proposal-final-apr2020-en.pdf [accessed 2025-01-17]
17. Hill KD, Hunter SW, Batchelor FA, Cavalheri V, Burton E. Individualized home-based exercise programs for older people to reduce falls and improve physical performance: a systematic review and meta-analysis. Maturitas 2015 Sep;82(1):72-84. [doi: 10.1016/j.maturitas.2015.04.005] [Medline: 25989701]
18. Katsanos CS, Kobayashi H, Sheffield-Moore M, Aarsland A, Wolfe RR. Aging is associated with diminished accretion of muscle proteins after the ingestion of a small bolus of essential amino acids. Am J Clin Nutr 2005 Nov;82(5):1065-1073. [doi: 10.1093/ajcn/82.5.1065] [Medline: 16280440]
19. Global strategy on human resources for health: workforce. : World Health Organization; 2016 URL: https://www.who.int/publications/i/item/9789241511131 [accessed 2025-01-17]

XSL•FO
RenderX

20. Dorhout BG, de Groot L, van Dongen EJI, Doets EL, Haveman-Nies A. Effects and contextual factors of a diet and resistance exercise intervention vary across settings: an overview of three successive ProMuscle interventions. BMC Geriatr 2022 Mar 9;22(1):189. [doi: 10.1186/s12877-021-02733-6] [Medline: 35264105]
21. Bautmans I, Van Puyvelde K, Mets T. Sarcopenia and functional decline: pathophysiology, prevention and therapy. Acta Clin Belg 2009;64(4):303-316. [doi: 10.1179/acb.2009.048] [Medline: 19810417]
22. Raghupathi V, Raghupathi W. The influence of education on health: an empirical assessment of OECD countries for the period 1995-2015. Arch Public Health 2020;78:20. [doi: 10.1186/s13690-020-00402-5] [Medline: 32280462]
23. Coelho-Júnior HJ, Milano-Teixeira L, Rodrigues B, Bacurau R, Marzetti E, Uchida M. Relative protein intake and physical function in older adults: a systematic review and meta-analysis of observational studies. Nutrients 2018 Sep 19;10(9):1330. [doi: 10.3390/nu10091330] [Medline: 30235845]
24. Nelson ME, Layne JE, Bernstein MJ, et al. The effects of multidimensional home-based exercise on functional performance in elderly people. J Gerontol A Biol Sci Med Sci 2004 Feb;59(2):154-160. [doi: 10.1093/gerona/59.2.m154] [Medline: 14999030]
25. Rosenberg IH. Sarcopenia: origins and clinical relevance. J Nutr 1997 May;127(5 Suppl):990S-991S. [doi: 10.1093/jn/127.5.990S] [Medline: 9164280]
26. Johnson NR, Kotarsky CJ, Mahoney SJ, et al. Evenness of dietary protein intake is positively associated with lean mass and strength in healthy women. Nutr Metab Insights 2022;15:11786388221101829. [doi: 10.1177/11786388221101829] [Medline: 35734029]
27. van den Helder J, Mehra S, van Dronkelaar C, et al. Blended home-based exercise and dietary protein in community-dwelling older adults: a cluster randomized controlled trial. J Cachexia Sarcopenia Muscle 2020 Dec;11(6):1590-1602. [doi: 10.1002/jcsm.12634] [Medline: 33103379]

Abbreviations

SEP: socioeconomic position

Edited by A Mavragani; submitted 14.11.23; peer-reviewed by B Egan, M Stein; revised version received 15.11.24; accepted 02.12.24; published 30.01.25.

Please cite as:
Dorhout BG, Wezenbeek N, de Groot LCPGM, Grootswagers P
Web-Based Exercise and Nutrition Intervention to Improve Leg Muscle Strength and Physical Functioning in Older Adults: Pre-Post Pilot Study
JMIR Form Res 2025;9:e54392
URL: https://formative.jmir.org/2025/1/e54392
doi:10.2196/54392

© Berber Gijsbertha Dorhout, Nick Wezenbeek, Lisette C P G M de Groot, Pol Grootswagers. Originally published in JMIR Formative Research (https://formative.jmir.org), 30.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Identifying Digital Markers of Attention-Deficit/Hyperactivity Disorder (ADHD) in a Remote Monitoring Setting: Prospective Observational Study

Heet Sankesara[1], BTech; Hayley Denyer[2], PhD; Shaoxiong Sun[1,3], PhD; Qigang Deng[2], MSc; Yatharth Ranjan[1], MSc; Pauline Conde[1], BSc; Zulqarnain Rashid[1], PhD; Philip Asherson[2], MBBS, PhD; Andrea Bilbow[4]; Madeleine J Groom[5,6], PhD; Chris Hollis[6], PhD FRCPsych; Richard J B Dobson[1,7,8,9,10], PhD; Amos Folarin[1,7,8,9,10], PhD; Jonna Kuntsi[2], PhD

[1]Department of Biostatistics & Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, 16 De Crespigny Park, London, United Kingdom

[10]NIHR Biomedical Research Centre, University College London Hospitals NHS Foundation Trust, London, United Kingdom

[2]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[3]Department of Computer Science, University of Sheffield, Sheffield, United Kingdom

[4]Attention Deficit Disorder Information and Support Service (ADDISS), Edgware, Middlesex, United Kingdom

[5]School of Medicine, Mental Health & Clinical Neurosciences, Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom

[6]NIHR MindTech Healthcare Technology Co-operative, Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom

[7]Institute of Health Informatics, University College London, London, United Kingdom

[8]NIHR Biomedical Research Centre, South London and Maudsley NHS Foundation Trust and King's College London, London, United Kingdom

[9]Health Data Research UK London, University College London, London, United Kingdom

Corresponding Author:
Heet Sankesara, BTech
Department of Biostatistics & Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, 16 De Crespigny Park, London, United Kingdom

Abstract

Background: The symptoms and associated characteristics of attention-deficit/hyperactivity disorder (ADHD) are typically assessed in person at a clinic or in a research lab. Mobile health offers a new approach to obtaining additional passively and continuously measured real-world behavioral data. Using our new ADHD remote technology (ART) system, based on the Remote Assessment of Disease and Relapses (RADAR)–base platform, we explore novel digital markers for their potential to identify behavioral patterns associated with ADHD. The RADAR-base Passive App and wearable device collect sensor data in the background, while the Active App involves participants completing clinical symptom questionnaires.

Objective: The main aim of this study was to investigate whether adults and adolescents with ADHD differ from individuals without ADHD on 10 digital signals that we hypothesize capture lapses in attention, restlessness, or impulsive behaviors.

Methods: We collected data over 10 weeks from 20 individuals with ADHD and 20 comparison participants without ADHD between the ages of 16 and 39 years. We focus on features derived from (1) Active App (mean and SD of questionnaire notification response latency and of the time interval between questionnaires), (2) Passive App (daily mean and SD of response time to social and communication app notifications, the SD in ambient light during phone use, total phone use time, and total number of new apps added), and (3) a wearable device (Fitbit) (daily steps taken while active on the phone). Linear mixed models and *t* tests were employed to assess the group differences for repeatedly measured and time-aggregated variables, respectively. Effect sizes (*d*) convey the magnitude of differences.

Results: Group differences were significant for 5 of the 10 variables. The participants with ADHD were (1) slower (*P*=.047, *d*=1.05) and more variable (*P*=.01, *d*=0.84) in their speed of responding to the notifications to complete the questionnaires, (2) had a higher SD in the time interval between questionnaires (*P*=.04, *d*=1.13), (3) had higher daily mean response time to social and communication app notifications (*P*=.03, *d*=0.7), and (4) had a greater change in ambient (background) light when they were actively using the smartphone (*P*=.008, *d*=0.86). Moderate to high effect sizes with nonsignificant *P* values were additionally observed for the mean of time intervals between questionnaires (*P*=.06, *d*=0.82), daily SD in responding to social and communication app notifications (*P*=.05, *d*=0.64), and steps taken while active on the phone (*P*=.09, *d*=0.61). The groups did not differ in the total phone use time (*P*=.11, *d*=0.54) and the number of new apps downloaded (*P*=.24, *d*=0.18).

XSL•FO
RenderX

Conclusions: In a novel exploration of digital markers of ADHD, we identified candidate digital signals of restlessness, inconsistent attention, and difficulties completing tasks. Larger future studies are needed to replicate these findings and to assess the potential of such objective digital signals for tracking ADHD severity or predicting outcomes.

(JMIR Form Res 2025;9:e54531) doi:10.2196/54531

KEYWORDS

ADHD; smartphones; wearable devices; mobile health; mHealth; remote monitoring; surveillance; digital markers; attention-deficit/hyperactivity disorder; behavioral data; real world; adult; adolescent; participants; digital signals; restlessness; severity; predicting outcomes

Introduction

The core clinical symptoms of attention-deficit/hyperactivity disorder (ADHD), which affects 5.9% of children and 2.5% of adults, are inattention, impulsivity, and hyperactivity [1]. The diagnosis relies on a semistructured interview designed to evaluate the diagnostic criteria for ADHD symptoms and associated functional impairments. Cognitive tasks, which are often used in research, also reveal behavioral responses that suggest lapses in attention or impulsive responses in people with ADHD [2-4]. For example, ADHD is strongly associated with increased reaction time variability: people with ADHD are highly inconsistent in their speed of responding to cognitive tasks, which is linked to difficulties with the regulation of attention and arousal [5,6].

Beyond the conventional in-person clinic- or lab-based assessments on clinical symptoms and cognitive functioning, the field of mobile health (mHealth) offers a new approach to obtaining additional, more detailed, and longer-term behavioral data in the real world. Passive monitoring using smartphone sensors enables ongoing, objective, and unobtrusive data collection while the participant continues with their everyday activities [7-10]. The real-world behaviors captured using such passive monitoring may, for example, provide digital signals of ADHD severity or predict outcomes. Emerging evidence from research on other disorders, such as depression and schizophrenia, indicates the potential for passive smartphone data to identify markers related to symptom severity or warning signs of relapse [11,12].

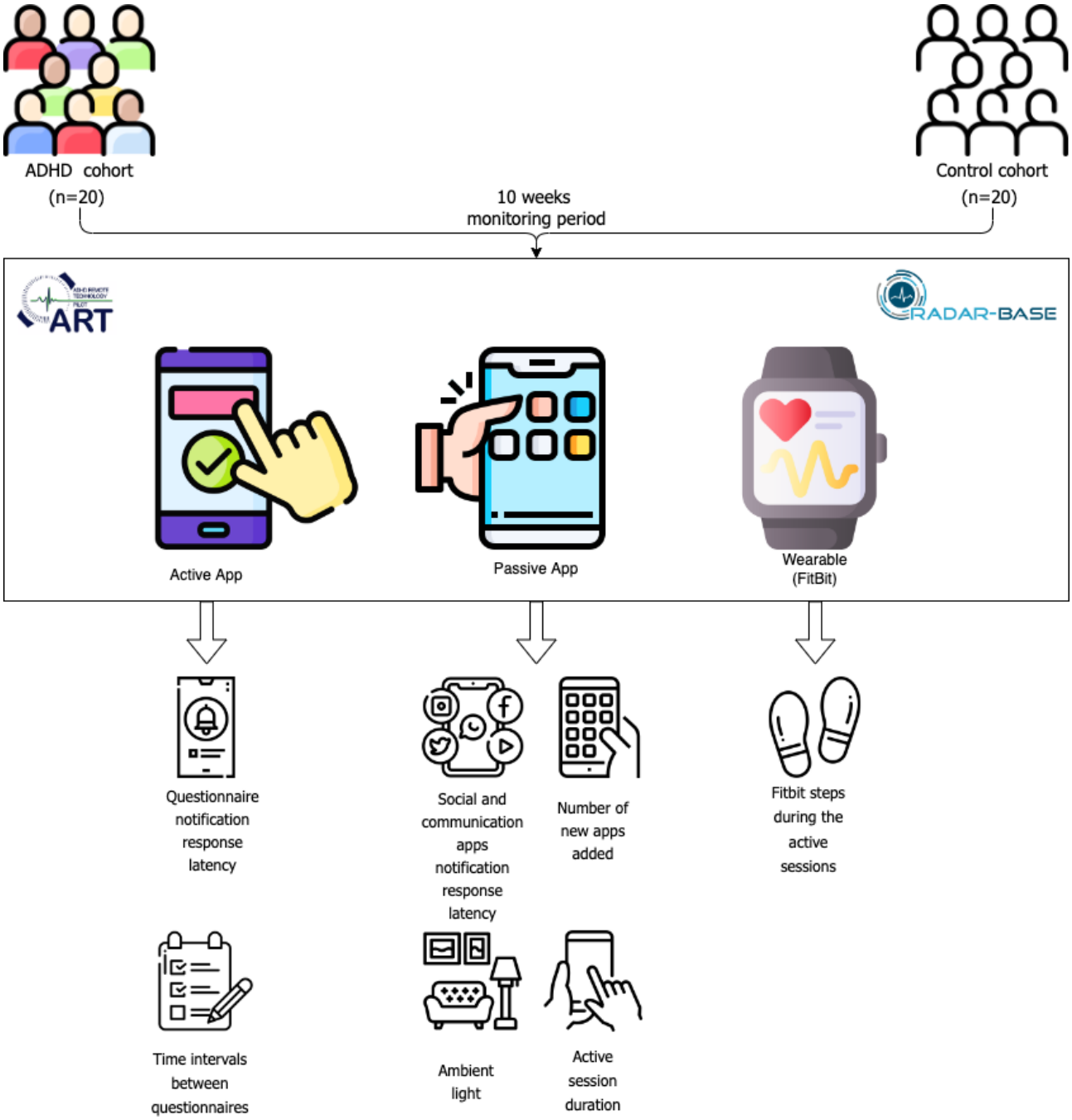
We recently developed a new remote measurement technology system for adults and adolescents with ADHD (older than 16 years)—the ADHD remote technology (ART) system—which uses the Remote Assessment of Disease and Relapses (RADAR)–base mHealth platform [13]. RADAR-base is an open-source platform to leverage data from wearable devices and mobile technologies. It provides scalable and customizable capabilities for remote real-time data collection from a wide range of sources and apps, providing a unified system for researchers to store, manage, and analyze the collected data. The RADAR-base Passive App collects background data from several sensors on modern smartphones, such as ambient noise, ambient light, phone use information (eg, which apps have been used and for how long), passive audio, GPS (relative) location, local Bluetooth device connectivity, battery life, gyroscope, steps, and acceleration from smartphone sensors. The Active App collects data requiring conscious effort (eg, questionnaires and other tasks), which are customizable according to study requirements. The platform sends notifications via the Active App to remind study participants to fill out the questionnaires and other prompts; related timestamps are recorded when the notification is delivered when participants start filling out the questionnaires and finish them.

Data previously collected using the Passive App in the RADAR-CNS (Remote Assessment of Disease and Relapse – Central Nervous System) long-term remote monitoring project [14] on adults with major depressive disorder revealed digital biomarkers that have the potential to predict depression severity [15]. For example, analyses focusing on a median follow-up period of 4 months showed that nearby Bluetooth device count—a proxy for social isolation—was associated with depression questionnaire scores [16]. The nearby Bluetooth device count can reflect the participants' social connections and interactions with family, friends, co-workers, and strangers. Therefore, the data can also reflect participants' time at home, mobility, social isolation, working status, and the number of other Bluetooth devices in the house and working environment. Further analyses of geolocation data over 24 months showed that an increase in depression symptom severity was preceded by an increase in time spent at home and a reduction in the average time spent spread across different locations. These findings from the RADAR-major depressive disorder study illustrate how data from the Passive App can be used to identify digital markers of the severity of psychiatric symptoms, supporting its application to other disorders, such as ADHD.

We aimed to quantify and investigate whether adolescents and adults with ADHD differ from people without ADHD on a range of measures hypothesized to capture lapses in attention or restless or impulsive behaviors. As shown in Figure 1, we focus on analyzing passively collected data from the wearable device and a dedicated Passive App. Similar to the behavioral data gathered by the Passive App, the analysis of data from the Active App explores participants' behavior during questionnaire completion. While the content of the questionnaires and cognitive data themselves are valuable for assessing ADHD and associated symptoms, they fall outside the scope of this particular paper. We aimed to investigate digital traces passively collected from participants' smartphones and wearable devices during their daily lives.

XSL•FO
RenderX

Figure 1. Graphical abstract showing study methodology and feature generation. A total of 20 participants with and 20 without ADHD were recruited in the United Kingdom between August and November 2020 and matched on age and gender. Participants completed the clinical symptom questionnaires via the Active App, while the smartphone sensor data were collected from the Passive App and the wearable sensor data from Fitbit. From the Active App data, the variables questionnaire notification response latency and time intervals between questionnaires are derived. Social and communication apps notification response latency, ambient light, number of new apps added, and active session duration are derived from the Passive App, whereas the Fitbit steps during the active sessions are derived from the wearable device. ART: attention-deficit/hyperactivity disorder remote technology; ADHD: attention-deficit/hyperactivity disorder; RADAR: Remote Assessment of Disease and Relapses.



Methods

Participants

Between August and November 2020, we recruited 20 individuals with ADHD and 20 comparison individuals without ADHD between the ages of 16 and 39 years into the study (Table 1). Participants were recruited from previous studies (where they had indicated that they were willing to be contacted regarding future research studies) via the Attention Deficit Disorder Information and Support Service using an advert placed on their website and emailed to members, social media (Facebook and Twitter), and placing an advert on the King's College London Volunteer circular and on the "Call for Participants" website. Exclusion criteria were as follows: (1) having psychosis, major depression, mania, drug dependency, or a major neurological disorder; (2) any other major medical condition that might impact upon the individual's ability to participate in normal daily activity; (3) pregnancy; and (4) IQ of <70. Participants in the ADHD group additionally met the diagnostic criteria for ADHD, and participants in the comparison

XSL•FO
RenderX

group could not meet the diagnostic criteria for ADHD based on the self-report on the Barkley Adult ADHD Rating Scale (BAARS) on current symptoms (BAARS-IV) and Barkley ADHD functional impairment questionnaire.

Table . Demographics for the groups of participants with and without attention-deficit/hyperactivity disorder (ADHD).

Group	Female, n (%)	Age (years), mean (SD)	WASI-II[a] vocabulary subscale, mean (SD)
ADHD (n=20)	15 (75)	27.49 (6.04)	57.85 (7.53)
Without ADHD (n=20)	15 (75)	27.79 (6.17)	56.80 (8.35)

[a]WASI-II: Wechsler Abbreviated Scale of Intelligence.

Ethical Considerations

The study was approved by the North East – Tyne and Wear South Research Ethics Committee (REC reference: 20/NE/0034). Informed consent was obtained from participants before the assessments started. To maintain participant anonymity, their data were pseudonymized, and participants were identified in the study data with a code. Data collected from the apps, wearable devices, and interviews were only associated with this code and stored separately from any personally identifiable information. Participants were compensated £30 (US $37) after completion of the baseline sessions, £20 (US $25) after the first remote active monitoring follow-up (end of week 5), and a further £50 (US $62) at the study end point (end of week 10). The participants did not receive additional compensation for completing the debrief interviews.

Procedure

ART is an observational, nonrandomized, noninterventional study using commercially available wearable technology and smartphone sensors, representing no change to the usual care or treatments of participants due to participation.

Participants attended 2 remote baseline sessions with a research worker using Microsoft Teams. The first remote baseline session with the participants with ADHD included the administration of the following assessments: (1) the Diagnostic Interview for ADHD in adults [17] to confirm ADHD diagnosis, (2) vocabulary and digit span subscales from the Wechsler Abbreviated Scale of Intelligence, and Wechsler Adult Intelligence Scale, respectively, and (3) web-based Research Electronic Data Capture (REDCap) [18] baseline questionnaires. The baseline REDCap questionnaires for the participants with ADHD included the About You (demographic) questionnaire, AQ-10 (autism), and the COVID-19 baseline questionnaire. The same was true for participants without ADHD however, in addition, they were required to complete the BAARS/Barkley Functional Impairment Scale (BFIS) (ADHD) to ensure they did not reach the research threshold for ADHD symptoms.

The second session was administered once participants had received their wearable device and smartphone by post, approximately a week after the first session. The second session included: (1) the administration of 2 cognitive tasks (combined cued continuous performance test and Go/NoGo task, and the Fast task) and (2) a training session on the use of the wearable device (Fitbit) and the smartphone Active and Passive Apps. The participant also received a leaflet summarizing key information (Participant Technology User Guide) and researcher contact details for future reference. Each session lasted for approximately 1.5 hours. Comparison participants were assessed in the same way, except that instead of the full ADHD diagnostic interview, they completed the ADHD symptom and impairment questionnaire [19].

During the baseline session, the participants with ADHD were asked to identify a partner, parent, or close friend who could complete informant-report versions of the active monitoring questionnaires on ADHD symptoms, impairment, and irritability using the web-based REDCap. The research worker contacted the informant and invited him/her to complete the questionnaires at each of the 3 remote active monitoring time points.

Each participant was, therefore, in the study for 10 weeks. The content of questionnaires and cognitive data are beyond the scope of this analysis. We describe below measures for variables used in this analyses.

Measures

Active App: Clinical Questionnaires

Active monitoring involved the participant completing clinical symptom questionnaires on the RADAR-base smartphone Active App and 2 cognitive tasks on their home PC or laptop, which took place 3 times: at 2 weeks (the first remote self-administered assessment), 6 weeks (the second remote self-administered assessment) and 10 weeks (the third remote self-administered assessment) after the baseline remote researcher-led session.

Participants were requested to fill out the following questionnaires every 2 weeks: Patient Health Questionnaire depression scale[19], Affective Reactivity Index questionnaire [20], BAARS-IV [21], Generalized Anxiety Disorder-7 [22], and Reactive-Proactive Aggression Questionnaire [23]. Participants with ADHD also completed a medication use questionnaire; however, we excluded this questionnaire from our analyses, as the comparison group was not asked to complete it. Participants were informed that the questionnaires would be available for 3 days; however, they were told to complete them as soon as it was safe to do so. As the aim of this paper focuses on the passive measures of ADHD classification, the responses to the questionnaires are beyond the scope of this paper.

Passive Monitoring Measures

Passive monitoring, where data are collected without conscious effort from the participants using the smartphone and the wearable device (Fitbit Charge 3), started from the second

XSL•FO
RenderX

baseline session and continued for 10 weeks for each participant. Passive data were collected continuously on a 24/7 basis. Participants with an existing Android phone had the option of either using their existing phone (1 participant with ADHD and 4 participants without ADHD) or upgrading it to a study Android phone (Motorola G7 Play or G7 Power; 3 participants with ADHD and 6 without ADHD). Participants who had an iPhone were asked to replace the iPhone with a study phone for the duration of the study (16 participants with ADHD and 10 without ADHD). Before enrolling in the study, participants were told they could only use an Android phone during the study period due to the availability of sensors and use data on Android devices. It was explained to them that due to our study design, they must only use one phone. During the study, the research team monitored phone battery status and phone app use to ensure compliance. Furthermore, we analyzed the battery discharging rate and found no significant differences in the discharging rate throughout the study period, confirming that the study phone was being used as the primary phone. Participants were asked to wear the wearable device on their nondominant hands.

The Passive App collected data on ambient noise, ambient light, phone use information (eg, which apps were used and how long, when the phone was unlocked), GPS location, Bluetooth connectivity, battery life, gyroscope, steps, and acceleration. Several passive measures were used to derive features to assess the smartphone use behavior of individuals with ADHD.

One such measure is phone status data. It logs the state of a smartphone at any given moment. The state can be "STANDBY" (device on standby), "UNLOCKED" (the device is unlocked), "SHUTDOWN" (the device is shut down), or "BOOTED" (the device is booting or starting). During the period between the device being unlocked and standby again, it is assumed that the participant is actively using the smartphone. We named this period an active phone use session.

The active phone use session can help us understand how long and when a participant is active on their smartphone. Another Passive App measure is the app event data. An "event" can represent any activity in the phone app. The app event data capture background (events that happen without the help of any active user interference), interactive (events that are initiated by the user using the app's user interface), and foreground (events that are shown in the app notification section) events. The app event data can be used to understand the user's smartphone activity (ie, what apps they use and how long they use them).

Combining this with the aforementioned active phone use data can be used to understand the smartphone behavior of individuals with ADHD. We calculated how long a participant took to open the phone after receiving a notification from any social and communication apps. We defined social and communication as applications that are labeled as "social" (Instagram, Facebook, etc) or "communication" (WhatsApp, Facebook Messenger, SMS, etc) in the Google Play store.

The smartphone also uses its light sensor to measure ambient light exposure level (ie, illuminance). We used the light sensors to assess the ambient information of participants. The data collection and management were handled by the open-source mHealth platform RADAR-base.

Feature Generation

Using the measures described, we derived the following features that we hypothesize capture ADHD tendencies (Table 2).

XSL•FO
RenderX

Table . List of features for the analyses and their descriptions.

Number	Feature name	Feature description
1	Questionnaire notification response latency	The time taken by the participant to start completing the questionnaire after receiving the notification.
2	SD in questionnaire response latency	SD in the time taken by the participant to start completing the questionnaire after receiving the notification.
3	Mean interval between questionnaires	The mean time interval between finishing one questionnaire and starting the next one.
4	SD in intervals between questionnaires	SD in the time interval between finishing one questionnaire and starting the next one.
5	Daily mean of social and communication apps' notification response latency	Daily mean of the response time to notifications from social and communication apps.
6	Daily SD in social and communication apps' notification response latency	SD in the response time to notifications from the social and communication apps.
7	SD in ambient light	SD in ambient light while actively using the smartphone.
8	Fitbit steps during the active sessions	Fitbit step count during each active phone use session.
9	Active session duration	The time participant spends actively using their smartphone device in each active phone use session.
10	Number of new apps added	The total number of new apps added daily.

Questionnaire Response Time Features

The variable capturing the response latency of the questionnaire notification enables us to assess whether individuals with ADHD are slower to start answering the questionnaires following the notification. The variable mean intervals between questionnaires allow us to explore whether individuals with ADHD have a longer time interval between finishing one questionnaire and starting the next one. The SD in questionnaire response latency and intervals between questionnaires assess the variability in the response time and questionnaire intervals.

Passive Phone Use Features

We further created variables that allow us to assess whether participants with ADHD are more variable and differ overall in their speed of responding to notifications from social and communication apps. The notification data obtained using the Passive app is used to determine how individuals with ADHD respond to notifications from social and communication apps compared with the participants without ADHD.

When a notification from a social or communication app occurs, the phone will record an OTHER event type followed by a FOREGROUND or INTERACTION event if the participant has interacted with the notification. We have excluded all the notifications from the analyses that appeared when the participants were, indicated by their Fitbit data, to be sleeping to remove instances as it is highly unlikely that they could have seen the notifications. We analyzed the time it took for participants to unlock their phones after receiving a notification from a social or communication app.

Specifically, we computed the daily mean and SD of the notification data to derive the participants' daily aggregate of smartphone behaviors.

Ambient Light Features

The variables on the SD in ambient light and steps while active on their smartphones were hypothesized to be related to restlessness by allowing us to assess whether individuals with ADHD move around more than participants without ADHD when they are actively using their devices.

Phone Activity Features

We also created variables measuring the length of each active phone use session and the number of new apps downloaded to assess whether shorter active phone use sessions and a greater number of new apps downloaded are observed in the participants with ADHD.

Statistical Methods

We used linear mixed models (LMMs) to examine differences between the groups for repeatedly measured variables over the study period. These variables include questionnaire notification response latency, mean intervals between questionnaires, the SD in intervals between questionnaires, daily mean of social and communication apps' notification response latency, the daily SD in social and communication apps' notification response, the SD in ambient light, Fitbit steps during the active sessions, active session duration and the number of new apps added.

We chose this approach as it takes the nonindependence of data into account. Therefore, we can use repeated measures from the same participant. LMMs are also highly robust and do not need

XSL•FO
RenderX

to satisfy the prerequirements necessary for ANOVA models [24]. Moreover, we investigated the interaction effect between the days since the recruitment and the control group with respect to questionnaire notification response latency and intervals between questionnaires to evaluate how being in the study for a longer duration impacts these behavioral variables.

For nontemporal aggregated variables (SD in questionnaire notification response latency), we used the standard *t* test to compare the group differences. The *P* values are noted by *P* in the text, and variables are defined as statistically significant if $P<.05$. Along with *P* values, we have also reported the Standard Score (Z value) and effect size (*d*). For repeatedly measured variables where we used LMM, we have computed the effect size in Formula 1 [25]. The effect size of nontemporal aggregated variables is computed using Cohen *d* [26].

Formula 1, measuring effect sizes on the results of LMM [25], is as follows:

$$d = \text{difference between means} / \sqrt{\text{varinterceptpart} + \text{varinterceptitem} + \text{varslopepart} + \text{varslopeitem} + \text{varresidual}}$$

where

d=effect size

varinterceptpart=the intercept coefficient per participant

varinterceptitem=the intercept coefficient per item

varslopepart=the slope coefficient per participant

varslopeitem=the slope coefficient per item

varresidual=the residual variance coefficient

The effect size $d>0.6$ has been employed to determine nonsignificant trends [27]. The cut-off was used to correspond to moderate to high effect sizes. As this is primarily an exploratory study and to inform future larger-scale studies, multiple testing corrections were not undertaken to reduce the chance of type-two errors (ie, false negative results) and to avoid over-correction for multiple comparisons involving multiple correlated variables.

We identified and removed outliers by using the IQR proximity rule, a well-established method for outlier detection in skewed distributions [28]. We calculated the upper limit (UL) and the lower limit (LL), a statistical threshold computed from data points according to Formula 2. Any values in the data above UL and below LL are deemed outliers and removed from the analyses. The derived limits serve as a reference point to identify and manage outliers or unusual data points in a data set. To assess the influence of potential outliers, we conducted a sensitivity analysis. This involved reanalyzing the data after excluding the 2 participants with the highest SD in ambient light values.

Formula 2, the IQR proximity rule [28], is as follows:

data:List of data points

Q1=Percentile(data,25)

Q3=Percentile(data,75)

IQR=Q3−Q1

LL=Q1−1.5×ILR;UL=Q3+1.5×ILR

Results

Group differences were significant for 5 out of the 10 variables: questionnaire notification response latency, the SD in questionnaire notification response latency, the SD in intervals between questionnaires, daily mean of social and communication apps' notification response latency, and SD in ambient light (Table 3). While not significant, the mean intervals between questionnaires, the daily SD in social and communication apps' notification response latency and Fitbit steps during the active session show results from moderate to high effect size, that is, $d>0.60$.

XSL•FO
RenderX

Table . Summary of results for all the 10 derived features.

Number	Feature name	ADHD[a], mean (SD)	Without ADHD, mean (SD)	*P* value	Standard score (Z)	Effect size (*d*)
1	Questionnaire notification response latency (in hours)[b]	19.14 (12.98)	11.69 (10.31)	.047	−1.99	1.05
2	SD in questionnaire notification response latency (in hours)[b]	15.88 (9.27)	8.97 (6.83)	.01	2.68	0.84
3	Mean intervals between questionnaires (in hours)	1.12 (2.17)	0.18 (0.64)	.06	−1.18	0.82
4	SD in intervals between questionnaires (in hours)[b]	2.01 (5.04)	0.35 (2.23)	.04	−2.03	1.13
5	Daily mean of social and communication apps' notification response latency (in seconds)[b]	2304.27 (1406.62)	1516.45 (785.78)	.03	−2.17	0.7
6	Daily SD in social and communication apps' notification response latency (in seconds)	2736.38 (1091.17)	2125.15 (840.86)	.05	−1.96	0.64
7	SD in ambient light (in lux)[b]	83.96 (155.57)	43.63 (73.57)	.008	−2.67	0.86
8	Fitbit steps during the active sessions (in steps)	313.14 (263.88)	179.20 (177.26)	.09	−1.67	−0.61
9	Active session duration (in seconds)	1778.07 (1653.87)	1047.68 (1060.45)	.11	−1.60	0.54
10	Number of new apps added	1.18 (0.40)	1.43 (0.81)	.24	1.18	0.18

[a]ADHD: attention-deficit/hyperactivity disorder.

[b]Statistically significant result.

For questionnaire notification response latency, mean intervals between questionnaires and SD in intervals between questionnaires, we observed the interaction of these features with time and found that the features increased slightly, though non significantly, with each day. We did not find any significant interaction between the control group and the day for the features. The negative coefficient observed corresponding to the interaction effect for the features implies that the increase in features per day is smaller in the comparison group compared with the ADHD group.

Figures 2-6 illustrate the results for questionnaire notification response latency, SD in questionnaire notification response latency, SD in intervals between questionnaires, daily mean of social and communication apps' notification response latency, and SD in ambient light, respectively. Figure 5 shows the SD in ambient light while participants are actively using their smartphones. This reveals that participant IDs 5 and 36, who have ADHD, are potential outliers. The sensitivity analysis confirmed that the group difference remained statistically significant after these exclusions. Therefore, we have chosen to include all participants in the final analysis for transparency and robustness.

XSL•FO
RenderX

Figure 2. The figure presents the differences in the time taken by the participant to start completing the questionnaire after receiving the notification (questionnaire notification response latency) in participants with and without ADHD. ADHD: attention-deficit/hyperactivity disorder.

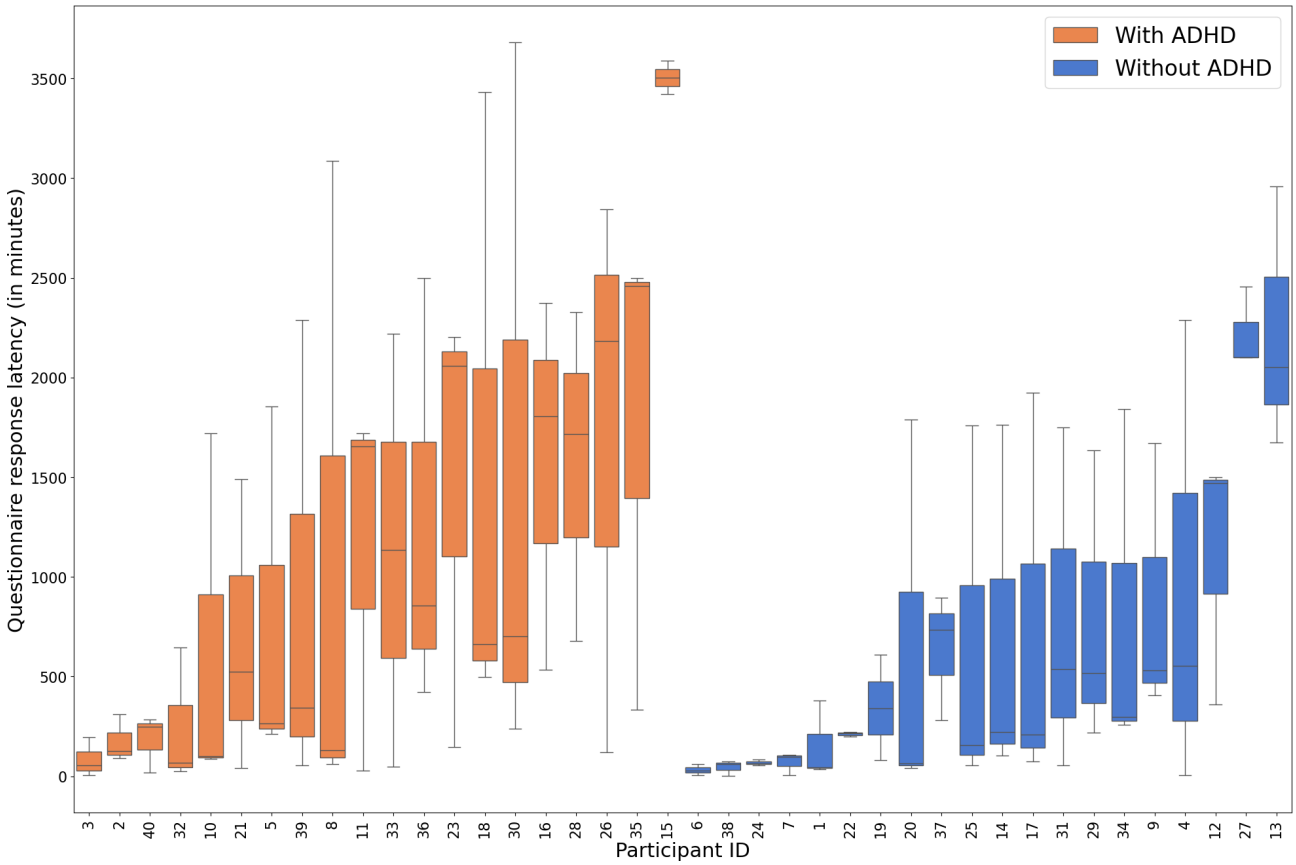
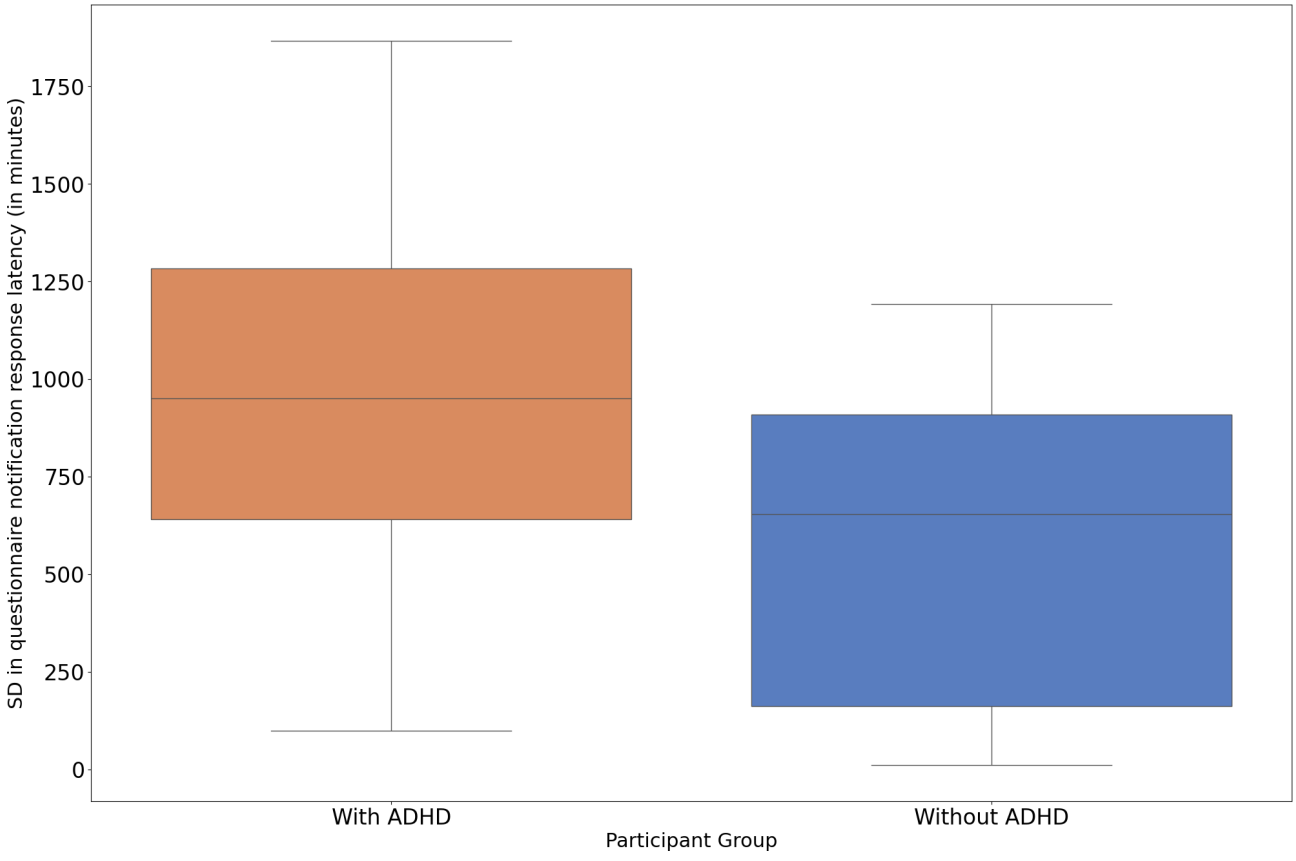


Figure 3. SD in the time taken by the participant to start completing the questionnaire after receiving the notification (SD in questionnaire notification response latency) in individuals with and without ADHD. The boxplot represents the median and IQR of SD in questionnaire notification response latency for both groups. ADHD: attention-deficit/hyperactivity disorder.



XSL•FO
RenderX

Figure 4. Log of SD in the interval between finishing one questionnaire and starting the next one (SD in intervals between questionnaires) in participants with and without ADHD. ADHD: attention-deficit/hyperactivity disorder.

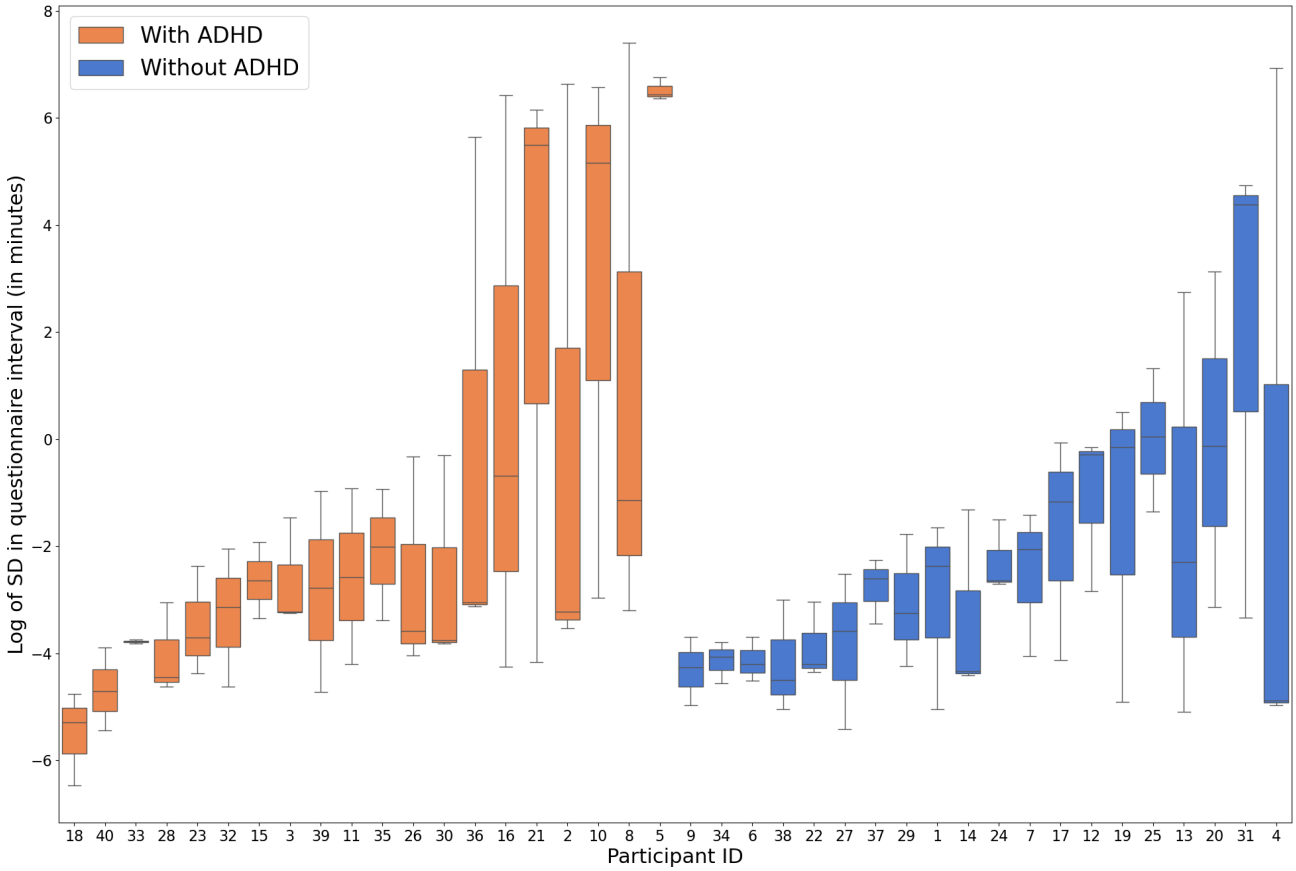
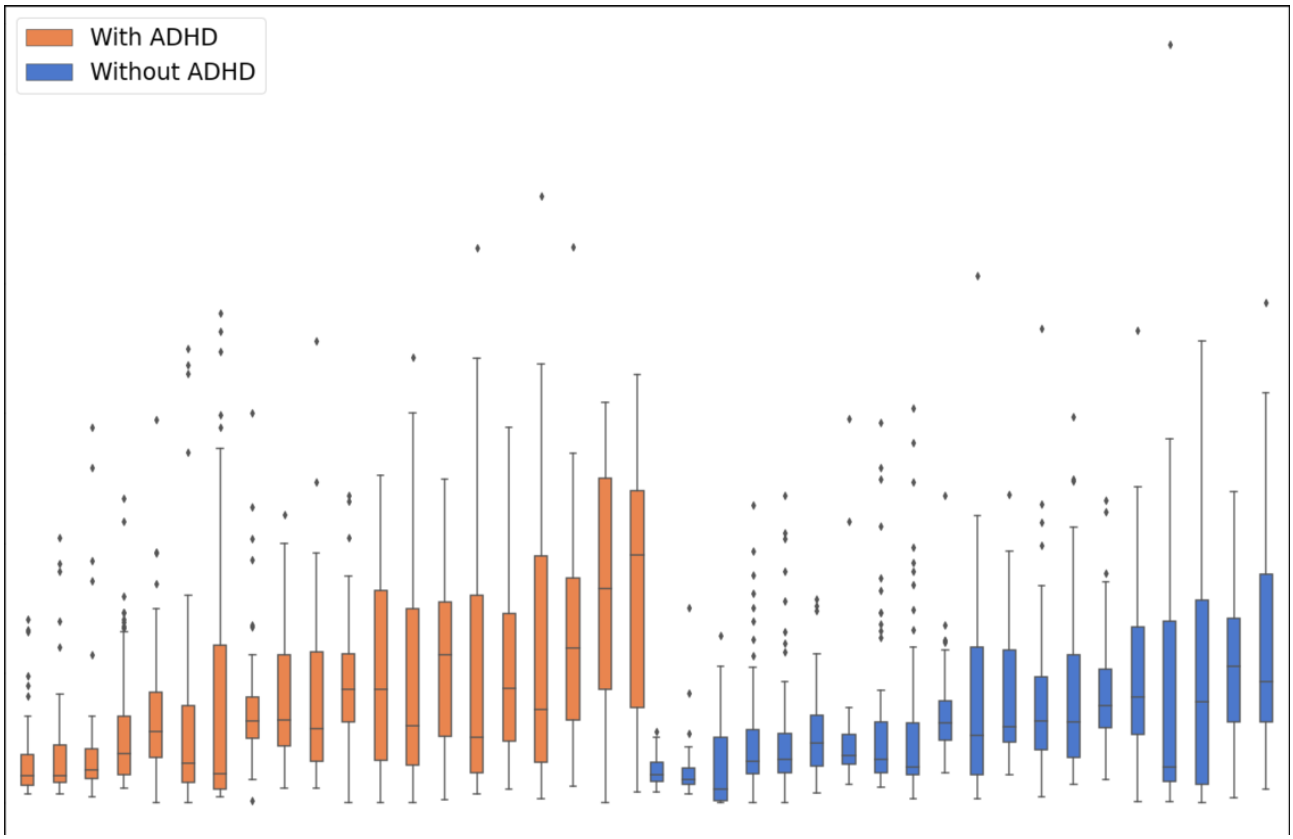
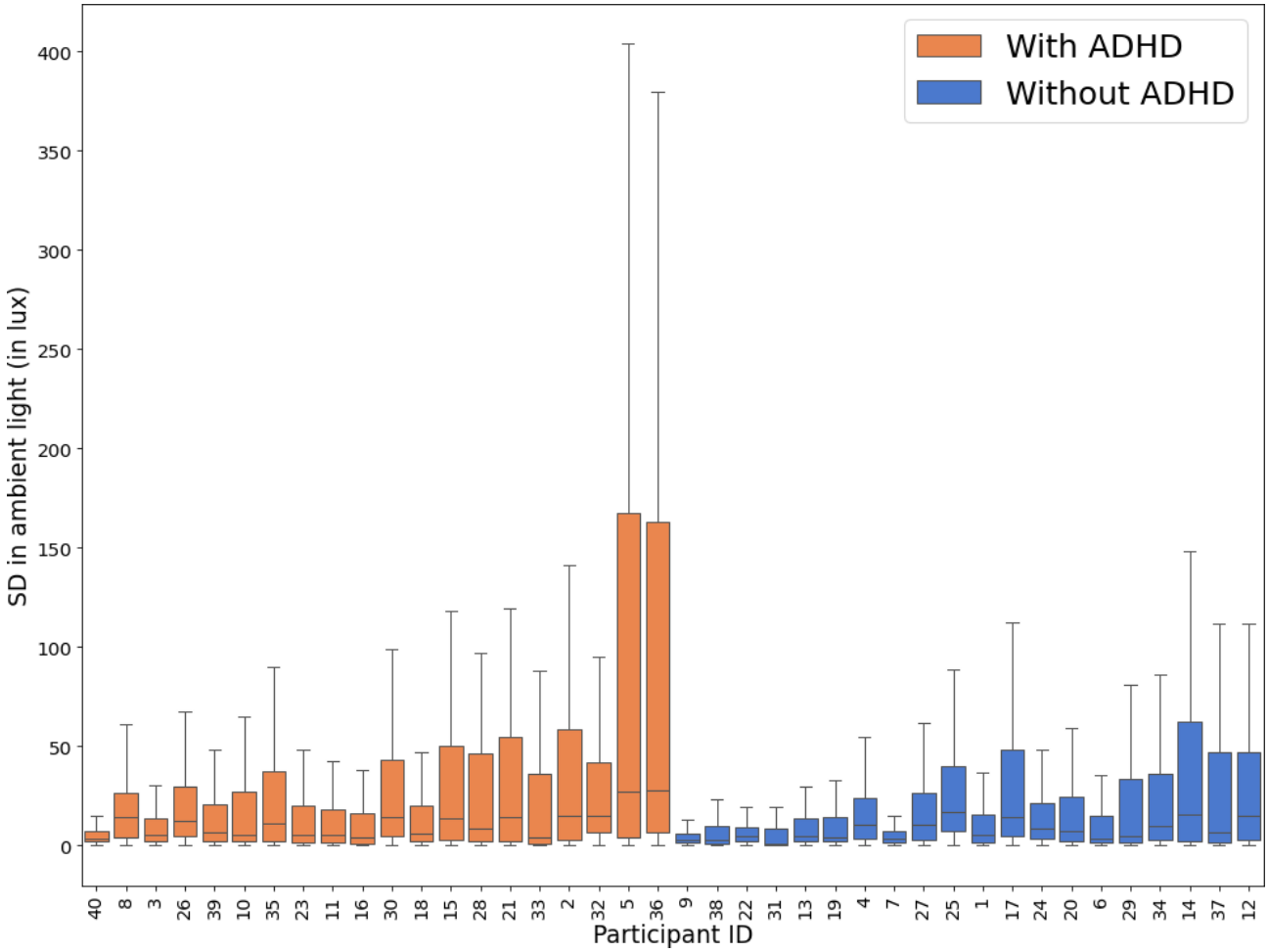


Figure 5. Daily mean of response time to notifications from social and communication apps (daily mean of social and communication apps' notification response latency) in participants with and without ADHD. ADHD: attention-deficit/hyperactivity disorder.



XSL•FO
RenderX

Figure 6. SD in ambient light while actively using the smartphone in participants with and without ADHD. ADHD: attention-deficit/hyperactivity disorder.



Discussion

Principal Findings

In a novel exploration of digital markers of ADHD, we identified candidate digital signals of restlessness, inconsistent attentional focusing, and difficulties completing tasks as part of a 10-week remote monitoring pilot study on our ART system.

Analyses of the data from the smartphone Active App, where participants were requested to complete a series of questionnaires at 4-week intervals, revealed that individuals with ADHD were slower and more variable in starting to complete questionnaires after receiving a notification, and more variable in the time interval between finishing one questionnaire and starting the next one, than the participants without ADHD. The Passive App data further showed that the participants with ADHD were slower to open their smartphones following a notification from a social or communication app compared with the group without ADHD. These novel findings—as well as further potential signals that emerged at trend-level only (greater mean in the ADHD group in the intervals between questionnaires and variability in the time taken to open the phone after receiving a social or communication app notification)—await replication in larger future studies.

The above findings are in line with the frequently reported difficulties in completing tasks among people with ADHD [3] and may further reflect characteristic inconsistent attentional focusing. Future research is needed to explore whether the slower and more variable speed of responding captured on the smartphone data reflects the same difficulties with the regulation of attention and arousal reported in lab-based cognitive studies in individuals with ADHD. Such cognitive studies strongly link ADHD with inconsistency in the speed of responding to cognitive tasks, which has been linked in neurophysiological studies of the regulation of attention and arousal [5,6]. Whether passive remote data collection using smartphones could offer a real-world passive alternative or augmentation to the more intensive cognitive task administration to measure behavioral characteristics associated with ADHD requires careful further study. Given the general availability of smartphones in the general population, the possibility of using this approach for symptom monitoring remains intriguing.

An additional group difference from our analyses of the Passive App data was the SD in ambient (background) light concurrent with the use of their mobile phone, which was significantly higher among the participants with ADHD. Further investigation would be useful to understand this better, but we presently hypothesize it may relate to participants with ADHD moving between differentially lit environments while using their smartphones. A related group comparison on Fitbit step count while active on their smartphone was not significant, but the high effect size observed again emphasizes the need for future

XSL•FO
RenderX

research with increased statistical power. We did not observe a difference between participants with and without ADHD in the number of new apps downloaded or in the length of active smartphone use sessions, which suggests that these markers do not capture impulsivity, restlessness, or other symptoms of ADHD.

Limitations

Since this was a pilot study, the sample size was small (n=40), and the findings require replication in future studies with larger samples. While this sample size provided sufficient power to explore initial relationships between ADHD and phone-based behavioral markers, it limits the generalizability of the findings to the broader ADHD population. Smaller sample sizes can increase the likelihood of chance findings and decrease the ability to detect subtle but meaningful effects. Therefore, replicating these findings in larger and more diverse samples is necessary to confirm their validity and generalizability. A strength of our design was, however, the large amount of data collected per participant: 10 weeks' worth of multimeasure remote monitoring data, which, in the case of passive monitoring data, was collected 24 hours a day. During the 10 weeks of the study period, we collected an average of 99,180 Fitbit step data points per participant. We collected an average of 149,821 different app activities per participant from 1855 unique apps during the same period.

Another limitation is that the data were collected during the COVID pandemic when restrictions were in place that may have affected daily activities (eg, physical activity, sleep, and smartphone activity) and well-being more generally. Importantly for our analyses, however, we note that such restrictions were concurrent for both groups.

Another limitation to consider is that 75% of our participants are women, and hence, there is a gender imbalance in our study. However, the groups were matched on gender. Future larger studies should aim to recruit samples with more representative gender distributions.

Furthermore, we provided several participants (n=35) who owned an Apple smartphone or a noncompatible Android smartphone with a new compatible Android smartphone. Despite instructing participants to use the new Android phone as their primary phone during their remote monitoring period, as participants had access to their Apple smartphone and noncompatible smartphone, it is difficult to assess how much they may have used their nonstudy phones.

Future Research

We are currently conducting a large remote monitoring study on adults with ADHD—the ADHD remote technology study of cardiometabolic risk factors and medication adherence (ART-CARMA) [29]. In ART-CARMA, we obtain remote monitoring data from the participants with ADHD both before and after initiation of ADHD medication treatment, which will allow us to examine the effects of medication on the digital markers of ADHD that we identified here. In total, we will obtain remote monitoring data from each participant for 12 months, enabling detailed over-time analyses of the multiple measures. Another direction for future research would be to explore potential age- and gender-related differences in digital markers, from childhood through adolescence and young adulthood, to older adulthood, in both men and women.

Conclusions

Using the ART system in a 10-week remote monitoring study of participants with and without ADHD, we identified novel candidate digital markers of ADHD: digital signals that we propose to capture ADHD symptoms of restlessness, inconsistent attentional focusing, and difficulties completing tasks. Future research can assess the potential for the candidate digital markers of ADHD, as identified here, to track ADHD severity or predict outcomes.

Acknowledgments

We thank all who made this research possible: all participants who contributed their time to the study and the funders. The

attention-deficit/hyperactivity disorder (ADHD) remote technology (ART) pilot study was funded by a King's Together Strategic Grant. This research was funded in part by the National Institute for Health Research (NIHR), University College London Hospitals Biomedical Research Centre, Maudsley Biomedical Research Centre at South London and Maudsley NHS (National Health Service) Foundation Trust, and King's College London. HD is supported by the UK Medical Research Council (MR/N013700/1) and King's College London member of the MRC Doctoral Training Partnership in Biomedical Sciences. QD is supported by the China Scholarship Council (201908440362). Generative artificial intelligence tools like ChatGPT were not used in writing this manuscript. However, Grammarly Pro was used for editing, grammar, and formatting purposes.

Data Availability

For open access, the author has applied a Creative Commons Attribution (CC BY) license to any author-accepted manuscript version. The data supporting this study are not openly available but may be shared on request following a formal review process by the chief investigator and coinvestigators.

Conflicts of Interest

The authors disclosed the following potential conflicts of interest related to the research, authorship, and publication of this article: JK has delivered talks at educational events sponsored by Medice, with all funds directed to King's College London to support

XSL•FO
RenderX

attention-deficit/hyperactivity disorder (ADHD) research. PA has received research funding from Vifor Pharma, delivered sponsored talks, and served as an advisor for Shire, Janssen-Cilag, Eli Lilly, Flynn Pharma, and Pfizer on the diagnosis and treatment of ADHD; all funds were received by King's College London and allocated to ADHD studies. The remaining authors have reported no conflicts of interest.

References

1. Faraone SV, Banaschewski T, Coghill D, et al. The World Federation of ADHD international consensus statement: 208 evidence-based conclusions about the disorder. Neurosci Biobehav Rev 2021 Sep;128:789-818. [doi: 10.1016/j.neubiorev.2021.01.022] [Medline: 33549739]
2. Kuntsi J, Wood AC, Rijsdijk F, et al. Separation of cognitive impairments in attention-deficit/hyperactivity disorder into 2 familial factors. Arch Gen Psychiatry 2010 Nov;67(11):1159-1167. [doi: 10.1001/archgenpsychiatry.2010.139] [Medline: 21041617]
3. Franke B, Michelini G, Asherson P, et al. Live fast, die young? A review on the developmental trajectories of ADHD across the lifespan. Eur Neuropsychopharmacol 2018 Oct;28(10):1059-1088. [doi: 10.1016/j.euroneuro.2018.08.001]
4. Sun S, Denyer H, Sankesara H, et al. Remote administration of ADHD-sensitive cognitive tasks: a pilot study. J Atten Disord 2023 Jul;27(9):1040-1050. [doi: 10.1177/10870547231172763] [Medline: 37269091]
5. James SN, Cheung CHM, Rijsdijk F, Asherson P, Kuntsi J. Modifiable arousal in attention-deficit/hyperactivity disorder and its etiological association with fluctuating reaction times. Biol Psychiatry Cogn Neurosci Neuroimaging 2016 Nov;1(6):539-547. [doi: 10.1016/j.bpsc.2016.06.003] [Medline: 27840854]
6. Cheung CHM, McLoughlin G, Brandeis D, Banaschewski T, Asherson P, Kuntsi J. Neurophysiological correlates of attentional fluctuation in attention-deficit/hyperactivity disorder. Brain Topogr 2017 May;30(3):320-332. [doi: 10.1007/s10548-017-0554-2] [Medline: 28289850]
7. Girolamo G, Barattieri di San Pietro C, Bulgari V, et al. The acceptability of real‐time health monitoring among community participants with depression: a systematic review and meta‐analysis of the literature. Depress Anx 2020 Sep;37(9):885-897. [doi: 10.1002/da.23023]
8. Ranjan Y, Rashid Z, Stewart C, et al. RADAR-base: open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, and mobile devices. JMIR Mhealth Uhealth 2018;7(8):e11734. [doi: 10.2196/11734] [Medline: 29674305]
9. Pratap A, Atkins DC, Renn BN, et al. The accuracy of passive phone sensors in predicting daily mood. Depress Anxiety 2019 Jan;36(1):72-81. [doi: 10.1002/da.22822] [Medline: 30129691]
10. Dalla Costa G, Leocani L, Montalban X, et al. Real-time assessment of COVID-19 prevalence among multiple sclerosis patients: a multicenter European study. Neurol Sci 2020 Jul;41(7):1647-1650. [doi: 10.1007/s10072-020-04519-x] [Medline: 32617741]
11. Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela JP. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. Neuropsychopharmacology 2018 Jul;43(8):1660-1666. [doi: 10.1038/s41386-018-0030-z] [Medline: 29511333]
12. Saeb S, Zhang M, Karr CJ, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015 Jul 15;17(7):e175. [doi: 10.2196/jmir.4273] [Medline: 26180009]
13. Ranjan Y, Rashid Z, Stewart C, et al. RADAR-base: open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, and mobile devices. JMIR Mhealth Uhealth 2019 Aug 1;7(8):e11734. [doi: 10.2196/11734] [Medline: 31373275]
14. Matcham F, Leightley D, Siddi S, et al. Remote assessment of disease and relapse in major depressive disorder (RADAR-MDD): recruitment, retention, and data availability in a longitudinal remote measurement study. Eur Psychiatr 2022 Jun;65(S1):S112-S112. [doi: 10.1192/j.eurpsy.2022.315]
15. Zhang Y, Folarin AA, Sun S, et al. Relationship between major depression symptom severity and sleep collected using a wristband wearable device: multicenter longitudinal observational study. JMIR Mhealth Uhealth 2021 Apr 12;9(4):e24604. [doi: 10.2196/24604] [Medline: 33843591]
16. Zhang Y, Folarin AA, Sun S, et al. Predicting depressive symptom severity through individuals' nearby bluetooth device count data collected by mobile phones: preliminary longitudinal study. JMIR Mhealth Uhealth 2021 Jul 30;9(7):e29840. [doi: 10.2196/29840] [Medline: 34328441]
17. Kooij J, Francken MH. DIVA 2.0 diagnostic interview for ADHD in adults. : DIVA Foundation; 2010 URL: www.advancedassessments.co.uk/resources/ADHD-Screening-Test-Adult.pdf
18. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
19. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]

XSL•FO
RenderX

20. Stringaris A, Goodman R, Ferdinando S, et al. The affective reactivity index: a concise irritability scale for clinical and research settings. J Child Psychol Psychiatry 2012 Nov;53(11):1109-1117. [doi: 10.1111/j.1469-7610.2012.02561.x] [Medline: 22574736]
21. Barkley RA, Murphy KR. Attention-Deficit Hyperactivity Disorder: A Clinical Workbook: Guilford Press; 2006.
22. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder. Arch Intern Med 2006 May 22;166(10):1092. [doi: 10.1001/archinte.166.10.1092]
23. Raine A, Dodge K, Loeber R, et al. The reactive-proactive aggression questionnaire: differential correlates of reactive and proactive aggression in adolescent boys. Aggress Behav 2006 Apr 1;32(2):159-171. [doi: 10.1002/ab.20115] [Medline: 20798781]
24. Schielzeth H, Dingemanse NJ, Nakagawa S, et al. Robustness of linear mixed ‐ effects models to violations of distributional assumptions. Methods Ecol Evol 2020 Sep;11(9):1141-1152. [doi: 10.1111/2041-210X.13434]
25. Brysbaert M, Stevens M. Power analysis and effect size in mixed effects models: a tutorial. J Cogn 2018 Jan 12;1(1):9. [doi: 10.5334/joc.10] [Medline: 31517183]
26. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863. [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
27. Nead KT, Wehner MR, Mitra N. The use of “Trend” statements to describe statistically nonsignificant results in the oncology literature. JAMA Oncol 2018 Dec 1;4(12):1778-1779. [doi: 10.1001/jamaoncol.2018.4524] [Medline: 30347042]
28. Vinutha HP, Poornima B, Sagar BM. Detection of Outliers Using Interquartile Range Technique from Intrusion Dataset Information and Decision Sciences: Springer; 2018:511-518. [doi: 10.1007/978-981-10-7563-6_53]
29. Denyer H, Ramos-Quiroga JA, Folarin A, et al. ADHD remote technology study of cardiometabolic risk factors and medication adherence (ART-CARMA): a multi-centre prospective cohort study protocol. BMC Psychiatry 2022 Dec 20;22(1):813. [doi: 10.1186/s12888-022-04429-6] [Medline: 36539756]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ART: attention-deficit/hyperactivity disorder remote technology
ART-CARMA: attention-deficit/hyperactivity disorder remote technology study of cardiometabolic risk factors and medication adherence
BAARS: Barkley Adult Attention-Deficit/Hyperactivity Disorder Rating Scale
BFIS: Barkley Functional Impairment Scale
LL: lower limit
LMM: linear mixed model
mHealth: mobile health
RADAR: Remote Assessment of Disease and Relapses
REDCap: Research Electronic Data Capture
UL: upper limit

Edited by A Mavragani; submitted 14.11.23; peer-reviewed by CI Campbell, HA Golkashani, TA Reddy Sure; revised version received 24.10.24; accepted 29.10.24; published 29.01.25.

Please cite as:
Sankesara H, Denyer H, Sun S, Deng Q, Ranjan Y, Conde P, Rashid Z, Asherson P, Bilbow A, Groom MJ, Hollis C, Dobson RJB, Folarin A, Kuntsi J
Identifying Digital Markers of Attention-Deficit/Hyperactivity Disorder (ADHD) in a Remote Monitoring Setting: Prospective Observational Study
JMIR Form Res 2025;9:e54531
URL: https://formative.jmir.org/2025/1/e54531
doi:10.2196/54531

© Heet Sankesara, Hayley Denyer, Shaoxiong Sun, Qigang Deng, Yatharth Ranjan, Pauline Conde, Zulqarnain Rashid, Philip Asherson, Andrea Bilbow, Madeleine J Groom, Chris Hollis, Richard J B Dobson, Amos Folarin, Jonna Kuntsi. Originally published in JMIR Formative Research (https://formative.jmir.org), 29.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Development of an Instagram Reel-Based Bystander Intervention Message Among College Students: Formative Survey and Mixed Methods Pilot Study

Leticia Couto, PhD

College of Communication, DePaul University, Daley Bldg, 14 E Jackson Blvd, Chicago, IL, United States

Corresponding Author:
Leticia Couto, PhD
College of Communication, DePaul University, Daley Bldg, 14 E Jackson Blvd, Chicago, IL, United States

Abstract

Background: Bystander intervention is a common method to address the ubiquitous issue that is sexual violence across college campuses. Short messages that incentivize bystander intervention behavior can be another tool to fight sexual violence.

Objective: This study aimed to conduct formative research surrounding social norms and bystander barriers to pilot and develop Instagram (Meta) reel-based messages addressing bystander intervention among college students.

Methods: The first step was to conduct a formative survey to identify peer norms and actual behavior of the intended population. Once that data were collected, a mixed methods message pilot was conducted by a survey where participants randomly saw 5 of the 12 messages developed, assessing them for credibility, perceived message effect, and intended audience.

Results: The formative survey was conducted among 195 college students from the same institution, and the pilot test was conducted among 107 college students. The formative survey indicated a discrepancy between perceived peer behavior and actual behavior of the participants in all 3 measures, allowing for the development of normative messaging. The pilot testing indicated the credibility was acceptable (eg, mean 3.94, SD 1.15 on a 5-point scale) as well as the perceived message effect (eg, mean 4.26, SD 0.94 on a 5-point scale). Intended audiences were also identified and reached. Qualitative results indicated that the messages may have lacked credibility, although the quantitative results suggest otherwise.

Conclusions: Participants understood the messages concerned bystander intervention, and perceived message effects results indicated the messages to be effective in assisting bystander intervention engagement by normative messaging. Messages were considered credible and reached the intended audience. The qualitative results provided further insights on how the messages can be adapted before being tested for effects. Future research should focus on further adapting the messages and testing their effects among the studied population.

(JMIR Form Res 2025;9:e66769) doi:10.2196/66769

KEYWORDS

bystander intervention; message development; sexual health; college; student; sexual violence; bystander; reel-based; Instagram; social media; short message; formative research; mixed methods; social norms; perceived behavior; qualitative; behavioral health; digital health

Introduction

Background

Bystander intervention trainings and messages are two of the main ways universities in the United States attempt to address the sexual violence crisis in college campuses [1,2]. Although it is common for bystander intervention messaging (here, *bystander intervention* is defined as the ability to identify situations in which a third party may assist someone in a risky and act to intervene [3]) to use bystander models as a theoretical framework [4,5], the social norms approach is also used to address sexual violence interventions [6]. For the social norms approach to be successful, it is necessary that the population of the message experiences pluralistic ignorance, where there is a difference between their own behavior and their estimate of peer behavior [6]. The social norms approach suggests that people who believe that their peers are engaging in a behavior are more likely to engage in that behavior themselves [6], whereas bystander models, in particular bystander barriers, suggest that the belief that others will take responsibility removes their own responsibility to engage in bystander intervention [7]. Many trainings and programs exist that focus on stimulating bystander intervention behavior [8,9], but less is known about the role of supporting media messages. In addition, it is necessary to investigate how these 2 opposing perspectives may be perceived as messaging mechanisms for college students.

XSL•FO
RenderX

Sexual Violence and Bystander Intervention

There is a sexual assault crisis on college campuses. Estimates indicate that approximately 1 in 4 college women are sexually assaulted during their time in college [1], which is high compared with the rate of 1 in 5 among the general population of women in the United States [2]. In addition, certain populations seem to be more at risk, such as first-year students [10] and LGBTQ+ (lesbian, gay, bisexual, transgender, queer+) students [1]. Universities across the country have tried to curb the problem by using a series of prevention efforts against sexual violence, with one of the most popular ones being bystander intervention [11].

Bystander intervention training and campaigns, including in-person and online training, are effective ways of increasing intervention intentions among college students in multiple scenarios [5,12]. An example of these campaigns is Green Dot, which features a focus on motivational speech and in-person training [12]. Over 500 universities in the United States have adopted Green Dot [13]. It is worth noting, however, that the bystander effect has been brought into question [14,15]. This can be in part due to its questionable origins in the Kitty Genovese issue [16]. Even still, this event led to an eventual refinement of what it means to intervene as a bystander, and the trainings developed from there proved to be successful [4,5,12,17,18].

Instagram as a Channel for Interventions

Using Instagram as a platform for public health concerns has been studied mostly in an exploratory manner [19]. However, some studies suggest that the issue is not necessarily reach but the ability of the entities governing these accounts to communicate with their publics [20]. In addition, such messaging developed for Instagram can be easily adapted for other channels, facilitating eventual scalability efforts.

Instagram reels, in particular, are more popular than regular Instagram posts and stories because they have better reach and can be found more easily than regular posts [21]. They also have more editing features that can be used to make messages engaging for the audience [21]. Instagram is also highly used among college-aged populations, more so than other platforms [22], such as Twitter (now X), Facebook, and Snapchat. Thus, the Instagram reel-inspired format can be an optimal channel for short and targeted messaging.

This Study

In this study, a formative survey was conducted to assess if pluralistic ignorance exists in a college population in relation to bystander intervention behavior in sexual assault situations. This step needed to be established first to ensure a social norms–based message would be feasible. In the presence of pluralistic ignorance, 3 theoretically based Instagram reel-based video messages with short scripts messages were developed and piloted in terms of credibility, intended audience, perceived message effectiveness, and a manipulation check of the conditions. One condition on bystander behavior barriers, focusing on showing participants only the actual behavioral data reported in the survey, was called the message readjustment condition in this study; another condition was based on the social norms approach, which was called the norm reinforcement condition in this study, as it focused on showing participants only the normative information from the data collected in the survey; and 1 combined both theoretical frameworks, showing both the normative information and the behavioral information with the intent of highlighting the discrepancy between the 2. The goal of this study is to assess the perceived message effectiveness, credibility, and intended audience reach of these messages pertaining to bystander intervention.

Methods

Formative Survey

In the first semester of 2023 in a large, public, American university, data were collected from 195 undergraduate students aged 18 to 29 years (mean 20.52, SD 2.22 years), most of whom identified as women (118/195, 60.5%), cisgender (184/195, 94.4%), heterosexual (123/195, 63.1%), White (139/195, 71.6%), not Hispanic (110/195, 56.7%), and middle class (124/195, 63.6%). Participants were recruited from a randomized list of 2000 emails from the university's registrar's office, representative of the university as whole.

Formative Research Measurements

Bystander Intervention Behavior

Three items were adapted from Hust et al [23]. Participants were asked to indicate how much they agree (5) or disagree (1) with statements related to perceived norms. Specifically, these items concerned the participants own behavior. These items included "I have made sure my friends are ok if I see him/her in an uncomfortable sexual situation at a party" and "I have discouraged a friend who said they planned to get someone drunk to have sex." The Cronbach α was assessed for the entire sample, for a value of .83.

Bystander Intervention Perceived Norms

Three items were adapted from Hust et al [23]. Participants were asked to indicate how much they agree (5) or disagree (1) with statements related to perceived norms. Specifically, these items refer to how they perceived their peers would act in these situations. Items include the stem "most of my friends would," followed by items such as "make sure their friend is ok if they see him/her in an uncomfortable sexual situation at a party" and "discourage a friend who said they planned to get someone drunk to have sex." The Cronbach α was assessed for the entire sample, for a value of .8.

Formative Research Results

Table 1 demonstrates that there was a discrepancy between peer perception and actual behavior of participants. The means were higher among the normative statements (bottom 3) as compared with their actual behavior items counterparts, indicating that participants said they more strongly agreed with the normative statements. This indicates that there was an opportunity for a social norms message to be developed, as pluralistic ignorance was identified.

XSL•FO
RenderX

Table . Means and SDs of items used for message development.

Items	Mean (SD)	n/N[a] (%)
I have made sure my friends are ok if I see him/her in an uncomfortable sexual situation at a party.	4.43 (1.08)	163/195 (83.6)
I have warned someone if I saw a drug being slipped into their drink.	3.63 (1.41)	93/192 (47.7)
I have discouraged a friend who said they planned to get someone drunk to have sex.	3.73 (1.43)	107/193 (55.5)
Most of my friends would make sure their friend is ok if they see him/her in an uncomfortable sexual situation at a party.	4.86 (0.41)	191/195 (98)
Most of my friends would warn someone if they saw a drug being slipped into their drink.	4.89 (0.38)	191/195 (98)
Most of my friends would discourage a friend who said they planned to get someone drunk to have sex.	4.82 (0.47)	188/195 (96.4)

[a]Percentage that indicated either “agree” (5) or “somewhat agree” (4) for each of these statements.

Pilot Testing

Messages were developed based on the items tested in the formative research survey step. A total of 12 messages were developed, 3 per condition. The experimental conditions and the theoretical basis for each can be found on Table 2.

Table . Experiment conditions based on theoretical frameworks.

Condition	Theoretical framework	Outcomes
Norm readjustment	Bystander barriers [7]	Intention to take responsibility
Norm reinforcement	Social norms approach [6]	Perceived peer norm
Norm readjustment and reinforcement	Social norms approach and bystander barriers [6,7]	Intentions
Control	—[a]	—

[a]Not applicable.

Data were collected in October of 2023. This data collection was exempt from institutional review board (IRB) review for similar reasons as the formative survey, where the university’s IRB reviewed the proposal and indicated that a full-board review was not necessary considering the minimal risks. To be eligible to participate, participants had to be currently in the United States, be an undergraduate student at the university being studied, and be between the ages of 18 and 29 years. Participants were recruited through a pool of students taking communication courses and were compensated with extra credit. Other recruitment methods were attempted, such as social media recruitment, but they yielded very few participants. These participants were compensated with Starbucks gift cards. These 3 conditions were added as screeners to ensure only participants who conformed to them participated in the pilot. This type of pilot testing strengthens the validity and reliability of the intervention, and it has been used in similar efforts [24,25].

Participants saw 5 of 12 Instagram reel-based short video messages developed in random order (3 per condition: norm readjustment, norm reinforcement, combined condition, and control). The following instructions were provided to the participants before they started the pilot: “Next, you will see 5 short videos. You will answer some questions about those videos and then some overall questions. Please make sure you can hear the videos as you watch them. Please click on the button on the bottom right to continue.” Participants were also provided with the following description of bystander intervention: “Definition of bystander intervention: Bystander Intervention is recognizing a potentially risky situation or interaction and choosing to respond in a way that could positively affect the outcome. For the purposes of this study, we are mostly referring to intervening in situations that could escalate to sexual violence.” As these messages had to be highly tailored, the messages referred to the students with the name of the university’s mascot (eg, if the university’s mascot was a tiger, the participants were called tigers). These references were removed from this paper to ensure the anonymity of the participants and an unbiased peer review process.

A total of 107 individuals participated in the pilot, but 81 participants responded to demographics items. Participants’ ages ranged from 18 to 24 years (mean 20.16, SD 1.33 years). Most identified as women (52/81, 64.2%) and cisgender (80/81, 98.8%). Most identified as heterosexual (60/81, 74.1%), but a considerable portion of the sample either chose a sexual minority (15/81, 18.4%) or chose not to answer (6/81, 7.4%). Most participants were non-Hispanic/Latine (44/81, 54.3%) and were

XSL•FO
RenderX

Caucasian/White (61/81, 75.3%). Most participants also self-identified as middle class (58, 71.6%).

Pilot Measures

Perceived Message Effectiveness

An adaptation of the 3 items from the University of North Carolina's Perceived Message Effectiveness Scale (UNC PMES) was used to measure perceived message effectiveness [26]. Example items include "This message makes me concerned about the lack of bystander intervention engagement on campus" and "This message discourages me from not engaging in bystander intervention." Participants were asked to indicate how much they agree (5) or disagree (1) with these statements. The Cronbach α for the messages varied between .68-.92. The average α was .87.

Credibility

To measure credibility, participants were asked how much they agreed (5) or disagreed (1) that the message was trustworthy, credible, believable, and realistic. The 4 items were combined to form a credibility scale. This scale was self-developed based on previous literature exploring credibility of messages [27]. The Cronbach α range for the credibility composite scores was .89-.95 with an average of .92.

Intended Audience for the Message

Participants were asked to indicate how much they disagree (1) or agree (5) that the following groups are the intended audience of the message: college students, people between 18 - 29 years old, and [university] students. The goal of this item was to assess whether the message resonates with its intended audience (ie, college students, [university] students, people aged between 18 - 29 years). These items were self-developed to address the specific intended audiences of these messages.

Manipulation Checks

A total of 6 manipulation checks were developed. The participants were asked to indicate how much they agreed (5) or disagreed (1) with the statements that had the stem "the message I saw," followed by "encouraged me to be an active bystander," "encouraged me to be part of the group that helps" (items that distinguished control messages from the other messages), "suggests very few [university mascot] help others at parties" (refers to readjustment condition messages and combination condition messages), "suggests a lot of [university mascot] help others at parties" (refers to reinforcement condition), "suggests [university mascot] trust their friends" (refers to reinforcement condition and combination condition messages), and "contains tips on how to party safely" (refers to the control messages). These were evaluated based on means and how responses compared to one another.

Open-Ended Items

Participants were asked to indicate what they liked and disliked about the messages, as well as to identify what they believed to be the purpose of the message.

Initial Message Design

The messages were designed on Canva, using available music and visuals. The script was created based on the bystander items described in this paper. The scripts are approximately 10 - 15 words from each other in terms of length. Please refer to Textbox 1 for a complete breakdown of each initial message, the condition that it belongs to, and the number by which the message is referred to in the *Results* section. Figure 1 shows an example of the message.

XSL•FO
RenderX

Textbox 1. Initial messages and the conditions they belong to (message numbers are those referred to in the Results section).

Norm readjustment

Message 1:

BE AN ACTIVE BYSTANDER

Only 48% of [UNIVERSITY MASCOT] have told someone that their drink had been drugged

IT'S ON YOU!

Be part of the group that helps!

Message 2:

BE AN ACTIVE BYSTANDER

Only 84% of [UNIVERSITY MASCOT] have made sure their friend was ok if they seem to be in an uncomfortable situation at a party

IT'S ON YOU!

Be part of the group that helps!

Message 3:

BE AN ACTIVE BYSTANDER

Only 55% of [UNIVERSITY MASCOT] have discouraged a friend from getting someone drunk to have sex with them

IT'S ON YOU!

Be part of the group that helps!

Norm reinforcement

Message 4:

BE AN ACTIVE BYSTANDER

More than 97% of [UNIVERSITY MASCOT] believe their friends would tell someone that their drink had been drugged

IT'S ON YOU!

Be part of the group that helps!

Message 5:

BE AN ACTIVE BYSTANDER

More than 97% of [UNIVERSITY MASCOT] believe that their friends would make sure another friend was ok if they seem to be in an uncomfortable situation at a party

IT'S ON YOU!

Be part of the group that helps!

Message 6:

BE AN ACTIVE BYSTANDER

More than 95% of [UNIVERSITY MASCOT] believe that their friends would discourage a friend from getting someone drunk to have sex with them

IT'S ON YOU!

Be part of the group that helps!

Norm readjustment and reinforcement

Message 7:

BE AN ACTIVE BYSTANDER

More than 97% of [UNIVERSITY MASCOT] believe their friends would tell someone that their drink had been drugged, but only 48% of [UNIVERSITY MASCOT] have done so

IT'S ON YOU!

Be part of the group that helps!

Message 8:

BE AN ACTIVE BYSTANDER

More than 97% of [UNIVERSITY MASCOT] believe that their friends would make sure another friend is ok if they seem to be in an uncomfortable situation at a party, but only 84% of [UNIVERSITY MASCOT] have done so

XSL•FO
RenderX

IT’S ON YOU!

Be part of the group that helps!

Message 9:

BE AN ACTIVE BYSTANDER

More than 95% of [UNIVERSITY MASCOT] believe that their friends would discourage a friend from getting someone drunk to have sex with them, but only 55% of [UNIVERSITY MASCOT] have done so

IT’S ON YOU!

Be part of the group that helps!

Control

Message 10:

BE SAFE AT YOUR NEXT PARTY!

At parties, make sure you know where the nearest exit is in case of an emergency. You never know when you might have to bounce!

PLAN AHEAD!

Safe partying is more fun!

Message 11:

BE SAFE AT YOUR NEXT PARTY!

At parties, make sure you know where you place important belongings like wallet and keys. You never know when you may need to bounce!

PLAN AHEAD!

Safe partying is more fun!

Message 12:

BE SAFE AT YOUR NEXT PARTY!

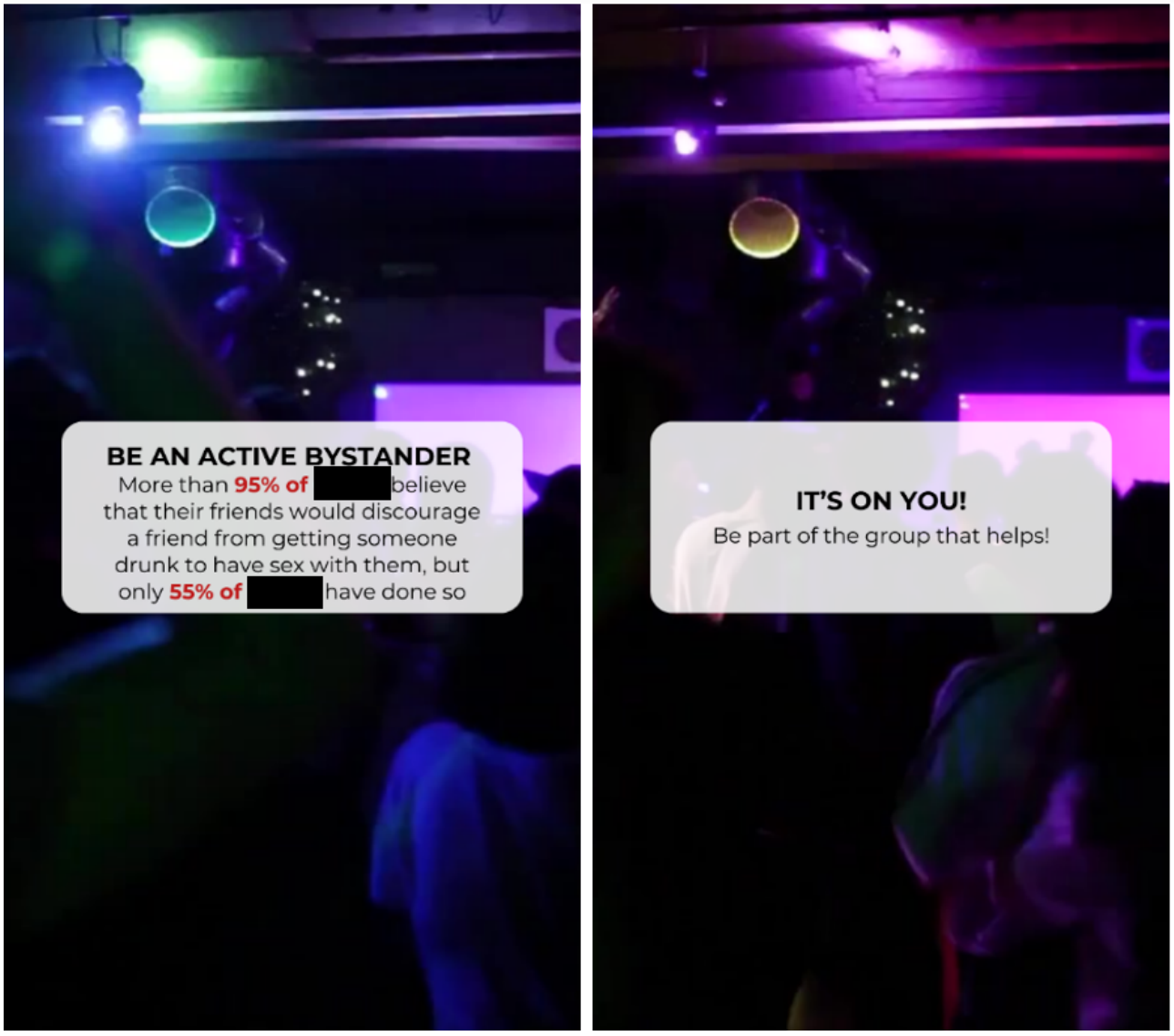
At parties, make sure you know how you are getting back home. You never know when you may need to bounce!

PLAN AHEAD!

Safe partying is more fun!

XSL•FO
RenderX

Figure 1. Piloted message example.



Qualitative Analysis

To evaluate the qualitative open-ended questions, thematic analysis was used [28,29]. Thematic analysis was chosen instead of other potential qualitative methods to allow for a more organic identification of themes and discussion among coders. The coders were the author and a colleague in the same discipline as the author. The first step in Braun and Clarke's [28,29] thematic analysis is to become familiarized with the data. The coders read all the responses before starting to code them. After that, the responses were prepped in a spreadsheet format. The coders highlighted cells based on how they may represent a theme. If a segment may be referring to 2 or more themes, coders were to add a column next to the column where the segments were and highlight the cell next to the original segment. This procedure repeated itself for each question (likes, dislikes, and purpose) and per message (9 messages total). Although 12 messages were tested, the qualitative coding only happened for the messages that were related to bystander intervention and not the control messages (messages 10 - 12). Thus, there is a total of 24 units to analyze, where each has between 28 - 37 segments as responses. A total of 930 segments were coded. Then, we used 1 set of messages to train, where we coded the qualitative responses related to message 1 while we were together, but the coding happened separately. During the training, we established the scope of the themes, and we established that for the priming question, we were coding for 2 themes: participants understood the message, and participants did not understand the message. After coding the segments associated with message 1, we found that we agreed in most instances on the development of themes and how we approached the priming segments. We used the segments associated with messages 1 - 4 to establish the most common themes and used those themes to code the rest of the segments in messages 5 - 9. Thus, approximately 44% of the data were used to establish the themes, and those themes were reinforced using the rest of the data. We then moved on to code the segments associated with messages 2 - 9 by ourselves, and compared the coding of the coder and my coding to assess if there was a need for us to meet again and discuss any major disagreements. Minor disagreements were addressed by email, phone calls, and text messages.

XSL•FO
RenderX

Statistical Analysis

The pilot was assessed using descriptive statistics only, as the goal of this stage of the messages is to check for credibility, perceived audience, manipulation checks, and perceived effects. Means and frequencies were used to make these assessments.

Ethical Considerations

Participants were compensated with the opportunity to enter a drawing to receive a gift card by Tango Cards in the value of US$100 (three available cards), US$50 (six available cards), or US$25 (eight available cards) in the formative survey and with extra credit in a communication course of their choice for their participation in the pilot for most participants, with very few receiving Starbucks gift cards instead. Data collection happened through an online survey and was collected anonymously. The drawing information was completely unlinked from the survey responses. This study was exempt from review by the author's institution. Exempt review means the protocol was reviewed by a member of the university's IRB, but a full-board review was not deemed necessary, as risks were considered minimal.

Results

Qualitative Results

Eye-Catching Statistics and Information

In general, participants indicated that they appreciated how informative the messages were and that statistics were provided. One participant wrote that they liked that "[The message] gives a statistic that catches your attention," and another participant mentioned, "I like the percentages displayed in the video." This was a pattern across all test messages where these were things participants appreciated in the messages.

Production Passed the Vibe Check

What is conceptualized here as production value is both aspects, such as the video chosen and the music in the background, as well as descriptions of "vibes" and entertainment value. Participants indicated that the message was "upbeat" and that they liked the background music often. They also appreciated how short the message was while still providing valuable information. One participant said that "[The message] was simple and effective," and a different participant said that "the background video and music was engaging."

No Source to Back It Up? Participants Disliked the Lack of Source for the Information Provided

The most common theme of dislike toward the message was the lack of source for the statistics provided. Multiple participants expressed disdain or skepticism toward the statistics, which could come from a lack of source but also a reluctance to accept that they were true in relation to fellow college students in their institution. One participant said that they disliked that there was "no source to back it up," while another one explained that they disliked that the message "did not show how they got the statistic."

Maybe a Bit Boring

This theme was developed from participants mentioning that the messaging is "boring" or "not very engaging." This theme is hard to conciliate with the presence of participants who did express they appreciated that the message was short and kept them interested throughout. One participant said that "[The message] is boring and unengaging," and another indicated that "[The message] was boring to watch."

Message Understood!

In general, participants understood what the messages were trying to achieve with their call to action and information. One participant explained that their understanding of the message was that "this message is telling people that [UNIVERSITY MASCOT] intervene when it comes to getting people drunk for sex, and so should they," while another participant indicated that "Be an active bystander" was the main message. This is important information to be taken together with the results from the quantitative analysis described in the following paragraphs.

Quantitative Results

Perceived Message Effectiveness

To assess the quantitative results, the mean of the scales was used in reference to the midpoint of the scale, that is, above the midpoint represented acceptance and below midpoint represented rejection. All bystander-related messages (1-9) were above the midpoint (3 on a 1 - 5 scale, where 1=disagree and 5=agree) in terms of perceived message effectiveness. The bottom 2 means in perceived message effectiveness belonged to 2 control messages: message 10 (mean 2.77, SD 1.36) and message 12 (mean 2.85, SD 1.12). Although all condition messages (1-9) were above the midpoint, the lowest mean among them belonged to message 5 (mean 3.19, SD 1.33), and that was lower than the mean for message 11 (mean 3.26, SD 1.27), which is a control message. The highest means belonged to message 6 (mean 3.94, SD 1.15) and message 9 (mean 3.85, SD 1.18). The results can be seen in their completion in Table 3.

XSL•FO
RenderX

Table . Perceived message effectiveness results.

Condition and message		Number of participants per condition	Mean (SD)
Readjustment			
	1	37	3.46 (1.2)
	2	38	3.56 (1.24)
	3	38	3.76 (1.12)
Reinforcement			
	4	34	3.79 (1.13)
	5	36	3.19 (1.33)
	6	36	3.94 (1.15)
Combination			
	7	36	3.67 (1.29)
	8	37	3.69 (1.24)
	9	35	3.85 (1.18)
Control			
	10	40	2.77 (1.36)
	11	35	3.26 (1.27)
	12	36	2.85 (1.12)

Credibility

When assessing the means for credibility for each message, we see that generally, all messages were above the midpoint (3 on a 1 - 5 scale). The highest was achieved by message 6, at mean 4.26 (SD 0.94), and the lowest were obtained for messages 5 and 11, at mean 3.90 (SD_5 0.96, SD_{11} 0.98). The mean (SD) values are presented in Table 4.

Table . Message credibility results.

Condition and message		Number, n	Mean (SD)
Readjustment			
	1	36	4.20 (0.74)
	2	38	3.95 (0.87)
	3	38	3.82 (1.03)
Reinforcement			
	4	34	4.30 (0.87)
	5	36	3.90 (0.96)
	6	36	4.26 (0.94)
Combination			
	7	36	4.22 (0.94)
	8	37	4.18 (0.79)
	9	35	4.10 (0.97)
Control			
	10	40	4.06 (0.75)
	11	34	3.90 (0.98)
	12	36	4.11 (1)

XSL•FO
RenderX

Message Audiences

Participants were asked to assess what audiences they believed the messages they saw would be most suited to. That was done using a Likert-type scale ranging from 1 (disagree) to 5 (agree). All major intended audiences for the message, college students, [university] students, and people who are 18 - 29 years old, had high means in terms of participants perceiving the messages as being appropriate for them as an audience. Among the condition messages, the highest one was Message 7 among college students (mean 4.86, SD 0.42), and the lowest one was Message 3 among people between ages 18 - 29 (mean 4.37, SD 1.08). Table 5 shows a full breakdown of results.

XSL•FO
RenderX

Table . Audience perception means and SDs.

Audience, condition, and message		Number, n	Mean (SD)
College students			
Readjustment			
	1	36	4.78 (0.54)
	2	38	4.58 (0.79)
	3	38	4.68 (0.7)
Reinforcement			
	4	34	4.79 (0.54)
	5	36	4.67 (0.63)
	6	35	4.77 (0.55)
Combination			
	7	36	4.86 (0.42)
	8	37	4.73 (0.65)
	9	35	4.77 (0.49)
Control			
	10	40	4.67 (0.66)
	11	34	4.62 (0.74)
	12	36	4.78 (0.59)
People aged between 18 - 29 years			
Readjustment			
	1	36	4.47 (1.06)
	2	38	4.53 (0.73)
	3	38	4.37 (1.08)
Reinforcement			
	4	34	4.56 (0.93)
	5	36	4.44 (0.91)
	6	35	4.6 (0.74)
Combination			
	7	36	4.61 (0.87)
	8	37	4.57 (0.77)
	9	35	4.57 (0.95)
Control			
	10	40	4.6 (0.78)
	11	34	4.41 (0.86)
	12	36	4.81 (0.4)
[University name] students			
Readjustment			
	1	36	4.67 (0.76)
	2	38	4.61 (0.76)
	3	38	4.68 (0.7)
Reinforcement			
	4	34	4.74 (0.71)
	5	36	4.81 (0.47)

XSL•FO
RenderX

Audience, condition, and message		Number, n	Mean (SD)
	6	35	4.74 (0.7)
Combination			
	7	36	4.81 (0.53)
	8	37	4.73 (0.65)
	9	35	4.74 (0.61)
Control			
	10	40	4.65 (0.62)
	11	34	4.38 (0.99)
	12	36	4.67 (0.72)

Manipulation Checks

The goal of the manipulation checks was to ascertain if the message conditions were properly different from each other in what they were trying to achieve. A total of 6 statements were developed to assess different aspects of the messages and check that manipulation efforts are working. Results indicated that, indeed, all message conditions were properly differentiated from each other. For example, the control condition message 11, which had the highest mean among the control messages, had a lower mean in the manipulation check "the message I saw encouraged me to be an active bystander" (mean 3.15, SD 1.46) in a 5-point scale than the lowest noncontrol message (message 3: mean 3.74, SD 1.35). Full results can be found in the Table 6.

XSL•FO
RenderX

Table . Manipulation check descriptive statistics.

Manipulation check, condition, and message	Number, n	Mean (SD)
The message I saw encouraged me to be an active bystander		
Readjustment		
1	36	4 (1.24)
2	38	3.82 (1.18)
3	38	3.74 (1.35)
Reinforcement		
4	34	4.18 (1.24)
5	35	4.03 (1.2)
6	35	4.06 (1.35)
Combination		
7	36	4.17 (1.23)
8	37	4.16 (1.07)
9	34	4.09 (1.26)
Control		
10	39	2.97 (1.58)
11	34	3.15 (1.46)
12	36	2.86 (1.48)
The message I saw encouraged me to be part of the group that helps		
Readjustment		
1	36	4.5 (0.74)
2	38	4.08 (1)
3	38	4.13 (0.99)
Reinforcement		
4	34	4.41 (0.96)
5	35	4.37 (0.81)
6	35	4.63 (0.81)
Combination		
7	36	4.47 (0.94)
8	37	4.49 (0.77)
9	34	4.47 (0.75)
Control		
10	39	3.05 (1.62)
11	34	3.65 (1.28)
12	36	3.17 (1.52)
The message I saw suggests very few [university mascot] help others at parties		
Readjustment		
1	36	3.94 (1.04)
2	38	2.95 (1.16)
3	38	3.84 (1.24)
Reinforcement		
4	34	3.06 (1.67)
5	35	3.03 (1.67)

XSL•FO
RenderX

Manipulation check, condition, and message		Number, n	Mean (SD)
	6	35	2.8 (1.71)
Combination			
	7	36	3.97 (1.21)
	8	37	3.38 (1.34)
	9	34	3.53 (1.24)
Control			
	10	39	2.49 (1.3)
	11	34	2.68 (1.36)
	12	36	2.5 (1.4)
The message I saw suggests a lot of [university mascot] help others at parties			
Readjustment			
	1	36	3 (1.24)
	2	38	3.87 (0.84)
	3	38	2.55 (1.13)
Reinforcement			
	4	34	3.97 (1.31)
	5	35	4.23 (1.14)
	6	35	4.49 (0.7)
Combination			
	7	36	2.89 (1.47)
	8	37	3.73 (1.17)
	9	34	3.06 (1.25)
Control			
	10	39	2.59 (1.23)
	11	34	2.97 (1.34)
	12	36	2.67 (1.29)
The message I saw suggests [university mascot] trust their friends			
Readjustment			
	1	36	3.11 (1.17)
	2	38	3.74 (0.86)
	3	38	2.89 (1.23)
Reinforcement			
	4	34	4.32 (0.95)
	5	35	4.37 (0.94)
	6	35	4.43 (0.85)
Combination			
	7	36	3.44 (1.3)
	8	37	4.11 (0.84)
	9	34	3.74 (1.11)
Control			
	10	39	2.69 (1.22)
	11	34	3.41 (1.33)
	12	36	2.83 (1.28)

XSL•FO
RenderX

Manipulation check, condition, and message		Number, n	Mean (SD)
The message I saw contains tips on how to party safely			
Readjustment			
	1	36	3.22 (1.36)
	2	38	3.11 (1.43)
	3	38	2.61 (1.37)
Reinforcement			
	4	34	3.12 (1.45)
	5	35	3.2 (1.28)
	6	35	3.83 (1.29)
Combination			
	7	36	3.06 (1.55)
	8	37	3.19 (1.43)
	9	34	3.15 (1.44)
Control			
	10	39	4.31 (1.22)
	11	34	4.35 (0.98)
	12	36	4.44 (0.84

Discussion

Principal Findings

In this study, the feasibility of a social norms message was assessed by the means of a formative survey study, demonstrating that there was indeed a discrepancy in terms of pluralistic ignorance. The presence of pluralistic ignorance allows for the development of social normative messaging. After establishing the presence of pluralistic ignorance, a message was piloted, and results indicated that the messages reached the intended audience, had the intended perceived effect, were perceived as credible, and the manipulation checks indicated the conditions were properly differentiated from each other while representing their purpose. The feasibility of messages focusing on bystander barriers and combining bystander barriers and the social norms approach was also explored. These messages were also highly accepted among the participants, as well as highly understood. Further research is necessary to explore which theoretical approach is ideal to reach this population in terms of the topic of bystander intervention in sexual assault situations. This was the first study to consider combining opposing approaches to bystander intervention messaging—one where it is suggested that people should consider the behavior of their peers [6] and one where they should take more responsibility for their own actions [7].

Although all 3 approaches showed appropriate levels of perceived message effect and credibility, it is necessary to underscore that the reinforcement condition and the combination condition were particularly successful. This suggests that the bystander barrier associated with taking responsibility may not be as strong of a factor in bystander intervention messaging as it is an attitude to be changed by messaging. The fact that the common denominator among the more successful messages is the presence of normative information is in line with the current literature surrounding sexual violence prevention in the form of media messages [30] as well as theoretical frameworks used to address sexual violence prevention [6,7], further building on it by adding a component that goes beyond norms only. However, more must be explored in terms of combined message by proper experimental designs.

The discrepancy between the qualitative and quantitative results of the pilot in terms of credibility highlighted the need to conduct mixed methods pilots to holistically assess messages that concern public health. Although the quantitative results indicated that credibility was appropriate as is, the qualitative data indicated that there were ways to improve credibility even more. The message should be adjusted before further evaluation of its effects. Possible reasons for this discrepancy could be that, although participants find the message credible overall, as in, the message had face validity to them, they would expect proper sourcing of this information. When asked quantitatively if the messages were credible, they may tend to say it is credible, but when they have the space to expand on that, they explain how this credibility could become more formal.

In terms of practical implications, this study shows the possibility for universities and other organizations that conduct bystander intervention training to provide their publics with messages that can underscore these trainings. It is unlikely that an Instagram reel-based message alone can change one's behavior. However, media messages can serve to reinforce adopted norms [30-32]. This can, in turn, maintain a change in behavior from a more complex intervention that goes beyond media exposure.

XSL•FO
RenderX

Limitations

This study, although novel, is not without limitations. The messages were highly tailored for the students of the university where the study was conducted. Although this is ideal for social normative messaging, it can interfere with the scalability of the study. Furthermore, more assessments using inferential statistics are necessary to establish the potential effects of these messages. These are the next steps in this line of research. In addition, the messages lacked proper sourcing, which participants identified particularly in the qualitative results as a flaw. Future developments of these messages should provide proper sourcing to enhance the credibility of this intervention.

Conclusion

Overall, this study explores how different theories can contribute in different ways to the same public health concern. The messages assessed here have proven to be adequate for the next stage of evaluation with an experimental design. However, this next step would not be as rigorous if the procedures discussed here had not been conducted first.

Data Availability

Data can be made available upon request at the discretion of the author.

Conflicts of Interest

None declared.

Checklist 1
CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist of items that apply to this study.
[PDF File, 27 KB - formative_v9i1e66769_app1.pdf]

References

1. Cantor D, Fisher B, Chibnall S, et al. Report on the AAU campus climate survey on sexual assault and sexual misconduct. Westat. 2019. URL: https://cpb-us-w2.wpmucdn.com/voices.uchicago.edu/dist/6/294/files/2019/10/2019-Campus-Climate-Survey-Results.pdf [accessed 2024-12-25]
2. Smith SG, Zhang X, Basile KC, et al. The National Intimate Partner and Sexual Violence Survey: 2015 data brief — updated release. Centers for Disease Control and Prevention. 2018. URL: https://stacks.cdc.gov/view/cdc/60893 [accessed 2025-01-21]
3. Your role in preventing sexual assault. RAINN. URL: https://www.rainn.org/articles/your-role-preventing-sexual-assault [accessed 2023-08-12]
4. Coker AL, Bush HM, Fisher BS, et al. Multi-college bystander intervention evaluation for violence prevention. Am J Prev Med 2016 Mar;50(3):295-302. [doi: 10.1016/j.amepre.2015.08.034] [Medline: 26541099]
5. Kleinsasser A, Jouriles EN, McDonald R, Rosenfield D. An online bystander intervention program for the prevention of sexual violence. Psychol Violence 2015 Jul;5(3):227-235. [doi: 10.1037/a0037393] [Medline: 26240776]
6. Berkowitz AD. An overview of the social norms approach. In: Changing the Culture of College Drinking: A Socially Situated Health Communication Campaign: Hampton Press; 2005:193-214.
7. Burn SM. A situational model of sexual assault prevention through bystander intervention. Sex Roles 2009 Jun;60(11-12):779-792. [doi: 10.1007/s11199-008-9581-5]
8. Orchowski LM, Merrill JE, Oesterle DW, et al. Integrated alcohol use and sexual assault prevention program for college men who engage in heavy drinking: randomized pilot study. JMIR Form Res 2023 Nov 23;7:e47354. [doi: 10.2196/47354] [Medline: 37995129]
9. Pant I, Kang BA, Rimal R. Improving bystander self-efficacy to prevent violence against women through interpersonal communication using mobile phone entertainment education: randomized controlled trial. JMIR Form Res 2023 Apr 14;7:e38688. [doi: 10.2196/38688] [Medline: 37058330]
10. Kimble M, Neacsiu AD, Flack WF, Horner J. Risk of unwanted sex for college women: evidence for a red zone. J Am Coll Health 2008;57(3):331-338. [doi: 10.3200/JACH.57.3.331-338] [Medline: 18980890]
11. McMahon S, Banyard VL. When can I help? a conceptual framework for the prevention of sexual violence through bystander intervention. Trauma Violence Abuse 2012 Jan;13(1):3-14. [doi: 10.1177/1524838011426015] [Medline: 22096017]
12. Coker AL, Cook-Craig PG, Williams CM, et al. Evaluation of green dot: an active bystander intervention to reduce sexual violence on college campuses. Violence Against Women 2011 Jun;17(6):777-796. [doi: 10.1177/1077801211410264] [Medline: 21642269]
13. Green Dot. Culture of Respect. 2014. URL: https://cultureofrespect.org/program/green-dot/ [accessed 2023-08-11]
14. Fischer P, Krueger JI, Greitemeyer T, et al. The bystander-effect: a meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. Psychol Bull 2011 Jul;137(4):517-537. [doi: 10.1037/a0023304] [Medline: 21534650]
15. Ganti N, Baek S. Why people stand by: a comprehensive study about the bystander effect. J Stud Res 2021;10(1). [doi: 10.47611/jsrhs.v10i1.1390]

XSL•FO
RenderX

16. Lurigio AJ. Crime narratives, dramatizations, and the legacy of the Kitty Genovese murder. Crim Just Behav 2015 Jul;42(7):782-789. [doi: 10.1177/0093854814562954]
17. Coker AL, Bush HM, Cook-Craig PG, et al. RCT testing bystander effectiveness to reduce violence. Am J Prev Med 2017 May;52(5):566-578. [doi: 10.1016/j.amepre.2017.01.020] [Medline: 28279546]
18. Salazar LF, Schipani-McLaughlin AM, Sebeh Y, Nizam Z, Hayat M. A web-based sexual violence, alcohol misuse, and bystander intervention program for college women (RealConsent): randomized controlled trial. J Med Internet Res 2023 Jun 21;25:e43740. [doi: 10.2196/43740] [Medline: 37342080]
19. Fung ICH, Blankenship EB, Ahweyevu JO, et al. Public health implications of image-based social media: a systematic review of Instagram, Pinterest, Tumblr, and Flickr. Perm J 2020;24:18.307. [doi: 10.7812/TPP/18.307] [Medline: 31852039]
20. Pinto PA, Antunes MJL, Almeida AMP. Public Health on Instagram: an analysis of health promotion strategies of Portugal and Brazil. Procedia Comput Sci 2021;181:231-238. [doi: 10.1016/j.procs.2021.01.142]
21. Gandola K. Instagram reels VS instagram stories. Net Influencer. 2022. URL: https://www.netinfluencer.com/instagram-reels-vs-instagram-stories/ [accessed 2023-08-11]
22. Social media fact sheet. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2023-08-12]
23. Hust SJT, Marett EG, Lei M, et al. Health promotion messages in entertainment media: crime drama viewership and intentions to intervene in a sexual assault situation. J Health Commun 2013;18(1):105-123. [doi: 10.1080/10810730.2012.688241] [Medline: 23020668]
24. Couto L, Willoughby JF. #LoveYourBody: an experimental test of the effects of objectification and body appreciation content on Instagram fitness and health posts among young women. Health Commun 2024;39(11):2298-2306. [doi: 10.1080/10410236.2023.2265647]
25. Willoughby JF, Hust SJT, Li J, Couto L. Social media, marijuana and sex: an exploratory study of adolescents' intentions to use and college students' use of marijuana. J Sex Res 2022 Jan;59(1):85-97. [doi: 10.1080/00224499.2020.1827217] [Medline: 33034543]
26. Baig SA, Noar SM, Gottfredson NC, Boynton MH, Ribisl KM, Brewer NT. UNC perceived message effectiveness: validation of a brief scale. Ann Behav Med 2019 Jul 17;53(8):732-742. [doi: 10.1093/abm/kay080] [Medline: 30321252]
27. Appelman A, Sundar SS. Measuring message credibility: construction and validation of an exclusive scale. J Mass Commun Q 2016;93(1):59-79. [doi: 10.1177/1077699015606057]
28. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
29. Braun V, Clarke V. Conceptual and design thinking for thematic analysis. Qual Psychol 2022;9(1):3-26. [doi: 10.1037/qup0000196]
30. Hust SJT, Kang S, Couto L. Using narrative-based social norms messaging to affect group norms. In: Orchowski LM, DeJong W, editors. Health Behavior and Social Norms: Theory, Research, and Strategies for Change (Forthcoming): American Psychological Association.
31. Hust SJT, Kang S, Couto L, Li J, Rodgers KB. Explaining college men's rape myth acceptance: the role of sports media, masculine norms & fraternity membership. J Health Commun 2023;28(8):477-486. [doi: 10.1080/10810730.2023.2222077]
32. Hust SJT, Marett EG, Lei M, Ren C, Ran W. Law & Order, CSI, and NCIS: the association between exposure to crime drama franchises, rape myth acceptance, and sexual consent negotiation among college students. J Health Commun 2015;20(12):1369-1381. [doi: 10.1080/10810730.2015.1018615] [Medline: 26418170]

Abbreviations

IRB: institutional review board
LGBTQ+: lesbian, gay, bisexual, transgender, queer+
UNC PMES: University of North Carolina's Perceived Message Effectiveness Scale

Edited by A Mavragani; submitted 22.09.24; peer-reviewed by K Fortson, T Vertommen; revised version received 25.11.24; accepted 26.11.24; published 27.01.25.

Please cite as:
Couto L
The Development of an Instagram Reel-Based Bystander Intervention Message Among College Students: Formative Survey and Mixed Methods Pilot Study
JMIR Form Res 2025;9:e66769
URL: https://formative.jmir.org/2025/1/e66769
doi:10.2196/66769

XSL•FO
RenderX

© Leticia Couto. Originally published in JMIR Formative Research (https://formative.jmir.org), 27.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluating Older Adults' Engagement and Usability With AI-Driven Interventions: Randomized Pilot Study

Marcia Shade[1*], BS, PhD; Changmin Yan[2*], PhD; Valerie K Jones[2*], PhD; Julie Boron[3*], PhD

[1]College of Nursing, University of Nebraska Medical Center, 985330 Nebraska Medical Center, Omaha, NE, United States

[2]College of Journalism and Mass Communications, University of Nebraska Lincoln, Lincoln, NE, United States

[3]Department of Gerontology, University of Nebraska at Omaha, Omaha, NE, United States

[*]all authors contributed equally

Corresponding Author:
Marcia Shade, BS, PhD
College of Nursing, University of Nebraska Medical Center, 985330 Nebraska Medical Center, Omaha, NE, United States

Abstract

Background: Technologies that serve as assistants are growing more popular for entertainment and aiding in daily tasks. Artificial intelligence (AI) in these technologies could also be helpful to deliver interventions that assist older adults with symptoms or self-management. Personality traits may play a role in how older adults engage with AI technologies. To ensure the best intervention delivery, we must understand older adults' engagement with and usability of AI-driven technologies.

Objective: This study aimed to describe how older adults engaged with routines facilitated by a conversational AI assistant.

Methods: A randomized pilot trial was conducted for 12-weeks in adults aged 60 years or older, self-reported living alone, and having chronic musculoskeletal pain. Participants (N=50) were randomly assigned to 1 of 2 intervention groups (standard vs enhanced) to engage with routines delivered by the AI assistant Alexa (Amazon). Participants were encouraged to interact with prescribed routines twice daily (morning and evening) and as needed. Data were collected and analyzed on routine engagement characteristics and perceived usability of the AI assistant. An analysis of the participants' personality traits was conducted to describe how personality may impact engagement and usability of AI technologies as interventions.

Results: The participants had a mean age of 79 years, with moderate to high levels of comfort and trust in technology, and were predominately White (48/50, 96%) and women (44/50, 88%). In both intervention groups, morning routines (n=62, 74%) were initiated more frequently than evening routines (n=52, 62%; z=−2.81, P=.005). Older adult participants in the enhanced group self-reported routine usability as good (mean 74.50, SD 11.90), and those in the standard group reported lower but acceptable usability scores (mean 66.29, SD 6.94). Higher extraversion personality trait scores predicted higher rates of routine initiation throughout the whole day and morning in both groups (standard day: B=0.47, P=.004; enhanced day: B=0.44, P=.045; standard morning: B=0.50, P=.03; enhanced morning: B=0.53, P=.02). Higher agreeableness (standard: B=0.50, P=.02; enhanced B=0.46, P=.002) and higher conscientiousness (standard: B=0.33, P=.04; enhanced: B=0.38, P=.006) personality trait scores predicted better usability scores in both groups.

Conclusions: he prescribed interactive routines delivered by an AI assistant were feasible to use as interventions with older adults. Engagement and usability by older adults may be influenced by personality traits such as extraversion, agreeableness, and conscientiousness. While integrating AI-driven interventions into health care, it is important to consider these factors to promote positive outcomes.

Trial Registration: ClinicalTrials.gov NCT05387447; https://clinicaltrials.gov/study/NCT05387447

(JMIR Form Res 2025;9:e64763) doi:10.2196/64763

KEYWORDS

voice assistant; interventions; usability; engagement; personality; older adults; aging; technology; artificial intelligence; AI; self-management; pilot trial; chronic; musculoskeletal pain; AI assistant; Alexa; user experience; digital health; digital intervention; mobile phone

Introduction

Digital technologies may be beneficial for helping older adults stay connected and directing them to health resources [1]. Newer, technology-based interventions may provide a way to assist with the self-management of chronic illnesses and improve social connections for older adults [2]. Traditional technologies such as computers and mobile tablets have been adopted by aging adults but can be abandoned depending on acceptance and ease of use [3]. Newer technologies such as conversational

XSL•FO
RenderX

voice assistants are a type of artificial intelligence (AI) that have the capacity to assist and have basic interactions with individuals. Older adults can engage with conversational AI through voice, which offers an advantage over spending extensive time learning a new program or manually operating small sized device features. Among Americans aged 18 years and older, 62% use an AI conversational voice assistant on devices such as smart speakers, smartphones, TV remotes, in-car systems, computers, and tablets among others [4]. With the increasing population of older adults, conversational voice assistants may be an optimal modality to deliver interventions.

The use of conversational voice assistants has been studied as an intervention for chronic conditions such as heart failure, lung disorders, and mental health [5]. Their use has been explored for promoting behaviors such as physical activity and for health care support [6,7]. Studies are emerging in relation to conversational voice assistants independently influencing loneliness, social isolation, and pain [8-13]. These studies have primarily focused on acceptability, barriers, and the design of conversational voice assistants for older adult populations. In addition, a recent review of the literature only focused on the use and acceptability of conversational voice assistants to reduce loneliness; none of the conversational voice assistant-based interventions were personalized [13,14]. Kocabelli et al [15] found several conversational voice assistants delivered interventions that were personalized, but lacked theoretical framework or evidence-based approach for implementation. Further exploration is needed to understand older adults' engagement with conversational voice assistants to ensure optimal outcomes when used as a modality for intervention delivery.

An individual's personality can influence acceptance and needs to be considered as a part of usability and the engagement with technology-driven interventions [16]. In younger populations, openness to experience and agreeableness, have been shown to be positively related to usability of technology [17].

Personality traits are characteristics of individuals that typically remain stable over the adult life course. Adults from their mid-to-late 20s and older were found to score similarly across the lifespan on the Big Five traits (ie, openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism [18]. The OCEAN acronym is used to describe these traits. The "O" personality trait of being open to experience is the appreciation of a variety of experiences, "C" conscientiousness is the exhibition of self-discipline, and "E" extraversion is the amount of engagement one has with the external world. The trait of "A" is, agreeableness, which is a concern for social harmony and the "N" is neuroticism trait, which is the tendency of a person to exhibit negative emotions. An individual's personality traits play a role in health care, particularly when implementing patient-centered care and encouraging intervention adherence [19]. Loneliness, which is prevalent in older adults, may influence the presentation of personality traits. A meta-analysis documented negative associations between loneliness and extraversion, agreeableness, conscientiousness, and openness, and a positive association between loneliness and neuroticism [20]. Another study found that loneliness was associated with varying levels of extraversion, neuroticism, and agreeableness [21]. When using technology to deliver interventions in older adults, engagement may differ in those that self-report loneliness; however, this warrants further exploration.

An individual's personality may influence the engagement and user experience with AI technology. Within the realm of traditional technologies, the strongest correlates of older adults' perceived usefulness and ease of use of computers were the personality traits of agreeableness and openness [22,23]. Openness to experience has been associated with increased probability of internet use [16]. Furthermore, higher agreeableness and lower neuroticism in older adults have been associated with perceived usability with an automated vehicle [24].

Since the AI boom, the relationship between the personality traits of older adults and use of conversational voice assistants has not been extensively explored and may impact engagement with AI driven interventions. A commercially available AI voice assistant, that is, Alexa, was used to deliver prescribed interactive routines as nonpharmacological interventions for pain and loneliness. In this analysis of the data, the following research questions were explored:

1. When do older adults initiate the various prescribed routines?
2. What are the older adults' perceived usability scores with different intervention routines delivered by AI through conversational voice assistants?
3. What is the relationship between older adults' personality and their engagement with the different intervention routines?
4. What is the relationship between older adults' personality and the perceived usability of different intervention routines delivered by AI through conversational voice assistants?

Methods

Overview

This study was a 12-week pilot randomized trial conducted with a convenience sample of participants who were 60 years of age or older, self-reported living alone, spoke English, and experienced pain symptoms. A sample size calculation was performed a priori for this pilot study. A sample of 59 participants was calculated to account for 10% study attrition, 80% power, 0.35 effect size, and 0.05 significance. The Mini Cog was used to screen individuals to ensure cognitive ability of participants [25]. Individuals with memory loss or with previous or current use of conversational voice assistants were excluded. Informed consent was obtained prior to initiating study related activities. Once participants were enrolled, they were given a study identification number, and a computerized sorter randomly assigned them to one of the two intervention groups. Participants were compensated for their time during this study. One $25 dollar gift card was given to participants after completing baseline questionnaires, one $25 gift card after completing 12-week questionnaires, and were allowed to keep the Amazon Echo Dot device.

XSL•FO
RenderX

Intervention Groups

Prescribed routines were used for the intervention, such as the frequency of prescription medications (eg, daily, twice daily, as needed). The routines were created based on nonpharmacologic strategies [26] and selected from evidence-based interventions that have a positive influence on pain and loneliness [2,27,28]. After screening and enrollment, participants were randomly assigned to intervention groups.

The 2 intervention groups included the conversational voice assistant standard (CVA-S) routine and the conversational voice assistant enhanced (CVA-E) routine. Standard routines were more generalized and static, while the enhanced routines were tailored, as it has been well documented that individualized or tailored interventions increase engagement and adherence [29-31]. The AI routines in the CVA-S group included a generalized (without personalization) Alexa greeting such as stating, “good morning” or “good evening.” The routines consisted of a combination of a generalized (1) meditation, (2) game, (3) music, and (4) jokes. The CVA-S group’s “as needed” routine was a general greeting and interactive activity. The AI routines in the CVA-E group included more sociable greetings that were personalized to each participant’s name preference such as stating, “hi there, Jim” or “good evening, Betty.” The CVA-E routine consisted of a combination of personalized (1) pain location and pain type meditation (2) interactive game preferences, (3) genre or musical artist, and (4) choice of preferred joke style, devotionals, interactive stretching and or other interactive activities.

Routine Programming

The conversational voice assistant “Alexa” was used for this study because the 65+ population prefers Alexa among commercially available voice assistants [32]. Routines were programmed within the Alexa app using available actions and skills created by software app developers. Actions are standard preprogrammed interactions such as playing music or telling a joke, and skills are voice-activated apps such as health fitness, education, games, podcasts, and meditations that can be used with Alexa and the smart speaker. Before skill selection for this study, research personnel reviewed ratings and tested the skills to make sure they worked and would be appropriate for older populations. Routines were initiated by participants speaking to an Amazon Echo Dot device and stating “Alexa, start my routine,” which then cued the sequence of selected actions and skills.

Measures and Instruments

Data collected from participants was de-identified and included characteristics such as demographics, health history, and comfort with and trust of technology, assessed on a 4-point scale ranging from 1 (none) to 4 (high). The Big Five Inventory (BFI) was used to measure personality traits of the participants. The BFI contains 44 questions that represent the different traits of openness (Cronbach α=0.77), conscientiousness (Cronbach α=0.72), extraversion (Cronbach α=0.79), agreeableness (Cronbach α=0.78), and neuroticism (Cronbach α=0.70) [18,33].

Feasibility outcomes included capturing engagement with the AI assistant routines and self-reported usability. Engagement with routines was measured at the end of the study. Participants’ voice assistant profiles were securely accessed to track the date and time stamps on the incidence of routine initiation with the standard and enhanced loneliness routines. Usability was measured using the System Usability Scale (SUS) administered at the end of the study. A 10-item system usability scale (ranging from 0 to 100) was created (Cronbach α=0.75) [34].

Data Analysis

The data analysis strategy for this study was multifaceted, aimed at understanding the impact of different intervention conditions on routine initiation rates, the influence of personality traits on engagement with Alexa-based routines, and perceptions of system usability. Given the nature of the data and the research questions, the analysis used both descriptive statistics and inferential statistical tests to explore these relationships and test hypotheses. This comprehensive data analysis strategy, incorporating both nonparametric tests for hypothesis testing and regression analysis for exploring relationships between personality traits and key outcomes, facilitated a nuanced understanding of the factors influencing routine initiation and system usability in the context of Alexa-based interventions.

Descriptive statistics in addition to normality tests were conducted. Initial analysis involved descriptive statistics (mean, median, SD, skewness, and kurtosis) to summarize the central tendency and dispersion of routine initiation rates and system usability scores. The Shapiro-Wilk test was used to assess the normality of distributions for these variables. Given the significant results from the Shapiro-Wilk test indicating nonnormal distributions for both routine initiation rates and system usability scores, nonparametric tests were chosen for hypothesis testing.

Nonparametric tests were applied to test the differences in routine initiation rates and SUS scores across conditions (morning vs evening and standard vs enhanced). The Wilcoxon Signed-Rank test was used to compare morning and evening routine initiation rates within participants. Mann-Whitney *U* test was used to test differences between the standard and enhanced conditions in terms of routine initiation rates and SUS scores. This test was chosen due to its suitability for comparing 2 independent samples without assuming normal distribution.

Personality Traits Analysis

Multiple regression analyses were conducted to explore the relationship between participants’ personality traits and their engagement with Alexa routines, as well as the perceived usability of the system. These analyses were performed for the total sample and separately for each condition (standard vs enhanced) to determine if personality traits differentially influenced outcomes based on the intervention type. The choice of multiple regression was guided by the interest in assessing the predictive power of multiple independent variables (personality traits) on a continuous dependent variable (number of routine initiations or system usability scores).

Ethical Considerations

The University of Nebraska Medical Center Institutional Review Board approved the study protocol (IRB #0177 - 21-EP).

XSL•FO
RenderX

Results

Demographic findings

The study included 50 participants with mean age 79.24 (SD 7.85; median 78, range 65 - 98). The majority (44/50, 88%) were female (male: 6/50, 12%) and most (48/50, 96%) participants identified as White (African American: 2/50, 44%). Most of the older adults were either widowed (22/50, 44%) or legally divorced or separated (21/50, 42%), while the remaining participants were either living apart from their partner due to health reasons (5/50, 10%) or single (2/50, 4%). Most ((40/50, 80%) participants had completed a college degree or higher, while 16% (8/50) had completed high school, and (2/50, 4%) had attended trade school. The majority (32/50, 64%) reported incomes between US $10,000-$50,000, while 4% (2/50) earned below US $10,000, 14% (7/50) earned between US $50,001-$90,000, and 18% (9/50) earned above US $90,000. Participants reported moderate to high levels of comfort (mean 2.75, SD 0.67) and trust in technology (mean 2.94, SD 0.91) on a scale of 1-5.

Engagement and Perceived Usability of Prescribed Routines

Participants were encouraged to engage with and initiate the routines twice daily for 12 weeks, which equated to 168 interactions in total. As shown in Table 1, the rates of routine initiations were significantly higher among participants in the CVA-E group than the CVA-S group. While observing the prescribed time of day for engagement, the routine initiation rates were higher in the morning.

Table . Comparisons of daily routine initiations of participants in the CVA-S[a] and CVA-E[b] groups in this pilot study of older adults' engagement with 12-week artificial intelligence (AI)-driven interventions.

Condition	Routine initiation (Time of day)	Routine initiation frequency (actual number of interactions/total number of possible interactions, %)	Number of initiations, mean (SD)	*z* score	*P* value
Full sample (n=50)	Morning	62/84[c], 74	62.13 (32.85)	−2.81	.005
Full sample (n=50)	Evening	52/84, 62	51.94 (32.13)		
CVA-E (n=26)	Entire day	126/168[d], 75	126.53 (59.9)	−1.98	.048
CVA-S (n=24)	Entire day	101/168, 60	100.56 (54.88)		
CVA-E (n=26)	Morning	62/84, 74	62.24 (31.86)	−3.16	.002
CVA-S (n=24)	Morning	49/84, 58	49.01 (29.17)		
CVA-E (n=26)	Evening	44/84, 52	52.29 (32.3)	−0.13	.90
CVA-S (n=24)	Evening	44/84, 52	51.55 (32.63)		

[a]CVA-S: conversational voice assistant standard.

[b]CVA-E: conversational voice assistant enhanced.

[c]Number of interactions during either morning and evening.

[d]Total number of interactions during the day.

Given the variances in rates of routine initiations, and a notable preference for morning over evening initiations across conditions, group comparisons were conducted using Wilcoxon analyses to examine the interaction between intervention timing (morning vs evening) and intervention groups (CVA-S vs CVA-E using the Mann-Whitney *U* test). As presented in Table 1, the CVA-E group initiated "morning routines" significantly more often than those in the CVA-S group. In the evening, the rates of routine initiation were not significantly different between the groups.

The perceived usability of the AI assistant showed favorable results. The SUS scores of the overall sample of participants were "good" (mean 70.56, SD 10.58). When conditions were analyzed separately, the CVA-E group's SUS scores were good (mean 74.50, SD 11.90), while the CVA-S group scores were "OK" (mean 66.29, SD 6.94).

Personality and Engagement With Prescribed Routines

As displayed in Figures 1 and 2, the overall model including all Big Five personality traits was not a good fit for predicting of the daily number of routine initiations ($F_{5,44}$=1.45, P=.23, R^2=.14), morning routine initiation ($F_{5,44}$=2.19, P=.07, R^2=.20), or evening routine initiation ($F_{5,44}$=0.72, P=.61, R^2=.08) at α=.05.

However, the extraversion personality trait was significantly associated with routine initiations. It was a significant predictor of overall daily routine initiations (B=0.42, t=2.65, P=.01) and initiation of morning routines (B=0.47, t=3.07, P=.004), but not the evening routines (B=0.28, t=1.73, P=.09). The other 4 personality traits did not predict routine initiations.

XSL•FO
RenderX

Figure 1. Personality trait predictors of routine initiations throughout the day in this pilot study of older adults' engagement with 12-week artificial intelligence (AI)-driven interventions.

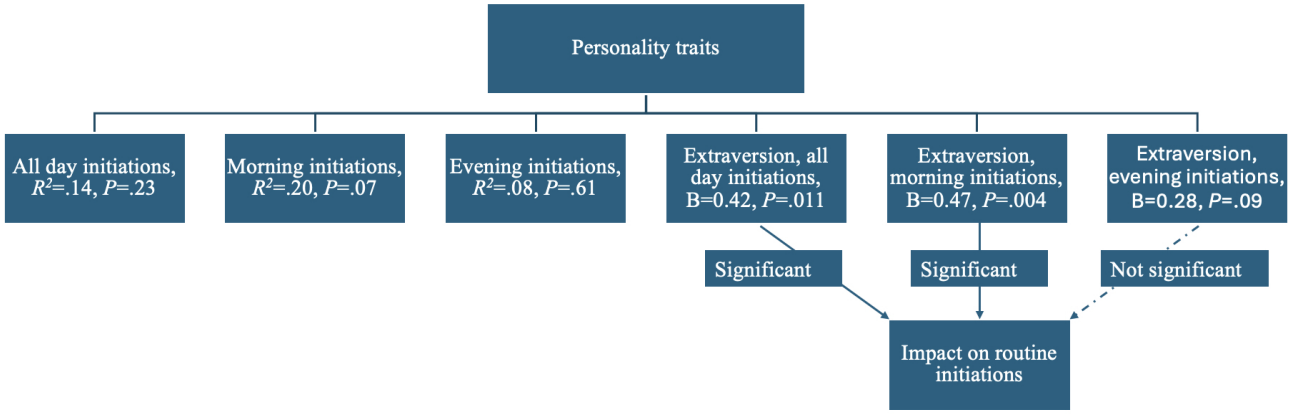
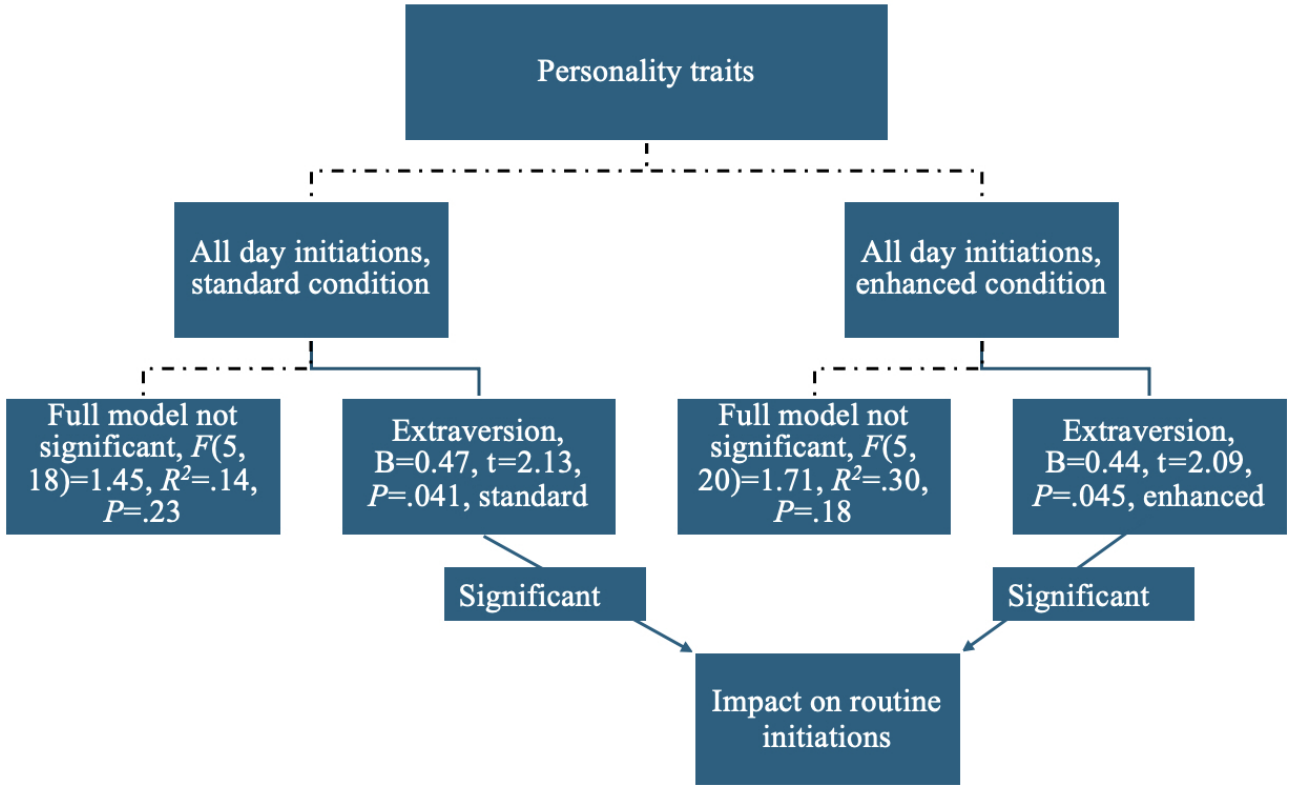


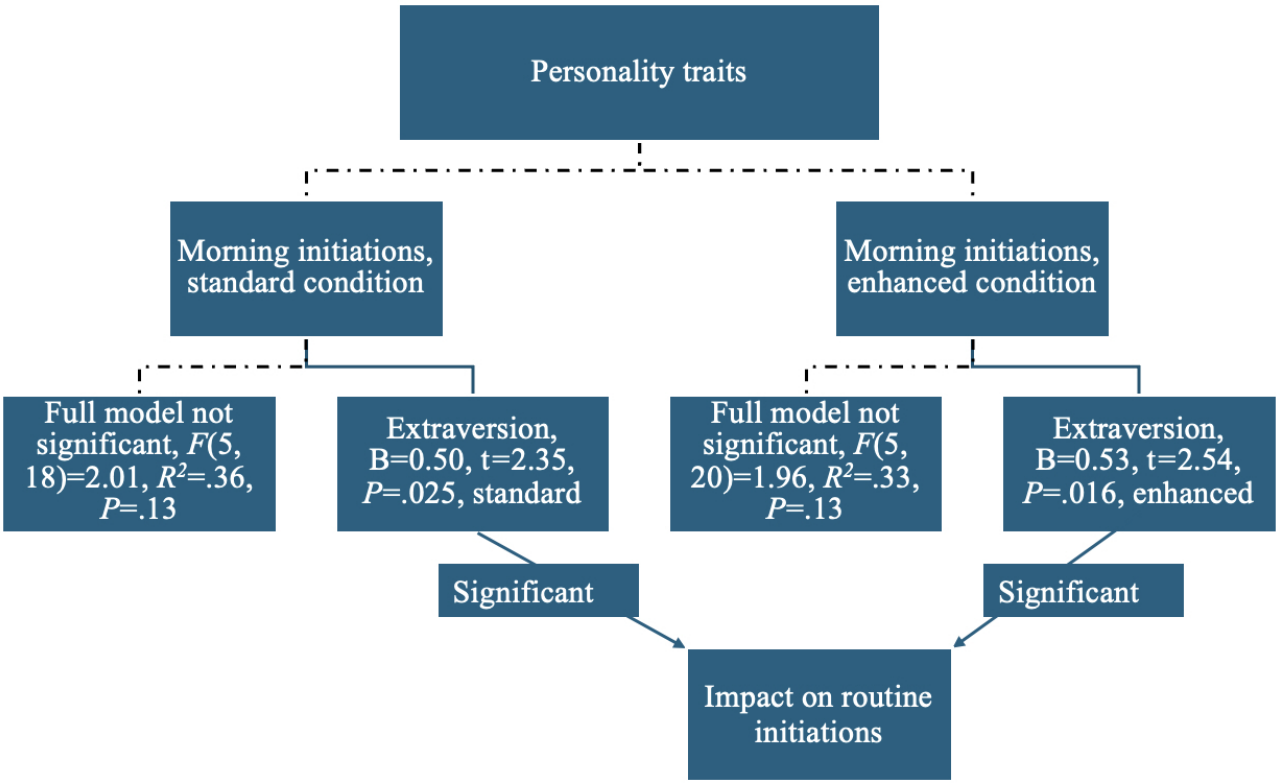
Figure 2. Personality trait predictors of daily routine initiations of participants in the CVA-S (conversational voice assistant standard) and CVA-E (conversational voice assistant enhanced) groups in this pilot study of older adults' engagement with 12-week artificial intelligence (AI)-driven interventions.



Similar patterns were observed when comparing the CVA-S and CVA-E groups. Extraversion significantly predicted all routine initiations in the CVA-S group, (B=0.47, t=2.13, P=.041 and the CVA-E group, B=0.44, t=2.09, P=.045). As displayed in Figure 3, extraversion was a significant predictor of morning routine initiations in the CVA-S group, B=0.50, t=2.35, P=.03, and the CVA-E group, B=0.53, t=2.54, P=.02. None of the Big Five personality traits significantly predicted evening routine initiations in either group.

XSL•FO
RenderX

Figure 3. Personality trait predictors of morning routine initiations of participants in the CVA-S (conversational voice assistant standard) and CVA-E (conversational voice assistant enhanced) groups in this pilot study of older adults' engagement with 12-week artificial intelligence (AI)-driven interventions.



Personality and Perceived Usability of Prescribed Routines

As depicted in Figure 4, among the Big Five personality traits, only agreeableness (B=0.43, t=4.67, P<.001) and conscientiousness (B=0.35, t=3.99, P<.001) positively predicted SUS scores in the overall sample ($F_{5,44}$=29.60, P<.001, R^2=0.77).

As shown in Figure 5, when analyzed separately, both the CVA-S (B=0.50, t=2.49, P=.02) and CVA-E (B=0.46, t=3.31, P=.002) groups reported agreeableness respectively and conscientiousness (B=0.33, t=2.15, P=.04; B=0.38, t=2.97, P=.006 respectively) were positively associated with SUS scores.

XSL•FO
RenderX

Figure 4. Personality trait predictors of participant self-reported usability in this pilot study of older adults' engagement with 12-week artificial intelligence (AI)-driven interventions.

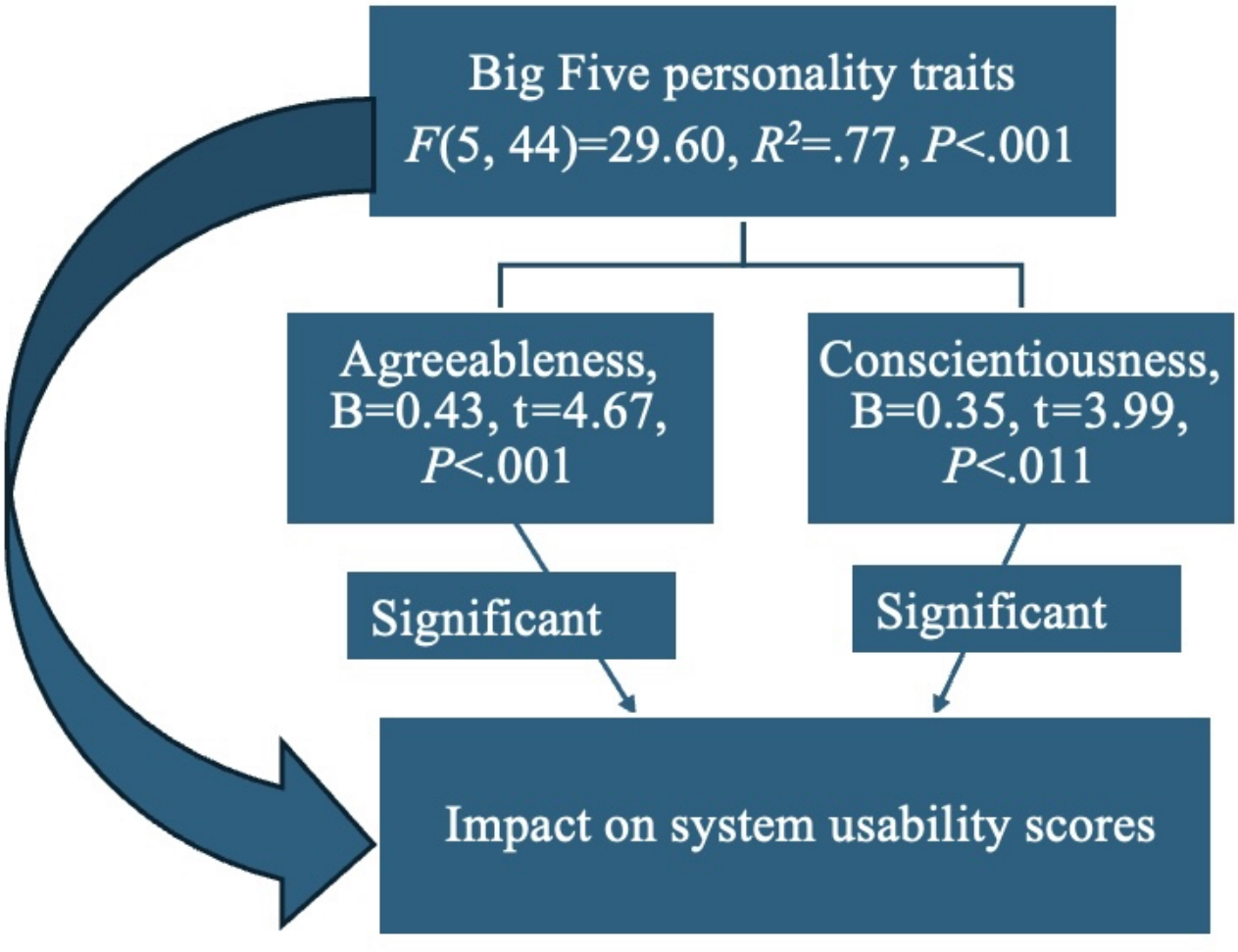
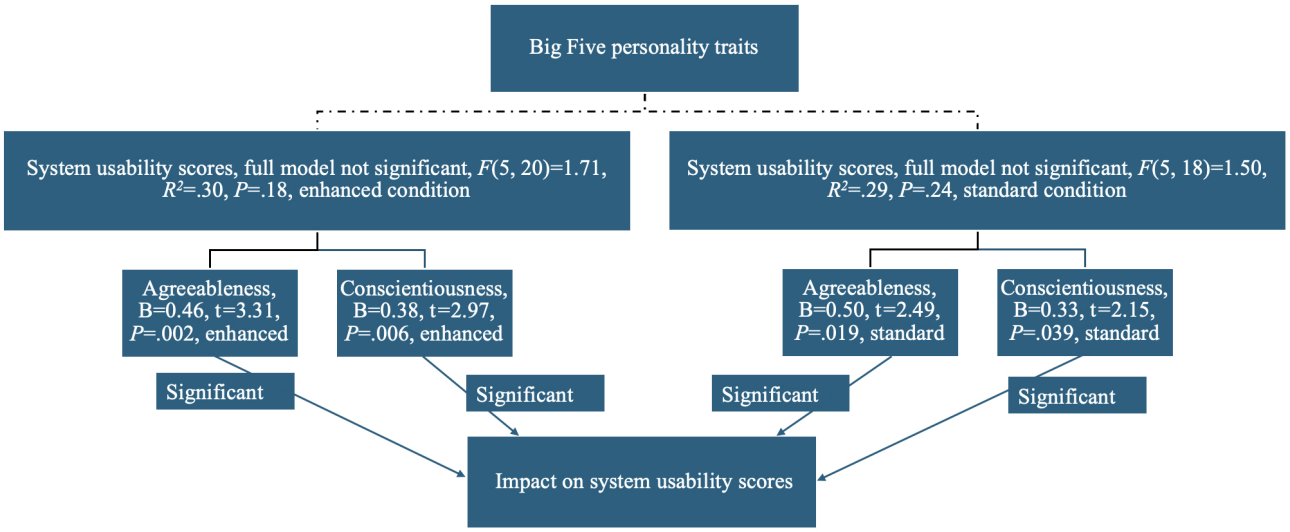


Figure 5. Personality trait predictors of participants self-reported usability in the CVA-S (conversational voice assistant standard) and CVA-E (conversational voice assistant enhanced) groups in this pilot study of older adults' engagement with 12-week artificial intelligence (AI)-driven interventions.



Discussion

Principal Findings

Technology is becoming a popular vehicle for delivering interventions for health care and self-management. In this analysis, we examined older adults' engagement with and usability of AI-delivered routine interventions. Overall, the older participants engaged with the assistant throughout the 12 weeks of the study. Perceived usability was acceptable, with routine initiation rates being higher in the morning. Personality

XSL•FO
RenderX

traits of the participants in this study may have impacted engagement with the prescribed routines. The following discussion provides additional insights into the findings.

Older Adults' Engagement With Prescribed Routines and Perceived Usability

Our study revealed significant differences in routine initiation rates between morning and evening sessions, with participants demonstrating a greater propensity to engage in prescribed routines in the morning. A potential explanation is behavioral habit. Engagement and adherence may be greater in the morning than in the evening because of internal behavioral cues [35]. Another potential explanation for this finding lies within circadian rhythm science. Circadian rhythms are internal processes that regulate the sleep-wake cycle, repeating roughly every 24 hours. They influence physical, mental, and behavioral changes in living organisms. Circadian rhythms are crucial for determining human sleep patterns and have significant implications for designing technology that aligns with these natural processes [36]. Research suggests that morning hours are associated with higher alertness and cognitive functioning, which could facilitate the initiation of health-related routines [37]. The enhanced intervention components may have amplified these effects, indicating that tailored interventions can effectively increase engagement with health-promoting behaviors. The integration of technology with human circadian rhythms or habits may enhance intervention engagement and adherence.

Another possible explanation for the higher initiation rates in the morning could be arthritic pain. Inclusion criteria was self-reported pain symptoms and most of the participants reported a diagnosis of arthritis. Commonly, individuals experience stiffness in the morning that may impact pain symptoms and need for management [38]. In older adults with musculoskeletal pain, technology-based routines or interventions may be best encouraged in the morning.

Older Adults' Personality and Engagement

The results of this study highlight the influence of personality traits on engagement and usability. Extraversion was a significant predictor of routine initiation, particularly in the morning, which is consistent with previous research indicating that extraverted individuals are more likely to engage in health-promoting activities [39]. Extraverts, characterized by their energy, enthusiasm, and sociability, might have the most inherent motivation for social interaction and stimulation. When they perceived the routines as opportunities for engaging experiences or personal growth, it may have made the process of initiating routines more appealing [40,41]. The finding that this effect was more pronounced in the morning suggests that, consistent with their circadian rhythm, extraverts may be prone to engage with technology during early hours, setting a positive tone for the rest of their day.

Consistent with Kortum and Oswald [17], agreeableness was one personality trait that aligned with usability in our sample of older adults. Agreeable and conscientious personality predicted higher perceived usability, suggesting that these traits may influence individuals' perceptions and interactions with technology-based health interventions. It is worth noting the older participants in this study reported moderate to high levels of comfort and trust in technology. Agreeable individuals, known for their cooperative and trusting nature, may be more predisposed to view the Alexa-based intervention in a positive light, attributing any potential usability issues to their own learning curve rather than the system's design flaws. This optimistic viewpoint may lead to higher perceived usability. Conscientious individuals, with their focus on diligence, organization, and reliability, may appreciate systems that are consistent and efficient. The participants' positive evaluation of the system's usability might stem from a recognition of its potential to support their goal-oriented behavior and routine adherence. The structured nature of the preset routines could align well with their preference for orderly environments, further enhancing their perception of the system's usability.

The study reveals a nuanced differentiation between participants' engagement with prescribed routines and their perceptions of system usability, emphasizing the distinct roles of personality traits. Extraversion was found to drive the frequency of routine initiations, particularly in the morning, suggesting that an action-oriented engagement is influenced by a desire for stimulation and interaction. Older adults with extraverted personality traits may be more likely to seek that interaction and treat AI with anthropomorphic properties [42]. Anthropomorphism is when a human attributes human traits to entities or devices that are not human [43]. Researchers have found that people interact with conversational voice assistants in anthropomorphic ways, greeting them, asking them personal questions, exhibiting polite behaviors such as saying "please" and "thank you," and reacting to what the voice assistant did or did not say [12]. Thus, there is a possibility that anthropomorphic interactions with the conversational voice assistants may have impacted engagement and usability of conversational voice assistants to deliver interventions.

Conversely, agreeableness and conscientiousness were associated with more favorable perceptions of system usability, indicating a cognitive evaluation process that values trust, adaptability, and reliability. This distinction underscores the importance of differentiating the motivational drivers behind the use of technology-based health interventions and the cognitive evaluations that shape user satisfaction. By acknowledging the interplay between action-oriented engagement and cognitive evaluations, designers can create more effective, user-friendly interventions tailored to diverse user preferences and personality profiles.

Extraverts, who are stimulated by social interaction and are more action-oriented, particularly in the morning, may benefit from interventions scheduled during this time. Health platforms can incorporate adaptive scheduling features that customize the timing of prompts based on the user's personality profile, thus optimizing the likelihood of engagement and adherence which could impact health outcomes.

Furthermore, understanding the role of personality traits in the acceptance and usability of health technologies can inform the development of more personalized, effective, and user-friendly interventions. The significant role of agreeableness and

XSL•FO
RenderX

conscientiousness in shaping perceptions of system usability suggests that these traits should be considered when designing user interfaces. Users high in agreeableness and conscientiousness value trust, adaptability, and reliability. Therefore, health interventions should prioritize these design elements to cater to their cognitive evaluation processes. For example, features that provide consistent feedback, allow for personalization, and are easy to navigate could be more appealing to these users. Researchers have found that conversational voice assistants exhibit "personality type" traits. The Google and Alexa AI assistants have personality traits of functional intelligence, sincerity, sociability, and creativity [44]. The personality and voice of the assistant has shown to enhance engagement [45]. Future interventions could be tailored for optimal engagement when matching AI assistants and human dominant personalities.

Although the results of this pilot study are informative there are limitations worth noting. There were an even number of participants in each intervention group, however, our sample was relatively small and lacked diversity. The findings are not generalizable to all older adults who report loneliness and pain. Our participants self-reported loneliness and pain, but our study may not have enrolled individuals who were experiencing more intense loneliness and pain. Primary care providers could help us identify older adults with more severe pain and or loneliness. Those without regular internet access in their home due to location (eg, rural areas) or lack of broadband access (due to personal choice or financial restraints) would not be able to interact with conversational voice assistants. Some of these sources of variation could also interact with personality differences. For example, those that choose not to have internet in their homes for personal reasons, may differ from those that do. Furthermore, exploring how to engage those with differing personality characteristics is an important area for future research, so that routines are individualized enough to produce an impact on daily life. Finally, our study was 12-weeks in duration; a longer study would be beneficial to explore the length of time needed to integrate routines that impact well-being and to sustain long term impact.

Conclusions and Implications

Commercially available conversational voice assistants have the potential to deliver nonpharmacologic interventions that are individualized for older adults. Prescribed interactive routines were feasible to use in older adults who self-reported loneliness and pain. Routine engagement may be influenced by the older adults' pain characteristics and personalities. Tailoring these routines to align with individual pain profiles and personality traits may improve the efficacy of interventions and user satisfaction. A personalized approach can enhance adherence and overall outcomes by addressing specific needs and preferences of older adults, making technology-assisted interventions more relevant and efficient in managing their conditions.

Acknowledgments

We acknowledge the role of Victoria Dube, Hyeon Jung Heselton, Taylor Bacon, and our wonderful study participants. This work was supported by Great Plains IDeA-CTR–NIH (NCT05387447).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Checklist 1
CONSORT guidelines for randomized pilot study.
[PDF File, 94 KB - formative_v9i1e64763_app1.pdf]

References

1. Sen K, Prybutok G, Prybutok V. The use of digital technology for social wellbeing reduces social isolation in older adults: a systematic review. SSM Popul Health 2022 Mar;17:101020. [doi: 10.1016/j.ssmph.2021.101020] [Medline: 35024424]
2. Fakoya OA, McCorry NK, Donnelly M. Loneliness and social isolation interventions for older adults: a scoping review of reviews. BMC Public Health 2020 Feb 14;20(1):129. [doi: 10.1186/s12889-020-8251-6] [Medline: 32054474]
3. Bhattarai P, Newton-John TRO, Phillips JL. Quality and usability of arthritic pain self-management apps for older adults: a systematic review. Pain Med 2018 Mar 1;19(3):471-484. [doi: 10.1093/pm/pnx090] [Medline: 28541464]
4. Serrano KJ, Yu M, Riley WT, et al. Willingness to exchange health information via mobile devices: findings from a population-based survey. Ann Fam Med 2016;14(1):34-40. [doi: 10.1370/afm.1888]
5. Bérubé C, Schachner T, Keller R, et al. Voice-based conversational agents for the prevention and management of chronic and mental health conditions: systematic literature review. J Med Internet Res 2021 Mar 29;23(3):e25933. [doi: 10.2196/25933] [Medline: 33658174]

XSL•FO
RenderX

6. Luo TC, Aguilera A, Lyles CR, Figueroa CA. Promoting physical activity through conversational agents: mixed methods systematic review. J Med Internet Res 2021 Sep 14;23(9):e25486. [doi: 10.2196/25486] [Medline: 34519653]
7. Tudor Car L, Dhinagaran DA, Kyaw BM, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug 7;22(8):e17158. [doi: 10.2196/17158] [Medline: 32763886]
8. Kim S, Choudhury A. Exploring older adults' perception and use of smart speaker-based voice assistants: a longitudinal study. Comput Human Behav 2021 Nov;124:106914. [doi: 10.1016/j.chb.2021.106914]
9. Shade MY, Rector K, Soumana R, Kupzyk K. Voice assistant reminders for pain self-management tasks in aging adults. J Gerontol Nurs 2020 Oct 1;46(10):27-33. [doi: 10.3928/00989134-20200820-03] [Medline: 32852047]
10. Shade MY, Hama RS, Eisenhauer C, Khazanchi D, Pozehl B. "Ask, 'when you do this, how much pain are you in?'": content preferences for a conversational pain self-management software application. J Gerontol Nurs 2023 Jan;49(1):11-17. [doi: 0.3928/00989134-20221205-04]
11. Jones VK, Yan C, Shade MY, et al. Reducing loneliness and improving social support among older adults through different modalities of personal voice assistants. Geriatrics (Basel) 2024 Feb 22;9(2):22. [doi: 10.3390/geriatrics9020022] [Medline: 38525739]
12. Jones VK, Hanus M, Yan C, Shade MY, Blaskewicz Boron J, Maschieri Bicudo R. Reducing loneliness among aging adults: the roles of personal voice assistants and anthropomorphic interactions. Front Public Health 2021;9:750736. [doi: 10.3389/fpubh.2021.750736] [Medline: 34957013]
13. Corbett CF, Wright PJ, Jones K, Parmer M. Voice-activated virtual home assistant use and social isolation and loneliness among older adults: mini review. Front Public Health 2021;9:742012. [doi: 10.3389/fpubh.2021.742012] [Medline: 34708017]
14. Marziali RA, Franceschetti C, Dinculescu A, et al. Reducing loneliness and social isolation of older adults through voice assistants. Lit Rev and Bibl Anal J Med Internet Res 2024;26:e50534. [doi: 10.2196/50534]
15. Kocaballi AB, Berkovsky S, Quiroz JC, et al. The personalization of conversational agents in health care: systematic review. J Med Internet Res 2019 Nov 7;21(11):e15360. [doi: 10.2196/15360] [Medline: 31697237]
16. Özbek V, Alnıaçık Ü, Koc F, Akkılıç ME, Kaş E. The impact of personality on technology acceptance: a study on smart phone users. Proc Soc Behav Sci 2014 Sep;150:541-551. [doi: 10.1016/j.sbspro.2014.09.073]
17. Kortum P, Oswald FL. The impact of personality on the subjective assessment of usability. Int J Hum-Comput Interact 2018 Feb;34(2):177-186. [doi: 10.1080/10447318.2017.1336317]
18. John OP, Srivastava S. The big-five trait taxonomy: history, measurement, and theoretical perspectives. 1999 URL: https://www.academia.edu/13361041/The_Big_Five_trait_taxonomy_History_measurement_and_theoretical_perspectives
19. Redelmeier DA, Najeeb U, Etchells EE. Understanding patient personality in medical care: five-factor model. J Gen Intern Med 2021 Jul;36(7):2111-2114. [doi: 10.1007/s11606-021-06598-8] [Medline: 33506393]
20. Barańczuk U. The Five Factor Model of personality and social support: A meta-analysis. J Res Pers 2019 Aug;81:38-46. [doi: 10.1016/j.jrp.2019.05.002]
21. Schutter N, Koorevaar L, Holwerda TJ, Stek ML, Dekker J, Comijs HC. "Big Five" personality characteristics are associated with loneliness but not with social network size in older adults, irrespective of depression. Int Psychogeriatr 2020 Jan;32(1):53-63. [doi: 10.1017/S1041610219000231] [Medline: 30968789]
22. Mitzner TL, Rogers WA, Fisk AD, et al. Predicting older adults' perceptions about a computer system designed for seniors. Univers Access Inf Soc 2016 Jun;15(2):271-280. [doi: 10.1007/s10209-014-0383-y] [Medline: 31186624]
23. Mitzner TL, Boron JB, Fausset CB, et al. Older adults talk technology: technology usage and attitudes. Comput Human Behav 2010 Nov 1;26(6):1710-1721. [doi: 10.1016/j.chb.2010.06.020] [Medline: 20967133]
24. Voinescu A, Morgan PL, Alford C, Caleb-Solly P. The utility of psychological measures in evaluating perceived usability of automated vehicle interfaces – a study with older adults. Transp Res Part F: Traffic Psychol Behav 2020 Jul;72:244-263. [doi: 10.1016/j.trf.2020.05.003]
25. Borson S, Scanlan JM, Chen P, Ganguli M. The Mini-Cog as a screen for dementia: validation in a population-based sample. J Am Geriatr Soc 2003 Oct;51(10):1451-1454. [doi: 10.1046/j.1532-5415.2003.51465.x] [Medline: 14511167]
26. Ambrose KR, Golightly YM. Physical exercise as non-pharmacological treatment of chronic pain: why and when. Best Pract Res Clin Rheumatol 2015 Feb;29(1):120-130. [doi: 10.1016/j.berh.2015.04.022] [Medline: 26267006]
27. Welsh TP, Yang AE, Makris UE. Musculoskeletal pain in older adults: a clinical review. Med Clin North Am 2020 Sep;104(5):855-872. [doi: 10.1016/j.mcna.2020.05.002] [Medline: 32773050]
28. Arnstein P, Shade M, Herr KA, Young HM, Fishman SM. Managing older adults' chronic pain: lower-risk interventions. AJN Am J Nurs 2023;123(2):46-52. [doi: 10.1097/01.NAJ.0000919740.00088.93]
29. Schubart JR, Stuckey HL, Ganeshamoorthy A, Sciamanna CN. Chronic health conditions and internet behavioral interventions: a review of factors to enhance user engagement. Comput Inform Nurs 2011 Feb;29(2):81-92. [doi: 10.1097/NCN.0b013e3182065eed] [Medline: 21164337]
30. Dworschak C, Heim E, Kuhn N, Schwager J, Tröster A, Maercker A. User-centered development of an internet-based CBT intervention for the treatment of loneliness in older individuals. Internet Interv 2024 Mar;35:100720. [doi: 10.1016/j.invent.2024.100720] [Medline: 38328277]

XSL•FO
RenderX

31. Martorella G, Boitor M, Berube M, Fredericks S, Le May S, Gélinas C. Tailored web-based interventions for pain: systematic review and meta-analysis. J Med Internet Res 2017 Nov 10;19(11):e385. [doi: 10.2196/jmir.8826] [Medline: 29127076]
32. EMARKETER. Siri, Alexa, and Google Assistant Popularity Varies Across US Age Groups. Secondary siri, alexa, and google assistant popularity varies across US age groups 2024. 2021. URL: https://www.emarketer.com/content/siri-alexa-google-assistant-popularity-varies-across-us-age-groups
33. Goldberg LR. The structure of phenotypic personality traits. Am Psychol 1993 Jan;48(1):26-34. [doi: 10.1037//0003-066x.48.1.26] [Medline: 8427480]
34. Brooke J. System usability scale (SUS): a quick-and-dirty method of system evaluation user information reading, UK: digital equipment co ltd U10 - exported. 1986 URL: https://www.researchgate.net/publication/228593520_SUS_A_quick_and_dirty_usability_scale
35. Phillips LA, Burns E, Leventhal H. Time-of-day differences in treatment-related habit strength and adherence. Ann Behav Med 2021 Mar 20;55(3):280-285. [doi: 10.1093/abm/kaaa042] [Medline: 32542355]
36. Roenneberg T, Kuehnle T, Juda M, et al. Epidemiology of the human circadian clock. Sleep Med Rev 2007 Dec;11(6):429-438. [doi: 10.1016/j.smrv.2007.07.005] [Medline: 17936039]
37. Valdez P. Circadian rhythms in attention. Yale J Biol Med 2019 Mar;92(1):81-92 [FREE Full text] [Medline: 30923475]
38. COBB S, MERCHANT WR, RUBIN T. The relation of symptoms to osteoarthritis. J Chronic Dis 1957 Feb;5(2):197-204. [doi: 10.1016/0021-9681(57)90135-2] [Medline: 13405980]
39. Rizzuto D, Mossello E, Fratiglioni L, Santoni G, Wang HX. Personality and survival in older age: the role of lifestyle behaviors and health status. Am J Geriatr Psychiatry 2017 Dec;25(12):1363-1372. [doi: 10.1016/j.jagp.2017.06.008] [Medline: 28711464]
40. Chapman BP, Duberstein PR, Sörensen S, Lyness JM. Personality and perceived health in older adults: the five factor model in primary care. J Gerontol B Psychol Sci Soc Sci 2006 Nov;61(6):362-365. [doi: 10.1093/geronb/61.6.p362] [Medline: 17114306]
41. Duberstein PR, Sörensen S, Lyness JM, et al. Personality is associated with perceived health and functional status in older primary care patients. Psychol Aging 2003 Mar;18(1):25-37. [doi: 10.1037/0882-7974.18.1.25] [Medline: 12641310]
42. Kaplan AD, Sanders T, Hancock PA. The relationship between extroversion and the tendency to anthropomorphize robots: A Bayesian analysis. Front Robot AI 2018;5:135. [doi: 10.3389/frobt.2018.00135] [Medline: 33501013]
43. Epley N, Waytz A, Cacioppo JT. On seeing human: a three-factor theory of anthropomorphism. Psychol Rev 2007 Oct;114(4):864-886. [doi: 10.1037/0033-295X.114.4.864] [Medline: 17907867]
44. Poushneh A. Humanizing voice assistant: The impact of voice assistant personality on consumers' attitudes and behaviors. J Retail Consum Serv 2021 Jan;58:102283. [doi: 10.1016/j.jretconser.2020.102283]
45. Zhong R, Ma M. Effects of communication style, anthropomorphic setting and individual differences on older adults using voice assistants in a health context. BMC Geriatr 2022 Sep 15;22(1):751. [doi: 10.1186/s12877-022-03428-2] [Medline: 36104775]

Abbreviations

AI: artificial intelligence
BFI: Big Five Inventory
CVA-E: conversational voice assistant enhanced
CVA-S: conversational voice assistant standard
SUS: System Usability Scale

Edited by A Mavragani; submitted 25.07.24; peer-reviewed by E Ross, N Hird; revised version received 20.11.24; accepted 26.11.24; published 24.01.25.

Please cite as:
Shade M, Yan C, Jones VK, Boron J
Evaluating Older Adults' Engagement and Usability With AI-Driven Interventions: Randomized Pilot Study
JMIR Form Res 2025;9:e64763
URL: https://formative.jmir.org/2025/1/e64763
doi:10.2196/64763

© Marcia Shade, Changmin Yan, Valerie K Jones, Julie Boron. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete

XSL•FO
RenderX

bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Characterizing Telehealth Barriers and Preferences to Promote Acceptable Implementation Strategies in Central Uganda: Multilevel Formative Evaluation

Michael Kizito[1*], BSCS, MSCS, PhD; Erina Nabunjo Mugabi[2*], BEd, MBA; Sabrina Ford[3*], BScPT, PhD; Bree Holtz[4*], BA, MSIS, PhD; Kelly Hirko[5], BScB, MPH, PhD

[1]Department of Computer Science, Makerere University Kampala, Kampala, Uganda

[2]Department of Health Policy Planning and Management, School of Public Health, Makerere University Kampala, Kampala, Uganda

[3]College of Human Medicine, Michigan State University, East Lansing, MI, United States

[4]Department of Advertising & PR, College of Communication Arts & Sciences, Michigan State University, East Lansing, MI, United States

[5]Department of Epidemiology and Biostatistics, College of Human Medicine, Michigan State University, East Lansing, MI, United States

[*]these authors contributed equally

Corresponding Author:
Michael Kizito, BSCS, MSCS, PhD
Department of Computer Science, Makerere University Kampala, Kampala, Uganda

Abstract

Background: Telehealth approaches can address health care access barriers and improve care delivery in resource-limited settings around the globe. Yet, telehealth adoption in Africa has been limited, due in part to an insufficient understanding of effective strategies for implementation.

Objective: This study aimed to conduct a multi-level formative evaluation identifying barriers and facilitators for implementing telehealth among health service providers and patients in Central Uganda.

Methods: We collected surveys characterizing telehealth perceptions, barriers, and preferences from health care providers and patients seeking primary care in the Central Region of Uganda from January 2022 to July 2022. Survey development was informed by the technology acceptance model and evaluated predictors of technology acceptance (ie, perceived usefulness, ease of use, and attitudes). We used descriptive statistics to characterize telehealth perceptions and examined differences according to provider and patient characteristics using Student *t* tests.

Results: Nearly 79% (n=48) of 61 providers surveyed had used telehealth, and perceptions were generally favorable. While 93.4% (n=57) reported that telehealth adds value to clinical practice, less than half (n=30, 49.2%) felt telehealth was more efficient than in-person visits. Provider-reported barriers to telehealth included technology challenges for the patient (34/132, 26%), low patient engagement (25/132, 19%), and lack of implementation support (24/132, 18%). Telehealth use was lower among the 91 surveyed patients, with only 19.8% (n=18) having used telehealth. Although 89% (n=81) of patients reported saving time with telehealth approaches, 33.3% (n=30) of patients reported that telehealth made them feel uncomfortable, and 43.8% (n=39) reported concerns about confidentiality. Over 72% (n=66) of patients who had used telehealth previously reported satisfaction with the telehealth services they received. Several differences in perceptions of telehealth according to patient's self-reported health status were observed.

Conclusions: Perceptions of telehealth were generally favorable, although higher among providers than patients. Barriers impeding telehealth use include technology challenges and the lack of infrastructure and implementation support. Findings from this study can inform the implementation of acceptable telehealth approaches to address disparities propagated by health care access barriers in Sub-Saharan Africa.

(JMIR Form Res 2025;9:e60843) doi:10.2196/60843

KEYWORDS

telehealth; telemedicine; health care; disparities; technology; barriers; resource-limited; preferences; Uganda; Africa; barrier; formative evaluation; health service provider; primary care; satisfaction; Sub-Saharan Africa; survey; utility

XSL•FO
RenderX

Introduction

Access to quality health care services is critical to achieving optimal health outcomes and reducing health disparities [1]. Individuals residing in resource-limited settings often encounter various economic, social, cultural, and geographic barriers to health care, including provider shortages, inadequate health insurance coverage, transportation challenges, and limited health literacy [2-4]. Health care access barriers are especially stark in many countries in Sub-Saharan Africa, which have 0.23 doctors for every 10,000 people, compared to the highest ratio of 84.2 doctors in the most high-income countries [5]. Increasing access to quality health care and preventive services via telehealth is a promising strategy to reduce health disparities in low- and middle-income countries.

Telehealth refers broadly to the remote delivery of health care services using electronic and telecommunications technology, including voice, chat, and video-based platforms. There are numerous types of telehealth applications, such as live video conferencing for the provision of clinical services to patients (ie, telemedicine), provider-to-provider health care consultation, and remote delivery of nonclinical services, including the use of mobile health (mHealth) tools to support health promotion and wellbeing efforts [6]. Telehealth approaches offer a practical and often cost-effective method to expand the reach of health care services and address health care access barriers in resource-limited settings [7]. Despite the widespread adoption of mobile phones and other digital tools that can support telehealth applications, and the acknowledged potential of using these technologies to improve health care service delivery [8], telehealth adoption in clinical settings remains low. This is particularly true in health care professional shortage areas, which arguably stand to benefit most from these technologies. Indeed, results from a recent review demonstrate the promise of telemedicine approaches in revolutionizing health care delivery in Africa by improving access, efficiency, and patient outcomes [9]. Yet, significant challenges related to the effective implementation and adoption of telehealth remain.

The COVID-19 pandemic highlighted the need for innovative, cost-effective, technology-enabled systems for health care delivery and resulted in a rapid increase in telehealth adoption worldwide. Indeed, the rise in innovative telehealth start-up companies is beginning to transform health care access across Africa, despite concerns about data privacy, inadequate infrastructure, and a lack of policy frameworks [9,10]. For example, Rocket Health based in Kampala, Uganda [11], offers telehealth services as a licensed and registered clinic, laboratory, and pharmacy, including doctor consultations through phone calls and SMS text messages, as well as laboratory sample pickups and medicine deliveries. However, these services can be costly and may not be accessible to those residing in rural areas of Africa with limited internet access. Thus, continued efforts are urgently needed to ensure the broad and equitable reach of telehealth applications in Africa.

An insufficient understanding of effective strategies for implementation has limited telehealth adoption in under-resourced settings. Indeed, very little is known about specific challenges and preferences for telehealth program implementation in African countries, particularly in Uganda [12]. This dearth of knowledge severely limits the ability to implement effective and appropriate telehealth solutions to improve health care access in Africa equitably. Thus, the purpose of this study was to conduct a multi-level formative evaluation identifying barriers and facilitators for implementing telehealth among health service providers and patients in Central Uganda.

Methods

Study Population

The study population included health service providers (ie, doctors, nurses, midwives, and lab technicians) and patients seeking health care services at regularly scheduled clinic visits between January and July 2022 at the following health care facilities in the Central Region of Uganda: Namulonge Health Centre III, Wattuba Health Centre III, Kasangati Health Centre IV, and Buwambo Health Centre IV. Uganda's health care system works on a referral basis and includes small village health teams, health centers (II, III, and IV), district general hospitals, and regional and national referral hospitals, with more intensive services provided at larger centers or hospitals [13]. A health center III facility has approximately 18 staff, including a senior clinical officer, and provides laboratory services and a general outpatient clinic and maternity ward. Health centers IV are small hospitals, with all of the services available at health center III, along with additional medical providers, the ability to admit patients, and a theater for surgical procedures. Although phone and internet access is widespread in Uganda's health center IV facilities, telemedicine integration is minimal, and technology has been used largely in referral hospital settings to initiate referrals and consultations and for knowledge sharing between providers [14]. Of note, the adoption of telecommunication technologies in Uganda is relatively low, with mobile phone coverage at approximately 60% and mobile internet subscriptions at around 9% in 2016, and primarily among urban residents [15]. Although marked variations in internet coverage across different regions in Uganda are evident, the vast majority of the population in Central Uganda has access to at least 2G networks [16].

Study Design

We used a cross-sectional study design engaging providers and patients through surveys using a convenience sampling approach. Inclusion criteria required participants to be providers at the study sites or patients receiving health care at the study sites at the time of study recruitment. Study participants were not selected based on prior telehealth use. Recruitment was facilitated through partnership with the Wakiso District Health Officer (DHO). The research team met with the DHO to describe the study objectives, and the DHO drafted a letter of Introduction to the Health centers for the research team to initiate contact with health centers in the study region. The research team then met with the medical superintendents at health centers in the Central Region of Uganda (Kasangati HCIV, Namulonge Health Centre III, Wattuba Health Centre III, and Buwambo Centre HCIV) to provide the introductory letter from the DHO and to

XSL•FO
RenderX

schedule site visits with the designated nurse at each site to determine the best approaches for distributing surveys at each site. Research team members distributed paper copies of the patient and provider surveys at each site visit with the designated nurse. Surveys were distributed to all patients during health care encounters at the health centers during the research team's site visit. The designated nurse at each site helped distribute and collect the surveys from providers. Through our existing contacts, digital surveys were also emailed to service providers from academia, health IT professionals, and the Infectious Diseases Institute (IDI) from January through June 2022. Given the descriptive nature of this cross-sectional study, we did not conduct power analyses but aimed to have a minimum sample of 50 participants in each group to ensure sufficient data for descriptive analyses.

Ethical Considerations

This study was reviewed for human subject research ethics and was approved as exempt by the Institutional Review Board at Michigan State University (Study ID 00006715). An informed consent waiver, including a description of the study purpose and contact information for the study investigators, was provided on the study surveys. The consent waiver on the surveys specified that participation was voluntary and that responses were confidential. Surveys were anonymous and did not collect any personally identifiable information. Participants were not compensated for survey completion.

Data Collection

Study surveys were developed for patients and providers. The patient survey questions were based on the validated Service User Technology Acceptability Questionnaire, which measures user perceptions about the acceptability of telehealth in domains of enhanced care, increased acceptability, privacy and discomfort, care personnel concerns, telehealth as a substitution, and satisfaction [17]. The provider survey was based on a questionnaire developed by the American Medical Association to assess telehealth preferences and impact on clinical practice, including perceptions of clinical value and efficiency of telehealth services [18]. This modified questionnaire was deployed in our previous study assessing telehealth satisfaction among providers in rural Northwest Michigan, United States of America [19]. Informed by the Technology Acceptance Model [20], the surveys also assessed the following factors influencing an individual's intention to use new technology: perceived usefulness and perceived ease of use. Survey respondents provided perceptions of telehealth using a 5-point Likert scale, where 1 indicated "strongly agree" and 5 indicated "strongly disagree." Additionally, the survey ascertained demographic information, including age range (<30, 30 - 49, 50 - 65, and 65+ years), gender, length of time in practice (providers only), and self-reported health status (patients only). The survey also asked whether respondents had heard about telemedicine and used telemedicine prior to completing the survey. The provider survey also ascertained information on barriers to telehealth using the question: "What are your biggest challenges, if any, that you face regarding telehealth visits? Select all that apply." The following options were available for selection, lack of reimbursement, licensure, technology challenges for the patients, technology challenges for the health care provider/practice, low patient engagement, lack of implementation support, no challenges, or other. This list was based on a prior review of barriers to telehealth in Sub-Saharan Africa [21], with reimbursement issues reflecting barriers due to differential financial return for telehealth visits and licensure indicating legal and regulatory barriers to telehealth implementation. Survey data were entered into Qualtrics, either directly by participants completing the online survey or by study staff for paper-based surveys.

Statistical Analysis

The survey data were downloaded from Qualtrics and collated in an Excel spreadsheet for analysis. We conducted descriptive statistics of the survey items, describing participant characteristics, perceptions of telehealth, and barriers to using telehealth among providers. Descriptive statistics were calculated to describe provider-reported barriers to telehealth. We also assessed whether telehealth perceptions varied according to participant age, and previous telehealth use, length of time practicing medicine (providers only), and health status (patients only) using Student *t* tests. In secondary analyses, we assessed patient perceptions among the subset of individuals who had used telehealth previously. Missing data were excluded from analysis. All statistical analyses were performed using SAS (version 9.4; SAS Institute) and statistical significance was defined at $P<.05$.

Results

Surveys from 61 providers and 91 patients were completed and included in the analysis. For providers, 41 surveys were completed on paper and 20 were completed online. For patients, all 91 surveys were completed on paper. As shown in Table 1, most providers were between the ages of 30 - 49 years (n=43, 71.1%) and reported female gender (n=38, 63.3%). Providers varied in the length of time in practice, with 20% (n=12) of respondents practicing less than 5 years, 43.3% (n=25) practicing between 5 and 10 years, and 36.7% (n=22) practicing more than 10 years. Most of the surveyed providers had heard about telemedicine before (n=42, 68.9%) and used telemedicine previously (n=48, 78.7%). Of note, several providers (n=6) noted that they had not heard of telemedicine previously but reported using telemedicine after reading the description that telemedicine included telephone-based visits. Patient survey respondents were younger than providers, with 57.8% (n=52) of patients under the age of 30 years. Most patient participants were female (n=67, 74.4%) and reported health status as "Excellent" or "Very good" (n=26, 28.9% and n=23, 25.6%, respectively). Compared to providers, a lower percentage of patients had heard of telemedicine (n=32, 35.2%) or used telemedicine previously (n=18, 19.8%).

XSL•FO
RenderX

Table . Characteristics of provider and patient survey respondents. Missing data on age, gender, length of time in practice and self-reported health status for 1 patient survey.

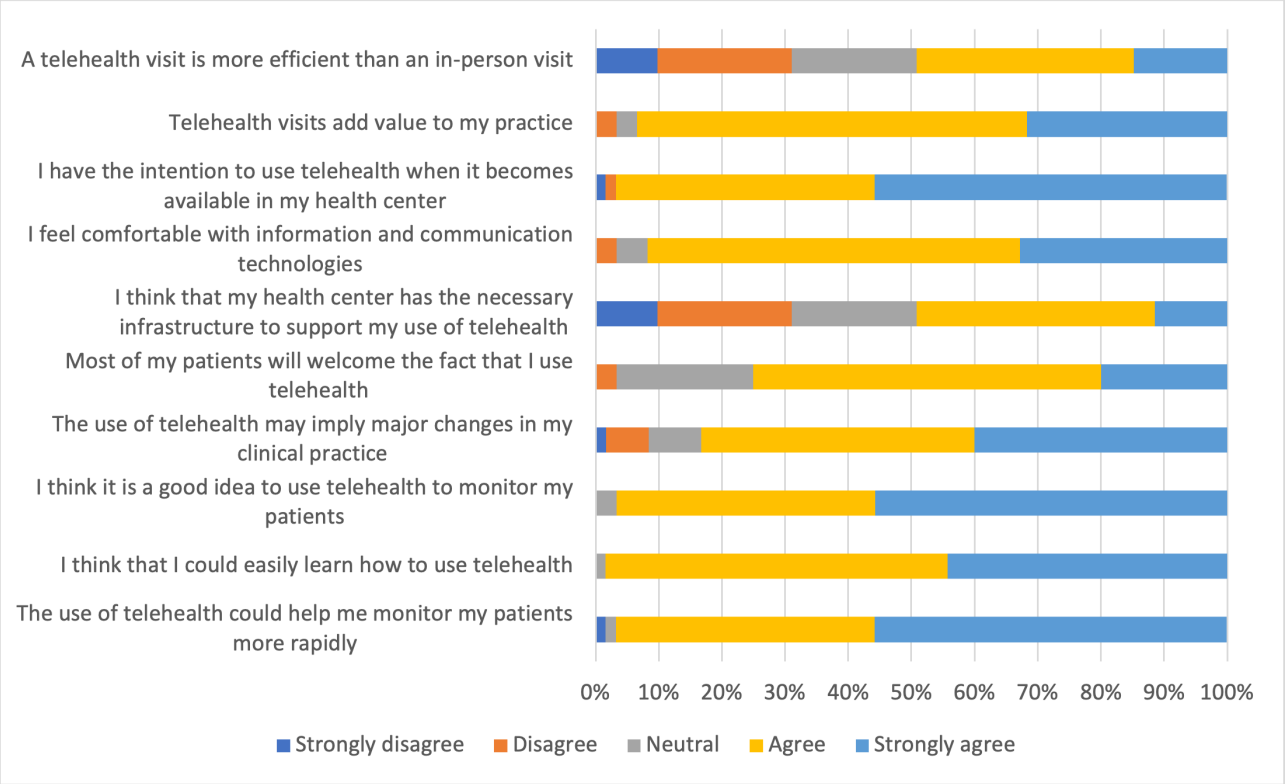
	Providers (n=61), n (%)	Patients (n=91), n (%)
Age, years		
<30	10 (16.7)	52 (57.8)
30 - 49	43 (71.7)	22 (24.4)
50 - 65	6 (10)	10 (11.1)
>65	1 (1.7)	6 (6.7)
Gender		
Female	38 (63.3)	67 (74.4)
Male	22 (36.7)	23 (25.6)
Length of time in practice, years		
<5	12 (20)	—[a]
5 - 10	25 (43.3)	—
11 - 20	16 (26.7)	—
>20	6 (10.0)	—
Self-reported health status		
Excellent	—	26 (28.9)
Very good	—	23 (25.6)
Good	—	17 (18.9)
Fair	—	20 (22.2)
Poor	—	4 (4.4)
Heard about telemedicine previously		
Yes	42 (68.9)	32 (35.2)
No	19 (31.2)	59 (64.8)
Used telemedicine previously		
Yes	48 (78.7)	18 (19.8)
No	13 (21.3)	73 (80.2)

[a]Not applicable

As shown in Figure 1, provider perceptions of telehealth were generally positive, with 96.7% (n=59) agreeing or strongly agreeing that telehealth could help them monitor patients more rapidly and 98.4% (n=60) reporting that they could easily learn how to use telehealth. While 93.4% (n=57) reported that telehealth adds value to clinical practice, less than half (n=30, 49.2%) felt that telehealth was more efficient than in-person visits. Most providers (n=56, 91.8%) reported feeling comfortable with information and communication technologies, but only 49.2% (n=30) thought their health center had the necessary infrastructure to support telehealth use. Indeed, 83.3% (n=51) of providers reported that telehealth use would imply major changes in their clinical practice. Finally, all 48 providers who had used telehealth would recommend telehealth services to others, and all 13 providers who had not used telehealth would be interested in using telehealth services in the future.

XSL•FO
RenderX

Figure 1. Provider perceptions of telehealth from a cross-sectional survey conducted in Central Uganda, January 2022-July 2022 (n=61).



Perceptions of telehealth were less favorable among patients compared to providers (Figure 2). Nearly 78% (n=70) agreed or strongly agreed that telehealth services should be recommended to people in a similar condition, and 76% (n=68) agreed or strongly agreed that telehealth services would allow their caretakers to better monitor their health condition. Although approximately 33% (n=30) of patient survey respondents agreed or strongly agreed that telehealth services have made them feel uncomfortable, and 43.8% (n=39) agreed or strongly agreed that telehealth services make them worried about the confidentiality of private information. Roughly half of the respondents (n=46, 51.1%) did not agree that telehealth could replace regular care, and 43.4% (n=39) agreed or strongly agreed that telehealth services are not as suitable as regular face-to-face consultations. In secondary analyses among the subset of patients who had used telehealth previously (n=18), favorable perceptions were more pronounced, with nearly 90% (n=82) reporting that telehealth services increased their access to care, 89% (n=81) reporting that telehealth saved time, and 88% (n=80) reporting that telehealth services have helped them improve their health. Over 72% (n=66) of patients who had used telehealth previously reported satisfaction with the telehealth services they received. All 18 patients who had used telehealth would recommend telehealth services to others. However, only 27.8% (n=25) of those who had used telehealth agreed or strongly agreed that telehealth services allowed them to be less concerned about their health status (data not shown). Of the 73 patients who had not used telehealth, 60 (82.2%) would be interested in using telehealth services in the future.

Provider-reported barriers to telehealth, shown in Table 2, included technology challenges for the patient (34/132, 26%), low patient engagement (25/132, 19%), and lack of implementation support (24/132, 18%). Lack of reimbursement (19/132, 14%) and technology challenges for the health care provider or practice (19/132, 14%) were also noted.

XSL•FO
RenderX

Figure 2. Patient perceptions of telehealth from a cross-sectional survey conducted in Central Uganda, January 2022-July 2022 (n=91).

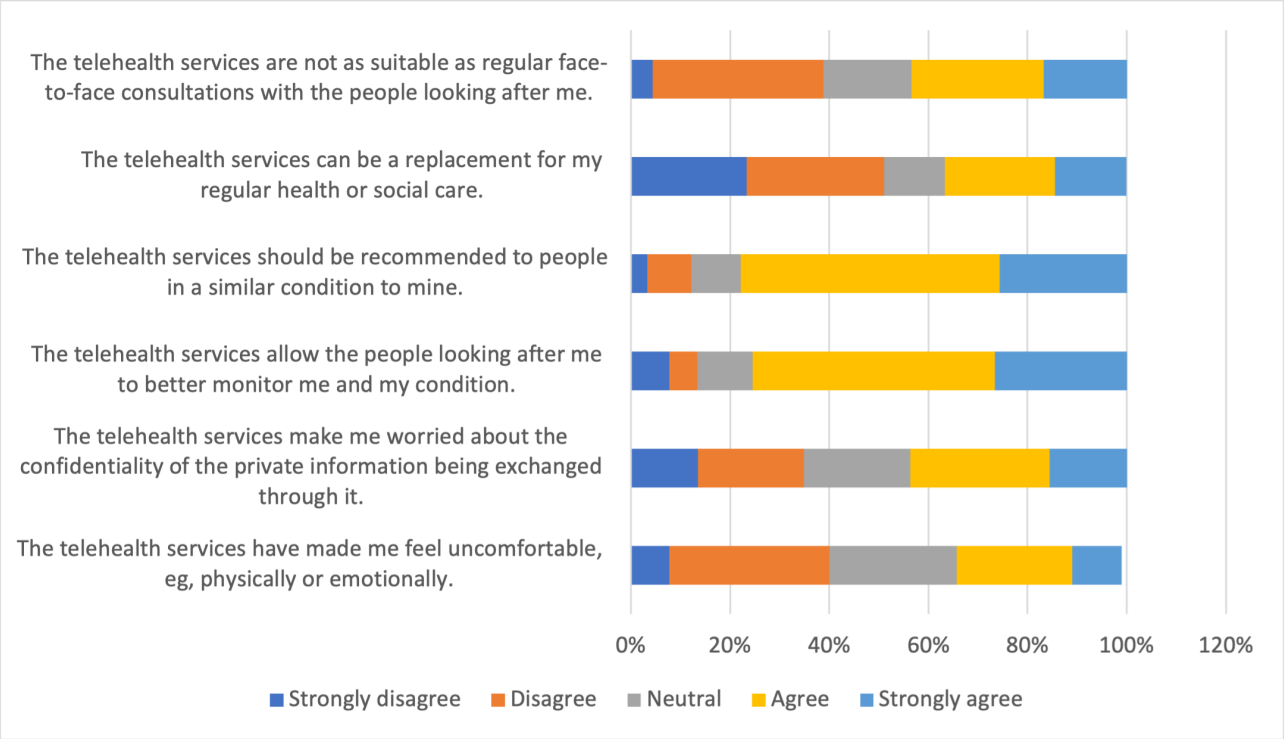


Table . Provider-reported barriers to telehealth from a cross-section survey conducted in Central Uganda, January 2022-July 2022.

	Number of provider-reported barriers (n=132), n (%)
Lack of reimbursement	19 (14)
Licensure	5 (4)
Technology challenges for the patient	34 (26)
Technology challenges for the health care provider or practice	19 (14)
Low patient engagement	25 (19)
Lack of implementation support	24 (18)
No challenges	2 (2)
Other	4 (3)

Overall, provider perceptions of telehealth did not differ according to age, gender, length of time practicing medicine, or based on prior telehealth use (Table 3). However, intentions to use telehealth when it becomes available were slightly higher among male versus female providers (4.73 vs 4.34 on Likert 1-5 scale; *P*=.03). Conversely, several differences in perceptions of telehealth according to patient characteristics were observed (Table 4). For example, concerns about confidentiality with telehealth were slightly higher among patients who did not use telehealth compared to users (3.29 vs 2.35; *P*=.01). Additionally, patients with better self-reported health status were more likely to agree that telehealth should be recommended to people in a similar condition to theirs compared to those with poorer health status (4.14 vs 3.53; *P*=.003). Although not statistically significant, males were slightly more likely than females to report feeling uncomfortable with using telehealth (3.30 vs 2.80; *P*=.06). Interest in participating in telehealth visits in the future was higher among younger (<50 y) versus older participants (3.22 vs 2.56; *P*=.07), among females versus males (4.08 vs 3.43; *P*=.05) and among those with better versus worse self-reported health status (4.12 vs 3.64; *P*=.04).

XSL•FO
RenderX

Table . Provider perceptions overall and according to provider characteristics.

	Overall (n=61), mean (SD)	Age			Sex			Length of time practicing			Telehealth use		
		<50 years (n=53), mean (SD)	≥50 years (n=7), mean (SD)	*P* value	Male (n=22), mean (SD)	Female (n=38), mean (SD)	*P* value	≤10 years (n=37), mean (SD)	>10 years (n=22), mean (SD)	*P* value	Used telehealth (n=48), mean (SD)	Did not use telehealth (n=13), mean (SD)	*P* value
The use of telehealth could help me monitor my patients more rapidly	4.49 (0.7)	4.47 (0.7)	4.57 (0.5)	.73	4.41 (0.6)	4.53 (0.8)	.54	4.58 (0.6)	4.32 (0.9)	.22	4.52 (0.7)	4.38 (0.5)	.54
I think that I could easily learn how to use telehealth	4.43 (0.5)	4.42 (0.5)	4.57 (0.5)	.47	4.45 (0.5)	4.42 (0.6)	.82	4.45 (0.6)	4.41 (0.5)	.79	4.38 (0.5)	4.62 (0.51)	.15
I think it is a good idea to use telehealth to monitor my patients	4.52 (0.6)	4.49 (0.6)	4.71 (0.5)	.33	4.55 (0.5)	4.50 (0.6)	.77	4.47 (0.6)	4.59 (0.5)	.45	4.48 (0.6)	4.69 (0.5)	.23
The use of telehealth may imply major changes in my clinical practice	4.13 (0.9)	4.15 (0.9)	4.43 (0.5)	.43	4.14 (0.9)	4.22 (0.8)	.73	4.24 (0.8)	4.09 (0.9)	.52	4.09 (1.0)	4.31 (0.9)	.65
Most of my patients will welcome the fact that I use telehealth	3.92 (0.7)	3.94 (0.8)	3.71 (0.8)	.45	3.82 (0.7)	3.97 (0.8)	.45	3.84 (0.8)	4.05 (0.7)	.32	4.00 (0.8)	3.62 (0.5)	.10

XSL•FO
RenderX

	Overall (n=61), mean (SD)	Age			Sex			Length of time practicing			Telehealth use		
		<50 years (n=53), mean (SD)	≥50 years (n=7), mean (SD)	*P* value	Male (n=22), mean (SD)	Female (n=38), mean (SD)	*P* value	≤10 years (n=37), mean (SD)	>10 years (n=22), mean (SD)	*P* value	Used telehealth (n=48), mean (SD)	Did not use telehealth (n=13), mean (SD)	*P* value
I think that my health center has the necessary infrastructure to support my use of telehealth	3.20 (1.2)	3.19 (1.2)	3.14 (1.1)	.93	3.18 (1.2)	3.18 (1.2)	.99	3.13 (1.3)	3.27 (1.1)	.66	3.27 (1.2)	2.92 (1.2)	.36
I feel comfortable with information and communication technologies	4.21 (0.7)	4.21 (0.7)	4.29 (0.8)	.78	4.18 (0.7)	4.24 (0.7)	.77	4.29 (0.7)	4.09 (0.7)	.29	4.23 (0.7)	4.15 (0.6)	0.73
I have the intention to use telehealth when it becomes available in my health center	4.48 (0.7)	4.43 (0.8)	4.86 (0.4)	.16	4.73 (0.5)	4.34 (0.8)	.03	4.53 (0.6)	4.41 (0.9)	.56	4.46 (0.7)	4.54 (0.9)	.73
Telehealth visits add value to my practice	4.22 (0.7)	4.21 (0.7)	4.29 (0.5)	.79	4.18 (0.7)	4.24 (0.6)	.74	4.13 (0.7)	4.38 (0.6)	.17	4.21 (0.7)	4.23 (0.8)	.93
A telehealth visit is more efficient than an in-person visit	3.23 (1.2)	3.28 (1.2)	3.00 (1.4)	.57	2.95 (1.2)	3.42 (1.2)	.16	3.21 (1.2)	3.32 (1.3)	.75	3.23 (1.2)	3.23 (1.4)	.99

XSL•FO
RenderX

Table . Patient perceptions overall and according to patient characteristics.

	Overall (n=91), mean (SD)	Age			Sex			Health status			Telehealth use		
		<50 years (n=74), mean (SD)	≥50 years (n=16), mean (SD)	*P* value	Male (n=23), mean (SD)	Female (n=67), mean (SD)	*P* value	<Very good (n=41), mean (SD)	≥Very good (n=49), mean (SD)	*P* value	Used telehealth (n=18), mean (SD)	Did not use telehealth (n=73), mean (SD)	*P* value
The telehealth services have made me feel uncomfortable, eg, physically, or emotionally.	2.96 (1.1)	3.74 (1.1)	3.75 (1.0)	.97	3.30 (1.1)	2.80 (1.1)	.06	2.98 (1.0)	2.90 (1.2)	.75	2.56 (1.1)	3.06 (1.1)	.09
The telehealth services make me worried about the confidentiality of the private information being exchanged through it.	3.11 (1.3)	3.94 (0.9)	3.50 (0.9)	.08	3.09 (1.0)	3.11 (1.4)	.96	2.90 (1.4)	3.27 (1.2)	.18	2.35 (1.1)	3.29 (1.3)	.01
The telehealth services allow the people looking after me to better monitor me and my condition.	3.81 (1.1)	3.88 (1.0)	3.56 (0.8)	.24	4.00 (1.1)	3.76 (1.1)	.38	3.70 (1.2)	3.92 (1.1)	.37	4.06 (1.2)	3.75 (1.1)	.31
The telehealth services should be recommended to people in a similar condition to mine.	3.88 (1.0)	2.86 (1.1)	3.25 (0.9)	.21	3.65 (1.1)	3.94 (1.0)	.24	3.53 (1.1)	4.14 (0.9)	.003	4.11 (0.7)	3.82 (1.1)	.16

XSL•FO
RenderX

	Overall (n=91), mean (SD)	Age			Sex			Health status			Telehealth use		
		<50 years (n=74), mean (SD)	≥50 years (n=16), mean (SD)	*P* value	Male (n=23), mean (SD)	Female (n=67), mean (SD)	*P* value	<Very good (n=41), mean (SD)	≥Very good (n=49), mean (SD)	*P* value	Used telehealth (n=18), mean (SD)	Did not use telehealth (n=73), mean (SD)	*P* value
The telehealth services can be a replacement for my regular health or social care.	2.77 (1.4)	3.14 (1.3)	3.38 (0.9)	.48	2.96 (1.4)	2.68 (1.4)	.42	2.75 (1.4)	2.76 (1.4)	.99	3.11 (1.6)	2.68 (1.4)	.25
The telehealth services are not as suitable as regular face to face consultations with the people looking after me.	3.17 (1.2)	3.88 (0.9)	3.50 (0.9)	.15	3.26 (1.3)	3.14 (1.2)	.67	3.28 (1.1)	3.08 (1.3)	.46	2.89 (1.3)	3.24 (1.2)	.28
I would be interested in continuing to participate in telehealth visits in the future.	3.92 (1.1)	3.22 (1.3)	2.56 (1.3)	.07	3.43 (1.1)	4.08 (0.9)	.05	3.64 (1.1)	4.12 (1.0)	.04	4.11 (1.0)	3.87 (1.1)	.42

Discussion

Findings from this study demonstrate generally favorable perceptions of telehealth among providers and patients in Central Uganda. Providers were more likely than patients to have used telehealth, and most noted the clinical value of telehealth and reported feeling comfortable using the technology. Importantly, this study also identified provider barriers impeding telehealth use, including technology challenges for the patient and the lack of infrastructure and implementation support. Patient perceptions of telehealth were largely neutral to positive, and concerns around confidentiality of information and comfort level using telehealth services were reported. Overall, our findings suggest the strong potential of telehealth approaches to improve health care service delivery in Central Uganda. Our results demonstrate the need for future telehealth implementation efforts that address technology and infrastructure-related challenges.

Overall, the favorable perceptions of telehealth among providers in this study align with results from another cross-sectional study in Uganda, where attitudes toward telemedicine were overwhelmingly positive, yet gaps in technology readiness were pronounced [22]. Results from the providers in this study demonstrate evidence for perceived usefulness of telehealth in monitoring patient care and providing clinical value to patients. These findings align with results from prior studies [23-25], including our prior study of rural providers in the United States

XSL•FO
RenderX

[19]. Additionally, most providers in this study reported ease of use in telehealth technology platforms. However, concerns around the efficiency of telehealth visits compared to in-person clinic visits were also commonly reported. Concerns around the efficiency of telehealth approaches were also noted in our team's prior study of rural providers in the United States [19]. These findings differ, however, from results in other study settings where improvements in efficiency with telehealth approaches were observed [23,24]. The lack of supportive infrastructure for telehealth visits, which was noted as an important barrier to telehealth among survey respondents, may contribute to the discrepancies in findings across studies. Indeed, most providers in this study reported that telehealth approaches would require major changes in current clinical practice and workflows. Such changes may help improve overall efficiencies with the telehealth system and overcome this reported barrier.

In this study, patient perceptions of telehealth were less favorable compared to providers. This may be due in part to lower levels of telehealth utilization and comfort in using the technology among patients. Indeed, while most providers in this study had used telehealth, both awareness and telehealth use among patients were much lower. This suggests the need to better understand patient-level barriers to telehealth use and to target patient education efforts to increase overall knowledge about telehealth applications from the patient perspective. Importantly, our findings also demonstrate satisfaction among patients who had previously used telehealth, and most previous users also recognized the clinical utility of telehealth for care monitoring. However, there were concerns among patients related to the comfort of using telehealth and the confidentiality of the private information shared through the technology. These findings emphasize the need for future efforts to increase awareness of telehealth services, offer educational trainings on digital literacy, and address concerns related to the confidentiality of information.

Perceptions of telehealth in this study did not differ substantially by most provider and patient characteristics. Interestingly, we did not observe any significant differences in telehealth perceptions according to provider or patient age. This differs from findings in U.S. study populations, where telehealth perceptions were more favorable among younger patients and providers [19,26,27]. In these prior studies, age-related differences in telehealth perceptions were largely attributed to lower technology access and digital literacy among older adults [28]. Participants in this study were younger overall, and this may have contributed to the lack of observed differences in telehealth perceptions by age in our study. Intentions to use telehealth in clinical practice were modestly higher among male compared to female providers in our study. This finding differs from those assessing mental health providers in the United States, where no gender differences in intention to continue using telehealth after the COVID-19 pandemic were observed [29]. It is important to note that the intention to use telehealth was high among female and male providers in our study. Thus, future studies are needed to assess whether this gender difference persists in other study settings. Interestingly, patients with better self-reported health status were more favorable about recommending telehealth to people in similar conditions and were also more interested in continuing to participate in telehealth visits in the future. These findings highlight the growing recognition that telehealth may not be appropriate or feasible for all patients or all clinical care situations [30].

The strengths of this study include the focus on multiple perspectives, including those of both providers and patients across multiple clinical sites in this understudied region. Moreover, our study used a community-engaged approach, working in partnership with community members and incorporating input on study processes and outcomes from clinical and community partners. This approach can help guide implementation efforts toward those with the most community relevance and potential impact. We also held dissemination workshops in the community to discuss the findings and inform next steps to ensure that research efforts address the relevant needs and preferences of the community members in this region of Uganda. This research provides a replicable and scalable model that can help inform telehealth implementation efforts in other under-resourced settings. This study did have several limitations. First, using a convenience sampling strategy, we were unable to calculate a response rate given resource limitations for tracking survey distribution across multiple study sites. While the survey development was informed by the technology acceptance model [20] and modeled on surveys used in prior studies [31,32], the survey questions were not specifically validated. Additionally, we were unable to assess whether perceptions of telehealth differed according to the type of health care provider, as this information was not collected on the provider survey. This should be assessed in future studies, particularly given that readiness to integrate telemedicine services varied across health facility types and job roles in a prior study conducted in Uganda [14]. Finally, this study focuses on a single region in Uganda, and results may not be broadly generalizable across other populations and settings.

Overall, our findings demonstrate both satisfaction with telehealth among users and willingness to use it among nonusers for both providers and patients. Indeed, providers and patients who previously used telehealth unanimously reported recommending the service to others. All surveyed providers and most patients who had not used telehealth were interested in using this approach in the future. These findings suggest positive attitudes and behavioral intentions to use telehealth, which are key components to technology acceptance according to the technology acceptance model [20]. Taken together, our results suggest that the perceived usefulness, ease of use, and attitudes toward telehealth are supportive of future implementation efforts to increase the reach of telehealth services in Uganda and in other low- and middle-income countries.

Findings from this study can inform the implementation of acceptable telehealth approaches to address disparities propagated by health care access barriers in Sub-Saharan Africa.

XSL•FO
RenderX

Acknowledgments

Funding for this study was received from Michigan State University, Alliance for African Partnerships: Partnerships for Innovative Research in Africa, Early Career Investigators Award (MK and KH).

Data Availability

The datasets generated during this study are not publicly available due to participants not consenting to the sharing of data beyond the study team.

Conflicts of Interest

None declared.

References

1. Institute of Medicine (US) Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care. In: Smedley BD, Stith AY, Nelson AR, editors. Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care: National Academies Press; 2003.
2. Douthit N, Kiv S, Dwolatzky T, Biswas S. Exposing some important barriers to health care access in the rural USA. Pub Health (Fairfax) 2015 Jun;129(6):611-620. [doi: 10.1016/j.puhe.2015.04.001]
3. Syed ST, Gerber BS, Sharp LK. Traveling towards disease: transportation barriers to health care access. J Community Health 2013 Oct;38(5):976-993. [doi: 10.1007/s10900-013-9681-1]
4. Nutbeam D, Lloyd JE. Understanding and responding to health literacy as a social determinant of health. Annu Rev Public Health 2021 Apr 1;42:159-173. [doi: 10.1146/annurev-publhealth-090419-102529] [Medline: 33035427]
5. Hsiao A, Vogt V, Quentin W. Effect of corruption on perceived difficulties in healthcare access in sub-Saharan Africa. PLoS ONE 2019;14(8):e0220583. [doi: 10.1371/journal.pone.0220583] [Medline: 31433821]
6. Kamulegey L, Ssebwana J, Abigaba W, Bwanika J, Musinguzi D. Mobile health in Uganda: a case study of the Medical Concierge Group. EASci 2019 Mar 25;1(1):9-14. [doi: 10.24248/easci.v1i1.12]
7. LeRouge C, Garfield MJ. Crossing the telemedicine chasm: have the U.S. barriers to widespread adoption of telemedicine been significantly reduced? Int J Environ Res Public Health 2013 Nov 28;10(12):6472-6484. [doi: 10.3390/ijerph10126472] [Medline: 24287864]
8. Hoque MR, Bao Y, Sorwar G. Investigating factors influencing the adoption of e-Health in developing countries: a patient's perspective. Inf Health Soc Care 2017 Jan 2;42(1):1-17. [doi: 10.3109/17538157.2015.1075541]
9. Mahmoud Z, Sanusi M, Nartey C, Adedinsewo D. Using technology to deliver cardiovascular care in African countries. Curr Cardiol Rep 2023 Dec;25(12):1823-1830. [doi: 10.1007/s11886-023-01988-2] [Medline: 37486571]
10. Ogodo O. Dozens of telehealth companies are innovating in Africa. N Med 2023 Feb;29(2):281-282. [doi: 10.1038/s41591-022-02089-3] [Medline: 36759675]
11. RocketHealth. URL: https://rockethealth.africa/ [accessed 2022-08-08]
12. Isabalija D, Mayoka KG, Rwashana DA, Mbarika PVW. Factors affecting adoption, implementation and sustainability of telemedicine information systems in Uganda. J Health Inform Dev Ctries 2011;5(2).
13. Medard T, Yawe BL, Bosco OJ. Health care delivery system in Uganda: a review. Tanzan J Health Res 2023;24(2):57. [doi: 10.4314.thrb.v24i2.5]
14. Kiberu VM, Scott RE, Mars M. Assessment of health provider readiness for telemedicine services in Uganda. Health Inf Manag 2019 Jan;48(1):33-41. [doi: 10.1177/1833358317749369] [Medline: 29359588]
15. Paul KJ, Munuulo J, Nagujja S. Empowerment digital divide: case of internet adoption in Uganda's rural-urban areas. Adv Soc Sci Res J 2015 Nov;2(11):11. [doi: 10.14738/assrj.211.1567]
16. Uganda: driving inclusive socio-economic progress through mobile-enabled digital transformation. Global System for Mobile Communications Association. 2019. URL: https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-for-development/wp-content/uploads/2019/03/GSMA_Connected_Society_Uganda_Overview.pdf [accessed 2025-01-17]
17. Hirani SP, Rixon L, Beynon M, et al. Quantifying beliefs regarding telehealth: development of the whole systems demonstrator service user technology acceptability questionnaire. J Telemed Telecare 2017 May;23(4):460-469. [doi: 10.1177/1357633X16649531] [Medline: 27224997]
18. AMA digital health research: physicians' motivations and requirements for adopting digital health. American Medical Association. 2020 Feb. URL: https://www.ama-assn.org/system/files/2020-02/ama-digital-health-study.pdf [accessed 2025-01-20]
19. Klee D, Pyne D, Kroll J, James W, Hirko KA. Rural patient and provider perceptions of telehealth implemented during the COVID-19 pandemic. BMC Health Serv Res 2023 Sep 12;23(1):981. [doi: 10.1186/s12913-023-09994-4] [Medline: 37700286]

XSL•FO
RenderX

20. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
21. Dodoo JE, Al-Samarraie H, Alzahrani AI. Telemedicine use in Sub-Saharan Africa: barriers and policy recommendations for Covid-19 and beyond. Int J Med Inform 2021 Jul;151:104467. [doi: 10.1016/j.ijmedinf.2021.104467]
22. Kiberu VM, Scott RE, Mars M. Assessing core, e-learning, clinical and technology readiness to integrate telemedicine at public health facilities in Uganda: a health facility - based survey. BMC Health Serv Res 2019 Apr 29;19(1):266. [doi: 10.1186/s12913-019-4057-6] [Medline: 31035976]
23. Beiser M, Lu V, Paul S, et al. Electronic health record usage patterns: assessing telemedicine's impact on the provider experience during the COVID-19 pandemic. Telemed E-Health 2021 Aug 1;27(8):934-938. [doi: 10.1089/tmj.2020.0490]
24. Saiyed S, Nguyen A, Singh R. Physician perspective and key satisfaction indicators with rapid telehealth adoption during the coronavirus disease 2019 pandemic. Telemed J E Health 2021 Nov;27(11):1225-1234. [doi: 10.1089/tmj.2020.0492] [Medline: 33513045]
25. Gentry MT, Puspitasari AJ, McKean AJ, et al. Clinician satisfaction with rapid adoption and implementation of telehealth services during the COVID-19 pandemic. Telemed J E Health 2021 Dec;27(12):1385-1392. [doi: 10.1089/tmj.2020.0575] [Medline: 33606560]
26. Ladin K, Porteny T, Perugini JM, et al. Perceptions of telehealth vs in-person visits among older adults with advanced kidney disease, care partners, and clinicians. JAMA Netw Open 2021 Dec 1;4(12):e2137193. [doi: 10.1001/jamanetworkopen.2021.37193] [Medline: 34870680]
27. Fischer SH, Ray KN, Mehrotra A, Bloom EL, Uscher-Pines L. Prevalence and characteristics of telehealth utilization in the United States. JAMA Netw Open 2020 Oct 1;3(10):e2022302. [doi: 10.1001/jamanetworkopen.2020.22302] [Medline: 33104208]
28. Gordon NP, Crouch E. Digital information technology use and patient preferences for internet-based health education modalities: cross-sectional survey study of middle-aged and older adults with chronic health conditions. JMIR Aging 2019 Apr 4;2(1):e12243. [doi: 10.2196/12243] [Medline: 31518291]
29. Wilczewski H, Paige SR, Ong T, et al. Providers' perspectives on telemental health usage after the COVID-19 pandemic: retrospective analysis. JMIR Form Res 2022 Nov 11;6(11):e39634. [doi: 10.2196/39634] [Medline: 36322787]
30. Chaet D, Clearfield R, Sabin JE, Skimming K, Council on Ethical and Judicial Affairs American Medical Association. Ethical practice in telehealth and telemedicine. J Gen Intern Med 2017 Oct;32(10):1136-1140. [doi: 10.1007/s11606-017-4082-2] [Medline: 28653233]
31. Holtz B, Mitchell K, Hirko K, Ford S. Using the technology acceptance model to characterize barriers and opportunities of telemedicine in rural populations: survey and interview study. JMIR Form Res 2022 Apr 15;6(4):e35130. [doi: 10.2196/35130] [Medline: 35436207]
32. Holtz BE. Patients perceptions of telemedicine visits before and after the coronavirus disease 2019 pandemic. Telemed J E Health 2021 Jan;27(1):107-112. [doi: 10.1089/tmj.2020.0168] [Medline: 32614689]

Abbreviations

DHO: District Health Officer

Edited by A Mavragani; submitted 23.05.24; peer-reviewed by M Mars; revised version received 19.11.24; accepted 20.11.24; published 23.01.25.

Please cite as:
Kizito M, Mugabi EN, Ford S, Holtz B, Hirko K
Characterizing Telehealth Barriers and Preferences to Promote Acceptable Implementation Strategies in Central Uganda: Multilevel Formative Evaluation
JMIR Form Res 2025;9:e60843
URL: https://formative.jmir.org/2025/1/e60843
doi:10.2196/60843

© Michael Kizito, Erina Nabunjo Mugabi, Sabrina Ford, Bree Holtz, Kelly Hirko. Originally published in JMIR Formative Research (https://formative.jmir.org), 23.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Testing an Electronic Patient-Reported Outcome Platform in the Context of Traumatic Brain Injury: PRiORiTy Usability Study

Christel McMullan[1], PhD; Grace Turner[2], PhD; Ameeta Retzer[1], PhD; Antonio Belli[3], MD, FRCS, FRCS (SN); Elin Haf Davies[4], PhD; Laura Nice[5], PhD; Luke Flavell[1]; Jackie Flavell[1]; Melanie Calvert[1], PhD

[1]Centre for Patient Reported Outcomes Research, Institute of Applied Health Research, University of Birmingham, Birmingham, United Kingdom

[2]Sport, Exercise and Rehabilitation Sciences, University of Birmingham, Birmingham, United Kingdom

[3]Institute of Inflammation and Ageing, University of Birmingham, Birmingham, United Kingdom

[4]Aparito Ltd, Wrexham, United Kingdom

[5]Centre for Conflict Wound Research, University of Birmingham, Birmingham, United Kingdom

Corresponding Author:
Christel McMullan, PhD
Centre for Patient Reported Outcomes Research, Institute of Applied Health Research, University of Birmingham, Birmingham, United Kingdom

Abstract

Background: Traumatic brain injury (TBI) is a significant public health issue and a leading cause of death and disability globally. Advances in clinical care have improved survival rates, leading to a growing population living with long-term effects of TBI, which can impact physical, cognitive, and emotional health. These effects often require continuous management and individualized care. Traditional paper-based assessments can be cumbersome, potentially impeding regular monitoring of patient-reported outcomes (PROs). Electronic PROs (ePROs) offer a promising alternative by enabling real-time symptom tracking, which can facilitate early identification of issues, support shared decision-making, and improve outcomes for patients with TBI.

Objective: This study evaluates the usability of an ePRO platform—Atom5—for individuals with TBI. By analyzing how patients use the system to report their symptoms, the study aims to identify usability issues, assess user satisfaction, and determine the potential of Atom5 to support ongoing patient-centered care.

Methods: Atom5 was customized to enable individuals with TBI to report their symptoms. Usability testing was conducted through one-on-one sessions with participants recruited from Headway UK—an organization supporting brain injury survivors. Each participant took part in cognitive interviews using with the "Think Aloud" method, encouraging them to verbalize their thoughts and experiences while using the platform. This approach provided qualitative insights into areas of difficulty, usability strengths, and accessibility barriers. User satisfaction was quantitatively assessed with a brief 4-item questionnaire based on the System Usability Scale. Usability outcomes were analyzed for critical and noncritical errors, focusing on user experience and overall satisfaction.

Results: In total, 9 participants completed a single usability testing session using Atom5, including 4 men, 4 women, and 1 nonbinary individual; 4 participants were under 55 years old, and 6 had their TBI <10 years ago. Finally, 8 participants used an Android device. The platform included measures for anxiety (Generalized Anxiety Disorder-2 item), depression (Patient Health Questionnaire-2), posttraumatic stress disorder (Posttraumatic Stress Disorder checklist 2), and TBI-specific quality of life (Traumatic Brain Injury – Quality of Life Short form) and a total of 26 questions. Overall, all participants were satisfied with the system, noting that it was easy to navigate and accessible despite difficulties in understanding some questions. Further, 6 participants encountered no errors, while 1 participant reported one critical error and 2 others reported one noncritical error each. The participants rated their overall satisfaction with the platform at an average score of 3.9 (SD 0.49) out of 5.

Conclusions: This usability study suggests that individuals living with TBI can effectively report symptoms using the Atom5 ePRO platform, with generally high satisfaction and few usability issues, thereby enabling continuous monitoring and proactive symptom management. Future ePRO development should focus on inclusivity and adaptability to address the diverse needs of patients with TBI, ensuring these tools can effectively support a wide range of users.

(JMIR Form Res 2025;9:e58128) doi:10.2196/58128

KEYWORDS

usability study; usability; patient reported outcome; PRO; electronic patient reported outcome; ePRO; traumatic brain injury; TBI; think aloud; cognitive interviews; early warning; early detection; mobile phone

XSL•FO
RenderX

Introduction

Background

A traumatic brain injury (TBI) is defined as "a traumatically induced structural injury and/or physiological disruption of brain function as a result of an external force" [1]. Each year, it is estimated that over 50 million TBIs occur worldwide [2]. In the United Kingdom, around 900,000 people attend accident and emergency departments (immediate critical care) with a head injury every year, with approximately 1.3 million people living with the effects of a TBI [3].

Although most TBIs result in mild symptoms and do not require hospitalization, many individuals experience disabling long-term symptoms [4-7]. The impact of a TBI falls into several categories: physical (headaches, dizziness, and blurred vision), cognitive (trouble with memory, concentration and attention, and impaired reasoning), emotional (personality changes, mood swings, depression, and anxiety), and behavioral (fatigue and anger) [8-11]. This means that patients with mild TBI have different needs, and health care must be tailored to each patient [12].

In addition, advances in management and guidelines [13] mean that people now live longer with the effects caused by their TBI. It is therefore important for clinicians to monitor these effects over time to improve the quality of life of people living with the effects of a TBI.

As a result, TBI is being increasingly regarded as a chronic condition, as it impacts on multiple health domains and functions [4,14]. Research has shown that TBI is a risk factor for dementia [14], psychiatric, cardiovascular, genitourinary and neurodegenerative conditions [15-17].

One way to monitor the effects of TBI on patients' symptoms and quality of life over time is by using patient-reported outcomes measures (PROMs). PROMs are defined as "questionnaires completed by patients to assess the effects of disease or treatment (or both) on symptoms, functioning, and health related quality of life from their perspective" [18]. PROMs have traditionally been administered in a paper format. Research has shown that there are several drawbacks to paper PROMs. These include lengthy completion time, risk of poor data quality, and being prone to human error [19]. The growth of eHealth technologies and the omnipresence of smartphones and tablets [20] has allowed patients to play a more active role in their health care [21-23].

Digital health is a rapidly evolving field and provides a cost-effective way to remotely monitor patients [12,23]. Often referred to as mobile health or "mhealth," recent research has shown that the use of these new technologies in health care has been received positively both by patients [12,24] and clinicians [24-26]. In addition to this, new technologies have given clinicians and patients the opportunity to use electronic PROMs (ePROMs) [20]. Unlike paper PROMs, ePROMs can facilitate real-time data, remote monitoring of symptoms and send or receive feedback [27], and can improve acceptance rate [28]. Examples of successful ePROs systems include AmbuFlex in Denmark with patients with renal failure [29] and the eRAPID system in the United Kingdom to monitor side effects of chemotherapy [30]. However, disadvantages of ePROMs include concerns around privacy issues, technical difficulties, a large initial financial investment, and the potential digital divide [19].

Like any new product being developed, it is vital to test the usability of a new electronic patient-reported outcome (ePRO) system in order to improve and optimize the final product [31,32]. Usability testing is defined by the International Organisation for Standardization (ISO) as "the extent to which a product can be used by specified user to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [33].

Research Context: The PRiORiTy Study

This usability study is the second stage of the wider National Institute for Health and Care Research Surgical Reconstruction and Microbiology Research Centre (NIHR SRMRC)–funded PRiORiTy (Patient-Reported Outcomes Research in Trauma) study. The PRiORiTy study aimed to develop and test the usability and feasibility of an ePRO system with people with a TBI. Prior to this usability study, we conducted a qualitative study on the views and perspectives of patients living with the effects of a TBI, their carers or family members, and health care professionals on using PROMs or ePROMs. The findings of the qualitative study were reported in an article published elsewhere [27].

Aim

The aim of this study was to test the usability of using an ePRO system (Atom5) with individuals who had a TBI.

Methods

Study Participants and Settings

People who had a TBI were recruited in July 2021 from Headway UK, a UK-based charity aiming to promote understanding of all aspects of brain injury and provide information, support and services to survivors, their families and carers. The sample size was informed by previous usability studies, which were in favor of the 10±2 rules, which advocated that around 10 participants could identify 80% of the issues around the system being tested [34-36]. Other studies showed that as low as 5 people are needed if conducting a series of testing sessions [31,37]. However, since we conducted only one testing session, we decided to increase our sample size and it was clear that we reached data saturation by the time we conducted our last testing session. Inclusion criteria were being 18 years of age and older, having the ability to converse in English, and to give informed consent. We excluded anyone who was not able to give informed consent.

Recruitment Process

A lead contact within Headway UK identified people who were deemed eligible (purposive sampling), approached them in person, and introduced them to the study. The Headway UK staff member briefly explained the study to eligible people and gave them the information sheet. They then sent the researcher (CM), via a secure email address, the contact details of the people who had verbally agreed to take part in the usability

XSL•FO
RenderX

study. CM phoned these people to arrange a mutually suitable date and time to conduct the testing session. Written informed consent was obtained from all the participants and data were anonymized.

ePRO Platform (Atom5)

The ePRO platform used in this study was Atom5, which was developed by Aparito Ltd [38], a UK-based medical technology company. Atom5 consists of two interfaces (Multimedia Appendix 1):

1. A clinician dashboard accessed via a web browser that can be accessed by specific members of the research team with appropriate permission.
2. A patient-facing interface accessed via an app on Android or iOS devices onto which study participants input their patient-reported outcome data.

The Atom5 app for the PRiORiTy study was designed in conjunction with our Patient and Public Involvement and Engagement group, the research team (including a consultant neurosurgeon) and was informed by the findings from the project's qualitative study [27].

Aparito Ltd are accredited for ISO 13485:2016 for medical device quality management system, ISO 27001:2001 for information security management systems, and Cyber Essential Plus recertification.

PROMs Selection

The selection of PROMs was informed by the findings of our previous qualitative research [27], literature searches, and input from our patient partners and clinical collaborators. Our qualitative research showed that people who had a TBI often experienced anxiety, depression, and posttraumatic stress disorder, and this was further evidenced by the existing literature [8,9,27]. Therefore, the PROMs included in this study were the Generalized Anxiety Disorder-2 item (GAD-2), Patient Health Questionnaire-9 (PHQ-9), Posttraumatic Stress Disorder checklist 2 (PCL-2) and Traumatic Brain Injury – Quality of Life Short form (TBI QOL SF; ability to participate in social roles and satisfaction with social roles and activities). Quality of Life after Brain Injury (QOLIBRI) was considered as a potential TBI-specific measure. However, because it includes 37 items, the Patient and Public Involvement and Engagement and clinical collaborators advised against it, as they felt participants with TBI would not be able to complete it due to their inability to stay focused for long periods of time. Instead they suggested using 2 submeasures of the TBI QOL SF (ability to participate in social roles and satisfaction with social roles and activities) for a total of 20 questions. In addition, clinical collaborators had already been using TBI QOL questionnaire in their clinical practice.

The details of the four PROMs selected are mentioned in Table 1.

XSL•FO
RenderX

Table . Patient-reported outcome measures included in the Patient-Reported Outcomes Research in Trauma usability study.

PROM[a] category and PROMs	Description
General population	
GAD-2[b] (Anxiety)	• Screening tool for generalized anxiety disorder [39] derived from GAD-7[c] • Two items • Recall period of 2 weeks • Validated measure which has retained the same psychometrics properties of the GAD-7 (86% sensitivity and 83% specificity) [40]
PHQ-2[d] (Depression)	• Screening tool for depression derived from the PHQ-9[e] • Two items • Recall period of 2 weeks • Validated measure with sensitivity of 83% and specificity of 92% [41]
PCL-2[f] (PTSD[g])	• Abbreviated version of the PTSD Checklist – Civilian version (PCL-C) • Two items • Screening tool for PTSD • Recall period of 1 month • Validated measure with sensitivity of .97 and specificity of .58 [42]
Disease-specific	
TBI QOL SF[h] (ability to participate in social roles; satisfaction with social roles and activities)	• Health-related Quality of Life • 20 items • Recall period of 7 days • Validated measures with excellent psychometric characteristics, although more research is needed to establish evidence of construct validity and evaluate the measures' sensitivity to change [43]. In addition, the measures need to be validated when using via mobile health.

[a]PROM: patient-reported outcome measures.

[b]GAD-2: Generalized Anxiety Disorder-2 item.

[c]GAD-7: Generalized Anxiety Disorder-7 item.

[d]PHQ-2: Patient Health Questionnaire-2.

[e]PHQ-9: Patient Health Questionnaire-9.

[f]PCL-2: Posttraumatic Stress Disorder checklist 2.

[g]PTSD: posttraumatic stress disorder.

[h]TBI QOL SF: Traumatic Brain Injury – Quality of Life Short form.

Data Collection and Testing Procedure

Testing Sessions

The usability testing aimed at assessing the effectiveness and efficiency of the ePRO platform. This consisted of one-to-one sessions conducted either online on a University of Birmingham secure Zoom account or face-to-face on Headway premises in the West Midlands. The testing sessions and interviews were conducted by CM, PhD, who is a female qualitative research fellow with extensive experience in qualitative data collection, including conducting interviews. She did not know the participants prior to the start of the research. The participants were fully informed of the reasons for doing the research prior to the data collection.

At the start of the session, participants were given a patient leaflet that outlined the main stages involved in navigating through the app and completing the questionnaires. They were then asked to download the app and perform a series of tasks on the app while sharing their thoughts about the platform (Think Aloud technique) [44]. Finally, the participants took part in a short interview (10 - 15 min) on their experience of using the electronic platform. A carer or family member was allowed to attend the testing session. A topic guide was developed and pilot-tested with a patient with TBI. Field notes were kept during the testing sessions.

The tasks performed by the participants are detailed in Textbox 1.

XSL•FO
RenderX

Textbox 1. Tasks performed by participants during the testing sessions.

- Download the Atom5 app
- Onboard
- Consent
- Navigate through homepage
- Navigate through Questionnaires module
- Complete and submit Generalized Anxiety Disorder-2
- Complete and submit Patient Health Questionnaire-2
- Complete and submit Posttraumatic Stress Disorder checklist 2
- Complete and submit Traumatic Brain Injury – Quality of Life Short form
- Navigate through frequently asked questions module

Satisfaction Questionnaire

Finally, the participants were asked to answer a brief 4-question satisfaction questionnaire to rate their satisfaction with the usability of the electronic platform (5-point scale: 1–poor or never to 5–excellent or yes) [36,45].

Data Saturation

There has been an ongoing debate about sample sizes for usability testing. Our sample size has been been informed by previous research, which shows that 8 - 10 participants are required to detect over 80% of issues.

Data Analysis

Data were analyzed by two independent researchers. The participant characteristics were summarized as frequency and percentage. The testing sessions were audio-recorded. The participants' comments during the testing sessions and the researcher's observations were collated into a table. Errors were recorded and divided into critical (requiring the researcher's input to be able to continue) and noncritical (not requiring the researcher's input to be able to continue). Suggestions to improve the electronic platform were also recorded. Participant ratings for the 4 satisfaction questions were used to calculate a mean score per question. Overall mean score and SD was also calculated.

Data collected from the ePROMs was used for the purpose of the usability study only and was not downloaded.

The COREQ (Consolidated Criteria for Reporting Qualitative Research) reporting checklist was used to report findings (Checklist 1).

Patient and Public Involvement and Engagement

One person who was living with the effects of a TBI and a family member were involved throughout the study. We met regularly (on average every 3 mo) and they reviewed patient-facing documents (patient information sheet, consent form, and instructions leaflet), assisted with recruitment and provided advice on the design and development of the Atom5 platform. In addition to meeting regularly, we were in regular email contact in order to quickly address any urgent issues.

Ethical Considerations

The study was approved the University of Birmingham Ethics committee (reference ERN_17 - 1253). Informed consent was obtained from all participants. The participants were not compensated for taking part in the study. All data were deidentified to ensure participant privacy and confidentiality.

Results

Participants' Characteristics

A total of 9 people who had a TBI took part in a testing session. This included 4 men, 4 women, and 1 person identifying as nonbinary. A total of 4 out of the 9 participants (44%) were under 55 years old, and 5 were over 55 years old. The majority of the participants (6/9 participants, 66%) had their TBI less than 10 years ago. Finally, 8 out of 9 participants (89%) used an Android device, and 1 used an iOS device. The participants' characteristics are summarized in Table 2.

In total, 5 testing sessions took place remotely by video using Zoom and the other 4 took place face-to-face on Headway UK premises. No one else was present in the interview room, although Headway UK staff were in the building. The testing sessions lasted between 42 and 63 minutes.

Among the 4 male participants, 3 of them were visually impaired and needed some help to read the questions. The researcher (CM) read the questions for them verbatim and helped them to input their answers onto the electronic device. These 3 male participants did not own an electronic device and had to borrow one from friends.

XSL•FO
RenderX

Table . Participants' characteristics.

Variables		Values, n (%)
Age (years)		
	<55	4 (44)
	>55	5 (56)
Sex		
	Male	4 (44)
	Female	4 (44)
	Nonbinary	1 (12)
Ethnicity		
	White	9 (100)
Number of years since traumatic brain injury		
	<10 years ago	6 (67)
	10 - 20 years ago	2 (22)
	>20 years ago	1 (11)
Device used		
	iOS	1 (11)
	Android	8 (89)

Effectiveness and Efficiency of the Atom5 App

One critical error and 2 noncritical errors were recorded and all the participants completed the testing sessions. The critical error was due the participant not knowing what to do after inputting the onboarding code (please note that after downloading the Atom5 platform, users need to either scan a QR code or enter a numerical code to be able to access the questionnaires. The QR code or numerical code was given to them by the researcher (CM). The two noncritical errors were due to downloading the app on an iPad not compatible with the platform and not being able to answer a question because the question was missing. Table 3 shows how each participant addressed the critical and noncritical errors.

Table . Descriptions of critical and noncritical errors encountered by participants and how they were addressed during testing sessions.No differences were found between iOS and Android users.

Error (error type)	How participants addressed noncritical error
Not knowing what to do after inputting the onboarding code (critical error)	Researcher advised the participant
Downloading the app on an iPad not compatible with the platform (non-critical error)	Participant used a newer device
Not being able to answer a question because the question was missing (noncritical error)	Participant had to go back to previous page to read the question

It was difficult to record how long it took participants to complete the PROMs on Atom5 as they discussed their experience with the researcher while carrying out the tasks required. However, the researcher observed that while the majority had very little difficulty completing GAD-2, PHQ-2, and PCL-2 and completed them in under 30 seconds, most of the participants struggled with the TBI QOL SF questionnaire and took longer to complete it. One of the reasons for this is that it is a longer questionnaire (20 questions for TBI QOL SF vs 6 questions for GAD-2, PHQ-2, and PCL-2 combined). However, another reason for struggling with it was the difficulty understanding some of the questions, such as the following questions, and not knowing how to answer them:

- "I am able to do all the community activities that I want to do"
- "I can do everything for work that I want to do (including work at home)."
- "I am satisfied with the amount of time I spend doing work (including work at home)."

Issues and Suggestions for Improvement

The participants identified a number of issues and made some suggestions to improve their experience of using the Atom5 platform (Table 4). These issues and suggestions related to font size, lack of clarity, names of questionnaires, questions numbers, clarity of some the text, what to do after onboarding, communication with clinical team, length of questionnaires, questions wording, colors, and name of the app.

XSL•FO
RenderX

Table . Summary of issues and suggestions for improvement made by participants during the testing sessions.

Issue	Details	Suggestions made by participants
Font size too small	Some items were difficult to read: • Questions in GAD-2[a] were difficult to read • FAQs[b]	• Increase font
Lack of clarity	Difficulty reading and understanding: • Consent text • FAQs • Thank you message after questionnaires • Unsure what the FAQs are about	• Reformat using bullet points or add paragraphs • Include information in FAQs about study and what participants have to do
Lack of information about study	• What participants need to do after onboarding was not always obvious to all participants	• Add instructions on app
Communication with clinical team	• Need for some participants to leave comments	• Add free text boxes at the end of each questionnaire
Length of questionnaires	• TBI QOL SF[c] was lengthy	• Add the question number at the start of TBI QOL SF question (eg, 1/20; 2/20, etc)
Questions wording	• Difficulty understanding some of the TBI QOL SF questions	• Reword the questions[d] • Add emojis/pictures
Orange and white colors	• Most participants did not understand the significance of the orange and white colors • The white color was difficult to see by one participant	• Add explanation in FAQs • Increase font size • Bold words that are on white background
Name of the app Atom5	• Name of app was not related to TBI[e]	• Change the name to make it more relevant to TBI

[a]GAD-2: Generalized Anxiety Disorder-2 item.

[b]FAQ: frequently asked questions.

[c]TBI QOL SF: Traumatic Brain Injury – Quality of Life Short form.

[d]This may not be possible, as it may impact on validity of measures and would require work with the developers.

[e]TBI: traumatic brain injury.

Participants' Overall Comments and Researchers' Observations

Overall, participants were positive about using Atom5 to report their symptoms. They found the app easy to download and use. Most of them liked the white and orange colors of the app, even though they did not understand the significance of it. The main issues mentioned by the participants was regarding the TBI QOL SF questions, which they had difficulty understanding and therefore were not sure how to answer them. One participant reported they would be unlikely to use the platform again because it would be too difficult to use on their own as they would forget to use it (Textbox 2).

XSL•FO
RenderX

Textbox 2. Participants' overall comments on using the Atom5 platform and researchers' observations.

Participants' comments

- Very easy to use.
- App is quite good and looks smart.
- Easier than a piece of paper.
- Things are laid out very well in the app.
- Unlikely to use it again because it would be difficult to use on my own.
- I wouldn't remember to use it.
- Downloading the app might be an issue for some people with a traumatic brain injury.
- Made sense but had to think about the questions.

Researchers' observations

- Several participants had to read the Traumatic Brain Injury – Quality of Life Short form (TBI QOL SF) questions several times, difficulty understanding what the questions mean.
- Two participants were getting tired halfway through completing the TBI QOL SF questionnaire.

Participants' Satisfaction

The summary scores for participants' satisfaction and rating of usability of Atom5 are shown in Table 5. The overall mean score was 3.9 (SD 0.33) and the mean score for individual questions was between 3.4 (SD 0.52) and 4.6 (SD 0.72).

Table . Participants' usability and satisfaction scores of using the Atom5 electronic platform (range: 1 - 5).

Question	Average score (SD)
Q1. How easy was the system to use and navigate?	3.9 (0.33)
Q2. How satisfied are you with the content?	4.6 (0.72)
Q3. How satisfied are you with the display?	3.4 (0.52)
Q4. How likely are you to use it again or recommend it to others?	3.9 (1.16)
Average usability and satisfaction score	3.9 (0.49)

Discussion

Summary of Findings

This article reports the results of the usability testing of the Atom5 platform used by people living with the effects of a TBI. Our ePRO platform was informed by existing literature, patients, and a clinician. A total of 4 PRO measures were selected and programmed in Atom5. The results of the interviews show that most of the people living with the effects of a TBI who took part in the usability study were able to report their symptoms using the Atom5 platform. The number of errors were low (1 critical and 2 noncritical errors) and were due to incompatibility of a device, not knowing what to do after onboarding, or a missing question. Overall, the platform was found to be effective. Suggestions to improve the electronic platform included increasing the font, reformatting the text and layout to make it clearer, and adding instructions or study information and free text boxes.

Findings in Relation to Existing Literature

The results of our usability study reflect results from previous usability studies to some extent [24]. Although our participants were positive about their overall experience of using Atom5, their satisfaction scores were slightly lower than in other studies [36,45,46], reminding us that barriers to the use of ePRO platforms still exist. Participants' lower satisfaction scores could be attributed to several factors, including visual difficulty, fatigue, cognitive impairment, and some questions difficult to understand and answer.

As we were aware that a TBI, even a mild TBI, can cause fatigue and attention deficit, among many other symptoms [47,48], we deliberately kept the number of questions as low as possible while being useful enough for the clinical team. However, a few participants started to feel tired a few minutes into the testing sessions. This is sometimes known as "respondent fatigue," and it refers to the way respondents lose concentration as they complete questionnaires, especially in the latter part of the questionnaires [49]. The fact that all the participants managed to complete the questionnaires and the tasks required is encouraging and suggests that they believed in the potential of the ePRO platform enough to complete the testing sessions [50].

In addition, several participants had to read some of the questions several times, especially questions from the TBI QOL SF questionnaire. Even after reading the questions several times, they did not always understand them. While this could be due to a lack of concentration, it is also worth wondering if the

XSL•FO
RenderX

wording of the questionnaires is suitable for individuals living with the effects of a TBI and whether the measure is suitable for use in routine care. As this issue might not directly relate to the ePRO platform, it would be important to explore this further in a future study.

One of the main risks of using ePRO platforms is excluding certain groups of patients [19,51]. This concerned 2 specific groups in our study. First, 3 participants who were visually impaired after their TBI had to be helped with reading the questions. Visual impairment is a common consequence of TBI [52], and therefore it is essential to include individuals who have visual impairment after a TBI. The second group of people who risks of being excluded is people who do not own an electronic device or are not familiar with technology [19]. A total of 3 of our participants did not have a smartphone or tablet and borrowed someone else's device in order to take part in the usability study. This could, however, raise issues around data privacy and security.

Finally, 1 participant felt they would not remember to use the electronic platform unless they were sent a reminder and suggested being sent a reminder, which has been shown to increase response rate [53].

Implications for ePRO Developers, Health Care Professionals, and Researchers

The fact that some of our participants were visually impaired should not exclude them from using ePROMs, as demonstrated in our study. In the case of our participants with visual impairment, the researcher read the questions. Therefore, ePRO developers, health care providers, and researchers should ensure electronic platforms are accessible for all, to ensure inclusivity and maximize participation [46] by allowing family members or carers to complete proxy measures [54] or by providing a "Read aloud" option for patients.

Similarly, although this was not the case with our participants, the suggestion that some people who had a TBI might struggle with downloading the app suggests issues around digital literacy. ePRO developers and researchers should ensure participants receive adequate training or support. The fact that 3 of our participants did not have access to a mobile device indicates that electronic devices should be provided to patients to allow them to complete ePROMs. This would also remove privacy and security issues if using their own devices or borrowing someone else's. Maximizing participation could also be increased if ePRO developers sent reminders to participants.

Several participants mentioned they would have liked to see their previous answers to the questionnaires in order to view their progress over time. Therefore, adding visual representations of users' previous answers would help them to do so.

Finally, it is recommended for health care professionals and researchers to use validated PROMs.

Strengths and Limitations

The main strength of our study is the Think Aloud technique used for data collection; allowing participants to share their views while carrying out the requested tasks provides real-time feedback [44,55], in particular, about the questions from the TBI QOL that they had difficulty understanding and responding to.

There are several limitations to our study. The first one is the lack of ethnic diversity among participants, meaning that study findings might not be representative of the wider population. The research team recognizes the need to promote inclusive data collection. In addition, the lack of text-to-speech or voice commands in the app made it more difficult for participants with vision impairment to use the app.

The time it took participants to complete the tasks was not representative of the time it would take to complete the ePROs, as they shared their views while carrying out the tasks. This made it difficult to assess efficiency of the electronic platform.

The fact that, despite excellent psychometric characteristics and a reading age of fifth-grade level (10/11 years old), some of the participants struggled to understand and answer some of the questions, implying that their answers did not reflect how they felt.

Finally, the thinking aloud approach used in this usability study means that some of the participants could become a little tired during the testing sessions because of the cognitive issues some of them faced. Researchers should ask participants to reflect on their experience after each task rather than at the end of the session.

Conclusion

Usability testing is a vital stage of any new product development. In our case, it showed that although individuals living with the effects of a TBI can report their symptoms on an electronic platform, ePRO developers, health care providers, and researchers should ensure that electronic platforms are inclusive and can be adapted to people's needs.

Future research should include testing the Atom5 platform in TBI clinic settings and with ethnic minorities.

Acknowledgments

The authors wish thank Headway UK for their contribution to recruitment.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

CM, GT, AR, TB, LN, JF, LF, and MC designed the study. CM completed the data collection and data analysis. CM drafted the manuscript. EHD/Aparito developed the Atom5 platform and provided technical support. All authors reviewed and provided feedback on the manuscript.

Conflicts of Interest

CM receives funding from the National Institute for Health Research (NIHR) Surgical Reconstruction and Microbiology Research Centre, the NIHR Blood and Transplant Research Unit (BTRU) in Precision Transplant and Cellular Therapeutics, Innovate UK. GT receives funding from the NIHR. AR receives funding from the NIHR Birmingham Biomedical Research Centre, University of Birmingham, and the NIHR ARC West Midlands, University of Birmingham. MC is an NIHR senior investigator and receives funding from the NIHR Birmingham Biomedical Research Centre, the NIHR Surgical Reconstruction and Microbiology Research Centre, and NIHR ARC West Midlands at the University of Birmingham and University Hospitals Birmingham NHS Foundation Trust, Health Data Research UK, UKRI, Macmillan Cancer Support, UCB Pharma, Gilead, Janssen, and GSK. MC has received personal fees from Astellas, Aparito Ltd, CIS Oncology, Takeda, Merck, Daiichi Sankyo, Glaukos, GSK, and the Patient-Centred Outcomes Research Institute (PCORI) outside the submitted work. DK reports grants from Macmillan Cancer Support, the NIHR, NIHR Birmingham Biomedical Research Centre, and NIHR SRMRC at the University of Birmingham and University Hospitals Birmingham NHS Foundation Trust, and personal fees from Merck and GSK outside the submitted work.

Multimedia Appendix 1
Screenshots of Atom5 app and clinical dashboard.
[DOCX File, 395 KB - formative_v9i1e58128_app1.docx]

Checklist 1
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[PDF File, 599 KB - formative_v9i1e58128_app2.pdf]

References

1. Head injury. Quality standards and indicators: briefing paper. : National Institute for Health and Care Excellence; 2014 URL: https://www.nice.org.uk/guidance/qs74/resources/head-injury-briefing-paper2 [accessed 2024-11-29]
2. Feigin VL, Theadom A, Barker-Collo S, et al. Incidence of traumatic brain injury in New Zealand: a population-based study. Lancet Neurol 2013 Jan;12(1):53-64. [doi: 10.1016/S1474-4422(12)70262-4] [Medline: 23177532]
3. Traumatic brain injury across the life course: priorities, challenges, and opportunities. : UK Research and Innovation/Medical Research Council; 2022 URL: https://www.ukri.org/wp-content/uploads/2022/12/MRC-07122022-MRC-Traumatic-Brain-Injury-Workshop-report_June-2022.pdf [accessed 2024-11-29]
4. Maas AIR, Menon DK, Manley GT, et al. Traumatic brain injury: progress and challenges in prevention, clinical care, and research. Lancet Neurol 2022 Nov;21(11):1004-1060. [doi: 10.1016/S1474-4422(22)00309-X] [Medline: 36183712]
5. Levin HS, Diaz-Arrastia RR. Diagnosis, prognosis, and clinical management of mild traumatic brain injury. Lancet Neurol 2015 May;14(5):506-517. [doi: 10.1016/S1474-4422(15)00002-2] [Medline: 25801547]
6. Varner C, Thompson C, de Wit K, Borgundvaag B, Houston R, McLeod S. Predictors of persistent concussion symptoms in adults with acute mild traumatic brain injury presenting to the emergency department. CJEM 2021 May;23(3):365-373. [doi: 10.1007/s43678-020-00076-6] [Medline: 33555602]
7. Masel BE, DeWitt DS. Traumatic brain injury: a disease process, not an event. J Neurotrauma 2010 Aug;27(8):1529-1540. [doi: 10.1089/neu.2010.1358] [Medline: 20504161]
8. Ryan LM, Warden DL. Post concussion syndrome. Int Rev Psychiatry 2003 Nov;15(4):310-316. [doi: 10.1080/09540260310001606692] [Medline: 15276952]
9. Traumatic brain injury (TBI). Headway. 2022. URL: https://www.headway.org.uk/about-brain-injury/individuals/types-of-brain-injury/traumatic-brain-injury [accessed 2024-12-04]
10. Tran S, Kenny B, Power E, et al. Cognitive-communication and psychosocial functioning 12 months after severe traumatic brain injury. Brain Inj 2018;32(13-14):1700-1711. [doi: 10.1080/02699052.2018.1537006] [Medline: 30346807]
11. Voormolen DC, Polinder S, von Steinbuechel N, Vos PE, Cnossen MC, Haagsma JA. The association between post-concussion symptoms and health-related quality of life in patients with mild traumatic brain injury. Injury 2019 May;50(5):1068-1074. [doi: 10.1016/j.injury.2018.12.002] [Medline: 30554897]
12. Karvandi E, Barrett L, Newcombe V, Hutchinson P, Helmy A. Digital health interventions for remote follow-up after mild traumatic brain injury. Br J Neurosurg 2024 May 6;6(1-7):1-7. [doi: 10.1080/02688697.2024.2346564] [Medline: 38711206]
13. Lawrence T, Helmy A, Bouamra O, Woodford M, Lecky F, Hutchinson PJ. Traumatic brain injury in England and Wales: prospective audit of epidemiology, complications and standardised mortality. BMJ Open 2016 Nov 24;6(11):e012197. [doi: 10.1136/bmjopen-2016-012197] [Medline: 27884843]

XSL•FO
RenderX

14. Dams-O'Connor K, Juengst SB, Bogner J, et al. Traumatic brain injury as a chronic disease: insights from the United States Traumatic Brain Injury Model Systems Research Program. Lancet Neurol 2023 Jun;22(6):517-528. [doi: 10.1016/S1474-4422(23)00065-0] [Medline: 37086742]
15. Izzy S, Chen PM, Tahir Z, et al. Association of traumatic brain injury with the risk of developing chronic cardiovascular, endocrine, neurological, and psychiatric disorders. JAMA Netw Open 2022 Apr 1;5(4):e229478. [doi: 10.1001/jamanetworkopen.2022.9478] [Medline: 35482306]
16. Brett BL, Gardner RC, Godbout J, Dams-O'Connor K, Keene CD. Traumatic brain injury and risk of neurodegenerative disorder. Biol Psychiatry 2022 Mar 1;91(5):498-507. [doi: 10.1016/j.biopsych.2021.05.025] [Medline: 34364650]
17. Livingston G, Huntley J, Liu KY, et al. Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. Lancet 2024 Aug 10;404(10452):572-628. [doi: 10.1016/S0140-6736(24)01296-0] [Medline: 39096926]
18. Calvert M, Kyte D, Price G, Valderas JM, Hjollund NH. Maximising the impact of patient reported outcome assessment for patients and society. BMJ 2019 Jan 24;364:k5267. [doi: 10.1136/bmj.k5267] [Medline: 30679170]
19. Meirte J, Hellemans N, Anthonissen M, et al. Benefits and disadvantages of electronic patient-reported outcome measures: systematic review. JMIR Perioper Med 2020 Apr 3;3(1):e15588. [doi: 10.2196/15588] [Medline: 33393920]
20. Glenwright BG, Simmich J, Cottrell M, et al. Facilitators and barriers to implementing electronic patient-reported outcome and experience measures in a health care setting: a systematic review. J Patient Rep Outcomes 2023 Feb 14;7(1):13. [doi: 10.1186/s41687-023-00554-2] [Medline: 36786914]
21. Shaw T, McGregor D, Brunner M, Keep M, Janssen A, Barnet S. What is eHealth (6)? Development of a conceptual model for eHealth: qualitative study with key informants. J Med Internet Res 2017 Oct 24;19(10):e324. [doi: 10.2196/jmir.8106] [Medline: 29066429]
22. Cao J, Lim Y, Sengoku S, Guo X, Kodama K. Exploring the shift in International trends in mobile health research from 2000 to 2020: bibliometric analysis. JMIR Mhealth Uhealth 2021 Sep 8;9(9):e31097. [doi: 10.2196/31097] [Medline: 34494968]
23. Guo X, Chen S, Zhang X, Ju X, Wang X. Exploring patients' intentions for continuous usage of mHealth services: elaboration-likelihood perspective study. JMIR Mhealth Uhealth 2020 Apr 6;8(4):e17258. [doi: 10.2196/17258] [Medline: 32250277]
24. Bonn MM, Graham LJ, Marrocco S, Jeske S, Moran B, Wolfe DL. Usability evaluation of a self-management mobile application for individuals with a mild traumatic brain injury. Dig Health 2023;9:20552076231183555. [doi: 10.1177/20552076231183555] [Medline: 37426589]
25. Ganefianty A, Songwathana P, Damkliang J, Imron A, Latour JM. A mobile health transitional care intervention delivered by nurses improves postdischarge outcomes of caregivers of patients with traumatic brain injury: a randomized controlled trial. World Neurosurg 2024 Apr;184:191-201. [doi: 10.1016/j.wneu.2024.01.078] [Medline: 38244683]
26. Sánchez-Rodríguez MT, Pinzón-Bernal MY, Jiménez-Antona C, et al. Designing an informative app for neurorehabilitation: a feasibility and satisfaction study by physiotherapists. Healthcare (Basel) 2023 Sep 14;11(18):2549. [doi: 10.3390/healthcare11182549] [Medline: 37761746]
27. McMullan C, Retzer A, Slade A, et al. Care providers' and patients' attitudes toward using electronic-patient reported outcomes to support patients with traumatic brain injury: a qualitative study (PRiORiTy). Brain Inj 2020 May 11;34(6):723-731. [doi: 10.1080/02699052.2020.1740944] [Medline: 32223338]
28. Velikova G, Wright EP, Smith AB, et al. Automated collection of quality-of-life data: a comparison of paper and computer touch-screen questionnaires. J Clin Oncol 1999 Mar;17(3):998-1007. [doi: 10.1200/JCO.1999.17.3.998] [Medline: 10071295]
29. Schougaard LMV, Larsen LP, Jessen A, et al. AmbuFlex: tele-patient-reported outcomes (telePRO) as the basis for follow-up in chronic and malignant diseases. Qual Life Res 2016 Mar;25(3):525-534. [doi: 10.1007/s11136-015-1207-0] [Medline: 26790427]
30. Holch P, Warrington L, Bamforth LCA, et al. Development of an integrated electronic platform for patient self-report and management of adverse events during cancer treatment. Ann Oncol 2017 Sep 1;28(9):2305-2311. [doi: 10.1093/annonc/mdx317] [Medline: 28911065]
31. Aiyegbusi OL. Key methodological considerations for usability testing of electronic patient-reported outcome (ePRO) systems. Qual Life Res 2020 Feb;29(2):325-333. [doi: 10.1007/s11136-019-02329-z] [Medline: 31691202]
32. Jaspers MWM. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009 May;78(5):340-353. [doi: 10.1016/j.ijmedinf.2008.10.002] [Medline: 19046928]
33. 20282-2:2013(en) IT. Usability of consumer products and products for public use — part 2: summative test method. : International Organization for Standardization; 2013 URL: https://www.iso.org/standard/62733.html [accessed 2024-11-29]
34. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003 Aug;35(3):379-383. [doi: 10.3758/bf03195514] [Medline: 14587545]
35. Hwang W, Salvendy G. Number of people required for usability evaluation. Commun ACM 2010 May;53(5):130-133. [doi: 10.1145/1735223.1735255]
36. McMullan C, Hughes SE, Aiyegbusi OL, Calvert M. Usability testing of an electronic patient-reported outcome system linked to an electronic chemotherapy prescribing and patient management system for patients with cancer. Heliyon 2023 Jun;9(6):e16453. [doi: 10.1016/j.heliyon.2023.e16453] [Medline: 37260889]

XSL•FO
RenderX

37. Macefield R. How to specify the participant group size for usability studies: a practitioner's guide. J Usability Stud 2009;5(1):34-45.
38. Aparito. URL: https://www.aparito.com [accessed 2024-12-24]
39. Sapra A, Bhandari P, Sharma S, Chanpura T, Lopp L. Using Generalized Anxiety Disorder-2 (GAD-2) and GAD-7 in a primary care setting. Cureus 2020 May 21;12(5):e8224. [doi: 10.7759/cureus.8224] [Medline: 32582485]
40. Plummer F, Manea L, Trepel D, McMillan D. Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. Gen Hosp Psychiatry 2016;39:24-31. [doi: 10.1016/j.genhosppsych.2015.11.005] [Medline: 26719105]
41. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
42. Lang AJ, Stein MB. An abbreviated PTSD checklist for use as a screening instrument in primary care. Behav Res Ther 2005 May;43(5):585-594. [doi: 10.1016/j.brat.2004.04.005] [Medline: 15865914]
43. Heinemann AW, Kisala PA, Boulton AJ, et al. Development and calibration of the TBI-QOL ability to participate in social roles and activities and TBI-QOL satisfaction with social roles and activities item banks and short forms. Arch Phys Med Rehabil 2020 Jan;101(1):20-32. [doi: 10.1016/j.apmr.2019.07.015] [Medline: 31473208]
44. Eccles DW, Arsal G. The think aloud method: what is it and how do I use it? Qual Res Sport Exerc Health 2017 Aug 8;9(4):514-531. [doi: 10.1080/2159676X.2017.1331501]
45. Aiyegbusi OL, Kyte D, Cockwell P, et al. Development and usability testing of an electronic patient-reported outcome measure (ePROM) system for patients with advanced chronic kidney disease. Comput Biol Med 2018 Oct 1;101:120-127. [doi: 10.1016/j.compbiomed.2018.08.012] [Medline: 30130638]
46. McMullan C, Retzer A, Hughes SE, et al. Development and usability testing of an electronic patient-reported outcome (ePRO) solution for patients with inflammatory diseases in an Advanced Therapy Medicinal Product (ATMP) basket trial. J Patient Rep Outcomes 2023 Oct 9;7(1):98. [doi: 10.1186/s41687-023-00634-3] [Medline: 37812323]
47. Maas AIR, Menon DK, Adelson PD, et al. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. Lancet Neurol 2017 Dec;16(12):987-1048. [doi: 10.1016/S1474-4422(17)30371-X] [Medline: 29122524]
48. Nelson LD, Temkin NR, Dikmen S, et al. Recovery after mild traumatic brain injury in patients presenting to US level I trauma centers: a Transforming Research and Clinical Knowledge in Traumatic Brain Injury (TRACK-TBI) study. JAMA Neurol 2019 Sep 1;76(9):1049-1059. [doi: 10.1001/jamaneurol.2019.1313] [Medline: 31157856]
49. Hochheimer CJ, Sabo RT, Krist AH, Day T, Cyrus J, Woolf SH. Methods for evaluating respondent attrition in web-based surveys. J Med Internet Res 2016 Nov 22;18(11):e301. [doi: 10.2196/jmir.6342] [Medline: 27876687]
50. O'Reilly-Shah VN. Factors influencing healthcare provider respondent fatigue answering a globally administered in-app survey. PeerJ 2017;5:e3785. [doi: 10.7717/peerj.3785] [Medline: 28924502]
51. Calvert MJ, Cruz Rivera S, Retzer A, et al. Patient reported outcome assessment must be inclusive and equitable. Nat Med 2022 Jun;28(6):1120-1124. [doi: 10.1038/s41591-022-01781-8] [Medline: 35513530]
52. Armstrong RA. Visual problems associated with traumatic brain injury. Clin Exp Optom 2018 Nov;101(6):716-726. [doi: 10.1111/cxo.12670] [Medline: 29488253]
53. Triplet JJ, Momoh E, Kurowicki J, Villarroel LD, Law TY, Levy JC. E-mail reminders improve completion rates of patient-reported outcome measures. JSES Open Access 2017 Mar;1(1):25-28. [doi: 10.1016/j.jses.2017.03.002] [Medline: 30675535]
54. Holden RJ, Valdez RS, Schubert CC, Thompson MJ, Hundt AS. Macroergonomic factors in the patient work system: examining the context of patients with chronic illness. Ergonomics 2017 Jan;60(1):26-43. [doi: 10.1080/00140139.2016.1168529] [Medline: 27164171]
55. Usability. Digital.gov. 2013. URL: https://www.usability.gov/get-involved/blog/2013/04/moderating-usability-tests.html [accessed 2024-12-04]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
ePRO: electronic patient-reported outcomes
ePROM: electronic patient-reported outcome measures
GAD-2: Generalized Anxiety Disorder-2 measure
ISO: International Organisation for Standardization
mTBI: mild traumatic brain injury
NIHR SRMRC: National Institute for Health and Care Research Surgical Reconstruction and Microbiology Research Centre
PCL-2: Posttraumatic Stress Disorder Checklist 2
PHQ-2: Patient Health Questionnaire-2
PRiORiTy : Patient-Reported Outcomes Research in Trauma
PRO: patient-reported outcomes

XSL•FO
RenderX

PROM: patient-reported outcome measure
QOLIBRI: Quality of Life after Brain Injury
TBI: traumatic brain injury
TBI QOL SF: Traumatic Brain Injury – Quality of Life Short form

Edited by A Mavragani; submitted 12.03.24; peer-reviewed by H Mühlan, S Juengst; revised version received 08.11.24; accepted 10.11.24; published 23.01.25.

Please cite as:
McMullan C, Turner G, Retzer A, Belli A, Davies EH, Nice L, Flavell L, Flavell J, Calvert M
Testing an Electronic Patient-Reported Outcome Platform in the Context of Traumatic Brain Injury: PRiORiTy Usability Study
JMIR Form Res 2025;9:e58128
URL: https://formative.jmir.org/2025/1/e58128
doi:10.2196/58128

© Christel McMullan, Grace Turner, Ameeta Retzer, Antonio Belli, Elin Haf Davies, Laura Nice, Luke Flavell, Jackie Flavell, Melanie Calvert. Originally published in JMIR Formative Research (https://formative.jmir.org), 23.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Association of Psychological Factors With Willingness to Share Health-Related Data From Technological Devices: Cross-Sectional Questionnaire Study

Marijn Eversdijk, MSc; Emma Rixt Douma, MSc; Mirela Habibovic, PhD; Willem Johan Kop, PhD

Department of Medical and Clinical Psychology, Center of Research on Psychological Disorders and Somatic Diseases (CoRPS), Tilburg University, Tilburg, the Netherlands

Corresponding Author:
Marijn Eversdijk, MSc
Department of Medical and Clinical Psychology, Center of Research on Psychological Disorders and Somatic Diseases (CoRPS), Tilburg University, Tilburg, the Netherlands

Abstract

Background: Health-related data from technological devices are increasingly obtained through smartphone apps and wearable devices. These data could enable physicians and other care providers to monitor patients outside the clinic or assist individuals in improving lifestyle factors. However, the use of health technology data might be hampered by the reluctance of patients to share personal health technology data because of the privacy sensitivity of this information.

Objective: This study investigates to what extent psychological factors play a role in people's willingness to share personal health technology data.

Methods: Data for this cross-sectional study were obtained by quota sampling based on age and sex in a community-based sample (N=1013; mean age 48.6, SD 16.6 years; 522/1013, 51.5% women). Willingness to share personal health technology data and related privacy concerns were assessed using an 8-item questionnaire with good psychometric properties (Cronbach's α=0.82). Psychological variables were assessed using validated questionnaires for optimism (Life Orientation Test—Revised), psychological flexibility (Psychological Flexibility Questionnaire), negative affectivity (Type D Scale-14—Negative Affectivity), social inhibition (Type D Scale-14—Social Inhibition), generalized anxiety (Generalized Anxiety Disorder-7), and depressive symptoms (Patient Health Questionnaire-9). Data were analyzed using multiple linear regression analyses, and network analysis was used to visualize the associations between the item scores.

Results: Higher levels of optimism (β=.093; *P*=.004) and psychological flexibility (β=.127; *P*<.001) and lower levels of social inhibition (β=−.096; *P*=.002) were significantly associated with higher levels of willingness to share health technology data when adjusting for age, sex, and education level in separate regression models. Other associations with psychological variables were not statistically significant. Network analysis revealed that psychological flexibility clustered more with items that focused on the benefits of sharing data, while optimism was negatively associated with privacy concerns.

Conclusions: The current results suggest that people with higher levels of optimism and psychological flexibility and those with lower social inhibition levels are more likely to share health technology data. The magnitude of the effect sizes was low, and future studies with additional psychological measures are needed to establish which factors identify people who are reluctant to share their data such that optimal use of devices in health care can be facilitated.

(JMIR Form Res 2025;9:e64244) doi:10.2196/64244

KEYWORDS

health data sharing; privacy concerns; wearable health technology; personality; psychological flexibility; optimism; social inhibition; psychological factors; willingness; health-related data; mobile phone

Introduction

Recent advances in health technology, including wearable devices and smartphone apps, have facilitated remote patient health monitoring [1-3]. Sharing data from these devices with care providers, researchers, or device developers has multiple advantages for the patient, including early detection of health-related abnormalities, risk stratification, individualized health interventions, and overall better disease management [4,5]. The health care system might also benefit by applying machine learning algorithms to the large amounts of collected data to improve future health care and increase cost-efficiency [6].

Despite the potential health benefits, uptake of these health technologies is often impeded by patient and other user concerns related to the privacy and confidentiality of sharing sensitive

XSL•FO
RenderX

personal data [7-9]. These privacy-related concerns are relevant because they influence the willingness to share their health technology data [10,11]. While respect for personal autonomy in the decision to share data is an important cornerstone in medical ethics, privacy-related objections could be overcome by adjusting the technological features of a device based on specific desires or patient characteristics [12]. Assessing individual differences in willingness to share technology data could help develop a personalized approach toward sharing health data to facilitate lower privacy concerns, higher user satisfaction, higher satisfaction with care, and better health outcomes.

Willingness to share personal health technology data can be defined as an individual tendency to allow others to have access to personal biobehavioral (eg, heart rate) or other personal data (eg, geographic location) collected using health technology (eg, wearable devices and smartphone apps) [11,13,14]. In general, the extent to which a person is willing to share or protect personal data is determined by an interaction between the sensitivity of the information (eg, medical or socioeconomic information, purchase history, and location), the context of data sharing (eg, health care or commercial setting), and personal dispositions such as privacy-related attitudes [15]. These personal dispositions are most relevant for the question of who is willing to share personal health technology data. Research has identified not only digital literacy, older age, and privacy knowledge as factors associated with willingness to share technology data but also psychological factors such as personality traits and generalized trust [13,16,17]. However, there is a knowledge gap regarding the magnitude of associations between psychological factors and willingness to share personal health technology data.

Current evidence regarding positive psychological factors, although scarce, indicates that patients who are overall more optimistic are more likely to expect positive outcomes from sharing their data (eg, better health monitoring) or are more likely to believe that they are less at risk of privacy breaches than others [18,19]. Optimism is also associated with higher psychological flexibility, which is the ability to distance from current mindsets and consider other possible mindsets [20,21]. People with high levels of psychological flexibility might be better able to act upon goals that involve new technologies to facilitate better personalized care that might also help other patients by sharing data even when this comes at the cost of less privacy for the user.

In addition to positive psychological traits, negative psychological factors are also expected to be associated with the willingness to share health-related data. Experiencing distress (eg, anxiety and depressive symptoms) is linked to threat-avoidant behaviors and reduced engagement in reward-seeking behaviors [22]. These tendencies might translate into overvaluing the risks and undervaluing the potential rewards of sharing data, leading to a lower willingness to share health technology data. Prior research has shown that anxiety and depressive symptoms are more likely to occur in patients with negative personality traits such as negative affectivity (NA) and social inhibition (SI) [23]. The relative importance of positive psychological characteristics versus negative characteristics with the willingness to share health technology data has yet to be researched.

This study explores associations between psychological factors and the willingness to share personal health technology data. It was expected that positive psychological characteristics (ie, optimism and psychological flexibility) would be associated with a higher willingness to share health technology data, whereas negative psychological characteristics (ie, anxiety, depressive symptoms, and measures of distress-related personality traits: NA and SI) would be associated with less willingness to share health technology data. The role of sociodemographic background factors (age, sex, and education level) and whether or not participants have previously used health technology apps was also investigated using multivariable models. Exploratory network analyses on the questionnaire items were conducted to reveal potential psychological mechanisms between the participant responses on the psychological constructs and items that reflect the willingness to share data.

Methods

Participants and Procedure

This cross-sectional study was based on data obtained from a Dutch community-based sample (N=1013) as part of an annually recurring survey study (for additional details, see the study by Kop et al [24]). Research assistants were instructed to apply quota sampling within their personal network by approaching a minimum number of participants from each sex and age group to ensure equal representation (eg, to ensure the number of men aged between 31 and 40 years in the total sample is equal to the number of women aged between 31 and 40 years in the total sample). Inclusion criteria were age above 18 years and being fluent in Dutch. Although there were no geographical restrictions, most participants lived in the south of the Netherlands, as a result of the sampling procedure. Data for this study were collected between February 2022 and April 2022. Data collection took place online using Qualtrics.

Ethical Considerations

Ethical approval was obtained from the Tilburg School of Social and Behavioral Sciences ethical review board (protocol RP55). All participants gave informed consent before taking part in this study, and their data were deidentified prior to data analysis. Participants were not compensated for their participation.

Measures

Data collection was conducted using a web-based survey after participants provided informed consent for their participation. All data were obtained using validated questionnaires or purpose-designed questions (see Multimedia Appendix 1 for an overview of the used questionnaires).

Health Technology Data-Sharing Inventory

The willingness to share health technology data was assessed using an 8-item questionnaire designed for the purposes of this study. Six questions focused on the potential benefits of sharing data (eg, “I would allow my personal data to be shared if it would help me”). Two items were derived from the Service

XSL•FO
RenderX

User Technology Acceptability Questionnaire and focused on the feeling that health technology violates confidentiality or people's privacy [25]. Participants were asked to indicate their level of agreement on a 5-point Likert scale (1=strongly disagree, 5=strongly agree). After reverse-coding the 2 negatively phrased items, a continuous total score was calculated by adding up the items (range 8-40). A higher score indicates a higher willingness to share health technology data. The Cronbach's α coefficient for this scale was 0.82 in this study, which indicates a good internal consistency of this inventory.

Optimism

Optimism was measured using the 10-item version of the Life Orientation Test—Revised (LOT-R) [26]. The LOT-R includes 3 items about optimism, 3 items about pessimism, and 4 filler items. Examples of questions are "I'm always optimistic about my future" (ie, optimism) and "I hardly ever expect things to go my way" (ie, pessimism). Participants were asked to indicate their level of agreement on a 5-point Likert scale (0=strongly disagree, 4=strongly agree). A continuous total score for optimism (range 0 - 24) was calculated by adding the scores on optimism to the reversed scores on pessimism. A higher score indicates higher levels of optimism. The published internal consistency of the LOT-R total score is acceptable (Cronbach's α=0.78) [26], which was similar and acceptable in the present sample (Cronbach's α=0.74).

Psychological Flexibility

The Psychological Flexibility Questionnaire (PFQ) [27] was used to measure psychological flexibility. Participants were asked to indicate their degree of agreement on a 5-point Likert scale (1=strongly disagree, 5=strongly agree), enabling participants to enter a neutral response (3=neither agree nor disagree). Examples of items are "I feel open to changes" and "Concepts may possess different meanings when perceived in different contexts." The answers to all 20 questions were added up to a continuous total score (range 20-100). A higher score indicates higher psychological flexibility. Internal consistency of the questionnaire was good in the present sample (Cronbach's α=0.89).

Negative Affectivity and Social Inhibition

NA and SI were measured using the DS-14 (Type D Scale-14) [28]. These scales have been developed to assess a construct identifying individuals with a distressed personality trait (ie, type D), but the subscales are used as separate dimensions in this paper, as the focus was not on the interaction of these 2 subscales, which is needed when investigating the type D construct [29]. The questionnaire contains a 7-item subscale assessing NA and a 7-item scale assessing SI. An example of a question for NA is "I often feel unhappy," and for SI, an example is "I am a closed kind of person." Two questions of the SI subscale are reverse-phrased and recoded before calculating total scores. Participants rated their agreement on a 5-point Likert scale (0=strongly disagree, 4=strongly agree). A continuous total score was calculated by adding the scores on each subscale (range 0 - 28). A higher score indicates a higher presence of the personality trait. The Cronbach's α value for NA in this sample was 0.88 and for SI 0.88, indicating good internal consistency.

Generalized Anxiety and Depressive Symptoms

The Generalized Anxiety Disorder-7 (GAD-7) questionnaire [30] was used to measure the presence of anxiety-related complaints over the last 2 weeks (example items are "Feeling nervous, anxious or on edge" and "Not being able to stop or control worrying"). Participants were asked to indicate the presence of these complaints on a 4-point Likert scale (0=not at all, 3=nearly every day). A continuous total score was calculated by adding the scores of the 7 items (range 0 - 21). A higher score indicates a higher presence of anxiety-related complaints. The Cronbach's α value in this study was 0.87, indicating good internal consistency.

Depressive symptoms were assessed using the Patient Health Questionnaire-9 (PHQ-9) [31], asking about the presence of depressive symptoms such as "Little interest or pleasure in doing things" or "Feeling down, depressed, or hopeless." Questions focused on the presence of complaints over the past 2 weeks and had to be filled in on a 4-point Likert scale (0=not at all, 3=nearly every day). A continuous total score was calculated by adding the scores on all 9 items (range 0-27). A higher score indicates a higher presence of depressive symptoms. Internal consistency was considered good, with a Cronbach's α value of 0.85 in the present sample.

Background Variables and Covariates

Demographic measures (age, self-reported sex, and education level) were obtained from the self-report questionnaire. The dataset contained only a measurement for sex and not for gender. Although gender would have better matched the psychological and behavioral variables in the dataset, the authors decided to report sex since this measure better represents the construct measured in the questionnaire. Level of education was divided into low (completed secondary school or less), middle (completed vocational education or high school only), or high (completed college or higher). Participants also reported their health status by indicating the presence of a lifetime diagnosed medical condition (eg, cardiovascular, cancer, lung disease, gastrointestinal disorder, and diabetes mellitus). In addition, an assessment was made concerning participants' use of smartphone-based health apps or related wearable technology.

Statistical Analysis

Descriptive statistics were obtained for all demographic and psychological variables by calculating the mean and SD for continuous variables and the number of occurrences and percentages for all categorical variables. Pearson and nonparametric Spearman correlation analyses were used to obtain an overview of the bivariate associations between the psychological measures and willingness to share health technology data. For illustrative purposes, odds ratios (ORs) with 95% CIs were also calculated using tertile-based dichotomized data of the psychological measures and the willingness to share health data (a value higher than 1 indicates a higher chance of willingness to share data based on the upper tertile of the psychological measure).

XSL•FO
RenderX

Multiple linear regression analyses were performed to assess which variables were independently associated with the willingness to share health technology data. Assumptions for regression analysis, including linearity, normality of residuals, homoscedasticity, and the presence of influential cases, were evaluated using the plot function in R (version 4.2.2; R Foundation for Statistical Computing) (findings are shown in Table S1 and Figure S1 in Multimedia Appendix 2). Data are presented in terms of the overall explained variance of the model (R^2) and the standardized regression weights per variable (β). Analyses first focused on the results of the full sample (N=1013). Subsequent sensitivity analyses were conducted for participants who had used or currently used health technology apps (N=244) and those not using such apps (N=769).

Total scores were coded as missing values if participants answered <60% of all items of that specific scale. If at least 60% of the items of a questionnaire were completed, then missing items of that scale were imputed by the mean of the answered questions of that scale. The percentage of missing values among all the psychological item scores was <0.1%, and there appeared to be no patterns in missingness. For the multiple regression models, listwise deletion was used to deal with missing values on the total scores of each predictor. The potential of biased parameters estimated as a result of mean imputation was limited, as it was applied only within variables, the percentage of missing values was extremely low, and listwise deletion was used [32].

Exploratory network analyses were performed to further investigate the associations between the item scores of all the psychological scales. Network analyses were visualized with the "qgraph" package (version 1.9.3) in R [33]. Each node represents an individual item of the questionnaire, while edges represent the associations between 2 nodes (ie, questionnaire items).

Results

Participants and Demographics

Table 1 displays the participant characteristics. The mean age was 48.6 (SD 16.6) years, and 51.5% (522/1013) of the participants were women. Levels of education were relatively high (52.3% higher education). A current or history of a medical disorder was reported by 36.9% (374/1013) of the participants, and average anxiety and depressive symptom levels were relatively low.

Table . Characteristics of all included participants.

	Total sample (N=1013)	Technology users (N=244)	Not using health technology (N=769)	*P* value
Age (years), mean (SD)	48.6 (16.6)	46.3 (15.0)	49.4 (17.1)	.01
Sex (female), n (%)	522 (51.5)	127 (52.0)	395 (51.4)	.91
Education, n (%)				<.001
Low	129 (12.7)	26 (10.7)	103 (13.4)	
Middle	354 (34.9)	65 (26.6)	289 (37.6)	
High	530 (52.3)	153 (62.7)	377 (49.0)	
Lifetime medical disorder, n (%)	374 (36.9)	94 (38.5)	280 (36.4)	.60
Variables (range), mean (SD)				
Optimism (0-24)	15.5 (3.5)	16.0 (3.5)	15.4 (3.4)	.02
Psychological flexibility (20-100)	75.1 (8.6)	76.6 (9.1)	74.6 (8.4)	.002
Negative affectivity (0-28)	7.7 (5.2)	7.2 (4.8)	7.9 (5.4)	.09
Social inhibition (0-28)	8.7 (5.3)	8.7 (5.1)	8.7 (5.4)	.87
Generalized anxiety (0-21)	3.0 (3.4)	3.0 (3.3)	2.9 (3.4)	.77
Depressive symptoms (0-27)	3.5 (3.9)	3.4 (3.2)	3.6 (4.1)	.42
Health data sharing (8-40)	23.8 (6.0)	25.0 (5.6)	23.4 (6.0)	<.001

The variable willingness to share health technology was approximately normally distributed, with scores ranging from 8 to 40 (mean 23.8, SD 6.0, median 24.0) (Figure S2 in Multimedia Appendix 2). Furthermore, 24.1% (244/1013) of the participants were currently using digital health apps or other forms of health technology (ie, wearable devices or smartphone apps aimed at improving physical activity, sleep, and medication adherence, or at reducing stress, alcohol use, and smoking). Participants who used technological devices were younger, were more likely to have a high education level, and had higher psychological flexibility levels than participants who did not use such devices, whereas no differences were found on the other variables listed in Table 1.

Associations of Psychological Factors With Willingness to Share Health Technology Data

As shown in Table 2, high levels of optimism (r=0.07; P=.03) and psychological flexibility (r=0.13; P<.001) were significantly

XSL•FO
RenderX

correlated with higher levels of willingness to share health technology data. In addition, higher levels of SI were associated with a lower willingness to share health technology data (r=−0.08; P=.01). Other bivariate correlations between psychological factors and willingness to share data were nonsignificant. When repeating these analyses with dichotomized scores, it was found that participants scoring in the highest tertile of psychological flexibility were 63% more likely to score high (upper tertile) on willingness to share health technology data (OR 1.63, 95% CI 1.25-2.12). Participants scoring in the lowest tertile of SI were also more likely to be willing to share their data (OR 1.53, 95% CI 1.17-2.00), whereas none of the other dichotomized psychological measures revealed significant ORs with a high level of willingness to share data.

Table . Correlation and regression coefficients of a linear regression model with all psychological measures and willingness to share health technology data.

Variable	Spearman ρ	*P* value	Pearson *r*	*P* value	*B* (95% CI)	β	*P* value
Optimism (LOT-R[a])	0.07	.03	0.07	.03	0.094 (−0.043 to 0.231)	.054	.18
Psychological flexibility (PFQ[b])	0.13	<.001	0.13	<.001	0.067 (0.017 to 0.117)	.097	.008
Negative affectivity (DS-14: NA[c])	−0.04	.20	−0.04	.26	0.050 (−0.067 to 0.166)	.043	.41
Social inhibition (DS-14: SI[d])	−0.10	.002	−0.08	.01	−0.068 (−0.149 to 0.013)	−.061	.10
Generalized anxiety (GAD-7[e])	0.01	.80	0.01	.84	0.115 (−0.072 to 0.301)	.065	.23
Depressive symptoms (PHQ-9[f])	−0.01	.71	−0.02	.62	−0.097 (−0.257 to 0.063)	−.063	.23
Age	−0.08	.01	−0.09	.004	−0.031 (−0.055 to −0.007)	−.086	.01
Sex[g]	0.07	.04	0.07	.03	0.873 (0.124 to 1.622)	.073	.02
Education level[h]	−0.06	.05	−0.06	.06	−0.629 (−1.160 to −0.096)	−.074	.02

[a]LOT-R: Life Orientation Test—Revised.

[b]PFQ: Psychological Flexibility Questionnaire.

[c]DS-14 NA: Type D Scale-14 Negative Affectivity

[d]DS-14 SI: Type D Scale-14 Social Inhibition

[e]GAD-7: Generalized Anxiety Disorder-7.

[f]PHQ-9: Patient Health Questionnaire-9.

[g]Sex coded as 0 (female) and 1 (male; a positive value indicates that men have higher willingness to share health technology data than women).

[h]Education level coded as 0 (low), 1 (middle), and 2 (high; a positive value indicates that a higher education is associated with a higher willingness to share health technology data).

Of the background factors, male participants (Cohen d=−0.14; P=.03) and technology users (Cohen d=−0.25; P<.001) scored significantly higher on willingness to share technology data (Table S2 in Multimedia Appendix 2). When adjusting for age, sex, and education level in separate linear regression models, optimism (β=.093; P=.004), psychological flexibility (β=.127; P<.001), and SI (β=−.096; P=.002) remained significantly associated with willingness to share health technology data.

To evaluate which psychological factor was associated with willingness to share health technology data, independent of other psychological factors, a multiple linear regression analysis was used entering all psychological factors simultaneously while also adjusting for age, sex, and education level (adjusted R^2=0.03; P<.001). Higher psychological flexibility was the psychological variable that was independently associated with more willingness to share data (β=.097; P=.008) while adjusting for covariates. Other correlates of willingness to participate in health technology data sharing were younger age (β=−.086; P=.01), male sex (β=.073; P=.02), and lower education level (β=−.074; P=.02) (Table 2).

When repeating the analyses after stratifying the sample by the use of health technology, a similar pattern of results was found for both groups. Among health technology users (N=244), higher levels of psychological flexibility were independently associated with more willingness to share data (β=.204; P=.006). Other associations with the willingness to share data among technology

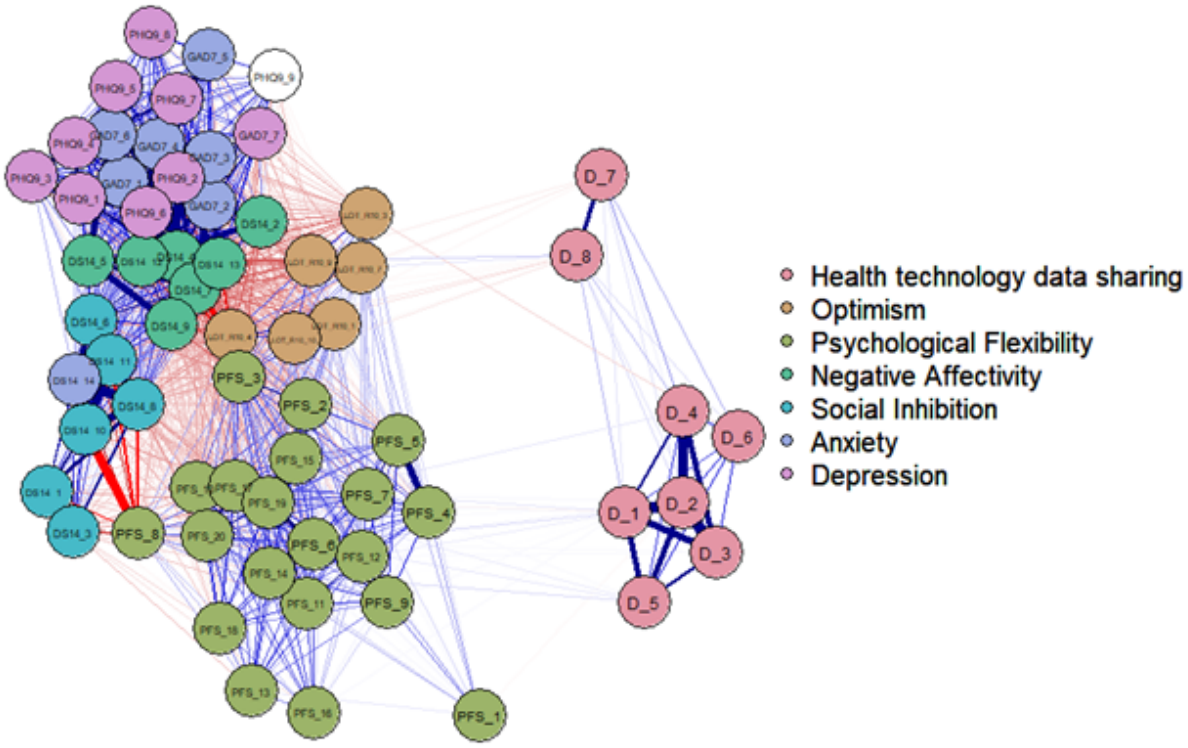
XSL•FO
RenderX

users were nonsignificant (adjusted R^2 for the full model=0.04; *P*=.03). In the subset of participants who did not use health technology (N=769), lower SI was associated with a higher willingness to share data (β=−.092; *P*=.03), while male sex (β=.076; *P*=.04) and lower education level (β=−.089; *P*=.02) were also independently associated with willingness to share data (Table S3 in Multimedia Appendix 2).

Network Analysis of the Interrelationship Between Psychological Factors and Willingness to Share Health Technology Data

Figure 1 shows the results of the network analysis based on the correlational structure of all questionnaire items. Every edge represents a correlation larger than 0.1. Each underlying construct is represented by a different color. The network visualizes the extent to which items cluster together, indicating the measurement of the same underlying construct. The network visualizes that the items on data sharing do not cluster together with most of the psychological predictors. Within the health technology data-sharing items, it appeared that items 7 ("I am worried about the confidentiality of the private information being exchanged by the health technology") and 8 ("Health technology will violate my privacy") were distinct from the other items (eg, "I would allow my personal data to be shared with my physician if it would improve my treatment"), based on their relative spatial position. The item-level comparisons show that items 7 and 8 cluster more toward items on optimism, while the other data-sharing items cluster more toward psychological flexibility. A regularized network analysis can be found in Figure S3 in Multimedia Appendix 2.

Figure 1. Network correlation structure of all psychological measurements and willingness to share health technology data.



Discussion

Principal Findings

This study shows that higher levels of optimism and psychological flexibility and lower levels of SI are associated with higher levels of willingness to share health technology data. Multivariable analyses further indicated that psychological flexibility was the strongest and independent psychological factor associated with the willingness to share health technology data. Younger age, male sex, and lower education level were also independently associated with a higher willingness to share health technology data. Item-level network analysis revealed that optimism was associated with less concerns about confidentiality and privacy, whereas psychological flexibility was associated with a higher willingness to share health technology data if it could benefit the user or others.

Integration of the Present Findings With the Existing Literature

The current findings are consistent with previous research indicating that positive psychological traits, such as optimism and trust, are associated with a higher willingness to share health data [17,19]. The association of psychological flexibility with willingness to share health technology data has not yet been described in the scientific literature. People high on psychological flexibility are able to adapt to situational demands and balance competing desires, needs, and life domains [20].

XSL•FO
RenderX

Furthermore, they are more open toward new forms of behavior or shifting behavioral repertoires to achieve goals related to their personal values, which could play a central role in the decision to share personal health technology data with others. When balancing between giving up a part of their privacy compared and helping themselves or others by sharing their data, their prosocial values can take precedence over their protective intuitions, resulting into a higher willingness to share data. By shifting behaviors toward sharing data, people accept potential privacy concerns and remain true to personal values such as helping others, pursuing optimal health outcomes, or other values related to the benefits of data sharing [34].

Limitations and Strengths

The findings of this study need to be interpreted in the context of several limitations. The questionnaire to evaluate willingness to share health technology data has not been validated in prior settings, although the psychometric properties of the questionnaire were good. There was a lack of diversity in the sample with regard to education level and ethnic background. The levels of generalized anxiety and depressive symptoms were also low in this community-based sample, potentially resulting in a floor effect for their relationship with willingness to share health data. The cross-sectional design precludes causal inferences and a more in-depth qualitative research approach, such as conducting individual interviews or focus groups, might provide additional information about the underlying values and arguments in the decision to share health technology data. However, even with additional knowledge on these motivations, there is a gap between data-sharing intention and action [35]. This observation indicates the need for future studies measuring actual privacy-related behaviors instead of evaluating intentions to engage in such behaviors. A strength of this study is the large sample covering a wide range of ages and an equal number of women and men, in which the relative importance of multiple psychological constructs could be studied. However, the set of psychological constructs was limited in the already existing dataset, and the models indicated very small explained variance, which pleads for an extension of other relevant psychological factors, such as technological self-efficacy, performance expectancy, trust, or other personality traits [14,17,36-39].

Conclusions and Future Directions

This study documents a significant association of optimism, psychological flexibility, and low SI with a higher likelihood of individuals being willing to share health technology data. Of these variables, psychological flexibility appeared to be the strongest factor, particularly among individuals who use health technology. In addition, younger age, male sex, and low education levels were associated with a higher willingness to share health technology data, particularly among people who currently do not use health technology. Examining the underlying network structure of a large set of psychological variables revealed that psychological flexibility is mainly associated with the benefits of sharing health data, while people low on optimism show more concerns about confidentiality and privacy. Still, the magnitude of effect sizes was low, and additional work is needed on the underlying psychological mechanisms and reasons behind the disclosure and sharing of health-related data in general and health technology data in particular. Future studies that investigate psychological factors with reasons for sharing health data could identify which specific reasons are important for whom on an individual level instead of a population level. Additional studies with a broader set of psychological variables should be conducted to identify those variables that contribute the most to the willingness to share health data before specific clinical implications can be formulated. To increase the explained variance of the regression models, a broader set of psychological measures should be considered in studying their relationship with the willingness to share health technology data. Findings from these studies would provide important building blocks for the design of future health technology and promote individually tailored functionalities. This would enable customization of the data-sharing settings of personal health technology data sharing in accordance with individual preferences, rather than a binary decision to share all data or no data at all. The possibility to retract or modify prior consent to health technology data sharing with minimal effort could also increase people's initial participation in data sharing [40]. This personalized approach might result in less privacy concerns, higher user satisfaction with the technology, higher satisfaction with health care, and potentially also better health outcomes.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of questionnaire items used for analysis.
[DOCX File, 26 KB - formative_v9i1e64244_app1.docx]

Multimedia Appendix 2
Supplemental tables and figures with further data on continuous psychological variables, dichotomous background factors, willingness to share data, and the regularized correlation structure of all scales.
[DOCX File, 147 KB - formative_v9i1e64244_app2.docx]

XSL•FO
RenderX

References

1. Afshin A, Babalola D, Mclean M, et al. Information technology and lifestyle: a systematic evaluation of internet and mobile interventions for improving diet, physical activity, obesity, tobacco, and alcohol use. JAHA 2016 Sep;5(9). [doi: 10.1161/JAHA.115.003058]
2. Chaudhry B, Wang J, Wu S, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752. [doi: 10.7326/0003-4819-144-10-200605160-00125] [Medline: 16702590]
3. Habibović M, Rollman B. Technological innovations in biobehavioral and psychosomatic medicine. Psychosom Med 2023 Sep 1;85(7):565-567. [doi: 10.1097/PSY.0000000000001240] [Medline: 37506296]
4. Banerjee S, Hemphill T, Longstreet P. Wearable devices and healthcare: data sharing and privacy. Inf Soc 2018 Jan 1;34(1):49-57. [doi: 10.1080/01972243.2017.1391912]
5. Bayoumy K, Gaber M, Elshafeey A, et al. Smart wearable devices in cardiovascular care: where we are and how to move forward. Nat Rev Cardiol 2021 Aug;18(8):581-599. [doi: 10.1038/s41569-021-00522-7] [Medline: 33664502]
6. Bauer M, Glenn T, Monteith S, Bauer R, Whybrow PC, Geddes J. Ethical perspectives on recommending digital technology for patients with mental illness. Int J Bipolar Disord 2017 Dec;5(1):6. [doi: 10.1186/s40345-017-0073-9] [Medline: 28155206]
7. Martinez-Martin N, Luo Z, Kaushal A, et al. Ethical issues in using ambient intelligence in health-care settings. Lancet Digit Health 2021 Feb;3(2):e115-e123. [doi: 10.1016/S2589-7500(20)30275-2] [Medline: 33358138]
8. Segura Anaya LH, Alsadoon A, Costadopoulos N, Prasad PWC. Ethical implications of user perceptions of wearable devices. Sci Eng Ethics 2018 Feb;24(1):1-28. [doi: 10.1007/s11948-017-9872-8] [Medline: 28155094]
9. Baines R, Stevens S, Austin D, et al. Patient and public willingness to share personal health data for third-party or secondary uses: systematic review. J Med Internet Res 2024 Mar 5;26:e50421. [doi: 10.2196/50421] [Medline: 38441944]
10. Wetzels M, Broers E, Peters P, Feijs L, Widdershoven J, Habibovic M. Patient perspectives on health data privacy and management: “Where Is my data and whose is it?”. Int J Telemed Appl 2018;2018:3838747. [doi: 10.1155/2018/3838747] [Medline: 30631347]
11. Kim TK, Choi M. Older adults’ willingness to share their personal and health information when adopting healthcare technology and services. Int J Med Inform 2019 Jun;126:86-94. [doi: 10.1016/j.ijmedinf.2019.03.010] [Medline: 31029268]
12. Rubeis G, Schochow M, Steger F. Patient autonomy and quality of care in telehealthcare. Sci Eng Ethics 2018 Feb;24(1):93-107. [doi: 10.1007/s11948-017-9885-3]
13. Karampela M, Ouhbi S, Isomursu M. Connected health user willingness to share personal health data: questionnaire study. J Med Internet Res 2019;21(11):e14537. [doi: 10.2196/14537]
14. Rising CJ, Gaysynsky A, Blake KD, Jensen RE, Oh A. Willingness to share data from wearable health and activity trackers: analysis of the 2019 Health Information National Trends Survey Data. JMIR Mhealth Uhealth 2021 Dec 13;9(12):e29190. [doi: 10.2196/29190] [Medline: 34898448]
15. Brough AR, Martin KD. Critical roles of knowledge and motivation in privacy research. Curr Opin Psychol 2020 Feb;31:11-15. [doi: 10.1016/j.copsyc.2019.06.021] [Medline: 31376581]
16. Baruh L, Secinti E, Cemalcilar Z. Online privacy concerns and privacy management: a meta-analytical review. J Commun 2017 Feb;67(1):26-53. [doi: 10.1111/jcom.12276]
17. Bansal G, Zahedi FM, Gefen D. Do context and personality matter? Trust and privacy concerns in disclosing private information online. Inf Manag 2016 Jan;53(1):1-21. [doi: 10.1016/j.im.2015.08.001]
18. Kokolakis S. Privacy attitudes and privacy behaviour: a review of current research on the privacy paradox phenomenon. Comput Secur 2017 Jan;64:122-134. [doi: 10.1016/j.cose.2015.07.002]
19. Cho H, Lee JS, Chung S. Optimistic bias about online privacy risks: testing the moderating effects of perceived controllability and prior experience. Comput Human Behav 2010 Sep;26(5):987-995. [doi: 10.1016/j.chb.2010.02.012]
20. Kashdan TB, Rottenberg J. Psychological flexibility as a fundamental aspect of health. Clin Psychol Rev 2010 Nov;30(7):865-878. [doi: 10.1016/j.cpr.2010.03.001] [Medline: 21151705]
21. Demirtaş AS. Optimism and happiness in undergraduate students: cognitive flexibility and adjustment to university life as mediators. Anales de Psicología 2020;36(2):320-329. [doi: 10.6018/analesps.381181]
22. Bishop SJ, Gagne C. Anxiety, depression, and decision making: a computational perspective. Annu Rev Neurosci 2018 Jul 8;41:371-388. [doi: 10.1146/annurev-neuro-080317-062007] [Medline: 29709209]
23. Mols F, Denollet J. Type D personality among noncardiovascular patient populations: a systematic review. Gen Hosp Psychiatry 2010 Jan;32(1):66-72. [doi: 10.1016/j.genhosppsych.2009.09.010]
24. Kop WJ, Toussaint A, Mols F, Löwe B. Somatic symptom disorder in the general population: associations with medical status and health care utilization using the SSD-12. Gen Hosp Psychiatry 2019;56:36-41. [doi: 10.1016/j.genhosppsych.2018.10.004] [Medline: 30578985]
25. Hirani SP, Rixon L, Beynon M, et al. Quantifying beliefs regarding telehealth: Development of the Whole Systems Demonstrator Service User Technology Acceptability Questionnaire. J Telemed Telecare 2017 May;23(4):460-469. [doi: 10.1177/1357633X16649531]

XSL•FO
RenderX

26. Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. J Pers Soc Psychol 1994 Dec;67(6):1063-1078. [doi: 10.1037//0022-3514.67.6.1063] [Medline: 7815302]
27. Ben-Itzhak S, Bluvstein I, Maor M. The Psychological Flexibility Questionnaire (PFQ): development, reliability and validity. WebmedCentral Psychology 2014;5(4):WMC004606.
28. Denollet J. DS14: standard assessment of negative affectivity, social inhibition, and type D personality. Psychosom Med 2005;67(1):89-97. [doi: 10.1097/01.psy.0000149256.81953.49] [Medline: 15673629]
29. Lodder P. Modeling synergy: how to assess a Type D personality effect. J Psychosom Res 2020 May;132:109990. [doi: 10.1016/j.jpsychores.2020.109990] [Medline: 32151819]
30. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
31. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
32. Lodder P. To impute or not impute: that's the question. In: Mellenbergh GJ, Adler HJ, editors. Advising on Research Methods: Selected Topics 2013: Johannes van Kessel Publishing; 2014.
33. Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D. qgraph: network visualizations of relationships in psychometric data. J Stat Softw 2012;48(4). [doi: 10.18637/jss.v048.i04]
34. Kalkman S, van Delden J, Banerjee A, Tyl B, Mostert M, van Thiel G. Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. J Med Ethics 2022 Jan;48(1):3-13. [doi: 10.1136/medethics-2019-105651] [Medline: 31719155]
35. Kim J, Im E, Kim H. From intention to action: the factors affecting health data sharing intention and action. Int J Med Inform 2023 Jul;175:105071. [doi: 10.1016/j.ijmedinf.2023.105071]
36. Chandrasekaran R, Katthula V, Moustakas E. Too old for technology? Use of wearable healthcare devices by older adults and their willingness to share health data with providers. Health Informatics J 2021;27(4):14604582211058073. [doi: 10.1177/14604582211058073] [Medline: 34802315]
37. Tolentino DA, Costa DK, Jiang Y. Determinants of American adults' use of digital health and willingness to share health data to providers, family, and social media: a cross-sectional study. Comput Inform Nurs 2023 Nov 1;41(11):892-902. [doi: 10.1097/CIN.0000000000001025] [Medline: 37310724]
38. Cherif E, Bezaz N, Mzoughi M. Do personal health concerns and trust in healthcare providers mitigate privacy concerns? Effects on patients' intention to share personal health data on electronic health records. Soc Sci Med 2021 Aug;283:114146. [doi: 10.1016/j.socscimed.2021.114146] [Medline: 34242891]
39. Bansal G, Zahedi FM, Gefen D. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. Decis Support Syst 2010 May;49(2):138-150. [doi: 10.1016/j.dss.2010.01.010]
40. Budin-Ljøsne I, Teare HJA, Kaye J, et al. Dynamic consent: a potential solution to some of the challenges of modern biomedical research. BMC Med Ethics 2017 Jan 25;18(1):4. [doi: 10.1186/s12910-016-0162-9] [Medline: 28122615]

Abbreviations

DS-14: Type D Scale-14
GAD-7: Generalized Anxiety Disorder-7
LOT-R: Life Orientation Test—Revised
NA: negative affectivity
OR: odds ratio
PFQ: Psychological Flexibility Questionnaire
PHQ-9: Patient Health Questionnaire-9
SI: social inhibition

Edited by A Mavragani; submitted 12.07.24; peer-reviewed by J Odumegwu; revised version received 02.12.24; accepted 02.12.24; published 23.01.25.

Please cite as:
Eversdijk M, Douma ER, Habibovic M, Kop WJ
The Association of Psychological Factors With Willingness to Share Health-Related Data From Technological Devices: Cross-Sectional Questionnaire Study
JMIR Form Res 2025;9:e64244
URL: https://formative.jmir.org/2025/1/e64244
doi:10.2196/64244

XSL•FO
RenderX

© Marijn Eversdijk, Emma Rixt Douma, Mirela Habibovic, Willem Johan Kop. Originally published in JMIR Formative Research (https://formative.jmir.org), 23.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

AI Machine Learning–Based Diabetes Prediction in Older Adults in South Korea: Cross-Sectional Analysis

Hocheol Lee, PhD; Myung-Bae Park, PhD; Young-Joo Won, PhD

Department of Health Administration, College of Software and Digital Healthcare Convergence, Yonsei University, Changjogwan, Yonseidae-gil 1, Wonju, Republic of Korea

Corresponding Author:
Young-Joo Won, PhD
Department of Health Administration, College of Software and Digital Healthcare Convergence, Yonsei University, Changjogwan, Yonseidae-gil 1, Wonju, Republic of Korea

Abstract

Background: Diabetes is prevalent in older adults, and machine learning algorithms could help predict diabetes in this population.

Objective: This study determined diabetes risk factors among older adults aged ≥60 years using machine learning algorithms and selected an optimized prediction model.

Methods: This cross-sectional study was conducted on 3084 older adults aged ≥60 years in Seoul from January to November 2023. Data were collected using a mobile app (Gosufit) that measured depression, stress, anxiety, basal metabolic rate, oxygen saturation, heart rate, and average daily step count. Health coordinators recorded data on diabetes, hypertension, hyperlipidemia, chronic obstructive pulmonary disease, percent body fat, and percent muscle. The presence of diabetes was the target variable, with various health indicators as predictors. Machine learning algorithms, including random forest, gradient boosting model, light gradient boosting model, extreme gradient boosting model, and k-nearest neighbors, were employed for analysis. The dataset was split into 70% training and 30% testing sets. Model performance was evaluated using accuracy, precision, recall, F1 score, and area under the curve (AUC). Shapley additive explanations (SHAPs) were used for model interpretability.

Results: Significant predictors of diabetes included hypertension (χ^2_1=197.294; P<.001), hyperlipidemia (χ^2_1=47.671; P<.001), age (mean: diabetes group 72.66 years vs nondiabetes group 71.81 years), stress (mean: diabetes group 42.68 vs nondiabetes group 41.47; t_{3082}=−2.858; P=.004), and heart rate (mean: diabetes group 75.05 beats/min vs nondiabetes group 73.14 beats/min; t_{3082}=−7.948; P<.001). The extreme gradient boosting model (XGBM) demonstrated the best performance, with an accuracy of 84.88%, precision of 77.92%, recall of 66.91%, F1 score of 72.00, and AUC of 0.7957. The SHAP analysis of the top-performing XGBM revealed key predictors for diabetes: hypertension, age, percent body fat, heart rate, hyperlipidemia, basal metabolic rate, stress, and oxygen saturation. Hypertension strongly increased diabetes risk, while advanced age and elevated stress levels also showed significant associations. Hyperlipidemia and higher heart rates further heightened diabetes probability. These results highlight the importance and directional impact of specific features in predicting diabetes, providing valuable insights for risk stratification and targeted interventions.

Conclusions: This study focused on modifiable risk factors, providing crucial data for establishing a system for the automated collection of health information and lifelog data from older adults using digital devices at service facilities.

(JMIR Form Res 2025;9:e57874) doi:10.2196/57874

KEYWORDS

diabetes; prediction model; super-aging population; extreme gradient boosting model; geriatrics; older adults; aging; artificial intelligence; machine learning

Introduction

With advancements in medical technology contributing to longer life expectancies, the world is witnessing a rapid acceleration in population aging. The United Nations projects that the global older adult population will increase from 10% in 2022 to 16% by 2025 [1]. This demographic shift significantly drives the increased prevalence of noncommunicable diseases such as diabetes, hypertension, hyperlipidemia, renal failure, arthritis, and Alzheimer disease, which burden the primary health care system [2]. Specifically, the aging population presents challenges in geriatric care within primary health care systems, including shortages of caregiving personnel, financial constraints, and psychological stresses associated with family caregiving.

Among the Organisation for Economic Co-operation and Development (OECD) countries, South Korea is expected to become the first super-aged society—a society where the older

XSL•FO
RenderX

adult population accounts for more than 20% of the total population [3]. The rising older adult population and concurrent increase in chronic diseases represent a critical public health issue in Korea. Notably, diabetes is a serious issue, with 39.2% of older adults experiencing diabetes [4]. Diabetes management is crucial, as inadequate control can lead to severe complications, including hypertension and hyperlipidemia [5]. However, the diabetes management rate among the Korean older adult population stands at a mere 30.3%, significantly lower than the awareness rates (84%) and treatment rates (74.8%) [6]. Older adults are particularly vulnerable to diabetes due to aging and consequent physiological changes, as well as lifestyle modifications resulting from physical decline and other medical conditions [7].

The early prevention of diabetes is imperative in older adults. Previous studies have shown that early intervention in older adults can reduce the risk of complications such as cardiovascular diseases, renal failure, and vision impairment [8]. Yet, clinical interventions such as fast blood glucose tests are often needed for diabetes management, and these may not be feasible due to older adults' physical frailty. Thus, health education and behavioral interventions, not clinical treatments and diagnostics, are vital for preventing diabetes and other chronic conditions in this population [9]. Research indicates that the factors affecting diabetes in older adults differ from those in the general population [10]. While obesity is a risk factor in individuals in their 40s, being underweight elevates the risk for various chronic conditions, including obesity in older adults [11]. The differences in diabetes risk factors for older adults highlight the importance of specific research on the older adult population.

With the global advancement of computer technology, machine learning and deep learning have been used across various fields. Machine learning algorithms can analyze complex and large datasets and identify patterns and risk factors that might not be apparent through traditional statistical methods. In medicine, the potential of machine learning has been demonstrated through an increase in research on disease prediction, personalized medicine, and personalized public health services in clinical and public health care. Particularly, machine learning algorithms can incorporate a broader array of factors and use a wide range of data types to produce generalizable results compared to conventional statistical approaches [12].

However, research using machine learning algorithms to predict diabetes in the Korean older adult population remains sparse. Studies targeting older adults are particularly lacking, partly due to challenges in accessing physical and mental health data, as well as daily lifelog data, due to the mobility constraints of the older population. Thus, a system for the automated collection of health information and lifelog data from older adults must be established using digital devices at service facilities they frequent [10]. This study determines diabetes risk factors among older adults aged ≥60 years by using machine learning algorithms and selects an optimized model. We hypothesize that older adults may be affected by different and identical risk factors compared to younger generations and provide evidence to facilitate policy-making. The findings could serve as a model for other countries with similar demographic changes and health care challenges.

Methods

Data Collection

This cross-sectional study was conducted on older adults aged ≥60 years, and the survey was conducted from January to November 2023. Participants aged ≥60 years living in Seoul, South Korea, were recruited among individuals who had participated in the Mind Care Provider Project. Participants in the project voluntarily collected data through public institution promotions aimed at users interested in using health measurement services. To collect data from the participants, a mobile app called Gosufit was developed, which was installed on participants' digital devices [13].

First, the app measured indicators such as depression, stress, anxiety, basal metabolic rate (BMR), oxygen saturation, heart rate, and average daily step count, and the data were stored on a server. Second, a health coordinator (registered nurse) measured and recorded data on diabetes, hypertension, hyperlipidemia, chronic obstructive pulmonary disease (COPD), percent body fat, and percent muscle. In total, 3674 older adults aged ≥60 years participated in the survey from January 1 to November 30, 2023. After excluding 590 participants due to nonresponse, dropout, or missing data, 3084 participants were included in the final analysis. Nonresponse and dropout occurred when participants did not wish to continue or when data transmission was interrupted due to issues with the app during the survey.

Instruments

The target variable was the presence of diabetes. Those diagnosed with diabetes by a physician were coded as 1, and those who have not been diagnosed were coded as 0. The predictor variables included hypertension, posttraumatic stress disorder (PTSD), stress, anxiety, depression, BMR, oxygen saturation, average daily step count, hyperlipidemia, COPD, percent body fat, and percent muscle. Hypertension, hyperlipidemia, and COPD were coded as 1 for a physician diagnosis or 0 otherwise. PTSD, stress, and anxiety and depression were assessed using the 5-item PTSD checklist (PCL-5), the short version of the Geriatric Depression Scale (SGDS), and the Hospital Anxiety and Depression Scale (HADS), respectively, with 100-point–based scoring. Oxygen saturation was determined using a digital oximeter per 100% saturation. The average daily step count was measured via the app.

Statistical Analysis

The Boruta-based feature selection method (FSM) was used for feature selection, which is a wrapper-based FSM that uses the random forest classification algorithm [14]. The entire dataset was randomly divided into a 70% training set and a 30% testing set using a stratified sampling procedure [15].

Machine Learning Algorithm

The random forest algorithm is an ensemble learning method used for classification and regression. It constructed multiple

XSL•FO
RenderX

decision trees during training and output the class or mean prediction (for regression) of individual trees. Random forests corrected overfitting to the training set that is often seen in decision trees. The gradient boosting model (GBM) is a machine learning technique used for regression and classification problems. It generated a predictive model as an ensemble of weak prediction models (typically decision trees). Like other boosting methods, it built the model in a stage-wise fashion and optimized any differentiable loss function to generalize the model. The light gradient boosting model (LGBM) is a gradient-boosting framework that uses a tree-based learning algorithm. It was designed for distributed and efficient operation with a faster training speed, higher efficiency, lower memory use, and better accuracy. The LGBM was capable of processing large-scale data with numerous features and data points while maintaining performance.

The extreme gradient boosting model (XGBM) is an algorithm for tree-based ensemble learning that addresses the slow performance speed and overfitting regularization issues of gradient boosting models. It featured built-in cross-validation and the automatic handling of missing values. The k-nearest neighbors (KNN) model is a simple, versatile, and easy-to-implement supervised machine learning algorithm used for classification and regression. It classified data points based on how their neighbors were classified, stored all available cases, and classified new cases based on a majority vote of its k neighbors. The case assigned to the class was the most common class among the k-nearest neighbors.

Performance Evaluation Criteria

The performance of the machine learning models was evaluated based on accuracy, precision, recall, F1 score, and area under the curve (AUC). The confusion matrix evaluated classification models by dividing them into positive or negative categories based on the match between actual and predicted classes.

Accuracy represented the proportion of correctly classified data among the total predictions for the hypertension risk group. It measured how accurate the predictions for the hypertension risk group were. The mathematical equation for accuracy is as follows:

$$\text{Accuracy}\% = tp+tntp+tn+fp+fn \times 100$$

Precision represented the proportion of actual positive samples among the cases that were predicted as positive by the machine learning model. In other words, it indicated the ratio of samples positive for hypertension to those predicted to have hypertension by the model. The mathematical equation for precision is as follows:

$$\text{Precision}\% = tptp+fp \times 100$$

Recall measured the proportion of those predicted by the machine learning model to have hypertension within the actual hypertension group. It provided the percentage of cases predicted to be at risk for hypertension from the entire hypertension risk group. The mathematical equation is as follows:

$$\text{Recall}\% = tptp+fn \times 100$$

The F1 score was the harmonic mean of precision and recall. The mathematical equation is as follows:

$$\text{F1score}\% = 2tp2tp+fp+fn \times 100$$

The AUC referred to the area under the receiver operating characteristic curve, which was used to evaluate the performance of binary classification models, and an AUC close to 1 indicated better model performance. The AUC equation is as follows:

$$\text{AUC} = \int X=01 \text{TPRFPR-1} xdx$$

Model Interpretability

Shapley additive explanations (SHAPs) interpret the prediction outcomes of machine learning models. SHAPs were introduced by Lundberg and Lee [16] in 2017 and were designed based on a game theory concept known as Shapley values. These values supported prediction interpretation, enabling the assessment of the relative importance among features. Additionally, they helped understand the characteristics with the greatest influence on the model's predictions, assisting in model improvement or decision-making processes. The mathematical equation is as follows:

$$\varnothing k(v) = 1M! \sum S \subseteq M/\{k\} |S|! (M-|S|-1)! [v(S \cup \{k\}) - v(S)]$$

Ethical Considerations

Ethical approval for this study was obtained from Yonsei University Mirae Institutional Review Board (No. 1041849 - 202401-SB-021-01), including a supplementary application for expanded data collection. All procedures and data management were conducted following the General Data Protection Regulation and ethical principles outlined in the Helsinki Declaration. Informed consent was obtained from all study participants regarding data collection and the analysis of the data. The questionnaires were submitted entirely anonymously. No form of compensation was provided to the participants.

Results

Participant Information

In total, 3084 individuals participated in this study (Table 1). The study population comprised of 895 (29%) individuals with diabetes and 2189 (71%) individuals without. In addition, 1730 (56.1%) had hypertension, and diabetes prevalence significantly differed according to hypertension (χ^2_1=197.294; P<.001). In total, 1803 (58.5%) participants had hyperlipidemia, and diabetes prevalence significantly differed according to hyperlipidemia (χ^2_1=47.671; P<.001). The mean age was 72.66 years in the diabetes group and 71.81 years in the nondiabetes group. The mean stress score was 42.68 in the diabetes group and 41.47 in the nondiabetes group and significantly differed between the two groups (t_{3082}=−2.858; P=.004). The mean heart rate was 75.05 beats/min in the diabetes group and 73.14 beats/min in the nondiabetes group and significantly differed between the two groups (t_{3082}=−7.948; P<.001).

XSL•FO
RenderX

Table . Respondent characteristics.

Risk factor	Diabetes (n=895)	Nondiabetes (n=2189)	Chi-square or *t* test (*df*)	*P* value
Hypertension, n (%)			197.294[a] (1)	<.001
No	209 (23.4)	1115 (50.9)		
Yes	686 (76.6)	1074 (49.1)		
Hyperlipidemia, n (%)			47.671[a] (1)	<.001
No	286 (32)	995 (45.5)		
Yes	609 (68)	1194 (54.5)		
COPD[c], n (%)			7.764[b] (3082)	.005
No	856 (95.6)	2135 (97.5)		
Yes	39 (4.4)	54 (2.5)		
Sex, n (%)			3.472[b] (3082)	.06
Male	60 (6.7)	191 (8.7)		
Female	835 (93.3)	1998 (91.3)		
Age (years), mean (SD)	72.66 (6.31)	71.81 (6.32)	−3.395[b] (3082)	.001
PTSD[d], mean (SD)	14.45 (10.02)	14.21 (9.87)	−0.614[b] (3082)	.5
Stress, mean (SD)	42.68 (9.83)	41.47 (10.97)	−2.858[b] (3082)	.004
Anxiety, mean (SD)	15.58 (12.09)	15.67 (12.38)	0.179[b] (3082)	.86
Depression, mean (SD)	19.39 (12.48)	19.42 (12.92)	0.063[b] (3082)	.95
Percent body fat, mean (SD)	31.00 (8.74)	30.75 (8.66)	−0.727[b] (3082)	.47
Percent muscle, mean (SD)	35.73 (5.30)	35.95 (5.35)	−0.465[b] (3082)	.31
BMR[e], mean (SD)	1192.54 (159.05)	1187.71 (272.54)	−0.496[b] (3082)	.62
Oxygen saturation (%), mean (SD)	97.12 (3.48)	97.28 (2.44)	1.455[b] (3082)	.15
Heart rate (beats/min), mean (SD)	75.05 (6.29)	73.14 (5.97)	−7.948[b] (3082)	<.001
Daily step count, mean (SD)	11789.99 (18061.89)	13710.84 (34980.68)	1.560[b] (3082)	.12

[a]Chi-square.

[b]1-tailed *t* test.

[c]COPD: chronic obstructive pulmonary disease.

[d]PTSD: posttraumatic stress disorder.

[e]BMR: basal metabolic rate.

Risk Factor Selection Using Boruta

The importance of features was measured using the Boruta-based FSM (Multimedia Appendix 1). Hypertension, age, percent body fat, heart rate, hyperlipidemia, BMR, stress, and oxygen saturation were identified as important features. These features were included in the machine learning model for diabetes prediction.

Performance Comparison of Machine Learning Models

The performances of the 5 machine learning models used in the study were compared based on accuracy, precision, recall, F1 score, and AUC. Model performance was the highest for the XGBM, followed by the LGBM, random forest model, GBM, and KNN model (Table 2, Figure 1). The XGBM had an accuracy of 84.88%, precision of 77.92%, recall of 66.91%, F1 score of 72.00, and AUC of 0.7957, showing a high performance for its prediction.

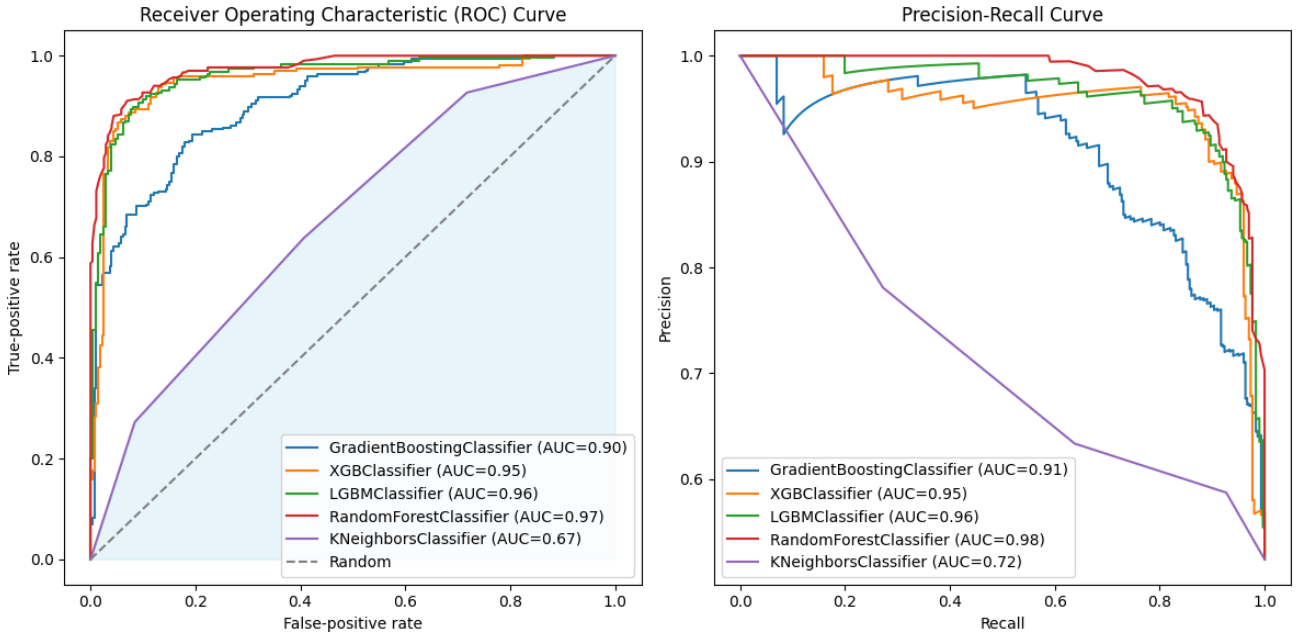
XSL•FO
RenderX

Table . Performance of ml methods.

Ranking	Models	Accuracy (%)	Precision (%)	Recall (%)	F1 score	AUC[a]
1	Extreme gradient boosting model	84.88	77.92	66.91	72.00	0.7957
2	Light gradient boosting model	84.77	78.57	65.42	71.39	0.7906
3	Random forest model	81.53	78.82	49.81	61.04	0.7216
4	Gradient boosting model	77.32	77.57	30.85	44.14	0.6360
5	K-nearest neighbors model	66.95	42.06	36.43	39.04	0.5794

[a]AUC: area under the curve.

Figure 1. Receiver operating characteristic curves and precision-recall curves of 4 predictive models. AUC: area under the curve; LBGM: light gradient boosting model; ROC: receiver operating characteristic; XBGM: extreme gradient boosting model.

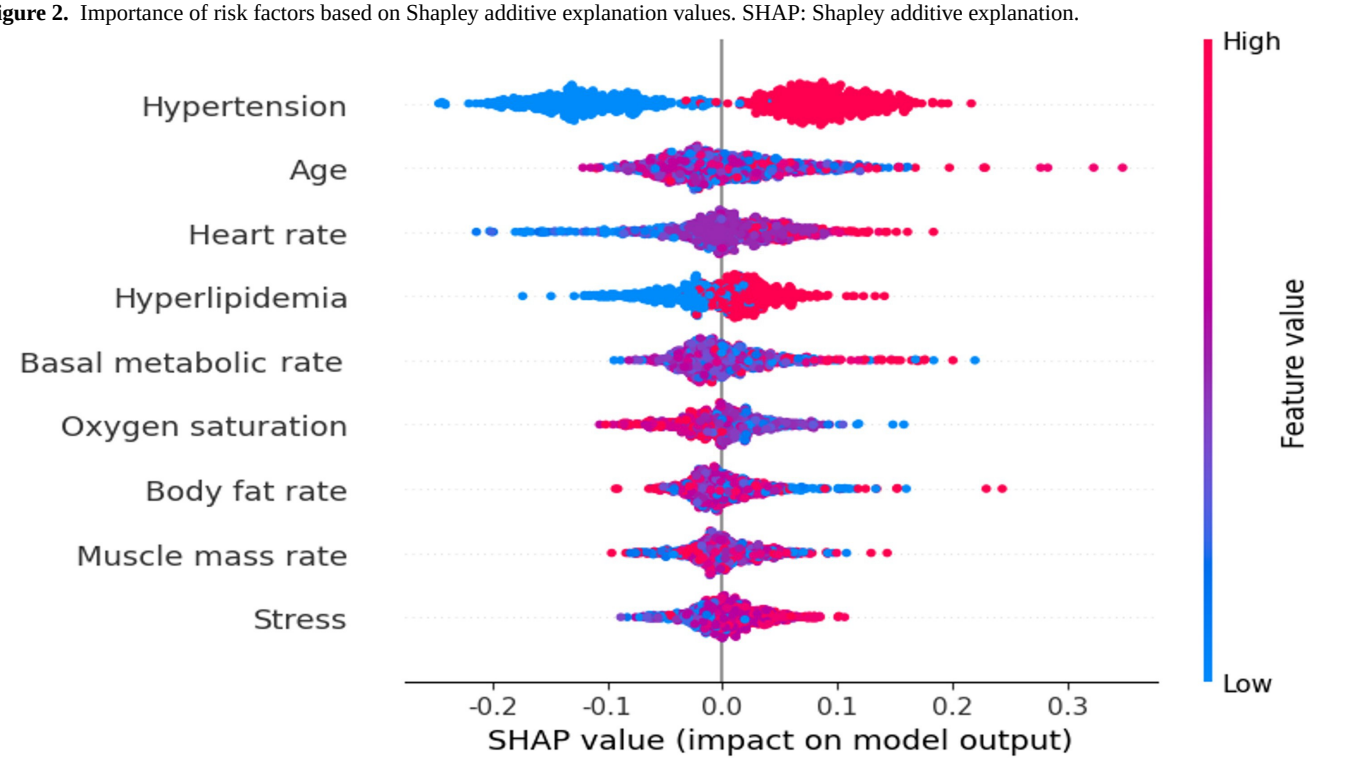


Interpretable Risk Factors of Hypertension

SHAP analysis was performed for the XGBM—the best-performing diabetes prediction model. The features selected through the FSM (hypertension, age, percent body fat, heart rate, hyperlipidemia, BMR, stress, and oxygen saturation) were included in the SHAP analysis. Red SHAP values indicated a higher impact on diabetes prediction, and blue SHAP values suggested a greater influence on nondiabetes outcomes (Figure 2). The findings indicated positive SHAP values for hypertension, indicating a stronger prediction of having diabetes, while not having hypertension was strongly linked to not having diabetes. Age presented mixed SHAP values, but notably, higher ages correlated with increased diabetes prediction. Similarly, higher heart rates and the presence of hyperlipidemia were associated with increased diabetes risk. Additionally, elevated stress levels were linked to a higher probability of diabetes. The SHAP analysis revealed a mixed impact of age on diabetes prediction. Particularly, as age increased, the likelihood of a diabetes prediction also increased. Additionally, a higher heart rate and the presence of hyperlipidemia were also linked to a higher diabetes probability. Moreover, an increase in stress levels elevated the risk of developing diabetes.

XSL•FO
RenderX

Figure 2. Importance of risk factors based on Shapley additive explanation values. SHAP: Shapley additive explanation.



Discussion

Principal Results

This study analyzed the performance of 5 machine learning–based algorithms in predicting diabetes risk in older Korean adults. The XGBM performed the best, supporting previous research that shows XGBM as the superior model for diabetes prediction with an AUC of 84% [17], which was also similar to this study. In contrast, other studies have assessed the predictive accuracy of decision trees, naive Bayes algorithms, and random forest algorithms for diabetes risk factors, finding the random forest algorithm to be the top performer with 94% accuracy and precision [18], which was identical to our predictive algorithm. Systematic reviews of machine learning approaches also mention that support vector machines, artificial neural networks, and decision trees are frequently used prediction classification models [19]. However, these were excluded due to the anticipated challenges in managing nonlinear patterns with linear approaches [20].

In this study, hypertension, hyperlipidemia, old age, heart rate, and stress were identified as diabetes risk factors, which was similar to previous reports. Existing studies establish a close link between hypertension and diabetes, with a stronger correlation observed in older adults [21], and this was also seen in our results. A study analyzing diabetes risk factors using large-scale health care data from Kuwait demonstrated that the logistic regression had the highest accuracy at 80.7% and that hypertension, obesity, and sex were also strongly associated with diabetes risk [22]. Further, an increase in heart rate was predicted to elevate diabetes risk. A study of 30,000 participants reported that diabetes risk increased with every 10 beats per minute rise in pulse rate [23], which was consistent with the findings of this study. Hyperlipidemia was also identified as a diabetes risk factor. Hyperlipidemia, hypertension, and diabetes are 3 major chronic conditions that are interrelated as mutual risk factors, and this study confirmed that hyperlipidemia is a diabetes risk factor [24].

This study predicted diabetes in older adults, and certain factors differed from previously known diabetes risk factors. In the existing literature, diabetes risk is elevated with increasing obesity [25], but in this study, the SHAP value for percent body fat was not clear, with no significant differences in diabetes risk according to percent body fat in the independent *t* test. Thus, the severity of obesity was not a significant risk factor for diabetes. On the contrary, diabetes risk increased with decreasing body weight. An array of study results has been reported regarding this issue. One such report suggested that weight loss positively impacts diabetes in older adults but also complicates the treatment for optimal blood glucose regulation in patients with type 2 diabetes [26]. Here, weight loss in older adults entails loss of muscles and bone density, which may have a detrimental impact on diabetes in the long term [27].

Certain predictor variables identified as significant diabetes risk factors in previous studies were not included in this study. Particularly, family history is a known risk factor for diabetes [28]. One important factor of this study was that it focused on modifiable risk factors, rather than family history, which is an unmodifiable risk factor. Particularly, we examined other risk factors besides unmodifiable risk factors, such as family history, age, and sex, as being able to provide interventions for these risk factors is essential. The Korean government should emphasize preventive health education in primary health care facilities, senior welfare centers, community centers, and public health centers to address modifiable diabetes risk factors in the older adult population, such as hypertension, heart rate, hyperlipidemia, and stress. Additionally, while education for older adults is crucial, prevention is more important. Thus,

XSL•FO
RenderX

health education should be provided for those in their 40s to prepare for healthy older adulthood.

Limitations

This study had a few limitations. First, the study population comprised older adults aged ≥60 years who reside in Seoul, Korea. Thus, it did not represent the entire 60-and-over population in the country. Further, while existing studies set the age criterion for older adults as 65 years and older, we set the age to 60 years and older because the retirement age in Korea is 60 years. Second, we could not obtain personally identifiable information such as personal income and education level due to legal regulations. Thus, even though these factors may predict diabetes, this study was limited to other characteristics. Lastly, this study randomly split the dataset into training and testing subsets (70% and 30%). This split was not stratified, which means the distribution of target variables in the subsets may not perfectly match the original dataset's distribution. This could potentially impact the generalizability of the model's performance to other datasets. Future studies could consider employing a stratified splitting method to ensure a balanced representation of target variables across subsets.

Conclusions

This study analyzed machine learning algorithms for diabetes prediction in older adults in Korea. Hypertension, hyperlipidemia, and stress were identified as modifiable diabetes risk factors. Additionally, body fat percentage did not significantly predict diabetes in older adults, presumably because body fat loss is closely linked to muscle strength and bone mass loss. The findings suggest that targeted interventions focusing on managing hypertension, hyperlipidemia, and stress can significantly reduce diabetes risk in this population.

Furthermore, these diabetes predictors in older adults could be mitigated by promoting healthier lifestyle choices and behaviors, such as regular physical activity, balanced nutrition, and stress management techniques. The government should implement comprehensive health education programs across various facilities, including primary health care facilities and welfare centers, to raise awareness about these modifiable risk factors.

Moreover, educational interventions should be initiated at a younger age—particularly for individuals in their 40s—to foster proactive health management and prevent the onset of diabetes in older adulthood. By adopting a preventive approach and addressing modifiable risk factors early, we can enhance the overall health and quality of life for the aging population in Korea. It is necessary to develop personalized modeling that predicts major chronic diseases such as diabetes, hypertension, hyperlipidemia, and obesity through the advancement of prediction algorithms. This will provide a foundation for creating personalized health promotion education and programs. Future research should focus on refining machine learning–based models by incorporating diverse datasets and longitudinal data to improve generalizability and predictive performance. Additionally, exploring the integration of behavioral and environmental factors into machine learning algorithms may further enhance the accuracy and applicability of these models in real-world settings.

Acknowledgments

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (NRF-2021R1C1C2005464).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Risk factors selection using the Boruta-based feature selection method.
[JPEG File, 127 KB - formative_v9i1e57874_app1.jpeg]

References

1. World population prospects 2022: summary of results. United Nations. 2022. URL: https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wpp2022_summary_of_results.pdf [accessed 2024-12-31]
2. Chen C, Ding S, Wang J. Digital health for aging populations. N Med 2023 Jul;29(7):1623-1630. [doi: 10.1038/s41591-023-02391-8] [Medline: 37464029]
3. Kim KW, Kim OS. Super aging in South Korea unstoppable but mitigatable: a sub-national scale population projection for best policy planning. Spat Demogr 2020 Jul;8(2):155-173. [doi: 10.1007/s40980-020-00061-8] [Medline: 34222615]
4. Kwon HS. Prevalence and treatment status of diabetes mellitus in Korea. J Korean Med Assoc 2023 Jul;66(7):404-407. [doi: 10.5124/jkma.2023.66.7.404]
5. de Boer IH, Bangalore S, Benetos A, et al. Diabetes and hypertension: a position statement by the American Diabetes Association. Diabetes Care 2017 Sep;40(9):1273-1284. [doi: 10.2337/dci17-0026] [Medline: 28830958]

XSL•FO
RenderX

6. Ha KH, Lee KA, Han KD, Moon MK, Kim DJ. Diabetes screening in South Korea: a new estimate of the number needed to screen to detect diabetes. Korean J Intern Med 2023 Jan;38(1):93-100. [doi: 10.3904/kjim.2022.283] [Medline: 36420563]
7. Flack KD, Davy KP, Hulver MW, Winett RA, Frisard MI, Davy BM. Aging, resistance training, and diabetes prevention. J Aging Res 2010 Dec 15;2011:127315. [doi: 10.4061/2011/127315] [Medline: 21197110]
8. Rooney MR, Rawlings AM, Pankow JS, et al. Risk of progression to diabetes among older adults with prediabetes. JAMA Intern Med 2021 Apr 1;181(4):511-519. [doi: 10.1001/jamainternmed.2020.8774] [Medline: 33555311]
9. Kramer MK, Vanderwood KK, Arena VC, et al. Evaluation of a diabetes prevention program lifestyle intervention in older adults: a randomized controlled study in three senior/community centers of varying socioeconomic status. Diabetes Educ 2018 Apr;44(2):118-129. [doi: 10.1177/0145721718759982] [Medline: 29514568]
10. Caprani N, Gurrin C, O'Connor N. I like to log: a questionnaire study towards accessible lifelogging for older users. Presented at: 12th International ACM SIGACCESS Conference on Computers and Accessibility; Oct 25-27, 2010; Orlando, FL p. 263-264. [doi: 10.1145/1878803.1878862]
11. Wilson PWF, Kannel WB. Obesity, diabetes, and risk of cardiovascular disease in the elderly. Am J Geriatr Cardiol 2002;11(2):119-123. [doi: 10.1111/j.1076-7460.2002.00998.x] [Medline: 11872970]
12. Henglin M, Stein G, Hushcha PV, Snoek J, Wiltschko AB, Cheng S. Machine learning approaches in cardiovascular imaging. Circ Cardiovasc Imaging 2017 Oct;10(10):e005614. [doi: 10.1161/CIRCIMAGING.117.005614] [Medline: 28956772]
13. GoSuFit. Google Play. URL: https://play.google.com/store/apps/details?id=com.healthbridge.hfact20&pcampaignid=web_share&pli=1 [accessed 2024-12-31]
14. Pudjihartono N, Fadason T, Kempa-Liehr AW, O'Sullivan JM. A review of feature selection methods for machine learning-based disease risk prediction. Front Bioinform 2022;2:927312. [doi: 10.3389/fbinf.2022.927312] [Medline: 36304293]
15. May RJ, Maier HR, Dandy GC. Data splitting for artificial neural networks using SOM-based stratified sampling. Neural Netw 2010 Mar;23(2):283-294. [doi: 10.1016/j.neunet.2009.11.009] [Medline: 19959327]
16. Lundberg SM, Su-In L. A unified approach to interpreting model predictions. Presented at: 31st International Conference on Neural Information Processing Systems; Dec 4-9, 2017; Long Beach, CA p. 4768-4777 URL: https://dl.acm.org/doi/10.5555/3295222.3295230 [accessed 2024-12-31]
17. Lai H, Huang H, Keshavjee K, Guergachi A, Gao X. Predictive models for diabetes mellitus using machine learning techniques. BMC Endocr Disord 2019 Oct 15;19(1):101. [doi: 10.1186/s12902-019-0436-6] [Medline: 31615566]
18. Yuvaraj N, SriPreethaa KR. Diabetes prediction in healthcare systems using machine learning algorithms on Hadoop cluster. Cluster Comput 2019 Jan;22(S1):1-9. [doi: 10.1007/s10586-017-1532-x]
19. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. Comput Struct Biotechnol J 2017;15:104-116. [doi: 10.1016/j.csbj.2016.12.005] [Medline: 28138367]
20. Tang Y. Deep learning using linear support vector machines. arXiv. Preprint posted online on Feb 21, 2015. [doi: arXiv:1306.0239v4]
21. Ye C, Fu T, Hao S, et al. Prediction of incident hypertension within the next year: prospective study using statewide electronic health records and machine learning. J Med Internet Res 2018 Jan 30;20(1):e22. [doi: 10.2196/jmir.9268] [Medline: 29382633]
22. Farran B, Channanath AM, Behbehani K, Thanaraj TA. Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity: machine-learning algorithms and validation using national health data from Kuwait--a cohort study. BMJ Open 2013 May 14;3(5):e002457. [doi: 10.1136/bmjopen-2012-002457] [Medline: 23676796]
23. Lee DH, de Rezende LFM, Hu FB, Jeon JY, Giovannucci EL. Resting heart rate and risk of type 2 diabetes: a prospective cohort study and meta-analysis. Diabetes Metab Res Rev 2019 Feb;35(2):e3095. [doi: 10.1002/dmrr.3095] [Medline: 30378246]
24. Sullivan PW, Ghushchyan VH, Ben-Joseph R. The impact of obesity on diabetes, hyperlipidemia and hypertension in the United States. Qual Life Res 2008 Oct;17(8):1063-1071. [doi: 10.1007/s11136-008-9385-7] [Medline: 18777200]
25. Du J, Yang S, Zeng Y, Ye C, Chang X, Wu S. Visualization obesity risk prediction system based on machine learning. Sci Rep 2024 Sep 28;14(1):22424. [doi: 10.1038/s41598-024-73826-6] [Medline: 39342032]
26. Kyrou I, Tsigos C. Obesity in the elderly diabetic patient: is weight loss beneficial? No. Diabetes Care 2009 Nov;32(Suppl 2):S403-S409. [doi: 10.2337/dc09-S348] [Medline: 19875589]
27. Newman AB, Lee JS, Visser M, et al. Weight change and the conservation of lean mass in old age: the Health, Aging and Body Composition Study. Am J Clin Nutr 2005 Oct;82(4):872-878. [doi: 10.1093/ajcn/82.4.872] [Medline: 16210719]
28. Harrison TA, Hindorff LA, Kim H, et al. Family history of diabetes as a potential public health tool. Am J Prev Med 2003 Feb;24(2):152-159. [doi: 10.1016/s0749-3797(02)00588-3] [Medline: 12568821]

Abbreviations

AUC: area under the curve
BMR: basal metabolic rate
COPD: chronic obstructive pulmonary disease

XSL•FO
RenderX

FSM: feature selection method
GBM: gradient boosting model
HADS: Hospital Anxiety and Depression Scale
KNN: k-nearest neighbors
LGBM: light gradient boosting model
OECD: Organisation for Economic Co-operation and Development
PCL-5: 5-item PTSD checklist
SGDS: short version of the Geriatric Depression Scale
SHAP: Shapley additive explanation
XGBM: extreme gradient boosting model

Edited by A Mavragani; submitted 28.02.24; peer-reviewed by BM Kunar, HB Drissi, J Nalepa; revised version received 09.12.24; accepted 09.12.24; published 21.01.25.

Please cite as:
Lee H, Park MB, Won YJ
AI Machine Learning–Based Diabetes Prediction in Older Adults in South Korea: Cross-Sectional Analysis
JMIR Form Res 2025;9:e57874
URL: https://formative.jmir.org/2025/1/e57874
doi:10.2196/57874

©Hocheol Lee, Myung-Bae Park, Young-Joo Won. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Insights From Diverse Perspectives on Social Media Messages to Inform Young Adults With Cancer About Clinical Trials: Focus Group Study

Melissa P Beauchemin[1], PhD, RN, CPNP, CPON; Desiree Walker[2], BA; Allison Rosen[3], MS; Maria Frazer[4], BSE; Meital Eisenberger[5], BS; Rhea K Khurana[1], BSc; Edward Bentlyewski[5], MSN, APN, NP-C, AOCNP, CCRP; Victoria Fedorko[6], BA; Corey H Basch[7], EdD, MPH; Grace C Hillyer[6], EdD, MPH

[1]Columbia University School of Nursing, New York, NY, United States

[2]Afiya, New York, NY, United States

[3]Advocates for Collaborative Education, Santa Cruz, CA, United States

[4]Pediatrics, Hematology/Oncology, University of Iowa Health Care, Iowa City, IA, United States

[5]Cancer Clinical Protocol and Data Management, Columbia University Herbert Irving Comprehensive Cancer Center, New York, NY, United States

[6]Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY, United States

[7]Department of Public Health, William Paterson University, Wayne, NJ, United States

Corresponding Author:
Melissa P Beauchemin, PhD, RN, CPNP, CPON
Columbia University School of Nursing, New York, NY, United States

Abstract

Background: Low rates of adolescent and young adult (YA; aged 15-39 y) clinical trial enrollment (CTE), particularly among underserved groups, have resulted in a lack of standardized cancer treatments and follow-up guidelines for this group that may limit improvement in cancer treatments and survival outcomes for YAs.

Objective: To understand and address unique barriers to CTE, we conducted focus groups to learn about informational, financial, and psychosocial needs of YAs surrounding CTE and identify strategies to address these barriers.

Methods: We conducted 5 focus groups in 2023 among a diverse sample of YA patients from across the United States. An interview guide was developed collaboratively with YA advocates. Specifically, informational needs, financial concerns, and psychosocial issues were explored, and participants were probed to suggest strategies, especially those that leverage technology, to address these barriers. Sessions were audio recorded, transcribed, and coded using direct content analysis. Findings were synthesized through consensus discussions.

Results: We confirmed the previously proposed thematic barriers regarding YA CTE and identified 9 subthemes: awareness, lack of clear and accessible CTE information, fear of the unknown, assumptions about costs, insurance coverage, navigating financial responsibilities, clinical trial discussions, clinical trial misconceptions, and desire for a support network. Throughout, YAs mentioned needs that might be addressed through informational outreach leveraging digital technology, the internet, and social media.

Conclusions: This study expands knowledge of YA perceived barriers to CTE. These findings suggest that leveraging digital technology to disseminate reliable information to address needs may be an effective strategy to improve clinical trial participation in the YA population.

(JMIR Form Res 2025;9:e64265) doi:10.2196/64265

KEYWORDS

young adults; adolescent and young adult cancer; cancer; cancer treatment; clinical trials; clinical trial awareness; clinical trial enrollment; clinical trial knowledge; clinical trial attitudes; clinical trial enrollment barriers; social media; social media messages; psychosocial; United States

Introduction

Among adolescents and young adults (AYAs; aged 15 - 39 y) with cancer [1], clinical trial enrollment (CTE) rates range from 3% to 14%, a number disproportionately below the observed annual cancer incidence in this age group [2]. Previous research shows the disparity in CTE among AYAs is stark when compared to the cancer incidence rates across other age groups. For instance, pediatric and older adult patients demonstrate higher CTE rates compared to AYAs, although AYAs

XSL•FO
RenderX

experience higher cancer incidence [3-5]. Despite cancer incidence being 3 times greater and with lower survival rates across all major tumor types for AYAs [3-8], there remains insufficient progress in discovering and developing new therapies, investigating molecular differences, and CTE for this population [9,10]. Inferior survival outcomes are further compounded among AYAs from marginalized groups (eg, non-White, publicly insured, low-income, rural residents, and non-English speakers [9,11]).

Young adults (YAs; aged 18-39 y) with cancer are a distinct group with needs that differ greatly from pediatric and adult populations. A cancer diagnosis as a YA can derail developmental transitions and impact emotional, cognitive, and social development [12,13]. As digital natives, a term used to describe individuals who have grown up with digital technology as an integral part of their daily lives, YAs approach health information-seeking differently than their younger and older counterparts [14]. Their proficiency with digital technology makes them especially adept at accessing medical information online using a wider variety of digital resources than older adults [15,16]. Interventions to increase CTE have demonstrated feasibility and acceptability among YAs with cancer [17], however, interventions developed thus far incorporate little input from YA patients.

Abrahão et al [18] explored CTE barriers from the perspectives of physicians and AYA patients using individual interviews and found that poor understanding of clinical trials, financial issues, and psychosocial problems were the greatest obstacles to CTE for AYAs. As the first phase of a larger study (TALK TO ME: Tailored YA Trial Messaging, HF CU22-2020), we conducted focus groups (FGs) to capitalize on interpersonal communication among AYAs with cancer and inform the development of tailored, relevant, and salient social media messages to address barriers to CTE among YAs with cancer.

Methods

Study Design

To build on prior work understanding barriers and facilitators to CTE among YAs, we used a qualitative descriptive approach to conduct FGs among YAs with cancer, as this methodology is ideal for providing a rich description of an experience by participants rather than interpreting or theorizing [19,20]. As the goal of this phase of this study was to develop social media messages for YAs regarding CTE, this methodologic approach was ideal to stay close to the participant experiences and perceptions. We used FGs as the mode of data collection to tap into the communication, interactions, and processes that are more readily observed among groups than in individual interviews [21].

Study Population and Recruitment

FG participants were recruited by leveraging the networks and social media platforms (eg, Facebook support groups and Instagram) of this study team's YA advocates via social media posts or by direct outreach. Outreach messages explained that we were recruiting individuals who were diagnosed with cancer between 18 and 39 years of age to participate in a FG to explore YA perceptions about clinical trials with the goal of informing the development of social media messages that aim to improve understanding and address concerns about clinical trials for YAs with cancer. Interested respondents were asked to provide basic sociodemographic (current age, gender identity, race or ethnicity, and city or state of residence, rural vs urban), whether they were active social media users, clinical information, including age at diagnosis and cancer type and availability (weekday vs weekend and daytime vs evening) for group participation, via a Google Forms survey link. As we sought to understand reasons for both participation and nonparticipation in cancer clinical trials, participation in a clinical trial was not an eligibility criterion.

This study team's YA advocates contacted potential participants to explain this study's procedures, confirm eligibility (diagnosed with cancer between 18 and 39 years of age, active social media user, and English speaker), and invite participation. Potential participants were then purposively sampled to represent a range of experiences (1 group all female and any race or ethnicity; 1 group all male and any race or ethnicity; 3 groups mixed gender identity, race or ethnicity, and any cancer type).

FG Data Collection

An FG guide was developed during weekly study team meetings guided by the Abrahaō framework from individual interviews among AYA participants [18]. We applied an iterative process to the guide development; pilot testing was carried out by this study's team members, including patient advocate partners [22]. To ensure study participants were oriented to the topic, a brief, 2‐ to 3-minute overview was prepared by this study team to introduce and define a clinical trial [23].

FGs were conducted virtually over Zoom between May and October 2023 and were 1 to 1½ hours in duration. Moderators included female YA advocates and, for the all-male group, a male oncology nurse practitioner. An information sheet was distributed via email to all participants before the FG.

The sessions began with a brief explanation of clinical trials and important definitions. Participants shared their experiences with cancer clinical trials including experiences of family members or friends and were asked about reasons for participating or not, and the support needed for trial involvement specifically as related to information, finances, and their psychosocial needs [18]. They also discussed their trust in social media information, preferred platforms, where they often seek health information, and suggested ways to use it to share clinical trial information with YAs.

Data Analysis

All FG sessions were audio-recorded and transcribed verbatim using NVivo Transcription, Lumivero, LLC. Participants were assigned a speaker ID, and transcripts were deidentified and cleaned using a denaturalized approach [24] to facilitate analysis of the participants' words rather than a more interpretive approach. We used a directed content analysis [25] to enrich and expand our understanding of barriers and facilitators to refine existing themes rather than identify new ones. Study team members trained (MPB, ME, RKK, and GCH) in qualitative analysis coded the transcripts first individually; then consensus

XSL•FO
RenderX

discussions were held to discuss coding discrepancies, synthesis of findings, and to ensure intercoder agreement. Saturation was anticipated given the deductive nature of the analytic approach but was tracked during consensus discussions [26].

Ethical Considerations

Verbal consent, including consent to be audio recorded, was obtained at the start of each FG. Participants were informed of their right to withdraw from this study at any time without penalty. For their time and effort, participants were compensated with a gift card valued at US $50. Regulatory approval was obtained from Columbia University Institutional Review Board (ID# AAAU4337).

Results

Overview

A total of 33 individuals participated in the FGs (maximum 8, minimum 4). The mean age was 32.2 (SD 5.7) years. Males and females were equally represented; 42.4% (n=14) were White and 42.4% (n=14) were Black or African American. Most (n=24, 72.7%) resided or sought treatment in urban areas, and 27.3% (n=9) were diagnosed with breast cancer (Table 1). Findings are grouped by themes (informational, financial, and psychosocial) and subtheme. Exemplary quotes further supporting subthemes are displayed in Table 2.

Table . Characteristics of focus group participants (N=33).

Demographic	Value
Age (years)	
Mean (SD)	32.2 (5.7)
Range	24-42
Gender, n (%)	
Female	17 (51.5)
Male	16 (48.5)
Race, n (%)	
White	14 (42.4)
Black or African American	14 (42.4)
Asian	1 (3)
Mixed race	3 (9.1)
Did not specify	1 (3)
Rural versus urban, n (%)	
Rural	9 (27.3)
Urban	24 (72.7)
Cancer type, n (%)	
Skin cancer	2 (6.1)
Breast cancer	9 (27.3)
Lymphoma	2 (6.1)
Other	20 (60.6)

XSL•FO
RenderX

Table . Interview themes and categories with additional illustrative quotes.

Theme and categories		Representative quote
Theme 1: Information or knowledge		
	Awareness	"I was given two options as far as radiation and chemo. And basically, they were like, these are the two options you have, which one do you want? Clinical trials never even brought up." [FG1[a]]
	Lack of standardized or central system	"One of the biggest issues, I think upon diagnosis as a patient and as a as a caregiver is information overload." [FG1]
	Fear of the unknown	"Being reminded that...my health is more important than just a number or an experiment or a statistic. And then in terms of. Like in a form of social media that I would go to, I think like visual media tends to be more compelling to me like to be able to see your face and and a name and to know you know who that person is giving me information." [FG4]
Theme 2: Psychosocial		
	Clinical trial discussions	"You know, being upfront about the risks obviously is important, but the potential benefits too, I think just prevent presenting balanced information about the nature of the trial is probably all you can do, really." [FG5]
	Clinical trial misconceptions	"I feel like clinical trials have been sort of this last-ditch effort after thought" [FG1]
	Desire for support network	"Social media is the connection to community" [FG2]
Theme 3: Financial		
	Assumptions or knowledge about clinical trials	"And another one said, In order to find out if you're eligible, you need to pay $750 out of pocket to do this consultation. So that just makes people I mean, it just it's just not fair." [FG1]
	Young adult specific concerns	"The trial sponsor may be able to offer some form of compensation or some sort of resource, but there's also the sort of unintended financial impact that can happen, especially to this age group taking time off from work, finding childcare, meeting the needs for other bills, that kind of stuff." [FG1]
	Communication	"And I think that the visits they go by so quickly anyways, there's so much to cover. There probably wouldn't even be enough time for a provider to sit down and have a full conversation about what is a clinical trial." [FG5]

[a]FG: focus group.

Theme 1: Informational

Subtheme 1: Awareness

Across all FGs, YAs cited varying levels of exposure to information and lack of awareness of cancer generally, their specific cancer, and treatment options that may have included clinical trials. Many reported facing the challenges of gathering important information about their diagnosis and treatment options on their own and conveying their informational needs to others. One participant reflected saying, "when I was diagnosed, I was still on my parent's health insurance and my dad handled all of the insurance...I didn't know anything like, I feel like if I hadn't had that support, like I can't even imagine."

Participants also noted that they wanted to understand their disease and treatment options to better participate in their own care, however finding relatable and accessible information was difficult. One participant remarked "I thought I was really educated until I got diagnosed with cancer and then I just felt like an idiot." Some participants felt that they were tasked with processing information about their cancer diagnosis and

XSL•FO
RenderX

treatment that was not easily digestible or so complex that "no normal person is just going to know."

The process of deciding on treatment, whether the standard of care or a potential clinical trial, was described as "draining." In some cases, participants said that they were completely unaware that a clinical trial may have been an option. As one person recalled, "I was given two options as far as radiation and chemo. And basically, they were like, these are the two options you have, which one do you want? Clinical trials [were] never even brought up."

More specifically, YAs wanted to be better informed about their possible treatment options and how they might impact them in the future. They expressed the need to comprehend the risks and benefits associated with each treatment option to make an informed decision. One participant summarized this as "there's so many different things that you have to consider when making these decisions and side effects are no joke for cancer treatment…there's just it's stuff that's very lasting and those are other things you have to consider when making the decision." Most importantly, participants conveyed that they wanted their providers to consider them as a whole person when treating their cancer, specifically considering their reproductive health concerns and fertility preservation. One participant told us "I had like seven days to decide if I wanted to freeze my eggs or not." They believed that the provider should make these YA-unique issues a top priority when creating cancer treatment plans and ensure that YAs fully understand these implications at the start.

Subtheme 2: Lack of Standardized or Central System

Participants repeatedly spoke of the difficulty in finding centralized databases with information on their cancer or resources relevant to their experiences. When trying to understand the concept of a clinical trial, participants found the information to be complex and filled with medical terminology they did not understand only added to the stress of their illness. Many felt intimidated, overwhelmed, and frightened by the information they found, especially for YAs with rare cancers. Among participants who searched these databases, the consensus was there needs to be "some sort of database that like pulls that information [together]...clinical trial wise…it would be nice to know that there are options for people that have like, not so ideal cancer circumstances." Additionally, participants felt this information should be engaging, compassionate, and written in plain language.

When researching clinical trials, YAs said that they would like information to come from trusted sources such as their medical team or the organization implementing the clinical trial. It helps, they said when "someone is maybe [a] medical personnel…or a governmental organization[s] that you can then be able to trust that information." One participant who participated in a clinical trial explained that "the biggest differentiator for me was my oncologist explaining to me all of the options for clinical…and really went into detail…he explained to me in a way that I could understand...and I trusted him. I trusted him implicitly."

Participants addressed the role of caregivers seeking treatment information for their YAs. They emphasized the need for resources to comfort both YAs and their caregivers throughout treatment.

> *The caregiver is often the person doing this research or supporting this. And so, I think it's really important to consider them as your equal audience.*

They also warned that "One of the biggest issues, I think upon diagnosis as a patient and as a caregiver is information overload."

Subtheme 3: Fear of the Unknown

A common fear cited by participants regarding clinical trials was adverse side effects of the experimental treatment. They also wanted to connect with others like themselves to hear firsthand about their experience with cancer and the treatment options they were considering to put their minds at ease. They stated that when they could not access the support and information they needed through their care team, they sought it on social media using Instagram hashtags and Facebook groups.

> *I think the biggest thing for me, where I found the most helpful information was searching hashtags on Instagram. I searched my chemo regimen hashtags. I searched my diagnosis. I searched by stage. I searched AYA cancer...And I would find people there that would have a community on their pages and resources on their pages.*

YAs that used social media to supplement their understanding and find support expressed concern over the lack of reliable sources and the spread of misinformation on some platforms. Participants suggested that it is difficult to differentiate between helpful information and potentially harmful content on social media. Another issue related to the use of social media reported by participants was "doom scrolling," which can cause increased conflict within the patient. One participant explained that "because of the information overload, I've really had to limit my information intake from social media."

Theme 2: Psychosocial Issues

Subtheme 1: Clinical Trial Discussions

YAs cited that care teams should remember that patients usually have limited knowledge of clinical trials, including the risks and benefits. As one participant explained, "I actually refused [experimental] treatment because I thought my only option was chemotherapy and radiation...if I knew that I had other options, I probably would have been more proactive, no matter what the cost was like." Participants stated that initiating simple dialogue that normalizes CTE in the context of cancer treatment would motivate patients to consider clinical trials as a treatment option.

Subtheme 2: Clinical Trial Misconceptions

Participants explained that common misconceptions about clinical trials were that they were reserved for patients with advanced disease and considered a "last ditch effort" for only those with terminal disease. These misconceptions add to the fear surrounding clinical trials and can present as a key barrier to CTE and psychological distress. Participants explained that

XSL•FO
RenderX

providers should anticipate and address misconceptions about clinical trials during treatment discussions.

Subtheme 3: Desire for a Support Network

Participants reported that the effects of their illness and treatment took a significant toll on their mental health and that additional check-ins with the medical team would have been beneficial. For those who participated in a clinical trial, the additional contact and attention by the clinical trial care team eased the burden. One participant who participated in a clinical trial explained that "the additional support for me was one of the best things that came out of the clinical trial because I was at such a large hospital…This connection is especially important during trial enrollment as this is an uncertain time for patient[s]."

Participants explained that they would have appreciated a peer support network while navigating life and treatment simultaneously. As one participant explained, "cancer is the worst part-time job on the entire planet." They use social media as a "connection to community." Participants expressed interest in a social media account or online group that connects patients more easily since it can be valuable for communication with those who have had similar experiences.

Some participants mentioned that discomfort while considering a clinical trial can be further exacerbated by family members' lack of understanding or distrust of medical research. This may lead family members to caution patients away from participating. One family responded to the idea of a clinical trial their son was considering by saying "No, don't do it. You know, they're going to do x y z. It's not safe. You know, you're not that sick." The participant explained that their family saw a clinical trial as "a last-line option and not a first-line option." The importance of providing caregivers and others surrounding the patient with accurate information to support the YAs treatment decisions was stressed.

Theme 3: Financial Concerns

Subtheme 1: Assumptions About Clinical Trials Costs

Financial hardship was mentioned by many participants as a key deterrent to CTE. Several reported that the fear of financial problems prevented them from inquiring about clinical trials because they assumed resources would not be provided or that the resources provided would not be sufficient to cover additional costs. Patients applauded treatment centers that provide financial navigation and resources to patients considering clinical trials. They believed this could ease barriers such as travel expenses and limited insurance coverage and potentially increase CTE.

Subtheme 2: Insurance Coverage

Some participants indicated that the insurance framework is difficult to navigate on its own and can be further complicated by seeking nonstandard of care treatment. One participant explained that they "wanted to get enrolled in a clinical trial that was at another health institution" but they did not proceed because "my insurance wouldn't cover it because it was out of network." Alongside this, the struggle to balance their schooling, family, and work responsibilities impacts the state of their finances.

Subtheme 3: Navigating Financial Responsibilities

Some of the participants assumed that the patient was responsible for additional costs which deterred them from inquiring about clinical trials as a treatment option. Many participants were unaware that most health insurers cover many, if not all, costs associated with clinical trials, that costs related to increasingly more clinical trials are covered by the sponsor, or that there are occasions when sponsors offer financial assistance to address costs not covered by insurance or the clinical trial. One participant remarked, "I think a lot of us don't even realize that our hospital might have financial navigators or patient navigators who exist to help answer those questions." YAs believed providers and any online sources should educate patients about their financial options while addressing common misperceptions.

Discussion

In this study, we explored categorical individual-level barriers [18] to CTE among a diverse group of individuals diagnosed with cancer as YAs to extend our knowledge and inform future strategies to address these barriers. Through FGs representative of gender identity, race, and geographic region in the United States, we found that lack of knowledge and imperfect communication were the greatest barriers to CTE in this population. Participants indicated that limited awareness about clinical trials, commonly held misperceptions, misinformation on websites and social media, and a lack of clear and accessible information about clinical trial risks and benefits likely contribute to underrepresentation among YAs. Proactively providing anticipatory guidance may circumvent distress if the topic is raised and reduce patient hesitancy in participation.

Consistent with other literature, our study supports that YAs often seek information on the web or social media [27,28] and participants favored an intervention using social media to address CTE barriers. They also provided clear guidance on how information might be tailored to YAs and what specific information may be most helpful in addressing concerns and barriers to CTE.

In many ways, the financial concerns and psychosocial issues reflected a subset of the informational gaps reported by participants. Finances were a significant concern, and several participants were unaware of insurance coverage requirements or that resources such as financial navigators exist to help with the costs of CTE [29]. Having financial information readily available during the cancer treatment decision-making process can support discussions around a clinical trial option. A digital support network may also provide a connection with others sharing a similar lived experience during times of loneliness and fear.

We followed rigorous qualitative methodology to conduct our study, however, there are limitations to acknowledge. Potential selection bias may be present, as participants were recruited through our YA patient advocates on the internet and, thus, may have been more technologically proficient and familiar with social media and clinical trials than the general YA patients with cancer community. However, the vast majority of YAs

XSL•FO
RenderX

have an online presence and use social media and the internet regularly. Further, as with other qualitative research, the transferability (ie, generalizability) of these data may be limited although great effort was expended to ensure that our sample was diverse in terms of sociodemographic characteristics to yield FGs representative of the YA population. More than half of our FG participants were non-White which provides a deeper view of experiences from underrepresented YAs regarding engagement with and perceptions of CTE. Although we based our discussions a priori on themes identified by Abrahão et al [18], moderators followed the discussions and probed further when any topic arose organically to ensure all potential barriers were explored and thus, resulted in the expansion of subthemes described.

Our research enhances our understanding of the obstacles that YAs face when making decisions about cancer treatment, including the option to participate in a clinical trial. It also provides further evidence that using a combination of approaches at different levels may be most effective in engaging YAs in discussions about CTE. While barriers to CTE at the provider, institutional, and system levels have been extensively studied, the YAs in our study highlight individual-level challenges that need attention. Our findings indicate that a social media intervention focused on sharing reliable information about CTE could help YAs process complex information and make informed treatment decisions, thus improving CTE in this population. Engaging with YAs through social media allows researchers and medical professionals to connect with them in a way that is familiar and comfortable. Further research should address the gaps in information identified in our study to develop resources and educational materials that can enhance understanding of CTE among YAs, their families, and caregivers. When developing treatment plans, it is crucial to prioritize patients' informational needs and provide all YAs with the tools and resources to have meaningful discussions with their health care providers about CTE.

Acknowledgments

This study was funded by a grant to GCH from the Hope Foundation (HF CU22-2020).

Conflicts of Interest

None declared.

References

1. Unger JM, Beauchemin M, Hershman DL. Adolescent and young adult enrollment to a National Cancer Institute-sponsored National Clinical Trials Network Research Group over 25 years. Cancer 2021 Dec 15;127(24):4574-4584. [doi: 10.1002/cncr.33855] [Medline: 34351619]
2. Jacob SA, Shaw PH. No improvement in clinical trial enrollment for adolescents and young adults with cancer at a children's hospital. Pediatr Blood Cancer 2017 Dec;64(12). [doi: 10.1002/pbc.26638] [Medline: 28509440]
3. Keegan THM, Ries LAG, Barr RD, et al. Comparison of cancer survival trends in the United States of adolescents and young adults with those in children and older adults. Cancer 2016 Apr 1;122(7):1009-1016. [doi: 10.1002/cncr.29869] [Medline: 26848927]
4. Bleyer A, Budd T, Montello M. Adolescents and young adults with cancer: the scope of the problem and criticality of clinical trials. Cancer 2006 Oct 1;107(7 Suppl):1645-1655. [doi: 10.1002/cncr.22102] [Medline: 16906507]
5. Bleyer A, Viny A, Barr R. Cancer in 15- to 29-year-olds by primary site. Oncologist 2006 Jun;11(6):590-601. [doi: 10.1634/theoncologist.11-6-590] [Medline: 16794238]
6. Bleyer A OLM, Barr R, Ries LAG. Cancer Epidemiology in Older Adolescents and Young Adults 15 to 29 Years of Age, Including SEER Incidence and Survival: 1975-2000: National Cancer Institute (NCI); 2006.
7. Berkman AM, Livingston JA, Merriman K, et al. Long-term survival among 5-year survivors of adolescent and young adult cancer. Cancer 2020 Aug 15;126(16):3708-3718. [doi: 10.1002/cncr.33003] [Medline: 32484922]
8. Kahn JM, Keegan THM, Tao L, Abrahão R, Bleyer A, Viny AD. Racial disparities in the survival of American children, adolescents, and young adults with acute lymphoblastic leukemia, acute myelogenous leukemia, and Hodgkin lymphoma. Cancer 2016 Sep 1;122(17):2723-2730. [doi: 10.1002/cncr.30089] [Medline: 27286322]
9. Collins CL, Malvar J, Hamilton AS, Deapen DM, Freyer DR. Case-linked analysis of clinical trial enrollment among adolescents and young adults at a National Cancer Institute-designated comprehensive cancer center. Cancer 2015 Dec 15;121(24):4398-4406. [doi: 10.1002/cncr.29669] [Medline: 26393950]
10. de Rojas T, Neven A, Terada M, et al. Access to clinical trials for adolescents and young adults with cancer: a meta-research analysis. JNCI Cancer Spectr 2019 Dec;3(4):kz057. [doi: 10.1093/jncics/pkz057] [Medline: 32337483]
11. Roth M, Beauchemin M, Kahn JM, Bleyer A. Patterns of National Cancer Institute-sponsored clinical trial enrollment in Black adolescents and young adults. Cancer Med 2021 Nov;10(21):7620-7628. [doi: 10.1002/cam4.4292] [Medline: 34592782]
12. Rueda A, MacIntyre J. Challenges faced by adolescent and young adult patients with cancer. Oncology Nursing News. 2022. URL: https://www.oncnursingnews.com/view/challenges-faced-by-adolescent-and-young-adult-patients-with-cancer [accessed 2024-05-27]

XSL•FO
RenderX

13. Silin C. Help AYA survivors and patients with cancer navigate infertility. ONS Voice. 2020. URL: https://voice.ons.org/stories/help-aya-survivors-and-patients-with-cancer-navigate-infertility [accessed 2024-05-27]
14. Catley J. Here's why teens and young adults are healthcare digital natives. Health Works Collective. 2018. URL: https://www.healthworkscollective.com/heres-why-teens-and-young-adults-are-healthcare-digital-natives/ [accessed 2025-01-10]
15. Digital natives. American Library Association. 2015. URL: https://www.ala.org/future/trends/digitalnatives [accessed 2024-10-31]
16. Corzo C. Tech savvy plays a role on two generations' quest for helath information, FIU research finds. FIU College of Business. 2024. URL: https://business.fiu.edu/news/2024/tech-savvy-plays-a-role-on-two-generations-quest-for-health-information.html [accessed 2024-10-31]
17. Avutu V, Monga V, Mittal N, et al. Use of communication technology to improve clinical trial participation in adolescents and young adults with cancer: consensus statement from the children's oncology group adolescent and young adult responsible investigator network. JCO Oncol Pract 2022 Mar;18(3):224-231. [doi: 10.1200/OP.21.00554] [Medline: 34905405]
18. Abrahão R, Alvarez EM, Waters AR, et al. A qualitative study of barriers and facilitators to adolescents and young adults' participation in cancer clinical trials: oncologist and patient perspectives. Pediatr Blood Cancer 2022 Apr;69(4):e29479. [doi: 10.1002/pbc.29479] [Medline: 34913583]
19. Sandelowski M. What's in a name? Qualitative description revisited. Res Nurs Health 2010 Feb;33(1):77-84. [doi: 10.1002/nur.20362] [Medline: 20014004]
20. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
21. Kitzinger J. The methodology of focus groups: the importance of interaction between research participants. Soc Health & Illn 1994 Jan;16(1):103-121 [FREE Full text] [doi: 10.1111/1467-9566.ep11347023]
22. Kallio H, Pietilä AM, Johnson M, Kangasniemi M. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. J Adv Nurs 2016 Dec;72(12):2954-2965. [doi: 10.1111/jan.13031] [Medline: 27221824]
23. What are clinical trials? National Cancer Institute. 2024. URL: https://www.cancer.gov/research/participate/clinical-trials/what-are-clinical-trials [accessed 2024-10-01]
24. Sandelowski M. Notes on transcription. Res Nurs Health 1994 Aug;17(4):311-314. [doi: 10.1002/nur.4770170410] [Medline: 8036279]
25. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
26. Morse JM. "Data were saturated...". Qual Health Res 2015 May;25(5):587-588. [doi: 10.1177/1049732315576699] [Medline: 25829508]
27. Psihogios AM, Lane-Fall MB, Graham AK. Adolescents are still waiting on a digital health revolution: accelerating research-to-practice translation through design for implementation. JAMA Pediatr 2022 Jun 1;176(6):545-546. [doi: 10.1001/jamapediatrics.2022.0500] [Medline: 35404381]
28. Rideout V, Fox S, Trust WB. Digital health practices, social media use, and mental well-being among teens and young adults in the US. Providence Digital Commons. 2018. URL: https://digitalcommons.providence.org/publications/1093/ [accessed 2024-06-16]
29. Doherty MJ, Thom B, Gany F. Evidence of the feasibility and preliminary efficacy of oncology financial navigation: a scoping review. Cancer Epidemiol Biomarkers Prev 2021 Oct;30(10):1778-1784. [doi: 10.1158/1055-9965.EPI-20-1853] [Medline: 34341051]

Abbreviations

AYA: adolescent and young adult
CTE: clinical trial enrollment
FG: focus group
YA: young adult

Edited by A Mavragani; submitted 12.07.24; peer-reviewed by JG Barnard, L Super, S Chantziara; revised version received 21.11.24; accepted 24.11.24; published 20.01.25.

Please cite as:
Beauchemin MP, Walker D, Rosen A, Frazer M, Eisenberger M, Khurana RK, Bentlyewski E, Fedorko V, Basch CH, Hillyer GC
Insights From Diverse Perspectives on Social Media Messages to Inform Young Adults With Cancer About Clinical Trials: Focus Group Study
JMIR Form Res 2025;9:e64265
URL: https://formative.jmir.org/2025/1/e64265
doi:10.2196/64265

XSL•FO
RenderX

© Melissa P Beauchemin, Desiree Walker, Allison Rosen, Maria Frazer, Meital Eisenberger, Rhea K Khurana, Edward Bentlyewski, Victoria Fedorko, Corey H Basch, Grace C Hillyer. Originally published in JMIR Formative Research (https://formative.jmir.org), 20.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Sentiments of Individuals with Interstitial Cystitis/Bladder Pain Syndrome Toward Pentosan Polysulfate Sodium: Infodemiology Study

Yulin Hswen[1], MPH, ScD; Qiuyuan Qin[2], MS, PhD; Pressley Smith[3], MS; Alison Swierczynski[3], BS; Stuart Bauer[4,5], MD; Erika Ladson[6], MS, PhD; Amanda Leigh Garrett[6], PhD; Catherine A Brownstein[3,7], MPH, PhD

[1]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, United States

[2]Department of Public Health and Sciences, University of Rochester, Rochester, NY, United States

[3]Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, United States

[4]Department of Urology, Boston Children's Hospital, Boston, MA, United States

[5]Department of Surgery, Harvard Medical School, Boston, MA, United States

[6]Inspire, Arlington, VA, United States

[7]Department of Pediatrics, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Catherine A Brownstein, MPH, PhD
Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, United States

Abstract

Background: Interstitial cystitis/bladder pain syndrome (IC/BPS) is a multifactorial, chronic syndrome involving urinary frequency, urgency, and bladder discomfort. These IC/BPS symptoms can significantly impact individuals' quality of life, affecting their mental, physical, sexual, and financial well-being. Individuals sometimes rely on peer-to-peer support to understand the disease and find methods of alleviating symptoms. The only US Food and Drug Administration–approved medication to treat IC/BPS is pentosan polysulfate sodium (PPS). However, ocular pigmentary maculopathy has been described in some individuals, with greater severity associated with prolonged PPS exposure.

Objective: While prior research has separately assessed the benefits and side effects of PPS, this study sought to identify (1) sentiments of individuals with IC/BPS toward PPS and (2) topics discussed by individuals with IC/BPS in conjunction with PPS through use of an internet peer-to-peer forum.

Methods: Data were collected from Inspire—an anonymous web-based health community where individuals gather by condition to find support and information. Sentiment analysis and percentages of negative, positive, and neutral sentiment for PPS discussions encompassing each topic was conducted using VADER (Valence Aware Dictionary for Sentiment Reasoning). Topic modeling was conducted using latent Dirichlet allocation. Words with the highest probability were ranked to categorize each topic, and authors manually investigated and labeled discussions.

Results: There were 354 forum posts related to PPS. Topic modeling with latent Dirichlet allocation revealed 5 topic categories: "ineffectiveness or discontinued use," "alternative treatments," "personal treatment suggestions based on experience," "severe side effects," and "risk of long-term use." Topics related to "severe side effects" and "risk of long-term use" garnered less discussion, with the former also having the lowest positive sentiment (4.28, 14.29%). The topic "ineffectiveness or discontinued use" was most frequently discussed. This topic also had the highest percentage of negative posts (52/152, 34.21%). However, the average compound score was within the neutral compound score range (−0.094, SD 0.625). In addition, forum data highlighted individuals' acknowledgment of the efficacy of PPS in improving their quality of life, with statements such as "saved my sanity" being representative. The overall compound individuals' sentiment toward PPS was −0.083, split across 32.49% (115/354) negative, 22.03% (78/354) positive, and 45.48% (161/354) neutral sentiment categories.

Conclusions: The overall authentic sentiment toward PPS is broad but balances to neutral. This neutral sentiment suggests that while some individuals express concerns about the side effects and long-term risks associated with PPS, others appreciate its positive impact on their quality of life. This research confirms that individuals with IC/BPS actively engage with health forums like Inspire to seek information, share their experiences, and explore different treatment options. As IC/BPS remains a complex syndrome, this study highlights the value of patient-led discussions in informing treatment decisions. Furthermore, these findings suggest that health care providers might benefit from considering the insights shared on peer-to-peer forums to better understand individual preferences, concerns, and expectations.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e54209) doi:10.2196/54209

KEYWORDS

interstitial cystitis; IC; painful bladder syndrome; bladder pain syndrome; BPS; social media; social network; pain; treatment; chronic condition; chronic disease; chronic illness; Elmiron; pentosan polysulfate sodium; PPS; internet forum

Introduction

Interstitial Cystitis/Bladder Pain Syndrome and Current Treatment

Interstitial cystitis/bladder pain syndrome (IC/BPS) is a multifactorial chronic syndrome that affects the urinary bladder [1]. The condition is characterized by urinary frequency, urgency, and bladder pain [1], involving both women and men [2], that impact their emotional, psychological, and social well-being [1]. The etiology of IC/BPS remains elusive; the disease is diagnosed not with a single test but a combination of multiple investigations to rule out other diagnoses [3]. Many individuals with IC/BPS present heterogeneous syndromes from several causes [4] suggesting it should be classified into different subtypes [5]. Recent investigations of genetics and biomarkers have begun to uncover the molecular mechanisms of IC/BPS [4,6]. First-degree relatives of women with IC/BPS have a greater prevalence of the disease than women without affected relatives, indicating a genetic component [7,8]. Twin studies have shown a greater concordance of IC/BPS among monozygotic than dizygotic twins [7,9]. The Golgi-localized adenosine triphosphatase gene *ATP2C1* has also been identified as having increased rare variant burden in an IC/BPS cohort (OR [odds ratio]=6.76, 95% CI 1.1-75.9) [6] and is a candidate gene.

Pentosan Polysulfate Sodium–Associated Pigmentary Maculopathy

Currently, the only US Food and Drug Administration (FDA)–approved treatment is oral pentosan polysulfate sodium (PPS). Use of PPS has been demonstrated to improve symptoms [10], although recent studies link PPS to a unique presentation of pigmentary maculopathy [11]. Pigmentary maculopathy can produce moderate visual impairment, macular hyperpigmented spots, yellow-orange deposits, and patchy retinal pigment epithelium atrophy [12]. The first study detailed 6 individuals over 2 years who developed new visual symptoms after exposure to PPS [11]. Several additional studies have replicated the association between PPS use and pigmentary maculopathy [13-16].

Support Groups and the Inspire Platform

Patients and caregivers often join web-based health communities to find support for incurable or rare diseases. Through interpersonal exchanges, patients and caregivers can develop an understanding of their illnesses. Social Comparison Theory postulates that social behaviors could be predicted on the assumption that individuals seek the maintenance of normalcy and accuracy in their world by using opinions of others to evaluate their thinking and feeling [17]. The benefits of support groups can be explained by social support proximal factors, for example, attachment, group membership, and collective identity [18]. Secure adult attachment is described as having comfort, intimacy, and trust such that another person can provide support, empathy, and care. Thus, individuals with chronic illnesses seek comfort from others who can provide this type of attachment [18]. There is also evidence that attachment security benefits individuals with health-related distress, as it was found to be negatively associated with depression among individuals with chronic pain. Therefore, illness support groups can provide security and a positive effect on individuals [18]. Many forums allow users to maintain anonymity, which can encourage more honest and candid discussions, especially about sensitive topics, while being easily accessible to anyone with internet access.

Group membership also describes the evolution of humans having an automatic tendency to categorize themselves and others as in-group and out-group based on similar characteristics, for example, having the same chronic illness [18]. Individuals with chronic illnesses sometimes deviate from social norms, that may impact their physical appearance and actions and can lead to stigmatization, discrimination, and social exclusion, threatening a sense of belonging. Due to these effects, individuals look toward in-group membership, leading to beneficial cohesion [18].

The perceived benefits of participating in support communities include meeting and befriending others with the same condition and similar experiences, learning about the disease and related treatments, giving and receiving emotional support to group members, and accountability for reaching health-related goals [19-21]. Digital communities often serve as spaces where users collaborate to find solutions to common challenges, drawing on the collective wisdom of the group [22].

The Inspire platform is the world's largest and fastest-growing health community designed as an environment for individuals to feel comfortable sharing personal experiences and sensitive health information on an anonymous forum. More than 1o million annual users on the Inspire platform have joined hundreds of disease-specific community groups with patients and caregivers with similar diseases [23]. The Inspire platform maintains anonymity while allowing individuals to obtain information about, and participate in, clinical trials and research efforts [23]. The Interstitial Cystitis Association community on Inspire has over 40,000 members and is the most active IC/BPS forum [23].

Web-based communities provide a rich source of information for health research [24]. For example, monitoring web-based social networks has been an effective tool for identifying adverse events [25] and barriers to accessing health care [26]. The use of web-based social networks for disease surveillance is well documented [27-30]. Notably, data from web-based networks has also been used to obtain a deeper understanding of patient sentiment [31]. Therefore, this study sought to identify (1) sentiments of individuals with IC/BPS toward PPS and (2) topics

XSL•FO
RenderX

discussed by individuals with IC/BPS in conjunction with PPS through use of an internet peer-to-peer forum.

Methods

Study Design and Data Sources

To examine both the sentiment and the topics discussed concerning PPS, the Inspire forum for individuals with IC/BPS was used to derive quantifiable data points. We used the IC/BPS discussion forum posts and replies from the United States from January 12, 2019, to November 30, 2022. In our analytical sample, 3232 patient forum posts were collected from the Inspire forum, and 354 discussed PPS. The data corpus was created to gain insight into the triggers or factors flaring or worsening IC/BPS symptomology. As such, specific text terms and razor tags were used to bring these topics to the surface: flare-, flair-, worse-, increase-, more, trigger-, caus-, inducing, induce-, producing, produce-, factor, factors, provok-, activat-, exacerbate-, irritat-, hypersensitive-, sensiti-, react-, risky, acute, #flare, #riskfactor, #environmentalfactor, #cause, #causality, #hypersensitivity, and #sensitization. Texts were analyzed for specific discussions regarding PPS and other concerns of individuals with IC/BPS. Conversations and discussion posts were limited to sentences with the term PPS and the trade name for PPS. These sentences were used for further analysis alongside the previous sentence relative to the usage of “PPS” and the following sentence relative to the usage of “PPS”.

All analyses used this definition of PPS-related discussions and were conducted using Python version 3.6 (Python Software Foundation). In addition, a combination of topic modeling and sentiment analysis was used in the retroactive investigation.

Topic Modeling

Topic modeling is statistical modeling for discovering abstract “topics” in a text collection. Latent Dirichlet allocation (LDA) is a statistical topic model used to discover and classify text to a particular collection. LDA builds a topic per document model and words per topic model, creating Dirichlet distributions. Each topic has a probability of generating different words in the text. Topics are surfaced based on the likelihood of word co-occurrence and the probability of words making up each topic.

Before using LDA for topic modeling, raw data were preprocessed (lower cased, removing URL links, nonletter characters, and abbreviations) and tokenized (divided into sets) by returning a list of strings for each post to prepare for analysis using the GenSim package in Python [32]. Furthermore, bigrams were used on the tokenized data to combine frequently occurring words to analyze topics. Bigrams were also implemented on the tokenized data to combine the words that occurred together frequently. Common stop words such as “the,” “a,” “an,” “in,” and so on, were filtered out using the Natural Language Toolkit in Python, which contains a list of stop words stored in the English language [33]. The spacy package was used to lemmatize words (reducing words to their root word, eg, “running” and “ran” can be reduced to “run”) and ensure the thorough removal of stop words. The words “PPS” (and the associated trade name) were also excluded from the LDA analysis because the data were filtered explicitly for discussions about “PPS.” For the intertopic distance map, multidimension scaling is used for the visualization of the topics in a 2D space with the 2 principal components, prinicipal component 1 and principal component 2. We used coherence score to select the potential numbers of topics for LDA model. Coherence score measures the consistency, clarity, and relevance of each topic in LDA modeling [34,35]. We then selected the maximum number of topics that could be produced without overlapping the intertopical distance map. We also used perplexity score to examine the model performance. Perplexity score measures the predication performance of a LDA model on held-out documents [36,37]. The size of the topic represented by circles, is proportional to the proportion of posts about PPS by individuals on Inspire. Topics having more words in common are represented closer together, thus the distance between the topics is related to how different the topics are in the posts.

After preprocessing and tokenizing the data, each forum post was categorized to the LDA topic with the highest proportion to identify hidden topics within LDA modeling. Words with the highest probably were ranked for each topic. Finally, topics were manually labeled to determine the overarching theme for each topic (AS, EL, CAB, and ALG).

Sentiment Analysis

Sentiment analysis was used to computationally identify and categorize opinions expressed on the IC/BPS Inspire forum to find the emotional tone behind PPS-associated text. Sentiment analysis was conducted using the Valence Aware Dictionary for Sentiment Reasoning (VADER). VADER is a lexicon and rule-based sentiment analysis tool to identify the sentiment attributed to a topic of interest. It is also considered the standard for analyzing sentences to determine the sentiment of a piece of literature.

The results generated by VADER include a dictionary of 4 percentage scores classified as negative, neutral, positive, and compound. The negative, neutral, and positive scores are the relative percentages of negative, neutral, and positive sentiments surrounding the topic discussion. The compound score is a continuous score ranging from −1 (most negative) to +1 (most positive). Definitions most commonly used for compound scores are: −1.0 to −0.5 (a negative score), −0.5 to +0.5 (a neutral score), and +0.5 to +1.0 (a positive score). Sentiment analysis was conducted on all PPS-related discussions and on all topics determined by LDA topic.

Ethical Considerations

Inspire community members may anonymously engage with others through public discussion posts in the forum. The participants agree to have their posts or discussion comments used for research purposes upon registering for Inspire. Deidentified community posts, with all personally identifiable information removed, are available for secondary data analysis. Individuals were not compensated for participating in the forum. This study was approved by the Boston Children’s Hospital Institutional Review Board (IRB-P00040767).

XSL•FO
RenderX

Results

Overview

Results are based on the 354 posts that related to PPS. Examples of forum posts with sentiment categorization by VADER are listed in Table 1.

Table . Select interstitial cystitis/bladder pain syndrome forum posts describing their experiences with pentosan polysulfate sodium. Posts with positive, negative, and neutral sentiment are shown.

Example notable quotes	Sentiment categorization by VADER[a]
"IC is not just an inflamed bladder. It is a disease with no known cause or cure. It is a lifetime of pain and misery. I know I have had it for over 20 years now. IC causes the lining of your bladder to shed each time you urinate. It is so painful it feels like lava coming through your thighs."	Overall neutral sentiment toward topic 1 (ineffectiveness or discontinuing use)
"Interstitial Cystitis is insidious, terribly painful, and has been a direct causality of what I can and could not do throughout the entire duration of my lifetime."	Overall negative sentiment toward topic 3 (personal treatment suggestions based on the experience)
"You must be patient and not give up hope, be willing to try different remedies as not one thing works for everyone. Relief can be fleeting and not every day since the condition is chronic."	Overall positive sentiment toward topic 3 (personal treatment suggestions based on the experience)
"…with all the class action suits, I got scared and stopped taking [...] I have had several flare-ups since I stopped and no quality of life, as someone else said. I think eyesight is a big side effect to consider, but the alternative is pain [...] still deciding if I should go back on it?"	Overall neutral sentiment toward topic 1 (ineffectiveness or discontinuing use)
"[...] actually would give me a reasonable baseline of normal, but I have chosen to go off of it due to the new vision concerns. I am praying that they will determine the vision issue to be very low risk, and I will consider going back on because it is really what works best for me."	Overall positive sentiment toward topic 2 (alternative treatment)
"I foolishly went off [...] in January because of the possible retinal eye damage and of course the high cost. Really big mistake. I'm back on [...] and hoping that it doesn't take 3 - 4 months for it to kick in again."	Overall negative sentiment toward topic 3 (personal treatment suggestions based on the experience)
"a lifesaver because [they] also [were] in severe pain, burning, etc, and running to the bathroom. After [...], [they] had no pain, [they] could eat anything [they] wanted."	Overall negative sentiment toward topic 5 (risk of long-term use)
"I am so glad to find someone besides myself that isn't stopping [...] I was suicidal before [...] went to work on me. The fear of returning to the hot mess I was before taking [...] terrifies me. I am comforted to find out I am not the only one sticking with it."	Overall negative sentiment toward topic 5 (risk of long-term use)
"[...] is garbage and has terrible side effects."	Overall negative sentiment toward topic 1 (ineffectiveness or discontinuing use)
"antihistamines, amitriptyline, [...], and an ultra-strict diet seem to help for a few days, and then I am right back to where I started."	Overall negative sentiment toward topic 1 (ineffectiveness or discontinuing use)

[a]VADER: Valence Aware Dictionary for Sentiment Reasoning.

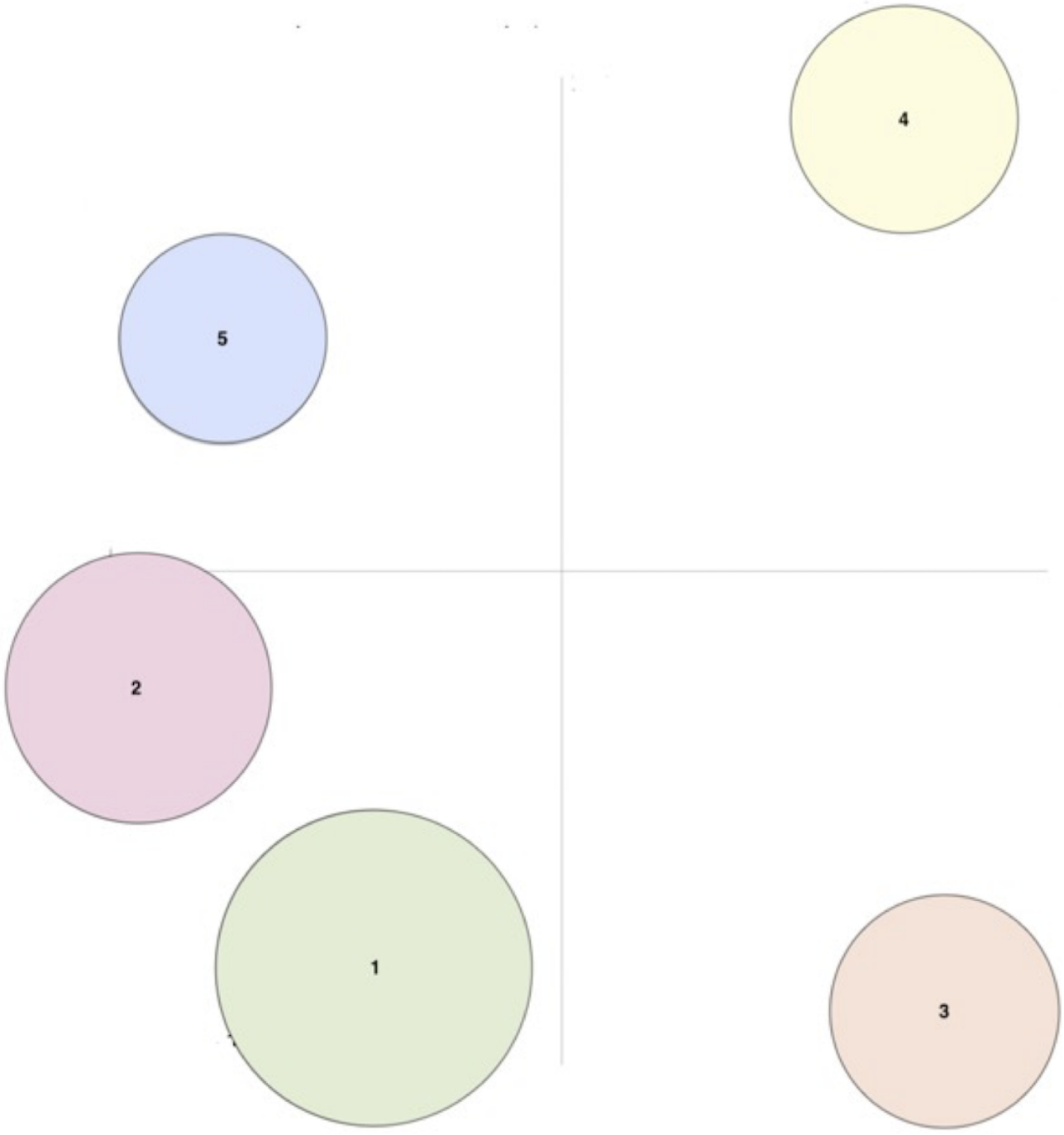
In Table 1, all posts were analyzed using VADER for sentiment categorization. To determine sentiment, VADER begins by splitting the text into individual words. It then assigns each word a score to indicate whether it is positive or negative. Using these predefined scores, VADER computes an overall sentiment score for the entire text. VADER also accounts for sentiment intensity (eg, capitalization and punctuation) and looks for modifiers that might alter the meaning of nearby words.

LDA Modeling Output

LDA topic modeling revealed 5 topics, the minimum number without overlap. Figure 1 displays the intertopic distance map, the area of the topic circles proportional to the number of words that belong to each topic. The distance between topics represents the degree of difference between each. Topics having more words in common are closer together. The greatest proportion of discussions was topic 1 (152/354), then topic 2 (92/354), topic 3 (56/354), topic 4 (28/354), and topic 5 (26/354). In Figure 1, topics 1, 2, and 5 were more closely clustered than topics 3 and 4, indicating they were most dissimilar to topics 1, 2, and 5. The highest probability of co-occurring words making up each LDA topic is shown in Table 2. Using the co-occurence of words and manual investigation, Topic 1 was identified as "ineffectiveness or discontinued use," topic 2 as "alternative treatments," topic 3 as "personal treatment suggestions based on experience," topic 4 as "severe side effects," and topic 5 as "risk of long-term use."

XSL•FO
RenderX

Figure 1. Display of intertopic distance and marginal distribution of the 5 topics. Instead of representing the divergence of topics as a phylogentic tree in 1D, this map allows visualization of how topics relate in 2D space–based principal components. Intertopic distance map measures the distance between topics. Similar topics are closer to each other, while very different topics are farther apart. The size of the circle represents the number of words in each topic; larger circles mean more words belong to that topic.



XSL•FO
RenderX

Table . Latent Dirichlet allocation word occurrence probability for pentosan polysulfate sodium–related discussions in the interstitial cystitis/bladder pain syndrome forum.

Topic 1: ineffectiveness or discontinued use (152/354, 42.9%)		Topic 2: alternative treatments (92/354, 26%)		Topic 3: personal treatment suggestions based on the experience (56/354, 15.8%)		Topic 4: severe side effects (28/354, 7.9%)		Topic 5: risk of long-term use (26/354, 7.3%)	
Probability	Word	Probability	Word	Probability	Word	Probability	Word	Probability	Word
.08	Take	.08	Go	.12	Eye	.05	Try	.05	Night
.08	Year	.08	Also	.09	Damage	.04	Bladder	.04	Problem
.04	Stop	.07	Pain	.07	Cause	.04	Start	.04	Really
.04	Flare	.04	Help	.06	Retinal	.04	Have Been	.04	Think
.03	IC	.04	Month	.05	See	.03	Back	.03	Pigmentary
.03	Time	.03	Day	.04	Issue	.03	Continue	.03	Vision
.03	Work	.02	Capsule	.04	Hear	.03	Make	.03	Maculopathy
.03	Find	.02	Get	.03	Vision	.03	Still	.03	Macular degeneration
.02	Thing	.02	Med	.03	Retina	.03	Medication	.02	End
.02	Know	.02	Diet	.03	Take	.03	Burn	.02	Cost
.02	*Help*	.02	Bladder	.03	Ask	.02	Due	.02	Symptom
.02	People	.02	Give	.03	Drug	.02	Want	.02	Many
.02	Try	.02	Back	.03	Condition	.02	Give	.02	Much
.02	Dr	.01	Week	.02	Doctor	.02	Ever	.02	Risk
.02	Hope	.01	Mg	.02	Long	.02	Instillation	.02	Course
.02	Lot	.01	Seem	.02	Hair	.02	Fall	.02	Patient
.02	Life	.01	Urologist	.02	Light	.02	Free	.02	Hour
.02	Month	.01	Time	.01	Specific	.02	Develop	.02	Concern
.01	Possible	.01	Food	.01	Term	.02	Little	.02	Hydroxyzine
.01	Day	.01	Eat	.01	Stop	.02	Pill	.02	Helpful

In Table 2, the LDA modeling method is used to reveal the main topics and how they are distributed across a dataset (ie, the IC/BPS forum). Five main topics were identified. The proportion, number of posts discussing each topic, and manual topic labels for each LDA topic are shown.

Sentiment and Topic Analysis

The overall sentiment scores for each topic are presented in Table 3. Overall, the VADER compound sentiment of discussions about PPS was −0.083, (negative sentiment 32.49% [115/354], positive 22.03% [78/354], and neutral 45.48% [161/354]).

Table . Sentiment scores for discussions surrounding pentosan polysulfate sodium (n=354). Posts were stratified by the identified topics from the latent Dirichlet allocation analysis.

Topic number and label	Average compound score (SD)	Negative, n (%)	Positive, n (%)	Neutral, n (%)
Total (n=354)	−0.083 (0.603)	115 (32.49)	78 (22.03)	161 (45.48)
Topic 1: ineffectiveness or discontinuing use (n=152)	−0.094 (0.625)	52 (34.21)	35 (23.03)	65 (42.76)
Topic 2: alternative treatments (n=92)	−0.082 (0.620)	31 (33.7)	23 (25)	38 (41.3)
Topic 3: personal treatment suggestions based on the experience (n=56)	−0.004 (0.573)	15 (26.79)	12 (21.43)	29 (51.79)
Topic 4: severe side effects (n=28)	−0.162 (0.523)	9 (32.14)	4 (14.29)	15 (53.57)
Topic 5: risk of long-term use (n=26)	−0.119 (0.581)	8 (30.77)	4 (15.38)	14 (53.85)

XSL•FO
RenderX

In Table 3, each topic number represents a specific theme that emerged during the analysis of the text corpus. The average compound score indicates whether the overall tone of discussions for each topic is positive, negative, or neutral. The number and percentage of scored negative, postitive, and neutral posts is shown. By comparing sentiment across these topic numbers, we can gain insights into how different aspects of PPS are perceived within the discussions.

Topic 1 identified by the LDA topic modeling was manually labeled "ineffectiveness" based on the total posts about PPS on the IC/BPS Inspire forum using its relative size in Figure 1 and number of posts in Table 3. In Topic 1, the highest probability of co-occurring and relevant words compromising this topic were: take (.08), year (.08), stop (.04), and flare (.04). Its distance in relationship to topics 2 (alternative treatments) and 5 (risk of long-term use), seen in Figure 1, indicates a higher degree of similarity based on the types of words they have in common. Topics 4 (severe side effects) and 3 (personal treatment suggestions based on experience) were furthest from topic 1. shows the overall compound scores, and percent negative, positive, and neutral for the entire dataset and for each topic of PPS posts by individuals from Inspire. Topic 1 had the highest percentage of negative posts with 34.21% (52/152), 23.03% (35/152) positive sentiment posts, while the percent of neutral sentiment was 42.76% (65/152). However, despite having the most percent negative posts compared with any other topic, the average compound score was −0.094 which falls within the neutral compound score range of −0.50 to +0.50.

Topic 2 was identified by LDA modeling and manually labeled as "alternative treatments." This topic had the most overlap of words in common with topic 1 Figure 1. Topic 2 was the second largest topic with 26% (92/354) of the total posts about PPS on the IC/BPS Inspire forum. The highest probability of co-occurring words most relevant terms to the topic was: pain (.07), help (.04), diet (.02), urologist (.01), time (.01). Moreover, it had the second-largest proportion of discussions on the IC/BPS Inspire forum and the second most posts associated with the topic; note its relative size (Figure 1). Compared with the other 4 topics, topic 2 had the second-highest percentage of negative posts at 33.7% (31/92). However, it also had the highest percentage positive sentiment score (23/92, 25%) and percentage of neutral (38/92, 41.3%). The overall compound score in topic 2 was −0.082, the second most negative, falling within the neutral compound score range of −0.50 to +0.50.

Topic 3 was identified by LDA modeling and manually labeled as "personal treatment suggestions based on experience." It was the third largest topic with 15.8% (56/354) about PPS on the IC/BPS Inspire forum. The highest probability of co-occurring words most relevant to the topic was: eye (.12), damage (.09, retinal (.06), vision (.03), retina (.03), and doctor (.02). Compared with the other 4 topics, topic 3 had the least negative compound sentiment (−0.004) and the least percent of negative (15/26, 26.79%), positive (12/56, 21.43%), and (29/56, 51.79%) neutral posts. This is reflective of the distance of topic 3 in the intertopic distance map whereby it is further from to topic 2.

Topic 4 was identified by the LDA modeling and manually labeled as "severe side effects." It was the second smallest topic with 7.9% (28/354) of the total posts about PPS on the IC/BPS Inspire forum. The highest probability of co-occurring words most relevant to the topic was bladder (.04), medication (.03), burn (.02), and instillation (.02). It had the highest intertopic distance compared with the topics 2, 1, and 3, making it the least similar topic based on word commonality. Topic 4 had the most negative compound score, −0.162 with 32.14% (9/28) negative, 14.29% (4/28) positive, and 53.57% (15/28) neutral posts.

Topic 5 was identified by the LDA modeling and manually labeled as "risk of long-term use." It was the smallest topic with 7.3% (26/354) about PPS on the IC/BPS Inspire forum. The highest probability of co-occurring words most relevant to the topic was: night (.05), problem (.04), pigmentary (.03), vision (.03), maculopathy (.03), and macular degeneration (.03). The average compound score for was −.119 (SD 0.581) which falls within the neutral compound range of −0.50 to +0.50. It had the highest percentage of neutral sentiment posts (14/26, 53.85%), compared with the other topics.

Discussion

Principal Findings

PPS is widely prescribed to relieve IC/BPS symptoms [1]; it is the only FDA-approved oral medication available for symptom alleviation. However, PPS has been linked to a unique presentation of pigmentary maculopathy [16]. This potential side-effect has created a need for individuals exposed to PPS to be consistently monitored by an optometrist for early detection [11,13].

Although previous studies have assessed the benefits and side effects of PPS, they have done so separately without capturing individuals' thoughts on if or why they are willing to assume the side effects for the relief of pain. This study provides information as to the sentiment of the IC/BPS community toward PPS, which was determined to be neutral, with 5 topics discussed in the forum all producing a neutral compound score.

Studying sentiment toward medications offers several important benefits, beginning with improved patient-centered care that aligns with individual preferences, concerns, and quality of life. An understanding of patient sentiment may assist health care professionals in assessing whether the perceived benefits of a medication outweigh the risks from the patient's perspective. As the data here indicate that patient sentiment is neither strongly for nor against PPS, patients may be less influenced by emotional or external biases, making it easier to assess the medication's effects based on objective outcomes rather than preconceived notions. This can inform shared decision-making and guide personalized treatment options. It is arguable that understanding patient's sentiment toward their medication could reduce anxiety and improve trust, both of which are related to treatment adherence [38].

As a whole, the forum data describe a willingness among some individuals to risk the potential side effects associated with PPS due to its effectiveness at relieving pain. Multiple posts indicate PPS, "saved [their] sanity" or, "changed [their] life" (Table 1). However, the proportion of negative sentiment is close to

XSL•FO
RenderX

one-third (115/354, 32.49%), with positive sentiment taking up a proportion of just over one-fifth (78/354, 22.03%) and neutral just under one half (161/354, 45.48%). The mean compound sentiment score in the forum is −0.083 (SD 0.603), indicating an overall neutral sentiment.

Topics 4 (severe side effects) and 5 (risk of long-term use), were the least discussed. While topic 4 did not surface as often in the LDA model and had a lower discussion frequency, it did have the lowest positive sentiment. Topic 5 had the highest neutral sentiment of the topics.

Limitations

Some drawbacks of using web-based health communities such as Inspire include difficulty controlling the quality of information shared on such sites. Uncertain and unknowable generalizability of these results is another limitation. Other potential drawbacks include privacy risks to patients and caregivers utilizing the platform, as "data intruders" could lead to discrimination by employers, insurance companies, family, and friends. Although there are policies, such as the Genetic Information Nondiscrimination Act, which prohibits genetic information discrimination in employment, privacy is still a risk that must be considered [20] and individuals may be unwilling to speak freely.

VADER uses a lexicon-based approach system. In such an approach, while broader term coverage is made available, words have a finite assignment to sentiment orientation, which does not consider figures of speech or negations, as mentioned above [39]. LDA topic modeling limitations include the number of topics that can be generated and the inability to depict correlations [40,41]. Topics are not developed over time, and the modeling assumes that words are exchangeable and that sentence structure is not modeled.

Confounding factors, such as current symptoms and overall IC/BPS management, should be included in future studies to further the understanding of the underlying determinants of sentiment toward PPS. Comparitive attitudes toward different IC/BPS treatments should also be explored to assess varying disease management strategies.

Conclusions

Individuals use health forums such as Inspire to learn more about their disease, medications, and treatments. Therefore, this study aimed to use forum posts from Inspire to investigate individuals with IC/BPS' overall sentiment toward PPS. Using VADER, sentiments regarding PPS were found to be neutral. These results also suggest that patients are open to considering both the benefits and risks of PPS, which can lead to informed and balanced decision-making regarding their IC/BPS treatment. Both the breadth of discussion topics and sentiment demonstrate the complexity inherent in individual consideration of a medication with publicized side effects. Future studies of sentiment analyses toward other IC/BPS treatments and their comparison to PPS may provide additional context and direction in the treatment of this painful condition.

Acknowledgments

We thank the patients, families, and caregivers for their participation in this study. This publication was supported by the Centers for Disease Control and Prevention of the U.S. Department of Health and Human Services (HHS) as part of a financial assistance award totaling $2,400,000 with 100 percent funding from the CDC/HHS (U01DP006634-01). The contents are those of the author(s) and do not necessarily represent the official views of, nor are they endorsed by, the CDC/HHS or the U.S. Government.

Data Availability

Aggregated and anonymized data are available upon reasonable request to Inspire [42].

Conflicts of Interest

None declared.

References

1. Lim Y, O'Rourke S. Interstitial Cystitis: StatPearls Publishing; 2022.
2. Messing EM, Stamey TA. Interstitial cystitis: early diagnosis, pathology, and treatment. Urology 1978 Oct;12(4):381-392. [doi: 10.1016/0090-4295(78)90286-8] [Medline: 213864]
3. Hanno P, Nordling J, van Ophoven A. What is new in bladder pain syndrome/interstitial cystitis? Curr Opin Urol 2008 Jul;18(4):353-358. [doi: 10.1097/MOU.0b013e3282fcea88] [Medline: 18520754]
4. Liu HT, Kuo HC. Biomarkers for patients with interstitial cystitis/bladder pain syndrome. Urol Sci 2015 Dec;26(4):225-229. [doi: 10.1016/j.urols.2015.02.002]
5. Akiyama Y, Hanno P. Phenotyping of interstitial cystitis/bladder pain syndrome. Int J Urol 2019 Jun;26 Suppl 1:17-19. [doi: 10.1111/iju.13969] [Medline: 31144756]
6. Estrella E, Rockowitz S, Thorne M, et al. Mendelian disorders in an interstitial cystitis/bladder pain syndrome cohort. Adv Genet (Hoboken) 2023 Mar;4(1):2200013. [doi: 10.1002/ggn2.202200013] [Medline: 36910591]
7. Warren JW, Jackson TL, Langenberg P, Meyers DJ, Xu J. Prevalence of interstitial cystitis in first-degree relatives of patients with interstitial cystitis. Urology 2004 Jan;63(1):17-21. [doi: 10.1016/j.urology.2003.08.027] [Medline: 14751339]

XSL•FO
RenderX

8. Dimitrakov JD. A case of familial clustering of interstitial cystitis and chronic pelvic pain syndrome. Urology 2001 Aug;58(2):281. [doi: 10.1016/s0090-4295(01)01138-4] [Medline: 11489726]
9. Warren JW, Keay SK, Meyers D, Xu J. Concordance of interstitial cystitis in monozygotic and dizygotic twin pairs. Urology 2001 Jun;57(6 Suppl 1):22-25. [doi: 10.1016/s0090-4295(01)01120-7] [Medline: 11378045]
10. Parsons CL, Schmidt JD, Pollen JJ. Successful treatment of interstitial cystitis with sodium pentosanpolysulfate. J Urol 1983 Jul;130(1):51-53. [doi: 10.1016/s0022-5347(17)50948-9] [Medline: 6191049]
11. Pearce WA, Chen R, Jain N. Pigmentary maculopathy associated with chronic exposure to pentosan polysulfate sodium. Ophthalmology 2018 Nov;125(11):1793-1802. [doi: 10.1016/j.ophtha.2018.04.026] [Medline: 29801663]
12. Parkinson R, Thurau S, Maag KP. Pigmentary maculopathy in interstitial cystitis/bladder pain syndrome treated with oral pentosan polysulfate: a review. EMJ Urol 2022 Aug 15:74-81. [doi: 10.33590/emjurol/10054331]
13. Kailavasan M, Goddard JC. Association of Elmiron (pentosanpolysulphate sodium) with pigmented maculopathy: an update for urologists and patients. J Clin Urol 2023 Sep;16(5):458-462. [doi: 10.1177/20514158211053699]
14. Doiron RC, Bona M, Nickel JC. Possible drug-induced, vision-threatening maculopathy secondary to chronic pentosan polysulfate sodium (Elmiron®) exposure. Can Urol Assoc J 2020 Feb;14(2):10-11. [doi: 10.5489/cuaj.6401] [Medline: 31999543]
15. Higgins K, Welch RJ, Bacorn C, et al. Identification of patients with pentosan polysulfate sodium-associated maculopathy through screening of the electronic medical record at an academic center. J Ophthalmol 2020;2020:8866961. [doi: 10.1155/2020/8866961] [Medline: 33489347]
16. Philip AM, Wannamaker KW, Miller DM. Prevalence and dose dependency analysis of pentosan polysulfate sodium maculopathy. Ophthalmic Epidemiol 2023 Feb;30(1):82-87. [doi: 10.1080/09286586.2022.2031227] [Medline: 35081852]
17. Davison KP, Pennebaker JW, Dickerson SS. Who talks? The social psychology of illness support groups. Am Psychol 2000 Feb;55(2):205-217. [Medline: 10717968]
18. Lehardy EN, Fowers BJ. Ultimate (evolutionary) explanations for the attraction and benefits of chronic illness support groups: attachment, belonging, and collective identity. Curr Psychol 2020 Aug;39(4):1405-1415. [doi: 10.1007/s12144-018-9841-7]
19. Delisle VC, Gumuchian ST, Rice DB, et al. Perceived benefits and factors that influence the ability to establish and maintain patient support groups in rare diseases: a scoping review. Patient 2017 Jun;10(3):283-293. [doi: 10.1007/s40271-016-0213-9] [Medline: 28004275]
20. Solberg LB. The benefits of online health communities. Virtual Mentor 2014 Apr 1;16(4):270-274. [doi: 10.1001/virtualmentor.2014.16.04.stas1-1404] [Medline: 24735575]
21. Wicks P, Keininger DL, Massagli MP, et al. Perceived benefits of sharing health data between people with epilepsy on an online platform. Epilepsy Behav 2012 Jan;23(1):16-23. [doi: 10.1016/j.yebeh.2011.09.026] [Medline: 22099528]
22. Wicks P, Massagli M, Frost J, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010 Jun 14;12(2):e19. [doi: 10.2196/jmir.1549] [Medline: 20542858]
23. Inspire. URL: https://www.inspire.com [accessed 2025-01-02]
24. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection--harnessing the web for public health surveillance. N Engl J Med 2009 May 21;360(21):2153-2155. [doi: 10.1056/NEJMp0900702] [Medline: 19423867]
25. Lee JY, Lee YS, Kim DH, Lee HS, Yang BR, Kim MG. The use of social media in detecting drug safety-related new black box warnings, labeling changes, or withdrawals: scoping review. JMIR Public Health Surveill 2021 Jun 28;7(6):e30137. [doi: 10.2196/30137] [Medline: 34185021]
26. Kline E, Garrett AL, Brownstein C, et al. Using social media listening to understand barriers to genomic medicine for those living with Ehlers-Danlos syndromes and hypermobility spectrum disorders. Health Expect 2023 Aug;26(4):1524-1535. [doi: 10.1111/hex.13755] [Medline: 37062887]
27. Bisanzio D, Kraemer MUG, Brewer T, Brownstein JS, Reithinger R. Geolocated Twitter social media data to describe the geographic spread of SARS-CoV-2. J Travel Med 2020 Aug 20;27(5):taaa120. [doi: 10.1093/jtm/taaa120] [Medline: 32701135]
28. Harris JK, Hawkins JB, Nguyen L, et al. Using Twitter to identify and respond to food poisoning: The Food Safety STL Project. J Public Health Manag Pract 2017;23(6):577-580. [doi: 10.1097/PHH.0000000000000516] [Medline: 28166175]
29. Young SD, Rivers C, Lewis B. Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. Prev Med 2014 Jun;63:112-115. [doi: 10.1016/j.ypmed.2014.01.024] [Medline: 24513169]
30. Clemente L, Lu F, Santillana M. Improved real-time influenza surveillance: using internet search data in eight Latin American countries. JMIR Public Health Surveill 2019 Apr 4;5(2):e12214. [doi: 10.2196/12214] [Medline: 30946017]
31. Sewalk KC, Tuli G, Hswen Y, Brownstein JS, Hawkins JB. Using Twitter to examine web-based patient experience sentiments in the United States: longitudinal study. J Med Internet Res 2018 Oct 12;20(10):e10043. [doi: 10.2196/10043] [Medline: 30314959]
32. Haider MM, Hossin M, Mahi HR, Arif H. Automatic text summarization using gensim word2vec and k-means clustering algorithm. Presented at: 2020 IEEE Region 10 Symposium (TENSYMP); Jun 5-7, 2020; Dhaka, Bangladesh. [doi: 10.1109/TENSYMP50017.2020.9230670]

XSL•FO
RenderX

33. Bird S, Klein E, Loper E. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit: O'Reilly Media, Inc; 2009, Vol. ISBN.
34. O'Callaghan D, Greene D, Carthy J, Cunningham P. An analysis of the coherence of descriptors in topic modeling. Expert Syst Appl 2015 Aug;42(13):5645-5657. [doi: 10.1016/j.eswa.2015.02.055]
35. Syed S, Spruit M. Full-text or abstract? Examining topic coherence scores using latent Dirichlet allocation. Presented at: 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA); Oct 19-21, 2017; Tokyo, Japan. [doi: 10.1109/DSAA.2017.61]
36. Zhao W, Chen JJ, Perkins R, et al. A heuristic approach to determine an appropriate number of topics in topic modeling. BMC Bioinformatics 2015;16:S8. [doi: 10.1186/1471-2105-16-S13-S8] [Medline: 26424364]
37. Rüdiger M, Antons D, Joshi AM, Salge TO. Topic modeling revisited: new evidence on algorithm performance and quality metrics. PLoS ONE 2022;17(4):e0266325. [doi: 10.1371/journal.pone.0266325] [Medline: 35482786]
38. Costello K, Kennedy P, Scanzillo J. Recognizing nonadherence in patients with multiple sclerosis and maintaining treatment adherence in the long term. Medscape J Med 2008;10(9):225. [Medline: 19008986]
39. D'Andrea A, Ferri F, Grifoni P, Guzzo T. Approaches, tools and applications for sentiment analysis implementation. IJCA 2015;125(3):26-33. [doi: 10.5120/ijca2015905866]
40. Barde BV, Bainwad AM. An overview of topic modeling methods and tools. Presented at: 2017 International Conference on Intelligent Computing and Control Systems (ICICCS); Jun 15-16, 2017; Madurai, India. [doi: 10.1109/ICCONS.2017.8250563]
41. Lafferty J, Blei D, editors. Correlated Topic Models: MIT Press; 2005.
42. Inspire research. Inspire. URL: https://inspireresearch.com/ [accessed 2025-01-07]

Abbreviations

FDA: Food and Drug Administration
IC/BPS: interstitial cystitis/bladder pain syndrome
LDA: latent Dirichlet allocation
OR: odds ratio
PPS: pentosan polysulfate sodium
VADER: Valence Aware Dictionary for Sentiment Reasoning

Edited by A Mavragani; submitted 03.11.23; peer-reviewed by E Said-Hung, R Taneja; revised version received 04.10.24; accepted 05.10.24; published 17.01.25.

Please cite as:
Hswen Y, Qin Q, Smith P, Swierczynski A, Bauer S, Ladson E, Garrett AL, Brownstein CA
Sentiments of Individuals with Interstitial Cystitis/Bladder Pain Syndrome Toward Pentosan Polysulfate Sodium: Infodemiology Study
JMIR Form Res 2025;9:e54209
URL: https://formative.jmir.org/2025/1/e54209
doi:10.2196/54209

© Yulin Hswen, Qiuyuan Qin, Pressley Smith, Alison Swierczynski, Stuart Bauer, Erika Ladson, Amanda Leigh Garrett, Catherine A Brownstein. Originally published in JMIR Formative Research (https://formative.jmir.org), 17.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Using Health Information Resources for People With Cognitive Impairment (digiDEM Bayern): Registry-Based Cohort Study

Florian Weidinger[1], MSc; Nikolas Dietzel[1], PhD; Elmar Graessel[2], MD, PhD; Hans-Ulrich Prokosch[3], PhD; Peter Kolominsky-Rabas[1], MBA, MD, PhD

[1]Interdisciplinary Centre for Health Technology Assessment and Public Health, Friedrich-Alexander-Universität Erlangen-Nürnberg, Schwabachanlage 6, Erlangen, Germany

[2]Department of Psychiatry and Psychotherapy, Center for Health Services Research in Medicine, Universitätsklinikum Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

[3]Department of Medical Informatics, Biometrics and Epidemiology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

Corresponding Author:
Florian Weidinger, MSc
Interdisciplinary Centre for Health Technology Assessment and Public Health, Friedrich-Alexander-Universität Erlangen-Nürnberg, Schwabachanlage 6, Erlangen, Germany

Abstract

Background: Dementia is a growing global health challenge with significant economic and social implications. Underdiagnosis of dementia is prevalent due to a lack of knowledge and understanding among the general population. Enhancing dementia literacy through improved health information–seeking behavior is crucial for the self-determined management of the disease by those affected. Understanding the relationship between dementia literacy, health information–seeking behavior, and the use of various information sources among individuals with cognitive impairment is of high importance in this context.

Objective: The aim of this study was to analyze the relevance of different sources of health information from the perspective of people with cognitive impairment, while also evaluating differences based on age, gender, and disease progression.

Methods: This study is part of the ongoing project "Digital Dementia Registry Bavaria – digiDEM Bayern." The Digital Dementia Registry Bavaria is a multicenter, prospective, longitudinal register study in Bavaria, Germany. People with cognitive impairment rated several information sources by using Likert scales with the values unimportant (1) to very important (5). Data were analyzed descriptively, and multiple 2-sample, 2-tailed *t* tests were used to evaluate differences by cognitive status and gender and using multiple one-way ANOVA to evaluate differences by age group.

Results: Data of 924 people with cognitive impairment (531 with dementia, 393 with mild cognitive impairment) were evaluated. The most relevant health information sources were "Personal visit to a medical professional" (mean 3.9, SD 1.1) and "Family / Friends" (mean 3.9, SD 1.2). "Internet" was 1 of the 2 lowest-rated information sources by people with cognitive impairment (mean 1.6, SD 1.1), with nearly three-quarters (684/924, 74%) of the participants rating the source as unimportant. The age-specific analyses showed significant differences for the sources "Internet" ($F_{2,921}$=61.23; P<.001), "Courses / Lectures" ($F_{2,921}$=18.88; P<.001), and "Family / Friends" ($F_{2,921}$=6.27; P=.002) for the 3 defined age groups. There were several significant differences between people with mild cognitive impairment and dementia whereby the first group evaluated most sources higher, such as "Internet" (mean difference=0.6; t_{640}=7.52; P<.001). The only sources rated higher by the dementia group were "TV / Radio" and "Family / Friends," with none of them showing significant differences. Gender-specific analyses showed women with cognitive impairment valuing every evaluated source higher than men apart from "Internet" (mean difference=0.4; t_{685}=4.97; P<.001).

Conclusions: To enhance health and dementia literacy, the best way to communicate health information to people with cognitive impairment is through interpersonal contact with medical professionals and their friends and family. Slight changes in valuation should be considered as the medical condition progresses, along with variations by age and gender. In particular, the evaluation and use of the internet are dependent on these factors. Further research is needed to capture potential changes in the valuation of the internet as a health information source.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2020-043473

(JMIR Form Res 2025;9:e54460) doi:10.2196/54460

KEYWORDS

dementia; mild cognitive impairment; cognitive impairment; information sources; health information; health information–seeking behavior; Digital Dementia Registry Bavaria; digiDEM

XSL•FO
RenderX

Introduction

Dementia is one of the most significant public health challenges of our time, and it will continue to be so in the future. To date, there are an estimated 57.4 million people with dementia worldwide, with estimates of 78 million people affected by 2030 and up to 152.8 million cases by 2050 [1,2]. In financial terms, these figures translate into an estimated global cost of dementia of US $1313.4 billion, which equates to an annual cost of US $23,796 per person with dementia [3]. Estimations also show that 75% of people with dementia worldwide live without a confirmed diagnosis. One of the reasons for this underdiagnosis is a lack of knowledge and understanding of dementia and its symptoms [1].

One way to address this problem of underdiagnosis of dementia is to improve health literacy and, therefore, dementia literacy [4]. However, levels of health literacy are low across Europe, particularly in Germany. According to the European Health Literacy Population Survey 2019 - 2021, 46% of the European population and 72% of the German population had limited health literacy in 2021. These figures were even more pronounced for vulnerable subpopulations such as people of advanced age or people with chronic diseases [5].

Nutbeam [6] defines health literacy as "the ability of individuals to gain access to, understand, and use information to promote and maintain good health." Dementia literacy is a subcategory and describes the knowledge for the recognition, management, and prevention of dementia [7]. There are multiple conceptual frameworks that describe the model of health literacy with most of them recognizing the process of using information as an essential element of health literacy. One example is the integrated model of health literacy by Sørensen et al [8], which is based on the 4 competencies of access, understanding, appraisal, and application. The first competency includes the ability to seek, find, and obtain health information. This implies that health information–seeking behavior also plays an important role in the process of how health literacy affects health outcomes [9].

Zimmerman et al [10] describe health information–seeking behavior as the purposeful behavior by an individual to find health information. The strategies for finding health information can take different forms, for example, through active seeking or more passive measures. An important aspect of health information–seeking behavior is which sources people use to obtain health information [10]. This implies a relationship between health literacy and the use of health information sources [9].

According to a recent survey conducted in 2021, 55% of Europeans aged 16 - 74 years have sought health-related information via the web [11]. The US Health Information National Trends Survey indicates a similar situation for the American general population, with 72.7% using the internet first for their most recent search for health information in 2019 [12]. However, health information–seeking behavior and preferences for information sources depend heavily on various individual aspects, such as gender and age, as well as situational aspects, such as having a chronic disease like dementia [13,14]. Women are generally more interested in health information than men and also use health information sources more frequently [15]. For chronic conditions, Oh and Cho [16] found that chronic patients were generally more likely to search for health information than healthy individuals. To date, no studies have investigated which sources of health information people with cognitive impairment use and how they value them.

Therefore, the following research questions (RQs) were asked: RQ1 (How do people with cognitive impairment evaluate different sources of health information?), RQ2 (Are there differences in the evaluation based on the age of the person affected?), RQ3 (Are there differences in the evaluation based on the gender of the person affected?), and RQ4 (Are there differences in the evaluation based on the progress of the medical condition of the person affected?).

Methods

Study Design

This study is part of the ongoing project "Digital Dementia Registry Bavaria – digiDEM Bayern." The Digital Dementia Registry Bavaria is a multicenter, prospective, longitudinal register study conducted in all administrative regions of Bavaria. The detailed methodology of the project is described elsewhere [17,18].

Study Population

Participants are people with mild cognitive impairment (MCI) and people with mild or moderate dementia living in Bavaria, who are in the following collectively referred to as people with cognitive impairment. MCI is a memory impairment that results in below the expected performance for the patient's age and level of education. People with MCI are not demented and generally maintain their independence in functional abilities of daily life [19,20]. To identify eligible participants, people have to undergo a screening based on the Mini-Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA) prior to inclusion [21,22]. If available, a family caregiver is included with the person with cognitive impairment [17,18].

Recruitment and Data Collection

Participants are recruited by specially trained research partners in all of the 7 administrative regions in Bavaria, beginning in August 2020. Research partners are institutions that are specialized and have experience in the management and care of people with cognitive impairment and their family caregivers. Data collection is conducted through standardized face-to-face interviews using a web-based data entry system [17,18].

Ethical Considerations

The Digital Dementia Registry Bavaria is performed following the ethical standards of the Declaration of Helsinki, and it obtained ethical approval from the ethics committee of the Medical Faculty of the Friedrich-Alexander-Universität Erlangen-Nürnberg (application number: 253_20 B). Informed consent from the participants or their authorized representative is acquired before screening and study inclusion. The project collects and stores all personal data separately from the registry

XSL•FO
RenderX

data on different stand-alone systems to ensure data protection. All participants are pseudonymized. The data protection concept was approved by the local data protection supervisor of the Friedrich-Alexander-Universität Erlangen-Nürnberg and authorized by the Bavarian Data Protection Commissioner. Participation is voluntary, and participants are not compensated [17].

Measures

Sociodemographic data of the people with cognitive impairment were collected, such as age, gender, educational degree, family status, score in cognitive assessment, and whether they had a family caregiver.

The relevance of 8 different sources of health information was rated by people with cognitive impairment using Likert scales ranging from unimportant (1) to very important (5). Based on previous studies, the rated health information sources were "Internet," "TV / Radio," "Books / Brochures," "Courses / Lectures," "Newspaper / Journals," "Family / Friends," "Pharmacy," and "Personal visit to a medical professional" [23,24]. In order to further evaluate the target group's use of the internet, the participants were asked about the last time they had used the internet. Participants who had last used the internet 3 months ago or had never used it were also asked about their reasons for not using it.

Statistical Analysis

Subgroup analyses were performed for the cognitive status and gender of the people with cognitive impairment using multiple 2-sample, 2-tailed *t* tests and for the age of the participants using multiple one-way ANOVA to evaluate potential differences depending on the progress of the medical condition and gender- and age-specific differences. Based on the age distribution in the sample, the participants were divided into 3 age groups (<70 y, 70 - 79 y, and 80 y or older) for the analysis of age-specific differences. In order to minimize or avoid the probability of a type 1 error, we applied the Bonferroni correction for multiple testing. Accordingly, the adjusted significance level of $P<.002$ was set for all subgroup analyses (nonadjusted significance level: $P<.05$) [25]. Effect sizes (Cohen *d* or η^2) were determined [26]. Statistical analysis was performed using IBM SPSS Statistics (version 28; IBM Corp).

Results

Study Population

In the period between August 2020 and July 2023, in total, 924 people with cognitive impairment were included in the Digital Dementia Registry Bavaria. The following presented analyses refer to the baseline survey.

Table 1 shows the characteristics of the participants in total and for the 2 subgroups "MCI" and "dementia." The age of the study population ranged from 54 to 102 years and was 80.5 (SD 7.7) years on average. Around 59.4% (549/924) of the participants were female, with a mean MMSE score of 22.8 (SD 3.8) points and a mean MoCA result of 19.9 (SD 2.7) points. With 58.2% (538/924), most participants had a middle education, and 50.3% (465/924) were in a partnership. Around 64.1% (592/924) of the people with cognitive impairment regularly received help from a family caregiver at least once per week. Based on their MMSE score, 42.5% (393/924) of the study population was assigned to the MCI group and 57.5% (531/924) of the study population was assigned to the dementia group. There were significant differences between the 2 subgroups for their age ($P<.001$), their MMSE result ($P<.001$), their education ($P<.001$), their family status ($P=.003$), and whether they regularly received support from a family caregiver ($P<.001$).

XSL•FO
RenderX

Table . Description of this study population with the classification into the groups "MCI" (mild cognitive impairment) and "dementia" in a registry-based cohort study in Germany from 2020 to 2023.

		Total (N=924)	MCI (n=393; 42.5%)	Dementia (n=531; 57.5%)	*P* value
Age (years), mean (SD)		80.5 (7.7)	78.3 (7.8)	82.1 (7.2)	<.001
Gender, n (%)					.46
	Female	549 (59.4)	228 (58.0)	321 (60.5)	
	Male	375 (40.6)	165 (42.0)	210 (39.5)	
Family caregiver, n (%)					<.001
	No family caregiver	332 (35.9)	201 (51.1)	131 (24.7)	
	Has family caregiver	592 (64.1)	192 (48.9)	400 (75.3)	
Cognitive assessment, mean (SD)					
	MMSE[a]	22.8 (3.8)	26.5 (1.7)	20.1 (2.5)	<.001
	MoCA[b]	19.9 (2.7)	19.9 (2.7)	—[c]	—
Education, n (%)					<.001
	Lower education	197 (21.3)	62 (15.8)	135 (25.4)	
	Middle education	538 (58.2)	234 (59.5)	304 (57.3)	
	Higher education	189 (20.5)	97 (24.7)	92 (17.3)	
Family status, n (%)					.003
	No partnership	459 (49.7)	173 (44.0)	286 (53.9)	
	In partnership	465 (50.3)	220 (56.0)	245 (46.1)	

[a]MMSE: Mini-Mental State Examination.

[b]MoCA: Montreal Cognitive Assessment.

[c]Not applicable.

General Valuation of Health Information Sources

The most relevant health information sources for people with cognitive impairment were "Family / Friends" as well as "Personal visit to a medical professional" (eg, doctors or physiotherapists) (Table 2). With an average score of 3.9 (SD 1.1), 33.0% (305/924) of the people with cognitive impairment rated "Personal visit to a medical professional" as a very important source, and around three-quarters (694/924, 75.1%) rated it as at least "important." A similar result was found for "Family / Friends," with an average score of 3.9 (SD 1.2), and 39.6% (366/924) of the people with cognitive impairment rated them as a very important source of health information. Going by the highest average score, these 2 sources were followed by "TV / Radio," "Newspaper / Journals," "Pharmacy," and "Books / Brochures." The 2 lowest-rated information sources by people with cognitive impairment were "Internet" and "Courses / Lectures," with an average score of 1.6 (SD 1.1) and 1.6 (SD 1.0) on the Likert scale. Nearly three-quarters (684/924, 74.0%) of the participants rated "Internet" as an unimportant source, with a similar amount (680/924, 73.6%) stating that they had not used the internet in the last 3 months; the majority of those participants (569/680, 83.7%) stated that they had never used the internet before. The most frequently cited reasons for not using the internet were "generally no interest in the internet" (446/680, 65.6%) and "too complicated" (386/680, 56.8%) (Multimedia Appendix 1).

XSL•FO
RenderX

Table . General evaluation of health information sources by people with cognitive impairment using Likert scales ranging from unimportant (1) to very important (5) in a registry-based cohort study in Germany from 2020 to 2023.

	Likert scale ratings (% of participants who selected each rating)					Likert scale score, mean (SD)
	Unimportant	Less important	Partially	Important	Very important	
Internet	74.0	6.8	8.0	7.4	3.8	1.6 (1.1)
TV/Radio	13.7	9.3	20.9	38.5	17.5	3.4 (1.3)
Books/Brochures	41.0	19.9	18.1	14.0	7.0	2.3 (1.3)
Courses/Lectures	69.0	13.6	9.4	6.5	1.4	1.6 (1.0)
Newspaper/Journals	18.8	11.9	19.4	33.7	16.2	3.2 (1.4)
Family/Friends	8.0	7.3	13.2	31.9	39.6	3.9 (1.2)
Pharmacy	26.8	12.3	19.7	28.6	12.6	2.9 (1.4)
Personal visit to a medical professional	6.6	5.7	12.6	42.1	33.0	3.9 (1.1)

Valuation of Health Information Sources by Age

The age-specific analysis shows heterogeneous results for the 8 health information sources (Multimedia Appendix 2). The sources "Internet" ($F_{2,921}$=61.23; P<.001), "Courses / Lectures" ($F_{2,921}$=18.88; P<.001), and "Family / Friends" ($F_{2,921}$=6.27; P=.002) showed significant differences for the 3 defined age groups. People with cognitive impairment valued "Internet" and "Courses / Lectures" lower as age increased, from people under the age of 70 years ("Internet": mean 2.7, SD 1.5; "Courses / Lectures": mean 2.1, SD 1.3), to people aged 70 to 79 years ("Internet": mean 1.8, SD 1.3; "Courses / Lectures": mean 1.7, SD 1.0), and to people aged 80 years and older ("Internet": mean 1.3, SD 0.9; "Courses / Lectures": mean 1.4, SD 0.9). In comparison, the valuation of the information source "Family / Friends" grew with increasing age, from people younger than 70 years (mean 3.7, SD 1.2) and people between the ages of 70 and 79 years (mean 3.7, SD 1.3), to people aged 80 years and older (mean 4.0, SD 1.2). While age-specific differences in the information sources "Family / Friends" (η^2=0.01) and "Courses / Lectures" (η^2=0.04) showed small effect sizes, the effect size for the information source "Internet" (η^2=0.12) was medium.

Valuation of Health Information Sources by Gender

Concerning the gender-specific evaluation of information sources, the results show that apart from "Internet" and "Personal visit to a medical professional," women with cognitive impairment valued every explored source more than men with cognitive impairment on average (Multimedia Appendix 3). Two of the information sources showed significant gender-specific differences. As already described, men with cognitive impairment rated "Internet" 0.4 points higher on average than women with cognitive impairment (t_{685}=4.97; P<.001), with a small effect size (Cohen d=0.35). Female participants valued "Pharmacy" higher than male participants by 0.4 points on average (t_{772}=3.98; P<.001), with a small effect size (Cohen d=0.27). Another information source that was valued higher by women with cognitive impairment was the source "Family / Friends," with a mean difference of 0.2 (t_{756}=2.49; P=.013) and a small effect size (Cohen d=0.17).

Valuation of Health Information Sources by Cognition

Regarding the relevance of health information sources categorized by cognitive status, there were several significant differences between the MCI and dementia groups (Multimedia Appendix 4). People with MCI rated the sources "Internet," "Books / Brochures," "Courses / Lectures," "Newspapers / Journals," "Pharmacy," and "Personal visit to a medical professional" higher. Except for the sources "Books / Brochures" and "Newspapers / Journals," all of those group differences showed significant results. The source "Internet" showed the most considerable mean difference between both groups, with 0.6 points (t_{640}=7.52; P<.001) and a medium effect size (Cohen d=0.53). The only sources rated higher by the dementia group were "TV / Radio" and "Family / Friends," with none of them showing significant differences.

Discussion

Principal Results

To the best of our knowledge, this is the first study that describes the relevance of different sources of health information from the perspective of people with cognitive impairment.

Our analyses show that interpersonal contact with family and friends and that with medical professionals are the most important sources of health information for people with cognitive impairment. This interpersonal communication is followed closely by traditional media formats such as TV, radio, newspapers, and journals. People with cognitive impairment rated the internet and attending courses or lectures the lowest, with each being valued as less important or unimportant.

General Valuation of Health Information Sources

There are a number of possible reasons for the low relevance of the internet for people with cognitive impairment. The participants of our study mainly reported that they were generally not interested in the internet as a medium and that the

XSL•FO
RenderX

internet was too complicated for them. Other frequently cited reasons were that relatives take care of matters on the internet, no advantage is seen in using it, or traditional media are sufficient. These findings are partially in line with the literature, although it should be noted that the majority of the participants in our study had never used the internet before and therefore had no experience with the internet. Dixon et al [27] defined 4 main barriers that prevented people with cognitive impairment from seeking the internet for health information. These are the scarcity of relevant information and the inaccessibility, inaccuracy, and distrust of the information that was found [27]. More importantly, Dixon et al [28] see a chance of the internet becoming a more relevant source for people with cognitive impairment in the future, especially if their close relatives and health professionals support them in validating internet-based health information.

Our findings are in line with the large-scale survey by Weber et al [29] who assessed adults older than 60 years and reported high use of interpersonal communication with medical professionals, family, and friends as well as traditional mass media or brochures and infrequent use of the internet for seeking health information.

Stehr et al [30] came to similar results and showed that interpersonal contact and traditional mass media are the most used health information sources for people aged 65 years or older while they also confirmed that the internet is used little or not at all. Further, they noticed that the use and valuation of communication with medical professionals increases for people with chronic diseases, which include cognitive impairment [30].

The above results can be explained by, among other things, the trust that older adults or people with cognitive impairment place in the various sources. Baumann et al [31] reported that medical professionals, in this case doctors, are the most trusted health information source followed by family and friends, while the internet is one of the least trusted sources. These findings implicate that the trustiness of the various sources is a significant factor in the valuation by people with cognitive impairment.

Pertl et al [32] also stated that people with MCI show problems with understanding numerical health information. These problems are likely to increase for people with dementia, as they found a statistically significant correlation with global cognitive status [32]. These findings suggest that people with MCI and dementia might require explanation through communicating with medical professionals and their friends or family to understand their health issues.

Valuation of Health Information Sources by Age

The age-specific results show that the relevance of the information sources “Internet” and “Courses / Lectures” decreased significantly with age. Contrary to this, the evaluation of the source “Family / Friends” enlarged with increasing age. No significant differences were found for the 5 other sources based on participants’ age. The difference in the information source “Internet” is in particular noteworthy, with a medium effect size and high differences in the mean values between the 3 age groups.

Our age-specific findings regarding the internet align with the results of the “D80+ report” by Reissmann et al [33], which found low internet use among people older than 80 years in Germany, at 37%. Weber et al [29] also found that age is a decisive factor influencing internet use as the tendency of infrequent use of the internet for seeking health information became more pronounced with increasing age.

Reasons for these observations may include the often limited digital health literacy or generally limited digital media literacy and internet skills of the older population [34]. These limitations become even more pronounced with increasing age, especially for people with cognitive impairment [34,35]. The high average age of 80.5 years in this sample could therefore also explain the generally low rating of the internet as a source of health information.

Valuation of Health Information Sources by Gender

Our gender-specific analyses show that women and men with cognitive impairment differ in assessing the relevance of health information sources. Except for the internet, all our evaluated sources were more important to women with cognitive impairment, though these differences between men and women were small. The sources “Pharmacy” and “Family / Friends” showed the most significant differences.

These gender-specific findings are in line with current scientific research. For the general population, Manierre [15] discovered that women seek health information more often than men. In the group of adults aged 65 years and older, Stehr et al [30] found higher use of most health information sources by women. They also noted the most remarkable differences between the sources “Pharmacy” and “Internet,” with the former being much more used by women and the latter by men [30]. The same applies to the results of the KomPaS study by Horch et al [36], who found similar differences in the use of these 2 sources.

Valuation of Health Information Sources by Cognition

When categorized by cognitive status, our results show that people with MCI rate most of the health information sources evaluated more highly than people with dementia. People with dementia only value contact with family and friends as well as watching TV or listening to the radio more.

It is noticeable that the mean values for health information sources that require some activity to obtain information, for example, leaving home to see a doctor, are all higher for the group of people with MCI. The same applies to the mean for sources of information that can be used passively, such as watching television or receiving advice from a friend, which is higher for people with dementia. This could be explained by the progressive social withdrawal of people with cognitive impairment as the medical condition progresses [37].

Dixon et al [28] also found similar results with their research on the changing strategies of people with cognitive impairment seeking health information. They reported that most people with cognitive impairment apply an active search technique after receiving their diagnosis and then transition to a monitoring strategy for ongoing seeking of health information. With the progression of their medical condition, they transition to a proxy

XSL•FO
RenderX

information search, as they have problems accessing the health information needed when just actively searching by themselves. With further progression, the authors even describe a strategy of information avoidance and selective exposure to health information by people with cognitive impairment [28]. These changes in information behavior could explain the distribution of different sources in the 2 cognitive groups. While people with MCI are likely to use one of the first, more active search strategies described, people with dementia are more likely to use one of the latter strategies due to their disease progression.

Limitations

This study has limitations. First, our sample was a convenience sample with a specific target group of people with cognitive impairment in Bavaria. We cannot exclude that we may have a selection bias as a result of the different referral patterns in the different regions of Bavaria. Although our research partners were institutions specialized in managing and caring for people with cognitive impairment in all of the 7 administrative regions in Bavaria, this study is not representative of the total population. This limits the comparability and transferability of the results to the entirety of people with cognitive impairment.

Furthermore, the results of the group differences mostly show small effect sizes, except for the group differences between people with MCI and dementia and the 3 age groups for the information source “Internet.” This limits the power and generalizability of the results while highlighting the need for further research with even larger case numbers.

Conclusions

To enhance health and dementia literacy in people with cognitive impairment, our results suggest that interpersonal contact with medical professionals and their friends and family is the best way to communicate health information to people with cognitive impairment. Another way is to disseminate information through traditional mass media. As the disease progresses, the importance of family and friends as sources of information increases, while the relevance of other sources decreases. Women with cognitive impairment are generally easier to reach and more interested in health information. At the same time, men and younger people with cognitive impairment also use the internet to some extent and are open to receiving information via this medium.

While the internet was of little importance as a source of health information to most of our study population of people with cognitive impairment older than 70 years, this may change in the future, as recent surveys from the European Union and the United States indicate [11,12]. Due to the continuing improvement in digital literacy in older adults, the internet might become more valuable as a source of health information. This implicates a need for further research in the future to capture these potential changes.

Acknowledgments

The project "Digital Dementia Registry Bavaria –digiDEM Bayern" is funded by the Bavarian State Ministry of Health, Care and Prevention as part of the funding initiative "BAYERN DIGITAL II" (grant number: G42d-G8300-2017/1606-83). The present work was performed by FW in (partial) fulfillment of the requirements for obtaining the degree “Dr. rer. biol. hum.” at the Medical Faculty of the Friedrich-Alexander-Universität Erlangen-Nürnberg. No generative artificial intelligence tools were used in the preparation and writing of the manuscript.

Data Availability

The datasets used and analyzed during this study are not publicly available since this is an ongoing research project and data will be used in the course of future analysis. Deidentified or anonymized data are available from the Bavarian State Ministry of Health, Care and Prevention upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Frequency of reasons given for not using the internet by people with cognitive impairment in a registry-based cohort study in Germany from 2020 to 2023.
[PNG File, 35 KB - formative_v9i1e54460_app1.png]

Multimedia Appendix 2
Evaluation of health information sources by people with cognitive impairment, with differentiation by age, using Likert scales ranging from unimportant (1) to very important (5) in a registry-based cohort study in Germany from 2020 to 2023.
[PNG File, 83 KB - formative_v9i1e54460_app2.png]

Multimedia Appendix 3
Evaluation of health information sources by people with cognitive impairment, with differentiation by gender, using Likert scales ranging from unimportant (1) to very important (5) in a registry-based cohort study in Germany from 2020 to 2023.

XSL•FO
RenderX

[PNG File, 74 KB - formative_v9i1e54460_app3.png]

Multimedia Appendix 4
Evaluation of health information sources by people with cognitive impairment, with differentiation by their cognitive status, using Likert scales ranging from unimportant (1) to very important (5) in a registry-based cohort study in Germany from 2020 to 2023.
[PNG File, 75 KB - formative_v9i1e54460_app4.png]

References

1. Gauthier S, Rosa-Neto P, Morais JA, Webster C. World Alzheimer Report 2021: Journey Through the Diagnosis of Dementia: Alzheimer's Disease International; 2021.
2. Nichols E, Steinmetz JD, Vollset SE, et al. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the global burden of disease study 2019. Lancet Public Health 2022 Feb;7(2):e105-e125. [doi: 10.1016/S2468-2667(21)00249-8]
3. Wimo A, Seeher K, Cataldi R, et al. The worldwide costs of dementia in 2019. Alzheimers Dement 2023 Jul;19(7):2865-2873. [doi: 10.1002/alz.12901] [Medline: 36617519]
4. Aihara Y, Maeda K. Dementia literacy and willingness to dementia screening. Int J Environ Res Public Health 2020 Nov 4;17(21):33158014. [doi: 10.3390/ijerph17218134] [Medline: 33158014]
5. The HLS19 Consortium of the WHO Action Network M-POHL. International report on the methodology, results, and recommendations of the european health literacy population survey 2019-2021 (HLS19) of M-POHL. : Austrian National Public Health Institute: Vienna; 2021.
6. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Promot Int 2000 Sep 1;15(3):259-267. [doi: 10.1093/heapro/15.3.259]
7. Low LF, Anstey KJ. Dementia literacy: recognition and beliefs on dementia of the Australian public. Alzheimers Dement 2009 Jan;5(1):43-49. [doi: 10.1016/j.jalz.2008.03.011] [Medline: 19118808]
8. Sørensen K, Van den Broucke S, Fullam J, et al. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80. [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
9. Chen X, Hay JL, Waters EA, et al. Health literacy and use and trust in health information. J Health Commun 2018;23(8):724-734. [doi: 10.1080/10810730.2018.1511658] [Medline: 30160641]
10. Zimmerman MS, Shaw G Jr. Health information seeking behaviour: a concept analysis. Health Info Libr J 2020 Sep;37(3):173-191. [doi: 10.1111/hir.12287] [Medline: 32052549]
11. EU citizens: over half seek health information online. Eurostat. 2022. URL: https://ec.europa.eu/eurostat/en/web/products-eurostat-news/-/edn-20220406-1 [accessed 2023-08-22]
12. The most recent time you looked for information about health or medical topics, where did you go first?- HINTS 5 cycle 3. National Cancer Institute Health Information National Trends Survey. 2019. URL: https://hints.cancer.gov/view-questions/question-detail.aspx?PK_Cycle=12&qid=688 [accessed 2023-08-22]
13. Rossmann C, Lampert C, Stehr P, Grimm M. Nutzung Und Verbreitung von Gesundheitsinformationen: Ein Literaturüberblick Zu Theoretischen Ansätzen Und Empirischen Befunden. Use and Dissemination of Health Information: A Literature Review of Theoretical Approaches and Empirical Findings: Bertelsmann Stiftung; 2018. [doi: 10.11586/2017051]
14. Chasiotis A, Wedderhoff O, Rosman T, Mayer AK. Why do we want health information? the goals associated with health information seeking (GAINS) questionnaire. Psychol Health 2020 Mar;35(3):255-274. [doi: 10.1080/08870446.2019.1644336] [Medline: 31340675]
15. Manierre MJ. Gaps in knowledge: tracking and explaining gender differences in health information seeking. Soc Sci Med 2015 Mar;128(151-8):151-158. [doi: 10.1016/j.socscimed.2015.01.028] [Medline: 25618604]
16. Oh YS, Cho Y. Examining the relationships between resources and online health information seeking among patients with chronic diseases and healthy people. Soc Work Health Care 2015;54(2):83-100. [doi: 10.1080/00981389.2014.987940] [Medline: 25674723]
17. Dietzel N, Kürten L, Karrer L, et al. Digital Dementia Registry Bavaria-digiDEM Bayern: study protocol for a multicentre, prospective, longitudinal register study. BMJ Open 2021 Feb 8;11(2):e043473. [doi: 10.1136/bmjopen-2020-043473] [Medline: 33558357]
18. Reichold M, Dietzel N, Chmelirsch C, Kolominsky-Rabas PL, Graessel E, Prokosch HU. Designing and implementing an IT architecture for a digital multicenter dementia registry: digiDEM Bayern. Appl Clin Inform 2021 May;12(3):551-563. [doi: 10.1055/s-0041-1731286] [Medline: 34134149]
19. Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to alzheimer's disease: recommendations from the National Institute on Aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. Alzheimers Dement 2011 May;7(3):270-279. [doi: 10.1016/j.jalz.2011.03.008] [Medline: 21514249]
20. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999 Mar;56(3):303-308. [doi: 10.1001/archneur.56.3.303] [Medline: 10190820]

XSL•FO
RenderX

21. Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”. a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
22. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
23. Gräßel E, Bleich S, Meyer-Wegener K, Schmid U, Kornhuber J, Prokosch HU. Das Internet als Informationsquelle für pflegende Angehörige eines Demenzpatienten. The internet as an information source for informal caregivers of dementia patients. Psychiat Prax 2009 Apr 3;36:115-118. [doi: 10.1055/s-2008-1067550]
24. Lausen B, Potapov S, Prokosch HU. Health-related use of the internet in Germany 2007. GMS Med Inform Biometr Epidemiol 2008 Sep;4(2).
25. Armstrong RA. When to use the Bonferroni correction. Ophthalmic Physiol Opt 2014 Sep;34(5):502-508. [doi: 10.1111/opo.12131] [Medline: 24697967]
26. Cohen J. Statistical power analysis. Curr Dir Psychol Sci 1992 Jun;1(3):98-101. [doi: 10.1111/1467-8721.ep10768783]
27. Dixon E, Anderson J, Blackwelder D, Radnofsky M, Lazar A. Barriers to online dementia information and mitigation. Proc SIGCHI Conf Hum Factor Comput Syst 2022 Apr;2022:513. [doi: 10.1145/3491102.3517554] [Medline: 35789135]
28. Dixon E, Anderson J, Blackwelder DC, Radnofsky ML, Lazar A. The human need for equilibrium: qualitative study on the ingenuity, technical competency, and changing strategies of people with dementia seeking health information. J Med Internet Res 2022 Aug 11;24(8):e35072. [doi: 10.2196/35072] [Medline: 35969426]
29. Weber W, Reinhardt A, Rossmann C. Lifestyle segmentation to explain the online health information-seeking behavior of older adults: representative telephone survey. J Med Internet Res 2020 Jun 12;22(6):e15099. [doi: 10.2196/15099] [Medline: 32530433]
30. Stehr P, Weber W, Rossmann C. Gesundheitsinformationsverhalten 65+: Erreichbarkeit älterer Zielgruppen. Health seeking behavior 65+: reaching older target groups. Bundesgesundheitsbl 2021 Jan;64(1):37-44. [doi: 10.1007/s00103-020-03253-1]
31. Baumann E, Czerwinski F, Rosset M, Seelig M, Suhr R. Wie informieren sich die Menschen in Deutschland zum Thema Gesundheit? erkenntnisse aus der ersten Welle von HINTS Germany. How do people in Germany seek health information? insights from the first wave of HINTS Germany. Bundesgesundheitsbl 2020 Sep;63(9):1151-1160. [doi: 10.1007/s00103-020-03192-x]
32. Pertl MT, Benke T, Zamarian L, et al. Do patients with mild cognitive impairment understand numerical health information? J Alzheimers Dis 2014;40(3):531-540. [doi: 10.3233/JAD-131895] [Medline: 24473188]
33. Reissmann M, Oswald V, Zank S, Tesch-Römer C, Cologne Center for Ethics, Rights, Economics, and Social Sciences of Health (ceres). Digitale Teilhabe in der Hochaltrigkeit (D80+ Kurzberichte, 6). Digital participation in very old age (D80+ abridged reports, 6). Social Science Open Access Repository. 2022. URL: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-78429-7 [accessed 2024-10-07]
34. Hargittai E, Piper AM, Morris MR. From internet access to internet skills: digital inequality among older adults. Univ Access Inf Soc 2019 Nov;18(4):881-890. [doi: 10.1007/s10209-018-0617-5]
35. Talbot CV, Briggs P. The use of digital technologies by people with mild-to-moderate dementia during the COVID-19 pandemic: a positive technology perspective. Dementia (London) 2022 May;21(4):1363-1380. [doi: 10.1177/14713012221079477] [Medline: 35333111]
36. Horch K, Jordan S. Sachbericht zur studie "Kommunikation und Information im Gesundheitswesen aus Sicht der Bevölkerung. patientensicherheit und informierte Entscheidung" (KomPaS). Report on the study "Communication and information in the healthcare system from the perspective of the population. patient safety and informed decision making" (KomPaS). Bundesministerium für Gesundheit. 2019. URL: https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/5_Publikationen/Praevention/abschlussbericht/2019-08-01_Sachbericht_KomPaS.pdf [accessed 2023-08-22]
37. Porcelli S, Van Der Wee N, van der Werff S, et al. Social brain, social dysfunction and social withdrawal. Neurosci Biobehav Rev 2019 Feb;97:10-33. [doi: 10.1016/j.neubiorev.2018.09.012] [Medline: 30244163]

Abbreviations

MCI: mild cognitive impairment
MMSE: Mini-Mental State Examination
MoCA: Montreal Cognitive Assessment
RQ: research question

XSL•FO
RenderX

Edited by A Mavragani; submitted 10.11.23; peer-reviewed by A Reinhardt, N Döring, T Brkovic; revised version received 07.10.24; accepted 15.10.24; published 15.01.25.

Please cite as:
Weidinger F, Dietzel N, Graessel E, Prokosch HU, Kolominsky-Rabas P
Using Health Information Resources for People With Cognitive Impairment (digiDEM Bayern): Registry-Based Cohort Study
JMIR Form Res 2025;9:e54460
URL: https://formative.jmir.org/2025/1/e54460
doi:10.2196/54460

© Florian Weidinger, Nikolas Dietzel, Elmar Graessel, Hans-Ulrich Prokosch, Peter Kolominsky-Rabas. Originally published in JMIR Formative Research (https://formative.jmir.org), 15.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile Phone App to Promote Lifestyle Change in People at Risk of Type 2 Diabetes: Feasibility 3-Arm Randomized Controlled Trial

Gyri Skoglund[1], MSc; Gunvor Hilde[1], PhD; Pernille Lunde[1], PhD; Venessa Vera Cruz Naceno[1], PT; Cecilie Fromholt Olsen[2], PhD; Birgitta Blakstad Nilsson[3], PhD

[1]Department of Rehabilitation Science and Health Technology, Faculty of Health Sciences, Oslo Metropolitan University, PB 4, St. Olavs plass, Oslo, Norway

[2]Section for Allied Health Professionals, Department of Neurology, Oslo University Hospital, Oslo, Norway

[3]Section for Physiotherapy, Division of Medicine, Oslo University Hospital, Oslo, Norway

Corresponding Author:
Gyri Skoglund, MSc
Department of Rehabilitation Science and Health Technology, Faculty of Health Sciences, Oslo Metropolitan University, PB 4, St. Olavs plass, Oslo, Norway

Abstract

Background: The use of mobile health interventions, such as apps, are proposed to meet the challenges faced by preventive health care services due to the increasing prevalence of type 2 diabetes (T2D). Thus, we developed and conducted initial feasibility testing of the Plunde app for promoting and monitoring individual goals related to lifestyle change for people at risk of T2D.

Objective: The primary aim of this study was to assess the feasibility of an app for promoting lifestyle change in people at risk of T2D. The secondary aim was to assess recruitment rate, resource requirements, and change in potential outcomes for a full scale randomized controlled trial (RCT) study .

Methods: A 3-arm feasibility RCT lasting 12 weeks was designed. Participants were recruited from 9 general practitioners in Norway. Eligible participants were randomized to either (1) app follow-up; (2) app follow-up and referral to care as usual in Healthy Life Centers; or (3) referral to care as usual in a Healthy Life Center, only. The primary outcome was feasibility and was measured by app adherence (actual usage of the app), the System Usability Scale, and app motivation score gained from a questionnaire designed for this study. Criteria for success were preset based on these measures. Secondary outcomes included recruitment rate, resource requirements, and potential primary outcomes of a full-scale RCT. This included change in body weight, waist circumference, and self-evaluated functional health status, assessed with the Dartmouth Primary Care Cooperative Research Network/World Organization of Family Doctors (COOP/WONCA) functional health assessment chart.

Results: Within 8 months, 9 general practitioners recruited a total of 54 participants, of which 45 were eligble for participation in the study. Mean age was 61 (SD 13) years and 53% (n=24) were female. App adherence was 86%, the mean System Usability Scale score was 87.3 (SD 11.9), and the mean app motivation score was 74.8 (SD 30.3). Throughout the intervention period, health care professionals spent on average 3.0 (SD 1.0) minutes per participant per week providing follow-up. Statistically significant reduction in body weight and waist circumference was shown in group 1 and 3.

Conclusions: Based on the preset criteria for success, the Plunde app is feasible in providing support for lifestyle change. The Plunde app had excellent user satisfaction. The amount of time spent on monitoring and promoting lifestyle change through the app was low; however, the recruitment was slow. Results from this study will guide the development of further research within this field.

Trial Registration: ClinicalTrials.gov NCT06117098; https://clinicaltrials.gov/study/NCT06117098

(JMIR Form Res 2025;9:e63737) doi:10.2196/63737

KEYWORDS

prevention; risk of type 2 diabetes; lifestyle change; feasibility; app adherence; mobile health; mHealth; mobile phone

Introduction

There is an urgent need to prevent and combat the global epidemic of type 2 diabetes (T2D). The number of adults with diabetes is expected to reach 67 million by 2030 and 69 million by 2045 [1]. Prediabetes, an intermediate stage of glucose dysregulation that may precede T2D, affected approximately 720 million individuals worldwide in 2021 and is estimated to

XSL•FO
RenderX

affect estimated 1 billion people by 2045 [1,2]. Prediabetes is diagnosed when blood glucose levels are above the normal range but not high enough to be classified as diabetes [2].

Healthy diet, regular physical activity, and maintaining a normal body weight are established cornerstones of prevention or delayed onset of T2D [3-5]. Hence, lifestyle modification is the preferred initial approach when having confirmed prediabetes or being at high risk of T2D [2]. Siginficant risk reduction in delaying or preventing progression into T2D has been found in people receiving lifestyle interventions [3,6]. A post-hoc analysis of randomized, controlled, multicentre Prediabetes Lifestyle Intervention Study showed that a weight loss of 5 percent or more, could decrease the risk of T2D with 72%, and that full remission of T2D should be the target aim [7]. A larger waist circumference (WC), independent of overall adiposity, has been found to be strongly and linearly associated with a higher risk of T2D [8]. Consequently, both body weight and WC may be important predictors of the development for T2D.

Healthy Life Centers (HLCs) is a part of the public health care service in the municipalities in Norway, providing in person lifestyle modification programs for people with, or in high risk of disease [9]. Considering the burden of noncommunicable diseases (NCDs), together with an aging population, the health care system faces fundamental challenges in delivering optimal care [10]. Thus, the primary health care services are at a threshold limit.

The use of mobile health (mHealth) technologies has been proposed to meet challenges related to delivery of sustainable prevention strategies [11]. The beneficial effects of mHealth interventions regarding lifestyle factors have been demonstrated in patients with NCDs in several systematic reviews [12-14]. However, few studies have examined the effect of using an app to deliver support and guidance for risk reduction in patients at risk of T2D.

In 2022, we investigated barriers and facilitators for lifestyle change in people at risk of T2D in a qualitative meta-synthesis [15]. We found that central barriers were time, cost, and availability [15]. Additionally individual tailoring of lifestyle interventions was highlighted to facilitate the process of making healthy changes [15]. The advantages of digital health solutions lie, in its potential to overcome the barriers for lifestyle change and at the same time enabling personalized, empowering, and person-centered health interventions [16]. Therefore, based on former research, user orientation- and involvement, we developed the Plunde app (People Living UNDEr change) to monitor and guide individuals in initiating and maintaining lifestyle change and healthy behavior [17].

Randomized controlled trials (RCTs) are needed to assess potential effect of lifestyle interventions delivered by an app. Before commencing such a trial, investigating the feasibility of using an app to promote lifestyle change is warranted. Therefore, the primary aim of this study was to assess if the Plunde app was feasible to use for promoting lifestyle change in people at risk of T2D. Preset criteria for success was based on app adherence (actual usage of the app), the System Usability Scale (SUS) and app motivation score which was gained from a questionnaire designed for this feasibility study. Secondary, we aimed to assess recruitment rate, resource requirements and change in body weight, WC, and self-evaluated functional health status.

Methods

Study Design

This 3-arm feasibility RCT, had an intervention period lasting for 12 weeks. The participants were randomized to receive: lifestyle intervention delivered by the Plunde app (group 1) or referral to group-based lifestyle intervention at the local HLCs in addition to follow-up with the Plunde app (group 2) or to usual care which was referral to group-based lifestyle intervention at the local HLCs (group 3). A computer-generated block randomization scheme was used to allocate the participants to 1 of the 3 groups via concealed envelopes after baseline assessment. The block sizes varied by 3 and 6.

Reporting follows the Consilidated Standard of Reporting Trials (CONSORT) 2010 statement: extension to randomized pilot and feasibility trials [18]. Additionally, principles from the Consilidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth (CONSORT-eHEALTH) statement [19] were applied (Checklist 1).

Setting and Recruitment

Participants were recruited from general practitioner (GP) clinics in the eastern part of Norway from May 2023 and throughout December 2023. Initially the patients were recruited from 2 GP clinics. From September 2023 we included additionally 3 GPs clinics. Prior to recruitment, information meetings at the clinic sites were held. Eligible participants were asked by their GP regarding willingness to participate in the study, and with their consent they were contacted by a researcher (BBN or VVCN). The baseline assessments and follow-ups were conducted at 3 collaborating HLCs by 2 of the project members (BBN and VVCN).

Participants

Eligible participants were women and men aged 18 years or older, who were assessed by their GP to be at high risk of developing T2D. The criteria for being at high risk was HbA_{1c} within the threshold level of 42-48 mL/mmol. In addition, participants had to own and use an Android or iOS smartphone and be able to read and understand Norwegian or English.

By September 2023 only 25 participants were included; therefore, we invited 3 more GP clinics to recruit participants to the project. To increase the recruitment rate, we also widened the criteria for being at high risk of T2D. Hypertension, hypercholesteremia, and being obese were set as criteria in addition to, or instead of, HbA_{1c} levels between 42 mL/mmol and 48 mL/mmol.

Baseline Assessment

At baseline we collected demographic and descriptive data within different lifestyle domains from the standardized “HLC Startup” questionnaire developed by the Norwegian Health Directorate [20]. The demographic data collected included

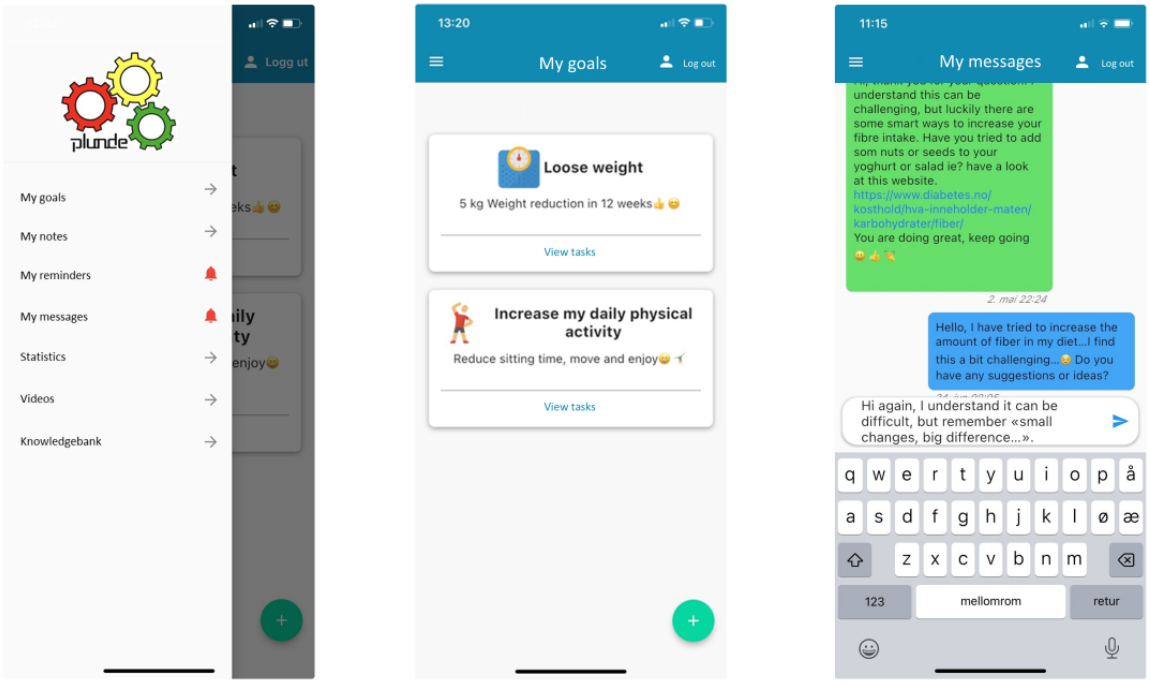
XSL•FO
RenderX

gender, age, educational level, and occupational status. The descriptive data included physical activity level, dietary habits, tobacco, alcohol, sleep habit, and social support. This questionnaire was primarily used to guide and structure the plan for lifestyle change and setting of goals and tasks. Anthropometric data (body weight, WC, and height) were collected at baseline and at follow-up by the same researcher.

The Plunde App

Participants randomized to the app intervention groups, group 1 and 2, received access to Plunde and guiding in how to use it after baseline assessment. Plunde was developed to promote initiation of behavioral change and adherence to healthy behavior [17]. The app is highly individualized as it is based on the patient's own goals to initiate or maintain a lifestyle change [17]. Plunde permits the user to set personal goals (Figure 1B) with tasks and accompanying reminders. Goals and tasks decided during the baseline consultation were added to the app. The goals set by the participants were related to the different areas of lifestyle they wanted to work on during the intervention period. The main lifestyle domains presented (through the questionnaire) were dietary habits, physical activity level and exercise, sleep, mental well-being, alcohol, and smoking habits. Each participant set different tasks to their goals, and 1 goal could have 1 or more related tasks. To each of the tasks, the participant set a reminder and decided when and how often the reminders of their tasks should appear. Additionally, they could choose the wording of their reminders. Plunde provided push notifications related to the individual tasks, and the participants replied "completed" or "not completed."

Figure 1. Screenshots showing the user interface.



Other features and functions in Plunde include a note function, messages, videos, and a knowledge bank (Figure 1A). The note function allowed the participant to write relevant notes or comments on their goal achievement process, or for example writing an exercise diary. The message function allowed the participants to chat with their supervisor during the follow-up period (Figure 1C). Relevant videos based on the participants' goals were added under the video function. In the knowledge bank, the supervisor could add written educational content based on the participants' goals and requests.

A supervisor had access to an administrator interface and monitored the goals, tasks, and notes of each participant throughout the intervention period. During the follow-up period the supervisor could send weekly encouraging messages to support their process of lifestyle change. If the participants asked any questions in the message function, these were answered within 2 working days. The 2 supervisors in this study were both physiotherapists, 1 being a researcher with comprehensive experience in cardiac rehabilitation and cardiac research and the other with experience from HLCs and community health services. The same supervisor who conducted the baseline consultation monitored and provided feedback throughout the study period.

Referral to Healthy Lifestyle Center

Participants randomized to group 2 and 3 were referred to the HLC by the respective supervisor immediately after the baseline consultation, with an expected enrollment to the HLC program within 2 weeks. The lifestyle modification program at the HLC included individual and group-based counseling and courses for increased physical activity, healthy diet, improved sleep, and tobacco cessation. The participants were encouraged to attain at least 1 or 2 exercise sessions per week at the HLC. The participants in all the 3 study arms, received counseling based on elements from motivational interviewing, which is a client-centered method of intervention focused on enhancing

XSL•FO
RenderX

intrinsic motivation and behavioral change [21]. Goal setting and planning were crucial parts of this conversation, based on the needs, values, and perspectives of the participants.

Primary Outcome

The primary outcome was feasibility measured as app adherence, user satisfaction, and app motivation. Data related to feasibility of Plunde were collected from patients in groups 1 and 2.

Adherence to the app was registered in terms of actual use, which we defined to be the percentage of reminders answered per week throughout the 12-week intervention period. These data were collected from the administrator interface.

The user satisfaction with the app was assessed with the SUS. The SUS is a technology independent, 10-item questionnaire with a score between 0 and 100, where 0 represents low usability and 100 represent high usability [22]. The elements of the SUS scale cover aspects of usability such as ease of use, integration of functions, learning curve, satisfaction, user confidence, and complexity [22,23].

App motivation was measured with a questionnaire designed for this study. This questionnaire consisted of 17 questions—13 questions with Likert scale (0 to 100) options, 3 multiple-choice questions, and 1 open-ended question (Multimedia Appendix 1). The main questions defining app motivation score was based on 2 of the questions; to what extent Plunde was experienced as motivating and if the individual feedback provided from the supervisor was experienced as motivating. In addition, we asked for how long the participants would have liked to continue using the app and, in this period, how often they would like to receive feedback from their supervisors. The criteria for feasibility success of the app that were preset were (1) at least 80% of the participants reminders were answered per week (app adherence), (2) mean SUS score ≥65, and (3) app motivation score ≥75.

Secondary Outcomes

To inform a potential full-scale RCT, data on recruitment rate and resource requirements were collected throughout the study period. We calculated the recruitment rate as the number of participants successfully recruited for the study divided by the total amount of time required for recruitment. Resource requirements in groups having the app intervention was assessed by the supervisors who logged all time spent on monitoring and providing feedback to participants. Additionally, potential primary outcomes in a future scale RCT were evaluated. This included body weight, WC, and self-evaluated functional health status. These outcomes were assessed to get experience with the measurements and to examine a potential change.

Body weight was measured in kilogram using a digital scale (Body Composition Analyzer, BC-418, ADE MeWa GMbh, Schwerin and Beurer GS400 Signature Line weight). The evaluation was carried out in the “standard mode,” introducing the participant’s age, sex, and height. We strived to use the same equipment for measuring body weight at baseline and at follow-up. WC was measured with a measuring tape (in cm; Seca 201) at the end of several consecutive natural breaths, midpoint being between the top of the iliac crest and the lower margin of the last palpable rib in the mid axillary line [24]. The cut off points for abdominal obesity and increased risk of T2D is WC ≥80 in women and WC ≥90 in men [25,26].

Self-evaluated functional health status was assessed with the Dartmouth Primary Care Cooperative Research Network/World Organization of Family Doctors (COOP/WONCA) chart [27]. The questionnaire is a generic questionnaire comprising 6 domains—physical fitness, feelings (mental well-being), daily activities, social activities, change in health, and overall health. Each dimension is illustrated pictorially, numerically, and in writing inquiring about the patient’s status during the past 2 weeks. The response categories were scored from 1 to 5, where higher score indicate worse health status.

Statistical Analysis

Data were analysed using SPSS Statistics for Windows (version 29.0.0; IBM Corp). Descriptive statistics were used to assess feasibility measures. Descriptive statistics are reported in mean and SD or median and range for continuous variables and in number and percentages for categorical variables. To examine a potential change throughout the intervention period for body weight, WC, and self-evaluated functional health status, we performed within-group analyses on mean change from baseline to 12-weeks follow-up. If the data qualified for a normal distribution, the paired sample *t* test was used. If not normally distributed, the paired Wilcoxon signed rank test was used. All statistical tests were 2-sided and *P* values <.05 were considered statistically significant. To ensure a robust sample for the purpose of this feasibility study, we aimed to include a total of 60 participants [18].

Ethical Considerations

The Regional Committee for Medical Research Ethics in Southeast Norway did not find approval to be required. The study is approved by The Norwegian Data Protection Service for Research and the study has been registered on ClinicalTrials.gov (NCT06117098). The participants provided written informed consent before participation.

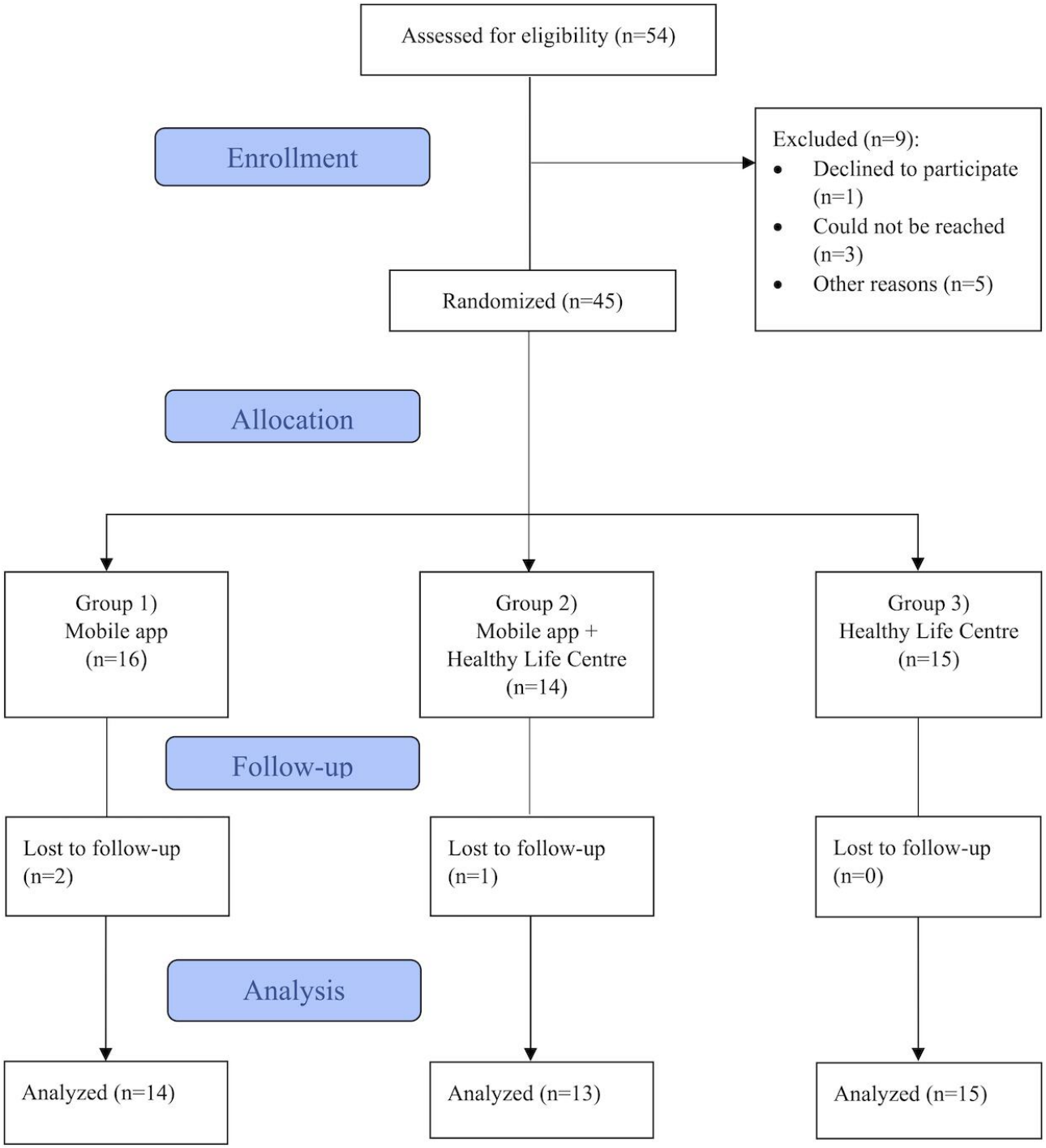
Results

Recruitment

From May 2023 and throughout December 2023, 9 GPs recruited a total of 54 participants of which 45 (82%) were eligible and included in this study. As shown in Figure 2, three participants dropped out during the intervention period. No reasons for drop out were provided.

XSL•FO
RenderX

Figure 2. Recruitment and participant flow.



Participant Characteristics

Baseline characteristics are presented in Table 1. In total, 53% (n=24) were female and the mean age was 61 (SD 13) years with a range from 31 to 83 years. A total of 49% (n=22) of the participants had BMI ≥30, 88% (n=39) of the women had WC ≥80 cm, and 95% (n=43) of the men had WC ≥90 cm. The goals of the participants pertained to the 5 lifestyle domains of weight reduction, reduced HbA_{1c}, healthy diet, physical activity, and sleep.

XSL•FO
RenderX

Table . Baseline characteristics.

Characteristics	Study sample (N=45)	App (group 1; n=16)	App + Healthy Life Center (group 2; n=14)	Healthy Life Center (group 3; n=15)
Female, n (%)	24 (53)	6 (38)	9 (38)	9 (38)
Age (years), mean (SD)	60.8 (12.6)	54.5 (12.4)	59.1(12.7)	68.1 (9.1)
Body weight (kg), mean (SD)	91.0 (22.3)	96.6 (19.6)	84.6 (24.3)	89.7 (22.8)
BMI, mean (SD)	31.0 (6.6)	32.3 (5.4)	30.0 (7.0)	30.0 (7.0)
Waist circumference (cm), mean (SD)	105.2 (14.9)	108.5 (12.1)	99.0 (16.0)	106.7 (15.4)
Educational level, n (%)				
Elementary or high school	20 (46)	7 (44)	7 (50)	7 (47)
University or college >3 years	9 (20)	3 (19)	2 (14)	4 (27)
University or college ≤3 years	14 (33)	5 (31)	5 (36)	4 (27)
Employment, n (%)				
Full-time work or part-time work	19 (43)	9 (56)	8 (57)	4 (27)
Disability benefits	8 (18)	5 (31)	0 (0)	3 (20)
Retired	17 (39)	2 (13)	6 (43)	8 (53)
COOP/WONCA[a] score, mean (SD)				
Physical fitness	2.3 (1.0)	2.0 (0.9)	2.2 (1.1)	2.8 (1.0)
Feelings	1.8 (1.2)	1.8 (1.1)	1.8 (1.1)	1.8 (1.3)
Daily activities	1.5 (0.9)	1.4 (0.7)	1.5 (0.9)	1.7 (1.1)
Social activities	1.4 (0.8)	1.4 (0.8)	1.1 (0.3)	1.7 (1.0)
Change in health	2.6 (0.7)	2.5 (0.6)	2.6 (0.6)	2.8 (0.7)
Overall health	2.4 (0.8)	2.5 (0.7)	2.2 (0.8)	2.5 (0.7)
Total score, mean (SD)	11.8 (3.2)	11.6 (2.8)	10.6 (3.3)	13.1 (3.0)

[a]COOP/WONCA: Dartmouth Primary Care Cooperative Research Network/World Organization of Family Doctors functional health assessment chart.

Participants randomized to Plunde (groups 1 and 2; n=27) had a mean of 1.5 (SD 0.7) goals, 1 - 3 tasks related to each goal and a mean of 7.7 (SD 7.1) weekly reminders of their goal related tasks. The most frequent goal was related to weight reduction and reduced HbA_{1c}. Further, 17 (61%) participants randomized to Plunde had iOS smartphone and 11 (39%) had Android smartphone. For 2 (7%) participants, data regarding the operative system were not collected. No technical problems or bugs was reported during the intervention period.

Primary Outcome

At the 12 weeks follow-up, user log data could be retrieved from 27 participants. The overall app adherence score was 86% (n=27) for the whole intervention period. The mean SUS score was 86.6 (SD 12.0), indicating a grade A and excellent usability (n=22).

The mean app motivation score was 74.8 (SD 30.3). The motivational messaging from the supervisor was reported (n=26) to be highly motivating, with a mean score of 93.5 (SD 19.8).

When asked for how long the participants would prefer to be followed up through the Plunde app, the most frequent answer was 12 months. On the question on how often the participants would find it useful to get feedback from their supervisor, every other week was the most frequent answer.

Secondary Outcomes

From May 2023 and throughout December 2023, 54 participants were recruited from 9 GP clinics within 8 months, of these 45 participants were eligible for the study. The recruitment rate was 1.4 participants per week.

The first consultation with the participants in group 1 and 2 (n=27) lasted in average for 91.3 (SD 26.5) minutes. Monitoring and follow-up through the app during the study period, took an average of 3.0 (SD 1.0) minutes per participant per week.

As presented in Table 2, preliminary pre-post interventions findings showed statistically significant within-group changes in 1 or more outcomes in all 3 groups. A significant reduction in body weight and WC was shown in group 1 and 3, and

XSL•FO
RenderX

improved changes in self-evaluated health was shown in all 3 groups. In the data analysis related to weight reduction, there was 1 data point missing from group 1. For WC, 1 data point was missing from both group 1 and 2. Meanwhile, for the COOP/WONCA analysis, group 3 had 1 missing data point.

Table . Changes in body weight, waist circumference, and self-evaluated functional health status at baseline (N=45) and after 12 weeks follow-up (n=42).

Outcomes and measures	Baseline, mean (SD)	12 weeks, mean (SD)	Mean difference (95% CI)	*P* value
App (group 1; n=14)				
Body weight (kg)	96.7 (20.1)	93.8 (18.5)	2.8 (0.8 to 4.9)	.011
Waist circumference (cm)	108.9 (12.4)	104.3 (11.3)	4.6 (2.0 to 7.1)	.002
COOP/WONCA[a] total score	11.9 (2.8)	9.4 (1.9)	2.5 (1.0 to 4.0)	.003
Physical fitness	2.1 (0.9)	1.8 (0.7)	0.4 (−0.2 to 0.9)	.174
Feelings	1.9 (1.1)	1.5 (0.9)	0.4 (−0.2 to 1.0)	.139
Daily activities	1.4 (0.8)	1.3 (0.7)	0.1 (−0.4 to 0.6)	.547
Social activities	1.4 (0.9)	1.0 (0.0)	0.4 (−0.6 to 0.9)	.082
Change in health	2.5 (0.7)	1.6 (0.6)	0.9 (0.5 to 1.2)	<.001
Overall health	2.6 (0.8)	2.1 (0.8)	0.4 (−0.1 to 0.9)	.082
App + Healthy Life Center (group 2; n=13)				
Body weight	81.5 (22.3)	80.2 (21.5)	1.3 (−0.5 to 3.1)	.136
Waist circumference (cm)	98.1 (13.8)	95.5 (15.3)	2.6 (−0.7 to 6.0)	.114
COOP/WONCA total score	10.3 (3.2)	9.4 (2.3)	0.9 (−1.5 to 3.3)	.423
Physical fitness	2.1 (1.0)	1.8 (0.8)	0.3 (−0.1 to 0.7)	.104
Feelings	1.8 (1.2)	1.8 (1.1)	0.0 (−0.7 to 0.7)	>.99
Daily activities	1.5 (1.0)	1.1 (0.3)	0.4 (−0.2 to 1.0)	.209
Social activities	1.1 (0.3)	1.1 (0.3)	__[b]	—
Change in health	2.5 (0.7)	2.2 (0.9)	0.9 (−0.3 to 1.5)	.004
Overall health	2.2 (0.8)	1.6 (0.7)	0.5 (0.1 to 0.9)	.012
Healthy Life Center (group 3; n=15)				
Body weight (kg)	89.7 (22.8)	86.9 (21.4)	2.7 (0.9 to 4.6)	.007
Waist circumference (cm)	106.7 (15.4)	103.2 (13.9)	3.5 (1.1 to 6.0)	.009
COOP/WONCA total score	13.1 (3.1)	11.4 (3.3)	1.8 (-0.2 to 3.7)	.070
Physical fitness	2.8 (1.0)	2.4 (1.1)	0.4 (−0.3 to 1.1)	.233
Feelings	1.8 (1.3)	2.1 (1.5)	−0.3 (−0.9 to 0.3)	.265
Daily activities	1.8 (1.1)	1.4 (0.9)	0.4 (−0.4 to 1.1)	.336
Social activities	1.7 (1.1)	1.1 (0.4)	0 (−0.02 to 1.2)	.055
Change in health	2.8 (0.7)	2.1 (0.8)	0.6 (−0.02 to 1.3)	.057
Overall health	2.5 (0.8)	2.0 (0.7)	0.5 (0.1 to 0.9)	.029

[a]COOP/WONCA: Dartmouth Primary Care Cooperative Research Network/World Organization of Family Doctors functional health assessment chart.
[b]The paired sample effect size could not be produced.

Discussion

Principal Findings

According to our preset criteria for feasibility, we found the Plunde app to be feasible to use for promoting and monitoring lifestyle change for people at risk of T2D. App adherence was high, and Plunde was rated with excellent user satisfaction. Furthermore, the app was evaluated as motivating, and feedback from the supervisors was scored as the most motivating factor. The amount of time spent on monitoring and promoting lifestyle change through the app was low. However, the recruitment was time consuming.

XSL•FO
RenderX

To our knowledge this is the first study assessing the feasibility of a mobile phone app to promote and monitor lifestyle change in people at risk of T2D. In our study, the adherence to the app was high. This is an important finding as the effectiveness of mHealth interventions have been proven to be closely related to app adherence and engagement [28]. Recently, Jakob et al [28] found that 5 intervention-related factors indicated positive effects on adherence to mHealth apps for prevention or management of NCDs. This included tailoring and personalization to the individual needs of the user, reminders in the format of individualized push notifications, being user-friendly, a technically stable app design, and personal support that complemented the digital intervention [28]. In line with this, 2 systematic reviews by Dugas et al [29] and Asbjørnsen et al [30], showed that the most frequent behavior change techniques used in mHealth interventions were goal-setting and planning, personalization, feedback and monitoring as well as prompts and cues. All these factors were integrated in the Plunde app and could, therefore, have contributed to the high adherence to the app. The integration of digital features and behavior change techniques may optimize T2D prevention interventions to achieve clinically significant weight loss [31] and improve self-management [32].

The app motivation scores in our study were high. Particularly, participants had high scores on the motivational feedback from the supervisor. Having “a person behind the app” can also be aligned with a so-called blended care model, characterized by a combination of digital interventions and therapist-guided interventions [33-35]. Blended care intervention has shown promising results on promoting lifestyle change in terms of weight loss and physical activity [34-36]. In previous research, the communication with the supervisor was highlighted as crucial, because of the possibility to communicate and receive individualized feedback by a real person [37]. On the contrary, the lack of direct contact and involvement with health care professionals has in previous study been pointed out as a main barrier for feasibility of an app-based intervention for cardiovascular disease and diabetes risk awareness and prevention [38]. An important component of this form of blended care model is the “onboarding” phase of the app intervention [39]. This is likely to be another factor having contributed to the high app adherence in this study. The initial face-to-face consultation enabled further tailored communication through the app, as it focused on the individual’s goals and tasks, and the associated barriers and facilitators to reach them. The freedom of choice and flexibility in a tailored program allows participants to set personalized and meaningful goals [15]. A systematic review and meta-analysis by Joiner et al [12] showed that technology combined with online health coaching resulted in greater weight loss compared to fully automated electronic health interventions. Thus, it is reasonable to assume that a crucial factor for adherence to the app is the individualized follow-up provided by a supervisor. Consequently, to succeed in a full-scale RCT, each participant should have a supervisor which can serve as the person behind the app and provide personalized and tailored feedback.

Weight reduction is the most important factor in preventing T2D, delaying the onset of T2D, and even remission of prediabetes and micro- and macrovascular risk [6,7]. In this feasibility study we assessed body weight, WC, and self-evaluated health at baseline and at 12 weeks follow-up. This was done to get experience with the measurements and to examine a potential change in these outcomes for a potential full-scale trial. In such a trial, one could consider to include the outcome waist-to-height ratio, as this variable has been shown to have an even stronger association with T2D risk than BMI, WC, and the waist-to-hip ratio [25,40,41]. Importantly, this study was designed as a feasibility RCT and did not aim to detect possible differences between groups, thus between-group analysis was not performed. We got valuable experiences with conducting the measurements and collecting data on body weight and WC.

Strengths and Limitations

A strength of this study is the heterogeneity of our sample in terms of age, gender, educational level, and sociodemographic status [42]. This increases the reliability and generalizability of our results, particularly regarding app adherence, app motivation, and user satisfaction. However, participation bias can be considered as a limitation in this study, as the people who agreed to participate are more likely to have a higher motivation for lifestyle change and being more aware of their risk of T2D. It is also important to consider that there is a potential risk in failing to recruit people at the highest risk [43]. However, considering the baseline characteristics of our study sample we seem to have reached the targeted population for risk reduction of T2D.

An important limitation to this study is that we changed our inclusion criteria during the recruitment process, jeopardizing a correct picture of the recruitment process. However, widening the inclusion criteria was a pragmatic choice, to achieve an adequate sample for being able to evaluate the feasibility of Plunde. If we had used the widened criteria from the beginning, the recruitment rate could have been faster. Thus, this has given us valuable considerations to the recruitment process in a potential full-scale RCT.

According to the recommendations for complex interventions and assessing feasibility [44] mixed methods is recommended [45]. Thus, not having qualitative interviews at the follow-up after ended intervention may present a limitation. Qualitative interviews could have provided us with valuable and more extensive information to better understand the complexity of adherence and engagement in use of the app. However, researchers from our group have previously investigated participants experiences with app-based follow-up over a period of 1 year [37]. Qualitative data and results from their work prepared the ground when designing the Plunde app in terms of functionality and likewise the role of the app supervisor to promote adherence and engagement to the app [17].

From a Feasibility Study to a Full-Scale RCT Study

Despite the overall feasibility of the Plunde, our study revealed some adjustments to be made prior commencing a full-scale RCT. To address the challenge of the slow recruitment rate, involvement of a substantial number of GP clinics seems necessary. However, it is important to consider that people at

XSL•FO
RenderX

risk of T2D, do not necessarily visit their GP very often, being in a state often without clinical symptoms [2].

Several recruitment strategies could be considered to enhance the recruitment rate and to increase the number of people available for recruitment. This could be done by developing a campaign for recruitment of participant into a full RCT, mediated or passive recruitment by using a wide variety of social media platforms. In person recruitment of people at risk of T2D has been conducted in venues like public spaces, libraries, churches, and different meeting places to reach people assessable for eligibility [46]. Rapid risk assessments and point-of-care testing of HbA_{1c} levels in community settings, for example GP offices, has been suggested to increase access to testing, facilitating early diagnosis, glycemic awareness, and lifestyle change [47], and could potentially increase the recruitment rate. Furthermore, using electronic health record-supported recruitment approaches has been found to be feasible and promising, and have the potential to increase the reach of eligible participants [48]. This could also be an important approach to facilitate the engagement of GPs and health care professionals in the research [48].

In terms of conducting a potential future 3-arm full-scale RCT, this feasibility study has provided us with valuable insights to the final decision of the design. We got experience with the administration of 3 intervention groups, including acceptability of the randomization procedure, and importantly, we gained important experience on the overall recruitment rate. In terms of conducting a potential future 3-arm full-scale RCT, we consider it as a strength to have evaluated feasibility of the app when used as "stand alone" intervention (group 1) and when used in combination with face-to-face group intervention (group 2). However, the complexity and resource requirements required to conduct a 3-arm design in a full-scale RCT must be taken under consideration. Regarding the usual care group (group 3), the variation in modalities offered by the HLCs and the risk of prolonged inclusion time to the HLC program, due to capacity and resource challenges, adds to the complexity and underscores the importance of an integrated stakeholder involvement in the planning and implementation of a full-scale RCT.

An advantage of conducting a full scale RCT with a 3-arm design, would be the potential to investigate whether an app intervention alone is as good as and not inferior to usual care (HLC). In addition, an evaluation whether a lifestyle intervention delivered as combination of app and HLC would be superior to usual care alone. Follow-up through an app instead of participating in HLC could be the preferred alternative for people because of the flexibility and accessibility an app provides. As a significant proportion of the targeted population for risk reduction of T2D are people committed to work and family responsibilities, and in terms of finding optimal prevention strategies, it is a challenge that the HLCs for the most offer their program during daytime and main working hours.

Despite the growing evidence on the effectiveness of mHealth interventions and lifestyle change for risk reduction of T2D, more evidence is needed regarding long-term adherence and effectiveness [28,49,50]. Considering our positive findings on the low amount of time spent on monitoring and providing follow-up, an assessment of long-term adherence and effectiveness implemented in a full-scale RCT would be feasible and sustainable in terms of resource requirements. When we asked the participants in this study how long they would have continued to use the app, most of the participants answered 12 months. In terms of individualized feedback from the supervisor, most of the app participants answered that they found feedback every month appropriate, suggesting that long-term follow does not necessarily need to be time-consuming from a health care professionals' perspective.

The high speed of technology, innovative app development and the potential of artificial intelligence chatbots as supervisors provides a huge potential in providing scalable mHealth interventions for people at risk of T2D. The need for more research to understand which interventions and components have the greatest reach, promotes engagement and adherence and being most effective in the long term is consistently elucidated [50]. However, this feasibility study leads us to suggest that the potential of digital interventions or mHealth is also to be found beyond the technology itself. The blended care model design of the Plunde app contributes to the research on finding the optimal interaction model between health care professionals and people at risk of T2D.

Conclusions

Based on our preset criteria for success the Plunde app is feasible to monitor and promote lifestyle change in patients at risk of T2D. The amount of time spent on monitoring and promoting lifestyle change through the app was low; however, recruitment was slow. This feasibility study has provided important information and guidance for planning and executing a potential randomized controlled trial.

Acknowledgments

The authors would like to express their gratitude to the people who participated and made this study possible. The authors confirm that all personal identifiers have been removed or disguised so that those described are not identifiable and cannot be identified. They would like to thank the collaborating general practitioners and their secretaries for recruiting participants to our study and the Healthy Life Centers for including the participants to their service and following up in their program. They would like to thank associate professor Way Kiat Bong and MSc student Duy Lee at the Faculty of Technology, Art and Design (TKD) for providing standby IT technical support and resource. This project was internally funded by Oslo Metropolitan University.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questions related to the experience with the setup and use of the smartphone application Plunde.
[PDF File, 106 KB - formative_v9i1e63737_app1.pdf]

Checklist 1
CONSORT-eHEALTH checklist (V1.6.1).
[PDF File, 8365 KB - formative_v9i1e63737_app2.pdf]

References

1. Sun H, Saeedi P, Karuranga S, et al. IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022 Jan;183:109119. [doi: 10.1016/j.diabres.2021.109119] [Medline: 34879977]
2. Echouffo-Tcheugui JB, Perreault L, Ji L, Dagogo-Jack S. Diagnosis and management of prediabetes: a review. JAMA 2023 Apr 11;329(14):1206-1216. [doi: 10.1001/jama.2023.4063] [Medline: 37039787]
3. Magkos F, Hjorth MF, Astrup A. Diet and exercise in the prevention and treatment of type 2 diabetes mellitus. Nat Rev Endocrinol 2020 Oct;16(10):545-555. [doi: 10.1038/s41574-020-0381-5] [Medline: 32690918]
4. Kim HJ, Kwon O. Nutrition and exercise: cornerstones of health with emphasis on obesity and type 2 diabetes management-a narrative review. Obes Rev 2024 Aug;25(8):e13762. [doi: 10.1111/obr.13762] [Medline: 38715378]
5. Uusitupa M, Khan TA, Viguiliouk E, et al. Prevention of type 2 diabetes by lifestyle changes: a systematic review and meta-analysis. Nutrients 2019;11(11):2611. [doi: 10.3390/nu11112611]
6. Glechner A, Keuchel L, Affengruber L, et al. Effects of lifestyle changes on adults with prediabetes: a systematic review and meta-analysis. Prim Care Diabetes 2018 Oct;12(5):393-408. [doi: 10.1016/j.pcd.2018.07.003] [Medline: 30076075]
7. Sandforth A, von Schwartzenberg RJ, Arreola EV, et al. Mechanisms of weight loss-induced remission in people with prediabetes: a post-hoc analysis of the randomised, controlled, multicentre Prediabetes Lifestyle Intervention Study (PLIS). Lancet Diabetes Endocrinol 2023 Nov;11(11):798-810. [doi: 10.1016/S2213-8587(23)00235-8]
8. Jayedi A, Soltani S, Motlagh SZT, et al. Anthropometric and adiposity indicators and risk of type 2 diabetes: systematic review and dose-response meta-analysis of cohort studies. BMJ 2022 Jan 18;376:e067516. [doi: 10.1136/bmj-2021-067516] [Medline: 35042741]
9. Healthy Life Centres in Norway. Helsedirektoratet. 2017. URL: https://www.helsedirektoratet.no/english/healthy-life-centres [accessed 2024-12-27]
10. Tromp J, Jindal D, Redfern J, et al. World Heart Federation roadmap for digital health in cardiology. Glob Heart 2022;17(1):61. [doi: 10.5334/gh.1141] [Medline: 36051317]
11. mHealth: New Horizons for Health through Mobile Technologies: World Health Organization; 2011.
12. Joiner KL, Nam S, Whittemore R. Lifestyle interventions based on the diabetes prevention program delivered via eHealth: a systematic review and meta-analysis. Prev Med 2017 Jul;100:194-207. [doi: 10.1016/j.ypmed.2017.04.033]
13. Lunde P, Nilsson BB, Bergland A, Kværner KJ, Bye A. The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: systematic review and meta-analyses. J Med Internet Res 2018 May 4;20(5):e162. [doi: 10.2196/jmir.9751] [Medline: 29728346]
14. Barengo NC, Diaz Valencia PA, Apolina LM, et al. Mobile health technology in the primary prevention of type 2 diabetes: a systematic review. Curr Diab Rep 2022 Jan;22(1):1-10. [doi: 10.1007/s11892-021-01445-w] [Medline: 35113334]
15. Skoglund G, Nilsson BB, Olsen CF, Bergland A, Hilde G. Facilitators and barriers for lifestyle change in people with prediabetes: a meta-synthesis of qualitative studies. BMC Public Health 2022 Mar 21;22(1):553. [doi: 10.1186/s12889-022-12885-8] [Medline: 35313859]
16. Zangger G, Bricca A, Liaghat B, et al. Benefits and harms of digital health interventions promoting physical activity in people with chronic conditions: systematic review and meta-analysis. J Med Internet Res 2023;25:e46439. [doi: 10.2196/46439]
17. Lunde P, Skoglund G, Olsen CF, Hilde G, Bong WK, Nilsson BB. Think aloud testing of a smartphone app for lifestyle change among persons at risk of type 2 diabetes: usability study. JMIR Hum Factors 2023 Nov 15;10(1):e48950. [doi: 10.2196/48950] [Medline: 37966894]
18. Eldridge SM, Chan CL, Campbell MJ, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016 Oct 24;355:i5239. [doi: 10.1136/bmj.i5239] [Medline: 27777223]
19. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126. [doi: 10.2196/jmir.1923]
20. Tilbud ved frisklivssentraler og veilederkurs [Article in Norwegian]. Helsedirektoratet. 2023. URL: https://www.helsedirektoratet.no/tema/frisklivssentraler/tilbud-ved-frisklivssentraler-og-veilederkurs [accessed 2024-12-27]

XSL•FO
RenderX

21. Armstrong MJ, Mottershead TA, Ronksley PE, Sigal RJ, Campbell TS, Hemmelgarn BR. Motivational interviewing to improve weight loss in overweight and/or obese patients: a systematic review and meta-analysis of randomized controlled trials. Obes Rev 2011 Sep;12(9):709-723. [doi: 10.1111/j.1467-789X.2011.00892.x] [Medline: 21692966]
22. Brooke J. SUS: a quick and dirty usability scale. In: Jordan PW, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry: Taylor & Francis; 1996:189-194.
23. Hyzy M, Bond R, Mulvenna M, et al. System Usability Scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290. [doi: 10.2196/37290] [Medline: 35980732]
24. Waist circumference and waist-hip ratio: report of a WHO expert consultation, Geneva, 8-11 December 2008. World Health Organization Institutional Repository for Information Sharing. 2008. URL: iris.who.int/bitstream/handle/10665/44583/?sequence=1 [accessed 2024-12-27]
25. Ke JF, Wang JW, Lu JX, Zhang ZH, Liu Y, Li LX. Waist-to-height ratio has a stronger association with cardiovascular risks than waist circumference, waist-hip ratio and body mass index in type 2 diabetes. Diabetes Res Clin Pract 2022 Jan;183:109151. [doi: 10.1016/j.diabres.2021.109151] [Medline: 34863718]
26. Siren R, Eriksson JG, Vanhanen H. Waist circumference a good indicator of future risk for type 2 diabetes and cardiovascular disease. BMC Public Health 2012 Aug 9;12(1-6):631. [doi: 10.1186/1471-2458-12-631] [Medline: 22877354]
27. Kinnersley P, Peters T, Stott N. Measuring functional health status in primary care using the COOP-WONCA charts: acceptability, range of scores, construct validity, reliability and sensitivity to change. Br J Gen Pract 1994 Dec;44(389):545-549. [Medline: 7748664]
28. Jakob R, Harperink S, Rudolf AM, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
29. Dugas M, Gao GG, Agarwal R. Unpacking mHealth interventions: a systematic review of behavior change techniques used in randomized controlled trials assessing mHealth effectiveness. Dig Health 2020;6:2055207620905411. [doi: 10.1177/2055207620905411] [Medline: 32128233]
30. Asbjørnsen RA, Smedsrød ML, Solberg Nes L, et al. Persuasive system design principles and behavior change techniques to stimulate motivation and adherence in electronic health interventions to support weight loss maintenance: scoping review. J Med Internet Res 2019 Jun 21;21(6):e14265. [doi: 10.2196/14265] [Medline: 31228174]
31. Van Rhoon L, Byrne M, Morrissey E, Murphy J, McSharry J. A systematic review of the behaviour change techniques and digital features in technology-driven type 2 diabetes prevention interventions. D Health 2020;6:2055207620914427. [doi: 10.1177/2055207620914427] [Medline: 32269830]
32. El-Gayar O, Ofori M, Nawar N. On the efficacy of behavior change techniques in mHealth for self-management of diabetes: a meta-analysis. J Biomed Inform 2021 Jul;119:103839. [doi: 10.1016/j.jbi.2021.103839] [Medline: 34139330]
33. Obro LF, Heiselberg K, Krogh PG, et al. Combining mHealth and health-coaching for improving self-management in chronic care. A scoping review. Patient Educ Couns 2021 Apr;104(4):680-688. [doi: 10.1016/j.pec.2020.10.026] [Medline: 33143907]
34. Berry MP, Sala M, Abber SR, Forman EM. Incorporating automated digital interventions into coach-delivered weight loss treatment: a meta-analysis. Health Psychol 2021;40(8):534-545. [doi: 10.1037/hea0001106]
35. Hohberg V, Fuchs R, Gerber M, Künzler D, Paganini S, Faude O. Blended care interventions to promote physical activity: a systematic review of randomized controlled trials. Sports Med Open 2022 Jul 30;8(1):100. [doi: 10.1186/s40798-022-00489-w] [Medline: 35907158]
36. Kouwenhoven-Pasmooij TA, Robroek SJW, Kraaijenhagen RA, et al. Effectiveness of the blended-care lifestyle intervention “PerfectFit”: a cluster randomised trial in employees at risk for cardiovascular diseases. BMC Public Health 2018 Jun 19;18(1):766. [doi: 10.1186/s12889-018-5633-0] [Medline: 29921255]
37. Lunde P, Bye A, Bruusgaard KA, Hellem E, Nilsson BB. Patients’ experiences of using a smartphone app after cardiac rehabilitation: qualitative study. JMIR Hum Factors 2022 Mar 23;9(1):e34294. [doi: 10.2196/34294] [Medline: 35319476]
38. Buss VH, Varnfield M, Harris M, Barr M. Remotely conducted app-based intervention for cardiovascular disease and diabetes risk awareness and prevention: single-group feasibility trial. JMIR Hum Factors 2022 Jul 1;9(3):e38469. [doi: 10.2196/38469] [Medline: 35776504]
39. Bostrøm K, Børøsund E, Varsi C, et al. Digital self-management in support of patients living with chronic pain: feasibility pilot study. JMIR Form Res 2020 Oct 23;4(10):e23893. [doi: 10.2196/23893] [Medline: 33094734]
40. Moosaie F, Fatemi Abhari SM, Deravi N, et al. Waist-to-height ratio is a more accurate tool for predicting hypertension than waist-to-hip circumference and BMI in patients with type 2 diabetes: a prospective study. Front Public Health 2021;9:726288. [doi: 10.3389/fpubh.2021.726288] [Medline: 34692623]
41. Chen N, Hu LK, Sun Y, et al. Associations of waist-to-height ratio with the incidence of type 2 diabetes and mediation analysis: two independent cohort studies. Obes Res Clin Pract 2023 Jan;17(1):9-15. [doi: 10.1016/j.orcp.2022.12.005]
42. Kontochristopoulou AM, Karatzi K, Karaglani E, et al. Sociodemographic, anthropometric, and lifestyle correlates of prediabetes and type 2 diabetes in europe: the Feel4Diabetes study. Nutr Metab Cardiovasc Dis 2022 Aug;32(8):1851-1862. [doi: 10.1016/j.numecd.2022.03.021] [Medline: 35610084]

XSL•FO
RenderX

43. Bayley A, Stahl D, Ashworth M, et al. Response bias to a randomised controlled trial of a lifestyle intervention in people at high risk of cardiovascular disease: a cross-sectional analysis. BMC Public Health 2018 Sep 4;18(1):1092. [doi: 10.1186/s12889-018-5939-y] [Medline: 30180833]
44. Craig P, Dieppe P, Macintyre S, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655. [doi: 10.1136/bmj.a1655] [Medline: 18824488]
45. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:2061. [doi: 10.1136/bmj.n2061] [Medline: 34593508]
46. Paglialunga S, Bond R, Jaycox SH. Evaluation of HbA1c screening during outreach events for prediabetes subject recruitment for clinical research. Trials 2018 Jan 22;19(1):60. [doi: 10.1186/s13063-018-2459-0] [Medline: 29357915]
47. Gourlay A, Sutherland C, Radley A. Point-of-care testing of HbA1c levels in community settings for people with established diabetes or people at risk of developing diabetes: a systematic review and meta-analysis. Prim Care Diabetes 2024 Feb;18(1):7-16. [doi: 10.1016/j.pcd.2023.10.011] [Medline: 37925311]
48. Aroda VR, Sheehan PR, Vickery EM, et al. Establishing an electronic health record-supported approach for outreach to and recruitment of persons at high risk of type 2 diabetes in clinical trials: the vitamin D and type 2 diabetes (D2d) study experience. Clin Trials 2019 Jun;16(3):306-315. [doi: 10.1177/1740774519839062] [Medline: 31007049]
49. Mönninghoff A, Kramer JN, Hess AJ, et al. Long-term effectiveness of mHealth physical activity interventions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2021 Apr 30;23(4):e26699. [doi: 10.2196/26699] [Medline: 33811021]
50. Stowell M, Dobson R, Garner K, Baig M, Nehren N, Whittaker R. Digital interventions for self-management of prediabetes: a scoping review. PLoS One 2024;19(5):e0303074. [doi: 10.1371/journal.pone.0303074] [Medline: 38728296]

Abbreviations

CONSORT: Consilidated Standard of Reporting Trials
CONSORT-eHEALTH: Consilidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
GP: general practitioner
HLC: Healthy Life Center
mHealth: mobile health
NCD: noncommunicable disease
RCT: randomized controlled trial
SUS: System Usability Scale
T2D: type 2 diabetes
WC: waist circumference

Edited by A Mavragani; submitted 28.06.24; peer-reviewed by S Bonn; revised version received 29.10.24; accepted 30.10.24; published 15.01.25.

Please cite as:
Skoglund G, Hilde G, Lunde P, Cruz Naceno VV, Fromholt Olsen C, Blakstad Nilsson B
Mobile Phone App to Promote Lifestyle Change in People at Risk of Type 2 Diabetes: Feasibility 3-Arm Randomized Controlled Trial
JMIR Form Res 2025;9:e63737
URL: https://formative.jmir.org/2025/1/e63737
doi:10.2196/63737

© Gyri Skoglund, Gunvor Hilde, Pernille Lunde, Venessa Vera Cruz Naceno, Cecilie Fromholt Olsen, Birgitta Blakstad Nilsson. Originally published in JMIR Formative Research (https://formative.jmir.org), 15.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Considering Comorbidities and Individual Differences in Testing a Gaming Behavioral Activation App for Perinatal Depression and Anxiety: Open Trial Pilot Intervention Study

Gabriella E Hamlett[1,2], MA; Chloe Schrader[1,2], BA; Craig Ferguson[3], MS; Lauren A Kobylski[1,4], MPH; Rosalind Picard[3], ScD; Joseph J Locascio[5,6], PhD; Richard J McNally[2], PhD; Lee S Cohen[1,7], MD; Rachel Vanderkruik[1,7], MSc, PhD

[1]Center for Women's Mental Health, Massachusetts General Hospital, Boston, MA, United States

[2]Department of Psychology, Harvard University, Cambridge, MA, United States

[3]MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA, United States

[4]Department of Psychological & Brain Sciences, George Washington University, Washington, DC, United States

[5]Harvard Catalyst Biostatistical Consulting Group and Department of Neurology, Massachusetts General Hospital, Boston, MA, United States

[6]Department of Neurology, Harvard Medical School, Boston, MA, United States

[7]Department of Psychiatry, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Rachel Vanderkruik, MSc, PhD
Center for Women's Mental Health, Massachusetts General Hospital, Boston, MA, United States

Abstract

Background: There is increasing interest in the development of scalable digital mental health interventions for perinatal populations to increase accessibility. Mobile behavioral activation (BA) is efficacious for the treatment of perinatal depression; however, the effect of comorbid anxiety and depression (CAD) on symptom trajectories remains underexplored. This is important given that at least 10% of women in the perinatal period experience CAD.

Objective: We assessed whether there were differences in symptom trajectories in pregnant participants with CAD as compared to those with depression only (ie, major depressive disorder [MDD]) during intervention with a BA mobile gaming app.

Methods: Pregnant adults with either CAD (n=10) or MDD (n=7) used a BA app for 10 weeks and completed biweekly symptom severity questionnaires for depression and anxiety. We assessed whether baseline diagnoses were associated with differential symptom trajectories across the study with mixed effects longitudinal models.

Results: When controlling for baseline symptoms, results revealed a significant interaction between baseline diagnosis and the quadratic component of study week on anxiety (β=.18, SE 0.07; t_{62}=2.61; *P*=.01), revealing a tendency for anxiety in the CAD group to increase initially and then decrease at an accelerated rate, whereas MDD symptoms were relatively stable across time. There was a significant effect of linear time on depression (β=−.39, SE 0.11; t_{68}=−3.51; *P*=.001), showing that depression declined steadily across time for both groups. There was a significant effect of baseline diagnosis on depression (β=−8.53, SE 3.93; t_{13}=−2.17; *P*=.05), suggesting that those with MDD had higher follow-up depression compared to those with CAD when holding other predictors constant.

Conclusions: The app was beneficial in reducing depression symptoms in perinatal individuals with different comorbidity profiles. With respect to anxiety symptom trajectories, however, there was more variability. The app may be especially effective for the treatment of anxiety symptoms among individuals with CAD, as it encourages in-the-moment ecologically relevant exposure to anxiety-provoking stimuli. Despite no significant group difference in baseline anxiety symptoms, the MDD group did not have a significant reduction in their anxiety symptoms across the study period, and some individuals had an increase in anxiety. Findings may point to opportunities for the augmentation of BA gaming apps for those with MDD to more effectively target anxiety symptoms. Overall, findings suggest there may be value in considering comorbidities and individual variations in participants when developing scalable mobile interventions for perinatal populations.

(JMIR Form Res 2025;9:e59154) doi:10.2196/59154

KEYWORDS

perinatal anxiety; perinatal depression; behavioral activation; digital mental health; mobile phone

XSL•FO
RenderX

Introduction

During the perinatal period, 19% - 40% of women experience anxiety or depression, and at least 10% experience comorbid anxiety and depression (CAD [1,2]). Despite its prevalence, CAD in perinatal populations is relatively underexplored [3]. CAD is associated with greater symptom severity, decreased treatment response, and longer episode duration [2-5]. Therefore, more research is needed to assess treatment considerations for perinatal individuals with CAD.

Behavioral activation (BA) is a behavioral intervention that aims to boost engagement in pleasurable activities, emphasizing value-driven behaviors and avoidance reduction, and is efficacious for the treatment of perinatal depression and anxiety [6]. While there is a strong need for the treatment of anxiety and depression in perinatal populations, access to treatment remains challenging and thus further highlights the need for scalable treatments that are accessible for use in ecologically valid settings [7]. Due to accessibility difficulties and maternity health care deserts [8], there is increasing interest in the development of scalable digital mental health interventions for perinatal populations [9-13].

Vanderkruik et al's [14] pilot study on the feasibility and acceptability of a BA gaming app for pregnant women suggests its potential to decrease depression symptoms during pregnancy. However, the effect of comorbidity on symptom trajectories remains unexplored, and few studies focus on individual differences in mobile BA treatment outcomes [14,15]. In this pilot study, we take a hypothesis-generating approach to assess differences in anxiety and depression symptom trajectories in participants with either CAD or major depressive disorder (MDD).

Methods

Recruitment

Participants (pregnant adults, native English speakers, smartphone users, and at least moderately depressed [14]) were recruited throughout the United States between 2021 and 2022 via clinician referrals, the Massachusetts General Hospital Center for Women's Mental Health website [16], and social media advertising. Ineligibility criteria included imminent risk of self-harm, current substance abuse, psychotic disorder, or active mania. All participants (n=18) met the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM-5*) criteria for a current major depressive episode, 10 of whom additionally met criteria for generalized anxiety disorder (GAD) as assessed by the Mini International Neuropsychiatric Interview (MINI [17]). One participant dropped from the study due to active substance use, resulting in a final sample of 17.

Participants were instructed to use the app for the 10-week study period and invited to complete daily "Adventures" (eg, exercising and cleaning) in their real life to earn points and progress in the game. Participants completed surveys for depression and anxiety symptoms biweekly over the 10-week study period [14].

Ethical Considerations

This study is a secondary analysis of deidentified data from the parent study. All original study procedures were approved by the Mass General Brigham institutional review board (2021P001400). All participants provided informed consent in the parent study, which stipulated that deidentified data may be analyzed in secondary analyses without additional consent. Participants were not provided compensation for their participation.

Measures

Eligibility was assessed at baseline with the MINI [17], a semistructured interview conducted by a trained research assistant [14]. Participant grouping (CAD vs MDD) was determined at baseline using MINI diagnostic criteria. The CAD group included individuals who met *DSM-5* criteria for GAD (excessive, difficult to control anxiety and worry for most days for at least 6 months and 3 of the following symptoms: restlessness, being easily fatigued, difficulty concentrating, irritability, muscle tension, and sleep disturbance [18]) and MDD (a depressed mood for most of the day or anhedonia for at least 2-weeks, as well as 5 of the following symptoms: significant unintentional weight or appetite change, insomnia or hypersomnia, psychomotor agitation or retardation, fatigue, feelings of worthlessness or inappropriate or excessive guilt, decreased ability to concentrate or make decisions, and recurrent thoughts of death or suicidal ideation [19]). Those in the MDD group only met the criteria for MDD. Self-reported anxiety and depression symptoms within the past two weeks were measured with the Generalized Anxiety Disorder-7 (GAD-7 [18]) and Patient Health Questionnaire-9 (PHQ-9 [14]), respectively. Study completers were defined as participants who completed the final assessment at the end of the 10-week study period. As published previously [14], app engagement was measured by assessing days using the app (ie, the number of days a participant logged on) and activities completed (ie, the total number of activities a participant completed in the game).

Statistical Analysis

Given the scope of this pilot study, we chose a hypothesis-generating approach with the intent of informing future research, though we expected to see less favorable depression and anxiety symptom trajectories (according to PHQ-9 and GAD-7, respectively) for CAD relative to the MDD group. The analyses were conducted in R (version 4.3.0; R Core Team). Demographic variables, group differences in reductions of anxiety and depression symptoms, app engagement, study completion, and individual symptom trajectories by group were analyzed with descriptive statistics. Independent group *t* tests (2-tailed) explored differences in baseline measures of depression and anxiety by baseline diagnosis.

To assess individual variation in depression and anxiety symptom decreases over time within each group, we plotted individual participant data to visualize trajectories of change for anxiety and depression symptoms across time. To determine whether treatment outcomes changed differentially across time by baseline diagnosis while controlling for baseline depression and anxiety symptom levels, we used mixed effects longitudinal

XSL•FO
RenderX

models. In separate models, PHQ-9 and GAD-7 (hereafter referred to as depression and anxiety symptoms, respectively) were the dependent variables. We controlled for baseline depression and anxiety to ensure that trajectories were due to the app and diagnostic group rather than differences in baseline symptom levels. Fixed predictors were baseline diagnosis group (CAD vs MDD), time (linear and quadratic components of the week in study), baseline depression or anxiety symptoms for the analysis of their respective dependent variables, and all 2 and 3-way interactions of group, time, and baseline depression and anxiety symptoms. Participant-level covariates of baseline age and baseline week gestation were also included to ensure effects were due to the app and not differences in age or weeks gestation. The random effect was participant intercept. We progressively removed nonsignificant covariates and higher-order terms and reran the model. We checked the model residuals to ensure conformance to model assumptions of normality and calculated the proportion of variance in the dependent variable attributable to the fixed effects.

Results

Out of the 17 participants, 59% (n=10) had CAD and 41% (n=7) had MDD at baseline. Participants were primarily White, heterosexual, married, employed, and educated, with private health insurance, as seen in Table 1.

XSL•FO
RenderX

Table . Demographics and participant characteristics.

		Comorbid anxiety and depression (n=10)	Major depressive disorder (MDD; n=7)	Total (N=17)
Age (years), mean (SD)		33.4 (3.0)	35.8 (2.7)	34.4(3.1)
Weeks gestation, mean (SD)		18.1 (8.4)	15.1 (5.9)	16.8 (7.3)
Race or ethnicity, n (%)				
	White	10 (100)	2 (29)	12 (71)
	Black or African American	0 (0)	3 (43)	3 (18)
	Asian	0 (0)	2 (29)	2 (12)
	Non-Hispanic or Latina	9 (90)	6 (86)	15 (88)
	Hispanic or Latina	1 (10)	1 (14)	2 (12)
Sexual orientation, n (%)				
	Heterosexual	7 (70)	7 (100)	14 (82)
	Bisexual	2 (20)	0 (0)	2 (12)
	Queer	1 (10)	0 (0)	1 (6)
Marital status, n (%)				
	Married	9 (90)	3 (43)	12 (71)
	Divorced	0 (0)	1 (14)	1 (6)
	Never married	1 (10)	3 (43)	4 (24)
Employment status, n (%)				
	Employed	7 (70)	7 (100)	14 (82)
	Student	3 (30)	0 (0)	3 (18)
	Disabled or unable to work	0 (0)	0 (0)	0 (0)
Insurance status, n (%)				
	Private health insurance	10 (100)	6 (86)	16 (94)
	Medicaid	0 (0)	1 (14)	1 (6)
Education level, n (%)				
	Postgraduate training	7 (70)	5 (71)	12 (71)
	Bachelor's degree	3 (30)	1 (14)	4 (24)
	Some college	0 (0)	0 (0)	0 (0)
	High school diploma	0 (0)	0 (0)	0 (0)
	Some high school	0 (0)	1 (14)	1 (6)
Treatment, n (%)				
	Psychiatric medication in the last 2 months	6 (60)	2 (29)	8 (47)
	Psychosocial treatment in the last 2 months	5 (50)	3 (43)	8 (47)

Table 1 shows the demographics and participant characteristics of our participants in a pilot study conducted between 2021 and 2022 in the United States testing whether baseline symptom and diagnostic differences influence the symptom trajectory of pregnant individuals using a BA gaming app for perinatal depression and anxiety.

There were no significant group differences in age ($mean_{CAD}$ 33.4, SD_{CAD} 3.0 years and $mean_{MDD}$ 35.8, SD_{MDD} 2.7 years), baseline weeks gestation ($mean_{CAD}$ 18.1, SD_{CAD} 8.4 and $mean_{MDD}$ 15.1, SD_{MDD} 5.9), or baseline depression symptoms ($mean_{CAD}$ 12.8, SD_{CAD} 3.3 and $mean_{MDD}$ 11.7, SD_{MDD} 3.5). There was no significant difference in baseline past 2-week anxiety symptoms ($mean_{CAD}$ 10.7, SD_{CAD} 2.7 and $mean_{MDD}$ 7.0, SD_{MDD} 2.7; Table 2). That is, while both diagnostic groups started with similar past 2-week levels of anxiety, only the CAD group met full *DSM-5* criteria for GAD, which requires a duration of at least 6 months of symptoms. The CAD group was 3.6 times more likely to be using psychotropic medications (6

XSL•FO
RenderX

out of 10, 60% of participants with CAD, 2 out of 7, 29% of participants with MDD) though with this sample size, it did not reach statistical significance (*P*=.32). The CAD group was 1.2 times more likely to be in therapy (5 out of 10, 50% participants with CAD and 3 out of 7, 43% participants with MDD); findings did not reach significance (*P*=.63).

Table . Mean changes from baseline to follow-up in anxiety and depression for completers.

		PHQ-9[a] (mean, SD)	GAD-7[b] (mean, SD)
Comorbid anxiety and depression			
	Baseline (n=10)	12.8 (3.7)	10.7 (2.7)
	Baseline completers (n=7)	12.7 (3.9)	10.0 (2.9)
	Follow-up (n=7)	7.0 (3.4)	6.0 (2.6)
	Completers mean change score	−5.7 (4.3)	−4.0 (4.2)
Major depressive disorder			
	Baseline (n=7)	11.7 (3.5)	7.0 (2.7)
	Baseline completers (n=3)	13.7 (4.5)	7.3 (0.6)
	Follow-up (n=3)	7.7 (4.9)	9.7 (6.5)
	Completers mean change score	−6.0 (1.7)	2.3 (6.5)

[a]PHQ-9: Patient Health Questionnaire-9.

[b]GAD-7: Generalized Anxiety Disorder-7.

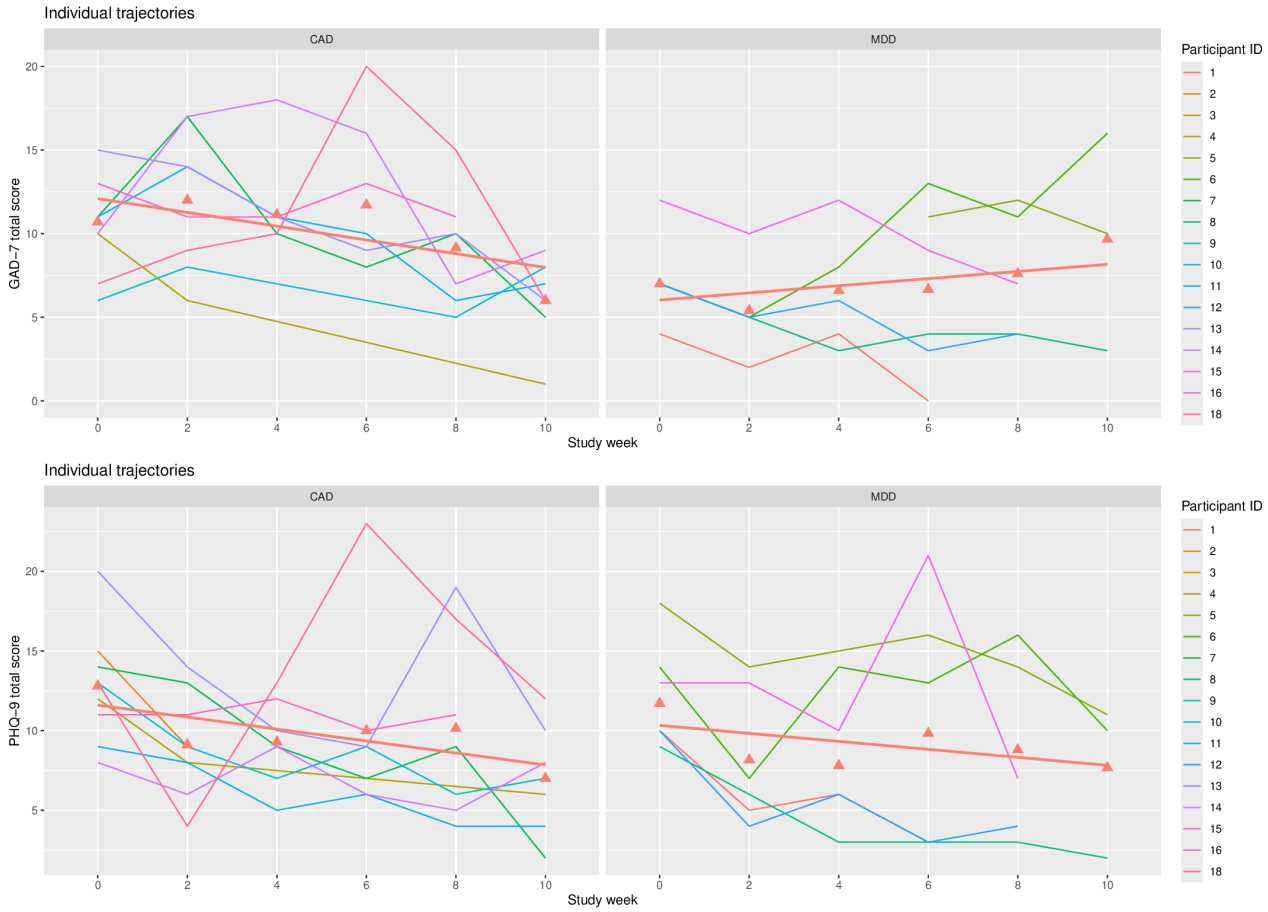
Of the 10 participants with CAD, 7 (70%) completed the last study assessment and of the original 7 participants with MDD, 3 (43%) completed the last study assessment. Baseline diagnosis (CAD vs MDD) was not a significant predictor of whether participants completed the last assessment (β=−1.36, SE 1.00; *z* score=−1.35; *P*=.18). There was also no significant difference in the number of days using the app between the CAD group and the MDD group ($mean_{CAD}$ 18.0, SD_{CAD} 16.20; $mean_{MDD}$ 11.6, SD_{MDD} 19.79; $t_{7.04}$=0.62; *P*=.56; 95% CI −18.09 to 30.89). There was no significant difference in the number of tasks completed between the CAD group and the MDD group ($mean_{CAD}$ 13.22, SD_{CAD} 14.07; $mean_{MDD}$ 6.0, SD_{MDD} 12.86; $t_{9.08}$=0.97; *P*=.36; 95% CI −9.55 to 23.99).

There was no significant difference between the depression symptom mean change score for completers from baseline to follow-up for the CAD group compared to the MDD group ($mean_{CAD}$ −5.7, SD_{CAD} 4.3; $mean_{MDD}$ −6.0, SD_{MDD} 1.7; t_8=0.11; *P*=.92; 95% CI −5.8 to −6.4). There was also no significant difference in the anxiety symptom mean change score for completers from baseline to follow-up for the CAD group compared to the MDD group ($mean_{CAD}$ −4.0, SD_{CAD} 4.2; $mean_{MDD}$ 2.3, SD_{MDD} 6.5; t_8=−1.87; *P*=.09; 95% CI −14.1 to 1.5).

When assessing individual-level trajectories of depression and anxiety symptom scores across study week by participant, we see within-group variation both for depression and anxiety symptom scores (Figure 1). In the MDD group, most individuals' anxiety symptoms decreased over time, though 1 individual demonstrated an increase in anxiety. The anxiety of some individuals with CAD increased before decreasing over time. While overall depression symptoms decreased over time, a few individuals in both groups displayed increases in symptoms.

XSL•FO
RenderX

Figure 1. Individual symptom trajectories of anxiety (GAD-7) and depression (PHQ-9) by baseline diagnosis. Figure 1 shows raw score symptom trajectories for each participant on anxiety (GAD-7) and depression (PHQ-9) symptoms by baseline diagnosis in a pilot study conducted between 2021 and 2022 in the United States testing whether baseline symptom and diagnostic differences influence symptom trajectory in pregnant individuals using a behavioral activation gaming app for perinatal depression and anxiety. Red triangles denote the mean score at each study week, with the red line representing a least squares regression line fit to these means. CAD: comorbid anxiety and depression; GAD-7: Generalized Anxiety Disorder-7; MDD: major depressive disorder; PHQ-9: Patient Health Questionnaire-9.

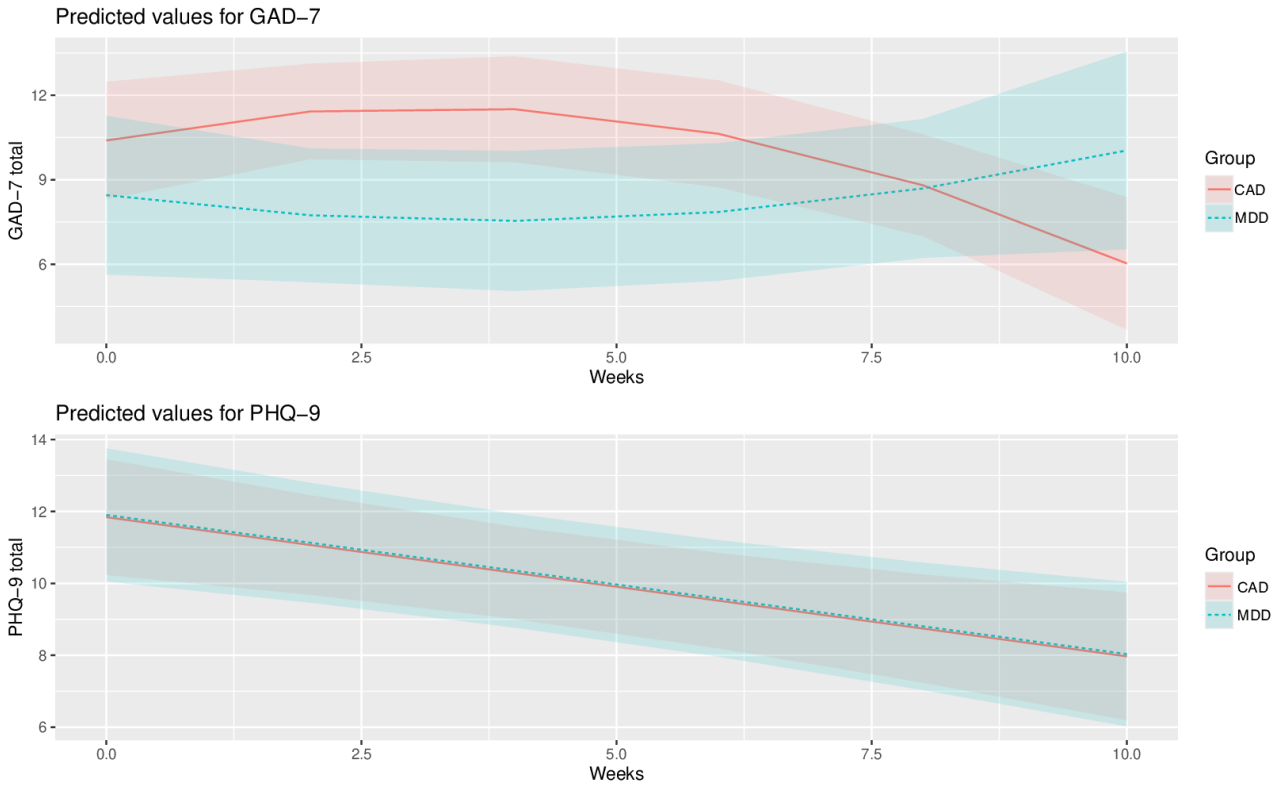


For mixed effects models, nonsignificant higher-order terms and covariates were removed, and models rerun. Relevant to the assessment of any association of baseline diagnosis to change in anxiety across time controlling for baseline anxiety, there was a significant interaction between baseline diagnosis and quadratic study week (β=.18, SE 0.07; t_{62}=2.61; *P*=.01). This interaction reflects a tendency for the CAD group anxiety symptoms to increase initially and then decrease at an accelerated rate, whereas the MDD groups' anxiety symptoms were relatively stable across time or even showed a slight "U" shape pattern across time (Figure 2). Overall, the fixed effects accounted for 33.9% of the total variance in the anxiety symptom scores. All other fixed effects of interest were nonsignificant (Table 3).

For the assessment of baseline diagnosis of depression across time controlling for baseline depression, there was a significant effect of linear time on depression symptom scores (β=−.39, SE 0.11; t_{68}=−3.51; *P*<.001), showing that depression symptoms decline steadily across time for both groups. We also found a significant effect of baseline diagnosis on depression symptoms at follow-up (β=−8.53, SE 3.93; t_{13}=−2.17; *P*=.05), suggesting that being in the MDD group was associated with slightly higher follow-up depression symptom scores compared to the CAD groups' follow-up depression symptom scores across time (Figure 2), holding other predictors (eg, baseline depression symptoms), constant. There was also a marginally significant interaction of baseline diagnosis with baseline depression symptoms (β=.65, SE 0.31; t_{13}=2.12; *P*=.054) suggesting that the difference in follow-up depression scores by baseline diagnosis difference was slightly stronger among those with lower baseline depression levels than at higher levels.

XSL•FO
RenderX

Figure 2. Model fixed effect predicted values of GAD-7 and PHQ-9 over time. Figure 2 shows effects plots from the 2 mixed effects models predicting anxiety (GAD-7; top panel) and depression (PHQ-9; bottom panel) across time in a pilot study conducted between 2021 and 2022 in the United States testing whether baseline symptom and diagnostic differences influence symptom trajectory in pregnant individuals using a behavioral activation gaming app for perinatal depression and anxiety. Predicted values from model fixed effects are shown with 95% confidence bands around the lines, controlling for baseline symptoms. CAD: comorbid anxiety and depression; GAD-7: Generalized Anxiety Disorder-7; MDD: major depressive disorder; PHQ-9: Patient Health Questionnaire-9.



XSL•FO
RenderX

Table . Mixed models fixed effects results. Table 3 shows fixed effects results from both mixed effects models in a pilot study conducted between 2021 and 2022 in the US testing whether baseline symptom and diagnostic differences influence symptom trajectory in pregnant individuals using a behavioral activation gaming app for perinatal depression and anxiety.

		β (SE)	*t* test (*df*)	*P* value
GAD-7 [a] model				
	Intercept	5.06 (2.62)	1.93 (15)	.07
	Baseline diagnosis	−1.94 (1.81)	−1.07 (32)	.29
	Time[b]	.75 (0.44)	1.72 (63)	.09
	Time^{2}[c]	−.12 (0.04)	−2.76 (62)	.007
	Baseline GAD-7	.53 (0.23)	2.35 (13)	.04
	Baseline diagnosis time[b]	−1.24 (0.69)	−1.81 (62)	.08
	Baseline diagnosis time^{2}[c]	.18 (0.07)	2.61 (62)	.01
PHQ-9 [d] model				
	Intercept	4.37 (2.52)	1.73 (12)	.11
	Baseline diagnosis	−8.53 (3.93)	−2.17 (13)	.049
	Baseline PHQ-9	.57 (0.18)	3.07 (11)	.01
	Time	−.39 (0.11)	−3.51 (68)	.001
	Baseline diagnosis baseline PHQ-9	.65 (0.31)	2.12 (13)	.054

[a]GAD-7: Generalized Anxiety Disorder-7.

[b]Time: linear time.

[c]Time^{2}: quadratic time.

[d]PHQ-9: Patient Health Questionnaire-9.

Discussion

Principal Results

In this pilot study, our objective was to take a hypothesis-generating exploratory approach to assess differences in anxiety and depression symptom trajectories in participants with either CAD or MDD. While we expected to see less favorable depression and anxiety symptom trajectories (according to PHQ-9 and GAD-7, respectively) for CAD relative to the MDD group, our first main finding was that anxiety in the CAD group increased initially and then decreased at an accelerated rate, whereas in the MDD group, anxiety symptoms remained relatively stable across time despite no difference in baseline anxiety symptoms. Our other main finding was that depression symptoms declined steadily across time for both groups; however, those with MDD had higher follow-up depression scores compared to those with CAD, despite no differences in baseline depression symptoms. Given that there were no group differences in app engagement, these findings may not be explained by adherence or engagement. Our findings provide preliminary support for considering individual differences in baseline diagnostic characteristics when developing personalized mobile interventions, particularly in the context of treating perinatal individuals with psychological comorbidities.

With respect to the anxiety symptom trajectories, our findings are in line with the literature showing that compared to active controls, BA has a large effect on the reduction of depressive symptoms and a small effect on the reduction of anxiety symptoms when anxiety is secondary to depression [20]. In our study, it appears that the BA gaming app may be associated with a substantial reduction of anxiety symptoms for individuals with CAD. For these individuals, it may be that the BA components of the app target avoidance behaviors, which could result in a temporary increase in anxiety followed by reductions in anxiety potentially due to increased reward, habituation, and violation of expectations [6,21]. These findings suggest that the app may be especially effective for individuals with CAD, as it encourages in-the-moment ecologically relevant exposure to anxiety-provoking stimuli [7]. Particularly for individuals with CAD, for whom daily external stimuli may be especially provoking, the app may be especially useful and serve as an exposure.

Despite no significant group difference in baseline anxiety, the MDD group did not have a significant reduction in their anxiety symptoms across the study period, and some individuals had an increase in anxiety. Future studies should explore whether the anxiety experienced by the MDD group is maintained by other cognitive mechanisms to improve targeting anxiety symptoms when anxiety symptoms are secondary to depression symptoms. Findings may point to opportunities for the

XSL•FO
RenderX

augmentation of BA gaming apps for those with MDD to more effectively target anxiety symptoms. Future treatment personalization studies should also assess whether notifying patients of common treatment trajectories based on baseline presentations may improve treatment adherence. Findings of individual variability in anxiety and depression over time point to an opportunity for future more highly powered studies investigating the risk for increasing anxiety and depression across treatment during pregnancy.

Regarding our hypothesis that the CAD group may have less favorable depression symptom trajectories, we instead found that depression symptoms decreased linearly across time for both groups and that the MDD group had higher depression symptom scores at follow-up compared to the CAD groups' follow-up symptoms, suggesting that the app may be beneficial in reducing depression in perinatal individuals with different comorbidity profiles. These findings are in line with results from traditional and mobile BA interventions that BA is efficacious for the reduction of depressive symptoms (eg, [6,20,22]) even for individuals with different comorbidity profiles.

Limitations

This pilot study has several limitations (eg, modest sample size and no control group; discussed further in Vanderkruik et al [14]). Future fully powered studies should consider app engagement when assessing differences in symptom trajectories by baseline diagnosis. Due to the small sample size and lack of a control group, the results should be interpreted with caution and be considered as hypothesis generating.

Conclusions

The findings provide preliminary support for considering individual differences and comorbidities when developing scalable mobile interventions for perinatal populations. This pilot study has potentially important broad implications. Demand for treatment, particularly in perinatal populations, is high and yet can be hard to access for a variety of reasons, including the fact that many traditional interventions require presence in a formal and controlled therapy environment. This study represents an initial step toward developing scalable interventions and highlights the need to consider baseline differences in future studies that may personalize the app experience to increase adherence and effectiveness. Further research involving an adequately powered randomized controlled trial may further illuminate individual differences affecting the impact of mobile interventions.

Acknowledgments

The authors would like to thank all study participants for their participation in this study. This study was funded by the Harvard Medical School Kaplen Fellowship on Depression to RV at the MGH Center for Women's Mental Health. Generative artificial intelligence was not used for this paper.

Authors' Contributions

LSC, RV, and GEH contributed to the study conception and design. Material preparation and data collection were performed by RV, LAK, and CF. Analyses were performed by GEH and CS. JJL provided consultation. The first draft of the paper was written by GEH with contributions from RV and CS, and all authors commented on previous versions of the study. All authors read and approved the final paper. Funding was acquired by RV, and supervision was provided by JJL, RP, RJM, LSC, and RV.

Conflicts of Interest

LSC, as of February 26, 2024, receives research support as an employee of Massachusetts General Hospital (MGH) and works with the MGH National Pregnancy Registry. The current sponsors of MGH National Pregnancy Registry are Alkermes, Inc. (2016 to present); Dr. Reddy's Laboratories, Inc. (2023 to present); Eisai Inc. (2022 to present); Otsuka America Pharmaceutical, Inc. (2008 to present); Supernus Pharmaceuticals (2021 to present); and Teva Pharmaceutical Industries Ltd. (2018 to present). As an employee of MGH, LSC works with the MGH
Clinical Trials Network and Institute, which has had research funding from multiple pharmaceutical companies and the National Institute of Mental Health. LSC reports other research support from National Institutes of Health and SAGE Therapeutics but no support from advisory or consulting roles, no speaking fees or honoraria, no royalties or patents, and no other income. RV, as of March 14, 2024, receives research support as an employee of MGH and works on projects funded by The J Willard and Alice S Marriott Foundation and the National Eating Disorder Association. RV consults for the World Health Organization and Soulside, Inc.

References

1. Dennis CL, Falah-Hassani K, Shiri R. Prevalence of antenatal and postnatal anxiety: systematic review and meta-analysis. Br J Psychiatry 2017 May;210(5):315-323. [doi: 10.1192/bjp.bp.116.187179] [Medline: 28302701]
2. Falah-Hassani K, Shiri R, Dennis CL. The prevalence of antenatal and postnatal co-morbid anxiety and depression: a meta-analysis. Psychol Med 2017 Sep;47(12):2041-2053. [doi: 10.1017/S0033291717000617] [Medline: 28414017]

XSL•FO
RenderX

3. Goodman JH, Tyer-Viola L. Detection, treatment, and referral of perinatal depression and anxiety by obstetrical providers. J Womens Health (Larchmt) 2010 Mar;19(3):477-490. [doi: 10.1089/jwh.2008.1352]
4. Cuijpers P, Miguel C, Ciharova M, et al. Psychological treatment of depression with other comorbid mental disorders: systematic review and meta-analysis. Cogn Behav Ther 2023 May;52(3):246-268. [doi: 10.1080/16506073.2023.2166578] [Medline: 36718645]
5. Hopwood M. Anxiety symptoms in patients with major depressive disorder: commentary on prevalence and clinical implications. Neurol Ther 2023 Apr;12(Suppl 1):5-12. [doi: 10.1007/s40120-023-00469-6] [Medline: 37115459]
6. Dimidjian S, Goodman SH, Sherwood NE, et al. A pragmatic randomized clinical trial of behavioral activation for depressed pregnant women. J Consult Clin Psychol 2017;85(1):26-36. [doi: 10.1037/ccp0000151]
7. Ebrahimi OV, Asmundson GJG. Scaling up psychological interventions into the daily lives of patients with anxiety and related disorders. J Anxiety Disord 2024 Aug;106:102916. [doi: 10.1016/j.janxdis.2024.102916] [Medline: 39178811]
8. DiPietro Mager NA, Zollinger TW, Turman JE Jr, Zhang J, Dixon BE. Routine healthcare utilization among reproductive-age women residing in a rural maternity care desert. J Community Health 2021 Feb;46(1):108-116. [doi: 10.1007/s10900-020-00852-6] [Medline: 32488525]
9. Hadfield H, Wittkowski A. Women's experiences of seeking and receiving psychological and psychosocial interventions for postpartum depression: a systematic review and thematic synthesis of the qualitative literature. J Midwife Womens Health 2017 Nov;62(6):723-736. [doi: 10.1111/jmwh.12669]
10. Chen C, Wang X, Xu H, Li Y. Effectiveness of digital psychological interventions in reducing perinatal depression: a systematic review of meta-analyses. Arch Womens Ment Health 2023 Aug;26(4):423-439. [doi: 10.1007/s00737-023-01327-y] [Medline: 37231116]
11. Lara-Cinisomo S, Ramirez Olarte A, Rosales M, Barrera AZ. A systematic review of technology-based prevention and treatment interventions for perinatal depression and anxiety in Latina and African American women. Matern Child Health J 2021 Feb;25(2):268-281. [doi: 10.1007/s10995-020-03028-9] [Medline: 33389589]
12. Lewkowitz AK, Whelan AR, Ayala NK, et al. The effect of digital health interventions on postpartum depression or anxiety: a systematic review and meta-analysis of randomized controlled trials. Am J Obstet Gynecol 2024 Jan;230(1):12-43. [doi: 10.1016/j.ajog.2023.06.028]
13. Franco P, Olhaberry M, Muzard A, Harismendy Á, Kelders S. Developing a guided web app for postpartum depression symptoms: user-centered design approach. JMIR Form Res 2024 Aug 19;8(1):e56319. [doi: 10.2196/56319] [Medline: 39159447]
14. Vanderkruik RC, Ferguson C, Kobylski LA, et al. Testing a behavioral activation gaming app for depression during pregnancy: multimethod pilot study. JMIR Form Res 2024 Jan 26;8:e44029. [doi: 10.2196/44029] [Medline: 38277191]
15. Mancinelli E, Dell'Arciprete G, Pattarozzi D, Gabrielli S, Salcuni S. Digital behavioral activation interventions during the perinatal period: scoping review. JMIR Pediatr Parent 2023 Feb 28;6:e40937. [doi: 10.2196/40937] [Medline: 36853756]
16. Women's mental health across the life cycle. Massachusetts General Hospital Center for Women's Mental Health. URL: https://womensmentalhealth.org/ [accessed 2024-10-01]
17. Lecrubier Y, Sheehan D, Weiller E, et al. The MINI International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. Eur psychiatr 1997;12(5):224-231. [doi: 10.1016/S0924-9338(97)83296-8]
18. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
19. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care 2004 Dec;42(12):1194-1201. [doi: 10.1097/00005650-200412000-00006] [Medline: 15550799]
20. Stein AT, Carl E, Cuijpers P, Karyotaki E, Smits JAJ. Looking beyond depression: a meta-analysis of the effect of behavioral activation on depression, anxiety, and activation. Psychol Med 2021 Jul;51(9):1491-1504. [doi: 10.1017/S0033291720000239] [Medline: 32138802]
21. Hamlett GE, Foa EB, Brown LA. Exposure therapy and its mechanisms. In: Milad MR, Norrholm SD, editors. Fear Extinction: From Basic Neuroscience to Clinical Implications: Springer International Publishing; 2023:273-288. [doi: 10.1007/78542023428]
22. Dahne J, Wahlquist AE, Kustanowitz J, et al. Behavioral activation-based digital smoking cessation intervention for individuals with depressive symptoms: randomized clinical trial. J Med Internet Res 2023 Nov 1;25:e49809. [doi: 10.2196/49809] [Medline: 37910157]

Abbreviations

BA: behavioral activation
CAD: comorbid anxiety and depression
***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*
GAD: generalized anxiety disorder

XSL•FO
RenderX

GAD-7: Generalized Anxiety Disorder-7
MDD: major depressive disorder
MINI: Mini International Neuropsychiatric Interview
PHQ-9: Patient Health Questionnaire-9

Edited by A Mavragani; submitted 29.04.24; peer-reviewed by M Zimmermann; revised version received 01.11.24; accepted 05.11.24; published 14.01.25.

Please cite as:
Hamlett GE, Schrader C, Ferguson C, Kobylski LA, Picard R, Locascio JJ, McNally RJ, Cohen LS, Vanderkruik R
Considering Comorbidities and Individual Differences in Testing a Gaming Behavioral Activation App for Perinatal Depression and Anxiety: Open Trial Pilot Intervention Study
JMIR Form Res 2025;9:e59154
URL: https://formative.jmir.org/2025/1/e59154
doi:10.2196/59154

© Gabriella E Hamlett, Chloe Schrader, Craig Ferguson, Lauren A Kobylski, Rosalind Picard, Joseph J Locascio, Richard J McNally, Lee S Cohen, Rachel Vanderkruik. Originally published in JMIR Formative Research (https://formative.jmir.org), 14.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Verification of the Reliability of an Automated Urine Test Strip Colorimetric Program Using Colorimetric Analysis: Survey Study

Keigo Inagaki[1*], DPhil; Daisuke Tsuriya[2*], DPhil; Takuya Hashimoto[2], DPhil; Katsumasa Nakamura[2], Prof Dr Med

[1]Clinical Nursing, Gerontological Nursing, Faculty of Nursing, Hamamatsu University School of Medicine, 1-20-1 Handayama, Chuo-ku, Hamamatsu city, Shizuoka, Japan

[2]Hamamatsu University School of Medicine, Hamamatsu City, Chuo-ku, Japan

[*]these authors contributed equally

Corresponding Author:
Keigo Inagaki, DPhil
Clinical Nursing, Gerontological Nursing, Faculty of Nursing, Hamamatsu University School of Medicine, 1-20-1 Handayama, Chuo-ku, Hamamatsu city, Shizuoka, Japan

Abstract

Background: One method for noninvasive and simple urinary microalbumin testing is urine test strips. However, when visually assessing urine test strips, accurate assessment may be difficult due to environmental influences—such as lighting color and intensity—and the physical and psychological influences of the assessor. These complicate the formation of an objective assessment.

Objectives: This study developed an "automated urine test strip colorimetric program" (hereinafter referred to as "this program") to objectively assess urine test strips. Using this program may allow urine tests to be conducted at home. In this study, urine samples from hospitalized or outpatient patients were randomly obtained, and the reliability of this program was verified by comparing the agreement rate between this program and an automatic urine analyzer (US-3500 [Eiken Chemical Co, Ltd] and LABOSPECT 006 [Hitachi High-Tech Co, Ltd]). Furthermore, the sensitivity and specificity of the urine albumin test were investigated, and its applicability to screening for microalbuminuria was verified.

Methods: A urine test strip was placed in a photography box with constant light intensity and color temperature conditions. The image taken with a smartphone camera on top of the photography box was judged by this program. This program used Accelerated KAZE to perform image-matching processing to reduce the effect of misalignment during photography. It also calculated and judged the item with the smallest color difference between the color chart and the urine test strip using the CIEDE2000 color difference formula. The agreement rate of the results of this program was investigated using the results of an automatic urine analyzer as the gold standard.

Results: Compared with the judgments of an automatic urine analyzer, the average agreement rate for 12 items (protein, glucose, urobilinogen, bilirubin, ketone bodies, specific gravity, occult blood, pH, white blood cells, nitrite, creatinine, and albumin) was 78.6%. Furthermore, the average agreement rate of the 12 items within ±1 rank was 95.4%. The results showed a sensitivity of 100% and a specificity of 58.6% in determining albumin in urine, which is important for determining the stage of diabetic nephropathy. Finally, the area under the curve (0.907) derived from the receiver operating characteristic curve was satisfactory.

Conclusions: The program developed by the authors can determine urine test strips without requiring calibration in a certain shooting environment. If this program can be used at home to perform urinary microalbumin tests, the early detection and treatment of diabetic nephropathy may prevent the condition from becoming severe.

(JMIR Form Res 2025;9:e62772) doi:10.2196/62772

KEYWORDS

urine test strip; reliability; automatic urine analyzer; quasi-experimental study; colorimetric analysis; colorimetric; urinalysis; urinary; urine; evaluation; mobile phone

Introduction

Urinalysis testing displays vast physiological information and may be an effective tool for diagnosing various human diseases, including metabolic dysfunctions (eg, diabetes, urinary tract infections, and malignancies) and renal, cardiovascular, or hepatic disease [1-3]. Urinalysis test strips are an inexpensive, easy-to-use method that has long been used in clinical settings [4,5].

Urinalysis test strips change color in response to the concentration of the analyte present in the urine sample. Urinalysis test strips that react with urine analytes are checked against a reference color chart of several color blocks, each representing a specific analyte concentration.

XSL•FO
RenderX

The interpretation of colorimetric changes can be influenced by the health and psychological state of the observer [6,7]. Due to the lack of standardization and reading methods, it is not routinely used at home. The concept of self-screening at home has been progressively explored from multiple perspectives, and various types of devices have been developed depending on the application. For example, a Food and Drug Administration–approved product (Healthy.io) that uses a smartphone app to test items on urinalysis test strips for the early detection of chronic kidney disease and urinary tract infections is on the market. Various technologies exist for objectively judging urine test strips using smartphone apps such as “Vivoo” and “Uchek.”

This study developed an automated urine test strip colorimetric program to automatically judge urinalysis test strips from smartphone images [8]. This program automatically identified the item with the most negligible color difference between the reference color chart and urinalysis test strip and displayed the urinalysis results. Since the assessment was unaffected by the measurer’s physical condition or psychological state, the proposed method provided an objective rather than a subjective visual evaluation.

This study verified the reliability of an automated urine test strip colorimetric program, investigated the sensitivity and specificity of urinary albumin testing, and verified its applicability for microalbuminuria screening. Using real patient urine specimens, the concordance rate between creatinine and albumin and the 10 previously validated items was assessed. The concordance rate for each judgment item and the sensitivity and specificity of albumin were further investigated.

Methods

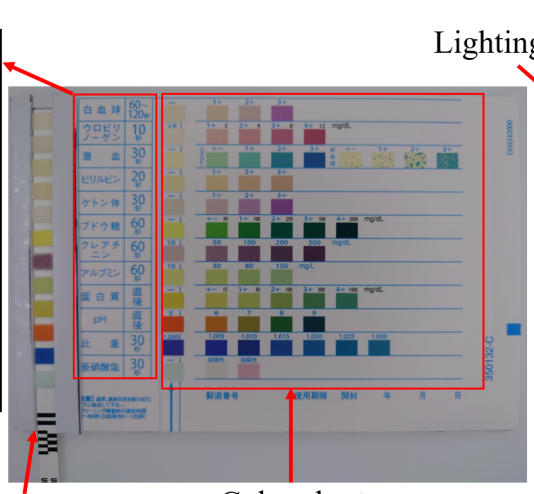
Overview

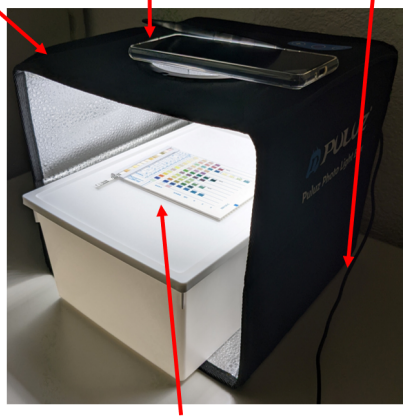
Figure 1 depicts this study’s experimental environment. A photography box (PU5025B; PULUZ) was used to maintain a constant lighting environment. An Xperia 5II (XQ-AS42, SONY) smartphone was used. The smartphone had a standard triple-lens camera (ZEISS lens) with a 1/1.7-inch large format sensor, approximately 12.2 million pixels/24-mm focal length, and a 1.7/dual photodiode/hybrid image stabilization (optical+ electronic) *F* value. A urinalysis test strip (Uropaper III Eiken [E-UR97]; Eiken Chemical Co, Ltd) capable of testing 12 items was used as the test strip. The reference color chart, commonly used in clinical practice, was used as a comparison target for the urinalysis test strips (Figure 1). From June 2022 to January 2023, 1184 urine specimens were randomly extracted from specimens handled by the laboratory department of a hospital in Hamamatsu City. The inclusion criteria were urine samples from people aged 20-90 years who were evaluated at the laboratory department of a hospital in Hamamatsu City. The exclusion criteria were urine samples from people admitted to the pediatric ward and people with cloudy urine or gross hematuria. Basic information on the specimens is shown in Table 1. The US-3500 (Eiken Chemical Co, Ltd) and LABOSPECT 006 (Hitachi High-Tech Co, Ltd) were used as automatic urine analyzers for comparison with the results of the proposed method. The US-3500 was used to compare urobilinogen, occult blood, protein, glucose, ketone bodies, bilirubin, nitrite, specific gravity, white blood cells, and pH, and the LABOSPECT 006 was used to compare creatine and albumin levels. All 1184 specimens of urobilinogen, occult blood, protein, glucose, ketone bodies, bilirubin, nitrite, specific gravity, and pH were obtained using an automatic analyzer. Automated analyzer data were obtained for 1182 leukocyte, 214 creatine, and 42 albumin samples. The automatic urine test strip colorimetric program developed in this study was created using a desktop computer (HP EliteDesk 800 G4 TWR, HP), with Python as the development language and OpenCV library for image processing.

Figure 1. Reference color chart and shooting environment used for color comparison of urine test strips.

Item	Measurement time
Leukocytes	60-120 seconds
Urobilinogen	10 seconds
Occult blood	30 seconds
Bilirubin	20 seconds
Ketone	30 seconds
Glucose	60 seconds
Creatinine	60 seconds
Albumin	60 seconds
Protein	Immediately after
pH	Immediately after
Specific gravity	30 seconds
Nitrite	30 seconds

For all items, the measurement time for screening tests is 30-60 seconds.





XSL•FO
RenderX

Table . Basic information on 1184 urine samples randomly extracted from specimens handled by the testing department of a hospital in Hamamatsu City from June 2022 to January 2023.

Content	Individuals (N=1184)
Sex, n (%)	
Female	648 (54.7)
Male	536 (45.3)
Sex, Average age (SD)	
Female	48.16 (21.4)
Male	58.00 (22.8)
Clinical department, n (%)	
Department of rehabilitation	1 (0.1)
General surgery	12 (1)
Lower gastrointestinal surgery	14 (1.2)
Liver internal medicine	36 (3)
Hepato-biliary-pancreatic surgery	3 (0.3)
Ophthalmology	57 (4.8)
Hematology	55 (4.6)
Vascular surgery	1 (0.1)
Respiratory surgery	8 (0.7)
Respiratory medicine	31 (2.6)
Obstetrics and gynecology	263 (22.2)
Oral surgery	5 (0.4)
Otorhinolaryngology	6 (0.5)
Cardiology	24 (2)
Pediatrics	86 (7.3)
Gastroenterology	23 (1.9)
Upper gastrointestinal surgery	53 (4.5)
Cardiovascular surgery	18 (1.5)
Nephrology	73 (6.2)
Orthopedics	66 (5.6)
Psychiatry	4 (0.3)
Endocrinology and metabolism	131 (11.1)
Breast surgery	5 (0.4)
Neurosurgery	10 (0.8)
Urology	169 (14.3)
Dermatology	14 (1.2)
Immunology/rheumatology	15 (1.3)
Clinical pharmacology	1 (0.1)

XSL•FO
RenderX

Experimental Methods

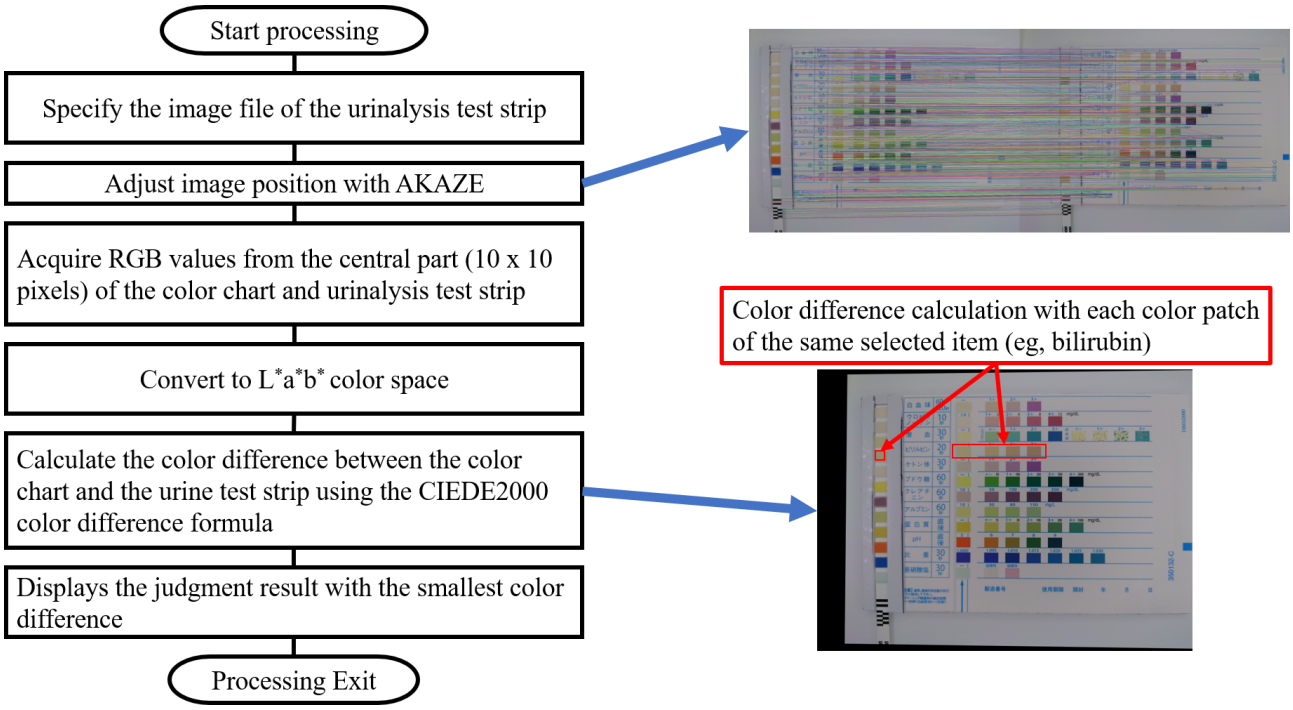
Judgment of the Automatic Urine Dipstick Colorimetric Program

Automated Urine Dipstick Colorimetric Program

Figure 2 shows the processing flow of the automatic urine dipstick colorimetric program. The colors of the test paper for the urine test and color chart were compared according to the following procedure:

1. Specify the image file of the urinalysis strip.
2. Adjust the image position using Accelerated KAZE (AKAZE).
3. Acquire RGB (red, green, and blue; from the RGB color model) values from the central part (10×10 pixels) of the color chart and urinalysis test strip.
4. Convert to the L*a*b* color space.
5. Calculate the color difference between the color chart and the urine test strip using the CIEDE2000 color difference formula (color difference calculation with each color patch of the same selected item).
6. Display the judgment results with the smallest color difference.

Figure 2. Process flow for comparing the color of urine test strips and color charts. RGB: red, green, and blue.



Filming Environment and Procedures

The color was evaluated using the following procedure in the same environment using a photography box (PU5025B):

1. Turn on the shooting box. Leave for 30 minutes to stabilize the lighting.
2. Start shooting 30 minutes after turning on the power.
3. Install the smartphone with the camera activated on top of the shooting box. Set the camera color temperature to sunlight, which is the same as that of the lighting of the shooting box.
4. Soak the urine test strip completely in the urine (approximately 1 or 2 seconds).
5. Lightly touch the side of the test paper with the tissue paper to remove excess urine on the test paper.
6. Place the test paper in a shooting box.
7. Take a picture 50 seconds after removing it from the urine sample (ie, when the paper is removed from the Spitz).
8. Judge the photographed images using an automatic urine dipstick colorimetric program.
9. Acquire age, sex, clinical department, and urinalysis results from electronic medical records.
10. Match the results with those of the automatic urine analyzer.

Accelerated KAZE

In this study, AKAZE, which is available in the OpenCV library, was used in image matching to reduce the effects of misalignment during shooting. AKAZE is an accelerated version of the algorithm used in KAZE [9], which is a 2D feature detection and description algorithm in nonlinear scale space developed by Alcantarilla et al [10]. This feature extraction algorithm is robust to changes in scaling, rotation, and lighting, and is similar to scale-invariant feature transform (SIFT). The KAZE algorithm is an improved version of SIFT and uses a nonlinear diffusion filter instead of a Gaussian filter to extract features while preserving the original image features and edges. SIFT uses a Gaussian filter to detect features. The center pixel of interest to be processed has the largest weight, and the outer pixels have smaller weights. This allows for images to be blurred while retaining more information around the pixels.

Judging Method for Urine Test Strips

The following methods are used for judging:

1. Approximate selection method: select the color on the color chart that is closest to the color of the test strip.

XSL•FO
RenderX

2. Rounding down method: if the color of the test strip does not reach the color on the color chart, judge it as a color with a lower density.
3. Rounding up method: if the color of the test strip is even slightly darker than that of the color chart, judge it as a color with a higher density.

Currently, there is no standard judgment method, but the Urinary Protein Measurement Committee has reported that "The approximate selection method seems to be theoretically correct, but it is prone to differences due to lighting and differences in each person's vision" [11]. Therefore, if the effects of lighting and individual differences can be reduced using this program, the "approximate selection method" may be suitable as an evaluation method. In this study, the "approximate selection method" was used.

Statistical Analysis

The rate of agreement was calculated using the results of the automatic urine analyzer. A concordance rate of ±1 was calculated as the concordance rate within 1 rank before and after the matching item. IBM SPSS Statistics 28th edition was used for the receiver operating characteristic (ROC) curve and area under the curve (AUC).

Ethical Considerations

This research was reviewed and approved by the ethics committee of the Hamamatsu University School of Medicine (approval number: 21 - 239). This study used existing samples in addition to existing information. However, as the samples were submitted to outpatient and other clinics, it was difficult to obtain direct consent. As such, information about this study, including the purpose of use of the samples and information, was made public on the Hamamatsu University School of Medicine's website and posted on notices within the hospital. Research subjects were given an opt-out form via information disclosure document and were guaranteed the opportunity to refuse participation in the study.

The samples and information provided, including personal information, were managed responsibly by the principal investigator (KI) and were anonymized such that specific individuals could be identified using a correspondence table or other means. This study collected only existing samples and information accumulated in the course of normal medical treatment and care, and there were no adverse events or malfunctions caused by the implementation of the study. This study targeted only the collection of existing samples and information accumulated during routine medical treatment and care, and there was no possibility of health damage occurring as a result of the study, so compensation for health damage and insurance enrollment did not apply.

Results

Agreement Rates and ±1 Rank Agreement Between Automated Urine Analyzers and Automated Urine Test Paper Colorimetric Programs

In this study, the US-3500, which is capable of qualitative evaluation using urine test strips, and the LABOSPECT 006, which is capable of quantitative evaluation of creatinine and albumin, were used. The results of the automated urine analyzers (US-3500 and LABOSPECT 006) were referred to the values recorded in the electronic medical records of the cooperating hospitals in the study. Table 2 shows the concordance rate of the automated urine analyzer and automated urine test strip colorimetric program, a concordance rate of ±1 rank, and the number of specimens. Urobilinogen, occult blood, protein, glucose, ketone bodies, bilirubin, nitrite, specific gravity, and pH were evaluated for concordance in 1184 samples. For white blood cells, 1182 specimens were evaluated for concordance. The amount of data obtained for creatinine and albumin is small due to the small number of daily test requests at the hospitals cooperating in the study. Therefore, the concordance rate was evaluated using 214 specimens for creatine and 42 specimens for albumin, which was in line with the data obtained from the automatic urine analyzer. The concordance rate for each item was more than 60%, except for specific gravity. The mean concordance for the 12 items was 78.6%. The matching rate of the ±1 rank for each item exceeded 90% for items other than specific gravity. The mean ±1 rank agreement for the 12 items was 95.4%. Multimedia Appendix 1 shows the concordance rates for each inspection item. The results indicate that an automated urine test paper colorimetric program could be applied to screening within ±1 rank.

XSL•FO
RenderX

Table . The concordance and ±1 rank match rate of the output results of the automated urine analyzer and the automated urine test strip colorimetric program and the number of samples obtained[a].

Item	Concordance rate, %	±1 rank match rate, %	Number of samples, n
Protein	83.1	100	1184
Glucose	94.8	100	1184
Urobilinogen	86	99.9	1184
Bilirubin	72.4	99.4	1184
Ketone	98	100	1184
Specific gravity	14.4	49.7	1184
Occult blood	87.8	97.8	1184
pH	80.7	100	1184
Leukocytes	86.6	99.1	1182
Nitrite	98.5	100	1184
Creatinine	64	98.6	214
Albumin	76.2	100	42
Average of 12 items	78.6	95.4	N/A[b]

[a]The samples were randomly selected from samples handled in the testing department of a hospital in Hamamatsu City from June 2022 to January 2023.
[b]N/A: not applicable.

Sensitivity and Specificity for Albumin

Trace albumin is a useful biomarker for early detection of nephropathy in patients with diabetes. Therefore, if trace albumin, which is important in determining the second stage of nephropathy, can be detected, it may lead to early detection of the progression of nephropathy. To confirm the applicability of our automated urine test paper colorimetric program for screening trace albuminuria, we calculated the sensitivity and specificity of the program for albumin. Since this program is a qualitative evaluation using urine test strips, the albumin value cannot be corrected for creatinine. Therefore, only the albumin value in urine was used for validation. Table 3 shows a cross-table for calculating the sensitivity and specificity of albumin. A diagnostic factor for diabetic nephropathy is microalbuminuria, which ranges from 30 to 299 mg/L. The cross-table was created assuming that patients with albumin levels of ≥30 and <30 were classified as having and not having the disease. Program judgment results show the program's output; microalbumin (urine automatic analysis data) shows the results recorded by LABOSPECT 006 in the electronic medical records of the collaborating hospitals. Cross-tab calculations yielded a sensitivity and specificity of 100% and 58.6%, respectively. Figure 3 shows the ROC curves. The AUC was 0.907, which is very favorable. Therefore, the results indicate that this program could be applied to screening for stage 2 nephropathy in diabetic nephropathy.

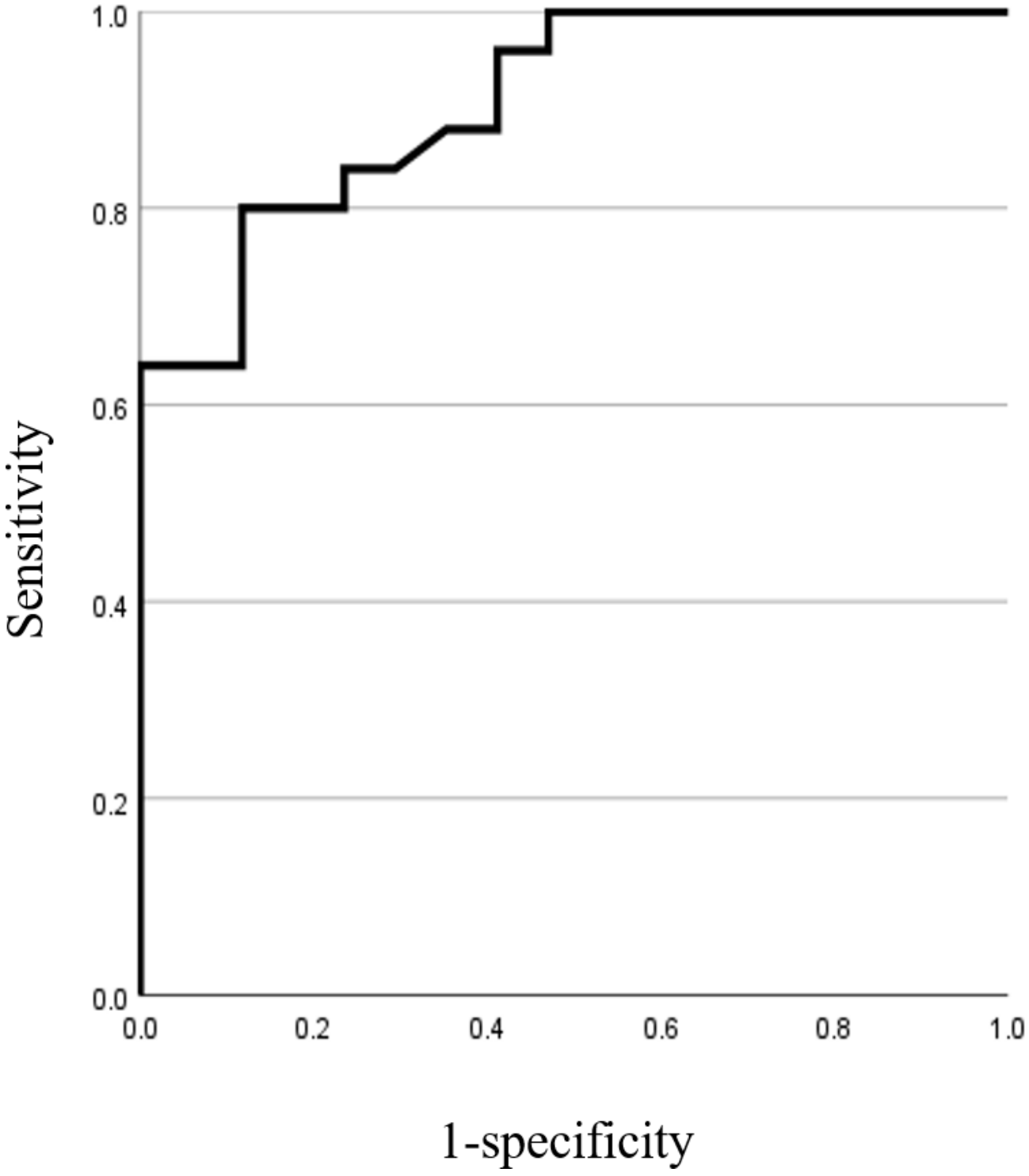
Table . Albumin sensitivity and specificity[a].

	Microalbumin (urine automatic analysis data)		Total
	With disease (albumin ≥30 mg/L)	With no disease (albumin <30 mg/L)	
Program judgment result			
Positive (albumin ≥30 mg/L)	13	12	25
Negative (albumin <30 mg/L)	0	17	17
Total	13	29	

[a]Patients with albumin levels of ≥30 and <30 were classified as those with diabetes and with no diabetes, respectively. Sensitivity and specificity were calculated by creating a cross-tabulation using the output of the automated urine analyzer and the automated urine test strip colorimetric program. From June 2022 to January 2023, 42 urine specimens were extracted from specimens handled by the laboratory department of a hospital in Hamamatsu City.

XSL•FO
RenderX

Figure 3. Receiver operating characteristic curve showing sensitivity and specificity of albumin. Patients with albumin levels of ≥30 were classified as those with diabetes, and those with albumin levels <30 were classified as patients with no diabetes. Sensitivity and specificity were calculated by creating a cross-table. From June 2022 to January 2023, 42 urine specimens were extracted from specimens handled by the laboratory department of a hospital in Hamamatsu City.



Discussion

Principal Results

This study used real patient urine specimens to investigate the agreement between an automated urine test strip colorimetric program and an automated urine analyzer. The concordance rate for each judgment item and the sensitivity and specificity of albumin were also evaluated. The concordance rate of the ±1 rank was 90% or more; thus, the concordance rate could be improved by devising the judgment threshold and judgment algorithm of the program.

The concordance rate of specific gravity obtained with this program was very low and was consistent with previous studies [12]. In the visual judgment, the colors of the reference color chart and test paper for urinalysis differed. Determining specific gravity with a low matching rate by simple color comparison

XSL•FO
RenderX

was challenging; therefore, a determination algorithm should be considered.

The detailed results for each item showed many negative and few positive judgments. However, despite the small number of positive tests, many items matched, supporting the effectiveness of the proposed method.

The proposed method determined the matching rate using a program that identifies the color closest to the reference color chart. Since the proposed method used the "approximate selection method," some items may present a low matching rate. Therefore, other evaluation methods should be considered in the future.

Previous reports have included many studies evaluating the reliability of programs using real urine samples, artificial urine, and reagents [13-17]. In a report using real urine samples, urine testing was performed using a smartphone app developed by Healthy.io, and its acceptability and feasibility were evaluated [16,17]. The satisfaction rate was very high among patients and caregivers, and the app could greatly enhance patient-centered care [16]. The potential of remote digital urine testing has also been reported [17].

Rahmat et al [18] validated the scanner color comparison using real urine specimens and reported an accuracy of 95.45% using a high-resolution scanner. The low-resolution scanner reported 83%, comparable with our automated urine test strip colorimetric program. Africa and Velasco [19] also verified the accuracy of color comparison using 45 samples from an accredited urinalysis laboratory and reported an accuracy of 96.519%. However, besides evaluating 10 items (ie, protein, glucose, ketone, specific gravity, occult blood, leukocytes, nitrite, urobilinogen, bilirubin, and pH), these 2 studies did not provide the breakdown of the concordance rate (ie, negative and positive breakdown and item-by-item classification). Urinary albumin and creatinine levels, necessary for determining diabetic nephropathy, have not yet been evaluated in the literature. Determining albumin and creatinine may help screen for microalbuminuria, enabling intervention before diabetic nephropathy becomes severe.

The sensitivity and specificity of this method for albumin detection were 100% and 58.6%, respectively. These suggest potentially high and low rates of false positives and negatives, respectively, allowing for people with diseases to be identified. Furthermore, the AUC derived from the ROC curve was 0.907, indicating satisfactory results. Therefore, the proposed method may be used as initial screening.

The proposed method may also support point-of-care (POC) testing at home to prevent the progression of diabetic nephropathy. Thakur et al [15] reported a method for albumin screening using smartphones and urine test strips. The study reported a 92% accuracy rate, validated only with reagents. In contrast, this study used actual urine specimens, and excellent screening results were obtained with a concordance rate of 76.2%, sensitivity of 100%, and specificity of 58.6%, relative to the results of an automated urine analyzer. Due to the COVID-19 pandemic, many people refrained from seeing doctors. As such, the proposed method is expected to be particularly useful for POC tests for home health management. Therefore, we are currently developing an application that uses the proposed system.

The proposed automatic urine test strip colorimetric program focuses only on color differences but does not require a large amount of data from machine learning. In addition, if standard color chart data from each manufacturer become available, urinalysis test strips from various manufacturers worldwide may be used without requiring calibration. In the future, the accuracy of determination may be improved by using data obtained from specimens, adjusting the threshold for determination, and examining determination methods.

Limitations

Despite its contributions, this study has limitations. The results do not apply to all smartphones because verification was performed using only 1 smartphone model. In the future, the program should be tested on other operating systems, smartphones, and cameras. In addition, owing to lighting limitations due to the shooting box, development did not progress sufficiently enough that judgment could be formulated with just a smartphone. This system is not suitable as a POC test; therefore, future investigations should verify whether similar results can be obtained using the flash function of a smartphone. In the future, the differences between multiple models and methods that do not use shooting boxes should be considered.

In addition, this study used urine test strips to evaluate the agreement between the white blood cell count obtained by an automated urine analyzer and the program developed in this study. Normally, the agreement with sediment analysis is verified. Future studies should compare the results of the formulated program with those from sediment analysis to compare the accuracy.

When urine samples were randomly collected, many samples did not have albumin analysis in the doctor's orders. Therefore, the albumin results from the automated urine analyzer were not recorded in the electronic medical record, and the match rate could not be calculated for many samples. The study protocol should be adjusted so that an adequate sample of albumin tests can be collected.

Normally, creatinine should be corrected in urinalysis results. However, given that the coincidence rate of creatinine was not high, only albumin was highlighted. In the future, accuracy can be improved by evaluating the albumin-creatinine ratio.

The analyzers used in this study (US-3500 and LABOSPECT 006) can evaluate reflectance, allowing for a more accurate evaluation of the consistency rate of urine test strips. However, this study used data stored in electronic medical records, and reflectance data could not be obtained. In the future, reflectance data should also be analyzed.

In this study, the agreement rate by qualitative evaluation was verified using urine test strips. However, sensitivity is low when using urine test strips. Therefore, in the future, reliability should be verified by comparing data that can be quantitatively measured, such as immunoturbidimetric methods. Finally, to

XSL•FO
RenderX

avoid over- or underestimating clinical scenarios, sample collection periods should span several years.

Conclusions

This study used real patient urine specimens to investigate the agreement between an automated urine test strip colorimetric program and an automated urine analyzer. It also investigated the concordance rate for each judgment item and the sensitivity and specificity of the program in evaluating albumin. The average concordance rate and rate of ±1 rank for the mean of the 12 inspected items were 78.6% and 95.4%, respectively. These results clarified that colorimetric analysis in a fixed imaging environment could be used to determine urine test strips without calibration.

Urinary albumin determination, essential for determining the stage of diabetic nephropathy, resulted in a sensitivity of 100% and a specificity of 58.6%. In addition, the AUC derived from the ROC curve was 0.907, indicating good results. The proposed method may be used for the screening of microalbuminuria to help prevent the exacerbation of diabetic nephropathy.

Acknowledgments

The authors thank Dr Masato Maekawa, Mr Keita Yamashita, Ms Misaki Hidaka, and the laboratory staff of Hamamatsu University Hospital for providing urine specimens and coordinating schedules in the midst of their busy work schedules. Professor Katsumasa Nakamura provided oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team. Dr Daisuke Tsuriya and Dr Takuya Hashimoto were responsible for the management and coordination of the planning and execution of research activities. This work was supported by JSPS KAKENHI (grant 24K15836) and Terumo Life Science Foundation (21-Ⅲ7009), Hamamatsu University School of Medicine grant-in-aid. The funder played no role in the study design, data collection, data analysis, and interpretation, or manuscript writing.

Disclaimer

No artificial intelligence, language modeling, machine learning, or similar technologies have been used to create or assist in the creation or editing of any content (eg, text, tables, figures, and video) included in this submission.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary Table
[DOCX File, 43 KB - formative_v9i1e62772_app1.docx]

References

1. Coppens A, Speeckaert M, Delanghe J. The pre-analytical challenges of routine urinalysis. Acta Clin Belg 2010;65(3):182-189. [doi: 10.1179/acb.2010.038] [Medline: 20669786]
2. Devillé WLJM, Yzermans JC, van Duijn NP, Bezemer PD, van der Windt DAWM, Bouter LM. The urine dipstick test useful to rule out infections. A meta-analysis of the accuracy. BMC Urol 2004 Jun 2;4:1-14. [doi: 10.1186/1471-2490-4-4] [Medline: 15175113]
3. Simerville JA, Maxted WC, Pahira JJ. Urinalysis: a comprehensive review. Am Fam Physician 2005 Mar 15;71(6):1153-1162. [Medline: 15791892]
4. Gong MM, Sinton D. Turning the page: advancing paper-based microfluidics for broad diagnostic application. Chem Rev 2017 Jun 28;117(12):8447-8480. [doi: 10.1021/acs.chemrev.7b00024] [Medline: 28627178]
5. Reboud J, Xu G, Garrett A, et al. Paper-based microfluidics for DNA diagnostics of malaria in low resource underserved rural communities. Proc Natl Acad Sci U S A 2019 Mar 12;116(11):4834-4842. [doi: 10.1073/pnas.1812296116] [Medline: 30782834]
6. Hong JI, Chang BY. Development of the smartphone-based colorimetry for multi-analyte sensing arrays. Lab Chip 2014 May 21;14(10):1725-1732. [doi: 10.1039/c3lc51451j] [Medline: 24671456]
7. Memişoğulları R, Yüksel H, Yıldırım HA, Yavuz Ö. Performance characteristics of dipstick and microscopic urinalysis for diagnosis of urinary tract infection. Eur J Gen Med 2010;7(2):174-178. [doi: 10.29333/ejgm/82846]
8. Inagaki K. Comparison of judgment results between the color comparison method of urine analysis test papers using the CIEDE2000 color-difference formula and the automatic urine analyzer. J Nurs Sci Eng 2021;8:170-176. [doi: 10.24462/jnse.8.0_170]

XSL•FO
RenderX

9. Alcantarilla PF, Nuevo J, Bartoli A. Fast explicit diffusion for accelerated features in nonlinear scale spaces. Presented at: British Machine Vision Conference 2013; Sep 9-13, 2013; Bristol, United Kingdom. [doi: 10.5244/C.27.13]
10. Alcantarilla PF, Bartoli A, Davison AJ. KAZE features. In: Fitzgibbon A, Lazebnik S, Perona P, Sato Y, Schmid C, editors. Computer Vision - ECCV Lecture Notes in Computer Science: Springer; 2012, Vol. 7577:214-227. [doi: 10.1007/978-3-642-33783-3_16]
11. Urinary Protein Measurement Committee. Report of the committee on renal function (GFR) and urinary protein measurement of the Japanese Society of Nephrology [Article in Japanese]. J Jpn Soc Nephrol 2001;43(1):1-19. [doi: 10.14842/jpnjnephrol1959.43.1]
12. Inagaki K, Nakamura K. Verification of individual differences in visual judgment of urine test strips and proposal of objective evaluation method. Lab Med Int 2024;3(2):42-49. [doi: 10.51041/lmi.3.2_42]
13. Kim SC, Cho YS. Predictive system implementation to improve the accuracy of urine self-diagnosis with smartphones: application of a confusion matrix-based learning model through RGB semiquantitative analysis. Sensors (Basel) 2022 Jul 21;22(14):5445. [doi: 10.3390/s22145445] [Medline: 35891125]
14. Ra M, Muhammad MS, Lim C, Han S, Jung C, Kim WY. Smartphone-based point-of-care urinalysis under variable illumination. IEEE J Transl Eng Health Med 2018;6:2800111. [doi: 10.1109/JTEHM.2017.2765631] [Medline: 29333352]
15. Thakur R, Maheshwari P, Datta SK, Dubey SK, Shakher C. Machine learning-based rapid diagnostic-test reader for albuminuria using smartphone. IEEE Sensors J 2021;21(13):14011-14026. [doi: 10.1109/JSEN.2020.3034904]
16. Erez DL, Derwick H, Furth S, et al. Dipping at home: is it better, easier, and more convenient? A feasibility and acceptability study of a novel home urinalysis using a smartphone application. Pediatr Nephrol 2023 Jan;38(1):139-143. [doi: 10.1007/s00467-022-05556-8] [Medline: 35445981]
17. Stauss M, Dhaygude A, Ponnusamy A, Myers M, Woywodt A. Remote digital urinalysis with smartphone technology as part of remote management of glomerular disease during the SARS-CoV-2 virus pandemic: single-centre experience in 25 patients. Clin Kidney J 2022 May;15(5):903-911. [doi: 10.1093/ckj/sfab286] [Medline: 35498887]
18. Rahmat RF, Muchtar MA, Adnan S, et al. Automated color classification of urine dipstick image in urine examination. J Phys Conf Ser 2018 Mar 1;978(1):012008. [doi: 10.1088/1742-6596/978/1/012008]
19. Africa AD, Velasco J. Development of a urine strip analyzer using artificial neural network using an android phone. ARPN J Eng Appl Sci 2017;12(6):1706-1712.

Abbreviations

AKAZE: Accelerated KAZE
AUC: area under the curve
POC: point-of-care
RGB: red, green, and blue
ROC: receiver operating characteristic
SIFT: scale-invariant feature transform

Edited by A Mavragani; submitted 10.06.24; peer-reviewed by AM Matos, J Favresse; revised version received 12.11.24; accepted 13.11.24; published 14.01.25.

Please cite as:
Inagaki K, Tsuriya D, Hashimoto T, Nakamura K
Verification of the Reliability of an Automated Urine Test Strip Colorimetric Program Using Colorimetric Analysis: Survey Study
JMIR Form Res 2025;9:e62772
URL: https://formative.jmir.org/2025/1/e62772
doi:10.2196/62772

© Keigo Inagaki, Daisuke Tsuriya, Takuya Hashimoto, Katsumasa Nakamura. Originally published in JMIR Formative Research (https://formative.jmir.org), 14.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Determinants of Dropping Out of Remote Patient-Reported Outcome–Based Follow-Up Among Patients With Epilepsy: Prospective Cohort Study

Sofie Bech Vestergaard[1,2*], MHSc; Mette Roost[3*], MHSc; David Høyrup Christiansen[2,4,5], MHSc, PhD; Liv Marit Valen Schougaard[2,3], MHSc, PhD

[1]Department of Rheumatology, Aarhus University Hospital, Aarhus N, Denmark

[2]Department of Clinical Medicine, Aarhus University, Aarhus N, Denmark

[3]AmbuFlex, Centre for Patient-Reported Outcomes, Gødstrup Hospital, Herning, Denmark

[4]Centre for Research in Health and Nursing, Research, Regional Hospital Central Jutland, Viborg, Denmark

[5]University Clinic, Centre of Elective Surgery, Regional Hospital Silkeborg, Silkeborg, Denmark

[*]these authors contributed equally

Corresponding Author:
Sofie Bech Vestergaard, MHSc
Department of Rheumatology, Aarhus University Hospital, Aarhus N, Denmark

Abstract

Background: The use of patient-reported outcome (PRO) measures is an emerging field in health care. In the Central Denmark Region, epilepsy outpatients can participate in remote PRO-based follow-up by completing a questionnaire at home instead of attending a traditional outpatient appointment. This approach aims to encourage patient engagement and is used in approximately half of all epilepsy outpatient consultations. However, dropout in this type of follow-up is a challenging issue.

Objective: This study aimed to examine the association between potential self-reported determinants and dropout in remote PRO-based follow-up for patients with epilepsy.

Methods: This prospective cohort study (n=2282) explored the association between dropout in remote PRO-based follow-up for patients with epilepsy and 9 potential determinants covering 3 domains: health-related self-management, general and mental health status, and patient satisfaction. The associations were examined using multiple logistic regression analyses with adjustment for sex, age, education, and cohabitation.

Results: A total of 770 patients (33.7%) dropped out of remote PRO-based follow-up over 5 years. Statistically significant associations were identified between all potential determinants and dropouts in PRO-based follow-up. Patients with low social support had an odds ratio of 2.20 (95% CI 1.38-3.50) for dropout. Patients with poor health ratings had an odds ratio of 2.17 (95% CI 1.65-2.85) for dropout. Similar estimates were identified for the remaining determinants in question.

Conclusions: Patients with reduced self-management, poor health status, and low patient satisfaction had higher odds of dropout in remote PRO-based follow-up. However, further research is needed to determine the reasons for dropout.

(JMIR Form Res 2025;9:e58258) doi:10.2196/58258

KEYWORDS

patient-reported outcome measures; dropouts; digital solutions; outpatient care; epilepsy; seizure disorder; neurological condition; cohort study; health care; Denmark; self-reported; self-management; mental health; patient satisfaction; logistic regression; social support

Introduction

Chronic conditions are a global burden that is increasing in most countries. In 2019, more than 65% of the population in Denmark had one or more chronic conditions [1]. To manage the worldwide development of these chronic conditions, it is crucial that health care resources are being optimally used. The application of telehealth initiatives could be a cost-effective solution for providing high-quality health care [2]. Over the past 50 years, the availability and use of telehealth have expanded and include various tools to manage health care [3]. The World Health Organization (WHO) defines telehealth as *"the delivery of health care services, where distance is a critical factor, by all health care professionals using information and communication technologies for the exchange of valid information for diagnosis, treatment, and prevention of disease and injuries, research and evaluation, and for the continuing education of health care providers, all in the interests of*

XSL•FO
RenderX

advancing the health of individuals and their communities" [4]. After the WHO declared COVID-19 as a pandemic in March 2020, the use of telemedicine has rapidly increased [5,6]. The pandemic forced clinicians to rethink the course of treatment for their patients in various clinical settings to be effectively managed from a distance [6]. One way of using telehealth could be using patient-reported outcomes (PROs) in remote symptom monitoring. This could transform the delivery of the traditional treatment for patients with chronic conditions with prescheduled appointments by optimizing the use of resources and ensuring patient-centered care [7-10].

The use of PROs was originally developed for utilization in research but has, over time, evolved to be utilized at a patient level in clinical practice [11,12]. The use of PROs in clinical practice has been documented to improve patient survival, provide individualized care, and ensure efficient usage of health care resources [13]. In Denmark, a generic clinical PRO system for data collection in clinical practice, called AmbuFlex, offers tailor-made PRO solutions [8]. The overall purpose is to improve patient involvement and quality of treatment while optimizing resource allocation [8,9]. As of September 2023, AmbuFlex has developed 65 PRO solutions for various chronic and malignant conditions in remote outpatient care [8]. The use of PROs in remote outpatient follow-up provides real-time symptom monitoring, early detection of health issues, and rapid clinical intervention [11]. This type of follow-up requires the patient to be willing and capable of taking more responsibility as well as playing an active role in monitoring and identifying symptoms of disease [14]. Patient experiences with this course of treatment are diverse. For patients with epilepsy, it has been documented to promote health-related self-management [15]. Research on patients with rheumatoid arthritis has shown that some patients find remote PRO-based follow-up to be a flexible, time- and resource-saving solution. Others worry about the absence of face-to-face communication with a clinician and find the responsibility to cause insecurity [14].

Dropout and nonuse of telehealth solutions is a common reported challenge [16,17]. A scoping review highlights that the reasons for the nonuse of digital PRO solutions are diverse and, among others, cover the ability to use PRO, engagement, emotional issues, lack of time, and technical barriers [18]. Several contributing factors such as sex [18-21], age [19,22,23], education [18,22], and cohabitation [24-27] have been found to influence the dropout rate in digital solutions. Previous research on PRO-based follow-up for patients with epilepsy has shown that socioeconomically advantaged patients with a high level of health literacy, self-efficacy, patient activation, good well-being, and general health are more likely to be referred to PRO-based follow-up than socioeconomically disadvantaged patients [24]. Whether the same determinants apply to dropout has not been investigated. With the limited amount of literature covering dropout in remote PRO-based follow-up, further research is needed. Thus, this study aimed to investigate potential determinants for dropout in remote PRO-based follow-up among patients with epilepsy. We examined the association between dropout in PRO-based follow-up and determinants from the following three domains: health-related self-management, general and mental health status, and patient satisfaction. We hypothesized that reduced health-related self-management, poor general and mental health status, and low patient satisfaction were associated with an increased risk of dropout in PRO-based follow-up.

Methods

Remote PRO-Based Follow-Up for Patients With Epilepsy

PRO-based follow-up was implemented for patients with epilepsy at 3 neurological departments in the Central Denmark Region in 2012. In PRO-based follow-up, patients complete a questionnaire at home instead of having prescheduled appointments at the outpatient clinic. To attend PRO-based follow-up, patients are individually referred by a health professional. Individual referral requires that the patient must be 15 years of age or older, have no cognitive impairments, and be capable of reading and writing in Danish [28]. The questionnaire is sent at a predefined time interval of 3, 6, 12, or 24 months either electronically through a secure electronic platform or as a paper version [28].

The epilepsy questionnaire includes 47 items and deals with the frequency and intensity of seizures, medical treatment, and the patient's well-being. The AmbuFlex system, based on an automated decision algorithm, can assess whether the patient needs clinical attention [8]. To identify patients in need of clinical attention, all response options in the questionnaire are categorized as red, yellow, or green. Red indicates that the patient needs or wants personal contact with the clinic. Yellow indicates that the patient might need personal contact. If all the patients' answers are categorized as green, the system automatically sends a new questionnaire within the predefined period. All answers that indicate a need for clinical attention are reviewed by a nurse within 14 days. The patients' questionnaire responses were displayed in a graphical overview available in the patients' electronic health records. Where clinical attention is needed, the questionnaire data are used to support patient-clinician communication and clinical decision-making [28]. If a patient wants to discontinue or a clinician decides to end a PRO-based follow-up treatment, the clinician must actively register the patient as "deregistered" in the AmbuFlex system. Deceased patients are automatically deregistered in the system. PRO-based follow-ups are offered to approximately 50% of the entire patient population at the 3 neurological departments and have been used since 2012 [24]. In January 2016, approximately 3000 epilepsy outpatients attended the PRO-based follow-up program at the 3 departments.

Study Population and Setting

We conducted a prospective cohort study among patients with epilepsy who attended PRO-based follow-ups at 3 neurology departments in the Central Denmark Region. During the period of January 1st to December 31st, 2016, patients were invited to answer a research questionnaire in addition to their usual scheduled epilepsy questionnaire from the outpatient clinic. The research questionnaire contained information about aspects related to self-management, general and mental health status, and patient satisfaction. Patients could respond to either a

XSL•FO
RenderX

web-based or paper version of the questionnaire. Nonresponders received up to 3 reminders [29]. Clinicians assessed the usual epilepsy questionnaire according to their normal routine, but they were blinded to the research questionnaire.

Outcome

The outcome was the dropout status in remote PRO-based follow-up in February 2021. This information was retrieved from the AmbuFlex system. As a clinician actively has to register the patient as "deregistered" in the AmbuFlex system, we defined dropouts in this study as "patients who, after using PRO-based follow-up, were deregistered by a clinician." The reason for the dropout was not registered. If the patients were deregistered due to death, this information was shown in the AmbuFlex system. Patients who died during follow-up were excluded from the study.

Potential Determinants

Self-Management

We assessed patient's health-related self-management using 3 different constructs: health literacy, self-efficacy, and patient activation. Health literacy was measured using part of the multi-dimensional Health Literacy Questionnaire (HLQ) [30]. The questionnaire includes 9 subscales covering different areas of health literacy to assess strengths and challenges. The HLQ has well-documented psychometric properties considering the original questionnaire and the Danish translation, with strong content and construct validity as well as high reliability [30,31]. In this study, the following 3 subscales were used: subscale #4, "Social support for health"; subscale #6, "Ability to actively engage with health care providers," and subscale #9, "Understanding health information well enough to know what to do." Subscale #4 includes 5 items with a 4-point ordinal response option ranging from, 1, "strongly disagree," 2, "disagree," 3, "agree," to 4, "strongly agree." Subscales #6 and #9 both include 5 items as well a 5-point ordinal response option ranging from, 1, "cannot do," 2, "very difficult," 3, "quite difficult," 4, "quite easy," to 5, "very easy." For all 3 subscales, a lower score indicates lower health literacy. An average score was calculated for each of the 3 subscales based on a guide provided when acquiring the license. If 1 item was missing, the mean score of the existing items was used to estimate the score. If 2 or more items were missing, the average score was not calculated. As the variables did not exhibit a log-linear relationship, the score was dichotomized into high and low health literacy. For subscale #4, high health literacy was a score >2, and low health literacy was a score ≤2. For subscales #6 and #9, high health literacy was a score >3, and low health literacy was a score ≤3.

Self-efficacy was measured using the General Self-efficacy Scale (GSES) [32]. The psychometric properties have been assessed across many nations and populations, showing a strong construct validity and a high reliability [32,33]. The GSES includes 10 questions regarding one's belief in the ability to cope with stress and challenges. Each question has a 4-point ordinal response option ranging from 1, "not at all true," 2, "hardly true," 3, "moderately true," to 4, "exactly true." The total GSES score ranges from 10 to 40. A lower score indicates a lower degree of self-efficacy [32]. If 1 or more items were missing, the total score was not calculated. The score was included as a continuous variable.

Patient activation was assessed using part of the Patient Activation Measure (PAM) [34]. The original questionnaire includes 13 items that assessed activation in illness or health conditions [35]. In this study, 2 single items were used in a modified form: "I am confident that I can tell when I need to get outpatient care" and "I am confident I can figure out solutions when new situations or problems arise with my health condition." Both items had 4 response options "strongly disagree," "disagree," "agree," and "strongly agree." Both items were dichotomized into strongly disagree/disagree and agree/strongly agree.

General and Mental Health Status

Mental well-being was measured using the WHO-5 Well-being Index (WHO-5) [36]. The WHO-5 includes 5 positively worded items focusing on mental well-being over the previous 2 weeks [36]. The questionnaire has well-documented psychometric properties for various diseases, including epilepsy [28,36,37]. Each item has a 6-point ordinal response option ranging from 5, "All the time," 4, "most of the time," 3, "more than half of the time," 2, "less than half of the time," 1, "some of the time," to 0, "at no time." To calculate the total score, the point for each item was added and afterward multiplied by 4, resulting in a total score between 0 and 100. A score of 100 indicates the best imaginable well-being [36]. If 1 or more items were missing, the total score was not calculated. The score was included as a continuous variable.

The patients' general health status was assessed using a single item from Short Form-36 (SF-36) [38]. The SF-36 covers 8 different domains related to health-related quality of life [39]. It has been documented that a single item from the questionnaire can predict general health [40]. In this study, the single item "In general, would you say your health is: excellent, very good, good, fair, or poor?" was therefore used. The 5 response options were categorized into "Excellent/very good," "good," and "fair/poor."

Patient Satisfaction

The domain of patient satisfaction included patient involvement, confidence, comfort, and satisfaction. Information regarding these determinants was collected using questions inspired by the Danish Cancer Society's Barometer Survey [41] and modified to fit the epilepsy outpatient setting. Information about patient involvement, confidence, and safety was obtained via the following questions: "Do you feel sufficiently involved in your course of treatment?," "Are you confident that the epilepsy outpatient clinic contributes in the best possible way in your course of treatment?," and "Are you comfortable with the follow-up you are receiving?." These questions had five response options: "Yes, to a great extent," "Yes, to some extent," "To a minor extent," "No, not at all," or "Do not know." Information about satisfaction was obtained through: "Overall, how would you assess your treatment in the epilepsy outpatient clinic?" with the following 5 response options: "Very good," "Mostly good," "Mostly bad," "Very bad," or "Do not know,"

XSL•FO
RenderX

For all questions, the category "Do not know" was excluded in the analyses. Patient involvement, confidence, and safety were dichotomized into "Yes, to a great extent/Yes, to some extent" and "To a minor extent/not at all." Satisfaction was dichotomized into "Very good/Mostly good" and "Mostly bad/Very bad."

Statistical Analyses

A Wilcoxon rank-sum test was utilized to compare age and sex between responders and nonresponders. Characteristics for the study population were presented by outcome. Continuous data were all nonnormally distributed and presented as the median and IQR. Categorical and dichotomized data were presented as numbers (n) and percentage distribution (%). The association between the potential determinants and dropouts in PRO-based follow-up was analyzed through multiple logistic regression analysis. All the odds ratio (OR) estimates were tested at a 5% significance level and reported with 95% CIs. Variables were continuously included if the linearity requirements were met. The assumption on linearity was deemed fulfilled for self-efficacy and well-being. Sex, age, education, and cohabitation were included in the adjusted analyses. The confounder variables were defined a priori based on a systematic literature review [18-27]. For the patient satisfaction domain, most of the determinants were controlled only for sex and cohabitation, as some categories had a limited number of observations. Each potential determinant was examined separately as the determinants were assessed as interdependent. All analyses were performed using the available observations giving varying observations for crude and adjusted estimates. An examination of all missing data was conducted prior to this. Upon inspection, no systematic or significant missing data were identified. For the continuous variables, postestimation analyses were performed to nuance the significance of different scores and dropouts in PRO-based follow-up. The statistical analyses were performed using Stata/IC 16.1 (StataCorp LLC).

Ethical Considerations

The study was approved by the Danish Data Protection Agency (number 1-16-02-691-14). The Ethics Committee of the Central Denmark Region was consulted and decided that the study did not require their approval. According to Danish law, only studies involving human biological material require committee approval. Correspondence with the ethics committee is available from the authors upon request. Informed consent for participation in the study was obtained in accordance with guidelines from the Danish Data Protection Agency. Information was provided to patients in a letter along with the questionnaire. Patients were informed that responding to the questionnaire constituted active consent for their participation and the use of their data in the research project. They were also informed that they could withdraw their consent at any time. All data were stored and handled with confidentiality.

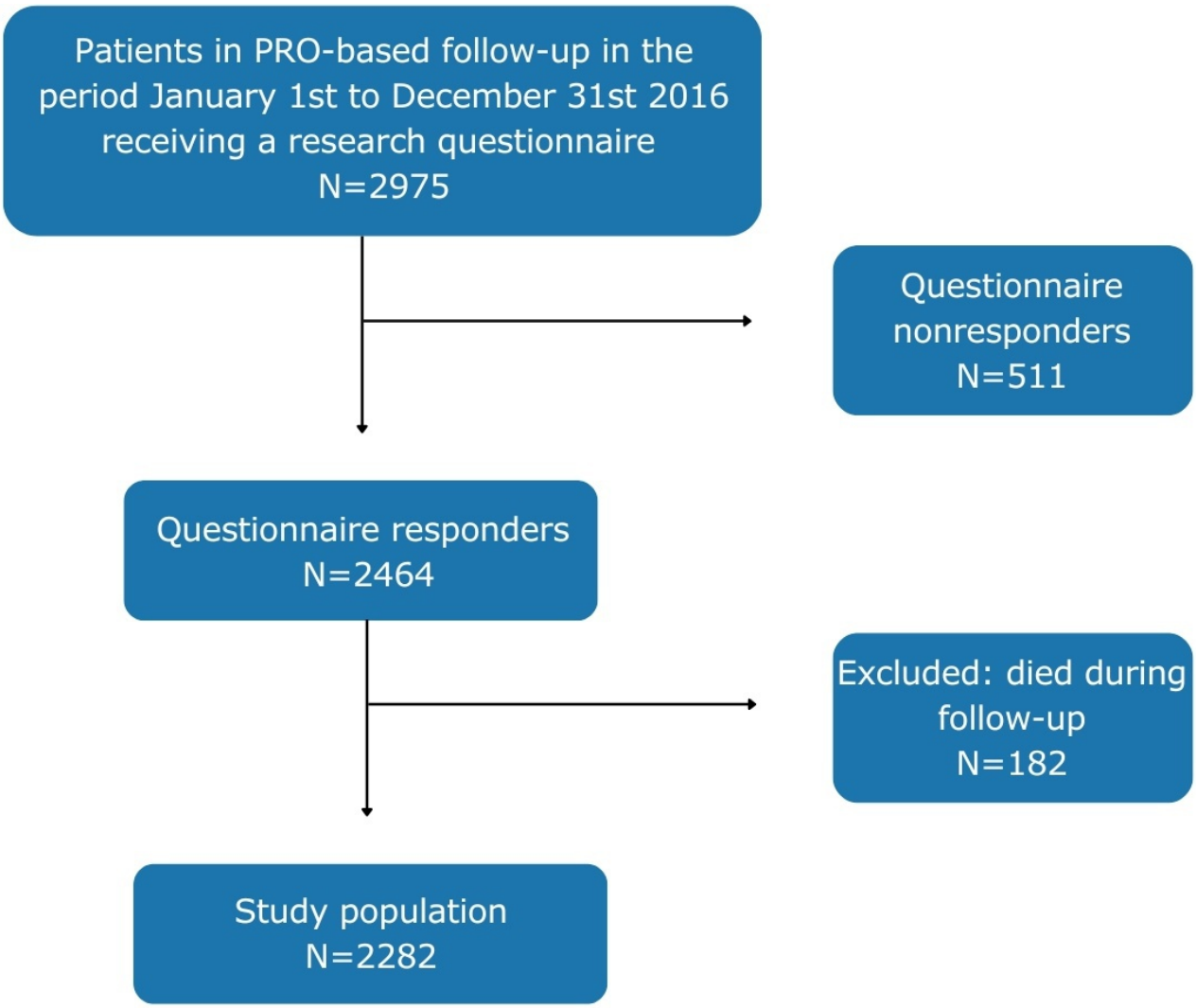
Results

Patient Characteristics

From January 1st to December 31st, 2016, a total of 2975 patients received a paper or web-based questionnaire. Of these, 2464 patients answered the questionnaire, resulting in a response rate of 83%. During the follow-up period, 182 patients died. The final study population included 2282 patients (Figure 1). The median age in the study population was 49 (IQR 29.7) years, 1125 (49.3%) were male, 982 (43%) had a low education, and 1645 (72.1%) were not living alone (Table 1). A total of 770 patients (33.7%) had dropped out of PRO-based follow-up in February 2021. No statistical differences were found between responders and nonresponders in sex distribution (P=.34) or age distribution (P=.70). Missing data ranged from 1.5% (general health) to 7.8% (general self-efficacy) (Table 1).

XSL•FO
RenderX

Figure 1. Flowchart of patients included in the study.



XSL•FO
RenderX

Table . Baseline characteristics and patient-reported potential determinants by status in patient-reported outcome (PRO)–based follow-up among patients with epilepsy in February 2021.

Characteristics	Total, n=2282	Active, n=1512	Dropout, n=770
Sex, n (%)			
Male	1125 (49.3)	757 (50.1)	368 (47.8)
Female	1157 (50.7)	755 (49.9)	402 (52.2)
Age, years, n (%)			
15 - 24	245 (10.7)	128 (8.5)	117 (15.2)
25 - 39	538 (23.6)	340 (22.5)	198 (25.7)
40 - 54	586 (25.7)	407 (26.9)	179 (23.3)
55 - 69	579 (25.4)	421 (27.8)	158 (20.5)
≥70	334 (14.6)	216 (14.3)	118 (15.3)
Education, n (%)			
None	512 (22.4)	317 (21.0)	195 (25.3)
Low	982 (43.0)	674 (44.6)	308 (40.0)
Medium/high	622 (27.3)	426 (28.2)	196 (25.5)
Missing	166 (7.3)	95 (6.3)	71 (9.2)
Cohabitation, n (%)			
Living alone	526 (23.1)	327 (21.6)	199 (25.8)
Not living alone	1645 (72.1)	1126 (74.5)	519 (67.4)
Missing	111 (4.9)	59 (3.9)	52 (6.8)
Health literacy #4[a], n (%)			
Low ≤2	77 (3.4)	37 (2.5)	40 (5.2)
High >2	2049 (89.8)	1395 (92.3)	654 (84.9)
Missing	156 (6.8)	80 (5.3)	76 (9.9)
Health literacy #6[b], n (%)			
Low ≤3	398 (17.4)	222 (14.7)	176 (22.9)
High >3	1728 (75.7)	1209 (80.0)	519 (67.4)
Missing	156 (6.8)	81 (5.4)	75 (9.7)
Health literacy #9[c], n (%)			
Low ≤3	323 (14.2)	172 (11.4)	151 (19.6)
High >3	1796 (78.7)	1253 (82.9)	543 (70.5)
Missing	163 (7.1)	87 (5.8)	76 (9.9)
Self-efficacy			
General Self-efficacy Scale score, median (IQR)	30 (25 - 33)	30 (26 - 34)	29 (24 - 32)
Missing, n (%)	177 (7.8)	91 (6.0)	86 (11.2)
PAM #5[d], n (%)			
Strongly disagree/disagree	221 (9.7)	118 (7.8)	103 (13.4)
Agree/strongly agree	1926 (84.4)	1323 (87.5)	603 (78.3)
Missing	135 (5.9)	71 (4.7)	64 (8.3)
PAM #12[e], n (%)			
Strongly disagree/disagree	342 (15.0)	193 (12.8)	149 (19.4)

XSL•FO
RenderX

Characteristics	Total, n=2282	Active, n=1512	Dropout, n=770
Agree/strongly agree	1797 (78.8)	1246 (82.4)	551 (71.6)
Missing	143 (6.3)	73 (4.8)	70 (9.1)
Mental well-being			
WHO-5 Well-being Index score, median (IQR)	76 (60 - 80)	76 (64 - 80)	72 (52 - 80)
Missing, n (%)	64 (2.8)	35 (2.3)	29 (3.8)
General health[f], n (%)			
Excellent/very good	1030 (45.1)	719 (47.6)	311 (40.4)
Good	886 (38.8)	597 (39.5)	289 (37.5)
Fair/poor	333 (14.6)	181 (12.0)	152 (19.7)
Missing	33 (1.5)	15 (1.0)	18 (2.3)
Patient involvement, n (%)			
A great extent/some extent	1820 (79.8)	1259 (83.3)	561 (72.9)
A lesser extent/not at all	189 (8.3)	102 (6.8)	87 (11.3)
Do not know	140 (6.1)	82 (5.2)	58 (7.5)
Missing	133 (5.8)	69 (4.6)	64 (8.3)
Confidence, n (%)			
A great extent/some extent	1945 (85.2)	1324 (87.6)	621 (80.7)
A minor extent/not at all	99 (4.3)	56 (3.7)	43 (5.6)
Do not know	116 (5.1)	68 (4.5)	48 (6.2)
Missing	122 (5.4)	64 (4.2)	58 (7.5)
Comfort, n (%)			
A great extent/some extent	1896 (83.1)	1307 (86.4)	589 (76.5)
A minor extent/not at all	108 (4.7)	56 (3.7)	52 (6.8)
Do not know	148 (6.5)	80 (5.3)	68 (8.8)
Missing	130 (5.7)	69 (4.6)	61 (7.9)
Satisfaction, n (%)			
Very good/mostly good	1863 (81.6)	1288 (85.2)	575 (74.7)
Mostly bad/very bad	93 (4.1)	49 (3.2)	44 (5.7)
Do not know	197 (8.6)	106 (7.0)	91 (11.8)
Missing	129 (5.7)	69 (4.6)	60 (7.8)

[a]Health Literacy Questionnaire subscale #4 “Social support for health.”

[b]Health Literacy Questionnaire subscale #6 “Ability to actively engage with health care providers.”

[c]Health Literacy Questionnaire subscale #9 “Understanding health information well enough to know what to do.”

[d]Patient Activation Measure item #5: “I am confident that I can tell when I need to get outpatient care.”

[e]Patient Activation Measure item #12: “I am confident I can figure out solutions when new situations or problems arise with my health condition.”

[f]Assessed using a single item from Short Form-36.

Associations Between Self-Management and Dropout in PRO-Based Follow-Up

The ORs for the self-management determinants and dropout in PRO-based follow-up are presented in Table 2. Statistically significant associations were found between all potential self-management determinants and dropout. A lower degree of self-management was associated with dropout. Patients reporting having low social support for health (HLQ4) had higher odds of dropout (OR 2,02, 95% CI 1.38-3.50). Similar associations were identified for the remaining determinants regarding health literacy (Table 2). Patients who had lower confidence in knowing when to get outpatient care (PAM5) had higher odds of dropout (OR 1.94, 95% CI 1.44-2.62). A high level of

XSL•FO
RenderX

self-efficacy had decreased odds for dropout per point (OR 0.96, 95% CI 0.95-0.98). When comparing a patient with the lowest GSES score (score of 10) to a patient with the highest possible GSES score (score of 40), the patient with the low self-efficacy had an OR of 3.21 (95% CI 2.05-5.03) for dropout (Table 3).

Table . Crude and adjusted odds ratio (OR) for potential determinants of self-management and drop-out in patient-reported outcome (PRO)–based follow-up among patients with epilepsy.

Potential determinants	Crude		Adjusted[a]	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Social support for health (HLQ 4)	n=2126		n=2099	
High >2	ref		ref	
Low ≤2	2.31(1.46-3.64)	<.001	2.20 (1.38-3.50)[b]	<.001
Ability to actively engage with health care providers (HLQ 6)	n=2126		n=2050	
High >3	ref		ref	
Low ≤3	1.85 (1.48-2.31)	<.001	1.82 (1.44-2.31)	<.001
Low ability to understand health information well enough to know what to do (HLQ9)	n=2119		n=2046	
High >3	ref		ref	
Low ≤3	2.03 (1.59-2.58)	<.001	2.15 (1.65-2.80)	<.001
Self-efficacy	n=2105		n=2032	
General Self-efficacy Scale score [c]	0.96 (0.95-0.98)[d]	<.001	0.96 (0.95-0.98)[d]	<.001
PAM #5	n=2147		n=2066	
Strongly agree/agree	ref		ref	
Disagree/strongly disagree	1.92 (1.45-2.54)	<.001	1.94 (1.44-2.62)	<.001
PAM #12	n=2139		n=2060	
Strongly agree/agree	ref		ref	
Disagree/strongly disagree	1.75 (1.38-2.21)	<.001	1.84 (1.43-2.36)	<.001

[a]Adjusted for sex, age, education, and cohabitation.

[b]Adjusted for sex and cohabitation due to low observations in some categories.

[c]General Self-Efficacy Scale score in the interval 10 - 40.

[d]Per point increase in the General Self-Efficacy Scale score. The score 0 is reference.

Table . Adjusted odds ratio (OR) for dropout in patient-reported outcome (PRO)–based follow-up among patients with epilepsy at different scores relative to a patient with the highest possible score (40) on the General Self-Efficacy Scale.

General Self-efficacy scale score	OR (95% CI)[a]	*P* value
40 points	ref	
30 points	1.47 (1.27-1.71)	<.001
20 points	2.18 (1.61-2.94)	<.001
10 points	3.21 (2.05-5.03)	<.001

[a]OR for dropout relative to a reference with a score on 40 adjusted for sex, age, education, and cohabitation

XSL•FO
RenderX

Associations Between General and Mental Health Status and Dropout in PRO-Based Follow-Up

We found that patients with good general health status had statistically significantly higher odds for dropout (OR 1.28, 95% CI 1.65-2.85) compared to patients who reported excellent/very good health status. Similarly, patients with fair/poor general heath status had statistically significantly higher odds for dropout (OR 2.17, 95% CI 1.65-2.85) compared to patients who reported excellent/very good health status (Table 4). Moreover, patients with lower mental health status had higher odds of dropout. A patient with the lowest possible WHO-5 score (0) had an OR of 3.16 (95% CI 1.94-5.15) for dropout relative to a patient with the highest score (100) (Table 5).

Table . Crude and adjusted Odds ratio (OR) for potential determinants of general and mental health status and drop-out in PRO-based follow-up among patients with epilepsy.

Potential determinants	Crude		Adjusted[a]	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Mental well-being	n=2218		n=2063	
WHO-5 Well-being Index score[b]	0.99 (0.98-0.99)[c]	<.001	0.99 (0.98-0.99)[c]	<.001
General health (SF-36)[d]	n=2249		n=2088	
Excellent/very good	ref		ref	
Good	1.12 (0.92-1.36)	.25	1.28 (1.04-1.57)	.02
Fair/poor	1.94 (1.51-2.50)	<.001	2.17 (1.65-2.85)	<.001

[a]Adjusted for sex, age, education, and cohabitation.

[b]WHO-5 Well-being Index score in the interval 0 - 100.

[c]Per point increase in the WHO-5 Well-being Index score. The score 0 is reference.

[d]Using a single item from the Short Form-36.

Table . Adjusted odds ratio (OR) for dropout in patient-reported outcome (PRO)-based follow-up among patients with epilepsy at different scores relative to a patient with the highest possible score (100) on the WHO-5 Well-being Index score.

WHO-5 Well-being Index score	OR (95% CI)[a]	*P* value
0 points	3.16 (1.94-5.15)	<.001
25 points	2.37 (1.64-3.42)	<.001
50 points	1.78 (1.39-2.27)	<.001
75 points	1.33 (1.18-1.51)	<.001

[a]OR for dropout relative to a reference with a score on 100 adjusted for sex, age, education, and cohabitation.

Associations Between Patient Satisfaction and Dropout in PRO-Based Follow-Up

The associations between patient satisfaction and dropout in PRO-based follow-up are presented in Table 6. Patients who, to a minor extent/not at all, felt sufficiently involved in the course of treatment had an OR of 1.93 (95% CI 1.42-2.63) of dropping out compared to patients who felt sufficiently involved. Similar associations were found for the determinants of confidence and comfort. Patients who were not satisfied and found their overall treatment to be mostly bad/very bad had an OR of 2.00 (95% CI: 1.31;3.06) for dropout compared to satisfied patients.

XSL•FO
RenderX

Table . Crude and adjusted odds ratio (OR) for potential determinants for patient satisfaction and drop-out in patient-reported outcome (PRO)–based follow-up among patients with epilepsy.

Potential determinants	Crude		Adjusted[a]	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Patient involvement	n=2009		n=1935	
A great extent/some extent	ref		ref	
A minor extent/not at all	1.91 (1.41-2.59)	<.001	1.93(1.42-2.63)[b]	<.001
Confidence	n=2044		n=2021	
A great extent/some extent	ref		ref	
A minor extent/not at all	1.64 (1.09-2.46)	.02	1.62 (1.07-2.44)	.02
Comfort	n=2004		n=1981	
A great extent/some extent	ref		ref	
A minor extent/not at all	2.06 (1.40-3.04)	<.001	2.04 (1.38-3.02)	<.001
Satisfaction	n=1956		n=1934	
Very good/mostly good	ref		ref	
Mostly bad/very bad	2.01(1.32-3.06)	<.001	2.00 (1.31-3.06)	<.001

[a]Adjusted for sex and cohabitation.

[b]Adjusted for sex, education, and cohabitation.

Discussion

Principal Findings and Comparison With Previous Works

The majority of the patients reported a high level of self-management, good health status, and a high degree of patient satisfaction. A strong association was found between all 9 self-reported potential determinants and dropout in remote PRO-based follow-up for patients with epilepsy in the Central Denmark Region. Patients with reduced self-management, poor health status, and low patient satisfaction had higher odds of dropout in PRO-based follow-up. The association appeared independent, as the correction for sex, age, education, and cohabitation showed only a minor impact on the relation between the potential determinants and dropout.

To the best of our knowledge, previous research has not primarily focused on investigating dropout rates; instead, dropout has typically been included as a secondary finding [18,22]. Furthermore, past research has not given exclusive attention to PRO-based follow-up but has focused on telehealth more broadly, encompassing app-based interventions [18,22]. Using a meta-analysis, Meyerowitz-Katz et al determined that app-based chronic disease management interventions exhibit a dropout rate of 49% (95% CI 27%-70%) [22]. A scoping review of 31 studies on patient reasons for not using digital PRO concepts identified a dropout range of 2.4% to 72.3% [18]. Both studies reported higher dropout rates compared to the dropout rate of 33.7% in this study. The studies included follow-up periods with varying response frequencies, ranging from daily to every 3 months [18,22]. This variability in follow-up schedules may have contributed to the higher dropout rates observed in the comparative studies, along with the contextual differences between dropout rates in digital interventions and dropout rates in continuous chronic disease follow-up. Adherence to technology is another concept that can indicate how patients use digital solutions. A recently published systematic review investigated factors associated with adherence to tele-monitoring using electronic PRO measures in patients with chronic diseases, including heart failure, rheumatoid arthritis, and chronic pain [42]. This review found adherence rates to range between 61% and 96% in the included studies [42], which aligns with the dropout rate of approximately 30% found in our study.

Patient characteristics associated with dropout have been investigated mainly in relation to sociodemographic factors, such as sex, age, education, and cohabitation status in other studies [18,22,42]. Self-reported aspects, including self-management and health literacy, have to our knowledge not been reported in other studies regarding the dropout rate in digital interventions. Thus, our study contributes novel knowledge in this field and underlines the importance of considering different individual needs and competencies when implementing digital solutions in clinical care. Nielsen et al also investigated reasons for not using digital PRO concepts in clinical practice among patients with long-term conditions. They found several reasons for digital nonuse of PRO; among others, the themes covered the ability to use PRO data, engagement, emotional distress, and technical barriers [18]. From the patient's point of view, these reasons are important to consider before implementing a digital PRO solution in a clinical setting. Reasons for dropout in PRO-based follow-up will be further

XSL•FO
RenderX

investigated in future research by interviewing patients as well as clinicians.

The clinicians' perspective is important to investigate because dropout could be related not only to digital skills or engagement by patients. From the health care provider's point of view, there may also be concerns that arise during the treatment of patients included in a digital solution. Qualitative research regarding the clinicians' perspective of using PRO-based follow-up in patients with epilepsy has shown reluctance among clinicians [43]. For example, some of the clinicians found the lack of interpersonal contact as a negative consequence of PRO-based follow-up and felt unsure about some of the patients' capabilities to participate even though the patients were already included in PRO-based follow-up [43]. There is no standardized guideline regarding referral to PRO-based follow-up in patients with epilepsy in the Central Denmark Region; thus, the personal preferences amongst clinicians may play a role when deciding to refer or exclude patients from PRO-based follow-up. Prior research investigating factors associated with referral to PRO-based follow-up among patients with epilepsy and patients with rheumatoid arthritis supports this statement since patients with a higher socioeconomic background more often were referred to PRO-based follow-up compared to patients with a lower socioeconomic background [24,44]. Similar to our finding, the study in patients with epilepsy also found a positive association between a high level of health literacy and referral to PRO-based follow-up [24]. These findings in addition to our current study support the importance on focusing on the involvement of vulnerable patient groups when designing digital PRO solutions in clinical practice.

Strengths and Limitations

We conducted a large Danish prospective cohort study among a representative epilepsy population enrolled in remote PRO-based follow-up. Self-reported data were collected before the outcome of interest; thus, potential misclassification would be nondifferential and not bias the results. The outcome (dropout) was register-based, and missing data were acceptable for both self-reported and register-based data. Thus, the internal and external validity of this study are considered sufficient. However, some limitations should be noted. First, we possessed only baseline information about the potential determinants. It is possible that the patients' perceptions of self-management, health status, and patient satisfaction changed during the follow-up period.

Second, the initial response rate in this study was high (83%); however, the only data available regarding the nonrespondents were sex and age. No differences were found between responders and nonresponders, but we cannot rule out that nonresponders did differ from the respondents considering other characteristics. In general, we found high levels of self-management, health status, and satisfaction amongst the responders. Nondifferential selection bias could occur if a lower level of these aspects may have been reported by the nonresponders. Third, an issue occurred with the interpretation of the question regarding the confounder variable education. This was evident for some younger patients who reported having a higher educational level than plausible for their age. It was not possible to determine if this issue was present throughout the study, and there is a possibility that this misclassification may have contributed to insufficient correction for education. Furthermore, despite adjusting for the most common sociodemographic confounders, a risk of residual confounding cannot be completely ruled out. Finally, differentiating the reasons for dropout would have enhanced the strength of this study. However, this information was not available and will be explored in future research.

Conclusions

In this study, we found that reduced self-management, poor health status, and patient dissatisfaction were associated with an increased odds of dropout in remote PRO-based follow-up. Thus, findings from this study support that a remote approach in outpatient follow-up may not be suitable to the whole patient population, but a selected group of patients. However, further research is necessary to explore the reasons for patient dropout. The findings underline the importance of involving patients with different needs and competencies when designing and implementing digital PRO solutions in the health care system.

Acknowledgments

This manuscript has been written without artificial intelligence technology.

Data Availability

The datasets generated and analyzed during this study are available in anonymous form from the corresponding author on request or contact the AmbuFlex-Center for Patient-reported Outcome [45].

Authors' Contributions

Data Analysis: MR and SBV
Supervison: DHC and LMVS
Writing – original draft: MR and SBV
Writing – review & editing: MR, SBV, DHC, and LMVS

References

XSL•FO
RenderX

1. Hvidberg MF, Johnsen SP, Davidsen M, Ehlers L. A nationwide study of prevalence rates and characteristics of 199 chronic conditions in Denmark. Pharmacoecon Open 2020 Jun;4(2):361-380. [doi: 10.1007/s41669-019-0167-7] [Medline: 31342402]
2. Ryu S. Telemedicine: opportunities and developments in member states: report on the second global survey on eHealth 2009 (Global Observatory for eHealth Series, Volume 2). Healthc Inform Res 2012;18(2):153. [doi: 10.4258/hir.2012.18.2.153] [Medline: 22509475]
3. Weinstein RS, Krupinski EA, Doarn CR. Clinical examination component of telemedicine, telehealth, mHealth, and connected health medical practices. Med Clin North Am 2018 May;102(3):533-544. [doi: 10.1016/j.mcna.2018.01.002] [Medline: 29650074]
4. Scalvini S, Vitacca M, Paletta L, Giordano A, Balbi B. Telemedicine: a new frontier for effective healthcare services. Monaldi Arch Chest Dis 2004;PMID(4):226-233. [doi: 10.4081/monaldi.2004.686]
5. Scott A, Bai T, Zhang Y. Association between telehealth use and general practitioner characteristics during COVID-19: findings from a nationally representative survey of Australian doctors. BMJ Open 2021 Mar 24;11(3):e046857. [doi: 10.1136/bmjopen-2020-046857] [Medline: 33762248]
6. Bashshur R, Doarn CR, Frenk JM, Kvedar JC, Woolliscroft JO. Telemedicine and the COVID-19 pandemic, lessons for the Future. Telemed J E Health 2020 May;26(5):571-573. [doi: 10.1089/tmj.2020.29040.rb] [Medline: 32275485]
7. Gandrup J, Ali SM, McBeth J, van der Veer SN, Dixon WG. Remote symptom monitoring integrated into electronic health records: a systematic review. J Am Med Inform Assoc 2020 Nov 1;27(11):1752-1763. [doi: 10.1093/jamia/ocaa177] [Medline: 32968785]
8. Schougaard LMV, Larsen LP, Jessen A, et al. AmbuFlex: tele-patient-reported outcomes (telePRO) as the basis for follow-up in chronic and malignant diseases. Qual Life Res 2016 Mar;25(3):525-534. [doi: 10.1007/s11136-015-1207-0] [Medline: 26790427]
9. Hjollund NHI, Larsen LP, de Thurah AL, et al. Patient-reported outcome (PRO) measurements in chronic and malignant diseases: ten years' experience with PRO-algorithm-based patient-clinician interaction (telePRO) in AmbuFlex. Qual Life Res 2023 Apr;32(4):1053-1067. [doi: 10.1007/s11136-022-03322-9] [Medline: 36639598]
10. Aiyegbusi OL, Hughes SE, Peipert JD, Schougaard LMV, Wilson R, Calvert MJ. Reducing the pressures of outpatient care: the potential role of patient-reported outcomes. J R Soc Med 2023 Feb;116(2):44-64. [doi: 10.1177/01410768231152222] [Medline: 36758615]
11. Calvert M, Kyte D, Price G, Valderas JM, Hjollund NH. Maximising the impact of patient reported outcome assessment for patients and society. BMJ 2019 Jan 24;364:k5267. [doi: 10.1136/bmj.k5267] [Medline: 30679170]
12. Black N. Patient reported outcome measures could help transform healthcare. BMJ 2013 Jan 28;346:f167. [doi: 10.1136/bmj.f167] [Medline: 23358487]
13. Aiyegbusi OL, Nair D, Peipert JD, Schick-Makaroff K, Mucsi I. A narrative review of current evidence supporting the implementation of electronic patient-reported outcome measures in the management of chronic diseases. Ther Adv Chronic Dis 2021;12:20406223211015958. [doi: 10.1177/20406223211015958] [Medline: 34104376]
14. Knudsen LR, de Thurah A, Lomborg K. Experiences with telehealth followup in patients with rheumatoid arthritis: a qualitative interview study. Arthritis Care Res (Hoboken) 2018 Sep;70(9):1366-1372. [doi: 10.1002/acr.23483] [Medline: 29195011]
15. Mejdahl CT, Schougaard LMV, Hjollund NH, Riiskjær E, Thorne S, Lomborg K. PRO-based follow-up as a means of self-management support - an interpretive description of the patient perspective. J Patient Rep Outcomes 2017;2:38. [doi: 10.1186/s41687-018-0067-0] [Medline: 30238083]
16. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11. [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
17. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 5;13(3):e52. [doi: 10.2196/jmir.1772] [Medline: 21821503]
18. Nielsen AS, Kidholm K, Kayser L. Patients' reasons for non-use of digital patient-reported outcome concepts: a scoping review. Health Informatics J 2020 Dec;26(4):2811-2833. [doi: 10.1177/1460458220942649] [Medline: 32731773]
19. Pedersen DH, Mansourvar M, Sortsø C, Schmidt T. Predicting dropouts from an electronic health platform for lifestyle interventions: analysis of methods and predictors. J Med Internet Res 2019 Sep 4;21(9):e13617. [doi: 10.2196/13617] [Medline: 31486409]
20. Multone E, Vader JP, Mottet C, et al. Characteristics of non-responders to self-reported questionnaires in a large inflammatory bowel disease cohort study. Scand J Gastroenterol 2015;50(11):1348-1356. [doi: 10.3109/00365521.2015.1041150] [Medline: 25921677]
21. Højmark K, Støttrup C, Carreon L, Andersen MO. Patient-reported outcome measures unbiased by loss of follow-up. Single-center study based on DaneSpine, the Danish spine surgery registry. Eur Spine J 2016 Jan;25(1):282-286. [doi: 10.1007/s00586-015-4127-3] [Medline: 26208938]
22. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease. Syst Rev Meta-Analysis J Med Internet Res 2020;22(9):e20283. [doi: 10.2196/20283] [Medline: 32990635]

XSL•FO
RenderX

23. Portz JD, LaMendola WF. Participation, retention, and utilization of a web-based chronic disease self-management intervention among older adults. Telemed J E Health 2019 Feb;25(2):126-131. [doi: 10.1089/tmj.2017.0208] [Medline: 29782228]
24. Schougaard LMV, de Thurah A, Christensen J, et al. Sociodemographic, personal, and disease-related determinants of referral to patient-reported outcome-based follow-up of remote outpatients: a prospective cohort study. Qual Life Res 2020 May;29(5):1335-1347. [doi: 10.1007/s11136-019-02407-2] [Medline: 31900763]
25. Christensen AI, Ekholm O, Glümer C, et al. The Danish National Health Survey 2010. Study design and respondent characteristics. Scand J Public Health 2012 Jun;40(4):391-397. [doi: 10.1177/1403494812451412] [Medline: 22786925]
26. Christensen AI, Ekholm O, Gray L, Glümer C, Juel K. What is wrong with non-respondents? Alcohol-, drug- and smoking-related mortality and morbidity in a 12-year follow-up study of respondents and non-respondents in the Danish Health and Morbidity Survey. Addiction 2015 Sep;110(9):1505-1512. [doi: 10.1111/add.12939] [Medline: 25845815]
27. Parai C, Hägg O, Willers C, Lind B, Brisby H. Characteristics and predicted outcome of patients lost to follow-up after degenerative lumbar spine surgery. Eur Spine J 2020 Dec;29(12):3063-3073. [doi: 10.1007/s00586-020-06528-y] [Medline: 32666273]
28. Schougaard LMV, de Thurah A, Christiansen DH, Sidenius P, Hjollund NH. Patient-reported outcome (PRO) measure-based algorithm for clinical decision support in epilepsy outpatient follow-up: a test-retest reliability study. BMJ Open 2018 Jul 25;8(7):e021337. [doi: 10.1136/bmjopen-2017-021337] [Medline: 30049693]
29. Schougaard LMV, Mejdahl CT, Christensen J, et al. Patient-initiated versus fixed-interval patient-reported outcome-based follow-up in outpatients with epilepsy: a pragmatic randomized controlled trial. J Patient Rep Outcomes 2019 Sep 13;3(1):61. [doi: 10.1186/s41687-019-0151-0] [Medline: 31520247]
30. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 2013 Jul 16;13:658. [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
31. Maindal HT, Kayser L, Norgaard O, Bo A, Elsworth GR, Osborne RH. Cultural adaptation and validation of the Health Literacy Questionnaire (HLQ): robust nine-dimension Danish language confirmatory factor model. Springerplus 2016;5(1):1232. [doi: 10.1186/s40064-016-2887-9] [Medline: 27536516]
32. Luszczynska A, Scholz U, Schwarzer R. The general self-efficacy scale: multicultural validation studies. J Psychol 2005 Sep;139(5):439-457. [doi: 10.3200/JRLP.139.5.439-457] [Medline: 16285214]
33. Scholz U, Doña BG, Sud S, Schwarzer R. Is general self-efficacy a universal construct? Psychometric findings from 25 countries. Eur J Psychol Assess 2002;18(3):242-251. [doi: 10.1027//1015-5759.18.3.242]
34. Maindal HT, Sokolowski I, Vedsted P. Translation, adaptation and validation of the American short form Patient Activation Measure (PAM13) in a Danish version. BMC Public Health 2009 Jun 29;9:209. [doi: 10.1186/1471-2458-9-209] [Medline: 19563630]
35. Maindal HT, Vedsted P, Mikkelsen EM. Oversættelse og kulturel tilpasning af den danske version af "The Patient Activation Measure". KS 2011 Mar 2;25(1):7-17. [doi: 10.18261/ISSN1903-2285-2011-01-03]
36. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176. [doi: 10.1159/000376585] [Medline: 25831962]
37. Heun R, Burkart M, Maier W, Bech P. Internal and external validity of the WHO Well-Being Scale in the elderly general population. Acta Psychiatr Scand 1999 Mar;99(3):171-178. [doi: 10.1111/j.1600-0447.1999.tb00973.x] [Medline: 10100911]
38. Bjorner JB, Thunedborg K, Kristensen TS, Modvig J, Bech P. The Danish SF-36 Health Survey: translation and preliminary validity studies. J Clin Epidemiol 1998 Nov;51(11):991-999. [doi: 10.1016/s0895-4356(98)00091-2] [Medline: 9817117]
39. Bjorner JB, Damsgaard MT, Watt T, Groenvold M. Tests of data quality, scaling assumptions, and reliability of the Danish SF-36. J Clin Epidemiol 1998 Nov;51(11):1001-1011. [doi: 10.1016/s0895-4356(98)00092-4] [Medline: 9817118]
40. Macias C, Gold PB, Öngür D, Cohen BM, Panch T. Are single-item global ratings useful for assessing health status? J Clin Psychol Med Settings 2015 Oct 22. [doi: 10.1007/s10880-015-9436-5] [Medline: 26492891]
41. Bekæmpelse K. Kræftpatienters behov og oplevelser med sundhedsvæsenet under udredning og behandling -kræftens bekæmpelses barometerundersøgelse (in Danish). : Danish Cancer Society; 2017 URL: https://www.cancer.dk/om-os/udgivelser-og-rapporter/2017kraeftpatienters-behov-og-oplevelser-med-sundhedsvaesenet-under-udredning-og-behandling-2017/ [accessed 2024-12-19]
42. Wiegel J, Seppen B, van der Leeden M, van der Esch M, de Vries R, Bos W. Adherence to telemonitoring by electronic patient-reported outcome measures in patients with chronic diseases: a systematic review. Int J Environ Res Public Health 2021 Sep 27;18(19):10161. [doi: 10.3390/ijerph181910161] [Medline: 34639463]
43. Mejdahl CT, Schougaard LMV, Hjollund NH, Riiskjær E, Lomborg K. Exploring organisational mechanisms in PRO-based follow-up in routine outpatient care - an interpretive description of the clinician perspective. BMC Health Serv Res 2018 Jul 13;18(1):546. [doi: 10.1186/s12913-018-3352-y] [Medline: 30001729]
44. Schougaard LMV, Knudsen LR, Grove BE, et al. Socioeconomic, disease-related, and personal factors associated with participation in remote follow-up in rheumatoid arthritis: a cross-sectional study. Arthritis Care Res (Hoboken) 2023 Oct;75(10):2044-2053. [doi: 10.1002/acr.25105] [Medline: 36785998]

XSL•FO
RenderX

45. AmbuFlex, center for PRO. Health in the Central Denmark Region. URL: www.ambuflex.dk [accessed 2025-01-13]

Abbreviations

CI: confidence interval
GSES: General Self-Efficacy Scale
HLQ: Health Literacy Questionnaire
IQR: Interquartile range
OR: odds ratio
PAM: patient activation measure
PRO: patient-reported outcome
SF-36: Short Form-36
WHO: World Health Organization
WHO-5: WHO-5 Well-being Index

Edited by A Mavragani, C Hoving; submitted 11.03.24; peer-reviewed by A Hassan, J Mistry; revised version received 04.11.24; accepted 06.11.24; published 14.01.25.

Please cite as:
Vestergaard SB, Roost M, Christiansen DH, Schougaard LMV
Determinants of Dropping Out of Remote Patient-Reported Outcome–Based Follow-Up Among Patients With Epilepsy: Prospective Cohort Study
JMIR Form Res 2025;9:e58258
URL: https://formative.jmir.org/2025/1/e58258
doi:10.2196/58258

© Sofie Bech Vestergaard, Mette Roost, David Høyrup Christiansen, Liv Marit Valen Schougaard. Originally published in JMIR Formative Research (https://formative.jmir.org), 14.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impact of a Virtual Care Navigation Service on Member-Reported Outcomes Among Lesbian, Gay, Bisexual, Transgender, and Queer Populations: Case Study

Seul Ki Choi[1], MPH, PhD; Jaclyn Marshall[2], MS; Patrina Sexton Topper[3], RN, MS, PhD; Andrew Pregnall[4], MSc, MPhil; José Bauermeister[1], MPH, PhD

[1]School of Nursing, University of Pennsylvania, 418 Curie Blvd, Philadelphia, PA, United States

[2]Included Health, San Francisco, CA, United States

[3]College of Nursing, Villanova University, Villanova, PA, United States

[4]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Seul Ki Choi, MPH, PhD
School of Nursing, University of Pennsylvania, 418 Curie Blvd, Philadelphia, PA, United States

Abstract

Background: While the significance of care navigation in facilitating access to health care within the lesbian, gay, bisexual, transgender, queer, and other (LGBTQ+) communities has been acknowledged, there is limited research examining how care navigation influences an individual's ability to understand and access the care they need in real-world settings. By analyzing private sector data, we can bridge the gap between theoretical research findings and practical applications, ultimately informing both business strategies and public policy with evidence grounded in real-world efficacy.

Objective: The objective of this study was to evaluate the impact of specialized virtual care navigation services on LGBTQ+ individuals' ability to comprehend and access necessary care within a national cohort of commercially insured members.

Methods: This case study is based on the experience of commercially insured members, aged 18 or older, who used the LGBTQ+ Health Care Navigation (LGBTQ+ Navigation) service by Included Health between January 26 and July 31, 2023. Care coordinators assisted members by connecting them with vetted identity-affirming in-network providers, helping them navigate and understand their LGBTQ+ health benefits, and providing education and advocacy for clinical and nonclinical needs. We examined the impact of navigation on 5 member-reported outcomes. In addition to reporting the proportion who agreed or strongly agreed, we calculated an impact score that averaged assigned numerical values to all 5 question responses (1=strongly disagree to 5=strongly agree) for each respondent. We used ANOVA with Tukey post hoc tests and *t* tests to explore the relationships between the impact score and member characteristics, including optional self-reported demographics.

Results: Out of 4703 LGBTQ+ Navigation cases, 7.53% (n=354) had member-reported outcomes. A large majority of LGBTQ+ members agreed or strongly agreed that care navigation resulted in less stress (315/354, 89%), less care avoidance (305/354, 86.2%), higher confidence in finding an identity-affirming provider (327/354, 92.4%), improved ability to comprehend health care information (312/354, 88.1%), and improved ability to engage with providers (308/354, 87%). The average impact score was 4.44 (SD 0.69), with statistically significant differences by gender identity (*P*=.003), race (*P*=.01), ethnicity (*P*=.008), and pronouns (*P*=.02). The scores were highest for members with multiple gender identities (mean 4.56, SD 0.37), and members who did not provide their race, ethnicity, or their pronouns (mean 4.55, SD 0.64). Impact scores were lowest for transgender members (mean 4.11, SD 0.95).

Conclusions: The LGBTQ+ Navigation service, by enhancing members' comprehension and use of necessary care, demonstrates potential public health utility and value. Continuous evaluation of navigation services can serve as a supplementary tool for employers seeking to promote health equity and improve belonging among employees. This is particularly important as discrimination and stigma against LGBTQ+ communities persist in the United States. Therefore, scalable and system-level changes that use navigation services are essential to reach a larger proportion of the LGBTQ+ population.

(JMIR Form Res 2025;9:e64137) doi:10.2196/64137

KEYWORDS

health care navigation; LGBTQ+; lesbian; gay; bisexual; transgender; queer; access to care; care avoidance; identity-affirming care; minority; stress; stigma; health outcomes

XSL•FO
RenderX

Introduction

Compelling evidence indicates that communities comprised of lesbian, gay, bisexual, transgender, queer, and a broader spectrum of sexual orientations and/or gender identities (LGBTQ+) disproportionately endure a range of health issues [1-3]. According to minority stress theory, the stigma associated with sexual and/or gender minority identities acts as a stressor, potentially serving as a foundational contributor to health inequities among LGBTQ+ people [4]. This inequity includes adverse health outcomes, such as various forms of cancer, mental health disorders including depression and anxiety, HIV, and higher rates of substance use compared to their heterosexual and cisgender counterparts [5-10].

Research has identified impeded access to health care services as a key driver of adverse health outcomes within the LGBTQ+ communities, further amplifying health inequities [11-13]. Drawing upon the comprehensive framework for health care access of Levesque et al [14], it is imperative to examine health care accessibility from a multifaceted perspective. This perspective can involve accounting for factors like lack of inclusive policies, limited availability of health insurance, financial challenges that make health care unaffordable, insufficient awareness of available health care services, a sense of mistrust toward health care providers, and poor quality health care related to discrimination based on sexual orientation and gender identity, as well as a lack of health care providers' understanding of LGBTQ+ health care needs [15-24].

Policies and interventions that take this multidimensional view of health care access [14] are essential for improving access to health care, as well as improving access to the evidence-based medical information, all of which play a significant role in facilitating health care access [25]. One such intervention is care navigation, acknowledged as a successful model for increasing access to quality health care, particularly for marginalized populations [26]. Recently, small studies have demonstrated the ability of patient navigation to connect and engage individuals within diverse LGBTQ+ communities with HIV care [27] and facilitate gender-affirming care for transgender populations [28, 29].

While the significance of care navigation in facilitating access to care within the LGBTQ+ communities has been acknowledged [27-29], research in this area is limited. One gap that remains is understanding how care navigation influences an individual's ability to understand and access the care they need in real-world settings. Evaluating these services that are delivered by the private sector offers unique value that complements traditional research studies by providing real-world, actionable insights. Unlike controlled environments in research, private sector evaluations reflect the complexities of actual consumer behaviors, operational challenges, and market conditions. This allows for the observation of how services perform in diverse, dynamic settings where human behavior is less predictable and influenced by a range of contextual factors. Moreover, private sector data often encompasses larger, more varied populations, offering a broader scope of generalizability. By analyzing this data, we can bridge the gap between theoretical research findings and practical applications, ultimately informing both business strategies and public policy with evidence grounded in real-world efficacy. Therefore, the objective of the study was to evaluate the use of an LGBTQ+-specialized virtual care navigation service among Included Health's members and assess the impact of these services on members' abilities to understand and access the care they need among a national, commercially-insured cohort.

Methods

This study follows STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) reporting guidelines for cohort studies.

LGBTQ+ Virtual Health Care Navigation Service

This study is based on the experience of commercially insured members, aged 18 or older, who used the LGBTQ+ Health Care Navigation (LGBTQ+ Navigation) service by Included Health. LGBTQ+ Navigation is Included Health's phone- and chat-based health care navigation service for LGBTQ+ communities. It is offered by self-insured employers as a health care benefit for employees and their dependents and by health plans for their insured members. The LGBTQ+ Navigation service is consistent across all members who are eligible for the service. Employers share information about the navigation service with their employees and their family members similarly to other health benefits, which may include mailings, webinars, and their intranet. Members then proactively initiate a case by completing an intake form on Included Health's website or within the digital health app. A team of care coordinators, 90% of whom identify as LGBTQ+, provide concierge advocacy and guidance for members across 3 areas: connections to vetted identity-affirming in-network providers, benefits navigation focused on LGBTQ+ health, and education and advocacy for clinical and nonclinical needs.

These providers, ranging from primary care physicians to specialists, have been vetted by Included Health to ensure familiarity with LGBTQ+ health needs and inclusive practices. Coordinators, trained in investigative interviewing, gather information by speaking directly to providers and their staff on the provider's experience with population-level health needs, the treatment of LGBTQ+ members, and confirmation of inclusive practices, such as using the correct name and pronouns for patients regardless of their gender assigned at birth. If there are not multiple provider options for the member in the directory, coordinators conduct additional research and vetting to find suitable providers.

Included Health uses a rigorous vetting process to ensure that providers recommended for LGBTQ+ care offer a high standard of cultural competence and inclusivity. Rather than relying on binary yes or no questions or allowing providers to self-identify as inclusive, the care coordinators—who bring specialized training and lived experience—conduct in-depth conversations with providers. Care coordinators are trained over a 4 to 5 week program where they are prepared to address various topics including benefits plans, social determinants of health, family building, and common systemic health conditions faced by

XSL•FO
RenderX

LGBTQ+ individuals. In addition, they receive training in patient advocacy and engage in intuitive listening practices to round out their deep practical knowledge with the skills to uncover needs and empower members. This approach allows coordinators to evaluate providers' practical experience with LGBTQ+ communities and their commitment to delivering culturally affirming, high-quality care. To further ensure consistency in care standards, Included Health actively solicits feedback from LGBTQ+ members who visit these recommended providers. Providers who do not meet member expectations for culturally competent care are reviewed and, if necessary, removed from the directory. This continuous feedback loop, guided by both professional assessment and member experiences, helps to maintain a directory of providers who align with Included Health's standards for LGBTQ+ inclusive care.

Care coordinators also assist members with understanding their benefits related to LGBTQ+ health, such as family expansion and gender-affirming care. For example, members interested in family expansion may have access to adoption support and reproductive endocrinology interventions. In addition, care coordinators assist the member in completing all the steps along gender-affirming care pathways, which may include any necessary documentation such as World Professional Association for Transgender Health (WPATH) letters and access to travel benefits if needed. They are trained to uncover additional needs, allowing them to connect members to benefits they might not know they have. In addition, the care coordinators support members through education and advocacy for clinical and nonclinical needs. This support may include helping the member navigate a claim denial, interpret the workplace policy guidance, identify community resources (eg, parents of transgender children support group), complete required paperwork prior to receiving care, and inform the member on how to change their legal name.

Data Source

Our study sample consisted of all completed service requests (closed cases) of LGBTQ+ Navigation from January 26 to July 31, 2023. We merged data from member intake forms and the postcase member surveys. Members initiated one or more LGBTQ+ Navigation cases by completing an LGBTQ+ Navigation member request form. The form requested information to confirm LGBTQ+ Navigation eligibility, how the member could be supported (eg, find a provider or resource request), and geographic location, as well as optional fields for race, sexual orientation, pronouns, and gender identity. At case closure, the LGBTQ+-specialized care coordinator emailed an optional survey, without an incentive offered for completing the survey, to the member specific to the LGBTQ+ Navigation offering.

Ethical Considerations

The University of Pennsylvania Institutional Review Board (IRB) reviewed the study's protocol and concluded it was exempt as it did not meet the regulatory definition of human subjects research. The committee stated that "the IRB determined that this project qualifies as quality/program improvement meaning that the project does not meet the regulatory definition of human participants research and therefore does not require formal IRB review." When submitting the form to request navigation services, the member agreed to the terms of service which included the Included Health privacy policy, which stated anonymous information may be used for research. Data were deidentified which ensured that the authors did not have access to any identifiable information about members. Compensation was not provided for study participation.

Measures

Member and Case Characteristics

The member's geographic region was defined as Midwest, Northeast, South, or West based on the US census regions. We also included the type of case request as a binary variable (a provider request or seeking resource/support) and whether the member had more than one case. Race, ethnicity, gender identity, sexual orientation, and pronouns were optional fields within the member intake form. The race and ethnicity question allowed respondents to select all that apply, aligning with US Census Bureau recommendations [30]. Due to limited sample size, we derived 2 variables: one indicating whether they identified as White and the other indicating whether they identified as Hispanic. For pronouns, individuals were presented with the option to select from he/him/his, she/her/hers, they/them/theirs, and ze/zir/zirs using a choose-all-that-apply format. Subsequently, we derived a variable indicating whether they chose a single set of pronouns or multiple pronouns.

In the original survey, participants could select all applicable options for gender identity and sexual orientation (see Supplement 2 in Multimedia Appendix 1) [31]. To create mutually exclusive variables for gender identity and sexual orientation, we compiled all the selected options and categorized them into distinct sexual orientations and gender identities. If respondents chose a single identity for either sexual orientation or gender identity, they were merged into a new variable informed by the National Academies' Consensus Report [32]. For both gender identity and sexual orientation, some distinct identities had low sample sizes. In those instances, we combined the respondents into one category. If they selected multiple options for sexual orientation or gender identity, they were placed in a category for multiple options. Consequently, we established mutually exclusive categories for gender identity (cisgender; transgender; other gender identities including agender, gender fluid, intersex, or nonbinary; and multiple gender identities) and sexual orientation (bisexual; gay/lesbian; other sexual orientations including aromantic, asexual, pansexual, queer, or heterosexual; and multiple sexual orientations). For cases in which members did not complete the optional form fields, their member characteristics were presented as not available to indicate missing information. Details on the original responses for gender identity and sexual orientation, as well as how this data was merged into new mutually exclusive categories, can be found in the Supplement 2 in Multimedia Appendix 1.

XSL•FO
RenderX

Member-Reported Outcomes

Our study outcome examined members' evaluation of LGBTQ+ Navigation using a 5-point scale (1=strongly disagree to 5=strongly agree). The questions began, "Since working with Included Health, I am…" and the 5 fragments were (1) less stressed about using health care services, (2) less likely to avoid care, (3) better able to find, understand, and use the information I need to inform my health care choices, (4) better able to find health care providers who understand my needs, and (5) more prepared to engage with health care providers.

In addition to the individual responses, we calculated a composite score for each member to denote the LGBTQ+ Navigation impact. Responses were averaged across the 5 questions (Cronbach α=0.94). A score of 4 or higher was considered an overall positive impact of LGBTQ+ Navigation, as it meant the respondent selected agree (4 points) or strongly agree (5 points) on average, across the 5 separate questions. A score less than 4 was considered a neutral to no impact of LGBTQ+ Navigation, since it indicated that the member was more likely to select neither agree or disagree, disagree, or strongly disagree across the questions.

Statistical Analyses

We used descriptive statistics, including means and distributions, to report the study sample characteristics and the member-reported outcomes. To understand the differences between those with and without member-reported outcomes, we used Fisher exact tests or χ^2 tests for categorical variables and 2-tailed *t* tests for continuous variables. Among the subset of individuals who completed a follow-up survey, we explored the relationships between the LGBTQ+ Navigation impact score and member characteristics. We used ANOVA with Tukey post hoc tests and *t* tests. The α level was defined as .05. All analyses were conducted in SAS version 9.4 (SAS Institute).

Results

There were 4703 LGBTQ+ Navigation cases in the study sample, of which 354 (7.53%) cases were linked to postcase surveys with member-reported outcomes (Table 1). LGBTQ+ Navigation cases with member-reported outcomes were more likely to be for members with multiple cases (202/354, 57.1% vs 1761/4349, 40.49%; *P*<.001) and less likely to be from members who live in the South (98/354, 27.8% vs 1635/4349, 37.97%; *P*<.003) compared to cases without member-reported outcomes. Moreover, members who provided member-reported outcomes were less likely to provide their demographic characteristics, including race and ethnicity (178/354, 50.3% vs 1484/4349, 34.12%), pronouns (178/354, 50.3% vs 1484/4349, 34.12%), gender identity (205/354, 57.9% vs 1869/4349, 42.98%), and sexual orientation (190/354, 53.7% vs 1663/4349, 38.24%), in comparison to those who did not submit member-reported outcomes (*P*<.001). There was not a statistically significant difference in the proportion of cases that requested assistance in finding an identity-affirming provider (350/354, 98.9% vs 4229/4349, 97.24%; *P*=.07).

Among cases linked to member-reported outcomes, half of the respondents chose not to provide their race (178/354, 50.3%) and ethnicity (178/354, 50.3%), pronouns (178/354, 50.3%), gender identity (205/354, 57.9%), or sexual orientation (190/354, 53.7%). For the cases where race, ethnicity, and pronouns were provided (176/354, 49.7%), a majority identified as White (113/176, 64.2%) and non-Hispanic (141/176, 80.1%), and they used singular pronouns (149/176, 84.7%). For the cases where gender identities were provided (149/354, 42.1%), half of the members identified as cisgender (75/149, 50.3%), followed by transgender (33/149, 22.2%); other gender identities including agender, gender fluid, intersex, or nonbinary (30/149, 20.1%); or multiple gender identities (11/149, 7.4%). For the cases where sexual orientations were provided (164/354, 46.3%), 46.3% (76/164) identified as gay or lesbian, followed by multiple sexual orientations (40/164, 24.4%); other sexual orientations including aromantic, asexual, pansexual, queer, or heterosexual (30/164, 18.3%); or bisexual (18/164, 11%).

We found that a large majority of respondents agreed or strongly agreed that LGBTQ+ Navigation had a positive impact on their ability to understand and use the care they needed. Roughly 89% (315/354) felt less stressed about accessing health care services, and 86.2% (305/354) reported they were less likely to avoid care (Table 2). Furthermore, 88.1% (312/354) of respondents agreed or strongly agreed that they improved their ability to find, comprehend, and use the information necessary for making informed health care decisions. Over 92% (327/354) agreed or strongly agreed they had increased confidence in finding health care providers, and 87% (308/354) felt more prepared to engage with these providers.

The average composite LGBTQ+ Navigation impact score was 4.44 (SD=0.69). Approximately 44.6% (158/354) of members had a composite score of 5, indicating they responded "strongly agree" to each of the 5 statements. In addition, 36.2% (128/354) of scores were 4 or more but less than 5, while 19.2% (68/354) were less than 4 (Table 3).

The LGBTQ+ Navigation impact scores varied by case and member characteristics (Table 4). The mean impact scores were higher for members in the South and Midwest (South: 4.69; Midwest: 4.59; Northeast: 4.39; West: 4.20; *P*<.001); members with multiple gender identities (multiple gender identities: 4.56; not available: 4.55; cisgender: 4.36; other gender identities: 4.25; transgender: 4.11; *P*=.003); and members who did not provide their race (not available: 4.55; non-White: 4.40; White: 4.30; *P*=.01), their ethnicity (not available: 4.55; Hispanic: 4.46; non-Hispanic: 4.30; *P*=.008), or their pronouns (not available: 4.55; multiple pronouns 4.36; single set of pronouns: 4.33; *P*=.02). The impact scores did not differ significantly by sexual orientation.

XSL•FO
RenderX

Table . Associations between case and member characteristics and the completion status of member-reported outcomes (n=4703).

Variable	Total (n=4703), n (%)	Member-reported outcomes, n (%)		*P* value
		No (n=4349)	Yes (n=354)	
Region				.003
Midwest	694 (14.9)	635 (14.75)	59 (16.8)	
Northeast	809 (17.37)	738 (17.14)	71 (20.2)	
South	1733 (37.2)	1635 (37.97)	98 (27.8)	
West	1422 (30.53)	1298 (30.14)	124 (35.2)	
Requester service				.07
Provider request	4579 (97.36)	4229 (97.24)	350 (98.9)	
Resources/support	124 (2.64)	120 (2.76)	4 (1.1)	
Member had more than one case				<.001
No	2740 (58.26)	2588 (59.51)	152 (42.9)	
Yes	1963 (41.74)	1761 (40.49)	202 (57.1)	
Race				<.001
Non-White	1302 (27.68)	1239 (28.49)	63 (17.8)	
White	1739 (36.98)	1626 (37.39)	113 (31.9)	
N/A[a]	1662 (35.34)	1484 (34.12)	178 (50.3)	
Ethnicity				<.001
Non-Hispanic	2484 (52.82)	2343 (53.87)	141 (39.8)	
Hispanic	558 (11.86)	523 (12.03)	35 (9.9)	
N/A	1661 (35.52)	1483 (34.1)	178 (50.3)	
Pronouns				<.001
Single pronouns	2725 (57.94)	2576 (59.23)	149 (42.1)	
Multiple pronouns	316 (6.72)	289 (6.65)	27 (7.6)	
N/A	1662 (35.64)	1484 (34.12)	178 (50.3)	
Gender identity				<.001
N/A	2074 (44.1)	1869 (42.98)	205 (57.9)	
Cisgender	1677 (35.66)	1602 (36.84)	75 (21.2)	
Multiple gender identities	86 (1.83)	75 (1.72)	11 (3.1)	
Other[b]	422 (8.97)	392 (9.01)	30 (8.5)	
Transgender	444 (9.44)	411 (9.45)	33 (9.3)	
Sexual orientation				<.001
N/A	1853 (39.4)	1663 (38.24)	190 (53.7)	
Bisexual	276 (5.87)	258 (5.93)	18 (5.1)	
Gay/lesbian	1402 (29.81)	1326 (30.49)	76 (21.5)	
Multiple sexual orientations	497 (10.57)	457 (10.51)	40 (11.3)	
Other[c]	675 (14.35)	645 (14.83)	30 (8.5)	

[a]N/A: not available.

[b]Other gender identities included agender, gender fluid, intersex, or nonbinary.

[c]Other sexual orientations included aromantic, asexual, pansexual, queer, or heterosexual.

XSL•FO
RenderX

Table . Unadjusted average and distribution of responses by member-reported outcomes (n=354).

Response item	Response score, mean (SD)	Responses, n (%)				
		Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Less stress	4.44 (0.78)	4 (1.1)	3 (0.8)	32 (9)	109 (30.8)	206 (58.2)
Less likely to avoid care	4.37 (0.92)	9 (2.5)	7 (2)	33 (9.3)	100 (28.2)	205 (57.9)
Better able to find and use health care information	4.45 (0.76)	2 (0.6)	3 (0.8)	37 (10.4)	102 (28.8)	210 (59.3)
Better able to find provider	4.51 (0.74)	4 (1.1)	3 (0.8)	20 (5.6)	108 (30.5)	219 (61.9)
More prepared to engage with physician	4.43 (0.78)	3 (0.8)	2 (0.6)	41 (11.6)	102 (28.8)	206 (58.2)

Table . Unadjusted composite LGBTQ+[a] Health Care Navigation impact score (n=354).

Composite scores[b]	LGBTQ+ Navigation impact values
Average score, mean (SD)	4.44 (0.69)
Score of less than 4, n (%)	68 (19.2)
Score of 4 or more but less than 5, n (%)	128 (36.2)
Score or 5, n (%)	158 (44.63)

[a]LGBTQ+: lesbian, gay, bisexual, transgender, queer, and others.

[b]Composite scores are the unadjusted average of the 5 individual member-reported outcomes for each member.

XSL•FO
RenderX

Table . Association between LGBTQ+ Health Care Navigation impact scores and case and member characteristics (n=354).

	LGBTQ+ Health care navigation impact score, mean (SD)	*P* value
Region		<.001[b]
Midwest	4.59 (0.52)	
Northeast	4.39 (0.72)	
South	4.69 (0.51)	
West	4.20 (0.79)	
Requester service		.41[c]
Provider request	4.45 (0.67)	
Resources/support	3.6 (1.77)	
Returning user		.054[c]
No	4.36 (0.67)	
Yes	4.50 (0.71)	
Race		.01[b]
Non-White	4.40 (0.74)	
White	4.30 (0.73)	
N/A[d]	4.55 (0.64)	
Ethnicity, n (%)		.008[b]
Non-Hispanic	4.30 (0.74)	
Hispanic	4.46 (0.69)	
N/A	4.55 (0.64)	
Pronouns, n (%)		.02[b]
Single pronouns	4.33 (0.76)	
Multiple pronouns	4.36 (0.51)	
N/A	4.55 (0.64)	
Gender identity, n (%)		.003[b]
N/A	4.55 (0.64)	
Cisgender	4.36 (0.66)	
Multiple gender identities	4.56 (0.37)	
Other[e]	4.25 (0.77)	
Transgender	4.11 (0.95)	
Sexual orientation, n (%)		.05[b]
N/A	4.54 (0.65)	
Bisexual	4.48 (0.59)	
Gay/lesbian	4.31 (0.69)	
Multiple sexual orientations	4.34 (0.68)	
Other[f]	4.27 (0.97)	

[a]LGBTQ+: lesbian, gay, bisexual, transgender, queer, and others.

[b]Determined using ANOVA with Tukey post hoc tests.

[c]Determined using 2-tailed *t* tests.

[d]N/A: not available.

[e]Other gender identities included agender, gender fluid, intersex, or nonbinary.

XSL•FO
RenderX

[f]Other sexual orientations included aromantic, asexual, pansexual, queer, or heterosexual.

Discussion

Principal Findings

To our knowledge, this is the first study to empirically measure the impact of a national, virtual LGBTQ+-specialized care navigation service on member-reported outcomes among commercially insured adults. We found the navigation service, which was considered a comprehensive framework to health care access [14], had a meaningful impact on multiple process outcomes that were associated with the use of preventive services, improved health, and lower long-term costs.

While LGBTQ+ individuals are more likely to receive preventive screenings and have better management of mental health conditions when under the care of an identity-affirming provider [33], the challenge lies in finding LGBTQ+ affirming providers [33,34]. Our study highlighted that members who used LGBTQ+ Navigation reported improved confidence in finding identity-affirming providers and reported being prepared to interact with them. Care avoidance is a significant health care concern as well, leading to the reduced use of preventive services, compromised health outcomes, and elevated long-term health care expenses [35,36]. This is particularly pertinent to the LGBTQ+ communities, which are more prone than other populations to avoid seeking care due to barriers and discrimination [37].

Our study found that 86.2% (305/354) of members reported they were less likely to avoid care after using the LGBTQ+ Navigation service. Moreover, LGBTQ+ Navigation increased members' ability to find, understand, and use health care information, which may lead to potential improvements in medication adherence, increased use of preventive care, reduced emergency room visits, and enhanced management of chronic care [38]. These results echo existing literature discussing the positive impact of LGBTQ+ navigation services on health care access and health care outcomes [28,39,40]. On the other hand, a fifth of members had an LGBTQ+ Navigation impact score that was less than 4, which means they were more likely to respond with neutral, disagree, or strongly disagree to the survey questions. This may indicate the presence of systemic barriers, such as laws, legislation, and insurance policies, that LGBTQ+ Navigation services are unable to address. Further analysis of members who did not respond affirmatively is needed to enhance services and better understand their unmet needs.

We found that care navigation positively impacted all member subgroups, with scores above 4 on the LGBTQ+ Navigation impact, though there were statistically significant differences between the groups. Notably, we found statistically significant differences by gender identity, with the lowest score for transgender individuals. In the context of historical stigma and continual discrimination toward transgender individuals compared to cisgender individuals, the lower scores suggest the need for additional gains toward more inclusive care environments. We also found the highest impact scores were among members who did not provide their demographic information on intake forms. This makes it difficult to interpret differences and highlights an area for future research when more data is available.

Research to date on the impact of care navigation for LGBTQ+ communities has been limited to small samples and focused on subsets of the communities, specifically those receiving care for chronic diseases (eg, HIV) or gender-affirming care [27,28]. This formative evaluation's results are encouraging, and similar tailored care navigation services have the potential to broaden the reach and scalability of health care access, particularly benefiting LGBTQ+ populations who often face challenges in navigating their health care needs. While promising, we recognize that the observed differences between respondents and nonrespondents could introduce biases into our findings, particularly concerning regional disparities and engagement levels. For instance, even though we observed no differences between respondents and nonrespondents in the use of care navigation services, the participants included in our analyses were less likely to be from the South. Moreover, members with multiple cases were more likely to provide feedback. These patterns suggest that certain groups may face barriers to participation, which could skew the overall results. To address this, we recommend targeted outreach efforts aimed at increasing survey participation among underrepresented groups, particularly those in the South and individuals with fewer cases. By implementing these strategies, we aim to create a more balanced dataset in future research, which would ultimately improve the generalizability and inclusivity of the study's conclusions.

Overall, continuous evaluation of these services can serve as a supplementary tool for employers seeking to promote health equity and improve belonging among employees. This is particularly important as discrimination and stigma against these communities persist in the United States. Such services may help reduce barriers by lowering out-of-pocket costs, allowing employees to access care without financial strain, while also addressing both insurance policy limitations and state-level policy constraints [28]. Furthermore, they show promise for enhanced health literacy among employees [41], which may empower them to navigate the health care system and make informed decisions about their health. Therefore, scalable and system-level changes that use navigation services are essential to reaching a larger proportion of the LGBTQ+ population.

Limitations

This study is not without limitations. First, the results may lack generalizability. Included Health was used as a case study, and the members who completed member-reported outcomes differed from those who did not. Second, there may be survey-response bias or a ceiling effect in member-reported outcomes. We recognize that a subset of members provided follow-up feedback through the surveys. Future research examining strategies to encourage increased survey completion rates among members may offer greater insights into patient-reported outcomes related to the Included Health LGBTQ+ Navigation service. Third, while members were asked to report the impact since working with Included Health, member-reported outcomes were solicited at case completion

XSL•FO
RenderX

only and do not reflect changes from a presurvey to postsurvey. Fourth, some members had access to additional Included Health services, such as virtual care, which may have influenced the responses to the survey. However, less than 7% of respondents used virtual care services and the survey came directly from their LGBTQ+ care coordinator after their case was closed, limiting the likelihood that the survey captured their experience with other Included Health services. Fifth, we were unable to differentiate the impact of LGBTQ+ Navigation by the multiple race and ethnicity, sexual orientation, and gender identities captured in the intake form due to small sample sizes and the need for mutually exclusive categories to support statistical testing. However, we felt it was important to examine if any differences were present by subgroup definitions with the data available. By consolidating members into broader race, sexual orientation, and gender identity categories, we were able to maximize data availability and better understand the differences between groups. Finally, most members did not provide optional demographic characteristics, which limited our ability to draw conclusions from the subgroup analyses. Moreover, we were unable to distinguish whether the examined case was for the policy member or for their family members since the data was not initially collected for research purposes. Despite these limitations, the strengths of the study include a national sample of members from the LGBTQ+ communities and an initial assessment of the impact of navigation on key process outcomes that can lead to improved health and reduced costs.

Conclusion

Health care navigation services can offer advocacy and guidance to members by facilitating connections to verified in-network providers, focusing on benefits navigation tailored to LGBTQ+ health needs and providing education and advocacy for both clinical and nonclinical needs. The Included Health LGBTQ+ Navigation service demonstrates potential public health utility and value. Further research is needed to examine the use of care navigation within segments of the LGBTQ+ communities. Moreover, investigating how care navigators addressed specific barriers will inform future navigation opportunities and policy development.

Acknowledgments

The authors are grateful to all the members and care coordinators of Included Health. We also recognize the contributions of staff in Included Health and the University of Pennsylvania. In particular, the authors would like to thank Fiona Weeks, Cass Wahl, and Zachary Taylor at Included Health and James Wolfe and Xandro Xu at Eidos LGBTQ+ Health Initiative, University of Pennsylvania.

Data Availability

The data sets generated and analyzed for this study are not publicly available due to them containing proprietary information of Included Health. The data is confidential and was used under the data use agreement between University of Pennsylvania and Included Health for the current study. Data access can be granted to authorized individuals subject to the company's approval and a nondisclosure agreement.

Authors' Contributions

SKC performed the analysis and led the writing of the manuscript. JM gathered the data and contributed to the conceptualization and drafting of this manuscript. PST contributed to the conceptualization and drafting of this manuscript. AP cleaned and analyzed the data and contributed to the drafting of this manuscript. JB contributed to the conceptualization and drafting of this manuscript. All authors provided feedback and helped shape the research, analysis, and manuscript.

Conflicts of Interest

JM is employed by Included Health. All other authors declare that they do not have competing interests.

Multimedia Appendix 1
Original member responses to optional gender identity and sexual orientation fields within member request form (n=176).
[DOCX File, 20 KB - formative_v9i1e64137_app1.docx]

Checklist 1
STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist of items that should be included in reports of cohort studies.
[DOCX File, 34 KB - formative_v9i1e64137_app2.docx]

References

1. Zeeman L, Sherriff N, Browne K, et al. A review of lesbian, gay, bisexual, trans and intersex (LGBTI) health and healthcare inequalities. Eur J Public Health 2019 Oct 1;29(5):974-980. [doi: 10.1093/eurpub/cky226] [Medline: 30380045]

XSL•FO
RenderX

2. Bariola E, Lyons A, Leonard W, Pitts M, Badcock P, Couch M. Demographic and psychosocial factors associated with psychological distress and resilience among transgender individuals. Am J Public Health 2015 Oct;105(10):2108-2116. [doi: 10.2105/AJPH.2015.302763] [Medline: 26270284]
3. Fredriksen-Goldsen KI, Kim HJ, Barkan SE, Muraco A, Hoy-Ellis CP. Health disparities among lesbian, gay, and bisexual older adults: results from a population-based study. Am J Public Health 2013 Oct;103(10):1802-1809. [doi: 10.2105/AJPH.2012.301110] [Medline: 23763391]
4. Meyer IH. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. Psychol Bull 2003 Sep;129(5):674-697. [doi: 10.1037/0033-2909.129.5.674] [Medline: 12956539]
5. Ferlatte O, Salway T, Samji H, et al. An application of syndemic theory to identify drivers of the syphilis epidemic among gay, bisexual, and other men who have sex with men. Sex Transm Dis 2018 Mar;45(3):163-168. [doi: 10.1097/OLQ.0000000000000713] [Medline: 29420444]
6. Kamen C, Palesh O, Gerry AA, et al. Disparities in health risk behavior and psychological distress among gay versus heterosexual male cancer survivors. LGBT Health 2014 Jun;1(2):86-92. [doi: 10.1089/lgbt.2013.0022] [Medline: 26789618]
7. Kamen CS, Smith-Stoner M, Heckler CE, Flannery M, Margolies L. Social support, self-rated health, and lesbian, gay, bisexual, and transgender identity disclosure to cancer care providers. Oncol Nurs Forum 2015 Jan;42(1):44-51. [doi: 10.1188/15.ONF.44-51] [Medline: 25542320]
8. Chan ASW, Leung LM, Wong FKC, et al. Needs and experiences of cancer care in patients' perspectives among the lesbian, gay, bisexual, transgender and queer community: a systematic review. Soc Work Health Care 2023;62(8-9):263-279. [doi: 10.1080/00981389.2023.2226182] [Medline: 37338823]
9. Hsieh N, Shuster SM. Health and health care of sexual and gender minorities. J Health Soc Behav 2021 Sep;62(3):318-333. [doi: 10.1177/00221465211016436] [Medline: 34528481]
10. National Academies of Sciences, Engineering, and Medicine. Understanding the Well-Being of LGBTQI+ Populations: National Academies Press; 2020.
11. Dodge B, Herbenick D, Friedman MR, et al. Attitudes toward bisexual men and women among a nationally representative probability sample of adults in the United States. PLoS ONE 2016;11(10):e0164430. [doi: 10.1371/journal.pone.0164430] [Medline: 27783644]
12. Grant J, Mottet L, Tanis J. National transgender discrimination survey report on health and health care. National LGBT Cancer Network. 2010. URL: https://cancer-network.org/wp-content/uploads/2017/02/National_Transgender_Discrimination_Survey_Report_on_health_and_health_care.pdf [accessed 2025-01-02]
13. Lund EM, Burgess CM. Sexual and gender minority health care disparities: barriers to care and strategies to bridge the gap. Prim Care 2021 Jun;48(2):179-189. [doi: 10.1016/j.pop.2021.02.007] [Medline: 33985698]
14. Levesque JF, Harris MF, Russell G. Patient-centred access to health care: conceptualising access at the interface of health systems and populations. Int J Equity Health 2013 Mar 11;12:18. [doi: 10.1186/1475-9276-12-18] [Medline: 23496984]
15. Cruz TM. Assessing access to care for transgender and gender nonconforming people: a consideration of diversity in combating discrimination. Soc Sci Med 2014 Jun;110:65-73. [doi: 10.1016/j.socscimed.2014.03.032] [Medline: 24727533]
16. Garbers S, Heck CJ, Gold MA, Santelli JS, Bersamin M. Providing culturally competent care for LGBTQ youth in school-based health centers: a needs assessment to guide quality of care improvements. J Sch Nurs 2018 Dec;34(6):424-429. [doi: 10.1177/1059840517727335] [Medline: 28830333]
17. Conron KJ, Mimiaga MJ, Landers SJ. A population-based study of sexual orientation identity and gender differences in adult health. Am J Public Health 2010 Oct;100(10):1953-1960. [doi: 10.2105/AJPH.2009.174169] [Medline: 20516373]
18. Durso LE, Meyer IH. Patterns and predictors of disclosure of sexual orientation to healthcare providers among lesbians, gay men, and bisexuals. Sex Res Social Policy 2013 Mar 1;10(1):35-42. [doi: 10.1007/s13178-012-0105-2] [Medline: 23463442]
19. Holland D, White LCJ, Pantelic M, Llewellyn C. The experiences of transgender and nonbinary adults in primary care: a systematic review. Eur J Gen Pract 2024 Dec;30(1):2296571. [doi: 10.1080/13814788.2023.2296571] [Medline: 38197305]
20. Chong LSH, Kerklaan J, Clarke S, et al. Experiences and perspectives of transgender youths in accessing health care: a systematic review. JAMA Pediatr 2021 Nov 1;175(11):1159-1173. [doi: 10.1001/jamapediatrics.2021.2061] [Medline: 34279538]
21. Wittgens C, Fischer MM, Buspavanich P, Theobald S, Schweizer K, Trautmann S. Mental health in people with minority sexual orientations: a meta-analysis of population-based studies. Acta Psychiatr Scand 2022 Apr;145(4):357-372. [doi: 10.1111/acps.13405] [Medline: 35090051]
22. Yu H, Flores DD, Bonett S, Bauermeister JA. LGBTQ + cultural competency training for health professionals: a systematic review. BMC Med Educ 2023 Aug 9;23(1):558. [doi: 10.1186/s12909-023-04373-3] [Medline: 37559033]
23. Holt NR, Eldridge-Smith ED, Griffin JA, Stepleman LM, Drescher CF, Casanova T. Differences in health care access, utilization, and experiences among LGBTQ+ subgroups in the Southern United States. Fam Community Health 2023;46(1):58-68. [doi: 10.1097/FCH.0000000000000340] [Medline: 35943219]
24. Hatzenbuehler ML, Lattanner MR, McKetta S, Pachankis JE. Structural stigma and LGBTQ+ health: a narrative review of quantitative studies. Lancet Public Health 2024 Feb;9(2):e109-e127. [doi: 10.1016/S2468-2667(23)00312-2] [Medline: 38307678]

XSL•FO
RenderX

25. Butkus R, Rapp K, Cooney TG, Engel LS, Health and Public Policy Committee of the American College of Physicians. Envisioning a better U.S. health care system for all: reducing barriers to care and addressing social determinants of health. Ann Intern Med 2020 Jan 21;172(2 Suppl):S50-S59. [doi: 10.7326/M19-2410] [Medline: 31958803]
26. Johnson CV, Mimiaga MJ, Bradford J. Health care issues among lesbian, gay, bisexual, transgender and intersex (LGBTI) populations in the United States: introduction. J Homosex 2008;54(3):213-224. [doi: 10.1080/00918360801982025] [Medline: 18825859]
27. Roland KB, Higa DH, Leighton CA, Mizuno Y, DeLuca JB, Koenig LJ. Client perspectives and experiences with HIV patient navigation in the United States: a qualitative meta-synthesis. Health Promot Pract 2020 Jan;21(1):25-36. [doi: 10.1177/1524839919875727] [Medline: 31597497]
28. Dozier R. How navigators influence insurance coverage for gender-affirming surgeries: a qualitative study. Perm J 2023 Mar 15;27(1):72-76. [doi: 10.7812/TPP/22.115] [Medline: 36916175]
29. Zamudio-Haas S, Koester K, Venegas L, et al. "Entre Nosotras:" a qualitative study of a peer-led PrEP project for transgender latinas. BMC Health Serv Res 2023 Sep 20;23(1):1013. [doi: 10.1186/s12913-023-09707-x] [Medline: 37730598]
30. Matthews K, Phelan J, Jones M, et al. 2015 national content test race and ethnicity analysis report. US Census Bureau.: U.S. Department of Commerce; 2017. URL: https://www.documentcloud.org/documents/4316468-2015nct-Race-Ethnicity-Analysis [accessed 2023-12-01]
31. Bauermeister J, Choi SK, Bruehlman-Senecal E, et al. An identity-affirming web application to help sexual and gender minority youth cope with minority stress: pilot randomized controlled trial. J Med Internet Res 2022 Aug 1;24(8):e39094. [doi: 10.2196/39094] [Medline: 35916700]
32. National Academies of Sciences, Engineering, and Medicine. Measuring Sex, Gender Identity, and Sexual Orientation: National Academies Press; 2022. [doi: 10.17226/26424]
33. McKay T, Tran NM, Barbee H, Min JK. Association of affirming care with chronic disease and preventive care outcomes among lesbian, gay, bisexual, transgender, and queer older adults. Am J Prev Med 2023 Mar;64(3):305-314. [doi: 10.1016/j.amepre.2022.09.025] [Medline: 36460525]
34. Khalili J, Leung LB, Diamant AL. Finding the perfect doctor: identifying lesbian, gay, bisexual, and transgender–competent physicians. Am J Public Health 2015 Jun;105(6):1114-1119. [doi: 10.2105/AJPH.2014.302448] [Medline: 25880937]
35. Hostetter J, Schwarz N, Klug M, Wynne J, Basson MD. Primary care visits increase utilization of evidence-based preventative health measures. BMC Fam Pract 2020 Jul 28;21(1):151. [doi: 10.1186/s12875-020-01216-8] [Medline: 32718313]
36. Gao J, Moran E, Grimm R, Toporek A, Ruser C. The effect of primary care visits on total patient care cost: evidence from the Veterans Health Administration. J Prim Care Community Health 2022;13:21501319221141792. [doi: 10.1177/21501319221141792] [Medline: 36564889]
37. Finneran P, Toribio MP, Natarajan P, Honigberg MC. Delays in accessing healthcare across the gender spectrum in the All of Us research program. J Gen Intern Med 2024 May;39(7):1156-1163. [doi: 10.1007/s11606-023-08548-y] [Medline: 38049655]
38. U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion. National action plan to improve health literacy. US Department of Health and Human Services. 2010. URL: https://health.gov/sites/default/files/2019-09/Health_Literacy_Action_Plan.pdf [accessed 2023-12-14]
39. Doblecki-Lewis S, Butts S, Botero V, Klose K, Cardenas G, Feaster D. A randomized study of passive versus active PrEP patient navigation for a heterogeneous population at risk for HIV in South Florida. J Int Assoc Provid AIDS Care 2019;18:2325958219848848. [doi: 10.1177/2325958219848848] [Medline: 31131679]
40. Hostetter CR, Call J, Gerke DR, Holloway BT, Walls NE, Greenfield JC. "We are doing the absolute most that we can, and no one is listening": barriers and facilitators to health literacy within transgender and nonbinary communities. Int J Environ Res Public Health 2022 Jan 22;19(3):1229. [doi: 10.3390/ijerph19031229] [Medline: 35162254]
41. Loignon C, Dupéré S, Fortin M, Ramsden VR, Truchon K. Health literacy - engaging the community in the co-creation of meaningful health navigation services: a study protocol. BMC Health Serv Res 2018 Jun 28;18(1):505. [doi: 10.1186/s12913-018-3315-3] [Medline: 29954407]

Abbreviations

IRB: Institutional Review Board
LGBTQ+: lesbian, gay, bisexual, transgender, queer, and others
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
WPATH: World Professional Association for Transgender Health

XSL•FO
RenderX

Edited by A Mavragani; submitted 09.07.24; peer-reviewed by DZ Nowaskie, M Cabell Jonas; revised version received 02.12.24; accepted 02.12.24; published 09.01.25.

Please cite as:
Choi SK, Marshall J, Sexton Topper P, Pregnall A, Bauermeister J
Impact of a Virtual Care Navigation Service on Member-Reported Outcomes Among Lesbian, Gay, Bisexual, Transgender, and Queer Populations: Case Study
JMIR Form Res 2025;9:e64137
URL: https://formative.jmir.org/2025/1/e64137
doi:10.2196/64137

© Seul Ki Choi, Jaclyn Marshall, Patrina Sexton Topper, Andrew Pregnall, José Bauermeister. Originally published in JMIR Formative Research (https://formative.jmir.org), 9.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Mobile App for Promoting Breastfeeding-Friendly Communities in Hong Kong: Design and Development Study

Heidi Sze Lok Fan[1], BNurs, MPhil, PhD; Emily Tsz Yan Leung[2], BA, BEd; Ka Wing Lau[2], BSc; Janet Yuen Ha Wong[3], BNurs, MNurs, PhD; Edmond Pui Hang Choi[2], BNurs, MPhil, PhD; Christine Lam[4], RM, BHS, MMid, IBCLC; Marie Tarrant[1], BN, MN, MPH, PhD; Hextan Yuen Sheung Ngan[5], MBBS, MD; Patrick Ip[6], MBBS, DCH, Dip Med; Chia Chin Lin[2], BSN, MSN, EMBA, PhD; Kris Yuet Wan Lok[2], BSc, MSc, PhD

[1]School of Nursing, Faculty of Health and Social Development, The University of British Columbia, Kelowna, BC, Canada

[2]School of Nursing, Li Ka Shing Faculty of Medicine, University of Hong Kong, 5/F, Academic Building, Pokfulam, Hong Kong, China (Hong Kong)

[3]School of Nursing and Health Sciences, Hong Kong Metropolitan University, Hong Kong, China (Hong Kong)

[4]Department of Obstetrics and Gynaecology, Queen Elizabeth Hospital, Kowloon, Hong Kong, China (Hong Kong)

[5]Department of Obstetrics and Gynaecology, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, China (Hong Kong)

[6]Li Ka Shing Faculty of Medicine, Department of Paediatrics and Adolescent Medicine, University of Hong Kong, Hong Kong, China (Hong Kong)

Corresponding Author:
Kris Yuet Wan Lok, BSc, MSc, PhD
School of Nursing, Li Ka Shing Faculty of Medicine, University of Hong Kong, 5/F, Academic Building, Pokfulam, Hong Kong, China (Hong Kong)

Abstract

Background: Breastfeeding is vital for the health and well-being of both mothers and infants, and it is crucial to create supportive environments that promote and maintain breastfeeding practices.

Objective: The objective of this paper was to describe the development of a breastfeeding-friendly app called “bfGPS” (HKU TALIC), which provides comprehensive territory-wide information on breastfeeding facilities in Hong Kong, with the goal of fostering a breastfeeding-friendly community.

Methods: The development of bfGPS can be categorized into three phases, which are (1) planning, prototype development, and preimplementation evaluation; (2) implementation and updates; and (3) usability evaluation. In phase 1, a meeting was held with experts, including maternal and child health researchers, app developers, breastfeeding individuals, and health professionals, to discuss the focus and functionality of the breastfeeding app. A prototype was developed, and breastfeeding facilities in various public venues in Hong Kong were assessed using a structured checklist. For the preimplementation evaluation, 10 focus groups and 19 one-on-one interviews were conducted between May 2019 and October 2020 with staff working in public premises (n=29) and breastfeeding individuals (n=29). For phase 2, bfGPS was published on iOS (Apple Inc) and Android (Google) platforms in September 2020. App updates were launched in September 2021 and May 2022 based on the suggestions provided by the participants in the preimplementation evaluation. For the usability evaluation, semistructured, in-depth, one-to-one interviews were conducted with breastfeeding individuals (n=30) to understand their experiences of using bfGPS. Content analysis was used to analyze the data.

Results: bfGPS is a mobile app that was developed to assist breastfeeding individuals in locating breastfeeding facilities in public venues in Hong Kong. In the preimplementation evaluation, the participants gave comments on the layout and interface of bfGPS, and suggestions were given on incorporating new functions into the app. Based on the suggestions of the participants in the preimplementation evaluation, a few additional functions were added into bfGPS, including allowing the users to rate and upload recent information about breastfeeding facilities and an infant tracker function that encourages users to record infant development. In the usability evaluation, 3 main themes emerged—bfGPS improves the community experience for breastfeeding individuals, facilitates tracking the infant’s growth, and provides suggestions for further development.

Conclusions: The bfGPS app is the first user-friendly tool designed to assist users in locating breastfeeding facilities within the community. It stands as a guide for similar health care app developments, emphasizing the importance of accurate, current data to ensure user adoption and long-term use. The app’s potential lies in the support and reinforcement of breastfeeding practices coupled with self-management strategies.

(JMIR Form Res 2025;9:e64191) doi:10.2196/64191

XSL•FO
RenderX

KEYWORDS

Baby-Friendly Community Initiative; Baby-Friendly Hospital Initiative; breastfeeding; community; stakeholders; mobile app; friendly communities; baby-friendly; well-being; mother; infant; application; mHealth; qualitative; user-friendly; self-management

Introduction

Background

Breastfeeding provides numerous benefits to both parents and infants [1]. It is considered one of the most effective health interventions for preventing infant and child mortality [2]. The World Health Organization (WHO) recommends that infants be exclusively breastfed for the first 6 months of life. After 6 months of age, complementary feeding should be provided with continuing breastfeeding up to 2 years of age or beyond [3]. However, breastfeeding individuals face numerous challenges that can lead to early weaning. Less than one-half of infants under 6 months old were exclusively breastfed globally, and only 45% of infants can be breastfed up to 2 years of age [4]. The determinants of breastfeeding continuation are multifactorial, with various demographic [5,6], psychological [7], and social factors [8].

Breastfeeding support is crucial for long-term breastfeeding. Technology-based interventions effectively improve breastfeeding outcomes [9,10]. However, a meta-analysis of randomized controlled trials found that mobile app usage was not significantly associated with prolonged breastfeeding duration. However, only a few studies were included in the meta-analysis, and there was high heterogeneity among the interventions [11]. Recently published randomized controlled trials showed that the use of mobile apps can significantly increase [12] exclusive breastfeeding duration [13].

The Baby-Friendly Hospital Initiative (BFHI) was launched by the WHO and the United Nations International Children's Emergency Fund in 1991 [14]. A large body of evidence shows that the BFHI can significantly improve breastfeeding outcomes [15,16]. The BFHI comprises 10 steps that aim to promote, protect, and support breastfeeding in a hospital setting [14]. Step 10 of the BFHI is about fostering the establishment of breastfeeding support groups in the community and referring breastfeeding individuals to them on discharge from the hospital or clinic [14]. Community support is essential for extending breastfeeding duration [17]. The Baby-Friendly Community Initiative (BFCI) is an extension of the BFHI [18] and is focused on providing community-based support for breastfeeding individuals [19]. However, the BFCI is implemented in only a few countries and resources from the government are needed to support the BFCI [19].

Globally, breastfeeding in public is not yet normalized. In Western countries, the proportion of individuals supporting breastfeeding in public ranges from 50% to 75% [20-23]. In China, although 65% of the community agrees that breastfeeding in public is acceptable, 30% believe that women should breastfeed only in private spaces [24]. Furthermore, breastfeeding individuals face numerous challenges when breastfeeding in public. Embarrassment has been identified as a key barrier to public breastfeeding [25] and unwanted attention and unsuitable structural environments are common challenges faced outside of the home [26]. Studies show that some public areas are unwelcoming to breastfeeding individuals [27], highlighting the lack of comfortable breastfeeding spaces [28]. A study has shown that breastfeeding in public is associated with a longer duration of breastfeeding [29]. It is crucial for breastfeeding individuals to find breastfeeding-friendly places when they visit public venues. Several mobile apps are available for breastfeeding individuals to locate breastfeeding facilities in their communities, such as "FeedFinder (GarbTech)," primarily used in the United Kingdom [30] and "Moommae (Moon Active)" in Thailand [31]. These apps can increase the confidence of breastfeeding individuals to go out and breastfeed [30] while supporting breastfeeding on demand [32]. However, users have emphasized the need for more breastfeeding facilities to be included in the app [30,32]. Therefore, we developed a social media mobile app with comprehensive territory-wide information on breastfeeding facilities in Hong Kong to promote breastfeeding-friendly attitudes within the community.

Breastfeeding Practices and Public Breastfeeding in Hong Kong

In Hong Kong, approximately 84.7% of postpartum individuals initiate breastfeeding [33]. However, only 23.9% of parents are able to continue to breastfeed for 12 months and only 22.2% exclusively breastfeed until 6 months [34]. Breastfeeding individuals face various barriers in public places. Frequently reported unpleasant experiences included being stared at by others and being asked to breastfeed in unsanitary or unsuitable places, such as public restrooms [35]. A significant number of women stated that they reduced going out with their children and resorted to feeding them with infant formula to avoid public breastfeeding [35]. However, given that supplementing with infant formula is strongly associated with breastfeeding cessation [36,37], this is not an optimal solution. One study demonstrated breastfeeding in public was associated with longer breastfeeding duration [29]. Therefore, it is crucial to offer community support.

The objective of this paper was to describe the development of a breastfeeding-friendly app called "bfGPS (HKU TALIC)," which provides comprehensive territory-wide information on breastfeeding facilities in Hong Kong, with the goal of fostering a breastfeeding-friendly community. We also gain insights into how breastfeeding individuals and public venue staff perceive bfGPS through qualitative interviews.

Methods

Design

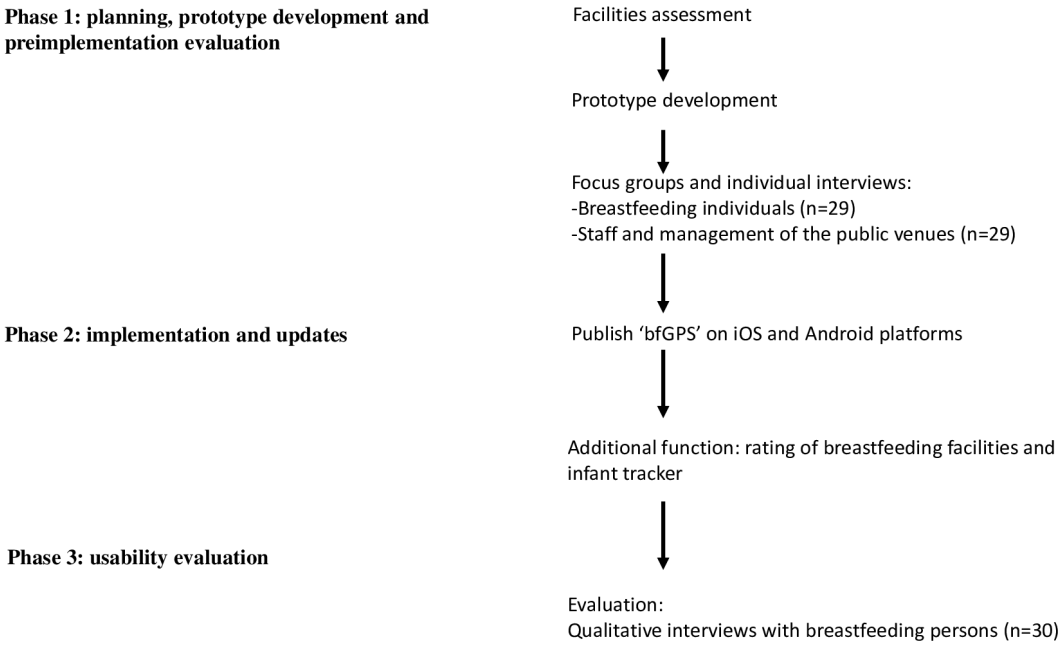
This study was conducted as part of the BFCI, which aims to promote breastfeeding-friendly attitudes in the community. The BFCI program encompasses the following activities: (1) providing training to the staff and management of public venues to support breastfeeding individuals, (2) offering breastfeeding support to pregnant and breastfeeding individuals, and (3)

XSL•FO
RenderX

developing the bfGPS mobile app that aims to assist breastfeeding individuals in locating breastfeeding facilities in public venues in Hong Kong. The details of the BFCI program were described elsewhere [38].

In this study, we report the development of bfGPS, which is divided into three phases: (1) planning, prototype development, and preimplementation evaluation; (2) implementation and updates; and (3) usability evaluation (Figure 1). Phase 1 involves facilities assessment, the development of the prototype of bfGPS, and a preimplementation evaluation. The preimplementation evaluation involves focus groups and individual interviews to collect feedback on the app. For phase 2 of the development, we published bfGPS on iOS (Apple Inc) and Android (Google) platforms, and we updated the apps to add additional functions, which are rating of breastfeeding facilities and infant tracker. Phase 3 is a usability evaluation. It includes qualitative interviews with breastfeeding individuals to understand their experiences of using bfGPS.

Figure 1. Flow diagram of the development of bfGPS.



Phase 1: Planning, Prototyping, and Preimplementation Evaluation

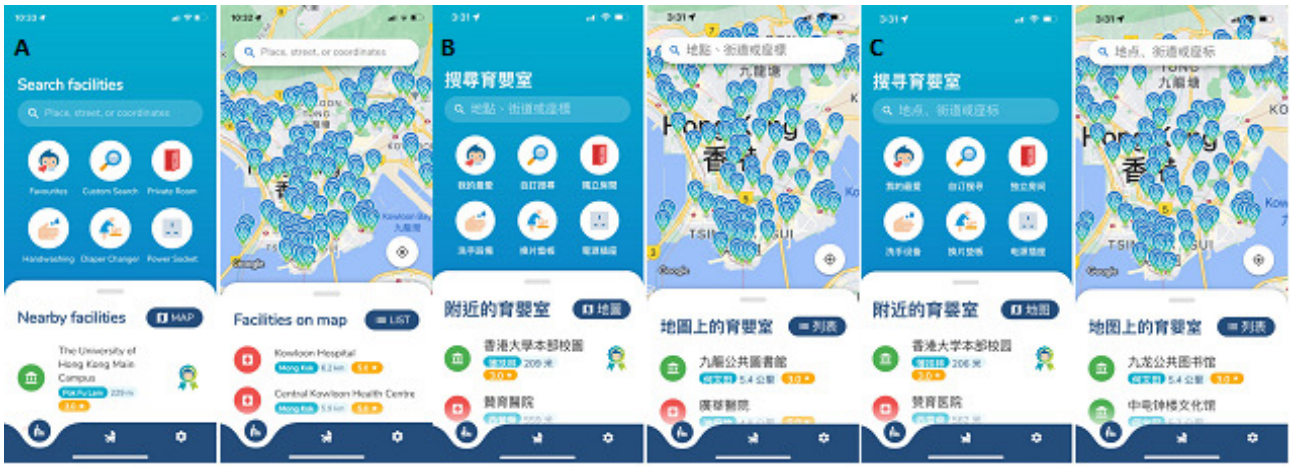
Setting and Relevant Context

A meeting was held with experts, including maternal and child health researchers, app developers, breastfeeding individuals, and health professionals, to discuss the focus and functionality of the breastfeeding app. Based on the results of the Department of Health (2016) survey [35], the team decided to design an app that could offer user-friendly guidance on breastfeeding to the general public.

The content of the app was outlined, and a prototype of the bfGPS app was developed (Figure 2). bfGPS is available in English, Traditional Chinese, and Simplified Chinese. Between June 2019 and June 2020, undergraduate nursing students assessed breastfeeding facilities in various public venues in Hong Kong, such as government premises, shopping malls, and metro stations, using a structured checklist.

XSL•FO
RenderX

Figure 2. The interface of bfGPS for locating breastfeeding facilities is available in (**A**) English, (**B**) traditional Chinese, and (**C**) simplified Chinese. bfGPS provides a search bar that enables users to search breastfeeding facilities across Hong Kong. Users can view nearby facilities and the distance from their current location by enabling location access. The details of nearby facilities are shown with icons and descriptions. In addition, bfGPS allows for customized searches to enable users to filter results based on their needs, such as private rooms or handwashing facilities.



After developing the bfGPS prototype, breastfeeding individuals and staff from public venues were invited to participate in preimplementation evaluation. This preimplementation evaluation included focus groups and one-on-one interviews to gather feedback on the app's functionality, design, performance, and suggestions for further development of bfGPS. Both public venue staff and breastfeeding individuals were recruited, as it was important to confirm the clarity of the content related to the public premises.

Recruitment and Data Collection

A total of 29 public facility staff members and 29 breastfeeding individuals were invited to participate in the focus groups and one-on-one interviews between May 2019 and October 2020. Convenience sampling was used to recruit participants. Public venue staff were recruited through social media advertisements or by participating in staff training within the BFCI program. Staff training for the BFCI program was described elsewhere [38]. In brief, a convenience sample of shopping malls and premises was identified throughout Hong Kong using the internet search. Invitation letters were sent to senior managers and customer service of the identified shopping malls and premises. The emails and letters of invitation include a description of the BFCI program, which included exposure to a breastfeeding mobile app, bfGPS. The shopping malls and premises employees were invited to attend a 2-hour training course based on the WHO breastfeeding counseling course. The employees who attended the BFCI training were invited to participate in the preimplementation evaluation through SMS text messages or emails.

The staff members had diverse occupations and job roles, ranging from frontline staff to managers. The inclusion criteria for staff members were individuals who were 18 years of age or older, worked in public venues, and interacted with the public as part of their jobs. Breastfeeding individuals were recruited if they had breastfed their infants in public or had used breastfeeding facilities in public venues. The recruitment was conducted through the researchers' network. Participants received an HK $100 (approximately US $12.80) gift card as compensation.

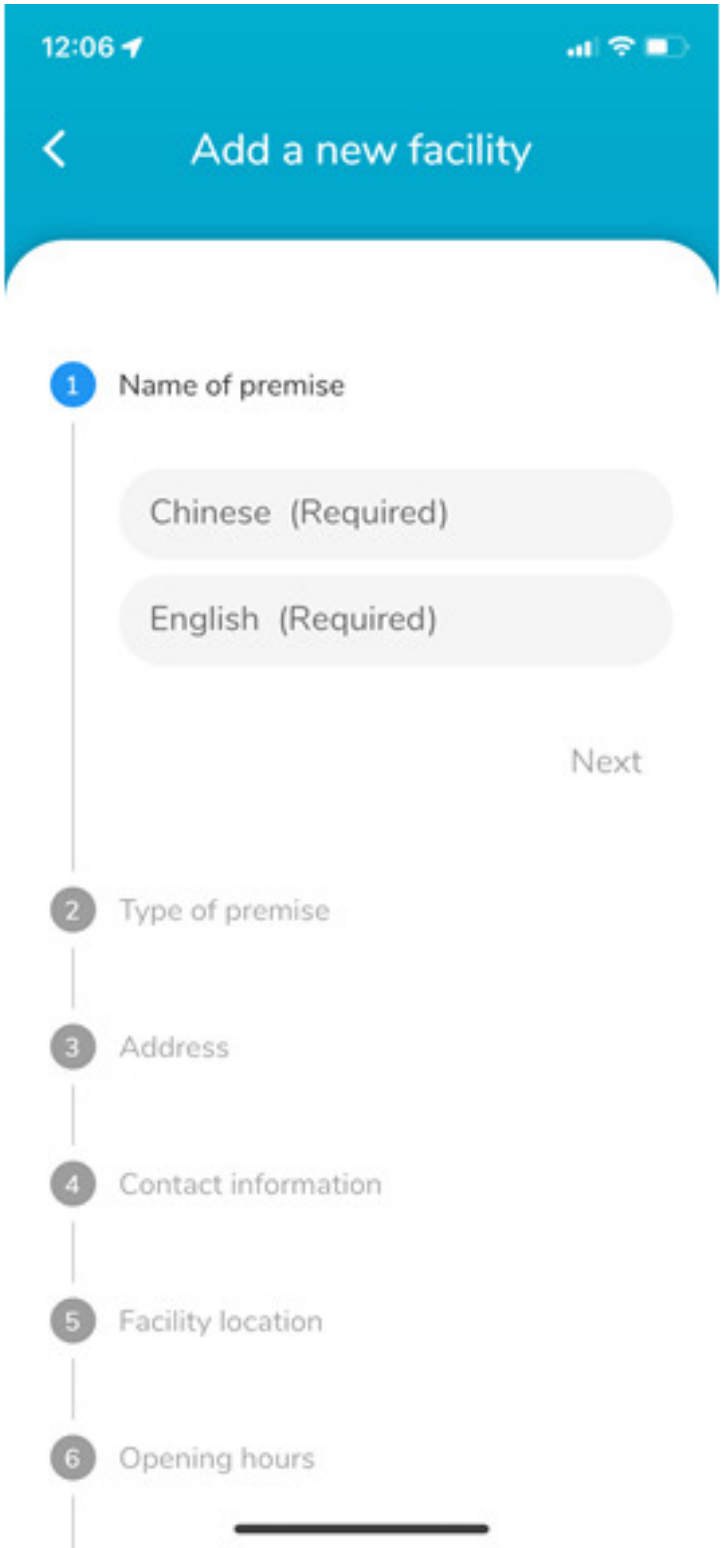
Phase 2: Implementation and Updates (Setting and Relevant Context)

Beta testing was conducted to provide early versions of the app to experienced app testers who examined it for errors, crashes, layout issues, software bugs, or other problems. The research team and app developers performed beta testing. bfGPS was published on iOS and Android platforms in September 2020. An app update was launched in September 2021 which allowed users to rate and evaluate breastfeeding facilities and included an online function for the public to upload recent information about breastfeeding facilities.

Based on the participants' suggestions in the focus groups and one-on-one interviews, the ratings and evaluation stage were introduced in September 2021, along with an online function for the public to upload recent information about breastfeeding facilities (Figure 3). The infant tracker was a new add-on feature launched in May 2022 to encourage users to record infant development, including the amount of breast milk, breastfeeding times, sleep patterns, and reminders for feeding and vaccinations (Figure 4).

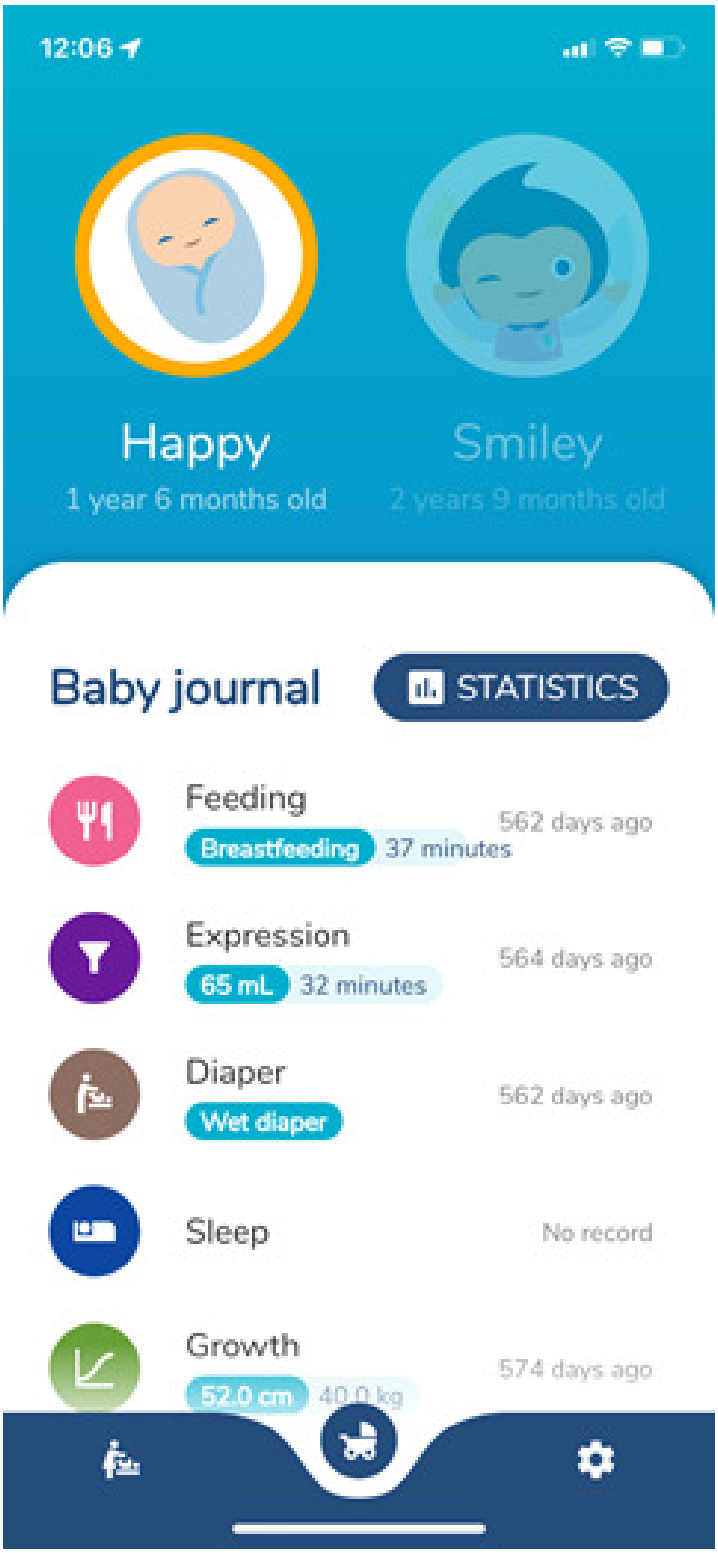
XSL•FO
RenderX

Figure 3. Interactive function for the users to upload information related to breastfeeding facilities in bfGPS. Users can submit information about breastfeeding facilities, such as name, type of premise, address, and photos of the premises. The study team then reviews the information to ensure accuracy before updating the app. This approach helps maintain up-to-date and reliable data for bfGPS users.



XSL•FO
RenderX

Figure 4. The infant tracker function of bfGPS. The infant tracker enables users to create multiple infant profiles. Key functions include recording infant feeding, human milk expression, diaper changes, sleep patterns, growth metrics, and immunization schedules. The growth patterns are plotted on growth charts using data from the Hong Kong Growth Study 2020 [39]. In addition, the infant tracker calculates upcoming vaccination dates based on the infant's birth date, following the schedule of the Hong Kong Childhood Immunization Programme [40]. The infant record can be exported in a PDF file.



Phase 3: Usability Evaluation (Setting and Relevant Context, Recruitment, and Data Collection)

Following the introduction of the infant tracker function, qualitative interviews were conducted with 30 breastfeeding individuals who used location mapping and the infant tracker function from August to September 2022. Breastfeeding individuals were recruited through Facebook (Meta) and Instagram (Meta) posts containing brief descriptions of the study and a link for study registration. The following criteria were used: (1) 18 years of age or older, (2) Cantonese speaking, (3) pregnant or recently gave birth to a live newborn, (4) Hong

XSL•FO
RenderX

Kong residents, and (5) had ever used bfGPS. In the study registration form, participants were asked to indicate which functions did they used in bfGPS. Data saturation was achieved, as no new information emerged from the interviews [41]. Participants received HK $100 (approximately US $12.80) in coupons for a local grocery store as an honorarium for participating in the interviews.

Data Collection

Semistructured interview guides with open-ended questions were used to collect the data. The participants' experiences of using bfGPS, their insights into the strengths and limitations of bfGPS, recommendations for future development, and significant features for improving the user experience were comprehensively examined. The interview guides were refined as new insights, and perspectives were obtained throughout the study. Focus groups and one-on-one interviews were scheduled based on participants' preferences and conducted at various locations. The interviews were conducted by the authors, who are experienced in conducting qualitative interviews (HSLF and KWL). For the preimplementation evaluation (phase 1), 7 focus groups and 1 staff one-on-one interview were conducted in public commercial rooms and at the university campus. The 3 focus groups and 18 one-on-one interviews with breastfeeding individuals took place in public commercial rooms. The usability evaluation interviews (phase 3) were conducted at public venues convenient for the interviewees (n=30). Each focus group and the one-on-one interview lasted approximately 30 minutes and were audio-recorded with the participants' written permission.

Data Analysis

Descriptive statistics were used to summarize the participants' characteristics. Trained transcriptionists transcribed the audio recordings verbatim into Chinese. The transcripts were reviewed multiple times to increase the researcher's familiarity with the participants' experiences of using bfGPS. The transcripts were then denaturalized and translated into English. Content analysis was used to analyze the transcripts, as it allows for the identification of core consistencies and meanings in qualitative data [42]. The research team reviewed each transcribed interview and developed an open code list derived directly from the data. All textual data were coded and grouped into conceptually similar and overarching themes [43]. To enhance familiarity with the data, a manual data management approach was used, aided by using Microsoft Word [44].

Ethical Considerations

Ethical approval was obtained from the Institutional Review Board of the University of Hong Kong and Hospital Authority Hong Kong West Cluster (UW18-489). The study was conducted in accordance with the principles of the Helsinki Declaration. Participants were informed that they could withdraw from the study at any time without any repercussions. They received HK $100 (approximately US $12.80) in grocery store coupons as an honorarium for interview participation. Written informed consent was obtained from all participants. No specific identifiers were included in the data used for analysis, and participants' confidentiality was maintained at all times.

Results

Breastfeeding Facilities Assessment and Function of bfGPS

A total of 658 breastfeeding facilities in 502 sites were assessed, and the prototype of bfGPS was developed. bfGPS uses the GPS function to locate the nearest breastfeeding facilities. Users can use the advanced search function to narrow their search based on the type of premises and specific equipment provided by the breastfeeding facilities. In addition, photographs of the listed breastfeeding facilities were made available, along with equipment details.

Preimplementation Evaluation of bfGPS

Tables 1 and 2 present the characteristics of the staff members and breastfeeding individuals who participated in the focus groups and one-on-one interviews. Approximately 40% of the staff members were 35 years of age or above (Table 1). The majority of them were female (24/29, 83%), worked full-time (28/29, 97%), and were employed in the service industry (17/29, 59%). Table 2 provides the characteristics of the breastfeeding individuals, revealing that the majority were aged 35 years or older (15/27, 56%), primiparous (16/27, 59%), and returned to work within 6 months post partum (14/27, 52%).

XSL•FO
RenderX

Table . Characteristics of the public facility staff who participated in the preimplementation evaluation.

Demographic variable		Total (N=29), n (%)
Age (years)		
	18 - 29	10 (37)
	30 - 34	6 (22)
	≥35	11 (41)
	Not reported	2 (7)
Sex		
	Female	24 (83)
	Male	5 (17)
Employment status		
	Work full-time	28 (97)
	Work part-time	1 (4)
Occupation		
	Service industry	17 (59)
	Catering industry	4 (14)
	Property management	6 (21)
	Others	2 (7)

Table . Characteristics of the breastfeeding individuals who participated in the preimplementation evaluation.

Demographic variable		Total (N=27[a]), n (%)
Maternal age (years)		
	18 - 29	2 (7)
	30 - 34	10 (37)
	≥35	15 (56)
Parity		
	Primiparous	16 (59)
	Multiparous	11 (41)
Age of the child during the interview		
	<3 months	5 (19)
	3 to <6 months	5 (19)
	6 to <12 months	4 (15)
	≥12 months	13 (48)
Return to work within 6 months post partum		
	No	13 (48)
	Yes	14 (52)

[a]Of total, 2 participants had missing sociodemographic characteristics.

Participants provided feedback on various aspects of bfGPS, including layout, app performance, functionality, and suggestions for further development. Overall, the participants expressed satisfaction with the functions of bfGPS but preferred a more visually appealing layout. They also provided suggestions to enhance user experiences, such as implementing a rating function for the breastfeeding facilities and an infant tracker.

Layout and App Performance

Most participants stated that the layout and interface of bfGPS were user-friendly and easy to use. The app effectively helped users locate breastfeeding facilities. While the participants found the facility photos useful, some expressed concerns about the loading speed and the app's ability to quickly provide the information they needed, such as the location, name, and

XSL•FO
RenderX

equipment details. Furthermore, participants recommended enhancing the visual appeal by improving the layout of bfGPS.

> *With just one click to enter the page, you can choose locations and facilities, which are actually quite convenient. Moreover, opening hours were provided in the app. Parents might find it inconvenient to locate breastfeeding rooms, especially at night, which can be very helpful for them.* [Staff #3, Focus group #1]

> *When using the app for the first time, the premises' name and address should load first…. Waiting for photos to load before showing the premises' name and location can be time-consuming. We want to be fast.* [Breastfeeding individual #7]

> *bfGPS would attract me easily if it has an appealing appearance.* [Breastfeeding individual #6]

Areas of Improvement

Participants were asked to suggest ways to maximize the usage of bfGPS. In addition to the facilities locator function, participants recommended incorporating the following functions into the app: an infant tracker, more up-to-date information, increased uniqueness, and an interface for commenting on and reading other users' feedback.

All participants expressed concerns about the accuracy and timeliness of the information provided by the app. They emphasized that having inaccurate or outdated information would cause inconvenience to them and discourage them from using the app. To address this, participants suggested adding a channel for app users to report updates on breastfeeding facilities. The system administrators can then verify and approve the updates, ensuring the app provides accurate and up-to-date information.

> *I suggest the option for users to upload new baby-care facilities. Alternatively, companies can provide information directly to you (the research team) if they have opened new nursing rooms. You can check the information and upload it to the app after approval. If you need to do all the work by yourself, you may be unable to have a timely update…. This could significantly affect me if the information is outdated.*

> *If I go to the baby-care facility and discover it has been demolished, it will waste my time and mess up my plan.* [Breastfeeding individual #8]

To prolong the usage of the app, participants suggested allowing them to enter feeding or baby information, such as feeding time or body weight of the infants. This feature would incentivize them to continue using the app and make them less likely to delete it.

> *Honestly, my phone always runs out of storage space, making it easy to uninstall and reinstall bfGPS. However, if bfGPS provides more personalized functions, e.g. body weight record, feeding schedule, or which shopping mall I go to. I won't delete the app. If I do not go out with the baby or only go to the shopping malls I am familiar with, I don't see the need to keep the app as it uses up my storage space.* [Breastfeeding individual #19]

> *If the app includes a function for entering the baby's height, weight, and head circumference, we are more likely to use it and keep it for an extended period.* [Breastfeeding individual #1]

In addition to the standard information uploaded by the developers, participants expressed their desire to see the photos uploaded by other users and read their comments about the baby-care facilities.

> *I would like to view photos taken by users.* [Staff #4, Focus group #2]

Based on participants' feedback from the preimplementation evaluation, we integrated an infant tracking function, incorporated public ratings, and added a function for uploading recent breastfeeding facility information in bfGPS. These enhancements were introduced in phase 2 of the study, as outlined in the methods section.

Usability Evaluation

The characteristics of the participants are shown in Table 3. Approximately two-thirds were working full-time, and a majority had a bachelor's degree or higher (17/30, 57%), were 30 - 34 years of age (16/30, 53%), and lived in Hong Kong since birth (27/30, 90%).

XSL•FO
RenderX

Table . Characteristics of participants in the end-user evaluations.

Demographic variable		Total (N=30), n (%)
Maternal age (years)		
	18 - 29	3 (10)
	30 - 34	16 (53)
	≥35	11 (37)
Employment status		
	Full-time	20 (67)
	Part-time	2 (7)
	Stay-at-home parent	8 (27)
Maternal education		
	Lower than university degree	13 (43)
	University degree or above	17 (57)
Length of residence in Hong Kong		
	15 years or above	3 (10)
	Since birth	27 (90)

bfGPS Improves Community Experience

The majority of participants reported that bfGPS is very convenient to use, providing prompt and detailed information about nearby breastfeeding facilities. bfGPS enables users to assess the breastfeeding facilities and plan ahead before going out with their babies. It offers flexibility to parents in their community experiences, as they can quickly locate the breastfeeding rooms.

> *Before using bfGPS, I wouldn't stay outside for too long. I would return home to feed within a few hours. With bfGPS, I know the locations of nursing rooms, allowing me to plan ahead for where I can go later. This enables me to explore farther places and extend the duration of my time going out.* [Participant #14]

In addition, participants appreciated the rating function and the inclusion of photos of breastfeeding facilities, as it allowed them to gain insights into the settings of these facilities.

> *The best feature of bfGPS is the provided photos, which clearly show the condition of the nursing rooms to me.… I think the 'ratings' function is beneficial because only those who have used the rooms would rate. The rating is important to me…I clicked on the page to see the issues of the breastfeeding facilities with 1-star ratings.* [Participant #32]

bfGPS Helps Track the Infants' Growth

bfGPS allows people to track the growth of infants and helps them retrieve infants' feeding records easily, especially during health checkups.

> *bfGPS allows me to input my baby's growth data, and it can generate a growth chart similar to the one in the maternal and child health center's booklet. I find it beneficial, so I continue to record (the growth data) here…. The breastfeeding timer is also useful and convenient. It enables me to track how much milk or the duration that my baby was fed. I can observe the amount is increasing with the graph.* [Participant #1]

> *If my family member is willing to record in bfGPS, I can easily access information about my baby's sleeping times. This allows me to plan her sleep schedule, especially on weekends, as I am the primary caregiver. My family member handwrites it, and I find it less convenient as I have to check the notebook each time.* [Participant #18]

Suggestions on Further Development

Similar to the findings in the preimplementation evaluation, the majority of participants emphasized the importance of providing written comments. In addition, participants highlighted the significance of sharing data from the infant tracker with family members. They suggested giving information on infant feeding and knowledge about infants' growth and parenting into bfGPS.

> *Users would benefit more from leaving comments after using the nursing rooms instead of solely relying on rating with the number of stars. For example, if the room is clean but the music is too loud, I can describe it clearly with words and leave other users to choose whether to use it. When you (the research team) receive the comments, you may communicate them to the person in charge and let them improve it.* [Participant #19]

> *It is important to have the sharing function, especially when other family members or helpers take care of your baby. This allows me to stay informed about how my baby is doing…. It is good to provide parenting information for parents. You may consider collaborating with the maternal and child health centers as they have a webpage on this.* [Participant #9]

XSL•FO
RenderX

> *(I would like to have information related to) the optimal time for introducing solid food or links to articles on the benefits of breastfeeding or other relevant breastfeeding.… It will be too monotonous if bfGPS is solely for record-keeping. It is better to have additional reference material. It will be challenging to continue using the app without information.* [Participant #7]

Evaluation of bfGPS

bfGPS was ranked number 3 in the Hong Kong App Store's "Health and Fitness" free app category in September 2020. The total downloads in the iOS and Android platforms were 11,542 in December 2023. As of December 2023, the app usage data demonstrated impact and behavior change, with more than 11,000 unique users having used the app and performing more than 484,000 facilities searches. On average, each user conducted 50 (SD 7.25) searches, with an average of 4 (SD 2.47) searches per app usage session. All users viewed the facility details over 150,000 (SD 39,827.06) times in total. Each user viewed the detailed information of 19 (SD 3.43) facilities on average, and each session of app usage involved 1.5 (SD 0.77) views on average. The app attained average user ratings of 4.8 and 4.111 (out of 5.0) on the App Store and Google Play, respectively. We collected 2287 ratings of breastfeeding facilities and 217 uploads of breastfeeding facility information.

Discussion

Principal Results

This study provides an overview of the app development process of bfGPS, which primarily focuses on locating breastfeeding facilities in the community. We have refined the app based on the feedback provided by participants in the preimplementation evaluation, incorporating features such as ratings for breastfeeding facilities and the ability to upload recent breastfeeding facility information.

The infant tracker has the potential to increase the participants' engagement. A study conducted in the United States showed that approximately 57% of participants used a mobile app to track infant feeding [45]. An infant tracker provides objective data to breastfeeding individuals regarding infant feeding methods and other aspects of baby care [46]. However, some breastfeeding individuals may express concerns about relying too heavily on the infant tracker feature [46]. Therefore, a large-scale quantitative study may be necessary to understand the effect of infant tracking on the well-being of breastfeeding individuals and its impact on breastfeeding outcomes.

For the future development of bfGPS and further research, we aim to transform it into a comprehensive platform that empowers users with access to breastfeeding knowledge and information, ultimately enhancing their self-efficacy and improving breastfeeding outcomes. We plan to collaborate with lactation consultants, registered nurses, midwives, breastfeeding individuals, and maternal and child health researchers to develop videos and informational leaflets. These resources will provide valuable information to help breastfeeding individuals overcome challenges. Furthermore, we intend to evaluate the effectiveness of bfGPS in supporting breastfeeding individuals within the community. The option to provide written comments on breastfeeding facilities is currently unavailable due to resource constraints, as close monitoring of the comments and additional staff is required. We will explore methods to enable a more comprehensive rating or comment system in the future.

The participants suggested the inclusion of breastfeeding information in the app. Studies have shown that incorporating breastfeeding information in the app can significantly improve breastfeeding outcomes [47,48]. While providing breastfeeding knowledge, it is essential to ensure the accuracy of the information. A study indicates that a high proportion of infant feeding apps provide incomplete or incorrect information [49]. In addition, most of these apps are provided by commercial sectors, which may result in biased information regarding infant feeding methods [49]. Beyond the app's content, gamification can be an important element to incorporate, as it is associated with hedonic well-being [50] and can encourage long-term app usage [51]. By maintaining user engagement, gamification is an effective component for promoting behavior change [52].

The development of bfGPS involved cross-disciplinary and multidisciplinary collaboration among public, nongovernmental, governmental, and commercial organizations. The fundamental goal is to have an impact on health in Hong Kong and beyond, influencing public, nongovernmental, governmental, and commercial organizations, as well as policies to develop an infrastructure that transforms breastfeeding culture. To date, only a limited number of studies have been conducted to understand the effects of the BFCI on breastfeeding outcomes [19]. Further research is needed to understand the effective elements in creating a breastfeeding-friendly community.

The engagement of community partners and breastfeeding individuals during the app development phase is vital. This approach allows us to understand the community's needs and ensures a user-centered design. Providing accurate information about breastfeeding facilities is important in encouraging breastfeeding individuals to use the app. Including commercial sectors in the BFCI program is important in gaining insights into the perspectives of the community members. This involvement enhances community engagement, encouraging community partners to offer breastfeeding-friendly services, and maintain up-to-date information on breastfeeding facilities. Ultimately, this collaborative effort contributes to fostering a breastfeeding-friendly attitude within the community.

Limitations

Our study has several limitations. First, the end users who participated in the usability evaluation were recruited from the BFCI project's social media page, which may have resulted in a selection bias, as participants were potentially more favorable toward breastfeeding, limiting the generalizability of the study findings. Second, conducting focus groups with public premises staff might have constrained them from freely expressing their individual opinions, as fellow participants were their colleagues or supervisors. Third, the objective of the usability evaluation was to understand the experiences of using the GPS feature. However, some participants did not use either the breastfeeding facilities locations or infant tracker functions, limiting their

XSL•FO
RenderX

ability to provide feedback on these specific features. In addition, some focus groups and one-on-one interviews were conducted before the COVID-19 pandemic, while others were conducted during the pandemic. Participants' concerns, needs, and expectations regarding breastfeeding facilities may have changed. This may limit the transferability of the findings.

Conclusion

This paper presents the user-centered process used in developing the bfGPS app, which aims to improve the breastfeeding experiences of families in the community. In addition, the app includes an infant tracker feature as an add-on, which can enhance various aspects of infant care. Our objective is for this work to serve as a model or guide for applying a user-centered methodology when developing new health care apps. Ensuring the app provides accurate and up-to-date information on breastfeeding facilities is crucial to promoting its continued and widespread use among breastfeeding individuals. This approach holds promise as an effective method for enhancing breastfeeding experiences in public and the community. The app has the potential to both reinforce and support the breastfeeding information provided to parents while also offering strategies for self-management to users. The next step involves testing the app as a therapeutic tool through a randomized clinical trial to verify its effectiveness in sustaining breastfeeding duration and addressing immediate breastfeeding issues. This ultimate goal serves as the motivation behind the creation and development of the app.

Acknowledgments

We would like to thank the participants in this study for sharing their experiences. During the preparation of this work, ChatGPT was used to enhance readability and improve the language. Following the use of this tool, HSLF and KYWL thoroughly reviewed and edited the content as necessary and take full responsibility for the accuracy and integrity of the final publication. This research is funded by the Health and Medical Research Fund (HKU-BF).

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Victora CG, Bahl R, Barros AJD, et al. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. Lancet 2016 Jan 30;387(10017):475-490. [doi: 10.1016/S0140-6736(15)01024-7] [Medline: 26869575]
2. Jones G, Steketee RW, Black RE, Bhutta ZA, Morris SS, Bellagio Child Survival Study Group. How many child deaths can we prevent this year? Lancet 2003 Jul 5;362(9377):65-71. [doi: 10.1016/S0140-6736(03)13811-1] [Medline: 12853204]
3. World Health Organization, UNICEF. Global strategy for infant and young child feeding. : World Health Organization; 2003.
4. World Health Organization. Global breastfeeding scorecard 2023: rates of breastfeeding increase around the world through improved protection and support. : World Health Organization; 2023.
5. Holowko N, Jones M, Koupil I, Tooth L, Mishra G. High education and increased parity are associated with breast-feeding initiation and duration among Australian women. Public Health Nutr 2016 Oct;19(14):2551-2561. [doi: 10.1017/S1368980016000367] [Medline: 26996672]
6. Tarrant M, Fong DYT, Wu KM, et al. Breastfeeding and weaning practices among Hong Kong mothers: a prospective study. BMC Pregnancy Childbirth 2010 May 29;10:27. [doi: 10.1186/1471-2393-10-27] [Medline: 20509959]
7. Brown A, Rance J, Bennett P. Understanding the relationship between breastfeeding and postnatal depression: the role of pain and physical difficulties. J Adv Nurs 2016 Feb;72(2):273-282. [doi: 10.1111/jan.12832] [Medline: 26494433]
8. Bai DL, Fong DYT, Lok KYW, Tarrant M. Relationship between the infant feeding preferences of Chinese mothers' immediate social network and early breastfeeding cessation. J Hum Lact 2016 May;32(2):301-308. [doi: 10.1177/0890334416630537] [Medline: 26887843]
9. Almohanna AA, Win KT, Meedya S. Effectiveness of internet-based electronic technology interventions on breastfeeding outcomes: systematic review. J Med Internet Res 2020 May 29;22(5):e17361. [doi: 10.2196/17361] [Medline: 32469315]
10. Lau Y, Htun TP, Tam WSW, Klainin-Yobas P. Efficacy of e-technologies in improving breastfeeding outcomes among perinatal women: a meta-analysis. Matern Child Nutr 2016 Jul;12(3):381-401. [doi: 10.1111/mcn.12202] [Medline: 26194599]
11. Ziebart M, Kammermeier M, Koletzko B, Patro-Golab B. Mobile applications for promoting and supporting breastfeeding: systematic review and meta-analysis. Matern Child Nutr 2024 Oct 11:e13733. [doi: 10.1111/mcn.13733] [Medline: 39390945]
12. Vila-Candel R, Mena-Tudela D, Franco-Antonio C, Quesada JA, Soriano-Vidal FJ. Effects of a mobile application on breastfeeding maintenance in the first 6 months after birth: randomised controlled trial (COMLACT study). Midwifery 2024 Jan;128:103874. [doi: 10.1016/j.midw.2023.103874] [Medline: 37979550]

XSL•FO
RenderX

13. Koçak V, Ege E, İyisoy MS. The development of the postpartum mobile support application and the effect of the application on mothers' anxiety and depression symptoms. Arch Psychiatr Nurs 2021 Oct;35(5):441-449. [doi: 10.1016/j.apnu.2021.06.009] [Medline: 34561057]
14. World Health Organization, UNICEF. Baby-friendly hospital initiative: revised, updated and expanded for integrated care. : World Health Organization; 2009.
15. Guajardo-Villar A, Pelat C, Blondel B, et al. The impact of the Baby-Friendly Hospital Initiative on breastfeeding rates at maternity units in France. Int J Epidemiol 2024 Apr 11;53(3):dyae080. [doi: 10.1093/ije/dyae080] [Medline: 38857529]
16. Pérez-Escamilla R, Martinez JL, Segura-Pérez S. Impact of the Baby-friendly Hospital Initiative on breastfeeding and child health outcomes: a systematic review. Matern Child Nutr 2016 Jul;12(3):402-417. [doi: 10.1111/mcn.12294] [Medline: 26924775]
17. Fair FJ, Morrison A, Soltani H. The impact of Baby Friendly Initiative accreditation: an overview of systematic reviews. Matern Child Nutr 2021 Oct;17(4):e13216. [doi: 10.1111/mcn.13216] [Medline: 34076344]
18. Kavle JA, Ahoya B, Kiige L, et al. Baby-Friendly Community Initiative-from national guidelines to implementation: a multisectoral platform for improving infant and young child feeding practices and integrated health services. Matern Child Nutr 2019 Jan;15 Suppl 1(Suppl 1):e12747. [doi: 10.1111/mcn.12747] [Medline: 30748118]
19. Walsh A, Pieterse P, Mishra N, et al. Improving breastfeeding support through the implementation of the Baby-Friendly Hospital and Community Initiatives: a scoping review. Int Breastfeed J 2023 Apr 15;18(1):22. [doi: 10.1186/s13006-023-00556-2] [Medline: 37061737]
20. Huang Y, Osborne D, Sibley CG. Public support and sociodemographic correlates of public breastfeeding support in New Zealand. N Z Med J 2018 Dec 14;131(1487):44-53. [Medline: 30543611]
21. Meng X, Daly A, Pollard CM, Binns CW. Community attitudes toward breastfeeding in public places among Western Australia Adults, 1995-2009. J Hum Lact 2013 May;29(2):183-189. [doi: 10.1177/0890334413478835] [Medline: 23504475]
22. Mulready-Ward C, Hackett M. Perception and attitudes: breastfeeding in public in New York City. J Hum Lact 2014 May;30(2):195-200. [doi: 10.1177/0890334414524988] [Medline: 24614263]
23. Russell K, Ali A. Public attitudes toward breastfeeding in public places in Ottawa, Canada. J Hum Lact 2017 May;33(2):401-408. [doi: 10.1177/0890334417695203] [Medline: 28418805]
24. Zhao Y, Ouyang YQ, Redding SR. Attitudes of Chinese adults to breastfeeding in public: a web-based survey. Breastfeed Med 2017 Jun;12:316-321. [doi: 10.1089/bfm.2017.0013] [Medline: 28440676]
25. US Department of Health Human Services. Executive summary: the surgeon general's call to action to support breastfeeding. Breastfeed Med 2011 Feb;6(1):3-5. [doi: 10.1089/bfm.2011.9996]
26. Hauck YL, Kuliukas L, Gallagher L, Brady V, Dykes C, Rubertsson C. Helpful and challenging aspects of breastfeeding in public for women living in Australia, Ireland and Sweden: a cross-sectional study. Int Breastfeed J 2020 May 12;15(1):38. [doi: 10.1186/s13006-020-00281-0] [Medline: 32398087]
27. Grant A. Breasts and the city: an urban ethnography of infant feeding in public spaces within Cardiff, United Kingdom. Int Breastfeed J 2021 Apr 29;16(1):1-9. [doi: 10.1186/s13006-021-00384-2] [Medline: 33926506]
28. Grant A, Pell B, Copeland L, et al. Views and experience of breastfeeding in public: a qualitative systematic review. Matern Child Nutr 2022 Oct;18(4):e13407. [doi: 10.1111/mcn.13407] [Medline: 35914544]
29. Scott JA, Kwok YY, Synnott K, et al. A comparison of maternal attitudes to breastfeeding in public and the association with breastfeeding duration in four European countries: results of a cohort study. Birth 2015 Mar;42(1):78-85. [doi: 10.1111/birt.12138] [Medline: 25339304]
30. Balaam M, Comber R, Jenkins E, Sutton S, Garbett A, editors. FeedFinder: a location-mapping mobile application for breastfeeding women. Presented at: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems; Apr 18-23, 2015; New York, NY. [doi: 10.1145/2702123.2702328]
31. Chaovalit P, Pongnumkul S, editors. Moommae: A location-aware mobile information system for breastfeeding mothers in thailand (invited paper). Presented at: 2017 10th International Conference on Ubi-media Computing and Workshops (Ubi-Media); Aug 1-4, 2017; Pattaya, Thailand. [doi: 10.1109/UMEDIA.2017.8074104]
32. Wang CJ, Chaovalit P, Pongnumkul S. A breastfeed-promoting mobile app intervention: usability and usefulness study. JMIR Mhealth Uhealth 2018 Jan 26;6(1):e27. [doi: 10.2196/mhealth.8337] [Medline: 29374000]
33. World breastfeeding week 2023: annual survey summary. : Baby Friendly Hospital Initiative Hong Kong Association; 2023 URL: https://www.babyfriendly.org.hk/wp-content/uploads/2023/07/2023-WBW-Annual-Survey_E.pdf [accessed 2025-01-06]
34. Breastfeeding survey 2021. Department of Health. 2021. URL: https://www.fhs.gov.hk/english/reports/files/BF_survey_2021.pdf [accessed 2025-01-06]
35. Survey on mothers' breastfeeding experience in public places. Department of Health. 2016. URL: https://www.fhs.gov.hk/english/archive/files/reports/Survey_on_Mothers_views_on_BF_exp_in_pubplace_2016_final.pdf [accessed 2025-01-06]
36. Howel D, Ball H. Association between length of exclusive breastfeeding and subsequent breastfeeding continuation. J Hum Lact 2013 Nov;29(4):579-585. [doi: 10.1177/0890334413492908] [Medline: 23821654]
37. Scott J, Ahwong E, Devenish G, Ha D, Do L. Determinants of continued breastfeeding at 12 and 24 months: results of an Australian cohort study. Int J Environ Res Public Health 2019 Oct 18;16(20):3980. [doi: 10.3390/ijerph16203980] [Medline: 31635280]

XSL•FO
RenderX

38. Lok KYW, Fan HSL, Ip HL, et al. Building a baby-friendly community: development, strategy, implementation and evaluation of a BFCI program. J Adv Nurs 2024 Jun 14. [doi: 10.1111/jan.16267] [Medline: 38877670]
39. Hong Kong 2020 growth references 2024. Department of Health HKSAR. URL: https://www.dh.gov.hk/english/useful/useful_HP_Growth_Chart/files/growth_charts.pdf [accessed 2025-01-06]
40. Schedule of Hong Kong Childhood Immunisation Programme. Family Health Service Department of Health HKSAR. 2022. URL: https://www.fhs.gov.hk/english/main_ser/child_health/child_health_recommend.html [accessed 2024-12-12]
41. Fusch PI, Ness LR. Are we there yet? Data saturation in qualitative research. TQR 2015;20(9):1408. [doi: 10.46743/2160-3715/2015.2281]
42. Patton MQ. Qualitative Research & Evaluation Methods, 3rd edition: Sage Publications; 2002. URL: https://aulasvirtuales.wordpress.com/wp-content/uploads/2014/02/qualitative-research-evaluation-methods-by-michael-patton.pdf [accessed 2025-01-06]
43. Bengtsson M. How to plan and perform a qualitative study using content analysis. NursingPlus Open 2016;2:8-14. [doi: 10.1016/j.npls.2016.01.001]
44. Webb C. Analysing qualitative data: computerized and other approaches. J Adv Nurs 1999 Feb;29(2):323-330. [doi: 10.1046/j.1365-2648.1999.00892.x] [Medline: 10197931]
45. Dinour LM. Infant feeding tracker applications: cross-sectional analysis of use. J Nutr Educ Behav 2022 Sep;54(9):835-843. [doi: 10.1016/j.jneb.2022.03.012] [Medline: 35643748]
46. Dienelt K, Moores CJ, Miller J, Mehta K. An investigation into the use of infant feeding tracker apps by breastfeeding mothers. Health Informatics J 2020 Sep;26(3):1672-1683. [doi: 10.1177/1460458219888402] [Medline: 31793799]
47. Doan TTD, Pham NM, Binns C, et al. Effect of a mobile application on breastfeeding rates among mothers who have cesarean deliveries: a randomized controlled trial. Breastfeed Med 2022 Oct;17(10):832-840. [doi: 10.1089/bfm.2022.0088] [Medline: 36083249]
48. Seyyedi N, Rahmatnezhad L, Mesgarzadeh M, Khalkhali H, Seyyedi N, Rahimi B. Effectiveness of a smartphone-based educational intervention to improve breastfeeding. Int Breastfeed J 2021 Sep 20;16(1):70. [doi: 10.1186/s13006-021-00417-w] [Medline: 34544429]
49. Cheng H, Tutt A, Llewellyn C, et al. Content and quality of infant feeding smartphone apps: five-year update on a systematic search and evaluation. JMIR Mhealth Uhealth 2020 May 27;8(5):e17300. [doi: 10.2196/17300] [Medline: 32459187]
50. Lin Y, Tudor-Sfetea C, Siddiqui S, Sherwani Y, Ahmed M, Eisingerich AB. Effective behavioral changes through a digital mHealth app: exploring the impact of hedonic well-being, psychological empowerment and inspiration. JMIR Mhealth Uhealth 2018 Jun 15;6(6):e10024. [doi: 10.2196/10024] [Medline: 29907557]
51. White B, Giglia RC, White JA, Dhaliwal S, Burns SK, Scott JA. Gamifying breastfeeding for fathers: process evaluation of the milk man mobile app. JMIR Pediatr Parent 2019 Jun 20;2(1):e12157. [doi: 10.2196/12157] [Medline: 31518324]
52. El-Hilly AA, Iqbal SS, Ahmed M, et al. Game on? Smoking cessation through the gamification of mHealth: a longitudinal qualitative study. JMIR Serious Games 2016 Oct 24;4(2):e18. [doi: 10.2196/games.5678] [Medline: 27777216]

Abbreviations

BFCI: Baby-Friendly Community Initiative
BFHI: Baby-Friendly Hospital Initiative
WHO: World Health Organization

Edited by A Mavragani; submitted 17.07.24; peer-reviewed by A Eisingerich, T Ma; revised version received 05.11.24; accepted 06.11.24; published 10.01.25.

Please cite as:
Fan HSL, Leung ETY, Lau KW, Wong JYH, Choi EPH, Lam C, Tarrant M, Ngan HYS, Ip P, Lin CC, Lok KYW
A Mobile App for Promoting Breastfeeding-Friendly Communities in Hong Kong: Design and Development Study
JMIR Form Res 2025;9:e64191
URL: https://formative.jmir.org/2025/1/e64191
doi:10.2196/64191

© Heidi Sze Lok Fan, Emily Tsz Yan Leung, Ka Wing Lau, Janet Yuen Ha Wong, Edmond Pui Hang Choi, Christine Lam, Marie Tarrant, Hextan Yuen Sheung Ngan, Patrick Ip, Chia Chin Lin, Kris Yuet Wan Lok. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete

XSL•FO
RenderX

bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Predicting Age and Visual-Motor Integration Using Origami Photographs: Deep Learning Study

Chien-Yu Huang[1,2], PhD; Yen-Ting Yu[1,3], PhD; Kuan-Lin Chen[3,4,5], PhD; Jenn-Jier Lien[6], PhD; Gong-Hong Lin[7], PhD; Ching-Lin Hsieh[1,2,8], PhD

[1]School of Occupational Therapy, College of Medicine, National Taiwan University, Taipei, Taiwan

[2]Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, Taipei, Taiwan

[3]Department of Occupational Therapy, College of Medicine, National Cheng Kung University, Tainan, Taiwan

[4]Institute of Allied Health Sciences, College of Medicine, National Cheng Kung University, Tainan, Taiwan

[5]Department of Physical Medicine and Rehabilitation, National Cheng Kung University Hospital, Tainan, Taiwan

[6]Department of Computer Science and Information Engineering, College of Electrical Engineering and Computer Science, National Cheng Kung University, Tainan, Taiwan

[7]International PhD Program in Gerontology and Long-Term Care, School of Nursing, Taipei Medical University, No. 250 Wu-Xing Street, Taipei, Taiwan

[8]Department of Occupational Therapy, College of Medical and Health Sciences, Asia University, Taipei, Taiwan

Corresponding Author:
Gong-Hong Lin, PhD
International PhD Program in Gerontology and Long-Term Care, School of Nursing, Taipei Medical University, No. 250 Wu-Xing Street, Taipei, Taiwan

Abstract

Background: Origami is a popular activity among preschool children and can be used by therapists as an evaluation tool to assess children's development in clinical settings. It is easy to implement, appealing to children, and time-efficient, requiring only simple materials—pieces of paper. Furthermore, the products of origami may reflect children's ages and their visual-motor integration (VMI) development. However, therapists typically evaluate children's origami creations based primarily on their personal background knowledge and clinical experience, leading to subjective and descriptive feedback. Consequently, the effectiveness of using origami products to determine children's age and VMI development lacks empirical support.

Objective: This study had two main aims. First, we sought to apply artificial intelligence (AI) techniques to origami products to predict children's ages and VMI development, including VMI level (standardized scores) and VMI developmental status (typical, borderline, or delayed). Second, we explored the performance of the AI models using all combinations of photographs taken from different angles.

Methods: A total of 515 children aged 2-6 years were recruited and divided into training and testing groups at a 4:1 ratio. Children created origami dogs, which were photographed from 8 different angles. The Beery–Buktenica Developmental Test of Visual-Motor Integration, 6th Edition, was used to assess the children's VMI levels and developmental status. Three AI models—ResNet-50, XGBoost, and a multilayer perceptron—were combined sequentially to predict age *z* scores and VMI *z* scores using the training group. The trained models were then tested using the testing group, and the accuracy of the predicted VMI developmental status was also calculated.

Results: The R^2 of the age and the VMI trained models ranged from 0.50 to 0.73 and from 0.50 to 0.66, respectively. The AI models that obtained an R^2>0.70 for the age model and an R^2>0.60 for the VMI model were selected for model testing. Those models were further examined for the accuracy of the VMI developmental status, the correlations, and the mean absolute error (MAE) of both the age and the VMI models. The accuracy of the VMI developmental status was about 71%-76%. The correlations between the final predicted age *z* score and the real age *z* score ranged from 0.84 to 0.85, and the correlations of the final predicted VMI *z* scores to the real *z* scores ranged from 0.77 to 0.81. The MAE of the age models ranged from 0.42 to 0.46 and those of the VMI models ranged from 0.43 to 0.48.

Conclusion: Our findings indicate that AI techniques have a significant potential for predicting children's development. The insights provided by AI may assist therapists in better interpreting children's performance in activities.

(JMIR Form Res 2025;9:e58421) doi:10.2196/58421

XSL•FO
RenderX

KEYWORDS

artificial intelligence; origami; child development screening; child development; visual motor integration; children; developmental status; activity performance; deep learning

Introduction

Origami is a common activity for preschool children [1]. From an occupational therapist's perspective, the quality of origami products may reflect a child's developmental age and visual-motor integration (VMI) [2]. For example, a 2-year-old child with typical development is able to roughly fold a sheet of paper in half, and a 3-year-old child is able to fold one diagonally and precisely. Therefore, origami activities and their products could help therapists identify children's development levels, including developmental age and VMI development, through the activity's process and the quality of the origami products.

Origami can be used by therapists as an evaluation tool to identify children's development in clinics because it is easy to implement, attractive to children, requires little time, and uses a simple material (only a piece of paper). However, therapists typically judge children's origami products based mainly on their own background knowledge and clinical experience, providing subjective and descriptive comments on the products (descriptive information). A few developmental assessments have used origami to assess children's fine motor skills. For example, the Peabody Developmental Motor Scales, Second Edition [3]; the Comprehensive Developmental Inventory for Infants and Toddlers [4]; the Stanford-Binet Intelligence Scale, Fifth Edition [5]; and the Bruininks-Oseretsky Test of Motor Proficiency, Second Edition [6] all contain one or more items involving paper folding to assess VMI, visual-spatial skills, or fine motor skills. However, these items are relatively simple and cannot solely be used to predict a child's developmental status. A study by Travers et al [7] investigated how children fold paper across early childhood, and they found that paper folding emerges as early as 27 months and becomes more accurate with age. Moreover, at least 50% of children between 4 and 5.54 years of age completed folds. However, their study only demonstrated children's ability to fold paper at various ages and provided no information on other developmental statuses. As a result, whether origami products can be used to predict children's development remains unsupported by evidence.

Artificial intelligence (AI) can be defined as a computer that acts or thinks like a human. Two AI techniques widely used in the medical and rehabilitation fields are computer vision and data prediction [8-10]. These two techniques may provide objective and quantitative data on origami products. Computer vision can transform image features into numerical data for data prediction, whereas the AI technique of data prediction uses data mining and probability to predict or estimate specific outcomes. Compared to traditional statistical analyses, AI data prediction has more flexibility to deal with categorical, ordinal, and numerical data. By combining the techniques of computer vision and data prediction, the important features of origami images can be extracted and further used to predict outcomes of interest for occupational therapists (eg, the children's age and scores of the standardized VMI assessment in our study).

Some studies have shown that AI techniques can analyze images and can identify children's behaviors and diagnoses with high accuracy [10-12]. In our preliminary study, we employed a simplified AI model (a single AI model) with a small sample size (n=119) to use photographs of origami products to predict children's ages [13]. The AI models showed promising performances (R^2=0.10 - 0.69). To extend upon our previous works, there were two purposes of our study. First, we refined the AI models, using a combination of 2 types of models, with a larger sample of origami products to predict children's ages (*z* scores), VMI levels (*z* scores), and VMI developmental status (typical, borderline, or delayed). Second, we explored the performances of the AI models using all combinations of photographs taken from different angles.

Methods

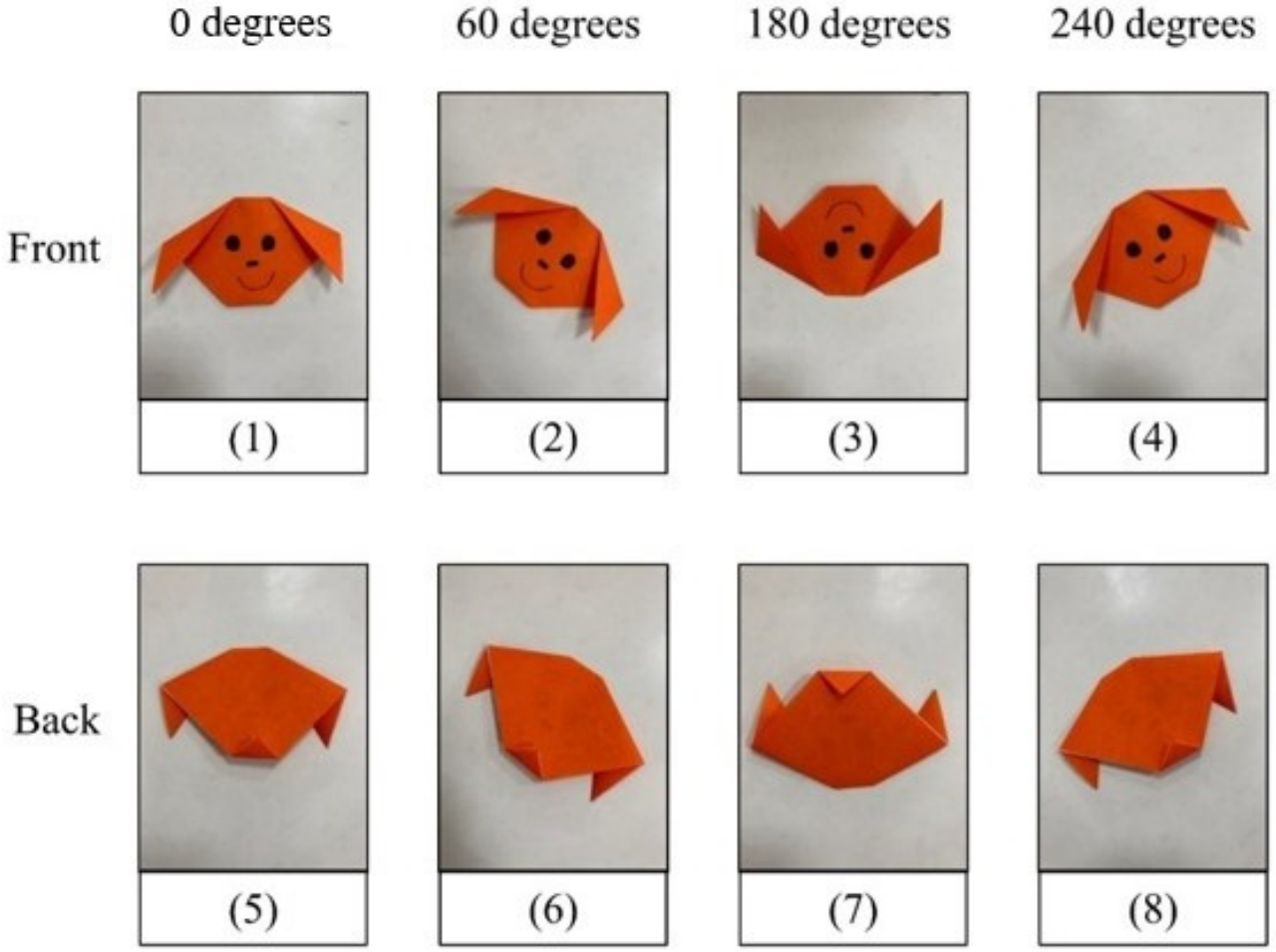
Participants and Procedure

The data included in our study were collected as part of an ongoing AI project (An Artificial Intelligence System for Assessing Gross Motor and Fine Motor, Handwriting, Attention and Emotion in Children with Developmental Delay), and some of the data were used in the preliminary study [13]. Preschool children enrolled in kindergarten were recruited to the study. The inclusion criteria were as follows: aged 2-6 years, ability to complete group administrations, and informed consent given by caregivers. The exclusion criterion was any obvious physical or mental disability.

The researchers first explained research purposes and procedures to the kindergarten principals. After the kindergarten principals agreed to participate in this study, researchers then contacted the children's parents to invite them to participate in the study. The researchers visited the children's classroom at the kindergarten and led them in two group activities (ie, origami and the administration of the VMI subtest). During the group activities, one researcher was a leader, and the other researchers (2-3 people) were coleaders who provided assistance to the children. Each coleader helped 3 - 6 children, depending on the number of children in the class. After leading the group activities, the researchers took photographs of the products. The researchers took 8 photographs of each origami work, including 4 angles (0, 60, 180, and 240 degrees) from the front side and the same 4 angles from the back side (Figure 1; the photographs are numbered 1 to 8 for each child depending on the shooting angle).

XSL•FO
RenderX

Figure 1. Origami photographs captured from different angles.



Measures

Origami

An origami dog was made by each of the children (see example in Figure 1). The dog was constructed in three steps: (1) fold a square paper into a triangle shape; (2) from the 2 corners of the long side of the triangle, diagonally fold 2 small triangles as 2 ears; and (3) turn the paper over, and from the other corner, fold a small triangle as the dog's chin. After the dog's face was made, the origami was turned to the front side. Facial features were then drawn on the dog by the child.

BEERY VMI Beery-Buktenica Visual-Motor Integration, Sixth Edition

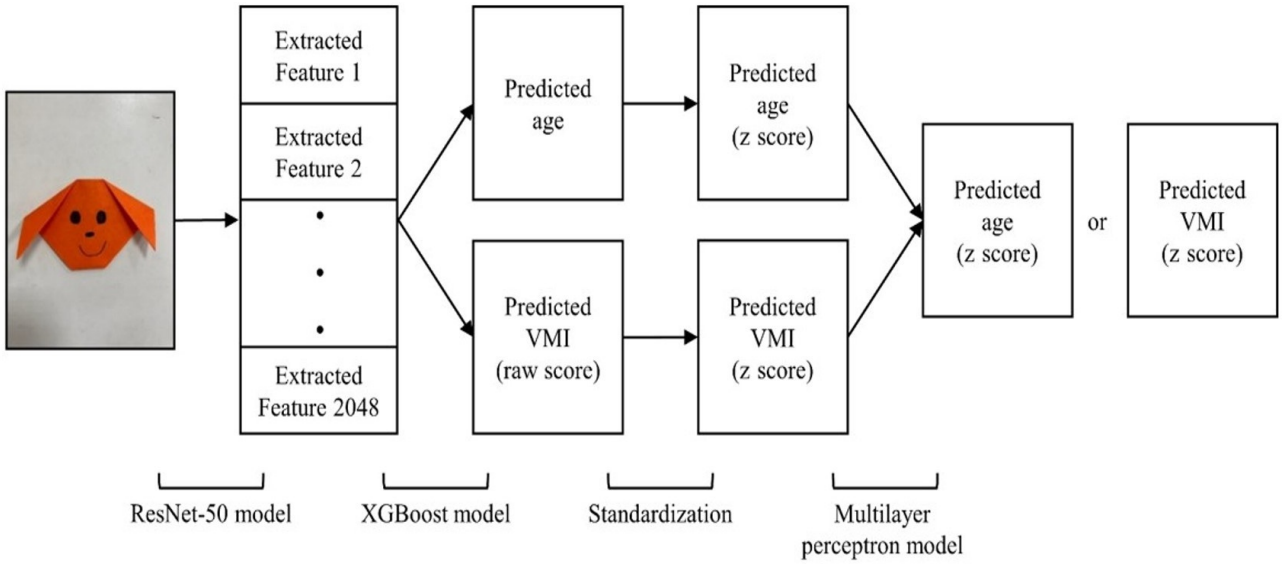
The BEERY VMI Beery-Buktenica Visual-Motor Integration, Sixth Edition (BEERY VMI-6) was used to assess the children's VMI development [14]. The BEERY VMI-6 consists of three subtests: VMI, visual perception, and motor coordination. Only the VMI subtest was used in this study. The VMI subtest contains 30 pictures. A child needed to copy each picture, and the copy was rated as 0 (incorrect) or 1 (correct). The total score of the VMI test ranged from 0 to 30. The total score of the VMI test was transformed into standard scores based on the VMI norm, and the standard scores were classified into 3 developmental statuses: typical, borderline, or delayed. The total VMI scores (raw VMI scores were referred to as the VMI level in this study) and the developmental status were used for the analyses. The BEERY VMI-6 has good psychometric properties [14].

Statistical Analysis

The participants' products were randomly divided into a training group and a testing group at a 4:1 ratio for the model training and testing. For the training group, 4 steps were conducted to build the AI models that would use origami photographs to respectively predict the children's age *z* scores and VMI *z* scores (Figure 2). First, we used a deep residual network of 50 layers, which is called ResNet-50 (developed by He et al [15]), to extract origami features of various photos at different angles [16]. Second, we trained XGBoost 1.0.0 models (developed by Chen and Guestrin [17]) to use a child's photo features to respectively predict the child's age and raw VMI score. ResNet-50 extracted 2048 features from the photos and the XGBoost model predicted age and raw VMI scores. All combinations of the extracted photo features (from 1 photo alone to 8 photographs together) were respectively fitted to the XGBoost models. The parameters of the XGBoost model were set as follows: n_estimators=100, random_state=1, max_depth=3, and subsample=0.8. The parameters of the multilayer perceptron model were set as follows: random_state=1, max_iter=300, tol=$1e^{-3}$, and n_iter_no_change=5.

XSL•FO
RenderX

Figure 2. An artificial intelligence model example of the deep learning procedure. VMI: visual-motor integration.



Third, the ages and the raw VMI scores predicted by the XGBoost models were respectively transformed into *z* scores (ie, a predicted age *z* score and a predicted VMI *z* score). We adopted the mean and SD of the training group's age and raw VMI scores for the transformation of *z* scores. We did so because the variables of age and raw VMI score were rated on different scales, and the standardization step ensured that the scales of the age and VMI score would be consistent (both between –1 and 1), which facilitated the subsequent model of data prediction [18,19]. Last, we trained multilayer perceptron models using both the predicted age *z* score and the predicted VMI *z* score to predict the real age *z* score and the real VMI *z* score, respectively.

After the two models (the XGBoost model and the multilayer perceptron model) were trained, the testing group was applied to the two trained models to examine the prediction performance of the models. The R^2, the correlation coefficient, and the mean absolute error (MAE) were used to identify the fitness of the models. The MAE represented the absolute score differences (ie, errors) between the real scores and the predicted scores. It was calculated as follows:

$$\sum i=1nyi-xi\vee n$$

y_i = predicted scores (eg, predicted age and predicted raw VMI scores);

x_i = real scores (eg, real age and real VMI scores).

Moreover, we intended to identify if the predicted VMI scores could reflect children's VMI developmental status. The predicted VMI *z* scores were further used for the analyses. First, the final predicted VMI *z* scores were transformed back to the predicted raw VMI scores. The predicted raw VMI scores were than transformed into predicted standard scores according to the norm of the VMI. Based on the predicted standard scores, we identified children's VMI developmental status as typical, borderline, or delayed. To examine the accuracy of the AI models in predicting the VMI developmental status, we compared the predicted developmental status to the child's real developmental status. Python 3.9 (Python Software) was used to perform the model training and the model testing.

Ethical Considerations

The ethics committee of I-Da hospital approved this study (REC number: EMRP46107N). This study was also approved by the Institutional Review Board of National Taiwan University Hospital (202111093RINB). All methods were conducted in accordance with relevant guidelines and regulations. Informed consent was obtained from the caregivers of the participating children. The study data were deidentified and securely stored in a private cloud space. Participants received a 200 New Taiwan dollar ($6.14 US) coupon after completing the evaluation. No identifying information of individual participants was present in any images included in the manuscript. Participation in this study is entirely voluntary. Participants had the right to withdraw at any time without providing a reason and without facing any negative consequences.

Results

Sample Descriptions

A total of 515 children were included in our study. The data for 412 children were used for the training group, and data for the other 102 children were used for the testing group. The sample characteristics and the VMI scores of the children are presented in Table 1. There were no significant differences in age and sex between the training and testing groups.

XSL•FO
RenderX

Table . Characteristics of the preschool children recruited from kindergartens in the cities of Kaohsiung and Tainan (n=515).

Variables	Training group (n=412)	Testing group (n=103)	*P* value
Age (years), mean (SD)	57.14 (14.46)	57.49 (12.03)	.79
Sex, n (%)			.63
Male	211 (51.2)	50 (48.5)	
Female	201(48.8)	53 (51.5)	
Raw VMI[a] scores, mean (SD)	11.22 (3.74)	11.34 (4.13)	.78
VMI age-equivalent scores, mean (SD)	58.02 (15.04)	58.70 (16.57)	.99
VMI standard scores, mean (SD)	99.06 (10.32)	99.04 (10.08)	.97
VMI percentiles, mean (SD)	47.51 (22.38)	47.61 (22.77)	.69
VMI developmental status, n (%)			.29
Typical	338 (82)	74 (71.8)	
Borderline	63 (15.3)	27 (26.2)	
Delayed	11 (2.7)	2 (2)	

[a]VMI: visual-motor integration.

AI Model Testing

All combinations of photo features were fitted to the XGBoost models, and then the predicted age *z* scores and the predicted VMI *z* scores were fitted to the multilayer perceptron models to produce the final predicted age *z* score and the VMI *z* score. We first inspected the R^2 values of the multilayer perceptron models for predicting the age *z* score and the VMI *z* score to select the better-performing models for further analyses. The R^2 of the age and the VMI models ranged from 0.50 to 0.73 and from 0.50 to 0.66, respectively. We selected AI models that had an R^2>0.70 for the age model and R^2>0.60 for the VMI model, which indicated better performance.

Those models were further examined for their accuracy in predicting the VMI developmental status, the correlations, and the MAE of both the age and the VMI models. Table 2 presents the final performances of the selected models. The accuracy of VMI developmental status was about 71%-76%. The correlations between the final predicted age *z* score and the real age *z* score ranged from 0.84 to 0.86, and the correlations of the final predicted VMI *z* scores to the real VMI *z* scores ranged from 0.7 to 0.81. The MAEs of the age models were between 0.42-0.46, and the MAEs of the VMI models were from 0.43 to 0.48.

Table . Comparison of origami prediction models using photos with different origami angles.

Photo numbers[a]	R^2 (age)	R^2 (VMI[b])	Accuracy (VMI)	Correlation (age)	Correlation (VMI)	MAE[c] (age)	MAE (VMI)
2, 5, 6	0.70	0.61	0.76	0.85	0.79	0.44	0.46
2, 5, 8	0.72	0.66	0.75	0.86	0.81	0.42	0.43
2, 7, 8	0.71	0.61	0.74	0.84	0.78	0.42	0.47
2, 3, 5, 6, 7	0.70	0.63	0.74	0.85	0.79	0.44	0.46
1, 3, 5, 7, 8	0.71	0.60	0.72	0.85	0.78	0.46	0.48
1, 3, 4, 5, 6, 7	0.73	0.61	0.71	0.86	0.78	0.44	0.47
2, 3, 4, 5, 6, 8	0.70	0.62	0.72	0.84	0.79	0.44	0.46
1, 2, 3, 5, 7, 8	0.70	0.62	0.73	0.84	0.79	0.45	0.46
2, 3, 4, 5, 7, 8	0.71	0.63	0.71	0.85	0.79	0.44	0.45
1, 2, 4, 5, 6, 7, 8	0.71	0.60	0.74	0.85	0.78	0.44	0.47

[a]Numbers correspond to the origami photographs shot from different sides and angles. Please see Figure 1 for the side and angle of each photo.

[b]VMI: visual-motor integration.

[c]MAE: mean absolute error.

Among all the combinations of photo features, the combination of photographs 2 (front 60 degrees), 5 (back 0 degrees), and 6

XSL•FO
RenderX

(back 60 degrees) and the combination of photographs 2, 5, and 8 (back 240 degrees) had better performances for predicting children's age *z* scores and VMI *z* scores.

Discussion

Principal Findings

This is the first study to apply AI techniques to origami images for predicting age, VMI level, and VMI developmental status in children. The results showed that AI models could be trained using photographs of children's origami creations to predict their ages (R^2 ranging from 0.70 to 0.73) and VMI levels (R^2 ranging from 0.60 to 0.66). Furthermore, the accuracy of predicting VMI developmental status was approximately 71%-76%. These findings suggest that AI techniques hold significant potential for analyzing origami images to further assess children's VMI development.

Analyzing origami photographs using AI techniques to predict children's age, as done in this study, presents two advantages. First, age is an interval indicator, and an interval indicator is an appropriate outcome measure for testing a prediction model. Second, children's age, to some extent, reflects their developmental level. As mentioned in the introduction, a 2-year-old child with typical development is supposed to be able to roughly fold a paper in half, and a 3-year-old child should be able to fold a paper diagonally and precisely. Therefore, predicting children's age through origami may help caregivers and therapists identify children's developmental status. For example, if an AI model prediction reports that a 5-year-old boy has the origami skills of a 3-year-old, the boy may have a developmental delay in some aspects (eg, fine motor manipulation, visual perception, or VMI). Thus, the boy can be referred for further evaluation and possible intervention. In summary, predicting children's ages could not only benefit researchers testing the AI model but also provide information about a child's development for therapists and caregivers.

Comparing all photo combinations for the final models showed that the combinations of photographs 2, 5, and 6 and of photographs 2, 5, and 8 performed better in predicting children's development. These results indicate that those photo combinations provide more useful information (or important features) about the origami products. This may be because the images of the origami dogs taken from those angles fully capture the dog's facial features, including the 2 ears, the chin, and the hand-drawn facial features, which could reflect the quality of the origami products. Using a particular and limited number of photographs can promote the efficiency and clinical utility of AI applications.

It is noted that the final predictions of the AI model used the age *z* score and the VMI *z* score as the input and output values. The reason was that the age and the raw VMI scores were two variables based on different scales, and normalization of the two variables could facilitate the model training process [18,19]. Moreover, the predicted age *z* score and the predicted VMI *z* score could be transformed back to the predicted age and predicted raw VMI scores based on the calculation formula. Therefore, future users could obtain the predicted age and the predicted VMI score from the AI model for use in the interpretation of origami performances.

Based on our findings, an evaluation tool for VMI development could be developed and used in both community and clinical settings. This tool would be an app designed to implement an AI model. By allowing users to upload photographs of a child's origami dogs, the app would immediately calculate and report the child's predicted age and raw VMI score. This app could provide 4 key benefits. First, the app would offer quantitative data, such as the predicted age, VMI level, and VMI developmental status, derived from a child's performance in folding origami dogs. Therapists could combine this quantitative information with personal observations to better interpret children's origami performance. Second, caregivers and teachers, who are not medical professionals, could use the app to better recognize children's development. The app's reports could help identify children with delayed VMI development or a predicted age significantly behind their chronological age, prompting caregivers and teachers to refer those children to medical facilities for further evaluation. Third, the app's internet connectivity would ensure easy accessibility, particularly benefiting children in remote regions or those who cannot undergo standardized evaluations. Fourth, as more people use the app, the continuous input of new data would help refine the AI model, thereby increasing the accuracy of the predicted age, VMI level, and VMI developmental status. This iterative process would enhance the app's effectiveness over time. Given these advantages, the app has significant potential for application in education and rehabilitation fields. Future research could explore incorporating other origami products into the model prediction to enhance its robustness. Examples include origami airplanes, frogs, and flowers. Allowing children to choose products based on their motivation and preferences could increase the model's utility and make its predictions of VMI development more robust.

Limitations

Several limitations were noted in this study. First, a small sample size of children aged 24 to 29 months (n=10) was used. This limited dataset may have resulted in insufficient data for AI learning, potentially increasing prediction errors for this age group. Second, most children recruited in the study exhibited typical development, and those with developmental conditions had mixed diagnoses. Consequently, the results may predominantly reflect the performance of typically developing children. Moreover, the model may tend to predict no developmental issues, increasing the likelihood of false negative results. Third, the images used for the model prediction were restricted to dog origami products, as only these were used to train the model. This limitation restricts the model's applicability and generalizability to other origami forms. Fourth, due to constraints in evaluation time and research personnel, only the standardized VMI score was used for model prediction. However, executing origami activities also requires manual dexterity, cognition, and visual perception, which were not included in the assessment. Fifth, since the predictions were primarily based on an AI model and conducted via the internet, there are concerns about children's privacy and data security. Furthermore, it is recommended that clinicians ensure

XSL•FO
RenderX

appropriate use and understanding when interpreting the results. Future studies should consider using standardized assessments of other developmental dimensions as model prediction outcomes to provide a more comprehensive understanding of children's development through origami activities. Additionally, increasing the sample size of young children and children with development conditions and incorporating different origami products in model training would enhance the model's utility and robustness.

Conclusion

Overall, it was previously known that origami activities and their products could help therapists identify children's development levels, including developmental age and VMI development, and photographs of origami could predict children's age and VMI. However, we have shown that AI techniques can provide quantitative information about activity products, and photographs of origami dogs can be used as an development evaluation tool for children. In conclusion, our study applied AI techniques to images of origami products to predict children's ages and VMI development. The results demonstrated that children's ages and VMI development could be predicted using photographs of origami dogs. Given that origami served as a successful example, AI techniques appear to have the significant potential for application to various products of children's activities, which could provide more quantitative insights.

Acknowledgments

We gratefully appreciate the children and caregivers who participated in our study. We also thank the principals and teachers from the Nian-Tze Preschool, the affiliated Da-She Elementary School Preschool, the affiliated Yan-Chao Elementary School Preschool, Mary Help of Christians Preschool, the affiliated Dong-Guang Elementary School Preschool, the affiliated Da-Tong Elementary School Preschool, and the New Generation Preschool. This work was supported by the Ministry of Science and Technology, Republic of China (110-2628-B-038-013, 109-2636-B-214-001, and 110-2636-B-214-001).

Data Availability

The datasets generated or analyzed during this study are available from the first author, CYH, upon reasonable request.

Authors' Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by GHL, CLH and CYH. The first draft of the manuscript was written by CYH and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. CYH is the guarantor of this paper.

Conflicts of Interest

None declared.

References

1. Arief YS, Binarwati D, Ismiwiranti R. Playing origami enhance the creativity of school aged children. J Ners 2008;3(1):42-48. [doi: 10.20473/jn.v3i1.4979]
2. Imaroonrak S, Phunwutikorn P, Phattharayuttawat S, Ngamthipwatthana T, Sumalrot T, Auampradit N. The effects of origami training on creativity and visual-motor integration in preschool children. J Med Assoc Thailand 2018;101(1):S85-S89. [doi: 10.6594/JTOTA.202112_39(2).0005]
3. Folio MR, Fewell RR. Peabody Developmental Motor Scales, 2nd edition: Pearson; 2000.
4. Hwang AW, Weng LJ, Liao HF. Construct validity of the comprehensive developmental inventory for infants and toddlers. Pediatr Int 2010 Aug;52(4):598-606. [doi: 10.1111/j.1442-200X.2010.03102.x] [Medline: 20202156]
5. Roid GH, Pomplun M. The Stanford-Binet Intelligence Scales, Fifth Edition. In: Contemporary Intellectual Assessment: Theories, Tests, and Issues: The Guilford Press; 2012.
6. Bruininks RH, Bruininks BD. BOT2: Bruininks-Oseretsky Test of Motor Proficiency, 2nd edition: Pearson; 2005.
7. Travers BG, Kirkorian HL, Jiang MJ, et al. Knowing how to fold 'em: paper folding across early childhood. J Mot Learn Dev 2018 Jun;6(1):147-166. [doi: 10.1123/jmld.2016-0044] [Medline: 33981774]
8. Li X, Shi Y. Computer vision imaging based on artificial intelligence. Presented at: 2018 International Conference on Virtual Reality and Intelligent Systems (ICVRIS); Aug 10-11, 2018; Changsha, Hunan, China p. 22-25. [doi: 10.1109/ICVRIS.2018.00014]
9. Hamet P, Tremblay J. Artificial intelligence in medicine. Metab Clin Exp 2017 Apr;69S:S36-S40. [doi: 10.1016/j.metabol.2017.01.011] [Medline: 28126242]
10. Agrawal A, Gans JS, Goldfarb A. Exploring the impact of artificial intelligence: prediction versus judgment. Inf Econ Policy 2019 Jun;47:1-6. [doi: 10.1016/j.infoecopol.2019.05.001]

XSL•FO
RenderX

11. Zhang Y, Ma Y. Application of supervised machine learning algorithms in the classification of sagittal gait patterns of cerebral palsy children with spastic diplegia. Comput Biol Med 2019 Mar;106:33-39. [doi: 10.1016/j.compbiomed.2019.01.009] [Medline: 30665140]
12. Ihlen EAF, Støen R, Boswell L, et al. Machine learning of infant spontaneous movements for the early prediction of cerebral palsy: a multi-site cohort study. J Clin Med 2019 Dec 18;9(1):5. [doi: 10.3390/jcm9010005] [Medline: 31861380]
13. Yu YT, Lin GH, Huang CY, Hsieh CL. Using deep learning to analyze pictures of origami to pedict the age of three-to-six-year-old children. TOTA 2021;39(2):211-231. [doi: 10.6594/jtota.202112_39(2).0005]
14. Beery KE, Buktenica NA, Beery NA. Beery VMI: The Beery-Buktenica Developmental Test of Visual-Motor Integration: Pearson; 2004.
15. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Jun 27-30, 2016; Las Vegas, NV p. 770-778. [doi: 10.1109/CVPR.2016.90]
16. Hershey S, Chaudhuri S, Ellis DPW, et al. CNN architectures for large-scale audio classification. Presented at: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); Mar 5-9, 2017; New Orleans, LA p. 131-135. [doi: 10.1109/ICASSP.2017.7952132]
17. Chen T, Guestrin C. Xgboost: a scalable tree boosting system. Presented at: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13-17, 2016; San Francisco, CA p. 785-794. [doi: 10.1145/2939672.2939785]
18. Salimans T, Kingma DP. Weight normalization: a simple reparameterization to accelerate training of deep neural networks. Presented at: The 30th Conference on Neural Information Processing Systems; Dec 5-10, 2016; Barcelona, Spain URL: https://dl.acm.org/doi/10.5555/3157096.3157197 [accessed 2024-12-11]
19. Sola J, Sevilla J. Importance of input data normalization for the application of neural networks to complex industrial problems. IEEE Trans Nucl Sci 1997;44(3):1464-1468. [doi: 10.1109/23.589532]

Abbreviations

AI: artificial intelligence
BEERY VMI-6: BEERY VMI Beery-Buktenica Visual-Motor Integration, Sixth Edition
MAE: mean absolute error
VMI: visual-motor integration

Edited by A Mavragani; submitted 15.03.24; peer-reviewed by D Kohen; revised version received 07.10.24; accepted 10.10.24; published 10.01.25.

Please cite as:
Huang CY, Yu YT, Chen KL, Lien JJ, Lin GH, Hsieh CL
Predicting Age and Visual-Motor Integration Using Origami Photographs: Deep Learning Study
JMIR Form Res 2025;9:e58421
URL: https://formative.jmir.org/2025/1/e58421
doi:10.2196/58421

© Chien-Yu Huang, Yen-Ting Yu, Kuan-Lin Chen, Jenn-Jier Lien, Gong-Hong Lin, Ching-Lin Hsieh. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Comparative Evaluation of Consumer Wearable Devices for Atrial Fibrillation Detection: Validation Study

Femke Wouters[1,2*], MSc; Henri Gruwez[1,2,3,4*], MD, PhD; Christophe Smeets[1,2], MSc, PhD; Anessa Pijalovic[2], MSc; Wouter Wilms[2], MSc; Julie Vranken[1,2], MSc; Zoë Pieters[5], MSc, PhD; Hugo Van Herendael[3], MD; Dieter Nuyens[3], MD, PhD, Prof Dr; Maximo Rivero-Ayerza[3], MD, PhD; Pieter Vandervoort[1,2,3], MD, PhD, Prof Dr; Peter Haemers[4], MD, PhD, Prof Dr; Laurent Pison[1,3], MD, PhD, Prof Dr

[1]Limburg Clinical Research Center/Mobile Health Unit, Faculty of Medicine and Life Sciences, Hasselt University, Hasselt, Belgium

[2]Department Future Health, Ziekenhuis Oost-Limburg, Genk, Belgium

[3]Department of Cardiology, Ziekenhuis Oost-Limburg, Genk, Belgium

[4]Department of Cardiovascular Sciences, KU Leuven, Leuven, Belgium

[5]Data Science Institute, Hasselt University, Hasselt, Belgium

[*]these authors contributed equally

Corresponding Author:
Laurent Pison, MD, PhD, Prof Dr
Limburg Clinical Research Center/Mobile Health Unit, Faculty of Medicine and Life Sciences, Hasselt University, Hasselt, Belgium

Abstract

Background: Consumer-oriented wearable devices (CWDs) such as smartphones and smartwatches have gained prominence for their ability to detect atrial fibrillation (AF) through proprietary algorithms using electrocardiography or photoplethysmography (PPG)–based digital recordings. Despite numerous individual validation studies, a direct comparison of interdevice performance is lacking.

Objective: This study aimed to evaluate and compare the ability of CWDs to distinguish between sinus rhythm and AF.

Methods: Patients exhibiting sinus rhythm or AF were enrolled through a cardiology outpatient clinic. The participants were instructed to perform heart rhythm measurements using a handheld 6-lead electrocardiogram (ECG) device (KardiaMobile 6L), a smartwatch-derived single-lead ECG (Apple Watch), and two PPG-based smartphone apps (FibriCheck and Preventicus) in a random sequence, with simultaneous 12-lead reference ECG as the gold standard.

Results: A total of 122 participants were included in the study: median age 69 (IQR 61-77) years, 63.9% (n=78) men, 25% (n=30) with AF, 9.8% (n=12) without prior smartphone experience, and 73% (n=89) without experience in using a smartwatch. The sensitivity to detect AF was 100% for all devices. The specificity to detect sinus rhythm was 96.4% (95% CI 89.5%-98.8%) for KardiaMobile 6L, 97.8% (95% CI 91.6% - 99.5%) for Apple Watch, 98.9% (95% CI 92.5% - 99.8%) for FibriCheck, and 97.8% (95% CI 91.5% - 99.4%) for Preventicus (*P*=.50). Insufficient quality measurements were observed in 10.7% (95% CI 6.3%-17.5%) of cases for both KardiaMobile 6L and Apple Watch, 7.4% (95% CI 3.9% - 13.6%) for FibriCheck, and 14.8% (95% CI 9.5% - 22.2%) for Preventicus (*P*=.21). Participants preferred Apple Watch over the other devices to monitor their heart rhythm.

Conclusions: In this study population, the discrimination between sinus rhythm and AF using CWDs based on ECG or PPG was highly accurate, with no significant variations in performance across the examined devices.

Trial Registration: ClinicalTrials.gov NCT06023290; https://clinicaltrials.gov/study/NCT06023290

(JMIR Form Res 2025;9:e65139) doi:10.2196/65139

KEYWORDS

atrial fibrillation; AF; mobile health; photoplethysmography; electrocardiography; smartphone; consumer wearable device; wearable devices; detection; electrocardiogram; ECG; mHealth

Introduction

Atrial fibrillation (AF) is the most prevalent cardiac arrhythmia among adults and is associated with an increased risk of stroke and mortality [1-4]. Since AF is often asymptomatic and paroxysmal, its diagnosis can be challenging. Asymptomatic AF frequently remains undetected until a thromboembolic event occurs. The diagnosis of AF requires an electrocardiogram

XSL•FO
RenderX

(ECG) showing irregular R-R intervals (when atrioventricular conduction is not impaired) and the absence of distinct repeating P waves [4].

Historically, ECG diagnostics were confined to medical settings, using devices administered or prescribed by health care professionals. However, the 21st century has witnessed a rapid surge in the availability and use of consumer-oriented wearable devices (CWDs) capable of monitoring heart rhythm [5-8]. Although this trend represents a significant democratization of access to heart rhythm monitoring, it also presents unique challenges for practicing clinicians to diagnose AF based on CWD data, due to the heterogeneous technologies and evolving algorithms used by these devices [9].

The underlying CWD technology is mainly based on electrocardiography or photoplethysmography (PPG). While an ECG measures electrical signals from the heart using surface electrodes on the skin, PPG analyzes the heart rhythm through an optical technique that measures the peripheral pulse [10]. This PPG technology has been adopted by smartwatches and smartphones, using light-emitting diodes and sensors or smartphone flashlights and cameras to generate PPG waveforms. The CWDs analyze ECG or PPG measurements using proprietary algorithms to notify the users of potential AF episodes [9,11,12]. However, only a few CWDs have received medical certification from the US Food and Drug Administration or the Conformité Européenne mark [13].

The emergence of CWDs equipped with automatic heart rate analysis presents a significant opportunity for the early detection of AF. These devices empower patients to monitor their heart rhythms independently, beyond traditional health care settings or without the need for physicians' prescriptions [9,14-16]. This shift enables more frequent and timely heart rhythm assessments, potentially leading to the earlier identification of AF episodes. However, this democratization of health monitoring also implies that physicians have limited involvement in selecting devices used by patients for heart rhythm analysis. Consequently, it becomes imperative for health care providers to adeptly interpret data from CWDs and assess the reliability of automated rhythm classifications of these devices. These skills are crucial to counsel patients who receive AF detections from CWDs effectively, especially as patients increasingly seek guidance on the accuracy of CWD measurements and on selecting the most suitable device.

Despite numerous validation studies of individual CWDs, a comprehensive comparison of their methodologies and outcomes is challenging, which leaves a gap in our understanding of their comparative effectiveness. This study aims to bridge this knowledge gap by evaluating and comparing the ability of digital CWDs to classify heart rhythms, specifically focusing on their accuracy in distinguishing between sinus rhythm and AF compared to the gold standard 12-lead ECG.

Methods

Study Design

Patients were recruited from July to November 2023. The VALIDATION (Validation of Digital Heart Rhythm Devices in the Detection of Atrial Fibrillation) study was a monocentric, prospective, randomized, interventional study that validated the use of PPG- and ECG-based CWDs to detect AF compared to a simultaneous 12-lead ECG in ambulatory cardiology patients in a supervised hospital environment.

Ethical Considerations

The study protocol is in accordance with the Declaration of Helsinki and the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use – Good Clinical Practice guidelines and was approved by 2 medical ethics committees (ie, Ziekenhuis Oost-Limburg, Genk, Belgium, Z-2023023; and Hasselt University, Hasselt, Belgium, B3712023000006). The study was registered at ClinicalTrials.gov (NCT06023290) and meets all STARD 2015 criteria. Informed consent was obtained in writing from all participants. Moreover, they were informed about their right to opt out of the study at any time without consequences. Data collected during the study was pseudonymized to protect participant confidentiality. Unique identifiers were used, and direct personal information was stored separately, accessible only to authorized personnel. Participants were not compensated for their involvement in the study as participation did not require an additional hospital visit and involved minimal additional time or effort.

Study Population

The study employed two distinct methods for participant inclusion, targeting adult patients from the cardiology outpatient clinic (Multimedia Appendix 1). First, all patients visiting the outpatient cardiology clinic were screened by chart review for eligibility on selected days when study personnel were available. Second, to enrich the study population with a higher prevalence of AF, all patients scheduled for electrical cardioversion at the cardiology outpatient clinic were also identified for screening by chart review.

Patients who met the following inclusion and exclusion criteria were invited to participate in the study: (1) age 18 years or older and (2) exhibit sinus rhythm or AF. Exclusion criteria were (1) presence of a pacemaker, (2) participation in another clinical trial that might interfere with the study protocol, and (3) physical or cognitive restrictions, including language barriers that hindered study measurements.

Study Measurements

After the participants provided informed consent, they were instructed by study personnel to perform heart rhythm measurements using 4 CWDs in a random sequence consisting of 2 ECG- and 2 PPG-based devices, according to the manufacturer's instructions. During all measurements, the patients were also instructed to stay still to avoid motion artifacts. All CWDs were provided by the study team to prevent device-induced bias.

KardiaMobile 6L (AliveCor Inc, Mountain View, CA) and a smartwatch, Apple Watch (Apple Inc, Cupertino, CA), were used to derive a single-lead ECG of 30 seconds. A lead-1 ECG was recorded using KardiaMobile 6L by positioning the device on the left knee with both thumbs on the device's front

XSL•FO
RenderX

electrodes [17]. To capture a lead-1 ECG using the Apple Watch, the participant wore the watch on the left wrist, ensuring skin contact with the electrode on the watch's underside. Measurement was initiated by the participant by touching a second electrode on the watch's crown with the right hand's index finger.

The PPG measurements were performed with two smartphone apps: (1) FibriCheck (Qompium NV, Hasselt, Belgium) and (2) Preventicus Heartbeats App (Preventicus GmbH, Jena, Germany). While initiating a PPG measurement in either smartphone app, the camera's flashlight was activated. Participants were instructed by study personnel to place an index finger over the smartphone camera. Using the camera, the app then captured skin color variations caused by heartbeat-induced blood flow changes. These variations were analyzed by the app to generate the PPG waveform.

The final rhythm diagnosis was made by the device's proprietary algorithm for both ECG and PPG measurements. Although 6 leads were measured using the KardiaMobile 6L device, its algorithm only took lead 1 into account. To minimize the occurrence of missing data, measurements were repeated up to 3 attempts until a good-quality measurement was obtained with each device. If good-quality measurements could not be obtained with a specific device, the patient's results were not excluded from the analyses. Good quality was defined as sufficient quality for analysis by the device algorithm. Rhythm diagnosis from CWDs were provided by the proprietary device algorithms and categorized as sinus rhythm or AF (Textbox 1). Concurrently with the CWD measurements, a continuous 12-lead ECG was performed for reference monitoring (MAC5000, GE Healthcare, Chicago, IL). Participants were excluded from the analyses if the heart rhythm changed on the reference ECG during the study conduct. The reference diagnosis on the 12-lead ECG was provided by two cardiologists blinded to the CWD classifications and the participant's clinical information.

XSL•FO
RenderX

Textbox 1. Categorization of consumer-oriented wearable device outcome labels.

Categorized as sinus rhythm

- KardiaMobile 6L
- Normal
- Bradycardia
- Tachycardia
- Apple Watch
- Sinus rhythm
- High heart rate - no atrial fibrillation
- Low or high heart rate
- FibriCheck
- Normal
- Mild irregularities
- Preventicus
- Normal
- Mild irregularities

Categorized as atrial fibrillation

- KardiaMobile 6L
- Possible atrial fibrillation
- Apple Watch
- Atrial fibrillation
- Atrial fibrillation - high heart rate
- FibriCheck
- Possible atrial fibrillation
- Preventicus
- Possible atrial fibrillation

Categorized as insufficient quality

- KardiaMobile 6L
- No analysis
- Unclassified
- Unreadable
- Apple Watch
- Inconclusive
- Poor reading
- FibriCheck
- Insufficient quality

XSL•FO
RenderX

- Preventicus
- Insufficient quality

Study Questionnaire

Following the completion of study measurements, participants were provided a 5-point Likert scale questionnaire designed to assess their experiences with and preferences regarding the devices used in the study. Additionally, the questionnaire aimed to evaluate the participants' digital literacy levels prior to their involvement in the study.

Data Analysis

Data analysis was performed using SPSS software (version 29.0; IBM Corp) and SAS (version 9.4; SAS Institute). $P<.05$ was considered statistically significant. The Shapiro-Wilk statistic was used to assess the normality of continuous data. Continuous variables are presented as median and IQR as appropriate and were compared using the Mann-Whitney *U* test. Discrete variables are presented as absolute numbers and percentages (%) and compared using Pearson χ^2 test. The performance of the CWDs were evaluated by calculating their sensitivity, specificity, and accuracy compared to the 12-lead reference ECG. These metrics were estimated and compared between CWDs using a generalized estimation equation model, accounting for the correlated nature of the measurements of the same patient. Measurements of insufficient quality were not included in these analyses.

Results

Overview

On designated study days, 200 patients (145 attending outpatient consultations and 55 undergoing elective cardioversions) were invited to participate in the study. Among these, 123 subjects were included, of which 122 completed the study protocol (see Multimedia Appendix 1). The remaining 77 patients either declined participation due to a lack of interest or were unwilling to commit the additional time required for study procedures. One participant with paroxysmal AF, who exhibited both sinus rhythm and AF on the reference ECG during the study, was excluded from the analysis to avoid confounding results. This decision was made because the specific rhythm present during the CWD recordings could not be reliably determined.

Study Population Characteristics

The median age of the study population was 69 (IQR 61-77) years, with 44 of the 122 (36.1%) participants being women. All participants were White. Detailed demographic and clinical characteristics are given in Table 1, and information on drug therapy is provided in Multimedia Appendix 2. There were no adverse events related to the use of the CWDs. Based on the 12-lead reference ECG, AF was present in 30 (24.6%) patients. The majority of the study population demonstrated a fair level of digital literacy, with 91 (74.6%) participants reporting daily internet usage and 96 (78.7%) using smartphones daily, respectively, as illustrated in Figure 1. However, 16 (13.1%) participants were identified as potentially digitally illiterate, having no prior experience with the internet or a smartphone. Furthermore, 89 (73%) participants reported no previous experience with CWDs, except smartphones.

XSL•FO
RenderX

Table . Demographic and clinical characteristics of the study population.

Participant characteristics	Participants			*P* value[a]
	Overall (n=122)	SR[b] (n=92)	AF[c] (n=30)	
Age (years), median (IQR)	69 (61-77)	68 (60.25 - 77.0)	72.5 (63.5 - 78.5)	.17
Gender (women), n (%)	44 (36.1)	34 (37)	10 (33.3)	.72
BMI (kg/m^2), median (IQR)	26.9 (24.1 - 30.8)	26.9 (23.5 - 30.3)	28.2 (25.7 - 33.6)	.02
Hypertension, n (%)	69 (56.6)	51 (55.4)	18 (60)	.66
Diabetes, n (%)	21 (17.2)	17 (18.5)	4 (13.3)	.52
Hypercholesterolemia, n (%)	83 (68)	65 (70.7)	18 (60)	.28
Heart failure, n (%)	21 (17.2)	12 (13)	9 (30)	.03
Vascular disease history, n (%)	26 (21.3)	22 (23.9)	4 (13.3)	.22
Stroke, n (%)	19 (15.6)	15 (16.3)	4 (13.3)	≥.99
CHA_2DS_2-VASc scores[d], n (%)				.50
0	12 (9.8)	9 (9.8)	3 (10)	
1	22 (18)	16 (17.4)	6 (20)	
2	21 (17.2)	18 (19.6)	3 (10)	
3	33 (27)	23 (25)	10 (33.3)	
4	21 (17.2)	18 (19.6)	3 (10)	
5	6 (4.9)	3 (3.3)	3 (10)	
6	7 (5.7)	5 (5.4)	2 (6.7)	

[a]Chi-square test, Fisher exact test, and Mann-Whitney *U* test were used in the analyses.

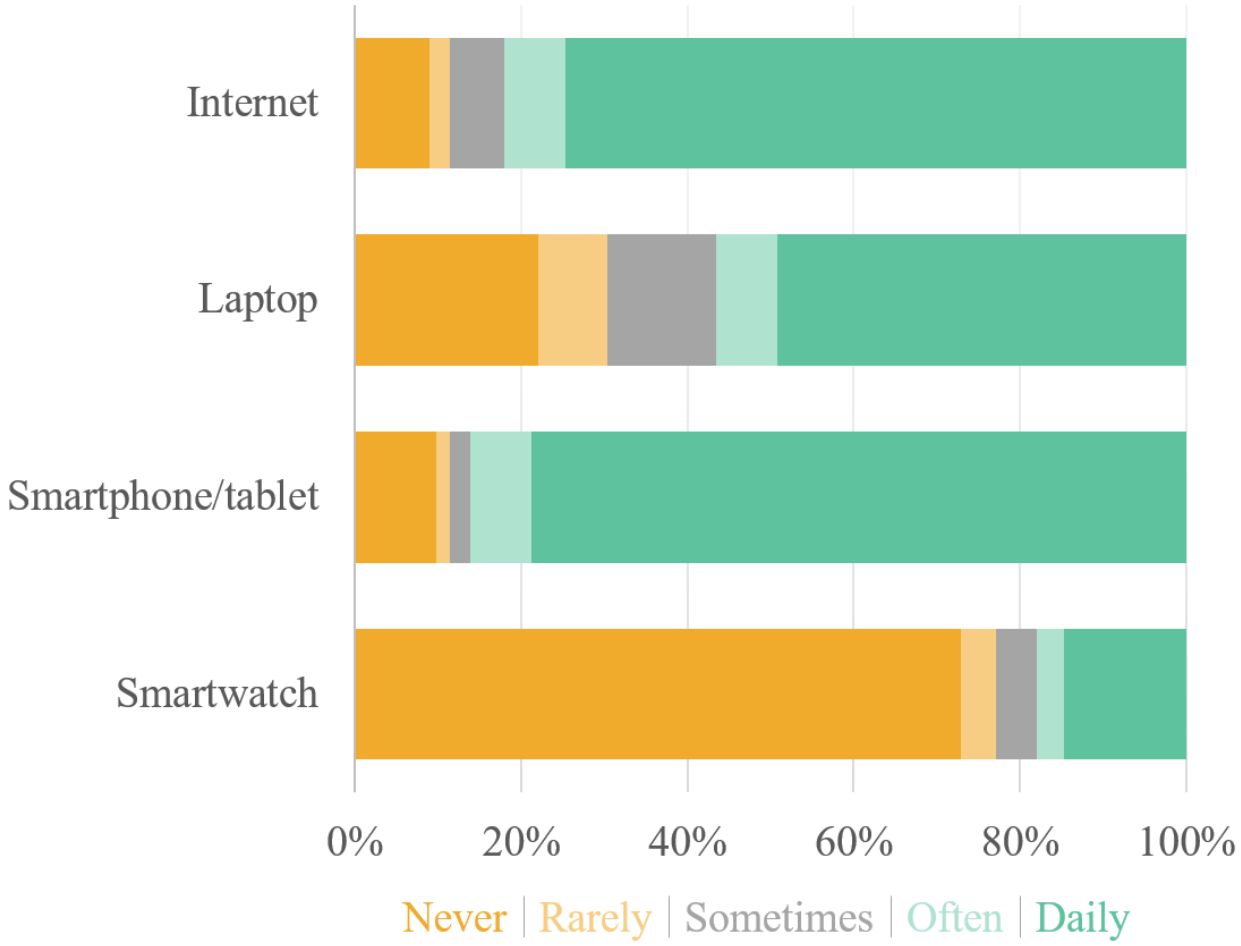
[b]SR: sinus rhythm.

[c]AF: atrial fibrillation.

[d]CHA_2DS_2-VASc is a points-based system used to stratify the risk of stroke in patients with AF. CHA_2DS_2-VASc is calculated as congestive heart failure, hypertension, age ≥75 years (doubled), diabetes, stroke (doubled), vascular disease, aged 65-74 years, and sex category (female) [18].

XSL•FO
RenderX

Figure 1. Experience with digital technology. Questions: I use the internet; I use a computer and/or laptop; I use a smartphone and/or tablet; I use a wearable device (eg, fitness tracker, smartwatch).

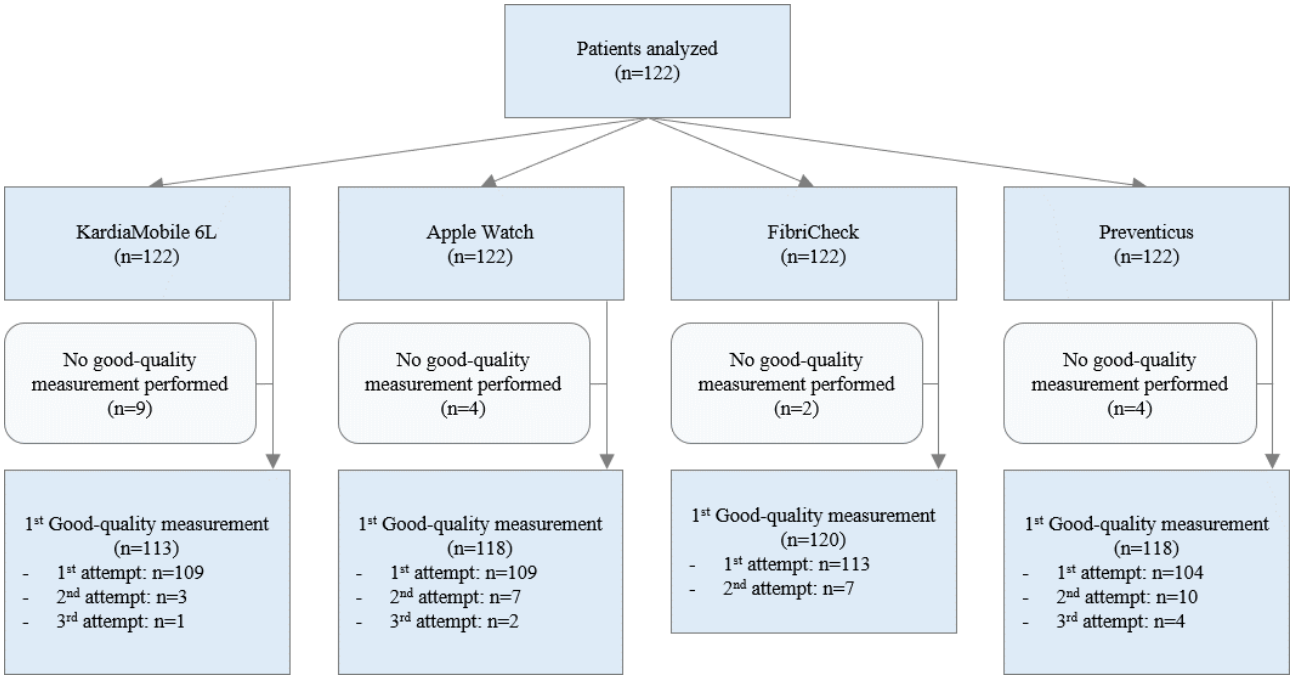


Signal Quality

Measurements of insufficient quality were repeated until the CWD algorithm provided a rhythm diagnosis. The number of additional measurements required for diagnosis is illustrated in Figure 2. Good-quality measurements could not be obtained after three attempts in 9 (7.4%) patients with KardiaMobile 6L, 4 (3.3%) with Apple Watch, 2 (1.6%) with FibriCheck, and 4 (3.3%) with Preventicus. Achieving good-quality measurements required 5 additional attempts for KardiaMobile 6L, 11 for Apple Watch, 7 for FibriCheck, and 18 for Preventicus. Consequently, the prevalence of insufficient quality measurements was 10.7% (95% CI 6.3%-17.5%) for both KardiaMobile 6L and Apple Watch, 7.4% (95% CI 3.9%-13.6%) for FibriCheck, and 14.8% (95% CI 9.5%-22.2%) for Preventicus, with no significant differences between the devices (P=.21). Furthermore, the distribution of insufficient quality measurements was comparable between patients in sinus rhythm and AF (11.7% vs 8.3%, P=.31).

XSL•FO
RenderX

Figure 2. Signal quality for each digital device. The flowchart describes the number of patients in which no good-quality measurement could be obtained after three attempts and the number of attempts necessary to obtain a good-quality measurement for each device.



Patients with insufficient quality recordings from both ECG and PPG devices were significantly older, with a median age of 83 (IQR 77-86) years compared to 68 (IQR 60-76) years (*P*<.001). However, there was no significant age difference between patients with insufficient quality recordings from either ECG or PPG devices and those with only good-quality recordings with a median age of 71 (IQR 62-80) years and 68 (IQR 60-75) years, respectively (*P*=.14). Similarly, no significant age difference was observed among patients with insufficient quality recordings (median age 74, IQR 62-80 years) versus those with only good-quality recordings (median age 68, IQR 60-75 years) from PPG devices specifically (*P*=.14). For ECG devices, measurements of insufficient quality were more often performed by older patients with a median age of 74 (IQR 64-82) years versus 68 (IQR 60-75) years (*P*=.02). Despite reduced digital literacy in older patients (*P*<.001), there was no significant association between digital literacy and the occurrence of insufficient data quality from either ECG- or PPG-based CWDs (*P*=.14).

Heart Rhythm Classification

Based on the first good-quality measurement, all patients with AF were correctly identified by all CWDs (Table 2). Therefore, the sensitivity for all CWDs to detect AF was 100%. The specificity rates for identifying sinus rhythm were 96.4% for KardiaMobile 6L, 97.8% for Apple Watch, 98.9% for FibriCheck, and 97.8% for Preventicus. The differences in specificity between devices were not statistically significant (*P*=.50). The diagnostic performance of all study devices is detailed in Table 3.

XSL•FO
RenderX

Table . Cross-tabulation of the consumer-oriented wearable device results compared to the 12-lead reference electrocardiogram (ECG).

	12-lead ECG rhythm (n=113)	
	SR[a]	AF[b]
KardiaMobile 6L, n/N[c] (%)		
SR	81/84 (96.4)	0/29 (0)
AF	3/84 (3.6)	29/29 (100)
Apple Watch, n/N (%)		
SR	89/91 (97.8)	0/27 (0)
AF	2/91 (2.2)	27/27 (100)
FibriCheck, n/N (%)		
SR	89/90 (98.9)	0/30 (0)
AF	1/90 (1.1)	30/30 (100)
Preventicus, n/N (%)		
SR	87/89 (97.8)	0/29 (0)
AF	2/89 (2.2)	29/29 (100)

[a]SR: sinus rhythm.

[b]AF: atrial fibrillation.

[c]N represents the number of detections by the 12-lead reference ECG for the SR and AF columns.

Table . Comparison of performance of the consumer-oriented wearable devices compared to the 12–lead reference electrocardiogram.[a]

Performance metrics	KardiaMobile 6L[b]	Apple Watch[b]	FibriCheck[c]	Preventicus[c]	*P* value
Good-quality tracing, % (95% CI)	89.3 (82.5-93.7)	89.3 (82.5-93.7)	92.6 (86.4-96.1)	85.3 (77.8-90.5)	.21
Sensitivity, % (95% CI)	100	100	100	100	—[d]
Specificity, % (95% CI)	96.4 (89.5-98.8)	97.8 (91.6-99.5)	98.9 (92.5-99.8)	97.8 (91.5-99.4)	.50
Accuracy, % (95% CI)	97.4 (92.1-99.1)	98.3 (93.5-99.6)	99.2 (94.3-99.9)	98.3 (93.5-99.6)	.48

[a]Generalized estimation equation model was used for the analysis.

[b]Single-lead electrocardiogram (lead 1).

[c]Smartphone photoplethysmography (finger over the camera).

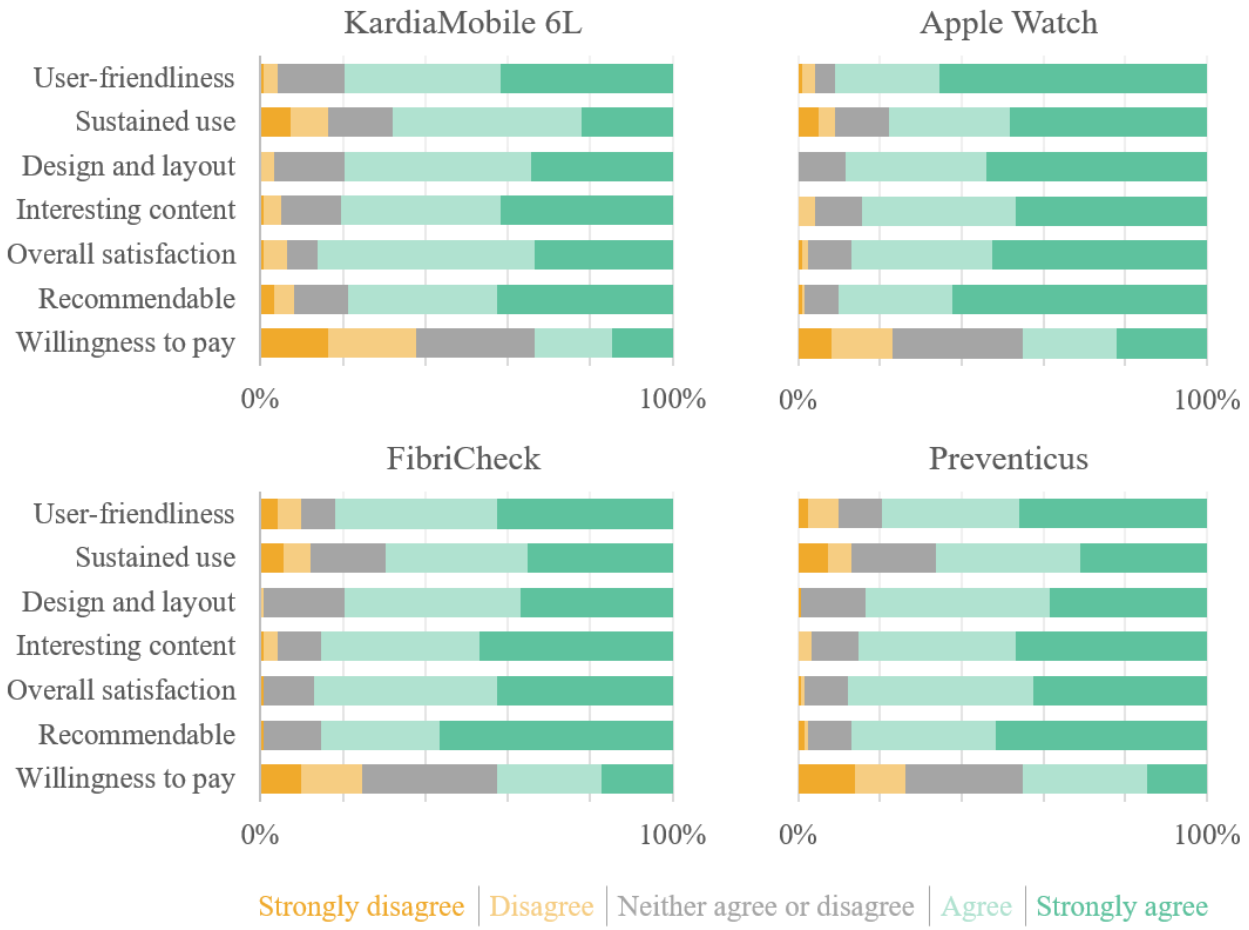
[d]Not applicable.

Study Questionnaire

After performing at least two measurements with each CWD, a significant majority of participants (over 80%) rated the study devices favorably across various criteria, including user-friendliness, likelihood of sustained use, design and layout, interesting content, overall satisfaction, and likelihood of recommending the devices to others (Figure 3). In these evaluations, the Apple Watch scored higher on the questionnaire, followed by FibriCheck, Preventicus, and KardiaMobile 6L. Despite these positive assessments, cost emerged as the principal barrier to broader adoption and implementation of these devices.

XSL•FO
RenderX

Figure 3. Study questionnaire responses. Questions: This device / app was easy to use; I would use this device / app again; I like the design and layout of this device / app; the content of this device / app is interesting to me; in general, I am satisfied with this device / app; I would recommend this device / app to people who would benefit from it; I would pay for this device / app (no price disclosed).



Discussion

Principal Findings

This study validated the ability of four CWDs to differentiate between sinus rhythm and AF using ECG or PPG technology and their proprietary algorithms. The devices were compared against a concurrently obtained 12-lead reference ECG, demonstrating very high sensitivity and specificity, thus confirming a near-perfect association with the diagnosis made using the 12-lead reference ECG.

To our knowledge, this is the first study to directly compare the accuracy of automated AF detection by ECG- and PPG-based CWDs against simultaneous 12-lead reference ECG monitoring. These findings align with prior studies that separately validated PPG and single-lead ECG CWDs [17,19-21]. However, the exclusion of patients and handling of uninterpretable results can significantly impact outcomes and hamper comparisons between studies [17]. A meta-analysis of 28 validation studies of PPG-based CWDs revealed significant biases, mainly concerning participant selection and insufficient information on study flow, design, and timing [19].

Addressing the limitations of previous studies, this trial adopted an inclusive approach by including all patients with a scheduled visit on designated study days, conducting simultaneous reference and index tests, and randomizing the sequence of index tests to minimize biases. However, a large number of the patients could not be screened or did not meet inclusion criteria, limiting findings to a specific demographic of adult White patients without a pacemaker in AF or sinus rhythm without frequent extrasystoles, who were capable of performing CWD measurements. Unlike previous studies, which have shown a greater incidence of uninterpretable measurements in up to one-third of cases, this study limited exclusions to 7.4%-14.8% by iterating measurements until a diagnosis was obtained. These measurements were excluded from this validation study, leading to an overestimation of the algorithm's accuracy [17,21,22]. This approach reflects real-world use, where users typically repeat measurements until a diagnosis is obtained from the device. Older patients, expectedly, were more likely to produce recordings of insufficient quality, frequently due to tremors.

Future validation studies are needed to validate these results in non-White populations and in individuals with rhythms other than sinus rhythm or AF, or with frequent extrasystoles that may challenge the algorithm's specificity. In our study, false-positive results from both ECG- and PPG-based CWDs were more likely when the 12-lead reference ECG showed premature atrial or ventricular contractions. These findings suggest that frequent extrasystoles can lead to rhythm misclassifications, highlighting the need for further research into how these irregularities affect device performance,

XSL•FO
RenderX

particularly in populations with diverse arrhythmias or more frequent extrasystoles. The study population included patients with both high and low digital literacy, which did not affect their performance in a supervised hospital environment. However, CWDs are mostly used in unsupervised settings. Recent real-world validation studies in unsupervised environments have exhibited similar results [23,24].

This study has implications for clinical practice as it provides physicians with information about the accuracy of CWDs. Physicians are likely to encounter CWDs, either by the patient's choice or owing to these devices being marketed and sold directly to the consumer. This comparative validation study of several devices with heterogeneous technologies and algorithms enables physicians to counsel patients on measurement results or the selection of medically certified devices. Similar studies should be repeated over time as algorithms are frequently updated, potentially improving diagnostic accuracy.

Given the high prevalence of uninterpretable device ECGs, mass utilization by the general population may lead to an increased and unnecessary health care burden [25]. To support clinical decisions based on CWD heart rhythm measurements, physicians should not rely solely on algorithm results, instead they should interpret traces manually when required. Although manual interpretation of CWDs was not within the scope of this study, it was investigated in the INTERPRET-AF study which reported similar accuracy for PPG- and ECG-based CWDs upon interpretation by physicians, with only modest reductions compared to the 12-lead ECG, even without prior PPG training [26].

The potential of CWDs in detecting new AF cases in high-risk populations (eg, secondary prevention in cryptogenic stroke patients) or monitoring AF recurrence in patients with an established diagnosis highlights the evolving landscape of cardiac care [23,27]. Despite increased AF detection and cost-effectiveness of these implementations, the stroke risk and therapeutic implications for AF detected by screening remain debatable [28]. Patients' willingness to adopt these CWDs was reported in over 80% of the questionnaire responses, signaling an increasing likelihood of implementation in the future.

Limitations

This study has several limitations. First, tracings deemed to be of insufficient quality were excluded from the analyses, potentially overestimating the performance and influencing the generalizability of the results. Second, patients with frequent extrasystoles or rhythms other than sinus rhythm or AF, as documented in their medical records, were excluded from this study, which could have reduced the specificity. Nevertheless, extrasystoles that occurred during the study procedures were included in the analyses, leading to false positive findings. The results of the CWD's algorithms were categorized into only two categories (ie, sinus rhythm and AF), without accounting for a broader range of differential diagnoses. Third, participants received instructions on device usage and were corrected when necessary to optimize signal quality. Fourth, this study could not evaluate the effect of skin color on the diagnostic accuracy of these devices. Fifth, the algorithm of the CWD with multiple ECG leads (ie, KardiaMobile 6L), which may have increased the specificity, was evaluated using 1 lead [17].

Conclusions

In this selected population, the distinction between sinus rhythm and AF using CWDs based on ECG or PPG was highly accurate, with no discernible variations observed across the examined devices.

Acknowledgments

FW is supported as a predoctoral student by the Bijzonder Onderzoeksfonds of Hasselt University (BOF20DOC17). HG is supported as a predoctoral strategic basic research fellow by the Fund for Scientific Research Flanders (FWO 1S83221N). This study is part of Limburg Clinical Research Center, supported by the foundation Limburg Sterk Merk, province of Limburg, Flemish government, Hasselt University, Ziekenhuis Oost-Limburg, and Jessa Hospital. The consumer-oriented wearable devices were provided by Qompium NV. The authors would like to thank Daimy Roebroek for her contribution to this work.

Authors' Contributions

Methodology: FW, HG, CS
Investigation: WW, AP, HG, FW
Statistical analysis: ZP, FW, HG
Writing – original draft: FW, HG
Writing – review and editing: FW, HG, CS, JV, HVH, DN, MRA, PV, PH, LP

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flowchart of study screening and inclusion. AF: atrial fibrillation; SR: sinus rhythm; SVES: supraventricular extrasystoles; VES: ventricular extrasystoles.
[PNG File, 116 KB - formative_v9i1e65139_app1.png]

XSL•FO
RenderX

Multimedia Appendix 2
Medication use of the study population. Chi-square test was used in the analyses. ACE: angiotensin-converting enzyme; AF: atrial fibrillation; ARB, angiotensin receptor blockers; CCB: calcium channel blockers; OAC: oral anticoagulation; SR: sinus rhythm.
[PNG File, 134 KB - formative_v9i1e65139_app2.png]

References

1. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. Stroke 1991 Aug;22(8):983-988. [doi: 10.1161/01.str.22.8.983] [Medline: 1866765]
2. McAllen PM, Marshall J. Cardiac dysrhythmia and transient cerebral ischaemic attacks. Lancet 1973 Jun 2;1(7814):1212-1214. [doi: 10.1016/s0140-6736(73)90527-8] [Medline: 4122563]
3. Chugh SS, Blackshear JL, Shen WK, Hammill SC, Gersh BJ. Epidemiology and natural history of atrial fibrillation: clinical implications. J Am Coll Cardiol 2001 Feb;37(2):371-378. [doi: 10.1016/s0735-1097(00)01107-4] [Medline: 11216949]
4. Hindricks G, Potpara T, Dagres N, et al. Corrigendum to: 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Eur Heart J 2021 Oct 21;42(40):4194-4194. [doi: 10.1093/eurheartj/ehab648]
5. Dagher L, Shi H, Zhao Y, Marrouche NF. Wearables in cardiology: here to stay. Heart Rhythm 2020 May;17(5 Pt B):889-895. [doi: 10.1016/j.hrthm.2020.02.023] [Medline: 32354455]
6. Seneviratne MG, Connolly SB, Martin SS, Parakh K. Grains of sand to clinical pearls: realizing the potential of wearable data. Am J Med 2023 Feb;136(2):136-142. [doi: 10.1016/j.amjmed.2022.10.006] [Medline: 36351523]
7. Hong W. Advances and opportunities of mobile health in the postpandemic era: smartphonization of wearable devices and wearable deviceization of smartphones. JMIR Mhealth Uhealth 2024 Jan 22;12:e48803. [doi: 10.2196/48803] [Medline: 38252596]
8. Nazarian S, Lam K, Darzi A, Ashrafian H. Diagnostic accuracy of smartwatches for the detection of cardiac arrhythmia: systematic review and meta-analysis. J Med Internet Res 2021 Aug 27;23(8):e28974. [doi: 10.2196/28974] [Medline: 34448706]
9. Svennberg E, Tjong F, Goette A, et al. How to use digital devices to detect and manage arrhythmias: an EHRA practical guide. Europace 2022 Jul 15;24(6):979-1005. [doi: 10.1093/europace/euac038] [Medline: 35368065]
10. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 2007 Mar;28(3):R1-39. [doi: 10.1088/0967-3334/28/3/R01] [Medline: 17322588]
11. Varma N, Cygankiewicz I, Turakhia MP, et al. 2021 ISHNE/HRS/EHRA/APHRS collaborative statement on mHealth in arrhythmia management: digital medical tools for heart rhythm professionals: from the International Society for Holter and Noninvasive Electrocardiology/Heart Rhythm Society/European Heart Rhythm Association/Asia Pacific Heart Rhythm Society. Cardiovasc Digit Health J 2021 Feb;2(1):4-54. [doi: 10.1016/j.cvdhj.2020.11.004] [Medline: 35265889]
12. Narayan SM, Wan EY, Andrade JG, et al. Visions for digital integrated cardiovascular care: HRS Digital Health Committee perspectives. Cardiovasc Digit Health J 2024 Apr;5(2):37-49. [doi: 10.1016/j.cvdhj.2024.02.003] [Medline: 38765620]
13. Baman JR, Mathew DT, Jiang M, Passman RS. Mobile health for arrhythmia diagnosis and management. J Gen Intern Med 2022 Jan;37(1):188-197. [doi: 10.1007/s11606-021-07007-w] [Medline: 34282532]
14. Lopez Perales CR, Van Spall HGC, Maeda S, et al. Mobile health applications for the detection of atrial fibrillation: a systematic review. EP Europace 2021 Jan 27;23(1):11-28. [doi: 10.1093/europace/euaa139]
15. Bonini N, Vitolo M, Imberti JF, et al. Mobile health technology in atrial fibrillation. Expert Rev Med Devices 2022 Apr;19(4):327-340. [doi: 10.1080/17434440.2022.2070005] [Medline: 35451347]
16. Klier K, Koch L, Graf L, Schinköthe T, Schmidt A. Diagnostic accuracy of single-lead electrocardiograms using the Kardia Mobile App and the Apple Watch 4: validation study. JMIR Cardio 2023 Nov 23;7:e50701. [doi: 10.2196/50701] [Medline: 37995111]
17. Scholten J, Jansen WPJ, Horsthuis T, et al. Six-lead device superior to single-lead smartwatch ECG in atrial fibrillation detection. Am Heart J 2022 Nov;253(53-8):53-58. [doi: 10.1016/j.ahj.2022.06.010] [Medline: 35850242]
18. Lip GYH, Nieuwlaat R, Pisters R, Lane DA, Crijns H. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the Euro Heart Survey on Atrial Fibrillation. Chest 2010 Feb;137(2):263-272. [doi: 10.1378/chest.09-1584] [Medline: 19762550]
19. Gill S, Bunting KV, Sartini C, et al. Smartphone detection of atrial fibrillation using photoplethysmography: a systematic review and meta-analysis. Heart 2022 Sep 26;108(20):1600-1607. [doi: 10.1136/heartjnl-2021-320417] [Medline: 35277454]
20. O'Sullivan JW, Grigg S, Crawford W, et al. Accuracy of smartphone camera applications for detecting atrial fibrillation: a systematic review and meta-analysis. JAMA Netw Open 2020 Apr 1;3(4):e202064. [doi: 10.1001/jamanetworkopen.2020.2064] [Medline: 32242908]

XSL•FO
RenderX

21. Mannhart D, Lischer M, Knecht S, et al. Clinical validation of 5 direct-to-consumer wearable smart devices to detect atrial fibrillation: BASEL wearable study. JACC Clin Electrophysiol 2023 Feb;9(2):232-242. [doi: 10.1016/j.jacep.2022.09.011] [Medline: 36858690]
22. Seshadri DR, Bittel B, Browsky D, et al. Accuracy of Apple Watch for detection of atrial fibrillation. Circulation 2020 Feb 25;141(8):702-703. [doi: 10.1161/CIRCULATIONAHA.119.044126] [Medline: 32091929]
23. Gruwez H, Ezzat D, Van Puyvelde T, et al. Real-world validation of smartphone-based photoplethysmography for rate and rhythm monitoring in atrial fibrillation. Europace 2024 Mar 30;26(4):38630867. [doi: 10.1093/europace/euae065]
24. Fernstad J, Svennberg E, Åberg P, Kemp Gudmundsdottir K, Jansson A, Engdahl J. Validation of a novel smartphone-based photoplethysmographic method for ambulatory heart rhythm diagnostics: the SMARTBEATS study. Europace 2024 Mar 30;26(4):euae079. [doi: 10.1093/europace/euae079] [Medline: 38533836]
25. Perez MV, Mahaffey KW, Hedlin H, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917. [doi: 10.1056/NEJMoa1901183] [Medline: 31722151]
26. Gruwez H, Evens S, Proesmans T, et al. Accuracy of physicians interpreting photoplethysmography and electrocardiography tracings to detect atrial fibrillation: INTERPRET-AF. Front Cardiovasc Med 2021;8(734737):734737. [doi: 10.3389/fcvm.2021.734737] [Medline: 34616786]
27. Wouters F, Gruwez H, Vranken J, et al. The potential and limitations of mobile health and insertable cardiac monitors in the detection of atrial fibrillation in cryptogenic stroke patients: preliminary results from the REMOTE trial. Front Cardiovasc Med 2022;9:848914. [doi: 10.3389/fcvm.2022.848914] [Medline: 35498000]
28. Lowres N, Neubeck L, Salkeld G, et al. Feasibility and cost-effectiveness of stroke prevention through community screening for atrial fibrillation using iPhone ECG in pharmacies. The SEARCH-AF study. Thromb Haemost 2014 Jun;111(6):1167-1176. [doi: 10.1160/TH14-03-0231] [Medline: 24687081]

Abbreviations

AF: atrial fibrillation
CWD: consumer-oriented wearable device
ECG: electrocardiogram
PPG: photoplethysmography
VALIDATION: Validation of Digital Heart Rhythm Devices in the Detection of Atrial Fibrillation

Edited by A Mavragani; submitted 06.08.24; peer-reviewed by D Sokas; revised version received 05.11.24; accepted 19.11.24; published 09.01.25.

Please cite as:
Wouters F, Gruwez H, Smeets C, Pijalovic A, Wilms W, Vranken J, Pieters Z, Van Herendael H, Nuyens D, Rivero-Ayerza M, Vandervoort P, Haemers P, Pison L
Comparative Evaluation of Consumer Wearable Devices for Atrial Fibrillation Detection: Validation Study
JMIR Form Res 2025;9:e65139
URL: https://formative.jmir.org/2025/1/e65139
doi:10.2196/65139

©Femke Wouters, Henri Gruwez, Christophe Smeets, Anessa Pijalovic, Wouter Wilms, Julie Vranken, Zoë Pieters, Hugo Van Herendael, Dieter Nuyens, Maximo Rivero-Ayerza, Pieter Vandervoort, Peter Haemers, Laurent Pison. Originally published in JMIR Formative Research (https://formative.jmir.org), 9.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Real-Time Analytics and AI for Managing No-Show Appointments in Primary Health Care in the United Arab Emirates: Before-and-After Study

Yousif Mohamed AlSerkal[1*], BMed, MSc, MBBS, BAO; Naseem Mohamed Ibrahim[1*], MBBCH, Arab Board Speciality; Aisha Suhail Alsereidi[1*]; Mubaraka Ibrahim[1*]; Sudheer Kurakula[2*]; Sadaf Ahsan Naqvi[2*], MBBS, SQIL; Yasir Khan[2*]; Neema Preman Oottumadathil[2*]

[1]Emirates Health Services, Dubai, United Arab Emirates

[2]Oracle, The Edge Building, Al Falak Street, Dubai Internet City, Dubai, United Arab Emirates

[*]all authors contributed equally

Corresponding Author:
Sadaf Ahsan Naqvi, MBBS, SQIL
Oracle, The Edge Building, Al Falak Street, Dubai Internet City, Dubai, United Arab Emirates

Abstract

Background: Primary health care (PHC) services face operational challenges due to high patient volumes, leading to complex management needs. Patients access services through booked appointments and walk-in visits, with walk-in visits often facing longer waiting times. No-show appointments are significant contributors to inefficiency in PHC operations, which can lead to an estimated 3%-14% revenue loss, disrupt resource allocation, and negatively impact health care quality. Emirates Health Services (EHS) PHC centers handle over 140,000 visits monthly. Baseline data indicate a 21% no-show rate and an average patient wait time exceeding 16 minutes, necessitating an advanced scheduling and resource management system to enhance patient experiences and operational efficiency.

Objective: The objective of this study was to evaluate the impact of an artificial intelligence (AI)-driven solution that was integrated with an interactive real-time data dashboard on reducing no-show appointments and improving patient waiting times at the EHS PHCs.

Methods: This study introduced an innovative AI-based data application to enhance PHC efficiency. Leveraging our electronic health record system, we deployed an AI model with an 86% accuracy rate to predict no-shows by analyzing historical data and categorizing appointments based on no-show risk. The model was integrated with a real-time dashboard to monitor patient journeys and wait times. Clinic coordinators used the dashboard to proactively manage high-risk appointments and optimize resource allocation. The intervention was assessed through a before-and-after comparison of PHC appointment dynamics and wait times, analyzing data from 135,393 appointments (67,429 before implementation and 67,964 after implementation).

Results: Implementation of the AI-powered no-show prediction model resulted in a significant 50.7% reduction in no-show rates (P<.001). The odds ratio for no-shows after implementation was 0.43 (95% CI 0.42-0.45; P<.001), indicating a 57% reduction in the likelihood of no-shows. Additionally, patient wait times decreased by an average of 5.7 minutes overall (P<.001), with some PHCs achieving up to a 50% reduction in wait times.

Conclusions: This project demonstrates that integrating AI with a data analytics platform and an electronic health record systems can significantly improve operational efficiency and patient satisfaction in PHC settings. The AI model enabled daily assessments of wait times and allowed for real-time adjustments, such as reallocating patients to different clinicians, thus reducing wait times and optimizing resource use. These findings illustrate the transformative potential of AI and real-time data analytics in health care delivery.

(JMIR Form Res 2025;9:e64936) doi:10.2196/64936

KEYWORDS

electronic health record; EHR; artificial intelligence; AI; no-show appointments; real-time data; primary health care; risk prediction; clinic waiting time; operational efficiency

XSL•FO
RenderX

Introduction

Primary health care (PHC) centers are the initial contact point for health care and are considered the gateway to the health care system for a large proportion of the population [1,2]. PHC centers experience a high daily patient influx, making PHC operations complex and requiring efficient systems. While many patients book appointments in advance, others rely on walk-ins, often leading to longer wait times for older and pediatric patients who require quicker attention. However, a significant issue is no-show appointments, where patients fail to attend their scheduled appointments without prior cancellation.

No-show appointments create a significant burden by unnecessarily blocking health care resources. They can negatively impact the financial and operational planning of health care organizations, leading to an estimated revenue loss of 3% - 14% [3,4]. Various interventions have been proposed to address this issue, focusing on understanding patient behaviors and reducing no-show rates [5]. Many organizations use data analytics to predict no-show probabilities and devise interventions to improve clinical efficiencies [6-8]. Literature suggests that while machine learning models have been explored to predict no-show likelihoods [9], a gap exists in operationalizing these insights at a large scale, especially in settings like the Emirates Health Services (EHS), which manages a vast network of health care facilities.

Another key indicator of patient satisfaction is waiting time, which directly affects the quality of care and patient experience [10-12]. Numerous studies have demonstrated that reducing patient waiting times can significantly enhance patient satisfaction and overall service quality. These studies highlight that multiple interventions, such as resource optimization and system redesign, are necessary to improve operational efficiency. Identifying bottlenecks and implementing practical interventions can significantly reduce waiting times [13-17]. With the evolution of artificial intelligence (AI), organizations can now leverage real-time data analytics to develop predictive models and suggest proactive interventions.

While there are examples from various health care settings demonstrating the successful use of AI models and patient reminders to reduce no-shows [18,19], there is limited research on how such systems can be optimized specifically for large health care networks like the EHS. Several studies highlight the importance of real-time analytics in improving health care outcomes, allowing organizations to respond promptly to emerging challenges and optimize resource allocation [20-22]. Hence, this study aims to address this gap by providing an electronic health record (EHR)-driven, real-time analytics solution tailored to the operational needs of PHC centers.

The EHS is a vast health care network across 6 emirates in the United Arab Emirates, managing hospitals, clinics, specialized centers, and supporting services [23]. The PHC centers within the EHS handle over 140,000 visits each month, resulting in a high patient turnover. While this often leads to longer-than-expected wait times negatively impacting patient experience, some appointment slots remain underused due to an average no-shows rate of 21%. Baseline data show the average physician waiting time exceeds 16 minutes, and while resources are distributed based on center needs, occasional shortages contribute to operational delays.

Considering these challenges, our study aims to evaluate the impact of a real-time dashboard that projects operational metrics and integrates an AI model to predict no-show appointments in PHC centers at the EHS. By using this AI-enriched dashboard, clinic administrators can identify high-risk patients (risk of no-show) and proactively contact them to emphasize the importance of their scheduled visits. The primary goals of this evaluation were to assess if the dashboard improved operational efficiency, reduced no-show appointments, optimized resource allocation, and decreased patient waiting times.

Methods

Study Design

This project was conceived as an innovative data-driven intervention using our data intelligence platform seamlessly integrated with the EHR (Multimedia Appendix 1). Using real-time PHC data, we created a dashboard with an AI model to guide the operational workflow and to reduce clinic wait times and no-show appointments.

This was a before-and-after study that evaluated the impact of a real-time dashboard with an AI-powered no-show prediction model at EHS PHCs 3 months before and after implementation. By leveraging this dashboard, clinic coordinators could proactively manage PHC traffic flow, enhancing operational efficiency and reducing patient waiting times. To capture key metrics accurately, several data definitions were established to track patient flow within the PHC setting. Clinic waiting time was defined as the intervals between various points in a patient's journey, which were all recorded as time stamps in the EHR system. Specifically, this included:

- Nurse waiting time: time from registration to the nurse station.
- Physician waiting time: time from being ready to be seen by the physician to the actual consultation.
- Check-in to checkout time: total time from check-in to final checkout by the physician.
- Appointment no-show: defined as a patient missing a scheduled appointment without prior notice to the PHC.

For comprehensive analysis, all visit and attendance data were sourced from the EHR and the human resources attendance system, both of which were updated in real time to ensure data accuracy and relevance.

Research Road Map Intervention and Dashboard Implementation

Our research followed a structured path from data acquisition to intervention and evaluation. We first integrated patient and clinic staff data from the EHR into a real-time dashboard that incorporated an AI model trained on historical EHR data to predict no-show appointments. This model, along with other key PHC metrics—such as footfall, resource availability, and wait times—were displayed on the dashboard to support clinic coordinators in managing appointments proactively. The AI

XSL•FO
RenderX

model projected daily no-show risk, allowing coordinators to reach out to high-risk patients for confirmation or rescheduling. Additionally, real-time resource constraints were visible on the dashboard, enabling operational adjustments as needed. All users received training on the new dashboard and workflow, and continuous monitoring and evaluation processes were established to assess ongoing performance. Finally, we analyzed the results to measure the dashboard's impact on clinic operations (Multimedia Appendices 2 and 3).

AI Model

Our AI model was developed to predict no-show appointments using historical EHR data from EHS PHCs, incorporating 16 distinct features: 4 demographic, 3 on patient history, and 9 appointment-specific factors (Multimedia Appendix 4). We used a random forest classification technique to assess the model's effectiveness in handling high-dimensional data, achieving an accuracy rate of 86%. The model categorized appointments into a high, medium, or low risk of a no-show, based on probability thresholds of ≥90%, 89% - 80%, and 79% - 70%, respectively. To validate the AI model, we used bootstrapping to ensure robustness and to reduce overfitting.

Statistical Analysis and Experimental Validation

Data were analyzed using SPSS version 17 (IBM Corp) and SAS version 03.05 software (SAS Institute). Descriptive statistics were computed for all variables. Independent 2-tailed sample *t* tests were conducted to compare mean waiting times before and after the intervention. To analyze changes in appointment no-show rates, *z* tests were performed. Logistic regression was used to estimate odds ratios (OR) and determine the likelihood of a no-show after implementation. Statistical significance was set at an α level of .05, with 95% CI computed for all estimates. The statistical validation involved performing hypothesis tests to determine the significance of changes in waiting times and no-show rates.

Outcome Measures

The outcome measures were the mean difference in waiting times, which was calculated using the waiting times before and after implementation and the percent reduction in no-shows, which was calculated using the no-show rates before and after implementation.

Ethical Considerations

Ethical approval was obtained from the Dubai Research Ethics Committee at the Ministry of Health and Prevention, United Arab Emirates. (Approval Reference No. MOHAP/DXB-REC/S.O.O/No.136/2023). The study involved deidentified, retrospective data negating the need for individual patient consent. Patient compensation was not applicable for this study. Bias was minimized by using deidentified data and implementing objective data quality control measures.

Results

We implemented our project across all PHCs in the EHS in October 2022. To compare the outcomes of this tool, we analyzed the data 3 months before (July to September 2022) and 3 months after (November 2022, to January 2023) implementation, and the outcomes were grouped and labeled as before or after, respectively. Descriptive analytics revealed a total of 67,429 booked visits before implementation compared to 67,964 visits after implementation with a similar sex and nationality distribution between both periods (Multimedia Appendix 5). Our analysis showed that most of the visits were through walk-ins instead of booked appointments. There was a slight decrease in walk-in appointments between the two periods, but the percentage of these visits remained relatively stable (80.9% to 80.7%). There was a notable change in the number of no-show appointments, drastically reducing from 20.82% to 10.25% (Table 1).

Table . Visit frequencies and dynamics before and after program implementation in primary health care.

	Before implementation, n/N (%)	After implementation, n/N (%)
Walk-ins	285,700/353,129 (80.91)	284,015/351,979 (80.69)
Appointments	67,429/353,129 (19.09)	67,964/351,979 (19.31)
No-shows	14,038/67,429 (20.82)	6966/67,964 (10.25)

Using a *z* test, our analysis showed that the implementation of the no-show AI model led to a significant 50.7% reduction in no-show rates across the PHC department (*P*<.001). Binary logistic regression further assessed the dashboard's impact on reducing no-shows, yielding an OR of 0.43 (95% CI 0.42 - 0.45; *P*<.001). This OR being less than 1 indicated that the intervention effectively decreased no-show occurrences, reducing the likelihood of no-shows by 57% after implementation of the AI model (Table 2).

Table . Mean difference and impact analysis of appointment no-show rates before and after program implementation in primary health care centers.

Measurements	Values	*P* value
Difference in no-show rate after implementation, n/N (%)	50.7	<.001
Likelihood of no-shows after implementation, odds ratio (95% CI)	0.43 (0.42-0.45)	<.001

Table 3 reflects the different waiting times for the nurse assessment, physician assessment, and overall consultation time (Multimedia Appendix 6). We observed reduced average waiting times at all stations across the PHC journey. We found a

XSL•FO
RenderX

decrease in patient wait times by an average of 5.7 minutes (from 54 to 49 minutes; $P<.001$). Based on the average time saved per visit (5.7 minutes), a total of 387,394.8 minutes were saved in the 3 months after implementation. This is equivalent to approximately 6,456.6 hours saved during this period. The reduction in waiting time from check-in to checkout has been consistent since implementing this solution.

Table . Comparison of patient waiting times at different stations during patient visits before and after program implementation at primary health care centers.

Waiting times	Before implementation (min), mean (SE)	After implementation (min), mean (SE)	Difference following implementation (min)	*P* value
Nurse waiting time	8.4 (0.02)	7.2 (0.02)	1.2	<.001
Physician waiting time	16.3 (0.03)	15.0 (0.03)	1.3	<.001
Physician consultation to checkout time	29.3 (0.06)	26.2 (0.06)	3.1	<.001
Check-in to checkout time	54.2 (0.08)	48.5 (0.07)	5.7	<.001

Discussion

Principal Results and Comparison With Prior Work

The implementation of an AI-powered no-show prediction model, integrated with a real-time operational dashboard at PHC centers within the EHS, has demonstrated significant improvements in operational efficiency and patient care. Our main findings revealed a significant 50.7% reduction in no-show appointments *(P*<.001), and a 5.7-minute reduction in overall patient waiting times, highlighting the direct impact of our intervention on streamlining health care delivery. These results underlined the value of leveraging predictive analytics and real-time data in optimizing resource allocation, reducing manual processes, and enhancing the patient experience. Unlike other studies, where the emphasis has been to explore the factors and behaviors contributing to patient no-shows using machine learning and other techniques [24], we successfully operationalized our machine learning model and connected it to a closed loop workflow which helped reduce absenteeism.

A key achievement of this intervention was the substantial reduction in no-show appointments. By proactively contacting high-risk patients identified by the AI model, clinic administrators were able to significantly reduce missed appointments. This was a notable success, as it not only improved the efficiency of the appointments but also allowed walk-in patients to fill the available slots, further maximizing clinic capacity. Additionally, the use of real-time dashboards provided clinic managers with immediate insights into patient volumes, waiting times, and staff availability, enabling swift adjustments in resource allocation during peak hours. These operational efficiencies, particularly the reduction in waiting times at multiple stages of the patient journey (nurse assessment, physician consultation, and overall clinic visit), underscore the importance of real-time analytics in improving the flow of care. Our robust analysis reassured that the findings of this before and after implementation were significant and could be attributed to our intervention.

The impact of these factors has been previously studied in isolation. There are successful examples of AI models to predict no-show appointments, which have helped reduce the burden of these instances [25-27]. Our observations were in alignment with these studies. Like other studies, we have seen that connecting with patients before their scheduled appointment reduces their chances of missing it [28]. Several strategies, like text messages, patient portal reminders, and phone calls, have been examined, and it has been found that calling patients before their scheduled visit is effective for reducing no-shows [29].

Operationally, real-time dashboards and decision support tools for clinic managers are helpful for identifying operational bottlenecks. Numerous studies have examined the impact of real-time dashboards on reducing patient waiting times in health care settings. These studies suggest that real-time dashboards are crucial for improving patient flow and reducing waiting times across various health care settings, emphasizing their potential as valuable tools in optimizing health care delivery [30-33].

As recommended by other studies, using data to calculate and predict interday scheduling precisely is undeniably beneficial for large health care organizations [30,34]. Trend analysis helps estimate the expected number of unbooked versus booked visits and helps allocate resources accordingly. Many PHC centers reserve a proportion of their scheduling slots for walk-in visits. Since the EHS has many walk-in visits, we should follow a similar strategy and use predictive analysis to help improve our resource allocation. We must carefully design this strategy to reduce our patient waiting times. In the future, we could further enhance the outcomes achieved in this study. By scaling up the AI-model to other services and departments, we could increase the benefits of this program for the EHS. The flexibility and the scalability of our real-time data intelligence platform makes it a valuable tool to resolve operational challenges in our health care system.

Strengths

This study presents a novel approach by integrating an AI-driven predictive model with a real-time dashboard to manage no-show appointments and reduce waiting times in a PHC setting. Unlike previous studies, which primarily focus on predictive modeling without an operational framework, our approach goes a step further by embedding predictive insights into an actionable, real-time dashboard that enables proactive intervention. By operationalizing predictive analytics within a large-scale health care network, this study contributes a scalable and practical

XSL•FO
RenderX

solution to the current state of digital health interventions aimed at improving health care efficiency and patient outcomes.

The major strengths of our study include big data analysis, the use of case mix, the contextualization to the local population, crossfunctional analytics for the clinician-to-patient ratio, real-time analysis for immediate identification of prolonged waiting times, and AI-based modeling for no-show appointment prediction that is coupled with patient contact information. Our holistic approach to use an AI model and real-time insights for a guided operational management workflow is unique in its nature and it can be used by organizations with similar resources.

Limitations

Certain limitations of this study must be acknowledged. First, the mitigation workflow is outside of EHRs, so we cannot quantify the efforts accurately. It is an established fact that a reduction in no-show appointments result in improved financial outcomes [35,36], and we acknowledge that a detailed analysis of the financial implications of reducing no-shows was beyond the scope of this study. Future research could address this aspect to provide a more comprehensive understanding of the intervention's economic impact.

Conclusion

In conclusion, this project represents the first instance of the EHS using a real-time, AI-driven analytics platform to improve operational efficiency, reduce no-show appointments, and enhance patient experience. The success of our solution in reducing no-shows and waiting times highlights the importance of integrating advanced analytics into daily health care operations. By providing clinic administrators with real-time actionable data, this platform enables better decision-making, optimized resource use, and improved patient flow. The scalability of this application across other health care settings further emphasizes its potential to drive widespread improvements in health care delivery, making it a valuable tool for health care organizations looking to modernize their operations through data-driven solutions. As we continue to refine this system, we expect further reductions in waiting times and improvements in patient satisfaction.

Acknowledgments

We would like to acknowledge Benjamin Vinodson. The study was fully funded by the Emirates Health Services, with no external sources involved.

Data Availability

The data sets generated during and/or analyzed during this study are available from the corresponding author on reasonable request

Authors' Contributions

NMI was the principal investigator of this project. All authors contributed to the design, analysis, and manuscript creation according to the sequence of authorship assigned.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data architecture diagram representing the layers of data transition from source of information to the dashboard.
[PPTX File, 718 KB - formative_v9i1e64936_app1.pptx]

Multimedia Appendix 2
Program workflow for using the real-time artificial intelligence–powered dashboard to manage no-shows and reduce waiting times in primary health care.
[PPTX File, 45 KB - formative_v9i1e64936_app2.pptx]

Multimedia Appendix 3
A screenshot of the data dashboard showing real-time analytics for primary health care visits in a day.
[PPTX File, 1025 KB - formative_v9i1e64936_app3.pptx]

Multimedia Appendix 4
Major predictors contributing to the no-show prediction artificial intelligence model from primary health care visit data.
[PPTX File, 139 KB - formative_v9i1e64936_app4.pptx]

Multimedia Appendix 5

XSL•FO
RenderX

Frequency distribution of booked visits before and after the artificial intelligence program implementation in primary health care centers (July 2022, to January 2023).
[PDF File, 84 KB - formative_v9i1e64936_app5.pdf]

Multimedia Appendix 6
Reduction in patient waiting times at different stages of care following artificial intelligence program implementation in primary health care centers.
[PPTX File, 48 KB - formative_v9i1e64936_app6.pptx]

References

1. Primary health care. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/primary-health-care [accessed 2024-10-09]
2. Leslie LK, Mehus CJ, Hawkins JD, et al. Primary health care: potential home for family-focused preventive interventions. Am J Prev Med 2016 Oct;51(4 Suppl 2):S106-S118. [doi: 10.1016/j.amepre.2016.05.014] [Medline: 27498167]
3. Guzek LM, Gentry SD, Golomb MR. The estimated cost of “no-shows” in an academic pediatric neurology clinic. Pediatr Neurol 2015 Feb;52(2):198-201. [doi: 10.1016/j.pediatrneurol.2014.10.020] [Medline: 25497123]
4. Moore CG, Wilson-Witherspoon P, Probst JC. Time and money: effects of no-shows at a family practice residency clinic. Fam Med 2001;33(7):522-527. [Medline: 11456244]
5. Harris SL, Samorani M. On selecting a probabilistic classifier for appointment no-show prediction. Decis Support Syst 2021 Mar;142:113472. [doi: 10.1016/j.dss.2020.113472]
6. Lin Q, Betancourt B, Goldstein BA, Steorts RC. Prediction of appointment no-shows using electronic health records. J Appl Stat 2020;47(7):1220-1234. [doi: 10.1080/02664763.2019.1672631] [Medline: 35707022]
7. Denney J, Coyne S, Rafiqi S. Machine learning predictions of no-show appointments in a primary care setting. SMU Data Sci Rev 2019;2(1) [FREE Full text]
8. Harvey HB, Liu C, Ai J, et al. Predicting no-shows in radiology using regression modeling of data available in the electronic medical record. J Am Coll Radiol 2017 Oct;14(10):1303-1309. [doi: 10.1016/j.jacr.2017.05.007] [Medline: 28673777]
9. Yang Y, Madanian S, Parry D. Enhancing health equity by predicting missed appointments in health care: machine learning study. JMIR Med Inform 2024 Jan 12;12:e48273. [doi: 10.2196/48273] [Medline: 38214974]
10. Dansky KH, Miles J. Patient satisfaction with ambulatory healthcare services: waiting time and filling time. Hosp Health Serv Adm 1997;42(2):165-177. [Medline: 10167452]
11. Huang XM. Patient attitude towards waiting in an outpatient clinic and its applications. Health Serv Manage Res 1994 Feb;7(1):2-8. [doi: 10.1177/095148489400700101] [Medline: 10133292]
12. Leddy KM, Kaldenberg DO, Becker BW. Timeliness in ambulatory care treatment. An examination of patient satisfaction and wait times in medical practices and outpatient test and treatment facilities. J Ambul Care Manage 2003;26(2):138-149. [doi: 10.1097/00004479-200304000-00006] [Medline: 12698928]
13. Michael M, Schaffer SD, Egan PL, Little BB, Pritchard PS. Improving wait times and patient satisfaction in primary care. J Healthc Qual 2013;35(2):50-59. [doi: 10.1111/jhq.12004] [Medline: 23480405]
14. Vinitsky A, David B, Michalik L, et al. Reducing wait time in a high-volume pediatric neuro-oncology clinic by optimizing process flow: a quality improvement project. Pediatr Qual Saf 2022;7(3):e557. [doi: 10.1097/pq9.0000000000000557] [Medline: 35720861]
15. Santibáñez P, Chow VS, French J, Puterman ML, Tyldesley S. Reducing patient wait times and improving resource utilization at British Columbia Cancer Agency’s ambulatory care unit through simulation. Health Care Manag Sci 2009 Dec;12(4):392-407. [doi: 10.1007/s10729-009-9103-1] [Medline: 20058528]
16. Le V, Wagar EA, Phipps RA, Del Guidice RE, Le H, Middleton LP. Improving patient experience of wait times and courtesy through electronic sign-in and notification in the phlebotomy clinic. Arch Pathol Lab Med 2020 Jun;144(6):769-775. [doi: 10.5858/arpa.2019-0139-OA] [Medline: 31755777]
17. Almomani I, AlSarheed A. Enhancing outpatient clinics management software by reducing patients’ waiting time. J Infect Public Health 2016;9(6):734-743. [doi: 10.1016/j.jiph.2016.09.005] [Medline: 27663518]
18. Saeed S, Somani N, Sharif F, Kazi AM. Evaluating the effectiveness of text messaging and phone call reminders to minimize no show at pediatric outpatient clinics in Pakistan: protocol for a mixed-methods study. JMIR Res Protoc 2018 Apr 10;7(4):e91. [doi: 10.2196/resprot.9294] [Medline: 29636321]
19. Horvath M, Levy J, L’Engle P, Carlson B, Ahmad A, Ferranti J. Impact of health portal enrollment with email reminders on adherence to clinic appointments: a pilot study. J Med Internet Res 2011 May 26;13(2):e41. [doi: 10.2196/jmir.1702] [Medline: 21616784]
20. Milosevic Z, Chen W, Berry A, Rabhi F. Real-time analytics. In: Big Data 2016:39-61. [doi: 10.1016/B978-0-12-805394-2.00002-7]
21. Wills MJ. Decisions through data: analytics in healthcare. J Healthc Manag 2014;59(4):254-262. [doi: 10.1097/00115514-201407000-00005] [Medline: 25154123]

XSL•FO
RenderX

22. Kaur J, Kulwinder Singh Mann D. AI based healthcare platform for real time, predictive and prescriptive analytics using reactive programming. J Phys Conf Ser 2018 Jan 3;933(1):012010. [doi: 10.1088/1742-6596/933/1/012010]
23. Emirates Health Services - UAE. URL: https://www.ehs.gov.ae/en/home [accessed 2024-12-20]
24. Salazar LHA, Parreira WD, Fernandes ADR, Leithardt VRQ. No-show in medical appointments with machine learning techniques: a systematic literature review. Inf 2022;13(11):507. [doi: 10.3390/info13110507]
25. AlMuhaideb S, Alswailem O, Alsubaie N, Ferwana I, Alnajem A. Prediction of hospital no-show appointments through artificial intelligence algorithms. Ann Saudi Med 2019;39(6):373-381. [doi: 10.5144/0256-4947.2019.373] [Medline: 31804138]
26. Goffman RM, Harris SL, May JH, et al. Modeling patient no-show history and predicting future outpatient appointment behavior in the veterans health administration. Mil Med 2017 May;182(5):e1708-e1714. [doi: 10.7205/MILMED-D-16-00345] [Medline: 29087915]
27. Dunstan J, Villena F, Hoyos JP, et al. Predicting no-show appointments in a pediatric hospital in Chile using machine learning. Health Care Manag Sci 2023 Jun;26(2):313-329. [doi: 10.1007/s10729-022-09626-z] [Medline: 36707485]
28. Liu C, Harvey HB, Jaworsky C, Shore MT, Guerrier CE, Pianykh O. Text message reminders reduce outpatient radiology no-shows but do not improve arrival punctuality. J Am Coll Radiol 2017 Aug;14(8):1049-1054. [doi: 10.1016/j.jacr.2017.04.016] [Medline: 28583321]
29. Childers RE, Laird A, Newman L, Keyashian K. The role of a nurse telephone call to prevent no-shows in endoscopy. Gastrointest Endosc 2016 Dec;84(6):1010-1017. [doi: 10.1016/j.gie.2016.05.052] [Medline: 27327847]
30. Cayirli T, Yang KK, Quek SA. A universal appointment rule in the presence of no‐shows and walk‐ins. Prod Oper Manag 2012 Jul;21(4):682-697. [doi: 10.1111/j.1937-5956.2011.01297.x]
31. Franklin A, Gantela S, Shifarraw S, et al. Dashboard visualizations: supporting real-time throughput decision-making. J Biomed Inform 2017 Jul;71:211-221. [doi: 10.1016/j.jbi.2017.05.024] [Medline: 28579532]
32. Marmor YN, Wasserkrug S, Zeltyn S, et al. Toward simulation-based real-time decision-support systems for emergency departments. Presented at: The 2009 Winter Simulation Conference (WSC); Dec 13-16, 2009; Austin, TX p. 2042-2053. [doi: 10.1109/WSC.2009.5429416]
33. Alaeddini A, Murty KG. DSS (decision support system) for allocating appointment times to calling patients at a medical facility. In: Murty KG, editor. Case Studies in Operations Research: Springer; 2015:83-109. [doi: 10.1007/978-1-4939-1007-6_5]
34. Wiesche L, Schacht M, Werners B. Strategies for interday appointment scheduling in primary care. Health Care Manag Sci 2017 Sep;20(3):403-418. [doi: 10.1007/s10729-016-9361-7] [Medline: 27000079]
35. Berg BP, Murr M, Chermak D, et al. Estimating the cost of no-shows and evaluating the effects of mitigation strategies. Med Decis Making 2013 Nov;33(8):976-985. [doi: 10.1177/0272989X13478194] [Medline: 23515215]
36. Kheirkhah P, Feng Q, Travis LM, Tavakoli-Tabasi S, Sharafkhaneh A. Prevalence, predictors and economic consequences of no-shows. BMC Health Serv Res 2016 Jan 14;16:13. [doi: 10.1186/s12913-015-1243-z] [Medline: 26769153]

Abbreviations

AI: artificial intelligence
EHR: electronic health record
EHS: Emirates Health Services
OR: odds ratio
PHC: primary health care

Edited by A Mavragani; submitted 12.08.24; peer-reviewed by A Ala, F Yang; revised version received 14.11.24; accepted 14.11.24; published 06.01.25.

Please cite as:
AlSerkal YM, Ibrahim NM, Alsereidi AS, Ibrahim M, Kurakula S, Naqvi SA, Khan Y, Oottumadathil NP
Real-Time Analytics and AI for Managing No-Show Appointments in Primary Health Care in the United Arab Emirates: Before-and-After Study
JMIR Form Res 2025;9:e64936
URL: https://formative.jmir.org/2025/1/e64936
doi:10.2196/64936

© Yousif Mohamed AlSerkal, Naseem Mohamed Ibrahim, Aisha Suhail Alsereidi , Mubaraka Ibrahim, Sudheer Kurakula, Sadaf Ahsan Naqvi, Yasir Khan, Neema Preman Oottumadathil. Originally published in JMIR Formative Research (https://formative.jmir.org), 6.1.2025. This is an open-access article distributed under the terms of the Creative Commons

XSL•FO
RenderX

Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effects of Smart Goggles Used at Bedtime on Objectively Measured Sleep and Self-Reported Anxiety, Stress, and Relaxation: Pre-Post Pilot Study

Sharon Danoff-Burg, PhD; Elie Gottlieb, PhD; Morgan A Weaver, BA; Kiara C Carmon, BA; Duvia Lara Ledesma, MA; Holly M Rus, PhD

Scientific Research, SleepScore Labs, 2175 Salk Avenue, Suite 150, Carlsbad, CA, United States

Corresponding Author:
Sharon Danoff-Burg, PhD
Scientific Research, SleepScore Labs, 2175 Salk Avenue, Suite 150, Carlsbad, CA, United States

Abstract

Background: Insufficient sleep is a problem affecting millions. Poor sleep can trigger or worsen anxiety; conversely, anxiety can lead to or exacerbate poor sleep. Advances in innovative consumer products designed to promote relaxation and support healthy sleep are emerging, and their effectiveness can be evaluated accurately using sleep measurement technologies in the home environment.

Objective: This pilot study examined the effects of smart goggles used at bedtime to deliver gentle, slow vibration to the eyes and temples. The study hypothesized that objective sleep, perceived sleep, self-reported stress, anxiety, relaxation, and sleepiness would improve after using the smart goggles.

Methods: A within-participants, pre-post study design was implemented. Healthy adults with subclinical threshold sleep problems (N=20) tracked their sleep nightly using a polysomnography-validated noncontact biomotion device and completed daily questionnaires over two phases: a 3-week baseline period and a 3-week intervention period. During the baseline period, participants followed their usual sleep routines at home. During the intervention period, participants used Therabody SmartGoggles in "Sleep" mode at bedtime. This mode, designed for relaxation, delivers a gentle eye and temple massage through the inflation of internal compartments to create a kneading sensation combined with vibrating motors. Each night, the participants completed questionnaires assessing relaxation, stress, anxiety, and sleepiness immediately before and after using the goggles. Daily morning questionnaires assessed perceived sleep, complementing the objective sleep data measured every night.

Results: Multilevel regression analysis of 676 nights of objective sleep parameters showed improvements during nights when the goggles were used compared to the baseline period. Key findings include sleep duration (increased by 12 minutes, P=.01); duration of deep sleep (increased by 6 minutes, P=.002); proportion of deep sleep (7% relative increase, P=.02); BodyScore, an age- and gender-normalized measure of deep sleep (4% increase, P=.002); number of nighttime awakenings (7% decrease, P=.02); total time awake after sleep onset (reduced by 6 minutes, P=.047); and SleepScore, a measure of overall sleep quality (3% increase, P=.02). Questionnaire responses showed that compared to baseline, participants felt they had better sleep quality (P<.001) and woke feeling more well-rested (P<.001). Additionally, participants reported feeling sleepier, less stressed, less anxious, and more relaxed (all P values <.05) immediately after using the goggles each night, compared to immediately before use. A standardized inventory administered before and after the 3-week intervention period indicated reduced anxiety (P=.03), confirming the nightly analysis.

Conclusions: The use of smart goggles at bedtime significantly improved objectively measured sleep metrics and perceived sleep quality. Further, participants reported increased feelings of relaxation along with reduced stress and anxiety. Future research expanding on this pilot study is warranted to confirm and expand on the preliminary evidence presented in this brief report.

(JMIR Form Res 2025;9:e58461) doi:10.2196/58461

KEYWORDS

relaxation; stress; anxiety; sleep; health technology; intervention

Introduction

Insufficient sleep affects approximately one-third of the population [1] and is associated with adverse health outcomes and impaired performance [2]. Poor sleep can trigger or worsen anxiety, while conversely, anxiety can lead to or exacerbate poor sleep [3]. Technological advances in unobtrusive sleep measurement enable intervention studies to be conducted in the comfort of research participants' own bedrooms, providing ecologically valid results while capturing accurate objective

XSL•FO
RenderX

data [4,5]. Concomitantly, the development of innovative consumer products designed to promote relaxation and support healthy sleep is emerging. Their effectiveness can be evaluated in field studies using ambulatory measurement technologies [6].

The aim of this research was to examine the effects of smart goggles used at bedtime to deliver gentle vibration to the eyes and temples, on sleep as well as on perceived stress, anxiety, and relaxation. Although a variety of evidence-based relaxation techniques [7] and sleep-enhancing products [8] already exist, devices such as smart goggles may appeal to individuals who wish to use such technological tools as an option within their repertoire of strategies for winding down at bedtime. Previous research suggests that vibration can increase relaxation [9], induce drowsiness [10], and may be a useful nonpharmacological intervention for poor sleep [11-13]. For example, a preliminary study documented the use of vibration in improving objectively measured sleep outcomes in people with mild to moderate symptoms of insomnia [11]. In this pilot study on adults with subclinical threshold sleep problems, we hypothesized that objective and perceived sleep outcomes as well as self-reported stress, anxiety, relaxation, and sleepiness would improve after using smart goggles delivering slow vibrations at bedtime.

Methods

Participants

Invitations to complete an eligibility questionnaire for a study testing smart goggles were emailed to registrants in a large database of individuals interested in sleep research and using SleepScore technology. Eligible respondents were invited to enroll based on the following selection criteria: difficulty falling or staying asleep, no history of sleep disorders, absence of other medical issues affecting sleep, substance use that could affect sleep, and no lifestyle issues such as shift work that might influence their sleep. The study included adults (N=20) with subclinical threshold sleep problems who were willing to track their sleep and use the smart goggles as instructed.

Design and Procedures

A within-participants, pre-post study design was implemented. The participants were aware that the intervention had the potential to affect their sleep. Following a 3-week baseline period during which participants measured their sleep data at home without any intervention, they used the Therabody SmartGoggles (Therabody Inc, Los Angeles, CA) at bedtime for 15 minutes (within 30 minutes of their intended sleep time) over a 3-week intervention period. The participants were instructed to use “Sleep” mode, designed for relaxation and inducing sleepiness. This mode delivers slow and gentle massage to the eye and temple areas by the inflation of internal compartments to create a kneading sensation and vibrating motors. Additionally, two other modes (SmartRelax and Focus) are available; however, since these provide different experiences, they were not used in this study.

During the entire 6-week study, nightly objective sleep measurements were collected and participants completed online questionnaires each morning and evening. Data collection was synchronized across all participants to account for weekday or weekend variation.

Measurement

Objective sleep data were collected with SleepScore Max (Consumer Sleep Solutions LLC, Carlsbad, CA), a noncontact monitoring device that uses respiratory signal and motion sensing to detect sleep. The device is placed next to the bed and controlled using a companion app. It uses ultra-low power radiofrequency waves to monitor body movement and respiration patterns when in bed; the measurement is unaffected by bedding or nightwear. If a partner is present, only the sleep of the individual closest to the device is measured. The device captures high-resolution magnitude and duration data of gross movements, micromovements, and full breathing cycles, which are transformed into 30-second epoch sleep stage data (wake, light, deep, rapid eye movement [REM]) using proprietary algorithms. Studies have shown good agreement with gold-standard polysomnography [14,15], exceeding the accuracy typically reported for actigraphy-based devices [16].

Using the 30-second epoch data, standard sleep metrics were calculated. In addition, 3 SleepScore technology proprietary sleep metrics reflecting sleep quality, all ranging from 0 to 100 and normalized for age and gender using reference values from the meta-analysis of quantitative sleep parameters by Ohayon and colleagues [17], were calculated:

- *SleepScore* is an overall sleep quality metric that includes objectively measured total sleep time, sleep onset latency, and sleep stage durations.
- *BodyScore* reflects the age- and gender-normalized amount of deep (non–rapid eye movement stage 3 [NREM-3]) sleep.
- *MindScore* reflects the age- and gender-normalized amount of REM sleep.

Self-reported data were collected daily, across the entire study via 100-point visual analog scales. Morning assessments measured perceived sleep quality and feeling well-rested upon waking. At night, the 100-point visual analog scales assessed relaxation, stress, anxiety, and sleepiness before and after goggle use. The construct validity and discriminate sensitivity of visual analog scales to assess perceived stress and related constructs have been documented [18]. Participants completed these scales immediately before using the goggles at bedtime and after using the goggles for 15 minutes, and then went to sleep. A 6-item version of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI) [19] was administered once before and once after the intervention period.

Statistical Analyses

Nightly objective sleep data and daily self-reported data were analyzed using multilevel regression with random intercept, accounting for nights nested within participants, comparing nights during the baseline period to nights during the intervention period for each outcome. The regression model was $Sleepmeasure_{ij} = \beta_0 + \beta_1 * TestPeriod_{ij} + u_{0j} + e_{ij}$; TestPeriod, coded as 0 for observations during baseline and 1 for nights during the intervention period. Similarly, analysis of the nightly self-reported data used the same model to compare pre- and

XSL•FO
RenderX

post-goggle use. A 2-tailed paired-samples *t* test was used to analyze changes in the 6-item STAI scores.

Discrepancies in sample sizes (N=20 for objective and self-reported sleep, n=17 for 6-item STAI) were due to incomplete data sources. Participants tracked their sleep at home and, at times, were not fully compliant with the use of measurement tools or the completion of online surveys. This is common in field research collecting longitudinal and daily assessment data. All reported results reflect the largest available sample for each set of analyses.

Ethical Considerations

The study was approved by Sterling Institutional Review Board (ID 11012), and all procedures were conducted in accordance with the ethical standards of the Declaration of Helsinki. All participants provided informed consent using a secure online platform to review, electronically sign, and return a copy of the document to the research coordinator. They were informed that the study was voluntary, and of their right to withdraw at any time. Both objective and self-reported study data were deidentified prior to analysis and accessible only to members of the research team. As compensation, participants kept the smart goggles and sleep measurement device used during the study.

Results

Of the 20 participants, 40% (n=8) were women, and the age range was 26-75 (mean 50.41, SD 13.12) years. Further demographic details are provided in Table 1.

XSL•FO
RenderX

Table . Demographic information of participants at baseline (N=20).

Demographics	Values
Age (years)	
Mean (SD)	50.41 (13.12)
Range	26 - 75
Gender, n (%)	
Men	8 (40)
Women	12 (60)
Other identities	0 (0)
Race/Ethnicity, n (%)	
American Indian or Alaska Native	0 (0)
Asian	1 (5)
Black or African American	0 (0)
Hispanic/Latino	4 (20)
Native Hawaiian or other Pacific Islander	1 (5)
White	14 (70)
Household composition, n (%)	
I live alone	7 (35)
I live with my partner - (currently) no children in the house	3 (15)
I live with my children - (currently) no partner in the house	1 (5)
I live with my partner and child(ren)	6 (30)
I live with a family member(s)	3 (15)
I live with a roommate(s)	0 (0)
Other	0 (0)
Employment, n (%)	
Working full-time	14 (70)
Working part-time	2 (10)
Homemaker	2 (10)
Full-time student	1 (5)
Retired	1 (5)
Education, n (%)	
High school degree or equivalent (eg, GED[a])	1 (5)
Some college, no degree	5 (25)
Associate degree	1 (5)
Bachelor’s degree	8 (40)
Master’s degree	3 (15)
Doctoral degree	2 (10)
Annual household income (US$), n (%)	
<$25,000	0 (0)
$25,000-$49,999	3 (15)
$50,000-$74,999	4 (20)
$75,000-$99,999	3 (15)
$100,000-$124,999	2 (10)
$125,000-$149,999	2 (10)

XSL•FO
RenderX

$150,000-$174,999	0 (0)
$175,000-$199,999	0 (0)
≥$200,000	5 (25)
Would rather not answer this question	1 (5)

[a]GED: General Educational Development.

Objective Sleep

Nightly objective measurement of sleep (n=676 nights nested within 20 participants) revealed multiple improvements when participants used the goggles at bedtime. Key findings included increased sleep duration (+12 min, *P*=.01); increased deep sleep, reflected both in duration (+6 min, *P*=.002) and proportion of the night (7% relative increase, *P*=.02); enhanced BodyScore (+4%, *P*=.002); fewer nighttime awakenings (−7%, *P*=.02); reduced total wake time at night after sleep onset (−6 min, *P*=.047); and improved SleepScore, indicating overall sleep quality (+3%, *P*=.02). Detailed objective sleep metrics are displayed in Table 2.

Table . Multilevel regression results for objective sleep (n=676 nights nested within 20 participants), comparing the baseline period (sleep at home prior to intervention) to the intervention period (sleep at home using smart goggles at bedtime).

Outcomes	Objective measurement of sleep, observed mean (SD)[a]		Estimated marginal means[b]		
	Baseline period	Intervetion period	Intercept (SE)	β[c]	*P* value
Total sleep time in minutes	331.55 (57.49)	343.48 (62.69)	331.52 (4.87)	12.04	.01
Sleep onset latency in minutes	23.34 (18.33)	23.20 (18.17)	23.28 (1.64)	−0.13	.92
Number of awakenings	5.51 (2.10)	5.14 (2.12)	5.51 (0.17)	−0.38	.02
Time spent awake after sleep onset in minutes	60.50 (31.30)	54.32 (31.38)	60.17 (2.72)	−5.39	.047
Light sleep in minutes	215.36 (49.64)	219.86 (50.57)	215.47 (3.99)	4.54	.26
Deep sleep in minutes	57.52 (19.92)	63.19 (24.27)	57.66 (1.79)	5.58	.002
REM[d] sleep in minutes	58.67 (23.54)	60.44 (25.55)	58.52 (1.93)	1.94	.32
Percentage of time spent awake after sleep onset	16 (7)	14 (7)	15.36 (0.58)	−1.02	.08
Percentage of time in light sleep	54 (8)	55 (8)	54.50 (0.60)	0.02	.97
Percentage of time in deep sleep	15 (6)	16 (7)	15.10 (0.49)	1.14	.02
Percentage of time in REM sleep	15 (6)	15 (6)	14.92 (0.46)	−0.12	.80
SleepScore[e]	69.57 (10.89)	71.65 (10.87)	69.61 (0.87)	2.02	.02
BodyScore[e]	73.14 (10.38)	76.13 (11.15)	73.42 (0.86)	2.64	.002
MindScore[e]	69.30 (13.76)	69.68 (14.48)	69.18 (1.13)	0.55	.62

[a]For the baseline and intervention periods, each mean was calculated by averaging nights across participants, then averaging those participants' averages to a single simple average.

[b]These are the outcomes of separate multilevel regression analyses. Each row shows results from a different single-predictor, single-outcome model.

[c]The beta values are unstandardized and can therefore be interpreted on the same scale as the original data.

[d]REM: rapid eye movement.

[e]These scores range from 0 to 100. SleepScore is an age- and gender-normalized measure of overall sleep quality, BodyScore is an age- and gender-normalized measure of deep sleep, and MindScore is an age- and gender-normalized measure of REM sleep.

Self-Reported Sleep Quality and Well-Restedness

Multilevel regression analyses of daily self-reported sleep data (N=723 nights nested within 20 participants) showed that participants perceived better sleep quality (β=12.37, *P*<.001) and felt more rested in the morning (β=12.13, *P*<.001) when

XSL•FO
RenderX

using the goggles at bedtime compared to baseline, as assessed on a scale from 0 to 100.

Sleepiness, Anxiety, Stress, and Relaxation

Across 334 nights of the intervention period, multilevel regression analyses comparing responses immediately after using the goggles to those reported immediately before showed that participants felt sleepier (β=9.98, $P<.001$), less stressed (β=−10.38, $P<.001$), less anxious (β=−12.87, $P<.001$), and more relaxed (β=11.76, $P<.001$), all rated on a scale from 0 to 100.

At the end of the intervention period, compared to the end of the baseline period, participants' scores on the 6-item STAI showed reduced anxiety ($t_{16}=2.31$, $P=.03$), reflecting a 10% decrease and confirming the nightly analyses.

Discussion

Nonpharmacological techniques for promoting relaxation and improving sleep have the potential to help millions of individuals who experience suboptimal sleep [20]. This study evaluated the effectiveness of smart goggles designed to induce relaxation and support healthy sleep. Outcomes were measured using both self-reported data and a polysomnography-validated, noncontact biomotion device. The study population included a nonclinical sample of adults reporting poor sleep in the absence of diagnosed sleep disorders.

Compared to baseline, using the goggles at bedtime led to objective improvements in both sleep quality and duration. Although total sleep time remained less than 6 hours per night on average, objective improvements were seen in several parameters, including:

- sleep duration;
- increased deep sleep, both in duration and as a proportion of the night;
- reduced number of nighttime awakenings;
- decreased time spent awake at night after initially falling asleep; and
- enhanced sleep quality.

Aligning with these objective results, questionnaire data showed that participants perceived improvement in their sleep quality and felt more well-rested in the morning. In addition, immediately after using the goggles, participants felt sleepier, less stressed, less anxious, and more relaxed, compared to their experience immediately before using the goggles. A standardized inventory administered before and after the 3-week intervention period also indicated reduced anxiety, confirming the nightly analysis.

The observed improvements in the objective and self-reported sleep data may be attributed to increased relaxation resulting from the use of the goggles at bedtime. Vibration has been previously shown to be able to induce physiological relaxation [9,10] and support sleep [11-13]. This interpretation is supported by the changes in perceived relaxation, stress, and anxiety. However, objective parameters of physiological relaxation prior to sleep were not assessed, presenting an interesting avenue for future research.

Further studies could explore the intervention through a controlled trial, including comparison to goggles without vibration or to other relaxation techniques. This study assessed changes in parameters from the baseline to the intervention period while using the product at home, resembling how it is used outside of a research setting. The within-participants study design, which included long-term product use following a baseline period without the intervention, provides confidence that significant effects are due to the intervention itself. Although this study design has limitations, it reflects the real-world experience of introducing an intervention into the home environment, while also accounting for night-to-night variations in sleep patterns.

While longitudinal data provides an advantage by enabling the detection of nightly within-person differences, this pilot study is limited by its small sample size. Although we compared participants' improvements in sleep and related outcomes to their own baseline, the small sample size could have introduced a potential bias, and therefore these results should be interpreted with caution. Future research should investigate the effects of this intervention using larger sample sizes. Another limitation is that contactless technology is not ambulatory and therefore does not capture objective measures of daytime activity. However, the noncontact sleep measurement system used in this study offers the advantage of significantly higher overall accuracy, particularly in specificity, compared to traditional research-based actigraphy [16].

To conclude, this study found that using vibration therapy administered via smart goggles before going to sleep was associated with improvements in both objective sleep measures and self-reported outcomes. While these findings must be interpreted with caution, our data suggest that a device delivering gentle vibrations to the eye and temple area may have the potential to promote relaxation, decrease anxiety, and support healthy sleep when used at bedtime by adults with suboptimal sleep. Not only did sleep quality and sleep duration increase relative to baseline, but there was a decrease in the number of awakenings and the duration of time spent awake during the night. Further, improvement was seen in multiple metrics of deep sleep, which is vital for brain health and physical recovery. Future research expanding on this pilot study is warranted to confirm the preliminary evidence presented in this study.

Acknowledgments

Therabody provided the SmartGoggles and funding support to SleepScore Labs to design, conduct, and report the study.

XSL•FO
RenderX

Data Availability

The data analyzed in this study may be available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors are employed by SleepScore Labs.

References

1. Grandner MA. Epidemiology of insufficient sleep and poor sleep quality. In: Grandner MA, editor. Sleep and Health: Sleep Health Acad Press; 2019:11-20. [doi: 10.1016/B978-0-12-815373-4.00002-2]
2. Consensus Conference Panel, Watson NF, Badr MS, et al. Recommended amount of sleep for a healthy adult: a joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society. J Clin Sleep Med 2015 Jun 15;11(6):591-592. [doi: 10.5664/jcsm.4758] [Medline: 25979105]
3. Chellappa SL, Aeschbach D. Sleep and anxiety: from mechanisms to interventions. Sleep Med Rev 2022 Feb;61:101583. [doi: 10.1016/j.smrv.2021.101583] [Medline: 34979437]
4. Guillodo E, Lemey C, Simonnet M, et al. Clinical applications of mobile health wearable-based sleep monitoring: systematic review. JMIR Mhealth Uhealth 2020 Apr 1;8(4):e10733. [doi: 10.2196/10733] [Medline: 32234707]
5. Ko PRT, Kientz JA, Choe EK, Kay M, Landis CA, Watson NF. Consumer sleep technologies: a review of the landscape. J Clin Sleep Med 2015 Dec 15;11(12):1455-1461. [doi: 10.5664/jcsm.5288] [Medline: 26156958]
6. Watson NF, Lawlor C, Raymann RJEM. Will consumer sleep technologies change the way we practice sleep medicine? J Clin Sleep Med 2019 Jan 15;15(1):159-161. [doi: 10.5664/jcsm.7596] [Medline: 30621844]
7. Bertisch SM, Wells RE, Smith MT, McCarthy EP. Use of relaxation techniques and complementary and alternative medicine by American adults with insomnia symptoms: results from a national survey. J Clin Sleep Med 2012 Dec 15;8(6):681-691. [doi: 10.5664/jcsm.2264] [Medline: 23243402]
8. Maust DT, Solway E, Clark SJ, Kirch M, Singer DC, Malani P. Prescription and nonprescription sleep product use among older adults in the United States. Am J Geriatr Psychiatry 2019 Jan;27(1):32-41. [doi: 10.1016/j.jagp.2018.09.004] [Medline: 30409547]
9. Zhou Y, Murata A, Watanabe J. The calming effect of heartbeat vibration. Presented at: 2020 IEEE HAPTICS Symposium (HAPTICS); Mar 28-31, 2020; Crystal City, VA. [doi: 10.1109/HAPTICS45997.2020.ras.HAP20.157.5a2e1551]
10. Zhang N, Fard M, Bhuiyan MHU, Verhagen D, Azari MF, Robinson SR. The effects of physical vibration on heart rate variability as a measure of drowsiness. Ergonomics 2018 Sep;61(9):1259-1272. [doi: 10.1080/00140139.2018.1482373] [Medline: 29871584]
11. Himes BJ, Blotter JD, Kay DB, et al. The effect of beat frequency vibration on sleep latency and neural complexity: a pilot study. IEEE Trans Neural Syst Rehabil Eng 2021;29:872-883. [doi: 10.1109/TNSRE.2021.3076983] [Medline: 33950842]
12. Choi SH, Yoon H, Jin HW, et al. Effect of closed-loop vibration stimulation on heart rhythm during naps. Sensors (Basel) 2019 Sep 24;19(19):4136. [doi: 10.3390/s19194136] [Medline: 31554268]
13. Choi SH, Kwon HB, Jin HW, et al. Weak closed-loop vibrational stimulation improves the depth of slow-wave sleep and declarative memory consolidation. Sleep 2021 Jun 11;44(6):zsaa285. [doi: 10.1093/sleep/zsaa285] [Medline: 33367712]
14. O'Hare E, Flanagan D, Penzel T, Garcia C, Frohberg D, Heneghan C. A comparison of radio-frequency biomotion sensors and actigraphy versus polysomnography for the assessment of sleep in normal subjects. Sleep Breath 2015 Mar;19(1):91-98. [doi: 10.1007/s11325-014-0967-z] [Medline: 24614968]
15. Schade MM, Bauer CE, Murray BR, et al. Sleep validity of a non-contact bedside movement and respiration-sensing device. J Clin Sleep Med 2019 Jul 15;15(7):1051-1061. [doi: 10.5664/jcsm.7892] [Medline: 31383243]
16. Chinoy ED, Cuellar JA, Huwa KE, et al. Performance of seven consumer sleep-tracking devices compared with polysomnography. Sleep 2021 May 14;44(5):zsaa291. [doi: 10.1093/sleep/zsaa291] [Medline: 33378539]
17. Ohayon MM, Carskadon MA, Guilleminault C, Vitiello MV. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. Sleep 2004;27:1255-1273. [doi: 10.1093/sleep.27.7.1255] [Medline: 15586779]
18. Lesage FX, Berjot S, Deschamps F. Clinical stress assessment using a visual analogue scale. Occup Med (Lond) 2012 Dec;62(8):600-605. [doi: 10.1093/occmed/kqs140] [Medline: 22965867]
19. Marteau TM, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). Br J Clin Psychol 1992 Sep;31(3):301-306. [doi: 10.1111/j.2044-8260.1992.tb00997.x] [Medline: 1393159]
20. Maness DL, Khan M. Nonpharmacologic management of chronic insomnia. Am Fam Physician 2015 Dec 15;92(12):1058-1064. [Medline: 26760592]

Abbreviations

NREM-3: non–rapid eye movement stage 3

XSL•FO
RenderX

REM: rapid eye movement
STAI: State-Trait Anxiety Inventory

Edited by A Mavragani; submitted 15.03.24; peer-reviewed by GR Elkins, R Chamorro; revised version received 21.10.24; accepted 29.10.24; published 03.01.25.

Please cite as:
Danoff-Burg S, Gottlieb E, Weaver MA, Carmon KC, Lara Ledesma D, Rus HM
Effects of Smart Goggles Used at Bedtime on Objectively Measured Sleep and Self-Reported Anxiety, Stress, and Relaxation: Pre-Post Pilot Study
JMIR Form Res 2025;9:e58461
URL: https://formative.jmir.org/2025/1/e58461
doi:10.2196/58461

© Sharon Danoff-Burg, Elie Gottlieb, Morgan A Weaver, Kiara C Carmon, Duvia Lara Ledesma, Holly M Rus. Originally published in JMIR Formative Research (https://formative.jmir.org), 3.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Use of Social Media on Enhancing Dental Care and Practice Among Dental Professionals: Cross-Sectional Survey Study

Joseph Macadaeg Acosta[1,2], MDE, DMD; Palinee Detsomboonrat[3], MSc, DDS, PhD; Pagaporn Pantuwadee Pisarnturakit[3], MSc, DDS, PhD; Nipaporn Urwannachotima[3], MEd, DDS, PhD

[1]International Graduate Program in Dental Public Health, Department of Community Dentistry, Faculty of Dentistry, Chulalongkorn University, Bangkok, Thailand

[2]School of Dentistry, University of Baguio, Baguio City, Philippines

[3]Faculty of Dentistry, Department of Community Dentistry, Chulalongkorn University, 34 Henri Dunant Road, Pathumwan, Bangkok, Thailand

Corresponding Author:
Nipaporn Urwannachotima, MEd, DDS, PhD
Faculty of Dentistry, Department of Community Dentistry, Chulalongkorn University, 34 Henri Dunant Road, Pathumwan, Bangkok, Thailand

Abstract

Background: As digitalization continues to advance globally, the health care sector, including dental practice, increasingly recognizes social media as a vital tool for health care promotion, patient recruitment, marketing, and communication strategies.

Objective: This study aimed to investigate the use of social media and assess its impact on enhancing dental care and practice among dental professionals in the Philippines.

Methods: A cross-sectional survey was conducted among dental practitioners in the Philippines. The study used a 23-item questionnaire, which included 5 questions on dentists' background and demographic information and 18 questions regarding the use, frequency, and purpose of social media in patient advising and quality of care improvement. Data were analyzed using SPSS software, with frequency distributions and χ^2 tests used to assess the association between social media use and demographic variables and the impact on dental practice.

Results: The 265 dental practitioners in this study were predominantly female (n=204, 77%) and aged between 20 - 30 years (n=145, 54.7%). Most of the participants were general practitioners (n=260, 98.1%) working in a private practice (n=240, 90.6%), with 58.5% (n=155) having 0 - 5 years of clinical experience. Social media use was significantly higher among younger practitioners (20 - 30 years old) compared to older age groups (P<.001), though factors such as sex, dental specialty, and years of clinical practice did not significantly influence use. The majority (n=179, 67.5%) reported using social media in their practice, primarily for oral health promotion and education (n=191, 72.1%), connecting with patients and colleagues (n=165, 62.3%), and marketing (n=150, 56.6%). Facebook (n=179, 67.5%) and YouTube (n=163, 61.5%) were the most frequented platforms for clinical information, with Twitter (subsequently rebranded X) being the least used (n=4, 1.5%). Despite widespread social media engagement, only 8.7% (n=23) trusted the credibility of web-based information, and 63.4% (n=168) perceived a potential impact on the patient-dentist relationship due to patients seeking information on the internet. Social media was also perceived to enhance practice quality, with users reporting significant improvements in patient care (P=.001).

Conclusions: The findings highlight that social media is widely used among younger dental practitioners, primarily for education, communication, and marketing purposes. While social media use is associated with perceived improvements in practice quality and patient care, trust in information on social media remains low, and concerns remain regarding its effect on patient relationships. It is recommended to establish enhanced guidelines and provide reliable web-based resources to help dental practitioners use social media effectively and responsibly.

(JMIR Form Res 2025;9:e66121) doi:10.2196/66121

KEYWORDS

social media; oral health promotion; oral health education; dentists; dental practice; dental professionals; dental practitioners

Introduction

Traditionally, dental practices advertised their services primarily through local yellow pages, such as telephone directories listing local businesses or by placing promotional notices in the windows of dental offices [1]. Word-of-mouth was also a common method, where information about the quality of service was shared from one person to another [2]. However, with the advent of technological advancements, there has been a significant shift toward the use of social media. Social media

XSL•FO
RenderX

platforms offer the potential to reach diverse audiences, including those seeking information on oral health management, thereby promoting oral health [3].

Social media refers to a group of web-based applications built on the foundations of Web 2.0 [4,5]. According to the Merriam-Webster Dictionary, social media is defined as "any form of electronic communication through which users create web-based communities to share information, personal messages, ideas, and other content such as photos and videos" [6]. Popular platforms include social networking sites like Facebook and Twitter (subsequently rebranded X), media-sharing sites like YouTube and Instagram, blogs, and microblogging sites [7]. These platforms are widely used for searching, sharing, and communicating information, including health-related content [8,9]. Facebook fosters connections through its friend-based network, creating a web-based community where users can stay in touch and share content [10]. Twitter's microblogging feature allows users to share ideas and opinions in real time [11,12]. YouTube is a leading video-sharing platform [13-15], while Instagram, a photo and video social networking site, facilitates real-time communication through comments and direct messages and supports professional engagement and knowledge sharing [16-18]. TikTok, known for its short video content, has also gained popularity [19,20].

Social media serves as a dynamic tool for exchanging information and communicating with colleagues, patients, and the public on health-related issues. Public health professionals and organizations use these platforms for a variety of purposes, including health education, telemedicine, social marketing, scientific research, recruitment, career development, and professional networking [9]. Dental practitioners are increasingly leveraging social media to educate and inform the public by sharing detailed content on various dental procedures and educational activities [8,21-25], providing oral hygiene instructions and promoting oral health [26,27], enhancing communication [28-30], and engaging in marketing efforts [1,25,31,32]. Given the widespread use of social media, using these platforms responsibly and effectively in health care is crucial to ensure that they contribute positively to patient health outcomes.

As of April 2024, there were 5.44 billion internet users worldwide, representing 67.1% of the global population. Of this total, 5.07 billion, or 62.6% of the world's population, were social media users [33]. Based on the number of monthly active users, the most popular social networks globally are Facebook, YouTube, Instagram, WhatsApp, and TikTok [34]. In the Philippines, approximately 68.72 million people are social media users, with an average daily use of over 3 hours, the highest across the Asia-Pacific region. Facebook, Instagram, TikTok, and X (formerly known as Twitter) are currently some of the most popular platforms, providing opportunities for connecting with family and friends, sharing content via digital platforms, and expanding the reach of promotional marketing through web-based banners [35]. Social media in the Philippines has gained significant traction for entertainment, communication, marketing, and professional education. In dentistry, practitioners use these platforms to share content related to oral health, modern treatment options, trends in dental materials, and treatment costs.

Although several reports have highlighted the use of social media among dentists and its influence on dental practice [3,8,9,22,30,32,36,37], as well as its impact on dental education among faculty and students [7,21,23,38,39], relatively few studies have examined the use and influence of social media on dental practice in Southeast Asia, particularly in the Philippines where these platforms are widely used. This study aims to investigate the use of social media and assess its impact on enhancing dental care and practice among dental professionals in the Philippines.

Methods

Design and Study Population

This study used a cross-sectional survey design with purposive sampling, targeting all eligible licensed dental professionals, including both general dentists and specialists working in private or public clinical settings and who were involved in patient care, education, and practice management. Participants were also required to be active members of the Philippine Dental Association-Baguio City Chapter for the 2023 - 2024 term. The study population consisted of 265 dental practitioners. Practitioners who were not active members of the chapter were excluded from the study.

The survey questionnaire was adapted from previously published literature related to social media use [22,36,38]. For this study, social media included popular platforms such as Facebook, YouTube, Twitter, Instagram, TikTok, and Viber, which dental practitioners may have used for professional purposes like patient education, practice marketing, peer communication, and access to clinical information. A panel of experts validated the questionnaire items and the questionnaire underwent pretesting to ensure applicability and reliability. Revisions and minor adjustments were made based on the pilot study results and expert feedback. The 23-item questionnaire was divided into two sections. The first section gathered background and demographic information, including age, sex, dental specialty, work sector, and years of clinical practice. The second section comprised 18 questions designed to gather information on social media use; frequency of use; and purpose in dental practice, particularly its role in advising patients and improving the quality of care. This study was conducted within dental practice settings, with participants recruited through web-based methods. Recruitment emphasized voluntary participation, with all respondents providing informed consent before inclusion in the study. The questionnaire was administered via Google Forms, with the survey link disseminated through email, a Facebook page, and the Baguio City Chapter group chat. Respondents were encouraged to provide suggestions, assist in recruitment, and further share the questionnaire with other dental practitioners to reach more participants. They were given one day to complete the form.

Data Analysis

Data were analyzed using SPSS software version 28.0 (IBM Corp). The analysis included frequency distributions that were

XSL•FO
RenderX

expressed in numbers and percentages, as well as χ^2 tests to assess the association between social media use and demographic or social variables and the impact on dental practice. Statistical significance was set at an α level of .05. This study adhered to the reporting guidelines for cross-sectional studies following the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement.

Ethical Considerations

Ethical approval was obtained from the Human Research Ethics Committee of the Faculty of Dentistry, Chulalongkorn University (Study Code: HREC-DCU 2023 - 108). Informed consent was obtained from all participants through a written consent form, which provided details about the study's objectives and potential benefits. Participation in the study was entirely voluntary, with participants having the right to withdraw or discontinue at any time without any penalties or loss of benefits. All information was kept strictly confidential and anonymized to ensure participants could not be identified. No financial rewards or incentives were provided for participation.

Results

The background and demographic details of the 265 study participants are summarized in Table 1. The majority of participants were female (n=204, 77%), with over half (n=145, 54.7%) aged between 20 to 30 years. This was followed by participants aged 31 - 40 years (n=52, 19.6%), and both groups of participants aged 41 - 50 years and over 50 years represented 12.8% (n=34) of participants. Most participants (n=260, 98.1%) were general practitioners, and 90.6% (n=240) worked as private dentists. Regarding clinical experience, 58.5% (n=155) had 0 to 5 years of practice, while 29.8% (n=79) had over 10 years, and 11.7% (n=31) had 6 to 10 years of practice.

Table . Demographic variables and their association with the use of social media among dental practitioners.

Variable	Use of social media, n (%)[a]			Total (n=265), n (%)	*P* value[b]
	Yes	Sometimes	No		
Age (years)					<.001
20 - 30	102 (70.3)	35 (24.1)	8 (5.5)	145 (54.7)	
31 - 40	31 (60)	21 (40)	0 (0)	52 (19.6)	
41 - 50	27 (79)	2 (6)	5 (15)	34 (12.8)	
>50	19 (56)	15 (44)	0 (0)	34 (12.8)	
Sex					.66
Male	40 (66)	19 (31)	2 (3)	61 (23)	
Female	139 (68.1)	54 (26.5)	11 (5.4)	204 (77)	
Dental specialty					.25
General practitioners	177 (68.1)	70 (26.9)	13 (5)	260 (98.1)	
Specialists	2 (40)	3 (60)	0 (0)	5 (1.9)	
Working sector					.30
Government	10 (71)	2 (14)	2 (14)	14 (5.3)	
Private	160 (66.6)	69 (28.8)	11 (4.6)	240 (90.6)	
Both (government and private)	9 (82)	2 (18)	0 (0)	11 (4.2)	
Number of years of clinical practice					.19
0 to 5	109 (70.3)	38 (24.5)	8 (5.2)	155 (58.5)	
6 to 10	24 (77)	7 (23)	0 (0)	31 (11.7)	
>10	46 (58)	28 (35)	5 (6)	79 (29.8)	

[a]Percentages were calculated based on the number of participants corresponding to each variable.

[b]Chi-square test; α=.05.

There was a significant increase in social media use among the younger age group (20 - 30 years) compared to older age groups (31 - 40 years, 41 - 50 years, and >50 years), with a *P* value of <.001. However, sex, dental specialty, work sector, and years of clinical practice did not significantly influence social media use.

Table 2 illustrates the frequency distribution of social media use among dental practitioners. Of the 265 participants, most reported moderate use of social media with 37.4% (n=99) using it 1 - 2 days per week. However, nearly two-thirds (n=164, 61.9%) reported using social media at least 3 days weekly. Facebook (n=179, 67.5%) and YouTube (n=163, 61.5%) were the most commonly visited platforms for obtaining clinical information, followed by Instagram (n=115, 43.4%). Twitter

XSL•FO
RenderX

(n=4, 1.5%) was the least popular platform. Social media was mainly used for oral health education and promotion (n=191, 72.1%); communication with patients, friends, and family (n=165, 62.3%); and marketing and advertising (n=150, 56.6%).

Table . Frequency, preferred platforms, and purpose of social media use.

Variables	Frequency (n=265), n (%)
Social media use	
Never	2 (0.8)
Seldom (average of 1 - 2 d/wk)	99 (37.4)
Occasionally (average of 3 - 5 d/wk)	96 (36.2)
Frequently (average of 6 - 7 d/wk)	68 (25.7)
Social media website used[a]	
Twitter	4 (1.5)
Facebook	179 (67.5)
Instagram	115 (43.4)
YouTube	163 (61.5)
TikTok	38 (14.3)
Viber	23 (8.7)
Others	75 (28.3)
Purpose of social media use[a]	
Oral health promotion and education	191 (72.1)
Connect and communicate with patients, friends, and family	165 (62.3)
Exchange opinions and views regarding cases with colleagues	113 (42.6)
Marketing and advertising	150 (56.6)

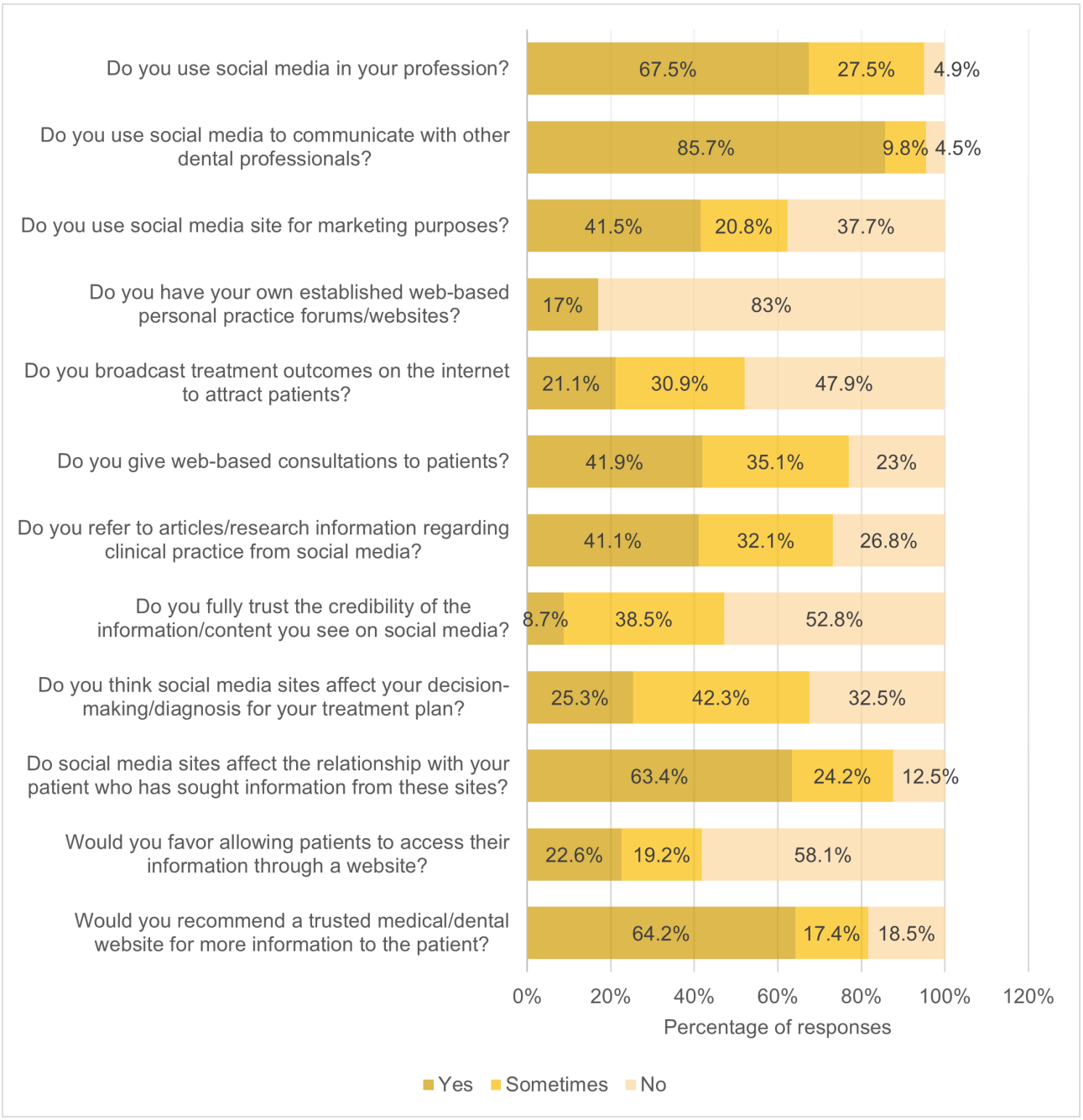
[a]Given that study participants could indicate multiple responses, the sum of the percentages is not 100.

The responses of dental practitioners to questions regarding social media use are presented in Figure 1. Of the 265 participants, 67.5% (n=179) indicated that they use social media in their dental practice. Despite this, 83% (n=220) of dentists did not have their own web-based practice forums or websites. However, the majority (n=227, 85.7%) reported using social media to communicate with other dental professionals, and 41.5% (n=110) used it for marketing purposes. Although 41.1% (n=109) of dentists referred to articles or research information from social media for clinical practice, nearly half (n=127, 47.9%) were reluctant to broadcast treatment outcomes on the internet to attract patients, and 41.9% (n=111) were open to providing web-based consultations. Only 8.7% (n=23) of dentists trusted the credibility of information on social media, while 63.4% (n=168) believed that social media can affect the patient-dentist relationship when patients seek information on the internet. Additionally, most participants (n=154, 58.1%) did not allow patients to access their information through a website, but 64.2% (n=170) would recommend a trusted website to patients.

Table 3 explores the relationship between social media use and its impact on dental practice. Participants who used social media in their practice believed that it enhanced the quality of care provided to patients (*P*=.001) and significantly reported that it improved their dental practice (*P*=.02).

XSL•FO
RenderX

Figure 1. Percentage of dental practitioners' responses to questions on social media use (n=265).



XSL•FO
RenderX

Table . Use of social media and respondents' perceived impact on enhancing patient care quality and dental practice.

Variable	Use of social media, n (%)[a]			Total (n=265), n (%)	*P* value[b]
	Yes	Sometimes	No		
Social media improves the quality of care delivered to patients					.001
Yes	126 (74.1)	40 (23.5)	4 (2.4)	170 (64.2)	
Maybe	42 (56)	24 (32)	9 (12)	75 (28.3)	
No	11 (55)	9 (45)	0 (0)	20 (7.5)	
Social media improves dental practice					.02
Yes	106 (74.7)	30 (21.1)	6 (4.2)	142 (53.6)	
Maybe	48 (57)	29 (35)	7 (8)	84 (31.7)	
No	25 (64)	14 (36)	0 (0)	39 (14.7)	

[a]Percentages were calculated based on the number of participants corresponding to each variable.

[b]Chi-square test was used with an α level of .05.

Discussion

Principal Findings

This study aimed to investigate the use of social media and assess its impact on enhancing dental care and practice among dental professionals. The findings revealed that over half of the participants aged 20 - 30 years are actively using social media in their dental profession, aligning with other studies that indicate younger dentists or health care professionals are more likely to use social media compared to their older counterparts [36,38,40-42]. Variables such as sex, dental specialty, work sector, and years of clinical practice did not significantly influence social media use.

Most respondents reported using social media 1 - 2 days per week to engage with the public and other dental professionals. This is in contrast to other studies where social media was used at least once daily for obtaining information [32,42]. Among social media users in the Philippines, Facebook and YouTube were identified as the primary platforms for accessing clinical information, which is consistent with The Digital Report 2024, highlighting that Facebook has 86.75 million active users in the Philippines [43]. These findings are similar to those reported among dentists in the United Kingdom and the United States, where Facebook is the most commonly used platform in dental practice [1,31]. This was also true as among dental students [7]. In comparison, dentists in Saudi Arabia predominantly use Twitter [22,25] and Instagram [32], while WhatsApp is the preferred platform among Indian dentists [37]. Similarly, South African dentists primarily use Google Plus and Facebook for personal rather than professional purposes [42].

Previous studies have shown that marketing is a common reason for using social media [1,25,31,32]. In our study, 72.1% (191/265) of participants indicated that their primary reason for using social media was for oral health promotion and education. This finding is consistent with earlier research, which found that a high proportion of respondents use social media for dental education and prefer to share clinical work [21]. Other studies also support the notion that education and sourcing oral health information are primary purposes for social media use among dental professionals [22-25]. Social media interventions have positively impacted various aspects of oral health, particularly among adolescents, where platforms like YouTube, WhatsApp, Instagram, and Telegram have been effectively used to promote oral health [27]. For instance, a study demonstrated that oral health education via Telegram significantly improved oral health outcomes among secondary school students [44] and adolescents [45] in Iran. Additionally, Instagram has been shown to support orthodontists in motivating young patients to maintain proper oral hygiene [46], and WhatsApp has been integrated successfully into oral hygiene protocols to enhance patient compliance and oral health during orthodontic treatment [47]. Presenting audiovisual information through YouTube has also significantly improved patient knowledge [48].

Another key purpose of social media use is communication and the exchange of opinions with colleagues. According to the survey, 62.3% (165/265) of participants used social media to connect with patients, friends, and family, while 85.7% (227/265) used it to communicate with other dental professionals. This is supported by studies showing that social media communication, such as Instagram, can reduce patient anxiety prior to dental procedures [28], enhance dentists' communication skills with other professionals [38], and improve patient interactions [30]. Another study found that a significant proportion of the younger population engages with social media to discuss dentistry. This highlights the need for careful management to ensure the dissemination of accurate dental health information [25].

The respondents of the survey also indicated that social media had the potential to enhance dental practice and the quality of care provided to patients. This finding is in agreement with study that shows that dentists who use social media in their practice not only recommend it to their peers but also believe it enhances their practice [22]. Social media can improve various aspects of dental practice, including service provision, advertising, counseling, and oral health education, and it can also be a tool for professional development [49]. Dental organizations and educators can use social media to disseminate information and updates [36,50]. Furthermore, multidimensional health care approaches, including social media, have been

XSL•FO
RenderX

proven highly successful in improving patient care, increasing public knowledge, facilitating research, connecting health care providers, improving medical education, and addressing public health crises [29].

This study has several limitations that should be considered when interpreting the results. First, the study population was limited to dental practitioners in Baguio City, Philippines, which may restrict the generalizability of the findings to the broader dentist population in the country. The uneven distribution of survey participation across various demographic variables, such as age, sex, dental specialty, working sector, and years of clinical practice, may further impact the ability to extrapolate the results to the entire dental community. Additionally, the use of an electronic survey as the sole data collection method presented inherent challenges. A significant portion of the target population may not respond to or use electronic surveys, potentially introducing nonresponse bias and affecting the representativeness of the sample. This bias could be particularly relevant in the context of social media use, as those who are less inclined to participate in electronic surveys may also be less likely to engage with social media platforms. Furthermore, the cross-sectional nature of the study limits the ability to establish causal relationships between social media use and its impact on dental practice and patient care. Longitudinal studies would be necessary to better understand the long-term effects of social media integration in dental settings. Despite these limitations, this study provides valuable insights into the current use of social media among dental practitioners in the Philippines and its perceived impact on dental care and practice. Future research should aim to include larger, more representative samples across different regions and settings to enhance the generalizability of the findings. Additionally, mixed-methods approaches combining quantitative and qualitative data collection techniques could provide a more comprehensive understanding of the topic.

Conclusions

Social media is widely used among younger dental practitioners, primarily for education, communication, and marketing, with a notable impact on enhancing practice quality and patient care. Social media platforms, especially Facebook and YouTube, were commonly used by practitioners to promote oral health, engage with patients and colleagues, and support marketing efforts.

Social media offers a convenient space for professional growth, patient education, and community interaction, which many dentists perceive as beneficial to their practice. However, a low trust in social media information and concerns about its influence on patient-dentist relationships indicate the need for clear guidelines and quality web-based resources. Enhanced support in these areas can help dental professionals maximize the positive impacts of social media on dental care and practice.

Future studies should compare social media with traditional marketing methods in the dental practice such as print ads, community outreach, and patient referrals. These traditional methods remain valuable for building trust, local credibility, and reaching offline patients. Comparing these strategies can help find the best way to combine digital and traditional approaches for better patient engagement and practice success.

Acknowledgments

The authors would like to thank all participants who participated in this study. We also appreciated the funding support from the Faculty of Dentistry at Chulalongkorn University.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Parmar N, Dong L, Eisingerich AB. Connecting with your dentist on Facebook: patients' and dentists' attitudes towards social media usage in dentistry. J Med Internet Res 2018 Jun 29;20(6):e10109. [doi: 10.2196/10109] [Medline: 29959108]
2. Jung YS, Yang HY, Choi YH, et al. Factors affecting use of word-of-mouth by dental patients. Int Dent J 2018 Oct;68(5):314-319. [doi: 10.1111/idj.12387] [Medline: 29572927]
3. Mary AV, Kesavan R, V V, Pi J, G D, S S. Impact of social media on dental treatment needs: a cross sectional pilot study. Int J Appl Dent Sci 2023 Jan 1;9(4):313-317. [doi: 10.22271/oral.2023.v9.i4e.1887]
4. Afful-Dadzie E, Afful-Dadzie A, Egala SB. Social media in health communication: a literature review of information quality. Hlth Inf Manag 2023 Jan;52(1):3-17. [doi: 10.1177/1833358321992683] [Medline: 33818176]
5. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of social media. Bus Horiz 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
6. Social media. Merriam-Webster Dictionary. URL: https://www.merriam-webster.com/dictionary/social%20media [accessed 2024-12-17]

XSL•FO
RenderX

7. Kenny P, Johnson IG. Social media use, attitudes, behaviours and perceptions of online professionalism amongst dental students. Br Dent J 2016 Nov 18;221(10):651-655. [doi: 10.1038/sj.bdj.2016.864] [Medline: 27857111]
8. Sivaramakrishnan G, AbdulAmeer F, Faisal F, et al. Use of social media to view and post dentistry-related information in Bahrain: a cross-sectional study. Healthc Inform Res 2023 Jan;29(1):31-39. [doi: 10.4258/hir.2023.29.1.31] [Medline: 36792099]
9. Freire Y, Sánchez MG, Suárez A, Joves G, Nowak M, Díaz-Flores García V. Influence of the use of social media on patients changing dental practice: a web-based questionnaire study. BMC Oral Health 2023 Jun 6;23(1):365. [doi: 10.1186/s12903-023-03078-9] [Medline: 37277765]
10. MacDonald J, Sohn S, Ellis P. Privacy, professionalism and Facebook: a dilemma for young doctors. Med Educ 2010 Aug;44(8):805-813. [doi: 10.1111/j.1365-2923.2010.03720.x] [Medline: 20633220]
11. Malik A, Heyman-Schrum C, Johri A. Use of Twitter across educational settings: a review of the literature. Int J Educ Technol High Educ 2019 Dec;16(1):1-22. [doi: 10.1186/s41239-019-0166-x]
12. Forgie SE, Duff JP, Ross S. Twelve tips for using Twitter as a learning tool in medical education. Med Teach 2013;35(1):8-14. [doi: 10.3109/0142159X.2012.746448] [Medline: 23259608]
13. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: a systematic review. Health Informatics J 2015 Sep;21(3):173-194. [doi: 10.1177/1460458213512220] [Medline: 24670899]
14. Simsek H, Buyuk SK, Cetinkaya E. YouTube as a source of information on oral habits. J Indian Soc Pedod Prev Dent 2020;38(2):115-118. [doi: 10.4103/JISPPD.JISPPD_357_19] [Medline: 32611855]
15. Guo J, Yan X, Li S, Van der Walt J, Guan G, Mei L. Quantitative and qualitative analyses of orthodontic-related videos on YouTube. Angle Orthod 2020 May 1;90(3):411-418. [doi: 10.2319/082019-542.1] [Medline: 33378439]
16. Ooi HL, Kelleher MGD. Instagram dentistry. Prim Dent J 2021 Mar;10(1):13-19. [doi: 10.1177/2050168420980964] [Medline: 33722130]
17. Patil P, Pawar AM, Agustin Wahjuningrum D, Maniangat Luke A, Reda R, Testarelli L. Connectivity and integration of Instagram use in the lives of dental students and professionals: a country-wide cross-sectional study using the InstaAA Questionnaire. J Multidiscip Healthc 2023;16:2963-2976. [doi: 10.2147/JMDH.S430973] [Medline: 37818291]
18. Buyuk SK, Alpaydın MT, Imamoglu T. Social media and orthodontics: an analysis of orthodontic-related posts on Instagram. Aust Orthod J 2020 Jan 1;36(2):153-159. [doi: 10.21307/aoj-2020-017]
19. Fraticelli L, Smentek C, Tardivo D, et al. Characterizing the content related to oral health education on TikTok. Int J Environ Res Public Health 2021 Dec 16;18(24):13260. [doi: 10.3390/ijerph182413260] [Medline: 34948869]
20. Kılınç DD. Is the information about orthodontics on Youtube and TikTok reliable for the oral health of the public? A cross sectional comparative study. J Stomatol Oral Maxillofac Surg 2022 Oct;123(5):e349-e354. [doi: 10.1016/j.jormas.2022.04.009] [Medline: 35470122]
21. Kurian N, Varghese IA, Cherian JM, et al. Influence of social media platforms in dental education and clinical practice: a cross-sectional survey among dental trainees and professionals. J Dent Educ 2022 Aug;86(8):1036-1042. [doi: 10.1002/jdd.12914] [Medline: 35257384]
22. Hamasha AAH, Alghofaili N, Obaid A, et al. Social media utilization among dental practitioner in Riyadh, Saudi Arabia. Open Dent J 2019 Feb 28;13(1):101-106. [doi: 10.2174/1874210601913010101]
23. Aboalshamat K, Alkiyadi S, Alsaleh S, et al. Attitudes toward social media among practicing dentists and dental students in clinical years in Saudi Arabia. Open Dent J 2019 Mar 28;13(1):143-149. [doi: 10.2174/1874210601913010143]
24. Maharani DA, El Tantawi M, Yoseph MG, Rahardjo A. The use of internet platforms for oral health information and associated factors among adolescents from Jakarta: a cross sectional study. BMC Oral Health 2021 Jan 7;21(1):22. [doi: 10.1186/s12903-020-01387-x] [Medline: 33413315]
25. Althunayan A, Alsalhi R, Elmoazen R. Role of social media in dental health promotion and behavior change in Qassim province, Saudi Arabia. Int J Med Health Res 2018;4(2):98-103 [FREE Full text]
26. El Tantawi M, Bakhurji E, Al-Ansari A, AlSubaie A, Al Subaie HA, AlAli A. Indicators of adolescents' preference to receive oral health information using social media. Acta Odontol Scand 2019 Apr;77(3):213-218. [doi: 10.1080/00016357.2018.1536803] [Medline: 30632864]
27. Farrokhi F, Ghorbani Z, Farrokhi F, Namdari M, Salavatian S. Social media as a tool for oral health promotion: a systematic review. PLoS ONE 2023;18(12):e0296102. [doi: 10.1371/journal.pone.0296102] [Medline: 38113216]
28. Sivrikaya EC, Yilmaz O, Sivrikaya P. Dentist-patient communication on dental anxiety using the social media: a randomized controlled trial. Scand J Psychol 2021 Dec;62(6):780-786. [doi: 10.1111/sjop.12769] [Medline: 34333783]
29. Farsi D. Social media and health care, part I: literature review of social media use by health care providers. J Med Internet Res 2021 Apr 5;23(4):e23205. [doi: 10.2196/23205] [Medline: 33664014]
30. Ghandhi D, Bodani N, Lal A, et al. Evaluation of social media usage by dental practitioners of Pakistan for professional purposes - a cross-sectional study. Clin Cosmet Investig Dent 2022;14:245-252. [doi: 10.2147/CCIDE.S374111] [Medline: 35975279]
31. Henry RK, Molnar A, Henry JC. A survey of US dental practices' use of social media. J Contemp Dent Pract 2012 Mar 1;13(2):137-141. [doi: 10.5005/jp-journals-10024-1109] [Medline: 22665737]

XSL•FO
RenderX

32. Bahabri RH, Zaidan AB. The impact of social media on dental practice promotion and professionalism amongst general dental practitioners and specialists in KSA. J Taibah Univ Med Sci 2021 Jun;16(3):456-460. [doi: 10.1016/j.jtumed.2020.12.017] [Medline: 34140874]
33. Number of internet and social media users worldwide as of July 2024 (in billions). Statista. URL: https://www.statista.com/statistics/617136/digital-population-worldwide/ [accessed 2024-12-17]
34. Most popular social networks worldwide as of April 2024, by number of monthly active users (in millions). Statista. URL: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ [accessed 2024-12-17]
35. Social media in the Philippines - statistics & facts. Statista. URL: https://www.statista.com/topics/6759/social-media-usage-in-the-philippines/#topicOverview [accessed 2024-12-17]
36. Al-Khalifa KS, Al-Swuailem AS, AlSheikh R, et al. The use of social media for professional purposes among dentists in Saudi Arabia. BMC Oral Health 2021 Jan 12;21(1):26. [doi: 10.1186/s12903-021-01390-w] [Medline: 33435911]
37. Jamkhande A, Pangarikar AB, Khairnar MR, Gavali N, Patil Y, Wadgave U. Social media use in dental practice by dental practitioners of India: a cross-sectional survey. J Clin Diagn Res 2022;16(6). [doi: 10.7860/JCDR/2022/56404.16533]
38. Makkar N, Lawrence D, Veeresh D, Banik A, Kumar A, Mehar AS. Assessment of social media usage and its influence among dental faculty members in Davangere city – a cross sectional survey. J Indian Assoc Public Health Dent 2018;16(3):246. [doi: 10.4103/jiaphd.jiaphd_143_17]
39. Kamarudin Y, Mohd Nor NA, Libamin AC, et al. Social media use, professional behaviors online, and perceptions toward e-professionalism among dental students. J Dent Educ 2022 Aug;86(8):958-967. [doi: 10.1002/jdd.12912] [Medline: 35247218]
40. Almaiman S, Bahkali S, Al Farhan A, Bamuhair S, Househ M, Alsurimi K. The prevalence of using social media among healthcare professionals in Saudi Arabia: a pilot study. Stud Health Technol Inform 2015;213:263-266. [Medline: 26153010]
41. Alsobayel H. Use of social media for professional development by health care professionals: a cross-sectional web-based survey. JMIR Med Educ 2016 Sep 12;2(2):e15. [doi: 10.2196/mededu.6232] [Medline: 27731855]
42. Snyman L, Visser JH. The adoption of social media and social media marketing by dentists in South Africa. SADJ 2014 Jul;69(6):258. [Medline: 26548200]
43. Kemp S. Digital 2024: the Philippines. Datareportal. 2024 Feb 21. URL: https://datareportal.com/reports/digital-2024-philippines [accessed 2024-12-17]
44. Zareban I, Rahmani A, Allahqoli L, et al. Effect of education based on trans-theoretical model in social media on students with gingivitis; a randomized controlled trial. Health Educ Health Prom 2022;10(1):1-7 [FREE Full text]
45. Scheerman JFM, Hamilton K, Sharif MO, Lindmark U, Pakpour AH. A theory-based intervention delivered by an online social media platform to promote oral health among Iranian adolescents: a cluster randomized controlled trial. Psychol Health 2020 Apr;35(4):449-466. [doi: 10.1080/08870446.2019.1673895] [Medline: 31621423]
46. Scribante A, Gallo S, Bertino K, Meles S, Gandini P, Sfondrini MF. The effect of chairside verbal instructions matched with Instagram social media on oral hygiene of young orthodontic patients: a randomized clinical trial. Appl Sci (Basel) 2021;11(2):706. [doi: 10.3390/app11020706]
47. Zotti F, Dalessandri D, Salgarello S, et al. Usefulness of an app in improving oral hygiene compliance in adolescent orthodontic patients. Angle Orthod 2016 Jan;86(1):101-107. [doi: 10.2319/010915-19.1] [Medline: 25799001]
48. Al-Silwadi FM, Gill DS, Petrie A, Cunningham SJ. Effect of social media in improving knowledge among patients having fixed appliance orthodontic treatment: a single-center randomized controlled trial. Am J Orthod Dentofacial Orthop 2015 Aug;148(2):231-237. [doi: 10.1016/j.ajodo.2015.03.029] [Medline: 26232831]
49. Al-Nahdi S, Alanzi T, Alhodaib H. Use of social media by healthcare providers in saudi arabia: reasons and usefulness for their professional development. Presented at: 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES); Oct 16-19, 2020; Xuzhou, China. [doi: 10.1109/DCABES50732.2020.00091]
50. Terrasse M, Gorin M, Sisti D. Social media, e-health, and medical ethics. Hast Cent Rep 2019 Jan;49(1):24-33. [doi: 10.1002/hast.975] [Medline: 30790306]

Abbreviations

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

XSL•FO
RenderX

Edited by A Mavragani; submitted 05.09.24; peer-reviewed by J Hunsrisakhun, K Sadasiva; revised version received 19.11.24; accepted 20.11.24; published 03.01.25.

Please cite as:
Acosta JM, Detsomboonrat P, Pisarnturakit PP, Urwannachotima N
The Use of Social Media on Enhancing Dental Care and Practice Among Dental Professionals: Cross-Sectional Survey Study
JMIR Form Res 2025;9:e66121
URL: https://formative.jmir.org/2025/1/e66121
doi:10.2196/66121

© Joseph Macadaeg Acosta, Palinee Detsomboonrat, Pagaporn Pantuwadee Pisarnturakit, Nipaporn Urwannachotima. Originally published in JMIR Formative Research (https://formative.jmir.org), 3.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Artificial Intelligence–Powered Training Database for Clinical Thinking: App Development Study

Heng Wang[1*], MEd; Danni Zheng[1*], PhD; Mengying Wang[2], MEng; Hong Ji[2], MEng; Jiangli Han[3], MD; Yan Wang[4], MD; Ning Shen[5], MD; Jie Qiao[6], MD

[1]Education Department, Peking University Third Hospital, Beijing, China

[2]Information Management and Big Data Center, Peking University Third Hospital, Beijing, China

[3]Department of Cardiology, Peking University Third Hospital, Beijing, China

[4]Department of Obstetrics and Gynecology, Peking University Third Hospital, Beijing, China

[5]Department of Pulmonary and Critical Care Medicine, Peking University Third Hospital, Beijing, China

[6]Centre for Reproductive Medicine, Department of Obstetrics and Gynecology, Peking University Third Hospital, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Jie Qiao, MD
Centre for Reproductive Medicine, Department of Obstetrics and Gynecology, Peking University Third Hospital, Beijing, China

Abstract

Background: With the development of artificial intelligence (AI), medicine has entered the era of *intelligent medicine*, and various aspects, such as medical education and talent cultivation, are also being redefined. The cultivation of clinical thinking abilities poses a formidable challenge even for seasoned clinical educators, as offline training modalities often fall short in bridging the divide between current practice and the desired ideal. Consequently, there arises an imperative need for the expeditious development of a web-based database, tailored to empower physicians in their quest to learn and hone their clinical reasoning skills.

Objective: This study aimed to introduce an app named "XueYiKu," which includes consultations, physical examinations, auxiliary examinations, and diagnosis, incorporating AI and actual complete hospital medical records to build an online-learning platform using human-computer interaction.

Methods: The "XueYiKu" app was designed as a contactless, self-service, trial-and-error system application based on actual complete hospital medical records and natural language processing technology to comprehensively assess the "clinical competence" of residents at different stages. Case extraction was performed at a hospital's case data center, and the best-matching cases were differentiated through natural language processing, word segmentation, synonym conversion, and sorting. More than 400 teaching cases covering 65 kinds of diseases were released for students to learn, and the subjects covered internal medicine, surgery, gynecology and obstetrics, and pediatrics. The difficulty of learning cases was divided into four levels in ascending order. Moreover, the learning and teaching effects were evaluated using 6 dimensions covering systematicness, agility, logic, knowledge expansion, multidimensional evaluation indicators, and preciseness.

Results: From the app's first launch on the Android platform in May 2019 to the last version updated in May 2023, the total number of teacher and student users was 6209 and 1180, respectively. The top 3 subjects most frequently learned were respirology (n=606, 24.1%), general surgery (n=506, 20.1%), and urinary surgery (n=390, 15.5%). For diseases, pneumonia was the most frequently learned, followed by cholecystolithiasis (n=216, 14.1%), benign prostate hyperplasia (n=196, 12.8%), and bladder tumor (n=193, 12.6%). Among 479 students, roughly a third (n=168, 35.1%) scored in the 60 to 80 range, and half of them scored over 80 points (n=238, 49.7%). The app enabled medical students' learning to become more active and self-motivated, with a variety of formats, and provided real-time feedback through assessments on the platform. The learning effect was satisfactory overall and provided important precedence for establishing scientific models and methods for assessing clinical thinking skills in the future.

Conclusions: The integration of AI and medical education will undoubtedly assist in the restructuring of education processes; promote the evolution of the education ecosystem; and provide new convenient ways for independent learning, interactive communication, and educational resource sharing.

(JMIR Form Res 2025;9:e58426) doi:10.2196/58426

KEYWORDS

artificial intelligence; clinical thinking ability; virtual medical records; distance education; medical education; online learning

XSL•FO
RenderX

Introduction

Clinical thinking refers to the ability of doctors to understand, diagnose, and treat diseases through analysis, synthesis, judgment, reasoning, and other thinking activities by using their medical knowledge, skills, and experience [1]. Clinical reasoning skills include a series of strategies such as correlating information, conducting a comprehensive analysis to form one or more diagnostic hypotheses, weighing the risks and benefits of diagnostic tests and treatments, as well as formulating a reasonable diagnosis and treatment plan [2]. This detailed process includes collecting and evaluating clinical information, selecting diagnostic tests, evaluating test results, developing diagnostic hypotheses, and weighing treatment options and risks [3].

Teaching clinical thinking ability is a challenging task even for experienced clinical teachers. As clinical knowledge is the basis of clinical thinking, the cultivation of clinical thinking needs to go through the links of guidance, deliberate practice, and feedback, and it takes a long time to cultivate efficient clinical thinking ability. The level of clinical thinking of residents is directly related to the safety and quality of medical treatment [4]. Therefore, we should constantly improve the understanding of the clinical thinking training and improve the training quality of residents.

Therefore, the cultivation of clinical thinking ability is a necessary condition for the transition from medical students to clinicians and is currently the key point and difficulty among resident training programs. Among the training processes, the most effective method is based on real or virtual case training, combined with feedback and reflection to continuously improve. Previous studies have shown that clinical thinking ability is scarce among physicians, while offline training cannot meet the gap between the present reality and the ideal requirements [5]. Therefore, there is an urgent need for the development of a web-based database to help physicians learn and exercise clinical thinking ability.

Artificial intelligence (AI) is a novel technology that can be utilized to research and develop the theories, methods, technologies, and applications used to simulate, extend, and expand human intelligence. AI belongs to the interdisciplinary field of natural science, social science, and technological science [6]. AI has a history of more than 70 years, and its application in the medical field mainly focuses on five aspects: medical imaging, auxiliary diagnosis, drug research and development, health management, and disease prediction.

There have been several reports and studies on the use of AI in medical teaching and training of clinical thinking ability. For example, Diagnostic Reasoning (DxR) Clinician (DxR Development Group, Inc.) was developed in 1994 by Myers and Dorsey in the United States to train clinical thinking among students [7]. In 2006, the BP network neural algorithm in AI was applied to teach quality monitoring instead of manually teaching daily monitoring [8]. In 2011, Knewton, an adaptive education platform, offered personalized learning services by extracting learning data from students. In 2017, Tsinghua University launched a study system combined with AI and full quantitative reality virtual technology. In this system, patients' image data can be converted into a holographic three-dimensional anatomical structure of the human body and mapped in the virtual space [9]. In 2020, a competency-based learning and assessment system (CBLAS) for residency training was designed to provide resident physicians with clinical assessments and learning in order to enhance the learning of trainees and reduce the burden of assessments [10]. However, in the era of AI, there are still issues that exist in medical teaching, such as the lack of general education, the lack of classroom interaction, and the lag in curriculum construction [11].

On this basis, the present study introduces an app named "XueYiKu," which includes consultations, physical examinations, auxiliary examinations, and diagnosis, by incorporating AI and actual complete hospital medical records to build an online-learning platform using human-computer interaction. This application could cultivate students' ability to solve clinical problems as well as improve teaching quality and efficiency, thereby achieving the sharing of high-quality teaching resources.

Methods

Design and Study Population

The "XueYiKu" app was designed as a contactless, self-service, trial-and-error system application based on actual complete hospital medical records and natural language processing technology to build an online-learning platform using human-computer interaction. Training participants included medical undergraduates, medical graduate students, residents, chief residents, and attending doctors. The purpose of the app was to improve clinical thinking ability through independent study accompanied by real-time feedback. The XueYiKu app was launched on the Android platform on May 14, 2019, and on the IOS platform on August 3, 2019. After 27 updates, the last version of the XueYiKu app was updated on January 12, 2022.

Data Collection

The general database of the XueYiKu app was based on a collection of real medical records and data from general teaching, including a standard clinical question database, virtual medical record database, and student learning record database. Using information extraction technology in natural language processing from real medical records, the XueYiKu app could poststructure the content from real electronic medical records, generate questions, and match answers online through machine learning. In the teaching cases, the automatically extracted answers matched 90% of the original medical records. The structured cases in the electronic medical record system of the hospital were then screened by a teaching team from Peking University Third Hospital, and appropriate medical records were selected and reviewed by the teaching team. Once the teaching case was determined, relevant inspection results were also extracted, including inspections, auxiliary inspections, and ultrasound, imaging, and pathology findings.

XSL•FO
RenderX

Outcome Measures

Clinical thinking ability involves extremely high requirements for the objectivity and standardization of evaluation. Hence, a multidimensional automatic evaluation system is required. With reference to the residents' standardized training report, the whole process of student learning was recorded systematically, and clinical thinking was evaluated in multiple dimensions according to the learning situation of the student, including preciseness, logic, systematicness, agility, knowledge expansion, and multidimensional and comprehensive application (Table 1). Based on the scores of students in various dimensions of learning, a radar chart of clinical thinking ability is drawn to assess the compliance of medical students' clinical thinking ability.

Table . Seven elements for clinical thinking evaluation in the XueYiKu app.

Elements	Contents
Preciseness	Comprehensive evaluation of diagnosis and differential diagnosis
Logic	The logical sequence evaluation of inquiry, physical examination, auxiliary examination, and diagnosis
Systematicness	Comprehensive evaluation of the completion rate of inquiry, physical examination, auxiliary examination, diagnosis, and differential diagnosis
Agility	Comprehensive evaluation of the time of inquiry, physical examination, auxiliary examination, diagnosis, and differential diagnosis
Knowledge expansion	Accuracy and logical evaluation of diagnosis and differential diagnosis
Multidimensional evaluation indicators	Comprehensive evaluation of diagnosis and treatment programs
Comprehensive application	Comprehensive application ability of basic and clinical knowledge (for evaluation of level 4 individually)

Evaluation and Feedback

We recruited all students in Peking University Hospital to evaluate the teaching effect of the XueYiKu app, including medical undergraduates, medical graduate students, residents, chief residents, and attending doctors. Both the questionnaire method and interview method were utilized. Web-based surveys, developed by the study team, were conducted upon completion of the case study so that students could fill in and evaluate their learning effect. Through interviews with relevant teachers and experts, the common clinical path extracted using AI technology was determined. Meanwhile, the range of clinical thinking ability evaluation was confirmed according to the requirements of different training stages. The questionnaire method was used to investigate the following two factors: (1) the current status of clinical thinking training regarding the ability of medical students and residents to clarify the problems in clinical thinking training and improve the content of the database; (2) the experience of medical students and residents regarding the use of the case database and self-evaluation of learning effects to complete the effect evaluation of the XueYiKu app.

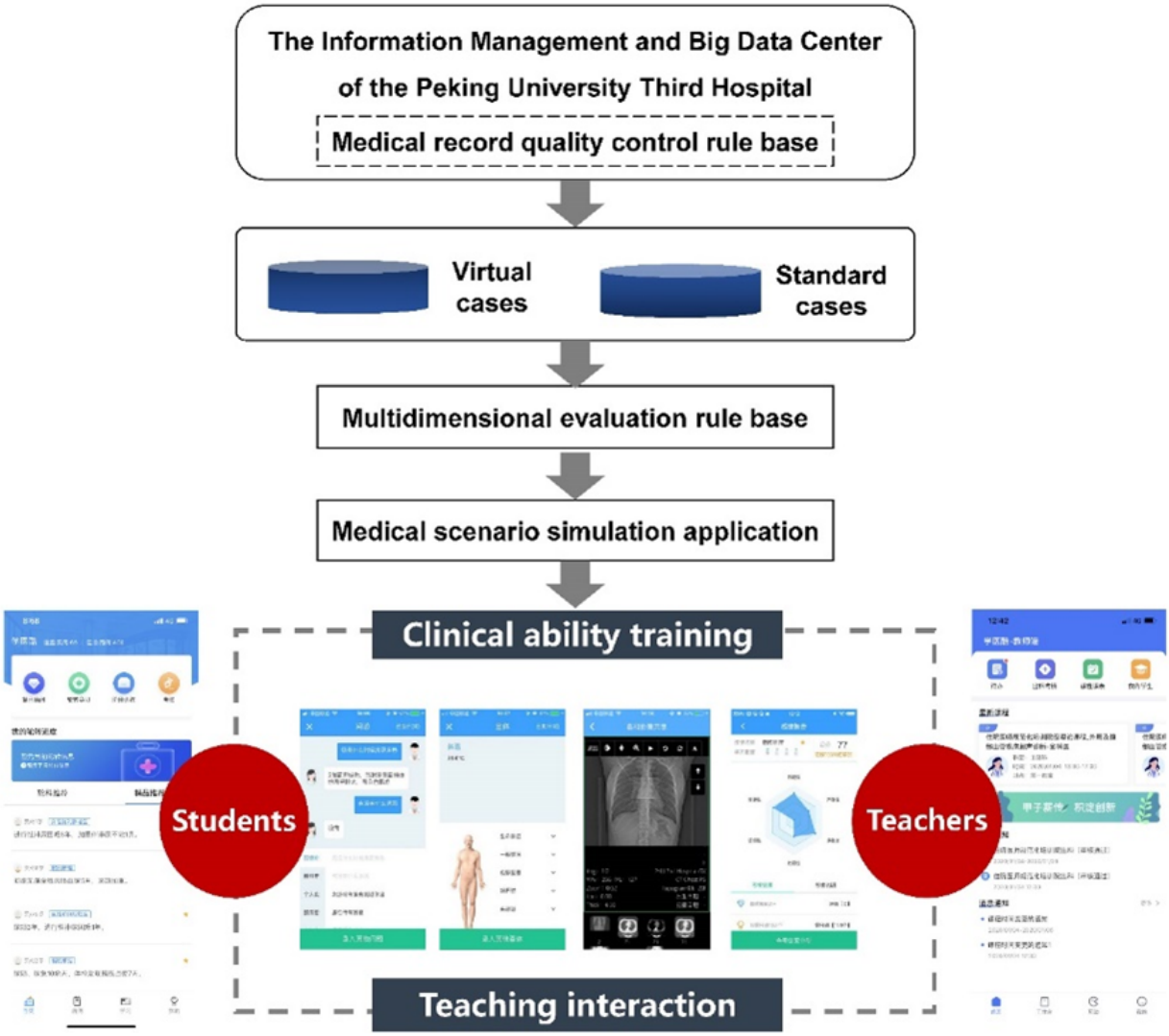
Data Processing

Details of the virtual case system and AI teaching development based on actual complete hospital medical records and natural language processing technology have been described previously [12]. A flow chart of the program design from data collection to interaction between students and teachers is shown in Figure 1. Case extraction was performed at a hospital's case data center, and the best-matching cases were extracted through natural language processing, word segmentation, synonym conversion, and sorting. A standard clinical questioning data module, virtual case data module, and student learning difficulty module were established to achieve simulation. Students can view the objective examination and inspection data of actual cases, including details of the consultation and physical examination, and automatically provide their learning response via a multidimensional evaluation system. In order to assess the changes in students' clinical thinking after using the XueYiKu app, 15 medical graduate students were subjected to two simulation tests before and after learning through the virtual case system. The tests, which included the full-process case examination of cases having the same difficulty level, examined core clinical thinking test points such as consultation, physical examination, and disposal, and generated multidimensional evaluation indicators (rigor, logic, system, agility, and knowledge expansion). Thus, a complete and credible evaluation system was developed.

For statistical analysis, standard descriptive statistics were used to describe the study population: continuous variables were presented as the mean and SD, and categorical variables were presented as percentages.

XSL•FO
RenderX

Figure 1. Flow chart of the program design from data collection to interaction between students and teachers in the XueYiKu app.



Ethical Considerations

This study was reviewed and approved by the Peking University Third Hospital Medical Science Research Ethics Committee (No. IRB00006761-M2022063). Invitations were dispatched to prospective participants via the XueYiKu app, accompanied by a comprehensive informed consent form outlining the nature of the investigation. Participants were clearly notified that the interviews would be audio-recorded solely for analytical purposes and would remain exclusively accessible to the research team. Furthermore, they were assured of their right to withdraw their consent at any juncture during the interview process. All information gathered about the participants was kept private and confidential. All study data were deidentified.

Results

Overview

After concealment processing, more than 400 teaching cases covering 65 kinds of diseases were released for students to learn from, and the subjects covered internal medicine, surgery, gynecology and obstetrics, and pediatrics (Table 2). The difficulty in learning cases was divided into 4 levels in ascending order: level 1 provided all score items for students to choose from; except for items from level 1, level 2 provided additional interference items; compared with level 2, level 3 displayed some score items and all interference items to enhance difficulty; for level 4, there was neither a score item nor an interference item, and the system would match the results from the students' voice or manual input of the questions.

XSL•FO
RenderX

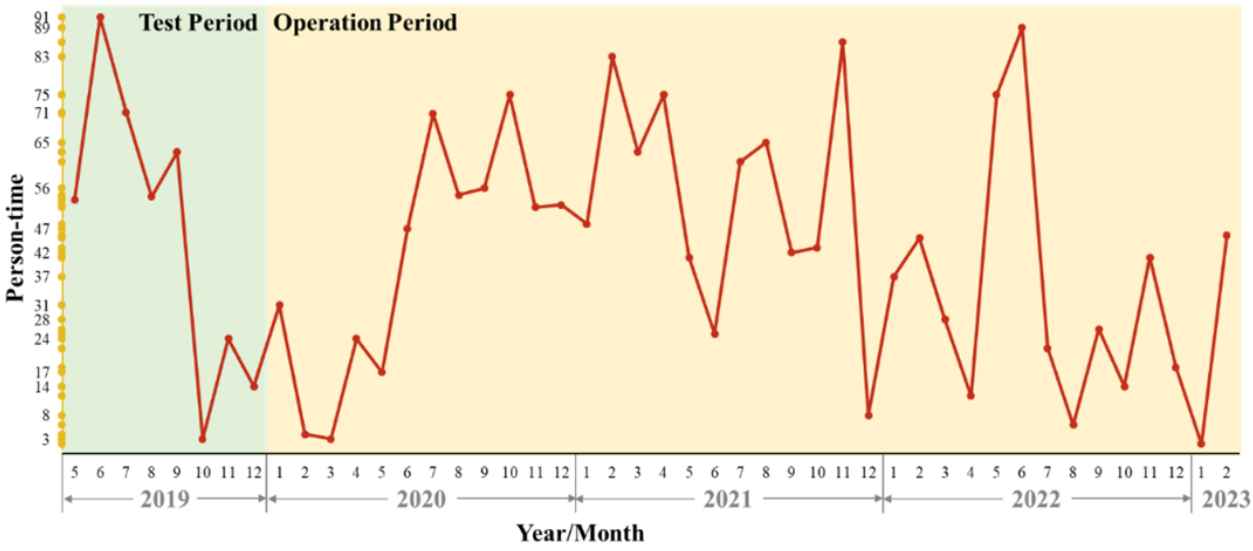
Table . Categories of subjects and diseases for learning and evaluation in the database of the XueYiKu app.

Subjects	Category	Diseases
Internal Medicine	20	Coronary heart disease, heart failure, arrhythmia, hypertension, pneumonia, chronic airway inflammatory disease, pleural effusion, pulmonary embolism, upper gastrointestinal bleeding, pancreatitis, inflammatory bowel disease, ascites, nephrotic syndrome, acute kidney injury, diabetes, hyperthyroidism, leukemia, anemia, systemic lupus erythematosus, rheumatoid arthritis
Surgery	20	Inguinal hernia, appendicitis, gallbladder stones, bile duct stones, stomach cancer, colon cancer, rectal cancer, liver cancer, pancreatic cancer, lung cancer, pneumothorax, pituitary tumor, coronary heart disease, benign prostatic hyperplasia, bladder tumor, acute respiratory distress syndrome, cervical myelopathy disease, lumbar disc herniation, knee osteoarthritis, ankle fracture
Gynecology and obstetrics	15	Ectopic pregnancy, uterine fibroids, ovarian tumors, dysfunctional uterine bleeding, endometrial cancer, cervical cancer, ovarian cancer, hypertension during pregnancy (including gestational hypertension, preeclampsia, eclampsia, chronic hypertension, and chronic hypertension combined with preeclampsia), gestational diabetes or diabetes (impaired glucose tolerance) combined with pregnancy, prenatal hemorrhage (including placenta previa and abruption of the placenta)
Pediatrics	10	Neonatal hyperbilirubinemia, neonatal respiratory distress syndrome, neonatal sepsis, neonatal hypoxic-ischemic encephalopathy, pneumonia, diarrhea, nutritional diseases, nervous system disease, Kawasaki disease, and kidney disease

App Use Status

From the first launch on the Android platform on May 14, 2019, to the last version of the XueYiKu app updated in May 2023, the total number of teacher and student users was 6209 and 1180, respectively. Figure 2 represents the use status of the XueYiKu app during the test period and operation period. In total, there were 3224 person-times in the last 3 years, with an average of 771 person-times per year. From 2019 to 2021, the person-times logged in the application were the highest in 2019, at 1032 (125.4 on average per month), and increased from 646 in 2020 (53.8 on average per month) to 994 in 2021 (76.6 on average per month).

Figure 2. Changes in the number of student users of the XueYiKu app from May 2019 to February 2023.



XSL•FO
RenderX

We calculated the frequency at which subjects and related diseases were learned by students, and the top 10 most frequently learned subjects and diseases are listed in Table 3 in descending order. For the second discipline, the most frequently learned subjects were respirology (n=606, 62.2% in internal medicine), general surgery (n=506, 48.8% in surgery), and urinary surgery (n=390, 37.6% in surgery). For diseases, pneumonia was the most frequently learned, with 452 person-times (58.4% in internal medicine), followed by cholecystolithiasis (n=216, 28.4% in surgery), benign prostate hyperplasia (n=196, 25.8% in surgery), and bladder tumor (n=193, 25.4% in surgery).

Table . The top 10 most frequently learned subjects and diseases in the XueYiKu app.

	Person-time	Percentage (%)
Subjects[a]		
Internal Medicine	975	38.8
Respirology	606	62.2
Cardiology	201	20.6
Gastroenterology	88	9
Endocrinology	80	8.2
Surgery	1036	41.2
General surgery	506	48.8
Urinary surgery	390	37.6
Thoracic surgery	140	13.5
Gynecology and obstetrics	338	13.5
Gynecology	175	51.8
Obstetrics	163	48.2
Pediatrics	163	6.5
Diseases[a]		
Internal Medicine	774	50.4
Pneumonia	452	58.4
Coronary heart disease	99	12.8
Lung cancer	79	10.2
Arrhythmia	76	9.8
Chronic airway inflammation	68	8.8
Surgery	761	49.6
Cholecystolithiasis	216	28.4
Benign prostate hyperplasia	196	25.8
Bladder tumor	193	25.4
Appendicitis	79	10.4
Inguinal hernia	77	10.1

[a]The top 10 subjects and diseases were listed in descending order due to quantitative limitations.

Evaluation of the App and its Learning Effects

To evaluate the learning effect of the XueYiKu app, the scores of 479 students were analyzed. On the basis of a full score of 100, 15% (n=73) of students scored less than a passing grade of 60 points, roughly a third (n=168, 35.1%) had a score in the 60 to 80 range, and half of them had a score of over 80 points (n=238, 49.7%). A radar chart (Figure 3) with specific scores (Table 4) was presented considering multiple dimensions to evaluate the learning situation of students: students who learned cases with difficulty levels 3 - 4 achieved excellent learning outcomes regarding preciseness (83.4 for level 3), multiple dimensions (87.4 for level 4), and knowledge expansion (86.5 for level 3). However, students who learned cases with a difficulty of level 4 did not score as high as the other students in systematicness (34.1 for level 4), agility (30.1 for level 4), and logic (21.1 for level 4). In contrast, students who learned from cases with level 3 difficulty had positive effects in many areas (86.5 in knowledge expansion and 83.4 in preciseness). In fact, there was a high level of satisfaction with knowledge expansion and preciseness in all learners, but it is worth noting

XSL•FO
RenderX

that all levels were completely negative about systematicness, agility, and logic.

Figure 3. Radar chart with specific scores of clinical thinking evaluation for the XueYiKu app by difficulty in learning cases.

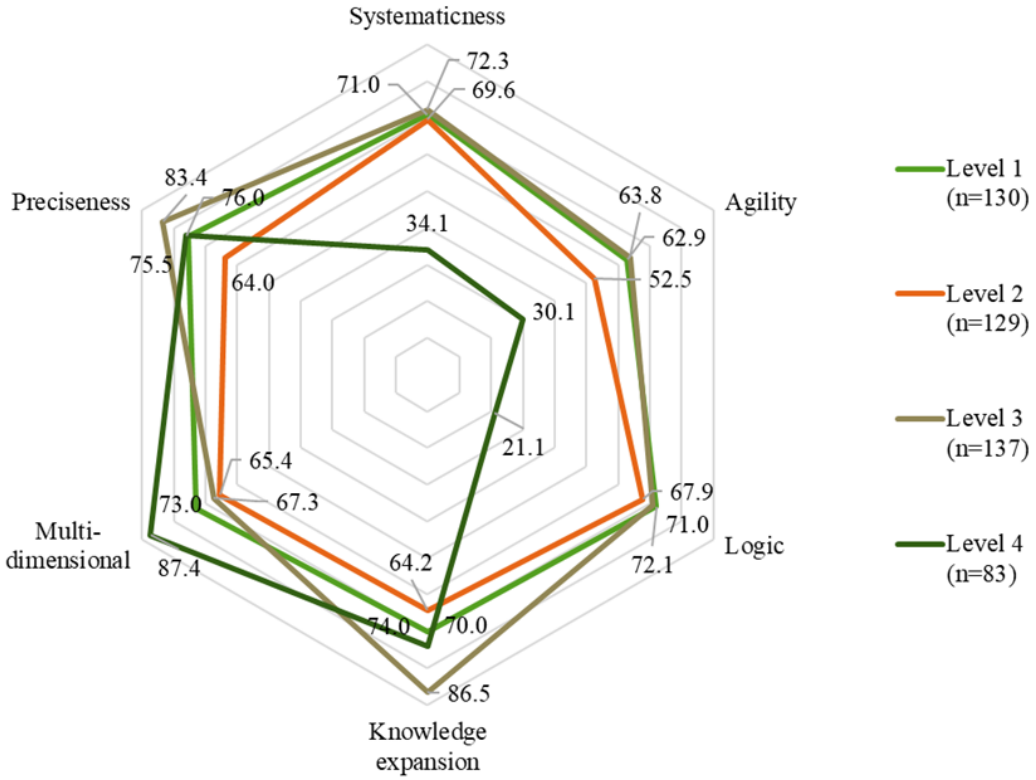


Table . Scores of clinical thinking for the XueYiKu app by the difficulty in learning cases and 6 dimensions according to the learning situation of the student.

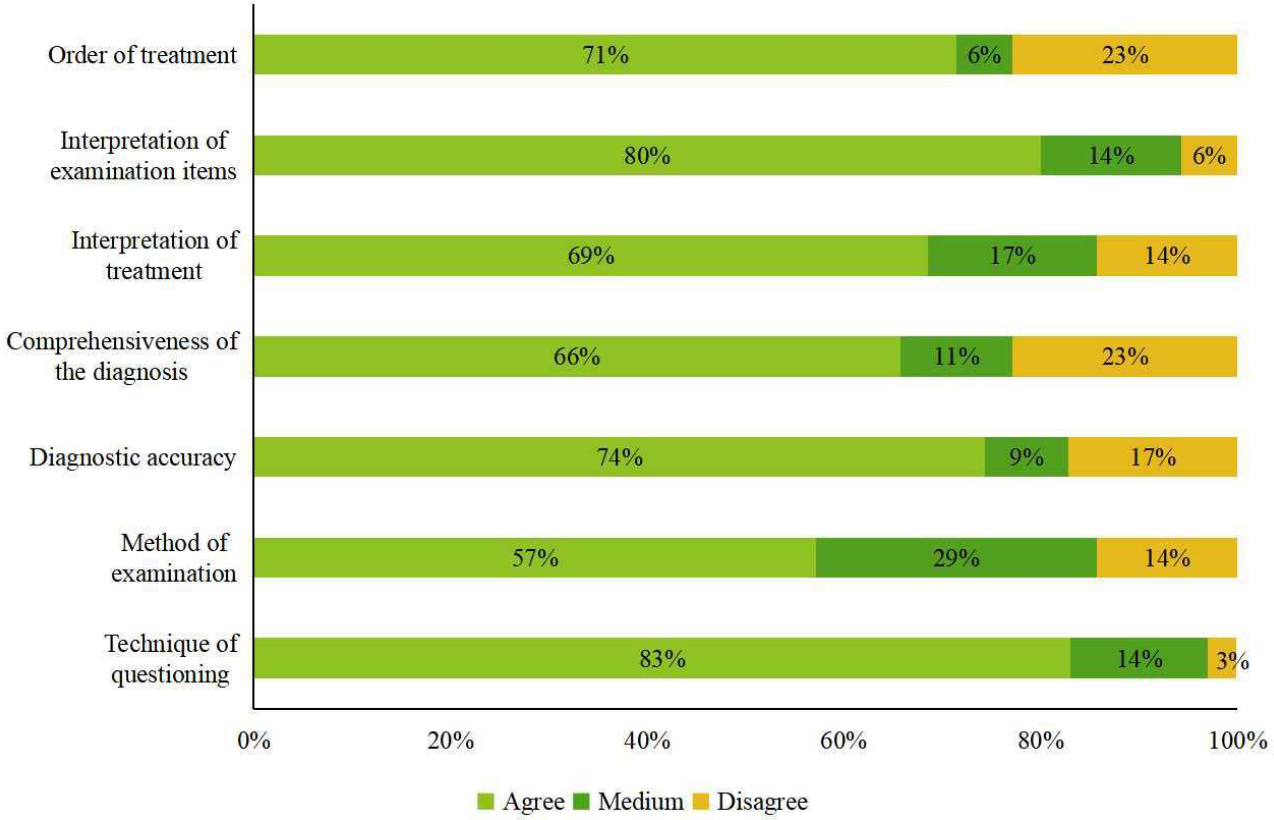
	Level 1 (n=130)	Level 2 (n=129)	Level 3 (n=137)	Level 4 (n=83)
Systematicness	71.0	69.6	72.3	34.1
Agility	62.9	52.5	63.8	30.1
Logic	72.1	67.9	71.0	21.1
Knowledge expansion	70.0	64.2	86.5	74.0
Multidimensional evaluation indicators	73.0	65.4	67.3	87.4
Preciseness	75.5	64.0	83.4	76.0

Evaluation of the App and Its Learning Effects

This section provides information regarding the 70 students' satisfaction of the XueYiKu app and their self-assessment of its learning effect (Figure 4). Overall, most students were satisfied with the app. The degree of satisfaction kept increasing considering the app function, the personal learning effect, and the case quality. For the self-assessment of learning outcomes, all sections received a rating of more than satisfactory, but the method of examination was rated the least satisfactory. In other sections, the technique of questioning received the highest satisfaction rating, followed by the interpretation of examination items, with both categories receiving more than 80% positive feedback. In contrast, the diagnostic accuracy and the order and interpretation of treatment ranked next in order (~70%), and the other factors were at approximately 60%. Most of the dissatisfaction was rated for the order of treatment and the comprehensiveness of the diagnosis, which were responsible for more than 20% of the negative ratings. Except for the questioning technique, the proportion of negative ratings was more pronounced in all sections. Additionally, for the method of examination, about 30% rated "medium," which differed greatly from other sections' ratings.

XSL•FO
RenderX

Figure 4. Self-assessment and satisfaction of 70 students towards the learning outcomes for the XueYiKu app.



Feedback on the App and Its Teaching Effects

According to the feedback, there were some issues associated with the app that are worthy of attention. First, the original login method should be simplified, as well as that for the online case communication and sharing section. Additionally, discussion and comment functions should be added. Regarding data processing, the AI answers could optimize the fuzzy search, and the system could be better connected to the residency system so that students can check and manage the patient's detailed medical record information. Regarding the case corpus, there is still no section in the app to read images. In addition, not only do the users of the app want to obtain clearer guidance or feedback in future updates, but it would also be beneficial if special departments fulfilled the auxiliary examination and clinical guidelines. Finally, some of the cases included in the app are so simple that they are only suited for undergraduate medical students; it would be beneficial to add cases that are rare, complex, or even life-threatening, and to include the postoperative complications, judgment, treatment, and other information relevant to the case.

Discussion

Principal Findings

Clinical thinking training for medical students is very important; however, few design cases rely on IT for self-learning. The current study uses AI and big data platforms to establish a virtual learning platform based on real clinical cases. The XueYiKu app can assist in the development of medical students' clinical thinking skills by allowing them to perform consultation, physical examination, auxiliary examination, diagnosis, and treatment on virtual patients; the app can also design real cases with different difficulties to meet the needs of residents' clinical thinking training at different stages. Through the evaluation of students' experiences with the app, this study concludes that medical students' learning can become more active and self-motivated, using a variety of formats, with real-time feedback through the platform's assessment. The learning effect is also satisfactory overall and provides important methods and ideas for establishing scientific models and methods for assessing clinical thinking skills in the future.

Comparison to Prior Work

Analysis of the data shows that the number of case studies was positively correlated with the number of residents in the specialty. The top 4 diseases most frequently learned included pneumonia, cholecystolithiasis, benign prostate hyperplasia, and bladder tumor, all of which are common, require adequate diagnosis and differential diagnosis, and meet the basic goal of clinical thinking training at the residency stage. The platform contains 65 common diseases and 400 study cases covering internal medicine, surgery, obstetrics and gynecology, and pediatrics. The design is reasonable and can meet the basic learning needs of residents in various specialties. In the 1990s, DxR Clinician was developed by DxR Development Group, Inc. [7] in the United States, and it was later introduced and translated into Chinese by the China Union Medical University Press in 2009 [13]. However, the use of this software in China is still in its infancy, with few relevant literature reports. A previous study showed that the cases included in the DxR software were virtual cases and were limited by the purchase [7]. Students needed to complete the study on the computer, and the average time for students to become familiar with the

XSL•FO
RenderX

software was as long as 2 hours. In contrast, the types of diseases and the number of cases covered in the XueYiKu app, as well as the ease of use, are advantages. In addition, the framework of this platform is reasonable, as new case types of diseases can be added according to students' needs, and the user interface is simple and fast. Therefore, this app is more suitable for residents' clinical thinking training in the global internet and postepidemic era.

There is a decline in the training and formation of clinical thinking in residents. Based on this, the real case base used in this app was designed to present 4 levels of difficulty for the same case. Evaluation and feedback in the process of clinical thinking training formation and refinement are also important. As a result, students were evaluated in multiple dimensions according to the learning situation. Analysis of the results showed that the learning process meets the clinical diagnosis and treatment reality. For first- to second-year residents, level 2 - 3 difficulty was the most appropriate, as illustrated by their better performance in knowledge expansion and logic, among other factors. The clinical learning process is in line with clinical practice and meets the learning needs of residents. Specific to the learning process at different levels of difficulty, the primary difficulty allowed for a complete diagnosis based on the completed clinical information, and as the difficulty level increased, there was an advancement of the learning model; the results of the 4 levels of difficulty were in line with the development of the resident's clinical thinking, realizing the intended purpose of the app. One similar app currently in use is Brianly, a medical tutoring and social networking platform that provides personalized medical learning tutoring through AI algorithms [14]. In Tsinghua University, "artificial intelligence + virtual reality" within Brianly is used for clinical training, which allows direct viewing of the anatomical details of a patient's real body structure in a virtual space, as well as virtual surgery. However, there is currently no learning software for medical students that is designed to match the growth and learning process, and that provides a difficulty gradient based on the clinical thinking ability. In this context, our current app will continue to develop, improve, and provide feedback related to the 4 levels of difficulty to be better used for the clinical thinking training of senior residents.

Based on current usage, the highest number of uses per capita was usually found during the admission and start of residency, that is, from September to February each year. The survey results show that students had a high level of satisfaction with the consultation aspect of the app. The case study on the app allows for both a generic history-taking process and a structured questioning for a particular disease, which is inextricably linked to the teaching case base from clinical consultations of real cases. However, there are some differences between the examination and the actual situation, and the contents of the specialist examination and the physical examination for a particular disease should be further optimized to be more concise and focused.

Strengths and Limitations

One of the greatest advantages of our app is the inclusion of real cases, which could provide a new model of clinical thinking training by building a system to comprehensively assess the "clinical competence" of residents at different stages. For students, the strongest advantage is that they can customize clinical competency training according to their needs without the limitations of time and space. For app designers, the XueYiKu app is inherently user-friendly and sustainable, which allows the addition of cases in bulk based on the existing framework.

In addition to the benefits of the app, there are several limitations that could be further improved. First, current cases and diseases could be further expanded to cover more subjects, which is the next step of development. Second, in terms of use, students currently use the app independently, while the frequency of use is mainly based on students' interests combined with required course study and clinical rotation. Third, current evaluation results are almost subjective evaluations by humans. Thus, the evaluation results may be biased.

Conclusion

The emergence of virtual network learning systems such as the XueYiKu app can not only rapidly cultivate students' ability to solve clinical problems but also save teaching costs, improve teaching quality and efficiency, and achieve the sharing of high-quality teaching resources. In addition to enriching the content of clinical teaching, students with diverse experience and innovative thinking are trained, which promotes the basic clinical knowledge of medical students and effectively cultivates their clinical thinking ability.

Acknowledgments

The authors would like to thank all participants for willingly sharing their experiences. This study was supported by grant #22-458 Development and implementation of an IT-based competency assessment system for Internal Medicine and Obstetrics & Gynecology resident training programs in China. The study funders had no rule in the study design, implementation, analysis, manuscript, preparation, or decision to submit this article for publication.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author upon reasonable request.

XSL•FO
RenderX

Authors' Contributions

YW, NS, and JQ contributed equally to this study. YW and HW conceptualized the study and collected the data. HW and DZ jointly analyzed the data and drafted the article, which was later refined by YW, who also drafted and revised the paper. MW and HJ contributed to data analysis and offered comments. JH and NS provided critical comments and reviewed the important intellectual content. JQ contributed to conception and design as well as the final approval of the version to be published as the corresponding author.

Conflicts of Interest

None declared.

References

1. Richards JB, Hayes MM, Schwartzstein RM. Teaching clinical reasoning and critical thinking: from cognitive theory to practical application. Chest 2020 Oct;158(4):1617-1628. [doi: 10.1016/j.chest.2020.05.525] [Medline: 32450242]
2. Reinus JF. Patient-data management and clinical reasoning: neglected essential skills. Am J Med 2022 Dec;135(12):1393-1394. [doi: 10.1016/j.amjmed.2022.08.022] [Medline: 36063858]
3. Karunaratne D, Sibbald M, Chandratilake M. Understanding cultural dynamics shaping clinical reasoning skills: A dialogical exploration. Med Educ 2024 Aug 1. [doi: 10.1111/medu.15479] [Medline: 39089689]
4. Locke R, Mason A, Coles C, Lusznat RM, Masding MG. The development of clinical thinking in trainee physicians: the educator perspective. BMC Med Educ 2020 Jul 16;20(1):226. [doi: 10.1186/s12909-020-02138-w] [Medline: 32678045]
5. Lee DS, Abdullah KL, Subramanian P, Bachmann RT, Ong SL. An integrated review of the correlation between critical thinking ability and clinical decision-making in nursing. J Clin Nurs 2017 Dec;26(23-24):4065-4079. [doi: 10.1111/jocn.13901] [Medline: 28557238]
6. Carin L. On artificial intelligence and deep learning within medical education. Acad Med 2020 Nov;95(11S):S10-S11. [doi: 10.1097/ACM.0000000000003630] [Medline: 32769462]
7. Myers JH, Dorsey JK. Using diagnostic reasoning (DxR) to teach and evaluate clinical reasoning skills. Acad Med 1994 May;69(5):428-429. [doi: 10.1097/00001888-199405000-00052] [Medline: 8086079]
8. Xu X, Liu F. Optimization of online education and teaching evaluation system based on GA-BP neural network. Comput Intell Neurosci 2021;2021:8785127. [doi: 10.1155/2021/8785127] [Medline: 34422036]
9. Nagi F, Salih R, Alzubaidi M, et al. Applications of artificial intelligence (AI) in medical education: a scoping review. Stud Health Technol Inform 2023 Jun 29;305:648-651. [doi: 10.3233/SHTI230581] [Medline: 37387115]
10. Hsiao CT, Chou FC, Hsieh CC, Chang LC, Hsu CM. Developing a competency-based learning and assessment system for residency training: analysis study of user requirements and acceptance. J Med Internet Res 2020 Apr 14;22(4):e15655. [doi: 10.2196/15655] [Medline: 32286233]
11. Webster CS. Artificial intelligence and the adoption of new technology in medical education. Med Educ 2021 Jan;55(1):6-7. [doi: 10.1111/medu.14409] [Medline: 33155316]
12. Wang M, Sun Z, Jia M, et al. Intelligent virtual case learning system based on real medical records and natural language processing. BMC Med Inform Decis Mak 2022 Mar 4;22(1):60. [doi: 10.1186/s12911-022-01797-7] [Medline: 35246134]
13. Meng Q, Zhaoxia T. Promoting effect of information products on case-based clinical teaching model. J Med Inform 2012(8):89-92. [doi: 10.3969/j.issn.1673-6036.2012.08.024]
14. Han ER, Yeo S, Kim MJ, Lee YH, Park KH, Roh H. Medical education trends for future physicians in the era of advanced technology and artificial intelligence: an integrative review. BMC Med Educ 2019 Dec 11;19(1):460. [doi: 10.1186/s12909-019-1891-5] [Medline: 31829208]

Abbreviations

AI: artificial intelligence
CBLAS: competency-based learning and assessment system
DxR: diagnostic reasoning

Edited by A Mavragani; submitted 26.04.24; peer-reviewed by D Hu, Q Xia; revised version received 23.09.24; accepted 24.09.24; published 03.01.25.

Please cite as:
Wang H, Zheng D, Wang M, Ji H, Han J, Wang Y, Shen N, Qiao J
Artificial Intelligence–Powered Training Database for Clinical Thinking: App Development Study
JMIR Form Res 2025;9:e58426
URL: https://formative.jmir.org/2025/1/e58426
doi:10.2196/58426

XSL•FO
RenderX

© Heng Wang, Danni Zheng, Mengying Wang, Hong Ji, Jiangli Han, Yan Wang, Ning Shen, Jie Qiao. Originally published in JMIR Formative Research (https://formative.jmir.org), 3.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Physical Activity at Work (PAW) Program in Thai Office Workers: Mixed Methods Process Evaluation Study

Katika Akksilp[1,2], MD, PhD; Thomas Rouyard[3], PhD; Wanrudee Isaranuwatchai[2], PhD; Ryota Nakamura[4], PhD; Falk Müller-Riemenschneider[1,5,6], PhD; Yot Teerawattananon[1,2], PhD, MD; Cynthia Chen[1,5], PhD

[1]Health Intervention and Technology Assessment Programme, Ministry of Public Health, Nonthaburi, Thailand

[2]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore

[3]Graduate School of Public Health & Health Policy, City University of New York, New York, NY, United States

[4]Hitotsubashi Institute for Advanced Study, Hitotsubashi University, Tokyo, Japan

[5]Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore

[6]Digital Health Center, Berlin Institute of Health, Charité-Universitätsmedizin Berlin, Berlin, Germany

Corresponding Author:
Katika Akksilp, MD, PhD
Health Intervention and Technology Assessment Programme
Ministry of Public Health
6th Floor, 6th Building, Department of Health
Tiwanon Rd., Muang
Nonthaburi, 11000
Thailand
Phone: 66 2 590 4549
Fax: 66 2 590 4369
Email: gochi.akk@gmail.com

Abstract

Background: An increasing number of multicomponent workplace interventions are being developed to reduce sedentary time and promote physical activity among office workers. The Physical Activity at Work (PAW) trial was one of these interventions, but it yielded an inconclusive effect on sedentary time after 6 months, with a low uptake of movement breaks, the main intervention component.

Objective: This study investigates the factors contributing to the outcomes of the PAW cluster randomized trial.

Methods: Following the Medical Research Council's guidance for process evaluation of complex interventions, we used a mixed methods study design to evaluate the PAW study's recruitment and context (how job nature and cluster recruitment affected movement break participation), implementation (dose and fidelity), and mechanisms of impact (assessing how intervention components affected movement break participation and identifying the facilitators and barriers to participation in the movement breaks). Data from accelerometers, pedometers, questionnaires, on-site monitoring, and focus group discussions were used for the evaluation. Linear mixed effects models were used to analyze the effects of different intervention components on the movement breaks. Subsequently, qualitative analysis of the focus group discussions provided additional insights into the relationship between the intervention components.

Results: The participation in movement breaks declined after the third week, averaging 12.7 sessions (SD 4.94) per participant per week for the first 3 weeks, and continuing to decrease throughout the intervention. On-site monitoring confirmed high implementation fidelity. Analysis of Fitbit data revealed that each additional movement break was associated with a reduction of 6.20 (95% CI 6.99-5.41) minutes in sedentary time and an increase of 245 (95% CI 222-267) steps. Regarding the mechanisms of impact, clusters with higher baseline sedentary time demonstrated greater participation in movement breaks, while those with frequent out-of-office duties showed minimal engagement. Moreover, clusters with enthusiastic and encouraging movement break leaders were associated with a 24.1% (95% CI 8.88%-39.4%) increase in participation. Environmental and organizational support components using posters and leaders' messages were ineffective, showing no significant change in percentage participation in movement breaks (4.49%, 95% CI –0.49% to 9.47% and 1.82%, 95% CI –2.25% to 5.9%, respectively). Barriers such as high workloads and meetings further hindered participation, while the facilitators included participants' motivation to feel active and the perceived health benefits from movement breaks.

XSL•FO
RenderX

Conclusions: Despite high fidelity, the PAW trial did not significantly reduce sedentary time, with limited uptake of movement breaks due to context-related challenges, ineffective environmental support, and high workloads during the COVID-19 pandemic.

(JMIR Form Res 2025;9:e57604) doi:10.2196/57604

KEYWORDS

process evaluation; sedentary behavior; physical activity; workplace; movement breaks

Introduction

Background

Extensive research has examined workplace interventions to reduce sedentary behavior and promote physical activity globally [1,2]. Numerous interventions have yielded noteworthy results, showcasing their effectiveness, whereas some have encountered challenges in achieving desired outcomes [2-5]. While early meta-analyses, such as one from 1998, suggested minimal impact of workplace interventions on physical activity [6], subsequent reviews, including one from 2003, began to recognize their potential benefits [7]. However, more recent evidence has provided a clearer picture, indicating that more successful interventions tended to use pedometers, apply internet-based approaches, and include activities at social and environmental levels [5]. One systematic review found that physical environment interventions, whether implemented at the workstation, building, or neighborhood level, were effective in promoting physical activity in the workplace [8]. Finally, complex or multicomponent interventions emerge as the most effective workplace strategy for reducing sitting time [4], highlighting the intricate process leading to successful behavioral changes [9].

However, multicomponent intervention trial reports often fail to elucidate the rationale and mechanisms behind their results, leaving readers with questions regarding the efficacy of the interventions. Many scholars have criticized these trials as being akin to a “black box” because the underlying reasons for their success or failure remain unknown [10]. As a result, process evaluations, which assess how interventions were implemented in practice, are essential for deciphering the implications of the results from multicomponent intervention trials. These evaluations highlight aspects of the intervention that may require improvement to increase the probability of success [11,12].

Several frameworks advocate for process evaluations to explore the context, implementation, and impact mechanisms of intervention programs, encompassing aspects such as recruitment, reach, dose, fidelity, and challenges [10-13]. Process evaluations have become particularly relevant because they can help determine whether success or failure lies in the implementation, the intervention itself, or a combination of both factors. Notably, the Medical Research Council’s guidance on process evaluation of complex interventions is particularly valuable for assessing multicomponent interventions [11]. By contrast, frameworks like the one proposed by Steckler and Linnan [13] focus more on specific elements such as dose and fidelity but may not provide the broader perspective necessary to understand the interactions between context, mechanisms, and implementation. The emphasis of the Medical Research Council’s guidance on context and mechanisms was pivotal in capturing the socioecological dynamics within our intervention.

In addition, there is increasing acknowledgment that incorporating both qualitative and quantitative data, as well as using theoretical frameworks within process evaluations, plays a crucial role in promoting evidence-based practice [14]; for example, the Older People’s Exercise Intervention in Residential and Nursing Accommodation trial [15] used a mixed methods process evaluation to assess whether the multicomponent intervention changed residents’ home culture to increase physical activity and whether residents engaged in exercise activities. The study found no notable cultural shifts or sufficient engagement, thus explaining the trial’s null results [15,16]. Another process evaluation explored the implementation and impact mechanisms of a park prescription intervention trial, revealing key mediators of the intervention effects, such as park physical activity levels, as well as barriers that may have weakened intervention effectiveness [17].

Objectives

In Thailand, it has been reported that the majority of adults who engage in sufficient physical activity reside in rural areas, typically due to their work in the agricultural sector, whereas physical activity levels tend to be lower among office workers [18]. Numerous nonresearch initiatives have been launched to promote physical activity and reduce sedentary behavior among office workers in Thai companies and organizations. The Physical Activity at Work (PAW) trial [19] marked the first comprehensive cluster randomized trial aimed at promoting physical activity and reducing sedentary behavior among Thai office workers. This trial incorporated a multicomponent intervention and used accelerometer-measured data, laying a solid foundation for physical activity research in Thailand [20]. However, the trial produced inconclusive findings, leaving critical questions regarding the factors contributing to its outcomes unanswered. We observed a suboptimal uptake of movement breaks, the primary intervention component, and inferred that this, coupled with a low recruitment rate, may explain the absence of statistically significant outcomes [20]. Understanding these underlying reasons is essential for making necessary improvements in both research and policy. To address this, we conducted a process evaluation to investigate the factors contributing to the trial’s outcome.

Following the Medical Research Council’s guidance for process evaluation of complex interventions [11], we used a mixed methods study design to evaluate the PAW study’s recruitment and context, implementation, and impact mechanisms. Specifically, we examined how job nature and recruitment processes within each cluster influenced participation in movement breaks (context), the overall dose and fidelity of the

XSL•FO
RenderX

intervention (implementation), and the effects of intervention components on participation in movement breaks as well as the facilitators and barriers to movement break participation (impact mechanisms).

Methods

Description of the Cluster Randomized Trial

The PAW study was conducted in a free-living setting at the Ministry of Public Health offices in Nonthaburi, Thailand. Detailed methodological information regarding the cluster randomized trial (Thai Clinical Trials Registry: TCTR20200604007) was published in the protocol [19] and main results [20] manuscripts. Briefly, between July and September 2021, we recruited 282 office workers (age: mean 38.6, SD 10.4, years; female: n=228, 80.9%) from 18 separate offices. After baseline data collection, the participants were randomized into 9 intervention offices (140/282, 49.6%) and 9 control offices (142/282, 50.4%). The 6-month intervention, which took place between September 2021 and March 2022, consisted of 6 components across 4 levels, as outlined in Textbox 1.

Textbox 1. Key elements of the Physical Activity at Work cluster randomized trial.

Individual-level components

- Providing a wearable device with real-time activity feedback (Fitbit Inspire HR; Google LLC)
- Using a Fitbit smartphone app
- Offering individual weekly lottery-based financial incentives, where 1 intervention group participant who participated in at least 70% of the movement breaks within the previous week was randomly selected to receive a reward of THB 500 (US $16)

Social-level components

- Team movement breaks of light to moderate intensity, lasting at least 4 minutes and occurring 4 times a day (alarm clocks with speakers were provided to movement break leaders to initiate sessions), which served as the primary intervention component
- Team-based incentives of an additional weekly lottery reward of THB 500 (US $16) given to the winner if at least 70% of the participants in the cluster also attended at least 70% of the movement breaks within the previous week; 4 alarm reminders were set (9:30 AM, 10:30 AM, 2:30 PM, and 3:30 PM), and trained movement break leaders managed the starting times, songs, and movements; participants working from home were encouraged to join the sessions via web conferencing

Environmental-level component

- Three types of posters providing information on health risks associated with high sedentary time and the benefits of physical activity, as well as examples of stretching exercises

Organizational-level component

- Leadership support in the form of messages sent by office directors twice a week via Line (LY Corporation) to encourage participants to reduce sedentary time using movement breaks and increase physical activity, as well as to announce reward winners with congratulatory messages and photographs of the reward ceremony

The multicomponent intervention was developed using the socioecological model [9,21]. The intervention components were strategically designed to complement one another, with a particular focus on the primary component: the movement breaks. In addition, department directors actively encouraged participants to attend more movement break sessions and announced the weekly lottery reward, which was contingent upon the frequency of participation in the movement breaks. Two movement break leaders per office were trained to oversee these sessions. Moreover, posters were displayed to inform participants about the adverse effects of prolonged sedentary behavior and to provide examples of strategies to break up sedentary time in the office. All intervention participants were equipped with a Fitbit Inspire HR smartwatch (Google LLC), preset to remind them to interrupt prolonged sitting every 30 minutes. Consequently, the intervention operated as illustrated in Figure 1.

We formulated our research questions to encompass the context, implementation, and impact mechanisms, with a specific emphasis on the primary intervention component: the movement breaks. This emphasis stemmed from two key considerations: (1) as previously noted, our intervention design places the movement break as the central intervention component, supported by other components; and (2) we observed a low participation rate (median 31.5%, IQR 20.4%-42.7%) in the movement break sessions. Moreover, 77.5% (107/138) of the intervention participants attended fewer than half of the breaks (240/480, 50%) throughout the intervention period. We hypothesized that the absence of a significant intervention effect could be attributed to low attendance in the movement break sessions [20]. Hence, our focus is to investigate the underlying reasons behind this issue. Table 1 describes the process evaluation components, specific questions, and the data sources.

XSL•FO
RenderX

Figure 1. The mechanism of associations between the Physical Activity at Work cluster randomized trial intervention components and outcomes.

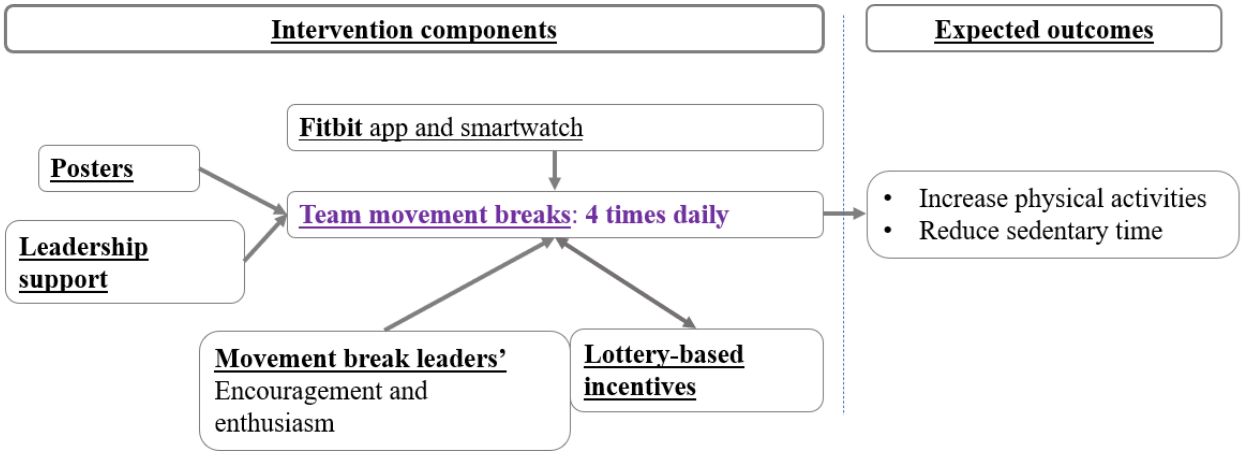


Table 1. Descriptions of the Physical Activity at Work process evaluation components, specific questions, and the data sources.

Process evaluation components	Specific process evaluation questions	Quantitative data	Qualitative data
Recruitment and context			
Recruitment	How does recruitment influence participation in movement breaks?	Recruitment rate	Recruitment fashion
Context	What impact do job descriptions and baseline characteristics of each cluster have on movement break participation?	Baseline summary statistics of each cluster	Job description of each cluster
Implementation			
Dose	What was the extent of intervention delivery?	Web monitoring: Fitbit wear time and conducts of movement breaks	On-site monitoring
Fidelity	Was the intervention delivered as intended?	Associations between participation in movement breaks and daily Fitbit sedentary time data as well as step counts during the intervention period	On-site monitoring
Mechanisms of impact			
Testing the assumptions behind the intervention design	How did supporting components influence participation in movement breaks?	Effects of intervention components on the percentage of weekly movement break participation	Focus group discussions
Participants' attitude	What were the facilitators and barriers to participation in movement breaks?	Additional questionnaire regarding facilitators and barriers to participation in movement breaks	Focus group discussions

Ethical Considerations

Ethics approval for both the quantitative and qualitative components of this study was granted by the Ethical Review Committee for Research in Human Subjects, Ministry of Public Health, Thailand (IRB00001629), in accordance with relevant ethical guidelines for human research. Any modifications to the approved protocol will be submitted to the ethics committee for review. All participants provided written informed consent before participation, which included consent for the use of deidentified data in secondary analyses and for the publication of results. To protect participants' privacy and confidentiality, all data were deidentified, and no personally identifiable information was included in the final dataset, with all published data presented in aggregate form. Participants' data (eg, case record forms, laboratory test results, information sheets, and consent forms) are stored in a locked cabinet in a researcher's office. All data will be destroyed by researchers within 5 years of publication. During the study, only deidentified data were used, and the data were only accessible to the research team. Participants received compensation of THB 250 (approximately US $7.5) for each data collection session in which they participated.

Dissemination Policy and Authorship Guidelines

In addition to disseminating our research findings to the funder of this study, the Ministry of Public Health, we will disseminate our findings to other countries, the study participants, and the research community. We have followed the authorship guidelines of the International Committee of Medical Journal Editing.

Data Collection

Baseline and Follow-Up Data Collection (Questionnaire)

An interviewer-administered questionnaire, based on the Thai National Statistical Office's health survey and capturing

XSL•FO
RenderX

sociodemographic data such as age and education [22], was used to collect participant data at both baseline and follow-up. In addition, intervention participants were asked supplementary questions regarding the implementation of the intervention during the follow-up data collection session (Multimedia Appendix 1).

Data Monitoring (Movement Break Schedule)

A data monitoring team, consisting of the trial implementer and an administrator, conducted weekly random field visits to 4 offices during the scheduled movement break sessions, adhering to the initial timetable. Monitored data included (1) the occurrence of sessions, (2) session quality, (3) participant attendance, and (4) any issues encountered during the session.

Movement break leaders were responsible for submitting weekly schedules via Microsoft Excel Online, indicating session timings and dates. Updates were required if a session did not fall within the 30-minute alarm windows (9:30 AM, 10:30 AM, 2:30 PM, and 3:30 PM), indicating absence or starting times outside the specified windows.

Fitbit Data and Movement Break Participation

During the 6-month intervention period, we collected Fitbit data, which included wear time, sedentary time, and step counts. Daily activity data were anonymized and used to compile weekly lists of potential reward recipients and to assess compliance with movement breaks. We determined participation in each movement break session by (1) analyzing a quadratic function of steps, (2) detecting quadratic functions within the 1-hour time frame of movement breaks, and (3) ensuring that participants had accumulated at least 100 steps.

Quantitative Data Analysis

Linear mixed models were used to examine the associations between intervention components and activity outcomes, as well as between the intervention components themselves. These models incorporated random intercepts from both individual and cluster levels, while also accounting for random slope by including intervention weeks. Weekly individual data were derived from daily Fitbit records, follow-up questionnaire responses, and monitoring data (Table 2). Data were analyzed using RStudio (version 4.0.3; Posit Software, PBC) and Stata software (version 14.2; StataCorp LLC), with significance level set at 5%.

XSL•FO
RenderX

Table 2. Variables used in the Physical Activity at Work process evaluation.

Variable	Explanation	Scale
Associations between movement break participation and daily Fitbit sedentary time and step count data during the intervention period		
Outcome variables		
Sedentary time	Daily sedentary time data from Fitbit	Continuous (min)
Steps	Daily step count data from Fitbit	Continuous (step count)
Exposure variables		
Movement break participation	Daily movement break participation	Ordinal; 1-4 (sessions)
Effects of intervention components on weekly movement break participation percentage		
Outcome variable		
Movement break participation percentage	The frequency of an individual's participation in movement break sessions within a specific week compared to the maximum number of sessions available during that week	Continuous (percentage)
Exposure variables		
Individual reward	We designated the weeks following individual reward wins as {1}, while all other weeks were coded as {0}, using the web monitoring data	1=won a reward last week; 0=others
Cluster reward	We assigned {1} to the weeks following office colleagues' reward wins and {0} to all other weeks, using the web monitoring data	1=a colleague in the same office cluster won a reward last week; 0=others
Fitbit wear time	We categorized the average weekly Fitbit wear time for each individual into a binary variable using the median	1=wore Fitbit at or above the median wear time; 0=wore Fitbit below the median wear time
Fitbit wear time (alternative)	How often Fitbit led individuals to engage in physical activity in the office in the last 2 weeks of the intervention	1=at least 4 d/wk; 0=<4 d/wk
Leadership support	How often the directors' support led individuals to engage in physical activity in the office in the last 2 weeks of the intervention	1=at least 4 d/wk; 0=<4 d/wk
Movement break leaders	How the movement break leaders' (1) encouragement and (2) enthusiasm contributed to movement break participation (We asked the movement break leaders and used the data as cluster-representative data)	3=both factors contributed *somewhat* or *a lot*; 2=only the encouragement contributed *somewhat* or *a lot*; 1=only the enthusiasm contributed *somewhat* or *a lot*; 0=both factors contributed *not at all*, *very little*, or *a little*
Posters	How often the posters led individuals to engage in physical activity in the office in the last 2 weeks of the intervention	1=at least 4 d/wk; 0=<4 d/wk
Posters (alternative)	How accurate individuals were in reporting the number of different poster designs in their office	1=3-4 designs; 0=0-2 designs

Focus Group Discussions

The qualitative aspect of the study involved focus group discussions aimed at exploring participants' perspectives on facilitators and barriers to engagement in movement breaks, as well as their attitudes toward intervention components. Intervention clusters were initially ranked by mean percentage participation in movement breaks. Subsequently, using purposive sampling, we invited up to 6 participants with the highest percentage participation from the top 2 clusters and up to 6 participants with the lowest percentage participation from the bottom 2 clusters for each of the 4 focus group discussions.

The 4 focus group discussions took place via 45- to 75-minute Zoom (Zoom Video Communications, Inc) meetings, with participants joining from their respective offices. The focus group discussions were recorded with participants' consent (both video and audio were captured), while notes were taken during the sessions. Before each focus group discussion, participants were briefed about its purpose, format, and estimated duration, with the option to interrupt as needed.

Verbatim transcriptions were manually completed by hired research assistants. The transcriptions were then subjected to deductive thematic analysis, with facilitator and barrier serving as overarching themes, and the socioecological model used as subtheme [9]. Two analysts (KA and WI) independently coded each transcript using the pre-established framework. References under themes and subthemes were compared and discussed to ensure consistency and accuracy. Intercoder reliability was not analyzed in this study.

An interview guide (Table S1 in Multimedia Appendix 2) was developed based on the research question "What are the facilitators and barriers to movement break participation?" The socioecological model informed this guide. A total of 4 interviewers (female: n=2, 50%) conducted the focus group discussions. They comprised the PAW trial implementer (first author; doctor of medicine), 2 study administrators (bachelor

XSL•FO
RenderX

of communication arts and master of political science), and a research assistant (bachelor of clinical pharmacology) who was not a trial staff member. All interviewers underwent a comprehensive 2-day training program tailored to the qualitative data collection and analysis requirements of the study. This training was led by a Thai senior researcher (PhD in anthropology). The focus group discussion guide served as a reference tool, ensuring that critical topics were not overlooked. During the focus group discussions, the researcher flexibly adjusted the sequence, content, and style based on individual responses. Emotions and nonverbal cues expressed by participants were carefully documented.

Code names were used in place of real names in the recorded data to protect participant privacy. Data collection and analysis proceeded concurrently throughout the study. After coding the transcripts from the 4 focus group discussions, the researchers and supervisors deliberated on data saturation. Subsequently, 2 additional focus group discussions were conducted with participants from clusters ranked in the top and bottom thirds, following the same procedures. No new themes emerged from these final 2 focus group discussions, indicating that data saturation had been achieved.

Results

Recruitment and Context

The PAW study recruited participants from all clusters within the Department of Medical Services building and the International Health Policy Program, Ministry of Public Health, Thailand, between July and September 2020. The ministry's governance structure is typically bureaucratic, with each department led by a single director general and each office headed by an office director. In total, 18 clusters were successfully recruited, with 15 (83%) falling under the Department of Medical Services purview; the remaining 3 (17%) clusters were under the jurisdiction of the Office of the Permanent Secretary (n=2, 67% were part of the International Health Policy Program and located in a distinct building complex; Figure S1 in Multimedia Appendix 2).

Recruitment procedures at the cluster level proceeded seamlessly. The directors of each office endorsed the active involvement of their respective teams in the trial, resulting in a cluster-level recruitment rate of 100% (18/18). For individual-level recruitment, we initiated the process by organizing office-specific group meetings to thoroughly explain the trial details. After these meetings, office workers interested in participating could engage with the trial staff to complete the individual informed consent process. To accommodate those unable to attend the scheduled meetings due to prior commitments, additional sessions were arranged.

Despite the bureaucratic governance style, the individual recruitment rate reached 62.9% (282/449). Job descriptions varied and included research-related roles and academic, finance, law, digital, and other administrative positions. At baseline, 3 (17%) of the 18 clusters exhibited a mean daily sedentary time exceeding 9 hours (mean sedentary time of 546, SD 90.3 min). Moreover, mean daily moderate to vigorous physical activity time was generally higher than the current physical activity guideline, reflecting that participants were relatively active. No significant associations were observed among job descriptions, baseline time spent in sedentary or moderate to vigorous physical activity, and monitoring data. However, clusters with higher baseline sedentary time demonstrated increased participation in movement breaks (Figure S2 in Multimedia Appendix 2). Furthermore, mean age and cluster size seemed to influence participation in movement breaks within the intervention group. Notably, individuals in the cluster with the youngest and smallest number of participants showed minimal participation in movement breaks (Table 3).

Table 3. Recruitment and context of intervention participants and the implementation of the Physical Activity at Work intervention (n=140).

	Baseline (n=140)			Data monitoring (6-mo intervention period) (n=138)			
Job description	Cluster size, n (%)	Recruitment rate n/N (%)	Sedentary time (min/d), mean (SD)	Lottery-incentive wins, n (%)	Movement break participation (%), mean (SD)	Maximum work from home, n/N (%)	Notes
Research related	23 (16.4	23/25 (92)	541 (90.5)	15 (63)	59.7 (14.7)	2.5/5 (50)	—[a]
Research related	8 (5.7)	8/44 (18)	483 (99.4)	0 (0)	25.8 (13.2)	5/5 (100)	Different building and department
Nursing	13 (9.3)	13/37 (35)	463 (117)	4 (17)	38.8 (16.6)	0/5 (0)	Different department
Finance	10 (7.1)	10/11 (90)	460 (95.9)	0 (0)	32.9 (5.49)	2/5 (40)	Broadcast from another cluster
Finance	14 (10.0)	14/17 (82)	481 (94.1)	1 (4.17)	29 (16.4)	2/5 (40)	—
Human resource	34 (24.3)	34/35 (97)	473 (113)	2 (8)	31.5 (11)	2.5/5 (50)	—
Human resource	18 (12.9)	18/31 (58)	394 (141)	1 (4)	25 (7.99)	2.5/5 (50)	Broadcast from another cluster
Digital	15 (10.7)	15/25 (60)	447 (116)	1 (4)	19.7 (9.83)	2/5 (40)	—
Inspection	5 (3.6)	5/18 (27)	350 (121)	0 (0)	3.88 (4.54)	2/5 (40)	n=3 (50%) participants at follow-up

[a]Not applicable.

XSL•FO
RenderX

Implementation

Fidelity: Was the Intervention Delivered as Intended?

On-Site Monitoring

The on-site monitoring team conducted 65 field visits to the intervention group participants' offices during their scheduled movement break sessions. During these visits, 35 movement breaks were observed, with 28 (80%) sessions featuring the Department of Medical Services theme song alongside another preferred Thai song. Importantly, there were no instances of cheating, such as merely shaking the Fitbit device without engaging in actual physical activity. However, 1 (8%) of the 13 participants from cluster 13 chose to take a 10-minute walk outside the office instead of participating in the team movement breaks, even when the sessions were not prompted. This participant managed to secure the weekly reward 3 times during the 24-week intervention period. Finally, no participants from the control group attended the sessions.

All alarm clocks were operational; nevertheless, leaders opted to broadcast the movement break songs using alternative devices, including the built-in broadcast system and office speakers. Notably, 2 (22%) of the 9 clusters did not independently initiate their movement break sessions; instead, they waited for another cluster to start the activity because they shared the same broadcast systems.

The movement break schedule reports submitted by the leaders were accurate; nonleaders who independently initiated sessions assisted with the reporting process. This collaboration ensured that the recorded data accurately reflected actual session initiation and participation. This also supports the finding that in some of the clusters (3/9, 33%), participants took the initiative to lead sessions without leader involvement, demonstrating shared responsibility in maintaining intervention fidelity.

Weekly rewards were distributed to winners by department directors. Participants received the team-based incentive on only 2 (8%) out of 24 occasions. A photograph capturing the reward ceremony, featuring face-to-face reward distribution, was shared with all intervention participants through Line OpenChat.

We received notifications regarding Fitbit issues, including syncing, freezing, and charging problems, both during and outside of our field visits. Fortunately, some participants with technological expertise were able to provide assistance, alleviating the need for the implementation team to respond to every notification. During our observations, we noticed that some of the participants (9/138, 6.5%) occasionally left their Fitbit devices at home when attending the movement break sessions. Despite understanding that their participation would not be recorded without the Fitbit, they still chose to engage in the sessions. In addition, 1 (0.7%) of the 138 participants was observed wearing the Fitbit only during the movement breaks.

All posters remained undamaged and visible, without any changes to their placement.

Due to the COVID-19 pandemic, a work-from-home policy was implemented. However, fewer than half of the participants (49/138, 35.5%) worked from home, with only 1 (11%) of the 9 clusters having a 100% work-from-home arrangement. This cluster used Zoom for meetings to conduct their movement breaks. Upon visiting the Zoom link twice, we found only 1 (0.7%) of the 138 participants in the Zoom meeting, with no movement breaks initiated. By contrast, another cluster coordinated their participation through a Line group chat when working from home, ensuring simultaneous session participation without requiring an online meeting.

Online Monitoring

We examined the correlations between engaging in movement breaks and daily Fitbit sedentary time and step count data. Each additional movement break was associated with a reduction of 6.20 (95% CI 6.99-5.41) minutes in sedentary time and an increase of 245 (95% CI 222-267) steps, adjusting for Fitbit wear time, number of public holidays in that week, and participants' age, sex, and education, with cluster and ID as random intercepts and intervention week number as a random slope.

This analysis, based on pedometer-measured outcomes, supported the fidelity of the movement break implementation because our prescribed minimum requirement for the duration of a single session was 4 minutes. Furthermore, the cumulative steps surpassed the eligibility criterion of 100 steps. The analysis also suggests successful data synthesis regarding movement break participation (Table 4).

Table 4. Associations between movement break participation on daily Fitbit sedentary time and step counts during the intervention period in the Physical Activity at Work cluster randomized trial.

	Model A		Model B	
	β[a] (95% CI)	*P* value	β[b] (95% CI)	*P* value
Sedentary time (min)	–6.32 (–7.06 to –5.57)	<.001	–6.20 (–6.99 to –5.41)	<.001
Steps (count)	263 (242 to 283)	<.001	245 (222 to 267)	<.001

[a]Linear mixed-effect model adjusted for Fitbit wear time, with cluster and ID as random intercepts and intervention week number as the random slope.
[b]Further adjusted for the number of public holidays in that week, age, gender, and education of the participants.

XSL•FO
RenderX

Dose: What Was the Extent of Intervention Delivery? (Online Monitoring)

According to the movement break leaders' weekly reports, many movement break sessions were never initiated by the leaders in each cluster. By the third week of the intervention, approximately 40% (72/180) of the scheduled movement break sessions had not been initiated (Figure S3 in Multimedia Appendix 2). Similarly, the overall average participation in movement breaks declined after the third week of the intervention, reducing to an average of 8 sessions per week per participant. Subsequently, the participation rate continued to decrease (Figure S4 in Multimedia Appendix 2).

We plotted the average daily Fitbit wear time to assess adherence during the intervention period (Figure S5 in Multimedia Appendix 2). We found that participants typically wore the devices for 10 to 15 hours per day at the start, with a slight decrease to 8 to 14 hours occurring 2 to 3 months into the intervention.

Mechanisms of Impact

Overview

Building on Figure 1, we present the effects of each component derived from a mixed methods analysis aimed at exploring the associations between each intervention component and movement break participation. We combine results from both quantitative and qualitative analyses to comprehensively address the question.

Quantitative Analysis

A total of 3200 participant-week data points were extracted by combining participants' demographics, Fitbit wear time, sedentary time and step count data, answers from the attitude-toward-intervention-components questionnaire, and online monitoring of rewards (Table 2).

Qualitative Analysis

Overview

There were 28 participants (n=3, 11% to n=6, 21% in each session) across 6 focus group discussions. The primary data analysts for the main results included 4 interviewers: the head of the data collection team (male), a program administrator (female), a researcher (male), and a staff member from the Health Intervention and Technology Assessment Program (a semiautonomous research unit under the Ministry of Public Health; female). Demographic details of all participants from the clusters (coded aby C12, C13, C5, C8, C16, and C17) involved in the focus group discussions are presented in Table S2 in Multimedia Appendix 2. Generally, the mean age of participants from the 3 top-ranked clusters was higher than that of participants from the bottom-ranked clusters. The majority of the participants (≥80%; 23/28, 82%) were female. Notably, the cluster with the lowest participation rate of 2.1% (10/480) comprised only 3 (2.2%) of the 138 participants at the 6-month follow-up (Table S2 in Multimedia Appendix 2).

How Did the Supporting Components Influence Participation in Movement Breaks?

Lottery-Based Incentives

Individuals who won the weekly lottery rewards the previous week demonstrated an 8.64% (95% CI 0.985%-16.3%) increase in movement break participation compared to other data points (Table 5). However, it is crucial to note the possibility of an overestimation when comparing data from winners against those from nonwinners. To address this, we conducted another subgroup analysis exclusively including winners, revealing a 5.1% (95% CI –3.44% to –13.6%) increase in movement break participation. However, this increase lacked statistical significance due to the small sample size (Table S3 in Multimedia Appendix 2).

Table 5. Effects of the Physical Activity at Work intervention components on weekly movement break participation percentage.

	β[a] (%; 95% CI)	Standardized β (%; 95% CI)	*P* value
Individual reward	8.64 (.985 to 16.3)	.0257 (.00293 to .0484)	.03
Cluster reward	–.325 (–2.64 to 1.99)	–.004 (–.0325 to .0245)	.78
Fitbit wear time	3.96 (2.28 to 5.65)	.0696 (.04 to .0993)	.001
Leadership support	1.82 (–2.25 to 5.9)	.0309 (–.0381 to .1)	.38
Enthusiastic	2 (–17.2 to 21.2)	0.077 (–.663 to .817)	.84
Encouraging	1.49 (–18.8 to 21.8)	0.0576 (–.724 to .84)	.89
Both enthusiastic and encouraging	24.1 (8.96 to 39.2)	0.929 (.346 to 1.51)	.002
Posters	4.49 (–.493 to 9.47)	0.0695 (–.00764 to .147)	.08

[a]Linear mixed effects model adjusted for the previous-week movement break participation percentage, number of public holidays in that week, age, sex, and education of the participants, with cluster and ID as random intercepts and intervention week number as the random slope.

During the focus group discussions, some of the participants suggested that rather than a single winner receiving a substantial cash prize, there should be multiple winners receiving more affordable rewards:

> *Rather than cash, it should be acknowledgments, such as showing who reach this many steps so we can compete with each other.* [C13]

> *[C]ould be something cheaper, don't have to be cash. Cheap shirts will do!* [C10]

XSL•FO
RenderX

> *[I]t was indeed a motivating factor, but the conditions for obtaining it should be somewhat more lenient. That is, to make it accessible to a broader audience, but the current conditions seem to be quite high. If it were reduced a bit, say to 100, I believe that providing rewards would be motivating enough.* [C8]

A participant in the worst-performing cluster did not remember the details of the financial incentive:

> *Sorry but how much was the reward?* [C17]

Team-Based Incentives

We observed no difference in movement break participation percentage (–0.325%, 95% CI –2.64% to 1.99%) in the weeks after a colleague from the same office won lottery rewards compared to other weeks. Contrary to our assumptions during the intervention development phase, peer pressure seems to have had no discernible effects on motivating movement break participation, as indicated by the quantitative analysis model (Table 5).

Different ideas emerged during the focus group discussions, as outlined in the following paragraphs.

Peer support was mentioned among members in the best-performing cluster as an important motivator:

> *Because it helped the team. If we dance, someone gets 1000THB, if we don't, it's 500THB only.* [C12]

However, others expressed discouragement:

> *When I see others got it and I never won for weeks, I was disheartened.* [C12]

Some of the participants viewed rewards as not motivating:

> *I joined the sessions because I want to be healthier. Rewards are meaningless for me.* [C13]

Fitbit Wear Time

There was a 3.96% (95% CI 2.28%-5.65%) increase in movement break participation among individuals with higher Fitbit wear time compared to those with lower wear time, using the median weekly wear time as the cutoff (Table 5). However, this increase may be attributed to motivation triggered by Fitbit notifications (eg, reminders to break prolonged sitting) or potentially to an information bias, wherein participants who wore their Fitbit devices for longer periods were detected more frequently in movement break participation.

To address potential bias, we compared the result with another model where Fitbit exposure was based on self-report, using the question "How often did you look at the Fitbit tracker during the last 2 weeks of the intervention period?" The analysis revealed that looking at the tracker more frequently (≥4 times/wk) was associated with a 1.97% (95% CI –1.42% to 5.36%) higher movement break participation percentage, without statistical significance (Table S4 in Multimedia Appendix 2).

Qualitative analysis indicated that participants perceived Fitbit as beneficial not necessarily for encouraging additional movement breaks but rather for motivating increased exercise and providing real-time feedback:

> *I joined short breaks a lot at first, but after a while, I forgot...However, I always wear the watch; look at the daily data.* [C13]
>
> *I set my goal to walk 10,000 steps a day after I got the watch, and I succeeded!* [C16]
>
> *[S]aying Fitbit encourages more movement breaks is wrong for me, but I feel the urge to exercise more from wearing the watch, like running after work.* [C17]

Leadership Support

Participants perceived the encouragement from directors as ineffective in motivating them to participate in more movement breaks.

Although office directors joined very few movement breaks, participants believed that their presence helped motivate everyone in the cluster:

> *He joined, I saw. But mostly he's busy.* [C16]
>
> *It made us stand up and dance...if we didn't it'd be awkward.* [C5]
>
> *We all danced every time our director was present (laugh).* [C8]

With regard to the movement break leaders' encouragement and enthusiasm, clusters with leaders who self-evaluated as more encouraging and enthusiastic had a significant increase of 24.1% (95% CI 8.88%-39.4%) in movement break participation compared to clusters with less encouraging and less enthusiastic leaders. However, because these exposure variables are self-reported by cluster leaders, the detected difference may be influenced by other cluster-specific factors.

Participants believed that movement break leaders' enthusiasm and encouragement helped them join more break sessions:

> *He was very active and always encourages everyone to stand up and dance.* [C8]
>
> *We were aware of the scheduled time, but if we were occupied, the leader would notify us.* [C13]

Nevertheless, some of the participants mentioned that they did not rely on leaders:

> *No matter how many people in the office, we danced. no leaders, no problem at all.* [C12]
>
> *I'm not sure...leaders always initiated the activity, but we did not always join.* [C16]
>
> *No one was available.* [C17]

Posters

Participants who reported that posters motivated them to engage in more movement breaks during the last 2 weeks of the intervention exhibited approximately a 5% higher participation rate, although this increase lacked statistical significance (Table 5). Notably, this increase may indicate participants' attitudes toward the intervention component rather than the direct effects of the posters themselves. Therefore, we compared the result by converting the variable to whether participants accurately identified the number of different styles of posters in their offices. The analysis revealed no significant difference in

XSL•FO
RenderX

movement break adherence between those who answered correctly and those who did not (standardized β=1.94, 95% CI –1.34 to 5.23; Table S5 in Multimedia Appendix 2).

Qualitative analysis showed that while the posters initially captured interest, over time, they failed to sustain attention:

> *At first I read them. I thought it was helpful and tried to follow some moves. After a while I just ignored them.* [C2]

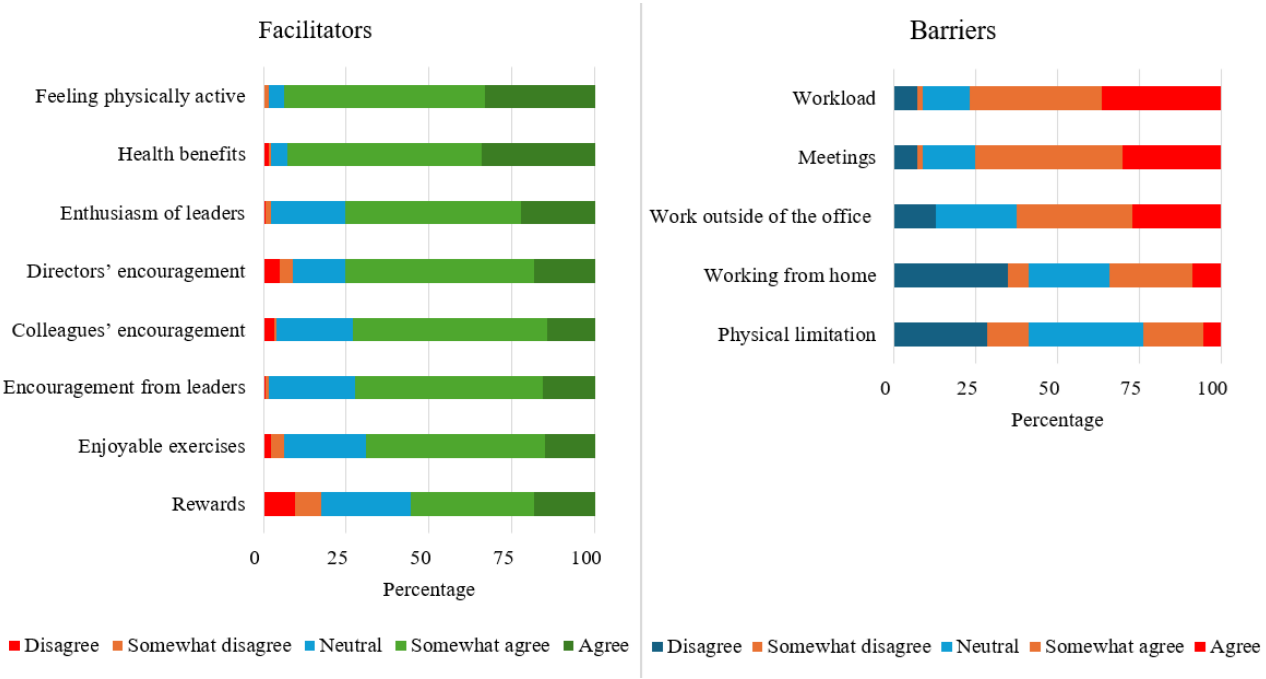
> *I didn't really read it that much...just walk past.* [C3]

The standardized β coefficients of all exposures indicate that the self-evaluated enthusiasm and encouragement of movement break leaders were the most important variables among those included, followed by Fitbit wear time and the individual reward (Table 5).

Facilitators and Barriers to Movement Break Participation

Figure 2 illustrates participants' attitudes regarding the facilitators and barriers to movement break participation. The top 2 facilitators were the positive feelings associated with being physically active and the perceived health benefits, followed by encouragement from leaders, directors, and colleagues. The appeal of engaging in enjoyable exercises also motivated their participation. Interestingly, weekly rewards ranked lowest among the facilitators.

Figure 2. Facilitators and barriers to movement break participation in the Physical Activity at Work cluster randomized trial.



Conversely, the primary barrier perceived by participants was their workload, followed by meetings and working outside of the office. Surprisingly, many participants did not view working from home as a barrier to their participation.

Thematic analysis revealed positive attitudes toward intervention components:

> *I think having Fitbit is a really good motivation to move.* [C17]

> *Previously when someone invited me to run, I wanted to but it was hard. This [movement breaks] is easier and fun.* [C12]

Participants also noted barriers to the intervention, such as the lottery reward design:

> *It [individual reward] somehow motivated me, but I guess I just could not do as good as others.* [C16]

Another important barrier was the monotonous design of the movement breaks:

> *Dancing with the same moves gets boring after a while.* [C8]

> *I think we can change the songs to make it more interesting.* [C16]

Table 6 presents thematic analysis results with references under each theme and subtheme from all 6 focus group discussions.

XSL•FO
RenderX

Table 6. Thematic analysis of focus group discussions on the facilitators and barriers to movement break participation in the Physical Activity at Work cluster randomized trial.

Themes and subthemes	Quotes
Facilitators	
Psychological capability	• "Previously when someone invited me to run, I wanted to, but it was hard. This is easier and fun." (C12)
Social opportunity	• "Both songs and moves...it was like mass hysteria; others enjoyed the sessions and I wanted to join" (C12) • "Our director joined, and we danced happily" (C8) • "When my boss dance and I just sat there working, it felt strange, so I joined" (C5)
Physical opportunity	• "I'm happy that there is a new opportunity for me to exercise...in workplace. Normally I don't have time." (C12) • "Our office is spacious. There are empty spaces for dancing." (C5)
Reflective motivation	• "My personal goal was actually health, not rewards" (C13) • "I set my goal to get the reward and lose weight and cholesterol level because the programme measured those" (C5)
Automatic Motivation	• "It was relaxing, both mind and body...especially the mind." (C17) • "Just that we got to dance...when the session started, it relieved the stress quite a lot" (C5)
Barriers	
Psychological capability	• "10 min. dancing is too short. 30 min. working out is better for your health for the day." (C13) • "My programming work never allows me to lose focus" (C16)
Social opportunity	• "Meetings definitely prevented us to join movement sessions" (C8) • "We were not really motivated by directors and colleagues" (C17)
Physical opportunity	• "Sometimes workspace is not wide enough...also I was afraid I'll annoy others who were not in the project" (C13) • "During high workload, we cannot join" (C5) • "The high workload never allows us to do anything else" (C17) • "Ever since COVID situation got worse, I've been sitting all the time at my desk. It's the workload, can't do anything else." (C16)
Reflective motivation	• "Rewards could be anything, like shoes, incentives don't need to be money, like shirts..." (C5)
Automatic motivation	• "Each programme shouldn't last long. I mean, we should always change the stimulant to avoid boredom..." (C13) • "Dancing with the same moves gets boring after a while" (C8) • "I think we can change the songs to make it more interesting." (C16)

Discussion

The PAW cluster randomized trial of a multicomponent intervention, developed based on the socioecological framework [9], showed a decrease in participants' waking sedentary time and an increase in moderate to vigorous physical activity, although without statistical significance [20]. We conducted a mixed methods process evaluation, following the Medical Research Council's guidance [11], to comprehensively describe the PAW trial's (1) recruitment and context, (2) implementation, and (3) impact mechanisms.

Recruitment and Context

Our team anticipated a high recruitment rate once the office directors approved the inclusion of their offices in the trial, given the bureaucratic nature of the ministry organization in Thailand. However, recruitment rates were low in many clusters, weakening the trial's statistical power and risking reduced intervention adherence, particularly in the social-level component.

Cluster 17 serves as a notable example, demonstrating minimal engagement in movement breaks (Table 3; Figure S4 in Multimedia Appendix 2). This could be attributed to the recruitment of too few, relatively young participants alongside a high number of nonparticipants (Tables 3 and 4). Despite its potential for easy qualification for the team-based incentive with just 3 participants (at follow-up), cluster 17 might have faced challenges due to the negative influence of an environment where breaks are discouraged. This aligns with a previous study reporting reduced break taking in disapproving work environments compared to more supportive ones [23]. Cluster 13 also faced challenges with a low recruitment rate; yet, the engagement of the participants in this cluster in movement breaks was high. Two reasons might explain this: (1) the participant who chose to walk outside the office and secured 3 weekly rewards increased the cluster's average participation

XSL•FO
RenderX

rate, and (2) recruiting at least 13 participants could create an environment conducive to making movement breaks feel enjoyable and secure to participate in.

Job characteristics might also play a role in movement break participation. Cluster 17, "the inspection office," frequently required its members to leave the office for inspections, which was reflected in their low sedentary time at baseline (Table 3) and echoed in the focus group interviews. In addition, cluster 16 represented a digital office where most participants worked on their laptop computers all day. While this setting might seem ideal for implementing the intervention, participants encountered challenges with breaks because they needed continual focus to code. This aligns with the suggestion that identifying the optimal fit between the organizational context and the intervention is crucial [24]. However, challenges remain when implementing such interventions across organizations with diverse offices, each characterized by unique job contexts and work styles.

Implementation

We implemented the trial with a comprehensive dose delivery, ensuring intervention fidelity by closely adhering to the original plan for each component. Moreover, the findings in Table 3 affirm the application of precise criteria for enumerating movement breaks, underscoring their correlations with the expected reductions in sedentary time and increases in step counts. Nevertheless, we encountered challenges related to the frequency of movement break sessions conducted by participants, which we refer to as "dose received" [13]. This inconsistency may be attributed to the performance of the movement break leaders, as illustrated in Figure S3 in Multimedia Appendix 2, which shows that many sessions were never initiated. However, several clusters relied on automatic timing mechanisms to initiate breaks, ensuring that sessions commenced even without leaders present. We discuss this further in the next subsection.

Mechanisms of Impact

Testing the Designed Intervention Components

Healthy behaviors are maximized when environments and policies support healthful choices, and individuals are motivated and educated to make them [25]. As one of the main principles of the ecological model, the interaction of influences means that variables within the system work together synergistically [26]. Hence, our study examined the impact of the supporting components on the movement break participation using multivariate linear mixed models (Table 4).

The enthusiasm and encouragement of the movement break leaders seem to have contributed to higher participation percentages. However, the cluster that exhibited the best performance challenged this observation, asserting that the members initiated sessions independently, even without leaders. Unlike other cluster randomized trials where exercise sessions were led by nonparticipants such as physiotherapists [16], the movement break leaders in our study were participants and could be replaced if absent. Hence, without the leaders, other members could initiate sessions automatically. Nevertheless, the influence of the leaders' encouragement and enthusiasm likely played a role in motivating the rest of the team. Prior research has also found that workplace team leaders play a significant role in facilitating the implementation of workplace interventions [27,28]. Moreover, a comprehensive review emphasizes that effective team performance underscores the fulfillment of leadership styles, supportive team behaviors, communication, and performance feedback [29].

The impact of the team-based incentive on movement break participation was evidently negligible. Regarding individual lottery-based incentives, their influence remains somewhat unclear. Systematic reviews suggest that financial rewards for physical activity have positive short-term effects, surpassing unconditional incentives [30,31]. Moreover, another study indicates that increasing reward values may lead to improved results [32]. Nevertheless, insights from online monitoring data and focus group discussions revealed that only 1 (11%) of the 9 clusters actively pursued rewards and engaged in friendly competition, while the others (8/9, 89%) were indifferent. Awarding only 1 winner per week might have demotivated participants who faithfully adhered to the intervention but never won. This phenomenon can be explained using goal-setting theory, which suggests that setting clear and appropriately challenging goals is crucial [33]. By contrast, setting unachievable goals may induce stress, anxiety, and perceived pressure [34]. Therefore, overly difficult goals, such as securing the weekly lottery reward, may have discouraged participants.

Fitbit was perceived as a helpful tool for real-time data monitoring and might help motivate leisure physical activity. However, current evidence indicates unfavorable outcomes regarding the effectiveness of pedometers in increasing physical activity within the workplace or in motivating sedentary breaks [35,36]. The component may have primarily served as a data collection tool rather than actively supporting movement breaks. By contrast, the posters, which served as the environmental component, proved ineffective and should be replaced with physical environment interventions, which have been shown to be more effective [4,8].

Finally, leadership support was considered helpful at the start of the intervention but provided no lasting effect over time. The component could be perceived as a nudge, which has been widely used to prompt behavioral change among participants by providing alternative options to sedentary habits in the workplace [37]. Leadership support was expected to facilitate these behavioral shifts. Management support has also been found to be an essential enabler for workplace intervention participation to reduce sitting time [27,38,39]. However, the sustainability of the effects depends on the meticulous design of the intervention component. Future studies evaluating long-term effects must enhance the intervention component to ensure sustainability.

Facilitators and Barriers

Automatic motivation, driven by the desire to feel active and relaxed, was ranked as the top facilitator of movement break participation, followed by reflective motivation, including perceived health benefits. The findings are in line with a recent systematic review [39]. Another review also supports the idea that microbreaks at work can improve well-being by boosting vigor and reducing fatigue [40]. In addition, participants

XSL•FO
RenderX

generally grasped the concept of health benefits associated with movement breaks. However, some of the individuals compared movement breaks to aerobic exercise, expressing doubt regarding their health benefits.

Workload and meetings were perceived as the main barriers to movement breaks, which aligns with a previous study's prediction that high workloads, although positively related to the desire to detach from work, would also deter employees from actually taking breaks [23]. In addition, recent findings from a systematic review supported the notion that microbreaks increase well-being but do not necessarily improve work performance, especially in tasks that require high cognitive engagement. Moreover, the review suggested that breaks of >10 minutes may be necessary to enhance work performance [40].

By contrast, meetings presented clear barriers to our intervention design. Future research should be dedicated to advocating feasible and context-specific active meetings within an active workplace; for example, an exploratory study found that standing meetings were feasible, well-received by employees, and may reduce sitting time among the population [41]. However, widespread adoption faces obstacles due to prevailing sedentary work cultures and concerns about self-consciousness in front of senior staff, highlighting the need for broader social behavior change initiatives [42].

We hypothesized that movement break participation might be low due to the work arrangement during the peaks of the COVID-19 pandemic [20]. However, participants did not consider working from home to be the main barrier to movement break participation. Instead, they thought that the COVID-19 pandemic did not significantly hinder their adherence to the intervention due to the work-from-home policy. Instead, they attributed the difficulty to increased workload. Nevertheless, working from home remained 1 of the barriers to movement break participation. This is in line with another workplace cluster randomized trial, which reported that engaging in physical exercise with colleagues during working hours was more effective than home-based exercise in enhancing vitality and managing pain-related concerns among health care workers [43].

Strengths and Limitations

We conducted a comprehensive mixed methods process evaluation of the PAW multicomponent intervention, covering recruitment and context, implementation, and impact mechanisms. Rigorous analyses were applied to both quantitative and qualitative data. The data are prospective, spanning 6 months of follow-up. Although the overall results show no significant impact of the intervention, this process evaluation offers insights that could be crucial for the future development and evaluation of intervention packages aimed at reducing sedentary time while improving physical activity levels.

Nevertheless, our study had some limitations. First, we could not test mediators to understand the underlying mechanisms through which one variable influences another due to the trial's lack of efficacy [20]. Second, constructing intervention theories will involve reconstructing a logic model [11,44], a task we plan to undertake in future studies. Third, we could not complete the management and analysis of process evaluation data before the conclusion of the PAW trial. As a result, the process evaluation was conducted post hoc to elucidate the trial outcomes. Furthermore, our evaluation only incorporated participants' data and perspectives, overlooking input from other stakeholders such as organizational directors and nonparticipants. Finally, the team members responsible for process evaluation were also involved in the outcome evaluation. While our team members possess the most comprehensive understanding of the trial details, potential bias in interpretations must be acknowledged [11].

Conclusions

The PAW trial did not significantly reduce sedentary time among Thai office workers. Although the trial implementation was satisfactory regarding dose delivery and fidelity, there was limited uptake of the movement breaks, the key intervention component. This limited uptake could be attributed to (1) context-related challenges (including jobs requiring high cognitive engagement or frequent out-of-office work and meetings), (2) the absence of goal-setting aspects in the detailed design of individual and social components, (3) the lack of effective and sustainable supporting components at the environmental and organizational levels, and (4) elevated workloads in specific clusters (exacerbated during peak periods of the COVID-19 pandemic) serving as a significant barrier.

Acknowledgments

The authors thank Buppa Pardang, Pimonphan Thongoun, Saudamini Dabak, Kewalin Chomrenoo, Nachawish Kittibovorndit, and Budsadee Soboon for their help with the focus group discussions, further helpful discussions, and administrative support that contributed to this work. The trial was funded by revenue from taxes on tobacco and alcohol via the Thai Health Promotion Foundation. This study was supported by the Singapore Ministry of Health's National Medical Research Council under its Population Health Research Grant (PHRGOC22Jul-0020, ID: MOH-001315). The Health Intervention and Technology Assessment Program's International Unit is supported by the International Decision Support Initiative to provide technical assistance on health intervention and technology assessment to governments in low- and middle-income countries. RN and TR are supported by the Japan Society for the Promotion of Science Core-to-Core Program (JPJSCCB20200002). The funders had no role in the study design, data collection or analysis, the preparation of the manuscript, or the decision to publish. The views and opinions expressed in this paper are those of the authors and do not necessarily reflect those of the funders or the institutions to which the authors belong.

XSL•FO
RenderX

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors contributed to the study design or delivery of the trial. KA was the principal investigator of the trial. FM-R and CC provided statistical expertise and data management. TR, RN, YT, WI, and FM-R provided expertise on conceptualizing the process evaluation aspects and the narrative of the manuscript. RN and TR provided expertise on behavioral economics, such as the impact of lottery-based and team-based incentives. TR and FM-R provided expertise on physical activity promotion theories and frameworks. KA, YT, and CC drafted the manuscript. All authors reviewed the manuscript draft and read and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional questions for intervention participants.
[DOCX File , 740 KB - formative_v9i1e57604_app1.docx]

Multimedia Appendix 2
Supplementary figures and tables.
[DOCX File , 15 KB - formative_v9i1e57604_app2.docx]

References

1. Buckingham SA, Williams AJ, Morrissey K, Price L, Harrison J. Mobile health interventions to promote physical activity and reduce sedentary behaviour in the workplace: a systematic review. Digit Health 2019 Mar 27;5:2055207619839883 [FREE Full text] [doi: 10.1177/2055207619839883] [Medline: 30944728]
2. Zhou L, Deng X, Xu M, Wu Y, Shang X, E F, et al. The effects of active workstations on reducing work-specific sedentary time in office workers: a network meta-analysis of 23 randomized controlled trials. Int J Behav Nutr Phys Act 2023 Jul 27;20(1):92 [FREE Full text] [doi: 10.1186/s12966-023-01467-5] [Medline: 37501138]
3. Mulchandani R, Chandrasekaran AM, Shivashankar R, Kondal D, Agrawal A, Panniyammakal J, et al. Effect of workplace physical activity interventions on the cardio-metabolic health of working adults: systematic review and meta-analysis. Int J Behav Nutr Phys Act 2019 Dec 19;16(1):134 [FREE Full text] [doi: 10.1186/s12966-019-0896-0] [Medline: 31856826]
4. Chu AH, Ng SH, Tan CS, Win AM, Koh D, Müller-Riemenschneider F. A systematic review and meta-analysis of workplace intervention strategies to reduce sedentary time in white-collar workers. Obes Rev 2016 May;17(5):467-481. [doi: 10.1111/obr.12388] [Medline: 26990220]
5. To QG, Chen TT, Magnussen CG, To KG. Workplace physical activity interventions: a systematic review. Am J Health Promot 2013;27(6):e113-e123. [doi: 10.4278/ajhp.120425-LIT-222] [Medline: 23631453]
6. Dishman RK, Oldenburg B, O'Neal H, Shephard RJ. Worksite physical activity interventions. Am J Prev Med 1998 Nov;15(4):344-361. [doi: 10.1016/s0749-3797(98)00077-4] [Medline: 9838977]
7. Proper KI, Koning M, van der Beek AJ, Hildebrandt VH, Bosscher RJ, van Mechelen W. The effectiveness of worksite physical activity programs on physical activity, physical fitness, and health. Clin J Sport Med 2003 Mar;13(2):106-117. [doi: 10.1097/00042752-200303000-00008] [Medline: 12629429]
8. Zhu X, Yoshikawa A, Qiu L, Lu Z, Lee C, Ory M. Healthy workplaces, active employees: a systematic literature review on impacts of workplace environments on employees' physical activity and sedentary behavior. Built Environ 2020 Jan;168:106455. [doi: 10.1016/j.buildenv.2019.106455]
9. Sallis J, Owen N, Fisher EB. Ecological models of health behavior. In: Glanz BK, K V, editors. Health Behavior and Health Education: Theory, Research, and Practice. Hoboken, NJ: Jossey-Bass; 2008.
10. Grant A, Treweek S, Dreischulte T, Foy R, Guthrie B. Process evaluations for cluster-randomised trials of complex interventions: a proposed framework for design and reporting. Trials 2013 Jan 12;14:15 [FREE Full text] [doi: 10.1186/1745-6215-14-15] [Medline: 23311722]
11. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: medical research council guidance. BMJ 2015 Mar 19;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
12. Oakley A, Strange V, Bonell C, Allen E, Stephenson J, RIPPLE Study Team. Process evaluation in randomised controlled trials of complex interventions. BMJ 2006 Feb 18;332(7538):413-416 [FREE Full text] [doi: 10.1136/bmj.332.7538.413] [Medline: 16484270]

XSL•FO
RenderX

13. Steckler A, Linnan L. Process Evaluation for Public Health Interventions and Research. San Francisco, CA: Jossey-Bass; 2022.
14. Liu H, Andersson G, Manchaiah V. Editorial: the process evaluation of clinical trials. Front Med (Lausanne) 2022 Jul 11;9:950637 [FREE Full text] [doi: 10.3389/fmed.2022.950637] [Medline: 35899205]
15. Ellard DR, Taylor SJ, Parsons S, Thorogood M. The OPERA trial: a protocol for the process evaluation of a randomised trial of an exercise intervention for older people in residential and nursing accommodation. Trials 2011 Feb 02;12:28 [FREE Full text] [doi: 10.1186/1745-6215-12-28] [Medline: 21288341]
16. Ellard DR, Thorogood M, Underwood M, Seale C, Taylor SJ. Whole home exercise intervention for depression in older care home residents (the OPERA study): a process evaluation. BMC Med 2014 Jan 03;12:1 [FREE Full text] [doi: 10.1186/1741-7015-12-1] [Medline: 24386900]
17. Petrunoff N, Yao J, Sia A, Ng A, Ramiah A, Wong M, et al. Activity in nature mediates a park prescription intervention's effects on physical activity, park use and quality of life: a mixed-methods process evaluation. BMC Public Health 2021 Jan 22;21(1):204 [FREE Full text] [doi: 10.1186/s12889-021-10177-1] [Medline: 33482787]
18. Tuangratananon T. Differences in physical activity levels between urban and rural adults in Thailand: findings from the 2015 National Health and Welfare Survey. Health Syst Res 2018;12(2018):27-41 [FREE Full text]
19. Chen C, Dieterich AV, Koh JJ, Akksilp K, Tong EH, Budtarad N, et al. The physical activity at work (PAW) study protocol: a cluster randomised trial of a multicomponent short-break intervention to reduce sitting time and increase physical activity among office workers in Thailand. BMC Public Health 2020 Sep 01;20(1):1332 [FREE Full text] [doi: 10.1186/s12889-020-09427-5] [Medline: 32873258]
20. Akksilp K, Koh JJ, Tan V, Tong EH, Budtarad N, Xueying G, et al. The physical activity at work (PAW) study: a cluster randomised trial of a multicomponent short-break intervention to reduce sitting time and increase physical activity among office workers in Thailand. Lancet Reg Health Southeast Asia 2022 Oct 19;8:100086 [FREE Full text] [doi: 10.1016/j.lansea.2022.100086] [Medline: 37384135]
21. Owen N, Sugiyama T, Eakin EE, Gardiner PA, Tremblay MS, Sallis JF. Adults' sedentary behavior determinants and interventions. Am J Prev Med 2011 Aug;41(2):189-196. [doi: 10.1016/j.amepre.2011.05.013] [Medline: 21767727]
22. The Health and Welfare Survey 2015. National Statistical Office of Thailand. 2023. URL: https://www.nso.go.th/nsoweb/storage/survey_detail/2023/20230505165728_20473.pdf [accessed 2024-12-17]
23. Phan V, Beck JW. Why do people (not) take breaks? An investigation of individuals' reasons for taking and for not taking breaks at work. J Bus Psychol 2023;38(2):259-282 [FREE Full text] [doi: 10.1007/s10869-022-09866-4] [Medline: 36593864]
24. Nielsen K, Taris TW, Cox T. The future of organizational interventions: addressing the challenges of today's organizations. Work Stress 2010 Oct 20;24(3):219-233. [doi: 10.1080/02678373.2010.519176]
25. Ottawa charter for health promotion. World Health Organization. 2012 Jun 16. URL: https://www.who.int/publications/i/item/WH-1987 [accessed 2024-10-31]
26. Glanz K, Viswanath K, Rimer BK. Health Behavior and Health Education: Theory, Research, and Practice. Hoboken, NJ: John Wiley & Sons; 2008.
27. Goode AD, Hadgraft NT, Neuhaus M, Healy GN. Perceptions of an online 'train-the-champion' approach to increase workplace movement. Health Promot Int 2019 Dec 01;34(6):1179-1190. [doi: 10.1093/heapro/day092] [Medline: 30452649]
28. Hopkins JM, Glenn BA, Cole BL, McCarthy W, Yancey A. Implementing organizational physical activity and healthy eating strategies on paid time: process evaluation of the UCLA WORKING pilot study. Health Educ Res 2012 Jun;27(3):385-398 [FREE Full text] [doi: 10.1093/her/cys010] [Medline: 22323279]
29. Salcinovic B, Drew M, Dijkstra P, Waddington G, Serpell BG. Factors influencing team performance: what can support teams in high-performance sport learn from other industries? A systematic scoping review. Sports Med Open 2022 Feb 22;8(1):25 [FREE Full text] [doi: 10.1186/s40798-021-00406-7] [Medline: 35192078]
30. Barte JC, Wendel-Vos GC. A systematic review of financial incentives for physical activity: the effects on physical activity and related outcomes. Behav Med 2017;43(2):79-90. [doi: 10.1080/08964289.2015.1074880] [Medline: 26431076]
31. Mitchell MS, Goodman JM, Alter DA, John LK, Oh PI, Pakosh MT, et al. Financial incentives for exercise adherence in adults: systematic review and meta-analysis. Am J Prev Med 2013 Nov;45(5):658-667. [doi: 10.1016/j.amepre.2013.06.017] [Medline: 24139781]
32. Finkelstein EA, Linnan LA, Tate DF, Birken BE. A pilot study testing the effect of different levels of financial incentives on weight loss among overweight employees. J Occup Environ Med 2007 Sep;49(9):981-989. [doi: 10.1097/JOM.0b013e31813c6dcb] [Medline: 17848854]
33. Latham GP, Locke EA. Self-regulation through goal setting. Organ Behav Hum Decis Process 1991 Dec;50(2):212-247. [doi: 10.1016/0749-5978(91)90021-k]
34. Latham GP, Locke EA. Enhancing the benefits and overcoming the pitfalls of goal setting. Organ Dyn 2006;35(4):332-340. [doi: 10.1016/j.orgdyn.2006.08.008]
35. Harrison J, O'Donoghue A, Hill J. The use of pedometers in the workplace to increase physical activity. Br J Card Nurs 2021 Aug 02;16(8):1-4 [FREE Full text] [doi: 10.12968/bjca.2021.0066] [Medline: 38125904]

XSL•FO
RenderX

36. Freak-Poli R, Cumpston M, Albarqouni L, Clemes SA, Peeters A. Workplace pedometer interventions for increasing physical activity. Cochrane Database Syst Rev 2020 Jul 21;7(7):CD009209. [doi: 10.1002/14651858.CD009209.pub3] [Medline: 32700325]
37. Forberger S, Wichmann F, Comito CN. Nudges used to promote physical activity and to reduce sedentary behaviour in the workplace: results of a scoping review. Prev Med 2022 Feb;155:106922. [doi: 10.1016/j.ypmed.2021.106922] [Medline: 34933021]
38. Mackenzie K, Such E, Norman P, Goyder E. Sitting less at work: a qualitative study of barriers and enablers in organisations of different size and sector. BMC Public Health 2019 Jul 04;19(1):884 [FREE Full text] [doi: 10.1186/s12889-019-7148-8] [Medline: 31272484]
39. Mackenzie K, Such E, Norman P, Goyder E. The development, implementation and evaluation of interventions to reduce workplace sitting: a qualitative systematic review and evidence-based operational framework. BMC Public Health 2018 Jul 04;18(1):833 [FREE Full text] [doi: 10.1186/s12889-018-5768-z] [Medline: 29973188]
40. Albulescu P, Macsinga I, Rusu A, Sulea C, Bodnaru A, Tulbure BT. "Give me a break!" A systematic review and meta-analysis on the efficacy of micro-breaks for increasing well-being and performance. PLoS One 2022 Aug 31;17(8):e0272460 [FREE Full text] [doi: 10.1371/journal.pone.0272460] [Medline: 36044424]
41. Danquah IH, Tolstrup JS. Standing meetings are feasible and effective in reducing sitting time among office workers-walking meetings are not: mixed-methods results on the feasibility and effectiveness of active meetings based on data from the "take a stand!" study. Int J Environ Res Public Health 2020 Mar 05;17(5):1713 [FREE Full text] [doi: 10.3390/ijerph17051713] [Medline: 32151044]
42. Biddle SJ, O'Connell SE, Davies MJ, Dunstan D, Edwardson CL, Esliger DW, et al. Reducing sitting at work: process evaluation of the SMArT Work (Stand More At Work) intervention. Trials 2020 May 13;21(1):403 [FREE Full text] [doi: 10.1186/s13063-020-04300-7] [Medline: 32404181]
43. Jakobsen MD, Sundstrup E, Brandt M, Andersen LL. Psychosocial benefits of workplace physical exercise: cluster randomized controlled trial. BMC Public Health 2017 Oct 10;17(1):798 [FREE Full text] [doi: 10.1186/s12889-017-4728-3] [Medline: 29017479]
44. Rush B, Ogborne A. Program logic models: expanding their role and structure for program planning and evaluation. Can J Program Eval 1991;6(2):95-106. [doi: 10.3138/cjpe.6.005]

Edited by A Mavragani; submitted 23.02.24; peer-reviewed by E Breckin; comments to author 09.09.24; revised version received 29.09.24; accepted 10.10.24; published 02.01.25.

Please cite as:
Akksilp K, Rouyard T, Isaranuwatchai W, Nakamura R, Müller-Riemenschneider F, Teerawattananon Y, Chen C
The Physical Activity at Work (PAW) Program in Thai Office Workers: Mixed Methods Process Evaluation Study
JMIR Form Res 2025;9:e57604
URL: https://formative.jmir.org/2025/1/e57604
doi:10.2196/57604
PMID:

©Katika Akksilp, Thomas Rouyard, Wanrudee Isaranuwatchai, Ryota Nakamura, Falk Müller-Riemenschneider, Yot Teerawattananon, Cynthia Chen. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

ChatGPT's Attitude, Knowledge, and Clinical Application in Geriatrics Practice and Education: Exploratory Observational Study

Huai Yong Cheng[1], MPH, MS, MD

Minneapolis VA Health Care System, Minneapolis, MN, United States

Corresponding Author:
Huai Yong Cheng, MPH, MS, MD
Minneapolis VA Health Care System
1 Veterans Dr.
Minneapolis, MN, 55417
United States
Phone: 1 6124672051
Fax: 1 6127252084
Email: wchengwcheng@gmail.com

Abstract

Background: The increasing use of ChatGPT in clinical practice and medical education necessitates the evaluation of its reliability, particularly in geriatrics.

Objective: This study aimed to evaluate ChatGPT's trustworthiness in geriatrics through 3 distinct approaches: evaluating ChatGPT's geriatrics attitude, knowledge, and clinical application with 2 vignettes of geriatric syndromes (polypharmacy and falls).

Methods: We used the validated University of California, Los Angeles, geriatrics attitude and knowledge instruments to evaluate ChatGPT's geriatrics attitude and knowledge and compare its performance with that of medical students, residents, and geriatrics fellows from reported results in the literature. We also evaluated ChatGPT's application to 2 vignettes of geriatric syndromes (polypharmacy and falls).

Results: The mean total score on geriatrics attitude of ChatGPT was significantly lower than that of trainees (medical students, internal medicine residents, and geriatric medicine fellows; 2.7 vs 3.7 on a scale from 1-5; 1=*strongly disagree*; 5=*strongly agree*). The mean subscore on positive geriatrics attitude of ChatGPT was higher than that of the trainees (medical students, internal medicine residents, and neurologists; 4.1 vs 3.7 on a scale from 1 to 5 where a higher score means a more positive attitude toward older adults). The mean subscore on negative geriatrics attitude of ChatGPT was lower than that of the trainees and neurologists (1.8 vs 2.8 on a scale from 1 to 5 where a lower subscore means a less negative attitude toward aging). On the University of California, Los Angeles geriatrics knowledge test, ChatGPT outperformed all medical students, internal medicine residents, and geriatric medicine fellows from validated studies (14.7 vs 11.3 with a score range of –18 to +18 where +18 means that all questions were answered correctly). Regarding the polypharmacy vignette, ChatGPT not only demonstrated solid knowledge of potentially inappropriate medications but also accurately identified 7 common potentially inappropriate medications and 5 drug-drug and 3 drug-disease interactions. However, ChatGPT missed 5 drug-disease and 1 drug-drug interaction and produced 2 hallucinations. Regarding the fall vignette, ChatGPT answered 3 of 5 pretests correctly and 2 of 5 pretests partially correctly, identified 6 categories of fall risks, followed fall guidelines correctly, listed 6 key physical examinations, and recommended 6 categories of fall prevention methods.

Conclusions: This study suggests that ChatGPT can be a valuable supplemental tool in geriatrics, offering reliable information with less age bias, robust geriatrics knowledge, and comprehensive recommendations for managing 2 common geriatric syndromes (polypharmacy and falls) that are consistent with evidence from guidelines, systematic reviews, and other types of studies. ChatGPT's potential as an educational and clinical resource could significantly benefit trainees, health care providers, and laypeople. Further research using GPT-4o, larger geriatrics question sets, and more geriatric syndromes is needed to expand and confirm these findings before adopting ChatGPT widely for geriatrics education and practice.

(JMIR Form Res 2025;9:e63494) doi:10.2196/63494

XSL•FO
RenderX

KEYWORDS

ChatGPT; geriatrics attitude; ageism; geriatrics competence; geriatric syndromes; polypharmacy; falls; aging, older adults

Introduction

Background

ChatGPT stands for Chat Generative Pretrained Transformer and was developed by an artificial intelligence (AI) research company, OpenAI. It is an AI chatbot technology that can process our natural human language and generate a response. ChatGPT was released to the public by OpenAI on November 30, 2022, and surpassed 1 million users in just 5 days. This performance set a record for the second fastest-growing user base after that of Threads. ChatGPT currently has >180 million users (as of October 26, 2024) [1]. ChatGPT users have increased exponentially, and it has been increasingly used in medicine, resulting in 4625 publications on PubMed (as of October 26, 2024) [2]. Notably, ChatGPT was not programmed originally to be used for medical education and clinical application [1] despite reports of its wide application for medical education [3-7] and clinical practice [5-12]. ChatGPT has been studied in multiple specialties and subspecialties, such as psychiatry [13], radiology [14], and cardiology [15]. Its rapid progress in such a brief time has raised challenges, concerns, and limitations, including privacy, ethnic and racial bias, and legal risks [16-21].

Any application of ChatGPT to medical education and clinical practice could be promising but needs to be well designed and investigated. In the context of a growing aging population, geriatrics education, geriatrics workforces, and geriatrics practice have been facing significant challenges for the last 4 decades [22-26]. ChatGPT could offer an opportunity to improve geriatrics education and clinical practice [27-30]. Preliminary studies have shown promise [27-30]. For example, older adults often have polypharmacy issues [31], and providers review and deprescribe medications [32]. A specifically trained large language model may provide useful clinical support in polypharmacy management for primary care physicians [29]. ChatGPT has scored better on general and theoretical questions in geriatrics than on complex decisions and end-of-life situations, with the lowest scores related to diagnosis and performing complex tests [30]. Overall, the application of ChatGPT to geriatrics education and clinical practice is sporadic and needs to be expanded and further investigated. Therefore, this study was designed to determine whether we can trust ChatGPT to be used in geriatrics education and clinical practice using 3 distinct approaches.

The first approach was to evaluate any age-biased outputs generated by ChatGPT. The aging narrative spans >210 years [33]. The term ageism, coined by Robert Butler in 1969, refers to age-based stereotyping, prejudice, and discrimination [34]. Self-perception of aging is another emergent concept related to ageism [35]. Ageism is global and prevalent in daily life [34,36,37], health care systems [34,38], and the news media [39,40]. It is present in various medicine specialties such as oncology [41] and cardiology [42] and was persistent during the COVID-19 pandemic [43,44]. Ageism is significantly associated with poor health outcomes and quality of life [34,35]. Interventions to reduce ageism have been widely studied [45,46]. ChatGPT has been found to exhibit privacy, ethnicity, gender, and racial bias and entail legal risks [16-21]. For example, ChatGPT recommends fewer female than male ophthalmologists [47]. However, whether ChatGPT generates age-biased outputs has not been reported. This study aimed to demonstrate whether ChatGPT exhibits ageism by testing its geriatrics attitude using a validated University of California, Los Angeles (UCLA) geriatrics attitude instrument [48,49] and comparing its results with those obtained by medical students residents, and geriatrics medicine fellows from published articles

The second approach was to evaluate the geriatrics knowledge of ChatGPT. ChatGPT has passed the United States Medical Licensing Examination (USMLE) [50,51] and many other medicine exams [52-56], demonstrating its competence in medical knowledge. However, it has been less studied in geriatrics. For example, in one Geriatrics and AI in Spain project in Spain [30], ChatGPT was prompted with 10 questions about geriatric medicine. Compared to 130 physicians who answered the questionnaire, ChatGPT scored better on general and theoretical questions than on complex decisions and end-of-life situations, with the lowest scores for diagnosis and performing complex tests [30]. AI is likely to be incorporated into some areas of geriatric medicine, but it still presents significant limitations, mainly in complex medical decision-making [30]. This study was designed to demonstrate the geriatrics competency of ChatGPT by examining its performance on the validated UCLA geriatrics knowledge test [48,49,57] and comparing it with that of medical students, internal medicine residents, and geriatric medicine fellows from previously published articles in the literature.

The third approach was to evaluate ChatGPT's knowledge of 2 geriatric syndromes and its clinical application to them (polypharmacy and falls). Multiple previous studies have shown that ChatGPT performs well on clinical vignette questions [58-64]. For example, ChatGPT achieved 71.7% accuracy overall on 36 clinical vignettes from the Merck Sharpe and Dohme clinical manuals and impressive accuracy in clinical decision-making [58]. In another study, GPT-4o was queried for diagnoses and management plans with 20 physician-written clinical vignettes in otolaryngology, showing high agreement with physicians [59]. In addition, ChatGPT demonstrated appreciable knowledge and interpretation skills in psychiatry through 100 clinical case vignettes [60]. In another study, 33 physicians across 17 specialties generated 284 medical questions. ChatGPT generated largely accurate information in response to diverse medical queries as judged by academic physician specialists, with improvement over time [61]. ChatGPT also offered accurate recommendations on managing hypertension based on clinical practice guidelines [62]. ChatGPT provided accurate recommendations for 8 out of 10 clinical scenarios in cardiology [63]. In another study, GPT-4o generated a good rehabilitation plan for a stroke case from a textbook [64]. Finally, several studies have reported the application of

XSL•FO
RenderX

ChatGPT to predict common drug-drug interactions [65], managing polypharmacy [29], clinical pharmacy [66-68], and medical pharmacology as a self-learning tool [69]. Taken together, all previous studies suggest that ChatGPT could potentially be applied to complex geriatric syndrome vignettes with comprehensive clinical questions. This study was designed to demonstrate whether ChatGPT has solid geriatrics knowledge and can apply it to 2 common complex geriatric syndrome vignettes (polypharmacy and falls) by responding to comprehensive questions.

Study Objectives

We designed these 3 distinct approaches to provide evidence of whether ChatGPT can be trusted to be potentially applied to geriatrics education and clinical practice as an assistive tool.

Methods

This observational study used the validated UCLA geriatrics attitude instrument [48,49] and geriatrics knowledge test [57].

The Geriatrics Attitude Instrument

The geriatrics attitude instrument comprises 16 questions (Multimedia Appendix 1) [48,49]. In total, 6 of these statements reflect positive attitudes toward aging, such as "Most old people are pleasant to be with" (question 1). A total of 10 statements reflect negative attitudes toward aging, such as "Treatment of chronically ill old patients is hopeless" (question 11). Responses are graded on a Likert scale from 1 (*strongly disagree*) to 5 (*strongly agree*). The mean total score of ChatGPT was calculated and compared to scores from medical students, internal medicine residents, and geriatric medicine fellows based on previously published studies in the literature. Subtotals for positive and negative attitudes toward aging were also calculated to compare them to those in previously published studies. For statements reflecting a positive attitude toward aging, higher scores indicated a more positive attitude toward aging. For statements reflecting a negative attitude toward aging, lower scores indicated a less biased attitude toward aging.

The Geriatrics Knowledge Test

The validated UCLA geriatrics knowledge test contains 18 questions (Multimedia Appendix 2) [57]. In total, 8 questions are true-or-false statements, such as "Most older people are living in nursing homes" (question 4; false). A total of 10 questions are short clinical vignettes, such as "A 78-year-old nursing home resident with mild dementia associated with Alzheimer's disease is best determined by their ability to understand treatment options" (question 9; correct answer: B).

To minimize unfounded guessing, the following scoring system was used: +1 for a correct answer, –1 for a wrong answer, and 0 for "don't know." Therefore, the total score ranged from −18 to +18 [57]. Correct answers were determined by the author based on the literature.

ChatGPT Inputs and Outputs

GPT-3.5 was prompted to undertake the validated UCLA geriatrics attitude instrument and geriatrics knowledge test. Prompts were derived from original published papers [48,49,57]. Each prompt was repeated 4 times to evaluate the consistency and obtain an average response. In addition, ChatGPT was prompted to respond to 2 common geriatric syndrome vignettes (polypharmacy and falls). These 2 vignettes with questions were used in the author's previously published geriatrics curricula [70,71] and are described fully in the following sections.

The Polypharmacy Vignette

This vignette was used from a previously published workshop of prescribing and deprescribing [70]:

> *Ms. Smith, an 85-year-old white woman, has HTN, CAD, HFpEF, A-Fib, DM (II), HLD, hypothyroidism, gout, OA. On Warfarin 2.5 mg/d, furosemide 20 mg/d, Glyburide 5 mg/d, Spironolactone 50 mg/d, Amiodarone 300mg/d, Carvedilol 6.25 mg twice a day, Lisinopril 40 mg/d, Lipitor 80 mg/d, Levothyroxine 250 mg/d, Allopurinol 100 mg three times a day, acetaminophen 500 mg two tabs q6h for joint pain, and several supplements. Her pulse is 62 beat/min. Her BP is 120/70 (sit) and 98/64 (upright), weight 60 kg. No JVD, HR regularly, no murmur, few crackles on lung bases. Abdomen exam is benign, + pitting edema of ankles. Serum Cr is 1.2 mg/dL and Bun is 15 mg/dL. INR is 3.5. LDL is 50. albumin 4, Hb1ac is 6.1. TSH is 2. EF is 60%.*

ChatGPT was prompted with the polypharmacy vignette following 2 steps. First, ChatGPT was prompted to answer 4 general questions on the appropriate use of medications to examine its geriatric pharmacology knowledge: (1) "Which of the commonly used 8 drugs are potentially inappropriate medication (PIM) and should be avoided in older adults?" (2) "What is PIM in older adults?" (3) "How to identify PIM in older adults?" (4) "What to do with PIM?" Second, ChatGPT was prompted to answer 4 additional specific questions related to the polypharmacy vignette to evaluate its application of geriatric pharmacology to a clinical vignette: (1) "Any drug-drug interactions for this patient?" (2) "Any drug is ineffective for this patient?" (3) "This patient was taking so many medications. Any way to reduce the number of her medications?" (4) "Any drug-disease interaction?"

The Fall Vignette

This vignette was used from a previously published workshop of prescribing and deprescribing [71]:

> *An 82-year-old community-dwelling woman with PMH of multiple falls, right hip fracture, Dx of hypertension, coronary heart disease, knee arthritis, came to my clinic after she fell after her meal at home this morning. She felt pain of R head. She had a cane and walker at home but didn't use that often. She thought she didn't need it. She has taken care of her husband for several years. Her medication included Doxazosin (Cardura) and diltiazem (Cardizem), and acetaminophen as needed. I asked many questions related to her fall. Her vital signs were normal without orthostatic hypotension. Her cardiovascular and neurological examinations were benign. Her mini-mental status examination (MMSE) was 30/30.*

XSL•FO
RenderX

> *She had independent activities of daily living (ADL). Her geriatric depression scale (GDS) was 0/15. She passed timed up and go test. Because of her right black eye and pain of right head, I sent her to UVA to rule out orbital fracture. Please ask the following questions: List all potential fall risk factors for this patient What are three fall screening questions you need to ask for elderly patients in an outpatient setting based on 2010 fall prevention guideline you want to follow for this patient? Which one of three pathways in figure 1 from 2010 fall prevention guideline you want to follow for this patient? List all physical examinations that you think are important to perform for this patient. List your recommendations as many as you think that could help her to prevent her falling.*

ChatGPT was prompted with the fall vignette following 3 steps. First, ChatGPT was prompted to answer 1 general question on fall risks in older adults. Second, ChatGPT was prompted to answer five pretest questions on fall prevention in older adults: (1) "When an older person falls, should you ask the following questions (one or more or none)? the circumstances of fall, mental status, medication changes, vision change, drinking"; (2) "Which of the following illnesses is a fall risk (one or more or none)? Pneumonia, Parkinson's disease, Diabetes mellitus, Dementia, Orthostatic hypotension"; (3) "Which of the following medications are not associated with an increased fall risk (one or more)? Levothyroxine, Diazepam, Oxycodone, Amitriptyline"; (4) "Which ones of the following are true (one or more or none)? An elderly person may become so fearful of falling that they restrict his or her mobility Most falls are associated with significant injuries Falls is a leading cause of accidental death in older population About 5% community-dwelling elderly person falls each year Survivors of fall-related hip fracture are rarely institutionalized Exercise improves function and reduces fall risk and injurious falls"; and (5) "Which of the following is NOT an environmental fall risk? (only pick one answer) Throw rug, freshly waxed kitchen floor, Grab bars, Electrical cord lying on the floor." Finally, ChatGPT was prompted to answer 5 additional questions to examine the application of fall prevention knowledge to the fall vignette: (1) "List all potential fall risk factors for this patient"; (2) "What are three fall screening questions you need to ask for elderly patients in an outpatient setting based on 2010 fall prevention guideline you want to follow for this patient?"; (3) "Which one of three pathways in figure 1 from 2010 fall prevention guideline you want to follow for this patient?"; (4) "List all physical examinations that you think they are important to perform for this patient"; and (5) "List your recommendations as many as you think that could help her to prevent her falling."

Evaluation of ChatGPT Outputs

ChatGPT outputs were collected and analyzed by the author. The author judged the correctness of the ChatGPT outputs on the polypharmacy and fall vignette questions based on evidence from clinical practice guidelines, systematic reviews, and other types of evidence in the literature. The descriptive analysis was performed using SPSS (version 28; IBM Corp).

In summary, we designed these 3 distinct approaches to provide evidence on whether we can trust ChatGPT to be potentially applied to geriatrics education and clinical practice as an assisting tool.

Ethical Considerations

This study was not human participant related, including the secondary data analysis. No informed consent was needed.

Results

Geriatrics Attitude

The total geriatrics attitude score of ChatGPT was significantly lower than that of trainees (medical students, internal medicine residents, and geriatric medicine fellows) from validation [49,57] and follow-up studies [48] (Multimedia Appendix 3). However, the mean subscore on positive geriatrics attitude of ChatGPT was higher than that of the trainees (medical students, internal medicine residents, and neurologists; Multimedia Appendix 4), where a higher score is better. The higher subscore on positive geriatrics attitude indicates a better attitude toward aging. Conversely, the mean subscore on negative geriatrics attitude of ChatGPT was lower than that of the trainees and neurologists, where lower subscores are better (Multimedia Appendix 4). The lower subscore on negative geriatrics attitude indicates a less age-biased attitude toward aging. Individual responses to 14 geriatrics attitude statements by ChatGPT, trainees, and neurologists are shown in Multimedia Appendix 4. The subscores of the comparison group in Multimedia Appendix 4 were based on previously published studies [72-74]. Notably, responses to statements 15 and 16 were rarely reported in previously published studies (Multimedia Appendix 4).

Of the 14 geriatrics attitude statements (Multimedia Appendix 4), 5 (36%) were positive, and 9 (64%) were negative. Regarding the positive geriatrics attitude statements, ChatGPT had 1 lower positive geriatric attitude response (indicating a less positive geriatrics attitude) and 5 similar or better positive geriatric attitude responses compared to medical students and residents (indicating a similar or better positive geriatrics attitude). Regarding the negative geriatrics attitude statements, ChatGPT had 1 similar negative geriatric attitude response and 8 better negative geriatrics attitude responses than medical students and residents, indicating a less negative geriatrics attitude.

The response from ChatGPT to the first prompt of the geriatrics attitude questions did not follow the Likert-scale format (1-5; 1=*strongly disagree*; 5=*strongly agree*; Multimedia Appendices 3 and 4) as ChatGPT provided only comments. The response to the third prompt included both the Likert-scale format (1-5; 1=*strongly disagree*; 5=*strongly agree*) and comments (Multimedia Appendix 4).

Geriatrics Knowledge

The total score of ChatGPT (mean 14.25) was significantly higher than the scores from all trainees (medical students, residents, and internal medicine fellows) in the validation studies [57] (mean 11.3; Multimedia Appendix 3) and was also significantly higher than that of first-year (mean 9.9) and second-year (mean 9.5) medical students (Multimedia Appendix

XSL•FO
RenderX

3) in the follow-up studies [48]. It was slightly higher than the scores of third-year medical students (mean 13.6) and internal medicine interns (mean 13.2), slightly lower than the scores of second- and third-year internal medicine residents (mean 14.7 and 14.9, respectively), and significantly lower than the scores of geriatric medicine fellows (mean 17.5) in the follow-up studies [48]. ChatGPT's performance on geriatrics knowledge in the first prompt was lower than that on the following 3 repeat prompts. For the first prompt, ChatGPT provided rationales without selecting options (Multimedia Appendix 4). For the third prompt, ChatGPT selected options and provided rationales for its choices (Multimedia Appendix 4). For the second and fourth prompts, ChatGPT select options without providing rationales (Multimedia Appendix 4).

To make it easy for the reader to replicate the results of this study, Multimedia Appendix 5 shows a few examples of prompts on geriatrics attitude and knowledge test questions and outputs by ChatGPT (all original outputs of ChatGPT are available upon request).

The Polypharmacy Vignette

Overall, ChatGPT performed well on the polypharmacy vignette involving a woman aged 85 years (Multimedia Appendix 6). ChatGPT provided appropriate responses to 4 general drug therapy questions and moderate responses to 4 specific drug therapy questions based on the vignette. ChatGPT correctly identified 5 drug-drug interactions and suggested deprescribing. However, it missed an ineffective medication (a supplement), 1 drug-drug interaction, and 3 drug-disease interactions (Multimedia Appendix 6). Despite the patient not having lung disease, ChatGPT provided 2 irrelevant drug-disease interactions specific to the vignette (ChatGPT hallucination; Multimedia Appendix 6).

The Fall Vignette

ChatGPT performed well on the fall vignette involving a woman aged 82 years (Multimedia Appendix 7). ChatGPT correctly summarized 10 common fall risks in older adults and correctly answered 3 out of 5 pretest questions. In the remaining 2 pretest questions, ChatGPT missed a few correct responses. In the fall vignette, ChatGPT provided appropriate responses to all 4 prompts (Multimedia Appendix 7). ChatGPT recognized most fall risk factors and responded perfectly to fall screening questions following the latest Centers for Disease Control and Prevention (CDC) fall guidelines even though the prompt specified the 2010 fall prevention guidelines. ChatGPT accurately followed the 3 pathways from the 2010 fall guidelines. It missed a few physical examinations. ChatGPT provided almost all recommended fall prevention strategies, with only a few omissions.

Discussion

Principal Findings

With the increasing use of ChatGPT in medical education [3-7] and clinical practice [5-12], the aim of this study was to evaluate whether we can trust ChatGPT to be used in geriatrics. The major findings of this study demonstrated that ChatGPT had less age-biased output than the trainees, outperformed the trainees on the validated UCLA geriatrics knowledge test, and reasonably applied geriatrics knowledge to 2 common geriatric syndrome vignettes (polypharmacy and falls). The preliminary findings of this study are promising and demonstrate that ChatGPT could be trusted and used as an assistant tool in geriatrics practice and education via 3 approaches, which are fully discussed in the following paragraphs.

The first approach was to evaluate ChatGPT's performance on a validated UCLA geriatrics attitude instrument [48,49]. To the best of our knowledge, this is the first study to demonstrate that ChatGPT is less likely to generate age-biased outputs, which contrasts with a few ethical and other concerns regarding ChatGPT [16-21]. This finding is significant because ageism is prevalent and associated with multiple poor outcomes, including health outcomes [34,35]. Various instruments exist to measure geriatrics attitude to assess ageism [34,35], with the UCLA geriatrics attitude instrument [48,49] being one of these validated tools [75], which we used in this study [48,49]. The UCLA geriatrics attitude instrument includes 16 items [48,49], 6 of which refer to a positive geriatrics attitude (eg, "most old people are pleasant to be with") where a higher score is better, and 10 of which refer to a negative geriatrics attitude where a lower score is better. Previous studies using the UCLA geriatrics attitude instrument have primarily used the total score on the geriatrics attitude scales without calculating and reporting the mean subscores on positive and negative geriatrics attitudes [48,49], which makes it difficult to interpret the significance of geriatrics attitude. However, a few studies have reported individual responses to the UCLA geriatrics attitude instrument, including positive and negative geriatrics attitude [48,49], for which we were able to calculate the subscores for comparison. We recommend that the subscores of the UCLA geriatrics attitude instrument should be reported to assess positive and negative geriatrics attitude. Overall, ChatGPT's individual responses to the UCLA geriatrics attitude statements were better than those of medical students and residents, with higher scores for positive attitude and lower scores for negative attitude (Multimedia Appendix 4), suggesting that ChatGPT has less age-biased outputs and could be trusted from an ethical perspective. This study suggests that ChatGPT is better than the news media and humans in providing less age-biased information [39,40]. However, a reliable and valid instrument with which to quantify modern medical student attitudes toward older people has not yet been developed. An adaptation of the Aging Semantic Differential scale for contemporary use has been recommended [76]. This study can be repeated using the Aging Semantic Differential scale to see whether ChatGPT still generates less age-biased outputs in the future.

The second approach was to evaluate ChatGPT's geriatrics knowledge. This study is the first to assess ChatGPT's performance on a validated UCLA geriatrics knowledge test [49,57]. ChatGPT has a substantial knowledge base and has been reported to pass the USMLE [50,51] and perform well on many other examinations [52-56]. Previous studies have suggested that ChatGPT could significantly impact medical education [3-7] and clinical practice, including clinical decision-making [5-12]. For example, a recent systematic review and meta-analysis on the performance of ChatGPT in medical

XSL•FO
RenderX

examinations showed that ChatGPT has been evaluated in multiple specialties, including plastic surgery, ophthalmology, neurosurgery, orthopedics, diabetes, gastroenterology, radiology, cardiology, dermatology, and anesthesia [61]. It can be reasonably believed that ChatGPT has good geriatrics knowledge. This study supported this belief and demonstrated that ChatGPT performed better than the average trainee in the validation studies (mean score 14.7 vs 11.3) and follow-up studies (mean score 14 vs 13.3; Multimedia Appendix 3). ChatGPT's performance was better than that of medical students and first-year internal medicine residents (PGY 1 in Multimedia Appendix 3), comparable to that of second- and third-year residents (PGY 2 and PGY 3 in Multimedia Appendix 3), but lower than that of geriatric medicine fellows. This suggests that ChatGPT could be a valuable supplemental resource for health professional trainees, health care providers, and laypeople. Given that GPT-4o has more parameters, it is reasonable to believe that GPT-4o could perform better on the geriatrics knowledge test. The UCLA geriatrics knowledge test only has 18 questions [57] and much less questions than the USMLE tested in previous studies [50,51]. Future studies should evaluate GPT-4o's performance on more multiple-choice questions from geriatrics board examinations to confirm the findings of this study before adopting ChatGPT for geriatrics education and clinical practice.

The third approach was to evaluate ChatGPT's responses to 2 common geriatric syndrome vignettes—polypharmacy and falls in older adults—from the author's previously published geriatrics curricula [70,71]. This study demonstrated that ChatGPT performed well on these 2 geriatric syndrome vignettes (Multimedia Appendices 6 and 7), which was promising and will be fully discussed in the following paragraphs.

For the polypharmacy vignette, ChatGPT was able to provide systematic recommendations for polypharmacy in older adults, which has not been investigated before. In this study, ChatGPT provided appropriate general principles for prescribing and deprescribing; identified PIMs; and suggested resources and criteria such as the Beers criteria [77], the Screening Tool of Older Persons' Prescriptions [75], the Drug Burden Index [78], and the Medication Appropriateness Index [79] (Multimedia Appendix 6). In addition, ChatGPT was able to provide 5 appropriate ways to manage PIMs in older adults (Multimedia Appendix 6) that are consistent with the evidence from geriatrics practice guidelines and prescribing principles [75,77-80]. In particular, this study demonstrated that ChatGPT can identify all 8 exemplary PIMs (Multimedia Appendix 6). These medications should be avoided in older adults [75,77-80]. ChatGPT was able to identify 4 drug-disease interactions but missed 3 drug-disease interactions in the polypharmacy vignette (Multimedia Appendix 6). This suggests room for improvement in training ChatGPT further. ChatGPT also identified drug-drug interactions accurately in the polypharmacy vignette (Multimedia Appendix 6), which was different from a previous study in which ChatGPT identified 39 out of 40 drug-drug interactions when it was given 40 pairs of drugs [65]. Therefore, the question is what the implications are of what ChatGPT can do for polypharmacy. Previous studies have shown that within the last 10 years medical students, postgraduate residents, primary care providers, and pharmacists were unaware of PIMs and the standard guidelines for older adults, such as the Beers criteria [81-84]. This study has shown that ChatGPT is more knowledgeable than humans. Both geriatricians and general internists still prescribed 7.2% to 8.7% of PIMs, respectively [84].

The persistent prevalence of PIMs and polypharmacy in older adults was well-reported [31,32]. What ChatGPT demonstrated its good knowledge and ability to do with polypharmacy in this study suggests that ChatGPT has great potential to be used by trainees and providers as an assistant tool. One suggested example will be to include an older patient's medical history and a list of all their medications in the ChatGPT prompt used in this study for medication review and to generate ChatGPT outputs which can be reviewed by providers at Clinic or other clinical settings. This will help nurses, other providers, and trainees at the point of care. It could help laypeople self-check PIMs. It could also help health professional trainees in the self-study of geriatric pharmacology.

Falling is another significant US and global problem [85,86] and one of the common geriatric syndromes in older adults [87]. ChatGPT has not been tested to provide systematic fall prevention recommendations for an older adult. This study demonstrated for the first time that ChatGPT has good knowledge of fall prevention as it performed well on pretests and provided a comprehensive summary of 10 common fall risks in older adults and specific recommendations in the fall vignette (Multimedia Appendix 7).

The 2024 US Preventive Services Task Force fall prevention guidelines listed several fall risks, including age; history of falls; cognitive and sensory deficits; presence of acute and chronic medical conditions; certain medications; environmental or occupational hazards; home or neighborhood features and alcohol or drug use; and impairments in mobility, gait, and balance [87]. The 2023 CDC Stopping Elderly Accidents, Deaths, and Injuries also has a long list of fall risks [88]. In addition, a systematic review and meta-analysis showed evidence-based fall risk factors among the aging population [89]. This study showed that ChatGPT identified most of fall risks for the fall vignette (Multimedia Appendix 7) consistent with the aforementioned fall prevention guidelines [87,88] and systematic review [89] except for missing risky behaviors such as standing on a chair instead of a step stool and loss of sensation in the feet.

This study also used 5 pretests to further evaluate ChatGPT's knowledge of fall prevention (Multimedia Appendix 7). ChatGPT generated responses consistent with the guidelines [88,89] and a systematic review [90]. In pretest 1, ChatGPT provided 5 good recommendations with reasoning for taking the medical history of an older adult. In pretest 2, ChatGPT correctly identified several medical conditions as fall risks, including Parkinson disease, dementia, and orthostatic hypotension but missing pneumonia and diabetes mellitus. This indicates that ChatGPT has some limitations. In pretest 3, ChatGPT correctly identified diazepam, oxycodone, and amitriptyline as fall risks [75,77-79] and levothyroxine as not a fall risk. In pretest 4, ChatGPT underestimated the prevalence

XSL•FO
RenderX

of falls in community-dwelling older adults (5%) but correctly answered most other true and false questions about fall risks and consequences. In pretest 5, ChatGPT correctly answered all questions about environmental fall risks and identified grab bars as not a risk. Overall, ChatGPT performed very well on 5 pretests with a few mistakes, indicating good knowledge of fall prevention for older adults.

To further test ChatGPT's clinical application to individual patients, the fall vignette was used to prompt ChatGPT. ChatGPT correctly identified 6 common fall risk factors except for potential home environment factors, consistent with the evidence from fall prevention guidelines [88,89] and a systematic review [90]. ChatGPT impressively recognized older caregiver burden, hypertension, and coronary heart diseases as fall risks, also consistent with the evidence from fall prevention guidelines [88,89] and a systematic review [90]. In response to fall screening questions, ChatGPT was correct if the 2023 CDC guidelines were used [89] but incorrect if the 2010 American Geriatrics Society guidelines were used [91]. This study indicates that ChatGPT can use the latest guidelines, although specific prompts might be unclear to ChatGPT. Regarding physical examinations relevant to the fall vignette, ChatGPT missed several but recommended multiple appropriate physical examinations [89,91], indicating a reasonably good performance. In the final prompt on fall prevention recommendations for the fall vignette, ChatGPT provided 6 appropriate fall prevention recommendations, including home environment–related prevention despite previously missing the home environment as a fall risk [88-91]. It was notable that the home environment was not on the list of fall risks in ChatGPT's response to the first prompt. However, ChatGPT correctly answered pretest 5 (home environment fall risk) and recommended home environment–related prevention for the patient in the fall vignette (Multimedia Appendix 7). This suggests some discordance between the fall assessment and action plans by ChatGPT, indicating a possibly different thinking and decision process within ChatGPT.

The crucial question is whether ChatGPT can be helpful and necessary in screening for fall risks and providing fall prevention recommendations to older adults in daily practice. Providers need knowledge and willingness to screen for fall risks in older adults, but multiple studies have shown that providers often lack fall prevention knowledge and are unwilling to screen for fall risks [92-96]. For example, one study showed that emergency providers lacked knowledge on which patients to be screened and were unwilling to spend more than a few minutes on screening for fall intervention [95]. Another study revealed that only 14% of providers at accountable care organizations were aware of the CDC's fall risk assessment algorithm (Stopping Elderly Accidents, Deaths, and Injuries) [93]. Furthermore, 43% of primary care providers did not agree that they had the expertise or time to perform fall risk assessments, and only a small percentage billed for fall risk screening despite being aware of Medicare reimbursement [96]. In addition, general practitioners were often unaware of their frail patients' fall history or fear of falling, with most patients not receiving fall prevention care. Less than 40% of providers asked most or all their older patients if they had fallen in the previous 12 months, and less than a quarter referred their patients to physical therapists for balance or gait training [92]. A recent CDC report showed some improvement, but gaps remain. Only 20% of providers were aware of any injury prevention resources, and a higher percentage screened for fall risk when older adults presented with specific concerns [94].

Given the high prevalence of fall among older adults and poor knowledge of fall prevention among providers, this study demonstrated that ChatGPT has better knowledge and ability to apply the fall prevention guidelines than many providers, suggesting that it can assist providers in assessing fall risks and preventing falls in older adults. For instance, entering an older patient's history into ChatGPT and reviewing the outputs generated can help providers identify patients with fall risk at the point of care. ChatGPT can also be used as an assistant tool to help health professional trainees with self-study.

Limitations and Quality Assessment of This Study

There are serval limitations to this study, which will be discussed in this section. ChatGPT's responses are based on the data that it was trained on, which may not include the most current geriatrics research, best practices, or region-specific guidelines, potentially limiting the trustworthiness of its advice. This study did not account for the quality of the information sources that ChatGPT draws from as ChatGPT cannot differentiate between authoritative and nonauthoritative content in real time like all other ChatGPT-based research. Despite GPT-4o having more data for training without releasing its data sources, this will not solve this basic problem. This study indicates that applying ChatGPT to geriatrics practice and education should be cautious and it should be used as an assistant tool only. ChatGPT will never or is unlikely to replace clinicians. As this study evaluated ChatGPT based on geriatric syndrome vignettes, it does not fully account for the complexities and nuances of real-world clinical decision-making and patient interactions, where additional context, history, and human judgment play critical roles. In addition, ChatGPT's responses might be overly generalized as it lacks the ability to perform individualized assessments or make nuanced clinical decisions based on patient-specific data. This is similar to vignette-based simulation education. This indicates that ChatGPT will not make a decision for someone but could assist with the decision-making. The effectiveness of ChatGPT's responses varies significantly depending on how well the user phrases the questions or provides necessary information. The input provided in this study was based on the author's geriatrics practice experience for >20 years, which might not mimic the way in which other clinicians would interact with the tool. Therefore, the results of this study may not be generalizable. Prompt engineering is developing to help improve the quality of the input. However, a standardized prompts among different providers can be very challenging. The author suggests practicing prompts to improve human-ChatGPT interaction and obtain reliable outputs from ChatGPT.

ChatGPT lacks clinical experience and the ability to engage in human clinical judgment, whereas it can process vast amounts of medical information. This may limit its application in critical decision-making scenarios or nuanced treatment planning for

XSL•FO
RenderX

complex geriatrics cases. Providers must use their clinical judgment integrated with clinical circumstances and not fully depend on the outputs of ChatGPT.

ChatGPT might face ethical and legal consequences. ChatGPT or AI hallucinations could cause ethical and legal risks. After the release of GPT-4o, ChatGPT often says, "I am AI and not clinician and cannot make decisions for you." It seems that ChatGPT developers recognize the potential legal and ethical consequences and protect ChatGPT.

Human-AI interaction is complex. Providers should be cautious in interpreting ChatGPT outputs. ChatGPT outputs should be interpreting with evidence-based clinical guidelines, systematic reviews, and randomized controlled trials shown in the Comments from the author columns in Multimedia Appendices 6 and 7. Whether the ChatGPT outputs are appropriate in this study should be judged and verified by a group of geriatricians

In addition, this study has several other limitations. This study used GPT-3.5. GPT-4o might produce better results. Comparisons between ChatGPT and human performance on geriatrics attitude and knowledge were based on previously reported results in the literature many years ago. A new study assessing geriatrics attitude and knowledge among current medical students, residents, and geriatric medicine fellows is needed to confirm the findings of this study. ChatGPT produced some errors, including ChatGPT hallucinations such as unrelated drug-disease interactions in Multimedia Appendix 6, consistent with previous reports [97,98]. We should be aware that ChatGPT can make mistakes, just like humans. ChatGPT should not be used alone in geriatrics practice and education at this point. ChatGPT needs to be pretrained specifically for medical education and clinical practice.

Like any other research, systematic appraisal of the current ChatGPT research is needed. A preliminary checklist, Model, Evaluation, Timing, Range and Randomization, Individual Factors, Counts, and Specificity of Prompts and Language (METRICS), to standardize the design and reporting of studies on generative AI–based models in health care education and practice has been recently developed [99]. The current study was started and completed before its development. However, it is worthwhile to use METRICS to evaluate the quality of this study. The following seven domains of METRICS will be used:

1. Model—ChatGPT was described and used.
2. Evaluation—both subjective and objective evaluation were described and used.
3. Timing—the study date was documented, but the duration was not documented; transparency—the geriatrics attitude instrument and knowledge test, the vignettes, and prompts were well described, and the outputs were exactly presented.
4. Range—the topics were well described, including geriatrics attitude and knowledge; randomization was not used.
5. Individual—there was subjective involvement with ChatGPT, such as judging the correctness of its outputs. This study was conducted by a single investigator.
6. Count—a total of 4 prompts were used. This study did not have a sample size.
7. Specificity of the prompts or language—for geriatrics attitude and knowledge, the prompts were from previously published studies. It is unclear whether they are appropriate for ChatGPT. For the 2 vignettes on polypharmacy and fall, the prompts were from previously published curricula. It is also unclear whether they are appropriate for ChatGPT.

Overall, the quality of this study is reasonably good based on METRICS. METRICS should be used in the reporting of any AI-based research in all journals.

In summary, this study took 3 distinct approaches to demonstrate the trustworthiness of ChatGPT to be used as an assistant tool in geriatrics practice and education. The implications of this preliminary study for geriatrics practice and education are further discussed in the following section.

Implications for Geriatrics Practice and Education

This study showed that ChatGPT outputs are not age-biased using the validated geriatrics attitude test and that ChatGPT performed well on the validated geriatrics knowledge test and on applied geriatrics knowledge tests (2 cases of common geriatric syndromes). These findings suggest that ChatGPT could potentially help geriatrics practice and education as an assistant tool in numerous ways. It is very important to know that ChatGPT is algorithm based. Clinical practice often uses algorithm- and pathway. Therefore, the underlying rationale for ChatGPT and clinical practice is similar. However, ChatGPT is not intended to replace physicians and other providers.

ChatGPT can be used as an assistant geriatrics tool and tutor to support self-learning and immediate feedback and self-assessment. For example, when trainees are studying geriatric syndromes, they can obtain responses on geriatric syndromes from ChatGPT. Therefore, ChatGPT could be incorporated to supplement the existing program in geriatrics education and support continuing medical education. ChatGPT can be potentially to be integrated to clinical reasoning and decision in a timely fashion and potentially improve the point of care in all clinical settings. For example, ChatGPT can help identify individual older patients who might have fall risks and drug-drug interactions.

ChatGPT can be used as an assistant or autonomous clinical tool to alleviate the geriatrics workforce crisis. For example, a nongeriatrician health provider could look for answers to clinical geriatrics questions, such as recommendations for preventing falls and reducing polypharmacy in older patients, from ChatGPT and help answer their patients' questions in a timely fashion. This is particularly helpful for rural practice where geriatrics consults are rare or nonexistent. This will improve physicians' and other providers' efficiency and accuracy in the care of older patients and provider-patient interactions.

It is expected that ChatGPT can help with post-visit summaries by providing patient education materials, such as home safety and environment modification for fall prevention. ChatGPT can also draft Clinic letters such as providing lab results and their interpretation to the patient in the Clinic letter.

ChatGPT can help older patients with self-screening or initial self-assessment when they are sitting in a clinic waiting room. For example, they can input their medical history into ChatGPT

XSL•FO
RenderX

such as age, physical function medications, and chronic conditions to ask whether they are at higher risk of fall.

Importantly, GPT-3.5 is free to use and can be cost-saving for providers and the health care system.

Finally, with more data training, ChatGPT will be more reliable and helpful in geriatrics practice and education.

Comparison With Previous Work

To the best of our knowledge, health professional trainees and providers but not ChatGPT have been assessed for geriatrics attitude [48,49]. ChatGPT has been assessed for geriatrics knowledge [28-30]. For example, ChatGPT was asked to answer 10 geriatrics questions [30]. The correctness of the outputs by ChatGPT were rated by a group of geriatricians [30]. In contrast, in this study, the correctness of the outputs by ChatGPT was determined by the author only, which could lead to errors. In another study, ChatGPT was prompted with 1 simple question [28]. The questions in the aforementioned studies were not validated. However, this study used validated UCLA geriatrics knowledge tests [57,72-74]. Several previous studies have demonstrated that ChatGPT has geriatric pharmacology knowledge [29,65-69]. For example, ChatGPT could identify drug-drug interactions from 40 drug-drug pairs [65]. In another study, ChatGPT could manage polypharmacy and make decisions on deprescribing 2 to 3 inappropriate medications based on a vignette [29]. Another study compared the answers from ChatGPT and clinical pharmacists to real clinical cases and clinical competency assessments [68]. However, these studies are different from this study, which used a systemic approach to polypharmacy (Multimedia Appendix 6). To the best of our knowledge, ChatGPT has not approached falls in older adults.

Future Directions

This study was a pilot with both promising findings and limitations. With the fast-growing use of ChatGPT and the new version of GPT-4o being released, the author suggests the following research directions.

A larger study using ChatGPT is needed to extend and confirm the findings of this study before adopting ChatGPT for geriatrics education and clinical practice. For example, more geriatrics knowledge questions and vignettes should be examined using GPT-4o. For example, the performance of ChatGPT on geriatrics certification exams should be tested.

Coinvestigators are needed to reduce biases. In particular, the correctness of the ChatGPT outputs should be evaluated and judged by a group of experts in geriatrics and integrated with the latest evidence, such as clinical practice guidelines, systematic reviews, and randomized controlled trials. For example, the ChatGPT outputs to a prompt should be rated from *strongly agree* to *strongly disagree* by a group of geriatrics experts in addition to determining their consistency with evidence-based clinical practice guidelines

Those who conduct ChatGPT research should receive training on prompt engineering. The author is doing this and feels that it is beneficial. This could reduce the variation in prompts.

The reliability of ChatGPT performance needs to be tested using different prompts to improve generalizability. The outcomes of the application of ChatGPT to geriatrics practice and education need to be further investigated.

Conclusions

This study suggests that ChatGPT could be a valuable assistant tool in geriatrics education and practice, helping health professional trainees and providers combat ageism, supplement geriatrics knowledge, and address common geriatric syndromes such as polypharmacy and falls. This study demonstrated that we could trust ChatGPT to be used in geriatrics practice and education by using 3 distinct approaches. One strength of this study is that Multimedia Appendices 5-7 provide the details of prompts to ChatGPT and the outputs generated by ChatGPT, which allows readers to apply a similar approach to their geriatrics education and practice studies. The findings of this study are promising but need more investigation before ChatGPT is widely adopted in geriatrics practice and education.

Data Availability

All data generated or analyzed during this study including all ChatGPT prompts and outputs are included in this published article and its supplementary information files on the journal website.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Geriatrics attitudes tests.
[DOCX File , 23 KB - formative_v9i1e63494_app1.docx]

Multimedia Appendix 2
University of California, Los Angeles, geriatrics knowledge test (questions 1-18).
[DOCX File , 25 KB - formative_v9i1e63494_app2.docx]

Multimedia Appendix 3

XSL•FO
RenderX

Comparison of geriatrics attitude and knowledge test performance between ChatGPT and trainees.
[DOCX File , 25 KB - formative_v9i1e63494_app3.docx]

Multimedia Appendix 4
Comparison of performance on the geriatrics attitude subscales by ChatGPT, trainees, and neurologists.
[DOCX File , 17 KB - formative_v9i1e63494_app4.docx]

Multimedia Appendix 5
Examples of ChatGPT outputs on geriatrics attitude and knowledge questions.
[DOCX File , 15 KB - formative_v9i1e63494_app5.docx]

Multimedia Appendix 6
ChatGPT output on the polypharmacy vignette about a woman aged 85 years.
[DOCX File , 17 KB - formative_v9i1e63494_app6.docx]

Multimedia Appendix 7
Fall risk identification and management in a woman aged 82 years.
[DOCX File , 20 KB - formative_v9i1e63494_app7.docx]

References

1. Number of ChatGPT users (Dec 2024). Exploding Topics. 2024 Dec 3. URL: https://explodingtopics.com/blog/chatgpt-users [accessed 2024-10-26]
2. PubMed homepage. PubMed. URL: https://pubmed.ncbi.nlm.nih.gov/ [accessed 2024-05-15]
3. Boscardin CK, Gin B, Golde PB, Hauer KE. ChatGPT and generative artificial intelligence for medical education: potential impact and opportunity. Acad Med 2023 Aug 31;99(1):22-27. [doi: 10.1097/acm.0000000000005439]
4. Preiksaitis C, Rose C. Opportunities, challenges, and future directions of generative artificial intelligence in medical education: scoping review. JMIR Med Educ 2023 Oct 20;9:e48785 [FREE Full text] [doi: 10.2196/48785] [Medline: 37862079]
5. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023 Mar 19;11(6):887 [FREE Full text] [doi: 10.3390/healthcare11060887] [Medline: 36981544]
6. Thirunavukarasu AJ, Ting DS, Elangovan K, Gutierrez L, Tan TF, Ting DS. Large language models in medicine. Nat Med 2023 Aug 17;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
7. Meng X, Yan X, Zhang K, Liu D, Cui X, Yang Y, et al. The application of large language models in medicine: a scoping review. iScience 2024 May 17;27(5):109713 [FREE Full text] [doi: 10.1016/j.isci.2024.109713] [Medline: 38746668]
8. Liu J, Wang C, Liu S. Utility of ChatGPT in clinical practice. J Med Internet Res 2023 Jun 28;25:e48568 [FREE Full text] [doi: 10.2196/48568] [Medline: 37379067]
9. Andrew A. Potential applications and implications of large language models in primary care. Fam Med Community Health 2024 Jan 30;12(Suppl 1):e002602 [FREE Full text] [doi: 10.1136/fmch-2023-002602] [Medline: 38290759]
10. Li J, Dada A, Puladi B, Kleesiek J, Egger J. ChatGPT in healthcare: a taxonomy and systematic review. Comput Methods Programs Biomed 2024 Mar;245:108013 [FREE Full text] [doi: 10.1016/j.cmpb.2024.108013] [Medline: 38262126]
11. Rao SJ, Isath A, Krishnan P, Tangsrivimol JA, Virk HU, Wang Z, et al. ChatGPT: a conceptual review of applications and utility in the field of medicine. J Med Syst 2024 Jun 05;48(1):59. [doi: 10.1007/s10916-024-02075-x] [Medline: 38836893]
12. Momenaei B, Mansour HA, Kuriyan AE, Xu D, Sridhar J, Ting DS, et al. ChatGPT enters the room: what it means for patient counseling, physician education, academics, and disease management. Curr Opin Ophthalmol 2024 May 01;35(3):205-209. [doi: 10.1097/ICU.0000000000001036] [Medline: 38334288]
13. Cheng SW, Chang CW, Chang WJ, Wang HW, Liang CS, Kishimoto T, et al. The now and future of ChatGPT and GPT in psychiatry. Psychiatry Clin Neurosci 2023 Nov 11;77(11):592-596 [FREE Full text] [doi: 10.1111/pcn.13588] [Medline: 37612880]
14. Temperley HC, O'Sullivan NJ, Mac Curtain BM, Corr A, Meaney JF, Kelly ME, et al. Current applications and future potential of ChatGPT in radiology: a systematic review. J Med Imaging Radiat Oncol 2024 Apr;68(3):257-264. [doi: 10.1111/1754-9485.13621] [Medline: 38243605]
15. Lautrup AD, Hyrup T, Schneider-Kamp A, Dahl M, Lindholt JS, Schneider-Kamp P. Heart-to-heart with ChatGPT: the impact of patients consulting AI for cardiovascular health advice. Open Heart 2023 Nov;10(2):e002455 [FREE Full text] [doi: 10.1136/openhrt-2023-002455] [Medline: 37945282]
16. Mohammad B, Supti T, Alzubaidi M, Shah H, Alam T, Shah Z, et al. The pros and cons of using ChatGPT in medical education: a scoping review. Stud Health Technol Inform 2023 Jun 29;305:644-647. [doi: 10.3233/SHTI230580] [Medline: 37387114]

XSL•FO
RenderX

17. Duffourc M, Gerke S. Generative AI in health care and liability risks for physicians and safety concerns for patients. JAMA 2023 Jul 25;330(4):313-314. [doi: 10.1001/jama.2023.9630] [Medline: 37410497]
18. Kanter GP, Packel EA. Health care privacy risks of AI chatbots. JAMA 2023 Jul 25;330(4):311-312. [doi: 10.1001/jama.2023.9618] [Medline: 37410449]
19. Hanna JJ, Wakene AD, Lehmann CU, Medford RJ. Assessing racial and ethnic bias in text generation for healthcare-related tasks by ChatGPT. medRxiv. Preprint posted online on August 28, 2023 [FREE Full text] [doi: 10.1101/2023.08.28.23294730] [Medline: 37693388]
20. Amin KS, Forman HP, Davis MA. Even with ChatGPT, race matters. Clin Imaging 2024 May;109:110113. [doi: 10.1016/j.clinimag.2024.110113] [Medline: 38552383]
21. Erren TC, Lewis P, Shaw DM. Brave (in a) new world: an ethical perspective on chatbots for medical advice. Front Public Health 2023 Aug 17;11:1254334 [FREE Full text] [doi: 10.3389/fpubh.2023.1254334] [Medline: 37663854]
22. Korcok M. Medical schools face challenge of preparing physicians to care for fast-growing elderly population. JAMA 1985 Mar 01;253(9):1225-7, 1231. [Medline: 3968847]
23. Besdine R, Boult C, Brangman S, Coleman EA, Fried LP, Gerety M, et al. Caring for older Americans: the future of geriatric medicine. J Am Geriatr Soc 2005 Jun;53(6 Suppl):S245-S256. [doi: 10.1111/j.1532-5415.2005.53350.x] [Medline: 15963180]
24. Institute of Medicine, Committee on the Future Health Care Workforce for Older Americans, Board on Health Care Services. Retooling for an Aging America: Building the Health Care Workforce. Washington, DC: National Academies Press; 2008.
25. Dawson CM, Abiola AO, Sullivan AM, Schwartz AW, Members of the GERI Team Research Group. You can't be what you can't see: a systematic website review of Geriatrics Online-Visibility at US medical schools. J Am Geriatr Soc 2022 Oct;70(10):2996-3005 [FREE Full text] [doi: 10.1111/jgs.17997] [Medline: 36082829]
26. Gurwitz JH. The paradoxical decline of geriatric medicine as a profession. JAMA 2023 Aug 22;330(8):693-694. [doi: 10.1001/jama.2023.11110] [Medline: 37540519]
27. Fear K, Gleber C. Shaping the future of older adult care: ChatGPT, advanced AI, and the transformation of clinical practice. JMIR Aging 2023 Sep 13;6:e51776 [FREE Full text] [doi: 10.2196/51776] [Medline: 37703085]
28. Haque N. Artificial intelligence and geriatric medicine: new possibilities and consequences. J Am Geriatr Soc 2023 Jun 17;71(6):2028-2031. [doi: 10.1111/jgs.18334] [Medline: 36930059]
29. Rao A, Kim J, Lie W, Pang M, Fuh L, Dreyer KJ, et al. Proactive polypharmacy management using large language models: opportunities to enhance geriatric care. J Med Syst 2024 Apr 18;48(1):41. [doi: 10.1007/s10916-024-02058-y] [Medline: 38632172]
30. Rosselló-Jiménez D, Docampo S, Collado Y, Cuadra-Llopart L, Riba F, Llonch-Masriera M. Geriatrics and artificial intelligence in Spain (Ger-IA project): talking to ChatGPT, a nationwide survey. Eur Geriatr Med 2024 Aug;15(4):1129-1136. [doi: 10.1007/s41999-024-00970-7] [Medline: 38615289]
31. Nicholson K, Liu W, Fitzpatrick D, Hardacre KA, Roberts S, Salerno J, et al. Prevalence of multimorbidity and polypharmacy among adults and older adults: a systematic review. Lancet Healthy Longev 2024 Apr;5(4):e287-e296 [FREE Full text] [doi: 10.1016/S2666-7568(24)00007-2] [Medline: 38452787]
32. Hung A, Kim YH, Pavon JM. Deprescribing in older adults with polypharmacy. BMJ 2024 May 07;385:e074892. [doi: 10.1136/bmj-2023-074892] [Medline: 38719530]
33. Ng R, Chow TY. Aging narratives over 210 years (1810-2019). J Gerontol B Psychol Sci Soc Sci 2021 Oct 30;76(9):1799-1807 [FREE Full text] [doi: 10.1093/geronb/gbaa222] [Medline: 33300996]
34. Global report on ageism. World Health Organization. 2021 Mar 18. URL: https://www.who.int/teams/social-determinants-of-health/demographic-change-and-healthy-ageing/combatting-ageism/global-report-on-ageism [accessed 2024-12-12]
35. Hu RX, Luo M, Zhang A, Li LW. Associations of ageism and health: a systematic review of quantitative observational studies. Res Aging 2021 Aug;43(7-8):311-322. [doi: 10.1177/0164027520980130] [Medline: 33317402]
36. Allen JO, Solway E, Kirch M, Singer D, Kullgren JT, Moïse V, et al. Experiences of everyday ageism and the health of older US adults. JAMA Netw Open 2022 Jun 01;5(6):e2217240 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.17240] [Medline: 35704314]
37. Chang ES, Kannoth S, Levy S, Wang SY, Lee JE, Levy BR. Global reach of ageism on older persons' health: a systematic review. PLoS One 2020;15(1):e0220857 [FREE Full text] [doi: 10.1371/journal.pone.0220857] [Medline: 31940338]
38. Araújo PO, Soares IM, Vale PR, Sousa AR, Aparicio EC, Carvalho ES. Ageism directed to older adults in health services: a scoping review. Rev Lat Am Enfermagem 2023;31:e4019 [FREE Full text] [doi: 10.1590/1518-8345.6727.4019] [Medline: 37820219]
39. Amundsen D. A critical gerontological framing analysis of persistent ageism in NZ online news media: don't call us "elderly"!. J Aging Stud 2022 Jun;61:101009 [FREE Full text] [doi: 10.1016/j.jaging.2022.101009] [Medline: 35654544]
40. Murphy E, Fallon A, Dukelow T, O'Neill D. Don't call me elderly: a review of medical journals' use of ageist literature. Eur Geriatr Med 2022 Aug;13(4):1007-1009. [doi: 10.1007/s41999-022-00650-4] [Medline: 35581468]
41. Haase KR, Sattar S, Pilleron S, Lambrechts Y, Hannan M, Navarrete E, et al. A scoping review of ageism towards older adults in cancer care. J Geriatr Oncol 2023 Jan;14(1):101385. [doi: 10.1016/j.jgo.2022.09.014] [Medline: 36244925]

XSL•FO
RenderX

42. Bowling A. Ageism in cardiology. BMJ 1999 Nov 20;319(7221):1353-1355 [FREE Full text] [doi: 10.1136/bmj.319.7221.1353] [Medline: 10567146]
43. Silva MF, Silva DS, Bacurau AG, Francisco PM, Assumpção D, Neri AL, et al. Ageism against older adults in the context of the COVID-19 pandemic: an integrative review. Rev Saude Publica 2021 Apr 05;55:4 [FREE Full text] [doi: 10.11606/s1518-8787.2021055003082] [Medline: 33886953]
44. Altın Z, Buran F. Attitudes of health professionals toward elderly patients during the COVID-19 pandemic. Aging Clin Exp Res 2022 Oct;34(10):2567-2576 [FREE Full text] [doi: 10.1007/s40520-022-02209-6] [Medline: 35986878]
45. Martínez-Arnau FM, López-Hernández L, Castellano-Rioja E, Botella-Navas M, Pérez-Ros P. Interventions to improve attitudes toward older people in undergraduate health and social sciences students. A systematic review and meta-analysis. Nurse Educ Today 2022 Mar;110:105269 [FREE Full text] [doi: 10.1016/j.nedt.2022.105269] [Medline: 35063781]
46. Apriceno M, Levy SR. Systematic review and meta-analyses of effective programs for reducing ageism toward older adults. J Appl Gerontol 2023 Jun 17;42(6):1356-1375 [FREE Full text] [doi: 10.1177/07334648231165266] [Medline: 37069824]
47. Oca M, Meller L, Wilson K, Parikh AO, McCoy A, Chang J, et al. Bias and inaccuracy in AI chatbot ophthalmologist recommendations. Cureus 2023 Sep;15(9):e45911 [FREE Full text] [doi: 10.7759/cureus.45911] [Medline: 37885556]
48. Kishimoto M, Nagoshi M, Williams S, Masaki KH, Blanchette PL. Knowledge and attitudes about geriatrics of medical students, internal medicine residents, and geriatric medicine fellows. J Am Geriatr Soc 2005 Jan 22;53(1):99-102. [doi: 10.1111/j.1532-5415.2005.53018.x] [Medline: 15667384]
49. Reuben DB, Lee M, Davis JWJ, Eslami MS, Osterweil DG, Melchiore S, et al. Development and validation of a geriatrics attitudes scale for primary care residents. J Am Geriatr Soc 1998 Nov 27;46(11):1425-1430. [doi: 10.1111/j.1532-5415.1998.tb06012.x] [Medline: 9809767]
50. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the united states medical licensing examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 08;9:e45312 [FREE Full text] [doi: 10.2196/45312] [Medline: 36753318]
51. Yaneva V, Baldwin P, Jurich DP, Swygert K, Clauser BE. Examining ChatGPT performance on USMLE sample items and implications for assessment. Acad Med 2024 Feb 01;99(2):192-197. [doi: 10.1097/ACM.0000000000005549] [Medline: 37934828]
52. Levin G, Horesh N, Brezinov Y, Meyer R. Performance of ChatGPT in medical examinations: a systematic review and a meta-analysis. BJOG 2024 Feb;131(3):378-380. [doi: 10.1111/1471-0528.17641] [Medline: 37604703]
53. Kung JE, Marshall C, Gauthier C, Gonzalez TA, Jackson JB3. Evaluating ChatGPT performance on the orthopaedic in-training examination. JB JS Open Access 2023;8(3):e23.00056 [FREE Full text] [doi: 10.2106/JBJS.OA.23.00056] [Medline: 37693092]
54. Passby L, Jenko N, Wernham A. Performance of ChatGPT on specialty certificate examination in dermatology multiple-choice questions. Clin Exp Dermatol 2024 Jun 25;49(7):722-727. [doi: 10.1093/ced/llad197] [Medline: 37264670]
55. Ali R, Tang OY, Connolly ID, Zadnik Sullivan PL, Shin JH, Fridley JS, et al. Performance of ChatGPT and GPT-4 on neurosurgery written board examinations. Neurosurgery 2023 Dec 01;93(6):1353-1365. [doi: 10.1227/neu.0000000000002632] [Medline: 37581444]
56. Long C, Lowe K, Zhang J, Santos AD, Alanazi A, O'Brien D, et al. A novel evaluation model for assessing ChatGPT on otolaryngology-head and neck surgery certification examinations: performance study. JMIR Med Educ 2024 Jan 16;10:e49970 [FREE Full text] [doi: 10.2196/49970] [Medline: 38227351]
57. Reuben DB, Lee M, Davis JW, Eslami MS, Osterweil DG, Melchiore S, et al. Development and validation of a geriatrics knowledge test for primary care residents. J Gen Intern Med 1997 Jul;12(7):450-452 [FREE Full text] [doi: 10.1046/j.1525-1497.1997.00078.x] [Medline: 9229284]
58. Rao A, Pang M, Kim J, Kamineni M, Lie W, Prasad AK, et al. Assessing the utility of ChatGPT throughout the entire clinical workflow. medRxiv. Preprint posted online on February 26, 2023 [FREE Full text] [doi: 10.1101/2023.02.21.23285886] [Medline: 36865204]
59. Qu RW, Qureshi U, Petersen G, Lee SC. Diagnostic and management applications of ChatGPT in structured otolaryngology clinical scenarios. OTO Open 2023;7(3):e67 [FREE Full text] [doi: 10.1002/oto2.67] [Medline: 37614494]
60. Franco D'Souza R, Amanullah S, Mathew M, Surapaneni KM. Appraising the performance of ChatGPT in psychiatry using 100 clinical case vignettes. Asian J Psychiatr 2023 Nov;89:103770. [doi: 10.1016/j.ajp.2023.103770] [Medline: 37812998]
61. Goodman RS, Patrinely JR, Stone CAJ, Zimmerman E, Donald RR, Chang SS, et al. Accuracy and reliability of chatbot responses to physician questions. JAMA Netw Open 2023 Oct 02;6(10):e2336483 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.36483] [Medline: 37782499]
62. Kassab J, El Dahdah J, Chedid El Helou M, Layoun H, Sarraju A, Laffin LJ, et al. Assessing the accuracy of an online chat-based artificial intelligence model in providing recommendations on hypertension management in accordance with the 2017 American College of Cardiology/American Heart Association and 2018 European Society of Cardiology/European Society of Hypertension Guidelines. Hypertension 2023 Jul;80(7). [doi: 10.1161/hypertensionaha.123.21183]
63. Rizwan A, Sadiq T. The use of AI in diagnosing diseases and providing management plans: a consultation on cardiovascular disorders with ChatGPT. Cureus 2023 Aug;15(8):e43106 [FREE Full text] [doi: 10.7759/cureus.43106] [Medline: 37692649]

XSL•FO
RenderX

64. Zhang L, Tashiro S, Mukaino M, Yamada S. Use of artificial intelligence large language models as a clinical tool in rehabilitation medicine: a comparative test case. J Rehabil Med 2023 Sep 11;55:jrm13373 [FREE Full text] [doi: 10.2340/jrm.v55.13373] [Medline: 37691497]
65. Juhi A, Pipil N, Santra S, Mondal S, Behera JK, Mondal H. The capability of ChatGPT in predicting and explaining common drug-drug interactions. Cureus 2023 Mar;15(3):e36272 [FREE Full text] [doi: 10.7759/cureus.36272] [Medline: 37073184]
66. Fournier A, Fallet C, Sadeghipour F, Perrottet N. Assessing the applicability and appropriateness of ChatGPT in answering clinical pharmacy questions. Ann Pharm Fr 2024 May;82(3):507-513. [doi: 10.1016/j.pharma.2023.11.001] [Medline: 37992892]
67. Al-Dujaili Z, Omari S, Pillai J, Al Faraj A. Assessing the accuracy and consistency of ChatGPT in clinical pharmacy management: a preliminary analysis with clinical pharmacy experts worldwide. Res Social Adm Pharm 2023 Dec;19(12):1590-1594. [doi: 10.1016/j.sapharm.2023.08.012] [Medline: 37696742]
68. Huang X, Estau D, Liu X, Yu Y, Qin J, Li Z. Evaluating the performance of ChatGPT in clinical pharmacy: a comparative study of ChatGPT and clinical pharmacists. Br J Clin Pharmacol 2024 Jan;90(1):232-238. [doi: 10.1111/bcp.15896] [Medline: 37626010]
69. Choi W. Assessment of the capacity of ChatGPT as a self-learning tool in medical pharmacology: a study using MCQs. BMC Med Educ 2023 Nov 13;23(1):864 [FREE Full text] [doi: 10.1186/s12909-023-04832-x] [Medline: 37957666]
70. Cheng HY, Bradley E. Improving identification of potentially inappropriate medications among older patients: a preliminary report of a single-blind, controlled trial teaching fourth year medical students value- and preference-based prescribing/de-prescribing. Med Sci Educ 2016 Jun 28;26:431-435. [doi: 10.1007/s40670-016-0283-2]
71. Cheng H. Falls and fall prevention. MedEdPORTAL 2014 Sep 26:9910. [doi: 10.15766/mep_2374-8265.9910]
72. Alqahtani R, Almuhaidib S, Jradi H. A cross-sectional study: exploring knowledge and attitude of medical and nursing students to Care for Elders in the future. BMC Geriatr 2022 Nov 14;22(1):856 [FREE Full text] [doi: 10.1186/s12877-022-03551-0] [Medline: 36376782]
73. Hogan TM, Chan SB, Hansoti B. Multidimensional attitudes of emergency medicine residents toward older adults. West J Emerg Med 2014 Jul;15(4):511-517 [FREE Full text] [doi: 10.5811/westjem.2014.2.19937] [Medline: 25035760]
74. Seferoğlu M, Yıldız D, Pekel NB, Güneş A, Yıldız A, Tufan F. Attitudes of neurology specialists toward older adults. Aging Clin Exp Res 2017 Aug;29(4):787-792. [doi: 10.1007/s40520-016-0606-6] [Medline: 27456680]
75. O'Mahony D, Cherubini A, Guiteras AR, Denkinger M, Beuscart JB, Onder G, et al. STOPP/START criteria for potentially inappropriate prescribing in older people: version 3. Eur Geriatr Med 2023 Aug;14(4):625-632 [FREE Full text] [doi: 10.1007/s41999-023-00777-y] [Medline: 37256475]
76. Wilson MA, Kurrle S, Wilson I. Medical student attitudes towards older people: a critical review of quantitative measures. BMC Res Notes 2018 Jan 24;11(1):71 [FREE Full text] [doi: 10.1186/s13104-018-3186-z] [Medline: 29361969]
77. By the 2023 American Geriatrics Society Beers Criteria® Update Expert Panel. American Geriatrics Society 2023 updated AGS Beers Criteria® for potentially inappropriate medication use in older adults. J Am Geriatr Soc 2023 Jul 04;71(7):2052-2081. [doi: 10.1111/jgs.18372] [Medline: 37139824]
78. Hilmer SN, Mager DE, Simonsick EM, Cao Y, Ling SM, Windham BG, et al. A drug burden index to define the functional burden of medications in older people. Arch Intern Med 2007 Apr 23;167(8):781-787. [doi: 10.1001/archinte.167.8.781] [Medline: 17452540]
79. Samsa GP, Hanlon JT, Schmader KE, Weinberger M, Clipp EC, Uttech KM, et al. A summated score for the medication appropriateness index: development and assessment of clinimetric properties including content validity. J Clin Epidemiol 1994 Aug;47(8):891-896. [doi: 10.1016/0895-4356(94)90192-9] [Medline: 7730892]
80. Pravodelov V. Thoughtful prescribing and deprescribing. Med Clin North Am 2020 Sep;104(5):751-765. [doi: 10.1016/j.mcna.2020.06.001] [Medline: 32773043]
81. Ramaswamy R, Maio V, Diamond JJ, Talati AR, Hartmann CW, Arenson C, et al. Potentially inappropriate prescribing in elderly: assessing doctor knowledge, confidence and barriers. J Eval Clin Pract 2011 Dec;17(6):1153-1159. [doi: 10.1111/j.1365-2753.2010.01494.x] [Medline: 20630004]
82. Abd Wahab MS. The relevance of educating doctors, pharmacists and older patients about potentially inappropriate medications. Int J Clin Pharm 2015 Dec;37(6):971-974. [doi: 10.1007/s11096-015-0203-3] [Medline: 26415737]
83. Rathore A, Sharma R, Bansal P, Chhabra M, Arora M. Knowledge, attitude, and practice of medical interns and postgraduate residents on American Geriatric Society updated Beers criteria. J Educ Health Promot 2023;12:1 [FREE Full text] [doi: 10.4103/jehp.jehp_769_22] [Medline: 37034871]
84. Vandergrift JL, Weng W, Leff B, Gray BM. Geriatricians, general internists, and potentially inappropriate medications for a national sample of older adults. J Am Geriatr Soc 2024 Jan;72(1):37-47. [doi: 10.1111/jgs.18489] [Medline: 37350649]
85. Moreland B, Kakara R, Henry A. Trends in nonfatal falls and fall-related injuries among adults aged ≥65 years - United States, 2012-2018. MMWR Morb Mortal Wkly Rep 2020 Jul 10;69(27):875-881 [FREE Full text] [doi: 10.15585/mmwr.mm6927a5] [Medline: 32644982]
86. Xiong W, Wang D, Ren W, Liu X, Wen R, Luo Y. The global prevalence of and risk factors for fear of falling among older adults: a systematic review and meta-analysis. BMC Geriatr 2024 Apr 05;24(1):321 [FREE Full text] [doi: 10.1186/s12877-024-04882-w] [Medline: 38580924]

XSL•FO
RenderX

87. Rausch C, van Zon SK, Liang Y, Laflamme L, Möller J, de Rooij SE, et al. Geriatric syndromes and incident chronic health conditions among 9094 older community-dwellers: findings from the lifelines cohort study. J Am Med Dir Assoc 2022 Jan;23(1):54-9.e2 [FREE Full text] [doi: 10.1016/j.jamda.2021.02.030] [Medline: 33798484]
88. US Preventive Services Task Force, Nicholson WK, Silverstein M, Wong JB, Barry MJ, Chelmow D, et al. Interventions to prevent falls in community-dwelling older adults: US Preventive Services Task Force recommendation statement. JAMA 2024 Jul 02;332(1):51-57. [doi: 10.1001/jama.2024.8481] [Medline: 38833246]
89. Clinical resources. Centers for Disease Control and Prevention STEADI. 2024 Apr 22. URL: https://www.cdc.gov/steadi/hcp/clinical-resources/index.html [accessed 2024-06-25]
90. Xu Q, Ou X, Li J. The risk of falls among the aging population: a systematic review and meta-analysis. Front Public Health 2022;10:902599 [FREE Full text] [doi: 10.3389/fpubh.2022.902599] [Medline: 36324472]
91. Panel on Prevention of Falls in Older Persons, American Geriatrics Society and British Geriatrics Society. Summary of the Updated American Geriatrics Society/British Geriatrics Society clinical practice guideline for prevention of falls in older persons. J Am Geriatr Soc 2011 Jan;59(1):148-157. [doi: 10.1111/j.1532-5415.2010.03234.x] [Medline: 21226685]
92. Smith ML, Stevens JA, Ehrenreich H, Wilson AD, Schuster RJ, Cherry CO, et al. Healthcare providers' perceptions and self-reported fall prevention practices: findings from a large New York health system. Front Public Health 2015;3:17 [FREE Full text] [doi: 10.3389/fpubh.2015.00017] [Medline: 25964942]
93. Howland J, Hackman H, Taylor A, O'Hara K, Liu J, Brusch J. Older adult fall prevention practices among primary care providers at accountable care organizations: a pilot study. PLoS One 2018;13(10):e0205279 [FREE Full text] [doi: 10.1371/journal.pone.0205279] [Medline: 30307974]
94. Ortmann N, Haddad YK, Beck L. Special report from the CDC: provider knowledge and practices around driving safety and fall prevention screening and recommendations for their older adult patients, DocStyles 2019. J Safety Res 2023 Sep;86:401-408. [doi: 10.1016/j.jsr.2023.06.011] [Medline: 37718068]
95. Davenport K, Cameron A, Samson M, Sri-On J, Liu SW. Fall prevention knowledge, attitudes, and behaviors: a survey of emergency providers. West J Emerg Med 2020 Jul 10;21(4):826-830 [FREE Full text] [doi: 10.5811/westjem.2020.4.43387] [Medline: 32726252]
96. Meekes WM, Leemrijse CJ, Weesie YM, van de Goor IA, Donker GA, Korevaar JC. Falls prevention at GP practices: a description of daily practice. BMC Fam Pract 2021 Sep 21;22(1):190 [FREE Full text] [doi: 10.1186/s12875-021-01540-7] [Medline: 34548022]
97. Siontis KC, Attia ZI, Asirvatham SJ, Friedman PA. ChatGPT hallucinating: can it get any more humanlike? Eur Heart J 2024 Feb 01;45(5):321-323. [doi: 10.1093/eurheartj/ehad766] [Medline: 38088452]
98. Hatem R, Simmons B, Thornton JE. A call to address AI "hallucinations" and how healthcare professionals can mitigate their risks. Cureus 2023 Sep;15(9):e44720 [FREE Full text] [doi: 10.7759/cureus.44720] [Medline: 37809168]
99. Sallam M, Barakat M, Sallam M. A preliminary checklist (METRICS) to standardize the design and reporting of studies on generative artificial intelligence-based models in health care education and practice: development study involving a literature review. Interact J Med Res 2024 Feb 15;13:e54704 [FREE Full text] [doi: 10.2196/54704] [Medline: 38276872]

Abbreviations

AI: artificial intelligence
CDC: Centers for Disease Control and Prevention
METRICS: Model, Evaluation, Timing, Range and Randomization, Individual Factors, Counts, and Specificity of Prompts and Language
PIM: potentially inappropriate medication
UCLA: University of California, Los Angeles
USMLE: United States Medical Licensing Examination

Edited by A Mavragani; submitted 30.06.24; peer-reviewed by J Mira, P Heyn; comments to author 17.10.24; revised version received 26.10.24; accepted 17.11.24; published 03.01.25.

Please cite as:
Cheng HY
ChatGPT's Attitude, Knowledge, and Clinical Application in Geriatrics Practice and Education: Exploratory Observational Study
JMIR Form Res 2025;9:e63494
URL: https://formative.jmir.org/2025/1/e63494
doi:10.2196/63494
PMID:39752214

XSL•FO
RenderX

©Huai Yong Cheng. Originally published in JMIR Formative Research (https://formative.jmir.org), 03.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Digital Parenting Intervention With Intimate Partner Violence Prevention Content: Quantitative Pre-Post Pilot Study

Moa Schafer[1], DPhil; Jamie Lachman[1,2], DPhil; Paula Zinser[1], MSc; Francisco Antonio Calderón Alfaro[1], MSc; Qing Han[1], PhD; Chiara Facciola[3], PhD; Lily Clements[3], PhD; Frances Gardner[1], DPhil; Genevieve Haupt Ronnie[2], MA; Ross Sheil[4], BSc

[1]Centre for Evidence Based Intervention, Department of Social Policy and Intervention, University of Oxford, Oxford, United Kingdom

[2]Centre for Social Science Research, University of Cape Town, Cape Town, South Africa

[3]Innovations in Development, Education, and the Mathematical Sciences (IDEMS) International, Reading, United Kingdom

[4]UNICEF Jamaica, Kingston, Jamaica

Corresponding Author:
Moa Schafer, DPhil
Centre for Evidence Based Intervention
Department of Social Policy and Intervention
University of Oxford
Barnett House
32-37 Wellington Square
Oxford, OX1 2ER
United Kingdom
Phone: 44 01865 2703
Email: moa.schafer@spi.ox.ac.uk

Abstract

Background: Intimate partner violence (IPV) and violence against children are global issues with severe consequences. Intersections shared by the 2 forms of violence have led to calls for joint programming efforts to prevent both IPV and violence against children. Parenting programs have been identified as a key entry point for addressing multiple forms of family violence. Building on the IPV prevention material that has been integrated into the parenting program ParentText, a digital parenting chatbot, this pilot study seeks to explore parents' engagement with the IPV prevention content in ParentText and explore preliminary changes in IPV.

Objective: This study aimed to assess parents' and caregivers' level of engagement with the IPV prevention material in the ParentText chatbot and explore preliminary changes in experiences and perpetration of IPV, attitudes toward IPV, and gender-equitable behaviors following the intervention.

Methods: Caregivers of children aged between 0 and 18 years were recruited through convenience sampling by research assistants in Cape Town, South Africa, and by UNICEF (United Nations Children's Fund) Jamaica staff in 3 parishes of Jamaica. Quantitative data from women in Jamaica (n=28) and South Africa (n=19) and men in South Africa (n=21) were collected electronically via weblinks sent to caregivers' phones using Open Data Kit. The primary outcome was IPV experience (women) and perpetration (men), with secondary outcomes including gender-equitable behaviors and attitudes toward IPV. Descriptive statistics were used to report sociodemographic characteristics and engagement outcomes. Chi-square tests and 2-tailed paired dependent-sample *t* tests were used to investigate potential changes in IPV outcomes between pretest and posttest.

Results: The average daily interaction rate with the program was 0.57 and 0.59 interactions per day for women and men in South Africa, and 0.21 for women in Jamaica. The rate of completion of at least 1 IPV prevention topic was 25% (5/20) for women and 5% (1/20) for men in South Africa, and 21% (6/28) for women in Jamaica. Exploratory analyses indicated significant pre-post reductions in overall IPV experience among women in South Africa (P=.01) and Jamaica (P=.01) and in men's overall harmful IPV attitudes (P=.01) and increases in men's overall gender-equitable behaviors (P=.02) in South Africa.

Conclusions: To the best of our knowledge, this is the first pilot study to investigate user engagement with and indicative outcomes of a digital parenting intervention with integrated IPV prevention content. Study findings provide valuable insights into user interactions with the chatbot and shed light on challenges related to low levels of chatbot engagement. Indicative results

XSL•FO
RenderX

suggest promising yet modest reductions in IPV and improvements in attitudes after the program. Further research using a randomized controlled trial is warranted to establish causality.

(JMIR Form Res 2025;9:e58611) doi:10.2196/58611

KEYWORDS

intimate partner violence; SMS text messaging; chatbot; user engagement; parenting; violence; mobile phone

Introduction

Background

Violence against women (VAW) and violence against children (VAC) are global issues with severe, long-lasting consequences, which affect individuals and communities worldwide. Global reports have revealed alarmingly high rates of both forms of violence [1,2]. Prevalence estimates of intimate partner violence (IPV), which is the most common form of VAW, have found that >27% of ever-partnered women aged >15 years have experienced physical IPV, sexual IPV, or both at least once in their life [1]. Reports of VAC are also of great concern, with systematic review findings suggesting that >50% of all children worldwide, that is, >1 billion children globally, have experienced past-year violence [3,4]. Research findings also indicate that prevalence rates are higher for both forms of violence in low- and middle-income countries [5,6].

A growing body of evidence shows that IPV and VAC frequently cooccur in the same families [7]. In recent years, the intersections between IPV and VAC have received increasing attention, with a growing demand for prevention efforts that target both forms of violence concurrently [8]. Intersections shared by both IPV and VAC include common risk factors, such as acceptability of family violence and parental history of physical abuse [9,10], similar short- and long-term consequences, including mental and physical health problems [11], and various intergenerational effects [12]. In the past, despite their overlaps, the fields of IPV and VAC prevention research have predominantly been separated [8]. Therefore, initiatives that seek to address these common risk factors may help shift violence prevention efforts from a siloed approach to one that targets multiple types of violence simultaneously. By targeting multiple and intergenerational forms of violence, this strategy is likely to be more sustainable long-term, helping to reduce costs and increasing opportunities for scale-up and impact [13].

In the field of IPV prevention, researchers and practitioners are increasingly adopting what is known as gender-transformative approaches in programs that aim to address multiple forms of violence, that is, both VAW and VAC [14]. More specifically, gender-transformative approaches can be defined as those that aim to address harmful social and gender norms and are "designed specifically to encourage men and boys to adopt and enact gender-equitable, nonviolent attitudes and behaviors" [15]. Many programs that adopt this approach include activities and modules that focus on topics such as improving interpersonal skills, redefining responsibilities in the home, and changing perceptions around gender roles and violence [14]. Emerging evidence suggests that gender-transformative approaches offer promising results in preventing IPV perpetration, shifting restrictive gender norms, and promoting gender-equitable behaviors and attitudes [16-18]. Gender-transformative approaches are being adopted in family and sexual health interventions, further highlighting the shift toward a more holistic approach toward violence prevention and recognition of the intergenerational elements of family violence and shared risk factors [19,20].

Alongside gender-transformative interventions, in the field of VAC, parenting programs have been identified as a major strategy for reducing and preventing violence [21]. Findings from a recent systematic review of parenting programs for reducing child maltreatment in low- and middle-income countries identified several promising results across child, parent, and family outcomes, with meta-analyses results revealing reductions in child maltreatment and harsh parenting [22]. One program that has integrated both VAC and IPV prevention content includes the gender-transformative Bandebereho program in Rwanda, which engaged men and partners with the aim of encouraging healthier couple relationships and enhancing caregiving skills [23]. The program used fatherhood as an avenue to promote gender equality and foster positive shifts in men's relations with both their children and partners [18]. A randomized controlled trial (RCT) of the program found promising short- and long-term effects, with reductions in IPV and physical punishment of children at both 21 months posttest and at a 6-year follow-up [18].

Despite the promising potential of parenting programs to tackle multiple forms of family violence, scaling up parenting interventions remains a challenge [24]. To address barriers to scale, digital parenting interventions have been gaining momentum as a means to increase reach, engagement, and accessibility [25]. One such program, developed by Parenting for Lifelong Health [26,27], is a chatbot intervention called ParentText that provides caregivers with social learning–based parenting content that seeks to improve parent-child interaction and prevent child maltreatment [28]. Given the emerging evidence based on the potential to integrate parenting and gender-transformative approaches to prevent multiple forms of family violence, this study seeks to explore how users engage with a parenting chatbot that contains integrated IPV prevention material. Building on the development of IPV prevention material designed and integrated into the ParentText intervention in 2021 [28], the purpose of this pilot study was to examine user engagement with the IPV prevention content in the chatbot as well as explore preliminary changes in IPV.

Objectives

The overarching aim of this pilot study is 2-fold. First, it seeks to assess parents' and caregivers' level of engagement with the integrated IPV prevention material in the ParentText chatbot.

XSL•FO
RenderX

Second, it aims to conduct exploratory analyses of preliminary pre-post changes in experiences and perpetration of IPV and related outcomes, including attitudes toward IPV and gender-equitable behaviors following the intervention.

Consequently, this study seeks to answer the following research questions:

1. What is the level of user engagement with the IPV prevention content integrated into the parenting chatbot, ParentText, among parents and caregivers with low income in South Africa and Jamaica?
2. What, if any, are the preliminary changes in experiences and perpetration of IPV (primary outcomes) and related risk factors, including attitudes toward IPV and gender-equitable behaviors (secondary outcomes) among caregivers with low income in South Africa and Jamaica, after using the ParentText chatbot?

Methods

Study Setting

The study was conducted in 2 countries where IPV is particularly widespread: South Africa and Jamaica. Despite differing in their cultural contexts and geographical regions, South Africa and Jamaica share similar challenges surrounding high rates of IPV, making them comparable settings for exploring the integration of IPV prevention strategies in a parenting chatbot. South Africa has one of the highest rates of IPV globally, with studies revealing prevalence rates of 20% to 50% in which women reported having experienced IPV sometime during their life [29-31]. In Jamaica, rates of VAW are also alarming, with 24% of women having experienced physical violence, sexual violence, or both, from an intimate partner sometime in their life [32]. These high rates of violence underscore that urgent prevention efforts that seek to prevent and reduce violence are crucial. Moreover, research findings also indicate a high prevalence of patriarchal social norms and restrictive beliefs around women's roles in both Jamaica [33] and South Africa [34], further highlighting the harmful sociocultural norms prevalent in both contexts and the overlapping challenges that both countries face. In Jamaica, this study was conducted in the parishes: Kingston, St Catherine, and St Elizabeth; and in South Africa, this study was conducted in Cape Town.

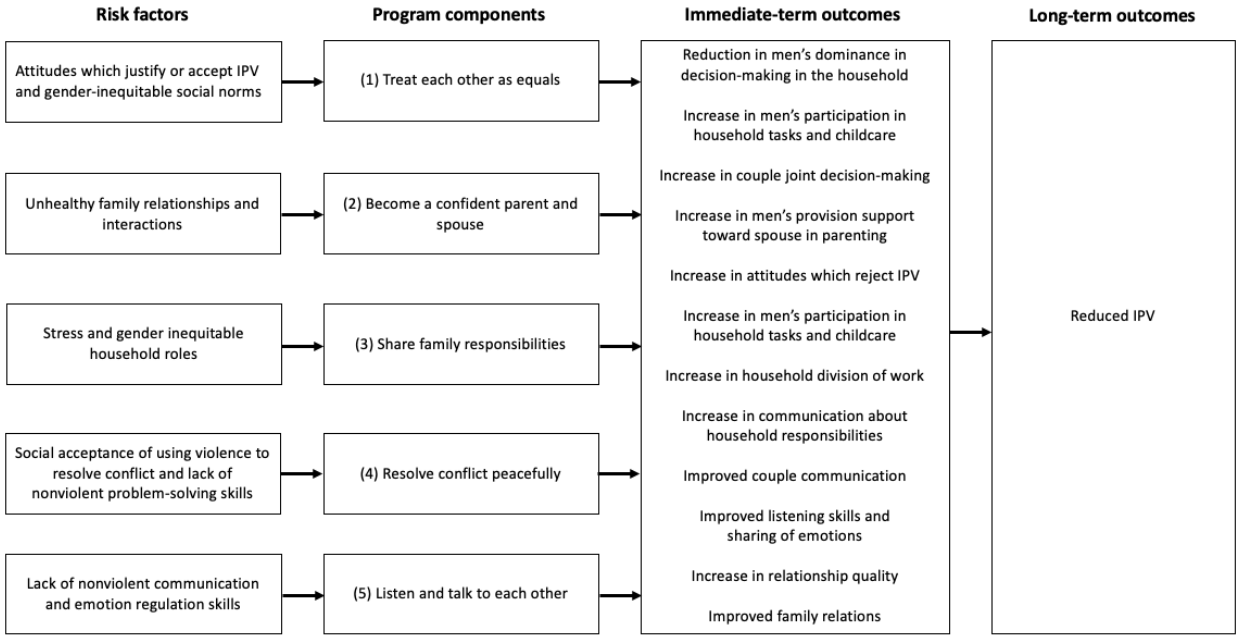
Participants and Recruitment

In Cape Town, South Africa, participants were recruited through convenience sampling by local research assistants through the organization Clowns Without Borders South Africa. Recruitment mainly took place in neighborhoods, townships, and communities in urban Cape Town. In Jamaica, recruitment also took place through convenience sampling by members of UNICEF (United Nations Children's Fund) Jamaica staff members in partnership with the National Parenting Support Commission, an agency of the Jamaican Ministry of Education and Youth. Participants were recruited primarily at schools and at women's centers for adolescent mothers in urban and rural communities. While the aim was to recruit both men and women, due to resource constraints, in the end, only women were recruited in Jamaica. Furthermore, while originally, the study sought to only include parents aged >18 years, following an interest expressed by UNICEF Jamaica and the Jamaican National Parenting Support Commission to address challenges faced by adolescent parents in the country, additional procedures were put in place (refer to the Ethical Considerations section for more details) to include adolescent parents aged >16 years in Jamaica. Unfortunately, due to staff and resource limitations in Cape Town, it was only possible to recruit parents aged >18 years in South Africa.

Consequently, the inclusion criteria for participants in this study were as follows: (1) age >18 years in South Africa and >16 years in Jamaica (where adolescent parents aged >16 years were also recruited), (2) currently caring for a child aged between 0 and 17 years, (3) being in a relationship (defined as either having a partner or being married—refer to Multimedia Appendix 1 for further details), (4) having access to a phone that could receive messages on WhatsApp (Meta Platforms) or Telegram (Telegram Messenger Inc) and that could connect to 3G or the internet, (5) provided consent to participate in the study, and (6) being able to speak either English or isiXhosa (South Africa only). Exclusion criteria were participants who were not currently in a partnered relationship and participants who were not caregivers, parents, or currently caring for a child. As an initial pre-post evaluation, this study was not designed to detect significant intervention effects but rather to assess user engagement and conduct exploratory analyses of intervention outcomes [35]. Due to funding limitations, the sample size of the study was restricted to 86 participants, with some lost due to drop out (Figure 1). Nevertheless, the study used a G*Power 3 calculator with a sensitivity power analysis to calculate the effect size (Cohen *d*) needed to obtain a significant intervention effect. Input parameters included the use of 2-tailed paired *t* tests based on the study's primary outcomes. Assuming a type I error of $P<.05$, 80% power, this sample size was sufficiently powered to detect a moderately significant intervention effect of Cohen $d=0.67$.

XSL•FO
RenderX

Figure 1. Theory of change for the ParentText intimate partner violence (IPV) prevention content, including risk factors, program components, immediate-term outcomes, and long-term outcomes.



Intervention

ParentText is a chatbot intervention for parents and caregivers of children aged between 0 and 17 years [28]. The technical architecture is open source and available for developers through GitHub. The digital parenting intervention sends automated messages to users through social messaging platforms, such as WhatsApp, Telegram, and Facebook messenger (Meta Platforms), and is also available via SMS text messaging for individuals without smartphone access (for an example of program components and example messages, refer to Figures 1-3 and Multimedia Appendix 2). The parenting content of ParentText is derived from the in-person Parenting for Lifelong Health programs [26,27]. Enrolled users receive 23 days of daily ParentText messages personalized for their chosen child's developmental stage (0-23 months, 2-9 years, and 10-17 years). Content is provided using a range of formats, including text, image, video, and audio. Structured and sequential parenting content is derived from common elements of social learning theory-based programs along two main themes as follows: (1) positive relationship building and (2) limit setting and nonviolent discipline [36]. In addition, content specifically targeting positive partner relationships and IPV prevention for users in partnered relationships was integrated into the chatbot by drawing on gender-transformative material from a range of interventions found to reduce IPV and improve gender-equitable attitudes [23,37,38]. It is the engagement with this content that is the primary focus of this study. The IPV prevention content in ParentText is divided into 5 topics (refer to program components in Figure 1). The proposed causal pathway between each IPV prevention topic, identified risk factor, and behavior change domain is illustrated in the theory of change in Figure 1. Schafer et al [28] and Multimedia Appendix 3 give further details on the development of the IPV prevention content and theory of change. Additional information and examples on how parents can interact with the chatbot, along with examples of the sequencing of the ParentText messages, are available in Figures 2 and 3.

XSL•FO
RenderX

Figure 2. Flow of the intimate partner violence (IPV) prevention content text messages in ParentText, which illustrates the flow of text messages that users in ParentText receive, including introduction text messages and text messages providing options to view further IPV prevention content.

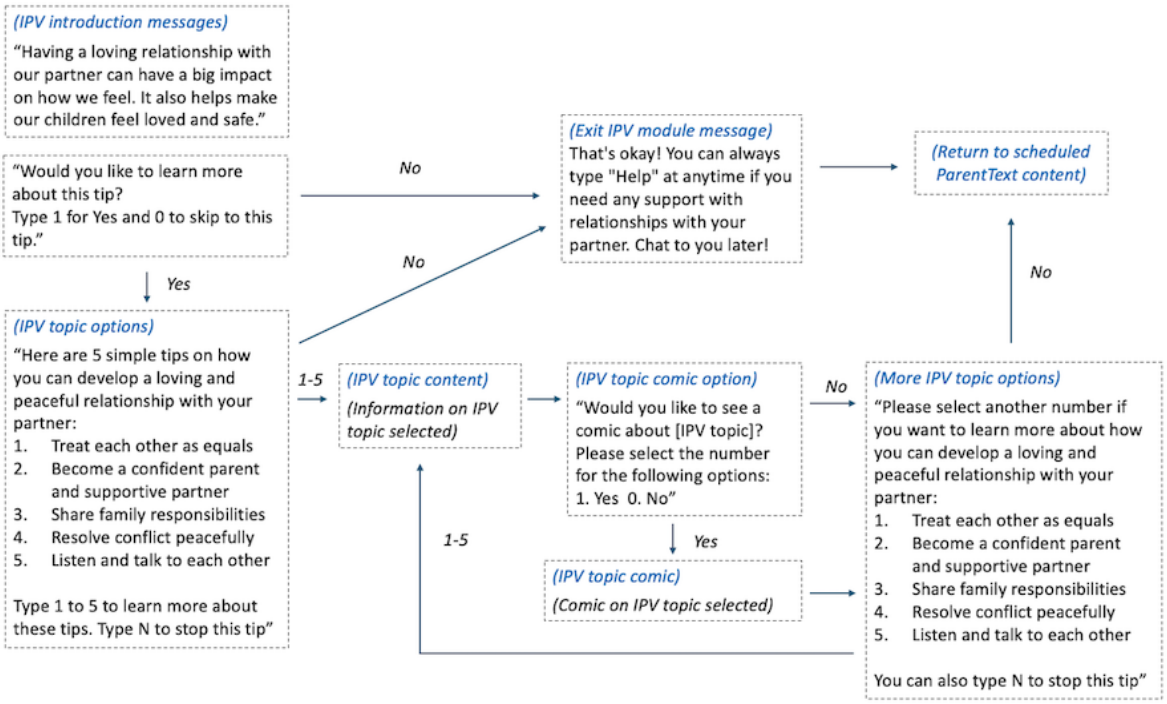
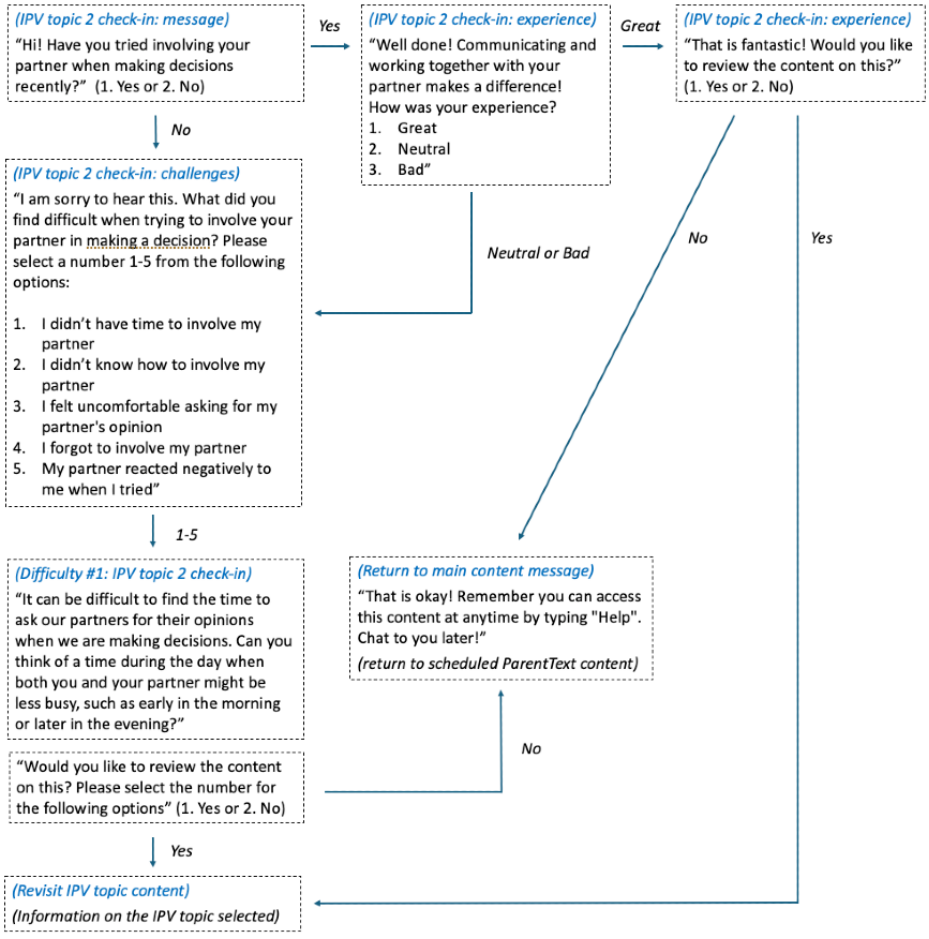


Figure 3. Flow of the check-in message on the intimate partner violence (IPV) topic 2 "joint decision-making," which illustrates the flow of the check-in text message that users in ParentText receive to provide an opportunity to reflect on the topic of "joint decision-making.".

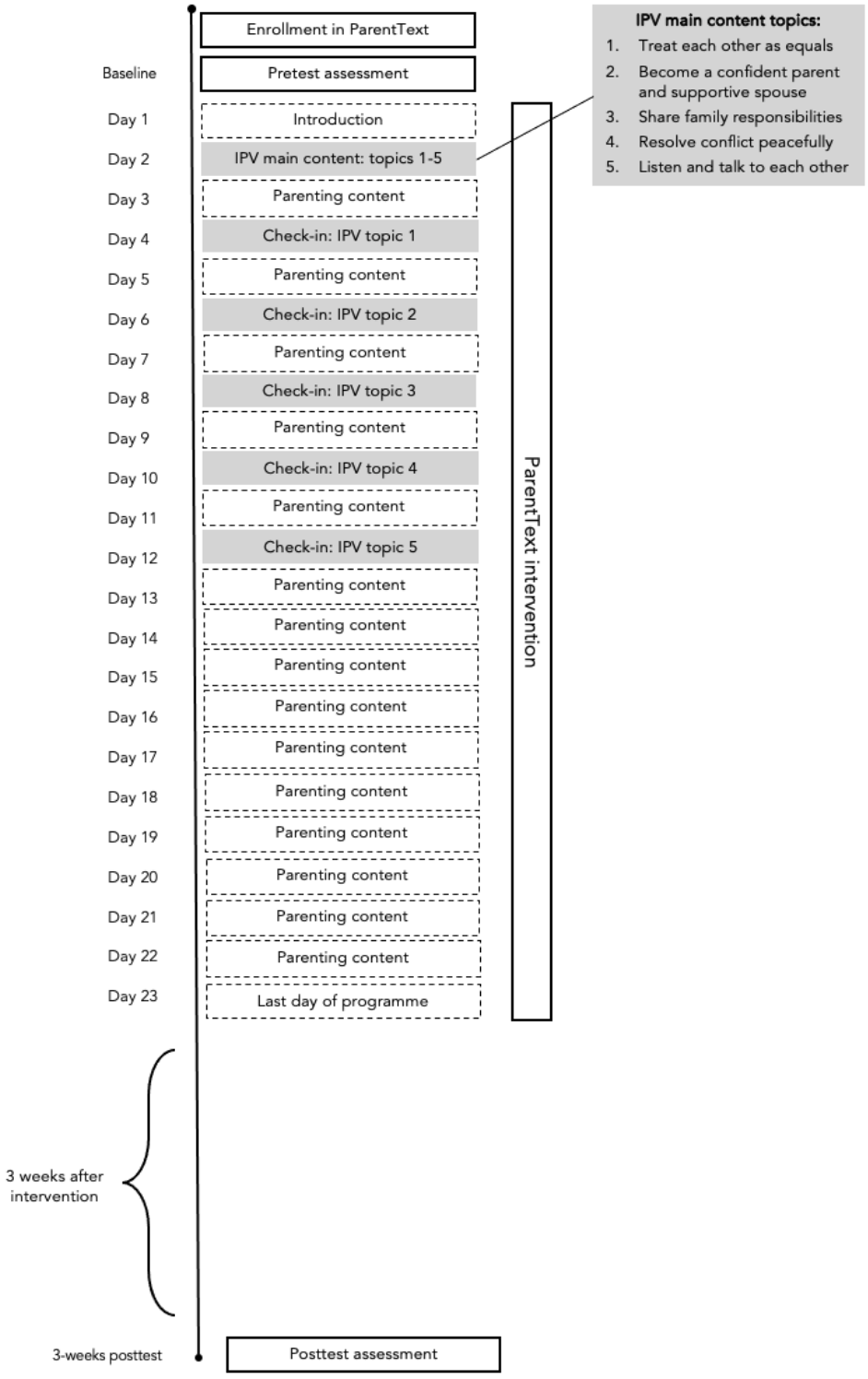


XSL•FO
RenderX

Design and Procedures

This study used a repeated-measures, single-arm design. Pre and posttest assessments were conducted at baseline (preintervention) and at 6 weeks after baseline (ie, 3 weeks after the end of the intervention) using the application Open Data Kit (ODK; refer to the Measures section and Multimedia Appendix 1 for more details). The posttest was planned as a 1-month posttest (1 month after the end of intervention); however, due to fieldwork challenges that arose and time constraints, it needed to be changed to 6 weeks after baseline (Figure 4).

Figure 4. Overview of the ParentText chatbot timeline, illustrating what content is delivered on what days of the program from day 1 to day 23, and demonstrating when the 3-week posttest assessment is delivered. IPV: intimate partner violence.



Reporting follows the CONSORT (Consolidated Standards of Reporting Trials) 2010 guidelines extension to randomized pilot and feasibility trials [39] (refer to Multimedia Appendix 4 for the CONSORT checklist). Recruitment, delivery, and

XSL•FO
RenderX

assessments in South Africa and Jamaica were conducted on a rolling basis from September 2022 to March 2023. Participants who enrolled in the intervention received data credit using a local service provider to ensure they would have access to 3G during the intervention. During recruitment, participants received a message that included a link to the program embedded within a WhatsApp or Telegram business account. After clicking the link, participants were directed to the ParentText chatbot account and asked to select their preferred language for the program. In keeping with recommendations from the formative research [28], the language was set to English in Jamaica, and the options were English and isiXhosa in South Africa. Participants were asked if they consented to participate in the intervention and evaluation and were provided with a link to a more detailed information sheet. Participants who did not consent exited the intervention. In alignment with guidelines for conducting statistical surveys on VAW [40], the survey included skip patterns, which determine the eligibility of specific questions that respondents are asked [40,41]. For example, in this study, men who responded that their partner was taking part in the intervention were not asked the IPV items in the IPV assessment and were instead only asked the questions on attitudes toward IPV and gender roles and on gender-equitable behaviors. Further information about this is provided in the Ethical Considerations section.

At baseline, participants were invited to complete the IPV assessment, which was optional, on an external server via Oxford's Linux virtual machine using ODK to ensure responses were not saved on their mobile devices. Participants provided their phone number when completing the IPV survey for their responses to be recorded and linked with a unique user ID. These data were then deidentified, exported as CSV files, and uploaded to a secure server in Oxford, which was only accessible by the research team and password protected. All personal identifying data were deleted once endline data collection was completed. Participants were invited to respond to the IPV assessment at 3 weeks postintervention via the phone number they provided when completing the IPV assessment at baseline.

Measures

The user engagement, demographic, primary outcomes, and secondary outcomes are provided in Table 1.

XSL•FO
RenderX

Table 1. Outcome measurements in the study, including user engagement, demographics, background variables, and primary and secondary outcomes, with descriptions of the study measurement used and the number of items used for each outcome.

Outcomes	Study measurement	Items, n
User engagement outcomes		
Number of days in the ParentText program	Measured by the final day the participant used the chatbot subtracted by the first day participant used the chatbot (including inactive days)	1
Number of active days the user actively interacted with the chatbot	Number of days participants actively interacted with ParentText as measured by chatbot interactions (excluding inactive days)	1
Overall chatbot interaction rate	The number of interactions logged per day divided by the total number of days participants actively interacted with ParentText	1
Participant engagement with each IPV[a] topic	The number of participants who viewed each IPV topic, in line with existing digital parenting intervention research where module viewing is frequently used as a proxy measure for the level of participant engagement [42,43]	5
Completion rate of IPV topics	Number of IPV topics completed out of the total 5 topics (measured by the number of IPV topics participants viewed that were recorded via RapidPro, UNICEF's open source framework developed to send and receive data using mobile phones [44])	5
Overall completion rate of ParentText parenting modules	Number of parenting modules completed (measured by the number of modules viewed that were recorded via the RapidPro (UNICEF) software)	1
Primary behavioral outcomes		
Women's experiences of IPV	Items adapted from the WHO[b] VAWI[c] and the core questionnaire in the WHO-MCS[d] [45]: physical (1 item), psychological (1 item), sexual (1 item), coercion (1 item), and economic abuse (1 item). Participants asked to give a frequency score on a scale of 0 to 8 or more times. The original VAWI tool measures past-year experiences of IPV [45]; however, a 1-year reporting time frame surpassed the duration of the intervention. The reporting timeframe and response codes were therefore adapted from the past-year (once, a few times, and many times) to the number of experiences of IPV in the past month (0, 1, 2, 3, 4, 5, 6, 7, and >8) to enhance the sensitivity of measuring changes in experiences of IPV. These adjustments are similar to those made to measure child maltreatment with the ISPCAN[e] Child Abuse Screening Tool (ICAST)-Trial measurement by Meinck et al [46]. Past-month time frame when measuring IPV has been used in other similar pieces of research, such as a study by Ebert and Steinert [47], which examined experiences of IPV during COVID-19 in Germany. Responses were dichotomously recoded, where 1 indicated any experience of each subsequent type of IPV and 0 indicated no experience of each type of IPV in the past month [48].	5
Men's perpetration of IPV	Items adapted from the WHO VAWI and the core questionnaire in the WHO-MCS [45]: physical (1 item), psychological (1 item), sexual (1 item), coercion (1 item), and economic abuse (1 item). Participants were asked to give a frequency score on a scale of 0 to >8 times, with responses dichotomized into any perpetration of IPV in the past month.	5
Secondary behavioral outcomes		
Men's experiences of IPV	Two items used and adapted from Abramsky et al [49], namely: "In the past month, during any potential times that you may have used violence against your partner, did they ever fight back physically to defend themselves?" and "In the past month, have you ever been hit or physically mistreated by your partner when you were not hitting or physically mistreating them?" Respondents were asked to give a frequency score on a scale of 0 to 8 or more times, with responses dichotomized.	2
Women's perpetration of IPV	Two items were adapted from a questionnaire developed by researchers at the London School of Hygiene and Tropical Medicine for a randomized controlled trial of a violence prevention intervention in Tanzania [49]. Items included: "During any potential times that you were hit in the past month, did you ever fight back physically to defend yourself?" and "In the past month, have you ever hit or physically mistreated your partner when they were not hitting or physically mistreating you?" Respondents were asked to give a frequency score on a scale of 0 to 8 or more times, with responses dichotomized.	2
Women's positive partner interactions	Item adapted from Abramsky et al [49] asking: "How many times in the past month did your partner show you they cared and respected your feelings even though they disagreed with you?" Participants were asked to give a frequency score on a scale of 0 to 8 or more times.	1
Men's positive partner interactions	One item adapted from Abramsky et al [49], that asks, "How many times in the past month did you show your partner you cared and respected their feelings even though you disagreed?." Participants asked to give a frequency score on a scale of 0 to 8 or more times.	1

XSL•FO
RenderX

Outcomes	Study measurement	Items, n
Gender-equitable behaviors	Measured using items adapted from questionnaires used in previous violence prevention interventions [23,49]. In total, 3 items were adapted from Abramsky et al [49], which included questions on couple communication, joint decision-making, and partner conflict resolution, with questions such as: "In the past week, how many times did you and your partner talk about your worries and feelings?" and "In the past week, how many times did you and your partner make a decision together?" The fourth item was adapted from a questionnaire developed by researchers of a randomized controlled trial of a gender-transformative violence prevention intervention in Rwanda [23], which asks: "In the past week, how many times did you and your partner share housework and caregiving tasks equally?" Respondents were asked to report whether in the past week the specific behavior occurred on a scale of 0 (never) to 3 (many times). The scores were summed for all items for the overall gender-equitable behavior score. The possible overall score ranged from 0 to 12.	4
Attitudes toward gender roles and IPV	Measured using 6 items in the "Attitudes toward gender roles" section of the WHO-MCS [45], such as "A woman should obey her husband's wishes even if she disagrees." Respondents were asked to indicate whether they agree or disagree with the statements using a Likert scale ranging from 1 (strongly agree) to 5 (strongly disagree). Negative items were reverse-coded so that higher scores denoted more gender-equitable attitudes. The scores were summed for all items for the overall attitude score. The possible overall score ranged from 6 to 30.	6
Background variables		
Parent gender, age, relationship status, living with partner, education level, employment status, partner employment status, and number of children	Sociodemographic information	8 (1 item each)
Women's past-year IPV experience (measured at baseline)	Measured using 1 item adapted from the WHO VAWI and the core questionnaire in the WHO-MCS [45], which asks, "How many times, in the past 12 months, did your partner do one of the following: insult or shout at you, hit or shove you, or force you to have sex?" Respondents were asked to give a frequency score on a scale of 0 to 8 or more times. Due to the skewed response patterns in the data for IPV, responses for IPV were dichotomously recoded, where 1 indicated any experience of IPV and 0 indicated no experience of IPV in the past year [48].	1

[a]IPV: intimate partner violence.

[b]WHO: World Health Organization.

[c]VAWI: Violence Against Women Instrument.

[d]MCS: Multi-Country Study on Domestic Violence.

[e]ISPCAN: International Society for Prevention of Child Abuse and Neglect.

Statistical Analysis

Data analyses were conducted using R (R Core Team). Descriptive statistics and frequencies were used to describe the sociodemographic characteristics of participants and to report engagement outcomes, summarized using means (SDs), medians (range), and N (%). To investigate changes in outcomes between pre and posttest, McNemar chi-square tests (for dichotomous variables) and paired dependent-sample *t* tests (for continuous variables) were conducted, with corresponding 95% CIs and *P* values [50,51]. Assumptions were examined to confirm no violations existed. Data were analyzed using complete case analysis.

Ethical Considerations

Ethical procedures were followed to ensure the safety and welfare of study participants. At the start of the IPV assessment, screening procedures ensured that men and women from the same household were not interviewed about IPV. This is based on the World Health Organization VAW research guidelines [52] and ethical and safety recommendations for intervention research on VAW [45] to prevent putting women at risk. Consequently, men who responded that their partner was taking part in the intervention were not asked the IPV items in the IPV assessment and were only asked instead the questions on attitudes toward IPV and gender roles and on gender-equitable behaviors. In addition, efforts were also in place to ensure responses to the IPV questions could not be viewed by others by delivering the IPV assessment using ODK. This procedure was implemented to protect participants by ensuring no responses were saved on their mobile devices. Participants who disclosed experiences of IPV were automatically provided referrals to local services, customized to their regional context that supported individuals experiencing violence. Local referral information could also be accessed by writing "Help Me" and choosing "Other Support" in the ParentText free text field. In line with UNICEF risk communication and community engagement guidelines [53], ParentText is also formatted to detect high-risk keywords to identify potential disclosure of dangerous situations via the free text field. Following detection, ParentText automatically offers the participant an empathetic and empowering reply with referral details that are customized to the country, which supports parent and child safety (eg, hotlines, police, and ambulance).

XSL•FO
RenderX

Further ethical considerations that were taken within the research project include recognitions surrounding power and limitations of positionality [54]. For instance, it is worth noting that while collaborators and members of the research team came from several countries, the primary author of this paper is from and is part of a research institution in the global north. Efforts to address limitations associated with this positionality were put in place, including integrating reflexivity across the stages of the research process [55]. For example, during the research design and data collection planning phase, members of local grass-roots organizations, communities, and stakeholders in Jamaica and South Africa were continuously consulted to review and revise the IPV prevention content and assessment materials [28]. Integrating these embedded consultations and opportunities for research amendments with members of the local community was a pivotal part of the research project, particularly given the first author's positionality as an "outsider" [55].

Ethics approval was obtained by the Department Research Ethics Committee at the University of Oxford (Central University Research Ethics Committee 2 Ref No: R69569/RE009). All methods were carried out in accordance with relevant guidelines and regulations. Informed consent to participate in the study was obtained from all participants before enrolling in the study and the intervention.

Results

Participants Characteristics

Participant characteristics are provided in Table 2 by country and with an overall summary. A study flow diagram illustrating the flow of participants through the trial is provided in Figure 5. The average age of participants across both countries was 34.4 (SD 8.74; range 16-57) years. The average age of women in South Africa was 34.9 (SD 6.77; range 25-49) years and men 28.8 (SD 2.68; range 25-36) years. In Jamaica, where only women were recruited into the study, the average age of women was 39.9 (SD 11.52; range 16-57) years. Most participants, across both countries, were partnered but not married (48/68, 71%). On average, participants had 2.23 (SD 1.19) children. Approximately half of the participants across both countries had a secondary education (36/68, 53%), and 7% (5/68) had only primary education or less. In South Africa, among women participants, 42% (8/19) were working and 58% (11/19) were unemployed or looking for work. Among men, slightly more were working (13/21, 62%), with 29% (6/21) unemployed or looking for work. In Jamaica, among women participants, this was similar, with 43% (12/28) working and 54% (15/28) unemployed or looking for work. Across both countries, the employment of partners was high (50/68, 74%), with only 18% (12/68) with partners who were unemployed or looking for work. Experience of IPV (women-report only due to safety reasons) at baseline was high: 21% (4/19) of women participants in South Africa and 29% (8/28) of women in Jamaica had experienced some form of IPV in the past year, respectively.

XSL•FO
RenderX

Table 2. Demographic characteristics of study participants in South Africa and Jamaica.

	South Africa		Jamaica	Overall
	Women (n=19)	Men (n=21)	Women (n=28)	Total (N=68)
Gender, n (%)				
Women	19 (48)	N/A[a]	28 (100)	47 (69)
Men	N/A	21 (53)	—[b]	21 (31)
Age (y), mean (SD)	34.9 (6.77)	28.8 (2.68)	39.9 (11.52)	34.4 (8.74)
Relationship, n (%)				
Married	6 (32)	2 (10)	12 (43)	20 (29)
Partnered but not married	13 (68)	19 (90)	16 (57)	48 (71)
Number of children, mean (SD)	2.53 (1.17)	1.19 (0.68)	2.61 (1.25)	2.23 (1.19)
Education, n (%)				
Primary	1 (5)	0 (0)	4 (14)	5 (7)
Secondary	9 (47)	10 (48)	17 (61)	36 (53)
Higher	9 (47)	11 (52)	7 (25)	27 (40)
Employment, n (%)				
Working	8 (42)	13 (62)	12 (43)	33 (49)
Unemployed or looking for work	11 (58)	6 (29)	15 (54)	32 (47)
Student	0 (0)	2 (10)	0 (0)	2 (3)
Retired	0 (0)	0 (0)	1 (4)	1 (2)
Partner's employment, n (%)				
Working	11 (58)	17 (81)	22 (79)	50 (74)
Unemployed or looking for work	7 (37)	2 (10)	3 (11)	12 (18)
Student	0 (0)	2 (10)	0 (0)	2 (3)
Retired	0 (0)	0 (0)	2 (7)	2 (3)
Refuse to answer	1 (5)	0 (0)	1 (4)	2 (3)
Experience of IPV[c], n (%)				
Experienced IPV in the past year[d]	4 (21)	X[e]	8 (29)	12 (26)[f]

[a]N/A: not applicable.

[b]No men were recruited in Jamaica.

[c]IPV: intimate partner violence.

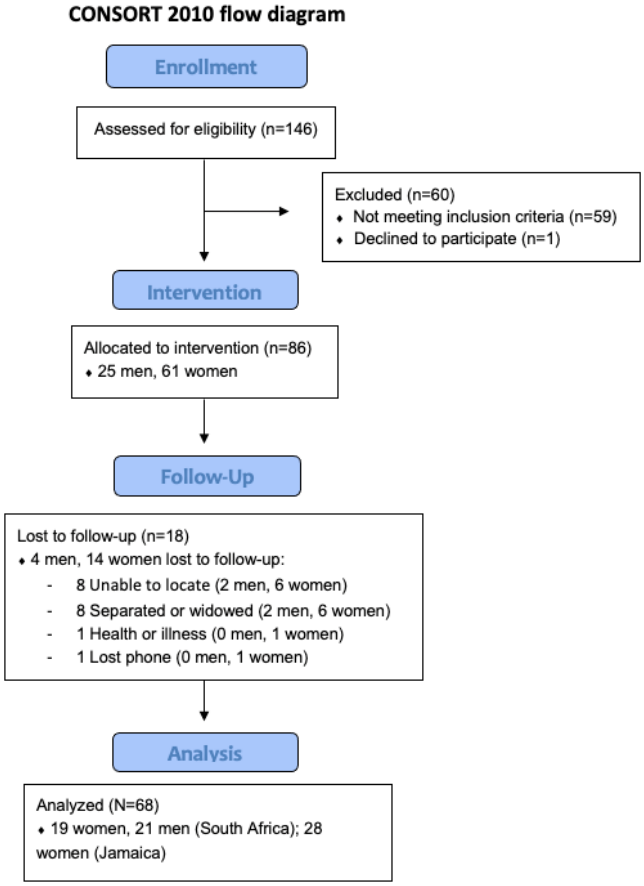
[d]For safety reasons, men were not asked this question at baseline.

[e]Men were not asked this question at baseline due to safety reasons.

[f]Overall IPV here only includes women.

XSL•FO
RenderX

Figure 5. Study flow diagram illustrating the flow of participants through the trial. CONSORT: Consolidated Standards of Reporting Trials.



User Engagement Outcomes

Overall engagement outcomes are provided in Table 3. The ParentText program lasted 23 days, with the main IPV prevention content delivered on day 2 and check-in messages delivered within the first 12 days (Multimedia Appendix 2 provides an overview of the program content delivered each day). In terms of engagement, in South Africa, the average number of days in the program was 13 (SD 23.90) and 3 (SD 3.78) days for women and men, respectively. The average number of days users interacted with the chatbot was 3 (SD 1.58) for women and 2.6 (SD 1.52) for men, each with an average interaction rate of 0.57 (SD 0.26) interactions and 0.59 (SD 0.17) interactions per day, respectively. In Jamaica, the average number of days in the program was 27 (SD 23.10) days, and the average total number of days users interacted with the chatbot was 5.2 (SD 5.45) days. The average interaction rate in Jamaica was 0.21 (SD 0.20) interactions per day. In terms of engagement across the 5 relationship topics, there was a greater engagement with topics in Jamaica where engagement with each topic ranged from 11% (3/28) to 29% (8/28) compared to South Africa, where it ranged from 0% to 11% (2/19). In terms of completion rates, in South Africa, 25% (5/20) and 5% (1/20) of women and men completed 1 out of the 5 IPV topics, and 5% (1/20) and 0% completed all 5 IPV topics, respectively. In comparison to the completion rate of overall ParentText modules, 16% (3/19) and 9% (2/21) of women and men completed all the parenting (ie, the non-IPV) ParentText modules in South Africa. In contrast, in Jamaica, 21% (6/28) completed 1 out of the 5 IPV topics, 11% (3/28) completed all 5 IPV topics, and 18% (5/28) completed all the parenting ParentText modules.

XSL•FO
RenderX

Table 3. ParentText user engagement outcomes of the 23-day program in South Africa and Jamaica, including overall retention, engagement per intimate partner violence (IPV) topic, and IPV topic completion rate.

Outcomes	South Africa		Jamaica
	Women (N=19)	Men (N=21)	Women (N=28)
Overall retention, mean (SD)			
Number of days in program[a]	13 (23.90)	3 (3.78)	27 (23.10)
Number of active days the user interacted with the chatbot[b]	3 (1.58)	2.6 (1.52)	5.2 (5.45)
Number of interactions per day[c]	0.57 (0.26)	0.59 (0.17)	0.21 (0.20)
Engagement per IPV topic[d]			
Participants who viewed each IPV topic, n (%)			
Relationship topic 1: “Treat each other as equals”	2 (11)	2 (10)	7 (25)
Relationship topic 2: “Become a confident parent and supportive spouse”	2 (11)	0 (0)	4 (14)
Relationship topic 3: “Share family responsibilities”	2 (11)	1 (5)	5 (18)
Relationship topic 4: “Resolve conflict peacefully”	2 (11)	1 (5)	3 (11)
Relationship topic 5: “Listen and talk to each other”	2 (11)	1 (5)	8 (29)
All relationship topics (viewed all relationship topics)	1 (5)	0 (0)	3 (11)
Completion rate of IPV topics completed, n (%)[e]			
1	5 (25)	1 (5)	6 (21)
2	0 (0)	0 (0)	3 (11)
3	0 (0)	0 (0)	1 (4)
4	0 (0)	1 (5)	0 (0)
5	1 (5)	0 (0)	3 (11)
Completion rate of ParentText parenting modules, mean (SD)	16 (0.19)	9 (0.06)	18 (0.20)

[a]Measured by the final day participant used the chatbot, subtracted by the first day the participant used the chatbot (note: this measure includes inactive days).

[b]Number of days participants actively interacted with ParentText (note: this measure excludes inactive days).

[c]Measured by the number of interactions logged per day divided by the total number of days the participant actively interacted with ParentText.

[d]Measured based on the number of participants who viewed each IPV topic.

[e]Number of IPV topics completed out of the total 5 topics.

Behavioral Outcomes

Experiences and Perpetration of IPV

Table 4 presents the descriptive results of the IPV outcomes, summarizing experiences of IPV among women and perpetration of IPV among men using past-month frequencies at pre and posttest and *P* values for each outcome. At posttest, 73% (11/15) of women in South Africa and 72% (13/18) of women in Jamaica reported experiencing some form of IPV, compared to 84% (16/19) and 96% (23/24) at baseline, reductions that were significant in South Africa (P=.01) and in Jamaica (P=.01). Experiences of all subtypes of IPV for women in both South Africa and Jamaica also showed reductions at posttest, and so did acts of self-defense and perpetration (Multimedia Appendix 5). For perpetration of IPV among men, at posttest, 75% (9/12) of men in South Africa perpetrated some form of IPV, compared to 85% (11/13) at baseline, which was nonsignificant for overall perpetration of IPV (P=.06). Analyses of subtypes of IPV, however, found significant reductions in psychological (P=.04), physical (P<.001), and sexual (P<.001) violence among men in South Africa (Multimedia Appendix 5). Detailed tables of the primary and secondary outcomes are provided in Multimedia Appendix 5.

XSL•FO
RenderX

Table 4. Primary behavioral outcomes for women's past-month experience of overall intimate partner violence (IPV) in South Africa and Jamaica and men's past-month perpetration of overall IPV in South Africa.

	Baseline overall IPV, n (%)[a]	Posttest overall IPV, n (%)	*P* value[b]
Women's experience of overall IPV (past month)			
South Africa	16/19 (84)	11/15 (73)	.01
Jamaica	23/24 (96)	13/18 (72)	.01
Men's perpetration of overall IPV (past month)			
South Africa	11/13 (85)	9/12 (75)	.06

[a]% used is proportional to the baseline and posttest samples, respectively.

[b]McNemar chi-squared *P* value.

Positive Partner Interactions

Table 5 presents outcomes on positive partner interactions measured through respectful behaviors, summarized using means and SDs for pre and posttest along with *P* values. There were no significant effects detected for change in positive partner interactions reported by men or women.

Table 5. Secondary behavioral outcomes for positive partner interactions in the past month, overall attitude toward gender roles and intimate partner violence (IPV), and overall gender-equitable behaviors in the past week for women and men in South Africa and women in Jamaica.

	Baseline, mean (SD)	Posttest, mean (SD)	*P* value (*t* test)
Positive partner interactions			
South Africa			
Women (n=15)	4.74 (2.84)	4.14 (2.98)	.57
Men (n=12)	3.92 (2.50)	5.67 (2.46)	.10
Jamaica			
Women (n=15)	5.08 (2.60)	3.89 (2.54)	.14
Overall attitude[a] toward gender roles and IPV			
South Africa			
Women (n=14)	10.11 (2.71)	10.13 (3.89)	.98
Men (n=16)	11.90 (3.82)	8.88 (2.63)	.01
Jamaica			
Women (n=15)	11.39 (3.45)	11.89 (2.35)	.56
Overall gender-equitable behaviors[b]			
South Africa			
Women (n=15)	7.37 (2.95)	7.93 (3.53)	.63
Men (n=16)	7.81 (1.94)	9.75 (1.61)	.02
Jamaica			
Women (n=15)	7.14 (3.49)	7.39 (2.85)	.80

[a]A lower score indicates less harmful attitudes toward IPV and gender roles.

[b]A higher score indicates more gender-equitable behaviors in the past week.

Attitudes Toward (Harmful) Gender Roles and IPV

Table 5 summarizes means and SDs for the attitude outcomes at pre and posttest, with *P* values for each attitude outcome also provided. Analyses indicate a significant reduction in harmful attitudes toward IPV and gender roles reported by men in South Africa (*P*=.01); however, no effects were reported among women in South Africa (*P*=.98) or in Jamaica (*P*=.56).

Gender-Equitable Behaviors

Analyses found a significant increase in gender-equitable behaviors for men (*P*=.02; Table 5). No significant effects were detected for women in South Africa (*P*=.63) or in Jamaica (*P*=.80).

XSL•FO
RenderX

Discussion

Principal Findings

To the best of our knowledge, this pre-post pilot study is the first that examines user engagement and indicative outcomes in a digital parenting intervention with integrated IPV prevention content [22]. The study offers valuable, preliminary insights across 2 countries on user engagement with a digital parenting intervention with integrated IPV prevention, as well as provides exploratory, indicative results on IPV experience and perpetration, attitudes toward IPV, and gender-equitable behavior outcomes. First, we will discuss the key findings on levels of user engagement. While the average daily interaction rate was higher in South Africa (0.57 and 0.59 interactions per day for women and men, respectively) compared to Jamaica (0.21 interactions per day), the average number of days users stayed in the program in Jamaica was greater (27 days), compared to South Africa (13 and 3 days for women and men, respectively). Due to WhatsApp messaging restrictions, "push" messaging was paused after 24 hours of user inactivity, until user activity resumed. Given research findings that highlight the importance of messaging reminders in digital interventions [56], this restriction may have contributed to some of the low engagement levels observed.

The completion rate of engaging with at least 1 IPV topic in the program was lower than expected across the intervention, with completion rates of 5% (1/20) for men and 25% (5/20) for women in South Africa, and 21% (6/28) for women in Jamaica. The completion rates of all 5 IPV topics were much lower, with 0% of men and 5% (1/20) of women in South Africa, and 11% (3/28) of women in Jamaica completing all 5 IPV prevention topics. Even though user engagement and retention rates have been noted as a challenge in digital interventions [57], completion rates in this study were lower than anticipated in comparison to findings in the field. For example, a review of digital interventions for parents with young children found that attrition rates in digital parenting programs ranged from 30% to 50% [58]. Accordingly, the low completion rates of the IPV topics observed in ParentText raise questions about whether user engagement in the program was sufficient to change complex behaviors and deeply rooted attitudes surrounding violence and gender roles.

There are various possible reasons for the low levels of engagement with the IPV content observed in ParentText. First is the format of the program schedule. On day 2 of the ParentText program, participants were provided with a list of all 5 IPV prevention topics, which they could choose to explore. After selecting a topic to view, users were given the option to view one of the other IPV topics. However, it is possible that viewing >1 IPV prevention topic per day may have been overwhelming for users. Therefore, this may have inhibited users from engaging with multiple IPV topics in the program. Another element that may have affected engagement was the timing of the program "check-in" reminders. Participants received "check-in" messages in the latter part of the program (days 4-12; Multimedia Appendix 2), where they were encouraged to revisit and explore the IPV topics they had not yet engaged with. However, given the already low average number of days for which the participants remained in the program, it is likely that many participants did not receive these messages. These findings raise important questions on how to increase engagement with the IPV modules integrated into the ParentText. For example, amending the timing of content delivery might be one way to increase engagement. For example, a study assessing a text-messaging intervention focusing on promoting health relationships found that the timing of message delivery played a key role in program engagement and that if delivered at the wrong time, users might be less likely to engage with the text messages [59].

Taken together, these study results give valuable insight into how users engage with IPV content integrated into a chatbot and suggest that rather than interacting with the chatbot content daily, users appear to interact with the chatbot at their own pace. Consequently, providing more options for user-led selection of program content would be important for future iterations of the chatbot to ensure users can access and revisit the information they need more readily. The importance of user-led content delivery has also been underscored by other digital parenting intervention research. For instance, findings from a study on the digital fatherhood program "SMS4dads" revealed that the length of time between users receiving a text message with a link to program content to that of users clicking the link was, on average, 2.11 (SD 3.94) days [60]. This highlights how engagement with digital intervention content is often not instantaneous, but rather asynchronous and driven by users' schedules [60]. Accordingly, in future versions of ParentText, combining "push" notification reminders with user-led selection of program content might help increase user engagement. Notably, since the present trial of ParentText in South Africa and Jamaica, newer versions of the WhatsApp settings have been released, which allow future iterations of the program to override the 24-hour limit inhibiting "push" messages after user inactivity.

More broadly, the relatively low levels of user engagement found in this pilot study raise questions as to whether a digital modality is sufficient when delivering intervention content related to complex gender-equitable beliefs and behaviors. It may be the case that a hybrid approach, combining both digital content delivery and in-person sessions, might be necessary. For instance, a study on enhancing father engagement in parenting programs revealed that in terms of intervention delivery, fathers prefer less intensive and low-dose programs, including internet - based interventions [61]. However, a recent systematic review of interventions promoting gender equality noted that in various studies it was found that dialog was critical for driving change in gender norms [62]. The importance of dialog is further underscored by research examining men's engagement in family and health interventions, revealing that men value and often seek the opportunity to connect with other men to discuss parenting and fatherhood, even though cultural norms sometimes hinder this. Together, these findings suggest that a combination of digitally delivered content and in-person discussion opportunities might be needed. Various studies also suggest that using face-to-face recruitment and participatory, peer-engaged methods is an effective approach for engaging

XSL•FO
RenderX

parents, particularly fathers [63,64]. Research findings also suggest that providing opportunities to critically discuss and engage in dialog with others, especially on difficult topics, plays a valuable role not only in terms of engagement but also in terms of creating change and shifting harmful gender attitudes [17,62,64]. Taken together, it may be possible that low-intensity, 1-way engagement with a chatbot is insufficient to create change in more deeply rooted behaviors and attitudes among users and that more interactive engagement is needed to shift behaviors and beliefs. While these hybrid options and amendments may improve engagement, it should be noted that they will also increase intervention costs and limit scalability, which are some of the benefits of digital-only interventions [24].

It is also possible that the low level of user engagement with the IPV prevention topic was due to other factors, such as the format of the program. The IPV content in ParentText was only delivered on 1 specific day, after which only check-in messages were sent (refer to Figure 4 for an overview of the content delivery schedule). Receiving the option to view all the IPV content on the same day may potentially have been overwhelming for participants. This may have reduced user engagement rates due to intervention fatigue [65], which is cognitive or emotional weariness resulting from competing demands of intervention engagement and other burdens in daily life [66] and has been linked to intervention adherence. Considerations for future iterations of ParentText include (1) delivering all the content automatically (since optional and supplemental content in digital interventions is often associated with low engagement [67]) and (2) adjusting the timing of content and spreading out the delivery of the IPV content to allow mental rests in between new material, strategies that have been suggested to reduce intervention fatigue [65]. Notably, this is already an amendment that future iterations of ParentText are incorporating and that is currently being implemented in a trial of an updated version of ParentText in Malaysia [68]. This integrated approach of implementing IPV prevention content beyond a single module is also a strategy that is favored and advocated by gender-transformative approaches and is one that is being advocated more broadly in the parenting field [69].

The exploratory analyses of the behavioral outcomes, while only preliminary, are also important to discuss. In terms of IPV, women in Jamaica reported a significant reduction in overall experiences of IPV at posttest. These tentative findings are in line with results from other parenting and gender-transformative programs that have found reductions in both IPV experience at posttest and even at follow-up [18,23,37]. For example, findings from an RCT of a gender-transformative intervention in Rwanda found that, compared to the control, women in the intervention group reported reductions in both physical and sexual IPV following the intervention [23].

In contrast to the significant reductions in IPV experience among women, no significant effects for overall IPV perpetration were detected among men in South Africa. However, analyses of IPV subtypes found significant reductions in psychological violence, physical violence, and sexual violence among men in South Africa (Multimedia Appendix 5). The lack of change in overall IPV perpetration might be related to the low level of male engagement. Notably, the user engagement and the IPV completion rate of men in South Africa were 50% less than those of women. Hence, the limited interaction men had with the chatbot may not have been sufficient to change these behaviors.

It may also be the case that more interactive engagement is needed to shift these complex behaviors. For instance, emerging findings suggest that content modality can impact digital intervention outcomes [42], as highlighted in a recent study on ChattyCuz, a chatbot intervention that aims to support young women in navigating intimate relationships in South Africa [70]. Interestingly, an RCT of ChattyCuz found that while the gamified version of the chatbot (treatment 1) led to modest but significant reductions in IPV experience compared to those without treatment, in the narrative treatment, there was no effect on IPV experience [70]. The authors hypothesize that the lack of measurable effect on IPV experience in the narrative treatment arm of the study might be related to the lack of user interaction, that greater engagement with intervention content is necessary for behavior change, and that an information-only approach is insufficient [70]. These findings highlight how low levels of user engagement in ParentText might play a role in the lack of behavior changes detected and underscore that, amendments to the program modality might be needed to make the content more interactive to shift behaviors. Recent studies have also demonstrated the strong impact that interpersonal relationships have on gender attitudes, underscoring the need for interventions to not only target individuals but also social networks, such as peers and family members through, for example, critical reflection and dialog among social groups [71]. Research suggests that programs that have shown evidence of change in gender norms involve the engagement of stakeholders at different levels of the social-ecological model, include activities that allow for active participation, and promote critical awareness [14]. Consequently, it is possible that greater interactive engagement is needed to see greater changes in IPV experience and perpetration.

This study also revealed a significant overall reduction in harmful attitudes toward IPV as well as a significant increase in overall gender-equitable behaviors among men in South Africa. These indicative effects are promising, and while caution is warranted given the small sample, the findings align with other interventions that have observed reductions in harmful attitudes among men at posttest and detected increases in gender-equitable behaviors following gender-transformative interventions [72,73].

Strengths and Limitations

The study has various strengths worth noting. As the first study of a digital parenting intervention with integrated IPV prevention content, the present research makes a valuable contribution to the field of violence prevention and targeted efforts seeking to address VAC and IPV concurrently. With a growing emphasis in policy and research on preventing multiple forms of violence simultaneously [74], this study sheds light on preliminary behavioral outcomes and how users engage in IPV prevention content that has been integrated into a digital parenting intervention. Another strength of the study is the use of data from 2 countries. By delivering the program in 2 different

XSL•FO
RenderX

regions and cultures, this study offers an insight into important considerations to bear in mind before scaling up the intervention. For instance, the challenges with recruiting men in Jamacia compared to South Africa suggest that alternative recruitment strategies might be needed when scaling up the intervention in the Caribbean region.

The study also has various limitations. The first relates to the lack of a control group, without which causality cannot be established from the pre-post data. The study also had a small sample size, which means that the research findings need to be interpreted with caution given that there is limited statistical power in detecting pre-post intervention effects. Consequently, the next step is a fully powered RCT to establish effectiveness. Future research would also benefit from incorporating a factorial experiment design, which can be used to examine which specific intervention components have an effect on outcomes and which components interact with each other [75]. Knowing which components should be kept to sustain impact when taking an intervention to scale or delivering it in a new setting is critical, especially when considering program length and implementation costs [76]. Including follow-up assessments to examine whether intervention effects remain, become reduced, or are delayed over time [77] would also be beneficial in future studies as this study was limited to a 6-week posttest due to resource and time constraints. In addition, the study used a convenience sampling approach, which may limit the generalizability of the study findings. For example, it is possible that caregivers who opted to take part in the parenting program were more willing to change, which may bias the results. The outcomes in the study were also self-reported, and consequently, they may be subject to disclosure bias, recall bias, and social-desirability bias, leading to under reporting of unfavorable behaviors and over reporting of what might be perceived as desirable answers. However, given that the self-reports were carried out on participants' phones and participants were told the reports were fully anonymized, this will likely have reduced desirability bias in the quantitative surveys.

Another important limitation is that while in this study, participants were provided with mobile credit, this may not always be realistic in a real-world setting. While access to the internet and mobile data are increasing globally [78], it is important to explore potential options of being able to use the program without mobile data, especially when planning strategies for scale-up and delivery. One digital intervention that has explored this is the parenting program ParentApp for Teens, which is a mobile app designed for offline use that targets parents of teenagers [79,80]. There are also assessment limitations related to the self-reported measurements used in this study. Given the sensitive nature surrounding IPV and attitudes toward violence, it is possible that participants were hesitant to disclose violence due to stigma or fear of being endangered, leading to under reporting and consequently leading to undetected changes in behaviors [81]. Similarly, social-desirability bias may have also impacted the measurement of behavioral outcomes, particularly the measurement of IPV perpetration, which has been found to be more prone to social-desirability bias than reports of IPV experience [82].

An additional limitation worth noting is that men were underrepresented in the study. Recruitment of male caregivers is a known difficulty with parenting interventions [83], and even though it was possible to recruit both male and female caregivers in South Africa, it was only possible to recruit female caregivers in Jamaica. The lack of male caregivers from Jamaica means that certain experiences may have been underestimated due to the low number of fathers included. This challenge highlights the need for future studies to ensure greater efforts are taken to adopt recruitment strategies that ensure men are also included. In future studies, it would also be valuable to include people aged between 16 and 17 years in the South African sample to better understand how adolescent parents in South Africa engage with the chatbot.

Conclusions

This research presents tentative findings on the first-ever attempt to research the integration of IPV and parenting content into a digital intervention. Some of the preliminary results of the intervention are promising, with indications of reductions in overall IPV for women and a potential decrease in harmful IPV attitudes among men. However, the high rates of IPV experiences and perpetration reported at the posttest are still concerning and suggest that greater intervention efforts and programmatic amendments are likely necessary to tackle deep-rooted attitudes and harmful behaviors. In addition, the unexpectedly low levels of engagement in the intervention raise questions about whether interactions with the chatbot were sufficient to shift behaviors and attitudes and suggest that a hybrid approach to program delivery might be necessary. Given the urgent need for scalable prevention efforts that can tackle multiple forms of violence concurrently [84], more research is necessary to further explore how to increase chatbot user engagement and achieve greater reductions in IPV and harmful attitudes toward gender roles and IPV.

Acknowledgments

This research was funded in whole, or in part, by the United Kingdom Research and Innovation (UKRI; grant ES/P000649/1). For the purpose of open access, the author has applied for a CC BY public copyright license to any author accepted manuscript version arising from this submission. This research was also funded by UNICEF (United Nations Children's Fund) Jamaica and the Economic and Social Research Council Grand Union Doctoral Training Partnership in collaboration with the University of Oxford Saven European Scholarship and the Balliol College Dervorguilla Scholarship in Social Sciences. This study is part of the Global Parenting Initiative, which is funded by the LEGO Foundation, Oak Foundation, the World Childhood Foundation (grant 16191), the Human Safety Net, and the UK Research and Innovation Global Challenges Research Fund (grant ES/S008101/1). Aside from UNICEF's involvement in guiding the selection of the research site and study time frame, none of the funders were

XSL•FO
RenderX

involved in study design, data collection, data analysis and interpretation, or the writing of the manuscript. The authors would like to thank those who made this research possible: the respondents who participated in this study, UNICEF Jamaica, the research assistants, the study funders, the researchers and stakeholders who provided input during intervention development and delivery, and the Department of Social Policy and Intervention at the University of Oxford for administrative support. Generative artificial intelligence was not used in any portion of the manuscript writing.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

MS, JL, and FG participated in the design of the study and oversaw data collection and cleaning. MS managed the data collection process, with support from PZ, FACA, GHR, and RS. MS designed the statistical analysis, with JL, FG, QH, FACA, CF, and LC providing guidance on the analysis approach. MS drafted the paper, with JL and FG contributing. All authors approved the final manuscript.

Conflicts of Interest

MS is a researcher at the University of Oxford, and JL and FG are employed as staff at the university. In addition, JL is the chief executive officer of Parenting for Lifelong Health, a charitable organization based in the United Kingdom that developed the ParentText intervention. MS, JL, PZ, QH, FACA, GHR, and FG have participated and are participating in several research studies involving the program as investigators at the universities of Oxford, Cape Town, and Glasgow. CF and LC are employees of IDEMS International, a technology development company that codeveloped the ParentText intervention and provided support on data analysis of user engagement. RS is an employee at UNICEF and provided support on data collection. Conflict is avoided by declaring these potential conflicts of interest and by conducting and disseminating rigorous, transparent, and impartial evaluation research on both this and other similar parenting programs. No profits will be made from this program.

Multimedia Appendix 1
Intimate partner violence assessment.
[DOCX File , 28 KB - formative_v9i1e58611_app1.docx]

Multimedia Appendix 2
Structure of the ParentText intervention.
[DOCX File , 24 KB - formative_v9i1e58611_app2.docx]

Multimedia Appendix 3
Overview of ParentText intimate partner violence prevention content.
[DOCX File , 20 KB - formative_v9i1e58611_app3.docx]

Multimedia Appendix 4
CONSORT (Consolidated Standards of Reporting Trials) checklist.
[DOCX File , 374 KB - formative_v9i1e58611_app4.docx]

Multimedia Appendix 5
Detailed primary and secondary outcome tables.
[DOCX File , 46 KB - formative_v9i1e58611_app5.docx]

References

1. Violence against women prevalence estimates, 2018. World Health Organization. 2021 Mar 9. URL: https://www.who.int/publications/i/item/9789240022256 [accessed 2024-12-10]
2. Whitten T, Tzoumakis S, Green MJ, Dean K. Global prevalence of childhood exposure to physical violence within domestic and family relationships in the general population: a systematic review and proportional meta-analysis. Trauma Violence Abuse 2024 Apr 10;25(2):1411-1430 [FREE Full text] [doi: 10.1177/15248380231179133] [Medline: 37300288]
3. Hillis S, Mercy J, Amobi A, Kress H. Global prevalence of past-year violence against children: a systematic review and minimum estimates. Pediatrics 2016 Mar;137(3):e20154079 [FREE Full text] [doi: 10.1542/peds.2015-4079] [Medline: 26810785]
4. Global status report on preventing violence against children 2020. World Health Organization. URL: https://www.unicef.org/media/70731/file/Global-status-report-on-preventing-violence-against-children-2020.pdf [accessed 2024-12-10]

XSL•FO
RenderX

5. Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence. World Health Organization. 2013 Oct 20. URL: https://www.who.int/publications/i/item/9789241564625 [accessed 2024-12-10]
6. Cerna-Turoff I, Fang Z, Meierkord A, Wu Z, Yanguela J, Bangirana CA, et al. Factors associated with violence against children in low- and middle-income countries: a systematic review and meta-regression of nationally representative data. Trauma Violence Abuse 2021 Apr;22(2):219-232 [FREE Full text] [doi: 10.1177/1524838020985532] [Medline: 33461441]
7. Pearson I, Page S, Zimmerman C, Meinck F, Gennari F, Guedes A, et al. The co-occurrence of intimate partner violence and violence against children: a systematic review on associated factors in low- and middle-income countries. Trauma Violence Abuse 2023 Oct;24(4):2097-2114 [FREE Full text] [doi: 10.1177/15248380221082943] [Medline: 35481390]
8. Guedes A, Bott S, Garcia-Moreno C, Colombini M. Bridging the gaps: a global review of intersections of violence against women and violence against children. Glob Health Action 2016 Jun 20;9:31516 [FREE Full text] [doi: 10.3402/gha.v9.31516] [Medline: 27329936]
9. Alhusen JL, Ho GW, Smith KF, Campbell JC. Addressing intimate partner violence and child maltreatment: challenges and opportunities. In: Korbin J, Krugman R, editors. Handbook of Child Maltreatment. Dordrecht, The Netherlands: Springer; 2014.
10. Gracia E, Rodriguez CM, Martín-Fernández M, Lila M. Acceptability of family violence: underlying ties between intimate partner violence and child abuse. J Interpers Violence 2020 Sep;35(17-18):3217-3236. [doi: 10.1177/0886260517707310] [Medline: 29294751]
11. Chisholm CA, Bullock L, Ferguson JE. Intimate partner violence and pregnancy: epidemiology and impact. Am J Obstet Gynecol 2017 Aug;217(2):141-144. [doi: 10.1016/j.ajog.2017.05.042] [Medline: 28551446]
12. Guedes A, Mikton C. Examining the intersections between child maltreatment and intimate partner violence. West J Emerg Med 2013 Aug;14(4):377-379 [FREE Full text] [doi: 10.5811/westjem.2013.2.16249] [Medline: 23997846]
13. Edwards KM, Kumar M, Waterman EA, Mullet N, Madeghe B, Musindo O. Programs to prevent violence against children in sub-Saharan Africa: a systematic review. Trauma Violence Abuse 2024 Jan;25(1):593-612 [FREE Full text] [doi: 10.1177/15248380231160742] [Medline: 36964686]
14. Levy JK, Darmstadt GL, Ashby C, Quandt M, Halsey E, Nagar A, et al. Characteristics of successful programmes targeting gender inequality and restrictive gender norms for the health and wellbeing of children, adolescents, and young adults: a systematic review. Lancet Glob Health 2020 Feb;8(2):e225-e236 [FREE Full text] [doi: 10.1016/S2214-109X(19)30495-4] [Medline: 31879212]
15. Brush LD, Miller E. Trouble in paradigm: "gender transformative" programming in violence prevention. Violence Against Women 2019 Nov;25(14):1635-1656. [doi: 10.1177/1077801219872551] [Medline: 31640536]
16. Marcos-Marcos J, Nardini K, Briones-Vozmediano E, Vives-Cases C. Listening to stakeholders in the prevention of gender-based violence among young people in Spain: a qualitative study from the positivMasc project. BMC Womens Health 2023 Jul 26;23(1):393 [FREE Full text] [doi: 10.1186/s12905-023-02545-3] [Medline: 37496067]
17. Casey E, Carlson J, Two Bulls S, Yager A. Gender transformative approaches to engaging men in gender-based violence prevention: a review and conceptual model. Trauma Violence Abuse 2018 Apr;19(2):231-246. [doi: 10.1177/1524838016650191] [Medline: 27197533]
18. Doyle K, Levtov RG, Karamage E, Rakshit D, Kazimbaya S, Sayinzoga F, et al. Long-term impacts of the Bandebereho programme on violence against women and children, maternal health-seeking, and couple relations in Rwanda: a six-year follow-up of a randomised controlled trial. EClinicalMedicine 2023 Sep 26;64:102233 [FREE Full text] [doi: 10.1016/j.eclinm.2023.102233] [Medline: 37781160]
19. van den Berg W, Hendricks L, Hatcher A, Peacock D, Godana P, Dworkin S. 'One man can': shifts in fatherhood beliefs and parenting practices following a gender-transformative programme in Eastern Cape, South Africa. Gend Dev 2013 Feb 01;21(1):111-125 [FREE Full text] [doi: 10.1080/13552074.2013.769775] [Medline: 24000276]
20. Ruane-McAteer E, Gillespie K, Amin A, Aventin Á, Robinson M, Hanratty J, et al. Gender-transformative programming with men and boys to improve sexual and reproductive health and rights: a systematic review of intervention studies. BMJ Glob Health 2020 Oct;5(10):e002997 [FREE Full text] [doi: 10.1136/bmjgh-2020-002997] [Medline: 33051283]
21. WHO guidelines on parenting interventions to prevent maltreatment and enhance parent–child relationships with children aged 0–17 years. World Health Organization. 2023 Feb 10. URL: https://www.who.int/publications/i/item/9789240065505 [accessed 2023-03-09]
22. Backhaus S, Gardner F, Schafer M, Melendez-Torres GJ, Knerr W, Lachman JM. World Health Organization guidelines on parenting interventions to prevent maltreatment and enhance parent–child relationships with children aged 0-17 Years: report of the systematic reviews of evidence. Center for Evidence-Based Intervention, University of Oxford. 2023. URL: https://pure.uva.nl/ws/files/192859093/systematic_reviews-for-the-who-parenting-guideline-jan-27th-2023.pdf [accessed 2024-12-10]
23. Doyle K, Levtov RG, Barker G, Bastian GG, Bingenheimer JB, Kazimbaya S, et al. Gender-transformative Bandebereho couples' intervention to promote male engagement in reproductive and maternal health and violence prevention in Rwanda: findings from a randomized controlled trial. PLoS One 2018 Apr 4;13(4):e0192756 [FREE Full text] [doi: 10.1371/journal.pone.0192756] [Medline: 29617375]

XSL•FO
RenderX

24. Sanders MR, Divan G, Singhal M, Turner KM, Velleman R, Michelson D, et al. Scaling up parenting interventions is critical for attaining the sustainable development goals. Child Psychiatry Hum Dev 2022 Oct;53(5):941-952 [FREE Full text] [doi: 10.1007/s10578-021-01171-0] [Medline: 33948778]
25. Harris M, Andrews K, Gonzalez A, Prime H, Atkinson L. Technology-assisted parenting interventions for families experiencing social disadvantage: a meta-analysis. Prev Sci 2020 Jul;21(5):714-727. [doi: 10.1007/s11121-020-01128-0] [Medline: 32415543]
26. Cluver LD, Meinck F, Steinert JI, Shenderovich Y, Doubt J, Herrero Romero R, et al. Parenting for lifelong health: a pragmatic cluster randomised controlled trial of a non-commercialised parenting programme for adolescents and their families in South Africa. BMJ Glob Health 2018 Jan 31;3(1):e000539 [FREE Full text] [doi: 10.1136/bmjgh-2017-000539] [Medline: 29564157]
27. Ward CL, Wessels IM, Lachman JM, Hutchings J, Cluver LD, Kassanjee R, et al. Parenting for lifelong health for young children: a randomized controlled trial of a parenting program in South Africa to prevent harsh parenting and child conduct problems. J Child Psychol Psychiatry 2020 Apr;61(4):503-512 [FREE Full text] [doi: 10.1111/jcpp.13129] [Medline: 31535371]
28. Schafer M, Lachman JM, Gardner F, Zinser P, Calderon F, Han Q, et al. Integrating intimate partner violence prevention content into a digital parenting chatbot intervention during COVID-19: intervention development and remote data collection. BMC Public Health 2023 Sep 04;23(1):1708 [FREE Full text] [doi: 10.1186/s12889-023-16649-w] [Medline: 37667352]
29. Shai NJ, Sikweyiya Y. Programmes for change: addressing sexual and intimate partner violence in South Africa. S Afr Crime Q 2015 Mar;51:31-41. [doi: 10.4314/sacq.v51i1.4]
30. Sere Y, Roman NV, Ruiter RA. Coping with the experiences of intimate partner violence among South African women: systematic review and meta-synthesis. Front Psychiatry 2021 May 26;12:655130 [FREE Full text] [doi: 10.3389/fpsyt.2021.655130] [Medline: 34122178]
31. Gordon C. Intimate partner violence is everyone's problem, but how should we approach it in a clinical setting? S Afr Med J 2016 Sep 10;106(10):962-965. [doi: 10.7196/samj.2016.v106i10.11408]
32. Violence against women. Organisation for Economic Co-operation and Development. URL: https://www.oecd.org/en/data/indicators/violence-against-women.html [accessed 2024-12-10]
33. Jamaica gender landscape. World Bank Group. URL: https://documents1.worldbank.org/curated/en/099106107012217020/pdf/IDU1101a2e811893114b471a77e16578ca7db7a4.pdf [accessed 2024-12-10]
34. Phaswana ED. Women, gender and race in post-apartheid South Africa. In: Yacob-Haliso O, Falola T, editors. The Palgrave Handbook of African Women's Studies. Cham, Switzerland: Palgrave Macmillan; Feb 09, 2021.
35. Thabane L, Cambon L, Potvin L, Pommier J, Kivits J, Minary L, et al. Population health intervention research: what is the place for pilot studies? Trials 2019 May 30;20(1):309 [FREE Full text] [doi: 10.1186/s13063-019-3422-4] [Medline: 31146768]
36. Kaehler LA, Jacobs M, Jones DJ. Distilling common history and practice elements to inform dissemination: Hanf-model BPT programs as an example. Clin Child Fam Psychol Rev 2016 Sep;19(3):236-258 [FREE Full text] [doi: 10.1007/s10567-016-0210-5] [Medline: 27389606]
37. Ashburn K, Kerner B, Ojamuge D, Lundgren R. Evaluation of the responsible, engaged, and loving (REAL) fathers initiative on physical child punishment and intimate partner violence in northern Uganda. Prev Sci 2017 Oct;18(7):854-864 [FREE Full text] [doi: 10.1007/s11121-016-0713-9] [Medline: 27738782]
38. Kapiga S, Harvey S, Mshana G, Hansen CH, Mtolela GJ, Madaha F, et al. A social empowerment intervention to prevent intimate partner violence against women in a microfinance scheme in Tanzania: findings from the MAISHA cluster randomised controlled trial. Lancet Glob Health 2019 Oct;7(10):e1423-e1434 [FREE Full text] [doi: 10.1016/S2214-109X(19)30316-X] [Medline: 31537372]
39. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016 Oct 24;355:i5239 [FREE Full text] [doi: 10.1136/bmj.i5239] [Medline: 27777223]
40. Guidelines for producing statistics on violence against women. United Nations. 2014. URL: https://unstats.un.org/unsd/gender/docs/guidelines_statistics_vaw.pdf [accessed 2024-12-10]
41. Gugerty MK, Karlan D. The Goldilocks Challenge: Right-Fit Evidence for the Social Sector. Oxford, UK: Oxford University Press; May 24, 2018.
42. Breitenstein SM, Gross D, Christophersen R. Digital delivery methods of parenting training interventions: a systematic review. Worldviews Evid Based Nurs 2014 Jun;11(3):168-176. [doi: 10.1111/wvn.12040] [Medline: 24842341]
43. Collins DA, Tully LA, Piotrowska PJ, Hawes DJ, Dadds MR. Perspectives on ParentWorks: learnings from the development and national roll-out of a self-directed online parenting intervention. Internet Interv 2019 Mar;15:52-59 [FREE Full text] [doi: 10.1016/j.invent.2018.12.002] [Medline: 30656140]
44. UNICEF's open source approach to innovation. UNICEF Innovation. 2020. URL: https://www.unicef.org./innovation/stories/unicefs-open-source-approach-innovation [accessed 2024-12-08]
45. García-Moreno C, Jansen HA, Watts C, Ellsberg M, Heise L. WHO multi-country study on women's health and domestic violence against women: summary report of initial results on prevalence, health outcomes and women's responses. World

XSL•FO
RenderX

Health Organization. 2005. URL: https://iris.who.int/bitstream/handle/10665/43310/9241593512_eng.pdf?sequence=1 [accessed 2024-12-10]
46. Meinck F, Boyes ME, Cluver L, Ward CL, Schmidt P, DeStone S, et al. Adaptation and psychometric properties of the ISPCAN Child Abuse Screening Tool for use in trials (ICAST-Trial) among South African adolescents and their primary caregivers. Child Abuse Negl 2018 Aug;82:45-58 [FREE Full text] [doi: 10.1016/j.chiabu.2018.05.022] [Medline: 29860107]
47. Ebert C, Steinert JI. Prevalence and risk factors of violence against women and children during COVID-19, Germany. Bull World Health Organ 2021 Jun 01;99(6):429-438 [FREE Full text] [doi: 10.2471/BLT.20.270983] [Medline: 34108753]
48. McNaughton Reyes HL, Maman S, Chen MS, Groves AK, Moodley D. Patterns of intimate partner violence victimization among South African women and their relation to emotional distress during pregnancy and postpartum. J Interpers Violence 2021 Apr 10;36(7-8):NP4230-NP4249 [FREE Full text] [doi: 10.1177/0886260518786738] [Medline: 29991312]
49. Abramsky T, Kapinga I, Mshana G, Lees S, Hansen CH, Hashim R, et al. Couples data from north-western Tanzania: insights from a survey of male partners of women enrolled in the MAISHA cluster randomized trial of an intimate partner violence prevention intervention. PLoS One 2020 Oct 2;15(10):e0240112 [FREE Full text] [doi: 10.1371/journal.pone.0240112] [Medline: 33006997]
50. Lachenbruch PA. McNemar test. In: Wiley StatsRef: Statistics Reference Online. Hoboken, NJ: John Wiley & Sons; 2014.
51. Trajman A, Luiz RR. McNemar chi2 test revisited: comparing sensitivity and specificity of diagnostic examinations. Scand J Clin Lab Invest 2008;68(1):77-80. [doi: 10.1080/00365510701666031] [Medline: 18224558]
52. Ethical and safety recommendations for intervention research on violence against women. World Health Organization. 2016 Nov 29. URL: https://www.who.int/publications/i/item/9789241510189 [accessed 2024-12-10]
53. COVID-19 Risk Communication and Community Engagement (RCCE) guidance note on quantitative data collection. United Nations International Children's Fund. 2020. URL: https://www.corecommitments.unicef.org/kp/rcce-covid19---quantitative-data-collection-guidance-note.pdf [accessed 2024-12-01]
54. Leung L, Miedema S, Warner X, Homan S, Fulu E. Making feminism count: integrating feminist research principles in large-scale quantitative research on violence against women and girls. Gend Dev 2019 Oct 29;27(3):427-447. [doi: 10.1080/13552074.2019.1668142]
55. Jamieson MK, Govaart GH, Pownall M. Reflexivity in quantitative research: a rationale and beginner's guide. Soc Personal Psychol Compass 2023 Feb;17(4):e12735. [doi: 10.1111/spc3.12735]
56. Richardson BP, van der Linde J, Pillay B, Swanepoel DW. Do text messages about health and development in young children affect caregiver behaviour and child outcomes? A systematic review. Health Educ J 2020 Oct 21;80(2):225-237. [doi: 10.1177/0017896920965992]
57. Dadds MR, Sicouri G, Piotrowska PJ, Collins DA, Hawes DJ, Moul C, et al. Keeping parents involved: predicting attrition in a self-directed, online program for childhood conduct problems. J Clin Child Adolesc Psychol 2019 Aug 01;48(6):881-893. [doi: 10.1080/15374416.2018.1485109] [Medline: 30067388]
58. Hall CM, Bierman KL. Technology-assisted interventions for parents of young children: emerging practices, current research, and future directions. Early Child Res Q 2015;33:21-32 [FREE Full text] [doi: 10.1016/j.ecresq.2015.05.003] [Medline: 27773964]
59. Guillot-Wright SP, Lu Y, Torres ED, Le VD, Hall HR, Temple JR. Design and feasibility of a school-based text message campaign to promote healthy relationships. School Ment Health 2018 Feb 9;10(4):428-436. [doi: 10.1007/s12310-018-9255-6]
60. Fletcher R, Kay-Lambkin F, May C, Oldmeadow C, Attia J, Leigh L. Supporting men through their transition to fatherhood with messages delivered to their smartphones: a feasibility study of SMS4dads. BMC Public Health 2017 Dec 13;17(1):953 [FREE Full text] [doi: 10.1186/s12889-017-4978-0] [Medline: 29237439]
61. Lechowicz ME, Jiang Y, Tully LA, Burn MT, Collins DA, Hawes DJ, et al. Enhancing father engagement in parenting programs: translating research into practice recommendations. Aus Psychol 2020 Nov 12;54(2):83-89. [doi: 10.1111/ap.12361]
62. Guthridge M, Kirkman M, Penovic T, Giummarra MJ. Promoting gender equality: a systematic review of interventions. Soc Just Res 2022 Sep 01;35(3):318-343. [doi: 10.1007/s11211-022-00398-z]
63. Yaremych HE, Persky S. Recruiting fathers for parenting research: an evaluation of eight recruitment methods and an exploration of fathers' motivations for participation. Parent Sci Pract 2023 Feb 25;23(1):1-32 [FREE Full text] [doi: 10.1080/15295192.2022.2036940] [Medline: 37346458]
64. Jeong J, Sullivan EF, McCann JK, McCoy DC, Yousafzai AK. Implementation characteristics of father-inclusive interventions in low- and middle-income countries: a systematic review. Ann N Y Acad Sci 2023 Feb;1520(1):34-52 [FREE Full text] [doi: 10.1111/nyas.14941] [Medline: 36482863]
65. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
66. Heckman BW, Mathew AR, Carpenter MJ. Treatment burden and treatment fatigue as barriers to health. Curr Opin Psychol 2015 Oct 01;5:31-36 [FREE Full text] [doi: 10.1016/j.copsyc.2015.03.004] [Medline: 26086031]
67. Hanno EC. Nudging early educators' knowledge, beliefs, and practices: an embedded randomized controlled trial of text message supports. Early Educ Dev 2022 Mar 20;34(2):530-550. [doi: 10.1080/10409289.2022.2053476]

XSL•FO
RenderX

68. Cooper H, Nadzri FZ, Vyas S, Juhari R, Ismail N, Arshat Z, et al. A hybrid digital parenting program delivered within the Malaysian preschool system: protocol for a feasibility study of a small-scale factorial cluster randomized trial. JMIR Res Protoc 2024 Apr 26;13:e55491 [FREE Full text] [doi: 10.2196/55491] [Medline: 38669679]
69. UNICEF Innocenti - Global Office of Research and Foresight, Prevention Collaborative, Equimundo. Parenting programmes to reduce violence against children and women. Why it is important. Brief 1.: UNICEF Innocenti; 2023. URL: https://www.unicef.org/innocenti/media/3456/file/UNICEF-Innocenti-Parenting-Programmes-Reduce-Violence-Against-Children-Women-Brief-1-2024-EN.pdf
70. De Filippo A, Bellatin P, Tietz N, Grant E, Whitefield A, Nkopane P, et al. Effects of digital chatbot on gender attitudes and exposure to intimate partner violence among young women in South Africa. PLOS Digit Health 2023 Oct 16;2(10):e0000358 [FREE Full text] [doi: 10.1371/journal.pdig.0000358] [Medline: 37844088]
71. Kågesten A, Gibbs S, Blum RW, Moreau C, Chandra-Mouli V, Herbert A, et al. Understanding factors that shape gender attitudes in early adolescence globally: a mixed-methods systematic review. PLoS One 2016 Jun 24;11(6):e0157805 [FREE Full text] [doi: 10.1371/journal.pone.0157805] [Medline: 27341206]
72. Falb KL, Asghar K, Blackwell A, Baseme S, Nyanguba M, Roth D, et al. Improving family functioning and reducing violence in the home in North Kivu, Democratic Republic of Congo: a pilot cluster-randomised controlled trial of Safe at Home. BMJ Open 2023 Mar 06;13(3):e065759 [FREE Full text] [doi: 10.1136/bmjopen-2022-065759] [Medline: 36878658]
73. Halim N, Steven Mzilangwe E, Reich N, Badi L, Simmons E, Servidone M, et al. Together to end violence against women in Tanzania: results of a pilot cluster randomized controlled trial to evaluate preliminary effectiveness of interpersonal and community level interventions to address intimate partner violence. Glob Public Health 2019 Dec 13;14(12):1653-1668 [FREE Full text] [doi: 10.1080/17441692.2019.1609062] [Medline: 31084264]
74. Gender action plan, 2022-2025: a vision for lasting, transformative change. United Nations Children's Fund. URL: https://www.unicef.org/gender-equality/gender-action-plan-2022-2025
75. Collins LM. Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: The Multiphase Optimization Strategy (MOST). Cham, Switzerland: Springer; 2018.
76. Starmann E, Collumbien M, Kyegombe N, Devries K, Michau L, Musuya T, et al. Exploring couples' processes of change in the context of SASA!, a violence against women and HIV prevention intervention in Uganda. Prev Sci 2017 Feb 29;18(2):233-244 [FREE Full text] [doi: 10.1007/s11121-016-0716-6] [Medline: 27682273]
77. van Aar J, Leijten P, Orobio de Castro B, Overbeek G. Sustained, fade-out or sleeper effects? A systematic review and meta-analysis of parenting interventions for disruptive child behavior. Clin Psychol Rev 2017 Feb;51:153-163. [doi: 10.1016/j.cpr.2016.11.006] [Medline: 27930935]
78. Broadband statistics. Organisation for Economic Co-operation and Development. URL: https://www.oecd.org/en/topics/sub-issues/broadband-statistics.html [accessed 2023-10-31]
79. Awah I, Green O, Baerecke L, Janowski R, Klapwijk J, Chetty AN, et al. 'It provides practical tips, practical solutions!': acceptability, usability, and satisfaction of a digital parenting intervention across African countries. Psychol Health Med 2022 Aug 18;27(sup1):107-123. [doi: 10.1080/13548506.2022.2113106] [Medline: 35980251]
80. Janowski R, Green O, Shenderovich Y, Stern D, Clements L, Wamoyi J, et al. Optimising engagement in a digital parenting intervention to prevent violence against adolescents in Tanzania: protocol for a cluster randomised factorial trial. BMC Public Health 2023 Jun 23;23(1):1224 [FREE Full text] [doi: 10.1186/s12889-023-15989-x] [Medline: 37353844]
81. Cullen C. Method matters: the underreporting of intimate partner violence. World Bank Econ Rev 2023 Feb;37(1):49-73. [doi: 10.1093/wber/lhac022]
82. Visschers J, Jaspaert E, Vervaeke G. Social desirability in intimate partner violence and relationship satisfaction reports: an exploratory analysis. J Interpers Violence 2017 May 09;32(9):1401-1420. [doi: 10.1177/0886260515588922] [Medline: 26058978]
83. McGirr S, Torres J, Heany J, Brandon H, Tarry C, Robinson C. Lessons learned on recruiting and retaining young fathers in a parenting and repeat pregnancy prevention program. Matern Child Health J 2020 Sep 20;24(Suppl 2):183-190 [FREE Full text] [doi: 10.1007/s10995-020-02956-w] [Medline: 32564249]
84. Bott S, Guedes A, Subrahmanian R, Chowdhury S, Rumble L, Daban RA. Gender dimensions of violence against children and adolescents. United Nations Children's Fund (UNICEF). 2020 Apr. URL: https://www.unicef.org/media/93986/file/Child-Protection-Gender-Dimensions-of-VACAG-2021.pdf [accessed 2024-12-10]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
IPV: intimate partner violence
ODK: Open Data Kit
RCT: randomized controlled trial
UNICEF: United Nations Children’s Fund
VAC: violence against children
VAW: violence against women

XSL•FO
RenderX

Edited by A Mavragani; submitted 20.03.24; peer-reviewed by M Wang, S Willan, RG Levtov; comments to author 25.06.24; revised version received 13.10.24; accepted 30.10.24; published 03.01.25.

Please cite as:
Schafer M, Lachman J, Zinser P, Calderón Alfaro FA, Han Q, Facciola C, Clements L, Gardner F, Haupt Ronnie G, Sheil R
A Digital Parenting Intervention With Intimate Partner Violence Prevention Content: Quantitative Pre-Post Pilot Study
JMIR Form Res 2025;9:e58611
URL: https://formative.jmir.org/2025/1/e58611
doi:10.2196/58611
PMID:

©Moa Schafer, Jamie Lachman, Paula Zinser, Francisco Antonio Calderón Alfaro, Qing Han, Chiara Facciola, Lily Clements, Frances Gardner, Genevieve Haupt Ronnie, Ross Sheil. Originally published in JMIR Formative Research (https://formative.jmir.org), 03.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Short Paper

Validation of Sleep Measurements of an Actigraphy Watch: Instrument Validation Study

Mari Waki[1*]; Ryohei Nakada[1*]; Kayo Waki[1]; Yuki Ban[1]; Ryo Suzuki[2]; Toshimasa Yamauchi[3]; Masaomi Nangaku[4]; Kazuhiko Ohe[1]

[1]Department of Biomedical Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[2]Department of Diabetes, Metabolism and Endocrinology, Tokyo Medical University, Tokyo, Japan

[3]Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[4]Division of Nephrology and Endocrinology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[*]these authors contributed equally

Corresponding Author:
Kayo Waki
Department of Biomedical Informatics
Graduate School of Medicine
The University of Tokyo
7-chōme-3-1 Hongō
Bunkyo City
Tokyo, 113-8654
Japan
Phone: 81 03 5800 9129
Fax: 81 03 5800 9173
Email: kwaki-tky@m.u-tokyo.ac.jp

Abstract

Background: The iAide2 (Tokai) physical activity monitoring system includes diverse measurements and wireless features useful to researchers. The iAide2's sleep measurement capabilities have not been compared to validated sleep measurement standards in any published work.

Objective: We aimed to assess the iAide2's sleep duration and total sleep time (TST) measurement performance and perform calibration if needed.

Methods: We performed free-living sleep monitoring in 6 convenience-sampled participants without known sleep disorders recruited from within the Waki DTx Laboratory at the Graduate School of Medicine, University of Tokyo. To assess free-living sleep, we validated the iAide2 against a second actigraph that was previously validated against polysomnography, the MotionWatch 8 (MW8; CamNtech Ltd). The participants wore both devices on the nondominant arm, with the MW8 closest to the hand, all day except when bathing. The MW8 and iAide2 assessments both used the MW8 EVENT-marker button to record bedtime and risetime. For the MW8, MotionWare Software (version 1.4.20; CamNtech Ltd) provided TST, and we calculated sleep duration from the sleep onset and sleep offset provided by the software. We used a similar process with the iAide2, using iAide2 software (version 7.0). We analyzed 64 nights and evaluated the agreement between the iAide2 and the MW8 for sleep duration and TST based on intraclass correlation coefficients (ICCs).

Results: The absolute ICCs (2-way mixed effects, absolute agreement, single measurement) for sleep duration (0.69, 95% CI –0.07 to 0.91) and TST (0.56, 95% CI –0.07 to 0.82) were moderate. The consistency ICC (2-way mixed effects, consistency, single measurement) was excellent for sleep duration (0.91, 95% CI 0.86-0.95) and moderate for TST (0.78, 95% CI 0.67-0.86). We determined a simple calibration approach. After calibration, the ICCs improved to 0.96 (95% CI 0.94-0.98) for sleep duration and 0.82 (95% CI 0.71-0.88) for TST. The results were not sensitive to the specific participants included, with an ICC range of 0.96-0.97 for sleep duration and 0.79-0.87 for TST when applying our calibration equation to data removing one participant at a time and 0.96-0.97 for sleep duration and 0.79-0.86 for TST when recalibrating while removing one participant at a time.

Conclusions: The measurement errors of the uncalibrated iAide2 for both sleep duration and TST seem too large for them to be useful as absolute measurements, though they could be useful as relative measurements. The measurement errors after calibration

XSL•FO
RenderX

are low, and the calibration approach is general and robust, validating the use of iAide2's sleep measurement functions alongside its other features in physical activity research.

(JMIR Form Res 2025;9:e63529) doi:10.2196/63529

KEYWORDS

actigraphy; sleep; Motion Watch 8; iAide2; total sleep time

Introduction

Background

The iAide2 (Tokai) is a physical activity monitoring system that measures intensity levels, step counts, temperature, and pulse rate [1,2] to assess activity and sleep, with minute-by-minute assessment of wake/sleep [3,4]. It is based on TDK's Silmee W20/W22 actigraphs [4-6] used in previous research [5,7-9]. The iAide2 supports long trials (via rechargeability and off-watch data storage) and timely behavior feedback (via a wireless EVENT button and near–real-time access to data). Its sleep measurements are not well validated; Kimura et al [7] compared its predecessor, the W20, to video-monitoring sleep assessment (which is itself unvalidated), and reported 5 sleep duration values, with detailed methodology not given [7].

Actigraphs are commonly validated against polysomnography (PSG) or another actigraph. PSG, though imperfect, is the gold standard of sleep measurement [10]. PSG's complexity requires sleep in a laboratory [11], which is costly and unrepresentative of free-living sleep. Watch-based actigraphy can efficiently collect multiple nights' data in a natural environment [12,13].

Objective

We assessed iAide2-measured sleep duration and total sleep time (TST) in participants without sleep disorders and added calibration when needed.

Methods

Approach

To assess free-living sleep, we validated iAide2 against a second actigraph, the MotionWatch 8 (MW8; CamNtech Ltd) [14], using methods similar to other actigraph-to-actigraph comparisons [11,15-17]. The MW8 is validated against PSG [18-20], with high correlations (r=0.920) and no significant bias (7.0 minutes) for TST [19]. Researchers have used the MW8 to assess sleep in people without sleep disorders [21-23]. We used validated MW8 settings (mode 1, 30-second epochs, and a threshold of 20 [24,25]).

Data Collection

We collected data from a convenience sample of members of the Waki DTx Laboratory at the Graduate School of Medicine, University of Tokyo. The laboratory is investigating lifestyle interventions to improve sleep and exercise and identified the iAide2 as a candidate measurement device. This validation study assessed the accuracy of iAide2 sleep monitoring in 6 individuals without known sleep disorders. They wore both devices on the nondominant arm, with the MW8 closest to the hand, all day except when bathing.

The participants used the MW8 EVENT-marker button to record bedtime (when they began trying to sleep) and risetime (when they stopped trying to sleep). The duration between these times is time in bed (TIB). Our MW8 and iAide2 assessments both used these nonactigraphy measurements.

MotionWare (version 1.4.20; CamNtech Ltd) provided sleep onset ("Fell Asleep" in MotionWare, the earliest time after bedtime when the participant was asleep), sleep offset ("Woke Up," the latest time prior to risetime when the participant transitioned from asleep to awake), and TST (the sum of all minutes of sleep between sleep onset and offset) [25]. From these measurements, we derived sleep onset latency (the duration between bedtime and sleep onset), sleep offset latency (the duration between sleep offset and risetime), and sleep duration (the duration between sleep onset and sleep offset).

We used a similar process with the iAide2, using iAide2 software (version 7.0). We defined sleep onset as the time of the first sleep after bedtime. Similarly, we defined sleep offset as 1 minute after the time of final sleep prior to risetime. We defined TST as the sum of minutes asleep within the sleep duration period.

Statistical Analysis

We compared measurements using intraclass correlation coefficients (ICCs; 2-way mixed effects, absolute agreement, single measurement, unless otherwise specified) [26,27]. We used the *irr* package 0.84.1 in R for point estimation of ICC and calculating 95% CIs, and R (version 4.2.1) for calibration.

Ethical Considerations

The University of Tokyo School of Medicine institutional review board (2024137NI) approved this study, which abided by the Declaration of Helsinki. Participants provided verbal consent for data collection upon recruitment. All participants provided written consent for the use of their data. Participants' data were anonymized. No compensation was provided.

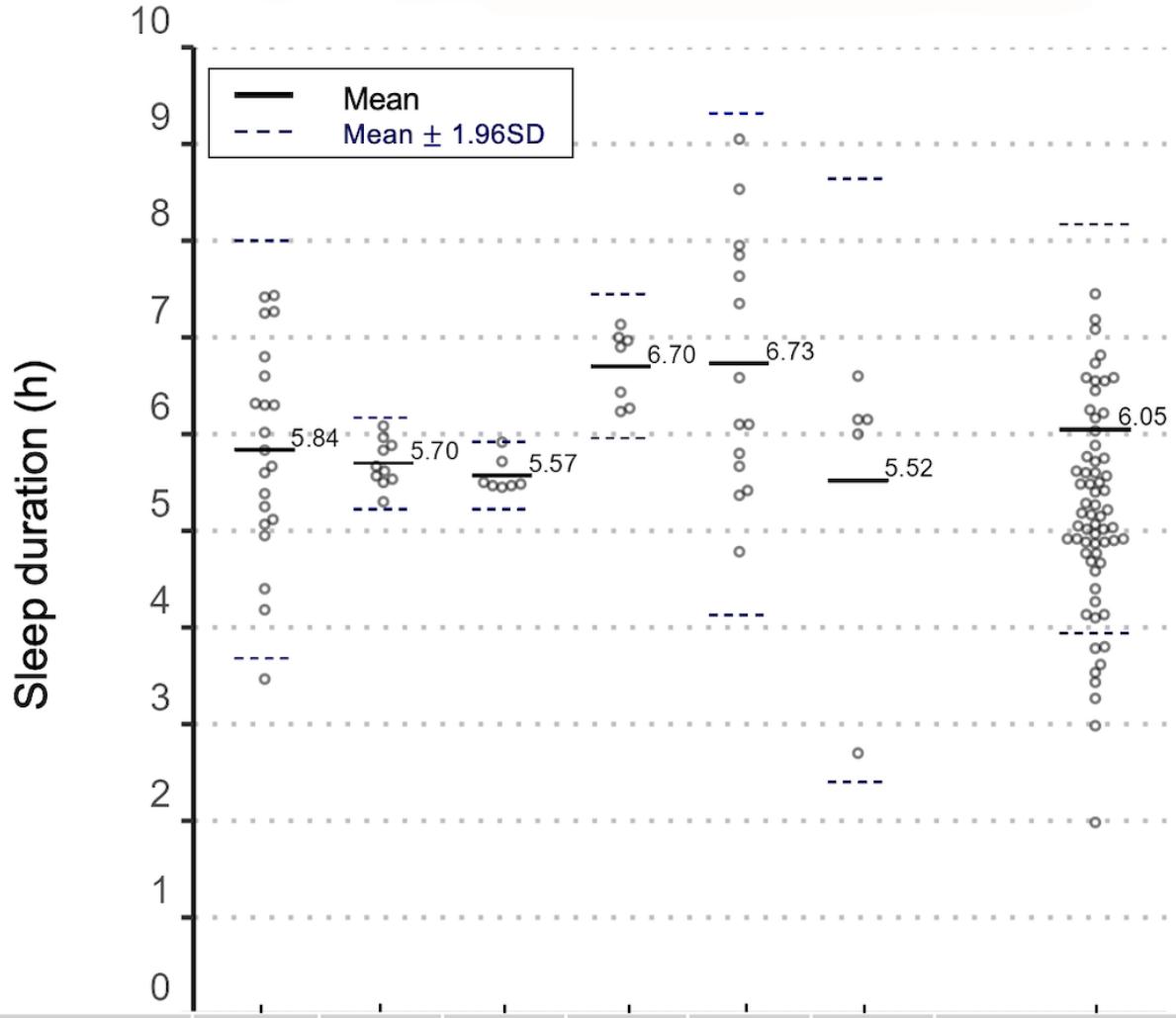
Results

Participants

We recruited 6 participants (aged 24-63 years) and recorded data for 64 nights (Figure 1). The number of nights per participant varied from 5 to 21. The average nightly sleep duration via MW8 per participant varied from 5.52 to 6.73 hours, and the participants differed in the regularity of their sleep durations.

XSL•FO
RenderX

Figure 1. Participants' information and sleep duration measured using the MotionWatch 8 actigraph (MW8; 6 participants, 64 nights).



ID	4	9	10	7	11	12	Total
Nights, n	21	10	7	7	14	5	64
Age (yrs)	24	50	54	63	55	35	
Sex	M	F	F	M	M	F	
Sleep duration (h), mean (95%CI)	5.84 [5.30-6.34]	5.70 [5.52-5.87]	5.57 [5.41-5.74]	6.70 [6.35-7.06]	6.73 [5.96-7.49]	5.52 [3.54-7.50]	6.05 [5.78-6.32]

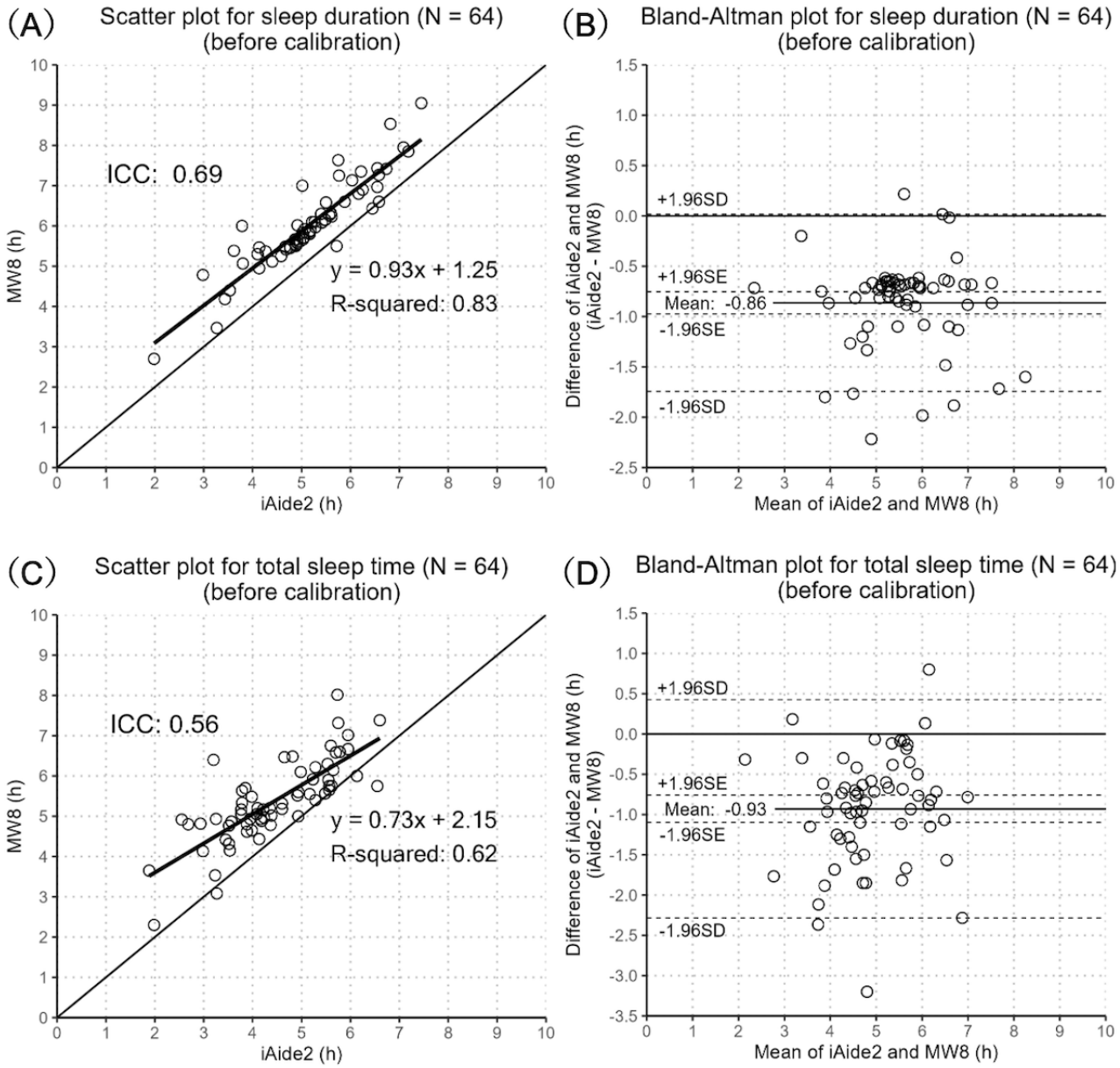
Measurements

The iAide2 underestimated duration relative to the MW8 and showed moderate ICC of 0.69 (95% CI –0.07 to 0.91; Figure 2A) [26]. Relative performance was excellent (ICC=0.91, 95% CI 0.86-0.95; 2-way mixed effects, consistency, single measurement) [26]. The mean, maximum, and minimum differences between iAide2 and MW8 were –0.86, 0.22, and –2.22 hours, respectively (Figure 2B).

Similarly, the iAide2 underestimated TST and showed moderate ICC of 0.56 (95% CI –0.07 to 0.82; Figure 2C) and a good consistency ICC (2-way mixed effects, consistency, single measurement) of 0.78 (95% CI 0.67-0.86). The mean, maximum, and minimum differences between iAide2 and MW8 were –0.93, 0.80, and –3.20 hours, respectively (Figure 2D).

XSL•FO
RenderX

Figure 2. Scatter plot and Bland-Altman plot for measured sleep duration and total sleep time using iAide2 and the MotionWatch 8 actigraph (MW8; 64 nights).



Calibration

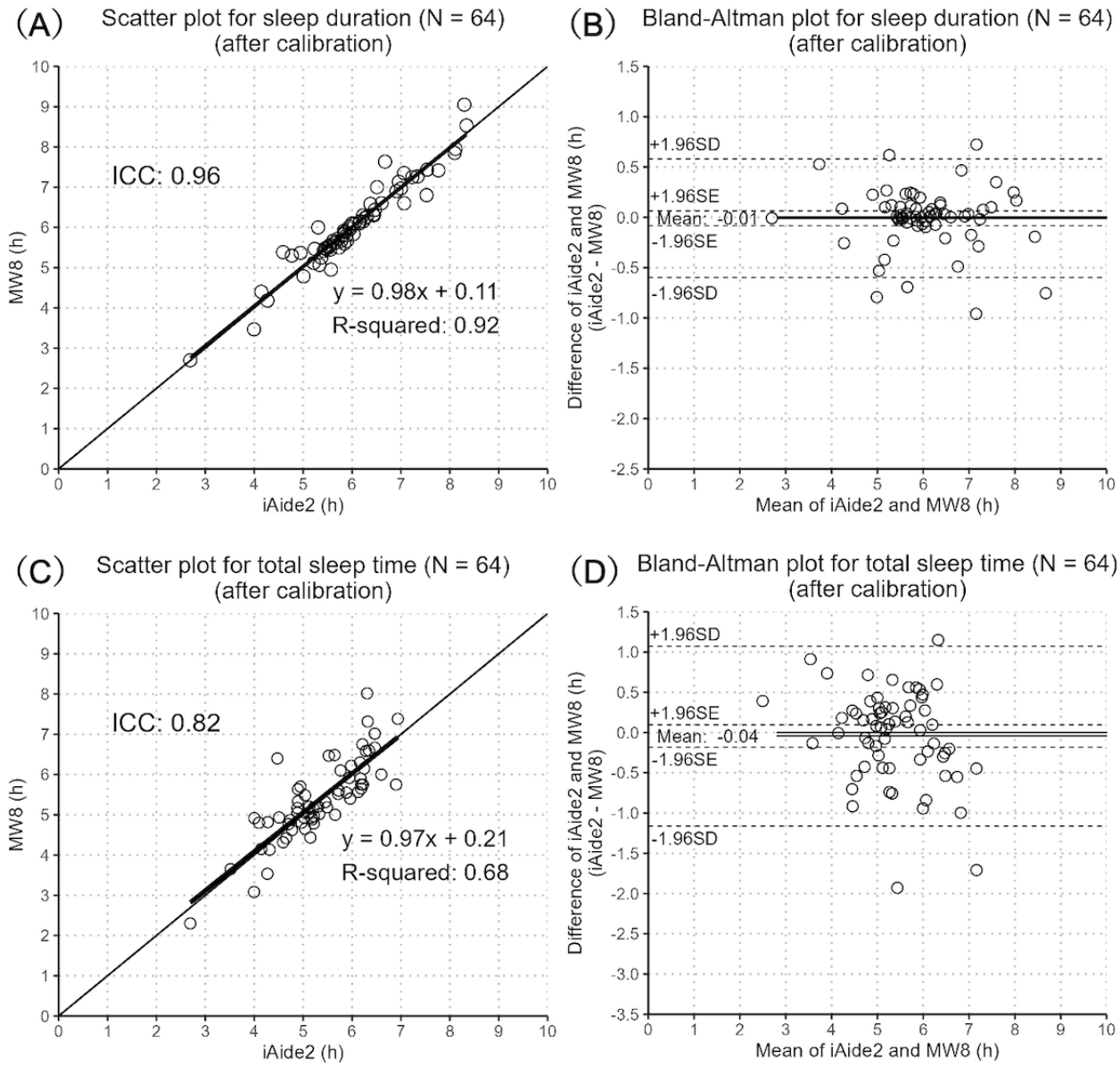
Our iAide2 data showed very small sleep offset latencies (mean 0.03 hours) but significant overestimates of sleep onset latency, so we corrected the sleep duration by the difference between the iAide2-measured sleep onset latency and an iAide2 sleep onset latency calibrated to the MW8 measurements via ordinary least squares (OLS). We added a small (<5 minute) bias correction, selected to optimize ICC, to account for small offset latency biases. To prevent physically impossible results, we limited the calibrated duration to be no longer than the TIB. The resulting calibration equation is D' = min {0.59S + D + 0.26, TIB} (hours), where D is the measured duration, S is the measured sleep onset latency, and D' is the calibrated duration.

For TST, we used OLS to calibrate the measured TST value, limiting the result to be not greater than the calibrated sleep duration. The resulting equation is T' = min {0.73T + 2.15, D'} (hours), where T is the measured TST and T' is the calibrated TST.

Calibration greatly improved accuracy. Sleep duration had excellent ICC (0.96, 95% CI 0.94-0.98; Figure 3A). The mean difference almost disappeared (–0.01 hours) and the difference between the maximum (0.72 hours) and minimum (–0.96 hours) values decreased (Figure 3B). TST similarly had greatly improved accuracy, with good ICC (0.82, 95% CI 0.71-0.88; Figure 3C). The mean difference became almost zero (–0.04 hours) and the difference between the maximum (1.15 hours) and minimum (–1.93 hours) values decreased (Figure 3D).

XSL•FO
RenderX

Figure 3. Scatter plot and Bland-Altman plot for calibrated sleep duration and total sleep time using iAide2 and the MotionWatch 8 actigraph (MW8; 64 nights).



Because our sample size is small with a varying number of nights per participant, we conducted a sensitivity analysis to determine how the ICC agreement rate changes based on the included participants. First, to show that our ICCs are not sensitive to the exact individuals, we performed trials excluding the data of one participant at a time, recalibrating for each trial. The calibrated ICC range was 0.96-0.97 for sleep duration, effectively unchanged from 0.96 in the baseline analysis, and 0.79-0.86 for TST, similar to 0.82 in the baseline analysis, demonstrating that our results are not highly influenced by the variability in number of nights per participant. To demonstrate the generality of our calibration equations, we applied the calibration derived from all participants' data to trials excluding one participant at a time. The ICC range was 0.96-0.97 for sleep duration and 0.79-0.87 for TST, again similar to baseline analysis values, showing that our calibration has high generality and is not highly dependent on any one participant's data.

Discussion

Principal Findings

The absolute accuracies of uncalibrated iAide2 measurements for sleep duration (ICC=0.69) and TST (ICC=0.56) are relatively poor. Uncalibrated data could be useful in applications needing relative measurements, with good values for the consistency version of the ICC (2-way mixed effects, consistency, single measurement) of 0.91 for sleep duration and 0.78 for TST.

The calibrated measurements have excellent absolute accuracy. The ICCs for sleep duration (0.96) and TST (0.82) are comparable to or better than those reported previously (0.67 and 0.68 for sleep duration and 0.77 and 0.84 for TST [28,29]). Sensitivity analysis revealed essentially unchanged performance when applying this calibration to data while removing one participant at a time and when recalibrating while removing one participant at a time. The calibration methods are simple

XSL•FO
RenderX

and have good statistical bases. The calibrated iAide2 measurements of sleep duration and TST seem usable in a wide range of applications.

Limitations

The MW8, our validation standard, has its own errors, with a reported underestimation of sleep onset latency (mean difference 8.9 minutes) and overestimation of TST (mean difference 47.1 minutes) relative to PSG [30]. Validation against other standards, including PSG, is warranted.

We used more nights (N=64) than some PSG-based validation studies (N=20, 37, and 38) but fewer than actigraph-based research (N=140) [17,18,31,32]. Our convenience sample was small (N=6); however, it included a mix of ages, sex, and sleeping regularity. Our results may not fully apply to other populations; and larger studies are warranted.

Conclusions

The iAide2 has interesting measurement capabilities with convenient features, and validating its sleep measurement accuracy adds to its usefulness. The absolute accuracies are low, but good relative accuracy may be suitable for some applications. Its postcalibration accuracies are excellent and well validated against the MW8.

Acknowledgments

We thank Daniel Lane for manuscript editing and scientific discussions.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request for noncommercial purposes.

Authors' Contributions

KW was responsible for the overall design and evaluation of this study. The model was designed and implemented by YB. Data were collected and organized by YB, RN, and MW. Experiments were conducted by YB, RN, and MW. KW and RS supervised the work. The first draft of the manuscript was written by MW. All authors read and approved the final manuscript.

Conflicts of Interest

None.

References

1. iAide2 (iAide2-W). Pharmaceuticals and Medical Devices Agency. 2021. URL: https://www.info.pmda.go.jp/downfiles/md/PDF/340205/340205_301AGBZX00020000_A_01_02.pdf [accessed 2024-02-26]
2. iAide. Tokai Corporation. URL: https://www.iaide-system.jp/iAide/Pages/Login.aspx [accessed 2024-04-10]
3. Watch-type activity meters that can be used in various businesses Silmee W22. TDK Product Center. URL: https://product.tdk.com/ja/products/biosensor/biosensor/silmee_w22/index.html [accessed 2024-02-26]
4. Miyamoto K, Hashimoto K, Kasaoka M, Kakumu M. Wearable sensors corresponding to various applications in medical/healthcare field. 2020 Presented at: 27th International Workshop on Active-Matrix Flatpanel Displays and Devices (AM-FPD); September 1-4, 2020; Kyoto, Japan. [doi: 10.23919/am-fpd49417.2020.9224474]
5. Detection of atrial fibrillation using a deep learning with wearable pulse wave sensor. 2019 Presented at: APHRS (ASIA Pacific Heart Rhythm Society scientific session); Oct 24-27, 2019; Bangkok, Thailand. [doi: 10.1002/joa3.12267]
6. Biometric Sensors: Business collaboration to commercialize 24-hour remote monitoring system for medical and nursing care using wristband-type wearable devices. TDK Corporation. 2017. URL: https://www.tdk.com/ja/news_center/press/20171002_01.html [accessed 2024-02-26]
7. Kimura N, Aso Y, Yabuuchi K, Ishibashi M, Hori D, Sasaki Y, et al. Modifiable lifestyle factors and cognitive function in older people: a cross-sectional observational study. Front Neurol 2019 Apr 24;10:401 [FREE Full text] [doi: 10.3389/fneur.2019.00401] [Medline: 31068892]
8. Tazawa Y, Liang K, Yoshimura M, Kitazawa M, Kaise Y, Takamiya A, et al. Evaluating depression with multimodal wristband-type wearable device: screening and assessing patient severity utilizing machine-learning. Heliyon 2020 Feb;6(2):e03274 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e03274] [Medline: 32055728]
9. Kadoya Y, Khan MSR, Watanapongvanich S, Binnagan P. Emotional status and productivity: evidence from the Special Economic Zone in Laos. Sustainability 2020 Feb 19;12(4):1544. [doi: 10.3390/su12041544]
10. Tryon WW. Nocturnal activity and sleep assessment. Clin Psychol Rev 1996 Jan;16(3):197-213. [doi: 10.1016/0272-7358(95)00059-3]
11. Asgari Mehrabadi M, Azimi I, Sarhaddi F, Axelin A, Niela-Vilén H, Myllyntausta S, et al. Sleep tracking of a commercially available smart ring and smartwatch against medical-grade actigraphy in everyday settings: instrument validation study. JMIR Mhealth Uhealth 2020 Nov 02;8(10):e20465. [doi: 10.2196/20465] [Medline: 33038869]

XSL•FO
RenderX

12. Fekedulegn D, Andrew ME, Shi M, Violanti JM, Knox S, Innes KE. Actigraphy-based assessment of sleep parameters. Ann Work Expo Health 2020 Apr 30;64(4):350-367 [FREE Full text] [doi: 10.1093/annweh/wxaa007] [Medline: 32053169]
13. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. Sleep 2003 May 01;26(3):342-392. [doi: 10.1093/sleep/26.3.342] [Medline: 12749557]
14. MotionWatch 8. CamNtech Ltd. URL: https://www.camntech.com/motionwatch-8/ [accessed 2024-02-26]
15. Morgan PL, Hampton S, Karatziotou A, Zaslona J, Middleton B. Validation of two new activity monitors: motionwatch 8 and pro-diary motion. 2012 Presented at: 21st Congress of the European-Sleep-Research-Society; September 4-8, 2012; Paris, France.
16. Topalidis P, Florea C, Eigl E, Kurapov A, Leon CAB, Schabus M. Evaluation of a low-cost commercial actigraph and its potential use in detecting cultural variations in physical activity and sleep. Sensors (Basel) 2021 May 29;21(11) [FREE Full text] [doi: 10.3390/s21113774] [Medline: 34072347]
17. Shah VV, Brumbach BH, Pearson S, Vasilyev P, King E, Carlson-Kuhta P, et al. Opal actigraphy (activity and sleep) measures compared to ActiGraph: a validation study. Sensors (Basel) 2023 Feb 18;23(4):2296 [FREE Full text] [doi: 10.3390/s23042296] [Medline: 36850896]
18. Elbaz M, Yauy K, Metlaine A, Martoni M, Leger D. Validation of a new actigraph motion watch versus polysomnography on 70 healthy and suspected sleep-disordered subjects. CamNtech Ltd. URL: https://www.camntech.com/Engineering/MotionWare/MotionWatchPSG.pdf [accessed 2023-10-19]
19. Dorffner G, Kemethofer M, Saletu-Zyhlarz G, Simeoni M, Rainer L, Parapatics S, et al. Measuring sleep: a comparison of polysomnography (PSG), wrist actigraphy and a reduced PSG solution. International Society for CNS Clinical Trials and Methodology. URL: https://isctm.org/public_access/Autumn2018/Abstracts/Dorffner-abstract.pdf [accessed 2023-10-17]
20. Yavuz-Kodat E, Reynaud E, Geoffray M, Limousin N, Franco P, Bourgin P, et al. Validity of actigraphy compared to polysomnography for sleep assessment in children with autism spectrum disorder. Front Psychiatry 2019 Aug 2;10:551 [FREE Full text] [doi: 10.3389/fpsyt.2019.00551] [Medline: 31428003]
21. Deantoni M, Reyt M, Baillet M, Dourte M, De Haan S, Lesoinne A, et al. Napping and circadian sleep-wake regulation during healthy aging. Sleep 2024 May 10;47(5):2023 [FREE Full text] [doi: 10.1093/sleep/zsad287] [Medline: 37943833]
22. Illingworth G, Sharman R, Harvey C, Foster RG, Espie CA. The Teensleep study: the effectiveness of a school-based sleep education programme at improving early adolescent sleep. Sleep Med X 2020 Dec;2:100011 [FREE Full text] [doi: 10.1016/j.sleepx.2019.100011] [Medline: 33458648]
23. Nédélec M, Chauvineau M, Guilhem G. On the road to Camarón: the sleep of an ultra-endurance athlete cycling 10,000 km in 24 days. Int J Environ Res Public Health 2022 Apr 09;19(8):4543 [FREE Full text] [doi: 10.3390/ijerph19084543] [Medline: 35457410]
24. The MotionWatch User Guide. CamNtech. 2024. URL: https://www.camntech.com/Products/MotionWatch/The%20MotionWatch%20User%20Guide.pdf [accessed 2024-02-26]
25. MotionWare Software User Guide. CamNtech. 2024. URL: https://www.camntech.com/Products/MotionWatch/The%20MotionWare%20Software%20User%20Guide.pdf [accessed 2024-02-26]
26. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163. [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
27. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996 Mar;1(1):30-46. [doi: 10.1037/1082-989X.1.1.30]
28. Plekhanova T, Rowlands AV, Davies MJ, Hall AP, Yates T, Edwardson CL. Validation of an automated sleep detection algorithm using data from multiple accelerometer brands. J Sleep Res 2023 Jun 31;32(3):e13760. [doi: 10.1111/jsr.13760] [Medline: 36317222]
29. Kainec KA, Caccavaro J, Barnes M, Hoff C, Berlin A, Spencer RMC. Evaluating accuracy in five commercial sleep-tracking devices compared to research-grade actigraphy and polysomnography. Sensors (Basel) 2024 Jan 19;24(2) [FREE Full text] [doi: 10.3390/s24020635] [Medline: 38276327]
30. Ameen MS, Cheung LM, Hauser T, Hahn MA, Schabus M. About the accuracy and problems of consumer devices in the assessment of sleep. Sensors (Basel) 2019 Sep 25;19(19) [FREE Full text] [doi: 10.3390/s19194160] [Medline: 31557952]
31. Kanady JC, Ruoff L, Straus LD, Varbel J, Metzler T, Richards A, et al. Validation of sleep measurement in a multisensor consumer grade wearable device in healthy young adults. J Clin Sleep Med 2020 Jun 15;16(6):917-924 [FREE Full text] [doi: 10.5664/jcsm.8362] [Medline: 32048595]
32. Kosmadopoulos A, Sargent C, Darwent D, Zhou X, Roach GD. Alternatives to polysomnography (PSG): a validation of wrist actigraphy and a partial-PSG system. Behav Res Methods 2014 Dec;46(4):1032-1041. [doi: 10.3758/s13428-013-0438-7] [Medline: 24442593]

Abbreviations

ICC: intraclass correlation coefficient
MW8: MotionWatch 8 actigraph
OLS: ordinary least squares

XSL•FO
RenderX

PSG: polysomnography
TIB: time in bed
TST: total sleep time

Edited by A Mavragani; submitted 23.06.24; peer-reviewed by Y Minami, JE Lee; comments to author 25.10.24; revised version received 11.11.24; accepted 24.11.24; published 06.01.25.

Please cite as:
Waki M, Nakada R, Waki K, Ban Y, Suzuki R, Yamauchi T, Nangaku M, Ohe K
Validation of Sleep Measurements of an Actigraphy Watch: Instrument Validation Study
JMIR Form Res 2025;9:e63529
URL: https://formative.jmir.org/2025/1/e63529
doi:10.2196/63529
PMID:

©Mari Waki, Ryohei Nakada, Kayo Waki, Yuki Ban, Ryo Suzuki, Toshimasa Yamauchi, Masaomi Nangaku, Kazuhiko Ohe. Originally published in JMIR Formative Research (https://formative.jmir.org), 06.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Differences in eHealth Access, Use, and Perceived Benefit Between Different Socioeconomic Groups in the Dutch Context: Secondary Cross-Sectional Study

Lucille Standaar[1,2,3], BSc, MSc; Lilian van Tuyl[2], MBA, PhD; Anita Suijkerbuijk[1], PhD; Anne Brabers[2], PhD; Roland Friele[2,3], Prof Dr

[1]Department of Population Health and Health Services Research, Centre for Public Health, Healthcare and Society, National Institute for Public Health and the Environment, Bilthoven, Netherlands

[2]Department Organisation and Quality of Care, Netherlands Institute for Health Services Research, Utrecht, Netherlands

[3]Tranzo Scientific Center for Care and Wellbeing, Tilburg University, Tilburg, Netherlands

Corresponding Author:
Lucille Standaar, BSc, MSc
Department of Population Health and Health Services Research
Centre for Public Health, Healthcare and Society
National Institute for Public Health and the Environment
PO Box 1
Bilthoven, 3723 BA
Netherlands
Phone: 31 302743243
Fax: 31 302744407
Email: lucille.standaar@rivm.nl

Abstract

Background: There is a growing concern that digital health care may exacerbate existing health disparities. Digital health care or eHealth encompasses the digital apps that are used in health care. Differences in access, use, and perceived benefits of digital technology among socioeconomic groups are commonly referred to as the digital divide. Current research shows that people in lower socioeconomic positions (SEPs) use eHealth less frequently.

Objective: This study aims to (1) investigate the association between SEP and eHealth access to, use of, and perceived benefit within the adult Dutch population and (2) evaluate disparities in eHealth access, use, and perceived benefit through three socioeconomic variables—education, standardized income, and the socioeconomic status of the neighborhood.

Methods: A secondary analysis was conducted on data from the Nivel Dutch Health Care Consumer Panel (response rate 57%, 849/1500), to assess access to, use of, and perceived benefits from eHealth. These data were collected to monitor eHealth developments in the Netherlands. eHealth was examined through two concepts: (1) eHealth in general and (2) websites, apps, and wearables. Results were stratified into 9 SEP populations based on 3 indicators—education, standardized income, and socioeconomic status level of the neighborhood. Logistic regression analyses were performed to evaluate whether the outcomes varied significantly across different SEP groups. Age was included as a covariate to control for confounding.

Results: This study confirms the association between eHealth and SEP and shows that low SEP respondents have less access (odds ratio [OR] 5.72, 95% CI 3.06-10.72) and use (OR 4.96, 95% CI 2.66-9.24) of eHealth compared to medium or high SEP respondents. Differences were most profound when stratifying for levels of education.

Conclusions: The access to and use of eHealth has a socioeconomic gradient and emphasizes that SEP indicators cannot be used interchangeably to assess eHealth access and use. The results underline the importance of activities and policies aimed at improving eHealth accessibility and usage among low SEP groups to mitigate disparities in health between different socioeconomic groups.

(JMIR Form Res 2025;9:e49585) doi:10.2196/49585

XSL•FO
RenderX

KEYWORDS

eHealth; digital divide; socioeconomic factors; education; income; neighborhood; health disparities; cross-sectional studies; digital health care; health equity; Netherlands

Introduction

Digital health care is expected to provide benefits for health care systems, providers, and patients and is considered a solution to address workforce shortages and rising health care costs [1-4]. Moreover, digital health care is anticipated to enhance the quality of care, stimulate patient self-management, and improve health accessibility and equity [1-4]. Digital health care or eHealth is defined as the digital apps that are used in health care. Health care policies focus on increased use of and dependency on the eHealth apps [4-6]. Concerns have been raised that digital health care may not be equally accessible for all [7-13]. Populations with lower socioeconomic positions (SEPs) are more likely to encounter financial, skill, or cognitive barriers to accessing and using eHealth, such as limited access to devices, limited digital health skills, or limited ability to take the initiative in using eHealth [7,11-16]. Additionally, these populations experience challenges in comprehending and implementing health information and healthy behaviors in daily life [17-21].

The demand for health care services is often higher in low-SEP populations, as people with a low SEP endure more often from chronic illnesses [18,19,22,23]. Research shows that low-SEP populations often have different views on health and the possible benefits of healthy behaviors compared to high-SEP populations [24-27]. Studies find that low-SEP populations have less time, more stress, and limited financial capacities to implement healthy behaviors [26]. Next to this, it is theorized that health beliefs often find origin in the health beliefs of previous generations [24-27]. In the context of low SEP, the expectation of a shorter life and the belief that their own behavior has limited influence on their longevity pose barriers to adhering to healthy behaviors [20]. Therefore, the digitalization of health care could seriously impact the access and use of health care for those who need it most [5,6,13,28-31].

National digital connectivity and policies that stimulate the transition toward a digital health care system could improve the implementation and accessibility of eHealth. The European Union and its member states deploy policies to realize the digital transition of health care systems [5,6]. Some states, such as the Netherlands, have been experimenting with digital health care for over a decade [4,6]. Most Dutch households (98%) have fast broadband coverage (2020) and 88% of the Dutch population uses mobile broadband (2019), which indicates the use of a mobile phone or other device with mobile internet access [5]. The level of connectivity in the Netherlands could facilitate the implementation of digital health care [5].

Several countries have developed national eHealth monitoring programs to monitor the uptake and effects of eHealth among health care professionals and citizens [32-34]. From a citizen's perspective, the monitoring programs focus on the use and evaluation of eHealth that involve citizen interaction [32-34]. This includes applications such as websites, apps, and wearables that citizens can use independently or involve eHealth tools that facilitate digital communication with a health care professional, such as video calls or messaging via patient portals [32-34]. This study is based on secondary data analysis of the Dutch eHealth monitoring program (2021) [29], which collects data about access, use, and perceived benefits from eHealth among Dutch citizens through questionnaires.

There is still limited understanding of the relationship between SEP and eHealth access, use, and perceived benefit and that understanding is generally limited to either access, use, or perceived benefit from eHealth, specific eHealth apps or specific subpopulations. Research showed the relation between SEP and the use of personal health records [35] and mobile apps [36]. Other research focuses on either the benefit from [37] or the use of eHealth [38,39] or specific patient groups such as cancer survivors [40,41], or citizens bound to specific locations [42]. To our knowledge, insight into differences in eHealth access, use, and perceived benefits and how different indicators for SEP display these differences within the Dutch general population aged 18 years and older are largely unknown. This study aims to assess differences in eHealth-related access, usage, and perceived benefits for different socioeconomic populations, based on education, standardized income, and socioeconomic status (SES) level of the neighborhood in the Netherlands. The findings of this study give insight into the disparities in access to, use of, and perceived benefit from eHealth in a highly connected country with an increasingly digitalized health care system. The results are insightful for other contexts that aim for or experience the same ambitions to transition to a digital health care system.

Methods

Panel

Data from the Dutch Health Care Consumer Panel (DHCCP) were used [43]. The DHCCP is a panel managed by Nivel (the Netherlands Institute for Health Services Research) and currently (as of September 2023) consists of approximately 11,500 panel members aged 18 years and older [43]. For this study, a study sample of 1.500 panel members was drawn by researchers from DHCCP. The study sample was representative of the Dutch population aged 18 years and older regarding age and sex [43]. Background characteristics of panel members, including their sex, age, level of education, net monthly income per household, and 4-digit postal codes were known. The panel was periodically renewed to ensure representative samples of the adult population in the Netherlands can be drawn. New panel members were recruited by buying an address file from an address supplier [43]. As a result, possible new members were sampled at random from the general population in the Netherlands [43]. The panel could only be joined through invitation. It was not possible for people to sign up on their own initiative [43]. Upon membership, panel members were informed of the purpose, scope, method,

XSL•FO
RenderX

and use of the panel [43]. Based on this information, participants could give permission to participate in the panel [43].

Ethical Considerations

According to the Dutch legislation, neither obtaining informed consent nor approval by a medical ethics committee was obligatory for doing research within the DHCCP [43,44]. Data analysis was conducted with pseudonymized data, according to the privacy regulations of the DHCCP, in compliance with the General Data Protection Regulation [44]. The privacy of the panel members was protected. All data were carefully stored by Nivel [43]. Personal information such as addresses was stored separately from the data of the questionnaires [43]. The privacy of the panel members in the study sample was guaranteed by DHCCP [43]. The researcher (LS) who analyzed the data had no access to the personal information of the panel members [43]. A written or digital informed consent was obtained at the time of registration of a new member to the panel [43]. Panel members were asked to participate approximately 4 or 5 times per year [43]. Participation was voluntary.

Data Collection

Data on the population's perspective on eHealth were collected via the DHCCP as part of a larger monitoring study into the perceptions, experiences, and usage of eHealth in the Netherlands [43]. A questionnaire was developed and reviewed by a team of representatives from the health care field in the Netherlands. The questionnaire was based on earlier distributed questionnaires of the monitoring study and was adjusted to reflect market developments [34]. The questionnaire was distributed via email and post (according to the preferences of the panel members) in May 2021. A digital reminder was sent after 1, 2, and 3 weeks after the start of the questionnaire and 1 written reminder was sent after 2 weeks. Panel members had 4 weeks to respond.

Socioeconomic Position Indicators

Overview

The concept of SEP is complex, as it results from the interaction between individual, social, economic, cultural, and societal factors [45,46]. In this study, 3 different operationalizations of SEP were used to study the digital divide in the context of eHealth—education as a historic starting point, standardized household income as a measure of current wealth, and SES level of the neighborhood to include environmental influence [46-49].

Education

The education levels were defined as low (none, primary school, or prevocational education); medium (secondary or vocational education level 1, 2, 3, or 4); and high (professional higher education or university) [50].

Standardized Income

Standardized income was defined as the net monthly income of the household adjusted for number of household members. The net income was converted to the equivalent of the net income of a single adult household by using equivalence factors from Statistics Netherlands (the Dutch Institute for Population Statistics) [51]. Some respondents acknowledged having children or other adults living in their household apart from their partner or children older than 18 years of age but did not specify the number. In the case of an unknown amount of children, 1.57 children were assumed. In case of an unknown amount of extra adults 1 extra adult was assumed. Assumptions were derived from the averages in Dutch households [50,52,53]. Information gathered from the panel members about their monthly net income was in ranges, and the mean of the range was taken as the monthly net income. Standardized income was divided into three categories (1) low (between €0 and €1659 [US $1718.56] per month; The used conversion rate was applicable on May 1, 2021), (2) medium (between €1660 [US $1719.59] and €2332 [US $2415.72] per month), and (3) high (more than €2332 [US $2415.72] per month). The categories were derived from the quartile distribution of the net income of the Dutch households (2020) [54].

Socioeconomic Status Level of the Neighborhood

The SES level of the neighborhood of all respondents was determined using the Social Economic Status-Wealth Education Employment (SES-WOA) score (2019) from the Statistics Netherlands. The SES-WOA score was based on the wealth, educational status, and recent employment history of households in the neighborhood [55,56]. The SES-WOA score was matched to the respondent by the 4-digit postal code. The neighborhoods of the respondents were categorized as (1) low (first tertile of SES score: –0.89 to 0.042), (2) medium (second tertile of SES score: 0.043-0.21), and (3) high (third tertile of SES score: 0.21-0.71). The average score of Dutch neighborhoods was 0.092 (SD 0.23; range –0.89 to 0.71) [55,56].

eHealth and the Digital Divide

For the interpretation of the data, the digital divide model was used. The digital divide model published by van Dijk et al [57] conceptualized that individuals' SEPs influence the available resources to access, use, and benefit from new digital media. In this study, the digital media in focus was eHealth.

Two concepts of eHealth were examined: (1) eHealth in general and (2) websites, apps, and wearables. Items in the questionnaire that informed these 2 concepts were matched to the 3 levels of the digital divide model [57], namely level 1—access, level 2—use, and level 3—perceived benefit. Operationalization of the variables measuring access, use, and perceived benefit for eHealth in general and websites, apps, and wearables can be found in Multimedia Appendix 1.

First, eHealth in general was studied to gain insight into an overall interest toward digital apps in health care. The levels of digital divide that were studied are access and the perceived benefit. Access was measured by motivation. Here respondents were asked what their general thoughts are about digital apps used in health care. The perceived benefit was operationalized by measuring to what extent the respondents perceived themselves as making more conscious decisions about their health as a result of eHealth use.

Second, eHealth in terms of websites, apps, and wearables was studied to gain insight into the use of these specific eHealth tools to improve health or provide support in coping with a disease. The digital divide levels that were taken into account

XSL•FO
RenderX

are access and use. Access was measured by motivation and physical access. Motivation was measured by asking the respondents if they have used or would like to use websites, apps, and wearables for their health. For physical access, the respondents were asked if they have access to an electronic device with internet.

The digital divide concept use was measured by barriers in use, diversity of use, and frequency of use. For the concept barriers in use, the respondents were asked if they experienced barriers in the use of websites, apps, and wearables. The use of websites, apps, and wearables was further operationalized in 2 variables—diversity of use and frequency of use. For these variables, the respondents were asked about 16 different websites, apps, or wearables if they used the app (once or more than once). Diversity entails the variety of websites, apps, and wearables used, while the frequency of use operationalizes the number of times (more than once) the apps were used.

Statistical Analysis

Descriptive analyses were used to describe the demographics and the outcome of the variables measuring the access, use, and perceived benefit for eHealth in general and websites, apps, and wearables. The variables measuring access, use, and perceived benefit were constructed by combining items from the original questionnaire. The operationalization of these variables can be found in Multimedia Appendix 1.

The outcomes were stratified by the 3 variables of SEP—educational level, standardized income, and SES level of the neighborhood. The differences in eHealth access and usage between SEP populations were investigated using logistic regression analysis. Ordered logistic regression was used for testing the differences in perceived benefit between SEP populations. Age was included in the analysis to test for confounding, as age is associated with both health and familiarity and use of digital media [10,58,59]. The correlation between the independent variables was determined via the Spearman rank order correlation coefficient. Data analysis was conducted using the Stata Statistical Software release (version 16.1; StataCorp) [60]. A $P<.05$ was considered statistically significant. Variables of physical access and diversity of use were not included in the logistic regression analysis because there were too few cases in outcome categories to meet the assumptions of the logistic regression analysis. The univariate outcomes are presented in Multimedia Appendix 2.

Results

Sample Characteristics

In total, 849 panel members responded to the questionnaire, resulting in a response rate of 56.6% (849/1500). Among the panel population of 1500, 8.9% (133/1500), 40% (600/1500), and 48.7% (731/1500) had a low, medium, and high level of education, respectively. Of these groups, 55.6% (74/133), 59.3% (356/600), and 54.3% (397/731) responded to the questionnaire. Regarding standardized income 35.8% (537/1500), 31.8% (477/1500), and 27.9 (419/1500) had low, medium, and high levels of standardized income, respectively. Of these groups, 57.2% (307/537), 56.4% (269/477), and 54.9% (230/269) had responded to the questionnaire. Finally, for SES-level of the neighborhood, 42.8% (642/1500), 32.6% (489/1500), and 23.1% (346/1500) had low, medium, and high levels of SES-level of the neighborhood in the panel population, respectively. Of these groups, 56.1% (360/642), 58.1% (284/489), and 55.2% (191/346) responded to the questionnaire.

The demographics of the study population can be found in Table 1. An overview of the study population stratified by education, standardized income, and SES level of the neighborhood can be found in Multimedia Appendix 3. Overall, the sample contained the same distribution of males or females as in the general population. When stratified for age category, our sample contained slightly more respondents aged 40 years and older and fewer respondents aged 18-39 years as compared to the general population. A frequency table of males and females in 3 age categories from the study population and the general Dutch population can be found in Multimedia Appendix 4 [61]. Compared to the Dutch general population, the study population had more males (11.3% general population, 14.9% study population) and females (13.1% general population, 15.4% study population) in the age category of 65 years and older. In terms of sex, the distribution was equal between both populations (Dutch general population of 49.3% male and 50.7% female, study population male 48.7% and female 51.3%). The mean age was 54 (SD 16.96) years. The lowest number of respondents was in the low educational level subpopulation (74/849, 8.72%) and the highest in the high educational level subpopulation (397/849, 46.76%). A high educational level was significantly and positively associated with a high standardized income level, as indicated by the correlations of educational level and standardized income level: $\rho=0.37$ ($P<.001$), educational level and SES level of the neighborhood: $\rho=0.021$ ($P=.55$), and standardized income level and SES level of the neighborhood: $\rho=0.035$ ($P=.33$) [62].

XSL•FO
RenderX

Table 1. Demographic description of the study population (n=849). Study population was sampled (2021) from a representative population (N=1500) of general Dutch population aged 18 years and older.

	Total (n=849)
Sex, n (%)	
Male	413 (48.6)
Female	435 (51.2)
Missing	1 (0)
Age (years)	
Mean (SD)	54 (16.96)
Range	19-92
Average household size	
Mean (SD)	2.29 (1.11)
Range	1-7
Number of households with children younger than 18 years, n (%)	
No	604 (71.1)
Yes	235 (27.7)
Missing	10 (1.2)
Number of households with children older than 18 years, n (%)	
No	763 (89.9)
Yes	76 (9)
Missing	10 (1.2)
Education	
Mean (SD)	2.39 (0.65)
Range	1-3
Low, n (%)	74 (8.7)
Medium, n (%)	356 (41.9)
High, n (%)	397 (46.8)
Missing, n (%)	22 (2.6)
Standardized income	
Mean (SD)	1.90 (0.81)
Range	1-3
Low, n (%)	307 (36.2)
Medium, n (%)	269 (31.7)
High, n (%)	230 (27.1)
Missing, n (%)	43 (5.1)
SES[b] level of the neighborhood	
Mean (SD)	1.80 (0.79)
Range	1-3
Low, n (%)	360 (42.4)
Medium, n (%)	284 (33.5)
High, n (%)	191 (22.5)
Missing, n (%)	14 (1.7)

[a]Education level—low (none, primary school or prevocational education); medium (secondary or vocational education level 1, 2, 3, or 4); and high (professional higher education or university). Standardized income was divided into 3 categories—low (between €0 and €1659 [US $1718.56] per month); medium (between €1660 [US $1719.59] and €2332 [US $2415.72] per month); and high (more than €2332 [US $2415.72] per month). The

XSL•FO
RenderX

SES level of the neighborhood was determined using the Social Economic Status-Wealth Education Employment (SES-WOA) score (2019) from Statistics Netherlands. The SES-WOA score was based on the wealth, educational status, and recent employment history of households in the neighborhood [55,56]. Categories: low (first tertile of SES score: –0.89 to 0.042); medium (second tertile of SES score: 0.043-0.21); and high (third tertile of SES score: 0.21-0.71). Not all values add up to 100% due to missing values.

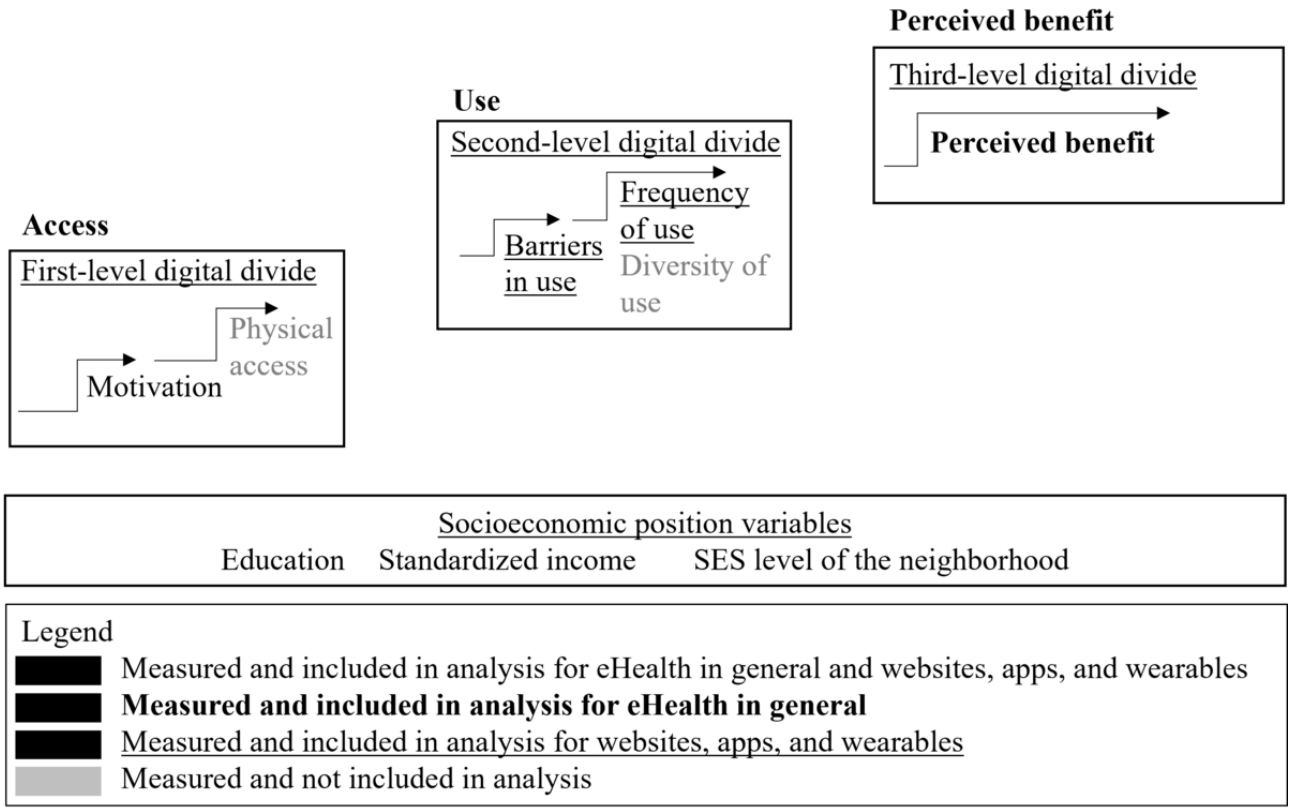
[b]SES: socioeconomic status.

Results of the Relation Between the Digital Divide Levels and the Socioeconomic Position Indicators

Overview

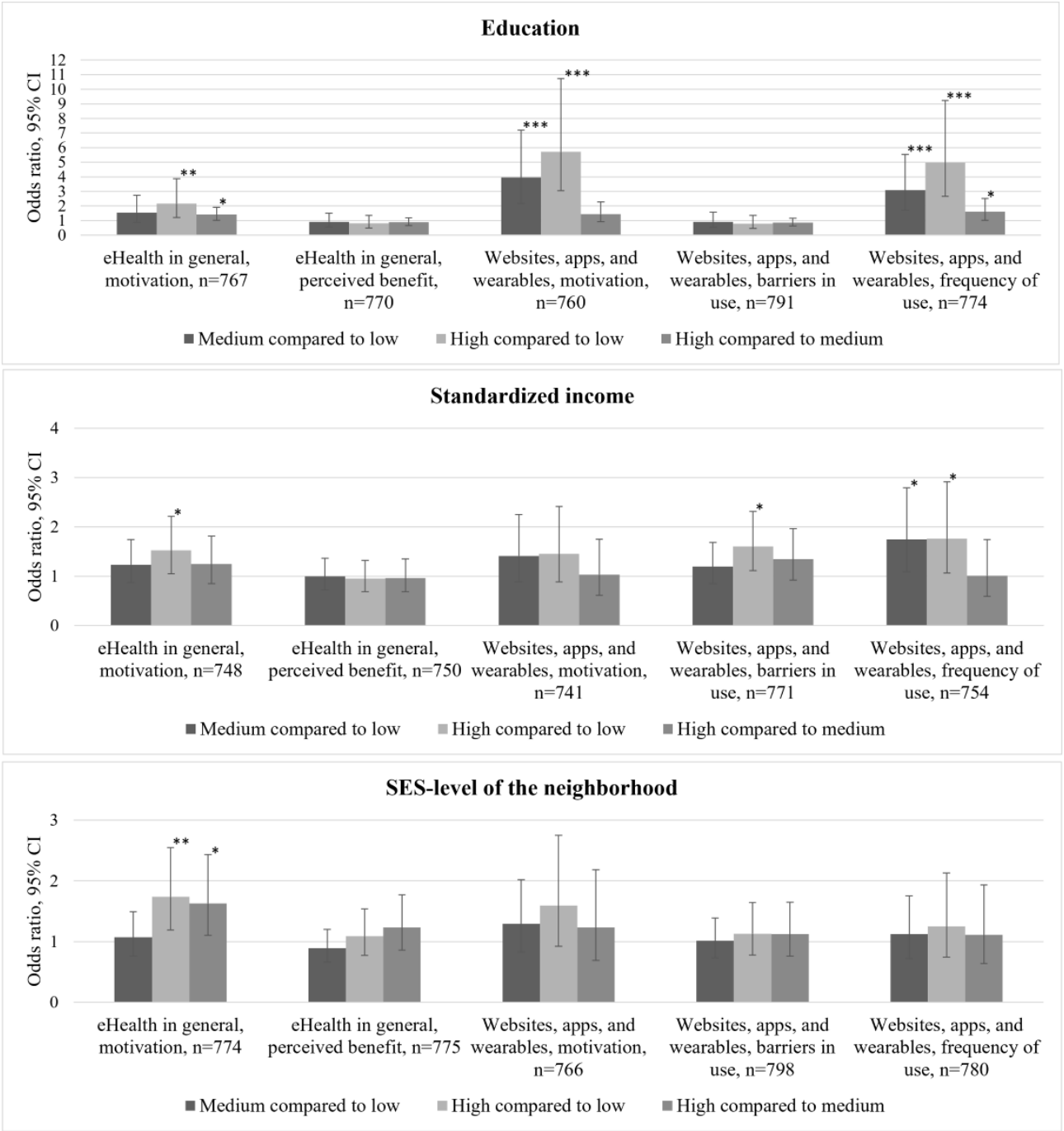
An overview of the measured and analyzed variables can be found in Figure 1. The frequencies of these outcomes can be found in Multimedia Appendix 2. Table 2 presents a descriptive overview of access, use, and perceived benefit, stratified by SEP indicators. The results showed that the outcome differed the most when the population was stratified by education. Figure 2 presents the associations between access; use; perceived benefit; low, medium, and high education; standardized income; and SES level of the neighborhood populations. The underlying data used for Figure 2 is presented in Multimedia Appendix 5.

Figure 1. Overview of the concepts of access, use, and perceived benefit, the variables matched to these concepts and the socioeconomic position indicators used in this study. Legend indicates whether the variables were matched for eHealth in general (bold), websites apps and wearables (underlined) or both. Second, the legend indicates whether the variables were viable, either included (black) or not included (grey), for logistic (ordered) regression analysis [57]. SES: socioeconomic status.



XSL•FO
RenderX

Figure 2. The relation between access, use, and benefit of eHealth and low, medium, and high levels of 3 SEP indicators—education, standardized income, and SES-level of the neighborhood. Results from a questionnaire (2021) answered by a representative study population (n=849) for the general Dutch population aged 18 years and older. Access: motivation, use: barriers in use and frequency of use, and perceived benefit: perceived benefit. For each SEP indicator, a bar graph of the results of logistic (ordered) regression analysis was presented. Each bar shows the odds ratio (OR) and the 95% CI for the difference between SEP levels for each digital divide concept. For each SEP indicator, the comparisons made were medium compared to low (dark grey), high compared to low (light grey), and high compared to medium (medium grey). The number of respondents (n) included in the analysis for each digital divide concept was indicated. The variable barriers in use were recoded to ensure that a positive score (1) reflected the outcome: no experienced barriers. Positive ORs (OR>1) should be interpreted that the primary group in the comparison, in the case of this study either medium or high SEP, had more likelihood of not experiencing barriers in use. *P>.05, **P>.01, ***P>.001. SEP: socioeconomic position; SES: socioeconomic status.



XSL•FO
RenderX

Table 2. Frequency distribution of access, use, and perceived benefit for (1) eHealth in general and (2) websites, apps, and wearables, stratified by education, standardized income, and SES[a] level of the neighborhood. Frequencies are derived from results of a questionnaire (2021) conducted among a sample of the general Dutch population (N=1500), final study population (n=849)[b,c].

	Total	Education			Standardized income			SES level of the neighborhood		
		Low	Med[d]	High	Low	Med	High	Low	Med	High
Population, n	849	74	356	397	307	269	230	360	284	191
eHealth in general										
Access—motivation, n (%)										
No motivation	386 (45.5)	46 (62.2)	177 (49.7)	152 (38.3)	153 (49.8)	124 (46.1)	90 (39.1)	174 (48.3)	141 (49.6)	68 (35.6)
Motivation	401 (47.2)	21 (28.4)	153 (43.0)	218 (54.9)	131 (42.7)	127 (47.2)	123 (53.5)	157 (43.6)	127 (44.7)	107 (56)
Perceived benefit—perceived benefit, n (%)										
Totally disagree—disagree	211 (24.9)	17 (23.0)	80 (22.4)	108 (27.2)	70 (22.8)	68 (25.3)	65 (28.3)	85 (23.6)	76 (26.7)	46 (24.1)
Not agree nor disagree	370 (43.6)	33 (44.5)	171 (48.0)	159 (40.1)	144 (46.9)	119 (44.2)	89 (38.7)	158 (43.9)	134 (47.2)	75 (39.3)
Agree—totally agree	205 (24.1)	17 (23.0)	81 (22.8)	104 (26.2)	69 (22.5)	65 (24.2)	61 (26.5)	87 (24.2)	61 (21.5)	53 (27.7)
Websites, apps, and wearables										
Access—motivation, n (%)										
No motivation	143 (16.8)	35 (47.3)	62 (17.4)	41 (10.3)	59 (19.2)	43 (16.0)	34 (14.8)	65 (18.1)	50 (17.6)	23 (12.0)
Motivation	636 (75.0)	29 (39.2)	265 (74.4)	328 (82.6)	221 (72.0)	206 (76.6)	178 (77.4)	263 (73.1)	215 (75.7)	150 (78.5)
Use—barriers in use, n (%)										
Experienced barriers	339 (40.0)	29 (39.2)	140 (39.3)	159 (40.1)	135 (44.0)	109 (40.5)	76 (33.0)	145 (40.3)	119 (41.9)	71 (37.2)
No experienced barriers	472 (55.6)	39 (52.7)	201 (56.5)	223 (56.2)	157 (51.1)	149 (55.4)	145 (63.0)	197 (54.7)	156 (54.9)	110 (57.6)
Use—frequency of use, n (%)										
No frequent use	131 (15.4)	29 (39.2)	60 (16.9)	38 (9.6)	59 (19.2)	36 (13.4)	29 (12.6)	57 (15.8)	47 (16.5)	24 (12.6)
Frequent use	662 (78.0)	34 (45.9)	274 (77.0)	339 (85.4)	222 (72.3)	219 (81.4)	189 (82.2)	277 (76.9)	226 (79.6)	149 (78.0)

[a]SES: socioeconomic status.

[b]The variables physical access and diversity of use for websites, apps, and wearables were not presented in Table 2 because these variables could not meet the assumptions for logistic regression analysis. The frequencies of these outcomes can be found in Multimedia Appendix 2.

[c]Not all values add up to 100% due to missing values.

[d]Med: medium level.

Access and Motivation

For eHealth in general, as well as for websites, apps and wearables, differences in motivation were found between different levels of education. Differences in motivation were most profound between low or medium versus highly educated respondents in both eHealth in general (odds ratio [OR] 2.18, 95% CI 1.22-3.88) and websites, apps, and wearables (OR 5.72, 95% CI 3.06-10.72). Regarding standardized income, a difference in motivation for eHealth in general was found between high and low standardized incomes (OR 1.52, 95% CI 1.05-2.21). A significant difference in motivation between low (OR 1.74, 95% CI 1.19-2.55) and medium (OR 1.63, 95% CI 1.1-2.43) versus high SES level of the neighborhood was found for eHealth in general.

Use, Barriers in Use, and Experienced Barriers

High standardized income was associated with no experience of barriers in use in comparison to low and medium levels of standardized income. This implies that fewer respondents with a high standardized income experienced barriers while using eHealth websites, apps, and wearables (OR 1.60, 95% CI 1.11-2.31). The frequency of eHealth use also differed between respondents with a low, medium, or high level of education,

XSL•FO
RenderX

with the most significant difference between high and low educational levels (OR 4.96, 95% CI 2.66-9.24). In terms of standardized income, there were significant differences between low and high (OR 1.76, 95% CI 1.06-2.91) and low and medium (OR 1.74, 95% CI 1.09-2.79) standardized income levels. High SEP respondents were more likely to frequently use an eHealth app, website, or wearable compared to medium or low SEP respondents.

Perceived Benefit

There were no significant differences found regarding the perceived benefits between low, medium, or high-SEP populations.

Discussion

Principal Findings

This study shows that low SEP respondents have less access to and use of eHealth compared to medium or high SEP respondents. The most significant digital divide observed in this study is related to educational background. The results of this paper contribute on a population level to previous findings that the access and use of eHealth has a socioeconomic gradient. Additionally, the results emphasize that SEP indicators cannot be used interchangeably to assess eHealth access and use.

The results of this study highlight that, across all 3 SEP indicators, the most substantial differences are found in access through motivation. Respondents from higher socioeconomic categories expressed greater motivation to use eHealth, including websites, apps, and wearables, in comparison to those from lower SEPs. Health equity researchers emphasize that comprehending how people perceive health and health care is a complex issue influenced by several societal, contextual, social, and biological individual factors. Therefore, a multidimensional and multicausal approach is necessary to comprehend these disparities [24,63,64]. Weiss et al [28] discuss existing literature and multiple theories as to why differences in eHealth access and use exist between socioeconomic groups. The literature described that the social position of individuals and the influence of the context and organizations surrounding the individual play a role in whether individuals choose to consume digital health care or not [28]. Other literature emphasized that the diffusion of digital health care in society will decrease the digital divide gap as the low-SEP population is assumed to be the latest to adopt [28]. However, the role of health care organizations, social policies, and political decisions that impact individuals' motivations toward eHealth has not been adequately studied. Current national Dutch policies concerning eHealth are focused on the development and interoperability of eHealth apps, the digital skills of health care professionals and the use of eHealth by older people at home, resulting in eHealth becoming an essential part of the health care system [65]. In countries where health care digitalization is progressing, it would be valuable to examine the potential influence of the government, health care organizations, and businesses on the motivation of low socioeconomic populations toward eHealth. Such research could offer valuable insights into the societal and policy changes required to make eHealth more appealing to low-SEP populations.

In this study, use was examined by studying the barriers in use and frequency of use of eHealth websites, apps, and wearables. The results demonstrated that respondents with a high standardized income level infrequently experienced barriers to using eHealth websites, apps, and wearables compared to lower standardized income levels. Previous studies show that highly educated people often have higher health literacy levels and digital skills and are more in contact with the digital world via their education or profession [10,31,59,66-70]. Other studies show that in the development of eHealth new apps are often pilot-tested by highly educated respondents and, therefore, might be more tailored to the needs of highly educated individuals [9,71,72]. In contrast, other studies point out the variety of health behaviors and health care use within SEP groups. De Boer et al [73] show that low SEP groups have more health care costs but that healthy lifestyle behavior such as smoking or being member of a sports club are attributed greatly to the variety in health care use in each socioeconomic group. In the light of eHealth, Agachi et al [74] showed that the user interface and the type of eHealth offered attributes to the use of eHealth between socioeconomic groups. Results revealed that for the same eHealth program, people living in a low socioeconomic neighborhood use the app-based tool more than people living high in a socioeconomic neighborhood. For the web-based version, results show the opposite emphasizing the importance of the user interface and the accessibility to digital devices, as is also theorized in the digital divide model [49,74]. Both De Boer [73] and Agachi [74] show that behavior-related and technical-related factors play a role in the use of health care and eHealth. The results of this study and other studies showed that understanding and creating insight into the existence of and possible solutions for health disparities is dependent on multiple dimensions. Future research into how the socioeconomic gradient in eHealth access and use are associated with other behavioral and technical factors is important to create an in-depth understanding of disparities in eHealth access and use that can inspire research, policy, and practice.

Results pointed out that high education and high standardized income levels were associated with frequent use of websites, apps, and wearables. This is in accordance with previous research, which shows that a high level of education and income is found to be associated with more access to and use of eHealth [12,13,31]. Surprisingly, no difference was found in the perceived benefits (making more conscious decisions in health because of eHealth use) in any of the SEP indicators, although the frequency of use of websites, apps, and wearables was high (78%) and did show significant differences. This implies that respondents with more frequent use of websites, apps, and wearables had the same perceived benefit, namely, making more conscious decisions due to eHealth use, compared to respondents who have not used websites, apps, and wearables once. Evaluation studies show that high SEP respondents have better outcomes from eHealth use than low SEP respondents [10,30,31,59]. The results of this study might indicate that in the real-life context, even though current eHealth apps might be more suitable for highly educated individuals, eHealth is not used appropriately or with similar discipline as in the clinical trial context.

XSL•FO
RenderX

Strengths and Limitations

The strengths of this study were the use of a large and representative sample of the Dutch population and the use of 3 SEP operationalizations to provide a broad insight into the effects of SEP on the digital divide. This study also has some limitations. Despite a large number of respondents, the skewed distribution of outcomes across SEP levels hampered the performance of multivariable logistic regression analysis [62]. Next to this, the SEP indicators used are focused on social demographic and economic aspects of SEP. Cultural and other social aspects, such as social networks and cultural background, have not been taken into account.

The questionnaire used in the study was not designed to measure all the different aspects of the digital divide model and is a secondary analysis of the data gathered. Although the majority of digital divide levels could be well matched with questionnaire items, data on the second level digital divide for eHealth in general and the third digital divide level for websites, apps, and wearables was lacking. In some cases, with emphasis on the concept barriers in use for websites, apps, and wearables more in-depth insights into the digital divide levels would have been desirable to improve the validation of the findings. Van Dijk [57] provides concepts to further define the digital divide levels. Access, the first level of digital divide, is described to entail the concepts of motivational access and physical access. Use, the second level of digital divide entails the concepts of digital skills and usage. Usage here encompasses both frequency and diversity of digital media use. Perceived benefit, the third level digital divide, is conceptualized by personal outcomes that are a result of the use of digital media [57]. In this study, the concept barriers in use were used instead of digital skills, as experienced barriers are not limited to barriers formed by a lack of digital skills.

Additionally, for the first and the second level digital divide, the technical design and the information and communication technology of eHealth are of importance [57]. The technical design and the information and communication technology imply factors such as accessibility, usability, mobility, quality, and accessibility of internet access and automation (self-learning devices or software tailored to serve the consumer better and automatically) of devices and apps. These factors are important to facilitate adherence and appropriate use of eHealth apps [57]. The questionnaire provided no insight into these factors.

Conclusions

The results of this study revealed that differences in motivation for eHealth use are most profound between different socioeconomic populations in the Dutch society, in which low-educated people are likely to be disadvantaged. A successful transition toward digital health care is a social issue that is dependent on the motivation to use eHealth and specific apps. It is imperative that future studies within academia and within the health care field focus on the motivations and needs associated with digital health care, specifically for low-SEP populations. Research on the societal changes stemming from the digital health care transition and the technical and design studies of digital care apps in single-intervention studies are both vital for the realization of an inclusive and comprehensive digital health care system. If eHealth takes a predominant role in the Dutch health care system, it might affect access and use of health care for the citizens who need it the most. The results of this study underscore the importance of policies aimed at facilitating and supporting low-SEP populations in the use of eHealth to reduce differences in health.

Acknowledgments

The authors would like to thank E de Vries and L van Deursen for providing comments and suggestions. This study was funded by the Dutch Ministry of Health, Welfare, and Sport. The funders had no role in the study design, data collection and analysis, data interpretation, writing of the paper, or approval for publication.

Data Availability

The datasets generated and analyzed during this study are available upon reasonable request from Prof Judith D de Jong (j.dejong@nivel.nl), project leader of the Dutch Health Care Consumer Panel, or the panel's secretary (conusmentenpanel@nivel.nl).

Conflicts of Interest

None to declare.

Multimedia Appendix 1
Operationalization of access, use, and perceived benefit of eHealth in general and websites, apps, and wearables.
[DOCX File , 24 KB - formative_v9i1e49585_app1.docx]

Multimedia Appendix 2
Frequency table of the outcomes of access, use, and perceived benefit for eHealth in general and websites, apps, and wearables gathered via a questionnaire (2021).
[DOCX File , 24 KB - formative_v9i1e49585_app2.docx]

XSL•FO
RenderX

Multimedia Appendix 3
Demographic description of the study population and the study population stratified by levels of education, standardized income, and SES level of the neighborhood.
[DOCX File , 31 KB - formative_v9i1e49585_app3.docx]

Multimedia Appendix 4
Frequency table of males and females aged above 18 in three age categories of the Dutch general population (2021) and the study population (2021).
[DOCX File , 14 KB - formative_v9i1e49585_app4.docx]

Multimedia Appendix 5
Results from logistic (ordered) regression analysis.
[DOCX File , 21 KB - formative_v9i1e49585_app5.docx]

References

1. Rutten P, Pruim J, Zijl NV, Merckelbach S. Digitale zorg in Nederland. Digitale zorg in Nederland. 2020. URL: https://www.mckinsey.com/nl/our-insights/digitale-zorg-in-nederland [accessed 2023-05-05]
2. Abernethy A, Adams L, Barrett M, Bechtel C, Brennan P, Butte A, et al. The promise of digital health: then, now, and the future. NAM Perspect 2022 Jun 27:202206e [FREE Full text] [doi: 10.31478/202206e] [Medline: 36177208]
3. Pagliari C, Sloan D, Gregor P, Sullivan F, Detmer D, Kahan JP, et al. What is eHealth (4): a scoping exercise to map the field. J Med Internet Res 2005 Mar 31;7(1):e9 [FREE Full text] [doi: 10.2196/jmir.7.1.e9] [Medline: 15829481]
4. Dutch digitisation strategy 2.0. Ministery of Economic Affairs and Climate Policy. 2019. URL: https://www.nederlanddigitaal.nl/documenten/publicaties/2019/11/13/english-version-of-the-dutch-digitalisation-strategy-2.0 [accessed 2023-05-05]
5. Digital economy and society index (DESI) 2021: The Netherlands. European Commission. 2021. URL: https://ec.europa.eu/newsroom/dae/redirection/document/80493 [accessed 2022-02-01]
6. 2030 digital compass: the European way for the digital decade. European Commission. 2021. URL: https://eufordigital.eu/library/2030-digital-compass-the-european-way-for-the-digital-decade/ [accessed 2022-02-01]
7. Boers SN, Jongsma KR, Lucivero F, Aardoom J, Büchner FL, de Vries M, et al. SERIES: eHealth in primary care. Part 2: Exploring the ethical implications of its application in primary care practice. Eur J Gen Pract 2019 Oct 30;26(1):26-32 [FREE Full text] [doi: 10.1080/13814788.2019.1678958] [Medline: 31663394]
8. European C. Protecting fundamental rights in the digital age - 2021 annual report on the application of the EU charter of fundamental rights. European Commission. 2021. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_21_6699 [accessed 2023-04-06]
9. Goedhart NS, Zuiderent-Jerak T, Woudstra J, Broerse JEW, Betten AW, Dedding C. Persistent inequitable design and implementation of patient portals for users at the margins. J Am Med Inform Assoc 2021 Feb 15;28(2):276-283 [FREE Full text] [doi: 10.1093/jamia/ocaa273] [Medline: 33463691]
10. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017 Apr 27;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
11. McAuley A. Digital health interventions: widening access or widening inequalities? Public Health 2014 Dec;128(12):1118-1120. [doi: 10.1016/j.puhe.2014.10.008] [Medline: 25458115]
12. Kontos E, Blake KD, Chou WYS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
13. Perez SL, Kravitz RL, Bell RA, Chan MS, Paterniti DA. Characterizing internet health information seeking strategies by socioeconomic status: a mixed methods approach. BMC Med Inform Decis Mak 2016 Aug 09;16:107. [doi: 10.1186/s12911-016-0344-x] [Medline: 27506607]
14. Scherer R, Siddiq F. The relation between students' socioeconomic status and ICT literacy: findings from a meta-analysis. Comput Educ 2019;138:13-32. [doi: 10.1016/j.compedu.2019.04.011]
15. Azzopardi-Muscat N, Sørensen K. Towards an equitable digital public health era: promoting equity through a health literacy perspective. Eur J Public Health 2019;29(Supplement_3):13-17 [FREE Full text] [doi: 10.1093/eurpub/ckz166] [Medline: 31738443]
16. Wong BLH, Maaß L, Vodden A, van Kessel R, Sorbello S, Buttigieg S, European Public Health Association (EUPHA) Digital Health Section. The dawn of digital public health in Europe: implications for public health policy and practice. Lancet Reg Health Eur 2022;14:100316 [FREE Full text] [doi: 10.1016/j.lanepe.2022.100316] [Medline: 35132399]
17. Raad VVES. Een eerlijke kans op gezond leven. Raad voor Volksgezondheid en Samenleving. 2021. URL: https://www.raadrvs.nl/documenten/publicaties/2021/04/07/een-eerlijke-kans-op-gezond-leven [accessed 2021-10-13]

XSL•FO
RenderX

18. de Boer WIJ, Buskens E, Koning RH, Mierau JO. Neighborhood socioeconomic status and health care costs: a population-wide study in the Netherlands. Am J Public Health 2019;109(6):927-933. [doi: 10.2105/AJPH.2019.305035] [Medline: 30998412]
19. Loef B, Meulman I, Herber GM, Kommer GJ, Koopmanschap MA, Kunst AE, et al. Socioeconomic differences in healthcare expenditure and utilization in the Netherlands. BMC Health Serv Res 2021;21(1):643 [FREE Full text] [doi: 10.1186/s12913-021-06694-9] [Medline: 34217287]
20. Petrovic D, de Mestral C, Bochud M, Bartley M, Kivimäki M, Vineis P, et al. The contribution of health behaviors to socioeconomic inequalities in health: a systematic review. Prev Med 2018;113:15-31 [FREE Full text] [doi: 10.1016/j.ypmed.2018.05.003] [Medline: 29752959]
21. Phelan JC, Link BG, Tehranifar P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. J Health Soc Behav 2010;51 Suppl:S28-S40. [doi: 10.1177/0022146510383498] [Medline: 20943581]
22. Agerholm J, Bruce D, Ponce de Leon A, Burström B. Socioeconomic differences in healthcare utilization, with and without adjustment for need: an example from Stockholm, Sweden. Scand J Public Health 2013;41(3):318-325. [doi: 10.1177/1403494812473205] [Medline: 23406653]
23. Lemstra M, Mackenbach J, Neudorf C, Nannapaneni U. High health care utilization and costs associated with lower socio-economic status: results from a linked dataset. Can J Public Health 2009;100(3):180-183 [FREE Full text] [doi: 10.1007/BF03405536] [Medline: 19507718]
24. Mackenbach JP. The persistence of health inequalities in modern welfare states: the explanation of a paradox. Soc Sci Med 2012;75(4):761-769. [doi: 10.1016/j.socscimed.2012.02.031] [Medline: 22475407]
25. Oude Groeniger J, Kamphuis CBM, Mackenbach JP, Beenackers MA, van Lenthe FJ. Are socio-economic inequalities in diet and physical activity a matter of social distinction? A cross-sectional study. Int J Public Health 2019;64(7):1037-1047 [FREE Full text] [doi: 10.1007/s00038-019-01268-3] [Medline: 31187165]
26. Pampel FC, Krueger PM, Denney JT. Socioeconomic disparities in health behaviors. Annu Rev Sociol 2010;36:349-370. [doi: 10.1146/annurev.soc.012809.102529] [Medline: 21909182]
27. Singh-Manoux A, Marmot M. Role of socialization in explaining social inequalities in health. Soc Sci Med 2005;60(9):2129-2133 [FREE Full text] [doi: 10.1016/j.socscimed.2004.08.070] [Medline: 15743660]
28. Weiss D, Rydland HT, Øversveen E, Jensen MR, Solhaug S, Krokstad S. Innovative technologies and social inequalities in health: a scoping review of the literature. PLoS One 2018;13(4):e0195447 [FREE Full text] [doi: 10.1371/journal.pone.0195447] [Medline: 29614114]
29. van der VR, van DL, Standaar LMB, Wouters EJM, Suijkerbuijk AWM, van TL. E-healthmonitor 2021: stand van zaken digitale zorg. 2022. URL: https://www.rivm.nl/documenten/e-healthmonitor-2021-stand-van-zaken-digitale-zorg [accessed 2023-05-05]
30. Western MJ, Armstrong MEG, Islam I, Morgan K, Jones UF, Kelson MJ. The effectiveness of digital interventions for increasing physical activity in individuals of low socioeconomic status: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2021;18(1):148 [FREE Full text] [doi: 10.1186/s12966-021-01218-4] [Medline: 34753490]
31. Hansen AH, Bradway M, Broz J, Claudi T, Henriksen Ø, Wangberg SC, et al. Inequalities in the use of eHealth between socioeconomic groups among patients with type 1 and type 2 diabetes: cross-sectional study. J Med Internet Res 2019;21(5):e13615 [FREE Full text] [doi: 10.2196/13615] [Medline: 31144669]
32. Verhellen A, Jewell C, Van Gils M, Jacobs A, Steenberghs E. Ehealthmonitor Citizens. 2019. URL: https://researchportal.vub.be/en/publications/ehealthmonitor-2019-citizens [accessed 2023-09-13]
33. Vehko T, Ruotsalainen R, Hyppönen H, Aalto A, Terveyden ja hyvinvoinnin laitos. E-health and e-welfare of Finland: check point 2018. 2019. URL: https://oula.finna.fi/Record/oy.9917519693906252?lng=en-gb [accessed 2023-09-13]
34. Huygens MWJ, Voogdt-Pruis HR, Wouters M, Meurs MM, van Lettow B, Kleijweg C, et al. The uptake and use of telemonitoring in chronic care between 2014 and 2019: nationwide survey among patients and health care professionals in the Netherlands. J Med Internet Res 2021;23(5):e24908 [FREE Full text] [doi: 10.2196/24908] [Medline: 33938808]
35. Paccoud I, Baumann M, Le Bihan E, Pétré B, Breinbauer M, Böhme P, et al. Socioeconomic and behavioural factors associated with access to and use of personal health records. BMC Med Inform Decis Mak 2021;21(1):18 [FREE Full text] [doi: 10.1186/s12911-020-01383-9] [Medline: 33435970]
36. Bol N, Helberger N, Weert JCM. Differences in mobile health app use: a source of new digital inequalities? Inf Soc 2018;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]
37. Van Deursen AJM, Helsper EJ. The third-level digital divide: who benefits most from being online. In: Robinson, L, Cotten SR, Schulz J, Hale TM, Williams A, editors. Communication and Information Technologies Annual. Leeds, England: Emerald Group Publishing Limited; 2015:29-52.
38. Chagpar AB. Sociodemographic factors affecting telemedicine access: a population-based analysis. Surgery 2022;171(3):793-798 [FREE Full text] [doi: 10.1016/j.surg.2021.08.059] [Medline: 34848076]
39. Aldhahir AM, Alqahtani JS, Althobiani MA, Alghamdi SM, Alanazi AF, Alnaim N, et al. Current knowledge, satisfaction, and use of E-Health mobile application (Seha) among the general population of Saudi Arabia: a cross-sectional study. J Multidiscip Healthc 2022;15:667-678 [FREE Full text] [doi: 10.2147/JMDH.S355093] [Medline: 35399807]

XSL•FO
RenderX

40. Moon Z, Zuchowski M, Moss-Morris R, Hunter MS, Norton S, Hughes LD. Disparities in access to mobile devices and e-health literacy among breast cancer survivors. Support Care Cancer 2022;30(1):117-126 [FREE Full text] [doi: 10.1007/s00520-021-06407-2] [Medline: 34236506]
41. Melhem SJ, Nabhani-Gebara S, Kayyali R. Digital trends, digital literacy, and E-Health engagement predictors of breast and colorectal cancer survivors: a population-based cross-sectional survey. Int J Environ Res Public Health 2023;20(2):1472 [FREE Full text] [doi: 10.3390/ijerph20021472] [Medline: 36674237]
42. Özdemir R, Bektemur G, Keles E, Baydili KN. Internet use, e-Heath literacy, and associated factors in Istanbul, Turkey: a cross-sectional study. J Consumer Health Internet 2023;27(1):1-11 [FREE Full text] [doi: 10.1080/15398285.2022.2129178]
43. Brabers AEM, de JJD. Nivel Consumentenpanel Gezondheidszorg Basisrapport met informatie over het panel 2022. NIVEL. 2022. URL: https://www.nivel.nl/nl/publicatie/nivel-consumentenpanel-gezondheidszorg-basisrapport-met-informatie-over-het-panel-2022 [accessed 2023-04-06]
44. Your research: is it subject to the WMO or not? Central Committee on Research Involving Human Subjects. URL: https://english.ccmo.nl/investigators/legal-framework-for-medical-scientific-research/your-research-is-it-subject-to-the-wmo-or-not [accessed 2024-06-13]
45. Giddens A. The Constitution of Society: Outline of the Theory of Structuration. Berkeley: University of California Press; 1984.
46. Bourdieu P. Outline of a Theory of Practice. Cambridge: Cambridge University Press; 1977.
47. Krieger N, Williams DR, Moss NE. Measuring social class in US public health research: concepts, methodologies, and guidelines. Annu Rev Public Health 1997;18:341-378. [doi: 10.1146/annurev.publhealth.18.1.341] [Medline: 9143723]
48. Scheerder A, van Deursen A, van Dijk J. Determinants of internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. Telematics Inform 2017;34(8):1607-1624. [doi: 10.1016/j.tele.2017.07.007]
49. Vrooman J, Boelhouwer J, Iedema J, van der T. Summary contemporary inequality. The Netherlands Institute for Social Research. 2023. URL: https://english.scp.nl/publications/publications/2023/03/07/summary-contemporary-inequality [accessed 2023-05-09]
50. Opleidingsniveau. Statistics Netherlands. 2019. URL: https://www.cbs.nl/nl-nl/nieuws/2019/33/verschil-levensverwachting-hoog-en-laagopgeleid-groeit/opleidingsniveau [accessed 2022-01-27]
51. What is my spendable income? Statistics Netherlands. 2018. URL: https://www.cbs.nl/en-gb/background/2008/50/what-is-my-spendable-income- [accessed 2023-04-06]
52. Geboorte; kerncijfers. StatLine. 2022. URL: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37422ned/table [accessed 2022-02-20]
53. Ruim 140 duizend samenwonende volwassen kinderen met ouders op leeftijd. Centraal Bureau van de Statistiek. 2014. URL: https://www.cbs.nl/nl-nl/achtergrond/2014/27/ruim-140-duizend-samenwonende-volwassen-kinderen-met-ouders-op-leeftijd [accessed 2022-08-02]
54. Income distribution (standardised income). Centraal Bureau van de Statistiek. 2019. URL: https://www.cbs.nl/en-gb/visualisations/income-distribution [accessed 2021-08-04]
55. Statusscore per wijk en buurt o.b.v. welvaart, opleidingsniveau en arbeid. Centraal Bureau van de Statistiek. 2022. URL: https://www.cbs.nl/nl-nl/achtergrond/2022/14/statusscore-per-wijk-en-buurt-o-b-v-welvaart-opleidingsniveau-en-arbeid [accessed 2022-08-01]
56. Sociaal-economische status; scores per wijk en buurt, regio-indeling 2021. Centraal Bureau van de Statistiek. 2022. URL: https://www.cbs.nl/nl-nl/cijfers/detail/85163NED [accessed 2022-08-01]
57. Van Dijk J. The Digital Divide. Cambridge: Polity Press; 2020.
58. Elena-Bucea A, Cruz-Jesus F, Oliveira T, Coelho PS. Assessing the role of age, education, gender and income on the digital divide: evidence for the European union. Inf Syst Front 2020;23(4):1007-1021. [doi: 10.1007/s10796-020-10012-9]
59. Reiners F, Sturm J, Bouw LJW, Wouters EJM. Sociodemographic factors influencing the use of eHealth in people with chronic diseases. Int J Environ Res Public Health 2019;16(4):645 [FREE Full text] [doi: 10.3390/ijerph16040645] [Medline: 30795623]
60. Stata statistical software: release 16. StataCorp. 2019. URL: https://www.stata-press.com/data/r16/ [accessed 2024-06-20]
61. Bevolking op 1 januari en gemiddeld; geslacht, leeftijd en regio. Statistics Netherlands. 2021. URL: https://www.cbs.nl/nl-nl/cijfers/detail/03759ned?dl=39E0B [accessed 2024-05-15]
62. Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med 2018;18(3):91-93 [FREE Full text] [doi: 10.1016/j.tjem.2018.08.001] [Medline: 30191186]
63. Øversveen E, Rydland HT, Bambra C, Eikemo TA. Rethinking the relationship between socio-economic status and health: making the case for sociological theory in health inequality research. Scand J Public Health 2017;45(2):103-112. [doi: 10.1177/1403494816686711] [Medline: 28078944]
64. Eikemo TA, Øversveen E. Social inequalities in health: challenges, knowledge gaps, key debates and the need for new data. Scand J Public Health 2019;47(6):593-597. [doi: 10.1177/1403494819866416] [Medline: 31512565]
65. Stimuleren gebruik e-health. Ministry of Public Health, Wellbeing and Sports. URL: https://www.rijksoverheid.nl/onderwerpen/e-health/overheid-stimuleert-e-health [accessed 2023-09-08]

XSL•FO
RenderX

66. Hiltz SR, Turoff M. Education goes digital: the evolution of online learning and the revolution in higher education. Commun ACM 2005;48(10):59-64. [doi: 10.1145/1089107.1089139]
67. Grossman LV, Masterson Creber RM, Benda NC, Wright D, Vawdrey DK, Ancker JS. Interventions to increase patient portal use in vulnerable populations: a systematic review. J Am Med Inform Assoc 2019;26(8-9):855-870 [FREE Full text] [doi: 10.1093/jamia/ocz023] [Medline: 30958532]
68. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
69. Bejaković P, Mrnjavac Z. The importance of digital literacy on the labour market. ER 2020;42(4):921-932. [doi: 10.1108/er-07-2019-0274]
70. Vasilescu MD, Serban AC, Dimian GC, Aceleanu MI, Picatoste X. Digital divide, skills and perceptions on digitalisation in the European union-towards a smart labour market. PLoS One 2020;15(4):e0232032 [FREE Full text] [doi: 10.1371/journal.pone.0232032] [Medline: 32324786]
71. Wagenaar KP, Hakim N, Broekhuizen BDL, Jaarsma T, Rutten FH, Hoes AW. Representativeness of participants in heart failure E-health trials: a report from the E-vita HF study. J Card Fail 2017;23(1):88-89 [FREE Full text] [doi: 10.1016/j.cardfail.2016.08.006] [Medline: 27565046]
72. Al-Dhahir I, Reijnders T, Faber JS, van den Berg-Emons RJ, Janssen VR, Kraaijenhagen RA, et al. The barriers and facilitators of eHealth-based lifestyle intervention programs for people with a low socioeconomic status: scoping review. J Med Internet Res 2022;24(8):e34229 [FREE Full text] [doi: 10.2196/34229] [Medline: 36001380]
73. de Boer WIJ, Dekker LH, Koning RH, Navis GJ, Mierau JO. How are lifestyle factors associated with socioeconomic differences in health care costs? Evidence from full population data in the Netherlands. Prev Med 2020;130:105929 [FREE Full text] [doi: 10.1016/j.ypmed.2019.105929] [Medline: 31778685]
74. Agachi E, Bijmolt THA, Mierau JO, van Ittersum K. Adoption of the website and mobile app of a preventive health program across neighborhoods with different socioeconomic conditions in the Netherlands: longitudinal study. JMIR Hum Factors 2022;9(1):e32112 [FREE Full text] [doi: 10.2196/32112] [Medline: 35107433]

Abbreviations

DHCCP: Dutch Health Care Consumer Panel
OR: odds ratio
SEP: socioeconomic position
SES-WOA: Social Economic Status-Wealth Education Employment
SES: socioeconomic status

Edited by A Mavragani; submitted 03.06.23; peer-reviewed by J Mierau, H Traino, PhD, MPH, M Earl; comments to author 29.08.23; revised version received 02.10.23; accepted 21.07.24; published 07.01.25.

Please cite as:
Standaar L, van Tuyl L, Suijkerbuijk A, Brabers A, Friele R
Differences in eHealth Access, Use, and Perceived Benefit Between Different Socioeconomic Groups in the Dutch Context: Secondary Cross-Sectional Study
JMIR Form Res 2025;9:e49585
URL: https://formative.jmir.org/2025/1/e49585
doi:10.2196/49585
PMID:

©Lucille Standaar, Lilian van Tuyl, Anita Suijkerbuijk, Anne Brabers, Roland Friele. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Trifecta of Industry, Academic, and Health System Partnership to Improve Mental Health Care Through Smartphone-Based Remote Patient Monitoring: Development and Usability Study

C Neill Epperson[1], MD; Rachel Davis[1], MD; Allison Dempsey[1], PhD; Heinrich C Haller[1], BS; David J Kupfer[2], MD; Tiffany Love[1], PhD; Pamela M Villarreal[3], MS; Mark Matthews[4], PhD; Susan L Moore[5], PhD; Kimberly Muller[5], MSc, JD; Christopher D Schneck[1], MD; Jessica L Scott[1], MA, MPH; Richard D Zane[3,6], MD; Ellen Frank[2,4], PhD

[1]Department of Psychiatry, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

[2]Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[3]UCHealth CARE Innovations Center, Anschutz Medical Campus, University of Colorado Health System, Aurora, CO, United States

[4]Health Rhythms, New York, NY, United States

[5]CU Innovations, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

[6]Department of Emergency Medicine, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

Corresponding Author:
C Neill Epperson, MD
Department of Psychiatry
School of Medicine
University of Colorado Anschutz Medical Campus
1890 N Revere Ct
Suite 4003, Mail Stop F546
Aurora, CO, 80045
United States
Phone: 1 303 724 4940
Email: neill.epperson@cuanschutz.edu

Abstract

Background: Mental health treatment is hindered by the limited number of mental health care providers and the infrequency of care. Digital mental health technology can help supplement treatment by remotely monitoring patient symptoms and predicting mental health crises in between clinical visits. However, the feasibility of digital mental health technologies has not yet been sufficiently explored. Rhythms, from the company Health Rhythms, is a smartphone platform that uses passively acquired smartphone data with artificial intelligence and predictive analytics to alert patients and providers to an emerging mental health crisis.

Objective: The objective of this study was to test the feasibility and acceptability of Rhythms among patients attending an academic psychiatric outpatient clinic.

Methods: Our group embedded Rhythms into the electronic health record of a large health system. Patients with a diagnosis of major depressive disorder, bipolar disorder, or other mood disorder were contacted online and enrolled for a 6-week trial of Rhythms. Participants provided data by completing electronic surveys as well as by active and passive use of Rhythms. Emergent and urgent alerts were monitored and managed according to passively collected data and patient self-ratings. A purposively sampled group of participants also participated in qualitative interviews about their experience with Rhythms at the end of the study.

Results: Of the 104 participants, 89 (85.6%) completed 6 weeks of monitoring. The majority of the participants were women (72/104, 69.2%), White (84/104, 80.8%), and non-Hispanic (100/104, 96.2%) and had a diagnosis of major depressive disorder (71/104, 68.3%). Two emergent alerts and 19 urgent alerts were received and managed according to protocol over 16 weeks. More than two-thirds (63/87, 72%) of those participating continued to use Rhythms after study completion. Comments from participants indicated appreciation for greater self-awareness and provider connection, while providers reported that Rhythms provided a more nuanced understanding of patient experience between clinical visits.

XSL•FO
RenderX

Conclusions: Rhythms is a user-friendly, electronic health record–adaptable, smartphone-based tool that provides patients and providers with a greater understanding of patient mental health status. Integration of Rhythms into health systems has the potential to facilitate mental health care and improve the experience of both patients and providers.

(JMIR Form Res 2025;9:e57624) doi:10.2196/57624

KEYWORDS

digital health; mobile intervention; telepsychiatry; artificial intelligence; psychiatry; mental health; depression; mood; bipolar; monitor; diagnostic tool; diagnosis; electronic health record; EHR; alert; notification; prediction; mHealth; mobile health; smartphone; passive; self-reported; patient generated

Introduction

Mental disorders are predicted to be the leading cause of disease burden worldwide by 2030 [1]. However, there is a tremendous mismatch between the number of psychiatrists, psychologists, and other mental health care providers and individuals in need of treatment for mental illness, including substance use disorders. In the United States alone, 160 million people reside where there is a shortage of mental health professionals [2]. Roughly 56% of counties are without a practicing psychiatrist, while 33% of counties are without a licensed psychologist [3]. The workforce is expected to be short 38,821 psychiatrists by 2024 [4]. In the United States, the majority of individuals seeking treatment for depression, anxiety, or any mental illness receive care from their primary care physician (PCPs), and PCPs provide almost a third of the care for serious mental illness [5].

On average, a PCP addresses between 3 to 5 problems, including psychiatric problems, during a typical encounter [6,7]. Furthermore, the majority of PCPs cannot see patients as frequently as may be necessary to adequately monitor a patients' mental illness and response to treatment, nor do they have a mental health specialist to whom they can refer their more concerning patients [8]. In the context of the current mental health landscape, remote monitoring of patient mental well-being between visits is poised to play a major role in the redesigning of mental health care, similar to what is occurring with remote management of myriad other conditions [9,10].

Ideally, remotely acquired data would be integrated into the electronic health record (EHR) and readily available to the patient's provider for interpretation and intervention if needed. With remote patient monitoring, the documentation of history since the previous appointment is quicker, the challenges of patient recall with respect to their signs and symptoms are obviated, and there is a foundation for patients to participate to a greater degree in their own psychiatric care by tracking their symptoms alongside their provider.

Clinical, patient-generated, app-acquired health data that are embedded and presented into the EHR workflow provide the opportunity for interprofessional use. This is particularly important when patients have multiple comorbid medical conditions that impact brain health and vice versa. Furthermore, patient engagement with health apps, including those that are mental health related, declines dramatically over the first few weeks of use, in some cases dropping below 10% within 2 weeks [11,12]. An ideal app may rely less upon active user engagement for data collection and to a greater degree on passively acquired data that reduce patient burden, which then trigger the completion of symptom self-reports using nationally accepted standardized ratings when passively acquired patient-generated health data (PGHD) indicate a relapse is about to occur.

To address these deficits in the digital mental health space, an academic-industry collaborative team was formed to test the feasibility and acceptability of industry start-up Health Rhythms' smartphone solution called Rhythms for use across a large health system. Previous research has demonstrated that Rhythms' proprietary artificial intelligence (AI) algorithm applied to passively collected data indicative of behavioral vital signs (eg, sleep patterns, activity measures, and sociability) can predict a relapse of depression 7 days before the patient's self-report of relapse with 90% accuracy [13-15]. This collaborative team reports the results of the EHR integration, results of the feasibility testing of the urgent and emergent alert algorithms, effectiveness of the 24/7 virtual patient monitoring system, user acceptability in terms of both patient use of and satisfaction with Rhythms, and provider impressions of the impact of Rhythms on patient care.

Methods

Collaborative Team

Health Rhythms is a technology company that was founded by experienced researchers and clinicians (EF and DK) in the assessment, treatment, and research of mental illness, particularly bipolar disorder and depression, in partnership with international leaders (MM) in behavioral sensing technology focused on mental health. The University of Colorado (CU) Anschutz Medical Campus Department of Psychiatry is a large department (500+ faculty) with a strong commitment to the incorporation of digital health and technology in the care of patients with serious mental illness. CU Innovations brings together industry partners, entrepreneurs, and investors to help CU researchers create biomedical technology that improves care delivery, health outcomes, and patient quality of life worldwide. University of Colorado Health (UCHealth) is a not-for-profit health care system across Colorado, southern Wyoming, and western Nebraska with an academic medical center, the University of Colorado Hospital (UCH), that is affiliated with the CU School of Medicine. The CARE Innovation Center at UCHealth partners with CU to create a comprehensive platform of resources both to conduct clinical and biomedical research and to transform real-world clinical practice together with industry partners or faculty entrepreneurs.

XSL•FO
RenderX

App Development

The Rhythms app was developed using a participatory design process [16], an approach that seeks to deeply involve end users in the design process. A weekly meeting between the Health Rhythms team and the UCHealth clinical team was the primary means to engage in this process. It focused on understanding end-user needs, existing clinical workflows, and seeking feedback on early prototypes.

A key decision resulting from this process was to integrate the Health Rhythms software development kit (SDK) into UCHealth's MyChart mobile app, rather than providing a distinct separate app that participants would have to download. Health Rhythms' SDK is a modular piece of software that can be integrated easily into native iOS or Android apps. Although this decision required significant technical work, the team anticipated that it would significantly streamline the patient end-user experience, since they would only have to install and use 1 app, and would help maintain user privacy and security.

Once integrated into the UCHealth MyChart app and end users have provided explicit permissions during onboarding, the SDK collects data continuously in the background from the activity, location, and display smartphone sensors. These data can then be used to automatically assess behavior patterns derived from these data streams including how much time a person is spending at home, levels of physical activity, estimates of sleep timing, and duration and interaction patterns with the smartphone itself.

Safeguarding patient privacy was paramount. Users were uniquely identified using a nonidentifiable identifier that was passed to Health Rhythms. The Rhythms app did not collect any identifiable data that falls under the protections of the Health Insurance Portability and Accountability Act (HIPAA) such as birth dates, home addresses, or social security numbers, and the SDK did not collect any data from texts, calls, photos, internet browsing, or other content. All data, including GPS location, are encrypted using state of the art encryption (ie, a 256-bit key) on the device, in transmission and in the cloud. The sensor data for each patient were computed into daily nonidentifiable behavioral summaries that were then transmitted into the EHR using standard Fast Healthcare Interoperability Resources end points. No unprocessed sensor data, including GPS, were transferred into the EHR.

While there is always a risk of data breach in such systems, the following measures were implemented to mitigate this risk. The team applied the principle of data minimization, reducing the collection, storage, and use of personal data to what is strictly necessary to assess individual behavioral health. Patient end users had the agency to rescind permissions for sensor data collection at any time by a "Manage Settings" screen, and transparent, easy-to-understand messaging during onboarding communicated clearly what data were collected and how they would be used. Finally, Health Rhythms maintained General Data Protection Regulation and California Consumer Privacy Act certification and conducted regular security tests by professional third parties.

AI Methods

The Rhythms system collects passive sensing data, comprising location, pedometer, activity, and device data, from the onboard sensors in smartphones. Each day, these data are processed into 65 behavioral inferences for each person. These inferences encompass measures related to physical and social activity, stability or instability of the person's weekly routine, sleep, and smartphone use. For example, time at home is an inference derived from location data about how much time a person is spending at home.

These behavioral inferences are used in 2 ways. First, heuristics related to these inferences are used to trigger a request for a self-report from participants. A total of 15 static heuristics, empirically derived from previous datasets, dynamically prompted individuals to complete self-reported assessments. For example, if an individual had a diagnosis of major depression and spent a long time at home for over 7 days, then a self-report was triggered. Second, the behavioral inferences were fed into a logistic regression machine learning model that provided a continuous and passive estimate of a person's risk for a relapse in depression, defined as scoring greater than 10 on the Patient Health Questionnaire 9-Item (PHQ-9). This model aims to be predictive of how a given person would answer specific questions on the PHQ-9. Model outputs with a high probability for a high PHQ-9 score were used to flag individuals as a medium alert on the dashboard if there was no corroborating self-report.

Participant Criteria and Recruitment

Potential participants were residents of Colorado; aged between 18 and 89 years; were the primary user of a mobile phone that supported iOS12 or above; enrolled in MyHealthConnection (MHC), which is UCH's EHR patient portal; and were willing to enroll their device into Rhythms.

The UCH's EHR (EPIC) was used to search the caseload of participating clinicians in the CU Department of Psychiatry to identify patients who carry the diagnosis of depressive disorder including major depressive disorder (MDD), dysthymia, depressive disorder due to another medical condition, or bipolar disorder. Providers reviewed the list to confirm that individuals were current patients and would be clinically appropriate for contact about Rhythms.

Recruitment was conducted virtually. Patients were contacted initially through MHC with a general message describing the study. Due to poor initial response to the MHC outreach, a personalized email letter was sent from their provider. Potential participants were contacted up to 3 times by a study coordinator, and screening appointments were scheduled if they expressed interest in the study. After consenting and screening, the study coordinator reviewed how to download the app and interact with the platform and placed an order for Rhythms in EPIC. Participants were considered lost to follow-up if they did not respond to 3 contacts by the study team after enrollment.

Throughout the study, participants provided data by completing electronic surveys sent by the study team (baseline, end of study [6 weeks], and when triggered by algorithm), as well as by active and passive use of the Rhythms platform. A purposively

XSL•FO
RenderX

sampled [17,18] group of participants participated in qualitative interviews with a member of the study team at the end of the study. An overview of the study design and procedures is presented in Figure 1.

Figure 1. Study flow chart: (A) potential participants identified using the electronic health record in collaboration with providers; (B) potential participants contacted up to three times; (C) participants (who met criteria and signed consent) downloaded Rhythms and were instructed in its use; (D) participants completed baseline assessments; (E) participants continued their daily activities, with Rhythms passively collecting data for 6 weeks; (F) participants completed baseline assessment again at end of study; (G) participants completed a survey of likes and dislikes with respect to Rhythms; and (H) selected group underwent qualitative interviews about Rhythms use.

Data Collected by Rhythms

Passively Measured Outcomes

Rhythms acquired proxy information regarding sleep quality, activity level, and sociability. The AI algorithm uses passively collected data to detect deviations from the individual's baseline and triggers a push to complete the self-report ratings described below. During the course of this study, the participant could not view their own passively acquired data.

Self-Report Measures

Per study protocol, self-report measures were administered at baseline and end of study. In addition, self-report measures were pushed by Rhythms to the participant during the study period when the AI algorithm detected a deviation in passively collected data indicative of worsening of symptoms of depression, anxiety, or mania. Participant self-ratings included the following:

PHQ-9 Instrument

The PHQ-9 asks the individual to rate the severity of symptoms over the past 2 weeks on a 4-point scale ranging from "not at all" to "nearly every day." Endorsement of either depressed mood or anhedonia (PHQ-2) triggers further evaluation with the PHQ-9, which assesses for the presence of other criteria for MDD. In multiple studies, PHQ-9 scores >10 had a sensitivity of 88% and a specificity of 88% for MDD [19-21].

Columbia-Suicide Symptom Severity Rating Scale

The Columbia-Suicide Symptom Severity Rating Scale (C-SSRS) is used to assess severity and immediacy of suicide risk using four constructs: (1) severity of suicidal ideation, (2) intensity of ideation, (3) behavior, and (4) lethality. The instrument has been reported to have a 67% sensitivity and 76% specificity for identifying suicidal behaviors [22,23].

Generalized Anxiety Disorder 7-Item

The Generalized Anxiety Disorder 7-Item (GAD-7) is used to measure the severity of generalized anxiety disorder symptoms by asking individuals to rate the severity of symptoms on a 4-point scale with ranging from "not at all" to "nearly every day." Scores of 5, 10, and 15 represent cutoff points for mild, moderate, and severe anxiety, respectively [24].

Altman Self-Rating Mania Scale

The Altman Self-Rating Mania Scale (ASRM) is a 5-item, self-reported scale to assess positive mood, self-confidence, sleep patterns, speech patterns and amount, and motor activity, which are rated on a scale of "0" or "normal" to "4" or "overtly manic." Total scores range from 0 to 20, with scores equal to or higher than 6 indicating a higher likelihood of manic or hypomanic symptoms [25].

Participant-Reported Measures Collected by Study Team

Postintervention Feasibility and Satisfaction Survey

At study end, participants completed a survey about what they liked or disliked most about Rhythms, their interest in continued use, challenges or barriers to use, and any concerns that they may have had regarding alerting and monitoring through the use of Rhythms.

Qualitative Interviews

The qualitative in-depth interviews were conducted by Zoom (Zoom Video Communications) using a semistructured interview guide to ensure key topics were discussed while also allowing for emergence and exploration of topics important to the participants. The main interview questions for patients focused on their perceived benefits from and drawbacks of the Rhythms' system and suggestions for improving Rhythms' features and use.

Providers also completed a feedback survey at study end. Main questions for providers included their experience integrating the Rhythms system into their workflow, how the system was perceived to impact the quality of care, and the efficiency of providing personalized treatment for patients.

Determining Alert Structure

Based upon previous research with Rhythms [14,15,17,26] and consensus of subject matter experts in the treatment of MDD, anxiety, and bipolar disorder, Rhythms was programmed to send urgent or emergent alerts to UCHealth's Virtual Behavioral Health Center (VBHC) or the Department of Psychiatry's care coordinator based upon passively acquired data and participant self-ratings (Textbox 1).

XSL•FO
RenderX

Textbox 1. Urgent and emergent alerts.

Urgent alert

- Physician Health Questionnaire-9 item score is 20-27, when the previous score was lower.
- Columbia-Suicide Severity Rating Scale question 3=yes.
- Altman Self-Rating for Mania Scale score is ≥11.
- Altman Self-Rating for Mania Scale score is <11 and score for question 3=3 or 4.
- If bipolar disorder diagnosis and Rhythms detect a 40% decrease in sleep.
- If bipolar disorder diagnosis and Rhythms monitoring shows no sleep in 2 days.
- Generalized Anxiety Disorder-7 item score ranges from 15-21, when the previous score was not in this range.

Emergent alert

- Columbia-Suicide Severity Rating Scale questions 4, 5, or 6=yes.
- Altman Self-Rating for Mania Scale total score >11 and answer to question 3=4.

The following specific self-rating questions pertained to the alert structure: C-SSRS question 3: Active suicidal ideation with any methods (not plan) without intent to act; C-SSRS question 4: Active suicidal ideation with some intent to act, without specific plan; C-SSRS question 5: Active suicidal ideation with specific plan and intent; C-SSRS question 6: Have you ever done anything, started to do anything, or prepared to do anything to end your life?; and ASRM question 3: Regarding Sleep Patterns: score of 3="I frequently need less sleep than usual." and score of 4="I can go all day and night without any sleep and not feel tired."

Per protocol, a care coordinator from the VBHC or the Department of Psychiatry called the participant up to 3 times within 24 hours in response to an *urgent* alert. The VBHC monitor covered alerts over weekends, holidays, and after hours, while the care coordinator supporting the participant's primary mental health care provider covered during regular work hours. When the participant was reached by the VBHC, a brief safety assessment was conducted; participants were offered the chance to speak to an on-call clinician for the VBHC and referred to their current provider if appropriate. In the case of clinic care coordinators, the participant was assessed with respect to safety, possible medication changes, and need for earlier appointment with their provider.

In response to *emergent* alerts, Health Rhythms sent a page to the VBHC monitor. The pager beeped every 5 minutes until the alert was acknowledged by the monitor in the EHR. All emergent alerts were managed by the VBHC monitor with the expectation that a call to the patient would be initiated within 15 minutes of the alert. Upon reaching the patient, a safety assessment was conducted, and an appropriate disposition was managed. If after 3 attempts there was no contact with the patient, the VBHC monitor called emergency services to conduct a welfare check at the address listed for the participant.

Initially, participants received a message each time they completed self-report measures that they could receive a call from the VBHC based upon how they answered the questionnaires. This caused confusion for participants and the process was changed such that only participants who had triggered an emergent or urgent alert would get the message that they would receive a call.

Data Analysis

Demographics and baseline behavioral rating scores as well as number and nature of alerts were analyzed descriptively with means (SDs) and frequencies (percentages). Postintervention survey data were summarized descriptively for fixed-choice items. Open-ended survey items and interview transcripts were analyzed qualitatively. Human-adjudicated transcription of interview recordings was performed with Otter.ai, a proprietary AI-powered note taking engine, to generate initial transcripts, after which a member of the study team reviewed the transcripts while simultaneously listening to the recordings to ensure there were no errors in transcription.

A rapid qualitative analysis approach was used for content analysis of open-ended survey and interview transcript data. We first deductively explored and identified common themes that appeared in participants' responses. Two members of the research team separately reviewed the data and identified keywords and topics that are relevant to our research questions (eg, most liked and disliked features of the Rhythms' system, facilitators and barriers to using the Rhythms system, and suggestions to improve the system). During this process, each reviewer independently summarized how frequently common items were reported by study participants. The reviewers compared their results, discussed and resolved any discrepancies, reached consensus on the latent themes that had emerged in analysis, and generated a summary.

Ethical Considerations

This project was approved by the University of Colorado Multiple Institution Review Board (COMIRB; 21-4903), and participants gave electronic informed consent. Participants received US $25 in compensation for completing the screening, enrollment, and baseline surveys, and an additional US $50 for completing the postintervention surveys as described below. All data were deidentified for analysis and presented anonymously.

XSL•FO
RenderX

Results

Participants

Of 394 patients who were informed about the study through MHC or email, 178 (45.2%) did not respond, 26 (6.7%) were contacted but lost to follow-up before enrollment, 86 (21.8%) were screened but did not enroll, and 104 (26.4%) were enrolled within 8 weeks of study initiation. The most common reason (60/86, 69.8%) for not meeting study criteria was having a phone using the Android operating system.

The study sample comprised 104 individuals; the majority identified as women (n=72 69.2%), non-Hispanic (n=100 96.2%), and White (n=84 80.8%). The average age was 42.1 (SD 15.9) years. The majority of participants had MDD (n=71 68.3%), while 23 (22.2%) individuals had either bipolar I or II disorder and 10 individuals (9.6%) had “other mood disorder” (Table 1).

Between enrollment and study completion, 17 (16.3%) of the 104 enrolled participants were lost to follow-up, withdrew from the study, or did not complete the final survey. Of note, among the 87 (83.7%) completers out of 104, a total of 72% (63/87) continued using Rhythms despite no longer being compensated for doing so.

Table 1. Participant characteristics and baseline behavioral ratings.

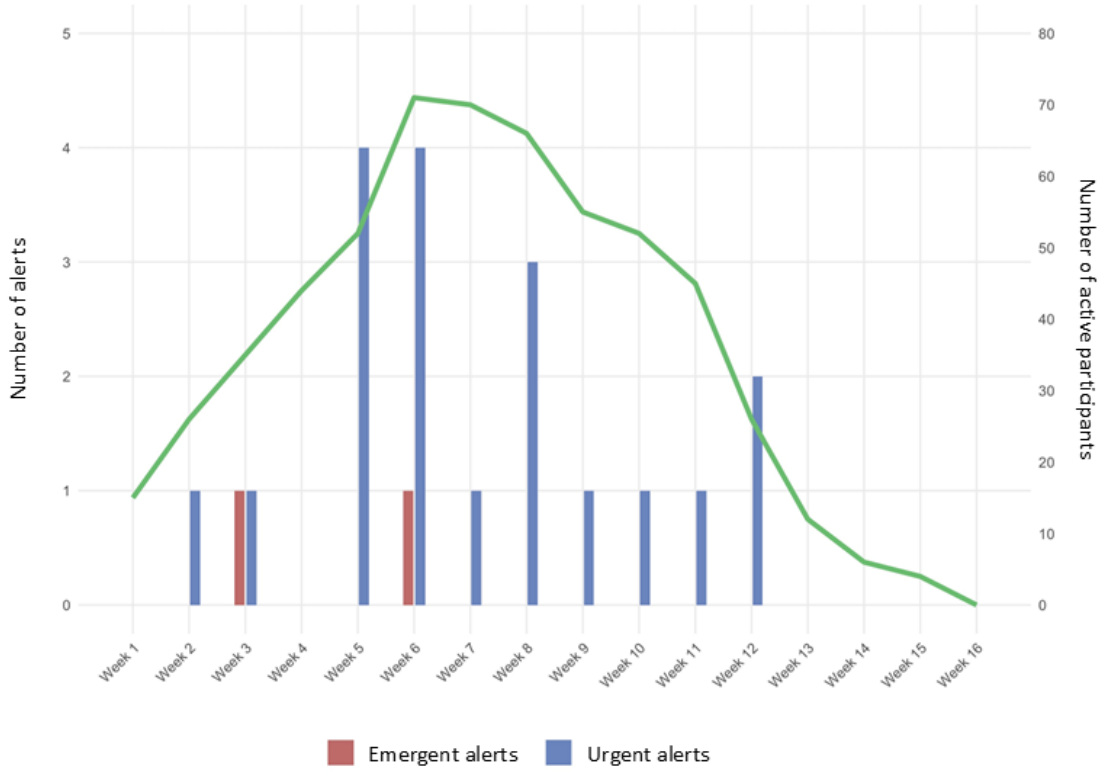
Variable	Value (n=104)
Age (years), mean (SD)	42.1 (15.9)
Gender, n (%)	
Women	72 (69.2)
Men	29 (27.9)
Transgender or nonbinary	3 (2.9)
Ethnicity, n (%)	
Hispanic	3 (2.9)
Non-Hispanic	100 (96.2)
Not answered	1 (1)
Race, n (%)	
Asian	9 (8.6)
Black or African American	1 (1)
American Indian or Alaska Native	1 (1)
White	84 (80.8)
More than one race	4 (3.8)
Other or prefer not to answer	5 (4.8)
Diagnosis, n (%)	
Major depressive disorder	71 (68.3)
Bipolar 1 disorder	12 (11.5)
Bipolar II disorder	11 (10.6)
Other mood disorder	10 (9.6)
Ratings, mean (SD)	
Physician Health Questionnaire 9-Item	6.5 (5.3)
Generalized Anxiety Disorder 7-Item	5.9 (4.8)
Altman Self-Rating Mania Scale	2.2 (2.7)

Urgent and Emergent Alerts

Over the course of the entire 16-week study period, the monitoring of 104 individuals with Rhythms for 6 weeks each resulted in 2 emergent alerts and 19 urgent alerts. The largest number of participants being monitored at 1 time was 92 and the average was 64. The greatest number of alerts on any given day was 3 (all urgent, 1 day only) when there were 58 participants being monitored (Figure 2).

XSL•FO
RenderX

Figure 2. Urgent and emergent alerts across time and by number of active participants.



The first emergent alert was triggered by a participant answering ASRM question 3 about sleep with a "4," indicating "I can go all day and night without any sleep and not feel tired." The second emergent alert was triggered by the participant answering yes to C-SSRS question 6, indicating that they "had ever done anything, started to do anything or prepared to do anything to end their life." The primary (n=15) trigger for urgent alerts was answering "yes" to question 3 on the C-SSRS, indicating that the participant had active suicidal ideation without method or intent to act. The second most common (n=12) reason for urgent alerts was a GAD-7 score indicating severe anxiety. Three urgent alerts occurred because of the ASRM indicating mania or hypomania based upon total ASRM or response to question 3 regarding need for sleep.

All calls to participants were managed within the specified alert time. Eleven (58%) of the 19 urgent alerts came through during work hours and were managed by the Department of Psychiatry's care coordinator. The 2 emergent alerts were managed by the VBHC, neither of which required an emergency room referral or in-person well-being check.

Postintervention Survey: Patients and Providers

Of the 87 participants who completed the postintervention feasibility and satisfaction survey, the median score for ease of starting the system, ease of use, and surveys being delivered at convenient times were 9, 10, and 9, respectively, on a scale of 0 (strongly disagree) to 10 (strongly agree). Median score for participant satisfaction with the Rhythms system within their MHC app was 8.

Themes from the free text indicated that respondents most liked that Rhythms gave them a greater awareness of their own mental health status, the feeling of being taken care of by their provider, and its ease of use. The most common complaint about Rhythms (mentioned by n=18, 21%) was having to keep the app open. Only 3 (3%) reported concerns about being "tracked."

When providers were surveyed about their experience with Rhythms, they noted that the platform provided them with information they would not otherwise have had and information that was directly actionable. For example, 1 provider commented that "...the sleep section was the most useful for me to review. I am very surprised that one of my patients is only averaging about 5 hours of sleep per night, so that is definitely something we will talk about." Another commented that "The more frequent assessment was great. I don't think patients remember their condition for more than a few days. I was able to reach out to a couple of patients based on worsening in [their] data."

Qualitative Interviews: Patients

Qualitative interviews with 19 (22%) of the 87 completers revealed that all but 2 were well established in the CU and UCH system for their mental health care, having received care there for at least 3 years. Most agreed to participate in the study because they felt positively about research (n=10, 11%), thought it would be easy (n=4, 5%), or thought it would be helpful to either themselves (n=3, 3%) or others (n=3, 3%). Setup and initial use of Rhythms was reported by 13 (68%) of those interviewed as being "simple, intuitive, and unobtrusive," although 5 (26%) reported some technical challenges, including difficulty with changing settings, granting permissions, or finding specific features. With respect to completing the self-report measures and their frequency, 63% (n=12) of respondents reported completing them at the time of the prompt. In general, respondents were okay with the frequency of measures (n=3, 3%) or agreed that weekly (n=3, 3%) or every 2 weeks (n=4, 5%) would be acceptable, although a few (n=2,

XSL•FO
RenderX

2%) preferred monthly or felt surveys were sent too often (n=1, 1%).

Overall, interview participants reported positive perceptions of Rhythms, finding it easy to use (n=8, 9%) and enjoying its use (n=5, 6%). With respect to what was liked most about Rhythms, there was consensus that the platform promoted greater provider connection (n=4, 5%) and provider awareness of patients' health status (n=6, 7%), contributing to a feeling of being "looked out for" (n=4, 5%). Participants also reported improved self-awareness of one's own mental health status (n=5, 6%). What was liked least about Rhythms was having to keep the app open (n=7, 8%), with related concern for battery life (n=4, 5%); the wording of a response message to completed surveys notifying patients that they might be contacted based on their responses (n=6, 7%); concerns about what would happen if answers to questions were negative (n=3, 3%); not receiving immediate feedback from Rhythms; and accuracy of the data if the person did not have their phone with them or were traveling outside of Colorado.

Roughly half of those interviewed endorsed having had a helpful discussion about their Rhythms data with their provider, while others indicated that their provider did not mention their data during their appointment, or they did not have an appointment during the 6 weeks of tracking. The most common recommendation for system improvement was the ability to access and track their own data in a visual format, an improvement that is now available to users. Illustrative quotations about positive aspects of Rhythms were as follows:

> *…I had one two-week period where it was like, everything's catastrophic…which solicited the call from [my provider], through the patient portal. "So I saw your responses. What's, the scoop?"…I liked the fact that somebody was looking out for me. It just—I'm, I'm an introvert. I don't want to bother people. You know, am I really so bad that I need to interrupt somebody today? I don't want to be a bother. But I was able to experience the effectiveness of the system, because [my provider] reached out to me. So, from a patient point of view, I can tell you, that particular item worked.*
>
> *And then there was one survey when I was starting to feel a little more anxious than I normally do. I got asked to complete a survey. And I did. And I know when I completed that survey, [my provider] reached out to me and said, "Hey, I noticed that you're feeling a little more anxious than you seem to be normally." So, I was. I'm not sure what prompted that. But I thought it was interesting that at a time I kind of had a need to communicate with [my provider], the system seemed to have recognized that.*
>
> *Overall, I think that it is a wonderful tool to have in one's toolbox to help support mental wellness.*

Illustrative quotations regarding concern about Rhythms were as follows:

> *Also, I remember first time and a few times after that, too, I got - because I know sometimes when you answer things in a way that is concerning, they say someone will call after every survey, I got that message. Sometimes it was someone may call and sometimes someone will call…I remember because one time I left my phone at home and I went out and I was like, oh my god, someone's gonna call and they're gonna message. Yeah, that message was a little scary sometimes…*
>
> *I didn't like hearing that I may or may not receive a call from a provider, I would have liked it if it was more definitive.*
>
> *Like, I don't sleep with my mobile device…and I don't always exercise with my phone.*

Reliability

The Rhythms system performed well during the pilot study with no outages and no major bugs detected.

Discussion

Principal Findings

Here, we describe the use, including acceptability, of the smartphone platform Rhythms by patients and providers in an academic department of psychiatry. Overall, the system integrated well with the health system's EHR and clinical workflow and was able to be deployed with very few technical issues to a group of 104 participants. The algorithm for urgent and emergent alerts allowed the team to identify concerning changes in patient mental health status and may be considered by other clinical groups wishing to use Rhythms. The lack of adverse outcomes during this period suggests that the algorithm was set conservatively enough to capture those who are in need of outreach, but not so sensitive as to lead to inappropriate alerts that could unnecessarily burden the patient or overwhelm the monitoring team.

An impressive percentage (63/87, 72%) of study participants remained on the platform and continued to use it without any compensation beyond the end of the 6-week study, suggesting they found the app helpful. This is in distinction to the findings from Kopka and colleagues [11] showing that use of 4 out of 10 popular mental health apps dropped below 20% within the 2 weeks of initiation. None of the 10 apps chosen by participants in Kopka and colleagues [11] collected passively acquired data or were integrated into the individual's health record. In our study, poststudy interviews revealed that passively collected data lead to a greater connection with one's provider and was a major strength of Rhythms. Participants also provided important feedback regarding the functionality (ie, ability to track ratings visually over time and receiving immediate feedback) of Rhythms that has now been incorporated into the platform. Importantly, privacy was not a major concern expressed by participants in this study, unlike those in some [26-29] but not all previous studies [12].

The concern about having to leave the app open is real, but this did not deter the majority of individuals from remaining on the app. While providers reported that Rhythms is a useful tool in their clinical practice, some participants reported that their provider did not discuss their Rhythms data during their

XSL•FO
RenderX

appointment. This may reflect the timing of Rhythms use and the availability of collected data in relation to participants' appointments, or it may be indicative of differing levels of provider adoption of the technology and how they incorporated it into their own practices. The use of objective measurements for patient care, which is quite common in most fields of medicine, is only recently making its way into the practice of psychiatry. We are encouraged that when the Rhythms data were brought into the session, the patients reported a benefit. While scalable technology holds great promise for improving psychiatric treatment, the acceptance and use of technology in psychiatry would require a culture change if measurement assisted care is to become a reality. Since the COVID-19 pandemic, many psychiatrists and other mental health specialists are now comfortable using videoconferencing for patient visits. However, incorporating data collected from apps and wearables is an unfamiliar step in the therapeutic interaction.

Likewise, many health systems have not yet embraced linkage of apps used to collect PGHD for mental health care purposes with their EHR. When there is collection of these data through medical records standards initiative such as Fast Healthcare Interoperability Resources and other platforms [30], information is not easily collated across multiple patient-derived sources. Recent advances have been made in this area through platforms, such as mindLAMP from the Digital Psychiatry Program at Beth Israel Deaconess Medical Center of Harvard Medical School, which has the power to incorporate PGHD collected by multiple apps and wearables and offers tools to help configure data such that they are more readily usable for research or clinical care [31]. Within this context, the benefit of Rhythms is that it can be seamlessly integrated into the patient's EHR such that the provider can view passively acquired data as well as ratings with 1 click in the patient's chart. The alert algorithm is strong and can help guide others in their use of Rhythms to detect emerging patient crises.

Those psychiatrists who incorporated Rhythms data into their clinical care reported that it improved their understanding of their patients' clinical status and that they used these data in their treatment planning. Likewise, patients appreciated the added connection to their provider, and there were no reports of increased provider burden across the study.

Several limitations should be considered. In total, 45.2% (178/394) of potential participants did not respond to either a general inquiry sent through MHC or an email letter sent by their provider. It is unclear why they did not respond to messages through the EHR, although previous research suggests that patient preferences for communication method are influenced by the type of information to be received [32]. Willingness to participate in a study that requires downloading an app onto one's personal phone may require direct interaction with one's provider and a clear discussion of the potential benefits for their treatment. The sample predominantly identified as women and White, limiting generalizability of these findings to other populations. The relatively short duration of this initial trial (6 weeks for each participant) and the relatively small number of patient and provider participants was also a limitation. Though the clinical value of Rhythms was quickly apparent, a longer period of monitoring is needed to determine the economic impact of remote patient monitoring of mental health status on health care use and costs. All study participants were currently in active mental health treatment and, on the whole, relatively stable. How Rhythms would perform in a more acutely ill population or those with psychosis within the greater health system is unknown but should be considered in future studies, as psychiatric disorders can negatively impact one's motivation and ability to engage with an app and their overall health care [33,34].

In summary, Rhythms is a user-friendly, digital platform that can be embedded into a health system's EHR and used to identify patients who are experiencing a worsening of their mental health between visits with their mental health care provider. The AI algorithms are based upon passively acquired data, limiting the need for patient surveys except when deviations in data suggest a deterioration in the patient's clinical status. The culture surrounding the use of technology as part of measurement assisted mental health care will become more welcoming as studies such as ours show the ease and clinical benefits of using remote patient monitoring tools.

Acknowledgments

This work was funded in part by a research grant from Health Rhythms. The authors thank Dr. Chloe E Page for assistance with submission and copyediting.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

CNE contributed to the study conception and design, data collection, analysis and interpretation of results, and manuscript preparation. RD contributed to study design, data collection, and manuscript preparation. AD contributed to data collection, interpretation of results, and manuscript preparation. HCH contributed to data collection, data management, and manuscript preparation. DJK participated in study conception and manuscript preparation. TL contributed to study design, data collection, and manuscript preparation. PMV participated in study design, data collection, and manuscript preparation. MM participated in data collection and manuscript preparation. SLM participated in study conception and design, data collection, analysis and interpretation of results, and manuscript preparation. KM participated in study conception and preparation. CDS participated in

XSL•FO
RenderX

study design, data collection, and manuscript preparation. JLS participated in study design, data collection, analysis and interpretation of results, and manuscript preparation. RDZ contributed to manuscript preparation. EF contributed to study conception and design and manuscript preparation.

Conflicts of Interest

CNE is a consultant to Skyland Trail, BabyScripts, and EmbarkNeuro. CNE was the principal investigator on this study, which was supported in part by a grant from Health Rhythms. EF is an employee of and holds equity in Health Rhythms. TL, MM, CNE, and DJK hold equity in Health Rhythms. AD received research grant support from Health Rhythms. KM is an investor in Health Rhythms.

References

1. Arias D, Saxena S, Verguet S. Quantifying the global burden of mental disorders and their economic value. EClinicalMedicine 2022;54:101675 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101675] [Medline: 36193171]
2. Counts N. Understanding the US behavioral health workforce shortage. The Commonwealth Fund. 2023. URL: https://www.commonwealthfund.org/publications/explainer/2023/may/understanding-us-behavioral-health-workforce-shortage [accessed 2024-11-02]
3. Beck AJ, Page C, Buche J, Rittman D, Gaiser M. Estimating the distribution of the U.S. psychiatric subspecialist workforce. Health Workforce Technical Assistance Center. 2018. URL: https://www.healthworkforceta.org/research-alerts/estimating-the-distribution-of-the-us-psychiatrist-subspecialist-workforce/ [accessed 2024-11-08]
4. Satiani A, Niedermier J, Satiani B, Svendsen DP. Projected workforce of psychiatrists in the United States: a population analysis. Psychiatr Serv 2018;69(6):710-713. [doi: 10.1176/appi.ps.201700344] [Medline: 29540118]
5. Jetty A, Petterson S, Westfall JM, Jabbarpour Y. Assessing primary care contributions to behavioral health: a cross-sectional study using medical expenditure panel survey. J Prim Care Community Health 2021;12:21501327211023871 [FREE Full text] [doi: 10.1177/21501327211023871] [Medline: 34109860]
6. Kyanko KA, A Curry L, E Keene D, Sutherland R, Naik K, Busch SH. Does primary care fill the gap in access to specialty mental health care? a mixed methods study. J Gen Intern Med 2022;37(7):1641-1647 [FREE Full text] [doi: 10.1007/s11606-021-07260-z] [Medline: 34993864]
7. Beasley JW, Hankey TH, Erickson R, Stange KC, Mundt M, Elliott M, et al. How many problems do family physicians manage at each encounter? a WReN study. Ann Fam Med 2004;2(5):405-410 [FREE Full text] [doi: 10.1370/afm.94] [Medline: 15506571]
8. Richman EL, Lombardi BM, Zerden LD. Mapping colocation: using national provider identified data to assess primary care and behavioral health colocation. Fam Syst Health 2020;38(1):16-23. [doi: 10.1037/fsh0000465] [Medline: 32202831]
9. Drincic A, Prahalad P, Greenwood D, Klonoff DC. Evidence-based mobile medical applications in diabetes. Endocrinol Metab Clin North Am 2016;45(4):943-965 [FREE Full text] [doi: 10.1016/j.ecl.2016.06.001] [Medline: 27823614]
10. Su D, Michaud TL, Estabrooks P, Schwab RJ, Eiland LA, Hansen G, et al. Diabetes management through remote patient monitoring: the importance of patient activation and engagement with the technology. Telemed J E Health 2019;25(10):952-959. [doi: 10.1089/tmj.2018.0205] [Medline: 30372366]
11. Kopka M, Camacho E, Kwon S, Torous J. Exploring how informed mental health app selection may impact user engagement and satisfaction. PLOS Digit Health 2023;2(3):e0000219 [FREE Full text] [doi: 10.1371/journal.pdig.0000219] [Medline: 36989237]
12. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
13. Frank E, Wallace ML, Matthews MJ, Kendrick J, Leach J, Moore T, et al. Personalized digital intervention for depression based on social rhythm principles adds significantly to outpatient treatment. Front Digit Health 2022;4:870522 [FREE Full text] [doi: 10.3389/fdgth.2022.870522] [Medline: 36120713]
14. Malkoff-Schwartz S, Frank E, Anderson BP, Hlastala SA, Luther JF, Sherrill JT, et al. Social rhythm disruption and stressful life events in the onset of bipolar and unipolar episodes. Psychol Med 2000;30(5):1005-1016. [doi: 10.1017/s0033291799002706] [Medline: 12027038]
15. Abdullah S, Murnane ELM, Matthews M, Choudhury T. Circadian computing: sensing, modeling, and maintaining biological rhythms. In: Rehg J, Murphy S, Kumar S, editors. Mobile Health: Sensors, Analytic Methods, and Applications. Cham, Switzerland: Springer; 2017:35-58.
16. Schuler D, Namioka A. Participatory Design: Principles and Practices. Boca Raton, FL: CRC press; 1993.
17. Patton MQ. Qualitative Research and Evaluation Methods. 3rd ed. Thousand Oaks, CA: Sage Publications; 2002.
18. Creswell JW, Clark VLP. Designing and Conducting Mixed Method Research. 2nd ed. Thousand Oaks, CA: Sage Publications; 2011.
19. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

XSL•FO
RenderX

20. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatric Annals 2002;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
21. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. J Gen Intern Med 2007;22(11):1596-1602 [FREE Full text] [doi: 10.1007/s11606-007-0333-y] [Medline: 17874169]
22. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry 2011;168(12):1266-1277 [FREE Full text] [doi: 10.1176/appi.ajp.2011.10111704] [Medline: 22193671]
23. Mundt JC, Greist JH, Jefferson JW, Federico M, Mann JJ, Posner K. Prediction of suicidal behavior in clinical research by lifetime suicidal ideation and behavior ascertained by the electronic Columbia-Suicide Severity Rating Scale. J Clin Psychiatry 2013;74(9):887-893. [doi: 10.4088/JCP.13m08398] [Medline: 24107762]
24. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
25. Altman EG, Hedeker D, Peterson JL, Davis JM. The Altman Self-Rating Mania Scale. Biol Psychiatry 1997;42(10):948-955. [doi: 10.1016/S0006-3223(96)00548-3] [Medline: 9359982]
26. Matthews M, Abdullah S, Gay G, Choudhury T. Tracking mental well-being: balancing rich sensing and patient needs. Computer 2014;47(4):36-43. [doi: 10.1109/mc.2014.107]
27. Melcher J, Camacho E, Lagan S, Torous J. College student engagement with mental health apps: analysis of barriers to sustained use. J Am Coll Health 2022;70(6):1819-1825. [doi: 10.1080/07448481.2020.1825225] [Medline: 33048626]
28. Di Matteo D, Fine A, Fotinos K, Rose J, Katzman M. Patient willingness to consent to mobile phone data collection for mental health apps: structured questionnaire. JMIR Ment Health 2018;5(3):e56 [FREE Full text] [doi: 10.2196/mental.9539] [Medline: 30158102]
29. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. JMIR Mhealth Uhealth 2019;7(4):e11223 [FREE Full text] [doi: 10.2196/11223] [Medline: 30990458]
30. Hassol A, Walker JM, Kidder D, Rokita K, Young D, Pierdon S, et al. Patient experiences and attitudes about access to a patient electronic health care record and linked web messaging. J Am Med Inform Assoc 2004;11(6):505-513 [FREE Full text] [doi: 10.1197/jamia.M1593] [Medline: 15299001]
31. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. J Am Med Inform Assoc 2016;23(5):899-908 [FREE Full text] [doi: 10.1093/jamia/ocv189] [Medline: 26911829]
32. Vaidyam A, Halamka J, Torous J. Enabling research and clinical use of patient-generated health data (the mindLAMP Platform): digital phenotyping study. JMIR Mhealth Uhealth 2022;10(1):e30557 [FREE Full text] [doi: 10.2196/30557] [Medline: 34994710]
33. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018;21(3):116-119 [FREE Full text] [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
34. Stiles-Shields C, Montague E, Lattie EG, Kwasny MJ, Mohr DC. What might get in the way: barriers to the use of apps for depression. Digit Health 2017;3:2055207617713827 [FREE Full text] [doi: 10.1177/2055207617713827] [Medline: 29942605]

Abbreviations

AI: artificial intelligence
ASRM: Altman Self-Rating Mania Scale
C-SSRS: Columbia-Suicide Symptom Severity Rating Scale
COMIRB: University of Colorado Multiple Institution Review Board
CU: University of Colorado
EHR: electronic health record
GAD-7: Generalized Anxiety Disorder 7-Item
HIPAA: Health Insurance Portability and Accountability Act
MDD: major depressive disorder
MHC: MyHealthConnection
PCP: primary care physician
PGDH: patient-generated health data
PHQ-9: Patient Health Questionnaire 9-Item
SDK: software development kit
UCH: University of Colorado Hospital
UCHealth: University of Colorado Health
VBHC: Virtual Behavioral Health Center

XSL•FO
RenderX

Edited by A Mavragani; submitted 21.02.24; peer-reviewed by D King, PH Liao; comments to author 22.06.24; revised version received 23.07.24; accepted 06.09.24; published 07.01.25.

Please cite as:
Epperson CN, Davis R, Dempsey A, Haller HC, Kupfer DJ, Love T, Villarreal PM, Matthews M, Moore SL, Muller K, Schneck CD, Scott JL, Zane RD, Frank E
The Trifecta of Industry, Academic, and Health System Partnership to Improve Mental Health Care Through Smartphone-Based Remote Patient Monitoring: Development and Usability Study
JMIR Form Res 2025;9:e57624
URL: https://formative.jmir.org/2025/1/e57624
doi:10.2196/57624
PMID:

©C Neill Epperson, Rachel Davis, Allison Dempsey, Heinrich C Haller, David J Kupfer, Tiffany Love, Pamela M Villarreal, Mark Matthews, Susan L Moore, Kimberly Muller, Christopher D Schneck, Jessica L Scott, Richard D Zane, Ellen Frank. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Co-Design of an Escape Room for e-Mental Health Training of Mental Health Care Professionals: Research Through Design Study

Joyce J P A Bierbooms[1], PhD; Wouter R J W Sluis-Thiescheffer[2], PhD; Milou Anne Feijt[3], MSc; Inge M B Bongers[1], PhD

[1]Tilburg University, Tilburg, Netherlands

[2]HAN University of Applied Sciences, Arnhem and Nijmegen, Netherlands

[3]Eindhoven University of Technology, Eindhoven, Netherlands

Corresponding Author:
Joyce J P A Bierbooms, PhD
Tilburg University
PO box 90153
Tilburg, 5000 LE
Netherlands
Phone: 31 13 466 31 39
Email: j.j.p.a.bierbooms@tilburguniversity.edu

Abstract

Background: Many efforts to increase the uptake of e-mental health (eMH) have failed due to a lack of knowledge and skills, particularly among professionals. To train health care professionals in technology, serious gaming concepts such as educational escape rooms are increasingly used, which could also possibly be used in mental health care. However, such serious-game concepts are scarcely available for eMH training for mental health care professionals.

Objective: This study aims to co-design an escape room for training mental health care professionals' eMH skills and test the escape room's usability by exploring their experiences with this concept as a training method.

Methods: This project used a research through design approach with 3 design stages. In the first stage, the purpose, expectations, and storylines for the escape room were formulated in 2 co-design sessions with mental health care professionals, game designers, innovation staff, and researchers. In the second stage, the results were translated into the first escape room, which was tested in 3 sessions, including one web version of the escape room. In the third stage, the escape room was tested with mental health care professionals outside the co-design team. First, 2 test sessions took place, followed by 3 field study sessions. In the field study sessions, a questionnaire was used in combination with focus groups to assess the usability of the escape room for eMH training in practice.

Results: An escape room prototype was iteratively developed and tested by the co-design team, which delivered multiple suggestions for adaptations that were assimilated in each next version of the prototype. The field study showed that the escape room creates a positive mindset toward eMH. The suitability of the escape room to explore the possibilities of eMH was rated 4.7 out of 5 by the professionals who participated in the field study. In addition, it was found to be fun and educational at the same time, scoring 4.7 (SD 0.68) on a 5-point scale. Attention should be paid to the game's complexity, credibility, and flexibility. This is important for the usefulness of the escape room in clinical practice, which was rated an average of 3.8 (SD 0.77) on a 5-point scale. Finally, implementation challenges should be addressed, including organizational policy and stimulation of eMH training.

Conclusions: We can conclude that the perceived usability of an escape room for training mental health care professionals in eMH skills is promising. However, it requires additional effort to transfer the learnings into mental health care professionals' clinical practice. A straightforward implementation plan and testing the effectiveness of an escape room on skill enhancement in mental health care professionals are essential next steps to reach sustainable goals.

(JMIR Form Res 2025;9:e58650) doi:10.2196/58650

XSL•FO
RenderX

KEYWORDS

serious gaming; mental health care professionals; e-mental health; skill enhancement; training

Introduction

Background

Despite a growing body of evidence for the effectiveness and benefits of e-mental health (eMH), many studies report a relatively low use of eMH in practice [1-3]. eMH means mental health services delivered or supported by digital technologies, such as video calling, self-management and self-tracking apps, web-based treatment modules, and virtual reality (VR) [4,5]. An important factor hampering the use of eMH is its relatively slow adoption by mental health care professionals [5-7]. To a large extent, this slow adoption results from a lack of knowledge and skills regarding eMH among mental health care professionals [7-9].

This lack of skills entails general digital skills but, more importantly, the ability to use an adequate communication approach compensating for a potential lack of nonverbal cues and contextual information. In addition, skills are needed to guide a professional in handling boundaries (eg, expectations regarding reaction time and attainability), choosing appropriate communication channels for each situation, and up-to-date knowledge of the availability of technological possibilities to enrich treatment in mental health care [10-12]. The COVID-19 pandemic caused a significant increase in video calling and, subsequently, the skills to apply such technologies for therapeutic purposes. However, mental health care professionals are still hesitant to pursue broader incorporation of eMH (ie, using different eMH tools in a context where remote treatment is a choice rather than an imposition) into their therapeutic practices because of the skills required [13]. Thus, to increase the adoption of a broad range of eMH tools, there is a persistent need to enhance mental health care professionals' skills in using eMH.

Many training options have recently been developed to capacitate mental health care professionals in using eMH [14,15]. The introduction and development of web-based treatment platforms have brought different programs to acquire the technical skills to use such a platform. Examples are applying web modules, written client interactions, and video sessions instead of real-life therapeutic interactions. Besides the classical training approach, there are several eLearning options to support mental health care professionals in digital skill development [16]. However, research on the effectiveness of different learning and training strategies in health care has shown that traditional forms, such as classical instructions and eLearning, generate insufficient learning outcomes [17,18]. This can be explained by a lack of experiential learning in conventional learning methods; there is a need for a "learning by doing" approach. The learning by doing approach is more engaging, more realistic, and contributes to a larger self-efficacy [17,18]. Thus, there is a need to pursue new, more innovative, and experiential solutions to train mental health care professionals in using eMH.

To gain a feeling of self-efficacy regarding the application of eMH, a potential strategy is to offer mental health care professionals training possibilities based on the concept of serious gaming [19]. Serious games are physical or digital games applied to learn or acquire skills and usually offer a combination of educational content and engagement [19,20]. Escape rooms have emerged as a popular instructional approach within serious gaming, attracting growing interest from researchers and educators [21]. An escape room is a game in which participants are confronted with a storyline, usually locked into a room, and need to solve puzzles, search for information, and follow leads to free themselves from the room within a limited amount of time. Escape room games are increasingly used for learning [22-24].

The benefit of an escape room may be that it offers the option to train skills by gaining hands-on experiences (experiential learning). Escape rooms create realistic learning environments that foster meaningful participation and provide a supportive atmosphere for experimentation [25]. In contrast, other methods (eg, classical instructions or eLearning) mainly focus on providing instructions on using an artifact or service [18,26]. In addition, an escape room is focused on the learning process while simultaneously being educational and fun, which makes it much more engaging to participate in training [27,28].

The underlying learning mechanisms of an educational escape room (EER) are often described in terms of the self-determination theory (SDT) [29,30] and the flow theory [30,31]. The SDT refers to fundamental psychological needs such as autonomy, competence, and relatedness to others [29,30]. EERs provide a safe environment that can be tuned to the player's competencies, and an appropriate storyline helps players relate the competencies demanded in the EER puzzles to competencies required in real-world situations. Together, they increase a player's sense of autonomy, and they can create a positive experience in self-determination. The flow theory explains the occurrence of someone's complete immersion into a specific task based on positive experiences [30,31]. This "flow" can be achieved in serious gaming if the challenges meet the player's knowledge and abilities. By attempting to address these fundamental needs in a playful environment, users can have a positive learning experience [32].

Theoretically, these mechanisms provide a solid argument for the use of EERs to train mental health care professionals' skills in successfully integrating eMH into their therapeutic practice [22,23]. However, as Vorderobermeier et al [30] indicate, the success of EERs is hardly ever tracked back to these mechanisms. To contribute to the emerging field of research into EERs, we include an evaluation of the participants' learning experience in relation to these learning mechanisms.

EERs in eMH

In eMH, the concept of EERs is recognized as a safe and interactive environment to explore eMH in different possible therapeutic situations [30,32] while providing a safe and social environment for training in mental health care. EERs provide

XSL•FO
RenderX

an opportunity to experiment with otherwise complex or expensive therapeutic situations [33,34]. As EERs are an emerging phenomenon, there are also some recent examples of research into EERs in the context of eMH.

In the context of raising awareness of severe mental illnesses (SMIs), Rodriguez-Ferrer et al [35] studied the effectiveness of an EER in educating first-year nursing students (a nonspecialized audience) in experiencing the stigmas associated with SMI. The EER was developed with professionals in SMI and the effect of raising awareness was tested using a randomized controlled trial experiment. Apart from the primary objective (raising awareness of SMI), the EER objectives were to determine whether an entirely web-based version was effective and whether participants were sufficiently immersed in the experience. The EER in this study is different in the sense that it is aimed at experiential learning, involving both digital and physical interactions. Moreover, the EER developed in this study is aimed at a professional audience, which means that the challenges are tuned toward expert knowledge and an in-depth skillset rooted in ample experience in treating mental illnesses [35].

Aragon [36] designed and studied the use of an EER for mental health nursing students attending an accelerated associate degree program. The topic of learning was communication and collaboration strategies in mental health. The design of the EER took a pragmatic, digital form to simplify the strategy for the development, distribution, and deployment of the EER in the educational program. The authors studied the potential learning effect of the digital EER but found no significant difference compared to students who were not exposed to the EER. The EER in this paper is different, similar to the EER studied by Rodriguez-Ferrer et al [35] mentioned earlier, as it targets a professional audience and aims to provide a more physical, collaborative learning experience. Moreover, the EER by Aragon [36] was solely designed by one researcher, whereas this paper proposes a co-design process to include the views, needs, and goals of the target audience in the design of the EER. Finally, the work by Aragon [36] briefly discusses the underlying mechanisms of the EER, namely (1) the potentially positive effects of collaboratively solving nursing challenges and (2) the strong relatedness of the puzzles with the curriculum thus providing the students with alternative means to consolidate what was learned during the program [36]. The EER in this paper makes a clear choice to define the underlying mechanisms in the SDT theory and the flow theory.

There are more EERs for eMH, such as the one proposed in the study by Petkari and Calvo [37], and the number of EERs is likely to increase over time as its potential is currently widely researched. Similar to the studies discussed earlier, Petkari and Calvo [37] focused on EERs as alternative or additional elements for existing curricula, targeted at soon-to-be professionals. In those contexts, the design of the EER is mostly in the hands of educators or researchers, while the professionals are only involved in evaluations [37].

In a previous paper, we described the careful considerations we took to aim for a successful design and subsequent deployment of an EER that is tailored toward experiential learning about (innovative) digital elements suitable for blended therapies [38]. This paper describes the execution of a preliminary version of the EER.

This study aims to (1) further pursue the ambitions of a co-designed EER for training mental health care professionals in eMH, (2) investigate whether an EER would be an acceptable solution for this purpose, and (3) determine whether the SDT theory and flow theory are recognized as effective mechanics based on the experience of professionals, thus making a contribution to the field of eMH and the emerging field of EERs in the context of health care.

Co-Designing an EER for eMH Therapy

On the basis of these encouraging arguments and inspiring examples of eMH EERs in pursuing the development of an escape room, we gauged practitioners' first reactions and expectations regarding this idea. This delivered mostly positive responses: mental health care professionals expect an escape room to be innovative, engaging, inviting, and much more reasonable than the traditional and familiar eMH training methods. Moreover, based on our user requirements analysis [39], we aimed to develop a physical escape room. One argument is that a primary obstacle mental health care professionals reported in the user requirement analysis is that they are unacquainted with the possibilities of eMH. On the basis of this, we assumed that a physical room, with a facilitator present to moderate and explain, if necessary, would yield better results. A physical escape room provides more opportunities for collaboration [40]. Moreover, we learned from the user requirements analysis that users preferred a learning context in which they met others and could discuss with colleagues. Finally, playing an escape room is not a stand-alone solution. It should be part of a holistic experience, including a clear introduction and a reflection afterward, where participants can transfer the value of their eMH experience in the escape room to their real-life work practice [39].

As we aimed to develop a serious game for training mental health care professionals in using eMH, it is important to classify an escape room according to the gameplay/purpose/scope model for serious games [41]. This allows a better understanding of an escape room as a serious game and whether it meets our aim of developing a game with educational content. According to the definition by Djaouti et al [41], which distinguishes game-based from play-based activities by noting that the latter lacks a clear goal, we can, we can classify the gameplay of our intended escape room as game-based. The goal of the escape room is to enhance knowledge of the use of eMH by mental health professionals and train them to use eMH tools in different situations. That immediately brings us to the purpose of the escape room, which, in the classification by Djaouti et al [41], would be training. The intention of the escape room is to improve the performance of mental health professionals in using eMH, including how to use certain techniques as well as how to use these techniques in a particular context. We could also argue that there is a data exchange component in our escape room, as we also aim to encourage the users to help each other solve the puzzles and learn from each other's experiences. The scope of the escape room is health care, with mental health

XSL•FO
RenderX

professionals as the main target audience. However, there is also the possibility that an escape room to train eMH skills is incorporated into educational content at vocational or higher professional education [41].

Objectives

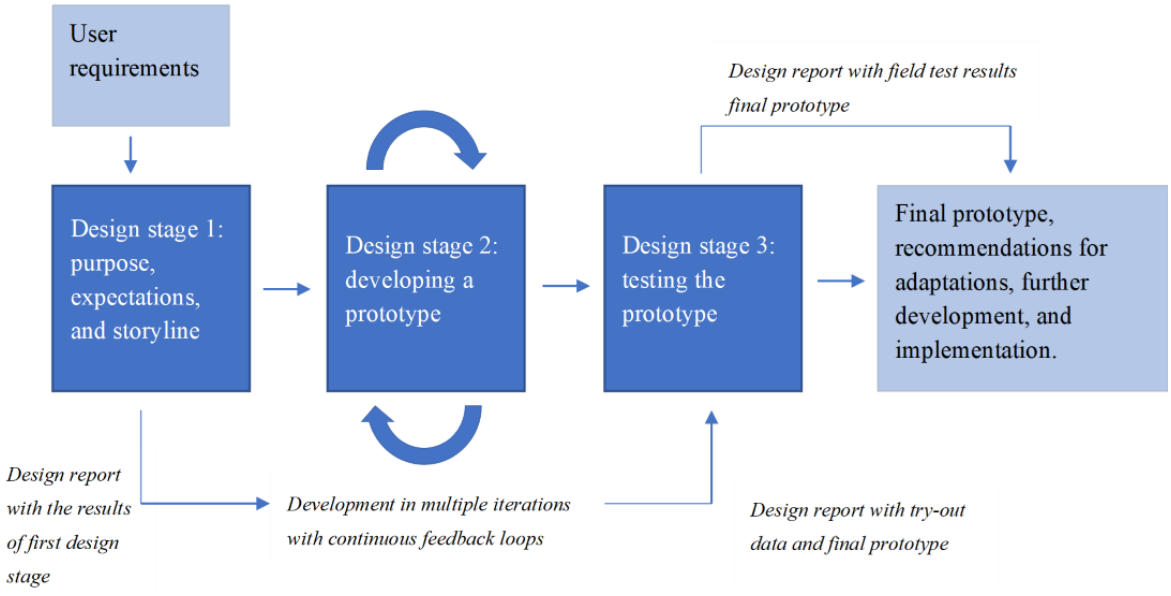
This paper presents the co-design process of developing an escape room for training mental health care professionals in eMH. By assessing the expectations and experiences regarding an EER in mental health care, we aim to evaluate the perceived usability and to gauge the effect of underlying mechanisms rooted in SDT and flow theory for training mental health care professionals' eMH skills.

Methods

Study Design

To shape the development and explore the perceived usability of an escape room, we used a process of 3 design stages, in which a testable prototype of an escape room for mental health care professionals was developed. In this "research through design" process, the design of the escape room is part of a research project in which the design activities contribute to generating knowledge [42]. Zimmerman et al [43] define research through design as "a research approach that employs methods and processes from design practice as a legitimate method of inquiry." This knowledge is applied at each stage to iteratively develop prototypes within the design process [42,43], leading to the final prototype. To explore the potential of an escape room, users' experiences and ideas are gathered on whether the escape room works as intended and whether it lives up to the promise of a concept for knowledge and skill enhancement. With this research through design approach, we seek to contribute to evidence-based decisions on the implementation, including potential adaptations and strategies to maximize the users' acceptance of the concept. This means that the design of an escape room is explored for suitability in the practice of mental health care and for its potential to address the aim of enhancing professionals' skills in using eMH. We organized 9 co-design sessions in 3 design stages (Figure 1). The escape room's purpose, expectations, and storyline ideas were formulated in the first design stage. These ideas were further translated into a gaming concept (ie, an escape room prototype), assessed for suitability in the second stage. In the third stage, a prototype was tested with end users (ie, mental health care professionals). Data from all stages within this process contributed to exploring the concept's perceived usability. In the Exploration of an Escape Room in 3 Design Stages subsection in the Results section, we report on the specific procedures and results of the co-design sessions. In doing this, we integrate the procedures and results to describe each design stage as a whole.

Figure 1. Three design stages in the development of an escape room for e-mental health (eMH) training.



Research Context

The study was conducted at a mental health care organization in Eindhoven (Geestelijke Gezondheidszorg Eindhoven en De Kempen, GGzE), in the southern part of the Netherlands. GGzE delivers specialized mental health care for people with various psychiatric and psychological issues (eg, depression, anxiety disorders, psychosis, autism, and personality disorders). Mental health care professionals working at GGzE include (specialized) psychologists, psychiatrists, nurses, and social workers. GGzE treats approximately 12,000 clients annually and has 2368 employees [44]. The research involved researchers from Tilburg University, Eindhoven Technical University, and Fontys University of Applied Sciences. Through this collaboration between universities, we combined expertise from social science, human and technology interaction, and game design. This combination of expertise vouches for a project in which all necessary aspects of a useful design process (eg, technique, human attitudes and behavior, and serious gaming concepts) and their interaction are considered.

XSL•FO
RenderX

Participants

Stages 1 and 2

The co-design sessions, including 1 web version in stage 2, were held with 5 mental health care professionals (psychologists: n=2, 40%; nurses: n=2, 40%; and social worker: n=1, 20%), 1 innovation staff member, 3 researchers, 3 research laboratory leaders, and 2 game designers. Together, they formed the co-design team for this stage in the design process (Table 1). In line with the existing literature on co-design, we involved stakeholders from the three areas of expertise that were needed for the design project [45-47]: (1) the mental health care professionals that represent the end users of the escape room, the innovation staff that will be involved in launching the escape room in practice; (2) the game designers who have expertise in the development of an escape room; and (3) researchers that are involved in exploring the usability of and experiences with an escape room. The mental health care professionals and the innovation staff members worked at GGzE. They were recruited through an announcement on the organization's internal communication platform and through network contacts of the project team, that is, a convenience sampling method was used. Mental health care professionals with varying levels of expertise were included in this stage of the development process. After their initial agreement, the participants were further informed about the project, the purpose of their participation, and how their data would be handled.

Table 1. Participants' roles in the development of an escape room for e-mental health (eMH) training.

Participants	Design stage 1	Design stage 2	Web version (design stage 2)	Design stage 3
Psychologists	• Participants (n=2)	• Participants (n=2)	—[a]	• Participant field study session (n=3)
Nurses and therapists	• Participant (n=1)	• Participants (n=2)	Participant (n=1)	• Participant field study session (n=2)
Social workers	• Participant (n=1)	—	—	• Participant field study session (n=5)
Mental health professionals' team for web-based treatment	—	—	—	• Participant test session 2 (n=4)
Innovation staff members	• Participant (n=1)	• Participants (n=2)	Participant (n=1)	• Participant test session 1 (n=2) • Facilitator field study sessions 1 to 3 (n=1)
Researchers	• Participants (n=3)	• Participant (n=1) • Observers (n=2)	Participant (n=1)	• Observant field study sessions 1-3 (n=2)
Research laboratory leaders	• Facilitator (n=1) • Participant (n=1) • Observer (n=1)	• Observers (n=3)	Participant (n=1)	—
Game designers	• Facilitator (n=1) • Participant (n=1)	• Facilitator (n=1) • Observant (n=1)	Facilitator (n=1)	—

[a]Not applicable.

Stage 3

In the third stage of the design project, we first asked a purposively sampled selection of mental health care professionals, namely 4 highly experienced eMH therapists from the team for web-based treatment and 2 members of the innovation staff of GGzE, to participate in 2 test sessions. The 2 test sessions secured a sufficient perceived flow of the escape room workshop (ie, including introduction and reflection). They eliminated further technical and practical issues concerning the logistics escape room (Table 1). Subsequently, we invited therapists from all disciplines within ambulatory specialized mental health care to participate in 3 field study sessions to test whether the final prototype of the escape room would be suitable as a concept to enhance mental health care professionals' skills in using eMH (Table 1). The first session included 2 psychologists, 1 nurse specialist, and 1 social worker. The second session involved 5 therapists engaged in social work and supporting clients' daily activities. Finally, a clinical psychologist and a verbal therapist tested the escape room in the third session. All mental health care professionals in the 3 field study sessions had some essential experience with eMH (ie, using a web-based module and video calling on a nonfrequent basis).

Data Collection and Analysis

Detailed notes were taken for all co-design sessions in all stages and were used to compose written reports. These reports were checked with the participants, after which themes were identified to describe, organize, and interpret the data [48]. In addition, video recordings were made of the sessions in stage 3, which were used to verify and complement the written reports. In the

XSL•FO
RenderX

first stage, the 2 sessions were analyzed individually because of the difference in nature of the sessions. In the second stage, the written reports were summarized into recommendations for the design team, which they used in the next iteration of the design. In the third stage, the first 2 test sessions were analyzed together, and the 3 field study sessions were also combined for analysis.

We used the Standards for Quality Improvement Reporting Excellence reporting guidelines to check the completeness and improve the quality of our manuscript [49]. A completed checklist can be found in Multimedia Appendix 1.

Ethical Considerations

This project was approved by the ethics review board of Tilburg University (EC-2013.14) and by the scientific committee of mental health care in Eindhoven (the Netherlands), where the participants were recruited. All participants were informed about the research and allowed to withdraw at any moment. All participants signed the informed consent form. Data in this project were anonymized and cannot be linked back to participants. Participation was voluntary, without financial compensation.

Results

Exploration of an Escape Room in 3 Design Stages

We describe the 3 stages of our research through the design process subsequently (Figure 1). As explained in the Methods section, we elaborate on the specific procedures we followed in each design stage to gather the data, the results generated, and how the findings within all 3 stages provide insights into the usability of an escape room as a game-based solution for eMH training for mental health care professionals (Figure 1).

Design Stage 1: Purpose, Expectations, and Storyline

Overview

In the first co-design session of design stage 1, which took place in December 2019, a joint purpose and common ground for the escape room were formulated. Participants were invited to work in subgroups to discuss the following: (1) the current situation, including the mental health care professionals' attitude regarding eMH; (2) a vision of how future mental health care professionals should be equipped and accommodated to integrate eMH into their daily practice; and (3) the requirements regarding an escape room to transition from the current to the future mental health care professional, in which eMH plays a central role, including the success and failure factors for this concept.

This was done by exchanging ideas, associating concepts, and explaining the relationship between eMH and the work of mental health care professionals.

Following this, a second co-design session was organized, aimed to deliver the storyline for the envisioned escape room. Information was gathered on the particulars of clients in mental health care, the mental health care setting in general, the various treatments and services available, including eMH, and the different client journeys that occur. This provided information on the requirements regarding the content of the game. The designers delineated a fictitious client and a story to form the basis of the escape room game. This initial storyline was verified with the mental health care professionals of the co-design team and was adjusted accordingly.

Purpose

In the first co-design session, the participants mentioned that many mental health care professionals are still convinced that the best way to treat a client is through face-to-face interaction. There is still a feeling of eMH being imposed by financial incentives, offering very limited job enrichment for mental health care professionals and very limited options to enrich treatment possibilities for clients. As a result, practitioners often need more interest or time to explore the possibilities of eMH. An important driver for mental health care practitioners to use eMH in the future would be that they see a clear-cut advantage in it for their clients, such as feeling less of a "patient" and gaining self-control and autonomy. This could be established by gaining positive experiences using eMH, for example, in training. Another aspect of the current attitude of many mental health care professionals mentioned in the co-design session was the lack of knowledge and skills to use eMH. Enhancing this knowledge requires time that needs to be facilitated by the organization they work for.

Despite this attitude, mental health care practitioners see the need to integrate eMH into therapy. The future mental health care professional is seen as someone for whom eMH is a natural part of the job. The session revealed that mental health care professionals consider their profession a craft, and some may see innovation as detrimental to this craft. According to the participants in the session, an escape room should demonstrate the opposite: instead of devaluating the craft, eMH has to reinforce it.

Expectations

On the basis of the purpose of the escape room, it is essential that sharing positive experiences and building confidence in using eMH are emphasized when designing a prototype. In the co-design session, it was stated that examples of positive experiences with eMH in clinical practice should be visible in the escape room, along with opportunities for professionals to share their positive experiences. Furthermore, the educational content of the escape room should seamlessly fit various real-life scenarios, making a transfer to everyday practice easy. Specific and well-defined learning goals should provide the opportunity to gain confidence in using technology, and the escape room should consist of customizable elements to fit the particular learning needs. Finally, the escape room should be engaging and contain continuous feedback loops, a recognizable storyline, and multiplayer options for a shared learning experience.

In addition to game-specific criteria, conditional requirements, such as perceived ease of use, a robust (technological) system, and perceived usefulness were considered significant. Furthermore, the social influence of colleagues, the perception of available time, endorsement by management, and an excellent incentive to engage in the escape room are issues that need to be considered. Finally, the escape room should be a complete learning experience with a clear introduction and reflection after

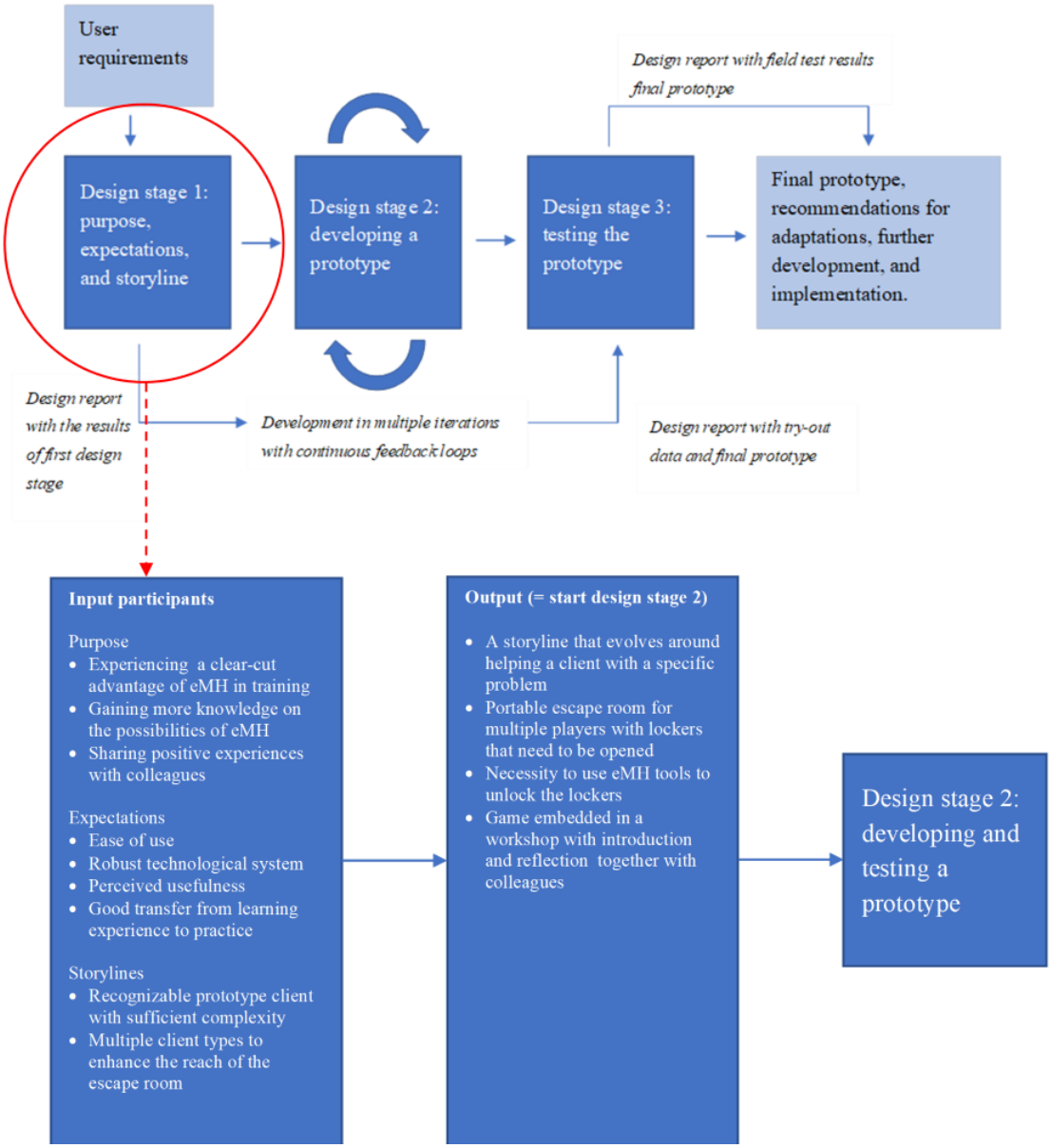
XSL•FO
RenderX

playing the game. Mental health care professionals must transfer from the learning experience to their therapeutic practice.

Storylines

On the basis of the input in the first session, the designers crafted a storyline for the escape room, which was presented to the co-design team. In their reflection on the storyline, the co-design team remarked that the mental health problems of the fictitious client should be more complex to be eligible for specialized mental health care. It should also contain more details to be sufficiently recognizable for therapists. In this stage of the design process, the co-design team also concluded that it would be valuable if there were sets of multiple storylines revolving around different settings or problems in mental health care to enhance the reach among the potential users of the escape room (Figure 2).

Figure 2. Results of the first design iteration in the development of an escape room for e-mental health (eMH) training.



Design Stage 2: Developing a Prototype

Overview

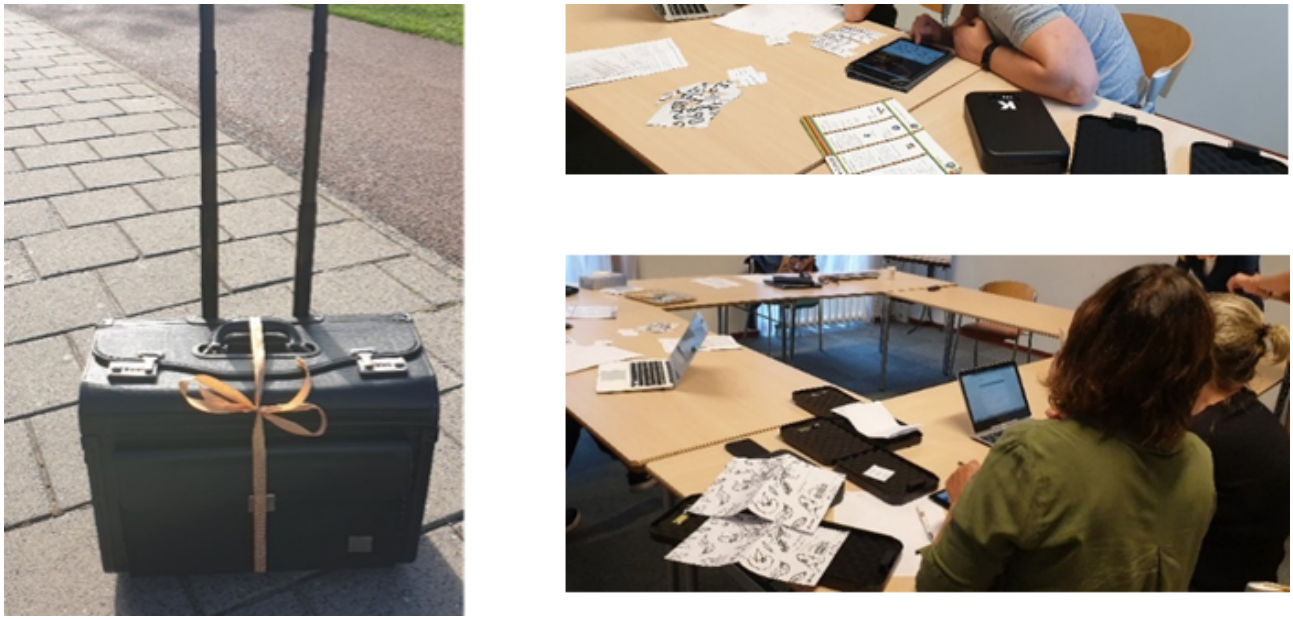
The prototype was presented to the co-design team in February 2020. The prototype of the escape room consists of 2 storylines, which are played separately. In each storyline, the participants are introduced to a client who experiences mental health problems. The challenge in the escape room is to answer questions in the client's electronic patient file, leading to a final solution to help the client take the next step. These questions can only be answered by finding clues in several locked safes, thoroughly analyzing the information given about the client, and using the eMH tools they see in the room. The eMH tools the participants provided to "help their client" were an electronic diary, a VR application for exposure therapy, biofeedback, a chat function, and web-based treatment platform. All materials used in the escape room are portable, which means playing the escape room is not dependent on a fixed place (Figure 3). The actual game is embedded in a workshop with an introduction to familiarize with the topic, goals, and expectations regarding

XSL•FO
RenderX

the escape room and a reflection afterward to discuss the experiences and take-home messages.

In a try-out session, the co-design team tested this prototype by playing the escape room in 3 small groups of 3 participants. This delivered additional data for our research regarding the expectations and experiences of an escape room. At the same time, it provided the designers with feedback for a second prototype. In each round, the participants were given 15 minutes to play the game and solve the puzzles. Each subgroup reflected individually on the prototype directly after playing the escape room. At the end of the session, a plenary discussion took place to summarize the results. The designers developed a second prototype for the escape room based on the feedback. In this phase, a second storyline was also developed. Different eMH tools were added and various connections were made between these eMH tools, the storyline, and the puzzles.

Figure 3. Picture of an escape room setting.



In March 2020, the world was confronted with an unprecedented global pandemic that caused a significant disruption of life as we knew it. Burdened by the consequences of the pandemic and the measures that national governments took to control the virus (eg, social distancing, lockdown, and quarantine), the activities in our design process were also heavily disturbed. Switching to this new situation, we sought to continue the design process by developing a web-based alternative to test the second prototype of the escape room in April 2020. This web version was a digital copy of the physical room, containing the same elements, puzzles, and eMH tools, except for the VR goggles. This second prototype included 2 storylines, each with a fictitious client, and was tested by the co-design team members. One storyline was tested while being moderated by the designers, the other storyline was tested by each member individually, without being moderated.

Try-Out Sessions

The feedback gathered during the try-out sessions delivered several expectations and experiences of mental health care professionals regarding an escape room. First, storylines in an escape room benefit from a good match with a client's pathway in mental health care and reflect truly relatable situations. Moreover, it is essential to focus on mental health problems as an integrated part of everyday life situations. This is important because otherwise, the suggested eMH tools cannot be appropriately linked, and a genuine eMH experience will be missed. This suggestion was also given in response to the final storyline in the previous phase of the design process. Second, there should not be too much focus on solving the puzzles in the escape room, as it diminishes the awareness of eMH in the game. More importantly, the eMH tools should be linked explicitly to the puzzles. The participants should experience a (positive) consequence when choosing, for example, finding the next clue when trying a specific eMH tool or experiencing that someone (ie, the fictitious client) benefits from you as a practitioner choosing to use eMH. Third, variation and differentiation in an escape room vouches for a stronger and fuller user experience. This could be established by multiple storylines (as was stated in the previous phase of the design process), different sets of eMH tools that can be used, and variations in the level of difficulty in the escape room. In this way, the escape room can be played multiple times and addresses different levels of expertise, either regarding eMH or playing an escape room. Fourth, there should be a clear conclusion to the escape room and, consequently, a euphoric feeling when finishing the game. To this end, the purpose at the beginning of the game may need to be more apparent. Fifth, it should be beyond any doubt that the experience of eMH in the escape room can also be transferred to the daily practice of a mental health care professional. Otherwise, the escape room will not be considered credible.

Web Version

From the tryout, it became clear that the experience of a web version of an escape room does not closely resemble the experience of an escape room in real life. Due to several technical hassles, the game's flow was hindered multiple times.

XSL•FO
RenderX

Most importantly, it impeded the possibility of team interactivity, working together, and exchanging ideas about eMH, which was mentioned multiple times in the design process as one of the critical success factors for an escape room.

Final Prototype

After receiving the feedback on the web version of the escape room, the designers continued developing a prototype that was fit for testing by end users. The co-design team subsequently assessed a final prototype in July 2020. After minor adaptations, this prototype was finalized in September 2020. Following this, a field study was planned (stage 3) to gather the user experiences of the escape room among the end users, that is, mental health care professionals from various disciplines (Multimedia Appendix 2).

Design Stage 3: Testing the Final Prototype

Test Sessions

In 2 test sessions (September and October 2020), we first checked the flow of the escape room workshop and the functioning of the materials and technology. This delivered a positive outcome regarding the potential of the concept for exploring eMH and, in that way, enhancing mental health care professionals' knowledge about the possibilities of technology in therapy. The participants of these sessions shared several thoughts on how to improve the workshop before starting the field study sessions with end users.

The first finding in the test sessions was the importance of good functioning of the content and technical aspects of the escape room to ensure a good flow in the game. Besides a good flow in the game, a facilitator should lead a good introduction and reflection to complement the learning experience.

According to the participants in the test sessions, a good explanation of the purpose and relevance of the escape room and presenting a game manual to maximize the effect of playing the escape room is necessary. One participant particularly mentioned the importance of stressing the educational purpose and preventing the feeling that the workshop is merely something for fun. After playing the escape room, reflecting on the purpose is important. According to the participants, there should be a clear take-home message, discussion on the experiences gained in the escape room, and acquired knowledge to share with colleagues afterward.

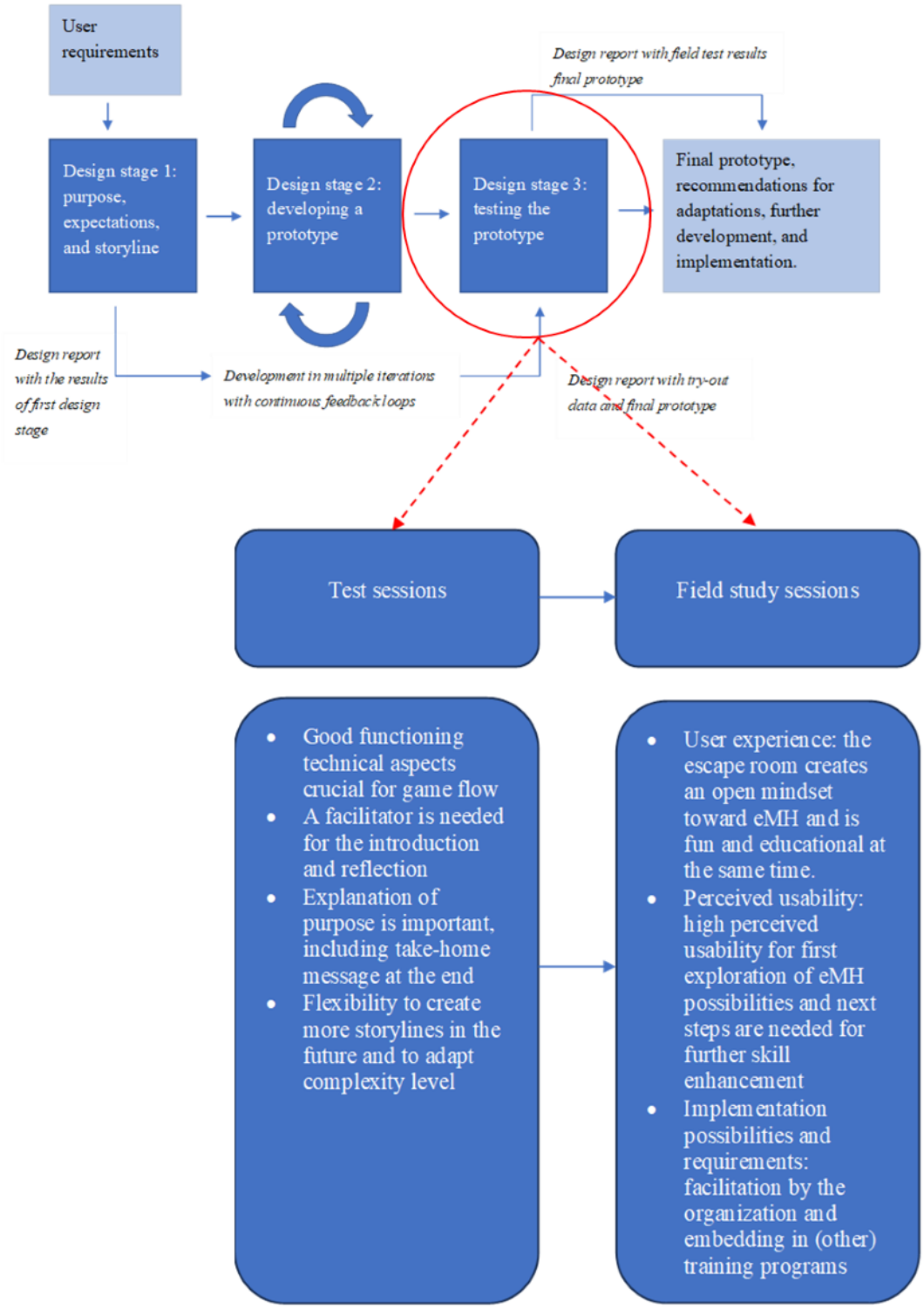
Another issue raised during the 2 test sessions was the need to continuously adapt the storylines or create the flexibility to insert additional storylines in the escape room game. Two storylines may be too few to address all therapeutic contexts, all levels of complexity in mental health care, and all eMH possibilities that could add value. The participants also marked a sense of reality in the storylines. Opportunism should be avoided because it should not imply that technology will always vouch for a positive treatment outcome. For example, in VR, the escape room should not pretend that it is something that instantly makes everything better. This was regarded as an essential issue to warrant the credibility of the escape room. However, the same session also mentioned that experiencing eMH tools in this manner created awareness about escape room's potential. At the same time, the biofeedback application also showed that it is possible to visually track panic attacks in such an application. The final aspect regarding the mentioned storylines is using the correct professional jargon in the escape room. At one point, a participant stated that the way things were written in the escape room did not reflect the language a therapist would use. In contrast, language and jargon appeared very personal, making it difficult to always be accurate in an escape room. This emphasizes the fact that there is a large diversity among mental health care professionals and the clients they are treating.

Finally, the participants discussed the complexity of the game itself. Some participants found figuring out the puzzles and clues challenging, but others remarked it was too simple. Some solutions were easily found without using the eMH tools in the escape room, resulting in an ineffective learning experience. Some participants in the test sessions suggested creating multiple levels of complexity in the escape room to address several types of game players (ie, level of experience with escape room games) and other kinds of therapists.

On the basis of the results of these 2 test sessions, minor adjustments were made regarding the complexity of the puzzles to ensure that clues and solutions could not be found without using eMH tools. Furthermore, the introduction and reflection were refined to embed the escape room game in a workshop format (Figure 4).

XSL•FO
RenderX

Figure 4. Results of the test and field study sessions (third iteration) in the development of an escape room for e-mental health (eMH) training.



Field Study Sessions

Overview

Hereafter, the workshop was tested in 3 sessions with mental health care professionals from different areas of expertise. A total of 11 mental health care professionals participated in these field study sessions (November and December 2020 and June 2021) and were asked to fill in a questionnaire (Multimedia Appendix 3) at the end of the workshop. Their answers formed the basis for a discussion about their user experiences. The questionnaire consisted of an overall experience rating (question 1), three parts to reflect on (1) the quality of the introduction (questions 2-4), (2) the game itself (questions 5-11), and (3) the reflection (questions 12-15), overall questions regarding the perceived usability (questions 16-19), and a question on strengths and suggestions (question 20). The questions in the 3 parts of the questionnaire were constructed based on the results of the co-design sessions in design stage 1. As we explained in design stage 1, mental health care professionals use eMH mainly because of its apparent benefits. To encourage this, the escape

XSL•FO
RenderX

room should therefore highlight positive experiences with eMH in clinical practice, in line with the flow theory that states that a positive learning experience can be reached if the challenges meet the abilities of the users and if they can share this with colleagues [31]. To explore the extent to which the escape room was experienced positively, we formulated questions 1, 3, 9, 10, and 13. Furthermore, from design stage 1 and SDT, we learned that the educational content must align with real-life scenarios, making it easy to apply in practice [29]. Questions 6, 8, and 15 were therefore constructed to assess whether the storylines are recognizable and whether the training with eMH tools in these storylines connects to their daily practice. In addition, questions 2 and 5 were formulated to explore if the learning goals become clear, and questions 7 and 12 assessed whether the escape room helps to build confidence in using technology, with customizable elements tailored to specific needs. Finally, the game should enable multiplayer options for shared learning [31]. To assess whether the escape room was experienced that way, we added questions 4, 11, and 14. Together this led to the questionnaire, which had 20 targeted questions. The answers to the questions were subsequently used as a guide in discussion with the participants to gain a more in-depth understanding of the possibilities of an escape room as an environment for mental health care professionals to experiment and practice with eMH.

User Experiences

From the questionnaire and discussion, we learned that the participants in the field study sessions were enthusiastic about the workshop they participated in. They stated that the workshop is inspiring and creates a positive mindset toward eMH. They recognized the potential to remove barriers they had at first, showing that eMH is accessible and not always that difficult to use. The participants experienced significant fun in the escape room, making it feel less obligatory. This is expected to help people engage more easily. In addition, the alternation between playing and reflecting was mentioned as meaningful. In total, 2 (18%) of the 11 participants mentioned they were pleased that there was no explicit element of competition in the escape room. According to these participants, too much competition, for example, ranking who has solved the escape room the fastest, would distract the users from the storyline and the eMH possibilities they come across.

All but one participant expressed that they experienced the storylines in the escape room as realistic. One mental health care professional did not recognize the storylines in the escape room. Another aspect related to this was the “credibility” of the suggested effect of eMH in the escape room. Although most (9/11, 82%) of the participants could describe the use of eMH to certain benefits for mental health care, it should be carefully suggested in the escape room. However, most (9/11, 82%) participants experienced how they could practically apply eMH in therapeutic situations. For example, VR was experienced in the escape room, resulting in a discussion on how VR could be used for clients in specific treatment situations. The participants were able to connect their VR experience in the workshop to real-life situations and could see the added value of using eHealth. All kinds of eHealth suggestions were exchanged, which illustrates the effect of being given the possibility to explore and experiment with eHealth in this type of playful setting.

Participants wondered whether an escape room would appeal enough to therapists who were skeptical about eMH. The participants also mentioned that the current storylines and eMH tools may not be recognizable for all mental health care professionals. It was considered important to explore eMH tools that are also accessible in real life to be able to transfer knowledge from training to reality. In contrast, it is also mentioned that mental health care professionals with vast experience in eMH will find it more attractive to explore more experimental eMH tools.

Perceived Usability of an Escape Room for eMH Training

To explore the perceived usability of an escape room for skill enhancement in eMH, we asked the participants to indicate whether the workshop provided them with a more elaborate and precise picture of the possibilities of eMH (Table 2). The feedback on this issue varied among the participants. In total, 2 (18%) of the 11 mental health care professionals reported that they had not gained a better picture of the options of eMH in the escape room. This was because they already had a relatively good basic knowledge about eMH beforehand. The participants who were less experienced users of eMH reflected that they had a much better idea about the possibilities of eMH after participating in the escape room workshop. It was suggested that the current escape room prototype would benefit from further development to create more options to try additional eMH tools because, in the current prototype, not all mental health care professionals would find tools suitable for their therapeutic practice.

Table 2. Perceived usability of an escape room (ER) for e-mental health (eMH) training (questionnaire results).

Items	Score on a scale of 1-5, mean (SD)
Suitability of an ER to explore possibilities of eMH	4.7 (0.92)
Possibilities for collaboration in the ER	4.7 (0.47)
Clarity of the storylines	4.3 (0.92)
Clarity of learning goal in the ER	4.6 (0.67)
Balance educational content and fun	4.7 (0.68)
Usefulness for daily practice	3.8 (0.77)

XSL•FO
RenderX

The participants largely agreed on the suitability of an escape room concept for exploring eMH (average score of 4.7, SD 0.92). They perceived the learning goal in the workshop as evident (average score of 4.6, SD 0.67). They also experienced a good balance between the educational and the fun part in the escape room (average score of 4.7, SD 0.68). It was mentioned in the discussion that suitability is supported when the bridge between the escape room and everyday practice is explicitly discussed in the workshop as well as the steps that can be taken after completing the workshop and returning to the workplace. Follow-up training is essential because otherwise the acquired knowledge will not be secured. According to the participants, this escape room offers an excellent first experience with the possibilities of eMH; more than that, it is just another educational workshop, "It is more 'getting to know eMH' than 'learning how to apply eMH." It was also suggested that more focus on the client groups that people are treating would be helpful in the reflection part of the workshop. The participants advocated that involving professionals with similar client groups or treatment contexts would be more effective in stimulating the exchange of experiences with eMH and the transferability of the experiences in the escape room to daily practice.

In line with the findings about follow-up training, the participants elaborately reflect on how to apply and secure knowledge on eMH once the first experiences are gained in the escape room. While this workshop helps to talk about eMH, keeping the conversation going in everyday practice is equally important. According to the participants in our study, mental health care professionals are focused on the content of their treatment and support, and the knowledge gained in the eMH escape room might lose its impact. It was stated that the organization should provide continuous information about what is new in the field of eMH and should provide time to invest in exploring new tools. The concept of an escape room or a game offers a very persuasive opportunity to invest in keeping knowledge up to date and motivating people. All workshops suggested that this would be an excellent fit for a teambuilding activity.

Implementation Possibilities and Requirements

In addition to gathering experiences with the escape room workshop, we asked the mental health care professionals who participated to indicate what they deemed essential for successfully implementing it. The first thing mentioned was that professionals in mental health care should be confronted with eMH from various directions, that is, in an escape room, but also by clients, colleagues, and organizational policies aimed at being innovative, requiring innovative and digitally skilled employees. Incorporating an escape room in an introductory meeting might be possible when people are new to the organization.

Another issue that emerged from the workshops concerns the approachability of the escape room. It is experienced as convenient that the escape room is currently a mobile version, which does not necessarily require people to go somewhere and spend much time traveling. It would be even more convenient if the escape room could be played without an external facilitator so that participants could be more flexible in their time planning. However, this remains the main concern of the participants in the workshops: will mental health care professionals be provided with enough time to explore eMH in the future, and will their production targets be adjusted accordingly? In addition, it should be decided who will provide training services and materials in the organization. Finally, the participants stated that a strong vision by the policy makers on the use of eMH in the organization and the expectations regarding the future mental healthcare professional is indispensable to warrant a successful implementation of an escape room or any serious gaming concept for eMH training in mental health care.

Discussion

Principal Findings

In this research through design study, we investigated the perceived usability of an escape room for training mental health care professionals in using eMH. On the basis of the results in each stage of our research and the overall findings, we can conclude that an escape room concept might be suitable for this type of training. We subsequently discuss our findings by reflecting on the experiences with our escape room prototype and what this means in light of the purpose of our design project: developing new training methods for mental health care professionals, with the ultimate goal of contributing to increasing the adoption of eMH. We conclude with an overall view of the possibilities of game-based training in mental health care based on the results of this design study.

What we know from earlier research, which also emerged in this study, is that many mental health care professionals are not convinced of the benefits of eMH, resulting in dismissive attitudes and behavior [50,51]. The escape room was experienced as both fun and educational; however, most importantly, it induced a more positive stance toward eMH tools. While an open mindset is a prerequisite for change, the escape room may offer a potential concept as the first step to changing attitudes and behavior among mental health care professionals regarding eMH [36,40].

For the escape room to be accepted as a valuable training method by its end users, the storylines, game elements (puzzles and clues), and eMH tools in the escape room needed to be credible and recognizable for mental health care professionals. At the same time, it became clear that it is tough to design storylines that seamlessly match the practices of mental health care therapists. In one of the first versions of the prototype, there was too much focus on everyday life situations instead of mental health problems. This is crucial because the participants need a genuine eMH experience to link to their treatment context properly. This feedback was given in different iterations in the design process, which marks the difficulty of designing a good and representative storyline and the importance of involving multiple disciplines in a co-design team [46,47].

Moreover, it became clear that mental health care professionals are diverse in terms of their client population, type of mental health care services, and level of adoption in eMH. This requires a flexible and adaptable training concept to expand the reach of the escape room. Greater diversification regarding the

XSL•FO
RenderX

storylines and available eMH tools in the escape room may be needed, along with various difficulty levels. The participants in the field study sessions indicated that the escape room in its current form might not be suitable for people who are already frequent users of eMH. This implies that the gains of the current escape room are probably highest for people who are relatively inexperienced eMH users, which applied to most mental health care professionals when we started the design process. Adding new, more innovative tools could, however, expand the reach of the escape room.

The participants welcomed the escape room as an engaging concept and noted that its design needs continuation and further development. In its current form, the escape room offers an environment for exploration and the first acquaintance with eMH rather than actual skill development in applying these tools. We believe that, whereas most mental health care professionals are—partly because of the coronavirus pandemic—familiar with standard eMH tools that are also used in everyday life (eg, videoconferencing and SMS text messaging), they remain unaware of the possibilities of more novel technologies, such as VR and biofeedback, and may be biased about its use in their daily clinical practice. Therefore, creating a different mindset remains an important foundation for further skill development in the future and increasing adoption in general. An extensive research base reports on various barriers that prevent accelerating the adoption of eMH [5,51,52]. This escape room addresses one of the critical barriers, a perceived lack of knowledge. However, issues such as compensating for the loss of nonverbal information, the lack of trust in technological systems, and the availability of time and resources to gain positive experiences in using eMH also need to be addressed.

When exploring ideas for further training for specific eMH tools, the question is whether an escape room would also be suitable for this particular type of training. Looking at escape rooms in health care, we find that escape rooms are most appropriate for exploration and awareness and less ideal for training-specific skills [23,24]. This was also why we chose an escape room concept in the first place; it suited the results of our user requirements analysis that revealed a specific need for exploring and enhancing knowledge rather than specific hands-on skills. Training these hands-on skills should be addressed for the next step, particularly because the participants in our study also indicated this need. Moreover, for both the possibility of exploration and specific skill training, the support in terms of time and resources provided by the organization and its management is crucial. The organization's commitment to a digital transition and the necessity for mental health care professionals to be equipped with 21st-century skills must be at the forefront. Finally, follow-up research is required to test the escape room's effectiveness regarding knowledge increase and skill enhancement.

The Perceived Usability of Serious Gaming for eMH Training in Mental Health Care Based on Users' Experiences

The results of our study show positive experiences with an escape room as a serious gaming concept for eMH training in mental health care. These positive experiences might be explained in terms of the SDT [29] and the flow theory [31]. By attempting to address the needs of competence and autonomy, the users may have had a positive experience. The flow of the escape room may have positively contributed to their motivation to enhance their knowledge of eMH and their general attitude toward using eMH in their clinical practice [30].

In the design of the escape room, we addressed several important game mechanics that we found to have influenced our users' experiences, for example, an adequate level of complexity in the game [19,53,54]. The participants in our study emphasized that when a puzzle in the escape room takes too much time or is too difficult, it may lead to frustration and an adverse effect. The opposite of this also works negatively; when clues are too obvious, people may not feel they are taken seriously. In general, our escape room was experienced as having an appropriate level of difficulty, which led to a positive experience. We also paid much attention to the importance of the story or narrative in the escape room. When this narrative is too farfetched or not relatable to clinical practice, it loses strength; when people can identify with the story in the game, it may be much more effective than any other learning method [29-31]. In addition, the aspect of social interaction [30] was predominantly present in our escape room and experienced as one of the critical factors for its success.

Strengths and Limitations

In this study, we particularly choose a research through design approach to develop and test an escape room as a serious gaming solution for eMH training among mental health care professionals. The strength of this approach is that it enabled us to use formative evaluations [55] between the stages of design to provide input for the next design stage. We strongly believe that without these intermediate assessments, we could not design a prototype suitable to the practice of mental health care providers. More importantly, the involvement of mental health care professionals in our co-design team ensured a prototype that matched the therapeutic practice we pursued to reflect in our escape room. However, there are some limitations. First of all, we used an inductive analysis method, which allowed us to gather any information deemed necessary for the development of the escape room but may have lacked the required structure to establish an evaluation of the effectiveness of the escape room, in terms of skill enhancement, and for other researchers to replicate the analyses. Analysis procedures that would allow an assessment of the escape room's success could be to use a pre- and posttest quasi-experimental design for the evaluation [56], a controlled experiment with naturalistic and summative evaluations to assess whether the purpose of the game is indeed reached [57] or a heuristic evaluation using predefined rules of thumb to measure the usability of the escape room [58]. A second limitation of this study is that we included a small sample of mental health care professionals who all worked at the same mental health care organization. In other words, we did not include participants from a diverse mental health care setting other than the differentiation in mental health care contexts within an organization. In addition, we are aware that for projects like this, where participants are recruited through open invitation, people who are already enthusiastic about the topic,

XSL•FO
RenderX

in this case, eMH, tend to be the ones who sign up for participation. In future research, it would be worthwhile to test the escape room or any other game-based training solution among mental health care professionals who are skeptical about eMH to see whether also, in this group, the effect of creating an open mindset through exploration and experience-based learning is achieved.

Conclusions

In general, we can conclude that experience-based training, such as game-based training in an escape room, helps people experience what they are being trained to do or use. This experience may be much stronger than the experience of learning through traditional learning (eg, classroom, textbook, or e-learning). When training is more fun and engaging, this will increase the learning experience. Moreover, it increases the participants' motivation and creates an open mindset. In our study, we learned that an escape room as a serious gaming concept stimulates the willingness to participate in training and be open to change. This openness to change is essential in increasing the use of eMH. However, it is important that there is a clear next step to this first exploration, providing professionals with the opportunity to thoroughly train their skills in specific eMH tools. Subsequently, it would be valuable to evaluate the effectiveness of game-based training in an escape room for enhancing particular skills. Furthermore, designing is a cooperation between different experts. The design process should be viewed from different angles and areas of expertise. The contribution of mental health care professionals was necessary to create a credible game in which both the storyline and set of integrated eMH tools can be transferred to their therapeutic practice. Finally, the success of a game-based training method, such as an escape room, depends on the commitment of the organization in which it is implemented. People need to feel that they have the time and support of their management to engage in training. This commitment is more than just telling them to "go and participate in training." The whole organization must enhance specific knowledge based on common ground and a shared vision of what is needed regarding competencies and skills by mental health care professionals.

Acknowledgments

The authors would like to thank all mental health care professionals and other participants who participated in this project. This work is supported by the Netherlands Organisation for Scientific Research grant 055.16.141, under the project titled Psychosocial Games Supporting Mental Health Professionals in Stress Reduction and Empathic Interactions in Remote/online Psychotherapy, under the Serious Games for Professional Skills Program.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to agreements about confidentiality and exchange of data that were made with the respondents in the study. The consent given by the participants does not allow for public availability. The data can be made available from the corresponding author on reasonable request and with the consent of the participants in this study.

Authors' Contributions

JB, MAF, and RJWST organized and conducted the co-design sessions. JB coded and analyzed the data and composed an initial draft for the manuscript. All authors contributed to multiple revisions. All authors agreed with the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
SQUIRE (Standards for Quality Improvement Reporting Excellence) checklist.
[DOCX File , 34 KB - formative_v9i1e58650_app1.docx]

Multimedia Appendix 2
Results of the second design iteration in the development of an escape room for e-Mental health training.
[DOCX File , 23 KB - formative_v9i1e58650_app2.docx]

Multimedia Appendix 3
Evaluation questions (English translation).
[PDF File (Adobe PDF File), 164 KB - formative_v9i1e58650_app3.pdf]

References

XSL•FO
RenderX

1. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct 01;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
2. de Grood C, Raissi A, Kwon Y, Santana M. Adoption of e-health technology by physicians: a scoping review. J Multidisciplin Healthc 2016 Aug;Volume 9:335-344. [doi: 10.2147/jmdh.s103881]
3. Glueckauf RL, Maheu MM, Drude KP, Wells BA, Wang Y, Gustafson DJ, et al. Survey of psychologists' telebehavioral health practices: technology use, ethical issues, and training needs. Prof Psychol Res Pract 2018 Jun;49(3):205-219. [doi: 10.1037/pro0000188]
4. Lal S, Adair CE. E-mental health: a rapid review of the literature. Psychiatr Serv 2014 Jan 01;65(1):24-32. [doi: 10.1176/appi.ps.201300009] [Medline: 24081188]
5. Feijt MA, de Kort YA, Bongers IM, IJsselsteijn WA. Perceived drivers and barriers to the adoption of eMental health by psychologists: the construction of the levels of adoption of eMental health model. J Med Internet Res 2018 Apr 24;20(4):e153 [FREE Full text] [doi: 10.2196/jmir.9485] [Medline: 29691215]
6. Vis C, Kleiboer A, Prior R, Bønes E, Cavallo M, Clark SA, et al. Implementing and up-scaling evidence-based eMental health in Europe: the study protocol for the MasterMind project. Internet Interv 2015 Nov;2(4):399-409. [doi: 10.1016/j.invent.2015.10.002]
7. Wozney L, Newton AS, Gehring ND, Bennett K, Huguet A, Hartling L, et al. Implementation of eMental Health care: viewpoints from key informants from organizations and agencies with eHealth mandates. BMC Med Inform Decis Mak 2017 Jun 02;17(1):78 [FREE Full text] [doi: 10.1186/s12911-017-0474-9] [Medline: 28577543]
8. Kilbourne AM. E-health and the transformation of mental health care. Psychiatr Serv 2012 Nov;63(11):1059. [doi: 10.1176/appi.ps.631103] [Medline: 23117506]
9. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
10. Chan CV, Kaufman DR. A framework for characterizing eHealth literacy demands and barriers. J Med Internet Res 2011 Nov 17;13(4):e94 [FREE Full text] [doi: 10.2196/jmir.1750] [Medline: 22094891]
11. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec 23;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
12. Schalken F. Handboek Online Hulpverlening: Met Internet Zorg en Welzijn Verbeteren. Houten, The Netherlands: Bohn Stafleu Van Loghum B.V; 2013.
13. Bierbooms JJ, van Haaren M, IJsselsteijn WA, de Kort YA, Feijt M, Bongers IM. Integration of online treatment into the "new normal" in mental health care in post-COVID-19 times: exploratory qualitative study. JMIR Form Res 2020 Oct 08;4(10):e21344 [FREE Full text] [doi: 10.2196/21344] [Medline: 33001835]
14. Crawford A, Gratzer D, Jovanovic M, Rodie D, Sockalingam S, Sunderji N, et al. Building eHealth and telepsychiatry capabilities: three educational reports across the learning continuum. Acad Psychiatry 2018 Dec 11;42(6):852-856. [doi: 10.1007/s40596-018-0989-0] [Medline: 30311178]
15. Echelard JF, Méthot F, Nguyen HA, Pomey MP. Medical student training in eHealth: scoping review. JMIR Med Educ 2020 Sep 11;6(2):e20027 [FREE Full text] [doi: 10.2196/20027] [Medline: 32915154]
16. Timman A, Caris JA. Leren in een andere realiteit. Onderwijs en Gezondheidszorg 2022;1:20-23. [doi: 10.24078/oeng.2022.2.128459]
17. Connolly TM, Boyle EA, MacArthur E, Hainey T, Boyle JM. A systematic literature review of empirical evidence on computer games and serious games. Comput Educ 2012 Sep;59(2):661-686. [doi: 10.1016/j.compedu.2012.03.004]
18. Squire K, Jenkins H. Harnessing the power of games in education. Insight 2003;3:5-33.
19. Wang R, DeMaria SJ, Goldberg A, Katz D. A systematic review of serious games in training health care professionals. Simul Healthc 2016 Feb;11(1):41-51. [doi: 10.1097/SIH.0000000000000118] [Medline: 26536340]
20. Susi T, Johannesson M, Backlund P. Serious games – an overview. Serious Games Cluster and Business Network Project. 2007 Feb 5. URL: https://www.diva-portal.org/smash/get/diva2:2416/FULLTEXT01.pdf [accessed 2024-12-16]
21. Makri A, Vlachopoulos D, Martina RA. Digital escape rooms as innovative pedagogical tools in education: a systematic literature review. Sustainability 2021 Apr 20;13(8):4587. [doi: 10.3390/su13084587]
22. Ritterfeld U, Cody M, Vorderer P. Serious Games: Mechanisms and Effects. Milton Park, UK: Taylor & Francis; 2009.
23. Adams V, Burger S, Crawford K, Setter R. Can you escape? Creating an escape room to facilitate active learning. J Nurses Prof Dev 2018;34(2):E1-E5. [doi: 10.1097/nnd.0000000000000433]
24. Mullen M, Seiler SO. Escaping the confines of traditional instruction: the use of the escape room in a transition to practice program. J Nurses Prof Dev 2019;35(6):354-356. [doi: 10.1097/nnd.0000000000000571]
25. Veldkamp A, van de Grint L, Knippels MC, van Joolingen WR. Escape education: a systematic review on escape rooms in education. Educ Res Rev 2020 Nov;31:100364. [doi: 10.1016/j.edurev.2020.100364]
26. Kolb DA, Boyatzis RE, Mainemelis C. Experiential learning theory: previous research and new directions. In: Sternberg RJ, Zhang LF, editors. Perspectives on Thinking, Learning, and Cognitive Styles. New York, NY: Routledge; 2014.

XSL•FO
RenderX

27. Prensky M. Fun, play and games: what makes games engaging. In: Digital Game-Based Learning. New York, NY: McGraw-Hill Education; 2001:5-31.
28. Grande-de-Prado M, García-Martín S, Baelo R, Abella-García V. Edu-escape rooms. Encyclopedia 2020 Dec 22;1(1):12-19. [doi: 10.3390/encyclopedia1010004]
29. Deci EL, Ryan RM. Handbook of Self-Determination Research. New York, NY: University of Rochester Press; 2004.
30. Vorderobermeier A, Abel J, Sailer M. Theoretical foundations and approaches in research on educational escape rooms: a systematic review. Educ Res Rev 2024 Aug;44:100625. [doi: 10.1016/j.edurev.2024.100625]
31. Csikszent M. Flow: The Psychology of Optimal Experience. New York, NY: HarperCollins; 1991.
32. Deen M. G.A.M.E., Games autonomy motivation & education: how autonomy-supportive game design may improve motivation to learn. Technische Universiteit Eindhoven. 2015. URL: https://pure.tue.nl/ws/files/3895781/784813.pdf [accessed 2024-12-07]
33. Okuda Y, Bryson EO, DeMaria SJ, Jacobson L, Quinones J, Shen B, et al. The utility of simulation in medical education: what is the evidence? Mt Sinai J Med 2009 Aug 29;76(4):330-343. [doi: 10.1002/msj.20127] [Medline: 19642147]
34. Reeves B, Read JL. Total Engagement: How Games and Virtual Worlds Are Changing the Way People Work and Businesses Compete. Brighton, MA: Harvard Business Review Press; 2009.
35. Rodriguez-Ferrer JM, Manzano-León A, Cangas AJ, Aguilar-Parra JM. A web-based escape room to raise awareness about severe mental illness among university students: randomized controlled trial. JMIR Serious Games 2022 May 05;10(2):e34222 [FREE Full text] [doi: 10.2196/34222] [Medline: 35511232]
36. Aragon CJ. Determining the effectiveness of an escape room in a mental health nursing course: a quantitative study. Northcentral University. 2023. URL: https://www.proquest.com/openview/afaa70dd57076e8ae4bab3fa3631ac20/1?pq-origsite=gscholar&cbl=18750&diss=y [accessed 2024-12-16]
37. Petkari E, Calvo A. Escaping psychodiagnostics: implementation of a Digital Educational Escape Room with future mental health professionals. Interact Learn Environ 2024 May 23:1-10. [doi: 10.1080/10494820.2024.2354429]
38. Bierbooms J, Feijt M, IJsselsteijn W, Sluis-Thiescheffer W, Weijmans M, van der Burg E, et al. Serious games for professional skills: the design of an escape room to explore the possibilities of eMental health. In: Proceedings of the Companion Publication of the 2020 ACM Designing Interactive Systems Conference. 2020 Presented at: DIS' 20 Companion; July 6-10, 2020; Eindhoven, The Netherlands. [doi: 10.1145/3393914.3395888]
39. Bierbooms J, Feijt MA, IJsselsteijn WA, Bongers IM. Design of a game-based training environment to enhance mental health care professionals' skills in using e-mental health: multiple methods user requirements analysis. JMIR Serious Games 2022 Jul 27;10(3):e34700 [FREE Full text] [doi: 10.2196/34700] [Medline: 35896032]
40. Pan R, Lo H, Neustaedter C. Collaboration, awareness, and communication in real-life escape rooms. In: Proceedings of the 2017 Conference on Designing Interactive Systems. 2017 Presented at: DIS '17; June 10-14, 2017; Edinburgh, UK. [doi: 10.1145/3064663.3064767]
41. Djaouti D, Alvarez J, Jessel JP. Classifying serious games: the G/P/S model. In: Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches. Hershey, PA: IGI Global; 2011.
42. Stappers PJ, Giaccardi E. Research through design. In: Soegaard M, Friis-Dam R, editors. The Encyclopedia of Human-Computer Interaction. Aarhus, Denmark: The Interaction Design Foundation; 2014.
43. Zimmerman J, Forlizzi J, Evenson S. Research through design as a method for interaction design research in HCI. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2007 Presented at: CHI '07; April 28-May 3, 2007; San Jose, CA. [doi: 10.1145/1240624.1240704]
44. Jaarverslag GGzE 2020: werken aan mentale vooruitgang. GGzE. URL: https://www.jaarverslagggze.nl/ [accessed 2021-10-22]
45. Donetto S, Pierri P, Tsianakas V, Robert G. Experience-based co-design and healthcare improvement: realizing participatory design in the public sector. Des J 2015 May 07;18(2):227-248. [doi: 10.2752/175630615X14212498964312]
46. Jessen S, Mirkovic J, Ruland CM. Creating gameful design in mHealth: a participatory co-design approach. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e11579 [FREE Full text] [doi: 10.2196/11579] [Medline: 30552080]
47. Robert G. Participatory action research: using experience-based co-design to improve the quality of healthcare services. In: Ziebland S, Coulter A, Calabrese JD, Locock L, editors. Understanding and Using Health Experiences: Improving Patient Care. Oxford, UK: Oxford University Press; 2013.
48. Boyatzis RE. Transforming Qualitative Information: Thematic Analysis and Code Development. Ann Arbor, MI: The University of Michigan; 1998.
49. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. BMJ Qual Saf 2016 Dec;25(12):986-992 [FREE Full text] [doi: 10.1136/bmjqs-2015-004411] [Medline: 26369893]
50. Connolly SL, Miller CJ, Lindsay JA, Bauer MS. A systematic review of providers' attitudes toward telemental health via videoconferencing. Clin Psychol (New York) 2020 Jan 06;27(2):10.1111/cpsp.12311 [FREE Full text] [doi: 10.1111/cpsp.12311] [Medline: 35966216]

XSL•FO
RenderX

51. Davies F, Shepherd HL, Beatty L, Clark B, Butow P, Shaw J. Implementing web-based therapy in routine mental health care: systematic review of health professionals' perspectives. J Med Internet Res 2020 Jul 23;22(7):e17362 [FREE Full text] [doi: 10.2196/17362] [Medline: 32706713]
52. Stoll J, Müller JA, Trachsel M. Ethical issues in online psychotherapy: a narrative review. Front Psychiatry 2019 Feb 11;10:993 [FREE Full text] [doi: 10.3389/fpsyt.2019.00993] [Medline: 32116819]
53. Ijaz A, Khan MY, Ali SM, Qadir J, Kamel Boulos MN. Serious games for healthcare professional training: a systematic review. Eur J Biomed Inform 2019;15(1):14-30 [FREE Full text]
54. Rooney P. Creating serious games at third level: evaluating the implications of an in-house approach. In: Proceedings of the European Conference on Game Based Learning. 2012 Presented at: ECGBL 2012; October 4-5, 2012; Cork, Ireland.
55. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
56. Moizer J, Lean J, Dell'Aquila E, Walsh P, Keary A, O'Byrne D, et al. An approach to evaluating the user experience of serious games. Comput Educ 2019 Jul;136:141-151. [doi: 10.1016/j.compedu.2019.04.006]
57. Pistono AM, dos Santos AM, Baptista RJ, Mamede HS. Framework for adaptive serious games. Comput Appl Eng Educ 2024 Mar 07;32(4):e22731. [doi: 10.1002/cae.22731]
58. Karavidas L, Apostolidis H, Tsiatsos T. Usability evaluation of an adaptive serious game prototype based on affective feedback. Information 2022 Sep 08;13(9):425. [doi: 10.3390/info13090425]

Abbreviations

EER: educational escape room
eMH: e-Mental health
GGzE: Geestelijke Gezondheidszorg Eindhoven en de Kempen
SDT: self-determination theory
SMI: severe mental illness
VR: virtual reality

Edited by A Mavragani; submitted 21.03.24; peer-reviewed by A Frey, IK Singgih, A Ferworn, E Vallefuoco; comments to author 10.05.24; revised version received 22.08.24; accepted 21.10.24; published 07.01.25.

Please cite as:
Bierbooms JJPA, Sluis-Thiescheffer WRJW, Feijt MA, Bongers IMB
Co-Design of an Escape Room for e-Mental Health Training of Mental Health Care Professionals: Research Through Design Study
JMIR Form Res 2025;9:e58650
URL: https://formative.jmir.org/2025/1/e58650
doi:10.2196/58650
PMID:

©Joyce J P A Bierbooms, Wouter R J W Sluis-Thiescheffer, Milou Anne Feijt, Inge M B Bongers. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Measuring Bound Attention During Complex Liver Surgery Planning: Feasibility Study

Tim Schneider[1*], MSc; Timur Cetin[1*], Dr rer nat; Stefan Uppenkamp[2], PD, Dr rer nat; Dirk Weyhe[1], Prof Dr Med; Thomas Muender[3], Dr-Ing; Anke V Reinschluessel[3,4], Dr-Ing; Daniela Salzmann[1], Dr phil; Verena Uslar[1], Dr rer nat

[1]University Hospital for Visceral Surgery, PIUS-Hospital, Department for Human Medicine, Faculty VI, University of Oldenburg, Oldenburg, Germany

[2]Department of Medical Physics and Acoustics, Faculty VI, University of Oldenburg, Oldenburg, Germany

[3]Digital Media Lab, Department of Mathematics and Computer Science, Faculty III, University of Bremen, Bremen, Germany

[4]Department of Computer and Information Science, University of Konstanz, Konstanz, Germany

[*]these authors contributed equally

Corresponding Author:
Tim Schneider, MSc
University Hospital for Visceral Surgery, PIUS-Hospital
Department for Human Medicine, Faculty VI
University of Oldenburg
Ammerländer Heerstraße 114-118
Oldenburg, 26129
Germany
Phone: 49 4412292922
Email: tim.schneider@uol.de

Abstract

Background: The integration of advanced technologies such as augmented reality (AR) and virtual reality (VR) into surgical procedures has garnered significant attention. However, the introduction of these innovations requires thorough evaluation in the context of human-machine interaction. Despite their potential benefits, new technologies can complicate surgical tasks and increase the cognitive load on surgeons, potentially offsetting their intended advantages. It is crucial to evaluate these technologies not only for their functional improvements but also for their impact on the surgeon's workload in clinical settings. A surgical team today must increasingly navigate advanced technologies such as AR and VR, aiming to reduce surgical trauma and enhance patient safety. However, each innovation needs to be evaluated in terms of human-machine interaction. Even if an innovation appears to bring advancements to the field it is applied in, it may complicate the work and increase the surgeon's workload rather than benefiting the surgeon.

Objective: This study aims to establish a method for objectively determining the additional workload generated using AR or VR glasses in a clinical context for the first time.

Methods: Electroencephalography (EEG) signals were recorded using a passive auditory oddball paradigm while 9 participants performed surgical planning for liver resection across 3 different conditions: (1) using AR glasses, (2) VR glasses, and (3) the conventional planning software on a computer.

Results: The electrophysiological results, that is, the potentials evoked by the auditory stimulus, were compared with the subjectively perceived stress of the participants, as determined by the National Aeronautics and Space Administration-Task Load Index (NASA-TLX) questionnaire. The AR condition had the highest scores for mental demand (median 75, IQR 70-85), effort (median 55, IQR 30-65), and frustration (median 40, IQR 15-75) compared with the VR and PC conditions. The analysis of the EEG revealed a trend toward a lower amplitude of the N1 component as well as for the P3 component at the central electrodes in the AR condition, suggesting a higher workload for participants when using AR glasses. In addition, EEG components in the VR condition did not reveal any noticeable differences compared with the EEG components in the conventional planning condition. For the P1 component, the VR condition elicited significantly earlier latencies at the Fz electrode (mean 75.3 ms, SD 25.8 ms) compared with the PC condition (mean 99.4 ms, SD 28.6 ms).

Conclusions: The results suggest a lower stress level when using VR glasses compared with AR glasses, likely due to the 3D visualization of the liver model. Additionally, the alignment between subjectively determined results and objectively determined results confirms the validity of the study design applied in this research.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e62740) doi:10.2196/62740

KEYWORDS

workload measurement; virtual reality; VR; augmented reality; AR; electroencephalography; EEG; event-related potential; ERP; auditory evoked potential; AEP; oddball experiment; National Aeronautics and Space Administration Task Load Index; NASA-TLX; surgical planning

Introduction

In modern surgery, new techniques are increasingly being adopted. One example of such innovative technologies is minimally invasive procedures, which aim to reduce surgical trauma and enhance patient safety [1]. However, while these techniques benefit the patient, they pose significant challenges—both cognitively and physically—for the surgical team, particularly the surgeon, due to limited visibility and increased physical strain. This, in turn, may jeopardize patient safety [2].

Another example of novel technologies applied in surgery is the use of augmented reality (AR) and virtual reality (VR) simulations. These simulations display 3D models of anatomical structures on a 3D screen, potentially aiding in their examination and analysis. Glasses that project these simulations can be used during the preoperative phase to assist with surgical planning.

To incorporate technologies such as AR and VR applications into everyday clinical practice or the operating room, these technologies must first be evaluated for safety, ergonomics, and particularly for their impact on workplace strain [3]. So far, AR and VR glasses have primarily been assessed for mental workload using subjective questionnaires, such as the National Aeronautics and Space Administration-Task Load Index (NASA-TLX), or by measuring the user's performance quality [4]. This study aims to assess mental workload both objectively, through electrophysiological recordings, and subjectively, using established questionnaires, to facilitate a comparison between the 2. The study is not intended to evaluate or rate the different techniques used.

From a neuroergonomic perspective, replacing subjective questionnaires with objective measurements offers distinct advantages. Subjective questionnaires rely on self-reporting, which can be influenced by various biases. By contrast, objective measurements provide more accurate and reliable data by directly assessing neurophysiological and behavioral responses in real time [5]. By minimizing reliance on subjective questionnaires, objective measurements enhance the scientific validity of neuroergonomic research, resulting in more robust findings and practical implications that could optimize human-machine interactions.

AR and VR glasses enable more accurate planning of invasive procedures through 3D visualization of anatomical structures. This allows the medical team to anticipate potential anomalies and develop a tailored plan for complex surgeries that considers the patient's individual anatomy more precisely than traditional preparation using conventional radiological images. Such technologies are particularly valuable in challenging interventions with limited direct visibility, such as laparoscopic procedures [6]. Additionally, VR and AR can serve as effective training tools for inexperienced surgeons, helping them gain proficiency and improve safety during interventions.

Several studies have subjectively evaluated the reduction of stress and cognitive burden during the use of VR and AR [7-9]. The results suggest that VR and AR can alleviate stress and cognitive load for users. These applications have been used to clarify anatomical conditions and facilitate the learning of practical tasks, thus reducing stress by providing presurgical training. Building on these findings, we hypothesized that in this study, using AR and VR simulations would be less taxing for participants when preparing for surgery, compared with the traditional method of using classical radiological images such as magnetic resonance imaging (MRI), CT, and X-rays.

Ghani et al [10] demonstrated that a passive oddball condition can be used to objectively assess mental workload through participants' electroencephalography (EEG) [10]. By combining subjective questionnaires with the objective evaluation of workload via EEG, we hypothesize that in this study, the objective results on mental workload will align with the subjective data from the NASA-TLX questionnaire [11].

Bernhardt et al [6], Nicolau et al [12], and Dixon et al [13] described significantly improved visualization capabilities through the use of AR and VR, particularly in minimally invasive procedures. Based on these findings, we hypothesize that AR and VR will facilitate mental visualization of anatomical conditions more effectively than traditional 2D images, due to their 3D displays, ultimately leading to better outcomes, such as improved surgical performance.

To test the following 3 hypotheses, a passive oddball experiment with continuous EEG monitoring was conducted with 9 visceral surgeons as participants. They performed mock preoperative planning for the resection of a cancerous liver tumor using AR and VR headsets, as well as traditional radiological images (MRI and computed tomography scans, resolution: 5 mm; arterial and venous phases available) displayed on a standard computer screen.

- Hypothesis 1: AR and VR simulations will reduce mental workload compared with traditional radiological images.
- Hypothesis 2: The mental workload measured objectively using EEG will match with subjective reports from the NASA-TLX questionnaire.
- Hypothesis 3: AR and VR use will improve anatomical visualization and surgical performance compared with traditional 2D imaging.

To successfully test hypothesis 1, we are using 2 real, anonymized surgical cases that have been adapted for display in traditional radiological, VR, and AR environments.

To evaluate hypothesis 2, a state-of-the-art mobile EEG setup is used on the same cases, after which participants are asked to

XSL•FO
RenderX

complete a commonly used, validated questionnaire on workload strain.

To evaluate performance in hypothesis 3, participants were required to assess the resectability of the tumor in each case, considering both its size and location to make an informed decision regarding surgical feasibility.

These hypotheses require both objective and subjective workload assessments, along with a comparative evaluation of task performance under different imaging conditions.

The design of the EEG task follows a standard passive auditory oddball paradigm, in which both a frequent and a rare auditory stimulus are presented. The EEG response to the target stimulus is analyzed, as it is expected to produce 2 distinct potentials: in addition to the early neuronal responses to the auditory stimulus (P1, N1, P2, and N2, collectively referred to as auditory evoked potentials [AEP]), a later response to the oddball modality—that is, the shift from a frequent to a rare stimulus—is expected to manifest as the P3 potential (hereafter referred to as event-related potential [ERP]). While the early AEPs are associated with basic neuronal processing, primarily reflecting the activity of working memory, the later ERPs are linked to more complex, higher-order cognitive processing of the presented stimulus [14].

The degree of these potentials is used to draw conclusions about the objective workload, specifically in terms of the attention required for a primary task. While the auditory stimuli are presented passively in the background, participants focus actively on a primary task—in our case, planning a surgery in different modalities. It has been demonstrated multiple times that EEG potentials are more pronounced when the primary task is less demanding, that is, requiring fewer neuronal resources: when the primary task is not resource-intensive, more mental resources are allocated to the passive background task, leading to higher EEG potentials for both AEPs and ERPs. If the primary task is more complex, fewer mental resources are allocated to the secondary, passive background task, resulting in less pronounced EEG potentials [10,14].

The objective workload in the 3 different conditions will be compared with the subjective workload measured using the NASA-TLX questionnaire. Additionally, the surgeons will be asked to perform simple tasks during the surgical planning process, such as measuring the diameter of lesions, to assess the quality of their work.

Methods

Ethical Considerations

All participants provided informed consent before the study. Participation was voluntary, and no form of compensation was offered. The study was approved by the Medical Ethics Committee (reference number 2020-150), the Data Protection and Information Security Management Unit at the University of Oldenburg, and the staff council at PIUS-Hospital. All participants were recruited through word of mouth by the medical staff of the University of Oldenburg's Department of Visceral Surgery, located at PIUS-Hospital, Oldenburg. The study was conducted in accordance with the Declaration of Helsinki and the guidelines of Good Clinical Practice. Data were treated confidentially at all times. The database, containing anonymized patient data, was stored on a password-protected computer, ensuring access to authorized personnel only. Data were stored in a pseudonymized manner, with each participant assigned a numerical code.

Participants

Nine participants with prior experience in liver surgery were recruited for this study. Although the sample size was relatively small, all participants were carefully selected from a well-defined target group—surgeons with experience in liver surgery—highlighting the study's quality. Inclusion criteria included attending status (or residents in their fourth year of training), signed informed consent, no significant coffee consumption, and no history of neurological damage. Exclusion criteria for participants included excessive caffeine consumption, the use of psychoactive drugs or medications, and any mental or neurological disorders. These criteria were assessed in advance using a standard questionnaire for EEG measurements.

All participants were male, right-handed individuals with an average age of 36.8 (SD 5.59) years. Six of the participating surgeons were attending surgeons (including 2 senior surgeons), and 3 were residents in the later stages of their training. On average, before the study, participants had assisted in an average of 43 (SD 65.63) liver surgeries and had led an average of 16 (SD 29.7) surgeries.

Stimuli

The acoustic oddball stimuli consisted of 2 sinusoidal tone pips, each with a duration of 50 ms and a rise/fall time of 5 ms. The standard tone, with a frequency of 500 Hz, had a probability of 88%, while the rarer target tone, at a frequency of 1 kHz, had a probability of 12%. Both tones were presented via loudspeakers (iLoud MTM; IK Multimedia), connected to an amplifier (the t.amp E4-130; Thomann), at a volume of 75 dB sound pressure level, with an interstimulus interval of 1 second. The tones were randomized with the constraint that a target tone was always followed by a standard tone. The stimuli were similar to those used by Williams et al [15]. The auditory stimuli were digitally created using MATLAB (version R2021a; The MathWorks, Inc.) and presented using Presentation Software (version 23.1; Neurobehavioral Systems).

Conditions

Three different conditions were tested to assess their impact on focused attention during liver surgery planning. These conditions were (1) an AR headset, (2) a VR headset, and (3) a PC with traditional radiological images. The VR setup included the VIVE Pro headset (HTC Corporation), 1 controller, and 2 tracking devices, known as lighthouses. For AR, the Microsoft HoloLens 2 was used (Microsoft Corp.). The software used for visualization in VR was a custom-built setup previously used in our department [16]. For the AR setup, a demonstrator developed as part of this project was used. Traditional radiological images were displayed using the Xero Viewer (Agfa Health care) on a conventional PC. Screenshots of the Xero Viewer software and videos of the VR and AR conditions are

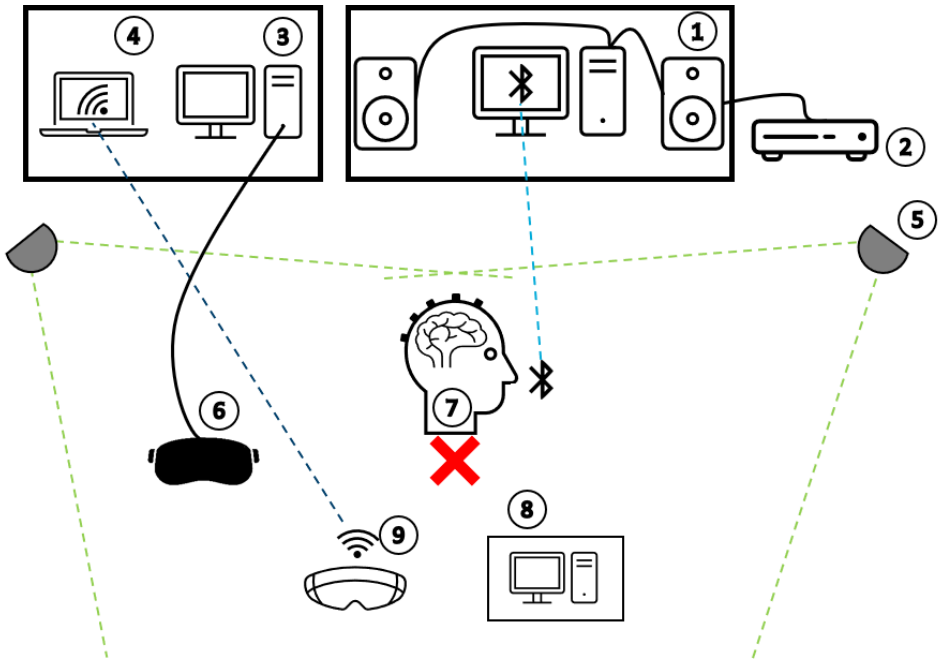
XSL•FO
RenderX

available in Multimedia Appendices 1-4, respectively. In the AR and VR conditions, participants were presented with previously segmented 3D models in combination with radiological images. In the PC condition, participants were only presented with the radiological images. In all 3 conditions, participants were instructed to plan a liver resection using all available functions specific to each condition. In the AR and VR conditions, participants could draw on the 3D model, set markers, and manually measure anatomical structures. In the VR condition, additional functionalities were available, such as volume cropping and the placement of vascular clips. In both the AR and VR conditions, participants could also view or superimpose the corresponding MRI images onto the 3D model. The control mechanisms differed between the 2 conditions due to distinct hardware setups: the VR condition was controlled using dedicated controllers, while the AR condition relied on gesture-based interaction. In all 3 conditions, participants were asked to assess the size and segment of the largest tumor, as well as the resectability of the tumor. The primary goal was to keep participants engaged for at least 3 minutes in each condition, using a basic task (ie, diagnosing based on the images). This task was consistent across all conditions, regardless of the specific interface or available functionalities. Evaluating the usability of the respective software and its functionalities was beyond the scope of this study, particularly because both the AR and VR setups were custom-built and still under development.

Experimental Setup

A mobile EEG amplifier (Smarting mobile; mbt) was connected to a standard 10/20 EEG cap (EasyCap GmbH) and recorded using software (Smarting Streamer v3.4.2; mbt) on a desktop computer running Windows 10 (Microsoft Corp.). The 2 speakers were placed next to the recording computer, with participants positioned 2 m in front of the speakers. During the experiment, care was taken to position the participants consistently in the same marked location across all 3 conditions to ensure that the stimuli were presented at the same sound level in each condition. A laptop was used to connect wirelessly to the AR glasses via Miracast, allowing the instructor to monitor what the participants were doing and seeing. The laptop was placed out of sight of the participants to avoid distractions. A second desktop computer was used to run the VR setup, with the VR headset connected to this computer and the 2 VR lighthouses mounted on tripods, enabling the participants to move freely while wearing the VR headset. To ensure that participants remained in the same position during all 3 conditions, the traditional radiological images presented in the PC condition were displayed on a computer connected to the hospital information system. This computer was placed on a mobile, height-adjustable table, with the screen positioned at eye level. This setup was designed to prevent the EEG cap from slipping due to excessive head movement and to allow participants to perform all conditions while standing (Figure 1). Controllers were used to operate the VR condition and were placed in the participants' hands. In the AR condition, control was achieved through gesture-based interaction using the participants' hands. No standardized training was provided before the experiment. However, if the conditions were unfamiliar to the participants or if they had any questions, a test run was conducted in which they could familiarize themselves with the controls and functions until they felt confident in handling the technology.

Figure 1. Experimental setup. 1: main computer for recording electroencephalography (EEG); 2: amplifier; 3: computer virtual reality (VR); 4: computer augmented reality (AR); 5: lighthouses for the VR headset; 6: VR headset; 7: EEG on marked position; 8: computer for conventional surgical planning; and 9: AR headset.

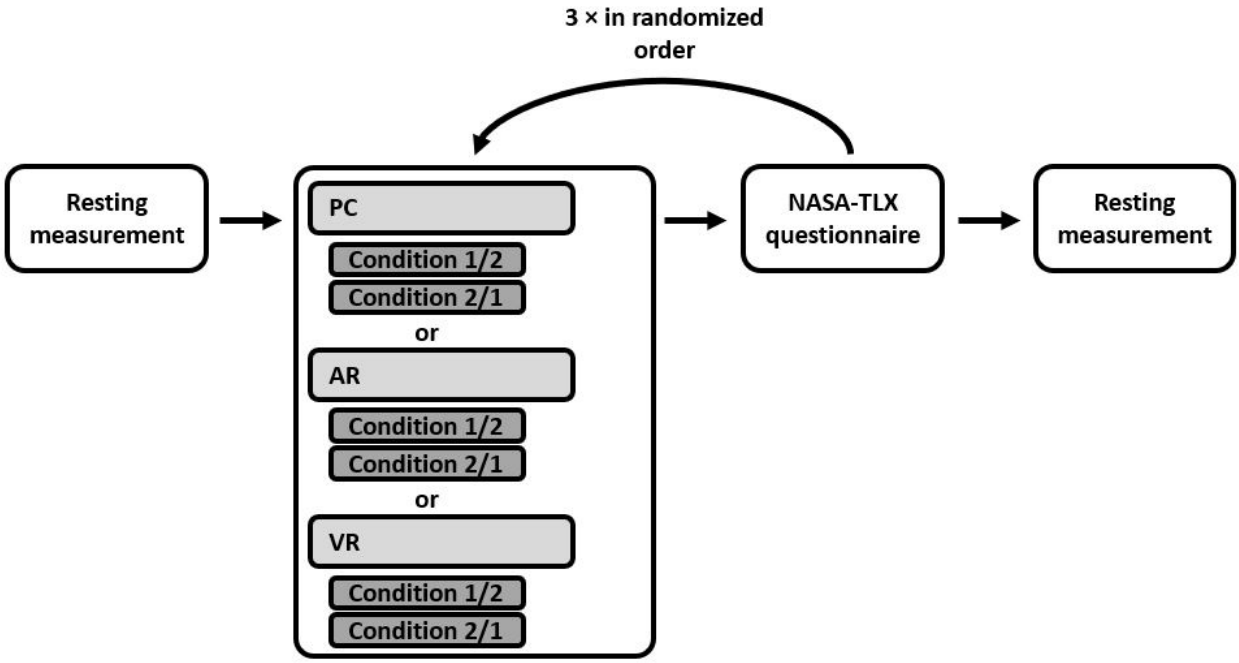


XSL•FO
RenderX

Procedure

Mobile EEG recordings (sampling rate: 250 Hz, Smarting Mobi; mbt) were obtained using a standard EasyCap EEG recording cap with 24 electrodes: Fp1, Fp2, AFz, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, CPz, M1, M2, P7, P3, Pz, P4, P8, POz, O1, and O2. The reference electrode was placed at a fronto-central position. Continuous EEG data were stored in .xdf format before being processed for analysis. For AEP analysis, EEG data were analyzed using EEGlab (V2020.0, open-source toolbox for MATLAB R2020a; MathWorks) on a standard PC running Windows 10 (Microsoft Corp.). For ERP analysis, EEG data were first converted to Brain Vision Analyzer format using MATLAB and then analyzed using Brain Vision Analyzer Software (version 2.2.1.8266, Professional Edition; Brain Products GmbH). EEG electrode impedances were adjusted to below 10 kΩ using 70% rubbing alcohol and abrasive conductive gel (EasyCap Abalyte Hi-Cl), following the manufacturer's recommended protocol. After the electrodes were properly placed, 2 minutes of resting-state EEG was recorded with the eyes open and eyes closed before actual testing began. For the testing, 1 of the 3 conditions—"AR," "VR," or "radiological images"—was selected in randomized order. In each of the 3 conditions, participants were required to plan 2 different surgeries, which included measuring the dimensions of the tumors and identifying their anatomical locations within the liver to determine whether the tumors were resectable. The 2 patient cases were presented in randomized order (see Figure 2 for the experimental procedure). By the end of the session, each participant had completed 6 sets of surgical planning (2 cases for each of the 3 conditions). The task for the participants was to plan a surgical procedure to remove cancerous tissue from the liver. The processing time for each condition (AR, VR, or radiology) was 10 minutes, with 5 minutes allocated to each patient case. Throughout the task, auditory stimuli were presented to the participants via 2 speakers in the background. Before switching to the next condition, the participants' subjective workload was assessed using the NASA-TLX questionnaire.

Figure 2. Experimental procedure flowchart: randomized order of conditions. AR: augmented reality; NASA-TLX: National Aeronautics and Space Administration-Task Load Index; VR: virtual reality.



NASA-TLX

The NASA-TLX workload questionnaire has been validated as an effective measure of subjective workload in medical work environments [17].

The NASA-TLX is a widely used tool for assessing subjective workload. Originally developed for aviation settings by NASA, it has since been adapted for use in various fields, including health care [17]. The NASA-TLX measures 6 dimensions of workload: mental demand, physical demand, temporal demand, effort, frustration, and perceived performance. Each dimension is rated on a scale from 0 (low) to 100 (high) [18].

Data Preprocessing of EEG Data

To retain the frequency components relevant for EEG analysis, the recordings were filtered using a bandpass filter (low cutoff: 0.05 Hz, high cutoff: 30 Hz, and filter order: 4). A notch filter was also applied at 50 Hz to remove power line interference. The continuous EEG data were divided into 1000 ms epochs, with 200 ms before and 800 ms after the stimulus. The data were then categorized into 2 categories: standard stimulus (nontarget: 0.5 kHz) and target stimulus (1 kHz). Automatic artifact rejection was performed by discarding all epochs where fluctuations greater than +100 µV or –100 µV occurred within 100 ms. Additionally, all epochs underwent manual visual inspection, focusing on signals from electrodes FP1 and FP2 to identify strong artifacts due to eye movements. Averages were

XSL•FO
RenderX

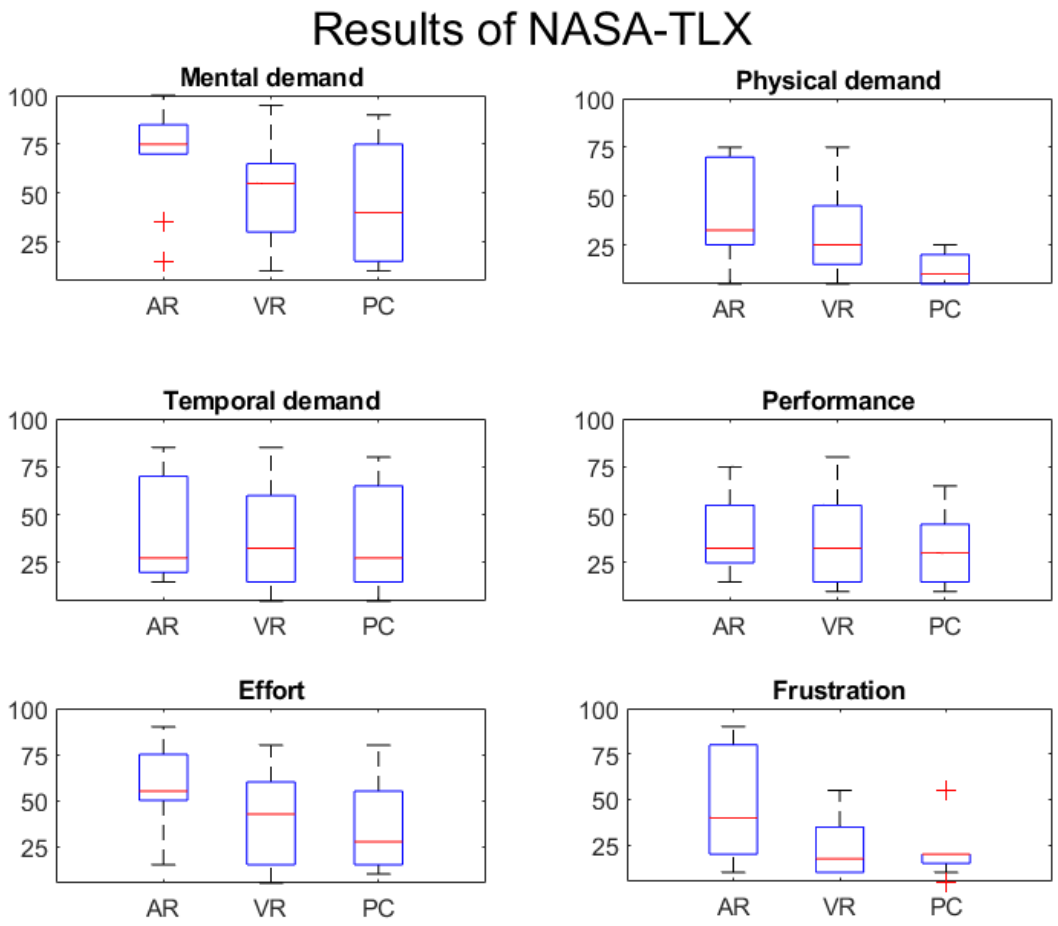
then calculated for the epochs recorded during the presentation of the target stimulus in each of the 3 conditions: 1 average for the EEG response to the target stimulus during AR, 1 for the EEG response during VR, and 1 for the EEG response during surgical planning using traditional radiological images. Averages were calculated for each participant to facilitate further statistical analyses. Baseline correction was applied to prevent potential linear drifts in the recorded signals and to ensure that the grand averages across participants and conditions were comparable. For AEP analysis, the data were re-referenced to electrodes T5 and T6, while for ERP analysis, the data remained referenced to the physical fronto-central reference.

Results

Subjective Workload

The median physical demand score for the AR condition was 30 (IQR 25-70) on the NASA-TLX scale, higher than the scores for the VR (median 25, IQR 15-45) and PC (median 10, IQR 5-20) conditions. For the perceived performance and temporal demand categories, all 3 conditions yielded the same NASA-TLX scores: 25 for temporal demand and 30 for perceived performance. Additionally, for the mental demand, effort, and frustration categories, the AR condition showed the highest median scores: 75 (IQR 70-85) for mental demand, 55 (IQR 50-75) for effort, and 40 (20-80) for frustration (Figure 3).

Figure 3. National Aeronautics and Space Administration-Task Load Index (NASA-TLX) scores. AR: augmented reality; VR: virtual reality.



Examination of the box plots reveals that the mental demand score for the AR headset (median 75, IQR 70-85) was higher than those for both the VR headset (median 55, IQR 30-65) and the PC (median 40, IQR 15-75) conditions. Similarly, higher scores for effort and frustration were observed in the NASA-TLX evaluations when using the AR headset (median 55, IQR 50-75 for effort; and median 40, IQR 20-80 for frustration) compared with using the VR headset (median 40, IQR 15-60 for effort; and median 15, IQR 10-35 for frustration) or the PC condition (median 25, IQR 15-55 for effort; and median 20, IQR 15-20 for frustration). In terms of physical demand, the score was lower in the PC condition (median 10, IQR 5-20) compared with the AR (median 30, IQR 25-70) or VR (median 25, IQR 15-45) condition.

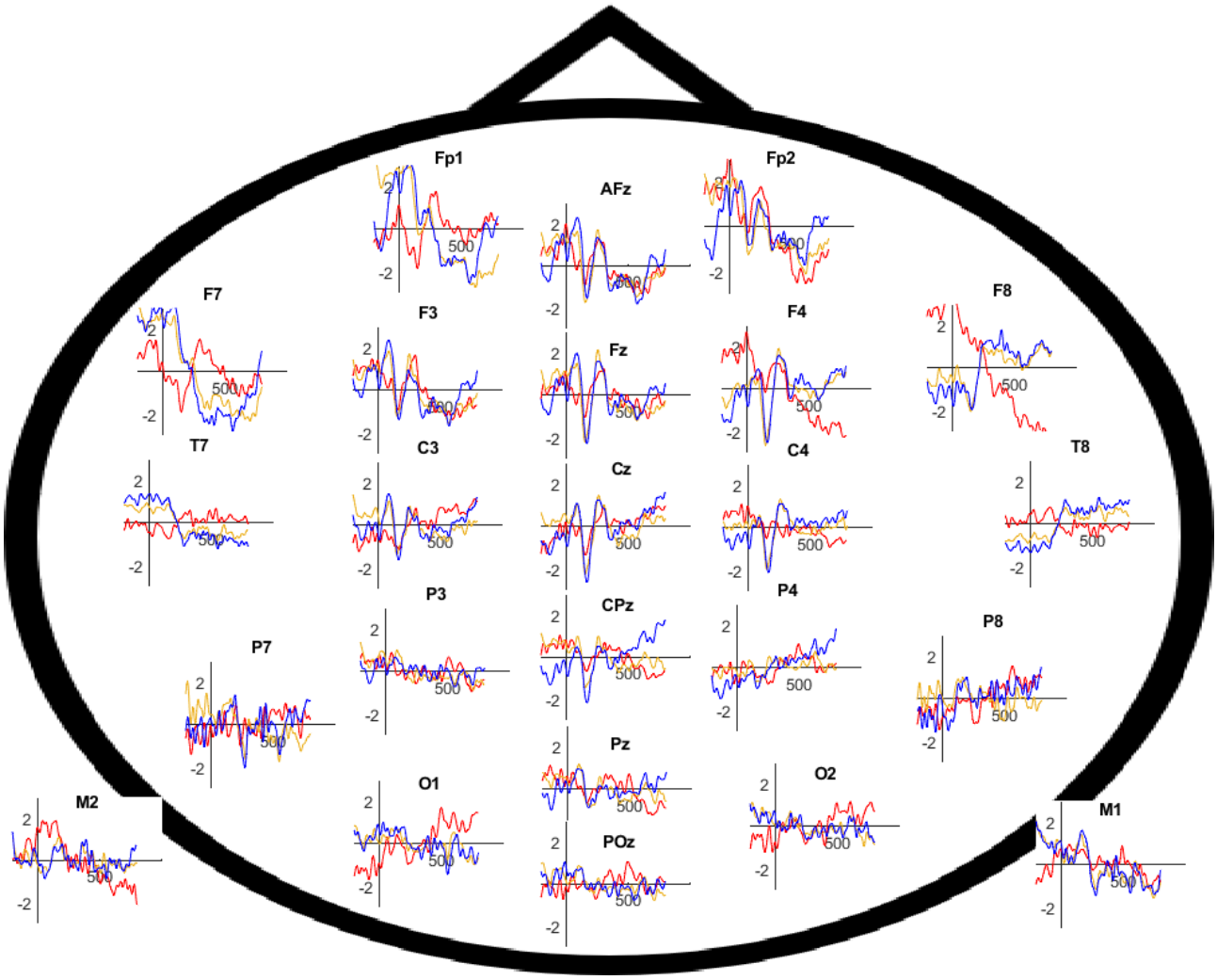
A Friedman test was performed to assess significance (significance level of .05), as the data did not follow a normal distribution according to the Kolmogorov-Smirnov test. To account for multiple comparisons, the Benjamini-Hochberg method was applied for correction using a web-based calculator [19]. Assuming a 95% CI, no significant differences (P=.10) were found.

XSL•FO
RenderX

Electrophysiological Recordings—Early Components

AEPs were primarily observed on the electrodes along the central row and midline (AFz to Pz, F3/C3, and F4/P4). Grand averages were calculated for all electrodes (Figure 4).

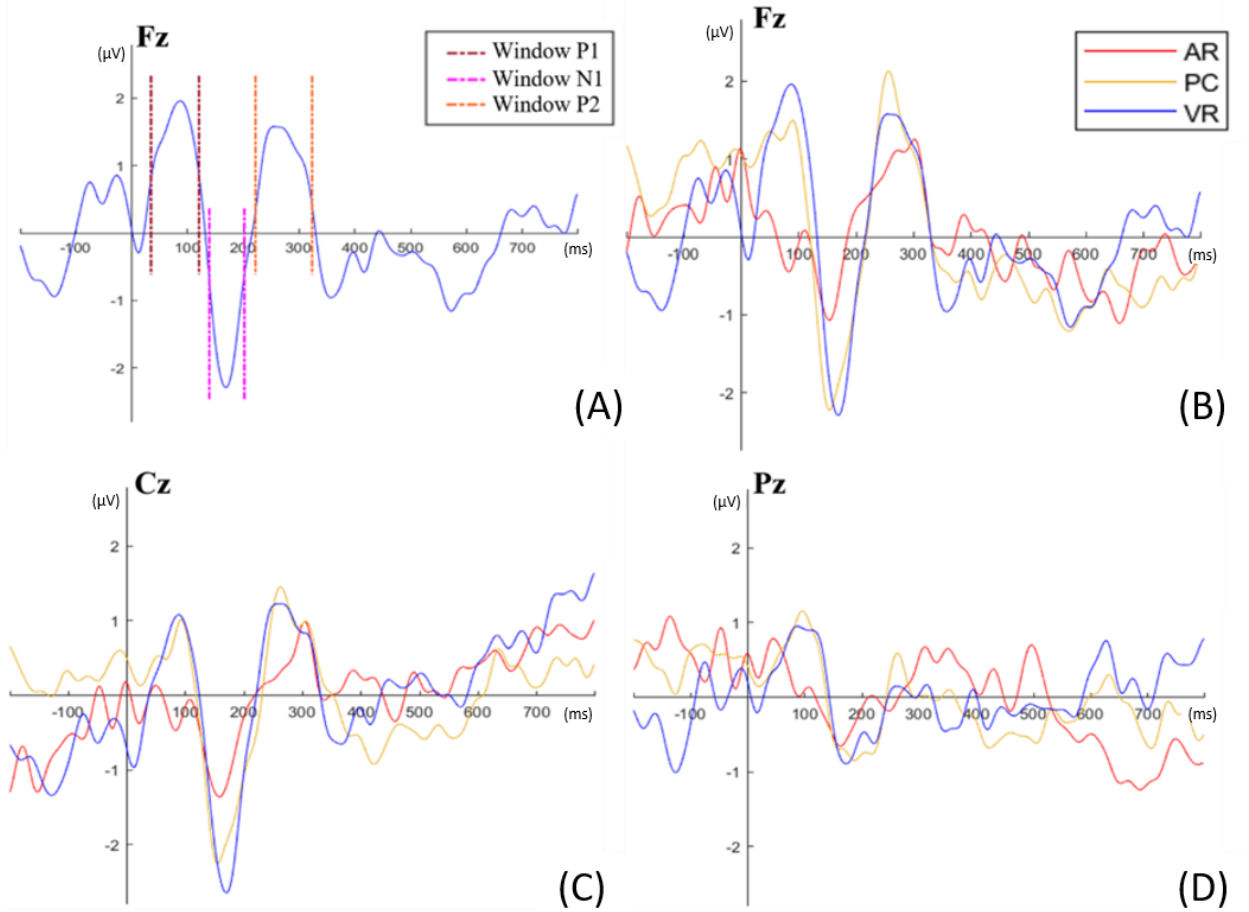
Figure 4. Event-related potentials as response to the target tone for all electrodes and all 3 conditions.



To statistically analyze the amplitudes for each participant, the peaks and their corresponding latencies were identified for all conditions at the central electrodes Fz, Cz, and Pz (Figure 5B-D). To statistically verify differences in the amplitudes of the grand averages, the approach outlined by Handy [20] was used as a guide. The intervals for P1, N1, and P2 were marked accordingly (Figure 5A). The resulting peaks and latencies for each participant were then analyzed using a repeated measures analysis of variance (rmANOVA) in SPSS (version 28.0.1.0; IBM Corp.) with the factors patient case, condition, electrode, and ERP components. A significant difference was observed between the ERP components ($F_{1.718,13.743}$=46.754, P<.001).

XSL•FO
RenderX

Figure 5. Grand averages of event-related potentials as a response to the target tone for statistically analyzed electrodes under all 3 conditions: (A) temporal windows for the event-related components P1, N1, and P2, exemplarily shown for the electrode position Fz; (B) grand averages of the electrode Fz; (C) grand averages of the electrode Cz; (D) grand averages of the electrode Pz. AR: augmented reality; VR: virtual reality.



The tendencies observed in the grand averages were not supported by the statistical analysis, as no significant differences were found between patient cases (P=.57), conditions (P=.71), or electrodes (P=.49). To determine the latencies of the AEP components (P1, N1, and P2), the method described by Handy [20] was also applied. The latency of the identified peaks within the defined intervals was measured and recorded for each participant. Subsequently, an rmANOVA was conducted with the factors patient case, condition, electrode, and ERP components to identify significant differences in the temporal occurrence of these components.

The tests of within-subject effects revealed a significant interaction among condition, electrode, and ERP components ($F_{3.401,27.211}$=6.747, P=.001). The results of the post hoc pairwise comparisons are presented in Table 1. According to these comparisons, the P1 component at the Fz electrode occurs significantly (P=.01) earlier in the VR condition (mean 75.3 ms, SD 25.8 ms) compared with when using traditional radiological images (mean 99.4 ms, SD 28.6 ms). The Cz electrode also demonstrated a significant (P=.04) difference in the P1 component between the VR condition (mean 111.1 ms, SD 9.9 ms) and the AR condition (mean 80.4 ms, SD 32.1 ms). Additionally, the P2 component in the AR condition (mean 290.6 ms, SD 23.5 ms) occurred significantly (P=.03) later than in the traditional radiological images condition (mean 265.3 ms, SD 13.1 ms). Significant differences were also observed for the P1 component at the Pz electrode, with the VR condition (mean 62.8 ms, SD 24.8 ms) occurring significantly earlier (P=.04) than both the AR condition (mean 89.1 ms, SD 14.6 ms) and the traditional radiological images condition (mean 92.8 ms, SD 28.4 ms; P=.04). No significant (P=.12) interactions were found for the N2 component.

XSL•FO
RenderX

Table 1. Significant results of the post hoc tests (pairwise comparisons).

Electrode and ERP[a] component	Condition (I)	Condition (J)	Latency (ms), mean		Difference between means (I–J)	*P* value
			Condition (I)	Condition (J)		
Fz						
P1	VR[b]	PC[c]	75.3	99.4	–24.1	.01
Cz						
P1	VR	AR[d]	111.11	80.4	30.71	.04
P2	AR	PC	280.7	265.3	25.4	.03
Pz						
P1	VR	AR	62.9	89.1	–26.2	.04
		PC	62.9	92.9	–30	.04

[a]ERP: event-related potential.

[b]VR: surgery planning using the virtual reality setup.

[c]PC: surgery planning using traditional radiological images on a desktop computer.

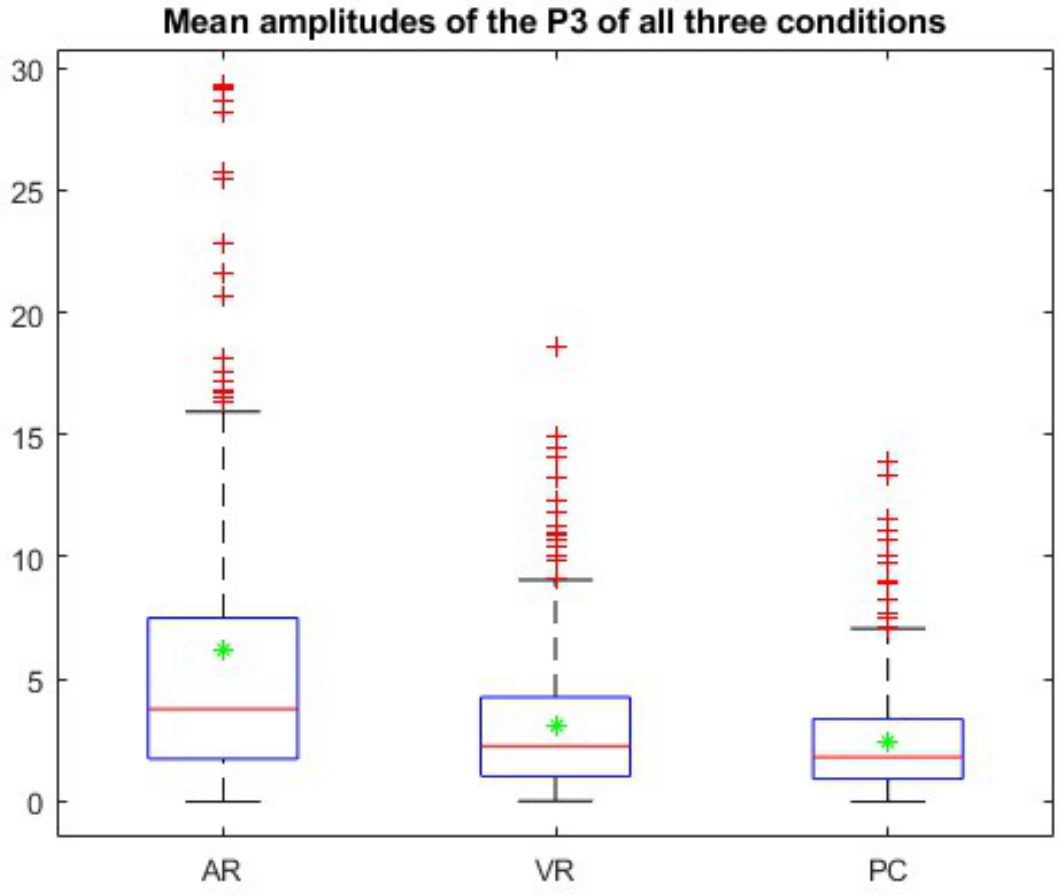
[d]AR: surgery planning using the augmented reality setup.

Electrophysiological Recordings (P3)

For the P3 component of the ERP, peak amplitudes within the time range of interest (300-400 ms poststimulus) were calculated. A grand average for all 3 conditions (radiology, VR, and AR) was computed using data from all electrodes (Figure 6). A clear trend was observed, with the classical radiological approach eliciting the lowest P3 amplitude (mean 2.45 µV, SD 2.26 µV), VR eliciting the second largest P3 amplitude (mean 3.08 µV, SD 2.80 µV), and the AR condition eliciting the highest P3 amplitude (mean 6.20 µV, SD 6.84 µV). SDs were high, and no statistical significance was found using an rmANOVA in SPSS (version 28.0.1.0) with the factors patient case, condition, and P3 amplitude ($F_{1.94,10.58}$=1.469, P=.26).

XSL•FO
RenderX

Figure 6. Amplitude of the P3 component of the event-related potential across all 3 conditions: the augmented reality (AR) approach (left), which showed the highest P3 amplitude (mean 6.20 µV, SD 6.84 µV); the virtual reality (VR) condition (center), with a mean amplitude of 3.08 µV (SD 2.80 µV); and the classical approach (right), which exhibited the lowest P3 amplitude (mean 2.45 µV, SD 2.26 µV). No statistically significant differences were found ($F_{(1.94,10.58)}$=1.469; P=.26).



Post Hoc Power Analysis

Given the small sample size (N=9), a post hoc power analysis was conducted to determine the effect size and assess whether a larger sample might yield different results. Partial eta-squared (η^2=0.114) was calculated for the interaction between experimental conditions using SPSS. To estimate the sample size required for significant results, a web-based calculator was utilized [19], indicating that a sample size of 101 participants would be needed to achieve a statistical power of 0.8 at a significance level of α=.05.

Correlation

A correlation analysis was conducted between the N1 component, which demonstrates a decreased amplitude with increased cognitive workload [10], and the "Mental Demand" and "Effort" dimensions of the NASA-TLX questionnaire. The results revealed a correlation coefficient of r=–0.0042 between the N1 component and "Mental Demand," and r=0.066 between the N1 component and "Effort."

Performance

All participants correctly identified the resectability of the tumors in all conditions and for every patient case, based on their accurate assessment of tumor size and localization.

Discussion

Findings and Implications

The aim of this study was to establish a method for objectively determining the additional workload generated using AR or VR glasses in a clinical context for the first time. For this purpose, 2 runs of planning a liver surgery were performed under each of the 3 conditions. In the background, a passive oddball experiment was conducted to produce clear early AEP components (P1, N1, and P2) as well as the P3 component of the ERP in the EEG recordings, guided by the work of Ghani et al [10].

Subjective Workload

NASA-TLX scores indicate a tendency toward higher scores, suggesting a greater subjective workload for AR planning, particularly in terms of "mental demands" and "frustration." However, this trend does not reach statistical significance when compared with VR or the traditional approach using radiological images, likely due to the small sample size. Moreover, the large variances observed in the data could have contributed to the lack of significant findings, as they indicate substantial variability in subjective workload perceptions among participants. The observed tendencies may be attributed to the

XSL•FO
RenderX

more complex and unfamiliar gesture controls of the AR headset. In addition, a poor internet connection may have resulted in higher latencies, potentially affecting gesture control and thus leading to higher workloads when using the AR demonstrator.

Higher physical demands were expected in the AR and VR setups compared with the traditional approach. A tendency was observed in the subjective ratings, as using a stationary mouse on a desktop PC required significantly less physical effort compared with using an AR or VR headset.

Objective Workload

The expectation that using an AR or VR setup would generally make planning surgery easier due to its 3D projection of the anatomical situation was not confirmed.

EEG data indicate a trend with more pronounced early AEPs when using VR and classical radiological images compared with AR, suggesting that AR results in the highest workload for the surgeon, while VR does not increase workload compared with classical radiological images. Differences are especially visible in the ERPs P1 and N1 components: no P1 component was found for the AR condition, whereas the N1 component showed a significant difference for the AR condition compared with radiological images and VR on the level of grand averages. This aligns with the finding of Ghani et al [10], indicating that the N1 component becomes less pronounced with increased task difficulty due to the more complicated gesture control of the AR headset and limited opportunity to familiarize with the gestures required to handle 3D objects in AR.

We would like to emphasize that the significance at the electrode Pz should be treated with caution, as the amplitudes here are generally lower. The results for the electrode Pz are clearly noisier, likely due to its position being obstructed by the fastening of the glasses, which are attached to the head near the position of electrode Pz. These results should therefore be interpreted with caution.

When examining the significances of the latencies of the different early AEP components, it is noticeable that there are occasional significances that do not follow a coherent pattern.

Concerning the event-related P3 wave in response to the target stimulus of the passive oddball paradigm, the pattern found differs from that seen in the early potentials: a trend from a low P3 in the radiology condition to a high P3 in the AR condition can be observed. The studies conducted by Thees et al [7], Küçük et al [8], and Chao et al [9], which asserted that AR and VR headsets result in reduced cognitive load, can thus only be partly substantiated in this investigation.

Based on our data, we conclude that the use of the passive oddball paradigm is a suitable tool to measure the objective workload between the radiology, AR, and VR setups in planning a complex surgery. There appear to be measurable differences in the neuronal processing of 2D screen images when compared with VR and AR images. However, the answer to why the different modalities of EEG components—early AEP versus ERP—yielded different findings is complex and likely rooted in distinct, complex neuronal and psychological processes responsible for producing these potentials. To better understand the different psychological processes behind the processing of 2D screen images versus VR and AR images, in addition to increasing the number of participants, a more in-depth analysis of the P3 potential is needed. In our analysis, we only looked at the grand average of the peak amplitude. However, it is also important to investigate the location of the most pronounced P3 potentials, as well as the dynamics—that is, the distribution of P3 peak latencies across different electrode localizations and how this distribution changes over time. Current research indicates that the P3 potential arises as a combination of different subcomponents, P3a and P3b, which may be involved in distinct neuropsychological processes. This differentiation needs to be addressed in further studies.

A simple but feasible explanation for the different findings between early AEPs and P3-ERPs could be attributed to the different mental processes involved: while the complex cognitive processing of AR pictures, represented by the P3 potential, might be less demanding for the user, aligning with findings from Thees et al [7], Küçük et al [8], and Chao et al [9], the difficult or less familiar navigation in the AR—and to a lesser extent, VR—environment may lead to higher demands on basic neuronal tasks, such as visuomotor coordination, which are more strongly represented by early neurological potentials than by late cognitive potentials. While this remains speculative at this point, we believe that our study should be regarded as a proof-of-principle, demonstrating a foundation for further research with a larger sample size and in more complex settings. This will allow for a deeper investigation into this issue.

The lack of a significant correlation between the N1 component and the subjective measures of mental demand and effort is likely influenced by the extremely small sample size in this study. However, this result also suggests that cognitive workload, as measured by physiological markers such as EEG, may not necessarily align with the subjectively perceived workload. This finding underscores the importance of using objective measures in assessing workplace stress and cognitive load. Relying solely on subjective assessments may not fully capture the cognitive demands placed on individuals, highlighting the need for more robust and comprehensive approaches in future research.

The use of a mobile EEG was particularly useful in this study. The high degree of mobility provided by the use of portable smartphones as recording devices allowed for recordings to be made in situations closely aligned with the operation without restricting participants' freedom of action. The validity of corresponding mobile EEGs has been previously described by Wascher et al [21]. Moreover, the work of El Basbasse et al [22], among others, has confirmed that reliable data can be derived using mobile EEG recordings, including the implementation of the EasyCap system in conjunction with VR headsets [21,22]. The results of this study support this research. Additionally, our findings demonstrate that the study design, when combined with mobile EEG technology, enables the investigation of the additional workload induced by various medical and clinical devices. Notably, even head-worn devices cannot be unequivocally excluded from this measurement approach.

XSL•FO
RenderX

The AR setup was perceived and measured as the most difficult of the 3 setups. As mentioned, this may be due to the complexity of its controls. While the classical approach was well-known to the participants, controlling the VR setup appeared to be more intuitive than using the AR setup. However, it is crucial to consider that AR headsets represent a relatively nascent technology, and gesture recognition is still in its developmental stages, with anticipated advancements in future iterations of AR headsets. It appeared that participants were overwhelmed by handling the headset without a controller or corresponding control element, as gesture control was a new experience for most of them. This difficulty may be overcome by more accurate training of users before working with the AR setup. We believe that our work could be used as a tool to assess how well users are trained in using the AR setup and may be used to generally investigate the need for training when navigating in an AR and VR environment, even beyond its medical applications. Additionally, more complex, subconscious explanations could be considered: The VR approach might benefit from the possibility of exclusively displaying the anatomical model, while the AR setup might strain participants' mental capabilities more due to the simultaneous display of the computer-generated 3D model along with the real-world environment. Furthermore, factors such as the different displays of the HoloLens and the VIVE Pro could contribute to the differences in participants' workload. However, this requires more detailed investigations. It must be considered in this discussion that the 2D viewing habits trained over decades are completely abandoned. New 3D viewing habits must be relearned in AR and VR, along with the integration of voice and gesture control.

This study makes a significant contribution to the existing body of research by specifically focusing on the mental workload of medical professionals. Unlike the studies listed in Table 2, which mainly explore cognitive load in nonmedical settings or involve participants lacking domain-specific expertise, our research is uniquely positioned within the context of clinical practice. By using objective EEG measurements, we were able to assess workplace strain in real-world scenarios, providing insights into the cognitive demands experienced by surgeons when using AR and VR technologies. This focus on medical professionals, coupled with the use of advanced neuroergonomic methods, highlights the practical relevance and novelty of our findings. Our study not only expands the understanding of how AR and VR technologies impact mental workload in high-stakes environments but also provides valuable data that could inform the optimization of these technologies for clinical use.

XSL•FO
RenderX

Table 2. Comparison of recent research studies investigating mental workload using EEG[a] measurements in AR[b] and VR[c] applications.

Study	Population	EEG application	Key findings
Alessa et al [23]	28 Healthy males (mean age 32.12 years)	EEG power analysis	• AR instructions reduced task time compared with paper instructions, particularly for high-demand tasks. • AR increased mental workload and information processing (germane load).
Binias et al [24]	Pilots in short-haul flight simulator sessions	EEG power analysis	• Positive correlation between theta power (frontal lobe) and reaction time. • Significant event-related changes in the frontal lobe for theta and beta waves.
Ji et al [25]	Pilots in VR-simulated missions	Focusing on β rhythms	• β rhythm changes correlated with high-difficulty tasks. • A significant relationship between mental workload and β rhythms.
Aksoy et al [26]	General population	Visual event–related potentials such as N1, P1, and P3	• ERP[d] waveforms were similar between VR HMD[e] and CS[f] • P3 amplitudes were comparable between VR HMD and CS. • N1 amplitudes were significantly higher in VR HMD compared with CS. • P1 amplitudes were higher at the occipital region in VR HMD.
Giorgi et al [27]	General population	Global field power, alpha, and theta	• A higher Workload Index was observed in the museum compared with VR. • Overall, the museum showed higher engagement potential, while VR potentially supported greater embodiment.
Marucci et al [28]	General population	P300 potentials with respect to the unimodal (visual) and visual and tactile bimodal stimulation	• Multisensory stimuli significantly improved performance and reduced EEG-based workload only in high load conditions. • Multimodal stimulation (visual-audio-tactile) decreased latency and increased P300 amplitude, indicating faster and more effective stimulus processing.
Atici-Ulusu et al [29]	Employees on an automobile assembly line	EEG and NASA-TLX were used to measure cognitive load	• Cognitive load (measured by EEG) was lower with AR glasses. • NASA-TLX scores decreased by 10% when using AR glasses, indicating reduced cognitive load. • No significant difference in cognitive load between the first and last days of AR glasses usage.
Trupti et al [30]	General population	EEG signals were analyzed using autoregressive modeling, fuzzy c-means clustering, hidden Markov model, and deep recurrent neural network classification	• The proposed method for classifying cognitive workload using EEG signals achieved 97.8% accuracy. • Handcrafted features combined with temporal dynamics modeling outperformed state-of-the-art methods. • The optimal performance was achieved with 4 electrodes.
Luque et al [31]	General population	Event-related potentials such as P3 and contingent negative variation	• VR effectively mimicked real-world cognitive load and decision-making processes. • Significant ERPs related to attention and decision-making were elicited in VR.

XSL•FO
RenderX

Study	Population	EEG application	Key findings
Mondellini et al [32]	General population	Calculated the MWL index[h]	• Different task complexity levels significantly affected MWL, increasing subjective measures and decreasing performance. • HMD generally resulted in lower EEG-derived MWL, indicating reduced cognitive load for some tasks.

[a]EEG: electroencephalography.

[b]AR: augmented reality.

[c]VR: virtual reality.

[d]ERP: event-related potential.

[e]HMD: head-mounted display.

[f]CS: computer screen.

[g]NASA-TLX: National Aeronautics and Space Administration-Task Load Index.

[h]MWL: mental workload.

To provide an outlook beyond the scope of this study, we believe that AR and VR can be very useful technologies in the clinical context when used for the appropriate tasks. For instance, VR could be more suitable in a teaching setting or for preoperative planning, whereas AR holds immense potential beyond preoperative planning, particularly in the operating room. Its ability to provide an open view of the real environment while maintaining sterility allows for virtual overlays of anatomical structures on a patient's body during surgery, enhancing visualization and accuracy. Whether the significant benefits of these technologies outweigh the potential additional workload for surgeons should be further investigated in future studies.

Furthermore, investigating the impact of training, the continuous development of software and hardware, and the increasing familiarity of the younger generation of physicians with these new technologies on measures of workload would be of great interest. Such research could provide valuable insights into the potential benefits and challenges posed by integrating advanced technologies such as AR and VR in medical practice. Further exploration in this area may also facilitate the design of tailored training programs and technological interventions to optimize the utilization of these innovations and improve overall health care outcomes.

Limitations

Concerning our hypothesis that the use of AR or VR might improve surgical performance, no conclusive results were found. Basic tasks, such as deciding whether a tumor is resectable or not, were performed with the same quality across all experimental conditions. This might indicate that AR and VR technology does not inherently lead to better performance in trained surgeons for simple tasks, but this needs to be evaluated further. Future studies should include more specialized criteria for evaluating performance, particularly to assess finer aspects of clinical decision-making in AR and VR environments. This could involve more complex cases or tasks that better reflect the challenges faced in surgical planning and execution, allowing for a deeper understanding of how these technologies impact surgical performance and outcomes. This was beyond the aim of this study. However, based on our findings, we suggest that increased performance might be found in a more challenging task, but not concerning simple anatomical measurements.

On a more general outlook, one might speculate that increased workload over a prolonged period might eventually lead to increased fatigue, potentially decreasing a surgeon's performance after an extended period. This is purely speculative and needs to be investigated in detail through further studies, possibly utilizing our suggested approach for workload measurement—correlating electrophysiological and performance results.

This study focused solely on the workload of the 3 conditions. It should be made clear that there are scenarios where a higher workload is acceptable when weighed against the benefits that such an approach would yield. While the VR setup might be suitable for surgical planning due to its lack of visibility into the real environment, the AR setup may be the more suitable approach during surgery—such as projecting radiological images as 3D models directly into the patient's situs during a procedure to localize, for example, tumors or blood vessels. In these situations, however, technologies such as AR glasses can offer significant advantages due to their see-through capabilities, allowing a slightly higher workload, as observed in this study, to be acceptable if parameters such as patient safety are enhanced through the use of such technologies.

Interpretation of our results should be conducted with caution due to the very low number of participants (N=9), resulting in limited statistical power. This work should be seen as a feasibility study on how to measure objective workload while using AR or VR in a surgical simulation setting. However, we believe the study still provides valuable data, as experienced surgeons in a clinical setting used real-life patient cases with state-of-the-art equipment for AR and VR presentation.

Conclusions

We aimed at developing a possible setup for objective measurement of workplace strain when using AR and VR technologies in the medical profession. We conclude that the use of a passive oddball paradigm is a feasible approach for this objective. The measurement of workload during surgical planning, and thereby likely also during other medical

XSL•FO
RenderX

applications when using AR or VR technology, can be sufficiently done using an oddball paradigm. This approach proves to be a good opportunity to objectively determine workplace strain in a medical setting.

In future studies, the use of this objective measurement could potentially replace subjective questionnaires and significantly enhance the significance of the measured workplace strain. This approach, with the objectivity provided, may even become a commercially available application for medical education and surgical training, utilizing automated EEG analysis. Furthermore, we consider that AR or VR approaches to displaying radiological images are not inherently superior to the classical approach. The use of AR technology, in particular, seems to require a certain amount of training, likely due to the novelty of gesture control. However, we are confident that with future developments in AR technology and gesture recognition, these issues will become less significant over time. Regardless, the differences in workload between AR and VR highlight the need for intuitive user interaction and ease of use for any new technology, especially when applied in a medical setting.

Additionally, integrating machine learning techniques could indeed enhance the analysis of physiological data, allowing for the development of predictive models that may generalize across different tasks and populations. Such approaches could provide deeper insights into cognitive workload and potentially lead to the creation of real-time monitoring tools to optimize performance in high-stakes environments such as surgery. This could eventually result in commercially available solutions for use in medical education or surgical training.

In conclusion, this paper demonstrates the feasibility of utilizing a mobile EEG setup in conjunction with a passive oddball paradigm to assess the potential additional cognitive load imposed by emerging technologies such as AR and VR headsets in a clinical setting.

Acknowledgments

We express our gratitude to Kilian Gloy for his contributions to the planning of the conducted study. His expertise and dedication have played a significant role in shaping the research design and ensuring its successful implementation. Our thanks also go to Fraunhofer MEVIS, Bremen, Germany, namely, Andrea Schenk and Christian Schumann, for providing us with segmented liver models used in this study. This study was conducted within the scope of the VIVATOP project and has thus received funding from the German Ministry of Education and Research (BMBF) in the grant program “Gesundes Leben” (healthy living; grant 16SV8079).

Data Availability

The data presented in this study are available on request from the corresponding author.

Conflicts of Interest

DS was employed in the involved research group at the University Clinic for Visceral Surgery at the University of Oldenburg during the project, and subsequently transitioned to apoQlar (Hamburg, Germany), a company also involved in the project for the development of the AR demonstrator. Other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Screenshot of the PC condition.
[PNG File , 201 KB - formative_v9i1e62740_app1.png]

Multimedia Appendix 2
Video part 1 of the virtual reality condition.
[MP4 File (MP4 Video), 8008 KB - formative_v9i1e62740_app2.mp4]

Multimedia Appendix 3
Video part 2 of the virtual reality condition.
[MP4 File (MP4 Video), 5818 KB - formative_v9i1e62740_app3.mp4]

Multimedia Appendix 4
Video of the augmented reality condition.
[MP4 File (MP4 Video), 143163 KB - formative_v9i1e62740_app4.mp4]

References

1. Dick J, Gebhard F. Zukunft der intraoperativen Bildgebung und Navigation. OP-JOURNAL 2015 Apr 23;30(03):188-195. [doi: 10.1055/s-0034-1383340]

XSL•FO
RenderX

2. Bogdanova R, Boulanger P, Zheng B. Depth perception of surgeons in minimally invasive surgery. Surg Innov 2016 Oct 09;23(5):515-524. [doi: 10.1177/1553350616639141] [Medline: 27009686]
3. Matern U, Büchel D. Gebrauchstauglichkeit von Medizinprodukten. In: Kramme R, editor. Medizintechnik. Berlin/Heidelberg, Germany: Springer; 2011:99-108.
4. Funk J, Schmidt L. Evaluation of an augmented reality in-struction for a complex assembly task. i-com 2021;20(1):63-72. [doi: 10.1515/icom-2021-0005]
5. Rosanne O, Albuquerque I, Cassani R, Gagnon J, Tremblay S, Falk TH. Adaptive filtering for improved EEG-based mental workload assessment of ambulant users. Front Neurosci 2021 Apr 7;15:611962. [doi: 10.3389/fnins.2021.611962] [Medline: 33897342]
6. Bernhardt S, Nicolau SA, Soler L, Doignon C. The status of augmented reality in laparoscopic surgery as of 2016. Med Image Anal 2017 Apr 18;37:66-90 [FREE Full text] [doi: 10.1016/j.media.2017.01.007] [Medline: 28160692]
7. Thees M, Kapp S, Strzys MP, Beil F, Lukowicz P, Kuhn J. Effects of augmented reality on learning and cognitive load in university physics laboratory courses. Computers in Human Behavior 2020 Jul;108:106316. [doi: 10.1016/j.chb.2020.106316]
8. Küçük S, Kapakin S, Göktaş Y. Learning anatomy via mobile augmented reality: effects on achievement and cognitive load. Anat Sci Educ 2016 Oct;9(5):411-421. [doi: 10.1002/ase.1603] [Medline: 26950521]
9. Chao C, Wu S, Yau Y, Feng W, Tseng F. Effects of three‐dimensional virtual reality and traditional training methods on mental workload and training performance. Hum Ftrs & Erg Mfg Svc 2017 May 24;27(4):187-196. [doi: 10.1002/hfm.20702]
10. Ghani U, Signal N, Niazi I, Taylor D. A novel approach to validate the efficacy of single task ERP paradigms to measure cognitive workload. Int J Psychophysiol 2020 Dec;158:9-15. [doi: 10.1016/j.ijpsycho.2020.09.007] [Medline: 33045292]
11. Hart S, Stavenland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Human Mental Workload. Amsterdam, The Netherlands: Elsevier; 1988:139-183.
12. Nicolau S, Soler L, Mutter D, Marescaux J. Augmented reality in laparoscopic surgical oncology. Surg Oncol 2011 Sep;20(3):189-201. [doi: 10.1016/j.suronc.2011.07.002] [Medline: 21802281]
13. Dixon BJ, Daly MJ, Chan H, Vescan A, Witterick IJ, Irish JC. Augmented real-time navigation with critical structure proximity alerts for endoscopic skull base surgery. Laryngoscope 2014 Apr;124(4):853-859. [doi: 10.1002/lary.24385] [Medline: 24122916]
14. Polich J. Updating P300: an integrative theory of P3a and P3b. Clin Neurophysiol 2007 Oct;118(10):2128-2148 [FREE Full text] [doi: 10.1016/j.clinph.2007.04.019] [Medline: 17573239]
15. Williams LM, Simms E, Clark CR, Paul RH, Rowe D, Gordon E. The test-retest reliability of a standardized neurocognitive and neurophysiological test battery: "neuromarker". Int J Neurosci 2005 Dec;115(12):1605-1630. [doi: 10.1080/00207450590958475] [Medline: 16287629]
16. Reinschluessel AV, Muender T, Salzmann D, Döring T, Malaka R, Weyhe D. Virtual reality for surgical planning - evaluation based on two liver tumor resections. Front Surg 2022;9:821060 [FREE Full text] [doi: 10.3389/fsurg.2022.821060] [Medline: 35296126]
17. Colligan L, Potts HWW, Finn CT, Sinkin RA. Cognitive workload changes for nurses transitioning from a legacy system with paper documentation to a commercial electronic health record. Int J Med Inform 2015 Jul;84(7):469-476. [doi: 10.1016/j.ijmedinf.2015.03.003] [Medline: 25868807]
18. Park S, Yoo J, Lee Y, DeGuzman PB, Kang M, Dykes PC, et al. Quantifying emergency department nursing workload at the task level using NASA-TLX: an exploratory descriptive study. Int Emerg Nurs 2024 Jun;74:101424. [doi: 10.1016/j.ienj.2024.101424] [Medline: 38531213]
19. Hemmerich W. Rechner zur adjustierung des α-niveaus. StatistikGuru. 2016. URL: https://statistikguru.de/rechner/adjustierung-des-alphaniveaus.html [accessed 2024-08-01]
20. Handy T. In: Handy T, editor. Event-Related Potentials. A Methods Handbook. Cambridge, MA: MIT Press; 2005:33-56.
21. Wascher E, Reiser J, Rinkenauer G, Larrá M, Dreger FA, Schneider D, et al. Neuroergonomics on the go: an evaluation of the potential of mobile EEG for workplace assessment and design. Hum Factors 2023 Feb;65(1):86-106 [FREE Full text] [doi: 10.1177/00187208211007707] [Medline: 33861182]
22. El Basbasse Y, Packheiser J, Peterburs J, Maymon C, Güntürkün O, Grimshaw G, et al. Walk the plank! Using mobile electroencephalography to investigate emotional lateralization of immersive fear in virtual reality. R Soc Open Sci 2023 May 31;10(5):221239 [FREE Full text] [doi: 10.1098/rsos.221239] [Medline: 37266038]
23. Alessa FM, Alhaag MH, Al-Harkan IM, Ramadan MZ, Alqahtani FM. A neurophysiological evaluation of cognitive load during augmented reality interactions in various industrial maintenance and assembly tasks. Sensors (Basel) 2023 Sep 06;23(18):7698 [FREE Full text] [doi: 10.3390/s23187698] [Medline: 37765755]
24. Binias B, Myszor D, Binias S, Cyran KA. Analysis of relation between brainwave activity and reaction time of short-haul pilots based on EEG data. Sensors (Basel) 2023 Jul 17;23(14):6470 [FREE Full text] [doi: 10.3390/s23146470] [Medline: 37514762]
25. Ji L, Zhang C, Li H, Zhang N, Guo C, Zhang Y, et al. A comprehensive experimental framework based on analysis of the pilot's EEG and NASA-TLX questionnaire in a VR environment. Proc Inst Mech Eng H 2023 Jul;237(7):869-878. [doi: 10.1177/09544119231179842] [Medline: 37313977]

XSL•FO
RenderX

26. Aksoy M, Ufodiama CE, Bateson AD, Martin S, Asghar AUR. A comparative experimental study of visual brain event-related potentials to a working memory task: virtual reality head-mounted display versus a desktop computer screen. Exp Brain Res 2021 Oct 04;239(10):3007-3022 [FREE Full text] [doi: 10.1007/s00221-021-06158-w] [Medline: 34347129]
27. Giorgi A, Menicocci S, Forte M, Ferrara V, Mingione M, Alaimo Di Loro P, et al. Virtual and reality: a neurophysiological pilot study of the sarcophagus of the spouses. Brain Sci 2023 Apr 07;13(4):635 [FREE Full text] [doi: 10.3390/brainsci13040635] [Medline: 37190600]
28. Marucci M, Di Flumeri G, Borghini G, Sciaraffa N, Scandola M, Pavone EF, et al. The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. Sci Rep 2021 Mar 01;11(1):4831 [FREE Full text] [doi: 10.1038/s41598-021-84196-8] [Medline: 33649348]
29. Atici-Ulusu H, Ikiz YD, Taskapilioglu O, Gunduz T. Effects of augmented reality glasses on the cognitive load of assembly operators in the automotive industry. International Journal of Computer Integrated Manufacturing 2021 Mar 21;34(5):487-499. [doi: 10.1080/0951192x.2021.1901314]
30. Taori TJ, Gupta SS, Gajre SS, Manthalkar RR. Cognitive workload classification: towards generalization through innovative pipeline interface using HMM. Biomedical Signal Processing and Control 2022 Sep;78:104010. [doi: 10.1016/j.bspc.2022.104010]
31. Luque F, Armada V, Piovano L, Jurado-Barba R, Santamaría A. Understanding pedestrian cognition workload in traffic environments using virtual reality and electroencephalography. Electronics 2024 Apr 11;13(8):1453. [doi: 10.3390/electronics13081453]
32. Mondellini M, Pirovano I, Colombo V, Arlati S, Sacco M, Rizzo G, et al. A multimodal approach exploiting EEG to investigate the effects of VR environment on mental workload. International Journal of Human–Computer Interaction 2023 Sep 19;40(20):6566-6578. [doi: 10.1080/10447318.2023.2258017]

Abbreviations

AEP: auditory evoked potential
AR: augmented reality
EEG: electroencephalography
ERP: event-related potential
MRI: magnetic resonance imaging
NASA-TLX: National Aeronautics and Space Administration-Task Load Index
rmANOVA: repeated measures analysis of variance
VR: virtual reality

Edited by A Mavragani; submitted 10.06.24; peer-reviewed by Y Gogna; comments to author 30.08.24; revised version received 14.10.24; accepted 07.11.24; published 08.01.25.

Please cite as:
Schneider T, Cetin T, Uppenkamp S, Weyhe D, Muender T, Reinschluessel AV, Salzmann D, Uslar V
Measuring Bound Attention During Complex Liver Surgery Planning: Feasibility Study
JMIR Form Res 2025;9:e62740
URL: https://formative.jmir.org/2025/1/e62740
doi:10.2196/62740
PMID:

©Tim Schneider, Timur Cetin, Stefan Uppenkamp, Dirk Weyhe, Thomas Muender, Anke V Reinschluessel, Daniela Salzmann, Verena Uslar. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development of Personas and Journey Maps for Artificial Intelligence Agents Supporting the Use of Health Big Data: Human-Centered Design Approach

Yoon Heui Lee[1], RN, MSN; Hanna Choi[2], RN, PhD; Soo-Kyoung Lee[3], RN, PhD

[1]Department of Nursing, Graduate School, Keimyung University, Daegu, Republic of Korea

[2]Department of Nursing Science, Nambu University, Gwangju, Republic of Korea

[3]Big Data Convergence and Open Sharing System, Seoul National University, Seoul, Republic of Korea

Corresponding Author:
Soo-Kyoung Lee, RN, PhD
Big Data Convergence and Open Sharing System
Seoul National University
1 Gwanak-ro
Seoul, 08826
Republic of Korea
Phone: 82 880 2344
Email: soo1005@snu.ac.kr

Abstract

Background: The rapid proliferation of artificial intelligence (AI) requires new approaches for human-AI interfaces that are different from classic human-computer interfaces. In developing a system that is conducive to the analysis and use of health big data (HBD), reflecting the empirical characteristics of users who have performed HBD analysis is the most crucial aspect to consider. Recently, human-centered design methodology, a field of user-centered design, has been expanded and is used not only to develop types of products but also technologies and services.

Objective: This study was conducted to integrate and analyze users' experiences along the HBD analysis journey using the human-centered design methodology and reflect them in the development of AI agents that support future HBD analysis. This research aims to help accelerate the development of novel human-AI interfaces for AI agents that support the analysis and use of HBD, which will be urgently needed in the near future.

Methods: Using human-centered design methodology, we collected data through shadowing and in-depth interviews with 16 people with experience in analyzing and using HBD. We identified users' empirical characteristics, emotions, pain points, and needs related to HBD analysis and use and created personas and journey maps.

Results: The general characteristics of participants (n=16) were as follows: the majority were in their 40s (n=6, 38%) and held a PhD degree (n=10, 63%). Professors (n=7, 44%) and health care personnel (n=10, 63%) represented the largest professional groups. Participants' experiences with big data analysis varied, with 25% (n=4) being beginners and 38% (n=6) having extensive experience. Common analysis methods included statistical analysis (n=7, 44%) and data mining (n=6, 38%). Qualitative findings from shadowing and in-depth interviews revealed key challenges: lack of knowledge on using analytical solutions, crisis management difficulties during errors, and inadequate understanding of health care data and clinical decision-making, especially among non–health care professionals. Three types of personas and journey maps—health care professionals as big data analysis beginners, health care professionals who have experience in big data analytics, and non–health care professionals who are experts in big data analytics—were derived. They showed a need for personalized platforms tailored to the user level, appropriate direction through a navigation function, a crisis management support system, communication and sharing among users, and expert linkage service.

Conclusions: The knowledge obtained from this study can be leveraged in designing an AI agent to support future HBD analysis and use. This is expected to further increase the usability of HBD by helping users perform effective use of HBD more easily.

(JMIR Form Res 2025;9:e67272) doi:10.2196/67272

XSL•FO
RenderX

KEYWORDS

analysis; health big data; human-centered design; persona; user journey map; artificial intelligence; human-AI; interviews; users' experiences

Introduction

Since COVID-19, the new normal era has brought innovative advances in digital technology, resulting in various changes in all areas of our lives. Health care is rapidly emerging as one of the most representative and top-priority areas of advanced digital technology [1-3]. Digital health is a generic term for health care–related fields that combine digital technology and medical care, including telehealth, telemedicine, mobile health, wearable and internet of things–based health care, and big data and artificial intelligence (AI)–based health care [4-6]. The purpose of digital health is to create specialized services for consumers based on data- and evidence-based judgment, and its goal is to realize precision health care [3,7,8] through sustainable digital transformation of health care [9].

Precision health care is a new health care paradigm that involves analyzing various information such as genetic and environmental factors, disease history, lifestyle, and individual health information to prevent diseases (preventive health care), maximize treatment effects, and minimize treatment side effects (personalized health care); this also involves the participation of patients in the decision-making process (participatory health care) [10-13]. Having an accurate prediction based on various health data is important in realizing such precision health care [14]. Health care is one of the fastest-growing sectors worldwide and requires an efficient and continuous digital transformation [15]. The digital transformation of health care should consider technological, political, social, cultural, and economic aspects and is a key and strategic field for improving health care quality and creating new value [16].

According to IBM, an individual generates more than 1100 terabytes of health data in their lifetime [17], and this huge amount of data that was previously just discarded can now be shared, transmitted, and stored with advances in the cloud, AI, and social media. Integrating and analyzing such health big data (HBD) can derive new insights for predicting, preventing, and treating diseases [18,19]. Korea has a huge amount of HBD; hence, the development of various government-level policy support and related technologies continues [20,21] and has favorable conditions for the use of HBD. However, the usability of HBD is low. The use of big data is mainly hindered by the absence of experts related to big data analysis, inconsistent data quality, and difficulty in analysis procedures [22,23].

Currently, there are only a few professionals who can manage and analyze big data in Korea. As interest in big data expands, analytic expert training courses are continuously being opened; however, they are still far from enough to meet demands [24]. In addition, considering future demands where the use and influence of big data will further increase, the workforce imbalance is also expected to increase [22,25]. Therefore, training more people with universal capabilities based on an understanding of data can be an effective alternative to meet insufficient workforce demand and increase the use of big data [26]. To this end, a system that supports the overall process of big data analysis is needed so that many people can access HBD more easily for analysis and use. Recently, human-centered design methodology, a field of user-centered design, has been expanded and used not only to develop types of products but also technologies or services [27,28]. This helps in specifically identifying the problem based on the user's experience, identifying the user's needs, and reflecting them in the human-centered design [29,30].

AI applications are becoming increasingly common in almost all areas of human life. In addition, the medical field is also accelerating to reflect AI. The rapid proliferation of AI requires new approaches for human-AI interfaces that are different from classic human-computer interfaces [31]. Persona and journey mapping techniques have been used in the literature and industry to elicit requirements for system development. Most of the software developers, users, and managers consider that the journey map and persona techniques are effective in helping understand the software requirements to be developed by the development teams [32].

In developing a system that supports the analysis and use of HBD, reflecting the empirical characteristics of users who actually perform big data analysis is the most important procedure to consider. Therefore, this study aims to understand the empirical characteristics and feelings of users related to big data analysis by building persona and user journey map included in the human-centered design methodology and to understand users' needs for systems that support HBD analysis and use. This can be used as a design concept for AI agents to support future HBD analysis and use.

Furthermore, this study aims to design an AI agent to support future HBD analysis and use by integrating and analyzing users' experiences along the HBD analysis and use journey and applying the human-centered design methodology as follows: (1) the persona type is derived by investigating the user's experience focusing on the HBD analysis and use journey and (2) a customer journey map is created according to the derived persona to identify the needs of users presented by type.

Methods

Research Design

This is a methodological study that synthesizes and analyzes users' experiences according to the big data analysis and use journey; generally, the human-centered design methodology is used before developing an AI agent that supports the analysis and use of HBD.

Participants

Research participants were secured through convenience and snowball sampling methods for adults aged 19 years or older living in Korea who have experience in big data analysis using HBD for research or work purposes or have experience using

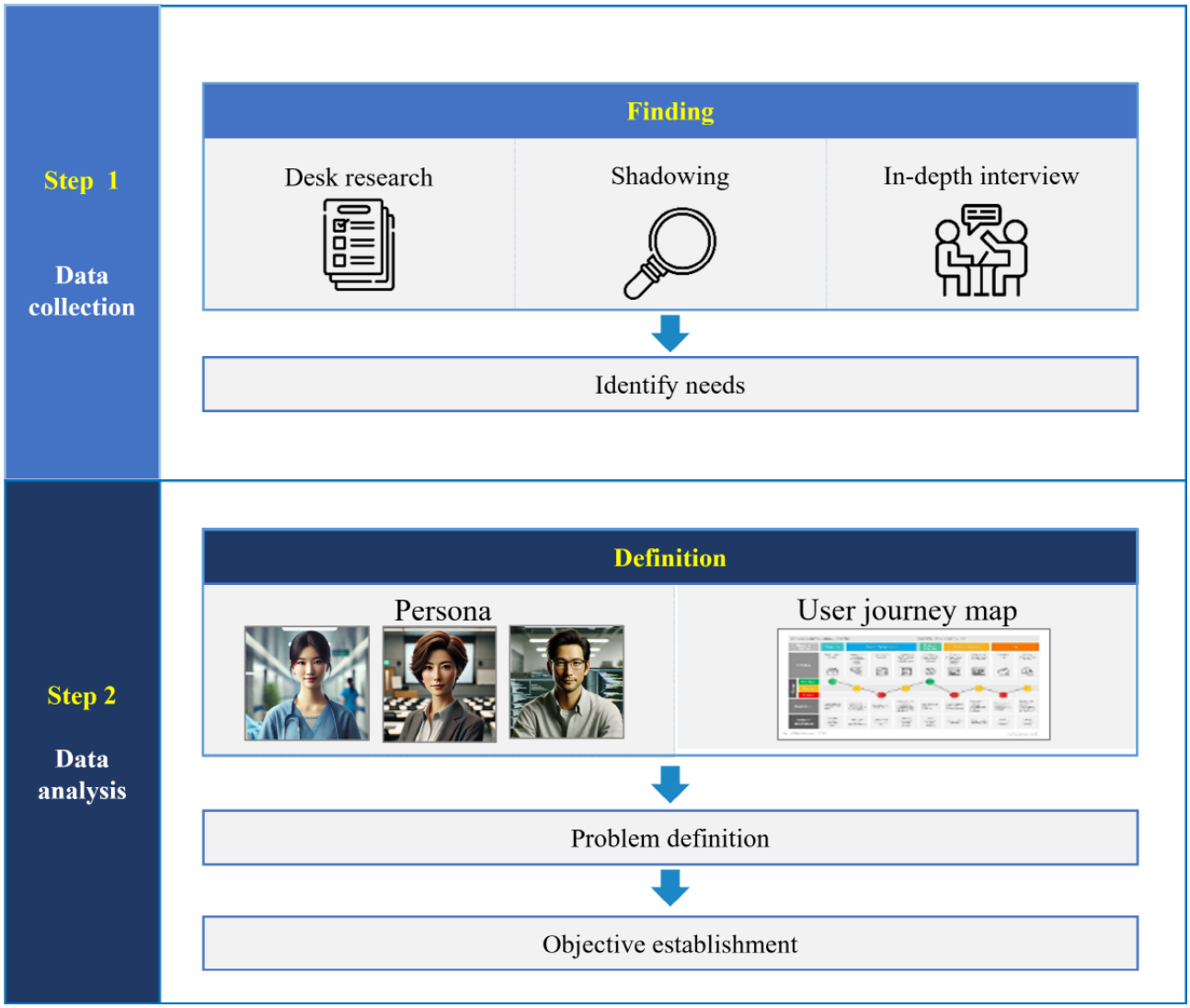
XSL•FO
RenderX

the results of the big data analysis. Participants for this study were recruited voluntarily by sending a recruitment notice to individuals identified as meeting the study's objectives from among those who had participated in a prior demand survey on the analysis and use of HBD [33]. This survey served as a preliminary study for this research, and the individuals identified through it were selected as the primary participants. Additional participants were then recruited through referrals from these primary participants. To ensure diversity within the participant pool, individuals were classified into 4 categories based on their level of analysis experience (beginner or proficient) and their occupational field (health and medical fields or other fields). Efforts were made to select participants evenly across these categories. There were 16 participants in this study. The number of participants was evaluated based on previous research that suggested that a saturation point without new discoveries could be reached when interviewing 6 to 12 participants, and recruitment of participants was terminated at the point when new topics were no longer derived [34]. Data were collected until no new topics emerged from the interviews, and the recruitment of new participants was stopped at the time of saturation [35].

Procedures

This study consisted of 2 stages based on a review of the big data analysis and use procedure through desk research and a case study on the HBD platform (Figure 1). The first stage, discover, involves examining and understanding the problem situation to find the difficulties and hidden needs of those who use the service. Data research was conducted through shadowing and in-depth interviews with participants working on HBD analysis. The second stage, define, involves deriving major issues related to the problem based on analysis and contextual understanding of user survey results and setting service goals. Based on the needs of users explored through persona and user journey map, the AI agent's design concept was proposed.

Figure 1. Research progress process: this figure outlines the research process, which comprises 2 stages: the first stage focuses on identifying users' challenges and hidden needs, while the second stage involves defining the problem and setting service goals to address those issues.



Data Collection

Desk research examined prior documents and public institution announcements and papers searching for HBD and big data analysis. Platforms in overseas (PubMed and Google Scholar) and domestic (DBPia, Riss, and Kiss) databases were also searched. In-depth interviews took place over 3 weeks from February 1 to 21, 2023. The interview was conducted using a

XSL•FO
RenderX

guideline consisting of 6 short-answer questions on the participant's basics and open questions related to their HBD analysis and use journey. Based on the related literature [36] and expert experience, some main categories of difficulties and coping methods experienced in each stage of big data analysis arose, as well as desired support services used to solve the problem (Textbox 1). This was done after the initial formulation and underwent expert review.

Textbox 1. In-depth interview guideline: open-ended questions regarding users' challenges and support needs at each stage of health big data analysis.

Collection

- Talk about the type of data you used, and how to select and collect data.
- What was the most difficult part of the data collection process?
- How did you solve the difficulty?
- If you had a magic wand to solve any problems during the data collection phase, what would you solve?

Preprocess

- Talk about the process of preprocessing collected data (integration, conversion, and refinement).
- What were the difficulties you experienced during the preprocessing process and what was the most difficult thing?
- How did you solve the difficulties that occurred during the preprocessing stage?
- If you had a magic wand that could solve any problems during the preprocessing phase, what would you solve?

Analyzing

- Did you conduct the analysis yourself? If not, why?
- Was there any problem when you asked an expert to analyze it?
- What were the difficulties you experienced in the analysis stage, and what was the most difficult thing?
- How did you solve the difficulties that occurred?
- If you had a magic wand to solve any problems during the analysis phase, what would you solve?

Visualizing

- Was there any difficulty in interpreting and applying the derived results?
- What were the difficulties you experienced in the analysis stage, and what was the most difficult thing?
- What support do you need most during the result interpretation and visualization phase?

Use

- What features do you think should be included in artificial intelligence agents that support the use of health big data analysis?

To increase the realistic value of the data, the researcher visited the user's studio where the computer was used by the participant for actual big data analysis and conducted one-on-one interviews. The individual interviews took about 1.5 hours. The process concluded when responses became repetitive, and no new information was generated. For aspects that were challenging to understand through verbal explanations alone, participants were asked to provide illustrative examples during the interview. With prior consent from the participants, the interviews were audio-recorded, and significant concepts identified during the discussions were documented in field notes. Each participant was interviewed once, and if further information was required during the data analysis phase, additional data were collected through follow-up phone calls or emails with the participants.

For 2 weeks, the researcher visited and shadowed 1 participant, a nurse researcher, who is currently conducting HBD analysis. While shadowing her in the laboratory, the researcher observed the participant's behavior, and they captured the contact point of inconvenience and the moment when a problem occurred. To check for difficulties encountered by participants under different stages, from preprocessing to visualization of HBD, the researcher visited 3 times for a total of 1 hour per visit and interviewed the participant when important issues were found. Observed behaviors and environmental factors were recorded in field notes.

Data Analysis

The interview data were transcribed by the researcher on the same day they were collected, and the transcriptions were analyzed at the levels of codes, subcategories, categories, and themes using qualitative content analysis procedures [37]. In the first stage, redundant content was excluded from the meaning units derived from the interviews, and codes were created to represent the core ideas using relevant concepts or phrases. In the second stage, the implicit meanings or content captured in the codes were abstracted to form subcategories. In the third stage, these subcategories were grouped into broader categories. Finally, in the fourth stage, themes were developed by

XSL•FO
RenderX

integrating participants' key experiences related to the HBD analysis process. To ensure the rigor of the analysis, the data were reviewed, analyzed, and discussed iteratively with the assistance of the second author (HC), a nursing professor with extensive experience in qualitative research. The findings were further validated for their universality and content validity through feedback from 2 research participants.

Based on insights confirmed through interviews and shadowing, personas and customer journey maps were designed to represent the participants' experiences. A persona is a vivid and specific description of a potential end user and a fictional character representing the type of group to use the service; this is created by adding attributes to multiple subgroups identified within a potential user group [38]. Identifying users' needs through personas formed by systematically classifying and embodying potential user groups can help present specific goals and directions for service development [39]. User journey maps are visualizations of the overall process and emotions of a persona experiencing the service over time; this helps in identifying and understanding the context of the problematic situation [30,40].

Ethical Considerations

This study was approved by the institutional review board of the University of Keimyung (40525-202210-HR-059-03). All participants received a detailed information sheet about the research, were given the opportunity to ask questions, and were informed of their right to withdraw at any time without any negative consequences. Written informed consent was obtained from all participants, and to ensure confidentiality, all data were anonymized. To ensure confidentiality, all audio-recorded data were anonymized by removing personal information and assigning unique identification codes to distinguish participants. These data were securely stored on the researcher's personal computer, equipped with security software, and were used solely for research purposes. Measures were taken to prevent any breach of personal information, violations of human rights, or ethical concerns, and the data will be permanently deleted after the study's conclusion. Upon the completion of interviews and observations, a small honorarium was provided to the participants.

Results

Participants' General Characteristics

Participants' general characteristics are as follows (Table 1). Regarding the age of the participants, 6 (38%) were in their 40s. In terms of education, 6 (38%) held a master degree, and 10 (63%) held a PhD degree. As far as professions, professors (n=7, 44%) and researchers (n=4, 25%) accounted for the majority. Among the participants, 10 (62%) were health care personnel, and 6 (38%) were non–health care personnel. Regarding the experience of big data analysis, 4 (25%) were beginners, 6 (38%) had minor experience, and 6 (38%) had extensive experience. Formal data from public institutions were the most common type of data used (n=10, 62%). The analysis methods were statistical analysis (n=7, 44%), data mining (n=6, 38%), and text mining (n=5, 31%). Big data were used mostly for academic and research purposes (n=14, 88%).

XSL•FO
RenderX

Table 1. General characteristics of participants about age, occupation, education, and experience in health big data analysis (N=16).

Characteristics and categories	Values, n (%)
Age (years)	
≤29	1 (6)
30-39	4 (25)
40-49	6 (38)
50-59	4 (25)
≥60	1 (6)
Degree of education	
Master degree	6 (38)
Doctoral degree	10 (62)
Job	
Doctor	1 (6)
Nurse	2 (12)
Professor	7 (44)
Researcher	4 (25)
Company worker	2 (12)
Experience of analysis	
Beginner	4 (25)
Minority	6 (38)
Majority	6 (37)
Method of analysis[a]	
Statistical analysis	7 (44)
Data mining	6 (38)
Text mining	5(31)
Type of used data[a]	
Structured data (public institution)	10 (62)
Dielectric data	2 (13)
Text data	5 (31)
Purpose of use	
Scholarship or research	14 (88)
Product or service development	2 (13)

[a]Multiple responses.

Shadowing

As a result of observing the preprocessing of text data, which were unstructured data, the natural language processing stage had more procedures for manual verification than automatic tasks; this involved visually checking numerous data individually using analytical solutions. Researchers call this work manual labor due to the enormous amount of time spent on the work and considerable energy consumption. This pretreatment is an important task that can affect the outcome later if it is not completely refined at that stage. However, it has been found that nonanalysis professionals often deal with huge amounts of work without understanding the step-by-step progress, which often leads to situations wherein important parts are omitted.

Most users are trained on how to use the analytical solution before conducting big data analysis; however, they often forget the necessary commands or use because they do not have sufficient knowledge and are unfamiliar with the command line interface. In this case, the user wants to inquire about the information to an analysis-related expert, but there is no expert around to answer; hence, it was found that it was mainly solved by books or internet searches. However, this method took too much time, and in most cases, the desired information could not be found.

XSL•FO
RenderX

In-Depth Interview

Overview

Based on the results of the in-depth interview, the user's difficulties in the process of analyzing HBD showed a clear difference depending on the job. In the case of health care personnel, there was a slight difference in their experience when performing analysis work directly and requesting analysis work from analysts; yet, in both cases, there was no difficulty in understanding health data or finding potential meanings in analysis results and applying them to clinical situations. However, it was found that the major problems commonly experienced involve selection, installation, and use of analytical solutions and lack of knowledge and understanding of step-by-step analysis tasks that vary widely by data. Meanwhile, for non–health care personnel, difficulties in decision-making at various stages derived from a lack of health care knowledge have been identified as a major problem. These difficulties can be divided into 3: lack of knowledge on how to use analytical solutions (health care personnel), lack of crisis management ability (health care personnel), and lack of understanding of health care data and clinical decision-making ability (non–health care personnel).

Type 1: Lack of Knowledge on How to Use Analytical Solutions

In the case of health care personnel, beginners who did not have much experience in analyzing big data answered that they had difficulty in selecting and installing appropriate analytical solutions. Moreover, in the preprocessing and analysis stages, they did not know the package or library required for each step or did not know the syntax or function. One participant stated that:

> *I thought it would be over if I installed the analytical solution, but it was really painful to have to reinstall the library I needed for each task. I don't know which package to use and how to install it, and how to enter the command after installing it. I don't remember what I learned in the past, and it's faster to do things manually, so most of the preprocessing were done manually.* [Participant 4]

Type 2: Lack of Crisis Management Skills

Participants responded that they had difficulty finding the cause and solution when an error occurred during big data analysis. In particular, if the error was not resolved, they could no longer proceed to the next step, and the whole process was completely interrupted; sometimes, even very simple problems that could be easily solved could not be found, so they spent days and days resolving them. One participant stated that:

> *I definitely put in the command as I learned, but I kept getting error messages. I don't know what's wrong, and I can't ask anyone because there's no one around me with this knowledge. I found Google and YouTubers to solve it, but it doesn't work out. "Should I really give up here?" I had all sorts of thoughts. Later, through the expert, I found that the cause of the error was a mistake in the command. When I corrected it, it was executed in less than five minutes, but I was disappointed. I wasted three days without doing anything because of that and tried all sorts of ways by Googling.* [Participant 7]

Type 3: Lack of Understanding of Health Care Data and Clinical Decision-Making Ability

In the case of non–health care personnel, it was difficult for participants to determine the clinical significance or usability in the stage of selecting the analysis topic. In addition, for some domestic public data, participants found the data collection procedure too complicated. They responded that it was difficult to grasp the meaning of various variables included in the collected data due to a lack of understanding of health data and that it was cumbersome to merge data separated by year or variable. Regarding preprocessing, it was found that difficulties in preprocessing procedures such as normalization and missing value treatment, which changed the health care meaning and resulted in difficulty in decision-making, were greater than the difficulties related to the analysis itself. In addition, it was confirmed that it was difficult for participants to select variables to be used for analysis. One participant stated that:

> *To use data from the National Health Insurance Service, you have to submit IRB in the research plan, which is so complicated and difficult. The data collection stage was more difficult than the analysis. "Why did the industrial complex data divide the files into so many?" The examination data are difficult to combine because the items conducted every year are slightly different. There is an explanation file for variables, but there are too many variables in the data, so it is difficult to grasp. The test results do not clearly show the extent of the numerical value, so I can't judge well. All the teachers who work in the hospital are busy, and there is no one close enough to approach personally and seek healthcare advice.* [Participant 12]

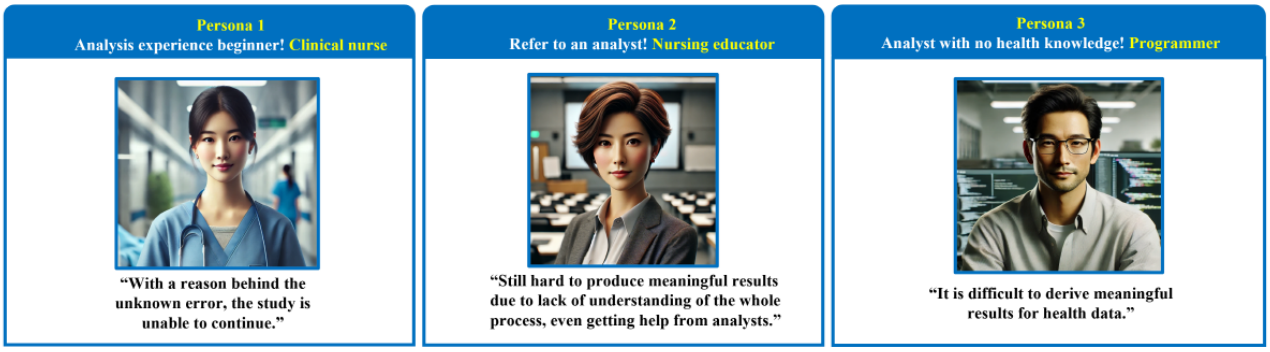
Definition of Persona and User Journey Map

Overview

In this study, when combining situations based on users' actual experiences obtained through in-depth interviews, it was confirmed that analysis experience, occupation, and direct performance of analysis work act as the main factors affecting the user's experience. Accordingly, 3 personas were set up: "health care professionals as big data analysis beginners," "health care professionals who have experience in big data analytics," and "non–health care professionals who are experts in big data analytics." Based on this, needs and pain points from the user's perspective were identified (Figure 2).

XSL•FO
RenderX

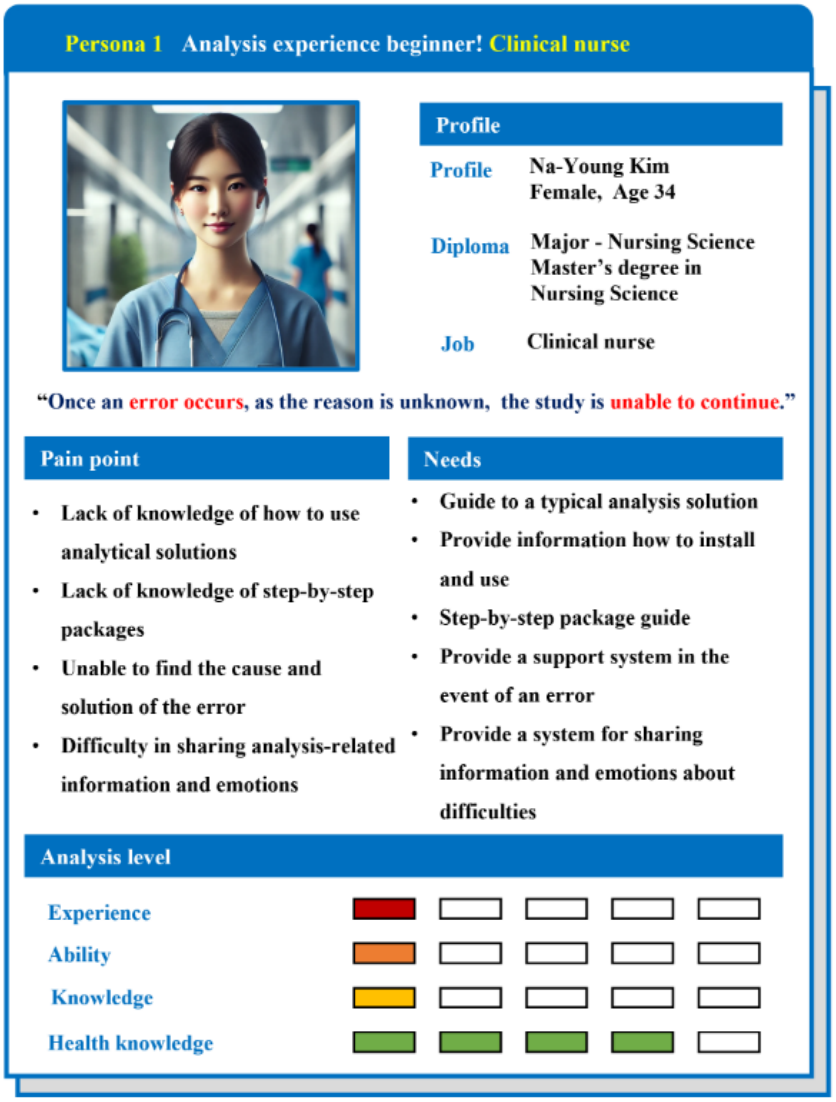
Figure 2. Three types of personas: persona 1: health care professionals as big data analysis beginners, persona 2: health care professionals who have experienced big data analytics, and persona 3: non–health care professionals who are experts in big data analytics.



Persona 1: Health care professionals as big data analysis beginners

Persona 1 is a persona that represents a health care practitioner with entry-level analytical experience (Figure 3). They lack knowledge of analytical solutions, experience frequent errors in progress due to unfamiliar use, and have difficulty finding solutions. Therefore, they need a support system that can provide information on how to use the analytical solutions and query errors or to be able to ask questions real time during the progress stage.

Figure 3. Persona 1: health care professionals as big data analysis beginners: represents novice medical personnel with limited experience in big data analysis. This persona lacks knowledge of analysis tools and solutions and has minimal experience using them.

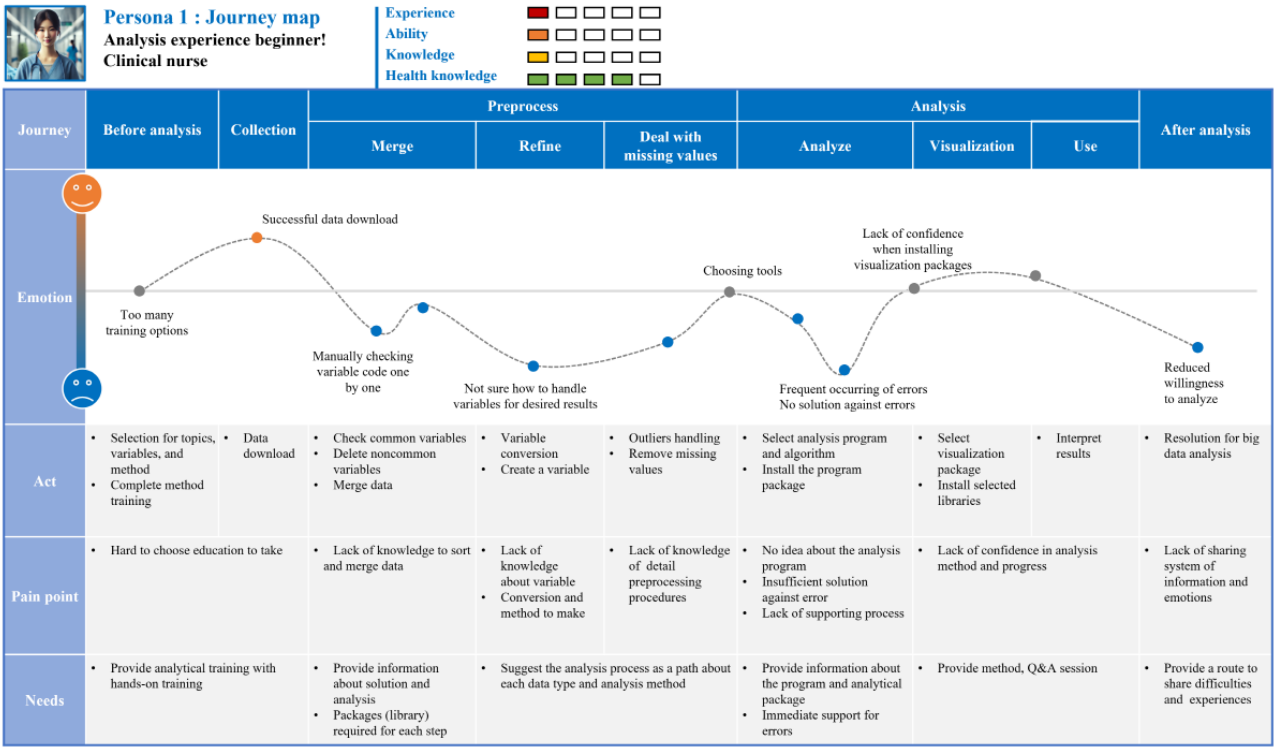


The user journey map prepared based on this persona is shown in Figure 4. "Na-Young Kim," who works as a clinical nurse at a university hospital, recently conducted a potential class analysis study on "risk factors for metabolic syndrome in

XSL•FO
RenderX

single-person households." Since she has never performed big data analysis before, she completed analysis-related training before conducting the research. She was unfamiliar with how to use the analytical solution in the preprocessing process, so she had to proceed with most of it manually, which consumes a lot of time and energy. In the complete analysis stage, an error occurred during the command input process, but the progress was stopped because the cause of the error could not be found. She searched for information through Google and tried to modify it by referring to the provided notice, but it was not solved, so she spent days and days working on the problem. The cause of the error, which was confirmed through experts who were barely connected after many twists and turns, was unexpectedly too simple, and the problem was solved in 5 minutes. She thinks several times that it would be nice to have someone who she can ask for advice on something. After many twists and turns, the analysis was completed, but with the idea that such a difficult experience is enough once in a lifetime, she decided that she will never do big data analysis research again.

Figure 4. User journey map of persona 1: highlights frequent errors occurring during the data preprocessing and analysis stages. The inability to identify the causes of these errors leads to significant time consumption and frustration. Q&A: question and answer.



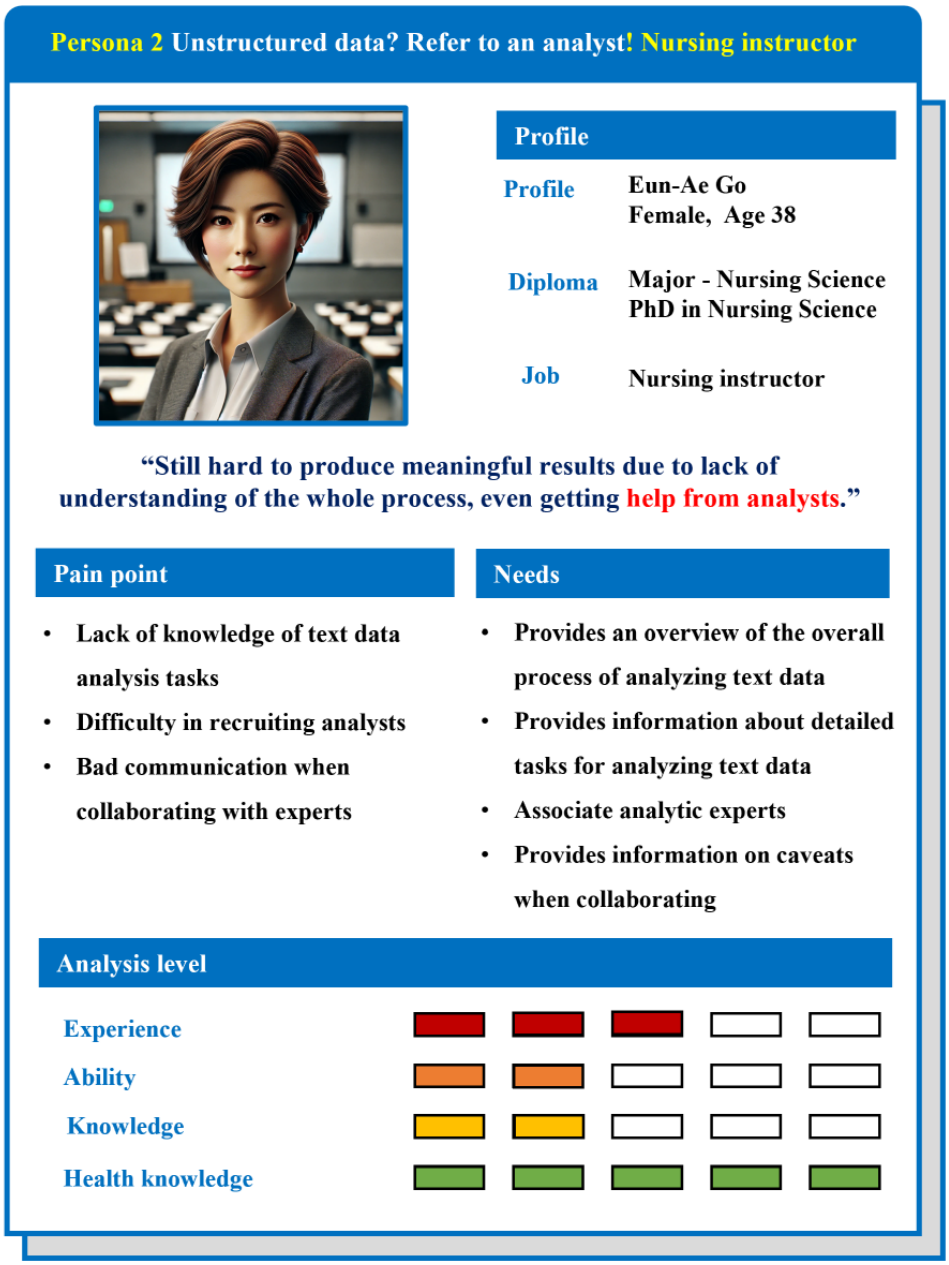
Which part of it is wrong? I don't know, there's no one I can ask, and once an error occurs like this, everything will be completely stopped, so it's really hopeless. I can't believe I held on to a problem that would be solved in five minutes for four days. It's so frustrating.

Persona 2: Health Care Professionals Who Have Experienced Big Data Analytics

Persona 2 is a persona that represents the experience of a health care professional who commissioned an analysis expert to perform the analysis (Figure 5). In this case, if the client lacks an understanding of the overall process of big data analysis or does not communicate sufficiently in advance, disagreement may occur during the process. Therefore, it was found that users who conduct big data analysis through collaboration with analysts need a system that provides information on the contents and precautions to be discussed in advance when requesting analysis, guides the entire process of big data analysis, and provides explanations.

XSL•FO
RenderX

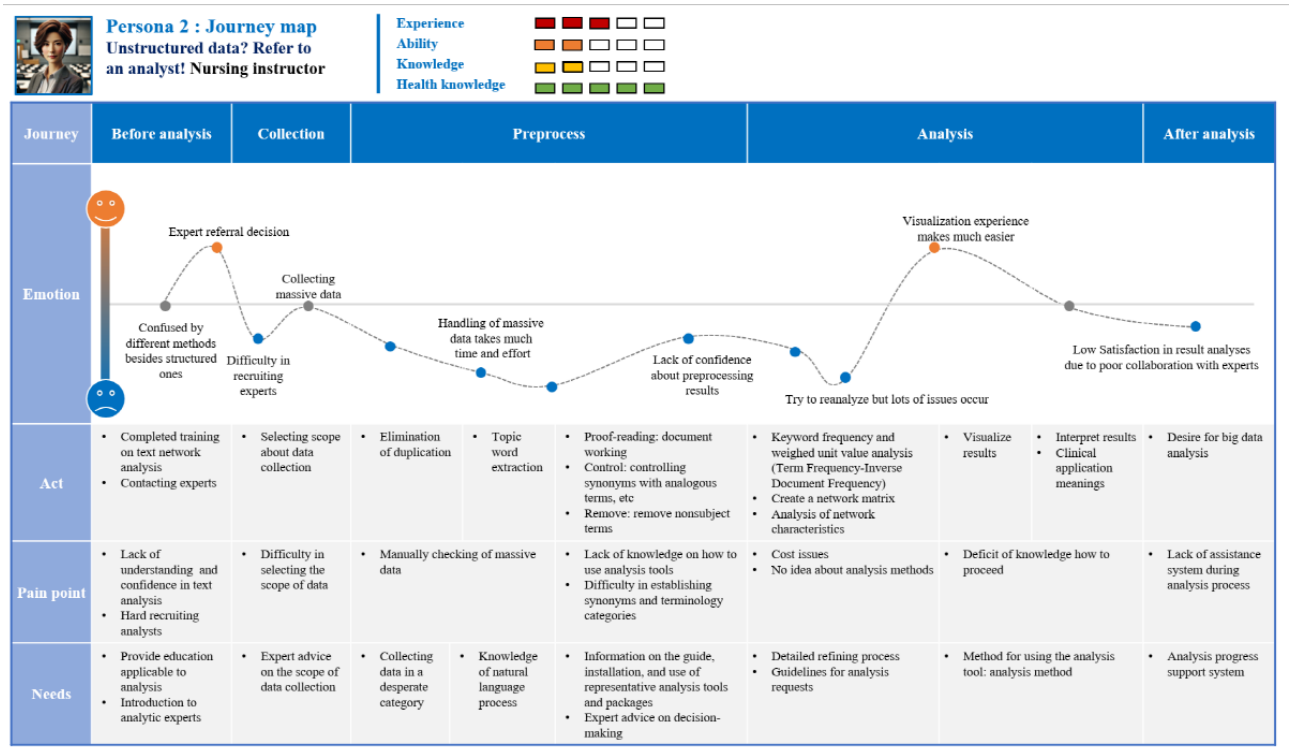
Figure 5. Persona 2: health care professionals who have experience in big data analytics: represents medical personnel who outsource big data analysis tasks to experts. This persona lacks an understanding of the overall analysis process and methods.



The itinerary map for persona 2 is shown in Figure 6. "Eun-Ae Go" is a lecturer who is currently teaching students at a nursing university and recently conducted a study to identify "severe COVID-19 risk factors" using HBD. Although she has previously conducted several studies using HBD, it was her first time analyzing a network using text data; hence, she completed several training sessions on using network analytical solutions before starting the analysis. She did not have an expert in big data analysis around her, and she could not make a decision because she did not trust the companies she found on the internet. She finally got in touch with an engineering major through a few acquaintances and asked them to do both data preprocessing and analysis. It was found after the end of the analysis that the resulting values were affected by texts that were not completely refined during the preprocessing process. She wanted to reanalyze by repurifying, but the analyst did not agree to the reanalysis due to time and cost issues. Through this, she realized that sufficient communication and consultation must precede when requesting analysis work from others; there is also a limit to discovering and interpreting the meaning of the data only with the analysis results conducted by the AI agent without understanding the analysis process or method. Next time, she intends to perform the pretreatment and analysis herself, even if she is not an expert.

XSL•FO
RenderX

Figure 6. User journey map of persona 2: illustrates difficulties in communication and disagreements during the collaboration process, stemming from limited knowledge of the analysis process and methods. These issues result in constraints on the quality and scope of the analysis outcomes.



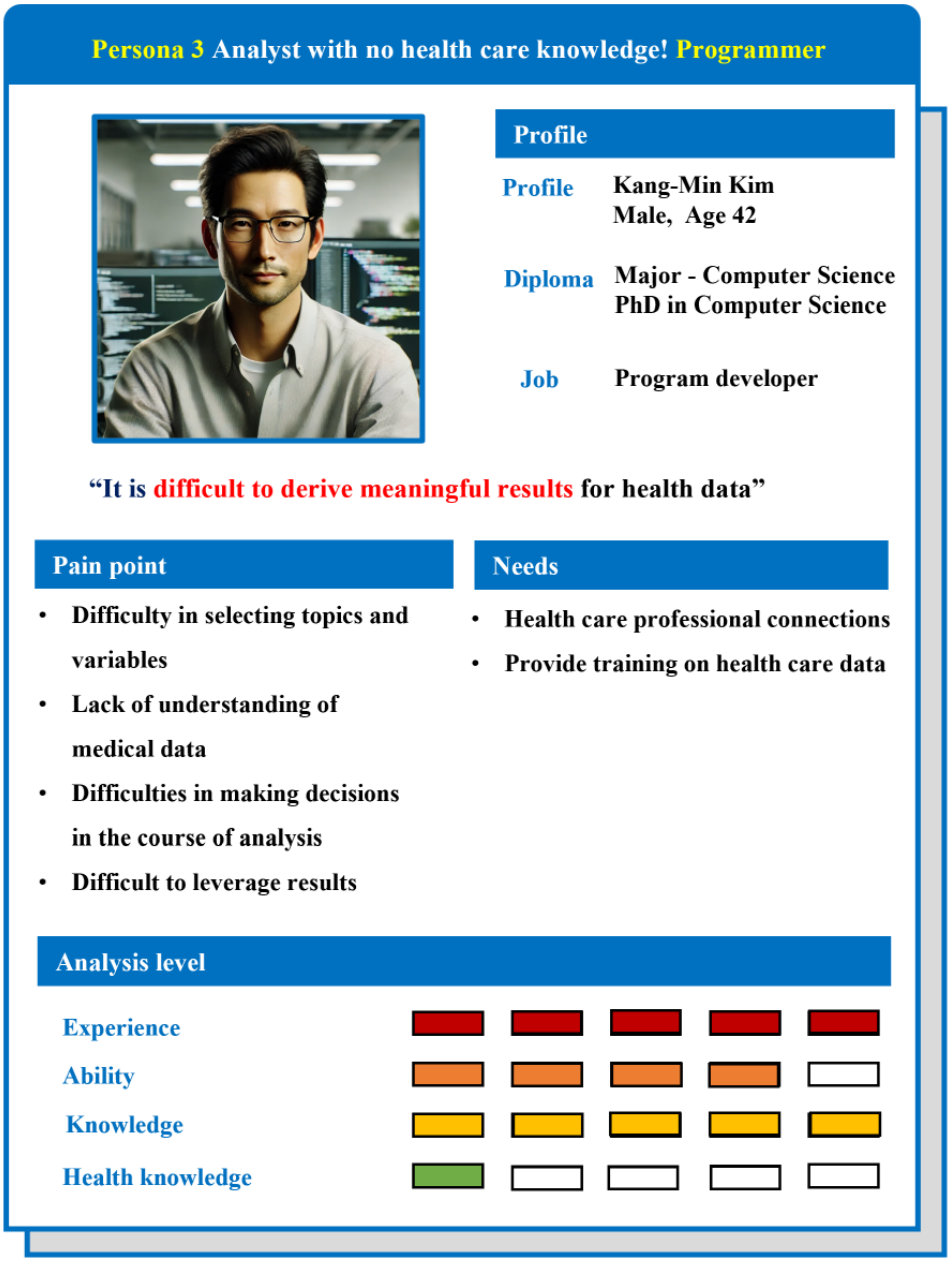
You need to know something when you ask others to do it. Even if we ask experts to analyze big data, we can produce meaningful results only when we have a basic understanding and knowledge of the overall analysis process and method.

Persona 3: Non–Health Care Professionals Who Are Experts in Big Data Analytics

Persona 3 is a persona representing people who are experts in big data analysis but have no knowledge of health care (Figure 7). They are good at big data and analysis, but they experience continuous doubts and confusion during the analysis process due to a lack of health care knowledge. Through this, it was found that non–health care users who are fluent in analysis need a system that provides information on health care data. They also need to seek advice from health care experts in decision-making situations that occur during the analysis process.

XSL•FO
RenderX

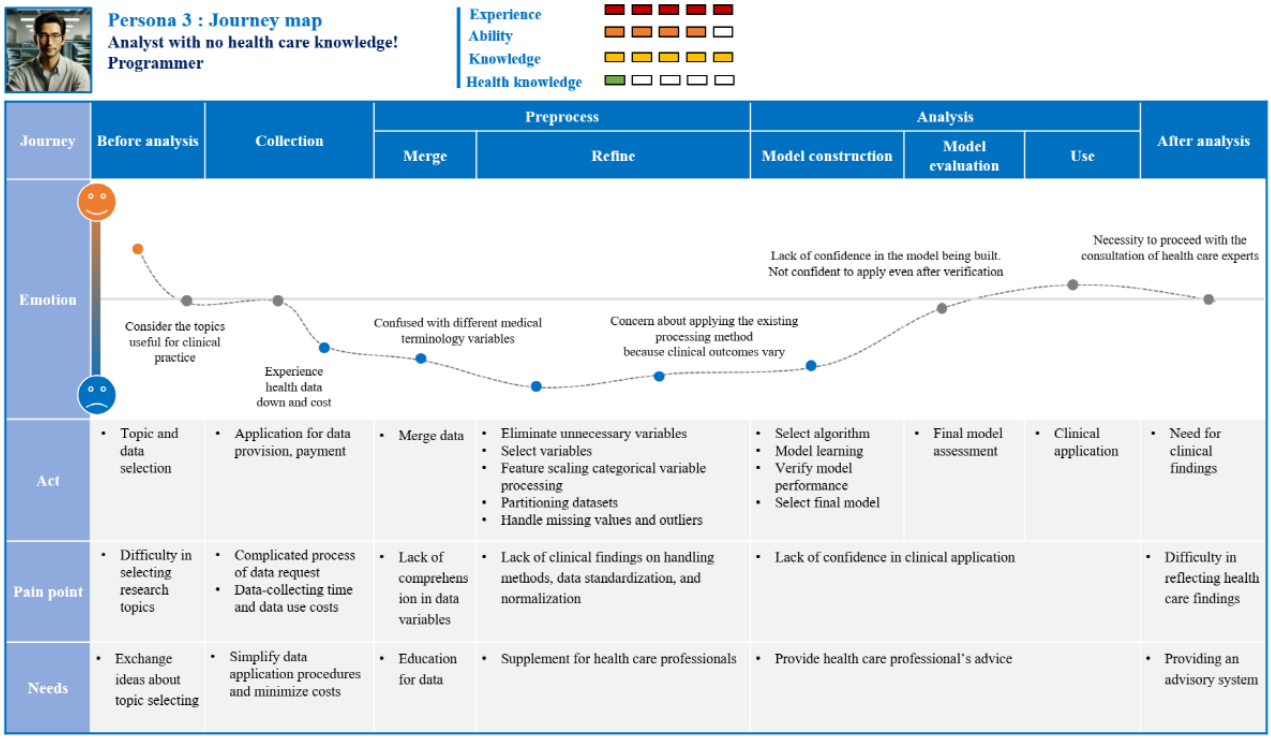
Figure 7. Persona 3: non–health care professionals who are experts in big data analytics: represents a nonmedical professional proficient in big data analysis but lacking expertise in health care. This persona struggles with understanding health data and clinical knowledge.



The itinerary map for persona 3 is shown in Figure 8. Currently working as a programmer, "Kang-Min Kim" is a computer science major and a big data analysis expert. Several studies related to big data analysis in various fields have been conducted. Recently, interest in HBD with infinite use value has increased, and research on "building a diabetes prediction model" has been conducted using the National Health Insurance Service's screening data. Since "Kang-Min Kim" is good at analyzing big data, he predicted that there will be no special variables or difficulties in analyzing big data in the process of HBD analysis. However, unlike previous data analysis work, he felt uncertainty and doubt throughout the analysis process due to the nature of the health data that the analysis findings of the machine were not 100% reliable. Due to a lack of understanding of the meaning of each variable in the health data, the impact of normalization and missing value processing methods on future analysis results, the selection of input variables when applying machine learning, and clinical meaning and application of derived results, among others, he had doubts that he could not be sure of the results of the work that he decided and implemented alone without communication with related experts. At the end of the work, it was deduced that clinical judgment and opinions of health care experts must be combined in the analysis process to properly use HBD. In the future, he intends to conduct HBD analysis through collaboration with health care experts.

XSL•FO
RenderX

Figure 8. User journey map of persona 3: describes experiences of confusion and uncertainty when situations requiring clinical judgment and decision-making arise at various stages, including preprocessing, analysis, and postanalysis.



This numerical value is an abnormal range, how much does this abnormal mean? Does this process affect the clinical meaning of data values? I can't judge because I don't have healthcare knowledge. A proper analysis of health big data requires not only being good at analysis itself.

Deduction of Needs

The user's needs derived for AI agent design that supports the use of big data analysis are as follows.

First, there is a need for a service that provides a guide for analytical solutions, which can be easily used by the public and provides information on how to use the solution to proceed with each step. For nonanalysis professionals, the knowledge of the various libraries (packages) and execution methods required for several tasks to be performed at each stage of preprocessing, analysis, and visualization was unclear; even those with experience did not know how to use the solution. Therefore, users will be able to conduct big data analysis more easily if they have a guide for analytical solutions that are easily available to the public and provide detailed explanations on how to use the solutions for step-by-step tasks according to the progress procedures for each data type.

Second, there is a need for a service that guides the overall process of big data analysis in detail. Big data analysis takes place through 4 stages: data collection, preprocessing, analysis, and visualization and use. Most of the big data analysis platforms in other fields currently operate these 4 stages and provide services such as data introduction and transaction services, data analysis tools introduction and installation guidance, analysis-related training guidance, data analysis and use case sharing. However, there is no system that guides detailed work procedures and specific methods required for each stage. Ordinary people, who are not big data analysis experts, need a service that guides them in their overall analysis journey and provides detailed explanations of the specific tasks to be performed at each stage, just like with navigation, wherein a person arrives at the destination if guided.

Third, there is a need for a system where one can ask or seek real-time advice from analysis experts about problem situations or questions that occur during analysis work. Users may encounter problem situations in the analysis process, and there may be questions that they want to be confirmed by experts. These problems should be dealt with quickly because if they are not solved, they cannot proceed to the next step. However, it is not easy for nonexperts in the field to find the cause of the problem, resulting in a significant waste of time.

Fourth, there is a need for a service that introduces and links analysis experts who can request analysis work or health care experts who can provide health care advice. There may be situations wherein health care personnel need to seek advice on big data analysis and non–health care personnel need to seek health care advice, but it is not easy for users to find these people on their own. It is necessary to provide a channel to meet necessary professionals and at the same time provide guidance on the contents of prior consultations or precautions necessary for collaboration.

Based on the earlier-mentioned problems and user needs, we present five design concepts: (1) personalized platform for the user level, (2) appropriate direction through a navigation function, (3) crisis management support system through question and answer, (4) communication and sharing opportunities between users, and (5) expert-linked services. In addition, it is

XSL•FO
RenderX

believed that these concepts will be able to meet users' needs by adding experience or information related to HBD analysis, providing guidance on health data introduction and access procedures, giving useful big data–related education introduction, and having one-on-one consultations.

Discussion

Principal Findings

To design appropriate and useful technologies or services, it is important to focus on the applicable user's experience and consider the user's individual needs [41]. Recently, this user-centered approach has been actively used in health care–related service development, such as in studies on the design of eHealth AI services [42], smart home system design [43], and web-based agents to support physical and mental health self-management in patients with chronic obstructive pulmonary disease [44]. In this study, before developing an AI agent that supports the use of HBD analysis, the experience and needs of users who have performed big data analysis by applying a human-centered design methodology were identified. The discussions on the 5 service concepts derived through this and the goals and contents to be achieved through each service are as follows.

First, the service is a personalized platform that fits the level of the first user. As a result of the study, the demand for a system that supports big data analysis was different for each user type. The analytical solution of the analysis environment provided in the existing big data platform focuses only on creating an analytic environment and building necessary analytical solutions rather than meeting the needs of users [45,46]. These platforms are not a problem for analysts to use, but in the case of ordinary users, they do not even know how to use a given analytical solution, so the platform created is practically irrelevant or meaningless. Therefore, it is important to provide an analytical environment that fits the user's level and includes detailed information required by the user [47].

In a study on web-based platform development to support participants and caregivers with memory problems, it was also found that the evaluation of the usefulness of the function was different for each group due to different priorities for platform functions between user groups. As a result, a customized platform was designed to meet the needs of a specific group [48]. Guiding cost-free, easily available analytical solutions, and providing necessary information according to the user's needs such as installation methods, necessary libraries, and commands, will reduce the fear of using analytical solutions by inexperienced users. This is supported by examples of successful results in weight, blood pressure, and blood sugar management by characterizing users in the development of a meal planning platform that supports healthy eating development and providing services such as customized nutrition assessment for each user, personalization of meal plans, and changing the environment of individual food purchases [49].

Second, the service provides an appropriate direction presentation system through a navigation function. Most health care professional users were found to lack knowledge on detailed big data analysis processes that vary by data type. If a detailed guide is provided for the analysis progress procedure for each type of representative data, like a road guide navigation, even inexperienced beginners can follow the guide and conduct the analysis on their own. In addition, this can improve understanding of the overall analysis process. In a digital platform designed for youth mental health services, the path of each service needs to be presented, and the content of the service provided by each route is different so that users map along the treatment path determined by case and are guided by the platform to continuously implement it according to the process [50].

Third, the service is a crisis management ability support system through question and answer. A service that allows users to raise queries and answer questions or errors that arise during analysis operations helps users solve problems quickly and reduces unnecessary time and energy waste. A platform development study, which supported operators who manage big data information processing systems, was also equipped with an intelligent web-based system based on AI that helps system operators cope with problems that occur occasionally and obtain the necessary information [51]. In the development of a mobile platform to support patient participation and reporting on lung cancer surgery, a channel was provided to ask patients using the program about questions or concerns that may arise during and after surgery or to report various suspected exacerbations. This allows users to hear immediate health care team explanations, promoting their ability to solve problems [52].

Fourth, the service creates a venue for communication and sharing that provides opportunities for exchange between users. There is a need for a channel of communication that can exchange experiences or information related to big data analysis and share effective big data use cases. Through this, the user may obtain important information and achieve new collaboration through communication. PatientsLikeMe (PatientsLikeMe Inc), a platform that provides information on rare diseases, provides a channel for users to share various information such as their health status and symptoms, drug effects, and side effects. It is not only useful to other users, but it has also resulted in an increase in the number of subscribers to the platform and further expansion of the platform's capabilities, with accumulated data being sold back to necessary institutions [5,6]. The survey on user needs for digital mental health platforms also revealed that the community was a major topic. It has been found that users do not hide their information about diseases or treatments but rather want to connect and share experiences with others, thereby developing the platform's capabilities and believing that new use cases can be developed [53]. In a digital platform designed to enhance cancer treatment support for adult survivors of cancer, survivors of cancer were also found to receive psychological support by expressing various interests such as eating habits and mental health to peer and patient support groups and sharing experiences [5,8].

Finally, the service provides a connection service with experts. Users who want to request analysis work, analysts, or those who need advice from health care experts are required to provide services that link related experts, and these services enable active

XSL•FO
RenderX

use of big data through collaboration. PatientsLikeMe also uses services that link institutions that run clinical trials on new drug development with patients who want the opportunity to use a drug being tested.

Strengths and Limitations

This study adopted a human-centered design approach to enhance the usability of HBD, successfully identifying the unique needs and challenges faced by various user groups. Through in-depth interviews and shadowing techniques, we gained significant insights into the emotional and cognitive aspects of users' experiences during the analysis process. The development of 3 distinct personas and journey maps revealed that each user group has specific requirements. For instance, novice health care professionals require more guidance and support during critical situations, whereas experienced data analysts prefer advanced features and customized settings tailored to various data types. These findings indicate that services provided by the system cannot be uniform and underscore the necessity for personalized user support. Additionally, while many users seek expert advice or communication with peers who have similar experiences, there is a notable lack of accessible communities dedicated to health data analysis and use. This highlights the need for developing user-friendly systems capable of offering practical support to users in the future. Our study has several limitations. First, the in-depth interview participants were primarily users who analyze and use widely used health data, which may limit the generalizability of the findings to users working with less common or emerging types of health data. As new types of health data continue to emerge, the specific demands and analytical experiences associated with these data have not been fully incorporated into the current personas and user journey maps. Additionally, this study may not be generalizable to users outside of the South Korean environment, as the findings are context-specific.

When the number of participants is expanded in future research, the personas and journey maps developed in this study may require reorganization, modification, or expansion beyond the 3 types presented. Like all in-depth studies, the results of this research cannot be standardized but aim to broadly capture the complex and challenging experiences of users regarding HBD analysis and use. Nevertheless, the personas and journey maps created in this study can serve as a foundational guide, gradually expanding and adapting to meet the needs of a wider range of users and data environments. Future research should conduct in-depth demand analyses targeting users of emerging data types to more comprehensively reflect diverse user experiences.

Conclusions

The ultimate purpose of precision health care is to provide optimal treatment and treatment tailored to consider individual genetic, environmental, and lifestyle factors based on health care information. This can be realized through an effective use of big data. This study proposes a human-centered design to develop an AI agent supporting the use of HBD to increase use. Based on the needs of users through shadowing and in-depth interviews, 5 key implications for HBD analysis and use support system (personal platform, navigation function, crisis management support, communication and sharing, and experts) were revealed.

In the future, we propose a follow-up study to develop a system that supports the overall process of big data analysis based on the implications derived through this study. This is expected to further increase big data use by helping many users who hesitate to use big data due to difficulties in the analysis process to analyze big data more easily.

Acknowledgments

The authors would like to acknowledge all participants who provided us with their time and shared their experiences in the interviews. This work was supported by the National Research Foundation of Korea; grant funded by the Korean government (Ministry of Science and ICT; 2021R1A2C2095271).

Data Availability

The datasets generated and analyzed during this study are not publicly available due to privacy restrictions but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. De Santis KK, Jahnel T, Sina E, Wienert J, Zeeb H. Digitization and health in Germany: cross-sectional nationwide survey. JMIR Public Health Surveill 2021;7(11):e32951 [FREE Full text] [doi: 10.2196/32951] [Medline: 34813493]
2. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of digital technologies in health care during the COVID-19 pandemic: systematic review of early scientific literature. J Med Internet Res 2020;22(11):e22280 [FREE Full text] [doi: 10.2196/22280] [Medline: 33079693]
3. Shin JY. What we have to prepare for being smart in healthcare: based on the Korean current situation and long-term direction of the national health insurance. HIRA Policy Brief. 2021. URL: https://repository.hira.or.kr/handle/2019.oak/2534 [accessed 2023-05-30]

XSL•FO
RenderX

4. Han JA, Kim EJ. Smart health care. KISTEP Technol. Trends Brief. 2020. URL: https://www.kistep.re.kr/board.es?mid=a10306060000&bid=0031&b_list=10&act=view&list_no=34992&nPage=4&keyField=&orderby= [accessed 2023-05-20]
5. Kostkova P. Grand challenges in digital health. Front Public Health 2015;3:149-199. [doi: 10.3389/fpubh.2015.00134]
6. Digital health center of excellence. US Food and Drug Administration. URL: https://www.fda.gov/medicaldevices/digitalhealth/#mobileapp [accessed 2023-06-02]
7. Choi YS. How will we respond to changes in the Age of digital technology? Global smart healthcare industry. In: Global Smart Healthcare Industry. Cheongju: Trend Publishing; Jan 12, 2016:1-12.
8. Gourraud PA, Henry RG, Cree BA, Crane JC, Lizee A, Olson MP, et al. Precision medicine in chronic disease management: the multiple sclerosis BioScreen. Ann Neurol 2014;76(5):633-642 [FREE Full text] [doi: 10.1002/ana.24282] [Medline: 25263997]
9. Garcia-Perez A, Cegarra-Navarro JG, Sallos MP, Martinez-Caro E, Chinnaswamy A. Resilience in healthcare systems: cyber security and digital transformation. Technovation 2023;121:102583. [doi: 10.1016/j.technovation.2022.102583]
10. Adams SC, Petersen C. Precision medicine: opportunities, possibilities, and challenges for patients and providers. J Am Med Inform Assoc 2016;23(4):787-790 [FREE Full text] [doi: 10.1093/jamia/ocv215] [Medline: 26977101]
11. Duffy DJ. Problems, challenges and promises: perspectives on precision medicine. Brief Bioinform 2016;17(3):494-504. [doi: 10.1093/bib/bbv060] [Medline: 26249224]
12. Jung IY, Goo WM. Data integration for healthcare innovation. Science Technology Policy. 2018. URL: https://www.ibric.org/bric/trend/bio-report.do?mode=view&articleNo=8692173&title=#!/list [accessed 2023-06-20]
13. Kim KH, Ko MS, Kim H. A study of establishment of medical CRM model in the post-corona era: focusing on the primary-level hospital. J Inf Technol Appl Manag 2021:28-12. [doi: 10.21219/jitam.2021.28.1.001]
14. Hopp WJ, Li J, Wang G. Big data and the precision medicine revolution. Prod Oper Manag 2018;27(9):1647-1664. [doi: 10.1111/poms.12891]
15. Dal Mas F, Massaro M, Rippa P, Secundo G. The challenges of digital transformation in healthcare: an interdisciplinary literature review, framework, and future research agenda. Technovation 2023;123:102716. [doi: 10.1016/j.technovation.2023.102716]
16. Santarsiero F, Schiuma G, Carlucci D, Helander N. Digital transformation in healthcare organisations: the role of innovation labs. Technovation 2023;122:102640. [doi: 10.1016/j.technovation.2022.102640]
17. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3 [FREE Full text] [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]
18. Lee D, Kim S. Digital healthcare innovation trends and policy implications. Science and Technology Policy Institute. 2018. URL: https://eiec.kdi.re.kr/policy/domesticView.do?ac=0000144167 [accessed 2023-06-13]
19. Wang L, Alexander CA. Big data analytics in medical engineering and healthcare: methods, advances and challenges. J Med Eng Technol 2020;44(6):267-283. [doi: 10.1080/03091902.2020.1769758] [Medline: 32498594]
20. Lee EJ. Applications of blockchain technology in healthcare: some notable cases abroad. HIRA. 2020. URL: https://repository.hira.or.kr/handle/2019.oak/2306 [accessed 2023-07-24]
21. Lim GC. Status and challenges of health and welfare data utilization. KIHASA 2020 Aug 01;8:2-5. [doi: 10.23062/2020.08.1]
22. Report for big data market. Big data platform newsletter. K-ICT Big Data Center. 2019. URL: https://kbig.kr/portal/kbig/knowledge/files.page [accessed 2023-07-15]
23. Hulsen T, Jamuar SS, Moody AR, Karnes JH, Varga O, Hedensted S, et al. From big data to precision medicine. Front Med (Lausanne) 2019;6:34 [FREE Full text] [doi: 10.3389/fmed.2019.00034] [Medline: 30881956]
24. Choi H, Choi YJ, Choi JM, Yoon SY. Big data education process of university optimal in the age of 4th Industrial Revolution. J Korea Institute Inf Commun Eng 2020;24:440-441.
25. Song YA. A case study on the big data analysis curriculum for the efficient use of data. J Pract Eng Educ 2020;12:23-29. [doi: 10.14702/JPEE.2020.023]
26. Park YS, Lee SJ. Study on the direction of universal big data and big data education-based on the survey of big data experts. J Korean Assoc Inf Educ 2020;24:201-214. [doi: 10.14352/jkaie.2020.24.2.201]
27. Karpathakis K, Libow G, Potts HW, Dixon S, Greaves F, Murray E. An evaluation service for digital public health interventions: user-centered design approach. J Med Internet Res 2021;23(9):e28356 [FREE Full text] [doi: 10.2196/28356] [Medline: 34494965]
28. Stickdorn M, Hormess ME, Lawrence A, Schneider J. This Is Service Design Doing: Applying Service Design Thinking in the Real World. Sebastopol, CA: O'Reilly Media, Inc; 2018.
29. Babione JN, Ocampo W, Haubrich S, Yang C, Zuk T, Kaufman J, et al. Human-centred design processes for clinical decision support: a pulmonary embolism case study. Int J Med Inform 2020;142:104196. [doi: 10.1016/j.ijmedinf.2020.104196] [Medline: 32947115]
30. Micheaux A, Bosio B. Customer journey mapping as a new way to teach data-driven marketing as a service. J Market Educ 2018;41(2):127-140. [doi: 10.1177/0273475318812551]
31. Holzinger A, Kargl M, Kipperer K, Regitnig P, Plass M, Muller H. Personas for artificial intelligence (AI) an open source toolbox. IEEE Access 2022 Feb 25;10(2):23732-23747. [doi: 10.1109/ACCESS.2022.3154776] [Medline: 20044162]

XSL•FO
RenderX

32. Canedo ED, Calazans ATS, Silva GRS, Costa PHT, Masson ETS. Use of journey maps and personas in software requirements elicitation. Int J Soft Eng Knowl Eng 2023;33(3):313-342. [doi: 10.1142/s0218194023300014]
33. Lee YH, Jang YJ, Lee SK. Obstacles to health big data utilization based on the perceptions and demands of health care workers in South Korea: web-based survey study. JMIR Form Res 2023;7:e45913 [FREE Full text] [doi: 10.2196/45913] [Medline: 37052992]
34. Glaser BA, Strauss AL. The Discovery of Grounded Theory: Strategies for Qualitative Research. 2nd Edition. London: A Division of Transaction Publishers; 2017.
35. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. J Big Data 2019;6(1):1-25. [doi: 10.1186/s40537-019-0217-0]
36. The guideline for big data analysis and usage version 1.2. K-ICT Big Data Center. 2014. URL: https://kbig.kr/portal/ [accessed 2024-12-19]
37. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
38. Williams I, Brereton M, Donovan J, Mcdonald K, Millard T, Tam A, et al. A collaborative rapid persona-building workshop: creating design personas with health researchers. Int J Sociotechnol Knowl Dev 2014;6(2):17-35. [doi: 10.4018/ijskd.2014040102]
39. Miaskiewicz T, Kozar KA. Personas and user-centered design: how can personas benefit product design processes? Des Stud 2011;32(5):417-430. [doi: 10.1016/j.destud.2011.03.003]
40. Meyer C, Schwager A. Understanding customer experience. Harvard Business Rev 2007;85:116 [FREE Full text]
41. Yang B, Liu Y, Liang Y, Tang M. Exploiting user experience from online customer reviews for product design. Int J Inform Manag 2019;46:173-186. [doi: 10.1016/j.ijinfomgt.2018.12.006]
42. He Q, Du F, Simonse LW. A patient journey map to improve the home isolation experience of persons with mild COVID-19: design research for service touchpoints of artificial intelligence in eHealth. JMIR Med Inform 2021;9(4):e23238. [doi: 10.2196/23238] [Medline: 33444156]
43. Kang HJ, Han J, Kwon GH. Determining the intellectual structure and academic trends of smart home health care research: coword and topic analyses. J Med Internet Res 2021;23(1):e19625 [FREE Full text] [doi: 10.2196/19625] [Medline: 33475514]
44. Easton K, Potter S, Bec R, Bennion M, Christensen H, Grindell C, et al. A virtual agent to support individuals living with physical and mental comorbidities: co-design and acceptability testing. J Med Internet Res 2019;21(5):e12996 [FREE Full text] [doi: 10.2196/12996] [Medline: 31148545]
45. Kim B. Big data platform case analysis and deployment strategies to revitalize the data economy. Converg Secur J 2021:21-78. [doi: 10.33778/kcsa.2021.21.1.073]
46. McPadden J, Durant TJ, Bunch DR, Coppi A, Price N, Rodgerson K, et al. Health care and precision medicine research: analysis of a scalable data science platform. J Med Internet Res 2019;21(4):e13043 [FREE Full text] [doi: 10.2196/13043] [Medline: 30964441]
47. Kim CW. Digital healthcare platform initiatives also designed from the perspective of users, not providers. Digital Health Innov 2016;215(2):116-126 [FREE Full text]
48. Zafeiridi P, Paulson K, Dunn R, Wolverson E, White C, Thorpe JA, et al. A web-based platform for people with memory problems and their caregivers (CAREGIVERSPRO-MMD): mixed-methods evaluation of usability. JMIR Form Res 2018;2(1):e4 [FREE Full text] [doi: 10.2196/formative.9083] [Medline: 30684403]
49. Shea B, Bakre S, Carano K, Scharen J, Langheier J, Hu EA. Changes in glycemic control among individuals with diabetes who used a personalized digital nutrition platform: longitudinal study. JMIR Diabetes 2021;6(4):e32298 [FREE Full text] [doi: 10.2196/32298] [Medline: 34661545]
50. Iorfino F, Ho N, Carpenter JS, Cross SP, Davenport TA, Hermens DF, et al. Predicting self-harm within six months after initial presentation to youth mental health services: a machine learning study. PLoS One 2020;15(12):e0243467 [FREE Full text] [doi: 10.1371/journal.pone.0243467] [Medline: 33382713]
51. Yan X, Yan J. Design and implementation of interactive platform for operation and maintenance of multimedia information system based on artificial intelligence and big data. Comput Intell Neurosci 2022;2022:4620930 [FREE Full text] [doi: 10.1155/2022/4620930] [Medline: 35571710]
52. Kneuertz PJ, Jagadesh N, Perkins A, Fitzgerald M, Moffatt-Bruce SD, Merritt RE, et al. Improving patient engagement, adherence, and satisfaction in lung cancer surgery with implementation of a mobile device platform for patient reported outcomes. J Thorac Dis 2020;12(11):6883-6891 [FREE Full text] [doi: 10.21037/jtd.2020.01.23] [Medline: 33282391]
53. Torous J, Wisniewski H, Bird B, Carpenter E, David G, Elejalde E, et al. Creating a digital health smartphone app and digital phenotyping platform for mental health and diverse healthcare needs: an interdisciplinary and collaborative approach. J Technol Behav Sci 2019;4(2):73-85 [FREE Full text] [doi: 10.1007/s41347-019-00095-w]

Abbreviations

AI: artificial intelligence

XSL•FO
RenderX

HBD: health big data

Edited by A Mavragani; submitted 07.10.24; peer-reviewed by A Holzinger, L He; comments to author 15.11.24; revised version received 28.11.24; accepted 28.11.24; published 08.01.25.

Please cite as:
Lee YH, Choi H, Lee SK
Development of Personas and Journey Maps for Artificial Intelligence Agents Supporting the Use of Health Big Data: Human-Centered Design Approach
JMIR Form Res 2025;9:e67272
URL: https://formative.jmir.org/2025/1/e67272
doi:10.2196/67272
PMID:

©Yoon Heui Lee, Hanna Choi, Soo-Kyoung Lee. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effect of a Narrative-Based Online Course Aimed at Reducing Stigma Toward Transgender Children and Adolescents: Longitudinal Observational Study

Merlin Greuel[1], MD, MGH; Van Kinh Nguyen[1], PhD; Doron Amsalem[2], MD; Maya Adam[1,3,4*], MD, PhD; Till Bärnighausen[1,5,6*], MD, MSci, MSci, SCD, Dr med

[1]Heidelberg Institute of Global Health, Heidelberg University, Heidelberg, Germany

[2]Department of Psychiatry, Irving Medical Center, Columbia University, New York, NY, United States

[3]Department of Pediatrics, Stanford School of Medicine, Stanford University, Stanford, CA, United States

[4]Center for Digital Health, Stanford School of Medicine, Stanford University, Stanford, CA, United States

[5]Department of Global Health and Population, Harvard T H Chan School of Public Health, Harvard University, Cambridge, MA, United States

[6]Africa Health Research Institute, Somkhele, South Africa

[*]these authors contributed equally

Corresponding Author:
Merlin Greuel, MD, MGH
Heidelberg Institute of Global Health
Heidelberg University
Im Neuenheimer Feld 130/3
Heidelberg, 69120
Germany
Phone: 49 6221565344
Email: greuel@stud.uni-heidelberg.de

Abstract

Background: Stigma toward transgender children and adolescents negatively impacts their health and educational outcomes. Contact with members of stigmatized groups can dismantle stereotypes and reduce stigma by facilitating exposure to the unique cognitive and emotional perspectives of individuals within the group. Recent evidence suggests that video-based contact interventions can be as effective as face-to-face encounters, but challenges lie in protecting the identities of transgender youth, since many of them live in stealth.

Objective: This study aims to evaluate the impact of an animated online course, rooted in authentic, personal narratives, on course participants' stigma toward transgender youth.

Methods: The online course was offered free of charge on Coursera and contained 19 teaching videos (3-7 minutes each), intermittent practice quizzes, and discussion prompts. Using real voice recordings of transgender children and their caregivers, the videos were designed to elicit empathy and transmit knowledge. All videos conveying the narratives of transgender youth were animated to protect their identities. A total of 447 course participants, distributed around the globe, completed pre- and postcourse surveys. While the course primarily targeted parents and caregivers of transgender youth, it was open to anyone with a Coursera account. The survey was based on the Transgender Attitudes and Beliefs Scale but modified to reflect the context of parents and caregivers. Using a 5-point Likert scale, it contained 5 questions that captured participants' levels of transgender stigma. Results of the pre- and postcourse surveys were then compared.

Results: The results were obtained in January 2023. Baseline levels of stigma were relatively low (18/25 across all questions, with 25 representing the lowest possible levels of stigma) and decreased further after completion of the course (to 19/25 across all questions, $P<.001$). A multivariate ordinal probit regression showed that, depending on the question, participants were 7%-34% more likely to endorse statements that indicated the lowest levels of stigma after completing the course. The course was equally effective across all demographics represented in our participant population.

Conclusions: Our findings document a significant reduction in stigma toward transgender youth in participants who chose to enroll in the first animated, open online gender health course, rooted in the authentic narratives of transgender youth. Stigma levels decreased significantly after taking the course, even among participants whose baseline levels of stigma were low. Future

XSL•FO
RenderX

interventions should include participants with more variable baseline levels of stigma, ideally in the setting of a randomized controlled trial. Despite its limitations, this evaluation adds to the existing evidence that digital, contact-based antistigma interventions, animated to protect the identity of the narrators, can effectively reduce stigma toward transgender youth.

(JMIR Form Res 2025;9:e59605) doi:10.2196/59605

KEYWORDS

stigma; transgender; children; adolescents; mental health; online course evaluation; entertainment-education; narratives

Introduction

Adolescence is a period marked by change, self-discovery, and significant vulnerability to mental health issues [1]. Children and young adults who realize that their gender identity differs from their sex assigned at birth are especially prone to poor mental health outcomes: transgender youth endure elevated rates of depression, self-harm, eating disorders, and suicidality compared with their nontransgender peers [2-4]. In addition, the educational outcomes of gender-diverse children often lag behind those of their cisgender peers [5]. Thus, the stressors that affect gender-diverse children and adolescents can have a profound effect on individuals' academic performance and the course of their future lives.

Stigma toward transgender youth contributes to the observed differences in mental health. According to the gender minority stress theory, many of the health disparities observed between transgender and cisgender individuals can be attributed to stressors emerging from the stigma faced by transgender people, also called "transphobia" [6,7]. The chronic stress associated with transphobia results in poor mental health outcomes with repercussions even on physical health, as evidenced by elevated rates of diabetes, hypertension, and death by suicide [7]. This is especially true for individuals with multiple disadvantaged statuses such as low-income, transgender women of color [7,8]. Stigma can affect individuals' health directly, for example, through chronic activation of the body's stress response and increased cortisol output [7], or indirectly by decreasing access to the housing market, education, and employment [7,9], all critical social determinants of health.

Depending on the academic discipline, stigma has been studied from different angles, using diverse terminology [7,9]. Stigma may be defined as "a severe social disapproval due to believed or actual individual characteristics, beliefs or behaviors that are against norms, be they economic, political, cultural, or social" [10]. More recently, neuroscience has offered a fresh perspective on stigma, describing its cognitive, emotional, and behavioral dimensions [11-13]. Stigma can be considered a defensive reaction resulting from an innate fear of individuals who differ from the self and collectively form an "outgroup" [11]. Functional neuroimaging studies show that stigma [14,15] and perceived outgroup membership [16,17] can decrease the levels of empathy felt for others. As a consequence, members of the outgroup are depersonalized and viewed as more homogeneous than ingroup members, thus encouraging prejudice, negative stereotyping, and antisocial behaviors [18].

Contact with a member of a stigmatized community can dismantle stereotypes and facilitate entry of the previously stigmatized person into the in-group [19,20]. Intergroup contact is considered the most effective means of reducing stigma [20,21]. Experimental evidence suggests that video-based contact interventions may be as effective as face-to-face encounters while being less expensive and easier to disseminate [20,22-24]. The conceptual framework of Jankowski et al [20] for social contact-based video interventions highlights the importance of addressing the cognitive, emotional, and behavioral dimensions of stigma [20]. By targeting all of these components, viewers can empathize and identify with the protagonists, thereby reducing intergroup anxiety [20] and with it the perceived barriers between groups.

By allowing the public to experience the cognitive and emotional perspectives of transgender youth, authentic, personal narratives lend themselves to antistigma interventions [18]. Narratives can provide a nonthreatening context to establish vicarious social contact with others who might previously have been met with anxiety or mistrust. Narrative stories have been shown to elicit more compassion, favorable attitudes, and beneficial behavioral intentions toward stigmatized groups when compared with nonnarrative stories [25]. High capacity for engagement, as well as the compassionate and affective reactions evoked by the narrative format, are thought to trigger empathic attitudes in the audience [25] that may translate into nonstigmatizing behavior. Through empathy and knowledge transfer, transgender individuals may leave the "outgroup" and enter the "ingroup," allowing their nontransgender peers to view and treat them as one of their own. Yet, few studies have explored the use of video-based narratives for reducing the stigma against transgender youth. These studies found that social contact–based videos have the potential to increase knowledge about transgender youth and to decrease transphobia [26,27]. However, care needs to be taken to protect the identities of transgender youth who share their stories, since many of them live in stealth [28]. To our knowledge, no study has examined the effect of integrating animated short videos, featuring authentic voices and narratives, into an online course aimed at reducing transphobia. Thus, this article contributes to the existing literature by integrating evidence-based methods to reduce transgender stigma into a novel, digitally scalable format that protects the identities of the narrators. The benefit of an online course over interventions with exposure to a single video is the possibility to expose participants to content over a longer time period, enabling a more in-depth discussion of course material, repetitions of core concepts, and assessments that stimulate subjects to actively engage with the topic of interest.

Here, we evaluate the effect of the online course "Health Across the Gender Spectrum," rooted in entertainment narratives, on participants' attitudes toward transgender youth. The aim of the

XSL•FO
RenderX

course was to promote a gender-inclusive society by establishing virtual contact with the transgender community. The narratives were unique in that they included real voice recordings of transgender children as well as their parents and caregivers and they were animated to conceal the identities of the storytellers. The course was made freely available to the public, with some targeting toward parents and caregivers of children whose gender identity differs from their sex assigned at birth. This target group was chosen because adolescents' "coming out" as transgender can significantly deteriorate family relations [2] due to a lack of acceptance of family members. Family rejection of transgender children and adolescents is strongly correlated with mental health issues, substance use, suicidality, and sexual risk [29-31]. On the other hand, family support of transgender youth protects against depression and improves their quality of life [32]. We analyzed pre- and postcourse survey data to determine whether participants' attitudes toward transgender youth changed after completing the course. Our hypothesis was that subjects would show lower levels of stigma toward transgender youth upon completion of the course. This evaluation may help pave the way for other innovative, empirically verified interventions to reduce family and caregiver stigma toward transgender individuals, thus improving their health and life outcomes [29-32].

Methods

Evaluation Design and Approval

This online course evaluation was designed as a virtual contact-based exposure study with a pre-post design, using a convenience sample. The CHERRIES (Checklist for Reporting Results of Internet E-Surveys) [33] was used to guide this article.

Online Course and Participants

The course "Health Across the Gender Spectrum," still available free of charge on the Coursera online platform as of publication [34], was set up as a 3-week program (8 hours of instruction in total) consisting of 3 modules: "What is Gender Identity?" "What is the Gender Spectrum?" and "How Do We Create a Gender-Inclusive Society?" Coursera is a global massive open online course provider based in the United States and used by universities and similar organizations to offer online courses, certificate programs, and degree courses, some of which are free of charge while others require a tuition fee. Participants had no fixed time schedule and could decide the pace of their progress on their own. The language of instruction was English. Conceived by the Stanford School of Medicine under the leadership of our coauthor (MA), learners were introduced to the experiences of 6 transgender children and their families. Through illustrated stories and 19 short teaching videos, participants explored the topics of gender identity and the gender spectrum. Stanford physicians, K-12 educators, and transgender faculty members offered practical advice for parents, health care providers, teachers, and others interested in exploring gender identity in children and adolescents.

Apart from a Coursera account, there were no prerequisites for enrollment. That is, while the course was geared toward parents and caregivers of transgender and gender-diverse children and adolescents, anyone was eligible to enroll. The course was announced to learners enrolled in the Stanford Child Nutrition and Cooking Course, also available on Coursera. Coursera also independently announces new courses to registered account holders. Each of the 3 modules was divided into subsections that sought to answer common questions such as "What's the difference between gender identity and sexual orientation?" The answers were then provided by a series of short, animated educational videos, interspersed with the narrative videos. Each subsection contained one practice quiz with one or two true-false questions, which required a response to move on to the next video. Before advancing to the next subsection, learners had the option to answer a voluntary discussion prompt. At the end of the last section of each module, participants were asked to complete a graded quiz of 4-6 questions. A score of 80% or higher was required to move on. Learners could repeat the quiz as often as they wished. An overview of all modules, sections, and titles of the videos and discussion prompts can be found in Multimedia Appendix 1.

We did not store any participant information other than the survey responses submitted by the learners. We had access to anonymized data from Coursera, such as the IP addresses of those who submitted surveys. These data did not personally identify any participants. Subjects were informed about the purpose of the surveys and their approximate length (3 minutes).

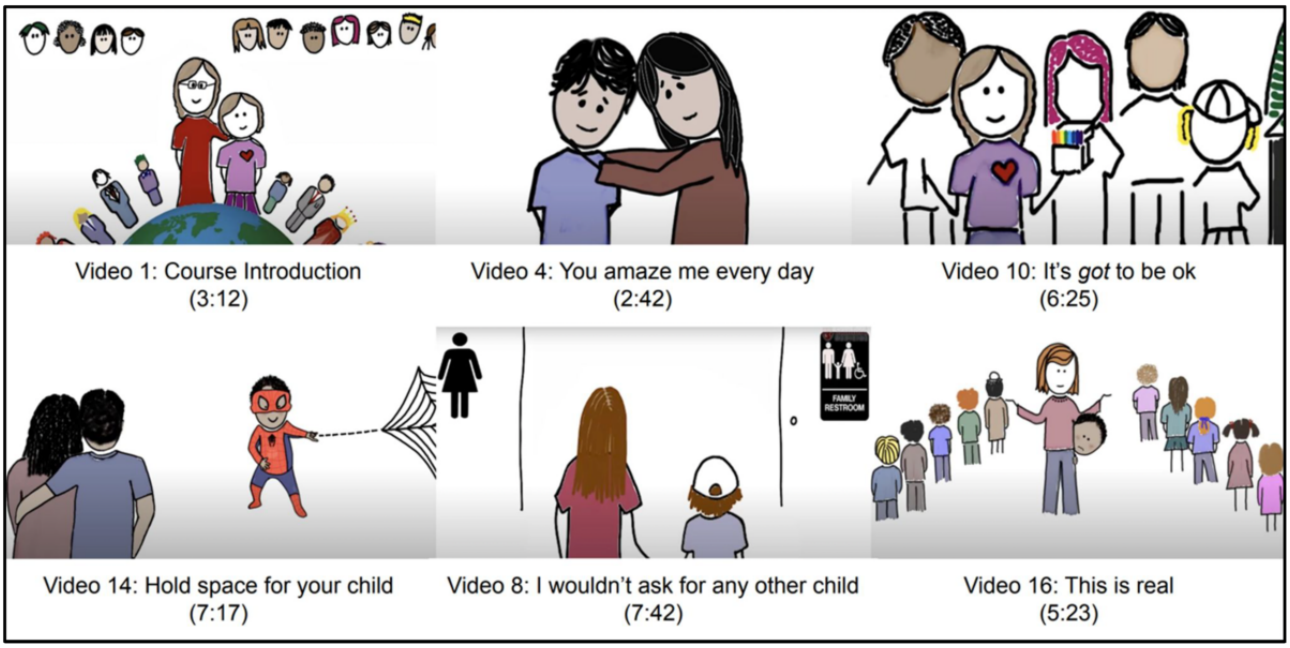
Video Animations

A total of 19 teaching videos (3-7 minutes each) were developed by our coauthor (MA) at the Stanford School of Medicine, with input from faculty at the Pediatric and Adolescent Gender Clinic. The videos featured universally accessible animations, appropriate for all cultural, ethnic, and racial contexts. This raised the likelihood that viewers could identify with the protagonists no matter their own cultural identity and origin. In some of the videos, MA interviewed faculty experts from the Stanford Pediatric Gender Clinic as well as 2 transgender faculty members, who appeared in person. The transgender faculty members shared personal anecdotes and experiences relevant to the course theme. However, depending on the 3 modules of the course, the animated content predominated and ranged between 75% and 100%. The narrative videos' hallmark was their story-telling approach, using the real voices of transgender children, parents, and caregivers to make the animations come alive. Rather than being didactic, these videos were designed to establish vicarious social contact between the viewers and transgender children as well as their families and caregivers, in a nonthreatening context. The use of animation effectively protected the identity of the young people who shared their stories and allowed us to visually emphasize the importance of diversity and inclusion in all vignettes. The emotional appeal of the video narratives relied heavily on the characters' descriptions of their feelings, which were mirrored in the animations and underscored by customized soundtracks. The aim was to enable viewers to experience the cognitive and emotional perspectives of the protagonists, thereby triggering compassionate and affective reactions and empathy. Some of the videos focused more on individuals' personal experiences, while others transmitted knowledge about gender identity and related topics. The language was simple and avoided medical jargon to reach viewers with diverse educational backgrounds.

XSL•FO
RenderX

Figure 1 shows select scenes from the narrative videos. A complete list of course videos can be found in Multimedia Appendix 2.

Figure 1. Sample scenes from narrative course videos with titles and run-times.



Survey

After enrollment, participants received an email in which they were asked to complete an open precourse survey. The survey was hosted by Stanford School of Medicine's Qualtrics account and appeared on a single page. Demographic information was collected on the second page. Qualtrics is a platform for online experiments and data collection [35]. To match pre- and postcourse surveys, we used Coursera's user ID hashtags. In addition, participants were asked to generate a unique survey ID as a backup. We also used IP addresses, locations, demographic information, and dates of submission to identify any falsely matched surveys or multiple submissions. The questionnaire was informed by the Transgender Attitudes and Beliefs Scale (TABS) [36]. We chose to abbreviate and modify the TABS for the context of parents and caregivers because the original TABS contains 29 items, which we deemed too long for a survey taken on a voluntary basis, with no financial incentives or compensation. Our questionnaire contained 5 items aimed at measuring preexisting stigma against transgender children and adolescents. The statements were based on the first 2 subscales of the TABS, which explore interpersonal comfort with transgender individuals and sex or gender beliefs. We modified the items to better reflect the perspective of parents and caregivers of transgender children and adolescents. For example, one of the items in the original TABS reads "All adults should identify as either male or female" (Q3.13, subscale 2) [36]. We changed this item to "It is important that all children learn to identify themselves as either a boy or a girl" (item 5 in our questionnaire). Participants responded through a 5-point Likert scale and an additional not applicable option. The answer choices ranged from "strongly disagree" to "strongly agree." Please refer to Multimedia Appendix 3 for the complete questionnaire, and to Multimedia Appendix 4 for an item-by-item overview of how the original TABS items were modified. The technical functionality of the survey was tested before its launch. The data were collected between 2017 and 2023.

After completing the course, learners received another email asking them to fill out the same questionnaire, thus allowing for pre-post comparisons. In both surveys, participants were instructed to generate a unique, anonymous identifier to match the surveys in the analysis.

Data Analysis

Whenever possible, the pre- and postcourse surveys were matched for each participant, creating a fully matched sample of individuals. We used a multivariate ordinal probit regression for matched data, in which the 5 questions were modeled jointly [37]. This approach treats the Likert response as an ordinal scale providing a better theoretical interpretation and statistical inferences (compared with a nonparametric or normally distributed response assumption, eg, in a change point analysis) [38]; furthermore, potential correlations between the questions are taken into account [39]. Matched individuals were modeled with a random intercept to allow for individual heterogeneity (Multimedia Appendix 5). An ordinal probit regression can model the relationship between one ordinal-dependent variable (in our case, the answer choices of the Likert scale) and one or more independent variables [40]. A multivariate ordinal probit model also takes into account the potential correlation among the responses across the 5 questions [37]. First, we defined that an answer was in the "positive" direction when it reflected low or very low levels of transgender stigma. For the first 3 questions, a "positive" answer meant that participants agreed (score of 4) or strongly agreed (score of 5) with the respective statements; for the last 2 survey items, a response in the "positive" direction was reflected by respondents' disagreeing (score of 4) or strongly disagreeing (score of 5) with the

XSL•FO
RenderX

statements. We then calculated the cumulative probabilities for each Likert category in the pre- and postcourse surveys; in other words, we determined the likelihood that a subject chose a specific answer (Likert category) for each question. Finally, we calculated the ratio between the probabilities for each question-and-answer choice (Likert category) in the post- versus precourse survey. This ratio describes the relative effect of the intervention. For instance, a relative effect (ratio) of "1" indicates that respondents' likelihood of choosing a respective answer choice after completing the course was just as high as before the course. A relative effect of "0.5" means that participants were half as likely to choose a respective answer after completing the course, compared with the precourse condition. Conversely, a relative effect of "2" indicates that subjects were twice as likely to select a respective answer choice after the course.

Since respondents often included erroneous identifiers, not all surveys could be matched. Thus, as a robustness check, we performed a secondary analysis on a propensity score 1:1 matched sample [41] (Multimedia Appendix 6). We used a probit regression as described above. The matching procedure used nearest neighbor matching and sampling with replacement [42].

Ethical Considerations

We obtained ethics approval from Stanford University's institutional review board (protocol #76972). Completion of the survey was optional and not a prerequisite for beginning the course, and no incentives were offered to participants. The data collected were anonymized and did not allow for the personal identification of subjects.

Results

Demographics

By January 2023, 3878 learners completed the course. Of these, we obtained 1591 survey records. A total of 447 participants (12% of all 3878 learners) could be fully matched based on subjects' unique survey IDs. Coursera's user-ID hashtags were not used due to a technical issue on the part of the platform. We also used data such as IP addresses, location, submission dates, and demographic information to identify potential mismatches or multiple submissions between pre- and postcourse surveys. If we found any mismatch or multiple submissions, we did not include the submission. Of the 447 learners, 377 (84%) were female, 406 (91%) had college-level education, and 283 (63%) identified as White. A total of 229 (51%) of participants completed the survey in 2017. A total of 200 participants (45%) completed both, the precourse and the postcourse survey within one day. Table 1 shows an overview of the subjects' social demographics for the matched sample. Figure 2 depicts the geographic distribution of all recorded participants, containing both pre- and postsurvey data.

XSL•FO
RenderX

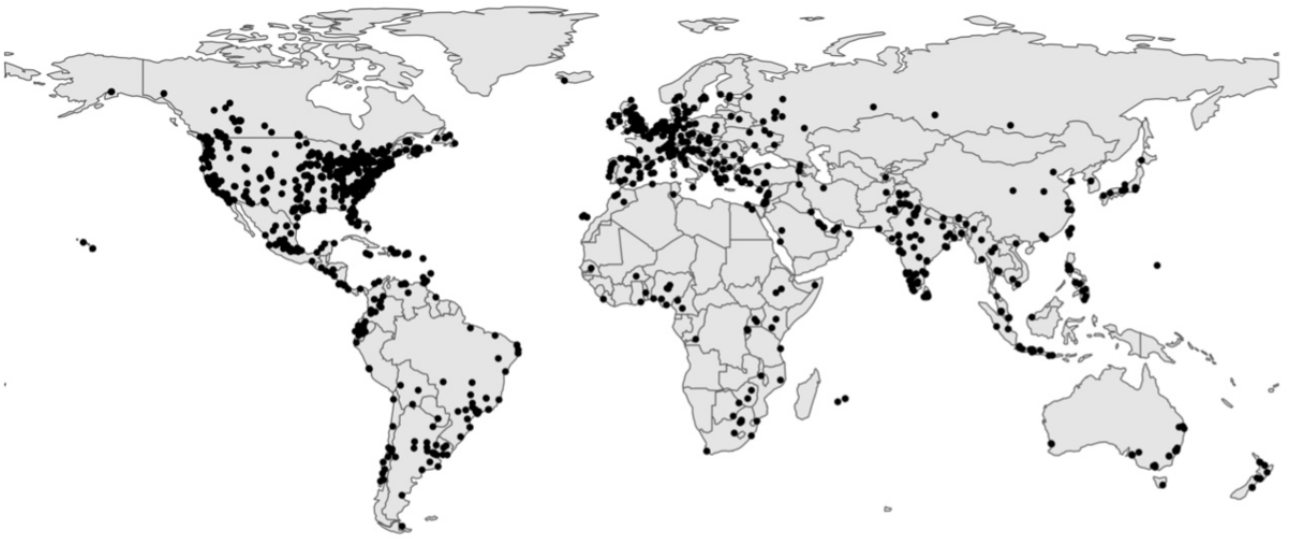
Table 1. Descriptive social demographics and responses of the fully matched sample.

Characteristics	Values
Age (years), mean (SD)	37.29 (14.85)
Sex, n (%)	
Female	377 (0.84)
Male	47 (0.11)
Gender-diverse	23 (0.05)
Education, n (%)	
College	406 (0.91)
High school	41 (0.09)
Race, n (%)	
Asian	45 (0.10)
Black	14 (0.03)
Latino	79 (0.18)
Mixed or other	26 (0.06)
White	283 (0.63)
US-based, n (%)	
No	280 (0.63)
Yes	167 (0.37)
Year of survey completion, n (%)	
2017	229 (0.51)
2018	55 (0.12)
2019	17 (0.04)
2020	70 (0.16)
2021	37 (0.08)
2022	28 (0.06)
2023	11 (0.02)
Time between pre- and postsurveys, n (%)	
Day	200 (0.45)
Week	66 (0.15)
Month	129 (0.29)
Month or more	52 (0.12)
Total score, median (IQR)[a]	
Precourse	18 (14-20)
Postcourse	19 (16-20)

[a]Wilcoxon signed rank *P* value <.001.

XSL•FO
RenderX

Figure 2. Geographic distribution of the sample, showing both pre- and postcourse participants, including those that could not be matched regarding their pre- and postcourse surveys.



Of the participants in the fully matched sample, 167 (36%) were located in the United States, the country most highly represented in the course. The US-based subjects were distributed across the country, following a pattern that roughly reflects the country's population density (Figure 2). However, 280 (63%) learners were located outside of the United States, distributed across all continents with larger clusters in Europe and Latin America.

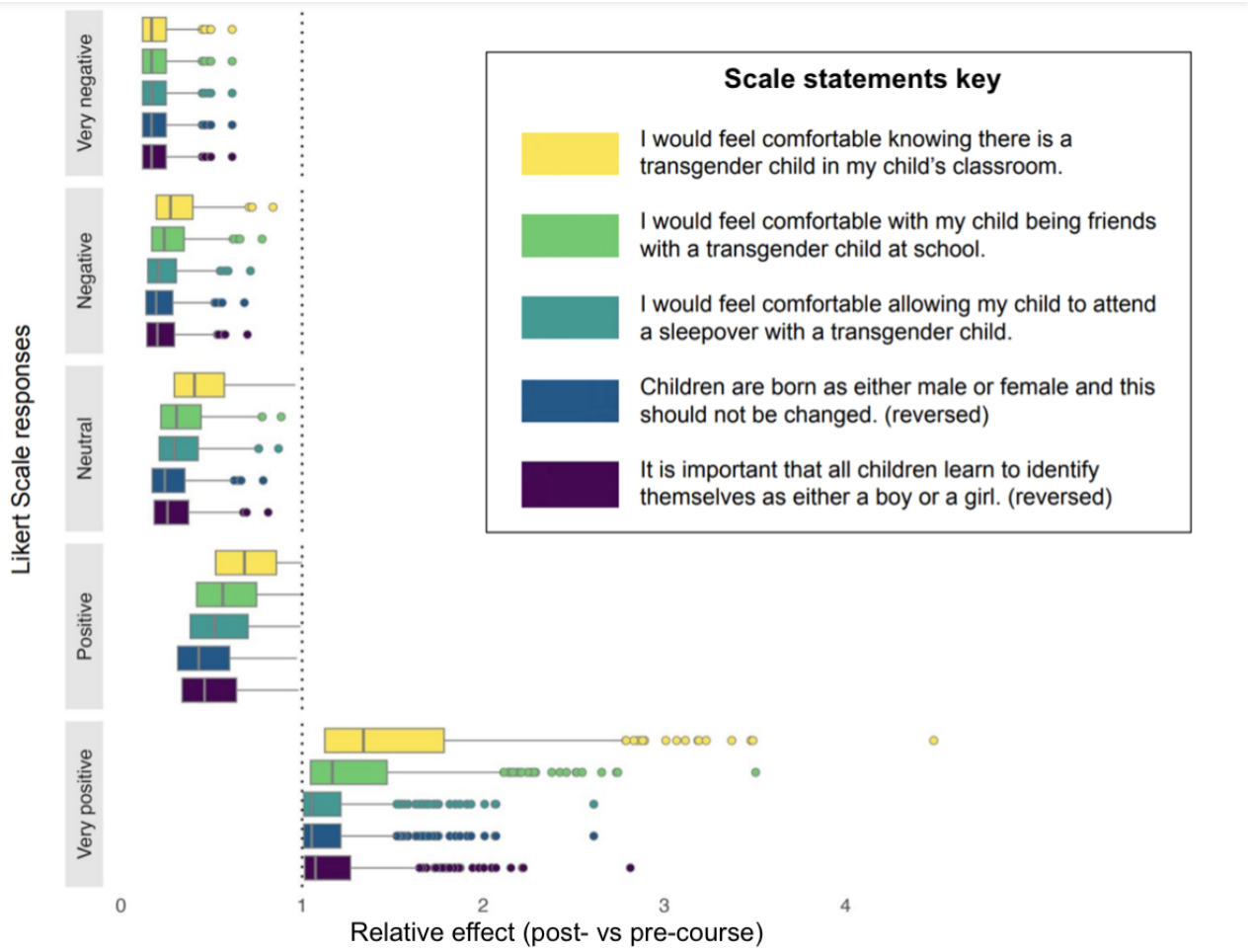
Levels of Stigma

The 5 questions had a Cronbach α of 0.83, reflecting the questionnaire's high level of internal consistency. Given that each of the 5 questions allowed for a maximum possible stigma score of 5 (and a minimum of 1), the maximum possible score for all 5 questions combined (indicating the lowest possible levels of stigma) was 25. The median of the total score across all 5 questions before the course was 18 (IQR 14-20). This median increased significantly to 19 (IQR 16-20) after the course (Wilcoxon signed rank test, $P<.001$).

Figure 3 shows the effect of the course on the fully matched data, showing separate results for each of the survey items. After the course, participants had a higher likelihood of answering very positively (strong agreement with the first 3 statements, strong disagreement with the remaining 2 items). At the same time, they were also less likely to provide negative, neutral, or less positive answers. The effect of the course varied largely between individual participants. While acknowledging that our sample was quite homogeneous in terms of demographic variables (refer to Table 1 and Discussion below), we found no significant differences in efficacy across the demographics represented in our participant population. The results of the secondary analysis on a propensity score 1:1 matched sample were consistent with those presented above (Multimedia Appendix 6).

XSL•FO
RenderX

Figure 3. Marginal effect of the course on the responses to the 5 questions. The relative effect was computed as the ratio between the postcourse and the precourse predicted probabilities for each survey item.



Discussion

Principal Findings

In this survey-based evaluation of the social contact–based online course Health Across the Gender Spectrum, we observed a significant decrease in participants' stigma toward transgender children and adolescents. Baseline levels of stigma were low but decreased further upon completion of the course. This effect was observed for all survey items and across all demographic groups. Despite its limitations, the findings of this evaluation provide the first documented evidence that an online course rooted in animated narratives may effectively reduce participants' stigma toward transgender youth. In addition, the fact that the online course reached subjects across all levels of education, gender, age, ethnicity, or location is another important revelation of this study with the potential to guide future interventions.

The course significantly reduced stigma toward transgender children and adolescents across all question items, despite low baseline stigma levels. After completion of the course, participants were more likely to respond "very positively" and less likely to choose answers in the other categories. This means that there was a general shift in responses toward the "very positive" category, which reflected the lowest levels of stigma. This effect was observed for all questions but especially pronounced for the statements regarding the innateness of gender and the importance that children learn to identify as a boy or a girl. These 2 statements were arguably the most complex to understand for a learner who had not been exposed to the course. Moreover, they represented the highest level of escalation together with the question regarding the acceptability of sleepovers with transgender children, meaning that they had been designed to maximize the discomfort of participants with high levels of stigma. The fact that the strongest change was observed for these statements implies that subjects with higher preexisting levels of stigma or less informed participants may have benefited the most from the course.

Not surprisingly, learners who enrolled in the course showed relatively low levels of transgender stigma even before completing the intervention. This makes sense because the course likely attracted learners with socially progressive attitudes and an openness to learn more about transgender individuals. This is a limitation of this study, as individuals with higher levels of transphobia are presumably less likely to enroll in this course. Nonetheless, the fact that we were able to document a measurable reduction in stigma, even among this enlightened group, suggests that the stigma-reducing effect of the course could be even greater in populations that are less accepting of gender diversity in children. Future studies, ideally in the form of randomized controlled trials, may explore the stigma-reducing effect of this course or similar content within a more

XSL•FO
RenderX

heterogeneous sample that is representative of the population as a whole.

The findings of our evaluation confirm previous research that demonstrated the stigma-reducing effect of social contact–based antistigma interventions [26,27]. This article adds to the literature by demonstrating that an online course with real learners (rather than participants recruited in a controlled experimental setting) achieved similar results. Furthermore, this is the first open online gender health course to experiment with the use of animation to protect the identities of transgender youth sharing authentic stories. Research in the field of neuroscience may help to explain why the course was effective. Stigma is thought to encompass a cognitive, emotional, and behavioral dimension [11-13]. The power of social contact–based video interventions is that they can address all of these components [20]. Emotion is a core feature of stigma, which often manifests as reflexive disgust, fear, and disapproval [11]. The brain regions responsible for emotional responses, the amygdala and insula, generally have a shorter latency to respond to aversive stimuli (eg, those inducing disgust) than cortical regions [11,43,44]. Thus, with the emotional response preceding cognitive modulation, emotions play a primary role in shaping individuals' proneness to exert stigmatizing behavior.

Direct contact with a member of a stigmatized community can reduce the negative emotions associated with stigma [20]. However, virtual contact may also be effective, as exemplified by an intervention that established intergroup contact with individuals enduring schizophrenia using an electronic chat [20,45]. Participants who had been in virtual contact with a schizophrenia patient exhibited lower levels of anger, fear, and stereotyping compared with those who had been in virtual contact with a person who did not suffer from a mental illness [45]. In fact, even just imagining a positive encounter with a schizophrenia patient can reduce anxiety and negative stereotypes [46], demonstrating the power of social contact—be it direct, virtual, or imagined—in shaping our emotions. In a similar way, the video-based virtual contact in our online course might have alleviated the negative emotions associated with transgender stigma, thereby contributing to the reduction in stigma scores observed in the postsurvey.

While emotions are key, stigma is more than just an affective response. It also includes cognition. Implicit stigma describes attitudes and biases that are outside of conscious control [47,48]. It is impulsive and depends on associative links [11]. In contrast, explicit stigma occurs within conscious awareness and is self-reportable and regulatable through cognitive processes [11,48]. Explicit stigma relies on factual knowledge and reflection [11]. Thus, an effective contact-based online course should also transmit knowledge. Contact-based interventions have proven effective when they portray characters that deviate from negative stereotypes in moderate ways; that is, by focusing on the characters' struggles and stories of hope instead of their accomplishments [20]. This is precisely what our videos do. In the video segments, animated children (narrated by real transgender youth) describe their journeys—their initial doubts about their gender identities, their feelings of shame and guilt, their fear of "coming out" and all of its consequences. Interviews with 2 transgender Stanford faculty role models give caregivers and parents hope by sharing how they successfully navigate the challenges of being transgender in academia. Finally, short clips embedded between the video sequences of transgender experiences also transmit facts about the concepts of gender identity, sexual orientation, the gender spectrum, the medical classification of transgender, gender-affirming management options, and the creation of a gender-inclusive society. A lack of subject knowledge supports the formation of stigma [49], so the aim of these supplementary videos was to provide facts that were easily comprehensible to a global audience. Thus, the effect of our video-based online course may have been underscored by the fact that its contact-based approach addressed both the emotion and cognition underlying stigma. The interaction between emotion and cognition gives rise to the concept of empathy—the ability to be sensitive to, comprehend, and vicariously experience the thoughts and feelings of others.

Fear of individuals who differ from the self are categorized internally into an "outgroup" [11]. Those who exert stigmatizing attitudes view "them" as decisively different from "us" [50,51]. This assumption prevents individuals from identifying with others: people use their self-knowledge, mental states, and experiences to infer what others might think or feel [50]; that is, to engage in the theory of mind [52]. If others are deemed fundamentally different, self-referential thinking is reduced [50], and with it, empathy. A neurobiological correlate of stigma is a reduction in the activity of the medial prefrontal cortex [52], a part of the brain that is thought to modulate empathy as it is involved in judgments about others and in viewing others as "human" [52]. As a consequence, members of the outgroup are depersonalized and viewed as more homogeneous than ingroup members, thus encouraging prejudice, negative stereotyping, and antisocial behaviors [18].

Since neural projections connect the amygdala and insula (responsible for emotion) to the ventral medial prefrontal cortex (modulating empathy), it is likely that affective responses are integrated into the theory of mind and empathy processes [52]. Moreover, emotions and empathy influence stigma-related cognition and behavior (and vice versa) [52,53] through a complex interplay involving other areas of the brain: cortical structures such as the temporal lobes and the inferior frontal gyrus encode and store stereotype concepts as semantic or social object (person) memory, which can be selected and transferred into working memory to influence judgments and behaviors [52]. The anterior temporal lobe, particularly involved in the representation of social knowledge (attributes describing people), is densely interconnected with the medial prefrontal cortex (the neural center for empathy) and the amygdala (implicated in emotion) [52,54]. Thus, it becomes evident that the various subcomponents of stigma—implicit versus explicit, with emotional, cognitive, and behavioral elements—can be traced to neurobiological correlates that work in a concerted action, determining individuals' capacity for empathy.

The videos in our course were designed to trigger a strong empathic response in viewers, containing emotionally salient stimuli such as the use of transgender children's original voices that explained their experiences, frustrations, and feelings of distress. Compelling visuals and soundtracks were intended to further support an empathic response in viewers. At the same

XSL•FO
RenderX

time, the knowledge transmitted throughout the course provided participants with the necessary context to understand the medical and social background of transgender adolescents. By acknowledging the multidimensional nature of stigma, the course notably decreased participants' stigma against transgender children and hopefully contributed to reducing the third component of stigma—the behavioral (enacted) one. Future studies designed as more rigorous randomized controlled trials may build on this article and include a behavioral endpoint. Furthermore, the emotional and empathic response of viewers may be tested in laboratory-based fMRI studies or by using online platforms that track viewers' facial reactions through a webcam. The former may elucidate the real-time impact of the intervention on the brain regions involved in stigma in a limited number of individuals, while the latter may be easier to implement in large-scale interventions with a greater sample size.

Limitations

This evaluation has several limitations. First of all, participants were matched ad hoc, resulting in a large number of subjects that could not be matched. Thus, the size of the matched sample represented only a third of the overall number of records. We partly alleviated this issue by performing a secondary analysis using propensity score matching to include more participants. This analysis supported our results. Second, our evaluation lacked a control group. Future studies should include a control group that is not exposed to the course. In addition, the time period between the pre- and postcourse surveys was variable among participants. It is unclear how this may have affected subjects' levels of stigma, although variable completion rates likely reflect the reality of most online courses. Another limitation was that participants enrolled voluntarily in the course. Learners who enrolled in the course likely shared characteristics that differ from the population as a whole, such as holding more socially progressive attitudes or an elevated interest in transgender issues. This is reflected by the fact that our sample was largely female, college-educated, and White. Those who enrolled had low baseline levels of stigma, and the course was likely unappealing to individuals with high levels of transphobia. Future studies should ensure random sampling and the inclusion of participants with diverse attitudes toward the transgender community and more balanced demographic features. Furthermore, as described above, stigma is inherently complex, and the questionnaire used is a simplistic method that might not capture the full dimension of stigma. The questionnaire had not been validated before this evaluation. We found no validated surveys that were appropriate for measuring transphobia toward youth within the parent-child context of the course content. To the best of our knowledge, there are no widely used, validated surveys measuring transphobia toward children and adolescents within this context. Another limitation of this intervention is that it remains unclear how much of the course's effect may be attributed to the narrated, animated content versus the appearance of guest speakers, discussion prompts, or quizzes. Yet, given that the majority of the content was animated, it is likely that the narrated, animated format of the course was the main active ingredient in the outcomes observed here. Finally, our intervention lacked any follow-up assessment. Future studies should include follow-up evaluations to explore if subjects' improvements in transgender stigma are stable over time.

Conclusion

This evaluation of transphobia in online learners enrolled in an animated, story-based course adds to the limited existing evidence that digital, contact-based antistigma interventions can be as effective as face-to-face approaches. Our findings make a case for readily scalable, digital contact-based interventions that use animation to protect the identities of individuals sharing their lived experiences. Ongoing advances in the neuroscience of stigma will further increase our knowledge of how animated educational content can be used to address stigma in historically stigmatized groups. Future interventions may build on the present one and feature more robust samples and study designs. The fact that attitudes toward transgender individuals vary highly across cultures provides an opportunity to assess the efficacy of interventions in diverse cultural contexts. We believe that stigma against transgender youth should be explored in more detail—first and foremost due to the demonstrated burden on the mental health of stigmatized children, who deserve the chance to express themselves authentically and thrive in an accepting society.

Acknowledgments

This study was funded by a faculty teaching innovation grant from the Stanford School of Medicine, a research grant from the Stanford Center for Digital Health, the Alexander von Humboldt Foundation, and the Heidelberg Institute of Global Health. For the publication fee, we acknowledge financial support from Heidelberg University.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

MG contributed to the study design, implementation, and analysis, as well as the drafting of this manuscript. MA conceived and designed the online course and all animated content. MA, KN, DA, and TB advised on the study design and data analysis. KN performed the statistical analysis. MG and DA provided context on transgender stigma.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of course content.
[DOCX File , 19 KB - formative_v9i1e59605_app1.docx]

Multimedia Appendix 2
Overview of video content.
[PDF File (Adobe PDF File), 469 KB - formative_v9i1e59605_app2.pdf]

Multimedia Appendix 3
Transgender stigma survey.
[PNG File , 345 KB - formative_v9i1e59605_app3.png]

Multimedia Appendix 4
Adaptation of the Transgender Attitudes and Beliefs Scale (TABS) to the context of parents and caregivers.
[DOCX File , 13 KB - formative_v9i1e59605_app4.docx]

Multimedia Appendix 5
Pre- and post-course survey matching based on the unique identifier.
[DOCX File , 16 KB - formative_v9i1e59605_app5.docx]

Multimedia Appendix 6
Pre- and post-course survey matching using a propensity score approach.
[DOCX File , 249 KB - formative_v9i1e59605_app6.docx]

References

1. Blakemore SJ. Adolescence and mental health. Lancet 2019;393(10185):2030-2031. [doi: 10.1016/S0140-6736(19)31013-X] [Medline: 31106741]
2. Clark TC, Lucassen MF, Bullen P, Denny SJ, Fleming TM, Robinson EM, et al. The health and well-being of transgender high school students: results from the New Zealand Adolescent Health Survey (Youth'12). J Adolesc Health 2014;55(1):93-99 [FREE Full text] [doi: 10.1016/j.jadohealth.2013.11.008] [Medline: 24438852]
3. Connolly MD, Zervos MJ, Barone CJ, Johnson CC, Joseph CL. The mental health of transgender youth: advances in understanding. J Adolesc Health 2016;59(5):489-495. [doi: 10.1016/j.jadohealth.2016.06.012] [Medline: 27544457]
4. Reisner SL, Vetters R, Leclerc M, Zaslow S, Wolfrum S, Shumer D, et al. Mental health of transgender youth in care at an adolescent urban community health center: a matched retrospective cohort study. J Adolesc Health 2015;56(3):274-279 [FREE Full text] [doi: 10.1016/j.jadohealth.2014.10.264] [Medline: 25577670]
5. Wilkinson L, Shifrer D, Pearson J. Educational outcomes of gender-diverse youth: a national population-based study. Gend Soc 2021;35(5):806-837 [FREE Full text] [doi: 10.1177/08912432211038689] [Medline: 34602743]
6. Hendricks ML, Testa RJ. A conceptual framework for clinical work with transgender and gender nonconforming clients: an adaptation of the minority stress model. Prof Psychol Res Pr 2012;43(5):460-467. [doi: 10.1037/a0029597]
7. White Hughto JM, Reisner SL, Pachankis JE. Transgender stigma and health: a critical review of stigma determinants, mechanisms, and interventions. Soc Sci Med 2015;147:222-231 [FREE Full text] [doi: 10.1016/j.socscimed.2015.11.010] [Medline: 26599625]
8. Denise EJ. Multiple disadvantaged statuses and health: the role of multiple forms of discrimination. J Health Soc Behav 2014;55(1):3-19. [doi: 10.1177/0022146514521215] [Medline: 24578393]
9. Major B, O'Brien LT. The social psychology of stigma. Annu Rev Psychol 2005;56:393-421. [doi: 10.1146/annurev.psych.56.091103.070137] [Medline: 15709941]
10. Lauber C. Stigma and discrimination against people with mental illness: a critical appraisal. Epidemiol Psichiatr Soc 2011;17(1):10-13. [doi: 10.1017/s1121189x0000261x]
11. Almeida OFX, Sousa N. Leveraging neuroscience to fight stigma around mental health. Front Behav Neurosci 2021;15:812184 [FREE Full text] [doi: 10.3389/fnbeh.2021.812184] [Medline: 35295248]
12. Rössler W. The stigma of mental disorders: a millennia-long history of social exclusion and prejudices. EMBO Rep 2016;17(9):1250-1253 [FREE Full text] [doi: 10.15252/embr.201643041] [Medline: 27470237]
13. Thornicroft G, Rose D, Kassam A, Sartorius N. Stigma: ignorance, prejudice or discrimination? Br J Psychiatry 2007;190:192-193. [doi: 10.1192/bjp.bp.106.025791] [Medline: 17329736]

XSL•FO
RenderX

14. Decety J, Echols S, Correll J. The blame game: the effect of responsibility and social stigma on empathy for pain. J Cogn Neurosci 2010;22(5):985-997. [doi: 10.1162/jocn.2009.21266] [Medline: 19425830]
15. Azevedo RT, Macaluso E, Viola V, Sani G, Aglioti SM. Weighing the stigma of weight: an fMRI study of neural reactivity to the pain of obese individuals. Neuroimage 2014;91:109-119. [doi: 10.1016/j.neuroimage.2013.11.041] [Medline: 24287441]
16. Hein G, Silani G, Preuschoff K, Batson CD, Singer T. Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. Neuron 2010;68(1):149-160 [FREE Full text] [doi: 10.1016/j.neuron.2010.09.003] [Medline: 20920798]
17. Xu X, Zuo X, Wang X, Han S. Do you feel my pain? Racial group membership modulates empathic neural responses. J Neurosci 2009;29(26):8525-8529 [FREE Full text] [doi: 10.1523/JNEUROSCI.2418-09.2009] [Medline: 19571143]
18. Chung AH, Slater MD. Reducing stigma and out-group distinctions through perspective-taking in narratives. J Commun 2013;63(5):894-911. [doi: 10.1111/jcom.12050]
19. Corrigan P. The Stigma Effect. New York, NY: Columbia University Press; 2019.
20. Jankowski SE, Yanos P, Dixon LB, Amsalem D. Reducing public stigma towards psychosis: a conceptual framework for understanding the effects of social contact based brief video interventions. Schizophr Bull 2023;49(1):99-107 [FREE Full text] [doi: 10.1093/schbul/sbac143] [Medline: 36190348]
21. Reinke RR, Corrigan PW, Leonhard C, Lundin RK, Kubiak MA. Examining two aspects of contact on the stigma of mental illness. J Soc Clin Psychol 2004;23(3):377-389. [doi: 10.1521/jscp.23.3.377.35457]
22. Amsalem D, Valeri L, Jankowski SE, Yang LH, Bello I, Nossel I, et al. Reducing public stigma toward individuals with psychosis across race and gender: a randomized controlled trial of young adults. Schizophr Res 2022;243:195-202. [doi: 10.1016/j.schres.2022.03.011] [Medline: 35397250]
23. Amsalem D, Yang LH, Jankowski S, Lieff SA, Markowitz JC, Dixon LB. Reducing stigma toward individuals with schizophrenia using a brief video: a randomized controlled trial of young adults. Schizophr Bull 2021;47(1):7-14 [FREE Full text] [doi: 10.1093/schbul/sbaa114] [Medline: 33484269]
24. Amsalem D, Markowitz JC, Jankowski SE, Yang LH, Valeri L, Lieff SA, et al. Sustained effect of a brief video in reducing public stigma toward individuals with psychosis: a randomized controlled trial of young adults. Am J Psychiatry 2021;178(7):635-642. [doi: 10.1176/appi.ajp.2020.20091293] [Medline: 33900809]
25. Oliver MB, Dillard JP, Bae K, Tamul DJ. The effect of narrative news format on empathy for stigmatized groups. J Mass Commun Q 2012;89(2):205-224. [doi: 10.1177/1077699012439020]
26. Martin A, Celentano J, Olezeski C, Halloran J, Penque B, Aguilar J, et al. Collaborating with transgender youth to educate healthcare trainees and professionals: randomized controlled trial of a didactic enhanced by brief videos. BMC Public Health 2022;22(1):2427 [FREE Full text] [doi: 10.1186/s12889-022-14791-5] [Medline: 36572857]
27. Amsalem D, Halloran J, Penque B, Celentano J, Martin A. Effect of a brief social contact video on transphobia and depression-related stigma among adolescents: a randomized clinical trial. JAMA Netw Open 2022;5(2):e220376 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.0376] [Medline: 35212749]
28. Olson KR, Durwood L, DeMeules M, McLaughlin KA. Mental health of transgender children who are supported in their identities. Pediatrics 2016;137(3):e20153223 [FREE Full text] [doi: 10.1542/peds.2015-3223] [Medline: 26921285]
29. Newcomb ME, LaSala MC, Bouris A, Mustanski B, Prado G, Schrager SM, et al. The influence of families on LGBTQ youth health: a call to action for innovation in research and intervention development. LGBT Health 2019;6(4):139-145 [FREE Full text] [doi: 10.1089/lgbt.2018.0157] [Medline: 30844341]
30. Bouris A, Guilamo-Ramos V, Pickard A, Shiu C, Loosier PS, Dittus P, et al. A systematic review of parental influences on the health and well-being of lesbian, gay, and bisexual youth: time for a new public health research and practice agenda. J Prim Prev 2010;31(5-6):273-309. [doi: 10.1007/s10935-010-0229-1] [Medline: 21161599]
31. Ryan C, Huebner D, Diaz RM, Sanchez J. Family rejection as a predictor of negative health outcomes in white and Latino lesbian, gay, and bisexual young adults. Pediatrics 2009;123(1):346-352. [doi: 10.1542/peds.2007-3524] [Medline: 19117902]
32. Simons L, Schrager SM, Clark LF, Belzer M, Olson J. Parental support and mental health among transgender adolescents. J Adolesc Health 2013;53(6):791-793 [FREE Full text] [doi: 10.1016/j.jadohealth.2013.07.019] [Medline: 24012067]
33. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
34. Coursera. Stanford. Health across the gender spectrum. 2017. URL: https://www.coursera.org/learn/health-gender-spectrum?irclickid=XdOxMRjKxyKRs4ynAzJ11rcUkCVWB1p7xuA0w0&irgwc=1&utm_medium=partners&utm_source=impact&utm_campaign=3919688&utm_content=b2c&gad_source=1 [accessed 2024-12-06]
35. QualtricsXM. URL: http://www.qualtrics.com [accessed 2024-09-05]
36. Kanamori Y, Cornelius-White JHD, Pegors TK, Daniel T, Hulgus J. Development and validation of the transgender attitudes and beliefs scale. Arch Sex Behav 2017;46(5):1503-1515. [doi: 10.1007/s10508-016-0840-1] [Medline: 27571743]
37. Schauberger G, Tutz G. Multivariate ordinal random effects models including subject and group specific response style effects. Stat Model 2021;22(5):409-429. [doi: 10.1177/1471082x20978034]
38. Bürkner PC, Vuorre M. Ordinal regression models in psychology: a tutorial. Adv Methods Pract Psychol Sci 2019;2(1):77-101. [doi: 10.1177/2515245918823199]

XSL•FO
RenderX

39. DeYoreo M, Kottas A. Bayesian nonparametric modeling for multivariate ordinal regression. J Comput Graph Stat 2017;27(1):71-84. [doi: 10.1080/10618600.2017.1316280]
40. Ananth CV, Kleinbaum DG. Regression models for ordinal responses: a review of methods and applications. Int J Epidemiol 1997;26(6):1323-1333. [doi: 10.1093/ije/26.6.1323] [Medline: 9447413]
41. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70(1):41-55. [doi: 10.1093/biomet/70.1.41]
42. Ho DE, Imai K, King G, Stuart EA. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Polit Anal 2017;15(3):199-236. [doi: 10.1093/pan/mpl013]
43. Krendl AC, Kensinger EA, Ambady N. How does the brain regulate negative bias to stigma? Soc Cogn Affect Neurosci 2012;7(6):715-726 [FREE Full text] [doi: 10.1093/scan/nsr046] [Medline: 21896496]
44. Krendl AC, Zucker HR, Kensinger EA. Examining the effects of emotion regulation on the ERP response to highly negative social stigmas. Soc Neurosci 2017;12(3):349-360 [FREE Full text] [doi: 10.1080/17470919.2016.1166155] [Medline: 26982459]
45. Maunder RD, White FA, Verrelli S. Modern avenues for intergroup contact: using E-contact and intergroup emotions to reduce stereotyping and social distancing against people with schizophrenia. Group Process Intergr 2018;22(7):947-963. [doi: 10.1177/1368430218794873]
46. Stathi S, Tsantila K, Crisp RJ. Imagining intergroup contact can combat mental health stigma by reducing anxiety, avoidance and negative stereotyping. J Soc Psychol 2012;152(6):746-757. [doi: 10.1080/00224545.2012.697080] [Medline: 23057193]
47. Stull LG, McGrew JH, Salyers MP, Ashburn-Nardo L. Implicit and explicit stigma of mental illness: attitudes in an evidence-based practice. J Nerv Ment Dis 2013;201(12):1072-1079 [FREE Full text] [doi: 10.1097/NMD.0000000000000056] [Medline: 24284643]
48. Nosek BA, Hawkins CB, Frazier RS. Implicit social cognition: from measures to mechanisms. Trends Cogn Sci 2011;15(4):152-159 [FREE Full text] [doi: 10.1016/j.tics.2011.01.005] [Medline: 21376657]
49. Simmons L, Jones T, Bradley E. Reducing mental health stigma : the relationship between knowledge and attitude change. Eur J Ment Health 2017;12(1):25-40. [doi: 10.5708/ejmh.12.2017.1.2]
50. Shin WG, Woo CW, Jung WH, Kim H, Lee TY, Decety J, et al. The neurobehavioral mechanisms underlying attitudes toward people with mental or physical illness. Front Behav Neurosci 2020;14:571225 [FREE Full text] [doi: 10.3389/fnbeh.2020.571225] [Medline: 33281570]
51. Link BG, Phelan JC. Conceptualizing stigma. Annu Rev Sociol 2001;27(1):363-385. [doi: 10.1146/annurev.soc.27.1.363]
52. Amodio DM. The neuroscience of prejudice and stereotyping. Nat Rev Neurosci 2014;15(10):670-682. [doi: 10.1038/nrn3800] [Medline: 25186236]
53. DeSteno D, Dasgupta N, Bartlett MY, Cajdric A. Prejudice from thin air: the effect of emotion on automatic intergroup attitudes. Psychol Sci 2004;15(5):319-324. [doi: 10.1111/j.0956-7976.2004.00676.x] [Medline: 15102141]
54. Thiebaut de Schotten M, Dell'Acqua F, Valabregue R, Catani M. Monkey to human comparative anatomy of the frontal lobe association tracts. Cortex 2012;48(1):82-96. [doi: 10.1016/j.cortex.2011.10.001] [Medline: 22088488]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
TABS: Transgender Attitudes and Beliefs Scale

Edited by A Mavragani; submitted 17.04.24; peer-reviewed by D Shumer, Y Perry; comments to author 14.08.24; revised version received 10.09.24; accepted 10.11.24; published 09.01.25.

Please cite as:
Greuel M, Nguyen VK, Amsalem D, Adam M, Bärnighausen T
Effect of a Narrative-Based Online Course Aimed at Reducing Stigma Toward Transgender Children and Adolescents: Longitudinal Observational Study
JMIR Form Res 2025;9:e59605
URL: https://formative.jmir.org/2025/1/e59605
doi:10.2196/59605
PMID:

©Merlin Greuel, Van Kinh Nguyen, Doron Amsalem, Maya Adam, Till Bärnighausen. Originally published in JMIR Formative Research (https://formative.jmir.org), 09.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The

XSL•FO
RenderX

complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

User Outcomes for an App-Delivered Hypnosis Intervention for Menopausal Hot Flashes: Retrospective Analysis

Katherine Scheffrahn[1], BS; Claire Hall[2], BS; Vanessa Muñiz[1], MA; Gary Elkins[1], PhD

[1]Mind-Body Medicine Lab, Department of Psychology and Neuroscience, Baylor University, Waco, TX, United States

[2]Mindset Health, Melbourne, Australia

Corresponding Author:
Gary Elkins, PhD
Mind-Body Medicine Lab
Department of Psychology and Neuroscience
Baylor University
801 Washington Ave
Waco, TX, 76701
United States
Phone: 1 254 296 0824
Email: gary_elkins@baylor.edu

Abstract

Background: Hypnotherapy has been shown to be a safe, nonhormonal intervention effective for treating menopausal hot flashes. However, women experiencing hot flashes may face accessibility barriers to in-person hypnotherapy. To solve this issue, a smartphone app has been created to deliver hypnotherapy. The Evia app delivers audio-recorded hypnotherapy and has the potential to help individuals experiencing hot flashes.

Objective: This study aims to determine user outcomes in hot flash frequency and severity for users of the Evia app.

Methods: This study is a retrospective analysis of a dataset of Evia app users. Participants were divided into 2 groups for analysis. The first group reported daytime hot flashes and night sweats, while the second group was asked to report only daytime hot flashes. The participants in the first group (daytime hot flashes and night sweats) were 139 women with ≥3 daily hot flashes who downloaded the Evia app between November 6, 2021, and June 9, 2022, with a baseline mean of 8.330 (SD 3.977) daily hot flashes. The participants in the second group (daytime hot flashes) were 271 women with ≥3 daily hot flashes who downloaded the Evia app between June 10, 2022, and February 5, 2024, with a baseline mean of 6.040 (SD 3.282) daily hot flashes. The Evia program included a 5-week program for all participants with daily tasks such as educational readings, hypnotic inductions, and daily hot-flash tracking. The app uses audio-recorded hypnosis and mental imagery for coolness, such as imagery for a cool breeze, snow, or calmness.

Results: A clinically significant reduction, defined as a 50% reduction, in daily hot flashes was experienced by 76.3% (106/139) of the women with hot flashes and night sweats and 56.8% (154/271) of the women with daily hot flashes from baseline to their last logged Evia app survey. On average, the women with hot flashes and night sweats experienced a reduction of 61.4% (SD 33.185%) in their hot flashes experienced at day and night while using the Evia app, and the women with daily hot flashes experienced a reduction of 45.2% (SD 42.567%) in their daytime hot flashes. In both groups, there was a large, statistically significant difference in the average number of daily hot flashes from baseline to end point (women with hot flashes and night sweats: Cohen d=1.28; t_{138}=15.055; P<.001; women with daily hot flashes: Cohen d=0.82; t_{270}=13.555; P<.001).

Conclusions: Hypnotherapy is an efficacious intervention for hot flashes, with the potential to improve women's lives by reducing hot flashes without hormonal or pharmacological intervention. This study takes the first step in evaluating the efficacy of an app-delivered hypnosis intervention for menopausal hot flashes, demonstrating the Evia app provides a promising app delivery of hypnotherapy with potential to increase accessibility to hypnotherapy.

(JMIR Form Res 2025;9:e63948) doi:10.2196/63948

KEYWORDS

hypnosis; hypnotherapy; hot flash; menopause; women's health; gynecology; smartphone app; applications; mobile health; mHealth; user; outcome; intervention; alternative; complementary; mind-body; mobile phone

XSL•FO
RenderX

Introduction

Background

During the menopause transition, approximately 80% of women experience hot flashes, and approximately 20% of women report hot flashes to be very bothersome [1,2]. Hot flashes negatively impact women's quality of life [3], and because the period during which women experience hot flashes lasts an average of 7.4 years, the impacts are long lasting [4].

In the past, the standard treatment for hot flashes, both premenopausal and postmenopausal, has been hormone therapy. However, hormone therapy has been associated with an increased risk of cancer, strokes, and thromboembolism [5-7]. Due to these concerns, many women do not wish to use hormone therapy; moreover, experts do not recommend hormone therapy due to these potential health issues [8].

Most of the options for nonhormonal treatments are either not as efficacious as would be desired or increase the burden on women significantly. Selective serotonin reuptake inhibitor, serotonin-norepinephrine reuptake inhibitor antidepressants, gabapentin, and paroxetine mesylate are nonhormonal medications approved by the US Food and Drug Administration for reducing hot flashes, but all these medications are associated with negative side effects [9-15]. A medication newly approved by the US Food and Drug Administration for menopausal hot flashes, fezolinetant (Veozah), has demonstrated evidence of effectiveness in reducing hot flashes [16,17]. However, fezolinetant's potential side effects include liver injury; to guard against this, patients must undergo routine blood work during the first 9 months of using this drug [17]. This risk of liver injury and the blood work required places undue burden on women seeking a beneficial nonhormonal treatment for hot flashes.

When turning to nonpharmacological, nonhormonal intervention options, a commonly used intervention is cognitive behavioral therapy, which does decrease the level of distress women feel from hot flashes; however, cognitive behavioral therapy does not reduce the number of hot flashes experienced [18,19].

Clinical hypnotherapy is a mind-body intervention that has been shown to be effective in reducing both the frequency and severity of menopausal hot flashes [15,20,21]. Hypnosis is a state of consciousness involving focused attention and reduced peripheral awareness characterized by an enhanced capacity for response to suggestion [22]. Hypnosis is a safe, nonhormonal, nonpharmacological treatment with minimal to no adverse events [15,20,21].

However, face-to-face hypnotherapy for hot flashes has limited accessibility for many women. Challenges such as financial concerns, a lack of proximity to a hypnotherapist, or the time commitment needed for appointments may all factor into women being unable to access hypnotherapy [23,24]. A smartphone app-delivered hypnotherapy intervention for menopausal hot flashes can help overcome these intervention implementation barriers. Mobile delivery of hypnotherapy provides convenient and affordable hypnotherapy for women experiencing hot flashes. With the significant growth of the mobile health industry, smartphone apps have become a commonly used method for intervention dissemination, and this approach can also be used to expand access to hypnotherapy for those in need [25].

To this end, a smartphone app to deliver hypnotherapy for hot flashes has been developed. Created by Mindset Health, the Evia app delivers hypnotic inductions and educational information for hot flashes. The Evia program was developed based on the hypnosis protocol for hot flashes tested in multiple randomized clinical trials [20,21]. The app offers a 5-week program with daily tasks such as educational readings, hypnotic inductions, and hot flash tracking. The app uses audio-recorded hypnotherapy and mental imagery for coolness (ie, imagery for cool breeze, snow, and calmness).

The Evia app holds significant potential to help individuals experiencing hot flashes and has already greatly increased women's access to hypnotherapy for hot flashes due to its availability through smartphone app stores. However, research on the app's effectiveness is lacking. A previous study analyzed Evia app users' characteristics and demographics and demonstrated that the age of Evia app users on average is consistent with the age of menopause onset and the emergence of menopausal symptoms [26]. In addition, this previous study demonstrated that on average app users experienced ≥5 hot flashes per day, a frequency similar to that of participants in previous randomized clinical trials.

Objectives

This study, which will be the first to provide data on Evia app users' outcomes, is a retrospective analysis of data collected from participants who downloaded and used the Evia app. The data, collected by Mindset Health, are from the questionnaire that users fill out before beginning the program, as well as the hot flash logs that users are encouraged to fill out throughout the program. There were 2 groups of participants. The first group was asked about daily hot flashes experienced during the day and night sweats (hot flashes that occur at night), while the second group was asked to report only daily hot flashes (daytime hot flashes). The data from each group were separately analyzed to fulfill this study's aims.

This study aims to (1) determine self-reported user outcomes regarding whether users experienced a clinically significant (50%) reduction in daily hot flashes, (2) determine self-reported user outcomes regarding changes in hot flash severity, (3) describe user demographics and characteristics, (4) describe self-reported Evia app use patterns, (5) determine factors associated with intervention outcomes, and (6) determine the association of self-reported hot flash outcomes with self-reported use patterns.

Methods

Overview

This study is a retrospective, cross-sectional analysis of data from the Evia app collected by its developers, Mindset Health. The data were collected at 1 time point from self-report questionnaires and app use data. See Multimedia Appendix 1 for the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) for this study [27]. All individuals who

XSL•FO
RenderX

downloaded the app were prompted to fill out questionnaires using an open survey method (anyone could download the app and fill out the questionnaire). Evia app users were required by the app design to fill out the baseline questionnaire when they downloaded the app (before beginning the intervention program); the study authors were not provided any information regarding whether there were potential Evia app users who did not complete the baseline questionnaire and therefore did not continue the program. Participants were prompted to optionally fill out hot flash tracker logs throughout the program. All questionnaires were filled out using the Evia app. The data used as the end point measure for this study were derived from the latest hot flash log filled out by each participant. The data were collected from November 6, 2021, to February 5, 2024.

The Evia app offers a 5-week program with daily tasks such as educational readings, hypnotic inductions, and hot flash tracking. The app uses audio-recorded hypnotherapy to deliver suggestions to induce mental imagery for coolness. During the program, users can access the following hypnotherapy audio files: moving snow, cool mountain stream, rock by a lake, cooling blue light, internal thermostat, hillside breeze, snowy mountain path, walking in summer rain, eating a cool treat, cold misty lake, air-conditioned cafe, lying in a hammock, cooling face mask, frosty morning, cold day at the beach, and relaxing in a pool. Each hypnotherapy audio file has a duration of 13 to 15 minutes. App users are encouraged to listen to a hypnotherapy audio file at least once per day.

All questionnaires were developed by Mindset Health for the Evia app, and the questionnaires were in use before the time frame from which this study used data; thus, all questionnaires were provided in a way that was usable for prior app users. Over the course of this study, the user interface and intervention program remained essentially consistent. During the time span of data collection, there was a change within the Evia app in how participants were asked to report their hot flashes. One group was asked to report daily hot flashes and night sweats (hot flashes that occur at night), and a second group was asked to report only daily hot flashes (hot flashes that occur during the day). The 2 groups are discussed further, with an explanation provided for why participants were divided into 2 groups in our analysis, in the Participant Groups subsection.

Ethical Considerations

The study protocol was reviewed by the institutional review board at Baylor University (2167670) and determined to be exempt from review because it qualified as non–human subjects research; therefore, institutional review board approval was not required. The data analyzed were provided in a deidentified form and contained no personal information. Users provided informed consent upon agreeing to the Evia app's terms and conditions, which included notification that their deidentified data may be provided for research focused on aggregated user data rather than individual information.

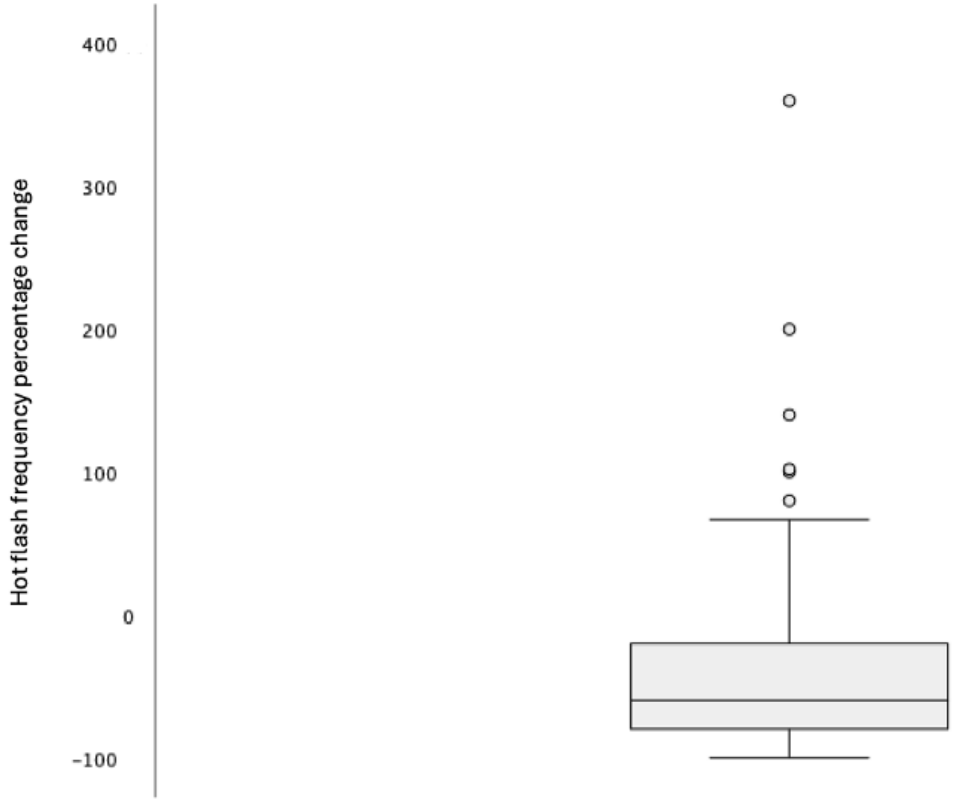
Participants

The participants are a convenience sample of women who downloaded and used the Evia app between November 6, 2021, and February 5, 2024. Given the contents of the Evia app, the respondents of the survey questionnaire are women experiencing hot flashes. A total of 835 Evia app users completed the initial questionnaire after downloading the app. Of these 835 women, 410 (49.1%) met the inclusion criteria and were included in the data analysis. Participants were included in the final analysis if they met the following criteria: (1) were aged ≥18 years and used the Evia app, defined as listening to >1 hypnotherapy audio file; (2) reported ≥3 daily hot flashes at baseline; and (3) filled out at least 1 hot flash diary log. These criteria ensured that the sample in the analysis only included women who used the Evia app past the initial download of the app, who experienced persistent hot flashes, and who had end point data to compare to baseline data. However, to better understand the differences between the people who downloaded and used the app and those who downloaded and did not use the app, demographic and characteristic data of the excluded users are also reported. Participants were not solicited or recruited for the study because the sample consists of women who downloaded and used the app. Participants provided consent digitally when downloading the Evia app.

After excluding app users based on the eligibility criteria, a box plot (Figure 1) was used to identify outliers based on the hot flash frequency change; this method was chosen because it is a standardized way to identify outliers when the data are not normally distributed. Of the 416 participants, 6 (1.4%) were identified as outliers and removed from the analysis. We chose to remove the statistical outliers because doing so allows for better concentration on the relevant data to fulfill the study aims. This study is meant to provide an understanding of what most women experience while using the Evia app and any patterns that may be present. Removing the statistical outliers ensures that the analysis is not skewed by extreme values. This resulted in 410 (49.1%) of the 835 Evia app users who completed the initial app questionnaire being included as participants.

XSL•FO
RenderX

Figure 1. Simple box plot of hot flash frequency change outliers. After excluding app users based on the eligibility criteria, a box plot was created to view the distribution of the changes in hot flash frequency from baseline to end point. This method revealed 6 participants who were statistical outliers. To best represent the relevant data to answer our study aims, these 6 outliers were removed from the final analysis in a retrospective data analysis on app use and outcomes in women experiencing menopausal hot flashes who downloaded the Evia app between November 6, 2021, and February 5, 2024.



Participant Groups

During the time span of data collection, the Evia app underwent several small changes to the entry questionnaire and the hot flash logs that users are prompted to respond to at the time of app download. When users downloaded the app, they were asked to report how many hot flashes they experienced. Users who downloaded the app up until June 9, 2022, were asked 2 questions about their hot flashes:

- "How many hot flashes do you experience each day?"
- "How many night sweats do you experience during the night?"

After June 9, 2022, users downloading the Evia app were only asked 1 question:

- "How many hot flashes do you experience each day?"

Due to this change in the app survey made by the developers during data collection, the 410 participants were divided into 2 groups: the first group (hot flashes and night sweats) included 139 (33.9%) women, and the second group (daily hot flashes) consisted of 271 (66.1%) women. This was necessary to accurately describe the data. This study reports the results of the 2 groups separately because their baseline hot flashes were measured differently.

Measures

Aim 1

Throughout the Evia program, participants were prompted to log their daily hot flashes by entering a number in the hot flash logs. Participants in both groups were prompted to report the daily number of hot flashes, while participants with hot flashes and night sweats were prompted to report night sweats (hot flashes experienced at night) as well. In the app, the daily hot flash log was optional for users, but due to the aims of this study, only individuals who had filled out at least 1 daily hot flash log were included in the final analysis. For the participants with hot flashes and night sweats, the total number of hot flashes experienced during the day and at night were summed together for each participant at baseline and end point to provide a final number of hot flashes experienced throughout a 24-hour period. The baseline number of daily hot flashes was reported by the user in the entry questionnaire, and the end point number of daily hot flashes was the number indicated by the user on their last recorded hot flash log.

Aim 2

When responding to the baseline questionnaire, before beginning the app-delivered intervention, participants provided the average intensity of their hot flashes. Participants were asked, "How would you rate the intensity of your daily hot flashes (on average)?" The response options included "mild," "moderate," "severe," and "very severe." This question provided descriptions of what each level of intensity would entail. The participants were given the option of logging their daily hot flash intensity

XSL•FO
RenderX

throughout the 5-week program with the hot flash tracker. The hot flash severity indicated on their last recorded hot flash log was considered the participants' hot flash severity at end point.

Aim 3

In the initial questionnaire, participants filled out questions on demographics and characteristics. The order of the items on the baseline questionnaire was not randomized, and 1 item per page was displayed on the app screen, with a total of 20 pages. Throughout the questionnaire, participants were able to go back and change any answers if needed. Participants were asked to report their age and whether their menstrual cycles were "regular," "irregular but haven't stopped," "stopped within the past year," "or "stopped over a year ago." Participants were also asked, "How would you classify your stage of menopause?" The response options included "perimenopause," "menopause," "postmenopause," and "I'm not sure." Participants were asked whether they were currently taking hormone replacement therapy and were given the options "yes," "yes, but weaning off it," or "no." Participants were asked at baseline, "How do you currently manage your hot flashes?" Answer options included "natural supplements or OTC drug," "other prescribed medications," "hormone therapy (HRT)," "cooling tools (water spray, fans, etc)," "soy consumption (soy isoflavones)," "lifestyle changes (exercise, etc)," "antidepressants (for hot flashes)," and "other." Participants were asked, "How important to you is a nonhormonal tool to help with hot flashes?" The answer options included "extremely important," "somewhat important," and "not important." Participants were asked, "Have you tried hypnotherapy before?" The answer options were "yes" or "no."

Aim 4

Each time an Evia app user listened to a hypnotherapy audio file on the app, the event was logged. To determine Evia app use patterns, this study used these listening logs to determine the number of times an individual listened to a hypnotherapy audio file as well as the number of days on which an individual listened to at least 1 audio file. Only listening events that occurred before each participant's end point were included in the analysis.

Aim 5

The aforementioned characteristics and demographics, including age, menopausal stage, whether they had tried hypnosis before, the importance of a nonhormonal tool, and other intervention methods, were used as measures to determine the characteristics associated with positive outcomes in hot flash reduction.

Aim 6

For this aim, the listening logs for each individual will be used in conjunction with the self-reported initial daily hot flashes and the self-reported daily hot flashes indicated in the end point hot flash tracker.

Data Analyses

In analyzing these data, descriptive statistics (means, SDs, and frequencies) were calculated for the variables of interest. For aim 1, a 2-tailed paired samples *t* test was used. Aim 2 used a Wilcoxon signed rank test. To carry out aims 5 and 6, Spearman rank correlation tests were used to test associations between the variables.

Results

Aim 1: Hot Flash Reduction Outcomes

This study included 410 women who used the Evia app and had ≥3 daily hot flashes. The 410 participants were divided into 2 groups: the first group (hot flashes and night sweats) included 139 (33.9%) women, and the second group (daily hot flashes) consisted of 271 (66.1%) women. At baseline, the women with hot flashes and night sweats experienced a mean of 8.330 (SD 3.977; range 3-24) daily hot flashes, while the women with daily hot flashes experienced a mean of 6.040 (SD 3.282; range 3-20) daily hot flashes. At end point, the women with hot flashes and night sweats experienced a mean of 3.070 (SD 2.883; range 0-15) daily hot flashes, while the women with daily hot flashes experienced a mean of 2.970 (SD 2.377; range 0-15) daily hot flashes. Tables 1 and 2 show the distribution, respectively, of total baseline and endpoint hot flashes reported by the group of participants with daily hot flashes and night sweats. Tables 3 and 4 show the distribution, respectively, of total baseline and endpoint hot flashes reported by the group of participants with daily hot flashes.

A paired samples *t* test indicated that there was a large, statistically significant difference in the average number of daily hot flashes experienced by the women with hot flashes and night sweats from baseline to end point (Cohen d=1.28; t_{138}=15.055; 2-tailed P<.001), as well as in the average number of daily hot flashes experienced by the women with daily hot flashes from baseline to end point (Cohen d=.82, t_{270}=13.555; 2-tailed P<.001).

Clinical significance in the reduction of hot flashes is defined as a reduction of 50% in daily hot flashes [5,28]. In this study, 76.3% (106/139) of the women with hot flashes and night sweats experienced at least a 50% reduction in daily hot flashes from when they completed their entry questionnaire to the latest hot flash log included in the dataset, while 56.8% (154/271) of the women with daily hot flashes experienced a clinically significant 50% reduction in daily hot flashes. The average percentage reduction in daily hot flashes for the women with hot flashes and night sweats was 61.4% (SD 33.185%), while for the women with daily hot flashes, it was 45.2% (SD 42.567%).

XSL•FO
RenderX

Table 1. Distribution of baseline hot flashes for women with daily hot flashes and night sweats (n=139).

Baseline total hot flashes	Participants, n (%)
3	8 (5.8)
4	10 (7.2)
5	21 (15.1)
6	15 (10.7)
7	6 (4.3)
8	13 (9.4)
9	12 (8.6)
10	39 (28.1)
11	1 (0.7)
12	1 (0.7)
13	1 (0.7)
14	0 (0)
15	5 (3.6)
16	0 (0)
17	0 (0)
18	0 (0)
19	0 (0)
20	6 (4.3)
21	0 (0)
22	0 (0)
23	0 (0)
24	1 (0.7)

Table 2. Distribution of end point hot flashes for women with daily hot flashes and night sweats (n=139).

End point total hot flashes	Participants, n (%)
0	22 (15.8)
1	21 (15.1)
2	28 (20.1)
3	22 (15.8)
4	17 (12.2)
5	12 (8.6)
6	3 (2.2)
7	2 (1.4)
8	4 (2.9)
9	0 (0)
10	5 (3.6)
11	0 (0)
12	1 (0.7)
13	0 (0)
14	0 (0)
15	2 (1.4)

XSL•FO
RenderX

Table 3. Distribution of baseline hot flashes with women with daily hot flashes (n=271).

Baseline total hot flashes	Participants, n (%)
3	30 (11.1)
4	35 (12.9)
5	135 (49.8)
6	14 (5.2)
7	3 (1.1)
8	8 (3)
9	0 (0)
10	29 (10.7)
11	0 (0)
12	2 (0.7)
13	0 (0)
14	1 (0.4)
15	8 (3)
16	1 (0.4)
17	0 (0)
18	0 (0)
19	0 (0)
20	5 (1.8)

Table 4. Distribution of end point hot flashes for women with daily hot flashes (n=271).

End point total hot flashes	Participants, n (%)
0	37 (13.7)
1	39 (14.4)
2	68 (25.1)
3	30 (11.1)
4	35 (12.9)
5	21 (7.7)
6	21 (7.7)
7	6 (2.2)
8	19 (7)
9	0 (0)
10	4 (1.5)
11	0 (0)
12	0 (0)
13	0 (0)
14	0 (0)
15	1 (0.4)

Aim 2: Hot Flash Severity Outcomes

Participants were asked to report their hot flash severity at baseline and end point. For participants with hot flashes and night sweats, 2 participants declined to answer regarding their hot flash severity at baseline, 42 (30.7%) reported mild hot flashes, 65 (47.4%) moderate hot flashes, 22 (16.1%) severe hot flashes, and 8 (5.8%) very severe hot flashes. At end point, of the 139 participants with hot flashes and night sweats, 22 (15.8%) reported no hot flash severity due to the absence of hot flashes, 58 (41.7%) reported mild hot flashes, 46 (33.1%)

XSL•FO
RenderX

reported moderate hot flashes, 11 (7.9%) reported severe hot flashes, and 2 (1.4%) reported very severe hot flashes.

At baseline, of the 271 participants with daily hot flashes, 73 (26.9%) reported mild hot flashes, 116 (42.8%) moderate hot flashes, 64 (23.6%) severe hot flashes, and 18 (6.6%) very severe hot flashes. At end point, of the 271 participants with daily hot flashes, 37 (13.7%) reported no hot flash severity due to the absence of hot flashes, 147 (54.2%) reported mild hot flashes, 74 (27.3%) reported moderate hot flashes, and 13 (4.8%) reported severe hot flashes. None reported very severe hot flashes.

For the participants with hot flashes and night sweats, a Wilcoxon signed rank test indicated that there was a significant (P<.001) difference between baseline and end point in daily hot flash severity, and the test demonstrated that 48.9% (67/137) of the women experienced a decrease in the severity of their daily hot flashes, 11.7% (16/137) had an increase in severity, and 39.4% (54/137) experienced no change in their daily hot flash severity.

For the participants with daily hot flashes, a Wilcoxon signed rank test indicated that there was a significant (P<.001) difference between baseline and end point in daily hot flash severity, and the test demonstrated that 64.2% (174/271) of the women experienced a decrease in the severity of their daily hot flashes, 4.4% (12/271) had an increase in severity, and 31.4% (85/271) experienced no change in their daily hot flash severity.

Aim 3: User Demographics and Characteristics

The mean age of the participants with hot flashes and night sweats was 52.600 (SD 5.577; range 38-72) years. The mean age of the participants with daily hot flashes was 52.400 (SD 5.483; range 30-77) years. When asked to classify their stage of menopause, of the 139 participants with hot flashes and night sweats, 51 (36.7%) indicated that they were unsure, 34 (24.5%) reported that they were perimenopausal, 23 (16.5%) reported that they were menopausal, and 31 (22.3%) reported that they were postmenopausal, while of the 271 participants with daily hot flashes, 96 (35.4%) indicated that they were unsure, 86 (31.7%) reported that they were perimenopausal, 29 (10.7%) reported that they were menopausal, and 60 (22.1%) reported that they were postmenopausal. When asked about their menstrual cycles, of the 77 women with hot flashes and night sweats who responded to this question, 35 (45%) reported that their menstrual cycles had stopped more than a year ago, 18 (23%) reported that they had stopped within the past year, 20 (26%) indicated that their cycles were irregular but had not stopped, and 4 (5%) reported having regular cycles. When asked about their menstrual cycles, of the 121 women with daily hot flashes who responded to this question, 62 (51.2%) reported that their menstrual cycles had stopped more than a year ago, 30 (24.8%) reported that they had stopped within the past year, 27 (22.3%) indicated that their cycles were irregular but had not stopped, and 2 (1.7%) reported having regular cycles.

Regarding using hormone therapy for hot flashes, of the 139 participants with hot flashes and night sweats, 128 (92.1%) were not using hormone therapy, 10 (7.2%) were using hormone therapy, and 1 (0.7%) was using hormone therapy but was weaning off it, while of the 271 participants with daily hot flashes, 234 (86.3%) were not using hormone therapy, 34 (12.5%) were using hormone therapy, and 3 (1.1%) were using hormone therapy but were weaning off it. In answer to the question “How important to you is a nonhormonal tool to help with hot flashes?” of the 63 women with hot flashes and night sweats who provided an answer, 44 (70%) responded “extremely important,” 17 (27%) responded “somewhat important,” and 2 (3%) responded “not important,” while of the 150 women with hot flashes who provided an answer, 104 (69.3%) responded “extremely important,” 39 (26%) responded “somewhat important,” and 7 (4.7%) responded “not important.”

Of the 410 participants in both groups, 209 (51%) provided information at baseline on what treatments they were currently using to manage their hot flashes. Of these 209 women, 111 (53.1%) used cooling tools such as a water spray or fans, 95 (45.5%) adopted lifestyle changes such as exercise, 76 (36.4%) used natural supplements or over-the-counter drugs, 30 (14.4%) consumed soy isoflavones, 14 (6.7%) used antidepressants for hot flashes, 10 (4.8%) took other prescribed medications, and 31 (14.8%) selected “other.” The majority of the participants (136/181, 75.1%) had not tried hypnotherapy before, while the rest (45/181, 24.9%) had tried it before.

Demographics and characteristics surveyed in the baseline questionnaire were also analyzed for the 425 users who downloaded the Evia app between November 6, 2021, and February 5, 2024, but these users were not included in the full data analysis due to not using the Evia app beyond downloading it, not experiencing at least 3 daily hot flashes, or not having end point data. It is important to note that not all individuals in this group filled out all questionnaire items; therefore, the percentages are based on those who provided an answer. The mean age of these users was 51.100 (SD 6.339; range 18-83) years. When asked to classify their stage of menopause, of the 401 users who responded, 118 (29.4%) indicated that they were unsure, 151 (37.7%) reported that they were perimenopausal, 50 (12.5%) reported that they were menopausal, 79 (19.7%) reported that they were postmenopausal, and 3 (0.7%) reported that they were not menopausal. When asked about their menstrual cycles, of the 231 women who responded, 103 (44.6%) reported that their menstrual cycles had stopped more than a year ago, 42 (17.7%) reported that they had stopped within the past year, 61 (26.4%) indicated that their cycles were irregular but had not stopped, and 25 (10.8%) reported having regular cycles. Regarding using hormone therapy for hot flashes, of the 419 users who responded, 362 (86.4%) were not using hormone therapy, 45 (10.7%) were using hormone therapy, and 12 (2.9%) were using hormone therapy but were weaning off it. In answer to the question “How important to you is a nonhormonal tool to help with hot flashes?” of the 166 users who provided an answer, 94 (56.6%) responded “extremely important,” 58 (34.9%) responded “somewhat important,” and 14 (8.4%) responded “not important.” Not all participants responded to this question because it was added to the Evia questionnaire partway through the data collection period. The majority of the participants (96/146, 65.8%) had not tried hypnotherapy before, while the rest (50/146, 34.2%) had tried it before.

XSL•FO
RenderX

Aim 4: Evia App Use Patterns

Evia app use can be evaluated by the number of days on which a user listened to at least 1 hypnotherapy audio file. Participants' end point data, based on their last logged hot flash diary, varied in the time elapsed since baseline (when the app was downloaded). For the participants with hot flashes and night sweats, end point ranged from day 1 (baseline) to day 730, with the average length of time from baseline being 110.31 (SD 152.86) days. For the participants with daily hot flashes, end point ranged from day 1 (baseline) to day 461, with the average length of time from baseline being 67.62 (SD 78.25) days. This differed from the number of days on which the participants used the app. The average number of distinct days on which the participants with hot flashes and night sweats listened to at least 1 audio file was 35.370 (SD 44.638; range 1-307). The average number of hypnotherapy audio files listened to overall by the participants with hot flashes and night sweats was 40.380 (SD 50.741; range 2-344). The average number of distinct days on which the participants with daily hot flashes listened to at least 1 audio file was 27.140 (SD 35.663; range 1-390). The average number of hypnotherapy audio files listened to overall by the participants with daily hot flashes was 31.280 (SD 39.816; range 2-414).

Aim 5: Factors Associated With Outcomes

The factors tested for association with the Evia app users' outcomes were participants' age, menopausal stage, hormone use, and past hypnotherapy use. A Spearman rank correlation test on the data of the participants with hot flashes and night sweats found no significant associations between participants' age, menopausal stage, hormone use, or past hypnotherapy use and the outcomes in hot flash frequency reduction. The results of the correlation test are shown in Table 5.

Table 5. Spearman rank correlation test for user characteristics and hot flash frequency percentage reduction among the participants with hot flashes and night sweats[a].

Characteristics	Correlation coefficient
Age (n=139)	–0.029
Menopausal stage (n=139)	0.027
Hormone use (n=139)	–0.113
Past hypnotherapy use (n=31)	0.070

[a]This table shows the correlational analysis results for the associations between user characteristics and user outcomes in a retrospective data analysis on app use and outcomes in women who experience menopausal hot flashes and downloaded the Evia app between November 6, 2021, and June 9, 2022.

A Spearman rank correlation test on the data of the participants with daily hot flashes found no significant associations between participants' age, hormone use, or past hypnotherapy use and their outcomes in hot flash frequency reduction. However, there was a small, positive, significant correlation between participants' menopausal stage and percentage reduction in hot flashes. The results of the correlation test are shown in Table 6.

Table 6. Spearman rank correlation test for user characteristics and hot flash frequency percentage reduction among the participants with daily hot flashes (n=271)[a].

Characteristics	Correlation coefficient
Age	–0.092
Menopausal stage	0.239[b]
Hormone use	0.047
Past hypnotherapy use	0.051

[a]This table shows the correlational analysis results for the associations between user characteristics and user outcomes in a retrospective data analysis on app use and outcomes in women who experience menopausal hot flashes and downloaded the Evia app between June 10, 2022, and February 5, 2024.

[b]$P<.001$.

Aim 6: Use Pattern Associations With Outcomes

A Spearman rank correlation test to assess the association between app use and percentage reduction in hot flashes among the participants with hot flashes and night sweats found a small, positive, statistically significant association between the number of days on which a user listened to at least 1 hypnotherapy audio file and the percentage reduction in hot flashes from baseline to end point. There was also a small, positive, statistically significant association between the total number of hypnotherapy audio files that users listened to and the percentage reduction in hot flashes from baseline to end point. The results of the correlation test are presented in Table 7.

XSL•FO
RenderX

Table 7. Spearman rank correlation test for app use and hot flash frequency percentage reduction among the participants with hot flashes and night sweats (n=139)[a].

App use	Correlation coefficient
Number of distinct days listened	0.209[b]
Number of audio file listens	0.199[c]

[a]This table shows the correlational analysis results for the associations between user outcomes and users' app use in a retrospective data analysis on app use and outcomes in the women with hot flashes and night sweats who experience menopausal hot flashes and downloaded the Evia app between November 6, 2021, and June 9, 2022. There was a small significant correlation between increased app use and a greater percentage reduction in hot flashes.

[b]P=.01.

[c]P=.02.

A Spearman rank correlation test to assess the association between app use and percentage reduction in hot flashes among the participants with daily hot flashes found no significant association between the number of days on which they listened to at least 1 hypnotherapy audio file and the percentage reduction in hot flashes from baseline to end point. There was also no statistically significant association between the total number of hypnotherapy audio files that users listened to and the percentage reduction in hot flashes from baseline to end point. The results of the correlation test are presented in Table 8.

Table 8. Spearman rank correlation test for app use and hot flash frequency percentage reduction among the participants with daily hot flashes (n=271)[a].

App use	Correlation coefficient
Number of distinct days listened	0.091
Number of audio file listens	0.080

[a]This table shows the correlational analysis results for the associations between user outcomes and users' app use in a retrospective data analysis on app use and outcomes in the women with daily hot flashes who experience menopausal hot flashes and downloaded the Evia app between June 10, 2022, and February 5, 2024. There were no significant correlations between increased app use and a percentage reduction in hot flashes.

Discussion

Principal Findings

The results demonstrated that the majority of Evia app users experienced a clinically significant reduction of at least 50% in their daily hot flashes, with 76.3% (106/139) of the participants with hot flashes and night sweats and 56.8% (154/271) of the participants with daily hot flashes achieving mean reductions of 61.4% and 45.2%, respectively. The difference in daily hot flashes from baseline to end point was statistically significant (P<.001) for both groups. This result demonstrates the potential benefits of the Evia app. Although this result should be interpreted in light of the fact that there was no control group in this dataset to compare with, the majority of women who began using the Evia app saw a decrease in their hot flashes of at least 50%. Using the measure of clinical significance to interpret these results will allow clinicians to provide clear information when discussing how the Evia app might benefit their patients. The participants with hot flashes and night sweats had a greater proportion reaching clinical significance due to more accurate reporting of hot flashes experienced during a 24-hour period (at day and night); thus, this group had a higher reported baseline mean of hot flashes than the participants with daily hot flashes. Although the baseline mean number of daily hot flashes for both groups differed (participants with hot flashes and night sweats: mean 8.330, SD 3.977; participants with daily hot flashes: mean 6.040, SD 3.282) because the participants with daily hot flashes were not being asked to report night sweats separately, both groups had a similar mean number of daily hot flashes at end point (participants with hot flashes and night sweats: mean 3.070, SD 2.883; participants with daily hot flashes: mean 2.970, SD 2.377).

In this study, many Evia app users also experienced decreases in their hot flash severity. Hot flash severity decreased in 48.9% (67/137) of the women with hot flashes and night sweats and 64.2% (174/271) of the women with daily hot flashes. Crucially, participants' severe and very severe hot flashes decreased: of the 410 women across both groups, 86 (21%) reported severe hot flashes at baseline, but only 24 (5.9%) reported severe hot flashes at end point; similarly, at baseline, 26 (6.3%) reported very severe hot flashes, but only 2 (0.5%) reported very severe hot flashes at end point. This result demonstrates that even women who do not experience a clinically significant reduction in their hot flash frequency may still experience significant benefit by a decrease in hot flash severity. The significant decreases in both hot flash frequency and severity are encouraging due to the need for accessible and nonpharmacological interventions for menopausal hot flashes. These results warrant further investigation in a randomized controlled trial of the Evia app to better understand if this is an efficacious mobile delivery method for hypnotherapy in treating menopausal hot flashes.

This study found a small, positive, significant association between app use and hot flash outcomes among the participants with hot flashes and night sweats. This may indicate that the more an individual uses the Evia app hypnotherapy audio files, the greater the percentage reduction in hot flashes. However, the correlation was small; therefore, this result should be

XSL•FO
RenderX

interpreted with caution. In addition, there was not a similarly significant result among the participants with daily hot flashes. App use varied widely between participants (ranging from 1 to 390 unique days), and this study did not evaluate whether participants used the app on consecutive days. As the Evia program is designed to be used daily for 5 weeks, it may be that some participants' lack of consistency made it challenging to understand the effects of use on outcome. A future study may be able to address this by examining within a more standardized study design whether use is associated with hot flash outcomes. Furthermore, this dataset relied on single-day measures of hot flash frequency for both baseline and end point. This may also have impacted the lack of a significant association between use and outcomes because women will naturally have some variability in their hot flash frequency from day to day.

Likewise, the lack of significant associations between participant characteristics such as age, hormone use, and past hypnotherapy use and outcomes is likely due to the lack of standardization across participants. There was a small, positive, significant association between participants' menopausal stage and hot flash reduction among the participants with daily hot flashes, with women further along in menopausal stage being more likely to have greater hot flash reductions. However, this association should be interpreted with caution, given that it is a small correlation within a large sample size, and there was no control group; no significant correlation was found among the participants with hot flashes and night sweats. A future study should further investigate what characteristics might be associated with achieving a clinically significant outcome through a randomized controlled trial design.

When comparing the users who met the inclusion criteria with users who did not continue using the app, the demographics and characteristics were generally comparable between the groups. One difference is that the mean number of daily hot flashes (6.820, SD 3.691) among the included users (n=410) at baseline was higher than the mean number of daily hot flashes (3.310, SD 3.799) among the excluded users (n=425). It may be that those who experienced more daily hot flashes were more likely to persist in using the app beyond the initial download.

Comparisons With Prior Work

In the retrospective study by Snyder and Elkins [26] on the demographics and characteristics of Evia app users, the authors indicated that a surprising number of women (3048/9103, 33.48%) reported that they were unsure of their menopausal stage. Our study found this proportion to be slightly higher when looking at the sample included in the analysis, with 35.9% (147/410) indicating that they were unsure of their menopausal stage. There may be a knowledge gap in the general population of women on the topic of menopause. An increase in educational content on menopausal stages in the Evia app may provide a needed resource for women.

This study found the mean number of daily hot flashes to be 8.330 (SD 3.977) among the participants with hot flashes and night sweats and 6.040 (SD 3.282) among the participants with daily hot flashes at baseline, both of which are lower than those found in an earlier clinical trial conducted by Elkins et al [21], where the authors found the mean number of daily hot flashes to be 10. However, the distribution of the baseline hot flash frequency is consistent with that in previous studies, with more women experiencing ≥5 hot flashes [21,26]. This may be due to the differences in measuring hot flashes between this study and the one by Elkins et al [21]. Our study had a single-item measure, with participants being asked, "How many hot flashes do you experience each day?" In the study by Elkins et al [21], participants tracked their daily hot flashes for a week, which was then averaged, thus providing a more accurate method due to women experiencing variable numbers of hot flashes from day to day. Future studies of the Evia app should explore the use of a weekly hot flash log.

As our study did not have a control group for comparison, we lack the ability to know how a control group's hot flash frequency and severity might have changed over time compared to the intervention group. However, by comparing our findings to those of prior research that included control groups, we can better understand the typical pattern of hot flash severity and frequency over time. Prior research has demonstrated that there is not a large change in hot flash frequency or severity over time for a control group. A 2013 randomized controlled trial examining face-to-face hypnosis for hot flashes found that the control group's hot flash score (a product of hot flash frequency × hot flash severity) only demonstrated a mean reduction of 8.32% compared to the mean reduction of 71.36% in the hot flash score of the hypnosis group [21]. Future studies should examine Evia app users compared to a matched control to understand whether the outcomes are significantly different.

Limitations

A primary limitation of this study concerns the change in the baseline questionnaire that resulted in 1 group of Evia app users not being specifically asked about their night sweats. This meant that the participants with daily hot flashes might have underreported their total number of hot flashes experienced during a 24-hour period (at day and night). However, we still observed that both the participants with hot flashes and night sweats and those with daily hot flashes had a very similar mean number of daily hot flashes at end point (3.070, SD 2.883 and 2.970, SD 2.377, respectively). In addition, since the completion of the data collection for this study, the Evia app has been updated to once again ask users specifically about night sweats, separate from daytime hot flashes.

Another primary limitation of this study is the lack of end point data for information collected in the baseline survey. Due to the retrospective nature of this analysis, participants only completed a baseline survey, after which they filled out hot flash tracker logs throughout the program; as a result, only hot flash frequency and severity could be compared between baseline and end point.

The data collected included the number of times a user listened to a hypnotherapy audio file. Listens were only logged in the app once a user had completed listening to the audio file. Thus, there are no data captured for users who may have partially listened to a hypnotherapy audio file. While the intervention is designed to have users listen to the hypnotherapy audio files in their entirety, future studies might consider including partial listens to the data collected.

XSL•FO
RenderX

In addition, this study is unable to fully assess the impacts of potential participant attrition. Participants who were more motivated to use all elements of the app, such as the hot flash diary logs (which served as end point measures), may have experienced greater benefit. Alternatively, those who benefited from the intervention may have reduced app use due to the absence of ongoing hot flashes. Given this limitation, future replication of this study is warranted.

Finally, this study was also limited by the lack of standardization across participants. Due to the retrospective nature of this analysis, the data included participants with varying levels of Evia app use. Some of the participants completed the 5-week Evia program, while others did not; some of the participants used the Evia app frequently, while others did not. A future study with a greater level of standardization across participants would be able to better provide a clearer understanding of the benefits of the Evia app.

Future Directions

Considering the positive clinical outcomes noted in this study, there is a need for future studies to address this study's limitations by conducting a randomized controlled trial of the Evia app for menopausal hot flashes. A randomized controlled trial will be able to clearly demonstrate whether the Evia app is an effective mobile delivery method for hypnotherapy in treating menopausal hot flashes. The randomized controlled trial should compare the Evia intervention to a placebo control. A future study should administer postintervention measures to provide end point data, ensure that all participants complete the 5-week program to reduce the variability in use among participants, and use weekly hot flash diaries to gain a more accurate measure of hot flash frequency. In addition, a randomized controlled trial of the Evia app should examine when during the 5-week program women begin to see results because this study only examined baseline and end point data.

Conclusions

This study is the first to report the hot flash frequency and severity outcomes of Evia app users. The results demonstrated that 76.3% (106/139) of the Evia app users who were asked about hot flashes during the day and night experienced a clinically significant reduction in their daily hot flashes, while 56.8% (154/271) of the Evia app users who were only asked about daily hot flashes achieved a clinically significant reduction in their daily hot flashes. In addition, 48.9% (67/137) of the women with hot flashes and night sweats and 64.2% (174/271) of the women with daily hot flashes experienced a decrease in their hot flash severity from baseline to end point. This study demonstrates the potential benefits of the Evia app as an accessible and nonpharmacological intervention for menopausal hot flashes. However, additional research is needed to explore the efficacy of the app in randomized controlled trial as well as to better understand the factors that moderate and mediate beneficial therapeutic outcomes.

Acknowledgments

The authors would like to acknowledge Mindset Health and its employees for collaborating in this project. Mindset Health funded the publication cost for this open access article but was not involved in the results or data analysis.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

GE is a consultant to Mindset Health but was not involved in data processing. GE receives a consultant salary from Mindset Health and holds stock options in the company. However, this salary and equity are not dependent on the results of this research. CH is an employee of Mindset Health and receives compensation, however CH was not involved in any of the data analysis of this project. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[DOCX File , 22 KB - formative_v9i1e63948_app1.docx]

References

1. Gold EB, Colvin A, Avis N, Bromberger J, Greendale GA, Powell L, et al. Longitudinal analysis of the association between vasomotor symptoms and race/ethnicity across the menopausal transition: study of women's health across the nation. Am J Public Health 2006 Jul;96(7):1226-1235. [doi: 10.2105/AJPH.2005.066936] [Medline: 16735636]
2. Kronenberg F. Hot flashes: epidemiology and physiology. Ann N Y Acad Sci 1990;592:52-86; discussion 123-33. [doi: 10.1111/j.1749-6632.1990.tb30316.x] [Medline: 2197954]
3. Stein KD, Jacobsen PB, Hann DM, Greenberg H, Lyman G. Impact of hot flashes on quality of life among postmenopausal women being treated for breast cancer. J Pain Symptom Manage 2000 Jun;19(6):436-445 [FREE Full text] [doi: 10.1016/s0885-3924(00)00142-1] [Medline: 10908824]

XSL•FO
RenderX

4. Avis NE, Crawford SL, Greendale G, Bromberger JT, Everson-Rose SA, Gold EB, et al. Duration of menopausal vasomotor symptoms over the menopause transition. JAMA Intern Med 2015 Apr;175(4):531-539 [FREE Full text] [doi: 10.1001/jamainternmed.2014.8063] [Medline: 25686030]
5. Loprinzi CL, Michalak JC, Quella SK, O'Fallon JR, Hatfield AK, Nelimark RA, et al. Megestrol acetate for the prevention of hot flashes. N Engl J Med 1994 Aug 11;331(6):347-352. [doi: 10.1056/NEJM199408113310602] [Medline: 8028614]
6. North American Menopause Society. The 2012 hormone therapy position statement of: the North American Menopause Society. Menopause 2012 Mar;19(3):257-271 [FREE Full text] [doi: 10.1097/gme.0b013e31824b970a] [Medline: 22367731]
7. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. JAMA 2002 Jul 17;288(3):321-333. [doi: 10.1001/jama.288.3.321] [Medline: 12117397]
8. Chlebowski RT, McTiernan A. Elements of informed consent for hormone replacement therapy in patients with diagnosed breast cancer. J Clin Oncol 1999 Jan;17(1):130-142. [doi: 10.1200/JCO.1999.17.1.130] [Medline: 10458226]
9. Nelson HD, Vesco KK, Haney E, Fu R, Nedrow A, Miller J, et al. Nonhormonal therapies for menopausal hot flashes: systematic review and meta-analysis. JAMA 2006 May 03;295(17):2057-2071. [doi: 10.1001/jama.295.17.2057] [Medline: 16670414]
10. Shams T, Firwana B, Habib F, Alshahrani A, Alnouh B, Murad MH, et al. SSRIs for hot flashes: a systematic review and meta-analysis of randomized trials. J Gen Intern Med 2014 Jan;29(1):204-213 [FREE Full text] [doi: 10.1007/s11606-013-2535-9] [Medline: 23888328]
11. Sun Z, Hao Y, Zhang M. Efficacy and safety of desvenlafaxine treatment for hot flashes associated with menopause: a meta-analysis of randomized controlled trials. Gynecol Obstet Invest 2013;75(4):255-262. [doi: 10.1159/000348564] [Medline: 23548358]
12. Wei D, Chen Y, Wu C, Wu Q, Yao L, Wang Q, et al. Effect and safety of paroxetine for vasomotor symptoms: systematic review and meta-analysis. BJOG 2016 Oct;123(11):1735-1743. [doi: 10.1111/1471-0528.13951] [Medline: 27062457]
13. Freeman EW, Guthrie KA, Caan B, Sternfeld B, Cohen LS, Joffe H, et al. Efficacy of escitalopram for hot flashes in healthy menopausal women: a randomized controlled trial. JAMA 2011 Jan 19;305(3):267-274 [FREE Full text] [doi: 10.1001/jama.2010.2016] [Medline: 21245182]
14. Yoon SH, Lee JY, Lee C, Lee H, Kim SN. Gabapentin for the treatment of hot flushes in menopause: a meta-analysis. Menopause 2020 Apr;27(4):485-493. [doi: 10.1097/GME.0000000000001491] [Medline: 32049930]
15. Maclaughlan David S, Salzillo S, Bowe P, Scuncio S, Malit B, Raker C, et al. Randomised controlled trial comparing hypnotherapy versus gabapentin for the treatment of hot flashes in breast cancer survivors: a pilot study. BMJ Open 2013 Sep 10;3(9):e003138 [FREE Full text] [doi: 10.1136/bmjopen-2013-003138] [Medline: 24022390]
16. Johnson KA, Martin N, Nappi RE, Neal-Perry G, Shapiro M, Stute P, et al. Efficacy and safety of Fezolinetant in moderate to severe vasomotor symptoms associated with menopause: a phase 3 RCT. J Clin Endocrinol Metab 2023 Jul 14;108(8):1981-1997 [FREE Full text] [doi: 10.1210/clinem/dgad058] [Medline: 36734148]
17. FDA approves novel drug to treat moderate to severe hot flashes caused by menopause. U.S. Food & Drug Administration. 2023 May 16. URL: https://www.fda.gov/news-events/press-announcements/fda-approves-novel-drug-treat-moderate-severe-hot-flashes-caused-menopause [accessed 2024-01-08]
18. Mann E, Smith MJ, Hellier J, Balabanovic JA, Hamed H, Grunfeld EA, et al. Cognitive behavioural treatment for women who have menopausal symptoms after breast cancer treatment (MENOS 1): a randomised controlled trial. Lancet Oncol 2012 Mar;13(3):309-318 [FREE Full text] [doi: 10.1016/S1470-2045(11)70364-3] [Medline: 22340966]
19. Ayers B, Smith M, Hellier J, Mann E, Hunter MS. Effectiveness of group and self-help cognitive behavior therapy in reducing problematic menopausal hot flushes and night sweats (MENOS 2): a randomized controlled trial. Menopause 2012 Jul;19(7):749-759. [doi: 10.1097/gme.0b013e31823fe835] [Medline: 22336748]
20. Elkins G, Marcus J, Stearns V, Perfect M, Rajab MH, Ruud C, et al. Randomized trial of a hypnosis intervention for treatment of hot flashes among breast cancer survivors. J Clin Oncol 2008 Nov 01;26(31):5022-5026 [FREE Full text] [doi: 10.1200/JCO.2008.16.6389] [Medline: 18809612]
21. Elkins GR, Fisher WI, Johnson AK, Carpenter JS, Keith TZ. Clinical hypnosis in the treatment of postmenopausal hot flashes: a randomized controlled trial. Menopause 2013 Mar;20(3):291-298 [FREE Full text] [doi: 10.1097/gme.0b013e31826ce3ed] [Medline: 23435026]
22. Elkins GR, Barabasz AF, Council JR, Spiegel D. Advancing research and practice: the revised APA Division 30 definition of hypnosis. Int J Clin Exp Hypn 2015;63(1):1-9. [doi: 10.1080/00207144.2014.961870] [Medline: 25365125]
23. Hasan SS, Vasant D. The emerging new reality of hypnosis teletherapy: a major new mode of delivery of hypnotherapy and clinical hypnosis training. Int J Clin Exp Hypn 2023;71(2):153-164. [doi: 10.1080/00207144.2023.2185527] [Medline: 37040191]
24. Hussain FA. Hypnotherapy as an adjunct to medical treatment: highlighting effectiveness and identifying barriers to further integrative treatment. J Health Manag 2021 May 07;23(2):185-196. [doi: 10.1177/09720634211011557]
25. mHealth apps market size, share and global growth report. Fortune Business Insights. URL: https://www.fortunebusinessinsights.com/mhealth-apps-market-102020 [accessed 2024-04-10]

XSL•FO
RenderX

26. Snyder M, Elkins GR. Characteristics of users of a digital hypnotherapy intervention for hot flashes: retrospective study. JMIR Form Res 2024 Mar 14;8:e53555 [FREE Full text] [doi: 10.2196/53555] [Medline: 38483465]
27. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet e-surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
28. Sloan JA, Loprinzi CL, Novotny PJ, Barton DL, Lavasseur BI, Windschitl H. Methodologic lessons learned from hot flash studies. J Clin Oncol 2001 Dec 01;19(23):4280-4290. [doi: 10.1200/JCO.2001.19.23.4280] [Medline: 11731510]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys

Edited by A Mavragani; submitted 03.07.24; peer-reviewed by J Marcus, Z Kekecs; comments to author 06.08.24; revised version received 11.10.24; accepted 24.11.24; published 09.01.25.

Please cite as:
Scheffrahn K, Hall C, Muñiz V, Elkins G
User Outcomes for an App-Delivered Hypnosis Intervention for Menopausal Hot Flashes: Retrospective Analysis
JMIR Form Res 2025;9:e63948
URL: https://formative.jmir.org/2025/1/e63948
doi:10.2196/63948
PMID:

©Katherine Scheffrahn, Claire Hall, Vanessa Muñiz, Gary Elkins. Originally published in JMIR Formative Research (https://formative.jmir.org), 09.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development and Beta Validation of an mHealth-Based Hearing Screener (SRESHT) for Young Children in Resource-Limited Countries: Pilot Validation Study

Vidya Ramkumar[1], MASLP, PhD; Deepashree Joshi B[1], BASLP, MSc.Aud; Anil Prabhakar[2], PhD; James W Hall[3], PhD; Ramya Vaidyanath[1], PhD

[1]Faculty of Audiology and Speech Language Pathology, Sri Ramachandra Institute of Higher Education and Research, Chennai, India

[2]Department of Electrical Engineering, Indian Institute of Technology Madras, Chennai, India

[3]Osborne College of Audiology, Salus University, Pennsylvania, PA, United States

Corresponding Author:
Deepashree Joshi B, BASLP, MSc.Aud
Faculty of Audiology and Speech Language Pathology
Sri Ramachandra Institute of Higher Education and Research
Sri Ramachandra Nagar, Porur
Chennai, 600116
India
Phone: 91 9894170609
Email: deepasrijoshi25@sriramachandra.edu.in

Abstract

Background: The prevalence of hearing loss in infants in India varies between 4 and 5 per 1000. Objective-based otoacoustic emissions and auditory brainstem response have been used in high-income countries for establishing early hearing screening and intervention programs. Nevertheless, the use of objective screening tests in low- and middle-income countries (LMICs) such as India is not feasible. Mobile health (mHealth) solutions have been demonstrated to be a viable option for hearing screening in LMICs.

Objective: This study aims to develop and beta-validate an affordable hearing screener for children younger than 6 years of age to identify moderately severe or higher degrees of hearing loss.

Methods: In phase 1, a mHealth-based hearing screener (SRESHT) was developed using a single board computer with wireless commercial headphones and speakers as transducers, which were calibrated according to the standard procedure. Three subjective hearing screening modules were conceptualized and developed for different age groups: (1) behavioral observation audiometry–screening for infants aged from 0 to 1 year; (2) speech spectrum awareness task–screening for children 1 to 3 years old; and (3) speech recognition task–screening for children 3 to 6 years old. Different auditory stimuli for the screening modules were generated and suitability was assessed: (1) noisemakers, animal sounds, and environmental sounds for infants (birth to 1 year old); (2) animal sounds and nonsense syllables for children (1 to 3 years old); and (3) eighteen picturable spondee words for children (3 to 6 years old). In phase 2, the SRESHT screener was beta-validated in children aged below 6 years to establish the agreement between SRESHT modules and the gold-standard procedure in identifying moderately severe and higher degrees of hearing loss.

Results: Off-the-shelf commercial speakers and headphones were selected and calibrated. On comparison of stimuli for behavioral observation audiometry on 15 children, Noisemaker stimuli were found suitable based on the average minimum response levels. On comparison of different stimuli for speech spectrum awareness task on 15 children, animal sounds were found to be suitable. On familiarity check of 18 spondee words for speech recognition task among 20 children, 12 spondee words had the eligibility cutoff (85%) and a presentation level of 5 dB SL (re-pure tone threshold) was sufficient to achieve 80% psychometric function. In phase 2, a total of 55 children aged 0 to 6 years (31 normal hearing and 24 hearing impairment) underwent SRESHT screening for beta validation. Cohen κ indicated that the overall SRESHT screener had a very good agreement (κ=0.82) with gold-standard audiometric screening for identifying moderately severe and higher degrees of hearing loss.

Conclusions: The development and beta validation of the SRESHT screener using the selected auditory stimuli showed that the stimuli were suitable for screening children.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e53460) doi:10.2196/53460

KEYWORDS

audiometry; mHealth; devices; wireless; tablet-based screening; childhood hearing loss; early hearing detection and intervention; tablets; children; neonates; hearing loss; infants; development; validation; mobile phones

Introduction

Childhood hearing loss can have wide-ranging effects on a child's language, speech, education, and social integration, making this a crucial issue to address. To achieve optimal outcomes using amplification devices, it is necessary to screen, identify, and intervene children before 6 months of age [1]. In high-income countries, there are often policies for newborn hearing screening and hearing screening of young school-aged children [1]. On the other hand, in low- and middle-income countries (LMICs), such mandatory policies do not exist.

In India, the prevalence of hearing loss among neonates and infants is 5 per 1000 [2], and it ranges between 6% and 16% among children [3]. However, procuring and maintaining the recommended objective screening instruments like otoacoustic emissions and auditory brainstem response has not been viable at scale. Even in several other LMICs, only paper-pencil–based checklists are used to screen hearing loss, for example, Bangladesh [4], Thailand [5], Ecuador [6], Kenya [7], and Brazil [8]. The problem is further magnified by the lack of adequate ear and hearing care providers for early identification and rehabilitation of children with hearing loss. For example, in India, the ratio of audiologists to the population as per recent data from the Rehabilitation Council of India is 4.41 per 100,000 population [9] and in Bangladesh, the ratio of audiometrists to the population is 0.77% [10].

Task shifting to community workers or health care volunteers is a strategy that has been recommended [11] and explored [12-14] to overcome the stark service gaps. Such task shifting has resulted in successful implementation by improving program outcomes like coverage, refer rate, and follow-up rate [12,14]. This has been well supported by innovations in mobile health (mHealth). mHealth apps have decentralized hearing care from tertiary and secondary care hospitals to primary and community settings in some of the LMICs [14-16]. However, many of the mHealth solutions are focused on adult hearing loss and hearing aid fitting [17-20]. Smartphone or tablet-based apps for children are only available for those over the age of 4 years [21-24] and use either pure tones [22], digit in noise [25], or speech stimuli [21]. Therefore, there is a need for an affordable mHealth screener for younger children that can be used by community workers to accurately screen hearing loss and support early intervention in LMICs.

In general, subjective measures are less preferred for screening due to their lower reliability. For example, behavioral observation audiometry (BOA) is reported to be less accurate as it is often performed informally using uncalibrated stimuli without standardization of intensity, frequency, and distance [11,26-28]. Yet, recognizing the inequity in the availability of testing tools, the World Health Organization's guidelines for implementing community-based childhood hearing screening include the recommendation of using subjective-based screening tools [29].

Therefore, this study describes the development of a tablet-based subjective hearing screener (SRESHT screener) to identify children younger than 6 years of age with moderately severe hearing loss or higher degrees of hearing loss as per the definition of the Global Experts Group [30]. Since the variability of subjective measures is higher only for lower degrees of hearing loss [31,32], measures such as BOA and speech awareness and speech recognition–based screening modules were considered affordable alternatives to screen for hearing loss (HL; 60 dB and greater in the better ear). This preliminary level (60 dB HL and greater) of screening was chosen as part of a larger feasibility study to screen for hearing losses that are eligible for publicly funded welfare schemes in India. Similarly, in many other LMICs (eg, Bangladesh, Brazil), moderate or higher degrees of hearing loss are considered for welfare schemes [33], aids, and appliances with public funds. Therefore, the screening device is designed for application in similar contexts.

The specific objectives were to (1) identify the most appropriate auditory stimuli to screen for hearing loss in children aged between 0 and 6 years and (2) to evaluate the agreement between the pass or refer results of from behavioral screening using gold standard audiometer and of screening device in screening for moderately severe and above degrees of hearing loss.

Methods

Ethical Considerations

This cross-sectional study was approved by the institutional ethics committee of Sri Ramachandra Institute of Higher Education and Research (IEC-NI/20/OCT/76/113). Informed consent was obtained from parents of children who participated in the study. The subjects were included on a voluntary basis and no monetary benefits were given to the participants. All subjects were anonymized during data entry, analysis, and reporting to maintain confidentiality.

Development of SRESHT Hearing Screener

Hardware and Software

A single board computer was used as the hardware with software coded using Python (Python Software Foundation) language. App modules were developed for hearing screening for children from birth to 6 years of age. The user interface was designed in both English and the local language with simple texts, images, and symbols such that nontechnical screening personnel can use the software. Three hearing screening test procedures were included: BOA-based screening (for children younger than 1 year), speech spectrum awareness task (SSAT)–based screening (for children older than 1 year and younger than 3 years), and

XSL•FO
RenderX

speech recognition task (SRT)–based screening (for children older than 3 years and younger than 6 years).

The test module gets automatically selected based on the age selection in the app by the tester. The test modules include screening frequencies of 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz at a specified moderate intensity level (selected based on calibration). For each test procedure, age-appropriate stimuli were developed (described under each test procedure in detail) and uploaded to the software. An inbuilt noise monitoring software was included which provides access to the screening module only when the ambient noise levels are less than 50 dB (A). A cloud-based back-end data management was integrated via Google Drive. The screening result is displayed automatically based on the response entered for each frequency.

Selection of Transducers

Suitable commercial transducers appropriate to the test module and age were selected based on frequency response at 500 Hz to 4 kHz, output intensity range, and distortions at intensities up to 80 dB SPL (sound pressure level). At first, 3 commercial grade supra-aural headphones (Sony WHCH510 headphones, Sennheiser HD 450 BT, and Boatrockers 410), and 3 speakers (JBL Go 2, Sony SRSXB12, and Fire Boltt Xplode 1500) were selected for output comparison based on manufacturer information on frequency, intensity range, cost, and ease of availability in the market.

Headphone

Headphones were connected via Bluetooth to the SRESHT screener at maximum volume setting. The headphone diaphragm was coupled to the calibration microphone with a 6CC coupler connected to the Larsen Davis sound level meter (SLM).

Speaker

Bluetooth speakers were connected to the SRESHT screener at maximum volume setting and placed at a distance of 1 m and at 0-degree azimuth from the SLM microphone. A subject was seated in a chair and the ear level was marked from the floor. The subject was then removed and the SLM microphone was suspended from a stand approximately at the ear level marked.

For all the transducers, warble tone stimuli were presented at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz at different volume settings (%) of the SRESHT screener hardware, and the corresponding dB SPL values were measured in the SLM. The software was initially coded to produce the desired output level for moderate intensity levels in SPL (40 dB, 50 dB, and 60 dB) across the frequencies.

The Sony WHCH510 headphones and JBL Go 2 pro speakers were selected based on their accuracy of output, consistency of output, and output range (described in Development of the SRESHT Screener section).

Reference Equivalent Threshold Sound Pressure Level of Transducers

Sony WHCH510 Headphones

The mean thresholds of 10 normal hearing individuals were measured on a standard audiometer using TDH39 headphones from 500 Hz to 4000 Hz. The same individuals' threshold volume levels (%) were then obtained using the SRESHT hearing screener and Sony WHCH510 headphones. Larson Davis SLM was used to calculate the dB SPL value corresponding to the threshold volume level for each frequency. The Wilber method was used to estimate the reference equivalent threshold sound pressure level (RETSPL) values for the Sony WHCH510 headphones [34].

JBL Go2 RETSPL

The standard RETSPL values for speakers at 0-degree azimuth [35] were applied for 500 Hz to 4 kHz. The software was recoded as per the RETSPL value derivation.

Calibration Tone

A 1000 Hz warble tone was generated and used as a calibration tone, which is introduced in the app prior to the commencement of the screening. The minimum audible level of the tone at a moderate noisy environment of 40 to 45 dBA for 10 normal hearing subjects was measured and used to set the calibration tone.

Stimulus Validation

BOA-Based Hearing Screening

Both tonal and nontonal stimuli were assessed for BOA [32,36]. Studies suggested that children's responses to nontonal stimuli were as effective as tonal stimuli [36]. Nontonal stimuli included environmental sounds, animal sounds, and noisemakers that were contextually relevant to screen children. The stimuli were downloaded from an open-source website [37,38]. The acoustic parameters of the stimuli were then modified and corrected or normalized using Audacity (version 3.5.0; Audacity Team) and Apple Logic Pro (version 10.6; Apple Inc). The stimuli and the characteristics are summarized in Table 1.

XSL•FO
RenderX

Table 1. Nontonal stimuli categorized into noisemakers, environmental sounds, and animal sounds with center frequencies of 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz and a bandpass filter with 500-Hz bandwidth (filter slope 24 dB/octave).

Frequencies (center frequency, Hz)	Noisemakers	Environmental sounds	Animal sounds	Bandwidth required	Duration (seconds)
500	Drum or wooden drum	Horn	Cow mooing	250 Hz to 750 Hz (18 or 24 dB/octave slope)	2-3
1000	Wooden rattles	Applause	Dog barking	750 Hz to 1250 Hz (24 dB/octave slope)	2-3
2000	Metallic rattle or squeaking toy	Laughing	Whinnying horse and goat meh	1750 to 2250 Hz (24 dB/octave slope)	2-3
4000	Bell	Telephone ringing	Bird chirp and cat meow	3750 to 4250 Hz (24 dB/octave slope)	2-3

Procedure for Identifying Suitable Stimuli for BOA

A sound-treated room was preferred for conducting the screening test at the SRIHER clinic or laboratory. Participants were children with normal hearing and hearing loss between 0 to 1 year of age. Participants' details are described in Table 2.

A standard diagnostic audiometer (Madsen Astera[2]) with calibrated speakers (Martin Audio C115) was used to conduct the test using the newly developed stimuli (warble, noisemakers, environmental sounds, and animal sounds) at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz for children. Behavioral responses such as eyeblink, eye widening, searching, and head turn toward the source were considered acceptable responses. This was noted by an audiology intern. Initially, the stimuli presentation started at 20 dB HL in 500 Hz for children with normal hearing and at a higher intensity level for children with hearing loss. If a response was obtained, the stimuli presentation was presented thrice in the same intensity level to confirm the response and if not, the presentation was increased by 5 dB steps (ascending method). Whenever the child got distracted or bored, tangible or visual reinforcement was given by the audiology assistant. The same procedure was repeated for each stimulus at each frequency. Minimum response levels or thresholds were obtained for each stimulus at each frequency and compared to identify the stimuli that resulted in the lowest thresholds.

Table 2. Characteristics of children involved in BOA[a] stimuli validation.

Participants	Sex	Age	Hearing age	Diagnosis	Type of amplification
Child 1	Male	6 months	6 months	Normal hearing	N/A[b]
Child 2	Female	4 months	4 months	Normal hearing	N/A
Child 3	Male	6 months	6 months	Normal hearing	N/A
Child 4	Male	6 months	6 months	Normal hearing	N/A
Child 5	Male	6 months	6 months	Normal hearing	N/A
Child 6	Female	4 months	4 months	Normal hearing	N/A
Child 7	Female	8 months	8 months	Normal hearing	N/A
Child 8	Male	8 months	8 months	Normal hearing	N/A
Child 9	Male	3 months	3 months	Normal hearing	N/A
Child 10	Female	9 months	9 months	Normal hearing	N/A
Child 11	Female	5 months	5 months	Normal hearing	N/A
Child 12	Male	2 years 5 months	8 months	Bilateral: Severe to profound hearing loss	Right: Cochlear implant
Child 13	Male	2 years 6 months	10 months	Bilateral: Severe to profound hearing loss	Right: Cochlear implant
Child 14	Female	1 year 8 months	6 months	Bilateral: Moderately severe hearing loss	Bilateral: Hearing aid
Child 15	Female	2 years 2 months	8 months	Bilateral: Moderately severe hearing loss	Bilateral: Hearing aid

[a]BOA: behavioral observation audiometry.

[b]N/A: not applicable.

XSL•FO
RenderX

SSAT-Based Hearing Screening

For children between 1 and 3 years of age, speech-like animal sounds, such as a cows' *moo*, dogs' *bow-wow*, goats' *meh*, and cats' *meow*, and nonsense syllables like *mamama* and *bababa* were included.

The spectrum of the animal sounds using PRAAT software (version 6.3.2; Softsonic International) was compared to the spectrum of consonants and vowels to verify that it resembled the properties of speech. The formant frequencies (F1, F2, and F3) of animal sounds were approximately similar to the corresponding formant frequencies of the target vowels. Table 3 shows the formant frequencies of recorded animal sounds and existing normative vowel formant frequencies, as adapted from Raphael et al [39].

Table 3. Comparison of formant frequencies of animal sounds and normative formant frequencies of vowels matched approximately[a].

Formant frequencies	Animal sounds			
	Cow moo/ /u/	Dog bark/ /a/	Goat meh/ /ae/	Cat meow/ /i/
F1	300 Hz/300 Hz	853 Hz/ 730-850 Hz	1565 Hz/ 850-1100 Hz	1566 Hz/ 1500 Hz
F2	693 Hz/800 Hz	1180 Hz/1100-1220 Hz	2060 Hz/2050-2300 Hz	2217 Hz/2500 Hz
F3	986 Hz/2100 Hz	1616 Hz/2100-2400 Hz	2507 Hz/ 2800-3200 Hz	2835 Hz/ 3000 Hz

[a]/u/,/a/,/ae/,/i/ are English vowel sounds represented in international phonetic alphabet (IPA).

Procedure for Identifying Suitable Stimuli for SSAT

For children between 1 and 3 years of age, SSAT-based screening was conducted with the standard audiometer (Madsen Astera[2]), with TDH 39 headphones with newly developed stimuli. The participants included children with normal hearing and hearing impairment between the ages of 1 and 3 years. Behavioral responses included head turn, locating the source of the sound, and imitating the sound. Stimuli was presented initially at 1 kHz at 60 dB HL for normal hearing and 80 dB HL for children with hearing impairment. The Hughson-Westlake method of threshold tracking was used to identify the minimum response levels for each stimulus at each frequency. The minimum response levels for these stimuli were compared and the stimuli yielding better responses (thresholds) were finalized for the screening test procedure.

SRT-Based Hearing Screening

SRT-based screening was developed for children between 3 and 6 years of age. Eighteen picturable spondee words in the local language (Tamil) were excerpted from a standardized spondee list of 50 words [40] that is routinely used in clinical practice. The words were selected based on the relevance of the word in present-day context, and its picturability. Pictures corresponding to the 18 words were downloaded from the open-source web-based repository named Wikimedia Commons [41].

Procedure for Identifying Suitable Stimuli for SRT

A familiarity test of all the words and their corresponding picture was conducted. Participants were both children with normal hearing and children with hearing impairment above the age of 3 years. For familiarity testing of words, each child was asked to point out the appropriate picture when the word was read out at the conversation level. Similarly, for picture familiarity, each child was asked to name the picture.

These 18 spondee words were further divided into 3 lists of 6 words each. Then, these spondee words were recorded with a female voice centered at 1000 Hz and with loudness balanced at –3 dB peak value, having an interstimulus interval of 5 seconds. A closed-choice response set was developed with 1 target picture and 3 other pictures. The 4 pictures were arranged in a sequential manner with an equal number of repetitions.

Procedure for Determining Suitable Presentation Level for SRT Screening in Children

Participants were children with normal hearing and children with hearing impairment between the age group of 3 and 9 years. Using the standard audiometer (Madsen Astera[2]) with TDH 39 headphones, each child's pure tone thresholds (PTA) were obtained using the Hughson-Westlake method. Then, each child underwent SRT with picture identification as response. This was performed at different sensation levels (0 dB SL, 5 dB SL, 10 dB SL, and 20 dB SL with re-PTA) to identify the level at which 80% correct response (5/6 words) was obtained.

Beta Validation of SRESHT Screening Modules

This beta validation was conducted prior to the full validation in which the suitability of the stimuli was evaluated and agreement between new stimuli versus standard stimuli was compared.

Sample

Two groups of children participated in the study. The first group consisted of children under 6 years of age with normal hearing. The other group consisted of children with hearing impairment having a hearing age of less than 6 years and these children were younger than 10 years of chronological age. Children with any other comorbid conditions or any other known disabilities were excluded from the study (eg, children with developmental delay, vision impairment, autism spectrum disorders).

Children were recruited from the outpatient clinic of the tertiary care hospital attached to the institution where the study was conducted, as well as in camps organized in 2 rural districts of Tamil Nadu for the period of 3 months.

Informed written consent was obtained from the parents of the children before the commencement of data collection.

XSL•FO
RenderX

Procedure

The procedure was conducted in any quiet room near the outpatient department or the campsites with an ambient noise level of less than 50 dB (A) as recorded through the noise monitoring software of the SRESHT screener. SRESHT hearing screening with the finalized stimuli and screening modules for each age group was performed for all children using age-appropriate screening procedures at 60 dB HL.

Then, the minimum response level was estimated for 500 Hz, 1 kHz, 2 kHz, and 4 kHz with the standard diagnostic audiometer (Madsen Astera[2]) for each test procedure, as a gold standard. Screening results for the SRESHT hearing screener at 60 dB HL (pass or refer) were compared to audiometric thresholds more or less than 60 dB HL.

Analysis

Nonparametric tests were used to analyze the data as the results of the Shapiro-Wilk test showed that the data were not normally distributed.

The Friedman test was used to compare the children's responses to different stimuli in BOA. The Wilcoxon signed rank test was used to compare the children's responses to different stimuli for SSAT. Percentage analysis was used for word familiarity for SRT screening. Agreement statistics were used to beta-validate the screening results.

Results

Development of SRESHT Hearing Screener

Transducers

The summary of output characteristics of various speakers and headphones is shown in Table 4. The Sony WHCH510 headphones and JBL Go 2 pro speakers were selected based on their accuracy of output, consistency of output, output range, ease of use, and cost.

Table 4. Characteristics of each transducer (speakers or headphones) for the hearing screening module.

Transducers	Frequency response	Minimum output	Maximum output	Distortion
Sony WHCH510 headphones	No frequency splattering from 500 Hz to 4 kHz	40 dB SPL	100 dB SPL	Within permissible limits (ANSI[a], 2020)
Sennheiser HD450BT headphones	No frequency splattering from 500 Hz to 4 kHz	40 dB SPL	102 dB SPL	Within permissible limits (ANSI, 2020)
Boat rockers 410 headphones	Frequency splattering present at 4 kHz	50 dB SPL	104 dB SPL	Distortion present at 2 kHz and 4 kHz at 80 dB SPL
JBL Go 2 speakers	No frequency splattering from 500 Hz to 4 kHz	40 dB SPL	80 dB SPL	Within permissible limits (ANSI, 2020)
Sony SRSXB12 speakers	Appropriate from 500 Hz to 4 kHz	40 dB SPL	77 dB SPL	Within permissible limits (ANSI, 2020)
Fire Boltt Xplode1500 speakers	Frequency splattering present 2 kHz and 4 kHz	50 dB SPL	87 dB SPL	Distortion presented 2 kHz and 4 kHz at 80 dB SPL

[a]ANSI: American National Standards Institute [42].

RETSPL of Transducers

The RETSPL values derived for the Sony WHCH510 headphones are shown in Table 5.

Table 5. Sony WHCH510 headphone RETSPL[a] values.

Frequency	RETSPL (dB)
500 Hz	11
1 kHz	1
2 kHz	3
4 kHz	–2

[a]RETSPL: reference equivalent threshold sound pressure level.

Stimuli Validation

BOA-Based Hearing Screening

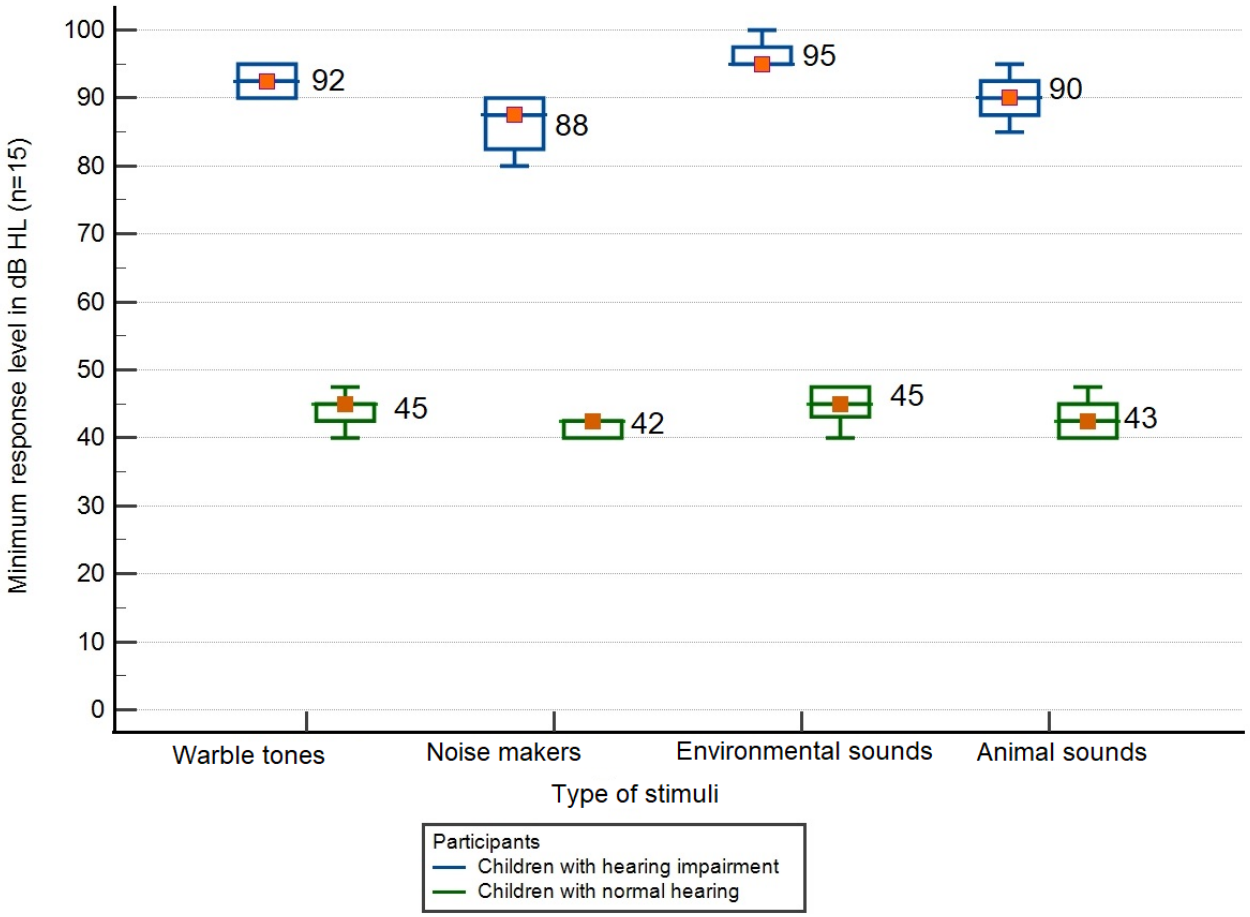
Fifteen children (11 Children with normal hearing and 4 Children with hearing impairment) in the age range of 0 to 1 year of age with a mean age of 6 (SD 1.9) months for children with normal hearing and a mean hearing age of 8 (SD 1.6) months for children with hearing impairment (Table 2), underwent the BOA using the newly developed stimuli. The Friedman test revealed no significant difference in the average minimum response levels of warble tone versus average

XSL•FO
RenderX

minimum response levels of noisemakers, animal sounds, and environmental sounds (n=15; χ^2_3=15.44; *P*=.089).

A comparison of the average minimum response levels for different stimuli for children with normal hearing and children with hearing impairment revealed that the average minimum responses to noisemakers for children with hearing impairment were superior to the responses to other stimuli, that is, animal sounds and environmental sounds (Figure 1). Therefore, noisemaker-based stimuli were selected for this age group, even though all 3 categories of non-tonal stimuli were appropriate for hearing screening.

Figure 1. Comparison of average minimum response levels for tonal and nontonal stimuli among children with normal hearing and children with hearing loss.



SSAT-Based Screening

A total of 15 children (10 children with normal hearing and 5 children with hearing impairment) in the age range of 1 to 3 years with a mean age of 2.3 (SD 0.37) years for children with normal hearing and 1.7 (SD 0.26) years of hearing age for children with hearing impairment (Table 6) underwent the SSAT. Using animal sound stimuli centered as frequencies corresponding to different speech sounds. Wilcoxon signed rank test showed a significant difference in the threshold for the nonsense syllables response and animal sounds response (*z*=–3.59; *P*<.001). The mean minimum response level difference between animal sounds was 5 dB better than the mean minimum response levels of nonsense syllables (Figure 2). Therefore, animal sounds were finalized for this age group.

XSL•FO
RenderX

Figure 2. Comparison of average minimum response levels for different stimuli among children with normal hearing and hearing loss (HL).

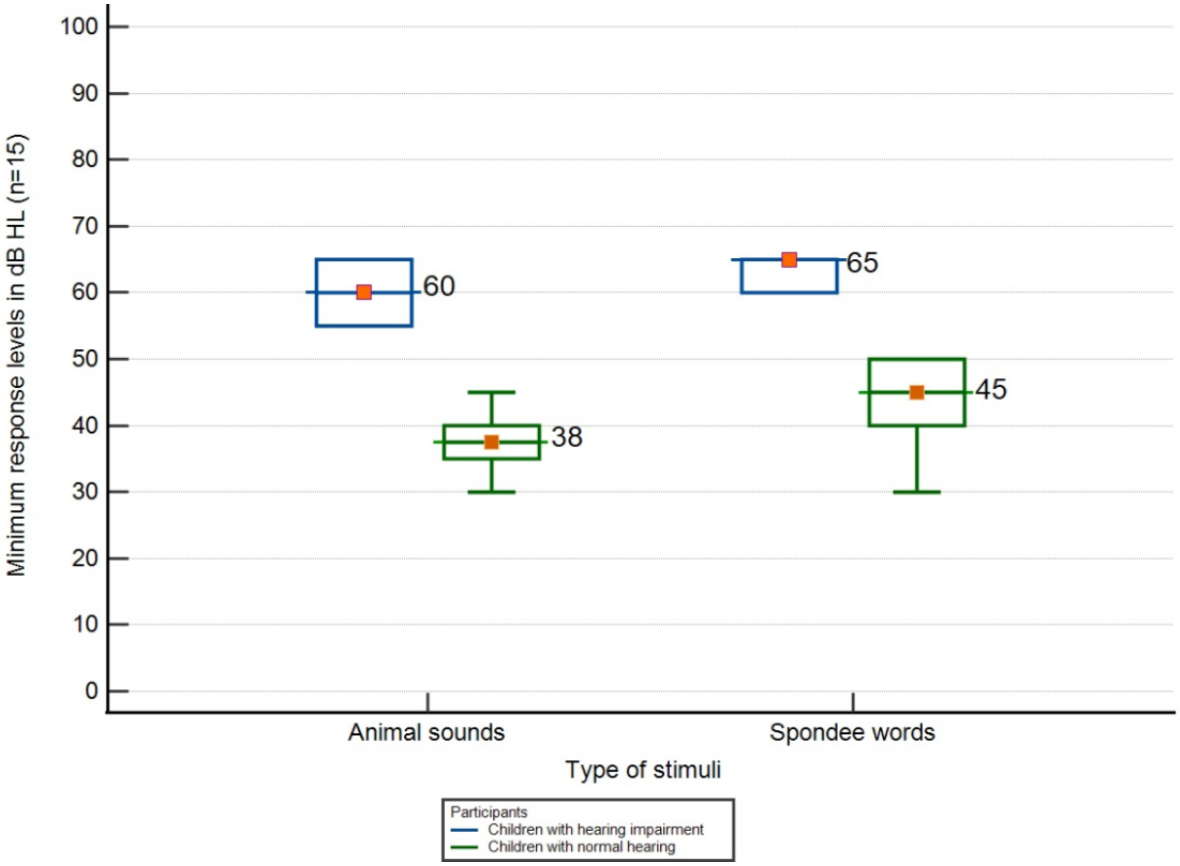


Table 6. Participants characteristics of children involved in SSAT[a] stimuli validation.

Participants	Sex	Chronological age (in years)	Hearing age (in years)	Diagnosis	Type of amplification
Child 1	Female	2.1	2.1	Normal hearing	N/A[b]
Child 2	Female	2.3	2.3	Normal hearing	N/A
Child 3	Male	2.5	2.5	Normal hearing	N/A
Child 4	Male	2.5	2.5	Normal hearing	N/A
Child 5	Male	2.5	2.5	Normal hearing	N/A
Child 6	Male	1.9	1.9	Normal hearing	N/A
Child 7	Female	2	2	Normal hearing	N/A
Child 8	Male	2.8	2.8	Normal hearing	N/A
Child 9	Male	2.9	2.9	Normal hearing	N/A
Child 10	Female	1.8	1.8	Normal hearing	N/A
Child 11	Male	3.7	1.5	Bilateral: Moderately severe hearing loss	Bilateral: Hearing aid user
Child 12	Female	4	1.8	Bilateral: Severe to profound hearing loss	Right: Cochlear implant
Child 13	Male	3.8	1.7	Bilateral: Severe hearing loss	Right: Cochlear implant; left: Hearing aid
Child 14	Female	4.9	2.2	Bilateral: Severe to profound hearing loss	Left: Cochlear implant
Child 15	Female	4.2	1.7	Bilateral: Moderately severe hearing loss	Bilateral: Hearing aid

[a]SSAT: speech spectrum awareness task.

[b]N/A: not applicable.

XSL•FO
RenderX

SRT-Based Screening

Twenty children (15 children with hearing impairment and 5 children with normal hearing between the ages of 3 to 9 years with a mean age of 3.7 (SD 0.44) years for children with normal hearing and 3.8 (SD 0.94) years of hearing age for children with hearing impairment (Table 7) underwent the speech recognition task using the spondee words.

Twelve of the 18 words could be identified by 85% of the normal-hearing children. Therefore, these words were included as 2 lists of 6 words each (Figure 3).

Thresholds for warble tone were compared with the speech recognition threshold to find the suitable presentation level for the spondee words for 80% psychometric function, 5 dB SL (re PTA) was found to be sufficient (Figure 4).

Table 7. Participants characteristics of children involved in SRT[a] stimuli validation.

Participants	Sex	Age (years)	Hearing age (years)	Diagnosis	Type of amplification
Child 1	Female	3.3	3.3	Normal hearing	N/A[b]
Child 2	Female	4.3	4.3	Normal hearing	N/A
Child 3	Male	4.2	4.2	Normal hearing	N/A
Child 4	Male	3.5	3.5	Normal hearing	N/A
Child 5	Male	3.6	3.6	Normal hearing	N/A
Child 6	Female	5.8	3	Bilateral: Severe to profound hearing loss	Right: Cochlear implant
Child 7	Male	4.1	3	Bilateral: Severe to profound hearing loss	Left: Cochlear implant
Child 8	Female	4.5	3.1	Bilateral: Moderate to moderately severe hearing loss	Bilateral: Hearing aid user
Child 9	Female	5	3.5	Bilateral: Profound hearing loss	Left: Cochlear implant
Child 10	Female	4.2	3	Right: Moderately severe hearing loss; left: Severe hearing loss	Left: Cochlear implant
Child 11	Female	8.5	5.2	Bilateral: Moderately severe hearing loss	Bilateral: Hearing aid user
Child 12	Male	7.1	4.5	Bilateral: Severe to profound hearing loss	Right: Cochlear implant
Child 13	Male	8	5.5	Bilateral: Moderately severe hearing loss	Right: Cochlear implant; left: Hearing aid
Child 14	Male	6.5	4.1	Left: Moderately severe hearing loss; right: Severe hearing loss	Bilateral: Hearing aid user
Child 15	Female	5	3.5	Bilateral: Moderately severe hearing loss	Bilateral: Hearing aid user
Child 16	Female	9	5	Bilateral: Severe to profound hearing loss	Left: Cochlear implant
Child 17	Female	7.5	3.2	Bilateral: Moderately severe hearing loss	Right: Cochlear implant
Child 18	Male	4.6	3	Bilateral: Severe hearing loss	Left: Cochlear implant
Child 19	Male	5.2	3.1	Bilateral: Severe to profound hearing loss	Right: Cochlear implant
Child 20	Male	4	3	Bilateral Severe to profound hearing loss	Left: Cochlear implant

[a]SRT: speech recognition task.

[b]N/A: not applicable.

XSL•FO
RenderX

Figure 3. Average percentage of correct responses by children for each word.

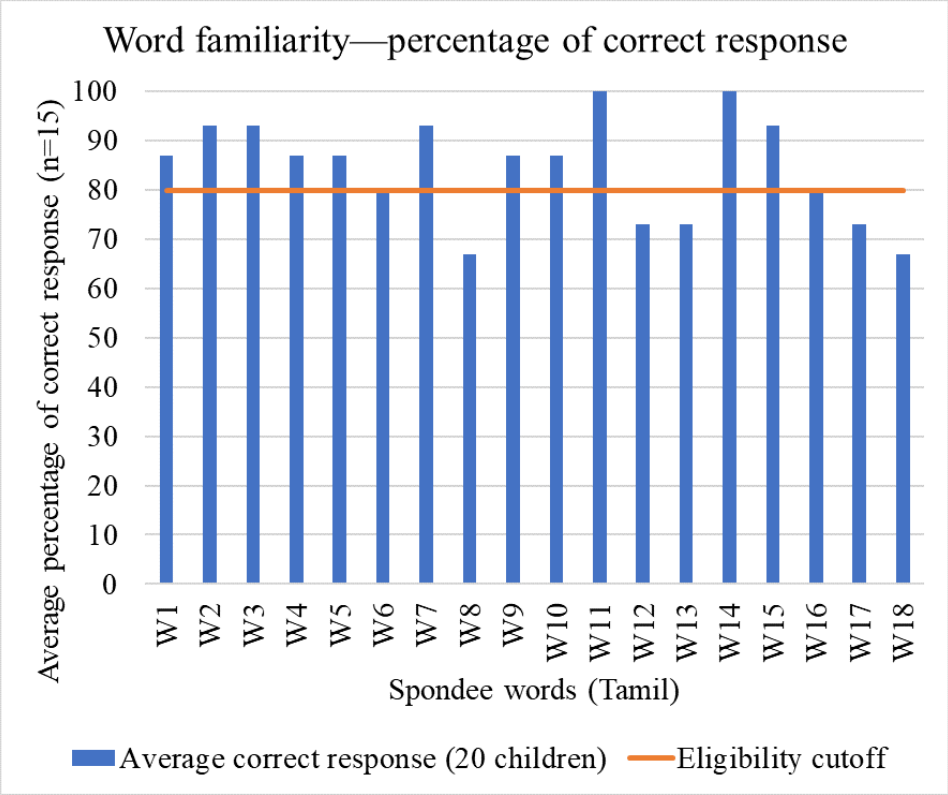
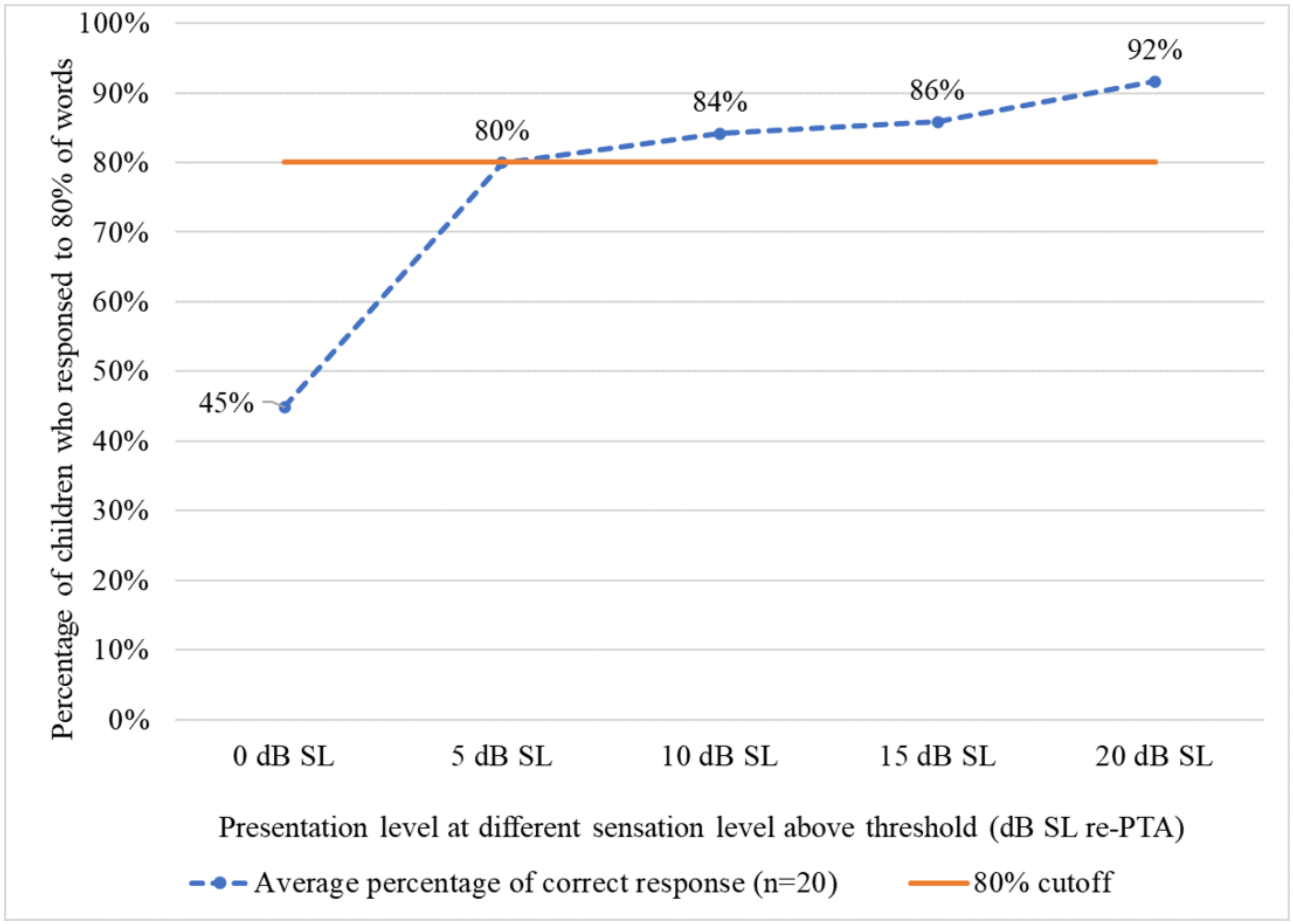


Figure 4. Comparison of percentage correct responses versus speech recognition threshold of 20 children. PTA: pure tone threshold.



XSL•FO
RenderX

Accuracy of SRESHT Screening Modules: Beta Validation

A total of 55 children (31 children with normal hearing and 24 children with hearing impairment underwent hearing screening using the SRESHT screener device. The chronological age range for children with normal hearing participants was 2 days to 6 years with a mean age of 3.2 (SD: 1.94) years, whereas the chronological age range for children with hearing impairment participants was 1 year to 9 years with a mean age of 4.9 (SD 1.42) years.

Of the 24 children with hearing impairment, 14 were unilateral cochlear implantees, 8 were bilateral hearing aid users, and 2 used bimodal hearing devices. The diagnostic reports of all 24 children with hearing impairment were reviewed to verify their unaided and most recent aided thresholds. Table 8 shows the agreement results.

Therefore, the SRESHT screener modules with new stimuli were as good as the standard audiometer in identifying moderately severe and higher degrees of hearing loss of 60 dB HL or higher.

Table 8. Agreement in findings between SRESHT screener and gold standard audiometric screening.

Test frequencies	κ value	95% CI	*P* value	Agreement
500 Hz	1.00	1.00-1.00	0.001	Perfect agreement
1000 Hz	0.87	0.75-1.00	0.001	Very good agreement
2000 Hz	0.72	0.47-0.97	0.005	Good agreement
4000 Hz	0.85	0.75-1.00	0.001	Very good agreement
Overall	0.82	0.66-0.97	0.001	Very good agreement

Discussion

Principal Findings

This study focused on the development and beta validation of an affordable subjective hearing screener (SRESHT) for identifying hearing losses greater than 60 dB HL in children younger than 6 years of age. The SRESHT hearing screener was developed in a single-board computer with calibrated off-the-shelf transducers. For children between 0 and 1 year of age, the BOA module was developed with noisemakers as suitable stimuli. For children between 1 and 3 years of age, the SSAT module with animal sounds was found as a suitable stimulus. For children between 3 and 6 years of age, the SRT module with 12 Tamil spondee words was found suitable. On beta validation of the SRESHT screener with age-appropriate modules among children between 0 to 6 years of age, it was found that the SRESHT screener was suitable to identify hearing losses of moderately severe and higher degrees.

The SRESHT screener is an affordable alternative with age-appropriate subjective hearing screening modules for children from birth to 6 years of age. The cost estimate is expected to be one-tenth of an objective hearing screener. Among the available smartphone-based hearing apps, only a few are designed specifically for children, but all are applicable only for children older than 4 years of age [18,21,24,43]. So far, to our best knowledge, the clinical population of children with hearing impairment has not been included in studies that describe mHealth hearing screeners.

In high-income countries, objective screening methods such as otoacoustic emissions and auditory brainstem response are routinely used as part of universal hearing screening as they are more accurate [26]. In LMICs, while there are efforts to use objective screeners by private hospitals and research teams, nationwide programs are negligible [2,44,45]. In a pilot government-initiated newborn hearing screening program in a southern state in India, objective screeners were initially procured, however, its long-term viability was called into question due to the high cost of screening equipment and frequent repair and consumables costs [46-48].

The semiautomated SRESHT screener is currently designed to screen “moderately severe” or higher degrees of hearing loss [49]. As a first step, this higher level of screening was chosen as part of a feasibility study to make a valid tool available to screen for 60 dB HL in the better ear (eligibility for publicly funded welfare schemes) within the public health system of a southern state in India. Not only in India, but in several LMICs (Bangladesh, Pakistan, and Kenya), a criterion of 50 dB HL or higher is used to provide welfare schemes, aid, and appliances with public funds. As a result, affordable tools that support this level of screening are useful in identifying this group of children who would otherwise miss out on welfare measures.

In the past, subjective screening using distraction and whisper tests in the United Kingdom [50,51], informal clap screening in Africa [52], noisemaker-based screening in India [53], and questionnaire-based screening in Thailand [5] and Kenya [7] have been used to support mass screening in the community. Some of the drawbacks of these screening methods are the lack of calibrated stimuli, the requirement of 2 personnel to administer the test, lack of standardization of intensity, frequency, distance, and test-retest reliability [11,26-28]. The SRESHT screening with age-appropriate calibrated stimuli reduced the extent of subjectivity. For the youngest babies, noisemakers are known to be robust [31,36,54] and this was reconfirmed from our beta validation. Stimuli such as animal sounds are known to be appealing to younger children for hearing testing [55], therefore, animal sounds were mapped to speech spectrums to develop the speech spectrum awareness test. These 2 modules are language independent and therefore can be used in several resource-limited settings across the world to support community-based screening efforts.

For the youngest age group (0 to 1 year), noisemakers are known to be robust [31,54], and this was reconfirmed from the results

XSL•FO
RenderX

of phase 1 of the study and used for the BOA screening module. In this study, BOA screening was conducted with noisemakers for infants under the age of 1 year. Visual reinforcement audiometry screening or distraction test is recommended for children aged 6 months to 1 year in a few studies [54] and guidelines [11,56], however, since the SRESHT screener is conceptualized for use by community workers and therefore the level of complexity and training required was an important consideration to choose BOA over visual reinforcement audiometry.

The initial version of the screening test for the age group of 1 year to 3 years consisted of a visual reinforcement task using warble tones. A simple light-emitting toy with a manual switch control was also developed for visual reinforcement. However, during field testing, the grassroots-level workers could not perform the screening task efficiently and provided feedback on practical challenges. This led to the modification of the module to a speech spectrum awareness task. The task was also tried in commercial grade insert earphones. However, insert earphones were not found suitable as children between 1 and 3 years of age were not comfortable to wear them.

For older children, the SRESHT screener uses picturable age-appropriate spondee words in the local language. While many apps have used pure tones [18,57]; warble tones [22,43], and very few exist with speech stimuli [21,24]. Picturable spondee words of different languages can be added to make the screener available elsewhere.

Affordable screeners are much needed to support nationwide early intervention programs in LMICs for long-term viability. The cost of screening a child for hearing loss using objective screening methods is estimated to be between US $17 and US $26 across high-income countries and LMICs like Brazil [58], India [59], the United States [60], and China, [61,62]. This cost can be lowered when device costs are less. Hearing screening apps for adults and older children range in price from US $600 to US $2000 [63], and some require standard audiometric headphones [18,22,24]. Commercial headphones with precise calibrations, on the other hand, can produce consistent output [16,17,21]. To keep costs low, the SRESHT screener was developed using off-the-shelf transducers that produce standard output with RETSPL derivation thereby reducing the cost to one-tenth the price of objective screening tools.

In community-based, school-based or mass screening programs, in-built noise monitoring is crucial due to the lack of sound-treated booths. The SRESHT screener performs noise monitoring prior to the commencement of each of the screening tests to limit the contamination of results by high ambient noise levels. Such features exist in some of the app-based screeners already [23,43,64].

The SRESHT screener is currently limited to permanent hearing losses of moderately severe and higher degrees in children under 6 years of age. Hence, further study is required to extend the scope of the screener to all types and degrees of hearing loss. While the scope of the current study is limited to the development of the screening modules with suitable beta validation, the full validation of the hearing screener with these finalized parameters has been undertaken and will be reported as a separate paper.

Future Directions

Full validation against the gold standard objective test is undertaken for this device and the sensitivity, specificity, and positive and negative predictive values will be reported in a separate publication. A project is also being undertaken to lower the screening intensity levels to screen for lower degrees of hearing loss.

Acknowledgments

This study was conducted as a part of a larger project funded by the Department of Biotechnology/Wellcome Trust India Alliance.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

VR and DJB researched articles and came up with the design of the software. AP was involved in the hardware and software development. Data collection and analysis was done by DJB. Reviewing of the analysis was done by VR. Manuscript writing was done by DJB and VR. The manuscript review was done by JWH and RV.

Conflicts of Interest

None declared.

References

1. American Academy of Pediatrics, Joint Committee on Infant Hearing. Year 2007 position statement: Principles and guidelines for early hearing detection and intervention programs. Pediatrics 2007 Oct;120(4):898-921. [doi: 10.1542/peds.2007-2333] [Medline: 17908777]

XSL•FO
RenderX

2. Joshi B, Ramkumar V, Nair LS, Kuper H. Early hearing detection and intervention (EHDI) programmes for infants and young children in low-income and middle-income countries in Asia: a systematic review. BMJ Paediatr Open 2023 Jan;7(1):1-7 [FREE Full text] [doi: 10.1136/bmjpo-2022-001752] [Medline: 36720503]
3. Verma R, Konkimalla A, Thakar A, Sikka K, Singh A, Khanna T. Prevalence of hearing loss in India. NMJI 2022 Jan 29;34:216-222 [FREE Full text] [doi: 10.25259/nmji_66_21]
4. Berg A, Papri H, Ferdous S, Khan NZ, Durkin MS. Screening methods for childhood hearing impairment in rural Bangladesh. Int J Pediatr Otorhinolaryngol 2006 Jan;70(1):107-114. [doi: 10.1016/j.ijporl.2005.05.029] [Medline: 16023223]
5. Poonual W, Navacharoen N, Kangsanarak J, Namwongprom S. Risk factors for hearing loss in infants under universal hearing screening program in Northern Thailand. J Multidiscip Healthc 2016;9:1-5 [FREE Full text] [doi: 10.2147/JMDH.S92818] [Medline: 26766912]
6. Muñoz K, Luzuriaga E, Callow-Heusser C, Ortiz E, White K. Evaluation of a hearing screening questionnaire for use with Ecuadorian school-aged children. Int J Audiol 2015;54(9):587-592. [doi: 10.3109/14992027.2015.1015689] [Medline: 25804301]
7. Newton VE, Macharia I, Mugwe P, Ototo B, Kan SW. Evaluation of the use of a questionnaire to detect hearing loss in Kenyan pre-school children. Int J Pediatr Otorhinolaryngol 2001 Mar;57(3):229-234. [doi: 10.1016/s0165-5876(00)00453-5] [Medline: 11223455]
8. Samelli AG, Rabelo CM, Vespasiano APC. Development and analysis of a low-cost screening tool to identify and classify hearing loss in children: a proposal for developing countries. Clinics (Sao Paulo) 2011;66(11):1943-1948 [FREE Full text] [doi: 10.1590/s1807-59322011001100015] [Medline: 22086526]
9. 35th Annual Report. Rehabilitation Council of India. 2022. URL: https://rehabcouncil.nic.in/node/467 [accessed 2024-11-25]
10. Al Imam MH, Jahan I, Das MC, Muhit M, Akbar D, Badawi N, et al. Situation analysis of rehabilitation services for persons with disabilities in Bangladesh: identifying service gaps and scopes for improvement. Disabil Rehabil 2022 Sep;44(19):5571-5584. [doi: 10.1080/09638288.2021.1939799] [Medline: 34176400]
11. Hearing screening: considerations for implementation. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/9789240032767 [accessed 2021-09-03]
12. de Kock T, Swanepoel DW, Hall JW. Newborn hearing screening at a community-based obstetric unit: Screening and diagnostic outcomes. Int J Pediatr Otorhinolaryngol 2016 May;84:124-131. [doi: 10.1016/j.ijporl.2016.02.031] [Medline: 27063767]
13. Shinn JR, Jayawardena ADL, Patro A, Zuniga MG, Netterville JL. Teacher Prescreening for Hearing Loss in the Developing World. Ear Nose Throat J 2021 Jun;100(3_suppl):259S-262S [FREE Full text] [doi: 10.1177/0145561319880388] [Medline: 31608682]
14. Yousuf Hussein S, Swanepoel DW, Mahomed F, Biagio de Jager L. Community-based hearing screening for young children using an mHealth service-delivery model. Glob Health Action 2018;11(1):1467077 [FREE Full text] [doi: 10.1080/16549716.2018.1467077] [Medline: 29764328]
15. Chu Y, Cheng Y, Lai Y, Tsao Y, Tu T, Young ST, et al. A Mobile Phone-Based Approach for Hearing Screening of School-Age Children: Cross-Sectional Validation Study. JMIR Mhealth Uhealth 2019 Apr 01;7(4):e12033 [FREE Full text] [doi: 10.2196/12033] [Medline: 30932870]
16. Dawood N, Mahomed Asmail F, Louw C, Swanepoel DW. Mhealth hearing screening for children by non-specialist health workers in communities. Int J Audiol 2021 Apr;60(sup1):S23-S29. [doi: 10.1080/14992027.2020.1829719] [Medline: 33043733]
17. Renda L, Selçuk ÖT, Eyigör H, Osma Ü, Yılmaz MD. Smartphone Based Audiometric Test for Confirming the Level of Hearing; Is It Useable in Underserved Areas? J Int Adv Otol 2016 Apr;12(1):61-66 [FREE Full text] [doi: 10.5152/iao.2016.1421] [Medline: 27340985]
18. Khoza-Shangase K, Kassner L. Automated screening audiometry in the digital age: exploring uhear™ and its use in a resource-stricken developing country. Int J Technol Assess Health Care 2013 Jan;29(1):42-47. [doi: 10.1017/S0266462312000761] [Medline: 23298579]
19. Tao KFM, Brennan-Jones CG, Jayakody DMP, Swanepoel DW, Fava G, Bellekom SR, et al. Validation of teleaudiology hearing aid rehabilitation services for adults: a systematic review of outcome measurement tools. Disabil Rehabil 2022 Aug;44(16):4161-4178. [doi: 10.1080/09638288.2021.1900928] [Medline: 33787435]
20. Samelli AG, Rabelo CM, Sanches SGG, Aquino CP, Gonzaga D. Tablet-Based Hearing Screening Test. Telemed J E Health 2017 Sep;23(9):747-752. [doi: 10.1089/tmj.2016.0253] [Medline: 28328389]
21. Dillon H, Mee C, Moreno JC, Seymour J. Hearing tests are just child's play: the sound scouts game for children entering school. Int J Audiol 2018 Jul;57(7):529-537. [doi: 10.1080/14992027.2018.1463464] [Medline: 29703099]
22. Kam ASC, Li LKC, Yeung KNK, Wu W, Huang Z, Wu H, et al. Automated hearing screening for preschool children. J Med Screen 2014 Jun;21(2):71-75. [doi: 10.1177/0969141314528322] [Medline: 24622246]
23. Swanepoel DW, Myburgh HC, Howe DM, Mahomed F, Eikelboom RH. Smartphone hearing screening with integrated quality control and data management. Int J Audiol 2014 Dec;53(12):841-849. [doi: 10.3109/14992027.2014.920965] [Medline: 24998412]

XSL•FO
RenderX

24. Yimtae K, Israsena P, Thanawirattananit P, Seesutas S, Saibua S, Kasemsiri P, et al. A Tablet-Based Mobile Hearing Screening System for Preschoolers: Design and Validation Study. JMIR Mhealth Uhealth 2018 Oct 23;6(10):e186 [FREE Full text] [doi: 10.2196/mhealth.9560] [Medline: 30355558]
25. Mahomed-Asmail F, Swanepoel D, Eikelboom RH, Myburgh HC, Hall J. Clinical Validity of hearScreen™ Smartphone Hearing Screening for School Children. Ear Hear 2016;37(1):e11-e17. [doi: 10.1097/AUD.0000000000000223] [Medline: 26372265]
26. Yoshinaga-Itano C, Manchaiah V, Hunnicutt C. Outcomes of universal newborn screening programs: systematic review. J Clin Med 2021;10(13):2784 [FREE Full text] [doi: 10.3390/jcm10132784] [Medline: 34202909]
27. Mott A, Emond A. What is the role of the distraction test of hearing? Arch Dis Child 1994 Jan;70(1):10-13 [FREE Full text] [doi: 10.1136/adc.70.1.10] [Medline: 8109999]
28. Weir CG. Use of behavioural tests in early diagnosis of hearing loss. Acta Otolaryngol Suppl 1985;421:86-92. [doi: 10.3109/00016488509121761] [Medline: 3862333]
29. World report on hearing. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/9789240020481 [accessed 2021-03-03]
30. Olusanya BO, Davis AC, Hoffman HJ. Hearing loss grades and the international classification of functioning, disability and health. Bull World Health Organ 2019;97(10):725-728 [FREE Full text] [doi: 10.2471/BLT.19.230367] [Medline: 31656340]
31. Ramesh A, Jagdish C, Nagapoorinima M, Suman Rao PN, Ramakrishnan AG, Thomas GC, et al. Low cost calibrated mechanical noisemaker for hearing screening of neonates in resource constrained settings. Indian J Med Res 2012;135(2):170-176 [FREE Full text] [Medline: 22446858]
32. Karikoski JO, Marttila TT, Jauhiainen T. Behavioural observation audiometry in testing young hearing-impaired children. Scand Audiol 1998;27(3):183-187. [doi: 10.1080/010503998422700] [Medline: 9728780]
33. Bangladesh Persons with Disablity Welfare Act. World Health Organization. 2001. URL: https://extranet.who.int/mindbank/item/3767 [accessed 2011-12-31]
34. Wilber LA, Kruger B, Killion MC. Reference thresholds for the ER-3A insert earphone. J Acoust Soc Am 1988 Feb;83(2):669-676. [doi: 10.1121/1.396162] [Medline: 3351125]
35. ASA/ANSI S3.6-2018 (R2023): Specification For Audiometers. American National Standards Institute. 2018. URL: https://blog.ansi.org/asa-ansi-s3-6-2018-r2023-audiometers-specification/ [accessed 2025-01-08]
36. Massie R, Dillon H, Ching T, Birtles G. Birtles G (2005) The evaluation of nontonal stimuli for hearing assessment of young children. Aust N Z J Audiol 2005;27(1):10-17 [FREE Full text] [doi: 10.1375/audi.2005.27.1.10]
37. Free sound effects downloads. ZapSplat. URL: https://www.zapsplat.com/ [accessed 2024-09-24]
38. Royalty free stock music. Mixkit. URL: https://mixkit.co/free-stock-music/ [accessed 2024-09-24]
39. Raphael LJ, Borden GJ, Harris KS. Speech Science Primer: Physiology, Acoustics, and Perception of Speech. Philadelphia, PA: Lippincott Williams and Wilkins; 2011.
40. Dayalan J. Development and Standardization of Phonetically Balanced Test Materials in Tamil Language [dissertation]. Mysore, India: Mysore University; 1976.
41. Wikimedia Commons. URL: https://commons.wikimedia.org/wiki/Main_Page [accessed 2024-12-12]
42. Procedure for Measuring the Ambient Noise Level in a Room. American National Standards Institute. 2020. URL: https://webstore.ansi.org/preview-pages/ASA/preview_ANSI+ASA+S12.72-2015.pdf [accessed 2022-03-05]
43. Yeung J, Javidnia H, Heley S, Beauregard Y, Champagne S, Bromwich M. The new age of play audiometry: prospective validation testing of an iPad-based play audiometer. J Otolaryngol Head Neck Surg 2013;42(1):21 [FREE Full text] [doi: 10.1186/1916-0216-42-21] [Medline: 23663317]
44. Olusanya BO. Neonatal hearing screening and intervention in resource-limited settings: an overview. Arch Dis Child 2012;97(7):654-659. [doi: 10.1136/archdischild-2012-301786] [Medline: 22611062]
45. Olusanya BO. Highlights of the new WHO Report on Newborn and Infant Hearing Screening and implications for developing countries. Int J Pediatr Otorhinolaryngol 2011 Jun;75(6):745-748. [doi: 10.1016/j.ijporl.2011.01.036] [Medline: 21728253]
46. Galhotra A, Sahu P. Challenges and solutions in implementing hearing screening program in India. Indian J Community Med 2019;44(4):299-302 [FREE Full text] [doi: 10.4103/ijcm.ijcm_73_19]
47. Olusanya B. Screening for neonatal deafness in resource-poor countries: challenges and solutions. RRN 2015 May:51. [doi: 10.2147/rrn.s61862]
48. Krishnan LA, Donaldson LK. Newborn hearing screening in developing countries: understanding the challenges and complexities of implementation. Perspect Glob Issues Commun Sci Relat Disord 2013;3(2):54-61 [FREE Full text] [doi: 10.1044/gics3.2.54]
49. Informal Working Group on Prevention of Deafness and Hearing Impairment Programme Planning, World Health Organization. Report of the Informal Working Group on Prevention of Deafness and Hearing Impairment Programme Planning, Geneva, 18-21 June 1991. World Health Organization. 1991. URL: https://iris.who.int/handle/10665/58839 [accessed 1991-07-21]

XSL•FO
RenderX

50. Owen M, Webb M, Evans K. Community based universal neonatal hearing screening by health visitors using otoacoustic emissions. Arch Dis Child Fetal Neonatal Ed 2001 May;84(3):F157-F162 [FREE Full text] [doi: 10.1136/fn.84.3.f157] [Medline: 11320040]
51. Pirozzo S, Papinczak T, Glasziou P. Whispered voice test for screening for hearing impairment in adults and children: systematic review. BMJ 2003 Oct 25;327(7421):967 [FREE Full text] [doi: 10.1136/bmj.327.7421.967] [Medline: 14576249]
52. Olusanya B. Early detection of hearing impairment in a developing country: What options? Int J Audiol 2001 Jan;40(3):141-147 [FREE Full text] [doi: 10.3109/00206090109073109]
53. Singh AK, Kumar R, Mishra CK, Khera A, Srivastava A. Moving from Survival to Healthy Survival through Child Health Screening and Early Intervention Services Under Rashtriya Bal Swasthya Karyakram (RBSK). Indian J Pediatr 2015 Nov;82(11):1012-1018. [doi: 10.1007/s12098-015-1823-2] [Medline: 26199076]
54. Valencia K. Testing small children's hearing with little or no equipment. Community Ear Hear Health 2015;12(16):1 [FREE Full text]
55. Nolte A, Coninx F, Müller F, Hess M, Wiesner T, Dudek N, et al. Frequency-specific Animal Sound Test (FAST) 4: A valid method for hearing screening. Int J Pediatr Otorhinolaryngol 2016 Feb;81:68-79. [doi: 10.1016/j.ijporl.2015.12.004] [Medline: 26810294]
56. Clinical practice guidelines: Childhood hearing screening. American Academy of Audiology. 2011. URL: https://www.audiology.org/practice-guideline/clinical-practice-guidelines-childhood-hearing-screening/ [accessed 2022-03-07]
57. Wu W, Lü J, Li Y, Kam ACS, Fai Tong MC, Huang Z, et al. A new hearing screening system for preschool children. Int J Pediatr Otorhinolaryngol 2014 Feb;78(2):290-295. [doi: 10.1016/j.ijporl.2013.11.026] [Medline: 24360950]
58. Bevilacqua MC, Alvarenga KDF, Costa OA, Moret ALM. The universal newborn hearing screening in Brazil: from identification to intervention. Int J Pediatr Otorhinolaryngol 2010 May;74(5):510-515. [doi: 10.1016/j.ijporl.2010.02.009] [Medline: 20303604]
59. Ramkumar V, John KR, Selvakumar K, Vanaja CS, Nagarajan R, Hall JW. Cost and outcome of a community-based paediatric hearing screening programme in rural India with application of tele-audiology for follow-up diagnostic hearing assessment. Int J Audiol 2018 Jun;57(6):407-414. [doi: 10.1080/14992027.2018.1442592] [Medline: 29490519]
60. Shulman S, Besculides M, Saltzman A, Ireys H, White KR, Forsman I. Evaluation of the universal newborn hearing screening and intervention program. Pediatrics 2010 Aug;126 Suppl 1:S19-S27. [doi: 10.1542/peds.2010-0354F] [Medline: 20679316]
61. Wenjin W, Xiangrong T, Yun LI, Jingrong L, Jianyong C, Xueling W, et al. Neonatal hearing screening in remote areas of China: a comparison between rural and urban populations. J Int Med Res 2018 Feb;46(2):637-651 [FREE Full text] [doi: 10.1177/0300060517706643] [Medline: 28606020]
62. Dai P, Huang L, Wang G, Gao X, Qu C, Chen X, et al. Concurrent Hearing and Genetic Screening of 180,469 Neonates with Follow-up in Beijing, China. Am J Hum Genet 2019 Oct 03;105(4):803-812 [FREE Full text] [doi: 10.1016/j.ajhg.2019.09.003] [Medline: 31564438]
63. Bright T, Pallawela D. Validated Smartphone-Based Apps for Ear and Hearing Assessments: A Review. JMIR Rehabil Assist Technol 2016 Dec 23;3(2):e13. [doi: 10.2196/rehab.6074] [Medline: 28582261]
64. Maclennan-Smith F, Swanepoel D, Hall JW. Validity of diagnostic pure-tone audiometry without a sound-treated environment in older adults. Int J Audiol 2013 Feb;52(2):66-73. [doi: 10.3109/14992027.2012.736692] [Medline: 23140522]

Abbreviations

BOA: behavioral observation audiometry
HL: hearing loss
LMIC: low- and middle-income country
mHealth: mobile health
PTA: pure tone threshold
RETSPL: reference equivalent threshold sound pressure level
SLM: sound level meter
SRT: speech recognition task
SSAT: speech spectrum awareness task

XSL•FO
RenderX

Edited by A Mavragani; submitted 09.10.23; peer-reviewed by K Ganapathy, C Frisby; comments to author 04.09.24; revised version received 24.09.24; accepted 10.10.24; published 13.01.25.

Please cite as:
Ramkumar V, Joshi B D, Prabhakar A, Hall JW, Vaidyanath R
Development and Beta Validation of an mHealth-Based Hearing Screener (SRESHT) for Young Children in Resource-Limited Countries: Pilot Validation Study
JMIR Form Res 2025;9:e53460
URL: https://formative.jmir.org/2025/1/e53460
doi:10.2196/53460
PMID:

©Vidya Ramkumar, Deepashree Joshi B, Anil Prabhakar, James W Hall, Ramya Vaidyanath. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Healthy Kai (Food) Checker Web-Based Tool to Support Healthy Food Policy Implementation: Development and Usability Study

Magda Rosin[1,2], PhD; Cliona Ni Mhurchu[1,2], PhD; Elaine Umali[3], MPH; Sally Mackay[1], PhD

[1]Department of Epidemiology and Biostatistics, School of Population Health, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand

[2]Centre for Translational Health Research: Informing Policy and Practice (TRANSFORM), Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand

[3]National Institute for Health Innovation, School of Population Health, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand

Corresponding Author:
Magda Rosin, PhD
Department of Epidemiology and Biostatistics
School of Population Health, Faculty of Medical and Health Sciences
University of Auckland
Building 507
22-30 Park Ave, Grafton
Auckland, 1023
New Zealand
Phone: 64 093737599 ext 88733
Email: m.rosin@auckland.ac.nz

Abstract

Background: Public health programs and policies can positively influence food environments. In 2016, a voluntary National Healthy Food and Drink Policy was released in New Zealand to improve the healthiness of food and drinks for hospital staff and visitors. However, no resources were developed to support policy implementation.

Objective: This study aimed to design, develop, and test a new web-based tool to support food providers implementing the National Healthy Food and Drink Policy in New Zealand.

Methods: The Double Diamond model, a structured framework with 4 design phases, was used to design and develop a web-based tool. Findings from our previous research, such as (1) systematic review of barriers and facilitators to workplace healthy food policy implementation; (2) scoping review of current tools and resources available in New Zealand, Australia, and Canada; (3) interviews with food providers and public health nutrition professionals; and (4) food and drink availability audit results in New Zealand hospitals were used in the "Discover" (understanding of current gaps) and "Define" (prioritizing functions and features) phases. Subsequent phases focused on generating ideas, creating prototypes, and testing a new web-based tool using Figma, a prototyping tool. During the "Develop" phase, project stakeholders (11 public health nutrition professionals) provided feedback on the basic content outline of the initial low-fidelity prototype. In the final "Deliver" phase, a high-fidelity prototype resembling the appearance and functionality of the final tool was tested with 3 end users (public health nutrition professionals) through interactive interviews, and user suggestions were incorporated to improve the tool.

Results: A new digital tool, Healthy Kai (Food) Checker—a searchable database of packaged food and drink products that classifies items according to the Policy's nutritional criteria—was identified as a key tool to support Policy implementation. Of 18 potential functions and features, 11 were prioritized by the study team, including basic and advanced searches for products, sorting list options, the ability to compile a list of selected products, a means to report products missing from the database, and ability to use on different devices. Feedback from interview participants was that the tool was easy to use, was logical to navigate, and had an appealing color scheme. Suggested visual and usability improvements included ensuring that images represented the diverse New Zealand population, reducing unnecessary clickable elements, adding information about the free registration option, and including more frequently asked questions.

Conclusions: Comprehensive research informed the development of a new digital tool to support implementation of the National Healthy Food and Drink Policy. Testing with end users identified features that would further enhance the tool's acceptability and

XSL•FO
RenderX

usability. Incorporation of more functions and extending the database to include products classified according to the healthy school lunches program policy in the same database would increase the tool's utility.

(JMIR Form Res 2025;9:e60447) doi:10.2196/60447

KEYWORDS

healthy food policy; policy implementation; nutrition; food environment; hospital; workplace; web-based tool; digital tool; database; user testing; food; drink; nutrition professionals; acceptability; usability

Introduction

Background

Despite the importance of healthy eating in the prevention of chronic diseases, the nutritional quality of diets worldwide does not align with evidence-based dietary guidelines, with widespread overconsumption of unhealthy and ultraprocessed foods and drinks [1,2]. However, foods are chosen in the context of local food environments [3,4], which in New Zealand are dominated by less healthy choices [5]. Workplace food environments, including hospitals, have been recognized as an important factor for employees' health promotion and well-being due to their potential for broad and sustainable population reach [6], as well as the fact that the majority of the adult population works [7-9]. When this study began, the New Zealand health system was structured into 20 district health boards (DHBs) responsible for delivering national health care services through hospitals and clinical centers. However, in 2022, the DHB system was disestablished as part of a national health system reform [10]. The provision of health care services was centralized under the newly established Health New Zealand—Te Whatu Ora agency that directly employs around 90,000 staff, making it a major nationwide employer [11].

A voluntary National Healthy Food and Drink Policy (the Policy) was released in New Zealand in 2016 [12] to encourage hospitals to provide healthier food and drink options for staff and visitors (not inpatients). The Policy includes a traffic light classification system for food and drinks, where *healthy* (*Green*) options should make up at least 55% of available items; the *less healthy* (*Amber*) options that might still provide some nutritive value should be less prevalent (<45% of available items); and the *unhealthy* (*Red*) items, such as confectionery, deep-fried foods and sugar-sweetened drinks, are not permitted [12]. Yet, 2021 evaluation findings showed that 5 years after the introduction of the voluntary Policy in 2016, only 8 of the 20 DHBs had adopted the Policy, resulting in regional inconsistency [13]. Furthermore, on-site audits of food and drink availability in 2021/2022 indicated that the implementation was inconsistent and largely unsuccessful, with *Red* and *Amber* choices predominating [14].

Previous Work

We conducted a systematic literature review of facilitators and barriers to implementing healthy food and drink policies in public sector workplaces. The review indicated that tailored supportive tools and resources facilitate the implementation of policies promoting healthy food environments [15]. A key reason why such support is required is because identifying healthier items can be challenging for food providers, who may find the policy's nutrition criteria too complex to understand without nutrition expertise [15]. The supportive tools identified in the systematic review [15] were closely aligned with the supporting materials recommended by the World Health Organization [16] in their 2021 action framework for the implementation of effective healthy food policies in public settings. Examples of useful tools included administrative materials (eg, contract templates and standardized procedural guidelines); customer communication resources, compliance monitoring tools; and databases of compliant recipes, suppliers, and products [15,16]. The most useful tools were policy specific, readily available, and centrally provided, limiting the burden on individual food providers [15]. Importantly, supportive tools needed to be practical, up-to-date, user-friendly, and useful to those implementing a healthy food policy [15].

As part of a scoping review, we identified and evaluated the currently available tools used in public sector workplace policy implementation in New Zealand, Australia, and Canada [17]. Just 2 paper-based tools were related to the New Zealand Policy: guidelines to make better pies [18] and an event and fundraiser guide aimed at schools and community groups [19], and they were likely insufficient to facilitate implementation of the Policy [17]. No other tools were developed to support the New Zealand Policy implementation in contrast to most Australian food environment policies. Nevertheless, the scoping review [17] provided information on the features, usability, and quality of the available tools, which can be used to guide the development of new tools or update of existing tools. For example, for digital tools, which were less common than paper-based tools, logical navigation, clear search and result structure, consistent style, and unambiguous indication of products' compliance in a web-based database were perceived as favorable features [17].

To identify which implementation tools were needed in New Zealand, we interviewed hospital food providers and public health nutrition professionals who were members of the National Food and Drink Environments Network (the Network) supporting implementation of the Policy in their respective DHBs [20]. Apart from customer communication materials, which are in development by Health New Zealand (the Network virtual meetings, 2023), interview participants suggested 2 digital tools that could support the New Zealand Policy implementation: a food and drink audit tool and an up-to-date database showing the traffic light classification of packaged products. Our research team previously developed a digital monitoring tool [21] with an embedded traffic light classification algorithm to systematically conduct the on-site audits of food and drinks available in New Zealand hospitals [14]. The second tool, a web-based product database, was requested because food providers and public health nutrition professionals spent

XSL•FO
RenderX

significant time looking for products on the web and manually assessing individual products against the Policy criteria [20]. Furthermore, there has been no platform for food manufacturers and suppliers to communicate with the hospital food providers about food and drink products that may comply with the Policy [20].

Rationale and Objective

Usability and appearance are likely to impact the usage of web-based tools. To develop the most optimal and feasible tool that meets the needs of its end users, user experience and user interface (UX/UI) design principles and user testing are needed [22]. A user-centered design approach is commonly used in fields that develop and use digital tools such as health care management [23], grocery shopping websites [22], and physical activity apps [24]. However, there is limited research on developing and testing tools designed to support the implementation of healthy food and drink policies.

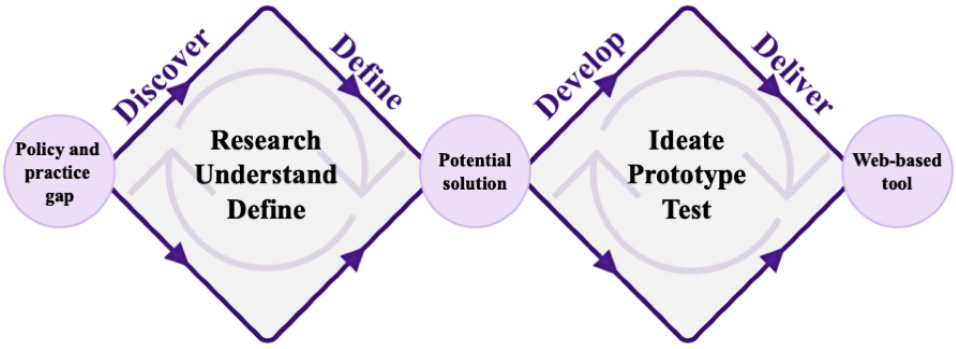
The tool development was a component of the Healthy Policy Evaluation (HYPE) study that comprehensively evaluated the implementation of the National Healthy Food and Drink Policy in New Zealand and determined which tools and resources were needed to support Policy implementation. This paper aims to describe the development and user testing of a new prototype web-based tool to facilitate classification of packaged products according to the National Healthy Food and Drink Policy traffic light criteria.

Methods

Overview

This research was underpinned by pragmatism as the methodological paradigm [25-27]. We focused on practical and carefully considered decisions to develop the most effective and efficient prototype tool to facilitate implementation of the New Zealand Policy by incorporating a user-centered approach that emphasized the needs and preferences of product end users (hospital food providers and Network members [public health nutrition professionals]) [28,29]. The design and development of the new tool were guided by the Double Diamond model (Figure 1) created by the UK Design Council [30]. The model outlines an iterative and flexible process of identifying a real-world problem or gap, defining a possible solution, and finally developing a tangible, valuable, and well-informed outcome addressing the identified problem or gap [30]. The framework is not linear or rigid, and the phases and diamonds can overlap to produce the most optimal tool [30,31] and represent the evolving innovation process and continually changing digital space [30]. The Double Diamond model has been revised over the years by integrating additional steps within the framework, although the principles and the 4 major phases (Discover, Define, Develop, and Deliver) remain largely unchanged [32]. The Double Diamond phases were completed with the assistance of a subcontracted UX/UI designer, and study authors reviewed and provided feedback on all design and development outputs to ensure accurate representation and alignment with end user needs and the aims and scope of the Policy.

Figure 1. The Double Diamond model, illustrating the 4 phases (Discover, Define, Develop, and Deliver) used to guide the development of the Healthy Kai (Food) Checker web-based tool. Arrows pointing toward the tips of the diamonds demonstrate the diverging phases of the process characterized by creative thinking and the expansion and exploration of different ideas and possibilities. Arrows pointing away from the tips of the diamonds reflect the converging phases of the process where progressive decision-making occurs by focusing, prioritizing, and narrowing down the ideas and possibilities toward the most feasible and relevant solution. Circular arrows within the diamonds illustrate the iterative nature of the design process (adapted from the UK Design Council's Double Diamond model [30] under Creative Common License CC BY 4.0 [52]).



The UX/UI design approach could be broadly defined as "the process of designing physical or digital products that are useful, easy to use, and provide a great experience in interacting with them" [28]. Key UX/UI design principles [28,29,34] applied in this study were as follows:

1. User-centered design: designing and developing focused on the needs and preferences of end users (prioritizing accessibility and usability) and collaborating and seeking feedback from key stakeholders throughout the process.
2. Clear information hierarchy: arranging and presenting information in the web-based tool in a logical and organized way to ensure easy and user-friendly navigation and understanding of the content.
3. Visual and interaction consistency: maintaining a consistent and harmonious visual style to achieve a cohesive and professional look and ensuring predictable and consistent performance of the same style elements.
4. Feedback and confirmation: providing instant feedback and confirmation messages on successful user actions and warnings to prevent user errors (eg, deletion).

XSL•FO
RenderX

5. Mobile responsiveness: creating a seamless experience on both desktop and mobile platforms by optimizing the design to adapt to various devices.

Ethical Considerations

Ethical approval for the HYPE study was granted by the Auckland Health Research Ethics Committee (reference AH2519). Subsequent locality approvals were sought from individual DHBs. All usability testing participants provided written consent to participate in the virtual interactive interviews. We approached potential participants up to 3 times by email between May and August 2023, explaining the purpose of the study and providing a detailed participant information sheet explaining the testing procedure. Participants who agreed to take part were sent a consent form, which they signed and returned by email. Participants had the option to have their *whānau* (family) present during the interview and were reimbursed for their time with a New Zealand $50 (US $30) grocery voucher. Study data were deidentified to ensure confidentiality and privacy.

Discover Phase

In this phase, we compiled findings from other components of the HYPE study outlined in the *Introduction* section [14,15,17,20] relevant to designing and developing the new tool. The findings included thematic analysis codes generated during the coding of studies in the systematic review [15] and interview data [20], as well as the features, usability, and quality of the tools assessed in the scoping review [17]. Furthermore, we obtained screenshots of the digital tools included in the scoping review [17] and the websites of the major food suppliers, distributors, and retailers in New Zealand. We collated the findings using the whiteboard function in the web-based Miro software tool (RealtimeBoard, Inc).

Define Phase

In the second phase, we analyzed and refined findings and outputs compiled in the Discover phase to define a feature and function requirement blueprint for the new tool. First, we created a user persona of a fictional end user; defined challenges they might face (pain points); and described their needs, actions they want to take, and expected benefits of using the new tool (a user story) [28,34]. Second, we further reviewed similar available web-based tools [17] in Australia [35,36] and Canada [37] and compared them on their UX/UI aspects of "first impressions," "website visual design" features, and "website interaction" (features, accessibility, user flow, and navigation) [28,29,34].

In the next step, we used the Must have, Should have, Could have, Won't have (MoSCoW) prioritization framework to define features and functions that were deemed "Must have" (essential and critical), "Should have" (important but not critical), "Could have" (desirable but not critical), and "Won't have" (out of scope and excluded in the current design but could be considered in the future) [38]. We considered features identified under the "Must have" and "Should have" as core elements to be incorporated into the new prototype web-based tool.

In the final step, we focused on understanding and carefully planning the path a user would take when interacting with the web-based tool from the initial starting point (eg, home page) to the final step (eg, finding suitable products and checking their classification) [28,34]. The user flow diagram was an important starting point for creating an efficient and easy-to-use web-based tool [39]. The corresponding site map outlined the hierarchical structure of the tool's pages, indicating how pages are linked, grouped, and prioritized in the structural tree [40], and we progressively simplified it to essential pages that aligned with the user flow diagram content.

Develop Phase

We used Figma, a prototyping and design tool, for the remaining Develop and Deliver phases. In the Develop phase, we created a wide range of ideas for the prototype web-based tool underpinned by the outputs from the Define phase and the UX/UI principles outlined earlier. We started with wireframes, a basic arrangement of text, images, and buttons on the web pages before any additional visual elements or content were included [28].

The subsequent design included low-fidelity prototypes containing more details than wireframes, allowing us to test and refine ideas and concepts [28,41]. We engaged with the Network (n=36 members, consisting of public health professionals and dietitians representing their respective DHBs, as well as nutrition and public health advisors from the Ministry of Health and New Zealand Heart Foundation) as key project stakeholders and asked for their feedback and input on the low-fidelity prototypes before committing to a specific design. The Network was provided with a document via email containing different design options alongside a small set of questions about their preferred features, structure, layout, and content (Multimedia Appendix 1). Feedback was collated during the subsequent monthly Network virtual meeting (11 members were present) and used to refine the prototype further.

Deliver Phase

Overview

In the Deliver phase, we developed a comprehensive high-fidelity prototype with interactive and functional elements, such as clickable content, buttons, and menus [28,41]. The tool's pages were linked, mirroring the navigational flow of a live website and providing a realistic user experience, which allowed us to test user interaction with the prototype and refine the user journey [28,41]. We tested the prototype multiple times and once with a colleague to further improve and refine the high-fidelity prototype before testing it with end users.

Interactive Interviews to Test Usability of the Prototype Web-Based Tool

We conducted usability testing that involved interactive virtual interviews in which user interaction and engagement with the prototype were evaluated to determine user preferences and behaviors related to the prototype tool [42]. The testing also identified UX/UI challenges and pain points that needed addressing or redesigning to improve the tool [42].

XSL•FO
RenderX

Participant Selection and Recruitment

Potential participants (end users) were food providers (for staff and visitors) and members of the Network (public health nutrition professionals) within any of the New Zealand hospitals who had previously participated in the stakeholder interviews (n=12) [20] and who expressed willingness to test the new web-based tool, as well as other key stakeholders who had previously used the Policy and were familiar with its traffic light classification system.

Data Collection

The usability test interview combined 2 methods: performance testing [43], in which participants were given tasks to complete while interacting with the prototype, and visual design testing [44], in which participants were asked to evaluate visual aspects of the prototype tool. The usability interview guide (Multimedia Appendix 2) included questions about the prototype tool's engagement, intuitive usage, functionality, navigation, quality, and user-friendliness, as well as scenarios and tasks related to the tool's functions.

At the beginning of the virtual interviews, we explained the research purpose (testing the usability and user experience of the design product and not the participants' ability to use the prototype), process, and participants' rights; answered any questions; and verbally confirmed the permission to video record the interviews (as previously agreed to on the consent form). We provided participants with a digital link to access the Figma prototype and asked them to share their screen to demonstrate their interaction with the prototype's features and functions. Participants were encouraged to think aloud during testing and to comment on the design, their preferences, and how the tool could be improved. All interviews were transcribed verbatim, and participants were given the opportunity to review and edit their transcripts.

Data Analysis

One author watched interview recordings multiple times and reviewed transcripts to analyze how participants interacted with the tool; how intuitive, engaging, efficient, and accessible the prototype tool was; and what changes were needed to improve the user experience. Participants' suggestions, comments, and interactions were recorded on the printouts of the prototype. The results were shared and discussed with the UX/UI designer, who incorporated the feedback to improve and refine the design of the prototype. Textbox 1 summarizes the methods used and outputs in the phases of the Double Diamond model.

Textbox 1. Phases and outputs of the Double Diamond model used in the design and development of the new prototype Healthy Kai (Food) Checker web-based tool to support implementation of the National Healthy Food and Drink Policy in New Zealand hospitals.

Discover—divergent research phase **(understanding user needs and challenges they face, exploring similar tools available)**

- A systematic review [15] and scoping review findings [17] about tools and resources used internationally in implementation of healthy food and drink policies.
- Healthy Policy Evaluation (HYPE) study interview findings [20] specific to the characteristics and needs of the target audience (food providers and Network members [public health nutrition professionals]).
- HYPE study food and drinks audit findings [14], including types of products available.
- Layout and structure characteristics of existing digital tools [17] and major retailer and supplier websites.

Define—convergent synthesis phase **(analyzing and refining insights generated in the Discover phase)**

- User persona of a fictional food provider, including their pain points and user story.
- Review and comparison of the user experience and user interface features of similar tools [35-37].
- Requirements blueprint for the new tool (refined by considering project timelines and the available budget).
- New tool features prioritized with the Must have, Should have, Could have, Won't have (MoSCoW) framework [38].
- Visual user flow diagram (outlining actions, decisions, and steps users take through the tool).
- Structural site map of the pages in the tool.

Develop—divergent ideation phase **(creating basic outlines of the prototype interface)**

- Wireframes, a basic skeletal outline of the website interface and elements.
- Low-fidelity prototype (a simple, monotone, and static representation of a prototype).
- Informal feedback from project stakeholders on the low-fidelity prototype.

Deliver—convergent implementation phase **(refining of the prototype and usability testing)**

- High-fidelity prototype (detailed, interactive, and realistic representation closely resembling the final web-based tool).
- Interactive usability testing with end users.

XSL•FO
RenderX

Results

Define Phase

The user persona of a fictional food provider, including their characteristics, challenges, and user story, is summarized in Figure 2. Based on the user persona, a new web-based tool, Healthy Kai (Food) Checker—a searchable database of packaged food and drink products showing their traffic light classification—was identified as a key tool to support the New Zealand Policy implementation. Textbox 2 summarizes positive UX/UI attributes [17] we identified based on the review of similar available tools from Australia, Healthy Food Finder [36] in New South Wales, and FoodChecker [35] in Victoria, as well as a Canadian tool, Brand Name Food List [37], available in British Columbia. The function and feature requirements of the Healthy Kai Checker prioritized with the MoSCoW framework as "Must have" and "Should have" are indicated in Textbox 3.

The Healthy Kai Checker user flow (Figure 3) and functionality design included elements closely related to the "Must have" and "Should have" features outlined in Textbox 3 [38], for example, registration to create a free account, personalization to apply search filters and create lists of favorite products, educational content to explain terms and provide tips on how to use the tool, and a feedback mechanism to report an issue or suggest improvements. The corresponding Healthy Kai Checker site map is available in Multimedia Appendix 3.

Figure 2. User persona, pain points, and user story of a fictional hospital food provider in New Zealand. The figure illustrates the characteristics, challenges, and narrative of a typical hospital food provider, highlighting their needs and experiences in implementing the National Healthy Food and Drink Policy and expected benefits of using the new tool.

XSL•FO
RenderX

Textbox 2. Favorable user experience and user interface features of web-based tools with product databases based on the review and comparison of similar existing web-based tools [17] in Australia [35,36] and Canada [37].

First impressions

- Visually appealing (clear, fresh, and crisp look) and clean design
- Good use of white space for a less compact layout
- Colorful
- Modern feel

Website visual design

- Consistent visual style
- Consistent and clear branding
- Clear font
- Simple design
- Good color scheme and use of images
- A good indication of clickable elements

Website interaction

- Automatic suggestion of products in the search bar (predictive text)
- A clear indication of product compliance/traffic light classification
- Results displayed based on relevance to the search term (eg, for misspelled words)
- Product information in separate columns (name, brand, and size)
- Good column sorting function (both directions)
- Relevant search filters
- Explanation of why the product does not meet the Policy criteria
- Option to report missing products
- Ability to create own product lists
- Ability to return to previous screens without loss of content
- Confirmation of completed action
- Intuitive navigation

XSL•FO
RenderX

Textbox 3. Features and functions of the Healthy Kai (Food) Checker web-based tool prioritized with the Must have, Should have, Could have, Won't have (MoSCoW) framework [38]. "Must have" features and functions were considered essential and critical, "Should have" were important but not critical, "Could have" were desirable but not critical, and "Won't have" were out of scope and excluded in the current design but could be considered in the future. "Must have" and "Should have" were core elements to be incorporated into the new prototype web-based tool to support implementation of the National Healthy Food and Drink Policy in New Zealand hospitals.

Must have

- A searchable list of packaged products (including *Red* products so that end users can quickly identify items that do not comply with the Policy instead of spending time classifying them manually).
- Ability to search by name, keyword, and brand.
- Basic and advanced search options (traffic light criteria and food group filters).
- A clear indication of the traffic light classification (text and color).
- Up-to-date product database (a database of products was created during the food and drink audits conducted as part of the HYPE study [14] that can be integrated).
- Engaging information about the Policy and Healthy Kai Checker.

Should have

- Product information separated into columns (name, size, and brand).
- Sort results function (alphabetically, product size, and traffic light classification).
- Ability to report a missing or incorrect products.
- Free registration requirement to view all products and unlock full access.
- Free registration requirement to create, save, export, and print product lists.
- No paid registration (free for all users).
- Mobile responsiveness for different devices.

Could have

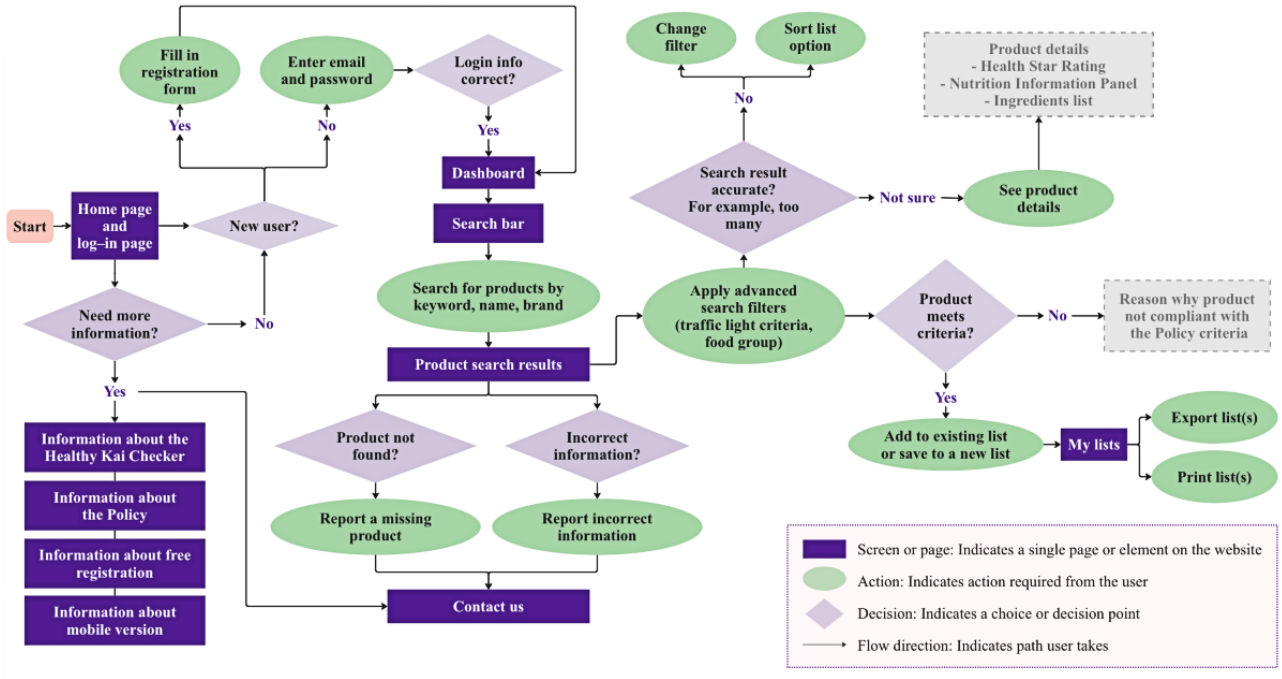
- Detailed information about each product (eg, nutritional, ingredient list, and allergens)
- Clear product photographs.
- Filters related to allergens, suitability for vending machines.
- Explanation or indication of why a product does not meet the *Green* criteria.

Won't have

- Self-assessment of products and recipes against the Policy criteria.
- Integration of healthy food and drink policies for New Zealand schools and early childhood education centers, or government-funded healthy school lunches program.
- Web-based Policy-relevant training modules.

XSL•FO
RenderX

Figure 3. User flow diagram of the Healthy Kai (Food) Checker web-based tool. This diagram illustrates the user journey through the tool, detailing steps from initial log-in to final product selection. The user flow is designed to support users in implementing the National Healthy Food and Drink Policy in New Zealand hospitals. The green ovals show actions taken by the users, while the light purple diamonds indicate points where users choose their next action. Grayed-out rectangles show features classified as "Could have" or "Won't have" according to the Must have, Should have, Could have, Won't have (MoSCoW) prioritization framework, reflecting their lower priority and exclusion from the initial prototype tool.

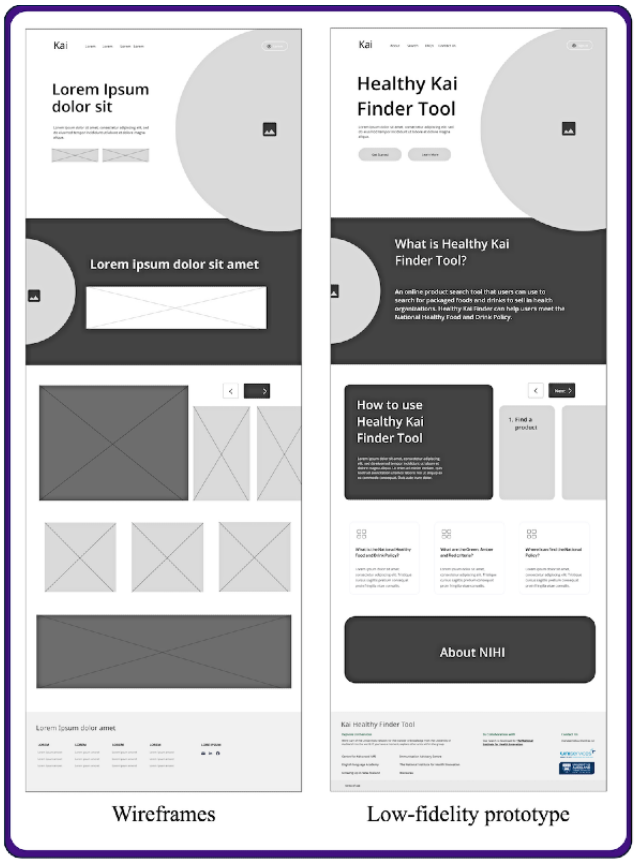


Develop Phase

The initial wireframe outline and the low-fidelity prototype for the home page are shown in Figure 4. The low-fidelity prototypes of the 2 design options for the product search and result dashboard are available in Multimedia Appendix 1. Feedback from the 11 Network members and the research team favored the option where the search box and result filters are at the top rather than on the side of the search page.

XSL•FO
RenderX

Figure 4. Wireframes and low-fidelity prototype of the Healthy Kai (Food) Checker web-based tool home page. This figure represents the initial design concepts for the tool, outlining the layout and structure of the home page and showcasing key elements of the interface design and basic functionality. The feedback from 11 public health nutrition professionals on the low-fidelity prototype was used to refine the web-based tool to support implementation of the National Healthy Food and Drink Policy in New Zealand hospitals.



Deliver Phase (Usability Testing)

Twelve individuals were invited to take part in the interactive interviews, of whom 1 declined due to lack of time, 2 were no longer employed in the same role, and 6 did not respond after 3 invitation emails. The final interview sample included 2 Network members (public health nutrition professionals) and 1 stakeholder who used the Policy in public health research on food environments. The interviews took an average of 49 (range 40-59) minutes to complete.

Participants generally found the Healthy Kai Checker prototype easy to use and logical to navigate. They appreciated the hyperlinks in the prototype that directed users to external web pages, such as the Ministry of Health website where the Policy was published. There was also an appreciation that the different pages within the tool could be reached through multiple buttons or links (ie, having multiple paths to 1 page), and participants provided additional suggestions on where specific buttons connecting to the various web pages of the tool or external links should be incorporated. All participants liked the color scheme and the purple and green palette, which made the tool look vibrant and fresh and differentiated it from other public health or government websites that typically use more muted colors. One suggestion was to ensure that images were representative of the diverse New Zealand population. The Healthy Kai Checker was positively perceived due to its simple and concise interface, which did not overwhelm users with too much information.

On the home page, participants suggested including more information about the free website registration and its benefits and placing that component closer to the top of the web page. The section informing users about the ability to use the tool on mobile or tablet devices was positively perceived and thought to be necessary in the current technology-driven working environments. However, a feature in the middle of the home page, “How to use the Healthy Kai Product Database” that contained a clickable carousel of moving parts from right to left, was deemed confusing. Participants suggested reducing the steps showing information on how to use the tool from 4 to 3 and displaying all of them at once on the main page, thus improving usability by eliminating the need to click on the various components to see the entire content.

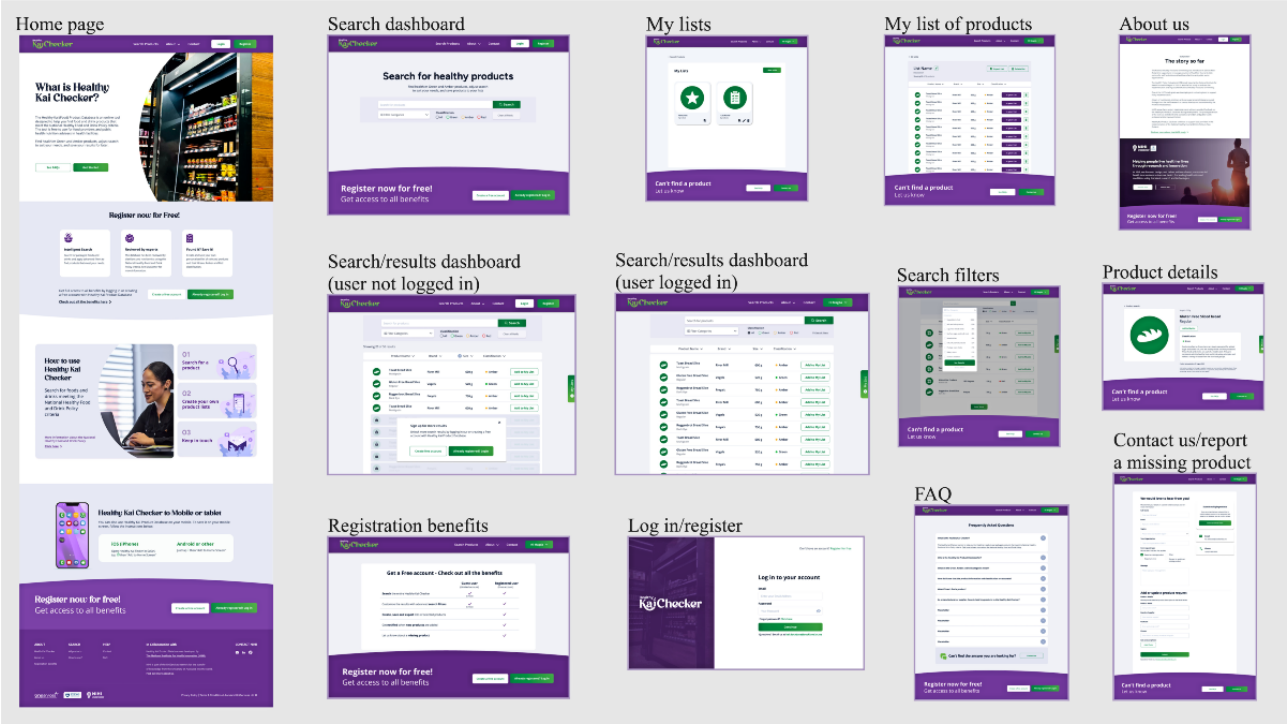
On the product search and results page and the pages showing products saved in the user’s individual lists, the indication of products’ categorization as *Green*, *Amber,* or *Red* was perceived as unambiguous. The favorable features were the ability to apply advanced search filters (traffic light classification and the Policy’s food group) to display only those products with the specified characteristics. Three suggestions for improvements were incorporating an indication of the number of products in the search results and personalized product lists, making the “My Lists” feature more prominent, and making some of the buttons and writing bigger to improve visibility. Participants also thought that allowing manufacturers and suppliers to submit their products for classification and inclusion in the database,

XSL•FO
RenderX

along with product photographs, was a good idea, and having a "Contact Us" page on the website that clearly outlined the steps required would facilitate this.

For the remaining pages, participants suggested some frequently asked questions that users would likely be interested in, since the Network members had experience working closely with food providers to implement the Policy. Furthermore, minor aesthetic adjustments, such as reducing plain text and adding graphical elements or photographs, were suggested to make the pages more visually appealing. Screenshots of the most recent version of the Healthy Kai Checker prototype tool are shown in Figure 5.

Figure 5. Screenshots of the various pages of the Healthy Kai (Food) Checker prototype web-based tool designed to support implementation of the National Healthy Food and Drink Policy in New Zealand hospitals. The tool is a searchable database of packaged food and drink products that classifies items according to the Policy's nutritional criteria, with functions such as basic and advanced searches for products, sorting list options, the ability to compile a list of selected products, a means to report products missing from the database, and responsiveness for usage on different devices. FAQ: frequently asked questions.



Discussion

Principal Findings

We aimed to develop and test a new evidence-based, web-based prototype tool to support the adoption and implementation of the National Healthy Food and Drink Policy in New Zealand. The new Healthy Kai Checker web-based tool will reduce the burden on individual food providers and Network members associated with identifying and classifying products according to the Policy's traffic light nutrition criteria. Until now, individual food providers and organizations had to undertake this essential implementation task manually, which is a repetitive and resource-intensive process.

The choice of a web-based tool in this study was driven by the increasingly digital or web-based practices and systems [45] that already saturate the grocery and retail market [22] and are increasingly used for individual behavioral change approaches [46] and research on food supply [47] and physical activity [24]. The scarcity of web-based or digital tools in New Zealand [17] for quickly checking the classification of packaged food and drink products represents an opportunity for the new Healthy Kai Checker to fill this significant gap and facilitate implementation of the Policy. Healthy Kai Checker was designed as an intuitive and logical web-based tool suitable for end users who may have limited nutrition knowledge and experience, and understanding of technical nutrition jargon, which can be overwhelming and complex [48-50]. Furthermore, the aim was to optimize user paths, accommodate varying levels of technological proficiency, and minimize the steps end users had to take [39] to search and identify suitable products.

The design and development of the Healthy Kai Checker followed the rigorous but flexible Double Diamond model [30], was guided by key UX/UI design principles [28,29,34], and was based on the comprehensive research our team conducted as part of the HYPE study [14,15,17,20]. While other design and development strategies could have resulted in a different or better tool, we are unable to make a direct comparison since the design and development process of similar tools in Australia [35,36] and Canada [37] had not been outlined in the literature. However, our approach provided key advantages of emphasizing and understanding user needs and stakeholder engagement, mitigating potential risks and challenges early in the process, and allowing iteration and creativity, leading to a unique, tailored, and effective solution to the identified resource gap [30,31].

XSL•FO
RenderX

Implementation Considerations

To ensure a seamless user experience for all intended end users, Healthy Kai Checker would need to be further tested with food providers in New Zealand hospitals. Furthermore, expanding the current features to include other healthy food policies in New Zealand, such as the government-funded healthy school lunches program policy (Ka Ora, Ka Ako) [51], would require user testing with food providers working in school environments. However, it should be noted that usability testing and the incorporation of end user feedback are an ongoing process in the evolving digital space and a critical component of the product's road map [30]. After Healthy Kai Checker is developed and released as a fully functioning web-based tool, it would need to be periodically updated to ensure that it maintains user-centered design and continues to offer advantages over other potential tools in the market.

Another key aspect of the Healthy Kai Checker is the ongoing management and maintenance of the food and drink product database that underpins the tool, as well as new updates of the algorithm embedded into the tool to facilitate the classification of products as *Green*, *Amber*, or *Red* [21]. The current database contains information on food and drinks collected during food availability audits in New Zealand hospitals as part of the HYPE study [14]. However, these data were captured in 2021/2022, and it is likely that, at some point, they will no longer accurately represent the supply on the market. Further manual field data collection, as used for the HYPE study, is likely too resource-intensive and costly to do regularly. Reaching out to the major New Zealand suppliers to encourage them to provide nutritional information regularly on their products could be a cost-effective strategy to streamline data collection and maintain the accuracy and currency of the dataset.

Another avenue includes obtaining product information directly from Healthy Kai Checker end users, as seen with crowdsourcing of supermarket data from the FoodSwitch [47] app users in Australia. The Healthy Kai Checker includes a feature where users can directly provide the necessary information and photographs about a missing or incorrect product for evaluation. An additional emerging method of data collection is the "web harvesting" or "data mining" approach used to automatically extract information from websites, such as web-based supermarket stores [52], and could also be explored to source data for the Healthy Kai Checker.

Regardless of the source of product data, classifying individual products into *Green*, *Amber*, and *Red* categories using the Policy's criteria can be complex because some of the required information is not provided on the food packaging. The challenging classification of some products with the Policy criteria was observed during food and drink availability audits as part of the HYPE study [14], although a robust algorithm was developed and included in the web-based audit tool to make the process systematic [21]. Simplifying the Policy's criteria could increase food provider and supplier understanding of the guidelines and streamline data collection to keep the product database up-to-date. Importantly, a classification algorithm would need to be revised if the Policy undergoes an update of its nutritional criteria, representing an additional cost associated with ongoing maintenance of the Healthy Kai Checker.

Strengths and Limitations

There are several strengths in this study. First, the design and development of the Healthy Kai Checker were informed by comprehensive research findings and aligned with the needs of the Policy implementers. Second, the UX/UI designer used the latest website design techniques and best practices to ensure the user-friendliness of the tool. Third, feedback and suggestions provided by stakeholders and participants were incorporated into the tool to improve its usability. However, there were some limitations. Although the Healthy Kai Checker prototype has been tested with 1 group of potential end users, testing has not yet been conducted with the New Zealand food providers because they did not respond to the invitation to take part in the research. Future research should explore the usability of the prototype tool with food providers and stakeholders associated with other policies or settings. User testing should also be conducted once the Healthy Kai Checker is a fully functioning website. Since the tool is still a prototype, this study did not explore the effectiveness and impact of the Healthy Kai Checker as a tool in the Policy implementation. Research on the effectiveness of tools and resources to support healthy food policy implementation is generally lacking, and future studies should focus on evaluating the effectiveness of tools in this space.

Conclusions

This study comprehensively designed and developed a new web-based Healthy Kai Checker tool, a searchable database of packaged food and drink products, to facilitate the implementation of the National Healthy Food and Drink Policy in New Zealand. The development of the tool was guided by the Double Diamond model and UX/UI design principles. Feedback received from project stakeholders was incorporated into the prototype, and comprehensive user-testing results further improved the design and usability of the prototype tool. The Healthy Kai Checker could provide valuable support to New Zealand hospital food providers and others who support implementation of the Policy.

Acknowledgments

The authors would like to thank and acknowledge Alexandra Rengifo, user experience and user interface (UX/UI) designer, for her help in conceptualizing the tool and developing the user-testing guide and the prototype tool, and thank the participants for taking part in the study. This research was funded by the Healthier Lives National Science Challenge as part of the Healthy Policy Evaluation (HYPE) study.

XSL•FO
RenderX

Data Availability

The interview data analyzed in this study are not publicly available due to presence of sensitive and personally identifiable qualitative data that could compromise the anonymity of the participants. The remaining datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

MR, CNM, and SM conceptualized the study. MR recruited participants and collected, managed, and analyzed the data. MR and EU conceptualized the tool and developed user-testing guide. MR wrote the first draft and is responsible for the paper's final content. All authors contributed to the interpretation of results, critically revised the paper, and approved the final version.

Conflicts of Interest

MR and CNM were members of the National DHB Food & Drink Environments Network. The Network had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. SM and EU declare no conflict of interest.

Multimedia Appendix 1
Low-fidelity prototype feedback document.
[PDF File (Adobe PDF File), 577 KB - formative_v9i1e60447_app1.pdf]

Multimedia Appendix 2
Usability testing of the Healthy Kai Checker interactive interview guide.
[PDF File (Adobe PDF File), 186 KB - formative_v9i1e60447_app2.pdf]

Multimedia Appendix 3
Healthy Kai Checker site map.
[PDF File (Adobe PDF File), 190 KB - formative_v9i1e60447_app3.pdf]

References

1. Imamura F, Micha R, Khatibzadeh S, Fahimi S, Shi P, Powles J, et al. Dietary quality among men and women in 187 countries in 1990 and 2010: a systematic assessment. Lancet Glob Health 2015;3(3):e132-e142 [FREE Full text] [doi: 10.1016/S2214-109X(14)70381-X] [Medline: 25701991]
2. Miller V, Webb P, Cudhea F, Shi P, Zhang J, Reedy J, et al. Global dietary quality in 185 countries from 1990 to 2018 show wide differences by nation, age, education, and urbanicity. Nat Food 2022;3(9):694-702. [doi: 10.1038/s43016-022-00594-9] [Medline: 37118151]
3. Healthy food environments. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nutrition/healthy-food-environments/ [accessed 2024-12-16]
4. Influencing food environments for healthy diets. Food and Agriculture Organization (FAO). 2016. URL: https://www.fao.org/fsnforum/resources/reports-and-briefs/influencing-food-environments-healthy-diets [accessed 2023-10-15]
5. Mackay S, Garton K, Gerritsen S, Sing F, Swinburn B. How Healthy Are Aotearoa New Zealand's Food Environments? Assessing the Impact of Recent Food Policies 2018–2021. Auckland, New Zealand: University of Auckland; 2021. URL: https://tinyurl.com/v2hysyht
6. Chu C, Breucker G, Harris N, Stitzel A, Gan X, Gu X, et al. Health-promoting workplaces—international settings development. Health Promot Int 2000;15(2):155-167. [doi: 10.1093/heapro/15.2.155]
7. Chu C, Driscoll T, Dwyer S. The health-promoting workplace: an integrative perspective. Aust N Z J Public Health 1997;21(4 Spec No):377-385 [FREE Full text] [doi: 10.1111/j.1467-842x.1997.tb01718.x] [Medline: 9308202]
8. Labour market statistics March 2023 quarter summary. Stats New Zealand. 2023. URL: https://tinyurl.com/54bzedhu [accessed 2023-10-15]
9. 2018 Census place summaries. Stats New Zealand. 2018. URL: https://www.stats.govt.nz/tools/2018-census-place-summaries/new-zealand#health [accessed 2023-09-08]
10. Changing the system. Te Whatu Ora—Health New Zealand. URL: https://www.tewhatuora.govt.nz/whats-happening/changing-the-system/ [accessed 2023-09-21]
11. Te Whatu Ora—Health New Zealand. Health workforce plan 2023/24. 2023. URL: https://www.tewhatuora.govt.nz/publications/health-workforce-plan-202324/ [accessed 2023-08-28]
12. National Healthy Food and Drink Policy (2nd ed). Ministry of Health. National District Health Board Food and Drink Environments Network. 2019. URL: https://www.health.govt.nz/publications/national-healthy-food-and-drink-policy [accessed 2020-10-10]

XSL•FO
RenderX

13. Gerritsen S, Kidd B, Rosin M, Shen S, Mackay S, Te Morenga L, et al. 2021 assessment of New Zealand district health boards' institutional healthy food and drink policies: the HealthY Policy Evaluation (HYPE) study. N Z Med J 2022;135(1560):67-76. [doi: 10.26635/6965.5698] [Medline: 35999800]
14. Ni Mhurchu C, Shen S, Kidd B, Jiang Y, Umali E, Gerritsen S, et al. Food and drink availability in New Zealand hospitals 2021/22: alignment with a voluntary National Healthy Food and Drink Policy. 2023 Presented at: ISBNPA 2023 Abstract Booklet. International Society of Behavioral Nutrition and Physical Activity; 2023 Jun 14-17; Uppsala, Sweden.
15. Rosin M, Mackay S, Gerritsen S, Te Morenga L, Terry G, Ni Mhurchu C. Barriers and facilitators to implementation of healthy food and drink policies in public sector workplaces: a systematic literature review. Nutr Rev 2024;82(4):503-535 [FREE Full text] [doi: 10.1093/nutrit/nuad062] [Medline: 37335857]
16. World Health Organization. Action Framework for Developing and Implementing Public Food Procurement and Service Policies for a Healthy Diet. Geneva, Switzerland: World Health Organization; 2021. URL: https://www.who.int/publications/i/item/9789240018341
17. Rosin M, Mackay S, Ni Mhurchu C. Tools and resources used to support implementation of workplace healthy food and drink policies: a scoping review of grey literature. Nutr Diet 2023;80(5):452-462. [doi: 10.1111/1747-0080.12844] [Medline: 37743098]
18. Appendix 3: how to make better pies guidelines (in the National Healthy Food and Drink Policy). The Pie Group. 2019. URL: https://www.health.govt.nz/publications/national-healthy-food-and-drink-policy [accessed 2020-10-10]
19. Healthy events and fundraiser guide. Canterbury District Health Board. URL: https://www.cph.co.nz/your-health/love-kai/ [accessed 2023-05-17]
20. Rosin M, Ni Mhurchu C, Mackay S. Implementing healthy food policies in health sector settings: New Zealand stakeholder perspectives. BMC Nutr 2024;10(1):119. [doi: 10.1186/s40795-024-00924-z] [Medline: 39244614]
21. Shen S, Kidd B, Mackay S, Te Morenga L, Gerritsen S, Jiang Y, et al. Designing an audit tool to evaluate the national healthy food and drink policy: the HYPE study. 2022 Presented at: Nutrition Society of New Zealand Annual Conference; 2021 December 2-3; New Zealand p. 27. [doi: 10.3390/msf2022009027]
22. Merritt K, Zhao S. An innovative reflection based on critically applying UX design principles. Journal Open Innov Technol Market Complexity 2021;7(2):129. [doi: 10.3390/joitmc7020129]
23. Stola K. User experience and design thinking as a global trend in healthcare. J Med Sci 2018;87(1):28-33. [doi: 10.20883/jms.2018.281]
24. Sanders JP, Gokal K, Thomas JJC, Rawstorn JC, Sherar LB, Maddison R, et al. Development of a mobile health snacktivity app to promote physical activity in inactive adults (SnackApp): intervention mapping and user testing study. JMIR Form Res 2023;7:e41114 [FREE Full text] [doi: 10.2196/41114] [Medline: 37213198]
25. Morgan DL. Integrating Qualitative and Quantitative Methods: A Pragmatic Approach. Thousand Oaks, CA: SAGE Publications, Inc; 2014.
26. Creswell JW, Poth CN. Qualitative Inquiry and Research Design: Choosing Among Five Approaches. Thousands Oaks, CA: SAGE; 2018.
27. Kaushik V, Walsh CA. Pragmatism as a research paradigm and its implications for social work research. Soc Sci 2019;8(9):255. [doi: 10.3390/socsci8090255]
28. Canziba E. Hands-on UX Design for Developers: Design, Prototype, and Implement Compelling User Experiences From Scratch. Birmingham, United Kingdom: Packt Publishing; 2018.
29. What are design principles? Interaction Design Foundation. 2016. URL: https://www.interaction-design.org/literature/topics/design-principles [accessed 2023-10-11]
30. The Double Diamond. Design Council. URL: https://www.designcouncil.org.uk [accessed 2023-08-29]
31. Design in the public sector? Design Council. URL: https://tinyurl.com/3x5jmntn [accessed 2023-10-11]
32. Nessler D. Medium. 2018. URL: https://tinyurl.com/aawvyx6x [accessed 2023-10-11]
33. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2024-12-20]
34. Rosenzweig E. Successful User Experience: Strategies and Roadmaps. Edinburgh, London: Elsevier Science; 2015.
35. Healthy eating advisory service—FoodChecker. State of Victoria. 2016. URL: https://foodchecker.au/ [accessed 2023-07-25]
36. Healthy food finder. New South Wales Ministry of Health. URL: https://www.foodfinder.health.nsw.gov.au/ [accessed 2022-02-14]
37. Province of British Columbia, HealthLinkBC. Brand name food list. URL: https://bnfl.healthlinkbc.ca/ [accessed 2022-02-11]
38. Chapter 10: MoSCoW prioritisation framework (The DSDM Agile Project Framework 2014 onwards). Agile Business Consortium. 2014. URL: https://www.agilebusiness.org/dsdm-project-framework/moscow-prioririsation.html [accessed 2023-08-29]
39. Frey C. Why user flows are pivotal and how to properly outline them. Medium. 2022. URL: https://tinyurl.com/3b238xwz [accessed 2023-10-12]
40. Jhang J. How to create a UX sitemap: a simple guideline. Medium. 2020. URL: https://tinyurl.com/ycx7e4ep [accessed 2023-10-12]

XSL•FO
RenderX

41. Kasturika K, Skrok D, Briggs C. 14 UX deliverables: what will I be making as a UX designer? Interaction Design Foundation. 2020. URL: https://www.interaction-design.org/literature/article/7-ux-deliverables-what-will-i-be-making-as-a-ux-designer [accessed 2023-10-12]
42. Moran K. Usability (user) testing 101. NN/g Nielsen Norman Group. 2019. URL: https://tinyurl.com/2rfr83et [accessed 2023-10-13]
43. Schroeter E. Usability testing: everything you need to know (methods, tools, and examples). Interaction Design Foundation. 2021. URL: https://careerfoundry.com/en/blog/ux-design/usability-testing-guide/#performance-testing [accessed 2023-10-13]
44. Whitenton K. How to test visual design. NN/g Nielsen Norman Group. 2018. URL: https://tinyurl.com/yr8fs8fn [accessed 2023-10-13]
45. Xu M, David JM, Kim SH. The fourth industrial revolution: opportunities and challenges. IJFR 2018;9(2):90-95. [doi: 10.5430/ijfr.v9n2p90]
46. Schoeppe S, Alley S, Rebar AL, Hayman M, Bray NA, van Lippevelde W, et al. Apps to improve diet, physical activity and sedentary behaviour in children and adolescents: a review of quality, features and behaviour change techniques. Int J Behav Nutr Phys Act 2017;14(1):83 [FREE Full text] [doi: 10.1186/s12966-017-0538-3] [Medline: 28646889]
47. Dunford E, Trevena H, Goodsell C, Ng KH, Webster J, Millis A, et al. FoodSwitch: a mobile phone app to enable consumers to make healthier food choices and crowdsourcing of national food composition data. JMIR Mhealth Uhealth 2014;2(3):e37 [FREE Full text] [doi: 10.2196/mhealth.3230] [Medline: 25147135]
48. Mackey MA, Metz M. Ease of reading of mandatory information on Canadian food product labels. Int J Consumer Studies 2009;33(4):369-381. [doi: 10.1111/j.1470-6431.2009.00787.x]
49. Dodson EA, Eyler AA, Chalifour S, Wintrode CG. A review of obesity-themed policy briefs. Am J Prev Med 2012;43(3 Suppl 2):S143-S148 [FREE Full text] [doi: 10.1016/j.amepre.2012.05.021] [Medline: 22898164]
50. Kastner M, Bhattacharyya O, Hayden L, Makarski J, Estey E, Durocher L, et al. Guideline uptake is influenced by six implementability domains for creating and communicating guidelines: a realist review. J Clin Epidemiol 2015;68(5):498-509 [FREE Full text] [doi: 10.1016/j.jclinepi.2014.12.013] [Medline: 25684154]
51. Nutrition Standards for Ka Ora, Ka Ako | Healthy school lunches programme. New Zealand Ministry of Education. 2022. URL: https://kaorakaako.education.govt.nz/file-download/Nutrition%20Standards%202022.pdf [accessed 2023-02-13]
52. Ahmed M, Schermel A, Lee J, Weippert M, Franco-Arellano B, L'Abbé M. Development of the food label information program: a comprehensive Canadian branded food composition database. Front Nutr 2021;8:825050 [FREE Full text] [doi: 10.3389/fnut.2021.825050] [Medline: 35187026]

Abbreviations

DHB: district health board
HYPE: Healthy Policy Evaluation
MoSCoW: Must have, Should have, Could have, Won't have
UX/UI: user experience and user interface

Edited by A Mavragani; submitted 11.05.24; peer-reviewed by L Cranney; comments to author 10.09.24; revised version received 15.09.24; accepted 15.11.24; published 13.01.25.

Please cite as:
Rosin M, Ni Mhurchu C, Umali E, Mackay S
Healthy Kai (Food) Checker Web-Based Tool to Support Healthy Food Policy Implementation: Development and Usability Study
JMIR Form Res 2025;9:e60447
URL: https://formative.jmir.org/2025/1/e60447
doi:10.2196/60447
PMID:

©Magda Rosin, Cliona Ni Mhurchu, Elaine Umali, Sally Mackay. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring the General Acceptability and User Experience of a Digital Therapeutic for Cognitive Training in a Singaporean Older Adult Population: Qualitative Study

Siong Peng Kwek[1,2,3], BSocSci; Qiao Ying Leong[1,2], MPH; V Vien Lee[1,2], PhD; Ni Yin Lau[1,2], BSocSci; Smrithi Vijayakumar[1,2], PhD; Wei Ying Ng[1,2], BSocSci; Bina Rai[2,3], PhD; Marlena Natalia Raczkowska[1,2], PhD; Christopher L Asplund[1,2,3,4,5*], PhD; Alexandria Remus[1,2,3,6*], PhD; Dean Ho[1,2,3,7,8,9*], PhD

[1]The Institute for Digital Medicine (WisDM), Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[2]The N.1 Institute for Health, National University of Singapore, Singapore, Singapore

[3]Department of Biomedical Engineering, College of Design and Engineering, National University of Singapore, Singapore, Singapore

[4]Division of Social Sciences, Yale-NUS College, National University of Singapore, Singapore, Singapore

[5]Department of Psychology, College of Humanities and Sciences, National University of Singapore, Singapore, Singapore

[6]Yong Loo Lin School of Medicine, Heat Resilience & Performance Centre, National University of Singapore, Singapore, Singapore

[7]Department of Pharmacology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[8]The Bia-Echo Asia Centre for Reproductive Longevity and Equality, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[9]Singapore's Health District @ Queenstown, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[*]these authors contributed equally

Corresponding Author:
Alexandria Remus, PhD
Yong Loo Lin School of Medicine
Heat Resilience & Performance Centre
National University of Singapore
27 Medical Drive, #03-01
Singapore, 117510
Singapore
Phone: 65 66017766
Email: alexandria.remus@nus.edu.sg

Abstract

Background: Singapore’s large aging population poses significant challenges for the health care system in managing cognitive decline, underscoring the importance of identifying and implementing effective interventions. Cognitive training delivered remotely as a digital therapeutic (DTx) may serve as a scalable and accessible approach to overcoming these challenges. While previous studies indicate the potential of cognitive training as a promising solution for managing cognitive decline, understanding the attitudes and experiences of older adults toward using such DTx platforms remains relatively unexplored.

Objective: This study aimed to characterize the general acceptability and user experience of CURATE.DTx, a multitasking-based DTx platform that challenges the cognitive domains of attention, problem-solving, and executive function in the Singaporean older adult population.

Methods: A total of 15 older adult participants (mean age 66.1, SD 3.5 years) were recruited for a 90-minute in-person session. This session included a 30-minute playtest of CURATE.DTx, followed by a 60-minute semistructured interview to understand their overall attitudes, experience, motivation, and views of the intervention. Interviews were audio-recorded and transcribed verbatim, then analyzed using an inductive approach. Thematic analysis was used to identify emerging patterns and insights.

Results: A total of 3 main themes, and their respective subthemes, emerged from the interviews: comprehension, with subthemes of instruction and task comprehension; acceptability, with subthemes of tablet usability, engagement and enjoyment, and attitude and perceived benefits; and facilitators to adoption, with subthemes of framing and aesthetics, motivation recommendations and the role of medical professionals. Our findings revealed that participants encountered some challenges with understanding certain elements of CURATE.DTx. Nevertheless, they were still highly engaged with it, finding the challenge to be enjoyable. Participants

XSL•FO
RenderX

also showed a strong awareness of the importance of cognitive training and expressed a keen interest in using CURATE.DTx for this purpose, especially if recommended by medical professionals.

Conclusions: Given the positive engagement and feedback obtained from Singaporean older adults on CURATE.DTx, this study can serve as a basis for future platform iterations and strategies that should be considered during implementation. Future studies should continue implementing an iterative codesign approach to ensure the broader applicability and effectiveness of interventions tailored to this demographic.

(JMIR Form Res 2025;9:e63568) doi:10.2196/63568

KEYWORDS

older adults; cognitive training; digital therapeutic; DTx; remote; usability; acceptance; interviews; gerontology; geriatric; elderly; experiences; attitudes; opinions; perceptions; perspectives; interview; cognition; digital health; qualitative; thematic

Introduction

As individuals age, cognitive abilities mature at different time points and eventually decline [1]. This age-related cognitive decline commonly affects memory, learning ability, fluid intelligence, and multitasking skills [2,3]. In Singapore, aggregate declines may be particularly significant given its rapidly aging population, whereby nearly 1 in 4 Singaporeans will be over the age of 65 by 2030 [4]. This decline in cognitive abilities not only poses challenges for the workforce but also imposes a heavy burden on the health care system [5]. Given these factors, exploring effective methods to manage the decline of cognitive capabilities within this population would be essential to overcome these challenges.

Research has shown that cognitive training can improve cognitive functioning in older adults and even potentially delay the onset or progression of neurodegenerative disorders such as Alzheimer disease [6,7]. In Singapore, where there is high internet and mobile device penetration [8], cognitive training can be administered as a digital therapeutic (DTx) at scale. DTx are evidence-based software designed to prevent, manage, or treat health conditions [9]. They can be delivered remotely and often incorporate rewards and gamification elements to enhance adherence [10].

We have developed CURATE.DTx, a DTx platform that delivers personalized digital cognitive training by a tablet-ready adaptation of the Multi-Attribute Task Battery (MATB) platform, a flight deck simulator initially developed by the National Aeronautics and Space Administration (NASA) [11] and later further adapted by the United States Air Force (USAF) [12,13]. The MATB platform requires users to simultaneously manage 4 different tasks, which include responding to indicator lights and changing scales, tracking a target using a joystick, following verbal instructions to tune to specific radio frequencies and channels, and managing fuel tank levels. It is a naturalistic multitasking paradigm that has commonly been used as a mental workload measure for different task intensities [14,15] and has strong face validity as, by design, it involves attention, problem-solving, and coordination amongst tasks. Using the platform challenges the executive function by demanding sustained vigilance and goal-directed action for managing multiple tasks while also training attention and problem-solving for each specific task [16]. MATB has also been shown to evoke local and network-level brain responses, characteristic of multitasking [17], and such training may enhance cognitive efficiency and coordination [18].

CURATE.DTx also harnesses CURATE.AI, an adaptive approach that has previously been validated for drug dose optimization in clinical conditions, to personalize a user's digital cognitive training by correlating difficulty level doses (inputs) to performance outcomes (outputs) using only their own data [19,20]. While this was initially developed for cognitive training in healthy adult and postradiotherapy patient populations, the same application may serve as a potential solution for managing cognitive decline in older adults [19-22].

At its core, CURATE.DTx was designed to retain the primary tasks and functionality of MATB. However, the interface was modernized with gamification, and usability elements were also incorporated into the platform to improve user engagement and overall experience, such as the detailed instructions screens explaining the MATB tasks and a tablet form factor with a touchscreen for intuitive control. The inclusion of these usability features could also improve the accessibility of CURATE.DTx for older adults. Indeed, the widespread adoption of tablets as the medium for game-based interventions for older adults underscores their versatility for this population [23]. In addition, while the use of DTx in older adults is a relatively recent development, initial studies have shown promising outcomes [24,25]. As older adults often have impaired cognitive performance compared to young adults, especially in multitasking-related tasks under high task loads [26], it would be useful for CURATE.DTx to be adapted for this population. Finally, older adults tend to show greater levels of disengagement than young adults when faced with higher task difficulty [27]. Such disengagement could also impact their enjoyment, which is a predictor of motivation when engaging with an intervention [28]. Considering these factors, it is important to understand older adults' attitudes and experiences with CURATE.DTx, as their feedback is also helpful for improving future iterations of the platform. Therefore, this study aimed to examine the general acceptability and user experience of the CURATE.DTx platform among older Singaporean adults.

Methods

Overview

The Consolidated Criteria for Reporting Qualitative Studies (COREQ) guidelines were followed for the reporting of this

XSL•FO
RenderX

study [29], with the COREQ checklist presented in Multimedia Appendix 1.

CURATE.DTx Platform

The CURATE.DTx platform is comprised of 4 tasks: Systems Monitoring, Communications, Tracking, and Resource Management, as depicted in Figure 1. Systems Monitoring requires the user to monitor the lights and gauges and respond as quickly as possible when they illuminate or change color. Communications requires the user to listen for the call sign "NGT504" and set the radio station and frequency according to the audio instructions, ignoring instructions for other call signs. Tracking requires the user to keep the ball as close to the center of the cross-hair as possible, dragging the ball back when it deviates away. Finally, Resource Management requires the user to maintain the two fuel tank levels within the score area by opening and closing the arrow valves, making appropriate adjustments when the valves occasionally turn off or fail. All 4 tasks can occur at the same time, and the user has to manage them simultaneously until the session ends.

Figure 1. Screenshot of the CURATE.DTx's user interface. Top left: Systems Monitoring—respond to indicator lights and gauges. Top right: Communications—follow verbal instructions to tune to specific radio frequencies and channels. Bottom left: Tracking—keep the target as close to the center of the crosshair as possible. Bottom right: Resource Management—maintain the fuel tank levels.

Recruitment

This study was approved by the National University of Singapore's institutional review board (reference NUS-IRB-2021-95). Participants were recruited through purposive sampling by age group from a community senior activity center in Singapore. The activity center coordinator identified potential participants, provided an overview of the study, and arranged for informed consent and study sessions. Participants were also told about the research team's involvement in the study and the reasons for doing the research. The research team (SPK, VVL, SV, and AR) consisted of full-time researchers, 3 female postdoctoral fellows, and 1 male research assistant, and had previous training in qualitative data collection and analyses. They screened potential participants against the following inclusion criteria: (1) Singaporean older adults aged 60 years old and older and (2) fluent in English. Participants with significant hearing, visual, or cognitive impairment that would affect their ability to complete the study were excluded. No relationships with the participants were established before the interview sessions. Recruitment took place over a period of 4 months and ended when the target sample of 15 was obtained. No participants refused to participate or dropped out after consenting to the study.

Data Collection

This study used semistructured interview sessions to understand the participants' experiences and attitudes toward CURATE.DTx. Each session was approximately 90 minutes and consisted of a 30-minute playtest where participants play-tested CURATE.DTx on a Samsung Galaxy Tab A7 (SM-T500) 10.4-inch tablet, followed by a 60-minute semistructured interview on their thoughts and experience. To reduce the risk of bias and enhance the credibility and robustness of the results, methods triangulation was used [30]. We compared the participants' actions observed during their playtest with their verbal qualitative feedback provided in the subsequent interview to ensure consistency of findings across different mediums. All sessions were in-person at the senior activity center's premises and were conducted in English with at least 2 members of the research team and no nonparticipants present.

Before the session commenced, the research team provided an overview of the study, and informed consent was obtained. All participants consented to the audio recording of the session for transcription purposes, while the participants' demographic background details were also collected. While performance score data from the sessions were automatically collected, they were not accessed or analyzed as they did not relate to the purpose of this study. Open-ended questions were asked during the interview based on a semistructured interview guide, which was developed to include guiding topics that broadly cover the factors of the Technology Acceptance Model (TAM) framework [31]. The guiding topics used for the interview discussion are presented in Table 1. While the research team's past familiarity with CURATE.DTx could have influenced them to engage with

XSL•FO
RenderX

participants about it in a more positive manner, these guiding topics ensured that both positive and negative feedback were fully captured. Short field notes were also taken where necessary to ensure understanding and clarity of participants' responses. At the end of the session, all participants were reimbursed for their time. All data collected, including signed consent forms, interview recordings, and demographic data, were stored in a secure database and deidentified for the purpose of this study.

Table 1. Semistructured interview topic guide.

Category	Guiding topics for discussion
Interface-specific	• Discuss the purpose of the tasks. • Discuss understanding of the tasks. • Follow up on slips and errors from the playtest session. • Discuss the ease of use and difficulties with using the interface.
Program-specific	• Discuss the expectations of joining this program. • Discuss factors that would increase motivation toward adoption. • Discuss barriers and concerns of adoption. • Discuss alternatives to this program.

Data Analysis

The interview recordings were transcribed verbatim, and inductive thematic analysis was used to identify patterns and emerging or recurring themes [32]. Initial coding was conducted on Microsoft Word. Participants' responses were descriptively labeled and summarized into codes independently by the 4 members of the research team. Any discrepancies in the codes were resolved through discussion. Codes were then organized into potential themes using Microsoft Excel and further reviewed and refined in relation to the entire data set. The final codes and broader themes were generated from multiple discussions and iterations by the research analysis team (SPK, VVL, SV, and AR). Data saturation was achieved when there were no new themes identified and confirmed through discussion within the research team. There were no repeat interviews, transcripts were not returned to participants for comments, and participants did not provide feedback on the findings.

Ethical Considerations

This study was approved by the National University of Singapore's institutional review board (NUS-IRB reference: NUS-IRB-2021-95). All participants provided informed consent and were informed of their right to withdraw from the study at any time. All data collected were deidentified for the purpose of this study. At the end of their sessions, participants were reimbursed SGD 45 (US $34) for their time and participation.

Results

Participant Characteristics

A total of 15 participants completed the study (10 males, 5 females; mean age 66.1, SD 3.5, range 62-73 years). Participants were from a range of education level backgrounds, with 5 O Levels (10+ years), 6 Diploma (13+ years), and 4 University (16+ years). All participants reported regular use of technology, such as mobile phones, computers, televisions, and tablets. Full participant demographics and characteristics are presented in Table 2.

XSL•FO
RenderX

Table 2. Participant demographics and characteristics.

Participant	Age (years)	Gender	Highest education level[a]	Current technology usage
1	63	Male	Diploma	Mobile phone, computer, television
2	67	Male	Diploma	Mobile phone, computer, television
3	67	Male	University	Mobile phone
4	68	Male	O Levels	Mobile phone
5	68	Male	Diploma	Mobile phone, computer, television, tablet
6	67	Male	O Levels	Mobile phone
7	72	Male	Diploma	Mobile phone
8	62	Female	University	Mobile phone
9	66	Male	University	Mobile phone, television, tablet
10	63	Female	Diploma	Mobile phone, television
11	64	Female	University	Mobile phone
12	69	Male	O Levels	Mobile phone, computer, tablet
13	73	Male	O Levels	Mobile phone, computer
14	73	Female	O Levels	Mobile phone
15	64	Female	Diploma	Mobile phone, television, tablet

[a]The years of education based on the Singapore education system is as follows: O Levels=10+ years, Diploma=13+ years, University=16+ years.

Interview Data

From the participants' responses, 3 main themes emerged—comprehension, acceptability, and facilitators to adoption. Comprehension describes the participants' understanding of the CURATE.DTx platform, acceptability covers their thoughts and attitudes towards the CURATE.DTx platform and the tablet hardware, and facilitators to adoption encompass feedback on several recommendations discussed with the participants.

Theme 1: Comprehension

Under the theme of comprehension, 2 subthemes emerged from participants' responses, namely instruction comprehension, which is the feedback specific to the "How To Play" instructions page on the web-based platform, and task comprehension, their feedback on the individual MATB tasks and overall gameplay.

Instruction Comprehension

Participants initially struggled to understand CURATE.DTx instructions, mainly regarding 2 aspects: the complexity of the instructions and a misconception regarding simultaneous task management.

Participants expressed that the provided instructions were long-winded and difficult to understand, often perceiving them as complicated. Participant 11 mentioned the length of the instructions that, "Firstly, the instructions are very long." Participant 13 further emphasized their initial struggle, stating, "It is quite complicated actually when you first look at it, I actually struggled initially to try and grasp what this whole thing is about."

In addition, participants misunderstood the instructions as they did not realize that all tasks had to be played simultaneously. Although it was explicitly mentioned early in the instructions, the sequential presentation of task instructions led some participants to expect CURATE.DTx is to be conducted in a single-task manner. Participant 5 explained this misunderstanding, "The first instruction says four tasks are running at the same time, and it shows during the explanation the four tasks will be coming out, but it is explaining one task at a time, [so I thought] okay so this [CURATE.DTx] is one task at a time."

Task Comprehension

Participants conveyed that it was challenging to understand the more complicated tasks from just the instructions, and that it took some time using the platform to get used to them. As Participant 11 expressed, "I never read instructions fully. I skim. So it's only as I play [that] I more or less get the gist of it, I know what's happening." However, participants also voiced concerns it might potentially be too complicated and challenging for some in their age group. Participant 9 cautioned, "I guess the elderly, if you give them too many instructions and too many things for multitasking, they might just switch off their mind, they find it a bit difficult for them."

Participants mentioned that using CURATE.DTx itself helped them understand how the tasks worked. They relied on the in-game task feedback to figure out if they were doing it correctly. Participant 9 remarked, "Initially, I just don't understand what this is, this whole thing. But after a while, oh yeah, I got the thing." Participant 5 emphasized the importance of effective task feedback, stating, "In terms of my understanding of the game, the task will give a person better feedback on what to [do], how to react."

However, participants still expressed a desire for additional assistance and guidance when they faced difficulties in

XSL•FO
RenderX

understanding CURATE.DTx. Participant 4 mentioned, “I think like when I’m jammed at a particular part, I ask you all, you all must help me then I [can clearly] understand the game.”

Theme 2: Acceptability

Under the theme of acceptability, 3 subthemes emerged from participants’ responses. The subthemes included table usability, engagement, and enjoyment of CURATE.DTx, and their attitude and perceived benefits of the intervention.

Tablet Usability

Participants mentioned that the usability of the tablet, namely its size and weight, were important factors to consider for the acceptability of CURATE.DTx in this population. While most were satisfied with the physical attributes of the provided tablet, some expressed concerns about other usability aspects of the tablet. Participant 12 shared that, “I think it’s okay. Too big, my age group will have a problem holding. So, lightweight [is better].” However, Participant 10 mentioned the potential issue of screen reflectivity, saying, “Just now the light was glaring, so I have to hold it up so that I can see better.”

Engagement and Enjoyment

Participants exhibited high engagement with CURATE.DTx, maintaining a strong level of focus throughout the entire playtest session. Participant 3 remarked afterward, “Like it keeps me on my toes. You didn’t see? If you look at this you will see me very attentive, listening and all this.”

Regarding their enjoyment of various aspects of the DTx, participants acknowledged that certain tasks were more enjoyable than others, with a general preference for the more challenging tasks over the easier ones. Participant 10 stated, “It’s nice to play... Challenging is nice.” Participant 12 had a similar opinion about a simpler task, stating, “This one [Systems Monitoring] is actually very simple, too simple already. Just pressing a button. No fun at all!”

Furthermore, some participants appreciated the presence of a mix of easier and more difficult tasks, as it added variety and enhanced their enjoyment. Participant 13 highlighted this aspect, “In a way, the four [tasks], there is variety... Not all like this [Systems Monitoring], this [is] kind of standard, then I think it's not so enjoyable.”

Attitude and Perceived Benefits

Participants expressed that they found CURATE.DTx as an intervention could be potentially beneficial as it was tailored to their prospective cognitive needs, with many considering cognitive training to be important for maintaining their mental capacity in their senior years. Participant 5 emphasized the significance of cognitive stimulation, “For me, my purpose is to stimulate the thinking, the brain, or else your brain deteriorates over time.” Participants viewed CURATE.DTx as a tool for cognitive stimulation and expressed strong intrinsic motivation and interest in using it for cognitive training to challenge themselves. Participant 15 remarked, “I might play this one, for training. Better than the sudoku.”

Theme 3: Facilitators to Adoption

Under the theme of facilitators to adoption, 3 subthemes emerged from participants’ responses. The subthemes include framing and aesthetics recommendations, motivation recommendations, and the role of medical professionals.

Framing and Aesthetics Recommendations

When discussing CURATE.DTx’s visual design, participants suggested that it would be more attractive to older adults if the tasks were framed as real-world, daily life analogies, as they found the current aviation-related tasks to be less relatable. Participant 11 illustrated, “I look at this one, you’ve got petrol, that one is the meter, maybe you can see this car dashboard? And then this one is your air-con slider. And from there at least I can associate, the person also can associate.”

Participants also expressed that the general visual appeal of CURATE.DTx could be improved. They found the color palette to be dull and suggested incorporating more visually appealing elements. Participant 8 elaborated, “If you say that you're going to do it in this color, it is very boring. Because for me, I am a very bright color person, blue, blue, blue? I don’t know.”

Motivation Recommendations

When discussing motivation, most participants showed strong intrinsic motivation towards using CURATE.DTx for cognitive training. However, participants also expressed approval for the addition of extrinsic motivation factors, such as the inclusion of a scoring system to CURATE.DTx. They believed that a scoring system would provide them with feedback on their performance and motivate them to strive for improvement. Participant 12 illustrated this viewpoint, saying, “For example, today I score 80, I want to aim for 85. 85, got to aim for 90, 90, 95! People will continue to chase. So if you don't have this, it's [just] another game.”

Some participants thought that CURATE.DTx could be boring over time due to the lack of variety of tasks and levels. As Participant 3 described it, “[it] can be quite boring, to play this everyday. Like there is no changes. Like you see the other games, you see the clearing the blocks game, that one every time change level. That one is different.” They also mentioned that CURATE.DTx lacked some common and attractive features compared to other more attractive existing games in the market. Participant 12 noted, “Most of the time, I play with the music. You see, [the music] attract[s] you, you know?”

The potential gameplay duration of each CURATE.DTx session also emerged as a factor that affected participants’ motivation to use the platform. Overall, most participants were satisfied with the playtest session’s duration, with Participant 12 considering it “not too long, not too short” and Participant 9 saying that, “I think it’s just right.” Participants also emphasized the importance of CURATE.DTx in fitting into their daily schedule. Participant 6 mentioned, “My time is very tight... If I got the time, I don’t mind.” Participant 10 echoed this opinion, saying, “Yes, I would if I have time.”

Role of Medical Professionals

When discussing the role of doctors and medical professionals in interventions, most participants expressed a strong belief in

XSL•FO
RenderX

the recommendations of their medical personnel, emphasizing that they would heed their advice. Participant 10 succinctly conveyed this sentiment, stating, "Anyone with medical [knowledge] who is looking after me medically, yes, I will take their advice seriously."

Furthermore, several participants were hopeful that the intervention could serve as a potential alternative to medication. Participant 3 expressed this sentiment, stating, "If this game can help, then why not? Instead of medication." Since the intervention was in the form of a game, participants generally did not expect any adverse effects. Participant 8 also commented on the absence of adverse effects typically associated with games, stating, "Generally games, the negative effects would be addiction, but I think this one doesn't have."

Furthermore, participants demonstrated a higher willingness to adopt the platform if it was scientifically proven to be beneficial for their health, even if they wouldn't have considered using it otherwise. Participant 3 elaborated on this point, explaining, "If you say by playing this twice a week, every day, for half an hour, you will improve your mental state of health, then I will definitely play since it is scientifically proven."

Discussion

Principal Findings

This study aimed to understand the acceptability and user experience of older Singaporean adults using CURATE.DTx, a digital cognitive training DTx platform. Overall, we found that participants faced some difficulties with understanding the CURATE.DTx platform. However, engagement with the platform was generally still strong, and participants were also enthusiastic about using it for cognitive training, especially if advised by medical professionals. In light of these findings, there is a need to address several key factors for the next iteration of the CURATE.DTx platform to ensure greater adoption and sustained usage among this older adult demographic group, with these considerations broadly applicable to the development of other similar DTx platforms in this space as well.

Comprehension Considerations

We identified the importance of considering the comprehension of the instructions on overall experience when developing a cognitive training DTx. Aging has been shown to be associated with declining cognitive abilities that could affect information processing, such as in visual attention [33] and working memory [34]. Specifically, older adults may have a harder time processing unfamiliar visual information and retaining it in memory [35]. We observed this in some of our participants, who struggled to fully understand the game instructions provided on CURATE.DTx and perceived them to be too long-winded and complicated. Some of this difficulty can be attributed to the complexity of several task explanations, as well as the continuous sequential presentation of the instructions. Therefore, older adults would benefit from having complex concepts broken down into smaller portions so that each step is within their capacity for processing, and this information should be summarized again for each task to ensure that it is adequately rehearsed and retained in their memory [36]. Providing accessible bite-sized chunks of information that are reiterated at various points of the instructions will ensure that they are clear and concise, which would prevent misunderstanding and improve their overall experience when using CURATE.DTx. In addition, participants may also benefit from the implementation of instructions explained as an interactive tutorial with the opportunity to practice all 4 tasks simultaneously at the end. This would provide them with the opportunity to familiarize themselves with the tasks and gain a better understanding of the mechanisms involved.

Our findings also showed that providing in-game feedback can be a useful tool for helping older adult users understand the tasks. In line with Nielsen's usability heuristics, which emphasizes recognition rather than recall [37], incorporating visual elements that provide small reminders of information can aid users in navigating the task without requiring them to remember specific task instructions. Our study participants themselves mentioned a preference for learning CURATE.DTx mechanics through gameplay, using in-game feedback to gauge their performance and adjust their strategies. This is in line with the literature, where feedback in educational games can provide players with schemas to help them learn from their misunderstandings [38]. Indeed, providing regular feedback about their performance enables users to engage in more critical learning [39], and clear and specific feedback—such as indicating correct and incorrect actions, highlighting areas for improvement, and offering suggestions or tips—can direct them to develop more effective game strategies. Furthermore, it is also important to present feedback in a way that is easily understandable to older adults, taking into account their cognitive abilities and visual acuity. Hence, designing a user-friendly interface with intuitive controls, clear visual cues, and accessible feedback would enhance their interaction and learning with CURATE.DTx, thus improving their overall experience.

Acceptability Considerations

Our findings showed that there was strong interest in using CURATE.DTx for cognitive training. Participants recognized the importance of cognitive training for maintaining cognitive functioning and believed that the CURATE.DTx platform would be useful and beneficial for this purpose. These stated views align with previous research showing that older adults in Singapore are open to adopting technology to achieve their health goals [40]. Furthermore, the intended and perceived benefits of the training platform emerged as the most common facilitators of its use. This finding aligns with the principles of the TAM framework, which suggests that the perceived usefulness of a technological tool strongly influences its acceptance as a learning and training resource [41]. Given the existing intrinsic motivation and expressed interest in CURATE.DTx, ensuring sustained engagement, would thus be key to its long-term adoption in older adults.

Engagement and enjoyment were also found to affect the acceptability of the CURATE.DTx platform. Specifically, participants expressed a preference for more challenging tasks, like Resource Management, which requires them to develop an

XSL•FO
RenderX

optimal strategy for balancing resources over simpler tasks, like Systems Monitoring, which merely requires them to click on lights when they illuminate. As facing challenges in a game has previously been reported to directly enhance learning and increase engagement with the game [42], incorporating more challenging tasks can offer participants a sense of accomplishment and keep them motivated. However, we also found that some tasks were reported to be overly complex and difficult to understand for this age group. Thus, striking a "sweet spot" trade-off between challenge and ease of understanding can lead to flow—the sense of engagement and immersion in the interactive experience [43]. While the use of an adaptive training approach such as CURATE.AI can partially aid in this by adjusting the difficulty to an optimal level based on personal performance [19], improving the design of the task elements based on participant feedback would help to ensure a more cohesively enjoyable experience.

The source of the challenge is also an important consideration. While there can be a motivating in-game frustration that causes the player to want to persevere and overcome the challenge, lack of comprehension of CURATE.DTx's tasks could instead lead to disheartening at-game frustration that causes them to disengage [44,45]. As these seemingly contradictory aspects are rated most highly in terms of importance for older adults [46], successfully finding this balance can make the platform highly appealing to this demographic group.

Finally, participants generally viewed using tablets as the physical medium for CURATE.DTx positively. However, the size and weight of the tablets were highlighted as important factors to consider, underscoring the need for a device that is easy to manage and handle. Despite some concerns about screen reflectivity, the overall design of the tablet was well-received. This is in line with existing literature reporting high satisfaction among older individuals using tablets in clinical settings [47] as well as their perceived benefit not just for cognitive training but also for enjoyment [48]. Given that the effective deployment of a tablet-based DTx in a remote setting use case requires addressing additional factors such as ensuring household internet access, it is important to note that in Singapore, where 99% of resident households, including 93% of senior-only households, have internet access [49], this is less of a concern compared with the digital divide at the global level [50], with disparities in internet access often linked poverty and unemployment [51]. Therefore, understanding the specific needs and circumstances of the target group is crucial for the successful adoption of the tablet-based platform.

Facilitators to Adoption Considerations

Our findings revealed that enhancing the framing and aesthetics of the current CURATE.DTx platform would make it more appealing and engaging to the older Singaporean population. Specifically, incorporating visually appealing elements and features found in popular games in the current market, such as vibrant colors and attractive visuals, can help to enhance enjoyment. Adhering to design principles, such as judicious color use with no more than 2 or 3 fully saturated intense colors, can assist in accomplishing this [52]. Participants also expressed a preference for tasks framed to be more relatable and familiar to their daily lives, such as a car dashboard, rather than the current MATB flight deck simulator interface incorporated into CURATE.DTx. Again, DTx developers should consider Nielsen's usability heuristics, specifically the principle of matching the system to the real world [37], to make the platform's task more intuitive. Using familiar symbols, illustrations, or other analogies that resonate with older adults' day-to-day lived experiences will ensure that they are able to understand the task requirements more easily. Such aesthetic changes can help to make the platform more appealing to older adults and increase their engagement with the platform [53].

Our findings also highlighted the importance of motivation in the DTx platform design. While participants already demonstrated intrinsic motivation in using the platform, they also welcomed extrinsic motivational factors to further strengthen their interest. Participants were particularly taken with the idea of gamification elements such as a scoring system, reporting that knowledge of their scores would drive them to improve. Enhancing scoring with a leaderboard may be an option to consider in a senior-friendly iteration of CURATE.DTx, as they have previously been shown to motivate players to strive for higher-difficulty goals [54]. Nevertheless, it is important to ensure that extrinsic motivation factors do not undermine participants' existing intrinsic motivation to use the CURATE.DTx platform but rather enhance it [55].

Participants also supported the introduction of a greater variety of tasks for CURATE.DTx. This can be a means to prevent boredom and also promote meta-learning, a concept of applying knowledge from past experiences to new, related tasks [56]. While the concept is mainly applied in machine learning and artificial intelligence, recent research has proposed that this idea of drawing from past learned experience can be applied in cognitive training, in particular with cognitive control [57]. Promoting such learning could be beneficial to the older adults' mental health while also increasing the appeal of the CURATE.DTx platform and improving its long-term engagement and adoption.

Finally, our findings revealed the role of medical professionals to be pivotal in the adoption process. In Singapore, health care providers are highly trusted sources for learning about health platforms [58,59], thus playing a critical role in influencing their patient's decision to adopt new technologies. Participants also highlighted the importance of scientific evidence demonstrating the benefits of the DTx, likely reflecting this population's trust in research [60]. They also viewed the DTx as a potential alternative to medication, indicating their willingness to explore nonpharmacological approaches to cognitive maintenance. Thus, collaborating with health care professionals and conducting further research studies to demonstrate the efficacy and cognitive benefits of the game could significantly increase trust and promote broader adoption among older Singaporean adults.

Limitations

Several limitations should be considered when interpreting the findings of this study. First, the study was restricted to English-speaking older adults, which may not fully capture the experiences and perspectives of older adults from

XSL•FO
RenderX

non–English-speaking backgrounds. Singapore is a multilingual country, and considering the diverse linguistic landscape, it is important to extend the research to include participants who speak languages other than English. This would provide a more comprehensive understanding of older adults' acceptability and experiences with the cognitive training digital therapeutics platform.

In addition, it is worth highlighting that this study is focused solely on the particular DTx cognitive training platform used in this research. It's important to acknowledge that this platform was not originally tailored for older adult users and that moving forward, it is recommended that future iterations be codesigned with the target demographic to develop a more bespoke solution. Future research should explore a wider range of cognitive training interventions and digital therapeutics platforms to gain a more comprehensive understanding of older adults' acceptability and experiences across different interventions.

Conclusion

Overall, this study investigated the general acceptability and user experience of a cognitive training DTx in an older adult Singapore population. Despite facing some difficulties in instructional and task comprehension, participants were still highly engaged with the platform and found the challenge provided by the DTx to be enjoyable. Participants also expressed interest and recognized the importance of cognitive training for their population. We also found that certain aesthetics and motivational features and the incorporation of medical professionals into the onboarding process would facilitate greater adoption and long-term usage in this population. These findings highlight several important considerations that can serve as a useful reference to inform the design and implementation strategies of the future iteration of CURATE.DTx as well as other cognitive training interventions for this population.

Acknowledgments

We gratefully acknowledge funding from the Singapore Cancer Society (grant SCS-GRA-2019-00063) for funding this study and had no influence on any part of it. DH gratefully acknowledges funding from the National Research Foundation Singapore under its AI Singapore Programme (award AISG-GC-2019-002) and the Singapore Ministry of Health's National Medical Research Council under its Open Fund- Large Collaborative Grant ('OF-LCG'; grant MOH-OFLCG18May-0028), Institute for Digital Medicine (WisDM) Translational Research Programme (grant R-719-000-037-733) at the Yong Loo Lin School of Medicine, National University of Singapore, Ministry of Education Tier 1 FRC Grant (grant R-397-000-333-114) and the Next-Generation Brain-Computer-Brain Platform—A Holistic Solution for the Restoration & Enhancement of Brain Functions (NOURISH) project from the RIE2020 ADVANCED MANUFACTURING AND ENGINEERING (AME) PROGRAMMATIC FUND (grant A20G8b0102/A-0002199-02-00). All funders have no influence on the study design, collection, management, analysis, interpretation of data, writing of the report, and the decision to submit the report for publication. We would also like to thank Jason Labbe for his assistance in the development of the CURATE.DTx platform and C&B for enabling a collaborative environment, and Lion Befrienders and Joshann Lee for their kind assistance during the data collection process.

Conflicts of Interest

CLA and DH are coinventors of intellectual property pertaining to personalized digital therapeutics and CURATE.AI. DH is a shareholder of KYAN Therapeutics, which has licensed intellectual property pertaining to AI-based oncology drug development. He has also filed a provisional patent on data-driven behavior change.

Multimedia Appendix 1
Consolidated Criteria for Reporting Qualitative Studies (COREQ) checklist.
[PDF File (Adobe PDF File), 416 KB - formative_v9i1e63568_app1.pdf]

References

1. Hartshorne JK, Germine LT. When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. Psychol Sci 2015;26(4):433-443 [FREE Full text] [doi: 10.1177/0956797614567339] [Medline: 25770099]
2. Bettio LEB, Rajendran L, Gil-Mohapel J. The effects of aging in the hippocampus and cognitive decline. Neurosci Biobehav Rev 2017;79:66-86. [doi: 10.1016/j.neubiorev.2017.04.030] [Medline: 28476525]
3. Kievit RA, Davis SW, Mitchell DJ, Taylor JR, Duncan J, Cam-CAN Research Team, Cam-CAN Research Team. Distinct aspects of frontal lobe structure mediate age-related differences in fluid intelligence and multitasking. Nat Commun 2014;5(1):5658 [FREE Full text] [doi: 10.1038/ncomms6658] [Medline: 25519467]
4. PMO | PM lee hsien loong at the 'Singapore ageing: issues and challenges ahead' Book Launch. Prime Minister's Office Singapore. Prime Minister's Office Singapore. URL: https://www.pmo.gov.sg/Newsroom/PM-Lee-Hsien-Loong-at-the-Singapore-Ageing-Issues-and-Challenges-Ahead-Book-Launch [accessed 2023-04-11]
5. Tan CC, Lam CSP, Matchar DB, Zee YK, Wong JEL. Singapore's health-care system: key features, challenges, and shifts. Lancet 2021;398(10305):1091-1104. [doi: 10.1016/S0140-6736(21)00252-X] [Medline: 34481560]

XSL•FO
RenderX

6. Rahe J, Petrelli A, Kaesberg S, Fink G, Kessler J, Kalbe E. Effects of cognitive training with additional physical activity compared to pure cognitive training in healthy older adults. CIA 2015;10:297. [doi: 10.2147/cia.s74071]
7. Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of alzheimer's disease. Alzheimers Dement 2007;3(3):186-191. [doi: 10.1016/j.jalz.2007.04.381] [Medline: 19595937]
8. World Bank. Individuals using the Internet (% of population) - Singapore. World Bank Open Data. URL: https://data.worldbank.org [accessed 2023-11-16]
9. Digital Therapeutics Alliance.. What is a DTx? Digital Therapeutics Alliance. URL: https://dtxalliance.org/understanding-dtx/what-is-a-dtx/ [accessed 2023-11-16]
10. Mahncke HW, Kim SJ, Rose A, Stasio C, Buckley P, Caroff S, et al. Evaluation of a plasticity-based cognitive training program in schizophrenia: results from the eCaesar trial. Schizophr Res 2019;208:182-189 [FREE Full text] [doi: 10.1016/j.schres.2019.03.006] [Medline: 30930034]
11. Comstock JR, Arnegard RJ. The Multi-Attribute Task Battery for Human Operator Workload and Strategic Behavior Research. 1992. URL: https://ntrs.nasa.gov/citations/19920007912 [accessed 1992-01-01]
12. Miller W. Air Force Research Laboratory. The US air force-developed adaptation of the multi-attribute task battery for the assessment of human operator workload and strategic behavior. 2010. URL: https://apps.dtic.mil/sti/citations/ADA537547 [accessed 2010-07-01]
13. Miller W, Schmidt K, Estepp J, Bowers M, Davis I. DTIC Document. An updated version of the US air force multi-attribute task battery (AF-MATB). URL: https://apps.dtic.mil/sti/citations/ADA611870 [accessed 2014-08-01]
14. Liu S, Nam CS. Quantitative modeling of user performance in multitasking environments. Computers in Human Behavior 2018;84:130-140. [doi: 10.1016/j.chb.2018.02.035]
15. Pontiggia A, Gomez-Merino D, Quiquempoix M, Beauchamps V, Boffet A, Fabries P, et al. MATB for assessing different mental workload levels. Front Physiol 2024;15:1408242 [FREE Full text] [doi: 10.3389/fphys.2024.1408242] [Medline: 39108543]
16. Kee T, Weiyan C, Blasiak A, Wang P, Chong JK, Chen J, et al. Harnessing CURATE.AI as a digital therapeutics platform by identifying N‐of‐1 learning trajectory profiles. Advanced Therapeutics 2019;2(9):1900023. [doi: 10.1002/adtp.201900023]
17. Lam TK, Vartanian O, Hollands JG. The brain under cognitive workload: neural networks underlying multitasking performance in the multi-attribute task battery. Neuropsychologia 2022;174:108350. [doi: 10.1016/j.neuropsychologia.2022.108350] [Medline: 35988804]
18. von Bastian CC, Belleville S, Udale RC, Reinhartz A, Essounni M, Strobach T. Mechanisms underlying training-induced cognitive change. Nat Rev Psychol 2022;1(1):30-41. [doi: 10.1038/s44159-021-00001-3]
19. Blasiak A, Khong J, Kee T. CURATE.AI: optimizing personalized medicine with artificial intelligence. SLAS Technol 2020;25(2):95-105. [doi: 10.1177/2472630319890316] [Medline: 31771394]
20. Remus A, Tadeo X, Kai GNS, Blasiak A, Kee T, Vijayakumar S, et al. CURATE.AI COR-Tx platform as a digital therapy and digital diagnostic for cognitive function in patients with brain tumour postradiotherapy treatment: protocol for a prospective mixed-methods feasibility clinical trial. BMJ Open 2023;13(10):e077219 [FREE Full text] [doi: 10.1136/bmjopen-2023-077219] [Medline: 37879700]
21. Wang P, Leong QY, Lau NY, Ng WY, Kwek SP, Tan L, et al. N-of-1 medicine. Singapore Med J 2024;65(3):167-175 [FREE Full text] [doi: 10.4103/singaporemedj.SMJ-2023-243] [Medline: 38527301]
22. Tan SB, Tan J, Raczkowska MN, Chean Wen Lee J, Rai B, Remus A, et al. Digital game-based interventions for cognitive training in healthy adults and adults with cognitive impairment: protocol for a two-part systematic review and meta-analysis. BMJ Open 2023;13(5):e071059 [FREE Full text] [doi: 10.1136/bmjopen-2022-071059] [Medline: 37142320]
23. Koivisto J, Malik A. Gamification for older adults: a systematic literature review. Gerontologist 2021;61(7):e360-e372 [FREE Full text] [doi: 10.1093/geront/gnaa047] [Medline: 32530026]
24. Al-Saleh S, Lee JK, Rogers WA, Insel KC. Translation of a successful behavioral intervention to a digital therapeutic self-management system for older adults. Ergon Des 2024;32(2):5-13 [FREE Full text] [doi: 10.1177/10648046211066409] [Medline: 38487251]
25. Bohle H, Rimpel J, Schauenburg G, Gebel A, Stelzel C, Heinzel S, et al. Behavioral and neural correlates of cognitive-motor interference during multitasking in young and old adults. Neural Plast 2019;2019 [FREE Full text] [doi: 10.1155/2019/9478656] [Medline: 31582967]
26. Tye-Murray N, Spehar B, Mauze E, Cardinal C. Hearing health care digital therapeutics: patient satisfaction evidence. Am J Audiol 2022;31(3S):905-913. [doi: 10.1044/2022_aja-21-00236]
27. Ennis GE, Hess TM, Smith BT. The impact of age and motivation on cognitive effort: implications for cognitive engagement in older adulthood. Psychol Aging 2013;28(2):495-504 [FREE Full text] [doi: 10.1037/a0031255] [Medline: 23421325]
28. Boot WR, Souders D, Charness N, Blocker K, Roque N, Vitale T. In: Zhou J, Salvendy G, editors. The Gamification of Cognitive Training: Older Adults? Perceptions ofAttitudes Toward Digital Game-Based Interventions. United States: Springer International Publishing; 2016.

XSL•FO
RenderX

29. Booth A, Hannes K, Harden A, Noyes J, Harris J, Tong A. COREQ (Consolidated Criteria for Reporting Qualitative Studies). In: Moher D, Altman DG, Schulz KF, Simera I, Wager E, editors. Guidelines for Reporting Health Research: A User's Manual. United States: Wiley; 2014.
30. Patton MQ. Enhancing the quality and credibility of qualitative analysis. Health Serv Res 1999;34(5 Pt 2):1189-1208 [FREE Full text] [Medline: 10591279]
31. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Mar;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
32. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
33. Jefferies LN, Roggeveen AB, Enns JT, Bennett PJ, Sekuler AB, Di Lollo V. On the time course of attentional focusing in older adults. Psychol Res 2015;79(1):28-41. [doi: 10.1007/s00426-013-0528-2] [Medline: 24337971]
34. Read CA, Rogers JM, Wilson PH. Working memory binding of visual object features in older adults. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2016;23(3):263-281. [doi: 10.1080/13825585.2015.1083937] [Medline: 26344033]
35. Ebaid D, Crewther SG. Visual information processing in young and older adults. Front Aging Neurosci 2019;11:116 [FREE Full text] [doi: 10.3389/fnagi.2019.00116] [Medline: 31156422]
36. Sharit J, Czaja SJ. Overcoming older adult barriers to learning through an understanding of perspectives on human information processing. J Appl Gerontol 2020;39(3):233-241 [FREE Full text] [doi: 10.1177/0733464818794574] [Medline: 30160573]
37. Nielsen J, Molich R. Heuristic evaluation of user interfaces. 1990 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 1990 April 5; Washington USA. [doi: 10.1145/97243.97281]
38. Clark RC, Mayer RE. e-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning. Hoboken: Wiley; 2016.
39. Erhel S, Jamet E. Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. Computers & Education 2013;67:156-167. [doi: 10.1016/j.compedu.2013.02.019]
40. Low STH, Sakhardande PG, Lai YF, Long ADS, Kaur-Gill S. Attitudes and perceptions toward healthcare technology adoption among older adults in singapore: a qualitative study. Front Public Health 2021;9:588590 [FREE Full text] [doi: 10.3389/fpubh.2021.588590] [Medline: 33659232]
41. Granić A, Marangunić N. Technology acceptance model in educational context: a systematic literature review. Brit J Educational Tech 2019;50(5):2572-2593. [doi: 10.1111/bjet.12864]
42. Hamari J, Shernoff DJ, Rowe E, Coller B, Asbell-Clarke J, Edwards T. Challenging games help students learn: an empirical study on engagement, flow and immersion in game-based learning. Comput Hum Behav 2016;54:170-179. [doi: 10.1016/j.chb.2015.07.045]
43. Abuhamdeh S, Csikszentmihalyi M. The importance of challenge for the enjoyment of intrinsically motivated, goal-directed activities. Pers Soc Psychol Bull 2012;38(3):317-330. [doi: 10.1177/0146167211427147] [Medline: 22067510]
44. Gilleade KM, Dix A. Using frustration in the design of adaptive videogames. 2004 Presented at: Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology; June 5, 2005; Singapore. [doi: 10.1145/1067343.1067372]
45. Miller MK, Mandryk RL. Differentiating in-game frustration from at-game frustration using touch pressure. 2016 Presented at: Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces; November 9, 2016; United States. [doi: 10.1145/2992154.2992185]
46. Salmon JP, Dolan SM, Drake RS, Wilson GC, Klein RM, Eskes GA. A survey of video game preferences in adults: building better games for older adults. Entertain Comput 2017;21:45-64. [doi: 10.1016/j.entcom.2017.04.006]
47. Ramprasad C, Tamariz L, Garcia-Barcena J, Nemeth Z, Palacio A. The use of tablet technology by older adults in health care settings-is it effective and satisfying? A systematic review and meta analysis. Clin Gerontol 2019;42(1):17-26. [doi: 10.1080/07317115.2017.1322162] [Medline: 28665775]
48. Chen K, Lou VWQ, Lo SSC. Exploring the acceptance of tablets usage for cognitive training among older people with cognitive impairments: a mixed-methods study. Appl Ergon 2021;93:103381. [doi: 10.1016/j.apergo.2021.103381] [Medline: 33578065]
49. Infocomm Media Development Authority. Household access to computer by household composition, 2017 - 2022. URL: https://www.imda.gov.sg/about-imda/research-and-statistics/digital-society [accessed 2023-11-16]
50. Seifert A, Cotten SR, Xie B. A double burden of exclusion? Digital and social exclusion of older adults in times of COVID-19. J Gerontol B Psychol Sci Soc Sci 2021;76(3):e99-e103 [FREE Full text] [doi: 10.1093/geronb/gbaa098] [Medline: 32672332]
51. Stringer S. Census and the city: Overcoming NYC's digital divide in the 2020 census. 2019. URL: https://comptroller.nyc.gov/reports/census-and-the-city/ [accessed 2019-07-23]
52. Sutcliffe A. Designing for User Engagment: Aesthetic and Attractive User Interfaces. Cham: Springer International Publishing; 2022.
53. Sayago S, Rosales A, Righi V, Ferreira SM, Coleman GW, Blat J. On the conceptualization, design, and evaluation of appealing, meaningful, and playable digital games for older people. Games and Culture 2015;11(1-2):53-80. [doi: 10.1177/1555412015597108]

XSL•FO
RenderX

54. Landers RN, Bauer KN, Callan RC. Gamification of task performance with leaderboards: a goal setting experiment. Comput Hum Behav 2017;71:508-515. [doi: 10.1016/j.chb.2015.08.008]
55. Hennessey B, Moran S, Altringer B, Amabile TM. Extrinsic and intrinsic motivation. Wiley Encyclopedia of Management. 2015. URL: https://onlinelibrary.wiley.com/doi/10.1002/9781118785317.weom110098 [accessed 2024-12-21]
56. Botvinick M, Ritter S, Wang JX, Kurth-Nelson Z, Blundell C, Hassabis D. Reinforcement learning, fast and slow. Trends Cogn Sci 2019;23(5):408-422 [FREE Full text] [doi: 10.1016/j.tics.2019.02.006] [Medline: 31003893]
57. Botvinick M, Wang JX, Dabney W, Miller KJ, Kurth-Nelson Z. Deep reinforcement learning and its neuroscientific implications. Neuron 2020;107(4):603-616 [FREE Full text] [doi: 10.1016/j.neuron.2020.06.014] [Medline: 32663439]
58. Lee ST, Dutta MJ, Lin J, Luk P, Kaur-Gill S. Trust ecologies and channel complementarity for information seeking in cancer prevention. J Health Commun 2018;23(3):254-263. [doi: 10.1080/10810730.2018.1433253] [Medline: 29436966]
59. Lee VV, Vijayakumar S, Lau NY, Blasiak A, Siah KTH, Ho D. Understanding the user: patients' perception, needs, and concerns of health apps for chronic constipation. Digit Health 2022;8 [FREE Full text] [doi: 10.1177/20552076221104673] [Medline: 35663236]
60. Striley CW, Lloyd S, Varma D, Vaddiparti K, Cottler LB. 3487 trust in research among older adults. J. Clin. Trans. Sci 2019;3(s1):98-98. [doi: 10.1017/cts.2019.223]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Studies
DTx: digital therapeutics
MATB: Multi-Attribute Task Battery
NASA: National Aeronautics and Space Administration
TAM: Technology Acceptance Model
USAF: United States Air Force

Edited by A Mavragani; submitted 17.07.24; peer-reviewed by N Buawangpong; comments to author 20.11.24; revised version received 04.12.24; accepted 09.12.24; published 13.01.25.

Please cite as:
Kwek SP, Leong QY, Lee VV, Lau NY, Vijayakumar S, Ng WY, Rai B, Raczkowska MN, Asplund CL, Remus A, Ho D
Exploring the General Acceptability and User Experience of a Digital Therapeutic for Cognitive Training in a Singaporean Older Adult Population: Qualitative Study
JMIR Form Res 2025;9:e63568
URL: https://formative.jmir.org/2025/1/e63568
doi:10.2196/63568
PMID:

©Siong Peng Kwek, Qiao Ying Leong, V Vien Lee, Ni Yin Lau, Smrithi Vijayakumar, Wei Ying Ng, Bina Rai, Marlena Natalia Raczkowska, Christopher L Asplund, Alexandria Remus, Dean Ho. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Proposed mHealth Intervention to Address Patient Barriers to Colposcopy Attendance: Qualitative Interview Study of Clinic Staff and Patient Perspectives

Jennifer R Hemler[1,2], PhD; Rachel B Wagner[3,4], MA; Brittany Sullivan[2], MPH; Myneka Macenat[5], MPH; Erin K Tagai[6], MPH, PhD; Jazmarie L Vega[6], BA; Enrique Hernandez[6], MD; Suzanne M Miller[6], PhD; Kuang-Yi Wen[7], PhD; Charletta A Ayers[8], MPH, MD; Mark H Einstein[9], MD; Shawna V Hudson[1,2,3], PhD; Racquel E Kohler[3,4], MSPH, PhD

[1]Department of Family Medicine and Community Health, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ, United States

[2]Center Advancing Research and Evaluation for Person-Centered Care, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ, United States

[3]Center for Cancer Health Equity, Rutgers Cancer Institute, New Brunswick, NJ, United States

[4]School of Public Health, Rutgers University, Piscataway, NJ, United States

[5]Cancer Prevention and Outcomes Data Support, Rutgers Cancer Institute, New Brunswick, NJ, United States

[6]Cancer Prevention and Control, Fox Chase Cancer Center, Temple University Health System, Philadelphia, PA, United States

[7]Department of Medical Oncology, Thomas Jefferson University, Philadelphia, PA, United States

[8]Department of Obstetrics, Gynecology and Reproductive Sciences, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ, United States

[9]Department of Obstetrics & Gynecology and Women's Health, Montefiore Medical Center, Bronx, NY, United States

Corresponding Author:
Jennifer R Hemler, PhD
Department of Family Medicine and Community Health
Rutgers Robert Wood Johnson Medical School
303 George St.
Suite 301, Rm 305
New Brunswick, NJ, 08901
United States
Phone: 1 848 932 0209
Email: hemlerje@rwjms.rutgers.edu

Abstract

Background: Cervical cancer disparities persist among minoritized women due to infrequent screening and poor follow-up. Structural and psychosocial barriers to following up with colposcopy are problematic for minoritized women. Evidence-based interventions using patient navigation and tailored telephone counseling, including the Tailored Communication for Cervical Cancer Risk (TC3), have modestly improved colposcopy attendance. However, the efficacious TC3 intervention is human resource-intense and could have greater reach if adapted for mobile health, which increases convenience and access to health information.

Objective: This study aimed to describe feedback from clinic staff members involved in colposcopy processes and patients referred for colposcopy regarding adaptions to the TC3 phone-based intervention to text messaging, which addresses barriers among those referred for colposcopy after abnormal screening results.

Methods: Semistructured depth qualitative interviews were conducted over Zoom [Zoom Communications, Inc] or telephone with a purposive sample of 22 clinic staff members (including clinicians and support staff members) and 34 patients referred for colposcopy from 3 academic obstetrics and gynecology (OB-GYN) clinics that serve predominantly low-income, minoritized patients in different urban locations in New Jersey and Pennsylvania. Participants were asked about colposcopy attendance barriers and perspectives on a proposed text message intervention to provide tailored education and support in the time between abnormal cervical screening and colposcopy. The analytic team discussed interviews, wrote summaries, and consensus-coded transcripts, analyzing output for emergent findings and crystallizing themes.

XSL•FO
RenderX

Results: Clinic staff members and patients had mixed feelings about a text-only intervention. They overwhelmingly perceived a need to provide patients with appointment reminders and information about abnormal cervical screening results and colposcopy purpose and procedure. Both groups also thought messages emphasizing that human papillomavirus is common and cervical cancer can be prevented with follow-up could enhance attendance. However, some had concerns about the privacy of text messages and text fatigue. Both groups thought that talking to clinic staff members was needed in certain instances; they proposed connecting patients experiencing complex psychosocial or structural barriers to staff members for additional information, psychological support, and help with scheduling around work and finding childcare and transportation solutions. They also identified inadequate scheduling and reminder systems as barriers. From this feedback, we revised our text message content and intervention design, adding a health coaching component to support patients with complex barriers and concerns.

Conclusions: Clinic staff members and patient perspectives are critical for designing appropriate and relevant interventions. These groups conveyed that text message-only interventions may be useful for patients with lesser barriers who may benefit from reminders, basic educational information, and scheduling support. However, multimodal interventions may be necessary for patients with complex barriers to colposcopy attendance, which we intend to evaluate in a subsequent trial.

(JMIR Form Res 2025;9:e55043) doi:10.2196/55043

KEYWORDS

cervical cancer screening; colposcopy; HPV; human papillomavirus; mHealth; health communication; qualitative research; cancer screening; cancer; cervical cancer; screening; women; clinic staff; barrier; messaging; privacy; text message; qualitative; colposcopic; mhealth intervention; mobile phone

Introduction

Over the last few decades, the incidence of invasive cervical cancer in the United States has decreased due to the high uptake of preventive screening and access to human papillomavirus (HPV) vaccination [1-3]. However, there are large and persistent disparities in cervical cancer incidence [4,5] and mortality [6] among minoritized populations [7-11]. Importantly, cancer cases are higher among women living in socially deprived communities (eg, high proportions living below poverty) with high proportions of racial and ethnic minorities and historical racial segregation [4,12,13]. There is variation in screening by race and ethnicity, geography, and insurance [14-18], and low follow-up after an abnormal test result, especially for Black and Hispanic women [19-21]. Indeed, the majority of invasive cervical cancer cases are diagnosed among women who have never or rarely screened [8,22-24]. Guideline-concordant screening has decreased over the last 2 decades, leaving many women under-screened and at risk [25,26]. Poor follow-up behaviors and barriers have also been documented among women diagnosed with cancer [8,27].

Follow-up management for abnormal screening results among minoritized groups is influenced by a myriad of barriers. Patients' low awareness or lack of knowledge [9,28-33], feelings of anxiety [29,33], and fear of receiving a fatal cancer diagnosis [31,34] are well-documented barriers to follow-up attendance. Patient confusion after receiving and discussing abnormal Papanicolaou or cotesting (ie, Papanicolaou and HPV tests) results with providers is common [35,36]. Delays in results delivery and challenges in appointment scheduling also contribute to low follow-up. Structural barriers such as lack of time, transportation, and childcare have also hindered colposcopy attendance [30]. Neighborhood factors, specifically living in more resource-deprived communities, and distance to care also influence time to follow-up, receipt of treatment, and ultimately outcomes, particularly among Black and Hispanic women [13,37-39]. Addressing the multiple, complex barriers to colposcopy remains critical to cervical cancer prevention.

Interventions using patient navigation [40,41] and tailored telephone counseling [42-44] have modestly improved colposcopy attendance. However, these human resource-intense interventions may have greater reach if they are adapted for new mobile health (mHealth) delivery platforms, which increase convenience and access to health information through automated direct communication [45,46]. Interventions using mHealth strategies have improved other aspects of cervical cancer prevention [47]. For example, online patient education tools administered before appointments increased knowledge about HPV or Papanicolaou testing and follow-up treatment [48,49]. In resource-limited settings, text messaging and app-based interventions were acceptable reminders [50], with some successfully increasing triage attendance [51,52] and others reducing psychosocial barriers to recommended follow-up [53-55]. However, many interventions have been developed outside of the United States and use one-way text messages to deliver cervical cancer screening results, which have not improved follow-up [54,55]. These studies highlight the need for additional features and content to address barriers and promote attendance. The flexibility [56,57] and ease of tailoring [58] interactive text messages have potential to increase an intervention's impact, but acceptability and appropriateness must be evaluated for women referred for colposcopy.

Therefore, as part of a multisite effort to adapt an efficacious phone-based counseling intervention, Tailored Communication for Cervical Cancer Risk (TC3) [42,59], to be delivered through preprogrammed text messages [60,61], we sought to incorporate feedback from those engaged in colposcopy services—clinicians and support staff members—and patients who had been referred for colposcopy in clinics that serve minoritized populations. In this manuscript, we report the major findings from interviews where participants provided insight on critical barriers to colposcopy attendance and feedback on the proposed SMS text

XSL•FO
RenderX

message intervention, useful functions the intervention should provide, and preferred delivery mode.

Methods

Intervention Adaptation

Our adapted intervention of the TC3 phone-based intervention, "Health Enhancement Resource System to address urban, cervical cancer disparities," is a pilot SMS text message intervention designed to deliver tailored educational and counseling messages to women who were referred for colposcopy after receiving abnormal cervical screening results. In TC3, patients completed a telephone-based session with a trained health educator. Using a computer-assisted telephone interviewing system, the health educator asked the patient a series of questions related to five Cognitive-Social Health Information-Processing (C-SHIP)–informed topic areas: (1) knowledge and perceptions of the implications of positive HPV and abnormal Papanicolaou tests, cervical cancer, and colposcopy; (2) expectancies and beliefs concerning colposcopy; (3) interfering affect in relation to cancer and colposcopy; (4) values and goals related to follow-up and physical well-being; and (5) self-regulation in terms of overcoming and coping with potential barriers to colposcopy attendance, including structural barriers [43,62]. Based on patients' responses, the computer-assisted telephone interviewing system algorithm gave the health educator a message to read to the patient for each C-SHIP barrier. In the Health Enhancement Resource System, patients awaiting colposcopy would receive a series of preprogrammed text messages adapted for text message delivery from TC3. For each domain, participants would receive a text asking them to assess their knowledge, level of concern, or perceived barrier, depending on the topic, on a 5-point scale. Participants would receive an intervention text message tailored to their response. For example, if expressing low knowledge or high concern about the barrier, they would receive an educational or counseling message; if expressing high knowledge or low concern about the barrier, they would receive an affirmation and move to the next text in the series. Participants could potentially receive all texts within 1 texting session or within a 2-week-long period before their colposcopy appointment.

Study Design and Setting

Qualitative semistructured depth interviews developed within a constructivist approach [63] were used to discuss facilitators and barriers to colposcopy attendance and to elicit feedback about the fit of the adapted text message intervention with clinic staff members and patients. We developed different interview guides for clinic staff members and patients, including additional questions on the patient interview guide in relation to domains of the C-SHIP model (Multimedia Appendix 1). The objectives for both sets of interviews were to determine whether the adapted intervention aligned with patients' needs and preferences and clinics' communication and counseling processes. For this study, we recruited a purposive sample of clinic staff members, including physicians, and patients. We asked participants about their perceptions and either personal or known barriers to colposcopy attendance as well as their thoughts about what would be useful in a text-based intervention aiming to help patients attend their colposcopy appointments. We included questions about demographics and social characteristics. In addition, we asked patient participants pointed questions about intervention modality preference, frequency, and timing of receiving messages.

The study included 3 obstetrics and gynecology (OB-GYN) and colposcopy clinics affiliated with 2 universities that serve patients predominantly of low socioeconomic status and minoritized groups in different urban locations in New Jersey (Site A, clinics 1 and 2) and Pennsylvania (Site B, clinic 3). The 3 clinics within the 2 sites were selected with the intent to pilot the adapted text message intervention after receiving feedback from this interview study. The clinics have similar patient care management processes and perform similar rates of screening and follow-up care, including colposcopy.

Participant Recruitment and Data Collection

Potential clinic staff member participants were referred to the study team by a clinic contact, who sent the team a list of clinic staff members involved in colposcopy services. At Site A, the clinics also sent the team a list of patients screened for eligibility and participation interest. Site B identified eligible patients through the electronic medical records, reviewing patients recently scheduled for colposcopy appointments. Clinic staff members were eligible to participate in interviews if they played a role in colposcopy services and were not a part of the research team. Patients were eligible if they were 18-70 years old, owned a cell phone with texting capability, and had received a colposcopy referral in the past year (completing the colposcopy was not a requirement). Patients were excluded if they were pregnant at the time of recruitment, had a history of malignancy, or had been diagnosed with invasive carcinoma of the cervix. Both groups of participants needed to be able to consent to participate. Site A participants were eligible if they spoke English or Spanish, whereas Site B participants were eligible if they spoke English.

A team of experienced health services researchers (JH, RB, BS, MM, JV, and Nicole Hernandez) trained in qualitative methods by a PhD health services researcher (JH and ET) and overseen by the larger team of cancer prevention and control, methodological, and clinical experts conducted qualitative semistructured depth interviews from July 2021 through March 2022 at Site A and until January 2023 at Site B. The team made up to 5 outreach attempts by phone and email to enroll potential participants into the study and emailed potential participants the study consent form to review ahead of the scheduled interview. Researchers answered all participant questions before asking participants for their consent. The trained researchers, all female and one bilingual in English and Spanish, conducted interviews in the participant's preferred language. Interviews were held over the telephone or via Zoom.

We received contact information for 50 eligible clinic staff members (n=16 at Site A; n=34 at Site B) and 145 eligible patient participants (n=29 at Site A; n=116 at Site B). Across sites, we were unable to reach 17 clinic staff members and 61 patients within the prescribed number of attempts. A total of 39 patients declined and 4 dropped out of the study after

XSL•FO
RenderX

providing consent. We were not able to schedule 4 clinic staff members and 7 patient interviews within the immediate time frame; saturation was achieved before being able to schedule these few interviews. Across sites, we interviewed 22 clinic staff members and 34 patients.

On average, clinician interviews were 53 minutes and patient interviews were 45 minutes. All interviews were recorded, professionally transcribed, translated from Spanish to English if necessary, and deidentified.

Data Analysis

At each site, the same team of health services researchers who conducted the interviews met routinely to discuss interviews on a rolling basis, writing summaries and creating a matrix of participant responses to C-SHIP–informed barriers and perspectives on the content and mode of the proposed intervention. From these materials, we created a codebook of a priori and emergent topics. Interviews were concluded upon reaching saturation at each site, which was determined upon discussing interviews and finding that information aligned with previous interviews and did not offer substantially different perspectives. After this, the teams concretized the code book and coded in pairs using Atlas.ti software (Lumivero; version 9) to reach consensus upon code application. Each site's analytic team discussed output and emergent themes with the larger study team, which included clinicians and health services researchers with expertise in cervical cancer prevention. The 2 sites compared and contrasted thematic findings, iterating themes until reaching crystallization [63]. The authors referred to the COREQ (Consolidated Criteria for Reporting Qualitative Research) [64] in reporting this research.

Ethical Considerations

This study was approved by Rutgers University's and Fox Chase Cancer Center's institutional review boards (Pro2020002750 and 21-1080, respectively). All participants provided informed consent prior to being interviewed. Data were deidentified. Participants were given a US $30 gift card for taking part in the study.

Results

Sample Characteristics

The 22 clinic staff members interviewed across clinics and sites were similar (Table 1). Most were physicians, female, and non-Hispanic White. Clinic participants from Site A averaged 48 years of age; however, age was not assessed at Site B. Clinic staff members from both sites reported serving various roles related to colposcopy services; some physicians performed both screening (ie, HPV and Papanicolaou tests) and colposcopies while others primarily performed colposcopies in specialized clinics or settings (ie, colposcopy clinic and gynecologic oncology team). Support staff members typically scheduled colposcopies for patients, while physicians or advanced practice providers communicated with patients directly about their colposcopy results and the next steps thereafter.

The 34 patients interviewed across the clinics were similar in most regards. The majority were non-Hispanic Black, worked full-time, and were single, divorced, or separated. We noted variation among patients in terms of age and education across the clinics. At Site A, Clinic 1, participants averaged 46 years old and were more highly educated than the other 2 clinics. Clinic 2 participants were younger on average than the other 2 clinics, averaging 32 years old; these participants completed lower levels of education than the other 2 clinics. Site B, Clinic 3 participants averaged 41 years of age and varied across income and education (Table 2).

All patient participants, from both sites and the 3 clinics, had received an abnormal Papanicolaou test or positive HPV test result and were referred for colposcopy in the past year. Half were referred for their first colposcopy. All had smartphones and reported using their cell phones for multiple activities, like talking, texting, making appointments, using health and fitness and other apps, and using social media. While all reported text messaging as one of the main activities they did with their phone, their estimated phone time varied widely.

Table 1. Obstetrics and gynecology (OB-GYN) and colposcopy clinic staff members' demographics by participating sites.

Characteristics	Site A		Site B	Total (N=22)
	Clinic 1 (n=8), n (%)	Clinic 2 (n=5), n (%)	Clinic 3 (n=9), n (%)	
Race and ethnicity, n (%)				
Non-Hispanic Black	0 (0)	2 (40.0)	2 (22.2)	4 (18.2)
Non-Hispanic White	5 (62.5)	2 (40.0)	6 (66.7)	13 (59.1)
Non-Hispanic Asian	2 (25.0)	0 (0)	0 (0)	2 (9.1)
Hispanic	1 (12.5)	1 (20.0)	0 (0)	2 (9.1)
More than one race	0 (0)	0 (0)	1 (10.1)	1 (4.5)
Job title, n (%)				
Physician	5 (62.5)	3 (60.0)	7 (77.8)	15 (68.2)
Nurse	1 (12.5)	0 (0)	2 (22.2)	3 (13.6)
Advanced practice provider	1 (12.5)	1 (20.0)	0 (0)	2 (9.1)
Patient support	1 (12.5)	1 (20.0)	0 (0)	2 (9.1)

XSL•FO
RenderX

Table 2. Demographic characteristics for patients referred for colposcopy at participating sites.

Characteristics	Site A		Site B	Total (N=34)
	Clinic 1 (n=9)	Clinic 2 (n=6)	Clinic 3 (n=19)	
Age (mean years, SD)	46.11 (15.01)	31.67 (7.39)	41.32 (10.06)	40.88 (11.88)
Race and ethnicity, n (%)				
Non-Hispanic Black	3 (33.3)	4 (66.6)	14 (73.7)	21 (61.8)
Non-Hispanic White	2 (22.2)	1 (16.7)	2 (10.5)	5 (14.7)
Non-Hispanic Asian	2 (22.2)	0 (0)	0 (0)	2 (5.9)
Hispanic	1 (11.1)	1 (16.7)	3 (18.8)	5 (14.7)
More than one race	1 (11.1)	0 (0)	0 (0)	1 (2.9)
Marital status, n (%)				
Single or never married	2 (22.2)	4 (66.6)	13 (68.4)	19 (55.9)
Married or living with partner	5 (55.6)	1 (16.7)	2 (10.5)	8 (23.5)
Divorced or separated or widowed	2 (22.2)	1 (16.7)	4 (21.1)	7 (20.6)
Employment status, n (%)				
Employed full-time	4 (44.4)	3 (50.0)	9 (47.4)	16 (47.0)
Employed part-time	0 (0)	1 (16.7)	4 (21.0)	5 (14.7)
Unemployed	1 (11.1)	2 (33.3)	4 (21.0)	7 (20.6)
Disabled	2 (22.2)	0 (0)	2 (10.5)	4 (11.8)
Retired	2 (22.2)	0 (0)	0 (0)	2 (5.9)
Annual income, n (%)				
US $15,000	0 (0)	0 (0)	6 (31.6)	6 (17.6)
US $15,001-US $30,000	0 (0)	4 (66.6)	6 (31.6)	10 (29.5)
>US $30,000	6 (66.7)	2 (33.3)	7 (36.8)	15 (44.1)
Don’t know or refused	3 (33.3)	0 (0)	0 (0)	3 (8.8)
Educational attainment, n (%)				
Less than a high-school diploma	0 (0)	0 (0)	2 (10.6)	2 (5.9)
High-school diploma or equivalent	0 (0)	5 (83.3)	8 (42.1)	13 (38.2)
Some college	0 (0)	0 (0)	4 (21.0)	4 (11.8)
Vocational school	0 (0)	0 (0)	3 (15.8)	3 (8.8)
Bachelor’s degree or higher	9 (100)	1 (16.7)	2 (10.6)	12 (35.3)
Follow-up appointment is for patient’s first abnormal cervical test, n (%)	5 (55.6)	4 (66.6)	8 (42.1)	17 (50.0)

Qualitative Findings

Below, we highlight the critical feedback participants gave regarding the text-based intervention. Clinicians and patients, despite patients’ demographic differences across clinics, were remarkably similar in their perspectives on the barriers they thought women encounter in attending their colposcopy appointments and in their thoughts on our proposed intervention concept, including the types of messages we should include. They had mixed preferences for the mode by which they thought counseling and informational content should be delivered. We report these themes and nuances below.

Address Variable Colposcopy Attendance Barriers

Clinic staff members and patients posited that women who do not attend their colposcopy appointments experience variable socioemotional and structural barriers (Table 3). They thought fear of having a biopsy or being afraid of receiving cancerous results prevented some women from attending appointments but might compel other women to follow up. Many clinic staff members stated that they thought patients would attend appointments if they understood the importance of the procedure in preventing cervical cancer. However, most of the patients we interviewed attended their appointments without feeling knowledgeable or prepared; they explained that they attended appointments to lessen their fear and take care of their health. Patients and clinic staff members acknowledged that structural barriers were difficult for many women to overcome, especially

XSL•FO
RenderX

inflexible or hourly work and lack of childcare, transportation, or inadequate health insurance. They noted that clinics not having evening or weekend appointment hours and needing patients to schedule appointments several months ahead likely increased the potential for patients not attending appointments. In addition, clinic staff members admitted their scheduling systems for making appointments were unnecessarily cumbersome, especially when patients needed to call back after receiving screening results rather than being able to schedule their colposcopy appointment during the same call. Patients thought being able to schedule through an automated online system, like a portal or website, would be more efficient and easier than having to call a scheduler, especially because they could use automated systems after clinic hours.

Table 3. Clinic and patient perspectives on common colposcopy attendance barriers.

Theme	Clinicians	Patients
Fear or avoidance	*Sometimes people are scared of the biopsies and sometimes they don't think that it's [the colposcopy] necessary*. [Site A, Clinic 1, Clinician, non-Hispanic Asian, 43 years old]	*They're fearful that it could be cancer, so that's why some people say, oh, just let it be. . . . They don't want to discover that they have positive results.* [Site A, Clinic 1, non-Hispanic Asian, 66 years old]
Structural barriers	*They face a lot of challenges and a lot of struggles, just being able to get to their appointments . . . . They're making sure the appointment is going to be covered with their insurance. . . or they don't have insurance. . . . being able to take off of work to come to clinic.* [Site B, Clinic 3, Clinician, non-Hispanic White, age not assessed]	*The childcare, because I'm a single mother and then I can't schedule it [at certain times]. . . . [In addition,] it takes me an hour to drive to the hospital.* [Site A, Clinic 2, non-Hispanic White, 28 years old]
Scheduling process	*Part of it is that our schedules book out a little bit far. Our colposcopy services, they're only offered at certain time of the day… scheduling can be important to our patients, especially ones who work non-traditional schedules.* [Site B, Clinic 3, Clinician, non-Hispanic Black, age not assessed]	*I was literally on the phone with her for ten minutes . . . between my work schedule and then you can't go there while you're having your period. And then I was also going on vacation. So really? I think we bounced between – I don't know – 10 different dates*. [Site A, Clinic 1, non-Hispanic Black, 40 years old]

Provide Practical Health Information and Supportive Messaging

All participants expressed a need for an intervention to provide pragmatic, educational, and counseling messages during the time between receipt of abnormal cervical screening results and colposcopy appointments to help patients overcome socioemotional and knowledge-based barriers to attendance. They thought sending appointment reminders through text was essential. They also suggested that sharing information with patients that emphasized prevention, prepared patients for the procedure, and reduced anxiety would be key to improving colposcopy attendance (Table 4).

Table 4. Clinic and patient perspectives on the types of messages intervention should include.

Theme	Clinic staff members	Patients
Provide patient appointment reminders	*I think our whole patient-reminder system, which is just a phone call with a message left, is probably not the most effective. . . . Text messages are better. . . . 99 percent of the other appointments I go to, I get a text message that says, "reply C to confirm.* [Site A, Clinic 1, Clinician, non-Hispanic White, 48 years old]	*It's sometimes unclear about when to follow up or if it's necessary . . . what I should do when I leave the doctor. . . . I think it helps just to – yeah, just like as a reminder and knowing more about what's going on with your appointments.* [Site A, Clinic 1, Biracial, non-Hispanic Black and White, 27 years old]
Emphasize prevention	*But even if there was a short little thing … would help them to understand why they need to go because I find that so many of our patients . . . they, a lot of times, are like, "Oh, I don't know. I just got a call that I have this appointment.* [Site B, Clinician, Biracial, non-Hispanic White and Asian, age not assessed]	*I don't know whether "cancer" will scare people off [but] maybe if you have an abnormal Pap smear and then you need a colposcopy maybe you need to tell them. You need this because this is a test that allows us to detect whether you have the beginnings, right.* [Site A, Clinic 1, non-Hispanic Black, 63 years old]
Reinforce information and prepare patients	*[A]ny reinforcement is excellent. . . . [A]s soon as you leave the office, you realize that you've got a question that you wanted to ask. And this may be something that helps.* [Site A, Clinic 2, Clinician, non-Hispanic Asian, 58 years old]	*I had to look everything up on my own. . . . [S]ome people might not even know where to start, or they may be too nervous to look it up. So if it was right there or provided, it may be a little easier.* [Site B, Clinic 3, non-Hispanic Black, 32 years old]
Reduce anxiety and provide support	*They've been anxious about all of these questions at home [and] these [texts] can relieve that. . . . This is exactly what we counsel them in the clinic….* [Site A, Clinic 1, Clinician, non-Hispanic Asian, 38 years old]	*I was so afraid of that appointment. So having a way to calm my nerves, to make me feel all right about the procedure, that is 100% for me.* [Site B, Clinic 3, Hispanic, 44 years old]

XSL•FO
RenderX

Provide Patient Appointment Reminders

When first asked about a potential text message intervention, most participants, clinic staff members, and patients assumed that the texts would be reminder texts. While clinics used patient portals, none had a text messaging system and relied on phone calls and letters to patients that clinic staff felt were "not the most effective." Clinic staff members thought text reminders could alert patients to schedule childcare, find transportation or make other arrangements that would help enable them to attend their appointments. In addition, some patients noted that reminders would be helpful for them to track follow-up needs. Some said they were confused about when, if, and what kind of follow-up appointment they needed. A text could help know "when to follow-up or if it's necessary . . . what I should do when I leave the doctor. . . ."

Emphasize That Cervical Cancer Can Be Prevented

Clinic staff members and patients also agreed that focusing on prevention and early detection was critical to increasing colposcopy attendance. Multiple participants noted that receiving text messages that explained why colposcopy was needed and how the procedure would prevent cancer–"help them understand 'why' they need to go"–would increase the likelihood of making an appointment. Clinic staff members and patients expressed that the texts needed to strike a delicate balance, presenting an adequate amount of information regarding cancer risk without instilling fear or raising new concerns; as 1 patient said, "I don't know whether [using the word] 'cancer' will scare people off…but maybe you need to tell them." Clinic staff members also suggested that messages could be an opportunity to promote HPV vaccination for cancer prevention and should be included in the intervention.

Reinforce Information Given by the Clinic and Prepare Patients for the Procedure

Clinic staff members said that the proposed messaging aligned with patients' frequently asked questions and what they tell patients. They thought redelivering this information was important, as they believed patients did not retain information about colposcopy when receiving their abnormal screening results: "any reinforcement is excellent." Patients said that texts about what to expect at the visit, including how long the procedure would take, if they would experience pain, and if they would need assistance post procedure, would help patients better prepare for the visit. Clinic staff members and patients thought including postcolposcopy instructions, such as how long to avoid sex if a biopsy is taken and the potential for light spotting to occur, would be helpful in preparing patients. Some patients wanted to know what follow-up plans could entail; however, some clinic staff members thought the purpose of the texts should be to encourage women to attend their immediate appointment and thought that projecting too far into the future would be too complicated and case-specific to capture in a brief text.

Reduce Anxiety and Provide Support

Patients reported not receiving enough information when they received their abnormal screening results and said that having more information would have eased some of the anxiety they felt while waiting for their colposcopy. They wanted information from a trustworthy source, such as their health care provider's office, instead of having to search online, and they wanted detailed information on types of HPV, HPV transmission, and potential consequences of their abnormal screening results. Younger women wanted information on fertility, especially. Clinic staff members acknowledged that patients have anxiety–"they've been anxious about all of these questions at home"–and that the texts may help address that anxiety: "this is exactly what we counsel them in the clinic." However, while some clinic staff members thought explaining some of these topics over text might be too complex or not generalizable to everyone receiving the texts, they agreed that explaining how the HPV virus may lie dormant for years and therefore may not be a direct result of a recent infidelity would help reduce the stigma of sexually transmitted infections. Some patients thought the intervention should include messages offering encouraging words or affirmations, which would help reduce stress: "I was so afraid… having a way to calm my nerves… that is 100% for me."

Enhance Text Messages With Additional Forms of Information or Support

While clinic staff members and patients responded positively to a patient-facing intervention that provided reminders, health information, and supportive messages, they had mixed feelings about a text-only intervention. Many in our sample preferred texts because they wanted "very direct," "short, quick pieces of information"; they thought texts would help patients "know quickly what [they] needed to do" (Table 5). Others liked the idea of texts, but suggested texts be supplemented with infographics and diagrams or links to credible websites where more detailed information could be found. Both clinic staff members and patients thought texts should provide a phone number to a clinic contact in the event a patient had additional questions and suggested linking patients to additional counseling services or a support group should patients experience elevated concern or distress. Some clinic staff members noted that patients could call the clinic if they needed help with insurance or transportation services, as well; some clinics have personnel like social workers, patient navigators, or front desk staff members who can work with patients to help them find a way to attend their appointments.

Some participants raised concerns about text messages in terms of fatigue and privacy. Some patients wanted the ability to opt in or out of threads, and a few preferred to have information sent by email or accessible through an app so they could peruse the intervention information in one place at one time at their convenience. Younger patients worried that someone else might see a message notification containing sensitive information on their phone screen. Clinic staff members worried about HIPAA (Health Insurance Portability and Accountability Act) violations or privacy and access issues for patients who shared phones with relatives. Some staff members thought older patients would not be as receptive to texts as younger patients, but we found no evidence for this from older patients in our sample.

XSL•FO
RenderX

Table 5. Clinic and patient preferences for mode of message delivery.

Theme	Clinicians	Patients
Text messages	*I think texting is a really good way to get short, quick pieces of information communicated.* [Site A, Clinic 1, Clinician, Non-Hispanic White, 48 years old]	*Texts tend to be shorter so I guess I would choose text. It's easier to read and then if I needed to do something, the message would be very direct so I would know quickly what I needed to do.* [Site B, Clinic 3, Non-Hispanic White, 37 years old]
Texts plus links to additional information	*Maybe text message ahead a link that you could click for information or also if you had any questions.* [Site A, Clinic 2, Support Staffs, Non-Hispanic White, 46 years old]	*Oh, that would be great, reliable information. . . . If there's a link that can explain to you what's happening and the reason why it's happening, that would be so helpful.* [Site B, Clinic 3, Hispanic, 44 years old]
Texts plus a person who can counsel	*I think certain things might need clarification. I just find that when I call patients . . . I don't know how that would translate through text. And I also wonder if it would feel a little impersonal as a patient to get like counseling through a text message.* [Site A, Clinic 1, Support Staff, Hispanic, 31 years old]	*There's some situations that text messages is not going to work. . . . [If] you need some information, I can send you something on text messages. You can read about it. [But] If you have more doubts and you feel like you need some support, I have a counselor that can assist you. . . .* [Site A, Clinic 1, Hispanic, 33 years old]

Discussion

Principal Results

We conducted this study to elicit feedback from clinic staff members and patients with colposcopy experience. Despite differences in our sample between clinic locations, participant roles, and patient age and educational characteristics, participants overwhelmingly agreed that an intervention providing colposcopy appointment reminders, abnormal screening results and colposcopy procedure information, and psychosocial and structural barriers counseling is needed for patients awaiting colposcopy appointments. As noted in the literature and echoed throughout this study, patients may experience varied and often multiple barriers to attendance [30]. In our adapted SMS text message intervention, we proposed to assess and address C-SHIP–related [62] barriers by providing information and counseling texts related to barriers patients endorse. This intervention has the potential to address many of the patient needs participants discussed, which align with C-SHIP constructs of knowledge and perception, expectancies and beliefs, and interfering affect. In addition, although there was interest in a text message intervention, participants expressed that texts alone may not be sufficient for or desired by all patients. Participants conveyed that patients who want additional information or who need additional support in overcoming barriers should be linked to counselors or other resources to enhance the intervention's potential to increase patient attendance.

mHealth strategies have been successful in improving prevention and health education among underserved minoritized populations [65], the population experiencing the greatest cervical cancer disparities. We heard from clinic staff members that health education is critical for follow-up attendance [66]. Our proposed text intervention is foundational for activated patients [67] who need basic information to assist in their decision-making about their colposcopy appointment. In our previous study, patients described being confused about the meaning of their screening results and having limited knowledge of HPV risk and colposcopy; seeking information on their own often increased anxiety and confusion [36]. In this study, clinic staff members reported that patients often do not absorb the information provided at the time of abnormal screening results delivery. Both patients and clinic staff members thought that messaging to reinforce the commonality of abnormal results but also the importance of colposcopy in cervical cancer prevention would be critical for patients seeking information to enhance their decision-making. Providing information of this kind may help reduce anxiety and increase knowledge, leading to increased colposcopy attendance for these types of patients. However, similar to other studies noting privacy concerns [68], the intervention messages will not contain patient information about their specific screening results.

Our findings also suggested a need for additional communication through other delivery modes for other kinds of patients. Texts may be limited in their ability to engage and support all patients and address all relevant topics. Participants acknowledged that patients with more complex or numerous psychosocial [69] or structural barriers [30] likely need additional support in overcoming their attendance barriers than text messages can provide. This group of patients would benefit from being connected to a point person at the practice or a professional counselor or support group, in addition to having the text message serve as a primer. Some clinic staff members noted that they can help connect patients to local resources for structural barriers like transportation or childcare or help navigate charity care or insurance needs. However, as staffing resources vary by clinic, this solution may not be scalable.

Importantly, the barrier experienced most by the patients we interviewed was scheduling, which is beyond the capacity of our proposed intervention. Work and childcare conflicts make scheduling difficult but having to schedule appointments months out aggravates these and other challenges especially pertinent to patients who have limited capacity to foresee or control their futures [18,70]. Clinic scheduling systems themselves were cumbersome for some patients. Including automated scheduling in conjunction with text reminders may help patients attend their follow-up appointments by reducing barriers to scheduling [71]. Patient-centered scheduling systems are essential structural elements that health systems can incorporate to improve

XSL•FO
RenderX

appointment attendance in general and colposcopy appointments specifically.

Future Research

Due to the strength and commonality of perspectives across clinic staff members and patient groups, and additional feasibility testing [61], the team expanded the intervention design to include brief health coaching sessions over the phone. Designing with efficiency and scalability in mind [72,73], this additional intervention component will be reserved for patients who experience complex barriers to attendance and missed appointments. Health coaching interventions, which are patient-centered and goal-focused, have improved health outcomes by empowering patients to self-manage their health or illness across multiple conditions and contexts [74-77]. The purpose of these calls will be to assess barriers, guide patients to existing resources, including connections to scheduling and structural barrier supports, and help them make a plan to attend the colposcopy appointment. The multimodal intervention will be tested this year.

Limitations

This study had several limitations to consider. As mentioned above, we interviewed patients referred for colposcopy, many of whom had already attended a colposcopy appointment. The patients and staff members we interviewed spoke about others they knew and could imagine scenarios in which they themselves would experience such barriers [42,60,78]. The small number of patients we recruited who did not attend their appointments (n=5) noted similar barriers and provided similar feedback on how to improve the intervention. We anticipate health coaching professionals may help this group address attendance barriers. Also, we interviewed participants from 2 urban academic OB-GYN sites that have support services like care coordinators for their patients. As this was a purposive sample to gather perspectives on our proposed intervention, we found these perspectives helpful. However, our results may not be generalizable to clinics and practices with lower attendance rates and fewer supports for patients.

Conclusions

Clinic staff members and patient perspectives are critical for designing appropriate and relevant interventions. These groups conveyed that text message-only interventions are useful for some patients but not for others. Multimodal interventions may be necessary for patients with complex barriers to colposcopy attendance, while patients with fewer barriers may benefit from reminders, basic educational information, and scheduling support.

Acknowledgments

This work was supported by a grant (R01 MD017675-01) and a Cancer Health Equity Pilot Award (REK and SVH) through Rutgers Cancer Institute of New Jersey, which is supported in part by an NIH/NCI Cancer Center Support Grant (P30CA072720). SVH was supported by the National Center for Advancing Translational Sciences (UL1TR003017). SMM was supported by the National Institutes of Health's National Cancer Institute (NIH/NCI) grant (P30CA006927) and the Marcia and Frank Carlucci Charitable Foundation Award (Prevent Cancer Foundation). REK was supported by the National Cancer Institute (K22CA258675). We would like to thank the patients who took part in this research as well as the staff at the OB-GYN practices, without whom this work would not have been possible. We would like to thank Nicole Hernandez for her work interviewing participants in Spanish.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sample Interview Questions from Patient and Healthcare Staff Interview Guides.
[DOCX File , 42 KB - formative_v9i1e55043_app1.docx]

References

1. Adegoke O, Kulasingam S, Virnig B. Cervical cancer trends in the United States: a 35-year population-based analysis. J Womens Health (Larchmt) 2012 Oct;21(10):1031-1037 [FREE Full text] [doi: 10.1089/jwh.2011.3385] [Medline: 22816437]
2. Mix JM, Van Dyne EA, Saraiya M, Hallowell BD, Thomas CC. Assessing impact of HPV vaccination on cervical cancer incidence among women aged 15-29 years in the United States, 1999-2017: an ecologic study. Cancer Epidemiol Biomarkers Prev 2021 Jan;30(1):30-37 [FREE Full text] [doi: 10.1158/1055-9965.EPI-20-0846] [Medline: 33082207]
3. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin 2023 Jan;73(1):17-48 [FREE Full text] [doi: 10.3322/caac.21763] [Medline: 36633525]
4. Roche LM, Niu X, Henry KA. Invasive cervical cancer incidence disparities in New Jersey--a spatial analysis in a high incidence state. J Health Care Poor Underserved 2015 Nov;26(4):1173-1185. [doi: 10.1353/hpu.2015.0127] [Medline: 26548671]
5. SEER*Explorer: an interactive website for SEER cancer statistics [Internet].: Surveillance Research Program, National Cancer Institute; 2024 Apr 17. URL: https://seer.cancer.gov/statistics-network/explorer/ [accessed 2024-09-06]

XSL•FO
RenderX

6. Beavis AL, Gravitt PE, Rositch AF. Hysterectomy-corrected cervical cancer mortality rates reveal a larger racial disparity in the United States. Cancer 2017 May 15;123(6):1044-1050 [FREE Full text] [doi: 10.1002/cncr.30507] [Medline: 28112816]
7. Benard VB, Lawson HW, Eheman CR, Anderson C, Helsel W. Adherence to guidelines for follow-up of low-grade cytologic abnormalities among medically underserved women. Obstet Gynecol 2005 Jun;105(6):1323-1328. [doi: 10.1097/01.AOG.0000159549.56601.75] [Medline: 15932824]
8. Benard VB, Jackson JE, Greek A, Senkomago V, Huh WK, Thomas CC, et al. A population study of screening history and diagnostic outcomes of women with invasive cervical cancer. Cancer Med 2021 Jun;10(12):4127-4137 [FREE Full text] [doi: 10.1002/cam4.3951] [Medline: 34018674]
9. Chase DM, Osann K, Sepina N, Wenzel L, Tewari KS. The challenge of follow-up in a low-income colposcopy clinic: characteristics associated with noncompliance in high-risk populations. J Low Genit Tract Dis 2012 Oct;16(4):345-351 [FREE Full text] [doi: 10.1097/LGT.0b013e318249640f] [Medline: 22622340]
10. Gaffney DK, Hashibe M, Kepka D, Maurer KA, Werner TL. Too many women are dying from cervix cancer: Problems and solutions. Gynecol Oncol 2018;151(3):547-554 [FREE Full text] [doi: 10.1016/j.ygyno.2018.10.004] [Medline: 30301561]
11. Olusola P, Banerjee HN, Philley JV, Dasgupta S. Human papilloma virus-associated cervical cancer and health disparities. Cells 2019;8(6):622 [FREE Full text] [doi: 10.3390/cells8060622] [Medline: 31234354]
12. Krieger N, Wright E, Chen JT, Waterman PD, Huntley ER, Arcaya M. Cancer stage at diagnosis, historical redlining, and current neighborhood characteristics: breast, cervical, lung, and colorectal cancers, Massachusetts, 2001-2015. Am J Epidemiol 2020;189(10):1065-1075 [FREE Full text] [doi: 10.1093/aje/kwaa045] [Medline: 32219369]
13. Sokale IO, Oluyomi AO, Montealegre JR, Thrift AP. Racial/Ethnic disparities in cervical cancer stage at diagnosis: mediating effects of neighborhood-level socioeconomic deprivation. Cancer Epidemiol Biomarkers Prev 2023;32(6):818-824 [FREE Full text] [doi: 10.1158/1055-9965.EPI-23-0038] [Medline: 37067295]
14. Suk R, Hong YR, Rajan SS, Xie Z, Zhu Y, Spencer JC. Assessment of US preventive services task force guideline-concordant cervical cancer screening rates and reasons for underscreening by age, race and Ethnicity, sexual orientation, rurality, and insurance, 2005 to 2019. JAMA Netw Open 2022;5(1):e2143582 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.43582] [Medline: 35040970]
15. Moss JL, Geyer NR, Lengerich EJ. Patterns of cancer-related healthcare access across pennsylvania: analysis of novel census tract-level indicators of persistent poverty. Cancer Epidemiol Biomarkers Prev 2024;33(4):616-623. [doi: 10.1158/1055-9965.EPI-23-1255] [Medline: 38329390]
16. McDaniel CC, Hallam HH, Cadwallader T, Lee HY, Chou C. Persistent racial disparities in cervical cancer screening with Pap test. Prev Med Rep 2021 Dec;24:101652 [FREE Full text] [doi: 10.1016/j.pmedr.2021.101652] [Medline: 34976700]
17. Sabatino SA, Thompson TD, White MC, Villarroel MA, Shapiro JA, Croswell JM, et al. Up-to-date breast, cervical, and colorectal cancer screening test use in the United States, 2021. Prev Chronic Dis 2023 Oct 26;20:E94 [FREE Full text] [doi: 10.5888/pcd20.230071] [Medline: 37884318]
18. Fuzzell LN, Perkins RB, Christy SM, Lake PW, Vadaparampil ST. Cervical cancer screening in the United States: challenges and potential solutions for underscreened groups. Prev Med 2021 Mar;144:106400. [doi: 10.1016/j.ypmed.2020.106400] [Medline: 33388330]
19. Ford S, Tarraf W, Williams KP, Roman LA, Leach R. Differences in cervical cancer screening and follow-up for black and white women in the United States. Gynecol Oncol 2021 Feb;160(2):369-374. [doi: 10.1016/j.ygyno.2020.11.027] [Medline: 33323276]
20. Boitano TKL, Ketch P, Maier JG, Nguyen CT, Huh WK, Michael Straughn J, et al. Increased disparities associated with black women and abnormal cervical cancer screening follow-up. Gynecol Oncol Rep 2022 Aug;42:101041 [FREE Full text] [doi: 10.1016/j.gore.2022.101041] [Medline: 35898199]
21. Spencer JC, Kim JJ, Tiro JA, Feldman SJ, Kobrin SC, Skinner CS, et al. Racial and ethnic disparities in cervical cancer screening from three U.S. healthcare settings. Am J Prev Med 2023 Oct;65(4):667-677 [FREE Full text] [doi: 10.1016/j.amepre.2023.04.016] [Medline: 37146839]
22. National Institutes of Health. National Institutes of Health Consensus Development Conference Statement: cervical cancer, April 1-3, 1996. National Institutes of Health Consensus Development Panel. J Natl Cancer Inst Monogr 1996(21):vii-xix. [Medline: 9023821]
23. Leyden WA, Manos MM, Geiger AM, Weinmann S, Mouchawar J, Bischoff K, et al. Cervical cancer in women with comprehensive health care access: attributable factors in the screening process. J Natl Cancer Inst 2005 May 04;97(9):675-683. [doi: 10.1093/jnci/dji115] [Medline: 15870438]
24. Landy R, Pesola F, Castañón A, Sasieni P. Impact of cervical screening on cervical cancer mortality: estimation using stage-specific results from a nested case-control study. Br J Cancer 2016 Oct 25;115(9):1140-1146 [FREE Full text] [doi: 10.1038/bjc.2016.290] [Medline: 27632376]
25. Rosenblum HG, Gargano JW, Cleveland AA, Dahl RM, Park IU, Whitney E, et al. U.S. women with invasive cervical cancer: characteristics and potential barriers to prevention. J Womens Health (Larchmt) 2024 May;33(5):594-603. [doi: 10.1089/jwh.2023.0462] [Medline: 38608239]

XSL•FO
RenderX

26. Holt HK, Peterson CE, MacLaughlan David S, Abdelaziz A, Sawaya GF, Guadamuz JS, et al. Mediation of racial and ethnic inequities in the diagnosis of advanced-stage cervical cancer by insurance status. JAMA Netw Open 2023 Mar 01;6(3):e232985 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.2985] [Medline: 36897588]
27. Subramaniam A, Fauci JM, Schneider KE, Whitworth JM, Erickson BK, Kim K, et al. Invasive cervical cancer and screening: what are the rates of unscreened and underscreened women in the modern era? J Low Genit Tract Dis 2011 Apr;15(2):110-113 [FREE Full text] [doi: 10.1097/LGT.0b013e3181f515a2] [Medline: 21263352]
28. Kulkarni A, Glynn S, Gamble CR, Shen MJ, Cantillo E, Frey MK, et al. Understanding perceived barriers to colposcopy follow-up among underserved women at an urban teaching hospital: a qualitative study. J Low Genit Tract Dis 2023;27(1):87-92. [doi: 10.1097/LGT.0000000000000700] [Medline: 36074132]
29. Hui SKA, Miller SM, Wen KY, Fang Z, Li T, Buzaglo J, et al. Psychosocial barriers to follow-up adherence after an abnormal cervical cytology test result among low-income, inner-city women. J Prim Care Community Health 2014;5(4):234-241 [FREE Full text] [doi: 10.1177/2150131914529307] [Medline: 24718518]
30. Glynn S, Cantillo E, Vasquez K, Shen M, Litvin D, Frey M, et al. Understanding knowledge gaps and perceived barriers to follow up after abnormal Pap smear among low-income women in NYC: a qualitative study. Gynecol Oncol 2021;162(1):S320-S321 [FREE Full text] [doi: 10.1016/S0090-8258(21)01259-2]
31. Nelson K, Geiger AM, Mangione CM. Effect of health beliefs on delays in care for abnormal cervical cytology in a multi-ethnic population. J Gen Intern Med 2002;17(9):709-716 [FREE Full text] [doi: 10.1046/j.1525-1497.2002.11231.x] [Medline: 12220368]
32. Nolan J, Renderos TB, Hynson J, Dai X, Chow W, Christie A, et al. Barriers to cervical cancer screening and follow-up care among black women in Massachusetts. J Obstet Gynecol Neonatal Nurs 2014;43(5):580-588. [doi: 10.1111/1552-6909.12488] [Medline: 25139164]
33. Percac-Lima S, Aldrich LS, Gamba GB, Bearse AM, Atlas SJ. Barriers to follow-up of an abnormal Pap smear in latina women referred for colposcopy. J Gen Intern Med 2010;25(11):1198-1204 [FREE Full text] [doi: 10.1007/s11606-010-1450-6] [Medline: 20652647]
34. Radecki Breitkopf C, Pearson HC. A theory-based approach to understanding follow-up of abnormal Pap tests. J Health Psychol 2009;14(3):361-371 [FREE Full text] [doi: 10.1177/1359105308101674] [Medline: 19293297]
35. Head KJ, Imburgia TM, Zimet GD, Shew ML. Women's understanding of their Pap and HPV test results: implications for patient–provider communication. J Healthc Commun 2017;10(1):37-46. [doi: 10.1080/17538068.2017.1282085]
36. Kohler RE, Hemler J, Wagner RB, Sullivan B, Macenat M, Tagai EK, et al. Confusion and anxiety in between abnormal cervical cancer screening results and colposcopy: "The land of the unknown". Patient Educ Couns 2023;114:107810. [doi: 10.1016/j.pec.2023.107810] [Medline: 37244133]
37. Spees LP, Brewster WR, Varia MA, Weinberger M, Baggett C, Zhou X, et al. Examining urban and rural differences in how distance to care influences the initiation and completion of treatment among insured cervical cancer patients. Cancer Epidemiol Biomarkers Prev 2019;28(5):882-889 [FREE Full text] [doi: 10.1158/1055-9965.EPI-18-0945] [Medline: 30733307]
38. Lim JW, Ashing-Giwa KT. Examining the effect of minority status and neighborhood characteristics on cervical cancer survival outcomes. Gynecol Oncol 2011;121(1):87-93. [doi: 10.1016/j.ygyno.2010.11.041] [Medline: 21183210]
39. Plascak JJ, Llanos AA, Pennell ML, Weier RC, Paskett ED. Neighborhood factors associated with time to resolution following an abnormal breast or cervical cancer screening test. Cancer Epidemiol Biomarkers Prev 2014;23(12):2819-2828 [FREE Full text] [doi: 10.1158/1055-9965.EPI-14-0348] [Medline: 25205516]
40. Glick SB, Clarke AR, Blanchard A, Whitaker AK. Cervical cancer screening, diagnosis and treatment interventions for racial and ethnic minorities: a systematic review. J Gen Intern Med 2012;27(8):1016-1032 [FREE Full text] [doi: 10.1007/s11606-012-2052-2] [Medline: 22798213]
41. Paskett ED, Dudley D, Young GS, Bernardo BM, Wells KJ, Calhoun EA, PNRP Investigators. Impact of patient navigation interventions on timely diagnostic follow up for abnormal cervical screening. J Womens Health (Larchmt) 2016;25(1):15-21 [FREE Full text] [doi: 10.1089/jwh.2014.5094] [Medline: 26625131]
42. Miller SM, Hui SA, Wen KY, Scarpato J, Zhu F, Buzaglo J, et al. Tailored telephone counseling to improve adherence to follow-up regimens after an abnormal pap smear among minority, underserved women. Patient Educ Couns 2013;93(3):488-495 [FREE Full text] [doi: 10.1016/j.pec.2013.08.008] [Medline: 24007767]
43. Miller SM, Siejak KK, Schroeder CM, Lerman C, Hernandez E, Helm CW. Enhancing adherence following abnormal Pap smears among low-income minority women: a preventive telephone counseling strategy. J Natl Cancer Inst 1997;89(10):703-708. [doi: 10.1093/jnci/89.10.703] [Medline: 9168185]
44. Lerman C, Hanjani P, Caputo C, Miller S, Delmoor E, Nolte S, et al. Telephone counseling improves adherence to colposcopy among lower-income minority women. J Clin Oncol 1992;10(2):330-333. [doi: 10.1200/JCO.1992.10.2.330] [Medline: 1732434]
45. Yang Q, Van Stee SK. The comparative effectiveness of mobile phone interventions in improving health outcomes: meta-analytic review. JMIR Mhealth Uhealth 2019;7(4):e11244 [FREE Full text] [doi: 10.2196/11244] [Medline: 30942695]

XSL•FO
RenderX

46. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
47. Bhochhibhoya S, Dobbs PD, Maness SB. Interventions using mHealth strategies to improve screening rates of cervical cancer: a scoping review. Prev Med 2021;143:106387. [doi: 10.1016/j.ypmed.2020.106387] [Medline: 33383069]
48. Lee HY, Koopmeiners JS, Rhee TG, Raveis VH, Ahluwalia JS. Mobile phone text messaging intervention for cervical cancer screening: changes in knowledge and behavior pre-post intervention. J Med Internet Res 2014;16(8):e196 [FREE Full text] [doi: 10.2196/jmir.3576] [Medline: 25164545]
49. Thiel de Bocanegra H, Dehlendorf C, Kuppermann M, Vangala SS, Moscicki AB. Impact of an educational tool on young women's knowledge of cervical cancer screening recommendations. Cancer Causes Control 2022;33(6):813-821 [FREE Full text] [doi: 10.1007/s10552-022-01569-8] [Medline: 35312891]
50. Romocki LS, Des Marais A, Cofie L, Anderson C, Curington T, Smith JS. Use of short messaging service to improve follow-up for abnormal pap test results in minority and medically underserved women in north carolina: questionnaire on attitudes and acceptability. JMIR Form Res 2019;3(3):e12675 [FREE Full text] [doi: 10.2196/12675] [Medline: 31389334]
51. Lee YH, Huang LH, Chen SH, Shao JH, Lai CH, Yang NP. Effects of mobile application program (App)-assisted health education on preventive behaviors and cancer literacy among women with cervical intraepithelial neoplasia. Int J Environ Res Public Health 2021;18(21):11603 [FREE Full text] [doi: 10.3390/ijerph182111603] [Medline: 34770117]
52. Arrossi S, Paolino M, Antelo VS, Thouyaret L, Kohler RE, Cuberli M, ATICA Study team. Effectiveness of an mHealth intervention to increase adherence to triage of HPV DNA positive women who have performed self-collection (the ATICA study): a hybrid type I cluster randomised effectiveness-implementation trial. Lancet Reg Health Am 2022;9:100199 [FREE Full text] [doi: 10.1016/j.lana.2022.100199] [Medline: 35655914]
53. Erwin E, Aronson KJ, Day A, Ginsburg O, Macheku G, Feksi A, et al. SMS behaviour change communication and eVoucher interventions to increase uptake of cervical cancer screening in the Kilimanjaro and Arusha regions of Tanzania: a randomised, double-blind, controlled trial of effectiveness. BMJ Innov 2019;5(1):28-34 [FREE Full text] [doi: 10.1136/bmjinnov-2018-000276] [Medline: 31645991]
54. Joseph NT, Namuli A, Kakuhikire B, Baguma C, Juliet M, Ayebare P, et al. Implementing community-based human papillomavirus self-sampling with SMS text follow-up for cervical cancer screening in rural, southwestern Uganda. J Glob Health 2021;11:04036 [FREE Full text] [doi: 10.7189/jogh.11.04036] [Medline: 35003710]
55. Linde DS, Andersen MS, Mwaiselage J, Manongi R, Kjaer SK, Rasch V. Effectiveness of one-way text messaging on attendance to follow-up cervical cancer screening among human papillomavirus-positive tanzanian women (Connected2Care): parallel-group randomized controlled trial. J Med Internet Res 2020;22(4):e15863 [FREE Full text] [doi: 10.2196/15863] [Medline: 32238335]
56. Kannisto KA, Koivunen MH, Välimäki MA. Use of mobile phone text message reminders in health care services: a narrative literature review. J Med Internet Res 2014;16(10):e222 [FREE Full text] [doi: 10.2196/jmir.3442] [Medline: 25326646]
57. Willcox JC, Dobson R, Whittaker R. Old-fashioned technology in the era of "Bling": is there a future for text messaging in health care? J Med Internet Res 2019;21(12):e16630 [FREE Full text] [doi: 10.2196/16630] [Medline: 31859678]
58. Suffoletto B. Text message behavioral interventions: from here to where? Curr Opin Psychol 2016;9:16-21 [FREE Full text] [doi: 10.1016/j.copsyc.2015.09.012] [Medline: 26665157]
59. Evidence-Based Cancer Control Programs (EBCCP) 2023. National Cancer Institute. 2023. URL: https://ebccp.cancercontrol.cancer.gov/programDetails.do?programId=26767713 [accessed 2024-09-09]
60. Sansoni M, Tagai EK, Lapitan E, Wen KY, Xu J, Belfiglio A, et al. Development of a text message-based intervention for follow-up colposcopy among predominately underserved Black and Hispanic/Latinx women. Cancer Causes Control 2022;33(6):861-873 [FREE Full text] [doi: 10.1007/s10552-022-01573-y] [Medline: 35334016]
61. Wen KY, Dayaratna S, Slamon R, Granda-Cameron C, Tagai EK, Kohler RE, et al. Chatbot-interfaced and cognitive-affective barrier-driven messages to improve colposcopy adherence after abnormal Pap test results in underserved urban women: A feasibility pilot study. Transl Behav Med 2024 Jan 11;14(1):1-12. [doi: 10.1093/tbm/ibad064] [Medline: 38014626]
62. Miller SM, Diefenbach MA. The cognitive-social health information-processing (C-SHIP) model: a theoretical framework for research in behavioral oncology. In: Technology and Methods in Behavioral Medicine. Mahwah: Lawrence Erlbaum Associates; 1998:219-244.
63. Crabtree BF, Miller WL. Doing Qualitative Research. Third Edition. Washington DC: SAGE Publications, Inc; 2023.
64. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
65. Anderson-Lewis C, Darville G, Mercado RE, Howell S, Di Maggio S. mHealth technology use and implications in historically underserved and minority populations in the United States: systematic literature review. JMIR Mhealth Uhealth 2018;6(6):e128 [FREE Full text] [doi: 10.2196/mhealth.8383] [Medline: 29914860]
66. Shokar NK, Calderon-Mora J, Molokwu J, Byrd T, Alomari A, Mallawaarachchi I, et al. Outcomes of a multicomponent culturally tailored cervical cancer screening intervention among underserved Hispanic women (De Casa en Casa). Health Promot Pract 2021;22(1):112-121. [doi: 10.1177/1524839919893309] [Medline: 31874564]

XSL•FO
RenderX

67. Miller SM, Tagai EK, Wen K, Lee M, Hui SKA, Kurtz D, et al. Predictors of adherence to follow-up recommendations after an abnormal pap smear among underserved inner-city women. Patient Educ Couns 2017;100(7):1353-1359 [FREE Full text] [doi: 10.1016/j.pec.2017.01.020] [Medline: 28190541]
68. Sanchez Antelo V, Kohler RE, Curotto M, Viswanath KV, Paolino M, Arrossi S. Developing SMS content to promote Papanicolaou triage among women who performed HPV self-collection test: qualitative study. JMIR Form Res 2020 Mar 06;4(3):e14652 [FREE Full text] [doi: 10.2196/14652] [Medline: 32032940]
69. Ilic I, Babic G, Dimitrijevic A, Sipetic Grujicic S, Ilic M. Psychological distress in women with abnormal Pap smear results attending cervical cancer screening. J Low Genit Tract Dis 2023 Oct 01;27(4):343-350. [doi: 10.1097/LGT.0000000000000761] [Medline: 37535069]
70. Atlas SJ, Tosteson ANA, Burdick TE, Wright A, Breslau ES, Dang TH, et al. Primary care practitioner perceptions on the follow-up of abnormal cancer screening test results. JAMA Netw Open 2022;5(9):e2234194 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.34194] [Medline: 36173627]
71. Feldman S, Lykken JM, Haas JS, Werner CL, Kobrin SC, Tiro JA, et al. Factors associated with timely colposcopy following an abnormal cervical cancer test result. Prev Med 2022;164:107307 [FREE Full text] [doi: 10.1016/j.ypmed.2022.107307] [Medline: 36270434]
72. Guastaferro K, Collins LM. Achieving the goals of translational science in public health intervention research: the multiphase optimization strategy (MOST). Am J Public Health 2019 Feb;109(S2):S128-S129. [doi: 10.2105/AJPH.2018.304874] [Medline: 30785800]
73. Guastaferro K, Collins LM. Optimization methods and implementation science: an opportunity for behavioral and biobehavioral interventions. Implement Res Pract 2021;2:26334895211054363 [FREE Full text] [doi: 10.1177/26334895211054363] [Medline: 37089990]
74. Hutchison AJ, Breckon JD. A review of telephone coaching services for people with long-term conditions. J Telemed Telecare 2011;17(8):451-458. [doi: 10.1258/jtt.2011.110513] [Medline: 22025743]
75. Kivelä K, Elo S, Kyngäs H, Kääriäinen M. The effects of health coaching on adult patients with chronic diseases: a systematic review. Patient Educ Couns 2014 Nov;97(2):147-157. [doi: 10.1016/j.pec.2014.07.026] [Medline: 25127667]
76. Lin WC, Chien HL, Willis G, O'Connell E, Rennie KS, Bottella HM, et al. The effect of a telephone-based health coaching disease management program on Medicaid members with chronic conditions. Med Care 2012 Jan;50(1):91-98. [doi: 10.1097/MLR.0b013e31822dcedf] [Medline: 21993059]
77. Simmons LA, Wolever RQ. Integrative health coaching and motivational interviewing: synergistic approaches to behavior change in healthcare. Glob Adv Health Med 2013 Jul;2(4):28-35 [FREE Full text] [doi: 10.7453/gahmj.2013.037] [Medline: 24416683]
78. Spaulding S. Simulation theory. In: The Routledge Handbook of Philosophy of Imagination. New York, NY: Routledge; 2016:262-273.

Abbreviations

C-SHIP: Cognitive Social Health Information Processing
COREQ: Consolidated Criteria for Reporting Qualitative Research
HIPAA: Health Insurance Portability and Accountability Act
HPV: human papillomavirus
mHealth: mobile health
OB-GYN: Obstetrics and gynecology
TC3: Tailored Communication for Cervical Cancer Risk

Edited by A Mavragani; submitted 30.11.23; peer-reviewed by S Xiong, SS Shaik, O Ezeigwe; comments to author 16.04.24; revised version received 08.06.24; accepted 04.09.24; published 14.01.25.

Please cite as:
Hemler JR, Wagner RB, Sullivan B, Macenat M, Tagai EK, Vega JL, Hernandez E, Miller SM, Wen KY, Ayers CA, Einstein MH, Hudson SV, Kohler RE
A Proposed mHealth Intervention to Address Patient Barriers to Colposcopy Attendance: Qualitative Interview Study of Clinic Staff and Patient Perspectives
JMIR Form Res 2025;9:e55043
URL: https://formative.jmir.org/2025/1/e55043
doi:10.2196/55043
PMID:39808485

XSL•FO
RenderX

©Jennifer R Hemler, Rachel B Wagner, Brittany Sullivan, Myneka Macenat, Erin K Tagai, Jazmarie L Vega, Enrique Hernandez, Suzanne M Miller, Kuang-Yi Wen, Charletta A Ayers, Mark H Einstein, Shawna V Hudson, Racquel E Kohler. Originally published in JMIR Formative Research (https://formative.jmir.org), 14.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Acceptance, Safety, and Effect Sizes in Online Dialectical Behavior Therapy for Borderline Personality Disorder: Interventional Pilot Study

Ruben Vonderlin[1,2], PhD; Tali Boritz[3,4], Prof Dr; Carola Claus[5], MSc; Büsra Senyüz[1,2], MSc; Saskia Mahalingam[1,2], MSc; Rachel Tennenhouse[4], BSc; Stefanie Lis[2,6], Prof Dr; Christian Schmahl[1,2], Prof Dr; Jürgen Margraf[5], Prof Dr; Tobias Teismann[5], Prof Dr; Nikolaus Kleindienst[1,2], Prof Dr; Shelley McMain[4,7], Prof Dr; Martin Bohus[1,2,5,8], Prof Dr

[1]Department of Psychosomatic Medicine and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[2]German Center for Mental Health (DZPG), Partner Site Mannheim - Heidelberg - Ulm, Mannheim, Germany

[3]Department of Psychology, York University, Toronto, ON, Canada

[4]Centre for Addiction and Mental Health, Borderline Personality Disorder Clinic, Toronto, ON, Canada

[5]Mental Health Research and Treatment Center, Ruhr-University Bochum, Bochum, Germany

[6]Department of Clinical Psychology, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[7]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[8]McLean Hospital, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Ruben Vonderlin, PhD
Department of Psychosomatic Medicine and Psychotherapy
Central Institute of Mental Health
Medical Faculty Mannheim, Heidelberg University
J5
Mannheim, 68159
Germany
Phone: 49 621 1703 4445
Email: ruben.vonderlin@zi-mannheim.de

Abstract

Background: The potential of telehealth psychotherapy (ie, the online delivery of treatment via a video web-based platform) is gaining increased attention. However, there is skepticism about its acceptance, safety, and efficacy for patients with high emotional and behavioral dysregulation.

Objective: This study aims to provide initial effect size estimates of symptom change from pre- to post treatment, and the acceptance and safety of telehealth dialectical behavior therapy (DBT) for individuals diagnosed with borderline personality disorder (BPD).

Methods: A total of 39 individuals meeting the *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders* [Fifth Edition]) criteria for BPD received 1 year of outpatient telehealth DBT at 3 sites in Germany and Canada. Effect size estimates were assessed using pre-post measures of BPD symptoms, dissociation, and quality of life. Safety was evaluated by analyzing suicide attempts and self-harm. Additionally, acceptance and feasibility, satisfaction with treatment, useability of the telehealth format, and the quality of the therapeutic alliance were assessed from both therapists' and patients' perspectives. All analyses were conducted on both the intention-to-treat (ITT) and according-to-protocol (ATP) samples.

Results: Analyses showed significant and large pre-post effect sizes for BPD symptoms (d=1.13 in the ITT sample and d=1.44 in the ATP sample; P<.001) and for quality of life (d=0.65 in the ITT sample and d=1.24 in the ATP sample). Dissociative symptoms showed small to nonsignificant reductions. Self-harm behaviors decreased significantly from 80% to 28% of all patients showing at least 1 self-harm behavior in the last 10 weeks (risk ratio 0.35). A high dropout rate of 38% was observed. One low-lethality suicide attempt was reported. Acceptance, feasibility, and satisfaction measures were high, although therapists reported only moderate useability of the telehealth format.

XSL•FO
RenderX

Conclusions: Telehealth DBT for BPD showed large pre-post effect sizes for BPD symptoms and quality of life. While the telehealth format appeared feasible and well-accepted, the dropout rate was relatively high. Future research should compare the efficacy of telehealth DBT with in-person formats in randomized controlled trials. Overall, telehealth DBT might offer a potentially effective alternative treatment option, enhancing treatment accessibility. However, strategies for decreasing drop-out should be considered.

Trial Registration: German Clinical Trials Register DRKS00027824; https://drks.de/search/en/trial/DRKS00027824

(JMIR Form Res 2025;9:e66181) doi:10.2196/66181

KEYWORDS

dialectical behavior therapy; borderline personality disorder; online psychotherapy; virtual psychotherapy; telehealth; personality disorders; mental; psychotherapy; online; internet; telemedicine; psychiatry; psychiatric; acceptance

Introduction

Background

Borderline personality disorder (BPD) is a severe mental disorder characterized by high instability in affect, identity, and interpersonal relationships [1,2]. The prevalence of BPD is estimated to be between 0.7% and 2.7% in community samples [3,4] and between 11% and 35.6% in clinical samples [4]. BPD is associated with significant health care use and costs, as well as profound individual suffering and impaired quality of life [5-7].

According to international guidelines [8,9], psychotherapy is the first-line treatment for BPD. However, the complexity of symptoms such as high levels of suicidality, and interpersonal difficulties present challenges for psychological treatments [10,11]. Dialectical behavior therapy (DBT) [12-14] is a specialized treatment that was developed to support therapists in dealing with the complex symptom constellations that characterize BPD. DBT includes a dynamic hierarchy of treatment foci and a large set of therapeutic strategies and skills-based interventions to support individuals in building a life worth living. DBT is a multimodal psychological treatment that consists of individual and group therapy. A comprehensive DBT treatment includes both individual and group sessions that occur weekly over a 1-year period. DBT is an empirically supported treatment that has a robust research base demonstrating its efficacy in the treatment of BPD [15].

The Potential Benefit of Telehealth DBT for BPD

While there are effective psychological treatments for BPD, DBT and other evidence-based treatments are difficult to access [16,17]. One significant barrier to accessing DBT is that given the complexity of the treatment (ie, its multiple components, length, and team-based model) the treatment is often localized in specialized treatment centers in urban settings; DBT-trained therapists and comprehensive DBT treatment programs are often not available for clients in remote or rural settings. This problem is even more pronounced in developing countries or low- and middle-income countries [18].

One solution for improving access to DBT is to expand the delivery options of the treatment to include telehealth treatment (ie, the online delivery of treatment via a video web-based platform). Propelled in large part by the COVID-19 pandemic, DBT practitioners around the world have already begun to shift to the telehealth delivery of the treatment, with many clinicians in both health care centers and in private practice currently offering telehealth DBT [19-21]. However, concerns have been raised regarding the efficacy, safety, and feasibility of telehealth-delivered psychotherapy, especially for patients with high levels of emotional and behavioral dysregulation. Although research has started to explore the benefits, barriers, and subjective experiences of clients and therapists engaged in telehealth DBT in routine care, to date there are no rigorous research trials that have examined the efficacy, safety, acceptance, and feasibility of telehealth DBT [19-29].

Challenges of Telehealth Psychotherapy for BPD

Of particular concern is the limited ability of therapists to effectively deal with crisis situations and suicidal behavior in telehealth treatments [20,25,30]. Zalewski et al [28] conducted a qualitative study that shed light on various telehealth treatment challenges with patients with BPD, including patients missing sessions, turning off their cameras prematurely, logging off unexpectedly, or engaging in distracting behaviors like smoking or drinking alcohol during telehealth group therapy sessions. They also reported that some patients appeared dissociative during group sessions. Dissociation poses a particular challenge to the safety and efficacy of telehealth treatments for BPD. Previous research has underscored how dissociative experiences can lead to neurobiological alterations and impede emotional learning, thereby hindering successful psychotherapeutic outcomes [31-33]. Ideally, a therapist noticing dissociation during a session would intervene and provide guidance on how to use distress tolerance skills to decrease or interrupt dissociation. However, implementing these skills often involves strong sensory input or physical activity, which make be harder to facilitate telehealth versus in-person. More research is needed to investigate if and how dissociation and other therapy-interfering behaviors can be managed in telehealth treatments for BPD [34].

Another critical aspect of telehealth psychotherapy is the quality of the therapeutic alliance [20]. The interpersonal difficulties that are common in BPD, such as mistrust and rejection sensitivity, can make it harder to develop a strong alliance [35]. Previous DBT research suggests that a weaker therapeutic alliance is associated with higher therapy dropout rates [36] and that alliance rupture and repair processes in the early stage of treatment differ between recovered and unrecovered patients with BPD [37]. There is an ongoing debate as to whether the telehealth setting hampers the effective establishment of a

XSL•FO
RenderX

therapeutic alliance and increases the likelihood of alliance ruptures. While some studies show comparability between telehealth and in-person treatment formats [38,39], others suggest that telehealth therapies may be less conducive to developing an alliance than in-person therapies. Despite the pivotal role of the therapeutic alliance in DBT, at this point, there is no conclusive evidence as to how well a therapeutic alliance can be established and maintained in telehealth DBT.

Previous Research on Telehealth DBT

Two scoping reviews have been conducted to document the research evidence associated with the online delivery of DBT for individuals with BPD [24,40]. van Leeuwen et al [40] conducted their review to document the research evidence regarding the efficacy and clinical use of telepsychology in DBT. The studies they included in their review were heterogenous and ranged from the online delivery of treatment (eg, telehealth defined as videoconferencing for individual and group sessions, to the use of telephone coaching between sessions, to the integration of mobile apps as stand-alone or hybrid interventions, to the integration of virtual reality elements in treatment). Overall, they included 41 studies; however, only 2 of these trials investigated telehealth DBT applied by therapists via videoconferencing systems for patients with BPD specifically [41,42]. In a second review, Lakeman et al [24] similarly reviewed the online delivery of DBT for a range of different mental disorders. They showed that attendance tended to be higher in telehealth format and clinical improvements were comparable to in-person formats. However, many of the original trials also highlighted challenges related to risk management, therapist preparedness, and technical difficulties. The authors concluded that despite various challenges, the telehealth delivery of DBT programs is more accessible and feasible and equally acceptable, safe, and effective as the in-person delivery. However, of the 11 trials included in this review, only 1 trial was conducted with patients with BPD [43], therefore it remains unclear whether their conclusions can be generalized to BPD-specific populations.

While the aforementioned results suggest the telehealth delivery of DBT may be promising, a closer look into the primary studies included in the 2 reviews reveals a number of limitations that impact the conclusions we can draw from them. For example, Lopez et al [41] conducted a nonrandomized trial comparing attendance, client satisfaction, and group cohesion measures in a telehealth DBT skills group (n=20) compared to an in-person skills group (n=15). They found that attendance rates were significantly higher in the telehealth format (91% vs 75%), though there were no significant differences in satisfaction between the groups. Cohesion measures between group members and their therapists showed no differences, although cohesion between group members was significantly higher in the in-person group. Additionally, clients were not randomized but instead allowed to select their preferred group format. Finally, their study did not report on clinical outcomes related to borderline symptoms, which means we cannot compare the clinical effectiveness of the 2 treatment formats on this domain.

Salamin et al [42] analyzed diary card data of 7 patients with BPD within a treatment that had to be shifted from in-person to telehealth format due to the COVID-19 pandemic. In the first 8 weeks of treatment, patients received weekly in-person individual sessions as well as group sessions. Due to COVID-19 restrictions, treatment shifted to telehealth, and patients received weekly individual sessions by telephone or videoconferencing and group therapy was replaced with personal coaching. Weekly monitoring of problem behaviors showed a decrease in binge-eating behaviors and a trend toward a decrease in alcohol consumption following the shift to telehealth treatment. Results also indicated that while clients reported a decrease in feelings of shame or guilt, fear and tension following the move to a telehealth format, their general level of distress increased. Given the small sample size and lack of a control group, as well as the significant confound of the start of the COVID-19 pandemic that spurred the change in format to begin with, results from this study cannot reliably be attributed to the treatment or delivery format.

Alavi et al [43] investigated the effectiveness of an email-based DBT intervention compared to an in-person DBT skills group in 107 patients with BPD. While there were no significant differences between the treatment formats, these findings must be interpreted with caution: the DBT intervention consisted of only 15 sessions, and less than 50% of patients in either group completed treatment and were included in the analysis.

In summary, there is limited evidence regarding the feasibility and effectiveness of telehealth comprehensive DBT.

Objective of This Study

In this study, we aimed to investigate the acceptance, safety, and effect sizes of a 1-year comprehensive outpatient DBT program delivered telehealth for individuals with BPD within a controlled research setting. We were interested in the effect sizes for BPD symptoms and quality of life, and whether self-harm and dissociative symptoms could be targeted and effectively managed in a telehealth setting. To understand the acceptability and feasibility of telehealth DBT, we collected clients' and therapists' perspectives on the delivery of telehealth treatment. Finally, to determine whether a strong therapeutic relationship could be established in a telehealth treatment, we measured both therapists' and clients' ratings of the quality of the alliance throughout treatment.

Methods

Study Design

The data originate from a pilot psychotherapy trial comparing 2 different forms of DBT (standard DBT and trauma-focused DBT) within a telehealth setting. This pilot trial was designed to test the feasibility, safety, and acceptance of telehealth DBT and to prepare for a larger randomized controlled trial (RCT) of trauma-focused DBT. In order to ensure comparability with existing studies in the field, we chose to solely analyze the study arm that delivered telehealth standard (comprehensive) DBT, as developed by Marsha Linehan [12-14]. The study was conducted across 2 sites in Germany (Mannheim and Bochum) and 1 site in Canada (Toronto) between February 2022 and June 2023. The pilot study was registered in the German Clinical

XSL•FO
RenderX

Trials Register (DRKS00027824) and received approval from the respective ethical review boards at each participating site.

Intervention and Treatment

Participants received a 12-month outpatient standard DBT intervention, which comprised weekly individual therapy sessions (each lasting 50 minutes) and weekly group sessions (each lasting 100 minutes). All individual and group sessions were delivered via telehealth using videoconferencing platforms. The treatment was based on Marsha Linehan DBT treatment manual [13,14]. The group therapy focused on skill-building in mindfulness, distress tolerance, emotion regulation, and interpersonal effectiveness. Individual session agendas were informed by a review of clients' diary cards. All skills were taught twice (ie, 2 rounds of a 20-session curriculum). In individual therapy, treatment targets were prioritized based on the dynamic hierarchy focus of DBT that addresses life-threatening behavior, therapy-interfering behavior, and quality-of-life-interfering behavior, as needed. After completing the 12-month therapy period, clients were offered up to 5 booster sessions, which could be scheduled flexibly within the 6 months following treatment termination. Treatment was considered concluded once the booster sessions were finished. No specific adjustments were made to the treatment content to accommodate the telehealth format. However, patients completed their diary cards electronically via an app on their personal smartphones to facilitate diary card review. Participants who missed 4 consecutive weeks of individual therapy were categorized as treatment dropouts. Completion according to protocol (ATP) was defined as attending a minimum of 32 individual sessions, which equaled 80% of the expected dosage. All therapists (n=16) were trained in DBT with at least 2 days of theoretical DBT training completed at the beginning of the study (mean 4.7, SD 2.2 days), and at least half a year of clinical DBT experience (mean 3.6, SD 3.4 years). All therapists participated in a DBT consultation team to guide the treatment, which occurred on a weekly basis, consisting of 1.5-hour sessions per week.

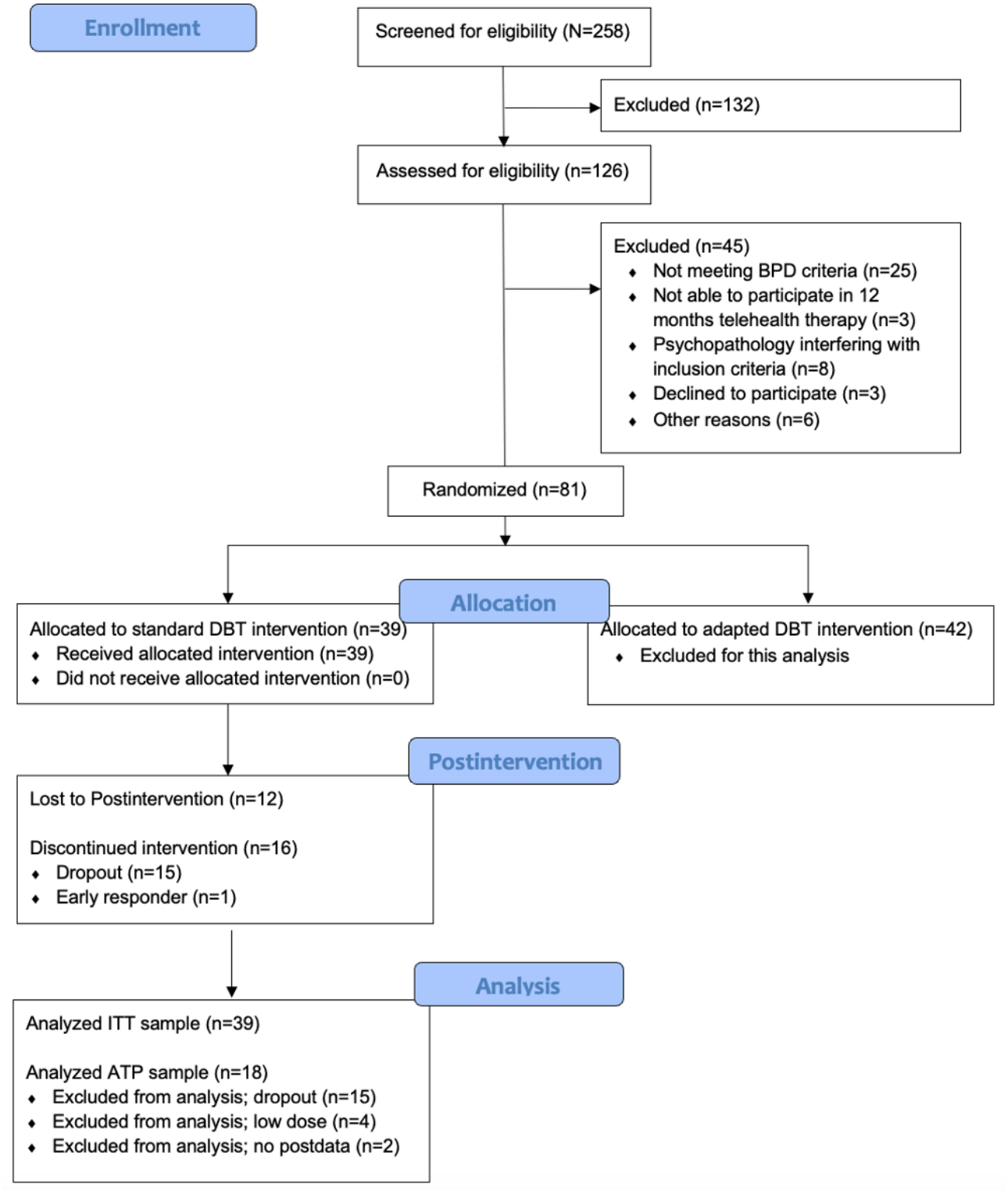
Participants

Participants were recruited from outpatient clinic waiting lists in Mannheim, Bochum, and Toronto, as well as through advertisements and therapist referrals. Inclusion criteria were individuals aged 18-65 years with a diagnosis of BPD, and meeting at least 5 of 9 criteria outlined in the *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders* [Fifth Edition]). Additionally, participants needed to commit to 1 year of outpatient treatment and be willing to engage in telehealth therapy (with internet-enabled devices provided to participants, if needed). Proficiency in English (for the Toronto site) or German (for the Mannheim and Bochum sites) and written informed consent were further prerequisites for participation. Exclusion criteria represented symptoms that indicated a different treatment model or setting was likely needed, such as a lifetime diagnosis of schizophrenia, bipolar I disorder, intellectual disability (IQ<70), posttraumatic stress disorder (PTSD), a BMI below 17.5, severe substance use disorder according to *DSM-5* requiring medical withdrawal support, and dementia. Additionally, individuals with more than 8 weeks of DBT in the preceding year were excluded.

As depicted in the flowchart (Figure 1), 258 individuals underwent telephone screening for eligibility, with 126 undergoing full assessment. Ultimately, 81 (64%) individuals met the inclusion or exclusion criteria and were randomized to receive either standard DBT treatment (n=39; 48%) or a newly adapted DBT treatment (trauma-focused DBT; n=42, 52%). As this study only analyzed acceptance, safety, and effect sizes for standard telehealth DBT, our final sample comprised 39 participants enrolled in the 1-year telehealth standard DBT program. Among these, 15 (38%) individuals discontinued treatment and were classified as dropouts. One client ended the treatment as an early responder. Additionally, 12 (31%) participants did not provide assessments postintervention and were excluded from the analysis ATP. Consequently, 18 (46%) participants were included in the ATP analysis post intervention. The intent-to-treat (ITT) sample comprised all 39 participants who were randomly assigned to telehealth standard DBT.

XSL•FO
RenderX

Figure 1. The flowchart shows the progression of participants throughout the study. ATP: according to protocol; BPD: borderline personality disorder; DBT: dialectical behavior therapy; ITT: intention-to-treat.



Measures

Primary Clinical Outcome

Borderline Symptom List-23

The Borderline Symptom List-23 (BSL-23) [44] was used to assess the borderline symptoms of participants throughout the trial. This self-report instrument comprises 23 items assessing borderline-typical symptoms on a 5-point Likert Scale, ranging from 0=not at all to 4=very strong. Widely used, the BSL-23 has demonstrated good psychometric properties [44,45]. In our sample, Cronbach α for the BSL-23 was 0.92. Assessment points included pre- and postintervention points, as well as every 3 months during the treatment period.

XSL•FO
RenderX

Secondary Clinical Outcomes

Deliberate Self-Harm Inventory

The Deliberate Self-Harm Inventory (DSHI) [46] is a self-report questionnaire designed to assess nonsuicidal self-injuries (NSSIs). Comprising 17 items, the DSHI captures data on the frequency, severity, duration, and type of self-harm behaviors. This instrument was administered both before and post intervention, as well as every 3 months during the treatment phase. In these analyses, we used the frequency score of the DSHI to determine the proportion of clients showing any NSSI in the last 10 weeks.

Dissociative Symptom Scale

The Dissociative Symptom Scale (DSS) [47] is a 20-item self-report questionnaire aimed at assessing dissociative symptoms. Participants are asked how often they have experienced dissociative symptoms over the last week ranging from 0=not at all to 4=more than once a day. It consists of 4 dimensions: depersonalization, derealization, gaps in awareness or memory, and dissociative reexperiencing. Since we were interested in the effect sizes on dissociative symptoms as a secondary outcome, we analyzed a global mean of the DSS across all dimensions. The DSS was administered before and post interventions as well as every 3 months during treatment.

Feasibility and Acceptability Outcomes

Acceptability of Intervention Measure

The Acceptability of Intervention Measure (AIM) [48] is a 15-item scale designed to assess therapists' perspectives on the acceptability, feasibility, and appropriateness of the delivered treatment. Widely used in implementation research, AIM has exhibited robust psychometric properties across various studies [48]. In our sample, intercorrelations between dimensions ranged from r=0.81 to 0.90. Therefore, we computed a mean score by combining the 3 dimensions for analysis. Administration occurred after 6 months of treatment and post intervention.

Client Satisfaction Questionnaire

The Client Satisfaction Questionnaire (CSQ-8) [49] is an 8-item scale that measures the client's satisfaction with the treatment and treatment provider on a 4-point Likert scale. The CSQ-8 was administered after 6 months of treatment as well as post intervention.

Usability of Technology

The unified theory of acceptance and use of technology (UTAUT) [50] assesses acceptance toward digital health care interventions based on the UTAUT framework. The questionnaire consists of 20 items assessing behavioral intention (4 items), performance expectancy (3 items), effort expectancy (3 items), social influence (2 items), facilitating conditions (2 items), internet anxiety (3 items), and experience with the internet (3 items) on a 5-point Likert scale ranging from 0=not at all to 4=very strong. We adapted items to also assess their usability throughout the treatment. The UTAUT was administered before and post interventions as well as at 6 months (midtreatment). Intercorrelations between dimensions ranged from r=0.80 to 0.91. Consequently, we computed a mean score to combine dimensions for analysis.

Severe Adverse Events

Severe adverse events (SAEs) were defined in accordance with the guidelines of good clinical practice. They encompassed suicide attempts or other events leading to death or posing an acute threat to life, where the study participant is in imminent danger of death at the time of the SAE. Therapists were queried weekly regarding the occurrence of potential SAEs. In the event of an SAE, therapists provided additional details to assess its intensity, outcome, and relationship to psychotherapeutic treatment. For safety purposes, all observed SAEs were promptly reported to the responsible ethics committee at the respective study site.

Working Alliance Inventory

The Working Alliance Inventory (WAI) [51] was used to assess the therapeutic alliance from both therapists' and clients' perspectives. The scale comprises 12 items assessing the therapeutic alliance on 3 dimensions: bonding, agreement on tasks, and agreement on goals. Therapists and their clients rate these items on a 7-point Likert scale ranging from 1=never to 7=always. Each therapist received separate questionnaires for each of their clients if they treated more than 1 client. The WAI was administered every 3 months during the treatment period. In our sample, intercorrelations between dimensions ranged from r=0.81 to 0.87 on the client level and from r=0.80 to 0.90 on the therapist level. Consequently, we computed a mean score to combine the 3 dimensions for analysis.

Statistical Analyses

To determine treatment effects, we analyzed BPD symptoms on the BSL-23 as the primary outcome. As secondary outcomes, we analyzed the quality of life, dissociation, and self-harm of participants throughout the treatment. Given the nonlinearity of treatment effects over time, we used a nonparametric Friedman test to assess differences among repeated measures. The Friedman test is robust and capable of handling nonnormal distributions and outliers. Effect sizes for pre-post comparisons for each outcome were analyzed using Cohen d, providing insights into the magnitude of treatment effects. We also calculated 95% CIs to test the robustness of effect size estimates. To compare the proportion of patients showing at least 1 NSSI in the last 10 weeks between pre- and posttreatments, we calculated risk ratio (RR) according to Altman [52] with the following equation, where a represented the number of patients with at least 1 NSSI at postintervention, b represented the number of patients without NSSI at postintervention, c represented the number of patients with at least 1 NSSI at preintervention, and d represented the number of patients without self-harm at preintervention.

To assess the safety of the telehealth trial, we examined SAEs (eg, serious and life-threatening suicide attempts) throughout the study, NSSI as well as individual symptom trajectories of clients on the BSL-23 to identify potential symptom deterioration throughout the treatment.

To evaluate the feasibility and acceptance of the telehealth trial, we conducted descriptive analyses of the client and therapist

XSL•FO
RenderX

feasibility and acceptance measures, as well as client and therapist ratings of the therapeutic alliance.

To analyze the intention-to-treat (ITT) sample, including all participants who had been randomized to the treatment, we imputed missing data by multiple imputation procedures [53]. Multiple imputation was based on the scale-level values of other assessment points and the dropout status (yes or no) to impute missing data. Since the missing pattern was not monotone, the Markov Chain Monte Carlo method [54] was used. Multiple imputation was based on the SAS (version 9.4; SAS Institute) multiple imputation procedures (1000 runs) and MIANALYZE. *P* values<.05 (2-tailed) were considered statistically significant.

Ethical Considerations

This study protocol was reviewed and approved by the applicable ethics committees at each study center, Medical Faculty Mannheim at Heidelberg University in Mannheim of the leading committee for all German sites including Ruhr University Bochum (2021-628-MA), and Center for Addiction and Mental Health, Toronto, Ontario, Canada (023-2023). A written informed consent was obtained from all participants in the study. At the start of the trial, all participants were assigned a randomly generated study code, ensuring that all data were processed and analyzed in a pseudonymized format. Participants did not receive financial compensation for data assessments conducted during the treatment. However, those who discontinued the treatment prematurely were compensated €20 (US $21) for each subsequent assessment time point. At the Canadian study site in Toronto, all clients received CAD $15 (US $10.5) per assessment hour.

Results

Participants

The mean age of participants was 31.1 (SD 10.6) years, with 76.9% (n=30) born female. On average, participants met 6.18 (SD 1.07) criteria for BPD according to *DSM-5*. Additionally, 74.4% of participants had a comorbid affective disorder, 53.8% had a comorbid anxiety disorder, and 23.1% (n=9) had a comorbid eating disorder. Psychotropic medication was received by 97.5% of clients. Table 1 presents the demographic characteristics of the sample.

XSL•FO
RenderX

Table 1. Demographic and clinical characteristics of participants.

Baseline characteristics	Full Sample	
	Values	Range
Age (years), mean (SD)	31.1 (10.6)	19-58
Sex (female), n (%)	30 (76.9)	N/A[a]
Highest educational level, n (%)		
No graduation	0 (0)	N/A
Lower secondary school (hauptschule)	6 (15.4)	N/A
Intermediate secondary school (realschule)	11 (28.2)	N/A
Higher secondary school (abitur)	13 (33.3)	N/A
University or postgraduate degree	9 (23)	N/A
Marital status, n (%)		
Single	26 (66.7)	N/A
Married or partnered	8 (20.5)	N/A
Divorced or widowed	5 (12.8)	N/A
Employment, n (%)		
Unemployed	5 (12.8)	N/A
Student	4 (10.3)	N/A
Employed	28 (71.8)	N/A
Other[b]	2 (5.1)	N/A
BPD[c] criteria (IPDE[d]), mean (SD)	6.18 (1.07)	5-8
Comorbidities[e], n (%)		
Affective disorders	29 (74.4)	N/A
Anxiety disorders	18 (46.2)	N/A
Eating disorders	9 (23.1)	N/A
Obsessive compulsive disorder	5 (12.8)	N/A
Psychotropic medications, n (%)		
Any psychotropic medication	31 (79.5)	N/A
Antidepressants	22 (56.4)	N/A
Neuroleptics	14 (35.9)	N/A
Mood stabilizers	1 (2.6)	N/A
Benzodiazepines	2 (5.1)	N/A
Others	16 (41)	N/A
1 and more suicide attempt (lifetime), n (%)	21 (54)	N/A
1 and more nonsuicidal self-injury in the last month, n (%)	31 (80)	N/A

[a]N/A: not applicable.

[b]n=1 identified as part-time student, n=1 reported current reintegration in the labor market.

[c]BPD: borderline personality disorder.

[d]IPDE: International Personality Disorder Examination.

[e]Comorbidities were categorized as follows: affective disorders (depressive disorder and bipolar II disorder); anxiety disorders (panic disorder, agoraphobia, social anxiety disorder and generalized anxiety disorder); and eating disorders (anorexia, bulimia, and binge eating).

Effect Size Estimations

The effect size estimations revealed a significant decrease in BPD symptoms (BSL-23), our primary outcome, with large effect sizes observed in both the ATP (χ^2_4=24.9; P<.001; d=1.44) and the ITT sample (χ^2_4=38.8; P<.001; d=1.04; Figure 2). As illustrated in Figure 3, the most substantial reduction in BPD symptoms occurred within the first 3 months of treatment,

XSL•FO
RenderX

with a further albeit smaller decrease observed throughout the remainder of the treatment period. Post–BSL-23 scores were a mean 0.91 (SD 0.60) in the ATP sample and a mean 1.20 (SD 0.65) in the ITT sample, indicating mild BPD symptoms in the ATP sample and moderate BPD symptoms in the ITT sample after treatment, respectively.

Figure 2. Course of borderline symptoms as the primary outcome in the ITT sample (n=39) as well as the ATP sample (n=18). Error bars represent 95% CIs using the bootstrapping method. ATP: according to protocol; BSL: Borderline Symptom List; ITT: intention-to-treat.

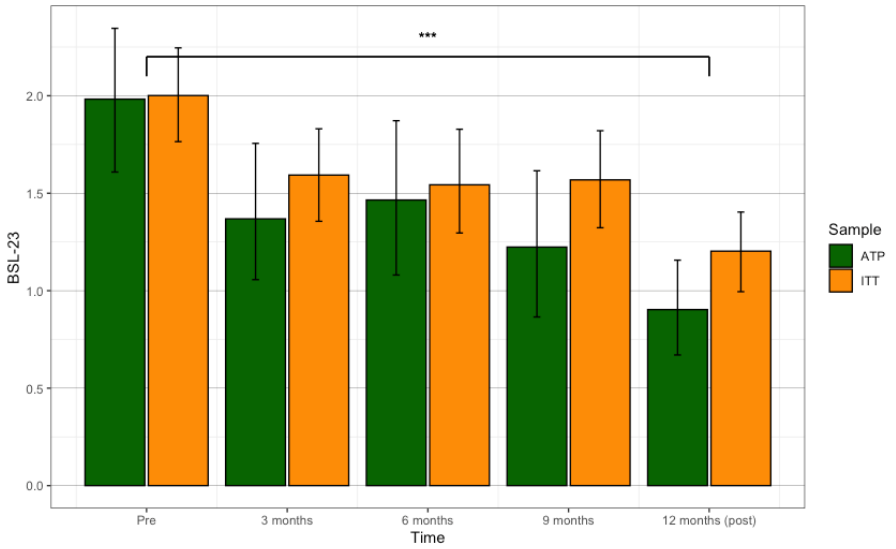
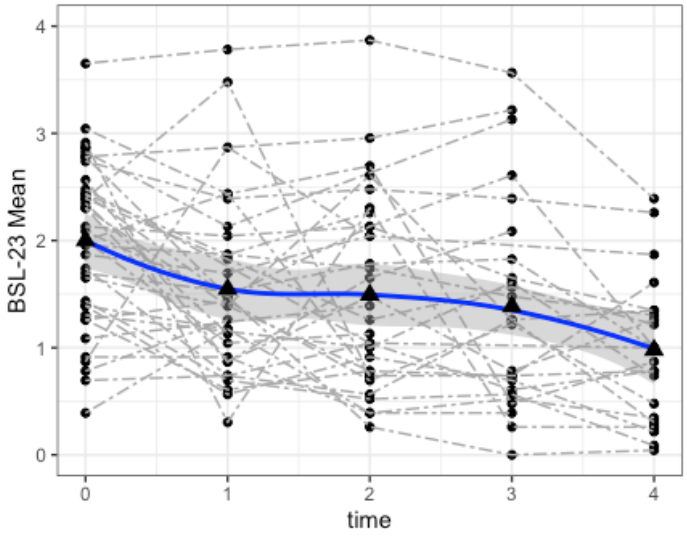


Figure 3. Individual data of participants' borderline symptoms over the course of the whole treatment. BSL: Borderline Symptom List-23.



Dissociative symptoms exhibited a small, but significant decline in the ATP (χ^2_4=11.1; *P*=.02; *d*=0.37) and ITT samples (χ^2_4=12.3; *P*=.02; *d*=0.45) with clients reporting fewer dissociative symptoms throughout the treatment course. However, all 95% CIs of dissociation effect sizes included zero. Self-harm (DSHI scores) significantly decreased over the course of treatment, with 80% of clients engaging in at least 1 self-harm event in the last 10 weeks before treatment compared to 28% after treatment (χ^2_4=16.60; *P*<.001; *RR* 0.35).

The significant reduction in clinical symptoms was mirrored by an increase in quality of life, displaying a large effect size in the ATP (χ^2_4=14.5; *P*<.001; *d*=1.24) and a medium effect size in the ITT sample (χ^2_4= 31.0; *P*<.001; *d*=0.65). The pre-post changes of borderline symptoms and quality of life were significantly intercorrelated by *r*=–0.55 (t_{37}=–4.04; *P*<.001). Detailed results are presented in Table 2.

XSL•FO
RenderX

Table 2. Mean, SD, Friedman Test, and effect size statistics for study variables for ITT[a] and ATP[b] samples[c].

Variable	Baseline, mean (SD)	3 months, mean (SD)	6 months, mean (SD)	9 months, mean (SD)	12 months, mean (SD)	Friedman test		Effect size *d* (pre-post 95% CI)
						Chi-square (*df*)	*P* value	
Effectiveness measures								
ITT (*n*=39)								
Borderline symptoms (BSL-23[d])	2.00 (0.76)	1.59 (0.78)	1.54 (0.84)	1.57 (0.86)	1.20 (0.65)	38.8 (4)	<.001	1.13 (0.64-1.61)
Quality of life (ReQoL[e])	1.48 (0.60)	1.83 (0.55)	2.01 (0.74)	1.71 (0.67)	1.93 (0.76)	31.0 (4)	<.001	0.65 (1.12-0.19)
Dissociation (DSS[f])	0.91 (0.61)	0.81 (0.59)	0.80 (0.68)	0.79 (0.44)	0.65 (0.55)	12.3 (4)	.015	0.45 (–0.01 to 0.90)
Per Protocol (ATP; n=18)								
Borderline symptoms (BSL-23)	1.98 (0.85)	1.38 (0.80)	1.47 (0.96)	1.22 (0.87)	0.91 (0.60)	24.9 (4)	<.001	1.44 (0.70-2.17)
Quality of life (ReQoL)	1.55 (0.58)	1.93 (0.48)	2.15 (0.81)	2.05 (0.65)	2.33 (0.68)	14.5 (4)	<.001	1.24 (0.64-1.61)
Dissociation (DSS)	0.99 (0.67)	0.83 (0.64)	0.88 (0.86)	0.68 (0.50)	0.72 (0.75)	11.1 (4)	.023	0.37 (–0.29 to 1.04)
NSSI[g] (%; DSHI[h])	0.80 (0.41)	0.68 (0.48)	0.40 (0.50)	0.42 (0.50)	0.28 (0.46)	16.6 (4)	<.001	0.35 (RR[i])
Acceptance Measures								
Client measures								
Client satisfaction (1-4; CSQ[j])	—[k]	—	3.33 (0.54)	—	3.40 (0.74)	—	—	—
Useability of technology (0-4; UTAUT[l])	2.81 (0.59)	—	2.83 (0.72)	—	3.13 (0.67)	—	—	—
Therapeutic alliance (1-7; WAI[m])	—	5.52 (1.08)	5.89 (0.75)	5.65 (1.28)	—	—	—	—
Therapist measures								
Acceptance (1-5; AIM[n])	—	—	4.21 (0.74)	—	4.53 (0.61)	—	—	—
Useability of technology (0-4; UTAUT)	2.48 (0.84)	—	—	—	2.57 (0.71)	—	—	—
Therapeutic alliance (1-7; WAI)	—	5.24 (0.98)	5.30 (0.99)	5.40 (0.88)	—	—	—	—

[a]ITT: intention-to-treat.

[b]ATP: according-to-protocol.

[c]Effect Sizes for continuous data were calculated using Cohen *d*. Effect sizes for categorical data (ie, NSSI) were calculated using risk ratio estimates.

[d]BSL-23: Borderline Symptom List-23.

[e]ReQoL: recovering quality of life.

[f]DSS: Dissociative Symptom Scale.

[g]NSSI: nonsuicidal self-injury.

[h]DSHI: Deliberate Self-Harm Inventory.

[i]RR: risk ratio

[j]CSQ: Client Satisfaction Questionnaire.

[k]Not applicable because variable was not assessed at this time-point

[l]UTAUT: unified theory of acceptance and use of technology.

[m]WAI: Working Alliance Inventory.

[n]AIM: Acceptability of Intervention Measure.

XSL•FO
RenderX

Safety

During the course of treatment, 1 SAE was recorded. One patient was hospitalized following a suicide attempt involving an overdose in week 12 of treatment. The therapist rated this SAE as moderate and potentially linked to the psychotherapeutic treatment. In addition, individual patient trajectories in BPD symptoms were observed graphically for safety assessment, as illustrated in Figure 3. The majority of clients demonstrated a consistent decrease in BPD symptoms throughout the treatment duration.

Acceptance, Feasibility, and Usability of Telehealth Treatment

The results regarding the acceptance and feasibility of clients are summarized in Table 2 and depicted in Multimedia Appendix 1. Clients reported high levels of satisfaction with the treatment. 97% (29/30) of clients reported high satisfaction with telehealth DBT treatment midway through therapy (mean 3.33, SD 0.54 on the CSQ-8) and 87% (20/23) reported high satisfaction at the end of treatment (mean 3.40, SD 0.74). Prior to treatment, 69% (27/39) of clients anticipated that the telehealth format would be very or mostly usable for receiving DBT treatment (mean 2.81, SD 0.59 on the UTAUT). This percentage increased slightly during treatment to 77% (23/30) at midtherapy (mean 2.83, SD 0.72) and further to 83% (19/23) at the end of treatment (mean 3.13, SD 0.67).

The results regarding the acceptance and feasibility of therapists are summarized in Table 2 and depicted in Multimedia Appendix 2. From therapists' perspective, 86% (12/14) found the telehealth DBT treatment to be very acceptable, feasible, and appropriate for their clients (mean 4.21, SD 0.74 on the AIM). This percentage rose to 91% (10/11) after the completion of treatment (mean 4.53, SD 0.61). Prior to treatment, 53% (8/15) of therapists anticipated that the telehealth format would be very much or mostly usable for delivering DBT treatment (mean 2.48, SD 0.84 on the UTAUT). However, this percentage slightly decreased during treatment to 45% (5/11), with therapists rating the telehealth format as very or mostly usable for delivering DBT after treatment (mean 2.57, SD 0.71).

Therapeutic Alliance

The therapeutic alliance, as assessed by both clients and therapists, is summarized in Table 2 and depicted in Multimedia Appendix 1 for clients and Multimedia Appendix 2 for therapists. From the client perspective, 82% (28/34) reported experiencing a (very) high affective bond with their therapists and demonstrated (very) high agreement on treatment-related tasks and goals (mean 5.52, SD 1.08 on the WAI). This percentage increased to 96% (27/28) at midtherapy (mean 5.89, SD 0.75) before slightly decreasing to 84% at 9 months into treatment (mean 5.65, SD 1.28).

Similarly, therapist perspectives revealed that 80% (28/25) reported a (very) high affective bond with their clients and rated (very) high agreement on treatment-related tasks and goals (mean 5.24, SD 0.98 on the WAI). This percentage increased slightly to 87% (16/19) at midtherapy (mean 5.30, SD 0.99) before decreasing slightly to 84% (21/25) at 9 months into treatment (mean 5.40, SD 0.88).

A moderate yet significant intercorrelation was observed between therapist and client ratings, with r=0.52 (t_{78}=5.35; P<.001) across all subscales and all 3 assessment times.

Discussion

Primary Results

This pilot study aimed to provide initial insights into the acceptance, safety, and effect sizes of standard (comprehensive) DBT treatment delivered via telehealth (ie, the online delivery of treatment via a video web-based platform) within a controlled research setting. Our findings indicate large pre-post effect sizes on BPD symptoms (d=1.13-1.44) and quality of life (d=0.65-1.24) in both the ITT and ATP samples, despite a relatively high dropout rate of 38% (n=15). One SAE was recorded during the study, attributed to hospitalization following a suicide attempt. Client ratings reflected high levels of satisfaction and usability, with 97% (29/30) expressing high satisfaction with the treatment and 83% (19/23) reporting high usability of the telehealth format. Therapists also expressed strong acceptance of the treatment, with 91% (10/11) indicating high acceptance. Interestingly, therapists rated the usability of the telehealth format comparatively lower than clients, with only 53% (8/15) reporting high usability. The majority of clients (27/28, 96%) and therapists (16/19, 87%) were able to establish a strong therapeutic relationship through telehealth contact.

Comparison With Prior DBT Studies

The dropout rate of 38% (n=15) in this telehealth trial was higher than the average dropout rate of 28% reported in the meta-analysis on in-person DBT trials [55]. One possible explanation for this is the influence of the telehealth setting. Given both the heterogeneity in dropout rates across psychotherapy trials and the lack of an in-person control group in this study to contextualize this result, this finding should be interpreted with caution. Past research has yielded conflicting results regarding attendance and dropout rates in telehealth versus in-person treatments, with some studies indicating no differences between telehealth and in-person DBT, while others have suggested even higher attendance rates in the telehealth format [26,41]. Future studies should aim to directly compare dropout rates between telehealth and in-person DBT programs within RCTs as a broader framework is necessary to determine which clients are more likely to benefit from telehealth settings.

The estimation of effect sizes in this trial demonstrated a significant reduction in BPD symptoms, with large effect sizes of d=1.13 in the ITT sample and d=1.44 in the ATP sample throughout the treatment period. According to the interpretation guidelines for the BSL-23 proposed by Kleindienst et al [56], participants experienced a decrease in average borderline symptom severity from high at baseline to moderate at post-intervention in the ITT sample, and mild in the ATP sample, respectively. Comparing these outcomes with previous in-person DBT trials using a similar treatment dosage, our results suggest that clinical effectiveness in our trial is comparable or even slightly higher. For instance, McMain et al [57] reported a pre-post reduction of d=0.67. Similarly, Priebe et al [58] observed a significant decrease in BPD symptoms, as

XSL•FO
RenderX

assessed by the Zanarini Rating Scale, with an effect size of *d*=0.71 from pre- to postintervention in a 12-month in-person DBT treatment. McMain et al [59] found that a 12-month in-person DBT treatment yielded a pre-post effect size of *d*=1.23 on the Zanarini Rating Scale. Sinnaeve et al [60] demonstrated a significant decrease in the Borderline Personality Disorder Symptom Index [61], with an effect size of *d*=1.15 from pre- to post-intervention in a 12-month outpatient in-person DBT treatment.

Of note, in our study, dissociative symptoms showed only small effect sizes of *d*=0.37 in the ATP and *d*=0.45 in the ITT sample with the 95% CIs containing 0. This finding may be due, in part, to the relatively low baseline dissociative scores of mean 0.91 (SD 0.61) and mean 0.99 (SD 0.67) likely related to our exclusion of patients with BPD with comorbid PTSD, who typically present higher dissociative symptoms compared to those without PTSD [62]. Our dissociation findings therefore continue the ongoing debate on the effectiveness of DBT in reducing dissociative experiences. According to the current Cochrane Review, DBT has a medium effect size for dissociation (*d*=0.43) across primary studies compared to waitlist and treatment-as-usual control groups [15]. Moreover, research suggests that greater effects on dissociation can be achieved by adapting a trauma-focused approach within DBT and explicitly teaching clients antidissociative skills, which were not included in our standard DBT protocol [63,64]. Consequently, our study does not allow us to conclude whether and how dissociative symptoms can be effectively addressed through the telehealth format. Future research should address this question by directly comparing the management of dissociative symptoms in telehealth versus in-person settings.

Self-harm and suicidal behavior in our study significantly decreased with a large reduction of participants' NSSI of RR 0.35 in the ATP sample. While 80% of participants reported engaging in at least 1 self-harm episode in the months preceding treatment initiation, this number decreased significantly to 28% postintervention. This finding aligns with prior research indicating the efficacy of DBT in reducing NSSI behaviors by empowering individuals to attain behavioral control [60,65,66]. Importantly, our study suggests that the reduction of self-harm can be also effectively achieved in a telehealth format. Notably, there was 1 SAE, a suicide attempt by overdose, during the treatment period, with no recorded completed suicides. In previous in-person DBT trials, the prevalence of suicide attempts requiring medical attention (eg, emergency department visits) over a 12-month treatment period has ranged from 3.5% to 6.7% [57,67,68]. Based on these prevalence rates, we would expect 1 to 3 suicide attempts in a cohort of 39 patients. Thus, the single suicide attempt recorded in our study aligns with this expected range and does not suggest an increased occurrence of suicide attempts in telehealth treatment settings.

In terms of the therapeutic alliance, our study showed that over 80% of clients and therapists reported a very strong alliance. Mean scores for the therapeutic alliance ranged from mean 5.24 (SD 0.98) to mean 5.89 (SD 0.75) on a 7-point Likert scale of the WAI. These scores are comparable to, or even higher than, those reported in studies administering the same questionnaire in in-person DBT trials [36,69]. However, this finding may be biased as participants who dropped out of treatment also discontinued completing the alliance measure, potentially inflating alliance ratings over the course of treatment. Nonetheless, our results indicate that a strong therapeutic alliance can be achieved in telehealth DBT settings. This finding challenges the argument that telehealth treatment leads to heightened disengagement or weak therapeutic alliances between clients and therapists [20,70].

Satisfaction and acceptance ratings in our sample were high, with 91% of therapists rating the acceptance of the treatment as high and 98% of participants indicating satisfaction with the treatment. These findings align with previous research demonstrating high acceptance and satisfaction rates of telehealth psychotherapy for patients with personality disorders [71]. From the participants' perspective, useability ratings of the telehealth format for delivering DBT were also high, with 69%-83% indicating high usability. Interestingly, participants' useability ratings significantly increased throughout the treatment, potentially indicating that those with initially low useability ratings may have dropped out of treatment. Unfortunately, our sample size was insufficient to robustly analyze useability as a predictor of therapy dropout during telehealth treatment. Future studies should explore potential predictors for treatment success related to the usability of the telehealth format. In contrast, therapists' perception of telehealth format usability was notably lower, with only half of the therapists (45%-53%) indicating high usability for delivering psychotherapy via telehealth. This discrepancy may be attributed to the greater adjustments required by therapists, transitioning from in-person to telehealth psychotherapy delivery. This finding is consistent with prior research identifying various barriers, challenges, and concerns among therapists delivering DBT via telehealth [20]. Thus, future research should explore strategies to enhance therapists' usability experiences during telehealth treatment. We would also argue that telehealth psychotherapy need not replace in-person psychotherapy, but rather can be used as a complement for clients in need. Both modalities offer valuable options for delivering effective and accessible treatment. Future studies should consider how to tailor treatment delivery format to individual preferences and circumstances.

Limitations

There are some limitations that should be considered when interpreting the study's results. First, the exclusion of patients with comorbid PTSD limits the representativeness of the studied population, particularly its applicability and generalizability to all patients with BPD. Therefore, we caution against generalizing these results to patients with BPD and comorbid PTSD. Second, the absence of a control group directly comparing telehealth and in-person conditions within the same trial and cohort impedes drawing inferential conclusions regarding the relative efficacy of telehealth versus in-person DBT. Lastly, the small sample size of 39 clients did not allow us to analyze potential moderator effects on treatment outcomes or dropout rates. Given the limitations of this pilot and feasibility study, future research should include RCTs with larger sample sizes to directly compare telehealth and in-person DBT treatments.

XSL•FO
RenderX

Conclusions

Our findings provide a first indication that telehealth DBT for BPD appears clinically efficacious, safe, feasible, and well accepted. However, the dropout rate was higher compared to other studies investigating in-person delivered DBT. Future research should compare the efficacy of telehealth DBT with in-person formats in RCTs. Overall, telehealth-delivered DBT offers a potentially effective alternative to in-person DBT, enhancing treatment accessibility and overcoming barriers to treatment engagement. This knowledge may be particularly useful for increasing access to treatment for those in rural areas or low- and middle-income countries where traveling distances to specialized DBT practitioners or treatment programs may be a large barrier.

Acknowledgments

The authors acknowledge the contribution and support of the staff at the study coordinating center and at the participating study sites, Leonie Staufner, Thirza Meyer, Hannah Stiefelhagen, and Eva-Maria Bütow from the Department of Psychosomatic Medicine and Psychotherapy, Central Institute of Mental Health Mannheim, Medical Faculty Mannheim, Heidelberg University; Jasmine Shirvani, Maya Mehrotra, and Askshayan Vimalanathan from the Center for Addiction and Mental Health, Toronto, Ontario, Canada. The authors thank the study therapists, as well as the supervisors, Hans Gunia and Angela Oermann.

Data Availability

All data generated or analyzed during this study are included in this article. Further inquiries can be directed to the corresponding author and will be made available on request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Acceptance, satisfaction, useability, and therapeutic alliance measures of clients during telehealth DBT. DBT: dialectical behavior therapy.
[PNG File , 307 KB - formative_v9i1e66181_app1.png]

Multimedia Appendix 2
Acceptance, satisfaction, useability, and therapeutic alliance measures of therapists during telehealth DBT. DBT: dialectical behavior therapy.
[PNG File , 229 KB - formative_v9i1e66181_app2.png]

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th edition. Washington, DC: American Psychiatric Association; 2013.
2. Bohus M, Stoffers-Winterling J, Sharp C, Krause-Utz A, Schmahl C, Lieb K. Borderline personality disorder. Lancet 2021;398(10310):1528-1540. [doi: 10.1016/s0140-6736(21)00476-1]
3. Eaton NR, Greene AL. Personality disorders: community prevalence and socio-demographic correlates. Curr Opin Psychol 2018;21:28-32. [doi: 10.1016/j.copsyc.2017.09.001] [Medline: 28961462]
4. Ellison WD, Rosenstein LK, Morgan TA, Zimmerman M. Community and clinical epidemiology of borderline personality disorder. Psychiatr Clin North Am 2018;41(4):561-573. [doi: 10.1016/j.psc.2018.07.008] [Medline: 30447724]
5. Bode K, Vogel R, Walker J, Kröger C. Health care costs of borderline personality disorder and matched controls with major depressive disorder: a comparative study based on anonymized claims data. Eur J Health Econ 2017;18(9):1125-1135. [doi: 10.1007/s10198-016-0858-2] [Medline: 27990594]
6. Soeteman DI, Verheul R, Busschbach JJ. The burden of disease in personality disorders: diagnosis-specific quality of life. J Pers Disord 2008;22(3):259-268. [doi: 10.1521/pedi.2008.22.3.259] [Medline: 18540798]
7. van Asselt A, Dirksen CD, Arntz A, Severens JL. The cost of borderline personality disorder: societal cost of illness in BPD-patients. Eur Psychiatry 2007;22(6):354-361. [doi: 10.1016/j.eurpsy.2007.04.001] [Medline: 17544636]
8. Borderline personality disorder: treatment and management. National Institute for Health and Care Excellence. 2009. URL: https://www.nice.org.uk/guidance/cg78 [accessed 2024-12-18]
9. [S3 guideline Borderline personality disorder]. Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenkeilkunde. URL: https://register.awmf.org/assets/guidelines/038-015l_S3_Borderline-Persoenlichkeitsstörungen_2022-11.pdf [accessed 2024-12-10]
10. Mehlum L. Clinical challenges in the assessment and management of suicidal behaviour in patients with bordeline personality disorder. Epidemiol Psichiatr Soc 2009;18(3):184-190. [doi: 10.1017/s1121189x00000440] [Medline: 20034194]

XSL•FO
RenderX

11. Pompili M, Girardi P, Ruberto A, Tatarelli R. Suicide in borderline personality disorder: a meta-analysis. Nord J Psychiatry 2005;59(5):319-324. [doi: 10.1080/08039480500320025] [Medline: 16757458]
12. Linehan MM, Kehrer CA. Borderline personality disorder. In: Barlow DH, editor. Clinical Handbook of Psychological Disorders: A Step-by-Step Treatment Manual. New York, NY: Guilford Press; 1999.
13. Linehan M. Skills Training Manual for Treating Borderline Personality Disorder. New York, NY: Guilford Press; 1993.
14. Linehan M. Cognitive-Behavioral Treatment of Borderline Personality Disorder. New York, NY: Guilford Press; 1993.
15. Storebø OJ, Stoffers-Winterling JM, Völlm BA, Kongerslev MT, Mattivi JT, Jørgensen MS, et al. Psychological therapies for people with borderline personality disorder. Cochrane Database Syst Rev 2020;5(5):CD012955 [FREE Full text] [doi: 10.1002/14651858.CD012955.pub2] [Medline: 32368793]
16. Lohman MC, Whiteman KL, Yeomans FE, Cherico SA, Christ WR. Qualitative analysis of resources and barriers related to treatment of borderline personality disorder in the United States. Psychiatr Serv 2017;68(2):167-172 [FREE Full text] [doi: 10.1176/appi.ps.201600108] [Medline: 27691382]
17. Spina DS, Levy KN. The availability of dialectical behavior therapy in partial hospitalization and residential services for borderline personality disorder: an exploratory longitudinal study of the National Mental Health Services Survey from 2014 to 2021. J Consult Clin Psychol 2024;92(3):176-186. [doi: 10.1037/ccp0000870] [Medline: 38059945]
18. Malhotra S, Chakrabarti S, Shah R. Telepsychiatry: promise, potential, and challenges. Indian J Psychiatry 2013;55(1):3-11 [FREE Full text] [doi: 10.4103/0019-5545.105499] [Medline: 23441027]
19. Bock MM, Graf T, Woeber V, Kothgassner OD, Buerger A, Plener PL. Radical acceptance of reality: putting DBT®-A skill groups online during the COVID-19 pandemic: a qualitative study. Front Psychiatry 2022;13:617941 [FREE Full text] [doi: 10.3389/fpsyt.2022.617941] [Medline: 35546945]
20. Cooney EB, Walton CJ, Gonzalez S. Getting DBT online down under: the experience of Australian and New Zealand dialectical behaviour therapy programmes during the Covid-19 pandemic. PLoS One 2022;17(10):e0275636 [FREE Full text] [doi: 10.1371/journal.pone.0275636] [Medline: 36201507]
21. Hood P, Turner CJ, Beggs B, Owens L, Chapman AL. Making lemonade out of lemons: dialectical behavior therapy via telehealth during a pandemic. Behav Ther 2023;54(5):876-891 [FREE Full text] [doi: 10.1016/j.beth.2023.03.007] [Medline: 37597964]
22. Conroy K, Kehrer SM, Georgiadis C, Hare M, Ringle VM, Shaw AM. Learning from adolescents and caregivers to enhance acceptability and engagement within virtual dialectical behavior therapy for adolescents skills groups: a qualitative study. Child Psychiatry Hum Dev 2024:1-15. [doi: 10.1007/s10578-023-01641-7] [Medline: 38183575]
23. Dunn N, Walton CJ, Matsunaga E, Williams C, Dimeff LA. Acceptability of telehealth for multidiagnostic suicidal patients in a real-world dialectical behavior therapy clinic during the COVID-19 pandemic. Telemed J E Health 2023;29(4):593-601. [doi: 10.1089/tmj.2022.0205] [Medline: 35984847]
24. Lakeman R, King P, Hurley J, Tranter R, Leggett A, Campbell K, et al. Towards online delivery of dialectical behaviour therapy: a scoping review. Int J Ment Health Nurs 2022;31(4):843-856 [FREE Full text] [doi: 10.1111/inm.12976] [Medline: 35048482]
25. Stetsiv K, Rebmann K, Wilks CR. Fears and challenges of dialectical behavior therapists using telehealth. J Behav Cogn Ther 2023;33(1):57-64. [doi: 10.1016/j.jbct.2023.02.001]
26. Walton CJ, Gonzalez S, Cooney EB, Leigh L, Szwec S. Engagement over telehealth: comparing attendance between dialectical behaviour therapy delivered face-to-face and via telehealth for programs in Australia and New Zealand during the Covid-19 pandemic. Borderline Personal Disord Emot Dysregul 2023;10(1):16 [FREE Full text] [doi: 10.1186/s40479-023-00221-4] [Medline: 37208784]
27. Wilks CR, Gurtovenko K. Telehealth delivery of dialectical behavior therapy. In: Weinberg H, Rolnick A, Leighton A, editors. Advances in Online Therapy: Emergence of a New Paradigm. New York, NY: Routledge; 2022.
28. Zalewski M, Walton CJ, Rizvi SL, White AW, Gamache Martin C, O'Brien JR, et al. Lessons learned conducting dialectical behavior therapy via telehealth in the age of COVID-19. Cogn Behav Pract 2021;28(4):573-587 [FREE Full text] [doi: 10.1016/j.cbpra.2021.02.005] [Medline: 34629837]
29. Hyland KA, McDonald JB, Verzijl CL, Faraci DC, Calixte-Civil PF, Gorey CM, et al. Telehealth for dialectical behavioral therapy: a commentary on the experience of a rapid transition to virtual delivery of DBT. Cogn Behav Pract 2022;29(2):367-380. [doi: 10.1016/j.cbpra.2021.02.006]
30. Stoll J, Müller JA, Trachsel M. Ethical issues in online psychotherapy: a narrative review. Front Psychiatry 2019;10:993 [FREE Full text] [doi: 10.3389/fpsyt.2019.00993] [Medline: 32116819]
31. Kleindienst N, Limberger MF, Ebner-Priemer UW, Keibel-Mauchnik J, Dyer A, Berger M, et al. Dissociation predicts poor response to dialectial behavioral therapy in female patients with borderline personality disorder. J Pers Disord 2011;25(4):432-447. [doi: 10.1521/pedi.2011.25.4.432] [Medline: 21838560]
32. Kleindienst N, Priebe K, Görg N, Dyer A, Steil R, Lyssenko L, et al. State dissociation moderates response to dialectical behavior therapy for posttraumatic stress disorder in women with and without borderline personality disorder. Eur J Psychotraumatol 2016;7(1):30375 [FREE Full text] [doi: 10.3402/ejpt.v7.30375] [Medline: 27396380]

XSL•FO
RenderX

33. Seitz KI, Sicorello M, Schmitz M, Valencia N, Herpertz SC, Bertsch K, et al. Childhood maltreatment and amygdala response to interpersonal threat in a transdiagnostic adult sample: the role of trait dissociation. Biol Psychiatry Cogn Neurosci Neuroimaging 2024;9(6):626-634. [doi: 10.1016/j.bpsc.2024.01.003] [Medline: 38280631]
34. Lyssenko L, Schmahl C, Bockhacker L, Vonderlin R, Bohus M, Kleindienst N. Dissociation in psychiatric disorders: a meta-analysis of studies using the dissociative experiences scale. Am J Psychiatry 2018;175(1):37-46. [doi: 10.1176/appi.ajp.2017.17010025] [Medline: 28946763]
35. Miano A, Barnow S, Wagner S, Roepke S, Dziobek I. Dyadic emotion regulation in women with borderline personality disorder. Cogn Ther Res 2021;45(6):1077-1092. [doi: 10.1007/s10608-021-10206-8]
36. Steuwe C, Berg M, Driessen M, Beblo T. Therapeutic alliance and dropout in patients with borderline pathology receiving residential dialectical behavior therapy. BMC Psychiatry 2023;23(1):605 [FREE Full text] [doi: 10.1186/s12888-023-05061-8] [Medline: 37596568]
37. Boritz T, Barnhart R, Eubanks CF, McMain S. Alliance rupture and resolution in dialectical behavior therapy for borderline personality disorder. J Pers Disord 2018;32:115-128. [doi: 10.1521/pedi.2018.32.supp.115] [Medline: 29388901]
38. Flückiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: a meta-analytic synthesis. Psychotherapy 2018;55(4):316-340 [FREE Full text] [doi: 10.1037/pst0000172] [Medline: 29792475]
39. Kaiser J, Hanschmidt F, Kersting A. The association between therapeutic alliance and outcome in internet-based psychological interventions: a meta-analysis. Comput Hum Behav 2021;114:106512. [doi: 10.1016/j.chb.2020.106512]
40. van Leeuwen H, Sinnaeve R, Witteveen U, Van Daele T, Ossewaarde L, Egger JIM, et al. Reviewing the availability, efficacy and clinical utility of telepsychology in dialectical behavior therapy (Tele-DBT). Borderline Personal Disord Emot Dysregul 2021;8(1):26 [FREE Full text] [doi: 10.1186/s40479-021-00165-7] [Medline: 34717772]
41. Lopez A, Rothberg B, Reaser E, Schwenk S, Griffin R. Therapeutic groups via video teleconferencing and the impact on group cohesion. Mhealth 2020;6:13 [FREE Full text] [doi: 10.21037/mhealth.2019.11.04] [Medline: 32270005]
42. Salamin V, Rossier V, Joye D, Nolde C, Pierrehumbert T, Gothuey I, et al. [Adaptations of an outpatient dialectical behavioral therapy during the COVID-19 pandemic and consequences of the confinement on patients with borderline personality disorder]. Ann Med Psychol 2021;179(2):131-136 [FREE Full text] [doi: 10.1016/j.amp.2020.08.006] [Medline: 32843771]
43. Alavi N, Stephenson C, Rivera M. Effectiveness of delivering dialectical behavioral therapy techniques by email in patients with borderline personality disorder: nonrandomized controlled trial. JMIR Ment Health 2021;8(4):e27308 [FREE Full text] [doi: 10.2196/27308] [Medline: 33835936]
44. Wolf M, Limberger MF, Kleindienst N, Stieglitz R, Domsalla M, Philipsen A, et al. [Short version of the borderline symptom list (BSL-23): development and psychometric evaluation]. Psychother Psychosom Med Psychol 2009;59(8):321-324. [doi: 10.1055/s-0028-1104598] [Medline: 19274605]
45. Bohus M, Kleindienst N, Limberger MF, Stieglitz R, Domsalla M, Chapman AL, et al. The short version of the borderline symptom List (BSL-23): development and initial data on psychometric properties. Psychopathology 2009;42(1):32-39. [doi: 10.1159/000173701] [Medline: 19023232]
46. Gratz K. Measurement of deliberate self-harm: preliminary data on the deliberate self-harm inventory. J Psychopathol Behav Assess 2001;23:253-263. [doi: 10.1037/t04163-000]
47. Carlson EB, Waelde LC, Palmieri PA, Macia KS, Smith SR, McDade-Montez E. Development and validation of the Dissociative Symptoms Scale. Assessment 2018;25(1):84-98. [doi: 10.1177/1073191116645904] [Medline: 27178761]
48. Weiner BJ, Lewis CC, Stanick C, Powell BJ, Dorsey CN, Clary AS, et al. Psychometric assessment of three newly developed implementation outcome measures. Implement Sci 2017;12(1):108 [FREE Full text] [doi: 10.1186/s13012-017-0635-3] [Medline: 28851459]
49. Attkisson CC, Greenfield TK. The Client Satisfaction Questionnaire (CSQ) scales and the Service Satisfaction Scale-30 (SSS-30). In: Sederer LL, Dickey B, editors. Outcomes Assessment in Clinical Practice. Baltimore, MD: Williams & Wilkins; 1996:120-127.
50. Philippi P, Baumeister H, Apolinário-Hagen J, Ebert DD, Hennemann S, Kott L, et al. Acceptance towards digital health interventions—model validation and further development of the unified theory of acceptance and use of technology. Internet Interv 2021;26:100459 [FREE Full text] [doi: 10.1016/j.invent.2021.100459] [Medline: 34603973]
51. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. J Counsel Psychol 1989;36(2):223-233. [doi: 10.1037//0022-0167.36.2.223]
52. Altman DG. Practical Statistics for Medical Research. New York, NY: Chapman and Hall/CRC; 1990.
53. Rubin DB. Multiple imputation after 18+ years. J Am Stat Assoc 1996;91(434):473-489. [doi: 10.1080/01621459.1996.10476908]
54. Schafer JL. Analysis of Incomplete Multivariate Data. New York, NY: Chapman and Hall/CRC; 1997.
55. Dixon LJ, Linardon J. A systematic review and meta-analysis of dropout rates from dialectical behaviour therapy in randomized controlled trials. Cogn Behav Ther 2020;49(3):181-196. [doi: 10.1080/16506073.2019.1620324] [Medline: 31204902]
56. Kleindienst N, Jungkunz M, Bohus M. A proposed severity classification of borderline symptoms using the Borderline Symptom List (BSL-23). Borderline Personal Disord Emot Dysregul 2020;7(1):11 [FREE Full text] [doi: 10.1186/s40479-020-00126-6] [Medline: 32514359]

XSL•FO
RenderX

57. McMain SF, Chapman AL, Kuo JR, Dixon-Gordon KL, Guimond TH, Labrish C, et al. The effectiveness of 6 versus 12 months of dialectical behavior therapy for borderline personality disorder: a noninferiority randomized clinical trial. Psychother Psychosom 2022;91(6):382-397 [FREE Full text] [doi: 10.1159/000525102] [Medline: 35738244]
58. Priebe S, Bhatti N, Barnicot K, Bremner S, Gaglia A, Katsakou C, et al. Effectiveness and cost-effectiveness of dialectical behaviour therapy for self-harming patients with personality disorder: a pragmatic randomised controlled trial. Psychother Psychosom 2012;81(6):356-365 [FREE Full text] [doi: 10.1159/000338897] [Medline: 22964561]
59. McMain SF, Links PS, Gnam WH, Guimond T, Cardish RJ, Korman L, et al. A randomized trial of dialectical behavior therapy versus general psychiatric management for borderline personality disorder. Am J Psychiatry 2009;166(12):1365-1374. [doi: 10.1176/appi.ajp.2009.09010039] [Medline: 19755574]
60. Sinnaeve R, van den Bosch LM, Hakkaart-van Roijen L, Vansteelandt K. Effectiveness of step-down versus outpatient dialectical behaviour therapy for patients with severe levels of borderline personality disorder: a pragmatic randomized controlled trial. Borderline Personal Disord Emot Dysregul 2018;5(1):12 [FREE Full text] [doi: 10.1186/s40479-018-0089-5] [Medline: 30002832]
61. Arntz A, van den Hoorn M, Cornelis J, Verheul R, van den Bosch WM, de Bie AJ. Reliability and validity of the borderline personality disorder severity index. J Pers Disord 2003;17(1):45-59. [doi: 10.1521/pedi.17.1.45.24053] [Medline: 12659546]
62. Cackowski S, Neubauer T, Kleindienst N. The impact of posttraumatic stress disorder on the symptomatology of borderline personality disorder. Borderline Personal Disord Emot Dysregul 2016;3(1):1-7 [FREE Full text] [doi: 10.1186/s40479-016-0042-4] [Medline: 27482383]
63. Harned MS, Korslund KE, Linehan MM. A pilot randomized controlled trial of dialectical behavior therapy with and without the dialectical behavior therapy prolonged exposure protocol for suicidal and self-injuring women with borderline personality disorder and PTSD. Behav Res Ther 2014;55:7-17 [FREE Full text] [doi: 10.1016/j.brat.2014.01.008] [Medline: 24562087]
64. Bohus M, Kleindienst N, Hahn C, Müller-Engelmann M, Ludäscher P, Steil R, et al. Dialectical behavior therapy for posttraumatic stress disorder (DBT-PTSD) compared with cognitive processing therapy (CPT) in complex presentations of PTSD in women survivors of childhood abuse: a randomized clinical trial. JAMA Psychiatry 2020;77(12):1235-1245 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.2148] [Medline: 32697288]
65. Carter GL, Willcox CH, Lewin TJ, Conrad AM, Bendit N. Hunter DBT project: randomized controlled trial of dialectical behaviour therapy in women with borderline personality disorder. Aust N Z J Psychiatry 2010;44(2):162-173. [doi: 10.3109/00048670903393621] [Medline: 20113305]
66. van den Bosch LM, Koeter MW, Stijnen T, Verheul R, van den Brink W. Sustained efficacy of dialectical behaviour therapy for borderline personality disorder. Behav Res Ther 2005;43(9):1231-1241. [doi: 10.1016/j.brat.2004.09.008] [Medline: 16005708]
67. Linehan MM, Comtois KA, Murray AM, Brown MZ, Gallop RJ, Heard HL, et al. Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. Arch Gen Psychiatry 2006;63(7):757-766. [doi: 10.1001/archpsyc.63.7.757] [Medline: 16818865]
68. Linehan MM, Korslund KE, Harned MS, Gallop RJ, Lungu A, Neacsiu AD, et al. Dialectical behavior therapy for high suicide risk in individuals with borderline personality disorder: a randomized clinical trial and component analysis. JAMA Psychiatry 2015;72(5):475-482. [doi: 10.1001/jamapsychiatry.2014.3039] [Medline: 25806661]
69. Hirsh JB, Quilty LC, Bagby RM, McMain SF. The relationship between agreeableness and the development of the working alliance in patients with borderline personality disorder. J Pers Disord 2012;26(4):616-627. [doi: 10.1521/pedi.2012.26.4.616] [Medline: 22867511]
70. Norwood C, Moghaddam NG, Malins S, Sabin-Farrell R. Working alliance and outcome effectiveness in videoconferencing psychotherapy: a systematic review and noninferiority meta-analysis. Clin Psychol Psychother 2018;25(6):797-808. [doi: 10.1002/cpp.2315] [Medline: 30014606]
71. Bailey RC, Knowles NG, Grenyer BF. Efficacy and recommendations for the delivery of telehealth psychotherapy for people with personality disorder. Australas Psychiatry 2024;32(2):125-134 [FREE Full text] [doi: 10.1177/10398562231222768] [Medline: 38103036]

Abbreviations

AIM: Acceptability of Intervention Measure
ATP: according to protocol
BPD: borderline personality disorder
BSL-23: Borderline Symptom List-23
CSQ-8: Client Satisfaction Questionnaire
DBT: dialectical behavior therapy
DSHI: Deliberate Self-Harm Inventory
DSM-5: *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition)
DSS: Dissociative Symptom Scale

XSL•FO
RenderX

ITT: intention-to-treat
NSSI: nonsuicidal self-injury
PTSD: posttraumatic stress disorder
RCT: randomized controlled trial
RR: risk ratio
SAE: severe adverse event
UTAUT: unified theory of acceptance and use of technology
WAI: Working Alliance Inventory

Edited by A Mavragani; submitted 06.09.24; peer-reviewed by B Mohajerin, DR Bartsch; comments to author 14.10.24; revised version received 28.10.24; accepted 01.11.24; published 14.01.25.

Please cite as:
Vonderlin R, Boritz T, Claus C, Senyüz B, Mahalingam S, Tennenhouse R, Lis S, Schmahl C, Margraf J, Teismann T, Kleindienst N, McMain S, Bohus M
Acceptance, Safety, and Effect Sizes in Online Dialectical Behavior Therapy for Borderline Personality Disorder: Interventional Pilot Study
JMIR Form Res 2025;9:e66181
URL: https://formative.jmir.org/2025/1/e66181
doi:10.2196/66181
PMID:

©Ruben Vonderlin, Tali Boritz, Carola Claus, Büsra Senyüz, Saskia Mahalingam, Rachel Tennenhouse, Stefanie Lis, Christian Schmahl, Jürgen Margraf, Tobias Teismann, Nikolaus Kleindienst, Shelley McMain, Martin Bohus. Originally published in JMIR Formative Research (https://formative.jmir.org), 14.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Cardiomeds, an mHealth App for Self-Management to Support Swiss Patients With Heart Failure: 2-Stage Mixed Methods Usability Study

Lisa Simioni[1]; Elena Tessitore[1,2], MD; Hamdi Hagberg[3]; Aurélie Schneider-Paccot[2]; Katherine Blondon[1,4], MD, PhD; Liliane Gschwind[5], PhD; Philippe Meyer[2], MD; Frederic Ehrler[3], PhD

[1]Faculty of Medicine, University of Geneva, Geneva, Switzerland

[2]Department of Cardiology, University Hospital of Geneva, Geneva, Switzerland

[3]Department of Computer Science, University Hospital of Geneva, Geneva, Switzerland

[4]Medical Directorate, University Hospital of Geneva, Geneva, Switzerland

[5]Department of Pharmacy, University Hospital of Geneva, Geneva, Switzerland

Corresponding Author:
Lisa Simioni
Faculty of Medicine
University of Geneva
Université de Genève Rue du Général-Dufour 24 CH
1211 Genève 4
Geneva
Switzerland
Phone: 41 786420657
Email: lisa.simioni@hotmail.com

Abstract

Background: Mobile health apps have shown promising results in improving self-management of several chronic diseases in patients. We have developed a mobile health app (Cardiomeds) dedicated to patients with heart failure (HF). This app includes an interactive medication list; daily self-monitoring of symptoms, weight, blood pressure, and heart rate; and educational information on HF delivered through various formats.

Objective: This study aimed to perform a mixed methods usability study of Cardiomeds.

Methods: Smartphone users with HF were recruited from the HF outpatient clinic at the University Hospital of Geneva. The usability test was conducted in 2 stages, with modifications made to the app after the first stage to address major usability issues. Each stage required 10 participants to perform 14 tasks, such as entering vital signs, entering a new medication and time of intake, or finding information about HF. Each task was timed, sessions were recorded, and all data were anonymized. After completing the tasks, patients completed the System Usability Scale 10-item questionnaire and answered 5 open questions about their perceptions of Cardiomeds.

Results: Twenty patients with HF, 75% (15/20) of whom were men, with a mean age of 55 years, were included in this study. The average time to complete all 14 tasks was 18 (SD 5.7) minutes. Manual medication entry was the most time-consuming task, taking an average of 154.40 (SD 68.08) seconds in the first stage, 103.10 (SD 42.76) seconds in the second stage, and 128 (SD 63) seconds overall. The mean overall success rate was 77% (SD 0.23%) for the first stage and 94% (SD 0.07%) for the second stage. A total of 30% (3/10) of participants in the first stage completed all tasks without any help compared with 50% (5/10) of participants during the second stage. The average System Usability Scale score was 80% (SD 17%), showing a slight increase from 79% (SD 16%) in the first stage to 80% (SD 28%) in the second stage, which qualifies the app as “good” in terms of usability. Between the 2 stages, part of the app interface was redesigned to address the key issues identified in the first stage. Despite these improvements, problems related to guidance were frequent and comprised 36% (8/22) of the problems in the first stage and 40% (6/15) in the second stage. In response to open questions, 85% (17/20) of the participants responded that they would like to use the app when it became available.

XSL•FO
RenderX

Conclusions: The usability test indicated that Cardiomeds is a suitable and user-friendly app for patients with HF. The app will be further tested in a randomized clinical trial (2022-00731) after acute HF hospitalization to assess its impact on patients' knowledge about HF, self-care, and quality of life.

(JMIR Form Res 2025;9:e63941) doi:10.2196/63941

KEYWORDS

usability; medication; mobile health; mHealth; Cardiomeds; mobile app; patient empowerment; eHealth; smartphone; heart failure; HF; chronic disease; interactive; self-monitoring; usability test; mobile phone

Introduction

Background

Heart failure (HF) impacts around 2% of the adult population in Europe and is the leading cause of hospital admissions among individuals aged >65 years. Within a year after being hospitalized for acute HF, over half of the patients are readmitted, and nearly a quarter of them die. This situation poses an enormous medical and economic challenge [1]. The cornerstone of HF management is pharmacotherapy, which involves the combination of multiple drug classes [2]. Patients with HF need to make lifestyle adjustments to prevent deterioration, keep track of symptom changes, and take daily medications. Support interventions for self-management behavior generally aim to provide patients with the skills to actively manage their chronic condition. These interventions focus on promoting symptom monitoring and improving problem-solving and decision-making abilities for medical treatment and a healthy lifestyle [3]. According to the latest guidelines of the European Society of Cardiology, self-management strategies are strongly recommended to reduce the risk of hospitalization and mortality related to HF [2]. Various self-management strategies are possible, such as multidisciplinary care, peer meetings, and the use of logs for symptom monitoring or setting goals. The longer these interventions are maintained, the greater the reduction in mortality, with a 1% to 4% decrease in risk for each additional month. In addition, these interventions work better when accompanied by regular contact with health professionals [4].

With advancements in technology, the introduction of therapeutic monitoring with telephone support is becoming increasingly prominent [5], which includes features to register symptoms, enter a treatment card, organize appointments, record alarms for taking medications, or support behavior reinforcement through educational materials [6].

Research consistently supports the effectiveness of structured telephone support and noninvasive telemonitoring. Several systematic reviews have concluded that these approaches significantly reduce all-cause mortality and HF-related hospitalizations [7-9]. Moreover, these interventions also showed enhancements in health-related quality of life, knowledge of HF, and self-care behaviors [7,10].

Smartphone-guided interventions in HF are growing in significance and effectiveness, with studies incorporating monitoring not only through text messages and phone calls but also via mobile health (mHealth) apps [11,12].

Thus, many mHealth apps are being developed and are the subject of various studies [5,6,10,12,13].

Several mHealth apps, providing valuable support to patients with HF are already available. For instance, the WOW ME 2000mg—Heart Failure Self-management Tool (AtlantiCare) focuses on helping patients and caregivers track symptoms and medications [6,14,15]. WebMD (WebMD Health Corp) offers a wide range of health resources, including HF management [14,16]. Similarly, My Cardiac Coach (American Heart Association) is designed for post–heart attack care and self-monitoring. Finally, "Heart Failure Health Storylines" (Heart Failure Society of America) aids in symptom tracking and treatment adherence [6,14,17]. Nevertheless, none of these apps refer to the Swiss medication database.

Experience with these existing mHealth apps has demonstrated a significant improvement in therapeutic adherence among patients requiring daily treatment [18,19]. In addition, patients understood their illness better, which allows them to feel more secure [10].

Despite the potential benefits of mHealth apps, studies have shown that patients with HF often face barriers to achieving good adherence. For many older patients, difficulties begin with technical issues, such as challenges in installing or navigating the apps on their smartphones [20]. These issues are often exacerbated by limited digital literacy, a lack of confidence in using technology, or cognitive impairments commonly associated with HF. In addition, usability concerns, such as complex interfaces, insufficient personalization, or lack of reminders, can further reduce the likelihood of long-term adherence [21]. Overcoming these barriers requires designing apps with simplicity, user-friendliness, and accessibility in mind while providing continuous support for patients [22].

In Switzerland, according to a study conducted by Pro Senectute, 69% of citizens older than 65 years own a smartphone, 81% of whom use it daily. Moreover, this study also discovered that their interest in mHealth apps is consequent [23]. Thus, the familiarity of older adults with smartphones makes the use of an mHealth app credible for conducting a study on patients with HF [24].

In this context, the cardiology department of the University Hospital of Geneva (HUG) has undertaken the development of a new mHealth app tailored for patients with HF, known as Cardiomeds. Designed to complement the physician-patient relationship, the app has been developed following a user-centric approach. We collected the information required to develop the app through interviews and focus groups that included patient groups, clinicians, and IT professionals. We created personas

XSL•FO
RenderX

and user journeys and refined our prototype iteratively using feedback from the stakeholders.

Therefore, it was important to conduct a usability test of this app before offering it to patients with HF. Indeed, these tests make it possible to point out the design problems of the apps, to improve the uncovering of opportunities, and to learn about the target user's behavior and preferences [25].

Objectives

The study aimed to evaluate the usability of the Cardiomeds mHealth app through qualitative and quantitative usability metric in laboratory settings. The main focus of the usability assessment was to test the app's ergonomic design, user-friendliness, comprehensibility, and appropriateness for the target population.

Methods

Study Design

The app's usability was evaluated through a usability test requiring participants to complete a sequence of tasks. The study was performed in 2 stages. After the completion of the first usability test by 10 participants, modifications were implemented in the app based on their feedback. Subsequently, a second usability test, featuring a new cohort of 10 new patients with the same protocol was conducted.

Participants and Setting

The study was conducted from May 6, 2022, to December 21, 2022, in a multipurpose room at the HUG to standardize the intervention. The evaluation was done through user-investigator interaction, deliberately omitting the user's real environments. Tasks were performed on an iPhone 6s smartphone (Apple Inc) at a resolution of 1334×750 pixels. For each of the 2 stages, we recruited 2 groups of 10 participants. Ten participants enabled to increase the number of comments and results in accordance with the recommendations which propose at least 5 participants for one arm of a study [26]. Participants were recruited on a voluntary basis from the HUG HF outpatient clinic. Participants had to be adults (aged ≥18 years), speak French, have and use a smartphone, and be a local resident. All enrolled participants had a diagnosis of HF.

The study was divided into 2 stages, each comprising 10 participants (resulting in a total of 20 participants), as proposed by the Nielsen Norman Group [27].

Cardiomeds Mobile App Overview

Cardiomeds is an mHealth app being developed since 2017 by HUG's cardiology, pharmacy, medical directorate, and IT departments, with the help of computer scientists, a nurse, pharmacists, physicians, and a student [28]. The app has been designed to fulfill the needs of patients with HF, allowing them to gather and record pertinent data concerning their condition, including treatment regimens, vital sign measurements, and the progression of their symptoms. In addition, the app facilitates data sharing with health care providers and enhances patients' understanding of their condition through informative quizzes and simplified content. The app also includes the option of setting reminders or alarms to ensure regular and accurate treatment, as well as monitoring patients' daily medication compliance.

The app's ultimate objective is to empower patients by fostering a comprehensive understanding of their medical condition, promoting effective self-management of daily treatments, facilitating continuous clinical monitoring, and enhancing patients' knowledge of their condition. Figures 1-5 provide an overview of the main menu and the app's appearance as tested by patients.

XSL•FO
RenderX

Figure 1. Screenshot of the main menu section of the Cardiomeds app. In this section, patients enter their vital signs of the day, program reminders, and validate their medication intake. Moreover, if their vital signs are not in a healthy range, the app suggests contacting their physician.

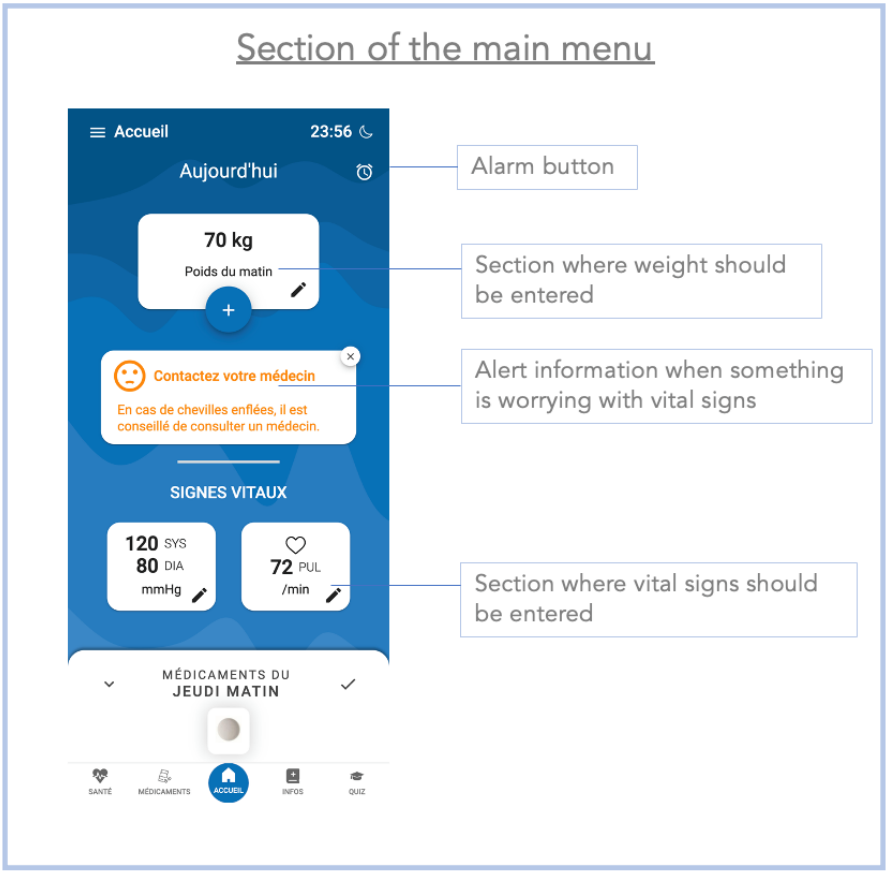


Figure 2. Screenshot of the section where vital signs are entered in the Cardiomeds app. Patients have to enter their systolic and diastolic blood pressure, their pulse, and the day's date.

XSL•FO
RenderX

Figure 3. Screenshot of the information section of the Cardiomeds app. Here, patients can find information about their illness, signs of worsening heart failure, their medications, and recommended lifestyle.

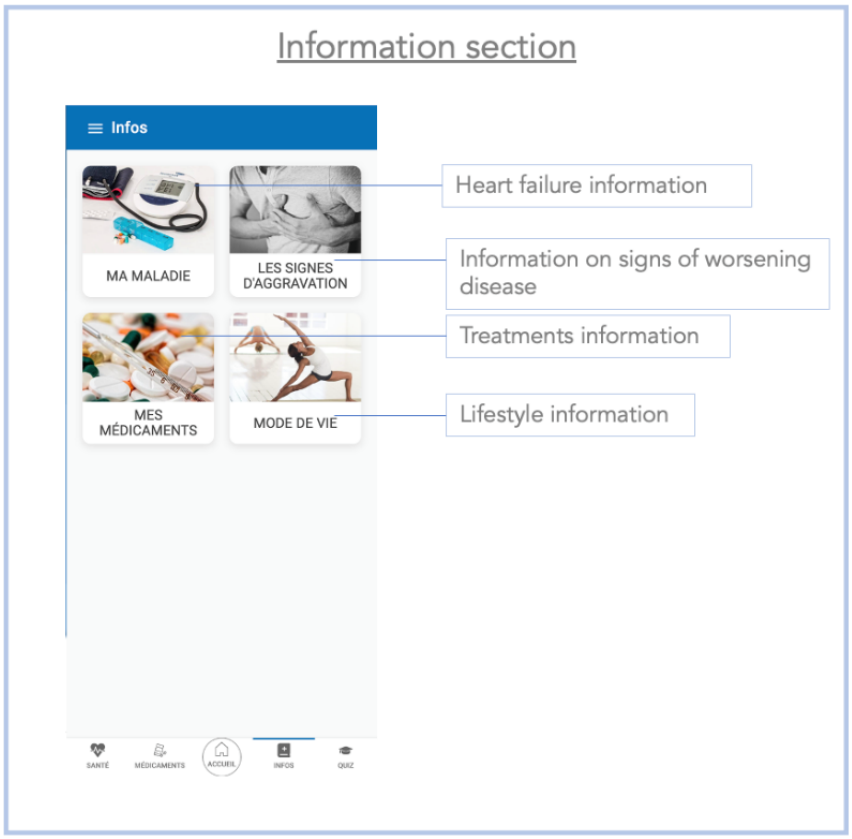
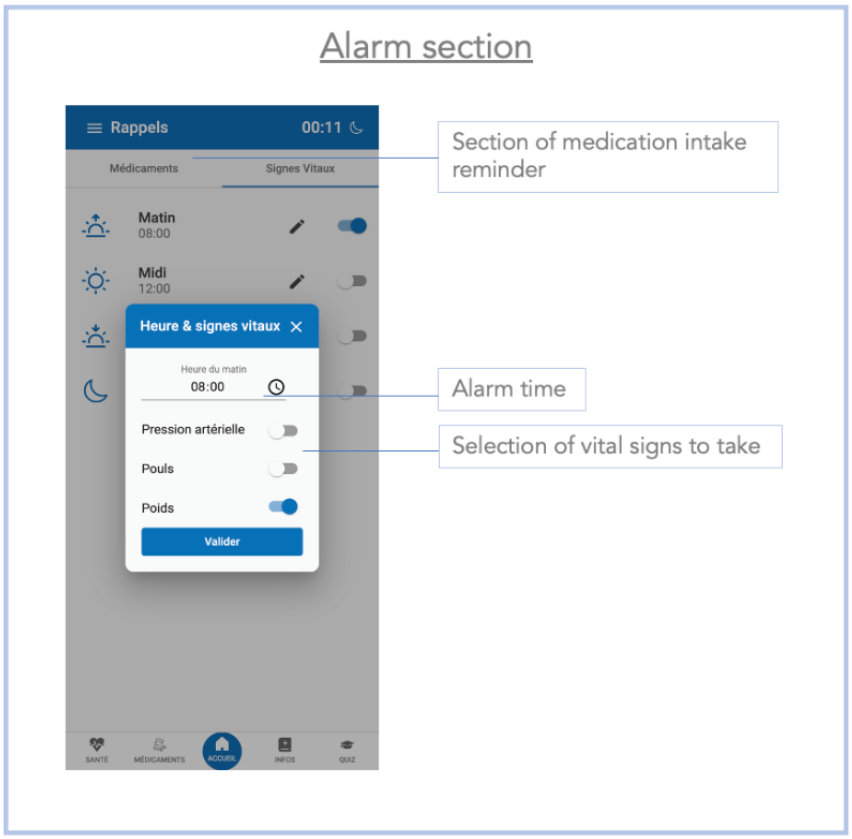


Figure 4. Screenshot of the disease course curve section of the Cardiomeds app. In this section, patients can see the course over the past few days for weight, blood pressure, and pulse. A button to export the curves is also present.

XSL•FO
RenderX

Figure 5. Screenshot of the alarm reminder section of the Cardiomeds app. This section allows the user to set an alarm for taking medication or measuring their vital signs.



Procedure

All participants were recruited from the HF outpatient clinic of the HUG. They were all invited to join the study via a phone call by a student (LS). Moreover, patients who were selected, were under hospital monitoring, with some having just completed the cardiac rehabilitation program following cardiac decompensation. All interested patients were invited on a given date to take the test. Participants were welcomed within a dedicated room at the HUG, where 2 assigned investigators, a student (LS) and a medical professional (ET), were responsible for overseeing the test procedure. Upon arrival and initial briefing by the investigators, participants signed a consent form, which explicitly stipulated the anonymization of their data.

All participants were provided with a loaned iPhone for test purposes. Then, the test was executed according to the protocol (Textbox 1). Collected measures included task completion timing, recording the activities performed on the smartphone screen, and capturing both the participants' vocal expressions and the investigator's commentary. The test comprised 14 predefined tasks exploring the functionalities of the app (Textbox 2). The investigator read out these tasks and precisely timed their execution. Task outcomes were categorized as follows:

- "Passed" if completed within a maximum of 2 minutes without requiring assistance.
- "Partially successful" if completed in more than 2 minutes or necessitating clues provided by the investigator.
- "Failed" if the task was abandoned or remained unsuccessful after 5 minutes.

Upon test completion, each participant was invited to complete the System Usability Scale (SUS) questionnaire [29,30], as well as 5 open questions about their personal feelings about the app and possible improvements (Textbox 3).

Subsequently, data collected from the tests were analyzed using Microsoft Excel. For in-depth insights into the procedures governing each test, please refer to the comprehensive procedure (Textbox 1), as well as the questionnaires distributed to participants (Textbox 3).

XSL•FO
RenderX

Textbox 1. Procedure of the usability test.

- Welcoming of the user of the app by the investigator.
- Arrival of the user and the worker in a booth.
- Verification of the user's identity and consent by the investigator.
- Asking for the user's patient ID, age, gender, education level, smart phone use time per day, and health app use (yes or no)
- Description of the booth:
  - The booth will contain a desk as well as 2 chairs side by side where the user and the investigator can sit.
  - On the desk, 2 cell phones (a Samsung [Samsung Electronics Co, Ltd] phone and an iPhone) will be placed. Therefore, the user will be able to choose the phone whose operating mode they know and with which they are most comfortable.
  - In addition, a sheet on which the tasks to be carried out during the session will be placed on the desk for the user to guide them.
  - Medication packages will also be placed on the desk to carry out the session's tasks.
- Explanation of the session to the user by the investigator:
  - The investigator will inform the user that their screen and their voice will be recorded (consent).
  - The investigator will explain to the user that the latter will have to carry out the requested tasks, describe their actions out loud, and inform the investigator of the start and end of their tasks.
  - The investigator will explain to the user that after 2 minutes, he or she will gradually give them clues specific to each task to help the user be able to finish the task. If after 5 minutes, the user is unable to complete the task despite help, the task will be marked as failed by the investigator. If the task is completed, the worker will mark it as successful. Once a task is completed, the user can move on to the next task.
  - The investigator will inform the user that he or she can interrupt the session at any time for any reason.
- Start of session:
  - At the start of the session the investigator will set up a screenshot as well as an audio recording if the user gives their consent.
  - The investigator will read the tasks aloud one after the other to the user.
  - Before starting to perform the tasks, the user will have 2 minutes to familiarize themselves with the design of the app.
  - At the beginning, the only questions the investigator will answer will be questions of understanding, if there is a misunderstanding of terminology.
  - The user can reread the task on the sheet placed on the desk for them if necessary.
  - As soon as the user starts each task, he or she will mention it out loud.
  - As soon as the user finishes each task, he or she will mention it out loud.
  - If the user has comments, the investigator will listen to them.
- Tasks: read aloud 14 tasks for the user. Please refer to the tasks below in Textbox 2.

XSL•FO
RenderX

Textbox 2. Description of the 14 goal-oriented test tasks to be performed by users during the usability study.

- Task 1: set alarm for vital signs (set a reminder to measure your vital signs at 8 AM).
- Task 2: decision-making (go to the home page and enter your morning weight of 65.0 kg). You must then answer 3 questions as follows:
  - Short of breath? As usual.
  - Swollen ankles? More than usual.
  - Any other symptoms? None of the above.
- Task 3: enter weight (you forgot to record your weight on 04.01.2022, and you now want to enter this value, which was 62.3 kg).
- Task 4: enter vital signs (enter your daily vital signs: blood pressure 120/80 mm Hg and pulse rate 70/min).
- Task 5: enter medication (add the medication to the app's medication list using the available package. Scan the barcode for Beloc ZOK 25 mg. The prescription for Beloc ZOK 25 mg is 2 1⁄2 tablets (2.5 tablets) in the morning every day. In addition, add the number of tablets found in the package).
- Task 6: enter medication without medication package (add a second medication to the app's medication list. You do not have the medication box. The prescription is as follows: Fosamax [bisphosphonate] 70 mg, 1 tablet on Sunday mornings only; 1 per week. The quantity in the pack is 4 tablets).
- Task 7: enter pro re nata (PRN; taken as needed without a set schedule) medication, without medication package (add a third medication to the app's medication list. You do not have the medication package. The prescription is as follows: zolpidem (any type proposed) 10 mg, maximum 1 tablet per day, in the evening if necessary. The number of tablets in the pack is 10).
- Task 8: set alarm for medication intake (set a reminder to take your medication at 8:00 AM).
- Task 9: confirm your medication intake (confirm that you have taken your morning medication).
- Task 10: check medication history (check your medication history).
- Task 11: find medication information (you no longer know why you are taking Beloc ZOK [β-blocker]. Where can you find this information? There are 2 ways to find this information).
- Task 12: export vital signs (export your health data [weight, blood pressure, and pulse] to your doctor by email).
- Task 13: quiz (test your knowledge on heart failure with the quiz).
- Task 14: find alcohol information (find information on the possibility of consuming alcohol in the context of your disease).

Textbox 3. Five open questions questionnaire given to users after passing the test.

1. What is your overall impression of the Cardiomeds app?
2. Do you plan to use Cardiomeds when the app is available, and why?
3. What improvements could we add to the Cardiomeds app?
4. Do you have any other comments about the Cardiomeds app or the usability test?
5. What differences have you noticed between Cardiomeds and other health apps?

Data Collection

Participants were audio-recorded during the test, and the screen of the test iPhone was recorded. The recording allowed us to retrospectively classify the usability problems encountered with the Bastian and Scapin criteria [31]. The audio was recorded using an iPad-mini-Dictaphone by a single investigator. Moreover, during the test, the investigator timed the tasks with another smartphone. The screen was recorded using the screen recording option built in the iPhone on which the participants were taking the tests. Subsequently, the usability metrics were transcribed onto Microsoft Excel spreadsheets. The audio records were analyzed by the investigator using thematic analysis. The SUS paper questionnaires were collected immediately after the intervention and transcribed onto Microsoft Excel spreadsheets. The investigator analyzed the success rates of each task and their duration. All of the collected data were anonymized. All data described above were collected and analyzed by a single investigator for the 20 participants. ChatGPT (GPT-4 version, OpenAI) was used as an assistance tool to generate text drafts, as a translation tool, and to synthesize concepts. The results were subsequently reviewed and edited by the authors to ensure accuracy and relevance to the topic.

Usability Analysis

Quantitative Evaluation

The participant's task-based performance was measured by several metrics. Effectiveness is defined by the number of tasks completed by each participant accurately [32]. In this study, effectiveness was calculated in 2 ways: "task completion rate" (TCR) per participant is the percentage of tasks successfully completed [33]. This was calculated using the following equation: TCR per participant=(number of tasks completed

XSL•FO
RenderX

successfully/total number of tasks undertaken)×100. (equation 1) When a task could not be started and evaluated (ie, because of a problem with the Wi-Fi connection), it was coded as "unavailable." TCR per participant was categorized according to 3 different levels of achievement as follows: (1) the task is considered completed when the user has successfully completed the task without any errors or difficulties; (2) completed with difficulty, but with difficulties that could have been solved with the help of the instructor in 5 minutes or less; and (3) failed to complete when the task was left incomplete, abandoned, or the participant took too much time to complete the task (>5 min) [33].

"Distribution of task success by task" was defined as the proportion of participants completing a task according to 3 possible levels of achievement: (1) the task is considered completed when the user has successfully completed the task without any errors or difficulties; (2) completed with difficulty, but with difficulties that could have been solved with the help of the instructor in 5 minutes or less; and (3) failed to complete and the task is left incomplete, abandoned, or the participant took too much time to complete the task (>5 min). When a task could not be started and evaluated (ie, because of a problem with the Wi-Fi connection), it was coded as unavailable [33].

"Efficiency" was defined as the time required for users to achieve specified goals in relation to accuracy and completeness [32]. In this study it was calculated as follows: "time on task" was defined as the average amount of time taken (in seconds) to complete a given task from the moment the participant finished hearing the instructions until the task was completed (whether with ease or with difficulty) or abandoned. When a task could not be started and evaluated (ie, because of a problem with the Wi-Fi connection), it was coded as unavailable.

Satisfaction was measured through the SUS questionnaire. After the completion of the usability test, the SUS questionnaire [29,30] was submitted to each participant. The SUS is a validated questionnaire designed to measure the usability of diverse products and services, including mHealth apps. This questionnaire, conceived by Brooke [29], comprises 10 questions, each offering 5 possible answers, ranging from "strongly disagree" to "strongly agree." In this ordinal scale, "strongly disagree" corresponds to a numerical value of 1, while "strongly agree" corresponds to a numerical value of 5. The even-numbered questions within the SUS are subjected to negative rotation, and their scores are calculated by subtracting the chosen value from 5. In contrast, the odd-numbered questions undergo positive rotation, and their scores are derived by subtracting 1 from their value. Once the individual scores for all 10 questions were calculated, they are added together and multiplied by a factor of 2.5. Consequently, the resultant SUS score falls within the range of 0 to 100. Usability is considered favorable when the SUS score attains a threshold of ≥70, signifying good usability for the tested product.

Qualitative Evaluation

Qualitative data were recorded through an open questionnaire. After completing the SUS questionnaire, participants were asked to answer a series of 5 open questions. The purpose of these questions, defined by the investigators, was to explore the participants' overall impression allowing us to better understand the strengths and weaknesses of our tools. The 5 open questions are described in Textbox 3.

The responses to the first question were classified into positive responses, negative responses, and no responses. The responses to the second question were classified into 4 categories: yes, no, maybe, and do not know. Questions 3 and 4 were analyzed using the frequency of propositions. All answers to the questions were recorded, sorted, and analyzed by a single person.

Bastien and Scapin Classification

We also used the criteria published by Bastien and Scapin [31] in 1993 to evaluate the problems of the app. These criteria cover 8 dimensions and enable the classification of problems between users and the interface.

Ethical Considerations

On April 1, 2022, a formal request was submitted to the Ethics Commission of the Canton of Geneva to conduct usability tests (Req-2022-00423). This request was subsequently approved on April 4, 2022. The app encompassed the testing protocol, a consent form to be executed by the study participants, and an evaluation sheet for each participant.

Results

Participant Characteristics

Between May and December 2022, a total of 20 participants participated in the study. Baseline demographic characteristics of the 2 study stages are shown in Table 1.

Demographic characteristics of study participants (N=10).

XSL•FO
RenderX

Table 1. Characteristics of the participants separated into the 2 stages of the study (N=20).

	Stage 1, n (%)	Stage 2, n (%)	Total, n (%)
Gender			
Women	3 (30)	2 (20)	5 (25)
Men	7 (70)	8 (80)	15 (75)
Age categories (years)			
31-40	1 (10)	0 (0)	1 (5)
41-50	3 (30)	1 (10)	4 (20)
51-60	4 (40)	5 (50)	9 (45)
61-70	0 (0)	4 (40)	4 (20)
71-80	2 (20)	0 (0)	2 (10)
Type of phone			
iPhone	3 (30)	1 (10)	16 (80)
Samsung	7 (70)	9 (90)	4 (20)
Educational level			
PRE[a] graduate school	5 (50)	5 (50)	10 (50)
Graduate school	2 (20)	1 (10)	3 (15)
University	3 (30)	4 (40)	7 (35)
Use of other medical apps			
Yes	4 (40)	2 (20)	6 (30)
No	6 (60)	8 (80)	14 (70)
Time of phone use (h)			
<1	1 (10)	0 (0)	1 (5)
1-3	8 (80)	7 (70)	15 (75)
>3	1 (10)	3 (30)	4 (20)

[a]PRE: federal diploma of vocational education and training.

First Stage of the Study

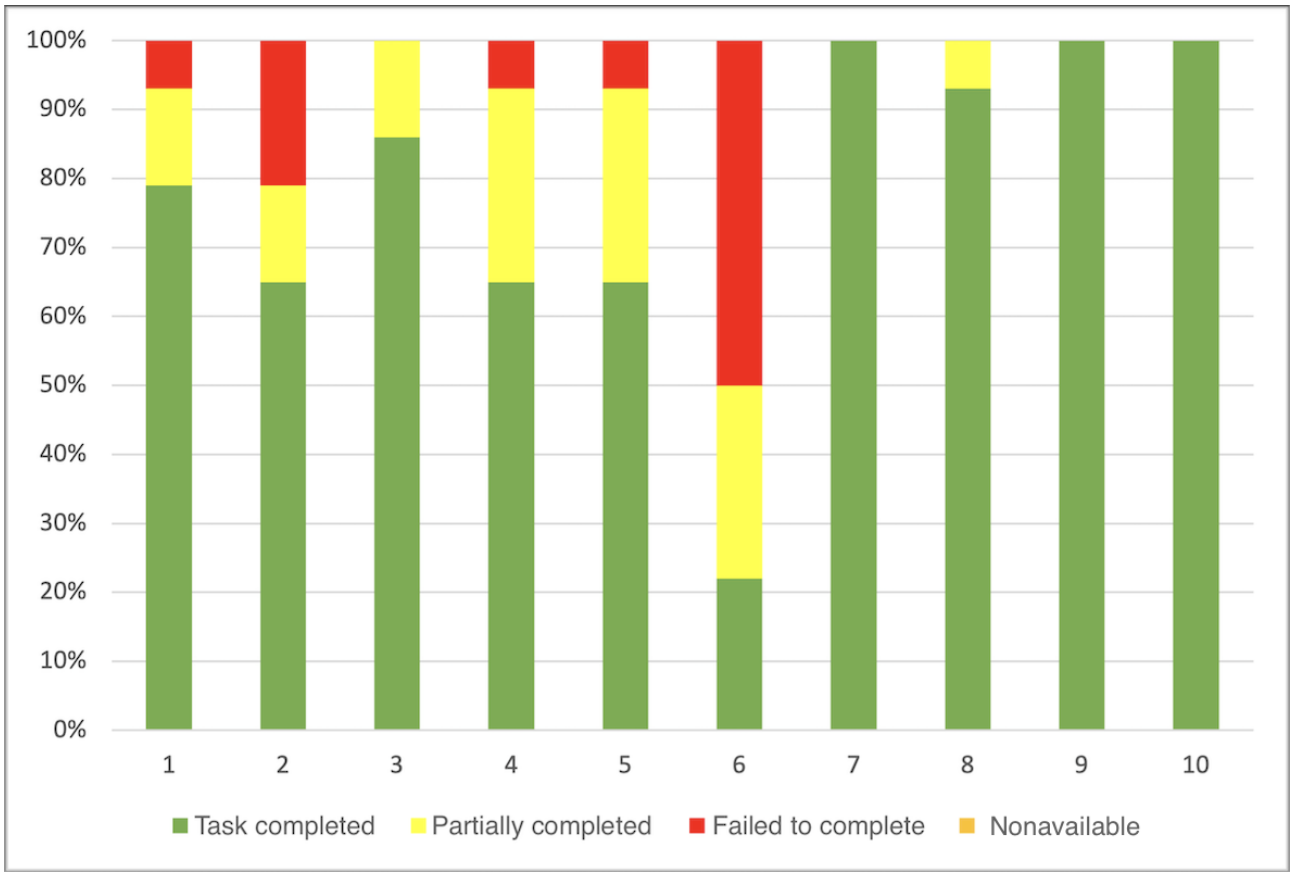
Quantitative Evaluation

Effectiveness per Participant

In the first part of the study, the overall completion rate (tasks completed, partially completed, and failed to complete) was 100% (140/140; Figure 6). The mean overall complete success rate was 50% (5/10), meaning that 50% of the participants solved all the tasks without making any mistakes (they did not fail to complete any task) and 30% (3/10) of the participants completed all the tasks without any help. Participant number 6 failed 50% (7/14) of the tasks and completed only 20% (3/14) of them without any help. Finally, on average, participants completed 77% (11/14; SD 0.23%) of tasks without any help. All these measurements are exposed and supported in Figure 6.

XSL•FO
RenderX

Figure 6. Task completion rate per participant for the 14 assigned tasks during the first stage of the study. "Task completed" represents the percentage of tasks successfully completed by a participant. "Failed to complete" is the percentage of tasks that participants failed to complete. "Partially completed" is defined as tasks completed, but with difficulty that could have been solved with the help of the instructor in 5 minutes or less. "Nonavailable" represents the percentage of missing data when a task could not be started and evaluated.

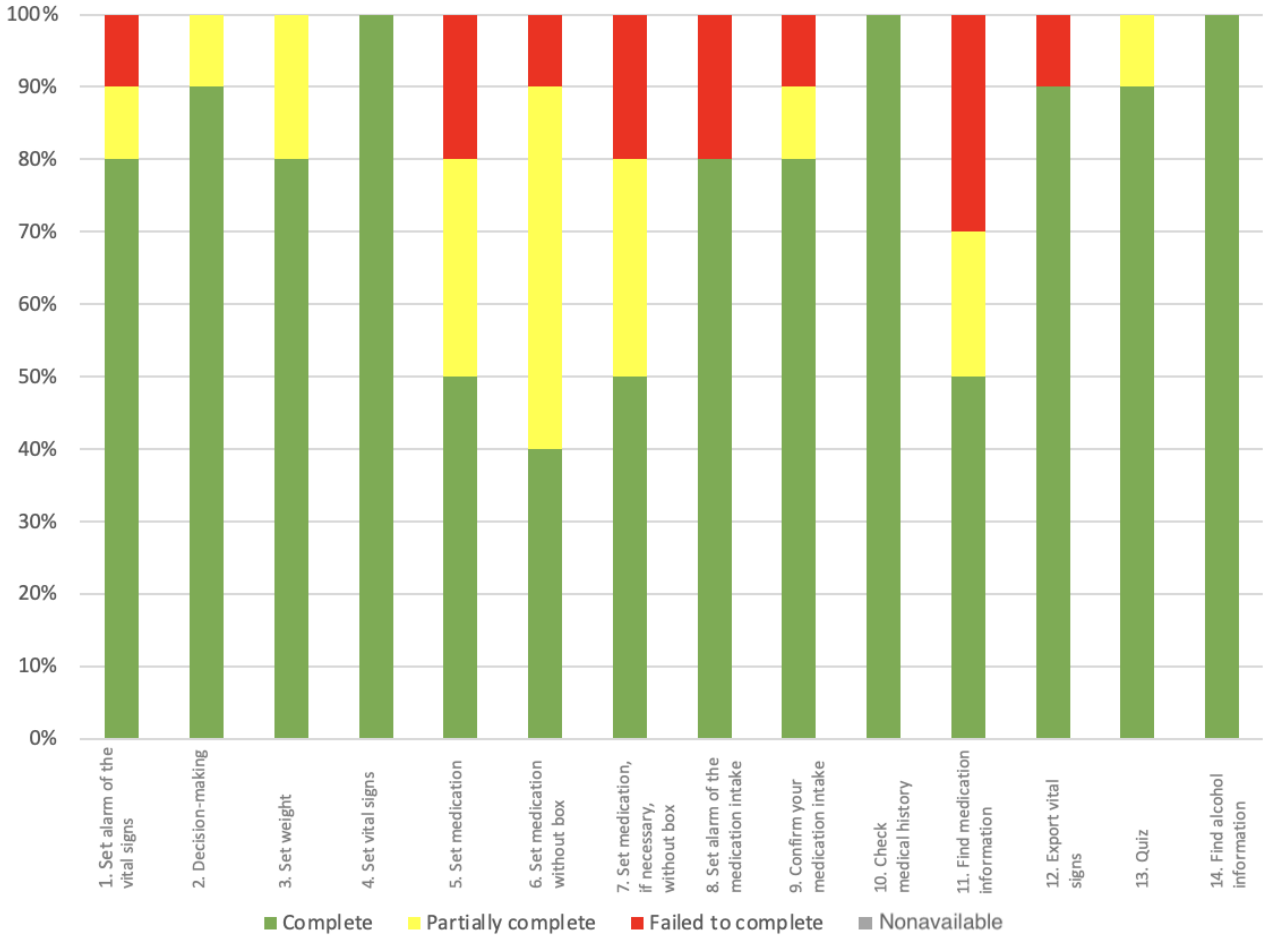


Task Success Distribution per Task

In the first part of the study, overall 77.1% (108/140) of tasks were completed successfully, 13.6% (19/140) were completed partially, and 9.3% (13/140) were not completed. A total of 21% (3/14) of the tasks (4, 10, and 14) were completed by all the participants without any help (10/10, 100%), followed by task 2 (9/10, 90%). Apart from tasks 4, 10, and 14, 79% (11/14) of the tasks led to difficulties for up to 60% (6/10) of the participants (task 6). At least 1 (1/10) participant failed to complete 8 (8/14) of the tasks (1, 5, 6, 7, 8, 9, 11, and 12), with up to 30% (3/10) of the participants failing to complete task 11, which consisted of finding information about medications in 2 different ways. All these measurements are exposed and supported in Figure 7.

XSL•FO
RenderX

Figure 7. The success distribution per task (N=10 participants) for the first stage of the study. "Complete" represents the percentage of tasks successfully completed by a participant. "Failed to complete" is the percentage of tasks that participants failed to complete. "Partially complete" is defined as tasks completed, but with difficulty that could have been solved with the help of the instructor in 5 minutes or less. "Nonavailable" represents the percentage of missing data when a task could not be started and evaluated.

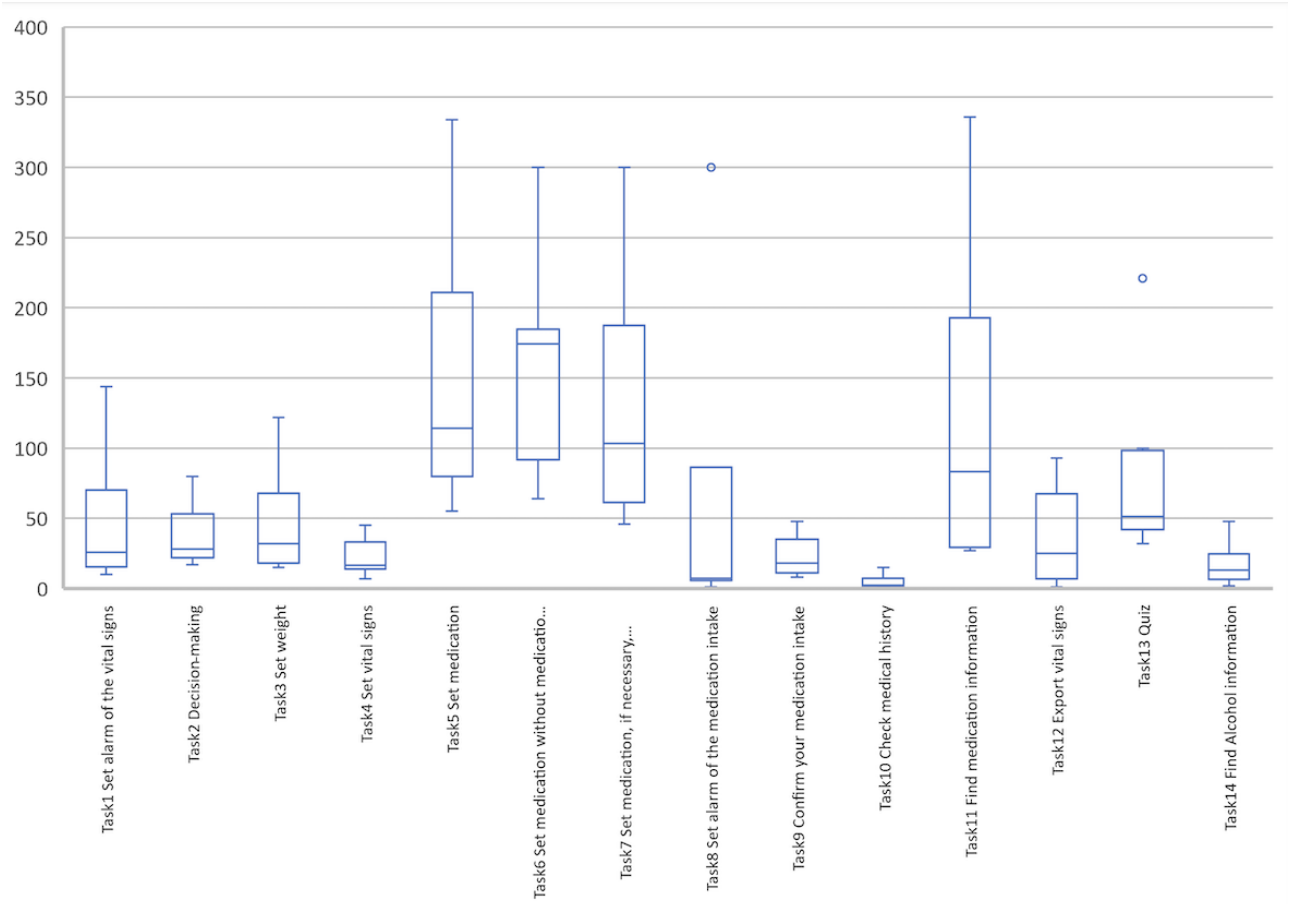


Time on Task per Study Task

In the first stage of the study, task 6, which required the participants to enter medication, took the longest time with an average of 154.40 (SD 68.08) seconds (Figure 8). The fastest task completed was checking medication history (task 10) with an average of 4.8 (SD 4.60) seconds.

XSL•FO
RenderX

Figure 8. Boxplot of the time taken (in seconds) to perform each task in stage 1.



Qualitative Evaluation

Five Open Questions Questionnaire

In stage 1, 80% (8/10) of the participants answered positively, 10% (1/10) answered negatively, and 10% (1/10) gave no answer about their overall impression. A total of 70% (7/10) of the participants said that they would like to use the app in the future, 20% (2/10) answered they would not, and 10% (1/10) did not answer the question. The participants who did not want to use the app mentioned that they already had an mHealth app or found that the app was too complicated for them. Participants suggested 3 different types of improvements for the app as follows: first, to increase the font size of the app; second, to change the medication input; and third, to improve the explanation given about the HF drugs. These comments are more described in Textbox 4.

Textbox 4. List of comments made by participants (N=10) that are useful in the short- and long-term for improving the app during the first stage of the test.

Useful comments for short-term improvements

1. The 2 (20%) oldest participants mentioned that the font in the app was too small to understand, which may be problematic for older users.
2. One (10%) participant was interested in transferring his data, but also in having contact with his cardiologist directly via the app.
3. Six (60%) participants found that entering the information on medications was complicated.
4. One (10%) participant wished to have the emergency number, the number of his physician, his cardiologist, and his pharmacy inside the app.

Useful comments for long-term improvements

1. One (10%) participant was interested in a search function in the information part.
2. One (10%) participant wished to have interactive blood pressure and saturation connected with his smartwatch.
3. One (10%) participant wanted this app to be connected with his other apps linked to the University Hospital of Geneva Swissmeds app.

Error Classification With the Bastien and Scapin Criteria

In stage 1, we recorded 22 problems during the completion of the 14 tasks. We have classified the problems into 9 categories according to the criteria of Bastien and Scapin [31] (Table 2). Task 6, requiring entering a medication, was the most problematic, with 6 (N=10, 60%) users experiencing problems.

XSL•FO
RenderX

The screen enabling manual medication recording was considered too dense with information.

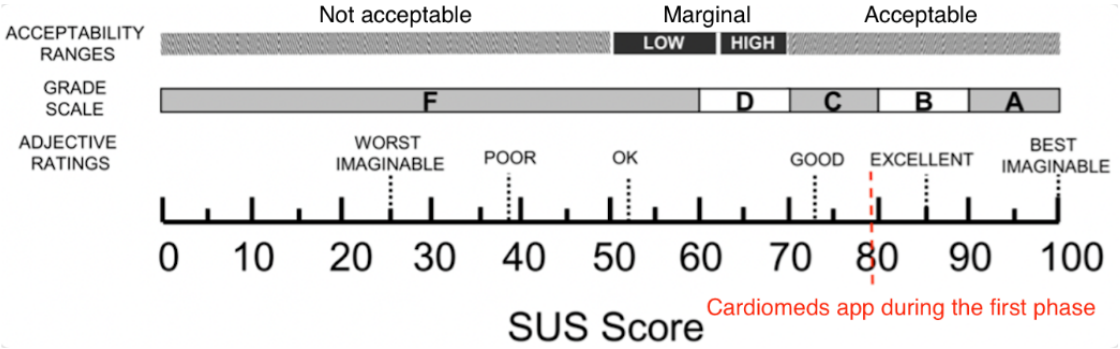
Table 2. List of problems encountered by participants (N=10) during the 14 tasks and description, categorization, and frequency of occurrence of these problems during task completion.

Item	Category (number according to the Bastien and Scapin classification)	Description of the problem	Task in which it occurs (frequency of occurrence)
1	1. Guidance	Five (50%) users did not find the place of the button to put a reminder. For most of them, it was too small, and its location was not easy to find.	Set alarm for the vital signs (5)
2	1. Guidance	Three (30%) users said the “information” symbol was too small next to the treatments list and that they did not find it easily	Find medication information (3)
3	7. Significance of codes	Two (20%) users did not find the scanning possibility option in the app explicit enough	Enter medication (2)
4	2.2 Information density	Six (60%) participants had difficulties recording medication manually in the app. They say they had to put too much information on one page.	Enter medication without medication package (6)
5	7. Significance of codes	One (10%) participant did not understand names of some drugs because the name of the medication was not written next to the active substance	Find medication information (1)
6	1.3 Immediate feedback	One (10%) participant found the validation of medication intake not explicit enough	Confirm your medication intake (1)
7	7. Significance of codes	Two (20%) participants did not understand the term “if necessary,” such as medication to take in reserve	Enter pro re nata, if necessary, without medication package (2)
8	8. Compatibility	One (10%) participant asked to set an end date for his treatment, but this function does not exist	Enter medication (1)
9	5.2 Error message quality	When entering the date, one participant (10%) did not write it correctly (separated by “/” symbols and not “.” symbols). Therefore, the date could not be validated and no explanation appeared	Enter weight (1)

Satisfaction: SUS Questionnaire

At the end of each test, all participants were asked to complete the SUS questionnaire [29,30]. The mean SUS score of the first stage of the study was 79.28 (SD 16.35), (Figure 9; Table 3).

Figure 9. Cardiomeds score on the System Usability Scale (SUS) at stage 1.



XSL•FO
RenderX

Table 3. System Usability Scale (SUS) questionnaire responses in the first stage of the study.

	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7	Question 8	Question 9	Question 10	SUS score (sum×2.5; maximum 100)
Participant 1	4	4	3	4	2	2	3	4	3	3	80
Participant 2	3	3	3	3	3	4	3	4	3	1	75
Participant 3	4	4	4	4	4	4	4	4	4	4	100
Participant 4	4	1	2	1	4	3	3	3	3	2	65
Participant 5	4	4	4	3	3	4	3	4	4	4	92.5
Participant 6	4	1	0	0	4	4	4	0	3	0	50
Participant 7	4	4	4	4	4	4	0	4	4	4	90
Participant 8	4	4	4	4	4	4	4	4	4	2	95
Participant 9	2	3	3	4	3	3	3	2	4	4	77.5
Participant 10	2	4	4	4	4	4	3	4	4	4	92.5
Mean (SD)	3.5 (0.82)	3.2 (1.19)	3.1 (1.22)	3.1 (1.40)	3.5 (0.81)	3.6 (0.82)	3 (1.1)	3.3 (1.29)	3.6 (0.52)	2.8 (1.40)	79.28 (16.35)
Median (IQR)	4 (4-3.25)	4 (4-3)	3.5 (4-3)	4 (4-3)	4 (4-3)	4 (4-3.25)	3 (3.75-3)	4 (4-3.25)	4 (4-3)	3.5 (4-2)	80 (92.5-70)

Improvements Made Between the 2 Stages of the Study

After completing the test with the first 10 users, we realized that many users encountered difficulties entering medication without scanning the package (task 6). In addition, the users provided comments suggesting the need to facilitate the recording of medications. During stage 1, the manual recording of medications was carried out in a single page (Figure 9). To simplify this process, the ergonomist proposed splitting the process into 3 pages, each requiring a validation (Figures 10-12).

We also made some minor changes, such as modifying the localization and visibility of the reminder and information sections to make them more accessible and more visible from the main menu. Once the modifications were done, we began stage 2 of the study. However, we did not increase the font size in the app, despite the many comments we received as mentioned earlier. This did not seem necessary because the font on the smartphone can be changed in its own settings, allowing each user to adapt it to their desires and needs.

XSL•FO
RenderX

Figure 10. Screenshot of the rolling menu for new medication, before improvements were made, during stage 1 of the study. All the process was presented on one page. First, the patient had to insert the name of their medication and then the time of the intake. There was also a scan button to scan the barcode of the medication.

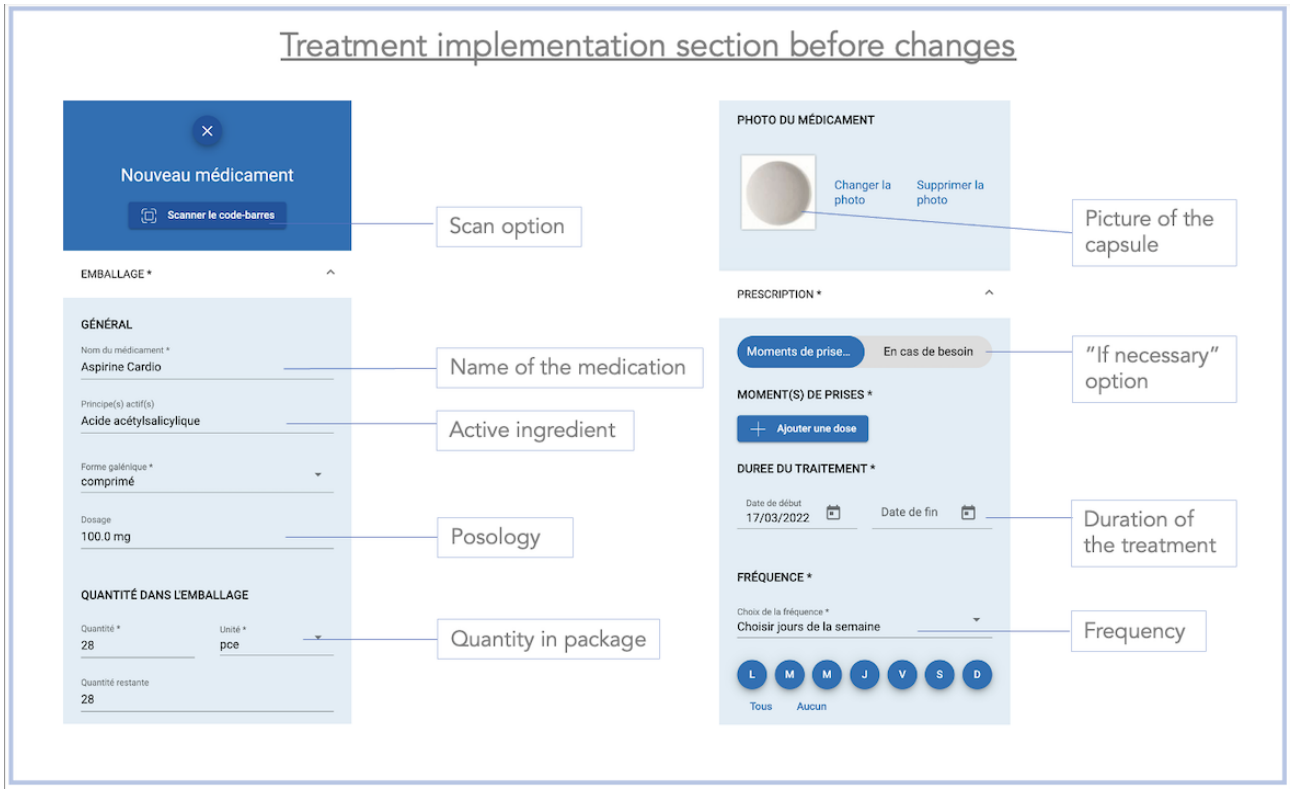
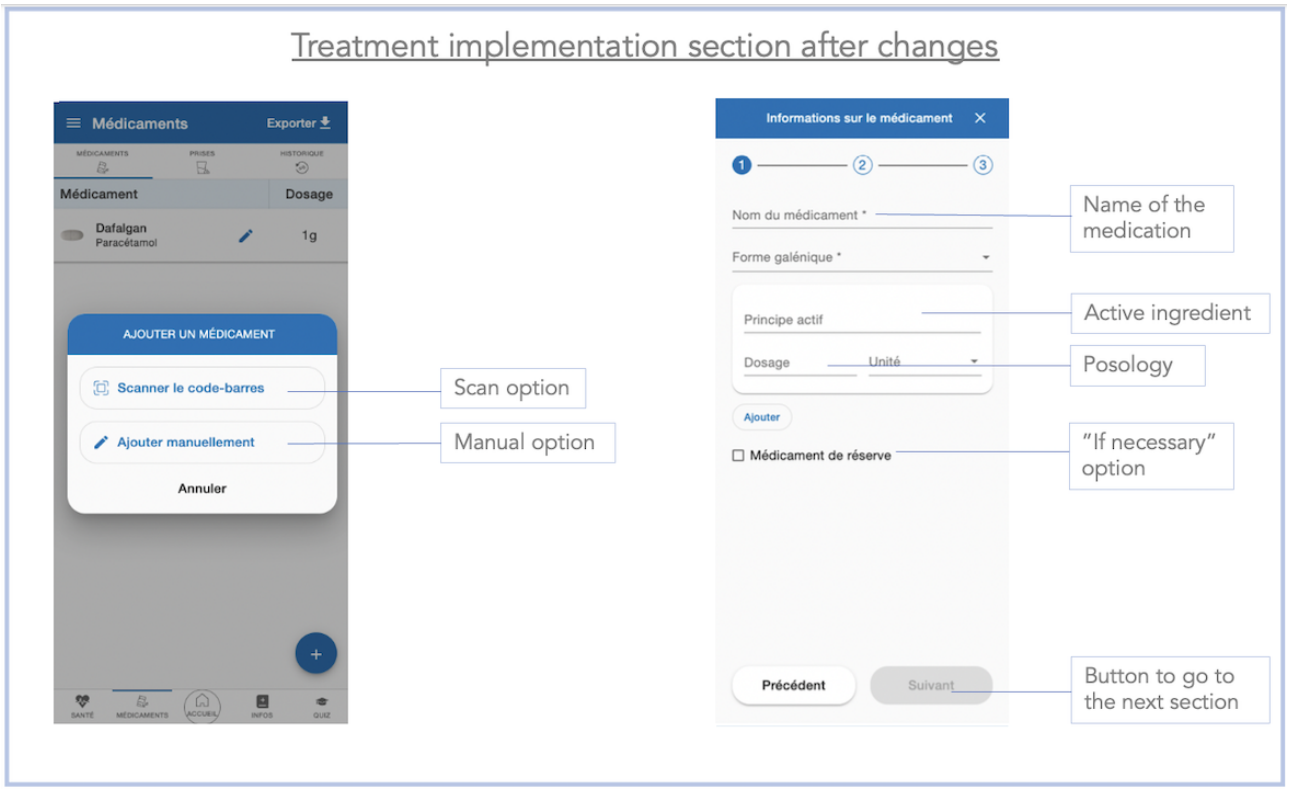
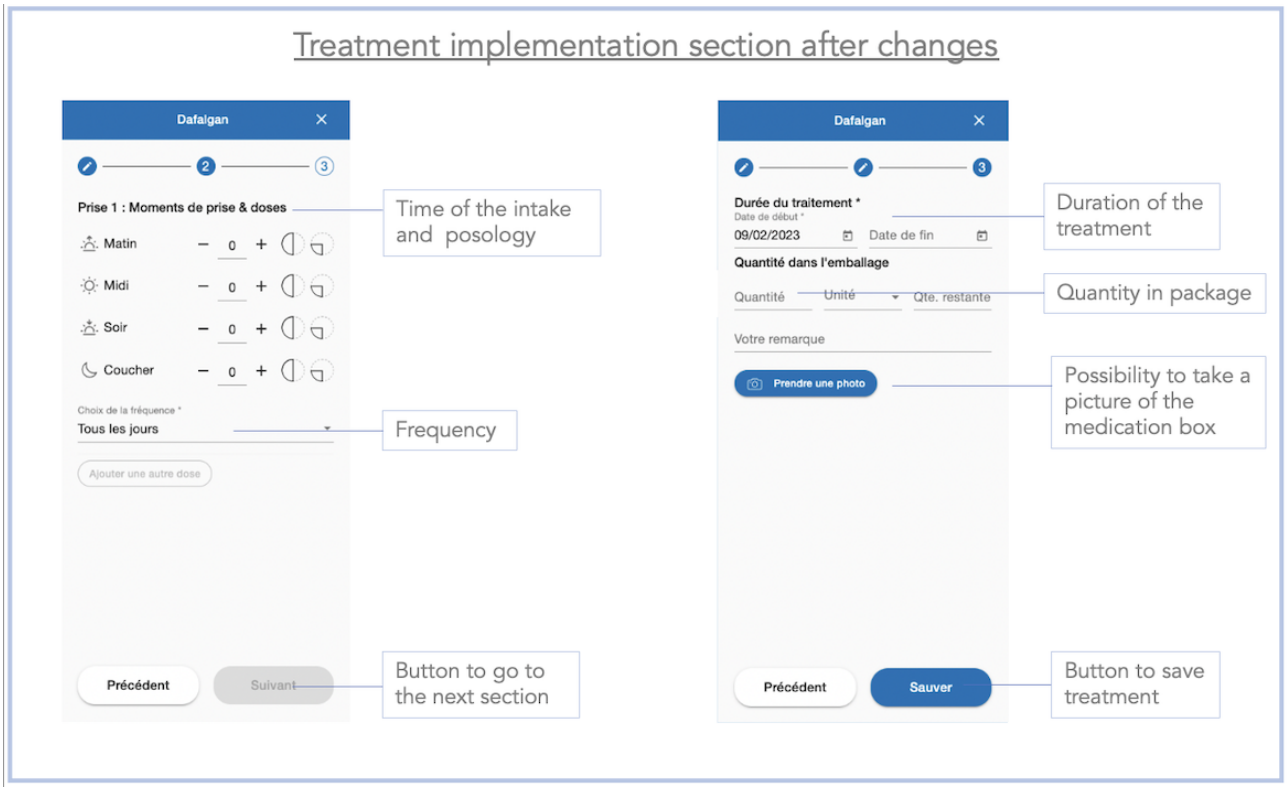


Figure 11. Screenshot of the medication intake section of the Cardiomeds app after the improvements were made during stage 2. First, the patient had to enter the name of the medication or scan the barcode. Then, they pressed the “next” button to go to the next section.



XSL•FO
RenderX

Figure 12. Screenshot of the medication section of the Cardiomeds app after implementation of the changes during phase 2. The patient reached this section after entering the name of the medication or after scanning the medication barcode. Here, the patient had to select the time to take the medication, the dosage, and the frequency.



Second Stage of the Study

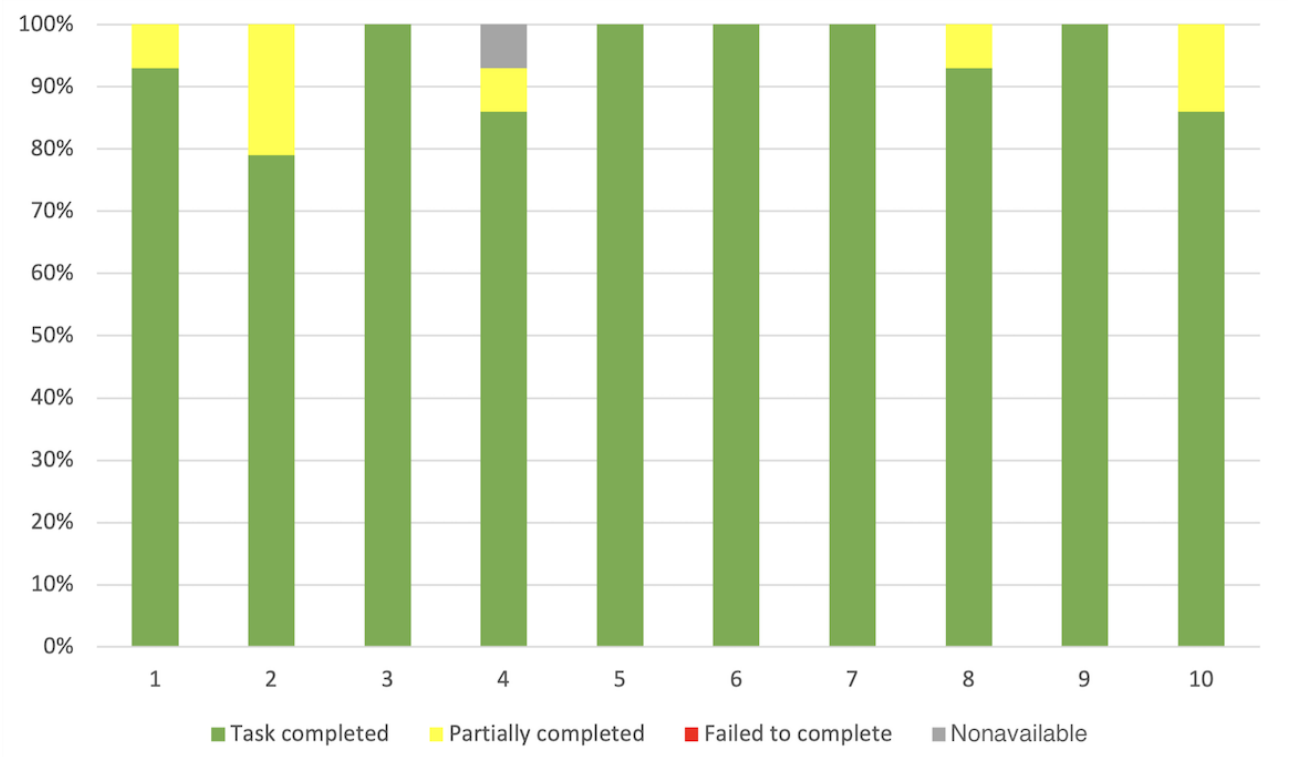
Quantitative Evaluation

Effectiveness per Participant

In stage 2, the overall completion rate of the test (tasks completed, partially completed, and failed to be completed) was 99.3% (139/140). Participant 4 did not perform task 14 during his test because of a software bug. None of the participants failed to complete any task. Only 50% (5/10) of the participants needed help to complete at least one task. Participant 2 needed the most help to finish the test, with 20% (3/14) of the tasks completed with help. All these measurements are exposed and supported in Figure 13.

XSL•FO
RenderX

Figure 13. Task completion rate per participant for the 14 assigned tasks during the second stage of the study. "Task completed" represents the percentage of tasks successfully completed by a participant. "Failed to complete" is the percentage of tasks that participants failed to complete. "Partially completed" is defined as tasks completed, but with difficulty that could have been solved with the help of the instructor in 5 minutes or less. "Nonavailable" represents the percentage of missing data when a task could not be started and evaluated.



Task Success Distribution per Task

In stage 2, overall 93.6% (131/140) of tasks were completed successfully, 5.7% (8/140) were completed partially, and 0% (0/140) failed. A total of 64% (9/14) of tasks 2, 3, 4, 7, 8, 9, 10, 12, and 14 were completed successfully by all participants (100%, 10/10), followed by task 13 (9/10, 90%). A total of 29% (4/14) of the tasks 1, 6, 11, and 13 were completed with difficulty with a rate ranging from 70% to 90%. In this stage of the study, no task completion error was made. Task 5 (enter medication) is the only one with 10% (1/10) missing data because of a technical problem during the test. During this task, the photo scan of the phone did not work. All these measurements are exposed and supported in Figure 14.

XSL•FO
RenderX

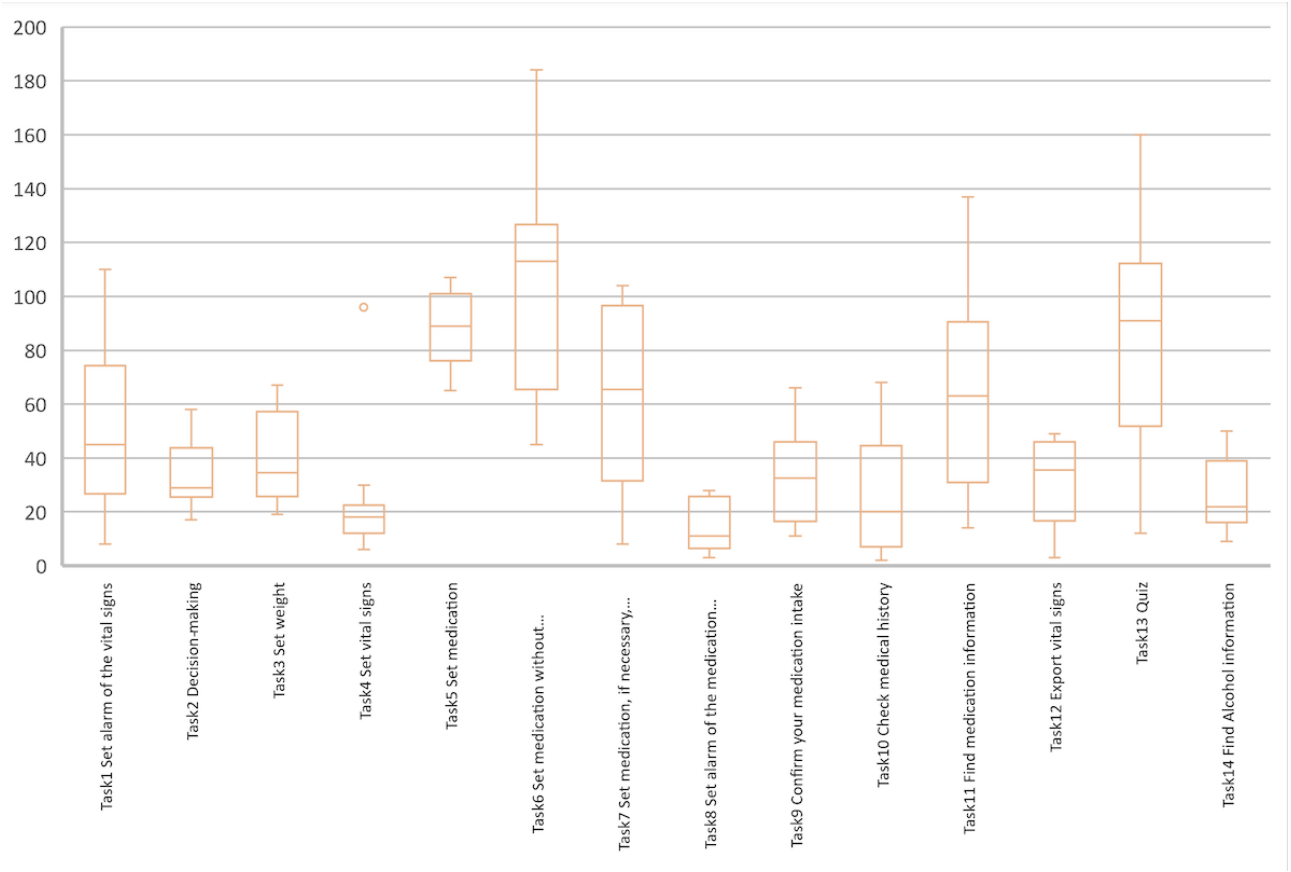
Figure 14. Task success distribution per task (N=10 participants) for the second stage of the study. "Complete" represents the percentage of participants who completed the task with ease. "Partially complete" represents the percentage of participants who completed the task with difficulty that could have been solved with the help of the instructor in 5 minutes or less. "Failed to complete" is defined as the percentage of participants who failed to complete the task. "Nonavailable" represents the percentage of missing data when a task could not be started and evaluated.

Time on Task per Study Task

In the second part of the study, entering medication details (task 6) remained the longest in duration with an average of 103.10 (SD 42.76) seconds (Figure 15). The task that was completed the fastest was task 8 with an average of 21.10 (SD 25.69) seconds.

XSL•FO
RenderX

Figure 15. Boxplot of the time taken (in seconds) to perform each task in stage 2.



Qualitative Evaluation

Five Open Questions Questionnaire

In stage 2, 80% (8/10) of participants answered that their impression of the app was "positive," 0% answered negatively, and 20% (2/10) did not answer this question.

Furthermore, 100% (10/10) of the participants answered that they would like to use the app.

During the second part of the study, participants suggested making the validation of medication intake more explicit, adding some emergency numbers inside the app, and putting the name of the drug next to the name of the active ingredient of the drug to increase patients' understanding of their treatments (Textbox 5).

Textbox 5. List of comments made by participants (N=10) that are useful in the short- and long-term for improving the app during the second stage of the test.

Useful comments for short-term improvements

- One (10%) participant suggested making the validation of drug intake more explicit
- One (10%) participant wished to have the emergency number and the number of his physician, his cardiologist, and his pharmacy inside the app

Useful comments for long-term improvements

- One (10%) participant wanted to have the name of the medication next to the active principle in the medication information

Error Classification With the Bastien and Scapin Criteria

In the second part of the study, we recorded a total of 15 problems. We classified them into 7 categories according to Bastien and Scapin's criteria [31] (Table 4). The most problematic were tasks 9 (confirm your medication intake) and 12 (export vital signs), with 4 (N=10, 40%) users experiencing problems with each task. Participants reported that the validation of daily medication intake was not explicit enough. The participants who encountered a problem with task 12 found that the "export" button was too small and therefore not legible and that the location of this button was not well placed.

XSL•FO
RenderX

Table 4. List of problems encountered by participants (N=10) of the second part of the study during the 14 tasks and description, categorization, and frequency of occurrence of these problems during task completion.

Item and problem	Category (number according to the Bastien and Scapin classification)	Description of the problem	Task in which it occurs (frequency of occurrence)
1	1. Guidance	Five (50%) users found the front text size too small and the localization of some symbols not easy to find	Enter medication (1) and export vital sign (4)
2	1.3 Immediate feedback	Four (40%) participants did not easily understand how to validate a medication intake	Confirm your medication intake (4)
3	7. Significance of codes	One (10%) participant did not understand some names of drugs because the name of the medication was not written next to the active substance	Find medication information (1)
4	1. Guidance	One (10%) user did not find the medication information in the principal menu	Find medication information (1)
5	5.2 Error message quality	When entering the date, two (20%) participants did not write it correctly (separated by "/" symbols and not "." symbols). Therefore, the date could not be validated, and no explanation appeared	Enter weight (2)
6	5.2 Error message quality	One (10%) participant made a mistake while entering his vital signs and could not go back, so he had to restart the task from the beginning	Enter vital signs (1)
7	2.2 Concision	One (10%) participant was troubled by the fact that there are two places to find information on drugs	Find medication information (1)

Satisfaction: SUS Questionnaire

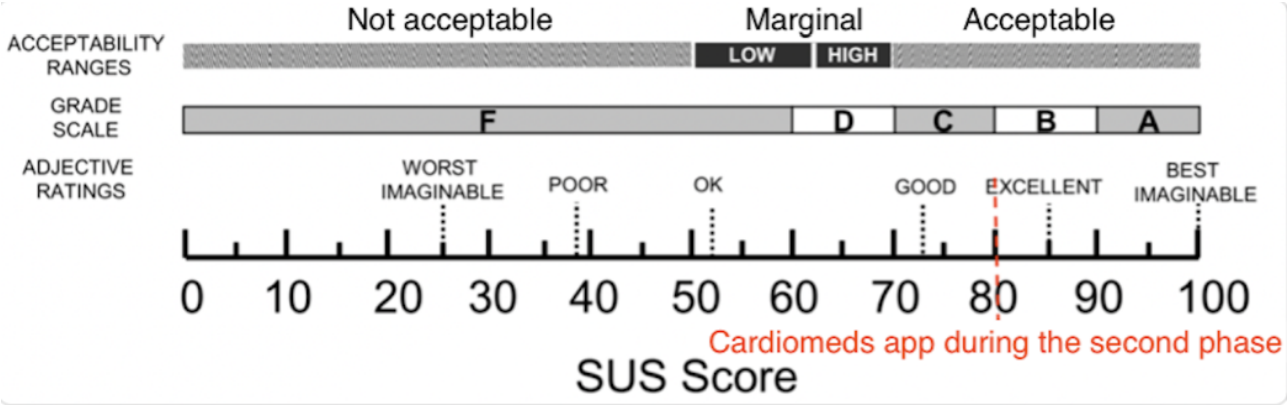
The mean SUS score of the second stage of the study was 80 (SD 28.3), described in Table 5 and classified in Figure 16.

Table 5. Responses to the System Usability Scale (SUS) questionnaire in the second stage of the study.

	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7	Question 8	Question 9	Question 10	SUS score (sum×2.5; maximum 100)
Participant 1	3	3	3	4	3	3	3	3	3	3	77.5
Participant 2	3	0	3	3	3	4	2	3	3	0	60
Participant 3	3	4	4	4	4	4	4	4	4	4	97.5
Participant 4	3	4	3	4	4	4	4	4	4	3	92.5
Participant 5	4	4	4	4	4	4	4	4	4	4	100
Participant 6	4	3	3	3	3	3	3	3	3	3	77.5
Participant 7	3	4	4	4	3	4	3	4	4	4	92.5
Participant 8	3	4	3	4	3	4	3	4	3	4	87.5
Participant 9	4	3	3	4	3	4	3	4	4	1	82.5
Participant 10	3	1	1	1	3	3	3	2	3	1	52.5
Mean (SD)	3.3 (0.47)	3 (1.34)	3.1 (0.83)	3.5 (0.93)	3.3 (0.46)	3.7 (0.5)	3.2 (0.6)	3.5 (0.68)	3.5 (0.52)	2.7 (1.52)	80 (28.3)
Median (IQR)	3 (3.75-3)	3.5 (4-3)	3 (3.75-3)	4 (4-3.25)	3 (3.75-3)	4 (4-3.25)	3 (3.75-3)	4 (4-3)	3 .5 (4-3)	3 (4-1.5)	80 (90-70)

XSL•FO
RenderX

Figure 16. System Usability Scale (SUS) score on the SUS [35] of the Cardiomeds app at stage 2.



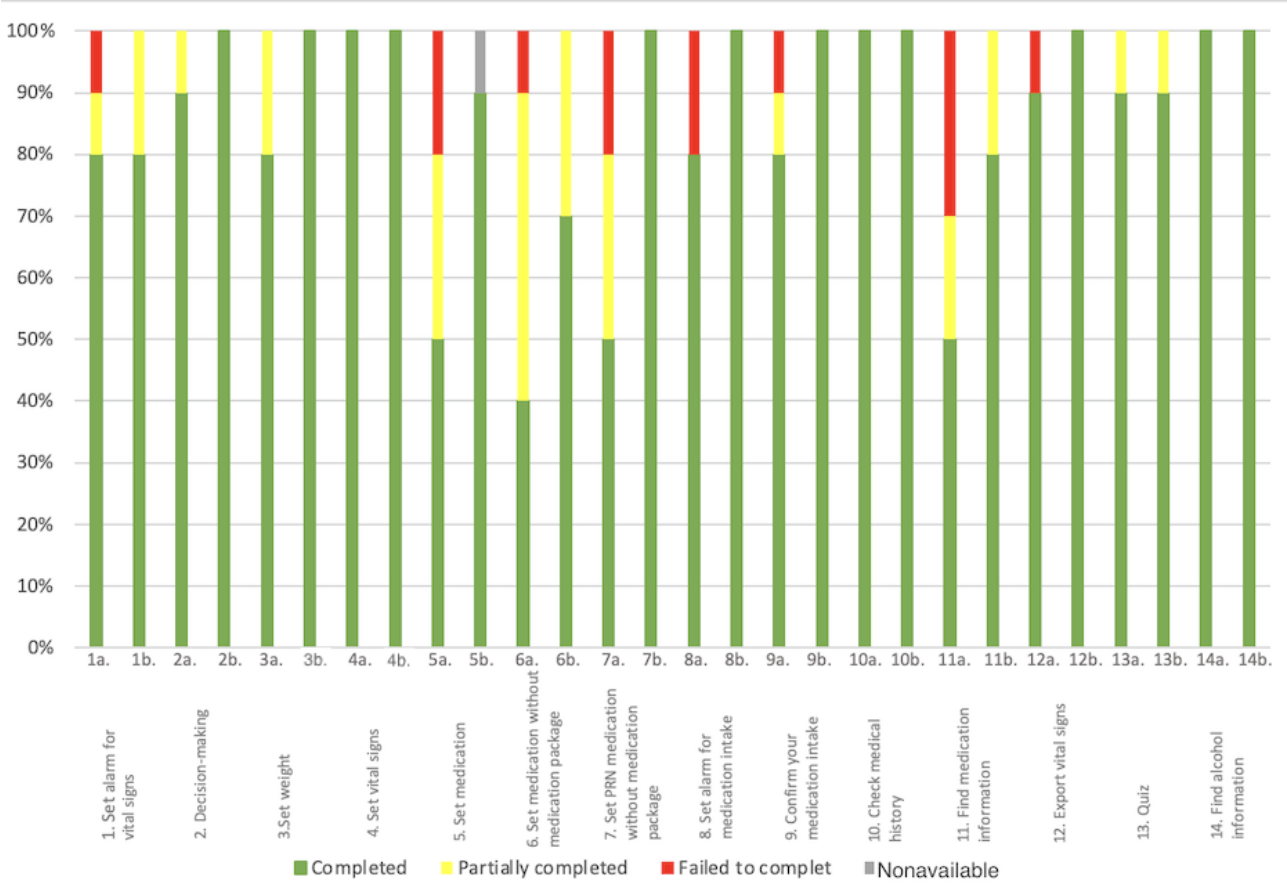
Comparison of First and Second Stages

Quantitative Evaluation

Task Success

On average, participants completed 86% (12/14; SD 0.19%) of tasks without any help overall. Conversely, on average, participants needed help with 14% (SD 0.11%) of tasks. As depicted in Figure 15, tasks 4, 10, and 14 maintained a success rate of 100% (20/20) in both stages of the study. Tasks 2, 3, 7, 8, 9, and 12 had a successful completion rate without any help of 100% in the second stage of the study. We also noticed that the failure rate for all tasks combined was reduced to 0% in the second stage of the study. All these measurements are exposed and supported in Figure 17.

Figure 17. Task success distribution per task (N=20 participants) for the overall study. "Completed" represents the percentage of participants who completed the task with ease. "Partially completed" represents the percentage of participants who completed the task with difficulty that could have been solved with the help of the instructor in 5 minutes or less. "Failed to complete" is defined as the percentage of participants who failed to complete the task. "Nonavailable" represents the percentage of missing data when a task could not be started and evaluated.



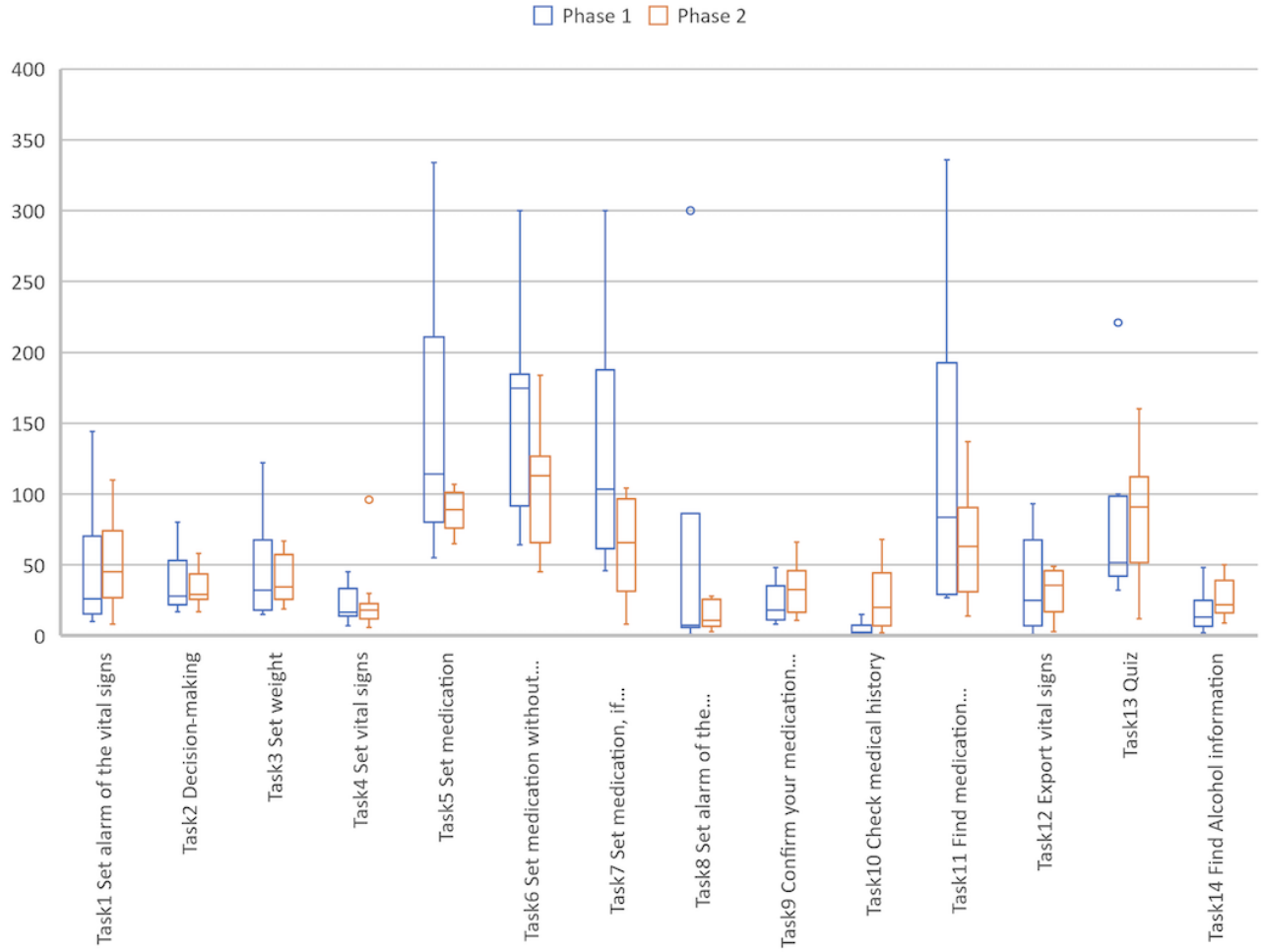
Time on Task per Study Task

The mean overall time on task for all tasks was 57.19 (SD 40.05) seconds. Tasks 5, 6, and 7 had a longer duration than the other tasks in both stages of the study (Figure 18). Nevertheless, task 11 was also one of the longest during the first stage. These findings showed that the task with the most failures (ie, task 6 [enter medication without medication package]) was also the

XSL•FO
RenderX

most time-consuming in both stages of the study, although the time taken was reduced during stage 2.

Figure 18. Boxplot of the comparison of the efficiency of the 2 stages of the study. Stage 1 is shown in blue and stage 2 is shown in orange.



Qualitative Evaluation

Five Open Questions Questionnaire

Regarding the qualitative part of the study, 80% (16/20) of participants responded positively to our open questions about general feelings in both stages of the study. Between the 2 stages, the number of participants who had negative feelings was reduced from 1 participant to 0 participant. In addition, 85% (17/20) of participants were inclined to use the app once it became available, indicating its ease of use. Another comment that emerged in both stages of the study was the wish to access various contact numbers (4/20, 20%), such as the emergency, physician, cardiologist, or pharmacy.

Error Classification

The first stage of evaluation highlighted various ergonomic problems as defined by the Bastien and Scapin criteria [31]. The reduction, from 22 problems of 6 different types in the first stage of the study to 15 problems of 5 different types identified during the second stage, demonstrated the benefits of the improvements made to the app.

We observed that in both stages of the study, the most frequent problem remained the problem of guidance (first stage; 8/22, 36%, and second stage; 6/15, 40%).

It seems that the change in localization of the reminder and information section led to a reduction of guidance problems, although 2 different types of problems remained, including 1 new problem (Table 6). Given that guidance problems appeared several times in both stages of the study for tasks 8, 11, and 12, we could improve the app further by making the buttons of the reminder and the information section even more visible in a future version (reminder, export vital signs, and find medication information). A recommendation would be to add a label under the reminder button and increase its size. As for the button in the information section, we would like to change its color and increase its size.

XSL•FO
RenderX

Table 6. Type of problems during stage 1, number of occurrences, number of new problems, and total type of problems in stage 2.

Category (number according to the Bastien and Scapin [31] classification)	Stage 1	Stage 2		
	Types of problems	Solved problems	New problems	Types of problems
1. Guidance	2	1	1	2
5.2 Error message quality	1	0	1	2
7. Significance of codes	3	2	0	1
8. Compatibility	1	1	0	0
2.2 Information density	1	1	0	0
1.3 Immediate feedback	1	0	0	1
2.2 Concision	0	0	1	1

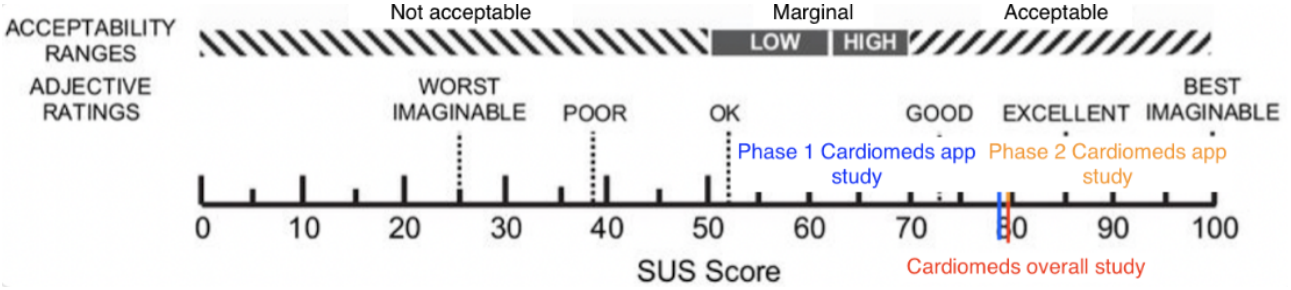
Moreover, we observed that the problem of the significance of codes decreased from 22% (5/22) to 7% (1/15) and from 3 such problems to only 1 (Table 6). These problems appeared during the task in which the participants had to scan the medication package, which we have made easier and more visible within the app, but also about the appellation by the component of certain medications that we have also tried to popularize as much as possible to reach a wider audience, for example, we have inserted the names of the drugs (metozerok) next to the class of the drug (β-blocker).

Problems related to error message quality (from 1/22, 5% to 3/15, 20%) and immediate feedback (from 1/22, 5% to 4/15, 26%) increased, although there was no new problem of the immediate feedback type, unlike the error message quality for which a new type of problem appeared. These problems did not increase significantly and are strongly dependent on the participant samples. Finally, a new problem of concision was mentioned during stage 2 by a single participant concerning the location of information concerning medications which could be found in 2 different places.

Satisfaction: SUS Questionnaire

Averaging the 2 results of the SUS questionnaire gave a score of 80 for the entire study. This shows that the usability of the Cardiomeds app was perceived as good in both stages of the study with a slight improvement during the second stage after some modifications. The scale in Figure 19 is used to interpret these results [34].

Figure 19. Cardiomeds score on the System Usability Scale (SUS) for the overall study.



Discussion

Principal Findings

Participants took 18 minutes on average to complete all 14 tasks, which means just over 1 minute per task. However, certain tasks took much longer than others. The shortest task took an average of 4.80 (SD 4.60) seconds to be completed in the first stage (task 10) and 21.10 (SD 25.69) seconds in the second stage (task 8). One possible reason that tasks 8 and 10 were the shortest in the study could be that these 2 tasks required minimal navigation within the app. In addition, for task 8 (set alarm for medication intake), participants had already set an alarm for task 1 (set an alarm for vital signs), which likely made this task easier. Similarly, task 10 (check medication history) followed task 9 (confirm your medication intake), where the same menu also provided access to the medication intake history button. Finally, because both tasks followed task 6, which was the longest in the study, participants may have become more familiar with the app's navigation, helping them complete the later tasks more efficiently. Furthermore, between the 2 stages of the study, we modified the localization and visibility of the reminder icon to make it more accessible and more visible from the main menu. This might explain why task 8 (set alarm for the medication intake) was the fastest task. However, we were unable to determine why the duration of task 10 (check medical history) was longer in the second part of the study than the first because we did not make any change that could influence retrieval time.

The maximum average time was 154.40 (SD 68.08) seconds during the first stage (task 6) and 103.10 (SD 42.76) seconds during the second stage (task 6). The entry of medications (task 6) took the longest time in both stages of the study, although we observed an improvement between the first and the second stages, suggesting that the app had been improved. Improving the medication entry process was crucial to reduce the time patients spent inputting or modifying their treatment in the app. We know that patients with HF with reduced left ventricular

XSL•FO
RenderX

ejection fraction usually take at least 4 different medications. In addition, the older the patients, the more they suffer from different comorbidities, such as type 2 diabetes [35] or hypertension [36], which further increases the number of medications required. Furthermore, the treatment for patients with HF is likely to change often, so we wanted to make this part as fluid, ludic, and easy to use as possible. Knowing these considerations, the time spent on the app could significantly increase which could discourage the target audience from using it. Therefore, it was important to improve the ergonomics of drug recording to make this process as short as possible. Feedback collected on task 6 (enter medication without medication package) led to app modifications aimed at simplifying the medication entry process. The main change was to divide the drug recording process into 3 different pages rather than having a single scrollable page. Nevertheless, we believe that after repeated use, the time taken will diminish, and the patient will become increasingly comfortable to use the app. Also, although entering medications is an important initial process in using the app, it is not needed in daily use.

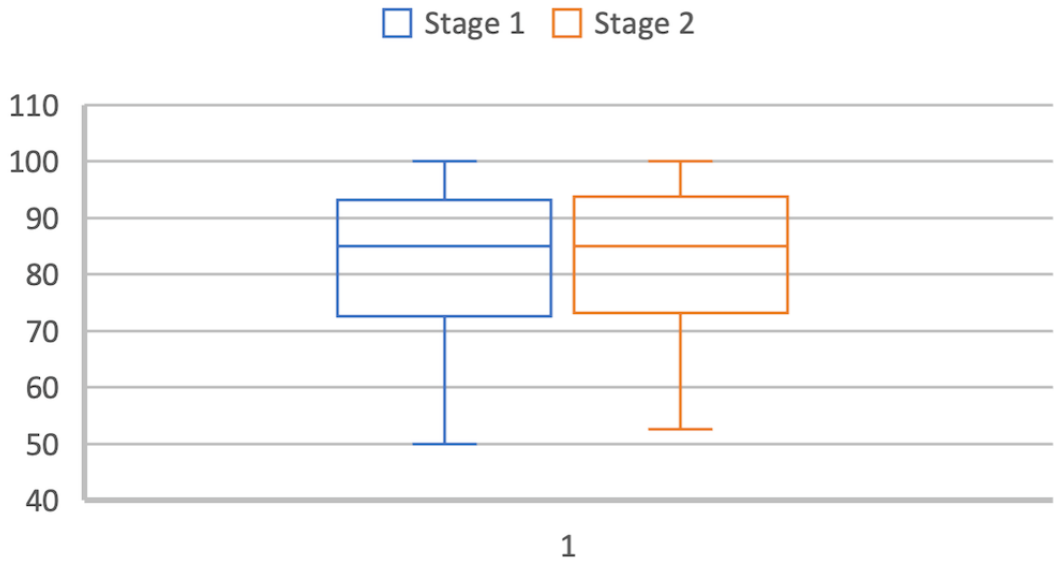
Concerning the quantitative results of the study, we observed that the mean overall success rate, which represents the effectiveness (completed the task without any help and completed the task with help) was 77.1% (108/140) for the first stage and 93.6% (131/140) for the second stage of the study. Furthermore, only 30% (3/10) of the participants in the first stage of the study managed to complete all tasks without any help versus 50% (5/10) during the second stage. We wonder if the differences observed between the 2 stages are partly due to the difference in characteristics of the participants in the 2 groups. Stage 1 included more participants aged between 71 and 80 years (1/10, 10%) than stage 2 (0/10, 0%).

We observed that 8 tasks (8/14, 57%) were failed by at least one participant during stage 1 compared to none during the second stage. The tasks failed or completed with difficulty in the first stage of the study included tasks 5 (5/10, 50%), 6 (6/10, 60%), and 7 (5/10, 50%). All these tasks were related to the recording of the medication. These same tasks induced fewer problems during the second stage of the study, with only 30% (3/10) of the participants completing task 6 with difficulty, and none of the participants making any errors for tasks 5 and 7. As these tasks aimed at the recording of medications, we can reasonably assume that the changes made to the app after the first stage contributed to the reduction of problems encountered during the second stage of the study.

Finally, we observe in Figure 19 that the efficiency of task 10 (check medical history) seems worse in the second stage of the study (mean time for the overall participants 27.10, SD 21.70 seconds) than in the first stage (mean time for the overall participants 4.80, SD 4.60 seconds). On the basis of our observations, we suspect that during the first stage of the study, participants took more time to complete the previous tasks, which gave them more time to navigate through the app. Thus, we hypothesize that the medication history page was already localized by the participants in stage 1.

It is surprising to observe that the significant increase in effectiveness per participant, from 50% (5/10) in the first stage of the study to 100% (10/10) in the second stage, is not reflected in an increased SUS score. Indeed, the Cardiomeds app was first evaluated with usability measured as “good” with a SUS score of 79% on average. It improved by 1% to reach 80% after modifications were made to the app between the 2 stages of the study. Nevertheless, when we look at the distribution of scores in the 2 groups in Figure 20, we see that they are very similar. We concluded that the participants mainly judged the concept of the app rather than the difficulty of the tasks carried out during the usability tests through the SUS. Therefore, it is logical that despite the changes made to the app, the SUS score did not change much.

Figure 20. System Usability Scale score distribution in both groups.



During the tests, we collected various participants’ comments. In the first stage of the study, the most common comment (2/10, 20%) was the extremely small font size, which may be a problem for older users. We did not change the font size in the app because it is possible to do so globally in the settings of

XSL•FO
RenderX

each device. During the second stage of the study, this remark did not emerge again, although we did not change the font size.

Furthermore, we can observe in Table 6 that between the 2 stages of the study, the problems of information density and compatibility were solved. Indeed, these problems mainly appeared while entering medications without a medication package (task 6), whose interface was improved between the 2 stages of the study. As mentioned earlier, this stage has been improved because we believe that the recording of medications must be simple given the number of treatments necessary and their regular change depending on the clinical condition of the patients and the course of their illness. We decided not to categorize guidance problems because it was not clear whether the guidance problems were localization problems or legibility problems. Indeed, the guidance problems were found in task 8 (set alarm for medication intake), 11 (find medication information), and 12 (export vital signs). It was not clear if the button was too small or poorly located, which would not make it visible enough.

It is important to point out that despite the availability of mHealth health interventions, most patients still wish to have follow-up appointments with their physician to maintain supervision and human contact. Many patients express the fear of being harmed when relying only on a health app [37]. However, with the current shortage of physicians and nurses in Switzerland's urban areas [38], the management of these patients is complex. The follow-up made possible with an app, such as Cardiomeds, could probably provide relief and help physicians to focus more on the relationship with the patient during medical appointments. Moreover, when patients are involved in their care and carry out self-checks, they are also more involved in their health, thus enabling in-depth monitoring of their illness to best prevent cardiac decompensations and avoid hospitalizations. Therefore, it is clear that physicians need to add eHealth to their daily practice by integrating it into their management, to offer optimal long-term follow-up.

Strengths

The first strong point of our study was that although common guidelines assume that a sample of 5 users is sufficient to reveal 85% of usability problems and 15 users is sufficient to discover almost 97% of problems [39], we evaluated the usability of the Cardiomeds app with 2 groups of 10 people each. In addition, we split the study into 2 stages, which enabled us to make modifications based on the feedback and then test the app with a new group to see if the changes improved the app's usability.

Second, we used mixed methods combined with usability testing methods to report and analyze quantitative and qualitative data in this study, which is recognized as a strong point in studies analyzing the usability of apps [40,41].

Finally, in our usability testing protocol, users had to complete tasks like those they would perform alone at home. Studies have shown that this is a good way to test an app's usability. In addition, the inclusion criteria for our study were broad, allowing a large panel of patients with HF to participate. In addition, even if most participants did not have an iPhone, it was still possible for Android users to test the app.

Limitations

A few limitations need to be considered. First, we can see that the most represented age group in our study (50-60 years) is younger than the age of the audience targeted by the app. Indeed, HF is more prevalent in patients >65 years of age [42]. Second, in both groups, women were underrepresented, with an average of 25% (5/20) of women, which is less than the prevalence of HF in women, which reaches 50% [42]. Third, only one investigator oversaw data collection in this study. Indeed, the usability tests, the SUS score, and the 5 open questions questionnaire were evaluated by one investigator, and all the recorded data were analyzed and timed by a single investigator. As only one person measured the times, there is a risk of bias in the results; however, this would be minimal given that the same investigator performed all the recordings.

Finally, the SUS questionnaire has limitations in assessing the usability of apps. Indeed, the SUS questionnaire does not provide specific insights into which parts of the app may cause issues for users. Moreover, it is a subjective measure of perceived usability. By adding some open questions to this questionnaire, we were able to obtain more information, but this does not reduce the inherent deficiencies of the SUS [43,44].

Conclusions

According to the SUS score obtained in both stages of the study, our usability study rates the Cardiomeds app as "good," with a slight improvement of 1% between the 2 stages of the study, to reach 80% in stage 2. The effectiveness per participant (the number of tasks completed or partially completed by each participant), went from 50% in stage 1% to 100% in stage 2. The most time-consuming task in both stages of the study was manual medication recording (task 6). Nevertheless, the improvement made after the first stage of the study enabled a reduction in the necessary time to complete this task. The most frequent ergonomic problems according to the Bastien and Scapin criteria remained problems of location and legibility within both stages of the study, with a slight improvement in stage 2. The problem of information density was solved because of the improvements made to this app by the IT professionals regarding medication recording. Most participants who tested the app recommended it and said they would use it when available. The multidisciplinary care required by patients with chronic HF is often complex because of the presence of many comorbidities. Therefore, it would be interesting to enable them to better understand their disease and increase their involvement, improve their therapeutic adherence, and finally enhance their quality of life through the future use of this mHealth app.

XSL•FO
RenderX

Acknowledgments

This study would not have been possible without the support of the University Hospital of Geneva. Funding for this usability study was provided by the cardiology department of the University Hospital of Geneva, and we thank the University of Geneva for their financial support.

We used the generative artificial intelligence tool, ChatGPT by OpenAI [45] to translate from French to English and to rephrase some sentences from the Introduction and Discussion of the study.

Conflicts of Interest

None declared.

References

1. Savarese G, Becher PM, Lund LH, Seferovic P, Rosano GM, Coats AJ. Global burden of heart failure: a comprehensive and updated review of epidemiology. Cardiovasc Res 2023 Jan 18;118(17):3272-3287 [FREE Full text] [doi: 10.1093/cvr/cvac013] [Medline: 35150240]
2. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, ESC Scientific Document Group. 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J 2021 Sep 21;42(36):3599-3726 [FREE Full text] [doi: 10.1093/eurheartj/ehab368] [Medline: 34447992]
3. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187 [FREE Full text] [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
4. Jonkman NH, Westland H, Groenwold RH, Ågren S, Anguita M, Blue L, et al. What are effective program characteristics of self-management interventions in patients with heart failure? An individual patient data meta-analysis. J Card Fail 2016 Nov;22(11):861-871 [FREE Full text] [doi: 10.1016/j.cardfail.2016.06.422] [Medline: 27374838]
5. Visseren FL, Mach F, Smulders YM, Carballo D, Koskinas KC, Bäck M, ESC National Cardiac Societies, ESC Scientific Document Group. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice. Eur Heart J 2021 Sep 07;42(34):3227-3337 [FREE Full text] [doi: 10.1093/eurheartj/ehab484] [Medline: 34458905]
6. Achury Saldaña DM, Gonzalez RA, Garcia A, Mariño A, Aponte L, Bohorquez WR. Evaluation of a mobile application for heart failure telemonitoring. Comput Inform Nurs 2021 May 17;39(11):764-771. [doi: 10.1097/CIN.0000000000000756] [Medline: 33993153]
7. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JG. Structured telephone support or non-invasive telemonitoring for patients with heart failure. Cochrane Database Syst Rev 2015 Oct 31;2015(10):CD007228 [FREE Full text] [doi: 10.1002/14651858.CD007228.pub3] [Medline: 26517969]
8. McAlister FA, Stewart S, Ferrua S, McMurray JJ. Multidisciplinary strategies for the management of heart failure patients at high risk for admission: a systematic review of randomized trials. J Am Coll Cardiol 2004 Aug 18;44(4):810-819 [FREE Full text] [doi: 10.1016/j.jacc.2004.05.055] [Medline: 15312864]
9. Zhu Y, Gu X, Xu C. Effectiveness of telemedicine systems for adults with heart failure: a meta-analysis of randomized controlled trials. Heart Fail Rev 2020 Mar;25(2):231-243 [FREE Full text] [doi: 10.1007/s10741-019-09801-5] [Medline: 31197564]
10. Lefler LL, Rhoads SJ, Harris M, Funderburg AE, Lubin SA, Martel ID, et al. Evaluating the use of mobile health technology in older adults with heart failure: mixed-methods study. JMIR Aging 2018 Dec 04;1(2):e12178 [FREE Full text] [doi: 10.2196/12178] [Medline: 31518257]
11. Gandapur Y, Kianoush S, Kelli HM, Misra S, Urrea B, Blaha MJ, et al. The role of mHealth for improving medication adherence in patients with cardiovascular disease: a systematic review. Eur Heart J Qual Care Clin Outcomes 2016 Oct 01;2(4):237-244 [FREE Full text] [doi: 10.1093/ehjqcco/qcw018] [Medline: 29474713]
12. Heiney SP, Donevant SB, Arp Adams SA, Parker PD, Chen H, Levkoff S. A smartphone app for self-management of heart failure in older African Americans: feasibility and usability study. JMIR Aging 2020 Apr 03;3(1):e17142 [FREE Full text] [doi: 10.2196/17142] [Medline: 32242822]
13. Bakogiannis C, Tsarouchas A, Mouselimis D, Lazaridis C, Theofillogianakos E, Billis A, et al. A patient-oriented app (ThessHF) to improve self-care quality in heart failure: from evidence-based design to pilot study. JMIR Mhealth Uhealth 2021 Apr 13;9(4):e24271 [FREE Full text] [doi: 10.2196/24271] [Medline: 33847599]
14. Masterson Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the mobile application rating scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74 [FREE Full text] [doi: 10.2196/mhealth.5882] [Medline: 27302310]
15. WOW ME 2000mg - heart failure self-management tool for patients and caregivers. CNET Download. URL: https://download.cnet.com/wow-me-2000mg-heart-failure-self-management-tool-for-patients-and-caregivers/3000-2129_4-76773751.html [accessed 2024-09-13]

XSL•FO
RenderX

16. WebMD: symptom checker. Google Play. URL: https://play.google.com/store/apps/details?id=com.webmd.android&hl=en [accessed 2024-09-13]
17. 8 smart mobile apps and gadgets for heart health. PCNA. 2018. URL: https://pcna.net/8-smart-mobile-apps-and-gadgets-for-heart-health/ [accessed 2024-09-13]
18. Mertens A, Brandl C, Miron-Shatz T, Schlick C, Neumann T, Kribben A, et al. A mobile application improves therapy-adherence rates in elderly patients undergoing rehabilitation: a crossover design study comparing documentation via iPad with paper-based control. Medicine (Baltimore) 2016 Sep;95(36):e4446 [FREE Full text] [doi: 10.1097/MD.0000000000004446] [Medline: 27603339]
19. Park JY, Li J, Howren A, Tsao NW, De Vera M. Mobile phone apps targeting medication adherence: quality assessment and content analysis of user reviews. JMIR Mhealth Uhealth 2019 Jan 31;7(1):e11919 [FREE Full text] [doi: 10.2196/11919] [Medline: 30702435]
20. Cajita MI, Gleason KT, Han HR. A systematic review of mHealth-based heart failure interventions. J Cardiovasc Nurs 2016;31(3):E10-E22 [FREE Full text] [doi: 10.1097/JCN.0000000000000305] [Medline: 26544175]
21. Bertolazzi A, Quaglia V, Bongelli R. Barriers and facilitators to health technology adoption by older adults with chronic diseases: an integrative systematic review. BMC Public Health 2024 Feb 16;24(1):506 [FREE Full text] [doi: 10.1186/s12889-024-18036-5] [Medline: 38365698]
22. Ahmad NA, Mat Ludin AF, Shahar S, Mohd Noah SA, Mohd Tohit N. Willingness, perceived barriers and motivators in adopting mobile applications for health-related interventions among older adults: a scoping review protocol. BMJ Open 2020 Mar 16;10(3):e033870 [FREE Full text] [doi: 10.1136/bmjopen-2019-033870] [Medline: 32184309]
23. Étude digital seniors. Pro Senectute Schweiz. URL: https://www.prosenectute.ch/fr/espace-pro/publications/etudes/digital-seniors.html [accessed 2023-03-01]
24. Trois Suisses sur quatre ont adopté le smartphone. Comparis. URL: https://fr.comparis.ch/comparis/press/medienmitteilungen/artikel/2016/telecom/smartphone-studie-2016/smartphone-verbreitungsstudie-2016 [accessed 2023-03-01]
25. Usability testing 101. Nielsen Norman Group. URL: https://www.nngroup.com/articles/usability-testing-101/ [accessed 2024-09-13]
26. Test d'utilisabilité: Réaliser un test d'utilisabilité (ou Test utilisateur) afin d'évaluer un logiciel ou une application. Usabilis: Conseil UX et Ergonomie Digitale. URL: https://www.usabilis.com/ressources/test-d-utilisabilite/ [accessed 2024-03-17]
27. Usability 101: introduction to usability. Nielsen Norman Group. URL: https://www.nngroup.com/articles/usability-101-introduction-to-usability/ [accessed 2024-03-17]
28. Ehrler F, Gschwind L, Meyer P, Blondon K. Swiss-Meds: an app fostering medication adherence of Swiss patient. Stud Health Technol Inform 2019;259:71-76. [Medline: 30923276]
29. Aubagna M. Comment utiliser SUS pour mesurer l'expérience utilisateur ? Skeepers. URL: https://skeepers.io/fr/blog/questionnaire-system-usability-scale-experience-client/ [accessed 2023-03-01]
30. System usability scale (SUS). U.S. General Services Administration. URL: https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html [accessed 2023-03-01]
31. Les critères Heuristiques de Bastien et Scapin. Le bloc-notes, UX & Design d'expérience utilisateur. URL: https://blocnotes.iergo.fr/concevoir/les-criteres-heuristiques-de-bastien-et-scapin/ [accessed 2024-03-17]
32. ISO 9241-11:1998(en) Ergonomic requirements for office work with visual display terminals (VDTs) — part 11: guidance on usability. International Organization for Standardization. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-1:v1:en [accessed 2024-03-17]
33. Ben Ramadan AA, Jackson-Thompson J, Schmaltz CL. Usability assessment of the Missouri cancer registry's published interactive mapping reports: round two. Online J Public Health Inform 2019;11(2):e3 [FREE Full text] [doi: 10.5210/ojphi.v11i2.9483] [Medline: 31632597]
34. Brooke J. SUS: a retrospective. J User Exp 2013;8(2):29-40 [FREE Full text]
35. Boyle L, Grainger R, Hall RM, Krebs JD. Use of and beliefs about mobile phone apps for diabetes self-management: surveys of people in a hospital diabetes clinic and diabetes health professionals in New Zealand. JMIR Mhealth Uhealth 2017 Jun 30;5(6):e85 [FREE Full text] [doi: 10.2196/mhealth.7263] [Medline: 28666975]
36. Kang H, Park HA. A mobile app for hypertension management based on clinical practice guidelines: development and deployment. JMIR Mhealth Uhealth 2016 Feb 02;4(1):e12 [FREE Full text] [doi: 10.2196/mhealth.4966] [Medline: 26839283]
37. Frederix I, Caiani EG, Dendale P, Anker S, Bax J, Böhm A, et al. ESC e-cardiology working group position paper: overcoming challenges in digital health implementation in cardiovascular medicine. Eur J Prev Cardiol 2019 Jul;26(11):1166-1177 [FREE Full text] [doi: 10.1177/2047487319832394] [Medline: 30917695]
38. Bernheim L, Balavoine JF. Pénurie de médecins en Suisse : mythe ou réalité ? Rev Med Suisse 2016;504:280-281 [FREE Full text] [doi: 10.53738/revmed.2016.12.504.0280]
39. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003 Aug;35(3):379-383 [FREE Full text] [doi: 10.3758/bf03195514] [Medline: 14587545]
40. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1 [FREE Full text] [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]

XSL•FO
RenderX

41. Alwashmi MF, Hawboldt J, Davis E, Fetters MD. The iterative convergent design for mobile health usability testing: mixed methods approach. JMIR Mhealth Uhealth 2019 Apr 26;7(4):e11656 [FREE Full text] [doi: 10.2196/11656] [Medline: 31025951]
42. Shah RU, Klein L, Lloyd-Jones DM. Heart failure in women: epidemiology, biology and treatment. Womens Health (Lond) 2009 Sep;5(5):517-527 [FREE Full text] [doi: 10.2217/whe.09.50] [Medline: 19702451]
43. What every client should know about SUS scores. Bentley University. URL: https://www.bentley.edu/centers/user-experience-center/what-every-client-should-know-about-sus-scores [accessed 2024-09-12]
44. Pros and cons of the system usability scale (SUS). Research Collective. URL: https://research-collective.com/pros-and-cons-of-the-system-usability-scale-sus/ [accessed 2024-09-13]
45. ChatGPT. URL: https://chatgpt.com [accessed 2024-10-14]

Abbreviations

HF: heart failure
HUG: University Hospital of Geneva
mHealth: mobile health
SUS: System Usability Scale
TCR: task completion rate

Edited by A Mavragani; submitted 10.07.24; peer-reviewed by R Rabiei, J Xia; comments to author 30.08.24; revised version received 18.09.24; accepted 15.10.24; published 15.01.25.

Please cite as:
Simioni L, Tessitore E, Hagberg H, Schneider-Paccot A, Blondon K, Gschwind L, Meyer P, Ehrler F
Cardiomeds, an mHealth App for Self-Management to Support Swiss Patients With Heart Failure: 2-Stage Mixed Methods Usability Study
JMIR Form Res 2025;9:e63941
URL: https://formative.jmir.org/2025/1/e63941
doi:10.2196/63941
PMID:39813081

©Lisa Simioni, Elena Tessitore, Hamdi Hagberg, Aurélie Schneider-Paccot, Katherine Blondon, Liliane Gschwind, Philippe Meyer, Frederic Ehrler. Originally published in JMIR Formative Research (https://formative.jmir.org), 15.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Teaching in the Digital Age—Developing a Support Program for Nursing Education Providers: Design-Based Research

Stefan Walzer[1*], MSc; Carolin Barthel[2*], MSc; Ronja Pazouki[2], MSc; Helga Marx[3], MSc; Sven Ziegler[3], MSc; Peter Koenig[1], Prof Dr; Christiane Kugler[2], Prof Dr; Stefan Jobst[2], MSc

[1]Care and Technology Lab, Furtwangen University, Furtwangen im Schwarzwald, Germany

[2]Institute of Nursing Science, Faculty of Medicine, University of Freiburg, Freiburg, Germany

[3]Nursing Direction, Medical Center, University of Freiburg, Freiburg, Germany

[*]these authors contributed equally

Corresponding Author:
Stefan Walzer, MSc
Care and Technology Lab
Furtwangen University
Robert-Gerwig-Platz 1
Furtwangen im Schwarzwald, 78120
Germany
Phone: 49 7723 920 2957
Email: stefan.walzer@hs-furtwangen.de

Abstract

Background: Health care systems and the nursing profession worldwide are being transformed by technology and digitalization. Nurses acquire digital competence through their own experience in daily practice, but also from education and training; nursing education providers thus play an important role. While nursing education providers have some level of digital competence, there is a need for ongoing training and support for them to develop more advanced skills and effectively integrate technology into their teaching.

Objective: This study aims to develop a needs-based support program for nursing education providers to foster digital competencies and to test this intervention.

Methods: We used a design-based research approach, incorporating iterative development with expert consultation to create and evaluate a support program for nursing education providers. Focus groups were conducted online to assess needs, and thematic content analysis was used to derive key insights. The support program was then refined through expert feedback and subjected to a feasibility and satisfaction test, with participant evaluations analyzed descriptively.

Results: Six main categories emerged from the focus groups, highlighting key areas, including the use of digital technology, ongoing support needs, and the current state of digitalization in nursing education. The support program was developed based on these findings, with expert validation leading to adjustments in timing, content prioritization, and platform integration. Preliminary testing showed good overall satisfaction with the support program, although participants suggested improvements in content relevance and digital platform usability.

Conclusions: Although the feasibility test showed high satisfaction with the support program, low participation rates and limited perceived knowledge gain were major concerns. The results suggest that while the program was well received, further refinements, including a focus on competency-based approaches and addressing workplace barriers, are needed to increase participation and effectiveness of such interventions. The findings of this research can be used as a basis for the development of similar programs in other educational and health care contexts.

(JMIR Form Res 2025;9:e66109) doi:10.2196/66109

KEYWORDS

digital competencies; nursing education; support program; needs assessment; design-based research; feasibility study; nursing education provider; qualitative research; nurse; health care; focus group; digital age; expert consultation; thematic content analysis; feasibility test; satisfaction; competency-based approach; workplace barrier; health care digitalization; digital technology

XSL•FO
RenderX

Introduction

Background

Technology and digitalization are transforming health care systems and the nursing profession globally [1]. The range of modern technologies used in health care practice is broad, with varying levels of complexity and includes, among others, electronic health records, sensors, tracking, artificial intelligence, and robotics [2,3]. Digital competence consists of knowledge, attitudes, and skills "required to use new technologies in a meaningful way and as a tool for learning, working and leisure time, understanding the essential phenomena of digital technologies in society as well as in one's own life, and the motivation to participate in the digital world as an active and responsible actor" (pp. 670-671) [4].

Digital competence for nurses is acquired through their own experience in daily practice, but also in continuing education and training, where the providers of this education play an important role [5,6]. In Germany, nursing is an apprenticeship-based profession. Educational activities during this apprenticeship are performed by nurse educators, who provide theoretical education to nursing students, and clinical mentors, who provide supervision and mentoring of trainees, students, or new professionals; these roles are referred to together as "nursing education providers" in this paper. The key aspects of the training and the characteristics of these job designations are summarized in Multimedia Appendix 1. Nursing education providers are responsible for designing and delivering educational programs that include the use of digital technologies [7]. These responsibilities involve developing digital teaching materials and resources for use in the classroom or online, in addition to providing hands-on training and support to help students or new professionals develop their digital competencies [8]. There are already several digital technologies being used in theoretical and practical nursing education, for example, laptops, headphones, and video cameras for online or hybrid teaching, digital blackboards, simulation dolls, and digital platforms [9]. The use of these technologies requires that nursing education providers are capable of imparting the necessary competencies and that they are also able to use digital technologies themselves through the demonstration of pedagogical digital competence in their work [10].

Internationally, several studies have investigated the evidence regarding the digital competence of nursing education providers. A study by Forman et al [11] showed barriers against the implementation of technology in teaching, as well as the need for better support for nursing education providers who want to use technology in educational activities. Furthermore, they pointed out a missing consensus about the definition of minimum levels of digital competence [11]. Similarly, Männistö et al [8] described competence as a multifaceted phenomenon, which depends on various factors. It is proven by multiple studies that a lack of knowledge regarding technology among nursing education providers leads to worse outcomes in the preparation of nursing students [12].

A survey of 169 nursing education providers in a German medical center assessed their self-perceived competence in using digital technologies for teaching [13]. The results showed that nurses generally feel competent and positive about using digital technology in educational activities. While a basic setup of digital teaching tools is available, respondents expressed a need for further training in digital competence. The study's quantitative approach provided insights into digital competence, but the authors recommend qualitative methods to deepen understanding and develop practical recommendations.

Overall, the evidence suggests that while nursing education providers have some level of digital competence, there is a need for ongoing training and support to help them develop more advanced skills and effectively incorporate technology into their teaching. Although the topic is attracting increasing international interest, there still seem to be no established training programs focused on digitalization in nursing education in several countries, including Germany [14].

Aim

The aims of this study were (1) to develop a needs-based support program for nursing education providers to foster digital competencies in a specific context at a university hospital in Germany and (2) to test this program regarding feasibility and acceptance.

Methods

Overview

This research report was written according to the GUIDED (Guidance for Reporting Intervention Development Studies in Health Research) [15] and the TIDieR (Template for Intervention Description and Replication) checklists [16]. Both checklists can be found in Multimedia Appendices 2 and 3.

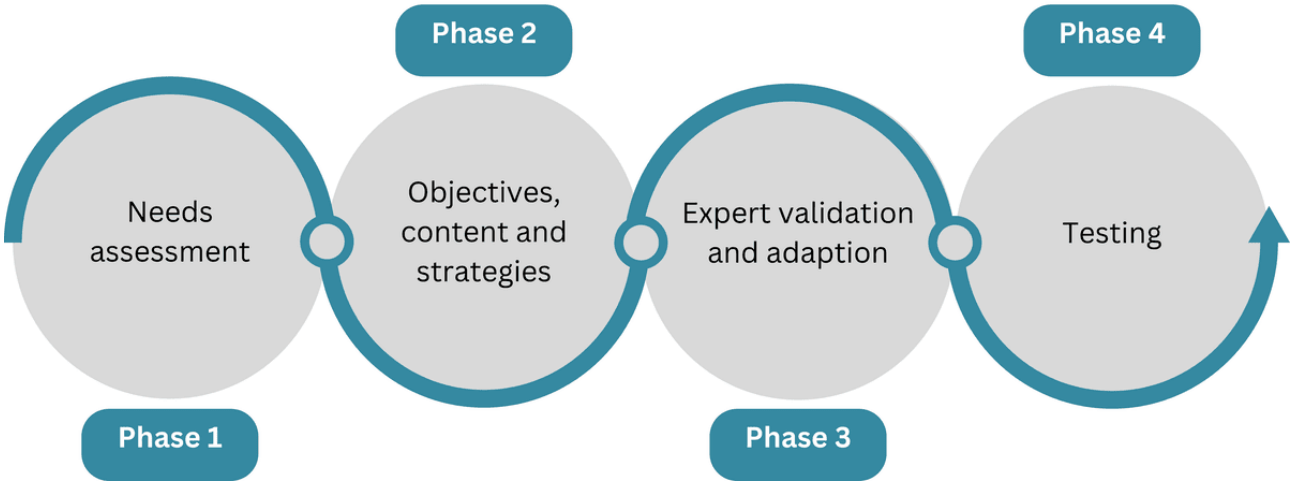
Design

The design of this study was based on the design-based research methodology as described by Anderson and Shattuck [17]. Design-based research is an iterative approach that aims to improve educational practices through the design, implementation, and refinement of interventions in real-world settings [17]. This methodology emphasizes stakeholder collaboration, iterative refinement, and empirical testing, making it well suited to the complexities of developing and testing a support program for nursing education providers [17] and has already proven effective for digital innovation in education [18]. Within design-based research methodology, the development of the support program followed a step-by-step, iterative, and participatory process with the consultative involvement of experts of the target group, based on the recommendations for the development of educational programs by Schneiderhahn et al [19], Schlutz [20], and the Medical Research Council framework for the development and evaluation of complex interventions in health care [21]. Schlutz [20] emphasizes the importance of developing educational services that respond to the specific needs of learners within their professional contexts. His focus on participatory approaches helps to ensure that the support program we develop is not only relevant but also tailored to the challenges faced by nursing education providers in the digital age. Similarly, Schneiderhan et al [19] present a systematic guide to curriculum development that emphasizes

XSL•FO
RenderX

the need to align educational objectives with learner needs and competencies. Their emphasis on stakeholder involvement and iterative revision is consistent with our design-based research methodology, enhancing the program's effectiveness and ensuring its relevance to nursing education. The Medical Research Council framework [20] further contributes to our study by providing a structured approach to the development and evaluation of complex interventions. It ensures that our support program is contextually appropriate, and incorporates iterative feedback and refinement throughout the development process.

Figure 1. Support program development steps.



Phase 1: Needs Assessment

Overview

Based on an explanatory sequential design [20], this phase adopted a qualitative research approach using focus groups to assess the needs of nursing education providers based on the results of the mentioned quantitative survey [13]. Focus groups are a special type of group interview where a moderator leads a discussion focused on a particular subject [22]. They are useful for obtaining information about beliefs, attitudes, thoughts, and perceptions on a specific topic from multiple individuals in a short amount of time. Participant interaction within the group plays a vital role in generating the data [22-24].

An interview guideline was developed by the research group following the recommendations of Kallio et al [25]. A semistructured interview guideline was chosen because it focused on the concepts of interest while still allowing flexibility. Sequential data integration [26] was performed by using the structure and content of the guideline as a base for the structure of the written survey questionnaire used in the study of Jobst et al [13] and the core findings to be explored in greater depth. During the development phase, preliminary versions of the questionnaire went through several internal reviews and adjustments by all members of the research team. The themes of the final guideline were as follows: (1) existence and use of digital technologies; (2) assessment of digital competencies; (3) support; and (4) age, gender, and specialty area. Each topic area was introduced in advance by a brief presentation of corresponding results from the written survey to provide context for participants and to encourage them to discuss their experiences related to the topic of interest. Due to hygiene regulations (caused by the COVID-19 pandemic) at the time the study was conducted, focus groups were scheduled as online sessions. For this purpose, the videoconferencing service WebEx (Cisco) was used which operated via the secure servers of the institution. The research team programmed technical support if necessary. Although the members of the research team and potential participants belong to the same professional group, there is no overlap between their respective fields of work. Thus, it was assumed that there is no dependency relationship or conflict of interest between the research team and the participants.

Participants

Nursing education providers were determined as the target group for participation in this study, as they had specific knowledge and experience related to the objectives under investigation. The teaching activity in nursing education can be considered as a common characteristic of the participants, which facilitates the creation of a pleasant atmosphere for discussion and exchange of opinions, as well as promoting interaction between them. A purposive sampling approach [27] was used to identify and recruit nursing education providers working at the above-mentioned medical center and the associated nursing school. Prerequisites for participation in the focus groups were membership in the professional group of nursing education providers employed at the University Medical Center, being

This study was conducted and coordinated by a group of nurse researchers with expertise in the fields of digitization (SW and SZ), adult education (CB and HM), and intervention development (SJ). One group member (SJ) was already involved in the previous quantitative survey mentioned above [13].

Procedures

In the context of this study, the phases for curriculum development according to Schneiderhan et al [19] and Schlutz [20] were integrated into the process of design-based research for the detailed design of the learning situation. The phases and the iterative approach are illustrated in Figure 1 and are outlined in the following sections.

XSL•FO
RenderX

aged at least 18 years, providing a signed consent form, and having to access to technology to use the videoconferencing service. Prior to starting the recruitment process, permission was obtained from the institution's staff council, which approved the involvement of nursing education providers. Potential participants were contacted via email using internal institutional mailing lists and asked to respond in writing if they were interested in participating in the study. Interested individuals received detailed written information about the study and participation, as well as a written informed consent form.

Data Collection

The two focus groups were led by two of the authors (SW and SJ), of whom one already had experience in conducting interviews. They were assisted by two other members of the research team who took field notes or asked supplementary questions. The focus group discussions were audio recorded and transcribed using a denaturalized approach in which idiosyncratic speech elements were omitted [28]. Field notes written during the discussions were merged and converted into a digital text format. Transcripts and field notes were pseudonymized prior to analysis. The participants received a demographic questionnaire prior to the focus group discussion.

Analysis

The qualitative data in textual form collected from the transcribed focus group discussions were analyzed using MAXQDA 2022 (VERBI Software). Content analysis using a deductive-inductive approach using the method of thematic qualitative text analysis was performed [29]. To analyze the data, a deductive categorization matrix was developed with four main categories based on the themes of the interview guideline. Content was assigned to these main categories and respective subcategories were inductively formed in the process. Codes that did not fit into the categorization matrix were combined into inductively formed new main categories and subcategories during the analysis process [30]. Prior to the coding process, the categorization matrix was pilot-tested by two members of the research team to check its fit and feasibility. Six coders divided into three groups sequentially coded the transcripts, each group building on the interpretive work of the previous ones. First, initial codes were developed in a process of condensation, followed by an abstraction process, where codes were then combined into categories that were allocated to the main categories of the predefined categorization matrix. After the end of this phase, coders met to discuss the resulting codes and categories, their classification in the categorization matrix, and any discrepancies, and to make modifications. Finally, summaries in narrative form were written for each category. Based on these summaries, thematic descriptions were created for the main categories [31]. For the development of the support program, content from categories that pointed out existing needs and requirements was used. The essential information contained in these descriptions is provided in the Results section of this report.

Phase 2: Objectives, Content, and Strategies

Overview

Phase 2 involved defining the content and objectives of the support service, as well as the format and methods with which the support service was to be implemented. The procedure as described in the model of Schneiderhan et al [19] was applied in an adapted form: steps 2 to 4—determining and prioritizing content, writing goals and objectives, and selecting teaching or educational strategies—were combined into one, following the example of Schlutz [20], as the content, objectives, and methods of a learning situation are highly interdependent.

The development of the support program has been guided by Bloom's [32] taxonomy, which categorizes learning into cognitive, affective, and social domains. To ensure support and acceptability of the support program by the targeted audience and key personnel (eg, nurse managers, leading clinical mentors, and school administration), current conditions, resources, and the results of both empirical studies (survey and focus groups) were also considered during the development of the support program.

Data Collection

The goals, objectives, and content were derived from the relevant results of the survey and the focus groups using sequential data integration [26]. The research group met regularly over a period of three months to plan the pilot implementation and testing of the support program. The following data sources were used: results of the web-based survey, results of the focus groups, experts within the institution, and knowledge of internal workflows at the hospital that would facilitate the testing of the support program.

Analysis

The research team produced a draft based on a matrix identifying the relevant components of a support program. This structure was linked to the steps of Schlutz [20]. It included information on the theoretical aspects of the support program [19,33], as well as information on practical implementation. Based on the results of the web-based survey and the focus groups, learning objectives were developed, which served as the basis for the content design of the support program. The analyses of the focus groups revealed the participants' learning needs for several topics. These were prioritized according to their importance and realizability. After that, topics were developed in a creative and iterative group process as frames for the organization of the content of the support program. Topics were operationalized using brief descriptions. At the same time, an educational strategy was defined in terms of format, teaching methods and media, quantity, and timing. Both, topic descriptions and educational strategies, were then entered into the matrix. Experts from the institution were identified and asked to act as potential lecturers at the planned times. Based on the subject areas and descriptions in the manual, they prepared short presentations according to the specified parameters (format, time). The content and methodology were agreed with the members of the research team.

XSL•FO
RenderX

Phase 3: Expert Validation and Adaptation

Overview

This phase included several feedback rounds in which the draft of the support program was presented to various experts from the target group. The feedback from the experts was intended to ensure the feasibility of the support service, as well as to increase the acceptance in the target group.

Participants

A total of five experts, all working at a University Medical Center, were consulted. They had the following expertise: adult education in health care, clinical mentoring, practical and theoretical training of nurses, or digitalization and e-learning.

Data Collection

The experts were interviewed in a total of four appointments, on three of the four appointments the experts were interviewed individually, on one appointment two clinical mentors were interviewed simultaneously. The interviews were conducted face to face and were structured as follows: first, the experts were presented with the draft of the support service, after which they were able to comment on it openly. Each meeting, and especially the feedback by the experts, was protocolled by a member of the research group.

Analysis

The protocols were anonymized; their contents were compiled and then analyzed using qualitative content analysis. Themes and categories resulting from this analysis were discussed within the research group and used for adjustments of the draft produced in phase 2 and finally expanded into a manual with detailed descriptions of all elements for implementing the support program.

Phase 4: Feasibility Testing (Feasibility and Acceptance)

Overview

In phase 4, the support program was implemented according to the manual to assess its feasibility and acceptance. Feasibility in this context refers to the extent to which the intervention can be practically implemented as intended. This includes factors such as the manageability of the program, the availability of necessary resources, and the ability of the target group to use the program as designed [33]. Acceptance was operationalized as participant satisfaction, meaning that the focus was on how well the program was received by users and whether it met their needs and expectations [33].

Prior to the implementation of the support program, experts from the institution were identified and asked to act as potential lecturers at the planned times. Based on the subject areas and descriptions in the manual, they prepared short presentations according to the specified parameters (format, time). The content and methodology were agreed with the members of the research team. A member of the research team was present at each session to provide technical support and moderation. Each short presentation was recorded with the consent of the lecturers so that it could be made available asynchronously at a later date. In order to comply with data protection regulations, the participants' cameras and microphones were deactivated during the short lecture. After the recording was finished, at the end of the lecture, the participants' cameras and microphones were activated so that they could take part in the discussion round.

Participants

All nursing education providers at the medical center had the opportunity to take part in the support program as part of the feasibility test. Gatekeepers forwarded an invitation to the target group via email with information on the support program. Participation in the support program and the survey at the end of phase 4 was voluntary.

Data Collection

A survey design was used to evaluate the support program afterward. A new questionnaire was developed by the research team especially for this study, as no suitable instrument was available in German. The development of the questionnaire was guided by the conceptualizations of Proctor et al [34], Sidani and Epstein [35], and integrated elements of the ABC-SAT questionnaire to assess affective, behavioral, and cognitive domains of satisfaction [36]. The questionnaire was designed as a web-based questionnaire comprising 18 questions on the domains of satisfaction, feasibility, and appropriateness, each of which could be rated using a 5-point Likert scale. Attendance at the relevant events was also requested and there was an option to enter a free text. The demographic information collected in the questionnaire was the participants' main area of work, age group, and gender. In addition, anonymized data on the number of participants during the sessions and the number of views of the video recordings of the sessions was collected by means of a protocol.

Analysis

Numerical data were analyzed descriptively using Microsoft Excel 2016. Textual data (comments of participants) were analyzed using qualitative content analysis.

Setting

The study was conducted at the University Medical Center Freiburg.

Ethical Considerations

The conceptualization and implementation of this study were based on the principles of the Declaration of Helsinki. By German law, survey studies with a focus on employees must be approved by the employee council at the respective institution. This body serves to fulfill ethical requirements and safeguard the rights of employees in this country. These responsibilities also refer to the protection of personal rights and data protection in the context of surveys. For this reason, a project description and information materials for potential participants were submitted to the relevant council for approval. The responsible committee of the employee council of the Freiburg University Medical Centre provided approval for this study in written format. Participation in the focus groups was voluntary, no personal data were collected, and anonymity was always maintained. All potential participants received written information on the study (reason for the study, objective,

XSL•FO
RenderX

processes, data protection), were informed about the decision of the employee council and its subcommittees, and had the opportunity to contact the investigators in case of questions at any time during the study. Informed consent to participate was assumed if individuals completed the questionnaire and were confirmed (by ticking a box) at the beginning of the questionnaire.

Results

The presentation of the results of the study follows the structure of the 4 steps of the development process of the support program.

Phase 1: Needs Assessment

Overview

Two focus group sessions with a total of 7 participants (clinical mentors: n=4; nurse educators: n=3) took place in April and May 2022. The first focus group discussion lasted 1 hour and 15 minutes, the second focus group discussion lasted 56 minutes. The characteristics of the participants are provided in Table 1.

Table 1. Sample characteristics.

Characteristics	Participants (N=7), n (%)
Job designation	
Clinical mentor	4 (57)
Nurse educator	3 (43)
Sex	
Female	4 (57)
Male	3 (43)
Age (in years)	
18-35	2 (29)
36-49	3 (43)
≥50	2 (29)

Six main categories emerged from the data, each with several subcategories: (1) existence and use of digital technologies; (2) assessment of (own) digital competencies; (3) support; (4) age, gender, and field of specialization; (5) attitude toward new technologies; and (6) current state of digitization. The last two main categories were derived through an inductive approach from the qualitative data, while the remaining main categories were derived from the interview guide. The development of the support program was based on the analysis of the relevant categories (1), (3), and (6). An overview of these categories and their subcategories is presented in Textbox 1.

Textbox 1. Categorization matrix of the content analysis of the focus groups with relevant content used for the development of the support program.

Main category 1: existence and use of digital technologies (deductive category)

- Existing and used technologies
- Existing and unused technologies
- Desired technologies
- Expectations of technologies

Main category 2: support (deductive category)

- Content-related aspects
- Organization and design of support
- Ongoing input and support
- Experiences with existing support

Main category 3: current state of digitalization (inductive category)

- Best practice
- Opportunities
- Poor practice
- Barriers

XSL•FO
RenderX

Existence and Use of Digital Technologies

This main category covers various aspects related to the use of technology in educational activities. Focus group discussions revealed a wide range of technologies that were already being used, such as digital blackboards, digital learning platforms, laptops, and software for online meetings. However, some existing technologies, like smart boards, were not used, or not used optimally due to a lack of knowledge or partially nonfunctional hardware:

> *Actually, the other parts [smart boards] are essentially used like whiteboards for writing on. You write on them and then wipe it away. And that approach is not really what one could do.* [Focus group 1, nurse educator]

Participants expressed a desire for better equipment in the rooms for hybrid teaching, including more laptops, tablets, simulation manikins, and robots. Specific expectations for technologies used in education were also mentioned. Participants want technology to be useful and have a reasonable cost-benefit ratio. Furthermore, it should increase students' interest and motivation and support the learning process. Additionally, the use of technology should be easy and accessible for collaborative work. The participants also mentioned the benefits of technologies that allow for the easy integration of media into teaching and the documentation of examination or instruction situations.

Support

Four subcategories emerged from the data as a result of the discussion about support. This category included participants' desire for ongoing input and support, their experience with existing support, as well as their ideas, wishes, and needs related to content, organizational, and design aspects of (new) support.

Several participants were already aware of support programs and training on new technologies provided by their institution and had taken advantage of it in individual cases. In this context, even small gains in knowledge were of great importance to the participants. Among other things, access to support programs through a complicated registration process or limited scheduling was viewed critically. Online formats and purely written information were perceived as challenging. The lack of practical exercises was also highlighted as a criticism. In addition, uneven levels of difficulty and limited applicability were named.

It was mentioned that support for nursing education providers during the implementation of new technologies in teaching must be oriented toward different elements, that is, inspiration and support for the correct use of technology in teaching, as well as training in the appropriate use of these technologies. In particular, opportunities for practice were desired. When organizing support, special attention should be paid to interprofessional group formation, different levels of competence, and the requirements of the particular professional groups in regard to new technologies:

> *I often had the feeling that meanwhile there was a huge gap. Either I get a very marginal introduction ... Or I have something high-end, which is not user-related.* [Focus group 1, nurse educator]

In addition, participants expressed a desire to contact persons who can provide support in the event of technical problems. Further training should also be programmed on an ongoing basis in addition to regular refresher training. Participants pointed out the challenge of reconciling a busy schedule with training on new technologies.

Current State of Digitalization

All participants gave examples and anecdotes of the current state of digitalization and related competencies in the institutions in which they work. In line with the two subcategories, the reports of participants can be divided into "best practice" examples and opportunities, as well as "poor practice" examples and barriers.

In regards to "best practice," the use of online platforms for communication and information exchange was reported. Other positive aspects mentioned were the training opportunities, internal staff with responsibility for training in certain technologies, and support and mutual collegial support in case of technical problems.

> *...for example, recently I did a video call into the dialysis unit. And I did the lesson with the colleague via webinar together. Something like that. It all works. That's the advantage. Thanks to Corona.* [Focus group 1, nurse educator]

Opportunities for improvement included the need for further development of existing training and high-quality hardware equipment, as well as expanding the monthly work-group meeting to include technical aspects. Participants noted that in order to increase motivation to use new technologies, a recognizable added value must be evident along with comprehensible introductions. In the case of the latter, one participant would like to see better training methods.

Clinical mentors reported examples of poor practice, such as the lack of simulation dolls due to high costs, as well as a lack of training opportunities and responsibilities, and underutilization of existing software despite having the knowledge for an appropriate use. Nursing education providers discussed inadequate use of digital blackboards, limited dates for training, and insufficient differentiation based on competence levels in the training programed. All participants identified common barriers to the use of new technologies, such as high costs, inadequate equipment, lack of time, and lack of practical training space.

> *Because when I do hybrid teaching with a laptop that is on the table in front, I then occasionally run over and ask if he [a student, note by the authors] can hear and see everything, which of course he cannot...* [Focus group 11, nurse educator]

Participants also noted individual barriers, such as negative attitudes and lack of experience among colleagues, as well as the challenge of implementing technical solutions into large companies or hospitals due to existing federalist structures. They highlighted the organizational effort of implementing technology, particularly in terms of adapting teaching methods. Additionally, the design and delivery of general IT training from instructors without a background in teaching or nursing

XSL•FO
RenderX

was identified as a barrier as the sessions failed to address the needs of the participants.

The study revealed that while a range of tools such as digital blackboards and online platforms are used, some technologies are underused due to limited knowledge or nonfunctional hardware. Participants expressed a need for better equipment, ongoing training, and support, particularly in practical applications of technology and in overcoming barriers such as high costs, inadequate time, and organizational challenges. They emphasized the importance of tailored, accessible support that considers different competence levels and professional needs. Additionally, participants noted both positive and negative experiences with digitalization in their institutions, highlighting the potential for improvement in training and infrastructure.

Phase 2: Objectives, Content, and Strategies

The following three learning objectives were formulated for the support program, which guided the development process:

- Raising awareness of the topic of digitalization in one's own field of work
- Promoting motivation to use technologies in nursing education
- Creating opportunities for networking between nursing education providers from different areas of work

In line with Bloom et al [32], participants are expected to develop competencies within the cognitive domain, by gaining essential knowledge about digitalization in nursing education and gaining an understanding of how to apply digital tools to enhance teaching practices. They are also anticipated to cultivate competence within the affective domain by fostering positive attitudes toward digital technologies and increasing their motivation to engage with these tools. Finally, social competence is expected to be strengthened through collaboration and critical engagement with digital solutions, enabling participants to contribute meaningfully to professional discourse on digitalization in nursing education. Based on the objectives, initial topic descriptions were developed for the support program. The draft of the newly developed support program consisted of a series of short online meetings (ie, sessions), which were planned to be programmed once a week over a period of five weeks via the videoconferencing program WebEx, each lasting 15-20 minutes. Each session would focus on one of the topics. The session would start with a topic-centered keynote speech (max 10 min) by a lecturer with expertise in the respective topic followed by a moderated discussion round with the opportunity to ask questions and exchange views on the topic. Every session would be recorded and would later be made available on an institutional digital learning platform. Participants would thus have the option of attending directly or asynchronously. In addition, a forum would be set up on the learning platform to enable participants to interact between sessions.

Phase 3: Expert Validation and Adaptation

The experts provided feedback on the following aspects of the support program: framework conditions, themes, format, timing, motivation, and potential obstacles. The draft of the support program was generally rated favorably by the experts surveyed. The online format, which does not require registration, the use of software already familiar to the institution, the interactivity of the program (forum and discussion group), and the possibility of addressing questions directly to the relevant experts were highlighted as particularly beneficial features. In particular, the provision of recorded sessions was emphasized, which allows great flexibility for the target group. According to the experts, fixed dates and the short duration of the sessions would enhance motivation and predictability of participation. Experts noted that scheduling the sessions between 1:30 PM and 2 PM would ensure the largest possible number of potential participants could be reached. In addition, experts recommended addressing opportunities for digital collaboration between nurse educators and clinical mentors, some of which are unknown. Based on this feedback, the timing of the program was adjusted accordingly and the topic "Collaboration on a digital learning platform" was added to the lecture series.

Finally, the following five topic descriptions were developed for the support program:

- New nursing technologies at the University Medical Center Freiburg
- Collaboration on a digital learning platform
- Further training programs or further training catalogs for teaching nurses
- Nuts and bolts of Cisco WebEx
- Overview of functions of a locally available online training platform

The manual for the support program was updated accordingly and the anchoring of the support program on the institutional digital teaching-learning platform was finalized.

Phase 4: Testing (Feasibility and Acceptance)

A total of 25 people enrolled in the course for the support program. Attendance at the four keynote presentations varied, with the second session (collaboration on a digital learning platform) and fourth session (nuts and bolts of Cisco WebEx) being the most attended (n=2), and the third session (educational programs or catalog for teaching nurses) receiving no attendance. The video of the first keynote (new nursing technologies) was viewed most often (n=12), with declining views for every following keynote. With one exception, the testing was conducted as initially planned. The last keynote speech could not be held due to the illness of the lecturer. Overall, the evaluation of the support program involved seven participants, consisting of four male participants, two female participants, and one nonbinary person. Participants included 4 clinical mentors and 2 nurse educators, with one abstention. The median age range was 40-49 years.

Satisfaction with the support program was rated overall as good. Positive responses noted a supportive atmosphere during presentations, high instructor motivation, and well-prepared instructors. However, cognitive feedback indicated less perceived knowledge gain and mixed opinions on content relevance and session length adequacy. Regarding feasibility, participants found the format suitable but noted challenges in integrating it into their daily routines. The digital conference format was generally viewed as appropriate, but the forum and

XSL•FO
RenderX

platform usability were deemed less satisfactory. Open-text responses highlighted strengths such as competent speakers and efficient time management, while suggestions for improvement included enhancing educational content relevance and platform usability.

Discussion

Principal Findings

In this study, a design-based research methodology was used to develop and test a needs-based support program for nursing education providers to foster digital competencies. The study was conducted in four phases: phase 1 used focus groups to assess the current use of digital technologies and the state of digitalization in nursing education. In phase 2, the objectives, content, and strategies for a support program were developed based on previous survey results and the focus group insights, leading to three main learning objectives: raising awareness of digitalization, motivating the use of technology, and fostering networking among nursing educators. This phase was framed by Bloom et al's [32] taxonomy, which emphasizes the development of learning in the cognitive, affective, and social domains—each of which is crucial to the development of digital competencies. Phase 3 involved expert validation through multiple interviews, resulting in the refinement of 5 key topics and the finalization of the support program. Finally, phase 4 tested the feasibility and acceptance of the support program using a web-based questionnaire based on the ABC-SAT [36], which showed high satisfaction but low participation and insufficient perceived learning outcomes.

In several countries, including Germany, there are no formal training programs regarding digitalization for nursing education providers [14]. The feasibility test of this support program represents the first step toward the implementation of such a formal training program. This program can be classified as level 1—"reaction" in Kirkpatrick's model of the "four levels of training" [37]. This means that this program remains rather rudimentary, but has the potential to be expanded and developed further using additional resources. In the absence of a specific guideline for the development of such a support program, the design-based research methodology was used to structure and guide the process of this study. This proved to be extremely helpful, as the steps were comprehensible and the quality of the support program could be improved through iterative and participatory processes. Involving representatives of the target group in the development process brought several advantages: not only did it contribute to increasing the quality of the support program, it also ensured the feasibility of the support program and provided the possibility of easy and low-threshold contact and information for the target group about the planned support program.

The support program aimed to address the entire target group of nursing education providers, which was legitimized with the aim of promoting cooperation between the subgroups of nurse educators and clinical mentors. When designing the content of the support program, a topic-centered approach was taken and an attempt was made to find topics that would appeal to both subgroups of the target group. We chose this approach for several reasons. First, it allowed us to address the immediate interests and needs of both subgroups within our target audience. By focusing on engaging topics, we aimed to maximize initial interest and participation, which is crucial for the success of any educational program. Our goal was to create a stimulating environment where participants would be motivated to actively engage with the content. As a consequence of the topic-oriented approach, it was not possible to design a competence-oriented approach. Although this is not evident from the qualitative and quantitative survey data, there are indications that a competence-oriented design would have been more suitable for the support program. To address this, potential improvements could include introducing competence assessments and providing additional resources while applying a broader variety of learning methods and concepts. This would lead to more needs-oriented programs, especially for the respective target group, and among other things, result in a differentiation of the presented topics into different competence levels (eg, beginner, advanced, and expert) or specific addressing of one of the professional groups. As noted by Sillat et al [38], the increasing number of technologies, frameworks, and strategies underscores the need for assessing and evaluating digital competence in various settings. Given the challenges in self-assessment, which are not limited to nursing settings, the provision of a screening or assessment of digital competencies before choosing a training level should be considered.

The extent to which differentiation into different competence levels would have led to higher participation figures and greater satisfaction with the support program remains to be clarified. With regard to the number of participants, it is noticeable that the recordings of the keynote speeches, which were made available online asynchronously, had higher participation rates than the synchronous online presentations. One possible reason for this could be the stressful and sometimes unpredictable working day of nursing education providers, especially if they work in clinical practice. In addition, the respondents previously reported that they had hardly any time resources, which may be another reason for the low level of participation. This is also confirmed by Dornan [39] stating that employees with a care mandate generally prioritize these tasks higher than participation in continuing education courses. After or during a stressful working day, some individuals may lack the mental capacity to absorb further input in the form of training, as they may already be in cognitive overload [40]. Another reason could be a lack of motivation to participate. In this respect, it is known from adult education that the willingness and motivation to participate in further training depends very much on the general environment of the workplace [41].

These are, for example, the values and mission statement of the organization, as well as the demands and requirements placed on employees in the context of lifelong learning. Learning behavior and thus participation in further training is also strongly dependent on individual learning behavior, learning habits, and personal attitudes [41]. Noteworthy in this context are the different levels of involvement, ranging from enthusiastic to completely absent. While the stakeholders, who also belong to the target group of nursing education providers but tend to work less in direct training and increasingly take on management

XSL•FO
RenderX

tasks, were very interested and committed to the development of the support program, there was almost no interest at all on the part of the nurses being involved in direct patient care. We can only speculate about the reason for this great difference in motivation and participation. Since this research project was part of a large, 5-year project with various other research projects, one can also speculate about a kind of fatigue or desensitization to other research projects. Such a phenomenon has already been described by health care professionals [42].

The evaluation recorded only a few participants. However, this number is certainly justifiable and positive in the context of a feasibility test. The evaluation results show an ambivalent picture. In addition to the low number of participants, the results showed little to no increase in knowledge among the participants, yet they stated that they were generally satisfied with the program. A further more in-depth survey of the target group could provide clarity here and answer questions such as which people did or did not take part in the sessions and for what reasons, and whether a different format or a different time would have been more suitable. To address the low participation rate, measures such as enhanced recruitment strategies (eg, engagement of leadership) and flexible participation options (eg, live, asynchronous, and on-demand) should be considered to encourage higher involvement. Additionally, a focus on both intrinsic motivation (eg, emphasizing the value of lifelong learning, team culture, professional identity, and commitment to quality care) and extrinsic motivation (eg, incentives) should be incorporated [43].

Implications for Practice and Research

Based on the results of this study, several implications for practice can be derived. The next step should be to determine the reasons for participation or nonparticipation in the support program, for example, through a follow-up survey of the target group. This should also address obstacles and facilitating factors.

The support program developed here should also be further developed on the basis of design-based research. When further developing existing support programs or designing new ones, workplace-related factors such as workload, time resources, and working hours should be taken into account. It is also recommended—especially for complex and wide-ranging topics—to pursue a competence-oriented approach in order to best meet the needs of the target group.

Limitations

This study has several strengths and limitations that should be considered when interpreting the results. First, the sample of this study was relatively small and consisted of individuals from only one medical center. This sample may restrict the generalizability of the findings. However, the evaluation showed an acceptable number of participants for a feasibility study and the issues and findings raised in this study may be still relevant on a national and international level, especially for nursing education and training programs that prioritize the digital competencies of nursing education providers. In this study, the sample was highly specific which improves the credibility [44].

Second, despite a review of existing instruments, no suitable questionnaire could be found, as existing instruments were mostly too generic. Therefore, a self-created questionnaire had to be used for the evaluation, which was based on the ABC-SAT questionnaire [36] and conceptualizations of feasibility and acceptance, but was not validated in its final form.

Third, it is important to be aware of the potential bias of the researchers in this study, as data collection and analysis were conducted by researchers with a background in nursing education [45]. This situation may lead to a better understanding of the context but also affect the interpretation of the data. To compensate for these biases, triangulation was performed during the analysis.

Conclusions

The research design, consisting of a quantitative survey, a qualitative exploration, and subsequent development with participatory elements of a support program, facilitated both the structured collection of nursing education providers' views and experiences regarding digital competence and the targeted development of a needs-based support program. However, the likely willingness of the target audience to use the programming in the future appears to be severely limited, as reflected in the low attendance during the synchronous keynote presentations. The reasons for this poor response are varied and may be due to factors such as lack of information or insufficient motivation on the part of the target group. Overall satisfaction was high, but concerns were expressed about the lack of opportunities for professional discourse and individual knowledge gain. Despite these challenges, the overall satisfaction with the pilot implementation suggests that digital and innovative programs have potential and can serve as a basis for further development. Moreover, the insights gained from this study, despite being derived from a localized setting, could be applied to guide the development and implementation of similar support programs in diverse educational and health care contexts.

Acknowledgments

We thank Sadhbh Byrne for data processing, statistical consultation, and proofreading the manuscript; as well as the stakeholders for their support in designing the support program. This study was conducted at the Center of Implementing Nursing Care Innovations, Freiburg, Germany. This joint project on the development and research of new care technologies in Germany is funded by the Federal Ministry of Education and Research. At the Freiburg Center, the focus is on integrating and evaluating innovative nursing technologies in the acute care setting. This entails evaluating the competencies of the users, and if necessary, supporting their development (grants 16SV7887, 16SV7888, and 16SV7886K).

XSL•FO
RenderX

Data Availability

The data sets generated during and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Presentation of the authors' contributions followed the recommendations of the Contributor Roles Taxonomy (CRediT) [46]. SW was responsible for the conceptualization of the study, developed the methodology, conducted the investigation, performed formal analysis, curated the data, and wrote the original draft of the manuscript. CB contributed to the study's conceptualization and methodology, participated in the investigation, performed formal analysis, curated the data, and wrote the original draft. RP curated the data and contributed to the writing of the manuscript by reviewing and editing it. HM worked on the conceptualization and methodology of the study, conducted the investigation, and contributed to the writing and editing of the manuscript. SZ played a role in the conceptualization and methodology, conducted the investigation, and contributed to the writing and editing of the manuscript. PK contributed to the conceptualization of the study, provided resources, contributed to the writing and editing of the manuscript, and supervised the project. CK was involved in the conceptualization, provided resources, contributed to writing and editing, and supervised the project. SJ contributed to the conceptualization, provided resources, contributed to the writing and editing of the manuscript, and supervised the study. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Key aspects of the training and the characteristics of educating nurses in Germany.
[DOCX File , 16 KB - formative_v9i1e66109_app1.docx]

Multimedia Appendix 2
Guidance for reporting intervention development studies in health research (GUIDED) checklist.
[PDF File (Adobe PDF File), 780 KB - formative_v9i1e66109_app2.pdf]

Multimedia Appendix 3
Template for Intervention Description and Replication (TIDieR) checklist.
[PDF File (Adobe PDF File), 502 KB - formative_v9i1e66109_app3.pdf]

References

1. Rouleau G, Gagnon MP, Côté J, Payne-Gagnon J, Hudson E, Dubois CA. Impact of information and communication technologies on nursing care: results of an overview of systematic reviews. J Med Internet Res 2017;19(4):e122 [FREE Full text] [doi: 10.2196/jmir.6686] [Medline: 28442454]
2. Krick T, Huter K, Domhoff D, Schmidt A, Rothgang H, Wolf-Ostermann K. Digital technology and nursing care: a scoping review on acceptance, effectiveness and efficiency studies of informal and formal care technologies. BMC Health Serv Res 2019;19(1):400 [FREE Full text] [doi: 10.1186/s12913-019-4238-3] [Medline: 31221133]
3. Huter K, Krick T, Domhoff D, Seibert K, Wolf-Ostermann K, Rothgang H. Effectiveness of digital technologies to support nursing care: results of a scoping review. J Multidiscip Healthc 2020;13:1905-1926 [FREE Full text] [doi: 10.2147/JMDH.S286193] [Medline: 33328736]
4. Ilomäki L, Paavola S, Lakkala M, Kantosalo A. Digital competence—an emergent boundary concept for policy and educational research. Educ Inf Technol 2014;21(3):655-679. [doi: 10.1007/s10639-014-9346-4]
5. Kaufman DM. Applying educational theory in practice. ABC BMJ 2003;326(7382):213-216 [FREE Full text] [doi: 10.1136/bmj.326.7382.213] [Medline: 12543841]
6. Bierema LL. Adult learning in health professions education. New Dir Adult Contin Educ 2018;2018(157):27-40. [doi: 10.1002/ace.20266]
7. Nes AAG, Steindal SA, Larsen MH, Heer HC, Lærum-Onsager E, Gjevjon ER. Technological literacy in nursing education: a scoping review. J Prof Nurs 2021;37(2):320-334 [FREE Full text] [doi: 10.1016/j.profnurs.2021.01.008] [Medline: 33867086]
8. Männistö M, Mikkonen K, Kuivila HM, Koskinen C, Koivula M, Sjögren T, et al. Health and social care educators' competence in digital collaborative learning: a cross-sectional survey. Sage Open 2020;10(4):215824402096278. [doi: 10.1177/2158244020962780]
9. Grimwood T, Snell L. The use of technology in healthcare education: a literature review. MedEdPublish 2020;9(1):137. [doi: 10.15694/mep.2020.000137.1]

XSL•FO
RenderX

10. From J. Pedagogical digital competence—between values, knowledge and skills. High Educ Stud 2017;7(2):43 [FREE Full text]
11. Forman TM, Armor DA, Miller AS. A review of clinical informatics competencies in nursing to inform best practices in education and nurse faculty development. Nurs Educ Perspect 2020;41(1):E3-E7. [doi: 10.1097/01.NEP.0000000000000588] [Medline: 31860501]
12. Webb L, Clough J, O'Reilly D, Wilmott D, Witham G. The utility and impact of information communication technology (ICT) for pre-registration nurse education: a narrative synthesis systematic review. Nurse Educ Today 2017;48:160-171. [doi: 10.1016/j.nedt.2016.10.007] [Medline: 27816862]
13. Jobst S, Lindwedel U, Marx H, Pazouki R, Ziegler S, König P, et al. Competencies and needs of nurse educators and clinical mentors for teaching in the digital age—a multi-institutional, cross-sectional study. BMC Nurs 2022;21(1):240. [doi: 10.1186/s12912-022-01018-6] [Medline: 36031618]
14. Tischendorf T, Hasseler M, Schaal T, Ruppert SN, Marchwacka M, Heitmann-Möller A, et al. Developing digital competencies of nursing professionals in continuing education and training—a scoping review. Front Med (Lausanne) 2024;11:1358398 [FREE Full text] [doi: 10.3389/fmed.2024.1358398] [Medline: 38947234]
15. Duncan E, O'Cathain A, Rousseau N, Croot L, Sworn K, Turner KM, et al. Guidance for reporting intervention development studies in health research (GUIDED): an evidence-based consensus study. BMJ Open 2020;10(4):e033516 [FREE Full text] [doi: 10.1136/bmjopen-2019-033516] [Medline: 32273313]
16. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014;348:g1687 [FREE Full text] [doi: 10.1136/bmj.g1687] [Medline: 24609605]
17. Anderson T, Shattuck J. Design-based research: a decade of progress in education research? Educ Res 2012;41(1):16-25. [doi: 10.3102/0013189x11428813]
18. Hall T. Bridging practice and theory: the emerging potential of design-based research (DBR) for digital innovation in education. Educ Res Perspect 2020;47:157-173 [FREE Full text]
19. Schneiderhan J, Guetterman TC, Dobson ML. Curriculum development: a how to primer. Fam Med Community Health 2019;7(2):e000046 [FREE Full text] [doi: 10.1136/fmch-2018-000046] [Medline: 32148703]
20. Schlutz E. Bildungsdienstleistungen und Angebotsentwicklung. Germany: Waxmann Verlag; 2006.
21. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of medical research council guidance. BMJ 2021;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
22. Plummer P. Focus group methodology. Part 1: design considerations. Int J Ther Rehabil 2017;24(7):297-301. [doi: 10.12968/ijtr.2017.24.7.297]
23. Doody O, Slevin E, Taggart L. Focus group interviews in nursing research: part 1. Br J Nurs 2013;22(1):16-19. [doi: 10.12968/bjon.2013.22.1.16] [Medline: 23299206]
24. Kuckartz U, Rädiker S. Analyzing Qualitative Data With MAXQDA: Text, Audio, and Video. New York, NY: Springer International Publishing; 2019.
25. Kallio H, Pietilä AM, Johnson M, Kangasniemi M. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. J Adv Nurs 2016;72(12):2954-2965. [doi: 10.1111/jan.13031] [Medline: 27221824]
26. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013;48(6 Pt 2):2134-2156 [FREE Full text] [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
27. Campbell S, Greenwood M, Prior S, Shearer T, Walkem K, Young S, et al. Purposive sampling: complex or simple? Research case examples. J Res Nurs 2020;25(8):652-661 [FREE Full text] [doi: 10.1177/1744987120927206] [Medline: 34394687]
28. Azevedo V, Carvalho M, Fernandes-Costa F, Mesquita S, Soares J, Teixeira F, et al. Interview transcription: conceptual issues, practical guidelines, and challenges. Rev Enf Ref 2017;IV Série(Nº14):159-168 [FREE Full text]
29. Graneheim UH, Lindgren B, Lundman B. Methodological challenges in qualitative content analysis: a discussion paper. Nurse Educ Today 2017;56:29-34. [doi: 10.1016/j.nedt.2017.06.002] [Medline: 28651100]
30. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
31. Lindgren B, Lundman B, Graneheim UH. Abstraction and interpretation during the qualitative content analysis process. Int J Nurs Stud 2020;108:103632. [doi: 10.1016/j.ijnurstu.2020.103632] [Medline: 32505813]
32. Bloom BS, Engelhart MD, Furst EJ, Hill WH, Krathwohl DR. Taxonomy of Educational Objectives: The Classification of Educational Goals. Green, NY: Longmans; 1964.
33. Sidani S, Braden C. Nursing and Health Interventions: Design, Evaluation and Implementation. 2nd Edition. Hoboken, NJ: Wiley-Blackwell; 2021.
34. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]

XSL•FO
RenderX

35. Sidani S, Epstein DR. Toward a conceptualization and operationalization of satisfaction with nonpharmacological interventions. Res Theory Nurs Pract 2016;30(3):242-257. [doi: 10.1891/1541-6577.30.3.242] [Medline: 28304269]
36. Meng M, Peter D, Mattner F, Igel C, Kugler C. Development and psychometric pilot-testing of a questionnaire for the evaluation of satisfaction with continuing education in infection control nurses. Nurse Educ Pract 2018;31:77-82. [doi: 10.1016/j.nepr.2018.05.003] [Medline: 29800763]
37. Kirkpatrick JD, Kirkpatrick WK. Kirkpatrick's Four Levels of Training Evaluation: Association for Talent Development. Alexandria, VA: ATD Press; 2016.
38. Sillat LH, Tammets K, Laanpere M. Digital competence assessment methods in higher education: a systematic literature review. Educ Sci 2021;11(8):402. [doi: 10.3390/educsci11080402]
39. Dornan T. Workplace learning. Perspect Med Educ 2012;1(1):15-23 [FREE Full text] [doi: 10.1007/s40037-012-0005-4] [Medline: 23316455]
40. McElhatton J, Drew C. Hurry-up syndrome. ASRS directline. 1993. URL: https://asrs.arc.nasa.gov/publications/directline/dl5_hurry.htm [accessed 2024-08-08]
41. West CP, Shanafelt TD. The influence of personal and environmental factors on professionalism in medical education. BMC Med Educ 2007;7:29 [FREE Full text] [doi: 10.1186/1472-6920-7-29] [Medline: 17760986]
42. Roxburgh M. An exploration of factors which constrain nurses from research participation. J Clin Nurs 2006;15(5):535-545. [doi: 10.1111/j.1365-2702.2006.01374.x] [Medline: 16629962]
43. Kalenda J, Kočvarová I. “Why don’t they participate?” Reasons for nonparticipation in adult learning and education from the viewpoint of self-determination theory. Eur J Res Educ Learn Adults 2022;13(2):193-208 [FREE Full text]
44. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
45. Kyngäs H, Mikkonen K, Kääriäinen M, editors. The Application of Content Analysis in Nursing Science Research. New York, NY: Springer International Publishing; 2020.
46. Brand AE, Allen L, Altman M, Hlava MMK, Scott J. Beyond authorship: attribution, contribution, collaboration, and credit. Learn Pub 2015;28(2):151-155 [FREE Full text]

Abbreviations

GUIDED: Guidance for reporting intervention development studies in health research
TIDieR: Template for Intervention Description and Replication

Edited by A Mavragani; submitted 04.09.24; peer-reviewed by S Kim, J Beaudin; comments to author 01.10.24; revised version received 18.10.24; accepted 31.10.24; published 15.01.25.

Please cite as:
Walzer S, Barthel C, Pazouki R, Marx H, Ziegler S, Koenig P, Kugler C, Jobst S
Teaching in the Digital Age—Developing a Support Program for Nursing Education Providers: Design-Based Research
JMIR Form Res 2025;9:e66109
URL: https://formative.jmir.org/2025/1/e66109
doi:10.2196/66109
PMID:

©Stefan Walzer, Carolin Barthel, Ronja Pazouki, Helga Marx, Sven Ziegler, Peter Koenig, Christiane Kugler, Stefan Jobst. Originally published in JMIR Formative Research (https://formative.jmir.org), 15.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Sexual and Gender Minority Adolescents' Preferences for HIV Pre-Exposure Prophylaxis Social Marketing Campaigns: Qualitative Preimplementation Study

Kathryn Macapagal[1,2,3], PhD; Juan Pablo Zapata[1,2], PhD; Junye Ma[4], MA, MS; Jacob D Gordon[1,5], MSW, PhD; Christopher Owens[6], MPH, PhD; Silvia Valadez-Tapia[7], MA; Peter Cummings[2,3], MA, MPH; Nathan Walter[8], PhD; Jim Pickett[9]

[1]Institute for Sexual and Gender Minority Health and Wellbeing, Northwestern University, Chicago, IL, United States
[2]Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
[3]Department of Psychiatry & Behavioral Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States
[4]Joint Doctoral Program in Clinical Psychology, San Diego State University/University of California San Diego, San Diego, CA, United States
[5]College of Nursing, University of Cincinnati, Cincinnati, United States
[6]Department of Health Behavior, School of Public Health, Texas A&M University, College Station, TX, United States
[7]AIDS Foundation Chicago, Chicago, IL, United States
[8]Department of Communication Studies, School of Communication, Northwestern University, Evanston, IL, United States
[9]Jim Pickett Consulting, Chicago, IL, United States

Corresponding Author:
Kathryn Macapagal, PhD
Institute for Sexual and Gender Minority Health and Wellbeing
Northwestern University
625 N Michigan Ave, Suite 1400
Chicago, IL, 60611
United States
Phone: 1 312 503 3605
Fax: 1 312 503 6700
Email: kathryn.macapagal@northwestern.edu

Abstract

Background: Sexual and gender minority (SGM) adolescents in the United States are disproportionately affected by HIV. Pre-exposure prophylaxis (PrEP) is a highly effective biomedical HIV prevention method, but its awareness and uptake among SGM adolescents are low. There are no adolescent-centered PrEP social marketing campaigns in the United States that have the potential to increase awareness and interest in PrEP.

Objective: To address this gap, this qualitative study aims to examine SGM adolescents' needs and preferences regarding adolescent-centered PrEP social marketing campaigns.

Methods: SGM adolescents from Chicago and its surrounding areas participated in web-based asynchronous focus groups from February to May 2021. Questions elicited their preferences for content, design, and delivery of SGM adolescent–centered PrEP campaigns. We used rapid qualitative data analysis and organized the findings around key components of social marketing, known as the 4 Ps: *product*, *price*, *place*, and *promotion*.

Results: Participants (N=56) were aged 14 to 19 years (mean 18.16, SD 1.22 y), and 64% (36/56) of them identified as a racial or ethnic minority. Among the 56 participants, 70% (n=39) were aware of PrEP; however, 95% (n=53) did not know that PrEP could be prescribed to those aged under 18 years. Adolescents expressed a need for PrEP campaign messaging that provides simple, accurate, and easily accessible information (eg, what is PrEP, for whom PrEP is indicated, and where and how to access PrEP). For product and price, SGM adolescents wanted a campaign to address barriers to, costs of, and how to access PrEP and desired to know about other adolescents' PrEP experiences to improve campaign relatability. For place and promotion, participants preferred digital campaigns on social media to reduce the possibility of embarrassment and stigma and increase the accessibility of health content.

XSL•FO
RenderX

Conclusions: These findings lay the groundwork for designing adolescent-centered educational PrEP campaigns that prioritize both user preferences in PrEP marketing design and strategies to overcome common barriers to PrEP awareness.

(JMIR Form Res 2025;9:e60398) doi:10.2196/60398

KEYWORDS

social marketing campaigns; sexual and gender minority; adolescent; HIV; pre-exposure prophylaxis; PrEP; human-centered design; implementation science; dissemination

Introduction

Background

Adolescents and young adults aged 13 to 24 years in the United States are disproportionately affected by HIV, accounting for 19.3% of new diagnoses in 2022 [1]. In 2019, of the 25,552 new HIV infections among sexual minority men in the United States, 6386 (25%) were aged 13 to 24 years [2]. Surveillance data in Cook County, Illinois, an Ending the HIV Epidemic priority area that includes Chicago and its surrounding areas, show that 82.3% of HIV cases among youths aged 13 to 29 years were among men who have sex with men, 59.7% of the cases were among Black sexual and gender minority (SGM) youths, and 20.3% of the cases were among Latino SGM youths [3].

Pre-exposure prophylaxis (PrEP) refers to HIV prevention medication regimens that currently include oral emtricitabine-tenofovir disoproxil fumarate (brand name Truvada) or emtricitabine-tenofovir alafenamide (brand name Descovy) and long-acting injectable cabotegravir (brand name Apretude). PrEP is a highly effective biomedical prevention strategy, preventing up to 99% of HIV infections when taken as prescribed [4,5,6]. Despite the US Food and Drug Administration's approval of oral PrEP for individuals aged under 18 years in 2018 and long-acting injectable PrEP in 2021 [7], PrEP uptake remains subpar in this age group [8,9]. One study that used US pharmacy claims data to estimate adolescent PrEP use [10] found that of the 6444 youths aged 13 to 19 years who were prescribed PrEP in 2021, just 12% were aged under 18 years, and most were adolescent girls. Across several studies using convenience samples, self-reported PrEP uptake among SGM youth aged 14 to 17 years has been consistently low, ranging from 2% to 5.9% [8,11-14].

Awareness and knowledge of PrEP are 2 significant barriers to its use among adolescents. In studies based in the United States, PrEP awareness among SGM adolescent samples has varied from 15% to 55% [12-15], with higher rates observed in studies consisting of participants who had previously taken part in HIV prevention research where they learned about PrEP. Several demographic, social, cultural, and structural factors relate to SGM youth PrEP awareness, including older age, greater educational attainment, and better access to sexual health care [16-20]. Research also suggests that adolescents have limited knowledge about PrEP beyond the fact that it prevents HIV infection and assume that PrEP is only available for cisgender men or those aged ≥18 years [11,12]. When adolescents are aware of PrEP, they report moderate to high levels of willingness to take it; for example, 2 survey studies showed that 64.9% [9] to 78.5% [8] of adolescent sexual minority men aged 13 to 18 years were willing to use PrEP. Furthermore, even when SGM adolescents are aware of PrEP, their willingness to use it may be moderated by a variety of developmental, socioeconomic, and sociocultural barriers to obtaining and adhering to PrEP. These include concerns about parental involvement, affordability, the need for regular clinic visits but the difficulty in accessing them independently, side effect concerns, and a lack of health care provider training in adolescent PrEP provision [8,10,12,17,21-25].

Social marketing campaigns, which integrate marketing, psychology, and public health principles, have immense potential to promote PrEP awareness and uptake among SGM adolescents, creating positive changes in attitudes, behaviors, and awareness for the greater good of society [26,27]. The key components of social marketing are often referred to as the 4 Ps [28,29]. *Product* includes messages about the desired behavior and its attributes, its health benefits (eg, PrEP awareness, uptake, and HIV prevention), and perceived nonhealth benefits of the behavior (eg, fulfilling social needs). *Price* refers to real and perceived financial or practical costs or other barriers to engaging in the desired behavior from the consumer's perspective and strategies to lower these costs. *Place* involves determining where the focal audience engages in the behavior (in this case, where one obtains PrEP) and implementing ways to make the desired behavior, service, or product more accessible. *Promotion* entails the ways persuasive messages are designed and delivered to the desired audience [28]. Other models also include a fifth P, *policy*, which includes policy-level barriers and facilitators to engaging in the desired behavior that should be communicated to the audience [29]. For instance, in some US states, PrEP is explicitly included in laws permitting adolescents to access certain types of health care without parental consent (eg, mental health and substance use treatment; prevention, testing, and treatment of sexually transmitted infections [STI] and HIV; and other sexual and reproductive health care) [30].

Social marketing campaigns take effect on both population and individual levels [31]. In the case of adolescents, a PrEP social marketing campaign could reach them quickly and directly through multiple forms of media (eg, social media, websites, and posters) that circumvent the need to rely on parents, health care providers, and educators who may or may not deliver or even know about this information. One example of such a campaign is PrEP4Love [32,33], a Chicago-based PrEP awareness initiative designed for Black adult gay and bisexual men and transgender and cisgender women. PrEP4Love reached 41 million unique viewers and was associated with a 3-fold increase in calls to a PrEP hotline over the 2 months following the campaign launch, most of whom were interested in learning

XSL•FO
RenderX

more about or being linked to PrEP [32,33]. The success of the campaign in Chicago led to its adoption by a French AIDS service organization to raise PrEP awareness across France [34].

Although adolescent-inclusive sexual health campaigns exist, such as Get Yourself Tested [35] and National Youth HIV/AIDS Awareness Day [36], few specifically include information about or an emphasis on PrEP. Social marketing draws from exchange theory, which suggests that individuals consider perceived costs and benefits when making decisions and focuses on segmenting the population into distinct audiences (ie, market segmentation) that are more likely to respond to a particular campaign or campaign component [37,38]. As such, social marketing campaigns can be designed to be developmentally appropriate for adolescents to highlight their specific preferences and needs. For example, the content and imagery of PrEP4Love [32,33] were clearly sensual and sexual in nature, which may make many adolescents uncomfortable; approaches that elicit different emotions, such as empowerment or affirmation, could be more resonant with younger audiences [39,40]. Moreover, where campaigns are placed is critical in ensuring that the campaign reaches its intended audience. For instance, PrEP4Love was often promoted in person (eg, in health clinics, public transit, and nightlife establishments) with a website, hotline, and social media presence. However, SGM adolescents may not be able to access the same physical or web-based places (eg, bars and adult-focused sexual health clinics) or may not pay attention to campaign materials in public spaces that portray intimacy among adults.

This Study

In this study, we used a "designing for dissemination and sustainability" approach, which uses methods and principles from dissemination and implementation science, communications, marketing, community engagement, and human-centered design. This interdisciplinary approach enables the tangible products of public health research to better meet the long-term needs of the intended audience and settings [41]. Following this approach, we directly engaged SGM adolescents in preparatory implementation research via web-based focus groups to inform whether and how to undertake an SGM adolescent–centered PrEP social marketing campaign in the Chicago area. Focus group questions sought to understand their perspectives on existing PrEP and sexual health campaigns and their preferences for campaign content, design, and implementation. At the end of the focus groups, we invited SGM adolescents to develop initial prototypes of digital marketing materials that they thought may resonate with a teenaged audience [42,43].

Methods

These data were collected in the context of a larger study that focused on lesbian, gay, bisexual, transgender, and queer (LGBTQ+) adolescents and supportive adults' perspectives on the need for an adolescent-centered PrEP campaign in the Chicago area and what each group desired out of such a campaign. Findings from interviews with adult participants are reported elsewhere [24].

Recruitment

SGM adolescents in Cook County, Illinois, were recruited from February to May 2021 via social media advertisements (eg, Instagram [Meta Platforms] and Snapchat [Snap Inc]); the Illinois PrEP Working Group listserv; registries of individuals who had expressed interest in research study participation through Northwestern University's Institute for Sexual and Gender Minority Health and Wellbeing and the Third Coast Center for AIDS Research; and through the research team's partnership with AIDS Foundation Chicago, a local HIV advocacy and service organization. Social media advertisements were designed to reach adolescents interested in SGM culture (eg, LGBTQ+ organizations and important figures), and study staff reached out to individuals in the registries via SMS text message or email to inform them of a new study opportunity. Social media advertisements and registry outreach directed prospective participants to the study's web-based screener. Eligible adolescents included those who (1) were aged 13 to 19 years; (2) lived in Cook County, Illinois; (3) identified as a sexual or gender minority individual; (4) self-reported HIV negative or unknown status; (5) had access to the internet; and (6) could read in English at an eighth-grade level. We used purposive sampling to enroll up to 20% transgender and nonbinary youths and at least 80% cisgender sexual minority youths to ensure representation of youths across the LGBTQ+ spectrum and aimed for a sample composition of at least 60% racial and ethnic minority youths to reflect the demographics of the Chicago area.

Ethical Considerations

All study procedures were approved by the Northwestern University Institutional Review Board (STU00213036). A waiver of parental permission was obtained for minor adolescents, as the study was determined to pose minimal risk. Following completion of the web-based screener, those who were preliminarily eligible were given a web-based consent form and responded to 4 multiple-choice items to assess their comprehension of the study procedures; those who failed these items twice after being given the opportunity to reread the consent form were deemed ineligible. Then, study staff contacted preliminarily eligible adolescents to participate in a Zoom (Zoom Communications, Inc) call, during which staff confirmed that the respondent was indeed an adolescent, met the eligibility criteria, understood the study procedures, and agreed to a code of conduct for focus group behavior (eg, no harassment or bullying, respectful language, and keeping discussion confidential). To protect their privacy, participants were encouraged to select a username for the focus groups that they did not use elsewhere on the Internet and were not permitted to include identifying information in their usernames. Throughout the study, participants were reminded that they could contact the study team at any time via study SMS text message or email with questions or concerns. At the end of the study, information about study payment (US $40 Amazon gift card sent via email) and HIV and PrEP resources were provided. Data presented here are deidentified for privacy.

XSL•FO
RenderX

Procedures

After enrolling, participants were sent a link to a 25-minute web-based baseline survey via email or text. Participants could select "I don't want to answer" for any item in the survey. Study staff then sent adolescents instructions to access an asynchronous focus group website developed using vBulletin (Internet Brands), web-based forum software that the study team had used in previous research [11,12]. Web-based focus groups have long been used in social science research due to their accessibility and flexibility (eg, they can be conducted synchronously or asynchronously), including in the research team's own work [44,45]. They have also been shown to provide data quality comparable to that of in-person focus groups [46]. Participants accessed the focus group website with unique log-ins. There, they could view links to an introductory forum and 7 forums focused on different study topics, with 1 to 5 questions within each topic (refer to Focus Group Guide section). The introductory forum included a video introducing the study team to participants, an overview of ground rules, and an icebreaker. Participants contributed by typing their responses or uploading content at their convenience. The focus group website was available for 7 days, and participants were strongly encouraged to complete all questions within the first 2 to 3 days to ensure sufficient time for the moderators to ask follow-up questions and for participants to interact with each other's written responses. To increase the likelihood of participant responses, probes used the @username function to trigger notifications. In addition, daily text messages or email reminders that included a direct link to questions and probes participants had not yet responded to were used to increase their engagement. We conducted 5 focus groups, with each group including 11 to 17 participants. Each group was facilitated by 2 moderators who were trained in qualitative methods. Adolescents who answered all questions on ≥4 of the 7 topics were considered completers and eligible for payment.

Focus Group Guide

The focus group guide assessed participants' reactions to existing PrEP campaigns and preferences for future adolescent PrEP campaigns. Questions were adapted from an unpublished needs assessment previously conducted by local community leaders working in HIV prevention and care, including one of the coauthors (JP), to inform the development of adult-focused English and Spanish PrEP social marketing campaigns in Chicago. Additional questions were added by the study team to be inclusive of adolescent experiences based on findings from empirical research on adolescent barriers and facilitators to PrEP and adolescents' preferences for receiving information about sexual health [11,13,25]. Textbox 1 provides examples of focus group questions and topics.

First, we presented existing PrEP and HIV campaigns (PrEP4Love [32-34], Get Yourself Tested [35], and the National Youth HIV/AIDS Awareness Day Campaign [36]) to give participants a reference point for public health campaigns focused on HIV and sexual health and those focused on youths versus adults, given that no PrEP campaigns for adolescents existed at the time. Questions assessed participants' reactions to these campaigns, areas of improvement, and comparisons. Next, we assessed adolescents' preferences for message delivery and dissemination, preferences on campaign design and content, developmental and cultural considerations, and perspectives on messaging for parents and injectable PrEP (not described in this study). Finally, a creative activity prompted adolescents to design mock PrEP campaign content; provide a sample image or video of a campaign they liked and an explanation; or develop a mood board reflecting colors, feelings, or content that might inspire a future campaign. At the end of the group, participants were invited to share their feedback, ask questions, or share additional thoughts they may have had that were not covered. Each focus group transcript was saved as a deidentified Microsoft Word document on password-protected servers. Upon downloading focus group transcripts, study staff disabled further user comments and visibility of the group to all users.

XSL•FO
RenderX

Textbox 1. Focus group topics and sample questions.

Topic 1: Your thoughts on the PrEP4Love campaign

- What works well in this campaign? What doesn't work so well in this campaign?
- PrEP4Love was designed for adults. What changes would you make to the campaign to make it more appealing and interesting to LGBTQ+ teens?

Topic 2: Your thoughts on the Get Yourself Tested and National Youth HIV/AIDS Awareness Day campaigns

- Compare and contrast the PrEP4Love, GYT, and NYHAAD campaigns. What aspects of each campaign do you like the best? What do you not like?
- Which of the aspects you mentioned above are the most important to you to include in a PrEP campaign for LGBTQ+ teens?
- What, if anything, might be missing from any of these campaigns that would be important to include in a PrEP campaign for LGBTQ+ teens?

Topic 3: Where should PrEP campaigns go?

- What in-person spaces would be good places to post ads/share info about PrEP for LGBTQ teens?
- What online spaces/platforms would be good places to post ads/share info about PrEP for LGBTQ teens?
- If you were to design a PrEP video for TikTok, what would your video look like?
- If you were to design a PrEP post or story on Instagram, what would it look like?

Topic 4: What should PrEP messages for LGBTQ+ teens in Chicago look like

- What words and/or phrases would capture your and other LGBTQ+ teens' attention?
- What types of images would you want to include in your PrEP campaign?
- If you wanted to show people in your PrEP campaign, what type of people would you show?

Topic 5: LGBTQ+ teens' feelings and reactions to PrEP campaigns

- How would you want a new PrEP campaign to make you and other LGBTQ+ teens feel?
- How (if at all) should teen PrEP campaigns differ for younger versus older LGBTQ+ teens? LGBTQ+ teens who ever had sex versus those who never had sex? LGBTQ+ teens who are out versus those who are not out? Teens who identify as LGBTQ+ versus teens who are questioning or do not identify as LGBTQ+ but have sex with same sex/gender partners?

Topic 6: PrEP messaging for adults and other forms of PrEP

- How do you think PrEP messages for parents and other adults need to be DIFFERENT from those aimed at LGBTQ teens? How should they be SIMILAR?
- In your opinion, what information is most important in an ad about teens and PrEP designed for parents/guardians?
- If you were designing a campaign for injectable PrEP, what is the one piece of information you'd want in it? How would it be different from a PrEP campaign focused on pills?

Topic 7: Your examples of LGBTQ+ teen PrEP campaigns!

- Option 1: Find an example of health-related (or PrEP/HIV/sexual health-related) content that you LIKE or LOVE! Then, explain what you like about it, and how it relates to your ideas for what a PrEP campaign for teens should be like.
- Option 2: The DIY route. Here are some ideas to get you started, but if you have other ideas feel free to run with it! Use apps like Instagram or Snapchat to edit/draw on/type on the campaigns we shared with you, create a mood board/collage to share the vibe and look you'd want your campaign to have, draw something on your phone.

Measures

Sociodemographic Characteristics

Participants self-reported their age, race, ethnicity, sex assigned at birth, sexual orientation identity, gender identity, outness to parents or guardians, parents' or guardians' acceptance of SGM identity, and sources of sexual health information. Additional investigator-created items assessed participants' HIV and STI testing history, previous sexual experience, and engagement in HIV transmission risk behavior, and other items assessed perceived HIV risk [47].

General PrEP Questions

Investigator-created items assessed perceived norms of PrEP uptake ("I'd like you to imagine 100 adolescents like you. How many of these 100 adolescents do you think are likely to use PrEP"), PrEP awareness before the study (*yes* or *no*), and lifetime uptake (*yes*; *no*). Participants also self-reported perceived difficulty (1=*not at all difficult* to 5=*very difficult*)

XSL•FO
RenderX

and confidence in their ability to access PrEP (1=*not at all confident* to 5=*very confident*).

PrEP Attitudes and Stigma

A scale developed by Walsh [48] in 2019 was used to assess participants' attitudes and perceived stigma toward PrEP. A total of 5 items evaluated perceived effectiveness, safety, and adherence to PrEP, while another 5 items addressed stigma concerns associated with PrEP (eg, "People who take PrEP are promiscuous"). Responses were on a 5-point Likert scale (1=*strongly disagree* to 5=*strongly agree*). PrEP attitude scores ranged from 5 to 25, with higher scores indicating a positive view of PrEP. Stigma scores also ranged from 5 to 25, with higher scores indicating greater perceived stigma.

Data Analysis

Descriptive analyses for all quantitative variables were conducted in R software (x64, version 3.6.1; R Foundation for Statistical Computing). Focus group data were analyzed deductively using rapid qualitative analysis [49], a method commonly used in public health and implementation research to quickly identify actionable themes in qualitative data. The results of such analyses have been shown to correlate highly with traditional qualitative analysis [50,51]. First, study staff exported participants' responses from the Microsoft Word group-level transcripts to tables in Microsoft Excel, with different tabs created for each focus group topic, and responses were cross-validated for accuracy. Two coders reduced data by emphasizing informative sections of participants' responses and removing redundant or filler words. This process resulted in concise data that spoke to the study objectives. Following data reduction, the coders independently examined all excerpts, assigned codes to all excerpts, refined the code definitions iteratively, and held discussions to reconcile differences. Data were then organized conceptually and reported according to the 4 Ps model. Quotes presented in the Results section include participants' age, gender, race, ethnicity, and sexual orientation.

Results

Screening, Enrollment, and Analytic Sample Characteristics

Overall, 348 individuals accessed the eligibility screener, of whom 126 (36.2%) completed the screener. Of the 126 participants who completed the screener, 96 (76.2%) were preliminarily eligible. Of the preliminarily eligible individuals, 68% (65/96) were enrolled after the Zoom call. Then, 94% (61/65) completed the baseline survey and 92% (56/61) answered ≥1 question in the focus group, and 89% (50/56) completed the focus group (ie, responded to all questions in the majority; 4/7, 57% of the topics); 94% (47/50) participants responded to all questions in all 7 topics. The characteristics of the analytic sample are in Table 1.

XSL•FO
RenderX

Table 1. Participant demographics (N=56).

Variables	Values
Age (y), mean (SD)	18.16 (1.22)
Age (y), n (%)	
14	1 (2)
15	2 (4)
16	5 (9)
17	16 (29)
18	16 (29)
19	16 (29)
Sex assigned at birth, n (%)	
Male	28 (50)
Female	28 (50)
Race and ethnicity, n (%)	
Asian	6 (11)
Black	9 (16)
Hispanic	12 (21)
Multiracial	9 (16)
Non-Hispanic White	19 (34)
No answer	1 (2)
Gender identity, n (%)	
Agender	1 (2)
Cisgender man	20 (36)
Cisgender woman	12 (21)
Transgender man	5 (9)
Transgender woman	3 (5)
Genderqueer	4 (7)
Gender nonconforming	3 (5)
Questioning	6 (11)
No answer	2 (4)
Sexual orientation identity, n (%)	
Asexual	1 (2)
Bisexual or pansexual	23 (41)
Gay	20 (36)
Lesbian	3 (5)
Queer	6 (11)
Questioning	3 (5)
Prior sexual experience (allowed multiple options), n (%)	
People with penises	33 (59)
People with vaginas	17 (30)
Never had penetrative sex	19 (34)

Participants could indicate multiple sources from where they heard of the study. Most participants heard of the study through paid social media advertisements (23/56, 41% via Instagram; 10/56, 18% via Facebook; and 6/56, 11% via Snapchat). Other sources included a registry email or text message (7/56, 13%), an LGBTQ+ organization (5/56, 9%), a friend (5/56, 9%), or a

XSL•FO
RenderX

clinic or hospital (2/56, 4%), and 2 (4%) participants indicated that they heard about the study elsewhere (a family member and an email newsletter for a different study for LGBTQ youth). Participants said that their main sources of sexual health information were social media (30/56, 54%), LGBTQ+ websites (15/56, 27%), and porn websites (14/56, 25%). More than half (37/56, 66%) of the participants had prior sexual experience, 48% (27/56) reported condomless penetrative sex, and 13% (7/56) had sex with someone who tested positive for HIV or who was of unknown status. Less than half self-reported that they had previously tested for HIV (24/56, 43%) or STIs (18/56, 32%).

One-third (19/56, 34%) of the participants worried about testing positive for HIV, and 24% (13/56) perceived themselves at risk for HIV infection. ANOVAs revealed significant group differences in perceived risk for HIV infection by gender identity ($F_{7,55}$=2.47; P=.03); Bonferroni-corrected pairwise comparisons showed that those who identified as men perceived greater risk than those who identified as women (mean 2.56, SD 1.26 vs mean 1.4, SD 0.63, respectively). There were also significant group differences in sexual orientation ($F_{5,55}$=2.79; P=.03) whereby participants identifying as gay perceived greater risk than those identifying as bisexual or pansexual (mean 2.80, SD 1.20 vs mean 1.78, SD 0.90, respectively). There were no significant differences in perceived HIV risk by age, race, ethnicity, sex assigned at birth, or having ever been tested for HIV or STIs.

Most (39/56, 70%) participants were aware of PrEP. Only 4% (2/56) had taken it, and 95% (53/56) were unaware that it could be prescribed to those aged <18 years. In addition, 38% (21/56) of the participants found accessing PrEP easy, while 50% (28/56) were confident in their ability to do so. Adolescents viewed PrEP positively regarding effectiveness, safety, and adherence (mean 19.93, SD 2.91; range 5-25) and had moderate stigma levels (mean 11.91, SD 4.39; range 5-25). More details are provided in Table 2.

Table 2. Pre-exposure prophylaxis (PrEP) awareness, use, and attitudes (N=56).

Variables	Values
Perceived norms (“Out of 100 adolescents like you how many do you think are likely to use PrEP?”), mean (SD)	37.88 (24.11)
PrEP attitudes, mean (SD; range)[a]	19.93 (2.91; 5-25)
PrEP stigma, mean (SD; range)[b]	11.91 (4.39; 5-25)
Participants’ awareness of PrEP before participating in the study, n (%)	
Yes	39 (70)
No	16 (29)
No answer	1 (2)
Do you know any teens aged under age 18 who are taking PrEP? n (%)	
Yes	3 (5)
No	53 (95)
Lifetime PrEP use, n (%)	
Yes	2 (4)
No	54 (96)
Perceived difficulty in accessing PrEP, n (%)	
Very easy	7 (12)
Somewhat easy	14 (25)
Neither difficult nor easy	17 (30)
Somewhat difficult	17 (30)
Very difficult	1 (2)
Confidence in the ability to access PrEP, n (%)	
Not at all confident	4 (7)
Somewhat unconfident	9 (16)
Neither confident nor unconfident	15 (27)
Somewhat confident	22 (39)
Very confident	6 (11)

[a]Higher scores indicate more positive attitudes toward PrEP.

[b]Higher scores indicate greater stigma against PrEP use.

XSL•FO
RenderX

Reactions to Existing PrEP Campaigns

Participants were presented with the PrEP4Love, Get Yourself Tested, and the National Youth HIV/AIDS Awareness Day Campaigns and were asked to react to the design, images, and written messages of each campaign. Adolescents generally favored campaigns that highlighted "diversity" in sexual and gender identities, body type, and couple types (eg, couples depicting transgender and cisgender individuals) and disliked campaigns that lacked representation of these groups, including adolescents:

> *I really like that the images show many different couples, regardless of race, gender, sexuality, etc.* [Asian gay man aged 19 years]

That said, they noted that the intentionally sensual nature of the PrEP4Love campaign was off-putting:

> *It seems too sexual/sensual and likely would make me look away before reading the [tag]line.* [Non-Hispanic White bisexual or pansexual man aged 17 years]

Regarding campaign taglines and copy, SGM adolescents favored "straightforward" messages, particularly those that challenged HIV stigma:

> *The simplicity of the campaign definitely works well. I think that it conveys simple information...while simultaneously being informative enough.* [Hispanic bisexual man aged 17 years]

Respondents expressed dislike for messages that were too broad, lacked clarity, or failed to provide PrEP-related resources or information. Adolescents noted that the National Youth HIV/AIDS Awareness Day campaign messages were "too busy and wordy" (non-Hispanic White bisexual or pansexual woman aged 17 years), and that the Get Yourself Tested campaign had a clear call to action but lacked further resources:

> *For GYT...I do think that there could have been a little bit more information.* [Hispanic questioning woman aged 19 years]

Participants preferred existing campaigns with "bright, eye-catching designs" and disliked campaigns with fonts that were difficult to read or those that did not feature teenaged models.

Preferences for Future Adolescent-Centered PrEP Campaigns

In this section, participants' preferences regarding the design of future adolescent-centered PrEP campaigns were organized based on the 4 Ps framework [29].

Product

Participants' comments suggested that social marketing campaigns should specify certain details about PrEP and HIV prevention for adolescents; mention its benefits among youths who have used PrEP; include messages that normalize the health, experiences, and relationships of SGM adolescents; and provide ample resources about PrEP, including information about how to talk to a health care provider about PrEP.

Participants described wanting campaigns to give information regarding the purpose and effectiveness of PrEP and how to access it:

> *I think it should include more...about HIV, what PrEP is, and the cost of PrEP.* [White queer, genderqueer adolescent assigned male at birth aged 19 years]

Participants also desired more tangible information about PrEP statistics and adolescents' lived experiences obtaining and being on PrEP:

> *Including statistics, quotes, and real experiences from adolescents would make it more relatable and emphasize the importance.* [Hispanic White questioning woman aged 19 years]

Other participants also emphasized the importance of including resources about PrEP access, such as offering a comprehensive list of local adolescent-friendly PrEP providers, information about talking to a provider about PrEP, and resources about insurance coverage and paying for PrEP. A , mentioned as follows:

> *I think you're missing financial information [is it free?], contact information...and you can say "talk to your doctor" all you want but like, when I was [younger], I didn't have a doctor. Who would I go to?* [Black transgender man questioning their sexual orientation aged 19 years]

A participant summarized common sentiments among the group regarding product information and how to frame this information to appeal to the audience:

> *I think that the dosing, how to get it, why it's good to take, and information regarding how it works [in a simplified manner], are all good things to include. If you just have a poster saying, "Take PrEP," nobody will take it, but if you say something closer to "Take PrEP, a once daily medication that keeps you 99% safe from HIV," it will be more effective.* [White bisexual woman aged 19 years]

Respondents noted that PrEP messaging centered on the needs of SGM adolescents, and not exclusively focusing on SGM populations, would likely be beneficial to adolescents regardless of background or identity. This was not only perceived to help normalize the health, experiences, and relationships of SGM adolescents but also showed cisgender heterosexual adolescents that PrEP could be for them as well (eg, campaign materials that say "PrEP is for everyone" and depict adolescents of various sexual orientations and genders). This focus on inclusive yet relatable messaging is reflected as follows:

> *The way that my friends explain PrEP to me is...you take the pill once a day, kinda like birth control [which is reassuring because we all know someone on birth control].* [Black gay man aged 18 years]

Furthermore, youths described how portraying SGM adolescents authentically, "just being themselves, like watching TV or enjoying outdoor activities" (Black gay man aged 17 years), can facilitate this sentiment in PrEP campaigns. Finally, another benefit of adolescent-centered PrEP campaigns is that they can

XSL•FO
RenderX

foster a sense of inclusion and empowerment among adolescents, which can help motivate them to consider taking action to use PrEP:

> *I want to feel valued, supported, and seen, where people genuinely care about my sexual health and well-being.* [Asian gay man aged 18 years]

Price

Most comments identified in this domain referenced the price of PrEP (eg, "A big part would be price of PrEP...and any side effects," as mentioned by a Hispanic gay man aged 16 years), the effort involved in accessing it, and the possibility of stigma and embarrassment with engaging with the campaign or seeking PrEP. Fewer adolescents expressed concerns about side effects or the medication regimen itself.

One transgender adolescent explained the potential concerns regarding PrEP that a campaign may help address by relating PrEP to their experiences seeking gender-affirming care:

> *Like when I was younger, I really wanted to go on hormones but...it felt like it would be impossible to get because I didn't fully understand how to access it and a lot of people were telling me how complicated it was. So, I think teens need reassurance that it is simple/easy to obtain prep and maybe even include a step-by-step guide on how to go on it?* [White bisexual transgender man aged 19 years]

Other participants described how engaging with a PrEP campaign or talking about PrEP to a provider or clinic staff member could be embarrassing or elicit judgment from others, which campaigns should proactively address:

> *I don't want to feel ashamed of talking about the topic. I want to feel respected and that I don't need to fear that others might judge me.* [Black gay woman aged 18 years]

Similarly, other youths reported parental involvement as a potential barrier to accessing PrEP:

> *If we go to our doctors for a routine checkup, we'd be scared to talk about sex in general...because most teens are going to be intruded on by parents...it's surely a roadblock for many teens.* [Black gay man aged 18 years]

Moreover, another participant described ensuring that the campaign had multiple ways to speak with a person about PrEP, such as a text line or web-based chat function, given young people's reluctance to speak on the phone:

> *I would include specific ways to get into contact about getting PrEP besides just a phone call because many teens would be way too anxious to call a stranger about that.* [Non-Hispanic White gay man aged 19 years]

Participants were asked to articulate possible drawbacks of adolescent PrEP campaigns focused on specific populations (eg, SGM individuals, racial and ethnic minority individuals, and younger adolescents). Although campaigns specifically for SGM individuals can ensure that PrEP awareness is increased among this population as well as address certain concerns that these populations may have in relation to PrEP ("Black and Latinx people lack medical access and are wary of the medical system because it has mistreated them lots of times," as mentioned by a Hispanic bisexual woman aged 18), participants described that this approach could downplay HIV susceptibility among cisgender heterosexual adolescents and perpetuate stigma within and toward the SGM community. A Black gay woman aged 18 years summarized this concern, stating the following:

> *...it may convey the message that HIV solely affects the LGBTQ community, rather than being a universal virus that impacts all of us.*

Others suggested how campaigns should be attuned to developmental differences in adolescents, with older adolescents being more comfortable with more explicit sexual information or references and younger adolescents possibly experiencing stigma in searching for sexual health information:

> *I feel like some younger people may feel embarrassment or shame in looking into preventative measures.* [Non-Hispanic White bisexual woman aged 19 years]

Place

For the domain of place, participants primarily provided recommendations on how specific individual "champions" or locations that can support youth well-being and health behavior could effectively communicate where and how adolescents could access PrEP, particularly for those who may be less familiar with navigating sexual health care. The following participant quoted below and others described how key influential adults and organizations (eg, schools and accepting churches) could play a role in supporting adolescents' access to PrEP, even if these locations did not offer PrEP themselves:

> *...every school should have guidance counselors or at least a support group for LGBTQ+ teens. I go to a catholic school and recently this year we made a new group called "Unidos" ['Together' in English]. That group helped me so much with resources and it made me feel represented.* [Hispanic gay man aged 15 years]

In addition, physical spaces recommended by participants included schools and public and entertainment venues, such as malls, public libraries, public transportation, and restaurants. These locations were favored due to their accessibility and convenience for SGM adolescents. For example, some youths identified public libraries as places that are accessible to and safe for adolescents, where they could learn more about PrEP with relative privacy:

> *I also think that libraries could work well because they are very trustworthy and when teenagers go to libraries, they are often alone to study or find books.* [White bisexual woman aged 17 years]

Participants also recommended advertising in LGBTQ+ serving organizations, clinics such as Planned Parenthood, pharmacies, and health care provider waiting rooms. These participants

XSL•FO
RenderX

indicated that the placement of marketing materials in waiting rooms could be beneficial:

> *Family doc waiting rooms...if you have to talk to your doctor about PrEP a reminder before going in would be good.* [White bisexual woman aged 18 years]

The placement of marketing materials in pharmacies and hospitals was echoed by other participants as follows:

> *...because it will generate a lot of trust from teens because it is in a healthcare facility. When you go get a check-up it will be easy to remember, and teens might even like to bring it up with their doctor in private.* [Black gay woman aged 18 years]

Promotion

Participants' feedback predominantly focused on the element of promotion, and they articulated strategies that can raise awareness about PrEP among SGM adolescents through carefully crafted content, appealing graphic design, and strategic placement. Finally, participants were also asked to give examples of what these campaign materials might look like.

Campaign Content

This element was distinct from "product" in that it focuses more on message framing, tone, word choice, and market segmentation than the features or benefits of the product itself. Regarding the phrasing of adolescent PrEP messages, participants expressed a preference for concise messages presented in casual language to increase engagement. Furthermore, adolescents highlighted that a campaign should have an affirming and empowering tone that is community-oriented, with several adolescents offering phrases such as "your health matters" or "let's stop HIV together."

Considering the influence of peers and role models on adolescents, participants also suggested including celebrities and social media influencers in future PrEP campaigns to increase campaign appeal:

> *I would feature a well-known celebrity who is attractive and has been a strong supporter of the LGBTQ community.* [Hispanic gay man aged 16 years]

Participants frequently recommended involving adult authority figures, such as scientists, health care providers, and educators, for their credibility and expertise in addressing concerns about PrEP side effects or potential long-term effects, provided their delivery was empowering and supportive rather than patronizing:

> *...doctors/healthcare providers. This would help make PrEP campaigns feel more reliable and trustworthy.* [Black bisexual or pansexual woman aged 17 years]

Moreover, SGM adolescents suggested featuring parents and other adults who may play influential roles in adolescents' health care (eg, "parents who feel more comfortable with their adolescents taking PrEP," as mentioned by an Asian gay man).

Although participants felt that adolescents should get the same access to basic PrEP information regardless of where they lived or who they were, they acknowledged the importance of content segmented by population to address key differences:

> *As someone who went to school in the 'burbs but lives in the city, it tends to be a bit more conservative up there, so I would just be aware of that.* [White queer man aged 19 years]
>
> *you'd have to advertise in different types of locations, as suburban teens don't take public transportation for example.* [Hispanic bisexual man aged 17 years]

This included creating Spanish language materials to reach the large Spanish-speaking community in Chicago, which could ultimately improve access:

> *...this community, which I'm also a part of, could benefit [by] having ads that are also in Spanish. If possible, just adapting ads to be more approachable and informative for other languages other than English.* [Hispanic bisexual woman aged 18 years]

Campaign Design

Regarding campaign design, adolescents preferred eye-catching graphics, including cartoons, avatars, and infographics:

> *...well thought out color schemes and easy-to-read fonts, along with statistical evidence that explains the effectiveness of PrEP [such as pie charts or bar graphs].* [White gay man aged 18 years]

Adolescents suggested that campaigns should avoid overly sexual images for discretion, with a Hispanic bisexual or pansexual woman aged 18 years stating, "staying away from anything that reinforces stereotypes or is overly sexual (due to parents)." Moreover, participants were clear that the campaign should avoid seeming too slick or corporate, which was viewed as untrustworthy, as described by a participant:

> *...you need to get that attention in a way that is personable, to show that this isn't just something being peddled by a brand, and that it is something genuine and important, from queer folk, to queer folk...Make it real, make it raw, adding polish at the end removes a level of authenticity that is necessary for this level of trust.* [White queer transgender woman aged 19 years]

Although adolescents acknowledged that there are benefits to market segmentation, they were emphatic that the campaign design should also be inclusive and diverse as they believed everyone could benefit from learning about PrEP:

> *I wouldn't necessarily exclude white bodies, but I would put an emphasis on people of color...it hurts literally no one to include them more and really only benefits. also, I think there shouldn't be much distinction between cis and trans [people regarding campaign content], but I do think that we need a bit more trans representation.* [Non-Hispanic White queer man aged 19 years]

Campaign Placement

Participants emphasized that an adolescent-centered PrEP campaign should predominantly be digital to reach this age group regardless of their location. They suggested using smartphone apps and LGBTQ+ websites; social media, such as

XSL•FO
RenderX

Instagram, TikTok, and Tumblr; dating apps (eg, Grindr and Tinder); music streaming apps (eg, Spotify); and mobile games. Among these, social media, particularly TikTok, emerged as the most widely discussed digital platform for campaign dissemination as it was perceived to have "a significant reach and tends to be LGBTQ+ friendly" (Hispanic questioning woman aged 19 years). Participants suggested avoiding Facebook ("I don't know many adolescents who actually use it. It's more popular with adults and the older generation," as mentioned by a Hispanic bisexual or pansexual woman aged 18 years) and digital spaces that feature skippable advertisements (eg, "ads on video streaming services because many of us skip them and don't even pay attention to them," as mentioned by a Hispanic questioning man aged 16 years) or have a reputation for tolerating more toxic Internet discourse (eg, Reddit and Twitter (subsequently rebranded X).

Regardless of medium, participants suggested that physical and digital campaign assets all lead to centralized resources:

> *Having a QR Code that leads to a website with more information on physical advertisements...the idea of a hotline is fantastic.* [White gay man aged 18 years]

They also suggested that such resources should be comprehensive enough so that adolescents do not have to hunt for information elsewhere:

> *The website should include all possible information about PrEP so that it's a one stop shop and so there is no way to walk into it blind...it will inform and persuade because they can just look at it and know, rather than having...to look all of it up on their own.* [White bisexual woman aged 19 years]

Example Campaigns From Focus Group Participants

At the end of the focus group, participants were given the following prompt: "We just asked you a bunch of questions about what a PrEP campaign for LGBTQ+ teens might look like. But some folks might prefer to share their thoughts visually."

Participants were given the option to share graphics or existing sexual health content that evoked what they wanted from a PrEP campaign or to design an example from scratch. In this section, we provide several representative examples and their rationales.

Participants who submitted an original entry most frequently gave examples formatted for social media, such as Instagram or TikTok. This participant described clarity, directness, and inclusion in their design:

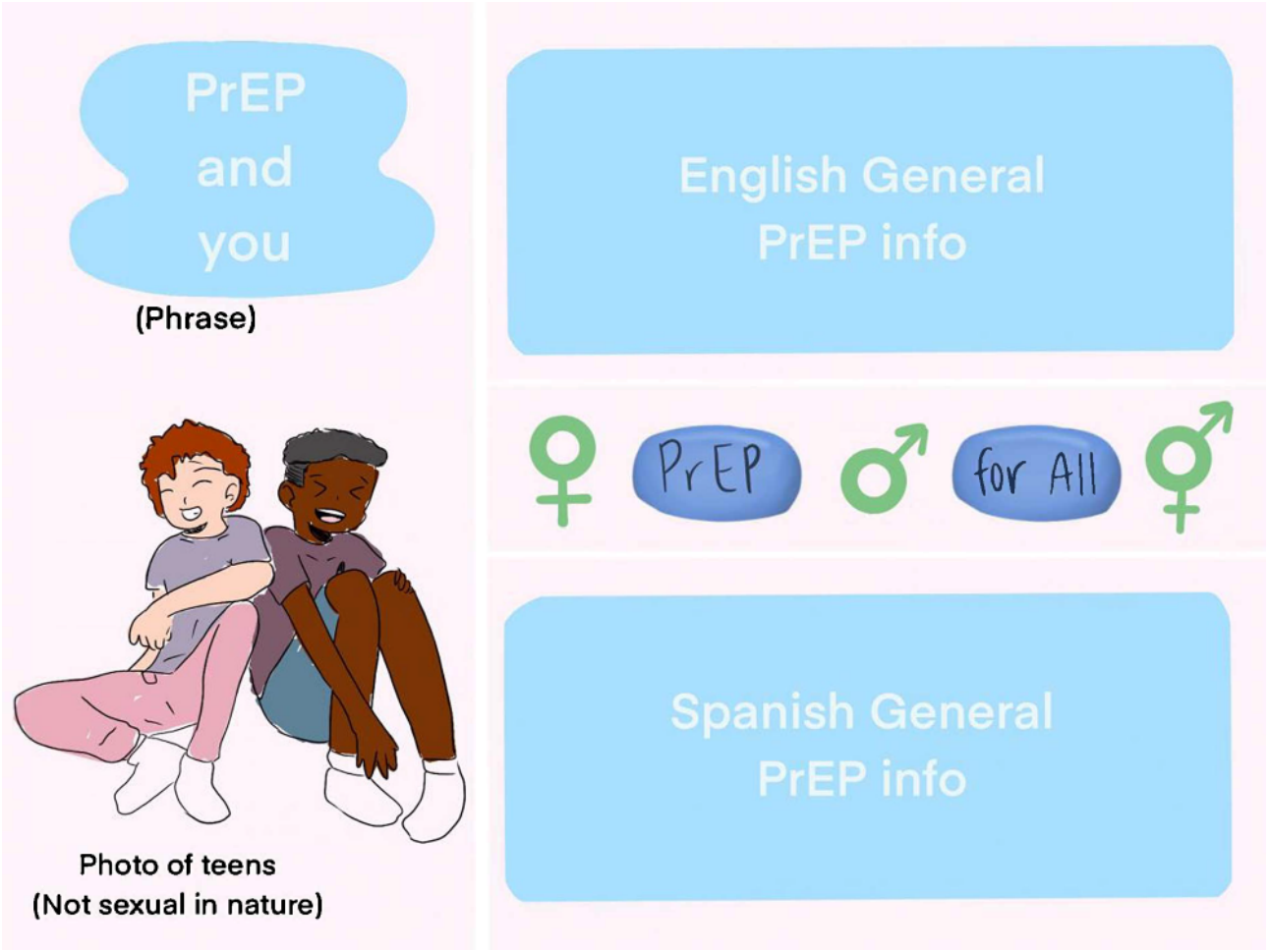
> *The idea behind this one [Figure 1] was an Instagram ad that has simplicity, there's a hook at the top, only 2 things to read, cute stickers, an option to swipe up for more info. And the two things I found most important were the hashtag that I was able to make rainbow! And the symbols for the other social media. I was able to get my point across that prep is for everyone, and a neat little reference is that the background somewhat resembles the bisexual flag, hinting that it doesn't matter if you have sex with men or women [or both or anywhere in between] and that prep is still for you.* [Hispanic gay man aged 17 years]

Others submitted examples that included information in both English and Spanish, referencing the conversation earlier in the focus group about the inclusion of Spanish speakers (Figure 2):

Figure 1. Social media pre-exposure prophylaxis (PrEP) campaign asset reflecting inclusive messaging and calls to action.

XSL•FO
RenderX

Figure 2. Social media pre-exposure prophylaxis (PrEP) campaign asset reflecting inclusive messaging and English- and Spanish-language information.



> *I figured that incorporating teens in the campaign in a nonsexual nature would...draw attention—these teens are also different, appearance-wise. For the information, I saw on a different forum that there should probably be information in both Spanish and English. Separating the information are gender symbols and PrEP pills that say, "PrEP for all" on them.* [Black gay woman aged 18 years]

One participant created a script for a commercial reminiscent of the US Trojan Man condom commercials [52] of the mid to late 1990s (Textbox 2):

> *I wrote a quick script for a hypothetical commercial. It's aimed at a [younger?] teen audience and ideal for platforms like TikTok, Snapchat, and Instagram. I think it could be either live-action or animated. This one would maybe be best aimed at teens who have not yet started having sex considering that's the situation in the commercial. The characters can be any gender/body type/sexual orientation. I also wrote it to be slightly raunchy since that will certainly appeal to/get teens attention.* [White bisexual man aged 19 years]

XSL•FO
RenderX

Textbox 2. Excerpt of proposed commercial or brief video for social marketing campaign.

INT. A TEEN BOY'S ROOM LATE AT NIGHT

Two teens make out in bed until TEEN 1 pulls away.

TEEN 1

You sure you're ready?

TEEN 2

(smiling warmly)

Mhm.

TEEN 2 leans into continue kissing TEEN 1 when CAPTAIN PrEP, a caped, over-the-top superhero with an anti-HIV logo or a pill logo on their shield, flies through the wall. The terrified teens scream.

CAPTAIN PrEP

Fear not, young horndogs. It is only Captain PrEP here to make sure you don't get HIV.

TEEN 1

What the f*ck, you destroyed my room!

Captain PrEP is unfazed.

CAPTAIN PrEP

(earnestly)

We all make sacrifices. But your sexual health shouldn't be one of them.

Discussion

Principal Findings

PrEP has the potential to dramatically reduce the number of HIV infections among adolescents in the United States. However, its use among young people remains low, with individual studies reporting that <0.5% of adolescents who may benefit from PrEP have used it [53,54]. Social marketing campaigns have the potential to improve PrEP awareness and uptake [24,26,32]; however, limited research has explored what adolescents want out of adolescent-centered PrEP campaigns. To our knowledge, this is among the first studies on SGM adolescents' perspectives on existing and future PrEP marketing and implementation. Our findings, framed around the 4 Ps of social marketing, suggest that previous sexual health campaigns tailored for SGM adults, or adolescents in general, do not meet the needs of SGM adolescents and that youths themselves can be involved in conceptualizing the content and implementation of PrEP campaigns [43]. In this section, we discuss concrete takeaways from the focus groups that public health practitioners and implementers may consider incorporating into adolescent PrEP campaigns.

Overall, what adolescents wanted from PrEP campaigns aligned with our findings from interviews with adults who worked with or parented LGBTQ+ adolescents about what should go into an adolescent-centered PrEP campaign [24]. For example, adults wanted adolescent-centered PrEP campaigns to offer basic information about PrEP (eg, safety, how it works, access, and costs) that was inclusive, destigmatizing, and aligned PrEP with adolescents' priorities (eg, romantic relationships). In addition, adults wanted concrete, actionable information for adolescents to act on (eg, call this hotline and go to this clinic), delivered through social media and physical spaces where adolescents spend most of their time. For instance, regarding *product* and *price*, SGM adolescents expressed a desire for basic PrEP information in these campaigns that address potential barriers and costs adolescents may encounter. Several years after PrEP's Food and Drug Administration approval for adolescents, access to consumer-facing, adolescent-specific information on PrEP in the United States remains sparse. Although there are some examples of consumer-facing information that indicates that adolescents can take PrEP [55], there are few details addressing the developmental issues that surfaced in this study and previous research on adolescents and PrEP [8,11,12]. An adolescent-focused social marketing campaign could offer answers to the following questions, some of which may need to be tailored to reflect current state laws and policies: Do I need to involve my parents in the decision to take PrEP, because in some states, PrEP is explicitly included in mature minors' right to health care, and in others, it is unclear or implied [56]? What health care providers or clinics can prescribe PrEP to minors, as web-based PrEP locator tools lack information about whether locations serve adolescents? I have never been to a health care provider on my own, so how do I prepare for a PrEP appointment? How do I pay for PrEP without my parents knowing?

Moreover, because PrEP requires consistent health care engagement (eg, for oral PrEP, quarterly appointments are recommended for HIV testing and lab work to check for side effects), campaigns should offer resources that help youth feel more comfortable independently accessing care given that they may have limited experience doing so. Options to cover or defray PrEP medication and related follow-up care should be made explicit (eg, coupons from pharmaceutical companies and requirements for insurance companies to cover PrEP prescriptions and services via the United States Affordable Care Act).

XSL•FO
RenderX

In general, participants in this study were more concerned about how to access PrEP than its side effects, suggesting that apart from basic information on PrEP to increase its awareness, campaigns should focus on demystifying the process of obtaining, paying for, and sustaining PrEP for adolescents. Highlighting real adolescents' experiences pursuing PrEP can make a social campaign more relatable, according to our results. In a different qualitative study by our team, 100 adolescents who had successfully accessed PrEP described different strategies, barriers, and facilitators they encountered as well as their experiences of using PrEP [57]. Sharing anecdotes from these youths and others could be incorporated into a future PrEP campaign to ground the messages in reality [57]. Regarding *place* and *promotion*, participants described wanting a primarily digital campaign implemented via social media. SGM adolescents are heavy users of social media [58-60], where they often seek and consume health content, and consequently, a digital-first campaign could reduce adolescents' embarrassment and stigma (which was relatively high in this sample) by reaching them where they already are. Moreover, additional suggestions indicated that participants wanted a campaign that acted as a "one-stop shop," short of visiting a physical clinic. In other words, adolescents' responses suggested that they wanted a campaign that not only raised awareness and offered education, but also included resources that could facilitate their ability to ask questions relatively anonymously and on demand (eg, a text line) as well as connections to PrEP care (eg, telehealth PrEP consultations). Nonetheless, adolescents also identified potential in-person spaces that could promote a PrEP campaign or connect adolescents to PrEP services, such as gender-sexuality alliances, offices of family medicine physicians or pediatricians, or public libraries. However, given rapidly increasing legislation proposing to restrict sexual, reproductive, gender-affirming, and LGBTQ+-inclusive care across the United States [61,62], one must consider the local social and cultural climate and policies to determine whether it is feasible, acceptable, and safe to implement campaign materials promoting adolescent sexual health services.

One important consideration for promotion is that at the time this study was conducted, self-service social media advertising platforms (eg, Meta, which includes Facebook and Instagram) permitted users to target advertisements to minors based on their demographics, such as gender, presumed sexual orientation identity, and race or ethnicity. However, since early 2023, in the United States, significant restrictions have been placed on social media advertising to minors that would preclude the ability to do market segmentation, in other words, focus educational campaigns on young populations who need it most. Thus, public health professionals wanting to reach minors via digital spaces may need to share their campaigns more broadly with adolescents and hope that LGBTQ+-inclusive messaging and imaging will reach their intended audiences. Alternatively, they will need to identify different means of sharing health campaign materials that do not solely rely on social media advertising, which is also increasingly more expensive. These means could include reaching youth-serving organizations with a digital presence instead of attempting to reach individual adolescents themselves or returning to a focus on in-person events where adolescents are present.

Regarding how content is displayed, participants advised against including sexualized images that are common in adult-focused sexual health campaigns. Instead, participants suggested bright, youthful imagery with concise and affirming messaging delivered by trusted messengers (eg, supportive parental figures, adolescents, and near peers). Results also indicated the need to strike a balance between purposefully inclusive, "PrEP is for all" campaign messaging and information intended to reach populations in demographic groups at higher HIV transmission risk or those with unique medical or health care access considerations, such as transgender, nonbinary, and rural or suburban youth. This preference stands in contrast to PrEP awareness campaigns that tend to adopt a targeted approach to specific populations at greatest risk of HIV transmission. Consistent with literature showing that adolescents value diversity in media [63], participants showed a strong desire for a PrEP campaign that emphasizes diversity across multiple dimensions. However, without some market segmentation that considers local context and needs regarding PrEP, a campaign can run the risk of seeming too generic and impersonal and, consequently, may not reach those who need the resources the most. Moreover, participants repeatedly expressed a distaste for marketing materials that appeared too "corporate" and polished, which suggested that community members who were young, LGBTQ+, and Black, Indigenous, and people of color were not as involved in campaign design. This speaks to the need for adolescent HIV prevention campaign designers to involve community members who are the intended recipients throughout the process, which can also increase trust and buy-in when the campaign is implemented.

Finally, to better understand what participants envisioned adolescent-centered PrEP campaign materials might look like, we included a participatory design exercise at the end of the focus group. Such approaches have been used previously in adolescent sexual health intervention design and have been suggested to improve both youth well-being and youth-centered programming [64,65]. This exercise enabled us to begin to understand how participants' text responses in the focus group data manifested in their design choices and priorities. Ultimately, these youth-generated materials can be used to inspire the direction of PrEP campaign components, ensuring that design choices are not guided solely by researchers' interpretations of focus group data but by adolescents' own visions as well. In fact, these findings, in part, formed the basis of our own adolescent-centered PrEP social marketing campaign, PrEP4Teens [66], launched in late 2023, the development of which will be described in a future publication.

Limitations

While this study has valuable insights, it is important to acknowledge its limitations. The participants were predominantly adolescents who had heard of PrEP before, with a larger proportion of transgender and nonbinary teenagers as well as cisgender sexual minority girls, compared to what is typically seen in HIV prevention research focusing on those at the highest behavioral risk for HIV acquisition (eg, sexually active sexual minority male individuals and transfeminine individuals who engage in unprotected anal sex). It is possible that adolescents who have not heard of PrEP before might want

XSL•FO
RenderX

different information or have different concerns than those who are already somewhat familiar with PrEP. Moreover, those in groups at the highest risk for acquiring HIV may have different perspectives on the campaigns. However, arguably, as nationwide clinical guidelines suggest that all sexually active adolescents and adults should be informed about PrEP regardless of sexuality or gender [67], this sample composition can also be seen as a strength. As this study took place in the Chicago area, which is a large urban area with a historically progressive social and political climate, findings may not generalize to other environments. Furthermore, as this study was conducted in February to May 2021, when most in-person social behavioral research studies had not yet resumed during the COVID-19 pandemic, it is possible that our use of asynchronous web-based focus groups may not have been able to capture certain nuances that live, in-person focus groups would have (eg, body language and tone of voice).

Conclusions

Developmentally tailored, SGM adolescent–inclusive information about PrEP and where to access PrEP services is critical to contribute to US efforts to end the domestic HIV epidemic. Social marketing campaigns that directly reach adolescents have great potential to address these needs quickly and equip them with the tools to make informed decisions about their sexual health. Public health professionals can consider developing their own adolescent-centered PrEP campaigns guided by our findings and make their campaigns responsive to local needs in partnership with local adolescents and adult stakeholders who support or parent adolescents. Public health dissemination and implementation researchers should consider ways to measure and evaluate the reach and impact of PrEP campaigns on adolescents' PrEP awareness, interest, and uptake—perhaps, in part, by adding questions about social marketing campaign exposure and PrEP uptake to large public health surveys of adolescents (eg, Youth Risk Behavior Surveillance Survey by Centers for Disease Control and Prevention [68])—and explore how such patterns may differ regionally and by social-political climate. Having said that, as parents, health care providers, and other supportive adults are also critical gatekeepers of adolescents' access to health care and health information, additional research should examine and design educational PrEP campaigns for these groups.

Acknowledgments

This project was made possible with support from an administrative supplement to the Third Coast Center for AIDS Research, a center funded by the National Institutes of Health (P30AI117943).

The authors would like to thank the participants in this study for sharing their time and for the valuable insights they provided. The authors would also like to thank Dr Sybil Hosek, Hadeis Safi, Jennifer Leininger, and Elijah McKinnon for their expert guidance.

Data Availability

The datasets generated during and analyzed during this study are not publicly available as they contain identifiable information about a socially marginalized group of minor adolescents, but deidentified data are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Understanding the current HIV epidemic. AIDSVu. URL: https://aidsvu.org/local-data/united-states/ [accessed 2024-01-14]
2. HIV Surveillance Report, 2020; vol. 33. Centers for Disease Control and Prevention. 2022 May. URL: https://stacks.cdc.gov/view/cdc/121127/cdc_121127_DS1.pdf [accessed 2024-12-25]
3. HIV Surveillance Update Reports. Illinois Department of Public Health. URL: https://dph.illinois.gov/topics-services/diseases-and-conditions/hiv-aids/hiv-surveillance/update-reports.html [accessed 2025-1-10]
4. Preventing HIV with PrEP. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/hiv/prevention/prep.html [accessed 2024-12-23]
5. Fonner VA, Dalglish SL, Kennedy CE, Baggaley R, O'Reilly KR, Koechlin FM, et al. Effectiveness and safety of oral HIV preexposure prophylaxis for all populations. AIDS 2016 Jul 31;30(12):1973-1983 [FREE Full text] [doi: 10.1097/QAD.0000000000001145] [Medline: 27149090]
6. McCormack S, Dunn DT, Desai M, Dolling DI, Gafos M, Gilson R, et al. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. Lancet 2016 Jan 02;387(10013):53-60 [FREE Full text] [doi: 10.1016/S0140-6736(15)00056-2] [Medline: 26364263]
7. U.S. Food and Drug Administration approves expanded indication for Truvada® (emtricitabine and tenofovir disoproxil fumarate) for reducing the risk of acquiring HIV-1 in adolescents. Gilead Sciences, Inc. 2018. URL: https://www.gilead.com/news/news-details/2018/us-food-and-drug-administration-approves-expanded-indication-for-truvada-emtricitabine-and-tenofovir-disoproxil-fumarate-for-reducing-the-risk-of-acquiring-hiv-1-in-adolescents [accessed 2024-12-23]

XSL•FO
RenderX

8. Moskowitz DA, Moran KO, Matson M, Alvarado-Avila A, Mustanski B. The PrEP cascade in a national cohort of adolescent men who have sex with men. J Acquir Immune Defic Syndr 2021 Apr 15;86(5):536-543. [doi: 10.1097/QAI.0000000000002613] [Medline: 33399311]
9. Dunville R, Harper C, Johns MM, Heim Viox M, Avripas S, Fordyce E, et al. Awareness and willingness to use biomedical prevention strategies for HIV among sexual and gender minority youth: results from a national survey. J Adolesc Health 2021 Jan;68(1):199-206 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.05.032] [Medline: 32693984]
10. Kimball AA, Zhu W, Leonard J, Wei W, Ravichandran I, Tanner MR, et al. HIV preexposure prophylaxis provision among adolescents: 2018 to 2021. Pediatrics 2023 Nov 01;152(5):2018 [FREE Full text] [doi: 10.1542/peds.2023-062599] [Medline: 37899721]
11. Macapagal K, Nery-Hurwit M, Matson M, Crosby S, Greene GJ. Perspectives on and preferences for on-demand and long-acting PrEP among sexual and gender minority adolescents assigned male at birth. Sex Res Social Policy 2021 Mar;18(1):39-53 [FREE Full text] [doi: 10.1007/s13178-020-00441-1] [Medline: 33456624]
12. Matson M, Nery-Hurwit M, Crosby S, Greene GJ, Macapagal K. Awareness and knowledge of pre-exposure prophylaxis among sexual and gender minority adolescents assigned male at birth in the United States. Arch Sex Behav 2021 May;50(4):1701-1711 [FREE Full text] [doi: 10.1007/s10508-020-01904-7] [Medline: 33982212]
13. Moskowitz DA, Macapagal K, Mongrella M, Pérez-Cardona L, Newcomb ME, Mustanski B. What if my dad finds out!?: assessing adolescent men who have sex with men's perceptions about parents as barriers to PrEP uptake. AIDS Behav 2020 Sep;24(9):2703-2719 [FREE Full text] [doi: 10.1007/s10461-020-02827-z] [Medline: 32157491]
14. Owens C, Moran K, Mongrella M, Moskowitz DA, Mustanski B, Macapagal K. "It's very inconvenient for me": a mixed-method study assessing barriers and facilitators of adolescent sexual minority males attending PrEP follow-up appointments. AIDS Behav 2022 Jan;26(1):21-34 [FREE Full text] [doi: 10.1007/s10461-021-03313-w] [Medline: 34081237]
15. Bauermeister JA, Meanley S, Pingel E, Soler JH, Harper GW. PrEP awareness and perceived barriers among single young men who have sex with men. Curr HIV Res 2013 Oct;11(7):520-527 [FREE Full text] [doi: 10.2174/1570162x12666140129100411] [Medline: 24476355]
16. Holloway IW, Dougherty R, Gildner J, Beougher SC, Pulsipher C, Montoya JA, et al. Brief report: PrEP uptake, adherence, and discontinuation among California YMSM using geosocial networking applications. J Acquir Immune Defic Syndr 2017 Jan 01;74(1):15-20 [FREE Full text] [doi: 10.1097/QAI.0000000000001164] [Medline: 27552158]
17. Khanna AS, Michaels S, Skaathun B, Morgan E, Green K, Young L, et al. Pre-exposure prophylaxis (PrEP) awareness and use in a population-based sample of young Black men who have sex with men. JAMA Intern Med 2016 Jan 01;176(1):136-138. [doi: 10.1001/jamainternmed.2015.6536] [Medline: 26571368]
18. Morgan E, Ryan DT, Newcomb ME, Mustanski B. High rate of discontinuation may diminish PrEP coverage among young men who have sex with men. AIDS Behav 2018 Nov;22(11):3645-3648 [FREE Full text] [doi: 10.1007/s10461-018-2125-2] [Medline: 29728950]
19. Strauss BB, Greene GJ, Phillips G, Bhatia R, Madkins K, Parsons JT, et al. Exploring patterns of awareness and use of HIV pre-exposure prophylaxis among young men who have sex with men. AIDS Behav 2017 May;21(5):1288-1298 [FREE Full text] [doi: 10.1007/s10461-016-1480-0] [Medline: 27401537]
20. Thoma BC, Huebner DM. Brief report: HIV pre-exposure prophylaxis engagement among adolescent men who have sex with men: the role of parent-adolescent communication about sex. J Acquir Immune Defic Syndr 2018 Dec 01;79(4):453-457 [FREE Full text] [doi: 10.1097/QAI.0000000000001837] [Medline: 30371531]
21. Macapagal K, Coventry R, Arbeit MR, Fisher CB, Mustanski B. "I won't out myself just to do a survey": sexual and gender minority adolescents' perspectives on the risks and benefits of sex research. Arch Sex Behav 2017 Jul 28;46(5):1393-1409 [FREE Full text] [doi: 10.1007/s10508-016-0784-5] [Medline: 27469352]
22. Ma J, Chase GE, Black A, Klaphake J, Garcia-Myers K, Baker JV, et al. Attitudes toward and beliefs in the effectiveness of biomedical HIV prevention strategies among emerging and young adult sexual minority men. Int J Behav Med 2023 Dec 19:10.1007/s12529-023-10244-4 [FREE Full text] [doi: 10.1007/s12529-023-10244-4] [Medline: 38114707]
23. Moore K, Dell S, Oliva M, Morgan A, Rothman R, Hsieh YH, et al. Does parental insurance impact willingness to take prep in adolescents and young adults? J Adolesc Health 2018 Feb;62(2):S23-S24 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.11.047]
24. Ma J, Owens C, Valadez-Tapia S, Brooks JJ, Pickett J, Walter N, et al. Adult stakeholders' perspectives on the content, design, and dissemination of sexual and gender minority adolescent-centered PrEP campaigns. Sex Res Soc Policy 2023 May 31;21(1):339-351 [FREE Full text] [doi: 10.1007/s13178-023-00826-y]
25. Macapagal K, Kraus A, Korpak AK, Jozsa K, Moskowitz DA. PrEP awareness, uptake, barriers, and correlates among adolescents assigned male at birth who have sex with males in the U.S. Arch Sex Behav 2020 Jan 10;49(1):113-124 [FREE Full text] [doi: 10.1007/s10508-019-1429-2] [Medline: 31602584]
26. Macapagal K, Ma J, Owens C, Valadez-Tapia S, Kraus A, Walter N, et al. 64. #PrEP4Teens: LGBTQ+ adolescent perspectives on content and implementation of a teen-centered PrEP social marketing campaign. J Adolesc Health 2022 Apr;70(4):S34 [FREE Full text] [doi: 10.1016/j.jadohealth.2022.01.024]
27. Grier S, Bryant CA. Social marketing in public health. Annu Rev Public Health 2005 Apr 21;26(1):319-339. [doi: 10.1146/annurev.publhealth.26.021304.144610] [Medline: 15760292]

XSL•FO
RenderX

28. Chau JY, McGill B, Thomas MM, Carroll TE, Bellew W, Bauman A, et al. Is this health campaign really social marketing? A checklist to help you decide. Health Promot J Austr 2018 Apr;29(1):79-83 [FREE Full text] [doi: 10.1002/hpja.13] [Medline: 29700937]
29. McDermott RJ. Social marketing: a tool for health education. Am J Health Behav 2000 Jan 01;24(1):6-10. [doi: 10.5993/AJHB.24.1.2]
30. Minor consent and confidentiality laws for HIV treatment and PrEP: Alabama. National Alliance of State and Territorial AIDS Directors. 2022. URL: https://nastad.org/sites/default/files/2022-02/PDF_Minor_Consent_and_Confidentiality_Laws.pdf [accessed 2024-12-23]
31. Evans WD. How social marketing works in health care. BMJ 2006 May 20;332(7551):1207-1210 [FREE Full text] [doi: 10.1136/bmj.332.7551.1207-a] [Medline: 16710002]
32. Keene LC, Dehlin JM, Pickett J, Berringer KR, Little I, Tsang A, et al. #PrEP4Love: success and stigma following release of the first sex-positive PrEP public health campaign. Cult Health Sex 2021 Mar;23(3):397-413. [doi: 10.1080/13691058.2020.1715482] [Medline: 32212962]
33. Dehlin JM, Stillwagon R, Pickett J, Keene L, Schneider JA. #PrEP4Love: an evaluation of a sex-positive HIV prevention campaign. JMIR Public Health Surveill 2019 Jun 17;5(2):e12822 [FREE Full text] [doi: 10.2196/12822] [Medline: 31210141]
34. #PrEP4Love. AIDSeS. URL: https://www.aides.org/campagne/prep4love [accessed 2024-12-27]
35. Friedman AL, Brookmeyer KA, Kachur RE, Ford J, Hogben M, Habel MA, et al. An assessment of the GYT: Get Yourself Tested campaign: an integrated approach to sexually transmitted disease prevention communication. Sex Transm Dis 2014 Mar;41(3):151-157 [FREE Full text] [doi: 10.1097/OLQ.0000000000000100] [Medline: 24521718]
36. National youth HIV/AIDS awareness day. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/healthyyouth/youth_hiv/resources.htm#:~:text=National%20Youth%20HIV%20%26%20AIDS%20Awareness%20Day%20(NYHAAD)%20is%20observed,health%20and%20safety%20of%20youth [accessed 2022-10-14]
37. Forthofer MS, Bryant CA. Using audience-segmentation techniques to tailor health behavior change strategies. Am J Health Behav 2000;24(1):36-43 [FREE Full text] [doi: 10.5993/AJHB.24.1.6]
38. Icard L, Bourjolly JN, Siddiqui N. Designing social marketing strategies to increase African Americans' access to health promotion programs. Health Soc Work 2003 Aug;28(3):214-223 [FREE Full text] [doi: 10.1093/hsw/28.3.214] [Medline: 12971285]
39. Walter N, Demetriades SZ, Murphy ST. Just a spoonful of sugar helps the messages go down: using stories and vicarious self-affirmation to reduce e-cigarette use. Health Commun 2019 Mar 13;34(3):352-360 [FREE Full text] [doi: 10.1080/10410236.2017.1407275] [Medline: 29236554]
40. Nabi RL, Walter N, Oshidary N, Endacott CG, Love-Nichols J, Lew ZJ, et al. Can emotions capture the elusive gain-loss framing effect? A meta-analysis. Commun Res 2019 Jul 13;47(8):1107-1130 [FREE Full text] [doi: 10.1177/0093650219861256]
41. Kwan BM, Brownson RC, Glasgow RE, Morrato EH, Luke DA. Designing for dissemination and sustainability to promote equitable impacts on health. Annu Rev Public Health 2022 Apr 05;43:331-353 [FREE Full text] [doi: 10.1146/annurev-publhealth-052220-112457] [Medline: 34982585]
42. Biroscak BJ, Schneider T, Martinez Tyson D, Aguado Loi CX, Bryant CA. Applying tools from human-centered design to social marketing planning. Soc Mark Q 2018 Mar 05;24(2):63-73 [FREE Full text] [doi: 10.1177/1524500418761624]
43. Riley WT, Glasgow RE, Etheredge L, Abernethy AP. Rapid, responsive, relevant (R3) research: a call for a rapid learning health research enterprise. Clin Transl Med 2013 May 10;2(1):10 [FREE Full text] [doi: 10.1186/2001-1326-2-10] [Medline: 23663660]
44. Stewart K, Williams M. Researching online populations: the use of online focus groups for social research. Qual Res 2005 Nov 01;5(4):395-416 [FREE Full text] [doi: 10.1177/1468794105056916]
45. Synnot A, Hill S, Summers M, Taylor M. Comparing face-to-face and online qualitative research with people with multiple sclerosis. Qual Health Res 2014 Mar;24(3):431-438 [FREE Full text] [doi: 10.1177/1049732314523840] [Medline: 24572010]
46. Woodyatt CR, Finneran CA, Stephenson R. In-person versus online focus group discussions: a comparative analysis of data quality. Qual Health Res 2016 May;26(6):741-749 [FREE Full text] [doi: 10.1177/1049732316631510] [Medline: 26935719]
47. Napper LE, Fisher DG, Reynolds GL. Development of the perceived risk of HIV scale. AIDS Behav 2012 May;16(4):1075-1083 [FREE Full text] [doi: 10.1007/s10461-011-0003-2] [Medline: 21785873]
48. Walsh JL. Applying the information-motivation-behavioral skills model to understand PrEP intentions and use among men who have sex with men. AIDS Behav 2019 Jul;23(7):1904-1916 [FREE Full text] [doi: 10.1007/s10461-018-2371-3] [Medline: 30554396]
49. Hamilton AB, Finley EP. Qualitative methods in implementation research: an introduction. Psychiatry Res 2019 Oct;280:112516 [FREE Full text] [doi: 10.1016/j.psychres.2019.112516] [Medline: 31437661]
50. Watkins DC. Rapid and rigorous qualitative data analysis: the “RADaR” technique for applied research. Int J Qual Methods 2017 Jun 08;16(1):1-9 [FREE Full text] [doi: 10.1177/1609406917712131]

XSL•FO
RenderX

51. Taylor B, Henshall C, Kenyon S, Litchfield I, Greenfield S. Can rapid approaches to qualitative analysis deliver timely, valid findings to clinical leaders? A mixed methods study comparing rapid and thematic analysis. BMJ Open 2018 Oct 08;8(10):e019993 [FREE Full text] [doi: 10.1136/bmjopen-2017-019993] [Medline: 30297341]
52. Marketing campaign case studies: Trojan man campaign. Marketing Campaign Case Studies. 2008 May 13. URL: https://marketing-case-studies.blogspot.com/2008/05/trojan-man-campaign.html [accessed 2023-07-01]
53. Magnuson D, Hawkins T, Mera R. Adolescent use of Truvada (FTC/TDF) for HIV pre-exposure prophylaxis (PrEP) in the United States: (2012-2017). In: Proceedings of the 22nd International AIDS Conference. 2018 Presented at: AIDS 2018; July 23-27, 2018; Amsterdam, The Netherlands.
54. Mustanski B, Saber R, Macapagal K, Matson M, Laber E, Rodriguez-Díaz C, et al. Effectiveness of the SMART Sex Ed program among 13–18 year old English and Spanish speaking adolescent men who have sex with men. AIDS and Behavior 2023 Feb 27:733-44.
55. Let's talk about it: preventing HIV with PrEP. US Preventive Services Task Force. URL: https://www.uspreventiveservicestaskforce.org/uspstf/sites/default/files/inline-files/hiv-prep-prevention-discussion-guide.pdf [accessed 2024-12-23]
56. Underhill K, Nelson KM. Consent laws for minors regarding sexually transmitted infection and HIV services-reply. JAMA 2022 Dec 20;328(23):2364 [FREE Full text] [doi: 10.1001/jama.2022.19040] [Medline: 36538312]
57. Gordon J, Wongsomboon V, Alvarado Avila A, Lorenzo J, Mustanski B, Macapagal K. Experiences of successful PrEP uptake among adolescent sexual minority men in the United States: a qualitative exploration. AIDS Educ Prev 2024 Feb;36(1):1-15 [FREE Full text] [doi: 10.1521/aeap.2024.36.1.1] [Medline: 38349349]
58. Phillips G2, Magnus M, Kuo I, Rawls A, Peterson J, Jia Y, et al. Use of geosocial networking (GSN) mobile phone applications to find men for sex by men who have sex with men (MSM) in Washington, DC. AIDS Behav 2014 Sep 29;18(9):1630-1637. [doi: 10.1007/s10461-014-0760-9] [Medline: 24682866]
59. Vogels EA, Gelles-Watnick R, Massarat N. Teens, social media and technology 2022. Pew Research Center. 2022 Aug 10. URL: https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022/ [accessed 2023-06-14]
60. McConnell E, Néray B, Hogan B, Korpak A, Clifford A, Birkett M. "Everybody puts their whole life on Facebook": identity management and the online social networks of LGBTQ Youth. Int J Environ Res Public Health 2018 May 26;15(6):1078 [FREE Full text] [doi: 10.3390/ijerph15061078] [Medline: 29861439]
61. Mapping attacks on LGBTQ rights in U.S. state legislatures in 2024. American Civil Liberties Union. URL: https://www.aclu.org/legislative-attacks-on-lgbtq-rights-2024?state=NV [accessed 2024-03-01]
62. State legislation tracker: major developments in sexual and reproductive health. Guttmacher Institute. URL: https://www.guttmacher.org/state-legislation-tracker [accessed 2024-03-01]
63. Ellithorpe ME, Bleakley A. Wanting to see people like me? Racial and gender diversity in popular adolescent television. J Youth Adolesc 2016 Jul 12;45(7):1426-1437 [FREE Full text] [doi: 10.1007/s10964-016-0415-4] [Medline: 26759131]
64. Ozer EJ. Youth-led participatory action research: developmental and equity perspectives. Adv Child Dev Behav 2016;50:189-207 [FREE Full text] [doi: 10.1016/bs.acdb.2015.11.006] [Medline: 26956074]
65. Fakoya I, Cole C, Larkin C, Punton M, Brown E, Ballonoff Suleiman A. Enhancing human-centered design with youth-led participatory action research approaches for adolescent sexual and reproductive health programming. Health Promot Pract 2022 Jan;23(1):25-31 [FREE Full text] [doi: 10.1177/15248399211003544] [Medline: 33858267]
66. PrEP4Teens homepage. PrEP4Teens. URL: https://prep4teens.com/about [accessed 2024-04-04]
67. Preexposure prophylaxis for the prevention of HIV infection in the United States – 2021 update. Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/hiv/pdf/risk/prep/cdc-hiv-prep-guidelines-2021.pdf [accessed 2022-12-01]
68. Youth risk behavior surveillance system (YRBSS). Centers for Disease Control and Prevention. URL: https://www.cdc.gov/yrbs/about/index.html [accessed 2024-06-28]

Abbreviations

LGBTQ+: lesbian, gay, bisexual, transgender, and queer
PrEP: pre-exposure prophylaxis
SGM: sexual and gender minority
STI: sexually transmitted infection

XSL•FO
RenderX

Edited by A Mavragani; submitted 16.05.24; peer-reviewed by PJ Coulaud, Y Cui; comments to author 21.08.24; revised version received 03.11.24; accepted 26.11.24; published 17.01.25.

Please cite as:
Macapagal K, Zapata JP, Ma J, Gordon JD, Owens C, Valadez-Tapia S, Cummings P, Walter N, Pickett J
Sexual and Gender Minority Adolescents' Preferences for HIV Pre-Exposure Prophylaxis Social Marketing Campaigns: Qualitative Preimplementation Study
JMIR Form Res 2025;9:e60398
URL: https://formative.jmir.org/2025/1/e60398
doi:10.2196/60398
PMID:

©Kathryn Macapagal, Juan Pablo Zapata, Junye Ma, Jacob D Gordon, Christopher Owens, Silvia Valadez-Tapia, Peter Cummings, Nathan Walter, Jim Pickett. Originally published in JMIR Formative Research (https://formative.jmir.org), 17.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Mobile Health Intervention to Support Collaborative Decision-Making in Mental Health Care: Development and Usability

Kristin Lie Romm[1,2], MD, PhD; Mari Skoge[1], ClinPsy; Elizabeth Ann Barrett[1], ClinPsy, PhD; Lars-Christian Berentzen[3], MD; Dagfinn Bergsager[4], BSc; Pål Fugelli[4], MEd, PhD; Thomas Bjella[2,5], MSc; Erlend Strand Gardsjord[3], MD, PhD; Kristine Kling[1], MA; Sindre Hembre Kruse[1], BSc; Kari Jorunn Kværner[6,7], MD, PhD; Ingrid Melle[2,8], MD, PhD; Erlend Mork[1], ClinPsy, PhD; Henrik Myhre Ihler[9], MD, PhD; Eline Borger Rognli[10], ClinPsy, PhD; Carmen Simonsen[1], ClinPsy, PhD; Tor Gunnar Værnes[1], ClinPsy, PhD; Sofie Ragnhild Aminoff[1], ClinPsy, PhD

[1]Early Intervention in Psychosis Advisory Unit for South-East Norway, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway

[2]Institute of Clinical Medicine, University of Oslo, Oslo, Norway

[3]Section for Early Intervention in Psychosis, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway

[4]University Center for Information Technology, University of Oslo, Oslo, Norway

[5]Department of Research and Innovation, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway

[6]Centre for Connected Care (C3), Oslo University Hospital, Oslo, Norway

[7]BI Norwegian Business School, Oslo, Norway

[8]Section for Clinical Psychosis Research, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway

[9]Nydalen DPS, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway

[10]Section for Clinical Addiction Research, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway

Corresponding Author:
Kristin Lie Romm, MD, PhD
Early Intervention in Psychosis Advisory Unit for South-East Norway
Division of Mental Health and Addiction
Oslo University Hospital
Sognsvannsveien 21
Oslo, 0372
Norway
Phone: 47 91 50 27 70
Email: k.l.romm@medisin.uio.no

Abstract

Background: Shared decision-making between clinicians and service users is crucial in mental health care. One significant barrier to achieving this goal is the lack of user-centered services. Integrating digital tools into mental health services holds promise for addressing some of these challenges. However, the implementation of digital tools, such as mobile apps, remains limited, and attrition rates for mental health apps are typically high. Design thinking can support the development of tools tailored to the needs of service users and clinicians.

Objective: This study aims to develop and beta test a digital tool designed for individuals with severe mental disorders or substance use disorders to facilitate shared decision-making on treatment goals and strategies within mental health services.

Methods: We used a user-centered design approach to develop iTandem, an app facilitating collaborative treatment between service users and clinicians. Through qualitative interviews and workshops, we engaged 6 service users with severe mental disorders or substance use disorders, 6 clinicians, and 1 relative to identify and design relevant app modules. A beta test of iTandem was conducted to refine the app and plan for a pilot trial in a clinical setting. After 6 weeks of app use, 5 clinicians and 4 service users were interviewed to provide feedback on the concept, implementation, and technical issues. Safety and ethical considerations were thoroughly discussed and addressed.

Results: To avoid overload for the service users, we applied a pragmatic take on module content and size. Thus, iTandem includes the following 8 modules, primarily based on the needs of service users and clinicians: Sleep (sleep diary), Medication

XSL•FO
RenderX

(intake and side effects), Recovery (measures, including well-being and personal recovery, and exercises, including good things and personal strengths), Mood (mood diary and report of daily feelings), Psychosis (level of positive symptoms and their consequences and level of negative symptoms), Activity (goal setting and progress), Substance use (weekly use, potential triggers or strategies used to abstain), and Feedback on therapy (of individual sessions and overall rating of the past week). For the beta testing, service users and clinicians collaborated in choosing 2-3 modules in iTandem to work with during treatment sessions. The testing showed that the app was well received by service users, and that facilitation for implementation is crucial.

Conclusions: iTandem and similar apps have the potential to enhance treatment outcomes by facilitating shared decision-making and tailoring treatment to the needs of service users. However, successful implementation requires thorough testing, iterative development, and evaluations of both utility and treatment effects. There is a critical need to focus on how technology integrates into clinical settings—from development to implementation—and to conduct further research on early health technology assessments to guide these processes.

(JMIR Form Res 2025;9:e57614) doi:10.2196/57614

KEYWORDS

eHealth; shared decision-making; user involvement; user-centered design; mental disorder; mobile technology; illness course; recovery; mobile apps; mHealth

Introduction

> *...in what way has it (iTandem) been helpful?*
>
> *Like right now, I just went into the "substance use" module to write that I managed to say 'no' to buying heroin. Or if I see needles and stuff, I choose to try and get away – move away from where there are needles and that kind of stuff available.* [Beta tester of iTandem]

The ability to communicate effectively with service users is essential in health care. Therapeutic communication can lead to several benefits, including a more accurate description of symptoms, improved detection of emotional states, more appropriate therapeutic measures, better identification of the service user's needs and understanding of their disorder and treatment, and an increased likelihood of shared decision-making between clinicians and service users [1-3]. Engaging individuals with severe mental disorders, substance use disorders, or both in their recovery process can be challenging. Various factors contribute to this difficulty, including a lack of trust in health care services [4], poor therapeutic alliance [5], reduced insight into their illness [6], diminished cognitive functioning [7], and issues with motivation and apathy [8,9].

Priebe and colleagues [10] outlined 5 guiding principles for effective communication in mental health services: focusing on the service users' concerns, demonstrating positive regard and personal respect, appropriately involving service users in decision-making, showing genuineness with a personal touch, and using a psychological treatment model. Emphasizing service users' concerns aligns closely with the concept of "patient-centeredness" in health care. A recent review [11] defined centeredness as encompassing attributes such as being treated as unique, being heard, and sharing responsibility. However, Feldthusen and colleagues [11] emphasize that centeredness is a complex concept, with its essence rooted in the interactions and communication between clinicians and service users. For individuals with severe mental illness, achieving a sense of positive regard and personal respect can be challenging. Therefore, adopting a nonjudgmental approach has been shown to be particularly valuable [12]. Shared decision-making is a widely recognized concept in medicine, but it remains challenging to implement effectively in mental health care. For individuals with severe mental disorders, barriers include clinicians' paternalistic approaches to care [13] and service users' internalized stigma [14]. Increasing service user involvement in activities such as goal setting, addressing values and preferences, and working with motivational factors is essential [15]. Additionally, exploring shared risk-taking can further enhance communication [16]. Genuineness and a personal touch involve openness and sharing but also reflect the therapist's behavior and characteristics, such as warmth and flexibility [17]. Finally, adopting a coherent client-centered therapy approach, such as solution-focused therapy, problem-solving therapy, or cognitive behavioral therapy, is often preferable [10].

Digitalization offers a potential solution to some of these challenges by making treatment more focused and efficient. There has been growing interest in the use of patient-reported outcomes and patient-reported outcome measures (PROMs) in health care [18], which are frequently digitalized. The inclusion of validated assessment tools in digital formats provides clear advantages. However, a significant barrier for health professionals is the time and expertise required to interpret and integrate the extensive findings into their workflow [19]. Additionally, Nguyen and colleagues [19] identified perceived irrelevance and lack of value for the service user as barriers to completing PROMs. As a result, both service users and clinicians emphasize the need for tools that are functional, straightforward, and concise, with as few questions as possible.

There has been a surge in smartphone apps designed to support individuals with mental health challenges, readily available on platforms such as the App Store (Apple Inc.) or Google Play (Alphabet Inc.). Most of these apps are standalone offerings in the free market, focusing on symptoms such as depression, anxiety, or stress. However, they are rarely developed to meet the safety standards typically required in health care settings, making integration into clinical practice difficult. Additionally, while many apps appear acceptable and feasible initially, attrition rates over time are often high [20]. Individuals with mental disorders may face unique challenges that demand

XSL•FO
RenderX

flexible solutions, as therapeutic targets can shift throughout the course of illness. Thus, assessing attrition alone is often insufficient without a deeper understanding of the context. Proposed solutions include placing greater emphasis on external motivating factors and user experience (UX) design, as well as integrating digital mental health solutions into existing care pathways [21,22].

KISS—"keep it simple, stupid"—is a design principle [23] frequently used in design processes and should be applied to service design as well [24]. This principle emphasizes that most systems function optimally when kept simple rather than overly complex [25]. Simpler innovations are more likely to be successfully deployed. Aligning with this concept, human-centered design has gained significant attention in recent years. It involves creative approaches that enable end users to influence the design of technology, tailoring it to better meet their needs [26]. One such approach is participatory design. A fundamental principle in this approach is that people affected by a decision should have the opportunity to influence it, with the goal of designing alternatives to improve quality of life [27]. Despite extensive research on the importance of applying design principles to technology to make it effective, only a limited number of studies in mental health care report using co-design and expert designers in their developmental process [26,28].

A previous service design project focused on the establishment of a low-threshold service for people with psychosis in health care services [29]. Based on insights from service users, caregivers, and clinicians, easy access to a first assessment of psychosis was established. During this project, it became evident that individuals with lived experience were interested in new ways of collaborating with health care services. They sought tools that could enhance the effects of treatment, make better use of the time between sessions, and help them stay focused on their goals. As one service user remarked, "I live my life between the appointments." This highlighted the need to explore how technology could strengthen the focus on treatment between appointments.

The main aim of the iTandem development was to investigate the needs described by service users and clinicians and to integrate them into an app designed to enhance shared decision-making for people with psychotic, bipolar, or substance use disorders in treatment. We applied design thinking principles to incorporate user needs and feedback throughout the process [30,31]. Our long-term goal is to explore the barriers and enablers for integrating technology into shared decision-making in mental health care and to examine how the integration of digital tools affects the mode of service delivery. In this paper, we will present (1) the design process behind the development of iTandem and (2) the results of beta testing. Finally, we will discuss the importance of future research on the digital enhancement of service delivery.

Methods

Project Setting

iTandem was developed at the Early Intervention in Psychosis Advisory Unit for South East Norway (TIPS SE) at Oslo University Hospital (OUH). TIPS SE works to support mental health services in detecting and treating patients with early psychosis in accordance with national guidelines, with a particular emphasis on how these services are organized.

Design Process

Overview

External funding from the South-Eastern Norway Regional Health Authority enabled the hiring of experts in service design and interaction design to lead the process. The process progressed through 3 phases, involving multiple stakeholders including patients, clinicians, developers, data protection officers, and clinicians, in addition to the original project team consisting of researchers and clinicians from OUH (KLR, EAB, KK, EM, EBR, CS, TGV, and SRA). The phases involved are discussed in the following subsections.

Preparation Phase

Qualitative interviews were conducted to inform the design process by understanding the needs and preferences regarding app content. Six patients diagnosed with psychotic or bipolar disorder, as well as 1 relative, were interviewed individually. In addition, 2 focus groups were held with 6 clinicians: 1 nurse practitioner, 1 clinical psychologist, 1 psychiatrist, and 1 medical doctor (all from an early intervention in a psychosis team); as well as 1 psychologist from a general psychiatric outpatient service and 1 from the child and adolescent services for early psychosis.

Insight Phase

A private design company managed the design process and developed a clickable app prototype to inform the developers. The design company, in collaboration with researchers, agreed upon a project plan that included several meetings with clinicians (the same clinicians who had been involved in phase 1) and 4 patients (the same as in phase 1). The designers utilized the insights collected during phase 1 and knowledge generated through workshops to focus on themes that appeared to be both significant and feasible to work with, given the project's budget and time constraints.

Developmental Phase

This phase was carried out in collaboration with the Mobile App team at the Center for Information Technology (USIT) at the University of Oslo (UiO), which developed a working app capable of transferring encrypted data to secure storage within Services for Sensitive Data (Tjenester for Sensitive Data or TSD). This was necessary to meet the safety requirements for the storage of sensitive data, as mandated by the data protection officer at OUH. A risk and vulnerability analysis was conducted at OUH. During this phase, 4 patients and 2 clinicians were invited to test the iTandem app during development to ensure it could be adjusted based on short iterations.

Project Organization

Overview of the Project Group

The project group included researchers and clinicians working with early psychosis treatment. TIPS SE also has a board composed of people with lived experience of bipolar and

XSL•FO
RenderX

psychotic disorders. These individuals were either currently in treatment or had previously been in treatment at mental health services in OUH during the development project. The roles of the different groups are presented below.

Researchers

During phase 1, researchers conducted individual qualitative interviews with 6 service users and 1 relative and led 1 focus group with 6 clinicians. The purpose was to explore the challenges participants believed digital tools could address. During phase 2, the researchers served as a bridge between the clinic and the design team, facilitating workshops, collaborating with user representatives, and ensuring ethical issues were handled appropriately. They also suggested improvements to the app content to ensure it aligned with clinical guidelines and recommendations regarding treatment. During phase 3, the researchers facilitated and oversaw the developmental process of the app at USIT.

Designers

During phase 2, designers planned and conducted workshops with clinicians and researchers. Throughout the project, they interviewed clinicians and user representatives, and user tested different options in the studio with both clinicians and user representatives in an iterative process. The process was based on design thinking, working with divergent and convergent thinking according to the Double Diamond framework [32]. During phase 3, UX designers, engaged in making products or services usable, enjoyable, and accessible, were involved at USIT to further develop the interface and add functionality based on tests from the prototyping phase.

Service Users

Service users with severe mental disorders or substance use disorders were involved throughout all phases of the project. Service users (aged 18-30 years) and 1 relative (parent) provided input on functionality and were invited to user test different prototypes. This involvement had an impact on the final app interface and functionality. Some of these service users were among those interviewed during phase 1. Additionally, we met with the board of user representatives in the Department of Substance Use and Addiction at OUH to discuss app content relevant to them. Service users not directly involved in development were invited to beta test iTandem.

Clinicians

During the process, involved clinicians highlighted important aspects essential for implementation, and shared their ideas on how iTandem could be valuable in their daily clinical work. Additionally, they user tested the app during iterative cycles and provided feedback on how they believed the app would be received by their patients.

Software

The iTandem app was developed using the React Native Open-Source UI Software Framework (Meta Platforms, Inc.) and is available for both iOS and Android platforms. It can be downloaded from the App Store or Google Play. Data are transferred via HTTPS as JSON files through a secure web survey provider (Nettskjema), hosted by USIT at the UiO. The "Activity," "List of Medication," and "Diary of Good Things" modules are stored on the user's phone for easy access, while aggregated data from all modules are encrypted and transferred to a secure server (TSD) at UiO once a module is completed. These data generate a report that can be accessed by clinicians during consultations and provide structured data available for researchers.

Ethical Considerations

The study was approved by the Data Protection Officer at OUH (approval numbers 2016/14408 and 20/18434). The regional ethical committee found that the development and testing fell outside their mandate (reference number 105851). All clinicians and service users who were interviewed signed informed consent forms. Service users who were engaged during the early phases of development, in collaboration with the service designers, also signed informed consent for their involvement in this work. All data were handled securely at TSD and by a secure research server at OUH. Service users were compensated with approximately US $23 per hour. All data used in this paper have been anonymized.

Results

The Development Process

Designing iTandem for Clinical Integration and Shared Decision-Making

We aimed to make the use of iTandem as true to a clinical setting as possible, keeping both the service user's and clinicians' experiences in mind when framing the app content. Our primary goal was to develop a tool to enhance shared decision-making. Besides supporting collaboration, we believe this is crucial for compliance. iTandem is not a standalone tool, and both its usage and the reasons for any attrition should be explored during treatment. We used phrases that closely mirror those used by clinicians. Furthermore, we chose to incorporate only a minimum of validated self-report measures, as the rationale behind the app was to serve as a communicative tool rather than a strict research assessment instrument.

The described needs were in line with current research, and based on the current literature, the rationale behind each module is outlined below (Table 1).

XSL•FO
RenderX

Table 1. Description of the modules of iTandem and their intended purposes.

Module	Frequency of notifications	Aim	Contents	Basis of measure/intervention
Sleep	• Twice daily	• Map sleep patterns and quality, link to the level of function	• Sleep diary	• Consensus sleep diary
Medication	• Once or twice daily	• Support service users in keeping track of and remembering to take medication at the right time	• Registration of medications taken and potential side effects • List of medications	• Partly developed by the project group and partly adapted from the UKU Side Effect Scale • Developed by the project group
Recovery	• Daily/weekly • Monthly	• Do positive psychology exercises • Map and focus on personal recovery and positive aspects of mental health	• Good things • Personal strengths • Psychological well-being • Emotional well-being • Cognitive well-being	• Exercises adapted from "Positive psychotherapy for psychosis" • Questionnaire about the Process of Recovery • Warrick-Edinburgh Mental Well-Being Scale • Lehman Life Satisfaction Scale
Mood	• Daily	• Map feelings and mood variability, link to sleep, medication, experiences, and other symptoms	• Mood diary • Emotions	• Adapted from the paper version of the mood diary from the Systematic Treatment Enhancement Program for Bipolar Disorder study • Developed by the project group
Psychosis	• Weekly • Monthly	• Follow symptom development and relation to difficulties in reaching goals, link to other variables, guide treatment	• Symptoms of psychosis • Negative symptoms	• Developed by the project group • Self-evaluation of negative symptoms (Self-Evaluation of Negative Symptoms Scale)
Activity	• Daily	• Increase/regulate activity level, set goals, plan activities, motivate to complete, bring personal experiences into treatment	• Registration of goals • Registration of activity	• Developed by the project group
Substance use	• Weekly	• Gain a better understanding of how substance use impacts symptoms and daily life, guide coping strategies	• Registration of substance use and coping strategies	• Developed by the project group
Feedback on treatment	• Before and after consultations	• Facilitate feedback to the clinician to strengthen patient involvement and quality of treatment, provide an overview of the treatment course	• Before consultation • After consultation	• Developed by the project group

Sleep

Sleep disturbances are common in psychotic and bipolar disorders [33]. In bipolar disorder, sleep disturbances are frequently observed during and between episodes and are associated with more severe symptoms [34]. Therefore, being able to detect and monitor disturbed sleep patterns and implement treatments that focus on improving sleep may be crucial. Both service users and clinicians considered it valuable to have a tool that collects daily data on sleep to avoid recall bias. Additionally, clinicians expressed a need for a tool to assist in diagnosing sleep disorders. The Sleep module was based on the consensus sleep diary, which enables the monitoring of subjective sleep measures and the diagnosing of common sleep disorders [35]. The diary consists of a questionnaire in the morning about how well you slept and a questionnaire in the evening where users rate how their day was. The intended purpose of the Sleep module is to provide an overview of service users' sleep patterns and, based on this, initiate relevant sleep treatment if necessary. Additionally, the module includes a link to a website [36] with information about common sleep problems in severe mental disorders and sleep hygiene advice.

Medication

Both service users and clinicians wanted a tool that could assist with reminders (notifications) for taking medication and monitoring medication side effects. Medication is a key component in the treatment of both psychotic and bipolar disorders. Antipsychotics are effective, especially during the early and acute phase of psychosis [37], and are also commonly used in people with bipolar disorder [38], along with mood stabilizers [39]. However, these medications may have side

XSL•FO
RenderX

effects that are not always properly addressed [40]. Recent studies have highlighted the importance of considering the user's experience level, autonomy processes, values and risk preferences, knowledge, and needs to accurately understand the service users' experience with medication [41]. The module was designed to enable service users to create a (revisable) list of their medications, set reminders for doses, and report side effects. Previous research has demonstrated that the language used to describe side effects can influence users' expectations and may contribute to experiencing nocebo-induced side effects [42]. Webster and colleagues [42] found this to be particularly significant for people with negative beliefs about side effects. To mitigate this, we used UX design to integrate the patient-rated UKU Side Effect Scale [43], which consists of 48 single symptom items, into the Medication module. Instead of requiring the service user to rate 48 individual items, we chose to present the scale as a visualization of the human body. Service users can click on the body part where they experience side effects and select the relevant side effect(s).

Recovery

Clinicians were primarily interested in how treatment impacted clinical recovery in terms of symptom remission and adequate functioning. However, they also emphasized the importance of considering the impact on quality of life. Service users, by contrast, highlighted the need to focus more on their positive outcomes in line with personal recovery. This is commonly defined as the subjective experience of connectedness, hope, identity, meaning in life, and empowerment [44]. Personal recovery has received increased attention in mental health research in recent years, particularly in relation to psychotic disorders [45,46]. This highlights the importance of monitoring treatment outcomes in terms of well-being, personal recovery, and clinical recovery. Positive psychology has been applied to the treatment of mental disorders and, more recently, to psychotic disorders through positive psychotherapy for psychosis [47]. The Recovery module has 2 parts: First, it monitors personal recovery using a validated assessment measure [48], focusing on good feelings and thoughts with a validated measure of mental well-being (Warrick-Edinburgh Mental Well-Being Scale) [49], and life satisfaction with The Life Satisfaction Scale [50]. Second, the module includes 2 positive psychology exercises, namely, (1) Good Things, which is a positive diary of good experiences. The diary allows for daily notes to be stored on the phone, which can be revisited by the service user. It is based on positive psychology, suggesting that gratitude interventions can positively affect well-being, happiness, and life satisfaction [51]. (2) Personal Strength Exercise, which emphasizes the need for strength-based interventions for people with severe mental disorders [52]. The service user is presented with 24 strengths visualized by icons. They are encouraged to select 4 of their personal strengths and plan activities that utilize these strengths. This approach is adapted from the Positive Psychotherapy for Psychosis group intervention developed by Slade et al [47], in accordance with the WELLFOCUS study [53].

Mood

Detecting fluctuations in mood symptoms, such as elevated and depressed moods, is crucial in the treatment and relapse prevention of bipolar disorder, as emphasized by the service users. However, mood symptoms are not limited to bipolar disorder; they also play a significant role in other mental disorders, as highlighted by all the service users and clinicians who were interviewed [54,55]. Monitoring mood and affective symptoms is also relevant for personal recovery [56]. The service users wanted the opportunity to log affective states or emotions as they emerge in daily life, with the option of linking such experiences to external triggers. Service users with bipolar disorder may benefit from using daily mood charts to monitor their symptoms [57]. This feature was incorporated into the Mood module based on a paper version adapted from the Systematic Treatment Enhancement Program for Bipolar Disorder [58]. Additionally, we developed a module to monitor daily emotions, potential triggers, and an evaluation of fluctuating or stable emotions.

Psychosis

The clinicians were interested in measures to detect how psychotic symptoms fluctuate and also the ability to detect triggers. This functionality was less important to the service users. This might seem paradoxical but aligns with the research literature on recovery [59]. However, most service users agreed that it could be relevant to log their symptoms to register the effects of medication, although they argued that the questions needed to be brief and relevant. As we defined iTandem as a tool aimed at facilitating dialogue in treatment sessions and not for research assessment, we chose to include questions about psychosis as typically phrased by clinicians in clinical practice. Thus, the Psychosis module includes a list of positive psychosis symptoms to choose from and brief follow-up questions regarding how symptoms affect daily function, triggers, etc. This may serve as a means to monitor treatment progression regarding positive psychotic symptoms. Additionally, the clinicians wanted to monitor negative symptoms. Negative symptoms are challenging to assess due to their overlapping features with, for example, depression and medication side effects [60]. However, the Self-Evaluation of Negative Symptoms Scale has been shown to be a valid self-report scale, and we chose to incorporate this into the Psychosis module [61].

Activity

Increasing goal-directed activity is important in the recovery process for severe mental disorders. However, achieving this in treatment can be challenging. The service users and clinicians wanted functionality that could be incorporated as part of treatment and motivate for change. This aligns with previous research, which states that goal planning is an important element in mental health service delivery, but we lack standardized approaches to support it [62]. Furthermore, both working alliance and goal achievement support the individual recovery process, and a strong working alliance seems to make people more capable of reaching their goals [63]. Potentially, digital tools could supplement and enhance this alliance. This module aims to support motivation for activity by allowing the service user to set long- and short-term goals. When an activity that

XSL•FO
RenderX

supports the goal has been executed, the service user selects an icon representing this goal. Each completed activity creates a streak.

Substance Use

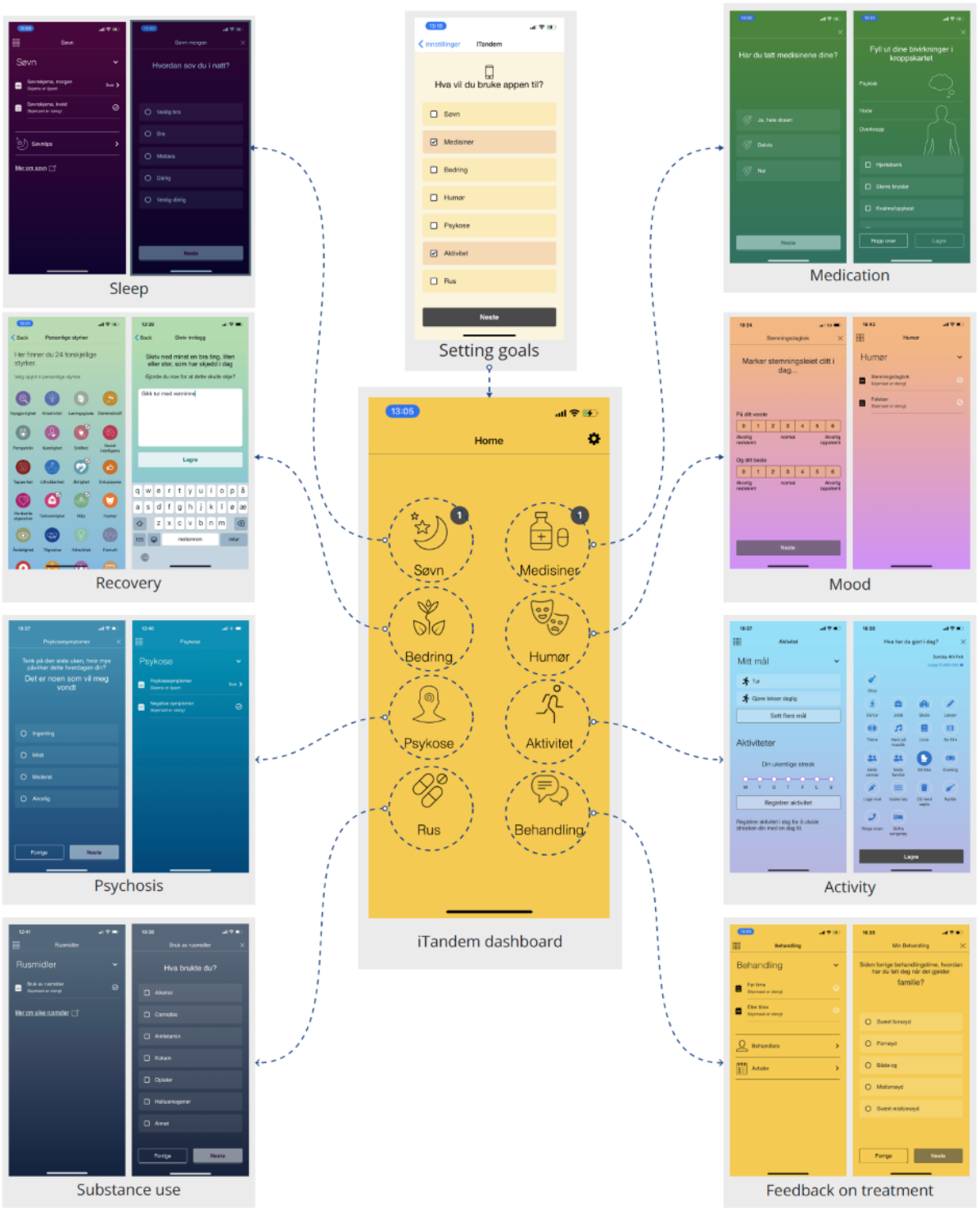
Substance use and abuse are of major concern in both psychotic [64] and bipolar disorders [65]. In addition, people with substance use disorder have a higher prevalence of mental disorders than the general population [66]. However, even though both service users and clinicians agree that this is a major problem, it became apparent that the service users did not want to focus on detailed intake of substances. They stated that thorough monitoring of their weekly substance use did not motivate change and argued for brief questioning on drug intake. However, they found it useful to be able to convey which strategies they had used to abstain from taking drugs during the week. Clinicians were more prone to ask for detailed accounts of substance use to get an overview but were also concerned that this could lead to demotivation and disengagement. In addition, we were guided by principles from motivational interviewing and aimed to engage the service user in strategies to abstain from using drugs when feeling a craving [67]. Hence, we chose the same strategy as with the Psychosis module and included only brief questions about use and abstaining strategies, along with a link to a recommended user-friendly web page about substance use.

Treatment

The final module of iTandem was included to adhere to the Norwegian guidelines for psychosis treatment, which underline the need for better tools to evaluate the experience of treatment. PROMs and patient-reported experience measures (PREMs) are widely used [68]. PROMs are primarily used for the assessment of treatment outcomes, while PREMs are used to evaluate patients' experiences with the service they receive [69]. They allow for more direct feedback to the provider. In a tool designed for therapeutic collaboration and alliance building, the service users' experience and ability to provide feedback on the sessions were considered important for iTandem. Feedback on treatment is also a fundamental principle in cognitive behavioral therapy, which is one of the most frequently used treatment approaches in mental health care. Therefore, including this module would also support cognitive behavioral therapy. The module consisted of questions about how the last week had been, and questions about the individual treatment sessions, to be answered before and after each session (Figure 1).

XSL•FO
RenderX

Figure 1. Screenshots illustrating the home page in iTandem and the page where you choose which modules you want to use (setting goals). The other pages show the various modules. Left: sleep, recovery, psychosis, and substance use. Right: medication, mood, activity, and feedback on treatment.



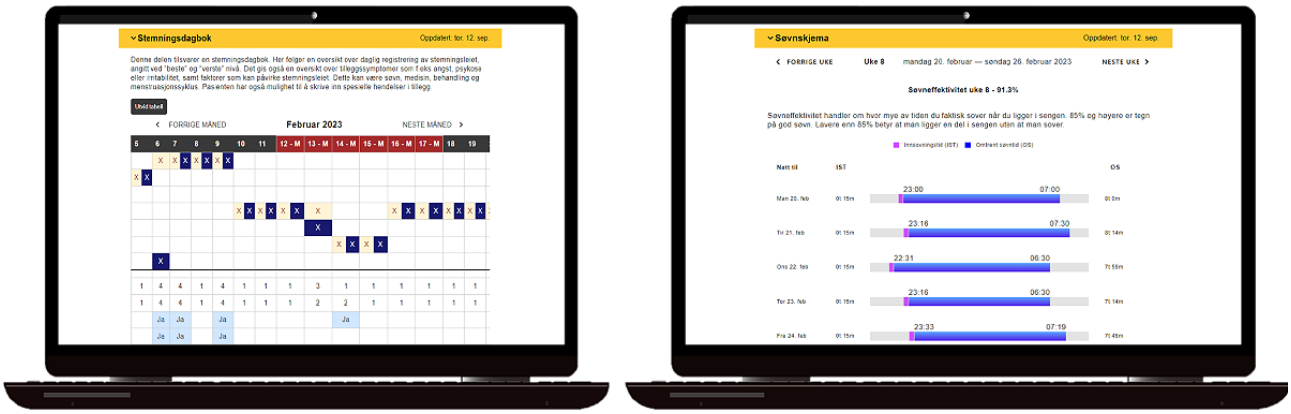
Report

Both clinicians and service users wanted a report to be able to log patterns. However, due to data security issues, only a limited amount of data could be stored on the phone. Except for the positive diary, activity, and medication list, data from all the modules were presented in the report. The positive diary was excluded from the report to allow service users to record positive events without feeling compelled to share them with clinicians. The Activity module incorporated elements of gamification, such as weekly streaks, which were believed to be a sufficient motivator. The other data entered were transferred to TSD. A report was available for clinicians, with the purpose of being shared with service users during sessions. Design principles were applied to make the report as easy to read as possible. Patients were not able to log into the report themselves. This decision was influenced by both technical and regulatory challenges, as well as how iTandem was intended to function

XSL•FO
RenderX

as a tool for collaboration and shared decision-making. By integrating the report reading into the sessions, we believe it better supports the main goal (Figure 2).

Figure 2. Illustration of 2 reports based on data logged into iTandem and how they will look on a screen. To the left is a mood diary illustrating data capturing the highest and the lowest mood during a day. To the right is a sleep diary, visualizing sleep during a week. The sleep period and sleep effectiveness are marked in different colors. A higher resolution version of this image is available as Multimedia Appendix 1.

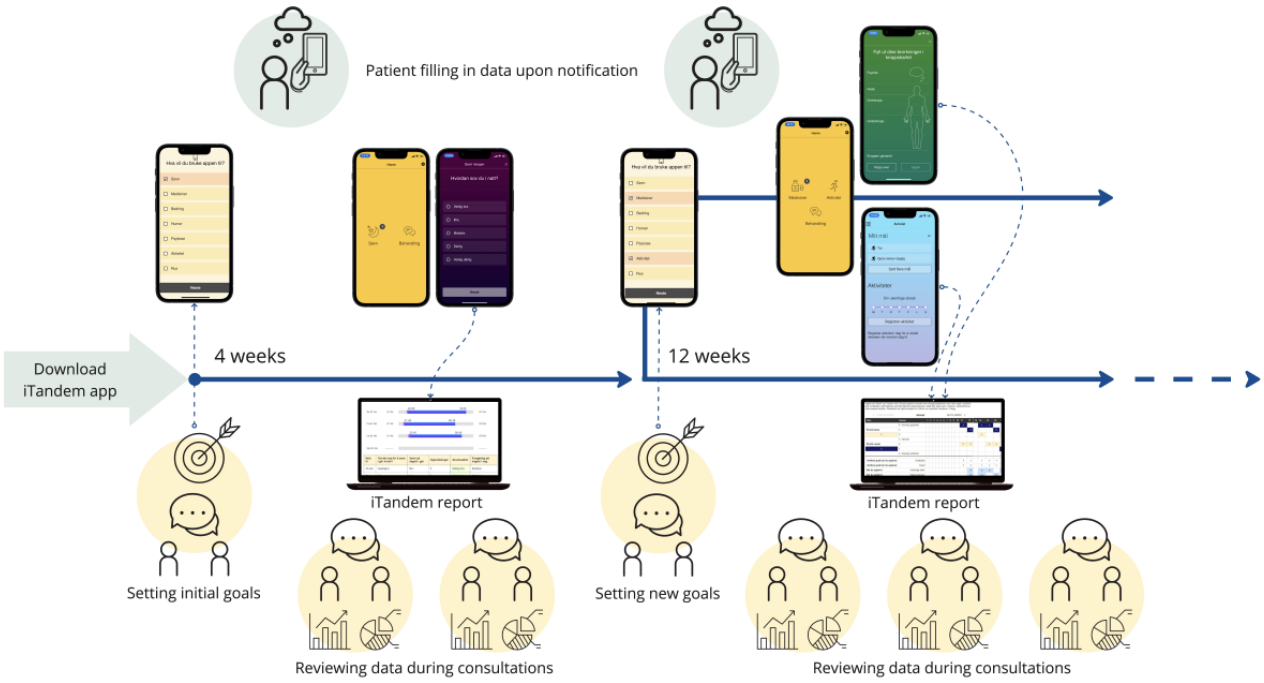


Beta Testing

Overview

Six clinicians (2 psychologists and 4 psychiatrists) were invited to participate in beta testing and attended a 3-hour training session on how to use iTandem. During this session, we demonstrated the different modules and their intended purposes. In addition, we provided an online protocol where participating clinicians and service users could access written information about the content and how to use the app. In collaboration with clinicians, service users were encouraged to choose 2-3 modules relevant to their treatment. It was not emphasized that this was set as an upper limit, but both clinicians and designers were concerned about overwhelming the patients with too many notifications. This could also undermine the overall goal of helping to focus sessions on issues considered particularly relevant to the patient. The patients used the app for 6 weeks (Figure 3).

Figure 3. A workflow with iTandem. After setting initial goals, the service user logs data, which are assessed during sessions. After monitoring sleep for 4 weeks, the service user and the clinician choose to change modules as they want to focus more on activities and medication.



After beta testing, 4 service users and 5 clinicians agreed to be interviewed about their experiences with iTandem. This served as a preliminary proof of concept to determine whether the concept appeared to have potential. We also wanted feedback to adjust both iTandem and the educational implementation program before proceeding with a larger test in a clinical setting. Several issues were addressed, and the responses were sorted based on what appeared to be the most relevant attitudes and sentiments following the test.

XSL•FO
RenderX

Feedback on Concept by Service Users

The service users were mainly positive toward iTandem and the possibilities it offered. They appreciated the ability to select different modules that captured issues beyond the symptoms of their primary diagnosis, working alongside their clinician. This approach made it easier for them to focus on subjects and issues that were important to them and to convey this effectively to their clinician, thus supporting collaboration on goal setting and shared decision-making. They had concerns about how the report was used during sessions and emphasized the importance of clinicians utilizing the logged data; otherwise, the aspect of collaboration would be undermined. Some service users felt that the questions in the app were a bit repetitive (mainly the validated forms), and called for more opportunities to use free text to convey important information to their clinicians. It should be noted that there were service users who signed the informed consent to participate but did not follow through. According to clinicians, this was related to either paranoid ideas, high general symptom load, or being transferred to other departments.

Feedback on the Concept by Clinicians

Clinicians found the data from iTandem in the report to be useful; it provided important insights for their therapeutic work. Several were surprised by how the focus of treatment shifted when they received new information through the app, aligning with the goal of shared decision-making and increased collaboration. However, they found it was not always easy to remember to access and read the report, and they called for the report to be made accessible in the regular hospital records. By contrast, when this was not an option, they found the final solution to be manageable.

Implementation Issues

Many service users were eager to try several modules at once. However, notifications from multiple modules were perceived as overwhelming, highlighting the importance of working with 2-3 modules at the same time. The time spent on each module was deemed appropriate. Clinicians also felt overwhelmed by the challenge of integrating a digital tool into treatment sessions while finding space for it in their clinical routine. They found it stressful not to be fully familiar with the tool, making it difficult for them to effectively explain its use to the service user.

Technical Issues

Several technical issues arose regarding both the data sent from the app to TSD and the presentation in the report. This was partly due to errors in the app, but also because the report functionality was newly developed by the web app group at USIT and had not been properly tested beforehand. This led to several periods of downtime for the report. Even though this was seen as a major problem that persisted for several weeks, participants continued to enter data and found it meaningful. Furthermore, there were issues regarding the user-friendliness of the clock in the Sleep module. It was too easy to enter the wrong date, which obviously resulted in incorrect sleep patterns in the report. We also discovered missing data in the report.

Revision After Beta Testing and Plans for Pilot Testing

Based on the beta testing, we managed to resolve the technical issues. We revised the information material and teaching manual, omitting text and focusing on video films to show and tell how the app is intended to be used. In addition, we only recommend using a limited number of modules at a time to avoid overwhelming notifications. The new information material, manual, and the revised iTandem have since been further tested in a pilot trial conducted during winter 2023. In this trial, we tested iTandem as part of an early intervention service in a rural area in Norway. Data from this trial are currently being analyzed and will be part of a PhD thesis on how digitalization impacts service delivery in mental health care.

Furthermore, several studies are planned, incorporating iTandem as part of studies that use single modules and investigate implementation research connected to digitalization in mental health care.

Discussion

We have developed and beta tested a digital tool, iTandem, based on principles from design thinking. The main aim has been to utilize technology to strengthen and focus shared decision-making between service users and clinicians in mental health care and to enhance goal setting. The demand for simple functionality was underlined by both service users and clinicians.

We believe the design process was helpful in designing a tool that accommodates shared decision-making. We found the tool to follow many of the principles recommended by Priebe and colleagues [10], such as a heightened focus on the service users' concerns, positive regard and personal respect, appropriate involvement of service users in decision-making, genuineness with personal touch, and the use of a psychological treatment model [10]. By allowing service users to choose modules to aid focus in treatment, we aim to empower them during their recovery process.

What is meaningful for the user should be the priority in treatment. This does not mean that clinical evaluation is less important, but to build an alliance and secure a therapeutic relationship, the service users' interests must remain in focus throughout therapy. This aligns with the demand for more patient-centered care. It is also in line with the shift from "patient-centered" to "person-centered" care, as discussed by Boardman and Dave [70]. There is a change in how we view the therapeutic relationship, where the therapeutic alliance is more about co-production as a key element in the clinician-service user collaboration.

Furthermore, we believe that the ability to choose themes to work on and the possibility to give feedback on treatment sessions, including questions about being met with respect, will underpin service users' experience of being respected and treated in a nonjudgmental way. We hope this tool will enhance clinicians' motivation to make changes in accordance with the service users' experienced needs. By using design both as a creative process and applied to the visual elements, we have strived to make the tool engaging and less "clinical." This may increase the feeling of a more personal touch. To accommodate

XSL•FO
RenderX

a more generic therapeutic use, the elements of iTandem are well suited to fit cognitive behavioral therapy, focusing on agenda setting, goal setting, and the execution of activities. Unsurprisingly, a design process led to solutions that align well with this therapy approach, as it is collaborative, pragmatic, structured, and focused on current problems.

We have previously worked with Boundary Objects Theory, where using a digital object can be seen as a way to transfer knowledge between different knowledge domains [71]. Based on this work, for an object to function effectively, it needs to inherit features such as dynamism, flexibility, standardization, and a shared structure [72]. We believe iTandem has the potential to work as a boundary object. It accommodates the needs of the service user, and it is simple to engage in, hence adhering to the KISS principle. iTandem allows for the transfer of knowledge between service user and clinician with an inherent flexibility yet standardized approach. The ability to collaborate via this object allows for a deeper understanding of the problem, leading to more focused problem-solving.

iTandem differs from the numerous assessment tools offered as PROMs in Norwegian mental health care services at the moment. The positive aspects of PROMs lie in their basis on validated measures that aim to track progress in therapy. However, the extensive number of questions to be answered can be overwhelming, and the amount of data may be challenging for clinicians to manage and oversee effectively. PROMs often cover broad categories, and their questions can sometimes seem less relevant to the service users' specific problems. Therefore, it may be necessary to scale down the assessments to focus on a minimum that allows for qualified evaluations of how treatment progresses. This approach can alleviate the burden on clinicians and service users by combining different tools, such as apps like iTandem and more focused classic PROMs. Utilizing language and questions that are closely aligned with a clinical setting might be sufficient to establish a rapport and gather relevant information about progress.

Another important aspect is how we provide flexibility in services. While we offer options such as video consultations, PROMs, apps, and similar tools, the solutions presented to service users often rely on traditional approaches, as if one size fits all. Alternating between different tools and formats during treatment sessions could, in our view, make the treatment more effective. Insights from user reviews of mental health apps indicate that highly rated apps tend to offer a variety of options, functionalities, and content for users to choose from [73]. This is particularly relevant to attrition rates, as different tools may be needed during various phases of the recovery process. A lack of use could even be a positive sign, indicating that the problem has been resolved and the user has moved on. Additionally, Alqahtani et al [73] identified several common weaknesses of mental health apps, including poor usability, lack of personalization, credibility issues, security concerns, and inadequate customer service. We believe these findings align with the need for greater flexibility in services.

To address the future needs of health care services, we believe it is essential to examine how digital tools can be integrated into regular workflows. Greater insight is needed into the barriers and facilitators of implementation, as well as how health technology assessment methods can be used to evaluate which technologies benefit service users, clinicians, and organizations while also being economically effective.

Acknowledgments

The development of iTandem was funded by the South-Eastern Norway Regional Health Authority (reference number 17/00264-14). We want to thank all the service users and clinicians who willingly have offered their time and interest in this development process. We also want to thank the agile team at USIT, which is not represented as co-authors in this paper, for good collaboration. We especially thank the senior developer, Magnus Alderslyst Nygaard, and the highly skilled, creative designers. Finally, we will thank Jonathan Romm for the illustrations and Trine Vik Lagerberg for valuable support along the way.

Data Availability

Because of the small sample size and limitations in the written informed consent, the data are not available to other researchers.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Higher Resolution version of Figure 2.
[PNG File , 409 KB - formative_v9i1e57614_app1.png]

References

1. Brédart A, Bouleuc C, Dolbeault S. Doctor-patient communication and satisfaction with care in oncology. Curr Opin Oncol 2005 Jul;17(4):351-354. [doi: 10.1097/01.cco.0000167734.26454.30] [Medline: 15933466]
2. Arora NK. Interacting with cancer patients: the significance of physicians' communication behavior. Soc Sci Med 2003 Sep;57(5):791-806. [doi: 10.1016/s0277-9536(02)00449-5] [Medline: 12850107]

XSL•FO
RenderX

3. Lee SJ, Back AL, Block SD, Stewart SK. Enhancing physician-patient communication. Hematology Am Soc Hematol Educ Program 2002:464-483. [doi: 10.1182/asheducation-2002.1.464] [Medline: 12446437]
4. Aftab A, LaGrotta C, Zyzanski SJ, Mishra P, Mehdi SMA, Brown K, et al. Impact of psychiatric hospitalization on trust, disclosure and working alliance with the outpatient psychiatric provider: a pilot survey study. Cureus 2019 Apr 22;11(4):e4515 [FREE Full text] [doi: 10.7759/cureus.4515] [Medline: 31259124]
5. Rübig LL, Fuchshuber J, Köldorfer P, Rinner A, Fink A, Unterrainer H. Attachment and therapeutic alliance in substance use disorders: initial findings for treatment in the therapeutic community. Front Psychiatry 2021;12:730876 [FREE Full text] [doi: 10.3389/fpsyt.2021.730876] [Medline: 34858223]
6. Reddy MS. Insight and psychosis. Indian J Psychol Med 2015;37(3):257-260 [FREE Full text] [doi: 10.4103/0253-7176.162909] [Medline: 26664072]
7. Harvey PD, Bosia M, Cavallaro R, Howes OD, Kahn RS, Leucht S, et al. Cognitive dysfunction in schizophrenia: an expert group paper on the current state of the art. Schizophr Res Cogn 2022 Sep;29:100249 [FREE Full text] [doi: 10.1016/j.scog.2022.100249] [Medline: 35345598]
8. Lyngstad SH, Gardsjord ES, Engen MJ, Haatveit B, Ihler HM, Wedervang-Resell K, et al. Trajectory and early predictors of apathy development in first-episode psychosis and healthy controls: a 10-year follow-up study. Eur Arch Psychiatry Clin Neurosci 2020 Sep;270(6):709-722 [FREE Full text] [doi: 10.1007/s00406-020-01112-3] [Medline: 32130475]
9. Ihler HM, Lyngstad SH, Mørch-Johnsen LE, Lagerberg TV, Melle I, Romm KL. A transdiagnostic approach to negative symptoms: exploring factor structure and negative symptoms in bipolar disorders. Front Psychiatry 2023;14:1136097 [FREE Full text] [doi: 10.3389/fpsyt.2023.1136097] [Medline: 37398600]
10. Priebe S, Dimic S, Wildgrube C, Jankovic J, Cushing A, McCabe R. Good communication in psychiatry--a conceptual review. Eur Psychiatry 2011 Oct;26(7):403-407. [doi: 10.1016/j.eurpsy.2010.07.010] [Medline: 21571504]
11. Feldthusen C, Forsgren E, Wallström S, Andersson V, Löfqvist N, Sawatzky R, et al. Centredness in health care: a systematic overview of reviews. Health Expect 2022 Jun;25(3):885-901 [FREE Full text] [doi: 10.1111/hex.13461] [Medline: 35261138]
12. Brown K, Parry S. How do people with first episode psychosis experience therapeutic relationships with mental health practitioners? A narrative review. Psychosis 2022 Dec 22;16(2):155-166. [doi: 10.1080/17522439.2022.2160487]
13. Hamann J, Heres S. Adapting shared decision making for individuals with severe mental illness. Psychiatr Serv 2014 Dec 01;65(12):1483-1486. [doi: 10.1176/appi.ps.201400307] [Medline: 25756970]
14. Simmons M, Hetrick S, Jorm A. Shared decision-making: benefits, barriers and current opportunities for application. Australas Psychiatry 2010 Oct;18(5):394-397. [doi: 10.3109/10398562.2010.499944] [Medline: 20863175]
15. Zisman-Ilani Y, Barnett E, Harik J, Pavlo A, O'Connell M. Expanding the concept of shared decision making for mental health: systematic search and scoping review of interventions. MHRJ 2017 Sep 11;22(3):191-213. [doi: 10.1108/mhrj-01-2017-0002]
16. Zisman-Ilani Y, Lysaker PH, Hasson-Ohayon I. Shared risk taking: shared decision making in serious mental illness. Psychiatr Serv 2021 Apr 01;72(4):461-463. [doi: 10.1176/appi.ps.202000156] [Medline: 33557595]
17. Castonguay LG, Constantino MJ, Holtforth MG. The working alliance: Where are we and where should we go? Psychotherapy (Chic) 2006;43(3):271-279. [doi: 10.1037/0033-3204.43.3.271] [Medline: 22122096]
18. Weldring T, Smith SMS. Patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs). Health Serv Insights 2013;6:61-68 [FREE Full text] [doi: 10.4137/HSI.S11093] [Medline: 25114561]
19. Nguyen H, Butow P, Dhillon H, Sundaresan P. A review of the barriers to using patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs) in routine cancer care. J Med Radiat Sci 2021 Jun;68(2):186-195 [FREE Full text] [doi: 10.1002/jmrs.421] [Medline: 32815314]
20. Nwosu A, Boardman S, Husain MM, Doraiswamy PM. Digital therapeutics for mental health: is attrition the Achilles heel? Front Psychiatry 2022;13:900615 [FREE Full text] [doi: 10.3389/fpsyt.2022.900615] [Medline: 35982936]
21. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
22. Mordecai D, Histon T, Neuwirth E, Heisler WS, Kraft A, Bang Y, et al. How Kaiser permanente created a mental health and wellness digital ecosystem. NEJM Catalyst 2021 Jan;2(1):1 [FREE Full text] [doi: 10.1056/cat.20.0295]
23. Dalzell T. The Routledge Dictionary of Modern American Slang and Unconventional English. London, UK: Routledge; Jul 25, 2008:595.
24. Alwin DF, Beattie BA. The KISS principle in survey design. Sociological Methodology 2016 Jun 21;46(1):121-152. [doi: 10.1177/0081175016641714]
25. Rich BR. Clarence Leonard (Kelly) Johnson 1910-1990: A Biographical Memoir. Washington, DC: The National Academies Press; 1995:221-241.
26. Vial S, Boudhraâ S, Dumont M. Human-centered design approaches in digital mental health interventions: exploratory mapping review. JMIR Ment Health 2022 Jun 07;9(6):e35591 [FREE Full text] [doi: 10.2196/35591] [Medline: 35671081]
27. Halskov K, Hansen N. The diversity of participatory design research practice at PDC 2002–2012. International Journal of Human-Computer Studies 2015 Feb;74:81-92 [FREE Full text] [doi: 10.1016/j.ijhcs.2014.09.003]

XSL•FO
RenderX

28. Bergin AD, Vallejos EP, Davies EB, Daley D, Ford T, Harold G, et al. Preventive digital mental health interventions for children and young people: a review of the design and reporting of research. NPJ Digit Med 2020;3:133 [FREE Full text] [doi: 10.1038/s41746-020-00339-7] [Medline: 33083568]
29. Romm KL, Gardsjord ES, Gjermundsen K, Aguirre Ulloa M, Berentzen L, Melle I. Designing easy access to care for first-episode psychosis in complex organizations. Early Interv Psychiatry 2019 Oct 27;13(5):1276-1282. [doi: 10.1111/eip.12802] [Medline: 30919597]
30. Roberts JP, Fisher TR, Trowbridge MJ, Bent C. A design thinking framework for healthcare management and innovation. Healthc (Amst) 2016 Mar;4(1):11-14. [doi: 10.1016/j.hjdsi.2015.12.002] [Medline: 27001093]
31. Altman M, Huang TT, Breland JY. Design thinking in health care. Prev Chronic Dis 2018 Sep 27;15:E117 [FREE Full text] [doi: 10.5888/pcd15.180128] [Medline: 30264690]
32. The Design Council. The Double Diamond. The Design Council. URL: https://www.designcouncil.org.uk/our-resources/the-double-diamond/ [accessed 2024-12-13]
33. Laskemoen JF, Simonsen C, Büchmann C, Barrett EA, Bjella T, Lagerberg TV, et al. Sleep disturbances in schizophrenia spectrum and bipolar disorders - a transdiagnostic perspective. Compr Psychiatry 2019 May;91:6-12 [FREE Full text] [doi: 10.1016/j.comppsych.2019.02.006] [Medline: 30856497]
34. Sylvia LG, Chang WC, Kamali M, Tohen M, Kinrys G, Deckersbach T, et al. Sleep disturbance may impact treatment outcome in bipolar disorder: a preliminary investigation in the context of a large comparative effectiveness trial. J Affect Disord 2018 Jan 01;225:563-568. [doi: 10.1016/j.jad.2017.08.056] [Medline: 28881294]
35. Carney CE, Buysse DJ, Ancoli-Israel S, Edinger JD, Krystal AD, Lichstein KL, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. Sleep 2012 Feb 01;35(2):287-302 [FREE Full text] [doi: 10.5665/sleep.1642] [Medline: 22294820]
36. www.psykose-bipolar.no. Psykose/Bipolar. Oslo, Norway: Oslo University Hospital URL: https://www.psykose-bipolar.no/bipolar-lidelse/tidlige-tegn-pa-bipolar-lidelse/ [accessed 2024-12-17]
37. Zhu Y, Li C, Huhn M, Rothe P, Krause M, Bighelli I, et al. How well do patients with a first episode of schizophrenia respond to antipsychotics: a systematic review and meta-analysis. Eur Neuropsychopharmacol 2017 Sep;27(9):835-844. [doi: 10.1016/j.euroneuro.2017.06.011] [Medline: 28669774]
38. Keramatian K, Chakrabarty T, Saraf G, Yatham LN. New developments in the use of atypical antipsychotics in the treatment of bipolar disorder: a systematic review of recent randomized controlled trials. Curr Psychiatry Rep 2021 May 08;23(7):39. [doi: 10.1007/s11920-021-01252-w] [Medline: 33963957]
39. Kishi T, Ikuta T, Matsuda Y, Sakuma K, Okuya M, Mishima K, et al. Mood stabilizers and/or antipsychotics for bipolar disorder in the maintenance phase: a systematic review and network meta-analysis of randomized controlled trials. Mol Psychiatry 2021 Aug;26(8):4146-4157 [FREE Full text] [doi: 10.1038/s41380-020-00946-6] [Medline: 33177610]
40. Haddad PM, Correll CU. The acute efficacy of antipsychotics in schizophrenia: a review of recent meta-analyses. Ther Adv Psychopharmacol 2018 Nov;8(11):303-318 [FREE Full text] [doi: 10.1177/2045125318781475] [Medline: 30344997]
41. Bjornestad J, Lavik KO, Davidson L, Hjeltnes A, Moltu C, Veseth M. Antipsychotic treatment - a systematic literature review and meta-analysis of qualitative studies. J Ment Health 2020 Oct;29(5):513-523. [doi: 10.1080/09638237.2019.1581352] [Medline: 30862219]
42. Webster RK, Weinman J, Rubin GJ. How does the side-effect information in patient information leaflets influence peoples' side-effect expectations? A cross-sectional national survey of 18- to 65-year-olds in England. Health Expect 2017 Dec;20(6):1411-1420 [FREE Full text] [doi: 10.1111/hex.12584] [Medline: 28618140]
43. Lindström E, Lewander T, Malm U, Malt UF, Lublin H, Ahlfors UG. Patient-rated versus clinician-rated side effects of drug treatment in schizophrenia. Clinical validation of a self-rating version of the UKU Side Effect Rating Scale (UKU-SERS-Pat). Nord J Psychiatry 2001;55 Suppl 44:5-69. [doi: 10.1080/080394801317084428] [Medline: 11860666]
44. Leamy M, Bird V, Le Boutillier C, Williams J, Slade M. Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. Br J Psychiatry 2011 Dec;199(6):445-452. [doi: 10.1192/bjp.bp.110.083733] [Medline: 22130746]
45. Slade M, Amering M, Farkas M, Hamilton B, O'Hagan M, Panther G, et al. Uses and abuses of recovery: implementing recovery-oriented practices in mental health systems. World Psychiatry 2014 Feb;13(1):12-20 [FREE Full text] [doi: 10.1002/wps.20084] [Medline: 24497237]
46. Åsbø G, Ueland T, Haatveit B, Bjella T, Flaaten CB, Wold KF, et al. The time is ripe for a consensus definition of clinical recovery in first-episode psychosis: suggestions based on a 10-year follow-up study. Schizophr Bull 2022 Jun 21;48(4):839-849. [doi: 10.1093/schbul/sbac035] [Medline: 35419608]
47. Slade M, Brownell T, Rashid T, Schrank B. Positive Psychotherapy for Psychosis: A Clinician's Guide and Manual. London, UK: Routledge; Dec 19, 2016.
48. Neil ST, Kilbride M, Pitt L, Nothard S, Welford M, Sellwood W, et al. The questionnaire about the process of recovery (QPR): a measurement tool developed in collaboration with service users. Psychosis 2009 Aug;1(2):145-155. [doi: 10.1080/17522430902913450]

XSL•FO
RenderX

49. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5:63 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
50. Lehman A. A quality of life interview for the chronically mentally ill. Evaluation and Program Planning 1988 Jul 03;11(1):51-62 [FREE Full text] [doi: 10.1016/0149-7189(88)90033-X]
51. Dickens LR. Using gratitude to promote positive change: a series of meta-analyses investigating the effectiveness of gratitude interventions. Basic and Applied Social Psychology 2017 May 30;39(4):193-208. [doi: 10.1080/01973533.2017.1323638]
52. Tse S, Tsoi EW, Hamilton B, O'Hagan M, Shepherd G, Slade M, et al. Uses of strength-based interventions for people with serious mental illness: a critical review. Int J Soc Psychiatry 2016 May;62(3):281-291. [doi: 10.1177/0020764015623970] [Medline: 26831826]
53. Riches S, Schrank B, Rashid T, Slade M. WELLFOCUS PPT: modifying positive psychotherapy for psychosis. Psychotherapy (Chic) 2016 Mar;53(1):68-77. [doi: 10.1037/pst0000013] [Medline: 25961372]
54. Upthegrove R, Marwaha S, Birchwood M. Depression and schizophrenia: cause, consequence, or trans-diagnostic issue? Schizophr Bull 2017 Mar 01;43(2):240-244 [FREE Full text] [doi: 10.1093/schbul/sbw097] [Medline: 27421793]
55. Alsheikh AM, Elemam MO, El-Bahnasawi M. Treatment of depression with alcohol and substance dependence: a systematic review. Cureus 2020 Oct 26;12(10):e11168 [FREE Full text] [doi: 10.7759/cureus.11168] [Medline: 33133799]
56. Skar-Fröding R, Clausen H, Šaltytė Benth J, Ruud T, Slade M, S Heiervang K. Associations between personal recovery and service user-rated versus clinician-rated clinical recovery, a cross-sectional study. BMC Psychiatry 2022 Jan 18;22(1):42 [FREE Full text] [doi: 10.1186/s12888-022-03691-y] [Medline: 35042494]
57. Yasui-Furukori N, Nakamura K. Bipolar disorder recurrence prevention using self-monitoring daily mood charts: case reports from a 5 year period. NDT 2017 Mar;Volume 13:733-736. [doi: 10.2147/ndt.s132355]
58. Sachs GS, Guille C, McMurrich SL. A clinical monitoring form for mood disorders. Bipolar Disord 2002 Oct;4(5):323-327. [doi: 10.1034/j.1399-5618.2002.01195.x] [Medline: 12479665]
59. Slade M. Implementing shared decision making in routine mental health care. World Psychiatry 2017 Jun;16(2):146-153 [FREE Full text] [doi: 10.1002/wps.20412] [Medline: 28498575]
60. Galderisi S, Mucci A, Dollfus S, Nordentoft M, Falkai P, Kaiser S, et al. EPA guidance on assessment of negative symptoms in schizophrenia. Eur Psychiatry 2021 Feb 18;64(1):e23 [FREE Full text] [doi: 10.1192/j.eurpsy.2021.11] [Medline: 33597064]
61. Dollfus S, Mach C, Morello R. Self-evaluation of negative symptoms: a novel tool to assess negative symptoms. Schizophr Bull 2016 May;42(3):571-578 [FREE Full text] [doi: 10.1093/schbul/sbv161] [Medline: 26564898]
62. Stewart V, McMillan SS, Hu J, Ng R, El-Den S, O'Reilly C, et al. Goal planning in mental health service delivery: a systematic integrative review. Front Psychiatry 2022;13:1057915 [FREE Full text] [doi: 10.3389/fpsyt.2022.1057915] [Medline: 36601527]
63. Rose G, Smith L. Mental health recovery, goal setting and working alliance in an Australian community-managed organisation. Health Psychol Open 2018;5(1):2055102918774674 [FREE Full text] [doi: 10.1177/2055102918774674] [Medline: 29785279]
64. Hartz SM, Pato CN, Medeiros H, Cavazos-Rehg P, Sobell JL, Knowles JA, Genomic Psychiatry Cohort Consortium. Comorbidity of severe psychotic disorders with measures of substance use. JAMA Psychiatry 2014 Mar;71(3):248-254 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.3726] [Medline: 24382686]
65. Hunt GE, Malhi GS, Cleary M, Lai HMX, Sitharthan T. Prevalence of comorbid bipolar and substance use disorders in clinical settings, 1990-2015: systematic review and meta-analysis. J Affect Disord 2016 Dec;206:331-349. [doi: 10.1016/j.jad.2016.07.011] [Medline: 27476137]
66. Richert T, Anderberg M, Dahlberg M. Mental health problems among young people in substance abuse treatment in Sweden. Subst Abuse Treat Prev Policy 2020 Jun 24;15(1):43 [FREE Full text] [doi: 10.1186/s13011-020-00282-6] [Medline: 32580732]
67. Galvani S. Motivational Interviewing: Helping People Change (3rd edition). Social Work Education 2014 Mar 04;33(6):854-855. [doi: 10.1080/02615479.2014.894351]
68. Friedel AL, Siegel S, Kirstein CF, Gerigk M, Bingel U, Diehl A, et al. Measuring patient experience and patient satisfaction-how are we doing it and why does it matter? A comparison of European and U.S. American approaches. Healthcare (Basel) 2023 Mar 08;11(6):797 [FREE Full text] [doi: 10.3390/healthcare11060797] [Medline: 36981454]
69. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. BJA Education 2017 Apr;17(4):137-144. [doi: 10.1093/bjaed/mkw060]
70. Boardman J, Dave S. Person-centred care and psychiatry: some key perspectives. BJPsych Int 2020 Aug;17(3):65-68 [FREE Full text] [doi: 10.1192/bji.2020.21] [Medline: 34287426]
71. Saidi T, Mork E, Aminoff S, Lobban F, Romm KL. Crossing boundaries in the delivery of healthcare - a qualitative study of an eHealth intervention in relation to boundary object theory. Digit Health 2023;9:20552076231196970 [FREE Full text] [doi: 10.1177/20552076231196970] [Medline: 37662674]
72. Star SL, Griesemer JR. Institutional Ecology, `Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. Soc Stud Sci 1989 Aug 01;19(3):387-420. [doi: 10.1177/030631289019003001]

XSL•FO
RenderX

73. Alqahtani F, Orji R. Insights from user reviews to improve mental health apps. Health Informatics J 2020 Sep;26(3):2042-2066 [FREE Full text] [doi: 10.1177/1460458219896492] [Medline: 31920160]

Abbreviations

KISS: keep it simple, stupid
OUH: Oslo University Hospital
PREM: patient-reported experience measure
PROM: patient-reported outcome measure
TIPS SE: Early Intervention in Psychosis Advisory Unit for South East Norway
TSD: Services for Sensitive Data Tjenester for Sensitive Data
UiO: University of Oslo
USIT: University of Oslo's Center for Information Technology
UX: user experience

Edited by A Mavragani; submitted 22.02.24; peer-reviewed by A Oostdyk, A Kurilla; comments to author 14.10.24; revised version received 28.10.24; accepted 06.11.24; published 17.01.25.

Please cite as:
Romm KL, Skoge M, Barrett EA, Berentzen LC, Bergsager D, Fugelli P, Bjella T, Gardsjord ES, Kling K, Kruse SH, Kværner KJ, Melle I, Mork E, Ihler HM, Rognli EB, Simonsen C, Værnes TG, Aminoff SR
A Mobile Health Intervention to Support Collaborative Decision-Making in Mental Health Care: Development and Usability
JMIR Form Res 2025;9:e57614
URL: https://formative.jmir.org/2025/1/e57614
doi:10.2196/57614
PMID:

©Kristin Lie Romm, Mari Skoge, Elizabeth Ann Barrett, Lars-Christian Berentzen, Dagfinn Bergsager, Pål Fugelli, Thomas Bjella, Erlend Strand Gardsjord, Kristine Kling, Sindre Hembre Kruse, Kari Jorunn Kværner, Ingrid Melle, Erlend Mork, Henrik Myhre Ihler, Eline Borger Rognli, Carmen Simonsen, Tor Gunnar Værnes, Sofie Ragnhild Aminoff. Originally published in JMIR Formative Research (https://formative.jmir.org), 17.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Mindfulness-Based App Intervention for Pregnant Women: Qualitative Evaluation of a Prototype Using Multiple Case Studies

Silvia Rizzi[1], MD; Maria Chiara Pavesi[2], MD; Alessia Moser[2], MD; Francesca Paolazzi[2], MD; Michele Marchesoni[3], MD; Stefania Poggianella[1], MD; Erik Gadotti[2], MD; Stefano Forti[4], MD

[1]Digital Health Research, Centre for Digital Health & Wellbeing, Fondazione Bruno Kessler, Trento, Italy

[2]Istituto Pavoniano Artigianelli, Trento, Italy

[3]Digital Health Innovation Lab, Centre for Digital Health & Wellbeing, Fondazione Bruno Kessler, Trento, Italy

[4]Centre for Digital Health & Wellbeing, Fondazione Bruno Kessler, Trento, Italy

Corresponding Author:
Silvia Rizzi, MD
Digital Health Research
Centre for Digital Health & Wellbeing
Fondazione Bruno Kessler
Via Sommarive 18
Trento, 38123
Italy
Phone: 39 0461312415
Email: srizzi@fbk.eu

Abstract

Background: Pregnancy is a complex period characterized by significant transformations. How a woman adapts to these changes can affect her quality of life and psychological well-being. Recently developed digital solutions have assumed a crucial role in supporting the psychological well-being of pregnant women. However, these tools have mainly been developed for women who already present clinically relevant psychological symptoms or mental disorders.

Objective: This study aimed to develop a mindfulness-based well-being intervention for all pregnant women that can be delivered electronically and guided by an online assistant with wide reach and dissemination. This paper aimed to describe a prototype technology-based mindfulness intervention's design and development process for pregnant women, including the exploration phase, intervention content development, and iterative software development (including design, development, and formative evaluation of paper and low-fidelity prototypes).

Methods: Design and development processes were iterative and performed in close collaboration with key stakeholders (N=15), domain experts including mindfulness experts (n=2), communication experts (n=2), and psychologists (n=3), and target users including pregnant women (n=2), mothers with young children (n=2), and midwives (n=4). User-centered and service design methods, such as interviews and usability testing, were included to ensure user involvement in each phase. Domain experts evaluated a paper prototype, while target users evaluated a low-fidelity prototype. Intervention content was developed by psychologists and mindfulness experts based on the Mindfulness-Based Childbirth and Parenting program and adjusted to an electronic format through multiple iterations with stakeholders.

Results: An 8-session intervention in a prototype electronic format using text, audio, video, and images was designed. In general, the prototypes were evaluated positively by the users involved. The questionnaires showed that domain experts, for instance, positively evaluated chatbot-related aspects such as empathy and comprehensibility of the terms used and rated the mindfulness traces present as supportive and functional. The target users found the content interesting and clear. However, both parties regarded the listening as not fully active. In addition, the interviews made it possible to pick up useful suggestions in order to refine the intervention. Domain experts suggested incorporating auditory components alongside textual content or substituting text entirely with auditory or audiovisual formats. Debate surrounded the inclusion of background music in mindfulness exercises, with opinions divided on its potential to either distract or aid in engagement. The target users proposed to supplement the app with some face-to-face meetings at crucial moments of the course, such as the beginning and the end.

XSL•FO
RenderX

Conclusions: This study illustrates how user-centered and service designs can be applied to identify and incorporate essential stakeholder aspects in the design and development process. Combined with evidence-based concepts, this process facilitated the development of a mindfulness intervention designed for the end users, in this case, pregnant women.

(JMIR Form Res 2025;9:e58265) doi:10.2196/58265

KEYWORDS

mindfulness; promoting well-being; pregnancy; eHealth; mHealth; mobile apps; development; usability; user-centered design; well-being; maternal health; digital health; intervention; design; preliminary testing; technology-based; interview; multidisciplinary approach; mother; women; WhatsApp; email; midwife

Introduction

Background

In a woman's life, the transition to motherhood is a time of profound change, requiring reorganizing her identity [1]. A woman's general well-being is affected by pregnancy, as various changes are expected at the physical, mental, and social levels [2]. Several studies are presented in the literature showing the presence of psychological symptoms in pregnant women. The presence of symptoms of anxiety, stress, and depression are most commonly emphasized [3,4]. Regarding anxiety and psychosocial stress, it is estimated that between 12.2% and 39% of pregnant women may suffer from these conditions [5]. Psychosocial anxiety is defined as an imbalance resulting precisely from the challenges of the new condition [6]. Therefore, a woman's quality of life and her psychological health are affected by how she reacts to all these changes. The vulnerable period of pregnancy and early motherhood also often involves increased worries and fears related to managing the impending changes associated with childbirth and parenting [7,8]. Woods and colleagues [9] found a spectrum of stress symptoms ranging from 6% (91/1522; high levels) to 78% (1190/1522; medium-low stress levels). Only 16% (241/1522) of the sample reported nothing. Another clinically significant condition in pregnant women is perinatal depression [10]. This occurs in 14.5% during pregnancy and reaches up to 49% in the first year after delivery [3].

Although psychoeducational interventions aimed at the psychological well-being of pregnant women are still scarce, ensuring a positive pregnancy experience is crucial, as it lays the foundation for cognitive motivation and a healthy future relationship between mother and child [11]. The use of digital tools represents a promising solution. Indeed, digital interventions offer accessibility and flexibility and overcome social stigma, making them an optimal choice for women seeking support [12,13].

Recognizing the critical need to promote psychological well-being during the transition to motherhood, especially in the digital age, becomes imperative.

According to Kabat-Zinn [14], mindfulness is defined as "the awareness that arises from paying attention, on purpose, in the present moment, and non-judgmentally." This practice involves actively noticing one's thoughts, emotions, and bodily sensations and fostering an attitude of openness and acceptance. Kabat-Zinn [15] emphasizes the significance of mindfulness in reducing stress and enhancing overall well-being. Mindfulness practices enable pregnant women to manage stress, anxiety, and depression actively, equipping them with valuable skills for various situations, including childbirth [14,16-19]. For this reason, mindfulness interventions delivered through web-based platforms, for example, have shown promise in improving the psychological well-being of pregnant women [17,20,21].

Therefore, psychoeducational interventions using mindfulness-based digital solutions have significant potential to promote psychological well-being, foster women's psychological adjustment during physiological pregnancies, and prevent clinically relevant symptoms [22].

This study illustrates the creation of a tailored mindfulness intervention for pregnant women using technology and a dedicated mobile app. Building on established mindfulness concepts proven effective in face-to-face interventions [23-25], the study adopted a user-centered design approach to ensure alignment of the intervention with users' needs and contextual use [26,27]. The fundamental principle of user-centered design is to actively involve users in the design and development stages, ensuring that they influence the evolution of the product [28].

The user-centered design approach ensures a holistic understanding of user perspectives by incorporating various methods that facilitate user participation at different development stages and involvement levels [29]. In addition, the study incorporated service design principles, emphasizing whole ecosystem experiences, such as how and when it is used, by whom, and where, rather than focusing exclusively on the end product [30]. This fusion of user-centered and service design approaches aimed to involve stakeholders, including pregnant women, health care providers, mindfulness and communication experts, psychologists, and midwives, throughout the design and development process. The overall goal was to create a final product that was easy to use, practical, engaging, and motivating, and fit seamlessly into the larger context of daily life and the challenges faced during pregnancy.

Mindfulness-Based Childbirth and Parenting

The Mindfulness-Based Childbirth and Parenting (MBCP) protocol was created in 1998 by Nancy Bardacke, a nurse, midwife, and mindfulness trainer. Through participation in the MBCP protocol, extrapolated from Kabat-Zinn's [15] Mindfulness-Based Stress Reduction (MBSR) protocol, pregnant women and their partners have the opportunity to learn mindfulness techniques to cope with the states of anxiety and stress ordinarily present during pregnancy, pain, and fear during childbirth, as well as to develop couple cooperation and parenting sensitivity [24,25,31]. MBCP was initially a group

XSL•FO
RenderX

course held for 3 hours once a week for 9 weeks. The 9-week MBCP course includes a day of silent meditation between weeks 6 and 7 and an additional reunion gathering after all women have given birth. Participants begin MBCP in the late second or early third trimester and are asked to commit to home practice using guided meditation CDs for 30 minutes a day, 6 days a week throughout the course. Formal mindfulness meditation instruction is given in each class, including sitting meditation, body scan, mindful yoga, and loving-kindness meditation, all tailored to include attention to mental and physical aspects of pregnancy (eg, noting fetal movements during the body scan) [24]. For a description of the main topics covered week by week, refer to Multimedia Appendix 1.

The literature indicates the extraordinary effectiveness of the MBCP protocol both during [23,32-36] and after pregnancy [20,37-39]. Creswell [22] found a moderate effect size (Cohen *d*=0.5) for mindfulness interventions in reducing anxiety and depression in pregnant women. Duncan et al [32] reported a pooled effect size of *d*=0.48 for mindfulness-based interventions in enhancing psychological well-being during pregnancy.

Some of the benefits of mindful birthing are (1) an increase in confidence and decrease in fear of childbirth; (2) tapping into deep inner resources for working with pain; (3) improving couple communication, connection, and cooperation; and (4) providing stress-reducing skills for greater joy and well-being [40].

Results suggest the effects of a group-based psychosocial intervention during pregnancy may endure for years, well beyond the initial perinatal period. Investing in prevention and intervention efforts for mental health during pregnancy may have long-term benefits for the well-being of women [39].

There are also several potential benefits to participating in a mindfulness program individually, especially during pregnancy. Participants can choose when and where to practice, making it easier to incorporate mindfulness into a busy schedule or daily routine. Practicing mindfulness alone can also help women concentrate more deeply on their thoughts, feelings, and sensations without distractions from group dynamics, encouraging deeper self-reflection and self-awareness, fostering a better understanding of personal experiences and emotions. For those who may feel anxious in group settings, individual participation can provide a more comfortable environment to explore mindfulness. Individuals can also select specific practices or techniques that resonate with them, creating a more personalized experience; this can empower participants to develop their mindfulness skills independently, fostering a sense of ownership over their mental well-being.

There are several digital apps that are based on MBCP principles. For example, Mindful Mama (Mindful Mamas Club) [41], an app that provides resources and guided meditations designed to support mothers during pregnancy and parenting, incorporating mindfulness techniques, or Mind the Bump (Smiling Mind) [42], which offers mindfulness practices, educational resources, and a community support feature for parents-to-be, integrating MBCP concepts. None of these involve the use of an online coach.

Our Study

MBCP intervention, which has already been validated and tested, will be fully available to users through digital tools. In particular, it will be delivered by a mobile app and guided by an online assistant, MAIA (developed by FBK). This research aims to assess the prototype of the MAIA app. This paper describes the prototype technology-based mindfulness intervention's design and development process for pregnant women, including the exploration phase, intervention content development, and iterative software development (including design, development, and formative evaluation of paper- and low-fidelity prototypes).

Methods

Overview

The design and development process encompassed a multidisciplinary approach and continuous, systematic evaluation throughout, as recommended in the Center for eHealth Research and Disease Management comprehensive roadmap approach to improve the uptake and impact of eHealth technologies [43].

The study's principal investigator (SR), a researcher and psychologist, led the intervention development work. During the design and development phase, the multidisciplinary project team, consisting of psychologists, mindfulness experts, communication experts, and information technology developers, had biweekly meetings. User-centered and service design methodologies were applied to ensure user involvement throughout the design and development process.

The mindfulness intervention was developed in iterative processes, following the ORBIT model [44], as shown in Figure 1, through a combination of (1) intervention content development (identified and adjusted from the MBCP program) and (2) iterative software development of a prototypes and formative evaluation. The development and iterative processes of the mindfulness intervention are shown in Figure 2.

XSL•FO
RenderX

Figure 1. The ORBIT model.

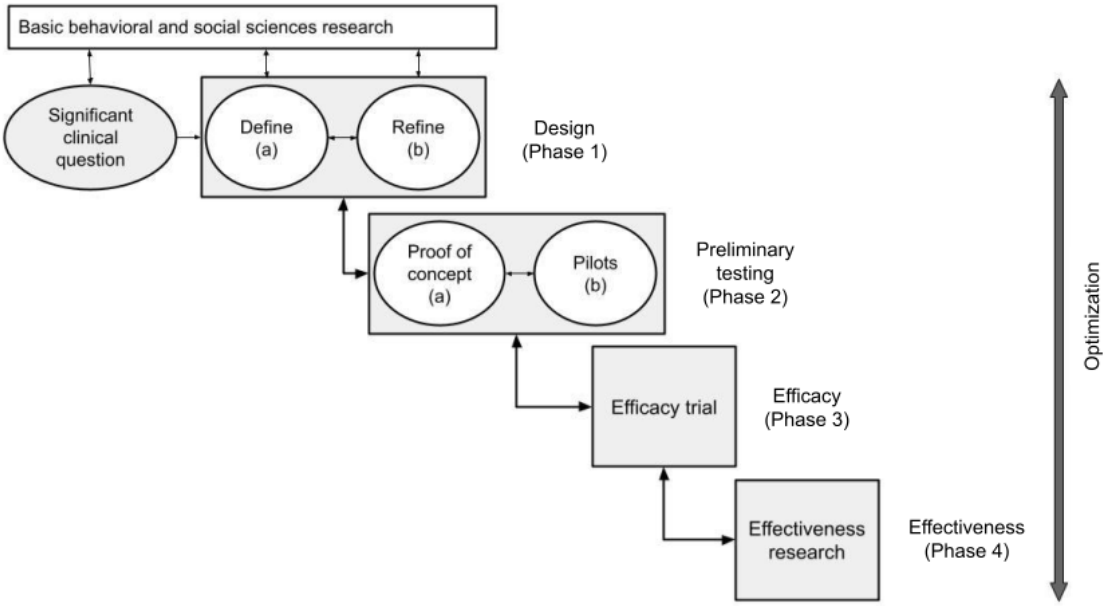
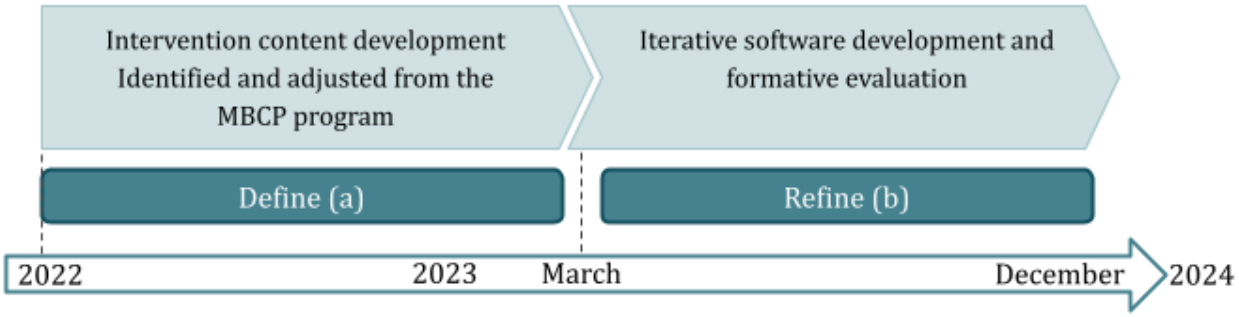


Figure 2. Development and iterative processes of the mindfulness intervention (Phase 1 of the ORBIT model—Design). MBCP: Mindfulness-Based Childbirth and Parenting.



Intervention Content Development

Adapting the protocol into a digital format involved an initial phase of studying the literature and the protocol itself. Once the main themes being presented week by week were identified, a team of psychologists with communication skills worked to adapt the content into an individual digital format delivered by chatbot. First, the team developed the online coach's script and then multimedia content (videos, audio, and images) useful for delivering critical information in a way that is easier for women to understand than just text format.

The initial intervention content for this study was first developed by the primary investigator, then adapted and tailored by the entire research team through iterative processes to fit an 8-session-based intervention in electronic format through text, audio, video, and images. The intervention was user-tested to make the content easily accessible, confirm adaptation to an app format, and ensure that the scientific foundation for the intervention was intact. Table 1 lists and briefly describes the 8 sessions and the topics covered in each session.

The intervention was constructed in such a way that MAIA is always the first to interact with the user. It is currently not possible for the woman to ask specific questions. The chatbot provides information but also asks questions to help the user in self-reflection. The user's response text is not analyzed at this stage of the study as it is not useful for the purpose of the study.

XSL•FO
RenderX

Table 1. MAIA intervention overview of sessions (8 sessions preceded by an introductory one) and their content divided into formal practices, informal practices, and readings or homework to be carried out independently between sessions.

Session	Formal practices	Informal practices	Readings or homework
S0	• Welcome	• Mindfulness definition. Formal practice and informal practice. • Rituality, posture, and intentionality.	Preparation of materials and space
S1	• Awareness of breath • Raisin meditation	• Being with the child	—[a]
S2	• Body scan • Awareness of breath	• Mindful eating	"The Body scan"
S3	• Body scan • Awareness of breath	• Mindfulness in daily life (brushing your teeth, dishes, cooking, and so on)	"Mindfulness in daily life," "The pain of childbirth," enjoyable events calendar
S4	• Body scan • Awareness of breath • Mindfulness meditation-related pain	• Informal pain practice	Calendar of unpleasant events, "Mindfulness practices for being with pain"
S5	• Body scan • Awareness of breath • Mindfulness meditation-related pain • Mindful touch-care meditation	• Physical activity and informal practice, being with the child	—
S6	• Body scan • Sitting meditation • Walking meditation	—	BRANN[b] (make informed decisions) and Mindfulness day planning
S7[c]	• Body scan • Sitting meditation • Loving-kindness meditation	• Inquiry families of origin	—
S8	• Personal choice (choose one or more meditation) • Loving-kindness meditation	• Continue informal practices	"Breastfeeding" and Mindfulness slogans, tips for continuing on your own

[a]Not applicable.

[b]BRANN: Benefits, Risk, Alternatives, Nothing, Now.

[c]A day of mindfulness practice was held between S6 and S7.

Iterative Software Development and Formative Evaluation

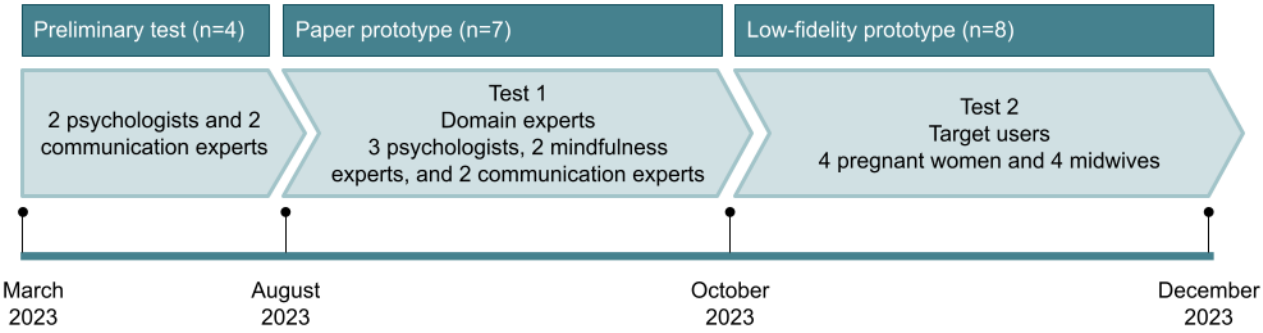
Adapting the MBCP interventions by implementing the MAIA chatbot provides a novel approach to delivering psychological support. Users can engage in a comprehensive and effective intervention through personalized sessions. Specifically, in the initial session, users are asked to select 2 days of the week on which they want to be contacted by MAIA to conduct mindfulness sessions, along with their preferred time slot (morning, afternoon, or evening). In addition, on the basis of some answers given, MAIA provided them with customized information (eg, MAIA asks whether or not the user wants to do a review of the previous session, whether they prefer to do a certain mindfulness practice now or later, and so on and on the basis of the answers the dialogue tree takes different directions). Therefore, considering different sections of the app that could be integrated was critical to supporting user engagement, adherence, and overall well-being.

In the first iteration, in spring 2023, 2 psychologists and 2 communication experts tested preliminary. They gave feedback on the prototype to ensure that the intervention program was logically built.

At this co-design stage, the involvement of experienced professionals, in addition to end users, is of paramount importance. For this reason, 3 other psychologists, 2 mindfulness experts, and 2 communication experts were involved in a later study (test 1) by creating a paper prototype. Later, after the development of a low-fidelity prototype, end users (4 women and 4 clinicians) were also involved (test 2). Thus, a multistage evaluation of the MAIA prototype was conducted between August and December 2023, involving 15 participants, as elaborated in Figure 3.

XSL•FO
RenderX

Figure 3. Timeline of software development and formative evaluation specifying the people involved (N=19).



Identification of Variables

To gather the necessary information from tests 1 and 2 with the paper prototype and the low-fidelity prototype, key variables were identified for investigation—communication, session structure, materials, and content. Given the study's objectives, a list of subvariables was chosen to be created ad hoc for the research. Table 2 shows the chosen variables and their respective subvariables. It was also decided not to ask each person everything. Still, based on the domain, a grid was created to determine which subvariables to investigate with the people involved.

Table 2. List of variables and subvariables investigated and people involved in evaluating individual variables.

Variables and subvariables	Psychologists	Mindfulness experts	Communication experts	Target	Midwives
Communication					
Empathy and listening	✓	✓	✓	✓	✓
Smoothness and fluidity	✓	✓	✓	✓	✓
Chatbot interaction	✓	—[a]	✓	✓	✓
Lexicon	✓	✓	✓	✓	✓
Session structure					
Interaction length	✓	✓	✓	✓	✓
Notifications and reminders	✓	—	✓	✓	—
external links	—	—	✓	✓	✓
Materials					
Diary	✓	✓	✓	✓	✓
Audio tracks	✓	✓	✓	✓	✓
Infographics and videos	✓	✓	✓	✓	✓
Content					
Adherence to MBCP[b] protocol	—	✓	—	—	—
content evaluation	✓	✓	—	✓	✓
content clarity	—	✓	✓	✓	✓

[a]Not applicable.

[b]MBCP: Mindfulness-Based Childbirth and Parenting.

These variables were assessed through the semantic differential tool [45] and a semistructured interview. The semantic differential is an instrument consisting of a series of scales, each composed of a pair of bipolar adjectives between which a rating scale (generally 5- or 7-position) is placed.

Interviews, instead, are suitable for a more in-depth investigation of users' attitudes and preferences toward new technological solutions since open-ended discussions with users can help researchers better understand the issues and concerns related to the possible future adoption of these solutions.

Iterative Development and Prototypes

Based on needed content adjustments, the first prototype version of the software was developed. This initial paper prototype consisted of the scripted dialogue of the chatbot and screens presenting the different sections of the app. All professionals

XSL•FO
RenderX

involved were assigned specific days of the mindfulness sessions to view. They were then asked to complete an online questionnaire based on the semantic differential. In this study, we used a 4-point rating instead of 5 or 7 points to prevent neutral feedback allocation. The rating scale displayed semantically opposing adjectives with negative connotations on the left and positive connotations on the right, representing the extremes. The pairs of adjectives chosen explicitly for this study correspond to 1 of the 4 macrovariables under investigation. Participants were tasked with indicating, for each adjective pair, which adjective better characterized the specified variable related to the chatbot. For instance, if the adjective pair is "passive listening-active listening," associated with the "communication" variable, participants would choose point 1 or 2 if they perceive the chatbot's listening as passive and point 3 or 4 if they perceive it as more active. Following the questionnaire, a short semistructured interview was conducted to investigate participants' impressions. The interviewer posed specific questions, and then an opportunity was given for the interviewee to share additional open reflections. To facilitate subsequent analyses, the interview was recorded after the participant's direct consent was obtained. Multimedia Appendix 2 shows the list of topics and questions posed during the interviews.

To recruit participants, we adopted criterion-based sampling. Specifically, participants had to be psychologists, mindfulness instructors, or communication experts. Participants were recruited through instant messaging tools (WhatsApp [Meta]) and email.

In the first iteration, psychologists, mindfulness, and communication experts tested and gave feedback on the prototype to ensure that the intervention program was logically built. Participants were given oral and written information about the study and provided written informed consent before user testing.

At this stage, the Personas methodology was also found useful. Personas is a user-centered and service design method used to create and visualize fictional representations of the target group [46]. Personas is an effective method for all project team members to understand the target group for which the app is built better. The personas in this study contained information about the pregnancy background, challenges, and technology use. Refer to Multimedia Appendix 3 for illustrated examples of study Personas. Participants had to identify with one of these Personas before the reading.

The data obtained from the administered questionnaires underwent aggregated analysis. Therefore, participants' evaluations on a 4-point scale for each adjective pair were summed. Once summed, the arithmetic mean for each adjective pair's evaluation was calculated. After obtaining all the average scores, a dot chart was generated using Datawrapper (Datawrapper GmbH), a web-based charting and mapping tool [47], to observe the overall trend of the scores better and, consequently, the semantic pole toward which they tend to orient.

The information obtained during the interviews has been transcribed for a more detailed analysis. The respondents' answers were then categorized by topics within the 4 initially identified variables—communication, session structure, materials, and content. In addition to these variables, a new theme emerged concerning the "postintervention" phase. The categorized responses within the variables were further divided into their respective subvariables (for the complete list, refer to Table 2). Finally, they were classified into positive and negative aspects or suggestions (areas for improvement). By "suggestions," we refer to those aspects of the app that might need modification in a forward-looking perspective.

Following minor adjustments, the paper prototype was transformed into an electronic tool to simulate the chatbot and app concept. The intervention was delivered through a software platform called TreC-PerseO (FBK), which supports the delivery of digital behavior change interventions. This proprietary software architecture is designed to ensure privacy and certified workflows in a production environment while also facilitating prototyping by adapting to the steps outlined in the chosen methodology. In this phase, Telegram (Telegram Messenger Inc) chat was used to deliver the intervention (to validate the content without the need to collect personal user data). During this second iteration (test 2), the final version of the low-fidelity prototype was tested by 2 pregnant women, 2 mothers, and 4 midwives. The participants were assigned different days of the mindfulness sessions to review, after which they completed an online questionnaire based on the semantic differential. A short interview followed the questionnaire.

To recruit participants, we adopted criteria-based sampling. Specifically, participants had to meet the following criteria: be pregnant or mothers with children aged less than 1 year or be midwives. Participants were recruited through instant messaging tools (WhatsApp) and email. The sample collected was 8 participants—2 pregnant women, 2 mothers, and 4 midwives. Participants were given oral and written information about the study and provided written informed consent before user testing.

The data collected during the interviews has been transcribed for a more detailed analysis. The respondents' answers were then categorized by topics within the 4 initially identified variables. The categorized responses within the variables were further divided into their respective subvariables, and then they were classified into positive and negative aspects or suggestions (areas for improvement). The resulting participant's feedback data were again used to evaluate, refine, iteratively adjust, and upgrade the prototype.

Ethical Considerations

At the time of enrollment, all participants who freely decided to take part in the study were asked to sign an informed consent form after a careful explanation of the study, its aims, the level of involvement required, and the duration of the research, as well as all ethical issues concerning confidentiality. All data collected were anonymized before analysis.

The participants will then be informed of the study's results through the dissemination of articles.

In line with Italian Law No. 3/2018, ethical review is mandatory for clinical trials, therapeutic procedures, or studies involving health risks or sensitive data. This study, being nonclinical and

XSL•FO
RenderX

focused on prototype development, did not meet these criteria, justifying the absence of ethics committee involvement.

No compensation was provided for participants.

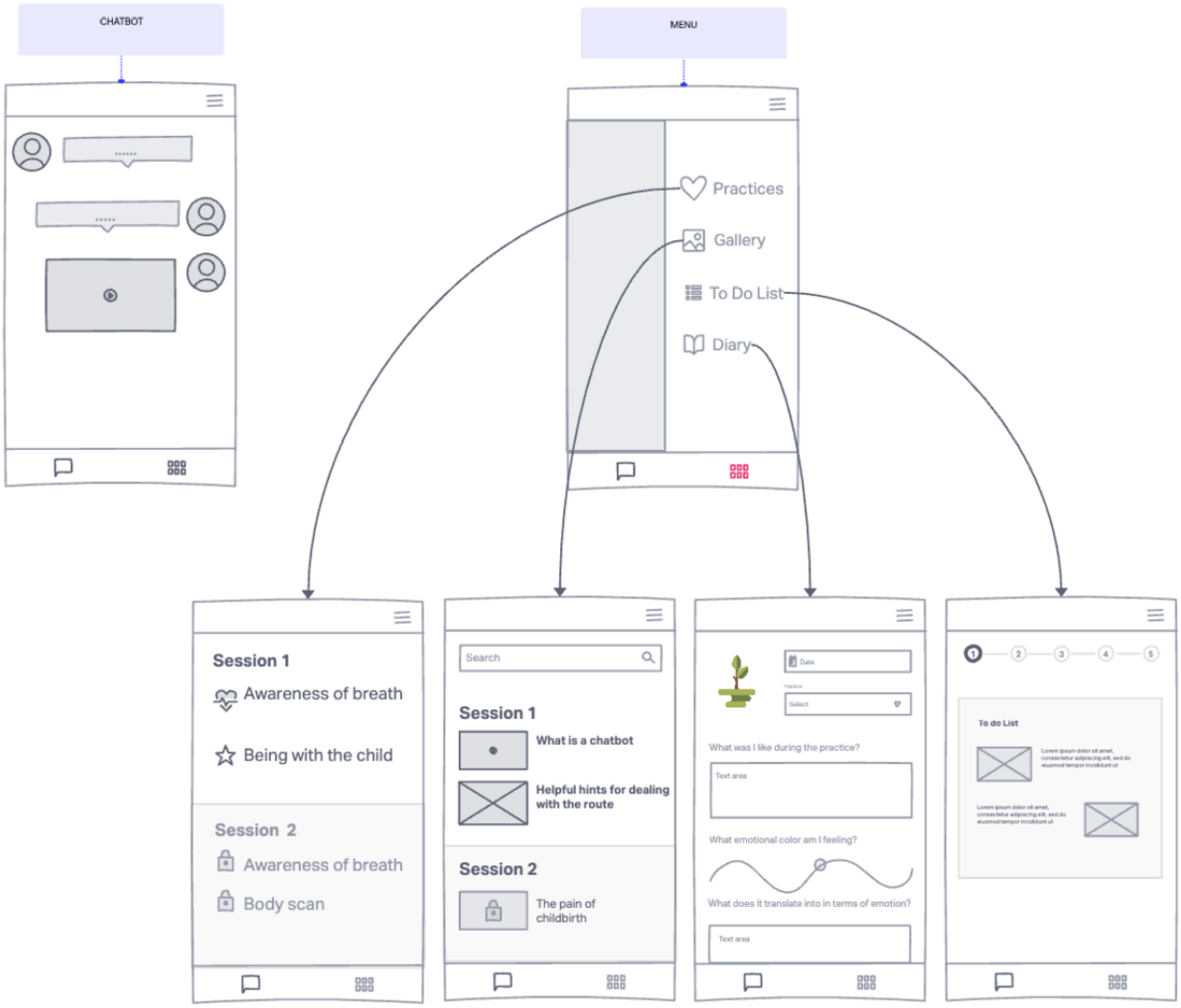
Results

Intervention Content Development

The iterative content development processes were parallel to app programming, and adjustments were made based on usability testing. Therefore, the app is designed to include 5 sections: the chatbot, Practices, Gallery, To Do List, and Diary.

Figure 4 shows some mockups of the app with its different sections.

Figure 4. Mockups of the MAIA app with its different sections.



MAIA assigns homework during the session that the users should carry out during the subsequent days, so the app Diary section will support self-monitoring. We have also provided for a tree that grows as the woman progresses. All the exercises women need to do during the week are listed in the To-Do List section to help her remember what she needs to do. All the practices, divided week per week, are shown in the Practices section. The Gallery page features all multimedia material delivered by MAIA, such as educational videos, images, and intervention introductions.

Iterative Software Development and Formative Evaluation

Iterative Development and Preliminary Test

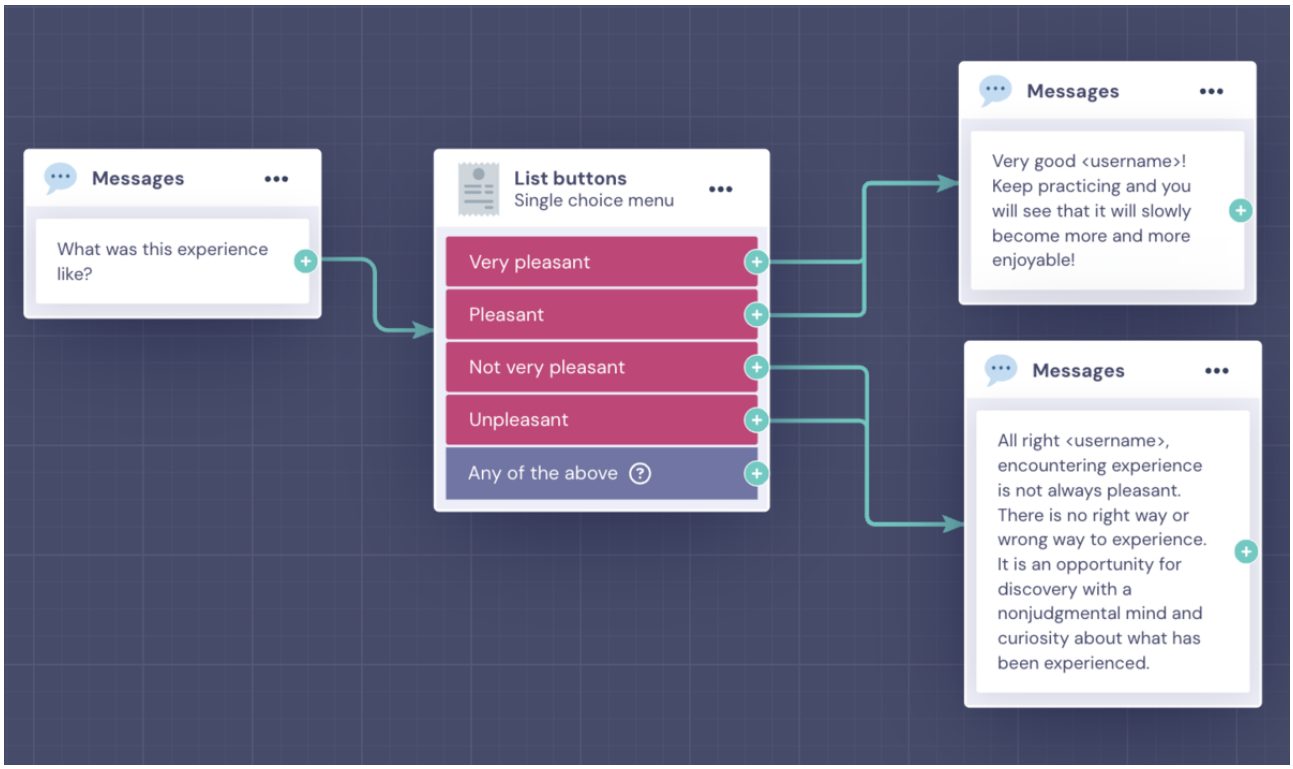
A total of 2 psychologists and 2 communication experts took part in this preliminary test phase. The objective of this evaluative phase was to test, through role-playing, the communicative style of dialogues—empathetic, welcoming, fluent, and smooth.

Therefore, it was possible to evaluate and revise MAIA's communication style through role-playing.

XSL•FO
RenderX

Regarding fluency and smoothness, the evaluation was primarily positive. To make the protocol more and more realistic, the dialogues were revised in some parts. Changes were made grammatically and syntactically to make the text more fluent and read smoothly. The changes also considered empathy. This makes it possible to create a feeling of trust in the relationship. Through empathic communication, the person can feel in a safe space, within which he or she is reassured and protected. Initially, the feeling while reading the dialogues could have been more apathetic communication. Therefore, multiple-choice questions were included about the individual practice (eg, How was this experience?) and the general course of action (eg, How did you do with mindfulness practices last time?). This allows MAIA to respond appropriately and more accurately based on the user's experience. Some examples are illustrated in Figure 5.

Figure 5. Examples of new multiple-choice questions included after preliminary tests.



A new practice not provided for in the MBCP initiate protocol, namely the practice of 3 breaths before each session, was also included. The user receiving notification from MAIA can be anywhere and under any conditions. To help the person focus on the present moment and be ready for the practices offered later, it was decided to provide this new initial practice at each session.

Finally, the correctness of the videos and images being sent was checked. For each video link, consistency was checked against the instructions, and the topic was addressed. Everything was correct, so no changes had to be made.

Iterative Development and Paper Prototype

After modifying some parts of the dialogue, as described above, the first paper prototype was developed. In this phase, the sample involved consisted of 7 professionals—3 psychologists, 2 mindfulness instructors, and 2 communication experts.

Using a semantic differential-based questionnaire facilitates observing the respondents' average positioning with respect to the 4 macrovariables under investigation.

As can be seen from Figure 6, a clear tendency toward the positive semantic pole (right pole) emerges in the feedback. This means that a significant portion of the chatbot-related feature pairs were evaluated in the positive rather than the negative sense. For instance, the vocabulary used by the chatbot was judged as understandable rather than abstruse. However, feedback related to 4 pairs of features occupied a middle position between the positive and negative poles, that is, passive listening–active listening, cold-warm, troublesome-welcome, and soporific-activator.

XSL•FO
RenderX

Figure 6. Graphical representation of mean values derived from semantic differential administered to domain experts (psychologists, mindfulness instructors, and communication experts).

In total, 2 psychologists (experts in communication) conducted the qualitative interview to gather additional information. Table 3 briefly highlights key positive and negative aspects that emerged during the domain experts' interviews.

In particular, the analysis of the interviews unearthed several notable observations that garnered attention due to their recurrent mention by more experts. Certain interviewees advocated augmenting written textual content with an auditory counterpart or substituting it entirely with auditory or audiovisual formats. The diary, initially conceived as a personal and confidential tool, prevents external control. However, the process of diary entry may elicit personal experiences, necessitating potential consultation with a designated professional if the need arises.

A further debated aspect is the inclusion of background music in mindfulness exercise tracks. According to certain interviewees, such music might prove distracting, diverting attention away from the individual's introspective focus. Conversely, others contend that it could serve as an aid, particularly in the absence of direct human interaction. In summation, these insights are deemed fundamental for the following stages of research, wherein the perspectives of pregnant women directly engaged in this study can be further investigated. In addition, we underscore the presence of numerous positive aspects delineated in the interviews, including the empathetic and welcoming communication style, the fluidity of discourse, and the use of imparted content, among others.

XSL•FO
RenderX

Table 3. Positive and negative aspects emerged from interviews with domain experts (psychologists, mindfulness instructors, and communication experts).

Variables and subvariables	Positive aspects	Negative aspects (and suggestions)
Communication		
Empathy and listening	• Empathic communication is comprehensive, conveys calm, reassuring, warm, welcoming, and puts at ease • Nice dialogue, relaxing, helpful • Customized	• Include more sentences that also address the child • Responses are repetitive and somewhat mechanical • "How did it go (the process)?" is too judgmental
Smoothness and fluidity	• Very smooth • Enjoyable to read	—[a]
Chatbot interaction	• If I read, I can go at the speed I want • It is not artificial, it works	• Needs to be more engaging • Reading can be tedious • Audio-only or audio and video or written and audio
Lexicon	• Accessible to everyone: intuitive, simple, straightforward, and understandable • Gentle, suitable, and consistent terms	• The term "discomfort" may cause concern • The term "pain" is repeated
Session structure		
Interaction length	• Not demanding • Self-management of dialogue times	—
Notifications and reminders	• Helpful in saving and reviewing it without entering the app	—
Materials		
Diary	• Putting into writing what is felt allows for processing and gaining clarity	• Option to print the diary and write it by hand
Audio tracks	• Useful content • "Being with the pain": very helpful for preparing for childbirth • Use of different voices	• Banal video graphics • Remove the music • Add music • Lack of initial welcoming moment • Option to choose a male or female voice
Infographics and videos	• Beautiful graphics, practical, and charming • Nice color used • Option to print PDFs	—
Content		
Adherence to MBCP[b] protocol	• Positive and negative experiences of other women • Liked the 8 attitudes	• Providing two options (same content presented differently) based on past pregnancy experiences
Content evaluation	• Interesting content • Practices dedicated to both mother and baby	• Having too much time to think is only sometimes positive
Content clarity	• Engaging protocol	—
Other	—	• Understand how it went postpartum in light of these weeks • Hold an in-person meeting at the end of the protocol to make it more engaging

[a]Not available.

[b]MBCP: Mindfulness-Based Childbirth and Parenting.

XSL•FO
RenderX

Iterative Development and Low-Fidelity Prototypes

In total, 2 pregnant women, 2 moms, and 4 midwives participated in this second test phase. To better understand the app and the entire intervention, more realistic mockups were created (Figure 7). This allowed users to represent the app in its final appearance better.

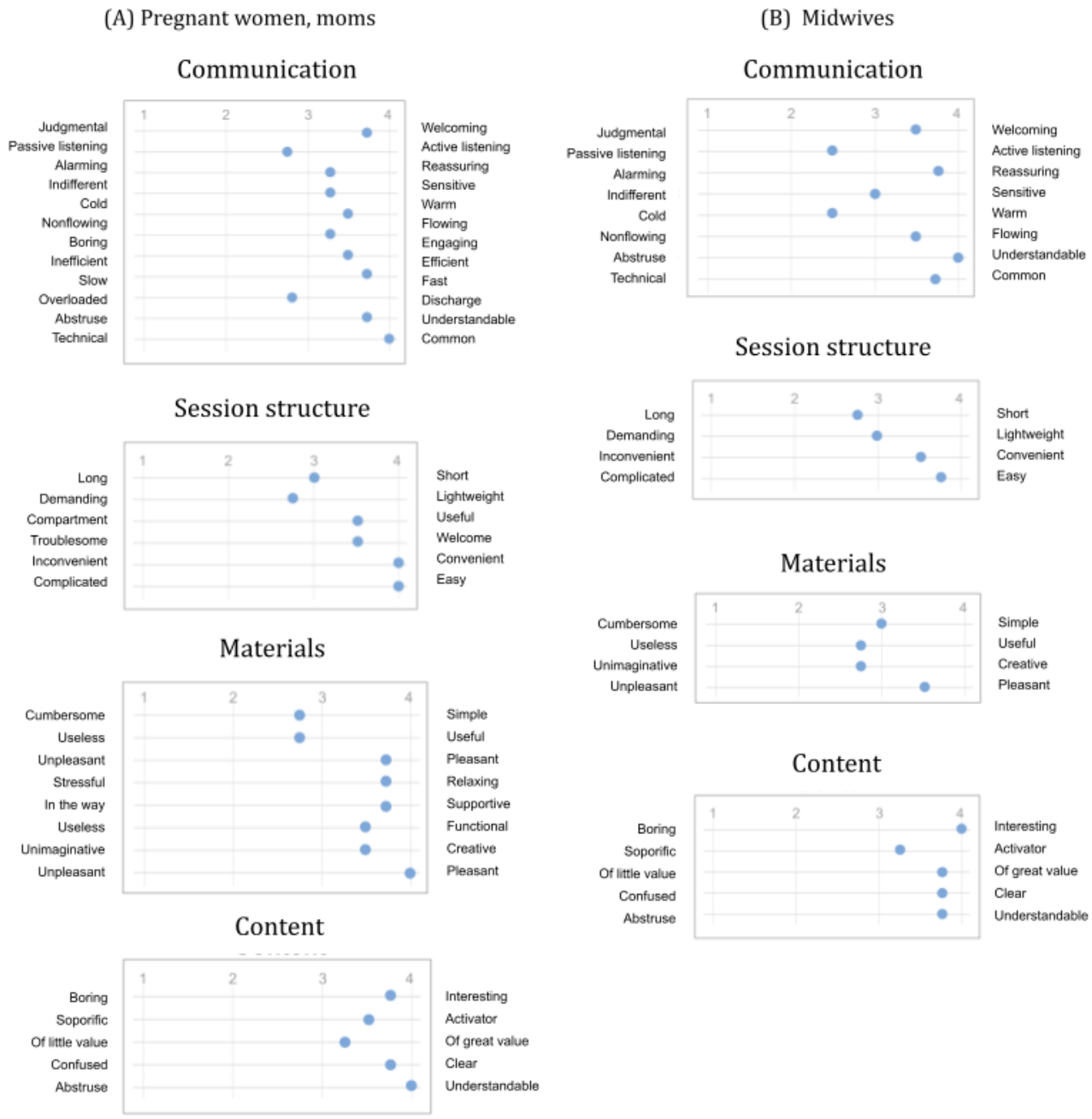
In this case, using a semantic differential-based questionnaire facilitates the observation of the respondents' average positioning with respect to the 4 variables under investigation.

As illustrated in Figure 8, a tendency toward the positive semantic pole (right pole) emerges in the feedback also in these groups of participants. However, feedback related to 5 pairs of features occupied a middle position between the positive and negative poles for the pregnant woman and moms' group and 5 for the midwives group.

Figure 7. More realistic mockups of the MAIA app with its different sections.

XSL•FO
RenderX

Figure 8. Graphical representation of the mean values obtained from administering the semantic differential to target users.



Furthermore, 2 psychologists (experts in communication) conducted the qualitative interview to gather additional information. Table 4 briefly highlights key positive and negative aspects that emerged during user interviews.

All respondents found the protocol content of great value, considering it interesting and useful. They appreciated the alternation of simple and streamlined proposals with more time-consuming ones. They consider it an added value to manage the timing of the course autonomously; with the app, they can postpone the performance of activities, reassured by the fact that they can catch up at another time. They also mostly consider the diary to be useful and usable as long as it is left as an option.

One of them suggested supplementing the stand-alone activities with a couple of face-to-face meetings (an initial one to explain the meaning and experience as a group and a central one for the more engaging exercises, such as the sultana and ice cube exercises).

The midwives dwelt more on the contents and found them correct and set out in a simple and comprehensible manner for all, giving useful advice to improve the video part in a couple of cases. One of them strongly urged the translation of the protocol into several languages.

Among the materials, the infographics were highly appreciated by both categories for their incisiveness and the fact that they were printable.

XSL•FO
RenderX

Table 4. Positive and negative aspects emerged from user interviews (pregnant women, moms, and midwives).

Variables and subvariables	Positive aspects	Negative aspects (and suggestions)
Communication		
Empathy and listening	• Nice dialogue • Communication puts at ease • Customized, nonjudgmental	• Responses are repetitive, somewhat mechanical • When the exercise does not work, it could be a question that goes into depth analyses and also advises for future exercises
Smoothness and fluidity	• Very smooth • Enjoyable to read	—[a]
Chatbot interaction	• It is not artificial, it works • It is fast and light, particularly when it proposes questions with multiple-choice answers	—
Lexicon	• Accessible to everyone: intuitive, simple, clear, and understandable	—
Session structure		
Interaction length	• Not demanding • Self-management of dialogue times	—
Notifications and reminders	• Useful even if you have no time at that moment	• Sometimes stressful because even at the scheduled time, you cannot do the exercises
Materials		
Diary	• Useful if optional • Interesting to focus on different exercises • Very useful to promote awareness	• It depends on the available time
Audio tracks	• Useful content • Good timing • Good voice	• Add music • Sometimes they are too long
Infographics and videos	• Beautiful graphics, effective, and charming • Very useful Infographics to focus on the most important information, and to repeat the exercise • Very useful images in the videos, especially for people with visual memory	• The pregnant woman is too thin • The eighth-week video on pain has a very heavy atmosphere and does not reflect the real situation in the delivery room (the bed is used very little, and the woman is never alone)
Content		
Content evaluation	• Interesting content • Engaging protocol • Very valuable content • The content on breastfeeding was very well done, it says the right things in the right way • The daily practical exercises are very interesting and engaging (eg, walking and teeth washing) • It is engaging because the commitment required is growing • The practice of three breaths is very useful and effective	—
Content clarity	• Very clear and simple to follow	• At first, it seems like the practice will always last 5 minutes

XSL•FO
RenderX

Variables and subvariables	Positive aspects	Negative aspects (and suggestions)
Other	• Very comfortable to have all you need in the app	• Desire to have other two weeks after the childbirth • Hold an in-person meeting at the beginning and at the end of the protocol to make it more engaging • It could be easier to follow the most demanding exercises in the team, with a guide

[a]Not available.

Discussion

Principal Findings

This study illustrates how user-centered and service designs can be applied to identify and incorporate essential stakeholder aspects in the design and development process. Combined with evidence-based concepts, this process facilitated the development of a mindfulness intervention designed for the end users, in this case, pregnant women.

Importing and adapting the intervention was a comprehensive and meticulous process. According to the ORBIT model, we examined a wide range of scientific studies on the efficacy of mindfulness during pregnancy and the postpartum period, focusing on the MBSR and the MBCP protocols. By integrating elements from both methodologies, we tailored key aspects of each protocol to create an intervention specifically designed for the needs of pregnant women. In addition, we considered other evidence and best practices in the field of mindfulness, ensuring the incorporation of empirically supported and culturally sensitive approaches (ie, the practice of the 3 breaths to anchor oneself in the present moment).

This integrative approach allowed us to develop a protocol that is scientifically grounded, practical, and suitable for the unique context of pregnancy.

Paying attention to the semantic differentials with lower values, we can see that in the case of listening, the answers of the users and midwives agree. In the interviews, 3 out of 4 women stated that they see the chatbot as the necessary tool to dispense material. Furthermore, 2 of them consider the dialogues to be valuable when it is fast and nondemanding (eg, when a multiple-choice answer is expected), which is also confirmed by the results on the overloaded or discharged differential. However, as seen from the other semantic differentials, the chatbot is welcoming, reassuring, and polite.

Consider the pair “passive listening–active listening,” referring to whether one feels actively or passively listened to by the chatbot; domain experts, on average, did not distinctly categorize the listening as active. This judgment may be due to the inherent limitation of digital apps—lacking a natural person on the other side of the screen responding to each user’s comment. The structure of this type of interaction, therefore, makes it difficult to feel fully heard, particularly in an active way.

One of the suggestions we received from users was to add audio in addition to the written text. This suggestion stems from sociocultural considerations. Indeed, nowadays, people increasingly use social networks, interacting with content that is short in duration and predominantly in the form of video rather than text. Reading may, therefore, appear more boring for some. Another focal point pertains to the diary feature, particularly the need to facilitate interaction with a designated professional.

Strengths, Limitations, and Future Directions

One of the main strengths of this study lies in the fact that it engaged both domain experts and target users from the very early stages. The suggestions that were gathered highlight different but equally important points of view for refining the intervention. Having midwives involved, in addition to pregnant women, once again allowed us to have an additional point of view. The midwives could give us suggestions (eg, editing a video so that it was more aligned with what is happening in the delivery room) that none of the other actors involved would have been able to provide.

One limitation, however, mainly concerns the development of the low-fidelity prototype and the number of users involved. Along with that prototype, we provided users with a mockup of the app, but interacting with such a prototype or with the final app is a very different experience. The purpose of this study, however, was not to evaluate the user experience with the app but to gather impressions of content, vocabulary, and communication style. For that reason, the tool used was deemed good enough. The next step will be to develop the application following the suggestions gleaned from this first study and then to set up a proof-of-concept study, following the orbit model, involving about 50 women to evaluate the acceptability, feasibility, and usability of the MAIA app [48].

Based on the data collected during this second study, we expect to have to make further modifications to the intervention and application. Following these modifications, we will finally be ready to field a randomized control trial to evaluate the efficacy of the intervention (the study protocol will be finalized by mid-2025). After conducting this randomized controlled trial, we also plan to conduct a real-world study in order to evaluate the effectiveness of the intervention and the app itself.

Conclusions

Developing a digital intervention is a complex procedure that requires work in different areas and involves different professionals. We aimed to build an intervention that was as suitable as possible for the target audience, both in terms of content and design. The contribution of specialists is undoubtedly an added value from the early stages in a coconstruction perspective. This survey is the first step in a

XSL•FO
RenderX

series of studies to validate the intervention. At this stage, it is still difficult to state how this research contributes to the broader evidence base supporting the use of mindfulness-based practices among pregnant women, as we are still at a preliminary stage. However, we can state that the impression of the users and experts involved is positive; as the results seem to be encouraging, the next step will be to conduct a study to assess the experience and engagement of pregnant women.

Acknowledgments

The research team would like to thank all the women involved, including mothers, midwives, and domain experts, who contributed to the success of this study in various capacities.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request. Before disclosure, data are aggregated and anonymized.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Brief description of the main topics covered week by week in the Mindfulness-Based Childbirth and Parenting original program.
[PNG File , 463 KB - formative_v9i1e58265_app1.png]

Multimedia Appendix 2
Questions posed to participants attending the evaluation.
[PNG File , 126 KB - formative_v9i1e58265_app2.png]

Multimedia Appendix 3
Examples of personas.
[PNG File , 80 KB - formative_v9i1e58265_app3.png]

References

1. Stern DN. The Motherhood Constellation. United Kingdom: Routledge; 2018. [doi: 10.4324/9780429482489-12]
2. Lutkiewicz K, Bieleninik Ł, Cieślak M, Bidzan M. Maternal-infant bonding and its relationships with maternal depressive symptoms, stress and anxiety in the early postpartum period in a Polish sample. Int J Environ Res Public Health 2020;17(15):5427 [FREE Full text] [doi: 10.3390/ijerph17155427] [Medline: 32731490]
3. Gavin NI, Gaynes BN, Lohr KN, Meltzer-Brody S, Gartlehner G, Swinson T. Perinatal depression: a systematic review of prevalence and incidence. Obstet Gynecol 2005;106(5 Pt 1):1071-1083. [doi: 10.1097/01.AOG.0000183597.31630.db] [Medline: 16260528]
4. Goodman JH, Guarino A, Chenausky K, Klein L, Prager J, Petersen R, et al. CALM pregnancy: results of a pilot study of mindfulness-based cognitive therapy for perinatal anxiety. Arch Womens Ment Health 2014;17(5):373-387 [FREE Full text] [doi: 10.1007/s00737-013-0402-7] [Medline: 24449191]
5. Dunkel Schetter C. Psychological science on pregnancy: stress processes, biopsychosocial models, and emerging research issues. Annu Rev Psychol 2011;62:531-558. [doi: 10.1146/annurev.psych.031809.130727] [Medline: 21126184]
6. Zarenejad M, Yazdkhasti M, Rahimzadeh M, Tourzani ZM, Esmaelzadeh-Saeieh S. The effect of mindfulness-based stress reduction on maternal anxiety and self-efficacy: a randomized controlled trial. Brain Behav 2020;10(4):e01561 [FREE Full text] [doi: 10.1002/brb3.1561] [Medline: 32162450]
7. Conde A, Figueiredo B, Tendais I, Teixeira C, Costa R, Pacheco A, et al. Mother's anxiety and depression and associated risk factors during early pregnancy: effects on fetal growth and activity at 20-22 weeks of gestation. J Psychosom Obstet Gynaecol 2010;31(2):70-82 [FREE Full text] [doi: 10.3109/01674821003681464] [Medline: 20236029]
8. Li H, Bowen A, Bowen R, Muhajarine N, Balbuena L. Mood instability, depression, and anxiety in pregnancy and adverse neonatal outcomes. BMC Pregnancy Childbirth 2021;21(1):583 [FREE Full text] [doi: 10.1186/s12884-021-04021-y] [Medline: 34429072]
9. Woods SM, Melville JL, Guo Y, Fan MY, Gavin AR. Psychosocial stress during pregnancy. Obstetric Anesthesia Digest 2010;30(4):237 [FREE Full text] [doi: 10.1097/01.aoa.0000389617.09252.47]
10. Fonseca A, Gorayeb R, Canavarro MC. Women s help-seeking behaviours for depressive symptoms during the perinatal period: socio-demographic and clinical correlates and perceived barriers to seeking professional help. Midwifery 2015;31(12):1177-1185. [doi: 10.1016/j.midw.2015.09.002] [Medline: 26433622]

XSL•FO
RenderX

11. Rilling JK. The neural and hormonal bases of human parental care. Neuropsychologia 2013;51(4):731-747. [doi: 10.1016/j.neuropsychologia.2012.12.017] [Medline: 23333868]
12. Barrera AZ, Wickham RE, Muñoz RF. Online prevention of postpartum depression for Spanish- and English-speaking pregnant women: a pilot randomized controlled trial. Internet Interv 2015;2(3):257-265 [FREE Full text] [doi: 10.1016/j.invent.2015.06.002] [Medline: 26273567]
13. Philippe TJ, Sikder N, Jackson A, Koblanski ME, Liow E, Pilarinos A, et al. Digital health interventions for delivery of mental health care: systematic and comprehensive meta-review. JMIR Ment Health 2022;9(5):e35159 [FREE Full text] [doi: 10.2196/35159] [Medline: 35551058]
14. Kabat-Zinn J. Full catastrophe living: using the wisdom of your body and mind to face stress, pain, and illness. New York City: New York Dell Publishing Co; 1990.
15. Kabat-Zinn J. Wherever You Go, There You Are. London, United Kingdom: Piatkus; 1994. [doi: 10.1037/a0012744]
16. Dhillon A, Sparkes E, Duarte RV. Mindfulness-based interventions during pregnancy: a systematic review and meta-analysis. Mindfulness (N Y) 2017;8(6):1421-1437 [FREE Full text] [doi: 10.1007/s12671-017-0726-x] [Medline: 29201244]
17. Sevilla-Llewellyn-Jones J, Santesteban-Echarri O, Pryor I, McGorry P, Alvarez-Jimenez M. Web-based mindfulness interventions for mental health treatment: systematic review and meta-analysis. JMIR Ment Health 2018;5(3):e10278 [FREE Full text] [doi: 10.2196/10278] [Medline: 30274963]
18. Lönnberg G, Jonas W, Unternaehrer E, Bränström R, Nissen E, Niemi M. Effects of a mindfulness based childbirth and parenting program on pregnant women's perceived stress and risk of perinatal depression-results from a randomized controlled trial. J Affect Disord 2020;262:133-142. [doi: 10.1016/j.jad.2019.10.048] [Medline: 31733457]
19. Vieten C, Astin J. Effects of a mindfulness-based intervention during pregnancy on prenatal stress and mood: results of a pilot study. Arch Womens Ment Health 2008;11(1):67-74. [doi: 10.1007/s00737-008-0214-3] [Medline: 18317710]
20. Pan WL, Chang CW, Chen SM, Gau ML. Assessing the effectiveness of mindfulness-based programs on mental health during pregnancy and early motherhood - a randomized control trial. BMC Pregnancy Childbirth 2019;19(1):346 [FREE Full text] [doi: 10.1186/s12884-019-2503-4] [Medline: 31601170]
21. Dayan J, Creveuil C, Herlicoviez M, Herbel C, Baranger E, Savoye C, et al. Role of anxiety and depression in the onset of spontaneous preterm labor. Am J Epidemiol 2002;155(4):293-301. [doi: 10.1093/aje/155.4.293] [Medline: 11836191]
22. Creswell JD. Mindfulness interventions. Annu Rev Psychol 2017;68:491-516. [doi: 10.1146/annurev-psych-042716-051139] [Medline: 27687118]
23. Zhang D, Tsang KWK, Duncan LG, Yip BHK, Chan DCC, Lee EKP, et al. Effects of the Mindfulness-Based Childbirth and Parenting (MBCP) program among pregnant women: a randomized controlled trial. Mindfulness 2023;14(1):50-65 [FREE Full text] [doi: 10.1007/s12671-022-02046-8]
24. Duncan LG, Shaddix C. Mindfulness-Based Childbirth and Parenting (MBCP): innovation in birth preparation to support healthy, happy families. Int J Birth Parent Educ 2015;2(2):30-33. [Medline: 29051821]
25. Van der Meulen RT, Veringa-Skiba I, Van Steensel FJA, Bögels SM, De Bruin EI. Mindfulness-Based Childbirth and Parenting for pregnant women with high fear of childbirth and their partners: outcomes of a randomized controlled trial assessing short- and longer-term effects on psychological well-being, birth and pregnancy experience. Midwifery 2023;116:103545 [FREE Full text] [doi: 10.1016/j.midw.2022.103545] [Medline: 36375411]
26. Abras C, Maloney-Krichmea D, Preece J. User-centered design. In: Bainbridge W, editor. Berkshire Encyclopedia of Human-Computer Interaction When Science Fiction Becomes Science Fact. Great Barrington, Massachusetts: Berkshire Publishing Group; 2004:757-768.
27. Dabbs ADV, Myers BA, Mc Curry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
28. Gulliksen J, Göransson B, Boivie I, Blomkvist S, Persson J, Cajander A. Key principles for user-centred systems design. Behaviour & Information Technology 2003;22(6):397-409 [FREE Full text] [doi: 10.1080/01449290310001624329]
29. Sanders EBN, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008;4(1):5-18 [FREE Full text] [doi: 10.1080/15710880701875068]
30. Stickdorn M, Schneider J. This is Service Design Thinking: Basics, Tools, Cases. Amsterdam: Bis Publishers; 2010.
31. Bardacke N. Mindfulness in Gravidanza. Perignano, Tuscany: Edra; 2023.
32. Duncan LG, Cohn MA, Chao MT, Cook JG, Riccobono J, Bardacke N. Benefits of preparing for childbirth with mindfulness training: a randomized controlled trial with active comparison. BMC Pregnancy Childbirth 2017;17(1):140 [FREE Full text] [doi: 10.1186/s12884-017-1319-3] [Medline: 28499376]
33. Warriner S, Crane C, Dymond M, Krusche A. An evaluation of mindfulness-based childbirth and parenting courses for pregnant women and prospective fathers/partners within the UK NHS (MBCP-4-NHS). Midwifery 2018;64:1-10. [doi: 10.1016/j.midw.2018.05.004] [Medline: 29843001]
34. Sbrilli MD, Duncan LG, Laurent HK. Effects of prenatal mindfulness-based childbirth education on child-bearers' trajectories of distress: a randomized control trial. BMC Pregnancy Childbirth 2020;20(1):623 [FREE Full text] [doi: 10.1186/s12884-020-03318-8] [Medline: 33059638]

XSL•FO
RenderX

35. Veringa-Skiba IK, de Bruin EI, van Steensel FJA, Bögels SM. Fear of childbirth, nonurgent obstetric interventions, and newborn outcomes: a randomized controlled trial comparing mindfulness-based childbirth and parenting with enhanced care as usual. Birth 2022;49(1):40-51 [FREE Full text] [doi: 10.1111/birt.12571] [Medline: 34250636]
36. Lönnberg G, Niemi M, Salomonsson B, Bränström R, Nissen E, Jonas W. Exploring the effects of mindfulness-based childbirth and parenting on infant social-emotional development. Mindfulness 2021;12(8):2009-2020 [FREE Full text] [doi: 10.1007/s12671-021-01658-w]
37. Kantrowitz-Gordon I, Abbott S, Hoehn R. Experiences of postpartum women after mindfulness childbirth classes: a qualitative study. J Midwifery Womens Health 2018;63(4):462-469. [doi: 10.1111/jmwh.12734] [Medline: 29924917]
38. Lönnberg G, Niemi M, Salomonsson B, Bränström R, Nissen E, Jonas W. Correction to: exploring the effects of mindfulness-based childbirth and parenting on infant social-emotional development. Mindfulness 2021;12(8):2021 [FREE Full text] [doi: 10.1007/s12671-021-01669-7]
39. Roubinov DS, Epel ES, Coccia M, Coleman-Phox K, Vieten C, Adler NE, et al. Long-term effects of a prenatal mindfulness intervention on depressive symptoms in a diverse sample of women. J Consult Clin Psychol 2022;90(12):942-949 [FREE Full text] [doi: 10.1037/ccp0000776] [Medline: 36441994]
40. Resources. Mindful Birthing and Parenting. URL: https://www.mindfulbirthing.org/resources [accessed 2023-11-29]
41. Self-care and mindfulness for moms. Mindful Mamas. URL: https://mindfulmamasclub.com/ [accessed 2024-10-14]
42. Mind the Bump. URL: https://www.mindthebump.org.au/ [accessed 2024-12-13]
43. van Gemert-Pijnen JEWC, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
44. Czajkowski SM, Powell LH, Adler N, Naar-King S, Reynolds KD, Hunter CM, et al. From ideas to efficacy: the ORBIT model for developing behavioral treatments for chronic diseases. Health Psychol 2015;34(10):971-982 [FREE Full text] [doi: 10.1037/hea0000161] [Medline: 25642841]
45. Charles EO, Suci GJ, Tannenbaum PH. The Measurement of Meaning. United States: University Of Illinois Press; 1957.
46. Adlin T, Pruitt J. The Essential Persona Lifecycle: Your Guide to Building and Using Personas. Cambridge, Massachusetts, United States: Morgan Kaufmann; 2010. [doi: 10.1016/b978-0-12-381418-0.00006-1]
47. Create charts and maps with Datawrapper. Datawrapper. URL: https://www.datawrapper.de/ [accessed 2019-01-30]
48. Rizzi S, Poggianella S, Pavesi M, Gios L, Bincoletto G, Scolari I, et al. A mindfulness-based app intervention for pregnant women: protocol for a pilot feasibility study. JMIR Res Protoc 2024;13:e53890 [FREE Full text] [doi: 10.2196/53890] [Medline: 38567964]

Abbreviations

MBCP: Mindfulness-Based Childbirth and Parenting
MBSR: Mindfulness-Based Stress Reduction

Edited by A Mavragani; submitted 26.03.24; peer-reviewed by L Paulos-Guarnieri, CME Fletcher; comments to author 05.10.24; revised version received 21.10.24; accepted 18.11.24; published 17.01.25.

Please cite as:
Rizzi S, Pavesi MC, Moser A, Paolazzi F, Marchesoni M, Poggianella S, Gadotti E, Forti S
A Mindfulness-Based App Intervention for Pregnant Women: Qualitative Evaluation of a Prototype Using Multiple Case Studies
JMIR Form Res 2025;9:e58265
URL: https://formative.jmir.org/2025/1/e58265
doi:10.2196/58265
PMID:

©Silvia Rizzi, Maria Chiara Pavesi, Alessia Moser, Francesca Paolazzi, Michele Marchesoni, Stefania Poggianella, Erik Gadotti, Stefano Forti. Originally published in JMIR Formative Research (https://formative.jmir.org), 17.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Reflections of Foster Youth Engaging in the Co-Design of Digital Mental Health Technology: Duoethnography Study

Ifunanya Ezimora[1*], BA; Tylia Lundberg[1*], BA; Dylan Miars[1], BA; Jeruel Trujeque[1]; Ashley Papias[1], AA; Margareth V Del Cid[1], PhD; Johanna B Folk[1], PhD; Marina Tolou-Shams[1], PhD

Department of Psychiatry and Behavioral Sciences, University of California, San Francisco, San Francisco, CA, United States

[*]these authors contributed equally

Corresponding Author:
Johanna B Folk, PhD
Department of Psychiatry and Behavioral Sciences
University of California, San Francisco
1001 Potrero Avenue
San Francisco, CA, 94110
United States
Phone: 1 4156029521
Email: johanna.folk@ucsf.edu

Abstract

Background: Current research on digital applications to support the mental health and well-being of foster youth is limited to theoretical applications for transition-aged foster youth and support platforms developed without intentional input from foster youth themselves. Centering the lived expertise of foster youth in digital solutions is crucial to dismantling barriers to care, leading to an increase in service access and improving mental health outcomes. Co-design centers the intended end users during the design process, creating a direct relationship between potential users and developers. This methodology holds promise for creating tools centered on foster youth, yet little is known about the co-design experience for foster youth. Understanding foster youth's experience with co-design is crucial to identifying best practices, knowledge of which is currently limited.

Objective: The aim of this paper is to reflect on the experiences of 4 foster youth involved in the co-design of FostrSpace, a mobile app designed through a collaboration among foster youth in the San Francisco Bay Area; clinicians and academics from the Juvenile Justice Behavioral Health research team at the University of California, San Francisco; and Chorus Innovations, a rapid technology development platform specializing in participatory design practices. Key recommendations for co-designing with foster youth were generated with reference to these reflections.

Methods: A duoethnography study was conducted over a 1-month period with the 4 transition-aged former foster youth co-designers of FostrSpace via written reflections and a single in-person roundtable discussion. Reflections were coded and analyzed via reflexive thematic analysis.

Results: In total, 4 main themes were identified from coding of the duoethnography reflections: power and control, resource navigation, building community and safe spaces, and identity. Themes of power and control and resource navigation highlighted the challenges FostrSpace co-designers experienced trying to access basic needs, support from caregivers, and mental health resources as foster youth and former foster youth. Discussions pertaining to building community and safe spaces highlighted the positive effect of foster youth communities on co-designers, and discussions related to identity revealed the complexities associated with understanding and embracing foster youth identity.

Conclusions: This duoethnography study highlights the importance of centering the lived expertise of co-designers throughout the app development process. As the digital health field increasingly shifts toward using co-design methods to develop digital mental health technologies for underserved youth populations, we offer recommendations for researchers seeking to ethically and effectively engage youth co-designers. Actively reflecting throughout the co-design process, finding creative ways to engage in power-sharing practices to build community, and ensuring mutual benefit among co-designers are some of the recommended core components to address when co-designing behavioral health technologies for youth.

(JMIR Form Res 2025;9:e53231) doi:10.2196/53231

XSL•FO
RenderX

KEYWORDS

foster youth; digital health technology; co-design; app development; mental health; adolescent; young adult; mobile health; mHealth; foster care; duoethnography

Introduction

Background

Over 391,000 youth are in foster care in the United States, with behavioral health being their largest unmet need [1]. Foster youth are young people removed from their caregivers due to safety risks or unforeseen circumstances (eg, death of caregivers) and placed in out-of-home care that aims to be safe and stable. Out-of-home placement, such as group homes and placement with a foster parent, can precede reunification with their original caregiver, permanent placement with a legal guardian or adoptive family, or other planned permanency arrangements [2]. Contact with the foster care system at key periods of development and abrupt placement with new caregivers can contribute to foster youth experiencing adverse behavioral health outcomes, increased risk of not having an accountable caregiver, and insufficient access to services that support their well-being [3]. Changes in placement can result in service disruption, further exacerbating unmet mental health needs. However, digital health technology may provide an intermediary behavioral health support solution while youth transition between placements or act as a longer-term support that can go with them anywhere from placement to placement. There is a need to target widespread inequities in behavioral health service access and involve foster youth voices in the development of new resources. Currently, there is a gap in the literature on the development, deployment, and evaluation of mobile apps designed for foster youth. Foster youth have been excluded from conversations regarding tools designed to support their well-being despite their lived expertise being crucial in targeting barriers that drive adverse mental health outcomes and underuse of services [3]. Co-design holds promise for creating youth-centered tools that are responsive to the needs of specific communities such as foster youth. Despite the potential benefits, little is known about how best to facilitate co-design with youth from underserved groups, such as foster youth, or their experiences working on co-design teams.

Co-Designing Mobile Apps

Interventions to address mental health and resource needs have increasingly shifted toward digital health technologies. Digital health methods hold great potential for addressing gaps in services for underserved communities, and using co-design methods may better meet the needs of these communities [4]. Co-design processes create a direct relationship between potential users and designers, where intended users from target communities actively participate in designing technologies for their community [4,5]. Co-design centers users in the design process and lies at one end of the user-centered design spectrum. Co-design processes with young people are meant to center youth involvement throughout the design process, but many technologies claiming to use co-design processes with youth use the term *co-design* to reference a continuum of user involvement ranging from minimal youth involvement at the level of consultation to more integrated involvement through mutual partnership (true co-design) [4].

Recent co-designed applications include mood-monitoring applications for youth experiencing depression; integrated digital crisis planning tools for youth; personal informatics apps; and open access, anonymous platforms with evidence-based, self-guided interventions for youth mental health [6-9]. Co-designed digital health technologies for both youth and adults have used human-centered co-design with studio methodology (ie, an iterative process of young people sketching out an idealized app and its features), adolescent-led and initiated co-design processes at a higher level in the "ladder of participation," and participatory technology development incorporating community-partnered research principles with technology platforms (eg, Chorus) [6,8,10,11].

Researchers have begun designing digital health technologies with greater intentionality in their work with young people by including them as early in the research design process as possible [5,12] and have called on others to do the same. However, there is still a dearth of mobile apps designed by young people for young people. Current data estimate that >10,000 existing mental health apps are available for download [13]. However, a 2020 systematic review identified only 30 mental health technologies that were co-designed digital mental health technologies for children and young people [4]. Another systematic review in 2020 surfaced only 10 publications specifically on mobile apps and web applications for young people, with most being focused on the dissemination of mental health information to specific populations [14]. With the increased use of co-design for the development of web-based mental health resources and the lack of literature on co-design with foster youth, it is imperative to focus efforts on pinpointing strategies that help center the voices of the intended beneficiaries in the creation of these technologies.

Apps and web-based platforms designed to be used by the foster youth community (and caregivers) exist. For example, FosterClub and iFoster are 2 web-based communities targeted toward empowering current and former foster youth through resources and opportunities [15,16]. Currently, iFoster is led by a former foster youth with team members with lived expertise working with children in care and transition-aged youth. FosterClub was founded by a caregiver of 2 foster youth to "help fill the gap she perceived her boys—and other children and youth in foster care in America—were experiencing: a lack of access to a peer support group and information to help them navigate the foster care system" [16]. However, these websites were designed and created by resource providers and professionals (some of whom were former foster youth that had long aged out of the system) and did not directly partner with current or transition-aged foster youth to address the emotional wellness needs of the foster youth community at creation. A recent study with current and former foster youth analyzed a conceptual framework for developing a mobile app to assist

XSL•FO
RenderX

youth in transitioning from foster care to adulthood; however, this study did not result in the formal creation of the app [17]. Thus, there remains a need to center the lived expertise of young adults presently in or recently aged out of the foster care system in designing and building emotional wellness tools. Particularly in this era defined by technology and increased accessibility, there is no room to exclude minoritized young people, such as foster youth, from participating in the creation of digital health technology meant to support them.

FostrSpace

FostrSpace is a mobile app designed and created through a collaboration among San Francisco Bay Area foster youth; clinicians and academics from the Juvenile Justice Behavioral Health (JJBH) research team at the University of California, San Francisco (UCSF); and Chorus Innovations, a rapid technology development platform specializing in participatory design practices. FostrSpace was created to take the stigma out of accessing mental health care and resources for young people who have been in foster care by offering access to digital psychoeducational materials, a personal resource navigator, and licensed clinicians who provide tele–mental health services. To our knowledge, FostrSpace is the only mobile app focused on improving emotional wellness resources for the foster youth community that was co-designed from inception to production to iteration with foster youth. FostrSpace is currently live and can be accessed at its website [18].

Aim of This Study

The aim of this paper is to reflect on the co-design experience of the 4 foster youth involved in the development of FostrSpace. The focus of this manuscript is not on the methodology of co-design as a practice but, rather, on the foster youth experience participating in the co-design process, including how lived experience in the foster care system informed what was cocreated and how the co-design process impacted the foster youth co-designers. These reflections are later leveraged to provide recommended practices for academics aiming to facilitate effective co-design spaces with foster youth. The preceding section briefly outlined the FostrSpace app and its features. The aim of this manuscript is to offer insights into the co-design process from the perspective of foster youth. Rather than providing an exhaustive description or outcomes of the resulting app, our emphasis remains on reflecting on the collaborative design experience. The FostrSpace co-design process reflects the critical duty of academics to collaborate with minoritized communities when developing technological resources. In general, academics without lived expertise in the foster care system can be at higher risk of recreating experiences that result in mental health and resource disparities or creating web-based spaces that exacerbate trauma experienced by foster youth. The duoethnography practice used in this study built on research by Porche et al [5] that aimed to promote more authentic partnerships between academics and minoritized communities. The ultimate goal is to dismantle barriers to resource access while nurturing mutually beneficial relationships among all participants. Through the duoethnography methodology, Porche et al [5] focused on the reflections of 4 psychologists and psychiatrists in academic medicine, 2 of whom are also authors of this paper. Porche et al [5] offer broad co-design reflections across various projects (including FostrSpace), highlighting barriers and challenges in sharing co-design responsibility with co-design partners. The paper by Porche et al [5] focuses solely on reflections from academic co-designers and does not include perspectives from community co-designers. This paper expands on the work by Porche et al [5] by examining the co-design process from the youth perspective and specifically with foster youth. The use of duoethnography in examining the FostrSpace co-design process helped reveal candid reflections on foster youth's lived expertise and involvement in co-design. These reflections underscore how positive dynamics within the co-design team led to an open and collaborative partnership.

Methods

Co-Design

The design of FostrSpace used a community-partnered participatory co-design process to ensure power sharing between co-designers. Techquity, the process of achieving health equity through strategic development and deployment of technology, guided the co-design [19]. Digital health technologies can help connect young people with care, addressing service use deficits and cost barriers. Techquity applies a critical lens to the inclusion and exclusion of individuals during the technology creation process, recognizing that excluding certain voices leads to gaps in technological advancements.

The co-design process used youth participatory action research (YPAR) principles with an antiracist framework. YPAR challenges assumptions about who is allowed to conduct research, placing young people in a central role to investigate social conditions (eg, the foster care system) that impact their lives [20]. YPAR principles underscore the significance of establishing a secure and inclusive environment for involving young people in research processes. Our commitment to safety and inclusion was manifested through practices such as ensuring that participants could make informed decisions about their engagement, highlighting voluntary participation, upholding the confidentiality of shared information, fostering active involvement in decision-making, and implementing continuous feedback and reflection mechanisms. Similarly, we were mindful of racialized research positionality and cultural and developmental contexts, deferring to the FostrSpace Advisory Board (FAB)—a name chosen by the members—co-designers for language and framing feedback whenever possible on the app. YPAR is a valuable approach to support young people wanting to make a positive impact, proving especially impactful for those facing marginalization rooted in racism, sexism, homophobia, transphobia, classism, ableism, or other oppressive factors. With this in mind, we chose YPAR as a framework to underscore the expertise of our foster youth co-designers and carefully manage power dynamics to promote power sharing. While all research projects within the JJBH research team incorporate antiracist principles, it requires a deliberate and sustained effort and, thus, was an intentional focus in the FostrSpace co-design. Particular efforts were made to value multiple ways of knowing and honor foster youth similar and

XSL•FO
RenderX

dissimilar lived expertise, as well as counter paternalism to ensure that participating foster youth were empowered to direct FostrSpace's creation from design to mechanics.

FostrSpace co-design included transition-aged foster youth, clinicians, academics, and technology developers in the same co-design space rather than holding separate parallel co-design processes, as is common in app development. By having all co-designers in the room, everyone could learn from one another and collaborate to integrate distinct experiences into the FostrSpace design process. Youth co-designers came with lived expertise and histories of foster youth advocacy work, whereas academic co-designers came with experience of co-designing with young people in other research projects [5]. Clinicians, some of whom held dual roles as academics, possessed clinical expertise in supporting youth well-being and facilitating group processes. Technology developers came with expertise creating digital health technologies through participatory design.

Co-design of FostrSpace started in the planning stages, with young people aged 16 to 26 years who were previously or currently involved in the foster care system being invited to participate in idea-sharing work groups. The work groups were advertised through organizations and resource providers that have ongoing collaborations with the JJBH research team (eg, the San Francisco Unified School District), who shared recruitment flyers within their networks. In total, 2 work groups took place in September 2021, qualitatively exploring how a mobile app could help support foster youth in identifying and navigating mental health and wellness resources, supports, and interventions. The first work group had 3 participants, and the second work group had 2 participants; each work group lasted 90 minutes and was held with a different group of foster youth. A call was put out to the 5 participants over the 2 work groups to join the co-design team in a more formal staff capacity. Once the team was finalized (ie, the FAB members were hired), formal co-design meetings for FostrSpace took place on a weekly basis.

Duoethnography

Overview

Duoethnography, as defined by Sawyer and Norris [21], was used to guide the analysis. This research method aims to look at "the cultural contexts of autobiographical experiences in order to gain insight into their current perspectives and experiences of issues related to personal and professional identities" [22], allowing researchers to use their "life curriculum" to collaboratively examine and apply meaning to the experience of interest. In this context, the duoethnography process focuses on the experience of 4 self-identified foster youth (the FAB members) who codeveloped the mental wellness app FostrSpace with a team of academics, clinicians, and technology developers. Duoethnography was regarded as the most appropriate approach for this study as it allowed for centering the voices of the FAB. Furthermore, the duoethnography reflection process promoted further collaboration between the FAB and the academic co-design team. The FAB's sharing of their lived experiences that motivated involvement in FostrSpace highlighted the importance of having young people with lived expertise contribute to interventions aimed at their community, as well as the practices that were considered to have made the co-design of FostrSpace a productive, reflective space.

Participants

Coauthors of this paper include 4 FAB co-designers (TL, DM, JT, and AP), with whom the duoethnography was performed, and 4 academic co-designers (IE, JBF, MVDC, and MT-S), who did not provide reflections for the duoethnography. As part of the duoethnography process, FAB co-designers provided both written reflections and verbal responses via a roundtable discussion regarding their experiences co-designing and cobuilding FostrSpace. All FAB members also contributed to the writing of the duoethnography in a multitude of ways (eg, co–first author and author) through providing data to shape the thematic analysis (TA), being involved in TA (eg, coding and writing), or providing feedback on the themes derived (eg, whether quotes were used appropriately or themes were derived appropriately). The collaborative academic co-designers and authors (n=4) were involved in providing structure to the co-design of the FostrSpace app, the reflection prompts, and the roundtable discussion. All academic co-designers were involved in the writing of this manuscript, in TA (eg, coding and writing), or in providing feedback on the themes derived. While the sample size of 4 FAB members contributing to the data collected for the TA is small, it reflects the number of young people with lived expertise in foster care who contributed to the development of FostrSpace and their reflections on the co-design process. Duoethnography does not have a set number of collaborators needed to engage in the methodology, rather focusing on the ability to maintain "intimacy, trust, and commitment" [22] to the tenets of duoethnography (eg, self-reflexivity) among 2 or more researchers. Similarly, reflexive TA dictates that the participant sample size for analysis depends on the size and scope of the project [23]. The FAB members' reflections provided a basis for engaging with foster youth in collaborative development of digital solutions and a framework for centering lived expertise in the development of and research on mental health technology.

The FAB included a White nonbinary foster youth, a White-passing mixed Asian-Hispanic foster youth, a Mexican-Latino foster youth, and a Chicana foster youth. The FAB members are a group of current and recently graduated multilingual, first-generation college students who, together, have experiences in both the traditional foster care system and kinship or relative care, making them intimately aware of the disparities present in the child welfare system. At the time of the duoethnography, 75% (3/4) of the FAB members identified as former foster youth, and 25% (1/4) identified as transition-aged foster youth; they maintained a consensus that they were not current foster youth due to being aged over 18 years but were still impacted by the identity. Through their upbringings, FAB members experienced poverty, low-income, working-class, and middle-class statuses, living in group homes, kinship caregiver households, community housing centers, campus housing, and individual housing. As a group, they hold both cisgender and queer identities, various levels of connection and separation from their cultural identities, different levels of disability status, and part-time positions in both FostrSpace and

XSL•FO
RenderX

other jobs in the community while maintaining statuses as students or full-time employees.

The academic co-designers included a first-generation Nigerian Black pangender and pansexual research coordinator; a first-generation Guatemalan American, neurodivergent, cisgender, and straight light-skinned Latina clinical psychologist; a White Jewish genderqueer and gay clinical psychologist; and a White-passing Iranian immigrant cisgender and heterosexual clinical psychologist. All are conducting community-engaged research to increase behavioral health service access for youth and families in contact with the juvenile legal and child welfare systems and have had personal ties to these communities.

Through connected but distinct lived expertise and various roles that we occupy in community and academic spheres, all the coauthors of the paper have an awareness of the types of hardship that foster youth can experience from childhood to adulthood and how those hardships interact with intersectional identities. These hardships included social and emotional factors such as parentification from a young age as well as structural and systemic barriers to resource navigation in the behavioral health, medical, and educational spheres. Our own identities and lived expertise drive us to address these barriers to improve conditions for foster youth and provide a lens for interpreting our results.

Procedure

Following duoethnography guidelines, FAB members reflected on their experiences in the foster care system and how this related to their experience in the FostrSpace app development co-design process (eg, what they personally got out of the process and how they conceptualized different features on the app) via written reflections and verbal discussions. Members spent 1 month (April 2022-May 2022) writing their own reflections and reading and responding to their fellow FAB members' reflections. Following the writing period, an in-person roundtable discussion about the co-design experience was held to allow members to verbally share their reflections as well (eg, because not all co-designers felt as comfortable with written reflections). The 4 FAB members engaged in an hour-long discussion with JBF as a facilitator and IE as notetaker. The facilitator and notetaker were academic co-designers that contributed to the FostrSpace app development and duoethnography paper writing; neither held a formal supervisory role over the FAB members (ie, the supervisor on record was not involved in the duoethnography process). We acknowledge that the power dynamics of having academic co-designers facilitate the roundtable discussion among the FAB members about their co-design experience could have impacted what was and was not shared. Keeping this in mind, academic co-designers provided structure to the discussion through the prompting of discussion questions while allowing space for the co-designers to talk with one another freely on the topics with limited interjection. In addition, FAB members and academic co-designers built mutual, trusting partnerships with one another during the co-design process, and the duoethnography took place after the bulk of the co-design process was concluded and the app was functional. As a result of the mitigation of the potential consequences of negative power dynamics through these methods, FAB members felt comfortable sharing sensitive data in the roundtable, as indicated by the vulnerability and honesty exhibited in the quotes used in the results as well as within co-design meeting discussions and one-on-one check-ins with staff. Reflection prompts were collaboratively developed by the coauthors before starting the duoethnography process, keeping in mind the research aim of understanding the experience of foster youth co-designing an app in collaboration with academics, clinicians, and technology developers. The same prompts were used for the written and verbal reflections. The prompts are available in Textbox 1.

Textbox 1. Reflection prompts used in the FostrSpace duoethnography with FostrSpace Advisory Board co-designers over the 1-month reflection period.

- Why did you get involved in this process? What encouraged you to keep going? After attending the first meeting, what kept you going to the second, third, and so on?
- How do you feel about the way the co-design was handled? In what ways would you have liked the process to go differently? What were your experiences of power dynamics in the co-design process and in what you bring to this work?
- Why is it important to center lived experience in this work? What parts of yourself and your experiences in foster care did you bring into this work? What inspired you throughout your time in foster care to have interest in what we are doing? What support did you lack in foster care? Remember: not every story is the same—not everyone would do the same things growing up and everyone's feelings are valid.
- How have you changed throughout the course of the co-design process? How have your goals or aspirations or mindset changed since starting this project?
- Has this project amplified your voice or feelings toward reform within the foster care system? In what ways?
- Why do you think this is a good resource for other foster youth to have?

Data Analysis

Reflexive TA [23-27] was used with 9 pieces (n=8, 89% written reflections and peer responses to reflections and n=1, 11% roundtable transcription) from the 4 FAB members. Reflexive TA was specifically chosen (as opposed to a coding reliability TA or a codebook approach to TA) to be congruent with the co-design value of honoring participant lived expertise. Reflexive TA allows for the analysis of participant subjective data through the lens of researcher reflexive interpretations [26]. An inductive, "bottom up" approach under an essentialist and realist paradigm was used by the first author coders (IE and TL) to ensure that identified patterns within the data were pulled from the data itself [23-27]. Initial codes aimed to provide a rich description of the full dataset to capture the breadth of

XSL•FO
RenderX

factors that should be at the forefront of academics' minds when co-designing with individuals with lived expertise in the foster care system. Themes were developed using a semantic approach in which surface meanings of the data were focused on in the initial coding process. This distinction was important for the coders as a way to acknowledge the prevalent history in the foster youth community of feeling disempowered by individuals and groups in power. By using a semantic approach to theme development, coders hoped that the process would be more empowering and respectful to the FAB members' experiences. A realist and essentialist paradigm rather than a constructionist approach was used to focus on individual psychologies as well as differences and similarities in lived expertise among the FAB members.

In line with the 6 guidelines by Braun and Clarke [23-27] for reflexive TA, the coders first familiarized themselves with the dataset before initiating coding. As an FAB member and participant in the verbal discussion and written prompts and reflections, TL was intimately familiar with the dataset. Similarly, as lead research coordinator on FostrSpace, IE read the FAB reflections as they were being written and acted as the notetaker during the roundtable discussion. All qualitative analysis was conducted in ATLAS.ti (ATLAS.ti Scientific Software Development GmbH) [28]. The 2 coders independently generated initial codes related to the co-design process and the foster youth experience. Preliminary codes were discussed collaboratively among the coders and coauthors (IE, TL, JBF, and MT-S) to develop a final list of codes. A total of 18 codes (initial broad themes) were developed and used to conduct final coding through which themes were developed (Textbox 2). Analysis involved more than one coder and, therefore, necessitated collaborative discussion.

Textbox 2. The 18 codes developed from thematic analysis and applied to the 4 FostrSpace Advisory Board members' reflections on the co-design process.

Codes

- App as a place for transparency
- App as a safe space
- Barriers to accessing resources that are not foster youth specific
- Barriers to accessing resources specifically for foster youth
- Co-design as a safe collaborative space
- Future goals for the app
- Identity (negative)—external gatekeeping
- Identity (negative)—internal gatekeeping
- Identity (positive)—finding, accepting, and embracing
- Initial app involvement
- Intersectionality
- Lack of clarity—foster care system
- Motivation for the app—paying it forward
- Negative experiences in kinship care
- Positive community support
- Positive impact of co-design participation on the self
- Resource stigma
- Stigma toward foster youth

The results include final analyses of the selected extracts and relevant themes, which were reviewed and approved by the coauthors of this paper, including the entire FAB team, who had the option of retracting extracts of their data that they did not feel comfortable sharing with the public. All FAB members endorsed the information presented in the manuscript and opted not to retract data.

Ethical Considerations

The institutional review board of the UCSF determined that this research is exempt from institutional review board oversight according to the federal regulations summarized in Title 45 of the Code of Federal Regulations Part 46.102(l). Participants from the work groups were compensated with a US $75 gift card for their participation. A total of 80% (4/5) of the individuals from the work groups applied for and joined the final co-design team, making up the FAB. FAB members were then hired as UCSF employees to co-design the FostrSpace app and were compensated at an hourly rate based on their individual compensation package, which was determined by educational level and experience as set by UCSF Human Resources for all the time spent completing tasks related to UCSF employment (eg, co-designing the FostrSpace app and engaging in the duoethnography reflections). The FAB as employees of UCSF must abide by federal and state employment laws as well as have access to human resources to support any workplace issues.

XSL•FO
RenderX

Results

Overview

Coders derived 4 main themes from organizing codes around core commonalities through reflexive TA. The final themes addressed the experience of taking part in the co-design process, including the FAB members' broad experiences in the foster care system, which directly influenced their approach to and navigation through the co-design process and app development. The number of quotes presented in the results from each FAB member ranged from 5 to 15 excerpts. Themes included (1) power and control (loss of power, reclamation of power, and power dynamics), (2) resource navigation, (3) building community and safe spaces, and (4) identity. All themes were rooted in explicit content from the coded data.

Power and Control

Loss of Power

FAB members described their experiences with a loss of power due to the nature of their foster youth status through childhood, adolescence, and even extending into young adulthood. Systems (ie, the child welfare system) allegedly designed to empower and protect them withheld crucial information necessary for FAB members to feel like they had agency over their own lives. To illustrate, the lack of transparency regarding monthly payments made to foster parents (both the monetary amount and transaction timeline) led to frustration and disempowerment stemming from FAB members feeling that they were left in the dark about their own care:

> *So, for a while, I was just like, "I'm homeless." Or I'm not homeless, "I'm poor." I was just like, "I'm poor. I can't ask for anything." And then 10-plus years later to find out that that person was still getting paid to take care of me? And I'm like—because they're family, you know what I mean, it's [sic] was like they didn't really owe you much?* [FAB member 3]

Existing foster care systems and policies (eg, foster youth not having a say in placement with family members) constantly take away power and control from foster youth, making decisions for them without their input. Decisions made by guardians without FAB members' input negatively impacted the ability of FAB members to access vital information necessary for their care:

> *I asked my grandparents, "Who is my social worker?" They were like, "Oh, we shredded all those documents because it didn't apply because they weren't coming anymore. And we were good." I'm like, "What?" So, I tried calling the county, they never got back to me. They just put me in this whole rigamarole.* [FAB member 2]

In addition, though part of a larger network of youth in the child welfare system, FAB members noted the isolation prevalent in the foster care system:

> *Growing up foster kids were never given a platform to speak and ask for help. I personally was never given resources growing up that could help connect me to other foster youth or provide mental health resources when I needed them.* [FAB member 1]

Reclamation of Power

By participating in the co-design of FostrSpace, FAB members were able to reclaim some of the power they had lost and their decision-making ability. Members felt empowered to reclaim control over their own trajectories as well as lay the groundwork for foster youth coming after them to assert agency. Through FostrSpace, FAB members felt that they could equip current and future generations of foster youth with information and resources, which could allow them to take control of their own lives and mental health:

> *I agree that the app is a great place for current and past foster youth to find a supportive community to share with as it is not always easy to reach out to the people around you for help. Especially when you're searching for a safe space to share struggles with mental health as so much stigma still surrounds mental illness in the United States as well as being a former or current foster youth.* [FAB member 4]

In addition to promoting resource access, co-designing FostrSpace strengthened FAB members' feelings of power and control by educating youth about the foster care system. Growing up in the isolation of the foster care system led to increased appreciation for the transparency available in FostrSpace:

> *[My caregiver] villainized everybody and it was like we cut contact off with everyone. It was really like isolating. So, I like that we have this resource where we can go search it up and be like, "Okay, well, what does it mean to be a foster kid? Where can I find a resource to go get some money for food?"* [FAB member 1]

FAB members emphasized the importance of giving users ample information to equip them with knowledge, encouraging them to be informed consumers. FAB members appreciated how FostrSpace leans toward having too much rather than too little information:

> *That's the thing—like the app. I remember I always like where we specify everything down to a T. Sometimes it gets a little wordy and that's annoying for some users of the app, but it explains everything to you clearly.* [FAB member 1]

With the opportunity to use their lived expertise in the development of the web-based platform, FAB members felt that they could reclaim the narrative over their life. FostrSpace gave them the opportunity to regain control after years of experiencing intentional withholding of control by the system:

> *It really points to just this greater idea of mitigating suffering. This greater idea of moving beyond and I think that's really important. I've always been focused on this as a greater abstract idea, but this is one of the ways I've actually nailed down providing support and making the world better, in a limited way that I*

XSL•FO
RenderX

> *can. And so, I've really appreciated having the space to allow me to actually do this.* [FAB member 2]

Power Dynamics

The interplay of power dynamics was key for FAB members regarding life experiences as detrimental versus supportive. Harmful power dynamics manifested in both navigating governmental assistance and the family unit. While navigating already mystifying systems, FAB members had to advocate for themselves to individuals in positions of power when submitting applications for food assistance, housing, and more. FAB members also expressed feeling as if their own realities and experiences were minimized or denied when trying to access resources:

> *...I had to call back for my EBT [electronic benefits transfer] because they denied it last month. I remember calling them and I was like, "Hey, you guys denied it. Even the app says I didn't have any money entered." And she's checking her records and she's like, "No. We put it in." And I was like, "Lady, like no you didn't."...I remember just being like, Come on. I'm already telling you I had to use money out of my own pocket that I've been trying to save for rent. I know what I'm using my money for.* [FAB member 1]

Government and school staff alike also denied FAB members' foster youth identities and ability to access foster youth–specific resources because they did not have proper documentation labeling them as foster youth:

> *In school, I remember I was constantly denied. I asked if they could get help finding foster youth resources and scholarships, and I literally had staff tell me, "You're not a foster kid. I don't have documentation."* [FAB member 1]

Consistently facing roadblocks and antagonism from the systems designed to help them led to feelings of disempowerment among the FAB members. Having to go up the chain of command to receive a response or going through several layers of forms, applications, and paperwork only for these to be rejected accumulated experiences of trauma and negatively impacted the motivation to access these resources for some FAB members:

> *I am and have been eligible for EBT [electronic benefits transfer], but after doing it for a year I don't do it...how many of these are actually foster youth-friendly? Because we can tell people to sign up for SNAP, but when we look at these programs the blatant villainization or willful ignorance to what we've gone through and the struggles we are going through makes it so it's harder to reach out for these resources...they don't take me seriously when I tell them about the trauma that comes from it.* [FAB member 2]

FAB members experienced kinship care as mirroring the unsupportive power dynamics in governmental assistance, with kinship care "family" behaving differently from how a "traditional" family or chosen family or community may be thought to behave. FAB members described feeling alienated due to the dynamics of being in a kinship care household:

> *...even though foster youth may have access to family in foster care through family members being their foster parent or having their sibling with them, it's still easy to feel alone. I personally felt that as the oldest sibling of three I couldn't be vulnerable when I was struggling as everyone else was also having a difficult time. In my experience being in foster care made it difficult for me to ask for support, I was uncomfortable being in a new parental dynamic with people I didn't know too well so it was hard to ask for things.* [FAB member 4]

FAB members experienced kinship caregivers withholding resources and information about the foster care system that FAB members needed even though they were meant to be trusted adults who could provide care:

> *She just wanted to do stuff where again she was the only one in control. And with me going and finding these resources? She didn't help me go do that. She was leaving me on my own and like, "Go do that." Again, it's just like she's leaving me out there and alone. It was really sad.* [FAB member 1]

Similarly, FAB members felt the weight of being the different party in a new caregiver-youth dynamic with mixed signals:

> *But looking back on [growing up], it was a heavy load kind of me just to like—well, my family is proud of me? No, they're not proud of me because I'm smarter than I guess they thought I'd be. So, for a while, it was just like wow. I was just so conflicted where it was like, "Am I doing the right thing? Should I keep pursuing my education?"...I'm like..."Why are you talking down on me and making me feel bad?...Isn't this what y'all wanted for me?"* [FAB member 3]

The ability of the co-design process to empower FAB members was in great part due to the intentional way in which power dynamics were attended to in the process:

> *I remember telling everybody like, "Dude, it's equal here...they don't talk on top of each other. They don't cut you off. They let you finish your idea even if it's dumb because it might be." [laughs]...here everyone is on an equal playing field, even your bosses. They're all chill with you...they're always, they're engaged. They want to listen, they want to do something, and they want to help the community. I think that's the most important part is the want to do something and then actually doing it.* [FAB member 1]

Notably, as individuals who had experienced systems designed to uplift them repeatedly ignoring their voices, FAB members went into FostrSpace without hopeful expectations:

> *I didn't necessarily anticipate having a big role in the development of this app, but I also didn't have many expectations about its potential either. I became ecstatic when I received my official offer letter to join*

XSL•FO
RenderX

> *UCSF and begin the process of developing an app as an ambassador.* [FAB member 3]

Over time in the co-design space, FAB members became comfortable, learning that the participating academics were working for the betterment of the community and uplifting individual voices within the group:

> *...the co-design team was wonderful to work with as it felt all the foster youth giving feedback were heard and treated as equals by superiors. Overall, the friendly space that was created by the entire team during meetings and the positive influence the finished app would have in the foster youth community keep me returning to the meetings.* [FAB member 4]

Resource Navigation

FAB members shared situational factors that contributed to inaccessibility of resources, including lack of knowledge of available resources and how to access them, responsive resources, adult figures to guide them toward resources, and documentation. FAB members expressed frustration over being stuck in a system with ambiguous directions:

> *Yes. Oh, my god. [collective agreement] All the [redacted name] county forms are like, "Worker ID this, ID this." Then you're like, okay, who do I call? Leaving voicemails for the workers because they never want to answer their phones. Then they never call you back. It's just stuff like that where you're just sitting there like, "Okay, well, what do I do?"* [FAB member 1]

Furthermore, FAB members mentioned that, when they did access resources, these were often not responsive to their lived expertise and status as foster youth, instead following a one-size-fits-all structure. FAB members recounted situations in which resources did not meet their needs due to limitations in access (eg, time frame):

> *I remember telling my whole life story at five years old to this lady and just telling her, "Oh, this is what happened," again the gory situation and everything, and she helped me cope with it. It was really good. I would walk out feeling refreshed. I felt like everything was just off my shoulders. Then after maybe like a few sessions it just stopped. And then I never got any resources.* [FAB member 1]

A lack of reliable adult figures during key periods of development was a common experience for FAB members, who suddenly lost support people who could direct them to needed resources. FAB members expressed that FostrSpace resonated with them as it attempts to connect foster youth with resources for mental health and basic needs:

> *I had received an email about outreach being done through UCSF to develop an app to better support the needs of foster youth, particular in the field of mental health and connecting youth to basic needs. I was previously in foster care in my early teens and lost both my parents in my teens, these experiences made me feel like I had no adult guidance to show me what resources I had available to me as well as support me when I was dealing with mental health issues.* [FAB member 4]

When resources required FAB members to provide documentation (eg, case IDs, permanent address, or social worker's name) that they did not have access to, they were hindered in submitting applications:

> *...they were doing scholarships or support funds for foster youth, and even like transition age. I remember I applied to it and they were like, "What's your social worker's name? What's your case number?" All this stuff. And I didn't have that information.* [FAB member 2]

FAB members also had difficulty identifying resources that they were entitled to as foster youth, such as Court Appointed Special Advocate workers:

> *Like, I'm going to be honest. I don't think I ever had a CASA person. So, just like you guys talking about CASA, it's like there's so many different aspects of just being a foster youth where we all have different experiences.* [FAB member 3]

The resource navigation challenges experienced by FAB members were described in contrast to the FostrSpace approach, which could benefit foster youth experiencing the same deterrents to resource access. The web-based platform provides foster youth with a directory of resources and live help on emotional wellness or resource navigation without the need for documentation. Due to FostrSpace's more straightforward approach to resource navigation, FAB members expressed that FostrSpace helps reinforce self-advocacy and facilitates positive resource navigation experiences:

> *Moreover, having good experiences accessing resources should further encourage foster youth to engage with resources as I have experienced and seen people turn away from help when their experience is negative.* [FAB member 4]

Building Community and Safe Spaces

FAB members discussed the importance of centering building community and safe spaces for foster youth. Membership in an established foster youth community positively impacted FAB members:

> *But I remembered when I got to UC Berkeley, my brother immediately was like, "Oh, my god, there's this program called Berkeley Hope Scholars"... I signed up—and [Director name redacted] like a week before was just like, "Hey, this is what we have. These are the resources. We're going to help you with everything." And I immediately felt a sense of just comfortability. I felt safe there and I finally found a community where I could find other people who are like me or who experienced some of the same stuff as me.* [FAB member 1]

Similarly, the FAB community was beneficial for creating a space to collaboratively give back to the foster youth community through FostrSpace:

XSL•FO
RenderX

> *After attending the first meeting and meeting my co-workers I instantly felt as if I had found a great community of people fighting for change for a community that is often hidden in the shadows. I was excited to finally have my voice heard and to speak for my friends who experienced the system in a multitude of ways.* [FAB member 1]

There was shared comradery among FAB members as FostrSpace allowed them to unite under a common goal: to curate a safe space for heterogeneous foster youth tied by a common thread of lived expertise:

> *We're dedicated to just creating a space so that people don't have to further suffer the way we might have already, or in ways that we may not have experienced, but we don't want anyone to ever risk going through.* [FAB member 2]

The importance of considering intersectionality was highlighted when building community and safe spaces for foster youth as foster youth experiences are shaped by how multitudes of identities, global world views, and inclusion interact. Through this duoethnography process, FAB members came to the realization that future iterations of FostrSpace's resource directory need to take into account other identities that foster youth may hold (eg, race and gender) to better target nuances of resource inaccessibility:

> *...I'm always talking about my multitude of identities and intersecting identities. But that's something where I don't want resources just to default onto foster youth because sometimes they're for White foster youth, Asian foster youth, or Black foster youth. It's like specific. So, I'd really like to find more resources for trans youth, and for other races, as well, like Black, White, Asian, Latino—because there's a huge Latino population in California. But just stuff like that where it's resources for women's reproductive health because again, I'm a woman and that's something I have to go through...Like just where it covers those identities and takes care of kids that not only align under foster youth, but are foster youth AND, you know foster youth-included.* [FAB member 1]

FAB members also felt fortunate to have found a foster youth community as well as a common desire to develop a community within FostrSpace, expressing that FostrSpace could help connect more foster youth with a supportive community:

> *But then when I actually saw what y'all were doing, I was really moved by it. Like providing mental health and wellness resources for like—a I just felt like it spoke to an earlier me that was in desperate need for those and needed support and a community aspect.* [FAB member 2]

Identity

FAB members engaged in rich discussions regarding the decision to identify as foster youth, the journey of coming to that decision, and the resulting consequences. Discussions highlighted that carrying a foster youth identity led to grappling with both positive and negative experiences. Within the context of negative experiences regarding identity, FAB members disclosed the strain that shame and stigma have on a foster youth's ability to embrace their identity. Shame was described as something that was both ascribed by others (eg, strangers and family members) and internalized. FAB members associated the identity with traumatic upbringing:

> *I was never told that I was even a foster kid. When I said—like when me and my brothers said that name, my aunt would roll her eyes and be like, "What's that? You're just making stuff up. You're dramatic." So, it was like things like that. It had a very negative connotation around it.* [FAB member 1]

While kinship care is recognized as a part of the foster care system, FAB members noted internal shame regarding placement in kinship care:

> *I think part of it is because I still feel a bit of shame about my identity where there's a part of me that's just like, "Yeah, but you were raised by your grandparents. And even though you were in the system, do you really count?"* [FAB member 2]

In addition, they experienced external shame when actively aligning with the foster youth identity:

> *...when I created my website, I was really open about it, I've written a lot about it. Then one day, one of my distant uncles calls me. I'll never forget this. I was getting ready and he was just like, "I saw that you posted an article on your Facebook about saying that you were a foster youth. You're not a foster youth." He was like, "How do you think your grandparents feel if you told them that you identify this way or that you say this?"* [FAB member 2]

Stigma was evident in the mischaracterization of the foster youth identity and pitying by others:

> *...when we were handing out flyers, one of my friends was like, "Watch out, they're going to say sorry a lot." I was like, "What do you mean?" And we were handing out flyers and the students were like, "Oh, I'm so sorry." They'll look at you with puppy-dog eyes and you're just like, "I'm good. I'm chilling, bro. I just got some free lemonade." [group laughter] It's just stuff where you feel the name "foster youth"—just you feel like has this weight on it. And that's what people add to it. It's just like not like that.* [FAB member 1]

While negative associations with identity stemmed from general community (eg, schoolmates), family (eg, kinship caregivers), and internalized anti–foster youth dialogue from living with kinship caregivers, positive associations with their foster youth identity were built through engagement in the FostrSpace co-design process and with a foster youth community. Being a part of the co-design process offered FAB members space to reflect on negative experiences related to identity as well as build a new, positive association with the foster youth identity (eg, through community building in meetings):

> *...this process has...given [me] an ability to understand and process the uniqueness of what I have*

XSL•FO
RenderX

> *gone through, but also the commonality of the struggles and how no matter who you are, we all share these ideas, these feelings. And so why not try together to overcome them?* [FAB member 2]

Moreover, experience in an established foster youth community encouraged FAB members to embrace and accept identity, as well as providing a venue to receive support regarding struggles of identity. Co-designers expressed that being members of the FAB and working on the development of FostrSpace together validated their identity and allowed them to recognize that they were not isolated in their identity or experiences:

> *So, talking with you guys has always been great because every meeting is like, "Damn, I didn't know that," or, "Yes, I've been there." Not that I'm glad that someone else has went through that with me, but it's just like knowing that now collectively we can put that insight and experience to fruition.* [FAB member 3]

FostrSpace also acted as a platform to support the dismantling of stigma regarding the foster youth identity as FAB members were given the space to call on their lived expertise in the development of FostrSpace:

> *I felt as if I had just gained a new purpose in life, which was to share my lived experience and encourage others in the foster care system to embrace their stories and share it with pride.* [FAB member 3]

Discussion

Principal Findings

Co-design processes hold significant potential to create meaningful interventions for underserved populations such as current and former foster youth. The co-design process for FostrSpace exemplifies the value of centering the lived expertise of foster youth in the creation of digital tools designed to support them. The FAB members actively brought their own lived experiences navigating resources in their personal lives into the co-design space and used those experiences in developing the resource library for FostrSpace. As active co-designers, FAB members identified potential barriers to accessing resources that foster youth may face as well as brainstorming and problem-solving ways to inform the FostrSpace approach to resource navigation. For example, one barrier to accessing resources that FAB members identified was applications requiring details that foster youth may not have access to (ie, their social worker's ID, particularly because some foster youth are not alerted to the fact that they have a social worker). Due to this, the FAB reflected as part of the co-design that FostrSpace needed to commit to providing resources made with foster youth in mind. Ideally, foster youth–tailored resources would not require information that foster youth do not have access to. To address this, the FAB recommended providing brief highlights on the information required to access certain resources (ie, CalFresh). This would equip foster youth with knowledge of potential barriers before trying to access the application. In line with this, during co-design, the FAB advocated for ensuring that foster youth did not face similar barriers and restrictions to accessing FostrSpace to those they faced in accessing other resources meant for foster youth (eg, as such, FostrSpace does not require foster youth to upload proof of status as a foster youth or social worker ID numbers).

Furthermore, reflections from the FAB members indicated that the co-design process was a safe and constructive space where all members worked together for the betterment of the wider foster youth community. As FAB member 3 stated, "Together, we can change the system and it all starts with collective effort of informing the broader community," showing the drive to collectively work toward the shared goal of creating FostrSpace. This duoethnographic process adds to the literature of co-designing with young people and specifically to the scholarship of co-designing formal mental health applications with foster youth.

Recommended Practices

Overview

The 4 main themes shared across the 4 FAB members demonstrate their reflections on their experiences with the FostrSpace co-design process, including motivations for participation and potential reasons why FAB members could not access the very resources they aimed to make more accessible through FostrSpace. From this reflection process, we recommend 3 key practices when co-designing with young people. Although we did not explicitly seek direct feedback from the FAB on these practices as part of the duoethnography process, the richness of the duoethnography data affirms the value of these practices in creating an enriching co-design space. In addition, while a number of different practices were brought into the FostrSpace co-design space, the key practices that we chose to list here are inspired by what was highlighted in the FAB members' duoethnography quotes in addition to their endorsements in co-design meetings and check-ins. These recommendations are built from practices that were implemented in the creation of FostrSpace and are believed to have made the co-design process successful and worthwhile for co-designers:

1. Engage in active reflection on the co-design process both during and after co-design.
2. Find creative ways to engage in power-sharing practices to build community across diverse experiences and positions of power and privilege.
3. Ensure mutual benefit among co-designers.

Engage in Active Reflection of the Co-Design Process Both During and After Co-Design

Engaging in active reflection throughout the co-design process allows for the intervention (in this case, FostrSpace, a digital health technology) to become more actively shaped by thoughtful reflections. In the co-design of FostrSpace, brief icebreakers were used at the start of every meeting to ease individuals into lighthearted discussion before talking about a specific feature of the app. In addition, brief check-in questions were used at the end of meetings as well as in occasional one-on-one check-ins between FAB members and UCSF academic staff to evaluate co-designer comfort and feedback on the app. Although these engagement and reflection methods were highly useful in building community among the

XSL•FO
RenderX

co-designers and allowing for free discussion, many of the reflections uncovered through the duoethnography were highly distinct and had not surfaced through previous reflection tools used during the co-design process. Specifically, by having a space designed solely for FAB co-designers to talk with one another and respond to questions not directly related to building the app through the in-person duoethnography process, we were able to reaffirm the most important features of the current app and important aspects to consider for iterations moving forward (eg, fully detailing a community element and the different ways in which this was crucial for this community).

Notably, in FostrSpace, we aimed to provide a robust resource directory for foster youth with food insecurity, legal or housing concerns, employment goals, or transportation goals, among others. We felt confident that we had developed strong resource lists within the counties that we initially served (eg, by tailoring resource pages with resource eligibility criteria, relevant website links, summaries, and office hours); however, we found through the duoethnography process that we were missing an essential element. The resources, though helpful when successfully acquired, often contained application or eligibility processes that were not foster youth friendly. This discovery, made possible only due to FAB members discussing their own personal frustrations with accessing public aid (eg, being given the "run-around," having their needs denied, or being shamed for their needs) throughout the duoethnography, highlighted the need to seek out resource providers knowledgeable of the foster youth experience. It also reinforced the need to continue pushing the personal navigator to the forefront of the app as someone "who takes the role of goal-collaborator and accountabili-buddy by providing connections to resources so folks don't have to navigate life alone" and ensuring that, for resources that are not foster youth friendly, app users have someone to support them throughout the process. Furthermore, FAB members' experiences navigating public resources highlighted the importance of self-advocacy when dealing with resource providers lacking knowledge about the foster youth experience; personal and peer navigators (FAB members) can connect with users to help build self-advocacy to support resource access. FAB members as peer navigators and continued contributors to the app have since made content on sharing knowledge with current and former foster youth about skills they learned transitioning into adulthood, essentially creating a "how to" guide to accessing some common resources (eg, electronic benefits transfer).

Find Creative Ways to Engage in Power-Sharing Practices to Build Community

As noted previously, co-design meetings began with icebreakers, some led by UCSF academics, some led by Chorus designers and developers, and some led by the FAB members. Co-designers were encouraged to think of icebreakers in between meetings, topics they were curious about in their coworkers. Many came to the meetings excited to share their icebreakers, as FAB member 1 shared in the duoethnography roundtable:

> *I think the intro questions are my favorite part. I think that really showed it, too. They were like, "Hey, what's your favorite taco spot?" It was just like small questions like that where you were like, Oh, okay, this is a cool space. Every time before each time, I'm like, Okay, what cool icebreaker questions will I ask my coworkers?*

These icebreakers aided the development of a relaxed space before diving into the FostrSpace app design and development. All members had a chance to contribute and share something before transitioning into potentially daunting topics regarding designing an app. Furthermore, engaging in conversations about non–work-related topics served to shift the focus away from work titles and hierarchies, allowing participants to connect on a more personal level beyond the confines of the work environment. In addition, as co-design meetings were often in the early evening after a potentially draining work or school day for co-designers, the icebreakers helped in providing a check-in for co-designer energy levels, which influenced how many breaks were taken in a meeting and what agenda items were prioritized. FostrSpace was an app built through a co-design process that involved internal meetings among UCSF academics and clinicians; collaborative meetings with UCSF and Chorus designers and technology staff; and co-design meetings with UCSF, Chorus, and the FAB. Many ideas went through various iterations, with the FAB voice being taken into the highest consideration. FAB members provided different viewpoints and made decisions on design elements such as name and color scheme, as well as content elements such as specific questions and text language. Furthermore, having a technology partner with extensive experience in co-designing digital health technologies with diverse communities and ensuring equitable power sharing in the development of FostrSpace contributed a great deal to the success of the co-design space.

Ensure Mutual Benefit Among Co-Designers

The maintenance of reciprocal relationships during the co-design process is ideal, with all participants (academics, foster youth, clinicians, and developers) benefiting, whether through greater understanding of the self, community building, professional development, or furthering research. This relationship can allow for continued engagement in co-design and progression of the product, elucidation of honest and often sensitive or personal feedback, and novelty and creativity in the design process. Although FAB members held their own legitimacy as individuals with lived expertise in the foster care system, UCSF academics went through the process of creating staff positions for the FAB members, formalizing their roles and providing a professional title to add to their career trajectories and resume. FAB members also actively engaged in professional development opportunities such as presentations to community members, colleges, and donors to discuss the FostrSpace development. In hiring the FAB as staff, UCSF academics reaffirmed that FostrSpace would be a longer-term commitment and partnership. In turn, UCSF academics derived benefit from the stable consistency of FAB staff, who not only were committed to their work but also had the stability through financial support that allowed for continued engagement in the work. All co-designers were able to benefit by learning from one another (eg, UCSF academics and clinicians and FAB members learning about the vast capabilities of Chorus),

XSL•FO
RenderX

building community, and having a safe space to share. All co-designers also benefited from building an emotional wellness app for foster youth—the FAB members by developing an app tailored to their needs and the needs of their fellow community members, UCSF academics by adding to the literature and development of foster youth interventions and services, and Chorus by partnering with another community to leverage participatory design practices.

Limitations

The foster youth lived expertise is highly heterogeneous, and thus, there are limits to the generalizability found in the themes we derived. The cohort of co-design partners was diverse in ethnic origin, academic background, sexuality, and lived expertise; due to the small size of the FAB group (4 members), there are limitations to broader application as lived expertise in the foster care system can vary widely. Within the small group of FAB members, there was notable overlap in their lived expertise, especially due to the overrepresentation of students (all 4 FAB members were students in higher-education institutions, and 3/4, 75% were in the same college). The common profiles of tertiary education institutions and academic studies in the San Francisco Bay Area provide another limit to the generalizability of these findings to foster youth outside of the region. All FAB members had also been placed in kinship care, suggesting that barriers to mental health resources that are faced in other forms of foster care placement may not be adequately highlighted. Expanding the co-design process to include more diverse academic backgrounds (including youth who are not pursuing higher education), partners from regions beyond Northern California, and different forms of foster care placements appears to be a promising next step for improving intervention impacts and understanding how foster youth contribute to co-design processes.

Conclusions

Lessons learned from this duoethnographic process can inform future researchers seeking to co-design digital health technologies with foster youth, as well as young people more broadly. The FAB's determination of roles and level of participation underwent constant revision to accommodate and support FAB members as their lives developed over the course of the project. As suggested by Porche et al [5], by tailoring the nature and style of power sharing to the interests and thematic motivations of the co-design partners rather than just increasing the degree of power sharing and authority alone, a greater depth of engagement can be achieved. To this end, early and concurrent partnership with the intervention population to improve impact and ensure not only a satisfactory co-design process for the partners but also a tailored end product is essential [29]. Future research may consider the nuances of the themes derived from this duoethnography and devise methods to ensure that future co-design processes address those themes both through the actual co-design and in the developed intervention or app. Foster youth who have had extensive experiences with negative power dynamics in different spaces, for example, may benefit more from greater intentionality toward the cultivation of safety and empowerment in the co-design space than may be typically granted for other young people. Similarly, foster youth such as those who participated in this duoethnography process may have increased levels of negative experiences accessing resources and may be put off from the idea of trying to access future resources through a different platform; in this case, someone creating a resource bank may seek to provide more details on the resources being offered, tips and tricks for successful resource access, and potentially a difficulty rating or timescale indicating how much bandwidth a foster youth would have to commit in trying to access the resource.

Due to the diversity within the foster youth community, it may be difficult to develop a single plan that ensures adequate empowerment and recognition of each co-designer's lived expertise in a co-design process, thereby reinforcing the need for active reflection throughout the co-design process. Continuous reflection co-design practices can be refined over time to meet the needs of a specific co-design team. The themes and subthemes gathered from the FostrSpace co-design team's duoethnography may provide a helpful framework in starting to think about how to best serve foster youth throughout the co-design process. Co-design practices are vital for creating digital health technologies that are helpful, intuitive, and healing for their intended users. Co-design practices can draw out nuanced community knowledge and result in a thoughtful, vetted product. We call on researchers to continue the work of co-designing with young people and specifically foster youth, with a commitment to active and in-depth reflection throughout the co-design process.

Acknowledgments

FostrSpace was funded by an anonymous private donor, and the time spent on this paper was funded in part by grants K24DA046569 (principal investigator: MT-S), R61DA059860 (principal investigator: MT-S), and K23DA050798 (principal investigator: JBF). The authors would like to thank Michelle Porche, EdD, for consultation on qualitative analysis plan development; Chorus Innovations for partnership related to technology development; and the wider FostrSpace team for their contributions in developing FostrSpace.

Data Availability

The datasets generated during and analyzed during this study is not publicly available due to ethical concerns about possible reidentification with this small study sample. Study metadata (eg, reflection questions and codebook) are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

IE, TL, JBF, DM, JT, AP, and MT-S conceptualized the paper. IE and TL carried out the selection of articles, data analysis, and data interpretation. IE, TL, DM, JT, and AP contributed to writing the original draft. IE, TL, JBF, DM, JT, AP, MVDC, and MT-S were involved in writing, reviewing, and editing the manuscript. MT-S and JBF contributed equally to the senior authorship role and as such are listed as co-senior authors.

Conflicts of Interest

None declared.

References

1. National foster care month outreach toolkit. Child Welfare Information Gateway. URL: https://www.childwelfare.gov/resources/national-foster-care-month-outreach-toolkit/ [accessed 2023-06-09]
2. Title IV-E foster care. Administration for Children & Families, U.S. Department of Health & Human Services. 2012 May 17. URL: https://www.acf.hhs.gov/cb/grant-funding/title-iv-e-foster-care [accessed 2023-04-28]
3. Kerker BD, Dore MM. Mental health needs and treatment of foster youth: barriers and opportunities. Am J Orthopsychiatry 2006 Jan;76(1):138-147 [FREE Full text] [doi: 10.1037/0002-9432.76.1.138] [Medline: 16569139]
4. Bevan Jones R, Stallard P, Agha SS, Rice S, Werner-Seidler A, Stasiak K, et al. Practitioner review: co-design of digital mental health technologies with children and young people. J Child Psychol Psychiatry 2020 Aug;61(8):928-940 [FREE Full text] [doi: 10.1111/jcpp.13258] [Medline: 32572961]
5. Porche MV, Folk J, Tolou-Shams M, Fortuna LR. Researchers' perspectives on digital mental health intervention co-design with marginalized community stakeholder youth and families. Front Psychiatry 2022;13:867460 [FREE Full text] [doi: 10.3389/fpsyt.2022.867460] [Medline: 35530032]
6. Hetrick SE, Robinson J, Burge E, Blandon R, Mobilio B, Rice SM, et al. Youth codesign of a mobile phone app to facilitate self-monitoring and management of mood symptoms in young people with major depression, suicidal ideation, and self-harm. JMIR Ment Health 2018 Jan 23;5(1):e9 [FREE Full text] [doi: 10.2196/mental.9041] [Medline: 29362208]
7. Hodson E, Dadashi N, Delgado R, Chisholm C, Sgrignoli R, Swaine R. Co-design in mental health; Mellow: a self-help holistic crisis planning mobile application by youth, for youth. Des J 2019 May 31;22(sup1):1529-1542 [FREE Full text] [doi: 10.1080/14606925.2019.1594975]
8. Potapov K, Marshall P. LifeMosaic: co-design of a personal informatics tool for youth. In: Proceedings of the Interaction Design and Children Conference. 2020 Presented at: IDC '20; June 21-24, 2020; London, UK. [doi: 10.1145/3392063.3394429]
9. Shroff A, Roulston C, Fassler J, Dierschke NA, Todd JS, Ríos-Herrera Á, et al. A digital single-session intervention platform for youth mental health: cultural adaptation, evaluation, and dissemination. JMIR Ment Health 2023 Feb 14;10:e43062 [FREE Full text] [doi: 10.2196/43062] [Medline: 36787180]
10. Arevian AC, O'Hora J, Jones F, Mango J, Jones L, Williams PG, et al. Participatory technology development to enhance community resilience. Ethn Dis 2018;28(Suppl 2):493-502 [FREE Full text] [doi: 10.18865/ed.28.S2.493] [Medline: 30202203]
11. Arevian AC, O'Hora J, Rosser J, Mango JD, Miklowitz DJ, Wells KB. Patient and provider cocreation of mobile texting apps to support behavioral health: usability study. JMIR Mhealth Uhealth 2020 Jul 29;8(7):e12655 [FREE Full text] [doi: 10.2196/12655] [Medline: 32723714]
12. Palomo-Carrión R, Romay-Barrero H, Pinero-Pinto E, Romero-Galisteo RP, López-Muñoz P, Martínez-Galán I. Early intervention in unilateral cerebral palsy: let's listen to the families! What are their desires and perspectives? A preliminary family-researcher co-design study. Children (Basel) 2021 Aug 30;8(9):750 [FREE Full text] [doi: 10.3390/children8090750] [Medline: 34572182]
13. Torous J, Roberts LW. Needed innovation in digital health and smartphone applications for mental health: transparency and trust. JAMA Psychiatry 2017 May 01;74(5):437-438. [doi: 10.1001/jamapsychiatry.2017.0262] [Medline: 28384700]
14. Liverpool S, Mota CP, Sales CM, Čuš A, Carletto S, Hancheva C, et al. Engaging children and young people in digital mental health interventions: systematic review of modes of delivery, facilitators, and barriers. J Med Internet Res 2020 Jun 23;22(6):e16317 [FREE Full text] [doi: 10.2196/16317] [Medline: 32442160]
15. The national network for youth in foster care. The National Foster Care Youth & Alumni Policy Council. URL: https://www.nationalpolicycouncil.org/ [accessed 2023-04-23]
16. iFoster homepage. iFoster. URL: https://www.ifoster.org/ [accessed 2023-04-27]
17. Miller JJ, Chih MY, Washington E. Conceptualizing a mobile app for foster youth transitioning to adulthood: a mixed-method approach. J Technol Human Serv 2016 May 06;34(2):145-170. [doi: 10.1080/15228835.2015.1108260]
18. FostrSpace homepage. FostrSpace. URL: https://www.joinchorus.com/fostrspace [accessed 2025-01-01]
19. Rhee K, Dankwa-Mullan I, Brennan V, Clark C. What is TechQuity? J Health Care Poor Underserved 2021;32(2S):xiii-viii. [doi: 10.1353/hpu.2021.0045]

XSL•FO
RenderX

20. Aldana A, Richards-Schuster K. Youth-led antiracism research: making a case for participatory methods and creative strategies in developmental science. J Adolesc Res 2021 Sep 18;36(6):654-685 [FREE Full text] [doi: 10.1177/07435584211043289]
21. Sawyer RD, Norris J. Duoethnography: Understanding Qualitative Research. Oxford, UK: Oxford University Press; 2012.
22. Breault RA. Emerging issues in duoethnography. Int J Qual Stud Educ 2016 Apr 04;29(6):777-794. [doi: 10.1080/09518398.2016.1162866]
23. Thematic analysis. The University of Auckland. URL: https://www.thematicanalysis.net/ [accessed 2022-10-01]
24. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
25. What is thematic analysis? Victoria Clarke YouTube page. URL: https://www.youtube.com/watch?v=4voVhTiVydc [accessed 2024-12-19]
26. Byrne D. A worked example of Braun and Clarke's approach to reflexive thematic analysis. Qual Quant 2021 Jun 26;56(3):1391-1412. [doi: 10.1007/s11135-021-01182-y]
27. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019 Jun 13;11(4):589-597. [doi: 10.1080/2159676x.2019.1628806]
28. ATLAS.ti Mac and Windows (version 22) [Qualitative data analysis software]. ATLAS.ti Scientific Software Development GmbH. URL: https://atlasti.com [accessed 2022-12-13]
29. Ding X, Wuerth K, Sakakibara B, Schmidt J, Parde N, Holsti L, et al. Understanding mobile health and youth mental health: scoping review. JMIR Mhealth Uhealth 2023 Jun 16;11:e44951 [FREE Full text] [doi: 10.2196/44951] [Medline: 37220197]

Abbreviations

FAB: FostrSpace Advisory Board
JJBH: Juvenile Justice Behavioral Health
TA: thematic analysis
UCSF: University of California, San Francisco
YPAR: youth participatory action research

Edited by A Mavragani; submitted 29.09.23; peer-reviewed by M Simmons, J D'Arcey; comments to author 18.12.23; revised version received 29.02.24; accepted 20.11.24; published 20.01.25.

Please cite as:
Ezimora I, Lundberg T, Miars D, Trujeque J, Papias A, Del Cid MV, Folk JB, Tolou-Shams M
Reflections of Foster Youth Engaging in the Co-Design of Digital Mental Health Technology: Duoethnography Study
JMIR Form Res 2025;9:e53231
URL: https://formative.jmir.org/2025/1/e53231
doi:10.2196/53231
PMID:

©Ifunanya Ezimora, Tylia Lundberg, Dylan Miars, Jeruel Trujeque, Ashley Papias, Margareth V Del Cid, Johanna B Folk, Marina Tolou-Shams. Originally published in JMIR Formative Research (https://formative.jmir.org), 20.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Developing a Machine Learning–Based Automated Patient Engagement Estimator for Telehealth: Algorithm Development and Validation Study

Pooja Guhan[1], PhD; Naman Awasthi[1], PhD; Kathryn McDonald[2], BS; Kristin Bussell[3], PhD; Gloria Reeves[2], MD; Dinesh Manocha[1], PhD; Aniket Bera[4], PhD

[1]Department of Computer Science, University of Maryland, College Park, MD, United States

[2]Department of Psychiatry, Child and Adolescent Division, University of Maryland, Baltimore, MD, United States

[3]School of Nursing, University of Maryland, Baltimore, MD, United States

[4]Department of Computer Science, Purdue University, West Lafayett, IN, United States

Corresponding Author:
Pooja Guhan, PhD
Department of Computer Science
University of Maryland
8125 Paint Branch Dr
College Park, MD, 20742
United States
Phone: 1 2406309133
Fax: 1 (301) 405 6707
Email: pguhan@umd.edu

Abstract

Background: Patient engagement is a critical but challenging public health priority in behavioral health care. During telehealth sessions, health care providers need to rely predominantly on verbal strategies rather than typical nonverbal cues to effectively engage patients. Hence, the typical patient engagement behaviors are now different, and health care provider training on telehealth patient engagement is unavailable or quite limited. Therefore, we explore the application of machine learning for estimating patient engagement. This can assist psychotherapists in the development of a therapeutic relationship with the patient and enhance patient engagement in the treatment of mental health conditions during tele–mental health sessions.

Objective: This study aimed to examine the ability of machine learning models to estimate patient engagement levels during a tele–mental health session and understand whether the machine learning approach could support therapeutic engagement between the client and psychotherapist.

Methods: We proposed a multimodal learning-based approach. We uniquely leveraged latent vectors corresponding to affective and cognitive features frequently used in psychology literature to understand a person's level of engagement. Given the labeled data constraints that exist in health care, we explored a semisupervised learning solution. To support the development of similar technologies for telehealth, we also plan to release a dataset called Multimodal Engagement Detection in Clinical Analysis (MEDICA). This dataset includes 1229 video clips, each lasting 3 seconds. In addition, we present experiments conducted on this dataset, along with real-world tests that demonstrate the effectiveness of our method.

Results: Our algorithm reports a 40% improvement in root mean square error over state-of-the-art methods for engagement estimation. In our real-world tests on 438 video clips from psychotherapy sessions with 20 patients, in comparison to prior methods, positive correlations were observed between psychotherapists' Working Alliance Inventory scores and our mean and median engagement level estimates. This indicates the potential of the proposed model to present patient engagement estimations that align well with the engagement measures used by psychotherapists.

Conclusions: Patient engagement has been identified as being important to improve therapeutic alliance. However, limited research has been conducted to measure this in a telehealth setting, where the therapist lacks conventional cues to make a confident assessment. The algorithm developed is an attempt to model person-oriented engagement modeling theories within machine learning frameworks to estimate the level of engagement of the patient accurately and reliably in telehealth. The results are encouraging and emphasize the value of combining psychology and machine learning to understand patient engagement. Further testing in the real-world setting is necessary to fully assess its usefulness in helping therapists gauge patient engagement during

XSL•FO
RenderX

online sessions. However, the proposed approach and the creation of the new dataset, MEDICA, open avenues for future research and the development of impactful tools for telehealth.

(JMIR Form Res 2025;9:e46390) doi:10.2196/46390

KEYWORDS

machine learning; mental health; telehealth; engagement detection; patient engagement

Introduction

Overview

The World Health Organization defines mental health as a "state of well-being" that allows a person to lead a fulfilling and productive life and contribute to society [1]. In addition, the World Health Organization estimates that one-fourth of the adult population is affected by a mental disorder [2]. However, there are only approximately 9 psychiatrists per 100,000 people in higher-income countries and only approximately 0.1 psychiatrists for every 1,000,000 people in lower-income countries [3,4]. As stress and pressure continue to increase, leading to poor mental health outcomes, the need for improved tele–mental health care has become critical. Tele–mental health care is the process of providing psychotherapy remotely, typically using Health Insurance Portability and Accountability Act (HIPAA)–compliant videoconferencing [5]. It offers an effective means of accessing mental health services and treatment, transcending geographical and cultural boundaries worldwide, and helps address the chronic shortage of psychotherapists. These services not only remove practical barriers to care, such as transportation, but also offer affordability and direct access to qualified therapists. Therefore, there has been an upward trend in the demand for such services [6]. Despite these undeniable benefits, this emerging treatment modality raises new challenges in patient engagement compared to in-person care. By "engagement," we refer to the connection between a therapist and patient, characterized by a sense of basic trust and a willingness or interest to collaborate. This connection is essential for the therapeutic process and fosters the development of a strong therapeutic relationship.

Background

Patient engagement is a critical priority in behavioral health care as it involves establishing effective rapport between health care providers and patients. In the context of mental health, patient engagement is an indicator of a successful therapy session. This multidimensional concept [7,8] encompasses the interaction of cognitive and emotion-related components of the patient's psychological state. However, one of the challenges in fully understanding and evaluating patient engagement lies in the difficulty of establishing a standardized measurement approach. Patient engagement is a complex and dynamic construct that defies a one-size-fits-all measurement. The diverse factors influencing patient engagement, such as individual characteristics, and health care contexts make it challenging to develop a universally accepted standard of measuring its effectiveness. Traditional quantitative metrics (appointment attendance rates, treatment adherence, and patient satisfaction) often fall short in capturing the depth and richness of patient engagement. While they can provide numerical data, they may overlook the subjective experiences of the interaction and do not necessarily reflect health care provider-patient alliance in treatment. In contrast, qualitative methods such as patient narratives and feedback provide valuable insights but lack scalability, objectivity, and uniformity across different health systems. Therefore, to address these limitations, machine learning techniques are being explored to complement and enhance these traditional methods by leveraging computational power, scalability, and ability to identify patterns, relationships, and insights that may not be otherwise apparent. Some prior works in engagement detection have focused on using a single modality, such as facial expressions [9,10], speech [11], body posture [12], gaze direction [13], or head pose [14], to detect engagement. Combining different modalities has been observed to improve engagement detection accuracy [15-17]. The authors expanded the work of Stanford PBL Lab's eRing [18] by including information streams such as facial expressions, voice, and other biometric data. Monkaresi et al [19] proposed an approach to detect engagement levels in students during a writing task by not only making use of facial features but also features obtained from remote video-based detection of heart rate. The dataset used was generated by the authors, and they used self-reports instead of external annotation for classification purposes. Chang et al [20] made use of facial expressions as well as body posture for detecting engagement in learners. Fedotov et al [21] proposed the use of audio, facial, and body pose features to detect engagement and disengagement for an in-the-wild dataset that had been created using videos obtained from uncontrolled environments or situations. The distribution of class labels in this dataset is not balanced.

Despite the existence of a variety of such algorithms to perform engagement detection, the results obtained from these approaches (particularly single modality–based approaches) could be misleading in a telehealth setting due to factors such as camera position and resistant or guarded clients. In telehealth appointments, therapists have limited visual data available to them, as they can only view the patient's face rather than their full body. Asking the patient to adjust their position to gain a better view of their body during telehealth appointments is a potential solution for therapists. However, this approach may lead to communication difficulties, as it can impact the patient's proximity to the microphone, potentially affecting the quality of audio communication between the patient and the therapist. Therefore, therapists must rely on verbal communication strategies to engage patients than in-person care because they cannot use typical nonverbal cues to convey interest and be responsive to the patient (eg, a handshake at the beginning of a session, adjusting the distance between the patient and health care provider by moving a chair closer or further away, and observing a patient's response to questions while maintaining eye contact). It is also more difficult for therapists to convey

XSL•FO
RenderX

attentiveness because eye contact requires the therapist to look at a camera rather than observe or look at a person. They may also have limited training on telehealth patient engagement, leading to a lack of guidance on measuring patient engagement. Therefore, this highlights the need for a system that can provide feedback on patient engagement by just using the data that are easily accessible, namely, text, audio, and face visuals. Engagement is critical for both retention of patients in care as well as effectiveness of mental health treatment. Developing such a system will make it possible to enhance the quality of tele–mental health and ultimately improve patient outcomes.

Objective

Taking all this into consideration, we propose and explore a machine learning–based approach that takes visual, audio, and text data as input and estimates the engagement level of a patient with mental health problems through a regression-based approach. We demonstrate the effectiveness of our method in tele–mental health sessions and provide a new dataset called Multimodal Engagement Detection in Clinical Analysis (MEDICA) to advance mental health research by understanding patient engagement levels during therapy sessions. MEDICA consists of 1229 short video clips from mock mental health therapy sessions used by medical schools in their psychiatry teaching curriculum. To the best of our knowledge, MEDICA is the first multimodal dataset that focuses on mental health research and consists of annotations useful for understanding the psychological state of patients with mental health problems during a therapy session.

Methods

Datasets

MEDICA Dataset

Overview

Engagement is an overloaded term, and the definition varies with the application, making it difficult and expensive to collect, annotate, and analyze such data. As a result, we find too few multimodal-based engagement detection datasets currently available for us to use. Our problem statement revolves specifically around detecting patient engagement during a tele–mental health session. In such a setting, the only information we can work with includes the patient's face and speech (audio and text). There exist datasets such as Carnegie Mellon University Multimodal Corpus of Sentimental Intensity (CMU-MOSI) [22], Carnegie Mellon University Multimodal Opinion Sentiment and Emotion Intensity (CMU-MOSEI) [23], and Stanford Emotional Narratives Dataset (SEND) [24] that capture such settings. However, they are not specifically for engagement detection. Given the lack of a dataset that allows researchers to use multimodal features (video, text, and audio) for engagement, we propose MEDICA, a novel dataset developed specifically to cater to engagement detection using tele–mental health session videos. To use these data to address a broader range of issues related to mental health, we also include labels pertaining to stress and emotions.

To the best of our knowledge, this dataset is one of the first publicly available datasets specifically catering to multimodal research in patient engagement in mental health. Table 1 presents a comparison between MEDICA and other related datasets. Despite the rise in telehealth services and the poor patient-to-therapist ratios for individuals with mental health problems, there are no datasets that even try modeling telehealth sessions to give the community an opportunity to innovate and develop new technologies. MEDICA is a humble attempt by us to kick-start interesting research opportunities. The MEDICA dataset is unique in that it contains multimodal data from multiple sensors including speech and video. These data have been annotated with the engagement levels exhibited by each participant at different points in time.

XSL•FO
RenderX

Table 1. Comparison of the Multimodal Engagement Detection in Clinical Analysis (MEDICA) dataset with other related datasets[a].

Dataset name	Samples	Unique speakers	Modes	Emotion labels	Engagement information	Other mental health cues
RECOLA[b] [25]	3400	27	{v,a}	Yes	No	Physiological (electrocardiogram and electrothermal activity)
CMU MOSEAS[c] [26]	715	Multiple	{v,a,t}	Yes	No	—[d]
CMU-MOSI[e] [22]	2199	Multiple	{v,a,t}	Yes	No	—
CMU-MOSEI[f] [23]	3228	1000	{v,a,t}	Yes	No	—
SEND[g] [24]	193	49	{v,a,t}	Yes	No	—
DAiSEE[h] [27]	9068	112	{v}	No	Yes	—
HBCU[i] [9]	120	34	{v}	No	Yes	—
In the wild [28]	195	78	{v}	No	Yes	—
SDMATH [29]	20	20	{v,a}	Yes	Yes	—
MEDICA	1229[b]	13	{v,a,t}	Yes	Yes	Hesitation, stress, and attentiveness

[a]Modes indicate the subset of modalities present from visual (v), audio (a), and text (t).

[b]RECOLA: Remote Collaborative and Affective.

[c]CMU MOSEAS: Carnegie Mellon University Multimodal Opinion Sentiment, Emotions and Attributes.

[d]Not applicable.

[e]CMU-MOSI: Carnegie Mellon University Multimodal Corpus of Sentiment Intensity.

[f]CMU-MOSEI: Carnegie Mellon University Multimodal Opinion Sentiment and Emotion Intensity.

[g]SEND: Stanford Emotional Narratives Dataset.

[h]DAiSEE: Dataset for Affective States in E-Environments.

[i]HBCU: historically black college and university.

[j]Current status of the dataset. The size of the dataset will be increased.

Acquisition

MEDICA has been developed by collecting publicly available mock therapy session videos created by different psychiatry medical schools for training their students. The patients in these videos are being advised for depression, social anxiety, and posttraumatic stress disorder. We collected 13 videos, each having a duration of approximately 20 to 30 minutes. We limit the videos to the setup where both the therapist and the patient are not visible together in the same frame. In addition, we take only those videos where there is only 1 patient. Each video has a unique English-speaking patient. The therapists in these videos are actual therapists, while the patients are trained actors who have been dictated to act and react in a particular way. Figure 1 shows the snapshots taken from various video clips in the MEDICA dataset.

Figure 1. Examples from the Multimodal Engagement Detection in Clinical Analysis (MEDICA) dataset created for mental health research. This dataset has been created using publicly available videos that are usually used for training purposes by different medical schools. The faces of the individuals have been blurred to protect their identity.

Processing and Annotation

Our primary focus was to create a dataset that captures the behavior of patients with mental health problems during their sessions. Therefore, we selected only portions of the videos where only the patient was visible in the frames, which appeared at various intervals and durations throughout the recordings. We took these scattered clips and divided them into smaller clips of 3 seconds each, resulting in a dataset of size 1229. We use Moviepy and speech-recognition libraries to extract audio and text from the video clips, respectively. Each video was annotated for attentiveness, stress, and engagement, which were

XSL•FO
RenderX

scored on a Likert scale ranging from –3 to 3; hesitation was a binary target variable (yes or no). Humans tend to have multiple emotions with varying intensities while expressing their thoughts and feelings. Therefore, the videos have been labeled for 8 emotions related to mental health: happy, sad, irritated, neutral, anxious, embarrassed, scared, and surprised. This will enable us to develop systems capable of understanding the various interacting emotions of the users.

Real-World Data

The videos collected to develop MEDICA were created by the psychiatry schools in controlled settings. By controlled, we mean that the therapists in these videos were real, while the patients were trained method actors who had been instructed in a very detailed manner regarding what to expect and how they should react. While these videos were approved by a team of psychotherapists as being inclusive of potential scenarios and patient reactions, the work would still be meaningless if not tested in the wild, that is, in settings where we have real patients who have not been instructed what to do or how to behave during the session with their real therapist. To accomplish this, we collaborated with 8 child psychotherapists. These psychotherapists are individuals who were trained and licensed as psychotherapists for both children and adults. The administrative staff identified patients that were scheduled for sessions with any of the participating therapists and provided the caregiver contact information. Our research assistant contacted eligible caregivers and provided information related to the purpose, potential risks and benefits, and responsibilities of study participation. They were informed of the study protocol that involved recording 1 to 2 telemental sessions to test our proposed approach. A total of 20 caregivers provided consent to participate in the study. They were provided with instructions on the setup of the equipment that we provided to ensure a clean recording. “Clean” refers to a recording executed with a camera of good quality with appropriate lighting conditions. The equipment included a smartphone with a good quality camera, a ring light with a stand to ensure that the session was recorded in a well-lit environment, and a hot spot internet connection to ensure that the session occurred smoothly without any network glitches. They were also given the assurance regarding preserving the confidentiality of the data being collected. All video recordings were labeled only by study number, with all identifiable information removed and stored in a double password–protected database. Participants were also informed that any facial images other than the caregiver who participated in the session would be deidentified (using methods such as blurring, etc). These steps were taken to ensure that only the caregiver’s level of engagement was measured and analyzed. In addition, this step protected the confidentiality of nonparticipants who did not provide consent.

On average, each of these sessions lasted approximately 20 minutes. The demographics of the caregivers who participated in our real-world experiments can be found in Multimedia Appendix 1. The entire data collection process can be divided into three parts:

1. Before each tele–mental health session of a caregiver with their therapist, a research assistant helped the caregiver with setting up the equipment to record their session. The assistant also ensured that the caregivers were comfortable using the equipment.
2. During the session, we made sure that the tele–mental health session proceeded in the same manner as a typical session. The caregivers were requested to use the equipment provided if possible, but there were no restrictions imposed by us on either the therapist or the caregiver regarding how the session should be conducted. After the presession process, the research assistant would log off. Therefore, during the session, it would be just the therapist and the caregiver having a conversation. No one else from the study would be a part of it. The only thing different about this session was that the caregiver was being recorded using the smartphone given to them. We did not record the therapist.
3. After the session, after the session was complete, a research assistant guided the participant regarding the steps to stop the recording and save the data collected.

Upon completion of the session, the therapist scored the quality of the collaborative relationship (therapeutic alliance) established between them during the session with the Working Alliance Inventory (WAI) [30,31]. It is used for both child and adult therapy and is a widely studied therapeutic alliance measure. The WAI was modeled on the theoretical work conducted by Bordin [32]. It captures 3 dimensions of the alliance: bond, task, and goals. Extensive tests showed 12 items per dimension to be the minimum length for effective representations of the inventory. A composite score is computed based on these 12 items for each of the sessions conducted. Henceforth, we refer to this score as the WAI score. The WAI scores can range from 12 to 84 [33,34]. Figure 2 shows frames from a few of the video recordings collected.

XSL•FO
RenderX

Figure 2. A few frames from the real-world videos we collected. The faces have been blurred here to protect the identity of the patients. However, the consent of the patients was taken to use their unblurred faces as input to our model.

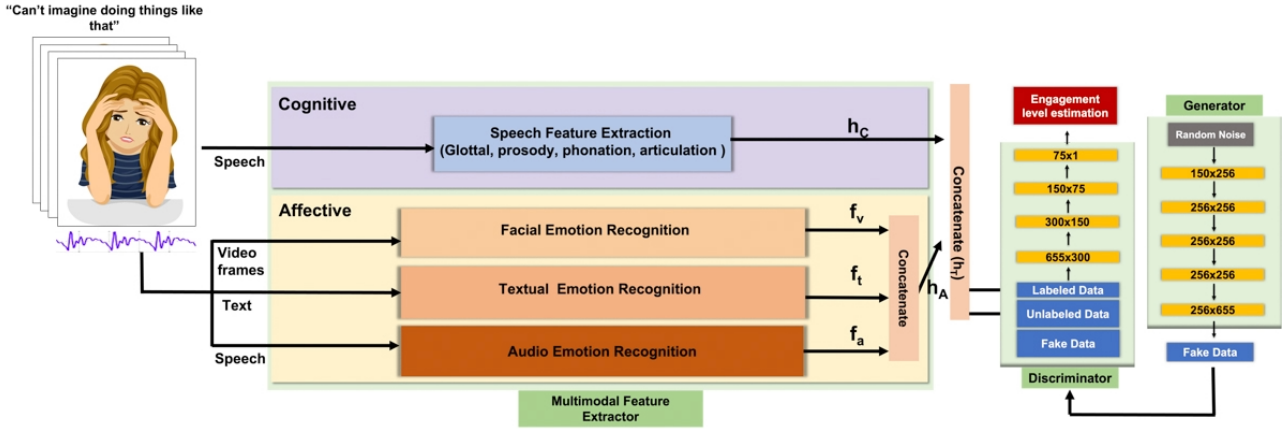
While efforts were taken to ensure clean video recordings, we still could not use 7 out of the 20 videos collected for any type of analysis. The reasons for it ranged from the video being extremely shaky due to the continuous movement of some participants during their session to not having the permissions to process and analyze videos where the child spoke for the entire session instead of the caregiver. The remaining videos were processed in the same way as MEDICA. At the end of data processing, we obtained approximately 438 clips, each lasting 3 seconds. Unlike MEDICA where we had annotations for every 3 seconds, we had only a single WAI score for the entire session. Therefore, all the 3-second clips obtained from a session were annotated with the WAI score obtained for the session. Any analysis conducted on these data treated all the 438 clips independently.

Proposed Model Design

Figure 3 presents an overview of our engagement estimation model. It has 2 key components: *multimodal feature extractor* and *semisupervised learning network*. We discuss each of them in detail in the subsequent sections.

XSL•FO
RenderX

Figure 3. Overall block diagram of the proposed architecture. The features obtained from the 2 modalities, cognitive and affective, are combined to form hT (real) of shape 655×1. The generator takes in random noise (Gaussian) and tries to generate a fake feature map that looks similar to hfakeT that looks similar to hT (real). The discriminator tries to distinguish between hT and hfakeT. The yellow bars in the discriminator and generator refer to linear layers with activation function as leaky rectified linear activation unit. The shape of the linear layers used has been written on the bars. The number of linear layers used can be changed depending upon the dataset size and task at hand.



Multimodal Feature Extractor

Overview

Multimodal feature extractor in general refers to a system or algorithm that extracts and combines features from multiple modalities, such as audio, video, text, and other types of data, to provide a more comprehensive representation of input data. The goal of multimodal feature extraction is to capture complementary information from different sources, which can lead to better performance in various tasks. In our model, the multimodal feature extractor is used to comprehend the psychological state of the patient with mental health problems. We take inspiration from literature in psychology and psychiatry to build an extractor that can enable the learning network to recognize, understand, and evaluate engagement as close as possible to that of a psychotherapist. Therefore, we proposed a multicomponent approach to build the feature extractor and introduce 2 modules, namely, cognitive and affective states, to capture the different cues used by psychotherapists to assess their patients. As mentioned earlier, the visual data available are limited. Therefore, apart from the visuals, we also need to make use of verbal information, that is, understanding the speaking style of the person (audio) and what the person said (text).

Therefore, given this information corresponding to a participant, the multimodal feature extractor outputs a feature vector h_T. h_T is obtained after concatenating the feature vectors h_C and h_A obtained from the cognitive and affective state modalities (components of the multimodal feature extractor), respectively. We now discuss the cognitive and affective state modalities in more detail in the subsequent sections.

Cognitive State Modality

The cognitive state involves comprehending complex concepts and issues and acquiring difficult skills. It conveys deep (rather than surface level) processing of information, where the person gains a critical or higher-order understanding of the subject matter and solves challenging problems.

Psychotherapists usually measure and evaluate the cognitive state of the person using the mental status examination, typically conducted via in-person interviews or self-evaluations, to gauge memory, thinking, and the extent of understanding of the topic of discussion. There has been a lot of work around determining biomarkers for detecting signs of a person's cognitive state. However, these methods are either offline or fail to consider various essential perceptual indicators. Recently, there has been a lot of work around using speech as a potential biomarker for detecting cognitive decline. For instance, stress negatively affects the cognitive functions of a person, and this can be easily detected using speech signals. Moreover, speech-based methods are attractive because they are nonintrusive, inexpensive, and can be done in real time. Four distinct features have proven to be highly effective in identifying signs of cognitive impairment and are increasingly utilized to detect conditions like Alzheimer's and Parkinson's diseases. These features include glottal (f_g), prosody (f_{pr}), phonation (f_{ph}), and articulation measures (f_{ar}), each offering unique insights into speech characteristics associated with cognitive decline.

Glottal features are acoustic measures used to characterize speech under stress. The glottal is the space between the vocal cords in the larynx, and the opening and closing of the vocal cords create sound during speech. During periods of stress, there is an aberration in the amount of tension applied in the opening (abduction) and closing (adduction) of the vocal cords. Prosody features characterize the speaker's intonation and speaking styles. Under this feature, we analyze variables such as timing, pitch, and loudness during the production of speech. Phonation refers to the production of vocal sounds by the vibration of the vocal cords. In individuals with cognitive decline, phonation is often characterized by bowing and inadequate closure of the vocal cords. These changes can lead to instability and irregularity in the vibration of the vocal cords, which in turn affects the production of speech. They are analyzed in terms of features related to perturbation measures, such as jitter (temporal perturbations of the fundamental frequency), shimmer (temporal perturbation of the amplitude of the signal), amplitude perturbation quotient, and pitch perturbation quotient. Apart from these, the degree of unvoiced is also included. Articulation

XSL•FO
RenderX

features refer to the movements of the lips, tongue, and jaw during speech production. Reduced amplitude and velocity of these movements can impact speech intelligibility. The analysis of articulation is primarily based on the first 2 vocal formants F1 and F2, which correspond to the frequency of the sound waves produced by the vocal tract.

For a given audio input, we extract these features using *librosa* [35] and *praat* [36] libraries. Therefore, we define features corresponding to cognitive state as the concatenation of these 4 audio features. Therefore, cognitive state features

$$h_c = \text{concat}\ (f_g, f_{pr}, f_{ph}, f_{ar})\ \textbf{(1)}$$

Affective State Modality

The affective state encompasses emotional reactions such as excitement, boredom, curiosity, and anger. The range of affective expressions will vary based on individual demographic factors (eg, age), cultural backgrounds or norms, and mental health symptoms.

To understand the affective state, we check if there exists any inconsistency between the emotions perceived and the statement the person made. Balomenos et al [37] and Porter and ten Brinke [38] suggest that when different modalities are modeled and projected onto a common space, they should point to similar affective cues; otherwise, the incongruity suggests distraction, deception, etc. In other words, if E1, E2, and E3 represent the emotions perceived individually from what the patient said (text), the way they said it or sounded (audio), and how they looked or expressed (visuals), respectively, then the patient would be considered engaged if E1, E2, and E3 are similar; otherwise, they would be considered disengaged. Therefore, we adopt pretrained emotion recognition models to extract affective features corresponding to audio, visuals, and text from each video sample separately.

Audio features (f_a) are extracted using mel-frequency cepstral coefficients (MFCC) from audio clips available in the dataset. MFCC is a commonly used audio feature extraction technique in signal processing and machine learning. It is based on the concept of the mel scale, which is a nonlinear transformation of frequency that is more perceptually relevant to human hearing. By applying MFCC to audio signals, we can extract features that capture important aspects of the sound, such as pitch, timbre, and spectral shape. These features are often used in speech and music analysis as well as in various machine learning applications, including emotion recognition and speaker identification. In our work, the extracted MFCC features from audio are then passed as input to a multilayer perceptron network trained for emotion recognition in speech, specifically using the data available in the Crowd-source Emotional Multimodal Actors Dataset (CREMA-D) to extract affective features from the audio data. Through this method, we obtained a feature vector f_a, for each audio clip, which provides valuable insights into the emotional characteristics of the speech.

Visual features (f_v) correspond to the affective features extracted from video frames to understand the emotions expressed through facial cues of the patient. We used a deep learning architecture called Visual Geometry Group-B (VCG-B), which has been pretrained as described in the study by Arriaga et al [39], to extract the affective features. We modified the output dimensions of the second last layer of the network to produce a feature vector f_v of length 100.

Text (f_t), characterizing the emotions, sentiment, and other affective cues present in the text, is obtained using a Bidirectional Encoder Representations from Transformers(BERT)–based model, which has been pretrained on the GoEmotions dataset [40]. The Bidirectional Encoder Representations from Transformers model works by processing the text input through a multilayer neural network that is trained to predict the context of each word based on its surrounding words. This allows the model to capture the complex relationships between words and their meanings, which are essential for understanding the affective context of the text.

Therefore, we represent the affective state of the patient as a concatenation of f_a, f_v, and f_t. Hence, affective state features

$$h_A = \text{concat}\ (f_a, f_v, f_t)\ \textbf{(2)}$$

Therefore, the overall output of the multimodal feature extractor is a feature vector h_T*which c*an be defined as

$$h_T = \text{concat}\ (h_C, h_A)$$

The learning network discussed in the next section makes use of this.

Learning Network

In this section, we discuss the second part of our proposed model, that is, the semisupervised learning network. Different machine learning techniques can be explored to solve the task of predicting the level of engagement. However, obtaining a large amount of high-quality labeled data to train a robust model for predicting patient engagement is inevitably laborious and requires expert medical knowledge. Considering that unlabeled data are relatively easy to collect, we propose a semisupervised learning–based solution. Semisupervised learning enables us to deploy machine learning systems in real-life applications (eg, image search [41], speech analysis [42,43], and natural language processing) where we have few labeled data samples and a lot of unlabeled data. Some prior works have also explored semisupervised learning for engagement detection in nonmedical domains. One of the earliest studies in this area is by Alyuz et al [44], where they consider the development of an engagement detection system, more specifically emotional or affective engagement of students, in a semisupervised fashion to personalize systems such as intelligent tutoring systems according to their needs. Nezami et al [45] conducted experiments to detect user engagement using a facial feature–based semisupervised model. Most state-of-the-art semisupervised learning methods use generative adversarial nets (GANs) [46].

GANs are a class of machine learning models and typically have 2 neural networks competing with each other to generate more accurate predictions. These 2 neural networks are referred to as the generator and the discriminator. The generator's goal is to artificially manufacture outputs (ie, fake data) that could easily be mistaken as real data. The goal of the discriminator is

XSL•FO
RenderX

to identify the real from the artificially generated data. In trying to generate high-quality outputs, the generator learns to capture the different possible variations in the input variables, and therefore, the data manifold well. This is extremely helpful when we may not be able to access data containing a wide variety of similar engagement-related cues visible across different patients. In general, the training process of GANs involves a dynamic interplay and competition between the generator and discriminator networks. Both the generator and discriminator networks improve their abilities iteratively. The generator aims to generate more realistic samples, while the discriminator aims to become better at distinguishing between real and generated (artificial) data. The eventual goal of the training process is for the generator and discriminator to reach Nash equilibrium [47]. Once the equilibrium is achieved, no player (generator or discriminator) has any incentive to deviate toward a more optimal position, that is, no player has a better strategy given the choices of the other player.

Taking all this into consideration, we propose using a multimodal semisupervised GAN as our learning network for regressing on the levels of patient engagement. This type of machine learning model combines multiple modalities (in this case, cognitive and affective modalities) to generate realistic samples (or features) that can fool the discriminator. The term *semisupervised* refers to the network's ability to use a limited amount of labeled data alongside a larger amount of unlabeled data to enhance its accuracy. Such a network will generalize better and be more robust compared to previously defined semisupervised learning approaches. It is different than the semisupervised GAN framework SR-GAN proposed by Olmschenk et al [48] in 2 main ways. First, our network is multimodal, whereas SR-GAN is not. Second, unlike the SR-GAN generator, which focuses on generating realistic images, our GAN's generator is designed to create realistic feature maps that mimic the multimodal feature vector h_T, derived from the fusion of cognitive and affective state modalities (as discussed in the previous section). The generator takes Gaussian noise as input and produces a synthetic feature vector h_{fakeT}, aiming to deceive the discriminator into classifying it as h_{fakeT}. The discriminator's role, in turn, is to distinguish h_T as a fake representation and correctly identify h_T as the genuine feature vector.

Overall, our multimodal semisupervised GAN can therefore be divided into 3 parts: multimodal feature extractor, generator, and discriminator.

For the purpose of training, we make use of the following four standard loss functions, namely, labeled loss (L_{lab}), unlabeled loss (L_{un}), fake loss (L_{fake}), and generator loss (L_{gen}):

1. L_{lab}—this loss is similar to typical supervised learning regression losses, where the goal is to assess the network's ability to estimate the engagement level accurately, as close as possible to the ground truth (ie, the actual engagement level annotated for the data sample). Therefore, we compute the mean squared error of model output (, ie, predicted level of engagement) with ground truth (y_t, ie, actual level of engagement).

2. L_{un}—in semisupervised GANs, the labeled dataset contains a small portion of data with labels, while most of the data are unlabeled. The goal of the L_{un} loss function is to improve the performance of the generator by minimizing the distance between the feature spaces of the labeled and unlabeled datasets. This helps to ensure that the generator produces feature maps that are consistent with both labeled and unlabeled data, thereby improving the generalization performance of the model. By minimizing the distance between the feature spaces of labeled and unlabeled datasets, the model can effectively learn from both labeled and unlabeled data, which is particularly useful in cases where labeled data is scarce.

3. L_{fake}—it is a loss function used to update the discriminator network. The objective of this loss is to maximize the distance between the feature space of unlabeled dataset and fake features generated by the generator. This enables the discriminator to distinguish the real and fake data features better.

4. L_{gen}—the generator network's objective is to produce feature vectors h_T (fake) that can be passed off as real feature vectors h_T: L_{gen} updates the generator network to do just that. L_{gen} encourages the generator to learn the underlying distribution of unlabeled data features (real) and generate fake features that match the feature statistics of the real features.

Similar to SR-GAN, we also make use of a gradient penalty (P) to keep the gradient of the discriminator in check, which helps convergence. The gradient penalty is calculated with respect to a randomly chosen point on the convex manifold connecting the unlabeled feature vector h_T to the fake feature vector h_T. We compute it as

Here, $p_{interpolate}$ examples are generated by $\alpha p_{unlabeled} + (1 - \alpha) p_{fake}$ for $\alpha \sim U(0,1)$.

$U(0,1)$ represents a uniform distribution over the range from 0 to 1.

The overall loss function used for training the network is:

$$L = L_{lab} + L_{un} + L_{fake} + L_{gen} + \lambda P \quad \textbf{(8)}$$

Overall Pipeline

In this section, we present an overview of how the various components of our proposed approach interact to estimate the engagement level of a patient with mental health problems during a tele–mental health session. We provide insights into the working mechanism of our model and shed light on the

XSL•FO
RenderX

underlying processes involved in engagement level estimation in a multimodal setting.

As depicted in Figure 3, the proposed model involves data preprocessing, which includes extracting video frames, audio signal, and text data from the input video clip. To capture engagement dynamics over small time scales, we analyze engagement at the microlevel, considering video clips of a few seconds duration. This aligns with the person-oriented analysis suggested by Sinatra et al [49]. Specifically, for the MEDICA dataset and real-world data, we consider video clips of 3 seconds duration. All samples, labeled and unlabeled, from the dataset are treated as *real* data for the semisupervised GAN. The multimodal feature extractor (consisting of the cognitive and affective state modalities) is used to obtain the h_T features from the preprocessed visual, audio, and text data. h_T is considered real. In parallel, the generator network creates h_{fakeT} (ie, fake h_T) features using gaussian noise as input. The labeled and unlabeled h_T features obtained from the multimodal feature extractor and the h_{fakeT} features obtained from the generator are then input to the discriminator network. In addition to distinguishing between real (h_T) and fake (h_{fakeT}) features, the discriminator network outputs the engagement level of the patient with mental health problems as a continuous value.

The subsequent sections outline the different studies being conducted to verify our approach.

Study 1: Testing Our Proposed Approach on MEDICA

The purpose of the first study is to test the applicability of our approach to estimate the levels of engagement exhibited by the participants present in the videos. The test was conducted on the MEDICA dataset. In addition, we compared the performance of the proposed model against other similar frameworks to understand its effectiveness. To test the models, we partitioned the dataset in a ratio of 70:10:20 for training, testing, and validation. Motivated by recent works in clinical psychotherapy [50-52], we used the standard evaluation metric of root mean square error (RMSE) to evaluate our approach. Smaller the RMSE value, better is the performance of the model for the engagement estimation task.

Study 2: Ablation Studies

The proposed model design involves the introduction of 2 modalities, namely, affective and cognitive. In this study, we explored the contribution of these 2 modalities for the purpose of engagement detection. We aimed to test the effectiveness of the model with and without these modalities. Therefore, we ran the overall pipeline described earlier by removing either one of the modalities corresponding to affective and cognitive states and reported our findings. The tests were conducted on the MEDICA dataset. Similar to our previous study, we performed the evaluation using RMSE.

Study 3: Analysis on Real-World Data

Evaluating the model only on MEDICA is not sufficient to prove the usability of the model in the real-world setting. Therefore, this study aimed to test our approach on data collected from real-world scenarios. The study involved testing the proposed model that has been trained on MEDICA using real-world data. We also compared the performance of our proposed model against other baselines explored in study 1 on the real-world data. Apart from testing our model's usability for the real world, we also tested its capability to perform on unseen real-world data, which is different from the data (MEDICA) it has been trained on. Contrary to previous experiments where we compared the performance on engagement estimation values, in this test, we will study the correlation of the model-estimated engagement values with the WAI score. This will help us to also test whether the scores estimated by our model are clinically useful for therapists.

Ethical Considerations

Both the MEDICA dataset and the real-world data collection processes have been approved by the institutional review boards (#HP-00092966) at both the University of Maryland College Park's computer science department and University of Maryland Baltimore medical school.

Participation of the caregivers while collecting the real-world data was completely voluntary. They could withdraw anytime without any negative consequences. They were informed about the purpose of the study and the analysis that would be conducted using the data collected.

The therapist and the participating caregiver were given the option of allowing us to store videotaped recordings for development of future studies or to request to have all videotape sessions destroyed after a 2-year period. The video recordings were stored in our laboratory databases and subject to strict University of Maryland privacy protocol, including encryption and password protection. While the research is ongoing, only select project personnel with internal clearance can have access to the data. In addition, we ensured that the data are untampered using standard cryptographic hash functions. During the video storage process, we deidentified any facial images beside the health care provider or caregiver who are recorded (eg, sibling interrupts session) so that they are not recognizable on the recorded video. All data related to enrollment, demographics, and questionnaires will be stored in an electronic database at the University of Maryland and will be double password protected with only select research members having access. The participating caregiver was given an individual ID number, and their data were deidentified to ensure it could not be linked back to them.

While the participating caregiver may not directly benefit from the study, they received 1-month free internet hot spot and a Fitbit, which were provided as part of the real-world study equipment. They were compensated with a US $20 gift card after returning the equipment.

Results

The semisupervised models were trained on an NVIDIA GeForce GTX 1080 Ti GPU with a batch size of 512 and a learning rate of 0.0001 for 50 epochs.

Study 1: Testing Our Proposed Approach on MEDICA

The purpose of the first study is to demonstrate the ability of our model to estimate the level of engagement exhibited by the

XSL•FO
RenderX

patient in the video. This study was performed on the MEDICA dataset. As our proposed methodology leverages a semisupervised approach, we extracted labeled samples from MEDICA and unlabeled samples from the CMU-MOSEI dataset. After preprocessing, we extracted 12,854 unlabeled data points from CMU-MOSEI. We split the 1229 labeled data points from MEDICA into 70:10:20 for training, validation, and testing, respectively. Therefore, the split of the labeled training data to unlabeled training data points was 860:12,854. In addition, the ground truth engagement levels corresponding to the training data points in the MEDICA dataset were normalized to fall within a range of 0 and 1. We compared our model with the following state-of-the-art methods for engagement detection:

1. Kaur et al [28] used a deep multiple instance learning–based framework for detecting engagement in students. They extracted local binary pattern on three orthogonal planes (LBP-TOP) features from the facial video segments and performed linear regression using a deep neural network to estimate the engagement scores.

2. Nezami et al [45] performed a semisupervised engagement detection using a semisupervised support vector machine.

In addition to being state-of-the-art, these methods can be used in a telehealth setting like ours. We used the publicly available implementation proposed by Kaur et al [28] and trained the entire model on MEDICA. Nezami et al [45] do not have a publicly available implementation. We reproduced the method to the best of our understanding.

Table 2 summarizes the RMSE values obtained for methods described by Kaur et al [28] and Nezami et al [45], including ours. We observed an improvement of at least 40%. Our approach is one of the first methods of engagement estimation built on the principles of psychotherapy. The modules used, specifically cognitive and affective states, help the overall model to effectively mimic the ways a psychotherapist perceives the patient's level of engagement. Similar to psychotherapists, these modules also look for specific engagement-related cues exhibited by the patient in the video.

Table 2. Comparison of RMSE values obtained from various methods for estimating levels of engagement on the Multimodal Engagement Detection in Clinical Analysis dataset.

Method	RMSE[a] for engagement
Kaur et al [28]	0.96
Nezami et al [45]	0.17
Our approach	0.10

[a]RMSE: root mean square error.

Study 2: Ablation Studies

To show the importance of the different components (affective and cognitive) used in our approach, we run our method on MEDICA by removing either one of the modules corresponding to affective or cognitive states and report our findings. Table 3 summarizes the results obtained from the ablation experiments. We observe that the ablated model (ie, only using affective [A] or cognitive [C] modules) does not perform as well as the model that includes both modules. To understand and verify the contribution of these modules further, we leveraged the other labels (stress, hesitation, and attention) available in MEDICA and performed regression tasks using our proposed architecture on all of them. We observed that mode C performed better when predicting stress and hesitation values. Mode A performed better in estimating a patient's level of attentiveness. These results agree with our understanding of cognitive state and affective state. Therefore, the combination of affective and cognitive state modes helps in efficiently predicting the engagement level of the patient.

Table 3. Ablation experiments on the Multimodal Engagement Detection in Clinical Analysis dataset. We ran our proposed model using only 1 modality at a time (either affective or cognitive) and compared the performance with that of using both modalities together.

Modality	RMSE[a] for engagement	RMSE for stress	RMSE for hesitation	RMSE for attention
Affective	0.24	0.15	0.146	0.07
Cognitive	0.3	0.13	0.16	0.08
Affective + cognitive (our approach)	0.10	0.12	0.14	0.1

[a]RMSE: root mean square error.

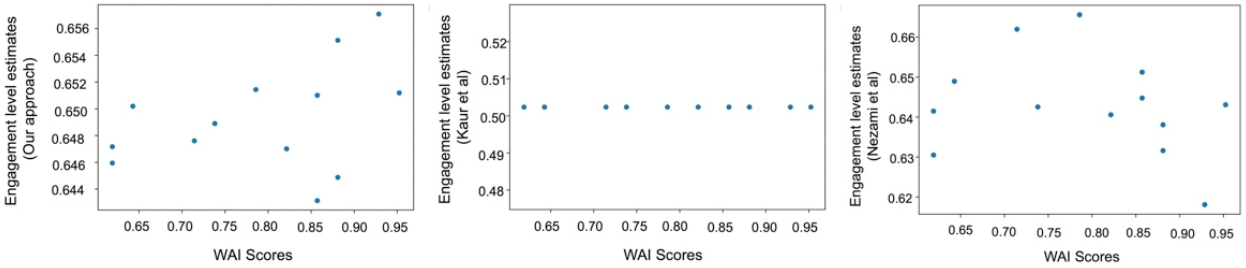
Study 3: Analysis on Real-World Data

Our model that has been trained for estimating engagement levels using the MEDICA dataset was tested on the processed real-world data. WAI scoring is based on certain observations the therapist makes during the session with the patient. The score obtained from our model was different from the WAI score, but we claim that similar to WAI, our estimates also captured the engagement levels of the patient well. If this is indeed the case, then both WAI and our estimates should be correlated. As discussed earlier, a single WAI score is reported by the therapist (health care provider) for the entire session. Unlike WAI, which reports a single score per session, our method, along with those proposed by Kaur et al [28] and Nezami et al [45], performs microanalysis, generating engagement estimates at multiple time points within a session.

XSL•FO
RenderX

To enable a fair comparison, we calculated the mean of the engagement estimates across each session for all methods. The mean engagement estimates across sessions were as follows: Kaur et al [28] had a mean of 0.50 (SD: 7.82e-17), Nezami et al [45] had a mean of 0.64 (SD: 0.017), and our approach had a mean of 0.65 (SD: 0.006). We subsequently examined each of their correlations with the corresponding WAI scores across sessions. Given that WAI scores and engagement estimates are measured on different scales, WAI scores were normalized to a 0-1 range for consistency with the engagement values from the methods. We found that Kaur et al [28] obtained a correlation score of –0.03, Nezami et al [45] obtained a score of –0.24 and our approach obtained a much better score of 0.38. We plot the engagement level estimations obtained for each session against its corresponding WAI score, as shown in Figure 4. The plot corresponding to our approach shows a noticeably better alignment with the WAI score trends compared to other methods [28,45].

Figure 4. Scatterplot between the engagement level estimations obtained from each method for each session: (A) the plot between engagement level estimations obtained from our approach and the Working Alliance Inventory (WAI) scores, (B) the plot between the engagement level estimations obtained by Kaur et al [28] for each session and the WAI scores, (C) the plot between the engagement level estimations for each session obtained by Nezami et al [45] and the WAI scores.



Additionally, instead of just taking the mean, we also took the median of the engagement level estimates available at different instances of the sessions. The average median of the engagement estimates across sessions for different methods were as follows: Kaur et al [28] had an average median of 0.50 (IQR 0.0), Nezami et al [45] had an average median of 0.64 (IQR 0.02), and our approach had an average median of 0.65 (IQR 0.007).

We then examined the median of the engagement estimates for each session, obtained using different methods, and assessed their correlation with the WAI scores. In this case, Kaur et al [28] obtained a correlation score of 0, Nezami et al [45] obtained a score of –0.18 and our approach obtained a relatively high score of 0.40. These findings reinforce that our model captures WAI patterns more effectively than prior methods. Its stronger positive correlation with the WAI scores indicates its capability to better estimate engagement levels during therapy sessions.

The conceptual model of our proposed approach is also supported by the theoretical work by Bordin [32]. According to this theory, the therapist-provider alliance is driven by 3 factors: bond, agreement on goals, and agreement on tasks; these factors align well with the features identified in this work. While bond would correspond with the affective state, goals and task agreement correspond with the cognitive state. The merit of the approach by Bordin [32] is that it has been used for child and adult therapy, and it is one of the more widely studied therapeutic alliance measures. Therefore, it is no surprise that our approach can work well to provide an estimate of engagement levels in a tele–mental health session.

Discussion

Principal Findings

In this study, we developed a machine learning–based model that integrates features related to cognitive and affective psychological states to predict patient engagement levels during tele–mental health sessions. The input to our proposed algorithm is the visual, audio, and text data available, while the output is the engagement level of the patient. Our investigation demonstrated the significant promise of our proposed method, achieving an average improvement of 40% in RMSE for predicting engagement levels on the newly introduced MEDICA dataset compared to prior works. In real-world settings, our method provided patient engagement estimates that closely aligned with the trends observed in WAI scores assigned by therapists based on their sessions [30,31]. This study highlights the potential of our multimodal engagement estimation algorithm to enhance the quality of telehealth services and ultimately improve patient outcomes.

Comparison With Prior Work

Patient engagement is recognized as a pivotal determinant of a successful therapy session, particularly in the realm of mental health care [53]. Current psychiatric literature [54-56] has extensively examined a range of strategies and methodologies designed to foster [57] and sustain patient engagement during therapy sessions. These strategies are crucial for achieving optimal therapeutic outcomes. There has been a growing interest in applying artificial intelligence (AI) to mental health [58-62], with innovative approaches being explored to enhance patient care. However, despite the proliferation of AI-based methods in mental health applications, there remains a noticeable gap in research specifically focused on understanding and quantifying engagement. This gap is even more evident in the context of tele–mental health [63]. The transition from traditional in-person therapy to digital platforms has introduced unique challenges, as conventional indicators of engagement, such as body language [64], eye contact [65], and physical presence, maybe less discernible or entirely absent in web-based settings.

XSL•FO
RenderX

Our research addresses this critical gap by pioneering an approach that estimates patient engagement specifically in the tele–mental health context. This work is the first of its kind to target engagement estimation in this setting, offering a novel perspective that has largely been overlooked in previous studies. While the broader concept of engagement has been explored by AI in various domains, existing methods [28,45,66] have not been adapted or tested within the unique demands of mental health care, particularly in a remote or web-based environment. From an algorithmic standpoint, previous works have explored both unimodal and multimodal learning-based networks to understand engagement. Unimodal approaches focus on individual modalities, such as speech, body posture, biometric data (eg, heart rate variability and skin conductance) [67,68], gaze direction [69], and head pose [70]. Multimodal approaches [71] integrate various signals, leveraging the complementary strengths of each modality to provide a more holistic understanding of engagement. In contrast to these works, our research focuses on using the types of information that are readily available during a tele–mental health session, specifically audio, video, and text data. This approach is designed to be practical and easily implementable in real-world telehealth settings. This also ensures that the engagement estimation process is minimally disruptive to the therapeutic experience during the telehealth session. We validated our approach using our newly introduced dataset, MEDICA. In addition, we tested it on data collected from real-world therapy sessions. Our method achieved state-of-the-art performance in both cases.

Strengths and Implications

In this work, we explored the multidimensional and temporally dynamic nature of patient engagement [72]. Our method estimates engagement at regular short intervals. These characteristics of our approach align perfectly with the person-oriented analysis discussed by Sinatra et al [49]. We released a new dataset, MEDICA, to enhance mental health research, specifically toward understanding the engagement levels of patients attending therapy sessions. To the best of our knowledge, there is no other multimodal dataset that caters specifically to the needs of the mental health–based research. In addition, while there are some image-based [9,27,28] or sensor information–based datasets, there is no dataset that addresses the possibility of exploring engagement detection using visual, audio, and text modalities. We verified the usefulness of our machine learning method using 3 studies. Study 1 demonstrates the ability of our model to estimate the engagement level exhibited by the patient in video. It was done to understand the usefulness of the proposed model in comparison to other existing engagement estimation methods such as those by Nezami et al [45] and Kaur et al [28]. Our approach reported an RMSE [50-52] of 0.1, which on an average is 40% less than the RMSE reported by other methods (Table 1). Our study is one of the first to estimate engagement specifically for patients with mental health problems in a tele–mental health setting. The modalities used, specifically cognitive and affective states, propel the overall algorithm to mimic the way a psychotherapist perceives the patient's level of engagement. As part of study 2, we were interested to understand the contribution of the 2 modalities (cognitive and affective states). The experiments conducted as part of this study are also termed as *ablation* as it involves running the network pipeline using only 1 of the 2 modalities available at a time. We leveraged the annotations in MEDICA on stress, hesitation, and attention to dig deeper regarding the contribution of the cognitive and affective state modalities. We ran the same ablation experiments on not only patient engagement but also stress, hesitation, and attention estimation. While the combination of affective and cognitive states gives better results (RMSE of 0.1), it was interesting to note that the observations obtained supported the theoretical discussion about the 2 states. Cognitive state modality was able to understand stress (RMSE of 0.13) better. Affective state that was built on the concept of capturing distraction using emotion inconsistency between visual, audio, and text data helped relate better with the understanding of attention (RMSE of 0.07). These results agree with our understanding of cognitive state and affective state discussed earlier.

WAI is a popular measure of patient engagement used by many mental health care providers. Therefore, in study 3, we wished to understand and check if the values being estimated by our method related well with the trends observed in WAI. We did so by computing the correlation scores between WAI and the values estimated by our proposed model. In addition, to test the effectiveness of our approach, we also compared the correlation scores obtained between WAI and patient engagement values estimated by other existing methods. Positive values are preferred because they indicate similarity in distribution behavior. Negative values would indicate that the trends observed in WAI and the engagement values predicted are different and opposite. The results (Table 2) revealed that our model was able to capture the trends in WAI much better than other engagement estimation methods, reinforcing its clinical relevance. The correlation strength observed between WAI and our method's estimated engagement values is better than other approaches, but it also indicates the need for further investigation. The trend is positive, suggesting that we have successfully modeled engagement to some extent. Further analysis will provide opportunities to enhance and strengthen the relationship between WAI and the engagement values estimated by machine learning.

Limitations and Future Work

One of the primary limitations of the proposed algorithm is that engagement predictions may not be optimal in case of occlusions or missing modalities, data corruption due to low internet bandwidth, and integration of wearable devices with our model. We recognize that this situation is likely to occur in a tele–mental health session and plan to incorporate solutions for it in our future work. In addition, we aim to explore ways to make the predictions more explainable, allowing psychotherapists to receive evidence-based suggestions to support their final decisions.

Conclusions

Telehealth behavioral services, delivered via videoconferencing systems, have emerged as a cost-effective, dependable, and secure option for mental health treatment, particularly in recent times. However, gauging patient engagement, a critical mental

XSL•FO
RenderX

health care standard, remains challenging in telehealth due to the absence of nonverbal cues and sensory data such as heart rate. To address this, we propose a novel multimodal semisupervised GAN that leverages affective and cognitive features from psychology to estimate engagement levels using visuals, audio, and text from video calls. This approach can significantly enhance social interactions and assist psychotherapists during tele–mental health sessions. Engagement is typically assessed through patient reports, which are susceptible to response bias, and current measures such as "show rate" and "patient satisfaction" do not accurately reflect the health care provider–patient alliance. Our model demonstrates its effectiveness on both our newly introduced dataset called MEDICA as well as real-world data. Given the lack of systematic engagement measurement and the limited training for health care providers on telehealth engagement, our proposed system offers a promising solution to these challenges, promoting better patient-health care provider interactions and making telehealth more effective.

Acknowledgments

This research was fully supported by the MPower grant.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Demographic information on the real-world experiment participants.
[DOCX File , 14 KB - formative_v9i1e46390_app1.docx]

References

1. WHO highlights urgent need to transform mental health and mental health care. World Health Organization. 2022 Jun 17. URL: https://www.who.int/news/item/17-06-2022-who-highlights-urgent-need-to-transform-mental-health-and-mental-health-care [accessed 2024-11-12]
2. The world health report 2001: mental health; new understanding, new hope. World Health Organization. 2001. URL: https://iris.who.int/handle/10665/42390 [accessed 2024-11-12]
3. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2197-2223. [doi: 10.1016/S0140-6736(12)61689-4] [Medline: 23245608]
4. Oladeji BD, Gureje O. Brain drain: a challenge to global mental health. BJPsych Int 2016 Aug;13(3):61-63 [FREE Full text] [doi: 10.1192/s2056474000001240] [Medline: 29093905]
5. Telemental health/providers. Anxiety & Depression Association of America. URL: https://adaa.org/find-help/treatment-help/telemental-health-providers [accessed 2024-11-12]
6. Greenwood H, Krzyzaniak N, Peiris R, Clark J, Scott AM, Cardona M, et al. Telehealth versus face-to-face psychotherapy for less common mental health conditions: systematic review and meta-analysis of randomized controlled trials. JMIR Ment Health 2022 Mar 11;9(3):e31780 [FREE Full text] [doi: 10.2196/31780] [Medline: 35275081]
7. Kelders SM, van Zyl LE, Ludden GD. The concept and components of engagement in different domains applied to eHealth: a systematic scoping review. Front Psychol 2020;11:926 [FREE Full text] [doi: 10.3389/fpsyg.2020.00926] [Medline: 32536888]
8. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
9. Whitehill J, Serpell Z, Lin YC, Foster A, Movellan JR. The faces of engagement: automatic recognition of student engagement from facial expressions. IEEE Trans Affective Comput 2014 Jan 1;5(1):86-98. [doi: 10.1109/taffc.2014.2316163]
10. Murshed M, Dewan MA, Lin F, Wen D. Engagement detection in e-learning environments using convolutional neural networks. In: Proceedings of the IEEE International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, International Conference on Cloud and Big Data Computing, International Conference on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech). 2019 Presented at: DASC/PiCom/CBDCom/CyberSciTech 2019; August 5-8, 2019; Fukuoka, Japan. [doi: 10.1109/dasc/picom/cbdcom/cyberscitech.2019.00028]
11. Yu C, Aoki PM, Woodruff A. Detecting user engagement in everyday conversations. arXiv Preprint posted online on October 13, 2004 [FREE Full text] [doi: 10.21437/interspeech.2004-327]

XSL•FO
RenderX

12. Sanghvi J, Castellano G, Leite I, Pereira A, McOwan PW, Paiva A. Automatic analysis of affective postures and body motion to detect engagement with a game companion. In: Proceedings of the 6th International Conference on Human-Robot Interaction. 2011 Presented at: HRI '11; March 6-9, 2011; Lausanne, Switzerland. [doi: 10.1145/1957656.1957781]
13. Nakano YI, Ishii R. Estimating user's engagement from eye-gaze behaviors in human-agent conversations. In: Proceedings of the 15th International Conference on Intelligent User Interfaces. 2010 Presented at: IUI '10; February 7-10, 2010; Hong Kong, China. [doi: 10.1145/1719970.1719990]
14. Sharma P, Joshi S, Gautam S, Maharjan S, Khanal SR, Reis MC, et al. Student engagement detection using emotion analysis, eye tracking and head movement with machine learning. arXiv Preprint posted online on September 18, 2019 [FREE Full text] [doi: 10.1007/978-3-031-22918-3_5]
15. Psaltis A, Apostolakis KC, Dimitropoulos K, Daras P. Multimodal student engagement recognition in prosocial games. IEEE Trans Games 2018 Sep;10(3):292-303. [doi: 10.1109/tciaig.2017.2743341]
16. Grafsgaard JF, Wiggins JB, Boyer KE, Wiebe EN, Lester LC. Embodied affect in tutorial dialogue: student gesture and posture. In: Proceedings of the 16th International Conference on Artificial Intelligence in Education. 2013 Presented at: AIED 2013; July 9-13, 2013; Memphis, TN. [doi: 10.1007/978-3-642-39112-5_1]
17. Aslan S, Cataltepe Z, Dundar O, Esme AA, Ferens R. Learner engagement measurement and classification in 1:1 learning. In: Proceedings of the 13th International Conference on Machine Learning and Applications. 2014 Presented at: ICMLA 2014; December 3-6, 2014; Detroit, MI. [doi: 10.1109/icmla.2014.111]
18. Ma J, Fruchter R. eRing: Body Motion Engagement Detection and Feedback in Global Teams. Stanford, CA: Stanford University; Apr 2015.
19. Monkaresi H, Bosch N, Calvo RA, D'Mello SK. Automated detection of engagement using video-based estimation of facial expressions and heart rate. IEEE Trans Affective Comput 2017 Jan 1;8(1):15-28. [doi: 10.1109/taffc.2016.2515084]
20. Chang C, Zhang C, Chen L, Liu Y. An ensemble model using face and body tracking for engagement detection. In: Proceedings of the 20th ACM International Conference on Multimodal Interaction. 2018 Presented at: ICMI '18; October 16-20, 2018; Boulder, CO. [doi: 10.1145/3242969.3264986]
21. Fedotov D, Perepelkina O, Kazimirova E, Konstantinova M, Minker W. Multimodal approach to engagement and disengagement detection with highly imbalanced in-the-wild data. In: Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data. 2018 Presented at: MCPMD '18; October 16, 2018; Boulder, CO. [doi: 10.1145/3279810.3279842]
22. Zadeh A, Zellers R, Pincus E, Morency LP. MOSI: multimodal corpus of sentiment intensity and subjectivity analysis in online opinion videos. arXiv Preprint posted online on June 20, 2016 [FREE Full text]
23. Zadeh AB, Liang PP, Poria S, Cambria E, Morency LP. Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. 2018 Presented at: ACL 2018; July 15-20, 2018; Melbourne, Australia. [doi: 10.18653/v1/p18-1208]
24. Ong DC, Wu Z, Zhi-Xuan T, Reddan M, Kahhale I, Mattek A, et al. Modeling emotion in complex stories: the Stanford Emotional Narratives Dataset. IEEE Trans Affect Comput 2021 Jul 1;12(3):579-594 [FREE Full text] [doi: 10.1109/taffc.2019.2955949] [Medline: 34484569]
25. Ringeval F, Sonderegger A, Sauer J, Lalanne D. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. In: Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). 2013 Presented at: FG 2013; April 22-26, 2013; Shanghai, China. [doi: 10.1109/fg.2013.6553805]
26. Zadeh A, Cao YS, Hessner S, Liang PP, Poria S, Morency LP. CMU-MOSEAS: a multimodal language dataset for Spanish, Portuguese, German and French. Proc Conf Empir Methods Nat Lang Process 2020 Nov;2020:1801-1812 [FREE Full text] [doi: 10.18653/v1/2020.emnlp-main.141] [Medline: 33969362]
27. Gupta A, D'Cunha A, Awasthi K, Balasubramanian V. DAiSEE: towards user engagement recognition in the wild. arXiv Preprint posted online on September 7, 2016 [FREE Full text]
28. Kaur A, Mustafa A, Mehta L, Dhall A. Prediction and localization of student engagement in the wild. arXiv Preprint posted online on April 3, 2018 [FREE Full text] [doi: 10.1109/dicta.2018.8615851]
29. Sathayanarayana S, Satzoda RK, Carini A, Lee M, Salamanca L, Reilly J. Towards automated understanding of student-tutor interactions using visual deictic gestures. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2014 Presented at: CVPRW 2014; June 23-28, 2014; Columbus, OH. [doi: 10.1109/cvprw.2014.77]
30. Berger T, Boettcher J, Caspar F. Internet-based guided self-help for several anxiety disorders: a randomized controlled trial comparing a tailored with a standardized disorder-specific approach. Psychotherapy (Chic) 2014 Jun;51(2):207-219. [doi: 10.1037/a0032527] [Medline: 24041199]
31. Gómez Penedo JM, Berger T, Grosse Holtforth M, Krieger T, Schröder J, Hohagen F, et al. The Working Alliance Inventory for guided Internet interventions (WAI-I). J Clin Psychol 2020 Jun;76(6):973-986. [doi: 10.1002/jclp.22823] [Medline: 31240727]
32. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. Psychother Theory Res Pract 1979;16(3):252-260. [doi: 10.1037/h0085885]
33. Herrero R, Vara MD, Miragall M, Botella C, García-Palacios A, Riper H, et al. Working alliance inventory for online interventions-short form (WAI-TECH-SF): the role of the therapeutic alliance between patient and online program in

XSL•FO
RenderX

therapeutic outcomes. Int J Environ Res Public Health 2020 Aug 25;17(17):6169 [FREE Full text] [doi: 10.3390/ijerph17176169] [Medline: 32854381]
34. Shinno S, Matsumoto K, Hamatani S, Inaba Y, Ozawa Y, Kawasaki Y, et al. Feasibility of guided internet-based cognitive behavioral therapy for panic disorder and social anxiety disorder in Japan: pilot single-arm trial. JMIR Form Res 2024 Feb 29;8:e53659 [FREE Full text] [doi: 10.2196/53659] [Medline: 38421717]
35. McFee B, Raffel C, Liang D, Ellis DP, McVicar M, Battenberg E, et al. librosa: audio and music signal analysis in python. In: Proceedings of the 14th Python in Science Conference. 2015 Presented at: SciPy 2015; July 6-12, 2015; Austin, TX. [doi: 10.25080/majora-7b98e3ed-003]
36. Magdin M, Sulka T, Tomanová J, Vozár M. Voice analysis using PRAAT software and classification of user emotional state. Int J Interact Multimed Artif Intell 2019;5(6):33-42. [doi: 10.9781/ijimai.2019.03.004]
37. Balomenos T, Raouzaiou A, Ioannou S, Drosopoulos A, Karpouzis K, Kollias S. Emotion analysis in man-machine interaction systems. In: Proceedings of the First International Workshop on Machine Learning for Multimodal Interaction. 2004 Presented at: MLMI 2004; June 21-23, 2004; Martigny, Switzerland. [doi: 10.1007/978-3-540-30568-2_27]
38. Porter S, ten Brinke L. Reading between the lies: identifying concealed and falsified emotions in universal facial expressions. Psychol Sci 2008 May;19(5):508-514. [doi: 10.1111/j.1467-9280.2008.02116.x] [Medline: 18466413]
39. Arriaga O, Valdenegro-Toro M, Plöger P. Real-time convolutional neural networks for emotion and gender classification. arXiv Preprint posted online on October 20, 2017.
40. Demszky D, Movshovitz-Attias D, Ko J, Cowen A, Nemade G, Ravi S. GoEmotions: a dataset of fine-grained emotions. arXiv Preprint posted online on May 1, 2020 [FREE Full text] [doi: 10.18653/v1/2020.acl-main.372]
41. Lu K, Zhao J, Xia M, Zeng J. Semi-supervised learning for image retrieval using support vector machines. In: Proceedings of the Second International Symposium on Neural Networks. 2005 Presented at: ISNN 2005; May 30-June 1, 2005; Chongqing, China. [doi: 10.1007/11427391_108]
42. Yu D, Varadarajan B, Deng L, Acero A. Active learning and semi-supervised learning for speech recognition: a unified framework using the global entropy reduction maximization criterion. Comput Speech Lang 2010 Jul;24(3):433-444. [doi: 10.1016/j.csl.2009.03.004]
43. Liu Y, Kirchhoff K. Graph-based semi-supervised learning for phone and segment classification. In: Proceedings of the INTERSPEECH 2013. 2013 Presented at: INTERSPEECH 2013; August 25-29, 2013; Lyon, France URL: https://www.isca-archive.org/interspeech_2013/liu13e_interspeech.html# [doi: 10.21437/interspeech.2013-453]
44. Alyuz N, Okur E, Oktay E, Genc U, Aslan S, Mete SE, et al. Semi-supervised model personalization for improved detection of learner's emotional engagement. In: Proceedings of the 18th ACM International Conference on Multimodal Interaction. 2016 Presented at: ICMI '16; November 12-16, 2016; Tokyo, Japan. [doi: 10.1145/2993148.2993166]
45. Nezami OM, Richards D, Hamey L. Semi-supervised detection of student engagement. In: Proceedings of the PACIS 2017. 2017 Presented at: PACIS 2017; July 16-20, 2017; Langkawi, Malaysia.
46. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2. 2014 Presented at: NIPS'14; December 8-13, 2014; Montreal, QC.
47. Saeed AQ, Sheikh Abdullah SN, Che-Hamzah J, Abdul Ghani AT. Accuracy of using generative adversarial networks for glaucoma detection: systematic review and bibliometric analysis. J Med Internet Res 2021 Sep 21;23(9):e27414 [FREE Full text] [doi: 10.2196/27414] [Medline: 34236992]
48. Olmschenk G, Zhu Z, Tang H. Generalizing semi-supervised generative adversarial networks to regression using feature contrasting. arXiv Preprint posted online on on November 27, 2018 [FREE Full text]
49. Sinatra GM, Heddy BC, Lombardi D. The challenges of defining and measuring student engagement in science. Educ Psychol 2015 Feb 27;50(1):1-13. [doi: 10.1080/00461520.2014.1002924]
50. Békés V, Aafjes-van Doorn K, Zilcha-Mano S, Prout T, Hoffman L. Psychotherapists' acceptance of telepsychotherapy during the COVID-19 pandemic: a machine learning approach. Clin Psychol Psychother 2021 Nov;28(6):1403-1415 [FREE Full text] [doi: 10.1002/cpp.2682] [Medline: 34723404]
51. Vail AK, Girard J, Bylsma L, Cohn J, Fournier J, Swartz H. Goals, tasks, and bonds: toward the computational assessment of therapist versus client perception of working alliance. In: Proceedings of the 16th IEEE International Conference on Automatic Face and Gesture Recognition. 2021 Presented at: FG 2021; December 15-18, 2021; Jodhpur, India. [doi: 10.1109/fg52635.2021.9667021]
52. Tracey TJ, Kokotovic AM. Factor structure of the Working Alliance Inventory. Psychol Assess J Consult Clin Psychol 1989 Sep;1(3):207-210. [doi: 10.1037/1040-3590.1.3.207]
53. Ewalds Mulliez AP, Pomey MP, Bordeleau J, Desbiens F, Pelletier JF. A voice for the patients: evaluation of the implementation of a strategic organizational committee for patient engagement in mental health. PLoS One 2018 Oct 24;13(10):e0205173 [FREE Full text] [doi: 10.1371/journal.pone.0205173] [Medline: 30356239]
54. Sagen JS, Smedslund G, Simonsen AE, Habberstad A, Kjeken I, Dagfinrud H, et al. Patient engagement in the development and delivery of healthcare services: a systematic scoping review. BMJ Open Qual 2023 Jun 27;12(2):e002309 [FREE Full text] [doi: 10.1136/bmjoq-2023-002309] [Medline: 37369560]

XSL•FO
RenderX

55. Khosravi M, Azar G. Factors influencing patient engagement in mental health chatbots: a thematic analysis of findings from a systematic review of reviews. Digit Health 2024 Apr 22;10:20552076241247983 [FREE Full text] [doi: 10.1177/20552076241247983] [Medline: 38655378]
56. Remmert JE, Destefano L, Chinman M, Oslin DW, Mavandadi S. Development of a peer-delivered primary care intervention to improve veteran mental health treatment engagement. Psychol Serv 2024 Aug 01. [doi: 10.1037/ser0000883] [Medline: 39088007]
57. Jørgensen K, Lerbæk B, Frederiksen J, Watson E, Karlsson B. Patient participation in mental health care - perspectives of healthcare professionals and patients: a scoping review. Issues Ment Health Nurs 2024 Aug 20;45(8):794-810. [doi: 10.1080/01612840.2024.2358931] [Medline: 38900284]
58. Elyoseph Z, Levkovich I. Comparing the perspectives of generative AI, mental health experts, and the general public on schizophrenia recovery: case vignette study. JMIR Ment Health 2024 Mar 18;11:e53043 [FREE Full text] [doi: 10.2196/53043] [Medline: 38533615]
59. Elyoseph Z, Levkovich I, Shinan-Altman S. Assessing prognosis in depression: comparing perspectives of AI models, mental health professionals and the general public. Fam Med Community Health 2024 Jan 09;12(Suppl 1):e002583 [FREE Full text] [doi: 10.1136/fmch-2023-002583] [Medline: 38199604]
60. Thieme A, Hanratty M, Lyons M, Palacios J, Marques RF, Morrison C, et al. Designing human-centered AI for mental health: developing clinically relevant applications for online CBT treatment. ACM Trans Comput Hum Interact 2023 Mar 17;30(2):1-50. [doi: 10.1145/3564752]
61. Rollwage M, Habicht J, Juechems K, Carrington B, Viswanathan S, Stylianou M, et al. Using conversational AI to facilitate mental health assessments and improve clinical efficiency within psychotherapy services: real-world observational study. JMIR AI 2023 Dec 13;2:e44358 [FREE Full text] [doi: 10.2196/44358] [Medline: 38875569]
62. Roy K, Khandelwal V, Goswami R, Dolbir N, Malekar J, Sheth A. Demo alleviate: demonstrating artificial intelligence enabled virtual assistance for telehealth: the mental health case. Proc AAAI Conf Artif Intell 2023 Jun 26;37(13):16479-16481. [doi: 10.1609/aaai.v37i13.27085]
63. Siegel A, Zuo Y, Moghaddamcharkari N, McIntyre RS, Rosenblat JD. Barriers, benefits and interventions for improving the delivery of telemental health services during the coronavirus disease 2019 pandemic: a systematic review. Curr Opin Psychiatry 2021 Jul 01;34(4):434-443 [FREE Full text] [doi: 10.1097/YCO.0000000000000714] [Medline: 33928918]
64. Foley GN, Gentile JP. Nonverbal communication in psychotherapy. Psychiatry (Edgmont) 2010 Jun;7(6):38-44 [FREE Full text] [Medline: 20622944]
65. Mitzkovitz C, Dowd SM, Cothran T, Musil S. The eyes have it: psychotherapy in the era of masks. J Clin Psychol Med Settings 2022 Dec 03;29(4):886-897 [FREE Full text] [doi: 10.1007/s10880-022-09856-x] [Medline: 35118604]
66. Wang Y, DeVito Dabbs A, Thomas TH, Campbell G, Donovan H. Measuring engagement in provider-guided digital health interventions with a conceptual and analytical framework using nurse WRITE as an exemplar: exploratory study with an iterative approach. JMIR Form Res 2024 Jul 22;8:e57529 [FREE Full text] [doi: 10.2196/57529] [Medline: 39037757]
67. Fragueiro A, Debroize RP, Bannier E, Cury C. Eye-tracking and skin conductance to monitor task engagement during neurofeedback sessions. In: Proceedings of the 9th Graz Brain Computer Interface Conference. 2024 Presented at: BCI 2024; September 9-12, 2024; Graz, Austria.
68. Ventura RB, Porfiri M. Galvanic skin response as a measure of engagement during play in virtual reality. In: Proceedings of the ASME 2020 Dynamic Systems and Control Conference. 2020 Presented at: ASME 2020; October 5-7, 2020; Virtual Event. [doi: 10.1115/DSCC2020-3177]
69. Varma M, Washington P, Chrisman B, Kline A, Leblanc E, Paskov K, et al. Identification of social engagement indicators associated with autism spectrum disorder using a game-based mobile app: comparative study of gaze fixation and visual scanning methods. J Med Internet Res 2022 Feb 15;24(2):e31830 [FREE Full text] [doi: 10.2196/31830] [Medline: 35166683]
70. Goldberg P, Sümer Ö, Stürmer K, Wagner W, Göllner R, Gerjets P, et al. Attentive or not? Toward a machine learning approach to assessing students' visible engagement in classroom instruction. Educ Psychol Rev 2019 Dec 18;33:27-49. [doi: 10.1007/s10648-019-09514-z]
71. Booth BM, Bosch N, D'Mello SK. Engagement detection and its applications in learning: a tutorial and selective review. Proc IEEE 2023 Oct;111(10):1398-1422. [doi: 10.1109/jproc.2023.3309560]
72. Bernard A, de Ossorno Garcia S, Salhi L, John A, DelPozo-Banos M. Patterns of engagement in a digital mental health service during COVID-19: a cohort study for children and young people. Front Psychiatry 2023 Jul 27;14:1143272 [FREE Full text] [doi: 10.3389/fpsyt.2023.1143272] [Medline: 37575580]

Abbreviations

AI: artificial intelligence
CMU-MOSEI: Carnegie Mellon University Multimodal Opinion Sentiment and Emotion Intensity
CMU-MOSI: Carnegie Mellon University Multimodal Corpus of Sentimental Intensity
GAN: generative adversarial net
HIPAA: Health Insurance Portability and Accountability Act

XSL•FO
RenderX

MEDICA: Multimodal Engagement Detection in Clinical Analysis
MFCC: mel-frequency cepstral coefficients
RMSE: root mean square error
WAI: Working Alliance Inventory
CREMA-D: Crowd-source Emotional Multimodal Actors Dataset
VGG-B: Visual Geometry Group-B
BERT: Bidirectional Encoder Representations from Transformers
LBP-TOP: local binary pattern on three orthogonal planes

Edited by A Mavragani; submitted 09.02.23; peer-reviewed by R Marshall, T Huang; comments to author 13.04.23; revised version received 30.06.23; accepted 03.09.24; published 20.01.25.

Please cite as:
Guhan P, Awasthi N, McDonald K, Bussell K, Reeves G, Manocha D, Bera A
Developing a Machine Learning–Based Automated Patient Engagement Estimator for Telehealth: Algorithm Development and Validation Study
JMIR Form Res 2025;9:e46390
URL: https://formative.jmir.org/2025/1/e46390
doi:10.2196/46390
PMID:

©Pooja Guhan, Naman Awasthi, Kathryn McDonald, Kristin Bussell, Gloria Reeves, Dinesh Manocha, Aniket Bera. Originally published in JMIR Formative Research (https://formative.jmir.org), 20.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Psychological and Behavioral Insights From Social Media Users: Natural Language Processing–Based Quantitative Study on Mental Well-Being

Xingwei Yang[1], PhD; Guang Li[2], PhD

[1]Information Technology Management, Ted Rogers School of Management, Toronto Metropolitan University, Toronto, ON, Canada

[2]Smith School of Business, Queen's University, Kingston, ON, Canada

Corresponding Author:
Xingwei Yang, PhD
Information Technology Management
Ted Rogers School of Management
Toronto Metropolitan University
55 Dundas St W
Toronto, ON, M5G 2C3
Canada
Phone: 1 416 979 5044
Email: nancy.yang@torontomu.ca

Abstract

Background: Depression significantly impacts an individual's thoughts, emotions, behaviors, and moods; this prevalent mental health condition affects millions globally. Traditional approaches to detecting and treating depression rely on questionnaires and personal interviews, which can be time consuming and potentially inefficient. As social media has permanently shifted the pattern of our daily communications, social media postings can offer new perspectives in understanding mental illness in individuals because they provide an unbiased exploration of their language use and behavioral patterns.

Objective: This study aimed to develop and evaluate a methodological language framework that integrates psychological patterns, contextual information, and social interactions using natural language processing and machine learning techniques. The goal was to enhance intelligent decision-making for detecting depression at the user level.

Methods: We extracted language patterns via natural language processing approaches that facilitate understanding contextual and psychological factors, such as affective patterns and personality traits linked with depression. Then, we extracted social interaction influence features. The resultant social interaction influence that users have within their online social group is derived based on users' emotions, psychological states, and context of communication extracted from status updates and the social network structure. We empirically evaluated the effectiveness of our framework by applying machine learning models to detect depression, reporting accuracy, recall, precision, and F_1-score using social media status updates from 1047 users along with their associated depression diagnosis questionnaire scores. These datasets also include user postings, network connections, and personality responses.

Results: The proposed framework demonstrates accurate and effective detection of depression, improving performance compared to traditional baselines with an average improvement of 6% in accuracy and 10% in F_1-score. It also shows competitive performance relative to state-of-the-art models. The inclusion of social interaction features demonstrates strong performance. By using all influence features (affective influence features, contextual influence features, and personality influence features), the model achieved an accuracy of 77% and a precision of 80%. Using affective features and affective influence features also showed strong performance, achieving 81% precision and an F_1-score of 79%.

Conclusions: The developed framework offers practical applications, such as accelerating hospital diagnoses, improving prediction accuracy, facilitating timely referrals, and providing actionable insights for early interventions in mental health treatment plans.

(JMIR Form Res 2025;9:e60286) doi:10.2196/60286

XSL•FO
RenderX

KEYWORDS

social media; natural language processing; social interaction; decision support system; depression detection

Introduction

Depression is characterized by a constant depressed mood and the inability to anticipate happiness or pleasure; [1] it has emerged as one of the major challenges globally [2]. Approximately 350 million people experience depression worldwide [3]. Without timely psychiatric interventions, a patient may have worsening symptoms, leading to more severe consequences such as self-harm, physical disability, or even suicide [4]. This problem has been exacerbated by the COVID-19 pandemic. Fear of COVID-19 infection, lack of social support, isolation, and limited access to medical resources during the COVID-19 pandemic have increased psychological distress [5,6]. Compared to the prepandemic period, depression rates almost doubled in 2020 to 2021 [7]. Despite its severity and seriousness, depression remains significantly underdiagnosed and undertreated [8]. Thus, there is an increasing need for intelligent clinical decision support tools to offer timely depression diagnoses and referral services so that people can receive early intervention or treatment before developing more serious mental health conditions.

Traditional methods of depression detection and treatment are based on questionnaires and personal interviews, which usually require long one-on-one physical engagements where psychiatrists gather information from patients and then recommend medications or treatments. This process can be both time consuming and costly, not only because of its complexity but also because of the expenses associated with the required trained personnel [9]. Thus, it is challenging for mental health professionals to offer immediate detection, referral services, proactive interventions, or treatments for those at risk and prepare for the next public health emergency, such as COVID-19, particularly when in-person assessments are difficult or costly. This challenge has motivated the exploration of more cost-effective and efficient methods, such as leveraging data from social media [10]. Platforms such as Meta are commonly used to share feelings and emotions. This makes social media a new venue for depression investigations [11-13] because they provide an unbiased exploration of language use and behavioral patterns of individuals [14]. Moreover, research has shown that depression can affect how individuals express themselves in their social media posts [15]. Therefore, there is an increasing trend toward using abundant text data and information from social media to examine individuals' sentiments and behaviors in mental health studies [16,17]. However, manually analyzing texts is not practical for timely psychiatric support due to the exponential growth of text data on social media and the increasing number of cases of mental illnesses. Thus, there is a rising need to use machine learning (ML) and natural language processing (NLP) techniques to deal with the large volume of data. Existing research in this domain primarily focuses on finding the key indicators, such as linguistic and sentiment features, to enhance the accuracy of depression detection [18,19]. For example, many studies focus on extracting textual content from social media data using linguistic features such as semantics and syntactic features [20], a bag of words [21], and n-grams [22]. Although these features can help yield reasonable results in depression detection, they overlook several critical aspects.

First, these linguistic features do not thoroughly reflect an individual's psychological patterns and states. Affective patterns (affect refers to emotions that change or influence the behaviors or thoughts of an individual) in language expressions are important dimensions of psychological patterns. They can reveal an individual's moods, sentiments, and emotional factors beyond simple linguistic features for analyzing mental disorders on social media. Studies show that frequent negative affect expression has higher connections with depressive states, while positive affect tends to have a lower connection to depression [23-25]. Thus, emotions play an essential role in diagnosing mental illnesses such as depression [11]. In addition, an individual's psychological states are directly related to depression. Personality has been used to distinguish social media users' activities, status updates content, and language use habits [26,27]. Moreover, Klein et al [28] and Hakulinen et al [29] identified the links between depression and personality types and found that certain personality traits may serve as predictive factors and influence the treatment of depression. Second, contextual information about the words is another critical dimension. Because pretrained language models have shown strong contextual modeling ability, they are used in understanding texts in various applications [30,31]. By considering the sequence of all words in the documents, sequence-level semantics can be learned to reflect the diverse contexts and meanings of words in sentences. For example, in the sentence "My car was left on the left-hand side," the word "left" is used 2 times, but it carries 2 different meanings. The first "left" is a verb suggesting the action performed on the car, while the second "left" is a noun specifying the car's location. The contextual meaning (eg, the meaning derived from the words' surroundings) can vary for the same word. Thus, contextual information is also a crucial aspect of understanding individuals' mental states.

Third, existing research often overlooks the impact of social interactions when analyzing features for mental disorders on social media. Social interactions reflect how users influence each other within an online social group. According to social influence theory, individuals in social environments significantly shape each other's attitudes, beliefs, and actions [32]. Recent psychological studies show that family, friends, or colleagues can directly or indirectly affect mental well-being [33]. For instance, in internet-based communities, recreational discussions, or chat groups, a negative mood can spread among different individuals [34,35]. Mental illnesses are influenced by an individual's daily social interactions [36]. This impact can be even more substantial on social media, where individuals are influenced by their friends' shared conventions and norms. Therefore, investigating the influence of social interaction is crucial for monitoring psychological patterns and states on social media. Studies by Murarka et al [37], Cui et al [38], and Ghosal

XSL•FO
RenderX

and Jain [39] analyzed emotions, personality, and pretrained models for mental health problems, but they did not account for the impact of social interactions. To study the mental well-being of social media users, an individual's mental health is reflected by more than just the content of their status updates. Thus, in our study, we not only account for psychological patterns and states and contextual information but also incorporate the influence of social interactions by considering the social network structure within these feature patterns.

This study aimed to offer a novel methodological language framework to enhance intelligent decision-making in the mental health domain. Our contributions are outlined as follows:

- We developed a unified language framework using NLP and ML techniques to enhance the detection of depression at the user level, making it more effective and accurate.
- Our framework advances previous studies by extracting and analyzing language patterns to comprehensively understand the contextual and psychological features (eg, affective patterns and personality) associated with depressive symptoms.
- We extracted features of social interaction influence, capturing how users influence their online social group. These features consider users' emotions, psychological states, and communication context from status updates and social network structure, differentiating individuals with depression from those without.

Our study aims for early-stage detection and initial screening to identify individuals at risk of getting depression and narrow down those needing closer observation or follow-up. It is designed to complement, not replace, clinical diagnoses, offering insights that support timely interventions and enhance patient care. Our framework can enrich the reliability and knowledge base of decision support systems for health care providers, enabling more precise diagnoses and timely referrals, interventions, and treatment plans. The framework is designed to be adaptable, enabling researchers to apply similar methods to their own datasets and explore psychological, contextual, and social influence factors. While it provides valuable insights, the framework is not intended to draw definitive conclusions but rather to offer a structured approach to understanding social interactions and their impact. Beyond its focus on mental illness, it can be extended to other domains, generating interpretable findings tailored to specific contexts.

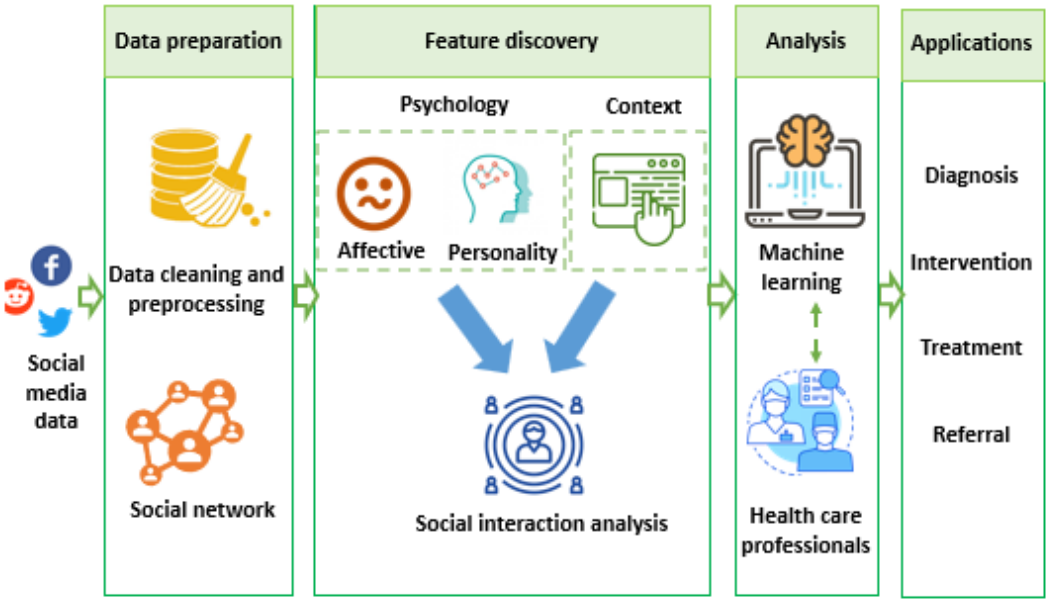
The paper is organized as follows. In the Methods section, we introduce our proposed framework and present the language feature discovery processes and the influence analysis. In the Results section, we empirically evaluate the effectiveness of our framework. We discuss the implications of our results, suggest directions for future research, and conclude the paper in the Discussion section.

Methods

Psychological and Contextual Analysis NLP Framework

This section presents our methods for inferring language patterns and conducting an analysis of social interactions from social media content. Figure 1 shows a conceptual representation of the proposed psychological and contextual analysis NLP framework to extract affective, personality, contextual, and social interaction influence features. Given the data collected from social media platforms, our framework starts with extensive data preparation and cleaning procedures. In addition, we establish a user-friendship network based on users' social connections on social media. This network shows if one user is connected to others (ie, if this user is a friend of other users).

Figure 1. The proposed psychological and contextual analysis natural language processing framework. It starts with extensive data preparation and cleaning procedures and establishes a user-friendship network based on users' social connections on social media. We then apply machine learning models to effectively manage high-volume and high-speed user-generated data to analyze the depression states of users within social media environments. This framework can be used in various applications.



XSL•FO
RenderX

In the feature discovery phase, our framework deduces depressive behaviors from social media users, considering both contextual and psychological perspectives. We first focus on extracting features, including (1) affective features—these suggest users' psychological patterns such as mood, sentiment, and emotional factors as expressed in their social media postings, (2) personality features—these reveal how users' psychological states (ie, personalities) affect depression states, and (3) contextual features—these show sequence-level semantics learned from users' social media postings to reflect different contexts and meanings of each word within sentences by considering the sequence of all words. On the basis of these features, we proceeded to perform a social interaction analysis to understand how the interactions between individuals with and without depression influence each other within their user-friendship network. We considered both the social network structure and users' psychological and contextual features for extracting social interaction influence features, where the structure of the user-friendship network shows the strength of connections among users in the social media environment. Then, we applied ML models to effectively manage high-volume and high-speed user-generated data to analyze the depression states of users within social media environments.

Data Cleaning and Preparation

We illustrated and evaluated our proposed framework using datasets from the study by Kosinski et al [40], which contain 197,230 Facebook (Meta Platforms, Inc) status updates from 1047 users, along with their associated Center for Epidemiologic Studies Depression (CES-D) Scale questionnaire scores. These datasets also include user postings, network connections, and personality responses. Note that datasets from other general social networks or online discussion forums can also be used. To minimize the privacy impact regarding using text data, we followed privacy-preserving procedures [41] to ensure anonymity throughout our analysis.

We obtained status updates from all users, regardless of their demographics, as our focus was on a general observation of social network users. Each user's status updates were then aggregated based on their anonymous IDs. We cleaned and preprocessed the data to ensure consistent content. Text preprocessing was conducted in Python (Python Software Foundation), involving correcting spelling errors with SpellChecker (package in Python), converting text to lowercase, removing special characters and numbers, and expanding contractions. We used the NLTK package (Natural Language Toolkit, a Python library for NLP) to remove stop words such as "the," "is," and "our," which are the most common words that appear in a sentence and do not contribute to its meaning. Thus, we removed stop words without sacrificing the meaning of the sentence. Our analysis assesses users' personality features based on the Big Five Factor Model using their responses to the International Personality Item Pool (IPIP) [42]. We determined users' depression levels by obtaining their responses to the CES-D Scale questionnaire, which is a commonly used approach to diagnose depression [43]. It is a 20-item screening questionnaire that captures users' responses to questions related to depressive states. These questions assess the frequency of feelings and behaviors over the past week, including loneliness, sadness, and restless sleep. The higher the score, the more likely an individual is to experience depression. We adopted a similar threshold of 25, as suggested by Lewinsohn et al [44] and Julian et al [45], to label users with depression (total CES-D score >25) and without depression.

We begin by extracting affective, personality, and contextual features from status updates, followed by deriving social interaction influence features based on social network connections. The subsequent sections detail the procedure to identify psychological (eg, affective and personality) and contextual features. These features provide insights into how emotional states, personality traits, and contextual factors interact with social influences. By analyzing these features, we can better understand the dynamics of social interactions and their broader implications for applications such as behavior modeling and targeted interventions.

Discover Affective Features

This section outlines how we extracted affective features. Emotions are essential in diagnosing mental illnesses [46]. Several studies show that depressive symptoms are more related to negative affect, while individuals without depression are more related to positive affect [23,24]. In our study, we extracted affective features to reveal individuals' emotional states and understand users' moods, sentiments, and emotional factors beyond simple linguistic features. Multimedia Appendix 1 shows the process of extracting emotion intensity features and emotion word features to represent the affective features. Because different words can convey affect to various degrees (eg, intensities), we used the NRC Emotion Intensity Lexicon [47] to extract the emotion words from users' social media postings. The lexicon includes >10,000 words related to emotions, including surprise, anger, joy, fear, trust, sadness, anticipation, and disgust, with different intensities. Both common English words and prominent words in social media are covered. For each word t, the intensity of a particular emotion e falls within the range of 0 to 1. A value of "1" indicates the highest intensity of a particular emotion e conveyed by the word t, while "0" indicates the lowest intensity of emotion e for the word t, where $e \in$ {fear, sadness, anger, joy, anticipation, trust, disgust, and surprise}. Thus, each emotion word has an emotion intensity vector $v(t)$ showing the intensities of 8 emotion dimensions. We first calculated how often each emotion word appears within individuals' social media postings to obtain an affective feature vector from a user's posting. Then, for each emotion word t, we multiplied the emotion intensity vector $v(t)$ with the frequency. After that, we computed the average intensity for all emotion words to obtain the affective feature vector.

For example, a user uses 3 emotion words "threat," "illness," and "sunny" in social media postings with a frequency of $freq(threat)=2$, and $freq(illness)=freq(sunny)=1$, respectively. The emotion intensity vector of each emotion word is $v(threat)=$ <0.742, 0.604, 0, 0, 0, 0, 0, 0 >, $v(illness)=$<0, 0.609, 0.688, 0, 0, 0, 0, 0 >, and $v(sunny)=$<0, 0, 0, 0, 0.562, 0.5, 0 >. For each emotion word, we multiplied the emotion intensity vector with the frequency of the word and then computed the average emotion intensity for all emotion words. The average of all emotion word intensity for this user is <0.495, 0.605, 0.229, 0,

XSL•FO
RenderX

0, 0.187, 0.167, 0 >. We repeated the process for each user to obtain the user-level average emotion intensity to obtain affective features for all users.

Research shows that the occurrence and distribution of particular words within a document can represent its main content [48]. This suggests that the occurrence of a particular emotion word appearing in social media postings can indicate individuals' emotional state, sentiment, and mood. Thus, we extracted emotion words from individual with and without depression to illustrate their respective emotional states and make a comparison between them. We divided the status updates into 2 datasets: one containing only the status updates of the individuals with depression and the other containing only the status updates of individuals without depression. We then computed the term frequencies of each emotion word based on these datasets from individuals with and without depression:

where $c_g(t)$ indicates the frequency of emotion words t in group g ($g=d$ for the depressed group and $g=nd$ for the nondepressed group).

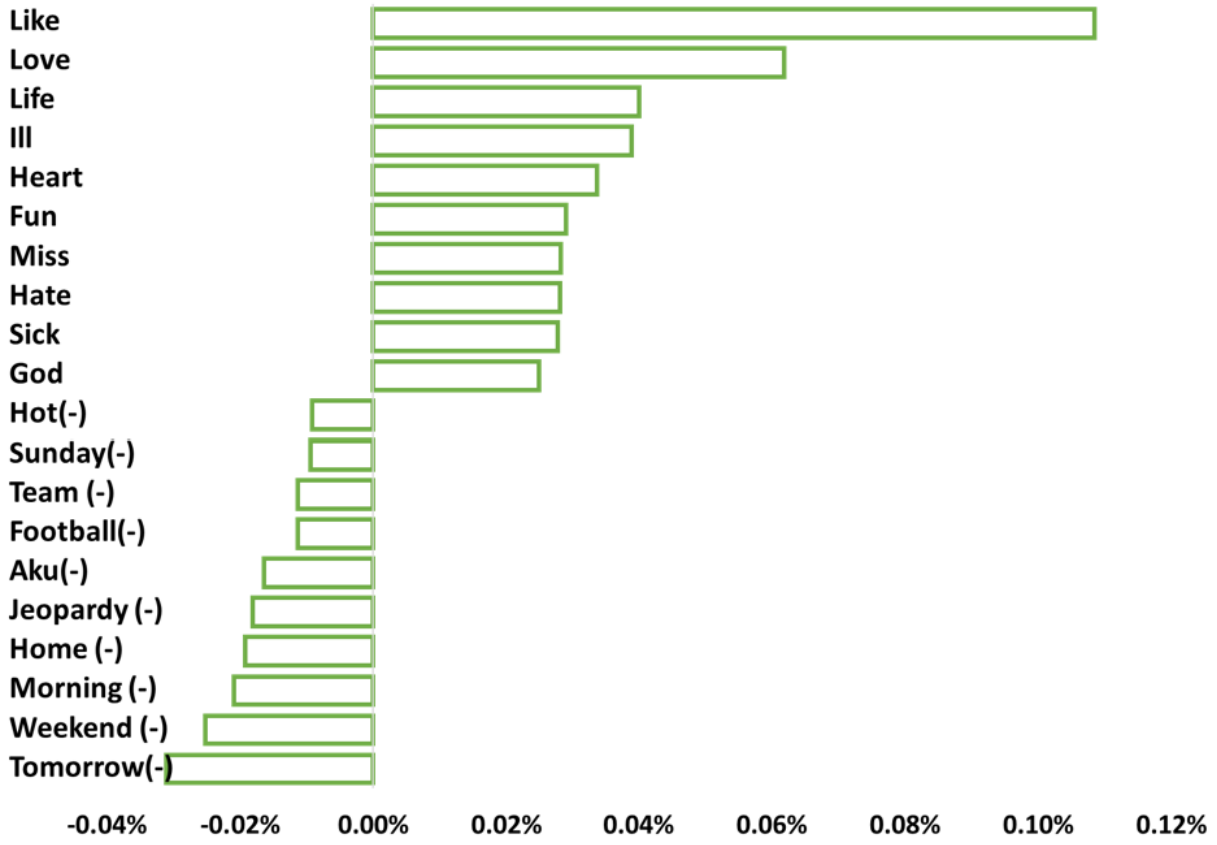
Then, we find df (t), the ratio differences in relative term frequency for each emotion word t that appears in the datasets from individuals with and without depression as follows:

$$df(t) = tf_d(t) - tf_{nd}(t) \quad \textbf{(2)}$$

We then sorted the ratio differences of all emotion words to find the most distinct emotion words across the depressed and nondepressed groups. We identify the top-k emotion words as distinct features of the depressed group and the bottom-k emotion words as distinct features of the nondepressed group, which help differentiate between the 2 groups.

Figure 2 illustrates an example of the top 10 and bottom 10 emotions words. The emotion words such as "like," "love," "life," "ill," and "heart" have the most positive disparities in term frequencies between the depressed group and the nondepressed group. These words indicate the depressed group's distinct emotion features. In contrast, "tomorrow," "weekend," "morning," "home," and "jeopardy" have the most negative differences, indicating the distinct emotion features of the nondepressed group. The distinct emotion features discovered through term frequency computation highlighted the contrasting emotional states between the depressed and nondepressed groups. These insights serve as invaluable resources for clinically intelligent decision-making, enhancing diagnostic accuracy and offering support tailored to the unique emotional needs of individuals, which, in turn, promotes more effective therapeutic outcomes and enhances overall well-being.

Figure 2. The top 10 distinct emotion words (positive) were associated with the group with depression, and the bottom 10 distinct emotion words (negative) were associated with the group without depression.



XSL•FO
RenderX

Once we identify the most distinct emotion features, we use term frequency and inverse document frequency (TFIDF) to indicate the importance of a word *t* to a document D (eg, users' aggregated status updates) in a corpus [49]. Each user will have 1 feature vector with all emotion words based on the TFIDF (t, D) value:

$$TFIDF\ (t, D) = TF\ (t, D) \times IDF\ (t) \quad \textbf{(3)}$$

where TF (t, D) stands for the term frequency of word *t* in a specific document *D*. The higher TF *(t, D)*, the more important the word *t* is for that document D. IDF (t) stands for inverse document frequency and shows how significant term *t* is in the entire collection of documents, with rarer words having higher IDF values.

Discover Personality Features

This section outlines how we extracted personality features. Personality shows the patterns of individuals' thoughts, feelings, and actions as well as their attitudes and reactions to external social environments. Numerous studies have shown a link between personality and depressive symptoms [28,29,50]. Therefore, many studies adopt personality for diagnosing depression [32,51].

Our study uses the IPIP questionnaire [42] to measure individuals' personalities based on 5 dimensions outlined in the Big Five Factor Model [48]. This model is the most commonly accepted personality theory in various research studies and clinical practice, offering a comprehensive approach to characterizing an individual's unique characteristics and patterns of behaviors. The model shows how people think, feel, and behave with 5 dimensions: extraversion, agreeableness, conscientiousness, neuroticism, and openness [52,53]. Preoţiuc-Pietro et al [54] used language-derived and user-level "big five" personality factor dimensions to predict the depression state of a user based on social media postings. The personality features show strong performance in differentiating between users with depression and without depression. Yang et al [55] also used personality indicators for depression detection, affirming their strong correlation with traits such as neuroticism, extraversion, and conscientiousness. The IPIP questionnaire includes 50 assessment questions designed to assess diverse individual differences. Participants indicate the extent of truthfulness regarding themselves on a scale from 1=disagree to 5=agree. Higher scores on a personality dimension indicate greater alignment with the respective trait. Multimedia Appendix 2 shows a detailed breakdown of each trait. Five numerical features are derived through responses to the IPIP questionnaire, each corresponding to a participant's personality trait.

Discover Contextual Features

This section shows how we extracted contextual features (contextual meaning of words derived from the words' surroundings in a sentence), which is another critical component of user-generated content. To extract contextual-level language features to reflect different contexts and meanings of words in sentences, we considered the sequence of words in users' social media postings to learn sequence-level semantics. Given that social media posts are often written in a general format, characterized by informal or everyday language, pretrained models—typically trained on general datasets—are particularly effective in capturing these patterns. Their strong contextual modeling capabilities have made pretrained language models well suited for analyzing such data, including the detection of different mental health conditions [51,56,57]. Thus, we take advantage of them to extract contextual-level language features from user generated content. Moreover, considering contextual information, we conducted cluster analysis to visualize specific words used by users with depression and without depression for further analysis by clinicians.

A state-of-the-art model such as Bidirectional Encoder Representations from Transformers (BERT) [58] is a pretrained language model designed to learn deep bidirectional representations of natural language from extensive unsupervised text corpora. BERT has the ability to capture information from both the preceding and following words, enabling it to grasp a more comprehensive understanding of the language context. BERT generates a representation for each word, called word embedding, which can be used for further ML tasks such as classification or clustering. However, to perform ML tasks such as classifying sentences, BERT cannot directly generate independent sentence embeddings (eg, representation for sentences). Thus, we adopted Sentence BERT (SBERT) [59], which effectively represents texts with >1 term. The SBERT model was fine-tuned and pretrained on large datasets [60,61]. Input sentences can be transformed into fixed-sized vectors (768 dimensions), serving as sentence representations. Moreover, sentences that are semantically similar will be close to each other; that is, if we use similarity measures such as Euclidean distance measure to compute the differences between 2 sentences, the more similar the 2 sentences are, the shorter the distance between them.

Multimedia Appendix 3 outlines the proposed procedure to discover contextual features. We first aggregated each user's social media postings. Then, we applied SBERT for each user's aggregated posting [59] to learn sentence embeddings based on each word's contextual information in a sentence to obtain user-level contextual features.

User-level contextual features alone may not offer interpretable information for aiding clinicians in making informed decisions. Hence, based on the sentence embedding, we conducted a cluster analysis to examine the content discussed by different user groups. First, we split the sentence embeddings (eg, user-level contextual features) into 2 datasets (depressed and nondepressed). Next, we applied K-means clustering separately to the datasets of user groups with depression and without depression to cluster sentence embeddings from each group. This helps differentiate the contents or topics discussed by different user groups.

To accomplish this, K-means clustering is conducted across a range of cluster sizes, computing the corresponding silhouette scores—a well-known measure indicating how effectively the resulting clusters separate from each other. The best model is chosen based on the highest silhouette score, with the associated number of clusters representing the optimal cluster count for K-means clustering on the dataset.

XSL•FO
RenderX

For each user group, users belonging to the same cluster are grouped together to obtain the aggregated social media postings. Next, we identified the main topics that user groups with depression and without depression talk about within each cluster. This is achieved by extracting the top-*k* words from each cluster based on term frequency, indicating how often a term appears in a specific cluster and group. Subsequently, we analyzed the words and main topics frequently discussed by different user groups.

Discover Social Interaction Influence Features

This section shows how we extracted social interaction influence features. Studies show that social interactions within social groups can directly or indirectly impact an individual's mental state [33]. Online social interactions are considered as intentional social actions that individuals engage in within the context of participation [62]. These interactions include users expressing their feelings and emotions by posting status updates and comments to share conventions, norms, or languages with positive or negative moods in online social groups. Research indicates that social interaction can facilitate the transfer of negative moods between individuals [34]. Thus, individuals' psychological patterns and the specific words and sentences they use in their online postings in different contexts may affect how friends or social connections in their online social groups react or express themselves on the internet. We conducted social interaction analysis to extract the social interaction influence features for each user in the social network and examine the impact from the psychological and contextual perspective of other users in social groups through social connections. Social interaction influences features that consider users' psychological and contextual features and network structure from social connections.

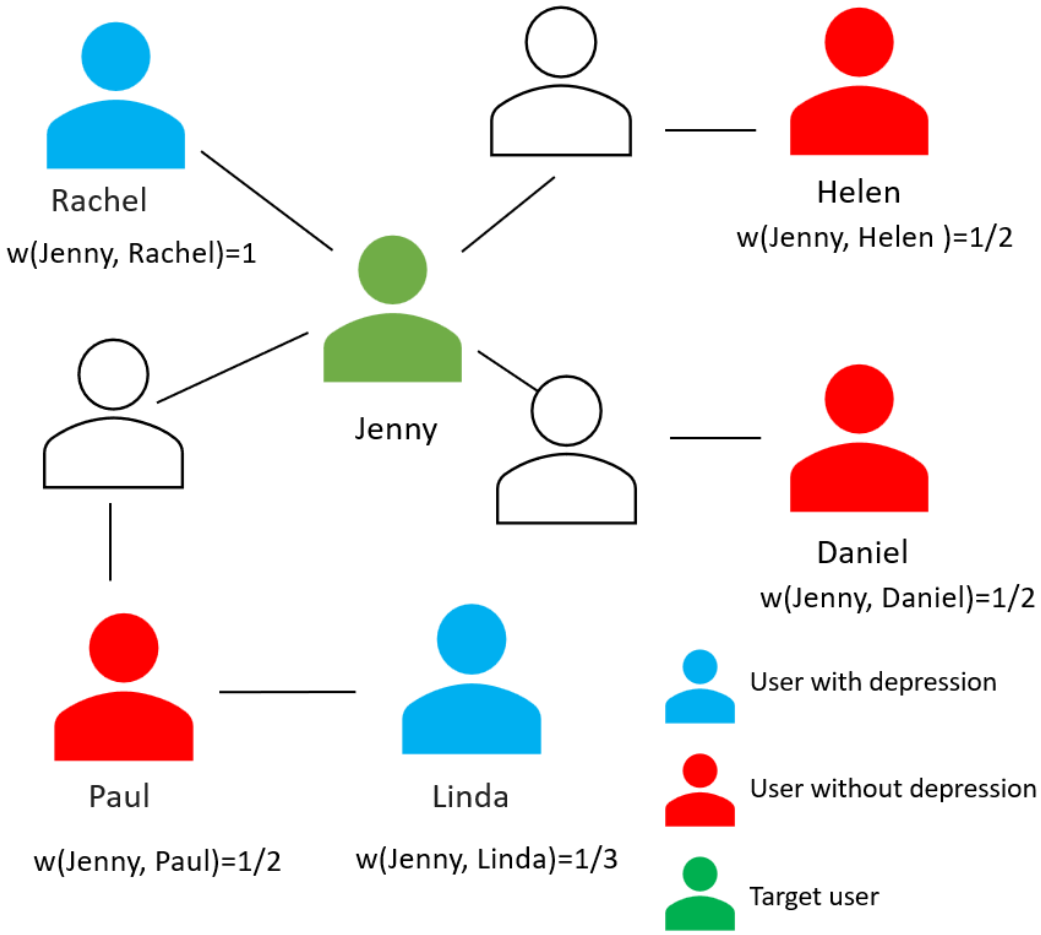
Multimedia Appendix 4 shows procedures for obtaining social interaction influence features. For each user, we obtained feature vectors from psychological and contextual features described in the Discover Affective Features, Discover Personality Features, and Discover Contextual Features sections. For each type of feature *f* (eg, emotion, personality, or context features), we have 2 feature vectors, $FV^{f}_{U,d}$, and $FV^{f}_{U,nd}$ for each user *U* where $g \in \{d, nd\}$. We then found the social connections of each pair of users via the user-friendship network that shows if one user is connected with the other. If user *U* is a friend of another user, *Y*, they are socially connected. Next, we examined the influence of users with depression and without depression on the target user *U*, respectively, based on the different features to obtain emotion, personality, and context influence features. We argue that the emotions, personality, and context information of users with depression and without depression on social media affect how other users express themselves through immediate and nonimmediate social connections and interactions. Therefore, if one user is more influenced by another, their features will be more similar. Moreover, the target user is more likely to be influenced by their close neighbors in a social network regarding distance. Thus, a more immediate neighbor influences the target user more than a faraway neighbor in the social network. Hence, we extract distance-weighted social interaction influence features from users with depression and without depression by considering both network structure (eg, distances between users) and different features (eg, emotions, personality, and contextual features) simultaneously to reflect the neighbors' impact on the target user. We first run the shortest paths for the whole social network between each pair of users. For target user *U*, we obtained the distance measures of the target user and the neighbor user *Y* as *Distance* (*U, Y)*. Then, we assigned weight to each neighbor user *Y*.

Then, the influence of depressed and nondepressed neighbors on the target user was computed using the weight and feature vectors. Consequently, we derived 2 separate social interaction influence features for the target user to represent the degree of similarities between the target user and their depressed and nondepressed neighbors. The same procedure was applied to all social network users.

We illustrate how the analysis works with an example for a particular user, Jenny. Figure 3 shows a partial user-friendship network Jenny. She has a direct connection to Rachel (immediate neighbor) and indirect connections to Paul, Helen, Daniel (second-level neighbors), and Linda (third-level neighbor) in her user-friendship network. Because the shortest path between Jenny and Rachel (first-level neighbor) is 1, we consider their distance to be 1 (ie, *Distance* [*Jenny, Rachel*]=1). Because the shortest path between Jenny and her second-level neighbors (ie, Paul, Helen, and Daniel) is 2, we consider the distance as 2. Likewise, the distance between Jenny and her third-tier neighbor (ie, Linda) is defined as 3, and so forth. Because we argue that the target user is more likely to be influenced by their immediate neighbors in a social network, we can infer that the first-level neighbor, Rachel, has a greater potential to impact Jenny than the second-level neighbors, Paul, Helen, and Daniel. The second-level neighbors have a stronger influence over Jenny than the third-level neighbor, Linda. Consequently, we allocated a greater influence on the first-level neighbor than the second-level neighbor to reflect the higher impact first-level neighbors have on the target user. Similarly, we assigned a higher weight to second-level neighbors than third-level neighbors. For each neighbor *Y* of user *U*, we assigned scores based on the following equation:

XSL•FO
RenderX

Figure 3. Partial user-friendship network of Jenny showing a direct connection to Rachel (immediate neighbor) and indirect connections to Paul, Helen, Daniel (second-level neighbors), and Linda (third-level neighbor) in her user-friendship network.



Thus, we assigned a weight of 1 to Rachel; a weight of 1/2 to Paul, Helen, and Daniel; and a weight of 1/3 to Linda to capture the decreasing influence with increasing distance between users.

In our analysis, we considered that the level of influence one user has on another is correlated with the similarity of features between the 2 users. Thus, we used Euclidean distance to assess the similarities of user *U* with their depressed and nondepressed neighbors separately. These similarities are determined by evaluating the depressed and nondepressed social media content of user *U* and their neighbor *Y* for each type of feature *f* (eg, emotions, personality, and contextual features). The similarity is calculated as the absolute difference between the feature vectors:

Sim $(U, Y)^f$ is calculated for all users in a social network with confirmed depression states. Then, we combined similarity measures with weights to obtain the social interaction score:

where α_Y is an indicator and α_Y=1 when the neighbor belongs to user group *g*. The higher the values of $SIC_g(U, Y)$, the more impact the specific neighbor user group (depressed or nondepressed) has on the target user in the social network. Multimedia Appendix 5 shows the pseudocode of the proposed social interaction analysis.

Ethical Considerations

The data used in this study were originally collected by researchers at Johns Hopkins University and received approval from their institutional review board. Permission to use the data solely for research purposes was obtained under a confidentiality agreement. This paper details the privacy-preserving procedures used to maintain participant anonymity in the Data Cleaning and Preparation section. Because this study only reanalyzed secondary, anonymized data, additional ethics approval for this study was not required. The original data host has obtained informed consent from all participants who voluntarily donated their data for research purposes, permitting reuse for academic research. The dataset used for this secondary analysis consisted of fully anonymized data and was restricted to noncommercial academic research purposes. This secondary analysis did not involve any direct participant interaction or additional compensation. There is no identification of individual participants or users in any images or supplementary material included in the manuscript. The dataset was deidentified, ensuring no identifiable information was presented.

XSL•FO
RenderX

Results

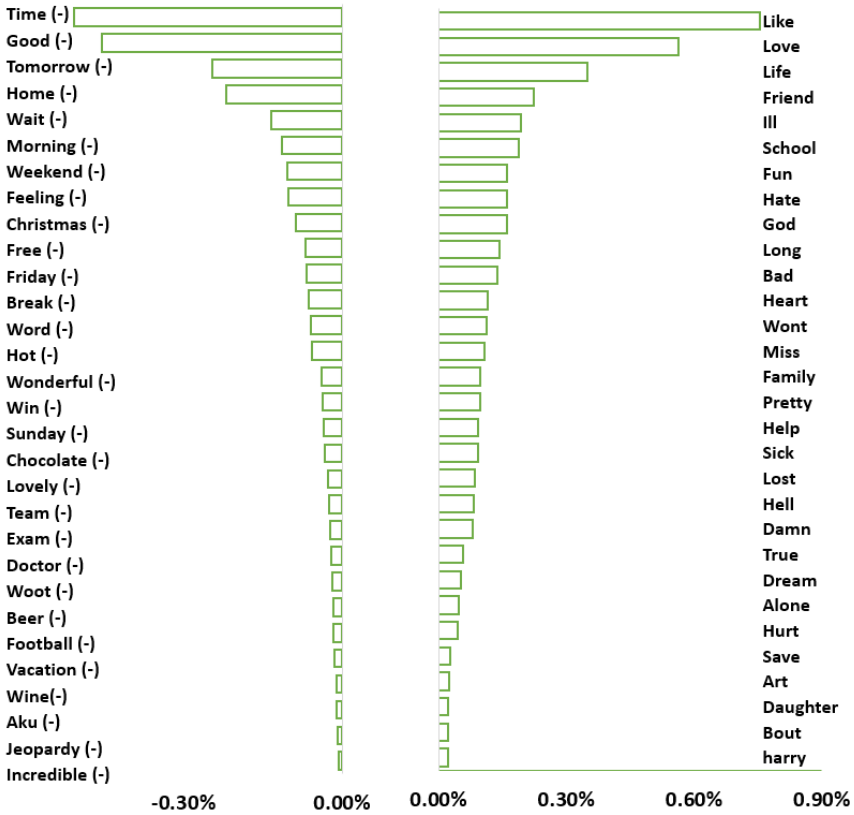
Overview

In this section, we follow the feature discovery procedure discussed in the Methods section and analyze the proposed features. We then conduct experiments to evaluate the proposed framework based on the datasets [40], which include Facebook users' status updates, network connections, users' associated depression scores measured by CES-D questionnaires, and personality measured by the IPIP personality questionnaire. We demonstrate how we extract psychological and contextual features as well as the resultant social interaction influence features. We explore the understandability of the developed features in explaining and understanding users' sentiments, moods, emotional behaviors, psychological states, and influence from social interactions in social networks. Our framework can be extended to include any social media and social platforms to enhance the capability of mental health professionals and clinicians to make explainable and timely diagnoses, referrals, interventions, and treatments for their potential patients.

Analysis of Affective Features

Given the procedure outlined in the Discover Affective Features section, we first obtained user-level emotion intensity by averaging the emotion intensities of the 8 emotion dimensions in users' social media postings. We further explored emotions, sentiments, and moods across different user groups by examining the top 30 and bottom 30 emotion words as distinct features, which were extracted from the depressed and nondepressed groups, respectively. Figure 4 illustrates the frequency distributions for each user group. Words with higher frequency values were more distinct emotion terms in each user group. The user group with depression frequently used emotion words with negative sentiments, such as "sick," "hurt," "bad," "alone," "ill," "lost," and "hate," and some swear words, which align with those reported in previous studies [63]. In contrast, users without depression expressed positive sentiments, including "good," "wonderful," "incredible," and "lovely." Words such as "family," "friends," and "daughter" were frequently mentioned in status updates by individuals in the user group with depression, indicating specific individuals or individuals that are related to them. In contrast, events such as "exams" and "vacation"; specific food, drinks, or items related to events including "beer," "wine," "football," "chocolate," and "hot"; and outcomes of events such as "win," "break," and "jeopardy" were discussed by the user group without depression.

Figure 4. Emotion frequency distributions of the user groups with depression and without depression. The top 30 distinct emotion words (positive) are associated with the depressed group, and the top 30 distinct emotion words (negative) are associated with the group without depression.



From Figure 4, we observe that the user group without depression is more likely to use emotion words associated with time, such as "time," "tomorrow," "morning," "weekend," "friday," "christmas," and "sunday." In contrast, the user group with depression expresses feelings by using words such as "like," "love," and "fun." They also care about more internal things such as "god," "life," "hear," and "dreams."

We calculated TFIDF (t, D) for each emotion word *t* in each user's social media postings *D* to generate the feature vectors

XSL•FO
RenderX

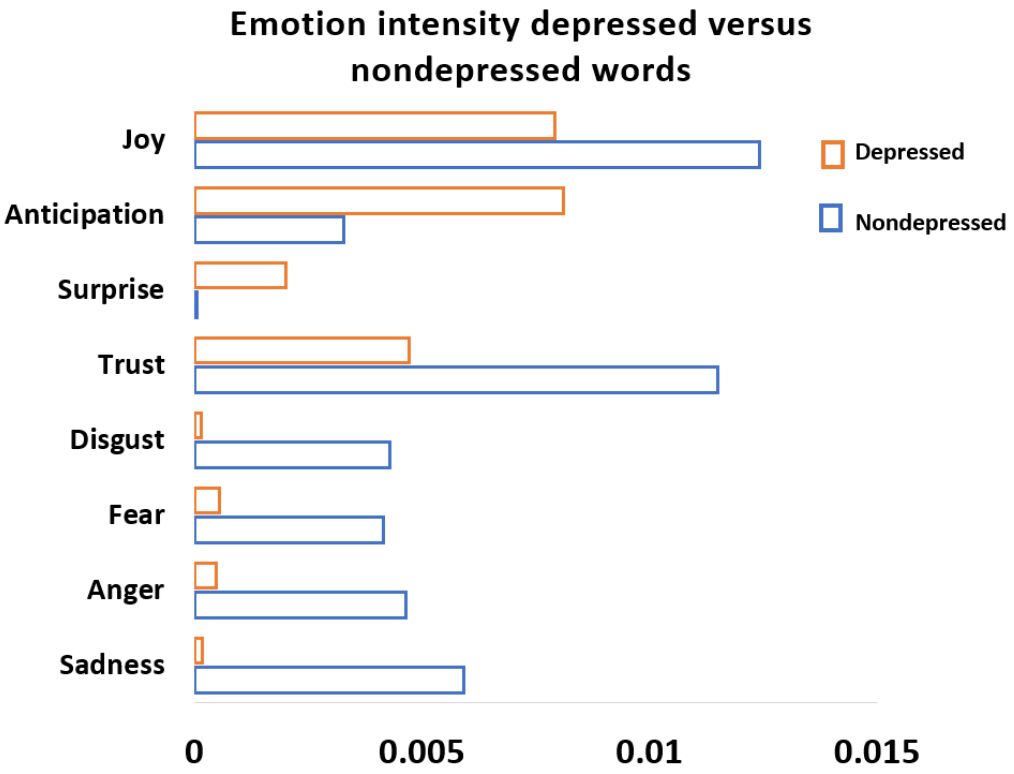
based on the most distinct depressed and nondepressed words as inputs for classifying the users with and without depression.

We conducted further analysis to examine the emotion intensities for each user group separately. To achieve this, we calculated the average intensity of each word in the emotion dimensions for the top 30 words in the depressed user group and the bottom 30 words in the nondepressed user group. The set of distinct emotion words used by user group g is $\{T_{g1}, T_{g2}, ..., T_{g30}\}$, where $g \in \{d, nd\}$. For each emotion word $t_g \in \{T_{g1}, T_{g2}, ..., T_{g30}\}$ in group g, we denoted the total count of word t_g's appearance in all users' social media postings in group g by TC_{tg} and the total number of words for all users as TC_U. Then, the average intensity of in group g is computed as: Sum of:

$$(TC_{tg} * V(t_g))/TC_U \quad \textbf{(7)}$$

We compared the average intensities of the most distinct emotions of these 2 user groups, as shown in Figure 5. Distinct depressed emotion words have high intensities in all emotion categories, suggesting that the user group with depression tends to have more emotional behavioral changes than the user group without depression. The user group with depression shows significant intensities in sadness, anger, fear, disgust, and trust compared to the user group without depression, suggesting that negative emotions are frequently expressed through social media postings by the user group with depression. The user group with depression also shows a higher intensity of joy than the user group without depression. This suggests that users with depression may try to use words related to joy to cheer themselves up, or they are seeking joy.

Figure 5. Comparison of average emotion intensities of the most distinct emotions of user groups with depression and without depression.



Analysis of Personality Features

As discussed in the Discover Personality Features section, we used the IPIP questionnaire to assess individuals on 5 dimensions of personality according to the Big Five Factor Model. For each user, we obtained a score for each of the 5 personality dimensions to form a feature vector for the personality feature of that user. Table 1 provides an example illustrating how we obtain the feature vector for each user. This personality feature vector comprises 5 values, each representing 1 of the 5 personality dimensions. The score for each dimension ranges from 1 to 5. For example, as shown in Table 1, the first user, U_1, scores 4 for openness, 2.25 for conscientiousness, 2.2 for extraversion, 3.6 for agreeableness, and 2.8 for neuroticism. Thus, U_1's feature vector for personality is <4, 2.25, 2.2, 3.6, 2.8 >. Similarly, we obtained the feature vector for U_2 as <3.8, 3.75, 3.4, 4.4, 2.15> and for U_3 as <3.55, 3.2, 3.2, 3.35, 2.9>, and so on. The feature vector for personality will be used to run experiments for depression detection.

XSL•FO
RenderX

Table 1. Examples of personality feature vectors for 3 users across 5 personality traits.

User	The feature vector <openness, conscientiousness, extraversion, agreeableness, neuroticism> is a numerical representation of personality traits based on the Big Five model used to describe and analyze individual characteristics in a structured format.
U_1[a]	<4, 2.25, 2.2, 3.6, 2.8>
U_2	<3.8, 3.75, 3.4, 4.4, 2.15>
U_3	<3.55, 3.2, 3.2, 3.35, 2.9>

[a]U_1 represents first user, U2 represent second user, and so on.

Analysis of Contextual Features

As outlined in the Discover Contextual Features section, we first used SBERT [59] to learn sentence embeddings based on each word's contextual information in a sentence for each user's social media posting to obtain user-level contextual features. The sentences users discuss in their social media postings are transformed into fixed-sized vectors (768 dimensions), which are the representation of the sentences. We conducted cluster analysis based on the SBERT feature vectors to obtain more valuable insights to support clinical decisions. We conducted K-means clustering separately for the user groups with and without depression to analyze different contents or topics different user groups discuss on social media. We found the optimal number of clusters for each user group based on the "silhouette" method. We chose 2 and 4 to be the optimal number of clusters for the user groups with and without depression, respectively, by experimenting with silhouette scores with regard to a range of clusters (ie, we choose the number of clusters with the highest silhouette score as the optimal number).

We then grouped users belonging to the same cluster together and obtained the aggregated social media postings for each cluster. Because the nondepressed group had fewer clusters than the depressed group, we obtained the top 20 words from each depressed cluster based on the term frequency and the top 40 words for each nondepressed cluster to have an equal number of top words from each user group for comparison.

The top words discussed in each cluster for each user group are shown in Multimedia Appendices 6 and 7.

On the basis of the top words, we conducted further analysis. Swear words and negative sentiments, such as "ass," "f**king," and "tear," were unique to the user group with depression, as shown in Textbox 1. This group also uses words related to family members, such as "dad" and "daughter," and spiritual words, such as "pray," "bless," and "trust." This may suggest that the user group with depression tends to seek comfort from family or the spiritual world. They also talk about school-related terms such as "teacher" and "college." This might suggest that the user group with depression has depressed symptoms due to pressure from school. The non-depressed group typically uses fewer negative sentiment words. The user group without depression uniquely mentioned words related to being forward-looking or future, such as "future," "forward," "chance," and "step"; optimistic actions, such as "pass," "win," and "worth"; or positive words, such as "power" and "perfect." They also used words suggesting the beginning of something, such as "awake" and "begin."

Textbox 1. Comparison of unique words used by users with and without depression.

Unique words used by users with depression

- "Awake", "begin", "burn", "chance", "comment", "cut", "exam", "Facebook", "fear", "final", "forward", "future", "human", "pass", "paste", "perfect", "point", "power", "problem", "self", "side", "step", "thought", "truth", "weather", "win", and "worth".

Unique words used by users without depression

- "Ass", "band", "beat", "bless", "dad", "daughter", "chocolate", "college", "easy", "fly", "fxxking", "kind", "omg", "pray", "repost", "seriously", "tear", "teacher", "text", and "trust".

Analysis of Social Interaction Influence Features

Next, we used the proposed psychological and contextual features, along with the social network structure derived from social connections, to extract social interaction influence features. We created social interaction influence features for each user and each type of proposed feature to show the extent of influence exerted by user groups with and without depression on the target user. As shown in Table 2, the example illustrates the affective, personality, and contextual influence of neighbors with and without depression, respectively, on 4 users, U_1 (user 1), U_2 (user 2), U_3 (user 3), and U_4.(user 4) We first extracted each user's affective, personality, and contextual features. For each user, U_i, and each feature, f, we obtained feature vectors $FV^f_{U,d}$ and $FV^f_{U,nd}$ that represent the features from user U_i 's social connections with and without depression, respectively.

XSL•FO
RenderX

Table 2. Examples of affective, personality, and contextual influences on four users, stemming from interactions with neighbors with depression and without depression, respectively. Each value represents the proportion or impact (as a fraction of 1) of the specific influence type (affective, personality, contextual) on the given user from neighbors with or without depression.

User	Feature type					
	Affective		Personality		Contextual	
	Depressed	Nondepressed	Depressed	Nondepressed	Depressed	Nondepressed
U_1[a]	0.736	0.264	0.677	0.323	0.668	0.332
U_2	0.566	0.434	0.514	0.486	0.396	0.604
U_3	0.282	0.718	0.405	0.595	0.464	0.536
U_4	0.206	0.794	0.191	0.809	0.265	0.735

[a]U_1 represent first user, U2 represent second user, and so on.

These feature vectors are inputs for extracting the social interaction influence features. In the case of affective features, the scores obtained from the users' group with depression are higher than those from the users' group without depression for U_1 and U_2. This suggests a stronger social interaction influence from the user with depression compared to the user without depression for U_1 and U_2. As a result, these 2 users exhibited emotions that are closer to those of their friends with depression and further from those of their friends without depression. In contrast, U_3 and U_4 exhibited greater similarity to their friends without depression compared to their friends with depression, suggesting they experienced a stronger influence from the friends without depression. Thus, $SIC_d(U,Y)$ affective is larger than $SIC_{nd}(U,Y)$ affective for user U_1 and U_2 for emotion features. Likewise, for personality and contextual features, higher scores suggest a higher influence. By comparing the scores between the user groups with depression and without depression, a higher score for a feature in 1 group suggests that the users are more aligned with that particular user group regarding that feature. This observation suggests that within the social networks, the user groups with depression and without depression may have different effects on their friends. Users who are more influenced by the user group without depression tend to display a higher degree of resemblance in affective, personality, or contextual information with the user group without depression, and vice versa.

Experiment Results: Depression Prediction Using ML Methods

The proposed psychological and contextual analysis NLP framework has the potential to enhance intelligent decision-making in the mental health domain. Our framework can enhance the reliability and knowledge base of decision support systems for health care providers, facilitating more accurate diagnoses and timely referrals and interventions, while also simplifying treatment plans. We found that incorporating affective, contextual, and personality features outperformed traditional baselines (improving performance compared to traditional baselines with an average improvement of 6% in accuracy and 10% in F_1-score) and achieved performance comparable to state-of-the-art baselines. Furthermore, the addition of social interaction influence enhanced performance across all metrics.

We conducted experiments using ML methods with the proposed features for depression detection to demonstrate the performance of our framework. All ML implementations were conducted using Python's *scikit-learn* library. The experiments were conducted on a ThinkPad (X1 Yoga Gen 8 Intel) with an Intel Core i7 CPU, 16 GB of RAM, running Windows 10. We used 5-fold cross-validation for model evaluation. Four ML methods were used: (1) logistic regression, (2) neural network, (3) k-nearest neighbors (KNN), and (4) random forest. We used an 80/20 split for training and testing the models. Specifically, we reported the accuracy of depression detection as part of our evaluation:

where TP=true positives, TN=true negatives, FP=false positives, and FN=false negatives.

In addition, we reported recall and precision scores. The recall represents the ratio of correctly predicted users with depression among all actual users with depression, while precision measures the proportion of users with depression correctly identified as depressed relative to the total number of users predicted as depressed, that is:

We also reported F_1-score, which is the combination of precision and recall:

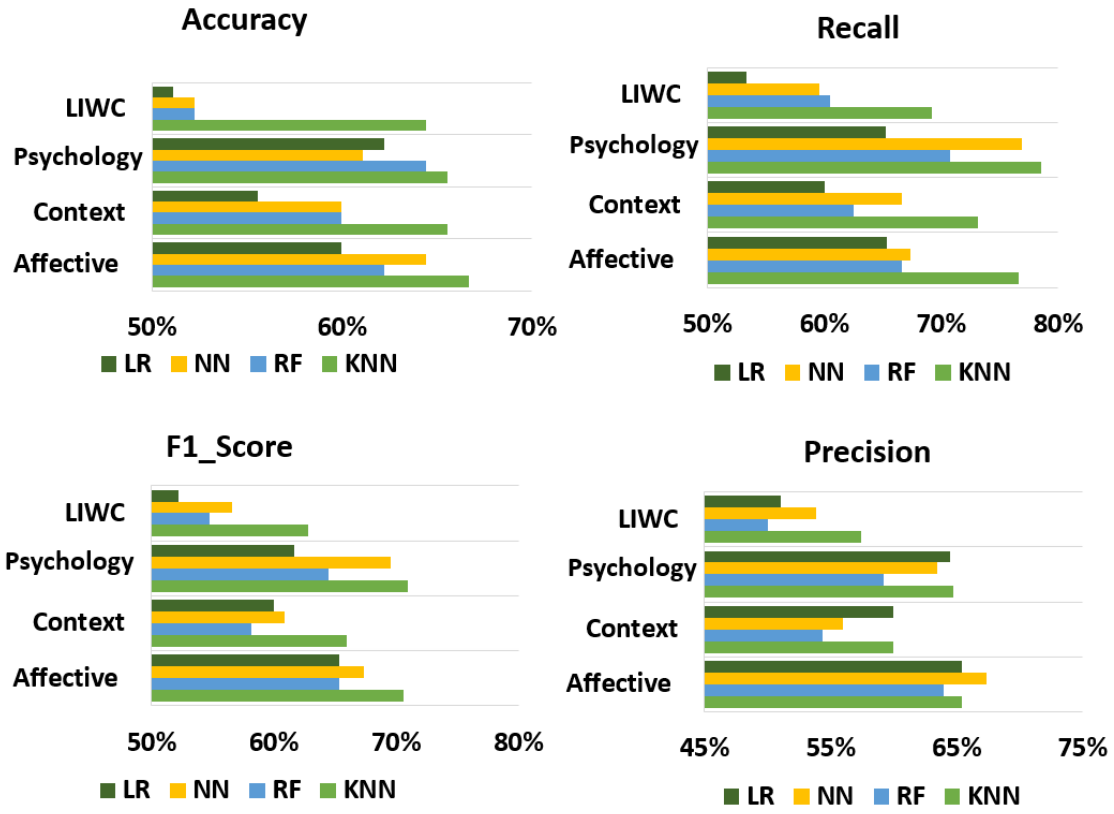
To assess the predictive capability of our proposed framework, we leveraged psychological (eg, affective and personality), contextual, and social interaction influence features to detect depression. We used the Linguistic Inquiry and Word Count (LIWC) [64] feature as the baseline, which is easily replicable to provide a comparative benchmark with the same data. The LIWC feature is widely used in depression detection because it offers reliable, unbiased, and effective measures [65-68].

We first conducted experiments using psychological (eg, affective and personality) and contextual features. The performance of various ML methods for different sets of features is illustrated in Figure 6. KNN showed the highest accuracy among all other ML models, with 69% for the LIWC feature,

XSL•FO
RenderX

76.6% for the affective features, 73% for contextual features, and 79% for the personality features. For precision, although neural network showed a higher performance of 67% compared to 66% under KNN for the affective features, KNN showed better performance compared to other ML methods for all other features. KNN also indicated the highest recall and F_1-score compared to other ML methods among all feature sets. Overall, KNN performed the best among all ML methods used in this study. Therefore, we conducted an additional set of experiments with different feature sets (F1-F9 correspond to different feature sets) listed in Table 3, using KNN, to assess the predictive power of depression detection based on the proposed features for each feature set.

Figure 6. Performance comparison of different machine learning methods across 4 metrics using Linguistic Inquiry and Word Count (LIWC), personality (psychology), contextual, and affective features. KNN: k-nearest neighbors; LR: logistic regression; NN: neural network; RF: random forest.



XSL•FO
RenderX

Table 3. Different feature sets using KNNa to evaluate the predictive power of depression detection.

Feature short name	Feature variables
Baseline1	LIWC[b] features
Baseline2	FastText
Baseline3	CNN[c]
F1	Affective features
F2	Contextual features
F3	Personality features
F4	Affective features, contextual features, and personality features
F5	Affective influence features, contextual influence features, and personality influence features
F6	Affective features and affective influence features
F7	Contextual features and contextual influence features
F8	Personality features and personality influence features
F9	Affective features, contextual features, and personality features and affective influence features, contextual influence features, and personality influence features

[a]KNN: k-nearest neighbors.

[b]LIWC: Linguistic Inquiry and Word Count.

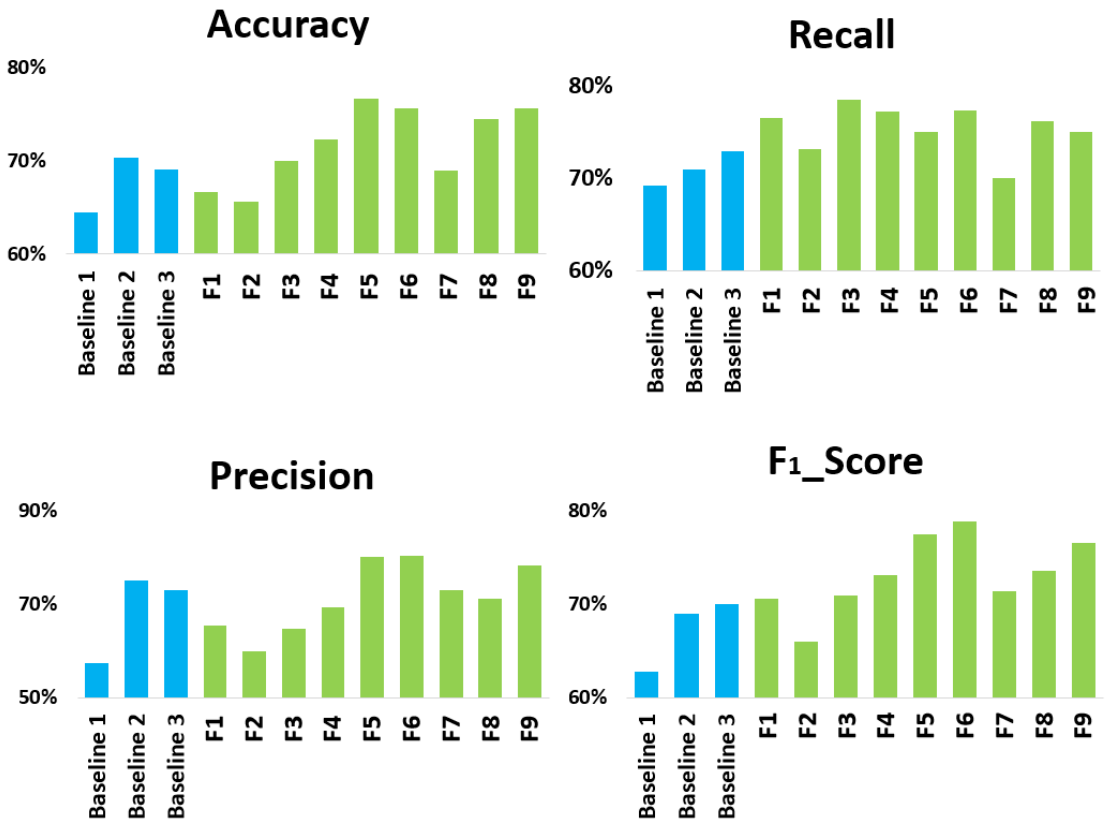
[c]CNN: convolutional neural network. F1-F9 correspond to different feature sets listed in Table 3, F1 is feature set 1 which is Affective Features as listed in Table 3, F2 is feature set 2 which is listed as Contextual Features in Table 3 and so on.

We treated the LIWC feature as baseline 1 and compared it against the proposed affective, contextual, personality, and corresponding social interaction influence features (eg, affective influence, contextual influence, and personality influence features) using KNN. We also compared our model with state-of-the-art approaches that used pretrained models for text processing combined with ML techniques for detection, such as FastText with XGBoost (baseline 2, as used in the study by Marriwala and Chaudhary [69]) and deep learning architectures, including convolutional neural networks for text processing and detection (baseline 3, as described in the study by Ghosal and Jain [39]).

As shown in Figure 7, the feature sets incorporating the proposed features significantly improved the detection capability compared to the LIWC feature across all performance measures. Our results are comparable to those achieved by baseline 2 (FastText with XGBoost) and baseline 3 (convolutional neural networks); however, the combination of all affective, personality, and contextual features, along with their influence features, yields even stronger performance. The inclusion of emotional and personality-driven factors, combined with social interaction features within our framework, offers a unique advantage. This approach delivers richer insights and a higher interpretive value compared to previous models, while maintaining high interpretability—a crucial aspect for mental health applications.

XSL•FO
RenderX

Figure 7. Performance comparison across different feature sets and baselines.



Among the proposed features, the model with affective features (F1) and the model with personality features (F3) performed better than the model with contextual features (F2). The model with affective features, personality features, and contextual features (F4) had the best performance in all measures except recall compared to F1, F2, and F3, while F3 showed better performance in recall compared to F4. The relatively strong performance of F2 showed the model's ability to effectively distinguish users with depression, suggesting its capacity to identify depression-specific language patterns. When considering the social interaction influence features for each feature type, the model with affective influence, personality influence, and contextual influence features (F5) showed higher performance compared to F4 with respect to accuracy, precision, and F_1-score. F3 and F4 showed better recall compared to F5. Compared to affective features (F1), the combination of affective features and the corresponding social interaction influence of affective features—affective influence features (F6)—performed better across all performance metrics. The same pattern applies for F7 with the combination of contextual features and the corresponding contextual influence features and for F8 with the combination of personality features and the corresponding personality influence features; they both performed better compared to the model with contextual features (F2) and personality features (F3) alone across all performance metrics. The model with all proposed features, including affective, personality, contextual, and the corresponding social interaction influence features (F9), showed high performance but not the highest among all models; this implies that simply including all features may not necessarily enhance performance. All proposed features outperformed baselines; notably, F5 and F6 demonstrated the strongest overall performance.

To examine how each feature impacts model performance, we performed permutation importance. This approach provides insights into how each feature affects the model's accuracy by measuring the change in performance when individual features are randomly shuffled [70]. Multimedia Appendix 8 shows the top 10 influential features, highlighting those most impactful for depression detection. The results indicate that personality features such as neuroticism, extraversion, and agreeableness are highly significant, followed by social interaction influence based on emotion, personality, and contextual measures. Finally, the affective feature, anticipation score, is the most influential among all emotions. F1-F9 represent short name of 9 feature sets we used in the experiments

Discussion

Principal Findings

We contribute by proposing a methodological framework using NLP and ML techniques to extract features and patterns, thereby comprehensively understanding depression-related symptoms from psychological and contextual perspectives. Building on the foundations of social influence theory, which points out that the way individuals behave, think, or act is influenced by their social environment, we extracted social interaction influence features based on social media postings. These features considered both psychological and contextual features (eg, users' emotions, psychological states, and context of communication) from users' social media postings and the social network structure of social media. They reflected the influence that a

XSL•FO
RenderX

user's social connections may have on their mental well-being. The combination of all affective, personality, and contextual features, along with their influence features, yields strong performance compared to baselines. F5 (which represent affective influence features, contextual influence features, and personality influence features in Table 3) and F6 (affective features and affective influence features) show particularly strong performances. Our study, designed to complement clinical diagnoses, delivers results that are sufficiently robust to support timely interventions and practical applications in mental health monitoring. By achieving comparable results to state-of-the-art methods, our approach demonstrates strong performance while adding value through richer insights and high interpretability. These patterns bridge the gap between model accuracy and interpretability, providing more actionable and inseparable insights that can aid clinicians in better understanding users' mentality and supporting their decision-making processes. This framework has the potential to expedite the diagnostic process and provide timely referral services to individuals at high risk, facilitating early intervention or treatment before more severe mental health conditions develop. With an improved understanding of users' mental states reflected in the psychological and contextual aspects, clinicians can more easily detect significant changes in users' mental well-being. This allows clinicians to provide well-supported guidance during early intervention and facilitate follow-ups in treatment plans.

Research in psychology has verified the theoretical connections between patients' mental health conditions and certain linguistic features. Various studies establish that individuals experiencing certain mental illnesses exhibit specific verbal behavior. Therefore, there is an increasing trend toward using rich sources of text data and information from social media to examine users' sentiments and behaviors in mental health studies [16,17]. Studies show that frequent negative affect expression has higher connections with depressive states, while positive affect tends to have a lower connection to depression [23-25]. Thus, emotions play an essential role in diagnosing mental illnesses such as depression [11]. Various studies have characterized emotions from language use of users disclosing their mental illnesses on social media to detect depression [68,71,72]. Personality is another important aspect of analyzing depression. Personality has been used to distinguish social media users' activities, posting content, and language use habits [26,27,73]. Several studies show a link between personality types and depression, and personality features can be used in depression diagnosis and treatments [28,50,55].

In addition to psychological factors, pretrained language models have shown strong contextual modeling ability; they are used in understanding texts in various applications [58]. For example, Haque et al [74] leveraged pretrained language models to detect early signs of suicidal thoughts, improve the diagnosis process, and provide people with appropriate treatment on time. Jiang et al [75] used deep contextualized pretrained word representation to build strong predictive models to detect different mental health problems from social media. Murarka et al [37] built a multiclass model using the pretrained language model to classify 5 prominent types of mental illnesses by analyzing unstructured user data on Reddit. Social media plays a critical role in shaping attitudes, beliefs, and behaviors through interactions [76,77]. Social influence theory suggests that individuals' mental states and social interactions affect their language and behaviors on social media [32,78]. While previous research has focused on linguistic, emotional, and personality factors, it often overlooks the impact of social interactions [37,68]. Our study aims to bridge this gap by analyzing social influence and interactions alongside psychological and contextual factors. This comprehensive framework can aid in improving mental health diagnosis and treatment plans.

Our study has several practical implications. By providing mental health professionals with a deeper and more comprehensive insight into a user's mental state, our framework can help accelerate the depression diagnosis process. It also enables timely referral services to individuals who may be unaware of their mental states or have limited access to mental health resources. Moreover, it is beneficial to alert health care management to offer early intervention or treatment to individuals at risk of developing more severe mental health conditions. Furthermore, by gaining a deeper insight into users' behaviors reflected in different aspects, mental health professionals can identify the significant changes in users' emotional, psychological, and mental health states to provide more justified recommendations during the early intervention and facilitate easier follow-ups in treatment plans.

Other domains that analyze social interactions and language use can also adopt our framework. In the general health domain, our framework offers valuable assistance in discovering positive and negative opinions by studying user-generated content from health-based social platforms. These platforms serve as spaces where users, including patients, physicians, and caregivers, can pose questions and seek advice from other users. By analyzing emotions, personality, and contextual information and the influence of social interaction, medical professionals and caregivers can effectively monitor different health conditions among patients and individuals at risk. This allows them to provide timely advice, referral services, or interventions, ultimately enhancing health care quality in these online communities. In the finance domain, for example, StockTwits is an online platform where investors exchange ideas, opinions, and sentiments regarding financial markets [79]. Our framework can be used to examine how emotional shifts, personality differences, and contextual information within sentences influence investors' decision-making processes. It also enables the examination of how social interactions among different investors may influence their investment decisions.

Limitations

While analyzing users' postings on social media enables the identification of mental states through means unavailable offline, our research does not consider the socioeconomic status, idiosyncratic behaviors "behind the scenes," or dynamic language data. For future studies, complementary sources of behavioral data and dynamic data, such as activity logs or search history, could be integrated with health records, such as records of prescriptions or health care claims data, to offer a more comprehensive understanding of the mental health status of an individual. Moreover, the expression of individuals with

XSL•FO
RenderX

depression may vary across different intensity levels of their depression. As a result, investigating how language use differs and evolves across different stages of depression could be a promising future direction, potentially leading to more personalized treatment plans. In addition, we plan to incorporate domain-specific lexicons and more specialized datasets to further enhance the model's accuracy in capturing nuanced depressive language. We also aim to fine-tune SBERT embeddings using mental health–specific corpora, which could improve performance in detecting depression-related language. Furthermore, after comprehensively understanding the patterns and trends of users with depression, a database system can be developed to store significant behaviors or language use changes. Updated classifiers could be trained using real-time data from online platforms or other sources. The system could automatically detect users who are at risk of depression even when they are unaware of their mental conditions and recommend suitable treatments based on information extracted from previous instances. We also plan to incorporate more recent and real-world datasets with similar features, extending the analysis beyond mental health contexts to various domains of social network analysis. This broader approach will facilitate a more comprehensive exploration across diverse scenarios, enhancing the framework's relevance and adaptability.

Conclusions

Our study illustrates a psychological and contextual analysis NLP framework that uses affective, personality, and contextual information. We further explore the influence of social interaction shaped by these elements to monitor depression behaviors on social media using NLP and ML. The proposed framework highlights the potential of using emotion intensities, personality measurement, pretrained language models, and social interactions among users on social media as features or patterns for predicting depression in individuals. Our research shows the potential of developing a better foundation for predicting the risk of depression and offering referrals, early intervention, and treatment plans via the analysis of social media data. It presents promising potential as an intelligent decision-making tool integrating information from multiple channels, such as social networks.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The process of extracting emotion intensity features and emotion word features to represent the affective features. We extracted emotion words from social media posts of users with and without depression and calculated the ratio differences in relative term frequency to identify distinct emotion words. We also combined the intensity levels across 8 emotion dimensions with emotion word frequency, computing the average intensity to form an emotion intensity feature.
[PNG File , 93 KB - formative_v9i1e60286_app1.png]

Multimedia Appendix 2
The Big Five Factor Model—personality traits with descriptions, showing the behavioral tendencies associated with low and high scores for each trait.
[PDF File (Adobe PDF File), 13 KB - formative_v9i1e60286_app2.pdf]

Multimedia Appendix 3
Discovery process of contextual features. We aggregated each user's social media postings and applied Sentence Bidirectional Encoder Representations (SBERT) from Transformers to generate sentence embeddings, capturing the contextual information of each word within a sentence. These embeddings serve as user-level contextual features. Then, we performed cluster analysis using K-means clustering, applied separately to the sentence embeddings of user groups with depression and without depression to analyze the content discussed by different user groups.
[PNG File , 108 KB - formative_v9i1e60286_app3.png]

Multimedia Appendix 4
Discovery process of social interaction influence features. For each user, we obtained feature vectors from the psychological and contextual features identified earlier. We then analyzed the influence of users with depression and without depression on the target user separately, based on the emotion, personality, and context influence features.
[PNG File , 19 KB - formative_v9i1e60286_app4.png]

Multimedia Appendix 5
Algorithm for discovering social interaction influence features.

XSL•FO
RenderX

[PNG File , 104 KB - formative_v9i1e60286_app5.png]

Multimedia Appendix 6
The most frequently discussed words across the four clusters for the user group with depression.
[PNG File , 254 KB - formative_v9i1e60286_app6.png]

Multimedia Appendix 7
The most frequently discussed words across the two clusters for the user group without depression.
[PNG File , 244 KB - formative_v9i1e60286_app7.png]

Multimedia Appendix 8
The top 10 influential features, highlighting the most impactful features for depression detection.
[PNG File , 21 KB - formative_v9i1e60286_app8.png]

References

1. Williams JW, First M. Diagnostic and statistical manual of mental disorders. In: McNutt JG, editor. Encyclopedia of Social Work. Oxford, UK: Oxford University Press; 2013.
2. Zhang T, Schoene AM, Ji S, Ananiadou S. Natural language processing applied to mental illness detection: a narrative review. NPJ Digit Med 2022 Apr 08;5(1):46 [FREE Full text] [doi: 10.1038/s41746-022-00589-7] [Medline: 35396451]
3. Depression. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2023-07-13]
4. Evans-Lacko S, Aguilar-Gaxiola S, Al-Hamzawi A, Alonso J, Benjet C, Bruffaerts R, et al. Socio-economic variations in the mental health treatment gap for people with anxiety, mood, and substance use disorders: results from the WHO World Mental Health (WMH) surveys. Psychol Med 2018 Jul;48(9):1560-1571 [FREE Full text] [doi: 10.1017/S0033291717003336] [Medline: 29173244]
5. Yang K, Zhang T, Ananiadou S. A mental state knowledge–aware and contrastive network for early stress and depression detection on social media. Inf Process Manage 2022 Jul;59(4):102961. [doi: 10.1016/j.ipm.2022.102961]
6. Motrico E, Mateus V, Bina R, Felice E, Bramante A, Kalcev G, et al. Good Practices in perinatal mental health during the COVID-19 pandemic: a report from task-force RISEUP-PPD COVID-19. Clínica y Salud 2020;31(3):155-160. [doi: 10.5093/clysa2020a26]
7. Coronavirus and depression in adults, Great Britain. Office for National Statistics. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/articles/coronavirusanddepressioninadultsgreatbritain/julytoaugust2021 [accessed 2023-07-29]
8. Sheehan DV. Depression: underdiagnosed, undertreated, underappreciated. Manag Care 2004 Jun;13(6 Suppl Depression):6-8. [Medline: 15293765]
9. Thombs BD, Kwakkenbos L, Levis AW, Benedetti A. Addressing overestimation of the prevalence of depression based on self-report screening questionnaires. CMAJ 2018 Jan 15;190(2):E44-E49 [FREE Full text] [doi: 10.1503/cmaj.170691] [Medline: 29335262]
10. Lin H, Jia J, Nie L, Shen G, Chua T. What does social media say about your stress? In: Proceedings of the 25th International Joint Conference on Artificial Intelligence. 2016 Presented at: IJCAI '16; July 9-15, 2016; New York, NY p. 3775-3781 URL: https://dl.acm.org/doi/abs/10.5555/3061053.3061147
11. Zhang T, Yang K, Ji S, Ananiadou S. Emotion fusion for mental illness detection from social media: a survey. Inf Fusion 2023 Apr;92:231-246. [doi: 10.1016/j.inffus.2022.11.031]
12. Zhang T, Yang K, Alhuzali H, Liu B, Ananiadou S. PHQ-aware depressive symptoms identification with similarity contrastive learning on social media. Inf Process Manag 2023 Sep;60(5):103417. [doi: 10.1016/j.ipm.2023.103417]
13. Issaka B, Aidoo EA, Wood SF, Mohammed F. "Anxiety is not cute" analysis of twitter users' discourses on romanticizing mental illness. BMC Psychiatry 2024 Mar 21;24(1):221 [FREE Full text] [doi: 10.1186/s12888-024-05663-w] [Medline: 38515062]
14. De Choudhury M. Role of social media in tackling challenges in mental health. In: Proceedings of the 2nd international workshop on Socially-aware multimedia. 2013 Presented at: SAM '13; October 21, 2013; Barcelona, Spain p. 49-52 URL: https://dl.acm.org/doi/10.1145/2509916.2509921 [doi: 10.1145/2509916.2509921]
15. Reece AG, Reagan AJ, Lix KL, Dodds PS, Danforth CM, Langer EJ. Forecasting the onset and course of mental illness with Twitter data. Sci Rep 2017 Oct 11;7(1):13006 [FREE Full text] [doi: 10.1038/s41598-017-12961-9] [Medline: 29021528]
16. Chancellor S, De Choudhury M. Methods in predictive techniques for mental health status on social media: a critical review. NPJ Digit Med 2020 Mar 24;3(1):43 [FREE Full text] [doi: 10.1038/s41746-020-0233-7] [Medline: 32219184]
17. Ríssola EA, Losada DE, Crestani F. A survey of computational methods for online mental state assessment on social media. ACM Trans Comput Healthcare 2021 Mar 17;2(2):1-31. [doi: 10.1145/3437259]

XSL•FO
RenderX

18. Amanat A, Rizwan M, Javed AR, Abdelhaq M, Alsaqour R, Pandya S, et al. Deep learning for depression detection from textual data. Electronics 2022 Feb 23;11(5):676. [doi: 10.3390/electronics11050676]
19. Ghosh S, Ekbal A, Bhattacharyya P. A multitask framework to detect depression, sentiment and multi-label emotion from suicide notes. Cogn Comput 2021 Feb 05;14(1):110-129. [doi: 10.1007/S12559-021-09828-7]
20. Yoo M, Lee S, Ha T. Semantic network analysis for understanding user experiences of bipolar and depressive disorders on Reddit. Inf Process Manag 2019 Jul;56(4):1565-1575. [doi: 10.1016/j.ipm.2018.10.001]
21. Aragón ME, López Monroy AP, González-Gurrola L, Montes M. Detecting depression in social media using fine-grained emotions. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019 Presented at: NAACL '19; June 2-7, 2019; Minneapolis, Minnesota p. 1481-1486 URL: https://aclanthology.org/N19-1151.pdf [doi: 10.18653/v1/n19-1151]
22. Shickel B, Siegel S, Heesacker M, Benton S, Rashidi P. Automatic detection and classification of cognitive distortions in mental health text. In: Proceedings of the 20th International Conference on Bioinformatics and Bioengineering. 2020 Presented at: BIBE '20; October 26-28, 2020; Cincinnati, OH p. 21-55 URL: https://ieeexplore.ieee.org/document/9288097 [doi: 10.1109/bibe50027.2020.00052]
23. Settanni M, Marengo D. Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts. Front Psychol 2015;6:1045 [FREE Full text] [doi: 10.3389/fpsyg.2015.01045] [Medline: 26257692]
24. Cavazos-Rehg PA, Krauss MJ, Sowles S, Connolly S, Rosas C, Bharadwaj M, et al. A content analysis of depression-related Tweets. Comput Human Behav 2016 Jan 01;54:351-357 [FREE Full text] [doi: 10.1016/j.chb.2015.08.023] [Medline: 26392678]
25. Park J, Lee DS, Shablack H, Verduyn P, Deldin P, Ybarra O, et al. When perceptions defy reality: the relationships between depression and actual and perceived Facebook social support. J Affect Disord 2016 Aug;200:37-44. [doi: 10.1016/j.jad.2016.01.048] [Medline: 27126138]
26. Golbeck J, Robles C, Turner K. Predicting personality with social media. In: Proceedings of the 2011 Conference on Human Factors in Computing Systems. 2011 Presented at: CHI EA '11; May 7-12, 2011; Vancouver, BC p. 253-262. [doi: 10.1145/1979742.1979614]
27. Tadesse MM, Lin H, Xu B, Yang L. Personality predictions based on user behavior on the Facebook social media platform. IEEE Access 2018;6:61959-61969. [doi: 10.1109/ACCESS.2018.2876502]
28. Klein DN, Kotov R, Bufferd SJ. Personality and depression: explanatory models and review of the evidence. Annu Rev Clin Psychol 2011;7:269-295 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032210-104540] [Medline: 21166535]
29. Hakulinen C, Elovainio M, Pulkki-Råback L, Virtanen M, Kivimäki M, Jokela M. Personality and depressive symptoms: individual participant meta-analysis of 10 cohort studies. Depress Anxiety 2015 Jul 26;32(7):461-470 [FREE Full text] [doi: 10.1002/da.22376] [Medline: 26014798]
30. Nguyen T, Yates A, Zirikly A, Desmet B, Cohan A. Improving the generalizability of depression detection by leveraging clinical questionnaires. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. 2022 Presented at: ACL '22; May 22-27, 2022; Virtual Event p. 8446-8459 URL: https://aclanthology.org/2022.acl-long.578.pdf [doi: 10.18653/v1/2022.acl-long.578]
31. Unlu A, Truong S, Tammi T, Lohiniva AL. Exploring political mistrust in pandemic risk communication: mixed-method study using social media data analysis. J Med Internet Res 2023 Oct 20;25:e50199 [FREE Full text] [doi: 10.2196/50199] [Medline: 37862088]
32. Kelman HC. Compliance, identification, and internalization three processes of attitude change. J Conflict Resolut 1958 Mar 01;2(1):51-60. [doi: 10.1177/002200275800200106]
33. About mental illness. Government of Canada. 2017. URL: https://www.canada.ca/en/public-health/services/about-mental-illness.html [accessed 2024-01-29]
34. Neumann R, Strack F. "Mood contagion": the automatic transfer of mood between persons. J Pers Soc Psychol 2000 Aug;79(2):211-223. [doi: 10.1037//0022-3514.79.2.211] [Medline: 10948975]
35. Gleick ME. The Information: A History, a Theory, a Flood. New York, NY: Pantheon; 2012.
36. Mental health myths and facts. MentalHealth.gov. 2020. URL: https://www.samhsa.gov/mental-health/myths-and-facts [accessed 2023-07-23]
37. Murarka A, Radhakrishnan B, Ravichandran S. Classification of mental illnesses on social media using RoBERTa. In: Proceedings of the 12th International Workshop on Health Text Mining and Information Analysis. 2021 Presented at: HTMIA '21; April 19, 2021; Virtual Event p. 59-68 URL: https://aclanthology.org/2021.louhi-1.7.pdf
38. Cui B, Wang J, Lin H, Zhang Y, Yang L, Xu B. Emotion-based reinforcement attention network for depression detection on social media: algorithm development and validation. JMIR Med Inform 2022 Aug 09;10(8):e37818 [FREE Full text] [doi: 10.2196/37818] [Medline: 35943770]
39. Ghosal S, Jain A. Depression and suicide risk detection on social media using fastText embedding and XGBoost classifier. Procedia Comput Sci 2023;218:1631-1639. [doi: 10.1016/j.procs.2023.01.141]
40. Kosinski M, Matz SC, Gosling SD, Popov V, Stillwell D. Facebook as a research tool for the social sciences: opportunities, challenges, ethical considerations, and practical guidelines. Am Psychol 2015 Sep;70(6):543-556. [doi: 10.1037/a0039210] [Medline: 26348336]

XSL•FO
RenderX

41. Mehmood A, Natgunanathan I, Xiang Y, Hua G, Guo S. Protection of big data privacy. IEEE Access 2016;4:1821-1834. [doi: 10.1109/ACCESS.2016.2558446]
42. Goldberg LR, Johnson JA, Eber HW, Hogan R, Ashton MC, Cloninger CR, et al. The international personality item pool and the future of public-domain personality measures. J Res Pers 2006 Feb;40(1):84-96. [doi: 10.1016/j.jrp.2005.08.007]
43. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977 Jun 01;1(3):385-401. [doi: 10.1177/014662167700100306]
44. Lewinsohn PM, Seeley JR, Roberts RE, Allen NB. Center for Epidemiologic Studies Depression scale (CES-D) as a screening instrument for depression among community-residing older adults. Psychol Aging 1997 Jun;12(2):277-287. [doi: 10.1037//0882-7974.12.2.277] [Medline: 9189988]
45. Julian LJ, Gregorich SE, Tonner C, Yazdany J, Trupin L, Criswell LA, et al. Using the Center for Epidemiologic Studies Depression Scale to screen for depression in systemic lupus erythematosus. Arthritis Care Res (Hoboken) 2011 Jun;63(6):884-890 [FREE Full text] [doi: 10.1002/acr.20447] [Medline: 21312347]
46. Bollen J, Gonçalves B, Ruan G, Mao H. Happiness is assortative in online social networks. Artif Life 2011;17(3):237-251. [doi: 10.1162/artl_a_00034] [Medline: 21554117]
47. Mohammad SM. Word affect intensities. In: Proceedings of the 11th Edition of the Language Resources and Evaluation Conference. 2018 Presented at: LREC '18; May 7-12, 2018; Miyazaki, Japan URL: https://aclanthology.org/L18-1027.pdf
48. Morinaga S, Yamanishi K, Tateishi K, Fukushima T. Mining product reputations on the web. In: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2002 Presented at: KDD '02; July 23-26, 2002; Edmonton, AB p. 341-349 URL: https://dl.acm.org/doi/10.1145/775047.775098 [doi: 10.1145/775094.775098]
49. Rajaraman A, Ullman JD. Mining of Massive Datasets. Cambridge, MA: Cambridge University Press; 2011.
50. Bagby RM, Psych C, Quilty LC, Ryder AC. Personality and depression. Can J Psychiatry 2008 Jan;53(1):14-25. [doi: 10.1177/070674370805300104] [Medline: 18286868]
51. Ghosh S, Ekbal A, Bhattacharyya P. VAD-assisted multitask transformer framework for emotion recognition and intensity prediction on suicide notes. Inf Process Manag 2023 Mar;60(2):103234. [doi: 10.1016/j.ipm.2022.103234]
52. McCrae RR, Costa PT. Personality in Adulthood: A Five-Factor Theory Perspective. New York, NY: The Guilford Press; 1990.
53. McCrae RR, John OP. An introduction to the five-factor model and its applications. J Pers 1992 Jun;60(2):175-215. [doi: 10.1111/j.1467-6494.1992.tb00970.x] [Medline: 1635039]
54. Preoţiuc-Pietro D, Eichstaedt J, Park G, Sap M, Smith L. The role of personality, age, and gender in tweeting about mental illness. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2015 Presented at: CLPsych '15; June 5, 2015; Denver, CO p. 21-30 URL: https://aclanthology.org/W15-1203.pdf [doi: 10.3115/v1/w15-1203]
55. Yang X, McEwen R, Ong LR, Zihayat M. A big data analytics framework for detecting user-level depression from social networks. Int J Inf Manage 2020 Oct;54:102141. [doi: 10.1016/j.ijinfomgt.2020.102141]
56. Muñoz S, Iglesias CA. A text classification approach to detect psychological stress combining a lexicon-based feature framework with distributional representations. Inf Process Manag 2022 Sep;59(5):103011. [doi: 10.1016/j.ipm.2022.103011]
57. Yao X, Chen S, Yu G. Effects of members' response styles in an online depression community based on text mining and empirical analysis. Inf Process Manag 2023 Mar;60(2):103198. [doi: 10.1016/j.ipm.2022.103198]
58. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2021 Presented at: NAACL-HLT '21; June 2-7, 2019; Virtual Event p. 4171-4186 URL: https://aclanthology.org/N19-1423.pdf [doi: 10.18653/v1/N19-1423]
59. Reimers N, Gurevych I. Sentence-BERT: sentence embeddings using siamese BERT-networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. 2019 Presented at: EMNLP-NLP '19; November 3-7, 2019; Hong Kong, China p. 3982-3992 URL: https://aclanthology.org/D19-1410.pdf [doi: 10.18653/v1/d19-1410]
60. Bowman SR, Angeli G, Potts C, Manning CD. A large annotated corpus for learning natural language inference. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015 Presented at: EMNLP '15; September 17-21, 2015; Lisbon, Portugal p. 632-642 URL: https://aclanthology.org/D15-1075.pdf [doi: 10.18653/v1/d15-1075]
61. Williams A, Nangia N, Bowman SR. A broad-coverage challenge corpus for sentence understanding through inference. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2018 Presented at: NAACL '18; June 1-6, 2018; New Orleans, LA p. 1112-1122 URL: https://aclanthology.org/N18-1101.pdf [doi: 10.18653/v1/n18-1101]
62. Bagozzi RP, Dholakia UM. Intentional social action in virtual communities. J Interact Mark 2002 May 01;16(2):2-21. [doi: 10.1002/dir.10006]
63. Slonim T. Verbal behavior in individuals with generalized anxiety disorder and depressive disorders. City University of New York. 2014. URL: https://academicworks.cuny.edu/gc_etds/386 [accessed 2024-07-23]

XSL•FO
RenderX

64. Liwc2015: how it works. Linguistic Inquiry and Word Count. 2020. URL: http://liwc.wpengine.com/how-it-works/ [accessed 2021-01-23]
65. Coppersmith G, Dredze M, Harman C. Quantifying mental health signals in Twitter. In: Proceedings of the 2014 Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2014 Presented at: CLPsych '14; June 27, 2014; Baltimore, MD p. 51-60 URL: https://aclanthology.org/W14-3207.pdf [doi: 10.3115/v1/w14-3207]
66. Coppersmith G, Dredze M, Harman C, Hollingshead K. From ADHD to SAD: analyzing the language of mental health on twitter through self-reported diagnoses. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2015 Presented at: CLPsych '15; June 5, 2015; Denver, CO p. 1-10 URL: https://aclanthology.org/W15-1201.pdf [doi: 10.3115/v1/w15-1201]
67. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. Proc Int AAAI Conf Web Soc Media 2021 Aug 03;7(1):128-137. [doi: 10.1609/icwsm.v7i1.14432]
68. Chen X, Sykora MD, Jackson TW, Elayan S. What about mood swings: identifying depression on Twitter with temporal measures of emotions. In: Proceedings of the 2018 Conference on World Wide Web. 2018 Presented at: WWW '18; April 23-27, 2018; Lyon, France p. 1653-1660 URL: https://dl.acm.org/doi/10.1145/3184558.3191624 [doi: 10.1145/3184558.3191624]
69. Marriwala N, Chaudhary D. A hybrid model for depression detection using deep learning. Measur Sens 2023 Feb;25:100587. [doi: 10.1016/j.measen.2022.100587]
70. Breiman L. Random forests. Machine Learn 2001;45:5-32 [FREE Full text] [doi: 10.1023/A:1010933404324] [Medline: 21816105]
71. Marerngsit S, Thammaboosadee S. A two-stage text-to-emotion depressive disorder screening assistance based on contents from online community. In: Proceedings of the 8th International Conference on Electrical Engineering Congress. 2020 Presented at: iEECON '20; March 4-6, 2020; Chiang Mai, Thailand p. 1-8 URL: https://ieeexplore.ieee.org/document/9077452 [doi: 10.1109/ieecon48109.2020.229524]
72. Wu W, Wu M, Yu K. Climate and weather: inspecting depression detection via emotion recognition. In: Proceedings of the 2022 IEEE International Conference on Acoustics, Speech and Signal Processing. 2022 Presented at: ICASSP '22; May 23-27, 2022; Singapore, Singapore p. 6262-6266 URL: https://ieeexplore.ieee.org/document/9746634 [doi: 10.1109/icassp43922.2022.9746634]
73. Sumner C, Byers A, Shearing M. Determining personality traits and privacy concerns from Facebook activity. Black Hat Brief 2011;11:197-221 [FREE Full text]
74. Haque F, Nur RU, Jahan SA, Mahmud Z, Shah FM. A transformer based approach to detect suicidal ideation using pre-trained language models. In: Proceedings of the 23rd International Conference on Computer and Information Technology. 2020 Presented at: ICCIT '20; December 19-21, 2020; Dhaka, Bangladesh p. 1-5 URL: https://ieeexplore.ieee.org/document/9392692 [doi: 10.1109/iccit51783.2020.9392692]
75. Jiang Z, Levitan SI, Zomick J, Hirschberg J. Detection of mental health from Reddit via deep contextualized representations. In: Proceedings of the 2020 International Workshop on Health Text Mining and Information Analysis. 2020 Presented at: HTMIA '20; November 20, 2020; Virtual Event p. 147-156 URL: https://aclanthology.org/2020.louhi-1.16.pdf [doi: 10.18653/v1/2020.louhi-1.16]
76. Kleinman SS. Social identification in a computer-mediated group for women in science and engineering. Sci Commun 2000 Jun 01;21(4):344-366. [doi: 10.1177/1075547000021004002]
77. McKenna KY, Bargh JA. Plan 9 from cyberspace: the implications of the internet for personality and social psychology. Pers Soc Psychol Rev 2000 Feb 01;4(1):57-75. [doi: 10.1207/s15327957pspr0401_6]
78. Park M, McDonald D, Cha M. Perception differences between the depressed and non-depressed users in Twitter. Proc Int AAAI Conf Web Soc Media 2021 Aug 03;7(1):476-485. [doi: 10.1609/icwsm.v7i1.14425]
79. Mahmoudi N, Docherty P, Moscato P. Deep neural networks understand investors better. Decis Support Syst 2018 Aug;112:23-34. [doi: 10.1016/j.dss.2018.06.002]

Abbreviations

BERT: Bidirectional Encoder Representations from Transformers
CES-D: Center for Epidemiologic Studies Depression
IPIP: International Personality Item Pool
KNN: k-nearest neighbors
LIWC: Linguistic Inquiry and Word Count
ML: machine learning
NLP: natural language processing
SBERT: Sentence Bidirectional Encoder Representations from Transformers
TFIDF: term frequency and inverse document frequency

XSL•FO
RenderX

Edited by A Mavragani; submitted 07.05.24; peer-reviewed by J Abbas, P Zadeh, S Shatabda; comments to author 25.10.24; revised version received 05.11.24; accepted 05.12.24; published 20.01.25.

Please cite as:
Yang X, Li G
Psychological and Behavioral Insights From Social Media Users: Natural Language Processing–Based Quantitative Study on Mental Well-Being
JMIR Form Res 2025;9:e60286
URL: https://formative.jmir.org/2025/1/e60286
doi:10.2196/60286
PMID:

©Xingwei Yang, Guang Li. Originally published in JMIR Formative Research (https://formative.jmir.org), 20.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Attitudes of German General Practitioners Toward eHealth Apps for Dementia Risk Reduction: Qualitative Interview Study

Adrian Schultz[1], BSc; Melanie Luppa[1], PD, Dr Rer Med; Markus Bleckwenn[2], Prof Dr Med; Steffi G Riedel-Heller[1*], MPH, Prof Dr Med; Andrea Zuelke[1*], Dr Rer Nat

[1]Institute of Social Medicine, Occupational Health and Public Health, Leipzig University, Leipzig, Germany

[2]Institute of General Practice, Leipzig University, Leipzig, Germany

[*]these authors contributed equally

Corresponding Author:
Andrea Zuelke, Dr Rer Nat
Institute of Social Medicine, Occupational Health and Public Health
Leipzig University
Philipp Rosenthal Str. 55
Leipzig, 04103
Germany
Phone: 49 03419715483
Email: andrea.zuelke@medizin.uni-leipzig.de

Abstract

Background: eHealth interventions constitute a promising approach to disease prevention, particularly because of their ability to facilitate lifestyle changes. Although a rather recent development, eHealth interventions might be able to promote brain health and reduce dementia risk in older adults.

Objective: This study aimed to explore the perspective of general practitioners (GPs) on the potentials and barriers of eHealth interventions for brain health. Understanding the perspective of GPs allows us to identify chances and challenges for implementing eHealth apps for dementia risk reduction.

Methods: We conducted semistructured expert interviews with 9 GPs working in an outpatient setting in and near Leipzig, Germany. Data were fully transcribed and analyzed using a process model of qualitative content analysis with codes and categories being constructed inductively and deductively.

Results: We found generally favorable but balanced views of eHealth apps for brain health. Eight themes were identified and elaborated on in the data as follows: “addressing dementia,” “knowledge about dementia,” “need for information,” “potential for prevention,” “chances for apps for prevention,” “development of apps for prevention,” and “barriers of apps for prevention.” GPs talked mostly about how and when to address dementia and the requirements for their use of eHealth apps for dementia prevention. GPs stated that they only addressed dementia once abnormalities were already present or less frequently when a patient or relative expressed a direct wish, while individual dementia risk or standardized diagnostic during routine check-ups were mentioned much less frequently. According to GPs, knowledge about dementia in patients was low; therefore, patients expressed little need for information on dementia risk factors and prevention in GP practices. Most patients wished for quick information regarding diagnostics, treatment options, and progression of the disease. GPs mentioned a lack of overview of the available eHealth apps and their content. They also expressed a fear of inducing health anxiety when talking to patients about risk factors and prevention.

Conclusions: GPs want patients to receive relevant and individualized information. Prerequisites for the use of eHealth apps for dementia prevention were app characteristics related to design and content. GPs need to address dementia more routinely, assess relevant risk factors, and aid patients in a preventive role. Concerns were expressed over limited effectiveness, overwhelming patients, limited use in clinical practice, and only targeting patients with an already low risk of dementia.

(JMIR Form Res 2025;9:e56310) doi:10.2196/56310

KEYWORDS

eHealth; dementia; primary care; lifestyle; risk factor; older adults; prevention; brain health

XSL•FO
RenderX

Introduction

Background

Dementia is one of the most common neurodegenerative diseases. In Germany alone, dementia is one of the 5 most common diseases prevalent at the time of death [1]. At the same time, dementia is connected with a high burden of illness and excess costs of illness that are 2.5 times higher when compared to people without dementia [2]. Without available curative therapies, risk factors for dementia are a primary focus for dementia prevention. According to the latest report of the Lancet Commission on Dementia Prevention, Intervention, and Care, 45% of dementia cases can be explained by 14 modifiable risk factors (low levels of education, hearing impairment, hypertension, smoking, obesity, depression, physical inactivity, diabetes, higher alcohol consumption, traumatic brain injury, air pollution, social isolation, high levels of low-density lipoprotein cholesterol, and visual impairment [3]). This underlines the potential of dementia risk reduction by risk factor modification. In Germany, 38% of dementia cases are estimated to be due to 11 modifiable risk factors [4]. The use of eHealth apps has garnered considerable attention as a potential tool to assist in promoting brain health and mitigating dementia risk [5].

General practitioners (GPs) working in primary care usually consult their patients regularly and could potentially identify risk factors and early markers of dementia. GPs are also perceived to be an important support body with a high level of competence and trust [6] and play a central role in advising patients on health-related matters. Therefore, the primary care setting is crucial for the assessment of dementia risk factors and early dementia diagnosis. However, existing research has shown that dementia risk reduction is not routinely conducted in primary care settings [7,8]. Only few existing studies examined the views of GPs on dementia risk reduction.

While older adults are the most rapidly growing share of internet users in Germany [9], their use of eHealth tools has been found to be significantly lower than in other (European) countries [10]. The general use of health apps was low in a population-based sample of older (aged ≥60 years) German adults conducted in 2016, with only 16.5% reporting previous use of health apps. Concerns about data protection and lack of trust in eHealth tools were the most commonly named barriers [11]. Although these findings indicate a rather skeptical view of older German adults on eHealth tools, findings from the @ktiv-trial, testing an eHealth intervention for depression, found higher levels of adherence in older than in younger users, indicating the potential benefits of eHealth interventions for older adults [12].

Digital interventions aimed at enhancing brain health and mitigating dementia risk factors are currently an important focus of research [13-15], which explores potential benefits and challenges of digital apps for dementia risk reduction, highlighting their capacity to deliver personalized interventions and facilitate self-management on a large scale. However, despite these advancements in the field, only limited attention has been directed toward understanding the perspective of GPs, who play a critical role in adoption and implementation of these apps into clinical practice. Therefore, knowing the potentials and barriers before the development and implementation of such measures can be important to enhance potential. To the best of our knowledge, no study so far has examined the attitudes of German GPs toward eHealth apps for dementia risk reduction.

This Study

Consequently, we conducted 9 qualitative expert interviews with German GPs practicing in the Leipzig area to gain insight into GPs' subjective experience with older patients, how GPs address topics of dementia risk factors, and the possibilities of eHealth apps for prevention and managing dementia risk. Qualitative research can provide an understanding of subjective experiences, attitudes, and challenges GPs face. Qualitative studies can help explore and explain social relations between the health care system and the way patients and providers interact. In addition, qualitative approaches allow depicting aspects that might be lost by transformation of experiences into numerical forms for proper statistical analysis [16,17]. We aimed to identify possible chances and hindrances of eHealth apps for prevention of dementia and challenges for implementation. Furthermore, we aimed to provide specific recommendations toward goals, content, challenges, and possible implementation of eHealth apps to reduce dementia risk.

Methods

Study Design

This study uses a qualitative semistructured interview design. We conducted interviews with GPs via telephone using guided interviews. Telephone-based interviews were conducted to increase convenience and flexibility for participating GPs. Furthermore, we aimed to reduce the risk of social desirability bias and keep technical requirements for GPs to a minimum, arguing for telephone- instead of video-based interviews. Guidelines for the interview were designed to cover a prespecified topic list focused on dementia prevention, attitudes toward eHealth apps for dementia prevention, and technology-related factors. Although participants were reminded that the interview focuses on dementia prevention, GPs frequently steered the conversation toward identifying and managing dementia. This divergence from the intended focus of prevention was noted and addressed in real time by interviewers. Participants were encouraged to address further topics not covered by the interview guidelines at the end of the interview. We used a process model of qualitative content analysis as developed by Mayring [18] to identify key themes using inductive and deductive coding.

Sample

Participants were GPs with practices in an outpatient setting (individual practice, joint practice, or medical center) in the city of Leipzig, Germany, and its surrounding region. No further inclusion or exclusion criteria were applied. Eligible GP practices were recruited from a database at the Institute of Social Medicine, Occupational Health and Public Health, including 53 GPs who had participated in previous studies or expressed interest in participating in future research. This sample contained

XSL•FO
RenderX

GP practices from all regions of the city of Leipzig, comprising urban and suburban areas, as well as practices from the environs of Leipzig. All 53 GPs were contacted via telephone and email, of which 5 (9%) refused participation. Of the remaining 48 (91%) GPs, 12 (23%) expressed interest in participation. In 3 (6%) cases, interviews could not be scheduled due to time constraints on the side of GPs; therefore, the final sample included 9 (17%) GPs. On the basis of previous experiences conducting qualitative interviews, as well as narrowly defined research questions and a relatively homogeneous study sample [19,20], we expected to reach data saturation after 8 interviews. Interviews were conducted in a timeframe of roughly 25 minutes each. Participating GPs received detailed study information, an informed consent form, account data forms for the transfer of the expense allowance, and a sociodemographic questionnaire in advance via postal service. At the beginning of the interviews, interviewers informed participants about the method of expert interviews and encouraged them to share their experiences and opinions openly.

Analysis

All interviews followed the same general structure. Before the interviews, participants were given a short introduction on the interview topic identical to the previously received detailed study information. Interviewers informed GPs that the goal of the study was to gather information on how informed older patients are about the risk and protective factors of dementia and their willingness to use eHealth apps to manage dementia risk. In addition, they were informed that there was an interview guideline and encouraged to share their own experiences and opinions. The interview guideline was developed for this study; a full English translation is provided in Multimedia Appendix 1.

First, interviewers asked GPs whether they address the topic of dementia and cognitive ability with their older patients. If so, they were further asked with which groups of patients they bring the topic up and whether older patients addressed dementia at appointments themselves. Second, GPs were asked to assess the knowledge of patients about dementia and respective risk or protective factors. In the third section of the interview, interviewers inquired if patients wished to receive information on dementia prevention. In the fourth section, interviewers gave a brief description of how modifiable risk factors, for example, managing cardiovascular risk, increasing physical and social activity, or optimizing diet, have a positive impact on cognitive ability. Interviewers then asked participants which risk factors they considered especially promising for dementia risk reduction, whether they bring up modifiable risk factors when talking about dementia with their patients, and which risk factors patients might already know. In the following section, participants were asked to assess the potential of eHealth apps for managing cognitive decline and reducing the risk of dementia. In the sixth section, participants were asked to imagine a respective eHealth tool that addressed modifiable risk factors (eg, diet, physical and social activity, cardiovascular risk) via standardized questionnaires and that allowed patients to receive individualized notifications and information on how to reduce their personal risk for dementia. Following this, GPs were asked to assess whether older patients would make use of this program, which patients they would recommend the program to, and whether they would recommend all of it to their own patients. In section 7, GPs were asked to elaborate on facilitators and barriers for such a program. In the last section, interviewers asked for the GP's opinion on the implementation of respective eHealth apps for the prevention of dementia and which information they would need to assess the use of the service. Technical requirements, content, criteria of quality, measurable effect, privacy, and cost were given as anchors. Closing the interview, participants were thanked for their collaboration and asked whether they had an opinion, question, or if any topic was not covered in the interview.

All interviews were audio-recorded, fully transcribed and content-analyzed using MAXQDA 2022. We formed codes inductively and deductively according to the process model developed by Mayring [18] and the analytic pathway 1 as identified by Raskind et al [21] on qualitative data analysis practices in health education and health behavior research. Three trained scientific project members with experience in qualitative research independently coded and revised all interviews. The coding scheme was derived from an initial coding template based on literature, the interview guide (deductive coding), and emerging codes from the interview material (inductive coding, second-level coding within a domain, and openness for open coding during this stage). Methodological rigor was ensured by agreeing on a rule and process model-based approach before coding, iterative processing of reviewing coded data, discussing coded data in a group setting to develop and reach agreement on salient themes, repeated comparisons among coded data, combining similar or related codes into themes, and extracting themes from within codes [18,21]. After the final group discussion, categories were reviewed and given examples from the transcripts.

Ethical Considerations

This study was performed in accordance with the principles of the Declaration of Helsinki in its revised version from 2000. The Ethics Committee of the Medical Faculty of Leipzig University, Germany, approved the study (587/21-ek). Participating GPs were informed of their right to withdraw from the study at any time without consequences and provided written informed consent before participation. Each participating GP was assigned an identification number and all data were deidentified for analyses. GPs received a reimbursement of €100 (US $103).

Results

Sample Characteristics

Table 1 shows the collected sociodemographic data and information for the GPs in the sample.

XSL•FO
RenderX

Table 1. Sample characteristics (N=9).

Characteristics	Values
Sex, n (%)	
Male	4 (44)
Female	5 (56)
Age (y), mean (SD; range)	46.3 (8.8); 33-57
Duration of practice as resident (y), mean (SD)	10.3 (5.5)
11-25, n (%)	3 (33)
26-50, n (%)	4 (44)
51-75, n (%)	2 (22)
Older patients treated per quarter, n (%)	
51-100	1 (11)
101-500	4 (44)
501-1000	3 (33)
>1000	1 (11)
Estimated percentage of older patients with dementia, mean (SD)	20.11 (11.1)
Form of practice, n (%)	
Individual practice	4 (44)
Joint practice (*Gemeinschaftspraxis*)	3 (33)
Medical center	2 (22)

Themes and Subthemes Identified

We identified 8 themes, each including 2 to 12 subthemes. Table 2 shows the overall frequency of quotes per subcategory.

The theme mentioned most frequently was addressing dementia. On the level of subthemes, requirements for the implementation of eHealth apps were mentioned more frequently than other subthemes. GPs emphasized personal requirements for using an eHealth app for dementia prevention in their care. Themes and subthemes with respective key quotations are described in Multimedia Appendix 2.

XSL•FO
RenderX

Table 2. Frequency of themes (N=8) and subthemes (N=28).

Themes and subthemes	Frequency, n[a]
Addressing dementia	54
Cause: abnormalities	17
Cause: high risk of dementia	1
Cause: routine checkup	3
Cause: affected relatives	3
Cause: initiative of patient or relatives	12
Expressed wishes toward GP[b]	18
Knowledge about dementia	13
High knowledge	2
Low knowledge	11
Need for information	17
For diagnostics, treatment options or progression	11
Psychosocial aspects	4
For risk factors	1
No need for information	1
Potential for risk reduction	35
Most promising risk factors	12
Addressing risk factors	11
Risk factors known by patients	9
Risk factors are disregarded	3
Chances for eHealth apps for dementia risk reduction	21
Advantages for patients	11
Advantages for practitioners	9
Low interest of patients	1
Development of eHealth apps for dementia risk reduction	25
Characteristics qualifying patients for apps for dementia risk reduction	13
Characteristics rendering patients unsuitable for eHealth apps for dementia risk reduction	10
Characteristics of eHealth apps for dementia risk reduction	2
Barriers to eHealth apps for dementia risk reduction	17
General barriers	3
Concerns	10
Limited use	4
Implementation of eHealth apps for dementia risk reduction	33
Prerequisite: accessibility	13
Prerequisites: characteristics of eHealth apps	10
Integration into GP care	10

[a]Frequency of reference to the respective theme or subtheme in the total sample (n=9 GPs); multiple mentions of themes or subthemes per GP possible.
[b]GP: general practitioner.

Addressing Dementia

GPs overwhelmingly addressed dementia in their care only when abnormalities were already present and noticeable by either the GP or their staff. Routine checkups or concerns of relatives played a minor role with only a singular mention of high dementia risk being the cause for addressing dementia in routine care:

XSL•FO
RenderX

> *So we address that routinely when doing check-ups, starting at 70. And in patients under 70 who have personal concerns or an increased risk of dementia from my perspective.* [GP5]

The initiative of patients or relatives was named frequently. GPs mentioned that patients rarely take initiative action regarding dementia risk reduction. Instead, GPs stated that relatives noticed changes in patients and referred them to GPs with a wish for diagnosis:

> *Oh well, when relatives register for consultation hours, they nearly always name somewhat of a reason, if it's acute. And then they'll say, "Dad is odd," or "My husband is odd" And surely also "can't we take a look?" and actively ask whether it's the case [dementia]. So yes, it's talked about quite directly.* [GP2]

Congruently, wishes expressed toward GPs had the most quotations within the theme. Wishes varied and ranged from explicit wishes for fast diagnosis to less specifically expressed emotional needs:

> *They [patients] surely want to hear that they are not forgetful when they are. Or can't remember certain things. And whether there is medication. And get their fears off their chest, which they surely can't articulate like that at home. Because it is a kind of weakness, how they feel about that. So this [GP practice] is a protected environment, where you can slowly move forward and ask whether it's still normal. So they slowly advance, maybe talk about an occurrence.* [GP4]

Knowledge About Dementia

All GPs described patients' knowledge about dementia as low. These included symptoms of dementia, especially in contrast to age-related cognitive decline, mental disorders in general, as well as types and progression of dementia:

> *I think patients don't know much about that [early signs of dementia]. Especially separating normal signs of aging and dementia is tough for a layman, I think. And in my experience, patients that worry a lot about having dementia more commonly suffer from senile depression or an anxiety disorder and not dementia. While the patients that do have dementia often negate that or trivialize that and think it's normal when there's surely an illness and a need for therapy, or at least help, from my perspective.* [GP5]

One GP mentioned that the level of education, personal experience, and degree of personal growth are relevant factors for high knowledge about dementia:

> *[Knowledge of elderly patients about dementia] is dependent on the level of education. It differs. Those who have the ability for differentiated thinking can assess that in a completely different way. And some confuse dementia with something else and simply say "Well, you're surely demented" or something like that to their relative even though he isn't. It's not only related to the level of education but personal experience that they've made, with other patients. And to their own personality.* [GP1]

Need for Information

In the experience of GPs, most patients want to receive information on personal diagnosis, treatment, and course of illness. Patients wish for standardized diagnostic as well as personally relevant assessments made by the GP. These assessments include personalized advice related to psychosocial factors, for example, living situation, relationships with family members, or concerns. Information regarding treatment options and the course of illness were mentioned more frequently than diagnostics. The need for information was only mentioned in regard to early diagnosis of dementia with only a singular mention of risk factors in the context of general health recommendations during check-ups:

> *For many it's about the question: If I had dementia, how would I continue to lead my life? Can I stay independent? Do I need to change my living situation? What happens to my relatives or close ones, commonly spouses? That's what I think concerns patients and is why dementia is omnipresent in our practice.* [GP5]

One GP elaborated that it is common for patients to have received misinformation about the use and limitations of pharmaceutical treatment for dementia by specialized health care providers. It is noteworthy that despite being included as an example in the interview guide and the inductive category system, no GP mentioned a need for information on the causes or risk factors of dementia:

> *If you suffer from dementia and see a psychiatrist, you get one or two different kinds of medication and that's that. There is nothing about lifestyle, nothing about counseling the [social] surroundings, nothing with alternative therapeutic options, what is still possible. You get your pills prescribed and that was that.* [GP6]

Potential for Dementia Risk Reduction

When asked about the most promising factors to be targeted for reducing dementia risk, GPs named social, cognitive, and physical activity. Cardiovascular risk factors were frequently named as part of routine care but are commonly not put into the context of dementia:

> *Measurable factors like blood sugar, blood pressure, and arteriosclerosis, that is definitely part of my job as a GP and an essential part of preventing dementia. The topic of social interactions, social contacts, exercise, I think those are very important. I talk about that, too, but I see my influence there as smaller because more initiative of the patient is required.* [GP5]

GPs stressed their concern over socially isolated older patients and their perceived limited influence on psychosocial factors (eg, social support). All participants stated that they addressed common risk factors in their practice but added that they are

XSL•FO
RenderX

often unsure whether patients follow their recommendations. In the experience of GPs, patients often know about broad components of a healthy lifestyle and cardiovascular risk factors for illnesses, but not for dementia in particular:

> *[I do not mention risk factors] at every contact, but I especially use check-ups and so on to ask: how do you exercise. I always ask about alcohol, and, of course, their daily routine and such things. There, I often talk about what would be good to do more of and what to do less of. But if me addressing it [risk and protective factors for dementia] is enough, that is always hard to say.* [GP8]

One GP mentioned that they addressed lifestyle factors in general but did not give patients any recommendations regarding diet, due to limited potential for improvement:

> *[Regarding diet] I wouldn't say that they [patients] should limit themselves or do less of anything. Disregarding alcohol, but nothing else. So, I wouldn't say they need to up their vitamins, because of Vitamin C... I don't see relevant potential there.* [GP9]

Chances for eHealth Apps for Dementia Risk Reduction

GPs perceived eHealth apps for dementia risk reduction mostly as a chance to standardize patient education and integrate social factors into care. Strengthening self-efficacy and motivating patients was another point mentioned by participating GPs:

> *It really is mostly about relatives who would be very interested in helping their spouse, for example. Or help their parents. And then, of course, if a patient isn't able to educate themselves about this application or read everything, with the help of relatives it would be very appreciated. I think that makes sense.* [GP3]

One GP mentioned the ability to self-administer screening tests before appointments as a chance to improve care:

> *[I would use it] just how I recommend a good depression scale if I know one, check yourself for depression. There it says a bit more, for example, moderate depression, and then I would offer to check with me again.* [GP9]

Development of eHealth Apps for Dementia Risk Reduction

GPs characterized suitable patients for eHealth apps for prevention as high in health literacy and self-efficacy. Especially younger patients and patients with high media literacy were named as suitable for eHealth apps for dementia risk reduction. Accordingly, patients who would be unsuitable for such measures were characterized as older individuals or those who lived alone without help:

> *But the older people who are alone, I don't think they could manage to do that. If I imagine two 85 year old seniors here, no. And one of those has dementia... They rather have questions like, how do I turn on the phone? And then getting to an application, that is really hard.* [GP6]

Barriers to eHealth Apps for Dementia Risk Reduction

GPs mentioned they were worried that early preventive measures for at-risk patients could increase anxiety and worries about dementia in patients. They also argued that patients who would be likely to use these measures are both already overly anxious and therefore already active. This perception would limit the effectiveness and likeliness of use for GPs severely.

One GP elaborated that it would be important for them to use eHealth only for supporting already existing care and explicitly communicate barriers of eHealth and limitations of prevention for dementia. As an example, this GP listed individualized recommendations for lifestyle changes from an eHealth app.

Implementation of eHealth apps for Dementia Risk Reduction

To implement eHealth apps for dementia risk reduction into their care, GPs mentioned that they would need to have an overview of the information provided by the app to administer its use. Furthermore, GPs emphasized a high degree of accessibility in terms of ease of use and low complexity. Reliable privacy policies and free use were mentioned as key factors for recommending such measures to patients. Most GPs advocated supervised use of preventive measures by qualified personnel (eg, GPs, psychologists, and nurses):

> *I imagine a certain type of older person at their computer, entering their data and the application says: "you have moderately severe dementia." And I don't want to imagine that patient alone with their computer.* [GP5]

Further ideas of GPs were integrating recommendations for eHealth apps for prevention into routine check-ups or cooperating with nursing services that can assess the need for lifestyle changes. Finally, one GP mentioned a lack of information regarding existing preventive measures for GPs and wished for vocational training, published guidance, and personalized letters or emails.

Discussion

Principal Findings

To the best of our knowledge, our study is the first to use a qualitative approach for examining GPs' views on eHealth apps for dementia risk reduction, interviewing 9 German GPs working in an outpatient setting. Overall, GPs named more positive than negative aspects of eHealth apps for dementia risk reduction and displayed a balanced but positive attitude toward them. It is noteworthy that participants frequently steered the conversation toward early dementia diagnosis or dementia care management, although the questions focused on dementia prevention. This was also reflected in the coding system, as GPs frequently noted that they only addressed dementia when abnormalities were present and that patients' need for information concerns only diagnostic and treatment options. Existing research has established that GPs are reluctant to perform early dementia diagnosis due to uncertainty, which seems to extend to dementia prevention [22]. Our findings may aid in the development and implementation of eHealth apps for

XSL•FO
RenderX

dementia prevention and to inform discussions about acceptance and attitudes of eHealth apps for dementia risk reduction with primary care givers in Germany.

According to the typology of eHealth implementation theories developed by Heinsch et al [23], respective theories can fall into the 5 categories of centering agency, structure, relations, meaning, and norms. Agency-centered theories focus on an individual's actions, beliefs, and attitudes as opposed to other theories that center, for example, on organizational environments or interactional processes. The topics covered by our participants are highly compatible with agency-centered theories, as there was no mention of structural, organizational, or other factors. Particularly, the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT) were reflected in our coding system. TAM proposes that technology acceptance and use are affected by perceived ease of use, perceived usefulness, and subjective norms, all of which were most frequently mentioned by our participants. UTAUT proposes that the intention to use eHealth apps is affected by effort expectancy, performance expectancy, social influence, facilitating conditions, and habit [23]. Effort expectancy and performance expectancy were factors GPs talked about in-depth, while social, environmental, and habitual factors only played a minor role, for example, for practical implementation purposes, such as delegating the use of eHealth app to staff. Addressing dementia due to high risk or during routine checkups was only mentioned 4 times. Therefore, many GPs seem to show little to no regard for primary prevention and dementia risk reduction, which is in line with recent findings from the Netherlands [6] and the United Kingdom [8], where GPs stated that they did not address dementia prevention or risk reduction with their patients in routine care.

According to the interviewed GPs, patients want to receive individualized information and apps they can work at their tempo. The social environment and relatives should be included in the use of the app. In addition, apps should be designed sensitively so as to not cause or reinforce worries regarding dementia and give realistic goals and expectations.

There seems to be a need for information on the part of patients, as GPs predominantly perceived their patients' knowledge about dementia as low. This indicates potential for increasing knowledge, for example, via eHealth apps for dementia risk reduction. Patients seem to have no interest in information on prevention of dementia, and risk factors can be attributed either to no interest in these aspects or, in our eyes more likely, to no knowledge about modifiable dementia risk factors [14,24]. When patients are unaware of a link between lifestyle and dementia risk, they have no reason to seek out their GP for information.

In an earlier study, older patients expressed the wish for eHealth services to be integrated into usual GP treatment and had favorable opinions toward eHealth services when they supplemented care [25]. Congruently, GPs were generally also in favor of individualized risk assessment and lifestyle recommendations. In a recent study [26], only a small number of interviewed GPs saw self-monitoring as a useful area for implementation of eHealth apps for prevention measures.

In accordance with these findings, the use of eHealth apps for dementia risk reduction was also only addressed by one GP in our study, while the use for psychosocial care was generally favored. In existing research on the perception of internet-based and digital health apps in general, physicians see a variety of advantages and potential uses, particularly for preventive measures and lifestyle modifications [26-29]. A meta-analysis reported small to medium effect sizes for eHealth interventions for brain health on subjective and cognitive performance as well as risk profiles for dementia [13]. More recent studies showed beneficial intervention effects on risk factors for dementia in middle aged and older adults [30,31]. Communicating respective findings regarding eHealth interventions for dementia risk reduction to GPs might increase practitioners' interest and convince them of their potential effectiveness. Against this background, raising awareness for the potential of dementia risk reduction and the role that eHealth interventions might play in this regard among GPs seems necessary to maximize the potential of risk reduction strategies.

Tying in with earlier studies [26], GPs stressed free access and ease of use. In addition, GPs felt unsure navigating the variety of digital health apps and wished for scientifically evaluated effectiveness to find appropriate apps. Despite the currently explorative stage of research, our results fit well into existing studies and allow us to draw narrow conclusions about the preventive use of internet-based or digital health apps for dementia risk reduction. A previous review identified wrongful data use, cost, and lack of quality of care as topics that concern patients the most regarding digital health apps and this fear seems to be warranted [32]. Therefore, potential eHealth apps for dementia risk reduction should prioritize no costs for patients in their financial planning if they want to ensure maximum use by GPs and patients alike. Similar to the studies that attribute the low use of digital health apps in practice to concerns regarding safety, interviewed GPs in our study also frequently mentioned security concerns and privacy issues. The most frequently named concerns in our study (limited use, low effects, and inducing anxiety and worries) can be explained both within the theoretical frameworks of TAM and UTAUT, as they seem to reflect the factors of perceived usefulness (TAM) and expected performance (UTAUT). A recent study with a similar design to our study from the United Kingdom also found that the fear of inducing health anxiety seems to be common in GPs regarding dementia prevention [8].

Another perceived barrier not mentioned in previous studies was that eHealth apps for prevention measures would only be used by patients with (1) high anxieties and worries regarding dementia, (2) high self-efficacy, or (3) high health literacy, thereby limiting potential. Generally, existing research suggests that digital health interventions improve health literacy [33,34]. However, the use of digital health apps is affected by structural inequalities, such as low socioeconomic status, ethnicity, and education status, which contribute to disparities in health literacy as well as use of additional health education programs [35]. In contrast, eHealth interventions can use various channels to provide information, for example, audiovisual or interactive formats, which may facilitate the inclusion of older adults with lower levels of education or health literacy [36]. This view was

XSL•FO
RenderX

also expressed by one of the GPs in our study (GP8). Previous studies indicate that the effects of digital health app use on health behaviors could be mediated by self-efficacy [37], which is a different effect than the one frequently proposed by our participating GPs (eHealth apps increase knowledge for already motivated patients) and could be communicated to GPs interested in the mechanisms of future internet-based preventive measures. An earlier study focusing on older adults in a low socioeconomic community found increased computer health literacy and self-efficacy because of a targeted computer intervention [38]. The GPs in our interviews also focused on increased self-efficacy and participation of patients and their social environments as one of the most promising factors for the use of eHealth apps for prevention measures. The barriers discussed by our participants and the concerns over implementation into routine care seem to reflect a fear of a lower quality of care or a loss of trust between the physician and the patient when relying too heavily on digital apps. The review by Ramachandran et al [39] on the impact of eHealth on relationships and trust in primary care shows that eHealth apps can even increase trust when patients perceive that they are receiving personalized and collaborative care. This indicates that GPs could be more in favor of eHealth apps for dementia prevention when used in a hybrid care model, where GPs use eHealth prevention tools when assessing high dementia risk in a patient. Our results allow a potential investigation of the role of eHealth apps that GPs would prefer, although the integration into routine care is a longstanding debate with many more factors to consider [40-42]. It would be interesting to see future evaluations of eHealth apps for prevention measures include constructs, such as self-efficacy, perceived usefulness, physician-patient trust, and measurement of possible covariates that might influence the effect of eHealth apps for prevention measures. Research based on social cognitive theory has consistently found that self-efficacy facilitates a change and uptake of change in health behaviors [43] and is, therefore, a relevant outcome for evaluative studies that test digital health apps.

GPs mentioned both a lack of overview of existing apps and their respective content and were unsure about how to integrate them into their clinical practice. Therefore, eHealth apps should be designed in a way that frequently refers to the GP in a primary care setting, for example, via having tips that direct patients to their GP. Furthermore, referral to eHealth apps by trusted sources like GPs has been suggested as a promising approach to increase older adults' interest in eHealth tools [35]. This could serve to strengthen the role of the GP in providing care and reduce unrealistic expectations of patients regarding the possibilities of the app, reducing 2 important barriers at once. To maximize cooperation with GPs, internet-based prevention measures should serve a clear and identifiable role in the care of patients, for example, as part of comprehensive psychosocial care provided by the GP or part of self-monitoring in-between appointments. Developers of eHealth interventions aimed at brain health and dementia risk reduction might benefit from cooperation with GPs early on in the process of designing interventions. Furthermore, offering concise information for GPs on the device, that is, contents, target group, administration in GP practice, and tutorials for GP practice could further enhance acceptance of eHealth tools among GPs and strengthen cooperation between researchers, developers, and practitioners toward the goal of dementia risk reduction [4]. A possible application for eHealth interventions for dementia risk reduction and other digital health apps should further focus on supplementing in-person health care to increase health literacy and mend care in areas with weak health infrastructure or particularly hard to reach patients [34,38,45,46]. Textbox 1 summarizes recommendations for the development and implementation of eHealth tools for dementia risk reduction, as expressed by participating GPs.

Textbox 1. Recommendations for design and implementation of eHealth tools for dementia risk reduction.

- Ease of use, low complexity: relevant information is readily traceable within the eHealth tool, intuitive navigation
- Individualized recommendations: provide guidance and recommendations based on individual dementia risk profiles and stage of motivation for behavior change
- Sincerity about the potential for dementia prevention: avoid overemphasizing individual means of risk reduction (eg, "risk reduction" instead of "prevention")
- Inclusion of partners and relatives: provide means for partners or other relatives to interact with the eHealth tool, for example, to coach patients in using the app
- Privacy and data protection: ensure privacy of user data, prohibit wrongful transfer of participant data to third parties
- Accessibility: provide the eHealth tool free of cost, easily findable in app stores
- Inclusion of general practitioners (GPs) and other health professionals: provide structured information on available eHealth tools for dementia risk reduction, including evidence of effectiveness, for GPs. Train health professionals (GPs, medical assistants, nurses, psychologists, etc) in the use of the app to facilitate supervision of patients' use of eHealth tools

Discussion of Methods

Qualitative research does not allow direct comparability via quantified scores and readily defined constructs. However, this tradeoff allows a more accurate depiction of ambivalences, subtleties, and possibly contradicting opinions that would be lost if focusing on rigorously established scales. Therefore, we decided on a structured, but explorative design to gain as broad of an insight as possible. Using the qualitative research approach of Mayring [18] in formulating the interview guide and data analysis with a very high degree of structure, validity, reliability, and generalizability, we hope to provide robust information. This procedure seems particularly appropriate, given the novelty of eHealth apps for dementia risk reduction.

XSL•FO
RenderX

Strengths and Limitations

Our study assessed a timely topic for means of dementia risk reduction, that is, the use of eHealth apps. GPs can play a major role in providing recommendations for lifestyle changes to reduce dementia risk. Therefore, assessing GPs' attitudes toward eHealth for dementia risk reduction may likely provide useful evidence for both the design and implementation of apps, as well as uptake and acceptance by patients. Still, certain limitations need to be considered when interpreting our findings. Although the interviews focused on primary prevention, that is, dementia risk reduction, several GPs referred to the use of eHealth for secondary prevention in our study. Furthermore, heterogeneity in participating GPs' age may have influenced our findings. It might be possible that younger participants may generally be more familiar with eHealth tools, while GPs in their 50s might be less accustomed to apps, possibly rendering them more skeptical toward eHealth for dementia risk reduction. To ensure the timely execution of interviews and due to time constraints of the project, GPs were recruited from established networks of GPs, raising the risk of selection bias and resulting in a rather homogeneous sample. Because we aimed to keep effort for GPs to a minimum, we only assessed limited sociodemographics and further information about GPs. A larger, more diverse sample recruited across various channels may improve the generalizability of findings in future studies. However, this study aimed to provide a first insight into German GPs' attitudes toward eHealth for a fairly new topic, that is, dementia risk reduction. We are confident that our findings provide valuable insights that can inform larger-scale surveys on the topic as well as the design of eHealth interventions aimed at reducing dementia risk.

Implications and Recommendations for Future Research

Future eHealth interventions for brain health could be developed and implemented as part of a blended care approach, similar to strategies already used in psychiatry and primary care [47-49]. Considering the intended form of use (eg, directly supervised, supported use in between appointments) before developing the app and its features would allow increased performance expectations and better evaluation. As lifestyle interventions aim to induce behavioral change, future studies for developing and evaluating eHealth apps for brain health should include measures that can assess the constructs that are intended to induce this behavioral change alongside relevant outcome measures. Along with general efficacy and effectiveness evaluations, this could allow a better understanding of not just if eHealth apps for prevention measures work but how they work, as few available interventions include comparable constructs (eg, changing health beliefs, increasing self-efficacy, and introducing or improving existing health behavior). A proof-of-concept study evaluating the Dutch MijnBreincoach app indicated no change in beliefs and attitudes despite positive evaluations [50], whereas in another study, the same app improved adherence to the Mediterranean diet, BMI, and knowledge of dementia risk reduction [30]. We recommend that future researchers rely on unified theory-based questionnaires for health beliefs, such as the Motivation to Change Lifestyle and Health Behaviors for Dementia Risk Reduction scale [51].

We further propose that researchers and developers consider the needs and concerns of GPs named in this study and similar studies to increase acceptance and adoption by practitioners. Similarly, the factors influencing user acceptance should be assessed to make future interventions valuable to practitioners and patients alike. The development of future interventions should prioritize tailoring and personalization of eHealth apps for dementia risk reduction. Features, recommendations, and information given should be relevant to the individual patient and provide a positive approach toward change of lifestyle factors, without raising or increasing fear of dementia in users. Finally, future research should pay attention to general health inequalities and access to eHealth apps for dementia risk reduction [34,35,46]. Understanding existing disparities in access, use, and outcomes among populations with different socioeconomic backgrounds can allow equitable distribution and ensure that interventions reach those who need them.

Acknowledgments

This work was funded by a junior research grant by the Medical Faculty of Leipzig University, granted to AZ. The funder had no role in the design and conduct of the study, collection, analyses, and interpretation of data, drafting of the manuscript, or the decision to submit the results. The authors wish to thank all participating general practitioners for their contribution to this study. They acknowledge support from Leipzig University within the program of open-access publishing.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

AZ, ML, and SGR-H participated in the conceptualization of the study. AS and AZ participated in developing the methodology. AS and AZ were responsible for validation. AS was responsible for formal analyses and investigation. SGR-H was responsible for procuring resources. AS and AZ participated in data curation. AS was responsible for writing the original draft. AS, ML, MB, SGR-H, and AZ were responsible for writing, reviewing, and editing; AS participated in visualization. ML, SGR-H, and AZ provided supervision. AZ, ML, SGR-H participated in project administration and AZ contributed to funding acquisition. No generative AI tools were used to prepare this manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File , 21 KB - formative_v9i1e56310_app1.docx]

Multimedia Appendix 2
Themes, subthemes, and quotations.
[DOCX File , 25 KB - formative_v9i1e56310_app2.docx]

References

1. Doblhammer G, Fritze T, Reinke C, Fink A. Can dementia become the most prevalent disease at the time of death in Germany? Projections up to the year 2060 for the five most important diseases at the time of death. J Popul Ageing 2022 May 16;15(2):523-540. [doi: 10.1007/s12062-022-09365-7]
2. Neubert L, König HH, Löbner M, Luppa M, Pentzek M, Fuchs A, et al. Excess costs of dementia in old age (85+) in Germany: results from the AgeCoDe-AgeQualiDe study. J Econ Ageing 2021 Oct;20:100346. [doi: 10.1016/j.jeoa.2021.100346]
3. Livingston G, Huntley J, Liu KY, Costafreda SG, Selbæk G, Alladi S, et al. Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. Lancet 2024 Aug 10;404(10452):572-628. [doi: 10.1016/S0140-6736(24)01296-0] [Medline: 39096926]
4. Blotenberg I, Hoffmann W, Thyrian JR. Dementia in Germany: epidemiology and prevention potential. Dtsch Arztebl Int 2023 Jul 10;120(27-28):470-476 [FREE Full text] [doi: 10.3238/arztebl.m2023.0100] [Medline: 37226316]
5. Solomon A, Stephen R, Altomare D, Carrera E, Frisoni GB, Kulmala J, et al. Multidomain interventions: state-of-the-art and future directions for protocols to implement precision dementia risk reduction. A user manual for Brain Health Services-part 4 of 6. Alzheimers Res Ther 2021 Oct 11;13(1):171 [FREE Full text] [doi: 10.1186/s13195-021-00875-8] [Medline: 34635167]
6. Wangler J, Jansky M. Factors influencing general practitioners' perception of and attitude towards dementia diagnostics and care-results of a survey among primary care physicians in Germany. Wien Med Wochenschr 2021 May;171(7-8):165-173 [FREE Full text] [doi: 10.1007/s10354-020-00803-9] [Medline: 33443611]
7. Godbee K, Gunn J, Lautenschlager NT, Curran E, Palmer VJ. Implementing dementia risk reduction in primary care: a preliminary conceptual model based on a scoping review of practitioners' views. Prim Health Care Res Dev 2019 Oct 23;20:e140 [FREE Full text] [doi: 10.1017/S1463423619000744] [Medline: 31640836]
8. Jones D, Drewery R, Windle K, Humphrey S, de Paiva AF. Dementia prevention and the GP's role: a qualitative interview study. Br J Gen Pract 2024 Apr;74(741):e242-e249 [FREE Full text] [doi: 10.3399/BJGP.2023.0103] [Medline: 37549993]
9. Müller A. D21-digital-index 2017/2018. Merz Medien Erziehung 2018 Apr 02;62(2):3 [FREE Full text] [doi: 10.21240/merz/2018.2.3]
10. Merkel S, Hess M. The use of internet-based health and care services by elderly people in Europe and the importance of the country context: multilevel study. JMIR Aging 2020 Jun 03;3(1):e15491 [FREE Full text] [doi: 10.2196/15491] [Medline: 32490837]
11. Rasche P, Wille M, Bröhl C, Theis S, Schäfer K, Knobe M, et al. Prevalence of health app use among older adults in Germany: national survey. JMIR Mhealth Uhealth 2018 Jan 23;6(1):e26 [FREE Full text] [doi: 10.2196/mhealth.8619] [Medline: 29362211]
12. Pabst A, Löbner M, Stein J, Luppa M, Kersting A, König HH, et al. Internet-based cognitive behavior therapy only for the young? A secondary analysis of a randomized controlled trial of depression treatment. Front Psychiatry 2020 Jul 24;11:735 [FREE Full text] [doi: 10.3389/fpsyt.2020.00735] [Medline: 32848915]
13. Wesselman LM, Hooghiemstra AM, Schoonmade LJ, de Wit MC, van der Flier WM, Sikkes SA. Web-based multidomain lifestyle programs for brain health: comprehensive overview and meta-analysis. JMIR Ment Health 2019 Apr 09;6(4):e12104 [FREE Full text] [doi: 10.2196/12104] [Medline: 30964438]
14. Heger I, Deckers K, van Boxtel M, de Vugt M, Hajema K, Verhey F, et al. Dementia awareness and risk perception in middle-aged and older individuals: baseline results of the MijnBreincoach survey on the association between lifestyle and brain health. BMC Public Health 2019 Jun 03;19(1):678 [FREE Full text] [doi: 10.1186/s12889-019-7010-z] [Medline: 31159779]
15. Richard E, Moll van Charante EP, Hoevenaar-Blom MP, Coley N, Barbera M, van der Groep A, et al. Healthy ageing through internet counselling in the elderly (HATICE): a multinational, randomised controlled trial. Lancet Digit Health 2019 Dec;1(8):e424-e434 [FREE Full text] [doi: 10.1016/S2589-7500(19)30153-0] [Medline: 33323224]

XSL•FO
RenderX

16. Mykhalovskiy E, Eakin J, Beagan B, Beausoleil N, Gibson BE, Macdonald ME, et al. Beyond bare bones: critical, theoretically engaged qualitative research in public health. Can J Public Health 2018 Dec;109(5-6):613-621 [FREE Full text] [doi: 10.17269/s41997-018-0154-2] [Medline: 30465286]
17. Greenhalgh T, Annandale E, Ashcroft R, Barlow J, Black N, Bleakley A, et al. An open letter to The BMJ editors on qualitative research. BMJ 2016 Feb 10;352:i563 [FREE Full text] [doi: 10.1136/bmj.i563] [Medline: 26865572]
18. Mayring P. Qualitative content analysis: demarcation, varieties, developments. Forum Qual Soc Res 2019;20(3). [doi: 10.17169/fqs-20.3.3343]
19. Schladitz K, Förster F, Löbner M, Welzel F, Stein J, Luppa M, et al. [Grief and loss in elderly people: a qualitative study regarding the user acceptance of an internet-based self-help program from user and expert perspective]. Z Evid Fortbild Qual Gesundhwes 2020 Apr;150-152:112-123. [doi: 10.1016/j.zefq.2020.01.007] [Medline: 32460967]
20. Baker SE, Edwards R. How many qualitative interviews is enough? Expert voices and early career reflections on sampling and cases in qualitative research. National Centre for Research Methods. URL: https://eprints.ncrm.ac.uk/id/eprint/2273/4/how_many_interviews.pdf [accessed 2024-09-26]
21. Raskind IG, Shelton RC, Comeau DL, Cooper HL, Griffith DM, Kegler MC. A review of qualitative data analysis practices in health education and health behavior research. Health Educ Behav 2019 Feb 18;46(1):32-39 [FREE Full text] [doi: 10.1177/1090198118795019] [Medline: 30227078]
22. Wangler J, Jansky M. Factors influencing general practitioners' perception of and attitude towards dementia diagnostics and care—results of a survey among primary care physicians in Germany. Wien Med Wochenschr 2021 Jan 14;171(7-8):165-173. [doi: 10.1007/s10354-020-00803-9]
23. Heinsch M, Wyllie J, Carlson J, Wells H, Tickner C, Kay-Lambkin F. Theories informing eHealth implementation: systematic review and typology classification. J Med Internet Res 2021 May 31;23(5):e18500 [FREE Full text] [doi: 10.2196/18500] [Medline: 34057427]
24. Zülke AE, Luppa M, Köhler S, Riedel-Heller SG. Knowledge of risk and protective factors for dementia in older German adults: a population-based survey on risk and protective factors for dementia and internet-based brain health interventions. PLoS One 2022 Nov 7;17(11):e0277037 [FREE Full text] [doi: 10.1371/journal.pone.0277037] [Medline: 36342935]
25. Nymberg VM, Bolmsjö BB, Wolff M, Calling S, Gerward S, Sandberg M. 'Having to learn this so late in our lives…' Swedish elderly patients' beliefs, experiences, attitudes and expectations of e-health in primary health care. Scand J Prim Health Care 2019 Mar;37(1):41-52 [FREE Full text] [doi: 10.1080/02813432.2019.1570612] [Medline: 30732519]
26. Wangler J, Jansky M. Gesundheits-apps als instrumente der prävention? – Eine interviewstudie zu potenzialen für das hausärztliche setting. Präv Gesundheitsf 2020 Mar 31;15(4):340-346. [doi: 10.1007/s11553-020-00769-x]
27. Bittner J. So denken Ärzte über digital health: eine synopse der aktuellen umfragen. Bertelsmann Stiftung. 2017 Aug 21. URL: https://blog.der-digitale-patient.de/synopse-aerztebefragungen-digital-health/ [accessed 2023-04-21]
28. Realitäts-check – nutzen Ärzte health apps? Coliquio. 2017 Jul 31. URL: https://www.coliquio-insights.de/nutzen-aerzte-health-apps-exklusive-umfrage/ [accessed 2023-04-21]
29. Rohlender B, Reinhardt K. Gesundheit 4.0 – wie Ärzte die digitale Zukunft sehen. Hartmannbund. 2017 Jun 8. URL: http://www.hartmannbund.de/fileadmin/user_upload/Downloads/Umfragen/2017_HB-Bitkom_Start-ups.pdf [accessed 2023-04-21]
30. Van Asbroeck S, Köhler S, Wimmers SC, Muris JW, van Boxtel MP, Deckers K. Multidomain dementia risk reduction in primary care is feasible: a proof-of-concept study. J Alzheimers Dis 2024;99(4):1455-1471 [FREE Full text] [doi: 10.3233/JAD-240229] [Medline: 38759017]
31. McMaster M, Kim S, Clare L, Torres SJ, Cherbuin N, D Este C, et al. Lifestyle risk factors and cognitive outcomes from the multidomain dementia risk reduction randomized controlled trial, body brain life for cognitive decline (BBL-CD). J Am Geriatr Soc 2020 Nov;68(11):2629-2637. [doi: 10.1111/jgs.16762] [Medline: 32909259]
32. Crossnohere NL, Weiss B, Hyman S, Bridges JF. Patient preferences for health information technologies: a systematic review. J Hosp Manag Health Policy 2021 Sep;5:25 [FREE Full text] [doi: 10.21037/jhmhp-20-105]
33. Jacobs RJ, Lou JQ, Ownby RL, Caballero J. A systematic review of eHealth interventions to improve health literacy. Health Informatics J 2016 Jun;22(2):81-98 [FREE Full text] [doi: 10.1177/1460458214534092] [Medline: 24916567]
34. Bodie GD, Dutta MJ. Understanding health literacy for strategic health marketing: eHealth literacy, health disparities, and the digital divide. Health Mark Q 2008;25(1-2):175-203. [doi: 10.1080/07359680802126301] [Medline: 18935884]
35. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021 Aug 17;21(1):1556 [FREE Full text] [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
36. Meppelink CS, van Weert JC, Haven CJ, Smit EG. The effectiveness of health animations in audiences with different health literacy levels: an experimental study. J Med Internet Res 2015 Jan 13;17(1):e11 [FREE Full text] [doi: 10.2196/jmir.3979] [Medline: 25586711]
37. Choi M. Association of eHealth use, literacy, informational social support, and health-promoting behaviors: mediation of health self-efficacy. Int J Environ Res Public Health 2020 Oct 28;17(21):7890 [FREE Full text] [doi: 10.3390/ijerph17217890] [Medline: 33126469]

XSL•FO
RenderX

38. Chu A, Huber J, Mastel-Smith B, Cesario S. "Partnering with seniors for better health": computer use and internet health information retrieval among older adults in a low socioeconomic community. J Med Libr Assoc 2009 Jan;97(1):12-20 [FREE Full text] [doi: 10.3163/1536-5050.97.1.003] [Medline: 19159002]
39. Ramachandran M, Brinton C, Wiljer D, Upshur R, Gray CS. The impact of eHealth on relationships and trust in primary care: a review of reviews. BMC Prim Care 2023 Nov 03;24(1):228 [FREE Full text] [doi: 10.1186/s12875-023-02176-5] [Medline: 37919688]
40. Carey M, Noble N, Mansfield E, Waller A, Henskens F, Sanson-Fisher R. The role of eHealth in optimizing preventive care in the primary care setting. J Med Internet Res 2015 May 22;17(5):e126 [FREE Full text] [doi: 10.2196/jmir.3817] [Medline: 26001983]
41. Stevens WJ, van der Sande R, Beijer LJ, Gerritsen MG, Assendelft WJ. eHealth apps replacing or complementing health care contacts: scoping review on adverse effects. J Med Internet Res 2019 Mar 01;21(3):e10736 [FREE Full text] [doi: 10.2196/10736] [Medline: 30821690]
42. Tossaint-Schoenmakers R, Versluis A, Chavannes N, Talboom-Kamp E, Kasteleyn M. The challenge of integrating eHealth into health care: systematic literature review of the Donabedian model of structure, process, and outcome. J Med Internet Res 2021 May 10;23(5):e27180 [FREE Full text] [doi: 10.2196/27180] [Medline: 33970123]
43. Warner LM, Schwarzer R. Self-efficacy and health. In: Gulliver SB, Cohen LM, editors. The Wiley Encyclopedia of Health Psychology. Hoboken, NJ: John Wiley & Sons; 2020.
44. Steyaert J, Deckers K, Smits C, Fox C, Thyrian R, Jeon YH, et al. Putting primary prevention of dementia on everybody's agenda. Aging Ment Health 2021 Aug;25(8):1376-1380. [doi: 10.1080/13607863.2020.1783514] [Medline: 32590910]
45. Hantgan SL, Jariwala SP. User-centered mobile health applications for asthma. Ann Allergy Asthma Immunol 2023 Feb;130(2):159-160 [FREE Full text] [doi: 10.1016/j.anai.2022.11.011] [Medline: 36400353]
46. Beacom AM, Newman SJ. Communicating health information to disadvantaged populations. Fam Community Health 2010;33(2):152-162. [doi: 10.1097/FCH.0b013e3181d59344] [Medline: 20216358]
47. Grätzel P. Blended treatment: theoretisch gut, in der umsetzung mühsam. InFo Neurologie Psychiatrie 2020 Mar 26;22(3):57-58. [doi: 10.1007/s15005-020-1314-5]
48. Maehder K, Löwe B, Härter M, Heddaeus D, Scherer M, Weigel A. Management of comorbid mental and somatic disorders in stepped care approaches in primary care: a systematic review. Fam Pract 2019 Jan 25;36(1):38-52. [doi: 10.1093/fampra/cmy122] [Medline: 30535053]
49. Löbner M, Pabst A, Stein J, Dorow M, Matschinger H, Luppa M, et al. Computerized cognitive behavior therapy for patients with mild to moderately severe depression in primary care: a pragmatic cluster randomized controlled trial (@ktiv). J Affect Disord 2018 Oct 01;238:317-326. [doi: 10.1016/j.jad.2018.06.008] [Medline: 29902736]
50. Heger I, Deckers K, de Vugt M, Verhey F, Oenema A, van Boxtel M, et al. Using mHealth for primary prevention of dementia: a proof-of-concept study on usage patterns, appreciation, and beliefs and attitudes regarding prevention. J Alzheimers Dis 2023;94(3):935-948 [FREE Full text] [doi: 10.3233/JAD-230225] [Medline: 37355903]
51. Kim S, Sargent-Cox K, Cherbuin N, Anstey KJ. Development of the motivation to change lifestyle and health behaviours for dementia risk reduction scale. Dement Geriatr Cogn Dis Extra 2014 May;4(2):172-183 [FREE Full text] [doi: 10.1159/000362228] [Medline: 25028583]

Abbreviations

GP: general practitioner
TAM: Technology Acceptance Model
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by A Mavragani; submitted 12.01.24; peer-reviewed by A Paiva, K Bond; comments to author 19.09.24; revised version received 24.10.24; accepted 14.11.24; published 22.01.25.

Please cite as:
Schultz A, Luppa M, Bleckwenn M, Riedel-Heller SG, Zuelke A
Attitudes of German General Practitioners Toward eHealth Apps for Dementia Risk Reduction: Qualitative Interview Study
JMIR Form Res 2025;9:e56310
URL: https://formative.jmir.org/2025/1/e56310
doi:10.2196/56310
PMID:

©Adrian Schultz, Melanie Luppa, Markus Bleckwenn, Steffi G Riedel-Heller, Andrea Zuelke. Originally published in JMIR Formative Research (https://formative.jmir.org), 22.01.2025. This is an open-access article distributed under the terms of the

XSL•FO
RenderX

Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Discrimination of Radiologists' Experience Level Using Eye-Tracking Technology and Machine Learning: Case Study

Stanford Martinez[1], PhD; Carolina Ramirez-Tamayo[1], MSc; Syed Hasib Akhter Faruqui[2], PhD; Kal Clark[3], MD; Adel Alaeddini[4], PhD; Nicholas Czarnek, PhD; Aarushi Aggarwal[3], MD; Sahra Emamzadeh[3], MD, MPH; Jeffrey R Mock[5], PhD; Edward J Golob[5], PhD

[1]Department of Mechanical Engineering, The University of Texas at San Antonio, San Antonio, TX, United States

[2]Department of Engineering Technology, Sam Houston State University, Huntsville, TX, United States

[3]Department of Radiology, University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

[4]Department of Mechanical Engineering, Southern Methodist University, Dallas, TX, United States

[5]Department of Psychology, The University of Texas at San Antonio, San Antonio, TX, United States

Corresponding Author:
Adel Alaeddini, PhD
Department of Mechanical Engineering
Southern Methodist University
3101 Dyer Street
Dallas, TX, 75205
United States
Phone: 1 214 768 3050
Email: aalaeddini@smu.edu

Abstract

Background: Perception-related errors comprise most diagnostic mistakes in radiology. To mitigate this problem, radiologists use personalized and high-dimensional visual search strategies, otherwise known as search patterns. Qualitative descriptions of these search patterns, which involve the physician verbalizing or annotating the order he or she analyzes the image, can be unreliable due to discrepancies in what is reported versus the actual visual patterns. This discrepancy can interfere with quality improvement interventions and negatively impact patient care.

Objective: The objective of this study is to provide an alternative method for distinguishing between radiologists by means of captured eye-tracking data such that the raw gaze (or processed fixation data) can be used to discriminate users based on subconscious behavior in visual inspection.

Methods: We present a novel discretized feature encoding based on spatiotemporal binning of fixation data for efficient geometric alignment and temporal ordering of eye movement when reading chest x-rays. The encoded features of the eye-fixation data are used by machine learning classifiers to discriminate between faculty and trainee radiologists. A clinical trial case study was conducted using metrics such as the area under the curve, accuracy, F_1-score, sensitivity, and specificity to evaluate the discriminability between the 2 groups regarding their level of experience. The classification performance was then compared with state-of-the-art methodologies. In addition, a repeatability experiment using a separate dataset, experimental protocol, and eye tracker was performed with 8 participants to evaluate the robustness of the proposed approach.

Results: The numerical results from both experiments demonstrate that classifiers using the proposed feature encoding methods outperform the current state-of-the-art in differentiating between radiologists in terms of experience level. An average performance gain of 6.9% is observed compared with traditional features while classifying experience levels of radiologists. This gain in accuracy is also substantial across different eye tracker–collected datasets, with improvements of 6.41% using the Tobii eye tracker and 7.29% using the EyeLink eye tracker. These results signify the potential impact of the proposed method for identifying radiologists' level of expertise and those who would benefit from additional training.

Conclusions: The effectiveness of the proposed spatiotemporal discretization approach, validated across diverse datasets and various classification metrics, underscores its potential for objective evaluation, informing targeted interventions and training strategies in radiology. This research advances reliable assessment tools, addressing challenges in perception-related errors to enhance patient care outcomes.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e53928) doi:10.2196/53928

KEYWORDS

machine learning; eye-tracking; experience level determination; radiology education; search pattern feature extraction; search pattern; radiology; classification; gaze; fixation; education; experience; spatio-temporal; image; x-ray; eye movement

Introduction

Lung cancer is the leading cause of cancer death, claiming 139,000 American lives yearly [1]. To mitigate its impact, the US Preventative Task Force recommends annual radiological screening for at-risk individuals [2]. Radiologists identify suspicious lung lesions (nodules) from patient chest images and recommend further management, including biopsy, continued surveillance, or further workup. Radiological surveillance reduces population mortality from lung cancer, but it is estimated that radiologists will make errors on 33% of abnormal chest exams, eliminating the chance for patients to start lifesaving treatment [3]. The predominant source of these errors is not deficient medical knowledge. Rather, errors primarily stem from the methods radiologists use to visually inspect the image, referred to as perceptual errors [4]. In other words, perceptual errors in radiology are mistakes that occur during the visual inspection and interpretation of medical images. They are distinct from cognitive errors, which involve incorrect reasoning or decision-making based on observed information. There are 2 primary patterns for overlooking a disease due to perceptual errors:

- Examining the affected area but ignoring the disease: This occurs when the radiologist inspects the region with the abnormality but fails to recognize it, possibly due to subtle presentation, distractions, or visual fatigue.
- Not examining the affected area: This happens when the radiologist misses the region with the abnormality entirely, often due to inefficient search patterns, incomplete scanning, or being misled by more prominent findings elsewhere.

Kundel [5] investigated the effects of perceptual errors in radiology and concluded that decisions and outcomes improve when radiologists' experiences are enhanced.

Radiologists and radiology educators understand the stakes associated with missed diagnoses due to perceptual errors but have limited tools to combat these errors. Classical educational texts include general concepts, for example, "...scan the areas of least interest first, working toward the more important areas" [6], which, unfortunately, are inadequate to improve radiologist performance meaningfully.

Eye-tracking technology has been previously proposed as a tool to evaluate radiologist perception. Eye trackers are powerful because they provide high (>30 Hz) temporal and spatial resolution (approximately 1 degree of error). With the aid of eye tracking, quantitative analyses can be performed to understand the cognitive and perceptual processes better. Eye-tracking technology has previously proven relevant in evaluating decision-making processes [7], attention interruption [8], skill level determination [9], and impact of search pattern education [10].

In 2017, van der Gijp et al [11] performed a systematic literature review outlining the current state of science concerning visual perception in radiology. A key tenet is the global-focal search model [12-14], which can be summarized as the generation of an initial, fast global impression followed by a more detailed focal search. Eye-tracking technology allows these principles to be tested and potentially optimized to evaluate all clinically relevant portions of the exam in greater detail. Of the 22 relevant articles van der Gijp et al [11] reviewed, a consensus "traditional" feature set consisting of 5 features that could be experimentally measured was found to be associated with expertise.

Despite the development of this consensus feature set, visual search complexity may not be adequately captured by simple, low-dimensional features that do not fully describe how visual perception relates to skill. Machine learning is well-suited to provide deeper insight into radiologist visual search behavior and how this relates to radiologist performance. Waite et al [3] highlighted the importance of understanding perceptual expertise in radiology and the potential use of eye-tracking and perceptual learning methods in medical training to improve diagnostic accuracy. Lim et al [15] identified several features that can be extracted from eye-tracking data, including pupil size, saccade, fixations, velocity, blink, pupil position, electrooculogram, and gaze point, to be used in machine learning models. Among these features, fixation was the most commonly used feature in the studies reviewed.

Shamyuktha et al [16] developed a machine learning framework using eye gaze data such as saccade latency and amplitude to classify expert and nonexpert radiologists. Harezlak et al [17] investigated eye movement traits to differentiate experts and laymen in a similar study. Akshay et al [18] proposed a machine learning algorithm to identify eye movement metrics using raw eye-tracking data. Rizzo et al [19] used machine learning to detect cognitive interference based on eye-tracking data. Öder et al [20] applied machine learning to classify familiar web users based on eye-tracking data. Indeed, these techniques can be used to enhance competency assessment and feedback techniques in radiologists.

Eye tracking also holds the potential for understanding the longitudinal aspects of competency progression in medical education, allowing for examining how interpretive and diagnostic skills develop over time. Karargyris et al [21] and Bigolin Lanfredi et al [22] created and validated chest x-ray datasets with eye-tracking data and report dictation for developing such artificial intelligence systems. These datasets aim to support the research community in developing more complex support tools for radiology research.

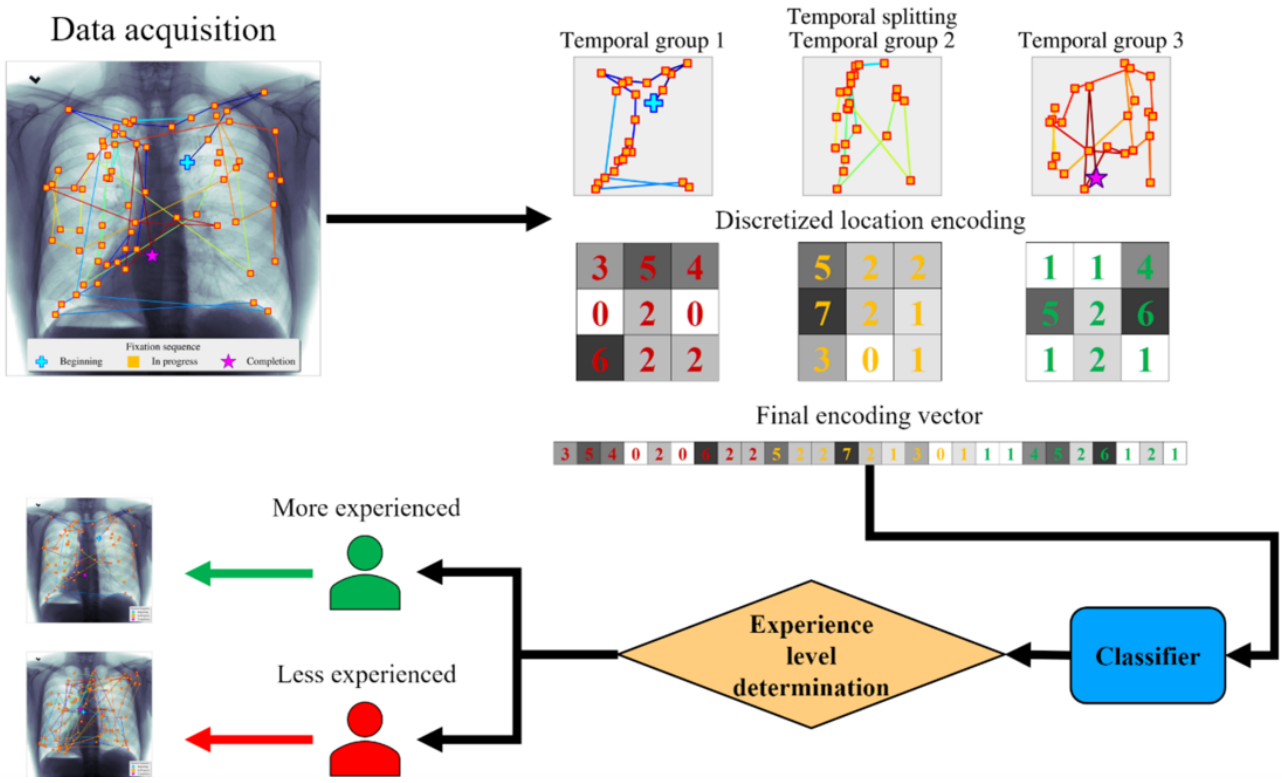
In this study, we use machine learning to compare the discriminability of 2 radiologists of different skill levels using, first, the aforementioned "traditional" gaze-based features (such

XSL•FO
RenderX

as time to scan, saccade length, the total number of fixations, and total regressive fixations) [11] and second, the "proposed" features that we developed to describe high-dimensional visual search patterns thoroughly and quantitatively. We curate the traditional feature sets to those that could be practically acquired without laborious manual ground truthing of exams, as this would permit large-scale deployment of this technology to health care institutions. To highlight the use of eye-tracking data and artificial intelligence, we term our general approach "biometric radiology artificial intelligence."

The driving hypothesis behind the work presented in this paper is that gaze patterns measurably differ among radiologists as a function of their experience level. To test this hypothesis, we proposed a novel discretized feature encoding method that condenses fixation data into a few representative spatiotemporal bins for descriptive and predictive analytics purposes (Figure 1). With spatiotemporal binning, fixations are divided into a predefined number of temporal segments (bins). Within each temporal bin, the fixations are counted within spatial subdivisions of the image. This process results in a vector that captures detailed and structured information about both where and when fixations occurred. By splitting fixations into temporal bins, we capture the evolution of the visual search process over time, providing insights into how radiologists allocate their attention during different phases of image inspection. Also, spatial binning allows us to understand which regions of the image are being focused on and how frequently. In addition, this method transforms raw fixation data into structured features that can be effectively used by machine learning models.

Figure 1. Overall algorithm: the steps required to generate proposed features from the raw dataset and build the proposed machine learning model.



We collected the gaze fixation data from radiologists while they were reading the x-rays. These data were then segmented into fixed temporal groups before discretizing them to convert them into final encoded vectors. The final encoded features were then used in training machine learning algorithms to classify radiologists.

Collecting data from 2 participants—1 faculty member (expert) and 1 resident (trainee)—we analyzed their behavior and level of experience using the proposed approach. Using stratified cross-validation over 10 folds, we compared the area under the curve (AUC) performance of several classifiers using the proposed methodology with the AUC performance of those same classifiers when using a traditional feature set (Table 1). We then confirmed our results using data from a second similarly designed, larger study evaluating 8 participants—4 faculty members (expert) and 4 residents (trainee). The remainder of the paper is structured as follows: *Methods* presents the data collection and preparation procedures and details of the proposed method; *Results* describes the simulation study and interpretation; and *Discussion* presents the discussion, concluding remarks, and advice to practitioners.

XSL•FO
RenderX

Table 1. State-of-the-art features from 22 relevant studies.

Attribute (per trial)	Attribute description	Association with high level of expertise (percentage of the total number of included studies)
Total time to scan	Measures the total duration spent scanning the chest x-ray image, indicating the thoroughness of the visual inspection	Decrease (45.45%)
Regressive fixation count	Counts the number of distinct locations revisited during the scan, suggesting areas of uncertainty or interest	Increase (4.55%) or decrease (4.55%)
Fixation count	Total number of fixations, reflecting the intensity of the visual scrutiny	Decrease (18.18%)
Total saccade length	Sum of all saccade lengths, indicating the extent and pattern of the visual search (time between fixations) [23] captured in a single chest x-ray scan	Increase (9.09%) or decrease (4.55%)
Coverage	Percentage of salient regions covered by the gaze, reflecting the comprehensiveness of the examination	Increase (9.09%) or decrease (9.09%)

Methods

Study Design, Data Collection, and Preparation

Overview

The study design was prospective, controlled, block-randomized, and Institutional Review Board (IRB) approved. Each study participant completed 4 roughly 1-hour sessions in a radiology reading room, including tutorial, calibration, assessment, and annotation periods. The tutorial included an overview of the assessment period and instructions on how to perform dictation and annotation consistently. Calibration was performed to ensure that recorded and actual gaze were consistent based on a 9-point custom calibration mapping script.

Nodule and normal cases were derived from the Shiraishi 2000 chest radiograph dataset [24], which includes 154 chest radiographs with 5 degrees of subtlety from level 1 (extremely subtle) to level 5 (obvious). Distractor cases were derived from the VinDr chest radiograph dataset [25]. A total of 3 sets of 6 nodule cases from the Japanese Society of Radiological Technology dataset, 1 set each from the intermediate difficulty levels (2, 3, and 4), and 1 set of 9 normal cases from the Japanese Society of Radiological Technology dataset were randomly sampled without replacement. In total, 2 cases each of pneumothorax, cardiomegaly, and consolidation from the VinDr dataset were randomly sampled without replacement to serve as distractor cases. These distractor cases functioned mainly to prevent control subject bias to the nodule detection task. Each participant reviewed the exams only once during the trial, and all study participants reviewed the same set of cases.

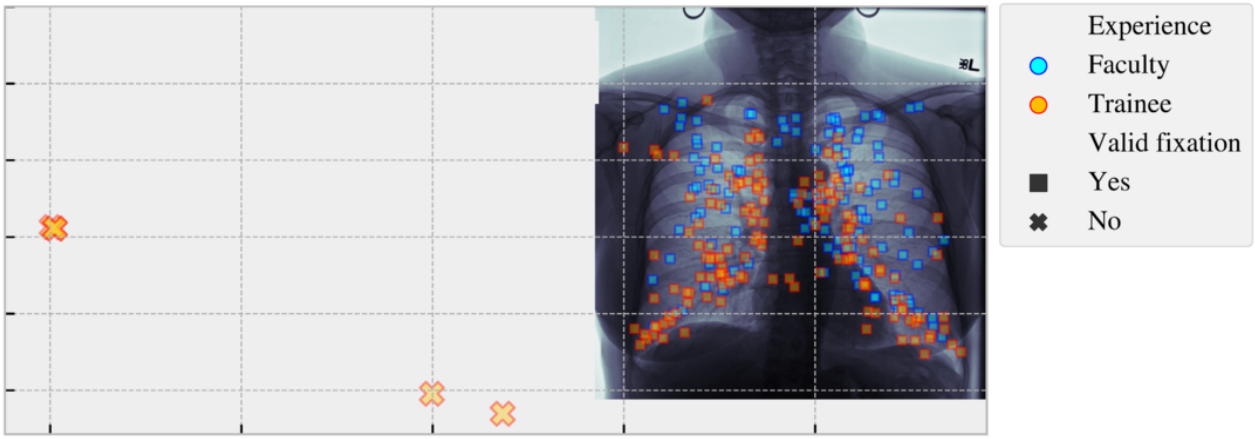
A custom software tool was developed to automatically display the study images and capture time-stamped bilateral gaze, bilateral pupil, head pose, voice, annotation, and image display configuration data. No chin rest was used to ensure that the study was performed in a manner that was as close as possible to a clinical setting. After each session, data were transferred to a database for further analysis.

Data Acquisition

In the first study, the EyeLink 1000 eye tracker and software were used to collect eye-tracking data [26]. A total of 2 participants—1 faculty member (9 years of faculty experience) and 1 resident (3 years of trainee experience)—observed a series of chest x-ray images, which contained a balanced class composition of normal scans (no abnormalities), abnormal scans (mass or nodule present), or abnormal scans with pleural effusion. A total of 110 trials (55 trials were studied by each participant) were recorded. We leveraged the EyeLink suite to remove most artifacts, such as blinks, from the eye-tracking data captured in each participant's trial and manually filtered remaining artifacts, such as off-screen distractions left unprocessed (eg, the far-displaced fixations in Figure 2).

XSL•FO
RenderX

Figure 2. Example of eye-tracking fixations for 1 trial processed by the EyeLink software. The fixations illustrated include participants 1 (blue) and 2 (red) superimposed on the image displayed during the trial. The "invalid" fixations that were not successfully filtered out are shown as "x" markers and were manually removed during data processing.



In the second study, a Tobii 5L eye tracker was used [27]. This second dataset included 8 participants (4 faculty with an average of 12.75 years of faculty experience and 4 trainees with 2.25 years of trainee experience), each scanning the same set of 27 images. The Tobii gaze data were unprocessed to evaluate the robustness of the proposed method to fixation postprocessing.

Common (Traditional) Features

To establish a necessary baseline to which the proposed methodology can compare, several attributes were established based on a meta-analysis done by van der Gijp et al [11] in 2017. The baseline is used for two main reasons: (1) traditional features are well-known and correlate with radiologist expertise, serving as a necessary reference point to evaluate our proposed method's effectiveness, and (2) comparing our novel discretized vector encoding method against this baseline demonstrates the added value, improved classification accuracy, and robustness of our new approach. We separated those features based on if they required ground truthing of exams. Meanwhile, ground truthing of medical exams is costly and time-consuming as it involves manual annotation by experts, which also limits the scalability of the method. In addition, reliance on ground truth annotations can introduce biases and errors, as the annotations themselves might vary between experts, which limits applicability and transferability to real-world applications. Consequently, features that required knowledge of the image abnormalities' ground truth location (ie, area of interest) were removed: fixation duration on the area of interest, number of fixations on the area of interest, and the time between trial start and the first fixation on the area of interest. This ensures that the proposed approach can be more easily and widely applied in clinical settings without the need for extensive preparatory work. It also helps mitigate potential variability and subjectivity in the training data, leading to more robust and generalizable models. Furthermore, by excluding features tied to known locations abnormally, we can better capture inherent differences in visual search patterns between experienced and less experienced radiologists.

Table 1 summarizes the remaining attribute names, descriptions, and expected association with levels of expertise. All features were used as originally defined except for coverage. Salient regions refer to areas of an image that are not part of a peripheral black background. This is typically necessary because users may be viewing scans with different amounts of background area. As noted previously, we used the Tobii gaze data without fixation postprocessing. For evaluating traditional features using fixations in the Tobii dataset, we substituted raw gaze data with fixation data. For purposes of clarity and brevity, we use fixations and gaze interchangeably for the remainder of the paper.

Proposed Approach: Discretized Vector Encoding for Fixation Data

Here, we describe the proposed method for directly using the fixation patterns as an alternative approach to using the current and previously described attributes in Table 1. The proposed strategy aims to extract information from fixations in the following 2 ways: first, geometric alignment: this involves mapping the coordinates of eye fixations on the chest x-ray images into a Cartesian grid. Each fixation is assigned to a specific grid cell based on its position, such as the Cartesian locations of the fixations when displayed on a chest x-ray image. Second, temporal order in which the fixations appear: the order in which fixations occur is crucial. Fixations are split into temporal bins, preserving the sequence of visual inspection. For each trial with recorded fixation data, we split the fixations into t number of temporal bins (each bin covers "total time divided by the number of bins in seconds") or groups before counting the number of fixations captured within square grids or subdivisions of size x. Then, the t number of x-by-x grids is encoded into a single vector of size 1-by-$(x \times x \times t)$. The overall procedure is described in pseudocode in algorithm 1 (Textbox 1) and illustrated in Figure 3.

XSL•FO
RenderX

Textbox 1. Discrete vector encoding for fixation data.

Algorithm 1: Vector encoding for fixation data

- Input
- n-Fixation coordinates of a single trial, F[n×2], number of x- and y-axis subdivisions, (x,y), number of temporal groups, (t)
- Output
- Encoded vector, V[1×(t×x×y)]
- Initialize
- Create array A[t×x×y] and centroids C[(x×y)×2] corresponding to the center of each grid subdivision (defined by second and third indices of A
- Evenly split fixations into t-groups, T = ([F1, F2,]1, , [, F(n-1) Fn]t)
- Procedure
- For $i = 1 \rightarrow >t$ do:

  $f = T_i$

  For $j = 1 \rightarrow len\ (f)$ do:

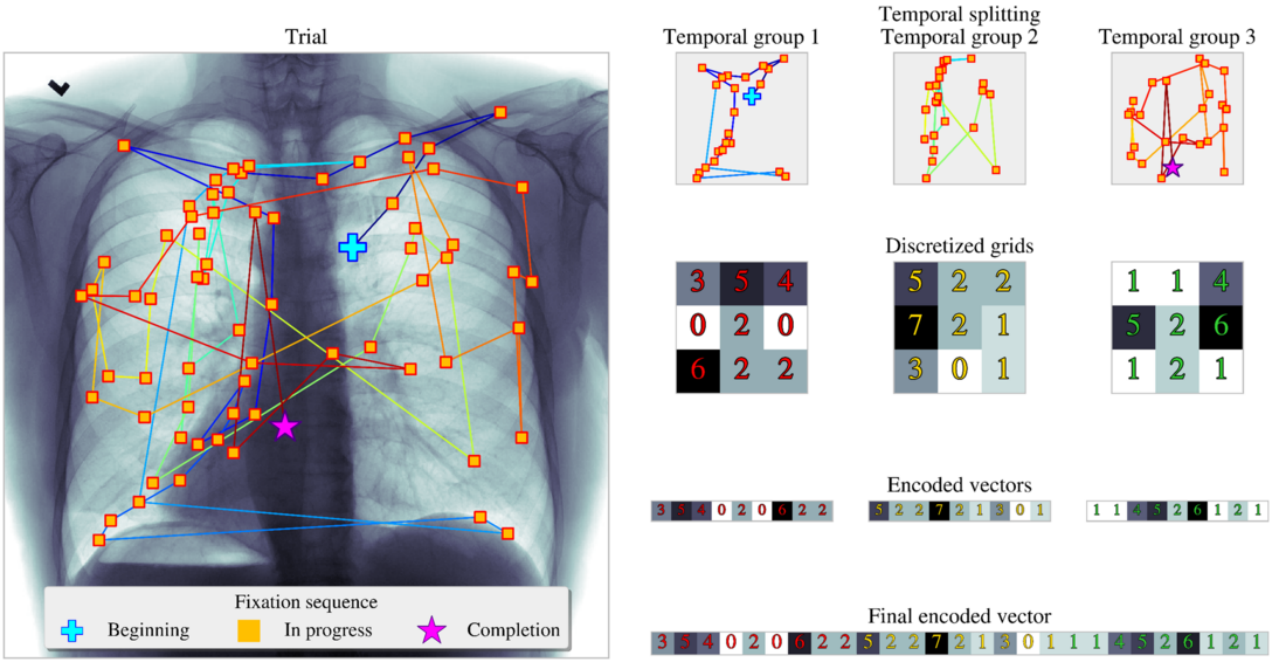
  $C^* = argmin\ (\| C - f_j \|)$

  $A\ [i, C_x^*, C_y^*] \mathrel{+}= 1$

  $V = vec(A)$

  Return V

Figure 3. Proposed discretized vector encoding for fixation data. Bins 1, 2, and 3 capture fixations in a preserved spatial dimension across different temporal windows. Each row represents a temporal bin, and within each bin, the chest x-ray image is divided into spatial grids. The fixations are counted within each grid cell, providing a detailed representation of the radiologist's visual search pattern over time.



In Figure 3, participant 1 inspects a single chest x-ray image; the processed fixations are captured as illustrated gold squares with red edges, with the first fixation labeled as a blue cross and the last fixation visualized as a magenta star. In this example, the fixations are split into 3 temporal segments (Step 1), in which 3-by-3 grids count the number of fixations within them (Step 2). Then, the proposed algorithm outputs the final encoding vector as the flattening and concatenation of the set of 3-by-3 grids (Step 3). For a given (square) grid of size *x* and *t* number of temporal segments, the final yielded output vector is of length, regardless of trial temporal duration. Segmenting the raw data into fixed temporal segments is one of the benefits of this approach and a strategy developed and imposed to generate consistent numbers of variables on the encoding output across different trials. As the number of fixations across each trial can vary between participants, fixing the number of temporal segments allows the capturing of trial duration while conforming to a prescribed number of grid subdivisions and

XSL•FO
RenderX

temporal groups. For example, with Figure 3 as a reference, the final encoded vector will yield a vector with larger values therein for longer trials and yield a sparse vector (with lower values or values of zero) for shorter trials. Users can increase the fidelity of the grid and the number of time groups to represent a continuous spatiotemporal domain more closely. It is notable that the proposed methodology possesses the capability for tensor configuration for use in deep learning architecture by using *t* layers of grids. This tensor configuration is not studied in the paper due to the small sample size. The introduced technique is designed with more accessible or simpler classifiers in mind.

Performance Metric and Simulation Setup

To evaluate the discriminability of participants using the proposed approach, we use a stratified-fold cross-validation to calculate the AUC metric for several classification models, where each of the folds contains 5 trials from both levels of experience as the hold-out set. The study was performed on the data acquired by the EyeLink and Tobii equipment separately, and the following sections will contain an elaboration on their respective results. We performed cross-validations on a full-factorial configuration of 5, 7, 10, and 15 square grid subdivisions and 3, 5, 10, and 20 temporal groupings and selected the settings for each classifier that yielded the best results. In the presentation of these results, the average scores were calculated by computing the AUC metrics at the lowest level (data acquisition method, classifier, data type, feature extraction method, grid-size, temporal-group, and cross-validation seed) and averaged to the presented levels of granularity. Given the small sample size of 110 (EyeLink dataset) and 216 (Tobii dataset) trials, and high dimensionality in the chosen configurations (up to 4500 encoded variables in our study), there are available pathways that we have used to alleviate the effects of the curse of dimensionality present [28], such as principal component analysis (PCA) [29] and kernel principal component analysis (KPCA) [30]. The feature extraction and dimensionality reduction methods used include reducing the input data to 2 dimensions (with varying amounts of explained variance) and fixing the amount of variance explained to 50%, 90%, and 99% (with varying numbers of dimensions). These techniques were used not only to reduce the density of the data but also to introduce an additional preprocessing step that leverages the spectral decomposition of data collected from each participant.

Some of the major reasons for considering PCA and KPCA instead of the other alternatives include the following: PCA and KPCA are among the most popular method of dimensionality reduction; most technical practitioners, especially in the field of medicine, are familiar with PCA and KPCA; PCA and KPCA have rigorous mathematical properties and commonly used baseline methods in statistical analysis; and PCA and KPCA have relatively low computational complexity compared with many of the other shallow and deep alternatives.

All the codes were written using Python. The used libraries and versions are as follows: *matplotlib* (3.7.1), *seaborn* (0.12.2), *tqdm* (4.65.0), *scipy* (1.8.0), *scikit-learn* (1.0.2), *xgboost* (1.7.5), *GPy* (1.10.0), *numpy* (1.21.6), *pandas* (2.0.1), and *joblib* (1.2.0).

Ethical Considerations

This study was conducted in full compliance with human participant research ethics and was reviewed and approved by the University of Texas Health San Antonio Institutional Review Board (20190533HU). All participants were fully informed about the purpose and procedures of the study, and informed consent was obtained before their inclusion. To ensure the privacy and confidentiality of participant data, all identifying information was removed to anonymize the dataset before analysis. Furthermore, participants were compensated $400 USD for their time and involvement in the study.

Results

Competing Algorithms and Training

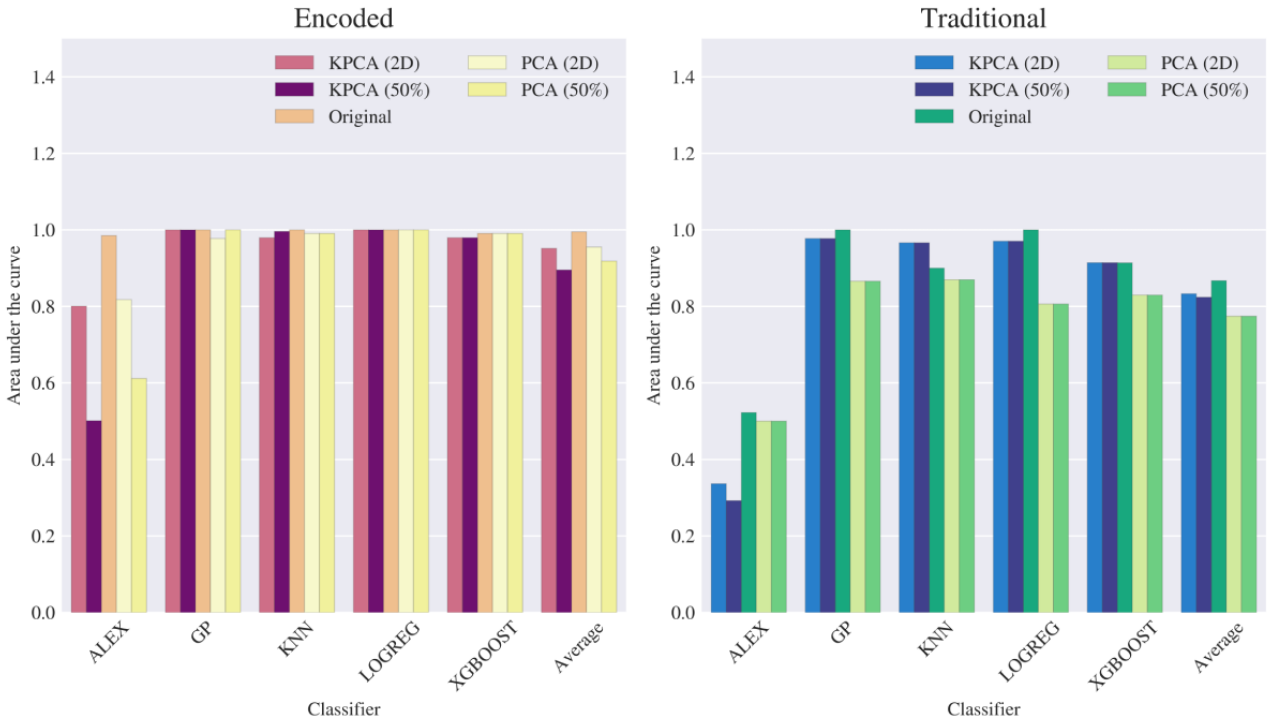
In this study, we use the Gaussian process, logistic regression, and k-nearest neighbors classifiers from the *Scikit-learn* [31] package; the extreme gradient boosting (XGBoost) [32] tree-based ensemble classifier; and a modified AlexNet [33] deep learning classifier. The *Scikit-learn* classifiers were selected for their accessibility to users, while the XGBoost and AlexNet-like neural networks were chosen as more complex classifiers. The logistic regression, k-nearest neighbors, and XGBoost classifiers used *Scikit-learn*'s *StratifiedKFold* and *GridSearchCV* packages to train on the balanced accuracy loss function (also defined by *Scikit-learn*), while the Gaussian process methodology used Laplace approximation as detailed in their documentation [31]. Finally, the AlexNet-like classifier used sparse categorical cross-entropy [34] for training.

EyeLink Dataset

Figure 4 illustrates the average AUC across each classifier tasked with distinguishing between 2 participants (particularly between 2 levels of experience) using either the traditional or the proposed encoded data types (features). Along with the original data types, we include the average AUC of the classifiers based on the usage of select feature extraction configurations. The encoded features extracted from the raw dataset, shown on the left, illustrate a consistently high AUC score compared with the traditional features shown on the right, implying that the model performance for each classifier (except for certain feature extraction configurations of the AlexNet model) has high discriminatory power under optimal spatiotemporal encoding settings.

XSL•FO
RenderX

Figure 4. Numerical study results on the area under the curve metric reported for each classifier when consuming the EyeLink dataset, organized by the aggregated average of classifier, data type, and select feature extraction levels using the original dataset, principal component analysis (PCA), and kernel principal component analysis (KPCA). Alex: AlexNet-like neural network classifier; Average: average of all classifiers; GP: Gaussian process; KNN: k-nearest neighbors; LR: logistic regression; XGBoost: extreme gradient boosting;.



We also present the AUC metrics from Figure 4 below in Table 2. The performance of the classifiers using the encoded data type consistently yielded higher discriminatory power than those using the traditional data type across all feature extraction methods. Encoding the fixation data into the proposed spatiotemporal elements provides more information each classifier can use to determine the experience level of a given participant more effectively than using traditional attributes. This table illustrates the original encoded data to possess the highest performance, with AUC scores consistently above 0.98 across all classifiers. However, usage of the traditional data yields roughly 0.522 at worst, as seen in the reported results for the AlexNet classifier. This trend of encoded data providing better results is also seen when using feature extraction; although a performance decrease is observable when reducing dimensions either through an information covariance matrix (PCA) or spatial relation (KPCA), the use of encoded data still outperforms those corresponding to the use of the traditional data. This suggests that the loss in information due to dimensionality reduction can be considered negligible in light of the benefits of using spatiotemporal encoding. The lower relative performance of the AlexNet-like classifier is likely caused by the number of training samples available in this study. The report on AUC in the table for the classifier is higher for the encoded data type, where it is observable that using the data without dimensionality reduction provides the best performance. This effect has been studied in Sumner and Alaeddini [35], in which neural networks already perform feature extraction throughout each present layer; this supportively evidences the reported results here, whereas (besides the small dataset) performing feature extraction beforehand may not provide enough information for the network to use its architecture to its fullest potential.

XSL•FO
RenderX

Table 2. Numerical tabulation of area under the curve scores across each classifier and data type and select feature extraction methods.

Feature extraction method and data type	Classifier for the EyeLink dataset, AUC[a]				
	AlexNet	Gaussian process	KNN[b]	LR[c]	XGBoost[d]
KPCA[e] (2D)					
Encoded	0.801	1.000	0.980	1.000	0.980
Traditional	0.336	0.978	0.967	0.970	0.914
KPCA (50%)					
Encoded	0.501	1.000	0.996	1.000	0.980
Traditional	0.292	0.978	0.967	0.970	0.914
Original					
Encoded	0.985	1.000	1.000	1.000	0.991
Traditional	0.522	1.000	0.900	1.000	0.914
PCA[f] (2D)					
Encoded	0.818	0.978	0.991	1.000	0.991
Traditional	0.500	0.866	0.870	0.806	0.830
PCA (50%)					
Encoded	0.611	1.000	0.991	1.000	0.991
Traditional	0.500	0.866	0.870	0.806	0.830

[a]AUC: area under the curve.

[b]KNN: k-nearest neighbor.

[c]LR: logistic regression.

[d]XGBoost: extreme gradient boosting.

[e]KPCA: kernel principal component analysis.

[f]PCA: principal component analysis.

By using the encoded vectors for classification, differences in eye-tracking patterns can more consistently be distinguished between the 2 participants. Figures 5 and 6 illustrate one such difference in search pattern behavior. The more experienced participant (participant 1, Figure 5) shows a more uniformly distributed search pattern across the chest x-ray. In contrast, the less-experienced participant (participant 2, Figure 6) focuses on regions where they suspect abnormalities. It is clear from a visual inspection that the behavior between these participants is markedly different and using the correct spatiotemporal configurations to capture the differences between the 2 participant's behavior by leveraging the proposed methodology (as reported numerically in Table 2) provides a consistent improvement of classification accuracy.

XSL•FO
RenderX

Figure 5. Scan of chest x-ray by participant 1 (faculty).

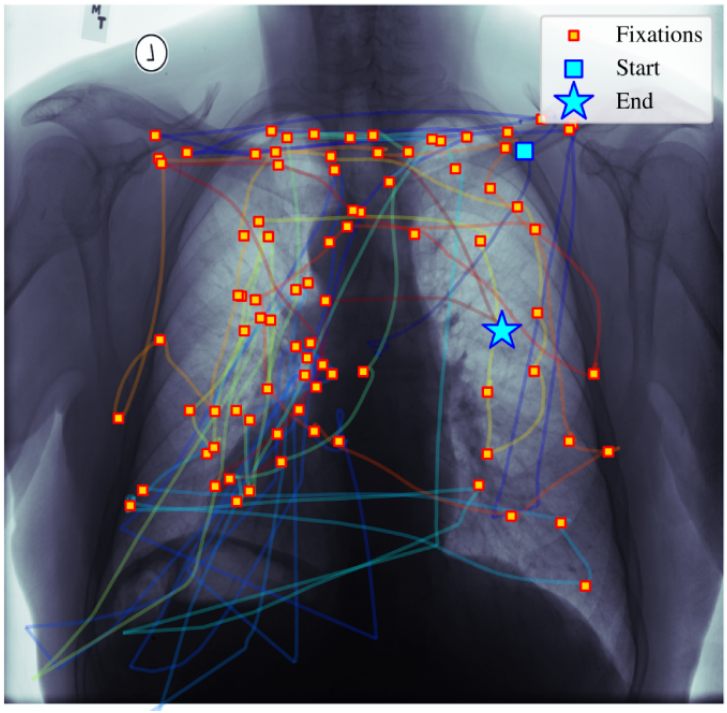
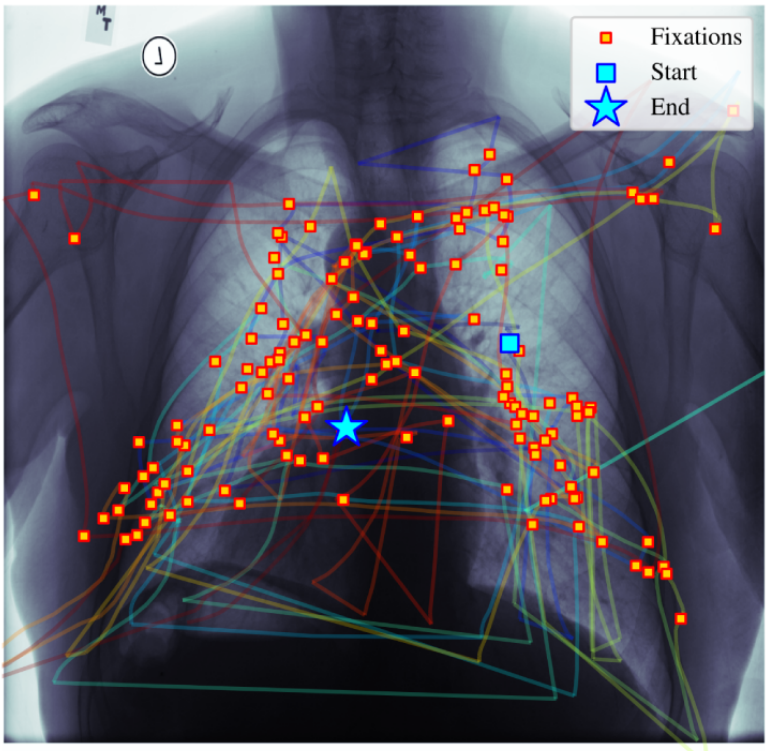


Figure 6. Scan of chest x-ray by participant 2 (trainee).

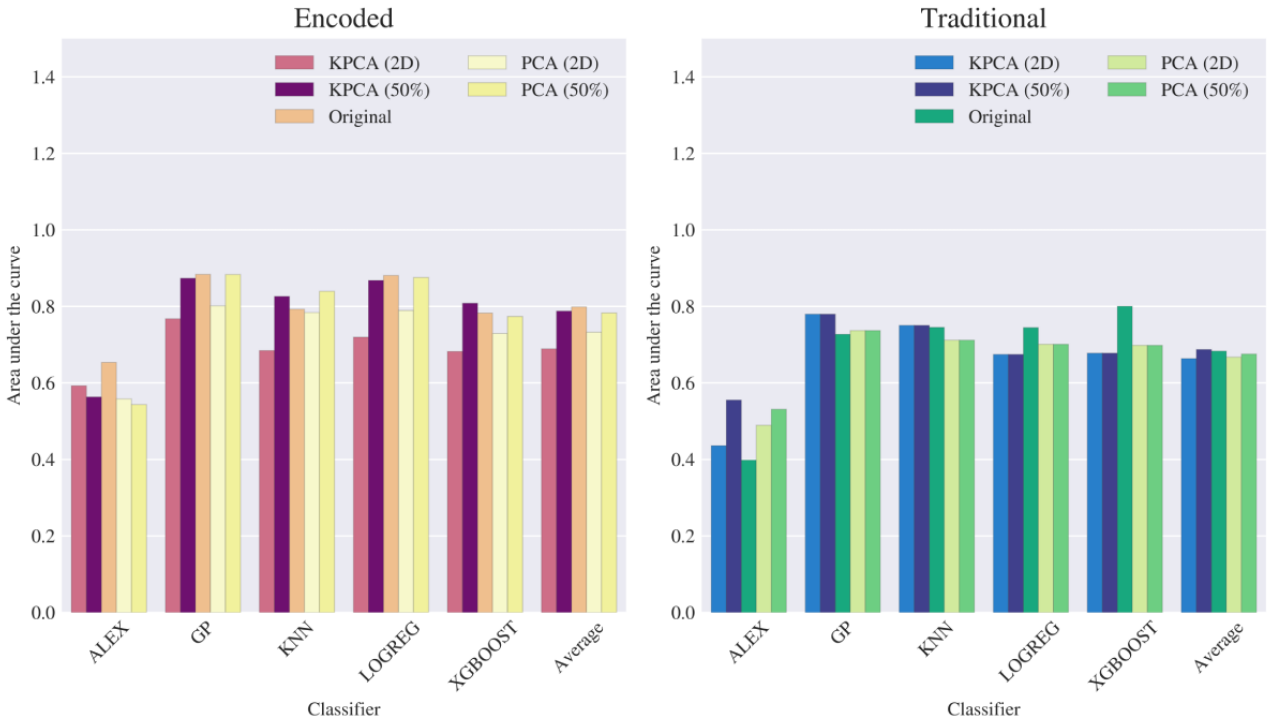


Tobii Dataset

We have performed the same analysis on the data acquired using Tobii eye-tracking equipment. It is notable that although the AUC scores from the EyeLink dataset are consistently high, natural anticipations allow one to observe more variation in classifier performance when more individual participants (classified as either a more-experienced faculty or less-experienced trainee) are introduced to the study. Figure 7 illustrates a report on AUC in a similar fashion to that in Figure 4, with lower scores across all classifying models for both data types. As seen in Figure 4, Figure 7 also suggests that the best performance for the encoded data on average is attained when using it without feature extraction, although, for several cases, we can observe that some form of feature extraction yields better results than their respective traditional dataset counterparts.

XSL•FO
RenderX

Figure 7. Numerical study results on the area under the curve metric reported for each classifier when consuming the Tobii dataset, organized by the aggregated average of classifier, data type, and selected feature extraction levels using the original dataset, principal component analysis (PCA), and kernel principal component analysis (KPCA). Alex: AlexNet-like neural network classifier; Average: average of all classifiers; GP: Gaussian process; KNN: k-nearest neighbors; LR: logistic regression; XGBoost: extreme gradient boosting;.



When inspecting Tables 3, we can numerically inspect the average (and variance of) AUC, F_1-score, accuracy, specificity, and sensitivity of each classifier when consuming each data type in both datasets. Within the Tobii dataset, the encoded data type generally outperformed (shown in italics) the traditional data type across most metrics and models. Although the encoded data type that was consumed within the Tobii dataset possessed more discriminatory capability than that in the traditional data, the performance gap was less pronounced than those observable in the EyeLink dataset. For example, the Tobii average (and variance) AUC scores for the encoded data type ranged from 0.55 (0.05) to 0.82 (0.04), while the traditional data type ranged from 0.51 (0.07) to 0.76 (0.05), and the EyeLink average (variance) AUC scores for the same data types ranged from 0.63 (0.07) to 1.0 (0.0) and from 0.52 (0.08) to 0.96 (0.01), respectively. There is a consistent trend across datasets that support the encoded data are capable of providing higher values of accuracy and performance; the F_1-score for the Gaussian process, k-nearest neighbors, logistic regression, and XGBoost were consistently higher when using the encoded data than when using traditional attributes in classification. This highlights the ability of the proposed encoding procedure to improve the balance between precision and recall in the classifiers and, as a result, the overall effectiveness of each model's predictions. In terms of specificity, the encoded data type is also shown to have a competitive edge in boosting a classifier's ability to correctly identify true negative class labels (experienced participants). As seen in Table 3, the average range of improvement lies between 0.01 to 0.04 for the EyeLink dataset and between –0.2 to +0.15 for the Tobii dataset; the negative value of the improvement is seen with the AlexNet-like model, which, as explained before, may have difficulty fitting well for classification on small datasets, made more difficult by the variation in subconscious behavior between participants that are recorded in spatiotemporal encodings by the proposed methodology.

XSL•FO
RenderX

Table 3. Numerical tabulation of mean and variance area under the curve, F_1-score, accuracy, specificity, and sensitivity across each data acquisition method, classifier, and data type.

Metric and classifier	Data acquisition system			
	EyeLink		Tobii	
	Data type		Data type	
	Encoded, average (SD)	Traditional, average (SD)	Encoded, average (SD)	Traditional, average (SD)
Area under the curve				
AlexNet	*0.63 (0.07)* [a]	0.52 (0.08)	*0.55 (0.05)*	0.51 (0.07)
GP[b]	*1 (0)*	0.96 (0.01)	*0.82 (0.04)*	0.76 (0.05)
KNN[c]	*0.96 (0.01)*	0.93 (0.01)	0.73 (0.05)	*0.74 (0.04)*
LR[d]	*1 (0)*	0.93 (0.01)	*0.82 (0.04)*	0.71 (0.08)
XGBoost[e]	*0.97 (0.0)*	0.91 (0.01)	*0.73 (0.05)*	0.71 (0.05)
F_1-score				
AlexNet	*0.43 (0.1)*	0.41 (0.09)	*0.39 (0.1)*	0.23 (0.1)
GP	*0.98 (0)*	0.90 (0.03)	*0.73 (0.07)*	0.71 (0.05)
KNN	*0.90 (0.03)*	0.88 (0.02)	0.61 (0.1)	*0.71 (0.04)*
LR	*0.98 (0)*	0.82 (0.04)	*0.74 (0.06)*	0.58 (0.09)
XGBoost	*0.96 (0)*	0.87 (0.02)	*0.68 (0.07)*	0.67 (0.05)
Accuracy				
AlexNet	*0.52 (0.02)*	0.49 (0.04)	*0.51 (0.0)*	0.5 (0.01)
GP	*0.99 (0.0)*	0.93 (0.01)	*0.76 (0.03)*	0.7 (0.03)
KNN	*0.93 (0.01)*	0.9 (0.01)	0.69 (0.03)	*0.7 (0.02)*
LR	*0.98 (0.0)*	0.88 (0.01)	*0.77 (0.03)*	0.64 (0.03)
XGBoost	*0.96 (0.0)*	0.9 (0.01)	*0.72 (0.03)*	0.67 (0.03)
Specificity				
AlexNet	*0.43 (0.22)*	0.42 (0.22)	0.46 (0.16)	*0.66 (0.12)*
GP	*0.98 (0)*	0.97 (0.01)	*0.80 (0.04)*	0.65 (0.07)
KNN	*0.97 (0.01)*	0.93 (0.01)	*0.76 (0.06)*	0.67 (0.04)
LR	*0.99 (0.0)*	0.97 (0.01)	*0.81 (0.04)*	0.67 (0.08)
XGBoost	*0.95 (0.01)*	0.92 (0.01)	*0.76 (0.05)*	0.63 (0.08)
Sensitivity				
AlexNet	*0.64 (0.22)*	0.57 (0.2)	*0.56 (0.17)*	0.32 (0.12)
GP	*0.99 (0)*	0.89 (0.04)	0.72 (0.07)	*0.76 (0.05)*
KNN	*0.89 (0.05)*	0.87 (0.03)	0.61 (0.1)	*0.74 (0.04)*
LR	*0.98 (0.01)*	0.77 (0.06)	*0.72 (0.06)*	0.6 (0.09)
XGBoost	*0.98 (0)*	0.86 (0.03)	0.68 (0.07)	*0.71 (0.05)*

[a]Superior values are italicized.

[b]GP: Gaussian process.

[c]KNN: k-nearest neighbor.

[d]LR: logistic regression.

[e]XGBoost: extreme gradient boosting.

XSL•FO
RenderX

Discussion

Principal Findings

In this study, we demonstrated the capacity of eye-tracking technology, combined with machine learning algorithms, to discriminate between radiologists' experience levels. For this purpose, we developed a novel feature encoding technique where fixations are first spatially arranged according to their Cartesian coordinates on chest x-ray images and temporally ordered. The fixations are then subdivided into predefined temporal bins, and within each bin, we count the number of eye fixations within each subdivision. These counted bins are then concatenated to form a vector encoding to be used as feature input for machine learning algorithms. Our experiments showed that the discretized vector encoding creates discriminative features that are not captured by conventional metrics. Using the encoding approach allows classifiers to better distinguish between participants in terms of experience level, which highlights performance gains (when compared with using traditional features for discrimination) of 6.9%, 7.11%, 9.14%, 9.59%, and 5.61% for AUC, accuracy, F_1-score, sensitivity, and specificity, respectively, aggregated across both EyeLink and Tobii datasets in Table 3. The Tobii dataset exhibits a lower performance gain (6.41%, 7.48%, 8.62%, 5.11%, and 9.45%) than observed using the EyeLink dataset (7.29%, 6.83%, 9.54%, and 13.13%) due to using a more diverse roster of participants; however, the trend in using the proposed eye-tracking encoding approach possessing the competitive edge is still present, highlighting the effectiveness of spatiotemporal assortment in the introduced method. These results validate our initial hypothesis that when appropriately encoded, eye-tracking data can provide nuanced insights into the difference between radiologist's expertise levels.

We can also observe the perceptual strategies radiologists use during diagnostic evaluations. Previous research has often focused on more general eye-tracking metrics without leveraging the full potential of machine learning to analyze the data. For example, studies by van der Gijp et al [11] and Waite et al [3] explored how visual search patterns correlate with diagnostic accuracy and expertise. With the help of the proposed encoding method, such machine learning models can be developed to determine expertise level and has the potential to identify and track potential features from eye fixations or gaze fixations.

Limitations

While our study has shown promising results and potential benefits, it is important to acknowledge limitations that may have a degree of effect on our findings. One such limitation is sample size; across both the EyeLink and Tobii datasets, there were 2 participants in one study (EyeLink) and 8 participants in the other (Tobii), with both containing small numbers of images scanned by each participant. Another condition involves the variation in data acquisition. In total, 2 different eye-tracking devices (EyeLink and Tobii) were used for data collection, and while serving the same overarching purpose of collecting data, some additional variability in the findings are notably attributed to the usage of 2 different hardware-software configurations. Another important consideration includes the difference in traditional feature sets between the EyeLink and Tobii datasets. Coupled with data acquisition differences, some features from the EyeLink software were not congruent with the Tobii dataset, such as the usage of fixations (EyeLink) versus gaze (Tobii). When applying the encoding approach to these datasets, the Tobii dataset had larger yielded values in each output vector. This did not affect the results substantially; however, it underscores the challenge of directly comparing data from 2 sources. One final consideration was our decision to remove certain metrics related to the location of abnormalities in chest x-rays as features in the traditional data type during performance evaluation. For example, we did not consider the time to first fixation on the region of abnormality. This and other like attributes possess statistical significance in previous works; however, their inclusion necessitates extensive labeling, validation, and other processing in order to establish ground truth information for each image scanned by each participant.

Conclusions

Despite the limitations above, this study holds significant promise and offers a range of benefits worthy of attention and consideration for use. By extracting spatiotemporal features from eye-tracking data, the proposed approach has demonstrated the capacity to differentiate users based on eye-tracking behavior alone instead of traditional methods and can be extended for use in fields ranging from medical to educational. The approach enables discriminability between users and offers a pathway to gaining deeper insights into generalized differences between levels of expertise. By exploring these pathways, this approach holds the potential to establish more effective educational programs that can assist users optimize their search patterns. Respective to the study conducted, by providing support to radiologists to find abnormalities quickly and accurately in chest x-rays, our approach seeks to reduce perceptual errors in medical diagnoses. In fields where the development of unique and precise search patterns is important, the proposed approach offers a valuable source of knowledge transfer. Transmission of expertise from more-experienced individuals to less-experienced individuals can be facilitated and result in increased streamlining during the learning process and yield more efficient and accurate search patterns. The potential benefits can apply to professionals and trainees or students alike.

In summary, we have shown the potential for spatiotemporal features extracted from eye-tracking data to be useful in discriminating between radiologists of different skill levels and opening the door to improving education. We plan to augment this research by increasing the number of radiologists to demonstrate generalizability and exploring additional types of spatiotemporal analyses. The implications of our findings extend beyond radiology, suggesting that similar methodologies could revolutionize training and assessment protocols in various fields that rely on visual cognition like aviation and ground transportation. Further research could explore the integration of these techniques into real-time training tools, potentially transforming educational paradigms in professions requiring visual expertise.

XSL•FO
RenderX

Acknowledgments

This research is funded by San Antonio Medical Foundation, and principal investigators KC (University of Texas Health Science Center at San Antonio) and EJG (University of Texas at San Antonio).

Data Availability

Data collected and analyzed in this study are available upon request. Interested researchers can obtain the data by completing a data request form, which includes their planned project outline and the details of the person responsible for handling the data. This process ensures that the data are used responsibly and for scientifically valid purposes. The request forms are reviewed on a case-by-case basis to ensure alignment with ethical and research standards. Images and annotations were obtained from the National Institutes of Health Clinical Center and can be downloaded on the web [36]. The Institutional Review Board restricts sharing of the gaze data. In case further information, please contact KC (clarkkl@uthscsa.edu).

Conflicts of Interest

KC holds equity in Zauron Labs Inc and NC is a co-owner of Zauron Labs.

References

1. An update on cancer deaths in the United States. Centers for Disease Control and Prevention. 2022 Feb 28. URL: https://stacks.cdc.gov/view/cdc/119728 [accessed 2022-02-28]
2. Screening for lung caner: U.S. Preventitive Services Task Force recommendation statement. Lung Cancer Research foundation. 2014. URL: https://www.lungcancerresearchfoundation.org/screening-for-lung-cancer-u-s-preventive-services-task-force-recommendation-statement/ [accessed 2024-09-23]
3. Waite S, Grigorian A, Alexander RG, Macknik SL, Carrasco M, Heeger DJ, et al. Analysis of perceptual expertise in radiology - current knowledge and a new perspective. Front Hum Neurosci 2019;13:213 [FREE Full text] [doi: 10.3389/fnhum.2019.00213] [Medline: 31293407]
4. Bruno MA, Walker EA, Abujudeh HH. Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction. Radiographics 2015;35(6):1668-1676. [doi: 10.1148/rg.2015150023] [Medline: 26466178]
5. Kundel HL. Perception errors in chest radiography. Seminars in Respiratory Medicine 1989;10(03):203-210. [doi: 10.1055/s-2007-1006173]
6. Goodman LR. Felson's Principles of Chest Roentgenology: A Programmed Text. 2nd ed. Radiology 2000 Mar;214(3):848. [doi: 10.1148/radiology.214.3.r00fe55848]
7. Tanoubi I, Tourangeau M, Sodoké K, Perron R, Drolet P, Bélanger ME, et al. Comparing the visual perception according to the performance using the eye-tracking technology in high-fidelity simulation settings. Behav Sci (Basel) 2021;11(3):31 [FREE Full text] [doi: 10.3390/bs11030031] [Medline: 33807673]
8. Wu CC, Wolfe JM. Eye movements in medical image perception: a selective review of past, present and future. Vision (Basel) 2019;3(2):32 [FREE Full text] [doi: 10.3390/vision3020032] [Medline: 31735833]
9. Kelly BS, Rainford LA, Darcy SP, Kavanagh EC, Toomey RJ. The development of expertise in radiology: in chest radiograph interpretation, "Expert" search pattern may predate "Expert" levels of diagnostic accuracy for pneumothorax identification. Radiology 2016;280(1):252-260. [doi: 10.1148/radiol.2016150409] [Medline: 27322975]
10. Wolfe JM, Lyu W, Dong J, Wu CC. What eye tracking can tell us about how radiologists use automated breast ultrasound. J Med Imaging (Bellingham) 2022;9(4):045502 [FREE Full text] [doi: 10.1117/1.JMI.9.4.045502] [Medline: 35911209]
11. van der Gijp A, Ravesloot CJ, Jarodzka H, van der Schaaf MF, van der Schaaf IC, van Schaik JPJ, et al. How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology. Adv Health Sci Educ Theory Pract 2017;22(3):765-787 [FREE Full text] [doi: 10.1007/s10459-016-9698-1] [Medline: 27436353]
12. Drew T, Evans K, Võ MLH, Jacobson FL, Wolfe JM. Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images? Radiographics 2013;33(1):263-274 [FREE Full text] [doi: 10.1148/rg.331125023] [Medline: 23104971]
13. Kundel HL, Nodine CF. Interpreting chest radiographs without visual search. Radiology 1975;116(3):527-532. [doi: 10.1148/116.3.527] [Medline: 125436]
14. Swensson RG. A two-stage detection model applied to skilled visual search by radiologists. Perception & Psychophysics 1980;27(1):11-16. [doi: 10.3758/bf03199899]
15. Lim JZ, Mountstephens J, Teo J. Eye-tracking feature extraction for biometric machine learning. Front Neurorobot 2022;15:796895 [FREE Full text] [doi: 10.3389/fnbot.2021.796895] [Medline: 35177973]
16. Shamyuktha RS, Amudha J, Aiswariya Milan K. A machine learning framework for classification of expert and non-experts radiologists using eye gaze data. : IEEE; 2022 Presented at: 2022 IEEE 7th International Conference on Recent Advances

XSL•FO
RenderX

and Innovations in Engineering (ICRAIE); December 01-03, 2022; Mangalore, India. [doi: 10.1109/icraie56454.2022.10054277]
17. Harezlak K, Kasprowska P, Kasprowska S. Eye movement traits in differentiating experts and laymen. 2017 Presented at: Man-Machine Interactions 5: 5th International Conference on Man-Machine Interactions, ICMMI 2017; October 3-6, 2017; Kraków, Poland. [doi: 10.1007/978-3-319-67792-7_9]
18. Akshay S, Megha YJ, Shetty CB. Machine learning algorithm to identify eye movement metrics using raw eye tracking data. : IEEE; 2020 Presented at: 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT); August 20-22, 2020; Tirunelveli, India p. 949-955. [doi: 10.1109/icssit48917.2020.9214290]
19. Rizzo A, Ermini S, Zanca D, Bernabini D, Rossi A. A machine learning approach for detecting cognitive interference based on eye-tracking data. Front Hum Neurosci 2022;16:806330 [FREE Full text] [doi: 10.3389/fnhum.2022.806330] [Medline: 35572006]
20. Öder M, Eraslan S, Yesilada Y. Automatically classifying familiar web users from eye-tracking data: a machine learning approach. Turkish Journal of Electrical Engineering and Computer Sciences 2022;30(1):233-248. [doi: 10.3906/elk-2103-6]
21. Karargyris A, Kashyap S, Lourentzou I, Wu JT, Sharma A, Tong M, et al. Creation and validation of a chest X-ray dataset with eye-tracking and report dictation for AI development. Sci Data 2021;8(1):92 [FREE Full text] [doi: 10.1038/s41597-021-00863-5] [Medline: 33767191]
22. Bigolin Lanfredi R, Zhang M, Auffermann WF, Chan J, Duong PAT, Srikumar V, et al. REFLACX, a dataset of reports and eye-tracking data for localization of abnormalities in chest x-rays. Sci Data 2022;9(1):350 [FREE Full text] [doi: 10.1038/s41597-022-01441-z] [Medline: 35717401]
23. Robinson KA. Dictionary of eye terminology. BMJ 1997;81(11) [FREE Full text] [doi: 10.1136/bjo.81.11.1021c]
24. Shiraishi J, Katsuragawa S, Ikezoe J, Matsumoto T, Kobayashi T, Komatsu K, et al. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. AJR Am J Roentgenol 2000;174(1):71-74. [doi: 10.2214/ajr.174.1.1740071] [Medline: 10628457]
25. Nguyen HQ, Lam K, Le LT, Pham HH, Tran DQ, Nguyen DB, et al. VinDr-CXR: an open dataset of chest X-rays with radiologist's annotations. Sci Data 2022;9(1):429 [FREE Full text] [doi: 10.1038/s41597-022-01498-w] [Medline: 35858929]
26. EyeLink 1000 plus. SR-Research. 2017. URL: https://www.sr-research.com/eyelink-1000-plus/ [accessed 2024-09-18]
27. Tobii eye tracker 5. Tobii. 2020. URL: https://gaming.tobii.com/product/eye-tracker-5/ [accessed 2024-09-18]
28. Aggarwal CC, Hinneburg A, Keim DA. On the surprising behavior of distance metrics in high dimensional space. 2001 Presented at: Database Theory - ICDT 2001; January 4-6, 2001; London, United Kingdom p. 8. [doi: 10.1007/3-540-44503-x_27]
29. Wold S, Esbensen K, Geladi P. Principal component analysis. Chemometrics and Intelligent Laboratory Systems 1987;2(1-3):37-52. [doi: 10.1016/0169-7439(87)80084-9]
30. Schölkopf B, Smola A, Müller KR. Kernel principal component analysis. 1997 Presented at: Artificial Neural Networks — ICANN '97: 7th International Conference; October 8–10, 1997; Lausanne, Switzerland. [doi: 10.1007/bfb0020217]
31. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. The Journal of Machine Learning Research 2011;12:2825-2830 [FREE Full text]
32. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. 2016 Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, CA. [doi: 10.1145/2939672.2939785]
33. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun. ACM 2017;60(6):84-90. [doi: 10.1145/3065386]
34. Cybenko G, O'Leary DP, Rissanen J. The Mathematics of Information Coding, Extraction and Distribution. Germany: Springer Science & Business Media; 1998.
35. Summer J, Alaeddini A. Analysis of feature extraction methods for prediction of 30-day hospital readmissions. Methods of Information in Medicine 2019;58(6):213-221. [doi: 10.1055/s-0040-1702159] [Medline: 32349155]
36. CXR8. NIH Clinical Center Box. URL: https://nihcc.app.box.com/v/ChestXray-NIHCC [accessed 2024-12-02]

Abbreviations

AUC: area under the curve
IRB: institutional review board
PCA: principal component analysis
KPCA: kernel principal component analysis
XGBoost: extreme gradient boosting

XSL•FO
RenderX

Edited by A Mavragani; submitted 24.10.23; peer-reviewed by N Silva, M Hirota; comments to author 05.12.23; revised version received 31.05.24; accepted 22.06.24; published 22.01.25.

Please cite as:
Martinez S, Ramirez-Tamayo C, Akhter Faruqui SH, Clark K, Alaeddini A, Czarnek N, Aggarwal A, Emamzadeh S, Mock JR, Golob EJ
Discrimination of Radiologists' Experience Level Using Eye-Tracking Technology and Machine Learning: Case Study
JMIR Form Res 2025;9:e53928
URL: https://formative.jmir.org/2025/1/e53928
doi:10.2196/53928
PMID:

©Stanford Martinez, Carolina Ramirez-Tamayo, Syed Hasib Akhter Faruqui, Kal Clark, Adel Alaeddini, Nicholas Czarnek, Aarushi Aggarwal, Sahra Emamzadeh, Jeffrey R Mock, Edward J Golob. Originally published in JMIR Formative Research (https://formative.jmir.org), 22.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Use of Go-Beyond as a Self-Directed Internet-Based Program Supporting Veterans' Transition to Civilian Life: Preliminary Usability Study

Karolina Katarzyna Alichniewicz[1,2], PhD; Sarah Hampton[1,3], DPsych(Clin); Madeline Romaniuk[2], DPsych(Clin); Darcy Bennett[1], BPsych(Hons); Camila Guindalini[1], PhD

[1]Greenslopes Private Hospital, Gallipoli Medical Research, Brisbane, Australia

[2]The University of Queensland, Faculty of Health and Behavioural Sciences, Brisbane, Australia

[3]School of Medicine, The University of Queensland, Brisbane, Australia

Corresponding Author:
Karolina Katarzyna Alichniewicz, PhD
Greenslopes Private Hospital
Gallipoli Medical Research
Newdegate Street
Greenslopes
Brisbane, QLD 4120
Australia
Phone: 61 7 33947613
Fax: 61 7 33297744
Email: alichniewiczkarolina@gallipoliresearch.org.au

Abstract

Background: The transition from military service to civilian life presents a variety of challenges for veterans, influenced by individual factors such as premilitary life, length of service, and deployment history. Mental health issues, physical injuries, difficulties in relationships, and identity loss compound the reintegration process. To address these challenges, various face-to-face and internet-based programs are available yet underused. This paper presents the preliminary evaluation of "Go-Beyond, Navigating Life Beyond Service," an internet-based psychoeducational program for veterans.

Objective: The study aims to identify the reach, adoption, and engagement with the program and to generate future recommendations to enhance its overall impact.

Methods: This study exclusively used data that were automatically and routinely collected from the start of the Go-Beyond program's launch on May 24, 2021, until May 7, 2023. When accessing the Go-Beyond website, veterans were asked to complete the Military-Civilian Adjustment and Reintegration Measure (M-CARM) questionnaire, which produces a unique M-CARM profile of results specifying potential areas of need on the 5 domains of the measure. Users were then automatically allocated to Go-Beyond modules that aligned with their M-CARM profile. Additionally, quantitative and qualitative data were collected from a survey on aesthetics, interactivity, user journey, and user experience, which was optional for users to complete at the end of each module.

Results: Results show a conversion rate of 28.5% (273/959) from the M-CARM survey to the Go-Beyond program. This rate is notably higher compared with similar internet-based self-help programs, such as VetChange (1033/22,087, 4.7%) and resources for gambling behavior (5652/8083, 14%), but lower than the MoodGYM program (82,159/194,840, 42.2%). However, these comparisons should be interpreted with caution due to the limited availability of published conversion rates and varying definitions of uptake and adoption across studies. Additionally, individuals were 1.64 (95% CI 1.17-2.28) more likely to enroll when they express a need in Purpose and Connection, and they were 1.50 (95% CI 1.06-2.18) times more likely to enroll when they express the need Beliefs About Civilians, compared with those without these needs. The overall completion rate for the program was 31% (85/273) and modules' individual completion rates varied from 8.4% (17/203) to 20% (41/206). Feedback survey revealed high overall user satisfaction with Go-Beyond, emphasizing its engaging content and user-friendly modules. Notably, 94% (88/94) of survey respondents indicated they would recommend the program to other veterans, family, or friends.

Conclusions: The Go-Beyond program may offer promising support for veterans transitioning to civilian life through digital technology. Our study reveals insights on user engagement and adoption, emphasizing the need for ongoing evaluation to further

XSL•FO
RenderX

address the diverse needs of military personnel. Future research should explore predictors of engagement, the addition of peer or facilitator support, and the use of outcome measures for effectiveness assessment.

(JMIR Form Res 2025;9:e60868) doi:10.2196/60868

KEYWORDS

military transition; web-based interventions; military-civilian adjustment; Go-Beyond; internet-based program; civilian; military service; veteran; premilitary life; mental health issues; physical injuries; adoption; quantitative analysis; survey; family; support; digital technology; user engagement; effectiveness; assessment

Introduction

The transition from military service to civilian life is a multifaceted process that can present challenges for veterans [1-3]. Upon leaving the military, veterans often face a period of adjustment as they navigate a vastly different sociocultural and professional landscape [4-6]. The process of transitioning from military service to civilian life is not uniform, as each veteran's experience is shaped by individual factors such as premilitary life, length of service, deployment history, and personal resilience [7]. While many individuals successfully navigate this change, a substantial proportion experience difficulty in adjusting to civilian life, with these difficulties persisting over the years [1,2,4].

In the literature, there is recognition of various challenges faced by military personnel when they transition into civilian life. Among these challenges, mental health issues such as posttraumatic stress disorder (PTSD), substance use disorders, depression, and anxiety are frequently reported [1]. In addition, physical injuries, including chronic pain and traumatic brain injuries, play a significant role [8]. Furthermore, difficulties in establishing and maintaining friendships and family relationships [4,5], loss of identity and purpose [2,3,9], disparities between military and civilian cultures, as well as employment factors [6] are all critical elements. Collectively, these multiple and dynamic factors have a significant impact on the reintegration of veterans into civilian life. It is important to note that the presence of hardships in various domains can compound the effects and exacerbate functional impairments beyond those caused by individual stressors alone [10]. Consequently, understanding the factors that influence transition and identifying effective accessible interventions is critical in supporting military personnel during this crucial phase.

Internet-based approaches can complement and expand the support systems available to service members and veterans [11-13]. By harnessing the potential of digital technology, personalized and accessible interventions that cater to the diverse needs of military personnel during their transition to civilian life can be provided. Moreover, internet-based interventions could present an effective way to overcome some of the barriers to care, such as long waiting periods, stigma, or lack of access to treatment due to remote location [12]. Internet-based interventions hold significant potential in supporting veterans during the critical transitioning phase. However, ensuring their uptake and effectiveness requires robust evaluations, covering aspects like utility, user satisfaction, usability, acceptability, and potential in addressing the diverse needs of transitioning military personnel [14]. Through such evaluations, policy makers, researchers, and support organizations can gain valuable insights into the most beneficial strategies for veterans [15]. Furthermore, a comprehensive assessment of internet-based interventions may enable tailored interventions, improvement of support mechanisms, and the development of evidence-based practices [16]. Therefore, understanding the impact of internet-based interventions is paramount in enhancing the well-being and successful integration of military veterans into civilian society.

In this paper, we present the preliminary evaluation of an internet-based psychoeducational program "Go-Beyond, Navigating Life Beyond Service" designed for veterans who are in the process of reintegration to civilian life. The aim of the study was to perform a user-centered evaluation of the Go-Beyond program using user engagement data routinely collected and assess the reach, adoption, engagement, and completion rates. Furthermore, the study aimed to gather user feedback to enhance the concept version of the program's modules, identify opportunities for further improvement, and generate future recommendations.

Methods

Program Description

"Go-Beyond, Navigating Life Beyond Service" is an evidence-informed, self-directed, web-based program designed to support veterans' adjustment to civilian life after separation from military service. The program was developed by Gallipoli Medical Research (GMR), a not-for-profit organization that has the mission to improve the quality of life for veterans and their families through research, translation, and implementation, in partnership with Returned and Services League of Australia Queensland (RSLQ), and was officially launched on May 24, 2021. Over the 24 months after the launch, all interested veterans were provided unlimited access to the Go-Beyond program. Former service members were made aware of the Go-Beyond program through a national implementation task force initiative that included flyers, posters, presentations, and a social media campaign and was conducted by GMR, RSLQ, Department of Veteran Affairs, Open Arms, and Commonwealth Super Corporation. The program was also promoted to various ex-serving organizations. Furthermore, Go-Beyond was accompanied by an applied practitioner portal that provided training on the use of the program and targeted clinicians, especially rehab providers. The implementation of the Go-Beyond program was part of a larger research agreement with RSLQ, and the associated costs for its development and preliminary implementation have been covered by the funding organization. This funding covered essential elements such as

XSL•FO
RenderX

platform design, content creation, technical support, and user testing to ensure the program's accessibility for veterans. In addition, future strategies are being developed to ensure the program's sustainability beyond the funding period in Australia.

When accessing the Go-Beyond website, veterans are asked to complete the Military-Civilian Adjustment and Reintegration Measure (M-CARM) questionnaire, which produces a unique M-CARM profile of results specifying potential areas of need on the 5 factors of the measure. The M-CARM is a valid and reliable self-report questionnaire designed to assess adjustment and reintegration to civilian life after service. It contains 21 questions, rated from 1 to 5, and takes approximately 5 minutes to complete. The measure is freely available on a dedicated website [17]. Users are automatically allocated to Go-Beyond modules that align with their M-CARM profile.

The Go-Beyond program aims to (1) provide education about topics that are important for psychological adjustment and reintegration to civilian life; (2) build psychological insight into personal challenges adjusting to civilian life; (3) offer practical first steps and tools to facilitate stronger adjustment and reintegration; and (4) provide links to additional services and support available in Australia.

The Go-Beyond program was developed by 2 clinical psychologists at GMR who have both research and clinical experience with ex-service personnel in Australia. The content of the program is based on previous research describing empirically derived psychological and cultural factors associated with adjustment and reintegration to civilian life, including (1) Purpose and Connection, (2) Help Seeking, (3) Beliefs About Civilians, (4) Resentment and Regret, and (5) Regimentation [18]. These factors were derived as part of the development of the M-CARM [18]. Go-Beyond consists of seven interactive modules entitled (1) Finding Purpose, (2) Help Seeking, (3) Regimentation, (4) Resentment and Regret, (5) Social Connection, (6) Beliefs About Civilians, and (7) an optional module available to anyone who signed up to Go-Beyond: (7) Effective Communication in Civilian Life. They are presented to participants in the above order. If a participant is not required to do a particular module, it gets omitted in that order. The session outline of the program is presented in Textbox 1. The program content uses evidence-based psychological therapeutic approaches including cognitive behavioral therapy and acceptance and commitment therapy. Both cognitive behavioral therapy and acceptance and commitment therapy have demonstrated transdiagnostic treatment efficacy within veteran populations [19]. Each module includes (1) psychoeducation for each topic; (2) hypothetical case studies and personal, video-recorded stories from Australian ex-service members; (3) personal reflection opportunities or written tasks; (4) helpful activities; and (5) direction to more specific services and resources for Australian ex-service personnel. The approximate time needed to complete each module is 45-60 minutes.

XSL•FO
RenderX

Textbox 1. Descriptions of Go-Beyond modules.

Beliefs About Civilians

- The module introduces the concept of beliefs, social identity theory, and common unhelpful beliefs about civilians that can hinder adjustment to civilian life. Participants engage in self-reflection to evaluate their beliefs and learn techniques to adapt them, including thought challenging, defusing exercises, and a "take your thought to court" activity. The importance of practicing these techniques is emphasized throughout.

Finding Purpose

- The module begins by exploring the importance of purpose, particularly in the military context, and the sense of loss that can occur during the transition to civilian life. It helps participants clarify their values, distinguish between values and goals, and engage in activities like the life compass, values-guided actions, and goal setting. The module also emphasizes finding purpose through volunteering or paid employment and provides information on services and resources that support veterans in securing meaningful work.

Help Seeking

- The module provides psychoeducation on the importance of help-seeking for mental health, addressing common barriers and highlighting the rates of help-seeking among Australian veterans. It covers how to find evidence-based treatments and recognize when help is needed. Through the "Staying Well" activity, participants learn to identify indicators of good mental health, and signs of decline, and create an action plan. Practical steps for accessing high-quality mental health support are also provided.

Social Connections

- This module covers the importance of social connections for health and well-being, focusing on their role in adjusting to transitions, especially for the "Defence family." It addresses common barriers such as cultural differences, misconceptions, and communication issues. Strategies for finding mateship outside the military, strengthening existing ties, and building new ones are discussed, alongside opportunities for social connection and volunteering.

Resentment and Regret

- This module explores resentment, covering its definition, functions, and impacts, and includes activities to identify underlying feelings, use self-compassion, and resolve related injustices. It also addresses regret, discussing its nature, impacts, and irrationality, with activities to manage regret by confronting hindsight bias, accepting imperfections, and clarifying values and goals.

Regimentation

- This module covers regimentation; its role in Defence; and its effects on personal life, family, and work. It includes activities to identify and evaluate regimentation's impact and emphasizes the importance of psychological flexibility and adaptability. Practical strategies are provided for embracing change, trying new things, and adjusting routines and thought processes.

Effective Communication in Civilian Life

- This module covers communication styles (aggressive, passive, passive-aggressive, and assertive) and includes a quiz to identify your style. It provides skills for meaningful conversations, including asking questions, self-disclosure, managing stereotypes, active listening, and overcoming barriers. The module also explores using compliments and empathy to build social connections.

Data Collection

This study exclusively used data that were automatically and routinely collected starting from the program's launch on May 24, 2021, until May 7, 2023, through websites [17] for M-CARM and [20] for Go-Beyond. Both websites were freely available to everyone, both veterans and civilian populations, in Australia. No specific inclusion or exclusion criteria were applied to this study. The reach of Go-Beyond was defined as the number of users who accessed and completed the M-CARM measure and the adoption was defined as the number of users who signed up to the Go-Beyond program. Completion was measured by the number of users who accessed and reached the end of the module, as tracked by the Absorb LMS system (Absorb Software Inc). The system records whether a user has started, is in progress, and has fully completed the module, including exiting the module properly. The modules were designed to be completed in 45-60 minutes.

In addition, quantitative and qualitative data were collected from a survey on aesthetics, interactivity, user journey, and user experience, which was optional for users to complete at the end of each module. The questions asked can be found in Table 1.

XSL•FO
RenderX

Table 1. User experience survey.

Question	Rating options
How did you find the duration of the module?	Option for comments
How easy was the Go-Beyond Platform to use? Were you able to navigate to what you needed? Please provide details.	Option for comments
How likely are you to implement changes in your life following this module?	Likert scale 1-5 • 1-Very Unlikely • 2-Unlikely • 3-Likely • 4-Very Likely • 5-Extremely Likely
If yes (likely as above), what changes would you make? If no (unlikely or very unlikely), why not?	Option for comments
How many minutes did it take for you to complete this module?	Option for comments
How useful did you find this module?	Likert scale 1-5 • 1-Completely dissatisfied • 2-Dissatisfied • 3-Neutral • 4-Satisfied • 5-Completely satisfied
How would you rate the aesthetics (layout, graphics, or visuals)?	Likert scale 1-5 • 1-Completely dissatisfied • 2-Dissatisfied • 3-Neutral • 4-Satisfied • 5-Completely satisfied
How would you rate the gestural design (interactions, taps, swipes, and scrolling) functions?	Likert scale 1-5 • 1-Completely dissatisfied • 2-Dissatisfied • 3-Neutral • 4-Satisfied • 5-Completely satisfied
How would you rate the performance of the module (features and components)?	Likert scale 1-5 • 1-Completely dissatisfied • 2-Dissatisfied • 3-Neutral • 4-Satisfied • 5-Completely satisfied
How would you rate the quality of content?	Likert scale 1-5 • 1-Completely dissatisfied • 2-Dissatisfied • 3-Neutral • 4-Satisfied • 5-Completely satisfied
How would you rate the usability or navigation of this module?	Likert scale 1-5 • 1-Completely dissatisfied • 2-Dissatisfied • 3-Neutral • 4-Satisfied • 5-Completely satisfied

XSL•FO
RenderX

Question	Rating options
Overall how engaging did you find the module (entertainment, interest, relatability, and interactivity)?	Likert scale 1-5 • 1-Completely dissatisfied • 2-Dissatisfied • 3-Neutral • 4-Satisfied • 5-Completely satisfied
Overall what would you rate this module?	Likert scale 1-5 • 1- Poor • 2-Satisfactory • 3-Good • 4-Very Good • 5-Excellent
What device did you use to access the module (desktop, mobile, or tablet)?	Options to choose from: • Laptop • Mobile • Tablet • Desktop • Combination
Would you recommend this module to other veterans or family or friends of a veteran?	Options to choose from: Yes, No
Did the audio and visual presentations resonate with you and your personal experiences?	Options to choose from: No, Somewhat, Yes, No comment
How did you find the use of personal stories, videos, animations, and illustrations?	Option for comments
Please make any further comments regarding the engagement and communication of the module.	Option for comments
Please provide any additional comments on the functionality of the module.	Option for comments
Was the visual information accurate (concepts are logical and correct)? Please provide any comments.	Option for comments
What further recommendations or comments would you make to improve the Go-Beyond Platform and the module?	Option for comments
What was the least helpful part of the module?	Option for comments
What was the most helpful part of the module?	Option for comments

Statistical Analysis

To determine the reach and adoption of the Go-Beyond program, a user-centered evaluation was performed using information routinely collected to monitor and improve the quality of the program. The system engagement data were automatically collected from users who accessed the M-CARM and Go-Beyond websites on a voluntary basis.

The data analysis was conducted using R (R Core Team) [21]. Descriptive statistics were used to gain insights into the distribution and central tendencies of the collected data. Subsequently, logistic regression analysis was used to discern factors influencing users' likelihood to progress to Go-Beyond following completion of the M-CARM.

To analyze the qualitative data gathered through the voluntary online surveys, a systematic approach was used. Thematic analysis was used to recognize patterns and identify core themes. Themes were extracted from the data and grouped under the questions posed adhering to the thematic analysis approach by Braun and Clarke [22]. The initial coding of responses was conducted manually by the primary researcher, who independently reviewed the data to develop preliminary codes. Responses were coded into categories and were organized thematically, allowing for the identification of recurring topics. These codes were then reviewed and refined through consensus discussions among the research team, ensuring that differing perspectives were considered. The researchers acknowledge that their perspectives and experiences may influence their interpretations, and they have taken steps to reflect on and mitigate potential biases throughout the research process. The positionality of the researchers was acknowledged throughout the analysis process, with the team reflecting on how their backgrounds, experiences, and potential biases might influence the interpretation of the data and this qualitative analysis [23]. All researchers were committed to evaluating a web-based intervention aiming for it to be not only useful but also feasible and acceptable to the intended users. The primary researcher brings over 10 years of experience working with clients from military backgrounds, providing a unique perspective informed by extensive professional practice. However, it is important to note that the primary and senior researchers were not involved in the creation or implementation of the version of the program under study, positioning them as outsiders in this context. This

XSL•FO
RenderX

outsider status allows for a more detached analysis, free from potential biases related to program design and execution. Consensus discussions among the research team also ensured that individual biases were minimized and enhanced the credibility of the interpretations. Representative quotations and exemplars were extracted to illustrate each identified theme.

Ethical Considerations

The Department of Defense and Veterans' Affairs Human Research Ethics Committee has deemed this study to be a quality assurance activity that upholds the principles of the National Statement on Ethical Conduct in Human Research and the Ethical Consideration in Quality Assurance and Evaluation Activities (approval number 516-23). In the privacy policy statement that participants have access to when deciding to sign in to the program, they are informed about the organization's commitment to comply with the Privacy Act 1988 (Cth) (Privacy Act), including the Australian Privacy Principles. The data were extracted from the Absorb LMS system by the LMS Administrator and deidentified before they were provided to the research team. After data cleaning and integration, analysis was conducted by the research team on deidentified data. All collected data were held confidential at all times, in a password-protected database, on a server that requires a unique log-in and password by all team members. Only the project team had access to the data. No compensation was provided to users as this service evaluation used routinely collected data.

Results

Go-Beyond User and M-CARM–Only User Profile

Out of the total of 959 users who voluntarily completed M-CARM, there were 686 who did not proceed to sign up for the Go-Beyond program (M-CARM–only users) and 273 who signed up for Go-Beyond, resulting in a conversion rate of 28.5% (273/959) and a preprogram dropout rate of 71.5% (686/959). Table 2 summarizes the demographic characteristics of both user groups. Only upon signing up for the Go-Beyond program, the data regarding time since discharge from the Australian Defence Force were recorded. Out of 273 Go-Beyond users, 215 were discharged from the Australian Defence Force and 43 reported that they were still serving. The median number of years since transition for Go-Beyond users who were discharged equaled 2.74 years, with a large proportion of users (165/273, 60.5 %) reporting being within 0-5 years from discharge.

XSL•FO
RenderX

Table 2. Demographic characteristics of the Go-Beyond users and Military-Civilian Adjustment and Reintegration Measure (M-CARM)–only users.

Variables	Go-Beyond users (n=273)	M-CARM–only users (n=686)
Sex, n (%)		
Male	192 (70.3)	498 (72.6)
Female	67 (25)	178 (26)
Prefer not to answer	1 (0.2)	10 (2)
N/A[a]	13 (4.8)	(0)
Age range (years), n (%)		
18-25	20 (7.3)	74 (11)
26-35	65 (24)	187 (27.3)
36-45	59 (21)	162 (23.6)
46-55	73 (26.7)	144 (21)
56-65	32 (12)	93 (14)
66-75	8 (3)	15 (2)
>75	2 (1)	6 (1)
Prefer not to answer	1 (0.2)	5 (1)
N/A	13 (4.8)	(0)
Service, n (%)		
Army	153 (56)	384 (56)
Navy	48 (18)	96 (14)
Airforce	31 (11)	117 (17)
Prefer not to answer	24 (9)	77 (11)
Army+Airforce	2 (1)	3 (0.4)
Army+Navy	1 (0.2)	1 (0.2)
Navy+Airforce	0 (0)	2 (0.3)
N/A	14 (5.1)	0 (0)
Type of needs as shown on M-CARM, n (%)[b]		
Beliefs About Civilians	197 (75.5)	441 (64.3)
Regimentation	182 (69.7)	452 (65.9)
Resentment and Regret	176 (67.4)	396 (57.7)
Purpose and Connection	146 (55.9)	287 (41.8)
Help Seeking	76 (29.1)	184 (26.8)

[a]N/A: not applicable.

[b]One person can have multiple needs.

There were no statistically significant differences observed between the Go-Beyond and M-CARM–only user groups concerning age, sex, or service type (all $P>.05$). However, a significantly higher number of individuals in the Go-Beyond user group were identified as having needs in the areas of Purpose and Connection ($\chi^2_4=14.59$, $P<.001$), Resentment and Regret ($\chi^2_4=7.05$, $P=.008$), and Beliefs About Civilians ($\chi^2_4=10.273$, $P=.001$) compared with the M-CARM–only users.

A logistic regression was then performed to assess the effects of M-CARM outcomes and service type on the likelihood of enrolling in the Go-Beyond program. Participant's age, sex, and service type were controlled for. The outcomes of logistic regression showed that participants with the need for Purpose and Connection and Beliefs About Civilians were found to be significantly more likely to sign up for Go-Beyond ($P=.003$ and $P=.03$ respectively). Individuals expressing a need for Purpose and Connection were approximately 1.64 times more likely to enroll (odds ratio 1.64, 95% CI 1.17-2.28) compared with those who did not. Similarly, those with the need Beliefs About Civilians were 1.50 times more likely to enroll (odds ratio 1.50, 95% CI 1.06-2.18) compared with those without this need. Individuals who were identified with a need for Resentment and Regret were not found to be significantly more likely to sign up for Go-Beyond ($P>.05$).

XSL•FO
RenderX

Go-Beyond User Engagement

Out of 273 users who signed up for Go-Beyond, 85 (31%) users completed at least 1 assigned module. The data on the enrollment status at the time of the data extraction for each of the modules and corresponding completion rates can be found in Table 3. Out of participants who completed at least 1 module, nearly all users who had been assigned 1 module, completed it (36/38, 94% users). Among users who were assigned 2 modules, 37% (14/38) have completed them both. For those who had been assigned 3 modules, 26% (9/34) have completed all 3. Of users who had been assigned 4 modules, 10% (3/31) have completed them, and of users who had been assigned 5 modules, 3% (2/63) have completed all 5. A total of 22% (10/45) of users who had assigned 6 modules have completed all 6. Overall, 3% (8/247) of participants who had been assigned other modules also completed the Communication module. Out of Go-Beyond users, who did not have any modules assigned, but still accessed the program (26 users), 4 (15%) users completed 1 module, 1 (4%) user completed 3 modules, 2 (8%) users completed 4 modules, 1 (4%) user completed all available modules, and 3 (12%) completed the Communication module. The module with the highest completion rate was Finding Purpose with 41 (20%) out of 206 users who signed up to this module, followed by Beliefs About Civilians with 48 (19.7%) out of 244 who signed up, Resentment and Regret with 38 (17.6%) out of 216 who signed up, Regimentation with 33 (16%) out of 205 who signed up, Help Seeking with 21 (15.5%) out of 156 who signed up, and Social Connection with 17 (8.4%) out of 220 signed up users. The lowest completion rate was the additional module Effective Communication in Civilian Life, in which everyone was enrolled automatically, with 11 (4%) out of 273 users who signed up.

Table 3. Enrollment status for each Go-Beyond module at the time of data extraction.

Module and status	Not started, n (%)	In progress, n (%)	Complete, n (%)
Beliefs About Civilians (n=244)	151 (61.9)	45 (18.4)	48 (19.7)
Finding Purpose (n=206)	128 (62)	37 (18)	41 (20)
Resentment and Regret (n=216)	163 (75.5)	15 (6.9)	38 (17.6)
Regimentation (n=205)	166 (81)	6 (3)	33 (16)
Help Seeking (n=156)	110 (81.5)	4 (3)	21 (15.5)
Social Connection (n=220)	175 (86)	11 (5.4)	17 (8.4)
Communication[a] (n=273)	262 (96)	0 (0)	11 (4)

[a]This module is available to everyone as an additional module regardless of their Military-Civilian Adjustment and Reintegration Measure profile.

The "time spent" refers to the time participants spent engaging with each module. This metric could reflect multiple factors, including module complexity, participant interest, or difficulty with the content. The module with the highest median time spent by the users who did complete it was Beliefs About Civilians (67, IQR 23.4-110.6 minutes). It was followed by Finding Purpose (median time of 32, IQR 9-55 minutes), Regimentation (median time of 25, IQR 11-39 minutes), Resentment and Regret (median time of 17, IQR 0.25-33.75 minutes), Social Connection (median time of 13, IQR 5.5-20.5 minutes) and Help Seeking (median time of 11, IQR 0-31.5 minutes). The median time for the additional module Effective Communication in Civilian Life was 19 (IQR 10.75-27.25 minutes) minutes.

No statistically significant differences were observed (all $P>.05$) in relation to age, sex, service type, or type of needs (Purpose and Connection, Help Seeking, Regimentation, Resentment and Regret, and Beliefs About Civilians) between participants who had completed at least 1 module and those who had not started or were still in progress at the time of data extraction.

Go-Beyond User Experience Survey Data

There were 94 self-report surveys assessing the Go-Beyond internet-based program completed. The survey responses provided insights regarding the program's subjective usefulness, user engagement, the platform's usability, and user recommendations.

Perceived Usefulness and Satisfaction

Out of the 94 internet-based self-report surveys evaluating Go-Beyond, 80% (74/94) of respondents acknowledged the usefulness of the module. Moreover, 85% (80/94) of the users were "satisfied" or "completely satisfied" with the program's features and components, and 79% (73/94) of survey respondents stated that they were "satisfied" or "completely satisfied" with the quality of the content. Furthermore, 86% (81/94) of survey respondents indicated a likelihood of implementing changes in their lives following the module, and 94% (88/94) reported they would recommend the program to other veterans or their family or friends.

User Assessment of Platforms Engaging Features

The users were asked how engaging overall the modules were in terms of entertainment, interest, relatability, and interactivity of the modules (see Textbox 1 for questions). A total of 78% (73/94) of the survey respondents found the modules engaging overall. In addition, 86% (81/94) of participants found the time required to complete the program to be satisfactory.

Platform Usability

A total of 91% (86/94) of users reported a high degree of satisfaction with the Go-Beyond Platform's usability. The audio and visual presentations resonated or somewhat resonated with 94% (88/94) of users. In relation to the use of personal stories, videos, animations, and illustrations, users reported that they were: "Very helpful and informative" and commented that they

XSL•FO
RenderX

enhanced the program: "I thought the quotes, illustrations, and videos were used really well in this module and very realistic." And helped them self-reflect: "It was good. Made me look at myself a little and accept that it is ok to be flexible." The survey indicated strong endorsements of the platform's aesthetics, rated as "satisfied" and completely satisfied" by 86% (81/94) of users. Furthermore, 85% (80/94) of users were "satisfied" or "completely satisfied" with gestural design functions of the platform.

User Recommendations

When questioned about the most beneficial features of the program, the majority of responses were related to (1) relatable psychoeducational content (ie, "The information behind emotions and why we feel them" and "The personal insights it provided along with leaving me with a sense of not being alone in feeling this way"), (2) opportunities for self-reflection (ie, "Given the option to self-reflect" and "Providing places to self-reflect on the information given"), and (3) practical strategies and techniques (ie, "strategies and thought techniques to widen perspectives and assist with positive actions"). On the other hand, the least helpful features included (1) a long reference list and links to resources (ie, "The links end up with more links that take you further and further from the main content"), (2) a lack of diversity in the personal videos (ie, "I don't recall any useful or memorable audio clips with information or personal stories, although I can appreciate that it is not simple to source authentic, vulnerable, and engaging stories"). Furthermore, some technical issues were also identified (ie, "No previous work was recorded. I had to wait to I got access to a printer to complete the module and all my previous work had not been saved. Having a complete module button that takes you back to the home page would be great," "Return to other courses button when completing the module would be great," and "Better audio for the doctor videos. Very echo'y").

Discussion

Principal Findings

The study aimed to assess the reach, adoption, and engagement with the "Go-Beyond, Navigating Life Beyond Service Program," a web-based program designed to aid veterans in transitioning to civilian life, incorporating evidence-informed content on psychological and cultural factors. The program aims to provide education, build psychological insight, offer practical tools, and offer users links to available support. The findings underscore the potential value of the Go-Beyond program as a web-based initiative, leveraging digital technology to provide a personalized and easily accessible resource for veterans navigating the transition to civilian life.

Most of the Go-Beyond users were male (192/273, 70.3%), within 46-55 years of age, and served in the Army (153/273, 56%). Median years since the transition for Go-Beyond users who were discharged equaled 2.74 years, with a large proportion of users (165/273, 60.5 %) reporting being within 0-5 years from discharge. It is noteworthy that the first few years after discharge are widely recognized as an extremely vulnerable period for ex-serving members. Suicide risk, in particular, is notably high during the initial year of transitioning to civilian life [24]. In addition, evidence in the literature points out that many individuals experience reintegration difficulties within 6 years after separation [25], underscoring the importance of the program during this critical phase. Furthermore, the analysis of users' needs highlights the need for a program that can address a wide array of transitional challenges and provide essential support during the transition process [18].

The overall conversion rate from completing the M-CARM questionnaire to signing up for the Go-Beyond program was 28.5% (273/959). Interestingly, among the analyzed factors, the study identified that having needs in Beliefs About Civilians and Purpose and Connection domains could be considered relevant predictors of engagement with the program. Users identified with these needs were up to almost 2 times more likely to sign up for Go-Beyond after completing the M-CARM survey. Interestingly, aligning with these findings, the module Finding Purpose had the highest completion rate (41/206, 20%), followed by Beliefs About Civilians (48/244, 19.7%). There is a limited number of studies providing conversion rates for internet-based, self-help programs for mental health, with the definition of uptake or adoption varying widely based on program structure and objectives [26]. A study assessing the public implementation of a web-based program for veterans dealing with risky alcohol use and PTSD, named VetChange, reported a conversion rate of 4.7% (1033/22,087) from unique homepage visits resulting in full account registrations [27]. Another study examining the impact of self-help resources on gambling behavior, found a conversion rate of 14% (5652/8083) from the program registration page to create an account. On the other hand, a study evaluating the MoodGYM program, which aims to prevent and manage symptoms of depression and anxiety, demonstrated a 42.2% (82,159/194,840) registration rate, considering the number of website visitors. Although further research is necessary to thoroughly assess ongoing participation after the initial sign-up to Go-Beyond, the conversion rate found in this study appears to be mostly higher than comparable published programs and indicates a decent level of interest and engagement among veterans.

The time spent on each module of the Go-Beyond Program indicates varying degrees of user engagement, with a notable proportion (31%) completing at least 1 assigned module. These outcomes also correspond to the findings of other studies [27]. A systematic review of digital self-help interventions for depression, low mood, or anxiety demonstrated that real-world uptake and engagement present a significant variability and may differ from adherence in controlled trials [27]. Across the studies analyzed in that research, minimal usage of the intervention ranged from 21% to 88%, with users either accessing it at least once or completing one module. However, completion or sustained use, defined as finishing all modules or the final assessment, or continuing to use the apps after a prolonged period of time was lower and ranged from 0.5% to 28.6%. Research has demonstrated that factors such as motivation, previous experience, and the perceived relevance of course content can significantly influence engagement and completion rates in online interventions [28]. Jardine et al [29] identified uncertainty about the intervention's usefulness and usability as the main barriers to its uptake. Previous studies also

XSL•FO
RenderX

revealed that forgetting about the intervention, not finding time for it, and not finding it useful posed main barriers to early engagement [28]. Further barriers to engagement included technological challenges, severe mental health issues, lack of personalization [30], lack of privacy, or difficulty in relating to the content [30,31]. Conversely, facilitators include among others user-friendly interfaces, personalized content, social connectedness facilitated by the intervention, and the perception of gaining better insight into the condition [30,31].

Beatty and Binnion [31] examined 36 studies to determine predictors for adherence to web-based interventions, such as demographic, psychological, characteristics of the presenting problem, and intervention and computer-related factors. The findings indicated that the female sex, higher treatment expectancy, sufficient time, and personalized intervention content were associated with increased adherence. Further, not having enough time, dissatisfaction with program content, perceiving content as impersonal, and computer difficulties were found to decrease adherence [32]. Sex was not found to play a significant role in predicting the completion of any module in the Go-Beyond program in our study. However, as identified in the users' feedback, technical challenges experienced by Go-Beyond users, such as difficulties with saving completed worksheets, accessing the homepage with all relevant modules, and audio quality, may have influenced veterans' frustration levels and their willingness or ability to complete the modules. On the other hand, the highly personalized content of the program, which offers users modules tailored to their identified M-CARM areas of need, may have contributed to the completion rates and engagement with the module evidenced by high ratings of quality of content and usefulness of the program. Moreover, the results of the user survey underscore overall user satisfaction with the Go-Beyond program, encompassing various aspects of module effectiveness, user engagement, platform usability, and user recommendations. They suggest that the Go-Beyond internet-based program engages users, providing valuable content with aesthetically pleasing and functionally efficient modules. Users also express a high likelihood of recommending the program to others, indicating a potential helpful impact on a broader audience. However, further studies evaluating specific barriers and facilitators for engagement with Go-Beyond are necessary.

While considering factors impacting adherence, it needs to be noted that Wojtowicz et al [32] suggest that levels of guidance or support through phone or email contribute to increased adherence to internet-based interventions. It is suggested that guided support may enhance motivation to participate or foster a greater sense of accountability to adhere [33]. Peer support and professional-assisted delivery of web-based interventions also pose a potential solution to the enhancement of engagement in web-based interventions [34]. Research suggests that peer-assisted internet-based interventions have proven to be feasible in the veteran population [35,36]. Since Go-Beyond was designed as a completely self-help tool, an entirely self-help format may be challenging for the ex-Defence cohort where low motivation and avoidance are key features in common mental health disorders experienced by veterans such as depression, anxiety, and PTSD [37-39].

The findings of this preliminary evaluation provide several lessons learned from the preliminary implementation and lay the groundwork for further investigation and refinement of the Go-Beyond program. In this context, it is crucial to emphasize the importance of using real-world data when evaluating adoption and engagement in digital mental health interventions. As previously highlighted, real-world settings often differ from controlled trial conditions, as individuals' interaction with these interventions can be influenced by factors such as accessibility, usability, and personal preferences. Therefore, comparing uptake and usage patterns across various interventions using real-world data could offer invaluable insights, enabling researchers and developers to identify trends, successes, and areas for improvement in the field of digital mental health. This, in turn, would facilitate making informed decisions to enhance user experience and outcomes [26].

Limitations and Future Directions

There are several limitations to this study. Veterans who chose to participate in the Go-Beyond program may not represent the entire veteran population, leading to selection bias. Since the M-CARM measure as well as the Go-Beyond program are available to anyone, there is a possibility that some of the users might have been members of the general population interested in the program. Those who completed the M-CARM and signed up for the Go-Beyond might have unique characteristics or experiences that differ from nonparticipants, impacting the generalizability of the findings. Therefore, the existing literature is the only source of information to draw conclusions as to why some chose to enroll in the program and complete assigned modules. Furthermore, only 31% of users who signed up for Go-Beyond completed at least 1 assigned module and provided feedback. As a result, the favorable quantitative and qualitative feedback received regarding the program may not fully represent the overall user experience. This highlights the importance of considering the perspectives of those who did not fully engage with or did not complete the program in future evaluations another limitation of our study is related to the interpretation of the "time spent" metric. While "time spent" refers to the duration participants engaged with each module, this measure does not account for various underlying factors that could influence the observed time. For instance, a longer duration may indicate that a module was more complex, more engaging, or potentially more challenging for participants. This variability can complicate the interpretation of whether differences in "time spent" are due to the inherent characteristics of the modules or other participant-specific factors.

Finally, this was a preliminary examination of the reach, user engagement, and completion rates of the Go-Beyond program, and further evaluations of the barriers and facilitators, as well as the effectiveness of the internet-based program on veterans' readjustment to civilian life are needed.

Conclusion

In conclusion, the Go-Beyond program may be a promising initiative in leveraging digital technology to support veterans during their transition to civilian life. The study's outcomes contribute valuable insights into user engagement, completion rates, and the program's overall adoption. As the program

XSL•FO
RenderX

continues to evolve, ongoing evaluation and refinement will be crucial to enhancing its effectiveness and addressing the diverse needs of military personnel in their reintegration journey. Our study supports the notion that web-based interventions hold substantial promise in enhancing the experience of transition into the civilian life of veterans. Future research should examine factors that predict the use and nonuse of Go-Beyond, how use changes over time, and what factors can serve as predictors of use. Furthermore, it is crucial to incorporate relevant outcome measures to gain a thorough understanding of its effectiveness and potential benefits.

Acknowledgments

We gratefully acknowledge Returned and Services League of Australia Queensland (RSLQ) for their committed support and ongoing partnership. RSLQ has funded a multiphase research project aimed at determining the factors that contribute to a successful and psychologically healthy transition. This collaboration has led to the development of the M-CARM and Go-Beyond program. Generative artificial intelligence was not used in any portion of the manuscript writing.

Data Availability

The datasets generated during and analyzed during this study are not publicly available due to ethical and privacy considerations but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Resnik L, Bradford DW, Glynn SM, Jette AM, Johnson Hernandez CJ, Wills S. Issues in defining and measuring veteran community reintegration: proceedings of the working group on community reintegration, VA Rehabilitation Outcomes Conference, Miami, Florida. J Rehabil Res Dev 2012;49(1):87-100 [FREE Full text] [doi: 10.1682/jrrd.2010.06.0107] [Medline: 22492341]
2. Romaniuk M, Kidd C. The psychological adjustment experience of reintegration following discharge from military service: a systemic review. Journal of Military and Veterans' Health 2018;26(2):60.
3. Sachdev S, Dixit S. Military to civilian cultural transition experiences of retired military personnel: a systematic meta-synthesis. Mil Psychol 2023:1-14 [FREE Full text] [doi: 10.1080/08995605.2023.2237835] [Medline: 37490333]
4. Black T, Papile C. Making it on civvy street: an online survey of canadian veterans in transition. Canadian journal of counselling and psychotherapy 2010;44(4):383.
5. Blackburn D. Out of uniform: psychosocial issues experienced and coping mechanisms used by veterans during the military–civilian transition. Journal of Military, Veteran and Family Health 2017;3(1):62-69 [FREE Full text] [doi: 10.3138/jmvfh.4160]
6. Romaniuk M, Kidd C, Banfield M, Batterham PJ. Obtaining and retaining employment post military service: a qualitative analysis of challenges experienced by Australian veterans. Journal of Employment Couns 2022;60(2):72-88 [FREE Full text] [doi: 10.1002/joec.12199]
7. Misca G, Augustus J, Russell J, Walker J. Meaning(s) of transition(s) from military to civilian life at the intersection with mental health: implications for clinical settings. Front Psychol 2023;14:1142528 [FREE Full text] [doi: 10.3389/fpsyg.2023.1142528] [Medline: 37228334]
8. Moriarty H, Winter L, Robinson K, True G, Piersol C, Vause-Earland T, et al. Exploration of individual and family factors related to community reintegration in veterans with traumatic brain injury. J Am Psychiatr Nurses Assoc 2015;21(3):195-211 [FREE Full text] [doi: 10.1177/1078390315591879] [Medline: 26156059]
9. Ahern J, Worthen M, Masters J, Lippman SA, Ozer EJ, Moos R. The challenges of afghanistan and iraq veterans' transition from military to civilian life and approaches to reconnection. PLoS One 2015;10(7):e0128599 [FREE Full text] [doi: 10.1371/journal.pone.0128599] [Medline: 26132291]
10. Pease JL, Billera M, Gerard G. Military culture and the transition to civilian life: suicide risk and other considerations. Soc Work 2016;61(1):83-86 [FREE Full text] [doi: 10.1093/sw/swv050] [Medline: 26898003]
11. Bröcker E, Suliman S, Olff M, Seedat S. Feasibility, acceptability, and effectiveness of web-based and mobile PTSD Coach: a systematic review and meta-analysis. Eur J Psychotraumatol 2023;14(2):2209469 [FREE Full text] [doi: 10.1080/20008066.2023.2209469] [Medline: 37226885]
12. Hobfoll SE, Blais RK, Stevens NR, Walt L, Gengler R. Vets prevail online intervention reduces PTSD and depression in veterans with mild-to-moderate symptoms. J Consult Clin Psychol 2016;84(1):31-42 [FREE Full text] [doi: 10.1037/ccp0000041] [Medline: 26322788]
13. Kuhn E, van der Meer C, Owen JE, Hoffman JE, Cash R, Carrese P, et al. PTSD coach around the world. Mhealth 2018;4:15 [FREE Full text] [doi: 10.21037/mhealth.2018.05.01] [Medline: 29963560]

XSL•FO
RenderX

14. McMahon B, Guindalini C, Mellor R. Computerised health interventions targeting Australian veterans and their families: a scoping review. Health Promot J Austr 2024;35(4):875-890 [FREE Full text] [doi: 10.1002/hpja.832] [Medline: 38082422]
15. Zanaboni P, Ngangue P, Mbemba GIC, Schopf TR, Bergmo TS, Gagnon MP. Methods to evaluate the effects of internet-based digital health interventions for citizens: systematic review of reviews. J Med Internet Res 2018;20(6):e10202-e10202 [FREE Full text] [doi: 10.2196/10202] [Medline: 29880470]
16. Fraynt R, Cooper D, Edwards-Stewart A, Hoyt T, Micheel L, Pruitt L, et al. An evaluation of mobile applications designed to assist service members and veterans transitioning to civilian life. Psychol Serv 2018;15(2):208-215 [FREE Full text] [doi: 10.1037/ser0000205] [Medline: 29723023]
17. Military-Civilian Adjustment and Reintegration Measure. 2021. URL: https://www.m-carm.org [accessed 2021-01-11]
18. Romaniuk M, Fisher G, Kidd C, Batterham PJ. Assessing psychological adjustment and cultural reintegration after military service: development and psychometric evaluation of the post-separation military-civilian adjustment and reintegration measure (M-CARM). BMC Psychiatry 2020;20(1):531 [FREE Full text] [doi: 10.1186/s12888-020-02936-y] [Medline: 33167907]
19. Lang AJ, Schnurr PP, Jain S, He F, Walser RD, Bolton E, et al. Randomized controlled trial of acceptance and commitment therapy for distress and impairment in OEF/OIF/OND veterans. Psychol Trauma 2017;9(Suppl 1):74-84 [FREE Full text] [doi: 10.1037/tra0000127] [Medline: 27322609]
20. Go-Beyond, navigating life beyond service®. 2021. URL: https://gobeyond.org.au/ [accessed 2024-11-23]
21. Team R. R: A language and environment for statistical computing. R Foundation for Statistical Computing 2021;12. [doi: 10.1201/b17972-15]
22. Braun V, Clarke V. Thematic Analysis In APA Handbook of Research Methods in Psychology. In: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological. Washington, DC: American Psychological Association; 2012.
23. Shaw RM, Howe J, Beazer J, Carr T. Ethics and positionality in qualitative research with vulnerable and marginal groups. Qualitative Research 2019;20(3):277-293 [FREE Full text] [doi: 10.1177/1468794119841839]
24. Ravindran C, Morley SW, Stephens BM, Stanley H, Reger MA. Association of suicide risk with transition to civilian life among US military service members. JAMA Netw Open 2020;3(9):e2016261 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.16261] [Medline: 32915235]
25. Sayer NA, Frazier P, Orazem RJ, Murdoch M, Gravely A, Carlson KF, et al. Military to civilian questionnaire: a measure of postdeployment community reintegration difficulty among veterans using Department of Veterans Affairs medical care. J Trauma Stress 2011;24(6):660-670 [FREE Full text] [doi: 10.1002/jts.20706] [Medline: 22162082]
26. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018;20(6):e199-e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
27. Enggasser JL, Hermos JA, Rubin A, Lachowicz M, Rybin D, Brief DJ, et al. Drinking goal choice and outcomes in a web-based alcohol intervention: results from VetChange. Addict Behav 2015;42:63-68 [FREE Full text] [doi: 10.1016/j.addbeh.2014.10.036] [Medline: 25671224]
28. Peng R, Li X, Guo Y, Ning H, Huang J, Jiang D, et al. Barriers and facilitators to acceptance and implementation of eMental-health intervention among older adults: a qualitative systematic review. Digit Health 2024;10:20552076241234628 [FREE Full text] [doi: 10.1177/20552076241234628] [Medline: 38444518]
29. Jardine J, Nadal C, Robinson S, Enrique A, Hanratty M, Doherty G. Between rhetoric and reality: real-world barriers to uptake and early engagement in digital mental health interventions. ACM Trans. Comput.-Hum. Interact 2024;31(2):1-59 [FREE Full text] [doi: 10.1145/3635472]
30. Borghouts J, Eikey E, Mark G, de Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
31. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016;23(6):776-794 [FREE Full text] [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
32. Wojtowicz M, Day V, McGrath PJ. Predictors of participant retention in a guided online self-help program for university students: prospective cohort study. J Med Internet Res 2013;15(5):e96 [FREE Full text] [doi: 10.2196/jmir.2323] [Medline: 23697614]
33. Possemato K, Wu J, Greene C, MacQueen R, Blonigen D, Wade M, et al. Web-based problem-solving training with and without peer support in veterans with unmet mental health needs: pilot study of feasibility, user acceptability, and participant engagement. J Med Internet Res 2022;24(1):e29559 [FREE Full text] [doi: 10.2196/29559] [Medline: 35023846]
34. Romaniuk M, Evans J, Kidd C. Evaluation of the online, peer delivered 'Post War: Survive to Thrive Program' for veterans with symptoms of posttraumatic stress disorder. Journal of Military and Veterans' Health 2019;27(2):55.
35. Nelson CB, Abraham KM, Walters H, Pfeiffer PN, Valenstein M. Integration of peer support and computer-based CBT for veterans with depression. Computers in Human Behavior 2014;31:57-64 [FREE Full text] [doi: 10.1016/j.chb.2013.10.012]

XSL•FO
RenderX

36. Boscarino JA, Hoffman SN, Pitcavage JM, Urosevich TG. Mental health disorders and treatment seeking among veterans in non-VA facilities: results and implications from the veterans' health study. Mil Behav Health 2015;3(4):244-254 [FREE Full text] [doi: 10.1080/21635781.2015.1077179] [Medline: 26640743]
37. Orcutt HK, Reffi AN, Ellis RA. Emotion in posttraumatic stress disorder. Etiology, Assessment, Neurobiology, and Treatment 2020:409-436 [FREE Full text] [doi: 10.1016/b978-0-12-816022-0.00014-4]
38. Smith B. Depression and motivation. Phenom Cogn Sci 2012;12(4):615-635 [FREE Full text] [doi: 10.1007/s11097-012-9264-0]
39. Svendsen MJ, Wood KW, Kyle J, Cooper K, Rasmussen CDN, Sandal LF, et al. Barriers and facilitators to patient uptake and utilisation of digital interventions for the self-management of low back pain: a systematic review of qualitative studies. BMJ Open 2020;10(12):e038800 [FREE Full text] [doi: 10.1136/bmjopen-2020-038800] [Medline: 33310794]

Abbreviations

GMR: Gallipoli Medical Research
M-CARM: Military-Civilian Adjustment and Reintegration Measure
PTSD: posttraumatic stress disorder
RSLQ: Returned and Services League of Australia Queensland

Edited by A Mavragani; submitted 23.05.24; peer-reviewed by A Barker, K Carra; comments to author 09.08.24; revised version received 13.09.24; accepted 17.10.24; published 23.01.25.

Please cite as:
Alichniewicz KK, Hampton S, Romaniuk M, Bennett D, Guindalini C
Use of Go-Beyond as a Self-Directed Internet-Based Program Supporting Veterans' Transition to Civilian Life: Preliminary Usability Study
JMIR Form Res 2025;9:e60868
URL: https://formative.jmir.org/2025/1/e60868
doi:10.2196/60868
PMID:39847427

©Karolina Katarzyna Alichniewicz, Sarah Hampton, Madeline Romaniuk, Darcy Bennett, Camila Guindalini. Originally published in JMIR Formative Research (https://formative.jmir.org), 23.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Usability and Usefulness of a Symptom Management Coaching System for Patients With Cancer Treated With Immune Checkpoint Inhibitors: Comparative Mixed Methods Study

Savannah Lucia Caterina Glaser[1,2], MSc; Itske Fraterman[3], MSc; Noah van Brummelen[1,4], MSc; Valentina Tibollo[5], MSc; Laura Maria Del Campo[6], PhD; Henk Mallo[7], MSN; Sofie Wilgenhof[7], MD, PhD; Szymon Wilk[8], PhD; Vitali Gisko[9], BA; Vadzim Khadakou[9], BA; Ronald Cornet[1,2], PhD, IR; Manuel Ottaviano[10], PhD; Stephanie Medlock[1,11], DVM, PhD

[1]Department of Medical Informatics, Amsterdam UMC - University of Amsterdam, Amsterdam, Netherlands

[2]Methodology & Digital Health, Amsterdam Public Health, Amsterdam, Netherlands

[3]Department of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Amsterdam, Netherlands

[4]Digital Health, Amsterdam Public Health, Amsterdam, Netherlands

[5]Laboratory of Informatics and Systems Engineering for Clinical Research, Istituti Clinici Scientifici Maugeri SpA SB IRCCS, Pavia, Italy

[6]Associazione Italiana Malati di Cancro, Rome, Italy

[7]Department of Medical Oncology, Antoni van Leeuwenhoek, Amsterdam, Netherlands

[8]Institute of Computing Science, Poznan University of Technology, Poznan, Poland

[9]BITSENS JSC, Vilnius, Lithuania

[10]Life Supporting Technologies, Universidad Politécnica de Madrid, Madrid, Spain

[11]Methodology & Aging & Later Life, Amsterdam Public Health, Amsterdam, Netherlands

Corresponding Author:
Stephanie Medlock, DVM, PhD
Department of Medical Informatics
Amsterdam UMC - University of Amsterdam
Amsterdam
Netherlands
Phone: 31 205669111
Email: s.k.medlock@amsterdamumc.nl

Abstract

Background: The prognosis for patients with several types of cancer has substantially improved following the introduction of immune checkpoint inhibitors, a novel type of immunotherapy. However, patients may experience symptoms both from the cancer itself and from the medication. A prototype of the eHealth tool Cancer Patients Better Life Experience (CAPABLE) was developed to facilitate symptom management, aimed at patients with melanoma and renal cell carcinoma treated with immunotherapy. Better usability of such eHealth tools can lead to improved user well-being and reduced risk of harm. It is unknown for usability evaluations whether certain usability problems would only be evident to patients whose condition closely resembles the target population, or if a broader group of patients would lead to the identification of a broader range of potential usability issues.

Objective: This study aims to evaluate the CAPABLE prototype by conducting tests to assess usability, user experience, and perceived acceptability among end users, and to assess any agreements or differences in the results of our wide range of participants.

Methods: This usability study was executed by interviewing participants with a melanoma or renal cell carcinoma diagnosis who have received immunotherapy and participants without direct experience with the targeted cancer types who have not received immunotherapy. Participants were asked to review the concept of the tool, perform think-aloud tasks, and complete the System Usability Scale and a Perceived Usefulness questionnaire. Usability problems were extracted from the interview data by independent coding and mapped to an eHealth Usability Problem Framework.

Results: We included 21 participants in the study, aged 29 to 73 years; 13 participants who had received immunotherapy and 8 participants who had not received immunotherapy. In total, 76 usability problems were identified. A total of 22 usability problems were in the task-technology fit category of the usability framework, mostly regarding the coaching and symptom functionality of the prototype. Critical problems regarding the symptom monitoring functionality were mainly found by participants who had

XSL•FO
RenderX

received immunotherapy. For 8 out of 10 statements in the Perceived Usefulness questionnaire, more than 75% of participants agreed or strongly agreed. The overall mean System Usability Scale score was 80 out of 100 (SD 11.3).

Conclusions: Despite identified usability issues, participants responded positively to the Perceived Usefulness questionnaire regarding the evaluated tool. Further analysis of the usability problems indicates that it was essential to include participants who matched the target end users. Participants treated with immunotherapy, specifically with previous experience in immune-related adverse events, encountered critical problems with symptom reporting that would not have been identified if these participants were not included. For other tasks and functionalities, it seems likely that loosening the inclusion criteria would have resulted in sufficient feedback without critical missing usability issues.

(JMIR Form Res 2025;9:e57659) doi:10.2196/57659

KEYWORDS

oncology; usability; usefulness; symptom management; coaching system; patients with cancer; immune checkpoint inhibitors; comparative qualitative study; medication; eHealth applications; caregivers; cancer treatment; patient education; well-being interventions; acceptability; melanoma; renal cell carcinoma; immunotherapy

Introduction

The prognosis for patients with several types of cancer has substantially improved following the introduction of immune checkpoint inhibitors, a novel form of immunotherapy that incites the patient's own immune system to attack the cancer cells [1-4]. The treatment has improved prognosis, but also often incites side effects, ranging from mild to potentially life-threatening. In addition, patients may experience symptoms from the cancer itself, as well as the stress of the disease and treatment on their mental well-being. Consequently, patients may experience a diminished health-related quality of life [2-5]. Previous research has shown that patients with cancer often have unmet care and information needs during their treatment and follow-up. eHealth apps may support health providers in addressing these needs [6,7].

For example, eHealth apps in cancer treatment are used to facilitate timely symptom reporting by the collection of patient-reported outcomes, by providing information for patients and caregivers on diagnosis, treatment, and side effects, and by giving patients access to home interventions for physical and mental well-being [8-12]. While there is conclusive evidence on the impact of eHealth on perceived support and knowledge levels, there are inconsistent findings for outcomes related to quality of life, self-management, and physical or mental well-being [6,13,14]. These potential benefits of eHealth apps are partially dependent on their ease of use, Perceived Usefulness, and eventual user acceptance [15,16].

It is known that user-centered design (UCD) processes for these types of eHealth apps may benefit system usability and user acceptance [16-18]. Better usability can lead to benefits such as enhanced user well-being and reduced risk of harm [19]. A significant part of UCD is evaluating apps using a usability evaluation method (UEM). Most UEMs highlight the importance of doing so with the intended end users of the product in order to accurately extract and understand the usability problems of target users [20]. The International Organization for Standardization standard for health and wellness apps states if the app is specifically designed to cater to individuals with a particular health condition, testing should involve participants with that condition [21]. However, the characteristics of the intended end users for health apps might not be strictly defined, or the intended end users might consist of a broadly defined, heterogeneous patient population [22]. Thus, it is not always clear which user characteristics should be considered when recruiting representative participants. In addition, it may be difficult for researchers to find patients who fit a particular profile, forcing them to adjust their inclusion criteria.

Prior to this study, we developed a prototype of the eHealth tool Cancer Patients Better Life Experience (CAPABLE). This tool facilitates symptom reporting, patient education, and well-being interventions [23]. The prototype was developed in an iterative manner based on UCD principles [6]. At the start of prototype development, preliminary and semistructured interviews were conducted with patients, caregivers, and health care professionals, to elicit their specific support needs and requirements for an eHealth tool [7]. These requirements were translated into functionalities of the prototype, followed by a preliminary usability test using heuristic evaluation.

The CAPABLE tool is intended for melanoma and renal cell carcinoma patients treated with immunotherapy, and their health care providers. During development, we aimed to consider various aspects of implementation for these different patient populations in different countries. The pilot study of CAPABLE will focus on these specific patient populations [24]. However, it is intended to be useful for a more general population of patients with cancer who are undergoing treatment. It is not known whether certain usability problems would only be evident to patients whose condition closely resembles patients in the trial, or if a broader group of patients would lead to the identification of a broader range of potential usability issues. This is particularly relevant for systems like CAPABLE, intended for use by patients with serious health complaints, who ethically should only be asked to participate in tasks where their time and effort are truly needed.

Therefore, we aimed to evaluate the CAPABLE prototype by conducting tests to assess usability, user experience, and perceived acceptability among end users. A secondary objective of the study is to assess any agreements or differences in the results of our wide range of participants in the usability studies, considering the target population of our eHealth intervention (patients with melanoma and renal cell carcinoma treated with immunotherapy) and participants that are part of a broader

XSL•FO
RenderX

population (patients with cancer with other types of cancer and informal caregivers).

Methods

Overview

The CAPABLE prototype was developed by a multidisciplinary Consortium, as part of the CAPABLE Project [23]. The overall aim of the project is to explore the effect, usability, and feasibility of the CAPABLE tool in a pilot study with patients with melanoma and renal cell carcinoma, during treatment with immune-checkpoint inhibitors [24]. This usability study was performed as part of the UCD process, through a think-aloud approach where participants were asked to execute scenario-based tasks using the CAPABLE prototype [20,25]. We followed the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist for reporting our results (Multimedia Appendix 1).

The CAPABLE Prototype

The CAPABLE prototype consists of a mobile app for patients and a web-based dashboard for their health care providers. This usability study focuses on the mobile app for patients. Software from Invision was used to create a full clickable prototype of the CAPABLE app. See Figure 1 for an excerpt of screens from the CAPABLE prototype app screens [26].

The mobile app for patients consists of five different sections to facilitate symptom reporting, well-being interventions, and patient education (Table 1).

Figure 1. Excerpt of screens from the CAPABLE mobile app prototype evaluated during the usability study. CAPABLE: Cancer Patients Better Life Experience.

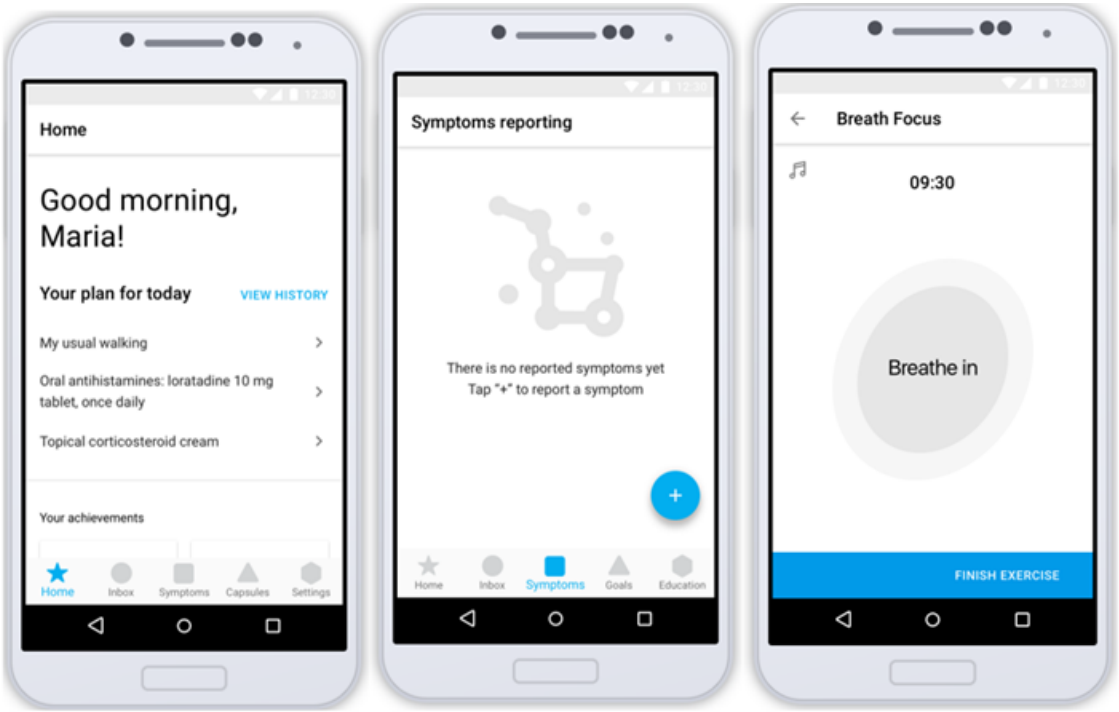


Table 1. Sections of the CAPABLE mobile app with their goals and content for supporting patients treated with immunotherapy, including symptom reporting, well-being interventions, and patient education.

Sections	Goals and contents
Home page	Contains a daily plan for patients to follow, including their hospital appointments, links to patient questionnaires, and suggestions for well-being interventions.
Inbox	Contains messages, recommendations, and reminders regarding their symptoms and well-being interventions.
Symptom reporting	Allows users to report their symptoms and provides feedback based on implemented computer-interpretable clinical guidelines for the management of symptoms related to immunotherapy side effects. Users report symptoms by selecting from a predefined set of descriptions that detail both the severity of symptoms and their impact on daily activities. After reporting, they receive feedback such as self-care instructions or advice to contact their health care provider.
Goals	Provides well-being goals such as improving sleep, physical well-being, and mental well-being, and offers corresponding theory-driven digital behavior change interventions including breathing exercises, meditation, and walking activities [27].
Education	Contains information about melanoma and renal cell carcinoma, treatments including targeted therapy and immunotherapy, side effects, supportive care, and nutrition.

XSL•FO
RenderX

Recruiting and Inclusion

Three recruiting organizations participated in the study, the Netherlands Cancer Institute-Antoni van Leeuwenhoek (NKI-AvL) in The Netherlands, Istituti Clinici Scientifici Maugeri (ICSM) in Italy, and the Italian Association of Cancer Patients, Relatives and Friends (AIMAC).

We aimed to recruit three participant groups; individuals diagnosed with melanoma previously treated with immunotherapy, those with renal cell carcinoma previously treated with immunotherapy, and participants without direct experience with the targeted cancer types nor with immunotherapy (thus patients with other types of cancer or informal caregivers). For each participant group, our goal was to recruit seven to nine participants. We did not specifically aim for data saturation as we considered the recommended sample sizes found in the literature, which range from five to ten participants. We chose seven to nine participants per group to reduce the risk of missing usability problems while being sensitive to time and resource constraints [28,29]. The inclusion criteria and recruitment strategies are described in Table 2.

Table 2. Inclusion criteria and recruitment strategies across the three participating centers.

	NKI-AvL[a]	ICSM[b]	AIMAC[c]
Participant	• Adult patients	• Adult patients	• Adult AIMAC members or their informal caregivers
Diagnosis	• High risk (resectable stage III) or advanced (stage IV and unresectable stage III) melanoma	• Renal cell carcinoma	• Any type of cancer
Treatment	• During or after treatment with immune checkpoint inhibitors	• During or after treatment with immune checkpoint inhibitors	• Any type of treatment
Language	• Sufficient understanding of the Dutch language	• Sufficient understanding of the Italian language	• Sufficient understanding of the Italian language
Recruitment	• Purposive sampling strategy to obtain a sample that varied in age • Participants were invited by their treating clinician, face-to-face or by telephone	• Purposive sampling strategy to obtain a sample that varied in age • Participants were invited by their treating clinician, face-to-face or by telephone	• Open enrollment recruitment from volunteering patient network

[a]NKI-AvL: the Netherlands Cancer Institute-Antoni van Leeuwenhoek.

[b]ICSM: Istituti Clinici Scientifici Maugeri in Italy.

[c]AIMAC: the Italian Association of Cancer Patients, Relatives, and Friends.

Ethical Considerations

This study was approved by the medical ethics committee of Amsterdam Universitair Medisch Centrum (Amsterdam; ID 2023.0944). Written informed consent was obtained from all participants prior to the interviews. All data collected for this study were pseudonymized with a new identifier. The research participants were not compensated for their study participation.

Interviews

Overview

The interviews were performed by two research teams. The interviews at NKI-AvL were performed in Dutch by two female PhD candidates with prior interviewing experience (IF and SLCG). The interviews at ICSM and AIMAC were performed in Italian by a female researcher (VT), a male senior researcher (MO), and a male research assistant with previous experience as a health care professional (Federico Dagostin). Two participants from NKI-AvL had participated in a previous interview with the NKI-AvL research team to elicit specific needs and requirements for the CAPABLE tool. The researchers had no clinical relationship with nor did know the remaining participants. A collaborative training session was held with all interviewers to streamline the UEMs used in the interviews as much as possible.

Due to the COVID-19 pandemic, interviews were done digitally using videoconferencing tools Teams (Microsoft Corp) and Zoom (Zoom Video Communications). The interviews were recorded using the screen- and audio-capture functionalities of these tools. The planned duration of the interviews was 45-60 minutes. The interviews were conducted in three phases: (1) the introduction, where participants were asked their opinion on the CAPABLE concept; (2) completing think-aloud tasks, and (3) filling in questionnaires. Participants were aware that the interviewers were researchers involved in the development of the CAPABLE prototype. See Multimedia Appendix 2 for the interview protocol.

Phase I: Introduction

The interview started with a few questions regarding patient demographics, and their previous experience with smartphones and technology. This was followed by a short presentation about the CAPABLE system and a few open-ended questions regarding the participant's opinion on the proposed goals and functionalities of the mobile app. Then, the interviewer explained the concept of the think-aloud phase of the interview. The participants were informed that the purpose of the

XSL•FO
RenderX

think-aloud tasks was to evaluate the app's performance and that the method required them to talk about what they were doing and thinking whilst using the app.

Phase II: Think-Aloud Tasks

The participants were asked to complete five tasks in total. These tasks were developed in cooperation with research team members and the developer. The tasks were (1) go through the introduction, check notifications, and report an activity; (2) report an itch symptom in the patient role; (3) report a fever symptom in the caregiver role; (4) find and perform a deep breathing exercise; and (5) find and review information about skin toxicity.

After performing the tasks, participants were asked about their final opinions, suggestions, or functionalities that they missed in the app.

Phase III: Questionnaires

Next, we administered the System Usability Scale (SUS) survey, a 10-item questionnaire aimed to assess the usability of a system considering effectiveness, efficiency, and satisfaction [30]. Finally, the participants were asked to fill out an 11-item questionnaire on Perceived Usefulness, based on the Technology Acceptance Model [31]. For both questionnaires, each item contains a statement that the participant is asked to rate based on a 5-point scale, ranging from strongly agree to strongly disagree (Multimedia Appendix 3).

Data Analysis

Participant Characteristics

The participants' characteristics were analyzed using descriptive statistics. We divided participants into two groups: treated with immunotherapy and not treated with immunotherapy.

Task Completion

We assessed the effectiveness of the participants performing the tasks according to three measures: (1) completed with ease, (2) completed with difficulty, and (3) failed to complete. We defined completed with difficulty as completed while needing to receive hints from the interviewer. Giving hints was done to ensure that the participant would have the opportunity to review all the content of the app and allow the researchers to obtain additional information about the usability of the app regardless of the completion status of the task.

Usability Problems

The Dutch interview recordings at NKI-AvL were transcribed verbatim (SLCG) as source data for our coding and analysis of the usability problems. During the Italian interviews at ICSM and AIMAC, notes were made by the interviewer that summarized the participants' feedback. These notes were revised and completed after the interview upon reviewing the recording. The transcripts and notes were not returned to the participants.

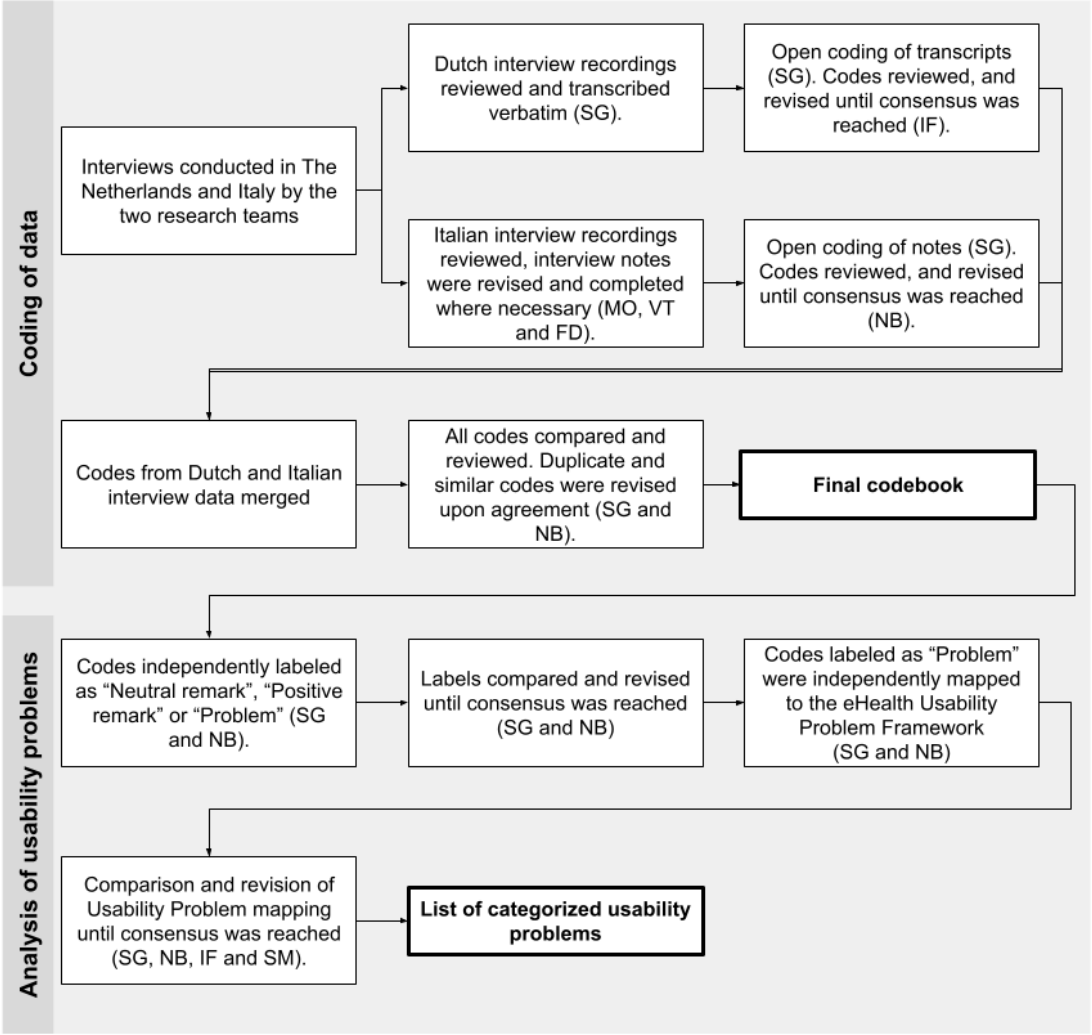
As two research teams performed the interviews and the data used to code the interviews varied (ie, transcripts and summarizing notes), we conducted a data quality check. Two Italian interviews and two Dutch interviews were transcribed and independently coded by researchers from both research teams (SLCG and VT). The codes based on the transcript were compared to those based on the summarizing notes of the Italian interviews to evaluate their consistency. In addition, the independently coded interviews were compared to ensure whether the interpretation of the data was similar between the two research teams. Transcripts and notes were considered comparable if both reviewers agreed that there were no substantial differences or omissions in the code's results.

Coding and Analysis of Interview Recordings

Several steps were taken to code and analyze the interview recordings. This was executed by the same researchers who conducted the Dutch interviews (SLCG and IF), plus NB, an MSc in Medical Informatics. The coding and analysis were done in English. See Figure 2 for an overview of these steps.

XSL•FO
RenderX

Figure 2. Steps executed for open coding of the usability interview transcripts and notes and analysis of the usability problems by mapping to the eHealth Usability Problem Framework. FD: Federico Dagostin.



First, the interviews were open-coded by SLCG. The codes from the Dutch and Italian interviews were reviewed by IF and NB, respectively, until a consensus was reached. After merging the data and codes from the Dutch and Italian interviews, an extra revision step was executed to ensure that duplicate and similar codes were revised, upon agreement of SLCG and NB. This resulted in the final codebook. All codes in the codebook were independently labeled by SLCG and NB as "Positive remark," "Neutral remark," or "Problem" to allow us to focus on the usability problems indicated by the participants. Then, we mapped the "Problem" codes in the codebook to an eHealth Usability Problem Framework developed by Broekhuis et al [32]. This framework specifies 21 usability factors in the following categories: basic system performance, task-technology fit, accessibility, interface design, navigation and structure, information and technology, guidance and support, and satisfaction. This mapping process was done by SLCG and NB, in three different cycles, independently mapping and comparing the mappings, reaching consensus. Any discrepancies were discussed with IF and SM.

To gain insight into the usability problems and the frequency with which they were encountered, we counted the distinct problems associated with each usability factor (eg, the number of issues related to the usability factor "design clarity" under the category "interface design." To address our secondary objective, we noted which participants had observed specific problems, and counted how many were identified by both groups (those treated with immunotherapy and those not treated with immunotherapy) versus problems identified by only one group. Furthermore, for each distinct usability problem, we noted the number of participants who identified the problem and compared how frequently the problem was reported across the two participant groups.

We provide a more detailed account of the usability problems in the "Qualitative Assessment" section. We adopted a UCD approach and a postpositivist perspective. The eHealth Usability Problem Framework was used as a theoretical framework for our analysis. We included the problems most frequently identified, as well as those where there was a notable difference in identification between the two participant groups.

The qualitative analysis was conducted without the use of any specialized software. The quantitative analysis of counts and frequency was performed using Excel (Microsoft Corp).

XSL•FO
RenderX

Perceived Usefulness and SUS Questionnaires

The SUS survey results were evaluated using the standard SUS scoring algorithm. We followed the general guideline for interpretation of the results, with scores below 68 indicating poor usability, and scores above 68 indicative of good usability. The Perceived Usefulness results were analyzed using frequency analysis.

Results

Participant Characteristics

We conducted 21 interviews in total. The interviews took place between June 2021 and April 2022. The average age of the participants was 53 (SD 12.1) years. We interviewed 7 patients with melanoma, 6 patients with renal cell carcinoma, 6 patients with other types of cancer (breast cancer, urinary cancer, lymphoma, and vestibular schwannoma), and 2 caregivers. See Table 3 for an overview of the participants' characteristics. The participants with melanoma and renal cell carcinoma were treated with immune checkpoint inhibitors. All participants used a smartphone and were familiar with mobile apps. Five participants mentioned having used a health-related mobile app before.

Table 3. Characteristics of participants that were interviewed for this usability study (between June 2021 and April 2022).

	Treated with immunotherapy?	
	Yes (n=13)	No, N/A[a] (caregiver; n=8)
Sex, n (%)		
Male	6 (29)	6 (29)
Female	7 (33)	2 (10)
Age (years)		
Mean (SD)	52 (11)	56 (14)
Range	32-71	29-73
Type of cancer, n (%)		
Melanoma III-IV	7 (33)	0 (0)
Kidney I-II	4 (19)	0 (0)
Kidney III-IV	2 (10)	0 (0)
Breast I-II	0 (0)	1 (5)
Breast III-IV	0 (0)	2 (10)
Lymphoma	0 (0)	1 (5)
Urinary IV	0 (0)	1 (5)
Vestibular schwannoma I-II	0 (0)	1 (5)
N/A (caregiver)	0 (0)	2 (10)
Treatment status, n (%)		
On treatment	8 (38)	4 (19)
Off treatment	5 (24)	2 (10)
N/A (caregiver)	0 (0)	2 (10)
Living situation, n (%)		
Alone	2 (10)	1 (5)
With relatives	11 (52)	6 (29)
Shared housing	0 (0)	1 (5)
Type of smartphone, n (%)		
Android	6 (29)	5 (24)
iOS	7 (33)	3 (14)

[a]Not applicable.

XSL•FO
RenderX

Task Completion

We measured the effectiveness of the participants performing tasks by completion rates (Table 4). Participants found Task 1, "Go through introduction, check notifications, and report an activity," and Task 3, "Report a fever symptom in a caregiver's role," the most challenging. Task 5, "Find and review information about skin toxicity," had the highest completion rate. Participants who were not treated with immunotherapy seemed to struggle more with the symptom-reporting tasks (Tasks 2 and 3).

Table 4. Task completion rates of tasks performed during the think-aloud part of the usability interview, divided by participant group (participants previously treated with immunotherapy and not treated with immunotherapy).

	Task 1: introduction and home page		Task 2: symptom reporting and response inbox (itch as patient)		Task 3: symptom reporting and response inbox (fever as caregiver)		Task 4: coaching, goals, and activities		Task 5: patient education	
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Completed chemotherapy with ease, n (%)	8 (62)	5 (63)	11 (85)	4 (50)	10 (77)	3 (38)	9 (69)	5 (63)	11 (85)	8 (100)
Completed chemotherapy with difficulty, n (%)	5 (38)	3 (37)	2 (15)	4 (50)	3 (23)	5 (62)	4 (31)	3 (37)	2 (15)	0 (0)

Usability Problems

In the four interviews used to assess data quality, no substantial differences were noted in the results from the Italian and Dutch research teams, nor were differences between the analyses based on transcripts compared to the analyses from notes. The participants identified 76 distinct usability problems and proposed 15 additional features missing in the CAPABLE app. The first task "Go through introduction, check notifications, and report an activity" resulted in 31 issues; the second and third tasks "Report an itch symptom in a patients' role" and "Report a fever symptom in a caregivers' role" resulted in 24 issues; the fourth task "Find and perform a breathing exercise" resulted in 16 issues; and the final fifth task "Find and review information about skin toxicity" resulted in 12 issues. Eight problems recurred in multiple tasks.

Table 5 shows an overview of the eHealth Usability Problem Framework [32] and the count of usability problems found per usability factor. The category with the highest number of usability problems is the task-technology fit category, which relates to the match between the system on the one hand, and the user, their context of use, and their health goals on the other hand. Of these, 23 usability problems were found by both participants who received immunotherapy and participants who did not receive immunotherapy, 31 problems were solely found by participants who received immunotherapy, and 22 problems were solely found by participants who did not receive immunotherapy. Table 6 shows an overview of all usability problems that were indicated by more than one participant, resulting in 32 problems. See Multimedia Appendix 4 for the complete table of usability problems. We discuss the usability problems that participants most frequently indicated during the study in the following section.

XSL•FO
RenderX

Table 5. Number of distinct usability problems identified during the think-aloud part of the usability interview[a].

Usability factors	Number of issues found by both groups	Number of issues found only by the immunotherapy group	Number of issues found only by the nonimmunotherapy group	Number of total issues
Basic system performance	1	1	0	2
General system interaction	1	1	0	2
Task-technology fit	6	9	7	22
Fit between system and context of use	1	2	1	4
Fit between system and health goals	3	2	3	8
Fit between system and user	2	5	3	10
Interface design	2	5	4	11
Design clarity	0	2	0	2
Interface organization	1	1	1	3
Readability of texts	1	0	2	3
Symbols, icons, and buttons	0	2	1	3
Navigation and structure	4	0	1	5
Navigation	3	0	1	4
Structure	1	0	0	1
Information and terminology	6	3	6	15
Health-related information	4	3	5	12
System information	2	0	1	3
Guidance and support	1	7	1	9
Procedural health-related information	1	6	1	8
Procedural system information	0	1	0	1
Satisfaction	3	6	3	12
Satisfaction with system	2	2	1	5
Satisfaction with the system's ability to support health goals	1	4	2	7
Total	23	31	22	76

[a]The counts are presented per usability factor and usability factor category.

XSL•FO
RenderX

Table 6. List of usability problems including the count of participants that identified these usability problems, divided by group (treated with immunotherapy vs not treated with immunotherapy)[a].

	Treated with immunotherapy?	
	Yes	No
Introduction and home page		
Task-technology fit		
Current list of hobbies not sufficient	2	2
Expects automatic detection of activities by smartwatch	3	1
Time of going to bed is different every day	2	0
Unclear what the added value is of recording exercises or activities	2	0
Allow user to write down why they did or did not like the challenge	2	0
Not possible to select multiple hobbies	1	1
Need for balance between coaching, support for symptoms, and support for cancer treatment in content	0	2
Interface design		
Text is too long	5	2
Participant prefers graphical explanations to textual explanations	0	2
Navigation and structure		
Link between times of waking up or going to bed and the symptom management unclear	1	1
Information and terminology		
Unclear what vital functions in home page are	2	0
Satisfaction		
Tone in the introduction text is not appreciated	1	1
Symptom reporting and response inbox		
Basic system performance		
Not clear that scrolling was necessary to view everything in screen	6	5
Task-technology fit		
Information missing about medication in recommendation (dosage, need for prescription, where to get it)	7	1
Symptom descriptions do not match experience of itch of the patient, would be difficult to choose	6	0
Feedback missing after report	1	5
Navigation and structure		
Participant cannot find symptom section easily	3	1
Information and terminology		
Term caregiver is unclear, caregiver can be a professional or family or friends visiting	1	1
Guidance and support	0	0
Unclear if the clinician will view the report, if patient will be contacted, and what is expected of patient	4	0
Missing from instructions that caregivers can report symptoms	2	0
Satisfaction		
Symptom reporting process seems long or steps redundant	1	1
Participant does not trust the feedback from the app	1	1
Coaching, goals, and well-being interventions		
Task-technology fit		
Expects automatic detection of activities by smartwatch	0	3
Unclear how users can set their own goals	0	2
Unmet expectation of the app recommending a schedule with activities, which can be personalized	2	2

XSL•FO
RenderX

	Treated with immunotherapy?	
	Yes	No
Navigation and structure		
Participant cannot find the exercise easily	7	2
Information and terminology		
Names of types of breathing exercises are not self-explanatory	4	1
Menu term “objectives or goals” does not match with content found	2	1
Guidance and support		
Unclear how goals relate to the activities	4	0
Invitation for well-being intervention is not clear, not self-explanatory what it is and how to proceed	10	7
Satisfaction		
Content, purpose, and benefits of well-being interventions not sufficient currently	3	0
Patient education		
Interface design		
Reorganize the categories and structure of the educational content list	1	3
Navigation and structure		
Participant cannot find educational section easily (clicks on symptoms first)	1	3
Information and terminology		
First part of the text is difficult to understand without in-depth knowledge	1	2
Education not the correct term for section	1	2
Rash and itching are not translated	1	1
Total	128	85

[a]Problems included in this table were identified by at least two participants.

Qualitative Assessment

Introduction and Home Page

Task-technology fit: the CAPABLE app asks the user to select their hobbies from a predefined list. This selection is used to recommend similar activities. Participants could not find their own hobbies in this list and were unable to select more than one hobby.

Interface design: in the introduction, the text to explain the purpose and functionalities of the CAPABLE app was deemed too long by some participants and participants stated a preference for more graphical explanations.

> *Yes, it [the text] was long, and the concentration of some patients might be decreased, so a video might be easier.* [Participant 20]

Symptom Reporting and Feedback Inbox

In general, it was not always clear to participants that scrolling was necessary to view everything on screen.

Task-technology fit: participants who received immunotherapy treatment predominantly experienced problems when reporting an “Itch” symptom using the symptom reporting function in the app. The app presents a set of symptom descriptions to choose from. Nearly half of the participants in the immunotherapy group found it difficult to choose one of those descriptions, which did not match their own experiences with itching caused by immunotherapy:

> *It’s very black-and-white, isn’t it? Yes, I had constant itching, all day long, but I could still do activities. And that’s not reflected in there. You really should give a lot more options ... I had constant itching, but of course, I could still go grocery shopping, and wash myself, and, sleeping was really difficult though, whole nights without sleep.* [Participant 21]

During the think-aloud task, the app recommended an emollient cream for the itch symptom reported by the participants. While participants appreciated the advice, a majority of the participants in the immunotherapy group found the message incomplete, as they were not sure what the exact dosage of the cream would be, if a prescription would be needed, and where to acquire the medication.

Guidance and support: after reporting the symptom, participants were confused about the next steps in the process. A subset of participants from the immunotherapy group expressed uncertainty on whether a clinician would view the report, in which cases they would be contacted by the clinician, and what was expected of them as a patient.

> *That may still need to be coordinated, like, when should the patient or caregiver take action themselves if they notice things are not going well, and when*

XSL•FO
RenderX

> *does the team proactively step in or call to check how things are going or if they can offer any help? would have expected if you report [symptoms] in this manner, that your care team would be informed.* [Participant 15]

Coaching, Goals, and Activities

Task-technology fit: the section "Goals" in the app contains a list of exercises and activities. These can be filtered by their goal, including supporting mental well-being, physical well-being, sleep, and acceptance. The connection between the name of the section and the list of activities was perceived as unclear and confusing by some participants. As a result, when asked to find a breathing exercise in the app, it was not always obvious for participants to click on the "Goals" button from the home page. In addition, while participants appreciated the ability to easily select and do an exercise, it was unclear how often to do these activities and how to schedule activities in the daily plan on the home page of the app:

> *[I expected] some sort of automatic coach within the system that informs you: you've chosen mental well-being as your focus, so every night before going to sleep, do fifteen minutes of meditation.* [Participant 8]

Participants were also confused about wearing a smartwatch while the app requires manual registration of exercises such as walking.

Guidance and support: users receive an invitation in the CAPABLE app inbox to participate in an activity, such as a daily walking challenge. While users are immediately prompted to accept this challenge, most participants found it unclear in the invitation what is expected of them, how they should proceed, where they can see the content of this program in the app, and whether and where they should register to do these activities.

Patient Education

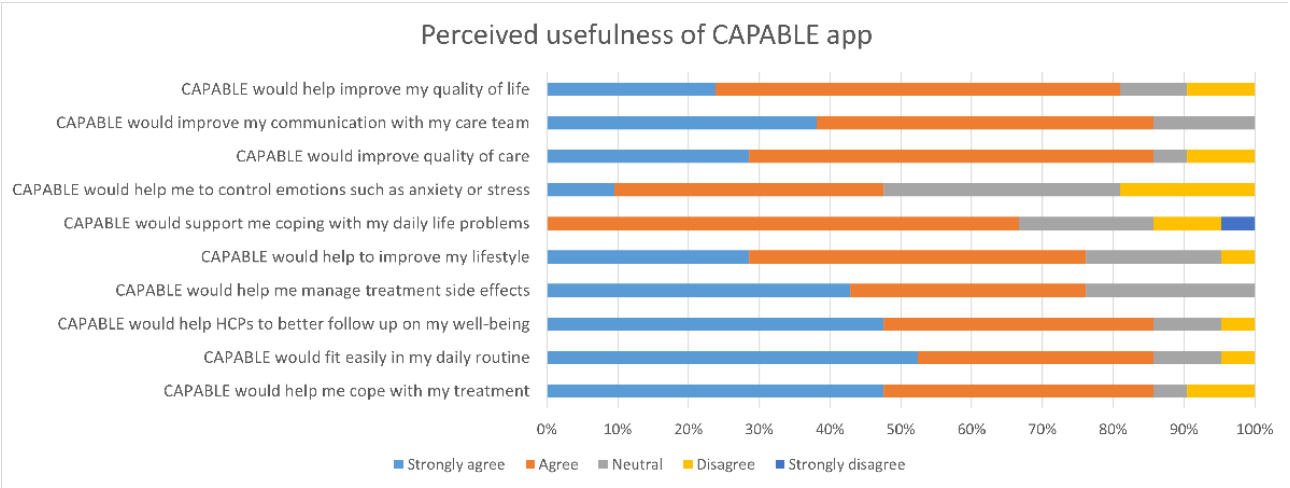
Navigation and structure: participants were asked to find information about skin adverse events due to immunotherapy. Some participants had trouble finding the correct section in the app and clicked on the "Symptom management" section instead of the "Education" section.

Information and terminology: some participants indicated that the text in the example section shown during the interview was challenging to understand without in-depth knowledge.

Perceived Usefulness and SUS Questionnaires

The participants filled in a questionnaire aiming to measure their acceptance and perceived usefulness of the CAPABLE app. Overall, participants were positive about nearly all statements. For eight out of ten questions, more than 75% of participants agreed or strongly agreed (Figure 3). The statements that were most widely agreed upon were that the system would easily fit in users' daily routines, could help health care professionals to follow up on patients' well-being, could improve communication with the care team, could help users cope with their treatment, and could help users support their quality of life. Participants were more skeptical about the CAPABLE app helping to manage emotions such as anxiety and stress and the ability of the app to help cope with daily life problems. The results of the questionnaire were comparable between the participant group treated with immunotherapy, versus the group not treated with immunotherapy. Finally, the participants were asked to fill in the SUS questionnaire. The mean SUS score was 80 (SD 11), which is classified as "good" in terms of usability.

Figure 3. Outcomes of the Perceived Usefulness questionnaire filled in at the end of the usability interview by participants, presented as a stacked bar chart. CAPABLE: Cancer Patients Better Life Experience; HCP: health care provider.



Discussion

Principal Findings

We evaluated the prototype of the CAPABLE app, a symptom monitoring and coaching system, with 21 participants divided into two groups, participants that had received immunotherapy (patients with melanoma and renal cell carcinoma) and that had not received immunotherapy (other cancer types), from The Netherlands and Italy. This evaluation was executed by use of think-aloud interviews and two questionnaires. In total, 76 usability issues were identified. Specifically, 23 usability issues were identified by both groups. The immunotherapy group

XSL•FO
RenderX

found 31 additional issues, and the nonimmunotherapy group found 22 additional issues.

Most usability problems were in the task-technology fit category of the eHealth Usability Problem Framework [33]. This is reflected in problems encountered in the coaching section of the app. The participants' mental model, defined as what a user knows, believes about, and expects from a system [34], did not match with the design and presentation of the digital behavior change interventions. Participants expected to be able to set their own goals, and thought that the app would suggest a recommended weekly schedule of activities that can be personalized according to their needs. Problems with task-technology fit were also encountered in creating a symptom report and with the subsequent advice given. Participants who had experienced itching due to immunotherapy found it difficult to choose one of the descriptions given to make a symptom report, as none of them matched their own experiences with the symptom. Additionally, participants felt that the advice provided was insufficient because it lacked information about the prescription, use, and dosage of the recommended medication, as well as whether to contact their clinician. This set of issues was mainly indicated by participants who had received immunotherapy. This is logical, since patients who received immunotherapy are more likely to have experienced a situation similar to the scenario used in the test, and could assess the fit of the app with their own lived experience. No differences were noted in the number and type of usability issues identified by caregivers compared to patients.

Despite identified usability issues, participants responded positively to the perceived acceptance and usefulness questionnaire regarding the CAPABLE app. Specifically, they would expect that the CAPABLE app would facilitate at-home monitoring, help patients cope with treatment, and support their quality of life.

Overall, our analysis of the usability problems seems to indicate a necessity to include participants with the characteristics of the intended end users for the evaluation of certain functionalities. In our case, this meant having the symptom-reporting functionality evaluated by patients who had been treated with immunotherapy, specifically with previous experience in immune-related adverse events. For the other tasks and functionalities, it seems likely that loosening the inclusion criteria would have resulted in sufficient feedback without critical missing usability issues.

Interpretation and Impact

This study adds to previously published information by extracting a specific description of the usability problems encountered by intended end users during the evaluation of symptom management and coaching eHealth intervention. These findings are relevant for researchers developing and evaluating tools like symptom monitoring apps, especially for patients with cancer. More generally, we noted that patients who had taken immunotherapy noticed problems that other patients with cancer (and caregivers) did not. This is useful information for researchers who are considering whether patients who have experienced the specific health problem under study need to be recruited. It can be that these patients are quite ill, and researchers do not want to burden them. Our results suggest that the participation of these patients added valuable information, at least in this study, and thus asking for their participation in usability studies is justified.

Comparison With Prior Work

This study found positive perspectives from participants on measured perceived acceptance and perceived usefulness. This might seem contradictory to the number and severity of usability issues reported. This is reflected in other studies, where participants are enthusiastic about the functionalities and perceived future benefits, while simultaneously encountering difficulties during the usability evaluation [35-37]. Patients' wants and needs seem to be identified clearly in research, while a gap between user needs and eHealth tool implementation might be caused by practical considerations, the adaptability of the tool to local context, complexity factors, and health professionals' uptake of the eHealth app [38].

In addition, we investigated the impact of different user characteristics on usability problems found during a think-aloud evaluation. Previous studies have researched the impact of domain knowledge, specified as the familiarity or expertise an individual has with a particular topic or subject area, on finding usability problems [39]. In these studies, the output of novices and experts, with previous expertise with the software or tools evaluated, was compared. In some cases, novices found more usability problems, less usability problems, or less usability problems but more severe problems, compared to experts [40-42]. However, this study would define previous domain knowledge not as previous experience with the software, but as previous experience with the treatment (immunotherapy) or previous experience with certain side effects (itch). Previous research found a significant difference between the novice and expert group with previous domain knowledge defined as previous pregnancy [43]. More research is needed on the impact of previous experience with a disease or treatment in the oncology field as part of the users' profile on usability evaluation outcomes.

Strengths and Limitations

In this study, we were able to recruit a large and diverse participant group from multiple centers in both The Netherlands and Italy. This was a benefit to our goal of collecting broad feedback on the CAPABLE tool and aiming to find a complete overview of potential usability issues. We had researchers from two different research teams performing the interviews. Nearly all researchers had previous interviewing experience, and a collaborative training session was held with all interviewers to streamline the UEMs used in the interviews as much as possible to prevent bias. The interviews were qualitative in nature; however, the data from the think-aloud tasks was specifically used to identify usability problems, rather than for thematic analysis. No pilot testing was performed, but no changes were needed during the course of the interview rounds.

The raw data used for coding the interviews varied: interviews with participants with renal cell carcinoma and other types of cancer were coded using summarizing interview notes, while interviews with melanoma participants were coded using

XSL•FO
RenderX

transcriptions. This discrepancy may have led to missing data from incomplete interview notes. However, our data quality check revealed that although transcripts provided more context for the codes, the codes derived from the interview notes were consistent with those from the transcripts. In addition, we had multiple researchers involved in the coding and analysis phase of the study, to reduce the evaluator effect [44].

We did not take health, smartphone literacy, or physical or mental impairments into account during recruitment. There might be a selection bias of participants agreeing to the interview as they have an interest in using their mobile phones, or invitees refusing to participate because of limited smartphone literacy. However, we recruited a varied group of 21 participants, with varying ages, different diagnoses, past treatments, and countries.

Future Work

The usability results of this study were used to improve the CAPABLE prototype. A prospective exploratory pilot study involving the final version of CAPABLE is currently being held in The Netherlands and Italy, with different patient populations treated with immune checkpoint inhibitors, including patients with melanoma and renal cell carcinoma [24]. The clinical impact and usability of the CAPABLE tool will be evaluated.

Our comparison of the usability problems generated by participants treated with immunotherapy versus the nonimmunotherapy group indicated a difference in output mainly for the symptom-reporting functionality of the CAPABLE tool. This comparison was based on qualitative data. Multiple factors may have been involved in the differences of usability problems found in addition to previous experience with the treatment or side effects, such as age, diagnosis, stadium of disease, health literacy, and country. However, our findings implore future studies to focus on exploring and comparing the usability evaluation output of participants with varying previous experiences. In addition, future research is needed to determine for which types of software, or which specific functionalities, there would be an added benefit or need for participants with specific characteristics.

Conclusions

While participants identified usability problems regarding task-technology fit, interface design, and overall satisfaction, they responded positively regarding the perceived impact of CAPABLE in monitoring patients from home, helping to cope with treatment, and supporting quality of life. Further analysis of the usability problems indicates that it was essential to include participants who matched the target end users. Participants treated with immunotherapy, specifically with previous experience in immune-related adverse events encountered critical problems with symptom reporting that would not have been identified if these participants were not included. For other tasks and functionalities, it seems likely that loosening the inclusion criteria would have resulted in sufficient feedback without critical missing usability issues. Future research is needed to determine for which types of software and which specific functionalities, there would be an added benefit or need for participants with specific characteristics.

Acknowledgments

This work was supported by the European Union’s Horizon 2020 research and innovation program (grant agreement 875052). The authors would like to thank all the participants who took part in the interviews.

Data Availability

The aggregated data is included in Multimedia Appendix 4. The unaggregated datasets generated during this study are not publicly available due to our institutions’ privacy policies but are available from the corresponding author upon reasonable request and with approval from the respective institution.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Consolidated criteria for reporting qualitative studies (COREQ): 32-item checklist.
[DOCX File , 22 KB - formative_v9i1e57659_app1.docx]

Multimedia Appendix 2
Think-aloud interview protocol.
[DOCX File , 27 KB - formative_v9i1e57659_app2.docx]

Multimedia Appendix 3
System Usability Scale and perceived impact questionnaire.
[DOCX File , 24 KB - formative_v9i1e57659_app3.docx]

Multimedia Appendix 4
Overview of all usability problems and suggestions.

XSL•FO
RenderX

[DOCX File , 38 KB - formative_v9i1e57659_app4.docx]

References

1. Larkin J, Chiarion-Sileni V, Gonzalez R, Grob JJ, Rutkowski P, Lao CD, et al. Five-year survival with combined nivolumab and ipilimumab in advanced melanoma. N Engl J Med 2019;381(16):1535-1546 [FREE Full text] [doi: 10.1056/NEJMoa1910836] [Medline: 31562797]
2. Robert C. A decade of immune-checkpoint inhibitors in cancer therapy. Nat Commun 2020;11(1):3801 [FREE Full text] [doi: 10.1038/s41467-020-17670-y] [Medline: 32732879]
3. Hargadon KM, Johnson CE, Williams CJ. Immune checkpoint blockade therapy for cancer: an overview of FDA-approved immune checkpoint inhibitors. Int Immunopharmacol 2018;62:29-39. [doi: 10.1016/j.intimp.2018.06.001] [Medline: 29990692]
4. Shiravand Y, Khodadadi F, Kashani SMA, Hosseini-Fard SR, Hosseini S, Sadeghirad H, et al. Immune checkpoint inhibitors in cancer therapy. Curr Oncol 2022;29(5):3044-3060 [FREE Full text] [doi: 10.3390/curroncol29050247] [Medline: 35621637]
5. Rogiers A, Boekhout A, Schwarze J, Awada G, Blank C, Neyns B. Long-term survival, quality of life, and psychosocial outcomes in advanced melanoma patients treated with immune checkpoint inhibitors. J Oncol 2019;2019:5269062 [FREE Full text] [doi: 10.1155/2019/5269062] [Medline: 31182961]
6. Slev VN, Mistiaen P, Pasman HRW, Verdonck-de LIM, van UCF, Francke AL. Effects of eHealth for patients and informal caregivers confronted with cancer: a meta-review. Int J Med Inform 2016;87:54-67. [doi: 10.1016/j.ijmedinf.2015.12.013] [Medline: 26806712]
7. Fraterman I, Glaser S, Wilgenhof S, Medlock S, Mallo H, Cornet R, et al. Exploring supportive care and information needs through a proposed eHealth application among melanoma patients undergoing systemic therapy: a qualitative study. Support Care Cancer 2022;30(9):7249-7260 [FREE Full text] [doi: 10.1007/s00520-022-07133-z] [Medline: 35589878]
8. Kotronoulas G, Kearney N, Maguire R, Harrow A, Di Domenico D, Croy S, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. J Clin Oncol 2014;32(14):1480-1501 [FREE Full text] [doi: 10.1200/JCO.2013.53.5948] [Medline: 24711559]
9. Lubberding S, van Uden-Kraan CF, Te Velde EA, Cuijpers P, Leemans CR, Verdonck-de Leeuw IM. Improving access to supportive cancer care through an eHealth application: a qualitative needs assessment among cancer survivors. J Clin Nurs 2015;24(9-10):1367-1379. [doi: 10.1111/jocn.12753] [Medline: 25677218]
10. Iivanainen S, Alanko T, Vihinen P, Konkola T, Ekstrom J, Virtanen H, et al. Follow-Up of cancer patients receiving anti-PD-(L)1 therapy using an electronic patient-reported outcomes tool (KISS): prospective feasibility cohort study. JMIR Form Res 2020;4(10):e17898 [FREE Full text] [doi: 10.2196/17898] [Medline: 33112242]
11. Hægermark EA, Kongshaug N, Raj SX, Hofsli E, Faxvaag A. Design, development, and evaluation of an mHealth app for reporting of side effects during cytostatic treatment: usability test and interview study. JMIR Form Res 2023;7:e47374 [FREE Full text] [doi: 10.2196/47374] [Medline: 37856183]
12. Schläpfer S, Schneider F, Santhanam P, Eicher M, Kowatsch T, Witt CM, et al. Engagement with a relaxation and mindfulness mobile app among people with cancer: exploratory analysis of use data and self-reports from a randomized controlled trial. JMIR Cancer 2024;10:e52386 [FREE Full text] [doi: 10.2196/52386] [Medline: 38819907]
13. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. BMC Health Serv Res 2013;13:211 [FREE Full text] [doi: 10.1186/1472-6963-13-211] [Medline: 23758898]
14. van Deursen L, Versluis A, van der Vaart R, Standaar L, Struijs J, Chavannes N, et al. eHealth interventions for Dutch cancer care: systematic review using the triple aim lens. JMIR Cancer 2022;8(2):e37093 [FREE Full text] [doi: 10.2196/37093] [Medline: 35699991]
15. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989;13(3):319. [doi: 10.2307/249008]
16. Ardito V, Golubev G, Ciani O, Tarricone R. Evaluating barriers and facilitators to the uptake of mHealth apps in cancer care using the consolidated framework for implementation research: scoping literature review. JMIR Cancer 2023;9:e42092 [FREE Full text] [doi: 10.2196/42092] [Medline: 36995750]
17. De Vito Dabbs A, Myers BA, Mc Curry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
18. Saparamadu AADNS, Fernando P, Zeng P, Teo H, Goh A, Lee JMY, et al. User-centered design process of an mHealth app for health professionals: case study. JMIR Mhealth Uhealth 2021;9(3):e18079 [FREE Full text] [doi: 10.2196/18079] [Medline: 33769297]
19. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]

XSL•FO
RenderX

20. Jaspers MWM. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009;78(5):340-353. [doi: 10.1016/j.ijmedinf.2008.10.002] [Medline: 19046928]
21. ISO/TS 82304-2:2021(en). Health software part 2: health and wellness apps. Quality and reliability [Internet]. URL: https://www.iso.org/obp/ui/en/#iso:std:iso:ts:82304:-2:ed-1:v1:en [accessed 2023-01-21]
22. Neal D, Engelsma T, Tan J, Craven MP, Marcilly R, Peute L, et al. Limitations of the new ISO standard for health and wellness apps. Lancet Digital Health 2022;4(2):e80-e82 [FREE Full text] [doi: 10.1016/S2589-7500(21)00273-9] [Medline: 35090677]
23. Home—CAPABLE [Internet]. URL: https://capable-project.eu/ [accessed 2021-04-13]
24. Fraterman I, Wollersheim BM, Tibollo V, Glaser SLC, Medlock S, Cornet R, et al. An eHealth app (CAPABLE) providing symptom monitoring, well-being interventions, and educational material for patients with melanoma treated with immune checkpoint inhibitors: protocol for an exploratory intervention trial. JMIR Res Protoc 2023;12:e49252 [FREE Full text] [doi: 10.2196/49252] [Medline: 37819691]
25. Jaspers MWM, Steen T, van den Bos C, Geenen M. The think aloud method: a guide to user interface design. Int J Med Inform 2004;73(11):781-795. [doi: 10.1016/j.ijmedinf.2004.08.003] [Medline: 15491929]
26. InVision. Collaborate better. URL: https://www.invisionapp.com/ [accessed 2023-10-13]
27. Lisowska A, Wilk S, Peleg M. SATO (IDEAS expAnded wiTh BCIO): workflow for designers of patient-centered mobile health behaviour change intervention applications. J Biomed Inform 2023;138:104276. [doi: 10.1016/j.jbi.2022.104276] [Medline: 36586499]
28. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003;35(3):379-383. [doi: 10.3758/bf03195514] [Medline: 14587545]
29. Nielsen J. Estimating the number of subjects needed for a thinking aloud test. Int J Hum-Comput Stud 1994;41(3):385-397. [doi: 10.1006/ijhc.1994.1065]
30. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum-Comput Interact 2008;24(6):574-594. [doi: 10.1080/10447310802205776]
31. Davis FD. Sloan School of Management. A technology acceptance model for empirically testing new end-user information systems: theory and results [Thesis]. 1985. URL: https://dspace.mit.edu/handle/1721.1/15192 [accessed 2023-07-18]
32. Broekhuis M, van Velsen L, Peute L, Halim M, Hermens H. Conceptualizing usability for the eHealth context: content analysis of usability problems of eHealth applications. JMIR Form Res 2021;5(7):e18198 [FREE Full text] [doi: 10.2196/18198] [Medline: 34313594]
33. Peute LW, Driest KF, Marcilly R, Bras Da Costa S, Beuscart-Zephir M, Jaspers MWM. A framework for reporting on human factor/usability studies of health information technologies. Stud Health Technol Inform 2013;194:54-60. [Medline: 23941930]
34. Nielsen Norman group. Experience WL in RBU. Mental Models and User Experience Design. URL: https://www.nngroup.com/articles/mental-models/ [accessed 2024-02-12]
35. Mirkovic J, Kaufman DR, Ruland CM. Supporting cancer patients in illness management: usability evaluation of a mobile app. JMIR Mhealth Uhealth 2014;2(3):e33 [FREE Full text] [doi: 10.2196/mhealth.3359] [Medline: 25119490]
36. O'Malley DM, Davis SN, Devine KA, Sullivan B, Bator A, Clemow L, et al. Development and usability testing of the e-EXCELS tool to guide cancer survivorship follow-up care. Psychooncology 2020;29(1):123-131 [FREE Full text] [doi: 10.1002/pon.5222] [Medline: 31626397]
37. Duman-Lubberding S, van Uden-Kraan CF, Peek N, Cuijpers P, Leemans CR, Verdonck-de Leeuw IM. An eHealth application in head and neck cancer survivorship care: health care professionals' perspectives. J Med Internet Res 2015;17(10):e235 [FREE Full text] [doi: 10.2196/jmir.4870] [Medline: 26489918]
38. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
39. Ju B. Does domain knowledge matter: mapping users' expertise to their information interactions. J Am Soc Inf Sci 2007;58(13):2007-2020. [doi: 10.1002/asi.20680]
40. Liu Y, Osvalder AL, Karlsson M. Considering the Importance of User Profiles in Interface Design. Croatia: NTECH; 2010.
41. Clarke MA, Belden JL, Kim MS. Determining differences in user performance between expert and novice primary care doctors when using an electronic health record (EHR). J Eval Clin Pract 2014;20(6):1153-1161. [doi: 10.1111/jep.12277] [Medline: 25470668]
42. Liu Y, Osvalder A, Karlsson M. User's expertise differences when interacting with simple medical user interfaces. In: Holzinger A, editor. HCI and Usability for Medicine and Health Care. Berlin, Heidelberg: Springer; 2007:441.
43. van den Berg L, van Beukering M, Velu A, Kok M. The impact of user characteristics on think aloud method results. Evaluation of a mHealth application for pregnancy-related work advice and information. URL: https://scripties.uba.uva.nl/ [accessed 2023-11-13]
44. Hertzum M, Jacobsen NE. The evaluator effect: a chilling fact about usability evaluation methods. Int J Hum-Comput Interact 2003;15(1):183-204. [doi: 10.1207/s15327590ijhc1501_14]

XSL•FO
RenderX

Abbreviations

AIMAC: Italian Association of Cancer Patients, Relatives and Friends
CAPABLE: Cancer Patients Better Life Experience
COREQ: Consolidated Criteria for Reporting Qualitative Research
ICSM: Istituti Clinici Scientifici Maugeri
NKI-AvL: Netherlands Cancer Institute-Antoni van Leeuwenhoek
SUS: System Usability Scale
UCD: user-centered design
UEM: usability evaluation method

Edited by A Mavragani; submitted 28.02.24; peer-reviewed by K Le Du, M Derksen; comments to author 09.09.24; revised version received 26.09.24; accepted 13.11.24; published 23.01.25.

Please cite as:
Glaser SLC, Fraterman I, van Brummelen N, Tibollo V, Del Campo LM, Mallo H, Wilgenhof S, Wilk S, Gisko V, Khadakou V, Cornet R, Ottaviano M, Medlock S
Usability and Usefulness of a Symptom Management Coaching System for Patients With Cancer Treated With Immune Checkpoint Inhibitors: Comparative Mixed Methods Study
JMIR Form Res 2025;9:e57659
URL: https://formative.jmir.org/2025/1/e57659
doi:10.2196/57659
PMID:

©Savannah Lucia Caterina Glaser, Itske Fraterman, Noah van Brummelen, Valentina Tibollo, Laura Maria Del Campo, Henk Mallo, Sofie Wilgenhof, Szymon Wilk, Vitali Gisko, Vadzim Khadakou, Ronald Cornet, Manuel Ottaviano, Stephanie Medlock. Originally published in JMIR Formative Research (https://formative.jmir.org), 23.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Improving Mental Health and Well-Being Through the Paradym App: Quantitative Study of Real-World Data

Athina Marina Metaxa[1], MSc; Shaun Liverpool[2], PhD; Mia Eisenstadt[3], PhD; John Pollard[4], MSc; Courtney Carlsson[4], MBA

[1]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

[2]Faculty of Health, Social Care and Medicine, Edge Hill University, Ormskirk, United Kingdom

[3]Evidence Based Practice Unit, Anna Freud National Centre for Children and Families, University College London, London, United Kingdom

[4]Paradym, London, United Kingdom

Corresponding Author:
Athina Marina Metaxa, MSc
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Observatory Quarter, Woodstock Rd
Oxford, OX2 6GG
United Kingdom
Phone: 30 6930228210
Email: athina.metaxa@hmc.ox.ac.uk

Abstract

Background: With growing evidence suggesting that levels of emotional well-being have been decreasing globally over the past few years, demand for easily accessible, convenient, and affordable well-being and mental health support has increased. Although mental health apps designed to tackle this demand by targeting diagnosed conditions have been shown to be beneficial, less research has focused on apps aiming to improve emotional well-being. There is also a dearth of research on well-being apps structured around users' lived experiences and emotional patterns and a lack of integration of real-world evidence of app usage. Thus, the potential benefits of these apps need to be evaluated using robust real-world data.

Objective: This study aimed to explore usage patterns and preliminary outcomes related to mental health and well-being among users of an app (Paradym; Paradym Ltd) designed to promote emotional well-being and positive mental health.

Methods: This is a pre-post, single-arm evaluation of real-world data provided by users of the Paradym app. Data were provided as part of optional built-in self-assessments that users completed to test their levels of depression (Patient Health Questionnaire-9), anxiety (Generalized Anxiety Disorder Questionnaire-7), life satisfaction (Satisfaction With Life Scale), and overall well-being (World Health Organization-5 Well-Being Index) when they first started using the app and at regular intervals following initial usage. Usage patterns, including the number of assessments completed and the length of time between assessments, were recorded. Data were analyzed using within-subjects *t* tests, and Cohen *d* estimates were used to measure effect sizes.

Results: A total of 3237 app users completed at least 1 self-assessment, and 787 users completed a follow-up assessment. The sample was diverse, with 2000 users (61.8%) being located outside of the United States. At baseline, many users reported experiencing strong feelings of burnout (677/1627, 41.6%), strong insecurities (73/211, 34.6%), and low levels of thriving (140/260, 53.8%). Users also experienced symptoms of depression (mean 9.85, SD 5.55) and anxiety (mean 14.27, SD 6.77) and reported low levels of life satisfaction (mean 12.14, SD 7.42) and general well-being (mean 9.88, SD 5.51). On average, users had been using the app for 74 days when they completed a follow-up assessment. Following app usage, small but significant improvements were reported across all outcomes of interest, with anxiety and depression scores improving by 1.20 and 1.26 points on average, respectively, and life satisfaction and well-being scores improving by 0.71 and 0.97 points, respectively.

Conclusions: This real-world data analysis and evaluation provided positive preliminary evidence for the Paradym app's effectiveness in improving mental health and well-being, supporting its use as a scalable intervention for emotional well-being, with potential applications across diverse populations and settings, and encourages the use of built-in assessments in mental health app research.

(JMIR Form Res 2025;9:e68031) doi:10.2196/68031

XSL•FO
RenderX

KEYWORDS

well-being; awareness; mental health; formative; mobile phone; well-being; apps; quantitative evaluation; real-world data; emotional well-being; pre-post; single arm; quantitative data

Introduction

Background

Emotional well-being has been widely recognized as a crucial factor in predicting overall health and quality of life [1]. Increased emotional well-being has been robustly associated with a lower risk of both physical and mental health problems, according to numerous studies [2,3]. Research findings from different countries indicate that emotional well-being has been declining within the general population over the last few years [4,5], a trend that was first observed following the beginning of the COVID-19 pandemic in 2020. Since then, problems with well-being and general satisfaction with life have intensified globally, and simultaneously, the demand for easily accessible, affordable, and effective support services has increased [4]. While research suggests that one effective way to enhance emotional well-being is by increasing emotional self-awareness, such a goal is not easily obtainable for many people facing mental health and well-being difficulties. Specifically, conventional face-to-face psychotherapeutic treatments are often inaccessible, usually due to high costs, long waiting times, and inflexible treatment schedules that are not accommodating to people's daily lives and work-related commitments [4,6]. As a result, mental health apps have been highlighted as a promising solution for delivering these services, given their affordability, ease of access, scalability, and ability to offer anonymous support [6].

In addition, mental health and well-being apps have also begun to be integrated into clinical practice; some therapists use such apps to enhance in-person treatment and provide patients with a tool they can use outside of face-to-face sessions to gain psychoeducational insights, monitor their symptoms, and reflect on their progress [7]. Therapists have expressed positive views about the integration of online mental health tools into their practice, as they can provide a streamlined way to organize and conduct therapy-related activities, improve patient's accessibility to therapy, and reduce the stigma associated with visiting a mental health professional [8]. Specific types of therapy, such as art therapy, can especially benefit from online mental health tools that enable the creation and sharing of visual arts media, which is a key method to fostering self-expression and well-being for patients [9]. However, concerns have been raised focusing on the high implementation costs associated with the integration of digital apps, both for the therapist and the patient, as well as the privacy, safety, and anonymity of app users, as apps collect sensitive personal information related to patients' mental and physical health and experiences [10]. In addition, clinicians have expressed some concerns about the robustness of the evidence base behind mental health apps, as some apps' content and structure are not clearly based on validated therapeutic techniques and have not been investigated in rigorous research studies [10,11]. However, emotional well-being apps have been increasingly studied and have mixed results with effectiveness. Of note, some studies show that, overall, such apps are valued by users for the privacy, personalization, and instant support that they can provide [12].

Emotional Well-Being

Emotional well-being is a key aspect of mental health that encompasses overall happiness and life satisfaction [13]. Emotional self-awareness, which refers to recognizing and expressing one's emotions as they occur [14], is crucial for achieving emotional well-being, as low levels of emotional awareness have consistently been linked to diminished emotional well-being, reduced quality of life, problems in interpersonal relationships, and a higher risk of mental health disorders like depression and anxiety [15,16]. In contrast, increased emotional awareness has been linked to better emotional regulation, increased positive emotions and quality of life, improvements in social functioning, and reductions in symptoms of depression and anxiety [17]. Based on these findings, both researchers and policymakers have highlighted emotional well-being as a means to improve overall mental and physical health and reduce the costs linked to poor mental health and the treatment of mental health disorders [1,2].

Emotional Well-Being Apps

In recent years, a growing number of apps have focused on the connection between emotional self-awareness and emotional well-being has led to a range of digital interventions that focus on this approach [18,19]. These apps, which can be easily accessed on smartphones and tablets, allow users to improve their emotional well-being at their convenience, offering privacy and flexibility, while they are usually either free or have a much lower annual cost than traditional face-to-face therapy [18,19].

A comprehensive review of 52 well-being apps revealed that most of these tools focus on helping users improve their emotional well-being and emotional awareness primarily through techniques such as mindfulness, meditation, cognitive behavioral therapy, mood tracking, and journaling [20]. Mindfulness-based apps, such as Headspace and Calm (Calm.com Inc), have gained popularity by encouraging users to incorporate meditation and breathing exercises into their daily routines. These practices are designed to heighten emotional awareness and help users manage their thoughts and feelings more effectively [21-23]. Similarly, other apps have integrated cognitive behavioral therapy strategies, which are aimed at helping users challenge and reframe negative thought patterns and engage in positive reappraisal [24,25]. Moreover, mood-tracking features in some apps enable users to record and monitor their emotions over time, often providing personalized suggestions for activities that can help regulate and improve their moods [26,27]. This combination of approaches highlights the diverse methods through which digital tools are addressing emotional well-being.

Despite the variety of techniques used, many well-being apps have been critiqued for not fully incorporating lived experiences or storytelling elements in their design [20]. When these

XSL•FO
RenderX

elements are included, they typically take the form of fictional narratives or real-life testimonials from individuals who have successfully navigated difficult emotional situations or life events [28]. These storytelling techniques can be powerful, offering users relatable examples and motivating them to engage more deeply with the app's content [29]. Research indicates that the incorporation of storytelling in therapeutic interventions can improve patient engagement and destigmatize mental health treatment [30]. Effective storytelling has also been shown to enable the delivery of novel health care interventions and capture the unmet needs of patients across demographic strata and treatment areas [31]. However, the limited use of these approaches suggests that there is untapped potential in using storytelling and lived experiences more extensively. In addition, very few apps encourage users to explore their emotional patterns over time. Such exploration could provide valuable insights into their personality traits and behavioral tendencies, offering a deeper understanding of their emotional landscape beyond the immediate moment [20].

Given the rapid growth of well-being apps and the limited research on those that use storytelling, lived experiences, or focus on emotional patterns, there is a clear opportunity for further exploration, which could help identify new strategies to increase user engagement and satisfaction with digital well-being tools, leading to better outcomes in emotional health management. As digital interventions continue to evolve, exploring these areas could play a crucial role in developing more holistic and user-centered tools that better meet the diverse needs of those seeking to improve their emotional well-being [20,32].

The Paradym App

The Paradym app was created in response to the increasing demand for scalable mental health solutions, driven by global challenges such as the COVID-19 pandemic and increased societal awareness of mental well-being. Paradym addresses these challenges by providing an accessible, evidence-based platform to help users increase their emotional awareness by exploring their emotional patterns, ultimately aiming to enhance emotional well-being, boost self-awareness, and improve overall life satisfaction [32]. The Paradym app introduces users to psychoeducational content through storytelling, covering key aspects of life such as love and relationships, body image, work and success, and identity. In addition, it incorporates elements of lived experience, which involves the user's personal, first-hand experiences and the significance these experiences bring to their current situations [33].

Paradym's unique use of emotional pattern recognition and storytelling aims to offer a novel approach within the field of digital mental health and use this combination of lived experience and evidence-based interventions to boost user engagement and emotional awareness. Finally, in contrast to many of the widely used mental health and well-being apps available, Paradym incorporates validated mental health questionnaires in its interface, such as the Satisfaction with Life Scale (SWLS), which allows users to monitor their mood and well-being as they go through the app's contents. To ensure effective implementation and guide further development, the app's design considered the principles of the Technology Acceptance Model [34]. This model suggests that users' engagement with a new digital tool is influenced by their perceptions of its usefulness and ease of use. In addition, the model recommends monitoring user anxiety, as previous research has identified anxiety as an external factor that can impact technology acceptance [35]. A 2021 acceptability and engagement study of the Paradym app [32] revealed that the majority of users found the app easy to use (64%) and used it daily (79%), and post hoc statistical analyses showed a significant increase in the well-being measures (d=0.73 for SWLS scores and d=0.51 for World Health Organization-5 Well-Being Index [WHO-5] scores) and reduction in symptoms of depression (d=–0.38 for Patient Health Questionnaire-9 [PHQ-9] scores) following app use. Importantly, no adverse effects were reported, providing encouraging preliminary evidence for the Paradym app's effectiveness and usability.

Aim of This Study

The aim of this study was to expand Paradym's evidence base by evaluating the app's effectiveness in addressing symptoms of anxiety and depression and promoting overall well-being and life satisfaction in a real-world context.

Methods

Participants and Procedure

Real-world data were collected from an international pool of Paradym app users, who had downloaded and were using the mobile app from the App Store or Google Play. The participant pool for this study consisted of the total of Paradym users who had completed the app's built-in assessments and questionnaires (Generalized Anxiety Disorder-7 [GAD-7], PHQ-9, WHO-5, and SWLS) at least once. These questionnaires are an optional part of the Paradym app experience, and users can complete them as many times as they wish while using the app to track their progress in their own well-being and life satisfaction. Owing to the fact that assessments are part of the app intervention a specific recruitment procedure was not required to obtain participants' data for this study. Instead, before completing the questionnaires, all Paradym app users were informed by a written statement that their data would be used for research purposes related to the effects of Paradym on anxiety, depression, general well-being, and satisfaction with life. Users provided consent when signing up to the app and informed that surveys were part of the app user experience. Following the principle of data minimization, users did not provide detailed demographic data beyond the country of residence. The data that informed this study from the in-app assessments was collected between March 2023 and June 2024. Participants did not receive any compensation for filling out the questionnaires as the measures were part of the app intervention.

The Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [36] was used to track the details of this study's methodology, this provides further detail on recruitment and administration and is provided in in Multimedia Appendix 1.

XSL•FO
RenderX

Intervention Overview

Paradym is an app designed to promote emotional well-being among the general population, specifically targeting adults aged 18 years and older. Its primary goal is to help users cultivate greater emotional awareness and achieve higher levels of life satisfaction. In its early development, Paradym was funded through crowdfunding, a choice made to avoid the influence of commercial interests during the initial stages of creating the intervention.

Paradym was developed based on evidence-based techniques and was tailored to address themes like success, body, identity, love, and relationships, which were identified by users as important. The platform's development involved clinical psychologists, coaches, and researchers who collaborated on crafting the content. This included selecting and reviewing psychoeducational materials, appropriate evidence-based strategies, and various activities like journaling and note-taking [32].

The foundation of Paradym is an integrative theoretical approach, enriched by therapeutic strategies like acceptance and commitment therapy and schema therapy, which shape the lived experience and storytelling aspects of the platform. User feedback further refined these approaches, leading to the integration of additional theory-driven strategies to generate content that resonated with the needs expressed by users. The core strategies of Paradym encompass these elements, providing a comprehensive toolkit for personal development.

Digital Lessons

Paradym's psychoeducational digital lessons are organized around 5 central themes: awareness, success, love, identity, and body. These areas were selected based on feedback gathered from users during the initial phase of development [32]. The educational content covers several topics, including building emotional awareness, aiding users in recognizing and understanding their emotions, and helping them identify and reinforce their personal values.

Each lesson starts by presenting the educational concept through narratives based on real-life experiences. The use of personal stories is intended to make the material more accessible and relatable, enhance empathy, and encourage users to see themselves in the stories being told. This storytelling approach is known to increase user engagement with both psychological interventions and digital applications. The lessons are available in various formats, including text chapters, audio recordings, and videos, catering to different learning preferences.

Emotional Patterns

At the conclusion of each digital lesson, users are prompted to examine their own emotional patterns in context with the app's material, incorporating elements from schema therapy, in order to enhance emotional insight. To facilitate this, the app presents various "emotional patterns" and encourages users to reflect on which pattern they resonate with most at the time of finishing the lesson. This allows users to track and potentially modify their emotional patterns as they progress through the lessons.

Reflections

Users receive reflections through push notifications, designed to enhance their daily interaction with the app by introducing a new reflection each day. Research has highlighted numerous advantages to such reflection exercises, including aiding individuals in maintaining focus on tasks amidst stressful situations, and reducing tendencies to ruminate, which is often associated with internal struggles and social conflicts. In addition, the use of notifications has been proven to boost user interaction with apps [32], and studies suggest that mobile apps delivering psychological insights are most effective when coupled with proactive user involvement.

Data Collection Tools

Demographics

In order to maintain users' anonymity and confidentiality, only information regarding their geographic location was collected.

Well-Being Measure

The WHO-5 was used to assess well-being in this study. The WHO-5 consists of a brief assessment (under 1 minute) of well-being over a 2-week period [37], in which individuals are asked to indicate how they felt over the past 2 weeks for each of 5 statements using a 6-point Likert scale, ranging from 0="at no time" to 5="all of the time." A higher total score on the WHO-5 indicates a higher level of well-being. The WHO-5 is derived from a 28-item questionnaire based on the Zung scales for depression, distress, and anxiety, as well as from the General Health Questionnaire and the Psychological General Well-Being Scale [38,39]. The WHO-5 has been validated as a measure of depression in both adolescents and older adults, with high measurement invariance [38,39].

Satisfaction With Life Measure

The SWLS [40] is a brief (approximately 1 minute) assessment consisting of 5 items designed to measure the global cognitive assessment of satisfaction with life. The SWLS has been shown to have very high construct validity, with Cronbach α=0.85-0.87 [41] and moderately high reliability (Cronbach α=0.78) [42]. A higher score on the SWLS indicates a higher level of life satisfaction.

Anxiety Measure

The GAD-7 is a 7-item, brief, widely-used clinical measure that assesses the presence and severity of generalized anxiety disorder. The self-report scale asks how often during the last 2 weeks individuals experienced symptoms of generalized anxiety disorder. Total scores range from 0-21, with cutoff scores of 5, 10, and 15 indicating mild, moderate, and severe anxiety, respectively. Increasing scores on the GAD-7 are strongly associated with greater functional impairment in real-world settings. Sample items are rated from 0 (not at all) to 3 (nearly every day). Scores over 10 have good sensitivity (89%) and specificity (82%) for diagnosis of generalized anxiety disorder by interview, and the scale has high internal reliability, as suggested by Cronbach α=0.92 [43]. The scale has been widely used and considered a valid and reliable screening tool in previous research, presenting good reliability, factorial, and

XSL•FO
RenderX

concurrent validity [44]. A higher GAD-7 score indicates more severe anxiety.

Depression Measure

The PHQ-9 is a widely used depression scale that scores each of the 9 Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition criteria as 0 (not at all) to 3 (nearly every day). The PHQ-9 has been validated for use in primary care [45]. It is not a screening tool for depression but can monitor the severity of symptoms and response to treatment. Scores over 10 have good sensitivity (88%) and specificity (88%) for the diagnosis of major depression by interview. It has high internal reliability, and Cronbach α=0.89 [46]. The construct validity of the PHQ-9 is also high for community and clinical samples [47]. A high score obtained from the PHQ-9 is indicative of a severe level of depression.

Data Analysis

Available demographic information (ie, geographic location) and emotional patterns (ie, burnout, sense of thriving, and insecurity) for app users were descriptively analyzed using count and percentages (n, %). Engagement was calculated based on the number of times users visited the app and the length of time the user had been registered on the Paradym app. The results for primary outcomes were largely descriptive in nature, and we evaluated success based on improvements in scores over time. Within-subjects *t* tests were conducted to estimate differences between time 1 (ie, the start of app use) and follow-up. A *P* value of <.001 was considered significant and Cohen *d* estimates were used to measure effect sizes. All statistical analyses were conducted using the SPSS (IBM Corp) software [48].

Ethical Considerations

This study was conducted with a strong commitment to ethical principles, ensuring the protection of participants' rights, safety, dignity, and well-being [49]. Data were gathered remotely from individuals who were not part of a clinical setting, and who freely gave their informed consent, allowing their anonymized and nonidentifiable in-app data to be used for research purposes. The research team made their contact information available, offering support and directing participants to additional well-being resources as needed. Participants did not receive any compensation for filling out the questionnaires as the measures were part of the app intervention.

The study was considered to be user testing or service evaluation, and as such, it was not subject to the usual requirements for formal ethical approval by the National Health Service Health Research Authority, aligning with practices in similar studies [50,51]. The certificate of exemption is provided in Multimedia Appendix 2. Since the evaluation included a nonclinical group, it was exempt from registration in a public trials registry, similar to other studies where formal ethical oversight was deemed unnecessary due to the low-risk nature of the research and its nonclinical focus [52,53]. The detailed Terms and Conditions regulations are provided in Multimedia Appendix 3; through these, users of the app agree to their data being used in research.

Results

Descriptive Data

A total of 3237 unique users of the Paradym app provided data included in this analysis, between March of 2023 and June of 2024. User uniqueness was determined based on the principal investigator and cookie data. The data were anonymized to protect participants' identities, so details on their age, gender, and race were not recorded. However, the users had the option to provide information on their geographic location (Table 1).

Table 1. Baseline demographic information of Paradym app users whose data were used in this real-world evaluation (N=3237).

Country	Participants, n (%)
United States	1237 (38.2)
Philippines	547 (16.9)
Great Britain	210 (6.5)
Canada	204 (6.3)
Australia	139 (4.3)
Other	900 (27.8)

Users had the option to self-report their perceived levels of burnout, sense of thriving, and level of insecurity before starting to use the app. A total of 1627 users provided information on their level of burnout, and almost half (n=677, 41.6%) reported experiencing strong burnout, while 641 (39.4%) experienced moderate levels of burnout and 309 (19%) experienced low levels of burnout. A total of 260 users rated their sense of thriving, and the majority reported either slight (n=140, 53.8%) or moderate (n=104, 40%) thriving, with only 16 (6.2%) reporting a strong sense of thriving. Among the 211 users who rated their level of insecurity, 34.6% (n=73) and 29.4% (n=62) reported strong and moderate insecurity, respectively, with 36.0% reporting slight insecurity (n=76; Table 2).

XSL•FO
RenderX

Table 2. Baseline data on Paradym app users' levels of burnout (n=1627), thriving (n=260), and insecurity (n=211).

Domain	Participants, n (%)
Burnout	
Slight	309 (19)
Moderate	641 (39.4)
Strong	677 (41.6)
Thriving	
Slight	140 (53.8)
Moderate	104 (40)
Strong	16 (6.2)
Insecurity	
Slight	76 (36)
Moderate	62 (29.4)
Strong	73 (34.6)

Efficacy Data

After downloading the app, the first-time users completed questionnaires assessing their levels of anxiety (GAD-7), depression (PHQ-9), life satisfaction (SWLS), and well-being (WHO-5). App usage among the participants in this study varied, but on average, users had been registered for 74 days at the time the data were collected and had visited the Paradym app 10 times, with over 30% of users reporting that they had been using the app somewhat frequently or very often. Of the 3237 app users, just over 20% (n=787) of app users returned to the app more than once to complete mental health and well-being assessments. This was also a significant indicator of those who used the app more frequently (119 times vs 379 times) and had been using the app over a longer period of time (mean 215 days vs mean 36 days). Notably, those who returned to monitor their mental health and well-being indicators at least once showed significant improvements in all 4 mental health and well-being measures. Using available data, within-subjects *t* tests were applied, and the improvement in anxiety (t_{521}=5.288, *P*<.001), depression (t_{521}=6.667, *P*<.001), satisfaction with life (t_{521}=3.191, *P*<.001), and general well-being (t_{521}=4.947, *P*<.001) scores were all significant at follow-up (Table 3; Figure 1).

Table 3. Baseline (first questionnaire completion) and follow-up (subsequent questionnaire completion) data on Paradym app users' levels of anxiety, depression, satisfaction with life, and general well-being (n=787).

Outcome measure	Baseline, mean (SD)	Follow-up, mean (SD)	Average baseline to follow-up score change	Effect size (Cohen *d*)
GAD-7[a]	14.27 (6.77)	13.07 (6.63)	1.20	0.231
PHQ-9[b]	9.85 (5.55)	8.6 (5.55)	1.26	0.292
SWLS[c]	12.14 (7.42)	12.85 (7.46)	0.71	0.140
WHO-5[d]	9.88 (5.51)	10.85 (5.5)	0.97	0.217

[a]GAD-7: Generalized Anxiety Disorder-7.

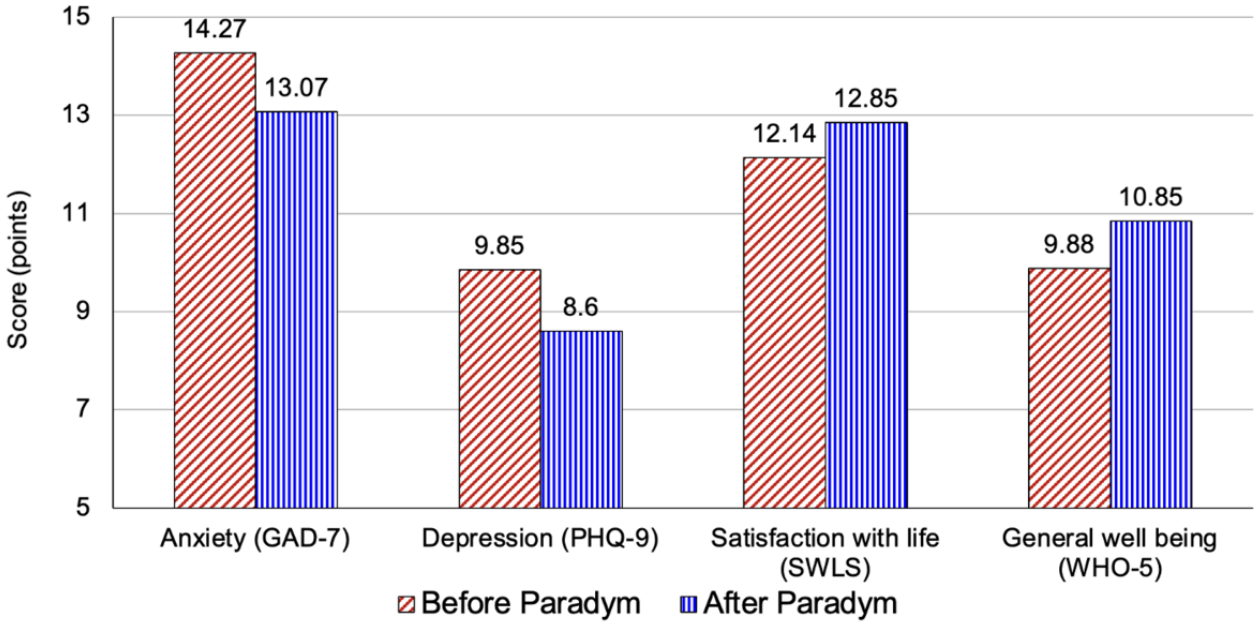
[b]PHQ-9: Patient Health Questionnaire-9.

[c]SWLS: Satisfaction With Life Scale.

[d]WHO-5: World Health Organization-5 Well-Being Index.

XSL•FO
RenderX

Figure 1. Baseline and follow-up data on participants' levels of anxiety, depression, satisfaction with life, and general well-being (n=787). GAD-7: Generalized Anxiety Disorder-7; PHQ-9: Patient Health Questionnaire-9; SWLS: Satisfaction with Life Scale; WHO-5: World Health Organization-5 Well-Being Index.



Discussion

Principal Findings

The aim of this study was to investigate the effects of the well-being app Paradym on the users' levels of anxiety, depression, satisfaction with life, and general well-being. The findings indicated that users who accessed the app's contents and completed the built-in mental health and well-being assessments (N=3237) generally reported problems with burnout, insecurity, and a lack of thriving in their lives. In addition, the app's user base scored highly on scales of depression (PHQ-9) and anxiety (GAD-7) but reported low levels of life satisfaction (SWLS) and well-being (WHO-5). A total of the 787 users chose to complete follow-up assessments of these measures, in which significant decreases in average scores of depression and anxiety were observed. In addition, both life satisfaction and well-being scores increased significantly, providing real-world data on Paradym's benefits for users with a relatively large sample. However, as this was not a controlled study, it is not possible to make any causal claims or explicitly attribute the findings to the use of the app. Nonetheless, these preliminary findings are overall positive and indicate that Paradym's digital format allows for wide accessibility and scalability, positioning it as an effective tool for both individual and public mental health strategies.

Despite these positive findings, the number of users who opted to not complete the follow-up mental health and well-being assessments was high (2450/3237), which is commonly observed in studies of mental health and well-being apps [54]. However, this dropout rate only refers to users who did not use the app's built-in questionnaires. Thus, it is possible that many of these users continued using the app and engaging with its contents, but did not feel the need to assess their progress with the app's built-in questionnaires. Alternatively, users who experienced higher levels of distress or more severe depression and anxiety symptoms found it more difficult to engage with the app or respond to the outcome measures, as has been observed in other similar studies [55]. In addition, mental health apps generally have high dropout rates, and some studies have found that once users learn a skill or knowledge from a particular app, they stop using it [56]. Further research would be required to understand the factors that influence users to complete the outcome measures beyond the first measure.

Comparison With Previous Work

This study's findings on Paradym's effectiveness in improving mental health and well-being are in line with those observed in studies of other similar apps and add to the growing literature supporting apps' benefits as mental health tools [57,58].

A significant decrease in symptoms of anxiety and depression has been observed in multiple other studies of popular apps, such as Headspace and Calm, using the same questionnaires as this study [22,59-61]. The same is true for studies examining well-being and life satisfaction [15,62]. In addition, the effect sizes observed for depression and anxiety are consistent with those reported in other studies and meta-analyses of mental health apps [21-23,58,60-63].

However, very few studies have examined the effects that multiple compared to stand-alone assessments or the timing of those assessments could have on the participants, and we did not identify any studies that studied this association using naturalistic data through assessments that were built into the apps' experience.

Strengths and Limitations

A key strength of the present analysis rests on its use of a large sample of users under real-world conditions from a diverse user

XSL•FO
RenderX

base over a long period of time. Naturalistic data were collected, as users were completing assessments as part of the app experience, and not as part of a controlled study environment. Such analyses could prompt app developers to integrate rigorous data collection processes within their apps' interface and core features. In addition, the high levels of burnout, insecurity, anxiety, and depression, and low levels of life satisfaction, thriving, and general well-being reported by users at baseline act as a strong indicator that Paradym reached users who were seeking support for their mental health and well-being.

Since no personal data related to the users' gender, race, or age were collected by the app, in order to ensure users' anonymity and confidentiality and follow the data minimization principle some limitations in our data analysis are acknowledged. Due to the absence of such detailed demographic data**,** subgroup analyses based on such characteristics were not conducted as part of this study. Subgroup analyses of effectiveness could provide important insights into how demographic factors affect the outcomes of interest, as has been observed in other studies [64,65].

Further, this study's single-arm design did not allow for the inclusion of comparators, which could limit the findings' robustness. Change scores were small but significant so this should be interpreted with caution. The change scores and resulting effect sizes were lower than those commonly accepted to indicate a clinically meaningful difference (approximately 4 points on the GAD-7 [66] and 3 points on the PHQ-9 [67]). However, it should be noted that these thresholds for clinically meaningful change are based on studies of clinical populations, including patients diagnosed with anxiety or depression who had mild to severe symptoms [68]; Paradym is not exclusively aimed at diagnosed patients, and thus, it is expected that the aforementioned clinically meaningful thresholds are not directly applicable to the app's users, who may have benefited substantially from smaller improvements in questionnaire scores [69].

Finally, while the number of users who dropped out in regards to the questionnaire filling was captured in the data, the exact patterns of app usage were not measured. For instance, some users may have stopped completing the questionnaires, but they maintained a subscription and continued using the app and engaging with its content daily. Collecting more specific data on user engagement patterns would allow for a clearer view of which features of the Paradym app users found more useful, beyond the questionnaires.

Future Directions

Future studies ought to gain more understanding of the mechanisms underpinning the observed reductions in anxiety and depression symptoms and the increase in life satisfaction and well-being that Paradym users experience. Assessments to capture the specific components of the app's content and structure that have a greater impact on users could be incorporated into future analyses of user data.

In addition, more detailed information on users' experiences with burnout, low levels of thriving, anxiety, and depression would help us better understand the needs of the app's user base, and tailor the app accordingly. For example, questions exploring the specific domains users struggle with (eg, work and personal relationships) could shed light on what kind of content they would benefit the most from (eg, self-assessments, reflection exercises, and guided lessons).

Paradym's benefits for mental health and well-being could be investigated through randomized controlled trials in the future, in order to understand whether improvements in the users' mental health and well-being are attributed to ongoing app usage. Comparison groups using similar apps or face-to-face therapy could also be introduced to enhance the results' validity, robustness, and generalizability.

While not collecting demographic data for this study was a deliberate choice made to preserve the users' anonymity, future studies could consider collecting data related to age, gender, and race, in order to allow for subgroup analyses of Paradym's efficacy based on these characteristics. This will be a significant next step in order to answer questions around "what worked, for whom, and under which circumstances."

Further studies could also explore Paradym's integration within broader systems, such as health care or educational settings, to evaluate the app's effectiveness at scale. For instance, patients could be offered access to Paradym if they are placed on a waiting list for in-person therapy, or if they do not feel comfortable initiating face-to-face therapy. In addition, future research could explore the app's impact across different populations and contexts, including workplace wellness programs or health care partnerships.

Finally, this study's findings indicate that the anxiety, depression, well-being, and life satisfaction data collection tools incorporated within the app are beneficial to the users. As a result, additional tools, including qualitative data collection measures, could be introduced to the app's interface, allowing participants to provide valuable and detailed information that could provide insight into their emotional and mental health journey while using the app, as well as a way for them to track their own progress. Feedback measures could also be added to capture the reasons why some participants decide to complete the existing assessments more or less frequently, and why some opt to not complete them at all. This would help to elucidate potential barriers to user engagement with these data collection tools, and whether these are related to the app's content and structure or to external factors that prevent users from fully engaging with Paradym's user journey.

Conclusions

Overall, this analysis highlighted the value of conducting formative research on mental health and well-being apps. The findings suggest that built-in assessments in a mental health app can lead to the collection of valid real-world data from users and promote organic engagement with the app's features. Among users who completed follow-up assessments in the Paradym app, levels of anxiety and depression decreased significantly, while life satisfaction and general well-being improved significantly. This study provided positive evidence for Paradym's effectiveness in improving mood and well-being and highlighted the value of the app's content, functionality,

XSL•FO
RenderX

and underlying principles. These encouraging results suggest the scope for further research on Paradym and the integration of built-in assessments in mental health apps.

Acknowledgments

The authors thank the participants who took the time to participate in this study. This study was funded by Paradym Ltd. Apart from commissioning a call for evidence in this area, CC and JP had no role in data collection, data analysis, or data interpretation. However, CC and JP were responsible for the oversight of the project and were involved in the study design and editing of the drafts of the manuscript.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

AMM, ME, CC, JP, and SL conceptualized the study and contributed to the study design. All authors contributed to user testing. To reduce the risk of bias in favor of the intervention, SL independently analyzed the data and interpreted the output with input from ME and AMM. All authors contributed to the writing and editing of the manuscript. All authors were involved in reviewing, editing, and final approval of the manuscript before submission.

Conflicts of Interest

CC is the founder and Chief Executive Officer of Paradym Ltd, and JP is the Head of Development lead at Paradym Ltd. AMM, ME, and SL were commissioned as external consultants during the analysis and write-up of this evaluation.

Multimedia Appendix 1
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[DOCX File , 19 KB - formative_v9i1e68031_app1.docx]

Multimedia Appendix 2
NHS Health Research Authority exemption certificate.
[PDF File (Adobe PDF File), 175 KB - formative_v9i1e68031_app2.pdf]

Multimedia Appendix 3
Captures of the Paradym app’s interface, including the onboarding screen, the privacy policy, and the Terms and Conditions.
[PDF File (Adobe PDF File), 527 KB - formative_v9i1e68031_app3.pdf]

References

1. Agteren J, Iasiello M. Advancing our understanding of mental well-being and mental health: the call to embrace complexity over simplification. Australian Psychologist 2021;55(4):307-316. [doi: 10.1111/ap.12440]
2. Huppert F. Psychological well‐being: evidence regarding its causes and consequences. Applied Psych Health & Well 2009;1(2):137-164. [doi: 10.1111/j.1758-0854.2009.01008.x]
3. Chida Y, Steptoe A. Positive psychological well-being and mortality: a quantitative review of prospective observational studies. Psychosom Med 2008;70(7):741-756. [doi: 10.1097/PSY.0b013e31818105ba] [Medline: 18725425]
4. Kaufman KR, Petkova E, Bhui KS, Schulze TG. A global needs assessment in times of a global crisis: world psychiatry response to the COVID-19 pandemic. BJPsych Open 2020;6(3):e48 [FREE Full text] [doi: 10.1192/bjo.2020.25] [Medline: 32250235]
5. Salari N, Hosseinian-Far A, Jalali R, Vaisi-Raygani A, Rasoulpoor S, Mohammadi M, et al. Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. Global Health 2020;16(1):57 [FREE Full text] [doi: 10.1186/s12992-020-00589-w] [Medline: 32631403]
6. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
7. Okoro YO, Ayo-Farai O, Maduka CP, Okongwu CC, Sodamade OT. The role of technology in enhancing mental health advocacy: a systematic review. Int. j. appl. res. soc. sci 2024;6(1):37-50. [doi: 10.51594/ijarss.v6i1.690]
8. Shojaei F, Osorio Torres J, Shih PC. Exploring the integration of technology in art therapy: insights from interviews with art therapists. Art Therapy 2024:1-7. [doi: 10.1080/07421656.2024.2383826]
9. Zubala A, Kennell N, Hackett S. Art therapy in the digital world: an integrative review of current practice and future directions. Front Psychol 2021;12:595536 [FREE Full text] [doi: 10.3389/fpsyg.2021.600070] [Medline: 33897522]

XSL•FO
RenderX

10. Berardi C, Antonini M, Jordan Z, Wechtler H, Paolucci F, Hinwood M. Barriers and facilitators to the implementation of digital technologies in mental health systems: a qualitative systematic review to inform a policy framework. BMC Health Serv Res 2024;24(1):243 [FREE Full text] [doi: 10.1186/s12913-023-10536-1] [Medline: 38408938]
11. Smith KA, Blease C, Faurholt-Jepsen M, Firth J, Van Daele T, Moreno C, et al. Digital mental health: challenges and next steps. BMJ Ment Health 2023;26(1):e300670 [FREE Full text] [doi: 10.1136/bmjment-2023-300670] [Medline: 37197797]
12. Pieritz S, Khwaja M, Faisal A, Matic A. Personalised recommendations in mental health apps: the impact of autonomy and data sharing. 2021 Presented at: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems; 2021 May 07; Yokohama Japan. [doi: 10.1145/3411764.3445523]
13. Keyes CLM. Mental health as a complete state: how the salutogenic perspective completes the picture. In: Bauer GF, Hammig O, editors. Bridging Occupational, Organizational and Public Health. New York: Springer; 2014.
14. Agnoli S, Mancini G, Andrei F, Trombini E. The relationship between trait emotional intelligence, cognition, and emotional awareness: an interpretative model. Front Psychol 2019;10:1711 [FREE Full text] [doi: 10.3389/fpsyg.2019.01711] [Medline: 31417455]
15. Hendryx MS, Haviland MG, Shaw DG. Dimensions of alexithymia and their relationships to anxiety and depression. J Pers Assess 1991;56(2):227-237. [doi: 10.1207/s15327752jpa5602_4] [Medline: 2056418]
16. Kauer S. School of Psychological Science, The University of Melbourne. Emotional self-awareness and depressive symptoms: an investigation of an early intervention mobile phone self-monitoring program for adolescents. URL: https://minerva-access.unimelb.edu.au/handle/11343/37899 [accessed 2025-12-01]
17. Guendelman S, Medeiros S, Rampes H. Mindfulness and emotion regulation: insights from neurobiological, psychological, and clinical studies. Front Psychol 2017;8:220 [FREE Full text] [doi: 10.3389/fpsyg.2017.00220] [Medline: 28321194]
18. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using health and well-being apps for behavior change: a systematic search and rating of apps. JMIR mHealth uHealth 2019;7(7):e11926 [FREE Full text] [doi: 10.2196/11926] [Medline: 31274112]
19. Roffarello AM, Russis LD. The race towards digital wellbeing. 2019 Presented at: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; 2019 May 9; Glasgow Scotland UK p. 4-9. [doi: 10.1145/3290605.3300616]
20. Eisenstadt M, Liverpool S, Infanti E, Ciuvat RM, Carlsson C. Mobile apps that promote emotion regulation, positive mental health, and well-being in the general population: systematic review and meta-analysis. JMIR Ment Health 2021;8(11):e31170 [FREE Full text] [doi: 10.2196/31170] [Medline: 34747713]
21. Avalos LA, Aghaee S, Kurtovich E, Quesenberry C, Nkemere L, McGinnis MK, et al. A mobile health mindfulness intervention for women with moderate to moderately severe postpartum depressive symptoms: feasibility study. JMIR Ment Health 2020;7(11):e17405 [FREE Full text] [doi: 10.2196/17405] [Medline: 33180028]
22. Comtois KA, Mata-Greve F, Johnson M, Pullmann MD, Mosser B, Arean P. Effectiveness of mental health apps for distress during COVID-19 in US unemployed and essential workers: remote pragmatic randomized clinical trial. JMIR mHealth uHealth 2022;10(11):e41689 [FREE Full text] [doi: 10.2196/41689] [Medline: 36191176]
23. Kubo A, Aghaee S, Kurtovich EM, Nkemere L, Quesenberry CP, McGinnis MK, et al. mHealth mindfulness intervention for women with moderate-to-moderately-severe antenatal depressive symptoms: a pilot study within an integrated health care system. Mindfulness (N Y) 2021;12(6):1387-1397 [FREE Full text] [doi: 10.1007/s12671-021-01606-8] [Medline: 33723491]
24. Bakker D, Rickard N. Engagement with a cognitive behavioural therapy mobile phone app predicts changes in mental health and wellbeing: MoodMission. Aust Psychol 2019;54(4):245-260. [doi: 10.1111/ap.12383]
25. Birney AJ, Gunn R, Russell JK, Ary DV. MoodHacker mobile web app with email for adults to self-manage mild-to-moderate depression: randomized controlled trial. JMIR mHealth uHealth 2016;4(1):e8 [FREE Full text] [doi: 10.2196/mhealth.4231] [Medline: 26813737]
26. Throuvala MA, Griffiths MD, Rennoldson M, Kuss DJ. Mind over matter: testing the efficacy of an online randomized controlled trial to reduce distraction from smartphone use. Int J Environ Res Public Health 2020;17(13):4842 [FREE Full text] [doi: 10.3390/ijerph17134842] [Medline: 32635650]
27. Bakker D, Rickard N. Engagement in mobile phone app for self-monitoring of emotional wellbeing predicts changes in mental health: MoodPrism. J Affect Disord 2018;227:432-442. [doi: 10.1016/j.jad.2017.11.016] [Medline: 29154165]
28. Daugherty DA, Runyan JD, Steenbergh TA, Fratzke BJ, Fry BN, Westra E. Smartphone delivery of a hope intervention: another way to flourish. PLoS One 2018;13(6):e0197930 [FREE Full text] [doi: 10.1371/journal.pone.0197930] [Medline: 29856805]
29. Zhu D, Al Mahmud A, Liu W. Digital storytelling intervention for enhancing the social participation of people with mild cognitive impairment: co-design and usability study. JMIR Aging 2024;7:e54138 [FREE Full text] [doi: 10.2196/54138] [Medline: 38231541]
30. Davidson T, Moreland A, Bunnell BE, Winkelmann J, Hamblen JL, Ruggiero KJ. Reducing stigma in mental health through digital storytelling. In Deconstructing stigma in mental health. USA: IGI Global; 2018.
31. Mehta J, Long E, Bynagari V, Shojaei F, Shojaei F, O'Brien ARW, et al. Creating demand for unmet needs: agile storytelling. Front Health Serv 2024;4:1376695 [FREE Full text] [doi: 10.3389/frhs.2024.1376695] [Medline: 39544454]

XSL•FO
RenderX

32. Eisenstadt A, Liverpool S, Metaxa AM, Ciuvat RM, Carlsson C. Acceptability, engagement, and exploratory outcomes of an emotional well-being app: mixed methods preliminary evaluation and descriptive analysis. JMIR Form Res 2021;5(11):e31064 [FREE Full text] [doi: 10.2196/31064] [Medline: 34569466]
33. Lindseth A, Norberg A. A phenomenological hermeneutical method for researching lived experience. Scand J Caring Sci 2004;18(2):145-153. [doi: 10.1111/j.1471-6712.2004.00258.x] [Medline: 15147477]
34. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989;13(3):319. [doi: 10.2307/249008]
35. Dönmez-Turan A, Kır M. User anxiety as an external variable of technology acceptance model: a meta-analytic study. Procedia Computer Science 2019;158:715-724. [doi: 10.1016/j.procs.2019.09.107]
36. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
37. Johansen KS. The use of well-being measures in primary health care/The DepCare Project. World Health Organization, Regional Office for Europe. URL: https://www.euro.who.int/__data/assets/pdf_file/0016/130750/E60246.pdf [accessed 2025-01-13]
38. Krieger T, Zimmermann J, Huffziger S, Ubl B, Diener C, Kuehner C, et al. Measuring depression with a well-being index: further evidence for the validity of the WHO well-being index (WHO-5) as a measure of the severity of depression. J Affect Disord 2014;156:240-244. [doi: 10.1016/j.jad.2013.12.015] [Medline: 24412323]
39. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 well-being index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176 [FREE Full text] [doi: 10.1159/000376585] [Medline: 25831962]
40. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. J Pers Assess 1985;49(1):71-75. [doi: 10.1207/s15327752jpa4901_13] [Medline: 16367493]
41. The satisfaction with life scale examining construct validity. Heerlen: Statistics Netherlands; 2012.
42. Vassar M. A note on the score reliability for the satisfaction with life scale: an RG study. Soc Indic Res 2007;86(1):47-57. [doi: 10.1007/s11205-007-9113-7]
43. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
44. Sousa TV, Viveiros V, Chai MV, Vicente FL, Jesus G, Carnot MJ, et al. Reliability and validity of the Portuguese version of the Generalized Anxiety Disorder (GAD-7) scale. Health Qual Life Outcomes 2015;13:50 [FREE Full text] [doi: 10.1186/s12955-015-0244-2] [Medline: 25908249]
45. Mazzocco PJ, Green MC, Sasota JA, Jones NW. This story is not for everyone: transportability and narrative persuasion. Social Psychological and Personality Science 2010;1(4):361-368. [doi: 10.1177/1948550610376600]
46. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
47. Kocalevent R, Hinz A, Brähler E. Standardization of the depression screener patient health questionnaire (PHQ-9) in the general population. Gen Hosp Psychiatry 2013;35(5):551-555. [doi: 10.1016/j.genhosppsych.2013.04.006] [Medline: 23664569]
48. IBM SPSS Statistics for Windows, version 27. URL: https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-27 [accessed 2021-05-21]
49. Koocher G, Keith-Spiegel P. Ethics in Psychology and the Mental Health Professions: Standards and Cases (3rd ed). New York: Oxford University Press; 2008.
50. Ly KH, Dahl J, Carlbring P, Andersson G. Development and initial evaluation of a smartphone application based on acceptance and commitment therapy. Springerplus 2012;1:11 [FREE Full text] [doi: 10.1186/2193-1801-1-11] [Medline: 23961342]
51. Morris ME, Kathawala Q, Leen TK, Gorenstein EE, Guilak F, Labhard M, et al. Mobile therapy: case study evaluations of a cell phone application for emotional self-awareness. J Med Internet Res 2010;12(2):e10 [FREE Full text] [doi: 10.2196/jmir.1371] [Medline: 20439251]
52. Ahtinen A, Mattila E, Välkkynen P, Kaipainen K, Vanhala T, Ermes M, et al. Mobile mental wellness training for stress management: feasibility and design implications based on a one-month field study. JMIR mHealth uHealth 2013;1(2):e11 [FREE Full text] [doi: 10.2196/mhealth.2596] [Medline: 25100683]
53. Ly KH, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. Internet Interv 2017;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
54. Schwartz K, Ganster FM, Tran US. Mindfulness-based mobile apps and their impact on well-being in nonclinical populations: systematic review of randomized controlled trials. J Med Internet Res 2023;25:e44638 [FREE Full text] [doi: 10.2196/44638] [Medline: 37540550]
55. Lui JH, Marcus DK, Barry CT. Evidence-based apps? A review of mental health mobile applications in a psychotherapy context. Professional Psychology: Research and Practice 2017;48(3):199-210. [doi: 10.1037/pro0000122]
56. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]

XSL•FO
RenderX

57. Omylinska-Thurston J, Aithal S, Liverpool S, Clark R, Moula Z, Wood J, et al. Digital psychotherapies for adults experiencing depressive symptoms: systematic review and meta-analysis. JMIR Ment Health 2024;11:e55500 [FREE Full text] [doi: 10.2196/55500] [Medline: 39348177]
58. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
59. Horwitz AG, Mills ED, Sen S, Bohnert ASB. Comparative effectiveness of three digital interventions for adults seeking psychiatric services: a randomized clinical trial. JAMA Netw Open 2024;7(7):e2422115. [doi: 10.1001/jamanetworkopen.2024.22115] [Medline: 39023893]
60. Strauss C, Dunkeld C, Cavanagh K. Is clinician-supported use of a mindfulness smartphone app a feasible treatment for depression? A mixed-methods feasibility study. Internet Interv 2021;25:100413 [FREE Full text] [doi: 10.1016/j.invent.2021.100413] [Medline: 34401372]
61. Huberty J, Eckert R, Larkey L, Kurka J, Rodríguez De Jesús SA, Yoo W, et al. Smartphone-based meditation for myeloproliferative neoplasm patients: feasibility study to inform future trials. JMIR Form Res 2019;3(2):e12662 [FREE Full text] [doi: 10.2196/12662] [Medline: 31033443]
62. Mahlo L, Windsor TD. Feasibility, acceptability, and preliminary efficacy of an app-based mindfulness-meditation program among older adults. Gerontologist 2021;61(5):775-786. [doi: 10.1093/geront/gnaa093] [Medline: 32663286]
63. Fish MT, Saul AD. The gamification of meditation: a randomized-controlled study of a prescribed mobile mindfulness meditation application in reducing college students' depression. Simulation & Gaming 2019;50(4):419-435. [doi: 10.1177/1046878119851821]
64. Lipschitz JM, Connolly SL, Miller CJ, Hogan TP, Simon SR, Burdick KE. Patient interest in mental health mobile app interventions: demographic and symptom-level differences. J Affect Disord 2020;263:216-220. [doi: 10.1016/j.jad.2019.11.083] [Medline: 31818779]
65. Petros NG, Hadlaczky G, Carletto S, Martinez SG, Ostacoli L, Ottaviano M, et al. Sociodemographic characteristics associated with an eHealth system designed to reduce depressive symptoms among patients with breast or prostate cancer: prospective study. JMIR Form Res 2022;6(6):e33734 [FREE Full text] [doi: 10.2196/33734] [Medline: 35675116]
66. Toussaint A, Hüsing P, Gumz A, Wingenfeld K, Härter M, Schramm E, et al. Sensitivity to change and minimal clinically important difference of the 7-item Generalized Anxiety Disorder Questionnaire (GAD-7). J Affect Disord 2020;265:395-401. [doi: 10.1016/j.jad.2020.01.032] [Medline: 32090765]
67. Lynch CP, Cha EDK, Jenkins NW, Parrish JM, Mohan S, Jadczak CN, et al. The minimum clinically important difference for Patient Health Questionnaire-9 in minimally invasive transforaminal interbody fusion. Spine 2021;46(9):603-609. [doi: 10.1097/BRS.0000000000003853] [Medline: 33290370]
68. Kounali D, Button KS, Lewis G, Gilbody S, Kessler D, Araya R, et al. How much change is enough? Evidence from a longitudinal study on depression in UK primary care. Psychol Med 2020;52(10):1875-1882. [doi: 10.1017/s0033291720003700]
69. Doi S, Ito M, Takebayashi Y, Muramatsu K, Horikoshi M. Factorial validity and invariance of the 7-item Generalized Anxiety Disorder scale (GAD-7) among populations with and without self-reported psychiatric diagnostic status. Front Psychol 2018;9:1741 [FREE Full text] [doi: 10.3389/fpsyg.2018.01741] [Medline: 30283386]

Abbreviations

GAD-7: Generalized Anxiety Disorder-7
PHQ-9: Patient Health Questionnaire-9
SWLS: Satisfaction With Life Scale
WHO-5: World Health Organization-5 Well-Being Index

Edited by A Mavragani; submitted 26.10.24; peer-reviewed by F Shojaei, TF Ojo; comments to author 11.11.24; revised version received 02.12.24; accepted 18.12.24; published 23.01.25.

Please cite as:
Metaxa AM, Liverpool S, Eisenstadt M, Pollard J, Carlsson C
Improving Mental Health and Well-Being Through the Paradym App: Quantitative Study of Real-World Data
JMIR Form Res 2025;9:e68031
URL: https://formative.jmir.org/2025/1/e68031
doi:10.2196/68031
PMID:

XSL•FO
RenderX

©Athina Marina Metaxa, Shaun Liverpool, Mia Eisenstadt, John Pollard, Courtney Carlsson. Originally published in JMIR Formative Research (https://formative.jmir.org), 23.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Mindfulness Training for Burnout Reduction in Physicians: Clinician-Driven Approach

Lia Antico[1], PhD; Judson Brewer[1], PhD, MD

Brown University, Department of Behavioral and Social Sciences, Providence, RI, United States

Corresponding Author:
Lia Antico, PhD
Brown University
Department of Behavioral and Social Sciences
121 S Main St
Providence, RI, 02903
United States
Phone: 1 401 328 0763
Email: lia_antico@brown.edu

Abstract

Background: Physician burnout is widespread in health care systems, with harmful consequences on physicians, patients, and health care organizations. Mindfulness training (MT) has proven effective in reducing burnout; however, its time-consuming requirements often pose challenges for physicians who are already struggling with their busy schedules.

Objective: This study aimed to design a short and pragmatic digital MT program with input from clinicians specifically to address burnout and to test its efficacy in physicians.

Methods: Two separate nonrandomized pilot studies were conducted. In the first study, 27 physicians received the digital MT in a podcast format, while in the second study, 29 physicians and nurse practitioners accessed the same training through a free app-based platform. The main outcome measure was cynicism, one dimension of burnout. The secondary outcome measures were emotional exhaustion (the second dimension of burnout), anxiety, depression, intolerance of uncertainty, empathy (personal distress, perspective taking, and empathic concern subscales), self-compassion, and mindfulness (nonreactivity and nonjudgment subscales). In the second study, worry, sleep disturbances, and difficulties in emotion regulation were also measured. Changes in outcomes were assessed using self-report questionnaires administered before and after the treatment and 1 month later as follow-up.

Results: Both studies showed that MT decreased cynicism (posttreatment: 33% reduction; $P \leq .04$; $r \geq 0.41$ and follow-up: 33% reduction; $P \leq .04$; $r \geq 0.45$), while improvements in emotional exhaustion were observed solely in the first study (25% reduction, $P=.02$, $r=.50$ at posttreatment; 25% reduction, $P=.008$, $r=.62$ at follow-up). There were also significant reductions in anxiety ($P \leq .01$, $r \geq 0.49$ at posttreatment; $P \leq .01$, $r \geq 0.54$ at follow-up), intolerance of uncertainty ($P \leq .03$, $r \geq .57$ at posttreatment; $P<.001$, $r \geq 0.66$ at follow-up), and personal distress ($P=.03$, $r=0.43$ at posttreatment; $P=.03$, $r=0.46$ at follow-up), while increases in self-compassion ($P \leq .02$, $r \geq 0.50$ at posttreatment; $P \leq .006$, $r \geq 0.59$ at follow-up) and mindfulness (nonreactivity: $P \leq .001$, $r \geq 0.69$ at posttreatment; $P \leq .004$, $r \geq 0.58$ at follow-up; nonjudgment: $P \leq .009$, $r \geq 0.50$ at posttreatment; $P \leq .03$, $r \geq 0.60$ at follow-up). In addition, the second study reported significant decreases in worry ($P=.04$, $r=0.40$ at posttreatment; $P=.006$, $r=0.58$ at follow-up), sleep disturbances ($P=.04$, $r=0.42$ at posttreatment; $P=.01$, $r=0.53$ at follow-up), and difficulties in emotion regulation ($P=.005$, $r=0.54$ at posttreatment; $P<.001$, $r=0.70$ at follow-up). However, no changes were observed over time for depression or perspective taking and empathic concern. Finally, both studies revealed significant positive correlations between burnout and anxiety (cynicism: $r \geq 0.38$; $P \leq .04$; emotional exhaustion: $r \geq 0.58$; $P \leq .001$).

Conclusions: To our knowledge, this research is the first where clinicians were involved in designing an intervention targeting burnout. These findings suggest that this digital MT serves as a viable and effective tool for alleviating burnout and anxiety among physicians.

Trial Registration: ClinicalTrials.gov NCT06145425; https://clinicaltrials.gov/study/NCT06145425

(JMIR Form Res 2025;9:e63197) doi:10.2196/63197

XSL•FO
RenderX

KEYWORDS

burnout; anxiety; empathy fatigue; physician; mindfulness; digital therapeutics; app; smartphone; podcast; compassion; health care provider; training; physician burnout; cynicism; efficacy; treatment; meditation; chronic; workplace stress; digital health; mHealth; mobile phone

Introduction

Background

Burnout is an occupational phenomenon that is highly prevalent among physicians globally. While the rates have increased because of the COVID-19 pandemic, this was only the acceleration of an existing trend over the last decade [1-4]. In a 2021 survey of 2500 US physicians, 62.8% reported at least 1 symptom of burnout, compared with 38.2% in 2020, 43.9% in 2017, 54.4% in 2014, and 45.5% in 2011 [4].

Defined by the World Health Organization as "a syndrome resulting from chronic workplace stress that has not been managed successfully," burnout is characterized by 3 dimensions: emotional exhaustion, cynicism, and reduced personal accomplishment [5]. *Emotional exhaustion* includes feeling overwhelmed by work demands and experiencing depletion of emotional and physical resources. *Cynicism* represents the interpersonal aspect of burnout and refers to callous and detached attitudes toward one's job and people. *Reduced personal accomplishment* refers to poor professional self-esteem and efficacy [6].

The consequences of burnout on physicians, patients, and health care organizations are significant and far reaching. A large number of studies showed that physicians reporting high rates of burnout are also those who report high levels of anxiety [7], sleep disturbances [8-10], depression, and substance abuse [11,12]. Suicide rates among physicians are twice as high as those in the general public [13]. The emotional toll is so high that in a 2020 survey, 31% of physicians planned to reduce their work hours, and 24% of physicians intended to leave their job within 2 years [14]. An estimation of the direct costs of physician turnover and reduced clinical work is approximately US $4.6 billion each year in the United States [15]. Meta-analyses found that physician burnout is also associated with increased risk of patient safety incidents and malpractice claims, poorer quality due to low professionalism, and diminished and ineffective communication between physicians and patients [16,17]. Indirect costs from medical errors, reduced patient satisfaction, damage to the organization's reputation, and decreased patient loyalty raise the burnout bill to approximately US $7.6 million annually per employed physician [15].

Associated Factors of Physician Burnout

Physician burnout is complex and multifactorial. Burnout among physicians arises from both institutional and individual factors that create an imbalance between job demands and available resources. For instance, high workload and increased administrative tasks (eg, the electronic medical record systems) lead to overworking and deprioritizing self-care. A study involving 7288 US physicians found that for every additional hour worked per week beyond 51.8 hours, burnout symptoms rose by approximately 2% [18]. Similarly, research on 1490 US oncologists revealed that each additional hour spent working at home led to a 2% rise in burnout [2]. Another major contributor to burnout is the loss of flexibility, autonomy, and control that occurs when health care organizations and leadership fail to support individual work goals. This leaves physicians excluded from decisions about patient visit durations and treatment approaches [19].

It may not be a coincidence that the term "burnout" was first used in 1974 to describe the exhaustion observed in people in "healing professions" [20]. Indeed, physicians are constantly exposed to highly stressful and demanding situations that are intrinsic to the clinical practice. Many aspects of medicine are uncertain and ambiguous. For example, diagnoses are often unclear, the natural course of illness is unpredictable, responses of patients to treatments vary, and the clinical information is complex. A growing body of literature indicates that physicians' reactions to uncertainty and their ability to tolerate it can predict stress [21]. Cross-sectional studies reported negative correlations between intolerance of uncertainty and well-being of physicians [22,23]. In particular, physicians with a high intolerance for uncertainty are more likely to experience higher rates of burnout and anxiety, make more referrals, and report lower job satisfaction [24]. However, to date, no studies have directly tested the effects of interventions aimed at reducing physician burnout on intolerance of uncertainty.

While alleviating patients' distress is among the main reasons why many physicians enter the healing professions, caring for the sick is not without consequences. Repeated exposure to illness, death, or social inequities may cause empathic distress that leads to empathy fatigue (sometimes called inappropriately compassion fatigue [25]) and burnout.

Empathy is defined as the capacity to understand and share others' emotions, feelings, and mental states by imagining or taking another person's perspective [25]. Clinical empathy is an essential element of quality care, associated with improved patient satisfaction; increased adherence to treatment; better surgical recovery; shorter hospital stays [26-30]; and increased physician health, well-being, and professional satisfaction [31,32]. However, overempathizing can lead to being overwhelmed by the patient's distress, which might result in withdrawal and contribute to empathy fatigue. Conversely, maintaining a professional perspective with compassionate concern allows physicians to engage with the patient's distress without taking it personally, fostering stronger connections while reducing the risk of burnout.

Interventions for Physician Burnout

Combating burnout is a shared responsibility between the institution and the individual. Effective initiatives have been developed and implemented at both levels [33]. Organization-directed interventions target workload, foster communication between members of the health care system,

XSL•FO
RenderX

and cultivate a sense of team cohesion and job control [33,34]. Individual-focused approaches aim to decrease perceived stress, increase resilience to stressful work environments, and enhance work engagement [33-35]. In particular, this paper focuses on the tangible support that could be provided to physicians at the individual level.

Mindfulness Training

Several studies have reported that mindfulness training (MT) is one of the most studied individual-focused interventions for reducing stress and enhancing well-being among physicians [33,36-40]. Mindfulness refers to being fully aware and present in the current moment, on purpose and nonjudgmentally, of thoughts, feelings, and sensations [41]. By fostering self-awareness and attitudes of openness and "nonjudgment," MT facilitates acceptance of the challenging components of experiences and reduces emotional reactivity [42], both of which are crucial in health care. Most studies have observed that MT reduces rumination, worry, and emotional reactivity among health care providers [40,43]. For instance, nonreactivity to inner experiences has been inversely associated with perceived stress in health care providers [44].

In addition, a meta-analysis found that MT also increases self-compassion in health care providers, with moderate to high effect sizes [45]. Self-compassion involves being touched by one's own distress and being motivated to alleviate it with kindness and understanding [46,47]. It has been theorized that self-compassion is pertinent to physicians' well-being at work and care for patients [48], as it was inversely associated with burnout and empathy fatigue among health care professionals [49-52]. For example, it has been found that physicians who were more self-compassionate felt less exhausted due to work demands and experienced greater work engagement and satisfaction with their professional life [53].

These findings suggest that MT may be promising as an intervention for physician burnout. Different formats have been used in clinical settings [54]; some of them are intensive, including weekly groups sessions of 2.5 hours, home practices of 45 minutes for 8 weeks, and a 7-hour retreat [39]. Therefore, shorter and more practical formats are needed to avoid inadvertently contributing to overload. For example, in a study, 44% of health care professionals randomized to an MT dropped the program because of "lack of time" [55]. Even when health care providers are willing to attend MT, work conflicts can often arise and affect attendance and home practice adherence [35]. In a qualitative study, physicians confirmed the role of mindfulness interventions in the development of self-awareness and highlighted the importance of adapting the course timing and materials to meet their schedule, thereby improving participation and adherence [56].

Furthermore, engaging physicians in the design and implementation of interventions might increase their sense of control and engagement ("user-centered design") [57] and might make the interventions more feasible and tailored to their specific needs. However, to our knowledge, no studies have yet addressed this issue. This study is one of the few that involve clinicians in designing an intervention to reduce physician burnout. Successively, we conducted 2 independent single-arm studies and assessed the impact of digital MT training on physician burnout. In study 1, the intervention was delivered in a podcast format, while in study 2, it was delivered via a free app-based platform.

Study 1

Overview

To develop a short and tailored intervention, we used a user-centered design approach by involving clinicians and focusing on their needs in each phase of the development process. After iterative content development, we delivered the digital intervention in the format of an audio podcast, with the reasoning that content could be consumed during one's commute to and from work and, thus, not add more to the physician's workload. The primary goal of this single-arm study was to examine whether the intervention reduces cynicism (ie, one dimension of burnout) in physicians. The secondary goal was to assess its effects on other variables, such as emotional exhaustion (ie, another dimension of burnout), anxiety, depression, mindfulness (ie, the aspects of nonreactivity and nonjudging of inner experiences), intolerance of uncertainty, empathy (ie, the dimensions of empathic concern, perspective taking, and personal distress), and self-compassion. Finally, we explored the relationship between the 2 dimensions of burnout (ie, cynicism and emotional exhaustion) and the other secondary outcomes.

We evaluated the following hypotheses:

- Hypothesis 1: there will be a main effect of time on cynicism. Physicians will report lower scores for cynicism at postintervention assessments compared to baseline.
- Hypothesis 2: there will be a main effect of time on the other variables. Physicians will report lower scores for emotional exhaustion, anxiety, depression, intolerance of uncertainty, and personal distress and higher scores for nonreactivity, nonjudging of inner experiences, empathic concern, perspective taking, and self-compassion at postintervention assessments compared to baseline.
- Hypothesis 3: cynicism and emotional exhaustion will positively correlate with anxiety.

Methods

Ethical Considerations

The study was approved by the Brown University Institutional Review Board (protocol 2022003296) and was conducted in accordance with the Declaration of Helsinki for experiments involving human subjects. Eligible individuals were directed to the informed consent form on Qualtrics (Qualtrics International, Inc). After enrolling by selecting "yes" (a waiver of documentation of informed consent was obtained), they were redirected to complete the web-based survey. Participants were compensated with a US $25 Amazon gift card for the completion of each survey (US $75 for 3 surveys). All collected data were anonymized, with each participant being assigned a unique alphanumeric code. For statistical analyses, the final database contained only these alphanumeric codes to ensure privacy.

XSL•FO
RenderX

Participants

Physicians were recruited using university hospital mailing lists and flyer advertisements targeting physicians who were interested in testing whether a short training program was effective at reducing burnout. We used the following text in the recruitment message: "Would you like to help us test a short mindfulness-based training to reduce burnout? If you are a physician and are interested in helping us learn about burnout, please contact us."

Inclusion criteria were as follows: direct patient interaction; currently employed as a physician; fluency in English; endorsed willingness to listen to a mindfulness audio course for 15 minutes per day for 7 days; and endorsed willingness to complete 3 web-based surveys, before and after the completion of the training and 1 month later as follow-up. The only exclusion criteria were changing the dose of psychotropic medication in the past 6 weeks and prior exposure to this specific podcast or its content.

Intervention

The mindfulness-based training teaches physicians how to acknowledge and support patients' suffering without becoming overwhelmed by it. It also teaches them to identify and work with empathy fatigue habits loops as well as anxious and worry thoughts that contribute to burnout (eg, not being able to stop thinking about a problem, worrying about patients, and taking work home). Specifically, this program trains awareness and curiosity by targeting fundamental learning processes, such as reinforcement learning, to alter entrenched habits [58]. Previous studies have demonstrated that these changes can lead to significant, clinically relevant effects [58-61].

This experiential education is delivered via an audio podcast, which consists of 7 modules of brief didactic and experience-based MT, each approximately 15 minutes long, and a summary module at the end. Each module includes real-world vignettes and practical mindfulness exercises designed to help identify habit loops in clinical practice and work-related tasks and to provide strategies for breaking these unhelpful loops (Textbox 1). We also provided links to guided meditations, including noting, breathing, grounding practices, as well as curiosity and loving-kindness meditations. Although participants were encouraged to practice these guided meditations, they were optional and not mandatory for participation in this study.

Textbox 1. Overview of the audio course: burnout to resilience themes and content.

Module 1: goals, empathy fatigue, and unhelpful habits

- Set goals and introduce how important it is to be empathetic with patients in their recovery and its limits. Help physicians reflect on the different coping strategies used to protect themselves from patients' distress that become habits and contribute to exhaustion. It discusses specific strategies for stepping out of unhelpful habit loops.

Module 2: empathy versus compassion

- Addresses specifically the difference between empathy and compassion in patterns of activations in the body and mind, and it introduces compassion as a new way to feel patients' distress without being emotionally exhausted by it.

Module 3: intrusive thoughts

- Explains that intrusive thoughts are distracting and can cause great distress. This module also provides strategies for handling them and coming back to the present moment.

Module 4: self-judgment and kindness

- Describes how the mind can be caught up in the loop of self-judgments that contribute to exhaustion by getting in the way of doing one's own tasks and interacting with patients. This module also focuses on how kindness can be a detox of self-judgment.

Module 5: stress and anxiety

- Discusses why the brain does not like uncertainty and how this causes stress and anxiety. This module also provides strategies to change habit patterns and reduce stress and anxiety.

Module 6: emotional contagion and curiosity

- Introduces emotional contagion and its impact on the interactions with patients and colleagues. This module also explains how curiosity can be used as a vaccine against emotional contagion.

Module 7: bias in the diagnosis and treatment of pain

- Describes the complexity of pain, including factors and biases that influence the diagnosis and treatment of pain in patients, and provides strategies for recognizing and dealing with those biases.

The content of this intervention was developed using a user-centered design approach, an iterative design process in which designers focus on the users and their needs in each phase of the design process. On the basis of the extant literature and interviews of the physicians, we identified common individual empathy fatigue "habits" that contribute to burnout [62-64]. Scripts were written by a physician with >20 years of clinical and mindfulness practice and >10 years of experience in

XSL•FO
RenderX

developing MT programs [59-61]. We developed a "minimum viable product" consisting of script-based content that users read, tried out, and commented. We refined the audio course and tested its suitability for a wide range of physicians and other health care providers by collecting feedback from 40 clinicians in 2 rounds. In round 1, a total of 10 physicians from different subspecialties were emailed a script, instructions, and a link to a feedback form. The feedback form included open-ended questions about what participants learned, what was helpful in their clinical practice and personal life, and any confusion or difficulties with understanding and completing home practices and slider questions (ranging from 0 to 100) about how well the module met their needs and those of their colleagues, the usefulness of the real-life stories shared, and the clarity of the home practice instructions. Once each module was completed (as indicated by a completed feedback form), the next module was emailed to the pilot tester. Feedback was collated, and common themes were extracted for training refinement. Vignettes from this round of pilot testers were also incorporated for story-based learning [65]. In round 2, a total of 30 physicians and other clinicians were recruited to repeat this process to confirm content fit and gather additional real-world vignettes for potential incorporation.

Intervention Orientation and Engagement

Participants were instructed to complete the modules at their own pace, with a maximum of 1 module per day. The project director reached out with check-in messages every 15 days from the start of the training to address technical difficulties and encourage continued engagement. During these check-ins, participants were asked about their experience with the podcast since the last contact. If participants expressed difficulties, efforts were made to resolve the problem.

Measures

Primary and secondary outcomes were assessed at 3 time points: baseline (before training), postintervention (after training completion), and follow-up (1 month after training completion).

Primary Outcome

Cynicism, a key dimension of burnout, was measured using a single item from the Maslach Burnout Inventory. This approach is supported by research from West et al [66,67], which demonstrated the validity of using a single item in relation to the full 22-item questionnaire. Participants rated how often they felt "more callous toward people since starting this job" on a 7-point scale (0="never" and 6="every day"). Higher scores indicate greater cynicism [68].

Secondary Outcomes

Emotional exhaustion was measured using a single Maslach Burnout Inventory item. Participants rated how often they felt "burned out from their work" on a 7-point scale, with higher scores indicating greater exhaustion [68].

Anxiety was measured using the *Generalized Anxiety Disorder-7* on a 4-point scale (0="not at all" and 3="nearly every day"). Scores range from 0 to 21, with higher scores indicating more severe anxiety. Total scores of 5 to 9, 10 to 14, and ≥15 correspond to cutoff points of mild, moderate, and severe anxiety, respectively [69].

Depression was measured using the *Patient Health Questionnaire-2* assessing depressive symptoms over the past 2 weeks. Participants rated each item on a 4-point scale (0="not at all" and 3="nearly every day"), with higher scores indicating more severe symptoms [70].

Intolerance *of uncertainty* was assessed by the *Intolerance of Uncertainty scale-12* measuring intolerance of uncertainty with 12 items rated on a 5-point scale (1="not at all" to 5="entirely"). Total scores range from 12 to 60, with higher scores indicating greater intolerance [71].

Self-compassion was measured by the *Self-Compassion Scale-Short Form* assessing self-compassion using 12 items rated on a 5-point scale (1="almost never" to 5="almost always"). Total scores range from 12 to 60, with higher scores indicating greater self-compassion [72].

Mindfulness was assessed using the *Five Facet Mindfulness Questionnaire* measuring nonreactivity and nonjudgment, with subscales rated on a 5-point scale (1="never true" to 5="always true"). Nonreactivity total scores range from 7 to 35, with higher scores indicating a greater ability to not react to or get caught up in thoughts or emotions. Nonjudgment total scores range from 8 to 40, with higher scores indicating a greater ability to not judge or criticize thoughts or emotions [73].

Empathy was assessed using the *Interpersonal Reactivity Index* that measured empathy using 4 subscales (empathic concern, perspective taking, personal distress, and fantasy). Items were rated on a 5-point scale (0="does not describe me" to 4="describes me well"). Each subscale consists of 7 items, and total scores range from 0 to 28, with higher scores indicating greater empathy across subscales. The fantasy subscale was excluded as not relevant [74].

Statistical Analysis

Statistical analyses were conducted in R (version 4.2.2; R Foundation for Statistical Computing). Friedman ANOVA, a nonparametric test, was used to analyze the overall change in cynicism scores (primary outcome) and the other scores (secondary outcomes) at the 3 time points due to the data having a nonnormal distribution. Post hoc analyses between the individual time points were analyzed using Wilcoxon signed rank tests and were corrected for multiple comparisons using the Benjamini-Hochberg procedure. The relationship between the dimensions of burnout and the other variables was evaluated using the Spearman rank correlation coefficient. Effect sizes were reported as Kendall $W = \chi^2/N(K-1)$ for Friedman tests and as Pearson $r = z/\sqrt{N}$ for pairwise Wilcoxon signed rank tests. Kendall W and Pearson *r* use Cohen criteria where 0.1 to <0.3 is a small effect, 0.3 to <0.5 is a medium effect, and ≥0.5 is a large effect [75].

Results

Participants

A total of 40 physicians replied to our recruitment message, and 32 (80%) participants met eligibility criteria and consented

XSL•FO
RenderX

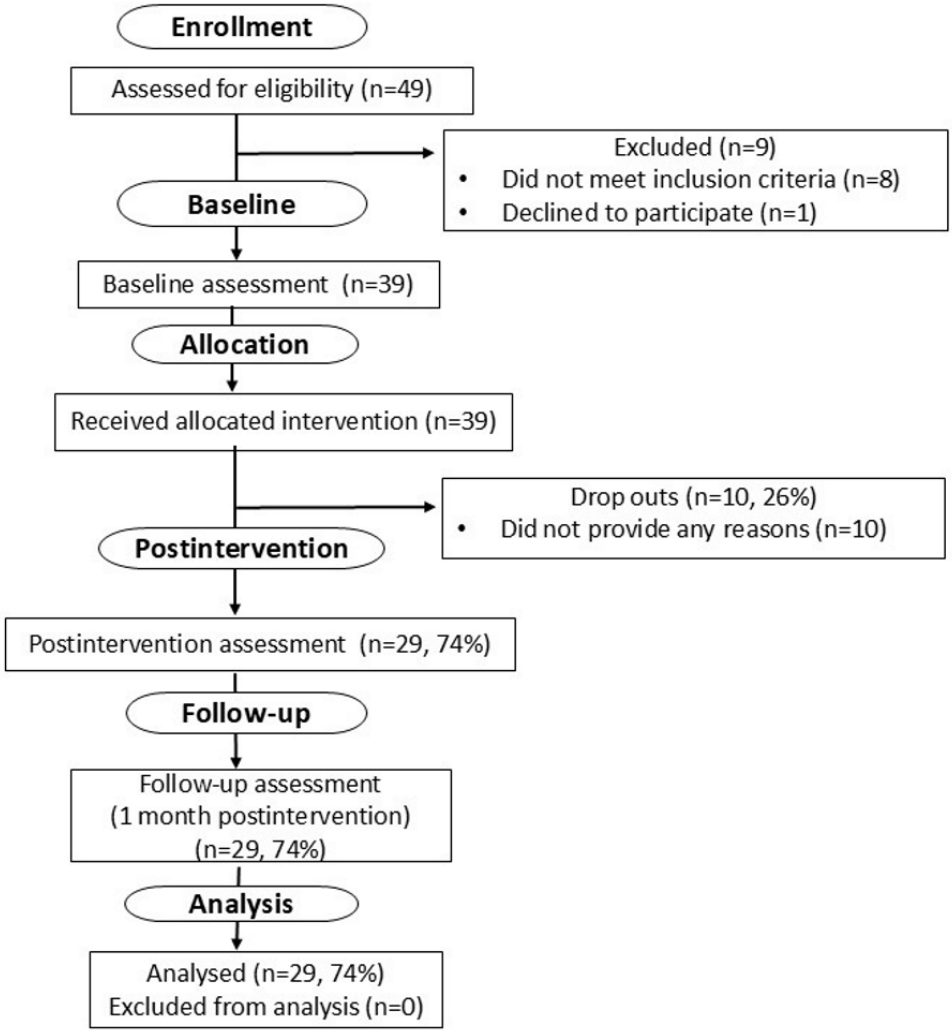
to participate between August 2022 and February 2023. Out of this group who received the allocated intervention, of the 32 physicians, 27 (84%) completed assessments at all 3 time points and were included in the final analysis (Figure 1). The participant population comprised 20 women and 7 men who worked in health care for an average of 15 (SD 9.76) years. The average age was 50.6 (SD 9.74) years (Table 1).

Table 1. Demographic characteristics of study 1.

	Participants (N=27)
Age (y), mean (SD)	50.6 (9.74)
Gender, n (%)	
Man	7 (26)
Woman	20 (74)
Race and ethnicity, n (%)	
Asian	5 (19)
Black or African American	1 (4)
Hispanic, Latinx, or Spanish	0 (0)
White	21 (77)
Work status, n (%)	
Employed full time	22 (82)
Employed part time	5 (8)
Years working in health care, mean (SD)	15 (9.76)
Medical specialty , n (%)	
Emergency medicine	2 (7)
Pediatric emergency medicine	3 (12)
Pediatric hematology oncology	2 (7)
Pulmonary and critical care medicine	2 (7)
Family medicine	6 (22)
Internal medicine	5 (19)
Neurology	1 (4)
Neurocritical care	1 (4)
Psychiatry	2 (7)
Infectious disease	1 (4)
Obstetrics-gynecology	2 (7)

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of study procedures of study 1.



Primary and Secondary Outcomes

The average time taken by participants to complete the program was 30 (SD 16) days.

Table 2 shows the outcome scores at each assessment point compared with baseline.

Friedman ANOVA tests indicated significant changes over time for cynicism (χ^2_2=6.96; *P*=.03; W=0.13), emotional exhaustion (χ^2_2=12.6; *P*=.002; W=0.23), anxiety (χ^2_2=10.7; *P*=.005; W=0.20), intolerance of uncertainty (χ^2_2=12.1; *P*=.002; W=0.22), self-compassion (χ^2_2=9.88; *P*=.007; W=0.18), nonreactivity (χ^2_2=16.7; *P*<.001; W=0.31), and nonjudgment of inner experiences (χ^2_2=19.5; *P*<.001; W=0.36) and personal distress (a subdimension of empathy; χ^2_2=6.74; *P*=.03; W=0.13). No significant change was observed across the 3 time points for the other subdimensions of empathy, such as perspective taking (χ^2_2=5.75; *P*=.06; W=0.106) and empathic concern (χ^2_2=90.3; *P*=.86; W=0.006), and for depression (χ^2_2=1.72; *P*=.42; W=0.032).

Post hoc Wilcoxon signed rank tests revealed 33% reduction of cynicism from baseline to postintervention (W_{26}=134; *P*=.04; *r*=0.41) and 1 month later (W_{26}=140; *P*=.04; *r*=0.47), 25% reduction of emotional exhaustion at both points (postintervention: W_{26}=116; *P*=.02; *r*=0.50 and follow-up: W_{26}=154; *P*=.008; *r*=0.62), and 43% reduction of anxiety at postintervention (W_{26}=244; *P*=.004; *r*=0.62) and 29% one month later (W_{26}=228; *P*=.01; *r*=0.54; Figure 2).

Wilcoxon signed rank tests revealed 26% reduction of intolerance of uncertainty at postintervention (W_{26}=314; *P*<.001; *r*=0.67) and 1 month after the end of the intervention (W_{26}=242; *P*<.001; *r*=0.70), 15% increase of self-compassion at both points (postintervention: W_{26}=73.5; *P*=.02; *r*=0.50 and follow-up: W_{26}=36; *P*=.006; *r*=0.59), 16% increase of nonreactivity (postintervention: W_{26}=22.5; *P*=.001; *r*=0.69 and follow-up: W_{26}=50; *P*=.004; *r*=0.58), and 15% increase of nonjudgment of inner experiences at postintervention (W_{26}=5; *P*<.001; *r*=0.81) and 12% increase 1 month later (W_{26}=47; *P*=.003; *r*=0.60; Figure 3).

XSL•FO
RenderX

Table 2. Descriptive statistics for primary and secondary outcomes (N=27 physicians).

	Values, mean (SD)	Values, median (IQR)	*P* value	Effect sizes (*r*)
Cynicism				
Baseline	2.93 (1.77)	3 (2)	—[a]	—
Postintervention	2.26 (1.26)	2 (1.5)	—	—
1-mo postintervention	2.07 (1.38)	2 (2)	—	—
Δ[b] at postintervention (%)	−0.67 (−23)	−1 (−33)	.04	0.41
Δ at 1-mo postintervention (%)	−0.86 (−29)	−1 (−33)	.04	0.47
Emotional exhaustion				
Baseline	3.89 (1.31)	4 (2)	—	—
Postintervention	3.22 (1.31)	3 (2)	—	—
1-mo postintervention	2.93 (1.41)	3 (2)	—	—
Δ at postintervention (%)	−0.67 (−17)	−1 (−25)	.02	0.50
Δ at 1-mo postintervention (%)	–0.96 (−25)	−1 (−25)	.08	0.62
GAD-7[c]				
Baseline	7.63 (5.25)	7 (8)	—	—
Postintervention	4.59 (4.45)	4 (4.5)	—	—
1-mo postintervention	4.67 (3.85)	5 (4)	—	—
Δ at postintervention (%)	−3.04 (−40)	−3 (−43)	.004	0.62
Δ at 1-mo postintervention (%)	−2.96 (−39)	−2 (−29)	.01	0.54
PHQ-2 [d]				
Baseline	0.56 (0.70)	0 (1)	—	—
Postintervention	0.63 (1.12)	0 (1)	—	—
1-mo postintervention	0.52 (0.80)	0 (1)	—	—
Δ at postintervention (%)	−0.07 (−12)	0 (0)	.99	0.03
Δ at 1-mo postintervention (%)	−0.04 (−7)	0 (0)	.97	0.11
Intolerance of uncertainty				
Baseline	32.6 (10)	34 (12.5)	—	—
Postintervention	27.4 (9.31)	25 (10)	—	—
1-mo postintervention	27.4 (8.97)	25 (13)	—	—
Δ at postintervention (%)	−5.2 (16)	−9 (−26)	<.001	0.67
Δ at 1-mo postintervention (%)	−5.2 (16)	−9 (−26)	<.001	0.70
Self-compassion				
Baseline	34.3 (7)	33 (11)	—	—
Postintervention	38.1 (6.97)	38 (7)	—	—
1-mo postintervention	39.2 (7.87)	38 (11.5)	—	—
Δ at postintervention (%)	3.8 (11)	5 (15)	.02	0.50
Δ at 1-mo postintervention (%)	4.9 (14)	5 (15)	.006	0.59
Nonreactivity				
Baseline	19.3 (4.65)	19 (7)	—	—
Postintervention	22.6 (3.90)	22 (4)	—	—
1-mo postintervention	21.9 (4.26)	22 (4)	—	—
Δ at postintervention (%)	3.3 (17)	3 (16)	.001	0.69

XSL•FO
RenderX

	Values, mean (SD)	Values, median (IQR)	*P* value	Effect sizes (*r*)
Δ at 1-mo postintervention (%)	2.6 (13)	3 (16)	.004	0.58
Nonjudgment				
Baseline	24 (6.66)	26 (10.5)	—	—
Postintervention	29.7 (5.46)	30 (5)	—	—
1-mo postintervention	28.3 (5.51)	29 (6)	—	—
Δ at postintervention (%)	5.7 (24)	4 (15)	<.001	0.81
Δ at 1-mo postintervention (%)	4.3 (18)	3 (12)	.003	0.60
Personal distress				
Baseline	8.52 (4.43)	8 (5.5)	—	—
Postintervention	6.96 (4.57)	7 (6)	—	—
1-mo postintervention	6.74 (4.64)	6 (4.5)	—	—
Δ at postintervention (%)	−1.56 (−18)	−1 (−12)	.03	0.43
Δ at 1-mo postintervention (%)	−1.78 (−21)	−2 (−25)	.03	0.46
Empathic concern				
Baseline	21.5 (5.02)	22 (6)	—	—
Postintervention	22.1 (3.47)	22 (5.5)	—	—
1-mo postintervention	21.6 (3.96)	22 (6)	—	—
Δ at postintervention (%)	0.6 (2.8)	0 (0)	.81	0.07
Δ at 1-mo postintervention (%)	0.1 (0.5)	0 (0)	.81	0.04
Perspective taking				
Baseline	20.3 (4.57)	21 (4.5)	—	—
Postintervention	21.2 (4.31)	21 (5.5)	—	—
1-mo postintervention	21.7 (4.31)	22 (5)	—	—
Δ at postintervention (%)	0.9 (4)	0 (0)	.18	0.26
Δ at 1-mo postintervention (%)	1.4 (7)	1 (5)	.07	0.43

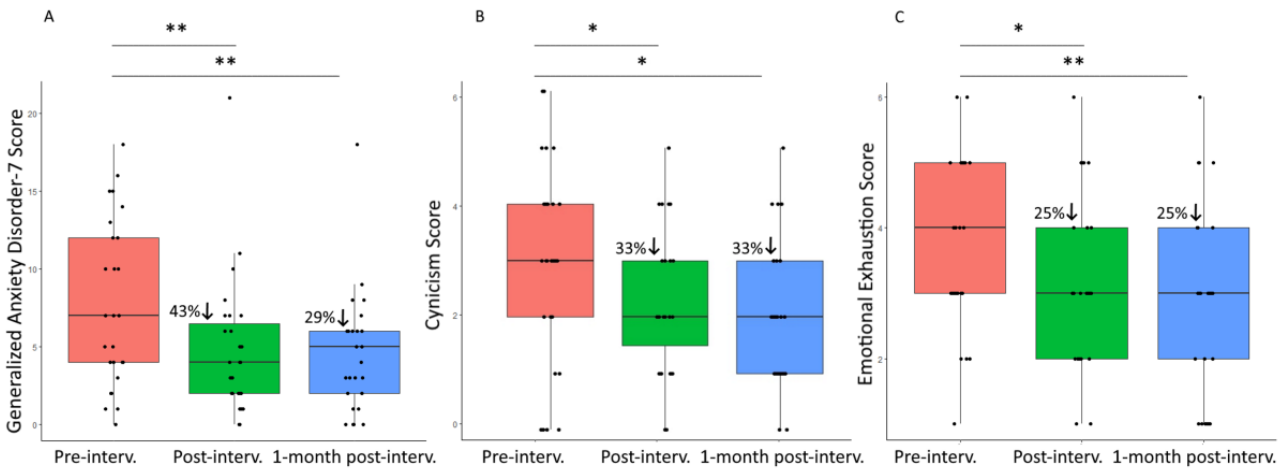
[a]Not applicable.

[b]Δ: change between baseline and stated time point.

[c]GAD-7: Generalized Anxiety Disorder-7.

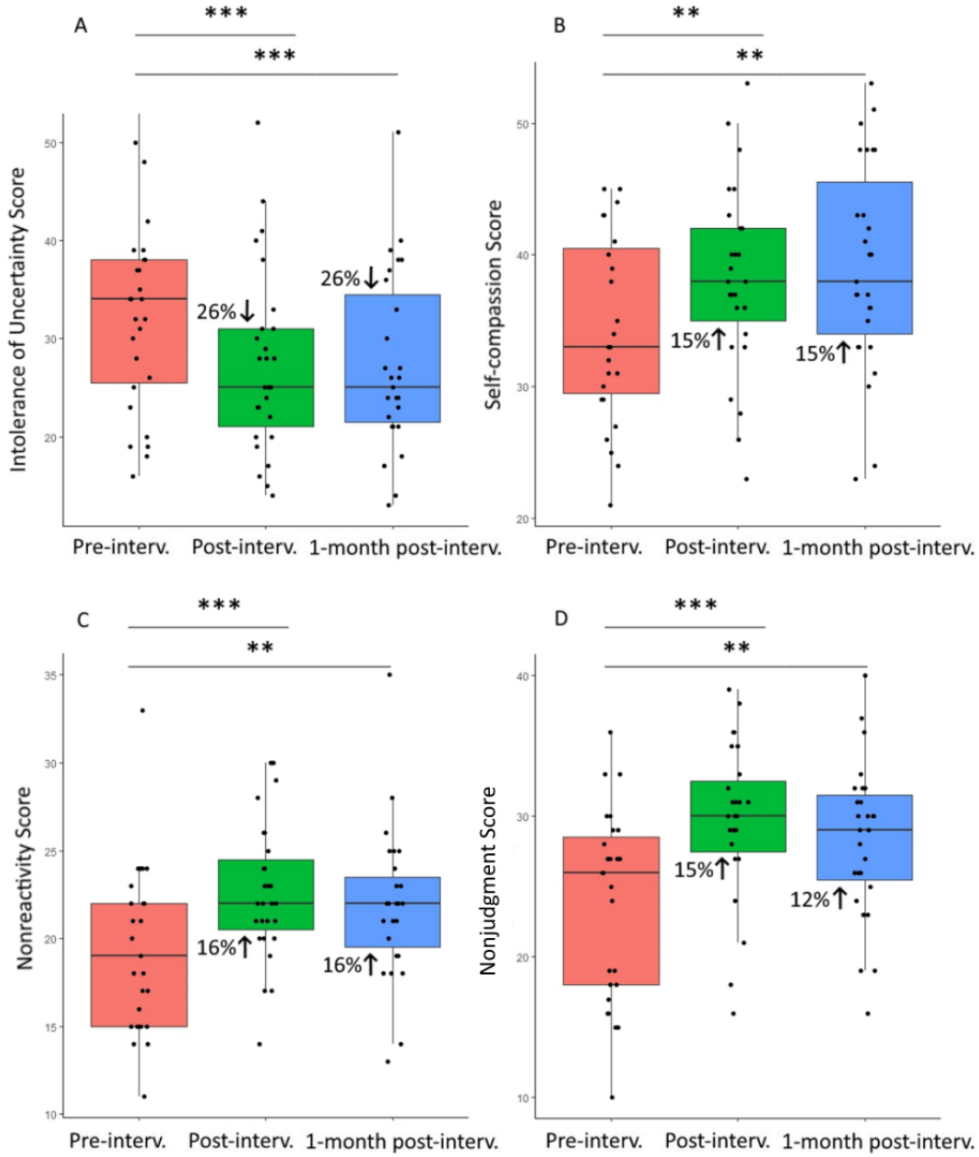
[d]PHQ-2: Patient Health Questionnaire-2.

Figure 2. Box and whisker plots at baseline, at postintervention, and 1 month after intervention completion for (A) Generalized Anxiety Disorder-7 scores, (B) cynicism scores, and (C) emotional exhaustion scores from the Maslach Burnout Inventory. **P*=.05, ***P*=.01.



XSL•FO
RenderX

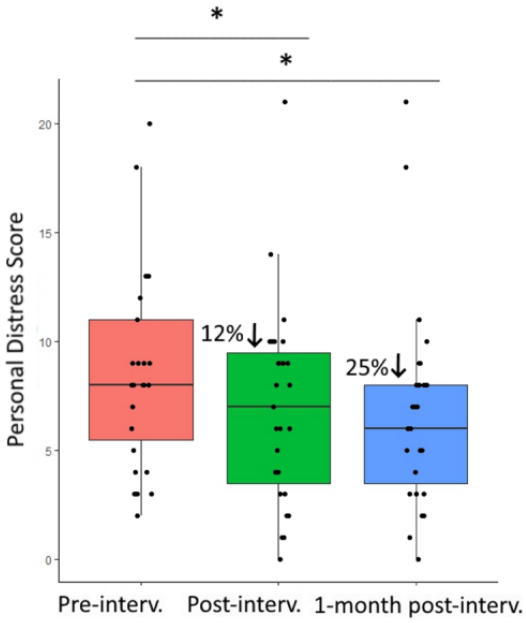
Figure 3. Box and whisker plots at baseline, at postintervention, and 1 month after intervention completion for (A) intolerance of uncertainty scores, (B) self-compassion scores, (C) nonreactivity, and (D) nonjudgment of inner experiences scores. *P=.05, **P=.01, ***P=.001.



Finally, Wilcoxon signed rank tests showed a 12% reduction in personal distress scores from baseline to postintervention (W_{26}=217; P=.03; r=0.43) and a 25% reduction at 1 month after the end of the intervention (W_{26}=200; P=.03; r=0.46; Figure 4).

XSL•FO
RenderX

Figure 4. Box and whisker plots at baseline, at post-intervention, and 1 month after intervention completion, for personal distress scores. Significance level is denoted by asterisks: *P=.05.



Correlations Between Burnout and Other Outcomes

Table 3 shows the correlations between the 2 burnout dimensions of cynicism and emotional exhaustion and the other outcomes at each assessment point. There were significant correlations between burnout and anxiety at baseline and postintervention (cynicism: r≥0.56; P≤.03 and emotional exhaustion: r≥0.53; P≤.004) but not 1 month after the end of the intervention. Burnout scores were also correlated with nonreactivity scores at baseline and postintervention (cynicism: r≥0.39; P≤.04 and emotional exhaustion: r≥−0.40; P≤.04) but not at follow-up. Emotional exhaustion score was also correlated with self-compassion score only at baseline (r=−0.39; P≤.043). We did not observe any other correlations (r≤|0.37|; P≥|.06|).

XSL•FO
RenderX

Table 3. Correlations between burnout scores and other outcomes at baseline, postintervention, and 1-month postintervention.

Variables	Baseline	Postintervention	1-mo postintervention
GAD-7[a] × cynicism			
Value	0.56	0.41	0.24
P value	.002	.03	.22
GAD-7 × emotional exhaustion			
Value	0.60	0.53	0.32
P value	<.001	.004	.99
PHQ-2[b] × cynicism			
Value	0.25	0.12	0.11
P value	.21	.56	.60
PHQ-2 × emotional exhaustion			
Value	0.17	0.08	0.21
P value	.39	.70	.28
IU[c] × cynicism			
Value	0.37	0.02	0.16
P value	.06	.91	.44
IU × emotional exhaustion			
Value	0.15	0.36	0.12
P value	.44	.07	.55
Self-compassion × cynicism			
Value	−0.22	−0.16	−0.05
P value	.27	.44	.80
Self-compassion × emotional exhaustion			
Value	−0.39	−0.17	−0.03
P value	.04	.41	.87
Nonreactivity × cynicism			
Value	−0.39	−0.69	−0.35
P value	.04	<.001	.07
Nonreactivity × emotional exhaustion			
Value	−0.40	−0.49	−0.35
P value	.04	.01	.07
Nonjudgment × cynicism			
Value	−0.34	−0.13	−0.17
P value	.09	.52	.41
Nonjudgment × emotional exhaustion			
Value	−0.37	−0.02	−0.06
P value	.06	.90	.76
Personal distress × cynicism			
Value	0.10	0.08	0.34
P value	.63	.67	.10
Personal distress × emotional exhaustion			
Value	0.06	0.15	0.23

XSL•FO
RenderX

Variables	Baseline	Postintervention	1-mo postintervention
P value	.77	.45	.24
Empathic concern × cynicism			
Value	−0.07	−0.01	−0.24
P value	.75	.95	.23
Empathic concern × emotional exhaustion			
Value	0.27	0.24	0.14
P value	.17	.24	.48
Perspective taking × cynicism			
Value	−0.17	−0.09	−0.10
P value	.38	.67	.62
Perspective taking × emotional exhaustion			
Value	−0.03	−0.25	−0.15
P value	.85	.21	.46

[a]GAD-7: Generalized Anxiety Disorder-7.

[b]PHQ-2: Patient Health Questionnaire-2.

[c]IU: intolerance of uncertainty.

Study 2

Overview

The results of study 1 provided evidence that physicians reported lower cynicism and emotional exhaustion at postintervention assessments compared to baseline. This was also the case for reduced anxiety, intolerance of uncertainty, and personal distress and increased self-compassion and mindfulness after the intervention. No difference was observed for depression and 2 subdimensions of empathy (ie, empathic concern and perspective taking) across time. This lack of change may be attributed to participants reporting low levels of depression and high levels of empathy at baseline.

To confirm the validity of these initial findings from study 1, we conducted a preregistered replication study. We used that same intervention but delivered it via a free smartphone-based platform called Unwinding by Sharecare (Behavioral Health team at Sharecare), which allows a progression of 7 daily modules of brief didactic; 1 summary at the end; and access to further features, such as experience-based MT, app-triggered check-ins, guided meditations (5-15 minutes), and clinician life stories.

Similar to study 1, participants completed web-based surveys at 3 time points: baseline, postintervention, and 1 month later as follow-up. To be more inclusive and expand to a broader range of clinicians, we recruited physicians and nurse practitioners, who provide direct care and medication prescriptions to their patients.

The same primary and secondary outcomes were included as in study 1, with the exception of the Interpersonal Reactivity Index, which was removed due to its lack of appropriateness for physicians, as indicated by the results of study 1.

In addition, based on the strong reduction of anxiety observed in study 1, we aimed to explore the effects of the intervention on further variables that have been suggested to be relevant for the onset and maintenance of anxiety, such as worry [76], sleep disturbances [77], and difficulties in emotion regulation [78].

We evaluated the following hypotheses:

- Hypothesis 1: there will be a main effect of time on cynicism. Physicians and nurse practitioners will report lower scores for cynicism at postintervention assessments compared to baseline.
- Hypothesis 2: there will be a main effect of time on the other variables. Physicians and nurse practitioners will report lower scores for emotional exhaustion, anxiety, depression, intolerance of uncertainty, worry, sleep disturbances, and difficulties in emotion regulation and higher scores for nonreactivity, nonjudging of inner experiences, and self-compassion at postintervention assessments compared to baseline.
- Hypothesis 3: cynicism and emotional exhaustion will positively correlate with anxiety.

Methods

Ethical Considerations

Similar to study 1, this study was approved by the Brown University Institutional Review Board (protocol 2022003296) and followed the Declaration of Helsinki. The same consent process was used: participants were directed to the informed consent form on Qualtrics, and after selecting “yes” (with a waiver of documentation), they completed the web-based survey. Data were anonymized with alphanumeric codes, and the final database used only these codes. Likewise, participants were compensated with a US $25 Amazon gift card for the completion of each survey (US $75 for 3 surveys). The study was preregistered at ClinicalTrials.gov (NCT06145425), except for

XSL•FO
RenderX

the exploratory outcome of difficulties in emotion regulation that we added later.

Participants

The same study design, recruitment process, and inclusion and exclusion criteria were used as in study 1, except for the additional recruitment of nurse practitioners. To be more inclusive, we recruited physicians and nurse practitioners, who are equally skilled and knowledgeable in providing direct care and medication prescriptions to their patients.

Intervention

The same intervention was delivered via the smartphone app Unwinding by Sharecare, which included 7 daily modules of MT, a final summary, app-triggered check-ins, guided meditations (5-15 minutes), and clinician life stories. Participants were encouraged to explore other app features; however, this was optional. Similar to study 1, participants completed 1 module per day at their own pace, with check-in messages from the project director every 15 days to address technical issues and encourage engagement.

Engagement

Data related to program engagement, as measured by the number of modules completed, were obtained directly from the Unwinding by Sharecare app.

Outcome Measures

The same approach was used as described in study 1: participants completed web-based surveys at 3 time points: baseline, postintervention, and 1 month later as follow-up. The same outcomes were included as described in study 1, with the following exceptions: the Interpersonal Reactivity Index was removed because the results of study 1 showed that this measure might not be appropriate for physicians. The Patient-Reported Outcomes Measurement Information System (PROMIS), the Penn State Worry Questionnaire, and the Difficulties in Emotion Regulation Scale were added as secondary outcomes to explore the effects of the intervention on sleep disturbances, worry, and difficulties in emotion regulation, respectively.

PROMIS was used to measure worry-related sleep disturbances. Five items from PROMIS were preselected based on their direct assessment of worry interfering with sleep [79]. Items were rated on a 5-point Likert scale (1="not at all" and 5="very much"). Scores range from 5 to 25, with higher scores indicating more sleep disturbances and greater sleep-related impairment [79,80].

Penn State Worry Questionnaire is a 16-item questionnaire to assess worry on a 5-point Likert scale (1="not at all typical of me" and 5="very typical of me"). Scores range from 16 to 80, with higher total scores indicating increased levels of worry [81].

Difficulties in Emotion Regulation Scale is a 36-item questionnaire used to assess emotion regulation challenges on a 5-point Likert scale (1="almost never" and 5="almost always"). Scores range from 36 to 180, with higher total scores indicating increased greater emotion regulation difficulty [82].

Statistical Analysis

The same data analysis approach described in study 1 was used.

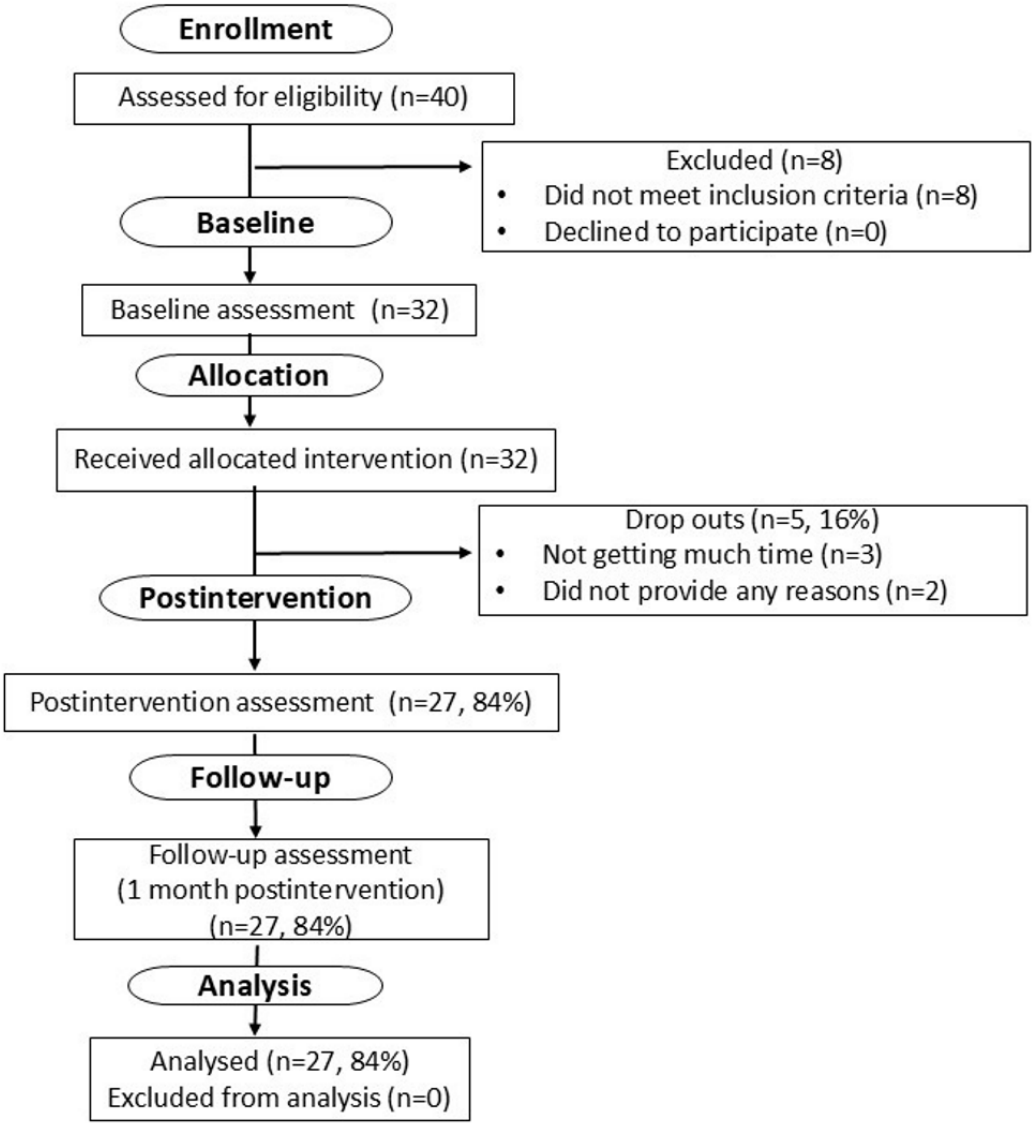
Results

Participants

A total of 40 physicians and 10 nurse practitioners replied to our recruitment message, and 39 (78%) participants (29 physicians and 10 nurse practitioners) met the eligibility criteria and consented to participate between June 2023 and May 2024. Out of this group who received the allocated intervention, of the N = 39 participants, 29 (74%, n=20, 69% physicians and n=9, 90% nurse practitioners) completed the 3-time point assessments and were included in the final analysis (Figure 5). The participant population comprised 23 women and 6 men who worked in health care for an average of 16.4 (SD 10.1) years. The average age was 49.6 (SD 9.8) years (Table 4).

XSL•FO
RenderX

Figure 5. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of study procedures of study 2.



XSL•FO
RenderX

Table 4. Demographic characteristics of study 2.

	Physicians+nurse practitioners (n=29)	Physicians (n=20)	Nurse practitioners (n=9)
Age (y), mean (SD)	49.6 (9.8)	49.8 (10.9)	49.1 (6.9)
Gender, n (%)			
Men	6 (21)	5 (25)	1 (11)
Women	23 (79)	15 (75)	8 (89)
Race and ethnicity, n (%)			
Asian	3 (10)	3 (15)	0 (0)
Black or African American	2 (7)	1 (5)	1 (11)
Hispanic, Latinx, or Spanish	0 (0)	0 (0)	0 (0)
White	24 (83)	16 (80)	8 (89)
Work status, n (%)			
Employed full time	27 (93)	18 (90)	9 (100)
Employed part time	2 (7)	2 (10)	0 (0)
Working duration in health care (y), mean (SD)	16.41 (10.14)	14.85 (8.83)	19.89 (11.86)
Medical specialty, n (%)			
Emergency medicine	3 (10.2)	2 (10)	1 (11)
Pediatrics	6 (21)	2 (10)	4 (45)
Oncology	1 (3.4)	0 (0)	1 (11)
Palliative medicine	2 (7)	2 (10)	0 (0)
Family medicine	2 (7)	1 (5)	1 (11)
Internal medicine	4 (14)	4 (20)	0 (0)
Nephrology	1 (3.4)	1 (5)	0 (0)
Hospitality	1 (3.4)	1 (5)	0 (0)
Psychiatry	1 (3.4)	0 (0)	1 (11)
Infectious disease	1 (3.4)	1 (5)	0 (0)
Obstetrics-gynecology	3 (10.2)	3 (15)	0 (0)
Endocrinology	1 (3.4)	1 (5)	0 (0)
Clinical immunology	1 (3.4)	1 (5)	0 (0)
Physical medicine	1 (3.4)	1 (5)	0 (0)
Orthopedics	1 (3.4)	0 (0)	1 (11)

Engagement

At posttreatment, 100% (29/29) of the participants completed the 7-module program; only 3 (10%) participants did not listen to the summary at the end. The average time taken by participants to complete the program was 23.8 (SD 18.7) days.

Primary and Secondary Outcomes

Multimedia Appendix 1 shows the outcome scores at each assessment point compared with baseline. Friedman ANOVA tests indicated significant changes over time for cynicism (χ^2_2=13.1; *P*=.001; W=0.23), emotional exhaustion (χ^2_2=2.39; *P*=.30; W=0.04), anxiety (χ^2_2=14.6; *P*<.001; W=0.25), intolerance of uncertainty (χ^2_2=15.6; *P*<.001; W=0.27), self-compassion (χ^2_2=22.2; *P*<.001; W=0.38), nonreactivity (χ^2_2=24.3; *P*<.001; W=0.42), and nonjudgment of inner experiences (χ^2_2=10.3; *P*=.006; W=0.18). Friedman ANOVA tests revealed significant changes over time for worry (χ^2_2=7.53; *P*=.02; W=0.13), sleep disturbances (χ^2_2=7.07; *P*=.03; W=0.122), and difficulties in emotion regulation (χ^2_2=15.4; *P*<.001; W=0.27). No significant change across the 3 time points was observed for depression (χ^2_2=1.77; *P*=.41; W=0.03).

Post hoc Wilcoxon signed rank tests revealed 33% reduction of cynicism at both postintervention (W_{28}=184; *P*=.008; *r*=0.58) and 1 month after the end of the intervention (W_{28}=196; *P*=.04; *r*=0.45) and 29% reduction of anxiety scores from baseline to postintervention (W_{28}=317; *P*=.01; *r*=0.49) and 57% reduction 1 month later (W_{28}=340; *P*=.006; *r*=0.59; Figure 6). We

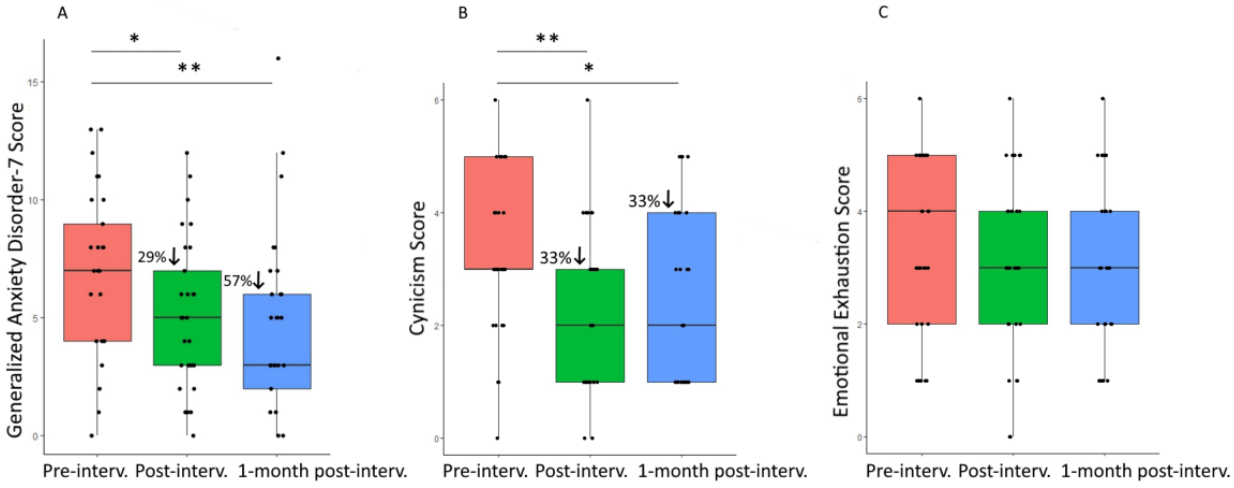
XSL•FO
RenderX

observed stronger results when we analyzed only the data from physicians: a significant change across the 3 time points for cynicism (χ^2_2=16.3; *P*<.001; W=0.41) and a 50% reduction in scores from baseline to postintervention (W_{19}=105; *P*=.003; *r*=0.81) and 33% reduction 1 month later (W_{19}=128; *P*=.02; *r*=0.58). Likewise, we observe similar results for anxiety over time (χ^2_2=10.5; *P*=.005; W=0.26) and 33% reduction from baseline to postintervention (W_{19}=150; *P*=.04; *r*=0.51) and 60% reduction 1 month later (W_{19}=165; *P*=.02; *r*=0.63).

Wilcoxon signed rank tests found 12% reduction of intolerance of uncertainty scores from baseline to postintervention (W_{26}=338; *P*=.003; *r*=0.57) and 8% reduction 1 month after the end of the intervention (W_{26}=310; *P*<.001; *r*=0.66) and 21% increase in self-compassion scores from baseline to postintervention (W_{28}=32.5; *P*<.001; *r*=0.69) and 27% increase 1 later (W_{28}=2; *P*<.001; *r*=0.84). We observed similar results with data collected only from physicians: significant changes across the 3 time points for intolerance of uncertainty (χ^2_2=9.24; *P*=.01; W=0.23) and for self-compassion (χ^2_2=15.5; *P*<.001; W=0.39). Intolerance of uncertainty scores reduced by 22% at postintervention (W_{19}=162; *P*=.01; *r*=0.59) and by 24% one month later (W_{19}=158; *P*=.003; *r*=0.72) and self-compassion scores increased by 31% at postintervention (W_{19}=9.5; *P*=.001; *r*=0.74) and 38% one month later (W_{19}=1.5; *P*<.001; *r*=0.82).

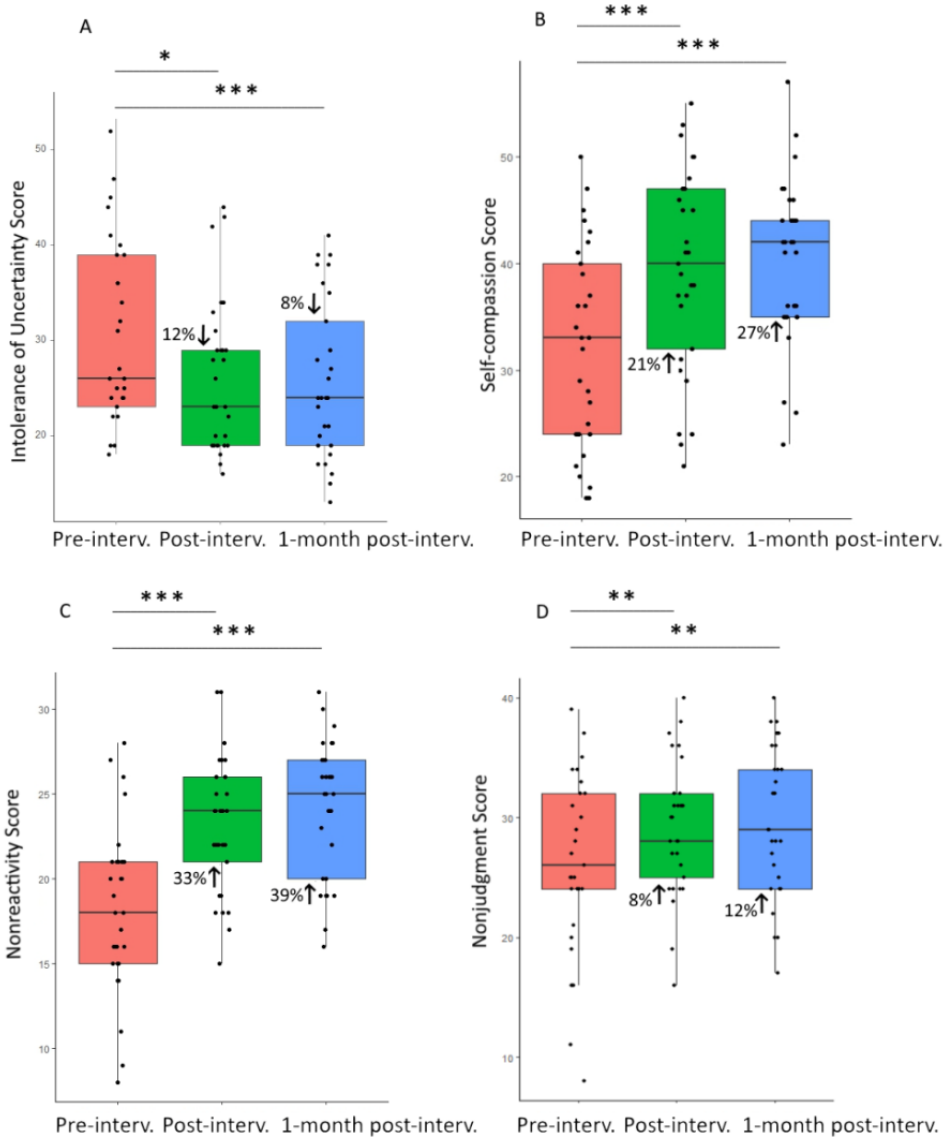
Wilcoxon signed rank tests also found 33% increase in nonreactivity scores from baseline to postintervention (W_{28}=38.5; *P*<.001; *r*=0.72) and 39% increase 1 month after the end of the intervention (W_{28}=9; *P*<.001; *r*=0.83) and 8% increase in nonjudgment scores at postintervention (W_{28}=60.5; *P*=.009; *r*=0.50) and 12% increase 1 month later (W_{28}=38.5; *P*=.003; *r*=0.62; Figure 7). We observed similar results with data only from physicians: a significant change across the 3 time points for nonreactivity (χ^2_2=18.8; *P*<.001; W=0.47) and for nonjudgment (χ^2_2=7.10; *P*=.03; W=0.18) and 28% increase in nonreactivity scores at postintervention (W_{19}=21; *P*=.003; *r*=0.71) and 39% increase 1 month later (W_{19}=4; *P*<.001; *r*=0.85) and 10% increase in nonjudgment scores at postintervention (W_{19}=40.5; *P*=.08; *r*=0.43) and 12% increase 1 month later (W_{19}=19.5; *P*=.02; *r*=0.62).

Figure 6. Box and whisker plots at baseline, at postintervention, and 1 month after intervention completion for (A) Generalized Anxiety Disorder-7 scores, (B) cynicism scores, and (C) emotional exhaustion scores from the Maslach Burnout Inventory in 29 physicians and nurse practitioners. **P*=.05, ***P*=.01.



XSL•FO
RenderX

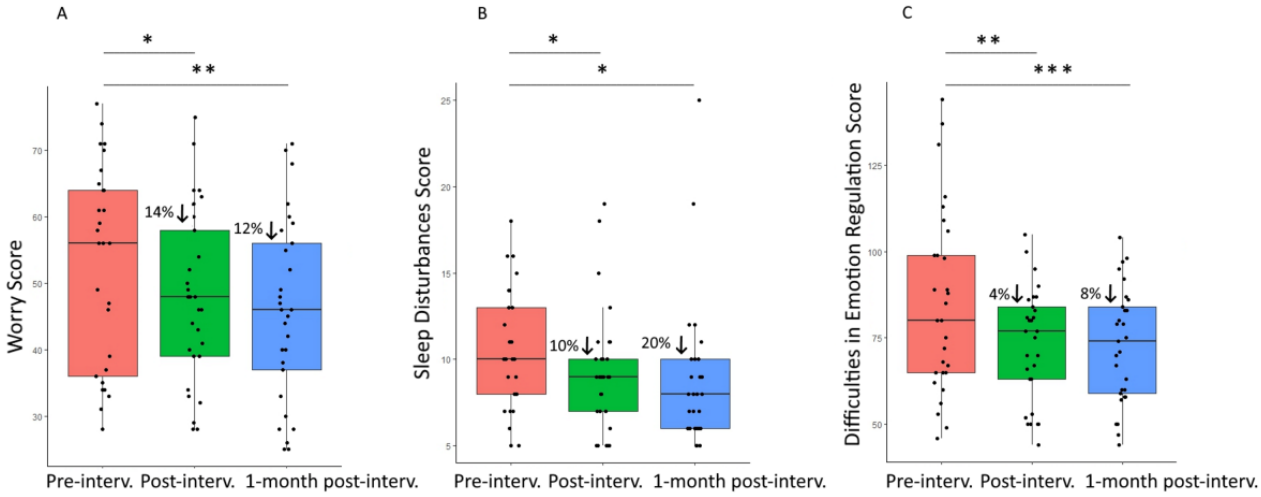
Figure 7. Box and whisker plots at baseline, at postintervention, and 1 month after intervention completion for (A) intolerance of uncertainty scores, (B) self-compassion scores, and (C) nonreactivity toward inner experiences scores in 29 physicians and nurse practitioners. *P=.05, **P=.01,***P=.001.



Finally, Wilcoxon signed rank tests revealed 14% reduction in worry scores from baseline to postintervention (W_{28}=247; P=.04; r=0.40) and 12% reduction 1 month after the end of the intervention (W_{28}=319; P=.006; r=0.58), 10% reduction in sleep disturbances at postintervention (W_{28}=283; P=.04; r=.42) and 20% reduction 1 month later (W_{28}=268; P=.01; r=0.53), and 4% reduction in difficulties in emotion regulation scores at postintervention (W_{28}=353; P=.005; r=0.54) and 8% reduction 1 month later (W_{28}=392; P<.001; r=0.70; Figure 8). We observe less strong results with data only from physicians: a trend change across the 3 time points for worry (χ^2_2=5.30; P=.07; W=0.13) and significant changes for sleep (χ^2_2=7.53; P=.02; W=0.188) and emotion regulation (χ^2_2=9.29; P=.009; W=0.23). We only observed 21% reduction in worry at 1 month after the end of the intervention (W_{19}=168; P=.01; r=0.66) but not at postintervention (W_{19}=116; P=.09; r=0.37). We found 22% reduction in sleep disturbances scores from baseline to postintervention (W_{19}=156; P=.02; r=0.54) and 30% reduction 1 month later (W_{19}=152; P=.01; r=0.64) and 8% reduction in difficulties in emotion regulation at postintervention (W_{19}=176; P=.01; r=0.59) and 11% reduction 1 month later (W_{19}=192; P=.04; r=0.73).

XSL•FO
RenderX

Figure 8. Box and whisker plots at baseline, at postintervention, and 1 month after intervention completion for (A) worry scores, (B) sleep disturbances scores, and (C) difficulties in emotion regulation scores in 29 physicians and nurse practitioners. *P=.05, **P=.01,***P=.001.



Correlations Between Burnout and Other Outcomes

Table 5 shows the correlations between the 2 burnout dimensions of cynicism and emotional exhaustion and the other outcomes at each assessment point. There were significant correlations between anxiety and burnout at baseline and 1 month after the end of the intervention (cynicism: r≥0.39; P≤.04 and emotional exhaustion: r≥0.58; P≤.001) but not right after the intervention. Burnout were correlated also with self-compassion scores (cynicism: r≥−0.42; P≤.02 and emotional exhaustion: r=−0.56; P=.002), intolerance of uncertainty scores (cynicism: r≥0.37; P≤.05 and emotional exhaustion: r≥0.40; P≤.03), and depression (cynicism: r=0.41; P=.03 and emotional exhaustion: r=0.41; P=.03). Burnout scores were also correlated with worry (cynicism: r≥0.45; P≤.01 and emotional exhaustion: r≥0.51; P≤.004) and emotion regulation (cynicism: r≥0.46; P≤.01 and emotional exhaustion: r=0.40; P=.03). Cynicism scores were also correlated with nonjudgment at baseline (r=−0.49; P=.006) and with sleep after the intervention (r=0.38; P=.04). Emotional exhaustion scores were also correlated with nonreactivity scores only at baseline (r=−0.56; P=.002). We did not observe any other correlations (r≤|0.36|; P≥|.054|).

XSL•FO
RenderX

Table 5. Correlations between anxiety and burnout scores at baseline, postintervention, and 1-month postintervention.

Variables	Baseline	Postintervention	1-mo postintervention
GAD-7[a] × cynicism			
Value	0.56	0.41	0.24
P value	.002	.03	.22
GAD-7 × emotional exhaustion			
Value	0.60	0.53	0.32
P value	<.001	.004	.10
PHQ-2[b] × cynicism			
Value	0.22	0.05	0.41
P value	.25	.81	.03
PHQ-2 × emotional exhaustion			
Value	0.35	0.23	0.41
P value	.06	.24	.03
IU[c] × cynicism			
Value	0.52	0.28	0.37
P value	.004	.13	.05
IU × emotional exhaustion			
Value	0.41	0.45	0.4
P value	.03	.01	.03
Self-compassion × cynicism			
Value	−0.52	−0.42	−0.62
P value	.004	.002	<.001
Self-compassion × emotional exhaustion			
Value	−0.35	−0.17	−0.56
P value	.06	.37	.002
Nonreactivity × cynicism			
Value	−0.27	−0.20	−0.3
P value	.15	.29	.11
Nonreactivity × emotional exhaustion			
Value	−0.56	−0.3	−0.26
P value	.002	.12	.12
Nonjudgment × cynicism			
Value	−0.49	−0.3	−0.34
P value	.006	.11	.07
Nonjudgment × emotional exhaustion			
Value	−0.25	−0.18	−0.34
P value	.19	.35	.08
Worry × cynicism			
Value	−0.49	−0.45	−0.64
P value	.007	.01	.001
Worry × emotional distress			
Value	0.51	0.32	0.54

XSL•FO
RenderX

Variables	Baseline	Postintervention	1-mo postintervention
P value	.004	.09	.003
Difficulties in emotion regulation × cynicism			
Value	0.49	0.33	0.53
P value	.01	.08	.003
Difficulties in emotion regulation × emotional exhaustion			
Value	0.35	0.31	0.4
P value	.07	.10	.03
Sleep disturbances × cynicism			
Value	−0.22	−0.38	−0.25
P value	.26	.04	.19
Sleep disturbances × emotional exhaustion			
Value	−0.29	−0.2	0.36
P value	.12	.30	.60

[a]GAD-7: Generalized Anxiety Disorder-7.

[b]PHQ-2: Patient Health Questionnaire-2.

[c]IU: intolerance of uncertainty.

Discussion

Principal Findings

To our knowledge, this is the first research to assess the efficacy of a digital MT program designed with input from clinicians specifically to address burnout. Two independent single-arm studies were conducted to evaluate its impact on physicians. In the first study, MT was delivered via podcast, while in the second study, it was administered through a free mobile app.

Results from both studies demonstrated that MT led to reductions in burnout, anxiety, intolerance of uncertainty, and personal distress, while also fostering increases in self-compassion and mindfulness, with effect sizes ranging from medium to large. In addition, the second study revealed decreases in worry, sleep disturbances, and difficulties in emotion regulation. However, no significant changes were observed over time for depression or perspective taking and empathic concern, potentially due to the already low levels of depression and high levels of perspective taking and empathic concern reported at baseline. Furthermore, significant positive correlations were found between burnout and anxiety at baseline in both studies, at postintervention in study 1, and at follow-up in study 2, thus confirming the relationship between these 2 variables.

Reduction of Physician Burnout

Preliminary findings suggest that this training could be an efficacious, accessible way for busy physicians to combat burnout. Both studies showed that it reduced cynicism, with the first also reducing emotional exhaustion. This may be due to the focus of the intervention on individual well-being, particularly addressing the interpersonal aspect of burnout. However, individual-focused interventions alone may not fully address physician burnout, which often requires collaborative efforts at both the individual and organizational levels [33,83].

Increase of Mindfulness and Self-Compassion

This study builds upon prior research demonstrating that in-person mindfulness-based interventions have been effective in reducing burnout and anxiety among physicians [37-40]. Importantly, we were able to replicate these findings using a digital program that is both free and less time consuming than the intensive in-person MT used in previous studies. This digital intervention may therefore offer a more feasible option for integration into the already hectic schedules of physicians. In addition, this intervention contributes to the growing field of digital therapeutics by demonstrating clinically meaningful effects [7,60,61,84].

Consistent with numerous studies, we found that MT increased nonreactivity and nonjudgmental attitudes toward thoughts, sensations, and emotions, while also enhancing self-compassion. Self-compassion has been theorized to provide self-care during times of distress, failure, and difficulty and to enable individuals to offer compassionate care to others. This concept comprises 3 interconnected components: mindfulness, self-kindness, and common humanity [46]. In practical terms, self-compassion can facilitate several key aspects: (1) recognizing and accepting feelings as they arise without suppression or avoidance (ie, mindfulness), (2) treating oneself with kindness and understanding (ie, self-kindness), and (3) acknowledging one's own pain and distress as part of the broader human experience (ie, common humanity [46]).

Therefore, self-compassion may enable health care providers, who often struggle with perfectionism and self-criticism in high-stakes environments where errors can have grave consequences, to better navigate the challenges they face. Such feelings of inadequacy have been linked to anxiety, depression, and even suicidal ideation among medical students [85,86] and physicians [87]. For instance, interventions promoting self-compassion among athletes have proven successful in

XSL•FO
RenderX

helping them manage concerns about mistakes, self-criticism, and rumination during setbacks and uncertainties, thereby improving both athletic performance and psychological well-being [88].

Reduction of Personal Distress and Intolerance of Uncertainty

Regarding empathy, we only observed a significant reduction in personal distress during postintervention assessments. This finding aligns with existing literature, which has consistently reported that personal distress is the empathic component positively associated with burnout [89]. However, we did not observe any effects on the other 2 dimensions of empathy: perspective taking and empathic concern. This outcome may be due to a ceiling effect, as previous research with health care professionals has also reported [90]. The high baseline scores suggest that the chosen questionnaire may not have been sensitive enough to capture changes in these dimensions among physicians.

Although previous literature has shown a link between high intolerance of uncertainty and an increased risk or presence of physician burnout [24,91] and work-related stress and satisfaction [21,92], these 2 studies are the first to investigate the effects of MT on intolerance of uncertainty. We found a significant reduction with a large effect size. Recognizing uncertainty and developing coping strategies are crucial in the field of medicine and are recognized as essential clinical competencies to be fostered throughout medical education and training [93,94].

For example, when faced with clinical situations where outcomes are difficult to predict, physicians with high intolerance of uncertainty may experience excessive worry and anxiety about the consequences of their decisions. They may also struggle with self-doubt and insecurity regarding their professional competence, leading to a sense of being “stuck” in uncertainty and hesitancy in decision-making [95,96]. It is estimated that 17% of the excessive costs in medical care result from physicians’ anxiety related to uncertainty management, leading to increased tendencies to order further tests [97], failures to adhere to evidence-based guidelines [98], and fear of malpractice litigation and defensive practices [99].

Reduction of Anxiety and Related Variables

Importantly, we observed that MT reduced anxiety by 40% postintervention, with moderate to large effect sizes. Furthermore, we confirmed the positive correlation between burnout and anxiety, although this correlation was not significant across all 3 time points in both studies. These results build upon previous findings regarding the relationship between burnout and anxiety [7].

Interestingly, MT also reduced variables previously associated with the onset and maintenance of anxiety, such as worry, sleep disturbances, and difficulties with emotion regulation. In previous research, app-based MT significantly reduced anxiety in individuals with generalized anxiety disorders [61] and improved sleep disturbances [84]. These effects were mediated by a reduction in worry [61,84]. MT appears to assist individuals in interrupting habitual worry patterns by fostering present-focused awareness of thoughts and emotions. Guiding physicians to observe and manage repetitive worry patterns rather than reacting to and reinforcing them could alleviate anxiety and burnout.

In addition, research indicates that being aware of one’s own emotions and regulating them in stressful work environments can not only reduce personal distress but also promote prosocial behavior [100,101]. For instance, when physicians encounter patients with sudden worsening symptoms, they need to manage their own discomfort to provide appropriate care for their patients considered distressed.

Limitations

These studies offer several strengths and significant findings, such as a freely accessible training program designed for real-world application, preregistered outcomes, and the inclusion of a diverse range of medical specialties in the samples. However, they also have several limitations.

First, both studies are single-arm trials with relatively short follow-up periods (ie, 1 month after completing the training). To validate the efficacy of this program and ascertain its long-term effects, future studies using a waiting list or active control intervention with longer follow-up periods are warranted. These studies could also investigate the underlying psychological mechanisms.

Second, the study samples are relatively small and not fully representative, with predominantly White and female physician participants. While future studies with balanced representation across sexes are necessary for generalizability, it is worth noting that women are more likely to be affected by burnout and anxiety compared to men [102-104]. In addition, it is important to recognize that burnout affects not only physicians but also other health care professionals and students. Moving forward, it is critical to adapt and test interventions across a wider population within the health care system. Addressing clinician burnout early in professional development and acknowledging specific stressors in the learning environment are essential steps in mitigating its impact. In addition, both studies involved self-selection of participants, meaning that those who chose to participate were likely already predisposed to favor the program. This suggests that the intervention might have been particularly beneficial for individuals who were attracted to the mindfulness component of the program. As a result, we do not know whether the intervention would affect those who were less interested in mindfulness or less inclined to participate.

Finally, the impact of the intervention was assessed exclusively through self-report measures and individual-related variables, without considering the individual-organization relationship. As we have noted, both individual and organizational factors contribute to burnout. Similar to recent findings in the field [105], future research should incorporate measures such as work engagement, teamwork, and feelings of being valued by the organization, in addition to patient-related ecological outcomes. While collecting data on mortality rates, complication rates, and disease progression might be challenging across different health care institutions, other measures such as patient compliance, the strength of the patient–health care provider

XSL•FO
RenderX

relationship, compassionate care, and testing and prescribing behaviors can be effectively implemented. Understanding these broader outcomes is crucial for developing effective interventions and strategies to address physician burnout and improve patient care and health outcomes.

Conclusions

Despite these limitations, both studies demonstrated that physicians were open to trying a pragmatically designed and delivered digital MT program and could potentially benefit from it. They suggest that digital MT holds promise as an accessible, evidence-based tool for addressing burnout and related aspects of anxiety among physicians. However, it is important to note that causal claims cannot be definitively made until future randomized controlled studies are conducted. These proof-of-concept studies serve as an initial exploration of potential mechanisms and effect size calculations for larger randomized controlled trials.

Furthermore, this evidence underscores the need for both individual-focused and organizational-focused approaches to addressing burnout, and it highlights the importance of involving physicians in the design of new interventions to create effective and sustainable strategies for preventing or mitigating burnout and promoting professional well-being.

Acknowledgments

The authors are grateful to Lauren Allister, Jacqui Barnett, Courtney Bilodeau, Kenneth Botelho, Robin Boudette, Ruth Chen, Christine Evans, Elizabeth Freire, Alisa Hale, Donna Hill Howes, Ana Hope Logghe, Katherine King, Lucia Larson, Nancy Lasson, Beth Lee-Herbert, Laurie Marbas, Gemma Majewski, Mary McLellan, Diana Meeks, Veronica Modise, Shantell Moreno, Chika Nwokenna, Samba Nyirenda, Nkenna Odom, Saley Pels, Heidi Rayala, Cynthia Rivera, Nickholas Ross, Nancy Quesada-Rodriguez, Chiara Saracci, Greg Salgueiro, Paul Sidandi, Rob Suhoza, David Scott Upton, Vinicius Vieira Neves, and Hannah Whitmore for their contribution to the development of the mindfulness intervention used in these studies. This research was financed by the Swiss National Science Foundation (P500PS_202924) and was made possible also in part by a grant from the Peter G Peterson Foundation (#008225). The statements and views expressed are solely the responsibility of the authors.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

LA and JB were responsible for the conceptualization of the study, and both authors were involved in securing funding for the research. The investigation process was carried out by LA and JB, who also played key roles in developing the methodology and overseeing the project administration. LA managed data curation and formal analysis, ensuring the accuracy and interpretation of the data. In addition, LA wrote the first draft of the manuscript, which was subsequently reviewed and edited by both authors.

Conflicts of Interest

JB owns stock in MindSciences, the company that developed the mindfulness app used in this study. This financial interest has been disclosed to Brown University and is being managed in accordance with its Conflict of Interest and Conflict of Commitment policies. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Descriptive statistics for primary and secondary outcomes (N=29 physicians+nurse practitioners).
[DOCX File , 20 KB - formative_v9i1e63197_app1.docx]

References

1. Shanafelt TD, Boone S, Tan L, Dyrbye LN, Sotile W, Satele D, et al. Burnout and satisfaction with work-life balance among US physicians relative to the general US population. Arch Intern Med 2012 Oct 08;172(18):1377-1385. [doi: 10.1001/archinternmed.2012.3199] [Medline: 22911330]
2. Shanafelt TD, Gradishar WJ, Kosty M, Satele D, Chew H, Horn L, et al. Burnout and career satisfaction among US oncologists. J Clin Oncol 2014 Mar 01;32(7):678-686 [FREE Full text] [doi: 10.1200/JCO.2013.51.8480] [Medline: 24470006]
3. Shanafelt TD, Hasan O, Dyrbye LN, Sinsky C, Satele D, Sloan J, et al. Changes in burnout and satisfaction with work-life balance in physicians and the general US working population between 2011 and 2014. Mayo Clin Proc 2015 Dec;90(12):1600-1613. [doi: 10.1016/j.mayocp.2015.08.023] [Medline: 26653297]
4. Shanafelt TD, West CP, Dyrbye LN, Trockel M, Tutty M, Wang H, et al. Changes in burnout and satisfaction with work-life integration in physicians during the first 2 years of the COVID-19 pandemic. Mayo Clin Proc 2022 Dec;97(12):2248-2258 [FREE Full text] [doi: 10.1016/j.mayocp.2022.09.002] [Medline: 36229269]

XSL•FO
RenderX

5. Burn-out an "occupational phenomenon": international classification of diseases. World Health Organization. 2019. URL: https://www.who.int/news/item/28-05-2019-burn-out-an-occupational-phenomenon-international-classification-of-diseases [accessed 2024-06-07]
6. Maslach C, Leiter MP. New insights into burnout and health care: strategies for improving civility and alleviating burnout. Med Teach 2017 Feb;39(2):160-163. [doi: 10.1080/0142159X.2016.1248918] [Medline: 27841065]
7. Roy A, Druker S, Hoge EA, Brewer JA. Physician anxiety and burnout: symptom correlates and a prospective pilot study of app-delivered mindfulness training. JMIR Mhealth Uhealth 2020 Apr 01;8(4):e15608 [FREE Full text] [doi: 10.2196/15608] [Medline: 32234708]
8. Melamed S, Ugarten U, Shirom A, Kahana L, Lerman Y, Froom P. Chronic burnout, somatic arousal and elevated salivary cortisol levels. J Psychosom Res 1999 Jun;46(6):591-598. [doi: 10.1016/s0022-3999(99)00007-0] [Medline: 10454175]
9. Vela-Bueno A, Moreno-Jiménez B, Rodríguez-Muñoz A, Olavarrieta-Bernardino S, Fernández-Mendoza J, De la Cruz-Troca JJ, et al. Insomnia and sleep quality among primary care physicians with low and high burnout levels. J Psychosom Res 2008 Apr;64(4):435-442. [doi: 10.1016/j.jpsychores.2007.10.014] [Medline: 18374744]
10. Weaver MD, Robbins R, Quan SF, O'Brien CS, Viyaran NC, Czeisler CA, et al. Association of sleep disorders with physician burnout. JAMA Netw Open 2020 Oct 01;3(10):e2023256 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.23256] [Medline: 33125492]
11. Ryan E, Hore K, Power J, Jackson T. The relationship between physician burnout and depression, anxiety, suicidality and substance abuse: a mixed methods systematic review. Front Public Health 2023 Mar 30;11:1133484 [FREE Full text] [doi: 10.3389/fpubh.2023.1133484] [Medline: 37064688]
12. Shanafelt TD, Balch CM, Dyrbye L, Bechamps G, Russell T, Satele D, et al. Special report: suicidal ideation among American surgeons. Arch Surg 2011 Jan;146(1):54-62. [doi: 10.1001/archsurg.2010.292] [Medline: 21242446]
13. Schernhammer ES, Colditz GA. Suicide rates among physicians: a quantitative and gender assessment (meta-analysis). Am J Psychiatry 2004 Dec;161(12):2295-2302. [doi: 10.1176/appi.ajp.161.12.2295] [Medline: 15569903]
14. Sinsky CA, Brown RL, Stillman MJ, Linzer M. COVID-related stress and work intentions in a sample of US health care workers. Mayo Clin Proc Innov Qual Outcomes 2021 Dec;5(6):1165-1173 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2021.08.007] [Medline: 34901752]
15. Han S, Shanafelt TD, Sinsky CA, Awad KM, Dyrbye LN, Fiscus LC, et al. Estimating the attributable cost of physician burnout in the United States. Ann Intern Med 2019 Jun 04;170(11):784-790. [doi: 10.7326/M18-1422] [Medline: 31132791]
16. Dewa CS, Loong D, Bonato S, Trojanowski L. The relationship between physician burnout and quality of healthcare in terms of safety and acceptability: a systematic review. BMJ Open 2017 Jun 21;7(6):e015141 [FREE Full text] [doi: 10.1136/bmjopen-2016-015141] [Medline: 28637730]
17. Salyers MP, Bonfils KA, Luther L, Firmin RL, White DA, Adams EL, et al. The relationship between professional burnout and quality and safety in healthcare: a meta-analysis. J Gen Intern Med 2017 Apr;32(4):475-482 [FREE Full text] [doi: 10.1007/s11606-016-3886-9] [Medline: 27785668]
18. Dyrbye LN, Varkey P, Boone SL, Satele DV, Sloan JA, Shanafelt TD. Physician satisfaction and burnout at different career stages. Mayo Clin Proc 2013 Dec;88(12):1358-1367. [doi: 10.1016/j.mayocp.2013.07.016] [Medline: 24290109]
19. Southwick FS, Southwick SM. The loss of a sense of control as a major contributor to physician burnout: a neuropsychiatric pathway to prevention and recovery. JAMA Psychiatry 2018 Jul 01;75(7):665-666. [doi: 10.1001/jamapsychiatry.2018.0566] [Medline: 29799941]
20. Freudenberger HJ. Staff burn‐out. J Soc Issues 2010 Apr 14;30(1):159-165. [doi: 10.1111/j.1540-4560.1974.tb00706.x]
21. Iannello P, Mottini A, Tirelli S, Riva S, Antonietti A. Ambiguity and uncertainty tolerance, need for cognition, and their association with stress. A study among Italian practicing physicians. Med Educ Online 2017;22(1):1270009 [FREE Full text] [doi: 10.1080/10872981.2016.1270009] [Medline: 28178917]
22. Begin AS, Hidrue M, Lehrhoff S, Del Carmen MG, Armstrong K, Wasfy JH. Factors associated with physician tolerance of uncertainty: an observational study. J Gen Intern Med 2022 May;37(6):1415-1421 [FREE Full text] [doi: 10.1007/s11606-021-06776-8] [Medline: 33904030]
23. Strout TD, Hillen M, Gutheil C, Anderson E, Hutchinson R, Ward H, et al. Tolerance of uncertainty: a systematic review of health and healthcare-related outcomes. Patient Educ Couns 2018 Sep;101(9):1518-1537. [doi: 10.1016/j.pec.2018.03.030] [Medline: 29655876]
24. Cooke GP, Doust JA, Steele MC. A survey of resilience, burnout, and tolerance of uncertainty in Australian general practice registrars. BMC Med Educ 2013 Jan 07;13(1):2 [FREE Full text] [doi: 10.1186/1472-6920-13-2] [Medline: 23294479]
25. Klimecki O, Singer T. Empathic distress fatigue rather than compassion fatigue? Integrating findings from empathy research in psychology and social neuroscience. In: Oakley B, Knafo A, Madhavan G, Wilson DS, editors. Pathological Altruism. Oxford, UK: Oxford University Press; 2012:368-384.
26. Burns DD, Nolen-Hoeksema S. Therapeutic empathy and recovery from depression in cognitive-behavioral therapy: a structural equation model. J Consult Clin Psychol 1992 Jun;60(3):441-449. [doi: 10.1037//0022-006x.60.3.441] [Medline: 1619098]

XSL•FO
RenderX

27. Del Canale S, Louis DZ, Maio V, Wang X, Rossi G, Hojat M, et al. The relationship between physician empathy and disease complications: an empirical study of primary care physicians and their diabetic patients in Parma, Italy. Acad Med 2012 Sep;87(9):1243-1249. [doi: 10.1097/ACM.0b013e3182628fbf] [Medline: 22836852]
28. Hojat M, Louis DZ, Markham FW, Wender R, Rabinowitz C, Gonnella JS. Physicians' empathy and clinical outcomes for diabetic patients. Acad Med 2011 Mar;86(3):359-364. [doi: 10.1097/ACM.0b013e3182086fe1] [Medline: 21248604]
29. Pereira L, Figueiredo-Braga M, Carvalho IP. Preoperative anxiety in ambulatory surgery: the impact of an empathic patient-centered approach on psychological and clinical outcomes. Patient Educ Couns 2016 May;99(5):733-738. [doi: 10.1016/j.pec.2015.11.016] [Medline: 26654958]
30. Rakel DP, Hoeft TJ, Barrett BP, Chewning BA, Craig BM, Niu M. Practitioner empathy and the duration of the common cold. Fam Med 2009;41(7):494-501 [FREE Full text] [Medline: 19582635]
31. Mercer SW, Reynolds WJ. Empathy and quality of care. Br J Gen Pract 2002 Oct;52 Suppl(Suppl):S9-12 [FREE Full text] [Medline: 12389763]
32. Shanafelt TD, West C, Zhao X, Novotny P, Kolars J, Habermann T, et al. Relationship between increased personal well-being and enhanced empathy among. J Gen Intern Med 2005 Jul;20(7):559-564. [doi: 10.1007/S11606-005-0102-8]
33. West CP, Dyrbye LN, Erwin PJ, Shanafelt TD. Interventions to prevent and reduce physician burnout: a systematic review and meta-analysis. Lancet 2016 Nov 05;388(10057):2272-2281. [doi: 10.1016/S0140-6736(16)31279-X] [Medline: 27692469]
34. Ruotsalainen JH, Verbeek JH, Mariné A, Serra C. Preventing occupational stress in healthcare workers. Cochrane Database Syst Rev 2015 Apr 07;2015(4):CD002892 [FREE Full text] [doi: 10.1002/14651858.CD002892.pub5] [Medline: 25847433]
35. Luberto CM, Wasson RS, Kraemer KM, Sears RW, Hueber C, Cotton S. Feasibility, acceptability, and preliminary effectiveness of a 4-week mindfulness-based cognitive therapy protocol for hospital employees. Mindfulness (N Y) 2017 Dec;8(6):1522-1531 [FREE Full text] [doi: 10.1007/s12671-017-0718-x] [Medline: 29335671]
36. Cascales-Pérez ML, Ferrer-Cascales R, Fernández-Alcántara M, Cabañero-Martínez MJ. Effects of a mindfulness-based programme on the health- and work-related quality of life of healthcare professionals. Scand J Caring Sci 2021 Sep;35(3):881-891. [doi: 10.1111/scs.12905] [Medline: 32865258]
37. Goodman MJ, Schorling JB. A mindfulness course decreases burnout and improves well-being among healthcare providers. Int J Psychiatry Med 2012;43(2):119-128. [doi: 10.2190/PM.43.2.b] [Medline: 22849035]
38. Gracia Gozalo RM, Ferrer Tarrés JM, Ayora Ayora A, Alonso Herrero M, Amutio Kareaga A, Ferrer Roca R. Application of a mindfulness program among healthcare professionals in an intensive care unit: effect on burnout, empathy and self-compassion. Med Intensiva (Engl Ed) 2019 May;43(4):207-216. [doi: 10.1016/j.medin.2018.02.005] [Medline: 29544729]
39. Krasner MS, Epstein RM, Beckman H, Suchman AL, Chapman B, Mooney CJ, et al. Association of an educational program in mindful communication with burnout, empathy, and attitudes among primary care physicians. JAMA 2009 Sep 23;302(12):1284-1293. [doi: 10.1001/jama.2009.1384] [Medline: 19773563]
40. Martín-Asuero A, García-Banda G. The Mindfulness-based Stress Reduction program (MBSR) reduces stress-related psychological distress in healthcare professionals. Span J Psychol 2010 Nov 10;13(2):897-905. [doi: 10.1017/s1138741600002547] [Medline: 20977037]
41. Kabat-Zinn J. Full Catastrophe Living: Using the Wisdom of Your Body and Mind to Face Stress, Pain, and Illness. New York, NY: Bantam Books; 1991.
42. Lindsay EK, Creswell JD. Mechanisms of mindfulness training: Monitor and Acceptance Theory (MAT). Clin Psychol Rev 2017 Feb;51:48-59 [FREE Full text] [doi: 10.1016/j.cpr.2016.10.011] [Medline: 27835764]
43. Watson T, Walker O, Cann R, Varghese AK. The benefits of mindfulness in mental healthcare professionals. F1000Res 2021;10:1085 [FREE Full text] [doi: 10.12688/f1000research.73729.2] [Medline: 35387270]
44. Benzo RP, Anderson PM, Bronars C, Clark M. Mindfulness for healthcare providers: the role of non-reactivity in reducing stress. Explore (NY) 2018 Nov;14(6):453-456 [FREE Full text] [doi: 10.1016/j.explore.2018.03.008] [Medline: 30292600]
45. Wasson RS, Barratt C, O'Brien WH. Effects of mindfulness-based interventions on self-compassion in health care professionals: a meta-analysis. Mindfulness (N Y) 2020;11(8):1914-1934 [FREE Full text] [doi: 10.1007/s12671-020-01342-5] [Medline: 32421083]
46. Neff KD. Self-compassion: an alternative conceptualization of a healthy attitude toward oneself. Self Identity 2003 Apr;2(2):85-101. [doi: 10.1080/15298860309032]
47. Neff KD, Hsieh YP, Dejitterat K. Self-compassion, achievement goals, and coping with academic failure. Self Identity 2005 Jul;4(3):263-287. [doi: 10.1080/13576500444000317]
48. Sinclair S, Beamer K, Hack TF, McClement S, Raffin Bouchal S, Chochinov HM, et al. Sympathy, empathy, and compassion: a grounded theory study of palliative care patients' understandings, experiences, and preferences. Palliat Med 2017 May;31(5):437-447 [FREE Full text] [doi: 10.1177/0269216316663499] [Medline: 27535319]
49. Duarte J, Pinto-Gouveia J, Cruz B. Relationships between nurses' empathy, self-compassion and dimensions of professional quality of life: a cross-sectional study. Int J Nurs Stud 2016 Aug;60:1-11. [doi: 10.1016/j.ijnurstu.2016.02.015] [Medline: 27297364]

XSL•FO
RenderX

50. Kemper KJ, McClafferty H, Wilson PM, Serwint JR, Batra M, Mahan JD, Pediatric Resident Burnout-Resilience Study Consortium. Do mindfulness and self-compassion predict burnout in pediatric residents? Acad Med 2019 Jun;94(6):876-884. [doi: 10.1097/ACM.0000000000002546] [Medline: 30520809]
51. Montero-Marin J, Zubiaga F, Cereceda M, Piva Demarzo MM, Trenc P, Garcia-Campayo J. Burnout subtypes and absence of self-compassion in primary healthcare professionals: a cross-sectional study. PLoS One 2016;11(6):e0157499 [FREE Full text] [doi: 10.1371/journal.pone.0157499] [Medline: 27310426]
52. Olson K, Kemper KJ, Mahan JD. What factors promote resilience and protect against burnout in first-year pediatric and medicine-pediatric residents? J Evid Based Complementary Altern Med 2015 Jul 18;20(3):192-198. [doi: 10.1177/2156587214568894] [Medline: 25694128]
53. Babenko O, Mosewich AD, Lee A, Koppula S. Association of physicians' self-compassion with work engagement, exhaustion, and professional life satisfaction. Med Sci (Basel) 2019 Feb 12;7(2):29 [FREE Full text] [doi: 10.3390/medsci7020029] [Medline: 30759845]
54. Scheepers RA, Emke H, Epstein RM, Lombarts KM. The impact of mindfulness-based interventions on doctors' well-being and performance: a systematic review. Med Educ 2020 Feb;54(2):138-149 [FREE Full text] [doi: 10.1111/medu.14020] [Medline: 31868262]
55. Shapiro SL, Astin JA, Bishop SR, Cordova M. Mindfulness-based stress reduction for health care professionals: results from a randomized trial. Int J Stress Manag 2005 May;12(2):164-176. [doi: 10.1037/1072-5245.12.2.164]
56. Valley M, Stallones L. A thematic analysis of health care workers' adoption of mindfulness practices. Workplace Health Saf 2018 Nov;66(11):538-544. [doi: 10.1177/2165079918771991] [Medline: 29806801]
57. Swensen S, Kabcenell A, Shanafelt T. Physician-organization collaboration reduces physician burnout and promotes engagement: the mayo clinic experience. J Healthc Manag 2016;61(2):105-127. [Medline: 27111930]
58. Brewer J. Mindfulness training for addictions: has neuroscience revealed a brain hack by which awareness subverts the addictive process? Curr Opin Psychol 2019 Aug;28:198-203. [doi: 10.1016/j.copsyc.2019.01.014] [Medline: 30785066]
59. Brewer JA, Mallik S, Babuscio TA, Nich C, Johnson HE, Deleone CM, et al. Mindfulness training for smoking cessation: results from a randomized controlled trial. Drug Alcohol Depend 2011 Dec 01;119(1-2):72-80 [FREE Full text] [doi: 10.1016/j.drugalcdep.2011.05.027] [Medline: 21723049]
60. Mason AE, Jhaveri K, Cohn M, Brewer JA. Testing a mobile mindful eating intervention targeting craving-related eating: feasibility and proof of concept. J Behav Med 2018 Apr 16;41(2):160-173 [FREE Full text] [doi: 10.1007/s10865-017-9884-5] [Medline: 28918456]
61. Roy A, Hoge EA, Abrante P, Druker S, Liu T, Brewer JA. Clinical efficacy and psychological mechanisms of an app-based digital therapeutic for generalized anxiety disorder: randomized controlled trial. J Med Internet Res 2021 Dec 02;23(12):e26987 [FREE Full text] [doi: 10.2196/26987] [Medline: 34860673]
62. Gleichgerrcht E, Decety J. Empathy in clinical practice: how individual dispositions, gender, and experience moderate empathic concern, burnout, and emotional distress in physicians. PLoS One 2013 Apr 19;8(4):e61526 [FREE Full text] [doi: 10.1371/journal.pone.0061526] [Medline: 23620760]
63. Gleichgerrcht E, Decety J. The relationship between different facets of empathy, pain perception and compassion fatigue among physicians. Front Behav Neurosci 2014 Jul 11;8:243 [FREE Full text] [doi: 10.3389/fnbeh.2014.00243] [Medline: 25071495]
64. Neumann M, Edelhäuser F, Tauschel D, Fischer MR, Wirtz M, Woopen C, et al. Empathy decline and its reasons: a systematic review of studies with medical students and residents. Acad Med 2011 Aug;86(8):996-1009. [doi: 10.1097/ACM.0b013e318221e615] [Medline: 21670661]
65. Rowe JP, Shores LR, Mott BW, Lester JC. Integrating learning and engagement in narrative-centered learning environments. In: Proceedings of the 10th International Conference on Intelligent Tutoring Systems. 2010 Presented at: ITS '10; June 14-18, 2010; Pittsburgh, PA p. 166-177 URL: https://link.springer.com/chapter/10.1007/978-3-642-13437-1_17 [doi: 10.1007/978-3-642-13437-1_17]
66. West CP, Dyrbye LN, Sloan JA, Shanafelt TD. Single item measures of emotional exhaustion and depersonalization are useful for assessing burnout in medical professionals. J Gen Intern Med 2009 Dec 3;24(12):1318-1321 [FREE Full text] [doi: 10.1007/s11606-009-1129-z] [Medline: 19802645]
67. West CP, Dyrbye LN, Satele DV, Sloan JA, Shanafelt TD. Concurrent validity of single-item measures of emotional exhaustion and depersonalization in burnout assessment. J Gen Intern Med 2012 Nov 24;27(11):1445-1452 [FREE Full text] [doi: 10.1007/s11606-012-2015-7] [Medline: 22362127]
68. Maslach C, Jackson SE, Leiter MP. Maslach burnout inventory. In: Zalaquett CP, Wood RJ, editors. Evaluating Stress: A Book of Resources. New York, NY: Scarecrow Education; 1997:191-218.
69. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
70. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
71. Carleton RN, Norton MA, Asmundson GJ. Fearing the unknown: a short version of the Intolerance of Uncertainty Scale. J Anxiety Disord 2007 Jan;21(1):105-117. [doi: 10.1016/j.janxdis.2006.03.014] [Medline: 16647833]

XSL•FO
RenderX

72. Raes F, Pommier E, Neff KD, Van Gucht D. Construction and factorial validation of a short form of the Self-Compassion Scale. Clin Psychol Psychother 2011 Jun 08;18(3):250-255. [doi: 10.1002/cpp.702] [Medline: 21584907]
73. Baer RA, Smith GT, Lykins E, Button D, Krietemeyer J, Sauer S, et al. Construct validity of the five facet mindfulness questionnaire in meditating and nonmeditating samples. Assessment 2008 Sep;15(3):329-342. [doi: 10.1177/1073191107313003] [Medline: 18310597]
74. Davis MH. Measuring individual differences in empathy: evidence for a multidimensional approach. J Pers Soc Psychol 1983 Jan;44(1):113-126. [doi: 10.1037/0022-3514.44.1.113]
75. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
76. Roemer L, Salters K, Raffa SD, Orsillo SM. Fear and avoidance of internal experiences in GAD: preliminary tests of a conceptual model. Cogn Ther Res 2005 Feb;29(1):71-88. [doi: 10.1007/S10608-005-1650-2]
77. Cox RC, Olatunji BO. Sleep in the anxiety-related disorders: a meta-analysis of subjective and objective research. Sleep Med Rev 2020 Jun;51:101282. [doi: 10.1016/j.smrv.2020.101282] [Medline: 32109832]
78. Cisler JM, Olatunji BO, Feldner MT, Forsyth JP. Emotion regulation and the anxiety disorders: an integrative review. J Psychopathol Behav Assess 2010 Mar 16;32(1):68-82 [FREE Full text] [doi: 10.1007/s10862-009-9161-1] [Medline: 20622981]
79. Yu L, Buysse DJ, Germain A, Moul DE, Stover A, Dodds NE, et al. Development of short forms from the PROMIS™ sleep disturbance and sleep-related impairment item banks. Behav Sleep Med 2011 Dec 28;10(1):6-24 [FREE Full text] [doi: 10.1080/15402002.2012.636266] [Medline: 22250775]
80. Buysse DJ, Yu L, Moul DE, Germain A, Stover A, Dodds NE, et al. Development and validation of patient-reported outcome measures for sleep disturbance and sleep-related impairments. Sleep 2010 Jun;33(6):781-792 [FREE Full text] [doi: 10.1093/sleep/33.6.781] [Medline: 20550019]
81. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State worry questionnaire. Behav Res Ther 1990 Jan;28(6):487-495. [doi: 10.1016/0005-7967(90)90135-6] [Medline: 2076086]
82. Gratz KL, Roemer L. Multidimensional assessment of emotion regulation and dysregulation: development, factor structure, and initial validation of the difficulties in emotion regulation scale. J Psychopathol Behav Assess 2008 Oct 17;30(4):315. [doi: 10.1007/S10862-008-9102-4]
83. Carrau D, Janis JE. Physician burnout: solutions for individuals and organizations. Plast Reconstr Surg Glob Open 2021 Feb;9(2):e3418 [FREE Full text] [doi: 10.1097/GOX.0000000000003418] [Medline: 33680666]
84. Gao M, Roy A, Deluty A, Sharkey KM, Hoge EA, Liu T, et al. Targeting anxiety to improve sleep disturbance: a randomized clinical trial of app-based mindfulness training. Psychosom Med 2022 Jun 01;84(5):632-642 [FREE Full text] [doi: 10.1097/PSY.0000000000001083] [Medline: 35420589]
85. Eley DS, Bansal V, Leung J. Perfectionism as a mediator of psychological distress: implications for addressing underlying vulnerabilities to the mental health of medical students. Med Teach 2020 Nov;42(11):1301-1307. [doi: 10.1080/0142159X.2020.1805101] [Medline: 32805157]
86. Enns MW, Cox BJ, Sareen J, Freeman P. Adaptive and maladaptive perfectionism in medical students: a longitudinal investigation. Med Educ 2001 Nov 04;35(11):1034-1042. [doi: 10.1046/j.1365-2923.2001.01044.x] [Medline: 11703639]
87. Craiovan PM. Correlations between perfectionism, stress, psychopathological symptoms and burnout in the medical field. Procedia Soc Behav Sci 2014 Apr;127:529-533. [doi: 10.1016/j.sbspro.2014.03.304]
88. Mosewich AD, Crocker PR, Kowalski KC, Delongis A. Applying self-compassion in sport: an intervention with women athletes. J Sport Exerc Psychol 2013 Oct;35(5):514-524. [doi: 10.1123/jsep.35.5.514] [Medline: 24197719]
89. Delgado N, Delgado J, Betancort M, Bonache H, Harris L. What is the link between different components of empathy and burnout in healthcare professionals? A systematic review and meta-analysis. Psychol Res Behav Manag 2023;16:447-463 [FREE Full text] [doi: 10.2147/PRBM.S384247] [Medline: 36814637]
90. Boellinghaus I, Jones FW, Hutton J. The role of mindfulness and loving-kindness meditation in cultivating self-compassion and other-focused concern in health care professionals. Mindfulness 2012 Oct 21;5(2):129-138. [doi: 10.1007/S12671-012-0158-6]
91. Kuhn G, Goldberg R, Compton S. Tolerance for uncertainty, burnout, and satisfaction with the career of emergency medicine. Ann Emerg Med 2009 Jul;54(1):106-13.e6. [doi: 10.1016/j.annemergmed.2008.12.019] [Medline: 19201058]
92. Bovier PA, Perneger TV. Stress from uncertainty from graduation to retirement--a population-based study of Swiss physicians. J Gen Intern Med 2007 May 9;22(5):632-638 [FREE Full text] [doi: 10.1007/s11606-007-0159-7] [Medline: 17443371]
93. Englander R, Cameron T, Ballard AJ, Dodge J, Bull J, Aschenbrener CA. Toward a common taxonomy of competency domains for the health professions and competencies for physicians. Acad Med 2013 Aug;88(8):1088-1094. [doi: 10.1097/ACM.0b013e31829a3b2b] [Medline: 23807109]
94. Simpson JG, Furnace J, Crosby J, Cumming AD, Evans PA, Friedman Ben David M, et al. The Scottish doctor--learning outcomes for the medical undergraduate in Scotland: a foundation for competent and reflective practitioners. Med Teach 2002 Mar;24(2):136-143. [doi: 10.1080/01421590220120713] [Medline: 12098432]
95. Lally J, Cantillon P. Uncertainty and ambiguity and their association with psychological distress in medical students. Acad Psychiatry 2014 Jun 10;38(3):339-344. [doi: 10.1007/s40596-014-0100-4] [Medline: 24718768]

XSL•FO
RenderX

96. Simpkin AL, Schwartzstein RM. Tolerating uncertainty - the next medical revolution? N Engl J Med 2016 Nov 03;375(18):1713-1715. [doi: 10.1056/NEJMp1606402] [Medline: 27806221]
97. Allison JJ, Kiefe CI, Cook EF, Gerrity MS, Orav EJ, Centor R. The association of physician attitudes about uncertainty and risk taking with resource use in a Medicare HMO. Med Decis Making 1998;18(3):320-329. [doi: 10.1177/0272989X9801800310] [Medline: 9679997]
98. Ghosh AK. On the challenges of using evidence-based information: the role of clinical uncertainty. J Lab Clin Med 2004 Aug;144(2):60-64. [doi: 10.1016/j.lab.2004.05.013] [Medline: 15322499]
99. Benbassat J, Pilpel D, Schor R. Physicians' attitudes toward litigation and defensive practice: development of a scale. Behav Med 2001;27(2):52-60. [doi: 10.1080/08964280109595771] [Medline: 11763825]
100. Eisenberg N, Fabes RA, Murphy B, Karbon M, Maszk P, Smith M, et al. The relations of emotionality and regulation to dispositional and situational empathy-related responding. J Pers Soc Psychol 1994 Apr;66(4):776-797. [doi: 10.1037//0022-3514.66.4.776] [Medline: 8189352]
101. Ochsner KN, Gross JJ. The cognitive control of emotion. Trends Cogn Sci 2005 May;9(5):242-249. [doi: 10.1016/j.tics.2005.03.010] [Medline: 15866151]
102. Catuzzi JE, Beck KD. Anxiety vulnerability in women: a two-hit hypothesis. Exp Neurol 2014 Sep;259:75-80. [doi: 10.1016/j.expneurol.2014.01.023] [Medline: 24518489]
103. Hoff T, Lee DR. Burnout and physician gender: what do we know? Med Care 2021 Aug 01;59(8):711-720. [doi: 10.1097/MLR.0000000000001584] [Medline: 34081678]
104. Lyubarova R, Salman L, Rittenberg E. Gender differences in physician burnout: driving factors and potential solutions. Perm J 2023 Jun 15;27(2):130-136 [FREE Full text] [doi: 10.7812/TPP/23.023] [Medline: 37303223]
105. Epstein RM, Marshall F, Sanders M, Krasner MS. Effect of an intensive mindful practice workshop on patient-centered compassionate care, clinician well-being, work engagement, and teamwork. J Contin Educ Health Prof 2022 Jan 01;42(1):19-27. [doi: 10.1097/CEH.0000000000000379] [Medline: 34459443]

Abbreviations

MT: mindfulness training
PROMIS: Patient-Reported Outcomes Measurement Information System

Edited by A Mavragani; submitted 12.06.24; peer-reviewed by MS Krasner, R Epstein; comments to author 27.08.24; revised version received 15.09.24; accepted 04.10.24; published 24.01.25.

Please cite as:
Antico L, Brewer J
Digital Mindfulness Training for Burnout Reduction in Physicians: Clinician-Driven Approach
JMIR Form Res 2025;9:e63197
URL: https://formative.jmir.org/2025/1/e63197
doi:10.2196/63197
PMID:

©Lia Antico, Judson Brewer. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Factors Influencing Telemedicine Adoption Among Health Care Professionals: Qualitative Interview Study

Fiona Schürmann[1*], PhD; Daniel Westmattelmann[1*], PhD; Gerhard Schewe[1], PhD

Center for Management, University of Münster, Münster, Germany

[*]these authors contributed equally

Corresponding Author:
Daniel Westmattelmann, PhD
Center for Management
University of Münster
Schlossplatz 3
Münster, 48149
Germany
Phone: 49 2518325010
Email: d.west@uni-muenster.de

Abstract

Background: Telemedicine is transforming health care by enabling remote diagnosis, consultation, and treatment. Despite rapid adoption during the COVID-19 pandemic, telemedicine uptake among health care professionals (HCPs) remains inconsistent due to perceived risks and lack of tailored policies. Existing studies focus on patient perspectives or general adoption factors, neglecting the complex interplay of contextual variables and trust constructs influencing HCPs' telemedicine adoption. This gap highlights the need for a framework integrating risks, benefits, and trust in telemedicine adoption, while addressing health care's unique dynamics.

Objective: This study aimed to adapt and extend the extended valence framework (EVF) to telemedicine, deconstructing factors driving adoption from an HCP perspective. Specifically, it investigated the nuanced roles of perceived risks, benefits, and trust referents (eg, technology, treatment, technology provider, and patient) in shaping behavioral intentions, while integrating contextual factors.

Methods: We used a qualitative research design involving semistructured interviews with 14 HCPs experienced in offering video consultations. The interview data were analyzed with deductive and inductive coding based on the EVF. Two coders conducted the coding process independently, achieving an intercoder reliability of 86.14%. The qualitative content analysis aimed to uncover the nuanced perspectives of HCPs, identifying key risk and benefit dimensions and trust referents relevant to telemedicine adoption.

Results: The study reveals the complex considerations HCPs have when adopting telemedicine. Perceived risks were multidimensional, including performance risks such as treatment limitations (mentioned by 7/14, 50% of the participants) and reliance on technical proficiency of patients (5/14, 36%), privacy risks related to data security (10/14, 71%), and time and financial risks associated with training (7/14, 50%) and equipment costs (4/14, 29%). Perceived benefits encompassed convenience through reduced travel time (5/14, 36%), improved care quality due to higher accessibility (8/14, 57%), and operational efficiency (7/14, 50%). Trust referents played a pivotal role; trust in technology was linked to functionality (6/14, 43%) and reliability (5/14, 36%), while trust in treatment depended on effective collaboration (9/14, 64%). Transparency emerged as a critical antecedent of trust across different referents, comprising disclosure, clarity, and accuracy. In addition, the study highlighted the importance of context-specific variables such as symptom characteristics (10/14, 71%) and prior professional experience with telemedicine (11/14, 79%).

Conclusions: This study expands the EVF for telemedicine, providing a framework integrating multidimensional risks, benefits, trust, and contextual factors. It advances theory by decomposing trust referents and transparency into actionable subdimensions and emphasizing context-specific variables. Practically, the findings guide stakeholders: policy makers should prioritize transparent regulations and data security, health care organizations should provide training and support for HCPs, and technology developers must design telemedicine solutions aligning with trust and usability needs. This understanding equips health care to address barriers, optimize adoption, and leverage telemedicine's potential for sustainable clinical integration.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e54777) doi:10.2196/54777

KEYWORDS

trust referents; multidimensional risk; benefits; transparency; technology adoption; telemedicine; extended valence framework

Introduction

Background

Generally, a rise in technological innovations is omnipresent and particularly impacts the health care sector [1,2]. Nowadays, telemedicine can provide a wide range of services, which is attributed to the continuous evolution of innovations [3]. Remote patient care, diagnosis, and treatment enabled by leveraging information and communication technologies offer health care services in addition to traditional face-to-face encounters of health care professionals (HCPs) with patients [4]. The use of technology allows new opportunities to be exploited in a wide range of medical fields [5]. Among researchers, the added value, such as cost efficiency, is widely agreed upon. As an example, the percentage of global gross domestic product accounted for by health care costs is around 10% and rising [6]. Correspondingly, studies have shown that low-cost improvements in patients' health may be achieved by using remote health services for consultation, evaluation, or treatment [7,8]. Furthermore, the implementation of telemedicine allows for increased accessibility to health care. Thus, by eliminating the proximity factor, patients can access basic and particularly specialist care from their home, which allows instant assessment or treatment [7,9-11]. The relevance of the accessibility gained via telemedicine is highlighted in diverse respects. Especially in rural areas, a shortage of health care providers represents a challenge that can be bridged by telemedicine [7]. Considering the disparity in health care among nations, the increase in telemedicine adoption has the potential to provide great benefits not only to patients but also to medical staff worldwide [10].

Despite the multitude of presented advantages, the adoption rate of telemedicine remains low [2]. Regarding diffusion, the recent COVID-19 pandemic had a strong impact on the adoption rate of telemedicine. The emergence of COVID-19 in 2020 and the restrictions that were enacted resulted in far-reaching challenges worldwide [12,13]. Telemedicine provided a viable solution to many of those challenges by avoiding direct contact, reducing the danger of transmission, and ultimately ensuring continuous care [5,14]. Consequently, the circumstances enhanced the adoption of telemedicine, inducing a rapid upscale of diffusion in 2020. However, many HCPs, especially in countries lagging in technology, remain suspicious, resulting in low telemedicine adoption rates [15]. For example, in countries such as Germany (industrialized economy), India (resource-limited economy), and Brazil (emerging economy), telemedicine has been slower to establish itself than in other countries such as Israel (industrialized economy) [16]. These countries are examples of countries lagging in technology in telemedicine adoption due to barriers such as regulatory hurdles, insufficient digital infrastructure, skepticism among HCPs, and limited access to technology in rural areas, despite the potential benefits [16,17].

Extensive research has been conducted on relevant factors for telemedicine adoption [5,18,19]. In this regard, trust has been repeatedly highlighted as a key determinant of acceptance and use of telemedicine services [20,21]. Relatedly, perceived usefulness and ease of use have also been found to influence telemedicine adoption substantially [22]. The adoption of technologies in health care brings uncertainties, particularly related to data security and privacy [23,24]. Such uncertainties often lead to perceived risks that pose significant barriers to adoption intentions [25,26]. Factors such as the HCP-patient relationship, cultural aspects, and the digital divide also serve a critical role in the overall adoption of telemedicine [27-29]. Taken together, the literature on trust, perceived risk, perceived benefits, and other factors related to telemedicine adoption has been researched in a variety of settings and from differing perspectives. Kuen et al [30], for example, highlighted the multidimensionality of risks and trust transfer effects between several trust referents in a telemedicine context from the patient's perspective. However, it is still to be determined to what extent these constructs are also relevant from an HCP's perspective.

The information systems (IS) literature perspective is particularly relevant for telemedicine adoption as it explores how technology is effectively implemented, managed, and accepted within organizations. Because telemedicine heavily relies on digital systems and interactions, IS literature provides valuable insights into user acceptance and integration of these technologies in health care. Indeed, several systematic literature reviews have shown that technology adoption models such as the technology acceptance model (TAM) and the unified theory of acceptance and use of technology (UTAUT) are the dominant theoretical frameworks used to study telemedicine adoption among HCPs [31,32]. While these models offer important insights into factors such as perceived usefulness and ease of use, they are limited in their ability to account for the full complexity of telemedicine adoption in health care settings. Specifically, they fail to adequately address the critical role of perceived risks, which are especially relevant due to the sensitive nature of health care services [31]. Moreover, these models often assume that the decision-making process is primarily driven by functional considerations, such as ease of integration and productivity improvements [33]. However, HCPs operate in a more complex environment, where ethical concerns, patient safety, and trust in the technology and technology provider play significant roles [30,34]. Traditional models such as TAM and UTAUT also overlook the emotional and psychological dimensions of decision-making, including anxiety about new technology or concerns about the erosion of the HCP-patient relationship [35]. In addition, these models are generally ill-suited to capture the intricate professional and legal responsibilities faced by HCPs, which can directly influence their adoption behavior [35]. In the health care context, the stakes are higher, and the consequences of technology adoption

XSL•FO
RenderX

can directly impact patient outcomes, creating a level of risk that these models fail to fully incorporate [36].

Therefore, these frameworks do not offer the depth needed to explore the multilayered, context-specific factors that influence telemedicine adoption among HCPs. For example, while Bakshi and Tandon [37] explored how various dimensions of risk—such as financial, social, technological, and privacy risks—affect physicians' intentions to adopt telemedicine, their work does not fully contextualize these risks or examine the factors that drive these perceptions. Furthermore, some studies that identify key adoption factors in telemedicine do not use specific theoretical frameworks or examine the relationships between these factors [38]. This creates a gap in the literature, as the existing models and frameworks (eg, TAM and UTAUT) do not sufficiently incorporate the multidimensional risks that are crucial in the health care context [31].

This gap underscores the need for a more comprehensive framework. While technology adoption models provide one perspective, they do not account for the broader factors that influence decision-making in health care. As Hong et al [39] argue, theoretical models must be adapted to specific contexts to provide meaningful insights. In health care, the adoption of telemedicine goes beyond technical functionality and is shaped by professional ethics, regulatory environments, and context-specific risks. Therefore, a framework that integrates these elements is needed to better explain telemedicine adoption from an HCP's perspective. Therefore, we pose the following research question: "Which factors are relevant in the context of telemedicine adoption from an HCP's perspective, and how can these be incorporated into an overarching framework?"

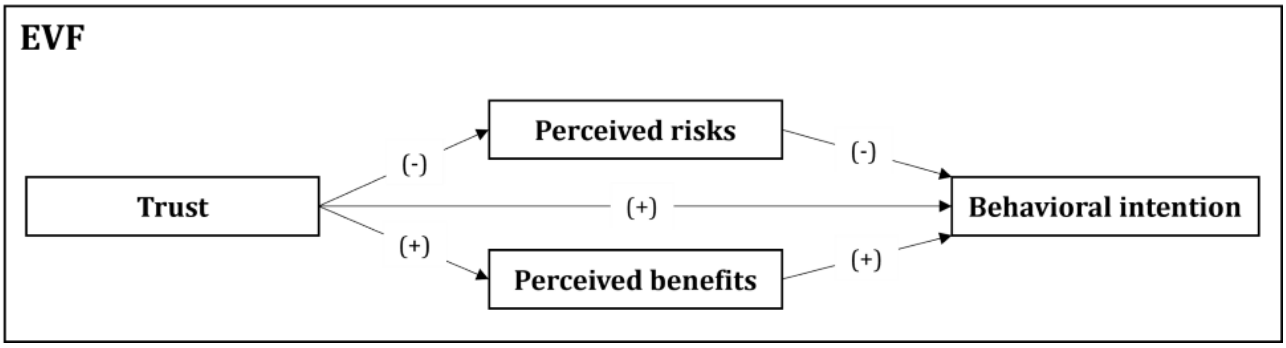
Objectives

The research goal of this paper is to identify the factors influencing the adoption of telemedicine in terms of perceived risk, perceived benefits, and trust from an HCP's perspective. The aim is to derive a contextualized framework for the telemedicine context, which can serve as a guideline for future research. We draw on the extended valence framework (EVF) by Kim et al [40], which was initially developed for the consumer context but has demonstrated applicability to technology adoption in various domains (eg, mobile payment solutions [41] and remote anti-doping testing [42]), including consumer acceptance of online health information services [43]. We adapt it to the context-specific characteristics of telemedicine, which include the exchange of sensitive and personal data, situations with high uncertainty and dependency for patients, the significant responsibility placed on HCPs, and the adoption of new technology.

Theoretical Lens

To answer the research question, we draw on previous literature on adoption behavior and provide a brief introduction to the relevant definitions of the related constructs. We argue that the EVF is well suited as an overarching framework to explore relevant factors for technology adoption as it provides a broad perspective on individuals' attitudes and perceptions toward a new technology [40]. Unlike traditional models such as the TAM and the UTAUT, which primarily emphasize positive factors such as perceived usefulness, the EVF integrates both positive drivers (eg, perceived benefits) and negative drivers (eg, perceived risks), offering a more balanced understanding of adoption decisions. This is particularly critical in health care, where concerns over patient safety, data security, and ethical implications play a central role in decision-making. The EVF integrates perceived risks, perceived benefits, trust, and behavioral intentions, all of which are critical to the adoption of new technologies [44]. Figure 1 [40] depicts the layout of the EVF in which trust positively affects perceived benefits and behavioral intentions but negatively affects perceived risks. In addition, perceived risks have a negative relationship, and perceived benefits have a positive relationship with behavioral intentions.

Figure 1. Extended valence framework (EVF) for telemedicine adoption (this depiction of the EVF is based on the study by Kim et al [40]).



In the context of telemedicine, the EVF is particularly appropriate because it not only captures the benefits of new technology but also addresses the risks associated with it, such as misdiagnoses, technical reliability, and patient confidentiality, which are often underexplored in models such as TAM and UTAUT. Perceived risk refers to the uncertainty and potential severity of negative consequences, a definition established by Dowling and Staelin [45] as well as Featherman and Pavlou [26]. Various dimensions of risk have been identified in the literature, underscoring its multifaceted nature [46]. While the risk dimensions used in this study were initially developed in consumer e-services and purchasing contexts [26,47,48], research has extended their applicability to health care and telemedicine [30]. Table 1 provides an overview of the risks identified by previous research and their respective definitions.

XSL•FO
RenderX

Table 1. Definitions of risk dimensions relevant to telemedicine adoption.

Risk dimension	Definition	Study
Performance risk	Reflects the perception that a product or service fails to meet the anticipated performance standards, consequently not delivering the desired benefits	Grewal et al [47], 1994
Privacy risk	Reflects the "potential loss of control over personal information"	Featherman and Pavlou [26], 2003
Financial risk	Reflects the potential monetary outlay due to the product's or service's costs	Featherman and Pavlou [26], 2003
Psychological risk	Reflects the mental stress that possibly arises through the use of technology	Jacoby and Kaplan [48], 1972, and Featherman and Pavlou [26], 2003
Time risk	Reflects the potential loss of time when making a bad purchasing decision by wasting time researching and making the purchase, learning how to use a technology or service only to must replace it if it does not perform to expectations	Featherman and Pavlou [26], 2003
Social risk	Reflects the potential loss of status in one's social group because of adopting a product or service, looking foolish or untrendy	Featherman and Pavlou [26], 2003

Contrary to the multidimensional definition of perceived risk, perceived benefit represents the extent to which an individual believes that they will benefit from using a product or service [23]. It is characterized by the user's motivation to maximize the expected value associated with a technology [40]. Therefore, in this study, perceived benefit is defined as the potential value that a user associates with the use of telemedicine. As telemedicine represents a context that is different from other topics addressed in IS research, commonly used constructs cannot be easily transferred [49]. Previous research on telemedicine has shown, for example, the impact of efficiency on perceived benefits in health contexts such as autism and diabetes [7,8], as well as the relevance of the perceived quality of the specific service or product under consideration [50-52]. A study on the patient's perspective toward telemedicine adoption additionally highlights convenience as an impacting factor [53].

Within the EVF, trust is integrated as another relevant variable, as studies have shown that it reduces risk perceptions and can increase perceived benefits [54-56]. Thus, we include it in the scope of our contextual decomposition approach and define trust according to Mayer et al [57] as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the truster, irrespective of the ability to monitor or control that other party." Within this definition, a relationship is considered to exist with another identifiable party that acts and reacts with volition toward the truster. Both parties—the truster and the trustee—are not limited to individuals; they can also include other entities, both physical and non-physical, such as technology [58]. Interpersonal trust is considered to be a trusting relationship between individuals. According to Mayer et al [57], individuals' trust is determined by their perception of the trustworthiness of the other party. Trustworthiness is considered a multidimensional construct, the dimensions (ability, benevolence, and integrity) that influence trust in another individual, groups of individuals, or an organization and lead to a higher willingness to take risks in the respective relationship [57]. In addition to interpersonal trust, the need to distinguish trust in technology emerged. While trust in an individual entails a moral and conscious interpersonal dynamic, technology is an artificial creation with limited capabilities and no moral agency. Therefore, further research on trust developed a multidimensional construct of trusting beliefs in technology [55,59]. Table 2 shows a compromised overview of the relevant dimension definitions.

As this study aims to decompose the EVF, we look at several potential and context-specific trust referents. In research on trust in telemedicine, 4 main stakeholders have been identified (technology, treatment, patient, and provider) [60]. The concept of trustworthiness and trusting beliefs can be applied to patients and providers (trustworthiness) as well as technology (trusting beliefs). As treatment is understood to be a process, it does not fall under the 2 trust constructs. However, trust in treatment can be characterized as an individual's belief that the telemedical treatment effectively addresses medical concerns, is clear, and results from a collaborative decision-making process between the patient and HCP [30,61].

XSL•FO
RenderX

Table 2. Definitions of trust dimensions for interpersonal and technological contexts.

Trust construct and dimension	Definition	Study
Interpersonal trust/trustworthiness		Mayer et al [57], 1995
Ability	The skills, competencies, and attributes that enable a party to have an impact in a specific domain.	
Benevolence	The degree to which a trustee intends goodwill toward the truster, alongside their perceived positive attitude toward the truster.	
Integrity	The truster's perception that the trustee adheres to a set of principles that the truster finds acceptable.	
Trust in technology/trusting beliefs		McKnight et al [59], 2011
Functionality	The belief that a specific technology has the appropriate functionalities or features to fulfill the requirement.	
Reliability	The belief that the specific technology will consistently operate properly.	
Helpfulness	The belief that adequate and responsive help is provided by the technology.	

Methods

Study Design

To determine how perceived risks, perceived benefits, and trust are related in the specific context of telemedicine, we referred to the guidelines by Hong et al [39] on context-specific theorizing. Accordingly, we chose a qualitative research approach and conducted semistructured interviews with HCPs who already offer video consultations to their patients, which is considered a widely applied method to obtain contextual and authentic accounts [62]. HCPs represent the primary party in adopting telemedicine, as they must be willing to offer telemedicine to enable patients to actively opt for it. Hence, they have the opportunity to shape the broad adoption of telemedicine [63]. Given the limited adoption of this technology, interviewing HCPs with extensive experience in telemedicine is expected to provide greater value, as their use and familiarity with it are essential for overall social acceptance.

As part of the study's rigorous design, a pilot study was conducted before the actual data collection to ensure the clarity and relevance of the interview guide [64]. This pilot phase involved conducting preliminary interviews with 4 HCPs who had prior experience with telemedicine. On the basis of this feedback, the interview guide only had to be marginally adapted. The interview guide was mainly guided by the EVF (Figure 1), covering key constructs such as trust, perceived risks, perceived benefits, use behavior, and intention (Table S1 in Multimedia Appendix 1). It also included contextual factors to capture additional insights that may not be covered by the EVF. The guide consisted of open-ended questions designed to explore how these factors influence telemedicine adoption from the HCP's perspective. This structure allowed for flexibility, enabling follow-up questions to be asked during the interviews as needed to explore emerging themes.

On the basis of the EVF, a combination of deductively derived and inductively elaborated relevant constructs was chosen as the research approach [65]. Thus, in line with the study by Hong et al [39], the various constructs can be decomposed, the appropriateness of the EVF for the specific context can be examined, and all relevant constructs can then be transferred into a context-specific variation of the EVF. In addition, newly emerging antecedents and context-specific variables are added.

Data Collection

To ensure the interview partners were sufficiently qualified, only HCPs who (1) have completed formal education in the medical or therapeutic field, (2) have experience with telemedicine through video consultations, and (3) practice in Germany (to ensure cultural comparability) were contacted. All medical specializations were included. As psychotherapists represent about 25% of HCPs in Germany, this group was included in our analysis [66]. This selection deliberately focused on gaining insights from HCPs who have practical experience with telemedicine, enabling a deeper exploration of its real-world effectiveness and challenges. Involving only experienced users ensures that the study captures practical insights that are directly relevant to improving the future implementation of telemedicine. Obtaining consent from interviewees and assuring anonymization, we audio-recorded the interviews and then transcribed them verbatim [67,68]. We chose to recruit via the medical website Jameda [69]. A total of 142 HCPs who met the requirements as experts were contacted by email. Out of these 142 professionals, 26 (18.3%) expressed interest in participating. The interviews were conducted online via Zoom (Zoom Video Communications) during March and April 2023. After 14 interviews, theoretical saturation was reached, suggesting that further interviews would be unlikely to add any new insights [70].

Sample Description

The resulting interview sample included 14 experts, half of whom were male (7/14, 50%) or female (7/14, 50%) and were, on average, aged 47 (range 34-68) years. The length of the interviews varied between 13 and 53 (mean 27, SD 10) minutes. Table 3 gives an overview of the relevant descriptives.

XSL•FO
RenderX

Table 3. Characteristics of the 14 health care professionals interviewed.

Number	Age (y)	Highest education	Field of practice
1	43	Doctoral degree	Palliative care
2	49	High school diploma	Psychotherapy
3	55	High school diploma	Psychotherapy
4	48	Doctoral degree	Gynecology
5	68	Master's degree	Psychotherapy
6	45	Doctoral degree	Psychiatry and psychotherapy
7	52	Doctoral degree	General medicine
8	50	Doctoral degree	Endocrinology
9	36	Bachelor's degree	Physiotherapy
10	42	Doctoral degree	Gynecology
11	68	Doctoral degree	General medicine
12	37	State examination	Psychotherapy
13	35	Doctoral degree	General medicine
14	34	State examination	General medicine

Data Analysis

We conducted a qualitative content analysis [65] using MAXQDA 2022 (version 22.7.0) software. The analysis was conducted by 2 authors (coder A and coder B) separately and using deductive as well as inductive coding methods. The deductive codes were derived from the relevant constructs already outlined in the theoretical lens based on the EVF. First, all transcripts were coded by coder A according to these categories. In addition, coder A inductively derived subcategories to allow for more in-depth explanations and context-specific extension, which were then discussed with coder B. Following coding, coding rules, definitions, and anchor examples were developed to ensure consistency for each category. Coder B performed the coding process independently following coder A. Subsequently, the coding was iteratively extended and analyzed in several steps. Transparency, for example, was identified as an antecedent to trust and coded following the dimensions defined by Schnackenberg and Tomlinson [71] (disclosure, clarity, and accuracy). We adhered to the standards set by Venkatesh et al [72] to ensure the quality of our qualitative research. The validity of our research design was confirmed through a thorough account of our research context, which increased the relevance and trustworthiness of our conclusions. To ensure a valid coding process, we took steps to ensure inferential validity. The interauthor agreement in the coding process resulted in an intercoder reliability of 86.14%. To account for the potential of random agreement at the segment level, we assessed coding reliability and calculated Cohen κ, which yielded a value of 0.86. These results suggest an "almost perfect" agreement between the 2 coders [73].

Ethical Considerations

The study received ethics approval from the ethics committee of the School of Business and Economics at the University of Münster (2020–02). The research adhered to ethical guidelines for human subject research. All participants provided written informed consent before participating in the study. The consent process included an explanation of the study objectives, procedures, and the right to withdraw at any time without any consequences. To protect participants' privacy, all interview data were anonymized during transcription. Identifiable information was removed, and participants were assigned unique codes to ensure confidentiality. Data were securely stored on encrypted devices accessible only to the research team. No monetary or material compensation was provided to participants for their involvement in the study. This study does not include any images of participants or supplementary material that could potentially identify individuals. As a precaution, the interview transcripts were carefully reviewed to exclude any inadvertent references that could lead to participant identification.

Results

Outcome Evaluation

By decomposing the constructs of the EVF, our results demonstrated their individual sublevels and relevance to the context of telemedicine. We found that HCPs clearly perceived both the risks and benefits of telemedicine. They were able to reflect on and consider these extensively. In addition to the decomposed constructs of the EVF, we identified context-specific variables; transparency was an inductively derived factor that preceded trust. The following sections presents our result according to each construct of the EVF. Although certain inductive codes were mentioned infrequently, their inclusion is justified by the nature of qualitative research. Qualitative analysis values the depth of insights over frequency, meaning even low-frequency mentions can offer significant contextual relevance. These rare codes can highlight emerging or underexplored challenges within telemedicine adoption, particularly in niche areas of health care practice [74]. Specific quote examples for the inductively generated level 3 codings can be found in Tables S2-S6 in Multimedia Appendix 1.

XSL•FO
RenderX

Perceived Risks

Limited treatment options due to visual restrictions were mentioned with regard to *performance risk*; that is, no holistic insight into the patient can be gained. Furthermore, physical contact is omitted, and further examinations for diagnosis might be neglected. This results in the "...danger...that something is overlooked in the diagnosis" (interviewee 11). Internet connection issues were additionally noted, especially concerning patients in rural areas. An interviewee expressed the following:

> *...the great issue is the poor internet connection of the clients most of the time....* [Interviewee 4]

In addition, *financial risks* were recognized, particularly associated with acquiring high-technology equipment. While the general cost of equipment was not seen as a risk, the purchase of advanced, high-technology tools was perceived differently. An interviewee stated the following:

> *You could also control it remotely and make a much better zoom than with the webcam. Such a tool is expensive, of course.* [Interviewee 6]

Concerning *time risks*, the investment of time required for staff to familiarize themselves with technology and its functions was underlined. One respondent mentioned that "[Telemedicine] is indeed a considerable investment of time, as you have to train the staff as a whole" (interviewee 10), pointing out the significant time commitment necessary for staff training. As far as *psychological risk* is concerned, it was rarely addressed and, if so, was associated with an uncomfortable feeling and a certain weight in digital communication. Concerning *social risks*, it was inferred that negative comments are mainly received from the social and professional environment. However, *privacy risk* was most frequently addressed by the interviewees. The main focus was on the perceived risk posed by patient data security:

> *The basic question is always that of data security since patient data is exchanged.* [Interviewee 1]

Table 4 shows the deductive and inductive codes for perceived risks according to the number of interviews in which these were addressed.

XSL•FO
RenderX

Table 4. Results of the qualitative content analysis for perceived risks (N=14).

Levels 1[a], 2[b], and 3[c]	Interviews in which this aspect was mentioned, n (%)
Perceived risk	
Performance risk	
Restriction in treatment [d]	7 (50)
Restriction in patient-HCP [e] *connection*	7 (50)
Internet connection issues	5 (36)
Technical issues	4 (29)
Technical proficiency of patient	5 (36)
Financial risk	
Need for equipment	4 (29)
Need for staff	1 (7)
Insurance coverage	3 (21)
Limited user group	2 (14)
Inadequate compensation	1 (7)
Time risk	
Familiarization with the technology	7 (50)
Adaptation to the new workflow	2 (14)
Research for alternatives	1 (7)
Psychological risk	
Uncomfortable feeling	3 (21)
Weight of communication	3 (21)
External interfering factors	2 (14)
Social risk	
Negative comments	5 (36)
Privacy risk	
Data security	10 (71)
HCPs' privacy	2 (14)
Patients' privacy	4 (29)

[a]Category derived from the study by Kim et al [40].

[b]Categories derived from the study by Featherman and Pavlou [26].

[c]Categories are inductive codes based on specifications from respondents.

[d]Inductive codes are italicized.

[e]HCP: health care professional.

Perceived Benefits

Interviewees expressed multiple benefits related to the *convenience of telemedicine*. A notable advantage was the reduction in travel time, which benefited HCPs by allowing them to manage their schedules more efficiently and handle a higher volume of consultations. One participant stated that video consultation had the following benefits: "No more traveling. Rehabilitation clinics, in particular, are always a little further out of town as far as our field is concerned." (interviewee 9). Moreover, convenience was accentuated in treating patients located in rural areas. Another interviewee emphasized that these patients particularly value telemedicine as it allows them to receive consultations at home. In addition, interviewees commented on the medical implications of reduced travel, mentioning that longer travel durations "may also not be harmless due to the illness" (interviewee 6). They suggested that maintaining distance, especially if patients are contagious, is sensible for both patient and HCP. Apart from convenience, there was a shared sentiment that telemedicine augments the *quality of care*. Accessibility was often spotlighted as a distinct advantage of telemedicine, with interviewees noting its significant role in improving patient care. In addition to the accessibility of medical treatment and consultation, the possibility of maintaining a safe space was mentioned several

XSL•FO
RenderX

times, particularly in the psychological field, as many patients might feel more comfortable at home:

> *...the more distance there is between medical staff and patients,...it gives patients a low-barrier opportunity to mention things that they might not dare to mention face-to-face.* [Interviewee 9]

Consequently, HCPs obtained more relevant information, thus ensuring a higher quality of diagnosis and treatment. Furthermore, telemedicine was observed to enhance HCP's *efficiency*. The emphasis was largely on the significant time savings for HCPs, allowing them to provide timely care while easily coordinating and managing their schedules through health platforms. This streamlined process enables HCPs to handle more appointments and optimize their workflow, ultimately improving their productivity. Table 5 shows the deductive and inductive codes for perceived benefits according to the number of interviews in which these were addressed.

Table 5. Results of the qualitative content analysis for perceived benefits (N=14).

Levels 1[a], 2[b], and 3[c]	Interviews in which this aspect was mentioned, n (%)
Perceived benefit	
Convenience of telemedical use [d]	
Space independence	8 (57)
Reduction of travel time	5 (36)
Time flexibility	6 (43)
Protection of HCPs' [e] *well-being*	4 (29)
Protection of patients' well-being	6 (43)
Quality of care	
Accessibility	8 (57)
Safe space for sensitive topics	4 (29)
Enhanced perception	3 (21)
Efficiency of telemedical service	
Time saving	7 (50)
More precise working time	2 (14)
Financial advantage	4 (29)

[a]Category derived from the study by Kim et al [40].

[b]Categories derived from the study by Featherman and Pavlou [26].

[c]Categories are inductive codes based on specifications from respondents.

[d]Inductive codes are italicized.

[e]HCP: health care professional.

Trust Referents

Trust referents were distinguished by the HCPs, and all were acknowledged, although not all were considered to be of high importance. Trust in technology was typically mentioned in terms of its functionality, such as the ability to provide audio and visual connection, schedule appointments, and even collect some anamnesis data; for example,

> *[Take] a quick picture of the throat with a smartphone so that you can get an idea. Are the tonsils swollen now? Are they coated? That works.* [Interviewee 13]

With regard to *trust in the treatment process*, 3 subcategories emerged, which can be subdivided into the initial necessary consensus, the subsequent effective collaboration of patient and HCP, and the final effectiveness of the treatment. The patient-HCP connection was particularly relevant for HCPs:

> *Definitely, if you manage to maintain the personal factor further [using video consultation].* [Interviewee 9]

> *Trust in the provider was supported above all by the data security it offers, which also corresponds to the high perceived risk of data security:*

> *I trust my specialist provider, so to speak, and assume that the solution he offers has also been checked with the data protection authorities in terms of data protection law.* [Interviewee 2]

It also became clear that the provider is expected to have both technical competence and knowledge of medical procedures, that is, an understanding of the particular application context. Compared to the other trust referents, the HCPs' perception of *trust in the patient* was rather limited. It was primarily based on the digital affinity assigned to the patient, which is relevant for the HCP to be able to use telemedicine effectively together with the patient. Table 6 presents the deductive and inductive codes for all trust referents.

XSL•FO
RenderX

Table 6. Results of the qualitative content analysis for trust referents (N=14).

Levels 1[a], 2[b], and 3[c]	Interviews in which this aspect was mentioned, n (%)
Technology as a trust referent	
Technology's functionality	
Audio connection [d]	5 (36)
Visual connection	6 (43)
Appointment scheduling	1 (7)
Anamnesis data taken	3 (21)
Technology's reliability	
Stable connection	5 (36)
Stable platform	4 (29)
Treatment as a trust referent	
Consensus understanding of the treatment process	
Practicality	4 (29)
Enabling adequate dialogue	4 (29)
Effectiveness of collaboration	
Consent of patient	5 (36)
Enabling compliance	4 (29)
HCP [e]*-patient connection*	9 (64)
Patient feedback	2 (14)
Effectiveness of treatment	
Provision of medical reports	3 (21)
Work facilitation	4 (29)
Possible guidance to self-examination	2 (14)
Provider as a trust referent	
Provider ability	
Contextual knowledge	2 (14)
Technical knowledge	2 (14)
Provider benevolence	
Proximity	1 (7)
Provider integrity	
Data security	11 (79)
Responsiveness	3 (21)
Certification	3 (21)
Patient as a trust referent	
Patient ability	
Digital affinity	5 (36)

[a]Category derived from the studies by Van Velsen et al [60] and Kim et al [40].

[b]Categories derived from the studies by Mayer et al [57] and McKnight et al [59].

[c]Categories as inductive codes based on respondents' specifications.

[d]Inductive codes are italicized.

[e]HCP: health care professional.

XSL•FO
RenderX

Transparency

In the coding process, transparency emerged as an important antecedent to several trust referents for HCPs. This is in line with previous literature that identified transparency as an antecedent to trust, as information shared in this way indicates the ability, benevolence, and integrity of a truster [75,76]. The definition by Schnackenberg and Tomlinson [71] is widely accepted and defines transparency as "the perceived quality of intentionally shared information from a sender." We followed the accepted multidimensional approach and accordingly subdivided transparency into its 3 dimensions: disclosure, clarity, and accuracy [71,76,77]. Thereby, we define "disclosure" as the belief that all relevant information is shared in a timely manner, "clarity" as the level of correspondence between the intended and the understood meaning of information, and "accuracy" as the extent to which information given matches reality and is free from intentional distortions [71,76].

Transparency was emphasized in the information provided about technology, treatment, and providers. The patient as a trust referent was not mentioned when it came to transparency. In addition to transparency in relation to the provider, the relevance for disclosure on technology also became apparent: "Now I'm not a technician and can't assess what it could do with this data, but I feel well informed." (interviewee 3) Similarly, the importance of transparency in the treatment process between the patient and the HCP was highlighted: "When you feel like you've had a lot explained to you and you have a lot of knowledge about it and you know how things work" (interviewee 9). Table 7 shows the deductive codes for the transparency recipients according to the number of interviews in which these were addressed for each transparency dimension.

Table 7. Results of the qualitative content analysis for transparency dimensions (N=14).

Levels 1, 2[a], and3[b]	Interviews in which this aspect was mentioned, n (%)
Transparency	
Transparency on technology	
Disclosure	2 (14)
Clarity	1 (7)
Accuracy	3 (21)
Transparency on treatment	
Disclosure	4 (29)
Clarity	2 (14)
Accuracy	3 (21)
Transparency on provider	
Disclosure	7 (50)
Clarity	7 (50)
Accuracy	5 (36)

[a]Categories are in accordance to the trust referents based on the studies by Van Velsen et al [60] and Kuen et al [30].

[b]Categories derived from the study by Schnackenberg and Tomlinson [71].

Context-Specific Variables

Regarding the need for contextual specification mentioned by Hong et al [39], we were able to identify variables that apply specifically to the telemedicine context. Especially, the relevance of prior consideration of the symptom characteristics and their suitability for treatment with telemedicine services, such as video consultation, was highlighted:

> *I can take a lot of responsibility for doing this via telemedicine or not....And I have a bad feeling when I hear patients report that...the patient is being sealed off even more.* [Interviewee 12]

It is not so much the distinction between mental and physical that plays a role here, but rather the necessary examination proximity and physical treatment approach:

> *What is more difficult is to examine you....Let's say one would have to palpate it or something else, then it stops.* [Interviewee 1]

Another context-specific variable is the previous experience or familiarization of the HCP and their patients with telemedicine:

> *[Video consultation] has also now worked out quite well over the last two years.* [Interviewee 8]

Frequent positive experiences, therefore, increase both the willingness to adopt and the level of trust in the technology. Table 8 shows the inductively derived codes for context-specific variables.

XSL•FO
RenderX

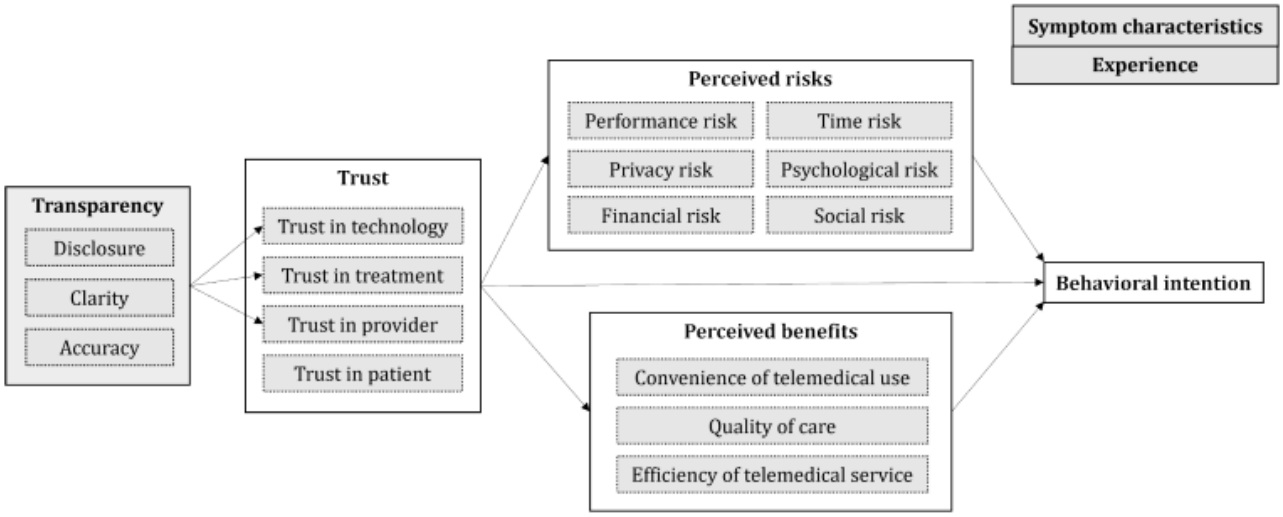
Table 8. Results of the qualitative content analysis for context-specific variables affecting telemedicine adoption.

Levels 1 and 2[a]	Interviews in which this aspect was mentioned, n (%)
Context variable	
Symptom characteristics (10)	10 (71)
Experience (11)	11 (79)

[a]Categories derived inductively.

In summary, our proposed expansion of the EVF is shown in Figure 2 [26,30,40,60,71]. We consider transparency a key multidimensional antecedent, along with context-specific variables such as symptom characteristics and experience with telemedicine services.

Figure 2. Telemedicine-contextualized valence framework. The extended valence framework is based on the study by Kim et al [40]. Perceived risk categories follow the study by Featherman and Pavlou [26]. Trust referents are derived from the studies by Van Velsen et al [61] and Kuen et al [30], and transparency dimensions are based on the study by Schnackenberg and Tomlinson [71].



Discussion

Interpretation of Results

Our study provides a multifaceted examination of the considerations HCPs have when adopting and using telemedicine services, revealing various facets of perceived risks and benefits, trust referents, transparency, and context-specific variables.

First, it is evident that HCPs have various concerns about telemedicine. Concerns primarily relate to performance risks such as visibility limitations that restrict holistic insight into the patient's well-being. Lack of physical contact and potential confusion in diagnosis were highlighted by the HCPs, pointing to the critical role of traditional face-to-face interaction in medical assessments. Financial risks, such as the time required to learn to use technology, and privacy risks, such as concerns about data security, emphasize the challenges that HCPs face. Previous studies have identified technological issues, data security, and privacy as primary risks associated with the telemedicine context [30,78]. Consistent with earlier research, this study reaffirms the existence and multidimensionality of perceived risk [26,30], further delineating it into subdimensions of performance risks (eg, visual restrictions), privacy risks (eg, data security), and financial risks (eg, uncertainty about insurance coverage) as the risks most often mentioned. The addition of these context-specific subdimensions is a nuanced insight, adding depth to the understanding of practical challenges faced by HCPs.

Second, among the perceived benefits, the convenience of the telemedical approach, quality of care, and efficiency of the telemedical service were identified as key factors. Particular emphasis was placed on minimizing the HCP's travel time and improving the protection of the patient's well-being by reducing the need for transportation. This convenience, along with increased comfort and protection through minimal physical contact, establishes a better experience and better safety using telemedicine. In addition, our findings indicate that telemedicine can foster an environment that leads to increased efficiency of the services the HCP offers. Telemedicine was often complimented for its convenience, accessibility, and potential to deliver health care services to remote or underserved areas [7]. Increased patient engagement and satisfaction were also commonly cited benefits [10]. Our results align with existing literature, emphasizing convenience, quality, and efficiency as significant benefits and adding context-specific dimensions by highlighting the potential for streamlined operations and enhanced HCP efficiency in subdimensions.

Third, we found that trust is a key driver of adoption behavior, with multiple trust referents such as technology, treatment process, provider, and patient. Our findings illustrate that while technology and treatment process were strongly and diversely

XSL•FO
RenderX

perceived trust referents, trust in patients was solely measured by the patient's abilities in terms of their digital affinity. This underscores the important role that technical reliability and effective treatment collaboration have in successful telemedicine practices. These findings corroborate prior research, signifying trust as a multifaceted concept involving technology, treatment processes, and patients [30,60]. However, our results contradict the relevance of trust in patients from an HCP's perspective, as they represent a given parameter in the general treatment process for the HCP, and changing to an online setting seems to adapt the relationship only to the extent that patients must be able to use the new technology in question. Furthermore, the subdivisions of the treatment process need to be considered. Therefore, relevancy for trust is seen in particular in the second part, the consensus of collaboration. This is consistent with the assumption that the highest level of interaction and reliance on functioning communication between patient and HCP occurs here. Therefore, this study offers a more nuanced perspective by identifying the different relevancy of trust referents from the HCP's perspective.

Fourth, transparency emerged as an antecedent for trust, indicating the need for disclosure, clarity, and accuracy in technological processes and interactions with providers to foster a trustworthy telemedicine environment. This underscores the critical role of clear communication and information sharing. Transparency, specifically regarding data use and security, has been identified as crucial in previous research in differing contexts [79-81]. In addition, transparency is often linked to increased trust and acceptance of telemedicine services [60]. Notably, the need for transparency was identified for all trust referents except the patient. This could be explained by the personal view of the HCPs, in that they do not feel any dependency on their patients, as they perceive themselves as service providers on whom the patients depend. This highlights the special nature of the medical sector and the attitude and unilateral dependent relationship between HCPs and patients. Our results align with prior research, highlighting the importance of transparency as an antecedent for certain trust referents. This underscores the need to tailor the information provided according to the HCPs' relevant trust referents when fostering a trustworthy telemedicine environment. We also show that the view of transparency as multidimensional and the 3 defined dimensions (disclosure, clarity, and accuracy) [71] are transferable to the context under consideration and that, with a qualitative research approach, no major deviations in the relevancy of the individual dimensions are apparent.

Fifth, context-specific variables such as symptom characteristics and prior experience with telemedicine have a significant role in its use and efficacy. An HCP's intention to use telemedicine is influenced by various contextual factors such as the characteristics of symptoms, patient demographics, and technological familiarity among users [82-84]. The results from our interview study resonated with these findings, emphasizing the impact of symptom characteristics and previous telemedicine experiences on its use and perceived effectiveness. In addition, adoption behaviors may vary across medical specializations, depending on how telemedicine software interacts with clinical workflows and the patient-practitioner dynamic. For example, therapy with clinical psychologists might involve a more personal and sensitive experience compared to discussions of physical symptoms with HCPs, which may not require the same level of intimacy. The distinct benefits telemedicine can offer in fields such as infectious diseases further highlight the need to consider how different clinical contexts impact telemedicine adoption. Thereby, we specify which type of symptom characteristic is decisive for the usability of telemedical services such as video consultation. We abandon a division into mental and physical symptoms, an approach taken in prior research [30], as we found no significant difference between mental and physical symptoms. Instead, distinctions should be made according to the severity of the need for direct intervention. This is similar to a distinction of symptom severity, which has also already been introduced [85], but should not be interpreted as synonymous with it. The demonstrated impact of repetition, leading to increased experience and familiarity, is consistent with other studies, such as the study by Venkatesh et al [33], which highlight the importance of the repetition cycle in building trust [86].

Implication for Research

Several theoretical implications can be derived from the results of our study. Arguably, the first and most straightforward is the applicability and relevance of the guidelines by Hong et al [39] for context-specific theorizing. Our results show that a systematic decomposition of the relevant constructs of adoption behavior should address context-specific characteristics, especially because the context of telemedicine is characterized by a special relationship among the patient, HCP, and technology used for the treatment process and its provider.

Our second theoretical contribution is that this study addresses the limitations of existing technology adoption models such as TAM and UTAUT by incorporating telemedicine-specific factors into a contextualized framework. While TAM and UTAUT focus primarily on general constructs such as perceived usefulness and ease of use, they often fail to capture critical dimensions relevant to health care, such as the multifaceted risks and trust referents that are crucial for HCPs. The developed contextualized model overcomes these limitations by integrating additional factors such as transparency and symptom characteristics, offering a more comprehensive understanding of telemedicine adoption from an HCP perspective. This framework provides a valuable additional perspective to existing technology adoption models, serving as a foundation for future research on telemedicine adoption. Moving forward, future research can build on this by conducting quantitative validation and refining the framework further.

Third, our study emphasizes the multidimensionality of relevant factors in telemedicine, such as risks [26,30] and trust [59,60]. Recognizing the multidimensional nature of these factors enables a more nuanced understanding and specific contextualization for telemedicine. This perspective can guide future research to explore these dimensions in different health care settings and with different stakeholders.

Fourth, the developed framework offers starting points for further analysis of stakeholders in health care. For example, the perspectives of patients or technology providers, who have

XSL•FO
RenderX

already been identified by Van Velsen et al [60] as trust referents in the context of telemedicine adoption, may be examined in future studies. Understanding their perspectives and relevant context-specific factors can enrich the theoretical understanding of telemedicine adoption and its broader implications in health care.

Practical Implications

Several practical implications for HCPs and their use of telemedicine can be derived from our findings. First, the need for additional training of medical staff in telemedicine technologies can be addressed by ensuring that HCPs can focus on medical examinations and treatments without being distracted by technical issues. Therefore, health care facilities and general practices should consider hiring specialized telemedicine support staff. This approach enables a more efficient division of labor and ensures that technical problems do not affect the quality of medical care. In addition, specific training programs can be developed or used to help staff mitigate risks and fully understand and leverage the benefits of telemedicine. Training should thus include best practice approaches and encourage exchange with other medical facilities. In addition, telemedicine should be formally integrated into medical school curriculums and clinical placements to prepare future HCPs for the evolving landscape of digital health care. This will ensure that medical students are equipped with both the technical and clinical skills necessary for using telemedicine effectively in their practice. However, it is important to note that the cost and feasibility of these recommendations may vary, particularly for smaller or resource-constrained organizations. Ideally, telemedicine software should be intuitive enough for existing staff, such as IT or administrative personnel, to manage with minimal training, although this may not always be feasible.

Second, the individually identified risks that are particularly relevant can be addressed in a more targeted manner. To better manage such risks, risk management guidelines could be developed to provide guidance and protocols to counteract the perceived risks of telemedicine, for example, by reducing the performance risk through enhanced remote examination with improved imaging technologies or remote monitoring tools to mitigate visibility limitations and diagnostic confusion. Data security protocols can also be strengthened, thus minimizing the privacy risk, as insurance with compliance privacy regulations sends a trustworthy signal.

Third, transparency is a significant antecedent that underlines the necessity to choose an appropriate provider and the right technology for the planned medical approach. On the basis of our results, HCPs should choose telemedicine platforms and technologies that fit their most important trust factors, for example, those that offer reliable, user-friendly interfaces and robust security features. Thus, professionals should look for providers that offer comprehensive support and training. In addition, health care institutions should aid HCPs in selecting the appropriate telemedicine platform and technology by choosing their recommendations based on the criteria that emerged as particularly relevant in this study.

Fourth, as not every symptom characteristic can be treated equally well or at all with telemedicine, providers should focus on developing solutions that are specialized for certain medical conditions and symptoms and can optimally support these. By tailoring their technologies and programs to diverse symptom characteristics, HCPs may more effectively manage treatment via telemedicine. However, developing specific policies and guidelines on when it is appropriate to use telemedicine may be a more practical approach, given the constraints of resources and the limitations of digital communication. These guidelines can help ensure that telemedicine is used optimally without compromising patient care.

Limitations and Further Research

First, the amendments to the framework are based on a qualitative approach that aimed to identify relevant constructs for the considered context. Next, they should be tested quantitatively for the strength of the proposed causal relationships and applied to other telemedicine applications, such as telemonitoring or videoconferencing between HCPs that are not diagnostic or a treatment. We have added the new context-adapted subcategories to already established constructs, such as multiple risks. Therefore, future research should review and adapt existing scales to ensure that they cover the relevant new subcategories. However, the current framework does not fully address key human factors such as task load burden and the user experience of interacting with telemedicine software in daily clinical practice. These factors represent ongoing, structural impacts on health care delivery rather than the transitional costs of implementation. Although this was not a focus of this study, it is important to acknowledge this limitation and consider it as an avenue for future research.

Second, we adjusted the sample size to suit the rarity of the niche population (ie, HCPs in Germany offering video consultations). Due to the low adoption rate, HCPs with an affinity for technology are more likely to use telemedicine. Therefore, the benefits could be overestimated and the perceived risks underestimated. In addition, the use of volunteer sampling may introduce self-selection bias, as those who choose to participate are likely to hold more favorable and supportive views of telemedicine. This could further increase the risk of overestimating its benefits and underestimating its risks. Thus, it has to be argued that there is a possible lack of generalizability of our results, which should be addressed in future studies by considering larger samples and subject triangulation [87]. However, we intentionally only interviewed HCPs with experience, ensuring that the perceived benefits and risks were based on real experience rather than assumptions. Moreover, we ensured a diverse sample in terms of age, gender, and specialization to take a variety of perspectives into account. Future research should compare our findings with the perspectives of HCPs who have not adopted telemedicine as well as their transferability to a patient's perspective. Furthermore, the transferability of our results to the patient's perspective ought to be examined, given that previous studies showed a greater disparity between HCPs and patients [88]. This will provide deeper insights into the barriers and facilitators to further implementation of telemedicine.

Third, our sample is limited to the German context with its national regulations and funding opportunities. Cultural factors

XSL•FO
RenderX

specific to Germany, such as high individualism and low uncertainty avoidance [89], may also influence perceptions of telemedicine. These biases could differ significantly in other cultural environments, particularly in low-income countries, where telemedicine adoption might be shaped by different cultural attitudes and resource constraints. Telemedicine is already more established in other countries, which might differ in terms of regulatory and organizational conditions [90,91]. Our study shows the specificity of early adopters, who may still face more prejudices, insecurities, and lack of habits. The results of this study can be adapted to other cultural environments with more established telemedicine offerings and reveal differing cultural-specific results. To address the relevance of the governmental framework, future research should analyze how recent and anticipated changes in health policy and regulations affect the adoption and effectiveness of telemedicine and how this might affect the constructs proposed in this study's adoption framework. In addition, the demographics and communities served by telemedicine users should be considered, as populations less affected by the digital divide may perceive lower risks and greater benefits compared to the groups considered disadvantaged.

Conclusions

Context-specific consideration is useful and shows which additional factors from generally applicable frameworks or theories are particularly relevant. It also reveals which factors, if any, do not seem relevant in the context of telemedicine from an HCP's perspective. Our results show that multidimensional transparency as an antecedent to trust, as well as the context-specific metavariables (symptom characteristics and experience), impact telemedicine adoption. The complexity and higher risk perception associated with telemedicine also lead to a multitude of perceived risks. For future research, the results of this study imply that prospective studies in the field of telemedicine should refer to the design of a modified telemedicine-contextualized valence framework to conduct a fully comprehensive analysis. The role of a broader systematic and organizational context should also be considered to understand how external factors shape telemedicine adoption. Especially when using quantitative approaches, existing scales should address telemedicine-specific characteristics of constructs such as trust referents, risks, and benefits, as well as antecedents such as transparency in their items. If they do not, the scales should be adapted accordingly.

Acknowledgments

The authors acknowledge the use of ChatGPT for assisting with grammar and language checks during the final preparation of this manuscript. This study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

FS was responsible for conceptualization, methodology, writing, project administration, and formal analysis. DW was responsible for conceptualization, methodology, writing, and supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide and coding schemes.
[DOCX File , 61 KB - formative_v9i1e54777_app1.docx]

References

1. Forkner-Dunn J. Internet-based patient self-care: the next generation of health care delivery. J Med Internet Res 2003;5(2):e8 [FREE Full text] [doi: 10.2196/jmir.5.2.e8] [Medline: 12857664]
2. Iyanna S, Kaur P, Ractham P, Talwar S, Najmul Islam A. Digital transformation of healthcare sector. What is impeding adoption and continued usage of technology-driven innovations by end-users? J Bus Res 2022 Dec;153:150-161. [doi: 10.1016/j.jbusres.2022.08.007]
3. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and facilitators to the implementation of eHealth services: systematic literature analysis. J Med Internet Res 2019 Nov 22;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]
4. Bashshur RL, Reardon TG, Shannon GW. Telemedicine: a new health care delivery system. Annu Rev Public Health 2000;21:613-637. [doi: 10.1146/annurev.publhealth.21.1.613] [Medline: 10884967]

XSL•FO
RenderX

5. Weißenfeld MM, Goetz K, Steinhäuser J. Facilitators and barriers for the implementation of telemedicine from a local government point of view - a cross-sectional survey in Germany. BMC Health Serv Res 2021 Sep 06;21(1):919 [FREE Full text] [doi: 10.1186/s12913-021-06929-9] [Medline: 34488753]
6. World development indicators. World Bank. URL: https://databank.worldbank.org/source/world-development-indicators [accessed 2023-11-06]
7. Harris MA, Freeman KA, Duke DC. Seeing is believing: using Skype to improve diabetes outcomes in youth. Diabetes Care 2015 Aug;38(8):1427-1434. [doi: 10.2337/dc14-2469] [Medline: 26033508]
8. Vismara LA, McCormick C, Young GS, Nadhan A, Monlux K. Preliminary findings of a telehealth approach to parent training in autism. J Autism Dev Disord 2013 Dec;43(12):2953-2969. [doi: 10.1007/s10803-013-1841-8] [Medline: 23677382]
9. Dorsey ER, Achey MA, Beck CA, Beran DB, Biglan KM, Boyd CM, et al. National randomized controlled trial of virtual house calls for people with Parkinson's disease: interest and barriers. Telemed J E Health 2016 Jul;22(7):590-598 [FREE Full text] [doi: 10.1089/tmj.2015.0191] [Medline: 26886406]
10. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. J Telemed Telecare 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]
11. Williams A, Larocca R, Chang T, Trinh NH, Fava M, Kvedar J, et al. Web-based depression screening and psychiatric consultation for college students: a feasibility and acceptability study. Int J Telemed Appl 2014;2014:580786 [FREE Full text] [doi: 10.1155/2014/580786] [Medline: 24799895]
12. Laato S, Islam AN, Farooq A, Dhir A. Unusual purchasing behavior during the early stages of the COVID-19 pandemic: the stimulus-organism-response approach. J Retail Consum Serv 2020 Nov;57:102224. [doi: 10.1016/j.jretconser.2020.102224]
13. Talwar S, Talwar M, Tarjanne V, Dhir A. Why retail investors traded equity during the pandemic? An application of artificial neural networks to examine behavioral biases. Psychol Mark 2021 Jul 26;38(11):2142-2163. [doi: 10.1002/mar.21550]
14. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 2020 Aug 01;20(1):1193 [FREE Full text] [doi: 10.1186/s12889-020-09301-4] [Medline: 32738884]
15. Webster P. Virtual health care in the era of COVID-19. Lancet 2020 Apr 11;395(10231):1180-1181 [FREE Full text] [doi: 10.1016/S0140-6736(20)30818-7] [Medline: 32278374]
16. Zeltzer D, Einav L, Balicer R. Healthcare delivery: what the rise of telemedicine means for you. World Economic Forum. URL: https://www.weforum.org/agenda/2021/07/telemedicine-covid-coronavirus-pandemic-healthcare/ [accessed 2024-09-28]
17. Bestsennyy O, Gilbert G, Harris A, Rost J. Telehealth: a quarter-trillion-dollar post-COVID-19 reality? McKinsey & Company. URL: https://www.mckinsey.com/industries/healthcare/our-insights/telehealth-a-quarter-trillion-dollar-post-covid-19-reality [accessed 2024-09-28]
18. Almathami HK, Win KT, Vlahu-Gjorgievska E. Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: systematic literature review. J Med Internet Res 2020 Feb 20;22(2):e16407 [FREE Full text] [doi: 10.2196/16407] [Medline: 32130131]
19. Dos Santos AD, Mata-Machado AT, Melo MD, Fonseca Sobrinho D, Araújo LL, Silva É, et al. Implementation of telehealth resources in primary care in Brazil and its association with quality of care. Telemed J E Health 2019 Oct;25(10):996-1004. [doi: 10.1089/tmj.2018.0166] [Medline: 30592699]
20. Katta M, Shao C, Herbey I, Modi S, Gleason L, Marques I, et al. IBD and the patient-provider relationship: does trust impact telemedicine utilization? Gastroenterology 2022 Feb;162(3):S83. [doi: 10.1053/j.gastro.2021.12.174]
21. Zheng S, Chang PY, Chen J, Chang YW, Fan HC. An investigation of patient decisions to use eHealth. J Organ End User Comput 2021;34(4):1-24. [doi: 10.4018/JOEUC.289433]
22. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. J Med Internet Res 2019 May 21;21(5):e13117 [FREE Full text] [doi: 10.2196/13117] [Medline: 31115340]
23. Kim DJ, Ferrin DL, Rao HR. A trust-based consumer decision-making model in electronic commerce: the role of trust, perceived risk, and their antecedents. Decis Support Syst 2008 Jan;44(2):544-564. [doi: 10.1016/j.dss.2007.07.001]
24. Wu TC, Ho CT. Barriers to telemedicine adoption during the COVID-19 pandemic in Taiwan: comparison of perceived risks by socioeconomic status correlates. Int J Environ Res Public Health 2023 Feb 16;20(4):23 [FREE Full text] [doi: 10.3390/ijerph20043504] [Medline: 36834205]
25. Jalali MS, Bruckes M, Westmattelmann D, Schewe G. Why employees (Still) click on phishing links: investigation in hospitals. J Med Internet Res 2020 Jan 23;22(1):e16775 [FREE Full text] [doi: 10.2196/16775] [Medline: 32012071]
26. Featherman MS, Pavlou PA. Predicting e-services adoption: a perceived risk facets perspective. Int J Hum Comput Stud 2003 Oct;59(4):451-474. [doi: 10.1016/S1071-5819(03)00111-3]
27. Alajlani M, Clarke M. Effect of culture on acceptance of telemedicine in Middle Eastern countries: case study of Jordan and Syria. Telemed J E Health 2013 Apr;19(4):305-311. [doi: 10.1089/tmj.2012.0106] [Medline: 23540280]

XSL•FO
RenderX

28. Luo J, Tong L, Crotty BH, Somai M, Taylor B, Osinski K, et al. Telemedicine adoption during the COVID-19 pandemic: gaps and inequalities. Appl Clin Inform 2021 Aug;12(4):836-844 [FREE Full text] [doi: 10.1055/s-0041-1733848] [Medline: 34496419]
29. Li Y, James L, McKibben J. Trust between physicians and patients in the e-health era. Technol Soc 2016 Aug;46:28-34. [doi: 10.1016/j.techsoc.2016.02.004]
30. Kuen L, Schürmann F, Westmattelmann D, Hartwig S, Tzafrir S, Schewe G. Trust transfer effects and associated risks in telemedicine adoption. Electron Mark 2023 Jul 21;33(1):6. [doi: 10.1007/S12525-023-00657-0]
31. Garavand A, Aslani N, Nadri H, Abedini S, Dehghan S. Acceptance of telemedicine technology among physicians: a systematic review. Inform Med Unlocked 2022;30:100943. [doi: 10.1016/j.imu.2022.100943]
32. Ghiwaa T, Khan I, White M, Beloff N. Telemedicine adoption for healthcare delivery: a systematic review. Int J Adv Comput Sci Appl 2023;14(11):1230-1243. [doi: 10.14569/IJACSA.2023.01411125]
33. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157. [doi: 10.2307/41410412]
34. Nittari G, Khuman R, Baldoni S, Pallotta G, Battineni G, Sirignano A, et al. Telemedicine practice: review of the current ethical and legal challenges. Telemed J E Health 2020 Dec;26(12):1427-1437 [FREE Full text] [doi: 10.1089/tmj.2019.0158] [Medline: 32049608]
35. Shachak A, Kuziemsky C, Petersen C. Beyond TAM and UTAUT: future directions for HIT implementation research. J Biomed Inform 2019 Dec;100:103315 [FREE Full text] [doi: 10.1016/j.jbi.2019.103315] [Medline: 31629923]
36. Meinert E, Alturkistani A, Brindley D, Knight P, Wells G, de Pennington N. Weighing benefits and risks in aspects of security, privacy and adoption of technology in a value-based healthcare system. BMC Med Inform Decis Mak 2018 Nov 13;18(1):100 [FREE Full text] [doi: 10.1186/s12911-018-0700-0] [Medline: 30424753]
37. Bakshi S, Tandon U. Understanding barriers of telemedicine adoption: a study in North India. Syst Res Behav Sci 2021 Jan 04;39(1):128-142. [doi: 10.1002/sres.2774]
38. Ly BA, Labonté R, Bourgeault IL, Niang MN. The individual and contextual determinants of the use of telemedicine: a descriptive study of the perceptions of Senegal's physicians and telemedicine projects managers. PLoS One 2017;12(7):e0181070 [FREE Full text] [doi: 10.1371/journal.pone.0181070] [Medline: 28732028]
39. Hong W, Chan FK, Thong JY, Chasalow LC, Dhillon G. A framework and guidelines for context-specific theorizing in information systems research. Inf Syst Res 2014 Mar;25(1):111-136. [doi: 10.1287/isre.2013.0501]
40. Kim DJ, Ferrin DL, Rao HR. Trust and satisfaction, two stepping stones for successful e-commerce relationships: a longitudinal exploration. Inf Syst Res 2009 Jun;20(2):237-257. [doi: 10.1287/isre.1080.0188]
41. Chin AG, Harris MA, Brookshire R. An empirical investigation of intent to adopt mobile payment systems using a trust-based extended valence framework. Inf Syst Front 2020 Nov 03;24(1):329-347. [doi: 10.1007/S10796-020-10080-X]
42. Stoffers B, Lanfer J, Westmattelmann D, Schewe G. One app to test them all -opportunities and challenges of implementing a remote sampling system in anti-doping work. In: Proceedings of the 2023 International Conference on Information Systems. 2023 Presented at: ICIS '23; December 10-13, 2023; Hyderabad, India p. 1-17 URL: https://www.researchgate.net/publication/375029604_One_App_to_Test_Them_All_-Opportunities_and_Challenges_of_Implementing_a_Remote_Sampling_System_in_Anti-Doping_Work
43. Mou J, Shin DH, Cohen J. Health beliefs and the valence framework in health information seeking behaviors. Inf Technol People 2016 Nov 07;29(4):876-900. [doi: 10.1108/ITP-06-2015-0140]
44. Venkatesh V, Thong JY, Xu X. Unified theory of acceptance and use of technology: a synthesis and the road ahead. J Assoc Inf Syst 2016;17(5):328-376. [doi: 10.17705/1JAIS.00428]
45. Dowling GR, Staelin R. A model of perceived risk and intended risk-handling activity. J Consum Res 1994 Jun;21(1):119. [doi: 10.1086/209386]
46. Lim N. Consumers' perceived risk: sources versus consequences. Electron Commer Res Appl 2003 Sep;2(3):216-228. [doi: 10.1016/S1567-4223(03)00025-5]
47. Grewal D, Gotlieb J, Marmorstein H. The moderating effects of message framing and source credibility on the price-perceived risk relationship. J Consum Res 1994 Jun;21(1):145. [doi: 10.1086/209388]
48. Jacoby J, Kaplan LB. The components of perceived risk. In: Proceedings of the 3rd Annual Conference of the Association for Consumer Research. 1972 Presented at: ACR '72; November 3-5, 1972; Chicago, IL p. 382-393 URL: https://www.researchgate.net/publication/247814928_The_Components_Of_Perceived_Risk
49. Van Velsen L, Wentzel J, Van Gemert-Pijnen JE. Designing eHealth that matters via a multidisciplinary requirements development approach. JMIR Res Protoc 2013 Jun 24;2(1):e21 [FREE Full text] [doi: 10.2196/resprot.2547] [Medline: 23796508]
50. Harris K, Schultz J, Feldman R. Measuring consumer perceptions of quality differences among competing health benefit plans. J Health Econ 2002 Jan;21(1):1-17. [doi: 10.1016/s0167-6296(01)00098-4] [Medline: 11845919]
51. Pitic L, Brad S, Pitic D. Study on perceived quality and perceived fair price. Procedia Econ Financ 2014;15:1304-1309. [doi: 10.1016/S2212-5671(14)00592-9]
52. Shah A. The impact of quality on satisfaction, revenue, and cost as perceived by providers of higher education. J Mark High Educ 2009 Dec 16;19(2):125-141. [doi: 10.1080/08841240903451324]

XSL•FO
RenderX

53. Call VR, Erickson LD, Dailey NK, Hicken BL, Rupper R, Yorgason JB, et al. Attitudes toward telemedicine in urban, rural, and highly rural communities. Telemed J E Health 2015 Aug;21(8):644-651. [doi: 10.1089/tmj.2014.0125] [Medline: 25839334]
54. Kim DJ, Song YI, Braynov S, Rao H. A multidimensional trust formation model in B-to-C e-commerce: a conceptual framework and content analyses of academia/practitioner perspectives. Decis Support Syst 2005 Aug;40(2):143-165. [doi: 10.1016/j.dss.2004.01.006]
55. McKnight DH, Chervany NL. What trust means in e-commerce customer relationships: an interdisciplinary conceptual typology. Int J Electron Commer 2014 Dec 23;6(2):35-59. [doi: 10.1080/10864415.2001.11044235]
56. Massey PA, Montoya-Weiss M. Unraveling the temporal fabric of knowledge conversion: a model of media selection and use. MIS Q 2006;30(1):99-114. [doi: 10.2307/25148720]
57. Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. Acad Manag Rev 1995 Jul;20(3):709. [doi: 10.2307/258792]
58. Kelton K, Fleischmann KR, Wallace WA. Trust in digital information. J Am Soc Inf Sci 2007 Dec 04;59(3):363-374. [doi: 10.1002/asi.20722]
59. Mcknight DH, Carter M, Thatcher JB, Clay PF. Trust in a specific technology: an investigation of its components and measures. ACM Trans Manage Inf Syst 2011 Jul 01;2(2):1-25. [doi: 10.1145/1985347.1985353]
60. Van Velsen L, Wildevuur S, Flierman I, Van Schooten B, Tabak M, Hermens H. Trust in telemedicine portals for rehabilitation care: an exploratory focus group study with patients and healthcare professionals. BMC Med Inform Decis Mak 2016 Jan 27;16:11 [FREE Full text] [doi: 10.1186/s12911-016-0250-2] [Medline: 26818611]
61. van Velsen L, Tabak M, Hermens H. Measuring patient trust in telemedicine services: development of a survey instrument and its validation for an anticoagulation web-service. Int J Med Inform 2017 Jan;97:52-58. [doi: 10.1016/j.ijmedinf.2016.09.009] [Medline: 27919395]
62. Schultze U, Avital M. Designing interviews to generate rich data for information systems research. Inf Organ 2011 Jan;21(1):1-16. [doi: 10.1016/j.infoandorg.2010.11.001]
63. van Velsen L, Flierman I, Tabak M. The formation of patient trust and its transference to online health services: the case of a Dutch online patient portal for rehabilitation care. BMC Med Inform Decis Mak 2021 Jun 12;21(1):188 [FREE Full text] [doi: 10.1186/s12911-021-01552-4] [Medline: 34118919]
64. Thabane L, Ma J, Chu R, Cheng J, Ismaila A, Rios LP, et al. A tutorial on pilot studies: the what, why and how. BMC Med Res Methodol 2010 Jan 06;10:1 [FREE Full text] [doi: 10.1186/1471-2288-10-1] [Medline: 20053272]
65. Mayring P. Qualitative content analysis. In: Flick U, von Kardoff E, Steinke I, editors. A companion to qualitative research. Thousand Oaks, CA: Sage Publications; 2004:159-176.
66. Tolle M, Stratkötter A. Die Geschlechtszugehörigkeit von Therapeutinnen und Therapeuten in der psychotherapeutischen Arbeit: Ein integratives Modell. In: Breuer F, editor. Qualitative Psychologie: Grundlagen, Methoden und Anwendungen eines Forschungsstils. Cham, Switzerland: Springer; 2022:251-266.
67. Myers MD, Newman M. The qualitative interview in IS research: examining the craft. Inf Organ 2007 Jan;17(1):2-26. [doi: 10.1016/j.infoandorg.2006.11.001]
68. Sarker S, Xiao X, Beaulieu T. Guest editorial: qualitative studies in information systems: a critical review and some guiding principles. MIS Q 2013;37(4):iii-xviii [FREE Full text]
69. Jameda GmbH. URL: https://pro.jameda.de/ [accessed 2024-04-29]
70. Bowen GA. Naturalistic inquiry and the saturation concept: a research note. Qual Res 2008 Feb 01;8(1):137-152. [doi: 10.1177/1468794107085301]
71. Schnackenberg AK, Tomlinson EC. Organizational transparency. J Manag 2016 Jul 09;42(7):1784-1810. [doi: 10.1177/0149206314525202]
72. Venkatesh V, Brown SA, Bala H. Bridging the qualitative-quantitative divide: guidelines for conducting mixed methods research in information systems. MIS Q 2013 Jan 1;37(1):21-54. [doi: 10.25300/MISQ/2013/37.1.02]
73. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159. [doi: 10.2307/2529310]
74. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
75. Tomlinson EC, Schnackenberg A. The effects of transparency perceptions on trustworthiness perceptions and trust. J Trust Res 2022 Apr 18;12(1):1-23. [doi: 10.1080/21515581.2022.2060245]
76. Schnackenberg AK, Tomlinson E, Coen C. The dimensional structure of transparency: a construct validation of transparency as disclosure, clarity, and accuracy in organizations. Hum Relat 2020 Jun 27;74(10):1628-1660. [doi: 10.1177/0018726720933317]
77. Lee YW, Strong DM, Kahn BK, Wang RY. AIMQ: a methodology for information quality assessment. Inf Manag 2002 Dec;40(2):133-146. [doi: 10.1016/S0378-7206(02)00043-5]
78. Orrange S, Patel A, Mack WJ, Cassetta J. Patient satisfaction and trust in telemedicine during the COVID-19 pandemic: retrospective observational study. JMIR Hum Factors 2021 Apr 22;8(2):e28589 [FREE Full text] [doi: 10.2196/28589] [Medline: 33822736]

XSL•FO
RenderX

79. Sabani A. Investigating the influence of transparency on the adoption of e-Government in Indonesia. J Sci Technol Policy Manag 2020 Nov 23;12(2):236-255. [doi: 10.1108/jstpm-03-2020-0046]
80. Märtins J, Kehl R, Westmattelmann D, Schewe G. A multi-dimensional theory of transparency, trust, and risk in technology adoption. In: Proceedings of the 2022 Conference on International Conference on Information Systems. 2022 Presented at: ICIS '22; December 16-20, 2022; Copenhagen, Denmark p. 1-17 URL: https://www.researchgate.net/publication/365490186_A_Multi-Dimensional_Theory_of_Transparency_Trust_and_Risk_in_Technology_Adoption
81. Oldeweme A, Märtins J, Westmattelmann D, Schewe G. The role of transparency, trust, and social influence on uncertainty reduction in times of pandemics: empirical study on the adoption of COVID-19 tracing apps. J Med Internet Res 2021 Feb 08;23(2):e25893 [FREE Full text] [doi: 10.2196/25893] [Medline: 33465036]
82. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A systematic review of the technology acceptance model in health informatics. Appl Clin Inform 2018 Jul;9(3):604-634 [FREE Full text] [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
83. Sintonen S, Immonen M. Telecare services for aging people: assessment of critical factors influencing the adoption intention. Comput Human Behav 2013 Jul;29(4):1307-1317. [doi: 10.1016/j.chb.2013.01.037]
84. Serrano KM, Mendes GH, Lizarelli FL, Ganga GM. Assessing the telemedicine acceptance for adults in Brazil. Int J Health Care Qual Assur (Forthcoming) 2020 Dec 24. [doi: 10.1108/IJHCQA-06-2020-0098] [Medline: 33369378]
85. Wu D, Gu H, Gu S, You H. Individual motivation and social influence: a study of telemedicine adoption in China based on social cognitive theory. Health Policy Technol 2021 Sep;10(3):100525. [doi: 10.1016/j.hlpt.2021.100525]
86. Bohnet I, Huck S. Repetition and reputation: implications for trust and trustworthiness in the short and in the long run. SSRN Journal. Preprint posted online February 6, 2004 [FREE Full text] [doi: 10.2139/ssrn.474544]
87. Patton MQ. Qualitative Evaluation and Research Methods. 2nd edition. Thousand Oaks, CA: Sage Publications; 2001.
88. Loriot A, Larceneux F, Guillard V, Bertocchio JP. Patients' representations of perceived distance and proximity to telehealth in France: qualitative study. J Med Internet Res 2024 Apr 26;26:e45702 [FREE Full text] [doi: 10.2196/45702] [Medline: 38669676]
89. Hofstede G. Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations. Thousand Oaks, CA: Sage Publications; 2001.
90. Nittari G, Savva D, Tomassoni D, Tayebati SK, Amenta F. Telemedicine in the COVID-19 era: a narrative review based on current evidence. Int J Environ Res Public Health 2022 Apr 22;19(9):23 [FREE Full text] [doi: 10.3390/ijerph19095101] [Medline: 35564494]
91. Luciano E, Mahmood MA, Mansouri Rad P. Telemedicine adoption issues in the United States and Brazil: perception of healthcare professionals. Health Informatics J 2020 Dec;26(4):2344-2361. [doi: 10.1177/1460458220902957] [Medline: 32072843]

Abbreviations

EVF: extended valence framework
HCP: health care professional
IS: information systems
TAM: technology acceptance model
UTAUT: unified theory of acceptance and use of technology

Edited by A Mavragani; submitted 21.11.23; peer-reviewed by D Singh, D Hua, G Sebastian, JP Bertocchio; comments to author 08.08.24; revised version received 01.10.24; accepted 01.12.24; published 27.01.25.

Please cite as:
Schürmann F, Westmattelmann D, Schewe G
Factors Influencing Telemedicine Adoption Among Health Care Professionals: Qualitative Interview Study
JMIR Form Res 2025;9:e54777
URL: https://formative.jmir.org/2025/1/e54777
doi:10.2196/54777
PMID:

©Fiona Schürmann, Daniel Westmattelmann, Gerhard Schewe. Originally published in JMIR Formative Research (https://formative.jmir.org), 27.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Digital Tool (Technology-Assisted Problem Management Plus) for Lay Health Workers to Address Common Mental Health Disorders: Co-production and Usability Study in Pakistan

Maham Saleem[1,2], MSPH; Shamsa Zafar[3], Prof Dr Med; Thomas Klein[4], MSc; Markus Koesters[5], Prof Dr; Adnan Bashir[6], MCS; Daniela C Fuhr[1], Prof Dr; Siham Sikander[7], FCPS, PhD; Hajo Zeeb[1,2], Prof Dr Med

[1]Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany

[2]Leibniz Science Campus Digital Public Health, Bremen, Germany

[3]Fazaia Medical College, Air University, Islamabad, Pakistan

[4]Department of Psychiatry and Psychotherapy II, District Hospital Guenzburg, Ulm University, Ulm, Germany

[5]Center for Evidence-Based Healthcare, Medical Faculty Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

[6]Health Information Systems Program, Islamabad, Pakistan

[7]Department of Primary Care and Mental Health, University of Liverpool, Liverpool, United Kingdom

Corresponding Author:
Maham Saleem, MSPH
Leibniz Institute for Prevention Research and Epidemiology – BIPS
Achterstraße 30
Bremen, 28359
Germany
Phone: 49 421 218 ext 56760
Email: saleem@leibniz-bips.de

Abstract

Background: Mental health remains among the top 10 leading causes of disease burden globally, and there is a significant treatment gap due to limited resources, stigma, limited accessibility, and low perceived need for treatment. Problem Management Plus, a World Health Organization–endorsed brief psychological intervention for mental health disorders, has been shown to be effective and cost-effective in various countries globally but faces implementation challenges, such as quality control in training, supervision, and delivery. While digital technologies to foster mental health care have the potential to close treatment gaps and address the issues of quality control, their development requires context-specific, interdisciplinary, and participatory approaches to enhance impact and acceptance.

Objective: We aimed to co-produce Technology-Assisted Problem Management Plus (TA-PM+) for "lady health workers" (LHWs; this is the terminology used by the Lady Health Worker Programme for lay health workers) to efficiently deliver sessions to women with symptoms of common mental health disorders within the community settings of Pakistan and conducted usability testing in community settings.

Methods: A 3-stage framework was used for co-producing and prototyping the intervention. Stage 1 (evidence review and stakeholder consultation) included 3 focus group discussions with 32 LHWs and 7 in-depth interviews with key stakeholders working in the health system or at the health policy level. Thematic analyses using the Capability, Opportunity, and Motivation for Behavioral Change (COM-B) model were conducted. Stage 2 included over eight online workshops, and a multidisciplinary intervention development group co-produced TA-PM+. Stage 3 (prototyping) involved 2 usability testing rounds. In round 1 conducted in laboratory settings, 6 LHWs participated in role plays and completed the 15-item mHealth Usability App Questionnaire (MUAQ) (score range 0-7). In round 2 conducted in community settings, trained LHWs delivered the intervention to 6 participants screened for depression and anxiety. Data were collected using the MUAQ completed by LHWs and the Patient Satisfaction Questionnaire (PSQ) (score range 0-46) completed by participants.

Results: Qualitative analysis indicated that a lack of digital skills among LHWs, high workload, resource scarcity for digitization (specifically internet bandwidth in the community), and need for comprehensive training were barriers for TA-PM+ implementation in the community through LHWs. Training, professional support, user guidance, an easy and automated interface, offline functionalities, incentives, and strong credibility among communities were perceived to enhance the capability, opportunity, and motivation of LHWs to implement TA-PM+. TA-PM+ was co-produced with features like an automated interface, a personal

XSL•FO
RenderX

dashboard, guidance videos, and a connected supervisory panel. The mean MUAQ score was 5.62 in round 1 of usability testing and improved to 5.96 after incorporating LHW feedback in round 2. The mean PSQ score for TA-PM+ was 40 in round 2.

Conclusions: Co-production of TA-PM+ for LHWs balanced context and evidence. The 3-stage iterative development approach resulted in high usability and acceptability of TA-PM+ for LHWs and participants.

(JMIR Form Res 2025;9:e59414) doi:10.2196/59414

KEYWORDS

digital mental health; digital technology; digital intervention; Problem Management Plus; lay health worker programme; common mental health disorders; low- and middle-income countries; co-production

Introduction

Background

Recent research has indicated that mental health remains among the top 10 leading causes of disease burden globally and constitutes 32.4% of years lived with disability [1]. Common mental health disorders are conditions that impact mood, thought processes, and behavior, often leading to challenges in daily functioning and overall well-being. These include disorders, such as depression, anxiety, and bipolar disorder, each characterized by unique symptoms and impacts on the lives of individuals [2]. Depression, affecting over 300 million people globally, is the most prevalent mental health disorder [3]. A major part of this burden (about 70%) is concentrated in low- and middle-income countries (LMICs). Additionally, a shortage of mental health specialists in LMICs contributes to a mental health treatment gap of 90% for almost all mental health conditions, meaning that only 1 out of 10 conditions requiring mental health treatment receive it from a mental health specialist [3]. Various factors contribute to the mental health treatment gap, including lack of human resources for mental health, stigma, accessibility, uneven mental health service distribution, and lack of a perceived need for treatment [4]. To address this gap, the World Health Organization (WHO) developed the Mental Health Gap Action Programme (mhGAP) by providing evidence-based interventions for 8 priority mental health disorders. These guidelines recommend involving lay health workers and community members in the early detection of mental health disorders and low-intensity psychological care of patients with mental health disorders. This can help LMICs address the treatment gap and health resource shortfall, and achieve WHO Mental Health Action Plan targets [3,4]. Building on the WHO's mhGAP, several initiatives in LMICs have been developed to address the mental health treatment gap [5-7]. These programs empower nonspecialist health workers to deliver mental health care through structured culturally adapted interventions. Examples include Problem Management Plus (PM+), a transdiagnostic intervention for managing stress and depression, and the Thinking Healthy Programme, which addresses perinatal depression through simplified behavioral strategies. Community-based programs like Friendship Bench in Zimbabwe further extend the reach of mhGAP by embedding mental health support within local settings, demonstrating significant improvements in access and patient outcomes. The Disease Control Priorities (DCP3) report endorses such scalable models as effective affordable solutions that bolster mental health care capacity in LMICs [6,7]. A comprehensive review [7] of 208 studies focusing on various mhGAP interventions in LMICs demonstrated their overall efficacy. The mhGAP interventions were shown to significantly improve the symptoms of common mental health disorders, reduce substance use, and positively impact parenting and child outcomes. Furthermore, mhGAP training across these interventions enhanced the clinical skills, knowledge, and competency of health workers, contributing to improved quality of care after training [7].

PM+ is a low-intensity brief psychological and transdiagnostic intervention for common mental health disorders. It is an evidence-based intervention that can address various mental health issues and address the limitations posed by scarce specialized treatments in low-resource settings [8-14]. PM+ has been designed for depression, anxiety, stress, grief, and self-identified practical issues (interpersonal conflict and unemployment) and is delivered by lay health workers [15]. The intervention consists of 5 weekly face-to-face sessions involving stress management and problem-solving approaches [15]. PM+ has been tested for effectiveness and cost-effectiveness in various LMICs [8-14]. It has proven to be particularly effective in reducing the symptoms of anxiety, depression, and general distress in settings affected by significant adversity, including conflict-affected regions or under-resourced settings [16]. PM+ is recommended for stepped-care models, especially in low-resource settings where it serves as an effective first-line intervention. In these contexts, PM+ addresses common mental health needs at a low cost, allowing higher-intensity services to be reserved for more severe cases. This approach optimizes resource use and enhances accessibility in areas with limited mental health infrastructure. Notably, PM+ is available in several languages including Urdu, and it has been contextualized and rigorously evaluated in Pakistan [10,12,13].

However, the implementation of PM+ can be challenging for lay health workers and can be hampered by issues of quality control in training, supervision, and delivery [8-11]. Lengthy sessions and the complex nature of systematic stress and patient problem management strategies further compound these challenges, along with the struggle to recall all core strategies and protocols [8-11]. This specific phenomenon is termed "program drift" in the field of implementation science, which refers to a situation where the quality and fidelity of an intervention diminish because of deviations from the manual protocol, resulting in a reduction in the potential benefits of the intervention [17]. These challenges and barriers can be overcome by using digital tools and training lay health workers for them [18]. A recent systematic review of studies from LMICs revealed

XSL•FO
RenderX

that the integration of digital technologies into mental health interventions led by nonspecialists typically results in positive outcomes and has the potential to address the following three gaps highlighted in mental health literature: (1) the treatment gap, (2) the treatment-quality gap by improving the competencies and knowledge of nonspecialists, and (3) the service utilization gap by increasing the use of mental health care services [18].

Few initiatives have pursued digital adaptations of PM+ to enhance accessibility and scalability, particularly in low-resource settings. The step-by-step intervention was initially conceptualized as an online self-help adaptation of PM+. While the step-by-step intervention initially incorporated the principles of PM+, the intervention diverged as the complexity of implementing problem-solving components within a digital app proved challenging. Consequently, the step-by-step intervention evolved into an independent transdiagnostic intervention focused on managing stress and addressing depression and anxiety through behavioral activation principles [19]. In response to the COVID-19 pandemic, additional digital training frameworks were developed to enable remote training of lay health workers, ensuring uninterrupted mental health support amid physical distancing measures [20].

PM+ has been rigorously evaluated in Pakistan, showing effectiveness in reducing depression and anxiety symptoms [10,12,13]. Despite higher initial costs than the standard care, it has demonstrated long-term cost-effectiveness and is recommended for integration into a stepped-care model in resource-limited settings like Pakistan [12]. In a stepped-care model [21], individuals begin with lower-intensity support and progress to more intensive treatments only if necessary, optimizing resources and tailoring care to individual needs. This approach is particularly relevant in Pakistan, where mental health services are limited and mostly concentrated in major urban centers [22]. Pakistan's health care system, with its community outreach through "lady health workers" (LHWs; this is the terminology used by the Lady Health Worker Programme for lay health workers) [23], who have established supervision and referral mechanisms, provides a strong foundation for implementing stepped-care mental health interventions like PM+. To address the challenges in implementing PM+ by applying technological assistance and to align with the recommended stepped-care model for Pakistan, we aimed to develop a technological platform Technology-Assisted Problem Management Plus (TA-PM+). TA-PM+ consists of an app designed to support nonspecialists in delivering the PM+ intervention within Pakistan's health system, with an established backend for supervision, monitoring, and referral processes. TA-PM+ seeks to ensure the consistent delivery of therapeutic components at the appropriate dosage and with high fidelity while being delivered face-to-face, fostering a therapeutic alliance, which is essential for maintaining the problem-management focus required for effective intervention delivery. This paper describes the stepwise co-production of TA-PM+ and its key features.

Objective

We aimed to co-produce TA-PM+ for LHWs to use in efficiently delivering sessions to women within community settings in Pakistan.

Methods

Setting and Context

The study was conducted in the Federal District of Islamabad Capital Territory (ICT), Pakistan. The study considered the Union Council (UC) of Tarlai, which has a population of 100,000 and is located about 10 kilometers outside the ICT. The UC is a predominantly low-income and lower middle–class area, with a significant number of migrants from different parts of Pakistan. The crude birth rate is estimated to be around 40 births per 1000 population, with families typically having 4-8 children. Traditional gender roles prevail in families. Among the poorer sections of the population, several adversities are common, including malnutrition, irregular income, and low literacy rates [24].

Each UC is served by the primary health care unit, which consists of a medical doctor, vaccinators, and a pharmacist, and serves as the base for the community outreach program of LHWs. The LHWs are supervised and monitored by a facility-based supervisor, known as the lady health supervisor (LHS). The LHWs provide family planning, and maternal and child preventive and promotive services to 250-300 households in the catchment area and establish referrals for high-risk cases. More recently, maternal and adolescent mental health awareness and prevention have been added to the curriculum. The UC of Tarlai has 32 LHWs who cover nearly 80% of the population of Tarlai.

Ethical Considerations

This study has been approved by the Institutional Review Board of Fazaia Medical College, Islamabad, Pakistan (approval number: IBD/FMC/1341/IRB), in accordance with the ethical guidelines for human subject research and the Declaration of Helsinki. Informed consent was obtained from all participants, and for secondary analyses, the original consent permitted data use without requiring additional permissions. All data were anonymized to ensure privacy and were securely stored on a protected cloud platform accessible only to the research team. Participants received modest compensation, such as financial allowances and travel reimbursements (US $5-7). No identifiable participant information has been included in the manuscript or supplementary materials.

TA-PM+ Co-production

In line with the 3-stage framework of co-production and prototyping of public health interventions [25], TA-PM+ was designed in three stages: (1) qualitative investigation, (2) co-production, and (3) prototyping and usability testing. Qualitative investigation used semistructured interviews and was designed and reported in line with the Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist (Multimedia Appendix 1).

XSL•FO
RenderX

Stage 1: Qualitative Investigation

A qualitative study was conducted to inform the development of TA-PM+. The main objectives were to explore implementation barriers; understand user needs (including training and supervision), particularly those related to technology use assisting intervention delivery; evaluate monitoring; and identify TA-PM+ functionalities that could be used to strategically address barriers and aspects of delivering PM+ sessions. The sample for qualitative investigation consisted of LHWs, an LHS, and participants working in the health system or at the health policy level. Focus group discussions and in-depth interviews were conducted as part of qualitative investigation. All participants were given information about the research aims, and rapport was established.

Focus Group Discussions

A trained female researcher conducted 3 focus groups between June and August 2022 with the LHWs (n=32). They were purposefully sampled from a range of settings to ensure diversity in demographic factors, age, and education. The focus group discussions were held at a primary health care facility in Islamabad, Pakistan. One focus group discussion lasted between 70 and 85 minutes. The focus group discussions were held in Urdu, and data on mental health, digitization, and intervention functions perceived to be useful for TA-PM+ were elicited using semistructured techniques (Multimedia Appendix 2).

In-Depth Interviews

Interviews were conducted with a purposive sample of 7 participants working in the health system (n=2) or at the health policy level (n=5) between July and August 2022. The participants included an LHS, doctors at the primary health care unit of the health system, health system specialists, individuals in Director General programs, program officers of noncommunicable diseases, and digital health and information specialists at the policy level. The interviews were conducted in Urdu and face-to-face at a location of the participant's choice by a trained female researcher. The duration of the interviews ranged from 35 to 45 minutes, during which a semistructured interview guide was employed (Multimedia Appendix 2).

Data Analysis

The focus group discussions and interviews were audio recorded and transcribed verbatim. The framework analysis approach was used for analyses consisting of familiarization, identification of the thematic framework, charting, mapping, and interpretation. A second author (SZ) was engaged to independently code a randomly selected subset of 20% of the transcripts. The deductive framework was constructed using the Capability Opportunity and Motivation for Behavioral Change framework and the Theoretical Domain Framework (TDF) [26]. The TDF was implemented in deductive analyses to include various elements that otherwise might have been ignored. However, the researcher realized that not all constructs would be used to code the data, and constructs that did not yield any findings were excluded. The thematic framework was refined through repetitive discussions between the researchers MS and SZ. To ensure the authenticity of the data and to minimize the risk of data loss, the coding process was carried out on the transcript written in Urdu. Once indexing and charting were completed using MaxQDA, the coded sections of the transcript were double translated into English.

Stage 2: Co-production of TA-PM+

During stage 2, TA-PM+ was collaboratively developed by an intervention development group comprising both local (n=5) and international (n=5) members. This group included professionals with diverse backgrounds, including software developers, clinical psychiatrists, epidemiologists, public health researchers, and health system specialists.

Co-production Workshops

Eight co-production workshops were conducted online on Zoom between July and October 2022. Each co-production workshop lasted 120 to 150 minutes and was conducted with the aim of planning TA-PM+ features and delivery mechanisms. The findings from stage 1 were presented, ideas were sought from all members, and a storyboard was formulated. Subsequently, software designers presented concepts, prototypes, and avatar samples, and this was followed by the collection of feedback and iterative refinement, ultimately resulting in the finalization of features for TA-PM+. All the workshops were recorded, and after each workshop, a researcher (MS) summarized the findings decided during the workshop, which included technological features, scripts of the videos from the PM+ manual, and text for different interfaces of sessions. Additionally, member checking [27] (ie, sharing the findings with all members of the workshop) was carried out to verify the authenticity of the summarized findings. The finalized summary was shared with the software designers who prepared diagrams, vignettes, icon pictures, and interface mock-ups and was presented at the next workshop where feedback was sought again. This process was replicated across all 8 workshops, resulting in the finalization of all features, interface mock-ups, and videos for TA-PM+.

TA-PM+ Technological Platform Development

TA-PM+ was developed using the Agile methodology fostering flexibility and adaptability throughout the app's development and maintenance phases [28] from October 2022 till May 2023. The Agile methodology is an iterative approach to software development that prioritizes flexibility, customer collaboration, and continuous improvement. It emphasizes adaptive planning, incremental delivery, and embracing change. Agile teams work in short iterations to deliver working software increments, gather feedback, and adapt to evolving requirements. TA-PM+ is designed to operate on standard mobile devices, including iOS and Android platforms, underpinned by a robust hardware infrastructure. The mobile app and web dashboard are hosted on servers that necessitate a configuration with a minimum of 4 GB of RAM and 20 GB of storage capacity, ensuring the smooth operation of the app. MySQL is employed as the database management system on the server, providing a reliable platform for data management.

The software components of TA-PM+ are developed using advanced programming languages. PHP 7.3 and PHP 8.0 are leveraged for web development, while React Native is used for the mobile app and Laravel 8.40 is used for web development.

XSL•FO
RenderX

The TA-PM+ mobile app, developed using React Native, serves as the primary interface for users. The server-side component, developed using PHP (Laravel), involves a RESTful API and is responsible for authenticating users, conducting assessments, and facilitating data retrieval. The web dashboard, built using PHP (Laravel), is accessible via web browsers and is primarily intended for administrators. It provides a user-friendly interface for content management, configuration settings, and report viewing.

Stage 3: Prototyping and Usability Testing

Prototyping

After the co-production, external experts (n=2) in the PM+ intervention and in clinical psychiatry prototyped the content, resources, and videos of TA-PM+. The content was updated according to their feedback. The intervention development team examined the application for heuristic evaluation employing Jakob Nielsen heuristics [29].

Cultural Adaptation

We used the culturally adapted and Urdu-translated PM+ manual developed for the Pakistan context [10,12,13]. To further enhance cultural relevance, psychoeducational videos for patients were developed with detailed adjustments, including culturally appropriate avatars, dress codes, names, video backgrounds, and case scenarios. Voiceovers were provided by a native Urdu speaker, and the videos were reviewed by clinical psychiatrists (n=2) in Pakistan for language refinement. LHWs (n=6) participated in usability testing of TA-PM+, providing feedback on wording, icon labels, and flowcharts in the app. Additionally, women (n=6) from the community together with LHWs (n=6) evaluated the videos for cultural appropriateness and sensitivity. All feedback was incorporated to optimize clarity, usability, and cultural alignment.

Usability Testing

To evaluate the preliminary acceptance and utility of TA-PM+, usability testing sessions were carried out with a convenience sample of LHWs (n=6) in 2 rounds. Nelson [30] has established that a total of 5 participants is sufficient to detect usability issues.

Recruitment

Six LHWs were randomly recruited from the primary health care unit at Tarlai for usability testing. The LHWs had a comprehensive training program of 15 hours, which focused on their proficiency in using TA-PM+. The training agenda included introduction to common mental health disorders, confidentiality, PSYCHLOPS (Psychological Outcome Profiles), introduction to the PM+ intervention, content of the first session of PM+ on stress management, and delivery of the first session using TA-PM+. Subsequently, for round 2 of usability testing, recruitment was conducted in community settings, and the LHWs identified individuals exhibiting symptoms indicative of a common mental health problem from their catchment areas. A researcher screened these women for the symptoms of common mental health disorders. The following two screening tools were used (specific cutoffs were applied): (1) General Health Questionnaire-12 (GHQ-12): a score of 3 or higher on the GHQ-12, which is a 12-item questionnaire measuring general psychological distress on a 4-point scale (possible range of 0 to 36) and (2) WHO Disability Assessment Schedule (WHODAS): a score of 16 or higher on the WHODAS 2.0, which is a screener for functional impairment consisting of 12 items measured on a scale ranging from 1 to 5 (possible range of 12 to 60).

A total of 6 women from community settings were included for round 2 of usability testing.

Data Collection

The LHWs completed round 1 of usability testing through role plays following the completion of the training, using the mHealth App Usability Questionnaire (MUAQ) for standalone apps for health care providers [31]. The MUAQ comprises 18 items rated on a 7-point scale, with a higher score indicating greater usability. To assess the usability of TA-PM+, the total average score of the responses to all statements was calculated. With a maximum score of 7, a higher overall average score indicates superior app usability. Feedback was sought from LHWs, and TA-PM+ was improved with feedback gathered before round 2 of usability testing was conducted in the community with 6 participants. After recruiting participants, 6 trained LHWs delivered the first session of TA-PM+ to the positively screened individuals. After session delivery, the MUAQ [31] was administered to the LHWs, while the Patient Satisfaction Questionnaire (PSQ) involving an mHealth intervention [32] adapted for TA-PM+ was administered to the participants. The PSQ comprises a total of 14 items. On a 5-point Likert scale, respondents indicate the degree of agreement with each item. Ten items are presented in the positive, while 4 are stated in the negative. In the analysis, the negatively stated items are inverted; therefore, a higher score indicates increased satisfaction with the intervention. The highest possible score is 46. Another round of feedback was gathered from the LHWs and integrated into the final version of the TA-PM+ app.

Results

Overview

This section is divided into three parts describing the results of (1) qualitative investigation, (2) co-production, and (3) prototyping and evaluation.

Qualitative Investigation

Participant Characteristics

A total of 40 participants (mean age 49 years; range 40-60 years) were recruited, out of which 34 were female and 6 were male. Among the 40 participants, 30 completed 10 years of education, 2 completed 14 years of education, and 8 completed over 16 years of education.

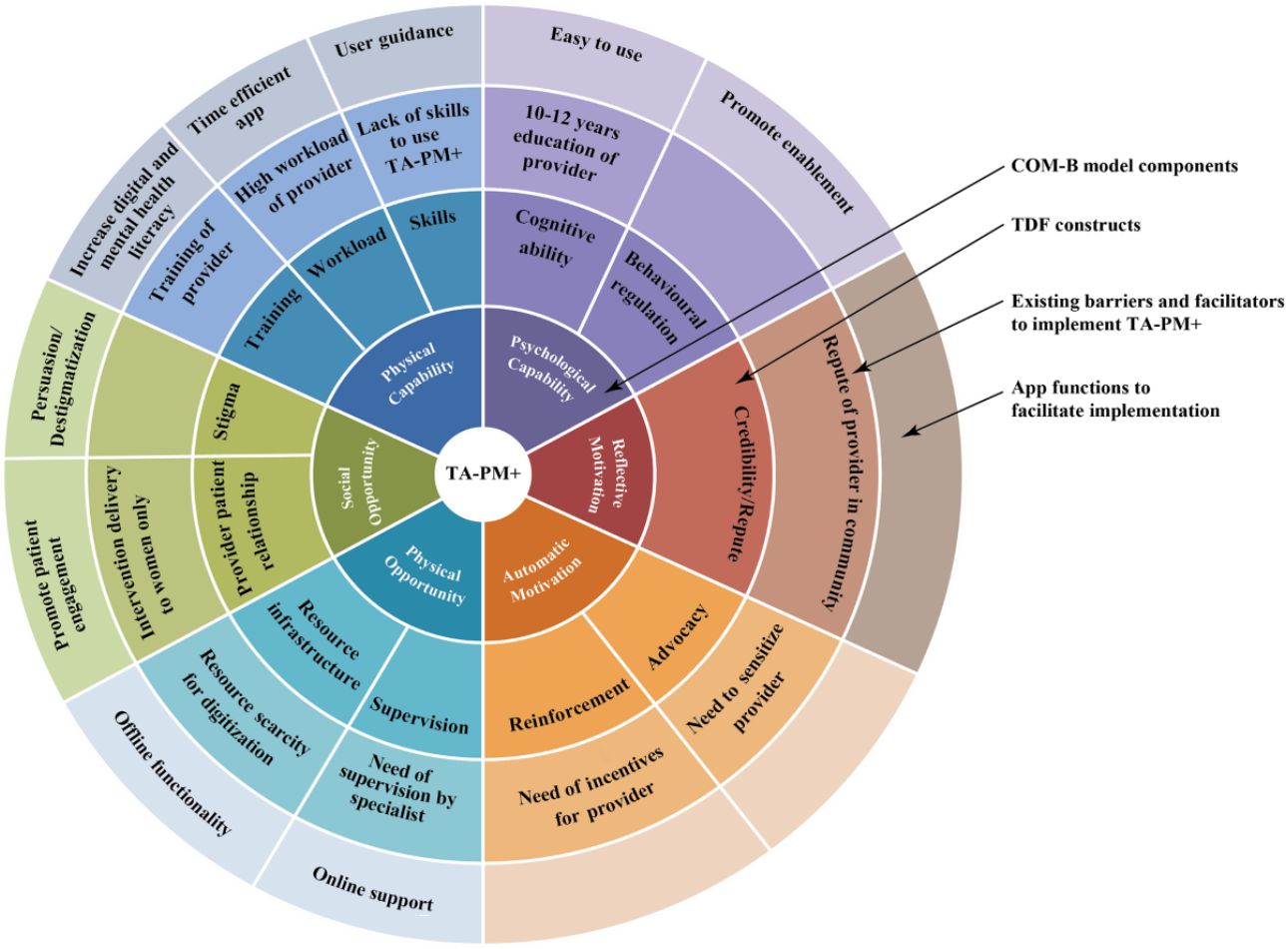
Perception of Factors Influencing the Implementation of TA-PM+ and the Features of TA-PM+ to Address These Factors Within the Capability, Opportunity, and Motivation-Behavior Model

Figure 1 provides an overview of the health system's barriers to the implementation of TA-PM+ and the TA-PM+ features

XSL•FO
RenderX

to facilitate implementation that were mapped under the constructs of the TDF and the components of the Capability, Opportunity, and Motivation for Behavioral Change (COM-B) model. The code chart with the description of codes is provided in Multimedia Appendix 3.

Figure 1. Perceived barriers and facilitators for the implementation of Technology-Assisted Problem Management Plus (TA-PM+) and the corresponding digital facilitators of TA-PM+ identified to address the barriers. Thematic analysis conducted using the Capability, Opportunity, and Motivation for Behavioral Change (COM-B) framework. TDF: Theoretical Domain Framework.



Barriers and Facilitators of the Health System Capacity to Implement TA-PM+ and the Features of TA-PM+ to Improve the Capacity Within the COM-B Framework

Knowledge and Skills

The significance of the knowledge and skill proficiency of LHWs in the effective implementation of TA-PM+ was highlighted by multiple participants. Although the digitization of the outpatient departments (OPDs) of LHWs has resulted in the development of some competencies to use digital tools, this aspect assumes utmost importance in the implementation process. Quotes have been provided, and the position of each quote in the transcript has been indicated.

> *There are some challenges for this intervention. Although the LHWs are given tablets, and their OPD is digitized…they receive proper digital training. But then again there might be problems regarding digital skills of LHWs to implement TA-PM+.* [Health policy participant_5, Position 10]

> *The challenges that can be faced include HR (Human Resources) and capacity issues.* [Health policy participant_3, Position 50]

Many participants perceived that in addition to training, the provision of user guidance and the incorporation of a fully automated guided app could effectively address the competency barrier.

> *When we open this app, it guides us from beginning till end. For example, when we are checking a child in a family, we open it all data comes and as we keep progressing, we are instructed what to do like its written 'check' and the name of the medicine, we click it, and it moves forward and tells us what to do next. This will make it easier for us.* [Focus group_1, LHW participant_6, Position 110]

Time/Workload of the Provider

All the participants agreed that a significant barrier to the implementation of TA-PM+ is the workload of LHWs. They were extensively involved in health promotion, disease prevention, and care delivery to women and children in their catchment area. The integration of a complex psychological intervention was perceived as an additional burden. Few respondents believed that the added workload associated with the additional digital tool complexity would outweigh the

XSL•FO
RenderX

assistance TA-PM+ is expected to provide in delivering the intervention.

> *We visit seven to eight homes daily, work on various aspects. We spend around 40 to 50 minutes in one household. In our app, which is dedicated to dengue, we send pictures. We also monitor the weight of children and assess the health of both mothers and children. When we encounter pregnant women, we provide them with cards and refer them accordingly. We give guidance, suggesting that they go to the center and obtain medication from there. In addition to all of this, we also record everything in our registers, and we enter data in the tablets.* [Focus group_3, LHW participant_1, Position 3]

The proposed design and strategies aimed at mitigating the strain experienced by LHWs were primarily focused on improving the time efficiency of TA-PM+. The participants strongly agreed on the importance of minimizing the amount of time spent using TA-PM+. Moreover, the participants emphasized the need for a simplified method that alleviates LHWs from the burden of typing. The ideal app, according to the participants, would be one that is fully automated and requires minimal input from LHWs. Almost all participants agreed to reduce session durations by leveraging the capabilities of the digital platform.

> *First, it should let them do everything with just a click. Why not create a form in such a way that these poor women don't need to type? So, I would suggest that you consider developing an app that is quick for them to use. The reason behind this suggestion is that these women are dealing with dengue, pneumonia, polio, health education, breast feeding, breast cancer, family planning, and post-abortion matters. So, think about how it could benefit them.* [Health system participant_6, Position 100]

Training

The importance of LHW training was acknowledged by the participants as a critical component of successful implementation. They emphasized the significance of comprehensive training covering the PM+ intervention as well as training to use TA-PM+. Some participants also highlighted the importance of field usability testing prior to implementation.

> *They should be guided on how to use it properly, and proper training and facilitation will be required. Train them for problem management plus and for the digital App. I think they shouldn't encounter any problems, but essentially, it's about sensitizing them first, then explaining the application thoroughly and conducting proper training. And, useful to conduct field test, then it should work well.* [Health policy participant_5, Position 16]

Regarding TA-PM+ functionalities, the participants felt that it should improve the TA-PM+ digital literacy of LHWs. Suggestions included the use of infographics or the provision of TA-PM+ introduction videos to serve as a refresher training for LHWs. This functionality was thought to be a helpful tool that LHWs could access at any time to address challenges in using the app.

> *Infographics should be designed in such a way that the LHWs can easily understand and get refresher training. For example, the LHW goes into the community, and sometimes someone asks her how to make ORS (Oral Rehydration Solution), and she doesn't remember how to make ORS. So, if she opens her learning platform and she can see how to make it. If there's a small infographic video uploaded, it will make it easier for community workers. Sometimes they may not remember everything, but with such resources, they can easily learn.* [Health policy participant_3, Position 56]

Cognitive Capacity

The cognitive capacity, including memory and attention, of LHWs has been identified as a crucial factor in the successful implementation of the TA-PM+ intervention. The majority of LHWs have 10 years of formal education. After training, LHWs need to remember and retain the information provided, pay attention to the needs and concerns of the individuals receiving the intervention, and actively engage with TA-PM+ to effectively deliver the intended support. Given the complexity of TA-PM+ and the cognitive demands associated with delivering it through a digital platform, it is crucial to develop an easy-to-use app that is not overly complex and employs plain language.

> *These poor women are not IT experts, they are all over 40, and while it seems easy, at this age, learning something new is challenging for them because they are accustomed to their old ways. So, I would suggest that you provide them with an app that is easy to download and use.* [Health system participant_6, Position 92]

Behavioral Regulation

One of the most important functionalities of TA-PM+ highlighted by LHWs was to promote enablement, which refers to the app's capacity to enhance precision in intervention delivery. To ensure that the intervention is implemented successfully, the TA-PM+ design must be optimized to promote and support the targeted actions.

> *At first, the name should appear, then their age, followed by their phone number, and then some other contact information. Also, session information like you've started a session, an entry has been made for that session or our visit data is recorded in it. The next session and date should be visible and whenever we click on patient name, the next session will come up.* [Focus group_3, LHW participant_6, Position 145]

XSL•FO
RenderX

Barriers and Facilitators for the Opportunity in Health Systems to Implement TA-PM+ and the Features of TA-PM+ to Increase the Opportunity Within the COM-B Framework

Resources/Infrastructure

The effective implementation of TA-PM+ relies on the availability of sufficient infrastructure and resources. The ongoing health reforms and digitization of health care in Pakistan provide a favorable environment for the implementation of a digital mental health intervention. However, the digitization of primary health systems is still in its early stages, and LHWs face numerous challenges. Despite the availability of digital tablets and the use of digital applications by LHWs for service delivery, concerns remain regarding device insurance, internet availability in different areas, and limited funds for internet access.

> *When you go to the community, sometimes it happens that you don't have access to the internet, and we cannot submit data. This internet, which is making life easy for people, has become a burden for us. It has become a source of trouble.* [Focus group_1, LHW participant_8, Position 254]

One of the key features considered important by all participants was offline functionality for TA-PM+. It was unanimously agreed that TA-PM+ should not rely on internet connectivity for its operation.

Monitoring, Supervision, and Support

The provision of supervision and continuous monitoring of LHWs was seen as crucial to the effective delivery of TA-PM+ to mental health patients. Furthermore, it is of utmost importance to ensure adequate support in case of any issues arising with the app.

The current model, which involves the supervision of LHWs by LHSs and referral from the community to primary health care facilities as needed, has been identified as a favorable approach for implementation. For TA-PM+ functionality, the online monitoring and referral features should be included to enhance the supervision process and ensure efficient communication between LHWs and health care facilities.

> *There should be a referral system in place where there is someone available who can assess if your patient is not benefiting from sessions and if their condition is severe, they should be referred through app. There should be a doctor available at the center, and there should be a clear referral process. It should also be made convenient so that if someone comes for help, it is beneficial for them, and there is no embarrassment for us.* [Focus group_3, LHW participant_2, Position 276]

Patient and Provider Relationship

TA-PM+ will be delivered by LHWs, and all participants perceived that the delivery can only be planned for women residing in the areas. Delivering the intervention to men is not possible given the focus of the LHW program on women's health. Furthermore, it was also perceived as a facilitator to implementation because LHWs have a positive reputation among the women of the community and the entire household has trust in them. This will facilitate the identification of cases of depression and stress by the LHW, as she will be fully informed of the circumstances a woman is experiencing.

> *When you visit someone's home, you get an idea of the internal situation there. You can estimate how they are doing, for instance... by observing their tone, silence, arguments, and fights. You can figure out that if a woman is distressed now, it might be because of her husband, or maybe because of her mother-in-law. In this case, they can obtain history and they can provide this lady intervention. And I think that it's not easy to go to a psychologist or psychiatrist for stress. But lady health worker can deal with this individual very comfortably.* [Health system participant_6, Position 108]

Participants indicated that features that improve patient engagement with the intervention were needed in TA-PM+ functionality. Features that are specifically tailored for patients and actively encourage and support patient engagement should be included in TA-PM+ to enhance the effectiveness of the intervention and further strengthen the patient-provider relationship.

> *We have worked on another digital project, there were videos for patients in that. We used to show them videos initially, and patient would understand quite well and used to be very convinced and focused.* [Focus group_3, LHW participant_2, Position 130]

Stigma

Stigma surrounding mental health disorders persists within the society in Pakistan, leading many individuals to perceive mental health treatment as stigmatized. Mental health services are generally provided at tertiary health hospitals, and psychiatric wards are locally known by stigmatized labels, reinforcing mental health's negative connotations. However, an alternative approach, that is, TA-PM+ provision in the community, specifically in the patient's home with LHWs, may reduce stigma and encourage treatment. This is because the community trusts these LHWs. One important component of TA-PM+ that has been identified is the use of persuasive strategies to reduce mental health stigma. In addition to promoting user engagement, TA-PM+ should also destigmatize mental health disorders and convince patients to accept support.

> *I really liked the feature of videos, especially for LHWs dealing with depression patients. This is very good. Additionally, there should be a face-to-face component in it. But overall, if you create something that includes engaging and persuading videos for patients, I think it would be very easily available to them all the time through the digital platform. I think it's good enough and efficient enough.* [Health policy participant_7, Position 11-12]

XSL•FO
RenderX

Barriers and Facilitators of Motivation for Health Systems to Implement TA-PM+ and the Features of TA-PM+ to Enhance Motivation Within the COM-B Framework

Credibility

Critical determinants of reflective motivation include credibility and reputation within the community. An LHW provides service to 150-200 households in the community where she resides. Improving her reputation in the community is directly linked to the provision of modern tools and expanded services, which leads to a sense of empowerment. It was perceived that her reputation and position in the community would be further enhanced by the implementation of the TA-PM+ program using a digital tablet to address depression and stress.

> *One thing to remember is that when we provide a tool to LHW, I give an example like this: if we introduce a new intervention that we are implementing throughout Pakistan, where they check blood pressure and refer patients with hypertension. When you provide them with a tool, their credibility in the community improves. In this, basically counseling, is included. But when you give them a digital tool, it motivates them and allows them to feel a bit empowered.* [Health policy participant_2, Position 42]

Co-production of TA-PM+

During the co-production of TA-PM+, the intervention development team reflected on the findings from stage 1 and decided on the key features of TA-PM+. Table 1 summarizes the TA-PM+ features planned during co-production workshops informed by the themes identified in the qualitative analyses of stage 1.

Table 1. Functionality of Technology-Assisted Problem Management Plus (TA-PM+) identified through a qualitative investigation and features developed in a co-production workshop.

TA-PM+[a] functionality identified in a qualitative investigation	TA-PM+ features developed in a co-production workshop
User guidance	• Various prompts and alerts in TA-PM+ to guide LHWs[b] and confirm completion of tasks • Instruction manual for TA-PM+
Easy and time efficient	• Basic consistent design across every session • Fully automated
Increase digital and mental literacy	• Two-minute informational videos for LHWs that provide information and outline the purpose of the session, its steps, the technique used, and how to use TA-PM+ to conduct the session • PDFs of the PM+[c] Urdu manual
Reduce cognitive load	• Fully automated PSYCHLOPS[d] displays questions in accordance with the scoring protocol • No typing required and audio recording available
Promote competence/enablement	• Personalized dashboard for each LHW, which includes patient details, name, address, session details, and next appointment date • Red, yellow, and green color codes to encourage regular appointments; the green code indicates that the session has just been delivered, the yellow code indicates that it is almost due, and the red code indicates that it is urgently due
Internet independent	• To enable use in places where internet access may be costly or where bandwidth is limited or unpredictable, the content's file size is kept as small as possible • All videos are stored in TA-PM+ and can be viewed without the internet
Supervision and monitoring	• Web-based supervisor backend; information is accessible regarding the session's date, duration, PSYCHLOPS responses, and score • Notification if the PSYCHLOPS score rises in subsequent sessions or suicidal ideation is noted in the patient • A locking feature that prevents session data from being altered once it is submitted • WhatsApp support group for urgent queries related to TA-PM+
Promote engagement	• Psychoeducation videos for patients that explain the physiological symptoms of anxiety and depression to the patients
Persuasion/destigmatization	• Psychoeducation videos for patients that include destigmatized vocabulary for symptoms • Psychoeducation videos for patients narrated by a female doctor to boost persuasiveness

[a]TA-PM+: Technology-Assisted Problem Management Plus.

[b]LHW: lady health worker.

[c]PM+: Problem Management Plus.

[d]PSYCHLOPS: Psychological Outcome Profiles.

XSL•FO
RenderX

TA-PM+ User Interface

TA-PM+ is an app that guides LHWs to deliver the PM+ intervention in a systematic manner. TA-PM+ features a personalized password-protected account and personal dashboard for each LHW (Figure 2), presenting essential patient details, session specifics, and upcoming appointment dates. Using a color-coded system, the app encourages the regular scheduling of appointments, indicating the urgency of upcoming sessions. By clicking on a patient name in the dashboard, the session home screen of the patient opens. TA-PM+ ensures the security and integrity of data through its locking features, preventing the alteration of submitted data.

TA-PM+ includes 5 informational videos, referred to as "Session Aid" (Figure 3), designed to provide LHWs with an outline of each session's purpose, the steps involved, and the techniques employed for each session. Moreover, the videos offer guidance on how to effectively use TA-PM+ to conduct these sessions. The videos have a duration of 2 minutes to promote utilization anytime uncertainty arises, without concern for schedule delays. Additionally, TA-PM+ provides access to the PM+ Urdu manual in PDF format.

To enhance understanding and engagement, and address mental health stigma, a series of psychoeducational videos (Figure 4) were developed and included in the session interface. These videos are narrated by a fictional character Dr Ayesha and employ a narrative approach, presenting the physiological symptoms of anxiety and depression in layperson's terms. The content is tailored to respect the cultural nuances and demographics of target communities and thus ensures maximum relevance. Using destigmatized vocabulary, the videos serve as a potent educational tool, addressing patient concerns and mitigating potential embarrassment or fear associated with their conditions. The primary objective is to develop trust and empathy, thereby promoting patient engagement and fostering acceptance of the sessions. Additionally, these videos play a vital role in efficiently explaining the active ingredients of PM+ sessions to the patients, empowering LHWs to establish context, and minimizing the time required to convey background information. The duration of each video is 4-5 minutes, and the content is designed to be concise and memorable for a diverse audience. The inclusion of a female doctor's narration adds a layer of persuasiveness, reinforcing and promoting mental health awareness. One video is developed for each session to deliver the active therapeutic ingredients for fidelity and ensure delivery at the right dose.

This design approach not only ensures the accurate delivery of active elements through patient-computer interaction, but also maintains the crucial patient-facilitator interaction for counseling and empathy throughout the intervention process. In addition, TA-PM+ enhances the counseling skills of LHWs and the therapeutic alliance between LHWs and patients.

Figure 2. User interface of the Technology-Assisted Problem Management Plus (TA-PM+) app developed to assist lady health workers in delivering mental health interventions within community settings in Pakistan. The first screen presents a dashboard with color-coded indicators (red, yellow, and green) to track and prioritize patient sessions. Subsequent screens illustrate session navigation, patient data entry, and automated PSYCHLOPS (Psychological Outcome Profiles) assessments in Urdu. Features have been co-designed for cultural relevance and usability among lady health workers in Pakistan.

XSL•FO
RenderX

Figure 3. Screenshots from session-aid videos developed for lady health workers to enhance digital literacy and support the delivery of Technology-Assisted Problem Management Plus (TA-PM+). These videos, designed in Urdu, provide step-by-step instructions for navigating the TA-PM+ app and conducting sessions effectively.

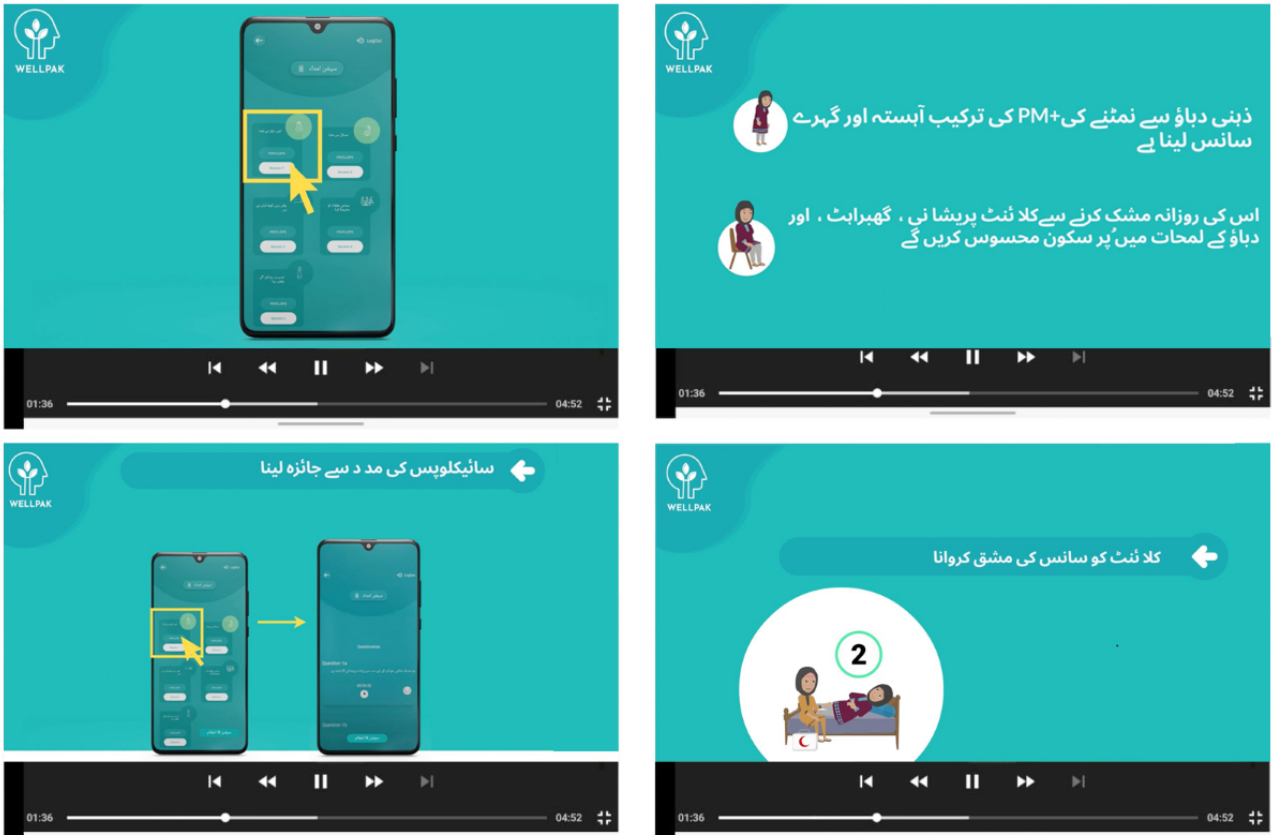
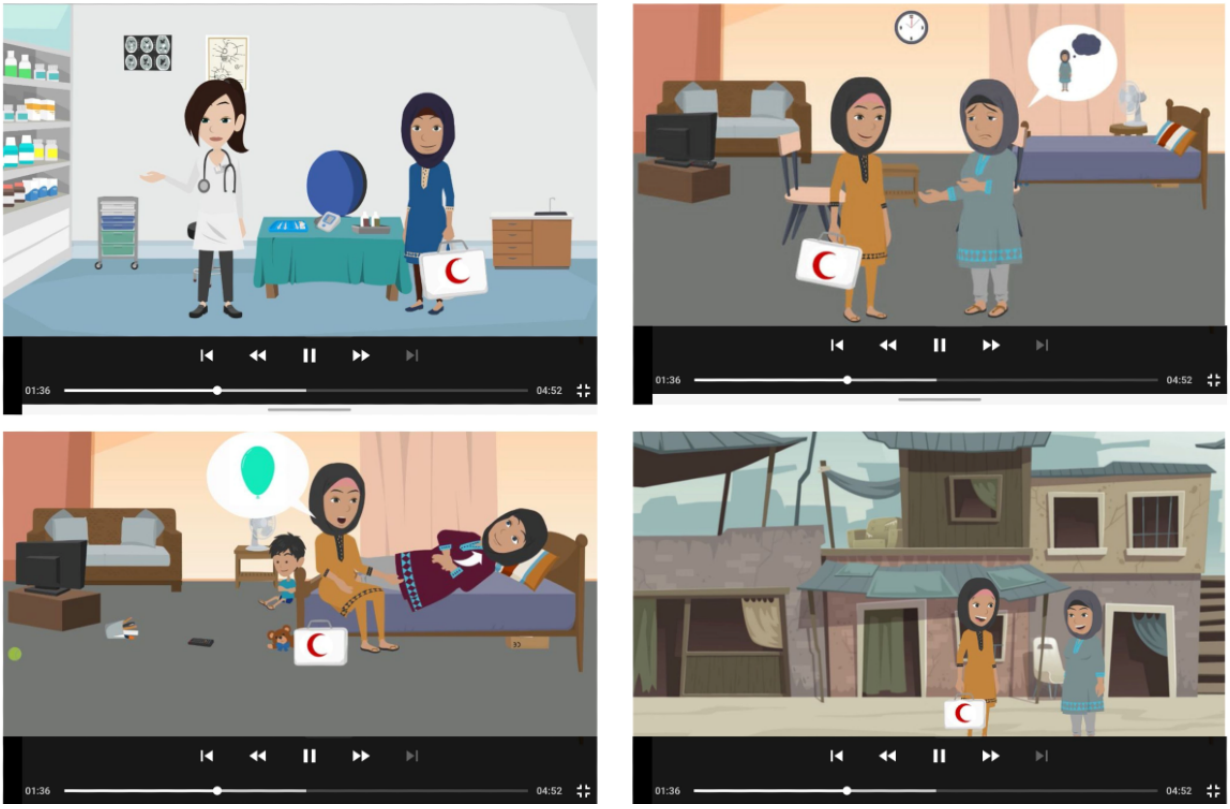


Figure 4. Screenshots from psychoeducational videos for patients, showing the characters Dr Ayesha, lady health worker Shazia, and patient Saima, which have been developed for sessions of Technology-Assisted Problem Management Plus (TA-PM+).



XSL•FO
RenderX

Web Dashboard for Supervisors

The online dashboard provides a user-friendly interface tailored for supervisors, streamlining the data from TA-PM+. This dashboard is structured to include an individual interface for each LHW, which in turn provides access to specific patient details. Supervisors can consistently monitor key information, such as session duration, delivery dates, and PSYCHLOPS scores. Moreover, the system incorporates notifications to alert supervisors in the event of a patient expressing suicidal thoughts, facilitating active case monitoring and prompt referral initiation.

Prototyping and Usability Testing

Employing the MUAQ designed for standalone mHealth apps used by health care providers, usability testing was performed twice, and it involved a cohort of 6 LHWs.

In the initial evaluation, 6 participating LHWs (mean age 44 years; range 41-58 years) were invited to fill the questionnaire and provide their feedback, which highlighted the issues of user interface, language, and pop-up alert settings.

The MUAQ item scores are shown in Table 2. After training, the mean usability score was recorded at 5.62 (highest possible score of 7). In the subsequent field assessment, the usability score increased to 5.96, which indicates improvement resulting from the incorporation of input from the first round. The usability scores demonstrated the system's overall usefulness.

Table 2 displays the average overall scores of the MUAQ according to the following 3 categories: ease of use, interface and satisfaction, and usefulness for delivering TA-PM+. After incorporating feedback from round 1 of usability testing into TA-PM+, scores of the categories ease of use and interface satisfaction increased in round 2. However, scores of the category usefulness for delivering PM+ did not change. This may be because comments were largely sought on the usability, interface, and functionality of TA-PM+ and not on its content, which was adapted from the PM+ manual.

In the subsequent assessment phase, field usability testing was conducted with 6 participants who were screened for depression using the GHQ-12 questionnaire and WHODAS. The 6 screened participants (mean age 29 years; range 25-37 years) were invited to fill the mHealth Satisfaction Questionnaire after TA-PM+ session delivery by LHWs. The TA-PM+ session client satisfaction score was 40 (highest possible score of 46) (Table 3), indicating that the patients expressed contentment with the session.

XSL•FO
RenderX

Table 2. Usability scores of the mHealth Usability App Questionnaire (MUAQ) for the Technology-Assisted Problem Management Plus (TA-PM+) app.

MUAQ[a] item	Score, mean (SD)	
	Round 1 (posttraining)	Round 2 (field)
Ease of use	4.33 (0.31)	5.13 (0.47)
The app was easy to use.	4.50 (1.97)	4.33 (2.65)
It was easy for me to learn to use the app.	5.00 (1.78)	5.83 (1.83)
The navigation was consistent when moving between screens.	3.00 (2.28)	5.16 (2.56)
The interface of the app allowed me to use all the functions (such as looking at the dashboard, filling the PSYCHLOPS[b], opening different sessions, and viewing videos) offered by the app.	4.16 (2.40)	4.00 (2.50)
Whenever I made a mistake using the app, I could recover easily and quickly.	5.00 (1.67)	6.33 (1.63)
Interface and satisfaction	6.01 (0.67)	6.21 (0.51)
I like the interface of the app.	5.16 (2.10)	5.83 (2.85)
The information in the app was well organized, so I could easily find the information I needed.	5.33 (1.86)	6.33 (1.63)
The app adequately acknowledged and provided information to let me know the progress of my action (delivered sessions).	6.33 (1.63)	6.33 (1.54)
I feel comfortable using this app while delivering the session to a patient.	5.83 (1.83)	6.33 (1.21)
The amount of time involved in using this app has been fitting for me.	6.33 (1.63)	6.33 (1.63)
I would use this app again.	7.00 (0.00)	6.33 (1.63)
Overall, I am satisfied with this app.	6.33 (1.63)	6.33 (1.63)
Usefulness	6.11 (0.89)	6.11 (0.61)
The app would be useful for my health care practice.	7.00 (0.00)	6.33 (1.63)
The app improved my access to delivering mental health service.	7.00 (0.00)	6.33 (1.63)
The app helped me manage my patients' mental health effectively.	6.33 (1.21)	6.16 (1.60)
This app has all the functions and capabilities I expected it to have.	6.00 (1.26)	6.50 (1.22)
I could use the app even when the internet connection was poor.	3.66 (2.65)	4.50 (2.25)
This mobile health app provides an acceptable way to deliver the PM+[c] intervention, such as accessing educational materials, tracking my own activities, and performing assessments.	6.66 (0.51)	6.83 (0.40)
Overall usability of the TA-PM+[d] app	5.62 (0.81)	5.96 (0.59)

[a]MUAQ: mHealth Usability App Questionnaire.

[b]PSYCHLOPS: Psychological Outcome Profiles.

[c]PM+: Problem Management Plus.

[d]TA-PM+: Technology-Assisted Problem Management Plus.

XSL•FO
RenderX

Table 3. Client satisfaction (mHealth Satisfaction Questionnaire) with the videos and sessions delivered by lady health workers using the Technology-Assisted Problem Management Plus (TA-PM+) app.

mHealth Satisfaction Questionnaire items	Score, mean (SD)
The video was helpful in receiving the sessions.	4.83 (0.40)
I enjoyed receiving the sessions through the app and videos.	5.00 (0.00)
The time spent receiving the session through the app was acceptable.	4.83 (0.40)
The language used in the videos was difficult to understand.	5.00 (0.00)
The information given in the video was in small chunks, which I was able to retain.	4.83 (0.40)
Videos were too time consuming.	2.33 (2.06)
I didn't like the videos or app usage during the session, and it felt like interruption.	1.66 (1.63)
It was so boring that I lost my interest during the sessions.	1.66 (0.81)
It was like a lecture, a disturbance, or too difficult to understand.	1.66 (0.00)
I can recommend it to others.	4.33 (1.03)
The avatars helped me to engage with the program. The scene settings were relatable, and it has motivated me to do the practices recommended for my well-being.	4.83 (1.02)
The examples in the videos helped me in understanding the information and helped me understand how the sessions will help my mental well-being.	4.83 (0.40)
The examples in the videos helped me understand how I need to implement session strategies in my life to feel better.	4.66 (0.51)
The videos and sessions have helped me set personal goals for my lifestyle habits in a way that I could not have done on my own.	4.83 (0.40)
Overall client satisfaction	40

Discussion

Principal Findings

The TA-PM+ app was designed to support LHWs in delivering the transdiagnostic PM+ intervention to patients with symptoms of stress, anxiety, and depression in Islamabad, Pakistan. The development process followed a thorough 3-stage methodology involving co-production with local stakeholders. Throughout the development of TA-PM+, careful consideration was given to the local health system perspective, available resources, and cultural sensitivities by involvement of experts. Usability testing confirmed that TA-PM+ is well-received by both LHWs and participants in community settings.

Urdu-translated PM+ was further adapted for technology-assisted delivery in our study using a comprehensive methodology involving local stakeholders and experts, a strategy recommended in the existing literature for PM+ adaptation. According to Heim et al [33], PM+ requires many adaptations to fit a given context before testing or implementing PM+ in that setting. Even changes to peripheral components of the intervention, such as instructional materials and illustrations, require the involvement of numerous experts in planning, translation, and cultural adaptation [34]. This procedural diligence is equally important when adapting the delivery style, for example, when PM+ is predominantly administered through web-based communication or when the length of treatment deviates from the originally specified 5 PM+ sessions [34].

The TA-PM+ app was developed through a co-production methodology, with the chosen delivery format being face-to-face interaction. The primary objective of the TA-PM+ app is to support LHWs in delivering the PM+ intervention in person, and patients only interact with the app to access psychoeducational videos. This particular design choice was carefully considered during the co-production to ensure a smooth integration with contextual elements and local digital and health resources, and the potential for a wider impact. Existing literature emphasizes that the success of digital technologies in mental health is dependent upon their alignment with sociopolitical conditions and cultural contexts. Failure to align with these factors often results in reduced user engagement and higher attrition rates. The term "nonalignment" encompasses various factors, such as values, mental health treatment methods, and specific health needs of the population [33,35].

The barriers to implementing TA-PM+ that were highlighted during the first round of qualitative investigation included skills and knowledge of LHWs to use TA-PM+, workload, time to implement TA-PM+, stigma associated with mental health, and lack of resources like internet availability. These barriers are in line with the current literature. A survey to explore barriers to implementation of digital mental health interventions in primary health care settings identified the availability of supporting infrastructure, organizational readiness and capacity, health care professional skills, and cultural appropriateness [36]. Research on digital mental health implementation [34] and digital health interventions through health systems [37] has shown that resource availability is a major barrier. The challenge posed by limited resource availability is particularly heightened in LMICs that are already struggling with resource constraints. A notable challenge in LMICs pertains to the availability of internet access and sufficient bandwidth. Addressing this barrier in digital health technologies is imperative, as highlighted both in our

XSL•FO
RenderX

study and in a concurrent investigation conducted in Nepal [38,39], advocating for the implementation of interventions that are independent of internet connectivity. One of the most frequent recommendations for TA-PM+ that emerged through the qualitative investigation was to shorten the sessions from 90 minutes to 45-60 minutes to increase feasibility. This is consistent with the existing research on the reasons for high attrition in PM+. Prior research has indicated that participants commonly had difficulties finishing all 5 sessions of PM+ because of the lengthy sessions, which frequently interfered with finding or maintaining a job, performing domestic duties, or raising children [8-10]. Top of Form

During the co-production workshop, customization of the content, especially tailoring of the psychoeducational videos to respect the cultural nuances and demographics of the target communities, was planned for TA-PM+, to address stigma and improve engagement. Likewise, recent reviews on engagement with digital mental health interventions [40,41] identified that tailored digital technologies with personalized videos and avatars increased user engagement and demonstrated effectiveness. To augment persuasiveness in TA-PM+, there was a strategic decision during co-production workshops to have the videos narrated by a fictional doctor avatar. This design feature of TA-PM+ is consistent with the self-help digital intervention recently developed by the WHO for depression, wherein all the videos feature a doctor as the narrator [19].

Limitations

There are limitations in this study, which have been divided into methodological limitations and design limitations of TA-PM+.

Methodological limitations include the potential lack of inclusivity and diversity in the co-production workshop sample, where people with lived experiences were not involved in the development of TA-PM+. Another limitation is the usability testing, which was conducted on a limited sample size and in the first session of TA-PM+, assuming interface uniformity across all sessions. The subsequent stage of our evaluation (feasibility study) will assess the feasibility, effectiveness, and usability of TA-PM+.

One of the main limitations in the design of TA-PM+ is that it requires the internet for the backend data storage function. Internet availability and bandwidth were identified as major barriers to the implementation of TA-PM+ in qualitative investigation. However, owing to backend storage of data and online monitoring, TA-PM+ cannot function without the internet, although some features requiring a high bandwidth, such as video playback, work without internet access. An additional limitation is that TA-PM+ will be administered by LHWs who provide services to women only. Thus, it will not be applicable to men, even though men require mental health care. TA-PM+ is currently in Urdu, necessitating technical and cultural adaptations to ensure its relevance and effectiveness in various sociocultural settings.

Conclusions

We successfully developed the TA-PM+ app for LHWs, with the potential to be integrated into primary health care to address common mental health conditions. This work has several implications even beyond the Pakistan setting. These include (1) an understanding of the process of developing a resource-appropriate digital health tool for high uptake; (2) a model for adaptation that is flexible and can be used to create similar tools for PM+ in other low-resource settings globally; and (3) the ability of the system to be expanded to function as a national tool to capture the prevalence of common mental health conditions at the population level.

Further feasibility testing and eventual up-scaling are envisaged based on the thorough development process outlined in this paper.

Acknowledgments

This work was supported by the German Federal Ministry of Education and Research (BMBF) grant 01KA2105A as part of the German Alliance for Global Health Research (GLOHRA).

Data Availability

All data generated and analyzed during this study are included in the main manuscript and its supplementary materials. The data are not deposited in a public repository due to the sensitive nature of participant information and ethical restrictions. Subject to ethical guidelines and institutional review board approval, the corresponding author can provide additional supporting data, such as comprehensive usability scores and satisfaction ratings, upon reasonable request.

Authors' Contributions

MS, SS, and HZ conceived and planned the study. MS and SZ conducted the qualitative interviews and analyses. MS, SZ, MK, TK, AB, DCF, SS, and HZ were involved in the co-production workshops and the planning of the TA-PM+ features. AB developed the software for TA-PM+. MS and SZ trained the lady health workers and conducted usability testing. MS and DCF wrote the manuscript. All authors commented on the manuscript.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
COREQ (COnsolidated criteria for REporting Qualitative research) Checklist.
[PDF File (Adobe PDF File), 403 KB - formative_v9i1e59414_app1.pdf]

Multimedia Appendix 2
Interview guides.
[DOCX File , 30 KB - formative_v9i1e59414_app2.docx]

Multimedia Appendix 3
Code chart.
[DOCX File , 21 KB - formative_v9i1e59414_app3.docx]

References

1. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry 2022 Feb;9(2):137-150 [FREE Full text] [doi: 10.1016/S2215-0366(21)00395-3] [Medline: 35026139]
2. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5. Washington, DC: American Psychiatric Publishing, Inc; 2013.
3. Depression and Other Common Mental Disorders: Global Health Estimates. World Health Organization. 2017. URL: https://iris.who.int/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf [accessed 2024-12-30]
4. WHO and Ministry of Health Pakistan. WHO-AIMS Report on Mental Health System in Pakistan. World Health Organization. 2009. URL: https://extranet.who.int/mindbank/item/1305 [accessed 2024-12-30]
5. Saxena S, Funk M, Chisholm D. World Health Assembly adopts Comprehensive Mental Health Action Plan 2013-2020. Lancet 2013 Jun 08;381(9882):1970-1971 [FREE Full text] [doi: 10.1016/S0140-6736(13)61139-3] [Medline: 23746771]
6. Verhey IJ, Ryan GK, Scherer N, Magidson JF. Implementation outcomes of cognitive behavioural therapy delivered by non-specialists for common mental disorders and substance-use disorders in low- and middle-income countries: a systematic review. Int J Ment Health Syst 2020;14:40 [FREE Full text] [doi: 10.1186/s13033-020-00372-9] [Medline: 32514304]
7. Keynejad RC, Spagnolo J, Thornicroft G. Mental healthcare in primary and community-based settings: evidence beyond the WHO Mental Health Gap Action Programme (mhGAP) Intervention Guide. Evid Based Ment Health 2022 Dec;25(e1):e1-e7 [FREE Full text] [doi: 10.1136/ebmental-2021-300401] [Medline: 35473750]
8. Acarturk C, Uygun E, Ilkkursun Z, Yurtbakan T, Kurt G, Adam-Troian J, et al. Group problem management plus (PM+) to decrease psychological distress among Syrian refugees in Turkey: a pilot randomised controlled trial. BMC Psychiatry 2022 Jan 04;22(1):8 [FREE Full text] [doi: 10.1186/s12888-021-03645-w] [Medline: 34983461]
9. de Graaff AM, Cuijpers P, McDaid D, Park A, Woodward A, Bryant RA, et al. Peer-provided Problem Management Plus (PM+) for adult Syrian refugees: a pilot randomised controlled trial on effectiveness and cost-effectiveness. Epidemiol Psychiatr Sci 2020 Aug 18;29:e162. [doi: 10.1017/s2045796020000724]
10. Khan MN, Hamdani SU, Chiumento A, Dawson K, Bryant RA, Sijbrandij M, et al. Evaluating feasibility and acceptability of a group WHO trans-diagnostic intervention for women with common mental disorders in rural Pakistan: a cluster randomised controlled feasibility trial. Epidemiol Psychiatr Sci 2019 Feb;28(1):77-87 [FREE Full text] [doi: 10.1017/S2045796017000336] [Medline: 28689511]
11. Maercker A, Heim E, Kirmayer LJ. Cultural Clinical Psychology and PTSD. Göttingen, Germany: Hogrefe Publishing GmbH; 2019.
12. Hamdani SU, Huma Z, Rahman A, Wang D, Chen T, van Ommeren M, et al. Cost-effectiveness of WHO Problem Management Plus for adults with mood and anxiety disorders in a post-conflict area of Pakistan: randomised controlled trial. Br J Psychiatry 2020 Nov 28;217(5):623-629. [doi: 10.1192/bjp.2020.138] [Medline: 32720628]
13. Rahman A, Riaz N, Dawson KS, Usman Hamdani S, Chiumento A, Sijbrandij M, et al. Problem Management Plus (PM+): pilot trial of a WHO transdiagnostic psychological intervention in conflict-affected Pakistan. World Psychiatry 2016 Jun;15(2):182-183 [FREE Full text] [doi: 10.1002/wps.20312] [Medline: 27265713]
14. Jordans MJD, Kohrt BA, Sangraula M, Turner EL, Wang X, Shrestha P, et al. Effectiveness of Group Problem Management Plus, a brief psychological intervention for adults affected by humanitarian disasters in Nepal: A cluster randomized controlled trial. PLoS Med 2021 Jun;18(6):e1003621 [FREE Full text] [doi: 10.1371/journal.pmed.1003621] [Medline: 34138875]
15. Dawson KS, Bryant RA, Harper M, Kuowei Tay A, Rahman A, Schafer A, et al. Problem Management Plus (PM+): a WHO transdiagnostic psychological intervention for common mental health problems. World Psychiatry 2015 Oct 25;14(3):354-357 [FREE Full text] [doi: 10.1002/wps.20255] [Medline: 26407793]
16. Schäfer S, Thomas LM, Lindner S, Lieb K. World Health Organization's low-intensity psychosocial interventions: a systematic review and meta-analysis of the effects of Problem Management Plus and Step-by-Step. World Psychiatry 2023 Oct 15;22(3):449-462 [FREE Full text] [doi: 10.1002/wps.21129] [Medline: 37713578]

XSL•FO
RenderX

17. Chambers DA, Glasgow RE, Stange KC. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. Implement Sci 2013 Oct 02;8:117 [FREE Full text] [doi: 10.1186/1748-5908-8-117] [Medline: 24088228]
18. Mudiyanselage KWW, De Santis KK, Jörg F, Saleem M, Stewart R, Zeeb H, et al. The effectiveness of mental health interventions involving non-specialists and digital technology in low-and middle-income countries - a systematic review. BMC Public Health 2024 Jan 03;24(1):77 [FREE Full text] [doi: 10.1186/s12889-023-17417-6] [Medline: 38172713]
19. Carswell K, Harper-Shehadeh M, Watts S, Van't Hof E, Abi Ramia J, Heim E, et al. Step-by-Step: a new WHO digital mental health intervention for depression. Mhealth 2018;4:34 [FREE Full text] [doi: 10.21037/mhealth.2018.08.01] [Medline: 30225240]
20. McBride K, Harrison S, Mahata S, Pfeffer K, Cardamone F, Ngigi T, et al. Building mental health and psychosocial support capacity during a pandemic: the process of adapting problem management plus for remote training and implementation during COVID-19 in New York City, Europe and East Africa. Intervention 2021;19(1):37. [doi: 10.4103/INTV.INTV_30_20]
21. Rivero-Santana A, Perestelo-Perez L, Alvarez-Perez Y, Ramos-Garcia V, Duarte-Diaz A, Linertova R, PECUNIA Group. Stepped care for the treatment of depression: a systematic review and meta-analysis. J Affect Disord 2021 Nov 01;294:391-409 [FREE Full text] [doi: 10.1016/j.jad.2021.07.008] [Medline: 34315101]
22. Hussain SS, Khan M, Gul R, Asad N. Integration of mental health into primary healthcare: perceptions of stakeholders in Pakistan. East Mediterr Health J 2018 May 03;24(2):146-153 [FREE Full text] [Medline: 29748943]
23. Hafeez A, Mohamud B, Shiekh M, Shah S, Jooma R. Lady health workers programme in Pakistan: Challenges, achievements and the way forward. J Pak Med Assoc 2011 Mar 1;61(3):210-215 [FREE Full text]
24. Population Census. Pakistan Bureau of Statistics. URL: http://www.pbs.gov.pk/content/population-census [accessed 2024-12-30]
25. Hawkins J, Madden K, Fletcher A, Midgley L, Grant A, Cox G, et al. Development of a framework for the co-production and prototyping of public health interventions. BMC Public Health 2017 Sep 04;17(1):689 [FREE Full text] [doi: 10.1186/s12889-017-4695-8] [Medline: 28870192]
26. Atkins L, Francis J, Islam R, O'Connor D, Patey A, Ivers N, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implement Sci 2017 Jun 21;12(1):77 [FREE Full text] [doi: 10.1186/s13012-017-0605-9] [Medline: 28637486]
27. Thomas DR. Feedback from research participants: are member checks useful in qualitative research? Qualitative Research in Psychology 2016 Aug 02;14(1):23-41. [doi: 10.1080/14780887.2016.1219435]
28. Williams L. Agile software development methodologies and practices. Advances in Computers 2010;80:1-44. [doi: 10.1016/S0065-2458(10)80001-4]
29. Joyce G, Lilley M. Towards the Development of Usability Heuristics for Native Smartphone Mobile Applications. In: Marcus A, editor. Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience. DUXU 2014. Lecture Notes in Computer Science, vol 8517. Cham: Springer; 2014:465-474.
30. Nielsen J. How Many Test Users in a Usability Study? NN/g. URL: https://www.nngroup.com/articles/how-many-test-users/ [accessed 2024-12-30]
31. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): Development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
32. Melin J, Bonn SE, Pendrill L, Trolle Lagerros Y. A questionnaire for assessing user satisfaction with mobile health apps: development using Rasch Measurement Theory. JMIR Mhealth Uhealth 2020 May 26;8(5):e15909 [FREE Full text] [doi: 10.2196/15909] [Medline: 32452817]
33. Heim E, Kohrt BA. Cultural adaptation of scalable psychological interventions: a new conceptual framework. CPE 2019 Dec 17;1(4):37679. [doi: 10.32872/cpe.v1i4.37679]
34. Williams MG, Stott R, Bromwich N, Oblak SK, Espie CA, Rose JB. Determinants of and barriers to adoption of digital therapeutics for mental health at scale in the NHS. BMJ Innov 2020 May 07;6(3):92-98. [doi: 10.1136/bmjinnov-2019-000384]
35. Connelly K, Stein KF, Chaudry B, Trabold N. Development of an ecological momentary assessment mobile app for a low-literacy, Mexican American population to collect disordered eating behaviors. JMIR Public Health Surveill 2016 Jul 14;2(2):e31 [FREE Full text] [doi: 10.2196/publichealth.5511] [Medline: 27418020]
36. Brantnell A, Temiz S, Baraldi E, Woodford J, von Essen L. Barriers to and facilitators of the implementation of digital mental health interventions as perceived by primary care decision makers: content analysis of structured open-ended survey data. JMIR Hum Factors 2023 Jun 26;10:e44688 [FREE Full text] [doi: 10.2196/44688] [Medline: 37358902]
37. Neher M, Nygårdh A, Broström A, Lundgren J, Johansson P. Perspectives of policy makers and service users concerning the implementation of eHealth in Sweden: interview study. J Med Internet Res 2022 Jan 28;24(1):e28870 [FREE Full text] [doi: 10.2196/28870] [Medline: 35089139]
38. Pokhrel P, Karmacharya R, Taylor Salisbury T, Carswell K, Kohrt BA, Jordans MJD, et al. Perception of healthcare workers on mobile app-based clinical guideline for the detection and treatment of mental health problems in primary care: a qualitative study in Nepal. BMC Med Inform Decis Mak 2021 Jan 19;21(1):21 [FREE Full text] [doi: 10.1186/s12911-021-01386-0] [Medline: 33468120]

XSL•FO
RenderX

39. Luitel N, Neupane V, Lamichhane B, Koirala G, Gautam K, Karki E, et al. Experience of primary healthcare workers in using the mobile app-based WHO mhGAP intervention guide in detection and treatment of people with mental disorders: A qualitative study in Nepal. SSM Ment Health 2023 Dec 15;4:100278 [FREE Full text] [doi: 10.1016/j.ssmmh.2023.100278] [Medline: 38155812]
40. Saleem M, Kühne L, De Santis KK, Christianson L, Brand T, Busse H. Understanding engagement strategies in digital interventions for mental health promotion: scoping review. JMIR Ment Health 2021 Dec 20;8(12):e30000 [FREE Full text] [doi: 10.2196/30000] [Medline: 34931995]
41. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from Critical Interpretive Synthesis. Telemed J E Health 2012 Mar;18(2):137-144 [FREE Full text] [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]

Abbreviations

COM-B: Capability, Opportunity, and Motivation for Behavioral Change
GHQ-12: General Health Questionnaire-12
LHS: lady health supervisor
LHW: lady health worker
LMIC: low- and middle-income country
mhGAP: Mental Health Gap Action Programme
MUAQ: mHealth App Usability Questionnaire
OPD: outpatient department
PM+: Problem Management Plus
PSQ: Patient Satisfaction Questionnaire
TA-PM+: Technology-Assisted Problem Management Plus
TDF: Theoretical Domain Framework
UC: Union Council
WHO: World Health Organization
WHODAS: WHO Disability Assessment Schedule

Edited by A Mavragani; submitted 11.04.24; peer-reviewed by W Zhang; comments to author 16.10.24; revised version received 06.11.24; accepted 02.12.24; published 28.01.25.

Please cite as:
Saleem M, Zafar S, Klein T, Koesters M, Bashir A, Fuhr DC, Sikander S, Zeeb H
A Digital Tool (Technology-Assisted Problem Management Plus) for Lay Health Workers to Address Common Mental Health Disorders: Co-production and Usability Study in Pakistan
JMIR Form Res 2025;9:e59414
URL: https://formative.jmir.org/2025/1/e59414
doi:10.2196/59414
PMID:39874072

©Maham Saleem, Shamsa Zafar, Thomas Klein, Markus Koesters, Adnan Bashir, Daniela C Fuhr, Siham Sikander, Hajo Zeeb. Originally published in JMIR Formative Research (https://formative.jmir.org), 28.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Technological Adjuncts to Streamline Patient Recruitment, Informed Consent, and Data Management Processes in Clinical Research: Observational Study

Jodie Koh[1], MS; Stacey Caron[2*], BSN; Amber N Watters[2,3*], MD, MS; Mahesh Vaidyanathan[2,3*], MD; David Melnick[2*], BS; Alyssa Santi[3*], BSN; Kenneth Hudson[2,3*], BS, MS; Catherine Arguelles[3*], BSN; Priyanka Mathur[4,5*], MD; Mozziyar Etemadi[2,3*], MD, PhD

[1]Kellogg School of Management, Northwestern University, Evanston, IL, United States

[2]Northwestern Medicine, Chicago, IL, United States

[3]Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[4]Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Boston, MA, United States

[5]Department of Obstetrics and Gynecology, Massachusetts General Hospital, Boston, MA, United States

[*]these authors contributed equally

Corresponding Author:
Jodie Koh, MS
Kellogg School of Management
Northwestern University
2211 Campus Drive
Evanston, IL, 60208
United States
Phone: 1 847 491 3300
Email: jodie.koh@kellogg.northwestern.edu

Abstract

Background: Patient recruitment and data management are laborious, resource-intensive aspects of clinical research that often dictate whether the successful completion of studies is possible. Technological advances present opportunities for streamlining these processes, thus improving completion rates for clinical research studies.

Objective: This paper aims to demonstrate how technological adjuncts can enhance clinical research processes via automation and digital integration.

Methods: Using one clinical research study as an example, we highlighted the use of technological adjuncts to automate and streamline research processes across various digital platforms, including a centralized database of electronic medical records (enterprise data warehouse [EDW]); a clinical research data management tool (REDCap [Research Electronic Data Capture]); and a locally managed, Health Insurance Portability and Accountability Act–compliant server. Eligible participants were identified through automated queries in the EDW, after which they received personalized email invitations with digital consent forms. After digital consent, patient data were transferred to a single Health Insurance Portability and Accountability Act–compliant server where each participant was assigned a unique QR code to facilitate data collection and integration. After the research study visit, data obtained were associated with existing electronic medical record data for each participant via a QR code system that collated participant consent, imaging data, and associated clinical data according to a unique examination ID.

Results: Over a 19-month period, automated EDW queries identified 20,988 eligible patients, and 10,582 patients received personalized email invitations. In total, 1000 (9.45%) patients signed consents to participate in the study. Of the consented patients, 549 unique patients completed 779 study visits; some patients consented to the study at more than 1 time period during their pregnancy.

Conclusions: Technological adjuncts in clinical research decrease human labor while increasing participant reach and minimizing disruptions to clinic operations. Automating portions of the clinical research process benefits clinical research efforts by expanding and optimizing participant reach while reducing the limitations of labor and time in completing research studies.

(JMIR Form Res 2025;9:e58628) doi:10.2196/58628

XSL•FO
RenderX

KEYWORDS

digital health; patient recruitment; consent; technological adjuncts; data management; clinical research processes; automation; digital platforms; data warehouse; patient data; imaging data; pregnancy; clinical research methods

Introduction

Clinical research processes, specifically in patient recruitment, are rife with logistical complexities that often determine the success or failure of a clinical trial [1]. Successful participant enrollment depends on organization, personnel, and participant factors as well as the features of the clinical trial at hand [2-5]. McDonald et al [6] cited reasons for research study failure as including heavy clinical and research workload of research team members, perceived imbalance between patient incentive and risk, and low status typically conferred to recruitment work. Of 114 studies in the review of McDonald et al [6], only 31% met their original recruitment goals, while 53% required extended periods for completion.

The resource-intensiveness of recruitment activities often dictates whether studies are completed successfully [7]. Recruitment processes have traditionally relied heavily on trained manual labor [8,9], which is costly. Moreover, these research personnel often face logistic barriers when recruiting in high-volume clinical settings, where their research activities may be restricted by clinical workflows and tight patient turnaround [10]. In general, in-person recruiting is personnel-intensive, as it requires research staff to repetitively explain the research study to each potential participant accurately and representatively [11,12]. Finally, traditional methods for recruiting patients rely on collecting and filing printed paperwork, especially in managing patient records and study results, which oftentimes require a high cognitive workload to manage and safeguard [2].

Following these observations, health care and federal institutions have increasingly demonstrated support for the integration of digital technological capabilities within clinical research [13-15], especially leveraging the capabilities brought about by automating clinical processes [16,17]. In 2019, the National Heart, Lung, and Blood Institute; National Institutes of Health; and National Science Foundation hosted a workshop calling for the digitization of clinical research using advanced analytics to facilitate patient screening and data management to increase the representation of diverse patient populations [18]. In a separate review of recruitment methods across 61 clinical trials, digital recruitment was found to be 52% more effective than conventional offline recruitment [19].

Such technological advancements present opportunities for overcoming the onerous complexities of clinical research workflows while also reducing the operational costs of research (by reducing research staff labor-intensiveness and expanding potential participant reach) as well as the potential for human error in recruitment processes (by digitizing data management). In this paper, we share our experience using technological adjuncts to streamline participant identification, recruitment, consent, and data management processes for an observational clinical research study. These strategies allowed us to complete our study recruitment goal ahead of schedule and in a cost-effective, resource-optimized way.

Methods

Ethical Considerations

All elements of the research study—which, in this portion, involved patient recruitment to achieve the goals of our broader study—were approved by Northwestern University's Institutional Review Board Office (study ID: STU00215717). This approval confirms that all aspects of the study meet the ethical standards of the review board. Invited patients were asked to provide consent if they agreed to participate in our study and were also given the option to opt out of the study with no consequences to the care provided to them. Patients were not financially compensated for their participation in our study, and all data collected were subsequently deidentified.

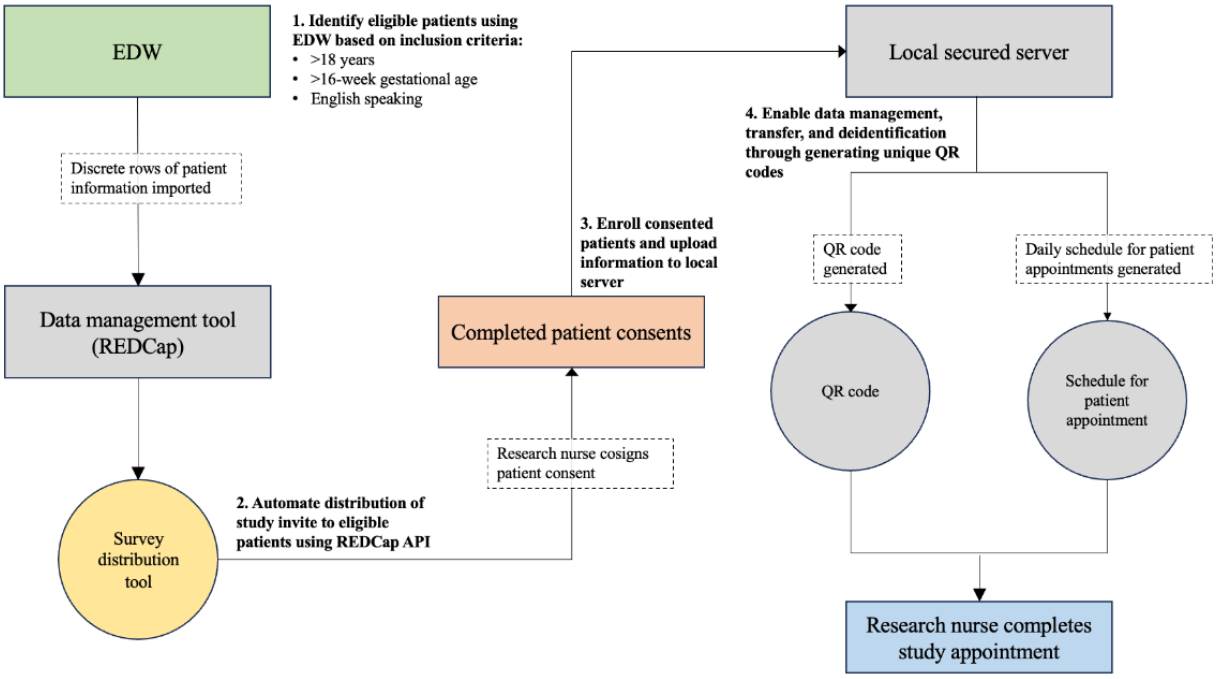
Purpose of Broader Research Study

The digitized recruitment processes that we describe in this paper are part of a prospective, observational research study. The research study's primary aim is to develop a database of ultrasonography images for use in the development of artificial intelligence technology with specific applications to perinatal care. To fulfill this goal, we planned to enroll 1000 patients across various time periods of pregnancy from within a single institution. To accomplish this task, we designed and executed a digitized process of patient recruitment and data management.

Given our ultimate goal to build algorithms capable of providing diagnostic recommendations and predictions of various maternal and fetal health outcomes, we engaged in strategic and targeted patient recruitment to ensure that we recruited an appropriate number of patients with specific health conditions across various outcomes while abiding by constraints including a temporally bounded funding structure. Our clinical research processes relied on 3 main technological adjuncts: a secured, centralized database of electronic medical records (enterprise data warehouse [EDW]), a research data management tool (REDCap [Research Electronic Data Capture]; Vanderbilt University), and a locally managed, Health Insurance Portability and Accountability Act–compliant QR code–based system that allowed us to coordinate all patient information, data, and enrollment-related documentation on a single, secured server. An overview of our clinical research processes is depicted in Figure 1. Each step of these processes is detailed below.

XSL•FO
RenderX

Figure 1. Workflow for clinical research processes across digital adjuncts. API: application programming interface; EDW: enterprise data warehouse; REDCap: Research Electronic Data Capture.



Identify Eligible Patients Using EDW Based on Inclusion Criteria

The patient recruitment process leveraged access to patient medical records collated from Epic and stored in a local EDW to identify eligible patients for the study. In doing so, we performed a SQL query, filtering for patients based on the inclusion and exclusion criteria and limited to participating clinical units. The inclusion criteria were patients aged 18 years and older, English speaking, and at least 16 weeks of gestational age. This query derived multiple rows of eligible patients to be invited to participate in our study. We set up this query function as an automated process with the list of eligible patients refreshed every 24 hours. This query generated a list of patient appointments that had been scheduled within the next 30 days across participating clinics.

Automate Distribution of Study Invite to Eligible Patients Using the REDCap Application Programming Interface

We used a REDCap-specific application programming interface to import the rows of eligible patient appointments generated from the EDW into the REDCap database. Each discrete row provided all necessary details for patient identification and recruitment including patient medical record numbers, personal and clinical information, contact information, and details including patient notes on upcoming appointments. Patients provide universal consent when first receiving care at our institution, thereby granting access to contact information for use in recruitment for research studies.

Within REDCap, we conducted an additional manual check on the list of eligible patients against our inclusion criteria using a filtering function. This was done to mitigate any errors that may have been made as a result of data compatibility issues during transfer between platforms. Given that patients are exclusively electronically recruited for this study, we also filtered patients based on the presence of an email address on file. Finally, we filtered to exclude patients who previously indicated disinterest in participating in this study, resulting in a narrowed list of eligible patients to recruit.

We used REDCap's survey distribution tool function to distribute invitations to eligible patients. We sent personalized emails with patient names as designated in their charts followed by all relevant information pertaining to the study including the title and purpose of clinical research, risks and benefits of participation, participant expectations for the study, and contact information for any inquiries. Patients were provided unique links that directed them to their personalized electronic consent forms embedded in REDCap. The email invitation text may be reviewed in Multimedia Appendix 1.

Enroll Consented Patients and Upload Information to Local Server

As the research nurse on the team cosigned patient consent forms, using a second SQL query, we uploaded the list of consented patients onto a local server managed by the project team. On the server, expired patient appointments were removed to optimize memory space. With the list of consented patients, we deployed a third SQL script that generated a list of patients who had an appointment in the clinical units each day. This script also generated a unique QR code for each patient to associate the collected data with the patient's medical records on the local server and in REDCap.

With the list of consented patients, the research nurse scheduled data collection appointments by prioritizing patients with outcomes of interest to our study, and also inconsideration of

XSL•FO
RenderX

where each patient was located and the providers' preference for data collection should be carried out. Scheduling accounted for travel time between appointment rooms and clinical units and the time taken to prepare for and wrap up data collection. This routine enabled the research nurse to carry data collection in multiple clinical units in a single day.

Enable Data Management, Transfer, and Deidentification Through Generating Unique QR Codes

As is standard for most data collection studies, the study workflow must culminate in generating a fully deidentified database. Alongside generating the QR code, a unique research examination ID that is not present in the patient's medical record is generated as a function for identifying collected data without associating it with any protected health information (PHI).

At the beginning of each data collection appointment, a research nurse would scan the unique QR code that would route the collected ultrasound scans to the appropriate folder on the local server to be stored. The folders on the server are labeled using the unique examination ID. This system allows the QR code to serve as a means for identification and reference as we link deidentified patient files with the associated ultrasound scans. Aside from a single "master list" linking the patient's "real" PHI information to their study information (as is standard in data collection studies), routing and parsing relevant data points are fully automated and rely on research-specific identifiers (unique examination ID) and QR codes, thus reducing the risk of study documentation being lost or otherwise mishandled. This contrasts current, manual approaches that oftentimes rely on names or date of births written on documents to associate data to a patient, which subsequently must be shredded and disposed of.

Results

Between December 1, 2021, and June 13, 2023 (19 months), we achieved our study goal of enrolling 1000 patients for the study. Our automated EDW queries identified 20,988 eligible patient appointments across 3 obstetric clinics throughout that time period. Each query was automatically run every 24 hours and generated between 1000 and 2000 eligible patient appointments.

Logistic constraints including limited study personnel (1 research nurse on the project) and limited work schedule availability to collect data resulted in only approximately 50.4% (10,582/20,988) of identified eligible patients receiving an invitation to participate in the study. We also restricted invites to 2 or 3 potential patients with overlapping appointment times across all clinical sites. Data collection appointments were only scheduled on weekdays between 8 AM and 4 PM.

Based on these constraints, personalized email invitations were subsequently issued to 10,582 potential participants. We enrolled 1000 patients, yielding a 9.5% enrollment rate over 19 months. Among those who were enrolled in the study, 779 patients completed the research study visit within the reported time frame. The difference between enrolled patients and those with complete data collection is a result of some consented patients not having completed their study appointments within our reported time frame. This is largely attributed to the rescheduling of data collection to a subsequent appointment to accommodate the time or logistic constraints of our 1 research nurse performing study data collection. During the study interval reported here, a total of 549 unique individual patients participated in the study; some patients participated more than once during their pregnancy.

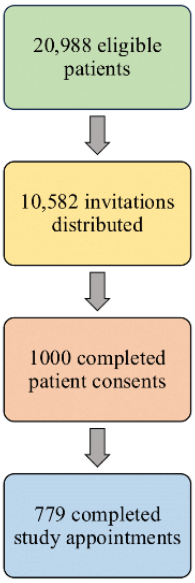
Table 1 presents the descriptive demographics of participating patients for race, ethnicity, age, gestational age, and insurance payor. The patient recruitment and enrollment process are summarized in Figure 2.

XSL•FO
RenderX

Table 1. Demographic data of the consented patient population at the time of study enrollment.

	Study population (raw), n (%)	Institution population (raw), n (%)
Race		
Asian	44 (8.0)	3958 (7.2)
Black	74 (13.5)	5452 (9.9)
White	284 (51.7)	35,930 (65.1)
Other	40 (7.3)	3363 (6.1)
Not specified	107 (19.5)	8028 (14.5)
Ethnicity		
Hispanic	80 (14.6)	11,403 (20.6)
Non-Hispanic	396 (72.1)	40,665 (73.6)
Not specified	73 (13.3)	3160 (5.7)
Age group (years)		
<20	4 (0.7)	1309 (2.4)
20-30	101 (18.4)	16,770 (30.4)
30-40	404 (73.6)	34,474 (62.4)
>40	40 (7.3)	2675 (4.8)
Gestational age (years)		
<16	2 (0.3)	794 (2.8)
16-20	167 (16.7)	3781 (18.2)
21-25	163 (16.3)	2666 (12.9)
26-30	196 (19.6)	3120 (15)
31-35	315 (31.5)	5502 (26.5)
36-40	157 (15.7)	4693 (22.6)
>40	0 (0)	183 (0.8)
Insurance payor		
Public	126 (23)	1286 (33.1)
Private	423 (77)	2595 (66.9)

Figure 2. Overview of patient recruitment data across the recruitment and enrollment process.



XSL•FO
RenderX

Discussion

Principal Findings

Overall, relying on digital adjuncts in this study helped us to achieve our patient enrollment goal in less than 19 months (as compared to an initial projection of 24 months) while optimizing operational costs and minimizing disruptions to participating clinical sites. For this study, using SQL automation within the EDW allowed us to accurately identify and regularly update lists of eligible participants without ongoing human input or intervention. Digitizing recruitment and consent expanded our reach to many more eligible participants than 1 research nurse could have realistically approached via in-person recruitment, thereby shifting the research nurse's efforts and availability to collecting ultrasound scans and ensuring a positive patient experience participating in the study. Finally, using a digital interface simplified and contained data management for the study to a single server, which reduced the human burden of data management and consequently improved data integrity.

Strengths

Consistent with the existing literature, we found digitizing clinical research processes to be effective in attaining goal recruitment rates, minimizing disruptions to clinic workflows, and ensuring transparency of information regarding research studies [9,19,20]. We also experienced the benefits of integrating the technological capabilities of the electronic medical record for broad-based identification of eligible participants [21-23].

Digitizing patient recruitment and enrollment allowed us the flexibility to strategically target patients with specific outcomes or appointments in specific locations. This also meant that our recruitment data were available to be analyzed frequently, thereby providing us with real-time feedback to improve our recruitment methods. Based on this feedback, we adjusted the time interval between sending an invitation and the scheduled appointment date to optimize participant consent rates.

Reducing aspects of clinical research that are highly dependent on human labor helped to alleviate financial constraints that have previously crippled clinical research and instead focused efforts on participant experience throughout the study [24]. Over 19 months of active patient recruitment with only 1 research nurse, our approach to patient recruitment enabled the research nurse to successfully collect data from patients across 3 clinical units that were in close geographic proximity, completing data collection appointments with up to 5 participants per day. In comparison, a study looking at consent rates using traditional in-person recruitment found that 9 recruiters were needed to approach 2498 patients over a 1-year period in a primary care physician office [24]. If we extended those ratios to our study, we would have needed 18 recruiters over 2 years to approach the 10,582 patients we reached by personalized email invitation.

In clinic settings where multiple clinical research studies are conducted concurrently, front-loading recruitment activities and consent limited disruption to the clinic workflows and minimized patient recruitment fatigue or coercion. We found that limiting research personnel presence in clinics with high volumes of patients and quick patient turnaround increased clinics' willingness to collaborate as recruitment sites for our study.

For patients, shifting recruitment processes onto digital platforms allowed them unlimited time to review holistically all the information pertaining to the study, their participation in the study, and address any data privacy concerns before agreeing to participate in our study. Digitizing patient recruitment inherently standardized the information shared with all eligible participants, giving us confidence in the ethical execution of informed consent.

Finally, our QR code system alleviated pressures on research nurses to maintain cumbersome paperwork associated with data management and instead focus on collecting data and the patient's experience in the study. Additionally, digitizing data management leveraged the compatibility between already digitized patient information (on Epic) and routed relevant study data, thus simplifying the deidentification process of PHI, especially when sharing data with external parties.

Limitations

A limitation of patient recruitment via digital platforms is the potential for selection bias, in our case by restricting recruitment to patients with an associated email address. To mitigate this, research teams may wish to complement digital recruitment activities with telephone or in-person outreach to directly target groups at risk of exclusion.

We recognize that our study was minimally invasive and posed no risk to participants. As such, our results using technological adjuncts for clinical research processes may not be generalizable to clinical research with different study features. Additionally, the scope of this study did not include patient feedback on the recruitment process, and patient experiences with digital recruitment processes are a valuable area for future consideration.

Finally, using digital adjuncts in clinical research inherently means that research teams must include team members well-versed in the technical skills and know-how for setting up digitized recruitment systems and troubleshooting when needed. This option may not be readily available for all research programs or may require consideration of the costs of such expertise.

Conclusions

Digital adjuncts are promising tools that can assist in streamlining patient identification, recruitment, and enrollment in clinical research studies. In our experience, digital adjuncts allowed us to reach out to many more patients than would have been possible with traditional one-on-one, in-person–based recruitment. Additionally, we believe that the benefit of privately reviewing information pertaining to the study without time constraints or any in-person influences optimizes an ethical consent process for patients. Finally, streamlining the data management necessary for our imaging-based observational study has reduced the personnel and resources required for the completion of the study. Using these digital tools could ease research and make it more equitable, leading to higher probabilities of study completion.

XSL•FO
RenderX

Acknowledgments

The study in which this digitized patient recruitment and enrollment process is for is funded by the Bill and Melinda Gates Foundation and Google.

Authors' Contributions

All authors contributed equally to the development of this paper. SC and DM were responsible for building the digitized patient recruitment process. SC, DM, KH, and AS were responsible for providing data for this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Personalized invitation sent to patients to participate in study.
[DOCX File , 15 KB - formative_v9i1e58628_app1.docx]

References

1. Chaudhari N, Ravi R, Gogtay N, Thatte U. Recruitment and retention of the participants in clinical trials: challenges and solutions. Perspect Clin Res 2020;11(2):64-69 [FREE Full text] [doi: 10.4103/picr.PICR_206_19] [Medline: 32670830]
2. Frank G. Current challenges in clinical trial patient recruitment and enrollment. SoCRA Source. (February) 2004;2(February):30-38. [doi: 10.1093/acprof:oso/9780195035681.003.0014]
3. Ohmann C, Kuchinke W. Meeting the challenges of patient recruitment: a role for electronic health records. Int J Pharm Med 2007;21(4):263-270. [doi: 10.2165/00124363-200721040-00002]
4. Thoma A, Farrokhyar F, McKnight L, Bhandari M. Practical tips for surgical research: how to optimize patient recruitment. Can J Surg 2010;53(3):205-210 [FREE Full text] [Medline: 20507795]
5. Adams M, Caffrey L, McKevitt C. Barriers and opportunities for enhancing patient recruitment and retention in clinical research: findings from an interview study in an NHS academic health science centre. Health Res Policy Syst 2015;13:8 [FREE Full text] [doi: 10.1186/1478-4505-13-8] [Medline: 25971302]
6. McDonald AM, Knight RC, Campbell MK, Entwistle VA, Grant AM, Cook JA, et al. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. Trials 2006;7:9 [FREE Full text] [doi: 10.1186/1745-6215-7-9] [Medline: 16603070]
7. Nipp RD, Hong K, Paskett ED. Overcoming barriers to clinical trial enrollment. Am Soc Clin Oncol Educ Book 2019;39:105-114 [FREE Full text] [doi: 10.1200/EDBK_243729] [Medline: 31099636]
8. Briel M, Elger BS, McLennan S, Schandelmaier S, von Elm E, Satalkar P. Exploring reasons for recruitment failure in clinical trials: a qualitative study with clinical trial stakeholders in Switzerland, Germany, and Canada. Trials 2021;22(1):844 [FREE Full text] [doi: 10.1186/s13063-021-05818-0] [Medline: 34823582]
9. Fitzer K, Haeuslschmid R, Blasini R, Altun FB, Hampf C, Freiesleben S, et al. Patient recruitment system for clinical trials: mixed methods study about requirements at ten university hospitals. JMIR Med Inform 2022;10(4):e28696 [FREE Full text] [doi: 10.2196/28696] [Medline: 35442203]
10. Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. Contemp Clin Trials Commun 2018;11:156-164 [FREE Full text] [doi: 10.1016/j.conctc.2018.08.001] [Medline: 30112460]
11. Goodyear-Smith F, York D, Petousis-Harris H, Turner N, Copp J, Kerse N, et al. Recruitment of practices in primary care research: the long and the short of it. Fam Pract 2009;26(2):128-136. [doi: 10.1093/fampra/cmp015] [Medline: 19251761]
12. Manca DP, O'Beirne M, Lightbody T, Johnston DW, Dymianiw DL, Nastalska K, et al. The most effective strategy for recruiting a pregnancy cohort: a tale of two cities. BMC Pregnancy Childbirth 2013;13:75 [FREE Full text] [doi: 10.1186/1471-2393-13-75] [Medline: 23521869]
13. Brøgger-Mikkelsen M, Ali Z, Zibert JR, Andersen AD, Thomsen SF. Online patient recruitment in clinical trials: systematic review and meta-analysis. J Med Internet Res 2020;22(11):e22179 [FREE Full text] [doi: 10.2196/22179] [Medline: 33146627]
14. Abernethy A, Adams L, Barrett M, Bechtel C, Brennan P, Butte A, et al. The promise of digital health: then, now, and the future. NAM Perspect 2022;2022 [FREE Full text] [doi: 10.31478/202206e] [Medline: 36177208]
15. Rosa C, Marsch LA, Winstanley EL, Brunner M, Campbell AN. Using digital technologies in clinical trials: current and future applications. Contemp Clin Trials 2021;100:106219 [FREE Full text] [doi: 10.1016/j.cct.2020.106219] [Medline: 33212293]
16. Inan OT, Tenaerts P, Prindiville SA, Reynolds HR, Dizon DS, Cooper-Arnold K, et al. Digitizing clinical trials. NPJ Digit Med 2020;3:101 [FREE Full text] [doi: 10.1038/s41746-020-0302-y] [Medline: 32821856]

XSL•FO
RenderX

17. Zayas-Cabán T, Okubo TH, Posnack S. Priorities to accelerate workflow automation in health care. J Am Med Inform Assoc 2022;30(1):195-201 [FREE Full text] [doi: 10.1093/jamia/ocac197] [Medline: 36259967]
18. Zayas-Cabán T, Haque SN, Kemper N. Identifying opportunities for workflow automation in health care: lessons learned from other industries. Appl Clin Inform 2021;12(3):686-697 [FREE Full text] [doi: 10.1055/s-0041-1731744] [Medline: 34320683]
19. Digital clinical trials workshop: creating a vision for the future. National Institutes of Health, National Heart, Lung, and Blood Institute. 2019. URL: https://www.nhlbi.nih.gov/events/2019/digital-clinical-trials-workshop-creating-vision-future [accessed 2023-10-19]
20. Mazurenko O, Sanner L, Apathy NC, Mamlin BW, Menachemi N, Adams MCB, et al. Evaluation of electronic recruitment efforts of primary care providers as research subjects during the COVID-19 pandemic. BMC Prim Care 2022;23(1):95 [FREE Full text] [doi: 10.1186/s12875-022-01705-y] [Medline: 35484491]
21. Embi PJ, Jain A, Harris CM. Physicians' perceptions of an electronic health record-based clinical trial alert approach to subject recruitment: a survey. BMC Med Inform Decis Mak 2008;8:13 [FREE Full text] [doi: 10.1186/1472-6947-8-13] [Medline: 18384682]
22. Dugas M, Lange M, Berdel WE, Müller-Tidow C. Workflow to improve patient recruitment for clinical trials within hospital information systems—a case-study. Trials 2008;9:2 [FREE Full text] [doi: 10.1186/1745-6215-9-2] [Medline: 18186949]
23. Dugas M, Lange M, Müller-Tidow C, Kirchhof P, Prokosch H. Routine data from hospital information systems can support patient recruitment for clinical studies. Clin Trials 2010;7(2):183-189. [doi: 10.1177/1740774510363013] [Medline: 20338903]
24. Felsen CB, Shaw EK, Ferrante JM, Lacroix LJ, Crabtree BF. Strategies for in-person recruitment: lessons learned from a New Jersey Primary Care Research Network (NJPCRN) study. J Am Board Fam Med 2010;23(4):523-533 [FREE Full text] [doi: 10.3122/jabfm.2010.04.090096] [Medline: 20616295]

Abbreviations

EDW: enterprise data warehouse
PHI: protected health information
REDCap: Research Electronic Data Capture

Edited by A Mavragani; submitted 12.04.24; peer-reviewed by Y Wang, JM Kerver; comments to author 28.11.24; revised version received 05.12.24; accepted 06.12.24; published 29.01.25.

Please cite as:
Koh J, Caron S, Watters AN, Vaidyanathan M, Melnick D, Santi A, Hudson K, Arguelles C, Mathur P, Etemadi M
Technological Adjuncts to Streamline Patient Recruitment, Informed Consent, and Data Management Processes in Clinical Research: Observational Study
JMIR Form Res 2025;9:e58628
URL: https://formative.jmir.org/2025/1/e58628
doi:10.2196/58628
PMID:39879093

©Jodie Koh, Stacey Caron, Amber N Watters, Mahesh Vaidyanathan, David Melnick, Alyssa Santi, Kenneth Hudson, Catherine Arguelles, Priyanka Mathur, Mozziyar Etemadi. Originally published in JMIR Formative Research (https://formative.jmir.org), 29.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Understanding Individual Differences in Happiness Sources and Implications for Health Technology Design: Exploratory Analysis of an Open Dataset

Edel Ennis[1*], PhD; Raymond Bond[2*], PhD; Maurice Mulvenna[2], PhD; Colm Sweeney[1], MSc

[1]School of Psychology, Ulster University, Coleraine, United Kingdom

[2]School of Computing, Ulster University, Belfast, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Edel Ennis, PhD
School of Psychology
Ulster University
Cromore Road
Coleraine, BT52 1SA
United Kingdom
Phone: 44 2870123892
Email: e.ennis@ulster.ac.uk

Abstract

Background: Psychologists have developed frameworks to understand many constructs, which have subsequently informed the design of digital mental health interventions (DMHIs) aimed at improving mental health outcomes. The science of happiness is one such domain that holds significant applied importance due to its links to well-being and evidence that happiness can be cultivated through interventions. However, as with many constructs, the unique ways in which individuals experience happiness present major challenges for designing personalized DMHIs.

Objective: This paper aims to (1) present an analysis of how sex may interact with age, marital status, and parental status to predict individual differences in sources of happiness, and (2) to present a preliminary discussion of how open datasets may contribute to the process of designing health-related technology innovations.

Methods: The HappyDB is an open database of 100,535 statements of what people consider to have made them happy, with some people asking to consider the past 24 hours (49,831 statements) and some considering the last 3 months (50,704 statements). Demographic information is also provided. Binary logistic regression analyses are used to determine whether various groups differed in their likelihood of selecting or not selecting a category as a source of their happiness.

Results: Sex and age interacted to influence what was selected as sources of happiness, with patterns being less consistent among female individuals in comparison with male individuals. For marital status, differences in sources of happiness were predominantly between married individuals and those who are divorced or separated, but these were the same for both sexes. Married, single, and widowed individuals were all largely similar in their likelihood of selecting each of the categories as a source of their happiness. However, there were some anomalies, and sex appeared to be important in these anomalies. Sex and parental status also interacted to influence what was selected as sources of happiness.

Conclusions: Sex interacts with age, marital status, and parental status in the likelihood of reporting affection, bonding, leisure, achievement, or enjoying the moment as sources of happiness. The contribution of an open dataset to understanding individual differences in sources of happiness is discussed in terms of its potential role in addressing the challenges of designing DMHIs that are ethical, responsible, evidence based, acceptable, engaging, inclusive, and effective for users. The discussion considers how the content design of DMHIs in general may benefit from exploring new methods informed by diverse data sources. It is proposed that examining the extent to which insights from nondigital settings can inform requirements gathering for DMHIs is warranted.

(JMIR Form Res 2025;9:e65658) doi:10.2196/65658

XSL•FO
RenderX

KEYWORDS

happiness; sexes; age; marital status; parents; affections; achievements; datasets; digital health; well-being; mental health; digital mental health interventions; regression analyses; evidence based

Introduction

Exploring Individual Differences in Sources of Happiness, and the Role of Open Datasets in Health Technology Design

Well-being is founded upon positive emotions, engagement, relationships, meaning, and accomplishments [1]. Different people will derive well-being from each of these 5 building blocks to varying degrees [1]. The premise is that individuals need to focus on positive emotions, engage with life and the activities around them, strive for meaningful relationships and social connections, try to find meaning in life, and have goals and ambitions. Later discussions of the model have concluded life is easier and more enjoyable when our health is optimal. Much literature uses terms such as “happiness,” “subjective well-being,” “thriving,” and “flourishing” interchangeably [2,3]. There is no complete consensus on the relationship between happiness and well-being, but happiness is at least considered to be a component of well-being.

At the very basic level, happiness is an example of a positive emotion that contributes to well-being, with joy, love, and gratitude representing additional examples [4]. The links between happiness and well-being have led to great interest in the science of happiness. This investigates what constitutes happiness; what makes people happy; and what we can do, to ourselves or others, to feel happier [5]. While genetics and life circumstances (eg, finances, job, and material belongings) play roles, 40% of our happiness is determined by behaviors that are under our control [6].

Open data represent a crucial element of the open science movement [7]. These datasets are publicly available, without any restrictions on their use or distribution [7]. The growth of the open science movement has led to an increased availability of open datasets and consideration of how we can reap maximum benefits from these [7]. Example benefits have included transparency and the capacity to amalgamate large datasets, which can give comprehensive insights to inform data-driven decision-making, for example, informing targeted interventions or data-driven resource allocation [7]. Open datasets also allow analysis from a variety of diverse perspectives and thus represent a powerful resource for developing insights into the underlying meaning of psychological constructs [8]. This in turn facilitates enhanced precision and clarity within theories and definitions, an understanding of the meaning of the construct across different cultures and contexts, the development of improved means of measurement, and the development of more effective interventions and treatment options [8].

The HappyDB is an example of a data source that has advanced our understanding of the individual meaning and sources of happiness. The HappyDB is a dataset containing over 100,000 happy statements [9] that is freely available on the web [10]. Some people were asked to consider what had brought them happiness in the past 24 hours, and some were asked to consider the last 3 months. Individuals could make more than 1 entry, but not more than 3 entries. Happiness does not always mean the same thing. Short-term happiness (24 hours) tends to be associated with more daily tasks such as “Exercise,” “Nature,” and “Leisure,” whereas longer-term happiness is more likely to come from loved ones or achievements [9]. Sources of happiness vary cross-culturally [11].

Based on studying happiness within a Latin-American sample, Pena-Lopez et al [12] stressed the need to understand the individual determinants of happiness. Focusing on the words used to express happiness within the HappyDB as opposed to the categories of the sources of happiness, Mohamed and Mostafa [13] used logistic regression, gradient boosting, and the fastText deep learning algorithm, and they found that demographics did influence the meaning and sources of happiness. For example, male individuals expressed greater interest in games and gadgets as sources of happiness, while female individuals discussed family and friends to a greater extent [13]. Parents placed more discussion around the well-being of their families and children, whereas nonparents placed their emphasis on friends, games, eating out, pets, and watching television [13]. Unmarried people focused mostly on dating, friendship, food, and exercise as sources of happiness, whereas married people focused mainly on children and family [13]. Age brackets also differed in how they expressed happiness [13]. Other research studies have also shown differences between demographic groups in the factors associated with happiness levels [14,15].

As noted earlier, sex influences the sources of happiness. Age, marital status, and parenthood status also influence the sources of happiness. The effects of age and marital status would also appear to interact with sex to influence the importance of individual sources of happiness [14]. This is not surprising given that they have been shown to interact in predicting overall levels of happiness or well-being. For example, some research has suggested that the transition to parenthood links to mental well-being, and this differs across the sexes [16,17]. However, other research suggests that the sex gap in the emotional implications of parenthood has dissipated [18]. On a similar note, sex has been found to interact with marital status in predicting levels of happiness [19].

This paper expands on existing research by examining how sex and time frame (last 24 hours vs last 3 months) interact with demographics like age, marital status, and parenthood in predicting sources of happiness. It offers new insights not only by exploring how these factors influence well-being over different time periods but also by considering them in relation to the 5 primary sources of happiness. Mohamed and Mostafa [13] examined the topics of sources of happiness, through natural language processing, and Asai et al [9] identified that statements of happiness can be classified within 7 categories (Table 1). Two of these are endorsed much less frequently than

XSL•FO
RenderX

the others (across 24-hour and 3-month time periods; Table 1), but the five most popular sources of happiness are (1) affection, (2) bonding, (3) leisure, (4) achievement, and (5) enjoying the moment [9]. This research focuses on these 5 most common sources of happiness.

These 5 sources shall be focused upon within this paper. The importance of affection aligns with research showing links between romantic relationships and mental health [20]. The importance of achievement concurs with Walsh et al [21] in their conclusion that happiness is associated with career success [21]. Indeed, while much research is correlational, they conclude that happiness actually often precedes career success, in that positive emotions lead to improved outcomes in the workplace [21].

Table 1. The categories of happy moments: definitions, examples, and frequency of occurrence.

Category	Definition	Examples	Happiness statements within the 24-hour category (n=49,831), n (%)	Happiness statements within the 3-month category (n=50,704), n (%)
Achievement	With extra effort to achieve a better-than-expected result	• Finish work • Complete marathon	15,398 (30.9)	18,507 (36.5)
Affection	Meaningful interaction with family, loved ones, and pets	• Hug • Cuddle • Kiss	16,295 (32.7)	17,797 (35.1)
Bonding	Meaningful interaction with friends and colleagues	• Have meals with coworker • Meet friends	5182 (10.4)	5476 (10.8)
Enjoy the Moment	Being aware or reflecting on present environment	• Have a good time • Mesmerize	6628 (13.3)	4513 (8.9)
Leisure	An activity done regularly in one's free time for pleasure	• Play games • Watch movies • Bake cookies	4335 (8.7)	3093 (6.1)
Nature	In the open air and in nature	• Garden • Beach • Sunset • Weather	1046 (2.1)	761 (1.5)
Exercise	With intent to exercise or workout	• Run • Bike • Do yoga • Lift weights	747 (1.5)	406 (0.8)

Objectives

This paper uses the HappyDB database to examine whether age category, marital status, and parenthood status predict the likelihood of selecting or not selecting each of the 5 most popular sources of happiness (affection, bonding, leisure, achievement, and enjoying the moment) and whether this differs by sex and time frame. Results will be discussed within a preliminary discussion of how open datasets such as HappyDB may represent a novel method to inform the content design of positive psychology interventions (PPIs), and indeed digital mental health interventions (DMHIs) in general. This domain brings great potential, but the ethical and responsible design of DMHIs brings many considerations and calls for creative approaches [22,23]. Best practice guidelines for the development of DMHIs recommend evidence-based approaches; theory-driven design; tailoring the DMHIs to the population and the context; and consideration of engagement, inclusivity, accessibility, and flexibility [22,24,25]. This represents a significant challenge, as collaboration is needed across many disciplines, and open datasets may have valuable contributions to make.

Methods

Materials

There are 100,535 statements within the HappyDB database, with 49,831 statements relating to what made the individual happy in the past 24 hours and 50,704 relating to what made the individual happy within the last 3 months [9]. Table 1 illustrates definitions of each of the 7 categories and examples of the topics discussed within each category. Examples of statements include "My son gave me a big hug in the morning when I woke him up," "I finally managed to make 40 pushups," "I had dinner with my husband," "Morning started with the chirping of birds and the pleasant sun rays," "The event at work was fun. I loved spending time with my good friends and laughing," or "I went to the park with the kids. The weather was perfect" [9].

XSL•FO
RenderX

Ethical Considerations

HappyDB is an open, publicly available dataset that has been used in many other research studies. No personally identifiable information is included in the datasets. Nonetheless, this study received ethical approval from the Filter Committee of the School of Psychology at Ulster University (FCPSYCH2020EE).

Participants

Sample details are outlined in Table 2.

Table 2. Sample details.

Variable	Overall sample (N=100,535), n (%)	24-hour data (n=49,831), n (%)	3-month data (n=50,704), n (%)
Sex			
Male	57,661 (57.4)	28,483 (49.4)	20,937 (49.8)
Female	42,036 (41.8)	29,178 (50.6)	21,099 (50.2)
Marital status			
Married	41,325 (41.2)	20,925 (42.1)	20,400 (40.3)
Divorced or separated	4446 (4.4)	2142 (4.3)	2302 (4.6)
Widowed	477 (0.5)	228 (0.5)	249 (0.5)
Single	54,091 (53.9)	26,441 (53.2)	27,650 (54.6)
Parental status			
No children	60,911 (60.6)	29,700(59.7)	31,211 (61.6)
Children	39,499 (39.3)	20,072 (40.3)	19,427 (38.4)
Age group (years)			
18-25	24,138 (24.1)	11,960 (24)	12,178 (24.1)
26-64	74,932 (74.7)	37,161 (74.7)	37,771 (74.6)
65 and older	1264 (1.3)	609 (1.2)	655 (1.8)

Analyses

Analyses took the form of binary regression analyses. Based on Asai et al [9] noting that sources of happiness varied across the time periods, the 24-hour data and 3-month data were analyzed separately. Analyses also considered male and female individuals separately. As noted earlier, exercise and nature were the least frequently endorsed sources of happiness (Table 1). Due to the small group sizes associated with these 2 categories, analyses focused on considering group differences within the remaining top-5 sources of happiness. Both sex and parental status were treated as dichotomous variables. Age was considered an ordinal variable, as it has a natural order but unequal intervals. Marital status, being categorical with no inherent order or ranking, was classified as nominal.

There were 20 binary logistic regression analyses in all. The overall dataset was separated into four separate datasets: (1) sources of happiness for female individuals over the previous 24 hours; (2) sources of happiness for male individuals over the previous 24 hours; (3) sources of happiness for female individuals over the previous 3 months; and (4) sources of happiness for male individuals over the previous 3 months. Within each of the 4 datasets, 5 binary logistic regression analyses were conducted. Each one of these five regressions considered one of the sources of happiness as a dependent variable: (1) the presence or absence of affection as a source of happiness; (2) the presence or absence of bonding as a source of happiness; (3) the presence or absence of leisure as a source of happiness; (4) the presence or absence of achievement as a source of happiness; and (5) the presence or absence of enjoying the moment as a source of happiness. The analysis examined whether each of these could be predicted by the demographic factors of age category, marital status, and parenthood status (predictors) and whether this differed by sex and time frame. Basically, there were 5 forms of regression, and each one was done within each of the 4 datasets, giving rise to 20 binary logistic regression analyses. For age as an independent variable, 18-25 years represented the reference category. For marital status as a predictor, married individuals represented the reference category. For parental status as a predictor, those with no children represented the reference category. The assumptions for the analyses were addressed as follows.

- Binary logistic regression assumes that observations are independent. While individuals could provide multiple reports about their sources of happiness, each report was treated as a separate data entry, ensuring no repeated measures
- Multicollinearity among independent variables was examined using variance inflation factor (VIF) values. VIF values below 5 indicate low multicollinearity, while values above 10 suggest a high degree of multicollinearity [26]. In this study, all VIF values were below 2, confirming no multicollinearity issues.
- The linearity of the logit: assumption applies to continuous independent variables. Since all independent variables in this study are categorical, this assumption is not relevant.
- Outlier detection is not applicable here due to the categorical nature of all variables.

XSL•FO
RenderX

Results

Affection

Predictors of affection as a source of happiness are outlined in Table 3. The first category (not selecting affection) was the reference category.

Within male individuals, the different age groups did not differ in their likelihood of selecting or not selecting affection as a source of their happiness. Within female individuals, all age groups were significantly more likely than the 18-25 years age group to select affection as a source of their happiness. This was evident within both the 24-hour (P<.001) and 3-month (P<.01) reports of sources of happiness.

Within both the male and female data, divorced or separated individuals were significantly more likely than married individuals to select affection as a source of their happiness (P<.001). However, across both male and female individuals, those who were widowed did not differ from married individuals in their likelihood of selecting affection as a source of happiness. Within male individuals, those who were single did not differ from married individuals in their likelihood of selecting affection as a source of happiness. Within the female individuals, this was the same within the 3-month data. However, within the data where female individuals had reported their sources of happiness within the last 24 hours, single female individuals were significantly (1.38 times; P<.05) more likely than married female individuals to select affection as a source of their happiness.

Within both the male and female data, those with children were significantly less likely than those without children to select affection as a source of their happiness (P<.001).

Table 3. ORs[a] and significance for logistic regression on choosing "affection" as a source of happiness by demographics and time frame[b].

	Male, OR (95% CI)		Female, OR (95% CI)	
	24-hour data	3-month data	24-hour data	3-month data
Age group (years) with "18-25 years" as reference				
26-64	1.13 (0.89-1.45)	1.14 (0.91-1.44)	2.13 (1.64-2.75)***	1.43 (1.12-1.83)**
65 and older	1.06 (0.84-1.35)	1.17 (0.94-1.46)	1.55 (1.21-1.98)***	1.21 (0.95-1.52)
Marital status with "Married" as reference				
Divorced or separated	1.54 (1.43-1.66)***	1.48 (1.38-1.59)***	1.49 (1.38-1.61)***	1.34 (1.24-1.45)***
Widowed	1.00 (0.85-1.19	1.13 (0.96-1.32)	1.04 (0.91-1.19)	0.98 (0.87-1.11)
Single	1.31 (0.71-2.41)	1.32 (0.81-2.13)	1.38 (1.01-1.88)*	1.26 (0.92-1.71)
Parenthood with "No children" as reference				
Children	0.60 (0.56-0.65)***	0.63 (0.59-0.68)***	0.56 (0.52-0.60)***	0.51 (0.48-0.55)***

[a]OR: odds ratio.

[b]Reference category: no selection of affection.

*P<.05.

**P<.01.

***P<.001.

Bonding

Predictors of bonding as a source of happiness are outlined in Table 4. The first category (not selecting bonding) was the reference category.

For age as a predictor of bonding as a source of happiness, 18-25 years represents the reference category. Within male individuals, the different age groups did not differ in their likelihood of selecting or not selecting bonding as a source of their happiness. Within female individuals, the different age groups did not differ in their likelihood of selecting or not selecting bonding as a source of their happiness over the past 3 months. However, when considering what made them happy over the past 24 hours, female individuals aged 26-64 years were significantly more (1.94 times; P<.05) more likely than female individuals aged 18-25 years to select bonding as a source of their happiness. Within reports of happiness in the past 24 hours, female individuals aged 65 years old and older did not differ significantly from those aged 18-25 years in their likelihood of selecting bonding as a source of their happiness.

Within both the male and female data, divorced or separated individuals were significantly less likely than married individuals to select bonding as a source of their happiness (P<.001). Within male reports of happiness within the past 24 hours, widowed male individuals were significantly less likely than married male individuals to select bonding as a source of their happiness (P<.05). However, widowed and married male individuals did not differ in their likelihood of selecting bonding as a source of their happiness over the past 3 months. Widowed and married female individuals did not differ in their likelihood of selecting bonding as a source of their happiness. Across both male and female individuals, those who were single did not differ from married individuals in their likelihood of selecting bonding as a source of happiness.

Male individuals with and without children did not differ in their likelihood of selecting bonding as a source of their happiness over the past 24 hours. However, within 3-month

XSL•FO
RenderX

reports of sources of happiness, male individuals with children were much more likely to endorse something within the bonding category in comparison with male individuals with no children (P<.001). Within female individuals, those with children were much more likely to endorse something within the bonding category in comparison with those with no children (P<.001).

Table 4. ORs[a] and significance for logistic regression on choosing "bonding" as a source of happiness by demographics and time frame[b].

	Male, OR (95% CI)		Female, OR (95% CI)	
	24-hour data	3-month data	24-hour data	3-month data
Age group (years) with "18-25 years" as reference				
26-64	0.85 (0.58-1.24)	1.03 (0.69-1.55)	1.94 (1.16-3.24)*	1.03 (0.68-1.58)
65 and older	0.75 (0.52-1.08)	0.98 (0.66-1.46)	1.52 (0.92-2.50)	0.93 (0.62-1.40)
Marital status with "Married" as reference				
Divorced or separated	0.69 (0.62-0.78)***	0.68 (0.61-0.75)***	0.78 (0.69-0.88)***	0.82 (0.72-0.92)***
Widowed	0.78 (0.61-1.00)*	0.83 (0.65-1.06)	0.96 (0.77-1.18)	1.13 (0.93-1.38)
Single	0.83(0.32-2.11)	1.10 (0.55-2.24)	1.20 (0.73-1.96)	0.88 (0.51-1.51)
Parenthood with "No children" as reference				
Children	1.07 (0.96-1.20)	1.24 (1.11-1.39)***	1.26 (1.11-1.42)***	1.61 (1.43-1.82)

[a]OR: odds ratio.

[b]Reference category: no selection of affection.

*P<.05.

**P<.01.

***P<.001.

Leisure

Predictors of leisure as a source of happiness are outlined in Table 5. The first category (not selecting leisure) was the reference category.

Table 5. ORs[a] and significance for logistic regression on choosing "leisure" as a source of happiness by demographics and time frame[b].

	Male, OR (95% CI)		Female, OR (95% CI)	
	24-hour data	3-month data	24-hour data	3-month data
Age group (years) with "18-25 years" as reference				
26-64	1.31 (0.80-2.15)	0.85 (0.53-1.36)	0.26 (0.17-0.38)***	0.44 (0.25-0.78)**
65 and older	1.40 (0.86-2.28)	0.94 (0.59-1.49)	0.44(0.30-0.63)***	0.85(0.49-1.47)
Marital status with "Married" as reference				
Divorced or separated	1.18 (1.05-1.31)**	1.08 (0.95-1.22)	0.95 (0.83-1.09)	1.27 (1.08-1.49)**
Widowed	1.24 (0.97-1.58)	0.94 (0.70-1.27)	0.95 (0.75-1.21)	1.28 (0.98-1.68)
Single	3.11 (1.48-6.53)**	0.95 (0.35-2.63)	0.31 (0.13-0.77)**	0.26 (0.06-1.04)
Parenthood with "No children" as reference				
Children	1.84 (1.63-2.07)	1.56 (1.36-1.78)***	1.87 (1.63-2.15)***	2.35 (2.00-2.75)***

[a]OR: odds ratio.

[b]Reference category: no selection of affection.

*P<.05.

**P<.01.

***P<.001.

Results show that within the male individuals of the sample, the different age groups did not differ in their likelihood of selecting or not selecting leisure as a source of their happiness. Within the female individuals of the sample, those aged 26-64 years and those aged 65 years and older were both significantly less likely than those aged 18-25 years to select leisure as a source of their happiness (P<.001). This was evident in reports of happiness over the past 3 months and the past 24 hours.

In relation to marital status, divorced or separated male individuals were significantly more (1.18 times; P<.01) likely than married male individuals to cite leisure as a source of their happiness over the past 24 hours. However, divorced or

XSL•FO
RenderX

separated and married male individuals did not differ in their likelihood of citing leisure as a source of their happiness over the past 3 months. The pattern was reversed among female individuals. Specifically, divorced or separated and married female individuals did not differ in their likelihood of citing leisure as a source of their happiness over the past 24 hours. However, divorced or separated female individuals were significantly more (1.27 times; $P<.01$) likely than married female individuals to cite leisure as a source of their happiness over the past 3 months.

Male individuals with and without children did not differ in their likelihood of selecting leisure as a source of their happiness over the past 24 hours. However, within 3-month reports of sources of happiness, male individuals with children were much more likely to endorse something within the leisure category in comparison with male individuals with no children ($P<.001$). Within female individuals, those with children were much more likely to endorse something within the leisure category in comparison with those with no children ($P<.001$).

Achievement

Predictors of achievement as a source of happiness are outlined in Table 6. The first category (not selecting achievement) was the reference category.

Results show that within the male individuals of the sample, the different age groups did not differ in their likelihood of selecting or not selecting achievement as a source of their happiness. Within the female individuals of the sample, those aged 26-64 years were significantly less likely than those aged 18-25 years to select achievement as a source of their happiness. This was evident in reports of happiness over the past 3 months ($P<.05$) and the past 24 hours ($P<.001$). However, female individuals aged 65 years and older only differed significantly from those aged 18-25 years in terms of them being less likely (0.74 times; $P<.05$) to cite achievement as a source of their happiness over the past 24 hours. The 2 groups did not differ significantly in their likelihood of endorsing achievement as a source of their happiness over the past 3 months.

In relation to marital status, those who were divorced or separated (male and female individuals) were significantly less likely than their counterparts to select achievement as a source of their happiness ($P<.001$). This was evident within 24-hour reports and those of the past 3 months. Within the 24-hour and 3-month reports, widowed individuals did not differ from their married counterparts in the likelihood of selecting achievement as a source of their happiness. Compared to married male individuals, single male individuals were significantly less (0.39 times; $P<.05$) likely to cite achievement as a source of their happiness over the past 24 hours. However, single and married female individuals did not differ significantly in their likelihood of endorsing achievement as a source of their happiness over the past 3 months. Single and married male individuals did not differ significantly in their likelihood of endorsing achievement as a source of their happiness over the past 24 hours or 3 months.

Within male and female individuals, and within the 24-hour reports and the 3-month reports, those with children were significantly more likely than those without children to select achievement as a source of their happiness ($P<.001$).

Table 6. ORs[a] and significance for logistic regression on choosing "achievement" as a source of happiness by demographics and time frame[b].

	Male, OR (95% CI)		Female, OR (95% CI)	
	24-hour data	3-month data	24-hour data	3-month data
Age group (years) with "18-25 years" as reference				
26-64	0.96 (0.74-1.25)	0.88 (0.70-1.11)	0.64 (0.49-0.84)***	0.73 (0.57-0.94)*
65 and older	1.02 (0.80-1.32)	0.84 (0.67-1.05)	0.74 (0.58-0.96)*	0.81 (0.63-1.03)
Marital status with "Married" as reference				
Divorced or separated	0.83 (0.77-0.89)***	0.82 (0.77-0.88)***	0.83 (0.76-0.90)***	0.76 (0.70-0.82)***
Widowed	0.99 (0.84-1.15)	1.08 (0.93-1.25)	1.12 (0.97-1.28)	0.90 (0.80-1.03)
Single	0.39 (0.18-0.83)*	0.92 (0.57-1.49)	1.08 (0.78-1.50)	0.84 (0.61-1.16)
Parenthood with "No children" as reference				
Children	1.17 (1.09-1.26)***	1.16 (1.08-1.24)***	1.30 (1.20-1.41)***	1.24 (1.15-1.34)***

[a]OR: odds ratio.

[b]Reference category: no selection of affection.

*$P<.05$.

**$P<.01$.

***$P<.001$.

Enjoy the Moment

Predictors of enjoying the moment as a source of happiness are outlined in Table 7. The first category (not selecting enjoying the moment) was the reference category.

Within male and female individuals, the age categories did not differ in their likelihood of selecting enjoying the moment as a source of their happiness. This was evident within both the 24-hour and 3-month reports of sources of happiness. Within the 24-hour data, divorced or separated male and female individuals were significantly less likely than their counterparts

XSL•FO
RenderX

to cite enjoying the moment as a source of their happiness over the past 24 hours (P<.001). However, within the 3-month data, divorced or separated male and female individuals did not differ significantly from their counterparts in their likelihood of citing enjoying the moment as a source of their happiness over the past 24 hours. Within male and female individuals, neither the widowed individuals nor single individuals differed from their married counterparts in their likelihood of selecting enjoying the moment as a source of their happiness. This was evident within both the 24-hour and 3-month reports of sources of happiness.

Within 24-hour reports of happiness, those with children (male and female individuals) did not differ from their counterparts in their likelihood of selecting enjoying the moment as a source of their happiness. However, within 3-month reports of happiness, those with children (male individuals: P<.01 and female individuals: P<.001) were significantly more likely than their counterparts in their likelihood of selecting enjoying the moment as a source of their happiness.

Table 7. ORs[a] and significance for logistic regression on choosing "enjoying the moment" as a source of happiness by demographics and time frame[b].

	Male, OR (95% CI)		Female, OR (95% CI)	
	24-hour data	3-month data	24-hour data	3-month data
Age group (years) with "18-25 years" as reference				
26-64	1.29 (0.89-1.87)	1.54 (0.95-2.52)	0.92 (0.63-1.34)	141 (0.87-2.27)
65 and older	1.26 (0.88-1.82)	1.57 (0.97-2.55)	0.94 (0.65-1.35)	1.13 (0.71-1.80)
Marital status with "Married" as reference				
Divorced or separated	0.82 (0.75-0.91)***	0.97 (0.87-1.09)	0.79 (0.71-0.89)***	1.00 (0.88-1.14)
Widowed	1.02 (0.83-1.26)	0.79 (0.60-1.04)	0.86 (0.71-0.89)	1.17 (0.95-1.46)
Single	1.58 (0.78-3.21)	0.46 (0.15-1.47)	0.70 (0.42-1.15)	1.39 (0.83-2.32)
Parenthood with "No children" as reference				
Children	0.97 (0.88-1.07)	1.20 (1.07-1.35)**	1.07 (0.95-1.19)	1.31 (1.15-1.49)***

[a]OR: odds ratio.

[b]Reference category: no selection of affection.

*P<.05.

**P<.01.

***P<.001.

Discussion

Principal Findings

This paper set out to (1) present an analysis of how sex may interact with age, marital status, and parental status to predict individual differences in sources of happiness; and (2) present a preliminary discussion of how open datasets may contribute to the process of designing health-related technology innovations. In terms of a brief summary of findings, sex interacted with age, marital status, and parental status in predicting sources of happiness, with some patterns varying across time periods.

Earlier research has suggested that sex interacts with other demographic factors to predict individual differences in sources of happiness [14]. Results partially supported this suggestion. First, sex and age interacted in predicting sources of happiness. Clear trends were evident in the data for male individuals. Male individuals across the 3 age categories did not differ in their likelihood of reporting affection, bonding, leisure, achievement, or enjoying the moment as sources of happiness. Trends were not as consistent within the female data, with the time period under consideration also exerting significant effects on the likelihood of affection, bonding, leisure, and achievement being reported as a source of happiness. The exact direction of effects is described in the *Results* section. The only exception was in relation to enjoying the moment as a source of happiness wherein, like the male group, the different female age groups did not differ in their likelihood of reporting enjoying the moment as a source of happiness. This was regardless of the time period under consideration. Results can therefore be said to only partially agree with a previous review by Buijs et al [27], which concluded that chronological age did not relate to affection, status, and other behavioral factors as sources of happiness.

Results partially tied in with earlier conclusions that sex interacts with marital status in predicting levels of happiness [19]. Across the board, differences in sources of happiness were predominantly between married individuals and those who are divorced or separated. However, these differences were the same for both sexes. Married, single, and widowed individuals were all largely similar in their likelihood of selecting each of the categories as a source of their happiness. However, there were some anomalies and sex appeared to be important in these anomalies. For example, in comparison to married male individuals, single male individuals were less likely to report achievement as a source of their happiness in the previous 24 hours. On the contrary, married and single female individuals did not differ in their likelihood of reporting achievement as a source of their happiness in the previous 24 hours. Compared to married male individuals, single male individuals were significantly more likely to report leisure as a source of their

XSL•FO
RenderX

happiness in the past 24 hours. Conversely, in comparison with married female individuals, single female individuals were significantly less likely to report leisure as a source of happiness in the previous 24 hours. Such patterns were not evident in the 3-month data. It must be acknowledged that the current study only looks at marital status, whereas other factors such as commitment are also important [20]. Understanding how marital status ties to happiness is important, as there is a bidirectional relationship between relationships and mental health, which is particularly strong when relationship status is used as a predictor of health [20].

Parental status and sex also appeared to interact in predicting some, but not all, sources of happiness. Compared to those who were not parents, those who were parents were significantly less likely to consider affection as a source of their happiness and significantly more likely to consider achievement as a source of their happiness. However, differences were apparent within both the male and female data, and across both the 24-hour and 3-month time periods. Compared to male individuals who were not parents, those male individuals who were parents were significantly more likely to consider enjoying the moment as a source of their happiness over the past 3 months, but not the past 24 hours. This exact same pattern was evident in the female data. However, as can be seen from the results, the sexes did differ in how they reported bonding and leisure as sources of happiness. Some past research has suggested sex differences in how the transition to parenthood links to mental well-being [16,17], and some have suggested that the sex gap in the emotional implications of parenthood has dissipated [18]. The current findings are consistent with the inconclusive nature of past research, and the recognized need to consider the complexity of how and why parents experience happiness [28].

Limitations

Limitations of course must be considered. As the HappyDB data were gathered via crowdsourcing, the representativeness of the sample is unknown. Obviously, data collection was dependent upon access to technology. Many other demographic factors link to happiness, for example, health and socioeconomic status. Sample details are not available for factors such as socioeconomic status or health conditions, but the US background and the limited representation of older participants are clear limitations in terms of the cultural and age group representativeness of this analysis. Open datasets often lack details on psychological factors, such as personality traits, that influence individual differences in happiness [29]. To maximize their use in developing DMHIs, researchers must balance the breadth and depth of data. While large sample sizes offer generalizable insights and scalability, incorporating psychometric measures (eg, personality or emotional well-being scales), and providing richer, more personalized data. A balanced approach that integrates demographic diversity with targeted, detailed subsamples would ensure both robust and effective DMHIs. Similarly, within the Positive Emotions, Engagement, Relationships, Meaning, Accomplishment, and Health model, happiness is just 1 example of positive emotion, and many other components link to well-being (eg, engagement, relationships, meaning, accomplishments, and health). This paper analyzes each source of happiness independently. However, exploring correlations between sources within demographics could offer deeper insights. For instance, responses about affection, bonding, leisure, achievement, or enjoying the moment might reflect the same “event” over different time frames, with patterns varying by demographic. It would be interesting to examine how many individuals fall into the same happiness category for both the 24-hour and 3-month statements by demographic group, and whether certain demographics are more likely to report consistent happiness across time frames.

Conclusions

There currently are no best practice guidelines to support developers and designers in gathering and defining the stakeholder-driven or humanity-centric requirements, use cases, and goals for DMHIs in select domains. The ethical and responsible design of DMHIs should be evidence-based approaches; use theory-driven design; be tailored to the population and the context; and be considerate of engagement, inclusivity, accessibility, and flexibility, but this represents a significant challenge [22,24,25]. However, the identification of universal and domain-specific needs is also central to user engagement and retention. Coproduction would typically include interviews, focus groups, literature reviews, workshops, panel discussions, and clinical expertise [23]. Of course, such processes are very time, and as such resource intensive [23].

Within their systematic review, Brotherdale et al [23] note the need for creative methods to be adopted in the coproduction of digital mental health intervention, and the use of open datasets may be a fruitful approach in this endeavor. Currently, little attention has been paid to open datasets as a creative approach to requirements gathering and informing content development of PPIs, and indeed DMHIs in general. The HappyDB analysis can be presented as a sample application.

PPIs can be tailored to address diverse challenges and promote well-being, as well as prevent mental ill health by promoting protective factors, enhancing resilience, and fostering well-being [30]. PPIs often involve suggested activities to enhance happiness or well-being [9]. PPIs generally focus on one of five categories: (1) savoring experiences and sensations; (2) cultivating (and sometimes expressing) gratitude; (3) engaging in kind acts; (4) promoting positive relationship processes; and (5) pursuing hope and meaning [31]. Their mode of operation is typically either diaries for self-reflection or applications that make suggestions based on responses to set questions [9]. Some examples relevant to this discussion might be virtual reality interventions that seek to promote happiness by immersing users in positive environments or guided meditations; strength-based platforms that help users to identify and use their personal strengths; positive psychology games that are designed to promote positive emotions; or gratitude apps where users can keep a digital gratitude journal, but there would also be reminders and prompts to help them focus on positive aspects of their lives.

User engagement difficulties across the board represent a major challenge to PPIs, and indeed DMHIs in general, achieving their maximum potential benefits [32]. Designing content that can be personalized based on user data, preferences, and progress

XSL•FO
RenderX

to increase engagement and effectiveness represents a major challenge for design developers, with collaboration across several disciplines required [32]. However, as can be seen earlier, notwithstanding the recognized limitations, open datasets such as the HappyDB provide great insight into how demographics link to differences in sources of happiness. Certain PPIs or suggested activities may have increased meaning for certain subgroups, and open datasets may provide a starting point for understanding what these might be. Research on the science of happiness is also relevant to the concept of intrinsic capacity for healthy aging and holistic well-being. This concept remains undervalidated, and a deeper understanding of happiness could help validate and practically apply it in mental health interventions [33]. Open datasets may also come in various formats. For example, there are many sources that outline counseling dialogues, and how these are rated by users. Through the use of machine learning, these may give insights into useful content generation for DMHIs.

When discussing DMHI content design, it is important to consider the lack of context in interpreting the significance of insights derived from open-source data. This is a key limitation in health technology design. Evaluating user satisfaction and engagement is essential to determine whether a version created using open data is received more positively than one developed through other approaches. Metrics such as usability, satisfaction, and effectiveness should be compared to assess the value of incorporating open data in DMHI development.

Many platforms host open datasets, with the Open Science Framework, Mendeley Data, PsychOpen Community, Archive, and Metadata for Open Science, or Kaggle representing examples. Nonetheless, supporting researchers in providing findable, accessible, interoperable, and reusable data poses significant challenges [8]. For example, what might a framework to describe the nature of the data that are typically gathered across different domains look like? What would the best methods be for structuring data (and metadata) that are human and machine readable but also useful for designers and analysts and for interpretability across different channels? How would datasets be described in terms of purposes, goals, and communication patterns, and what might a structured representational repository look like?

Huston [7] discusses the challenges associated with open data in terms of three categories: (1) making the technological shift; (2) social, cultural, legal, and ethical issues; and (3) avoiding the pitfalls. Nonetheless, a discussion of the extent to which our understanding of data and interactions from nondigital settings may inform requirement gathering for DMHIs is warranted. While the challenges associated with open data are substantial and should not be underestimated, the potential benefits of open data are very exciting [7]. They may play a role in the development of content for digital PPIs that are engaging, effective, and grounded in scientific research.

Data Availability

The HappyDB is a dataset containing over 100,000 happy statements [9] that is freely available on the web [10]. The data set analyzed during this study is available on the Megagon Labs website [34].

Conflicts of Interest

None declared.

References

1. Seligman M. PERMA and the building blocks of well-being. J Posit Psychol 2018;13(4):333-335. [doi: 10.1080/17439760.2018.1437466]
2. Butler J, Kern ML. The PERMA-Profiler: a brief multidimensional measure of flourishing. Int J Wellbeing 2016;6(3):1-48. [doi: 10.5502/ijw.v6i3.526]
3. Medvedev ON, Landhuis CE. Exploring constructs of well-being, happiness and quality of life. PeerJ 2018;6:e4903 [FREE Full text] [doi: 10.7717/peerj.4903] [Medline: 29876148]
4. Seligman MEP. Flourish: A Visionary New Understanding of Happiness and Well-Being. Mumbai, India: Free Press; 2011.
5. Seligman MEP. Positive psychology: a personal history. Annu Rev Clin Psychol 2019;15:1-23. [doi: 10.1146/annurev-clinpsy-050718-095653] [Medline: 30525996]
6. Diener E, Suh EM, Lucas RE, Smith HL. Subjective well-being: three decades of progress. Psychol Bull 1999;125(2):276-302. [doi: 10.1037//0033-2909.125.2.276]
7. Huston P, Edge VL, Bernier E. Reaping the benefits of open data in public health. Can Commun Dis Rep 2019;45(11):252-256 [FREE Full text] [doi: 10.14745/ccdr.v45i10a01] [Medline: 31647060]
8. Sadeh Y, Denejkina A, Karyotaki E, Lenferink LIM, Kassam-Adams N. Opportunities for improving data sharing and FAIR data practices to advance global mental health. Glob Ment Health 2023;10:e14 [FREE Full text] [doi: 10.1017/gmh.2023.7] [Medline: 37860102]
9. Asai A, Evensen S, Golshan B, Halevy A, Li V, Lopatenko A, et al. HappyDB: A corpus of 100,000 crowdsourced happy moments. 2018 Presented at: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018); May 1, 2018; Miyazaki, Japan URL: https://arxiv.org/abs/1801.07746

XSL•FO
RenderX

10. HappyDB: a happiness database of 100,000 happy moments. GitHub. URL: https://github.com/megagonlabs/HappyDB [accessed 2025-01-07]
11. Oishi S, Gilbert EA. Current and future directions in culture and happiness research. Curr Opin Psychol 2016;8:54-58. [doi: 10.1016/j.copsyc.2015.10.005] [Medline: 29506804]
12. Pena-López A, Rungo P, López-Bermúdez B. The "Efficiency" effect of conceptual referents on the generation of happiness: a cross-national analysis. J Happiness Stud 2020;22(6):2457-2483. [doi: 10.1007/s10902-020-00328-3]
13. Mohamed E, Mostafa SA. Computing happiness from textual data. Stats 2019;2(3):347-370. [doi: 10.3390/stats2030025]
14. Alkan Ö, Kavalci K. Gender differences in happiness during COVID-19 pandemic: evidence from Türkiye. J Psychiatry Behav Sci 2023;13(3):153-165. [doi: 10.5455/pbs.20230512053548]
15. Karaaslan K. Analysis of factors affecting individuals' sources of happiness with multinomial logistic model. J Meas Eval Educ Psychol 2021;12(3):286-302 [FREE Full text] [doi: 10.21031/epod.925631]
16. Metzger S, Gracia P. Gender differences in mental health following the transition into parenthood: longitudinal evidence from the UK. Adv Life Course Res 2023;56:100550. [doi: 10.1016/j.alcr.2023.100550] [Medline: 38054884]
17. Nomaguchi K, Milkie MA. Parenthood and well-being: a decade in review. J Marriage Fam 2020;82(1):198-223 [FREE Full text] [doi: 10.1111/jomf.12646] [Medline: 32606480]
18. Preisner K, Neuberger F, Bertogg A, Schaub JM. Closing the happiness gap: the decline of gendered parenthood norms and the increase in parental life satisfaction. Gender Society 2019;34(1):31-55. [doi: 10.1177/0891243219869365]
19. Maharlouei N, Cobb S, Bazargan M, Assari S. Subjective health and happiness in the United States: gender differences in the effects of socioeconomic status indicators. J Ment Health Clin Psychol 2020;4(2):8-17 [FREE Full text] [doi: 10.29245/2578-2959/2020/2.1196] [Medline: 32568256]
20. Braithwaite S, Holt-Lunstad J. Romantic relationships and mental health. Curr Opin Psychol 2017;13:120-125. [doi: 10.1016/j.copsyc.2016.04.001] [Medline: 28813281]
21. Walsh LC, Boehm JK, Lyubomirsky S. Does happiness promote career success? Revisiting the evidence. J Career Assess 2018;26(2):199-219. [doi: 10.1177/1069072717751441]
22. van Zyl LE, Efendic E, Rothmann S, Shankland R. Best-practice guidelines for positive psychological intervention research design. In: van Zyl L, Rothmann S, editors. Positive Psychological Intervention Design and Protocols for Multi-Cultural Contexts. Cham, Switzerland: Springer; 2019.
23. Brotherdale R, Berry K, Branitsky A, Bucci S. Co-producing digital mental health interventions: a systematic review. Digit Health 2024;10:20552076241239172 [FREE Full text] [doi: 10.1177/20552076241239172] [Medline: 38665886]
24. Berardi C, Antonini M, Jordan Z, Wechtler H, Paolucci F, Hinwood M. Barriers and facilitators to the implementation of digital technologies in mental health systems: a qualitative systematic review to inform a policy framework. BMC Health Serv Res 2024;24(1):243 [FREE Full text] [doi: 10.1186/s12913-023-10536-1] [Medline: 38408938]
25. Lipschitz JM, Pike CK, Hogan TP, Murphy SA, Burdick KE. The engagement problem: a review of engagement with digital mental health interventions and recommendations for a path forward. Curr Treat Options Psychiatry 2023;10(3):119-135 [FREE Full text] [doi: 10.1007/s40501-023-00297-3] [Medline: 38390026]
26. Alkan Ö, Güney E. Investigation of factors that affect the frequency of alcohol use of employees in Turkey. J Subst Use 2020;26(5):468-474. [doi: 10.1080/14659891.2020.1846811]
27. Buijs VL, Jeronimus BF, Lodder G, Steverink N, de Jonge P. Social needs and happiness: a life course perspective. J Happiness Stud 2020;22(2):1953-1978. [doi: 10.1007/s10902-020-00287-9]
28. Nelson SK, Kushlev K, Lyubomirsky S. The pains and pleasures of parenting: When, why, and how is parenthood associated with more or less well-being? Psychol Bull 2014;140(3):846-895. [doi: 10.1037/a0035444] [Medline: 24491021]
29. Abdullahi AM, Orji R, Rabiu AM, Kawu AA. Personality and subjective well-being: towards personalized persuasive interventions for health and well-being. Online J Public Health Inform 2020;12(1):e1 [FREE Full text] [doi: 10.5210/ojphi.v12i1.10335] [Medline: 32547678]
30. Carr A, Cullen K, Keeney C, Canning C, Mooney O, Chinseallaigh E, et al. Effectiveness of positive psychology interventions: a systematic review and meta-analysis. J Posit Psychol 2021;16(6):749-769. [doi: 10.1080/17439760.2020.1818807]
31. Schueller SM, Parks AC. The science of self-help: translating positive psychology research into increased individual happiness. Eur Psychol 2014;19(2):145-155. [doi: 10.1027/1016-9040/a000181]
32. Saleem M, Kühne L, de Santis KK, Christianson L, Brand T, Busse H. Understanding engagement strategies in digital interventions for mental health promotion: scoping review. JMIR Ment Health 2021;8(12):e30000 [FREE Full text] [doi: 10.2196/30000] [Medline: 34931995]
33. Hoogendijk EO, Dent E, Koivunen K. Intrinsic capacity: An under-researched concept in geriatrics. Age Ageing 2023;52(10):afad183. [doi: 10.1093/ageing/afad183] [Medline: 37782890]
34. HappyDB: a happiness database of 100,000 happy moments. Megagon Labs. URL: https://megagon.ai/happydb-a-happiness-database-of-100000-happy-moments/ [accessed 2025-01-27]

Abbreviations

DMHI: digital mental health intervention

XSL•FO
RenderX

PPI: positive psychology intervention
VIF: variance inflation factor

Edited by A Mavragani; submitted 21.08.24; peer-reviewed by Ö Alkan, N Hird; comments to author 10.12.24; revised version received 17.12.24; accepted 18.12.24; published 29.01.25.

Please cite as:
Ennis E, Bond R, Mulvenna M, Sweeney C
Understanding Individual Differences in Happiness Sources and Implications for Health Technology Design: Exploratory Analysis of an Open Dataset
JMIR Form Res 2025;9:e65658
URL: https://formative.jmir.org/2025/1/e65658
doi:10.2196/65658
PMID:

©Edel Ennis, Raymond Bond, Maurice Mulvenna, Colm Sweeney. Originally published in JMIR Formative Research (https://formative.jmir.org), 29.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Experiences of Peer Mentoring Sexual and Gender Minority Emerging Adults Who Are at Risk for Suicide: Mixed Methods Study

Jennifer T Tran[1], PhD; Jessica Webster[1], MSc; James R Wolfe[1], MSc; Jennifer Ben Nathan[1], BA; Lindiwe Mayinja[2], BS; Marin Kautz[2], PhD; Maria A Oquendo[2], MD, PhD; Gregory K Brown[2], PhD; David Mandell[2], ScD; Danielle Mowery[3], PhD; José A Bauermeister[1], MPH, PhD; Lily A Brown[2], PhD

[1]Department of Family and Community Health, School of Nursing, University of Pennsylvania, Philadelphia, PA, United States

[2]Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[3]Department of Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Jennifer T Tran, PhD
Department of Family and Community Health
School of Nursing
University of Pennsylvania
418 Curie Blvd
Philadelphia, PA, 19104
United States
Phone: 1 2158983616
Fax: 1 215 573 8439.
Email: jtgtran@nursing.upenn.edu

Abstract

Background: Sexual and Gender Diverse Youth (SGDY) are at increased risk for suicide due to unique experiences including discrimination, family or friend rejection, and low positive affect. Peer mentors (PMs) may offer a unique opportunity for intervention but are underutilized for suicide prevention among SGDY.

Objective: Little is known about the training needed for PMs when working with SGDY at risk for suicide. We developed an intervention, Supporting Transitions to Adulthood and Reducing Suicide (STARS), to improve suicide prevention among SGDY and increase social support, coping, and positive effects. PMs were trained by a licensed clinical therapist and provided a manual. PMs meet virtually for 6 weeks, providing social support, strategies to diminish the impact of discrimination, connection to safe spaces, and reinforcement of intentions to use Safety Plans with mentees.

Methods: To understand PMs' experiences in their role, including distress, fidelity to the manual, and perceptions of feasibility and acceptability of STARS and mentees' Safety Plan, we collected survey data from mentees and PMs as well as in-depth interviews with PMs after the completion of the intervention.

Results: As of September 2024, all peer mentees (N=64) have completed the study and all PMs have finished providing sessions for peer mentees. PMs (n=5) reported overall high comfort (8.52) and low distress (1.93) during sessions. All 5 PMs had high fidelity (>90%) to the PM intervention training. All 5 PMs reported high feasibility (17.50), acceptability (20), and appropriateness (20) of the STARS intervention. Mentees (n=27) reported high confidence ratings (3.54) in speaking with their PMs.

Conclusions: Peer mentorship for SGDY who are at risk for suicide was feasible and acceptable by PMs and mentees alike. PMs reported that they felt comfortable and confident during the sessions. Mentees also reported confidence in working with their PMs. Future research should explore the optimal strategies to support PMs and mentees as they engage in suicide prevention work as well as incorporate feedback from the PMs in this study to ensure optimal outcomes.

Trial Registration: ClinicalTrials.gov NCT05018143; https://clinicaltrials.gov/study/NCT05018143

International Registered Report Identifier (IRRID): RR2-10.2196/48177

(JMIR Form Res 2025;9:e67814) doi:10.2196/67814

XSL•FO
RenderX

KEYWORDS

suicide prevention; peer mentorship; LGBTQIA health; mental health

Introduction

Emerging adulthood is a critical period of suicide risk for Sexual and Gender Diverse Youth (SGDY) people [1]. A meta-analysis indicated 11%-20% of sexual and gender diverse (SGD) individuals have a lifetime prevalence of suicide attempts compared with 4% of heterosexual adults [2]. Rates of suicide attempts may be higher; however, as these statistics only include SGDY who have self-reported their sexual and gender minority identity and ignore those who have not "come out" and are at even greater risk for suicidal ideation [2]. A meta-analysis of SGDY and emerging adults (12-20 years old) found that bisexual and transgender youth were at the most significant risk of attempting suicide (odds ratios of 4.87 and 5.87, respectively) compared with their cisgender heterosexual peers [3].

Sexually diverse emerging adults (eg, lesbian, gay, bisexual, transgender, and queer/questioning [LGBTQ]) may present with unique individual-level distal risk factors for suicide, such as the age of coming out, violence, exposure to sexual orientation change or conversion efforts, and discrimination, and proximal risk factors for suicide, such as internalized homophobia, expectations of rejection, and level of social support, which all interact to either increase or decrease the risk for suicide depending on the social environment and the individual's perceived belongingness within said environment [4-7]. Rooted in both the minority stress theory and the interpersonal theory of suicide, previous work has indicated that sexual and gender minority stress for adolescents and emerging adults is associated with suicidal ideation and attempts through perceived burdensomeness and thwarted belongingness [8]. To address feelings of burdensome and thwarted belongingness, increasing a sense of connectedness for SGM emerging adults can be an important protective factor against suicide, even when accounting for experiences of discrimination and victimization [9].

PM may serve as a strategy to provide social support, promote positive affect and feelings of belongingness, and build a connection to the community. Given at-risk SGM individuals' tendency to disclose suicidal ideation to peers rather than professionals, peers are crucial actors in suicide prevention [10,11]. Previous research demonstrated that most individuals contemplating suicide, particularly adolescents and young adults, do not seek help from formal support structures due to stigma concerns, but instead use informal resources like friends and family [12]. Therefore, educating peers about suicide, depression, and the resources available for at-risk individuals may reduce their reluctance to intervene and increase their ability to do so appropriately.

A review shows that peer mentoring can effectively promote health behavior changes in adolescents and emerging adults through reinforcing coping skills, incorporating skill-building activities, and delivering social support [13]. In studies of suicide prevention models, peer mentorship decreased stigma related to help-seeking behaviors and reduced rates of repeated psychiatric hospitalization [14,15]. A pilot trial of a peer-delivered safety planning intervention (SPI) found that adults who received the peer-delivered version had fewer emergency room visits during the follow-up period than those who received the provider-delivered version [16]. Those with sexual minority identities reported reluctance to access mental health services for fear of discrimination and dismissal of their emotional distress [17,18]. Therefore, approaches using peer mentorship that decrease stigma and decrease the need for repeated service engagement may be critical for reducing suicide risk in the SGM community.

For SGDY, effective peer mentor (PM) interventions should incorporate several design characteristics to prevent future suicide risk. First, they should target social support [19]. Second, they should involve skill-based peer-administered interventions instead of purely supportive interventions [20,21]. Third, they should market interventions as LBGTQ (lesbian, gay, bisexual, transgender, and queer/questioning)-affirming [22]. Fourth, they should emphasize the use of a safety plan, which reduces the risk of suicide attempts [23]. When SGM emerging adults are reluctant to use their safety plan, peer mentorship should be enhanced with principles of motivational interviewing (MI). MI is a counseling method designed to elicit behavior change [24-27]. MI can be successfully used in peer-based interventions [28-30]. Given the need for peer mentor interventions to consider a multitude of important design characteristics and targets, there is a need to understand the challenges for PMs. Research to understand the experiences of PMs who deliver suicide prevention interventions to SGDY would be instructive.

In this paper, we describe the protocol for training PMs to fidelity in a suicide prevention intervention (Supporting Transitions to Adulthood and Reducing Suicide [STARS]) and detail their experiences in delivering MI- and cognitive behavioral therapy (CBT)–based content to SGDY mentees. We hypothesized that PMs would retain high fidelity to the protocol throughout their participation in the project, operationalized as greater than 80% fidelity to critical components of each session. We report on peer mentor distress and comfort ratings (completed after each peer mentor session) and evaluate changes in distress and comfort throughout the trial. We hypothesized that as PMs gained more experience with the protocol and the population, distress ratings would significantly decrease over time and comfort ratings would increase over time. Finally, we completed qualitative interviews with PMs about their experiences serving in this role. We hypothesized that PMs would report that the STARS intervention was acceptable, feasible, and appropriate.

Methods

The STARS Intervention

This study was part of a larger university randomized control trial (NCT05018143) evaluating an app-based intervention STARS aimed to reduce suicidal ideation and behaviors among SGDY (ages 18-24 years) [31]. Eligible participants were

XSL•FO
RenderX

between the ages of 18 and 24 years, lived in the Philadelphia Metro area, reported no psychotic symptoms (hallucinations or delusions), had access to a smartphone and Wi-Fi, and had active suicidal ideation in the past month. Participants were randomized to receive an in-person brief, evidence-based SPI (control arm) or to receive the SPI plus access to STARS (intervention arm). Mentees randomized to the STARS intervention had access to a mobile app focused on the provision of life skills and their safety plan and 6 peer mentor sessions. More detailed information on the more extensive study can be found in the study by Brown et al [31].

PM Training Plan

PMs completed eight 2-hour training sessions with 2 licensed mental health providers. The first session included an overall introduction to MI. PMs received training in core MI concepts, with an emphasis on skill use. We provided a 2-hour training covering ambivalence, MI spirit (ie, acceptance, partnership, compassion, and evocation), and the righting reflex (the inclination helpers have to give advice, correct what they see as wrong information, or the wrong reasoning, and generally fix things for the person they are helping). OARS (ie, open-ended questions, affirmations, reflections, and summaries) skills were covered in a second 2-hour training. A third 2-hour training covered the 4 MI processes (ie, engaging, focusing, evoking, and planning) and the Helping Roadmap model [32]. PMs were assigned exercises after each training session to support their acquisition of skills. These trainings were adapted from materials used in the iReach trial [28] and were led by a masters-prepared research staff member (project manager and research coordinator) with a background in mental health counseling and delivering MI training for PMs.

After training and practice in the style of MI, PMs received 62-hour training sessions on the content of PM sessions. These trainings included a didactic overview with questions and answers, modeled roleplay, and then an observed experiential role-play between the PMs with feedback from the study clinicians. After training in the MI-style and CBT content of the peer mentor intervention, additional time was dedicated to navigating difficult scenarios, including exacerbations in suicide risk, risky behaviors in general, boundary crossings, and building rapport and connection (or repairing ruptures in the peer mentor relationship).

After the original training, PMs attended a weekly group consultation meeting that was attended by 1 to 3 study clinicians. In the weekly meeting, PMs provided an update on the mentees to whom they were assigned, discussed difficulties in the session, and provided examples of successes in the PM relationship. PMs provided feedback and ideas to each other, and the study clinicians provided validation, coaching, and support around content areas that PMs indicated a need for more support (ie, what to do in a crisis and how to engage participants).

Approximately 6 months into the start of study recruitment, the team dedicated two 30-minute supervision sessions to reviewing and refreshing PMs' MI style.

Participant Measures: Mentee Feedback

In their follow-up survey (at 2 months), mentees were asked questions about their experience with PM sessions, including: "PM sessions were offered at times that worked for my schedule," and "I feel confident talking to a STARS PM to discuss what's going on in my life." Questions were answered on a Likert scale from 1 (strongly disagree) to 4 (strongly agree).

PM Measures: PM Comfort and Distress

At the end of each PM session, PMs were asked to fill out a post-session form that assessed issues that arose and distress and comfort during the PM sessions. The PM comfort rating included two questions: (1) "On a scale from 1-10 how comfortable did you feel during this PM session?" ("10" is "very comfortable" and "1" is "the most uncomfortable") and (2) "Please rate the highest level of distress you experienced during this session on a scale of 1-10." ("10" is "very distressed" and "1" is "not at all distressed"). Scores on the second question were reverse coded, and the scores for both questions were averaged for a total comfort score ranging from 1 to 10, with higher scores indicating higher comfort.

Implementation Measures

Overview

To determine the extent to which the STARS PM intervention was acceptable, appropriate, and feasible, we adapted the Acceptability of Intervention Measure (AIM), Intervention Appropriateness Measure (IAM), and Feasibility of Intervention Measure (FIM) [33]. The AIM, IAM, and FIM demonstrated strong psychometric properties in a series of 3 studies conducted by Weiner et al [33]. The AIM, IAM, and FIM are 4-item measures asking participants to answer questions on a scale of "1" ("completely disagree") to "5" ("completely agree"). We adapted the scales for the STARS PM intervention; for example, 1 item from the AIM states, "The STARS peer mentoring intervention meets my approval."

PM Qualitative Interview

Semistructured interviews were conducted with all PMs to gain a deeper understanding of PM experiences and views on the STARS PM intervention. A member of the research team, outside of those providing training or supervising the implementation of the peer mentoring intervention, conducted the interviews with PMs virtually. The interviews lasted about 60 minutes and were audio recorded and transcribed verbatim.

We used a Template Analysis approach, a form of thematic analysis emphasizing hierarchical coding [34-36]. An a priori codebook was created informed by the interview guide. Out of 3 researchers (JT, TB, and JW) coded the first transcript using the codebook as a guide and template. After the first transcript, a consensus was reached, and the codebook was edited with final details. The remaining transcripts were divided between 2 researchers and individually coded [34-36].

Fidelity Monitoring

PM fidelity to the intervention was assessed using a checklist that evaluated their adherence to content and style [31]. Content was assessed based on whether PMs covered key elements of

XSL•FO
RenderX

the session (eg, agenda setting for the day's session, checking in about safety plan use since the last session, and teaching the designated CBT skill for the session). Possible total content scores for the sessions ranged from 16 to 29, based on how many elements were meant to be covered in each session. The style was assessed based on the PMs' use of MI skills OARS [25] to facilitate participant learning and engagement with the session content. The style was scored on a scale of 0-5 for whether PMs asked open-ended questions, used affirming statements, used reflections, and used summaries during the session.

PMs' fidelity to the intervention was assessed across all sessions for the first 2 mentees they saw, allowing for early detection of challenges to fidelity and corrective supervision. After completing the intervention with their first 2 mentees, PMs were assessed randomly for 1 in every 6 sessions per mentee. Sessions were randomly selected for fidelity assessment using a randomized list generator. Out of 2 research team members were tasked with completing fidelity ratings, 10% of which were scored twice to assess interrater reliability.

Ethical Considerations

The study was reviewed by the university's institutional review board (protocol # 849500) and was deemed a human-subjects study. Informed consent was obtained from participants in person during the baseline visit. Participants were informed of all the risks and benefits of the study. All information is dedeidentified and all personal information collected is kept private and confidential. Participants received up to US $170 (US $50 for baseline, US $30 for 2 months, US $40 for 4 months, and US $50 for 6 months) through physical or electronic ClinCard.

Results

Mentee Description

As of September 2024, all peer mentees (N=64) have completed the study. A total of 32 mentees were randomized to receive the STARS app intervention. One mentee opted not to participate in the PM sessions. Out of 4 mentees did not complete all 6 recommended PM sessions. Of the 4 mentees who did not complete all 6 sessions, 2 only completed PM session 1, 1 completed 4 PM sessions, and 1 completed 5 PM sessions. A total of 27 mentees completed all 6 PM sessions.

PM Description

PMs were recruited from the Greater Philadelphia, Pennsylvania area, through job announcements through area university student employment boards, university bulletin boards, and network connections. Applicants were interviewed based on previous education, experience with mentoring, mental health, and LGBTQ advocacy. PMs selected needed to have some previous experience in mentoring; however, other aspects were not required as we intentionally wanted a heterogeneity of PMs in educational or professional experiences parallel to participants in the study. Although we never asked specific questions of applicants or employees, given human resources guidelines, we can provide descriptive information about the 5 PMs who delivered sessions, which we obtained from working closely with them. At hiring, PMs were between the ages of 22 to 26, self-identified as a part of a sexual and gender minority community orally, and over half identified as a racial and ethnic minority.

Out of 5 PMs delivered sessions after completing their training. PMs were assigned to study mentees who were randomized to the STARS intervention. PMs worked with no more than 3 STARS mentees at a time. STARS participants were assigned to PMs based on a list and caseload to prevent overburdening PMs.

Fidelity to PM Training

PM fidelity checks were completed by 2 researchers (LB and JW), both trained in therapeutic peer mentorship. PM fidelity to content was calculated for each of the 6 peer mentor sessions. Fidelity percentages were high, ranging from 95.7% to 100%. PM fidelity to style was calculated for each of the 6 PM sessions and was high, ranging from 92.1% to 97.4%.

Mentee Feedback on Intervention

STARS mentees provided ratings for PM sessions at 2 months. Mentees (N=28) rated their confidence in talking with a PM as 3.54 (SD 0.64) out of a possible confidence rating of 4.0. Mentees agreed that sessions were offered at times that worked for their schedule as 3.74 (SD 0.53) out of a score of 4.0.

PM Outcomes

PM Implementation Outcomes

PMs (n=5) rated the sessions with a feasibility score (FIM) of 16.40 (SD 3.21), an appropriateness score (IAM) of 18.40 (SD 3.58), and an acceptability score (AIM) of 18.40 (SD 3.58) with a possible maximum score of 20 for all implementation outcomes (Table 1).

XSL•FO
RenderX

Table 1. Average fidelity ratings to the training of 6 peer mentor (PM) sessions of a mixed methods study on suicide prevention in sexual and gender-diverse youth.

Measure	Session Time Point					
	1	2	3	4	5	6
Content total scores						
Mean (SD)	15.21 (0.70	20.93 (1.49)	20.33 (1.11)	21.38 (1.76)	17.54 (1.13)	26.71 (0.73)
Maximum	16	22	21	23	18	29
Style total scores						
Mean (SD)	20 (0)	19.80 (0.77)	19.87 (0.35)	19.92 (0.28)	19.77 (0.60)	19.14 (1.83)
Maximum	20	20	20	20	20	20

PM Comfort and Distress

PMs indicated high comfort ratings after each session and low distress. Paired samples *t* tests indicated that there were no significant differences in comfort (P=.44) or distress ratings (P=.70) from session 1 to session 6 (Table 2).

Table 2. Average comfort and distress ratings self-reported by peer mentors (PMs) for each of the 6 PM sessions in a mixed methods study on a suicide prevention intervention for sexual and gender-diverse youth.

Measure	Session time point					
	1	2	3	4	5	6
Comfort ratings						
N	31	29	29	29	28	27
Mean (SD)	8.81 (0.79)	8.45 (1.33)	8.38 (1.40)	8.41 (1.18)	8.89 (1.07)	8.52 (1.60)
Distress ratings						
N	31	29	29	29	28	27
Mean (SD)	1.81 (1.22)	2.52 (1.79)	2.59 (2.15)	2 (0.85)	1.96 (1.04)	1.93 (0.87)

PM Experiences

Thematic analysis of the interviews with PMs categorized their experiences into 6 themes: decision for peer mentoring, training experiences, use of the pm manual, interpersonal relationship with mentee, internal struggles and growth, and STARS App Thoughts.

Decisions for Peer Mentoring

This theme represents background information on the PMs and their motivations for becoming involved in STARS. PMs worked with STARS for 8 months to 18 months. All PMs indicated they were interested in the role due to their career goals related to mental health, LGBTQ+ research, or clinical work. Many PMs reported wanting to give back or support the LGBTQ+ community, due to their connection to the community. One PM stated, “I was really interested in the study itself. I thought it just sounded like a really cool and meaningful idea that would definitely be helpful for the participants.” While another PM said “reason is my self-identification as a cisgender gay man, and that aligns with my value to get back to the community to support all of them.”

Training Experiences

This theme represented information on PMs’ experiences of PM training and how prepared they felt to provide PM sessions to mentees for STARS. All PMs reported that their training was helpful, specifically the role-playing sessions. One PM stated, “I felt like I was very prepared to do everything in relation to the like discussing safety plans and sort of working to brainstorm a safety plan that wasn’t quite working for somebody that I felt like we did so much kind of prep work on, and that was always one of that was like one of my favorite parts, always.” Another PM said, “I think doing the mock sessions was definitely helpful. Where I would be pretending to be the participant in one, and then [another PM] would do the opposite for me. I think that was definitely really helpful.” Some PMs spoke about the importance of having weekly meetings (3/5 PMs), particularly having that space to talk to supervisors and other PMs about their experiences and troubleshooting. A PM stated, “It was nice to have, you know, the weekly meeting and be able to share difficult things that would come up.” One PM indicated a possibility for more practice during training periods. Several PMs noted specific skills or topics they needed clarification on, such as engagement with their STARS mentees. All PMs noted that the continued training (MI) was helpful.

Use of PM Manual

PMs were asked to describe their experiences delivering the STARS sessions to trial mentees. This theme includes feedback on barriers, facilitators, and suggestions for improvement in future STARS manualized session content implementation. Overall, PMs described having “good experiences” delivering the session content. Out of 2 noted session lengths varied, attributed to content and the interactions with individual

XSL•FO
RenderX

mentees. One PM explains, “I think that some of the sessions felt a little bit more like jam-packed with content than others, but that also depended on the person that you were with, and some people really had nothing to say-- and so you could get through it pretty easily. And with some people that was harder.” One PM expressed concern that mentees “zoned out” when delivering content where they had to “talk too much” per the session script. Out of 3 PMs described feeling “more comfortable” delivering the session scripts with each new mentee experience. For most PMs, the main barrier to delivering session content was covering session content within the 30-minute timeframe. Out of 3 PMs described an imbalance in the length of content from session to session as a barrier. One explained, “Delivering the content. Yeah, that would be one thing that I feel like I usually take a long time in delivering session 5 or 6, and I find it hard to squeeze the time to like 25 or 30. So that can be one thing.”

Out of 3 PMs identified the script in the manual as a primary facilitator to content delivery, describing it as “the way I guided people” and “easy to follow along” during sessions. Another explained, “How it was broken up I feel like it made sense. There was like a good flow to it. And it was nice to have examples of how to phrase things, or how to talk about something or questions to ask. Yeah. So that was helpful.” PMs identified flexibility in content and session structure as potential areas for improving STARS implementation. One PM suggested differentiating between essential and optional session content to help adhere to the time limit of sessions. Similarly, another expressed that increasing content flexibility would allow PMs to make sessions more relevant for individual STARS mentees. One PM noted that “if there were things that I felt like we couldn’t just breeze over or just wasn’t in line with where the script was headed. I think things like that made it hard for me to make sure I was getting all the fidelity items done.”

Interpersonal Relationship With Mentees

This theme described any interpersonal challenges experienced in the mentoring role as well as positive changes and growths that they observed in STARS mentees. Some challenges described by PMs included difficulties with engagement, challenges related to participant personality or behavior, missing appointments, texting during sessions, trouble relating to session content, and not using the safety plan. Of the 3 PMs had experiences where they felt mentees were not engaging with session content. One PM explains, “Some of the mentees who I felt weren’t that engaged would be examples that they brought up…And I feel like some mentees would like, give, just like, very like surface-level, or things that weren’t significant or gonna help them. I don’t know if they were avoiding getting deep with me. And, you know, I didn’t like that we couldn’t kinda get beyond that.”

Mentee personalities and behaviors were sometimes a challenge for PMs. Some PMs found it difficult to balance covering the session content while giving space for sharing with more talkative STARS mentees. During the sessions, a PM described challenges with STARS mentees doing other things on their computers and phones. Out of 2 PMs described challenges with STARS mentees who frequently missed appointments and needed to reschedule. A PM stated, “There were some who were like chronically late, or would forget, needed to reschedule.” Resolutions included sending reminders before sessions, changing times, and discussing barriers to scheduling.

While there were some interpersonal challenges, PMs also described growth in their STARS mentees. Out of 2 PMs identified an increase in openness to sharing experiences as an area of growth among STARS mentees. One PM described an example in which a mentee was initially resistant to being on camera at the first meeting due to negative body image and concerns of being seen by others but eventually engaged fully in video sessions as the relationship developed, indicating a relationship of safety and trust. All 5 PMs identified a willingness to practice and apply session content outside of weekly meetings as a primary growth area among STARS mentees. One PM noted, “The participants, mostly all of them I had like they were actually implementing [session content]. Like all of the participants, that kind of surprised me a bit. “Cause they actually were using the things that [STARS] was wanting to do. Which is the goal. Yeah. It was cool.” In one example, a PM noticed that STARS mentees who gave positive session feedback shared that they were also using the app and practicing content between sessions. Out of 2 PMs described a noticeable shift in mentees’ motivation and engagement around setting and following through on achieving goals. One PM explained, “I mean, a lot of people were reporting, I think, less suicidal ideation or more ease in using a safety plan and more sort of like feeling like they were able to ride the waves of something which was great and hmm. Yeah, I think just a lot of people developed coping skills.”

Internal Struggles and Growth

This theme represented information on the STARS PMs’ internal experiences throughout the study, including emotional responses and reactions to their interactions with mentees, both positive and negative. When discussing concerns about providing peer mentorship to mentees with a history of suicide risk, all 5 PMs mentioned a fear that a mentee would actively be in crisis during a session. They were comforted knowing that they had a script and protocol to follow if something did happen. One PM noted, “One major concern was just like how to go about like if they feel like they’re in crisis while they’re with me. And then another one was how to respond if someone says they want to harm themselves. I remember learning during the sessions.” Despite the safeguards in place, the potential for suicidal crises remained a source of anxiety. Some other internal, emotional, or personal challenges that PMs encountered included worries about their STARS mentees’ health, difficulty hearing about mentees’ history with suicidal ideation, and not being sure if they were making sense or being engaging to their STARS mentees. One person said that these challenges subsided the more “exposure” they had. Another PM felt frustrated by needing to balance helping the STARS mentee with sticking to the script to keep their fidelity rating high.

While PMs expressed some internal struggles, they managed difficult emotions that came up by mentioning them to the principal investigators during weekly team meetings and referring to the script in moments of uncertainty. Out of 2 PMs

XSL•FO
RenderX

mentioned the comfort that the script brought them. A PM stated, “I'm sticking to the scripts, and hitting all the items I need to hit… When those goals would kind of work against what I think is most important, to really just see what is most useful for the participant and see what I could do to help them.” Out of 2 other PMs said it was helpful to remind themselves of the boundaries of this role and that the STARS mentees’ lives outside of PM sessions were not in their control.

PMs also indicated that the experience was rewarding seeing some improvement in their STARS mentees. This ranged from simply feeling helpful and effective in their sessions to feeling like they were imparting something useful to seeing STARS mentees use their safety plan between sessions. A PM stated, “I think it was just like rewarding to be able to feel like I’m imparting something useful to the peer mentees. So it was nice when we would have like, you know, a fruitful discussion that felt like it was something that was helpful. And that they could use in their real life and use to make changes that were helpful to them.” One PM felt rewarded by getting encouraging supervisor feedback and hearing STARS mentees express gratitude. In addition, another PM appreciated the opportunity to interact more with the queer community and learn about experiences different from their own.

Discussion

Principal Findings

This study aimed to evaluate the experiences of PMs who deliver suicide prevention content to SGDY. The study was designed to provide PMs with comprehensive training for working with SGDY at risk for suicide, evaluate PM experiences (reported ratings of comfort and distress), and assess PM fidelity to the protocol (PM training).

It has been well-documented that there is anxiety in providing care for patients at risk for suicide in behavioral health practitioners [37]. Therefore, it is important to assess comfortability and distress with PMs working in suicide prevention interventions. Overall, PMs described high comfort and low distress ratings after each PM session. The comfort and distress ratings did not change over time. These ratings were aligned with the reported experiences of PMs during the follow-up interviews. PMs described having “good experiences” in delivering the session content and several reported that they felt “more comfortable” as they saw more mentees. Many PMs attributed their comfort with sessions by mentioning them during weekly team meetings. Some PMs stated that having the manual was a comfort during the sessions as a resource for them to refer. Therefore, having regular supervision and a thorough resource (such as a manual) are important components to include in future PM based interventions for SGDY at risk for suicide. To our knowledge, this is the first study to report on PM comfort and distress levels when providing support to mentees in a suicide prevention intervention. These results highlight the importance of structured supervision for PMs to aid in providing support to mentees at risk of suicide knowing the anxiety that can surround such a stigmatizing topic [37].

In examining the fidelity of PM sessions with the training, PMs had high fidelity overall (greater than 90% in content and style fidelity ratings). Fidelity (the degree to which an intervention, ie, peer mentoring, was delivered as intended) [38] plays an important role in how well an intervention is considered successful. Fidelity to an intervention provides information that can help determine if the intervention was delivered as intended and therefore any impacts can be more confidently associated with the intervention. PMs provided sessions to STARS mentees with high fidelity; however, PMs noted some challenges. PMs were made aware of fidelity to the training and session content were being assessed and expressed concerns about completing all the fidelity items in the allotted time limit (30-minute sessions). PMs noted that there was a lot of content to cover during the sessions, but they also wanted to provide space and time for their STARS mentees to share updates. PMs expressed frustration with sticking to the script and addressing all the fidelity content items while also preserving fidelity style. However, we did not interview or collect data pertaining to the STARS mentees experience with their PMs and cannot confirm if this was a similar experience for STARS mentees. These concerns from PMs have been noted by mental health professionals who advocate for the use of a manual in treatment [39]. One noted suggestion is the idea of flexibility in fidelity which refers to the implementation of an intervention protocol that contains the core attributes for fidelity but also allows for flexibility for each individual client [40]. A suggestion for future implementation should include highlighting content that is necessary. Determining what content is necessary can come from feedback provided by licensed mental health professionals, PMs, and mentees. Future iterations of the intervention should include flexibility in fidelity [40] within the manual and training by reducing content and increasing space for PMs and STARS mentees to connect for social support.

The STARS PM intervention was developed with several components for a culturally informed design including (1) targeting social support, (2) skill-based training for PMs, (3) focusing on SGM-affirming context, and (4) highlighting safety plan use [19-23]. In our interviews with PMs, we noted how PMs experience delivering the PM intervention to STARS mentees and how they align with the 4 factors incorporated into the intervention design. For the first factor of social support, PMs reported that mentees expressed gratitude and that they noticed improvement in their mentees. However, as mentioned above, we did not directly collect data from STARS mentees to determine if they felt supported by their PM. Future iterations of the study should look to collected data from STARS mentees and their experiences. Second, PMs highlighted the importance and helpfulness of the skill-based training that was provided, with MI and role-playing. All PMs noted that initial and continual training in MI skills were helpful in conducting sessions with STARS mentees. Some even noted even more training in MI would be helpful. Third, PMs all noted that they wanted to be mentors because they wanted to provide and support the SGD community. While we did not assess whether STARS mentees felt affirmed, we can note that the content delivered and provided for were SGD-affirming. Finally, the fourth component of the safety plan use was reiterated within the training and fidelity of the PMs’ sessions with STARS

XSL•FO
RenderX

mentees. During all sessions, PMs were asked to check in with STARS mentees and their use of the safety plan. Some PMs noted the difficulty in how to engage STARS mentees with the safety plan, particularly when mentees noted that they had not looked at or used their safety plan.

Our study provides needed context and implementation information regarding peer-based interventions for suicide prevention for SGDY. In a recent scoping review, Bowersox et al [41] identified very few peer-based interventions for suicide prevention overall. Researchers identified only 9 peer-based interventions aimed at crisis and relapse prevention: however, none of these focused on SGD communities. In a PM program for older adults, Van Orden et al [42] trained senior companions (55 years and older) on topics of confidentiality, reporting requirements, accommodations for disabilities, and common physical and mental health conditions but not in suicide risk. Another program, PREVAIL (Peers for Valued Living), is a trial for adapting peer support delivered intervention for veterans recently hospitalized for suicidal thoughts or behaviors [43]. In qualitative interviews with key parties (veterans with current or recent suicide risk, suicide prevention coordinator, clinicians, peer specialists, and a director of inpatient psychiatry), researchers noted that peers should be trained in topics specific to suicide prevention as well as more general clinical approaches such as MI. These findings inspired the focus in our project of offering practical skills training for reducing suicide risk using a style of MI. There are possible applications of the mentoring model developed in our study (ie, training structure, supervision support, and MI skills) for other communities at risk of suicide, including veterans and older adults.

Limitations and Future Directions

Limitations of our study include the small sample of PMs (n=5). The PMs may not have been representative of the whole pool of persons that would be PMs for our study as we were unable to recruit directly stating PMs with lived experiences of suicidal ideation and identify as a part of sexual and gender minority communities. Another limitation of the study is the lack of feedback from mentees on their experiences with PM sessions. Future studies should include gathering qualitative data from mentees on their experiences. Future directions should include simultaneous mixed methods study design of qualitative interviews with PMs and participants to understand the relationship between peers as it relates to quantitative outcomes (ie, thwarted belongingness) of participants identifying as SGM emerging adults.

Conclusion

PMs are outstanding candidates for delivering suicide prevention services, particularly when working with historically oppressed communities, such as SGDY. In this study, we demonstrated the feasibility and acceptability of training and implementing peer mentorship for SGDY who are at higher risk for suicide. PMs reported high ratings of comfort and low ratings of distress when implementing the STARS sessions. In addition, PMs had high fidelity ratings throughout the intervention. Finally, PMs offered helpful suggestions for improving our STARS app and PM content and style to improve outcomes in the future.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to the sensitive nature of the data collected but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Fish JN, Rice CE, Lanza ST, Russell ST. Is young adulthood a critical period for suicidal behavior among sexual minorities? Results from a US national sample. Prev Sci 2019;20(3):353-365 [FREE Full text] [doi: 10.1007/s11121-018-0878-5] [Medline: 29594980]
2. Hottes TS, Bogaert L, Rhodes AE, Brennan DJ, Gesink D. Lifetime prevalence of suicide attempts among sexual minority adults by study sampling strategies: a systematic review and meta-analysis. Am J Public Health 2016;106(5):e1-e12 [FREE Full text] [doi: 10.2105/ajph.2016.303088]
3. di Giacomo E, Krausz M, Colmegna F, Aspesi F, Clerici M. Estimating the risk of attempted suicide among sexual minority youths: a systematic review and meta-analysis. JAMA Pediatr 2018;172(12):1145-1152 [FREE Full text] [doi: 10.1001/jamapediatrics.2018.2731] [Medline: 30304350]
4. Blosnich JR, Henderson ER, Coulter RWS, Goldbach JT, Meyer IH. Sexual orientation change efforts, adverse childhood experiences, and suicide ideation and attempt among sexual minority adults, United States, 2016–2018. Am J Public Health 2020;110(7):1024-1030 [FREE Full text] [doi: 10.2105/ajph.2020.305637]
5. Hong JS, Espelage DL, Kral MJ. Understanding suicide among sexual minority youth in America: an ecological systems analysis. J Adolesc 2011;34(5):885-894 [FREE Full text] [doi: 10.1016/j.adolescence.2011.01.002] [Medline: 21303716]
6. Meyer IH, Russell ST, Hammack PL, Frost DM, Wilson BDM. Minority stress, distress, and suicide attempts in three cohorts of sexual minority adults: A U.S. probability sample. PLoS One 2021;16(3):e0246827 [FREE Full text] [doi: 10.1371/journal.pone.0246827] [Medline: 33657122]

XSL•FO
RenderX

7. Plöderl M, Sellmeier M, Fartacek C, Pichler EM, Fartacek R, Kralovec K. Explaining the suicide risk of sexual minority individuals by contrasting the minority stress model with suicide models. Arch Sex Behav 2014;43(8):1559-1570 [FREE Full text] [doi: 10.1007/s10508-014-0268-4] [Medline: 24573399]
8. Fulginiti A, Goldbach JT, Mamey MR, Rusow J, Srivastava A, Rhoades H, et al. Integrating minority stress theory and the interpersonal theory of suicide among sexual minority youth who engage crisis services. Suicide Life Threat Behav 2020;50(3):601-616 [FREE Full text] [doi: 10.1111/sltb.12623] [Medline: 32048340]
9. Busby DR, Horwitz AG, Zheng K, Eisenberg D, Harper GW, Albucher RC, et al. Suicide risk among gender and sexual minority college students: the roles of victimization, discrimination, connectedness, and identity affirmation. J Psychiatr Res 2020;121:182-188 [FREE Full text] [doi: 10.1016/j.jpsychires.2019.11.013] [Medline: 31837538]
10. Aseltine RH, DeMartino R. An outcome evaluation of the SOS suicide prevention program. Am J Public Health 2004;94(3):446-451 [FREE Full text] [doi: 10.2105/ajph.94.3.446] [Medline: 14998812]
11. Cross WF, Seaburn D, Gibbs D, Schmeelk-Cone K, White AM, Caine ED. Does practice make perfect? A randomized control trial of behavioral rehearsal on suicide prevention gatekeeper skills. J Prim Prev 2011;32(3-4):195-211 [FREE Full text] [doi: 10.1007/s10935-011-0250-z] [Medline: 21814869]
12. Michelmore L, Hindley P. Help-seeking for suicidal thoughts and self-harm in young people: a systematic review. Suicide Life Threat Behav 2012;42(5):507-524. [doi: 10.1111/j.1943-278X.2012.00108.x] [Medline: 22889130]
13. Petosa RL, Smith LH. Peer mentoring for health behavior change: a systematic review. American Journal of Health Education 2014;45(6):351-357 [FREE Full text] [doi: 10.1080/19325037.2014.945670]
14. Sledge WH, Lawless M, Sells D, Wieland M, O'Connell MJ, Davidson L. Effectiveness of peer support in reducing readmissions of persons with multiple psychiatric hospitalizations. Psychiatr Serv 2011;62(5):541-544 [FREE Full text] [doi: 10.1176/ps.62.5.pss6205_0541] [Medline: 21532082]
15. Wyman PA, Brown CH, Inman J, Cross W, Schmeelk-Cone K, Guo J, et al. Randomized trial of a gatekeeper program for suicide prevention: 1-year impact on secondary school staff. Journal of Consulting and Clinical Psychology 2008;76(1):104-115 [FREE Full text] [doi: 10.1037/0022-006x.76.1.104]
16. Wilson MP, Waliski A, Thompson RG. Feasibility of peer-delivered suicide safety planning in the emergency department: results from a pilot trial. Psychiatr Serv 2022;73(10):1087-1093 [FREE Full text] [doi: 10.1176/appi.ps.202100561] [Medline: 35502515]
17. McDermott E, Hughes E, Rawlings V. Norms and normalisation: understanding lesbian, gay, bisexual, transgender and queer youth, suicidality and help-seeking. Cult Health Sex 2018;20(2):156-172 [FREE Full text] [doi: 10.1080/13691058.2017.1335435] [Medline: 28641479]
18. Mirza SA, Rooney C. Center for American Progress, 18. Discrimination prevents LGBTQ people from accessing health care. URL: https://www.americanprogress.org/article/discrimination-prevents-lgbtq-people-accessing-health-care/ [accessed 2018-01-18]
19. McDonald K. Social support and mental health in LGBTQ adolescents: a review of the literature. Issues Ment Health Nurs 2018;39(1):16-29. [doi: 10.1080/01612840.2017.1398283] [Medline: 29333899]
20. Bryan AE, Arkowitz H. Meta-analysis of the effects of peer-administered psychosocial interventions on symptoms of depression. Am J Community Psychol 2015;55(3-4):455-471 [FREE Full text] [doi: 10.1007/s10464-015-9718-y] [Medline: 25861883]
21. Pfeiffer PN, Heisler M, Piette JD, Rogers MA, Valenstein M. Efficacy of peer support interventions for depression: a meta-analysis. Gen Hosp Psychiatry 2011;33(1):29-36 [FREE Full text] [doi: 10.1016/j.genhosppsych.2010.10.002] [Medline: 21353125]
22. Goldbach JT, Rhoades H, Green D, Fulginiti A, Marshal MP. Is there a need for LGBT-Specific suicide crisis services? Crisis 2019;40(3):203-208 [FREE Full text] [doi: 10.1027/0227-5910/a000542] [Medline: 30109965]
23. Stanley B, Brown GK, Brenner LA, Galfalvy HC, Currier GW, Knox KL, et al. Comparison of the safety planning intervention with follow-up vs usual care of suicidal patients treated in the emergency department. JAMA Psychiatry 2018;75(9):894-900 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.1776] [Medline: 29998307]
24. Budhwani H, Naar S. Training providers in motivational interviewing to promote behavior change. Pediatr Clin North Am 2022;69(4):779-794 [FREE Full text] [doi: 10.1016/j.pcl.2022.04.008] [Medline: 35934499]
25. Miller W, Rollnick S. Motivational Interviewing: Helping people change (3rd ed.). New York: Guilford Press; 2013.
26. Naar-King S. Motivational interviewing in adolescent treatment. Can J Psychiatry 2011;56(11):651-657 [FREE Full text] [doi: 10.1177/070674371105601103] [Medline: 22114919]
27. Naar-King S, Suarez M. Motivational interviewing with adolescents and young adults. New York: Guilford Press; 2011.
28. Bonar EE, Wolfe JR, Drab R, Stephenson R, Sullivan PS, Chavanduka T, et al. Training young adult peers in a mobile motivational interviewing-based mentoring approach to upstream HIV prevention. Am J Community Psychol 2021;67(1-2):237-248 [FREE Full text] [doi: 10.1002/ajcp.12471] [Medline: 33137221]
29. Mastroleo NR, Magill M, Barnett NP, Borsari B. A pilot study of two supervision approaches for peer-led alcohol interventions with mandated college students. J Stud Alcohol Drugs 2014;75(3):458-466 [FREE Full text] [doi: 10.15288/jsad.2014.75.458] [Medline: 24766758]

XSL•FO
RenderX

30. Naar-King S, Outlaw A, Green-Jones M, Wright K, Parsons JT. Motivational interviewing by peer outreach workers: a pilot randomized clinical trial to retain adolescents and young adults in HIV care. AIDS Care 2009;21(7):868-873 [FREE Full text] [doi: 10.1080/09540120802612824] [Medline: 20024744]
31. Brown LA, Webster JL, Tran JT, Wolfe JR, Golinkoff J, Patel E, et al. A suicide prevention intervention for emerging adult sexual and gender minority groups: protocol for a pilot hybrid effectiveness randomized controlled trial. JMIR Res Protoc 2023;12:e48177 [FREE Full text] [doi: 10.2196/48177] [Medline: 37773618]
32. Young ME. Learning the art of helping: building blocks and techniques (5th ed). Hoboken: Upper Saddle River, NJ: Prentice Hall; 2013.
33. Weiner BJ, Lewis CC, Stanick C, Powell BJ, Dorsey CN, Clary AS, et al. Psychometric assessment of three newly developed implementation outcome measures. Implement Sci 2017;12(1):108-112 [FREE Full text] [doi: 10.1186/s13012-017-0635-3] [Medline: 28851459]
34. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qualitative research in sport, exercise and health 2019;11(4):589-597. [doi: 10.1080/2159676x.2019.1628806]
35. Brooks J, McCluskey S, Turley E, King N. The utility of template analysis in qualitative psychology research. Qual Res Psychol 2015;12(2):202-222 [FREE Full text] [doi: 10.1080/14780887.2014.955224] [Medline: 27499705]
36. Byrne D. A worked example of braun and clarke's approach to reflexive thematic analysis. Qual Quant 2021;56(3):1391-1412 [FREE Full text] [doi: 10.1007/s11135-021-01182-y]
37. Becker-Haimes EM, Klein CC, Frank HE, Oquendo MA, Jager-Hyman S, Brown GK, et al. Clinician maladaptive anxious avoidance in the context of implementation of evidence-based interventions: a commentary. Front Health Serv 2022;2 [FREE Full text] [doi: 10.3389/frhs.2022.833214] [Medline: 36382152]
38. Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S. A conceptual framework for implementation fidelity. Implement Sci 2007;2:40-49 [FREE Full text] [doi: 10.1186/1748-5908-2-40] [Medline: 18053122]
39. Kendall PC, Beidas RS. Smoothing the trail for dissemination of evidence-based practices for youth: flexibility within fidelity. Professional Psychology: Research and Practice 2007;38(1):13-20 [FREE Full text] [doi: 10.1037/0735-7028.38.1.13]
40. Kendall PC, Frank HE. Implementing evidence-based treatment protocols: Flexibility within fidelity. Clin Psychol (New York) 2018;25(4):e12271 [FREE Full text] [doi: 10.1111/cpsp.12271] [Medline: 30643355]
41. Bowersox NW, Jagusch J, Garlick J, Chen JI, Pfeiffer PN. Peer-based interventions targeting suicide prevention: a scoping review. Am J Community Psychol 2021;68(1-2):232-248 [FREE Full text] [doi: 10.1002/ajcp.12510] [Medline: 33720444]
42. Van Orden KA, Stone DM, Rowe J, McIntosh WL, Podgorski C, Conwell Y. The senior connection: design and rationale of a randomized trial of peer companionship to reduce suicide risk in later life. Contemp Clin Trials 2013;35(1):117-126 [FREE Full text] [doi: 10.1016/j.cct.2013.03.003] [Medline: 23506973]
43. Lapidos A, Abraham KM, Jagusch J, Garlick J, Walters H, Kim HM, et al. Peer mentorship to reduce suicide attempts among high-risk adults (PREVAIL): rationale and design of a randomized controlled effectiveness-implementation trial. Contemp Clin Trials 2019;87:105850 [FREE Full text] [doi: 10.1016/j.cct.2019.105850] [Medline: 31525489]

Abbreviations

AIM: Acceptability of Intervention Measure
CBT: cognitive behavioral therapy
FIM: Feasibility of Intervention Measure
IAM: Intervention Appropriateness Measure
LGBTQ: lesbian, gay, bisexual, transgender, and queer/questioning
MI: motivational interviewing
OARS: open-ended questions, affirmations, reflections, and summaries
PM: peer mentor
PREVAIL: Peers for Valued Living
SGD: sexual and gender diverse
SGDY: sexual and gender diverse youth
SPI: safety planning intervention
STARS: Supporting Transitions to Adulthood and Reducing Suicide

XSL•FO
RenderX

Edited by A Mavragani; submitted 23.10.24; peer-reviewed by C Oko, AD Eaton, TF Ojo; comments to author 28.11.24; revised version received 09.12.24; accepted 11.12.24; published 29.01.25.

Please cite as:
Tran JT, Webster J, Wolfe JR, Ben Nathan J, Mayinja L, Kautz M, Oquendo MA, Brown GK, Mandell D, Mowery D, Bauermeister JA, Brown LA
Experiences of Peer Mentoring Sexual and Gender Minority Emerging Adults Who Are at Risk for Suicide: Mixed Methods Study
JMIR Form Res 2025;9:e67814
URL: https://formative.jmir.org/2025/1/e67814
doi:10.2196/67814
PMID:

©Jennifer T Tran, Jessica Webster, James R Wolfe, Jennifer Ben Nathan, Lindiwe Mayinja, Marin Kautz, Maria A Oquendo, Gregory K Brown, David Mandell, Danielle Mowery, José A Bauermeister, Lily A Brown. Originally published in JMIR Formative Research (https://formative.jmir.org), 29.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Self-Guided Smartphone App (Vimbo) for the Reduction of Symptoms of Depression and Anxiety in South African Adults: Pilot Quantitative Single-Arm Study

Sherrie Steyn[1*], BA, MSc; Meggan Slabbert[2*], MSocSci

[1]Vimbo Health SA (Pty) Ltd, Sandton, South Africa

[2]Private Practice, Ballito, South Africa

[*]all authors contributed equally

Corresponding Author:
Sherrie Steyn, BA, MSc
Vimbo Health SA (Pty) Ltd
20 Riverclub Mews, 7 Sycamore Avenue, Riverclub
Sandton, 2191
South Africa
Phone: 27 63 861 6780
Email: sherrie@vimbohealth.com

Abstract

Background: Barriers to mental health assessment and intervention have been well documented within South Africa, in both urban and rural settings. Internationally, evidence has emerged for the effectiveness of technology and, specifically, app-based mental health tools and interventions to help overcome some of these barriers. However, research on digital interventions specific to the South African context and mental health is limited.

Objective: This pilot study investigated the feasibility of using an app (Vimbo) to treat symptoms of anxiety and depression in South African adults recruited from a community sample. The Vimbo app is a self-guided, cognitive behavioral therapy–based digital intervention for common mental health difficulties developed for the South African context.

Methods: This pilot study used a naturalistic, single-arm design testing the Vimbo app over 12 weeks, from October 2020 to February 2021. Participants were recruited through the South African Depression and Anxiety Group and social media advertisements online. A 2-week retention period was used to allow for a minimum of 2 datasets. App usage and engagement metrics were extracted directly from the back end of the app. Based on the model, researchers expected many users to discontinue usage when their symptom levels entered a healthy range. Pre-post review of symptom levels was used to reflect on clinical recovery status at discontinuation after the retention period.

Results: A total of 218 applicants met study eligibility criteria and were invited to download the Vimbo app. Of these, 52% (114/218) of the participants registered with the app, who indicated multiple variances of depression and anxiety symptoms ranging in severity from mild to severe. Two participants users withdrew from the study. Moreover, 69% (77/112) of users were retained, including 8 who had technical issues with their treatment. When comparing broad uptake across all interested participants, chi-square analysis indicated significantly reduced uptake in participants identifying as “unemployed but seeking employment” (χ^2_4=10.47; N=251; *P*=.03). When considering app usage for the entire cohort (n=69, excluding participants with technical issues), there was a mean of 72.87 (SD 71.425) total module pages read, a mean of 30% (SD 29.473%) of prescribed content completed, and a mean of 19.93 (SD 27.517) times engaging with tools and skills.

Conclusions: Our findings support the case for continued exploration of app-based interventions for treating depression and anxiety in South Africa. Developing strategies to increase access and improve intervention uptake may prove essential to helping mobile health interventions make as significant an impact as possible. Future research should include a randomized controlled trial with a larger sample to further assess the efficacy of app-based interventions in treating mental health difficulties in South Africa.

(JMIR Form Res 2025;9:e54216) doi:10.2196/54216

XSL•FO
RenderX

KEYWORDS

treatment gap; mental health; health; depression; anxiety; South Africa; CBT; cognitive behavioral therapy; app-based intervention; mobile health; mental health app; smartphone; mobile phone

Introduction

South Africa currently has some of the poorest mental health outcomes globally, with 36% of the population feeling distressed or struggling [1]. Recent studies suggest that nationally, 25.7% (874/3402) of respondents reported scores ≥10 on the Patient Health Questionnaire-9 (PHQ-9) and 17.8% (606/3402) reported scores of >10 on the Generalized Anxiety Disorder-7 (GAD-7), indicating probable depression or anxiety [2]. Disorder prevalence varies widely by geographical province, which may reflect the variance in mental health resource availability and allocation [2,3]. These estimates are a marked increase from previous findings from the World Health Organization, which estimated that 4.6% of the population lived with a depressive disorder and 3.4% lived with an anxiety disorder [4]. This increase is part of a more significant global increase in cases of anxiety and major depressive disorders following the COVID-19 pandemic, including a 21.5% increase in the prevalence of these disorders in sub-Saharan Africa [5].

There is an ever-widening divide between increasing numbers of people experiencing mental health difficulties and access to adequate support or therapeutic intervention, known as the "treatment gap" [6,7]. Findings indicate that mental disorders often go untreated despite their high prevalence and associated impairment [6]. The treatment gap for depression and many anxiety disorders was more than 50%, increasing to as much as 80% in low- and middle-income countries for severe cases [6,8]. Government spending on mental health is below what is needed to adequately address this gap, with those in middle-income countries being the most affected [8].

South Africa is no exception. Despite having dedicated legislation, such as the Mental Health Care Act of 2002, there remains a significant treatment gap, especially in rural areas [9-11]. Recent studies suggest that for people in South Africa with mental health problems, only 1 in 10 receives the care they need [3]. This figure amounts to a staggering 92% treatment gap before the added burden of COVID-19. South Africa faces many barriers to bridging the treatment gap, including high levels of comorbidity and the relationship between common mental disorders (CMDs) and communicable diseases (such as HIV and tuberculosis), the apartheid legacy of institutional disparity, lack of adequate development of services, and a severe shortage of mental health professionals [9,11]. Research into cost-effective interventions is essential to inform policy decisions [9-11].

The use of technology and mental health apps in bridging barriers to this treatment gap is growing [7,9,12]. In 2018, approximately 99.5% of the population in South Africa had access to the 3G mobile network, and 85.7% had 4G/LTE coverage [13]. As of January 2020, South Africa had an estimated 3493 million active mobile web users [14].

The potential for app-based interventions to significantly reduce the cost of treatment, increase access in difficult-to-reach populations (such as rural communities), and reduce the need for clinician input and time is especially relevant to the South African context where these barriers have a more significant impact [6,9-11]. Several studies within South Africa have shown that mobile phones are already used as informal learning tools by patients and health care workers [15-20].

Despite this growing evidence and the number of mobile health (mHealth) apps, research is needed into apps designed specifically for the South African context and market. More generally, further research is needed to consider the cost-effectiveness and therapeutic efficacy of interventions for CMDs, especially mobile-based technology in South Africa [10,11].

Vimbo Health is a mental health assessment and treatment app, informed and developed based on a cognitive behavioral therapy (CBT) approach. The app streams users into courses for depression, anxiety and comorbid presentations using standardized psychometric measures, namely, the PHQ-9 and GAD-7. Subclinical users receive a well-being course focused on building awareness and resilience. CBT has proven effective in reducing symptom severity and functional impairment in both depressive and anxiety disorders [21-25].

This study aimed to investigate the feasibility of using the Vimbo app to reduce symptoms of anxiety and depression within an adult South African cohort. Using a single-arm, naturalistic design, we sought to pilot the app within a community sample as a self-guided intervention for those interested in supporting their mental health. We hypothesized that if offered a self-guided app (Vimbo), South African adults would voluntarily engage with the app to improve their mental health. We further hypothesized that Vimbo would reduce common symptoms of anxiety and depression symptom severity within 12 weeks of initial usage. Testing Vimbo's safety and efficacy provides valuable knowledge about using app-based mental health interventions in South Africa.

Methods

Study Design and Procedure

This study used a naturalistic, single-arm design to assess the acceptability of a self-guided, digital intervention to reduce symptoms of depression and anxiety in South African adults, using multiple repeat measures administered in the app over 12 weeks.

Interested individuals provided consent (Multimedia Appendix 1) and completed a demographic questionnaire and PHQ-9 online via a web form. Eligible participants were invited to participate via email, offered technical support, and given 2 weeks to download and register with the app (February 1-15, 2021). App registration included completing the PHQ-9, GAD-7, and Work and Social Adjustment Scale (WSAS). Participants self-managed app engagement with psychometric measures administered biweekly in the app. Data were collected directly

XSL•FO
RenderX

from the app over 12 weeks (ended May 10, 2021). At the study's end, participants received a user experience questionnaire to collect qualitative feedback and were offered an opportunity to debrief with a researcher. Participants had 2 weeks to complete and submit the user experience questionnaire on the web. This study does not include these qualitative data, which will be reported separately.

Ethical Considerations

This study received ethical clearance from the Human Sciences Research Council (HSRC) of South Africa on August 28, 2020. The HSRC Research Ethics Committee (REC) is registered with the South African National Health Research Ethics Council (REC-290808-015). The HSRC REC has the US Office for Human Research Protections Federalwide Assurance (organization 0000 6347). All participants signed a research study information and consent form via a web-based portal using a standardized form provided by the HSRC. Received consent forms were reviewed by researchers to ensure proper completion. All interested individuals were encouraged to contact the researchers should they have any question or wish to discuss their participation. Participation in the research was completely voluntary, with participants able to withdraw from the study at any point during the study period. The study was available to anyone comfortable communicating in English.

The Vimbo app is designed specifically for low data usage, both during download and regular use. The app has offline capability; as such, course content is downloaded once at the point of registration. Guided audio meditations are downloaded once when first played, so further playing does not consume any additional data. The app syncs to our back end when in daily use; however, expected use should not exceed 0.5 MB of data per day. The app is expected to use 126 MB over the 3-month study period. As such, participants were offered a once off reimbursement of R29.00 (approximately USD 2) to cover data usage based on local data costs.

Personally identifiable information stored in the mobile app back end was held on an encrypted Google cloud database, access to which was password protected and tightly controlled by the chief executive officer. Access to the back end is limited to only the chief executive officer and the software developer and restricted to critical use cases.

Data extracted from the back end were anonymized by removing or obfuscating personally identifiable information (name, email address, and phone number). Extracted data were kept in a password-protected and encrypted web-based folder hosted by Google with access granted only to the researchers. All electronic files generated by the research were maintained in the same electronic folders. Google Drive offers encryption and restricted access rights.

Informed consents and prescreening data were captured and stored electronically using the platform Jotform, which is password protected, is encrypted, and meets high security standards. Participants were anonymized through the use of coded identifiers, with any personally identifiable information including contact details stored in a separate, password-protected folder. Data were accessible only to the researchers and, due to necessity, the chief executive officer of Vimbo Health Ltd.

Participants

Participants were recruited through the South African Depression and Anxiety Group and paid social media advertisements on Facebook from October 2020 to February 2021. Advertisements informed potential volunteers that the study assessed the effectiveness of an app for anxiety and depression over 12 weeks. Participation inclusion criteria included the following:

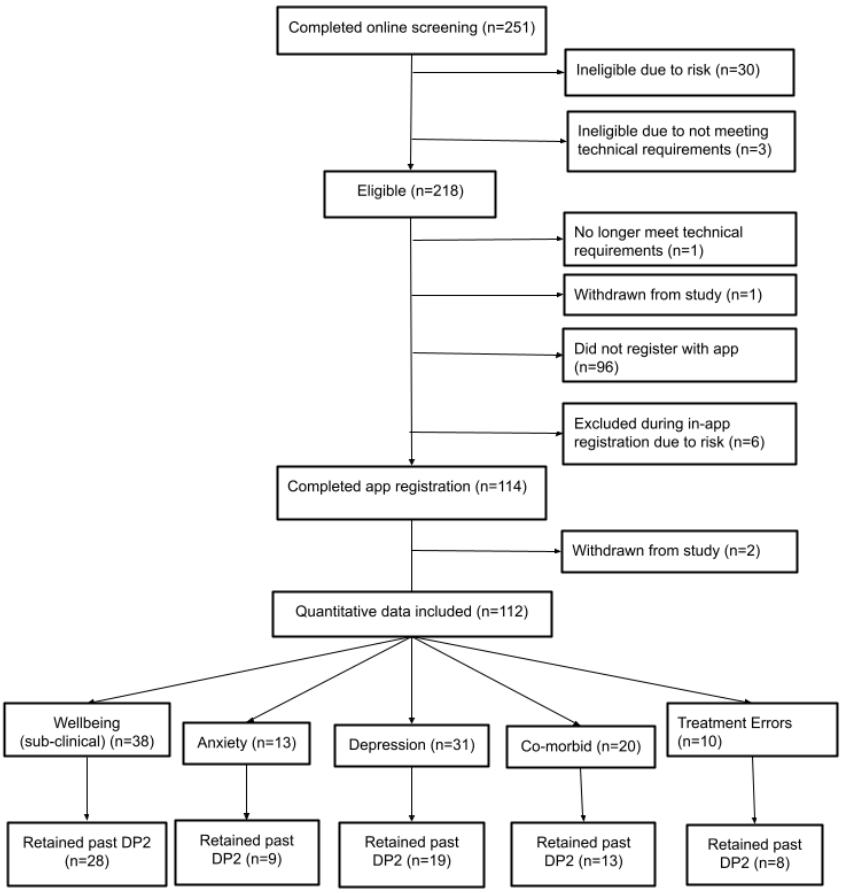
1. 18 years of age or older
2. South African residency
3. Access to an Android phone for the duration of the study
4. Comfortable using the app in English

Participants considered at risk of suicidality or self-harm as measured by a response of 2 and greater on question 9 of the PHQ-9 were excluded from participation in the intervention. Follow-up risk assessments and signposting were offered to these individuals.

A total of 4 participants were excluded based on technical criteria and 3 withdrew. An additional 36 were excluded due to risk concerns (PHQ-9 question 9 score of ≥2). Details of the flow of participants through recruitment and study are shown in Figure 1.

XSL•FO
RenderX

Figure 1. Flowchart of participants. DP2 is at 2 weeks after registration. DP2: data point 2.



Instrument and Procedure Overview

Interested persons were directed via a web-based link to a Jotform page for further study information and consent. Following digital receipt of study consent, individuals were asked to complete a demographic questionnaire and the PHQ-9 to help assess suitability for study participation. Following review of the consent form, demographic questionnaire, and PHQ-9 for inclusion and exclusion criteria, individuals were advised of their suitability for the study via their preferred communication mode.

Those participants who met the inclusion criteria and none of the exclusion criteria were invited to participate in the intervention and given 2 weeks to download and register with the app (February 1-15, 2021). Technical support was offered to participants to aid this process.

The registration process included the following:

1. Offering some brief information on the frequency of mental health problems and about the app, namely, that it is CBT based and aimed at assessing and treating mental health.
2. Entering their mobile number to get started. This number is used to text message a security code to authenticate the device.
3. Registering their details, including preferred name, email address, age, gender, and employment status.
4. Reading and agreeing to the terms and conditions.
5. Completing the PHQ-9 and receiving feedback on their score.
6. Completing the GAD-7 and receiving feedback on their score.
7. Completing the WSAS and receiving feedback on their score.
8. Completing a brief questionnaire on well-being.
9. Receiving feedback on the overall presentation.

After completing the registration process, participants self-managed their app engagement except for the psychometric measures, administered through pop-ups in the app every 2 weeks. Data were collected directly from the app over 12 weeks (ended May 10, 2021).

Interventions

Following in-app registration, participants were placed into 1 of the following 6 treatment courses, based on their symptom profile:

1. Depression courses (PHQ-9 >10 or WSAS >20)
2. Anxiety course (GAD-7 >10)
3. Comorbid depression and anxiety course, addressing depression first (PHQ-9 and GAD-7 both >10)
4. Comorbid depression and anxiety course, addressing anxiety first (PHQ-9 and GAD-7 both >10)
5. Subclinical course focused on general well-being (for users who present with none to minimal symptom severity)
6. Low activity (subclinical scores, WSAS >20)

Users in the low activity group receive the same treatment course as those in the depression group.

XSL•FO
RenderX

Our courses are structured around CBT-based modules, which focus on skills development and encourage real-world practice. To support participants, we provide in-app tools and skills pages that facilitate the completion of suggested practice exercises, reflection, and skill integration. Measures are readministered fortnightly in the app. These allow the user to reflect on the impact their new skills are having on their symptom profiles. While no additional support is provided outside technical help, 21 random pop-up notifications are strategically used to encourage engagement. Details of course content and tools are included in Multimedia Appendix 2.

Outcome Measures

App usage and engagement were measured by tracking interactions with the various aspects of the app, including content, skills, and tools. We also considered the percentage of overall content completed with the specific treatment course. Unfortunately, we could not reliably use the time stamp information to calculate the overall time spent on the app and individual session data due to possible inconsistencies in how these back-end data were collected. Participants were considered retained if they entered at least 2 complete sets of measures (PHQ-9, GAD-7, and WSAS), including those completed during registration.

In-app outcome measures allow the users to track their progress and symptom changes as a result of their efforts. These include the PHQ-9, GAD-7, and the WSAS [26-30].

1. PHQ-9: 9 items, 4-point Likert scale, measures depression severity (cutoffs: 5, 10, 15, and 20).
2. GAD-7: 7 items, 4-point Likert scale, measures anxiety severity (cutoffs: 5, 10, and 15).
3. WSAS: 5 items, 8-point Likert scale, measures functional impairment (cutoffs: 10 and 20).

Both the PHQ-9 and the GAD-7, widely used in research and clinical practice in health care and community settings, have shown reliable sensitivity within the general population [31,32]. These measures have also been tested for reliability in South Africa [33-36]. The WSAS, on the other hand, has proven to be a reliable and valid measure of impaired functioning, offering enough sensitivity to be an effective outcome measure [36,37]. All 3 measures were self-administered within the app, at registration, and every 2 weeks after that for the duration of the study, ensuring the credibility of our results. Reviewing pre-post changes on these measures allowed us to look at how clinical recovery may impact discontinuation with the app, with researchers theorizing that clinical recovery may coincide with discontinuation of the app in line with the underpinning theoretical model. Participants met clinical recovery if their final scores were below 10 on the GAD-7 and PHQ-9 after having initially scored above 10 on either measure.

Statistical Analysis

All data analysis was conducted using IBM SPSS Statistics software (version 28; IBM Corp). Unless otherwise stated, the significance level for all tests was set to α=.05.

Demographic differences based on intervention uptake, retention, and recovery status were analyzed using chi-square and independent 2-tailed *t* tests. Frequency tables captured app usage and engagement. Unfortunately, due to technical issues, we could not accurately collate time-relative session data and, as such, could not reliably calculate the total number of usage sessions or the total amount of time spent on the app. Due to the small number of participants in the 2 comorbid streams, these were collapsed into a single group for analysis. Pre-post comparisons of symptoms levels on the PHQ-9 and GAD-7 allowed us to establish clinical recovery status. For these calculations, data collected during registration were compared with the last full set of primary measures completed in the app, regardless of the time point at which this was submitted (so entry and exit measures only). Therefore, any participants who completed only some of the primary measures would have their previous complete set of measures used for analysis.

Results

Participant Characteristics

A total of 218 people were invited to download and register the app. Moreover, 47% (104/218) of those invited did not register or no longer met criteria to participate in treatment. A total of 114 participants registered with the app; however, 2 withdrew from the study, leaving a total sample size of 112 participants.

The mean age of participants (n=112) was 35.21 (SD 9.6) years. The average PHQ-9 score was 9.58 (SD 6.02). In terms of ethnicity, 35% (39/112) identified as Black African, 34% (38/112) as White, 10% (11/112) as Coloured, 5% (6/112) as Indian, and 16% (18/112) as belonging to another ethnic group. The gender distribution was predominantly female, with 83% (90/112) identifying as female and 17% (21/112) as male.

Regarding primary language, 11% (12/112) of the participants spoke Afrikaans, 64% (72/112) spoke English, 11% (12/112) spoke isiZulu, and 14% (16/112) spoke another language. When asked about their current health condition, 43% (48/112) reported having a health condition, 53% (59/112) reported no health condition, and 5% (5/112) preferred not to say. Furthermore, 65% (31/48) of those who reported having a health condition reported a mental health issue. Employment status among the participants showed that 61% (69/112) were in some form of employment, 19% (21/112) were unemployed but seeking work, and 20% (22/112) fell into other categories.

When comparing broad uptake across all interested participants, chi-square analysis indicated significantly reduced uptake in participants identifying as “Unemployed but seeking employment” (χ^2_4=10.47; N=251; *P*=.03). Participants identifying with this employment category had the highest representation of those who did not register with the app (49/139, 35.3%). Also, those who categorized themselves as having a mental health diagnosis were significantly more likely to register (31/112, 28%) than those who did not identify as having a mental health diagnosis (81/112, 72%; χ^2_1=8.56; N=251; *P*=.003). No other significant demographic differences were found based on intervention uptake (all *P*>.05).

Moreover, 69% (77/112) of users were retained using a 2-week metric, with 31% (35/112) dropping out before completing a

XSL•FO
RenderX

second set of measures in the app. Also, 45% (50/112) of users submitted a third set of measures, 29% (33/112) a fourth set, 22% (25/112) a fifth set, 12% (14/112) a sixth, and with only 5% (6/112) submitting a sixth set of measures at 12 weeks.

Table 1 looks at retention rates based on treatment group and symptom severity level. Moreover, 6% (2/35) of those participants who dropped out were missing treatment course data due to technical problems. Analysis of demographic information found no statistically significant differences associated with study retention (all P>.05). These percentages correspond to the overall proportionality in each treatment group.

Ten participants were missing treatment course information or had been placed in the incorrect treatment course due to technical error. Two (20%) of these 10 participants dropped out and 8 (80%) were retained. Of those who were retained, 7 (88%) of the 8 met the clinical symptom severity level criteria—neither of the participants who dropped out met the criteria for clinical symptom severity level (0/2, 0%).

Table 1. Comparison of treatment group allocation and symptom severity levels between retained and dropped out study participants.

Variable	Retained (n=77)	Dropped out (n=35)
Participants (n=112), n (%)	77 (69)	35 (31)
Treatment group, n (%)		
Subclinical group	32 (42)	10 (29)
Depression group	21 (27)	12 (34)
Comorbid group	13 (17)	7 (20)
Anxiety group	11 (14)	4 (11)
Unknown due to technical error	N/A[a]	2 (6)
Depressive symptom severity, n (%)		
Minimal	13 (17)	7 (20)
Mild	25 (33)	7 (20)
Moderate	21 (27)	17 (49)
Moderately severe	12 (16)	4 (11)
Severe	6 (8)	N/A
Generalized anxiety disorder symptom severity, n (%)		
Minimal	17 (22)	9 (26)
Mild	26 (34)	13 (37)
Moderate	23 (30)	9 (26)
Severe	11 (14)	4 (11)

[a]N/A: not applicable.

App Usage

When considering overall app usage for the entire cohort (n=69, excluding participants with any technical issues), there was a mean of 72.87 (SD 71.425; range 0-278; mode 0) total module pages read, a mean of 3.36 (SD 2.169; range 0-10; mode 3) total modules accessed, and a mean of 19.93 (SD 27.517; range 0-126; mode 0) times engaging with tools and skills. Users (n=69) completed a mean of 30% (SD 29.473%; range 0%-100%; mode 0%) of prescribed content. Table 2 compares app usage and engagement across treatment groups.

There were no significant differences in app usage between those who met clinical recovery criteria and those who did not. Similarly, no significant correlations were found between changes in primary measures and app usage (all P>.05).

XSL•FO
RenderX

Table 2. App usage across treatment groups (total n=69, excluding users with technical errors).

Group	Module pages read (n)	Tools/skills usage (n)	Prescribed content completed (%)
Depression group (n=19)			
Mean (SD)	69.68 (58.357)	19.05 (22.225)	29.40 (24.623)
Range	210	86	89
Mode	0	8[a]	0[a]
Anxiety group (n=9)			
Mean (SD)	77.67 (84.109)	31.22 (37.181)	28.77 (31.151)
Range	263	119	97
Mode	0[a]	4	0[a]
Comorbid group (n=13)			
Mean (SD)	66.85 (73.546)	12.54 (28.070)	20.07 (22.086)
Range	260	104	78
Mode	18[a]	3	5[a]
Subclinical group (n=28)			
Mean (SD)	76.29 (77.676)	20.32 (27.356)	34.21 (34.832)
Range	223	126	100
Mode	0	0[a]	0

[a]Multiple modes exist. The smallest value is shown.

Symptom Status at Discontinuation

The intervention follows a least intervention first approach, meaning that users are expected to engage with the app as needed until their symptoms return to a healthy level. Since CBT is time limited, the app focuses on real-world integration of the skills and techniques taught, with users expected to reduce engagement or discontinue use as their symptoms improve. Considering this, we have included a review of the recovery status of users based on their last complete set of measures submitted in the app, as an indicator of symptom status at point of discontinuation (last set of complete measures submitted in the app).

Of the 69 users who indicated clinically severe symptom levels at registration, 25% (17/69) met clinical recovery, 35% (24/69) did not meet clinical recovery, and 41% (28/69) did not have initial scores high enough to be considered before discontinuing use (last set of measures). Of those with clinical symptom levels at registration, 42% (17/41) met clinical recovery and 59% (24/41) did not meet clinical recovery based on their last submitted measures.

Moreover, 58% (11/19) of the depression course, 44% (4/9) of the GAD course, and 31% (4/13) of the comorbid course met recovery on the GAD-7 in their final scores. Also, 53% (10/19) of the depression course, 44% (4/9) of the GAD course, and 46% (6/13) of the comorbid course met recovery on the PHQ-9 in their final scores.

Based on current evidence, UK National Institute for Clinical Excellence guidelines recommend offering low-intensity CBT, including computerized CBT, for those individuals presenting with mild to moderate symptom levels and not for those individuals with moderate to severe symptom levels [38-41]. Considering this, Table 3 shows retention and clinical recovery status at discontinuation, when removing participants with severe symptom levels at registration.

Table 3. Retention and clinical recovery status at discontinuation or study end (excluding participants with severe symptom levels).

	Total participants, n	Dropped out, n (%)	Retained past DP2[a], n (%)	Clinical recovery met, n (%)
Clinical cohort	51	20 (39)	31 (61)	16 (52)
Depression group	30	12 (40)	18 (60)	9 (50)
Anxiety group	4	1 (25)	3 (75)	3 (100)
Comorbid group	17	7 (41)	10 (59)	4 (40)

[a]DP2: Data point 2.

XSL•FO
RenderX

Discussion

Principal Findings

This pilot study sought to test the feasibility of a self-guided, CBT-focused, mental health app–based intervention for South African adults who are experiencing mild to severe symptoms of depression and anxiety. Using an app specifically designed for the South African environment, we recruited 112 South African adults online and through support services, who were interested in using a mental health app. Participant engagement with the app was completely self-driven, with only technical support offered to users. Study retention rates and app usage indicate great potential for mHealth-based treatment; however, further technical development may be necessary to encourage greater engagement with content and tool use. In line with the treatment model, data suggest that most users who discontinued usage beyond the retention period did so after meeting clinical recovery. Although small, this pilot study is one of the first to test the usage case for an app-based intervention designed, developed for, and delivered within the South African context using a CBT model. Overall, these results support the continued exploration of using app-based interventions within the South African context to support the diagnosis and treatment of depression and anxiety.

In their survey of a low-income community in Johannesburg, van Rie and colleagues [42] found that 23% (283/1230) of the participants reported mild to severe depression. In our study, 27% (21/77) of the participants reported clinically significant depressive symptoms and 14% (11/77) reported clinically significant anxiety levels. Moreover, 17% (13/77) of the participants reported clinically significant depression and anxiety symptoms. Comparing our findings with a 2022 national survey where 25.7% (874/3402) of respondents reported PHQ-9 scores of 10 and greater and 17.8% (606/3402) reported scores of 10 and greater on the GAD-7 allows us to see better the impact the COVID-19 lockdown had on mental health [2].

Several authors have reflected on the disparity between individual enrollment and initiation in self-guided web-based interventions, with a systematic review finding that 33% to 88% of users who downloaded the app used it at least once [43-45]. Our uptake and retention rates are comfortably on the higher end of this scale and offer further evidence supporting the intervention. Other mental health app–based intervention studies report retention rates ranging from 5.5% to 70% after 30 days [46,47]. Our retention rates ranged from 69% at 2 weeks, 29% at 8 weeks, and 5% at 12 weeks. This steady reduction in retention is in line with the app model, which focuses on real-world integration of skills and learning with users expected to use the app for only as long as they need to reduce their symptoms to healthy levels. It further reflects the differing lengths of courses and individual progression through the intervention.

There was significantly reduced uptake in those identified as “unemployed but seeking employment.” This association may be due to issues with communication during the recruitment and app registration process, especially since subsequent analysis found no significant relationship between employment and retention. Several mHealth studies have found reduced retention in those who identified as unemployed [46,48]. The significantly increased uptake in those participants who reported having a mental health diagnosis suggests that already having a mental health diagnosis may act as a motivator for seeking support and engaging in research.

We found that users in a clinical course completed an average of 26% of their prescribed content, equating to approximately 2.1 modules completed per user. In a similar study, the participants completed an average of 2.6 (SD 2.7) modules. However, this included some access to coaching [49]. Aizenstros et al [46], in their study of a smartphone-delivered behavioral activation intervention, found that users completed an average of 5.46 (range 1-65; n=238) app-based activities over 30 days. Their findings were comparable with those of Watts et al [50], whose study participants’ mean aggregated score for homework completion within the mobile-based intervention was 11 (range 6-17; n=35).

Although usage is comparable, if not higher than other app-based interventions, numerous studies suggest that outcomes may be further improved if a mental health care professional guides the digital intervention [51-53]. Anton et al [49] found increased treatment outcomes and engagement in participants who completed at least 4 modules compared with those who completed fewer, suggesting a minimal dosage effect. However, it is tricky to know since interventions may not be directly comparable in amount and length of content [49]. However, in a similar small-scale pilot, Moghimi et al [54] found significant improvements in the GAD-7 and PHQ-9 for participants of their 9-week mHealth intervention. Their study had better overall retention, suggesting room for improving intervention efficacy through increased treatment adherence. Considering the scalability and impact on accessibility that mHealth interventions offer, even small gains can lead to considerable public health gains [23]. As other authors have noted, this is especially relevant to vulnerable populations, including those who may be geographically isolated and those with limited access to suitable resources [19,55].

As mentioned, there is a significant treatment gap related to the assessment and treatment of mental health disorders, specifically those individuals with a comorbid diagnosis, that is, evidenced by 2 or more mental health difficulties or disorders experienced simultaneously. The applicability of mHealth for people accessing mental health services and the possibility of reducing hospital admissions and improving symptom severity, engagement, and treatment are undeniable for clinical practice. Furthermore, this applicability extends to using mHealth as a stand-alone process or a process used in conjunction with other mental health care interventions, such as individual and group therapy or support groups and psychiatric interventions. The potential for app-based interventions to significantly reduce the cost of treatment, increase access in difficult-to-reach populations (such as low-income or rural communities), and reduce the need for clinician input and time is especially relevant to the South African context where these barriers have a more significant impact [6,9-11,19,56].

XSL•FO
RenderX

The app allows users to track their symptom progression easily, is designed for low data usage, and includes offline functionality. Furthermore, the app can be accessed using a voucher code and text message verification, allowing almost immediate access without professional guidance. This feature increases immediate access to treatment and reduces barriers such as wait times, travel and transport difficulties, and stigma. This is especially noteworthy in light of the findings by Forman-Hoffman et al [57], whose investigation found no significant differences between participants in underserved rural or metropolitan areas regarding usage, satisfaction, and treatment outcomes.

The Vimbo app has multiple usage cases, including as a stand-alone intervention or as a therapeutic tool and resource. This adaptability allows it to be integrated at various care levels within a stepped-care system to effectively create easily personalized intervention options into current health systems [12,23]. Using this stepped-care implementation directly into current health systems encourages a reduced caseload for care providers and better use and distribution of resources [7,10]. Furthermore, research suggests that using a blended CBT model that integrates a digital program may improve working alliance and outcomes, suggesting that treatment as usual may benefit by including access to the Vimbo app [58].

Strengths and Limitations

This pilot study possesses several strengths. One of the primary strengths is its contextual relevance, as it is among the first to explore app-based digital interventions within the South African context. This offers valuable insights into the scalability of such interventions to larger settings and services. The study's naturalistic setting, community orientation, and diverse participant age range offer realistic and generalizable insights into everyday usage and engagement. The research points to future digital intervention developments for rural and urban South Africa and has provided Vimbo Health with data for app improvements.

However, it is important to note the study's limitations, which could impact the generalizability of the findings. While qualitative data on user experiences were collected, they were not included in this report due to the significant amount of quantitative data, with plans to publish these findings separately. The small sample size limits the statistical power of the study, making it challenging to extend the findings broadly. The context of the pandemic may have influenced user registration, engagement, and outcomes, indicating a need for further research in nonpandemic conditions. The reliance on self-reported data could introduce biases and affect the accuracy of the results. The app's availability only in English excluded non-English speakers. Future iterations should consider localizing the app to include all South African languages.

Conclusions

This pilot study takes the first critical steps to establish the feasibility of using self-guided, app-based treatment for anxiety and depression to help address the substantial mental health treatment gap in South Africa. The 69% (77/112) retention of users for a minimum of 2 weeks shows that mHealth interventions are a viable treatment modality, acceptable to the South African community. However, only 52% (114/218) of individuals invited to participate in the research registered with the app, suggesting the potential for greater impact if the access and onboarding process can be refined and improved to increase uptake. Overall, our results show promise for the use of digital technologies in addressing symptoms of anxiety and depression in South African adults struggling with their mental health. However, due to the pilot nature of this study and the limited number of other studies in this field within South Africa that have examined the use of digital interventions, further studies are needed to better understand treatment parameters including efficacy and dosage response. Large-scale, longitudinal studies are necessary to better assess the impact digital interventions have on mental health symptoms. Findings from this initial study offer support for app-based psychological interventions to bridge the ever-widening treatment gap.

Data Availability

All data generated or analyzed during this study are included in Multimedia Appendices 1-2.

Conflicts of Interest

Both authors are affiliated with Vimbo Health Ltd through personal investment in the company. The lead researcher is both cofounder and chief science officer for Vimbo Health Ltd. The coresearcher remains an advisor to Vimbo Health Ltd. Both have received a salary or payment for their work on this research. Research funded by Vimbo Health Ltd. A potential conflict of interest exists as Vimbo Health Ltd sponsored the research, and the lead researcher is a cofounder and chief science officer. Despite this, strict protocols ensured unbiased data processing and analysis. Raw data are available for further analysis and cross-checking.

Multimedia Appendix 1
Participant information sheet and consent form.
[PDF File (Adobe PDF File), 111 KB - formative_v9i1e54216_app1.pdf]

Multimedia Appendix 2
Vimbo version 1 app outline.
[PDF File (Adobe PDF File), 708 KB - formative_v9i1e54216_app2.pdf]

XSL•FO
RenderX

References

1. Newson JJ, Pastukh V, Thiagarajan TC. Assessment of population well-being with the mental health quotient: validation study. JMIR Ment Health 2022;9(4):e34105 [FREE Full text] [doi: 10.2196/34105] [Medline: 35442210]
2. Craig A, Rochat T, Naicker SN, Mapanga W, Mtintsilana A, Dlamini SN, et al. The prevalence of probable depression and probable anxiety, and associations with adverse childhood experiences and socio-demographics: a national survey in South Africa. Front Public Health 2022;10:986531 [FREE Full text] [doi: 10.3389/fpubh.2022.986531] [Medline: 36388391]
3. Docrat S, Besada D, Cleary S, Daviaud E, Lund C. Mental health system costs, resources and constraints in South Africa: a national survey. Health Policy Plan 2019;34(9):706-719 [FREE Full text] [doi: 10.1093/heapol/czz085] [Medline: 31544948]
4. World Health Organization. Depression and Other Common Mental Disorders: Global Health Estimates. Geneva, Switzerland: World Health Organization; 2017.
5. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
6. Demyttenaere K, Bruffaerts R, Posada-Villa J, Gasquet I, Kovess V, Lepine JP, et al. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. JAMA 2004;291(21):2581-2590. [doi: 10.1001/jama.291.21.2581] [Medline: 15173149]
7. Prince M, Patel V, Saxena S, Maj M, Maselko J, Phillips MR, et al. No health without mental health. Lancet 2007;370(9590):859-877. [doi: 10.1016/S0140-6736(07)61238-0] [Medline: 17804063]
8. Saxena S, Thornicroft G, Knapp M, Whiteford H. Resources for mental health: scarcity, inequity, and inefficiency. Lancet 2007 Sep 08;370(9590):878-889. [doi: 10.1016/S0140-6736(07)61239-2] [Medline: 17804062]
9. Burns JK. The mental health gap in South Africa—a human rights issue. Equal Rights Trust. 2011. URL: https://www.equalrightstrust.org/ertdocumentbank/ERR06_special_Jonathan.pdf [accessed 2023-10-30]
10. Jack H, Wagner RG, Petersen I, Thom R, Newton CR, Stein A, et al. Closing the mental health treatment gap in South Africa: a review of costs and cost-effectiveness. Glob Health Action 2014;7:23431 [FREE Full text] [doi: 10.3402/gha.v7.23431] [Medline: 24848654]
11. Petersen I, Lund C. Mental health service delivery in South Africa from 2000 to 2010: one step forward, one step back. S Afr Med J 2011;101(10):751-757. [Medline: 22272856]
12. Wilhelm S, Weingarden H, Ladis I, Braddick V, Shin J, Jacobson NC. Cognitive-behavioral therapy in the digital age: presidential address. Behav Ther 2020;51(1):1-14 [FREE Full text] [doi: 10.1016/j.beth.2019.08.001] [Medline: 32005328]
13. O'Dea S. South Africa mobile coverage / smartphone penetration 2015-2018. Statista. 2023. URL: https://www.statista.com/statistics/1074377/south-africa-mobile-coverage-and-smartphone-penetration/ [accessed 2020-04-27]
14. Clement F. South Africa: digital population as of January. Statista. 2020. URL: https://www.statista.com/statistics/685134/south-africa-digital-population/ [accessed 2020-04-29]
15. Velghe F. Literacy acquisition, informal learning and mobile phones in a South African township. New York, NY: Association for Computing Machinery; 2013 Dec 07 Presented at: Sixth International Conference on Information and Communication Technologies and Development; December 7-10, 2013; Cape Town, South Africa p. 89-99. [doi: doi:10.1145/2516604.2516615]
16. Anstey Watkins JOT, Goudge J, Gómez-Olivé FX, Griffiths F. Mobile phone use among patients and health workers to enhance primary healthcare: a qualitative study in rural South Africa. Soc Sci Med 2018;198:139-147 [FREE Full text] [doi: 10.1016/j.socscimed.2018.01.011] [Medline: 29335160]
17. Tomlinson M, Rotheram-Borus MJ, Doherty T, Swendeman D, Tsai AC, Ijumba P, et al. Value of a mobile information system to improve quality of care by community health workers. S Afr J Inf Manag 2013;15(1):-9 [FREE Full text] [doi: 10.4102/sajim.v15i1.528] [Medline: 25147730]
18. Tsai AC, Tomlinson M, Dewing S, le Roux IM, Harwood JM, Chopra M, et al. Antenatal depression case finding by community health workers in South Africa: feasibility of a mobile phone application. Arch Womens Ment Health 2014;17(5):423-431 [FREE Full text] [doi: 10.1007/s00737-014-0426-7] [Medline: 24682529]
19. van Belle S, van Belle JPW, Weimann E. A South African mHealth app for depression management and treatment: a survey and requirements analysis. 2017 Presented at: African Conference on Information Systems & Technology (ACIST); September 12-13, 2024; Cape Town, South Africa URL: https://www.researchgate.net/publication/344453865_A_South_African_Mhealth_App_for_Depression_Management_and_Treatment_a_Survey_and_Requirements_Analysis_1
20. Barron P, Peter J, LeFevre AE, Sebidi J, Bekker M, Allen R, et al. Mobile health messaging service and helpdesk for South African mothers (MomConnect): history, successes and challenges. BMJ Glob Health 2018;3(Suppl 2):e000559 [FREE Full text] [doi: 10.1136/bmjgh-2017-000559] [Medline: 29713503]
21. Clark DM, Layard R, Smithies R, Richards DA, Suckling R, Wright B. Improving access to psychological therapy: initial evaluation of two UK demonstration sites. Behav Res Ther 2009;47(11):910-920 [FREE Full text] [doi: 10.1016/j.brat.2009.07.010] [Medline: 19647230]
22. Richards DA, Suckling R. Improving access to psychological therapy: the Doncaster demonstration site organisational model. Clin Psychol Forum 2008;181(181):9-16. [doi: 10.53841/bpscpf.2008.1.181.9]

XSL•FO
RenderX

23. Bennett-Levy J, Richards D, Farrand P, Christensen H, Lau MA, Proudfoot J, et al. Oxford Guide to Low Intensity CBT Interventions. New York, NY: Oxford University Press; 2010.
24. Dua T, Barbui C, Clark N, Fleischmann A, Poznyak V, van Ommeren M, et al. Evidence-based guidelines for mental, neurological, and substance use disorders in low- and middle-income countries: summary of WHO recommendations. PLoS Med 2011;8(11):e1001122 [FREE Full text] [doi: 10.1371/journal.pmed.1001122] [Medline: 22110406]
25. Overview: common mental health problems: identification and pathways to care: guidance. National Institute for Health and Care Excellence. 2011. URL: https://www.nice.org.uk/guidance/cg123 [accessed 2023-06-15]
26. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
27. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. Med Care 2004;42(12):1194-1201. [doi: 10.1097/00005650-200412000-00006] [Medline: 15550799]
28. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
29. Kroenke K, Spitzer RL, Williams JBW, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007;146(5):317-325. [doi: 10.7326/0003-4819-146-5-200703060-00004] [Medline: 17339617]
30. Mundt JC, Marks IM, Shear MK, Greist JM. The work and social adjustment scale: a simple measure of impairment in functioning. Br J Psychiatry 2002;180:461-464. [doi: 10.1192/bjp.180.5.461] [Medline: 11983645]
31. Kocalevent RD, Hinz A, Brähler E. Standardization of the depression screener Patient Health Questionnaire (PHQ-9) in the general population. Gen Hosp Psychiatry 2013;35(5):551-555. [doi: 10.1016/j.genhosppsych.2013.04.006] [Medline: 23664569]
32. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the generalized anxiety disorder screener (GAD-7) in the general population. Med Care 2008;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
33. Bhana A, Rathod SD, Selohilwe O, Kathree T, Petersen I. The validity of the Patient Health Questionnaire for screening depression in chronic care patients in primary health care in South Africa. BMC Psychiatry 2015;15:118 [FREE Full text] [doi: 10.1186/s12888-015-0503-0] [Medline: 26001915]
34. Cholera R, Gaynes BN, Pence BW, Bassett J, Qangule N, Macphail C, et al. Validity of the Patient Health Questionnaire-9 to screen for depression in a high-HIV burden primary healthcare clinic in Johannesburg, South Africa. J Affect Disord 2014;167:160-166 [FREE Full text] [doi: 10.1016/j.jad.2014.06.003] [Medline: 24972364]
35. Henn C, Morgan B. Differential item functioning of the CESDR-R and GAD-7 in African and white working adults. SA J Ind Psychol 2019;45(6):-10. [doi: 10.4102/sajip.v45i0.1663]
36. van Heyningen T, Honikman S, Tomlinson M, Field S, Myer L. Comparison of mental health screening tools for detecting antenatal depression and anxiety disorders in South African women. PLoS One 2018;13(4):e0193697 [FREE Full text] [doi: 10.1371/journal.pone.0193697] [Medline: 29668725]
37. Zahra D, Qureshi A, Henley W, Taylor R, Quinn C, Pooler J, et al. The Work and Social Adjustment Scale: reliability, sensitivity and value. Int J Psychiatry Clin Pract 2014;18(2):131-138. [doi: 10.3109/13651501.2014.894072] [Medline: 24527886]
38. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
39. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
40. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FMG, Furukawa TA, Cuijpers P, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 2021;78(4):361-371 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
41. National Institute for Health and Care Excellence. Common mental health problems: identification and pathways to care. 2011. URL: https://www.nice.org.uk/guidance/cg123 [accessed 2024-06-04]
42. van Rie A, West NS, Schwartz SR, Mutanga L, Hanrahan CF, Ncayiyana J, et al. The unmet needs and health priorities of the urban poor: generating the evidence base for urban community health worker programmes in South Africa. S Afr Med J 2018;108(9):734-740. [doi: 10.7196/SAMJ.2018.v108i9.13054] [Medline: 30182898]
43. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
44. Fleming TM, de Beurs D, Khazaal Y, Gaggioli A, Riva G, Botella C, et al. Maximizing the impact of e-therapy and serious gaming: time for a paradigm shift. Front Psychiatry 2016;7:65 [FREE Full text] [doi: 10.3389/fpsyt.2016.00065] [Medline: 27148094]

XSL•FO
RenderX

45. Hanano M, Rith-Najarian L, Boyd M, Chavira D. Measuring adherence within a self-guided online intervention for depression and anxiety: secondary analyses of a randomized controlled trial. JMIR Ment Health 2022;9(3):e30754 [FREE Full text] [doi: 10.2196/30754] [Medline: 35343901]
46. Aizenstros A, Bakker D, Hofmann SG, Curtiss J, Kazantzis N. Engagement with smartphone-delivered behavioural activation interventions: a study of the moodMission smartphone application. Behav Cogn Psychother 2021;49(5):569-581. [doi: 10.1017/S1352465820000922] [Medline: 34396942]
47. Bakker D, Kazantzis N, Rickwood D, Rickard N. A randomized controlled trial of three smartphone apps for enhancing public mental health. Behav Res Ther 2018;109:75-83. [doi: 10.1016/j.brat.2018.08.003] [Medline: 30125790]
48. Kannisto KA, Korhonen J, Adams CE, Koivunen MH, Vahlberg T, Välimäki MA. Factors associated with dropout during recruitment and follow-up periods of a mHealth-based randomized controlled trial for mobile.Net to encourage treatment adherence for people with serious mental health problems. J Med Internet Res 2017;19(2):e46 [FREE Full text] [doi: 10.2196/jmir.6417] [Medline: 28223262]
49. Anton MT, Greenberger HM, Andreopoulos E, Pande RL. Evaluation of a commercial mobile health app for depression and anxiety (AbleTo Digital+): retrospective cohort study. JMIR Form Res 2021;5(9):e27570 [FREE Full text] [doi: 10.2196/27570] [Medline: 34546170]
50. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 2013;13:49 [FREE Full text] [doi: 10.1186/1471-244X-13-49] [Medline: 23391304]
51. Duarte-Díaz A, Perestelo-Pérez L, Gelabert E, Robles N, Pérez-Navarro A, Vidal-Alaball J, et al. Efficacy, safety, and evaluation criteria of mHealth interventions for depression: systematic review. JMIR Ment Health 2023;10:e46877 [FREE Full text] [doi: 10.2196/46877] [Medline: 37756042]
52. Oehler C, Scholze K, Reich H, Sander C, Hegerl U. Intervention use and symptom change with unguided internet-based cognitive behavioral therapy for depression during the COVID-19 pandemic: log data analysis of a convenience sample. JMIR Ment Health 2021;8(7):e28321 [FREE Full text] [doi: 10.2196/28321] [Medline: 34115604]
53. Oser M, Wallace ML, Solano F, Szigethy EM. Guided digital cognitive behavioral program for anxiety in primary care: propensity-matched controlled trial. JMIR Ment Health 2019;6(4):e11981 [FREE Full text] [doi: 10.2196/11981] [Medline: 30946022]
54. Moghimi E, Stephenson C, Agarwal A, Nikjoo N, Malakouti N, Layzell G, et al. Efficacy of an electronic cognitive behavioral therapy program delivered via the online psychotherapy tool for depression and anxiety related to the COVID-19 pandemic: pre-post pilot study. JMIR Ment Health 2023;10:e51102 [FREE Full text] [doi: 10.2196/51102] [Medline: 37993984]
55. Forsythe AM, Venter C. Behavioral economics, motivating psycho-education improvements: a mobile technology initiative in South Africa. Front Psychol 2019;10:1560 [FREE Full text] [doi: 10.3389/fpsyg.2019.01560] [Medline: 31354576]
56. Kim J, Aryee LMD, Bang H, Prajogo S, Choi YK, Hoch JS, et al. Effectiveness of digital mental health tools to reduce depressive and anxiety symptoms in low- and middle-income countries: systematic review and meta-analysis. JMIR Ment Health 2023;10:e43066 [FREE Full text] [doi: 10.2196/43066] [Medline: 36939820]
57. Forman-Hoffman VL, Pirner MC, Flom M, Kirvin-Quamme A, Durden E, Kissinger JA, et al. Engagement, satisfaction, and mental health outcomes across different residential subgroup users of a digital mental health relational agent: exploratory single-arm study. JMIR Form Res 2023;7:e46473 [FREE Full text] [doi: 10.2196/46473] [Medline: 37756047]
58. Doukani A, Quartagno M, Sera F, Free C, Kakuma R, Riper H, et al. Comparison of the working alliance in blended cognitive behavioral therapy and treatment as usual for depression in Europe: secondary data analysis of the E-COMPARED randomized controlled trial. J Med Internet Res 2024;26:e47515 [FREE Full text] [doi: 10.2196/47515] [Medline: 38819882]

Abbreviations

CBT: cognitive behavioral therapy
CMD: common mental disorder
GAD-7: Generalized Anxiety Disorder Assessment-7
HSRC: Human Sciences Research Council
mHealth: mobile health
PHQ-9: Patient Health Questionnaire-9
REC: Research Ethics Committee
WSAS: Work and Social Adjustment Scale

XSL•FO
RenderX

Edited by A Mavragani; submitted 01.11.23; peer-reviewed by M Fernández-González; comments to author 18.02.24; revised version received 04.06.24; accepted 10.10.24; published 30.01.25.

Please cite as:
Steyn S, Slabbert M
Self-Guided Smartphone App (Vimbo) for the Reduction of Symptoms of Depression and Anxiety in South African Adults: Pilot Quantitative Single-Arm Study
JMIR Form Res 2025;9:e54216
URL: https://formative.jmir.org/2025/1/e54216
doi:10.2196/54216
PMID:

©Sherrie Steyn, Meggan Slabbert. Originally published in JMIR Formative Research (https://formative.jmir.org), 30.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Co-Designing a Digital Brief Intervention to Reduce the Risk of Prescription Opioid–Related Harm Among People With Chronic Noncancer Pain: Qualitative Analysis of Patient Lived Experiences

Rachel A Elphinston[1,2,3,4], PhD; Sue Pager[1,5], SpPath, MPH; Kelly Brown[1,4,6], MClinPsych; Michele Sterling[1,2], PhD; Farhad Fatehi[7,8], MD, PhD; Paul Gray[9,10], MBBS, PhD; Linda Hipper[4], Grad Dip, MPH; Lauren Cahill[1], BSc, BNurs; Jason P. Connor[10,11], PhD

[1]RECOVER Injury Research Centre, The University of Queensland, Herston, Australia

[2]National Health and Medical Research Council Centre for Research Excellence - Better Health Outcomes for Compensable Injury, The University of Queensland, Brisbane, Australia

[3]School of Psychology, The University of Queensland, Brisbane, Australia

[4]Metro South Addiction and Mental Health Services, Metro South Health Hospital and Health Service, Brisbane, Australia

[5]Health Equity & Access Unit, Metro South Hospital and Health Service, Brisbane, Australia

[6]Gallipoli Medical Research Foundation, Greenslopes Private Hospital, Brisbane, Australia

[7]Centre for Online Health, The University of Queensland, Brisbane, Australia

[8]School of Psychological Sciences & Turner Institute for Brain and Mental Health, Monash University, Melbourne, Australia

[9]Tess Cramond Pain and Research Centre, Metro North Hospital and Health Service, Brisbane, Australia

[10]Faculty of Medicine, The University of Queensland, Brisbane, Australia

[11]National Centre for Youth Substance Use Research, The University of Queensland, Brisbane, Australia

Corresponding Author:
Rachel A Elphinston, PhD
RECOVER Injury Research Centre
The University of Queensland
Level 7
296 Herston Road
Herston, 4006
Australia
Phone: 61 412662084
Email: rachelel@uq.edu.au

Abstract

Background: Opioid medications are important for pain management, but many patients progress to unsafe medication use. With few personalized and accessible behavioral treatment options to reduce potential opioid-related harm, new and innovative patient-centered approaches are urgently needed to fill this gap.

Objective: This study involved the first phase of co-designing a digital brief intervention to reduce the risk of opioid-related harm by investigating the lived experience of chronic noncancer pain (CNCP) in treatment-seeking patients, with a particular focus on opioid therapy experiences.

Methods: Eligible patients were those aged between 18 and 70 years with CNCP at a clinically significant level of intensity (a score of ≥4 of 10). Purposive sampling was used to engage patients on public hospital waitlists via mail or through the treating medical specialist. Participants (N=18; n=10 women; mean age 49.5 years, SD 11.50) completed semistructured telephone interviews. Interviews were transcribed verbatim, thematically analyzed using grounded theory, and member checked by patients.

Results: Eight overarching themes were found, listed in the order of their prominence from most to least prominent: limited treatment collaboration and partnership; limited biopsychosocial understanding of pain; continued opioid use when benefits do not outweigh harms; a trial-and-error approach to opioid use; cycles of hopefulness and hopelessness; diagnostic uncertainty; significant negative impacts tied to loss; and complexity of pain and opioid use journeys.

XSL•FO
RenderX

Conclusions: The findings of this study advance progress in co-designing digital brief interventions by actively engaging patient partners in their lived experiences of chronic pain and use of prescription opioid medications. The key recommendations proposed should guide the development of personalized solutions to address the complex care needs of patients with CNCP.

(JMIR Form Res 2025;9:e57208) doi:10.2196/57208

KEYWORDS

chronic noncancer pain; prescription opioid use; brief intervention; co-design; patient partners; lived experience; qualitative; digital health

Introduction

Opioid therapy is recommended as part of a multimodal approach to manage acute, postoperative, and cancer pain [1]. There is considerably less evidence on the effectiveness of opioid therapy for patients with chronic noncancer pain (CNCP) [2,3], and meta-analyses show modest CNCP relief [4]. The risks of adverse side effects [5], including opioid-related overdose death [6,7], remain a major public health concern [8-11], especially as (1) access to interdisciplinary rehabilitation is restricted by cost, treatment intensity, geography, demand [12-16], and patient fears about opioid discontinuation [17-19] and (2) the effectiveness of behavioral and pharmacological treatments to reduce the risks of opioid use in patients with CNCP is inconclusive [20,21].

Digital brief interventions (BIs) may offer a nonpharmacological option to reduce opioid-related harm [22,23]. BIs are widely used in substance use treatment, and they involve screening, feedback assessment, education, and goal setting, motivating those at risk to modify substance use and engage in further treatment if required [24]. BIs are effective in reducing substance use and related harm in various settings (eg, primary care and emergency departments) [25,26] and have shown similar effectiveness to more intensive psychological therapies [27,28]. Digital BIs are becoming an increasingly attractive, scalable technological innovation in the management of pain that may be just as effective as in-person BIs in reducing substance use and related harm [29]. The use of digital BIs in the CNCP population is still in its infancy, but there is interest among patients in these types of interventions [23,30].

Broadly speaking, as effective as digital health interventions (DHIs) are, they are not widely adopted in practice [31,32]. One of the main reasons is that key stakeholders (eg, patients and health professionals) have not been a part of the solution design [33-36]. Research co-design, that is, the meaningful involvement of end users in the design, implementation, and translation of research, has the potential to reduce the risk of research failing to translate into practice [37]. Co-design offers a means to increase the success of DHIs by aligning with user needs and preferences [38,39].

Co-design of DHIs for CNCP has not made significant progress over the past decade [40]. Most DHIs for pain or medication use are developed by the for-profit software industry [41], lack a theoretical framework [42] and evidence base [43]. Only 19% of studies involve end users in the development process, and this is usually in an ad hoc manner [44]; few studies support clinician access or involvement [42,44,45].

Using a co-design approach with patient partners makes it possible to design a fit-for-patient-purpose digital BI, which may accelerate its adoption and impact considerably. The first step of the co-design of DHIs is to obtain a rich and deep understanding of people's experiences within the disease management context while also sensitizing them to the problem space [46,47]. Despite the diverse and unique lived experiences of those with CNCP (eg, [48,49]), DHIs are rarely designed in partnership with this patient population. This study aimed to understand the lived experiences of patients with CNCP with a particular focus on opioid therapy—the first important step of co-design. The patients were informed of the overall goal to co-design a digital BI and then asked to describe their experiences of CNCP (past, present, and future) management, including opioid therapy. This new knowledge will inform future steps in the co-design process of digital BIs for the population that is most at risk of harmful use of opioid medicines.

Methods

Participants

Participants were eligible if they were aged between 18 and 70 years, living with CNCP (defined as a sensory and emotional experience that lasts or recurs for ≥3 months), and seeking treatment from public health specialist addiction or pain services. Participants were also required to report clinical levels of pain, as indicated by scoring ≥4 out of 10 on the Pain Numerical Rating Scale on average over the past week.

Participants were excluded if they were assessed as being highly distressed (based on scores ≥13 on the Kessler Psychological Distress Scale [50] and follow-up telephone risk assessments by the research team psychologists [RE and KB]); were non–English speaking; had a history of recent injecting drug use; or had currently tested positive with SARS-CoV-2. In addition, patients recruited from addiction services were excluded based on the clinical judgment from the addiction specialist if they had an acute, severe mental illness (eg, active psychosis).

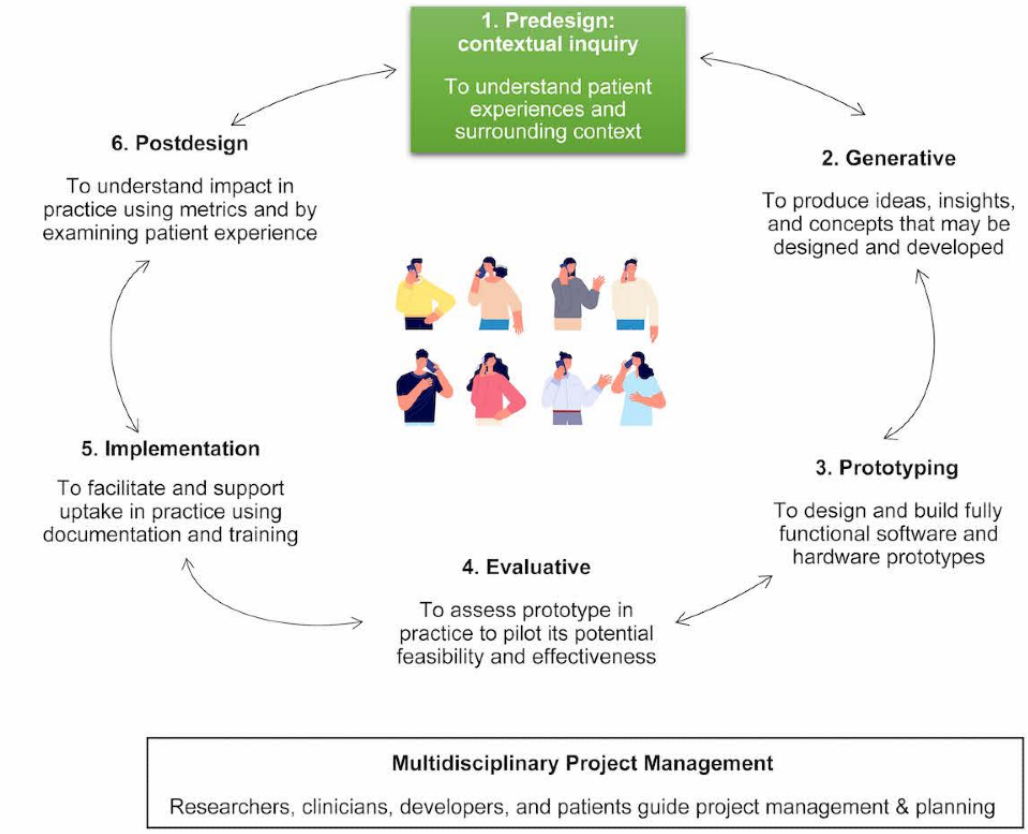
Design

A co-design approach [51] was used. We applied the co-design framework by Sanders and Stappers [47] and updated by Noorbergen et al [46] during the co-design of the digital BI. It outlines 6 iterative co-design phases: predesign, generative, prototyping, evaluative, implementation, and postdesign. Figure 1 presents an overview of the co-design framework. This study reports on the first phase of the iterative co-design development process – the exploration of patient experiences as part of predesign (ie, contextual inquiry). A systematic review

XSL•FO
RenderX

examining the state of the evidence and an adjunctive study examining patient preferences for BIs to reduce pain, opioid use, and related harm was also conducted as part of the first phase and is reported elsewhere [23].

Figure 1. Overview of the iterative co-design framework used in this study. The predesign phase, denoted by the box highlighted in green, is reported in this study.



Recruitment

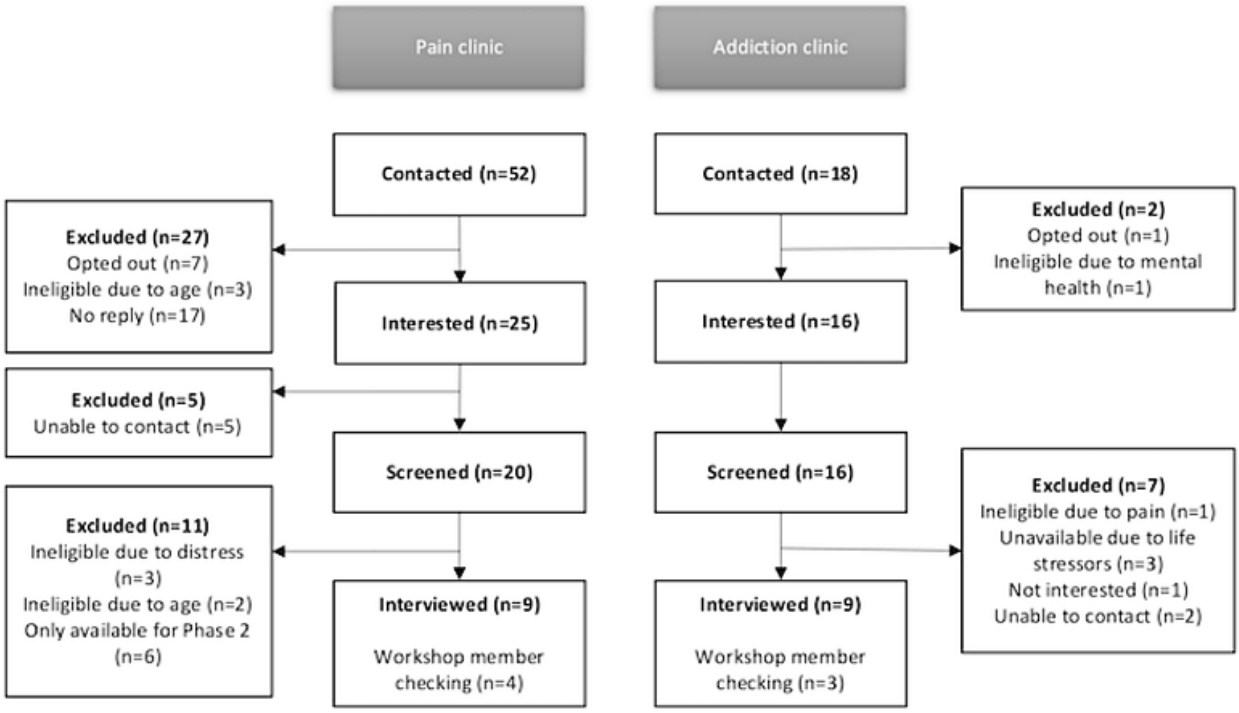
Treatment-seeking patients who were currently using prescription opioids or had a history of opioid use were purposively sampled. Participants with CNCP were recruited through a tertiary hospital specialist pain clinic or community addiction specialist service. Patients from the pain clinic were engaged by mailing an information pack inviting those on the waitlist to participate. Participants from the hospital addiction service were invited by the medical specialist and a member of the research team (RE) during their on-site appointment at the clinic. The final sample size was determined by thematic saturation using the approach outlined by Guest et al [52], applied during the data collection and inductive thematic analysis phase using 3 components: base size, run length, and new information threshold. The base size refers to the body of information identified in the initial dataset (ie, initial number of themes) and is used as a denominator of the saturation ratio. The run length refers to the number of interviews or other data collection events based on which new information or themes are identified. New information threshold refers to the level of information accepted as indicative of saturation. A new information threshold of ≤5% was selected to indicate that we had reached adequate saturation.

Procedure

Recruitment and interviews were completed from June to November 2020. Participants completed an initial telephone screening interview for us to determine eligibility. If eligible, patients completed a web-based questionnaire before the interview for descriptive purposes and to provide background to guide the interview. Interviews were audio-recorded, professionally transcribed verbatim, and thematically analyzed. Member checking of themes by patients who participated in the interviews was conducted during a follow-up workshop, which focused on the design components of the BI (not reported in this study). Figure 2 presents a summary of patient flow.

XSL•FO
RenderX

Figure 2. Patient flow.



Measures

Demographics

Demographic data, including age, sex, and cultural background, as well as information on pain location, duration, intensity, and current medications were collected. Data on history of mental health or substance use disorder were also collected.

Brief Pain Inventory

The Brief Pain Inventory [53,54] is an 8-item measure that assesses the pain experience, including location of pain, pain intensity and interference on a scale from 0 (no pain) to 10 (pain as bad as you can imagine). Scores are summed to produce 2 separate scores of pain intensity and pain interference, with higher levels indicating greater intensity and interference. An example item is as follows: “Please rate your pain by circling the one number that best describes your pain at its worst in the last week.” The Brief Pain Inventory is a validated and widely used measure in clinical settings with efficacy in both patients with cancer pain and those with chronic, noncancer pain [53,54].

Kessler Psychological Distress Scale

The Kessler Psychological Distress Scale [50] is a 6-item measure used to assess a person’s level of psychological distress. Participants were asked how often they have felt a range of emotions over the past 4 weeks (eg, “How often did you feel anxious?”). Responses are scored on a 5-point scale (0=none of the time; 4=all of the time). Scores ≥13 indicate the presence of a possible mental illness [50]. This scale has demonstrated excellent reliability [50].

Depression, Anxiety, and Stress

The 21-item Depression Anxiety Stress Scales [55] asked participants to rate to what degree depressive, anxiety, or stress symptoms have impacted them in the previous week on a scale from 0 (did not apply to me at all) to 3 (very much or most of the time). Items from each subscale were summed and multiplied by 2 to denote a patient’s current level of depression, with higher scores reflecting higher levels of depression. An example item from the depression subscale is as follows: “I felt that I had nothing to look forward to.” The 21-item Depression Anxiety Stress Scales has previously been shown to be a reliable and valid tool in a chronic pain sample (Cronbach α=0.96 for depression, Cronbach α=0.89 for anxiety, and Cronbach α=0.95 for stress) [56].

Current Opioid Misuse Measure

The Current Opioid Misuse Measure (COMM) [57] consists of 17 items measured on a 0 (never) to 4 (very often) scale, which asks patients about their current opioid use behaviors. Scores are summed, with higher totals reflecting higher levels of opioid misuse. An example item is as follows: “In the past 30 days, how often have you taken your medication differently to how they are prescribed?” The COMM is a validated and reliable measure within chronic pain samples, with previous research reporting Cronbach α=0.86 [57-59].

Semistructured Interview

Semistructured interview questions explored patient experiences and their initial perspectives of a digital BI (Multimedia Appendix 1). Participants were first informed about the overall aim of the research, to co-design a digital BI to reduce risk of unsafe use of prescription opioid medications. They were then asked about their past, current, and future experiences of pain and pain management; access to treatment services and programs including psychologists and psychological treatment; and specifically, their experience of using prescription opioids. Then, suggestions for the design and implementation of the digital BI

XSL•FO
RenderX

were explored; however, these will be reported in a separate paper as part of the second (generative) phase of the co-design process [46,47]. All interviews were conducted by RE, a clinical psychologist, and KB, a psychologist, both with expertise in treatment of chronic pain and qualitative research. Field notes were made after the interviews to record the interviewer's observations and impressions.

Data Analysis

Transcripts were thematically analyzed using grounded theory [60]. Two independent coders (RE and SP) read the transcripts to establish familiarity. An inductive approach to coding was used, with codes added as they were identified in the data. The coding structure consisted of first level codes recorded line-by-line, second level codes of possible subthemes, and third level codes of higher order themes. Key quotes were extracted. The analysis was conducted iteratively, with codes cross-checked between coders approximately every 3 to 5 transcripts to identify any discrepancies, which were resolved by further discussion. Theme refinement was undertaken between the coders and another interviewer (KB), integrating the interviewer's field notes to aid interpretation and reduce coder bias.

Member checking [61] was conducted in a subsample (n=7) of interviewed participants in subsequent workshops to explore the validity of the themes as part of the assessment of data saturation approach, that is, to check whether the themes were consistent with the participant's experience. This involved facilitators providing verbal feedback on the themes integrated with the workshop activities. We sought disconfirming voices (objectivism) and provided opportunities for reflection on personal experiences, creating opportunities to add data (constructivism) [62].

Ethical Considerations

Ethics approval was obtained from the relevant institutional review boards (HREC/2020/QMS/60695). All participants provided written informed consent. Study data were deidentified for analysis, and any identifying information was removed before publication, including patient numbers, to ensure complete anonymity. All participants were offered an Aus $50 (US $33) gift voucher for each session attended to thank them for their time.

Results

Data Saturation Analysis

Table 1 presents a summary of the data saturation analysis. We reached adequate saturation (≤5% threshold) after 16^{+2} in-depth interviews. The final 2 extra interviews did not add substantially to the body of information collected; however, these ensured a purposive sampling approach included a balance of male and female participants and people from various culturally and linguistically diverse backgrounds.

Table 1. Data saturation analysis.

Step	Outcome
1. Number of themes for the base interviews[a]	8 × PID_P[b] interviews
2. Number of unique themes for the first run	5 new themes from 4 × PID_A[c] interviews
3. Saturation ratio	Number of new themes/run: 5/30 = 17% (>5% threshold); continue with interviews
4. Number of new, unique themes for the next run	2 new themes from 5 × PID_A interviews and interviewer reflections: 2/30=7%
5. Updated saturation ratio	7% (>5% threshold); continue with interviews
6. Number of new unique themes for the next run	1 new theme from 2 × workshop member checking: 1/30=3%
7. Updated saturation ratio	3%, less than ≤5% threshold
8. Number of new unique themes for the next run (to address gaps in demographics)	1 new theme from 1 × PID_A and 1 × PID_P interview: 1/30=3%
9. Updated saturation ratio	3%, ≤5% threshold

[a]Denominator in the equation.

[b]PID_P: patients from the pain clinic.

[c]PID_A: patients from specialist addiction services.

Sample Characteristics

The sample included 10 women and 8 men, with a mean age of 49.5 (SD 11.50; range 25-62) years. One-third of the patients self-identified an ethnicity in addition to Australian descent, including 2 from Aboriginal and Torres Strait Islander descent. Four (22%) patients were employed either part time or full time. Approximately half of the participants reported their highest level of education as grade 12 and 4 (22%) participants had at least an undergraduate degree. More than half of the patients were in a relationship.

The sample predominantly reported back or neck pain (89%). Duration of pain ranged from 6 months to 42 years (mean 11.57, SD 11.55 years; median 7 years). Average pain severity in the past week was 5.82 (SD 1.51), and average pain interference in the past week was 5.80 (SD 1.91). In total, 16 (89%) patients were currently using prescription opioids (range 1.15 mg/d [pro re nata] to 480 mg/d), and 2 (11%) patients had used opioid

XSL•FO
RenderX

therapy in the past. Half of the participants reported current opioid misuse (scores ≥9 on the COMM) [57]. Three (19%) patients were prescribed high-dose opioids (≥100 mg/d) [63]. Seven (39%) patients reported moderate depressive symptoms. Three (17%) reported severe or extremely severe depressive symptoms; these patients were included as they were assessed to have adequate supports and risk management plans, which supported their involvement. Table 2 presents a summary of individual characteristics (see Multimedia Appendix 2 for a more detailed summary of individual patient characteristics).

Table 2. Sample characteristics of patients.

Age (y)	Sex	Pain duration (y)	Pain interference (past week)	Pain severity (past week)	COMM[a]	OME[b]
33	Female	15	2.71	2.25	4	1.15
52	Male	2	9.14	7	5	30
60	Female	3	5.00	6	7	45
31	Male	5	4.14	5	14[c]	60
47	Female	0.5	4.71	6	—[d]	—[d]
61	Female	4	2.86	5	8	—[d]
53	Female	14	4.71	4	10[c]	70
47	Female	0.83	6.57	6.75	20[c]	22.8
58	Female	42	8.86	8	13[c]	143.4
60	Female	2.5	6.43	7.25	2	37.9
54	Male	20	7.43	7.25	25[c]	480
25	Male	3	5.83	6.5	1	4.2
58	Male	11	6.43	5.5	15[c]	N/A[e]
53	Male	20	5.29	5.75	13[c]	N/A[e]
56	Female	28	8.00	6.75	3	N/A[e]
48	Male	9	4.14	6.5	11[c]	90
33	Female	3.5	4.29	3	27[c]	480
62	Male	25	7.71	6.25	16[c]	20

[a]COMM: Current Opioid Misuse Measure.

[b]OME: oral morphine equivalence (daily).

[c]These scores meet the cutoff for the scale as risk of unsafe opioid use.

[d]Missing.

[e]N/A: not applicable; patients were using methadone and dose could not be calculated.

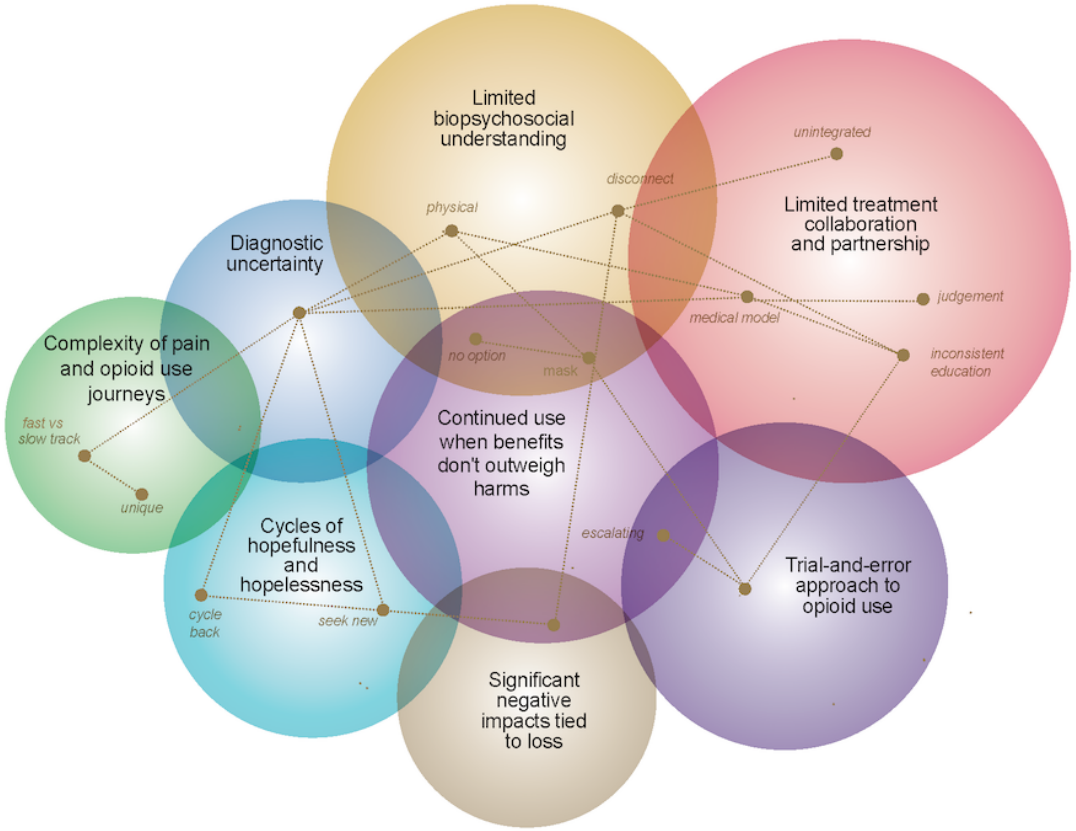
Thematic Analysis

Overview

Eight overarching themes and 13 subthemes were identified from the data, listed in the order of prominence from most to least prominent: limited treatment collaboration and partnership; limited biopsychosocial understanding of pain; continued opioid use when benefits do not outweigh harms; a trial-and-error approach to opioid use; cycles of hopefulness and hopelessness; diagnostic uncertainty; significant negative impacts tied to loss; and complexity of pain and opioid use journeys. Figure 3 provides an overview of the themes; larger circles reflect more prevalent themes. In subsequent sections, note that PID_P refers to patients from the pain clinic and PID_A refers to patients from addiction services. Patient numbers were removed before publication to ensure complete anonymity.

XSL•FO
RenderX

Figure 3. Overview of themes and subthemes. The overlap of circles represents connections between themes, and lines represent connections between subthemes.



Theme 1: Limited Collaboration and Partnership During Treatment

This theme reflected the lack of a patient–health professional collaborative partnership and an integrated approach among health professionals in the care of patients.

Patients Initially Followed the Medical Model, With Faith and Trust

Initial opioid prescription was usually prompted by the treating physicians in response to an acute episode of pain or following surgery. Patients with a long history of pain described early prescribing practices as including more passive rather than active involvement of patients. One patient (PID_P) reported the following:

> *So the Dr says, look here, take this, this will stop the pain. So, you go, ok, you take it.*

Patients said they often continued to follow the medical model, with faith and trust in the treatment approach guided by their physician. This suggested a lack of shared decision-making about whether opioids were the best or right treatment option and whether the benefits outweighed potential harms—"I'm in the hands of the Doctors" (PID_A). As their pain persisted over time, trust in opioid or medical treatment decreased and participants tried to seek alternatives treatments to manage pain. Some patients reported remaining very trusting of their general practitioner (GP) and were more actively involved in discussing options with them. Others reported becoming distrustful and even angry.

Inconsistent Opioid Use Education, Monitoring, and Support

There was variation in the amount of education about opioid use that the patients received when first prescribed. Some patients said they received education from their physician or pharmacist about the potential harmful effects of prescription opioids (eg, side effects and addiction potential). Other participants could not recall receiving any education in this regard:

> *A lot of doctors don't have communication or you know, they'll just walk in, they'll give you the script and walk out. But I think it's very important for the doctors, as the ones prescribing to actually sit there and go, all right, here's the side effects.* [PID_A]

The reported level of ongoing education and medical monitoring of opioid use over time also varied. One patient (PID_P) suggested, "They should have an opiate diary and they should make people keep diaries of when they take them and how many, that would allow them to monitor usage because I get given a prescription, but nobody actually monitors my usage."

Patients also described the discussions they had with their physician about possible changes to their opioid use. Some patients felt that doctors prescribed opioids when they were not sure about or had exhausted other treatment options. For example, one patient (PID_P) reported, "I have tried to ask my doctor is there anything else I can try to see if it works better...his advice was there's not really anything else that's going to be able to help." Lack of education appeared to fuel questions later during treatment. Some patients questioned why

XSL•FO
RenderX

they were still using opioids, given that they initially expected opioids to be a relatively short-term solution for pain: "I personally thought this (opioids) might be taken 6-8 months maybe a year, we'll fix it, back to work" (PID_P). One patient said that doctors had different opinions and that the tapering plan differed to information on the internet, so they did not progress with it. Some reported being encouraged to taper in the context of recent policy changes to accessing opioids but did not always feel they had adequate support to do this. One patient (PID_A) said the following:

> *You know, and they keep saying to you try and cut down, you try and cut down so much. You're in pain, cutting down, you show that you can try and cut down.... I know I don't want to be on them, I should be off them by now if it had of gone right. I think you just feel isolated because we don't have anybody we can really talk to about our pain anymore.*

Some patients described that they had not discussed tapering with their physician as they were worried that pain may increase and were fearful of the physician's response. One patient (PID_P) said the following:

> *I was too scared to say to him "look I want to start reducing" because I just thought if I couldn't cope with the pain on a reduced dose he many not let me go back up again and then I would be stuck with pain. So that stopped me sort of officially asking for help.*

Health Professional and System Judgment, Distrust, and Stigma

This theme related to the patients' general feeling of disillusionment and distrust of mainstream medicine because of continued pain and seeking to apportion blame (eg, "I don't trust doctors. I don't trust physios. I don't trust anybody because of what they've done to me" [PID_A]). When patients described more active attempts to discuss pain relief, concerns about addiction reportedly impacted the clinician's decision to prescribe opioid medications. One patient (PID_P) stated the following:

> *Drs are more bothered about other people's addiction rather than prescribing me pain relief.... My pain was not her concern when I went and asked for strong pain killers.*

These reports of mistrust were also observed when patients discussed the impact of recent policy changes in Australia that restricted access to over-the-counter codeine medicines (February 2018) and medications for CNCP (June 2020). There was an understanding that the system and the message had changed, leaving them in a difficult position: "well the medical community's clearly been saying for a long time 'you need to take this, you need to take this', and then they say 'well no, you're not allowed to take them anymore', you're left with 'well, nothing else works, what are you going to do?'" (PID_A). For patients from addiction services, there were cases where their physician had ceased prescription, which prompted referral to an addiction specialist and engagement in an opioid replacement program. This was accompanied by feelings of abandonment toward their previous prescriber and, broadly, the medical system and ambivalence about accessing an addiction service. For some, this led to consideration of other treatment alternatives, including psychological therapy and cannabis (both prescribed and nonprescribed).

Several patients reported feeling increasingly sensitive to questioning and comments about their opioid use and chronic pain condition:

> *She would sort of say well what medications are you on? Oh, that's a high dose. oh that's been a long time. I feel like they make you feel like you're doing the wrong thing but nobody is offering any sort of solutions.* [PID_P]

Some patients reported finding this confronting and exhibited some defensiveness to the possibility of addiction or dependency. One patient (PID_A) said, "So I don't think I was addicted to it, I just took it because I got pain."

This appeared to contribute to the possible stigma associated with addiction more broadly. A patient (PID_A) explained as follows:

> *My doctor sent me there.... It kind of made me feel like I had a habit or had an addiction. That kind of really put me on the spot, and I didn't really want to go. But I went anyway. So if you put someone that's willing to cut down, but then you send them to a place like that they might feel attacked in a certain way. Like they're in the same category as those people but they're trying to do it for good reasons, not because they're addicts.*

One patient (PID_A) also described being on methadone as stigmatizing as "only heroin users use methadone." Discussions indicated that some patients felt that addiction services were not the right place for them to seek treatment and that they were *different* from other drug users. There was also a difference for people between acknowledging that they may be addicted to opioids versus the label of "addict" or "junkie." It is possible that this may contribute to an identity crisis for patients (linked to the theme of significant negative impacts linked to loss): if they are not engaging in medical treatment for a medical problem, then they either are a "junkie" or have a mental health problem ("it's all in my head".

Lack of Coordination and Integration of Care

Patients reported seeking many different types of treatments and supports for their pain. However, there was little reported integration or coordination of care between different providers, the public and private sectors, or traditional and alternative treatments. One participant shared a positive experience with a multidisciplinary rehabilitation program:

> *It was a triple pronged approach with having the psychologist, the physiotherapist and the exercise physiologist. That was really good.* [PID_P]

Another participant who was referred to the pain clinic by their GP was not receptive to an integrated approach: "the guy saying 'you'll see physio and you'll see psychologist and you'll see this and you'll see that' and I'm thinking I've tried that, been there, done that" (PID_A).

XSL•FO
RenderX

Theme 2: Limited Biopsychosocial Understanding of Pain

Focus on Physical Aspects of Pain

Most participants focused on the physical aspects of pain and minimized the potential role of psychosocial factors in exacerbating pain. One patient (PID_P) said, "Well I know where my pain is, my pain is a deterioration of the bones...so I don't think it's psychological." There was also reference to experiences where one participant stated, "mine is physical pain not psychological pain" PID_P), with another participant describing pain as being perceived as separate from the mind:

> *Like if you're in pain, how can your mind tell you that you're not in pain? That's how we're looking at it. As if to say, well, you're going to go get pain management, and they're going to psychologically move your brain around to say that you're not really in pain, you don't really need these tablets.... but that still doesn't have to do with my pain. And I try and keep the two separate because pain is one thing, and keeping my mind sane is another thing.* [PID_A]

Some participants did describe how psychosocial factors influenced their pain experience. This knowledge often emerged in hindsight or in a precontemplative state of considering treatment options other than opioids. Examples of the possible role of biopsychological factors were often inconsistent or lacked detail. For example, one patient (PID_P) said, "I think there is a big mental side to it as well." Overall, patient connection with social influences was limited.

Disconnect Between Mental Health Problems and Pain and Engagement in Psychological Treatment

Several people described having a co-occurring mental health problem (eg, depression, anxiety, and posttraumatic stress disorder) for which they were receiving treatment from a psychiatrist or psychologist. This was also reflected in the clinical characteristics of the sample. Despite this, there was little evidence that they perceived their mental health as influencing the levels of pain experienced or vice versa.

Participants did not describe accessing psychological treatment specifically for pain or mention that psychological treatments were recommended. One participant (PID_P) reported:

> *With only recently going to the pain management sort of introduction that they told us that psychiatrist can...psychologist can sort of discuss pain with you and I just didn't know that that had anything to do with being on pain killers....I think it's fantastic because I think the brain is a powerful thing and I saw what cognitive behavioural therapy did with my son.*

Some patients reported positive experiences with psychological treatment for mental health problems but did not link how psychological treatments could be potentially effective for pain.

Theme 3: Continued Opioid Use When Benefits Do Not Outweigh Harms

Overview

Participants reported relatively more risks and consequences (eg, side effects and limited pain relief) versus benefits of their opioid use, but they continued to use them. Patients said the following:

> *I would reduce it down to nothing if I could...because I don't like taking them or just the way it makes me feel.* [PID_P]
>
> *Well if I could maybe reduce some, it would be good, But I can't see it happening.* [PID_P]

Patients also described frustration about taking opioids for pain, while at the same time, they reported holding out hope that there may be other options (eg, "it's a catch 22...I have to think clearly if there's a better way of doing it, than to be popping pills every day, then I'm open to anything" [PID_P]).

Opioids Mask the Pain

Some patients in this sample reported that opioids did not provide effective pain relief, but this varied from person to person; for example, "It's not like taking painkillers is actually stopping or making the pain better" (PID_P) and "(opioids) doesn't knock all the pain over" (PID_A). There were reports that opioids "take (pain) down a few notches" (PID_P); "take the edge off" (PID_A and PID_P); or "drowns out the pain" (PID_P). Others did report that medications got rid of the pain (PID_A) and that opioids work better than other medications in the market (PID_A).

In the cases of limited pain relief, opioid medications were often not reportedly taken as prescribed (theme 4): "...even the opioids aren't really helping with this pain. So that's why I haven't been taking them religiously, like I probably should have been" (PID_P). Another participant said, "it just puts me to sleep, I don't feel anything...I don't get the pain relief from the pill, you get pain relief from sleep" (PID_P).

No Options Other Than Opioids

Most participants reported that they had no other option but to use prescription opioids for pain. One patient (PID_P) said, "I can't not take it, I don't have any other option"; another patient (PID_A) said, "But with me, I'd rather not have them myself, you know, I don't know any other way you can deal with the pain when an operation has gone bad." For some patients, there was a view that they had no choice but to take opioids, which was related to feelings of frustration.

Escalating Dose and Need for Opioids

Participants described a history of escalating dose over time. Patients said the following: "it was two, then four, then six then eight Panadeine Forte...my body had totally adjusted to them...my body was saying, yep I'm having more and more of those" (PID_P); "That's the first time I was put on an opioids, Endone think it was, and then started off through there, and as the pain got worse over the years the medication got higher" (PID_A). Patients often initiated dose increases with their health professional. One patient demonstrated awareness of this:

XSL•FO
RenderX

"because in the long-run I'm going to need a detox clinic, then I'm going to need someone from this clinic...psychologists and psychiatrists" (PID_P). Some patients may have been experiencing withdrawal: "sometimes I just forget my lunchtime dose and what reminds me is the severe drop in mood" (PID_P). Continued use despite negative consequences (eg, social, physical, or psychological problems) and features of tolerance and withdrawal were also reported.

Theme 4: Trial-and-Error Approach to Opioid Use

Some patients reported that they were taking opioid medicines as a "trial-and-error" approach to get relief. This appeared to be related to efforts to take control of their pain. Some people said that they took their medication when they were in pain rather than as prescribed:

> *I try not to take the opioid if I can help it. Wait to see how bad the pain gets before I take it (opioids)...the pain is unbearable. So, I get to the stage where nothing's helping.* [PID_P]

> *I do try to do without them and as long as I possibly can and then sometimes, the pain is so bad...then waste half a day in bed, I get pissed off at myself about it.* [PID_P]

These reports reflected patients' delayed use of opioids when in pain to minimize reliance on opioid therapy; that is, patients were hoping or taking the risk that their pain would not become more severe, so they did not have to use opioids.

Others described overuse of opioid medications:

> *Take more... Yeah. Yeah. I have four of what he gives me for that day and then there's one for the morning...My overuse of Oxycontin.* [PID_A]

Some patients reported a history of overuse: "Yes, it was taking too much Targin and before that it was too much morphine, I was actually on morphine and I was taking too much of it" (PID_A). Opioids were also used for other reasons such as sleep or improved mood. For instance, a patient (PID_P) reported the following:

> *It just puts me to sleep, I don't feel anything...I don't get the pain relief from the pill, you get pain relief from sleep.*

Patients also reported using stockpiled opioid medications:

> *I had some old medication that I had from old surgeries...so I would take them when my pain was like 9 out of 10.* [PID_P]

There were also examples of patients overusing other medicines (eg, benzodiapines and antiepileptics) in conjunction with opioids when they were in severe pain. There were examples where patients decided to self-taper their opioids, without health professional support:

> *I tried to go cold turkey initially and that was shocking...I thought I would try and reduce on my own.* [PID_P]

Theme 5: Cycles of Hopefulness and Hopelessness in Desperation to Find Pain Relief

Overview

Patient reports reflected cycles of hopelessness and hopefulness associated with pain treatment in a desperate effort to find relief. Those with increased pain or pain moving to different parts of the body over time reported a sense of hopelessness about the future: "In the last 3 years, it's gotten worse and I don't know what else is going to work" (PID_P). Desperation to find pain relief was commonly reported among participants, which was reflected in the sheer number and variety of other strategies tried and tested over time. This theme linked to the unknown cause of pain and uncertainty that prompts engagement in help-seeking behaviors to either find the source of pain or physical treatments or other strategies to relieve pain (theme 6). While patients tended to use less treatments or strategies over time due to feelings of defeat or failure, burden of treatment, and financial reasons, some seemed to be always looking for the next treatment opportunity.

Cycle Back to Previous Treatment Options

Patients described a pattern where they engaged in various pain treatments (eg, physiotherapy and chiropractic treatment) often earlier in their pain journey and that it did not help or became increasingly less effective. This was associated with feelings of hopelessness and defeat, resulting in them taking a break from treatment or trying other strategies until they were prompted to engage in physiotherapy again. One patient (PID_P) reported as follows:

> *To be told, I want you to go back and do physio and then stay on tramadol and come back in 10 months.... I have already been through all of that.*

Some patients reported significant frustration that went along with what they assessed as ineffective treatments.

Seeking New Treatments Despite Hopelessness

Participants appeared to hold feelings of hopelessness about current pain treatment and hopefulness that something still may help in the future, at the same time. Some patients were optimistic for a cure and still wanted to be "fixed." One patient (PID_P) reported, "all I'm asking for is just to get fixed and just go back to work and get my life back on track." Patients said that hope tended to decrease over time, with one participant reporting, "I was definitely more hopeful earlier on, I'm just thinking there's going to be something that's going to be sorted out" (PID_P). Despite this, some patients were open to the idea of receiving psychological treatments as well as medical or surgical interventions or alternative treatments (eg, stem cell therapy) in the *hope* that this may relieve their pain.

Theme 6: Diagnostic Uncertainty and Associated Help-Seeking Behaviors to "Fix"

This theme reflects the participants' uncertainty about the underlying causes of pain which were varied and often unexplained, contributing to ongoing attempts to understand the origins of their pain. Distress, anger, and a sense of unfairness was often tied to the triggering pain event or onset of pain, which was still apparent many years later.

XSL•FO
RenderX

For some individuals, the main explanations for the originating causes of their CNCP were seemingly unusual and involved either internalization (eg, diet and posture) or externalization (eg, weather and surgical or medical interventions going wrong) of blame. For others, there was no known cause for primary pain and patients had difficulty pinpointing the triggering event to their CNCP condition. One patient (PID_P) reported, "I couldn't tell you because it has always been there on one level, but over the course of the years it's gotten to the level where it is now." Diagnostic uncertainty contributed to ongoing attempts to understand their CNCP (cause) and find a "cure" (treatment) through more assessments and tests (eg, "...let's see what happens with this MRI" [PID_A]). Help-seeking behaviors were associated with hopeful or hopeless feelings about pain treatments and their effects (theme 5). Particularly, in cases where medical diagnoses did not explain the cause of pain, this resulted in patients being in a constant holding pattern, waiting for treatment or further tests.

Theme 7: Significant Negative Impacts Linked to Loss

Participants reported a significant number and range of negative impacts of their pain and pain management. These impacts were biological and physical (eg, dental problems); psychological and functional (eg, cognitive and memory problems, sleep disturbance, difficulties engaging in daily tasks, and low mood); or social and economic (eg, social isolation, financial burden of treatment, can't drive, and intimacy and relationship problems) in nature. One participant (PID_A) spoke about pain as a disability. Another shared that these impacts were often centered around grief and loss: "Most days I feel like I'm letting everybody down in my family because I'm not out there making money" (PID_P). The interviewer observed that participants become the most distressed in the interview when they compared their life before the pain to how it is now—and what they had lost or had to try to change or adapt. This was interpreted as a sense of loss of participating in life and normality.

For some, this sense of loss was also related to a loss or changing of identity:

> *If I come off the 10 tablets, am I going to be in more pain or am I going to be in the same pain as I am now...but have me back I'm me, have me back, just be myself.* [PID_P]

This included changes to their identity over time—to a "chronic opioid user" or "addict." Others noted a more recovery-focused interpretation of their identity with opioid use. A patient (PID_A) said the following:

> *Well, that's just part of what you did. It's not who you are. It's what you do. It doesn't have to define you. You can live a normal freaking life, with or without opioids. You've got a choice.*

Importantly, the varying ways people used prescription opioids (theme 4) was interwoven with these impacts; that is, the use or misuse of opioids was often to reduce the impacts of pain on their life.

Theme 8: Complexity of Pain and Opioid Use Journeys

Participants described complex pain journeys that involved variations in pain experience over time (got better or worse) and either a slow or fast track to opioid use. The uniqueness of the pain experience was also highlighted.

Slow Versus Fast Track to Opioid Use

The slow track to opioids often started with physiotherapy and primary care efforts to relieve pain and, in some cases, for specialists to "fact find." Over the course of their pain journey, there was a common feeling that patients were exhausted trying to investigate the physical or organic cause of pain. Because of failed treatments, patients were often initiated on opioids. One patient (PID_P) said, "cause he's the one that ... organised me to go to physiotherapy and all this, and that's when it didn't work. That's when I really started on this, using that." This pathway was common among patients from the pain clinic.

For other participants, a fast track to prescription opioids was initiated as part of acute pain management, associated with acute injury or surgery. These individuals were initiated on opioids in hospital and on discharge. A patient (PID_A) described as follows:

> *Pretty much straight away when I was in pain (in hospital).... Yes, they gave me medication to take home and then take the deep heat and that's how it started.*

Participants described either positive or negative experiences of being prescribed opioids for the first time. Some patients reported immediate pain relief or euphoria. This sense of relief was reported as sometimes occurring after previously trying multiple types of opioids. Other patients described experiencing adverse side effects and limited longer-term pain relief:

> *I knew a little bit about it but it didn't really deter me from it because, you know, not going to work, not paying bills, seems more of a downside than just taking something throughout the day that can help you continue on with your life.* [PID_A]

Everyone's Pain Experience Is Unique

Patients described the experience of pain as unique to each individual. One participant (PID_A) said the following:

> *Everyone is different... Because it's a very individual thing, I think. Handling pain, dealing with pain.*

Additional statements from the workshops provided further validation of this subtheme (eg, "well I've always had pain...pain is a personal thing" [PID_P]).

Interviewer Observations

During the patient interviews, there was evidence that some patients benefitted from the interactions with the interviewers (psychologists). This was evident even for those patients who reported that it was really challenging to describe their pain journey and they would not have been able to do this if it was years ago. There was a view that listening, asking nonjudgmental questions, and showing empathy resulted in some patients softening throughout the interview when they appeared to be more defensive at the outset:

XSL•FO
RenderX

> *You went about it the correct way. I used to get annoyed at people quite easily. So yeah, you done it fine, you ask questions nicely, you didn't talk over, you stopped when someone else is talking, which is a nice thing. So, you conducted it very well so it made it a lot easier.* [PID_A]

Alternatively, for one patient (PID_A), the interview brought attention to past pain experiences, and this was related to an increased experience of pain: "Like my backs starting to hurt just thinking about it."

There appeared to be increased interest in understanding and openness to the biopsychosocial model at the end of the interview (eg, compared with starting the discussion around the aim of the study). Some patients appeared to reflect upon and repeat the points on the biopsychosocial model throughout the interview.

Validity Checking

The process of member checking resulted in several changes to the theme structure. These included the following:

1. Tapering fears and lack of patient-physician partnering was integrated in theme 1. Furthermore, lack of coordination and integration of care among health professionals; influence of policy change judgment; and mistrust of medical profession were also included in theme 1.
2. We removed seeking certainty for cause of pain and looking within and looking outward for explanations; these were absorbed in theme 6. Furthermore, we refined the theme to focus on diagnostic uncertainty and associated help-seeking behaviors to capture these elements.
3. Theme 8, previously named multiple pain journeys and pathways to opioid use, was reconceptualized as complexity of pain and opioid use journeys. This was to capture the complexities of pain management (either fast or slow track to opioids) and also the unique experience of pain (a subtheme that emerged from workshop 1).

Discussion

Main Findings

This study is one of the first to describe the lived experience of individuals with CNCP in the context of co-designing a digital BI to increase the safe use of pain medicines. In the sample of patients interviewed, half of the patients met the current opioid misuse thresholds on validated psychometric scales (COMM [57]). Our main findings were the following 8 overarching themes that emerged in this sample related to the lived experience of patients with CNCP regarding pain management including opioid therapy: limited treatment collaboration and partnership; limited biopsychosocial understanding of pain; continued opioid use when benefits do not outweigh harms; trial-and-error approach to opioid use; cycles of hopefulness and hopelessness; diagnostic uncertainty; significant negative impacts tied to loss; and complexity of pain and the opioid use journey.

Overall, patients felt they had limited choice but to continue opioid therapy, as there were few perceived alterative treatment options available, a common concern also shared by physicians [64]. Patients' fears about making changes to their opioid use may have also played a role. This state of "stuckness" may have contributed to patients' sense of hopelessness about pain and its management. In many ways, patients may be rightly consumed by their pain in a way that they may be ambivalent about many aspects of pain management and life in general. Opioid therapy was also linked to behaviors that could elevate patient risk (eg, not taking medication as prescribed, self-tapering, and combining different opioids and classes of drugs) in a trial-and-error approach. Opioid use generally provided a pain "mask" with limited pain relief, sometimes; consistent with the results from meta-analyses [4]. However, some patients from addiction services described more positive experiences of opioid medications. Altogether, these experiences signal gaps in our current models of care and suggest that BIs may provide an avenue for (1) patients to explore thoughts and feelings about opioid use, (2) treating clinicians to support patient-physician partnering, (3) opioid use education (risks and harms), and (4) exploring patient ambivalence (risks versus benefits of status quo versus change) as a precursor to resolving the push and pull of behavioral change. Whether these techniques translates into behavioral change requires further testing. Certainly, there is evidence for the effectiveness of BIs in reducing substance-related harm [26,27] and growing evidence for their role in treating pain [65].

BIs provide another avenue to improve patient understanding of the biopsychosocial model of pain, critical to effective pain care [66]. The biopsychosocial model is broad and complex, and our explanation during the interview was brief and over the telephone. While participants may have received some education in the past, it was not obvious during the interview that they had a clear prior understanding of the influence of physical and psychosocial factors in the development and management of pain. When discussing the biopsychosocial model as a part of member checking, the use of visual tools to explain the concepts, the group peer setting, and reinforced messages may have contributed to patients' more positive responses to the model. Educating patients in the biopsychosocial model of pain is an important part of multidisciplinary rehabilitation and discipline-specific treatments that can be effective (eg, [67]). Efforts to enhance the education of health professionals in the biopsychological model could be important in contributing to improved patient understanding [68-70]. The consideration of patient health literacy and eHealth literacy [71] could help tailoring pain education to patient needs. These approaches are especially important, given that the lack of understanding of the biopsychosocial model may be related to ongoing diagnostic uncertainty and attempts to find a (biological or physical) source of pain and treatments that "fix."

Patients indicated that specific psychological treatment for pain was rarely offered, although many patients had previously engaged with psychologists or psychiatrists for other co-occurring mental health reasons. These reports are consistent with survey data from the United States indicating that 37% of patients are not aware of this treatment option and only 33% had ever worked with a pain psychologist [69]. This is despite psychological treatments being one of the recommended

XSL•FO
RenderX

evidence-based treatment options for the treatment of chronic pain and in conjunction with opioid therapy [3,8,72]. Approaches that raise awareness and provide education on the role of psychology in the treatment of pain are clearly needed not only at the patient level but also at the provider and public health systems level [73]. Most patients reported being open to this approach. Therefore, it is necessary for the psychologist workforce to meet this need for psychoeducation [13]. This will be especially important, given the psychologist workforce shortages and growing demand in the post–COVID-19 pandemic era [74].

Patient–health care professional communication and relationships had a significant impact on patients' emotional well-being and in seeking appropriate pain care. Patients felt that various aspects of opioid therapy (eg, monitoring, dose reduction, and policy change) could have been managed differently. It is acknowledged that we do not know about the patient's prescription instructions or if there was a management plan to taper them (as recommended in clinical guidelines). Our findings suggest that partnerships could be optimized by enhancing the patient–health care professional alliance and shared decision-making. Assuring patients that they will not be neglected by the treating practitioner in their medication management is essential [75], as is the need to make patients feel heard and understood. There were several examples of effective patient–health care professional communication and effective multidisciplinary management of the patients' pain and health concerns. Our study found that these examples of patient care were positively received by patients.

Questions in our interviews explored the patient's pain journey, including participation in psychological treatments as well as the pros and cons of opioid use, consistent with a motivational interviewing framework. Applying patient-identified motivational BI strategies may give patients with complex chronic conditions a space to share and feel heard. This can often be the first important step toward facilitating behavioral change [76,77]. Motivational interviewing BI techniques might also be particularly beneficial in exploring ambivalence, that is, teasing out where adverse effects and consequences of opioid use outweigh the benefits (ie, positive effects on pain or daily functioning); pain triggers; different pain management strategies; or highlighting unintended consequences to those who do not know they are at elevated risk of opioid-related harm.

Recommendations for the Design of BIs

On the basis of our research findings, we provide 6 core recommendations (not in any particular order) for co-designing and codeveloping digital BIs. Textbox 1 presents a summary of the recommendations. Future research should consider the potential barriers and enablers to implementing these recommendations during the design and development process. For example, patients may be reluctant to consider psychological therapies or to see a psychologist (recommendation 1), as their pain experience may be perceived as "all in their head." Health professionals (especially GPs) may believe that they do not have enough time to address aspects of pain management, such as the ups and downs of the pain experience and pain treatments (recommendation 2), given the demands of primary care.

XSL•FO
RenderX

Textbox 1. Summary of recommendations for the co-design and co-development of digital brief interventions.

Recommendation 1: educate about the role of psychologists and behavioral therapies in pain care

Educate patients and health professionals about the role of psychologists and mental health professionals and the evidence and clinical recommendations for use of psychological therapies in pain care. This is necessary to ensure there are no missed opportunities for patients to access psychological treatments earlier in their pain journey or as a first-line treatment.

Recommendation 2: normalize the hopefulness and hopelessness cycles in pain treatment

Normalize the cycle of pain treatment and associated feelings of hopefulness and hopelessness in the proposed brief intervention (BI) by discussing expectations and potential challenges, such as the idea of setbacks. Informing patients that they may go back to strategies that they have tried before may be helpful. The degree to which they are experiencing hope and optimism about treatment could be acknowledged and explored. Sharing what works with peers could enhance pain self-efficacy and social connectedness.

Recommendation 3: communicate with compassion and empower patients in patient–health care professional partnerships

Incorporate key principles of partnership and shared decision-making between health professionals and patients. Inviting patients to share their story could strengthen patient–health professional relationships as well as acknowledging feelings of loss and frustration. Strategies to empower patients and communicate with compassion are needed. Regular monitoring and feedback of opioid therapy by health professionals may assist patients. Digital solutions could also optimize this goal.

Recommendation 4: inform about the risks of unsafe opioid use and explore any ambivalence

Weighing the individual benefits versus harms of current opioid use may assist in exploring ambivalence about change. Educate patients about the possible harms of opioid use and contraindications. Developing insight into potentially unsafe opioid use patterns may be aided by monitoring, assessment, and feedback. Motivational interviewing may provide a guiding framework.

Recommendation 5: educate about the biopsychosocial model of pain

Educate patients about the biopsychosocial model of pain in a simple and engaging way; that is, the dynamic interaction among physiology, psychological, and social factors that can contribute to and perpetuate pain. Provide textual and visual cues and supporting materials to aid explanation. It will also be important to address ongoing patient efforts to seek a medical or physical explanation for the onset and maintenance of pain. A clear definition of psychological treatments is also required as well as consideration of marketing of the BI, particularly when patients have a history of failed pain management strategies.

Recommendation 6: acknowledge stigma and the uniqueness of pain and personalize management to different care pathways

Acknowledge stigma and personalize management to different care pathways (eg, slow or fast track to opioids and different referral or service perceptions). Patients with pain see themselves as different from other patients of addiction services. Notwithstanding, patients from pain clinics also expressed that they experience stigma related to chronic noncancer pain and opioid use. Acknowledge the uniqueness of each pain experience. Tailoring of content may also be required. Distinguishing concepts, such as addiction versus physical and psychological dependence, may also be important. Consideration of recovery-focused language may assist in exploring impacts of diagnostic labels on identity (internal or social conflicts; sick role).

Strengths and Limitations

Several strategies were used to increase the generalizability of the findings to the complex pain population. A broad sampling strategy was included to capture diversity, including age, culture, language, sex, and history of opioid therapy. Despite these attempts, culturally and linguistically diverse patients were not well-represented. Ensuring cultural diversity is incorporated as much as possible in the subsequent steps of co-design and co-development of BIs could increase customizability of the solution to fit a wider range of users. Our strategy also considered the sample size required for data saturation prospectively during data collection using a new assessment method [52]. The validity of the themes and subthemes was further enhanced through patient member checking. Future research could explore whether these core themes generalize to individuals living in a community (not actively seeking specialist treatment) in a replication study. Patients from specialist addiction and pain clinics were recruited using different strategies. This was necessary to integrate recruitment into the existing models of care, but it may have affected initial patient engagement. Patient reports may have also been influenced by their position in the treatment pathway when they were engaged in this study (eg, before multidisciplinary pain clinic care vs during the initial visit vs review visit). Our study was conducted in Australia, where patients have access to a relatively high standard of public health care and access to heavily subsidized medications. Investigations in countries with different health care systems are needed. There remains a power differential between patients, health care providers, and researchers. It is unknown how this may have influenced our findings. Future work could acknowledge and explore relationship power imbalances in the initial design steps and jointly develop guiding ethical principles. Inviting patients to be members of the research team will be important.

Conclusions

BIs have a long history in chronic relapsing conditions [24,25]. It remains to be discovered whether this treatment can be successfully applied to patients with CNCP who are at risk of opioid-related harm. However, we have made progress in effectively engaging *patient partners* in the co-design of a digital BI, the first necessary step. Exploring lived experiences as part of contextual inquiry could be considered a predesign step to co-design [78]. Our findings add to the growing literature on pre- or co-design in DHIs [79] and extend research into the CNCP field. Co-design has advantages in designing fit-for-purpose solutions that have the potential to enhance treatment effectiveness, adoption, and reach for individuals with CNCP at risk of prescription opioid–related harm.

XSL•FO
RenderX

This important first step of partnering with patients who have CNCP to understand their lived experience is integral to the co-design of new and innovative digital psychological treatments. Without this, many of the challenges identified by patients (eg, lack of trust and accessibility of treatment options) would compromise later co-design phases. The next phase in co-designing personalized digital BIs involves the generation of early design concepts and prototyping, by bringing together scientific, patient, clinical, and technical expertise to meet our current and foreseeable complex challenge of patient-centered, safe opioid therapy in CNCP care.

Acknowledgments

The authors would like to sincerely thank the patients for participating in this project and sharing their experiences. They would also like to thank the specialist pain and addiction services for collaborating on this project. The authors extend special thanks to Dr Andrew Claus, Dr Geraldine Chew, and Dr Tabibul Khan for their assistance in patient engagement. They would also like to thank Ms Chelsea Geary for her contribution to the development of the interview questions and internet-based surveys. The Metro South Research Support Scheme Early Career Researcher Grant (CIA RE; RSS_2020_035) provided funding for this project. The RECOVER Injury Research Centre, University of Queensland; Addiction and Mental Health Services, Metro South Health; and the Tess Cramond Multidisciplinary Pain and Research Centre, Metro North Health provided in-kind support. RE is supported by a Research Stimulus Fellowship grant administered by the University of Queensland.

Data Availability

The datasets generated during and analyzed during this study are not publicly available due to limitations of ethics approval but are available from the corresponding author on reasonable request.

Authors' Contributions

RE, JPC, MS, KB, FF, and SP contributed to the conception and design of the project. RE, KB, and SP collected and analyzed the data. RE drafted the manuscript, and all authors reviewed and provided edits and comments on the revised versions. All authors were involved in the interpretation of the results and have approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient semi-structured interview questions.
[DOCX File , 20 KB - formative_v9i1e57208_app1.docx]

Multimedia Appendix 2
Sample characteristics of patients.
[DOCX File , 21 KB - formative_v9i1e57208_app2.docx]

References

1. Chou R, Gordon DB, de Leon-Casasola OA, Rosenberg JM, Bickler S, Brennan T, et al. Management of postoperative pain: a clinical practice guideline from the American Pain Society, the American Society of Regional Anesthesia and Pain Medicine, and the American Society of Anesthesiologists' Committee on Regional Anesthesia, Executive Committee, and Administrative Council. J Pain 2016 Feb;17(2):131-157 [FREE Full text] [doi: 10.1016/j.jpain.2015.12.008] [Medline: 26827847]
2. Chou R, Turner JA, Devine EB, Hansen RN, Sullivan SD, Blazina I, et al. The effectiveness and risks of long-term opioid therapy for chronic pain: a systematic review for a National Institutes of Health Pathways to Prevention Workshop. Ann Intern Med 2015 Feb 17;162(4):276-286 [FREE Full text] [doi: 10.7326/M14-2559] [Medline: 25581257]
3. Busse JW, Craigie S, Juurlink DN, Buckley DN, Wang L, Couban RJ, et al. Guideline for opioid therapy and chronic noncancer pain. CMAJ 2017 May 08;189(18):E659-E666 [FREE Full text] [doi: 10.1503/cmaj.170363] [Medline: 28483845]
4. Busse JW, Wang L, Kamaleldin M, Craigie S, Riva JJ, Montoya L, et al. Opioids for chronic noncancer pain: a systematic review and meta-analysis. JAMA 2018 Dec 18;320(23):2448-2460 [FREE Full text] [doi: 10.1001/jama.2018.18472] [Medline: 30561481]
5. Els C, Jackson TD, Kunyk D, Lappi VG, Sonnenberg B, Hagtvedt R, et al. Adverse events associated with medium- and long-term use of opioids for chronic non-cancer pain: an overview of Cochrane Reviews. Cochrane Database Syst Rev 2017 Oct 30;10(10):CD012509 [FREE Full text] [doi: 10.1002/14651858.CD012509.pub2] [Medline: 29084357]

XSL•FO
RenderX

6. Scholl L, Seth P, Kariisa M, Wilson N, Baldwin G. Drug and opioid-involved overdose deaths — United States, 2013–2017. MMWR Morb Mortal Wkly Rep 2018 Dec 21;67(5152):1419-1427. [doi: 10.15585/mmwr.mm6751521e1] [Medline: 30605448]
7. Roxburgh A, Hall WD, Dobbins T, Gisev N, Burns L, Pearson S, et al. Trends in heroin and pharmaceutical opioid overdose deaths in Australia. Drug Alcohol Depend 2017 Oct 01;179:291-298. [doi: 10.1016/j.drugalcdep.2017.07.018] [Medline: 28826104]
8. Dowell D, Haegerich TM, Chou R. CDC guideline for prescribing opioids for chronic pain--United States, 2016. JAMA 2016 Apr 19;315(15):1624-1645 [FREE Full text] [doi: 10.1001/jama.2016.1464] [Medline: 26977696]
9. Response to the opioid crisis. United Nations Office on Drugs and Crime. URL: https://www.unodc.org/unodc/en/opioid-crisis/index.html [accessed 2023-10-08]
10. Morin KA, Eibl JK, Franklyn AM, Marsh DC. The opioid crisis: past, present and future policy climate in Ontario, Canada. Subst Abuse Treat Prev Policy 2017 Nov 02;12(1):45 [FREE Full text] [doi: 10.1186/s13011-017-0130-5] [Medline: 29096653]
11. HHS acting secretary declares public health emergency to address national opioid crisis. U.S. Department of Health & Human Services. 2017. URL: https://www.proquest.com/docview/1955885383?sourcetype=Wire%20Feeds [accessed 2023-08-12]
12. Eccleston C, Morley SJ, Williams AC. Psychological approaches to chronic pain management: evidence and challenges. Br J Anaesth 2013 Jul;111(1):59-63 [FREE Full text] [doi: 10.1093/bja/aet207] [Medline: 23794646]
13. Darnall BD, Carr DB, Schatman ME. Pain psychology and the biopsychosocial model of pain treatment: ethical imperatives and social responsibility. Pain Med 2017 Aug 01;18(8):1413-1415 [FREE Full text] [doi: 10.1093/pm/pnw166] [Medline: 27425187]
14. Hogg MN, Gibson S, Helou A, DeGabriele J, Farrell MJ. Waiting in pain: a systematic investigation into the provision of persistent pain services in Australia. Med J Aust 2012 Apr 02;196(6):386-390. [doi: 10.5694/mja12.10140] [Medline: 22471539]
15. Crouch TB, Wedin S, Kilpatrick RL, Christon L, Balliet W, Borckardt J, et al. Pain rehabilitation's dual power: treatment for chronic pain and prevention of opioid-related risks. Am Psychol 2020 Sep;75(6):825-839 [FREE Full text] [doi: 10.1037/amp0000663] [Medline: 32915026]
16. Nicholas MK, Asghari A, Sharpe L, Beeston L, Brooker C, Glare P, et al. Reducing the use of opioids by patients with chronic pain: an effectiveness study with long-term follow-up. Pain 2020 Mar;161(3):509-519. [doi: 10.1097/j.pain.0000000000001763] [Medline: 31764391]
17. Larance B, Campbell G, Moore T, Nielsen S, Bruno R, Lintzeris N, et al. Concerns and help-seeking among patients using opioids for management of chronic noncancer pain. Pain Med 2019 Apr 01;20(4):758-769. [doi: 10.1093/pm/pny078] [Medline: 29762767]
18. Frank JW, Levy C, Matlock DD, Calcaterra SL, Mueller SR, Koester S, et al. Patients' perspectives on tapering of chronic opioid therapy: a qualitative study. Pain Med 2016 Oct 20;17(10):1838-1847 [FREE Full text] [doi: 10.1093/pm/pnw078] [Medline: 27207301]
19. Goesling J, DeJonckheere M, Pierce J, Williams DA, Brummett CM, Hassett AL, et al. Opioid cessation and chronic pain: perspectives of former opioid users. Pain 2019 May 16;160(5):1131-1145 [FREE Full text] [doi: 10.1097/j.pain.0000000000001493] [Medline: 30889052]
20. Eccleston C, Fisher E, Thomas KH, Hearn L, Derry S, Stannard C, et al. Interventions for the reduction of prescribed opioid use in chronic non-cancer pain. Cochrane Database Syst Rev 2017 Nov 13;11(11):CD010323. [doi: 10.1002/14651858.CD010323.pub3] [Medline: 29130474]
21. Garland EL, Brintz CE, Hanley AW, Roseen EJ, Atchley RM, Gaylord SA, et al. Mind-body therapies for opioid-treated pain: a systematic review and meta-analysis. JAMA Intern Med 2020 Jan 01;180(1):91-105 [FREE Full text] [doi: 10.1001/jamainternmed.2019.4917] [Medline: 31682676]
22. Darnall BD. Psychological treatment for chronic pain: improving access and integration. Psychol Sci Public Interest 2021 Sep;22(2):45-51 [FREE Full text] [doi: 10.1177/15291006211033612] [Medline: 34541966]
23. Elphinston RA, Sterling M, Leung J, Gray PD, Scotti Requena S, Connor JP. Brief psychological interventions for reducing prescription opioid use, related harm, and pain intensity in patients with chronic pain: a systematic review and assessment of patient interest. Clin J Pain 2021 Apr 01;37(4):270-280. [doi: 10.1097/AJP.0000000000000908] [Medline: 33323791]
24. Brief intervention for hazardous and harmful drinking (AUDIT): a manual for use in primary care. World Health Organization. 2001 Aug 5. URL: https://tinyurl.com/mwnftddd [accessed 2023-04-05]
25. Kaner EF, Beyer FR, Muirhead C, Campbell F, Pienaar ED, Bertholet N, et al. Effectiveness of brief alcohol interventions in primary care populations. Cochrane Database Syst Rev 2018 Feb 24;2(2):CD004148 [FREE Full text] [doi: 10.1002/14651858.CD004148.pub4] [Medline: 29476653]
26. Neighbors CJ, Barnett NP, Rohsenow DJ, Colby SM, Monti PM. Cost-effectiveness of a motivational intervention for alcohol-involved youth in a hospital emergency department. J Stud Alcohol Drugs 2010 May;71(3):384-394 [FREE Full text] [doi: 10.15288/jsad.2010.71.384] [Medline: 20409432]

XSL•FO
RenderX

27. Schmidt CS, Schulte B, Seo HN, Kuhn S, O'Donnell A, Kriston L, et al. Meta-analysis on the effectiveness of alcohol screening with brief interventions for patients in emergency care settings. Addiction 2016 May;111(5):783-794. [doi: 10.1111/add.13263] [Medline: 26637990]
28. Poston JM, Hanson WE. Meta-analysis of psychological assessment as a therapeutic intervention. Psychol Assess 2010 Jun;22(2):203-212. [doi: 10.1037/a0018679] [Medline: 20528048]
29. Kaner EF, Beyer FR, Garnett C, Crane D, Brown J, Muirhead C, et al. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. Cochrane Database Syst Rev 2017 Sep 25;9(9):CD011479 [FREE Full text] [doi: 10.1002/14651858.CD011479.pub2] [Medline: 28944453]
30. Darnall BD, Ziadni MS, Krishnamurthy P, Flood P, Heathcote LC, Mackey IG, et al. "My surgical success": effect of a digital behavioral pain medicine intervention on time to opioid cessation after breast cancer surgery-a pilot randomized controlled clinical trial. Pain Med 2019 Nov 01;20(11):2228-2237 [FREE Full text] [doi: 10.1093/pm/pnz094] [Medline: 31087093]
31. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, et al. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. BMJ 2015 Nov 11;351:h5627 [FREE Full text] [doi: 10.1136/bmj.h5627] [Medline: 26559241]
32. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research . . . and three things we can do about them. Psychiatr Serv 2017 May 01;68(5):427-429 [FREE Full text] [doi: 10.1176/appi.ps.201600541] [Medline: 28412890]
33. Ioannidis JP. Why most clinical research is not useful. PLoS Med 2016 Jun;13(6):e1002049 [FREE Full text] [doi: 10.1371/journal.pmed.1002049] [Medline: 27328301]
34. Svendsen MJ, Wood KW, Kyle J, Cooper K, Rasmussen CD, Sandal LF, et al. Barriers and facilitators to patient uptake and utilisation of digital interventions for the self-management of low back pain: a systematic review of qualitative studies. BMJ Open 2020 Dec 12;10(12):e038800 [FREE Full text] [doi: 10.1136/bmjopen-2020-038800] [Medline: 33310794]
35. Fernandes LG, Devan H, Fioratti I, Kamper SJ, Williams CM, Saragiotto BT. At my own pace, space, and place: a systematic review of qualitative studies of enablers and barriers to telehealth interventions for people with chronic pain. Pain 2022 Feb 01;163(2):e165-e181. [doi: 10.1097/j.pain.0000000000002364] [Medline: 34433776]
36. Schlieter H, Marsch LA, Whitehouse D, Otto L, Londral AR, Teepe GW, et al. Scale-up of digital innovations in health care: expert commentary on enablers and barriers. J Med Internet Res 2022 Mar 11;24(3):e24582 [FREE Full text] [doi: 10.2196/24582] [Medline: 35275065]
37. Slattery P, Saeri AK, Bragge P. Research co-design in health: a rapid overview of reviews. Health Res Policy Syst 2020 Feb 11;18(1):17 [FREE Full text] [doi: 10.1186/s12961-020-0528-9] [Medline: 32046728]
38. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
39. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
40. Keogh E, Rosser BA, Eccleston C. e-Health and chronic pain management: current status and developments. Pain 2010 Oct;151(1):18-21. [doi: 10.1016/j.pain.2010.07.014] [Medline: 20674174]
41. Tabi K, Randhawa AS, Choi F, Mithani Z, Albers F, Schnieder M, et al. Mobile apps for medication management: review and analysis. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e13608 [FREE Full text] [doi: 10.2196/13608] [Medline: 31512580]
42. Lalloo C, Jibb LA, Rivera J, Agarwal A, Stinson JN. "There's a pain app for that": review of patient-targeted smartphone applications for pain management. Clin J Pain 2015 Jun;31(6):557-563. [doi: 10.1097/AJP.0000000000000171] [Medline: 25370138]
43. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the mobile app rating scale. JMIR Mhealth Uhealth 2018 Oct 25;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]
44. Zhao P, Yoo I, Lancey R, Varghese E. Mobile applications for pain management: an app analysis for clinical usage. BMC Med Inform Decis Mak 2019 May 30;19(1):106 [FREE Full text] [doi: 10.1186/s12911-019-0827-7] [Medline: 31146739]
45. Devan H, Farmery D, Peebles L, Grainger R. Evaluation of self-management support functions in apps for people with persistent pain: systematic review. JMIR Mhealth Uhealth 2019 Feb 12;7(2):e13080 [FREE Full text] [doi: 10.2196/13080] [Medline: 30747715]
46. Noorbergen TJ, Adam MT, Teubner T, Collins CE. Using co-design in mobile health system development: a qualitative study with experts in co-design and mobile health system development. JMIR Mhealth Uhealth 2021 Nov 10;9(11):e27896 [FREE Full text] [doi: 10.2196/27896] [Medline: 34757323]
47. Sanders EB, Stappers PJ. Probes, toolkits and prototypes: three approaches to making in codesigning. CoDesign 2014 Mar 06;10(1):5-14. [doi: 10.1080/15710882.2014.888183]
48. Walker J, Holloway I, Sofaer B. In the system: the lived experience of chronic back pain from the perspectives of those seeking help from pain clinics. Pain 1999 Apr;80(3):621-628. [doi: 10.1016/S0304-3959(98)00254-1] [Medline: 10342423]

XSL•FO
RenderX

49. Liddle SD, Baxter GD, Gracey JH. Chronic low back pain: patients' experiences, opinions and expectations for clinical management. Disabil Rehabil 2007 Dec 30;29(24):1899-1909. [doi: 10.1080/09638280701189895] [Medline: 17852259]
50. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SL, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002 Aug;32(6):959-976. [doi: 10.1017/s0033291702006074] [Medline: 12214795]
51. Bate P, Robert G. Experience-based design: from redesigning the system around the patient to co-designing services with the patient. Qual Saf Health Care 2006 Oct;15(5):307-310 [FREE Full text] [doi: 10.1136/qshc.2005.016527] [Medline: 17074863]
52. Guest G, Namey E, Chen M. A simple method to assess and report thematic saturation in qualitative research. PLoS One 2020 May 5;15(5):e0232076 [FREE Full text] [doi: 10.1371/journal.pone.0232076] [Medline: 32369511]
53. Kumar SP. Utilization of brief pain inventory as an assessment tool for pain in patients with cancer: a focused review. Indian J Palliat Care 2011 May;17(2):108-115 [FREE Full text] [doi: 10.4103/0973-1075.84531] [Medline: 21976850]
54. Tan G, Jensen MP, Thornby JI, Shanti BF. Validation of the Brief Pain Inventory for chronic nonmalignant pain. J Pain 2004 Mar;5(2):133-137 [FREE Full text] [doi: 10.1016/j.jpain.2003.12.005] [Medline: 15042521]
55. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343 [FREE Full text] [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
56. Taylor R, Lovibond PF, Nicholas MK, Cayley C, Wilson PH. The utility of somatic items in the assessment of depression in patients with chronic pain: a comparison of the Zung Self-Rating Depression Scale and the Depression Anxiety Stress Scales in chronic pain and clinical and community samples. Clin J Pain 2005;21(1):91-100. [doi: 10.1097/00002508-200501000-00011] [Medline: 15599136]
57. Butler SF, Budman SH, Fernandez KC, Houle B, Benoit C, Katz N, et al. Development and validation of the current opioid misuse measure. Pain 2007 Jul;130(1-2):144-156 [FREE Full text] [doi: 10.1016/j.pain.2007.01.014] [Medline: 17493754]
58. Butler SF, Budman SH, Fanciullo GJ, Jamison RN. Cross validation of the current opioid misuse measure to monitor chronic pain patients on opioid therapy. Clin J Pain 2010;26(9):770-776 [FREE Full text] [doi: 10.1097/AJP.0b013e3181f195ba] [Medline: 20842012]
59. Meltzer EC, Rybin D, Saitz R, Samet JH, Schwartz SL, Butler SF, et al. Identifying prescription opioid use disorder in primary care: diagnostic characteristics of the Current Opioid Misuse Measure (COMM). Pain 2011 Feb;152(2):397-402 [FREE Full text] [doi: 10.1016/j.pain.2010.11.006] [Medline: 21177035]
60. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
61. Lincoln YS, Guba EG. Naturalistic Inquiry. Newbury Park, CA: SAGE Publications; 1985.
62. Charmaz K. Grounded theory. In: Smith JA, editor. Qualitative Psychology: A Practical Guide to Research Methods. Thousand Oaks, CA: SAGE Publications, Inc; 2003:81-110.
63. Prescribing drugs of dependence in general practice: part C2 - the role of opioids in pain management. The Royal Australian College of General Practitioners. URL: https://tinyurl.com/zs6ayr9y [accessed 2023-03-11]
64. Punwasi R, de Kleijn L, Rijkels-Otters JB, Veen M, Chiarotto A, Koes B. General practitioners' attitudes towards opioids for non-cancer pain: a qualitative systematic review. BMJ Open 2022 Feb 01;12(2):e054945 [FREE Full text] [doi: 10.1136/bmjopen-2021-054945] [Medline: 35105588]
65. Darnall BD, Roy A, Chen AL, Ziadni MS, Keane RT, You DS, et al. Comparison of a single-session pain management skills intervention with a single-session health education intervention and 8 sessions of cognitive behavioral therapy in adults with chronic low back pain: a randomized clinical trial. JAMA Netw Open 2021 Aug 02;4(8):e2113401 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.13401] [Medline: 34398206]
66. Gatchel RJ, Peng YB, Peters ML, Fuchs PN, Turk DC. The biopsychosocial approach to chronic pain: scientific advances and future directions. Psychol Bull 2007 Jul;133(4):581-624. [doi: 10.1037/0033-2909.133.4.581] [Medline: 17592957]
67. Siddall B, Ram A, Jones MD, Booth J, Perriman D, Summers SJ. Short-term impact of combining pain neuroscience education with exercise for chronic musculoskeletal pain: a systematic review and meta-analysis. Pain 2022 Jan 01;163(1):e20-e30. [doi: 10.1097/j.pain.0000000000002308] [Medline: 33863860]
68. Thompson K, Johnson MI, Milligan J, Briggs M. Twenty-five years of pain education research-what have we learned? Findings from a comprehensive scoping review of research into pre-registration pain education for health professionals. Pain 2018 Nov 26;159(11):2146-2158. [doi: 10.1097/j.pain.0000000000001352] [Medline: 30052574]
69. Darnall BD, Scheman J, Davin S, Burns JW, Murphy JL, Wilson AC, et al. Pain psychology: a global needs assessment and national call to action. Pain Med 2016 Feb;17(2):250-263 [FREE Full text] [doi: 10.1093/pm/pnv095] [Medline: 26803844]
70. Edmond SN, Heapy AA, Kerns RD. Engaging mental health professionals in addressing pain. JAMA Psychiatry 2019 Jun 01;76(6):565-566. [doi: 10.1001/jamapsychiatry.2019.0254] [Medline: 30942829]
71. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]

XSL•FO
RenderX

72. Qaseem A, Wilt TJ, McLean RM, Forciea MA, Clinical Guidelines Committee of the American College of Physicians, Denberg TD, et al. Noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2017 Apr 04;166(7):514-530 [FREE Full text] [doi: 10.7326/M16-2367] [Medline: 28192789]
73. Suman A, Armijo-Olivo S, Deshpande S, Marietta-Vasquez J, Dennett L, Miciak M, et al. A systematic review of the effectiveness of mass media campaigns for the management of low back pain. Disabil Rehabil 2021 Dec;43(24):3523-3551. [doi: 10.1080/09638288.2020.1743777] [Medline: 32242464]
74. Psychological expertise where we need it most: GPs not the only answer to our mental health crisis. Australian Psychological Society. 2022 Mar 1. URL: https://tinyurl.com/bde43nvv [accessed 2022-11-01]
75. Matthias MS, Johnson NL, Shields CG, Bair MJ, MacKie P, Huffman M, et al. "I'm not gonna pull the rug out from under you": patient-provider communication about opioid tapering. J Pain 2017 Nov;18(11):1365-1373 [FREE Full text] [doi: 10.1016/j.jpain.2017.06.008] [Medline: 28690000]
76. Rollnick S, Miller WR. What is motivational interviewing? Behav Cogn Psychother 2009 Jun 16;23(4):325-334. [doi: 10.1017/S135246580001643X]
77. Rubak S, Sandbaek A, Lauritzen T, Christensen B. Motivational interviewing: a systematic review and meta-analysis. Br J Gen Pract 2005 Apr;55(513):305-312 [FREE Full text] [Medline: 15826439]
78. Bird M, McGillion M, Chambers EM, Dix J, Fajardo CJ, Gilmour M, et al. A generative co-design framework for healthcare innovation: development and application of an end-user engagement framework. Res Involv Engagem 2021 Mar 01;7(1):12 [FREE Full text] [doi: 10.1186/s40900-021-00252-7] [Medline: 33648588]
79. Vandekerckhove P, de Mul M, Bramer WM, de Bont AA. Generative participatory design methodology to develop electronic health interventions: systematic literature review. J Med Internet Res 2020 Apr 27;22(4):e13780 [FREE Full text] [doi: 10.2196/13780] [Medline: 32338617]

Abbreviations

BI: brief intervention
CNCP: chronic noncancer pain
COMM: Current Opioid Misuse Measure
DHI: digital health intervention
GP: general practitioner

Edited by A Mavragani; submitted 23.02.24; peer-reviewed by M Keller; comments to author 25.06.24; revised version received 13.08.24; accepted 24.09.24; published 30.01.25.

Please cite as:
Elphinston RA, Pager S, Brown K, Sterling M, Fatehi F, Gray P, Hipper L, Cahill L, Connor JP
Co-Designing a Digital Brief Intervention to Reduce the Risk of Prescription Opioid–Related Harm Among People With Chronic Noncancer Pain: Qualitative Analysis of Patient Lived Experiences
JMIR Form Res 2025;9:e57208
URL: https://formative.jmir.org/2025/1/e57208
doi:10.2196/57208
PMID:

©Rachel A Elphinston, Sue Pager, Kelly Brown, Michele Sterling, Farhad Fatehi, Paul Gray, Linda Hipper, Lauren Cahill, Jason P. Connor. Originally published in JMIR Formative Research (https://formative.jmir.org), 30.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Human-Centered Design of an mHealth Tool for Optimizing HIV Index Testing in Wartime Ukraine: Formative Research Case Study

Nancy Puttkammer[1], MPH, PhD; Elizabeth Dunbar[1,2], PhD; Myroslava Germanovych[3], MA; Mariia Rosol[3], MA; Matthew Golden[4,5], MPH, MD; Anna Hubashova[3], BA; Vladyslav Fedorchenko[3], MD; Larisa Hetman[3], MPA, MD; Liudmyla Legkostup[3], MS; Jan Flowers[1,6], BS, MS; Olena Nesterova[3], MD

[1]Digital Initiatives Group at I-TECH, Department of Global Health, University of Washington, Seattle, WA, United States

[2]Department of Human Centered Design and Engineering, College of Engineering, University of Washington, Seattle, WA, United States

[3]The State Institution «Public Health Center of the Ministry of Health of Ukraine», Kyiv, Ukraine

[4]Department of Medicine, Division of Allergy and Infectious Diseases, University of Washington, Seattle, WA, United States

[5]STD Control Program, Public Health Seattle King County, Seattle, WA, United States

[6]Biobehavioral Nursing and Health Informatics Department, School of Nursing, University of Washington, Seattle, WA, United States

Corresponding Author:
Nancy Puttkammer, MPH, PhD
Digital Initiatives Group at I-TECH
Department of Global Health
University of Washington
325 Ninth Ave, Box # 359932
Seattle, WA, 98104
United States
Phone: 1 206 898 5892
Email: nputt@uw.edu

Abstract

Background: Assisted partner services (APSs; sometimes called index testing) are now being brought to scale as a high-yield HIV testing strategy in many nations. However, the success of APSs is often hampered by low levels of partner elicitation. The Computer-Assisted Self-Interview (CASI)–Plus study sought to develop and test a mobile health (mHealth) tool to increase the elicitation of sexual and needle-sharing partners among persons with newly diagnosed HIV. CASI-Plus provides client-facing information on APS methods and uses a standardized, self-guided questionnaire with nonjudgmental language for clients to list partners who would benefit from HIV testing. The tool also enables health care workers (HCWs) to see summarized data to facilitate partner tracking.

Objective: The formative research phase of the CASI-Plus study aimed to gather client and HCW input on the design of the CASI-Plus tool to ensure its acceptability, feasibility, and usability.

Methods: This study gathered input to prioritize features and tested the usability of CASI-Plus with HCWs and clients receiving HIV services in public health clinics in wartime Ukraine. The CASI-Plus study's formative phase, carried out from May 2023 to July 2024, adapted human-centered design (HCD) methods grounded in principles of empathy, iteration, and creative ideation. The study involved 3 steps: *formative HCD*, including in-depth individual interviews with clients, such as men who have sex with men and people who inject drugs, and internet-based design workshops with HCWs from rural and urban HIV clinics in Chernihiv and Dnipro; *software platform assessment and heuristic evaluation*, including assessment of open-source mHealth platforms against CASI-Plus requirements, prototype development, and testing of the REDCap (Research Electronic Data Capture) prototype based on usability heuristics; and *usability walk-throughs*, including simulated cases with HCWs and clients.

Results: The formative phase of the CASI-Plus study included in-depth individual interviews with 10 clients and 3 workshops with 22 HCWs. This study demonstrated how simplified HCD methods, adapted to the wartime context, gathered rich input on prioritized features and tool design. The CASI-Plus design reflected features that are both culturally sensitive and in alignment with the constraints of Ukraine's wartime setting. Prioritized features included information about the benefits of HIV index testing; a nonjudgmental, self-guided questionnaire to report partners; client stories; and bright images to accompany the text. Two-way

XSL•FO
RenderX

SMS text messaging between clients and HCWs was deemed impractical based on risks of privacy breaches, national patient privacy regulations, and HCW workload.

Conclusions: It was feasible to conduct HCD research in Ukraine in a wartime setting. The CASI-Plus mHealth tool was acceptable to both HCWs and clients. The next step for this research is a randomized clinical trial of the effect of the REDCap-based CASI-Plus tool on the number of partners named and the rate of partners completing HIV testing.

(JMIR Form Res 2025;9:e66132) doi:10.2196/66132

KEYWORDS

human-centered design; mobile health; mHealth; Ukraine; HIV testing; war and humanitarian settings

Introduction

Background

In December 2016, the World Health Organization recommended scale-up of HIV assisted partner services (APSs) as a strategy to increase the identification and testing of partners of persons with HIV [1,2]. APSs offer persons with HIV assistance to confidentially notify their sexual and needle-sharing partners of their exposure and link the partners to testing and treatment. Randomized controlled trials have shown that APSs increase HIV testing and case finding [3-6] and are cost-effective [7,8]. APSs are scalable in routine practice, but routine APS programs have typically demonstrated lower HIV case finding than was observed in randomized controlled trials [9-12]. At least in part, the low case finding observed with APSs in most settings reflects the small number of sexual and needle-sharing partners named by index cases. As of 2020, persons with HIV receiving APSs in Ukraine named an average of only 1.14 partners despite training of APS program personnel that emphasized the need to elicit more than 1 partner for each index client [13].

Computer-Assisted Self-Interview (CASI)–Plus seeks to improve partner elicitation and testing through a mobile health (mHealth) client engagement tool using a CASI. CASI-Plus supports the initial APS encounter by providing information on how APSs work, eliciting names of all sexual or injection partners, screening for risk of intimate partner violence, and planning for partner notification. It also facilitates case management at the point of partner testing and via repeat follow-up surveys to assess for barriers to notification, interest in provider assistance, and occurrence of intimate partner violence. In surveys of sexual behavior across diverse settings, the computer-assisted self-interview approach reduced social desirability bias compared with in-person interviews [14-17]. However, the computer-assisted self-interview approach has not yet been used in APS programs in low- or middle-income countries. The CASI-Plus study sought to assess whether the tool improves partner elicitation and HIV testing in a routine APS program operating at scale, with the goal of increasing the number of persons with HIV who know their status and are linked to either HIV pre-exposure prophylaxis or HIV antiretroviral therapy services. The Russian Federation's full-scale invasion and ongoing war in Ukraine has underscored the need for simple and scalable interventions to support the control of the HIV epidemic in a disrupted health care system. CASI-Plus offers the potential to improve APS partner elicitation and HIV case finding in wartime Ukraine, as well as in many other settings where routine APS programs face challenges with high-fidelity APS implementation and low contact (partner) elicitation.

Objectives

For the CASI-Plus study, we first carried out formative research with health care workers (HCWs) and clients from a human-centered design (HCD) perspective. As applied to mHealth tools, HCD is a pragmatic research orientation that seeks to overcome barriers to usability and integration with workflows that arise when tool designers ignore the sociotechnical complexity of health care systems [18,19]. HCD principles and methods emphasize engaging with mHealth users, starting with problem definition before solution design; supporting divergence and convergence of ideas through the design process; using participatory co-design techniques, prototyping, and iteration; and applying values of creativity, empathy, tolerance for ambiguity, and willingness to fail [18,20-22]. There are few research studies or frameworks discussing the application of HCD perspectives in mHealth projects in wartime and humanitarian contexts. The participatory engagement and time required to carry out HCD-oriented research may be challenging in wartime contexts [23], and more scholarship is needed on realistic approaches to HCD in such contexts. During the CASI-Plus formative research phase, we adapted HCD methods to Ukraine's wartime context. The aims of our formative research were to (1) describe how HCW and client perspectives and the wartime context of Ukraine shaped the design of the CASI-Plus mHealth tool and (2) identify HCD methods and principles for effective formative mHealth research in a wartime setting.

Methods

Study Design

The formative phase of the CASI-Plus study (ClinicalTrials.gov NCT05826977) supported the design and software adaptation of an mHealth tool for use in Ukraine's APS program. During the first step of the formative research, we conducted internet-based workshops with HCWs from 3 health facilities, as well as in-depth interviews with clients enrolled in HIV care at 1 health facility. During the second step, we developed the CASI-Plus tool, rated its usability based on common usability heuristics or "rules of thumb," and made improvements to the prototype. During the third step, we obtained feedback on the CASI-Plus tool prototype through simulated use of the tool by health workers and clients to further adapt our design. Table 1

XSL•FO
RenderX

presents an overview of the activities across the 3 steps of our formative research.

Table 1. Summary of the Computer-Assisted Self-Interview (CASI)–Plus formative research design—participants, methods, and activities for research steps 1 to 3 in wartime Ukraine (May 2023-July 2024).

	Step 1: formative design workshops and interviews—elicit and prioritize CASI-Plus features		Step 2: heuristic evaluation—critique platforms and prototype	Step 3: simulated walk-throughs—end user prototype testing	
	Health care workers	HIV care clients	Study team	Health care workers	HIV care clients
Method	Design workshops with focus group	In-depth individual interviews	Compare tools and review CASI-Plus usability	Review of prototype and focus group	Review of prototype and individual interviews
Activities	• Preworkshop exercise • Critique ideas for CASI-Plus features • Feature prioritization game • Group synthesis and feedback	• Discuss current APS[a] barriers • Critique CASI-Plus mock-up • Review client script	• Compare 5 open-source platforms with CASI-Plus requirements from step 1 • Critique CASI-Plus REDCap[b] prototype using usability best practices	• Add 5 fictional clients to prototype and test features • Discuss feedback with the group	• 5 current HIV care clients walked through the client portion of CASI-Plus • Gather feedback

[a]APS: assisted partner service.

[b]REDCap: Research Electronic Data Capture (Vanderbilt University).

Ukrainian Setting and War

Ukraine is an Eastern European nation that regained its independence from the Soviet Union in 1991. On February 24, 2022, the Russian Federation invaded Ukraine, inflicting the largest military assault in Europe since World War II. Of a prewar population of 41.6 million, an estimated 7.5 million (18%) fled the country as refugees, 5.9 million (14.2%) were internally displaced, 1.6 to 2.8 million were forcefully deported to Russia, and >10,000 civilians were killed [24-26]. As of mid-2023, when this study was conducted, the population was estimated to be 41 million, with 12.4% being internally displaced persons and 11.6% being refugees who had returned to their homes [27]. War crimes have included direct attacks on health care facilities, resulting in damage to >1000 facilities, complete destruction of >200 facilities, and deaths of >100 HCWs (with approximately one-third killed while working) [28-31]. Our study sites included health facilities in the cities of Chernihiv and Dnipro. Chernihiv is in northern Ukraine and was occupied by Russian forces from March 2022 to April 2022, leading to widespread destruction and trauma [32]. Dnipro is a leading industrial city located in southeastern Ukraine that has served as a hub for wartime logistics and humanitarian aid for the eastern and southern regions most affected by war and occupation.

HIV in Ukraine

Ukraine has a significant HIV burden, with an estimated 245,000 persons with HIV living in the country in early 2022, of whom roughly 79% knew their HIV status [28]. In 2019, the Ukraine Ministry of Health began scaling APSs, reaching 7533 index clients and 5038 newly tested partners in 2020 [13]. Clients who are newly diagnosed with HIV, as well as clients who were previously diagnosed but who experienced interruptions in HIV treatment or other conditions leading to having an unsuppressed HIV viral load, are invited to participate in APSs on a voluntary basis. Index clients are asked to name all recent sexual and needle-sharing partners as well as biological children to indicate whether these partners have recently been tested for HIV and to select a notification method for each partner. Health workers then follow up over time with the index clients to ensure that each partner is notified and provided information about how to get tested for HIV. Persons who inject drugs and men who have sex with men have high HIV prevalence (20.9% and 3.9%, respectively) and are estimated to account for 29.9% and 2.9% of all persons with HIV in Ukraine [33-36]. However, these key populations appear underrepresented in APSs as persons who inject drugs made up only 14.9% and men who have sex with men made up only 1.8% of APS index clients in 2020 [13,37,38]. An additional barrier to epidemic control in Ukraine was that approximately one-third of named partners with unknown status (32.4%) in the APS program did not complete HIV testing within 60 days [13].

In the wartime context, HIV programs have been stretched, and conditions have favored increased HIV transmission [39]. HIV testing declined by 36% in 2022 compared to 2021 before rebounding to prewar levels in 2023 [28]. The Ukraine Ministry of Health estimated that reduced coverage of HIV testing services, together with wartime occupation and population displacement, contributed to a 13% increase in HIV incidence in 2023 compared to 2022 [28]. In the wartime context, APSs remain an expected part of the basic package of HIV services. However, their prioritization during the war given other pressing needs has not been thoroughly assessed, including in the context of our study.

HCD Methods for the Design of CASI-Plus

We sought to design CASI-Plus as a tool to increase partner elicitation and testing that would be both tailored to Ukraine's specific wartime context and generalizable to APS programs worldwide. We selected HCD methods given their utility in leveraging end-user expertise and taking a system thinking

XSL•FO
RenderX

approach while designing this health care innovation [19,22,40]. Melles et al [18] discuss the importance of system thinking regarding how an innovation interacts across 3 levels: micro (eg, using tools to perform tasks), meso (eg, health workers as a part of teams that coordinate to deliver care), and macro (eg, health workers and clients as a part of organizations and societies). With HCWs, our inquiry focused on eliciting task-based features (micro level), as well as how to integrate a tool within their health care delivery workflow (meso level) and how to tailor the design for the Ukrainian cultural context and wartime environment (macro level). For index clients, our study focused on considerations to ensure that the tool was accessible (micro), aligned with patient health care experiences (meso), and aligned with client values (macro).

Our original study, conceived before the full-scale invasion of Ukraine in 2022, proposed in-person, participatory HCD workshops and interviews with HCWs and clients led jointly by American and Ukrainian investigators. However, in the wartime context, we needed to reconsider our HCD methods based on disrupted ground transportation, safety of Public Health Center (PHC) of the Ministry of Health of Ukraine and University of Washington study team members to travel to the study sites, understaffing at the study sites, and unpredictable client demand for health care services due to mass population displacement and military mobilization. Table 1 provides an overview of HCD activities across 3 steps, whereas Table 2 provides a more detailed description of the internet-based, tele-conferencing sessions with HCWs and in-person engagements with index clients and how the HCD methods were adapted to the wartime context.

XSL•FO
RenderX

Table 2. Detailed description of human-centered design (HCD) methods used in the Computer-Assisted Self-Interview (CASI)–Plus study, including HCD principles and adaptations to Ukraine's wartime context (May 2023-July 2024).

Step, participants, and activity	Description	HCD principles	Adaptations to wartime context
Step 1—HCWs[a]			
Individual preworkshop exercise	HCWs created an ordered list of the steps of APSs[b] and described care activities, who was involved, and the data collected. HCWs were reminded of their expertise and that there were no right or wrong answers. The study team consolidated individual workflow maps to produce a consensus version for display during the workshop session.	Empowering end users as experts—presession exercises can prime users' engagement and allow them to reflect independently and without the social pressures of the workshop [41,42]; crowdsource responses from multiple participants—this helps consolidate core steps, which is critical to understand context and activities that the tool could support [18,43]	Reduced the duration of the design workshop and allowed participants to complete when they had availability (thereby reducing burden). Allowed the study team to consolidate information for review at the workshop. Included explicit discussion of how war has impacted clients and services.
Critique ideas for CASI-Plus features	Reviewed mock-ups of 12 ideas for CASI-Plus features. Each mock-up contained key discussion question prompts (eg, "what are the advantages of this feature?" and "what are your concerns?"). Invited participants to share additional feature ideas and describe ideal feature designs that would be most useful.	Ideation and creativity [20]; co-design process supported design and validation of a solution based first on diverging perspectives and then on converging perspectives	Conducted the workshop via videoconference (Zoom [Zoom Video Communications]) where participants joined individually to minimize safety risks and the inconvenience of in-person meetings. Web-based tools for ideation and creative work (eg, Figma) were deemed too complex for HCWs to learn and use, so the study team relied on slides and discussion. Visual mock-ups supported reactions and discussions and stimulated creative thinking.
Feature prioritization game	HCWs completed 3 rounds of voting on priority features, with each HCW having 3 votes per round. Between rounds of voting, HCWs discussed why they prioritized specific features. One group received 3 extra votes for the final round when they did not change choices between the first and second round.	The prioritization game helped focus the feature set for health worker top preferences. The prioritization game was adapted to voting on features rather than a money-based game of purchasing options by the study team per recommendations that voting would resonate more with participants.	Used verbal roll call for voting along with marking each vote on the screen to avoid the complexity of web-based voting tools. HCWs were engaged in exercises, discussion, and explaining their choices.
Group synthesis and feedback	Semistructured focus group discussions on how CASI-Plus would impact workflow and what barriers it would and would not address	Consider user needs and experiences [43]; system perspective	HCWs required extra time to familiarize themselves with and practice using the tool after the internet-based orientation.
Step 1—clients			
Discuss current APS barriers	Semistructured questions on barriers to naming partners, notifying partners, using APSs, and partners completing testing. Reflections on what works well and what can be improved in APSs.	Empathy [20]	IDI[c] questions asked about barriers in the wartime context.
Critique CASI-Plus features and review client script	Semistructured questions on CASI-Plus features, including opinions on 2-way texting and reflections on what barriers it would and would not address. Clients reviewed 13 cards with visual mock-ups of the client-facing screens of the CASI-Plus tool. They shared reactions, including things that were confusing or inappropriate. They provided alternative suggestions for visual design.	Co-design; creativity	A single round of design feedback was conducted to minimize burden on client and researcher time in wartime context.
Step 2—study team			

XSL•FO
RenderX

Step, participants, and activity	Description	HCD principles	Adaptations to wartime context
Heuristic evaluation to critique platforms and prototype	Evaluated 5 open-source software platforms against the CASI-Plus requirements developed by the study team and refined during step 1. Conducted heuristic evaluation (usability "rules of thumb").	Heuristic evaluations are common in HCD in which study team members compare a tool against heuristics to fix common usability issues [44,45].	Ukrainian and American study team members independently and asynchronously completed this activity and documented the results in a spreadsheet to minimize the need for synchronous web-based meetings.
Step 3—HCWs			
Usability walk-throughs	Provided "test scripts" for 5 fictional clients for HCWs to enter data and a template for documenting issues and recommendations.	Consider user needs and experiences [43].	The study team met via videoconference with HCWs to debrief their walk-throughs.
Step 3—clients			
Usability walk-throughs	A total of 5 clients participating in APSs tested CASI-Plus using a "talk aloud method." HCWs documented issues and recommendations in a structured template.	Consider user needs and experiences [43].	Clients rejected being video recorded when piloting the use of the tool.

[a]HCW: health care worker.

[b]APS: assisted partner service.

[c]IDI: in-depth interview.

Step 1: Formative HCD (May 2023-July 2023)

During step 1, with HCWs, we asked participants to (1) complete a preworkshop exercise to describe current APSs, including which personnel and what data collection or data management tools were involved; (2) discuss the current state of APSs, including barriers to partner notification and testing; (3) review 12 ideas for CASI-Plus features; and (4) prioritize features during a voting game in which participants selected top features, discussed their choices, and repeated 3 rounds of voting to allow them to change their priorities. We also carried out individual in-depth interviews with clients in which we asked them to (1) discuss current barriers to partner notification and testing and (2) critique mock-ups showing the client-facing screens and CASI-Plus tool design.

HCW workshop participants included a purposive sample of site managers, infectious disease physicians, nurses, psychologists, and social workers who deliver APSs and HIV care services. The first 2 HCW workshops included HCWs from the regional AIDS centers at hospitals in Chernihiv and Dnipro. The third workshop included HCWs from smaller rural primary care clinics that operate as subsidiary locations for the Chernihiv site. Because of safety concerns with travel of researchers to the study sites, the 3 HCW workshops were held via Zoom videoconference (Zoom Video Communications), with voluntary video and audio recording of the sessions. In total, 2 Ukrainian researchers facilitated the videoconference workshops using a facilitator's manual and slide decks. The English-speaking study team members listened in silently with synchronous translation to follow the proceedings and observe group dynamics. Before the workshops, participants were given an individual preworkshop assignment to describe APS consultation and case management activities. Between the first and second workshops, given the HCD approach, the study team iterated on potential CASI-Plus features to reflect feedback and better tailor ideas for the subsequent 2 HCW workshops.

For index clients, the study team held 30- to 45-minute in-person interviews at a single site because we could not assume that clients had access to the internet or familiarity with videoconferencing methods. We also reduced the number of client touch points during step 1 from 2 to 1 session to reduce participant burden during wartime. Clients were recruited from the Chernihiv oblast AIDS centers using a convenience sample of clients coming for routine medical visits. Inclusion criteria were (1) being aged ≥18 years and (2) having participated in APSs within the previous 12 months at the study clinic. Because we were particularly interested in the perspectives of clients from vulnerable, stigmatized groups, including men who have sex with men and persons who inject drugs, study personnel worked with HCWs to ensure that clients from these groups were referred to the study.

All workshops and client interviews were conducted in the state language, Ukrainian.

Step 2: Software Platform Assessment and Heuristic Evaluation (October 2023-May 2024)

During step 2, the study team evaluated 5 open-source software platforms against the CASI-Plus functional and technical requirements. The study team developed the requirements based on work in a preliminary study and updated this based on feedback from step 1. We then adapted the selected platform (REDCap [Research Electronic Data Capture; Vanderbilt University]) [46,47] to become the CASI-Plus tool and completed heuristic evaluation to assess usability. Each heuristic, or rule of thumb, for what makes for a pleasing and usable electronic tool was drawn from frameworks by Nielsen [44] and Kientz et al [45] using concepts such as "helps users recognize and recover from errors or relies on recognition rather than recall." We created a matrix tool for rating usability

XSL•FO
RenderX

according to each heuristic applied to each section of the CASI-Plus tool, with coding of each issue as having low, medium, or high impact on usability (Multimedia Appendix 1). Study team members then conducted walk-throughs of the tool with 5 fictional client "test scripts," rating each part of the CASI-Plus tool as passing or failing each heuristic. We collated all issues and addressed all modifiable high- and medium-impact issues through a round of improvements to the tool.

Step 3: Usability Walk-Throughs (July 2024)

Step 3 involved HCWs and clients conducting simulated walk-throughs of CASI-Plus at the Chernihiv study site using a live test version of the tool with the Android tablets procured for use of CASI-Plus at study sites. HCWs used the 5 fictional client test scripts, where they registered clients in the live version of CASI-Plus, entered client data, viewed the summary results, and edited client data as needed. After these walk-throughs, the PHC team held a focus group to gather feedback on navigation, system performance, and concerns using CASI-Plus in practice. In addition, the study coordinator completed walk-throughs with 5 actual clients, again using the live version of the tool and the tablets, followed by brief individual interviews to gather feedback, learn whether they would recommend the tool's use by others, and share suggested changes to the tool. Participants received similar incentives as in step 1.

Data Analysis

Ukrainian study team members created English-language rapid debrief reports for the data collected in step 1 using a template that followed the semistructured discussion guides to support rapid qualitative analysis [48-51]. Rapid qualitative analysis enabled actionable insights to emerge from each HCW workshop so that we could iterate on the ideas and visuals to consider input from each session. Later, full written transcripts were created and translated into English for additional analysis to reveal themes related to Ukraine's wartime and social context and inform the tool design. One analyst (NP) developed an inductive thematic codebook after an initial read-through of the transcripts and then applied the codebook to each transcript using the Dedoose software (SocioCultural Research Consultants) [52]. Data from steps 2 and 3 were tracked in spreadsheets and used in conjunction with the software requirements and design documents to guide modifications to the software prototype.

Ethical Considerations

The protocol for the CASI-Plus study was reviewed and approved by the University of Washington Human Subjects Division (00015808) and the institutional review board for the PHC (222; August 12, 2022). All participants provided written informed consent using a Ukrainian-language consent process and form. Health workers and clients were each offered an incentive of US $10 for participation in the CASI-Plus formative research. All data collected from study participants were deidentified before analysis and stored in secure, password-protected electronic storage locations.

Results

Overview

There were 22 HCWs and 10 clients participating in step 1 data collection (Table 3). All HCW participants (22/22, 100%) were female, most (13/22, 59%) were infectious disease or other physicians, and most (18/22, 82%) had provided HIV services for ≥5 years. Clients interviewed were mostly male (8/10, 80%) and included both persons who inject drugs (6/10, 60%) and men who have sex with men (4/10, 40%). Most clients interviewed (6/10, 60%) were aged 35 to 44 years.

XSL•FO
RenderX

Table 3. Study participants for step 1 of the Computer-Assisted Self-Interview–Plus formative research in wartime Ukraine (May 2023-July 2023).

Characteristic	Participants, n (%)
Health workers (n=22 individuals in 3 workshops)	
Location of work	
Oblast AIDS center	15 (68)
Peripheral health center	7 (32)
Sex	
Female	22 (100)
Male	0
Professional cadre	
Infectious disease physician	11 (50)
Nurse	7 (32)
Other physician	2 (9)
Psychologist	1 (5)
Social worker	1 (5)
Time working in HIV program (y)	
>10	8 (36)
5-10	10 (45)
2-5	4 (18)
Age group (y)	
<35	4 (18)
35-44	8 (36)
≥45	10 (45)
Clients (n=10 in-depth interviews)	
Sex	
Male	8 (80)
Female	2 (20)
Key population group	
Persons who inject drugs	6 (60)
Men who have sex with men	4 (40)
Age group (y)	
<35	2 (20)
35-44	6 (60)
≥45	2 (20)

Step 1: Formative HCD

APS Workflow

HCWs across the 3 workshops identified common, standard steps for the APS workflow, matching steps in the index client journey through initial and follow-up visits (Multimedia Appendix 2). The workflow activity resulted in a consensus on a standardized workflow that harmonized perspectives of HCWs within and across study sites. Noting that APS clients are newly engaged in care, HCWs discussed how some clients needed to build trust in the health care team before disclosing information about their partners, which underscored that CASI-Plus should be designed for longitudinal partner elicitation rather than elicitation at the initial APS enrollment visit. HCWs at one site reported eliciting information about household members and friends (in addition to HIV-exposed partners and biological children); however, the PHC researchers determined that "social network testing" was not part of the standard APS model and should not be a focus of the tool.

APS Barriers and Effects of Wartime

HCWs and clients noted barriers to APSs and challenges of wartime; Table 4 summarizes these barriers and challenges with illustrative quotes. Barriers to partner notification included fear of blame, rejection, family dissolution, and aggressive reactions

XSL•FO
RenderX

or abuse by partners, as well as clients' inability to come to terms with their HIV diagnosis. Both HCWs and clients discussed stigma and discrimination in health care by family physicians and other medical providers as being a barrier to HIV testing. Both groups also cited practical barriers to partner notification and testing, such as difficulty traveling to health facilities due to disrupted transportation networks, difficulty tracking partners due to widespread population displacement, and hesitancy to accept calls from unknown numbers given the efforts of military recruiters to enlist citizens in the armed forces. HCWs at the Chernihiv site explained that they did no HIV testing for 2 months while under Russian occupation in 2022. They estimated that HIV testing levels had rebounded to approximately 75% of the level in 2019, before the war and the COVID-19 pandemic, while noting that understaffing was an ongoing challenge in delivering APSs.

XSL•FO
RenderX

Table 4. Barriers to HIV assisted partner services (APSs) in wartime Ukraine—results from step 1 in-depth client interviews and health worker workshops (May 2023-July 2023).

Barrier	Themes	Illustrative quotes
Lack of acceptance of HIV diagnosis	1. Denial of HIV test result 2. Anger and blame of others	• "To be honest, I didn't immediately accept it, because I was on my own, I had the Internet at first, I was searching for all possible options why the test could be wrong. Well, I had a feeling that I wanted to turn this diagnosis inside out." [Client; male; man who has sex with men; aged 32 years] • "[A] person is offended that he/she was punished in this way, so they want to punish others, I believe. I had a situation like that...there was a time when I blamed everyone, I wanted to infect the whole world...[A] psychologist worked with me and everything is normal now, touch wood." [Client; female; person who injects drugs; aged 44 years]
Concern about disclosure of HIV status	1. Fear of "word getting out" about one's HIV status 2. Even an anonymous notification by an HCW[a] could allow a partner to deduce the source of the exposure	• "So, you [health worker] are calling and saying: 'Here, we have been informed...that you had sexual contact during last year.' The thing is, that we don't know how many were there, right? But he does. He knows that he was only with me." [Client; male; man who has sex with men; aged 32 years]
Fear of partners' reactions	1. Women and clients with newer partners may have more trouble notifying their partners 2. A client who was a man who has sex with men reported greater difficulty notifying bisexual partners because of fear of blame if wives and children become affected	• "[Y]ou're not, like, the same person anymore. You're HIV-positive. It's honestly stressing me out...[A] lot of people kind of say don't be afraid of this, but I'm already a second-class person." [Client; male; person who injects drugs; aged 44 years] • "[I]f, for example, he was single, yes [I could notify him]. Whatever was to happen let it happen. [But if] he is married, it's ok if everything is fine. But if not, it would be chaos in the family...I'm afraid of the reaction, so to speak." [Client; male; man who has sex with men; aged 32 years]
Clients have other priorities	1. Clients who are persons who inject drugs are generally difficult to engage due to their preoccupation with obtaining drugs or money for drugs if they are actively using drugs 2. Fatalistic wartime attitude 3. Low obligation to notify partner if they do not have an ongoing relationship	• "Due to the war, some clients have adopted a fatalistic attitude towards life and do not wish to inform their partners, making an excuse that they will live as long as they live, and what if they die tomorrow from a rocket." [HCW; female; 8 years of professional experience] • "[W]hen...I met my partner just once or twice...we met, he texted me or called me, I answered him, we met, and I don't even sign this [phone] number. Especially if I didn't like the partner, then, well, what's the point?" [Client; male; man who has sex with men; aged 32 years]
Difficulty disclosing partnerships to HCWs due to stigma	1. Shame to admit to health workers that clients have sex partners who are men who have sex with men or that they have shared needles 2. Social stigma can prevent partners from accessing HIV testing services because they do not want to be seen entering the HIV clinic for testing	• "[T]hey do not name their partners in all cases, they say 'I have no partner,' and then, after a while, we conduct conversations, we figure out that a partner of this or that client presented to us...It happens after some time, not immediately when a patient has just found out his status, rather after a while." [HCW; nurse; 8 years of professional experience] • "[T]here are some people who are coming here and try to hide the fact that they are using [drugs]...or they may say the opposite, 'I'm quitting' or something else...[T]hey will be afraid, if they have to tell the doctor, the doctor will just swear or will be negligent...like: 'Oh, they're drug addicts?...[T]hey're so...annoying,'...so they ignore it and don't talk about it." [Client; female; person who injects drugs; aged 43 years]
Ethical concerns	1. Concern about sharing partner contact details without asking them first (client) 2. Preference for routine HIV testing as part of primary health care, alleviating responsibility for notifying partners (client)	• "[P]ersonally, I don't provide phone numbers to anyone without asking the person if I can. I don't know if other people do it." [Client; male; man who has sex with men; aged 39 years]

XSL•FO
RenderX

Barrier	Themes	Illustrative quotes
Practical barriers to partner notification and testing	1. People may not pay attention to phone calls or messages attempting notification (HCWs and clients) 2. Massive displacement of the population during war creates difficulties tracing and contacting partners 3. Disrupted transportation networks make it difficult for citizens from rural and semiurban areas to access HIV testing	• "Due to the war, male clients and male partners are much more hesitant to answer calls from unknown numbers. They fear mobilization and are apprehensive about receiving calls from the military, so very often they don't pick up or don't trust strangers on the phone." [HCW; female; >10 years of professional experience]
Skeletal, overburdened HCW staff is challenged to provide APSs	1. HCWs have limited bandwidth to track partners over multiple attempts due to workload 2. Too few social workers to help physicians with APSs	• "I have 2 social workers here. One works part-time 25 per cent, which is 1 hour and 28 minutes, the other works part-time 75 per cent. A patient comes for ARVs...you need to find out about their sexual partners, you need to ask them about complaints, enter it [data] into the system of socially significant diseases, simultaneously enter it into the NHSU, into e-Health, so that we are paid for this case, write them out a prescription, refer them to tests… we are given 12 minutes per each patient...[Y]ou have to do your best, talk through everything quickly...There is such a race with each patient. And how to do all this, to ask them about all partners...they sit and try to remember, it can only be done properly in the social worker's office, if they are free..." [HCW; female; unit head; 23 years of professional service]

[a]HCW: health care worker.

Perceptions of the CASI-Plus Tool

Both HCWs and clients had an overall favorable perception of the CASI-Plus tool. HCWs and clients both felt that it would help clients discuss their partnerships without shame or embarrassment:

> *...there are patients who are sometimes ashamed to tell the doctor how many partners they have, because the person lives in a marriage and does not really want to reveal that they have someone anyone else, you now, and when it will be on the tablet, we will not look into their eyes, into their soul, as they say, perhaps the person will then simply open up. Perhaps a person will not name them, but at least they will say the number. As if it will be useful for us.* [HCW; female; >10 years of professional experience]

One client felt that it would help providers remember to discuss partner testing with each client:

> *...it's a great idea with a tablet, why, if the doctor, let's say, forgets to ask something, then that's it, I'm gone, I'll go home and think "[shoot], but I still wanted to talk about my problem, well, I won't, okay, later."* [Client; female; person who injects drugs; aged 43 years]

Of the 10 clients interviewed, all (100%) expressed willingness to use CASI-Plus on a tablet within the clinic, and only 1 (10%) expressed a definitive preference for speaking with an HCW about APSs rather than using CASI-Plus. HCWs expressed concern about the physical security and maintenance of tablet devices and made recommendations for preventing theft.

Preferred Features for the CASI-Plus Tool

Figure 1 shows the original concept for the CASI-Plus features, input from HCWs and clients about the features, and how we modified the design to address the feedback. The most important changes were (1) dropping 2-way texting with clients and (2) modifying the case management dashboard to focus on person-level information about partner disposition as reported by index clients rather than APS program performance metrics. Multimedia Appendix 2 summarizes feedback and shows how the design evolved across each step of the APS workflow, whereas Multimedia Appendix 3 shows the results of the HCW prioritization voting exercise.

During the 3 HCW workshops, HCWs prioritized features that support the initial APS encounter, including educational messaging about the rationale for APSs and how they work, as well as an initial partner elicitation questionnaire. Some HCWs and clients liked the idea of client stories that shared realistic examples of clients who pursued partner notification and testing. Clients liked the idea of CASI-Plus including 2-way texting because they felt that it would be a convenient way to communicate with their health care team. They felt that encrypted messaging (eg, via the Viber platform) would protect the privacy of communication. However, HCWs were concerned that 2-way texting would add to their workloads even if messaging was partially automated. They were also concerned about operating outside the current national legislative framework, which narrowly limits the collection and use of private information such as names and phone numbers.

There was less consensus among HCWs about CASI-Plus functionality to support APS case management. HCWs were enthusiastic about having a dashboard display of information that would help them track whether partners had completed HIV testing. However, they did not value a dashboard with performance indicators such as the total number of clients served or the average number of partners named per index client. HCWs said that they could not handle the additional data entry of case updates about contact attempts or partner testing disposition within the tool given their existing workload. They

XSL•FO
RenderX

recommended that clients capture partner follow-up data via a self-administered follow-up CASI-Plus questionnaire to be used whenever they returned to the clinic. Clients requested more health promotion information, such as messages normalizing HIV and explaining benefits of HIV pre-exposure prophylaxis and antiretroviral therapy. Design preferences included a large font for readability and a bright and hopeful visual design, with images that could help clients easily follow the text (Figure 2).

Figure 1. Evolution of the Computer-Assisted Self-Interview–Plus functional design based on step 1 formative research in wartime Ukraine (May 2023-July 2023).

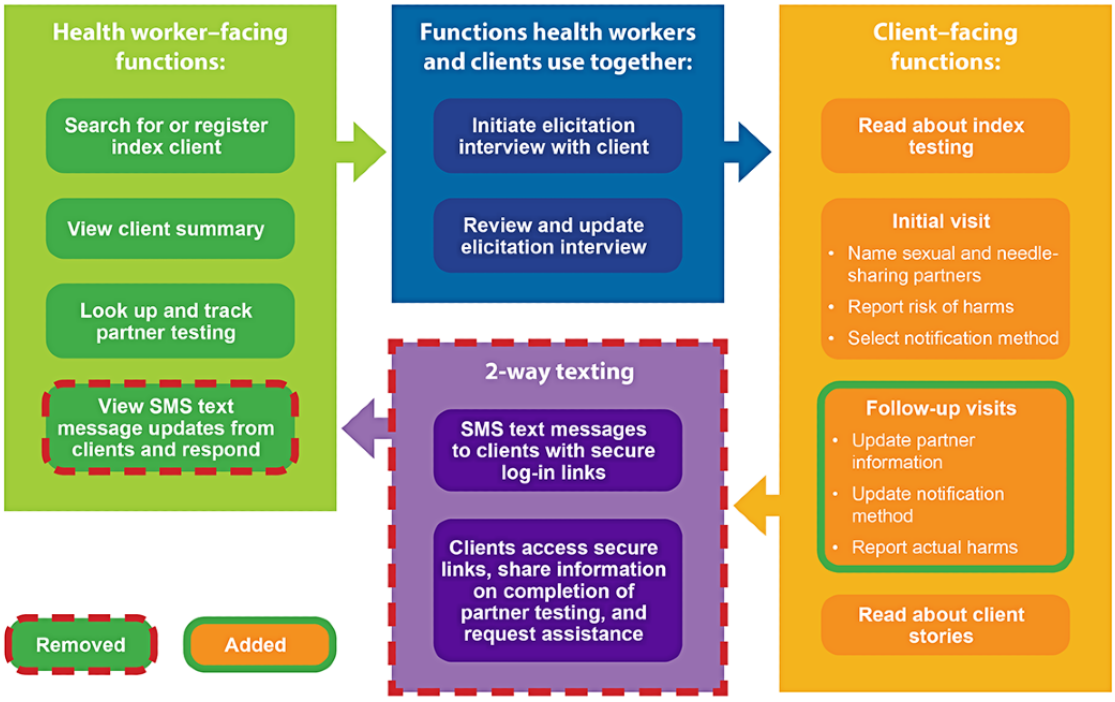
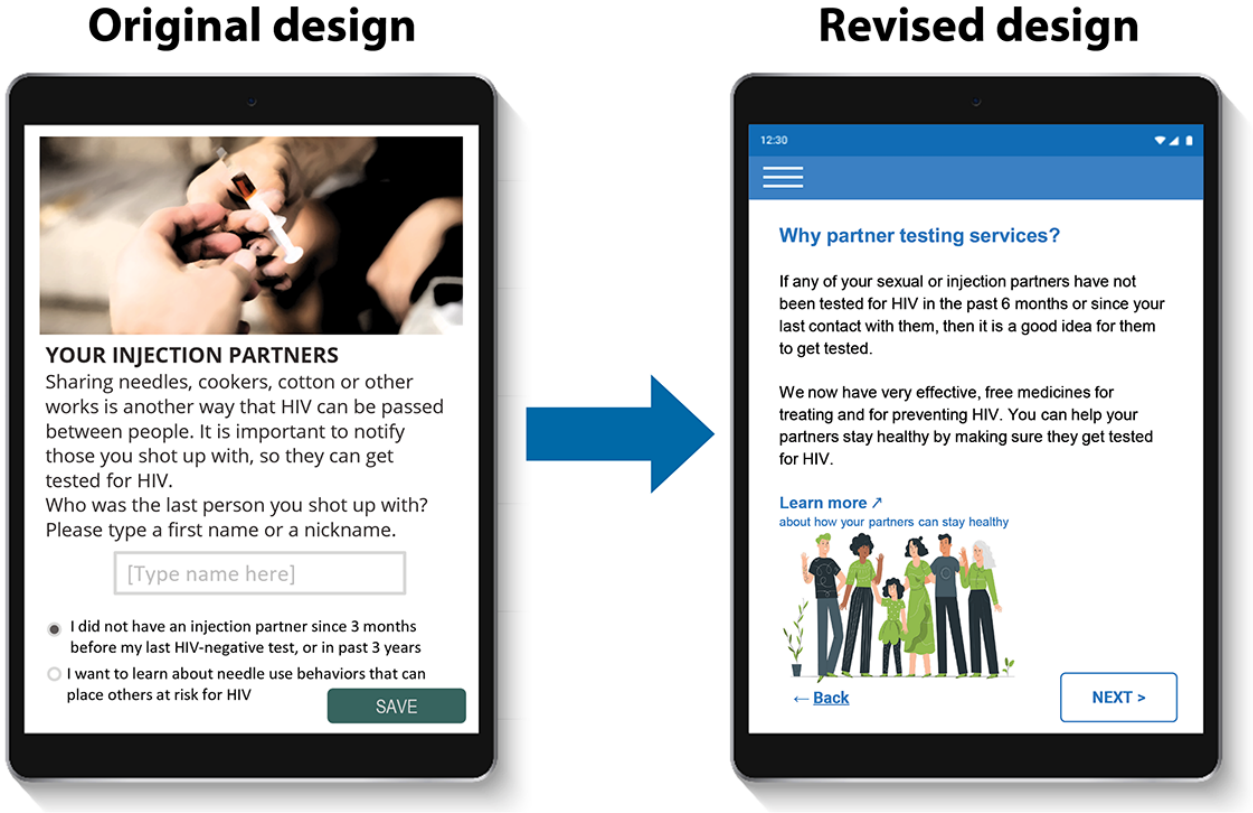


Figure 2. Evolution of the Computer-Assisted Self-Interview–Plus visual design based on step 1 formative research in wartime Ukraine (May 2023-July 2023).



XSL•FO
RenderX

Step 2: Software Platform Assessment and Heuristic Evaluation

An evaluation of 5 widely used open-source technology platforms for digital health resulted in the initial selection and adaptation of OpenSRP (Ona Systems) [53] as the optimal platform to meet the detailed functional and technical requirements developed by the team for CASI-Plus. The other 4 platforms considered were the Community Health Toolkit [54], DHIS2 Tracker [55], OpenMRS [56,57], and REDCap [46,47,58]. However, after building CASI-Plus using OpenSRP, the study encountered potential delays in implementing the tool due to risks posed by the ongoing war. The use of technology in Ukraine, particularly in wartime, required a careful approach to ensure data security and prevent leaks, reflecting a broader state-level policy on protecting personal and medical data. In the context of Russian cyber warfare threats, hosting CASI-Plus on secure, trusted servers within Ukraine became a top priority. Therefore, the study team emphasized the specific operational needs of having a platform with an appropriate license, a Ukrainian-language interface, and reliable technical support. These requirements were critical for aligning with national policy and ensuring the project's successful execution. Consequently, the team shifted to developing the final CASI-Plus prototype in REDCap, a platform already well established in Ukraine, even though it lacked certain desirable features identified during the requirement-gathering and design stages. The REDCap tool was hosted on PHC servers and was accessible as a web application to authorized users, meaning that internet connectivity at the sites was required for its use.

A total of 5 members of the PHC research team and 3 members of the University of Washington research team completed heuristic testing of the REDCap tool. Resulting improvements to CASI-Plus usability included (1) more consistent ways to advance between screens, with messages to inform clients about progress through the questionnaire; (2) improved instructions for questionnaire items with single versus multiple response options; (3) removal of duplicate items and content; and (4) improved page loading of images (Table 5).

Table 5. Illustrative examples of findings of the usability heuristic evaluation—step 2 of the Computer-Assisted Self-Interview–Plus formative research in wartime Ukraine (October 2023-May 2024).

Heuristic	Finding (severity)	Change to software
Consistency and standards—users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.	Inconsistent messages at the end of each section of the tool before moving on to the next section (medium)	Harmonized language showing progress through the sections of the tool
Error prevention—even better than good error messages is a careful design that prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.	Question flow allows users to enter a partner name while also indicating that they do not have a partner (eg, enter 2 mutually exclusive states at once; medium)	Modified question flow
Recognition rather than recall—minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.	For some questions, it is not obvious that multiple responses are allowed (low)	Added the description "Multiple answers allowed" to the questions with multiple response options
Aesthetic and minimalist design—dialogues should not contain information that is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units.	Confusing labeling and use of "show more" buttons in the section on methods of partner notification (high)	Changed style for "show more" buttons in REDCap[a]

[a]REDCap: Research Electronic Data Capture (Vanderbilt University).

Step 3: Usability Walk-Throughs

HCWs found that navigating the main menu to register and edit client records (an unmodifiable aspect of REDCap) was not intuitive but was learnable with practice and that navigating the client-facing sections of the tool was straightforward. They requested hosting separate instances of the REDCap tool for each study site to protect the confidentiality of client records across sites and prevent inappropriate editing of records by personnel from the other study sites. They recommended that future, scaled implementation of CASI-Plus would allow for selective access permissions by site, a feature built into the original OpenSRP-based tool but not available within REDCap. Overall, the HCWs were enthusiastic about testing the tool in clinical practice.

Clients also found CASI-Plus easy to use and understand. One user indicated that the tool "significantly simplifies notifying the partner; stories provide considerable support and confidence" (client; male).

There were several points of feedback. One client noted that the terms used for different partner notification methods were difficult to understand but that the explanations clarified the terminology. Another noted that it might be difficult to recall one's date of diagnosis. Client feedback did not require changes to the tool.

XSL•FO
RenderX

Discussion

Principal Findings

More interventions are needed to contain Ukraine's HIV epidemic and facilitate the effectiveness of large-scale APS programs in other low-resource settings. The CASI-Plus concept of an mHealth tool that gives persons with HIV a convenient way to learn about HIV index testing and report their partners and that helps HCWs track completion of partner testing was acceptable to both clients and HCWs in Ukraine. Prioritized features included information about the benefits of HIV index testing, a nonjudgmental self-guided questionnaire to report partners, client stories, and bright images to accompany text. Two-way SMS text messaging between clients and HCWs was not considered feasible to integrate within the workflow of understaffed public health clinics. Our final tool design and software platform choice (REDCap) reflected national policies and the principle of avoiding extra burden to clients and health workers, resulting in higher intention to engage with the tool.

Our focus was on developing a tool that would be of generalized relevance outside of Ukraine and that would also work well within the challenging conditions of Ukraine's wartime setting. This reflects the HCD design principle of "designing at the margins," in which tools that can address widespread problems are then adapted for marginalized use cases through an HCD process with the target population [59,60]. In our case, final choices for CASI-Plus features and design reflected the constraints of Ukraine's wartime setting but could be considered prudent choices for all contexts where health care resources are strained, workloads and time management are issues, and the target client population is stigmatized. The selection of the REDCap platform was driven by existing Ukraine Ministry of Health familiarity and capabilities. REDCap lacked several desired features that would have been possible with OpenSRP, such as hierarchical access permissions that would enable HCWs to view and edit the data only from their site while enabling scaled use across multiple sites and a data model supporting capture of record-level updates over time (rather than overwriting values through an editing process). Implementations of CASI-Plus in other settings might adapt the design but make a different technology choice.

This study contributes to the sparse literature on HCD methods for digital health interventions in war and humanitarian settings. The formative research phase of the CASI-Plus study demonstrated how simplified HCD methods adapted to the wartime context gathered rich input on the prioritized features and design of the CASI-Plus tool. There are few research studies or frameworks discussing HCD in mHealth projects in war or humanitarian settings. Existing published work has focused on tools for supporting the mental health of soldiers [61,62] or mHealth apps for health promotion among communities of refugees who have fled humanitarian crises [63-66]. Most existing work illustrates the role of the setting in shaping the problems and needs but not the specific ways in which the researchers adapted or shaped HCD methods based on a crisis context. For example, Beeman et al [65] described HCD methods applied to the design of menstrual health spaces for women living in a refugee settlement in Uganda. Their HCD methods involved standard steps of design research, rough prototyping, live prototyping, and piloting [65]. The resulting intervention design addressed the lack of privacy, tenuous access to disposable sanitary products, and other challenges facing women in the humanitarian setting. Another study of an mHealth app for Syrian refugees in Lebanon, a structured electronic health record with both client-facing features designed to improve health literacy and provider-facing features designed to support clinical decision-making, represents a cautionary tale in which design took place without participatory end-user input. The resulting tool failed to ease the workload of HCWs, and uptake was limited to approximately 20% of clinical consultations [67]. In another study, Bartlett et al [64] reported on an evaluation of an HCD process for a health care app for refugees. The evaluation findings noted power differentials in the design process and the need for diverse sampling techniques for client engagement so that design does not favor the perspectives of those with greater privilege [64].

Our formative study is unique in reporting on how we adapted mHealth HCD methods to a wartime context where the health system is under tremendous stress. Our study demonstrated several principles for successful HCD in a wartime context. First, we approached data collection so that it would be safe and feasible for both participants and study team members, all of whom were living through challenges of war, including displacement from their own homes; disrupted power supplies, including in winter; interrupted access to their offices, computers, and files; need to take shelter from bombs at a moment's notice; and having family members and friends risking their lives on the front lines. Drawing on the HCD principle of empathy, we relied on in-country investigators to guide the timing and design of participatory formative research activities, first brainstorming possible activities and then consulting with HCWs at the sites to confirm whether the activities would be feasible, hear their concerns, and modify plans. Second, we created a detailed facilitator guide for the HCD workshop and added guidance to workshop slides to support facilitators in real time in guiding the conversation toward the research aims. Third, we scheduled activities during the day (eg, in the afternoons before daily facility curfew closure times) and during weeks when wartime conditions (eg, fewer bombings and power outages) permitted.

Adoption of mHealth tools is always complex from a system perspective, affecting tasks by individuals, how teams work together, and how teams function within broader systems (the micro, meso, and macro levels discussed by Melles et al [18]). The complexity of absorbing system changes during war is amplified. Ukraine is struggling to maintain its health care delivery system in the face of direct attacks on health care infrastructure, departure of many health workers, and surging public health disease burden. Priorities for the health care system include carrying through with reforms started before the war (eg, decentralization of health care management and enabling of private providers to receive state funding to deliver approved HIV service packages) and adaptations resulting from the war (eg, modifications to the electronic data system for the HIV program [called the Information System for Socially Significant

XSL•FO
RenderX

Diseases] to enable remote use and tracking of patients receiving HIV care outside of Ukrainian borders). However, many adaptive responses in health care are still not formally recognized by the government, such as providers communicating with patients via messaging apps such as Viber or Telegram. In terms of CASI-Plus, our final design and prototype adapted to the existing APS workflows and formal regulations rather than introducing large changes that could be difficult to implement or sustain within understaffed public-sector clinics during wartime.

Future Priorities and Research

Future priorities for CASI-Plus technology development include features to reduce workloads for the already overworked HCWs, such as integrations with other existing routine data systems to avoid double data entry, and improved tracking of named partners. Future efforts are warranted to explore the use of 2-way SMS text messaging at a time when Ukrainian health policy makers can codify formal regulations for this type of technology. The next step in research on CASI-Plus for Ukraine's APS program is a randomized controlled trial of the effect of the REDCap-based CASI-Plus on the number of partners named and tested for HIV. CASI-Plus has the potential to complement other strategies being used in Ukraine to support HIV case finding in the wartime context. Previous studies have identified that APSs applied to social networks, particularly for persons who inject drugs, are effective for HIV case finding [33]. A recent study combined a social network approach to APSs with a machine learning algorithm applied to screening data from index clients to identify the networks most likely to yield new HIV cases [68]. Future research could study the use of CASI-Plus as part of social network testing and in conjunction with a prediction algorithm, as well as cultural adaptation of CASI-Plus in other countries.

Strengths and Limitations

Strengths of this study were the adaptation of HCD methods to Ukraine's wartime context, the inclusion of perspectives from 2 regions and both larger and smaller HIV service sites, and openness to HCW and client perspectives about the direction for CASI-Plus. Limitations of this study included the inability of HCWs and research team members to meet in person for creative co-design sessions; time pressure to complete the HCW workshops, where several participants had to drop out early when their worksites closed; and low engagement from rural, peripheral sites due to unfamiliarity with Zoom videoconferencing technology. Favorable opinions expressed about the CASI-Plus concept may reflect social desirability bias or a tendency to conform to the priorities of external funders because of the resources they provide [69]. Engaging participants from only 3 out of hundreds of primary health care sites may also have limited feature elicitation and utility across the spectrum of service delivery, particularly given the nuances in wartime settings. A limitation of pursuing CASI-Plus technology design during wartime was that national policies on technology platforms and patient electronic communications were less flexible than we anticipated when we conceptualized the study before the war, meaning that we were not able to implement 2-way SMS text messaging as part of the tool.

Conclusion

In conclusion, the CASI-Plus mHealth tool is an acceptable intervention to support HIV testing targeted at people most at risk of HIV exposure, namely, the sexual and needle-sharing partners of people who have been newly diagnosed with HIV. It was possible to adapt HCD methods grounded in principles of empathy, iteration, and creative ideation to Ukraine's wartime context using internet-based workshops with HCWs and individual interviews with clients. This work demonstrates the flexibility and value of HCD methods for adapting digital health tools in a wartime context.

Acknowledgments

The authors are grateful to Oleksandr Kurdus, head of the Information Technology Department, Public Health Center of the Ministry of Health of Ukraine, for assistance with implementation of the Computer-Assisted Self-Interview (CASI)–Plus tool in Ukraine and for review and input on this manuscript. They acknowledge Abert Namaya of the Digital Initiatives Group at the International Training and Education Center for Health at the University of Washington for contributions to the CASI-Plus technology assessment. This study would not have been possible without the support of all the health care workers and clients who chose to participate in the CASI-Plus formative study despite the disruptions and trauma of the war in Ukraine. The research reported in this publication was supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award R21EB032229. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Usability heuristic template for step 2 Computer-Assisted Self-Interview–Plus heuristic evaluation in wartime Ukraine (October 2023-May 2024).
[DOCX File , 50 KB - formative_v9i1e66132_app1.docx]

Multimedia Appendix 2
Feedback on the Computer-Assisted Self-Interview–Plus proposed features and resulting design modifications based on the step 1 formative research in wartime Ukraine (May 2023-July 2023).
[DOC File , 49 KB - formative_v9i1e66132_app2.doc]

Multimedia Appendix 3
Prioritization of Computer-Assisted Self-Interview–Plus features by health care workers during step 1 formative workshops in wartime Ukraine (May 2023-July 2023).
[DOCX File , 19 KB - formative_v9i1e66132_app3.docx]

References

1. Guidelines on HIV self-testing and partner notification: supplement to consolidated guidelines on HIV testing services. World Health Organization. 2016. URL: https://iris.who.int/handle/10665/251655 [accessed 2024-12-09]
2. Consolidated guidelines on HIV testing services for a changing epidemic. World Health Organization. 2019 Nov 27. URL: https://www.who.int/publications/i/item/WHO-CDS-HIV-19.31 [accessed 2024-12-09]
3. Cherutich P, Golden MR, Wamuti B, Richardson BA, Ásbjörnsdóttir KH, Otieno FA, et al. Assisted partner services for HIV in Kenya: a cluster randomised controlled trial. The Lancet HIV 2017 Feb;4(2):e74-e82. [doi: 10.1016/s2352-3018(16)30214-4]
4. Rosenberg NE, Mtande TK, Saidi F, Stanley C, Jere E, Paile L, et al. Recruiting male partners for couple HIV testing and counselling in Malawi's option B+ programme: an unblinded randomised controlled trial. Lancet HIV 2015 Nov;2(11):e483-e491 [FREE Full text] [doi: 10.1016/S2352-3018(15)00182-4] [Medline: 26520928]
5. Brown LB, Miller WC, Kamanga G, Nyirenda N, Mmodzi P, Pettifor A, et al. HIV partner notification is effective and feasible in sub-Saharan Africa: opportunities for HIV treatment and prevention. J Acquir Immune Defic Syndr 2011 Apr 15;56(5):437-442 [FREE Full text] [doi: 10.1097/qai.0b013e318202bf7d] [Medline: 22046601]
6. Landis SE, Schoenbach VJ, Weber DJ, Mittal M, Krishan B, Lewis K, et al. Results of a randomized trial of partner notification in cases of HIV infection in North Carolina. N Engl J Med 1992 Jan 09;326(2):101-106. [doi: 10.1056/NEJM199201093260205] [Medline: 1445500]
7. Rutstein SE, Brown LB, Biddle AK, Wheeler SB, Kamanga G, Mmodzi P, et al. Cost-effectiveness of provider-based HIV partner notification in urban Malawi. Health Policy Plan 2014 Jan 15;29(1):115-126 [FREE Full text] [doi: 10.1093/heapol/czs140] [Medline: 23325584]
8. Sharma M, Smith JA, Farquhar C, Ying R, Cherutich P, Golden M, et al. Assisted partner notification services are cost-effective for decreasing HIV burden in western Kenya. AIDS 2018 Jan 14;32(2):233-241 [FREE Full text] [doi: 10.1097/QAD.0000000000001697] [Medline: 29135576]
9. Myers RS, Feldacker C, Cesár F, Paredes Z, Augusto G, Muluana C, et al. Acceptability and effectiveness of assisted human immunodeficiency virus partner services in Mozambique: results from a pilot program in a public, urban clinic. Sexual Trans Dis 2016;43(11):690-695. [doi: 10.1097/olq.0000000000000529]
10. Henley C, Forgwei G, Welty T, Golden M, Adimora A, Shields R, et al. Scale-up and case-finding effectiveness of an HIV partner services program in Cameroon: an innovative HIV prevention intervention for developing countries. Sex Transm Dis 2013 Dec;40(12):909-914 [FREE Full text] [doi: 10.1097/OLQ.0000000000000032] [Medline: 24220349]
11. Shamu S, Farirai T, Kuwanda L, Slabbert J, Guloba G, Khupakonke S, et al. Comparison of community-based HIV counselling and testing (CBCT) through index client tracing and other modalities: outcomes in 13 South African high HIV prevalence districts by gender and age. PLoS One 2019 Sep 6;14(9):e0221215 [FREE Full text] [doi: 10.1371/journal.pone.0221215] [Medline: 31490938]
12. Katbi M, Adegboye A, Adedoyin A, Yunusa F, Kayode G, Bello M, et al. Effect of clients Strategic Index Case Testing on community-based detection of HIV infections (STRICT study). Int J Infect Dis 2018 Sep;74:54-60 [FREE Full text] [doi: 10.1016/j.ijid.2018.06.018] [Medline: 30100537]
13. Ihnatiuk AP, Shapoval AY, Kazanzhy AP, Kuzin IV, Riabokon SV, Shotorbani S, et al. Distinctive model for HIV index testing (IT) in Eastern Europe: results of Ukraine's physician-led, integrated IT programme. BMJ Open 2023 Sep 20;13(9):e070542 [FREE Full text] [doi: 10.1136/bmjopen-2022-070542] [Medline: 37730414]
14. Adebajo S, Obianwu O, Eluwa G, Vu L, Oginni A, Tun W, et al. Comparison of audio computer assisted self-interview and face-to-face interview methods in eliciting HIV-related risks among men who have sex with men and men who inject drugs in Nigeria. PLoS One 2014 Jan 8;9(1):e81981 [FREE Full text] [doi: 10.1371/journal.pone.0081981] [Medline: 24416134]

XSL•FO
RenderX

15. Beauclair R, Meng F, Deprez N, Temmerman M, Welte A, Hens N, et al. Evaluating audio computer assisted self-interviews in urban South African communities: evidence for good suitability and reduced social desirability bias of a cross-sectional survey on sexual behaviour. BMC Med Res Methodol 2013 Jan 31;13(1):11 [FREE Full text] [doi: 10.1186/1471-2288-13-11] [Medline: 23368888]
16. NIMH Collaborative HIV/STD Prevention Trial Group. The feasibility of audio computer-assisted self-interviewing in international settings. AIDS 2007 Apr;21 Suppl 2(Suppl 2):S49-S58. [doi: 10.1097/01.aids.0000266457.11020.f0] [Medline: 17413264]
17. John SA, Petroll AE, Walsh JL, Quinn KG, Patel VV, Grov C. Reducing the discussion divide by digital questionnaires in health care settings: disruptive innovation for HIV testing and PrEP screening. J Acquir Immune Defic Syndr 2020 Nov 01;85(3):302-308 [FREE Full text] [doi: 10.1097/QAI.0000000000002459] [Medline: 32701822]
18. Melles M, Albayrak A, Goossens R. Innovating health care: key characteristics of human-centered design. Int J Qual Health Care 2021 Jan 12;33(Supplement_1):37-44 [FREE Full text] [doi: 10.1093/intqhc/mzaa127] [Medline: 33068104]
19. Evans L, Evans J, Pagliari C, Källander K. Scoping review: exploring the equity impact of current digital health design practices. Oxford Open Digit Health 2023;1:oqad006. [doi: 10.1093/oodh/oqad006]
20. Beres LK, Simbeza S, Holmes CB, Mwamba C, Mukamba N, Sharma A, et al. Human-centered design lessons for implementation science: improving the implementation of a patient-centered care intervention. J Acquir Immune Defic Syndr 2019 Dec;82 Suppl 3(3):S230-S243 [FREE Full text] [doi: 10.1097/QAI.0000000000002216] [Medline: 31764259]
21. Bazzano AN, Martin J, Hicks E, Faughnan M, Murphy L. Human-centred design in global health: a scoping review of applications and contexts. PLoS One 2017;12(11):e0186744 [FREE Full text] [doi: 10.1371/journal.pone.0186744] [Medline: 29091935]
22. Holeman I, Kane D. Human-centered design for global health equity. Inf Technol Dev 2019 Sep 29;26(3):477-505 [FREE Full text] [doi: 10.1080/02681102.2019.1667289] [Medline: 32982007]
23. Cheney C. How organizations can apply human-centered design in crisis response. Devex. 2022 Apr 8. URL: https://www.devex.com/news/how-organizations-can-apply-human-centered-design-in-crisis-response-102778 [accessed 2024-07-06]
24. Ukraine refugee situation. United Nations High Commissioner for Refugees Operational Data Portal. 2024. URL: https://data.unhcr.org/en/situations/ukraine [accessed 2024-01-08]
25. From now on, Ukrainians forcibly deported to the Russian Federation can issue a certificate for return to Ukraine. State Migration Service, Ministry of Internal Affairs of Ukraine. 2022. URL: https://tinyurl.com/yhk644sa [accessed 2024-12-09]
26. Janowski K. Civilian deaths in Ukraine war top 10,000, UN says. United Nations Ukraine. 2023 Nov 21. URL: https://tinyurl.com/y54atac8 [accessed 2024-07-06]
27. Ukraine — snapshot report: population figures and geographic distribution — general population survey round 13 (11 - 23 May). International Organization for Migration. 2023. URL: https://tinyurl.com/4zmcrjhn [accessed 2024-12-03]
28. National response of HIV, TB, viral hepatitis and SMT programmes in the context of full-scale Russian invasion. Public Health Center of the MOH of Ukraine. 2023. URL: https://phc.org.ua/sites/default/files/users/user90/National_response_HIV_TB_VH_SMT_war_2023_ENG.pdf [accessed 2024-10-10]
29. About registered criminal offenses and the results of their pre-trial investigation. Office of the Prosecutor General, Government of Ukraine. 2022. URL: https://tinyurl.com/487f3as5 [accessed 2024-01-25]
30. Levy BS, Leaning J. Russia's war in Ukraine - the devastation of health and human rights. N Engl J Med 2022 Jul 14;387(2):102-105. [doi: 10.1056/NEJMp2207415] [Medline: 35767514]
31. Surveillance system for attacks on healthcare (SSA). World Health Organization. URL: https://tinyurl.com/2ms8z99w [accessed 2024-11-19]
32. McCarthy S, George S, Haq SN, Macaya M, Hayes M, Chowdhury M, et al. April 3, 2022 Russia-Ukraine news. CNN World. 2022 Apr 4. URL: https://www.cnn.com/europe/live-news/ukraine-russia-putin-news-04-3-22/index.html [accessed 2024-10-10]
33. Williams LD, Korobchuk A, Smyrnov P, Sazonova Y, Nikolopoulos GK, Skaathun B, et al. Social network approaches to locating people recently infected with HIV in Odessa, Ukraine. J Int AIDS Soc 2019 Jun 27;22(6):e25330 [FREE Full text] [doi: 10.1002/jia2.25330] [Medline: 31245917]
34. Booth RE, Davis JM, Dvoryak S, Brewster JT, Lisovska O, Strathdee SA, et al. HIV incidence among people who inject drugs (PWIDs) in Ukraine: results from a clustered randomised trial. The Lancet HIV 2016 Oct;3(10):e482-e489. [doi: 10.1016/s2352-3018(16)30040-6]
35. The key population atlas. Joint United Nations Programme on HIV/AIDS. URL: https://kpatlas.unaids.org/dashboard [accessed 2024-10-10]
36. Puttkammer N, Ihnatiuk A, Shapoval A, Kazanzhy A, Secor A, Shotorbani S, et al. Profile of partners who completed HIV testing and received a new HIV diagnosis in Ukraine's HIV index testing program: a retrospective cohort study to inform program improvement. BMC Infect Dis 2023 May 05;23(1):291 [FREE Full text] [doi: 10.1186/s12879-023-08281-1] [Medline: 37147618]
37. Dumchev K, Kornilova M, Kulchynska R, Azarskova M, Vitek C. Improved ascertainment of modes of HIV transmission in Ukraine indicates importance of drug injecting and homosexual risk. BMC Public Health 2020 Aug 26;20(1):1288 [FREE Full text] [doi: 10.1186/s12889-020-09373-2] [Medline: 32843008]

XSL•FO
RenderX

38. Dumchev K, Sazonova Y, Salyuk T, Varetska O. Trends in HIV prevalence among people injecting drugs, men having sex with men, and female sex workers in Ukraine. Int J STD AIDS 2018 Nov 27;29(13):1337-1344. [doi: 10.1177/0956462418784096] [Medline: 30049259]
39. Friedman SR, Smyrnov P, Vasylyeva TI. Will the Russian war in Ukraine unleash larger epidemics of HIV, TB and associated conditions and diseases in Ukraine? Harm Reduct J 2023 Sep 01;20(1):119 [FREE Full text] [doi: 10.1186/s12954-023-00855-1] [Medline: 37658448]
40. LaFond A, Cherney M. A theory of change for guiding the integration of human-centered design into global health programming. Glob Health Sci Pract 2021 Nov 29;9(Supplement 2):S209-S216. [doi: 10.9745/ghsp-d-21-00334]
41. Jewell C, Salvetti F. Towards a combined method of web usability testing: an assessment of the complementary advantages of lab testing, pre-session assignments, and online usability services. In: Proceedings of the CHI '12 Extended Abstracts on Human Factors in Computing Systems. 2012 Presented at: CHI EA '12; May 5-10, 2012; Austin, TX. [doi: 10.1145/2212776.2223720]
42. Jewell C, Salvetti S. Using pre-session homework to elicit more insights during web usability testing. In: Proceedings of the HCI International 2011 Posters' Extended Abstracts. 2011 Presented at: HCI International 2011; July 9-14, 2011; Orlando, FL. [doi: 10.1007/978-3-642-22098-2_32]
43. Dopp AR, Parisi KE, Munson SA, Lyon AR. Aligning implementation and user-centered design strategies to enhance the impact of health services: results from a concept mapping study. Implement Sci Commun 2020 Feb 26;1(1):17 [FREE Full text] [doi: 10.1186/s43058-020-00020-w] [Medline: 32885179]
44. Nielsen J. 10 usability heuristics for user interface design. Nielsen Norman Group. 1994 Apr 24. URL: https://www.nngroup.com/articles/ten-usability-heuristics/ [accessed 2024-07-06]
45. Kientz JA, Choe EK, Birch B, Maharaj R, Fonville A, Glasson C, et al. Heuristic evaluation of persuasive health technologies. In: Proceedings of the 1st ACM International Health Informatics Symposium. 2010 Presented at: IHI '10; November 11-12, 2010; Arlington, VA. [doi: 10.1145/1882992.1883084]
46. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
47. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
48. Vindrola-Padros C, Johnson GA. Rapid techniques in qualitative research: a critical review of the literature. Qual Health Res 2020 Aug 15;30(10):1596-1604. [doi: 10.1177/1049732320921835] [Medline: 32667277]
49. Taylor B, Henshall C, Kenyon S, Litchfield I, Greenfield S. Can rapid approaches to qualitative analysis deliver timely, valid findings to clinical leaders? A mixed methods study comparing rapid and thematic analysis. BMJ Open 2018 Oct 08;8(10):e019993 [FREE Full text] [doi: 10.1136/bmjopen-2017-019993] [Medline: 30297341]
50. Gale RC, Wu J, Erhardt T, Bounthavong M, Reardon CM, Damschroder LJ, et al. Comparison of rapid vs in-depth qualitative analytic methods from a process evaluation of academic detailing in the Veterans Health Administration. Implement Sci 2019 Feb 01;14(1):11 [FREE Full text] [doi: 10.1186/s13012-019-0853-y] [Medline: 30709368]
51. Simoni JM, Beima-Sofie K, Amico KR, Hosek SG, Johnson MO, Mensch BS. Debrief reports to expedite the impact of qualitative research: do they accurately capture data from in-depth interviews? AIDS Behav 2019 Aug 21;23(8):2185-2189 [FREE Full text] [doi: 10.1007/s10461-018-02387-3] [Medline: 30666522]
52. Dedoose hoempage. Dedoose. URL: https://www.dedoose.com/ [accessed 2024-04-17]
53. Open smart register platform homepage. OpenSRP. URL: https://opensrp.io/ [accessed 2024-10-10]
54. Tools - community health toolkit. Community Health Toolkit. URL: https://communityhealthtoolkit.org/tools [accessed 2024-10-10]
55. Individual data records with tracker. DHIS2. URL: https://dhis2.org/tracker/ [accessed 2024-10-10]
56. OpenMRS hoempage. OpenMRS. URL: https://openmrs.org/ [accessed 2024-10-10]
57. Mamlin BW, Biondich PG, Wolfe BA, Fraser H, Jazayeri D, Allen C, et al. Cooking up an open source EMR for developing countries: OpenMRS - a recipe for successful collaboration. AMIA Annu Symp Proc 2006;2006:529-533 [FREE Full text] [Medline: 17238397]
58. Research Electronic Data Capture (REDCap). Institute of Translational Health Sciences. URL: https://www.iths.org/investigators/services/bmi/redcap/ [accessed 2021-01-18]
59. Designing at the margins. Getting Smart. URL: https://www.gettingsmart.com/2023/10/05/designing-at-the-margins/ [accessed 2024-10-10]
60. Rigot A. Design from the margins. Belfer Center for Science and International Affairs. 2022. URL: https://tinyurl.com/3t9xk37s [accessed 2024-12-09]
61. Rizzo A, Pair J, McNerney PJ, Eastlund E, Manson B, Gratch J, et al. Development of a VR therapy application for Iraq war military personnel with PTSD. Stud Health Technol Inform 2005;111:407-413. [Medline: 15718769]
62. Mozgai SA, Femminella B, Hartholt A, Rizzo A. User-centered design model for mobile health (mHealth) applications: a military case study in rapid assessment process (RAP). In: Proceedings of the Extended Abstracts of the 2021 CHI Conference

XSL•FO
RenderX

on Human Factors in Computing Systems. 2021 Presented at: CHI EA '21; May 8-13, 2021; Yokohama, Japan. [doi: 10.1145/3411763.3443450]
63. Sikder A, Dickhoner J, Kysh L, Musheghyan L, Shekherdimian S, Levine B, et al. International collaboration to develop a remote monitoring web app for COVID-19 patients in Armenia: design and development with agile methodology. JMIR Hum Factors 2022 Nov 25;9(4):e40110 [FREE Full text] [doi: 10.2196/40110] [Medline: 36350739]
64. Bartlett R, Boyle JA, Simons Smith J, Khan N, Robinson T, Ramaswamy R. Evaluating human-centred design for public health: a case study on developing a healthcare app with refugee communities. Res Involv Engagem 2021 May 30;7(1):32 [FREE Full text] [doi: 10.1186/s40900-021-00273-2] [Medline: 34053451]
65. Beeman A, Kwesiga J, Ippoliti N, Bhandari T, Pandya G, Acam FA, et al. Using human-centered design to co-design dedicated menstrual health spaces with people who menstruate in Bidi Bidi refugee settlement, Uganda: learnings for further adaptation and scale in humanitarian settings. BMC Womens Health 2023 Jun 20;23(1):319 [FREE Full text] [doi: 10.1186/s12905-023-02421-0] [Medline: 37340385]
66. Black CJ, Berent JM, Joshi U, Khan A, Chamlagai L, Shrivastava R, et al. Applying human-centered design in global mental health to improve reach among underserved populations in the United States and India. Glob Health Sci Pract 2023 Feb 28;11(1):e2200312. [doi: 10.9745/ghsp-d-22-00312]
67. Doocy S, Paik K, Lyles E, Tam HH, Fahed Z, Winkler E, et al. Pilot testing and implementation of a mHealth tool for non-communicable diseases in a humanitarian setting. PLoS Curr 2017 Jun 05;9:ecurrents.dis.e98c648aac93797b1996a37de099be74 [FREE Full text] [doi: 10.1371/currents.dis.e98c648aac93797b1996a37de099be74] [Medline: 28744410]
68. Holt E. AI-assisted HIV case finding in Ukraine. The Lancet Microbe 2024 Jul;5(7):631. [doi: 10.1016/s2666-5247(24)00077-6]
69. Spicer N, Harmer A, Aleshkina J, Bogdan D, Chkhatarashvili K, Murzalieva G, et al. Circus monkeys or change agents? Civil society advocacy for HIV/AIDS in adverse policy environments. Soc Sci Med 2011 Dec;73(12):1748-1755. [doi: 10.1016/j.socscimed.2011.08.024] [Medline: 22036298]

Abbreviations

APS: assisted partner service
CASI-Plus: Computer-Assisted Self-Interview–Plus
HCD: human-centered design
HCW: health care worker
mHealth: mobile health
PHC: Public Health Center
REDCap: Research Electronic Data Capture

Edited by A Mavragani; submitted 18.10.24; peer-reviewed by A Zheluk, C Baxter; comments to author 24.11.24; revised version received 09.12.24; accepted 12.12.24; published 30.01.25.

Please cite as:
Puttkammer N, Dunbar E, Germanovych M, Rosol M, Golden M, Hubashova A, Fedorchenko V, Hetman L, Legkostup L, Flowers J, Nesterova O
Human-Centered Design of an mHealth Tool for Optimizing HIV Index Testing in Wartime Ukraine: Formative Research Case Study
JMIR Form Res 2025;9:e66132
URL: https://formative.jmir.org/2025/1/e66132
doi:10.2196/66132
PMID:

©Nancy Puttkammer, Elizabeth Dunbar, Myroslava Germanovych, Mariia Rosol, Matthew Golden, Anna Hubashova, Vladyslav Fedorchenko, Larisa Hetman, Liudmyla Legkostup, Jan Flowers, Olena Nesterova. Originally published in JMIR Formative Research (https://formative.jmir.org), 30.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluating the Impact of Pharmacotherapy in Augmenting Quit Rates Among Hispanic Adults in an App-Delivered Smoking Cessation Intervention: Secondary Analysis of a Randomized Controlled Trial

Margarita Santiago-Torres[1], PhD; Kristin E Mull[1], MS; Brianna M Sullivan[1], MS; Ana Paula Cupertino[2], PhD; Ramzi G Salloum[3], PhD; Matthew Triplette[1,4,5], MD, PhD; Michael J Zvolensky[6,7], PhD; Jonathan B Bricker[1,8], PhD

[1]Division of Public Health Sciences, Fred Hutchinson Cancer Center, Seattle, WA, United States

[2]Department of Public Health Sciences, University of Rochester Medical Center, Rochester, NY, United States

[3]Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, Gainesville, FL, United States

[4]Division of Pulmonary, Critical Care and Sleep Medicine, Department of Medicine, University of Washington, Seattle, WA, United States

[5]Clinical Research Division, Fred Hutchinson Cancer Center, Seattle, WA, United States

[6]Department of Psychology, University of Houston, Houston, TX, United States

[7]MD Anderson Cancer Center, University of Texas, Houston, TX, United States

[8]Department of Psychology, University of Washington, Seattle, WA, United States

Corresponding Author:
Margarita Santiago-Torres, PhD
Division of Public Health Sciences
Fred Hutchinson Cancer Center
1100 Fairview Avenue North
Seattle, WA, 98109
United States
Phone: 1 206 667 4780
Fax: 1 206 667 5977
Email: msantiag@fredhutch.org

Abstract

Background: Hispanic adults receive less advice to quit smoking and use fewer evidence-based smoking cessation treatments compared to their non-Hispanic counterparts. Digital smoking cessation interventions, such as those delivered via smartphone apps, provide a feasible and within-reach treatment option for Hispanic adults who smoke and want to quit smoking. While the combination of pharmacotherapy and behavioral interventions are considered best practices for smoking cessation, its efficacy among Hispanic adults, especially alongside smartphone app–based interventions, is uncertain.

Objective: This secondary analysis used data from a randomized controlled trial that compared the efficacy of 2 smoking cessation apps, iCanQuit (based on acceptance and commitment therapy) and QuitGuide (following US clinical practice guidelines), to explore the association between pharmacotherapy use and smoking cessation outcomes among the subsample of 173 Hispanic participants who reported on pharmacotherapy use. Given the randomized design, we first tested the potential interaction of pharmacotherapy use and intervention arm on 12-month cigarette smoking abstinence. We then examined whether the use of any pharmacotherapy (ie, nicotine replacement therapy [NRT], varenicline, or bupropion) and NRT alone augmented each app-based intervention efficacy.

Methods: Participants reported using pharmacotherapy on their own during the 3-month follow-up and cigarette smoking abstinence at the 12-month follow-up via web-based surveys. These data were used (1) to test the interaction effect of using pharmacotherapy to aid smoking cessation and intervention arm (iCanQuit vs QuitGuide) on smoking cessation at 12 months and (2) to test whether the use of pharmacotherapy to aid smoking cessation augmented the efficacy of each intervention arm to help participants successfully quit smoking.

Results: The subsample of Hispanic participants was recruited from 30 US states. They were on average 34.5 (SD 9.3) years of age, 50.9% (88/173) were female, and 56.1% (97/173) reported smoking at least 10 cigarettes daily. Approximately 22% (38/173) of participants reported using pharmacotherapy to aid smoking cessation at the 3-month follow-up, including NRT,

XSL•FO
RenderX

varenicline, or bupropion, with no difference between intervention arms. There was an interaction between pharmacotherapy use and intervention arm that marginally influenced 12-month quit rates at 12 months (*P* for interaction=.053). In the iCanQuit arm, 12-month missing-as-smoking quit rates were 43.8% (7/16) for pharmacotherapy users versus 28.8% (19/16) for nonusers (odds ratio 2.21, 95% CI 0.66-7.48; *P*=.20). In the QuitGuide arm, quit rates were 9.1% (2/22) for pharmacotherapy users versus 21.7% (15/69) for nonusers (odds ratio 0.36, 95% CI 0.07-1.72; *P*=.20). Results were similar for the use of NRT only.

Conclusions: Combining pharmacotherapy to aid smoking cessation with a smartphone app–based behavioral intervention that teaches acceptance of cravings to smoke (iCanQuit) shows promise in improving quit rates among Hispanic adults. However, this combined approach was not effective with the US clinical guideline–based app (QuitGuide).

Trial Registration: ClinicalTrials.gov NCT02724462; https://clinicaltrials.gov/study/NCT02724462

International Registered Report Identifier (IRRID): RR2-10.1001/jamainternmed.2020.4055

(JMIR Form Res 2025;9:e69311) doi:10.2196/69311

KEYWORDS

acceptance and commitment therapy; Hispanic or Latino; iCanQuit; QuitGuide; smartphone apps; smoking cessation; mobile phone

Introduction

Background

In the United States, 1 in 5 adults are Hispanic or Latino (referred to hereafter as Hispanic). Hispanic adults in the United States constitute the largest ethnic minority group [1,2]. In fact, it is estimated that by 2026, the Hispanic population will reach 111 million [3]. Presently, there are approximately 4 million Hispanic adults who smoke cigarettes [4]. Even though smoking rates are lower among Hispanic adults (7.7% vs 11.7%) when compared to the general population of adults in the United States who smoke [5], the prevalence of smoking varies across different Hispanic nationalities [6-8]. For instance, people from Puerto Rico and Cuba have higher than average smoking rates at 25.5% and 19.8%, respectively. Hispanic adults also face higher smoking-attributable morbidity and mortality compared to their non-Hispanic White counterparts [9-11]. Despite these health inequities in smoking and the rapid population growth of the Hispanic population in the United States, research on smoking cessation interventions for this group is sparse [12], with only 2 full-scale randomized controlled trials (RCTs) with 6-month follow-up or longer conducted to date [13,14].

While evidence-based smoking cessation treatments are generally within reach of the US general adult population [15,16], Hispanic adults often lack access to these treatments [17]. Reasons for these accessibility barriers include health providers being 50% less likely to offer quit-smoking advice to Hispanic adults compared to their non-Hispanic counterparts, which greatly contributes to significant inequities in smoking cessation efforts in this group [4,17]. Moreover, many Hispanic adults attempt to quit cigarette smoking cold turkey, which typically results in a very low likelihood of success of 2% to 5% [18,19]. Addressing these challenges requires research to develop culturally inclusive behavioral interventions that are within reach and tailored to the unique needs of Hispanic adults who smoke and want to quit smoking.

Free and remotely delivered digital interventions may improve access to evidence-based smoking cessation programs and thereby help improve smoking cessation outcomes [20]. Only 1 RCT, with a follow-up of 6 months or longer, has examined a digital smoking cessation intervention for Hispanic adults [21]. This study randomly assigned 457 English and Spanish-speaking Hispanic adults to receive either a culturally tailored mobile health smoking cessation intervention (Decídetexto) or standard of care, including written materials from the American Cancer Society and the National Cancer Institute (NCI), along with access to state-funded tobacco quitlines. Nicotine replacement therapies (NRTs; patches and gum) were offered to both arms at no cost. Results at 6 months showed higher (but nonsignificant) biochemically verified quit rates (Decídetexto: 33/229, 14.4% vs control group: 21/228, 9.2%; *P*=.09) and greater NRT use in the intervention group compared to standard of care (at least 1 day of NRT: 174/192, 90.6% vs 139/198, 70.2%; *P*=.01). However, it remains unclear if the use of pharmacotherapy to aid smoking cessation contributed to these quit rates, as both groups had access to free medication.

Digital interventions, like smartphone apps, offer an alternative approach to SMS text messaging or website-based interventions for smoking cessation [22], with over 500 smoking cessation apps currently available (R Nelson, personal communication, 2020). Although 91% of Hispanic adults in the United States own smartphones [23], to our knowledge, only 1 study has explored the potential efficacy of an app-based intervention for helping Hispanic adults quit smoking [24]. This study was a secondary analysis of a Hispanic adult subsample (n=210, 8.7% of the total sample) from a full-scale RCT comparing the iCanQuit app (Moby, Inc; based on acceptance and commitment therapy [ACT]) with the NCI's QuitGuide app (National Cancer Institute; based on US clinical practice guidelines [USCPGs]) for smoking cessation [24,25]. Results showed significantly higher 12-month quit rates in iCanQuit (28/82, 34.1%) versus QuitGuide (19/94, 20.2%; odds ratio [OR] 2.20, 95% CI 1.10-4.41; *P*=.02). This finding is significant, given the challenges faced by Hispanic adults in accessing evidence-based smoking cessation treatments, such as lack of medical insurance, ethnic discrimination, and cost [26,27].

Although app-based smoking cessation interventions hold promise, USCPGs emphasize that combining behavioral interventions with pharmacotherapy to aid smoking cessation

XSL•FO
RenderX

is more efficacious than either approach alone [28,29]. In the iCanQuit parent RCT, pharmacotherapy use was an exclusion criterion, but participants could opt to use it on their own after randomization [25]. In fact, both the iCanQuit and QuitGuide app–based interventions educate users on Food and Drug Administration (FDA)–approved pharmacotherapy to aid cessation [25]. In the full sample of the iCanQuit parent RCT, participants in the iCanQuit arm who used pharmacotherapy to aid smoking cessation on their own had significantly higher quit rates at 12 months compared to nonusers (users: 70/176, 39.8% vs nonusers: 192/691, 27.8%; *P*=.002) [30]. Whether the use of pharmacotherapy to aid cessation improved the quit rates of Hispanic participants, specifically, remains unknown.

Hispanic adults in the United States have the lowest rate of pharmacotherapy use to aid smoking cessation compared to other racial or ethnic groups [4,17,31]. For example, the rate of pharmacotherapy use is approximately 25.2% among non-Hispanic Black adults and 23.6% among non-Hispanic White adults but only 16.6% among Hispanic adults. In fact, reports from population-based studies of NRT use during quit attempts showed that an overwhelming 91% of Hispanic adults had never used NRT to aid their quit attempts [19]. There are various reasons why Hispanic adults may be more reluctant to use pharmacotherapy to aid their quit attempts. For instance, financial barriers and being medically uninsured (24% of US Hispanic adults) may play an important role [32,33]. Other reasons may include having a mistrust of the medical system [34] and having misconceptions about NRT use, such as NRT being addictive [31,33,35,36]. Prior research also highlights to a cultural emphasis on “voluntad propia” or the willpower to quit smoking that is unique to the Hispanic population [36]. In fact, when compared to non-Hispanic adults, Hispanic adults who smoke and have tried to quit smoking are twice as likely to cite willpower as their main reason for not trying pharmacotherapy to aid their quit attempts [19].

Objectives

To address existing knowledge gaps and build on prior studies [24,25], this secondary analysis used data from an RCT with long-term smoking cessation outcomes comparing 2 smoking cessation apps, iCanQuit (based on ACT) and QuitGuide (following USCPG) [25]. We explored the association between pharmacotherapy use and smoking cessation outcomes in the subsample of 173 English-speaking Hispanic adults recruited nationwide from 30 US states. Given the randomized design, we first tested the potential interaction of pharmacotherapy use and intervention arm on 12-month cigarette smoking abstinence. We then examined whether the use of any pharmacotherapy (ie, NRT, varenicline, or bupropion) augmented each app-based intervention efficacy to help Hispanic adults quit smoking. Finally, we compared NRT (eg, patch, gum, lozenge, and inhaler) users with nonusers, noting the scalability potential of mailed NRT for remote interventions compared to other smoking cessation medications.

Methods

Overview

The details of the iCanQuit RCT have been previously published [25]. Adults aged 18 years and older who reported smoking cigarettes every day, who had access to their own smartphones, and who had a desire to quit cigarette smoking within 30 days were eligible to participate. Exclusion criteria were an inability to read English, ongoing smoking cessation treatment including the use of pharmacotherapy at the time of enrollment only, prior use of iCanQuit or QuitGuide app–based interventions, or having a household member already enrolled in the trial. Participants were not excluded from the study if they chose to use pharmacotherapy to aid cessation on their own after randomization.

Ethical Considerations

All participants provided informed consent. Study procedures were approved by the Fred Hutchinson Cancer Center Institutional Review Board (IRB FHIRB0008317). Participants were compensated up to US $105 for completing data collection at the 3-, 6-, and 12-month follow-ups (US $35 per participant per follow-up). The data is de-identified.

Recruitment, Study Population, and Enrollment

Recruitment spanned from May 2017 to September 2018. Eligible participants were randomly assigned to iCanQuit or QuitGuide for 12 months. Follow-up data collection spanned from August 2017 to December 2019 via web-based surveys at the 3-, 6-, and 12-month follow-up after randomization. From the original 2415 adults in the iCanQuit parent RCT, a total of 173 Hispanic or Latino participants (173/2415, 7.2% of the total sample) who provided pharmacotherapy data at the 3-month follow-up were included in this paper’s secondary analysis (ie, pharmacotherapy data at the 3-month follow-up were missing for 37 Hispanic participants of 210 enrolled in the study).

Smartphone App–Based Interventions

The iCanQuit app, described in great detail elsewhere [25], comprises 8 ACT-based levels emphasizing two core processes of ACT: (1) acceptance of cravings to smoke, including physical sensations, emotions, and thoughts that cue smoking, and (2) aligning personal values, such as family, community, and health, with a smoke-free life. The “preparing to quit” phase fosters acceptance through mindfulness. It includes an “urge help” function and encourages users to track urges and cigarettes smoked daily. The “after you quit” phase focuses on motivation and relapse prevention.

QuitGuide, a free app developed by the NCI that is based on the USCPGs for smoking cessation [28], offers educational resources and techniques for quitting smoking and preventing relapse. It provides information on smoking cravings and quitting barriers. Unlike iCanQuit, which motivates through personal life values, QuitGuide focuses on health consequences and avoidance of cravings to smoke. Neither the content of iCanQuit nor QuitGuide was specifically designed for Hispanic adults who smoke. Neither arm provided pharmacotherapy to aid smoking cessation, coaching, or other interventions. Both

XSL•FO
RenderX

apps educated participants on FDA-approved medications to aid smoking cessation, highlighting benefits, where to obtain them, use instructions, and common side effects.

Measures

Baseline Characteristics

Participants completed baseline web-based surveys reporting sociodemographic, geographic, and smoking characteristics. Data collected included participants' self-reported age, sex, race, ethnicity, higher level of education attained, household annual income, employment status, marital status, sexual orientation, and household zip code [37-41]. Baseline smoking characteristics included daily cigarette consumption, cigarette dependence using the Fagerström Test for Cigarette Dependence [42], use of other nicotine-containing tobacco products including e-cigarettes, previous quit attempts, self-confidence in quitting, and friends and family who smoke. Alcohol consumption was assessed using the Quick Drinking Screen [43].

Pharmacotherapy Use

At the 3-month follow-up survey, participants were asked if they used NRTs or pharmacotherapy to aid smoking cessation since joining the study (ie, within the first 3 months following randomization). This time frame was selected because the content of each app is designed to be completed within approximately 6 weeks of initial use. Thus, assessing pharmacotherapy use by 3 months provides a relevant timeline to examine its potential synergistic effects with the content of the app-based interventions. Participants who reported using nicotine gum, patch, lozenge, or inhaler and varenicline or bupropion were classified as using pharmacotherapy. Participants selecting gum, patch, lozenge, or inhaler but not varenicline or bupropion were classified as using only NRT.

Smoking Abstinence

To assess cigarette smoking abstinence, participants were asked: "When was the last time you smoked, or even tried, a cigarette?" in the 12-month follow-up survey. Response choices were "earlier today," "24 hours ago," "2-7 days ago," "8-30 days ago," and "over 30 days ago" [25,44]. To be categorized as abstinence for the 30-day primary cessation outcomes, participants needed to report no cigarette consumption in the past 30 days at the time of the 12-month follow-up survey. The primary outcome was 30-day cigarette smoking abstinence analyzed through complete-case analysis and sensitivity analyses including missing-as-smoking and multiple imputation. Additional cessation outcomes included cessation of all nicotine and tobacco products (ie, cigarettes, e-cigarettes or vaping, chewing tobacco, snus, hookahs, cigars, cigarillos, tobacco pipes, and kreteks) and prolonged abstinence from cigarettes.

Statistical Analysis

Baseline characteristics were compared between groups by use of pharmacotherapy using Fisher exact tests for categorical variables and 2-sided *t* tests for continuous variables. To determine whether pharmacotherapy use or use of NRT alone by 3 months differentially impacted the effect of the app-based intervention arm on the primary 12-month cessation outcome, we used logistic regression models with a pharmacotherapy use or NRT use by intervention interaction term. Logistic regression models were then used to estimate the effect of pharmacotherapy use or NRT use on 12-month cessation, separately for each intervention arm [44]. For multiple imputation of missing 12-month smoking status, we used multivariate imputation by chained equations in the R package *mice* [45]. All models were adjusted for randomization stratification factors from the parent RCT: smoking frequency, education, and depression. Statistical analyses were carried out using R software (version 4.2.3; R Foundation for Statistical Computing), with 2-sided tests and α=.05.

Results

Baseline Characteristics

Hispanic participants were recruited from 30 US states, averaging 34.5 years in age, with 50.9% (88/173) being female; 20.2% (35/173) identifying as sexual or gender minority; and 30.6% (53/173) being married or partnered. Regarding race, 59% (102/173) identified as White, 15% (26/173) identified as multiracial, and 13.9% (24/173) did not report their race (Table 1). Regarding education, employment, and income, 46.8% (81/173) had a high school diploma or lower; 45.7% (79/173) were unemployed, disabled, or out of the labor force; and 42.8% (35/173) had annual household incomes of US $20,000 or less. More than half (94/173, 54.7%) screened positive for depression symptoms, 29.8% (51/173) for generalized anxiety, 31.8% (54/173) for panic disorder, 48.2% (82/173) for posttraumatic stress disorder, and 34.5% (59/173) for social anxiety disorder. Additionally, 19.1% (33/173) reported bipolar disorder or schizophrenia. On average, participants reported smoking an average of 16.2 (SD 15.9) cigarettes per day, 46.8% (81/173) had high cigarette dependence, 77.5% (134/173) smoked for at least 10 years, and 24.3% (42/173) were dual users of combustible cigarettes and e-cigarettes.

XSL•FO
RenderX

Table 1. Baseline characteristics of Hispanic participants.

Characteristic	Participants, n	Overall (N=173)	Did not use pharmacotherapy (n=135, 78%)	Used pharmacotherapy (n=38, 22%)	*P* value
Age (years), mean (SD)	173	34.5 (9.3)	34.2 (9.5)	35.5 (8.5)	.44
Sex, n (%)	173				.24
Female		88 (50.9)	68 (50.4)	20 (52.6)	
Male		82 (47.4)	65 (48.1)	17 (44.7)	
Transgender female		1 (0.6)	0 (0)	1 (2.6)	
Transgender male		2 (1.2)	2 (1.5)	0 (0)	
Race, n (%)	173				.19
American Indian or Alaska Native		8 (4.6)	8 (5.9)	0 (0)	
Black or African American		13 (7.5)	8 (5.9)	5 (13.2)	
Multiracial		26 (15)	21 (15.6)	5 (13.2)	
White		102 (59)	77 (57)	25 (65.8)	
Unknown (unreported) race		24 (13.9)	21 (15.6)	3 (7.9)	
LGBT[a], n (%)	173	35 (20.2)	28 (20.7)	7 (18.4)	.93
Education, n (%)	173				.93
GED[b], high school or less education		81 (46.8)	63 (46.7)	18 (47.4)	
Some college, no degree		63 (36.4)	50 (37)	13 (34.2)	
College degree or higher		29 (16.8)	22 (16.3)	7 (18.4)	
Employment status, n (%)	173				.04
Employed		94 (54.3)	68 (50.4)	26 (68.4)	
Unemployed		28 (16.2)	27 (20)	1 (2.6)	
Disabled		25 (14.5)	18 (13.3)	7 (18.4)	
Out of labor force		26 (15)	22 (16.3)	4 (10.5)	
Income (US $), n (%)	173				.17
Less than $20,000 per year		74 (42.8)	62 (45.9)	12 (31.6)	
$20,000-$54,999 per year		77 (44.5)	55 (40.7)	22 (57.9)	
$55,000 per year or more		22 (12.7)	18 (13.3)	4 (10.5)	
Married, n (%)	173	53 (30.6)	43 (31.9)	10 (26.3)	.65
Rural residence[c], n (%)	173	21 (12.1)	19 (14.1)	2 (5.3)	.23
Mental health positive screening, n (%)					
Depression (CES-D)[d]	172	94 (54.7)	70 (52.2)	24 (63.2)	.31
General anxiety (GAD-7)[e]	171	51 (29.8)	36 (27.1)	15 (39.5)	.20
Panic disorder (ANSQ)[f]	170	54 (31.8)	41 (31.1)	13 (34.2)	.87
PTSD (PCL-6)[g]	170	82 (48.2)	60 (45.5)	22 (57.9)	.24
Social anxiety (mini-SPIN)[h]	171	59 (34.5)	40 (30.1)	19 (50)	.04
Self-reported bipolar or schizophrenia[i], n (%)	173	33 (19.1)	26 (19.3)	7 (18.4)	>.99
Alcohol use, n (%)					
Drinks per drinking day, mean (SD)	166	2.3 (4.0)	2.3 (4.2)	2.4 (3.3)	.90
Heavy drinking[j], n (%)	166	29 (17.5)	22 (16.9)	7 (19.4)	.92
Smoking behavior					

XSL•FO
RenderX

Characteristic	Participants, n	Overall (N=173)	Did not use pharmacotherapy (n=135, 78%)	Used pharmacotherapy (n=38, 22%)	*P* value
Cigarettes smoked per day, mean (SD)	173	16.2 (15.9)	15.9 (16.2)	17.2 (15.1)	.67
FTCD score[k], mean (SD)	173	5.2 (2.0)	5.3 (1.9)	5.2 (2.5)	.97
High nicotine dependence (FTCD≥6), n (%)	173	81 (46.8)	60 (44.4)	21 (55.3)	.32
First cigarette within 5 minutes of waking, n (%)	173	75 (43.4)	58 (43)	17 (44.7)	.99
Smokes greater than one-half pack (10 cigarettes) per day, n (%)	173	97 (56.1)	72 (53.3)	25 (65.8)	.24
Smokes more than 1 pack (20 cigarettes) per day, n (%)	173	18 (10.4)	13 (9.6)	5 (13.2)	.74
Smoking for 10 or more years, n (%)	173	134 (77.5)	100 (74.1)	34 (89.5)	.07
Used e-cigarettes in the past month, n (%)	173	42 (24.3)	35 (25.9)	7 (18.4)	.46
Quit attempts in the past 12 months, mean (SD)	166	3.3 (15.8)	3.8 (17.8)	1.8 (3.2)	.51
Confidence to quit smoking[l], mean (SD)	173	67.9 (28.1)	68.0 (27.6)	67.8 (29.9)	.98
Friend and partner smoking					
Close friends who smoke, mean (SD)	173	2.8 (1.7)	2.7 (1.7)	2.9 (1.8)	.48
Housemates who smoke, mean (SD)	173	1.5 (1.1)	1.5 (1.1)	1.5 (0.8)	.96
Living with a partner who smokes, n (%)	173	56 (32.4)	44 (32.6)	12 (31.6)	>.99

[a]LGBT: lesbian, gay, bisexual, or transgender.

[b]GED: general education development.

[c]To ascertain the residential classification of participants into urban or rural areas within the United States, we used the *zipcode* R library [46] to link participants' zip codes to geographic locations. The classification was based on rural-urban commuting area codes at the subcounty level [37-41].

[d]Positive screening results for depression via Center for Epidemiologic Studies Depression Scale (CES-D)-20; threshold ≥16 [47].

[e]Positive screening results for generalized anxiety via General Anxiety Disorder (GAD)-7; threshold ≥10 [48].

[f]Positive screening results for panic disorder via Autonomic Nervous System Questionnaire (ANSQ) [49]. A positive screening was recorded for individuals reporting at least 1 panic attack in the past month, provided that at least 1 of these incidents occurred in a situation in which they were not in danger or the center of attention.

[g]Positive screening results for posttraumatic stress disorder (PTSD) via Post-Traumatic Checklist (PCL-6); threshold ≥14 [50].

[h]Positive screening results for social anxiety via Mini-Social Phobia Inventory (mini-SPIN); threshold ≥6 [51].

[i]Presence of bipolar disorder or schizophrenia was self-reported via question: "Do you have any of the following mental health disorders?" Multiselect response options included anxiety disorder, depression disorder, bipolar disorder, schizophrenia, alcohol abuse disorder, drug abuse disorder, or none of the above [52,53].

[j]Heavy drinking is defined as 4 or more drinks on a typical drinking day for women and 5 or more drinks on a typical drinking day for men within the past 30 days [43].

[k]Fagerström Test for Cigarette Dependence (FTCD) score [42].

[l]Range 0-100, where 0 indicates not at all confident and 100 indicates extremely confident.

Pharmacotherapy Use and Outcome Retention Rates

At the 3-month follow-up, 22% (38/173) of participants used pharmacotherapy (NRT, varenicline, or bupropion). There was no statistically significant difference in pharmacotherapy use between treatment arms (iCanQuit: 16/82, 19.5% vs QuitGuide: 22/91, 24.2%; *P*=.41). Approximately 73.6% (28/38) of participants who reported pharmacotherapy use opted for NRT only (ie, nicotine patch, gum, inhaler, or lozenge), whereas 26.3% (10/38) reported using a combination of NRT and medication (varenicline or bupropion). Outcome data retention at 12 months was high, with an overall rate of 93.6% (162/173). This rate did not differ by intervention arms (iCanQuit: 74/82, 90.2% vs QuitGuide: 88/91, 96.7%; *P*=.09) and pharmacotherapy use (users: 36/38, 94.7% vs nonusers: 126/135, 93.3%; *P*=.82).

Smoking Abstinence

Report of any FDA-approved pharmacotherapy use by 3 months had a marginal differential impact on 12-month quit rates by intervention arm (*P* for interaction=.053). In the iCanQuit arm, there was a signal indicating higher odds of quitting with pharmacotherapy. Specifically, 12-month cigarette smoking abstinence rates in the iCanQuit arm were 43.8% (7/16) for pharmacotherapy users versus 28.8% (19/66) for nonusers (OR 2.21, 95% CI 0.66-7.48; *P*=.20). In contrast, cigarette smoking abstinence rates in the QuitGuide arm were 9.1% (2/22) for pharmacotherapy users versus 21.7% (15/69) for nonusers (OR 0.36, 95% CI 0.07-1.72; *P*=.20). Similar trends were observed

XSL•FO
RenderX

for other secondary smoking cessation outcomes, but none reached statistical significance (Table 2).

The impact of using pharmacotherapy to aid cessation at 12 months was more pronounced when NRT was the sole form of pharmacotherapy used, although the interaction effect between NRT use and app-based intervention arm on the 12-month smoking cessation outcome was not statistically significant (P=.09). In the iCanQuit arm, 30-day cigarette smoking abstinence rates at 12 months were 45.5% (5/11) for NRT users versus 28.8% (20/69) for nonusers (OR 2.24, 95% CI 0.54-9.27; P=.26). In contrast, in the QuitGuide arm, 30-day cigarette smoking abstinence rates at 12 months were 11.8% (2/17) for NRT users versus 21.7% (15/69) for nonusers (OR 0.49, 95% CI 0.10-2.45; P=.36; Table 3).

Table 2. Association between any pharmacotherapy use and cessation outcomes in the iCanQuit and QuitGuide treatment arms among Hispanic participants.

12-Month cessation outcome	Interaction between treatment arm and use of pharmacotherapy		Treatment arm	Participants, n	Overall, n (%)	Did not use pharmacotherapy, n (%)	Used pharmacotherapy, n (%)	OR[a] (95% CI)[b]	*P* value
	β (SE)	*P* value							
30-Day PPA[c] from cigarettes, complete case	1.97 (1.02)	.053	iCanQuit	74	26 (35.1)	19 (31.7)	7 (50)	2.59 (0.69-9.73)	.16
30-Day PPA from cigarettes, complete case	1.97 (1.02)	.053	QuitGuide	88	17 (19.3)	15 (22.7)	2 (9.1)	0.34 (0.07-1.64)	.18
30-Day PPA from cigarettes, multiple imputation	1.86 (1.02)	.07	iCanQuit	820	290 (35.4)	214 (32.4)	76 (47.5)	2.27 (0.61-8.41)	.23
30-Day PPA from cigarettes, multiple imputation	1.86 (1.02)	.07	QuitGuide	910	177 (19.5)	157 (22.8)	20 (9.1)	0.34 (0.07-1.64)	.18
30-Day PPA from cigarettes, missing-as-smoking	1.79 (0.99)	.07	iCanQuit	82	26 (31.7)	19 (28.8)	7 (43.8)	2.21 (0.66-7.48)	.20
30-Day PPA from cigarettes, missing-as-smoking	1.79 (0.99)	.07	QuitGuide	91	17 (18.7)	15 (21.7)	2 (9.1)	0.36 (0.07-1.72)	.20
30-Day PPA from all nicotine and tobacco products[d]	1.90 (1.03)	.06	iCanQuit	74	21 (28.4)	15 (25)	6 (42.9)	2.79 (0.72-10.77)	.14
30-Day PPA from all nicotine and tobacco products[d]	1.90 (1.03)	.06	QuitGuide	88	16 (18.2)	14 (21.2)	2 (9.1)	0.37 (0.08-1.81)	.22
Prolonged abstinence from cigarettes[e]	1.06 (1.47)	.64	iCanQuit	54	10 (18.5)	8 (17.4)	2 (25)	1.67 (0.27-10.36)	.58
Prolonged abstinence from cigarettes[e]	1.06 (1.47)	.64	QuitGuide	69	6 (8.7)	5 (9.6)	1 (5.9)	0.58 (0.06-5.49)	.64

[a]OR: odds ratio.

[b]OR for use versus no use of pharmacotherapy by 3 months after randomization. All models were adjusted for factors used in stratified randomization: smoking >20 cigarettes per day, high school or lower education, and positive screening for depression.

[c]PPA: point prevalence abstinence.

[d]Including any kind of e-cigarettes or vaping, chewing tobacco, snus, hookahs, cigars, cigarillos, tobacco pipes, and kreteks.

[e]Prolonged abstinence is defined as no smoking since 3 months after randomization, using the self-reported date of the last cigarette.

XSL•FO
RenderX

Table 3. Association between use of nicotine replacement therapy (NRT) only and cessation outcomes in the iCanQuit and QuitGuide treatment arms among Hispanic participants.

12-Month cessation outcome	Interaction between treatment arm and use of NRT,		Treatment arm	Participants, n	Overall, n (%)	Did not use NRT, n (%)	Used NRT, n (%)	OR[a] (95% CI)[b]	*P* value
	β (SE)	*P* value							
30-Day PPA[c] from cigarettes, complete case	1.88 (1.11)	.09	iCanQuit	69	24 (34.8)	19 (31.7)	5 (55.6)	3.09 (0.62-15.41)	.17
30-Day PPA from cigarettes, complete case	1.88 (1.11)	.09	QuitGuide	83	17 (20.5)	15 (22.7)	2 (11.8)	0.47 (0.09-2.33)	.36
30-Day PPA from cigarettes, multiple imputation	1.69 (1.11)	.13	iCanQuit	770	270 (35.1)	214 (32.4)	56 (50.9)	2.46 (0.51-11.90)	.27
30-Day PPA from cigarettes, multiple imputation	1.69 (1.11)	.13	QuitGuide	860	177 (20.6)	157 (22.8)	20 (11.8)	0.47 (0.10-2.35)	.36
30-Day PPA from cigarettes, missing-as-smoking	1.56 (1.06)	.14	iCanQuit	77	24 (31.2)	19 (28.8)	5 (45.5)	2.24 (0.54-9.27)	.26
30-Day PPA from cigarettes, missing-as-smoking	1.56 (1.06)	.14	QuitGuide	86	17 (19.8)	15 (21.7)	2 (11.8)	0.49 (0.10-2.45)	.39
30-Day PPA from all nicotine and tobacco products[d]	1.66 (1.11)	.14	iCanQuit	69	19 (27.5)	15 (25)	4 (44.4)	2.73 (0.54-13.7)	.22
30-Day PPA from all nicotine and tobacco products[d]	1.66 (1.11)	.14	QuitGuide	83	16 (19.3)	14 (21.2)	2 (11.8)	0.51 (0.10-2.54)	.41
Prolonged abstinence from cigarettes[e]	0.29 (1.68)	.91	iCanQuit	51	9 (17.6)	8 (17.4)	1 (20)	1.20 (0.11-12.87)	.88
Prolonged abstinence from cigarettes[e]	0.29 (1.68)	.91	QuitGuide	64	6 (9.4)	5 (9.6)	1 (8.3)	0.90 (0.09-9.10)	.93

[a]OR: odds ratio.

[b]OR for use versus no use of NRT by 3 months after randomization. All models were adjusted for factors used in stratified randomization: smoking >20 cigarettes per day, high school or lower education, and positive screening for depression.

[c]PPA: point prevalence abstinence.

[d]Including any kind of e-cigarettes or vaping, chewing tobacco, snus, hookahs, cigars, cigarillos, tobacco pipes, and kreteks.

[e]Prolonged abstinence is defined as no smoking since 3 months after randomization, using the self-reported date of the last cigarette.

Discussion

Principal Findings

This paper aimed to address the critical knowledge gaps regarding the efficacy of smoking cessation app–based interventions combined with pharmacotherapy to help individuals quit cigarette smoking, particularly among Hispanic adults who tends to underuse evidence-based smoking cessation treatments, including pharmacotherapy [17,25,26,54]. The findings suggest that self-selected use of pharmacotherapy, including NRT, varenicline, or bupropion, improved 12-month quit rates among participants in the iCanQuit arm (users: 7/14, 50% vs nonusers: 19/60, 31.7%). However, this effect was not observed among participants in the QuitGuide arm (users: 2/22, 9.1% vs nonusers: 15/66, 22.7%).

The differential effect of using pharmacotherapy between participants in the iCanQuit and QuitGuide arms suggests that combining pharmacotherapy with an ACT-based intervention (iCanQuit) may have synergistically helped participants manage smoking cravings during quit attempts. The results suggest a combined and synergetic effect of managing short- and long-term withdrawal symptoms, resulting in a more effective treatment than ACT-based behavioral interventions alone. It is plausible that short-term withdrawal symptoms were ameliorated using pharmacotherapy, while the skills using iCanQuit to manage withdrawal symptoms may have helped in the long term. While the QuitGuide app–based intervention also mentions FDA-approved pharmacotherapy to aid cessation, the app usually portrays medications as separate rather than complementary to learning behavioral skills to quit smoking

XSL•FO
RenderX

[25,28]. It is then possible that participants in the QuitGuide arm may have not understood the potential complementary role of using pharmacotherapy to aid smoking cessation alongside the app. It is plausible that guidance is needed when pharmacotherapy is offered alongside behavioral interventions to have the desired synergetic effect [28]. Further investigation is warranted to understand the improved quit rates with pharmacotherapy in iCanQuit but not QuitGuide.

It is important to note that Hispanic adults in this study opted independently to use pharmacotherapy to aid smoking cessation, albeit at a nonsignificantly lower rate than non-Hispanic participants (38/173, 22% vs 581/1915, 30.3%; *P*=.33). This is contrary to population-based reports among US Hispanic adults being 50% less likely to use evidence-based smoking cessation treatments compared to the general population of adults who smoke in the United States [17,26,54]. Our results showed the potential of evidence-based smoking cessation interventions that are delivered remotely in improving awareness of FDA-approved pharmacotherapy to aid cessation.

Limitations

This study has several limitations. First, the reliance on secondary data inherently limits findings, as participants were not randomized to receive pharmacotherapy, resulting in associations rather than causal effects. Second, details on dosage, adherence, and duration of pharmacotherapy use are unknown; thus, the nature of pharmacotherapy usage remains uncertain. Third, Hispanic participants were English speakers who smoked daily, which limits the generalizability of the findings to Spanish-speaking Hispanic adults, or those who smoke less than daily [55]. Finally, reliance on self-reported use of pharmacotherapy and smoking abstinence introduces subjectivity and reporting bias. Future multimethod assessment protocols for smoking behavior would therefore be important.

This study also had several strengths. First, it is the first study to explore the effect of an app-based smoking cessation intervention combined with pharmacotherapy use as a potential treatment for Hispanic adults who smoke and wish to quit [17,26,54]. Second, the trial's long-term smoking cessation outcomes (12 months) and high retention rates that did not differ by intervention arm or pharmacotherapy use helped reduce any potential bias in the complete-case analysis and missing-as-smoking sensitivity analyses. Third, despite the modest sample size, Hispanic participants were recruited from 30 US states, underscoring the geographic diversity of the results and the potential of app-based interventions for widespread reach and broad dissemination.

The results from this secondary analysis have significant public health implications for designing future interventions to help Hispanic adults quit smoking. First, app-based smoking cessation interventions offer a viable approach for hard-to-reach population who are lacking access to evidence-based treatments, including 24% of uninsured Hispanic adults in the United States [32]. Second, app-based delivered interventions could help overcome accessibility barriers like transportation and time constraints. Third, our results show a notable increase in quit rates when using NRT alongside iCanQuit compared to nonuse (5/9, 55.6% vs 20/63, 31.7%). This is noteworthy, as over-the-counter NRT has shown high reach and efficacy [16,29]. Additionally, NRT is readily available and cost-effective and requires minimal medical oversight compared to other medications (varenicline or bupropion) [56,57]. Mailed NRT [58,59], as seen in previous trials, could hold potential for broad dissemination and scalability of smoking cessation programs at the population level. Moreover, future interventions for the Hispanic population could be offered in Spanish, expanding their reach to the 28% of Spanish-speaking Hispanic adults currently living in the United States [32].

Conclusions

This secondary analysis underscores the nuanced role of pharmacotherapy in enhancing the efficacy of app-based smoking cessation interventions among Hispanic adults. Pharmacotherapy use, particularly when paired with an ACT-based app (iCanQuit), appears to improve quit rates, highlighting a potential pathway to address disparities in smoking cessation outcomes. In contrast, the USCPG-based app (QuitGuide) did not result in a similar synergistic effect when combined with pharmacotherapy. Considering the scalability of NRT, future research should evaluate the efficacy and cost-effectiveness of remotely providing NRT alongside iCanQuit compared to iCanQuit alone.

Acknowledgments

The authors appreciate the tireless contributions of the entire study staff, most notably Eric Meier; Carolyn Ehret; Brianna Sullivan; the design services of Ayogo Inc; and the development services of Moby, Inc. The authors are very appreciative of the study participants. The study was funded by grant R01CA192849 awarded to JBB from the National Cancer Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

JBB, MS-T, and KEM conceptualized the study. MS-T led the manuscript writing. KEM led and conducted the data analysis. All authors assisted in manuscript writing and provided critical review. All authors have read and agreed to the published versions of the manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 966 KB - formative_v9i1e69311_app1.pdf]

References

1. Jones N, Marks R, Ramirez R, Ríos-Vargas M. Racial and ethnic diversity in the United States: 2010 census and 2020 census. U.S. Census Bureau. 2020. URL: https://www.census.gov/library/visualizations/interactive/racial-and-ethnic-diversity-in-the-united-states-2010-and-2020-census.html [accessed 2024-06-15]
2. Who is Hispanic? Pew Research Center. 2022. URL: https://www.pewresearch.org/fact-tank/2022/09/15/who-is-hispanic/ [accessed 2022-10-10]
3. Hispanic population to reach 111 million by 2060. U.S. Census Bureau. URL: https://www.census.gov/library/visualizations/2018/comm/hispanic-projected-pop.html [accessed 2018-10-09]
4. Babb S, Malarcher A, Schauer G, Asman K, Jamal A. Quitting smoking among adults—United States, 2000–2015. MMWR Morb Mortal Wkly Rep 2017;65(52):1457-1464 [FREE Full text] [doi: 10.15585/mmwr.mm6552a1] [Medline: 28056007]
5. Cornelius ME, Loretan CG, Jamal A, Davis Lynn BC, Mayer M, Alcantara IC, et al. Tobacco product use among adults—United States, 2021. MMWR Morb Mortal Wkly Rep 2023;72(18):475-483 [FREE Full text] [doi: 10.15585/mmwr.mm7218a1] [Medline: 37141154]
6. Kaplan RC, Bangdiwala SI, Barnhart JM, Castañeda SF, Gellman MD, Lee DJ, et al. Smoking among U.S. Hispanic/Latino adults: the Hispanic community health study/study of Latinos. Am J Prev Med 2014;46(5):496-506 [FREE Full text] [doi: 10.1016/j.amepre.2014.01.014] [Medline: 24745640]
7. Merzel CR, Isasi CR, Strizich G, Castañeda SF, Gellman M, Maisonet Giachello AL, et al. Smoking cessation among U.S. Hispanic/Latino adults: findings from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). Prev Med 2015;81:412-419 [FREE Full text] [doi: 10.1016/j.ypmed.2015.10.006] [Medline: 26515291]
8. Sisti JS, Jasek JP, Farley SM. Heterogeneity in current cigarette smoking among Hispanic/Latino Heritage Groups in New York City, 2003-2016. Ethn Dis 2020;30(1):97-108 [FREE Full text] [doi: 10.18865/ed.30.1.97] [Medline: 31969789]
9. Heron M. Deaths: leading causes for 2016. Natl Vital Stat Rep 2018;67(6):1-77. [doi: 10.15620/cdc:104186]
10. Henley SJ, Thomas CS, Sharapova SR, Momin B, Massetti GM, Winn DM, et al. Vital signs: disparities in tobacco-related cancer incidence and mortality—United States, 2004-2013. MMWR Morb Mortal Wkly Rep 2016;65(44):1212-1218 [FREE Full text] [doi: 10.15585/mmwr.mm6544a3] [Medline: 27832048]
11. Daviglus ML, Talavera GA, Avilés-Santa ML, Allison M, Cai J, Criqui MH, et al. Prevalence of major cardiovascular risk factors and cardiovascular diseases among Hispanic/Latino individuals of diverse backgrounds in the United States. JAMA 2012;308(17):1775-1784 [FREE Full text] [doi: 10.1001/jama.2012.14517] [Medline: 23117778]
12. Webb MS, Rodríguez-Esquivel D, Baker EA. Smoking cessation interventions among Hispanics in the United States: a systematic review and mini meta-analysis. Am J Health Promot 2010;25(2):109-118. [doi: 10.4278/ajhp.090123-lit-25]
13. Simmons VN, Sutton SK, Medina-Ramirez P, Martinez U, Brandon KO, Byrne MM, et al. Self-help smoking cessation intervention for Spanish-speaking Hispanics/Latinxs in the United States: a randomized controlled trial. Cancer 2022;128(5):984-994 [FREE Full text] [doi: 10.1002/cncr.33986] [Medline: 34679188]
14. Cartujano-Barrera F, Arana-Chicas E, Catley D, Cox LS, Diaz FJ, Ellerbeck EF, et al. Decídetexto: mobile cessation support for Latino smokers. Study protocol for a randomized clinical trial. Contemp Clin Trials 2020;99:106188 [FREE Full text] [doi: 10.1016/j.cct.2020.106188] [Medline: 33080379]
15. Smoking cessation: a report of the Surgeon General—executive summary. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. 2020. URL: https://www.hhs.gov/sites/default/files/2020-cessation-sgr-executive-summary.pdf [accessed 2020-02-23]
16. Patnode CD, Henderson JT, Coppola EL, Melnikow J, Durbin S, Thomas RG. Interventions for tobacco cessation in adults, including pregnant persons: updated evidence report and systematic review for the US Preventive Services Task Force. JAMA 2021;325(3):280-298. [doi: 10.1001/jama.2020.23541] [Medline: 33464342]
17. Babb S, Malarcher A, Asman K, Johns M, Caraballo R, VanFrank B, et al. Disparities in cessation behaviors between Hispanic and non-Hispanic White adult cigarette smokers in the United States, 2000-2015. Prev Chronic Dis 2020;17:E10 [FREE Full text] [doi: 10.5888/pcd17.190279] [Medline: 31999539]
18. Prochaska JJ, Benowitz NL. The past, present, and future of nicotine addiction therapy. Annu Rev Med 2016;67(1):467-486 [FREE Full text] [doi: 10.1146/annurev-med-111314-033712] [Medline: 26332005]
19. Cook-Shimanek M, Burns EK, Levinson AH. Medicinal nicotine nonuse: smokers' rationales for past behavior and intentions to try medicinal nicotine in a future quit attempt. Nicotine Tob Res 2013;15(11):1926-1933 [FREE Full text] [doi: 10.1093/ntr/ntt085] [Medline: 23817584]

XSL•FO
RenderX

20. Bold KW, Garrison KA, DeLucia A, Horvath M, Nguyen M, Camacho E, et al. Smartphone apps for smoking cessation: systematic framework for app review and analysis. J Med Internet Res 2023;25:e45183 [FREE Full text] [doi: 10.2196/45183] [Medline: 37440305]
21. Cartujano-Barrera F, Cox LS, Catley D, Cai X, Diaz FJ, Arana-Chicas E, et al. Decídetexto: mobile cessation support for Latino adults who smoke: a randomized clinical trial. Chest 2024:04904-3 [FREE Full text] [doi: 10.1016/j.chest.2024.07.160] [Medline: 39134144]
22. Wasil AR, Gillespie S, Shingleton R, Wilks CR, Weisz JR. Examining the reach of smartphone apps for depression and anxiety. Am J Psychiatry 2020;177(5):464-465. [doi: 10.1176/appi.ajp.2019.19090905] [Medline: 32354266]
23. Mobile fact sheets: internet and broadband. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2025-01-04]
24. Santiago-Torres M, Mull KE, Sullivan BM, Zvolensky MJ, Bricker JB. Can a smartphone application help Hispanic/Latinx adults quit smoking? A randomized trial secondary analysis. J Contextual Behav Sci 2022;26:261-270 [FREE Full text] [doi: 10.1016/j.jcbs.2022.11.001] [Medline: 36437818]
25. Bricker JB, Watson NL, Mull KE, Sullivan BM, Heffner JL. Efficacy of smartphone applications for smoking cessation: a randomized clinical trial. JAMA Intern Med 2020;180(11):1472. [doi: 10.1001/jamainternmed.2020.4055]
26. Li L, Zhan S, Hu L, Wilson KM, Mazumdar M, Liu B. Examining the role of healthcare access in racial/ethnic disparities in receipt of provider-patient discussions about smoking: a latent class analysis. Prev Med 2021;148:106584. [doi: 10.1016/j.ypmed.2021.106584] [Medline: 33930432]
27. Cohen RA, Cha AE, Terlizzi EP, Martinez ME. Demographic variation in health insurance coverage: United States. Natl Health Stat Report 2019;2021(159):1-15. [doi: 10.15620/cdc:106462]
28. Clinical Practice Guideline Treating Tobacco Use and Dependence 2008 Update Panel, Liaisons, and Staff. A clinical practice guideline for treating tobacco use and dependence: 2008 update. A U.S. Public Health Service report. Am J Prev Med 2008;35(2):158-176 [FREE Full text] [doi: 10.1016/j.amepre.2008.04.009] [Medline: 18617085]
29. Hartmann-Boyce J, Hong B, Livingstone-Banks J, Wheat H, Fanshawe TR. Additional behavioural support as an adjunct to pharmacotherapy for smoking cessation. Cochrane Database Syst Rev 2019;6(6):CD009670 [FREE Full text] [doi: 10.1002/14651858.CD009670.pub4] [Medline: 31166007]
30. Bricker JB, Santiago-Torres M, Mull KE, Sullivan BM, David SP, Schmitz J, et al. Do medications increase the efficacy of digital interventions for smoking cessation? Secondary results from the iCanQuit randomized trial. Addiction 2024;119(4):664-676. [doi: 10.1111/add.16396] [Medline: 38009551]
31. Levinson AH, Pérez-Stable EJ, Espinoza P, Flores ET, Byers TE. Latinos report less use of pharmaceutical aids when trying to quit smoking. Am J Prev Med 2004;26(2):105-111. [doi: 10.1016/j.amepre.2003.10.012] [Medline: 14751320]
32. 11 Facts about Hispanic origin groups in the U.S. Pew Research Center. 2023. URL: https://tinyurl.com/2ukr5hyp [accessed 2024-01-31]
33. Levinson AH, Borrayo EA, Espinoza P, Flores ET, Pérez-Stable EJ. An exploration of Latino smokers and the use of pharmaceutical aids. Am J Prev Med 2006;31(2):167-171. [doi: 10.1016/j.amepre.2006.03.022] [Medline: 16829334]
34. Alexander AC, Hébert ET, Businelle MS, Kendzor DE. Everyday discrimination indirectly influences smoking cessation through post-quit self-efficacy. Drug Alcohol Depend 2019;198:63-69. [doi: 10.1016/j.drugalcdep.2019.01.033] [Medline: 30878768]
35. Pacek LR, McClernon FJ, Bosworth HB. Adherence to pharmacological smoking cessation interventions: a literature review and synthesis of correlates and barriers. Nicotine Tob Res 2018;20(10):1163-1172 [FREE Full text] [doi: 10.1093/ntr/ntx210] [Medline: 29059394]
36. Carpenter MJ, Ford ME, Cartmell K, Alberg AJ. Misperceptions of nicotine replacement therapy within racially and ethnically diverse smokers. J Natl Med Assoc 2011;103(9-10):885-894 [FREE Full text] [doi: 10.1016/s0027-9684(15)30444-2] [Medline: 22364057]
37. Larson EH, Andrilla CHA, Garberson LA, Evans DV. Geographic access to health care for rural medicare beneficiaries in five states: an update. Policy brief. WWAMI Rural Health Research Center, University of Washington. URL: https://familymedicine.uw.edu/rhrc/wp-content/uploads/sites/4/2021/09/RHRC_PBSEP2021_LARSON.pdf [accessed 2025-01-04]
38. Ratcliffe M, Burd C, Holder K, Fields A. Understanding and using American Community Survey data. U.S. Census Bureau. 2016. URL: https://www.census.gov/content/dam/Census/library/publications/2019/acs/ACS_rural_handbook_2019_ch01.pdf [accessed 2024-07-07]
39. Brooks MM, Mueller JT, Thiede BC. County reclassifications and rural-urban mortality disparities in the United States (1970-2018). Am J Public Health 2020;110(12):1814-1816. [doi: 10.2105/AJPH.2020.305895] [Medline: 33058708]
40. Chen X, Orom H, Hay JL, Waters EA, Schofield E, Li Y, et al. Differences in rural and urban health information access and use. J Rural Health 2019;35(3):405-417 [FREE Full text] [doi: 10.1111/jrh.12335] [Medline: 30444935]
41. Unger JM, Moseley A, Symington B, Chavez-MacGregor M, Ramsey SD, Hershman DL. Geographic distribution and survival outcomes for rural patients with cancer treated in clinical trials. JAMA Netw Open 2018;1(4):e181235 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.1235] [Medline: 30646114]

XSL•FO
RenderX

42. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991;86(9):1119-1127. [doi: 10.1111/j.1360-0443.1991.tb01879.x] [Medline: 1932883]
43. Roy M, Dum M, Sobell LC, Sobell MB, Simco ER, Manor H, et al. Comparison of the quick drinking screen and the alcohol timeline followback with outpatient alcohol abusers. Subst Use Misuse 2008;43(14):2116-2223. [doi: 10.1080/10826080802347586] [Medline: 18825590]
44. Santiago-Torres M, Mull KE, Sullivan BM, Bricker JB. Relative efficacy of an acceptance and commitment therapy-based smartphone app with a standard US Clinical Practice Guidelines-based app for smoking cessation in dual users of combustible and electronic cigarettes: secondary findings from a randomized trial. Subst Use Misuse 2024;59(4):591-600. [doi: 10.1080/10826084.2023.2293732] [Medline: 38098199]
45. van Buuren B, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw 2011;45(3):1-67. [doi: 10.32614/cran.package.mice]
46. Breen J. zipcode: U.S. ZIP Code database for geocoding. R package version 1.0. R Project. URL: https://CRAN.R-project.org/package=zipcode [accessed 2012-03-12]
47. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977;1(3):385-401. [doi: 10.1177/014662167700100306]
48. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
49. Stein MB, Roy-Byrne PP, McQuaid JR, Laffaye C, Russo J, McCahill ME, et al. Development of a brief diagnostic screen for panic disorder in primary care. Psychosom Med 1999;61(3):359-364. [doi: 10.1097/00006842-199905000-00016] [Medline: 10367617]
50. Lang AJ, Wilkins K, Roy-Byrne PP, Golinelli D, Chavira D, Sherbourne C, et al. Abbreviated PTSD Checklist (PCL) as a guide to clinical response. Gen Hosp Psychiatry 2012;34(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2012.02.003] [Medline: 22460001]
51. Connor KM, Kobak KA, Churchill LE, Katzelnick D, Davidson JRT. Mini-SPIN: a brief screening assessment for generalized social anxiety disorder. Depress Anxiety 2001;14(2):137-140. [doi: 10.1002/da.1055] [Medline: 11668666]
52. Minimal data set. North American Quitline Consortium (NAQC). 2005. URL: https://www.naquitline.org/page/mds [accessed 2021-03-03]
53. Tedeschi GJ, Cummins SE, Anderson CM, Anthenelli RM, Zhuang Y, Zhu S. Smokers with self-reported mental health conditions: a case for screening in the context of tobacco cessation services. PLoS One 2016;11(7):e0159127 [FREE Full text] [doi: 10.1371/journal.pone.0159127] [Medline: 27391334]
54. Lee J, Contrera Avila J, Ahluwalia JS. Differences in cessation attempts and cessation methods by race/ethnicity among US adult smokers, 2016-2018. Addict Behav 2023;137:107523 [FREE Full text] [doi: 10.1016/j.addbeh.2022.107523] [Medline: 36279711]
55. Pagano A, Gubner N, Le T, Guydish J. Cigarette smoking and quit attempts among Latinos in substance use disorder treatment. Am J Drug Alcohol Abuse 2018;44(6):660-667 [FREE Full text] [doi: 10.1080/00952990.2017.1417417] [Medline: 29333890]
56. Cunningham JA, Kushnir V. Effectiveness of mass distribution of nicotine patches to promote tobacco control in rural versus urban settings. J Epidemiol Community Health 2017;71(5):519. [doi: 10.1136/jech-2016-208103] [Medline: 28396571]
57. Fu SS, van Ryn M, Nelson D, Burgess DJ, Thomas JL, Saul J, et al. Proactive tobacco treatment offering free nicotine replacement therapy and telephone counselling for socioeconomically disadvantaged smokers: a randomised clinical trial. Thorax 2016;71(5):446-453 [FREE Full text] [doi: 10.1136/thoraxjnl-2015-207904] [Medline: 26931362]
58. Kushnir V, Sproule BA, Cunningham JA. Mailed distribution of free nicotine patches without behavioral support: predictors of use and cessation. Addict Behav 2017;67:73-78. [doi: 10.1016/j.addbeh.2016.12.008] [Medline: 28039798]
59. Cunningham JA, Kushnir V, Selby P, Tyndale RF, Zawertailo L, Leatherdale ST. Effect of mailing nicotine patches on tobacco cessation among adult smokers: a randomized clinical trial. JAMA Intern Med 2016;176(2):184-190. [doi: 10.1001/jamainternmed.2015.7792] [Medline: 26809849]

Abbreviations

ACT: acceptance and commitment therapy
FDA: Food and Drug Administration
NCI: National Cancer Institute
NRT: nicotine replacement therapy
OR: odds ratio
RCT: randomized controlled trial
USCPG: US clinical practice guideline

XSL•FO
RenderX

Edited by A Mavragani; submitted 26.11.24; peer-reviewed by PS Olson; comments to author 24.12.24; revised version received 30.12.24; accepted 31.12.24; published 31.01.25.

Please cite as:
Santiago-Torres M, Mull KE, Sullivan BM, Cupertino AP, Salloum RG, Triplette M, Zvolensky MJ, Bricker JB
Evaluating the Impact of Pharmacotherapy in Augmenting Quit Rates Among Hispanic Adults in an App-Delivered Smoking Cessation Intervention: Secondary Analysis of a Randomized Controlled Trial
JMIR Form Res 2025;9:e69311
URL: https://formative.jmir.org/2025/1/e69311
doi:10.2196/69311
PMID:

©Margarita Santiago-Torres, Kristin E Mull, Brianna M Sullivan, Ana Paula Cupertino, Ramzi G Salloum, Matthew Triplette, Michael J Zvolensky, Jonathan B Bricker. Originally published in JMIR Formative Research (https://formative.jmir.org), 31.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of Cognitive Behavioral Therapy Provided Through a Web Application for Subthreshold Depression, Subthreshold Insomnia, and Subthreshold Panic: Open-Labeled 6-Arm Randomized Clinical Trial Pilot Study

Kayoko Taguchi[1], PhD; Mirai Miyoshi[2]; Yoichi Seki[1], PhD; Shiori Baba[3]; Eiji Shimizu[2], PhD

[1]Research Center for Child Mental Development, Chiba University, Chiba, Japan

[2]Graduate School of Medicine, Chiba University, Chiba, Japan

[3]Chiba University Hospital, Chiba, Japan

Corresponding Author:
Kayoko Taguchi, PhD
Research Center for Child Mental Development
Chiba University
chuo-ku
Inohana 1-8-1
Chiba, 260-8670
Japan
Phone: 81 43 226 2027
Fax: 81 43 226 2028
Email: k.taguchi@chiba-u.jp

Abstract

Background: A common definition of “subthreshold” is that the diagnostic threshold is not met but the individuals are not asymptomatic. Some symptoms are present, causing significant difficulty in functioning and negatively impacting quality of life. Despite the attention given to subthreshold symptoms and the interventions for subthreshold symptoms being efficient in preventing the transition to psychiatric disease in primary care, reports on specific interventions are insufficient.

Objective: This study aimed to verify the effectiveness of internet-delivered cognitive behavioral therapy (ICBT) for subthreshold depression (SD), subthreshold insomnia (SI), and subthreshold panic (SP). Additionally, this study aimed to explore the minimally important change (MIC) of each subthreshold group’s effectiveness outcome.

Methods: Participants aged 18-70 years from internet research monitors were categorized into SD, SI, and SP groups based on screening assessment. They were randomly assigned to intervention or control groups within each subthreshold symptom. The intervention groups worked on 4 weeks of nonguided ICBT (“Mentre”), while the control groups worked on a sham app. The primary outcome was the score change from screening (T1) to 4-week follow-up (T4) using the Center for Epidemiologic Studies Depression Scale (CESD) in the SD group, the Pittsburgh Sleep Quality Index (PSQI) in the SI group, and the Panic and Agoraphobia Scale (PAS) in the SP group. Secondary outcomes were score changes in the Generalized Anxiety Disorder-7 (GAD-7) scale, the Patient Health Questionnaire 9 (PHQ-9), the CESD, the PSQI, and the PAS, except the primary outcome in each group. Secondary outcomes were analyzed using complete-case analysis and repeated-measures ANOVA. Additionally, the MIC in the primary endpoint for each group was also calculated as an exploratory outcome.

Results: The SD, SP, and SI groups contained 846, 597, and 1106 participants, respectively. In the SD group, the difference in the CESD score change from baseline to follow-up between the intervention and control groups was significant (difference=0.52, 95% CI 1.29-4.66, *P*<.001). In the SI group, the difference in the PSQI score change was also significant (difference=0.53, 95% CI 0.11-0.94, *P*=.01). However, in the SP group, the difference in the PAS score change was not significant (difference=0.07, 95% CI –2.00 to 2.15, *P*=.94).

Conclusions: Our ICBT program Mentre contributes to the improvement of SI and SD. This suggests that nonguided ICBT may be effective in preventing SI and SD from progressing to the full threshold. However, appropriate definitions of subthreshold symptoms are necessary. In particular, it is difficult to define SP, and further research that considers the specific factors of each subthreshold symptom is necessary to accumulate evidence.

XSL•FO
RenderX

Trial Registration: University Hospital Medical Information Network (UMIN) UMIN000051280; https://tinyurl.com/2wyahhe3

(JMIR Form Res 2025;9:e63139) doi:10.2196/63139

KEYWORDS

minimally important change; nonguided cognitive behavioral therapy; subthreshold depression; subthreshold insomnia; subthreshold panic; cognitive behavioral therapy; CBT; psychiatric disease; primary care; interventions; depression; anxiety; insomnia; psychological therapy

Introduction

A common definition of "subthreshold" is that the diagnostic is not met but the individuals are not asymptomatic, that is, some symptoms are present, causing significant difficulty in functioning and negatively impacting the quality of life [1,2]. However, the concept of "subthreshold" symptoms remains unclear in the clinical setting. For instance, according to the latest scoping review of subthreshold depression (SD), there are many definitions based on the assessment tool cutoff point of the Center for Epidemiologic Studies Depression Scale (CESD), Beck's Depression Inventory II (BDI-II), and the Patient Health Questionnaire 9 (PHQ-9) and on the duration of symptoms. As a result, these definitions "seem to be arbitrary, with considerable overlap" [3]. Even the most popular subthreshold symptom is vague in its definition, and the other subthreshold symptoms, which are little mentioned, are even more so.

Although the definition is ambiguous, intervening for subthreshold symptoms is efficient in preventing the transition to psychiatric disease in primary care. Research has already suggested that shortening the duration of untreated illness (DUI) has a good therapeutic effect. The DUI is "measured as the interval between onset of the disorder and when the patient receives the first adequate treatment for that psychiatric disorder" [4], and many studies have claimed that the DUI affects treatment prognosis [5]. Furthermore, some studies have revealed the risk of transition from the subthreshold level to disease [6]. There is consensus that early intervention can be applied flexibly in different formats and across different target groups [7,8]. For instance, since SD and subthreshold anxiety occur twice as often as depression and anxiety [9], so appropriate definitions are, of course, essential for a concrete approach to primary care [10]. Furthermore, insomnia symptoms are closely associated with SD and subthreshold anxiety, and prevention programs for subthreshold insomnia (SI) have been reported to not only decrease depressive symptoms but also improve insomnia [11]. Namely, when considering primary care for subthreshold symptoms, overlapping symptoms also need to be considered; as a result, the treatment for specific subthreshold symptoms should also have the potential to cover a wide range of subthreshold symptoms. However, although there are some reports regarding SD interventions, there are insufficient studies on specific interventions in the subthreshold stage.

With limited reports, cognitive behavioral therapy (CBT) has been shown to be effective for subthreshold symptoms. CBT, which is a type of psychological therapy, improves symptoms by understanding patterns of maladaptive thinking and behavior and correcting the vicious cycle that perpetuates symptoms. It has been shown to be effective in improving diagnosed depression, anxiety, and insomnia [12-14]. However, nonguided internet-delivered cognitive behavioral therapy (ICBT) also has been shown to be effective for SD in a latest meta-analysis [15]. Since subthreshold symptoms are not severe, it is preferable that interventions be relatively short, easily accessible, and cost-effective. Nonguided CBT is more scalable and affordable [16,17], and it is suitable for people who have subthreshold symptoms but have no motivation to visit a hospital. However, there is insufficient evidence for ICBT as a subthreshold symptom treatment, and only 1 or 2 randomized controlled trials (RCTs) have verified that ICBT has valid effects on SD, SI, and subthreshold panic (SP). SI treatment verification over the internet has rarely been reported [18,19]. There are few reports on subthreshold anxiety, particularly reports on SP, which are mostly epidemiology and fact-finding surveys. Among the few papers on this topic, agreement has been reached concerning early intervention being valid and cost-effective for SP; however, they are not ICBTs [20,21].

Additionally, there are discussions of how patients perceive the significance of receiving ICBT for subthreshold symptoms. In recent years, interesting research results and suggestions concerning the minimally important change (MIC) have been reported [22]. The MIC is the smallest change or difference that patients perceive as significant. Researchers who reported on the MIC noted that the only satisfactory approach is for patients themselves to measure the health conditions they experience and know best. It is essential to discuss whether the MIC has implications for the significance of early intervention for subthreshold symptoms.

This study aimed to verify the effectiveness of the "Mentre" program as an ICBT for SD, SI, and SP in a 6-arm RCT. Additionally, this study aimed to explore the MIC of each subthreshold group's effectiveness outcome. These results will be useful as the first evidence accumulated about ICBT for subthreshold symptoms with stricter criteria and for knowledge of the MIC among patients with subthreshold symptoms.

Methods

Ethical Considerations

This study was conducted with the approval of the Institutional Review Board of Chiba University Hospital (approval number: G2022005). Additionally, the Clinical Research Ethics Review Committee oversaw proper implementation of the study at least once a year.

An explanatory document was presented to each web program site, and consent was obtained. Participants were informed that participation was voluntary, that they could decline at any time

XSL•FO
RenderX

for any reason, and that there would be no negative consequences for declining. Additionally, participants could withdraw their consent after providing informed consent.

For privacy, all participants' data were anonymized. Participants were granted points that would enable them to shop online as compensation for research participation. Each intervention group received 1000 points if the participants worked on the program and answered the outcome inventory completely, while the control groups received 800 points.

Trial Design and Procedure

The study was a prospective 6-arm, parallel-group RCT (Multimedia Appendix 1) [23]. The 6 arms were classified as SD, SI, and SP groups with 2 arms each. The trial was registered as "Effect of Mental Training Web Application for Subthreshold Insomnia, Subthreshold Depression and Subthreshold Panic Six Arms Randomized Controlled Trial" with the University Hospital Medical Information Network (UMIN; registration number: UMIN000051280). Participants were recruited from the internet research monitor of the NTT Com Online Marketing Solutions Corporation in May 2023. Participants who consented to the study completed a screening evaluation (T1) and a symptom persistence evaluation (T2) after 2 weeks. They were classified as having SD, SI, or SP and allocated to intervention or control groups. The intervention was a 4-week ICBT program Mentre, while the control program was a sham app, both of which started in June 2024 for each group. All participants were sent reminder emails every weekend to increase motivation and encourage continuous practice. After the program was complete, the participants were assessed for their symptoms postevaluation (T3), and 4 weeks later, follow-up evaluations were carried out (T4). Participants were awarded points for participating in the study and answering each time (T1-T4).

Participants

Participants were aged 18-70 years and did not visit a hospital for any disease. The exclusion and inclusion criteria for this study were as follows:

- Individuals who experienced little interest in things or had depressed thoughts almost every day for 2 weeks and met more than 5 items of depressive symptoms in the PHQ-9 were excluded as having depression. Of those who were not excluded, individuals with a CESD score of 16 or higher were eligible as those with SD.
- Individuals with SI who had PSQI scores greater than 6 and insomnia symptoms lasting less than 3 months were also eligible.
- Individuals with SP who had Panic and Agoraphobia Scale (PAS) scores greater than 9 and who experienced anxiety-related panic symptoms continuously lasting less than 1 month were also eligible. In the case of duplicate symptoms, if the panic score met the eligibility criteria, participants were allocated to the SP group preferentially because panic symptoms are more specific than depression and anxiety symptoms. SD sometimes overlaps with SI; hence, participants who showed SD and SI symptoms were allocated to the SD group.
- The exclusion criteria included individuals with a history of diagnosed depression, insomnia, or panic disorder and those who experienced difficulty or extreme difficulty in daily life owing to anxiety symptoms on the Generalized Anxiety Disorder-7 (GAD-7) scale or depressive thoughts on the PHQ-9.

Interventions

The intervention program Mentre was based on ICBT. We prepared 3 types of Mentre programs for each subthreshold symptom: SD, SI, and SP. Participants in the intervention groups undertook the relevant program for 4 weeks. The basic structure involved participants undergoing the first session at the beginning of the week and completing their homework throughout the remainder of the week. The control groups similarly recorded daily activities and weather conditions (sham app) in all subthreshold groups. All participants received guidance via email and accessed their program website through the URL provided on their mobile phones, personal computers, and tablets. The SD program included (1) cognitive restructuring about irrational thoughts, (2) psychoeducation for a depressive mood, (3) rumination distraction, and (4) behavioral activation to positive behavior. The SI program included (1) a sleep diary, (2) a review of one's sleeping behavior, (3) a review of one's sleep cognition, and (4) a review of one's sleep efficiency and sleep restriction methods. The SP program included (1) learning concerning panic disorder, (2) attention shift training for when facing a panic attack, (3) how to change the worst image of panic to a rational image, and (4) how to stop avoidant behavior. Each program involved paying attention to become aware of symptoms and to change irrational thoughts to rational thoughts and negative behavior to positive behavior. Participants could inquire about this study through each web program, and the researcher and manager of the web program responded accordingly. Reminder emails were sent to all participants at the end of each week to continue and not to forget working on the program.

Outcomes

Primary Endpoints

We set separate primary endpoints for each subthreshold symptom for the SD, SI, and SP groups: score change from screening (T1) to 4-week follow-up (T4) in each group:

- CESD in the SD group: The CESD is a brief self-report questionnaire for measuring the severity of depressive symptoms. It consists of 20 questions assessing numerous symptoms of depression experienced in the past week, with most items focusing on the emotional component of depression. Scores range from 0 to 60 points, with higher scores indicating more severe depressive symptoms. It has been used in many clinical trials involving a wide range of age groups [24,25].
- The Pittsburgh Sleep Quality Index (PSQI) in the SI group: The PSQI is a self-assessment consisting of 19 questions across 7 subscales (sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, hypnotic medication use, and daytime dysfunction). The total score indicates sleep quality and ranges between 0

XSL•FO
RenderX

(good sleep quality) and 21 (very poor sleep quality) points. The PSQI has been validated as a reliable and valid measure of subjective sleep quality in clinical practice and experimental studies [26,27].

- PAS in the SP group: The PAS is a 13-item questionnaire assessing panic attacks, composed of 5 factors (fear behavior, avoidance behavior, anticipatory anxiety, degree of disability, and health concerns) and evaluates the condition over the past week. Each item is scored from 0 to 4 points, with higher scores indicating more severe symptoms [28].

Secondary Endpoints

The secondary endpoints were the CESD, PSQI, and PAS score changes from screening (T1) to postevaluation (T3) and 4-week follow-up (T4) in all groups, as well as subsequent questionnaires. Additionally, the MIC in the primary endpoint for each group was also calculated.

- GAD-7 scale: It has been shown to have reliability, criterion, construct, factorial, and procedural validity. Cutoff points optimizing sensitivity (89%) and specificity (82%) were identified. The scale has 7 items assessing the severity of GAD in the past 2 weeks on a 4-point Likert scale (0=not at all, 1=1 episode, 2=on half or more days, and 3=almost daily). The minimum and maximum scores are 0 and 21, respectively (0-4, 5-9, 10-14, and 15-21 indicate no, mild, moderate, and severe symptoms, respectively). The cutoff score for clinically significant symptoms of anxiety is 10 [29].
- PHQ-9:It has diagnostic validity (for the diagnosis of 1 or more PHQ disorders; κ=0.65; overall accuracy=85,; sensitivity=75%, specificity=90%). It consists of 9 items scored on a 4-point Likert scale (0=not at all, 1=on several days, 2=on half or more days, and 3=almost daily). The minimum and maximum scores are 0 and 27, respectively (0-4, 5-9, 10-14, 15-19, and 20-27 indicate no, mild, moderate, moderate-to-severe, and severe symptoms, respectively). The PHQ cutoff score for clinically significant depressive symptoms is 10 [30].

Exploratory Endpoints

Minimally Important Change

The MIC is the minimum change threshold that the patient feels has improved or worsened and the change that is considered significant to the patient [20,29]. When patients evaluate their treatment effect using certain measures, it remains unclear how much change is meaningful. Hence, many studies suggest considering the MIC concept [32,33].

Sample Size

The sample size was calculated as follows: We needed to consider each subthreshold symptom, but there was insufficient evidence to estimate the effect size. We estimated the effect size of nonguided ICBT to be 0.2 based on a previous study [34]. Assuming an α level of .05 and a β level of .20, analysis was performed, resulting in a calculated sample size of 82. Based on the meta-analysis, more than 50% of the participants were expected to drop out. Hence, the sample size was set at 160 for each of the intervention and control groups.

Registration and Randomization

For case registration, participants who met the criteria were automatically registered after a preliminary screening on the server of an internet research company (NTT Com Online Marketing Solutions Corporation. This company assigned participants to the SI group, SD group, or SP group and further randomly assigned them to either the intervention or the control subgroup. The assigned control factors were CESD score≥23 or not for SD, PSQI score≥12 or not for SI, and PAS score≥22 or not for SP; sex was also assigned as a control factor for all groups. Participants were informed via email or other means about how to access the corresponding intervention program.

Protocol Deviations

Participants were provided the Mentre program through the web; therefore, dropouts and loss to follow-up were expected as main deviations. The number of not-evaluable cases was recorded. If any case reported a deviation due to an emergency crisis, the researcher immediately reported this to the Clinical Research Ethics Review Committee.

Statistical Methods

Main Analysis

Baseline variables were compared using the Fisher exact test and an unpaired 2-tailed *t* test for categorical and continuous variables, respectively. The significance level was set at .05. For the main analysis of treatment effects, the means of the least squares and their 95% CIs were estimated by repeated-measures ANOVA with each change in the CESD, PSQI, and PAS scores at the 4-week follow-up for each group. The model included intervention group, time, and intervention-by-time interaction as fixed effects. All comparisons were planned, and all *P* values were 2-sided. $P<.05$ was considered statistically significant. The primary outcome formed the basis of complete case analysis. No special complementary processing by statistical methods was performed for missing values. All statistical analyses were performed using R version 4.3.1 (R Foundation for Statistical Computing).

Secondary Analysis

To examine the MIC, we used an anchor. The patient-reported outcome (PRO) scores for each primary outcome were classified into the following anchor categories: “much better,” “slightly better,” “about the same,” “slightly worse,” and “much worse.” The average PRO score classified as “slightly better” was considered the MIC. We calculated the CESD, PSQI, and PAS scores of the MIC in this study.

Results

Participant Details

The flow of participants through the study is illustrated in Figure 1. This study was announced to 215,000 internet research monitors, and 25,418 (11.8%) monitors provided consent. Of them, 2549 (10%) participants with SD, SI, or SP symptoms were screened (T1) and assigned to 1 of 6 arms. Afterward,

XSL•FO
RenderX

2091 (82%) participants' symptoms continued after 2 weeks (T2). In total, 846 (33.2%) participants (n=424, 50.1%, intervention vs n=422, 49.9%, control participants) were classified as having SD, 1106 (43.4%) participants (n=552, 49.9%, intervention vs n=554, 50.1%, control participants) as having SI, and 597 (23.4%) participants (n=299, 50.1%, intervention vs n=298, 49.9%, control participants) as having SP according to screening evaluation scores. A total of 145 (34.2%) of 424 participants who were assigned to the SD group undertook the Mentre program, with 17 (11.7%) dropping out during the intervention. A total of 267 out of 552 participants who were assigned to the SI group undertook the Mentre program, with 74 dropping out during the intervention (27.7%). A total of 141 out of 299 participants who were assigned to the subthreshold panic group undertook the Mentre program, with 60 dropping out during the intervention (43.2%).

Figure 1. Flow diagram of the study participants.

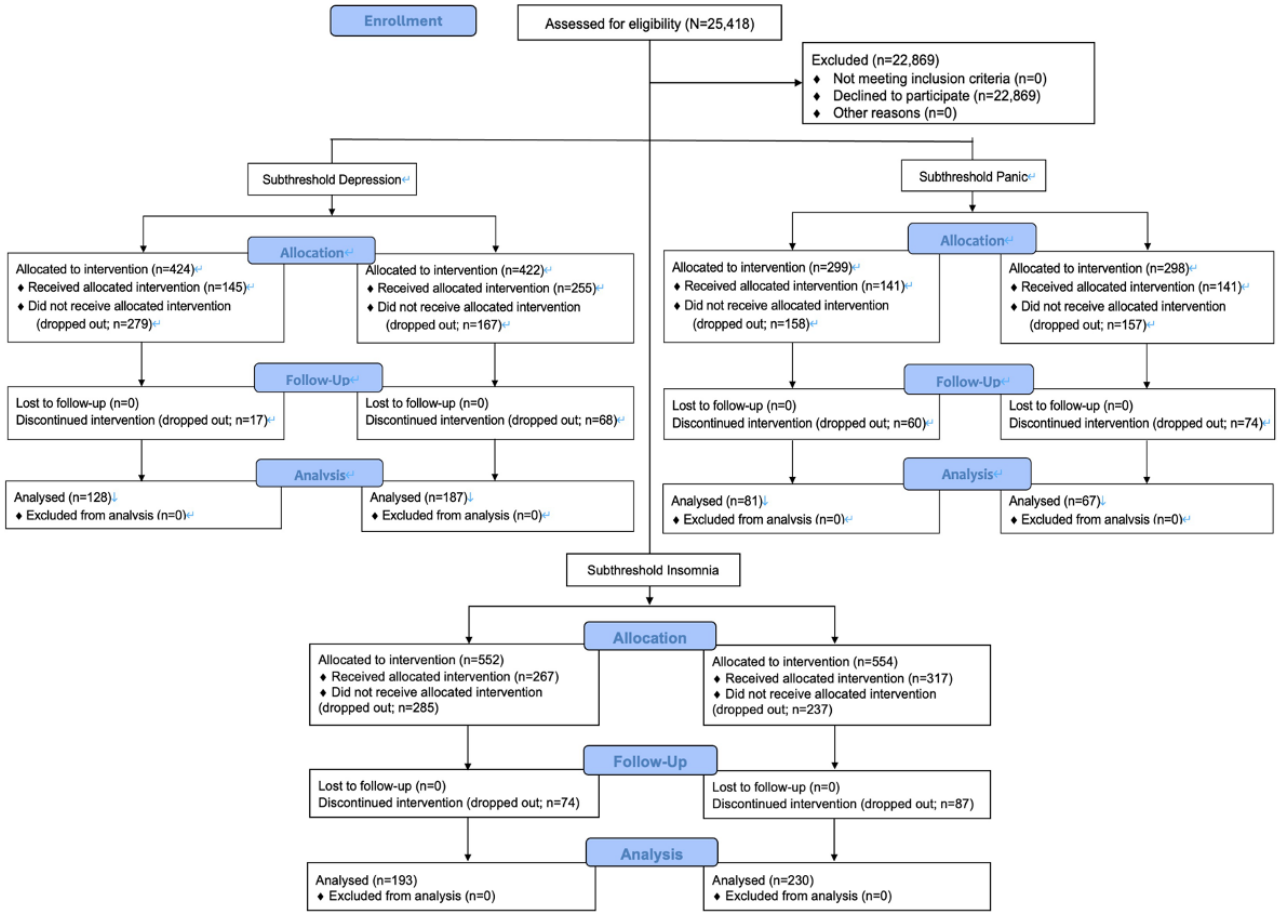


Table 1 presents the clinical characteristics of the patients in each group. In all 3 groups, no significant difference was found between the intervention and control subgroups in any average value, including sex, age, marital status, cohabitation rate, employment status, and income.

XSL•FO
RenderX

Table 1. Baseline demographic and clinical characteristics (N=2549).

Characteristics	SD[a]			SI[b]			SP[c]		
	Intervention (n=424)	Control (n=422)	*P* value	Intervention (n=552)	Control (n=554)	*P* value	Intervention (n=299)	Control (n=298)	*P* value
Sex (male), n (%)	130 (30.7)	130 (30.8)	.99	225 (40.7)	227 (41.0)	.951	106 (35.4)	107 (35.9)	.932
Age (years), mean (standard deviation)	41.3 (12.48)	42.5 (11.68)	.068	42.3 (11.74)	42.8 (12.10)	.546	41.2 (12.60)	42.1 (12.03)	.363
Marital status, n (%)									
Married	229 (54.0)	239 (56.6)	.448	318 (57.6)	331 (59.7)	.502	154 (51.5)	156 (52.3)	.870
Living with someone	322 (75.9)	317 (75.1)	.811	425 (77.0)	429 (77.4)	.886	236 (78.9)	229 (76.8)	.555
Education, mean (standard deviation)	14.8 (1.99)	14.7 (2.05)	.410	14.9 (1.95)	14.7 (1.92)	.059	14.6 (2.10)	14.3 (2.15)	.208
Employment, n (%)	284 (66.9)	291 (68.9)	.556	416 (75.3)	402 (72.6)	.304	195 (65.2)	186 (62.4)	.496
Yearly income (yen), n (%)	—[d]	—	.783	—	—	.320	—	—	.150
<1,000,000 (US $6322.53)[e]	131 (30.8)	119 (28.2)	—	126 (22.8)	143 (25.8)	—	105 (35.1)	112 (37.6)	—
1,000,001-5,000,000 (US $6322.54-$31,612.66)	203 (47.8)	214 (50.7)	—	284 (51.4)	268 (48.3)	—	137 (45.8)	135 (45.3)	—
5,000,001-10,000,000 (US $31,612.66-$63,225.31)	78 (18.4)	75 (17.7)	—	128 (23.2)	121 (21.8)	—	53 (17.7)	40 (13.4)	—
>10,000,000 (US $63,225.31)	12 (2.8)	14 (3.3)	—	14 (2.5)	22 (4)	—	4 (1.3)	11 (3.7)	—

[a]SD: subthreshold depression.

[b]SI: subthreshold insomnia.

[c]SP: subthreshold panic.

[d]Not applicable.

[e]An exchange rate of 1 yen=US $0.0063 has been applied.

Primary Outcome

Tables 2-4 and Figure 2 show the effects of the Mentre program as the primary outcome for each group. In the SD group, differences in the CESD score change from T1 to T4 between the intervention and control groups were significant (difference=0.52, 95% CI 1.29-4.66, *P*<.001). In the SI group, differences in the PSQI score change from T1 to T4 between the intervention and control groups were significant (difference=0.53, 95% CI 0.11-0.94, *P*=.01). In the SP group, differences in the PAS score change from T1 to T4 between the intervention and control groups were not significant (difference=0.07, 95% CI –2.00 to 2.15, *P*=.94).

XSL•FO
RenderX

Table 2. Primary outcome ANOVA results for the SD[a] group using the CESD[b].

Week[c]	Intervention (n=128), mean (standard deviation)	Control (n=187), mean (standard deviation)	Estimate (95% CI)	SE	*P* value	Cohen's d
T1	22.42 (9.03)	22.11 (8.48)	—[d]	—	—	—
T2	20.54 (9.70)	20.47 (8.97)	0.16 (–1.33 to 1.65)	0.76	.829	—
T3	18.41 (9.98)	20.18 (10.14)	2.01 (0.28 to 3.73)	0.88	.023[e]	0.18
T4	17.20 (9.33)	19.94 (10.05)	2.98 (1.29 to 4.66)	0.86	.001[f]	0.28

[a]SD: subthreshold depression.

[b]CESD: Center for Epidemiologic Studies Depression Scale.

[c]T1: screening; T2: symptom persistence evaluation; T3: postevaluation; T4: 4-week follow-up.

[d]Not applicable.

[e]*P*<.05.

[f]*P*<.01.

Table 3. Primary outcome ANOVA results for the SI[a] group using the PSQI[b].

Week[c]	Intervention (n=193), mean (standard deviation)	Control (n=230), mean (standard deviation)	Estimate (95% CI)	SE	*P* value	Cohen's d
T1	7.66 (1.63)	7.69 (1.87)	—[d]	—	—	—
T2	6.90 (2.26)	7.23 (2.49)	0.32 (–0.06 to 0.69)	0.19	.099	—
T3	6.48 (2.23)	7.00 (2.40)	0.50 (0.10 to 0.90)	0.20	.015[e]	0.22
T4	6.09 (2.26)	6.63 (2.42)	0.53 (0.11 to 0.94)	0.21	.013[e]	0.23

[a]SI: subthreshold insomnia.

[b]PSQI: Pittsburgh Sleep Quality Index.

[c]T1: screening; T2: symptom persistence evaluation; T3: postevaluation; T4: 4-week follow-up.

[d]Not applicable.

[e]*P*<.05.

Table 4. Primary outcome ANOVA results for the SP[a] group using the PAS[b].

Week[c]	Intervention (n=81), mean (standard deviation)	Control (n=67), mean (standard deviation)	Estimate (95% CI)	SE	*P* value	Cohen's d
T1	13.89 (4.73)	13.54 (4.40)	—[d]	—	—	—
T2	11.47 (6.68)	12.60 (7.32)	1.41 (–0.54 to 3.36)	0.99	.156	—
T3	12.30 (6.93)	10.43 (6.21)	–1.65 (–3.62 to 0.33)	1.00	.101	0.29
T4	10.69 (7.02)	10.52 (7.05)	0.07 (–2.00 to 2.15)	1.05	.945	0.02

[a]SP: subthreshold panic.

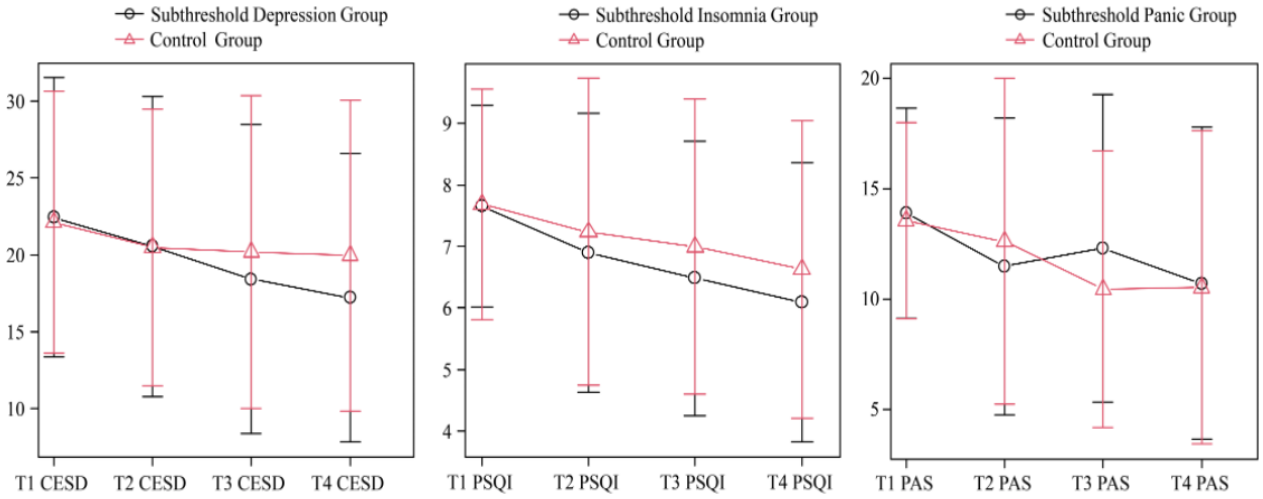
[b]PAS: Panic and Agoraphobia Scale.

[c]T1: screening; T2: symptom persistence evaluation; T3: postevaluation; T4: 4-week follow-up.

[d]Not applicable.

XSL•FO
RenderX

Figure 2. The effect of the Mentre program as the primary outcome for each group.



Secondary Outcomes

Tables 5-7 present the means of the secondary endpoints in each group and the differences between groups. In the SD group, differences in the mean PSQI and PAS scores at either time point were not significant; however, the difference in the PHQ-9 score at T4 was significant (difference=1.00, 95% CI 0.04-1.97, *P*=.04). Furthermore, at the time of postevaluation (T3), a significant difference was revealed in the PAS score (difference=2.05, 95% CI 0.02-4.08, *P*=.047). In the SI group, at the time of postevaluation (T3), a significant difference was found in the PSQI (difference=0.50, 95% CI 0.01-0.90, *P*=.02); however, between-group differences in changes in other endpoints were not significant at any time point. In the SP group, between-group differences in changes in all secondary endpoints were not significant.

XSL•FO
RenderX

Table 5. Secondary outcome ANOVA results for the SD[a] group.

Variables and weeks[b]	Intervention, mean (standard deviation)	Control group, mean (standard deviation)	SE (95% CI)	*P* value	Cohen's d
PSQI [c] (intervention: n=128; control: n=187)					
T1	6.01 (2.13)	6.47 (2.44)	—[d]	—	—
T3	5.83 (2.57)	6.53 (2.82)	0.26 (–0.12 to 0.89)	.13	0.26
T4	5.52 (2.48)	5.97 (3.13)	0.28 (–0.41 to 0.67)	.64	0.16
PAS [e] (intervention: n=38; control: n=48)					
T1	4.29 (3.70)	3.79 (2.21)	—	—	—
T3	4.24 (4.56)	5.94 (5.39)	1.02 (0.02 to 4.08)	.05	0.34
T4	4.29 (5.34)	4.60 (5.36)	1.09 (–1.50 to 2.84)	.54	0.06
PHQ-9 [f] (intervention: n=128; control: n=187)					
T1	7.81 (4.42)	7.80 (4.19)	—	—	—
T3	6.57 (5.18)	6.83 (5.23)	0.45 (–0.61 to 1.17)	.54	0.05
T4	5.69 (5.53)	6.68 (5.30)	0.49 (0.04 to 1.97)	.04[g]	0.18
GAD-7 [h] (intervention: n=128; control: n=187)					
T1	6.45 (4.26)	6.49 (4.23)	—	—	—
T3	5.09 (4.21)	5.42 (4.18)	0.39 (–0.46 to 1.07)	.43	0.08
T4	4.49 (4.14)	5.20 (4.22)	0.40 (–0.09 to 1.48)	.08	0.17

[a]SD: subthreshold depression.

[b]T1: screening; T3: postevaluation; T4: 4-week follow-up.

[c]PSQI: Pittsburgh Sleep Quality Index.

[d]Not applicable.

[e]PAS: Panic and Agoraphobia Scale.

[f]PHQ-9: Patient Health Questionnaire-9.

[g]$P<.05$.

[h]GAD-7: Generalized Anxiety Disorder-7.

XSL•FO
RenderX

Table 6. Secondary outcome ANOVA results for the SI[a] group.

Variables and weeks[b]	Intervention, mean (standard deviation)	Control group, mean (standard deviation)	SE (95% CI)	*P* value	Cohen's d
CESD [c] (intervention: n=193; control: n=230)					
T1	15.74 (7.74)	15.45 (7.18)	—[d]	—	—
T3	15.37 (7.99)	15.68 (8.46)	0.57 (–0.59 to 1.65)	.36	0.04
T4	14.49 (8.20)	15.02 (8.48)	0.60 (–0.44 to 1.92)	.22	0.06
PAS [e] (intervention: n=53; control: n=54)					
T1	3.91 (3.47)	3.43 (2.35)	—	—	—
T3	4.74 (4.72)	4.80 (4.59)	0.86 (–1.37 to 2.03)	.70	0.01
T4	4.26 (4.23)	3.83 (4.14)	0.78 (–1.75 to 1.34)	.79	0.10
PHQ-9 [f] (intervention: n=193; control: n=68)					
T1	3.98 (3.62)	5.01 (4.45)	—	—	—
T3	4.53 (4.27)	5.10 (4.85)	0.63 (–0.67 to 1.82)	.37	0.13
T4	4.13 (4.57)	4.19 (4.50)	0.65 (–1.21 to 1.35)	.91	0.01
GAD-7 [g] (intervention: n=193; control: n=230)					
T1	4.36 (4.13)	3.73 (3.46)	—	—	—
T3	3.54 (3.94)	3.54 (3.88)	0.31 (–0.22 to 0.99)	.22	0.00
T4	3.16 (3.68)	2.95 (3.68)	0.30 (–0.45 to 0.72)	.65	0.06

[a]SI: subthreshold insomnia.

[b]T1: screening; T3: postevaluation; T4: 4-week follow-up.

[c]CESD: Center for Epidemiologic Studies Depression Scale.

[d]Not applicable.

[e]PAS: Panic and Agoraphobia Scale.

[f]PHQ-9: Patient Health Questionnaire-9.

[g]GAD-7: Generalized Anxiety Disorder-7.

XSL•FO
RenderX

Table 7. Secondary outcome ANOVA results for the SP[a] group.

Variables and weeks[b]	Intervention, mean (standard deviation)	Control group, mean (standard deviation)	SE (95% CI)	*P* value	Cohen's d
PSQI [c] (intervention: n=99; control: n=98)					
T1	7.30 (2.76)	7.01 (2.79)	—[d]	—	—
T3	7.27 (2.98)	6.91 (2.93)	0.28 (–0.68 to 0.42)	.64	0.12
T4	6.80 (2.88)	6.73 (2.93)	0.28 (–0.39 to 0.72)	.56	0.02
CESD [e] (intervention: n=99; control: n=98)					
T1	24.68 (10.23)	21.80 (10.26)	—	—	—
T3	23.22 (9.53)	20.53 (9.21)	1.00 (–2.90 to 1.07)	.36	0.29
T4	20.33 (8.98)	20.51 (9.48)	1.02 (–0.16 to 3.87)	.07	0.02
PHQ-9 [f] (intervention: n=99; control: n=98)					
T1	6.04 (4.96)	4.92 (4.70)	—	—	—
T3	8.68 (5.83)	7.68 (5.71)	0.83 (–2.70 to 0.59)	.21	0.17
T4	8.01 (5.47)	7.30 (5.31)	0.78 (–2.35 to 0.72)	.30	0.13
GAD-7 [g] (intervention: n=99; control: n=98)					
T1	8.55 (4.64)	7.44 (3.96)	—	—	—
T3	7.26 (4.87)	6.28 (4.08)	0.52 (–1.32 to 0.73)	.57	0.22
T4	5.95 (4.23)	5.76 (3.98)	0.48 (–0.52 to 1.37)	.38	0.55

[a]SP: subthreshold panic.

[b]T1: screening; T3: postevaluation; T4: 4-week follow-up.

[c]PSQI: Pittsburgh Sleep Quality Index.

[d]Not applicable.

[e]CESD: Center for Epidemiologic Studies Depression Scale.

[f]PHQ-9: Patient Health Questionnaire-9.

[g]GAD-7: Generalized Anxiety Disorder-7.

Calculation of the MIC

Table 8 presents the MICs of the CESD, PSQI, and PAS. The mean MIC of the CESD for participants who undertook the Mentre program for SD was –5.23 (standard deviation 8.89). The mean MIC of the PSQI for participants who undertook the Mentre program for SI was –3.60 (standard deviation 6.34). The mean MIC of the PAS for participants who undertook the Mentre program for SP was –2.08 (standard deviation 2.44).

XSL•FO
RenderX

Table 8. MIC[a] of each primary endpoint.

CESD[b] in the SD[c] group			PSQI[d] in the SI[e] group		PAS[f] in the SP[g] group	
MIC	Participants, n (%)	Mean (standard deviation)	Participants, n (%)	Mean (standard deviation)	Participants, n (%)	Mean (standard deviation)
Better	23 (16.0)	–9.78 (6.381)	18 (8.4)	–3.78 (2.184)	11 (11.8)	–7.91 (5.991)
Slightly better	48 (33.3)	–5.23 (8.890)	71 (33.2)	–2.08 (2.436)	45 (48.4)	–3.60 (6.340)
About the same	66 (45.8)	–3.61 (7.831)	113 (52.8)	–1.1 (1.911)	33 (35.5)	–1.79 (5.732)
Slightly worse	4 (2.8)	–1.00 (1.414)	11 (5.1)	0.91 (2.548)	4 (4.3)	–3.00 (4.761)
Worse	3 (2.1)	8.00 (14.731)	1 (0.5)	0.00	0	—[h]
Total	144 (100.0)	–4.82 (8.470)	214 (100.0)	–1.59 (2.338)	93 (100.0)	–3.44 (6.220)

[a]MIC: minimally important change.

[b]CESD: Center for Epidemiologic Studies Depression Scale.

[c]SD: subthreshold depression.

[d]PSQI: Pittsburgh Sleep Quality Index.

[e]SI: subthreshold insomnia.

[f]PAS: Panic and Agoraphobia Scale.

[g]SP: subthreshold panic.

[h]Not applicable.

Adverse Events

None of the participants in either group experienced serious adverse events during the time of this study.

Discussion

Principal Findings

This study aimed to assess the effectiveness of the Mentre program, a 4-week ICBT, for SD, SI, and SP. We categorized participants into SD, SI, and SP groups based on their eligibility using previous research and conducted an RCT for each group. We verified differences in the score change from screening (T1) to 4-week follow-up (T4) as indicators of ICBT effectiveness. There was a significant difference between the SD and SI intervention and control groups but not in the SP groups. Additionally, significant differences in CESD score changes were observed from screening (T1) to postintervention (T3) in the SD intervention and control groups. The SI intervention and control groups also showed significant PSQI score changes from T1 to T3. Previous studies have yielded mixed results on ICBT for SD, indicating its effectiveness compared to usual treatment, with effectiveness maintained for more than 6 months [6,35]. Limited research on SI has also suggested the effectiveness of ICBT in improving insomnia symptoms [18]. Our findings of SD and SI are comparable to those of previous studies. However, there was no significant difference in PAS score changes in the SP intervention and control groups. One reason was the possibility that the eligibility of SP was not appropriate. Although no studies have yet properly defined and intervened on SP [36-40], a previous study [36] suggests that if agoraphobia is present, we should carefully attend to it even if it does not completely meet the panic disorder criteria. In another study [40], a participant was classified according to the severity based on the frequency of panic attacks. Our study participants were considered to have SP if they had no panic attacks and the duration of symptoms was less than 1 month but they scored more than 9 points on the PAS. Therefore, the Mentre program for SP did not work, because it focused on anticipatory anxiety or dealt with avoidance of a particular place.

At least in primary care, our Mentre app found that self-understanding depressive moods and thoughts and trying to change them into rational behavior are effective in SD and that recording and reviewing sleep are effective in SI. However, further research is needed for the definition of SP; in particular, it is necessary to verify whether ICBT programs, such as Mentre, are effective for SP accompanied by agoraphobia and panic attacks.

Regarding subthreshold criteria, it is not just SP that requires discussion. Previous studies [41,42] have emphasized the need to define the score indicating a healthy state for the definition of subthreshold symptoms. In our study, we considered the persistence of each subthreshold symptom for 2 weeks (T2), resulting in 458 participants no longer meeting the subthreshold criteria. Can we conclude that the 458 participants did not have subthreshold symptoms? It is worth considering whether the absence of symptoms for 2 weeks defines health. These patients were excluded from our study; however, we should consider the timing of the intervention for subthreshold symptoms, including medical histories, event factors, and the frequency of panic attacks or depressive episodes, in the definition of subthreshold symptoms in the future.

Given that the definition of the subthreshold state is ambiguous, it is difficult to estimate the appearance rate of subthreshold symptoms. In our study, we estimated 160 participants in each group as the target sample size, but in fact, there was considerable discrepancy. This may indicate that there are many more individuals with subthreshold symptoms than we predicted, possibly because there is substantial overlap across subthreshold states. In contrast, we also need to consider that the participants

XSL•FO
RenderX

earned reward points by registering over the internet, which may relate to the high dropout rate, and the participants' motivation to test the Mentre program should also be considered. Individuals with subthreshold symptoms may lack high motivation for treatment compared to diagnosed patients, because their symptoms are of a low level and do not impair everyday function. The Mentre program operates at the same level used for patients diagnosed with depression, insomnia, and panic disorder. Hence, daily participation may have been bothersome for the participants.

Future studies should be designed considering the concept of subthreshold symptoms and possible overlap across subthreshold conditions and having a structure that is manageable and appropriate for this particular patient population.

An essential consideration is the overlap across subthreshold symptoms. In this study, we verified all 3 subthreshold symptoms together, given the frequent comorbidity of depression and anxiety disorders, often with insomnia [43,44]. However, few previous studies have explored this overlap. A previous study [45] on universal prevention treatment for depression suggested that adolescents at low risk for depression benefit equally from preventive interventions as those at high risk; however, it did not specify a matching treatment for low risk [45]. In other words, patients with subthreshold symptoms may benefit from treatment, but the treatment is not specified, and a completely new approach may be necessary [46]. We hypothesize that if there is overlap across subthreshold symptoms, interventions targeting specific symptoms might also affect overlapping symptoms owing to their low severity. Unfortunately, our program only showed efficacy for the symptoms we pretargeted, but previous meta-analyses have shown that improvement in diagnosed insomnia is related to improved depressive symptoms [14]. Therefore, treatments focusing on identifying components of subthreshold symptoms and supplementing CBT techniques may be effective.

Additionally, our study provided evidence of the MIC for SD, SI, and SP, which is specific to the subthreshold condition. Despite being useful as an index for clinicians and patients, caution is necessary in generalizing to individuals, because the MIC provides a statistical measure of association and may not reflect clinical conditions [47]. Particularly, subthreshold symptoms were originally low in severity; hence, the MIC was also low. For the indicated MIC to be an important change, it is necessary to also examine the events and satisfaction levels that changed because of the MIC for patients who reported that they felt "slightly better."

Limitations

Several limitations of this study should be acknowledged. First, the participants were limited to internet research monitors; hence, it is possible that our sample may not have been representative of all individuals with subthreshold symptoms. Approximately 30% of the participants were not employed and had a low income, which also does not reflect the entire population. Second, the main analysis was a complete case analysis owing to concerns regarding the high dropout rate among patients with subthreshold symptoms. In this study, we needed to verify the effectiveness of the Mentre program for specialized subthreshold symptoms as a pilot study. In the future, more accurate efficacy verification research is needed by conducting RCTs on the general population and analyzing them using intent-to-treat methods.

Conclusion

Our study suggests that the use of the 4-week Mentre program as a nonguided ICBT is associated with improvements in symptoms of SD and SI in primary care. However, for SP, there is a need to first define it, and next, it is necessary to re-examine the effectiveness of ICBT. Additionally, although the MIC of subthreshold symptoms may be useful in clinical situations, individual adaptation needs to be approached with caution.

Acknowledgments

We would like to thank Editage for English language editing and all the researchers and staff who contributed to this study, as well as all the study participants. This work was supported by the Adaptable and Seamless Technology Transfer Program through Target-driven R&D (A-STEP) from the Japan Science and Technology Agency (JST; grant number: JPMJTR221A).

Data Availability

Raw data were generated at the research center of child mental developmental. Derived data supporting the findings of this study are available from the corresponding author (KT) upon request.

Authors' Contributions

KT and ES conceived and designed the study. KT, ES, and YS designed the Mentre program. KT calculated the sample size and developed an analysis plan. KT, MM. and SB wrote the first draft of the manuscript, and all other authors revised the manuscript critically. This manuscript was not used generative artificial intelligence in any part. All authors have approved the final version of the manuscript.

Conflicts of Interest

NTT Com Online Marketing Solutions Corporation was involved in this study, as mentioned in the manuscript. This company was in charge of recruitment, assignment randomization, data collection, and sending reminder emails to participants. The Mentre

XSL•FO
RenderX

program was developed by the authors, the subthreshold depression and subthreshold insomnia program was written using the programming code of JUNPUZI Co, Ltd, and the subthreshold panic program was written by NTT Communications.

Multimedia Appendix 1
CONSORT (Consolidated Standards of Reporting Trials) checklist.
[PDF File (Adobe PDF File), 131 KB - formative_v9i1e63139_app1.pdf]

References

1. Lipowski ZJ. Delirium (acute confusional states). JAMA 1987 Oct 02;258(13):1789-1792 [FREE Full text] [doi: 10.1001/jama.1987.03400130103041]
2. Rodríguez MR, Nuevo R, Chatterji S, Ayuso-Mateos JL. Definitions and factors associated with subthreshold depressive conditions: a systematic review. BMC Psychiatry 2012 Oct 30;12(1):181. [doi: 10.1186/1471-244x-12-181]
3. Volz H, Stirnweiß J, Kasper S, Möller HJ, Seifritz E. Subthreshold depression - concept, operationalisation and epidemiological data. A scoping review. Int J Psychiatry Clin Pract 2023 Mar 23;27(1):92-106 [FREE Full text] [doi: 10.1080/13651501.2022.2087530] [Medline: 35736807]
4. Albert U, Barbaro F, Bramante S, Rosso G, De Ronchi D, Maina G. Duration of untreated illness and response to SRI treatment in obsessive-compulsive disorder. Eur Psychiatry 2019 May 11;58:19-26 [FREE Full text] [doi: 10.1016/j.eurpsy.2019.01.017] [Medline: 30763828]
5. Vigne P, Fortes P, Dias RV, Laurito LD, Loureiro CP, de Menezes GB, et al. Duration of untreated illness in a cross-diagnostic sample of obsessive-compulsive disorder, panic disorder, and social anxiety disorder. CNS Spectr 2018 Nov 13;24(5):526-532. [doi: 10.1017/s1092852918001281]
6. Francesconi M, Minichino A, Carrión RE, Delle Chiaie R, Bevilacqua A, Parisi M, et al. Psychosis prediction in secondary mental health services. A broad, comprehensive approach to the "at risk mental state" syndrome. Eur Psychiatry 2017 Feb 16;40:96-104. [doi: 10.1016/j.eurpsy.2016.09.002] [Medline: 27992839]
7. Takagaki K, Okamoto Y, Jinnin R, Mori A, Nishiyama Y, Yamamura T, et al. Behavioral activation for late adolescents with subthreshold depression: a randomized controlled trial. Eur Child Adolesc Psychiatry 2016 Nov;25(11):1171-1182. [doi: 10.1007/s00787-016-0842-5] [Medline: 27003390]
8. Cuijpers P, Quero S, Dowrick C, Arroll B. Psychological treatment of depression in primary care: recent developments. Curr Psychiatry Rep 2019 Nov 23;21(12):129 [FREE Full text] [doi: 10.1007/s11920-019-1117-x] [Medline: 31760505]
9. Helmchen H. [Subthreshold psychological disorders]. Nervenarzt 2001 Mar 26;72(3):181-189. [doi: 10.1007/s001150050737] [Medline: 11268762]
10. Hanel G, Henningsen P, Herzog W, Sauer N, Schaefert R, Szecsenyi J, et al. Depression, anxiety, and somatoform disorders: vague or distinct categories in primary care? Results from a large cross-sectional study. J Psychosom Res 2009 Sep;67(3):189-197. [doi: 10.1016/j.jpsychores.2009.04.013] [Medline: 19686874]
11. Chan N, Li S, Zhang J, Lam SP, Kwok APL, Yu MWM, et al. A prevention program for insomnia in at-risk adolescents: a randomized controlled study. Pediatrics 2021 Mar;e2020006833(3):147. [doi: 10.1542/peds.2020-006833] [Medline: 33627370]
12. López-López JA, Davies SR, Caldwell DM, Churchill R, Peters TJ, Tallon D, et al. The process and delivery of CBT for depression in adults: a systematic review and network meta-analysis. Psychol Med 2019 Jun 10;49(12):1937-1947. [doi: 10.1017/s003329171900120x]
13. Papola D, Ostuzzi G, Tedeschi F, Gastaldon C, Purgato M, Del Giovane C, et al. CBT treatment delivery formats for panic disorder: a systematic review and network meta-analysis of randomised controlled trials. Psychol Med 2022 Dec 09;53(3):614-624. [doi: 10.1017/s0033291722003683]
14. Hertenstein E, Trinca E, Wunderlin M, Schneider CL, Züst MA, Fehér KD, et al. Cognitive behavioral therapy for insomnia in patients with mental disorders and comorbid insomnia: a systematic review and meta-analysis. Sleep Med Rev 2022 Apr;62:101597 [FREE Full text] [doi: 10.1016/j.smrv.2022.101597] [Medline: 35240417]
15. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FMG, Furukawa TA, Cuijpers P, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 2021 Apr 01;78(4):361-371 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
16. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 01;74(4):351-359 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
17. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther 2017 Jan;88:19-25 [FREE Full text] [doi: 10.1016/j.brat.2016.08.012] [Medline: 28110672]
18. Denis D, Eley TC, Rijsdijk F, Zavos HM, Keers R, Espie CA, et al. Is digital cognitive behavioural therapy for insomnia effective in treating sub-threshold insomnia: a pilot RCT. Sleep Med 2020 Feb;66:174-183. [doi: 10.1016/j.sleep.2019.10.007] [Medline: 31901759]

XSL•FO
RenderX

19. Theppornpitak W, Hemrungrojn S, Thienwiwatnukul K, Muntham D, Chirakalwasan N, Srisawart P. Effectiveness of internet-based CBT-I for the treatment of chronic subthreshold to moderate insomnia. Front Neurol 2023 Jun 2;14:1180339 [FREE Full text] [doi: 10.3389/fneur.2023.1180339] [Medline: 37346166]
20. Meulenbeek P, Spinhoven P, Smit F, van Balkom A, Cuijpers P. Cognitive mediation of panic reduction during an early intervention for panic. Acta Psychiatr Scand 2010 Jul;122(1):20-29. [doi: 10.1111/j.1600-0447.2009.01530.x] [Medline: 20070849]
21. Ophuis RH, Lokkerbol J, Hiligsmann M, Evers SMAA. Early intervention for subthreshold panic disorder in the Netherlands: a model-based economic evaluation from a societal perspective. PLoS One 2018 Feb 21;13(2):e0193338 [FREE Full text] [doi: 10.1371/journal.pone.0193338] [Medline: 29466470]
22. Wang Y, Devji T, Carrasco-Labra A, King MT, Terluin B, Terwee CB, et al. A step-by-step approach for selecting an optimal minimal important difference. BMJ 2023 May 26;381:e073822. [doi: 10.1136/bmj-2022-073822] [Medline: 37236647]
23. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
24. Lewinsohn PM, Seeley JR, Roberts RE, Allen NB. Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. Psychol Aging 1997;12(2):277-287. [doi: 10.1037/0882-7974.12.2.277]
25. Kim J, Chung H, Askew RL, Park R, Jones SMW, Cook KF, et al. Translating CESD-20 and PHQ-9 scores to PROMIS depression. Assessment 2017 Apr 28;24(3):300-307. [doi: 10.1177/1073191115607042] [Medline: 26423348]
26. Sato D, Sutoh C, Seki Y, Nagai E, Shimizu E. Treatment preferences for internet-based cognitive behavioral therapy for insomnia in Japan: online survey. JMIR Form Res 2019 May 15;3(2):e12635 [FREE Full text] [doi: 10.2196/12635] [Medline: 31094319]
27. Carpenter JS, Andrykowski MA. Psychometric evaluation of the Pittsburgh Sleep Quality Index. J Psychosom Res 1998 Jul;45(1):5-13. [doi: 10.1016/S0022-3999(97)00298-5]
28. Bandelow B. Assessing the efficacy of treatments for panic disorder and agoraphobia. II. The Panic and Agoraphobia Scale. Int Clin Psychopharmacol 1995 Jun;10(2):73-81. [doi: 10.1097/00004850-199506000-00003] [Medline: 7673659]
29. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
30. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
31. Terwee CB, Peipert JD, Chapman R, Lai J, Terluin B, Cella D, et al. Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. Qual Life Res 2021 Oct;30(10):2729-2754 [FREE Full text] [doi: 10.1007/s11136-021-02925-y] [Medline: 34247326]
32. Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. J Clin Epidemiol 2017 Feb;82:128-136 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.11.016] [Medline: 27986610]
33. Dawson J, Boller I, Doll H, Lavis G, Sharp R, Cooke P, et al. Minimally important change was estimated for the Manchester-Oxford Foot Questionnaire after foot/ankle surgery. J Clin Epidemiol 2014 Jun;67(6):697-705. [doi: 10.1016/j.jclinepi.2014.01.003] [Medline: 24709030]
34. Imamura K, Kawakami N, Furukawa TA, Matsuyama Y, Shimazu A, Umanodan R, et al. Effects of an internet-based cognitive behavioral therapy intervention on improving work engagement and other work-related outcomes: an analysis of secondary outcomes of a randomized controlled trial. J Occup Environ Med 2015 May;57(5):578-584. [doi: 10.1097/JOM.0000000000000411] [Medline: 25749132]
35. He R, Wei J, Huang K, Yang H, Chen Y, Liu Z, et al. Nonpharmacological interventions for subthreshold depression in adults: a systematic review and network meta-analysis. Psychiatry Res 2022 Nov;317:114897. [doi: 10.1016/j.psychres.2022.114897] [Medline: 36242840]
36. Pané-Farré CA, Fenske K, Stender JP, Meyer C, John U, Rumpf H, et al. Sub-threshold panic attacks and agoraphobic avoidance increase comorbidity of mental disorders: results from an adult general population sample. J Anxiety Disord 2013 Jun;27(5):485-493. [doi: 10.1016/j.janxdis.2013.06.008] [Medline: 23911432]
37. Politis S, Bellos S, Hadjulis M, Gournellis R, Petrikis P, Ploumpidis D, et al. Epidemiology of panic disorder and subthreshold panic symptoms in the Greek general population. Psychiatriki 2020 Oct 1;31(3):201-215 [FREE Full text] [doi: 10.22365/jpsych.2020.313.201] [Medline: 33099461]
38. Oral E, Aydin N, Gulec M, Oral M. Panic disorder and subthreshold panic in the light of comorbidity: a follow-up study. Compr Psychiatry 2012 Oct;53(7):988-994. [doi: 10.1016/j.comppsych.2012.01.009] [Medline: 22425529]
39. Skapinakis P, Lewis G, Davies S, Brugha T, Prince M, Singleton N. Panic disorder and subthreshold panic in the UK general population: epidemiology, comorbidity and functional limitation. Eur Psychiatry 2011 Sep;26(6):354-362. [doi: 10.1016/j.eurpsy.2010.06.004] [Medline: 20813508]

XSL•FO
RenderX

40. Bystritsky A, Kerwin L, Niv N, Natoli JL, Abrahami N, Klap R, et al. Clinical and subthreshold panic disorder. Depress Anxiety 2010 Apr;27(4):381-389 [FREE Full text] [doi: 10.1002/da.20622] [Medline: 20013957]
41. Schotte K, Cooper B. Subthreshold affective disorders: a useful concept in psychiatric epidemiology? Epidemiol Psichiatr Soc 1999 Oct 11;8(4):255-261. [doi: 10.1017/s1121189x00008162] [Medline: 10709316]
42. Helmchen H, Linden M. Subthreshold disorders in psychiatry: clinical reality, methodological artifact, and the double-threshold problem. Compr Psychiatry 2000 Mar;41(2 Suppl 1):1-7. [doi: 10.1016/s0010-440x(00)80001-2] [Medline: 10746897]
43. Hirschfeld RMA. The comorbidity of major depression and anxiety disorders: recognition and management in primary care. Prim Care Companion J Clin Psychiatry 2001 Dec 01;3(6):244-254 [FREE Full text] [doi: 10.4088/pcc.v03n0609] [Medline: 15014592]
44. Choi K, Kim Y, Jeon H. Comorbid anxiety and depression: clinical and conceptual consideration and transdiagnostic treatment. Adv Exp Med Biol 2020;1191:219-235. [doi: 10.1007/978-981-32-9705-0_14] [Medline: 32002932]
45. Cardemil EV, Reivich KJ, Seligman MEP. The prevention of depressive symptoms in low-income minority middle school students. Prev Treat 2002;5(1):8. [doi: 10.1037//1522-3736.5.1.58a]
46. McLaughlin KA. The public health impact of major depression: a call for interdisciplinary prevention efforts. Prev Sci 2011 Dec 6;12(4):361-371 [FREE Full text] [doi: 10.1007/s11121-011-0231-8] [Medline: 21732121]
47. Rennard SI. Minimal clinically important difference, clinical perspective: an opinion. COPD 2005 Mar 24;2(1):51-55 [FREE Full text] [doi: 10.1081/copd-200050641] [Medline: 17136962]

Abbreviations

CBT: cognitive behavioral therapy
CESD: Center for Epidemiologic Studies Depression Scale
DUI: duration of untreated illness
GAD-7: Generalized Anxiety Disorder-7
ICBT: internet-delivered cognitive behavioral therapy
MIC: minimally important difference
PAS: Panic and Agoraphobia Scale
PHQ-9: Patient Health Questionnaire 9
PRO: patient-reported outcome
PSQI: Pittsburgh Sleep Quality Index
RCT: randomized controlled trial
SD: subthreshold depression
SI: subthreshold insomnia
SP: subthreshold panic

Edited by A Mavragani; submitted 17.06.24; peer-reviewed by L Yang, R Nooripour; comments to author 04.11.24; revised version received 11.12.24; accepted 19.12.24; published 03.02.25.

Please cite as:
Taguchi K, Miyoshi M, Seki Y, Baba S, Shimizu E
Effectiveness of Cognitive Behavioral Therapy Provided Through a Web Application for Subthreshold Depression, Subthreshold Insomnia, and Subthreshold Panic: Open-Labeled 6-Arm Randomized Clinical Trial Pilot Study
JMIR Form Res 2025;9:e63139
URL: https://formative.jmir.org/2025/1/e63139
doi:10.2196/63139
PMID:39899369

©Kayoko Taguchi, Mirai Miyoshi, Yoichi Seki, Shiori Baba, Eiji Shimizu. Originally published in JMIR Formative Research (https://formative.jmir.org), 03.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of Electronic Quality Improvement Activities to Reduce Cardiovascular Disease Risk in People With Chronic Kidney Disease in General Practice: Cluster Randomized Trial With Active Control

Jo-Anne Manski-Nankervis[1,2,3], BSc (Hons), MBBS (Hons), PhD; Barbara Hunter[3], BA (Hons), PhD; Natalie Lumsden[3,4], BSc (Hon), PhD; Adrian Laughlin[3], BA (Hons), M (BioS); Rita McMorrow[2,3,5], MBBS, PhD; Douglas Boyle[2,3], BSc (Hons), PhD; Patty Chondros[3], BSc (Hons), MSc, PhD; Shilpanjali Jesudason[6], MBBS, PhD; Jan Radford[7], MBBS, MEd, MPsychMed; Megan Prictor[8], BA, MMus, PhD; Jon Emery[3], BA, MA, MBBS, PhD; Paul Amores[9], BEc (Hons); An Tran-Duy[9,10], MSc, PhD; Craig Nelson[4,11,12], MBBS, PhD

[1]Primary Care and Family Medicine, Lee Kong Chian School of Medicine, Singapore, Singapore

[2]Centre for Research Excellence in Interactive Digital Technology to Transform Australia's Chronic Disease Outcomes, Prahan, Australia

[3]Department of General Practice and Primary Care, University of Melbourne, Melbourne, Australia

[4]Western Health Chronic Disease Alliance, Western Health, Sunshine, Australia

[5]Department of General Practice, University College Cork, Cork, Ireland

[6]Central Northern Adelaide Renal and Transplantation Service, Royal Adelaide Hospital, University of Adelaide, Adelaide, Australia

[7]Launceston Clinical School, University of Tasmania, Launceston, Australia

[8]Melbourne Law School, University of Melbourne, Melbourne, Australia

[9]Centre for Health Policy, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia

[10]Australian Centre for Accelerating Diabetes, University of Melbourne, Melbourne, Australia

[11]Department of Medicine, University of Melbourne, Sunshine, Australia

[12]Department of Nephrology, Western Health, Sunshine, Australia

Corresponding Author:
Jo-Anne Manski-Nankervis, BSc (Hons), MBBS (Hons), PhD
Primary Care and Family Medicine
Lee Kong Chian School of Medicine
11 Mandalay Road
Singapore, 308232
Singapore
Phone: 65 65138572
Email: joanne.mn@ntu.edu.sg

Abstract

Background: Future Health Today (FHT) is a program integrated with electronic medical record (EMR) systems in general practice and comprises (1) a practice dashboard to identify people at risk of, or with, chronic disease who may benefit from intervention; (2) active clinical decision support (CDS) at the point of care; and (3) quality improvement activities. One module within FHT aims to facilitate cardiovascular disease (CVD) risk reduction in people with chronic kidney disease (CKD) through the recommendation of angiotensin-converting enzyme inhibitor inhibitors (ACEI), angiotensin receptor blockers (ARB), or statins according to Australian guidelines (defined as appropriate pharmacological therapy).

Objective: This study aimed to determine if the FHT program increases the proportion of general practice patients with CKD receiving appropriate pharmacological therapy (statins alone, ACEI or ARB alone, or both) to reduce CVD risk at 12 months postrandomization compared with active control (primary outcome).

Methods: General practices recruited through practice-based research networks in Victoria and Tasmania were randomly allocated 1:1 to the FHT CKD module or active control. The intervention was delivered to practices between October 4, 2021, and September 30, 2022. Data extracted from EMRs for eligible patients identified at baseline were used to evaluate the trial outcomes at the completion of the intervention period. The primary analysis used an intention-to-treat approach. The intervention

XSL•FO
RenderX

effect for the primary outcome was estimated with a marginal logistic model using generalized estimating equations with robust SE.

Results: Overall, of the 734 eligible patients from 19 intervention practices and 715 from 21 control practices, 82 (11.2%) and 70 (9.8%), respectively, had received appropriate pharmacological therapy (statins alone, ACEI or ARB alone, or both) at 12 months postintervention to reduce CVD risk, with an estimated between-trial group difference (Diff) of 2.0% (95% CI –1.6% to 5.7%) and odds ratio of 1.24 (95% CI 0.85 to 1.81; *P*=.26). Of the 470 intervention patients and 425 control patients that received a recommendation for statins, 61 (13%) and 38 (9%) were prescribed statins at follow-up (Diff 4.3%, 95% CI 0 to 8.6%; odds ratio 1.55, 95% CI 1.02 to 2.35; *P*=.04). There was no statistical evidence to support between-group differences in other secondary outcomes and general practice health care use.

Conclusions: FHT harnesses the data stored within EMRs to translate guidelines into practice through quality improvement activities and active clinical decision support. In this instance, it did not result in a difference in prescribing or clinical outcomes except for small changes in statin prescribing. This may relate to COVID-19–related disruptions, technical implementation challenges, and recruiting higher performing practices to the trial. A separate process evaluation will further explore factors impacting implementation and engagement with FHT.

Trial Registration: ACTRN12620000993998; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=380119

(JMIR Form Res 2025;9:e54147) doi:10.2196/54147

KEYWORDS

primary care; general practice; clinical decision support; chronic kidney disease; cardiovascular disease; cardiovascular; clinical decision; decision support; support; kidney; kidney disease; electronic medical record; risk; risk reduction; pharmacological; pharmacological therapy; medical records; logistic model

Introduction

More than 4 in 5 Australians visit their general practice at least once per year and 2 million attend every week [1,2]. As medical knowledge continues to increase exponentially, it is crucial that this knowledge is translated into practice efficiently and effectively. Australian general practitioners (GPs) need to have a good working knowledge of 167 conditions to manage 85% of presentations [3], but there is no way for GPs to easily keep up to date with a multitude of guidelines that are stored in different locations while delivering person-centered care. This is critically important for people at risk of, or with, chronic diseases such as chronic kidney disease (CKD), where early detection and management have the potential to reduce disease progression and the development of complications, improve quality of life, and reduce the burden on the health care system [4].

CKD is a common condition, primarily managed in the general practice setting, that is defined as kidney damage reduced function, or both present for at least 3 months [5]. In Australia, CKD affects up to 2 million people and costs the Australian economy more than Aus $5 billion per year (a currency exchange rate as of October 1, 2021, of Aus $1=US $0.726 is applicable) [6]. Untreated CKD can lead to end-stage kidney failure, which may require dialysis or transplant therapies, but the most common cause of death in people with CKD is cardiovascular disease (CVD) [4]. Kidney Health Australia guidelines support the detection of CKD in high-risk populations and recommend pharmacological treatment with angiotensin-converting enzyme inhibitor inhibitors (ACEI) or angiotensin receptor blockers (ARB), or statins or both statins and an ACEI or ARB to reduce CKD progression and CVD risk [5]. It has been identified that additional strategies are required to optimize CVD risk management in general practice [7,8].

Quality improvement (QI) and clinical decision support (CDS) software may help to optimize the care provided to people with CKD in general practice. Building on systematic review evidence for effective elements of these interventions [9,10], we collaborated with GPs, general practice nurses, practice managers, and patients to codesign Future Health Today (FHT) [11]. The FHT software program integrates with the electronic medical records (EMRs) used by over 90% of Australian general practices. It consists of 4 components: a dashboard, a CDS tool, access to resources and QI activities.

This trial aimed to determine whether the FHT platform, packaged with QI support and a series of webinars provided using the Extension for Community Healthcare Outcomes (Project ECHO) program [12] through the Zoom videoconferencing platform, was effective in increasing the proportion of patients with CKD receiving guideline-concordant care to reduce CVD risk at 12 months postrandomization in intervention practices compared with control practices which received an FHT module focused on cancer care. This is a timely trial as policy informing the use of CDS in Australian general practice is currently under development [13].

Methods

Overview

We conducted a stratified cluster randomized head-to-head trial with general practices randomly assigned 1:1 to the FHT CKD program or an active control (FHT cancer program). The full protocol was published [14] and registered with the Australia and New Zealand Clinical Trial Registry (ACTRN12620000993998) [15].

Setting

The FHT trial was conducted in 39 general practices in Victoria and 1 practice in Tasmania, Australia. In Australia, general

XSL•FO
RenderX

practice provides continuous, longitudinal care across the lifespan. The Australian government funds general practice through the Medicare Benefit Schedule (MBS) as a fee-per-service model. The MBS is a list of the medical services for which the Australian Government will make a payment [16], although this may not cover the full cost of the consultation.

The COVID-19 pandemic had a significant impact on general practices, service provision, and patient access during the study period. The government strategy was COVID-19 elimination with frequent lockdowns and border closures. The sixth lockdown in Victoria lasted from August 5, 2021, to October 22, 2021, crossing over the commencement of this trial, which had already been delayed by 6 months due to the pandemic. Significant workload related to COVID-19 infection control procedures and vaccination [17], a shift to telehealth [18], and a significant reduction in pathology services utilization were observed [19].

Recruitment

Forty general practices were recruited between October 2020 and August 2021. Practices were recruited through the VicREN practice-based research and education network at the Department of General Practice and Primary Care, The University of Melbourne [20], the University of Tasmania's Northern Tasmania General Practice-based research network, and through newsletters sent to practices through Primary Health Networks. Following the referral of interested practices by these networks to the study team, the research team liaised with general practice staff to explain the study further and to arrange consent and the completion of an information technology checklist to ensure that the general practice technology infrastructure that was consistent with the requirements of the FHT software.

General practice eligibility criteria are summarized in Textbox 1. Practices needed to have at least 2500 active patients in their EMR system (ie, not marked as deceased or no longer attending the practice) and use either the Best Practice or Medical Director EMR systems. These 2 systems are estimated to be used by more than 90% of Victorian general practices [21]. Practices also needed to agree to install the GRHANITE data extraction tool [22] so that they could contribute deidentified data from their EMR to the Patron dataset [23] to facilitate analysis of the trial outcome measures.

Practices were compensated Aus $2250 for participating in the trial; additional payments were made to GPs and general practice staff if they were nominated as practice champions (Aus $200) or participated in interviews (Aus $50 per interview).

There was no individual patient-level recruitment or consent required as outcomes were ascertained using deidentified EMR data. Inclusion criteria for patient records: Adults aged 18 to 80 years, inclusive, that were not marked as inactive or deceased, with a recorded diagnosis or pathology tests consistent with CKD as defined by Kidney Health Australia [5] that may benefit from pharmacological therapy to reduce CVD risk at baseline. Individuals with a recorded history of renal transplant or chronic dialysis were excluded. Pregnant individuals were also excluded.

Textbox 1. General practice eligibility criteria.

Inclusion criteria

- Agreed to install the GRHANITE data extraction tool [22] to contribute deidentified data from their electronic medical records (EMRs) to the Patron dataset [23].
- Consultations with at least 35 adults aged ≥18 years per day, >2500 active adult patients (defined as patients who attended the general practice at least 3 times in the last 2 years and not marked as deceased) recorded in their EMR or have at least 50 patients that fit cohort definitions for people with chronic kidney disease not on optimal medications and abnormal test results and additional clinical features placing them at risk of an undiagnosed cancer.
- Employed a general practice nurse.
- Contributed (or willing to contribute) data to the Patron dataset, a repository of data from EMR shared by general practices and curated by the University of Melbourne.
- Used Best Practice or Medical Director EMR software to record clinical consultations, prescription of medications, and ordering and receiving pathology results.
- Could identify a workstation (i5/i7 and 16GB RAM or upgradeable to 16GB) with Windows 10 (operating system) that would have GRHANITE, the data extraction tool, installed.
- Computers had Edge or Chrome web browsers installed.

Exclusion criteria

- Had previously participated in other Future Health Today projects.
- Intended to change to another EMR software during the trial period.
- Used a cloud-based EMR system.

Randomization

General practices were allocated 1:1 to the FHT CKD program or the FHT cancer program (active control), using a computer-generated schedule, stratified by Index of Relative Socioeconomic Disadvantage (IRSD) terciles [24] and the number of full-time equivalent (FTE) GPs (4 or fewer vs greater than 4), using random permuted block sizes of 2 and 4 within the stratum. Practices were randomly allocated after they were all recruited, and practices' baseline measures were collected.

XSL•FO
RenderX

The random allocation schedule was generated by the statistician not involved in the practice recruitment or data collection. After practices were recruited, the statistician randomly allocated the deidentified practices to one of the 2 intervention groups and then informed the clinical liaison staff, who advised practices of their allocated group. Practices were informed that it was a pragmatic cluster randomized controlled trial with 2 active groups.

Blinding of general practice participants and the team providing QI support was not possible. The statistician conducting the analysis and study investigators not involved in practice support and engagement remained blinded, and allocation of practices was only revealed after the data were analyzed and results interpreted.

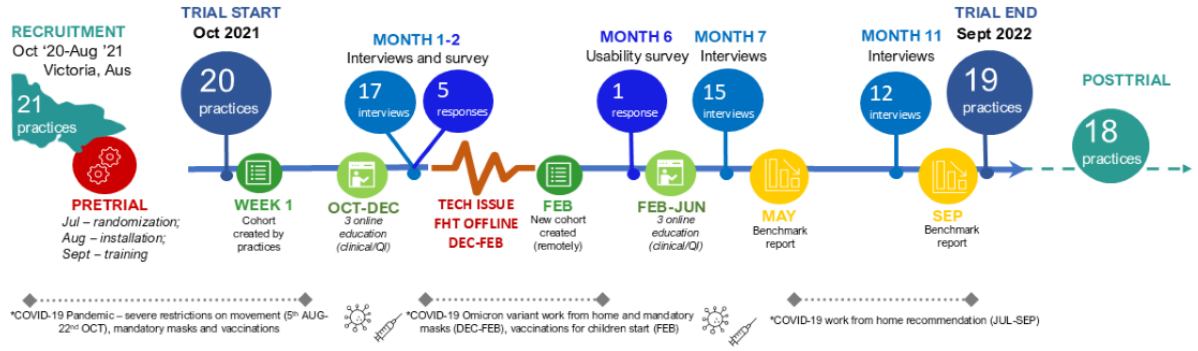
Intervention

Future Health Today Software

FHT is software installed in general practices to identify people who may benefit from the implementation of guideline-informed care. There are two main components: (1) a dashboard that can be used to create cohorts of patients that may benefit from review and potential optimization of care and (2) a CDS tool that deploys when a patient file is opened. Login details and passwords were provided for the dashboard. The CDS tool automatically deployed when a patient file was opened within the EMR without the need for the clinician to log in to FHT. FHT cohorts and CDS prompts are generated each night when code related to best practice guideline recommendations is run over EMR systems in the practice. Screenshots that illustrate the different components of FHT are provided in Multimedia Appendix 1. Implementation activities and timing for delivery are detailed in Figure 1. Following quality assurance testing of the FHT platform and the module algorithms in a virtual environment, FHT was installed in all practices (intervention and control) before the commencement of the trial. Practices were asked to install FHT on at least one GP and one practice nurse computer. A training FHT type 2 diabetes module was deployed and available to both intervention and control practices for approximately 4 weeks before the trial, to ensure that FHT was working correctly from a technical perspective. Pretrial training sessions covering the functionality of the FHT CDS and dashboard components were offered to all practices.

At trial commencement, the training FHT type 2 diabetes module was deactivated and the FHT CKD module was deployed to intervention practices. The CKD module consisted of 3 recommendations for medications that aim to reduce cardiovascular disease risk: “Consider initiation of ACEI or ARB,” “Consider initiation of statin for CKD management,” and “Consider initiation of ACEI or ARB (note: BP in target)” for patients who had an indication for these medications but had not been prescribed them. These recommendations were based on guidelines from Kidney Health Australia [5], the Royal Australian College of General Practitioners [25], and the National Vascular Disease Prevention Alliance [26] and were applied to patients who had a recorded diagnosis of CKD or pathology consistent with CKD, as defined by Kidney Health Australia guidelines [5]. Practices were asked to create a cohort of patients with these recommendations in the dashboard at baseline and asked to use FHT as they chose for the following 12 months.

Figure 1. Trial timeline and data sources (chronic kidney disease group).



Quality Improvement Supports

A practice liaison staff member was assigned to support the intervention practices. This staff member provided software training to practices, and 2 benchmarking reports (6 months and 12 months from trial start) and offered practice guidance in undertaking a quality improvement clinical audit activity related to the intervention. The benchmarking reports summarized the number of patients with recommendations for a specific practice and then compared these to the overall intervention cohort. The practice liaison also coordinated a series of webinars using the Project ECHO program [12], an evidence-based platform that facilitates case-based learning networks with primary care, facilitated by academic medicine departments, and delivered by the Zoom videoconferencing platform. Three separate 1-hour sessions were delivered with a focus on CKD management and reduction of cardiovascular disease risk, and a further three 1-hour sessions were delivered on how to conduct quality improvement activities. The clinical audit and webinar series were accredited by the Royal Australian College of General Practitioners and the Australian College of Rural and Remote Medicine as continuing professional development (CPD) activities. All participants, including nonclinical staff, were provided with certificates of participation in these activities.

Two deviations from the initially proposed protocol occurred, both of which related to the availability of the FHT software, (1) there was a pause in the provision of the CKD module in FHT software from December 16, 2021, to February 14, 2022, due to complexity of the FHT module algorithms and mappings

XSL•FO
RenderX

from CKD not being appropriately implemented within the FHT code. This meant that the software was not available for the full intended 12 months, (2) one practice did not have access to FHT modules until 2 weeks after the trial started due to challenges with software installation.

Consumer Involvement

A consumer advisory group and general practice advisory group (consisting of GPs, general practice nurses, and practice managers) provided input to the trial intervention and interpretation of the trial results.

Endpoints and Data Collection

All outcomes were measured at 12 months postrandomization. The primary outcome was the proportion of eligible patients with a prescription of ACE inhibitors, ARBs statins, or both according to guidelines. Patients were considered to have received guideline-concordant care if they had a recorded prescription for the classes of medication indicated by FHT recommendations within the study period. If a patient received recommendations for both classes of medication at baseline, for the primary outcome, they were required to have been prescribed both classes of medication. If a prescription was not recorded in the EMR, patients were considered not to have received guideline-concordant care. The secondary outcomes related to the prescription of individual classes of these medications, pathology results (including estimated glomerular filtration rate, albumin creatinine ratio, and cholesterol), and systolic blood pressure; refer to Multimedia Appendix 2 for additional details. The number and billing of general practice consultations were also captured. Costs to the government associated with consultations occurring during the trial were based on the Medicare Benefits Schedule (MBS) rebate for the service items retrieved from the Patron database [27].

Baseline and outcome measures were obtained using patient data extracted from general practice EMRs and stored in the Patron database [28]. Patron data are stored at the University of Melbourne and were only accessible by the study statisticians.

Practice characteristics, including the number of GPs and other general practice staff, billing method, and regionality were collected through a survey before randomization.

We also collected survey and qualitative data relating to implementation, to be reported as a process evaluation elsewhere.

Sample Size

We determined that 2580 patients from 20 practices per group are required to detect an absolute 10% increase in the difference in the percentage of patients on optimal pharmacological management in the intervention group at 12 months postintervention compared with the control group at 80% power for a 2-sided 5% significance level. Based on electronic health records from 77 general practices that contribute data to the Patron dataset in Victoria, Australia [28], we expected 55% of active control patients with CKD would be on optimal pharmacological management with an intracluster correlation (ICC) coefficient of 0.03 to account for clustering at the practice level. Further, we expected on average 139 eligible patients per practice (fixed cluster size), with a coefficient of variation of 0.41 to account for variable cluster sizes across the practices. The target of 40 practices included 4 additional practices to allow for attrition by 12 months (eg, practice closure, withdrawals, mergers) and the addition of one extra practice per group, for a degree of freedom correction. The effect size of 10% was considered a minimally important difference and worth implementing and is consistent with another Australian general practice study that aimed to improve CVD risk management, albeit not limited to patients with CKD [29].

Statistical Analysis

Descriptive statistics were used to summarize general practice, clinician, and patient characteristics by study group. The primary analysis used an intention-to-treat approach with practices analyzed according to their allocated group. All general practice patients were included in the analysis, regardless of the occurrence of intercurrent events (eg, death, pregnancy). For the primary outcome, the OR and between-group differences in proportions were estimated with a generalized linear regression model, with the logit and identity link functions, respectively. Generalized estimating equations (GEEs) were used with an exchangeable correlation structure and robust SE to allow for the correlation of outcomes within general practice. The ICC was estimated using a one-way analysis of variance. Binary secondary outcomes used the same method as the primary outcome. For continuous outcomes, a linear mixed effects model was used to estimate the between-group difference in mean change scores with random effects for general practice and individual and adjusted for their respective baseline measure with study group means at baseline constrained to be equal. A proportional odds logistic regression model with GEEs and robust SE was used to estimate cumulative OR for the albumin creatinine ratio level at 12 months. The between-group difference in the rate of consultations per year and rate ratio of visits were modeled using a negative binomial mixed effects model, with random effects for general practice, with GEEs and robust SE.

A 2-part model consisting of a logistic regression model for the probability of positive Medicare benefit and a generalized linear model with a log link and gamma distribution for the positive Medicare benefit was used to assess the association between the intervention and the Medicare benefit. The marginal effect of the intervention was computed in both the logistic regression model (as percentage point change) and the generalized linear regression model (as dollar change).

All regression models included the study group and randomization stratification factors (GP, FTE, and IRSD tercile) as covariates. For secondary binary outcomes, the recommendation to initiate statin and ACEI or ARB was adjusted for the recommendation to consider ACEI or ARB and statin initiation, respectively. Sensitivity analyses included (1) complete cases only (for continuous outcomes, if the linear mixed effects model failed to converge, linear regression with robust SE was used) and (2) adjusting for patient age and sex and previous practice involvement in formalized QI programs. For the binary outcomes, if the model used to estimate the risk difference failed to converge, the additional covariates were

XSL•FO
RenderX

excluded (noted in the results tables). Adherence-adjusted analysis was not conducted as we were unable to measure adherence to the intervention.

Estimated intervention effects were reported with 95% CI and *P* values. No corrections were made for multiple comparisons.

All analyses were conducted using Stata 17.0 (StataCorp). The full statistical analysis plan has been uploaded to the ANZCTR.

Ethical Considerations

The FHT trial was approved by the Faculty of Medicine, Dentistry and Health Sciences Human Ethics Sub-Committee at the University of Melbourne (Ethics ID:2056564). General practices signed a study agreement to participate in the trial. Patients were not recruited for this study, with data from the Patron dataset used to calculate the primary and secondary trial outcomes. The Patron dataset has a waiver of consent granted by the Faculty of Medicine, Dentistry and Health Sciences Human Ethics Sub-Committee at the University of Melbourne (ethics number 2023-23358-37538-6). Data within the Patron dataset is anonymized. General practices were paid Aus $2250 to compensate them for any IT support, meeting or training attendance, generating patient lists, mailouts, monthly key staff informant interviews, and participation in quality improvement activities and ECHO education activities.

Results

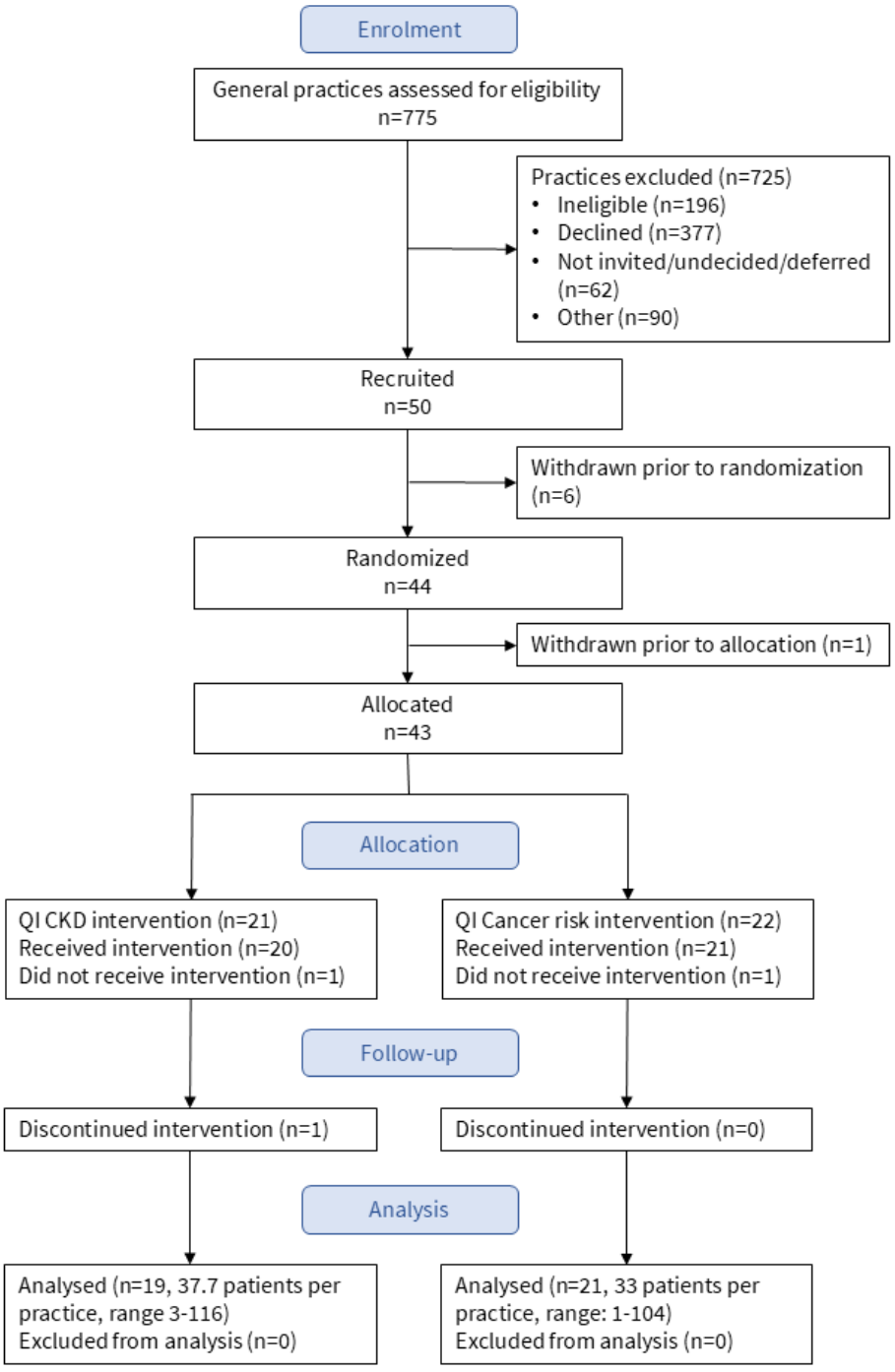
Participating Practices

Between October 2020 and August 2021, there were 775 general practices assessed for eligibility, following which 50 practices were recruited to participate in the trial (refer to Figure 2). Six practices withdrew before randomization (all these practices were owned by the same group), and a practice withdrew after randomization but before they were notified of their group allocation. Two randomized practices did not receive the allocated intervention and 1 discontinued the intervention during the trial, leaving forty participating practices that completed the trial. Patron data was only provided for the forty practices that completed the study.

Baseline characteristics of the 40 practices, clinicians, and staff (Table 1) and their eligible patients (Table 2) were balanced between the study groups. A total of 734 and 715 eligible patients in the intervention and control groups, respectively, were identified in the EMR with a recorded diagnosis or pathology tests consistent with CKD that may benefit from pharmacological therapy to reduce CVD risk. Of these, 17% had a recommendation for both ACEI or ARB and satins, 45% for statins only, and the remaining 38% for ACEI/ARBS only.

XSL•FO
RenderX

Figure 2. Practice recruitment, allocation, follow-up, and data analysis.



XSL•FO
RenderX

Table 1. Baseline characteristics of practices, clinicians, managers, and staff, overall and by study group.

	All practices (N=40)	Intervention group (N=19)	Control group (N=21)
Practice characteristics			
State, n (%)			
Victoria	39 (97.5)	19 (100)	20 (95.2)
Tasmania	1 (2.5)	0 (0)	1 (4.8)
Relative socioeconomic disadvantage index (terciles), n (%)			
1 (most disadvantaged)	12 (30)	6 (31.6)	6 (28.6)
2	12 (30)	6 (31.6)	6 (28.6)
3 (least disadvantaged)	16 (40)	7 (36.8)	9 (42.8)
Practice size			
4 or fewer FTE[a] GPs[b], n (%)	21 (52.5)	9 (47.4)	12 (57.1)
Greater than 4 FTE GPs, n (%)	19 (47.5)	10 (52.6)	9 (42.9)
Previous quality improvement program participation, n (%)	18 (45)	9 (47.4)	9 (42.9)
Eligible patients per practice, mean (range)	35.3 (1-116)	37.7 (3-116)	33.0 (1-104)
Workstations with Future Health Today per practice, mean (range)[c]	6.4 (1-20)	7.4 (1-20)	5.5 (1-19)
Future Health Today users per practice, mean (range)	31.8 (10-160)	30.2 (14-73)	33.2 (10-160)
Number of FTE GPs, median (IQR)	4.5 (3.5-6.0)	4.5 (3.5-7.5)	4.0 (3.5-5.5)
Number of FTE practice nurses, median (IQR)	1.8 (1.0-2.5)	1.5 (1.0-3.0)	2.0 (1.5-2.5)
Number of FTE practice managers or administrative staff, median (IQR)	4.0 (3.0-5.0)	4.5 (3.5-6.0)	3.5 (2.5-4.5)
General practitioners	309	164	145
Sex, n (%)			
Male	178 (57.6)	93 (56.7)	85 (58.6)
Female	131 (43.4)	71 (43.3)	60 (41.4)
Age, n (%)			
<35 years	40 (15)	19 (14.2)	21 (15.8)
35-50 years	129 (48.3)	56 (41.8)	73 (54.9)
>50 years	98 (36.7)	59 (44)	39 (29.3)
Nurses	123	64	59
Sex, n (%)			
Male	7 (5.7)	3 (4.7)	4 (6.8)
Female	116 (94.3)	61 (95.3)	55 (93.2)
Age[d], n (%)			
<35 years	60 (49.2)	33 (52.4)	32 (50)
35-50 years	31 (25.4)	14 (22.2)	17 (26.6)
>50 years	31 (25.4)	16 (25.4)	15 (23.4)
Practice managers or administrative staff	257	130	127
Sex, n (%)			
Male	27 (10.5)	14 (10.8)	13 (10.2)
Female	230 (89.5)	116 (89.2)	114 (89.8)
Age, n (%)			
<35 years	130 (51.6)	75 (60)	55 (43.3)
35-50 years	59 (23.4)	26 (20.8)	33 (26)

XSL•FO
RenderX

	All practices (N=40)	Intervention group (N=19)	Control group (N=21)
>50 years	63 (25)	24 (19.2)	39 (30.7)

[a]FTE: full-time equivalent.

[b]GPs: general practitioner.

[c]Six intervention and 2 control group practices ran a “terminal server” that pushed Future Health Today to all workstations.

[d]Percentages for age categories were calculated based on the number of staff who provided age data.

XSL•FO
RenderX

Table 2. Baseline characteristics of patients, overall and by study group.

Patient characteristics	All participants (N=1449)	Intervention group (n=734)	Control group (n=715)
Sex, n (%)			
Male	812 (56)	402 (54.8)	410 (57.3)
Female	637 (44)	332 (45.2)	305 (42.7)
Age (years), mean (SD)	65.9 (12.1)	66.3 (11.3)	65.4 (12.9)
Prescribed ACEI[a] or ARB[b] , n (%)			
ACEI only	215 (14.8)	109 (14.9)	106 (14.8)
ARB only	321 (22.2)	166 (22.6)	155 (21.7)
ACEI and ARB	6 (0.4)	4 (0.6)	2 (0.3)
None	907 (62.6)	455 (62)	452 (63.2)
Prescribed statins, n (%)			
Yes	424 (29.3)	192 (26.2)	232 (32.5)
No	1025 (70.7)	542 (73.8)	483 (67.5)
Guideline recommended initiation, n (%)			
Both ACEI or ARB and statins	246 (17)	131 (17.9)	115 (16.1)
ACEI or ARBs only	554 (38.2)	264 (36)	290 (40.5)
Statin only	649 (44.8)	339 (46.2)	310 (43.4)
Cardiovascular disease risk category, n (%)			
Total sample	1065 (73.5)	539 (73.4)	526 (73.8)
Low	85 (8)	43 (7.6)	42 (8)
Moderate	29 (2.7)	14 (2.7)	15 (2.9)
High	951 (89.3)	482 (89.7)	469 (89.2)
Type 2 diabetes n (%)	633 (43.7)	308 (42)	325 (45.5)
Systolic blood pressure (mmHg)			
Number (%)	843 (58.2)	434 (59.1)	409 (57.2)
Mean (SD)	138 (20)	138 (20)	137 (19)
Lipids (all mmol/L)			
Total cholesterol, n (%)	916 (63.2)	465 (63.4)	451 (63.1)
Total cholesterol, mean (SD)	4.8 (1.2)	4.8 (1.2)	4.8 (1.3)
Low-density lipoproteins, n (%)	844 (58.2)	431 (58.7)	413 (57.8)
Low-density lipoproteins, mean (SD)	2.7 (1.1)	2.7 (1.1)	2.6 (1.1)
High-density lipoproteins, n (%)	865 (59.7)	444 (60.5)	421 (58.9)
High-density lipoproteins, mean (SD)	1.3 (0.4)	1.3 (0.4)	1.3 (0.4)
Triglycerides, n (%)	914 (63.1)	464 (63.2)	450 (62.9)
Triglycerides mean (SD)	2.1 (1.8)	2.0 (1.4)	2.1 (2.1)
Urine albumin-creatinine ratio (mg/mmol)[c]			
Number (%)	702 (48.4)	323 (44)	379 (53)
Median (IQR)[c]	5.5 (1.6-21.7)	5.2 (1.4-22.5)	5.6 (1.7-20.2)
Macroalbuminuria, n (%)	137 (19.5)	64 (19.8)	73 (19.3)
Microalbuminuria, n (%)	343 (48.9)	150 (46.4)	193 (50.9)
Normal, n (%)	222 (31.6)	109 (33.8)	113 (29.8)
Estimated glomerular filtration rate (mL/min/1.73m^2)			

XSL•FO
RenderX

Patient characteristics	All participants (N=1449)	Intervention group (n=734)	Control group (n=715)
Number (%)	1175 (81.1)	592 (80.7)	583 (81.5)
Mean (SD)	55.3 (20.9)	53.5 (20.4)	57.1 (21.2)
G1 (≥90), n (%)	149 (12.7)	61 (10.3)	98 (15.1)
G2 (60-89), n (%)	224 (19.1)	105 (17.7)	119 (20.4)
G3a (45-59), n (%)	428 (36.4)	221 (37.3)	207 (35.5)
G3b (30-44), n (%)	267 (22.7)	146 (24.7)	121 (20.8)
G4 (15-29), n (%)	73 (6.2)	37 (6.3)	36 (6.2)
G5 (<15), n (%)	35 (2.9)	22 (3.7)	12 (2.1)

[a]ACEI: angiotensin-converting enzyme inhibitor.

[b]ARB: angiotensin receptor blocker.

[c]Median and IQR were reported due to skewed distribution.

Primary and Secondary Outcomes

Primary Outcome

Of patients identified as eligible for pharmacological therapy to reduce CVD risk, 11.2% (82/734) in the intervention group and 9.8% (70/715) in the control group had received appropriate pharmacological therapy at 12 months post-intervention, an increase of 2.0% (95% CI –1.6% to 5.7%) (Table 3), with the confidence bounds excluding our hypothesized minimally important difference of 10%. The ICC for the proportion of patients who were prescribed ACEI or ARB and Statin at baseline was estimated as 0.03 (95% CI 0.01 to 0.05).

Table 3. Primary outcome and secondary outcomes related to medication prescriptions

	All participants, n (%)	Intervention group, n (%)	Control group, n (%)	Estimated effect size (95% CI)		*P* value
				Diff[a] (95% CI), %	OR[b] (95% CI)	
Primary outcome						
Angiotensin-converting enzyme inhibitor or angiotensin receptor blocker and statins						
Number of participants	1449	734	715			
Primary analysis	152 (10.5)	82 (11.2)	70 (9.8)	2.0 (–1.6 to 5.7)	1.24 (0.85 to 1.81)	.26
Sensitivity analysis[c]	152 (10.5)	82 (11.2)	70 (9.8)	2.6 (0.0 to 6.1)	1.30 (0.90 to 1.87)	.17
Secondary outcomes						
Angiotensin-converting enzyme inhibitor or angiotensin receptor blocker						
Number of participants	800	395	405			
Primary analysis	100 (12.5)	48 (12.2)	52 (12.8)	0.0 (–3.7 to 4.6)	0.94 (0.62 to 1.42)	.76
Sensitivity analysis[a]	100 (12.5)	48 (12.2)	52 (12.8)	1.8 (–2.6 to 6.2)	1.04 (0.69 to 1.57)	.85
Statins						
Number of participants	895	470	425			
Primary analysis	99 (11.1)	61 (13)	38 (9)	4.3 (0 to 8.6)	1.55 (1.02 to 2.35)	.04
Sensitivity analysis[a]	99 (11.1)	61 (13)	38 (9)	4.4 (0 to 8.5)	1.56 (1.04 to 2.34)	.03

[a]Diff: difference in mean change score or percentages between the 2 groups.

[b]OR: odds ratio.

[c]Sensitivity analysis 1, as above, with adjustment for participant age and sex and general practitioner participation in formal quality improvement activities.

Secondary Outcomes

When we examined the patients with recommendations for statins and ACEI or ARBS, separately, 13% (61/470) of patients in the intervention group received a statin prescription compared to 9% (38/425) in the control group, an estimated between-group difference of 4.3% (95% CI 0%-8.6%). The sensitivity analyses

XSL•FO
RenderX

showed similar results. Primary analyses are generalized linear regression using generalized estimating equations with robust SE, adjusted for GP FTE and IRSD tercile.

Table 4 shows that there was no statistical evidence of an intervention effect for mean change in pathology measures from baseline between the intervention and control groups. Primary analyses for continuous measures used a linear mixed effects model with random effects for general practice and individuals, adjusted for baseline measure of the outcome. Primary analyses for binary outcomes used generalized linear regression using generalized estimating equations with robust standard errors. All primary analyses adjusted for GP FTE and IRSD tercile.

Table 5 summarizes the nonpathology measures of secondary outcomes. There was no evidence of an effect of the intervention on systolic blood pressure change, cardiovascular disease risk category, or rate of general practice visits of patients over the 12-month intervention period compared to the control group. Primary analyses for continuous measures used a linear mixed effects model with random effects for general practice and individuals and adjusted for baseline measures of the outcome. Primary analyses for ordinal outcomes used proportional odds logistic regression, using generalized estimating equations with robust standard errors. The primary analysis of the rate of visit data used a negative binomial mixed effects model with random effects for general practice. All primary analyses adjusted for GP FTE and IRSD tercile.

XSL•FO
RenderX

Table 4. Secondary outcomes for laboratory pathology measures.

	All participants	Intervention group	Control group	Estimated effect size, (95% CI)	*P* value
Cholesterol change (mmol/L), mean (SD)					
Sample size	563	260	303		
Primary analysis	0.14 (0.94)	0.07 (0.84)	0.20 (1.01)	Diff[a] –0.11 (–0.28 to 0.07)	.24
Sensitivity analysis[b]	0.14 (0.94)	0.07 (0.84)	0.20 (1.01)	Diff –0.09 (–0.27 to 0.09)	.32
Sensitivity analysis[c,d]	0.14 (0.94)	0.07 (0.84)	0.20 (1.01)	Diff –0.13 (–0.27 to 0.03)	.06
Low-density lipoproteins change (mmol/L), mean (SD)					
Sample size	484	230	254		
Primary analysis	0.13 (0.75)	0.05 (0.67)	0.20 (0.82)	Diff –0.10 (–0.26 to 0.06)	.21
Sensitivity analysis[b]	0.13 (0.75)	0.05 (0.67)	0.20 (0.82)	Diff –0.10 (–0.26 to 0.06)	.24
Sensitivity analysis[c]	0.13 (0.75)	0.05 (0.67)	0.20 (0.82)	Diff –0.13 (–0.26 to 0.00)	.05
High-density lipoproteins change (mmol/L), mean (SD)					
Sample size	518	247	271		
Primary analysis	0.00 (0.21)	–0.01 (0.20)	0.01 (0.22)	Diff –0.04 (–0.09 to 0.02)	.17
Sensitivity analysis[b]	0.00 (0.21)	–0.01 (0.20)	0.01 (0.22)	Diff –0.05 (–0.10 to 0.01)	.09
Sensitivity analysis[c]	0.00 (0.21)	–0.01 (0.20)	0.01 (0.22)	Diff –0.03 (–0.07 to 0.01)	.11
Triglycerides change (mmol/L), mean (SD)					
Sample size	560	258	302		
Primary analysis	0.07 (1.61)	0.06 (0.88)	0.07 (2.03)	Diff 0.09 (–0.11 to 0.29)	.36
Sensitivity analysis[b]	0.07 (1.61)	0.06 (0.88)	0.07 (2.03)	Diff 0.15 (–0.05 to 0.35)	.15
Sensitivity analysis[c]	0.07 (1.61)	0.06 (0.88)	0.07 (2.03)	Diff 0.09 (–0.11 to 0.29)	.39
Urine albumin-creatinine ratio change (mg/mmol), mean (SD)					
Sample size	398	178	220		
Primary analysis	–2.13 (41.80)	–6.56 (42.99)	1.45 (40.56)	Diff –9.11 (–21.36 to 3.14)	.15
Sensitivity analysis[b]	–2.13 (41.80)	–6.56 (42.99)	1.45 (40.56)	Diff –8.97 (–21.23 to 3.31)	.15
Sensitivity analysis[c]	–2.13 (41.80)	–6.56 (42.99)	1.45 (40.56)	Diff –6.69 (–14.33 to 0.95)	.09
Urine albumin-creatinine ratio 30% reduction					
Sample size	398	178	220		
Primary analysis, n (%)	117 (29.4)	54 (30.3)	63 (28.6)	Diff 3.1% (–5.7% to 12.0%); OR[e] 1.14 (.74 to 1.76)	.55
Sensitivity analysis[b], n (%)	117 (29.4)	54 (30.3)	63 (28.6)	Diff 2.4% (–6.8% to 11.5%); OR 1.09 (.69 to 1.73)	.70
Estimated glomerular filtration rate change (mL/min/1.73^2), mean (SD)					
Sample size	894	437	457		
Primary analysis	0.10 (8.47)	–0.27 (8.30)	0.44 (8.64)	Diff 1.35 (–1.15 to 3.86)	.29
Sensitivity analysis[b]	0.10 (8.47)	–0.27 (8.30)	0.44 (8.64)	Diff 1.04 (–1.37 to 3.45)	.40
Sensitivity analysis[c]	0.10 (8.47)	–0.27 (8.30)	0.44 (8.64)	Diff –0.56 (–1.75 to 0.64)	.36

[a]Diff: difference in mean change score or percentages between the 2 groups.

[b]Sensitivity analysis 1 additionally adjusts for participant age and sex, and practice participation in formal QI activities.

[c]Sensitivity analysis 2 using complete case only for outcomes with missing data.

[d]Linear regression model with robust SE used as an alternative to linear mixed effects model due to convergence issues.

[e]OR: odds ratio.

XSL•FO
RenderX

Table 5. Secondary outcomes for nonpathology measures.

	All participants	Intervention group	Control group	Estimated effect size (95% CI)	*P* value
Systolic blood pressure changes from baseline mmHg, mean (SD)					
Sample size	676	335	341		
Primary analysis	2.29 (20.26)	3.11 (20.66)	1.47 (19.86)	Diff[a] 0.75 (–1.71 to 3.21)	.55
Sensitivity analysis[b]	2.29 (20.26)	3.11 (20.66)	1.47 (19.86)	Diff 0.88 (–1.58 to 3.35)	.49
Sensitivity analysis[c]	2.29 (20.26)	3.11 (20.66)	1.47 (19.86)	Diff 1.00 (–1.49 to 3.48)	.43
Cardiovascular d isease risk category, n (%)					
Sample size	1019	516	503		
Primary analysis	—[d]	—	—	Cumulative OR[e] 0.89 (0.64 to 1.22)	.47
Low CVD[f] risk	101 (9.9)	48 (9.3)	53 (10.5)		
Medium CVD risk	39 (3.8)	23 (4.5)	16 (3.2)		
High CVD risk	879 (86.3)	445 (86.2)	434 (86.3)		
Sensitivity analysis[b]	—	—	—	Cumulative OR 0.81 (0.60 to 1.10)	.18
Rate of visits, n (%)					
Sample size	1447	734	713		
Number of visits, mean (SD)	17,237	7786	9451		
Primary analysis	11.91 (14.45)	10.61 (12.16)	13.26 (16.37)	Diff –0.14 (–0.38 to 0.11); RR[g] 0.87 (0.68 to 1.11)	.27
Sensitivity analysis[b]	11.91 (14.45)	10.61 (12.16)	13.26 (16.37)	Diff –0.15 (–0.38 to 0.08); RR 0.86 (0.69 to 1.08)	.20

[a]Diff: difference in mean change score or percentages between the 2 groups.

[b]Sensitivity analysis 1 additionally adjusts for participant age and sex and practice participation in formal quality improvement activities.

[c]Sensitivity analysis 2 using complete cases for outcomes with missing data.

[d]Not applicable.

[e]OR: odds ratio.

[f]CVD: cardiovascular disease.

[g]RR: rate ratio.

Health Care Costs

Table 6 summarizes the health economic outcomes. Of 734 patients in the intervention group, 541 (73.7%) had at least one general practice consultation associated with an MBS rebate, compared to 80.6% (576/715) patients in the control group. Patients meeting the criteria for the CKD FHT module in the practices receiving the intervention had a 6.7 percentage point lower probability of having at least one general practice consultation associated with an MBS rebate compared to those in the control group, but there was no evidence of a difference (95% CI –0.26 to 0.13).

Patients in the intervention group who claimed at least one MBS rebate during the trial period had Aus $38.55 lower Medicare benefits compared with the control group over 12 months, roughly equivalent to one standard consultation in general practice lasting up to 20 minutes. There was no evidence of a difference between groups (95% CI Aus $328.46 to Aus $251.35; *P*=0.79). Primary analyses for binary outcomes used logistic regression, with robust SE. Primary analysis of positive Medicare benefits was carried out with a generalized linear model with log link and gamma distribution, with robust standard errors. All primary analyses adjusted for GP FTE and IRSD tercile.

XSL•FO
RenderX

Table 6. Health economic outcomes.

	All participants	Intervention group	Control group	Estimated effect size (95% CI)	*P* value
Probability of incurring Medicare Benefits Schedule rebate (logistic regression), n (%)					
Sample size	1449	734	715		
Primary analysis	1117 (77.1)	541 (73.7)	576 (80.6)	Diff[a] –0.067 (–0.260 to 0.125); odds ratio 0.719 (0.285 to 1.82)	.49
Sensitivity analysis[b]	1117 (77.1)	541 (73.7)	576 (80.6)	Diff –0.094 (–0.337 to 0.148; odds ratio 0.659 (0.236 to 1.84)	.45
Positive Medicare Benefits Schedule rebate (generalized linear model), Aus $, mean (SE)					
Sample size	1117	541	576		
Primary analysis	847.28 (23.01)	850.96 (33.33)	843.84 (31.81)	Diff –38.555 (–328.463 to 251.353)	.79
Sensitivity analysis[b]	847.28 (23.01)	850.96 (33.33)	843.84 (31.81)	Diff 0.880 (–255.905 to 257.666)	.99

[a]Diff: difference in mean change score or percentages between the 2 groups.

[b]Sensitivity analysis 1 additionally adjusts for participant age and sex and practice participation in formal quality improvement activities (and a dummy variable for missing quality improvement practice participation).

Discussion

Principal Results

Our findings do not support that the intervention was effective in increasing the proportion of eligible patients with optimal pharmacological therapies at 12 months postintervention to reduce their CKD risk. Among general practice adult patients identified in the EMR at baseline with a recorded diagnosis or pathology tests consistent with CKD that may benefit from an ACEI or ARB or statin prescription or both to reduce the risk of CKD based on guidelines, we found that 11.2% and 9.8% of eligible patients received appropriate pharmacological therapy at 12 months post-randomization in those receiving the FHT CKD module compared to active control, an estimated 2% difference between the intervention and control groups.

For participants with a recommendation to initiate a statin, 13.0% received a prescription for statin medication in the intervention group compared to 9.0% of control patients, an estimated 4.3% (95% CI 0% to 8.6%) increase in statin prescribing when initiation was recommended. Although the confidence bound excluded the 10% effect size hypothesized as minimally important for optimal pharmacological therapies, the confidence bounds include improvement in statins prescribing when recommended which could potentially impact reducing cardiovascular events [30], particularly when considering the estimated effect size may be attenuated as a result of several extenuating circumstances.

Circumstances that may have impacted the trial findings include: pausing access of practices to the CKD module from December 16, 2021, to February 14, 2022, due to the complexity of the FHT module algorithms and mappings from CKD not being appropriately implemented within the FHT code; challenges with software installation in some practices with one practice not having access to FHT modules until two weeks after trial start; the number of patients identified in the module being approximately half that proposed in the target sample size; and variation in some of the coding implemented in FHT compared with that provided in the business requirements document. A sensitivity analysis conducted on the cohort of patients who would have been identified, had the algorithms been implemented without error, did not change the trial findings (data not shown).

Further, our trial was conducted during the COVID-19 pandemic which included stay-at-home orders and general practices needed to deliver COVID-19 vaccinations. As a result, there were unprecedented workforce pressures that may have impacted the ability of practices to use FHT and reduced patient attendance at general practice. GPs may have had more confidence in prescribing statins than ACEI or ARB as these medications do not require pathology tests postinitiation, which may have been an additional consideration during the pandemic.

Strengths and Limitations

Our study has several strengths. First, this was a pragmatic trial and so the results that are obtained are likely to reflect the impact of FHT in real-world practice. Forty practices were included in the analysis, although we had anticipated 36 after allowing for attrition. EMR data were used to capture patient data and outcome measures minimizing selection bias. Without this EMR data capture, the trial would have been logistically impossible due to the impact of COVID-19 on the general practice workforce and the COVID-19 restrictions that were in place restricted travel and access to medical practices by researchers during this period. Our light touch intervention was designed to overcome the known financial and scalability barriers to more traditional academic detailing. Finally, randomizing practices rather than individuals minimized the risk of contamination.

Study limitations included cluster sizes being smaller than anticipated and there was a lower percentage of eligible patients

XSL•FO
RenderX

receiving appropriate pharmacological therapy at 12 months in both groups than expected. Possible reasons include that we may have attracted higher performing practices to participate in the trial, so there were smaller proportions of patients with CKD who were not being managed according to guidelines; improvements in coding of the EMR data between the datasets used to estimate the parameters to determine the sample size calculations and then used to measure the primary outcome; variation in the quality of recording of data in EMRs within participating general practices; or fewer patients than anticipated may have attended the general practice during the COVID-19 pandemic due to the Victorian lockdowns. However, the smaller cluster sizes for the 40 practices and lower percentages receiving appropriate pharmacological therapy improved study efficiency, allowing us to address the research question and provide sufficient certainty to rule out the target 10% intervention effect. Further, among patients who did not receive appropriate pharmacological therapy, we were unable to determine if the GP did not consider that medications were indicated for other reasons or if the patient chose not to take the medications.

Comparison with Previous Work

QI activities and CDS form an important part of Australian and international primary care policy, with the aim of optimizing the care provided to patients through the strengthening of clinical decision-making and improving care processes. There is increased focus on QI programs in Australian general practices, with payments to practices for participating in QI activities commencing in 2019 [31] and exploration of the use and regulation of CDS in general practice as the focus of current Australian government consultation [13].

The advent of EMRs has resulted in the development of QI and CDS software internationally and across care settings. Australian general practices were early adopters of EMRs in the 1990s, with near universal computerization by 2006 [21]. Similar to our trial, the results of trials of the use of QI dashboards and clinical decision support software (CDSS) tools in primary care to facilitate guideline-recommended prescribing have been underwhelming. For example, the TORPEDO cluster randomized trial in Australian primary care found that the proportion of patients with a recording of risk factors for CVD increased, but the proportion of patients at high CVD risk that received recommended prescriptions was not significantly different from the control group [29]. Hespe's mixed methods descriptive study of a QI program in the Central and Eastern Sydney Primary Health Network area aimed to improve monitoring, prescribing practices, and attainment of BP and lipid targets for CVD risk reduction. The intervention included plan-do-study act cycles supported by a QI officer, workshops, an audit tool, and the HealthTracker guideline recommendation tool, but did not result in a change in the proportion of patients prescribed blood pressure and lipid-lowering medications over the 22-month intervention period.[24] The EBMeDS CDSS trial did not result in improvements in the management of diabetes in Belgian primary care despite high physician satisfaction [32].

Overall, the effectiveness of CDSS is generally small to moderate and highly variable [33] with the predictors of meaningful improvement unclear [34]. A systematic review of 11 trials incorporating clinical dashboards found limited evidence for their use, especially if they were not a component of a multifaceted intervention [35].

All of these studies highlight the challenges of implementation of QI dashboards and CDSS in general practice, including lack of time, practices' need for technical support, a perceived lack of value for QI work, difficulty disseminating knowledge across the practice team, tensions between the team and clinical staff, workforce, a need for strong IT infrastructure and technical support, organization of practices as providing reactive (as opposed to proactive, planned) care and suboptimal recording of patient data in the EMR [24,32,36].

Our experiences with this trial have resulted in more robust mechanisms to validate our algorithms that incorporate the use of comprehensive data from Patron. We think this, combined with greater utility through supporting a wider range of decision support capabilities, may increase utilization of FHT and offer a greater chance of demonstrating the effectiveness of the FHT CDSS going forward. We will explore factors relevant to the FHT trial in a process evaluation and implementation study to be published separately. Upcoming studies of different conditions, including the results of the cancer module active control in this trial, will provide further information regarding the implementation and effectiveness of FHT in Australian general practice.

Conclusions

In our trial, the CKD FHT intervention found no differences in trial primary or secondary outcomes except for a small improvement in prescribing of statin medication when initiation was recommended which may reduce future cardiovascular events. These results may be reflective of several factors, including the impact of the COVID-19 pandemic on general practice, technical implementation challenges, and sampling from high-performing practices. Which may have attenuated the intervention effect.

Acknowledgments

We would like to thank the general practices, health care professionals, and consumers that provided input to the development of FHT, piloted the tool in their practices, and participated in this trial. We are grateful to Christine Hallinan (University of Melbourne) who undertook the sample size calculations, and Antonia Arfaras, Ralph Audehm, Nicholas Brayshaw, Jade Cartwright, Deborah Daly, Ian Dennis, Julian Fidge, John Hodgson, George Jiang, Fiona Kleinitz, Eileen Lam, Ed Skinner and Richard Smith, who provided important insights to this trial and its interpretation through their involvement in our consumer and general practice advisory groups. Deidentified patient data from the Patron primary care data repository (extracted from consenting general

XSL•FO
RenderX

practices), which has been created and is operated by the Department of General Practice and Primary Care, The University of Melbourne was used to underpin trial feasibility, sample size calculations and to evaluate the outcomes of the trial.

The Paul Ramsay Foundation has provided funding for staff costs related to the technical development of FHT and the Implementation and Evaluation team members. They have also provided funding for professional staff support and general practice and consumer advisory groups. JDE is supported by an NHMRC (National Health and Medical Research Council) Investigator grant (APP1195302). The active control group of the trial was supported by the Cancer Australia-funded Primary Care Collaborative Cancer Clinical Trials Group (PC4). The trial is affiliated with the CanTest Collaborative (funded by Cancer Research UK C8640/A23385) of which JDE is an Associate Director. The design, management, analysis, and reporting of the study are entirely independent of the funders.

Generative AI was not used in any portion of the manuscript writing.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to them containing sensitive health information but may be available from the Department of General Practice and Primary Care, University of Melbourne subject to ethical and data governance approvals.

Conflicts of Interest

FHT software was developed by The University of Melbourne in collaboration with Western Health. These affiliations were displayed on the FHT dashboard and training materials. CN, JE, and JAM-N are the coleads of the Future Health Today program and DB leads the technical team that developed the software. Intellectual property related to FHT is owned by The University of Melbourne.

Multimedia Appendix 1
Screenshots of the Future Health Today dashboard and clinical decision support tool.
[PPTX File , 2886 KB - formative_v9i1e54147_app1.pptx]

Multimedia Appendix 2
Primary, secondary, and health economics outcomes at 12 months postrandomization for quality improvement chronic kidney disease program.
[DOCX File , 23 KB - formative_v9i1e54147_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2582 KB - formative_v9i1e54147_app3.pdf]

References

1. Hayes P. No one knows you like your GP. newsGP. 2018. URL: https://www1.racgp.org.au/newsgp/racgp/no-one-knows-you-like-your-gp [accessed 2024-12-24]
2. Britt H, Henderson J, Bayram C, Harrison C, Valenti L, Pan Y, et al. General practice activity in Australia 2014-15. In: General Practice Series No. 38. Sydney: Sydney University Press; 2015.
3. Cooke G, Valenti L, Glasziou P, Britt H. Common general practice presentations and publication frequency. Aust Fam Physician 2013;42(1-2):65-68 [FREE Full text] [Medline: 23529466]
4. Cardiovascular disease, diabetes and chronic kidney disease - Australian facts: prevalence and incidence, in cardiovascular, diabetes and chronic kidney disease series no. 2. Australian Institute of Health and Welfare (AIHW). 2014. URL: https://www.aihw.gov.au/reports/heart-stroke-vascular-diseases/cardiovascular-diabetes-chronic-kidney-mortality/summary [accessed 2014-10-24]
5. Kidney Health Australia. Chronic kidney disease (CKD) management in primary care (4th ed). Melbourne: Kidney Health Australia; 2020.
6. National Strategic Action Plan for Kidney Disease. Kidney Health Australia. 2019. URL: https://kidney.org.au/get-involved/advocacy/national-strategic-action-plan-for-kidney-disease [accessed 2024-12-24]
7. Razavian M, Heeley EL, Perkovic V, Zoungas S, Weekes A, Patel AA, et al. Cardiovascular risk management in chronic kidney disease in general practice (the AusHEART study). Nephrol Dial Transplant 2012;27(4):1396-1402. [doi: 10.1093/ndt/gfr599] [Medline: 22053091]
8. Hespe CM, Campain A, Webster R, Patel A, Rychetnik L, Harris MF, et al. Implementing cardiovascular disease preventive care guidelines in general practice: an opportunity missed. Med J Aust 2020;213(7):327-328. [doi: 10.5694/mja2.50756] [Medline: 32880959]

XSL•FO
RenderX

9. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev 2012;2012(6):CD000259. [doi: 10.1002/14651858.CD000259.pub3] [Medline: 22696318]
10. Van de Velde S, Heselmans A, Delvaux N, Brandt L, Marco-Ruiz L, Spitaels D, et al. A systematic review of trials evaluating success factors of interventions with computerised clinical decision support. Implement Sci 2018;13(1):114 [FREE Full text] [doi: 10.1186/s13012-018-0790-1] [Medline: 30126421]
11. Hunter B, Biezen R, Alexander K, Lumsden N, Hallinan C, Wood A, et al. Future Health Today: codesign of an electronic chronic disease quality improvement tool for use in general practice using a service design approach. BMJ Open 2020;10(12):e040228 [FREE Full text] [doi: 10.1136/bmjopen-2020-040228] [Medline: 33371024]
12. Arora S, Geppert CMA, Kalishman S, Dion D, Pullara F, Bjeletich B, et al. Academic health center management of chronic diseases through knowledge networks: Project ECHO. Acad Med 2007;82(2):154-160 [FREE Full text] [doi: 10.1097/ACM.0b013e31802d8f68] [Medline: 17264693]
13. General practice data and electronic clinical decision support - consultation regulation impact statement. Australian Government Department of Health and Aged Care. 2022. URL: https://consultations.health.gov.au/primary-health-network/gp-data-and-ecds-cris/ [accessed 2023-05-06]
14. Manski-Nankervis J, Hallinan CM, Hunter BL, Lumsden N, Martinez Gutierrez J, Chima S, et al. Future Health Today: protocol for a pragmatic stratified cluster randomised head-to-head trial of quality improvement activities in general practice compared to active control. University of Melbourne 2022 [FREE Full text] [doi: 10.2196/preprints.54147]
15. Manski-Nankervis J. Future Health Today: a cluster randomised controlled trial of quality improvement activities in general practice. 2020. URL: https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=380119&isReview=true [accessed 2021-03-14]
16. What is Medicare? healthdirect. 2023. URL: https://www.healthdirect.gov.au/what-is-medicare [accessed 2023-05-06]
17. Kidd M, de Toca L. The contribution of Australia's general practitioners to the COVID-19 vaccine rollout. Aust J Gen Pract 2021;50(12):871-872 [FREE Full text] [doi: 10.31128/AJGP-11-21-6235] [Medline: 34845474]
18. McGillion M, Swoboda K, Jones S, Canfield D, Wolba B, Vaughan A, et al. Expansion of telehealth services. In: The Auditor-General Auditor-General Report No.10 2022-23 Performance Audit. Canberra: Australian National Audit Office; 2023.
19. Gillam MH, Roughead E, Tavella R, Dodd T, Beltrame J, Ryan R, et al. Impact of COVID-19 restrictions on pathology service utilisation. Intern Med J 2022;52(1):42-48 [FREE Full text] [doi: 10.1111/imj.15501] [Medline: 34432345]
20. Soós M, Temple-Smith M, Gunn J, Johnston-Ata'Ata K, Pirotta M. Establishing the Victorian primary care practice based research network. Aust Fam Physician 2010;39(11):857-862 [FREE Full text] [Medline: 21301660]
21. McInnes DK, Saltman DC, Kidd MR. General practitioners' use of computers for prescribing and electronic health records: results from a national survey. Med J Aust 2006;185(2):88-91. [doi: 10.5694/j.1326-5377.2006.tb00479.x] [Medline: 16842064]
22. Liaw S, Boyle D. Secure data linkage and information sharing with GRHANITE. In: The Person in the Centre. Melbourne: Health Informatics Conference; 2008.
23. Boyle D, Sanci L, Emery J, Gunn J, Hocking J, Manski-Nankervis J, et al. PATRON primary care research data repository. The University Of Melbourne 2019 [FREE Full text] [doi: 10.26188/5c52934b4aeb0]
24. Hespe CM, Giskes K, Harris MF, Peiris D. Findings and lessons learnt implementing a cardiovascular disease quality improvement program in Australian primary care: a mixed method evaluation. BMC Health Serv Res 2022;22(1):108 [FREE Full text] [doi: 10.1186/s12913-021-07310-6] [Medline: 35078460]
25. The Royal Australian College of General Practitioners, Guidelines for Preventive Activities in General Practice 9th edition. East Melbourne: RACGP; 2018.
26. National Vascular Disease Prevention Alliance, Guidelines for the Management of Absolute Cardiovascular Disease Risk. Melbourne: National Stroke Foundation; 2012.
27. MBS Online. Australian Government Department of Health and Aged Care. URL: http://www.mbsonline.gov.au/internet/mbsonline/publishing.nsf/Content/Home [accessed 2023-06-04]
28. Manski-Nankervis J, Canaway R, Chidgey C, Emery J, Sanci L, Hocking JS, et al. Data resource profile: primary care audit, teaching and research open network (Patron). Int J Epidemiol 2024;53(1):dyae002 [FREE Full text] [doi: 10.1093/ije/dyae002] [Medline: 38302745]
29. Peiris D, Usherwood T, Panaretto K, Harris M, Hunt J, Redfern J, et al. Effect of a computer-guided, quality improvement program for cardiovascular disease risk management in primary health care: the treatment of cardiovascular risk using electronic decision support cluster-randomized trial. Circ Cardiovasc Qual Outcomes 2015;8(1):87-95. [doi: 10.1161/CIRCOUTCOMES.114.001235] [Medline: 25587090]
30. Baigent C, Landray MJ, Reith C, Emberson J, Wheeler DC, Tomson C, SHARP Investigators. The effects of lowering LDL cholesterol with simvastatin plus ezetimibe in patients with chronic kidney disease (Study of Heart and Renal Protection): a randomised placebo-controlled trial. Lancet 2011;377(9784):2181-2192 [FREE Full text] [doi: 10.1016/S0140-6736(11)60739-3] [Medline: 21663949]

XSL•FO
RenderX

31. Department of Health. Practice Incentives Program Quality Improvement Incentive Guidelines. Canberra: Australian Government Department of Health; 2019.
32. Heselmans A, Delvaux N, Laenen A, Van de Velde S, Ramaekers D, Kunnamo I, et al. Computerized clinical decision support system for diabetes in primary care does not improve quality of care: a cluster-randomized controlled trial. Implement Sci 2020;15(1):5 [FREE Full text] [doi: 10.1186/s13012-019-0955-6] [Medline: 31910877]
33. Ivers N, Brown B, Grimshaw J. Clinical performance feedback and decision support. In: Wensing R, Grol R, Grimshaw J, editors. Improving Patient Care: The Implementation of Change in Health Care. Hoboken, NJ: Wiley-Blackwell; 2020:235-251.
34. Kwan JL, Lo L, Ferguson J, Goldberg H, Diaz-Martinez JP, Tomlinson G, et al. Computerised clinical decision support systems and absolute improvements in care: meta-analysis of controlled clinical trials. BMJ 2020;370:m3216 [FREE Full text] [doi: 10.1136/bmj.m3216] [Medline: 32943437]
35. Xie CX, Chen Q, Hincapié CA, Hofstetter L, Maher CG, Machado GC. Effectiveness of clinical dashboards as audit and feedback or clinical decision support tools on medication use and test ordering: a systematic review of randomized controlled trials. J Am Med Inform Assoc 2022;29(10):1773-1785. [doi: 10.1093/jamia/ocac094] [Medline: 35689652]
36. Patel B, Usherwood T, Harris M, Patel A, Panaretto K, Zwar N, et al. What drives adoption of a computerised, multifaceted quality improvement intervention for cardiovascular disease management in primary healthcare settings? A mixed methods analysis using normalisation process theory. Implement Sci 2018;13(1):140 [FREE Full text] [doi: 10.1186/s13012-018-0830-x] [Medline: 30419934]

Abbreviations

ACEI: angiotensin-converting enzyme inhibitor
ARB: angiotensin receptor blocker
CDS: clinical decision support
CDSS: clinical decision support software
CKD: chronic kidney disease
CPD: continuing professional development
CVD: cardiovascular disease
ECHO: Extension for Community Healthcare Outcomes
EMR: electronic medical record
FHT: Future Health Today
GEE: generalized estimating equation
GP: general practitioner
ICC: intracluster correlation
IRSD: Index of Relative Socioeconomic Disadvantage
MBS: Medicare Benefits Schedule
QI: quality improvement
SBP: systolic blood pressure

Edited by A Mavragani; submitted 31.10.23; peer-reviewed by J Walsh, J Sussman; comments to author 09.08.24; revised version received 12.09.24; accepted 03.12.24; published 03.02.25.

Please cite as:
Manski-Nankervis JA, Hunter B, Lumsden N, Laughlin A, McMorrow R, Boyle D, Chondros P, Jesudason S, Radford J, Prictor M, Emery J, Amores P, Tran-Duy A, Nelson C
Effectiveness of Electronic Quality Improvement Activities to Reduce Cardiovascular Disease Risk in People With Chronic Kidney Disease in General Practice: Cluster Randomized Trial With Active Control
JMIR Form Res 2025;9:e54147
URL: https://formative.jmir.org/2025/1/e54147
doi:10.2196/54147
PMID:

©Jo-Anne Manski-Nankervis, Barbara Hunter, Natalie Lumsden, Adrian Laughlin, Rita McMorrow, Douglas Boyle, Patty Chondros, Shilpanjali Jesudason, Jan Radford, Megan Prictor, Jon Emery, Paul Amores, An Tran-Duy, Craig Nelson. Originally published in JMIR Formative Research (https://formative.jmir.org), 03.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is

XSL•FO
RenderX

properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Nurses' Perspectives and Experiences of Using a Bed-Exit Information System in an Acute Hospital Setting: Mixed Methods Study

Stefan Walzer[1], MSc; Isabel Schön[2], MA; Johanna Pfeil[2], BA; Sam Klemm[3], BSc; Sven Ziegler[3], MSc; Claudia Schmoor[4], Dr rer nat; Christophe Kunze[1], Prof Dr

[1]Care and Technology Lab, Furtwangen University, Furtwangen im Schwarzwald, Germany

[2]AGP Social Research, FIVE e.V., Freiburg, Germany

[3]Center of Implementing Nursing Care Innovations Freiburg, Nursing Direction, Medical Center, University of Freiburg, Freiburg, Germany

[4]Clinical Trials Unit, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany

Corresponding Author:
Stefan Walzer, MSc
Care and Technology Lab
Furtwangen University
Robert-Gerwig-Platz 1
Furtwangen im Schwarzwald, 78120
Germany
Phone: 49 7723 920 2957
Email: stefan.walzer@hs-furtwangen.de

Abstract

Background: Technology that detects early when a patient at risk of falling leaves the bed can support nurses in acute care hospitals.

Objective: To develop a better understanding of nurses' perspectives and experiences with a bed-exit information system (BES) in an acute care hospital setting.

Methods: BES was implemented on 3 wards of a university medical center. Nurses completed 2 online surveys at each time point (P0 and P1) and participated in focus groups before (P0) and after (P1) implementation. Additional patient data were collected. Descriptive statistics summarized the survey results, while content analysis was applied to focus group data. Patient rates and adverse events in both phases were compared using negative binomial models. Reporting of this study adhered to the GRAMMS checklist.

Results: A total of 30 questionnaires were completed at P0 (30/72, 42%) and 24 at P1 (24/71, 33%). Of the participants, 15 completed both questionnaires (complete cases). At P1, 64% (9/14) of participants agreed that their perceived workload and strain in caring for patients with cognitive impairment was reduced by the use of the BES. The adverse event rate per patient per day was reduced by a factor of 0.61 (95% CI 0.393-0.955; *P*=.03). In addition, 11 nurses participated in 4 focus groups before and after the intervention. Participants found it challenging to operationalize the use of the BES due to the heterogeneity of care settings, but certain behaviors of patients with cognitive impairment were recognized as indicating a need for intervention. Negative experiences included information overload and alarm fatigue, leading to occasional removal of the system.

Conclusions: While BES provides some support in managing patients with cognitive impairment, its impact remains limited to specific scenarios and does not significantly reduce nurses' workload or strain. Our findings highlight the need to manage expectations of BES performance to ensure alignment between expected and actual benefits. To improve BES effectiveness and long-term implementation, future research should consider both objective measures of patient care and subjective factors such as nurse experience, structural conditions, and technical specifications. Improving information mechanisms within call systems could help reduce alarm fatigue and increase perceived usefulness. Overall, successful integration of BES in acute care settings will require close collaboration with nursing staff to drive meaningful healthcare innovation and ensure that the technology meets the needs of both patients and nurses.

Trial Registration: German Register for Clinical Studies DRKS00021720; https://drks.de/search/de/trial/DRKS00021720

XSL•FO
RenderX

(JMIR Form Res 2025;9:e64444) doi:10.2196/64444

KEYWORDS

cognitive impairment; bed-exit; technology; fall prevention; inpatient; hospital; mixed methods; nurse; information system; acute hospital; support; online questionnaire; cognitively impaired; workload

Introduction

Background

Globally, the number of older adults admitted to hospitals in emergency situations is increasing, and this trend is expected to continue as a result of current demographic changes [1]. Currently, older adults account for approximately two-thirds of all hospital inpatients, and up to 50% of this group show signs of cognitive impairment, including forms associated with dementia [2,3]. Patients with cognitive impairment often find it challenging to adhere to standardized care plans in the hospital setting [1]. They are at increased risk of developing responsive behaviors (responsive behavior is a term, which represents the effect of cognitive impairment on actions, words, and gestures of a person) and continued cognitive and functional decline [4]. Typical behaviors include an inverted day-night cycle or lack of adherence to care plans [5]. We refer to such behaviors as adverse events (AEs) in the context of nursing care when they affect the health or well-being of patients or increase the workload of nurses (see the Adverse Events in the Context of Nursing Care section). A particular challenge for nurses in hospitals is motor agitation in and around the bed, as well as related incidents such as falls or the disconnection of catheters or vascular access [6]. Kang and Hur [7] reported in a qualitative meta-synthesis that these symptoms were associated with a significantly higher workload [7]. In light of this, technological innovations have been discussed as a potential support for nurses and caregivers [8-10]. Bed-exit information systems (BES) serve as warning systems designed to monitor bed exits or patient movements in bed. This should enable nurses to prevent (or respond more quickly to) falls or other related events [11]. Nurses' perspectives and experiences on a particular technology play a central role in whether it will be successfully used and adopted over time [12]. Understanding and addressing these attitudes is essential to promoting successful and sustained technology adoption in the acute care hospital setting. In this context, little attention has been paid to nurses' perspectives and experiences on the use of a BES [13,14]. Previous studies have predominantly addressed the efficacy and effectiveness of BES in terms of reducing falls or fall rates [15,16].

Aim and Research Questions

Based on these findings, the aim of this study was to develop a better understanding of nurses' experiences, views, and needs regarding BES in an acute care hospital setting.

In terms of the aim, this study addresses the following research questions:

- What are nurses' perspectives and experiences of using a BES in an acute hospital setting?
- What is the perceived workload and strain of nurses dealing with patients with cognitive impairments and how does it change with the use of a BES?
- What is the number of AEs before and during the use of a BES?

Methods

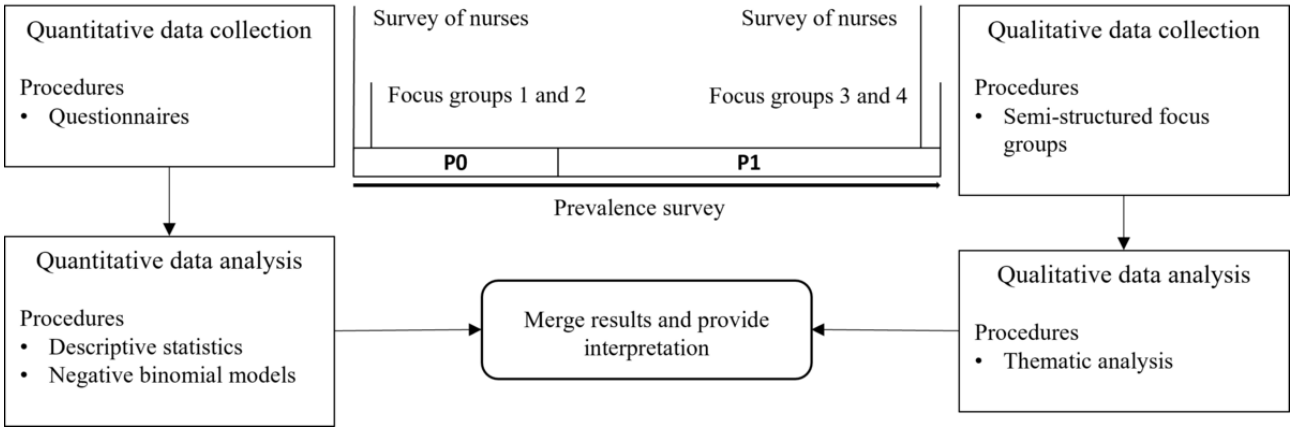
Reporting of the study followed the Good Reporting of a Mixed Methods Study (GRAMMS) criteria proposed by O'Cathain et al [17].

Design and Setting

A parallel mixed methods triangulation design (Figure 1) with quantitative surveys and qualitative focus groups (FGs) was used to gain more insight from multiple perspectives [18]. A fundamental premise of mixed methods research is that the merging of quantitative and qualitative data provides a more complete understanding of the research question than using either data source in isolation. As defined by Creswell [19], mixed methods research is "research in which the investigator collects both quantitative (closed) and qualitative (open) data, integrates the 2, and then draws interpretations based on the strengths of both data sets to understand the research problem" [19]. Integration of methods occurred in the discussion chapter of the study, where the authors collaboratively interpreted and consolidated the results. The study was conducted in 3 wards assigned to the departments of neurosurgery, surgery, and radiology at a university medical center in southern Germany with 2 phases of data collection over a total period of 190 days. The first phase without the use of BES (P0) lasted 64 days. In the second phase (P1), a BES was available for 126 days as a tool to assist nurses in caring for patients with cognitive impairment. During the study period, 912 adult patients were treated in the first phase of the study (P0). In the second phase (P1), 1748 patients were treated.

XSL•FO
RenderX

Figure 1. Study design.



Bed-Exit Information Systems

After a rigorous, multistage selection process, the research group chose a sensor mat in the form of a mattress topper from a Swiss manufacturer. By lowering the body weight on the sensors, the device transmits information about (potential) bed exits to the call system through a radio transmitter. The manufacturer has developed this technology specifically for people with cognitive impairments such as delirium, dementia, and sleep disorders. The sensor mat, equipped with a radio transmitter, was placed under the bed mattress and the corresponding radio receiver was connected to the call system. The call system of the participating wards was not able to distinguish between alarms triggered by the sensor mat and other calls. In this study, the assessment of patient cognitive impairment was performed by nurses. The decision to use the BES was also made independently by the nurses, as not all patients with cognitive impairment exhibit symptoms such as motor agitation. Training sessions were conducted for all wards, focusing on the handling and use of the equipment according to the manufacturer's intended scenarios. These sessions were held on separate days during the handover period from the early to the late shift. If additional training was deemed necessary, additional training sessions could have been scheduled. One of the authors conducted the training sessions.

Patients With Cognitive Impairment

For this study, cognitive impairment is defined as temporary or persistent problems with mental performance, regardless of age. Typical complaints include increasing forgetfulness, decreased attention, difficulty concentrating, speech problems, disorientation, or memory loss. Cognitive impairment can occur after surgery, as part of an illness, or as a result of general age-related decline [20].

Adverse Events in the Context of Nursing Care

Responsive behaviors can lead to events that (1) disrupt acute inpatient care and can directly or indirectly endanger the affected patient or fellow patients and (2) are related to the unintentional (potential) leaving of the bed without supervision [21]. As part of the prevalence survey, the number of the following AEs was recorded: (1) finding a patient out of bed; (2) falling of a patient; (3) finding a patient in a helpless situation; (4) disconnection or dislocation of accesses, drains, or catheters; (5) use of restraints; (6) leaving the ward or the hospital unsupervised, and (7) disturbing or endangering of fellow patients. These AEs were previously identified as particularly important in 2 workshops with nurses at the university hospital, which is consistent with recent studies [1,22].

Participants

Overview

The surveys and FGs were conducted among nurses who met the inclusion criteria. The target group consisted of all nurses who met the predefined criteria listed in Textbox 1, while individuals who did not meet these criteria were excluded. At P0, the total number of registered nurses used in all wards was 72, which decreased to 70 at P1. Potential participants were informed about the surveys and the FGs during ward meetings and by email using internal institutional mailing lists and were invited to participate. Interested nurses received detailed written information about the study and what participation entailed (see the Ethical Considerations section for further details). The surveys and FGs were completed during working hours with the permission of the respective organization, and no additional incentives or rewards were offered to participants.

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Age ≥18 years.
- Nurses with at least 3 years of apprenticeship or equivalent international training with professional recognition in Germany.
- Employees working on the included wards during the study period.

Exclusion criteria

- Employees of the included wards who belong to other professional groups.
- Employees of the reserve pool.
- Nurses from other wards helping out.
- Trainees.

Data Collection

The study followed a parallel mixed methods approach with the following elements:

1. Surveys and FGs of nurses at time points P0 and P1, before and after BES use
2. A prevalence survey on the number of patients with cognitive impairment and AEs related to their leaving the bed throughout the study period (P0/P1) and, in the intervention phase (P1), the number of patients with cognitive impairment who received the BES.

Surveys of Nurses

The surveys were conducted using a standardized questionnaire based on 2 questionnaires [23,24], supplemented by specific items related to the setting and technology. A total of 22 (P0) and 27 (P1) items assessed respondents' perceived strain in dealing with patients with cognitive impairment, expectations (P0) and experiences (P1) of the BES, and sociodemographic data. Items in the perceived strain category focused on AEs and the care process, which were associated with high levels of potential strain for nurses in the Isfort et al [23] study. With the exception of 2 unipolar-scaled items assessing general attitudes and acceptance of IT, items were expressed on a bipolar 7-point Likert scale.

Focus Groups

The 4 FGs were conducted by 2 moderators (SW and JP). Both moderators involved are experienced qualitative researchers, and one moderator (SW) is a registered nurse. FGs are a special type of group interview in which a moderator leads a discussion focused on a specific topic [25]. FGs are useful for obtaining information about beliefs, attitudes, thoughts, and perceptions about a particular topic from several people in a short period of time. The interaction of the participants within the group plays an important role in the generation of data [26]. A semistructured interview guide was used to structure the discussion. This guide was developed according to the recommendations of Kallio et al [27]. The structure and content of the guide are based on the current research findings mentioned above. A semistructured interview guide was chosen because it provided a framework for exploring the key concepts of interest while allowing for flexibility and openness in the discussion. During the development phase, several initial versions of the interview guide underwent multiple internal reviews and modifications by all members of the research team. Nurses were asked about their perspectives (P0) and experiences (P1), as well as barriers and opportunities for sustainable use of the BES, and the discussions were recorded and transcribed.

Prevalence Survey of Patients and Adverse Events

The prevalence survey was conducted retrospectively once a day for the last 24 hours by the nurses of the 3 wards. The number of patients with cognitive impairment and a tendency to get out of bed (first and second phases), the number of patients using a system (second phase), and the number and type of AEs that occurred in patients (first and second phases) were recorded anonymously on a documentation sheet (occupancy and fall prevalence statistics from the medical center's medical controlling and reporting systems were used as a reference for the quantitative data).

Analysis

Quantitative Data

The Clinical Trials Unit at the University Medical Center analyzed the quantitative data. Intraindividual changes in nurses' responses between phases were presented. AE rates of both phases were compared using negative binomial models [28]. Three different analyses were performed: (1) a raw overall comparison of both phases without considering individual wards, (2) a comparison of both phases separately for the 3 wards, and (3) an overall comparison of both phases adjusted for differences between wards. AE rates and multiplicative phase comparisons (ratio of AE rates in P1 vs P0) were obtained from the negative binomial models with 95% CIs, and tests of the hypothesis of equality of AE rates in P0 and P1 were performed at a 2-sided 5% significance level.

Qualitative Data

Content analysis was applied using a deductive-inductive approach [29]. The transcripts were read by 3 members of the research team (SW, IS, and JP) to familiarize themselves with the data. Codes were created through condensation and then abstracted into categories. The coders (SW and JP) reviewed the codes and categories, discussed any discrepancies, and made revisions. Finally, narrative summaries were written for each category, and thematic descriptions for the main categories were developed based on these summaries.

XSL•FO
RenderX

Setting

The study was conducted at the University Medical Center Freiburg.

Ethical Considerations

Approval was obtained from the Ethics Committee of the University of Freiburg (106/20 [MPG §23b]) and the Staff Council. Before the start of the study, all potential participants were provided with a comprehensive written document containing information about the study, including its purpose, objectives, procedures, and data protection measures. Participants were informed of the decision of the responsible committees and were given the opportunity to contact the investigators with any questions throughout the study. For participants who completed the surveys, they were explicitly informed at the beginning of the questionnaire that by submitting their responses they were indicating that they had read and understood the enclosed information letter and were giving their informed consent to participate in the survey. In the case of the FG, written informed consent was obtained from each participant at the beginning of each group session, confirming that they had been informed of the study details and were willing to participate in the study. Participation in the surveys and FGs was voluntary, no personal information was collected, and anonymity (surveys) or pseudonymization (FGs) was always maintained.

Results

Results of the Surveys

In total, 30 questionnaires were completed at P0 (30/72, 42%) and 24 at P1 (24/71, 33%). Among the total participants, 15 participants completed both questionnaires (complete cases). The sociodemographic characteristics of the participants are shown in Table 1. Approximately two-thirds of the respondents had at least 15 years of general nursing experience and the same amount of experience working with patients with cognitive impairment.

Table 1. Sociodemographic characteristics.

Characteristics	Before implementation (P0; n=30), n (%)	After implementation (P1; n=24), n (%)
Gender		
Women	20 (67)	12 (50)
Men	10 (33)	10 (42)
Age (years)		
≤40	10 (33)	9 (38)
>40	20 (67)	14 (58)
General work experience (years)		
≤15	8 (27)	6 (25)
>15	22 (73)	16 (67)
Work experience with patients with cognitive impairment (years)		
≤15	10 (33)	8 (33)
>15	20 (67)	15 (63)

The following results refer to the complete cases to allow a comparative view.

General Attitude Toward (Information) Technologies

The survey conducted at P0 showed that nurses generally had a positive attitude toward IT. This was reflected in the finding that 90% (27/30) of respondents recognized the benefits of IT, agreeing that it offered significant benefits and considering it essential in today's world.

Despite this overall positive outlook, some nurses expressed concerns and fears about using such technology. Overall, 14% (4/30) admitted that they feared making irreversible mistakes when using technology. In addition, 14% (4/30) found most technology-related issues challenging and 7% (2/30) said they were frightened by the possibility of using technology they had never worked with before.

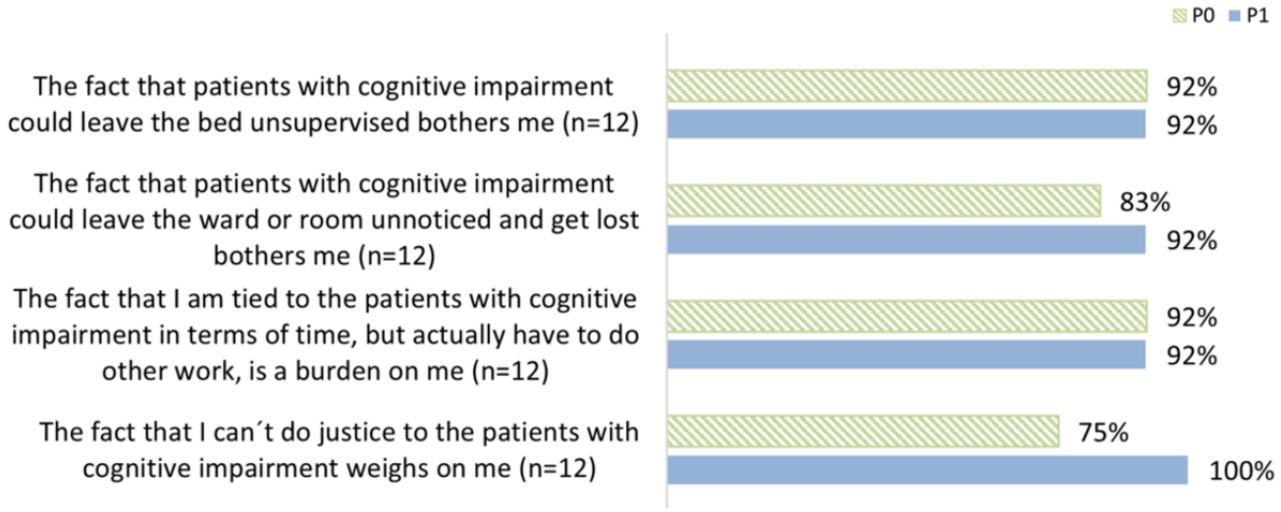
The following results refer to the complete cases to allow a comparative view. Multimedia Appendix 1 shows that the participants rated their experience in dealing with IT or acceptance of IT rather positively (experience: median 7 (IQR 6-8); acceptance: median 8 (IQR 7-9).

Perceived Strain

The vast majority of participating nurses felt moderately to very strongly strained by patients with cognitive impairment, both before and after the use of BES (Figure 2). For example, 75% (9/12) in P0 and 100% (12/12) in P1 of the participating nurses felt strained by not being able to do right by the patients, and 92% (11/12) in P0 and P1 felt strained by the fact that patients with cognitive impairment could leave the bed unattended.

XSL•FO
RenderX

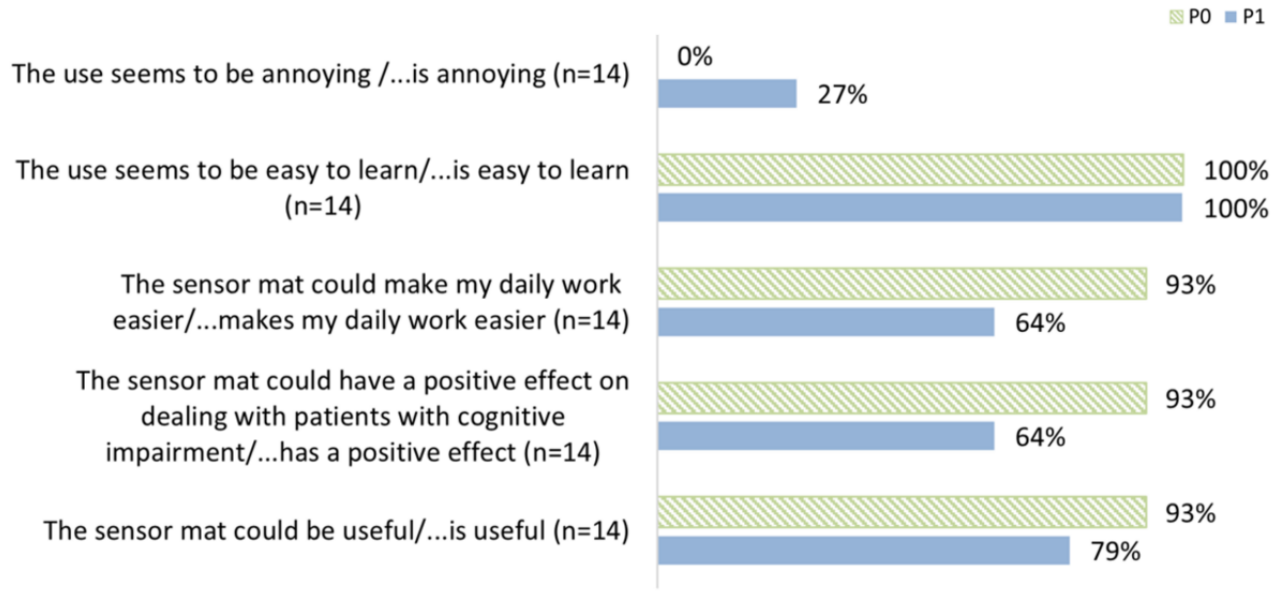
Figure 2. Perceived strain (ranging from moderate to very high).



Expectations and Experiences Regarding the Use of Bed-Exit Information Systems

Among the nurses who participated in both surveys, their expectations regarding the use of BES did not match their experiences in most categories (Figure 3). For example, after the intervention, 64% (93%, P0) agreed that BES can have a positive effect on the management of patients with cognitive impairment and that daily work becomes easier, and 79% (93%, P0) expected that BES could be useful.

Figure 3. Expectations and experiences regarding the use of technology (statements ranging from somewhat agree to strongly agree).



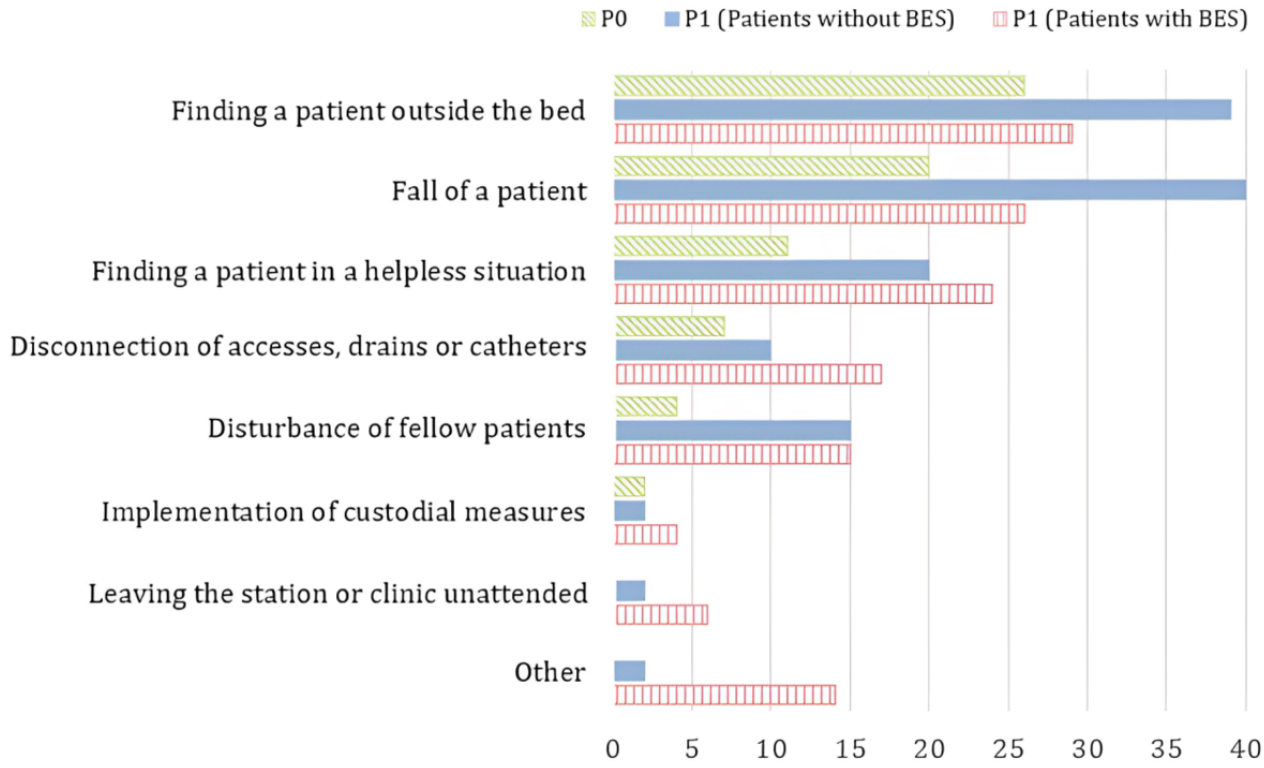
Results of the Prevalence Survey on Adverse Events

The collection of data through the questionnaires took place around 2 PM and referred to the period of the previous 24 hours. Analysis of the prevalence survey showed that in P0, patients with cognitive impairment were treated on 78% of the survey days (199/256). In P1, patients with cognitive impairment were treated on 60% of survey days (302/504), and BES was used by nurses on 20% of survey days (103 of 504). On 79 of 199 days in P0 (40%), 222 AEs were recorded in patients. In P1, 253 AEs were recorded on 82 of 302 days (27%).

The visualization in Figure 4 illustrates the occurrence of each defined AE. For each day, whether an event occurred at least once or not was recorded. It is important to note that the frequencies of the 3 categories presented can only be effectively compared within their respective contexts, as both the duration and the number of patients for each category vary. The most frequent event is the discovery of patients out of bed. This event may be due to a previous fall that left the patient in a vulnerable state. Conversely, a comparatively rare but meaningful event involves patients leaving the ward or clinic unattended, posing a potential risk of disorientation and “getting lost.”

XSL•FO
RenderX

Figure 4. Adverse events (type and number).



The average AE rates per patient per day in P0 and P1 and the ratio of AE rates in P1 versus P0 with 95% confidence intervals from the raw overall comparison without considering the individual wards shown in Table 2.

Table 2. Adverse events in the context of nursing care rates per patient and day in P0 and P1.

Phase	AE[a] rate (95% CI)	*P* value
0	0.49 (0.36-0.67)	—[b]
1	0.35 (0.27-0.46)	—
1 vs 0	0.73 (0.48-1.09)	.12

[a]AE: adverse event.

[b]Not applicable.

Table 3 provides a comparison of AE rates between P0 and P1 separately for the three wards.

XSL•FO
RenderX

Table 3. Adverse event rates per patient per day with interaction between phase and ward.

Ward	Phase	AE[a] rate (95% CI)	*P* value
1	0	0.29 (0.17-0.49)	—[b]
1	1	0.57 (0.39-0.82)	—
1	1 vs 0	1.97 (1.03-3.75)	.04
2	0	0.28 (0.13-0.60)	—
2	1	0.26 (0.08-0.08)	—
2	1 vs 0	0.92 (0.23-3.70)	.91
3	0	0.70 (0.47-1.05)	—
3	1	0.20 (0.13-0.29)	—
3	1 vs 0	0.28 (0.16-0.50)	<.001

[a]AE: adverse event; in the context of nursing care.

[b]Not applicable.

The picture was different for each ward. While ward 1 had a relatively low AE rate of 0.29 (95% CI 0.17-0.49) in P0 and increased by a factor of 1.97 (95% CI 1.03-3.75) in P1, ward 2 showed similarly low AE rates in P0 and P1. In ward 3 with 125 patient days in P0 and 248 patient days in P1, the relatively high AE rate in P0 decreased from 0.70 (95% CI 0.47-1.05) in P1 to 0.20 (95% CI 0.13-0.29). The statistical test for interaction between phase and ward showed a *P* value of .00. The overall comparison of AE rates between P0 and P1, adjusted for differences between the wards, is shown in Table 4.

Table 4. Adjusted adverse event rates per patient per day in P0 and P1.

Phase	AE[a] rate (95% CI)	*P* value
0	0.47 (0.33-0.67)	—[b]
1	0.29 (0.21-0.41)	—
1 vs 0	0.61 (0.39-0.95)	.03

[a]AE: adverse event; in the context of nursing care.

[b]Not applicable.

Thus, in a summary comparison of AE rates per cognitively impaired patient per day on each ward, there was a reduction in the AE rate by a factor of 0.613 (95% CI 0.393-0.955; *P*=.03).

Results of the FGs

A total of 11 nurses participated in the FG*s*. Two FGs with similar questions were held before (P0) and 2 were held after the use of the BES (P1). Each discussion lasted for 2 hours. The group size varied from 2 (see limitations for more details) to 5 participants, and 3 nurses participated in both P0 and P1. The key findings from the thematic descriptions are presented: (1) challenges in the use of BES and (2) opportunities in the use of BES; once from an expectation perspective (P0) and once from an experience perspective (P1).

Challenges

The nurses who participated in the preintervention FG identified several challenges in using BES. In the hospital setting, the exact causes of AEs are often not (or not fully) understood (also due to lack of biographical information). Therefore, concrete scenarios for the use of BES remain vague in advance. Consequently, the most specific expectations relate to the institutional framework: it is assumed that the system can be used in cases where continuous monitoring is needed, but no constant observers (sitters) and no other strategies (such as transfer to the intensive care unit) are available. Of course, nurses have also developed strategies to deal with AE independently of technical aids. Nontechnical alarms and safeguards such as mattresses on the floor in front of beds or plastic cups on door handles are used. However, the desire for sitters, who participants felt could best counteract AE, is mentioned very often.

> *[…] so, I see in general such a mat, a mat for movements ... I think, that [the BES] is the wrong way, because we actually need more people in the clinic […]* [P0_FG1]

After using the BES (P1), the participants confirmed that the diversity of conditions and symptoms makes operationalization a challenge. It became clear that the question, “Who gets a BES and for what reasons?” is not easy to answer. The heterogeneity of care situations also results in a variety of situations that are challenging for caregivers. Nevertheless, it was mostly specific behaviors of the patients with cognitive impairment that indicated intervention. One nurse summarized this as follows:

> *The cognitively altered [people], who are too mobile for the bed rail to protect [them]. [...] If this is someone who can perhaps still move a foot or even only half of it, he doesn’t necessarily need a sensor mat, because he can’t stand up anyway, yes. But [the*

XSL•FO
RenderX

> *one who is] just cognitively impaired, still very lively, very mobile, agile: he gets a mat.* [P1_FG1]

The discussions also revealed negative experiences with BES. Depending on the workload of the nurses, it became clear that the frequency of information could lead to delays and additional workload. After an initial rush to get information from the BES, information or alarm fatigue set in due to the frequency of information, and the systems were sometimes removed because nurses saw more disadvantages than benefits.

[…] *They said: 'Oh, that'll be the mat again.' But I mean - I think you still have to look. You can't say: 'My God, it's just the mat.' And just leave it there.* [P1_FG1]

Opportunities

Before using the BES (P0), the nurses discussed the need for more staff and sitters as mentioned earlier. Since sitters are not always available, despite the preference for sitters, a technological device such as the BES is seen as an alternative, especially at night.

> *And then we said: 'well better now than never' - sounds interesting and if it [BES] is of any use something […] then we'll cope with that. Because, I think for us, at night or also during the day depending on / - because sitters, that is another problem, you don't always get them and if this [BES] is of any help: great.* [P0_FG1]

As intended by the manufacturer, the BES is used for wandering patients. However, the participating nurses differentiated between whether the person felt a "harmless" urge to move, for example, to walk down the hallway, or whether there was an acute risk of falling. Participants indicated that BES was useful when patients with cognitive impairment were at risk of being lost if they left the building quickly and purposefully. In this context, BES has special potential at night.

I actually have the feeling that the situation [...] with people who absolutely have to go to the toilet at night and then they can't find the light - they don't even think about the bell-the night guards have already benefited from it and they also went into the room more quickly at night because they knew there was a mat in there […]. [P1_FG2]

After using the BES, the participating nurses were mostly positive about the reliability of the alarms and considered it acceptable that moderate technical problems occurred in a few cases, for example, false-positive alarms. From their point of view, the BES is a technical aid that can be used in individual situations, such as pressure mattresses or patient lifts, and as such has its place in everyday care.

> *I put that in once, plug it in and then it's there. When I put an infusion pump or an ultrasound or something else into the room, it's there, too. Maybe I'm too naive now, but of course it fits into the daily nursing routine. Because, as we have now experienced, it does prevent falls in some cases. Of course, it fits in there, for me.* [P1_FG1]

Even before use, participants identified potential support in the early detection of AE. This can help give them a sense of security and safety. BES can support and relieve nurses by alerting them to certain critical situations. One nurse summarized this as follows:

> *Then I already know how fast I can run. But that's the good thing, that I actually already know, ok, I don't hear it rumbling and then run, but it rings and I run and it doesn't rumble.* [P0_FG1]

This potential was also perceived by participants after use. In certain situations, the advantages of the BES may outweigh the disadvantages, such as constant alerting and the associated additional workload.

> *So, it was little more relaxed. Because you stop-if it rang, then there was something. Then you could go, more or less calmly. Sure, it rang all the time, but there wasn't this uncertainty of, 'What now?'*

Nonetheless, the question of overall benefit remains unanswered. Although nurses have had positive experiences, some question whether this technology is really what they need to improve their overall work situation and patient care.

Discussion

Principal Findings

This study used a parallel mixed methods triangulation design to explore nurses' perspectives and experiences of a BES in an acute care hospital. The BES was implemented in 3 wards in a university medical center. Data collection included standardized online surveys completed by nurses at two time points: before (P0) and after (P1) the implementation of the system. In addition, FGs were conducted at both time points, using a semistructured interview guide. Retrospective prevalence surveys anonymously documented patient-related data, including cognitive impairment, tendencies to get out of bed, and AEs such as falls, over 24-hour periods.

The results clearly show the emotional and psychological strain nurses experience when caring for and supporting patients with cognitive impairment. The reported workload was significantly higher than in a survey on the care of people with dementia in hospitals by Isfort et al [23]. In the survey, 66.9% of participants reported high levels of distress when patients with cognitive impairment were allowed to leave the ward unnoticed [23]. The same question in our survey resulted in over 90% agreement. The other statements also showed significantly higher reported distress. Although this is not a representative cross-section, the results are worth discussing. One explanation could be the different target group of respondents. Isfort et al [23] surveyed senior nurses who were at least no longer working full-time "at the bedside." The respondents in the current study were predominantly nurses working in direct patient care, so a greater potential for distress might be expected in this context. Another explanation may be that the participating units generally have an above-average number of patients with cognitive impairment. The nurses' general attitude and acceptance of IT and communication technology are positive. This is also consistent with the results of an online survey conducted by the German Center for Quality in Care [30].

XSL•FO
RenderX

Nearly half of nurses reported that the BES made their daily work easier or much easier. Depending on the situation, it can increase both patient safety and nurses' sense of safety. A BES can be a tool, but, like all technological innovations in highly complex environments, it does not solve complex, interrelated problems [12,31]. A BES can provide a sense of safety and increased security in selected situations, individually adapted to the setting and the patients. Looking at the complete cases, the experiences with the BES deviate negatively from the expectations. For example, the positive effect in dealing with patients with cognitive impairment, which 93% (13/14) of participants expected before using the BES, was observed by only 64% (9/14) of the participants after using the BES. This finding suggests that the experience of using the BES did not live up to previous expectations; helpfulness fell short of expectations. One way to deal with this is to manage expectations in advance (also on the part of the manufacturer) [32]. Whether the final perceived usefulness is sufficient for further use must be decided based on the fit between individual needs and benefits [33]. An improvement in perceived usefulness could possibly be achieved by implementing differentiated information mechanisms through the call system.

The most common AEs are getting patients out of bed and falls. A total of 66.9% (P0) and 59.9% (P1) of the respondents reported them as very stressful. These results are consistent with the findings of Isfort et al [23], who showed that these two events are the most stressful factors for nurses. Comparing the results of the cognitive impairment prevalence survey and AEs in the 2 phases, the rate of AEs per patient day decreased by a factor of 0.6. However, there were significant differences between the wards, with sometimes conflicting results. One explanation could be that the clinical conditions and comorbidities of the patients differ, which could lead to differences in the ability to use the BES very accurately.

Bed-leaving tendencies do not stand alone, but must always be evaluated in the context of other individual patient behaviors as well as the conditions of the particular ward. Focusing on a specific tool necessarily takes into account only a small part of the picture, which does not exist in this form for the nurses participating in the FGs. The situations described in the discussions show the complexity of supporting nurses in dealing with patients with cognitive impairment. Previous studies have predominantly addressed the efficacy and effectiveness of BES in terms of reducing falls or fall rates [15,34,35]. As a result, the primary focus of research has been on evaluating the functionality of the technology. However, there are many other factors that determine whether a technology will be successfully used and adopted over time in a particular setting [12,31]. This study provides a broader picture of the potential benefits of BES in the care and support of patients with cognitive impairments from a nurse's perspective.

The analysis of the FGs shows a generally open and pragmatic attitude toward technology in nursing. The benefits of BES from the nurses' point of view can be, for example, physical relief (eg, fewer rounds at night), psychological relief (eg, a greater sense of security), or an improvement in the quality of care (eg, avoidance of patient injury). At the same time, it is important to avoid new challenges, such as high alarm rates and limited response options, which can lead to alarm fatigue. In light of the survey results showing high expectations of the benefits of the BES and low experience of the benefits, the FGs showed that the high expectations were related to the reduction of falls. In the nurses' perception, this goal was not fully achieved. However, it helped to cope with the consequences and was therefore helpful in reducing stress. The FGs show that the BES is best suited for use at night and with patients at risk of falling and with mild cognitive impairment. This specification of use scenarios makes it possible to obtain benefits in the form of fall prevention or rapid response to falls without being overloaded by too many alarms. Our results suggest that the effectiveness of the intervention appears to be strongly influenced by several factors, such as the individual condition of the patient, the staff-to-patient ratio, and the timing of use.

Strengths and Limitations

This paper discusses the implementation of a specific BES. Due to the different systems and implementation aspects, it is not possible to generalize the results. The sample size, 30 people in P0, 24 in P1 and 15 complete cases, is a limitation for complex statistical procedures. However, the combination of quantitative and qualitative methods provides a comprehensive perspective and deeper insights into the research topic. We collected data for the prevalence survey using documentation forms at around 2 PM, referring to the previous 24-hour period. Overlaps occurred on 3 days per ward (study start, transition between study phases, and study end), but we ensured that patient and case counts were accurate to the day. The heterogeneous patient population, with complex and fluctuating symptom patterns and congruent but not equivalent wards, introduces potential bias in the evaluation of the prevalence survey. The significant reduction in the AE rate suggests that the use of the BES may be beneficial, although it does not allow causal conclusions. We must remember that leaving the bed unattended was also considered an AE. An increase in the AE rate could indicate a reduction in underreporting, paradoxically a desirable outcome. A more precise numerical presentation of the different AEs would have been helpful. Another limitation is the different conditions of use of the technology in different ward designs, which limits the direct applicability of the results to other wards. Since the focus of this study was to capture the perspectives of nursing staff specifically, the diversity of participants was limited by the study design. Future research could, however, consider including a broader range of health care professionals or expanding the diversity of nursing staff to explore how different perspectives, such as those from interdisciplinary teams, may influence the outcomes of BES implementation. Due to recruitment difficulties, we conducted a postintervention FG with only 2 participants, which may have changed the dynamics of the discussion, making it more like a dyadic interview. There are some similarities and differences between these methods [36]. In our case, each group discussion maintained the same thematic focus and similar dynamics.

Conclusions

Our results indicate that while the BES provides some assistance in managing patients with cognitive impairment, its impact is limited to specific scenarios and does not significantly reduce

XSL•FO
RenderX

nurses' workload or strain. In addition, our study highlights the importance of managing expectations regarding BES performance. Setting realistic expectations for both manufacturers and end users is essential to bridge the gap between expected and experienced benefits. To better evaluate the benefits of BES for nurses and increase the likelihood of successful long-term implementation, future studies should address not only objectively measurable changes in patient care but also subjective perspectives, structural factors, and technical specifications. Technical improvements, such as differentiated information mechanisms through the call system, could enhance its perceived usefulness and help reduce alarm fatigue. By examining these factors, a more comprehensive understanding of the BES's impact can be achieved, supporting its appropriate implementation in acute care settings. Finally, these findings emphasize the importance of integrating technology like the BES through close collaboration with nursing staff. Such collaboration is crucial for driving effective health care innovation and ensuring that new technologies meet the needs of both patients and nurses.

Acknowledgments

Our study team is particularly grateful to all the nurses in the participating wards for their cooperation. We would like to thank our colleagues and key persons of our project partners at the Center of Implementing Nursing Care Innovations, Freiburg, Germany. This joint project on the development and research of new care technologies in Germany is funded by the Federal Ministry of Education and Research. At the Freiburg Center, the focus is on integrating and evaluating innovative nursing technologies in the acute care setting (grants 16SV7888, 16SV7889, and 16SV7886K).

Authors' Contributions

SW contributed to conceptualization, methodology, investigation, formal analysis, data curation, and writing-original draft. IS was involved in conceptualization, methodology, investigation, formal analysis, data curation, and writing-original draft. JP contributed to data curation-reviewing and editing. SS contributed to the conceptualization, methodology, data curation, and investigation-reviewing and editing. SZ was involved in conceptualization, methodology, and investigation-reviewing and editing. CS contributed to conceptualization, supervision, resources-reviewing and editing. CK contributed to the conceptualization, supervision, resources-reviewing and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Self-assessment of general experience and acceptance (minimum 1, maximum 10).
[PNG File , 83 KB - formative_v9i1e64444_app1.png]

References

1. Mudge AM, Lee-Steere K, Treleaven E, Cahill M, Finnigan S, McRae P. Cognitive impairment in older hospital inpatients: prevalence, care needs and carer perceptions. Aust Health Rev 2022;46(2):244-250. [doi: 10.1071/AH20286] [Medline: 34856117]
2. Fogg C, Griffiths P, Meredith P, Bridges J. Hospital outcomes of older people with cognitive impairment: an integrative review. Int J Geriatr Psychiatry 2018;33(9):1177-1197 [FREE Full text] [doi: 10.1002/gps.4919] [Medline: 29947150]
3. Pinkert C, Faul E, Saxer S, Burgstaller M, Kamleitner D, Mayer H. Experiences of nurses with the care of patients with dementia in acute hospitals: a secondary analysis. J Clin Nurs 2018;27(1-2):162-172. [doi: 10.1111/jocn.13864] [Medline: 28426891]
4. Dookhy J, Daly L. Nurses' experiences of caring for persons with dementia experiencing responsive behaviours in an acute hospital: a qualitative descriptive study. Int J Older People Nurs 2021;16(4):e12379. [doi: 10.1111/opn.12379] [Medline: 33963674]
5. Motzek T, Büter K, Mächler K, Junge M, Marquardt G. Orientierungsschwierigkeiten, behaviorale und psychiatrische Symptome bei patienten mit kognitiven Einschränkungen im Krankenhaus. HBScience 2017;8(2):46-51. [doi: 10.1007/s16024-017-0297-8]
6. Choi SSW, Cajita MI, Gitlin LN. A review of measures of three common dementia-related behaviors: rejection of care, aggression, and agitation. Geriatr Nurs 2020;41(6):692-708. [doi: 10.1016/j.gerinurse.2020.04.003] [Medline: 32402574]
7. Kang Y, Hur Y. Nurses' experience of nursing workload-related issues during caring patients with dementia: a qualitative meta-synthesis. Int J Environ Res Public Health 2021;18(19):10448 [FREE Full text] [doi: 10.3390/ijerph181910448] [Medline: 34639748]
8. Torres RLS, Visvanathan R, Hoskins S, van den Hengel A, Ranasinghe D. Effectiveness of a batteryless and wireless wearable sensor system for identifying bed and chair exits in healthy older people. Sensors (Basel) 2016;16(4):546 [FREE Full text] [doi: 10.3390/s16040546] [Medline: 27092506]

XSL•FO
RenderX

9. Vandenberg AE, van Beijnum BJ, Overdevest VGP, Capezuti E, Johnson TM. US and Dutch nurse experiences with fall prevention technology within nursing home environment and workflow: a qualitative study. Geriatr Nurs 2017;38(4):276-282 [FREE Full text] [doi: 10.1016/j.gerinurse.2016.11.005] [Medline: 27956058]
10. Husebo BS, Heintz HL, Berge LI, Owoyemi P, Rahman AT, Vahia IV. Sensing technology to monitor behavioral and psychological symptoms and to assess treatment response in people with dementia. a systematic review. Front Pharmacol 2019;10:1699 [FREE Full text] [doi: 10.3389/fphar.2019.01699] [Medline: 32116687]
11. Seow JP, Chua TL, Aloweni F, Lim SH, Ang SY. Effectiveness of an integrated three-mode bed exit alarm system in reducing inpatient falls within an acute care setting. Jpn J Nurs Sci 2022;19(1):e12446. [doi: 10.1111/jjns.12446] [Medline: 34286920]
12. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
13. Hubbartt B, Davis SG, Kautz DD. Nurses' experiences with bed exit alarms may lead to ambivalence about their effectiveness. Rehabil Nurs 2011;36(5):196-199. [doi: 10.1002/j.2048-7940.2011.tb00195.x] [Medline: 21882797]
14. Wong Shee A, Phillips B, Hill K, Dodd K. Feasibility, acceptability, and effectiveness of an electronic sensor bed/chair alarm in reducing falls in patients with cognitive impairment in a subacute ward. J Nurs Care Qual 2014;29(3):253-262. [doi: 10.1097/NCQ.0000000000000054] [Medline: 24487696]
15. Cuttler SJ, Barr-Walker J, Cuttler L. Reducing medical-surgical inpatient falls and injuries with videos, icons and alarms. BMJ Open Qual 2017;6(2):e000119 [FREE Full text] [doi: 10.1136/bmjoq-2017-000119] [Medline: 29450285]
16. Røyset B, Talseth-Palmer BA, Lydersen S, Farup PG. Effects of a fall prevention program in elderly: a pragmatic observational study in two orthopedic departments. Clin Interv Aging 2019;14:145-154 [FREE Full text] [doi: 10.2147/CIA.S191832] [Medline: 30697039]
17. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008;13(2):92-98. [doi: 10.1258/jhsrp.2007.007074] [Medline: 18416914]
18. Ingham-Broomfield B. A nurses' guide to mixed methods research. Aust J Adv Nurs 2016;33(4):46-52. [doi: 10.37464/2016.334.1559]
19. Creswell JW. A Concise Introduction to Mixed Methods Research. Washington DC: SAGE Publications; 2021:1-168.
20. Roy E. Cognitive Impairment. In: Gellman MD, Turner R, editors. Encyclopedia of Behavioral Medicine. New York, NY: Springer New York, NY; 2013:449-451.
21. Watkin L, Blanchard MR, Tookman A, Sampson EL. Prospective cohort study of adverse events in older people admitted to the acute general hospital: risk factors and the impact of dementia. Int J Geriatr Psychiatry 2012;27(1):76-82. [doi: 10.1002/gps.2693] [Medline: 21360591]
22. Bickel H, Hendlmeier I, Heßler JB, Junge MN, Leonhardt-Achilles S, Weber J, et al. The prevalence of dementia and cognitive impairment in hospitals. Dtsch Arztebl Int 2018;115(44):733-740 [FREE Full text] [doi: 10.3238/arztebl.2018.0733] [Medline: 30565543]
23. Isfort M, Klostermann J, Gehlen D, Siegling B. Pflege-Thermometer 2014 Eine bundesweite Befragung von leitenden Pflegekräften zur Pflege und Patientenversorgung von Menschen mit Demenz im Krankenhaus. Köln: Deutsches Institut für angewandte Pflegeforschung e.V; 2014.
24. Spagnolli A, Guardigli E, Orso V, Varotto A, Gamberini L. Measuring user acceptance of wearable symbiotic devices: validation study across application scenarios. In: Jacucci G, Gamberini L, Freeman J, Spagnolli A, editors. Symbiotic Interaction. Cham, Switzerland: Springer International Publishing; 2014:978-973.
25. Rädiker S, Kuckartz U. Analyzing qualitative data with MAXQDA. In: Text, Audio, and Video. Cham, Switzerland: Springer; 2019:XIII-X290.
26. Plummer P. Focus group methodology. Part 1: design considerations. Int J Ther Rehabil 2017;24(7):297-301. [doi: 10.12968/ijtr.2017.24.7.297]
27. Kallio H, Pietilä AM, Johnson M, Kangasniemi M. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. J Adv Nurs 2016;72(12):2954-2965. [doi: 10.1111/jan.13031] [Medline: 27221824]
28. Baxter A, Huddleston E. SAS/STAT® 15.1 User?s Guide. Cary, NC, USA: SAS Institute Inc; 2018.
29. Kuckartz U. Qualitative text analysis: a systematic approach. In: Kaiser G, editor. Compendium for early career researchers in mathematics education. Cham, Switzerland: Springer; 2019:181-197.
30. Kuhlmey A, Blüher S, Nordheim J, Zöllick J. Technik in der Pflege?Einstellungen von professionell Pflegenden zu Chancen und Risiken neuer Technologien und technischer Assistenzsysteme. Zentrum für Qualität in der Pflege, Berlin 2019.
31. Walzer S, Armbruster C, Kunze C, Farin-Glattacker E. NASSS-CAT-D: Komplexität bei der Einführung von Gesundheitstechnologien erkennen, verstehen und beherrschen. In: Krick T, Zerth J, Rothgang H, Klawunn R, Walzer S, Kley T, editors. Pflegeinnovationen in der Praxis. Wiesbaden, Italy: Springer Gabler; 2023:17-35.
32. Hebb A, Kistler M, George E, Zamboni B. Satisfaction and technology acceptance of staff regarding use of continuous video monitoring in comparison with sitters. J Nurs Adm 2021;51(2):60-62. [doi: 10.1097/NNA.0000000000000970] [Medline: 33449593]

XSL•FO
RenderX

33. Tubaishat A. Perceived usefulness and perceived ease of use of electronic health records among nurses: application of technology acceptance model. Inform Health Soc Care 2018;43(4):379-389. [doi: 10.1080/17538157.2017.1363761] [Medline: 28920708]
34. Balaguera HU, Wise D, Ng CY, Tso HW, Chiang WL, Hutchinson AM, et al. Using a medical intranet of things system to prevent bed falls in an acute care hospital: a pilot study. J Med Internet Res 2017;19(5):e150 [FREE Full text] [doi: 10.2196/jmir.7131] [Medline: 28473306]
35. Barker AL, Morello RT, Wolfe R, Brand CA, Haines TP, Hill KD, et al. 6-PACK programme to decrease fall injuries in acute hospitals: cluster randomised controlled trial. BMJ 2016;352:h6781 [FREE Full text] [doi: 10.1136/bmj.h6781] [Medline: 26813674]
36. Szulc J, King N. The practice of dyadic interviewing: strengths, limitations and key decisions. Forum Qual Soc Res 2022;23(2). [doi: 10.17169/fqs-22.2.3776]

Abbreviations

AE: adverse event
BES: bed-exit information system
FG: focus group
GRAMMS: Good Reporting of a Mixed Methods Study

Edited by A Mavragani; submitted 28.07.24; peer-reviewed by J Walsh; comments to author 30.10.24; revised version received 27.11.24; accepted 09.12.24; published 05.02.25.

Please cite as:
Walzer S, Schön I, Pfeil J, Klemm S, Ziegler S, Schmoor C, Kunze C
Nurses' Perspectives and Experiences of Using a Bed-Exit Information System in an Acute Hospital Setting: Mixed Methods Study
JMIR Form Res 2025;9:e64444
URL: https://formative.jmir.org/2025/1/e64444
doi:10.2196/64444
PMID:

©Stefan Walzer, Isabel Schön, Johanna Pfeil, Sam Klemm, Sven Ziegler, Claudia Schmoor, Christophe Kunze. Originally published in JMIR Formative Research (https://formative.jmir.org), 05.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Status of Digital Health Technology Adoption in 5 Vietnamese Hospitals: Cross-Sectional Assessment

Duc Minh Tran[1], PMD; Nguyen Thanh Dung[2*], MD; Chau Minh Duc[3*], MD, PhD; Huynh Ngoc Hon[4*], MD; Le Minh Khoi[5*], MD, PhD; Nguyen Phuc Hau[2*], BSc; Duong Thi Thu Huyen[6*], BSc; Huynh Thi Le Thu[4*], BSc; Tran Van Duc[5*], BSc; VITAL (Vietnam ICU Translational Applications Laboratory) Investigators[7*]; Lam Minh Yen[1], MD; C Louise Thwaites[1,8], BSc, MBBS, MD; Chris Paton[9,10], BMBS, BMedSci, MBA

[1]Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam
[2]Hospital for Tropical Diseases, Ho Chi Minh City, Vietnam
[3]Dong Thap General Hospital, Dong Thap, Vietnam
[4]Trung Vuong Hospital, Ho Chi Minh City, Vietnam
[5]University Medical Center, Ho Chi Minh City, Vietnam
[6]Department of Information Technology, National Hospital for Tropical Diseases, Hanoi, Vietnam
[7]See Acknowledgements,
[8]Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, United Kingdom
[9]Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom
[10]Department of Information Science, University of Otago, Dunedin, New Zealand
[*]these authors contributed equally

Corresponding Author:
Duc Minh Tran, PMD
Oxford University Clinical Research Unit
764 Vo Van Kiet Street, Ward 1, District 5
Ho Chi Minh City
Vietnam
Phone: 84 2839237954
Email: ductm@oucru.org

Abstract

Background: Digital health technologies (DHTs) have been recognized as a key solution to help countries, especially those in the low- and middle-income group, to achieve the Sustainable Development Goals (SDGs) and the World Health Organization's (WHO) Triple Billion Targets. In hospital settings, DHTs need to be designed and implemented, considering the local context, to achieve usability and sustainability. As projects such as the Vietnam ICU Translational Applications Laboratory are seeking to integrate new digital technologies in the Vietnamese critical care settings, it is important to understand the current status of DHT adoption in Vietnamese hospitals.

Objective: We aimed to explore the current digital maturity in 5 Vietnamese public hospitals to understand their readiness in implementing new DHTs.

Methods: We assessed the adoption of some key DHTs and infrastructure in 5 top-tier public hospitals in Vietnam using a questionnaire adapted from the Vietnam Health Information Technology (HIT) Maturity Model. The questionnaire was answered by the heads of the hospitals' IT departments, with follow-up for clarifications and verifications on some answers. Descriptive statistics demonstrated on radar plots and tile graphs were used to visualize the data collected.

Results: Hospital information systems (HIS), laboratory information systems (LIS), and radiology information systems–picture archiving and communication systems (RIS-PACS) were implemented in all 5 hospitals, albeit at varied digital maturity levels. At least 50% of the criteria for LIS in the Vietnam HIT Maturity Model were satisfied by the hospitals in the assessment. However, this threshold was only met by 80% and 60% of the hospitals with regard to HIS and RIS-PACS, respectively. Two hospitals were not using any electronic medical record (EMR) system or fulfilling any extra digital capability, such as implementing clinical data repositories (CDRs) and clinical decision support systems (CDSS). No hospital reported sharing clinical data with other organizations using Health Level Seven (HL7) standards, such as Continuity of Care Document (CCD) and Clinical Document

XSL•FO
RenderX

Architecture (CDA), although 2 (40%) reported their systems adopted these standards. Of the 5 hospitals, 4 (80%) reported their RIS-PACS adopted the Digital Imaging and Communications in Medicine (DICOM) standard.

Conclusions: The 5 major Vietnamese public hospitals in this assessment have widely adopted information systems, such as HIS, LIS, and RIS-PACS, to support administrative and clinical tasks. Although the adoption of EMR systems is less common, their implementation revolves around data collection, management, and access to clinical data. Secondary use of clinical data for decision support through the implementation of CDRs and CDSS is limited, posing a potential barrier to the integration of external DHTs into the existing systems. However, the wide adoption of international standards, such as HL7 and DICOM, is a facilitator for the adoption of new DHTs in these hospitals.

(JMIR Form Res 2025;9:e53483) doi:10.2196/53483

KEYWORDS

electronic health record; electronic medical record; digital maturity; clinical decision support; digital infrastructure; Vietnam; health information technology; digital health technology; low- and middle-income country

Introduction

The COVID-19 pandemic demonstrated the need for countries worldwide to implement a national digital health infrastructure needed to respond to health emergencies and to achieve the Sustainable Development Goals (SDGs) [1]. As part of this digital transformation effort, digital health technologies (DHTs), such as electronic medical record (EMR) systems, laboratory information systems (LISs), and picture archiving and communication systems (PACSs), have become widespread. In recent years, Vietnam has developed and begun implementation of an ambitious national digital health strategy that includes the deployment of hospital-based EMRs and an electronic health insurance claim system [2].

As the cost of hardware continues to drop and the range and capabilities of DHTs continue to expand [3], digitization of health systems is becoming increasingly feasible in low- and middle-income countries (LMICs). New technologies, such as the Internet of Things, wearable devices, and artificial intelligence (AI), have begun to be adopted in health care settings and could be an important enabler for realizing the World Health Organization's (WHO) Triple Billion Targets (more than 1 billion people benefiting from universal health care; better protected from health emergencies; and enjoying better health and well-being) and to accelerate SDG achievement [4]. In 2023, WHO announced the Global Initiative for Digital Health to help LMICs achieve the aims of the Global Strategy on Digital Health by developing their national digital health infrastructure, adopting international data standards, and fostering increased international collaboration in digital health [1,4].

In global health research, there is increasing interest in implementing new technologies, such as AI or smart wearables, for clinical decision support [5,6]. However, for these systems to be practicable and to take advantage of increasingly available "big data," the local provision of technological infrastructure and implementation capabilities is necessary. For example, hospitals that have implemented EMR systems could use data from the EMRs to train AI models and enable the implementation of clinical decision support systems (CDSS) [7]. In addition, the adoption of international interoperability standards, such as Health Level Seven (HL7) Fast Healthcare Interoperability Resources (FHIR), is also necessary to allow efficient data exchange between information systems and new DHTs [8].

In hospital settings, digital maturity models, such as the Electronic Medical Record Adoption Model (EMRAM) from the Healthcare Information and Management Systems Society (HIMSS), are often used to guide DHT adoption and implementation [9]. These models can be used on a national or international scale and can target different dimensions, including technology, strategy, interoperability, analytics, and governance. Currently, there is no international consensus on how to best measure digital capabilities in health care institutions [9,10], and variations in digital maturity models across different countries may reflect an adaptation to national priorities and contexts, especially when being implemented in LMICs. In 2017, the Vietnam Ministry of Health (MoH) issued the Vietnam Health Information Technology (HIT) Maturity Model, originally known as Circular 54/2017/TT-BYT, as a roadmap for Vietnamese public hospitals toward a paperless and smart hospital model [2,11]. Although this guidance defines 7 digital maturity stages similar to EMRAM, the criteria to achieve a specific level have been significantly modified. For example, the Vietnam model includes the requirement for hospital information systems (HIS) to connect to the Vietnam Social Security claim portal and assesses the integration of the national standardized vocabularies into the digital systems [11].

The Vietnam ICU Translational Applications Laboratory (VITAL) project focuses on developing new technologies for improving outcomes of critically ill patients in Vietnam, such as CDSS using AI algorithms trained on local data [12-14]. In the national DHT policy landscape, regulations, infrastructure, and professional workflows are important elements [2,15] to ensure the sustainability and successful implementation of DHTs [6,16]. For this reason, in this study, we aimed to explore the current status and likely short-term development of the DHT infrastructure in 5 major public hospitals in Vietnam. Although the findings may not be generalizable to all Vietnamese hospitals regarding the state of their DHT adoption, we expect they can inform discussions around how new DHTs can be integrated with the existing IT infrastructure in Vietnamese public hospitals.

XSL•FO
RenderX

Methods

Study Design

This study used a structured questionnaire to explore the current adoption of DHTs in 5 public hospitals in Vietnam. The assessment items were developed based on the Vietnam HIT Maturity Model, which comprises 8 domains related to technology adoption and digital capabilities in hospitals [11]. The 8 domains are *IT infrastructure*, *HIS*, *LIS*, *radiology information systems–picture archiving and communication systems (RIS-PACS)*, *EMR*, *administrative and operation software*, *security and information safety*, and *nonfunctional criteria*. Except for security and information safety and nonfunctional criteria, each domain includes a set of infrastructure requirements, such as having a server room, or software functional criteria, such as having laboratory test management functionality. The model also sets out further digital capabilities focusing on assisting clinical practice, such as CDSS, clinical data repositories (CDRs), and medication management; these are named *extra capabilities*.

Following a stage-based approach, a hospital reaches a digital maturity level in a domain if a predefined set of criteria is met. The hospital can only move to a higher digital maturity level when all criteria of the lower levels have been satisfied. The domains of IT infrastructure and HIS have 7 digital maturity levels (1-7), while the other domains have 2 levels (*basic* and *advanced*). Multimedia Appendix 1 presents the digital maturity levels and their scoring criteria based on the domains and extra capabilities.

Questionnaire, Data Collection, and Analysis

Since the focus of this study was the adoption of digital health systems in hospitals, we developed an assessment tool based on the criteria from the IT infrastructure, HIS, LIS, RIS-PACS, EMR, and extra capabilities domains (see Multimedia Appendix 2). The administrative and operation software, security and information safety, and nonfunctional criteria domains were not included in this study as we considered these less relevant to the study's focus despite their importance to the overall digital maturity of a health care institution.

For each domain, we gathered information through the hospitals' IT departments to determine whether the hospitals met the component criteria or whether they planned to achieve the criteria in the next 3-5 years. Results of the assessments were discussed with the heads of the IT departments to ensure accuracy. Data collection was carried out from July 2022 to March 2023. Information about the digital maturity of each hospital was entered into Microsoft Excel for initial data cleaning and then analyzed using R software (R Foundation for Statistical Computing) [16].

Our aim was to provide a descriptive summary of what infrastructure and functionalities, or criteria, were available in each hospital across the domains rather than calculating scores for their digital maturity. The proportion of criteria that each hospital met in each domain was calculated. A radar plot was created using the *fmsb* package in R [17] to depict the difference in the proportions of criteria met among the hospitals by domain. Radar plots are regarded as an efficient tool to compare various groups on multiple variables, as discussed by Saary [18]. In addition to summary statistics, we further compared technology adoption between the 5hospitals by delving into the individual domains. Tile graphs were created using the *ggplot2* package in R [19] to illustrate the criteria that the hospitals had satisfied within each domain. These graphs allow for the visualization of individual data points (eg, which criterion was met by a specific hospital) as opposed to visualizing the counts or proportions of hospitals that met a particular criterion, as in the case of bar charts.

Study Population

Until the beginning of 2024, it was estimated that there were nearly 1500 public and over 300 private hospitals in Vietnam [20]. Five hospitals were purposively selected for this study based on previous interest in DHT innovation and participation in digital health research activities. Our approach aimed to sample hospitals where DHT innovation was likely to occur the soonest, rather than providing a comprehensive view of all government hospitals. Of the 5 hospitals that participated in this assessment, 4 (80%) were in Hanoi or Ho Chi Minh City, the 2 largest cities in Vietnam, and 1 (20%) was in a Mekong Delta (ie, rural) province. All are top-tier teaching public hospitals with a capacity of over 500 beds; 2 (40%) are tertiary hospitals that focus on a limited number of specialties, while 3 (60%) are general hospitals that provide care in a wide range of specialties. Most hospitals have between 8 and 12 IT staff per 1000 beds, except for 1 (20%) hospital that has an approximate ratio of over 50 per 1000 beds. Characteristics of the hospitals are shown in Table 1. To ensure confidentiality, we withheld information that could be used to identify the hospitals.

Table 1. Characteristics of the 5 hospitals participating in this assessment.

Hospital	Tier	Hospital type	Number of beds	Teaching hospital	IT staff:bed ratio
A	1	General hospital	>500	Yes	50:1000
B	1	Tertiary hospital	>500	Yes	13:1000
C	1	Tertiary hospital	>500	Yes	11:1000
D	1	General hospital	>500	Yes	8:1000
E	1	General hospital	>500	Yes	12:1000

XSL•FO
RenderX

Ethical Considerations

This project was exempted from ethical review as it constituted a service review, as defined by the Oxford Tropical Research Ethics Committee (OxTREC) [21,22]. The selected hospitals were identified from a professional network that the study's principal investigators (authors LT and CP) were part of. The hospitals were invited to participate in the assessment through an invitation letter explaining the purpose and scope of the study. Upon approval by the hospitals' authorities, the hospital IT directors were provided with an information sheet, the questionnaire, and an instruction document. Queries regarding the assessment, if any, were clarified by the first author (DMT).

The IT directors were informed in writing that they could refuse to provide any data if they wished and that the data collected would be securely managed and only be used for the purpose of the study. Participating departments were able to review and agree on the contents of the manuscript prior to submission. No personally identifiable data were collected during the research. The research data were managed as per the Oxford University Clinical Research Unit's data management policy. Information that can be used to identify the hospitals, such as names and the number of beds, was deidentified to ensure confidentiality. No compensation was given to the hospitals for participating in this study.

Information Systems in Vietnamese Public Hospitals

HIS, LIS, radiology information systems (RIS), PACS, and EMRs are popular information systems in health care institutions worldwide. However, their functionalities and use cases may differ to some extent depending on the national context that they are implemented in. In this section, we provide a Vietnam-oriented description of these information systems to support the interpretation of the study's findings.

Hospital Information Systems

Also known as hospital management systems, HIS are usually prioritized by Vietnamese hospitals among other information systems. They are typically used to manage outpatient services, as well as administrative and billing data emerging from all wards and departments in a hospital. HIS can be interfaced with RIS and LIS to send orders or receive imaging and laboratory results. A significant proportion of HIS in Vietnamese hospitals have been developed and provided by local vendors, such as the Corporation for Financing Promoting Technology (known as FPT), the Vietnam Posts and Telecommunications Group (known as VNPT), and Viettel [23].

Laboratory Information Systems

LIS allow clinical laboratories to track and use a wide range of data related to orders, specimens, results, and consumables. They reduce the turnaround time from ordering tests to collecting laboratory results through automatic processes, such as integrating with HIS to transfer orders to the laboratory and connecting with laboratory instruments to automatically receive results. Clinicians can use LIS installed in their wards or HIS interfaced with LIS to access laboratory results, when available. Many public hospitals in Vietnam have chosen LIS developed by local vendors, such as LABCONN by LABSoft (deployed in over 200 laboratories) and LABMDSOFT by MDsoft.

Radiology Information Systems–Picture Archiving and Communication Systems

RIS and PACS are often used together in medical imaging departments. RIS typically support managing orders, patients, and imaging procedure information. RIS can receive order information and send images and report back to HIS using HL7 messages. RIS can also send orders to imaging machines and receive processed images from PACS.

PACS comprise 4 components: imaging instruments, a secure network to transfer the images, workstations that allow doctors to view the images, and a storage system that can archive the images. PACS can thus facilitate image viewing, editing, storing, and sharing. With PACS, hospitals can eliminate physical film usage and provide multisite access to medical images through a web interface.

Electronic Medical Records

In Vietnam, EMR systems refer to systems implemented in hospitals, especially in inpatient wards, to manage patient care information, such as medical histories, problems, orders, tests, and medications, similar to paper charts [24]. EMRs are defined differently than electronic health records (EHRs) in Vietnam, which are defined as health records updated throughout a person's lifetime and can cover a comprehensive range of health information, such as allergies, vaccination, family history, and outpatient visits rather than specialized care information [25]. Although the data collected by EMR systems are independently managed by hospitals, EHR data are synchronized from multiple sources, such as hospital and primary care facility visits to a centralized platform managed by the MoH [26]. In the Vietnam HIT Maturity Model, EMR systems were assessed on 4 areas as follows:

- Clinical data management functions, such as medical histories, clinical documents, and test management
- Administrative and demographics functions, including managing information of health staff and patients and managing integrations between the EMR system and other information systems in a hospital
- Medical record storage capabilities, including the required record retention duration according to health care law, record synchronization, and medical record storage and restoring
- Technical administration functions, including security, supervision, standardized terminology management, standard-based data exchange, EMR workflow management, and database backup and recovery

Extra Capabilities

In the Vietnam HIT Maturity Model, each digital maturity level requires several extra digital capabilities, in addition to the principal information systems, such as HIS, LIS, and EMRs (see Multimedia Appendix 1). These capabilities can be grouped as follows:

- CDRs: These are centralized data stores that gather patient data generated from other clinical information systems. Data and resources from CDRs can also be used by other systems and applications in the same network, such as

XSL•FO
RenderX

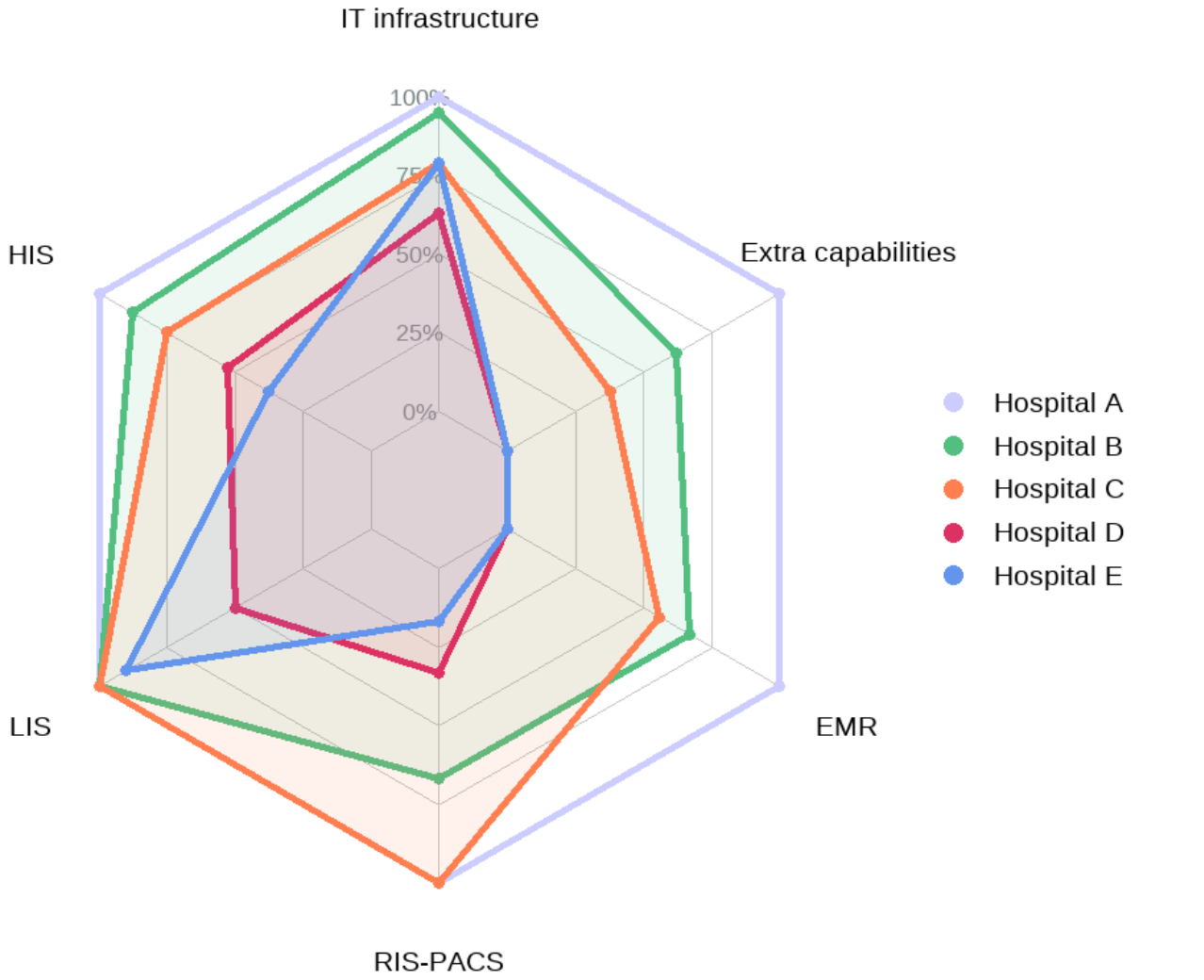
standardized laboratory order vocabularies can be jointly deployed in EMRs and LIS. This assessment examined CDR implementation based on the following criteria: (1) Hospitals can establish a CDR containing standardized vocabularies, medications, orders, and laboratory tests; (2) information in CDRs can be shared with stakeholders involved in patient care; and (3) CDRs contain data on vital signs, nursing notes, and clinical procedures.
- CDSS are categorized into 3 levels: CDSS level 1 can support drug prescribing, including new prescription and represcribing; CDSS level 2 can send alerts for basic conflicts in ordering and prescribing; and CDSS level 3 can inform clinicians' treatment plans and outcomes via appropriately customized alerts.
- Electronic ordering and inpatient order management.
- Digitizing clinical notes using electronic structured templates.
- Closed loop medication management using identification technologies such as radiofrequency identification and barcoding.
- Sharing clinical data with stakeholders involved in patient care through standardized electronic transactions, such as the HL7 Clinical Document Architecture (CDA) and the HL7 Continuity of Care Document (CCD).
- Continuous summaries of service usage data from all the departments, such as inpatient, outpatient, and emergency departments.

Results

Assessment Criteria

Although we were not able to collect information about 3-5-year investment plans from the 5 hospitals, we were able to collect data on all the assessment criteria from the maturity model. The radar plot in Figure 1 presents the percentage of criteria the 5 hospitals met in each domain.

Figure 1. Percentage of HIT criteria satisfied in each domain across the 5 hospitals (IT infrastructure, EMR, LIS, HIS, RIS-PACS, and extra capabilities). EMR: electronic medical record; HIS: hospital information systems; HIT: health information technology; LIS: laboratory information systems; RIS-PACS: radiology information systems–picture archiving and communication systems.



In general, HIS, LIS, and RIS-PACS were implemented in all 5 hospitals, albeit with varied digital maturity levels. Two hospitals were not using any EMR system or fulfilling any extra digital capability. One hospital reported meeting all the criteria across the 6 domains. In the following sections, we describe the

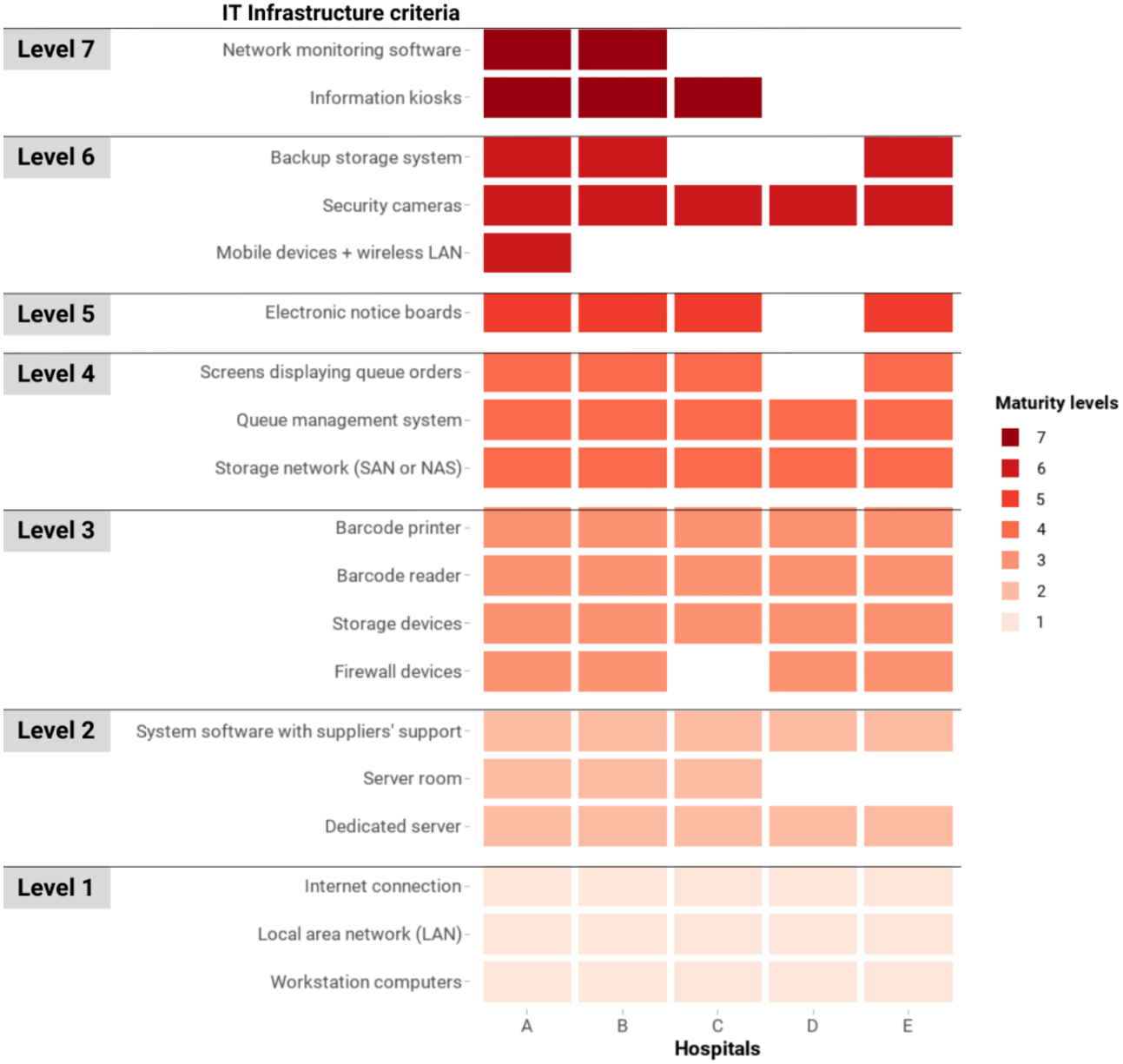
XSL•FO
RenderX

domains in more detail and compare the implementation strategies between the hospitals.

IT Infrastructure

Figure 2 shows the IT Infrastructure criteria that were met by the 5 hospitals. Level 1 criteria were met by all 5 hospitals, and most hospitals reported meeting level 2-5 criteria.

Figure 2. Criteria of IT infrastructure in the Vietnam HIT Maturity Model met by the 5 hospitals in this assessment, where a maximum of level 7 could be obtained. The color coding represents the maturity group that a criterion belongs to, not the entire domain's maturity level of a hospital. Although for official purposes, a hospital needs to meet all the criteria of a lower level to progress to a higher level, this figure represents any criteria attained. HIT: health information technology; LAN: local area network; NAS: network attached storage; SAN: storage area network.



All 5 hospitals had sufficient workstations for staff. They were connected to the internet and local area networks. All 5 hospitals' digital systems and databases were run on dedicated servers, and 3 (60%) hospitals had their own on-premise server rooms with physical controls. The hospitals' data were stored on storage devices that were supported by storage networks, such as network attached storage (NAS) or storage area network (SAN) systems. Backup storage systems were available in 3 (60%) of the 5 hospitals.

Almost all hospitals in this assessment adopted technologies to help reduce errors and improve patient flow. These included barcode readers, barcode printers, queue management systems, screens displaying queue orders, and electronic notice boards. Information kiosks were less common as only 3 (60%) of the 5 hospitals implemented this technology.

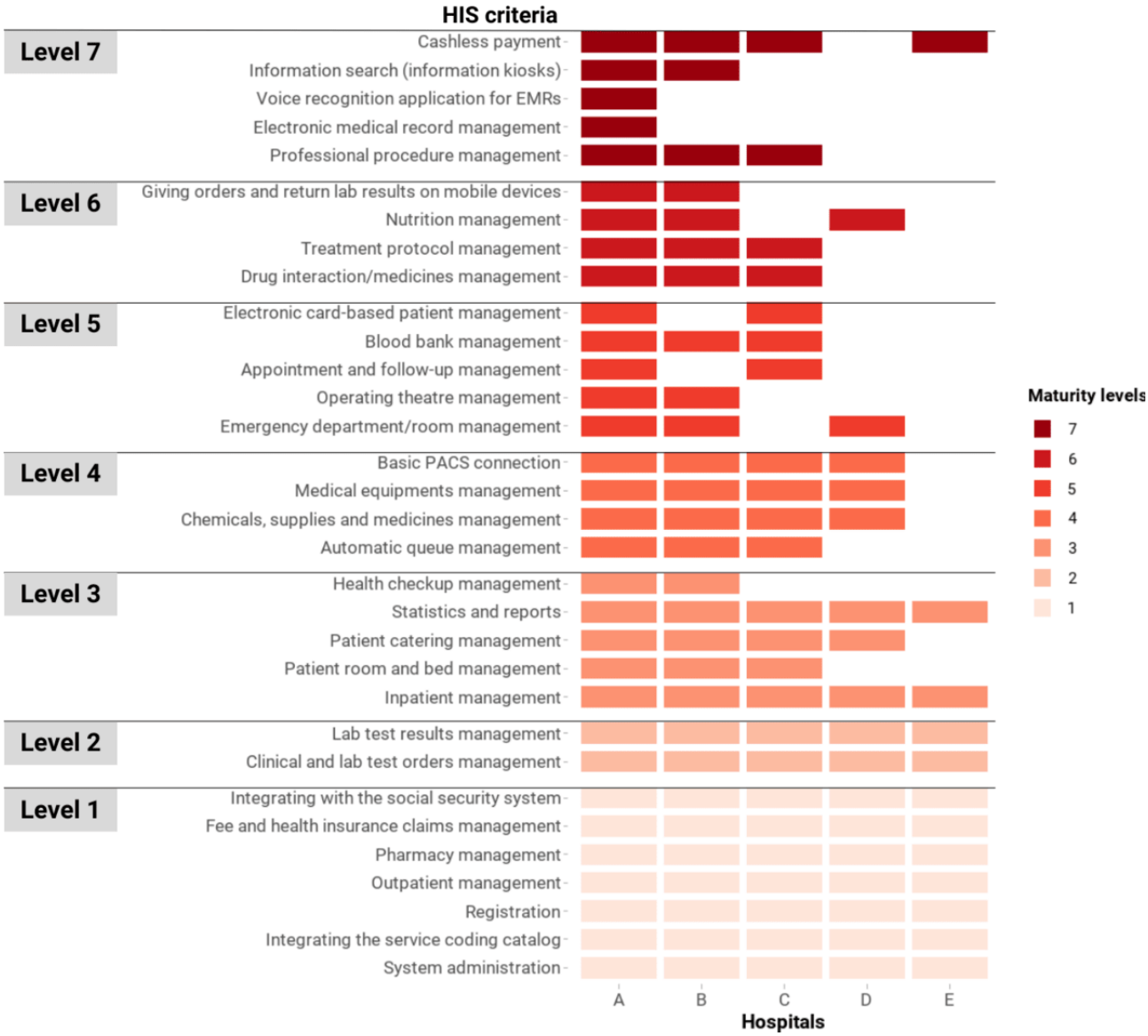
Level 6 and 7 criteria were less available in the hospitals. Only 1 (20%) hospital reported adopting mobile devices, such as tablets and smartphones, and a wireless local area network (LAN) to support hospital activities.

XSL•FO
RenderX

Hospital Information Systems

All 5 hospitals reported that their HIS fully met the level 1 and 2 criteria (Figure 3). These systems were able to support patient registration, outpatient services, and pharmacy stock management. Hospital fees, including social health insurance payments, were also managed in these HIS, where insurance claims were formatted into XML files following a national standard and routinely submitted to the Vietnam Social Security portal. The HIS were reportedly able to integrate Vietnam's national coding systems to standardize the classifications of some administrative units, such as the Vietnam ethnic classification system, and clinical services, such as the WHO International Classification of Disease [27]. Medical orders and test results were assigned unique patient IDs and uploaded to the HIS. Inpatient service management and reporting functionalities (level 3) were also available in the HIS across the 5 hospitals.

Figure 3. Criteria of HIS in the Vietnam HIT Maturity Model met by the 5 hospitals in this assessment, where a maximum of level 7 can be obtained. The color coding represents the maturity group that a criterion belongs to, not the entire domain's maturity level of a hospital. Although for official purposes, a hospital needs to meet all the criteria of a lower level to progress to a higher level, this figure represents any criteria attained. EMR: electronic medical record; HIS: hospital information systems; HIT: health information technology; PACS: picture archiving and communication systems.



Of the 5 hospitals, 4 (80%) reported that their HIS integrated with PACS and cashless payment technology. However, modules for treatment protocols, professional procedures, and drug interaction management were implemented in only 3 (60%) hospitals.

We found the functionalities that were least adopted were the ones that required integration with EMR systems (ie, voice recognition for EMRs and EMR management, available in 1, 20%, hospital). In addition, only 2 (40%) hospitals reported that their HIS could interface with information kiosks, electronic patient cards, and mobile devices.

Laboratory Information Systems

Results of our assessment showed that all 5 hospitals reported that their LIS met the basic criteria (Figure 4). All the LIS could manage hospital laboratory orders and results, as well as generate reports. Most of these systems could automatically send orders to and receive results from laboratory machines.

XSL•FO
RenderX

One hospital reported that its LIS could receive results from laboratory instruments, but the orders could only be entered manually. The national dictionary system for laboratory tests was adopted across all 5 hospitals to standardize laboratory coding. The LIS of 4 (80%) hospitals satisfied all the advanced criteria, which included the ability to integrate with HIS for exchanging orders and results, set alert thresholds for laboratory results, manage chemical supply stocks, and manage laboratory specimens.

Figure 4. Criteria of LIS in the Vietnam HIT Maturity Model met by the 5 hospitals in this assessment, where a maximum of 2 levels can be obtained. The color coding represents the maturity group that a criterion belongs to, not the entire domain's maturity level of a hospital. Although for official purposes a hospital needs to meet all the criteria of a lower level to progress to a higher level, this figure represents any criteria attained. HIS: hospital information systems; HIT: health information technology; LIS: laboratory information systems.

Radiology Information Systems—Picture Archiving and Communication Systems

All 5 hospitals were using RIS-PACS at the time of this assessment, although with different degrees of implementation (Figure 5). Of the 5 hospitals, 2 (40%; A and C) had RIS-PACS that met all the basic and advanced criteria, 1 (20%) hospital (B) nearly satisfied all the basic and advanced criteria, and RIS-PACS from the remaining 2 (40%) hospitals (D and E) were largely operated by external vendors and only implemented a limited number of functionalities of RIS-PACS.

XSL•FO
RenderX

Figure 5. Criteria of RIS-PACS in the Vietnam HIT Maturity Model met by the 5 hospitals in this assessment, where a maximum of 2 levels can be obtained. The color coding represents the maturity group that a criterion belongs to, not the entire domain's maturity level of a hospital. Although for official purposes a hospital needs to meet all the criteria of a lower level to progress to a higher level, this figure represents any criteria attained. DICOM: Digital Imaging and Communications in Medicine; HIS: hospital information systems; HIT: health information technology; HL7: Health Level Seven; RIS-PACS: radiology information systems–picture archiving and communication systems.

RIS-PACS meeting the basic criteria were able to retrieve Digital Imaging and Communications in Medicine (DICOM) images of common imaging modalities, such as X-ray, magnetic resonance imaging, and ultrasound, through 2-way interfaces. The RIS, PACS, and HIS could exchange orders and images with one another based on HL7 messaging standards. The PACS could convert DICOM images to JPEG format, and any changes to images in the PACS could be promptly updated to the HIS. Orders and radiologists' reading result management, measurement, and reporting were also ensured by these RIS-PACS.

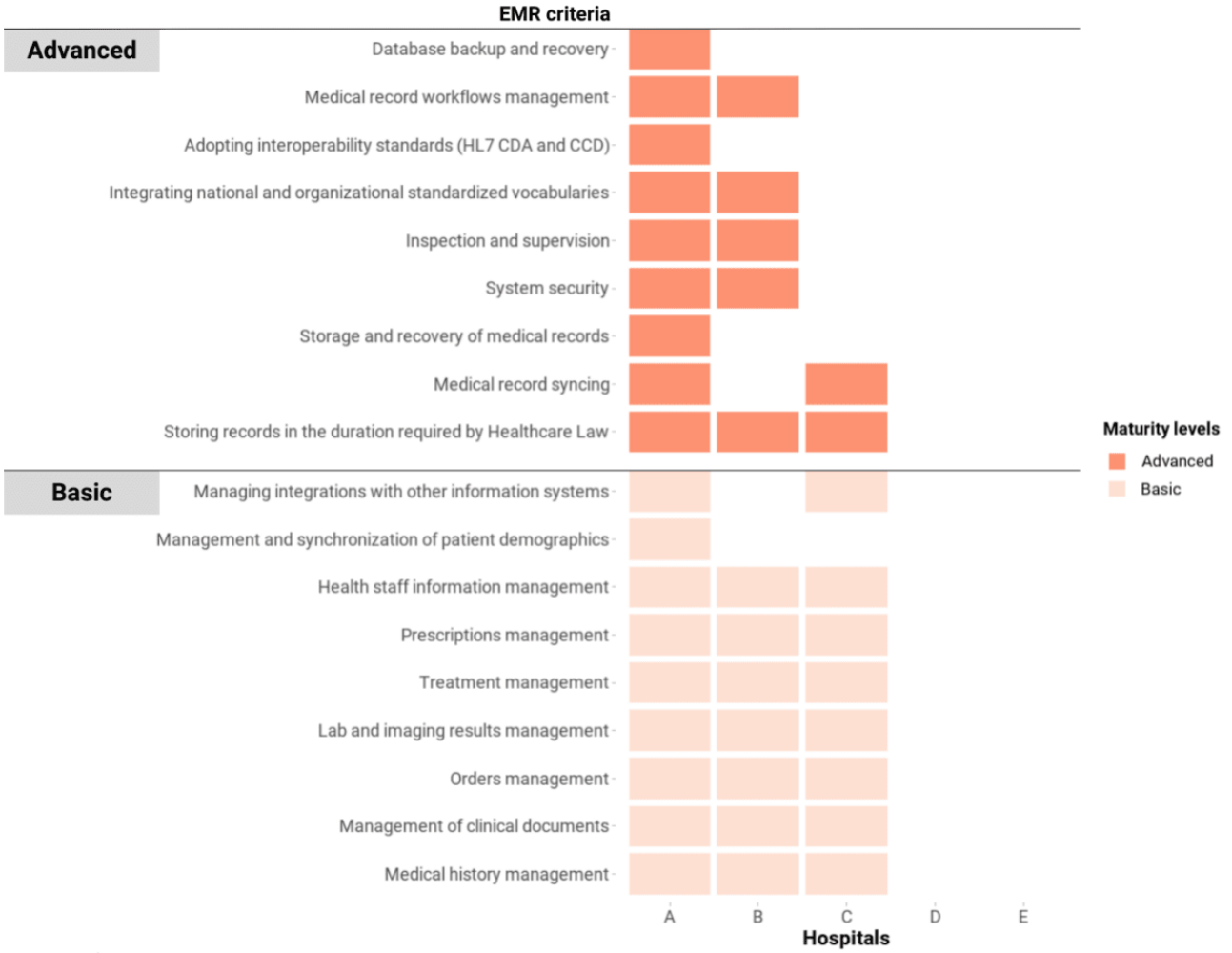
Hospitals meeting the advanced criteria of RIS-PACS could carry out multisite consultations through web-based access to DICOM images. One hospital reported that its PACS had adopted the HL7 FHIR standard.

Electronic Medical Records

Generally, the EMR status was less mature than other domains (Figure 6), with only 3 (60%) hospitals reporting that they were implementing EMR systems. Only 1 (20%) of these satisfied sufficient criteria in the regulations for EMR systems [24], which allowed the hospital to completely eliminate archiving of paper records for backup and legal purposes. Nevertheless, the other EMR systems could provide clinical data management functions, including managing medical histories, clinical documents, orders, laboratory and imaging results, treatments, and prescriptions.

XSL•FO
RenderX

Figure 6. Criteria of EMRs in the Vietnam HIT Maturity Model met by the 5 hospitals in this assessment, where a maximum of 2 levels can be obtained. The color coding represents the maturity group that a criterion belongs to, not the entire domain's maturity level of a hospital. Although for official purposes a hospital needs to meet all the criteria of a lower level to progress to a higher level, this figure represents any criteria attained. CCD: Continuity of Care Document; CDA: Clinical Document Architecture; EMR: electronic medical record; HIT: health information technology; HL7: Health Level Seven.



An EMR system fully compliant with the regulations for EMRs could ensure storage capacity, backup, and recovery capabilities. Interoperability standards, such as the HL7 CDA, the HL7 CCD, and HL7 FHIR, were implemented. Information security measures, such as authentication, audit trail, and encryption, were in place. Digital signatures were used to authenticate medical records and orders.

Extra Capabilities

The availability of these capabilities in the 5 hospitals is shown in Figure 7. Notably, 1 (20%) hospital reported achieving all the extra capabilities, with 2 (40%) hospitals confident about sharing clinical data using HL7 standards (although they were not currently sharing data at the time of the study). The adoption of CDRs and CDSS to support clinical tasks, such as giving orders, prescribing, and clinical decision-making, was only available in 1 (20%) hospital. However, all the hospitals with an EMR system reported being able to digitize all clinical documents and allow clinicians to give orders in the electronic environment.

XSL•FO
RenderX

Figure 7. Extra capability criteria in the Vietnam HIT Maturity Model met by the 5 hospitals in this assessment. CDR: clinical data repository; CDSS: clinical decision support systems; HIT: health information technology; HL7: Health Level Seven.

Discussion

Principal Findings

This study assessed the implementation of some key digital health systems in 5 top-tier hospitals in Vietnam. Overall, many processes related to administration, laboratory tests, and medical imaging were digitized in the hospitals through the implementation of HIS, LIS, and RIS-PACS on the basis of adequate basic IT infrastructures. The digital maturity of these systems, however, varied significantly between the hospitals, although all are large top-tier hospitals. EMR systems were implemented in 3 of the 5 hospitals. Although these systems could satisfy a wide range of data collection needs, their integration with CDRs and CDSS was sparse.

Comparison With Prior Works

This is the first original research that examined digital health system implementation in Vietnamese hospitals since the Vietnam MoH enacted the National EMR plan and the Vietnam HIT Maturity Model in 2018 as part of the sectorwide digital transformation agenda [11,21,24]. Academic publications on adoption and implementation of hospital-based digital health systems are limited [2]. Most of the previous studies have been conducted before 2014 and focused on the readiness to transition from paper-based processes to electronic systems rather than describing DHT adoption [28,29]. Other publications have described the development and pilots of technical solutions, such as LIS, data retrieval from medical devices, and data visualization for EMRs. We could not find any research exploring the digital systems currently used in hospitals in Vietnam [30-33].

Muinga et al [34] showed that most of the digital health systems adopted in Kenyan public hospitals from 2014 to 2016 were aimed at administrative and billing purposes rather than supporting clinical tasks. Systems such as LIS, RIS, and PACS, when available, were usually stand-alone systems and lacked interoperability with other systems in the same institution [34]. Our study found that all the Vietnamese hospitals assessed had gone beyond administrative systems and had used clinical information systems that are interoperable with each other, especially with HIS. EMRs, as a crucial clinical information system, had been implemented in 3 of the 5 hospitals to capture a wide range of data, such as medical histories, clinical documents, orders, and test results. However, the adoption of EMR capabilities, such as CDRs and CDSS, was limited, even though these are large teaching hospitals at the top tier of the Vietnamese health system. This is inferior to the digital maturity status of public hospitals in Turkey, as over 98% of the Turkish

XSL•FO
RenderX

hospitals surveyed reported having CDRs regardless of their size [35]. In addition, CDSS for medication orders and nonmedication orders were implemented in 71% and over 57% of Turkish public hospitals, respectively. The early use of EMRAM to guide the national digital health care transformation may be an important facilitator for this wide adoption of CDRs and CDSS in Turkish hospitals [36].

The Vietnam HIT Maturity Model [11] that this study's questionnaire was based on and the HIMSS EMRAM [9] share several similarities regarding assessment criteria. For example, CDRs, CDSS, closed-loop medication administration, and electronic documentation are included in both models, including the maturity stage that these systems belong to. However, EMRAM particularly focuses on EMR capabilities, in which systems such as laboratory, radiology, and pharmacy are seen as EMR ancillaries, while CDRs and CDSS are crucial components deciding the maturity of the EMR system. All the ancillary systems need to be implemented as early as in stage 1 of EMRAM. The Vietnam HIT Maturity Model, in contrast, splits the ancillaries into basic and more advanced maturity levels along the digital maturity roadmap. Systems such as CDRs and CDSS are the extra capabilities that should be met, alongside other domains, such as HIS, LIS, RIS, and PACS, so that a hospital can move up to a higher digital maturity level. Although the division of digital systems into smaller milestones may allow hospitals to better benchmark their digital maturity and plan for investments, the requirement to meet digital maturity criteria across multiple domains can be challenging for hospitals in a resource-limited setting such as Vietnam. Duncan et al [10] discussed that most published maturity models, including EMRAM, are not assessed in LMICs, questioning their applicability in cultures different from those of high-income countries, such as the United States and the United Kingdom. In addition, the implementation of EMRAM poses the risk of investing in complex systems without meeting the organization's local needs [37].

Integrating New DHTs Into the Existing System

Most hospitals participating in this assessment had a low adoption rate of EMRs, CDSS, and CDRs. The low number of complete EMR systems may be caused by the cost required to upgrade the infrastructure and a lack of the required technical capabilities to meet all the requirements in EMR regulations. It is difficult for hospitals with incomplete EMR systems and CDRs to embark on using CDSS. Integrating, implementing, and maintaining CDSS is usually complex and depends on multiple factors, such as the availability and interoperability of data, especially from EMRs; an updated knowledge base; and understanding of the hospital's workflow [38].

The wide availability of HIS, LIS, and RIS-PACS in the hospitals participating in our study suggests that digital systems for administrative and billing purposes and simpler systems to implement tend to be prioritized. Many hospitals may choose to implement clinical documentation functions before upgrading the other components to fully comply with EMR regulation. This leads to an increasing amount of medical data that are electronically available and useful for AI training and implementation even without full EMR implementation. However, the scarcity of CDSS, CDRs, and standard-based data-sharing experience found in this assessment suggests challenges that stakeholders may face when seeking to integrate AI-enabled solutions in Vietnamese hospitals. Although these challenges should be explored in further research, we believe a lack of national frameworks and detailed guidance for meaningful and secure clinical data exchange is 1 of the key barriers. In addition, recently, Chanh et al [39] mentioned exploring system interoperability, collaboration with local stakeholders from multiple disciplines, and understanding local data-sharing policies as some of the key considerations for applying AI in Vietnam.

Strengths and Limitations

We examined the adoption of digital health systems using the criteria published in the Vietnam HIT Maturity Model. This is a national benchmarking tool for HIT implementation in Vietnamese public hospitals, making it familiar to the hospitals participating in this study.

Only 5 public hospitals at the top-tier level were included in this assessment, so our findings do not represent the overall picture of HIT implementation in Vietnamese hospitals. The hospitals participating in this study were ones that had previous interest in DHT implementation and digital health innovation research. However, it is likely that hospitals at low tiers and located in less-funded regions have lower digital maturity across all the domains, especially in the EMR and extra capabilities domains.

Implications for Further Study

The adoption of digital health systems in this study was assessed based on technological criteria, such as the software functions that were available. Future work could address other dimensions related to the implementation of these systems, such as the organizational context and human factors [40]. The use of qualitative methods can be particularly suitable for such research questions as they can unveil the characteristics unique to technology adoption in hospitals in LMIC settings, such as Vietnam.

Conclusion

Several major public hospitals in Vietnam have a sound digital infrastructure in place to support fundamental administrative and clinical tasks, such as patient management, insurance claims, laboratory result management, and medical imaging inspection. Most of the hospitals have implemented EMR systems to a basic level that prioritizes data collection, management, and access. Nevertheless, the more advanced level of data management and use via CDRs and CDSS is not common. This can be seen as a barrier to the introduction of new DHTs in these hospitals. Along with the increased amount of data collected by the systems, the adoption of HL7, DICOM, and other international standards can be seen as a facilitator for new DHTs, such as AI-based CDSS, to be implemented in these hospitals.

XSL•FO
RenderX

Acknowledgments

This study was supported by funding from the Wellcome Trust United Kingdom (215010/Z/18/Z). The Vietnam Health Information Technology Maturity Model (Circular 54/2017/TT-BYT) was originally published in Vietnamese and was interpreted into English by DMT.

The members of *Vietnam ICU Translational Applications Laboratory (VITAL) investigators* are as follows (alphabetic order by surname):

OUCRU inclusive authorship list in Vietnam: Dang Phuong Thao, Dang Trung Kien, Doan Bui Xuan Thy, Dong Huu Khanh Trinh, Du Hong Duc, Ronald Geskus, Ho Bich Hai, Ho Quang Chanh, Ho Van Hien, Huynh Trung Trieu, Evelyne Kestelyn, Lam Minh Yen, Le Dinh Van Khoa, Le Thanh Phuong, Luu Hoai Bao Tran, Luu Phuoc An, Angela Mcbride, Nguyen Lam Vuong, Ngan Nguyen Lyle, Nguyen Quang Huy, Nguyen Than Ha Quyen, Nguyen Thanh Ngoc, Nguyen Thi Giang, Nguyen Thi Diem Trinh, Nguyen Thi Kim Anh, Nguyen Thi Le Thanh, Nguyen Thi Phuong Dung, Nguyen Thi Phuong Thao, Ninh Thi Thanh Van, Pham Tieu Kieu, Phan Nguyen Quoc Khanh, Phung Khanh Lam, Phung Tran Huy Nhat, Guy Thwaites, Louise Thwaites, Tran Minh Duc, Trinh Manh Hung, Hugo Turner, Jennifer Ilo Van Nuil, Vo Tan Hoang, Vu Ngo Thanh Huyen, Sophie Yacoub | Hospital for Tropical Diseases, Ho Chi Minh City: Cao Thi Tam, Ha Thi Hai Duong, Ho Dang Trung Nghia, Le Buu Chau, Le Mau Toan, Nguyen Hoan Phu, Nguyen Quoc Viet, Nguyen Thanh Dung, Nguyen Thanh Nguyen, Nguyen Thanh Phong, Nguyen Thi Cam Huong, Nguyen Van Hao, Nguyen Van Thanh Duoc, Pham Kieu Nguyet Oanh, Phan Thi Hong Van, Phan Tu Qui, Phan Vinh Tho, Truong Thi Phuong Thao | University of Oxford: Natasha Ali, James Anibal, David Clifton, Mike English, Ping Lu, Jacob McKnight, Chris Paton, Tingting Zhu | Imperial College London: Pantelis Georgiou, Bernard Hernandez Perez, Kerri Hill-Cawthorne, Alison Holmes, Stefan Karolcik, Damien Ming, Nicolas Moser, Jesus Rodriguez Manzano | King's College London: Liane Canas, Alberto Gomez, Hamideh Kerdegari, Andrew King, Marc Modat, Reza Razavi | University of Ulm: Walter Karlen | The University of Melbourne: Linda Denehy, Thomas Rollinson | Mahidol Oxford Tropical Medicine Research Unit (MORU): Luigi Pisani, Marcus Schultz

Data Availability

The data set generated and analyzed during this study is available upon request through the Oxford University Clinical Research Unit's managed data access scheme [41].

Authors' Contributions

All authors made substantial contributions to the conception and design of the study, drafted the work and revised it critically for important intellectual content, and provided final approval of the version to be published. DMT, NTD, CMD, HNH, LMK, NPH, DTTH, HTLT, TVD, and LMY were responsible for data acquisition and analysis, while DMT, CLT, and CP interpreted the data for the study. DMT, NTD, CMD, HNH, LMK, NPH, DTTH, HTLT, and TVD made the agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work were appropriately investigated and resolved. CLT and CP managed funding acquisition. The authors attest that there was no use of generative artificial intelligence technology in the generation of text, figures, or other informational content of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Vietnam HIT Maturity Model issued in 2017 by the Vietnam MoH. HIT: health information technology; MoH: Ministry of Health.
[DOCX File , 22 KB - formative_v9i1e53483_app1.docx]

Multimedia Appendix 2
The hospital health information systems assessment questionnaire.
[DOCX File , 21 KB - formative_v9i1e53483_app2.docx]

References

1. Global Initiative on Digital Health (GIDH) executive summary. World Health Organization. 2023 Feb 1. URL: https://cdn.who.int/media/docs/default-source/digital-health-documents/global-initiative-on-digital-health_executive-summary-31072023.pdf?sfvrsn=5282e32f_1 [accessed 2025-03-01]
2. Tran DM, Thwaites CL, Van Nuil JI, McKnight J, Luu AP, Paton C, Vietnam ICU Translational Applications Laboratory (VITAL). Digital health policy and programs for hospital care in Vietnam: scoping review. J Med Internet Res 2022 Feb 09;24(2):e32392 [FREE Full text] [doi: 10.2196/32392] [Medline: 35138264]

XSL•FO
RenderX

3. Moore GE. Cramming more components onto integrated circuits. Proc IEEE 1998 Jan;86(1):82-85. [doi: 10.1109/JPROC.1998.658762]
4. World Health Organization. Global Strategy on Digital Health 2020-2025. Geneva: World Health Organization; 2021.
5. Liverani M, Ir P, Perel P, Khan M, Balabanova D, Wiseman V. Assessing the potential of wearable health monitors for health system strengthening in low- and middle-income countries: a prospective study of technology adoption in Cambodia. Health Policy Plan 2022 Sep 13;37(8):943-951 [FREE Full text] [doi: 10.1093/heapol/czac019] [Medline: 35262172]
6. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. Lancet 2020 May;395(10236):1579-1586. [doi: 10.1016/s0140-6736(20)30226-9]
7. Lee S, Kim H. Prospect of artificial intelligence based on electronic medical record. J Lipid Atheroscler 2021 Sep;10(3):282-290 [FREE Full text] [doi: 10.12997/jla.2021.10.3.282] [Medline: 34621699]
8. Bender D, Sartipi K. HL7 FHIR: an agile and RESTful approach to healthcare information exchange. 2013 Presented at: 26th IEEE International Symposium on Computer-Based Medical Systems; June 20-22, 2013; Porto, Portugal. [doi: 10.1109/cbms.2013.6627810]
9. EMRAM: a strategic roadmap for effective EMR adoption and maturity. HIMSS Eurasia. 2024 Sep 10. URL: https://himsseurasia.com/en/himss-emram-standards/ [accessed 2025-01-07]
10. Duncan R, Eden R, Woods L, Wong I, Sullivan C. Synthesizing dimensions of digital maturity in hospitals: systematic review. J Med Internet Res 2022 Mar 30;24(3):e32994 [FREE Full text] [doi: 10.2196/32994] [Medline: 35353050]
11. Thông t 54/2017/TT-BYT v B tiêu chí ng d ng công ngh thông tin t i các c s khám b nh, ch a b nh. Vietnam Ministry of Health. 2017. URL: https://thuvienphapluat.vn/van-ban/Cong-nghe-thong-tin/Thong-tu-54-2017-TT-BYT-Bo-tieu-chi-ung-dung-cong-nghe-thong-tin-tai-cac-co-so-kham-chua-benh-373292.aspx [accessed 2023-07-15]
12. Ming DK, Hernandez B, Sangkaew S, Vuong NL, Lam PK, Nguyet NM, Vietnam ICU Translational Applications Laboratory (VITAL) investigators. Applied machine learning for the risk-stratification and clinical decision support of hospitalised patients with dengue in Vietnam. PLOS Digit Health 2022 Jan 18;1(1):e0000005 [FREE Full text] [doi: 10.1371/journal.pdig.0000005] [Medline: 36812518]
13. The Vietnam ICU Translational Applications Laboratory (VITAL). Oxford University Clinical Research Unit. URL: http://vital.oucru.org/ [accessed 2023-07-01]
14. Van HMT, Hao NV, Phan Nguyen Quoc K, Hai HB, Khoa LDV, Yen LM, et al. Vital sign monitoring using wearable devices in a Vietnamese intensive care unit. BMJ Innov 2021 Mar 31;7(Suppl 1):s7-s11. [doi: 10.1136/bmjinnov-2021-000707]
15. Nguyen QH, Ming DK, Luu AP, Chanh HQ, Tam DTH, Truong NT, Vietnam ICU Translational Applications Laboratory (VITAL) investigators. Mapping patient pathways and understanding clinical decision-making in dengue management to inform the development of digital health tools. BMC Med Inform Decis Mak 2023 Feb 02;23(1):24 [FREE Full text] [doi: 10.1186/s12911-023-02116-4] [Medline: 36732718]
16. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2023.
17. Nakazawa M. Package 'fmsb': functions for medical statistics book with some demographic data. Comprehensive R Archive Network. 2023. URL: https://cran.r-project.org/web/packages/fmsb/fmsb.pdf [accessed 2025-01-07]
18. Saary M. Radar plots: a useful way for presenting multivariate health care data. J Clin Epidemiol 2008 Apr;61(4):311-317 [FREE Full text] [doi: 10.1016/j.jclinepi.2007.04.021] [Medline: 18313553]
19. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York, NY: Springer-Verlag; 2016.
20. Le N. C n c ch 70 b nh vi n dùng b nh án đi n t . VN Express. 2024. URL: https://vnexpress.net/ca-nuoc-chi-70-benh-vien-dung-benh-an-dien-tu-4732748.html [accessed 2024-09-10]
21. OxTREC application process. University of Oxford. 2024. URL: https://researchsupport.admin.ox.ac.uk/governance/ethics/apply/oxtrec#collapse404361 [accessed 2024-10-17]
22. Definition of terms: research, audit and service review. University of Oxford. 2024. URL: https://researchsupport.admin.ox.ac.uk/governance/ethics/apply/oxtrec/definitions [accessed 2024-10-18]
23. KPMG Department for International Trade, Oxford University Clinical Research Unit. Digital health in Vietnam. Market intelligence report. KPMG. 2020 Dec. URL: https://assets.kpmg.com/content/dam/kpmg/vn/pdf/publication/2021/digital-health-vietnam-2020-twopage.pdf [accessed 2025-01-07]
24. Thông t 46/2018/TT-BYT quy đ nh v h s b nh án đi n t . Vietnam Ministry of Health. 2018. URL: https://thuvienphapluat.vn/van-ban/Cong-nghe-thong-tin/Thong-tu-46-2018-TT-BYT-su-dung-va-quan-ly-ho-so-benh-an-dien-tu-391438.aspx [accessed 2023-07-15]
25. Quy t đ nh 5349/QĐ-BYT năm 2019 v phê duy t K ho ch tri n khai h s s c kh e đi n t . Vietnam Ministry of Health. 2019. URL: https://thuvienphapluat.vn/van-ban/The-thao-Y-te/Quyet-dinh-5349-QD-BYT-2019-phe-duyet-Ke-hoach-trien-khai-ho-so-suc-khoe-dien-tu-428071.aspx [accessed 2023-07-16]
26. Quy t đ nh 1923/QĐ-BYT năm 2023 phê duy t K ho ch tri n khai N n t ng H s s c kh e đi n t năm. Vietnam Ministry of Health. 2023. URL: https://thuvienphapluat.vn/van-ban/Cong-nghe-thong-tin/Quyet-dinh-1923-QD-BYT-2023-Ke-hoach-trien-khai-Nen-tang-Ho-so-suc-khoe-dien-tu-564088.aspx [accessed 2023-07-17]

XSL•FO
RenderX

27. Quy t đ nh 2035/QĐ-BYT năm 2013 công b Danh m c k thu t v ng d ng công ngh thông tin. Vietnam Ministry of Health. 2013. URL: https://thuvienphapluat.vn/van-ban/Cong-nghe-thong-tin/Quyet-dinh-2035-QD-BYT-nam-2013-Danh-muc-ky-thuat-ung-dung-cong-nghe-thong-tin-trong-y-te-199624.aspx [accessed 2024-09-01]
28. Hochwarter S, Cuong DD, Chuc NTK, Larsson M. Towards an electronic health record system in Vietnam: a core readiness assessment. J Health Inform Dev Countries 2014 Sep 21;8(2):93-103 [FREE Full text]
29. Nguyen KT. Electronic health record readiness assessment in Thai Binh Hospital, Vietnam. IJSR 2018 Mar;7(3):1277-1281. [doi: 10.21275/ART2018949]
30. Landgraf KM, Kakkar R, Meigs M, Jankauskas PT, Phan TTH, Nguyen VN, et al. Open-source LIMS in Vietnam: the path toward sustainability and host country ownership. Int J Med Inform 2016 Sep;93:92-102 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.06.010] [Medline: 27435952]
31. Vo C, CAO T, Doan N, Tu S, Nguyen A, Nguyen B, et al. EMRVisualization for patient progress tracking: interface design and system implementation. IJAE 2019;18(3):117-128. [doi: 10.5057/ijae.IJAE-D-18-00025]
32. Vu DH, Nguyen DT. Design of laboratory information system for healthcare in Vietnam BK-LIS. 2010 Presented at: International Conference on Communications and Electronics 2010; August 11-13, 2010; Nha Trang, Vietnam. [doi: 10.1109/icce.2010.5670692]
33. Hai VD, Thuan ND, Ngoc PP. Automatic retrieving data from medical equipment to create electronic medical records (EMRs) for an e-hospital model in Vietnam. 2011 Presented at: 5th International Symposium on Bio and Medical Informatics and Cybernetics (BMIC); July 19-22, 2011; Orlando FL.
34. Muinga N, Magare S, Monda J, English M, Fraser H, Powell J, et al. Digital health Systems in Kenyan public hospitals: a mixed-methods survey. BMC Med Inform Decis Mak 2020 Jan 06;20(1):2 [FREE Full text] [doi: 10.1186/s12911-019-1005-7] [Medline: 31906932]
35. Kose I, Rayner J, Birinci S, Ulgu MM, Yilmaz I, Guner S, HIMSS Analytics Team, Team MoH. Adoption rates of electronic health records in Turkish hospitals and the relation with hospital sizes. BMC Health Serv Res 2020 Oct 21;20(1):967 [FREE Full text] [doi: 10.1186/s12913-020-05767-5] [Medline: 33087106]
36. Miliard M. HIMSS helps Turkish hospitals go digital. Healthcare IT News. 2013 Nov 15. URL: https://www.healthcareitnews.com/news/himss-helps-turkish-hospitals-go-digital [accessed 2024-09-11]
37. Cresswell K, Sheikh A, Krasuska M, Heeney C, Franklin BD, Lane W, et al. Reconceptualising the digital maturity of health systems. Lancet Digital Health 2019 Sep;1(5):e200-e201. [doi: 10.1016/s2589-7500(19)30083-4]
38. Middleton B, Sittig DF, Wright A. Clinical decision support: a 25 year retrospective and a 25 year vision. Yearb Med Inform 2018 Mar 06;25(S 01):S103-S116. [doi: 10.15265/iys-2016-s034]
39. Chanh HQ, Ming D, Nguyen QH, Duc TM, An LP, Trieu HT, et al. Applying artificial intelligence and digital health technologies, Viet Nam. Bull World Health Org 2023 Jul 01;101(07):487-492. [doi: 10.2471/blt.22.289423]
40. Cresswell K, Williams R, Sheikh A. Developing and applying a formative evaluation framework for health information technology implementations: qualitative investigation. J Med Internet Res 2020 Jun 10;22(6):e15068 [FREE Full text] [doi: 10.2196/15068] [Medline: 32519968]
41. Data sharing policy. Oxford University Clinical Research Unit. URL: https://www.oucru.org/data-sharing-policy/ [accessed 2025-01-07]

Abbreviations

AI: artificial intelligence
CCD: Continuity of Care Document
CDA: Clinical Document Architecture
CDR: clinical data repository
CDSS: clinical decision support systems
DHT: digital health technology
DICOM: Digital Imaging and Communications in Medicine
EMR: electronic medical record
EMRAM: Electronic Medical Record Adoption Model
FHIR: Fast Healthcare Interoperability Resources
HIMSS: Healthcare Information and Management Systems Society
HIS: hospital information systems
HIT: health information technology
HL7: Health Level Seven
LAN: local area network
LIS: laboratory information systems
LMIC: low- and middle-income country
MoH: Ministry of Health

XSL•FO
RenderX

NAS: network attached storage
PACS: picture archiving and communication systems
RIS: radiology information systems
RIS-PACS: radiology information systems–picture archiving and communication systems
SAN: storage area network
SDG: Sustainable Development Goal
WHO: World Health Organization

Edited by A Mavragani; submitted 04.12.23; peer-reviewed by A Hadianfard, S Lloyd; comments to author 27.06.24; revised version received 11.09.24; accepted 13.11.24; published 06.02.25.

Please cite as:
Tran DM, Thanh Dung N, Minh Duc C, Ngoc Hon H, Minh Khoi L, Phuc Hau N, Thi Thu Huyen D, Thi Le Thu H, Van Duc T, VITAL (Vietnam ICU Translational Applications Laboratory) Investigators, Minh Yen L, Thwaites CL, Paton C
Status of Digital Health Technology Adoption in 5 Vietnamese Hospitals: Cross-Sectional Assessment
JMIR Form Res 2025;9:e53483
URL: https://formative.jmir.org/2025/1/e53483
doi:10.2196/53483
PMID:

©Duc Minh Tran, Nguyen Thanh Dung, Chau Minh Duc, Huynh Ngoc Hon, Le Minh Khoi, Nguyen Phuc Hau, Duong Thi Thu Huyen, Huynh Thi Le Thu, Tran Van Duc, VITAL (Vietnam ICU Translational Applications Laboratory) Investigators, Lam Minh Yen, C Louise Thwaites, Chris Paton. Originally published in JMIR Formative Research (https://formative.jmir.org), 06.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Adapting a Mobile Health App for Smoking Cessation in Black Adults With Anxiety Through an Analysis of the Mobile Anxiety Sensitivity Program Proof-of-Concept Trial: Qualitative Study

Marshall K Cheney[1], PhD; Adam C Alexander[2], PhD; Lorra Garey[3], PhD; Matthew W Gallagher[3,4], PhD; Emily T Hébert[5], PhD; Anka A Vujanovic[6], PhD; Krista M Kezbers[2], PhD; Cameron T Matoska[3], BS; Michael J Zvolensky[3,4,7], PhD; Michael S Businelle[2,8], PhD

[1]Department of Health and Exercise Science, University of Oklahoma, Norman, OK, United States

[2]Tobacco Settlement Endowment Trust (TSET) Health Promotion Research Center, University of Oklahoma Health Sciences Center, Oklahoma City, OK, United States

[3]Department of Psychology, University of Houston, Houston, TX, United States

[4]Helping Everyone Achieve a LifeTime of Health (HEALTH) Institute, University of Houston, Houston, TX, United States

[5]Department of Health Promotion and Behavioral Sciences, UTHealth School of Public Health, Austin, TX, United States

[6]Department of Psychological and Brain Sciences, Texas A&M University, College Station, TX, United States

[7]Department of Behavioral Science, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

[8]Department of Family and Preventive Medicine, University of Oklahoma Health Sciences, Oklahoma City, OK, United States

Corresponding Author:
Marshall K Cheney, PhD
Department of Health and Exercise Science
University of Oklahoma
1401 Asp Ave
Norman, OK, 73019
United States
Phone: 1 4053255211
Email: marshall@ou.edu

Abstract

Background: At least half of smokers make a serious quit attempt each year, but Black adults who smoke are less likely than White adults who smoke to quit smoking successfully. Black adults who smoke and have high anxiety sensitivity (an individual difference factor implicated in smoking relapse and culturally relevant to Black adults) are even less successful. The Mobile Anxiety Sensitivity Program for Smoking (MASP) is a smoking cessation smartphone app culturally tailored to Black adults who smoke to increase smoking cessation rates by targeting anxiety sensitivity.

Objective: This study examined the acceptability and feasibility of the MASP smartphone app following a 6-week pilot test through postintervention qualitative interviews.

Methods: The MASP smoking cessation app was adapted from an evidence-based app by adding culturally tailored narration and images specific to the Black community, educational content on tobacco use in the Black community and the role of menthol, culturally tailored messages, and addressing tobacco use and racial discrimination. The MASP app was piloted with 24 adults with high anxiety sensitivity who identified as Black, smoked daily, and were not currently using medications or psychotherapy for smoking cessation. At the end of the 6-week pilot test, 21/24 participants (67% female; 95.2% non-Hispanic; mean age=47.3 years; 43% college educated; 86% single or separated) completed an audio-recorded semistructured interview assessing the acceptability and utility of the app, individual experiences, barriers to use, the cultural fit for Black adults who wanted to quit smoking, and identified areas for improvement. Transcribed interviews were coded using NVivo (Lumivero), and then analyzed for themes using an inductive, use-focused process.

Results: Most participants (17/21, 81%) had smoked for more than 20 years and 29% (6/21) of them smoked more than 20 cigarettes daily. Participants felt the MASP app was helpful in quitting smoking (20/21, 95%) and made them more aware of smoking thoughts, feelings, and behaviors (16/19, 84%). Half of the participants (11/21, 52%) thought the combination of medication and smartphone app gave them the best chance of quitting smoking. Themes related to participant experiences using

XSL•FO
RenderX

the app included establishing trust and credibility through the recruitment experience, providing personally tailored content linked to evidence-based stress reduction techniques, and self-reflection through daily surveys. The culturally tailored material increased app relevance, engagement, and acceptability. Suggested improvements included opportunities to engage with other participants, more control over app functions, and additional self-monitoring functions.

Conclusions: Adding culturally tailored material to an evidence-based mobile health (mHealth) intervention could increase the use of smoking cessation interventions among Black adults who want to quit smoking. Qualitative interviews provide mHealth app developers important insights into how apps can be improved before full study implementation and emphasize the importance of getting feedback from the target population throughout the development process of mHealth interventions.

Trial Registration: ClinicalTrials.gov NCT04838236; https://clinicaltrials.gov/ct2/show/NCT04838236

(JMIR Form Res 2025;9:e53566) doi:10.2196/53566

KEYWORDS

cultural tailoring; tailoring; African American; black; smoking cessation; mHealth; smartphone application; just in time adaptive intervention; qualitative; formative evaluation; app; application; anxiety; adult; qualitative analysis; smoking; mobile phone; tobacco

Introduction

Black or African American (hereafter referred to as Black) adults who smoke experience significant tobacco-related health disparities [1]. Black adults who smoke are at higher risk for tobacco-related death and disease, including lung cancer [2,3], heart disease, and stroke [4]. On average, Black adults who smoke initiate smoking later in life and smoke fewer cigarettes per day relative to their European American or White counterparts [5,6]. Yet, Black adults who smoke demonstrate greater cigarette dependence and are less likely to quit relative to other racial and ethnic groups [7,8], despite engaging in more quit attempts [8,9]. The difficulty this group faces in quitting successfully highlights the wide range of tobacco-related disparities that disproportionately impact Black adults who smoke (eg, tobacco advertising in predominately Black neighborhoods) [10-13].

One additional factor that can influence the success of a quit attempt and sustained smoking cessation among Black adults who smoke is anxiety sensitivity. Anxiety sensitivity is a relatively stable individual trait predisposing individuals to focus on internal bodily feelings of distress and to believe that anxiety and the bodily symptoms they experience from anxiety are uncommon or indicate an underlying threat to their health [14,15]. People who smoke and have high anxiety sensitivity can experience higher nicotine withdrawal symptoms and more intense cravings when they quit smoking [16], they are more prone to smoke when they have smoking-related thoughts, feelings, and sensations during cessation attempts, and they are more likely to smoke again soon after they have initiated a quit attempt [17-19]. Anxiety sensitivity is prevalent in up to a third of people who smoke across all races and ethnic groups [20,21]; however, Black people who smoke and have higher levels of anxiety sensitivity report experiencing the internal physical symptoms of nicotine withdrawal more intensely than Hispanic and White people who smoke [22,23]. Black people who smoke were also more likely to perceive these symptoms were out of their control or were worsened by environmental stressors such as racial discrimination [22,23]. Such processes could contribute to Black people being less successful in quitting smoking or put them more at risk of returning to smoking to help alleviate stress and anxiety [19].

One promising smoking cessation strategy for Black people with anxiety sensitivity is the delivery of tailored interventions through smartphone apps. Mobile health (mHealth) interventions can provide low-cost access to evidence-based smoking cessation strategies that extend the reach of traditional smoking cessation interventions. mHealth interventions can increase access for people who smoke but do not wish to engage in traditional smoking cessation interventions, those not interested in face-to-face interventions, and those who live in remote or underserved areas [24,25]. Smartphone smoking cessation apps have high acceptability in people who wish to quit smoking [25-27] and some smoking cessation apps have demonstrated efficacy in helping people to quit smoking, particularly for the most engaged app users [24,28-30]. While Black adults are underrepresented in mHealth studies [31], Black adults who have participated in mHealth studies viewed the interventions positively and found them to be acceptable and easy to use [32]. Thus, mHealth approaches may also reduce disparities in smoking cessation rates for Black people who smoke [33].

A benefit of mHealth smartphone apps is their potential to tailor intervention content to increase its relevance and acceptability for people who smoke. Tailoring involves selecting intervention content that corresponds to participant characteristics such as beliefs, identity, or readiness to change that can influence the success of a behavior change intervention and provide relevant content in the moments it is needed most [25,34-36]. Specifically, tailoring is the process of modifying the specific content a person receives in a mHealth smoking cessation intervention through *personalization* (based on regular participant assessments) and *feedback* (presenting participants with information about their behaviors over the course of the intervention) [34].

Cultural tailoring is an additional tailoring process of modifying an evidence-based mHealth behavioral intervention to increase the cultural relevance or cultural sensitivity of the intervention to a specific cultural group [37,38]. Cultural tailoring can include adjusting or adding intervention content to better match cultural and historical experiences, social norms and values, and

XSL•FO
RenderX

meanings of tobacco use to a specific cultural group [37]. Culturally tailored mHealth interventions can also address unique risk and protective factors, incorporate more inclusive language, and include culturally specific images and messages to decrease problems of acceptability, inclusion, and engagement [25,35,38,39]. Although it is not clear if culturally tailored materials increase the ultimate effectiveness of interventions, previous work has shown that Black people who smoke preferred culturally tailored content over standard content, which could ultimately increase acceptability and engagement with smoking cessation interventions, and improve retention rates [39-43]. Pilot testing culturally tailored interventions before full study implementation can verify if adapted content is acceptable and relevant to a cultural subgroup and can be used to obtain feedback on ways to improve the cultural relevance of the intervention [37,44].

The purpose of this study was (1) to assess participant reactions to an evidence-based smoking cessation smartphone app that was culturally tailored for Black people who smoke and have high anxiety sensitivity and (2) to provide recommendations for additional modifications to the app before full study implementation.

Methods

Mobile Anxiety Sensitivity Program for Smoking Adaptation

The Mobile Anxiety Sensitivity Program for Smoking (MASP) is a 6-week smoking cessation smartphone app for Black people with high anxiety sensitivity that was adapted from an evidence-based mHealth smartphone app for people with anxiety sensitivity [30,36,45,46]. The app contained a balance of standard intervention content for all participants (ie, on-demand tips, videos, and guided exercises) and individually tailored content based on the participant's responses to baseline, daily, and weekly assessments (ie, just-in-time messages; Figure 1) [46].

XSL•FO
RenderX

Figure 1. MASP (Mobile Anxiety Sensitivity Program) app components and cultural tailoring.

Component	Description	Cultural Tailoring	Participant Quotes (participant number)
Treatment Videos	16 4-6 minute intervention videos adapted from traditional anxiety sensitivity treatments: • relationship between stress and smoking • unhelpful thinking • myths about smoking • the importance of nicotine replacement therapy • strategies for cessation	Black male and female narrators for video content Education about menthol tobacco products, history of tobacco marketing to Black smokers, impact of racism and racial discrimination on stress and smoking	They gave me the history of African Americans of how things affect us differently…. How we wound up [with] menthol being the preferred cigarette… so that really made me wanna read more. P10
Coping Toolkit	• Guided relaxation and mindfulness exercises • challenging automatic thoughts • tips for coping with stress • managing uncomfortable sensations	Relating physical symptoms of the process of quitting smoking to anxiety, discrimination, and withdrawal.	just trying to get your body used to those sensations of being uncomfortable and being in that moment and it's OK to feel like this… it's gonna go away. P16
Quit Tips	On-demand education on nicotine replacement therapy and coping with smoking lapse triggers: • Nicotine medication tips • General Help for Quitting • Benefits of Quitting • Ways to Cope with Urges • Managing Stress/Mood • Motivate Me to Stay Quit • Harms of Smoking	Address concerns about pharmacotherapy, education about menthol	being around your spouse or your partner that smoke… social gatherings that smoke … it gave you an idea of how to deal with it. P5
Stress Management Trainings	3 x day participants received prompts for exercises, including: • over breathing • straw breathing • running in place • chair spinning	No tailoring – evidence-based standardized techniques	Just telling me how to handle this the stress that I was going through at the time and just breathing through it. P11
Daily Tailored Messages	Personally- tailored messages sent in response to EMA assessments. • inspirational messages • tailored to level of stress or craving • link to stress-management exercises	Mix of culturally tailored messages and evidence-based messages	it basically told me …how high the percentages from my culture and the bits of not smoking. P21

Based on the model proposed by Barrera and Castro [37], the MASP app was adapted for Black people who smoke with high anxiety sensitivity in an initial pilot study by incorporating culturally tailored content based on theoretical and empirical guidelines [47]. The MASP app was then reviewed by the University of Houston HEALTH Research Institute Community Research Advisory Board, which is part of a broader U54 addressing health equity. The Community Research Advisory Board provided input into video content and narration, clearer and more engaging language, and recruitment materials and strategies. From this input, key MASP app features were revised to include more culturally relevant content including treatment videos, messages, and on-demand content (Figure 1 and Multimedia Appendices 1 and 2). Examples of culturally tailored content included using Black male and female voices to narrate the videos, drawings, and illustrations of Black artists, the inclusion of additional videos addressing menthol tobacco products, the history of tobacco marketing to Black adults, and the impact of racism and discrimination on stress, smoking, and smoking relapse (Figure 1).

Expanded message content was created that related physical symptoms of quitting smoking to anxiety, discrimination, and severity of withdrawal symptoms, as well as messages relevant to Black culture (Textbox 1). Expanded educational content addressed topics relevant to Black people who smoke, including concerns about nicotine replacement therapy and medications to quit smoking. Several areas of the app were not modified, including skills-based approaches such as standard interceptive stress exposure videos, because of the universal applicability of the content [46]. The revised app reflected a balance of culturally tailored and standard material, consistent with other culturally adapted interventions [43].

XSL•FO
RenderX

Textbox 1. Examples of Mobile Anxiety Sensitivity Program culturally tailored messages for Black smokers.

- Your Tip: (FirstName), Did you know that 45,000 Black people die from smoking-related diseases each year in the US?
- Your Tip: (FirstName), 4 out of 5 Black smokers buy menthols. Chemicals are added to menthol cigarettes to give them a fresh, minty taste. This can make it easier for a smoker to inhale deeply, which may allow more harmful chemicals to enter the lungs.
- Your Tip: (FirstName), Did you know that the money that Black smokers spend on cigarettes in a single day could send 2500 Black students to college for an entire year?
- Your Tip: (FirstName), Tobacco companies have pushed menthol products on the Black community for decades. Did you know that there are significantly more menthol advertisements at stores in neighborhoods with a higher proportion of African American residents?
- Your Tip: (FirstName), Each year smoking kills more African Americans than AIDS, car accidents, fires, alcohol, drugs, and guns combined!
- Your Tip: (FirstName), Did you know African American children and adults are more likely to be exposed to secondhand smoke than any other racial or ethnic group? By quitting smoking you're not only making life better for yourself, but for the people around you too.
- Your Tip: (FirstName), Tobacco companies have targeted the African American community for decades. By quitting smoking, you're not only standing up for yourself but also for the members of your community.

Recruitment

Participants were recruited through Facebook (Meta) advertisements, mass emails through the University of Houston, and fliers posted in organizations and businesses in Houston communities with high concentrations of Black people. Individuals who saw the recruitment materials accessed the screening questionnaire through a posted link on the recruitment materials to assess their eligibility for the study. Individuals were eligible to participate in the pilot study if they (1) self-identified as Black or African American, (2) smoked daily (at least 10 cigarettes per day) for at least 2 years, (3) reported high Anxiety Sensitivity (Short Scale Anxiety Sensitivity Index score of at least 5), (4) endorsed a desire to quit smoking 2 weeks after completion of the baseline assessment, (5) were fluent in English, (6) had no current use of pharmacotherapy or psychotherapy for smoking cessation, and (7) had no cognitive impairment.

Fully enrolled participants (N=24) were instructed to download the INSIGHT (Insight Enterprises Inc) app to their personal smartphone, and those without a compatible phone were provided a study smartphone loaded with the MASP app. All participants were then asked to complete the baseline assessment in the app. Following baseline assessments and a final phone call appointment with project staff members, participants initiated the 6-week intervention. During the 6-week intervention, participants received 5 daily brief assessments (ie, 1 in the morning, 1 in the evening, and 3 at random times throughout the day) followed by individually tailored smoking cessation messages. Along with the daily surveys, participants were prompted to complete weekly follow-up assessments and self-reported smoking status. Participants also had access to stress management and relaxation videos, on-demand tips, strategies for smoking cessation and coping with cravings and urges, and no-cost nicotine replacement therapy. At the end of the intervention, participants used the app to complete a final assessment survey and were asked to participate in an online interview about their experiences with the app [46]. At the end of the 6-week intervention, participants were compensated US $45 for completing the week 6 survey, carbon monoxide breath test, and qualitative interview [46].

Qualitative Interviews

At the end of the 6-week pilot test, 21 of 24 (88%) participants agreed to participate in an interview about their experiences with the app. The interviews took place between March and August 2022. Two interviewers trained in qualitative interviewing techniques conducted the interviews over Zoom (Zoom Video Communications). The interviews with individual participants lasted an average of 44 minutes each. The interview question path was based on a previous study [48] and adapted to address the acceptability and utility of the app for smoking cessation, individual experiences with each component of the app, barriers to use, the relevance and fit for Black people who smoke, and identify areas for revision before full study implementation (Textbox 2 for question path excerpts).

Textbox 2. Excerpts from the Mobile Anxiety Sensitivity Program interview question path.

- What could have been done differently to help you complete more of the daily surveys?
- What are some specific things you liked about the app?
- What things did you like least about the app?
- If you had a slip-up during your quit attempt, was there ever a time that you didn't record that on the app?
- You received some different messages from the app during the study. Can you tell me about a time when you found the message to be really helpful when coping with stress?
- Can you tell me about a time when you found the message you received to be really helpful for quitting smoking?
- If you could add one thing to the app to make it more helpful to future smokers, what would you add?
- How relevant do you think the app is for Black Americans or African Americans who want to quit smoking?
- Do you have any suggestions for any way that we can improve the cultural tailoring of the app for Black Americans or African American smokers?

XSL•FO
RenderX

The audio-recorded interviews were transcribed verbatim, reviewed against the audio recording for accuracy, and then imported into NVivo (version 11; Lumivero) for data analysis. A qualitative researcher (MKC) first read through the transcripts, developed the codebook, and coded the transcripts in 2 coding passes. The codebook was developed using a usage-focused lens [49]. For the first coding pass, the individual interview questions were used as codes, which captured the specific functions of the MASP app, how participants interacted with the app, response to cultural tailoring, and the relevance of app content to Black people who wanted to quit smoking. The second coding pass used a pragmatism lens, coding for additional meanings and uses of the app across interview questions [50,51]. Following interview coding, a thematic analysis was conducted. The coded material was reviewed first within each code, then in groupings of related codes, and finally across all codes. Data displays were also created by grouping participant responses over a set of interview questions or codes to assess additional patterns of responses for individual participants [52]. A second qualitative researcher (AA) read the transcripts, reviewed the coding and primary data displays, and provided additional feedback. The qualitative researchers then arrived at the final determination of themes and quotes through an iterative process of feedback and discussion. A final review of the transcripts was then conducted for confirming and disconfirming evidence of themes [52]. Representative quotes were independently selected and distributed across participants.

Ethical Considerations

The University of Houston (reference number: 12747) and the University of Oklahoma Health Sciences Center (reference number: STUDY00000360) institutional review boards approved the protocol presented in this study. This trial was registered at ClinicalTrials.gov (NCT04838236).

Results

Participant Characteristics

Most (15/21, 71%) of the participants who completed the qualitative interviews were age 40 years or older (mean 47.33, SD=9.6), almost half (9/21, 43%) were college graduates, two-thirds were female (14/21, 67%), most participants were non-Hispanic (20/21, 95%), and reported being single (18/21, 86%). The majority smoked less than a pack of cigarettes per day (15/21, 71%), had smoked for 20 years or longer (17/21, 81%), smoked menthol cigarettes (20/21, 95%), and smoked an average of 15.48 (SD 5.02) cigarettes per day. Participants had moderate levels of nicotine dependence (mean 3.10, SD 1.45), and anxiety sensitivity (mean 11.00, SD 5.40). Most participants were highly motivated to quit smoking (Table 1).

XSL•FO
RenderX

Table 1. Interview participant characteristics (n=21).

Participant characteristics	Values
Age (years), mean (SD)	47.33 (9.6)
Age groups (years), n (%)	
30-39	6 (29)
40-49	6 (29)
50 or more	9 (43)
Education, n (%)	
Less than high-school degree	2 (10)
High-school graduate or General Educational Development degree	4 (19)
Some college	6 (29)
College or graduate degree	9 (43)
Sex, n (%)	
Female	14 (67)
Male	7 (33)
Ethnicity, n (%)	
Non-Hispanic	20 (95)
Hispanic	1 (5)
Marital status, n (%)	
Married or living with significant other	3 (14)
Single or separated	18 (86)
Employment, n (%)	
Regular full-time work	7 (33)
Regular part-time work	3 (14)
Stay-at-home caregiver	2 (10)
Unemployed or unable to work	9 (43)
Average number of cigarettes smoked per day, n (%)	
10-19	15 (71)
20 or more	6 (29)
Years smoked, n (%)	
10-19	4 (19)
20-29	12 (57)
30 or more	5 (24)
Motivation to quit (range 1-10), n (%)	
5-7.9	3 (14)
8-9.9	11 (52)
10	7 (33)

Before the start of the intervention, participants were asked about their preferred method for quitting smoking. While 17/21 indicated medications (81%) as their preferred cessation method, 11 (52%) also selected smartphone app. Half (n=11) of participants also indicated that the combination of medications and smartphone app would give them the best chance of quitting smoking (Table 2). Following the 6-week intervention, the majority of participants (16/19, 85%) thought that having the app and answering the questions increased their awareness of their smoking behaviors and 20/21 interviewees reported that the app was helpful in quitting smoking. The mean score on the African American Acculturation Scale was 15, indicating a medium level of Black cultural identification [53].

XSL•FO
RenderX

Table 2. Interview participant beliefs (n=21).

Participant beliefs	Values
Preintervention questions, n (%)	
If you were to quit smoking, which of the following would you prefer to receive?	
Medications	17 (81)
Group counseling	4 (19)
Smartphone app	11 (52)
In-person individual counseling	6 (29)
Smoking cessation telephone helpline	6 (29)
Which option would give you the best chance of quitting smoking?	
Medications	4 (19)
Both medications and counseling	4 (19)
Both smartphone apps and medications	11 (52)
Quitting “cold turkey” without counseling or medications	2 (10)
Postintervention questions, n (%)	
Consider the number of assessments that were automatically pushed by the smartphone app, was the number of assessments.	
Too high	11 (55)
About right	8 (40)
Not enough	1 (5)
Did carrying the phone and answering questions make you more aware of your thoughts, feelings, and behavior?	
Definitely or mostly no	3 (16)
Definitely or mostly yes	16 (84)
Overall, how helpful has the smartphone app been in helping you to quit smoking?	
Not at all helpful	1 (5)
Slightly or moderately helpful	5 (25)
Very or extremely helpful	14 (70)
African American Acculturation Scale score range (mean 15, SD 6.5, range 6-27)	
Low (score 1-9)	5 (24)
Medium (score 10-18)	9 (43)
High (score 19-27)	7 (33)

Themes

Thematic analysis identified 6 themes related to participant experiences using the app which are (1) recruitment experience, (2) overall experience with the app, (3) most helpful aspects of the app, (4) reaction to the culturally-tailored material, (5) experience with the daily surveys and other assessments, and (6) suggested changes to make the app more helpful or relevant to Black people who want to quit smoking (Figure 2 for a summary of themes).

XSL•FO
RenderX

Figure 2. Summary of themes.

Theme	Brief Summary	Example Quotes
Recruitment experience	Participants stressed the importance of determining study legitimacy and credibility through the professional appearance of the recruitment materials and linkage to an academic institution. Participants were also attracted to the novel intervention delivered by smartphone app.	*it was through the [University] … the lab was a legit thing, I felt much more confident. P12* *I've never seen a app to help you quit anything before* P17
Overall experience with the app	Ease of use and the flexibility and privacy the app provided. The supportive, non-judgmental tone of the messages and other content kept participants motivated and engaged even if they were not able to quit. Daily surveys and messages promoted self-reflection.	*it made me become accountable. It made me think about what I was doing… It made me just take a look at my patterns P10*
Most helpful aspects of the app	The integrated content linking messages and on-demand relaxation and stress reduction exercises increased coping skills in stressful situations and provided alternative strategies to smoking. The personally tailored daily messages increased the relevance of the intervention content. Having the app with them helped them avoid smoking in the moment.	*after I did my survey it was like do you want to do a coping exercise? P18* *I get in the house. First thing…and smoke a cigarette. With the app, I wouldn't do that. I would go to the app first …. using meditations on there. P8*
Reaction to culturally tailored material	Culturally tailored material increased relevance and engagement with the content. To increase relevance to Black smokers, participants suggested adding quotes or stories from Black former smokers, increasing the diversity of images, including more Black health care professionals, incorporate Black vernacular, and provide more information on tobacco use in the Black community.	*I think it's good is because you get the history, you get the fact that somebody understands. P9* *when you open up the app and being able to be it's talking like me it's looking like me. P12*
Experience with daily surveys and other assessments	Surveys increased self reflection and accountability. Suggestions to improve survey completion rates were to decrease survey frequency and question repetition, increase the window for response time, and make survey times more flexible to accommodate work schedules.	*it was more helpful to me answering them… made me not want to smoke. P15* *after so many weeks of doing the same thing ... that's when I started losing some focus. P11*
Suggested changes to improve the app	Provide an option to interact with other participants through the app. Increase self-monitoring functions to track progress.	*some way to be able to connect with …how has everyone done today? P18*

Recruitment Experience

Participants were asked about their initial reaction to the recruitment materials and what prompted them to complete the screening. Important factors in the recruitment process included the legitimacy of the study, the intervention was delivered using a smartphone app, and the inclusion of more familiar traditional nicotine replacement products.

Participants described the professional appearance of the recruitment materials, direct linkage with a university, and that the screening questions matched the stated purpose of the study as important factors in their initial participation. Participants were positive about being part of a real research study but did not want to be drawn into a for-profit research study. Two participants mentioned that they were skeptical but curious, looking up the university and the specific lab to make sure it was an actual study.

> *I thought this is probably a scam. But then I thought, okay, well I just won't give anything too personal, and as I started completing the survey, the questions*

XSL•FO
RenderX

> *were, they sounded legit. They were actually about my habit of smoking, and it just sounded scientific enough that I thought, okay yeah, this probably isn't a scam …. even though it was on Craigslist.* [Participant 12 (P12)]

Participants were interested in the study when they learned the intervention would be delivered using a smartphone app. This was a new strategy when others had not worked and would allow them more independence during the intervention.

Overall Experience With the App

Participants commented that they liked the app and found it easy to use and navigate, particularly for adults who are not "tech savvy" or who have not had much experience using smartphone apps. They reported that the app motivated them to keep trying to quit throughout the intervention, even if they were unable to completely quit smoking. Participants also said the supportive, encouraging, nonjudgmental tone of the app was very important to reporting their true behaviors and feelings.

> *the app … gives you the encouragement of why you don't have to do that [smoking]. There are other things you can do until you get better at not smoking… to help you along the way during that little minutes when it's difficult.* [P10]

> *it's not belittling you or anything and still giving you encouragement … they make you change your thought process into a positive one, so I think that's amazing.* [P16]

Over half of the participants also said that what they liked most was that the app helped them learn more about themselves and their smoking, made them aware of their smoking behaviors and habits, and "kept me accountable to myself."

> *The app made you think about the smoking you were doing. Normally you would just pick up a cigarette and smoke, but the app forces you like OK, do I really need that cigarette right now…. You're thinking about it more. Well, whereas smoking was just an extension of you before.* [P20]

About a third of participants mentioned what they liked the most was that the app was something they could do privately rather than with a group of others, trying different options when they were ready and at their own pace.

> *I was in a group before, and I didn't really get too much out of it… I'm by myself and I'm doing this alone. I'm with the app, I can think more clearly, and I think this is better than being in the group.* [P2]

> *It gives you anonymity. You don't have to tell anybody…It allowed me to be independent, and not have to speak to someone every day … it just allowed me to do it basically at my pace. I didn't feel pressured that I had to do it.* [P10]

A third of the participants mentioned the nicotine replacement products integrated into the intervention.

> *throughout the day, it would remind me to take more lozenges … that was cool.* [P6]

Most Helpful Aspects of the App

Participants found the tailored app content helpful for their current challenges with stress, anxiety, and smoking even if they were not able to quit smoking during the pilot study. Participants commented that the combination of app-directed content and access to complementary on-demand content was a real strength of the intervention. When participants mentioned a specific part of an app that was most helpful to them, they most often mentioned the relaxation and stress reduction exercises and videos, on-demand information that provided them with tips and suggestions, education about tobacco use, and the daily messages (Table 1).

> *They are targeted to what you put into the app yourself, you would say my anxiety is so much or I'm not able to deal with my anxiety. They would give you tips for that and I thought that was really helpful.* [P9]

> *It would tell you after you did the assessments, would you like to take a relaxation or would you like to go or just to go to your toolkit, and that would remind you after you took the assessments to actually do that.* [P14]

However, if the participant was at work, driving, or busy with family responsibilities, they found it difficult to read or act on a particular message at that time.

Across the different parts of the app, participants said that it was the content related to coping that was particularly helpful, either in the form of a tailored supportive message, developing a new skill to reduce stress or fight cravings, or through exercises to challenge unhelpful thoughts.

> *I feel like my anxiety leads into anger sometimes and that [the app] makes you sit there and think about it instead of you being in that moment and being so upset that you can't comprehend what's going on or be logical about your thinking … Just think about it…let's do this little activity.* [P16]

> *It would pose a question …It would give you a statement and ask you, do you believe this to be true … that really stuck with me. If you smoked one cigarette … do you believe that you could still not smoke the rest of the day … I really believe, once I've smoked one, I'm a smoker there's no stopping… I don't really challenge my way of thinking…There is a positive side if you would just open your eyes to it.* [P12]

Others said that the most helpful part of the intervention was having it with them all the time, to use at the moment they needed it.

> *Remove yourself from the situation… if you can just go to the app in that moment and have that put into your head, that's what's helpful … because it helps in that moment.* [P8]

> *When I was feeling stressed I pushed the button ... just stressed out like where you think you're gonna light up a cigarette and smoke … you push that button.* [P3]

XSL•FO
RenderX

Reactions to Cultural Tailoring

Participants were asked about the cultural tailoring of the app to reflect Black culture. Almost all participants said cultural tailoring contributed to their engagement with the app.

> *I really like this app being geared toward the Black people, I mean because it is a different thing. The high blood pressure and stress and the discrimination is a totally different thing than what White people face on a daily basis.* [P4]

Participants were then asked about the relevance of the tailored intervention components to smokers who are Black.

> *explaining about the history of cigarettes within the African American community, the effect that it had within the community … I feel like it's very relevant. There's a lot of facts in there that you wouldn't even think about as a Black person. There's a lot of things that were directed towards Black people…they portray it in media. I always thought that it was portrayed more for white people … for a Black person that actually smokes, I think it would be eye opening, it would be very, very helpful for them.* [P16]

> *A lot of the tips were directly towards African Americans.* [P17]

Participant recommendations to increase the relevance of the MASP app for Black people who want to quit smoking included hearing from Black former smokers, either through short videos added to the app, seeing former participant pictures added to the app, or brief quotes or stories from former participants as messages or on-demand content. These pictures and stories could help the participants connect to the project and the app and encourage them during the tough parts of the quitting process.

> *past users of the actual app that are African American that have been successful with it… if they [participants] have pictures of past people that have actually accomplished this, they'll be more inclined to use it…African American[s] really use this app, his picture's right there… it worked for him. It might work for me.* [P21]

Other suggestions to increase relevance for Black people who smoke included more overall diversity in photographs and videos throughout the app, especially more photographs and videos using Black people who smoke and former smokers, as well as Black health care professionals.

> *It's gotta be by Black people for black people to pay attention... I think a lot of African Americans are, if it's not something that looks familiar or sounds familiar, then they're scared to go around it.* [P12]

A few participants suggested additional information on menthol tobacco use that is highly prevalent.

> *I would just emphasize menthols and maybe Black and Milds? On top of cigarettes, you know that that those are just as deadly.* [P18]

Several participants also suggested incorporating Black vernacular in videos and messaging.

> *A more relatable language to the African American community …how we speak like there's certain words and lingo, maybe just incorporating that… it might appeal to more African Americans... they could read it and be like oh they know what I'm talking about.* [P16]

Experience With Daily Surveys and Other Assessments

All of the participants found the surveys easy to complete. Many of the participants thought the daily assessments were a part of the intervention content. Participants described the surveys as holding them more "accountable" and keeping them aware of their smoking behaviors and how much they were thinking about smoking. Several discussed how the survey questions helped them ask for help or remove themselves from a situation where they were more likely to smoke. Most responded quite positively to the idea of the daily surveys, but many participants did not like the amount of repetition of the survey questions and the number of surveys each day. The repetition of the questions over the 6 weeks of the study, which participants acknowledged as necessary to track their progress over time, was often experienced as frustrating or boring in the later weeks of the study.

> *I knew what questions were coming up and I didn't even have to read the questions anymore. I was just answering them.* [P20]

> *the amount of times you guys asked me a day if I'm depressed …it was a bit excessive. You're going to get the same answer 2 hours or three hours from now … so you might as well just ask me in the middle of the day … it was exhausting.* [P7]

Participants provided several suggestions to hold their interest in the surveys and help them complete more of the surveys throughout the study.

> *a different set of completely different set of questions weekly or change them every day.* [P4]

> *have more of a variety of questions…maybe add something, maybe switch them up a little bit.* [P18]

Participants expressed challenges with what they perceived as a short response window to start the surveys following notification, particularly the surveys that happened during the workday. Participants gave examples of missing surveys because they were driving, with customers, at work, or with family where they could not immediately stop to take the survey or press a button to indicate they would like to take the survey later. Participants also asked for more control over when to take the surveys during the day.

> *Give us more than five minutes to pull our cars over … there were quite a few times I'm on the highway or I'm on the other line with a doctor's office or kids school and I need more than five minutes.* [P7]

Suggested Changes to Improve the Mobile Anxiety Sensitivity Program for Smoking App

The most commonly suggested improvement was a feature to allow interaction with other participants. Other app improvements included more self-monitoring functions, more

XSL•FO
RenderX

pictures and color, and options for longer versions of the training, such as meditation and coping skills.

Hearing from other smokers trying to quit: Participants suggested connecting with other app users to chat, establish a buddy system within the app, or just talk through their challenges and hear from others. They suggested these contacts could take place within the app, be voluntary, and be anonymous to protect participant identities.

> *set up a buddy system when the person signs up for the study … get permission consent from both parties, if you need to talk to somebody live, this is your buddy.* [P10]

> *Maybe like a chat line…go into an open forum. Say we speak to 'em, just say whatever to chit chat, talk… moral support.* [P21]

Self-monitoring: Several participants asked if there was a way they could look at their survey data to see their trends in how many cigarettes they were smoking, how they were feeling at the time, and what seemed to trigger their smoking. Other participants suggested an optional journaling function where they could write about their experiences and review their entries later.

> *The option …to really track your smoking … more on an individual basis so you can actually see …how you were smoking or not or how long you were going without smoking, … especially your progress because I think that's the most motivating to see that you're actually doing it.* [P8]

Discussion

Principal Results

Black people are disproportionally affected by tobacco use and tobacco-related health conditions but historically have been underrepresented in tobacco cessation research [31]. The MASP smoking cessation app was shown to be acceptable and culturally fit with Black adults, and study participants provided recommendations to increase MASP's relevance to Black people who want to quit smoking.

Given the underrepresentation of Black adults in intervention research, generally, understanding what influences potential participants to complete study screening surveys or to contact study staff members is helpful in increasing participation in clinical trials among Black communities. Participants indicated that it is important that recruitment materials and app content be designed to establish study legitimacy. This includes the professional appearance of recruitment materials and linking the study with a specific university. One recommendation was to include testimonials from previous study participants, which is consistent with other mHealth studies [32,38,41].

Throughout the interviews, participants discussed trust and credibility regarding the intervention and its content. Similar to other studies focused on Black adults, the credibility of the information in the app was often assessed by how well the information in the app matched their own experiences and the source of the information [54,55]. This may be especially important for cultural groups such as Black adults who may experience suspicions or distrust of the medical community due to centuries of historical oppression and systemic racism [56]. To increase credibility and relevance to Black people who smoke, participants recommended including more images reflective of Black adults but also suggested increasing the diversity of the photos and videos beyond Black people to include more photographs and videos of Black health care professionals. However, participants in another study cautioned researchers not to use stock photos or stereotypical images, but to use images and photographs reflective of the app users [32,38].

MASP participants also suggested incorporating more language and statistics on Black people who smoke and more discussion of disparities and targeting by the tobacco industry into the intervention messages to increase engagement and acceptability of the app [32,41]. One novel strategy to increase the number of culturally relevant messages in other mHealth studies involved recruiting a group of Black people who smoke and those who have successfully quit smoking to generate motivational messages for other Black people trying to quit [57-59]. These messages could also be used to incorporate more culturally relevant content and Black vernacular into the app [59]. However, it should be noted that a small number of participants (2 out of 21 in the current study) did not react positively to the explicit cultural tailoring. Diversity of opinions and cultural identity would be expected in any cultural group, so testing messaging and educational content within the app and assessing a participant's degree of cultural identification at baseline would allow researchers to match culturally relevant messages by level of cultural identification so that participants could receive appropriately tailored messages for their level of cultural identification.

When asked what additional features they would add to the app, participants wanted a way to interact with other participants who were using the app to try to stop smoking. This was interesting, given participants' comments that they liked using the app because it allowed them privacy and flexibility. Interaction with other study participants within the app may increase cultural relevance as participants would interact with others from their cultural group who were embarking on a shared lived experience of smoking cessation. Other mHealth researchers have suggested a discussion board or group chats moderated by study staff members so that misinformation and unhealthy conversations could not be introduced [27,38]. A discussion board could be an easy option to allow participants to interact with each other [32,41,55,60]. These interactions could also be collected for additional qualitative data analysis. Researchers might want to consider a different discussion board for the control and intervention participants or only for intervention participants. This type of feature has not been well used in previous smoking cessation apps [61,62].

MASP participants also asked for tools that could summarize their progress within the app (eg, cigarette and other tobacco product tracking, money saved, and tracking triggers for smoking). This was also suggested by participants in other mHealth substance use studies [38,41,54,55,63]. Self-tracking or self-monitoring functions could help participants increase

XSL•FO
RenderX

their awareness of substance use and personal feelings of accountability for their behavior and develop plans to cope with personal smoking triggers [29]. Both personal accountability and increased awareness of behaviors were named by participants in this and other substance use mHealth studies as something they liked best about the study [54].

MASP participants also recommended prompting fewer surveys per day and adding or substituting questions within the surveys to increase variety and engagement throughout the study. One concern expressed by MASP participants was to make the surveys equally accessible to those who do not have as much control over their daily work schedule, as they cannot always stop what they are doing to complete surveys. This could apply to those who have changing shifts and scheduled work hours, those who drive a lot, are in customer service, and cannot stop serving customers to take surveys immediately, or those who are in hourly positions with less flexibility when they can take short breaks. This was also noted as a challenge in other mHealth interventions [30]. Similarly, MASP participants suggested an option to "try later" for surveys sent when participants could not or did not want to complete them; however, Bendotti et al [55] cautioned app developers to balance additional time to complete surveys with the number of daily notifications from the app as many users did not like frequent notifications. Rather than delaying surveys and messages, another option MASP participants suggested was to allow participants to keep the messages and surveys in a "bank" that they could come back to later that day. However, the MASP app prompted participants with 5 assessments and tailored messages throughout the day, so delaying them until the end of the day would mean participants would not be recording their experiences at the most stressful and potentially most meaningful moments during the day when tailored messages could be most needed to prevent smoking lapse.

Limitations

Limitations of the study's qualitative data analysis included the relatively small number of interviews, which limited generalizability to the larger population of Black people who smoke cigarettes. An additional limitation is that the 2 interviewers were not Black. This may have affected participant feedback, both constructive and positive. However, interviewers of the same cultural group as the participants can take unconscious "cultural shortcuts" in communication with participants or make assumptions about participant responses and move too quickly through participant responses without sufficient probing of responses. One suggestion is to have a diverse team of qualitative interviewers and analysts, which may lead to richer data collection and analysis.

Conclusions

In conclusion, MASP is a feasible and acceptable culturally tailored smoking cessation app for anxiety-sensitive Black adults who want to quit smoking. The development of evidence-based, culturally tailored mHealth apps such as MASP could help to reduce health disparities by improving the cultural relevance of mHealth interventions. Tailored mHealth interventions could decrease barriers to accessing smoking cessation interventions and could increase engagement and acceptability in population subgroups who experience higher burdens of health-harming behaviors and diseases [25,38].

Acknowledgments

AA, LG, MG, EH, AV, KK, CM, MZ, and MSB contributed to the intervention design and implementation; LG and CM collected the qualitative data and MKC and AA were the primary qualitative analysts with LG and CM. All authors contributed to the preparation of the manuscript.

This study was supported by funding from the National Institute on Minority Health and Health Disparities (U54MD015946: Principal Investigator Obasi; current study Multiple Principal Investigators (MPIs) Businelle and Zvolensky), National Cancer Institute (NCI) grant P30CA225520 awarded to the Stephenson Cancer Center, the Tobacco Settlement Endowment Trust (MSB), and the work used the Stephenson Cancer Center's mHealth Shared Resource. The manuscript preparation was additionally supported by grants to AA from the National Institute on Minority Health and Health Disparities (K01MD015295) and ETH from the National Institute on Drug Abuse (K99/R00DA046564). Work by LG on the manuscript was supported by the National Institute on Drug Abuse of the National Institutes of Health under Award Number R25DA054015 (MPIs: Obasi and Reitzel).

Conflicts of Interest

MSB is one of the inventors of the Insight mobile health platform and receives royalties related to the use of this platform by investigators external to the University of Oklahoma Health Sciences Center. As MSB is a multiple principal investigator in this study, he did not receive royalties for using the platform to create the Mobile Anxiety Sensitivity Program for Smoking app.

Multimedia Appendix 1
Examples of MASP (Mobile Anxiety Sensitivity Program) app menus.
[PNG File , 195 KB - formative_v9i1e53566_app1.png]

Multimedia Appendix 2
Sample MASP (Mobile Anxiety Sensitivity Program) video clip.
[MP4 File (MP4 Video), 2190 KB - formative_v9i1e53566_app2.mp4]

XSL•FO
RenderX

References

1. Martell BN, Garrett BE, Caraballo RS. Disparities in adult cigarette smoking-United States, 2002-2005 and 2010-2013. MMWR Morb Mortal Wkly Rep 2016;65(30):753-758 [FREE Full text] [doi: 10.15585/mmwr.mm6530a1] [Medline: 27491017]
2. Haiman CA, Stram DO, Wilkens LR, Pike MC, Kolonel LN, Henderson BE, et al. Ethnic and racial differences in the smoking-related risk of lung cancer. N Engl J Med 2006;354(4):333-342. [doi: 10.1056/NEJMoa033250] [Medline: 16436765]
3. Howlader N, Mariotto AB, Woloshin S, Schwartz LM. Providing clinicians and patients with actual prognosis: cancer in the context of competing causes of death. J Natl Cancer Inst Monogr 2014;2014(49):255-264 [FREE Full text] [doi: 10.1093/jncimonographs/lgu022] [Medline: 25417239]
4. Novotny TE, Warner KE, Kendrick JS, Remington PL. Smoking by blacks and whites: socioeconomic and demographic differences. Am J Public Health 1988;78(9):1187-1189. [doi: 10.2105/ajph.78.9.1187] [Medline: 3407817]
5. Caraballo RS, Giovino GA, Pechacek TF, Mowery PD, Richter PA, Strauss WJ, et al. Racial and ethnic differences in serum cotinine levels of cigarette smokers: third national health and nutrition examination survey, 1988-1991. JAMA 1998;280(2):135-139. [doi: 10.1001/jama.280.2.135] [Medline: 9669785]
6. Schoenborn CA, Adams PF, Peregoy JA. Health behaviors of adults: United States, 2008-2010. Vital Health Stat 10 2013(257):1-184 [FREE Full text] [Medline: 25116426]
7. Webb Hooper M, Baker EA, McNutt MD. Associations between coping, affect, and social support among low-income African American smokers. Addict Behav 2013;38(11):2736-2740. [doi: 10.1016/j.addbeh.2013.07.005] [Medline: 23934005]
8. Babb S, Malarcher A, Schauer G, Asman K, Jamal A. Quitting smoking among adults - United States, 2000-2015. MMWR Morb Mortal Wkly Rep 2017;65(52):1457-1464 [FREE Full text] [doi: 10.15585/mmwr.mm6552a1] [Medline: 28056007]
9. Bacio GA, Guzman IY, Shapiro JR, Ray LA. Differences in quit attempts between non-hispanic black and white daily smokers: the role of smoking motives. Addict Behav 2014;39(12):1769-1772 [FREE Full text] [doi: 10.1016/j.addbeh.2014.07.001] [Medline: 25123344]
10. Lee JGL, Henriksen L, Rose SW, Moreland-Russell S, Ribisl KM. A systematic review of neighborhood disparities in point-of-sale tobacco marketing. Am J Public Health 2015;105(9):e8-e18 [FREE Full text] [doi: 10.2105/AJPH.2015.302777] [Medline: 26180986]
11. Wheeler DC, Boyle J, Barsell DJ, Glasgow T, McClernon FJ, Oliver JA, et al. Associations of alcohol and tobacco retail outlet rates with neighborhood disadvantage. Int J Environ Res Public Health 2022;19(3):1134 [FREE Full text] [doi: 10.3390/ijerph19031134] [Medline: 35162162]
12. Cwalina SN, Ihenacho U, Barker J, Smiley SL, Pentz MA, Wipfli H. Advancing racial equity and social justice for black communities in US tobacco control policy. Tob Control 2023;32(3):381-384 [FREE Full text] [doi: 10.1136/tobaccocontrol-2021-056704] [Medline: 34526408]
13. Mattingly DT, Mezuk B, Elliott MR, Fleischer NL. Discrimination and tobacco use outcomes among US adults: effect modification by race/ethnicity. J Racial Ethn Health Disparities 2024;11(1):395-405 [FREE Full text] [doi: 10.1007/s40615-023-01527-3] [Medline: 36696079]
14. Leventhal AM, Zvolensky MJ. Anxiety, depression, and cigarette smoking: a transdiagnostic vulnerability framework to understanding emotion-smoking comorbidity. Psychol Bull 2015;141(1):176-212 [FREE Full text] [doi: 10.1037/bul0000003] [Medline: 25365764]
15. Langdon KJ, Farris SG, Hogan JB, Grover KW, Zvolensky MJ. Anxiety sensitivity in relation to quit day dropout among adult daily smokers recruited to participate in a self-guided cessation attempt. Addict Behav 2016;58:12-15 [FREE Full text] [doi: 10.1016/j.addbeh.2016.02.017] [Medline: 26896560]
16. Johnson KA, Stewart S, Rosenfield D, Steeves D, Zvolensky MJ. Prospective evaluation of the effects of anxiety sensitivity and state anxiety in predicting acute nicotine withdrawal symptoms during smoking cessation. Psychol Addict Behav 2012;26(2):289-297 [FREE Full text] [doi: 10.1037/a0024133] [Medline: 21644805]
17. Zvolensky MJ, Farris SG, Schmidt NB, Smits JAJ. The role of smoking inflexibility/avoidance in the relation between anxiety sensitivity and tobacco use and beliefs among treatment-seeking smokers. Exp Clin Psychopharmacol 2014;22(3):229-237 [FREE Full text] [doi: 10.1037/a0035306] [Medline: 24490706]
18. Johnson KA, Farris SG, Schmidt NB, Zvolensky MJ. Anxiety sensitivity and cognitive-based smoking processes: testing the mediating role of emotion dysregulation among treatment-seeking daily smokers. J Addict Dis 2012;31(2):143-157 [FREE Full text] [doi: 10.1080/10550887.2012.665695] [Medline: 22540436]
19. Assayag Y, Bernstein A, Zvolensky MJ, Steeves D, Stewart SS. Nature and role of change in anxiety sensitivity during NRT-aided cognitive-behavioral smoking cessation treatment. Cogn Behav Ther 2012;41(1):51-62. [doi: 10.1080/16506073.2011.632437] [Medline: 22375732]
20. Allan NP, Raines AM, Capron DW, Norr AM, Zvolensky MJ, Schmidt NB. Identification of anxiety sensitivity classes and clinical cut-scores in a sample of adult smokers: results from a factor mixture model. J Anxiety Disord 2014;28(7):696-703 [FREE Full text] [doi: 10.1016/j.janxdis.2014.07.006] [Medline: 25128664]

XSL•FO
RenderX

21. Farris SG, DiBello AM, Allan NP, Hogan J, Schmidt NB, Zvolensky MJ. Evaluation of the anxiety sensitivity index-3 among treatment-seeking smokers. Psychol Assess 2015;27(3):1123-1128 [FREE Full text] [doi: 10.1037/pas0000112] [Medline: 25894700]
22. Bello MS, Pang RD, Cropsey KL, Zvolensky MJ, Reitzel LR, Huh J, et al. Tobacco withdrawal amongst African American, hispanic, and white smokers. Nicotine Tob Res 2016;18(6):1479-1487 [FREE Full text] [doi: 10.1093/ntr/ntv231] [Medline: 26482061]
23. Reitzel LR, Childress SD, Obasi EM, Garey L, Vidrine DJ, McNeill LH, et al. Interactive effects of anxiety sensitivity and subjective social status on psychological symptomatology in black adults. Behav Med 2017;43(4):268-276 [FREE Full text] [doi: 10.1080/08964289.2016.1150805] [Medline: 26963636]
24. Smith AL, Carter SM, Chapman S, Dunlop SM, Freeman B. Why do smokers try to quit without medication or counselling? A qualitative study with ex-smokers. BMJ Open 2015;5(4):e007301 [FREE Full text] [doi: 10.1136/bmjopen-2014-007301] [Medline: 25933811]
25. Grundy Q. A review of the quality and impact of mobile health apps. Annu Rev Public Health 2022;43:117-134 [FREE Full text] [doi: 10.1146/annurev-publhealth-052020-103738] [Medline: 34910582]
26. Carreiro S, Newcomb M, Leach R, Ostrowski S, Boudreaux ED, Amante D. Current reporting of usability and impact of mHealth interventions for substance use disorder: a systematic review. Drug Alcohol Depend 2020;215:108201 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108201] [Medline: 32777691]
27. Monney G, Penzenstadler L, Dupraz O, Etter JF, Khazaal Y. mHealth app for cannabis users: satisfaction and perceived usefulness. Front Psychiatry 2015;6:120 [FREE Full text] [doi: 10.3389/fpsyt.2015.00120] [Medline: 26379561]
28. Nierkens V, Hartman MA, Nicolaou M, Vissenberg C, Beune EJAJ, Hosper K, et al. Effectiveness of cultural adaptations of interventions aimed at smoking cessation, diet, and/or physical activity in ethnic minorities. a systematic review. PLoS One 2013;8(10):e73373 [FREE Full text] [doi: 10.1371/journal.pone.0073373] [Medline: 24116000]
29. Herbst E, McCaslin SE, Hassanbeigi Daryani S, Laird KT, Hopkins LB, Pennington D, et al. A qualitative examination of stay quit coach, a mobile application for veteran smokers with posttraumatic stress disorder. Nicotine Tob Res 2020;22(4):560-569. [doi: 10.1093/ntr/ntz037] [Medline: 30874289]
30. Hébert ET, Ra CK, Alexander AC, Helt A, Moisiuc R, Kendzor DE, et al. A mobile just-in-time adaptive intervention for smoking cessation: pilot randomized controlled trial. J Med Internet Res 2020;22(3):e16907 [FREE Full text] [doi: 10.2196/16907] [Medline: 32149716]
31. James DC, Harville C, Sears C, Efunbumi O, Bondoc I. Participation of African Americans in e-Health and m-Health studies: a systematic review. Telemed J E Health 2017;23(5):351-364. [doi: 10.1089/tmj.2016.0067] [Medline: 27792475]
32. Enyioha C, Loufman LM, Grewe ME, Cené CW, Khairat S, Goldstein AO, et al. Black smokers' preferences for features of a smoking cessation app: qualitative study. JMIR Form Res 2023;7:e43603 [FREE Full text] [doi: 10.2196/43603] [Medline: 37252777]
33. Ray R, Sewell AA, Gilbert KL, Roberts JD. Missed opportunity? Leveraging mobile technology to reduce racial health disparities. J Health Polit Policy Law 2017;42(5):901-924. [doi: 10.1215/03616878-3940477] [Medline: 28663182]
34. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
35. Institute of Medicine. Speaking of Health: Assessing Health Communication Strategies for Diverse Populations. Washington, DC: The National Academies Press; 2002.
36. Businelle MS, Ma P, Kendzor DE, Frank SG, Vidrine DJ, Wetter DW. An ecological momentary intervention for smoking cessation: evaluation of feasibility and effectiveness. J Med Internet Res 2016;18(12):e321 [FREE Full text] [doi: 10.2196/jmir.6058] [Medline: 27956375]
37. Barrera M, Castro FG. A heuristic framework for the cultural adaptation of interventions. Clinical Psychology: Science and Practice 2006;13(4):311-316. [doi: 10.1111/j.1468-2850.2006.00043.x]
38. Trinidad SB, Shane AL, Guinn TR, Apok CR, Collier AF, Avey JP, et al. The cultural adaption of a sobriety support app for Alaskan Native and American Indian: qualitative feasibility and acceptability study. JMIR Form Res 2023;7:e38894 [FREE Full text] [doi: 10.2196/38894] [Medline: 36473107]
39. Webb MS. Does one size fit all African American smokers? The moderating role of acculturation in culturally specific interventions. Psychol Addict Behav 2008;22(4):592-596. [doi: 10.1037/a0012968] [Medline: 19071987]
40. Wall SP, Castillo P, Shuchat-Shaw F, Norman E, Brown D, Martinez-López N, et al. Targeting versus tailoring educational videos for encouraging deceased organ donor registration in black-owned barbershops. J Health Commun 2022;27(1):37-48 [FREE Full text] [doi: 10.1080/10810730.2022.2035021] [Medline: 35170401]
41. Brewer LC, Kumbamu A, Smith C, Jenkins S, Jones C, Hayes SN, et al. A cardiovascular health and wellness mobile health intervention among church-going African Americans: formative evaluation of the FAITH! app. JMIR Form Res 2020;4(11):e21450 [FREE Full text] [doi: 10.2196/21450] [Medline: 33200999]
42. Staffileno BA, Tangney CC, Fogg L. Favorable outcomes using an eHealth approach to promote physical activity and nutrition among young African American women. J Cardiovasc Nurs 2018;33(1):62-71. [doi: 10.1097/JCN.0000000000000409] [Medline: 28481823]

XSL•FO
RenderX

43. Matthews AK, Sánchez-Johnsen L, King A. Development of a culturally targeted smoking cessation intervention for African American smokers. J Community Health 2009;34(6):480-492 [FREE Full text] [doi: 10.1007/s10900-009-9181-5] [Medline: 19728056]
44. Resnicow K, Baranowski T, Ahluwalia JS, Braithwaite RL. Cultural sensitivity in public health: defined and demystified. Ethn Dis 1999;9(1):10-21. [Medline: 10355471]
45. Walters ST, Mun EY, Tan Z, Luningham JM, Hébert ET, Oliver JA, et al. Development and preliminary effectiveness of a smartphone-based, just-in-time adaptive intervention for adults with alcohol misuse who are experiencing homelessness. Alcohol Clin Exp Res 2022;46(9):1732-1741 [FREE Full text] [doi: 10.1111/acer.14908] [Medline: 35869820]
46. Businelle MS, Garey L, Gallagher MW, Hébert ET, Vujanovic A, Alexander A, et al. An integrated mHealth app for smoking cessation in black smokers with anxiety: protocol for a randomized controlled trial. JMIR Res Protoc 2022;11(5):e38905 [FREE Full text] [doi: 10.2196/38905] [Medline: 35635746]
47. Garey L, Hébert ET, Mayorga NA, Chavez J, Shepherd JM, Businelle MS, et al. Evaluating the feasibility and acceptability of a mobile-based health technology for smoking cessation: mobile anxiety sensitivity program. Br J Clin Psychol 2022;61 Suppl 1(Suppl 1):111-129 [FREE Full text] [doi: 10.1111/bjc.12294] [Medline: 33939190]
48. Casey M, Hayes PS, Glynn F, OLaighin G, Heaney D, Murphy AW, et al. Patients' experiences of using a smartphone application to increase physical activity: the SMART MOVE qualitative study in primary care. Br J Gen Pract 2014;64(625):e500-e508 [FREE Full text] [doi: 10.3399/bjgp14X680989] [Medline: 25071063]
49. Patton MQ. Utilization-Focused Evaluation. Thousand Oaks: Sage Publications, Inc; 2008.
50. Saldaña J. The Coding Manual for Qualitative Researchers. Thousand Oaks: Sage Publications, Inc; 2013.
51. Patton MQ. Qualitative Research and Evaluation Methods: Integrating Theory and Practice. Thousand Oaks: Sage Publications, Inc; 2014.
52. Miles MB, Huberman AM, Saldaña J. Qualitative Data Analysis. Thousand Oaks: SAGE Publications, Inc; 2019.
53. Snowden LR, Hines AM. A scale to assess African American acculturation. Journal of Black Psychology 1999;25(1):36-47. [doi: 10.1177/0095798499025001003]
54. Puddephatt JA, Leightley D, Palmer L, Jones N, Mahmoodi T, Drummond C, et al. A qualitative evaluation of the acceptability of a tailored smartphone alcohol intervention for a military population: information about drinking for ex-serving personnel (InDEx) app. JMIR Mhealth Uhealth 2019;7(5):e12267 [FREE Full text] [doi: 10.2196/12267] [Medline: 31127726]
55. Bendotti H, Lawler S, Ireland D, Gartner C, Hides L, Marshall HM. What do people want in a smoking cessation app? An analysis of user reviews and app quality. Nicotine Tob Res 2022;24(2):169-177. [doi: 10.1093/ntr/ntab174] [Medline: 34460922]
56. Brewer LC, Fortuna KL, Jones C, Walker R, Hayes SN, Patten CA, et al. Back to the future: achieving health equity through health informatics and digital health. JMIR Mhealth Uhealth 2020;8(1):e14512 [FREE Full text] [doi: 10.2196/14512] [Medline: 31934874]
57. Jackson DN, Sehgal N, Baur C. Benefits of mHealth co-design for African American and hispanic adults: multi-method participatory research for a health information app. JMIR Form Res 2022;6(3):e26764 [FREE Full text] [doi: 10.2196/26764] [Medline: 35262496]
58. Nagawa CS, Faro JM, Menon AJ, Ito Fukunaga M, Williams JH, Mourao D, et al. Written advice given by African American smokers to their peers: qualitative study of motivational messages. JMIR Form Res 2021;5(4):e21481 [FREE Full text] [doi: 10.2196/21481] [Medline: 33929332]
59. Sideropoulos V, Vangeli E, Naughton F, Cox S, Frings D, Notley C, et al. Mobile phone text messages to support people to stop smoking by switching to vaping: codevelopment, coproduction, and initial testing study. JMIR Form Res 2023;7:e49668 [FREE Full text] [doi: 10.2196/49668] [Medline: 37756034]
60. Struik L, Christianson K, Khan S, Sharma RH. Strengths and limitations of web-based cessation support for individuals who smoke, dual use, or vape: qualitative interview study. JMIR Form Res 2023;7:e43096 [FREE Full text] [doi: 10.2196/43096] [Medline: 38064266]
61. Barroso-Hurtado M, Suárez-Castro D, Martínez-Vispo C, Becoña E, López-Durán A. Smoking cessation apps: a systematic review of format, outcomes, and features. Int J Environ Res Public Health 2021;18(21):11664 [FREE Full text] [doi: 10.3390/ijerph182111664] [Medline: 34770178]
62. McClure JB, Heffner J, Hohl S, Klasnja P, Catz SL. Design considerations for mHealth programs targeting smokers not yet ready to quit: results of a sequential mixed-methods study. JMIR Mhealth Uhealth 2017;5(3):e31 [FREE Full text] [doi: 10.2196/mhealth.6845] [Medline: 28283465]
63. Attwood S, Parke H, Larsen J, Morton KL. Using a mobile health application to reduce alcohol consumption: a mixed-methods evaluation of the drinkaware track & calculate units application. BMC Public Health 2017;17(1):394 [FREE Full text] [doi: 10.1186/s12889-017-4358-9] [Medline: 28511698]

Abbreviations

MASP: Mobile Anxiety Sensitivity Program
mHealth: mobile health

XSL•FO
RenderX

NRT: Nicotine Replacement Therapy

Edited by A Mavragani; submitted 17.10.23; peer-reviewed by H Tateno; comments to author 23.03.24; accepted 24.04.24; published 07.02.25.

Please cite as:
Cheney MK, Alexander AC, Garey L, Gallagher MW, Hébert ET, Vujanovic AA, Kezbers KM, Matoska CT, Zvolensky MJ, Businelle MS
Adapting a Mobile Health App for Smoking Cessation in Black Adults With Anxiety Through an Analysis of the Mobile Anxiety Sensitivity Program Proof-of-Concept Trial: Qualitative Study
JMIR Form Res 2025;9:e53566
URL: https://formative.jmir.org/2025/1/e53566
doi:10.2196/53566
PMID:39918847

©Marshall K Cheney, Adam C Alexander, Lorra Garey, Matthew W Gallagher, Emily T Hébert, Anka A Vujanovic, Krista M Kezbers, Cameron T Matoska, Michael J Zvolensky, Michael S Businelle. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Improving the Utility, Safety, and Ethical Use of a Passive Mood-Tracking App for People With Bipolar Disorder Using Coproduction: Qualitative Focus Group Study

Laurence Astill Wright[1,2], MBBCH; Matthew Moore[1,3]; Stuart Reeves[4], PhD; Elvira Perez Vallejos[1,3], PhD; Richard Morriss[1,3,5,6], PhD

[1]Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom
[2]Centre for Academic Mental Health, Population Health Sciences, University of Bristol, Bristol, United Kingdom
[3]NIHR MindTech Medical Technology Collaborative, University of Nottingham, Nottingham, United Kingdom
[4]School of Computer Science, University of Nottingham, Nottingham, United Kingdom
[5]NIHR ARC East Midlands, University of Nottingham, Nottingham, United Kingdom
[6]Nottingham NIHR Biomedical Research Centre, University of Nottingham, Nottingham, United Kingdom

Corresponding Author:
Laurence Astill Wright, MBBCH
Institute of Mental Health
University of Nottingham
Jubilee Campus
Triumph Road
Nottingham, NG7 2TU
United Kingdom
Phone: 44 115 823 1294
Email: laurence.astillwright@nottingham.ac.uk

Abstract

Background: Coproduction with users of new digital technology, such as passive mood monitoring, is likely to improve its utility, safety, and successful implementation via improved design and consideration of how such technology fits with their daily lives. Mood-monitoring interventions are commonly used by people with bipolar disorder (BD) and have promising potential for digitization using novel technological methods.

Objective: This study aims to explore how a passive behavioral monitoring platform, Remote Assessment of Disease and Relapse, would meet the needs of people with BD by specifically considering purpose and function, diversity of need, personal preference, essential components and potential risks, and harms and mitigation strategies through an iterative coproduction process.

Methods: A total of 17 people with BD were recruited via national charities. We conducted 3 web-based focus groups as a part of an iterative coproduction process in line with responsible research and innovation principles and with consideration of clinical challenges associated with BD. Data were analyzed thematically. Results were cross-checked by someone with lived experience of BD.

Results: Focus groups were transcribed and analyzed using thematic analysis. Six themes were identified as follows: (1) the purpose of using the app, (2) desired features, (3) when to use the app, (4) risks of using the app, (5) sharing with family and friends, and (6) sharing with health care professionals.

Conclusions: People with BD who are interested in using passive technology to monitor their mood wish to do so for a wide variety of purposes, identifying several preferences and potential risks. Principally, people with BD wished to use this novel technology to aid them in self-managing their BD with greater insight and a better understanding of potential triggers. We discuss key features that may aid this functionality and purpose, including crisis plans and sharing with others. Future development of passive mood-monitoring technologies should not assume that the involvement of formal mental health services is desired.

(JMIR Form Res 2025;9:e65140) doi:10.2196/65140

XSL•FO
RenderX

KEYWORDS

mood monitoring; ecological momentary assessment; EMA; passive ecological momentary assessment; passive EMA; bipolar disorder; implementation; qualitative; mobile phone

Introduction

Background

Recent developments in technology have increased the possibility of determining the mental state of an individual via tracking behavioral data collected via smartphones [1]. This promises novel interventions based on passive ecological momentary assessment—a technology that can potentially frequently assess mood via data collected in the background of one's smartphone use [2]. While this hypothesis remains controversial, many research teams are trying to implement this new technology into usable and acceptable interventions [3]. One open-access research platform using passive monitoring technology to glean new insights into mental disorders is the Remote Assessment of Disease and Relapse (RADAR) [4]. Here, we explore the design and potential implementation of this technology, using the RADAR platform, in the form of a mood-tracking app for people with bipolar disorder (BD), considering purpose, function, components, and risks and harms in an iterative coproduction process centrally involving end users of the technology—people with BD interested in mood monitoring. However, clinical experts and experts in the design of digital technology involving human-computer interaction and responsible innovation were also involved in such coproduction, as we have previously proposed [5]. There is a tension between 2 models of technology explored here—one where mental state is predicted via smartphone use and another where technology enters our lives via coproduction. This paper builds on these conceptual ideas.

BD is a condition where individuals already apply various mood-tracking methods to try to gain insights into the disorder and prevent relapse [5,6]. There is some evidence suggesting that increasing awareness of mood fluctuations can improve insights into and the identification of early warning signs and can be useful in preventing relapse in BD [7]. It also suggests that novel technology increases the potential to harness this new technology to create more sophisticated and structured mood-tracking interventions for people with BD. Most of these existing techniques and methods of tracking mood require considerable effort from participants in logging their mood every day for many months [8,9] and may be subject to inaccurate information due to mood biases [10]. Many people with BD use approaches such as paper- and pen-based tracking procedures, mental and physical notation, conversations with their partner, or already existing active smartphone apps [5]. Passive data collection approaches (which are then interpreted actively by the user) may improve flexibility, usability, and acceptability and decrease the time requirement for people who monitor their moods. In a previous qualitative work, a passive mood-tracking app was perceived as potentially positive and helpful by people with BD [5], solving many of the issues with existing methods.

When applying novel technologies in potentially clinical and vulnerable populations, it is important to consider potential risks and harms [11,12]. Technical advances offer potential solutions to complex problems with lighter regulation than many existing approaches to mental health, for example, medication. There is an urgent need to embed responsible research and innovation (RRI) frameworks to maximize societal benefit and be able to anticipate and reflect on the medium- or long-term impact of technology on stakeholders as well as the environment [13]. For example, by engaging with users, developers can identify design properties that could minimize the potential for harm in a number of ways (eg, accessible interfaces, integration with conventional mental health services, and enhanced data privacy and security [14]) and also reduce other risks, for example, digital exclusion that might potentially increase inequalities [14-16].

RRI [11] provide an ethical approach that requires planning, anticipation, and co-design through technological development with end users and all those involved in developing and later using the technology. Along with mitigating harm and inequality, such a co-design process carried out in the framework helps to improve its acceptability and usability for all potential users and relevance to practice and provides some assurance of the safety of the product. For these reasons, a co-design process may lead to a faster spread of usable interventions and a quicker abandonment of less satisfactory developments. However, one element that has so far been given less consideration in the work of responsible innovation is the range of problems associated with different types of mental states when using different forms of technology, such as digital mood monitoring, particularly in conditions where change may be dramatically different, such as BD, where people with BD may experience more than one type of mental state, such as mania and depression. For instance, people with mania are likely to show dramatic changes in their social contacts and what they will communicate and share (particularly on social media) compared to when they are depressed or when they are in neither mental state. This paper applies principles from RRI to consider risk anticipation and risk mitigation in people with BD.

RADAR is an open-source platform and resource for self-tracking psychiatric and neurological conditions developed by a consortium of academic and industry organizations. Currently, the platform has been successfully developed for those with multiple sclerosis [17], depression [4], and epilepsy [18]. RADAR currently passively monitors the following smartphone data: phone use, local weather, step count, GPS location, app use and time duration, battery level, Bluetooth devices in the nearby vicinity, actigraphy, heart rate, call or SMS text message logs, and ambient light.

We are aiming to apply the technology to people with BD, to assess how it might be adapted and adjusted to meet the specific needs of this population. We will not only focus on the technology but also on the requirements needed to ensure it

XSL•FO
RenderX

fulfills the aspirations of people with BD, balanced against identifying potential risks. We will use expertise in responsible technology design to consider unintended consequences, potential risks, and how these might be mitigated.

Objectives

We conducted a series of focus groups to explore the opportunities and challenges of the app to ensure the acceptability and efficacy of passive mood-monitoring apps in the future. This work was part of an iterative coproduction process in line with the use of RRI. We aimed to determine opinions on the following: (1) the purpose and function of the app, (2) the essential and desired components of the digital tool and optimal methods of delivery and design frequency, and (3) the potential risks and harms of using the technology and how these might be mitigated

Methods

Overview

Participants were recruited via national and local charities (Bipolar UK and Bipolar LIFT CIC) via their research opportunity network and other snowballing techniques via an existing patient and public involvement (PPI) team based at the Institute of Mental Health and the University of Nottingham.

Previous research included 16 individuals with BD [5,19] in 2 workshops; however, this explored the utility of self-tracking in general before the development of the RADAR-based tool and did not explore the tool in detail. We chose to explore the study aims through 3 focus groups with potential end users of this technology to simulate discussion and exploration of these issues rather than conduct individual interviews. Previously, we recruited participants from a geographical area that was familiar with participation in research on BD and digital mental health. We wanted to check whether the findings would apply to participants who were not from an area where there was such exposure to research so that fresh ideas and challenges might emerge and also to check the transferability of previous findings. We wanted to test the credibility and confirmability by feeding back findings from previous workshops and focus groups. To facilitate this, one of the workshop leaders from our previous research (an individual with over a decade of experience in PPI and who also has experienced BD) took part in the development of the topic guides and write-up of the current project. All these approaches enhanced the trustworthiness of data collection and analysis. We hoped to recruit 5 to 8 individuals per focus group (>8 individuals would be inappropriate via video call). We planned to run as many focus groups as required until there was a saturation of themes emerging from the focus groups, and no new themes were emerging. Initially, we planned 3 focus groups based on our previous experience of carrying out workshops and focus groups.

The inclusion criteria were as follows: self-reported formal diagnosis of bipolar 1 disorder, bipolar 2 disorder, or bipolar spectrum disorder (BD not otherwise specified or cyclothymia) by a UK psychiatrist, owns and regularly uses a smartphone with an interest in self-tracking their mood via a digital tool, fluent in English, able to give informed consent, and aged >16 years. Exclusion criteria were as follows: participants who lacked the capacity to consent (eg, self-reported diagnosis of dementia and primary diagnosis of substance misuse) or were current inpatients in a psychiatric hospital.

The study was conducted between August and December 2023. We tried to ensure that our methods reflected current practices used by these organizations who organize workshops, focus groups, and group events, which was the reason behind our choice of data collection. Our 3 focus groups were held on the web to improve accessibility and to improve diversity recruitment. The COREQ (Consolidated Criteria for Reporting Qualitative Studies) checklist is presented in Multimedia Appendix 1.

Ethical Considerations

The study was subject to ethical review and approved by the faculty of medicine and health sciences at the University of Nottingham (FMHS 174-1222). Participants made first contact with the research team following advertising. Participants were initially provided with a participant information sheet and given an opportunity to discuss this and ask questions via phone call or email. After ≥48 hours, participants were invited to complete the consent form and again provided with an opportunity to ask questions via telephone call or an alternative method. The data in this study have been deidentified where applicable, and participants consented to the publishing of deidentified quotations. Participants were provided with a £30 (US $38) voucher following the completion of a single focus group. The participants had no relationship with the research team before joining the study. The participants knew about the reasons for conducting the research but had limited knowledge about the research team other than their qualifications. While the research team was diverse, including people with lived experience, some members were psychiatrists and therefore might have carried certain assumptions or biases about the delivery of interventions in people with BD in the National Health Service.

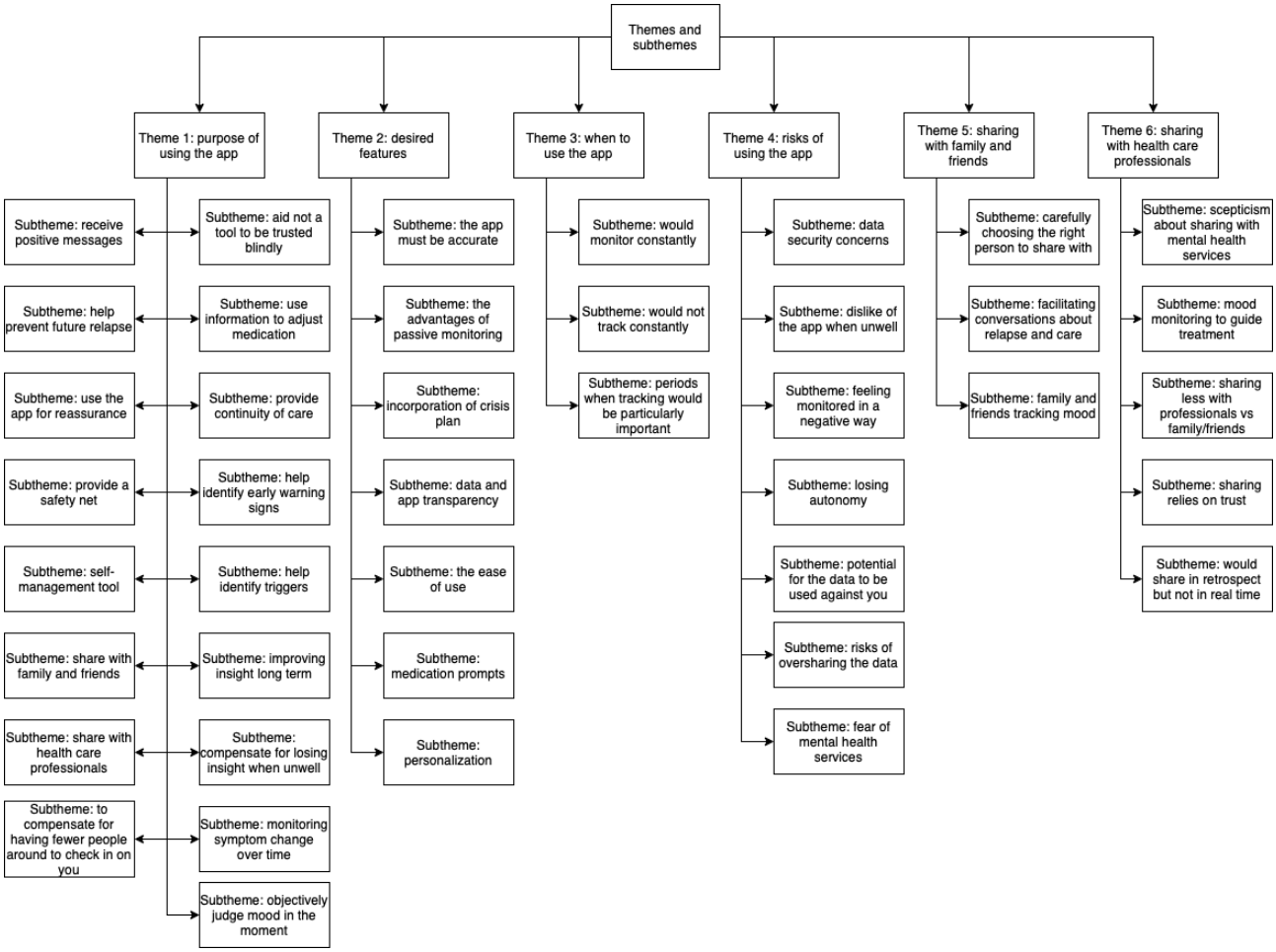
Focus Group Layout

Each focus group lasted 2 hours in total and followed a semistructured topic guide informed by previous research [5] and produced by 2 psychiatrists and 1 person with lived experience of BD (LAW, RM, and MM). Participants were given an explanation of the purpose of the focus groups before participation. The topic guide was used to ask open questions on (1) developing the function of the mood-tracking tool, (2) developing the content of the mood-tracking tool, and (3) how best to translate function and content into a usable and acceptable digital tool (Figure 1). The first 2 focus groups concentrated mostly on what participants wanted from the app and why, reaching saturation in relation to these themes. Therefore, the third focus group only briefly commented on these matters having been presented with these findings. The third focus group concentrated in more detail on how the app was delivered, although some of this content had been explored in the first 2 focus groups. Focus groups were video recorded followed by verbatim transcription. Focus groups were facilitated by LAW (male psychiatrist), who was trained in qualitative interviewing and focus group data collection. At the

XSL•FO
RenderX

time of the focus group, all participants knew LAW from their first contact when considering joining the study.

Figure 1. Overview of the main findings of the 6 themes.



Participants in all focus groups were given a verbal description of the RADAR platform, including the type and amount of data it was possible to collect presented pictorially (Figure S1 in Multimedia Appendix 2). Participants in focus group 1 were presented with the results from the previous round of workshops. The results are summarized in Figures S2 and S3 in Multimedia Appendix 2.

Results

Overview

A total of 17 participants joined 3 focus groups. The demographic characteristics of participants are reported in Table 1. A total of 97 people initially made contact via recruitment methods.

XSL•FO
RenderX

Table 1. Demographics and clinical characteristics of the participants (N=17).

Demographic and clinical characteristics	Participants, n (%)
Age group (y)	
<18	0 (0)
18-24	1 (6)
25-34	4 (24)
35-44	3 (18)
45-55	3 (18)
55-64	4 (24)
>65	0 (0)
Sex	
Male	5 (29)
Female	9 (53)
Nonbinary	1 (6)
Category of bipolar diagnosis	
Bipolar 1 disorder	5 (29)
Bipolar 2 disorder	8 (47)
Cyclothymia	1 (6)
Bipolar disorder not specified	1 (6)
Years of lived experience of bipolar disorder	
<1	0 (0)
1-3	2 (12)
3-5	5 (29)
5-7	1 (6)
>7	7 (41)

Thematic analysis resulted in 6 themes as follows: (1) purpose of using the app, (2) desired features, (3) when to use the app, (4) risks of using the app, (5) sharing with family and friends, and (6) sharing with health care professionals. For each theme, participants described positives and negatives. Tables 2-4 report themes, subthemes, and the participant identifiers.

All the results were cross-checked by someone with lived experience of BD type 1 and over a decade of experience in PPI. Participants did not provide feedback on the findings. Coding via NVivo (Lumivero) [20] was performed by 1 researcher (LAW).

XSL•FO
RenderX

Table 2. Quotes related to theme 1 (purpose of using the app).

Subthemes	Quotes
Subtheme: aid not a tool to be trusted blindly	• "But then I think most people would rather have someone check in and say ohh you know, your app suggests there's been a change and then for them to say Oh no actually you know, I'm fine it's just that this has happened that's happened." [ID15].
Subtheme: understanding triggers	• "What I have used it for is we'll got on. I'm I'm a bit anxious for no particular reason. Am I just being bloody minded and annoyed? But a lot of the time, it's actually. I've tripled my workload. I haven't had sleep." [ID13]
Subtheme: improving insight	• "So actually, I didn't sleep well and this happened or I did sleep well and I was fine and ohh I've been spending a lot of money doing a lot of stuff and then this happened. But also what was happening with my medication and where was I in my cycle? Where was I? What exercise was my level of exercise good at high or low or yeah, so I think it's for your own picture and your own health and everybody's different." [ID11]
Subtheme: objectively judging the mood	• "I think one of the positives about it for people with bipolar is that objectivity and you know the objective data that comes because it also takes a bit of that pressure off the person." [ID12]
Subtheme: prevention	• "Yes, that is fair to say and to catch it quickly, you know, I mean sort of the sooner you can pick up on it and if you've got a monitoring device that can alert you, that's even better." [ID8]. • "Bit hyper, you're a bit, you know, you're talking really fast or you're having loads of different ideas about different things and stuff, and it might be that I missed that until it's fed back to me, reflected back to me so." [ID9]
Subtheme: safety net and reassurance	• "So something in the background. Something in the background sort of saying actually, you're doing X, Y & Z and that's not normal for you." [ID5] • "It's a sense of security in a way. You know, it's almost like having someone there constantly keeping a check on your mood, even though they're not there virtually." [ID8]
Subtheme: self-management	• "If I'm struggling, I'm at 2. What activities could I do to boost it so I'll go for a walk or art activities so I will go back to that book that we made and try the strategies first before I went to the doctor and got a PRN so that would be how I'd monitor it." [ID2]
Subtheme: to compensate for having fewer people around to check in on you	• "I live alone so I don't have anybody to notice. A mood tracker would help with that." [ID5]
Subtheme: use information to adjust medication	• "Yeah, lifestyle changes. Probably tweaking my medication. It could be that because sleep deprivation is a big trigger for me and it could be that I'm lacking in sleep, so the may need to give me some sleep medication." [ID8]

XSL•FO
RenderX

Table 3. Quotes related to theme 2 (desired features), 3 (when to use the app), and 4 (risks of using the app).

Themes and subthemes	Quotes
Theme 2: desired features	
Subtheme: accuracy is important	"Depends how good the app is, whether it does actually perform well 'cause it could be something that is obviously not correct. In a sense, it might not necessarily me being a period of mania, it could be picking up something that is questionable." [ID9]
Subtheme: the advantages of passive monitoring	"And quite often I'm not even that. I'm not that aware of. Maybe what I've put into it 'cause I get up, do it, and then you know and go about my business. So maybe something that's a little bit more engaging and if it's an app that's working in the background more rather than you know, and then I can perhaps turn to that and realize stuff that I've not seen before." [ID9]
Subtheme: crisis plan	"We have a plan in place like that. He's allowed to call the doctors and the mental health team and the GP that needs to be called." [ID17]
Subtheme: data and app transparency	"I think I would like to see all the individual pieces of data. I feel like if that app is collecting those data points, I kind of feel like I'm entitled to see them." [ID10]
Subtheme: the ease of use	"If it's really like clunky and not that accessible or it doesn't feel very friendly. It's gotta be something that fits, doesn't it? The individual." [ID9]
Subtheme: medication prompts	"When I'm not well then I'm absolutely rubbish at taking medication. So yes, prompts to take medication and stuff would probably be for me a good idea." [ID8]
Subtheme: personalization	"I think that people are gonna have very different preferences and people are going to want to use this in very different ways. And I think that one size isn't gonna fit all. And if you're gonna do this, then you need to build in that adaptability within it." [ID15]
Theme 3: when to use the app	
Subtheme: would monitor constantly	"I think for me it would be useful to have it kind of on in the background all the time, so that then you can kind of see those periods where you're not 100% sure what it was like. You can go back and say oh actually I've had this amount of time...especially when the mood's low you can I find that I might think, oh, I've only been OK for a week, but it would give you a bit more of a realistic insight to say oh actually you've had a few months when you've been okay." [ID7]
Subtheme: not track all the time	"As for using it every day, I probably wouldn't. I would probably just dip in if I felt my mood starting to get a bit lower, a bit high." [ID14]
Subtheme: periods when tracking is particularly important	"I guess the idea would be that the app would learn about your behavior, so it would be able to predict when you know things might be difficult for you, or you know if certain times of the year are more difficult than being able to provide that support for you and then being able to toggle things on and off." [ID10]
Theme 4: risks of using the app	
Subtheme: data security concerns	"I think for me anyway, because you've got a lot of data, I'd be a little bit worried about where that data was going if it was being shared." [ID7]
Subtheme: dislike of the app when unwell	"I also got really paranoid when I was ill...so you, you know, would you get to the point where you didn't trust the app where you thought that people were trying to get you, you know, sectioned or trying to get your money off you or, you know, trying to do some sort of nefarious activities, you know, against you at that time?" [ID 14]
Subtheme: feeling monitored in a negative way	"Obviously there's a danger where you feel that you're being monitored." [ID13]
Subtheme: losing autonomy	"If the app sensed you know that things weren't going too well, you know, in whatever kind of data point that is, then it kinda feels like my, I dunno, sense of agency is removed, or autonomy...it doesn't feel good, it doesn't necessarily feel like an empowering thing." [ID10]
Subtheme: potential for the data to be used against you	"But it depends in certain situations, like if you got an abusive partner or something like that, you might not want them to know so." [ID2]
Subtheme: the risks of oversharing data	"Massive risk is taken...Overshare...there's a real risk the if you've got this sort of information you could just say to anybody any stranger: look at all this. The inhibitors aren't there. You lose the self-control, the inner voice, when you're high." [ID13]
Subtheme: the fear of mental health services	"I'd be scared of what they'd make of that information and whether they'd make decisions behind my back about my care. That yeah, that I wouldn't agree with." [ID4]

XSL•FO
RenderX

Table 4. Quotes related to theme 5 (sharing with family and friends) and theme 6 (sharing with health care professionals).

Themes and subthemes	Quotes
Theme 5: sharing with family and friends	
Subtheme: carefully choosing the right person to share with	"So I think the ability to share with other people is is good, but as a everyone said, you need to be really careful and have the options of who you share it with." [ID14]
Subtheme: facilitating conversations about relapse and care	"Yeah, I think it could create some difficult conversations, but the friend would usually be able to tell pretty quickly, especially if it's mania they'd be able to tell just by giving me a quick phone call." [ID4]
Subtheme: family and friends tracking mood	"That's it really, with me. And of course my partner does pull me back, do you know what I mean, when he sees it running out of, do you know what I mean, running of control." [ID3]
Theme 6: sharing with health care professionals	
Subtheme: skepticism about sharing with mental health services	"They don't take at face value what I'm telling them and therefore they are making comments like 'Ohh sorry, we don't believe you' and all this sort of stuff, so I'm opening up saying I'm ill and they're saying but you're talking to me, you're presenting a very strong individual, go away." [ID13]
Subtheme: mood monitoring to guide treatment	"I will say it amazed me actually as in I could see my highs and my lows and I'd be like you know where as if I hadn't been doing that, I wouldn't have been able to put that across to the health professional because it would kind of just gone you know, so this...was actually really beneficial to me." [ID16]
Subtheme: sharing less with professionals versus family and friends	"Do it myself first and share that sort of more directly with all my husband or whatever rather than sharing it sort of with the mental health team straight off." [ID7]
Subtheme: sharing relies on trust	"I think you'd have to be at one with who you're sharing with so you wouldn't just be sharing indiscriminately with health care professionals who that don't know you...my consultant psychiatrist, I've had, you know for 16, almost 16 years, so you know, it's lucky for me that I know her and true her and she knows me." [ID12]
Subtheme: would share in retrospect but not in real time	"Retrospectively, like it would work, I guess in the same way how I've been like using like the previous apps, I think in real time, I guess it could be like quite a few problems." [ID10]

Participants articulated a wide variety of different needs and preferences for the use of the app; some were commonly endorsed, and some were only endorsed by a few but for specific reasons, for example, because they lived alone and needed more objectivity (Tables 3 and 4). Most (35/45, 78%) subthemes were articulated, even if very briefly, in all 3 focus groups. Naturally, some subthemes were discussed in much more depth in certain focus groups. This is likely to represent important and generalizable data [21]. A minority (13/45, 29%) of subthemes were present in 1 or 2 focus groups (eg, medication prompts, providing continuity of care, positive messages, reassurance, data security concerns, feeling monitored [negative], potential for abuse, the risks of oversharing data, facilitating conversations about relapse and care, skepticism about sharing with mental health services, monitoring constantly, self-rating of mood can be unreliable, and stop self-monitoring when unwell).

Theme 1: Purpose of Using the App

Views on the purpose of the app were mixed (Table 3), and the participants wished to use the app in a variety of different ways. Some participants reported multiple uses for the app; others wished to use it for one specific purpose. The most popular was using the information they inputted into and reflecting back from the app to get a sense of where their mood was at that moment and to self-manage their BD. This could take the form of scaling back commitments, communicating with one's support network, changing medication, ensuring sufficient sleep, and doing more exercise, depending on the mood state. Many people wished to share information about their mood with trusted individuals around them and some with their health care professional.

Many wished to use the app to prevent future relapse by identifying early warning signs and triggers and improving long-term insight by examining both where their mood was now and the long-term changes in mood over many months and years. Many wanted a safety net for when they lost insights into how low or high their mood was at that moment, to both prevent the worsening of their mood and to avoid the negative behavioral consequences of being unwell, for example, excessive spending leading to the accumulation of long-term debt. There was an acknowledgment that the app might occasionally be wrong and that this information would be used to inform decision-making but not be the sole driver of it.

Theme 2: Desired Features

Participants were complimentary of passive monitoring (provided it was accurate and gave usable data), particularly regarding the ease of tracking their data over many years and when they may not wish to do so, for example, when they were unwell (Table 4). Many identified that a key barrier to them tracking their mood at present was the amount of time and effort it took, which they felt passive approaches would solve. Function and usability were key—the app being accurate in its core functions, for example, roughly identifying mood state. Participants wished to incorporate a crisis plan and personal alert system into the app; suggestions included writing things a person wished to be reminded of when they were well (such as a crisis plan) that could then be shown to them when they

XSL•FO
RenderX

were unwell, for example, around self-management of the illness. Another suggestion was receiving a notification from the app simply informing a person that they might be unwell as well as potentially sharing that with a predetermined trusted individual (explored in more detail in Theme 5: Sharing With Family and Friends section). The ability to customize and personalize aspects of the app seemed key, for example, toggling on-off certain aspects of data collection or any sharing. It was acknowledged that individuals would wish to use the app in a variety of different ways with different preferences around data security and the way they manage their BD.

Theme 3: When to Use the App

Opinions regarding when participants might use the app were mixed (Table 4). Some wished to monitor it constantly, acknowledging that doing so may help them triangulate their mood around important life events in the hope of managing their illness. Others disliked the idea of tracking all the time and did not wish to do so when well. Some mentioned that there were times, for example, around particularly stressful life events, when tracking might be more important.

Theme 4: Risks of Using the App

Concerns were raised around data security and possible breaches of data (Table 4), including from the individual who was unwell, for example, oversharing their information or data with someone who could potentially harm them. Susceptibility to abuse because of the collected data were a concern and some participants felt as the stigma of their illness had been used against them previously. Some participants were concerned they might feel monitored and that they would develop a dislike of the app if they felt particularly paranoid when unwell, potentially deleting it or removing their data from it. In total, 2 participants expressed concerns that the app could decrease their sense of autonomy, particularly if it was directing a person to do certain things or sharing their information with others. Many expressed concerns and fear of the involvement of mental health services deriving from previous negative experiences, including not feeling heard, being detained under the Mental Health Act, and a lack of trust in their health care professional.

Theme 5: Sharing With Family and Friends

We explored sharing with others in detail as it appeared to be a particularly important and potentially difficult-to-implement aspect of the app (Table 4). Participants felt that carefully choosing the right person to share any data with was extremely important. Some participants felt that their family and friends would wish to see their mood charts over long periods, while others felt that that information would not be of interest. Some wished for a single chosen person or a couple of chosen people to receive a notification when they were unwell as a safety net—building on previous nuanced conversations around safety or crisis planning, relapse, self-management, the involvement of mental health services, etc. Others would not choose this functionality and would not wish to share their data.

Theme 6: Sharing With Health Care Professionals

Many individuals did not wish to share their mood data with their health care professional in real time, due to concerns about an overreaction from mental health services, coercion, and a preference for self-management (Table 4). This seemed to be determined by the degree of trust participants had in their health care professional. However, most would share the data about their mood in retrospect to help with diagnosis, assessment, and subsequent treatment, for example, medication monitoring.

Discussion

Principal Findings

This qualitative study explored what people with BD wanted from a passive, mood-tracking app and tool to help them manage their BD. We explored the purpose, desired features, and usability in the context of ethical development of technology considering potential risks and safety of use and misuse. We explored the potential methods of incorporating desired features into the app in ways that would not compromise safety or cause unintended consequences, for example, sharing data with others.

Fundamentally, people with BD wanted to use passive mood-monitoring technology for a wide variety of purposes using different methods. They wished to use the data in a diverse and pragmatic way to help them self-manage their BD with greater insight and a better understanding of potential triggers. We discuss key features that may aid this functionality, for example, sharing with trusted individuals and reminders of crisis plans or personal alert systems. We discuss the future development of passive mood-monitoring technologies and particularly state that it should not be assumed that the involvement of formal mental health services is desired.

Most individuals principally wished to use the app to aid the already existing self-management of their BD. The way individuals wished to self-manage varied, and a preference for personalization was clear to accommodate the variety of strategies of using the app and the data it generated. This accords with previous work where insight, understanding, and self-management [6] to aid coping and prevention of relapse were the most salient reasons for engaging with self-monitoring [5,22-24]. Some subthemes were only articulated by a minority of participants, for example, using the information from the app to adjust their medication, providing continuity of care, and receiving positive or encouraging messages from the app. These less frequently endorsed subthemes highlight the multiple, different, and individual ways that the data generated by the app could be used. This kind of technological intervention could present data to individuals who can then use these data flexibly to prompt a variety of more nuanced behavioral changes.

Sharing with health care professionals appeared to be a less important reason for developing the app; again, this chimes with previous findings [5]. Individuals in this study felt that sharing their data and facilitating conversations around their illness with someone and people they trust was important, reinforcing what previous findings had demonstrated [5]. Communication with trusted people to improve relationships and manage the condition together seems important for many people with BD, while for others who do not have a trusted individual to check on them, the app seemed to provide an alternative safety net.

Fundamentally the app needed to be easy to use and accurate when determining mood. This presents programming and

XSL•FO
RenderX

technical challenges around performance, validity, and human-computer interaction. The RADAR platform, as used in depression, multiple sclerosis, and epilepsy, uses passive monitoring to obtain a general detection of important changes in the underlying condition, which then triggers active data completion to provide more granularity and greater sensitivity and specificity [17,25,26]. Participants articulated that the app does not necessarily have to be accurate all the time (subtheme: aid not a tool to be trusted blindly) and that any data from the app would be triangulated with contextual information, such as their own personal judgment on their mood and the opinions of those around them. It does seem important to have an awareness of how accurate the app is likely to be and this will factor into any decision-making that is reliant on the app's data [27].

Participants in our study suggested feasible ways of incorporating desired features and mitigating the risks of this technology. Crisis plans and personal alert systems could be set up when an individual is well, and then they could be reminded of crisis plans and personal alert systems when they are unwell (similar methods have been used in diabetes care [28] and epilepsy [29]). The person with BD could nominate 1 individual to receive a notification if the app considered that the person with BD might be unwell—again, similar to recent developments with seizure or diabetes tracking apps [28,29]. These features can all be chosen actively and toggled on and off, including all individual data collection methods. This would incorporate some aspects of personalization that seem important to people with BD, both in this study and in others, to account for differing needs, patterns of illness, and self-management strategies [5,22,24,30]. Data security should be considered carefully, and we should assume that many individuals will be skeptical about sharing their data with mental health services.

Participants in our focus groups did not feel that the app would feel confronting and potentially exacerbate any negative thoughts [31] via increasing awareness of depressive symptoms; however, other studies have suggested that this may be the case with mood-monitoring apps [32,33]. Some participants were concerned about the risks of using the app and there was a concern about technology exacerbating the risks that people with BD already face, for example, vulnerability, abuse, a loss of autonomy and independence, and stigma.

Limitations

The limitations of the study are mentioned subsequently. The sample reported here is small and is unlikely to be representative of those with BD more broadly, partly because we specifically recruited individuals with an interest in smartphone-based interventions and mood tracking. However, it did show diversity, both in terms of demographics and subthemes and views articulated. The recruitment was based on the kind of people we expected would ultimately use the app therefore was a pragmatic sampling method to best inform the final intervention. As most of these individuals were interested in mood-monitoring interventions, the degree of engagement and detailed discussion was likely higher than if we had selected individuals who did not have a particular interest in mood-tracking or digital interventions. Another limitation of this work was that we did not systematically record ethnicity data or previous experience with mood-tracking platforms. We were limited to how much personal data we collected and reported so that the data, combined with direction quotations, would not identify participants who were assured of anonymity through ethical review and individual consent.

The app was discussed hypothetically with a demonstration or discussion of possible features or usability; therefore, participants did not use the app. It is possible that following use, opinions may change; however, mostly the discussion around potential features, risks, and usability was theoretical, hence we decided to include a hypothetical discussion. Further work could iteratively develop the app, incorporating this feedback and these views into its development. A series of rounds of qualitative work could be conducted to gradually develop a feasible, usable, and acceptable app for people with BD. Furthermore, these results are specific to the RADAR platform, but there are many commonalities between passive behavior-tracking apps. Therefore, it is reasonable to consider the applicability of other platforms. These focus groups highlight the importance of accuracy and performance of digital mood-tracking apps [32,34,35].

Using the principles of RRI, we have considered and reflected on the issues identified by stakeholders as well as mitigation practices in the iterative development of this app; however, this is largely hypothetical. As we develop the app further into a useable prototype and the clinical utility is increasingly explored, we will continue engaging with stakeholders to anticipate and act upon risk detection and risk surveillance to anticipate potential misuse or unintended use as well as prioritizing safety, inclusivity, and responsible innovation [11]. This is particularly important considering the potential risks or harms highlighted here and previous concerns around the potential adverse effects of mood-monitoring interventions [32,33].

Conclusions

This study suggests that people with BD, who are interested in using passive technology to monitor mood, wish to do so for a wide variety of purposes and using a variety of different methods of using the data. Principally, people with BD wished to use this novel technology to aid them in self-managing their BD with greater insight and a better understanding of potential triggers. There are a few key features that may aid this functionality, for example, sharing with trusted individuals and the reminder of crisis plans and personal alert systems. Future development of passive mood-monitoring technologies should not assume that the involvement of formal mental health services is desired.

XSL•FO
RenderX

Acknowledgments

This work was supported by a Wellcome Trust Fellowship to LAW. RM was the chief investigator on a grant from the UK National Institute for Health and Care Research (NIHR) Applied Research Collaboration East Midlands Mental Health and Well-being Theme.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

LAW and RM were responsible for the original study design. LAW was responsible for data analysis and initial drafting of the manuscript. All authors were responsible for subsequent interpretation, editing, and rewriting of the manuscript.

Conflicts of Interest

RM has received funding from Novartis to serve on a data management and ethics committee for 2 trials on the treatment of depression.

Multimedia Appendix 1
COREQ (Consolidated Criteria for Reporting Qualitative Studies) checklist.
[DOCX File , 21 KB - formative_v9i1e65140_app1.docx]

Multimedia Appendix 2
Details of possible data collection and summary of previous works.
[DOCX File , 683 KB - formative_v9i1e65140_app2.docx]

References

1. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct 09;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]
2. Torous J, Choudhury T, Barnett I, Keshavan M, Kane J. Smartphone relapse prediction in serious mental illness: a pathway towards personalized preventive care. World Psychiatry 2020 Oct 15;19(3):308-309 [FREE Full text] [doi: 10.1002/wps.20805] [Medline: 32931109]
3. Depp C, Torous J, Thompson W. Technology-based early warning systems for bipolar disorder: a conceptual framework. JMIR Ment Health 2016 Sep 07;3(3):e42 [FREE Full text] [doi: 10.2196/mental.5798] [Medline: 27604265]
4. Matcham F, Barattieri di San Pietro C, Bulgari V, de Girolamo G, Dobson R, Eriksson H, et al. Remote assessment of disease and relapse in major depressive disorder (RADAR-MDD): a multi-centre prospective cohort study protocol. BMC Psychiatry 2019 Feb 18;19(1):72 [FREE Full text] [doi: 10.1186/s12888-019-2049-z] [Medline: 30777041]
5. Majid S, Morriss R, Figueredo G, Reeves S. Exploring self-tracking practices for those with lived experience of bipolar disorder: learning from combined principles of patient and public involvement and HCI. In: Proceedings of the 2022 ACM Designing Interactive Systems Conference. 2022 Presented at: DIS '22; June 13-17, 2022; Virtual Event p. 1907-1920 URL: https://dl.acm.org/doi/10.1145/3532106.3533531 [doi: 10.1145/3532106.3533531]
6. Murnane E, Cosley D, Chang P, Guha S, Frank E, Gay G, et al. Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: implications for the design of technologies to manage mental health. J Am Med Inform Assoc 2016 May;23(3):477-484. [doi: 10.1093/jamia/ocv165] [Medline: 26911822]
7. Morriss R, Vinjamuri I, Faizal MA, Bolton CA, McCarthy JP. Training to recognise the early signs of recurrence in schizophrenia. Cochrane Database Syst Rev 2013 Feb 28(2):CD005147. [doi: 10.1002/14651858.CD005147.pub2] [Medline: 23450559]
8. Bauer M, Glenn T, Alda M, Grof P, Bauer R, Ebner-Priemer UW, et al. Longitudinal digital mood charting in bipolar disorder: experiences with chronorecord over 20 years. Pharmacopsychiatry 2023 Sep 07;56(5):182-187 [FREE Full text] [doi: 10.1055/a-2156-5667] [Medline: 37678394]
9. Post RM, Denicoff KD, Leverich GS, Altshuler LL, Frye MA, Suppes TM, et al. Morbidity in 258 bipolar outpatients followed for 1 year with daily prospective ratings on the NIMH life chart method. J Clin Psychiatry 2003 Jun 15;64(6):680-738. [doi: 10.4088/jcp.v64n0610] [Medline: 12823083]
10. Bos F, Schreuder M, Doornbos B, Snippe E, Bruggeman R, van der Krieke L, et al. Prospective early warning signals to detect transitions to manic and depressive episodes in bipolar disorder. Eur Psychiatry 2021 Aug 13;64(S1):S78-S79. [doi: 10.1192/j.eurpsy.2021.237]

XSL•FO
RenderX

11. Eyre HA, Ellsworth W, Fu E, Manji HK, Berk M. Responsible innovation in technology for mental health care. Lancet Psychiatry 2020 Sep;7(9):728-730 [FREE Full text] [doi: 10.1016/S2215-0366(20)30192-9] [Medline: 32828153]
12. Stilgoe J, Owen R, Macnaghten P. Developing a framework for responsible innovation. Res Policy 2013 Nov;42(9):1568-1580. [doi: 10.1016/j.respol.2013.05.008]
13. Jirotka M, Grimpe B, Stahl B, Eden G, Hartswood M. Responsible research and innovation in the digital age. Commun ACM 2017 Apr 24;60(5):62-68. [doi: 10.1145/3064940]
14. Lopez-Campos G, Gabarron E, Martin-Sanchez F, Merolli M, Petersen C, Denecke K. Digital interventions and their unexpected outcomes - time for digitalovigilance? Stud Health Technol Inform 2024 Jan 25;310:479-483. [doi: 10.3233/SHTI231011] [Medline: 38269849]
15. Middle R, Welch L. Experiences of digital exclusion and the impact on health in people living with severe mental illness. Front Digit Health 2022 Nov 22;4:1004547 [FREE Full text] [doi: 10.3389/fdgth.2022.1004547] [Medline: 36483535]
16. Greer B, Robotham D, Simblett S, Curtis H, Griffiths H, Wykes T. Digital exclusion among mental health service users: qualitative investigation. J Med Internet Res 2019 Jan 09;21(1):e11696 [FREE Full text] [doi: 10.2196/11696] [Medline: 30626564]
17. Sun S, Folarin AA, Zhang Y, Cummins N, Liu S, Stewart C, et al. The utility of wearable devices in assessing ambulatory impairments of people with multiple sclerosis in free-living conditions. Comput Methods Programs Biomed 2022 Dec;227:107204 [FREE Full text] [doi: 10.1016/j.cmpb.2022.107204] [Medline: 36371974]
18. Simblett S, Matcham F, Curtis H, Greer B, Polhemus A, Novák J, et al. Patients' measurement priorities for remote measurement technologies to aid chronic health conditions: qualitative analysis. JMIR Mhealth Uhealth 2020 Jun 10;8(6):e15086 [FREE Full text] [doi: 10.2196/15086] [Medline: 32519975]
19. Polhemus A, Novak J, Majid S, Simblett S, Morris D, Bruce S, et al. Data visualization for chronic neurological and mental health condition self-management: systematic review of user perspectives. JMIR Ment Health 2022 Apr 28;9(4):e25249 [FREE Full text] [doi: 10.2196/25249] [Medline: 35482368]
20. Castleberry A. NVivo 10 [software program]. Version 10. QSR international; 2012. Am J Pharm Educ 2014 Feb;78(1):25. [doi: 10.5688/ajpe78125]
21. Guba EG. Criteria for assessing the trustworthiness of naturalistic inquiries. Educ Technol Res Dev 1981 Jun;29(2):75-91 [FREE Full text] [doi: 10.1007/bf02766777]
22. Bos FM, Snippe E, Bruggeman R, Doornbos B, Wichers M, van der Krieke L. Recommendations for the use of long-term experience sampling in bipolar disorder care: a qualitative study of patient and clinician experiences. Int J Bipolar Disord 2020 Dec 01;8(1):38 [FREE Full text] [doi: 10.1186/s40345-020-00201-5] [Medline: 33258015]
23. Saunders KE, Bilderbeck AC, Panchal P, Atkinson LZ, Geddes JR, Goodwin GM. Experiences of remote mood and activity monitoring in bipolar disorder: a qualitative study. Eur Psychiatry 2017 Mar 27;41(1):115-121 [FREE Full text] [doi: 10.1016/j.eurpsy.2016.11.005] [Medline: 28135594]
24. Bos FM, Snippe E, Bruggeman R, Wichers M, van der Krieke L. Insights of patients and clinicians on the promise of the experience sampling method for psychiatric care. Psychiatr Serv 2019 Nov 01;70(11):983-991. [doi: 10.1176/appi.ps.201900050] [Medline: 31434558]
25. Matcham F, Leightley D, Siddi S, Lamers F, White KM, Annas P, et al. Remote Assessment of Disease and Relapse in Major Depressive Disorder (RADAR-MDD): recruitment, retention, and data availability in a longitudinal remote measurement study. BMC Psychiatry 2022 Feb 21;22(1):136 [FREE Full text] [doi: 10.1186/s12888-022-03753-1] [Medline: 35189842]
26. Simblett SK, Biondi A, Bruno E, Ballard D, Stoneman A, Lees S, et al. Patients' experience of wearing multimodal sensor devices intended to detect epileptic seizures: a qualitative analysis. Epilepsy Behav 2020 Jan;102:106717. [doi: 10.1016/j.yebeh.2019.106717] [Medline: 31785481]
27. Andrews JA, Craven MP, Jamnadas-Khoda J, Lang AR, Morriss R, Hollis C, et al. Health care professionals' views on using remote measurement technology in managing central nervous system disorders: qualitative interview study. J Med Internet Res 2020 Jul 24;22(7):e17414 [FREE Full text] [doi: 10.2196/17414] [Medline: 32706664]
28. Debong F, Mayer H, Kober J. Real-world assessments of mySugr mobile health app. Diabetes Technol Ther 2019 Jun;21(S2):S2-35-S2-40. [doi: 10.1089/dia.2019.0019]
29. Alzamanan MZ, Lim KS, Akmar Ismail M, Abdul Ghani N. Self-management apps for people with epilepsy: systematic analysis. JMIR Mhealth Uhealth 2021 May 28;9(5):e22489 [FREE Full text] [doi: 10.2196/22489] [Medline: 34047709]
30. Majid S, Reeves S, Figueredo G, Brown S, Lang A, Moore M, et al. The extent of user involvement in the design of self-tracking technology for bipolar disorder: literature review. JMIR Ment Health 2021 Dec 20;8(12):e27991 [FREE Full text] [doi: 10.2196/27991] [Medline: 34931992]
31. van Genugten CR, Schuurmans J, Lamers F, Riese H, Penninx BW, Schoevers RA, et al. Experienced burden of and adherence to smartphone-based ecological momentary assessment in persons with affective disorders. J Clin Med 2020 Jan 23;9(2):322 [FREE Full text] [doi: 10.3390/jcm9020322] [Medline: 31979340]
32. Dubad M, Winsper C, Meyer C, Livanou M, Marwaha S. A systematic review of the psychometric properties, usability and clinical impacts of mobile mood-monitoring applications in young people. Psychol Med 2018 Jan 23;48(2):208-228. [doi: 10.1017/S0033291717001659] [Medline: 28641609]

XSL•FO
RenderX

33. Palmier-Claus J, Lobban F, Mansell W, Jones S, Tyler E, Lodge C, et al. Mood monitoring in bipolar disorder: is it always helpful? Bipolar Disord 2021 Jun 19;23(4):429-431. [doi: 10.1111/bdi.13057] [Medline: 33570820]
34. Faurholt-Jepsen M, Munkholm K, Frost M, Bardram JE, Kessing LV. Electronic self-monitoring of mood using IT platforms in adult patients with bipolar disorder: a systematic review of the validity and evidence. BMC Psychiatry 2016 Jan 15;16(1):7 [FREE Full text] [doi: 10.1186/s12888-016-0713-0] [Medline: 26769120]
35. Ortiz A, Maslej MM, Husain MI, Daskalakis ZJ, Mulsant BH. Apps and gaps in bipolar disorder: a systematic review on electronic monitoring for episode prediction. J Affect Disord 2021 Dec 01;295:1190-1200. [doi: 10.1016/j.jad.2021.08.140] [Medline: 34706433]

Abbreviations

BD: bipolar disorder
COREQ: Consolidated Criteria for Reporting Qualitative Studies
PPI: patient and public involvement
RADAR: Remote Assessment of Disease and Relapse
RRI: responsible research and innovation

Edited by A Mavragani; submitted 06.08.24; peer-reviewed by M Mcinnis; comments to author 28.11.24; revised version received 04.12.24; accepted 15.12.24; published 07.02.25.

Please cite as:
Astill Wright L, Moore M, Reeves S, Vallejos EP, Morriss R
Improving the Utility, Safety, and Ethical Use of a Passive Mood-Tracking App for People With Bipolar Disorder Using Coproduction: Qualitative Focus Group Study
JMIR Form Res 2025;9:e65140
URL: https://formative.jmir.org/2025/1/e65140
doi:10.2196/65140
PMID:39918865

©Laurence Astill Wright, Matthew Moore, Stuart Reeves, Elvira Perez Vallejos, Richard Morriss. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Social Media Use and Oral Health–Related Misconceptions in Saudi Arabia: Cross-Sectional Study

Rahaf Hamdan BinHamdan[1], BDS; Salwa Abdulrahman Alsadhan[1], BDS, MSc; Arwa Zohair Gazzaz[1], BDS, MPH, PhD; AlBandary Hassan AlJameel[1], BDS, MSc, PhD

Department of Periodontics and Community Dentistry, College of Dentistry, King Saud University, Riyadh, Saudi Arabia

Corresponding Author:
Rahaf Hamdan BinHamdan, BDS
Department of Periodontics and Community Dentistry
College of Dentistry
King Saud University
Diriyah 12372
Riyadh, 11545
Saudi Arabia
Phone: 966 0114677743
Email: rahafbinhamdan@gmail.com

Abstract

Background: Social media has become a central tool in health communication, offering both opportunities and challenges. In Saudi Arabia, where platforms like WhatsApp, Snapchat, and Instagram are widely used, the quality and credibility of oral health information shared digitally remain critical issues. Misconceptions about oral health can negatively influence individuals' behaviors and oral health outcomes.

Objective: This study aimed to describe the patterns of social media use and estimate the prevalence of oral health–related misconceptions among adults in Saudi Arabia. Additionally, it assessed the associations between engagement with oral health information, self-reported oral health, and the presence and count of these misconceptions.

Methods: A cross-sectional survey was conducted over 10 weeks, targeting adults aged 15 years and older in Saudi Arabia. Data were collected from a total sample size (n=387) via a questionnaire distributed through targeted advertisements on Instagram, TikTok, Snapchat, and X (Twitter). The prevalence of oral health–related misconceptions was estimated using descriptive statistics, including counts and percentages. Chi-square tests described sociodemographic, social media engagement, and self-reported oral health. Logistic and Poisson regression analyses were used to assess associations between engagement and self-reported oral health with misconceptions. Logistic regression models provided odds ratios and adjusted odds ratios with 95% CI to assess the presence of oral health misconceptions. Poisson regression was used to calculate mean ratios and adjusted mean ratios (AMRs) for the count of misconceptions.

Results: WhatsApp (n=344, 89.8%) and Instagram (n=304, 78.9%) were the most frequently used social media platforms daily. Common oral health misconceptions included beliefs that "Pregnancy causes calcium loss in teeth" (n=337, 87%) and "Dental treatment should be avoided during pregnancy" (n=245, 63.3%). Following dental-specific accounts was significantly associated with lower odds of having any misconceptions (adjusted odds ratio 0.41, 95% CI 0.22-0.78) and a lower count of misconceptions (AMR 0.87, 95% CI 0.77-0.98). Conversely, trust in social media as a source of oral health information was associated with a higher count of misconceptions (AMR 1.16, 95% CI 1.02-1.31).

Conclusions: Social media platforms are essential yet double-edged tools for oral health information dissemination in Saudi Arabia. Participants who followed dental-specific accounts had significantly lower misconceptions, while trust in social media as a source of information was linked to higher counts of misconceptions. These findings highlight the importance of promoting credible content from verified sources to combat misconceptions. Strategic collaborations with dental professionals are necessary to enhance the dissemination of accurate oral health information and public awareness and reduce the prevalence of oral health–related misconceptions.

(JMIR Form Res 2025;9:e70071) doi:10.2196/70071

XSL•FO
RenderX

KEYWORDS

social media; oral health; health misinformation; digital health; Saudi Arabia; public health; Instagram; Snapchat; TikTok; Twitter

Introduction

We live in a new era filled with rapid technological advancements that facilitate everyday life. The internet has been used worldwide for many purposes, and it has become an important tool that can improve the distribution of information and health care delivery and outcomes [1]. Web 2.0 is the second generation of the World Wide Web, emphasizing the shift from static web pages to more dynamic and participatory content, with social media becoming a key element shaping daily life [2]. Globally, social media platforms such as Facebook, Twitter, YouTube, and WhatsApp allow individuals to access and share information at a unique scale [3-5]. In Saudi Arabia, platforms like WhatsApp, Snapchat, and YouTube are heavily used for general communication and increasingly for seeking health-related information [6].

Social media platforms have emerged as powerful tools in health communication, offering both opportunities and challenges. In oral health, social media platforms can provide greater access to oral health information, engaging educational content, and web-based interaction with dental professionals [7,8]. For example, YouTube has been identified as a valuable resource for educating patients with leukemia on oral care [9]. Similarly, Snapchat-based interventions have effectively enhanced oral health knowledge among pregnant women in Saudi Arabia [10]. Oral health promotion campaigns delivered through social media platforms have been shown to improve oral health outcomes [11,12].

Despite social media benefits, information quality and reliability with sources often lack clear authorship or transparency, leading to difficulty in assessing information credibility for users [13-16]. This is particularly problematic in oral health, where misinformation—false or inaccurate health claims—could lead to negative oral health outcomes if it was taken by the users and distributed without consulting an expert [17,18]. Disinformation—information deliberately spread with the intent to harm—exacerbates these risks [19]. For example, practices such as using oil pulling to reduce mouth bacteria and prevent dental caries or relying on home remedy pastes to alleviate tooth pain are often shared digitally without scientific evidence [20]. This study uses “misconception” as a comprehensive term to encompass these various forms of inaccurate or misleading oral health information.

In addition to content concerns, sex-based differences in health behaviors, perceptions, and information-seeking patterns play a role in how individuals engage with oral health information on social media. It has been shown that women are more likely to seek web-based health information and engage with health care services compared to men [21]. They are also more likely to engage with and trust content shared on social media, making them a key demographic for addressing oral health misconceptions through these platforms [22]. Conversely, men may be less likely to search for web-based health information or disease prevention, which may lead to the persistence of misconceptions about oral health practices [23]. Sex differences explain that while women may be more susceptible to health misinformation due to their higher engagement, men may require distinct strategies to capture their attention and promote oral health awareness [24]. Such differences suggest that sex may influence the way individuals access and interpret oral health–related information on social media.

In Saudi Arabia, where 99% of the population are internet users actively engaging with social media platforms, this widespread use facilitates the diffusion of oral health misconceptions, leading users to accept inaccurate information and challenge health care professionals [25]. The challenge lies in the sheer quantity and variable quality of information available on social media platforms, which could lead to negative oral health outcomes if one misconception keeps spreading. For example, widespread beliefs about ineffective home remedies and misconceptions about neglecting primary teeth, dental treatment, and oral care during pregnancy continue to be prevalent among people in Saudi Arabia [26-28]. These misconceptions lead to public health implications, as untreated oral health issues can exacerbate systemic health conditions and increase health care costs [29]. To our knowledge, no previous study has specifically explored the context of oral health misconceptions on social media within this population, highlighting the need for this study.

This study aims to describe the patterns of social media use and to estimate the prevalence of oral health–related misconceptions among adults in Saudi Arabia. Additionally, the study aims to explore how social media engagement influences oral health misconceptions, with results stratified by sex to explore potential differences.

Methods

Study Design

This study is a cross-sectional survey. We collected the data over a 10-week period from September 9, 2024, to November 17, 2024, using an Arabic-language questionnaire developed on SurveyMonkey (SurveyMonkey Inc) among adults aged 15 years and older residing in Saudi Arabia. The questionnaire was initially developed in English and translated into Arabic to suit the target population. The questionnaire link was distributed through targeted web-based advertisements on the most popular social media platforms in Saudi Arabia [6], including Instagram, X (Twitter), Snapchat, and TikTok. Participants clicking on the advertisements were redirected to a dedicated web page containing a detailed study description and consent form, followed by the questionnaire. At the start of the survey, an introductory statement provided information on the study objectives, assured confidentiality, and included the researcher’s contact information. We obtained the consent via an “Agree” button, which participants clicked to access the survey. Participants were informed that the survey would take approximately five minutes to complete. They were assured

XSL•FO
RenderX

that participation is entirely voluntary, with the option to skip questions or withdraw at any time.

Sample Recruitment

The study was conducted in a digital setting with an open survey accessible to anyone who clicked on the advertisements, using widely used social media platforms in Saudi Arabia (Instagram, X [Twitter], Snapchat, and TikTok) for recruitment. These platforms were chosen due to their high penetration rates and active use in Saudi Arabia [6], ensuring ease of access and effective targeting through their advertising systems. We designed the social media–targeted advertisements to be culturally appropriate and visually engaging, highlighting the voluntary nature of participation and the study's focus on oral health–related information. The study used a simple random sampling approach facilitated by the advertisement algorithms. These algorithms were used to define the target group—adults aged 15 years and older residing in Saudi Arabia—ensuring that all individuals within this demographic had an equal chance of seeing the advertisement. According to 2022 population data from the General Authority of Statistics (GASTAT) [30] and 2023 internet penetration figures from the Communications, Space, and Technology Commission [6], the target population was estimated at 24,045,450 individuals, representing 99% of the total population in Saudi Arabia who are internet users within the targeted age group. This recruitment strategy aimed to: ensure representation of the population of interest by using an algorithm without prespecified interests, thereby targeting all individuals within the defined demographic; and maximize geographical diversity by specifying the algorithm to reach individuals residing in Saudi Arabia.

However, it is important to acknowledge that participation was voluntary, introducing self-selection bias that may affect the sample's generalizability. Despite this, the approach was a practical and effective method for recruiting a large, diverse population.

Effectiveness of the Selected Social Media Platforms

To effectively reach a broad and diverse sample, we launched advertisement campaigns across Instagram, X (Twitter), Snapchat, and TikTok. User interactions were monitored throughout the campaigns to evaluate reach and sample representativeness. Snapchat and TikTok provided us with strategic support to optimize campaign strategies. Snapchat emphasized targeting mechanisms such as predefined and custom audiences, advanced demographics, and accurate algorithmic delivery, which enhanced the ability to reach users based on demographics, location, and language. TikTok highlighted its robust targeting tools, including lookalike or custom audiences, ensuring advertisements were delivered to relevant users while monitoring user interactions like click-through and conversion rates. Although Instagram and X (Twitter) did not provide direct support, their platforms offered advanced targeting capabilities via Ads Manager tools. These tools allowed for audience segmentation by age, location, and sex, enabling effective recruitment across diverse demographics.

Sample Size

To estimate the prevalence of oral health–related misconceptions, with a 5% margin of error and a 95% CI, a sample size of 384 participants was required (assuming a 50% prevalence to maximize sample size requirements). Anticipating a low participation rate typical of web-based surveys, we assumed a 2% participation rate, leading us to target the reach of approximately 19,200 users in our advertisement campaigns across all platforms to ensure the required sample size. Reach refers to the estimated number of users who would see the advertisement, with metrics from Instagram, X (Twitter), Snapchat, and TikTok helping us track both views and link clicks.

Selection Criteria

We included both male and female participants aged 15 years and older residing in Saudi Arabia and those who provided informed consent to participate in the study. The age threshold of 15 was selected based on the classification by the GASTAT in Saudi Arabia, which classifies individuals aged 15 years and older as adults for statistical and demographic purposes [31]. Additionally, adolescents in this age group are active users of social media and are likely to be exposed to oral health information digitally, making their inclusion essential to assess the study objectives.

Questionnaire

We structured the questionnaire into five domains with a total of 29 fixed items distributed across six screens designed to align with the study's objectives and draw on previously published and validated cross-sectional studies. Face validity was established through expert review by a panel of dental public health specialists. Experts assessed the clarity, relevance, and domain alignment of each item, as well as the overall comprehensiveness of the questionnaire within the study's objectives. We implemented minor revisions to item wording and categorization based on their feedback (Multimedia Appendix 1). No adaptive questioning was used; all items required mandatory responses prior to proceeding to subsequent sections pages. Respondents could review and modify their answers using the "Back" button.

The Patterns of Social Media Use

The first domain assessed participants' use patterns and preferred platforms with one item [6,32-35]. Participants were asked to select their preferred social media platforms across the popular platforms in Saudi Arabia, which included X (Twitter), Snapchat, WhatsApp, Instagram, Facebook, TikTok, Telegram, YouTube, Line, and LinkedIn [6]. Responses were categorized as "more than once a day," "once a day," "2-3 times a week," "4-6 times a week," "once per week" or "once per month," and "I don't use this platform."

Engagement With Oral Health Information

The second domain, containing 6 items, focused on the engagement with oral health information obtained from social media [6,32,33,36,37]. Participants were asked binary (yes or no) questions, including whether they "use social media to search for oral health information," "ever received oral health

XSL•FO
RenderX

info from nondental professionals on social media," "following dental-specific accounts on social media," "trust in oral health information from social media," and "influence their trust by social media profile content." Perception of social media as a good source for oral health information was assessed with response options: "yes," "sometimes," and "no." We dichotomize it to combine "yes" and "sometimes" into a single category representing participants who perceived social media as a good source of oral health information.

Self-Reported Oral Health

To capture self-reported oral health, we included a single item in the third domain using the Global Self-Ratings of Oral Health question for participants to rate their own oral health status on a 3-point Likert scale as "good," "average," and "poor." [38]. For analysis, self-reported poor oral health was defined as participants selecting "poor," while better oral health perceptions included those reporting "good" oral health. Participants who neither rated their oral health as good nor poor, reflecting a moderate perception of their oral health, included those reporting "average" status.

Oral Health–Related Misconceptions

The fourth domain presented 14 statements on common oral health misconceptions in Saudi Arabia, derived from literature [27,28], with response options of "agree," "disagree," and "don't know." We combined responses of "disagree" and "don't know" into a single category representing participants who do not believe in the misconceptions. The category "agree" includes those who do agree with these misconceptions.

Sociodemographic Factors

In the fifth domain, we collected sociodemographic information across seven items: age, which was categorized into five groups: 15-24, 25-34, 35-44, 45-54, and 55+ years; sex; education, which was classified into "high school or less," "associate or intermediate diploma," "bachelor's degree," and "master's or PhD"; occupation, which was grouped into "student," "governmental worker," "private worker," and "unemployed"; marital status, "single," "married," or "other"; region of residence, included the 13 administrative regions of Saudi Arabia: "Riyadh," "Makkah," "Madinah," "Qassim," "Eastern Region," "Asir Region," "Tabuk," "Hail," "Northern Region," "Jazan," "Najran," "Al-Baha," and "Al-Jawf"; and nationality, "Saudi" and "Non-Saudi" within Saudi Arabia. Age and region categories were based on distributions from the GASTAT [30,39], and education levels followed the National Qualifications Framework 2023 [40].

Pilot

In a preliminary phase of the study, we conducted a pilot to confirm the validity and relevance of this study prior to full implementation. We targeted a sample group representing 10% of the total study sample size, with 10 participants recruited from each of the four social media platforms, which are Instagram, X (Twitter), Snapchat, and TikTok, resulting in a total of 40 participants. Participants responded positively to the questionnaire, navigating it without difficulty. Based on feedback, we made two key adjustments to improve the survey. First, we added a "sometimes" response option to the question in the second domain, "Do you think social media is a good source for obtaining oral health information?" to provide a wider range of responses. Second, we simplified certain Arabic terms in the misconception domain to make the statements clear to participants. This pilot phase was conducted from June 17, 2024, to September 2, 2024. Participants involved in this pilot were not included in the final study sample.

Statistical Analysis

Counts and percentages were reported to describe participants' sociodemographic characteristics, their engagement with oral health information, and their self-reported oral health. We used the chi-square test to examine potential associations between the 3 domains and sex, identifying statistically significant relationships where applicable.

For the multivariable analysis, the independent variables included engagement with the oral health information domain (the manner of participants' engagement with oral health–related information in social media measured via the six binary items, each item entered as a separate exposure in the model) and self-reported oral health (participants' subjective evaluation of their oral health status measured via 3-point Likert scale). The dependent variables were defined as:

- Presence of oral health misconceptions (binary): had at least one misconception based on the 14 statements on oral health misconceptions.
- Number of oral health misconceptions (count): the total number of misconceptions identified per participant.

To assess these relationships, we used logistic regression models to present both crude (odds ratio) and adjusted odds ratio (AOR), with respective 95% CI, to examine engagement with oral health information and self-reported oral health (exposure variables) with the presence of any oral health misconceptions among social media users (outcome). Logistic regression was selected due to its appropriateness for modeling binary outcomes, allowing us to examine the likelihood of participants holding at least one misconception. Poisson regression is used to report the mean ratio and adjusted mean ratio (AMR) of oral health misconceptions (outcome) by the engagement with oral health information and self-reported oral health (exposure variables). Poisson regression was chosen because it is suitable for modeling count data, particularly when the outcome variable represents a frequency count of misconceptions per participant. We confirmed the appropriateness of the Poisson model by assessing equidispersion using the Pearson chi-square statistic, which indicated no overdispersion. Sex, education (as a measure of socioeconomic status), and nationality were selected and adjusted as potential confounders based on their theoretical relevance [41] and observed associations with both the exposure and the outcome in exploratory analyses.

A sensitivity analysis was performed for both regression models to assess the potential impact on the results if the response "don't know" was reclassified as "agree" in the 14 statements addressing oral health misconceptions. We reclassified "don't know" responses based on the assumption that these individuals might be more susceptible to misconceptions due to a lack of knowledge or confidence. This provided a more conservative,

XSL•FO
RenderX

robust estimate of misconception prevalence, allowing us to examine the influence of uncertainty on our findings across different analyses. A 2-tailed α level of .05 was used to test relationships in both directions, and all statistical analyses were conducted using Stata/BE (version 17.0; StataCorp).

Ethical Considerations

This study was conducted in compliance with the ethical standards set by the institutional review board of King Saud University. Research ethical approval was obtained, and the study was registered with the Institutional Review Board from the Committee of Human and Social Research Ethics, King Saud University (registration KSU-HE-23848). Participants were informed about the study objectives, procedures, and their rights before participating in the survey. Informed consent was obtained electronically via a mandatory "Agree" button at the start of the questionnaire. Only individuals who provided consent were able to proceed with the survey. The survey content was specifically designed to be low-risk without including sensitive or intrusive questions. Data were collected anonymously, with no personally identifiable information linked to the responses. The platform (SurveyMonkey) complied with strict data security and privacy protocols to ensure the protection of participants' responses. IP addresses were monitored to identify duplicate entries within a 24-hour period and then omitted before analysis. Participation was voluntary, and no monetary or material compensation was provided to participants in this study. This study did not involve the use of images or media containing identifiable participants.

Results

Campaign Reach and Interaction Across Social Media Platforms

Using targeted social media campaigns across Instagram, X (Twitter), TikTok, and Snapchat, we observed notable differences in reach and interaction across platforms. TikTok achieved the highest reach (836,477) and link clicks (67,678) during its 30-day campaign, followed by Instagram, which reached (53,341) users and generated (1007) link clicks over 60 days. X (Twitter) and Snapchat had comparatively lower interactions. Across all platform advertisements, the overall view rate was 6.6%, and the participation rate was 0.7%.

Sociodemographic Characteristics

Out of 640 individuals who initially responded to the questionnaire, 387 (76%) individuals completed it fully; the remaining (n=253) either provided incomplete responses or exited the survey after clicking "disagree" to participate in the questionnaire. These partial responses were excluded from the analysis to ensure data completeness.

The final sample size consisted of 248 (64%) female and 139 (35.9%) male participants (Table 1). The largest age group was 15-24 years (n=221, 57.1%), followed by 25-34 years (n=77, 19.9%). Most participants held a bachelor's degree (n=184, 47.5%) or a high school diploma or less (n=138, 35.6%). Regarding occupation, more than half of the participants were students (n=199, 51.4%), while 97 (25%) participants were unemployed. In terms of marital status, 299 (77.2%) of participants were single, while 82 (21.1%) of participants were married. Saudi participants accounted for 72.3% (n=280) of the sample, and most participants resided in Riyadh (n=258, 66.6%). Significant sex-based differences were observed in occupation (P<.001) and marital status (P=.02), with male participants more likely to be employed and married compared to female participants, who were predominantly unemployed and single (Table 1).

XSL•FO
RenderX

Table 1. Sociodemographic characteristics and regional distribution among adults aged 15 years and older residing in Saudi Arabia (n=387), described by sex[a].

Sociodemographic characteristics	Overall (n=387, 100%), n (%)	Male (n=139, 35.9%), n (%)	Female (n=248, 64%), n (%)	*P* value
Age (in years)				.35
15-24	221 (57.1)	78 (35.2)	143 (64.7)	
25-34	77 (19.9)	22 (28.5)	55 (71.4)	
35-44	45 (11.6)	19 (42.2)	26 (57.7)	
45-54	36 (9.3)	16 (44.4)	20 (55.5)	
55+	8 (2)	4 (50)	4 (50)	
Education				.23
High school or less	138 (35.6)	49 (35.5)	89 (64.4)	
Associate—intermediate diploma	39 (10)	15 (38.4)	24 (61.5)	
Bachelor's degree	184 (47.5)	61 (33.1)	123 (66.8)	
Master's degree—PhD	26 (6.7)	14 (53.8)	12 (46.1)	
Occupation				<.001
Student	199 (51.4)	69 (34.6)	130 (65.3)	
Governmental worker	43 (11.1)	24 (55.8)	19 (44.1)	
Private worker	48 (12.4)	28 (58.3)	20 (41.6)	
Unemployed	97 (25)	18 (18.5)	79 (81.4)	
Marital status				.02
Single	299 (77.2)	98 (32.7)	201 (67.2)	
Married	82 (21.1)	40 (48.7)	42 (51.2)	
Other	6 (1.5)	1 (16.6)	5 (83.3)	
Nationality				.92
Saudi	280 (72.3)	101 (36)	179 (63.9)	
Non-Saudi	107 (27.6)	38 (35.5)	69 (64.4)	
Region of residence				.19
Riyadh	258 (66.6)	84 (32.5)	174 (67.4)	
Makkah	45 (11.6)	21 (46.6)	24 (53.3)	
Madinah	13 (3.3)	7 (53.8)	6 (46.1)	
Qassim	11 (2.8)	2 (18.1)	9 (81.8)	
Eastern Region	45 (11.6)	21 (46.6)	24 (53.3)	
Asir Region	5 (1.2)	1 (20)	4 (80)	
Tabuk	1 (0.2)	0 (0)	1 (100)	
Hail	2 (0.5)	0 (0)	2 (100)	
Northern Region	1 (0.2)	0 (0)	1 (100)	
Jazan	3 (0.7)	1 (33.3)	2 (66.6)	
Najran	1 (0.2)	1 (100)	0 (0)	
Al-Baha	1 (0.2)	1 (100)	0 (0)	
Al-Jawf	1 (0.2)	0 (0)	1 (100)	

[a]Data were collected in a cross-sectional survey on oral health misconceptions conducted from September to November 2024.

XSL•FO
RenderX

Engagement With Oral Health Information and Self-Reported Oral Health

Female participants were significantly more likely than male participants to use social media for oral health information (n=156, 68.7% compared to n=71, 31.2%; *P*=.02) and follow dental-specific accounts (n=126, 72% compared to n=49, 28%; *P*=.003; Table 2). For self-reported oral health, female participants were significantly more likely to rate their oral health as good (n=98, 73.1%) or average (n=125, 60.1%), while poor oral health was significantly slightly more common among male participants (n=20, 44.4%; *P*=.02).

Table 2. Engagement with oral health information and self-reported oral health described by sex among adults aged 15 years and older residing in Saudi Arabia (n=387)[a].

Characteristics	Overall (n=387, 100%), n (%)	Male (n=139, 35.9%), n (%)	Female (n=248, 64%), n (%)	*P* value
Engagement with oral health information				
Use of social media to search for oral health information	227 (58.6)	71 (31.2)	156 (68.7)	.02
Ever received oral health information from nondental professionals on social media	235 (60.7)	86 (36.6)	149 (63.4)	.73
Following dental-specific accounts on social media	175 (45.2)	49 (28)	126 (72)	.003
Trust in oral health information from social media	171 (44.1)	54 (31.5)	117 (68.4)	.11
Social media profile content influence on trust in oral health information	289 (74.6)	105 (36.3)	184 (63.6)	.77
Perception of social media as a good source for oral health information	349 (90.1)	126 (36.1)	223 (63.9)	.82
Self-reported oral health				.02
Good	134 (34.6)	36 (26.8)	98 (73.1)	
Average	208 (53.7)	83 (39.9)	125 (60.1)	
Poor	45 (11.6)	20 (44.4)	25 (55.5)	

[a]Data were collected in a cross-sectional survey on oral health misconceptions conducted from September to November 2024.

Patterns of Social Media Use

The patterns of social media use are summarized in Table 3. Of the total sample, the most frequently used platform on a daily basis was WhatsApp (n=344, 89.8%), followed by Instagram (n=304, 78.9%), Snapchat (n=239, 62%), and YouTube (n=196, 50.7%). The majority of our sample reported not using Line (n=374, 97.1%), Facebook (n=295, 76.4%), and LinkedIn (n=271, 70%). Telegram had a proportion of weekly users (n=134, 35.1%), while TikTok showed mixed use with 49.7% (n=192) reporting daily use and 31% (n=120) reporting no use.

XSL•FO
RenderX

Table 3. The patterns of social media use among adults aged 15 years and older residing in Saudi Arabia described by daily, weekly, monthly, and no use of commonly used platforms in Saudi Arabia[a].

Social media platforms	Daily use[b], n (%)	Weekly use[c], n (%)	Monthly use[d], n (%)	No use[e], n (%)
X (Twitter)	180 (46.8)	98 (25.5)	37 (9.6)	69 (17.9)
Snapchat	239 (62)	49 (12.7)	11 (2.8)	86 (22.3)
WhatsApp	344 (89.8)	32 (8.3)	2 (0.5)	5 (1.3)
Instagram	304 (78.9)	52 (13.5)	9 (2.3)	20 (5.1)
TikTok	192 (49.7)	61 (15.8)	13 (3.3)	120 (31)
Facebook	32 (8.2)	29 (7.5)	30 (7.7)	295 (76.4)
Telegram	146 (38.3)	134 (35.1)	51 (13.3)	50 (13.1)
YouTube	196 (50.7)	155 (40.1)	21 (5.4)	14 (3.6)
Line	4 (1)	5 (1.3)	2 (0.5)	374 (97.1)
LinkedIn	26 (6.7)	47 (12.1)	42 (10.8)	271 (70.2)

[a]Data were collected in a cross-sectional survey on oral health misconceptions conducted from September to November 2024.

[b]Daily use: users who use the platform once a day or more than once a day.

[c]Weekly use: users who use the platform once a week, 2-3 times a week, or 4-6 times a week. [d]Monthly use: users who use the platform once a month.

[e]No use: users who are not using the platform.

Oral Health Misconceptions

Table 4 highlights the prevalence of oral health misconceptions among participants. Overall, 96.9% (n=375) of participants reported having at least one or more of the identified misconceptions. When stratified by sex, 93.5% (n=130) of male and 98.8% (n=245) of female participants were found to hold at least one or more misconceptions. Common misconceptions included “During pregnancy, the baby absorbs calcium from the mother’s teeth and bones” (n=337, 87%) and “Pregnant women should only undergo dental treatment after childbirth” (n=245, 63.3%). Additionally, 59.9% (n=232) of participants believed that “You should not eat anything when going for tooth extraction.” Sex-based differences were observed for the misconception that “Using a hard-bristled toothbrush makes teeth whiter,” with male participants more likely to agree (n=38, 57.5%) than female participants (n=28, 43.4%; Table 4).

XSL•FO
RenderX

Table 4. The proportion of responses to common oral health misconceptions among adults aged 15 years and older residing in Saudi Arabia, described by sex[a].

Oral health misconceptions	Overall, n (%)	Male, n (%)	Female, n (%)
During pregnancy, the baby absorbs calcium from the mother's teeth and bones	337 (87)	108 (32)	229 (67.9)
Pregnant women should only undergo dental treatment after childbirth	245 (63.3)	89 (36.3)	156 (63.6)
You should not eat anything when going for tooth extraction	232 (59.9)	90 (38.7)	142 (61.2)
Scaling weakens the structure of the teeth	224 (57.8)	80 (35.7)	144 (64.2)
Upper teeth extraction affects the brain	218 (56.3)	66 (30.2)	152 (69.7)
Brushing with salt helps whiten teeth	183 (47.2)	63 (34.4)	120 (65.5)
It is better not to brush your teeth when you have bleeding gums	157 (40.5)	69 (43.9)	88 (56)
Extracted teeth do not need to be replaced with artificial teeth	146 (37.7)	59 (40.4)	87 (59.5)
Taking antibiotics will relieve the tooth pain	137 (35.4)	50 (36.6)	87 (63.5)
If there is no tooth pain, there is no need to visit the dentist	130 (33.5)	51 (39.2)	79 (60.7)
Leaving a milk bottle in the baby's mouth during sleep does not harm teeth	114 (29.4)	53 (46.4)	61 (53.5)
The appearance of wisdom teeth increases one's wisdom	82 (21.1)	31 (37.8)	51 (62.2)
Using a hard-bristled toothbrush makes teeth whiter	66 (17)	38 (57.5)	28 (43.4)
There is no need to care for primary teeth, as they will be replaced by permanent teeth anyway	64 (16.5)	38 (59.3)	26 (40.6)
Having at least one or more of the above-mentioned misconceptions	375 (96.90)	130 (93.53)	245 (98.79)

[a]Data were collected in a cross-sectional survey on oral health misconceptions conducted from September to November 2024.

Association Between Engagement With Oral Health Information and Self-Reported Oral Health With Oral Health Misconceptions

Logistic and Poisson regression analyses (Table 5) revealed significant associations between following dental-specific accounts on social media with lower odds of having any misconceptions (AOR 0.41, 95% CI 0.22-0.78) and a lower count of misconceptions (AMR 0.87, 95% CI 0.77-0.98). Trust in social media for oral health information was significantly associated with a higher count of misconceptions (AMR 1.16, 95% CI 1.02-1.31). Self-reported oral health showed that participants with poor oral health had a higher count of misconceptions (AMR 1.23, 95% CI 0.99-1.50), although the association was not statistically significant. After conducting the sensitivity analysis by combining "I don't know" with "agree" as a single variable, there was no difference in most of the observations, except that participants who perceived social media as a good source of oral health information had a statistically significant higher count of misconceptions (AMR 1.18, 95% CI 1.02-1.37). This reclassification suggests that uncertainty about oral health information may reflect susceptibility to misconception.

XSL•FO
RenderX

Table 5. Logistic and Poisson regression models of the association between engagement with oral health information and self-reported oral health with the presence and count of oral health misconceptions among adults aged 15 years and older residing in Saudi Arabia[a].

Characteristics	Any misconceptions		Number of misconceptions	
	Crude OR[b] (95% CI)	Adjusted OR (95% CI)	Crude MR[c] (95% CI)	Adjusted MR (95% CI)
Engagement with oral health information				
Use of social media to search for oral health information				
No	reference	reference	reference	reference
Yes	1.20 (0.67-2.16)	1.07 (0.59-1.97)	1.12 (0.99-1.28)	1.12 (0.99-1.28)
Ever received oral health info from nondental professionals on social media				
No	reference	reference	reference	reference
Yes	1.33 (0.74-2.39)	1.33 (0.73-2.44)	1.01 (0.89-1.15)	1.00 (0.89-1.14)
Following dental-specific accounts on social media				
No	reference	reference	reference	reference
Yes	0.49 (0.27-0.89)[d]	0.41 (0.22-0.78)[d]	0.85 (0.76-0.97)[d]	0.87 (0.77-0.98)[d]
Trust in oral health information from social media				
No	reference	reference	reference	reference
Yes	1.81 (0.98-3.35)	1.68 (0.89-3.17)	1.16 (1.02-1.31)[d]	1.16 (1.02-1.31)[d]
Social media profile content influence on trust in oral health information				
No	reference	reference	reference	reference
Yes	1.47 (0.79-2.77)	1.73 (0.89-3.35)	1.10 (0.95-1.28)	1.12 (0.97-1.30)
Perception of social media as a good source for oral health information				
No	reference	reference	reference	reference
Yes	0.95 (0.35-2.55)	0.93 (0.34-2.58)	1.21 (0.97-1.51)	1.20, (0.96-1.50)
Self-reported oral health				
Good	reference	reference	reference	reference
Average	1.31 (0.71-2.42)	1.33 (0.70-2.53)	1.17 (1.01-1.34)	1.14 (0.99-1.31)
Poor	2.01 (0.65-6.19)	2.11 (0.66-6.76)	1.26 (1.03-1.54)	1.23 (0.99-1.50)

[a]Data were collected in a cross-sectional survey on oral health misconceptions conducted from September to November 2024.

[b]OR: odds ratio.

[c]MR: mean ratio. Adjusted for sex, education, and nationality.

[d]Indicates statistical significance (P<.05).

Discussion

Main Findings

This study investigated the patterns of social media use and the prevalence of oral health misconceptions among Saudi adults, focusing on the relationship between engagement with oral health information on social media and self-reported oral health with oral health misconceptions. The study highlighted significant associations between following dental-specific accounts and lower oral health–related misconceptions, while higher trust in social media information was associated with a higher count of oral health–related misconceptions. Additionally, self-reported poor oral health was associated with higher counts of oral health–related misconceptions.

A significant portion of our sample consisted of students and women. The high proportion of students in our sample reflects that students are known to be highly active on social media and rely on these platforms for information seeking, including health-related information. This aligns with findings from a scoping review indicating that social media plays a dominant role in health education in Saudi Arabia and that younger individuals and women frequently seek medical and health-related advice through digital platforms [42].

The results of our targeted social media campaigns reveal distinct differences in audience reach and interaction across various platforms. TikTok, in particular, demonstrated a significant advantage in terms of reach. This aligns with previous studies that highlight TikTok's rapid rise as a dominant platform for user interaction, especially among younger audiences [43,44]. Instagram, known for its visual-centric content, also proved effective in reaching a wide audience and generating meaningful interactions in this study. Instagram has been shown to be effective for health communication because

XSL•FO
RenderX

of its ability to attract users through images, videos, and infographics, which enhance user interaction and retention [45]. In contrast, X (Twitter) and Snapchat showed lower interaction rates, which can be explained by their platform-specific characteristics. In contrast, both X (Twitter), with its focus on quick updates and real-time communication, and Snapchat, with its temporary content format appealing to younger users seeking informal and immediate interactions, limit their potential for sustained and broad interaction [46].

TikTok's effectiveness in disseminating health information can be attributed to its short-form video format, which caters to modern users' preference for quick, easily digestible content [47]. This format has not only captured the attention of younger audiences but also influenced other social media platforms to adopt similar features. The platform's popularity among younger demographics further enhances its role in health communication, as it serves as a primary source of information for this age group. Moreover, TikTok's participatory features, such as duets and challenges, encourage user participation and the viral spread of health-related content [48].

Our methodology assumed that using social media advertisement campaigns would effectively reach our target population, with advertisement reach aligning with the demographic distribution in Saudi Arabia, where men represent a larger proportion (61.2%) [49]. The term "ad recipient" refers specifically to individuals who were exposed to the advertisement (ie, saw the advertisement) without necessarily engaging with it by clicking. Men comprised the majority of advertisement recipients across all platforms: Snapchat (n=28,235, 56%), TikTok (n=625,665, 74.32%), Instagram (n=24,625, 57.2%), and X (Twitter; n=107,116, 61%), similar to the male demographic distribution in Saudi Arabia [49]. However, despite greater advertisement exposure among men, women represented the majority of questionnaire respondents. This difference in interaction may be explained by behavioral differences, as previous studies have found that women are more likely to interact with health-related advertisements and content on social media than men [50-52]. This trend may be driven by greater health awareness among women, a higher likelihood of managing family health matters, and cultural norms that encourage women to be more involved in health-related decision-making [53,54].

This study identified several common oral health–related misconceptions. The most prevalent misconceptions include the belief that pregnant women lose calcium during pregnancy, dental procedures should be postponed during pregnancy, eating before dental extractions is harmful, and scaling weakens tooth structure. These findings are consistent with earlier studies conducted in Saudi Arabia, which also documented widespread myths about oral health [26-28].

Many of these oral health misconceptions can negatively impact oral health by discouraging preventive behaviors and delaying necessary treatment. For example, myths about pregnancy and oral health, such as the belief that "Pregnancy causes calcium loss in teeth" or that "Dental treatment should be avoided during pregnancy," may lead to untreated dental issues, increasing the risk of adverse pregnancy outcomes like preterm birth and low birth weight [55]. Similarly, misconceptions about primary teeth being unimportant contribute to early childhood caries, which affects children's overall well-being and increases the likelihood of future dental problems [56]. Additionally, beliefs in ineffective home remedies, such as using salt for whitening or avoiding professional scaling, can lead to poor oral hygiene and a higher risk of periodontal disease [57,58]. Based on these observations, addressing these common oral health–related misconceptions through targeted public health campaigns and education is essential. Engaging the community to participate in the educational content, primarily through social media platforms, would be an active way of enhancing people's awareness.

The high prevalence of misconceptions underscores the challenges in addressing oral health information in the digital age. While social media platforms provide an avenue for disseminating information, their effectiveness depends on the quality and credibility of the content shared.

This study observed associations between trust in social media for oral health information and higher misconceptions are consistent with broader concerns about the role of social media in propagating misconceptions in health communication [59]. The user-driven nature of social media echo chambers, combined with algorithm-driven content personalization, can reinforce existing beliefs, making users more susceptible to content that aligns with their biases and less likely to evaluate or seek out alternative perspectives critically. Social media algorithms on platforms such as TikTok, Instagram, and Facebook select content based on users' past interactions, prioritizing posts that align with their preference history. As a result, users who have previously interacted with nonevidence-based health claims are more likely to be exposed to similar misconceptions [60,61]. Conversely, the protective effect of following dental-specific accounts aligns with prior studies that emphasize the importance of verified content from credible sources in mitigating health misconceptions [62-65]. On social media, accounts maintained by professionals or organizations with expertise in a given field act as anchors of credibility, reducing the likelihood that users will engage with inaccurate information [59].

Strengths and Limitations of the Study

Our research specifically investigates oral health–related misconceptions in Saudi Arabia, an area understudied despite social media's growing influence in shaping health beliefs. This study's multiplatform social media approach enabled a comprehensive analysis of social media's role in oral health information. The use of targeted advertisements across multiple platforms, including Snapchat, TikTok, Instagram, and X (Twitter), allowed for the random selection of participants, enhancing the representativeness of the sample. We provided robust insights into the associations between engagement behaviors and self-reported oral health with oral health misconceptions by using advanced statistical modeling, including logistic and Poisson regression analyses. However, the cross-sectional design limits the ability to establish temporal relationships between the variables. Moreover, the reliance on self-reported data introduces potential response and recall biases, as participants may have overstated their engagement with oral health information or underreported misconceptions due to

XSL•FO
RenderX

social desirability. Additionally, recall bias could have affected their ability to recall their pattern of social media use accurately. Furthermore, while we adjusted for key confounders, unmeasured variables may have influenced both the exposures and outcomes, potentially affecting the observed associations.

Implication of the Study

Our findings align with previous research that highlights social media's role as a valuable channel for disseminating health information and as a potential source of misconceptions [19,66]. Therefore, the highly used platforms like WhatsApp, Instagram, Snapchat, and YouTube, as observed in this study, highlight their potential as valuable tools for web-based dental public health campaigns in Saudi Arabia. These platforms are widely used daily, reflecting their integration into the community's digital habits and their capacity to reach broad and diverse audiences. To address oral health–related misconceptions, health authorities, dental professionals, and regulatory bodies should consider the following:

- Developing social media–targeted campaigns that collaborate with credible, well-followed dental professional creators to disseminate accurate and engaging oral health information. This can include posting participatory videos, photos, and Q and A sessions to address common oral health–related misconceptions identified in our findings. Use social media features tailored to user interaction—such as TikTok's targeting for reaching young adults, Instagram Reels for quick educational content, Snapchat Stories for oral health tips, and X (Twitter) threads to counter oral health–related misconceptions.
- Establishing Saudi health authorities verified social media pages that consistently share reliable oral health information, combat misconceptions, and engage with users. These dental-specific accounts should interact with trending dental topics to redirect audiences toward credible sources.
- Legalizing and monitoring both oral health–related advertisements and information in social media from Saudi regulatory bodies would help protect public health and ensure that individuals receive evidence-based information in social media. For example, the European Union's Digital Services Act mandates that web-based platforms prevent the spread of misinformation, including health-related content, by enforcing fact-checking measures and transparency requirements for web-based advertisements [67].

While this study provides valuable insights, future research is needed to investigate how different content formats (eg, videos, photos, and participatory posts) and features (eg, likes, comments, and shares) influence the dissemination and retention of oral health information on social media. Experimental studies can assess which formats are most effective in improving knowledge and correcting oral health–related misconceptions, while longitudinal research can explore how repeated exposure to accurate content impacts behavioral change over time.

Conclusions

The study underscores the role of social media in shaping oral health perceptions among adults in Saudi Arabia. Engagement with dental-specific accounts was associated with fewer misconceptions, while trust in oral health information in social media was associated with higher misconceptions. Platforms like WhatsApp, Instagram, TikTok, and Snapchat are widely used in Saudi Arabia and offer the potential for targeted dental public health campaigns. Strategic collaborations with credible dental professional creators to disseminate accurate and engaging oral health information. Creating Instagram reels for quick educational content or using TikTok's targeting features could enhance information retention and user interaction. Although this study focuses on Saudi Arabia, its findings have broader relevance, particularly in regions with high social media penetration and similar digital health-seeking behaviors, such as the Gulf countries. From a policy perspective, legalizing and monitoring oral health–related advertisements and information in social media should be implemented by regulatory bodies to protect public health.

Future research should build on these findings to investigate the effectiveness of varying social media platforms' content formats and features in improving oral health knowledge and correcting misconceptions through experimental studies. Longitudinal research should also explore the long-term impact of repeated exposure to accurate information on behavioral change. These insights provide a basis for stakeholders to formulate evidence-based, innovative strategies to improve public health in the context of the digital age.

Acknowledgments

The authors would like to thank the Deanship of Scientific Research at King Saud University for funding and supporting this research through the initiative of DSR Graduate Students Research Support. In addition, we would like to express our sincere gratitude to Muath Aldosari, BDS, MPH, DMSc, for his invaluable guidance throughout the analysis phase of this study. His expertise and willingness to provide feedback during multiple consultations were instrumental in navigating the complexities of data analysis.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to ethical restrictions and commitments that mandate strict confidentiality of participant information and prohibit the use of data for purposes beyond the approved research study outlined by the institutional review board at King Saud University, but are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

RHB conceived and designed the study, collected and analyzed the data, and wrote the manuscript draft. SAA contributed to the research design, methodology, results, and writing process. AZG contributed to the methodology, results, and discussion sections and edited the manuscript. AHA supervised the study, provided guidance on the research idea, and contributed to the research methodology, results, and discussion. All authors reviewed and revised the manuscript and approved the final version for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire used in the study to assess social media usage patterns, engagement with oral health information, self-reported oral health, oral health misconceptions, and sociodemographic characteristics among adults in Saudi Arabia.
[DOCX File , 22 KB - formative_v9i1e70071_app1.docx]

References

1. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the internet and e-mail for health care information: results from a national survey. JAMA 2003;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]
2. Vuori M. Exploring uses of social media in a global corporation. J Syst Inf Technol 2012;14(2):155-170. [doi: 10.1108/13287261211232171]
3. Gottfried J. Americans' social media use. Pew Research Center. 2024. URL: https://www.pewresearch.org/internet/2024/01/31/americans-social-media-use/ [accessed 2024-01-31]
4. Grajales FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. J Med Internet Res 2014;16(2):e13 [FREE Full text] [doi: 10.2196/jmir.2912] [Medline: 24518354]
5. Smailhodzic E, Hooijsma W, Boonstra A, Langley DJ. Social media use in healthcare: a systematic review of effects on patients and on their relationship with healthcare professionals. BMC Health Serv Res 2016;16(1):442 [FREE Full text] [doi: 10.1186/s12913-016-1691-0] [Medline: 27562728]
6. Saudi internet usage report 2023. Saudi Communications, Space, and Technology Commission. 2023. URL: https://cst.gov.sa [accessed 2024-11-20]
7. de Oliveira Júnior AJ, Oliveira JM, Bretz YP, Mialhe FL. Online social networks for prevention and promotion of oral health: a systematic review. Can J Dent Hyg 2023;57(2):83-97 [FREE Full text] [Medline: 37464997]
8. Sivrikaya EC, Yilmaz O, Sivrikaya P. Dentist-patient communication on dental anxiety using the social media: a randomized controlled trial. Scand J Psychol 2021;62(6):780-786. [doi: 10.1111/sjop.12769] [Medline: 34333783]
9. Şallı GA, Egil E. Are YouTube videos useful as a source of information for oral care of leukemia patients? Quintessence Int 2020;51(1):78-85. [doi: 10.3290/j.qi.a43665] [Medline: 31781693]
10. Aboalshamat K, Alharbi J, Alharthi S, Alnifaee A, Alhusayni A, Alhazmi R. The effects of social media (Snapchat) interventions on the knowledge of oral health during pregnancy among pregnant women in Saudi Arabia. PLoS One 2023;18(2):e0281908 [FREE Full text] [doi: 10.1371/journal.pone.0281908] [Medline: 36795745]
11. Deghatipour M, Ghorbani Z, Mokhlesi AH, Ghanbari S, Namdari M. Effect of oral health promotion interventions on pregnant women dental caries: a field trial. BMC Oral Health 2022;22(1):280 [FREE Full text] [doi: 10.1186/s12903-022-02292-1] [Medline: 35804346]
12. Zotti F, Dalessandri D, Salgarello S, Piancino M, Bonetti S, Visconti L, et al. Usefulness of an app in improving oral hygiene compliance in adolescent orthodontic patients. Angle Orthod 2016;86(1):101-107 [FREE Full text] [doi: 10.2319/010915-19.1] [Medline: 25799001]
13. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
14. Househ M, Borycki E, Kushniruk A. Empowering patients through social media: the benefits and challenges. Health Informatics J 2014;20(1):50-58 [FREE Full text] [doi: 10.1177/1460458213476969] [Medline: 24550564]
15. De Martino I, D'Apolito R, McLawhorn AS, Fehring KA, Sculco PK, Gasparini G. Social media for patients: benefits and drawbacks. Curr Rev Musculoskelet Med 2017;10(1):141-145 [FREE Full text] [doi: 10.1007/s12178-017-9394-7] [Medline: 28110391]
16. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
17. Sumayyia MD, Al-Madaney MM, Almousawi FH. Health information on social media. Perceptions, attitudes, and practices of patients and their companions. Saudi Med J 2019;40(12):1294-1298 [FREE Full text] [doi: 10.15537/smj.2019.12.24682] [Medline: 31828284]

XSL•FO
RenderX

18. Office of the Surgeon General (OSG). Confronting health misinformation: The U.S. Surgeon General's Advisory on building a healthy information environment. Washington (DC): US Department of Health and Human Services 2021:22 [FREE Full text] [Medline: 34283416]
19. Chou WYS, Gaysynsky A, Cappella JN. Where we go from here: health misinformation on social media. Am J Public Health 2020;110(S3):S273-S275. [doi: 10.2105/AJPH.2020.305905] [Medline: 33001722]
20. Lotto M, Jorge O, Machado MDA, Cruvinel T. Exploring online oral health misinformation: a content analysis. Braz Oral Res 2023;37:e049 [FREE Full text] [doi: 10.1590/1807-3107bor-2023.vol37.0049] [Medline: 37255069]
21. Bidmon S, Terlutter R. Gender differences in searching for health information on the internet and the virtual patient-physician relationship in Germany: exploratory results on how men and women differ and why. J Med Internet Res 2015;17(6):e156 [FREE Full text] [doi: 10.2196/jmir.4127] [Medline: 26099325]
22. Warner-Søderholm G, Bertsch A, Sawe E, Lee D, Wolfe T, Meyer J, et al. Who trusts social media? Comput Hum Behav 2018;81:303-315. [doi: 10.1016/j.chb.2017.12.026]
23. Stellefson M, Hanik B, Chaney B, Chaney D, Tennant B, Chavarria EA. eHealth literacy among college students: a systematic review with implications for eHealth education. J Med Internet Res 2011;13(4):e102 [FREE Full text] [doi: 10.2196/jmir.1703] [Medline: 22155629]
24. Pop LM, Iorga M, ipo LR, Iurcov R. Gender differences in healthy lifestyle, body consciousness, and the use of social networks among medical students. Medicina (Kaunas) 2021;57(7):648 [FREE Full text] [doi: 10.3390/medicina57070648] [Medline: 34202497]
25. Ruiz M, Kabani F, Cotter J. A review of the effects of oral health media hype on clients' perception of treatment. Can J Dent Hyg 2022;56(1):31-38 [FREE Full text] [Medline: 35401762]
26. Ajmal M, Assiri K, Al-Ahmari B. Oral hygiene myths and its association with gingival health status among patients in Aseer region of Saudi Arabia: a cross-sectional study. J Contemp Dent Pract 2021;22(5):506-510. [doi: 10.5005/jp-journals-10024-3062]
27. Al-Harthi MM, Holenarasipur AV, Al-Harthi HI, Al-Thobaiti FA, Al-Otaibi HA, Al-Dawood MM. Prevalence of myths regarding oral health, among patients attending a dental institute: a study from Taif, Saudi Arabia. JCDR 2019;13(9). [doi: 10.7860/jcdr/2019/41907.13125]
28. Gowdar IM, Alqahtani AM, Asiri AM, Aldossary SF, Alkhurayef IA, Alheneshi DI. Oral health myths among general population at Riyadh region, Saudi Arabia. J Pharm Bioallied Sci 2021;13(Suppl 1):S241-S245 [FREE Full text] [doi: 10.4103/jpbs.JPBS_700_20] [Medline: 34447085]
29. Duangthip D, Chu CH. Challenges in oral hygiene and oral health policy. Front Oral Health 2020;1:575428 [FREE Full text] [doi: 10.3389/froh.2020.575428] [Medline: 35047981]
30. Population estimates by gender, nationality, and region. General Authority for Statistics (GASTAT). 2022. URL: https://database.stats.gov.sa/home/indicator/535 [accessed 2024-11-20]
31. Manual of statistical definitions, concepts, and terms. General Authority for Statistics (GASTAT). 2024. URL: https://www.stats.gov.sa/sites/default/files/statistical_definitions_manual_english_2.pdf [accessed 2024-11-20]
32. Alhaddad MS. The use of social media among Saudi residents for medicines related information. Saudi Pharm J 2018;26(8):1106-1111 [FREE Full text] [doi: 10.1016/j.jsps.2018.05.021] [Medline: 30510470]
33. Almozainy M. Assessing the use of social media as a source of information related to dentistry in Saudi Arabia. JDHODT 2017;8(7):663-668. [doi: 10.15406/jdhodt.2017.08.00312]
34. Baik KM, Anbar G, Alshaikh A, Banjar A. Effect of social media on patient's perception of dental aesthetics in Saudi Arabia. Int J Dent 2022;2022:4794497 [FREE Full text] [doi: 10.1155/2022/4794497] [Medline: 35265132]
35. Barrera-Verdugo G, Villarroel-Villarroel A. Evaluating the relationship between social media use frequency and entrepreneurial perceptions and attitudes among students. Heliyon 2022;8(4):e09214 [FREE Full text] [doi: 10.1016/j.heliyon.2022.e09214] [Medline: 35434392]
36. Althunayan A, Alsalhi R, Elmoazen R. Role of social media in dental health promotion and bahavior change in qassim province, Saudi Arabia. Int J Med Health Res 2018;4:98-103 [FREE Full text]
37. ALrukban A. The health related uses of social media among users in Saudi Arabia. Int J Med Sci Public Health 2014;3(12):1492-1497. [doi: 10.5455/ijmsph.2014.170920141]
38. Lawal FB. Global self-rating of oral health as summary tool for oral health evaluation in low-resource settings. J Int Soc Prev Community Dent 2015;5(Suppl 1):S1-S6 [FREE Full text] [doi: 10.4103/2231-0762.156516] [Medline: 25984461]
39. Population by age groups and gender, mid-year 2020. General Authority for Statistics (GASTAT). 2020. URL: https://www.stats.gov.sa/sites/default/files/Population%20by%20Age%20Groups%20%2Cand%20Gender_0.pdf [accessed 2024-11-20]
40. National qualifications framework 2023: the national qualifications framework in the kingdom of Saudi Arabia. Education and Training Evaluation Commission (ETEC). 2023. URL: https://www.etec.gov.sa [accessed 2024-11-20]
41. Butani Y, Weintraub JA, Barker JC. Oral health-related cultural beliefs for four racial/ethnic groups: assessment of the literature. BMC Oral Health 2008;8:26 [FREE Full text] [doi: 10.1186/1472-6831-8-26] [Medline: 18793438]
42. El Kheir DYM, Boumarah DN, Bukhamseen FM, Masoudi JH, Boubshait LA. The Saudi experience of health-related social media use: a scoping review. Saudi J Health Syst Res 2021;1(3):1-12. [doi: 10.1159/000516473]

XSL•FO
RenderX

43. Lartey S. The future influence of TikTok on evolving social media trends: a comprehensive analysis. 2024. URL: https://tinyurl.com/5derawx6 [accessed 2024-12-12]
44. Gesmundo MAG, Jordan MDS, Meridor WHD, Muyot DV, Castano MCN, Bandojo AJP. TikTok as a platform for marketing campaigns: the effect of brand awareness and brand recall on the purchase intentions of millennials. JBMS 2022;4(2):343-361. [doi: 10.32996/jbms.2022.4.2.27]
45. Malik A, Khan ML, Quan-Haase A. Public health agencies outreach through instagram during the COVID-19 pandemic: crisis and emergency risk communication perspective. Int J Disaster Risk Reduct 2021;61:102346 [FREE Full text] [doi: 10.1016/j.ijdrr.2021.102346] [Medline: 36337987]
46. Alhabash S, Ma M. A tale of four platforms: motivations and uses of Facebook, Twitter, Instagram, and Snapchat among college students? Social Media + Society 2017;3(1):205630511769154. [doi: 10.1177/2056305117691544]
47. Sattora EA, Ganeles BC, Pierce ME, Wong R. Research on health topics communicated through TikTok: a systematic review of the literature. Journalism Media 2024;5(3):1395-1412. [doi: 10.3390/journalmedia5030088]
48. Basch CH, Hillyer GC, Jaime C. COVID-19 on TikTok: harnessing an emerging social media platform to convey important public health messages. Int J Adolesc Med Health 2022;34(5):367-369 [FREE Full text] [doi: 10.1515/ijamh-2020-0111] [Medline: 32776899]
49. Saudi census 2022: population summary report. General Authority for Statistics (GASTAT). URL: https://portal.saudicensus.sa/static-assets/media/content/20230531_GASTAT_Population_Report.pdf?crafterSite=gastat-portal [accessed 2024-11-19]
50. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? a systematic review. J Med Internet Res 2014;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
51. Ryan J, Lopian L, Le B, Edney S, Van Kessel G, Plotnikoff R, et al. It's not raining men: a mixed-methods study investigating methods of improving male recruitment to health behaviour research. BMC Public Health 2019;19(1):814 [FREE Full text] [doi: 10.1186/s12889-019-7087-4] [Medline: 31234825]
52. Asogwa CE, Okeke SV, Gever VC, Ezeah G. Gender disparities in the influence of social media advertisements on buying decision in Nigeria. Communicatio 2020;46(3):87-105 [FREE Full text] [doi: 10.1080/02500167.2020.1825100]
53. Courtenay WH. Key determinants of the health and well-being of men and boys. Int J Men's Health 2003;2(1):1-30. [doi: 10.3149/jmh.0201.1]
54. Addis ME, Mahalik JR. Men, masculinity, and the contexts of help seeking. Am Psychol 2003;58(1):5-14. [doi: 10.1037/0003-066x.58.1.5] [Medline: 12674814]
55. George A, Johnson M, Blinkhorn A, Ajwani S, Bhole S, Yeo AE, et al. The oral health status, practices and knowledge of pregnant women in South-Western Sydney. Aust Dent J 2013;58(1):26-33 [FREE Full text] [doi: 10.1111/adj.12024] [Medline: 23441789]
56. Anil S, Anand PS. Early childhood caries: prevalence, risk factors, and prevention. Front Pediatr 2017;5:157 [FREE Full text] [doi: 10.3389/fped.2017.00157] [Medline: 28770188]
57. Addy M, Hunter ML. Can tooth brushing damage your health? Effects on oral and dental tissues. Int Dent J 2003;53 Suppl 3:177-186. [doi: 10.1111/j.1875-595x.2003.tb00768.x] [Medline: 12875306]
58. Hujoel PP, Cunha-Cruz J, Banting DW, Loesche WJ. Dental flossing and interproximal caries: a systematic review. J Dent Res 2006;85(4):298-305. [doi: 10.1177/154405910608500404] [Medline: 16567548]
59. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccines Immunother 2020;16(11):2586-2593 [FREE Full text] [doi: 10.1080/21645515.2020.1780846] [Medline: 32693678]
60. Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, et al. The spreading of misinformation online. Proc Natl Acad Sci U S A 2016;113(3):554-559 [FREE Full text] [doi: 10.1073/pnas.1517441113] [Medline: 26729863]
61. Getman R, Helmi M, Roberts H, Yansane A, Cutler D, Seymour B. Vaccine hesitancy and online information: the influence of digital networks. Health Educ Behav 2018;45(4):599-606. [doi: 10.1177/1090198117739673] [Medline: 29267129]
62. Oehler RL. On measles, vaccination, social media activism, and how to win back our role as our patients' best advocates. Clin Infect Dis 2020;70(2):338-340. [doi: 10.1093/cid/ciz656] [Medline: 31309980]
63. Chen Q, Min C, Zhang W, Wang G, Ma X, Evans R. Unpacking the black box: how to promote citizen engagement through government social media during the COVID-19 crisis. Comput Human Behav 2020;110:106380 [FREE Full text] [doi: 10.1016/j.chb.2020.106380] [Medline: 32292239]
64. Basch CE, Basch CH, Hillyer GC, Jaime C. The role of YouTube and the entertainment industry in saving lives by educating and mobilizing the public to adopt behaviors for community mitigation of COVID-19: successive sampling design study. JMIR Public Health Surveill 2020;6(2):e19145 [FREE Full text] [doi: 10.2196/19145] [Medline: 32297593]
65. D'Souza RS, D'Souza S, Strand N, Anderson A, Vogt M, Olatoye O. YouTube as a source of medical information on the novel coronavirus 2019 disease (COVID-19) pandemic. Glob Public Health 2020;15(7):935-942 [FREE Full text] [doi: 10.1080/17441692.2020.1761426] [Medline: 32397870]

XSL•FO
RenderX

66. Adebesin F, Smuts H, Mawela T, Maramba G, Hattingh M. The role of social media in health misinformation and disinformation during the COVID-19 pandemic: bibliometric analysis. JMIR Infodemiol 2023;3:e48620 [FREE Full text] [doi: 10.2196/48620] [Medline: 37728981]
67. The EU's digital services act. European Commission. URL: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_en [accessed 2022-10-27]

Abbreviations

AMR: adjusted mean ratio
AOR: adjusted odds ratio
GASTST: General Authority of Statistics

Edited by A Mavragani; submitted 14.12.24; peer-reviewed by R Musa, A Noor; comments to author 07.01.25; revised version received 20.01.25; accepted 22.01.25; published 10.02.25.

Please cite as:
BinHamdan RH, Alsadhan SA, Gazzaz AZ, AlJameel AH
Social Media Use and Oral Health–Related Misconceptions in Saudi Arabia: Cross-Sectional Study
JMIR Form Res 2025;9:e70071
URL: https://formative.jmir.org/2025/1/e70071
doi:10.2196/70071
PMID:

©Rahaf Hamdan BinHamdan, Salwa Abdulrahman Alsadhan, Arwa Zohair Gazzaz, AlBandary Hassan AlJameel. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

mHealth App to Promote Healthy Lifestyles for Diverse Families Living in Rural Areas: Usability Study

Alejandra Perez Ramirez[1], MPH; Adrian Ortega[2], PhD; Natalie Stephenson[3], PhD; Angel Muñoz Osorio[1], BS; Anne Kazak[1,4], PhD; Thao-Ly Phan[1,4], MPH, MD

[1]Center For Healthcare Delivery Science, Nemours Children's Health, Wilmington, DE, United States

[2]Center for Behavior Intervention Technologies, Northwestern Feinberg School of Medicine, Northwestern University, Evanston, IL, United States

[3]Center for Health Delivery Innovation, Nemours Children's Health, Wilmington, DE, United States

[4]Department of Pediatrics, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA, United States

Corresponding Author:
Alejandra Perez Ramirez, MPH
Center For Healthcare Delivery Science
Nemours Children's Health
1600 Rockland Road
Wilmington, DE, 19803
United States
Phone: 1 302 358 5968
Email: alejandra.perezramirez@nemours.org

Abstract

Background: Mobile Integrated Care for Childhood Obesity is a multicomponent intervention for caregivers of young children with obesity from rural communities that was developed in collaboration with community, parent, and health care partners. It includes community programming to promote healthy lifestyles and address social needs and health care visits with an interdisciplinary team. A digital mobile health platform—the Healthy Lifestyle (Nemours Children's Health) dashboard—was designed as a self-management tool for caregivers to use as part of Mobile Integrated Care for Childhood Obesity.

Objective: This study aimed to improve the usability of the English and Spanish language versions of the Healthy Lifestyle dashboard.

Methods: During a 3-phased approach, usability testing was conducted with a diverse group of parents. In total, 7 mothers of children with obesity from rural communities (average age 39, SD 4.9 years; 4 Spanish-speaking and 3 English-speaking) provided feedback on a prototype of the dashboard. Participants verbalized their thoughts while using the prototype to complete 4 tasks. Preferences on the dashboard icon and resource page layout were also collected. Testing was done until feedback reached saturation and no additional substantive changes were suggested. Qualitative and quantitative data regarding usability, acceptability, and understandability were analyzed.

Results: The dashboard was noted to be acceptable by 100% (N=7) of the participants. Overall, participants found the dashboard easy to navigate and found the resources, notifications, and ability to communicate with the health care team to be especially helpful. However, all (N=4) of the Spanish-speaking participants identified challenges related to numeracy (eg, difficulty interpreting the growth chart) and literacy (eg, features not fully available in Spanish), which informed iterative refinements to make the dashboard clearer and more literacy-sensitive. All 7 participants (100%) selected the same dashboard icon and 71% (5/7) preferred the final resource page layout.

Conclusions: Conducting usability testing with key demographic populations, especially Spanish-speaking populations, was important to developing a mobile health intervention that is user-friendly, culturally relevant, and literacy-sensitive.

(JMIR Form Res 2025;9:e60495) doi:10.2196/60495

KEYWORDS

obesity; user testing; mHealth; mobile health; Spanish; child; rural population

XSL•FO
RenderX

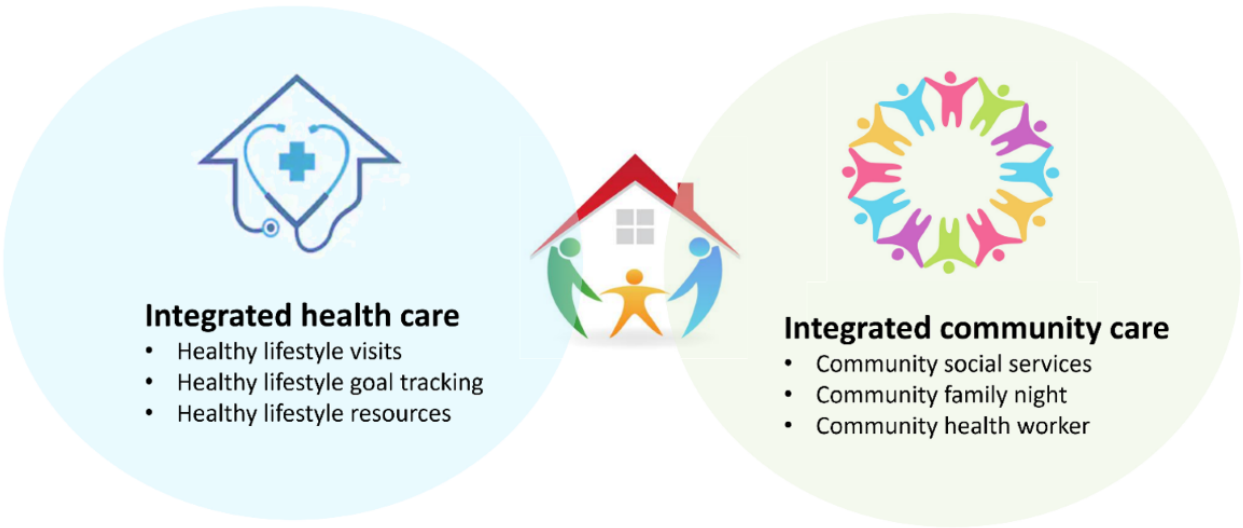
Introduction

Childhood obesity persists as a significant public health concern in the United States [1,2]. Notably, 1 in every 5 children in the United States lives with obesity. Obesity in childhood is linked to increased risks of comorbid and future health conditions such as diabetes, cardiovascular diseases, cancer, and psychosocial concerns, leading to increased health care costs and utilization [1,3]. This is especially the case for children from communities of color, communities with fewer resources, and rural communities, who continue to have the highest rates of obesity [2]. Challenges in affording and accessing nutritious food, accessing safe spaces for physical activity, having transportation to get to health care services, and managing family stressors are a few of the many systemic barriers that perpetuate disparities in childhood obesity [4,5]. While lifestyle behavior interventions are first-line treatments for improving health behaviors and obesity outcomes, most available interventions do not address these systemic barriers and are often not effective when implemented in the communities with the greatest need [6,7]. Thus, more studies are needed that consider the needs and preferences of diverse patient families when designing an intervention.

Digital interventions for the treatment of pediatric obesity can increase the convenience and access to treatment through the removal of common barriers such as financial, travel, and time constraints [8,9]. A recent systematic review and meta-analysis revealed that digital interventions for the treatment of pediatric obesity have a significant yet small effect on BMI *z* scores [10]. Still, there are important limitations to digital interventions that require consideration, especially regarding their uptake and implementation. For one, digital interventions are subject to high nonusage attrition such as low uptake and disengagement [11,12]. Although some digital interventions for the treatment of pediatric obesity have included diverse samples in their studies, there remain concerns about their cultural sensitivity [10]. For example, most of these interventions do not include content related to systemic barriers, social drivers of health, or community resources [13]. Rurality or place of residence is also not often reported or considered when implementing digital interventions [14]. In addition, there is only 1 published study of a digital intervention for pediatric obesity available in Spanish, which is a major gap given the prevalence of Hispanic or Latino children with overweight or obesity in the United States [5,15,16]. Therefore, there is a need to incorporate human-centered design principles into the design of digital interventions, such as partnering with target users to enhance the usability of these interventions. This way, digital intervention designers can maximize engagement by building an intervention that centers on the needs and preferences of its users. Key principles of human-centered design include but are not limited to the active involvement of users with lived experiences relevant to the intervention in design activities as well as iteratively refining and updating intervention features based on user feedback [17].

Drawing on the advantages of using a digital intervention, a digital mobile health (mHealth) platform was included as part of a larger, multicomponent intervention called Mobile Integrated Care for Childhood Obesity (MICCO). MICCO was designed for caregivers of children with obesity from rural communities (Figure 1). MICCO combines evidence-based interdisciplinary obesity treatment, monthly community-based programming for families to engage in physical activity and have a healthy affordable meal together, and community health worker outreach to address psychosocial needs and barriers to achieving lifestyle goals. Importantly, MICCO was developed with the input of a Council of Parent Partners, which included 6 parents of children with obesity from rural communities, and a Council of Community Partners, which included 5 primary care providers from clinics serving rural communities and 4 community organizations serving rural communities. These Councils met routinely by videoconference throughout the entirety of the study to provide valuable feedback on the development and implementation of the MICCO intervention, drawing on their knowledge about the needs of families of children with obesity and resources in the community to support families.

Figure 1. Mobile Integrated Care for Childhood Obesity (MICCO) care model, developed by a research team for families of young children with obesity.



XSL•FO
RenderX

As an additional component of MICCO, a custom Healthy Lifestyle dashboard on a mobile app linked to the Nemours Children's Health electronic health record (EHR; Nemours App [Nemours Children's Health]) was built, with input from interdisciplinary specialists from the Nemours Healthy Weight and Wellness Clinic and Council of Community Partners and in collaboration with app designers from the Nemours Center for Health Delivery Innovation. Psychoeducational tools based on social cognitive theory [18], behavior change techniques [19], and the socioecological model of obesity [20] were included on the dashboard, including a self-monitoring function to track daily progress toward lifestyle goals formed during treatment visits and culturally relevant videos to promote positive parenting and healthy lifestyle behaviors. The video resources, including 6 that were created specifically for the MICCO intervention, provided content related to mood and behavior, physical activity, and healthy eating; were available in English and Spanish; and featured families and health care providers of diverse backgrounds. Through the dashboard, caregivers can also track their child's weight using an integrated Bluetooth scale, send messages to their health care team, and connect to telehealth appointments. The dashboard also included links to community programs (including healthy lifestyle and social need resources) and monthly reminders about the monthly community-based program. Since many of the patient families with obesity served by Nemours Children's Health are Hispanic or Latino, the dashboard was developed in English and Spanish. A parallel dashboard was created for health care providers to access data entered by patient families about their goals and weight progress to assist with clinical decision support.

User testing was conducted to test and refine the MICCO Healthy Lifestyle dashboard to enhance the usability of the intervention for families of diverse backgrounds. No hypotheses were made a priori.

Methods

Overview

Parents of children with obesity living in rural communities provided feedback on the usability of the Healthy Lifestyle dashboard through Think Aloud testing. Think Aloud testing is a research method applied to an iterative human-centered design process to discover usability problems with a product, incorporate feedback, revise the product, and retest to validate the design. This evidence-based method for obtaining in vivo feedback has been used in the development of digital health interventions [21]. In a typical usability study, the most common usability problems are uncovered after testing with 5 participants [22].

Sample and Recruitment

Target users of the MICCO intervention are legal guardians of children ages 4-12 years with obesity (BMI percentile ≥95) receiving primary care from 1 of 3 Nemours Children's Health practices in Delaware serving rural communities. Rurality was determined based on a score of 4 or greater on the US Department of Agriculture 2010 Rural-Urban Commuting Area Codes [23]. Therefore, we recruited participants matching these characteristics for user testing. Legal guardians also had to be able to read English or Spanish and have access to the internet. A primary aim of the study was to incorporate diverse perspectives in refining the Healthy Lifestyle dashboard; therefore, purposeful sampling was used to recruit a diverse group of parents including Spanish-speaking parents of Hispanic or Latino ethnicity and English-speaking parents of Black and White race. A list of potentially eligible participants was pulled from the EHR of Nemours Children's Health, which is a large pediatric health care system with 12 primary care clinics in Delaware. A total of 196 parents (63 Spanish-speaking and 133 English-speaking) from this EHR list received an SMS text message informing them of the study. In total, 34 parents (15 Spanish-speaking and 19 English-speaking) responded to the text. A research coordinator then followed up with a call to discuss the study and procedures with each parent. Of these, 7 agreed to participate (4 Spanish-speaking and 3 English-speaking), 2 declined after hearing about the study, and the remainder did not respond to any additional contact by the study team. For those who agreed to participate, the research coordinator scheduled a Think Aloud session. Consent was collected at the time of the session before any data collection (refer to Ethical Considerations for details regarding the consent process).

Data Collection

Think Aloud testing was conducted via video calls through a secure Zoom (Zoom Communications) platform, accessible to participants either on their mobile device or computer. A web link to an interactive, high-fidelity prototype of the Healthy Lifestyle dashboard was shared with participants. The online prototype mirrored the app and allowed participants to easily access the design for Think Aloud testing. The sessions were done through a video call, similar to other studies [24], to increase the convenience for parents of young children and to simulate the environmental context within which the actual mobile app would be experienced. It also allowed for screen recording so the research coordinator could see which parts were easy or hard to navigate.

The testing sessions were video-recorded, and the audio was transcribed. Spanish language recordings were translated and transcribed into English. Rarely, participants encountered minor technical issues with Zoom, such as slow connectivity or difficulty using Zoom (eg, how to join the call or share their screen), but the bilingual research coordinator on the call was able to guide them through these issues.

Testing sessions were conducted using a semistructured interview guide (Multimedia Appendix 1) to facilitate the completion of tasks. Introductory questions were used to establish a connection with the participant before user testing of the dashboard. During the testing portion of the interview, participants were introduced to the dashboard at the time of testing so that first impressions could be observed. This also allowed for genuine reactions as they became familiar with the platform and navigated the tools. Participants were then asked to complete four major tasks, as laid out by the interview guide: (1) enter daily goal completion and review their weekly progress, (2) enter the child's height and weight and review a growth curve, (3) contact their health care team, and (4) find

XSL•FO
RenderX

resources for their family. Families were also asked about specific design layouts using preference testing (ie, participants were shown 2 design options and asked to pick the one they preferred). Throughout the testing, participants were encouraged to Think Aloud or verbalize their thoughts as they navigated independently through the tasks. The participants shared their screens with the research coordinator, which allowed the coordinator to see their navigation when using the app. Immediately following the testing portion, participants were asked closing questions to collect additional general feedback about the dashboard. Each session lasted 30 to 60 minutes.

A 3-phased approach was used to refine the dashboard and validate that iterative changes were acceptable. In phase 1, Think Aloud sessions with 4 Spanish-speaking families were conducted until no new feedback was received. The research and Nemours app team met to discuss feedback, and rapid changes were made to the user interface based on participant feedback. In phase 2, Think Aloud sessions with 3 English-speaking families were completed with the revised prototype until no substantive changes were recommended. In phase 3, a final prototype based on all of the Think Aloud feedback was shown in a video call to the larger study's Council of Parent Partners, which was started after user testing was complete. The Council of Parent Partners included 4 parents from phase 1, one parent from phase 2, and one parent who did not participate in Think Aloud testing. This meeting was not formal user testing. Instead, it presented an opportunity to obtain final approval from parents, including many who had participated as individuals in the user testing. Meetings were facilitated by a bilingual research coordinator, with whom the parents had developed a rapport, and participants received an additional US $50 gift card for participating. Parents were advised during the group meeting that they could also provide feedback on the final prototype privately if desired.

Data Analysis

All Think Aloud transcripts were uploaded into Dedoose [25], a qualitative data analysis software, to complete a formal qualitative analysis. A preliminary codebook of a priori codes was established based on the interview guide. The goal of the codebook was to capture the usability, understandability, and acceptability of the dashboard and its specific features during the Think Aloud sessions (Multimedia Appendix 2). Furthermore, 3 members of the research team (TLP, APR, and AO) met regularly to review the transcripts and codebook. The first transcript was coded by LP and APR together to test the codebook. All 3 members then independently coded 2 more transcripts. The team met to review discrepancies and revise the codebook as needed. The fourth transcript was used to test reliability among the coders and the established codebook. Given the strong intercoder reliability (83.5%), the codebook was finalized, and then all the transcripts were coded with the final codebook by APR and TLP. They met regularly to resolve any discrepancies. After all transcripts were coded, feedback for each of the 4 specific tasks and the overall app was pulled. The research team noted specific challenges and features highlighted by participants. This feedback was analyzed for each phase given that there were changes in-between phases. Quantitative data regarding the number of participants who reported overall acceptability of the dashboard and usability and understandability of each task were also extracted by coding specific mentions and counting the frequency of the codes.

Ethical Considerations

Considerations to protect participants were taken into account when designing study procedures. Before recruitment, the Nemours IRB (#1812070) approved all study procedures. During recruitment, families were given the option to decline participation or opt out of further communication. Interested participants were asked to complete a consent form using an e-consent process. At a convenient time for the participant, the research coordinator trained in collecting informed consent, called the participant on the phone to review the full consent form. The consent form listed the study procedures and risks and benefits involved with the study. Participants were reminded that their participation was voluntary and that they could ask to withdraw at any time. Spanish-speaking participants were consented by a native Spanish-speaking research coordinator and with an approved Spanish-language form. Study procedures, such as having participants turn their cameras off at the time of the Think Aloud session and removing all identifying information from transcripts, were followed to protect the participants' identities. All participants received a US $50 gift card after completing the Think Aloud session.

Results

In total, 7 mothers (average age 39, SD 4.9 years) participated in Think Aloud sessions (4 in phase 1 and 3 in phase 2) and 6 mothers (average age 42, SD 4.9 years, 5 of whom participated in Think Aloud testing) participated in the Council of Parent Partners meeting (phase 3, Table 1). No additional participants were recruited given that user feedback did not identify any new usability problems after 2 rounds of testing. These sample sizes are in line with a typical usability study [22].

Overall, MICCO was accepted and liked by 100% (N=7) of the participants. Participants shared that the Healthy Lifestyle dashboard would be an easy-to-use tool to help them develop healthier habits as a family. As 1 participant described:

> *It is a tool that is within my reach. I can contact my son's doctor to know what he can or cannot eat, and not have to make an appointment to go to the hospital. I can use the application to solve the doubts I have about some things that my son can or cannot get. For me, it would be an easy tool to handle as a mom.*

While it was generally well-received, participants shared some challenges with specific tasks and highlighted features that were important to them, as described further in this study.

XSL•FO
RenderX

Table 1. Demographic information of caregivers of children with obesity from rural clinics in Delaware, participating in 3 iterative phases of Think Aloud Testing of a mobile dashboard to promote healthy lifestyles.

Characteristics	Phase 1 (n=4)	Phase 2 (n=3)	Phase 3[a] (n=6)
Parent age in years, mean (SD)	38.8 (3.0)	40.3 (7.6)	42.0 (4.9)
Parent sex (female), n (%)	4 (100)	3 (100)	6 (100)
Race, n (%)			
White	2 (50)	1 (33)	2 (33)
Black	—[b]	—	—
Other[c]	2 (50)	1 (33)	3 (50)
More than 1 race	—	1 (33)	1 (17)
Ethnicity (Hispanic or Latino), n (%)	4 (100)	2 (67)	6 (100)
Insurance status (public)[c], n (%)	4 (100)	2 (100)[d]	5 (83)

[a]Five participants also participated in phases 1 and 2.

[b]Not applicable.

[c]"Other" was not specified by participants.

[d]Missing n=1; public insurance refers to a program run by the government to pay for some or all health care costs.

Phase 1

Overview

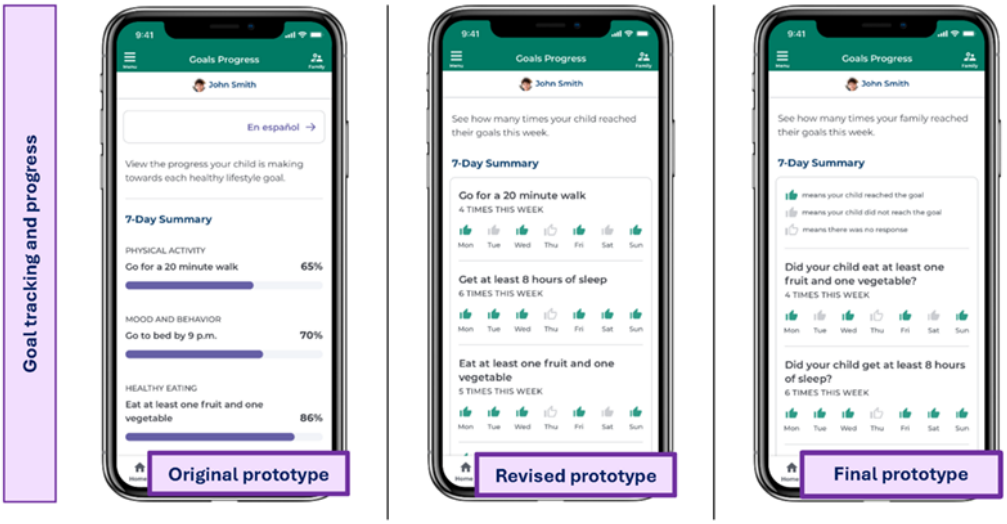
Spanish-speaking participants reported challenges surrounding understandability and usability of certain features of the Healthy Lifestyle dashboard. For example, 3 participants did not understand the lifestyle goal progress feature, 4 did not understand the growth curve feature, and 3 had trouble using the tool to communicate with the health care team. Themes surrounding numeracy and literacy challenges (Table 2) emerged and specific changes were made to the dashboard to ease these challenges after phase 1 (Figures 2 and 3, respectively).

Table 2. Challenges and solutions identified in Think Aloud testing of a mobile dashboard to promote healthy lifestyles by Spanish-speaking caregivers of children with obesity from rural clinics in Delaware.

Area and challenge	Solution
Numeracy challenges	
Interpretation of weekly goal completion	Changed the percentage bar to thumbs-up icons and provided a count instead of percentage
Interpretation of child's growth chart	Changed growth chart to show weight change over time instead of BMI
Literacy challenges	
Some components of the Nemours app (ie, the message center) were not available in Spanish	A pop-up note was added to provide the user with instructions in Spanish and a note that the following page would only be available in English
Despite the dashboard being in Spanish, there were concerns about the reading level	Language was simplified throughout the app

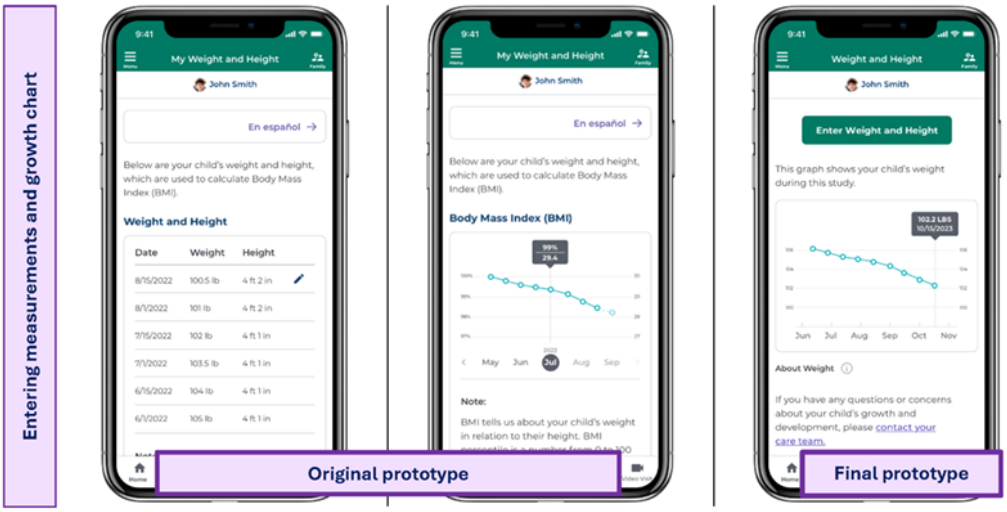
XSL•FO
RenderX

Figure 2. Revisions to dashboard to address numeracy challenges identified by 4 Spanish-speaking caregivers of children with obesity from rural clinics in Delaware during phase 1 of Think Aloud testing of a mobile app to promote healthy lifestyles.



"Why is 9:00 P.M… 70%? That's very – should it be earlier, or what? I don't understand." – Spanish-speaking participant

"It is telling me that what was done equals 65%, it is telling me that not 100% of what needs to be done has been done. The 70% is a little closer to the target, and healthy eating has a higher percentage than the other two questions that were asked during the day." – Spanish-speaking participant

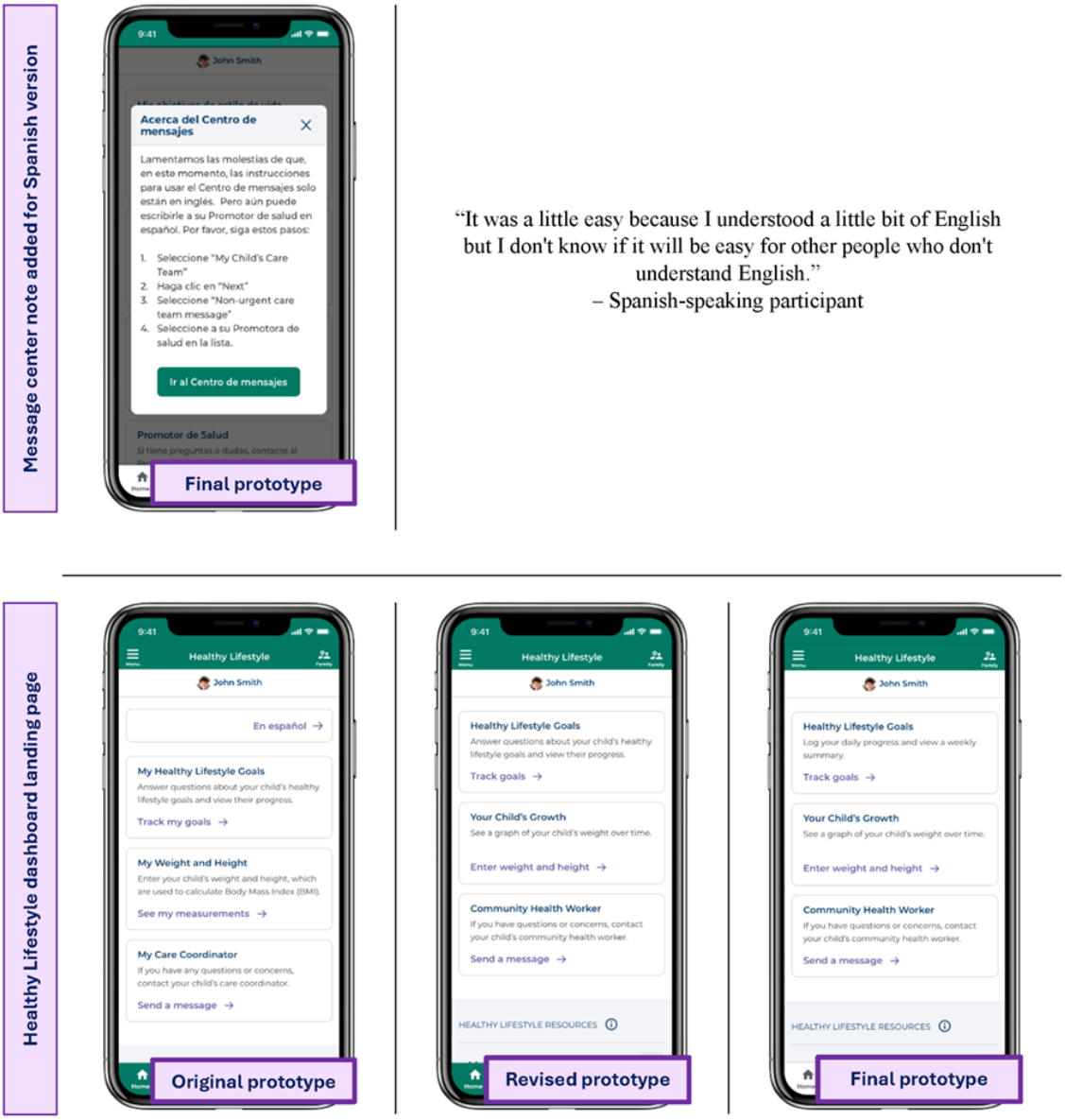


"Well, what I see is, I think the numbers – 100%. That's what I see. I don't know what it means – 28 and 100%. I don't know, maybe height, or weights, I don't know, maybe growth." – Spanish-speaking participant

"That I am reaching 99, maybe I am reaching the goal for which I am using the application. I don't understand what 29 is." – Spanish-speaking participant

XSL•FO
RenderX

Figure 3. Revisions to address literacy challenges experienced during phase 1 Think Aloud testing of a mobile app to promote healthy lifestyles by Spanish-speaking caregivers of children with obesity from rural clinics in Delaware.



Numeracy Challenges

All 4 participants reported numeracy issues when reviewing the lifestyle goal progress and growth curve in the dashboard. In particular, some of the participants did not understand how to interpret the percentage of goals reached over the past 7 days and some interpreted it as a "grade" or "score." To simplify the weekly summary and ensure it would not be seen as punitive, the percentages were removed and substituted with a weekly view that used a green thumbs-up icon to indicate completion of the goal for that day, a grey thumbs-up if the goal was not completed, or a white thumbs-up if there was no data entered for that day. In the first iteration of the growth curve, participants were shown their child's BMI over time. Despite having a note that explained how BMI was calculated and what it meant, all 4 participants struggled with the interpretation of the growth curve. Given this, the growth curve was simplified to show weight instead of BMI and to have the entry of height and weight appear on the same page as the weight curve.

Literacy Challenges

While all of the content in the Healthy Lifestyle dashboard was available in Spanish, there were some features of the overall Nemours app that were only available in English at the time of

XSL•FO
RenderX

testing. This created some challenges. For example, participants were asked to send a message to their community health worker during Think Aloud testing. Since the Nemours app message center was only available in English, the language barrier caused some difficulty for participants. To provide families with more guidance when using the message center, a note in Spanish was added to the revised prototype to provide instructions on how to use the message center. In addition to this messaging task, 2 participants shared concerns about language barriers. As 1 participant stated:

> *For example, I have my mom … she definitely has trouble with technology. One, because she doesn't really work with technology, and secondly, because there is a language barrier. Even when things are in Spanish, she still feels uncomfortable.*

In recognition of these barriers, changes were made to simplify the language throughout the Healthy Lifestyle dashboard.

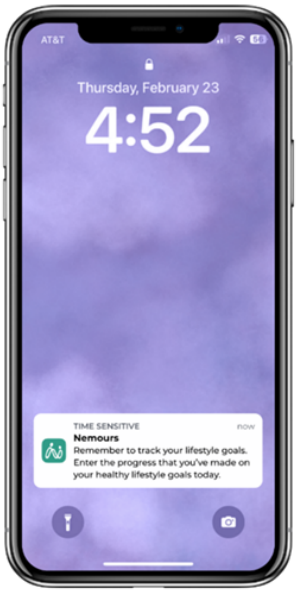
Phase 2

Overview

In phase 2, participants were shown a revised version of the Healthy Lifestyle prototype based on the feedback from phase 1. New changes were well-received by parents and 3 reported that the revised prototype was usable and understandable. Feedback during this phase highlighted the importance of certain features in the dashboard (Figure 4), with only minor suggestions provided.

XSL•FO
RenderX

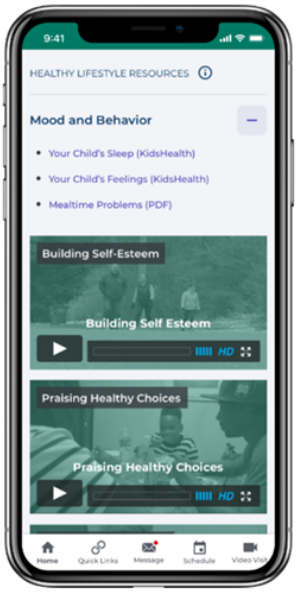
Figure 4. Features that caregivers of children with obesity from rural clinics in Delaware considered important to include during Think Aloud testing of a mobile app to promote healthy lifestyles.



Daily notifications

"I think that would be beneficial. I also think if there was a notification at the end that you could set for the end of the day to ask to fill it out to do the log for that day – so you're reminded. Like a reminder notification."
– English-speaking participant

"Yeah, as a parent, sometimes we need that reminder, so that would be great if the app could do that."
– English-speaking participant



Healthy Lifestyle resources

"I think it would help me a lot because I wouldn't be going through it Googling and trying to find out what I should be doing and what's healthier. It's all laid out there, and it would be a great resource to go to all at once and find out everything you need to know. And I wouldn't feel lost."
– English-speaking participant

"At the bottom of the page, there is Healthy Lifestyle Resource… I would select that, and it gives you all the things that you can do together, even videos – wow, this is nice – on how to be active together, which I would definitely watch and do."
– English-speaking participant

Important Features to Include

Participants liked many of the features already included in the Healthy Lifestyle dashboard and thought they would help motivate families. Participants found the multimedia resources (ie, videos, handouts, and links to external sources) embedded in the dashboard to be particularly useful. Participants especially liked the videos and thought they were a user-friendly way to view the resources. Having a set of curated resources in 1 centralized location was also an advantage because participants found it challenging to find reliable, evidence-based resources on their own. As 1 participant commented:

> *I would definitely promote all the things that app can help with – not just that it's about healthy living, healthy eating, and things like that, but also that it has resources for everything, because every family's different, and they deal with different things and different things within their children. There are great resources that you guys have and that you guys are offering.*

Finally, while participants liked the lifestyle goal feature and understood the importance of tracking their lifestyle goals, many participants talked about competing priorities that may get in the way of inputting their data. As a solution, they discussed

XSL•FO
RenderX

how an evening reminder to log their goals would help keep them on track.

Communication and Integration With Health Care Team

All participants discussed the importance of being able to communicate with their health care team through the dashboard, whether by sending messages or by having lifestyle goals and weight data available to their health care team. One participant highlighted how the transfer of the dashboard data to the child's care team could help keep families accountable for their goals:

> *I also think maybe to some parents, it would also be motivating to know – like if this is going to the provider, you want to show that as a parent, you're being proactive, and you're doing what you need to do for your child.*

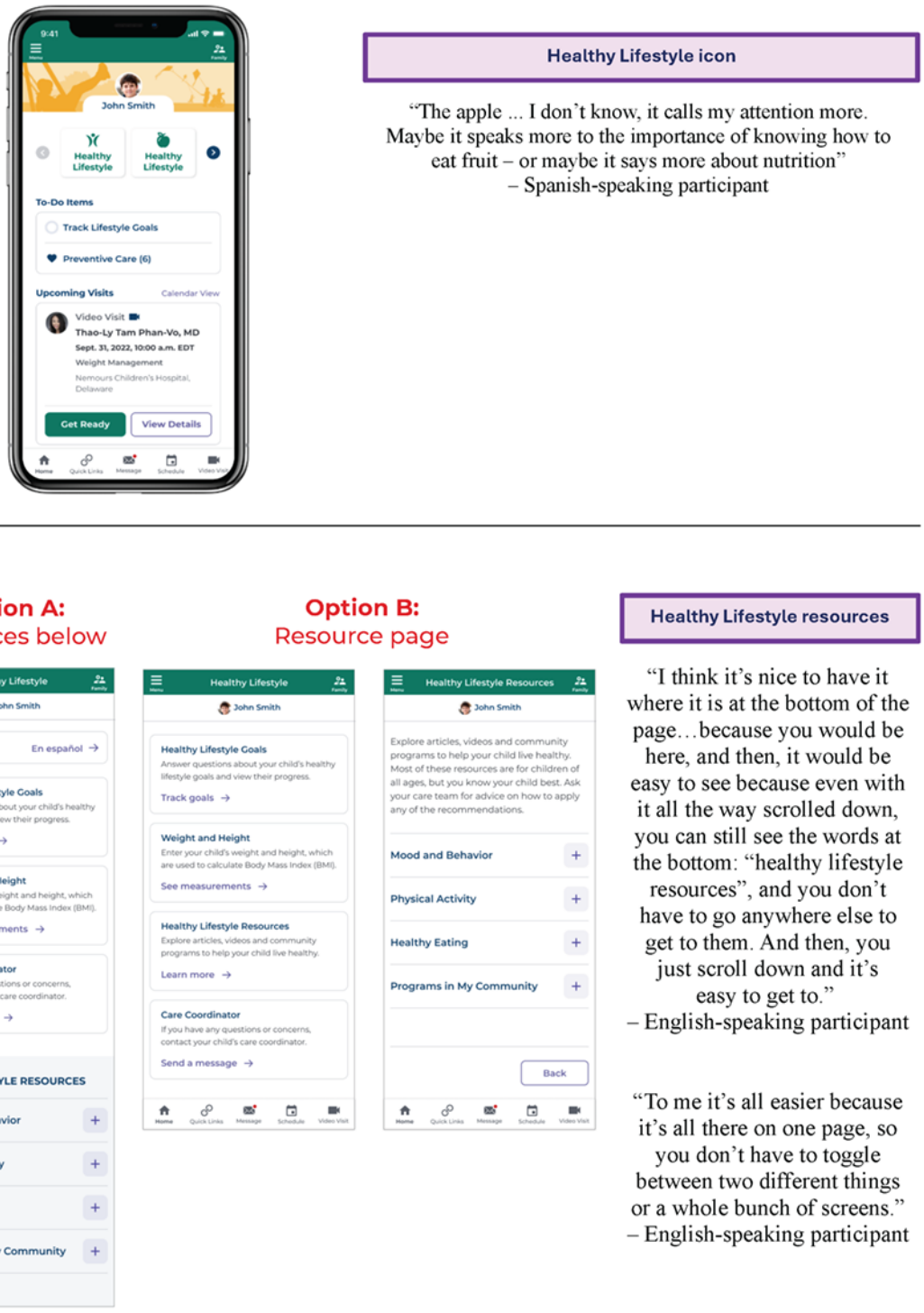
Despite the appeal of communicating with their health care team through the app, participants expressed a need for more clarity about who would view their data and how and when the health care team would be communicating back to them, which was incorporated into the final prototype.

Preference Testing

Preference testing is a research method during which participants compare 2 variations of a design, choose 1, and explain their preference [21], and it was used in both phases 1 and 2. This method was used for selecting an icon to represent the Healthy Lifestyle dashboard and determining the preferred way to incorporate resources (Figure 5).

XSL•FO
RenderX

Figure 5. Results of preference testing done during Think Aloud testing of caregivers of children with obesity from rural clinics in Delaware to determine a dashboard icon and resource page layout for a mobile app to promote healthy lifestyles.



Healthy Lifestyle Dashboard Icon

Since the Healthy Lifestyle dashboard is incorporated into the Nemours app, which is a much broader app designed for a range of health care needs, an icon was needed to launch the dashboard. Designers on the Nemours app team explored a range of concepts for the icon. Symbols that represent different aspects of a healthy lifestyle, like physical activity and nutrition, were considered. The top 2 designs—a joyous person (left) and an apple (right)—were shared during phase 1 testing. Based on feedback from all the participants, the apple icon was chosen to be the branding element for the Healthy Lifestyle dashboard. Participants cited liking the apple more because they associated it with nutrition, which was the primary domain that participants associated with healthy lifestyles.

XSL•FO
RenderX

Resource Page Layout

The list of Healthy Lifestyle resources included 19 items across 4 domains (healthy eating, being active, mood and behavior, and community programs). Preference testing was used to understand how participants preferred to access the resources. In total, 2 variations of the Healthy Lifestyle resources were designed. Option A made the resources available at the bottom of each page (like a footer) with expandable sections. Whereas option B involved adding them to the dashboard as a fourth link, which would open a new page for this section. Participants compared the 2 options and then indicated their preference. Though 2 of the 7 participants thought option B looked better, the majority picked option A. They liked having all the resources easily accessible on every page instead of having to navigate back and forth through the pages.

Phase 3

All 6 parents (refer to Table 1 for demographic information) who participated in phase 3 were in agreement that the revisions to the dashboard addressed their concerns from phase 1 or 2. They reported that the final prototype was useful and easy to understand. Given no further feedback about the dashboard, it was finalized for use in future research studies.

Discussion

Principal Findings

MICCO is a comprehensive, multicomponent intervention that was developed in collaboration with community, parent, and health care partners to improve the relevance and accessibility of pediatric obesity interventions for families from underserved and rural communities. Digital self-management tools are a key component of the MICCO intervention and are made available to families through a mobile app dashboard integrated with the EHR. This dashboard component of MICCO addresses notable limitations in the mHealth for childhood obesity literature by including community resources, being available in Spanish, incorporating culturally relevant resources, and being integrated into health care delivery. Human-centered design methods were applied to ensure that the needs and preferences of families of diverse backgrounds were incorporated. The resulting dashboard is user-friendly, literacy-sensitive, and culturally responsive and incorporates features that will facilitate its usability in future research and clinical practice.

Through a human-centered approach to design, the research team was able to better understand the challenges that families, especially those with limited English proficiency and from rural neighborhoods, may experience in using an app to facilitate healthy lifestyle behaviors. Challenges navigating health information are not unique to this study and have been described among larger cohorts of Spanish-speaking individuals of Hispanic or Latino ethnicity [26,27]. Even when health information is available in Spanish, studies have found that reading levels are typically above a sixth-grade level which is higher than the recommended reading level for health information in any language [28,29]. Difficulty understanding health information can contribute to low rates of engagement with programs and services among Spanish-speaking populations [27,30,31]. Therefore, testing mHealth interventions with Spanish-speaking users is critical to ensuring that the intervention is understood and well-received by these populations.

In this study, numeracy and literacy barriers were identified during Think Aloud testing with Spanish-speaking participants. Therefore, pages depicting weight and goal progress were simplified to convey relevant health information more clearly. In addition, information was added to help Spanish-speaking families navigate the 1 component (messaging function) of the app that was not fully available in Spanish at the time of testing. Not only were these adaptations made as part of this research study, but findings from this line of research in conjunction with other Diversity, Equity, and Inclusion initiatives at Nemours Children's Health, prompted an overhaul of the entire Nemours App to be fully available in Spanish after testing.

Feedback received during usability testing also helped the team develop a deeper understanding of families' busy lifestyles. In particular, the majority of participants expressed challenges balancing competing priorities when taking care of young children. Not surprisingly, time constraints are a commonly cited barrier to adherence in childhood obesity treatment [32,33]. Because of this, many participants reported that an evening reminder about tracking lifestyle goals would be helpful, so a push notification was implemented to remind families to track their lifestyle goals in the app. Participants also liked the curated resources available through the dashboard, especially the brief videos. Having resources that were easily accessible added to the convenience of the app for families; in addition to automated reminders, this may be a way to keep families engaged in treatment.

Finally, participants endorsed an appreciation for the integration of the dashboard with their overall health care. For example, they reported a desire for information tracked through the dashboard to be shared with their health care team and wanted to know more about how to communicate with their health care team. The Healthy Lifestyle dashboard was designed to facilitate this bidirectional communication. This approach is different than most apps designed for childhood obesity, which are stand-alone. While there are advantages to this, findings from this study suggest that mHealth interventions may be more effective if integrated with health care delivery, especially for rural communities that may experience lower access to health care facilities. Not only can this enhance satisfaction and motivation on the part of families, but it can also inform clinical decision-making on the part of the health care team [34,35].

Limitations

There were some limitations to this study. While the recruitment strategy was broad and a diverse sample was recruited, it is possible that our sample was biased toward those with more digital literacy and broadband access. In the future, it will be important to evaluate the app across families from varying socioeconomic backgrounds and with those who face greater barriers to care. Using in-person testing may also allow for testing of patients with limited digital access and provide a better understanding of nonverbal cues that may not have been captured in the video recording. While our sample size was

XSL•FO
RenderX

small, it is consistent with best practices for usability testing but there may be limitations in our methods of obtaining final input through an Advisory Council (vs repeated formal usability testing). Also, while participants endorsed the acceptability of the dashboard, the study's focus on usability does not allow for an understanding of whether participants would engage with it on a routine basis. A future study might conduct field testing of the MICCO app, wherein participants can use the app continuously for a shortened duration (eg, a month) and provide acceptability and usability feedback before it is deployed in a longer trial. In addition, previous studies have identified that digital interventions are subject to low uptake and engagement levels; therefore, further research using implementation frameworks will be needed to better understand factors that will facilitate routine use of the dashboard as part of the larger intervention. Finally, while the dashboard (and intervention more broadly) was designed for caregivers of young children, future studies should consider refining and testing the dashboard for youth with obesity who may find a digital intervention even more acceptable.

Conclusion

Collaborating with community and patient partners and conducting usability testing with key demographic populations, especially Spanish-speaking populations, was important to developing an mHealth intervention that is user-friendly, culturally relevant, and literacy-sensitive to promote pediatric healthy lifestyles among diverse families. Designing an intervention that is usable for intended users and in intended settings is a critical first step to trialing and implementing a successful intervention.

Acknowledgments

We thank the National Institutes of Health/National Institute of General Medical Sciences (P20GM144270) for supporting this project through the Center of Biomedical Research Excellence for the REACH Center. We also want to thank the parents who generously gave us their time to participate in this study.

Data Availability

Deidentified data are available upon reasonable request. All requests should be directed to the corresponding author.

Authors' Contributions

APR collected and analyzed data and conceptualized, drafted, reviewed, and revised the manuscript. TLP designed the study, coordinated, and supervised data collection, assisted with data analysis, and conceptualized, drafted, reviewed, and revised the paper. AK designed the study and reviewed and revised the manuscript. AMO collected data and drafted, reviewed, and revised the manuscript. AO conducted data analyses and drafted, reviewed, and revised the manuscript. NS designed and refined the prototype and drafted, reviewed, and revised the manuscript. All authors approved the final manuscript as submitted and agree to be accountable for the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mobile Integrated Care for Childhood Obesity (MICCO) interview guide used to guide Think Aloud sessions with caregivers of young children with obesity.
[DOCX File , 33 KB - formative_v9i1e60495_app1.docx]

Multimedia Appendix 2
Mobile Integrated Care for Childhood Obesity (MICCO) – Think Aloud code book.
[DOCX File , 27 KB - formative_v9i1e60495_app2.docx]

References

1. Hales CM, Carroll MD, Fryar CD, Ogden CL. Prevalence of obesity among adults and youth: United States, 2015-2016. NCHS Data Brief 2017(288):1-8 [FREE Full text] [Medline: 29155689]
2. Ogden CL, Fryar CD, Hales CM, Carroll MD, Aoki Y, Freedman DS. Differences in obesity prevalence by demographics and urbanization in US children and adolescents, 2013-2016. JAMA 2018;319(23):2410-2418 [FREE Full text] [doi: 10.1001/jama.2018.5158] [Medline: 29922826]
3. Trasande L, Chatterjee S. The impact of obesity on health service utilization and costs in childhood. Obesity (Silver Spring) 2009;17(9):1749-1754 [FREE Full text] [doi: 10.1038/oby.2009.67] [Medline: 19300433]

XSL•FO
RenderX

4. Sallis JF, Floyd MF, Rodríguez DA, Saelens BE. Role of built environments in physical activity, obesity, and cardiovascular disease. Circulation 2012;125(5):729-737 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.110.969022] [Medline: 22311885]
5. Yusuf ZI, Dongarwar D, Yusuf RA, Bell M, Harris T, Salihu HM. Social determinants of overweight and obesity among children in the United States. Int J MCH AIDS 2020;9(1):22-33 [FREE Full text] [doi: 10.21106/ijma.337] [Medline: 32123625]
6. Galani C, Schneider H. Prevention and treatment of obesity with lifestyle interventions: review and meta-analysis. Int J Public Health 2007;52(6):348-359. [doi: 10.1007/s00038-007-7015-8] [Medline: 18368998]
7. Janicke DM, Steele RG, Gayes LA, Lim CS, Clifford LM, Schneider EM, et al. Systematic review and meta-analysis of comprehensive behavioral family lifestyle interventions addressing pediatric obesity. J Pediatr Psychol 2014;39(8):809-825. [doi: 10.1093/jpepsy/jsu023] [Medline: 24824614]
8. Tate EB, Spruijt-Metz D, O'Reilly G, Jordan-Marsh M, Gotsis M, Pentz MA, et al. mHealth approaches to child obesity prevention: successes, unique challenges, and next directions. Transl Behav Med 2013;3(4):406-415 [FREE Full text] [doi: 10.1007/s13142-013-0222-3] [Medline: 24294329]
9. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
10. Azevedo LB, Stephenson J, Ells L, Adu-Ntiamoah S, DeSmet A, Giles EL, et al. The effectiveness of e-health interventions for the treatment of overweight or obesity in children and adolescents: a systematic review and meta-analysis. Obes Rev 2022;23(2):e13373. [doi: 10.1111/obr.13373] [Medline: 34747118]
11. Psihogios AM, Lane-Fall MB, Graham AK. Adolescents are still waiting on a digital health revolution: accelerating research-to-practice translation through design for implementation. JAMA Pediatr 2022;176(6):545-546. [doi: 10.1001/jamapediatrics.2022.0500] [Medline: 35404381]
12. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
13. Tully L, Burls A, Sorensen J, El-Moslemany R, O'Malley G. Mobile health for pediatric weight management: systematic scoping review. JMIR mHealth uHealth 2020;8(6):e16214 [FREE Full text] [doi: 10.2196/16214] [Medline: 32490849]
14. Partridge SR, Knight A, Todd A, McGill B, Wardak S, Alston L, et al. Addressing disparities: a systematic review of digital health equity for adolescent obesity prevention and management interventions. Obes Rev 2024;25(12):e13821. [doi: 10.1111/obr.13821] [Medline: 39256958]
15. Tsoi MF, Li HL, Feng Q, Cheung CL, Cheung TT, Cheung BMY. Prevalence of childhood obesity in the United States in 1999-2018: a 20-year analysis. Obes Facts 2022;15(4):560-569 [FREE Full text] [doi: 10.1159/000524261] [Medline: 35358970]
16. Soltero EG, Lopez C, Hernandez E, O'Connor TM, Thompson D. Technology-based obesity prevention interventions among Hispanic adolescents in the United States: scoping review. JMIR Pediatr Parent 2022;5(4):e39261 [FREE Full text] [doi: 10.2196/39261] [Medline: 36331547]
17. Maguire M. Methods to support human-centred design. International Journal of Human-Computer Studies 2001;55(4):587-634. [doi: 10.1006/ijhc.2001.0503]
18. Bandura A. Human agency in social cognitive theory. American Psychologist 1989;44(9):1175-1184. [doi: 10.1037/0003-066X.44.9.1175]
19. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
20. Davison KK, Birch LL. Childhood overweight: a contextual model and recommendations for future research. Obes Rev 2001;2(3):159-171 [FREE Full text] [doi: 10.1046/j.1467-789x.2001.00036.x] [Medline: 12120101]
21. Tomlin W. UX and Usability Testing Data: Combining Behavioral UX and Usability Testing Data to Optimize Websites. Berkeley, CA: Apress; 2018.
22. Nielsen J, Landauer T. A mathematical model of the finding of usability problems. 1993 Presented at: CHI '93: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems; 1993 April 24 - 29; USA. [doi: 10.1145/169059.169166]
23. USDA ERS - Rural-Urban Commuting Area Codes. URL: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/ [accessed 2024-10-11]
24. Liinasuo M, Kuula T, Goriachev V, Helin K. Choosing between remote and face-to-face features in a test setting: methodology used in two usability and user experience case studies. Behaviour & Information Technology Taylor & Francis 2023;42(3):298-306. [doi: 10.1080/0144929x.2023.2165453]
25. Dedoose. Los Angeles, CA: SocioCultural Research Consultants, LLC; 2021.

XSL•FO
RenderX

26. Millar RJ, Sahoo S, Yamashita T, Cummins PA. Literacy skills, language use, and online health information seeking among Hispanic adults in the United States. Patient Educ Couns 2020;103(8):1595-1600. [doi: 10.1016/j.pec.2020.02.030] [Medline: 32115313]
27. Chavarria EA, Christy SM, Feng H, Miao H, Abdulla R, Gutierrez L, et al. Online health information seeking and eHealth literacy among Spanish language-dominant Latino adults receiving care in a community clinic: secondary analysis of pilot randomized controlled trial data. JMIR Form Res 2022;6(10):e37687 [FREE Full text] [doi: 10.2196/37687] [Medline: 35238785]
28. Karthik N, Barekatain K, Vu H, Wu DTY, Ehrlich JR. Readability comparison of online Spanish and English patient education materials about vision health. Ophthalmic Epidemiol 2022;29(2):182-188. [doi: 10.1080/09286586.2021.1910316] [Medline: 33832394]
29. Mazmudar RS, Sheth A, Tripathi R, Scott JF. Readability of online Spanish patient education materials in dermatology. Arch Dermatol Res 2021;313(3):201-204. [doi: 10.1007/s00403-020-02036-7] [Medline: 32020323]
30. Ohtani A, Suzuki T, Takeuchi H, Uchida H. Language barriers and access to psychiatric care: a systematic review. Psychiatr Serv 2015;66(8):798-805. [doi: 10.1176/appi.ps.201400351] [Medline: 25930043]
31. Marquine MJ, Jimenez D. Cultural and linguistic proficiency in mental health care: a crucial aspect of professional competence. Int Psychogeriatr 2020;32(1):1-3 [FREE Full text] [doi: 10.1017/S1041610219000541] [Medline: 32008601]
32. Skelton JA, Martin S, Irby MB. Satisfaction and attrition in paediatric weight management. Clin Obes 2016;6(2):143-153 [FREE Full text] [doi: 10.1111/cob.12138] [Medline: 27008068]
33. McMaster CM, Cohen J, Alexander S, Neal R, Gow ML, Calleja E, et al. Satisfaction and acceptability of paediatric weight management services amongst parents and carers: a mixed-methods study. Clin Obes 2020;10(6):e12391. [doi: 10.1111/cob.12391] [Medline: 32830905]
34. Moon K, Sobolev M, Kane JM. Digital and mobile health technology in collaborative behavioral health care: scoping review. JMIR Ment Health 2022;9(2):e30810 [FREE Full text] [doi: 10.2196/30810] [Medline: 35171105]
35. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]

Abbreviations

EHR: electronic health record
mHealth: mobile health
MICCO: Mobile Integrated Care for Childhood Obesity

Edited by A Mavragani; submitted 13.05.24; peer-reviewed by T O'Connor, K Haas, B Lancaster; comments to author 30.09.24; revised version received 27.11.24; accepted 24.12.24; published 11.02.25.

Please cite as:
Perez Ramirez A, Ortega A, Stephenson N, Muñoz Osorio A, Kazak A, Phan TL
mHealth App to Promote Healthy Lifestyles for Diverse Families Living in Rural Areas: Usability Study
JMIR Form Res 2025;9:e60495
URL: https://formative.jmir.org/2025/1/e60495
doi:10.2196/60495
PMID:39932772

©Alejandra Perez Ramirez, Adrian Ortega, Natalie Stephenson, Angel Muñoz Osorio, Anne Kazak, Thao-Ly Phan. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Social Acceptance of Smart Glasses in Health Care: Model Evaluation Study of Anticipated Adoption and Social Interaction

Niek Zuidhof[1,2], PhD; Oscar Peters[3], PhD; Peter-Paul Verbeek[4], Prof Dr; Somaya Ben Allouch[5,6], Prof Dr

[1]Research Group Technology, Health & Care, Saxion University of Applied Sciences, Enschede, The Netherlands

[2]Research Group Employability Transition, Saxion University of Applied Sciences, Enschede, The Netherlands

[3]School of Finance and International Business, Saxion University of Applied Sciences, Enschede, The Netherlands

[4]Philosophy and Ethics of Science and Technology in a Changing World, University of Amsterdam, Amsterdam, The Netherlands

[5]Research Group Digital Interactions, University of Amsterdam, Amsterdam, The Netherlands

[6]Research Group Digital Life, Amsterdam University of Applied Sciences, Amsterdam, The Netherlands

Corresponding Author:
Niek Zuidhof, PhD
Research Group Technology, Health & Care
Saxion University of Applied Sciences
M.H. Tromplaan 28
Enschede, 7513 AB
The Netherlands
Phone: 31 682381157
Email: e.n.zuidhof@saxion.nl

Abstract

Background: Despite the growing interest in smart glasses, it is striking that they are not widespread among health care professionals. Previous research has identified issues related to social interactions involving the use of smart glasses in public settings, which may differ from those associated with their application in health care contexts.

Objective: Assuming that smart glasses mediate contact between the health care provider and patient, the objectives of this research are two-fold: (1) to develop an instrument that combines the adoption and mediation perspectives, and (2) to gain insights into how the intention to use is influenced through aspects of adoption and social interaction.

Methods: A questionnaire was administered to a target audience of health care professionals (N=450), with recruitment via MTurk. The sample primarily included male participants from the United States, with the majority aged 42 years or younger. Although a large portion of respondents were medical doctors, the sample also included nurses and other health care professionals. Data were analyzed by structural equation modeling.

Results: Regarding the aim of developing an instrument combining adoption and social interaction, the internal consistency was above the aspirational level (α>.70) for the instrument. Furthermore, regarding the second objective involving gaining insights into the influential constructs of the anticipated intention to use, the following results were highlighted: in testing the conceptual model, the measurement model generated a good fit and the respecified structural model also generated a good fit. The tested hypotheses confirmed that social interaction constructs could explain a higher variance of users' anticipated intention to use. Perceived social isolation and decreased attentional allocation did not have a significant effect on attitude. Furthermore, the intention to use smart glasses despite nonacceptance of smart glasses by the patient significantly influenced the anticipated intention to use. In summary, constructs that focus on social interaction could contribute to better explanation and prediction of the expected adoption of smart glasses in health care.

Conclusions: The empirical findings of this study provide new insights into how the mediation perspective can increase the explained variance compared to existing knowledge about adoption. Against expectations based on previous literature and despite the social issues raised earlier, these social aspects do play important roles for health care professionals but are ultimately not decisive for the intention to use. As a result, there are fewer threats to the adoption of smart glasses from the perspective of health care professionals than might be expected based on the previous literature. Therefore, the use of smart glasses can still be considered as an innovative way of working in health care.

(JMIR Form Res 2025;9:e49610) doi:10.2196/49610

XSL•FO
RenderX

KEYWORDS

smart glasses; technology adoption; social interaction; instrument development; structural equation modeling

Introduction

Background

The initial public reactions to smart glasses, notably Google Glass (Figure 1), were quite mixed. Google Glass was both praised for its innovative technologies, such as a see-through screen and hands-free use, and criticized for various reasons. Shortly after the introduction of Google Glass, cautionary stories about "Glassholes" emerged, a term coined to describe individuals who misuse the technology in socially inappropriate ways [1]. Examples of this phenomenon include people who wear Google Glass and ignore the people around them or stare off into space while operating their smart glasses. Google responded with rules to help users avoid these situations [2]. There are several examples of various emotional reactions to smart glasses in the media, and ethical questions are repeatedly raised about wearing smart glasses [3,4]. Although research into smart glasses is continuously expanding, current knowledge and theoretical foundations about the acceptance of smart glasses are still scarce.

Smart glasses have emerged as a growing research topic over the past years in various fields [5,6] and could bring about a revolution in health care [7]. Smart glasses are computerized head-mounted devices worn like eyeglasses, which can collect and provide information to the user and the environment. The majority of smart glasses incorporate a display through which information can be accessed while simultaneously maintaining visibility of the external environment. Furthermore, these devices often provide the advantage of hands-free operation [8]. Pilot studies on the use of smart glasses in health care have been conducted [7]; however, few researchers have addressed the issue of preadoption criteria, and identifying these criteria remains a critical issue [9].

Figure 1. Example of a health care professional wearing Google Glass smart glasses in a home environment to illustrate the appearance of the glasses.

Prior Research on the Adoption of Smart Glasses

The initial responses to smart glasses raise the question of how the adoption of smart glasses is progressing.

The growing body of research on the adoption of smart glasses was searched in the existing scientific literature. The following keywords were used: "smart glasses," "head-mounted display*," "wearable display*," "augmented reality (smart) glasses," "adoption," "acceptance," and "acceptability." The search was performed in the following databases on June 12, 2021: Scopus (11 hits), Web of Science (12 hits), ACM Digital Library (96 hits), IEEE Xplore (2 hits), and PubMed (0 hits). All synonyms were connected with the disjunction "OR," and both "smart glasses" and "adoption" were connected with the conjunction "AND." Keywords were searched in the title. The identified publications are listed in Multimedia Appendix 1.

Various methodologies have been employed by researchers to investigate the adoption of smart glasses. For an overview of these methods, please refer to Multimedia Appendix 1. The context of the studies varied, with the general context most often consisting of random individuals or students. A small number of studies were found in the context of health care and tourism; therefore, it can be concluded that knowledge about the adoption of smart glasses is still scarce in these fields. Adoption studies with industry, retail, and agriculture respondents are even more limited. Various methods have been applied, with the majority of studies being qualitative in nature. Most of the previous quantitative studies had a relatively low number of respondents.

XSL•FO
RenderX

The findings presented in Multimedia Appendix 1 offer foundational inputs for the development of a questionnaire. Furthermore, the social aspects surrounding smart glasses were mentioned multiple times in previous studies, and therefore, it could be assumed that social interaction is a relevant topic regarding the acceptance of smart glasses.

The Unified Theory of Acceptance and Use of Technology (UTAUT) [10], the Technology Acceptance Model (TAM) [11], and the Diffusion of Innovations [12] were found to be the most commonly accepted theories by scholars [13,14]. Some of the findings from prior research in Multimedia Appendix 1 might be related to the aforementioned models such as the influence on perceived usefulness, perceived ease of use, facilitating conditions, image, and perceived enjoyment [10,15]. However, other findings suggest themes regarding both the ethical and social contexts, which can be termed social acceptability [16]. Social acceptability can be divided into conversation partners, such as interpersonal communication, and spectators or third parties, with examples such as disturbance and controversial public use. The theme of privacy is also related to the social context and can be divided into the same categories: conversation partners and third parties. For example, wearing smart glasses is not only about the privacy of the user, but also about other people's privacy, and this can strongly influence the user's decision-making [17]. Figure 1 provides an illustrative example of an encounter with a health care professional wearing smart glasses. To summarize the findings of the prior research in Multimedia Appendix 1, it can be concluded that relevant aspects of the preadoption criteria should be informed by the current knowledge in adoption research, such as the TAM, and should also take into account the social context from the perspectives of both caregivers and patients.

Devices worn on the body in general influence the symbolic meaning in the natural environment of the person and the public surroundings and have psychological impacts on wearers and behavioral consequences for wearers [18]. Over the past decades, the TAM [19] had an incredible influence on empirical research. One might also say in general that current adoption models take a snapshot of adoption readiness, and the TAM consistently explains about 40% of the variance in the intention to use IT and the actual usage by individuals [15]. However, these adoption models have limitations [20], in which the current reality regarding smart glasses may not be fully involved. Current models are linear and strongly focused on the user's perspective alone. Moreover, according to recent studies, social influence, as it is considered in the UTAUT [10], may involve more than just the influence and desire of others in the future use of the one considering it [3]. Several studies have suggested that social context influences usage norms and that public use is controversial [21,22]. The use of smart glasses does not reproduce existing norms and understandings in turn-taking, knowledge, and identity and could reconfigure norms by creating new settings in which activities begin to compete and interfere with each other [23,24]. In addition, privacy is often mentioned to be influential [25]. The privacy of the user is important, and other people's privacy can also strongly influence the user's decision-making [17]. Therefore, smart glasses force us to take a broader view of adoption models, and we should not only consider the perspectives of potential users but also look at the surroundings. Hence, social influences might be currently underestimated in the context of the acceptance of smart glasses given the reports from various media and recent publications on the subject.

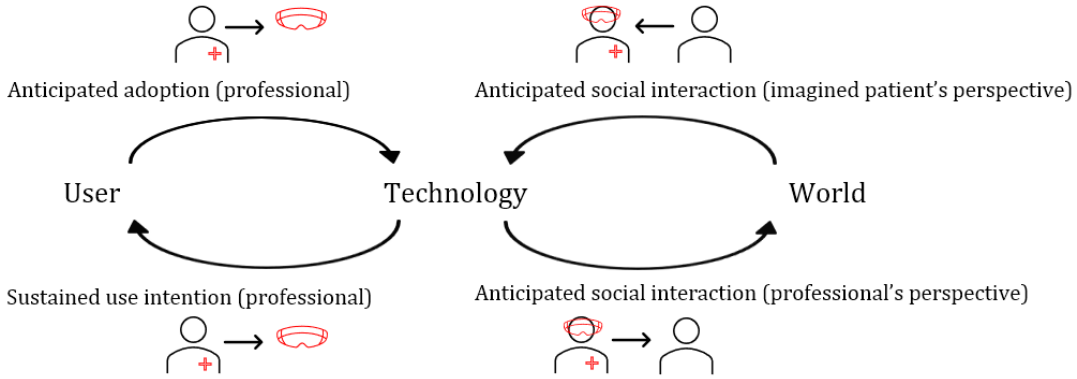
Turning to a perspective that takes a broader view of users and their surroundings and environment [26], the mediation perspective fits well with the identified gaps regarding the acceptance of smart glasses. The mediation perspective of the philosophy of technology considers the relationship between humans and technology from a phenomenological perspective, and it inspired the beginning of ubiquitous computing [27]. Furthermore, Mann [28] explained the use of smart glasses after over 35 years as a mediated version of reality. Those mediated relationships are reciprocal, and technology is perceived as the mediator between humans and the world. For example, a user looks *through* glasses rather than *at* glasses, and this is therefore termed as an embodiment relationship. In addition, smart glasses can represent information on a parallel screen while seeing the world, which is termed as a hermeneutic relationship [29]. A combined perspective of both adoption and mediation could complement each other in understanding the appropriation of smart glasses (Figure 2; derived and adapted from Zuidhof et al [30]). However, knowledge of the mediated experience of smart glasses is still scarce, and recent studies have reported strong influences of the social context in decision-making [17,21]. The theoretical framework of adoption and mediation can reveal both the user's perspective and the perspective of others who are related to the user, such as a patient.

In the field of the philosophy of technology, mediation is interpreted as determining human-technology relationships and interactions, and anticipating influences on human behavior [29]. Prior studies on the adoption of smart glasses, as illustrated in Multimedia Appendix 1, not only present familiar constructs but also have begun, since 2010, to delve into social themes from various perspectives, including those of the user, conversational partner, and third parties [26]. A large-scale survey in which the anticipated perspective of the health care professional and the anticipated social interaction are united is lacking in current knowledge. Based on previous research results and the theoretical framework of adoption and mediation (as part of a doctoral research project), it seems relevant to explore adoption themes and assess if social interaction influences the intention to use smart glasses.

Figure 2 proposes to adapt the theoretical framework of adoption and mediation to the context of the health care professional and the patient. It is considered that smart glasses for health care professionals are still in an early phase of dissemination, and therefore, it is needed to gather knowledge about anticipating the use of smart glasses. Moreover, health care professionals are expected to anticipate adoption while taking into account social interaction with patients, but they can also take into account the patient's perspective in social interaction. Therefore, this goes beyond mere adoption for self-interest and may lead to a higher or lower intention to use smart glasses.

XSL•FO
RenderX

Figure 2. Theoretical framework of anticipated adoption and social interaction.



Objectives and Hypotheses

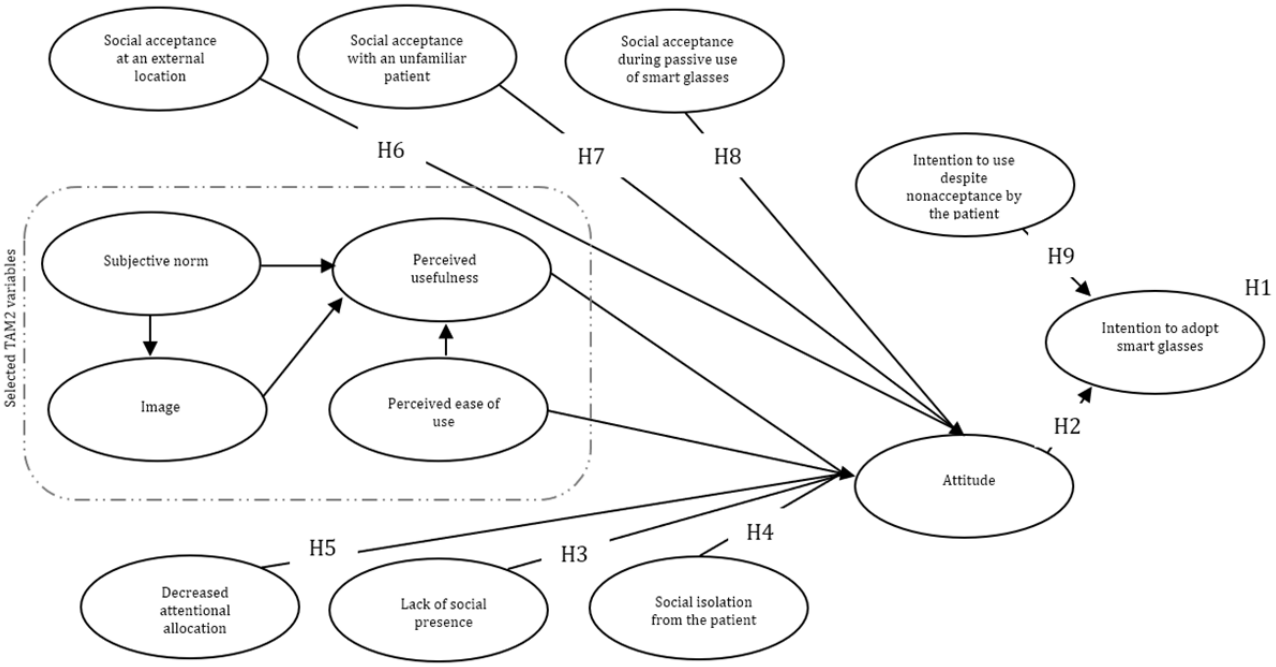
Prior research on the adoption of smart glasses has been mainly conducted in general contexts; however, some studies were carried out in the area of tourism or health care. Few of the recent qualitative studies on smart glasses in health care are part of the authors' doctoral research projects. Furthermore, it appears that there are no large-scale surveys conducted on the adoption of smart glasses and that most of the recent studies have an explorative and qualitative nature. Prior research on the adoption of smart glasses is presented in Multimedia Appendix 1. In addition, the emotional and ethical aspects that can play roles as issues to adoption described above are still involved to a limited extent in recent studies and might be especially relevant in professions where people are vulnerable or interact intensively like in health care.

The objectives of this study were two-fold: (1) to develop an instrument that combines the perspectives of adoption and mediation to measure the intention to use smart glasses by health care professionals, and (2) to obtain insights into how the intention to use is influenced by adoption and social interaction aspects. Regarding the state of the art and the objectives of this study, the following research question was formulated: "To what extent do constructs derived from a mediation perspective contribute to the established constructs of adoption to explain the intention to use smart glasses?"

Observed Variables Derived From Technology Acceptance Literature

In the proposed model (Figure 2), it was argued that technology adoption and social influence might influence each other and the user. The underlying idea here is that a decision to use smart glasses is explained by a linear model and is not only aimed at the self but also others, as is endorsed by the mediation perspective [30,31]. The recently developed conceptual model to explain the anticipated adoption and the social aspects of smart glasses has been outlined in Figure 3, and the aspects will be included.

Figure 3. Conceptual model with the hypotheses of the anticipated adoption and social interaction of smart glasses by health care professionals. H: hypothesis; TAM: Technology Acceptance Model.



XSL•FO
RenderX

Observed Variables Derived From the Technology Acceptance Model 2

As one of the most accepted theories by scholars, in the TAM [13,14], intention to use is proposed as a key independent variable. The constructs "perceived usefulness" and "perceived ease of use" are mainly aimed at their considerations and do not yet tell us much about how this relates to others. Therefore, we found the Technology Acceptance Model 2 (TAM2) to be the most appropriate because it builds on the basic model and extends this model with constructs that take others into account, such as the constructs "subjective norm" and "image." Other constructs from the TAM2 model, such as experience, voluntariness, job relevance, output quality, and result demonstrability, have been excluded as they extend beyond the scope of this study. These constructs do not specifically address social interaction, which is the central emphasis of our research. By adding specifically more social constructs, we aim to gain a more holistic perspective of not only personal considerations in own use but also how much a health care professional takes patients into account and how this influences the anticipated use of technology.

Hypothesis 1 is as follows: Additional social interaction constructs can explain the higher variance of users' anticipated intention to use smart glasses compared with subjective norm, image, perceived usefulness, and perceived usefulness on anticipated intention to use.

The attitude toward technology is a debatable construct in adoption studies because the added value in the model might not be good enough [10], or it may theoretically differ little from the intention to use. Similar to the Theory of Planned Behavior [32], we regard attitude as a predictor of the intention to use smart glasses, and we added the construct "attitude" because of previous findings [31,33], which may be related to the early phase of the adoption of smart glasses where we can only measure the anticipated intention to use.

Hypothesis 2 is as follows: Users' attitudes directly influence the anticipated intention to use smart glasses.

Observed Variables Regarding Social Interaction With Smart Glasses

Previous studies found several constructs under the themes that seemed theoretically relevant, such as attentional shift, social presence, and social isolation [14,31,34-36]. The items were translated as accurately as possible and applied to the anticipation of smart glasses. This resulted in the following constructs: lack of social presence, social isolation of the patient, and reduced attentional allocation to the patient.

Hypothesis 3 is as follows: The lack of social presence due to wearing smart glasses during a physical contact moment influences the attitude directly.

Hypothesis 4 is as follows: Social isolation due to wearing smart glasses during a physical contact moment influences the attitude directly.

Hypothesis 5 is as follows: The decreased attentional allocation due to wearing smart glasses during a physical contact moment influences the attitude directly.

Observed Variables Regarding Social Scenarios With Smart Glasses in Health Care

Furthermore, in previous qualitative studies, themes related to the context of smart glass use were also found, such as location, an unknown patient, and wearing smart glasses but not actively using them [31]. These scenarios were linked to social presence [36], resulting in social presence in 3 scenarios, namely, providing care with smart glasses on in an external location (outpatient care), providing care with smart glasses while in passive use, and providing care with smart glasses to an unfamiliar patient.

Hypothesis 6 is as follows: The use of smart glasses at an external location during a physical contact moment directly influences the attitude.

Hypothesis 7 is as follows: The use of smart glasses with an unknown person during a physical contact moment directly influences the attitude.

Hypothesis 8 is as follows: The passive use of smart glasses during a physical contact moment directly influences the attitude.

Observed Variables Regarding Nonacceptance of Smart Glasses by the Patient

The mediation approach visualizes the relationship with the outside world during adoption, while technology acts as a mediator in the work of the health care professional and the patient. To understand both the social interaction with patients during the use of smart glasses and the sustained use of smart glasses, the unintended consequences of smart glass use might also be relevant [26,37-39]. In particular, the patient's nonacceptance of the use of smart glasses is interesting, simply because the patient's acceptance of smart glasses would probably not create tension in the intentional use of smart glasses. The mediation perspective has led to the following construct in the model: intention to use despite social nonacceptance by others.

Hypothesis 9 is as follows: The intention to use smart glasses despite the nonacceptance of smart glasses by the patient directly influences the anticipated intention to use.

Methods

Study Design

To achieve the aims of this study, a summated rating scale development approach for social sciences was used [40]. The initial steps were aimed at defining constructs and designing the scale. The resulting scale was evaluated through an online survey among a sample of students.

Defining Constructs and Item Pool Development

To develop an item pool, existing literature was consulted, and the state-of-the-art knowledge, as described in the Introduction, was combined with the theoretical framework of adoption and mediation. A core aspect of the framework (Figure 2) is the reciprocal relationship among the user, the technology, and the world. In the next step, the aspects from Figure 2 were substantiated with existing literature and translated into items.

XSL•FO
RenderX

See Multimedia Appendix 2 for the item pool. All measurements were recorded using 5-point Likert scales. Back translation was performed by 2 researchers as a validation method for the survey [41] and to develop a robust item pool.

Item Pool Refinement

To check for comprehensibility, readability, and internal consistency, 2 pretest moments were organized. First, the questions were reviewed by 2 potential respondents. Next, data collection (as a pretest) was conducted at a health care institution in the Netherlands (n=17). Although the pretest was conducted on a small sample, internal consistency was acceptable. Initially, reversed items were included, but inconsistent response patterns were observed for those items. Respondents reported too much confusion with reversed items, and therefore, these items were removed from the questionnaire.

Data Collection

Respondents (N=450) were recruited in July 2022 on the internet using Amazon's Mechanical Turk. The questionnaire was made available to respondents who met the following inclusion criteria: (1) language: English; (2) employment industry: health care; and (3) location: Northwest Europe (Austria, Belgium, Denmark, Finland, France, Germany, Iceland, Ireland, Luxembourg, Netherlands, Norway, Sweden, and Switzerland), Canada, United Kingdom, and United States. The questionnaire was closed once the desired number of respondents was reached.

Demographic Information

A summary of the demographic information of the respondents is presented in Table 1. The sample pool consisted mainly of male participants, with over three-quarters being 42 years of age or younger. The majority of the respondents were from the United States. With regard to occupation, the participants were distributed as follows: physicians, 192 (42.6%); nurses, 165 (36.7%); and others, 93 (20.7%). Other professions included a diverse group of medical assistants (n=14), medical technical professionals (n=11), managers (n=10), health care professionals (n=9), IT professionals (n=6), and home care professionals (n=4), among other various professions in health care.

XSL•FO
RenderX

Table 1. Demographic information of the respondents.

Characteristic	Value (N=450), n (%)
Gender	
Female	303 (32.4)
Male	146 (67.3)
Prefer not to say	1 (0.2)
Age (years)	
>72	1 (0.2)
63-72	15 (3.3)
53-62	42 (9.3)
43-52	45 (10.0)
33-42	124 (27.6)
23-32	222 (49.3)
18-22	1 (0.2)
Country of residence	
Canada	4 (0.9)
United States of America	442 (98.2)
Northwest Europe	3 (0.7)
United Kingdom	1 (0.2)
Highest degree	
High school or less	5 (1.1)
Some college but no degree	22 (4.9)
Associate degree	32 (7.1)
Bachelor’s degree	235 (52.2)
Master’s degree	145 (32.2)
Professional or doctoral degree	11 (2.4)
Current profession	
Physician	192 (42.6)
Nurse	165 (36.7)
Other	93 (20.7)
Employment status	
Employed full time	401 (89.1)
Employed part-time	35 (7.8)
Student	4 (0.9)
Self-employed	10 (2.2)
Years of work experience	
<10	224 (49.8)
11-20	115 (25.6)
21-30	74 (16.4)
31-40	28 (6.2)
>40	9 (2.0)

XSL•FO
RenderX

Ethical Considerations

Before starting the survey, respondents provided active online consent after reading the opening statement (Multimedia Appendix 3). Furthermore, pictures were used with permission from health care professionals and patients. Respondents were rewarded with US $2 when they completed the survey. Study data were deidentified and anonymously used for analysis. This study was reviewed and approved by the Ethical Review Board of the University of Twente (approval number 211121).

Procedure

After the opening statement and consent, respondents were asked to watch 2 short videos about smart glasses to get an idea of the technology and its use. The first video showed an augmented reality platform for health care with English-spoken animations [42], and the second video showed an example of Vuzix M400 smart glasses in the context of health care [43]. After watching both videos, respondents were asked to fill out the questionnaire. Demographic information was requested at the end of the questionnaire. Some photos used in the questionnaire are presented in Figure 4.

Figure 4. Photos used in the questionnaire about examples of using smart glasses consecutively: (A) contact with patients, (B, C) face-to-face interaction in outpatient care, (D) inpatient care, (E) passive use of smart glasses, and (F) active use of smart glasses (published with permission).

Data Analysis

Before the analyses, data were checked for normality. No significant deviation from normality was found. All skewness and kurtosis values were below 1.96, and values between –2 and +2 are considered acceptable to prove normal distribution [44]. Structural Equation Modeling was performed with maximum likelihood estimation, as advocated by [45,46], to test the conceptual model of the intention to adopt smart glasses by health care professionals. To test if the conceptual model fits the data well, the next absolute and incremental model fit indices were used as suggested by Holbert and Stephensen [47] and Hoyle and Panter [48]. The most common absolute fit index is the χ^2 goodness-of-fit test, and to minimize the impact of sample size on the model chi-square, the relative/normed chi-square ($\chi^{2/df}$) was used [49]. The standardized root mean square residual (SRMR) was reported as an absolute fit index. The following incremental fit indices have been reported: Tucker-Lewis index (TLI) and root mean square error of approximation (RMSEA). The practitioner recommendations for model evaluation were followed, suggesting a cutoff value close to 0.95 for the TLI in combination with a cutoff value close to 0.09 for the SRMR to evaluate model fit, and a cutoff value close to 0.06 for the RMSEA [50]. The Fornell and Larcker discriminant validity criterion was used to test discriminant validity. The Fornell and Larcker criterion is satisfied when a construct is more closely related to its own indicators than to other constructs [51].

Results

Diffusion of Smart Glasses in Health Care

The questionnaire on the anticipated adoption and mediation of smart glasses was designed to examine the perceptions of future users regarding smart glasses. It was also designed to get a better understanding of how specific situations in the use of smart glasses influence the intention to use them. The research question was: “To what extent do constructs derived from a mediation perspective contribute to the established constructs of adoption to explain the intention to use smart glasses?”

Participants were queried regarding their familiarity and experience with smart glasses. The constructs “knowledge” and “persuasion” served as constructs in assessing how health care professionals perceive the diffusion of smart glasses. Overall, the respondents reported having substantial knowledge about smart glasses (eg, I know what someone can do with smart glasses; mean score 4.02, SD 0.84) and expressed understanding of the concept of smart glasses (mean score 4.08, SD 0.79). Close to 75% of the respondents (n=339) indicated their willingness to try smart glasses if they were available to them (mean score 3.94, SD 0.93).

XSL•FO
RenderX

Explaining Anticipated Adoption of Smart Glasses: Model Comparison

To investigate the added value of the observed variables in the conceptual model, a comparison was made between validated TAM2 [52] constructs used to create a baseline and the additional social interaction variables in this study. The constructs "perceived usefulness" and "perceived ease of use" were adapted to the context of smart glasses, and subjective norm and image were included because they represent social interaction. Although the construct "attitude" was not included in the TAM, there was a reason to include this validated construct based on previous studies on the early phase of the diffusion of smart glasses.

Interpreting Selected TAM2 Effects: Intention to Use, Perceived Usefulness, Perceived Ease of Use, Subjective Norm, and Image

Measurement Model

The initial measurement model generated a good fit (χ^2_{142}=183.55; $\chi^{2/df}$=1.29; SRMR=0.03; TLI=0.99; RMSEA=0.03 [CI 0.01-0.04]). The internal consistency was above the aspirational level (α>.70).

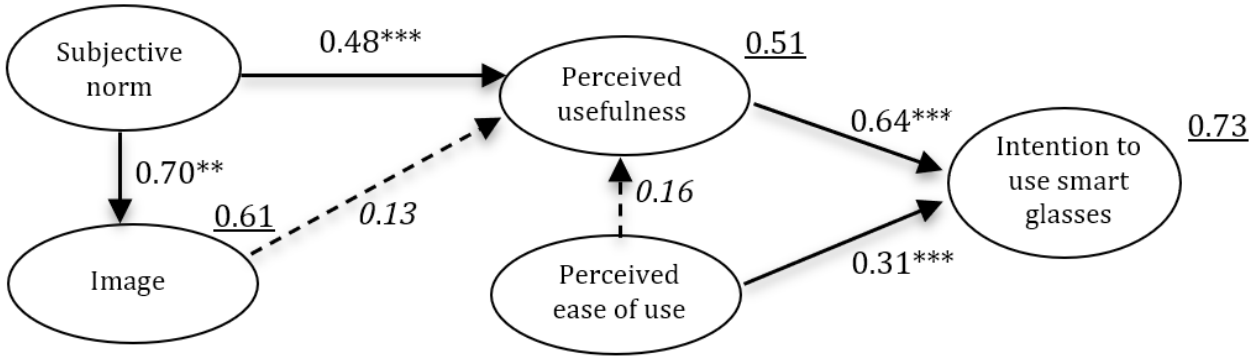
Structural Model

The results obtained from testing the validity of a causal structure of earlier validated variables showed a good fit (χ^2_{14}=218.314; $\chi^{2/df}$=1.52; SRMR=0.04; TLI=0.98; RMSEA=0.03 [CI 0.02-0.04]).

Path Model

The path model with standardized path coefficients is featured in Figure 5. The standardized path coefficients show significant direct effects of perceived usefulness and perceived ease of use on the intention to use smart glasses. Furthermore, subjective norm had a direct effect on perceived usefulness. Figure 5 also shows that subjective norm has a direct effect on image. The direct effect of perceived ease of use on perceived usefulness was not significant. The direct effect of image on perceived usefulness was also not significant. Subjective norm was the only contributor to perceived usefulness. In addition, perceived usefulness was the strongest contributor to the intention to use smart glasses. Squared multiple correlations showed that image accounted for 61%, perceived usefulness accounted for 51%, and intention to use smart glasses accounted for 73% of the explained variance.

Figure 5. Standardized path coefficients of the TAM2 variables for anticipated use of smart glasses by health care professionals (N=450). Dotted lines represent nonsignificant paths, italics represents nonsignificant factor loadings, and underline represents squared multiple correlations. TAM: Technology Acceptance Model. ***P<.001.



Interpreting Social Interaction Effects: Lack of Social Presence, Social Isolation, Decreased Attentional Allocation, External Location, Unfamiliar Person, Passive Use, Intention to Use Despite Nonacceptance by Others

Measurement Model

The initial measurement model generated a good fit (χ^2_{1461}=2146.696; $\chi^{2/df}$=1.47; SRMR=0.04; TLI=0.95; RMSEA=0.03 [CI 0.03-0.04]). The internal consistency was above the aspirational level (α>.70).

Structural Model

The results obtained from testing the validity of a causal structure of the conceptual model showed that the initial model did not fit the data (χ^2_{1525}=4896.900; $\chi^{2/df}$=3.21; SRMR=0.25; TLI=0.78; RMSEA=0.07 [CI 0.07-0.07]). Post-hoc modification indices suggested an improved fit by correlating error terms of perceived ease of use and social norm, social norm and image, lack of social presence and decreased attentional allocation, lack of social presence and social isolation, decreased attentional allocation and intention to use, external location and passive use, unfamiliar person and passive use, passive use and social norm, and passive use and intention to use despite social nonacceptance by others. Furthermore, post-hoc indication indices also suggested additional paths between perceived usefulness and intention to use, perceived ease of use and intention to use, perceived ease of use and image, perceived ease of use and social isolation, social norm and external location, social norm and lack of social presence, and lack of social presence and intention to use despite social nonacceptance by others. Some of the paths in the conceptual model were suggested differently, such as external location to intention to use instead of attitude, and the same applied for unfamiliar people and passive use. The paths suggested were from attitude to intention to use despite social nonacceptance by others and attitude to lack of social presence. The respecified model

XSL•FO
RenderX

generated a good fit (χ^2_{1337}=2113.035; $\chi^{2/df}$=1.58; SRMR=0.08; TLI=0.95; RMSEA=0.04 [CI 0.03-0.04]). Multimedia Appendix 2 summarizes the mean and SD, Cronbach α, factor loadings (β), and squared multiple correlations (R^2) of the observed variables to predict the anticipated intention to use smart glasses. The correlation matrix is available upon request.

Path Model

It was hypothesized that additional social interaction constructs could explain higher variance of users' anticipated intention to use compared to subjective norm, image, perceived usefulness, and perceived usefulness on anticipated intention to use (hypothesis 1). Squared multiple correlations are featured in Figure 6. It was shown that intention to use represented 96%, and it was raised with additional social interaction variables compared with squared multiple correlations, with representation of 73% from intention to use, perceived usefulness, perceived ease of use, subjective norm, and image. Attitude accounted for 89%, intention to use smart glasses despite social nonacceptance accounted for 65%, perceived usefulness accounted for 51%, and image accounted for 29%. These results supported the acceptance of hypothesis 1.

It was hypothesized that the user's attitudes directly influence the anticipated intention to use smart glasses (hypothesis 2). The standardized path coefficients in Figure 6 show a significant direct effect of attitude toward smart glasses on the intention to use smart glasses, which supported the acceptance of hypothesis 2.

It was stated that the lack of social presence due to wearing smart glasses during a physical contact moment influences the attitude directly (hypothesis 3). Figure 6 shows a significant negative direct effect of lack of social presence on attitude, and therefore, the results supported the acceptance of hypothesis 3. Surprisingly, the lack of social presence also had a positive direct effect on the intention to use despite social nonacceptance by patients.

Furthermore, it was hypothesized that social isolation due to wearing smart glasses during a physical contact moment influences the attitude directly. Social isolation did not have a significant effect on attitude (Figure 6), and this result supported the rejection of hypothesis 4.

In addition, it was hypothesized that the decreased attentional allocation due to wearing smart glasses during a physical contact moment influences the attitude directly (hypothesis 5). Figure 6 shows that decreased attentional allocation does not have a significant effect on attitude, and this result supported the rejection of hypothesis 5.

It was also hypothesized that the social acceptance of the use of smart glasses at an external location during a physical contact moment directly influences attitude (hypothesis 6). Figure 6 shows that use at an external location could not significantly explain attitude, and this result supported the rejection of hypothesis 6.

In addition, it was hypothesized that social acceptance regarding the use of smart glasses with an unknown person during a physical contact moment directly influences attitude. Figure 6 shows that there was no significant effect of the use of smart glasses with an unknown person on attitude, and this result supported the rejection of hypothesis 7.

Furthermore, it was hypothesized that the passive use of smart glasses during a physical contact moment directly influences attitude. Figure 6 shows that social acceptance regarding the passive use of smart glasses indeed directly influences attitude. Therefore, hypothesis 8 is accepted. In addition, social acceptance regarding the passive use of smart glasses has significant effects on the intention to use despite social nonacceptance by the patient, intention to use, and perceived usefulness.

Examining the nonacceptance of smart glasses by the patient, it was hypothesized that the intention to use despite the nonacceptance of smart glasses by the patient directly influences the anticipated intention to use (hypothesis 9). Figure 6 shows a significant effect, and therefore, hypothesis 9 is accepted. In addition, an overview of the accepted and rejected hypotheses can be found in Figure 7.

XSL•FO
RenderX

Figure 6. Standardized path coefficients of the adoption of smart glasses and the social interaction variables for the anticipated use of smart glasses by health care professionals (N=450). Dotted lines represent nonsignificant paths, italics represents nonsignificant factor loadings, and underline represents squared multiple correlations. TAM: Technology Acceptance Model. *P<.05; ***P<.001.

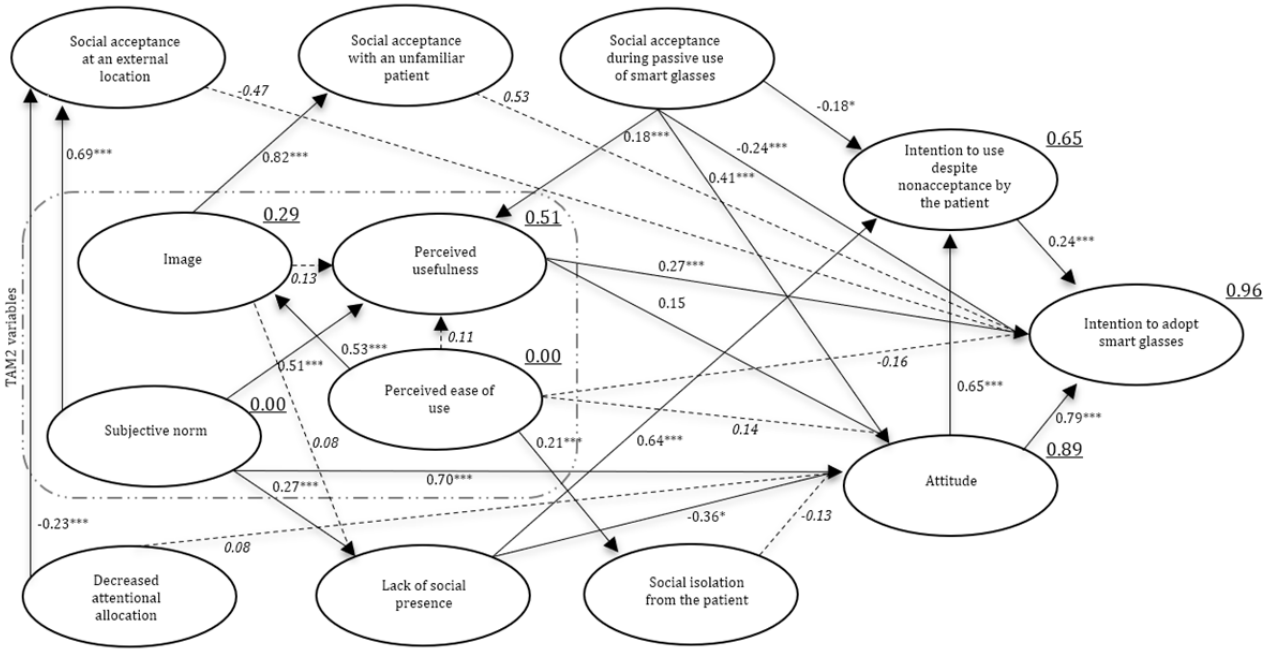
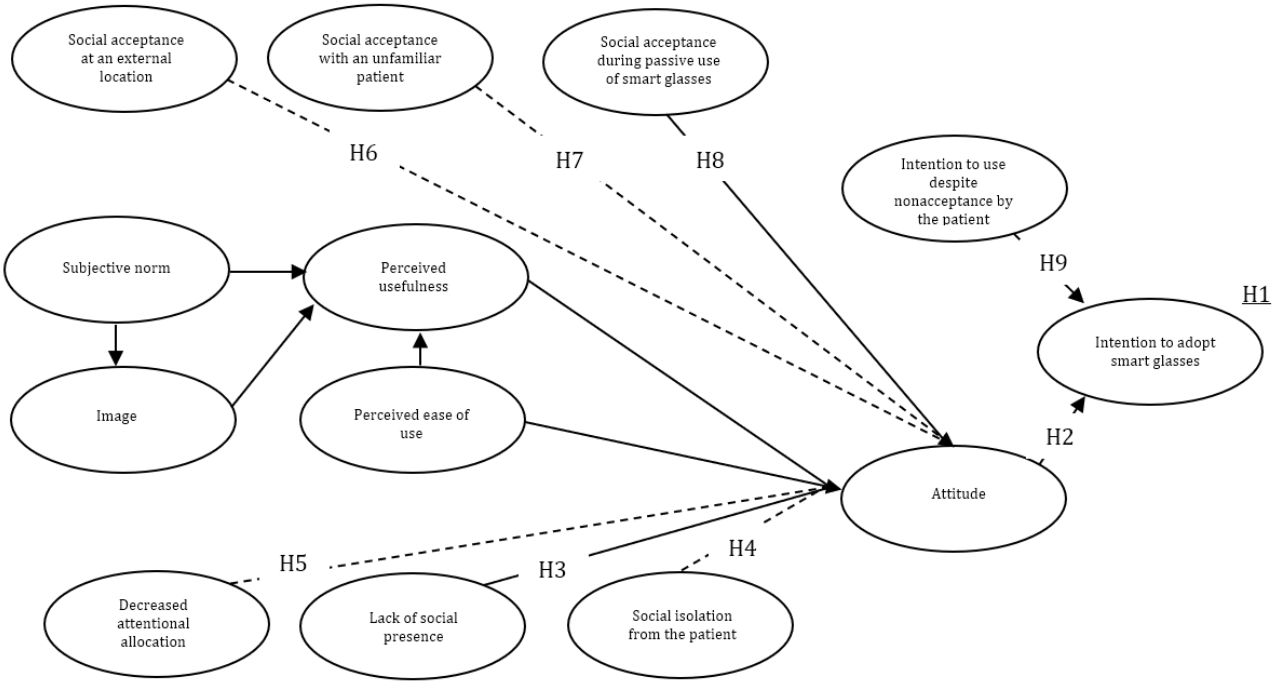


Figure 7. Overview of the tested hypotheses of the adoption of smart glasses and the social interaction variables for the anticipated use of smart glasses by health care professionals (N=450). Hypothesis 1 is underlined, and the hypothesis is accepted. Solid lines represent accepted hypotheses, and dotted lines represent rejected hypotheses. H: hypothesis.



Discussion

Overview

Considering the current state of the art and the objectives of this study, we formulated the following research question: "To what extent do constructs derived from a mediation perspective contribute to the established constructs of adoption to explain the intention to use smart glasses?" This study had 2 primary objectives. The first objective was to develop an instrument that combines the perspectives of adoption and mediation to measure the intention to use smart glasses by health care professionals. A conceptual model was created by drawing on previous findings related to smart glass adoption and testing hypotheses about their anticipated effects on attitudes and intentions to use. The instrument developed integrates both adoption and social interaction perspectives, highlighting smart glasses as a mediating factor between individuals, and it was validated with health care professionals. The approach for this instrument

XSL•FO
RenderX

builds on prior research and represents a step forward in understanding the role of smart glasses in health care.

The second objective was to obtain insights into how the intention to use is influenced by adoption and social interaction aspects. The structural equation analysis showed that integrating social interaction factors with selected TAM2 constructs led to a higher explained variance in the intention to use smart glasses than TAM2 alone. Key factors influencing the intention to use included attitude, perceived usefulness, intention to use despite social nonacceptance, and passive use of smart glasses. Social presence had an unexpected effect: as social presence increased, attitudes toward smart glasses decreased. Despite social nonacceptance by patients, health care professionals still showed willingness to use smart glasses, suggesting that this factor may be less decisive in health care. Passive use was linked to perceived usefulness and intention to use, indicating that wearing smart glasses on top of the head (as presented in Figure 4) could be convenient to have smart glasses close by, even when not actively in use. These findings suggest that social interaction plays a nuanced role in adoption, and further research is recommended to explore these relationships, especially regarding social presence and passive use.

Principal Findings

Development of the Instrument

Regarding the first objective of this study, there was an initial endeavor to create an instrument for investigating the anticipated adoption and mediation (as expressed through social interactions) of smart glasses within the health care sector. Consequently, a conceptual model was derived, which (1) drew upon prior findings related to the adoption of smart glasses, (2) tested hypotheses pertaining to the anticipated effects on attitudes and use intention, and (3) was applied to a sample of health care professionals for validation. We believe that the proposed conceptual model of expected adoption and social interaction provides a strong theoretical basis for further development of the model. The structural model in this study adds a higher explained variance but can be improved in the parsimony of the model. Furthermore, the focus in the model can be shifted in later research studies to other bystanders, such as colleagues, who may be confronted with a user of smart glasses.

The initial objective of this research was to develop an instrument that integrates the perspectives of both adoption and mediation. The resultant instrument was crafted to align closely with the state-of-the-art knowledge of smart glass adoption while simultaneously addressing aspects of social interaction. Central to this development was the assumption that smart glasses serve as a mediating factor between individuals. While there exist numerous methods to investigate social interaction factors, this study opted to ground its approach in findings from previous research.

Explaining the Anticipated Adoption of Smart Glasses

The second aim of this study was to gain insights into how the intention to use is influenced by adoption and social interaction aspects. The outcomes of the structural equation analysis revealed that our path model (Figure 6) integrating constructs from both TAM2 and social interaction variables demonstrated a higher explained variance in the anticipated intention to use smart glasses, compared to the isolated TAM2 constructs that were selected. This finding indicates that social interaction concerning smart glasses plays a role in anticipating the adoption of smart glasses in health care besides the perceived usefulness and attitude toward smart glasses. More specifically, the intention to use smart glasses is directly influenced by attitude, perceived usefulness, intention to use despite social nonacceptance, and passive use of smart glasses.

The presented model showed that attitude and intention to use both had high squared multiple correlations. This gives a reason to consider whether there is a theoretical difference between these constructs. In the Introduction section, it was also indicated that the attitude toward technology is a debatable construct in adoption studies because the added value in the model would not be substantial [10]. Original adoption models were developed for computer use in the 1980s, but smart glasses are different in several ways: they are worn on the body, are in the field of view, and often contain a camera. In this model, attitude is a mediating variable for subjective norm and the lack of social presence and was found to be important for the intention to use. Follow-up studies may investigate this relationship further.

Intention to use despite social nonacceptance shows that the influence of others is important for the health care professional but that it is not decisive to eventually use smart glasses. This can be understood because care actions can often be unpleasant for the patient but necessary for health purposes. This construct may score very differently in other disciplines where social relationships may be more important. Furthermore, the anticipated intention to use smart glasses is indirectly influenced by the lack of social presence and subjective norm.

A notable result of the lack of social presence is the decline in attitude. If the lack of social presence decreases (with an increase in social presence), the attitude toward smart glasses also decreases. This might indicate that respondents do not expect that smart glasses will ever provide more social presence and would not use them for that reason. One explanation that supports this finding is the positive regression coefficient with intention to use despite social nonacceptance by patients. This suggests that professionals would use smart glasses despite social nonacceptance by the patient.

The results of this study also indicate that there is a complex relationship between the passive use of smart glasses and 4 variables: a direct relationship with intention to use and an indirect relationship through perceived usefulness, attitude, and intention to use despite social nonacceptance. This can have 4 explanations, namely that passive use has a negative relationship with the intention to use, which indicates, for example, that health care professionals would rather not passively use smart glasses. However, this seems to contradict the finding of a weaker relationship with perceived usefulness, which could mean, for example, that it is convenient to have the glasses quickly at hand by wearing them on top of the head, just as a doctor places a stethoscope on the neck. Passive use also leads to an increased positive attitude toward glasses, which in turn leads to increased intention. In addition, there was a significant

XSL•FO
RenderX

negative regression between passive use and intention to use despite social nonacceptance by the patient. This could indicate, for example, that the health care professional is willing not to use technology or, in this case, is willing to remove it if the smart glasses are not actively used and the patient is not comfortable with them.

The second objective was to gain insights into how the intention to use smart glasses is influenced by the aspects of both adoption and social interaction. Based on our findings, it can be asserted that the presented model provides a higher explained variance for the intention to use smart glasses. The majority of the hypotheses were confirmed. However, some results were unexpected, such as the influence of social presence on attitudes, as discussed previously. Contrary to expectations drawn from prior research [31,53], we found that a lack of social presence did not result in a negative attitude or diminished intention to use. Therefore, we recommend further exploration in subsequent research, potentially using methodologies such as in-depth interviews and quasiexperimental studies. The same applies to the construct of social acceptance during the passive use of smart glasses. The model indicated that this factor also impacts perceived usefulness, suggesting that passive use of smart glasses could be beneficial in a work context and could consequently enhance the intention to use. Further research should shed more light on the underlying reasons for this finding.

Limitations

This study focused on the combination of adoption and social interaction. A limitation of this study lies in the comparison with the full TAM2 model. A priori–selected TAM2 constructs based on relevance and the aim of the study were included, namely to investigate social interaction and assess to what extent themes from the mediation perspective add something to current models. A comparison with the full TAM2 model is recommended for a follow-up study.

Second, in this study, social acceptance was examined through scenarios and not directly as a variable. The reason for this was to tie in with the previous findings and become more concrete in the specific context. This ignores social acceptance as an independent construct, and it is advisable to do so now that it appears that social interaction plays a substantial role in the model.

Third, as discussed in the conclusion, attitude and intention to use as variables score high on the explained variance, and it is questionable whether they are distinctive enough. Nevertheless, this difference is based on previous literature. Attitude was assumed as a preliminary state in the conceptual model, and it also fits the final model. It is recommended to investigate this further in a subsequent study, and it might be related to the last limitation.

The sample was limited to individuals who have access to MTurk. The distribution of respondents was mainly centered in the United States, and the questionnaire may have been completed mainly by people who were enthusiastic about the topic. As other researchers have addressed potential issues with sample representativeness and quality of the data [54,55], the results are more difficult to generalize and are not necessarily representative of the population of health care professionals. Therefore, it is important to revalidate the results with a reliable and representative dataset.

Furthermore, participants were presented with 2 videos to watch before proceeding further with the questionnaire. These videos, accessible to the public, were specifically chosen to provide a balanced depiction of what smart glasses are and their capabilities. It is worth noting, however, that as commercial videos, they may be designed to persuade viewers of the utility of smart glasses. The researchers selected these videos because they provided examples that were realistic and applicable in real-life work contexts at the time of the study and represented an attempt to prevent unrealistic imaging from smart glasses.

Although this model has been created based on previous literature and has remained as close as possible to the conceptual model, it is not a parsimonious model. From Cutting [56], a series of criteria can be derived to discuss the model evaluation, which is reflected by the accuracy, scope, and simplicity of the model. Parameter counting is widely used to indicate simplicity: if 2 models are both equally valid, the one with the fewest parameters is the best [56]. Several nonsignificant regression coefficients were found in the results of this study, and in combination with the previous recommendations, a subsequent simpler model can be developed.

Conclusions

Anticipating Adoption and Appropriation

The main finding is the answer to the following research question: “To what extent do constructs derived from a mediation perspective contribute to the established constructs of adoption to explain the intention to use smart glasses?” In the Introduction, the assumption was made that the target audience would likely be in an early stage of adoption of smart glasses. However, the responses gathered concerning knowledge and experience revealed that the majority of respondents demonstrated a fair understanding of smart glasses, and approximately half of them had already experimented with the technology.

The mediation perspective has provided insights into potentially useful variables for the anticipated adoption and social interaction of smart glasses. Social interaction with technology may be important because smart glasses are worn on the body. Venkatesh and Davis [52] found that the influence of subjective norm was only significant in mandatory usage contexts and not in voluntary contexts [52]. In this study, no obligatory or voluntary context was specifically referred to, yet subjective norm had a significant association with intention to use. That moderator can be added for control in a subsequent study, but it is also expected that with the current technology, such as smart glasses that are worn on the body, this is different than with personal computers. With smart glasses, there is human interaction where technology differently mediates in the physical world. In addition, social influences would have a nonsignificant influence on intention to use without moderators such as gender, age, experience, and use voluntariness [10]. This study showed that subjective norm and the lack of social presence influence

XSL•FO
RenderX

attitude and indirectly influence intention to use smart glasses in health care and can therefore make better predictions for anticipated use intention for smart glasses. Furthermore, despite many of the social issues addressed in the literature [21,26,31] and previous studies, most of the issues incorporated in this study did not have an effect on the anticipated intention to use smart glasses in the future. The findings reported here shed new light on the intention to use smart glasses in health care, where it could be argued that social interaction is perceived as an important aspect in providing care but not the most important construct for predicting future use. While this study does not provide evidence regarding the issues in social interaction surrounding smart glasses in health care as seen in the literature, a limitation in this study might be that respondents did not experience the use of smart glasses in real time. On the other hand, it may also be that the quality of care is more important than social interaction, and even if a patient feels uncomfortable when it comes to health, an intervention is sometimes necessary. Respondents might assume that smart glasses will help them in their work and that they will add value, regardless of the outcomes of social interaction.

Furthermore, although it is suggested that social acceptance is regularly included in evaluation sections in smart glasses research [18,57], it was observed that the adoption was only partly included in research designs in some studies. For example, in some cases, technology acceptance was evaluated by 3 constructs: perceived usefulness, perceived ease of use, and intention to use [58]. However, the first TAM [11] included attitude as well, and the more recent developments of the model included 13 factors influencing perceived usefulness and perceived ease of use and thereafter intention to use [15,59]. This example and current research in technology acceptance in related fields [60,61] show that there is more to study in adoption than 3 factors, and we argue to use all relevant aspects of adoption and acceptance frameworks to obtain more valid answers in acceptance research.

In summary, several theoretical perspectives have been used in this study, and it has been found that despite the concerns raised about social interaction with smart glasses, these aspects play important roles for health care professionals but are ultimately not decisive for the intention to use. As a result, there are fewer threats to implementation for health care than might be expected based on previous literature. It can also be concluded that the empirical findings of this study provide a new understanding of how the mediation perspective can increase the explained variance compared to existing constructs from the literature. This is why the mediation perspective adds new insights into a theoretical framework that can also be applied in other professional fields or related technologies.

Acknowledgments

This study would not have been possible without the respondents from a health care organization in the Netherlands and the workers of Amazon’s Mechanical Turk.

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

NZ led the design, data collection, and data analysis, and drafted the initial manuscript. SBA, OP, and PPV contributed to the design and data collection. OP contributed to data analysis as well. All authors reviewed the protocols, contributed to the conceptualization and methodology, and reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Prior research on the adoption of smart glasses.
[DOCX File , 50 KB - formative_v9i1e49610_app1.docx]

Multimedia Appendix 2
Item pool, descriptive statistics, factor loadings, squared multiple correlations, and Cronbach alpha of the observed indicators to explain and predict the anticipated use of smart glasses.
[DOCX File , 121 KB - formative_v9i1e49610_app2.docx]

Multimedia Appendix 3
Opening statement.
[DOCX File , 14 KB - formative_v9i1e49610_app3.docx]

References

XSL•FO
RenderX

1. Get Ready For Even More Google Glasshole Sightings. TechCrunch. 2013. URL: https://techcrunch.com/2013/01/28/glassholes/ [accessed 2024-12-26]
2. Google Glass advice: how to avoid being a glasshole. The Guardian. 2014. URL: https://www.theguardian.com/technology/2014/feb/19/google-glass-advice-smartglasses-glasshole [accessed 2024-12-26]
3. Hein DWE, Jodoin JL, Rauschnabel PA, Ivens BS. Are Wearables Good or Bad for Society?: An Exploration of Societal Benefits, Risks, and Consequences of Augmented Reality Smart Glasses. In: Wearable Technologies: Concepts, Methodologies, Tools, and Applications. Hershey, PA: IGI Global Scientific Publishing; 2018:1313-1337.
4. Hofmann B, Haustein D, Landeweerd L. Smart-glasses: exposing and elucidating the ethical issues. Sci Eng Ethics 2017 Jun 18;23(3):701-721. [doi: 10.1007/s11948-016-9792-z] [Medline: 27432401]
5. Kress B. See Through Optical Architectures for Wearable Displays. 2014 Presented at: Imaging and Applied Optics; July 13-17, 2014; Seattle, WA. [doi: 10.1364/AIO.2014.JTu1A.1]
6. Wortmeier A, Sousa Calepso A, Kropp C, Sedlmair M, Weiskopf D. Configuring augmented reality users: analysing YouTube commercials to understand industry expectations. Behaviour & Information Technology 2023 Jan 09;43(2):371-386. [doi: 10.1080/0144929X.2022.2163693]
7. Kolodzey L, Grantcharov PD, Rivas H, Schijven MP, Grantcharov TP. Wearable technology in the operating room: a systematic review. BMJ Innov 2016 Dec 21;3(1):55-63. [doi: 10.1136/bmjinnov-2016-000133]
8. Zuidhof N, Ben Allouch S, Peters O, Verbeek P. Defining smart glasses: a rapid review of state-of-the-art perspectives and future challenges from a social sciences' perspective. Augment Hum Res 2021 Oct 31;6(1):15. [doi: 10.1007/s41133-021-00053-3]
9. Lu J, Yao JE, Yu C. Personal innovativeness, social influences and adoption of wireless Internet services via mobile technology. The Journal of Strategic Information Systems 2005 Sep;14(3):245-268. [doi: 10.1016/j.jsis.2005.07.003]
10. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
11. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Management Science 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
12. Rogers E. Diffusion of Innovations, 3rd edition. New York, NY: Free Press; 1983.
13. Akbari M, Rezvani A, Shahriari E, Zúñiga M, Pouladian H. Acceptance of 5 G technology: mediation role of trust and concentration. Journal of Engineering and Technology Management 2020;57:101585 [FREE Full text] [doi: 10.1016/j.jengtecman.2020.101585]
14. Lee Y, Kozar KA, Larsen KR. The technology acceptance model: past, present, and future. CAIS 2003;12:1250. [doi: 10.17705/1CAIS.01250]
15. Venkatesh V, Bala H. Technology acceptance model 3 and a research agenda on interventions. Decision Sciences 2008 May 09;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
16. Koelle M, El Ali A, Cobus V, Heuten W, Boll S. All about Acceptability?: Identifying Factors for the Adoption of Data Glasses. In: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: 2017 CHI Conference on Human Factors in Computing Systems; May 6-11, 2017; Denver, CO. [doi: 10.1145/3025453.3025749]
17. Rauschnabel PA, He J, Ro YK. Antecedents to the adoption of augmented reality smart glasses: A closer look at privacy risks. Journal of Business Research 2018 Nov;92:374-384. [doi: 10.1016/j.jbusres.2018.08.008]
18. Nam C, Lee Y. Validation of the wearable acceptability range scale for smart apparel. Fash Text 2020 May 05;7(1):13. [doi: 10.1186/s40691-019-0203-3]
19. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
20. Bagozzi R. The legacy of the technology acceptance model and a proposal for a paradigm shift. JAIS 2007 Apr;8(4):244-254. [doi: 10.17705/1jais.00122]
21. Koelle M, Kranz M, Möller A. Don't look at me that way!: Understanding User Attitudes Towards Data Glasses Usage. In: MobileHCI '15: Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services. 2015 Presented at: 17th International Conference on Human-Computer Interaction with Mobile Devices and Services; August 24-27, 2015; Copenhagen, Denmark. [doi: 10.1145/2785830.2785842]
22. Rauschnabel P, Brem A, Ivens B. Mixing Physical and Virtual Realities: Implications of Augmented Reality Smart Glasses for the Management of Innovation. 2015 Presented at: AMA Summer Marketing Educators' Conference 2015; August 14-18, 2015; Chicago, IL.
23. Due B. The social construction of a Glasshole: Google Glass and multiactivity in social interaction. PsychNology Journal 2015;13(2-3):149-178 [FREE Full text]
24. Liinasuo M, Kuula T, Goriachev V, Helin K. Choosing between remote and face-to-face features in a test setting: methodology used in two usability and user experience case studies. Behaviour & Information Technology 2023 Jan 13;42(3):298-306. [doi: 10.1080/0144929x.2023.2165453]

XSL•FO
RenderX

25. Goken M, Basoglu A, Dabic M. Exploring adoption of smart glasses: Applications in medical industry. 2016 Presented at: Portland International Conference on Management of Engineering and Technology (PICMET); September 04-08, 2016; Honolulu, HI. [doi: 10.1109/picmet.2016.7806835]
26. Montero C, Alexander J, Marshall M, Subramanian S. Would you do that?: understanding social acceptance of gestural interfaces. In: MobileHCI '10: Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services. 2010 Presented at: 12th International Conference on Human Computer Interaction with Mobile Devices and Services; September 7-10, 2010; Lisbon, Portugal. [doi: 10.1145/1851600.1851647]
27. Weiser M, Gold R, Brown JS. The origins of ubiquitous computing research at PARC in the late 1980s. IBM Syst. J 1999;38(4):693-696. [doi: 10.1147/sj.384.0693]
28. Mann S. What I've learned from 35 years of wearing computerized eyewear. IEEE Spectr. 2013. URL: https://www.nxtbook.com/nxtbooks/ieee/spectrum_na_0313/index.php?startid=23#/1 [accessed 2024-12-26]
29. Verbeek P. Beyond interaction: a short introduction to mediation theory. Interactions 2015 Apr 27;22(3):26-31. [doi: 10.1145/2751314]
30. Zuidhof N, Ben AS, Peters O, Verbeek P. A theoretical framework to study long-term use of smart eyewear. In: UbiComp/ISWC '19 Adjunct: Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers. 2019 Presented at: 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and ACM International Symposium on Wearable Computers; September 9-13, 2019; London, UK. [doi: 10.1145/3341162.3348382]
31. Zuidhof N, Ben Allouch S, Peters O, Verbeek P. Perspectives on the acceptance and social implications of smart glasses: a qualitative focus group study in healthcare. International Journal of Human–Computer Interaction 2022 Aug 26;40(2):149-159. [doi: 10.1080/10447318.2022.2111046]
32. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
33. Zuidhof N, Ben Allouch S, Peters O, Verbeek P. Anticipated Acceptance of Head Mounted Displays: a content analysis of YouTube comments. 2019 Presented at: IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops); March 11-15, 2019; Kyoto, Japan. [doi: 10.1109/percomw.2019.8730658]
34. Han DD, Tom Dieck MC, Jung T. Augmented Reality Smart Glasses (ARSG) visitor adoption in cultural tourism. Leisure Studies 2019 Apr 21;38(5):618-633. [doi: 10.1080/02614367.2019.1604790]
35. Harms C, Biocca F. Internal consistency and reliability of the networked minds social presence measure. In: Alcaniz M, Rey B, editors. Seventh Annual International Workshop: Presence. Valencia: Universidad Politecnica de Valencia; 2004:246-251.
36. Karahanna E, Limayem M. E-mail and v-mail usage: generalizing across technologies. J of Organizational Comp & Elec Commerce 2000 Mar 1;10(1):49-66. [doi: 10.1207/s15327744joce100103]
37. Tenner E. Why Things Bite Back, Technology and the Revenge of Unintended Consequences. New York, NY: Vintage Books; 1996.
38. Verbeek P. Op de vleugels van Icarus. Hoe techniek en moraal met elkaar meebewegen. Rotterdam: Lemniscaat; 2014.
39. Merton RK. The unanticipated consequences of purposive social action. American Sociological Review 1936 Dec;1(6):894. [doi: 10.2307/2084615]
40. Spector P. Summated Rating Scale Construction, 82nd edition. Newbury Park, CA: Sage; 1992.
41. Tyupa S. A theoretical framework for back-translation as a quality assessment tool. New Voices in Translation Studies 2011;7(1):35-46 [FREE Full text]
42. Gemvision Healthcare platform | Augmented Reality Smart Glasses. YouTube. URL: https://youtu.be/N_QoKStFn_Q [accessed 2024-12-26]
43. M400 Telemedicine Medical. YouTube. URL: https://youtu.be/TyA_NVyK9iw [accessed 2024-12-26]
44. George D, Mallery M. SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 Update. Boston, MA: Allyn & Bacon, Inc; 2010.
45. Jöreskog KG. A general method for estimating a linear structural equation system*. ETS Research Bulletin Series 2014 Aug 08;1970(2):i-41. [doi: 10.1002/j.2333-8504.1970.tb00783.x]
46. Kline R. Principles and practice of structural equation modeling, 4th edition. New York, NY: Guilford Press; 2016.
47. Holbert RL, Stephenson MT. Structural equation modeling in the communication sciences, 1995-2000. Human Comm Res 2002 Oct;28(4):531-551. [doi: 10.1111/j.1468-2958.2002.tb00822.x]
48. Hoyle R, Panter A. Writing about structural equation models. In: Hoyle RH, editor. Structural equation modeling: Concepts, issues, and applications. Thousand Oaks, CA: Sage Publications; 1995:158-176.
49. Hooper D, Coughlan J, Mullen M. Structural equation modelling: guidelines for determining model fit. Electronic Journal of Business Research Methods 2008;6(1):53-60 [FREE Full text]
50. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
51. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research 1981 Feb;18(1):39. [doi: 10.2307/3151312]

XSL•FO
RenderX

52. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Management Science 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
53. Uhde A, Tretter S, von Terzi P, Koelle M, Diefenbach S, Hassenzahl M. Interaction in the Public: Aesthetics, Social Acceptability, and Social Context. Mensch und Computer 2021, Workshopband, Workshop on Interaction in the Public: Aesthetics, Social Acceptability, and Social Context. 2021. URL: https://alarithuhde.com/publications/Workshop-Interaction-in-the-Public.pdf [accessed 2024-12-26]
54. Aguinis H, Villamor I, Ramani RS. MTurk research: review and recommendations. Journal of Management 2020 Nov 19;47(4):823-837. [doi: 10.1177/0149206320969787]
55. Paolacci G, Chandler J. Inside the Turk. Curr Dir Psychol Sci 2014 Jun 03;23(3):184-188. [doi: 10.1177/0963721414531598]
56. Cutting JE. Accuracy, scope, and flexibility of models. J Math Psychol 2000 Mar;44(1):3-19. [doi: 10.1006/jmps.1999.1274] [Medline: 10733855]
57. Lee L, Hui P. Interaction methods for smart glasses: a survey. IEEE Access 2018;6:28712-28732. [doi: 10.1109/access.2018.2831081]
58. Lam K, Hang LL, Braud T, Hui P. M2A: A framework for visualizing information from mobile web to mobile augmented reality. 2019 Presented at: 2019 IEEE International Conference on Pervasive Computing and Communications (PerCom); March 11-15, 2019; Kyoto, Japan. [doi: 10.1109/percom.2019.8767388]
59. Lai P. The literature review of technology adoption models and theories for the novelty technology. JISTEM USP 2017 Jun 08;14(1):21-38. [doi: 10.4301/S1807-17752017000100002]
60. Ben Allouch S. The design and anticipated adoption of ambient intelligence in the home. University of Twente. 2008. URL: https://ris.utwente.nl/ws/files/6085272/thesis_Ben_Allouch.pdf [accessed 2024-12-26]
61. de Graaf M, Ben Allouch S, van Dijk J. Why Do They Refuse to Use My Robot?: Reasons for Non-Use Derived from a Long-Term Home Study. In: HRI '17: Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction. 2017 Presented at: 2017 ACM/IEEE International Conference on Human-Robot Interaction; March 6-9, 2017; Vienna, Austria. [doi: 10.1145/2909824.3020236]

Abbreviations

RMSEA: root mean square error of approximation
SRMR: standardized root mean square residual
TAM: Technology Acceptance Model
TAM2: Technology Acceptance Model 2
TLI: Tucker-Lewis Index
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by A Mavragani; submitted 03.06.23; peer-reviewed by D Nogueira-Leite, C Romare; comments to author 19.07.23; revised version received 01.08.23; accepted 28.11.24; published 11.02.25.

Please cite as:
Zuidhof N, Peters O, Verbeek PP, Ben Allouch S
Social Acceptance of Smart Glasses in Health Care: Model Evaluation Study of Anticipated Adoption and Social Interaction
JMIR Form Res 2025;9:e49610
URL: https://formative.jmir.org/2025/1/e49610
doi:10.2196/49610
PMID:

©Niek Zuidhof, Oscar Peters, Peter-Paul Verbeek, Somaya Ben Allouch. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Designing Patient-Centered Interventions for Emergency Care: Participatory Design Study

Woosuk Seo[1], BS; Shruti Jain[2], BS; Vivian Le[1], BS; Jiaqi Li[1], MS; Zhan Zhang[3], PhD; Hardeep Singh[4], MD, MPH; Kalyan Pasupathy[5], PhD; Prashant Mahajan[6], MPH, MBA, MD; Sun Young Park[7], PhD

[1]School of Information, University of Michigan, Ann Arbor, MI, United States

[2]Computer Science and Engineering, University of Michigan, Ann Arbor, United States

[3]Seidenberg School of Computer Science and Information Systems, Pace University, New York, NY, United States

[4]Center for Innovations in Quality, Effectiveness and Safety (IQuESt), Michael E. DeBakey Veterans Affairs Medical Center and Baylor College of Medicine, Houston, TX, United States

[5]Biomedical and Health Information Sciences, University of Illinois Chicago, Chicago, IL, United States

[6]Department of Emergency Medicine, University of Michigan Medical School, Ann Arbor, MI, United States

[7]School of Information, Stamps School of Art and Design, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Woosuk Seo, BS
School of Information
University of Michigan
105 S State St
Ann Arbor, MI
United States
Phone: 1 2063109264
Email: seow@umich.edu

Abstract

Background: Emergency departments (EDs) are high-pressure environments where clinicians diagnose patients under significant constraints, including limited medical histories, severe time pressures, and frequent interruptions. Current ED care practices often inadequately support meaningful patient participation. Most interventions prioritize clinical workflow and health care provider communication, inadvertently overlooking patients' needs. Additionally, patient-facing technologies in EDs are typically developed without meaningful patient input, leading to solutions that may not effectively address patients' specific challenges. To enhance both patient-centered care practices and the diagnosis process in EDs, patient involvement in technology design is essential to ensure their needs during emergency care are understood and addressed.

Objective: This study aimed to invite ED patients to participatory design sessions, identify their needs during ED visits, and present potential design guidelines for technological interventions to address these needs.

Methods: We conducted 8 design sessions with 36 ED patients and caregivers to validate their needs and identify considerations for designing patient-centered interventions to improve diagnostic safety. We used 10 technological intervention ideas as probes for a needs evaluation of the study participants. Participants discussed the use cases of each intervention idea to assess their needs during the ED care process. We facilitated co-design activities with the participants to improve the technological intervention designs. We audio- and video-recorded the design sessions. We then analyzed session transcripts, field notes, and design sketches.

Results: On the basis of ED patients' feedback and evaluation of our intervention designs, we found the 3 most preferred intervention ideas that addressed the common challenges ED patients experience. We also identified 4 themes of ED patients' needs: a feeling of inclusion in the ED care process, access to sources of medical information to enhance patient comprehension, addressing patient anxiety related to information overload and privacy concerns, and ensuring continuity in care and information. We interpreted these as insights for designing technological interventions for ED patients. Therefore, on the basis of the findings, we present five considerations for designing better patient-centered interventions in the ED care process: technology-based interventions should (1) address patients' dynamic needs to promote continuity in care; (2) consider the amount and timing of information that patients receive; (3) empower patients to be more active for better patient safety and care quality; (4) optimize human resources, depending on patients' needs; and (5) be designed with the consideration of patients' perspectives on implementation.

XSL•FO
RenderX

Conclusions: This study provides unique insights for designing technological interventions to support ED diagnostic processes. By inviting ED patients into the design process, we present unique insights into the diagnostic process and design considerations for designing novel technological interventions to enhance patient safety.

International Registered Report Identifier (IRRID): RR2-10.2196/55357

(JMIR Form Res 2025;9:e63610) doi:10.2196/63610

KEYWORDS

emergency department; participatory design; patient; technology; intervention

Introduction

Background

In the high-pressure environment of emergency departments (EDs), diagnosing and managing patients is inherently complex. This complexity arises from extensive interactions among various stakeholders, including health care providers, family caregivers, and patients. The dynamic and time-sensitive context of EDs forces stakeholders to make critical medical decisions promptly. The urgent nature of emergency care thus poses substantial challenges to ED providers, compounding the difficulty of achieving timely and accurate diagnoses and raising concerns regarding patient safety. While exact rates of diagnostic errors remain elusive, a conservative estimate of 5% errors in adults out of the 131 million annual ED visits translates to about 7 million cases of ED-based diagnostic errors [1]. Notably, nearly half of these diagnostic errors have the potential to cause harm to patients.

Many existing technological interventions focus on supporting the diagnostic process and promoting patient safety in the ED. For example, an ED provider-facing mobile app currently exists to deliver real-time patient information and allows for collaborative patient management for care teams [2]. Artificial intelligence (AI)–based software [3] has also been developed to provide practical guidance to novice nurses or scan operators in the ED who have limited experience in echocardiography. In addition, AI-based systems [4,5] have been used to support documentation for the nursing team, indirectly promoting the efficiency of the care process. Such systems could support providers by easing the documentation burden (eg, parsing long notes and automatically populating relevant text) so that they could focus more on patient care.

On the other hand, some technological interventions have been developed that focus more on ED patients. An indoor navigation system [6] was developed to assist patients unfamiliar with the ED environment, and a mobile display [7] was designed to present an interactive report detailing patients' progress, care plans, and care teams throughout their stay in the ED. To reduce patients' anxiety, some prior studies highlighted opportunities to use music apps [8] or virtual reality applications [9]. More recently, a study presented the opportunity for an AI-based application to provide medical triage to patients more efficiently [10]. Compared with human clinicians, the AI-based application made probable triage decisions. While these technological interventions improve providers' diagnosing work or patient safety, direct input from patients and providers working at the frontline is often overlooked when developing such solutions.

Participatory design (PD) [11] is a methodology that engages all stakeholders in the design process to create solutions that address their needs. PD has been widely used by researchers in prior studies to co-design user-centered technologies with health care providers. For instance, Kusunoki et al [12] conducted PD workshops with trauma team members to understand the different needs of awareness support among the various roles of team members and identify concrete design strategies to manage these differences. Pollack et al [13] organized a design session with 11 clinicians to develop a clinical information tool using PD techniques. On the basis of the session's findings, the authors identified benefits (eg, a high level of domain knowledge can be used to anticipate how design ideas can be applied to clinical processes and workflow) and potential challenges (eg, power dynamics between physicians) of leveraging PD techniques in designing a clinical information tool. The authors also outlined guiding principles for implementing these methods in health care organizations interested in advancing health information technology. These prior studies have presented how PD is helpful and practical for designing human-centered technology in health care settings.

Despite the benefits of PD in designing health technology, limited work has adopted a user-centered PD approach to develop technological interventions for ED specifically to enhance patient safety. Østervang et al [14] conducted PD workshops with health care providers and patients to design an ED information system. In the study, providers brainstormed initial design ideas for an ED information system in a workshop; then, patients provided feedback in separate workshops. The authors presented how the PD approach helps yield insights from ED providers and patients to create a more user-centered system. However, ED patients in the study had limited participation because the workshops were conducted one-on-one, and patients were only asked to provide feedback on intervention ideas developed by health care providers.

Objectives

Building on previous research, we aimed to engage ED patients in validating their needs and co-designing novel interventions to promote patient safety and enhance the ED diagnostic process. Because patients are vulnerable to potential errors, their involvement in developing patient-centered interventions is important to improve the diagnostic process. Through identifying patients' needs during their ED visits, this study presents design guidelines for more patient-centered technological interventions. We used preliminary technological intervention concepts to facilitate discussions about participants' experiences and needs. This approach helped expand our understanding of patient experiences in the ED. Rather than

XSL•FO
RenderX

proposing specific solutions, we aimed to refine intervention concepts based on patient input and establish broader guidelines for designing patient-facing interventions.

Methods

Overview

This study is part of a larger project that aims to identify ED patients' and health care providers' perspectives on technological interventions to support the ED diagnostic process and to craft design guidelines for interventions that meet both stakeholders' needs. In this paper, we focus on validating the needs of ED patients and caregivers during ED visits and their perspectives on technological interventions. We conducted 8 design sessions with 36 ED patients or their caregivers. On the basis of the analysis of transcripts, design sketches, and field notes, we identified 4 themes representing patient needs, strategies to mitigate common challenges, and design guidelines for potential interventions.

Design Idea Generation Phase for ED Care Interventions

Before our PD study with ED patients, the research team interviewed 8 patients and caregivers to better understand their experiences and challenges during ED visits. From the interviews, we identified difficulties, emerging patterns of complaints, and general levels of satisfaction with different aspects of the care process (Table 1). Common themes were discovered from each problem category through analysis of the interview data ranging from patient and caregiver difficulties, complaints, and satisfaction levels with various aspects of the ED care process. In addition to the known problems, such as long wait times and insufficient or poor communication with providers, ED patients and caregivers also faced difficulties because of the absence of caregivers, lack of ED process literacy, or inadequate understanding of discharge information. On the basis of the findings, the research team brainstormed numerous design ideas for each problem category, focusing on the stakeholders (eg, patients, caregivers, and providers). We then merged the design ideas based on feasibility and usefulness. Finally, we narrowed the list and finalized the 10 most effective intervention ideas (Table 2). These ideas were used to validate patient needs during each PD session, presented as storyboards, as shown in Figure 1. Each intervention idea aims to address at least one problem category.

Figure 1. Two sample storyboards of technological intervention ideas presented to the participants during the sessions (top: table-top robot; bottom: virtual specialist). ED: emergency department.

XSL•FO
RenderX

Table 1. A list of identified problem categories from the previous study's patient interview data.

Problem category	Examples
Waiting room general challenges	• Patients feel forgotten in the waiting room because health care providers do not check up on them • Patients are left in pain in the waiting room
Overcrowding challenges	• There are patients with mild symptoms who make the ED[a] more crowded
Information presentation and overload challenges	• Patients are inadequately informed of the ED process during the visit (ie, lack of ED process literacy) • Test results can be confusing for patients to understand because of the use of medical terminology
Information sharing challenges	• Patients often feel that there are errors in the physician's notes (eg, it is different from what was discussed in the visit) • Health care providers at hospitals sometimes wait for information transfer from the patient's original hospital, which can take time
Communication challenges	• Patients find it hard to recount all personal medical details in the ED because they are sick, stressed, or unaware • Patients sometimes have trouble communicating with providers because of challenges like being a nonnative English speaker
Contagious disease transfer in ED	• Crowded waiting rooms can have many sick patients in close proximity, which makes patients concerned • Contact with surfaces can cause disease transfer
Discharge and postvisit challenges	• Lack of postvisit resources for patients • Patients feel that they have been wrongly discharged

[a]ED: emergency department.

As probes, the storyboards incorporated user requirements and insights about patient-centered interventions to help participants better understand the use cases and contexts where the proposed interventions could be implemented. Each storyboard consisted of a 3-panel illustration, describing an example of a patient facing one of the challenges mentioned above and using the intervention to mitigate or prevent the problem (see Figure 1 for sample storyboards). Details of the intervention's features and functionality were left vague. We wanted participants to focus on the need rather than specific features to encourage more discussion about improving the intervention concept. Each storyboard included a lead question to highlight the user's needs (ie, patients) outlined in each intervention and to facilitate participant discussion about whether they had experienced the need, rather than directly referencing the intervention. Using storyboards allowed participants to have less biased and more open-ended discussions in the PD sessions. Table 2 describes each technological intervention and the patient need it addresses, along with storyboard samples.

XSL•FO
RenderX

Table 2. A list of 10 intervention ideas with brief descriptions and discussion questions to validate patient needs.

Intervention number	Technological intervention	Description	Sample of lead question used for the discussion of patient need
I-1	Virtual presence of a caregiver	A communication device for ED[a] patients that enables the virtual presence of the caregiver for the patient and their ED provider when the caregiver cannot physically be with the patient	Have you experienced any difficulty visiting the ED alone without a caregiver?
I-2	Transparent interactive display for ED-related guidance	An interactive display for ED patients in a waiting room that displays ED-related guidance and acts as a separator between patients for personal hygiene	During your ED visit, have you worried that you might be coming in contact with contagious diseases while in the waiting room?
I-3	Table-top robot for distraction and entertainment	A table-top robot for ED patients in the waiting room that provides distraction and entertainment as they wait	Have you felt distraction and entertainment is helpful during a long wait in the ED?
I-4	Wearable device for monitoring real-time condition	A wearable device for ED patients in the waiting room that tracks their real-time data (eg, vitals, such as heart rate and blood oxygen) and notifies providers if vitals indicate a patient needs immediate help	Have you felt worried about your condition worsening while waiting, and being unable to inform healthcare providers properly?
I-5	Assistant robot in the waiting room	A mobile robot that assists and checks on ED patients in the waiting room	Have you felt forgotten while waiting because of busy ED providers not checking up on your condition?
I-6	ED-specific chatbot for ED practice literacy	A mobile device–compatible chatbot that can provide ED-specific information directly from the hospital to patients	Have you ever felt confused about the meaning of the medical information you received during ED care?
I-7	Virtual communication with a specialist	A communication device for ED patients and providers that enables a virtual consultation with a specialist during ED care	Have you felt the need to consult with a specialist during your ED visit?
I-8	Pain expression device	A tactile device that helps ED patients understand and describe pain through vibrations and haptic feedback	Have you had any difficulties in describing your pain verbally?
I-9	Patient-facing screen for sharing physician's notes	A screen display that presents care-related information for patients to view during consultation or follow-up with providers about diagnosis or test results	Have you ever noticed errors or confusing details on physician's notes (usually on discharge notes or after-visit summaries)?
I-10	AI[b] caller for providing after-visit support	A follow-up call powered by AI to help answer patients' postdischarge questions regarding their ED visit	Have you had any challenges or questions that needed to be properly addressed after your ED visit?

[a]ED: emergency department.

[b]AI: artificial intelligence.

Participant Recruitment

We recruited patients and caregivers who had visited the ED within 6 weeks from the point of contact. Participants were mainly recruited from the adult and pediatric EDs at a university-affiliated hospital in 3 ways: calling or emailing them after their most recent ED visit, posting fliers with the study's contact information in the adult ED and the pediatric ED, and approaching patients during their ED visit. The inclusion criteria for participants were as follows: (1) aged ≥18 years old, (2) previously visited ED within 6 weeks from the point of contact, and (3) comfortable speaking English. During a patient's ED visit, researchers mainly focused on recruiting those having a triage level of 3 to 5, with 1 being critical and 5 being nonurgent, to avoid disrupting or risking their care. We contacted more than 150 potential participants via email and phone, as participant retention was challenging because of last-minute cancellations from patients. In total, 36 ED patients and caregivers participated in 8 design sessions (see Table 3 for their demographic information). All but one session had 4 or 5 participants, and one session had 3 participants. As compensation for participation, each participant was provided with a US $100 gift card for the session.

XSL•FO
RenderX

Table 3. Demographic information of study participants (N=36).

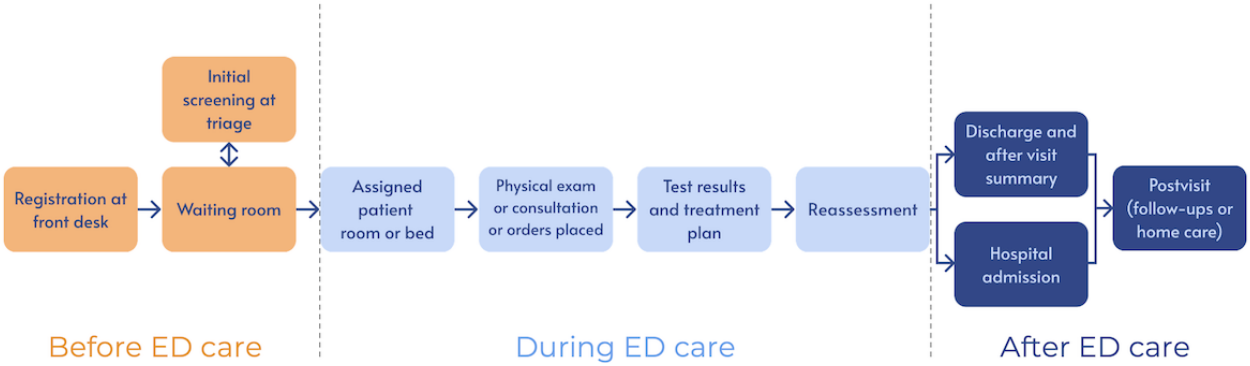
Participant demographics	Participants, n (%)
Race	
Asian	1 (3)
Black	7 (19)
White	27 (75)
Unknown	1 (3)
Sex	
Male	12 (33)
Female	24 (67)
Participant type	
Patient	23 (64)
Caregiver	13 (36)

PD Session Procedure

We conducted 8 PD sessions. Each session was roughly divided into three parts: (1) session introduction, (2) storyboard critique, and (3) co-design as a group. Each session took approximately 2 hours.

In the introductory session, participants were each given a paper packet outlining the current ED care process timeline, the 10 technological intervention storyboards for reference, and a sheet to rank the interventions based on their preferences. A presentation of the ED care process timeline (Figure 2) was shown to prompt participants to recall their previous or most recent ED experience. We created this timeline based on our last interview study [15] and prior studies on the ED care framework [16]. We first asked participants to write down one major challenge they encountered during this experience and specifically indicate where it occurred in the current ED care process timeline in their packet. Each participant was then asked to discuss their experiences with the group, establishing the foundation for the main session activities. Allowing participants to reflect on their previous or most recent ED visit prefaced discussions about how their challenges might or might not be addressed with one or more of the 10 technological interventions and the co-design activity.

Figure 2. A timeline of the emergency department (ED) care process illustrating steps before, during, and after care. This timeline, based on prior studies, was provided to the participants as a reference for recalling their ED visits.



For the storyboard critique activity, 10 intervention design ideas were presented via storyboards to prompt participant discussion about each intervention's feasibility in the current ED care process. Researchers encouraged participants to initiate the conversation with lead questions and additional follow-up questions related to each scenario were asked, as necessary, to maintain the discussion within the group. The discussion of the 10 technological interventions provided a space for participants to speak about their experiences and needs during ED visits and for researchers to better understand the underlying challenges patients face. After all the intervention storyboards were presented, participants ranked their top 3 preferred interventions based on preference and potential feasibility in ED care. While the main task for each participant was to consider the interventions from the patient or caregiver perspective, participants naturally framed these interventions from the provider's perspective as well. We did not set any restrictions regarding this task during the discussions and design activities to solicit as much feedback as possible regarding the feasibility of each intervention in the ED care context.

Each group of participants discussed their most preferred 3 interventions and then selected one for the co-design activity. Then, as a group, participants had approximately 25 to 30 minutes to augment their chosen technological intervention to improve the ED experience (Figure 3). Participants were tasked with adding, changing, and removing features or modifying the context of the technological intervention to help it better align

XSL•FO
RenderX

with their preferred, ideal future implementation. Researchers emphasized the importance of focusing on the roles and functionality of technological interventions rather than the aesthetic quality of the designs. The co-design activity helped us understand the features of the intervention technologies participants desired to address patient needs, along with other fundamental and feature-specific aspects that may not have been revealed in the storyboard critique. Through this, participants could also discuss their decisions in detail, out loud, and reach a consensus about the most critical design features to incorporate into an ideal intervention solution. Following the co-design activity, participants were asked to provide a summary presentation to the researchers and explain their reasoning behind the design decisions they made together.

Figure 3. Co-design activity with the participants in the middle of a discussion about improving idea 7 (I-7 in Table 2). Participants collaboratively drew ideas on a whiteboard and wrote notes about imagined functionality.

Data Analysis

During the PD sessions, at least 2 researchers presented and took separate notes on points discussed among groups and overall interpretations of participant reactions to ideas. After every session, the researchers collaborated to reflect on session notes and compiled a document with observations on critical insights and sentiments toward each design idea. For each PD session, a researcher also took pictures of the sketches created during the design activity and on participant handouts, which contained participant notes, feedback, and intervention rankings. We used transcripts as the primary data source, and all other collected data were used as supplementary sources. We also incorporated the co-design activity sketches as supplementary data to validate what participants discussed during the session and clarify participants' design suggestions. The above data collection was conducted with the participant's permission.

After organizing the collected data, we coded the first 3 PD transcripts. At least 2 researchers coded each transcript, and each code was reviewed to account for missed information. The codes were compiled into a growing codebook organized by their corresponding intervention or put into a general category. On the basis of the codebook, the remaining transcripts were then coded. If any new codes were found, they were either added to the codebook as a new code or incorporated into the definition of an existing similar code.

We then used affinity diagramming [17] to identify the emerging themes from the codes. Similar codes from all 10 interventions were first grouped together into subthemes specific to each intervention. Then, a team of researchers had a series of discussions to build larger, emerging themes across interventions that revealed the most important patient needs from the study (see Multimedia Appendix 1 [18] for the screen capture of our affinity diagramming). For example, a larger theme of patient-provider communication contained various subthemes, such as patients being unable to notify providers and improving the communication of pain. Codes that were ambiguous or needed further support were refined by referencing participant quotes from transcripts and participants' sketches from PD sessions. After multiple group discussions, we identified 4 themes that described patients' key needs during their ED visits.

Ethical Considerations

We complied with the following ethical considerations. First, we recruited and obtained consent from all participants before the study sessions described in the multisite institutional review board approved by the University of Michigan (HUM00156261). Participants were informed that their participation was voluntary, the sessions would be recorded, the collected information would be deidentified and protected at a secured storage, and they would be compensated with US $100 for their participation. Second, we ensured that the collected transcript data were deidentified by replacing their names with pseudonyms, and

XSL•FO
RenderX

participants' faces in photos or videos were blurred. All collected data are stored in secured storage that requires the University of Michigan accounts. Third, all the researchers completed research compliance training such as the Collaborative Institutional Training Initiative or Program for the Education and Evaluation in Responsible Conduct of Research to become trained on best practices and ethical considerations of interacting with providers and patients.

Results

Overview

On the basis of the analysis of the collected data, we identified the 3 most preferred intervention ideas that participants preferred and 4 emerging themes that described the participants' needs in the ED care process.

Across all sessions, I-4 (ie, wearable device), I-7 (ie, virtual specialist), and I-1 (ie, virtual presence of a caregiver) emerged as the most preferred intervention ideas that the participants selected to help enhance the overall quality of patient care in the existing ED process. Among 36 participants, 25 (69%) chose I-4; 22 (61%) chose I-7; and 21 (58%) chose I-1 among their top 3 preferred interventions. These preferences represent which patient needs the participants considered most important. Through I-4, participants valued provider accountability and prioritization of patient conditions in the ED waiting room. For I-7, participants valued real-time virtual interactions with providers to better use their visit time while receiving the specialized care they needed. Furthermore, for I-1, participants valued the role of the caregiver for assistance and emotional support in the ED process even if they could not be physically present.

We also report 4 emerging themes of specific needs that participants had during their ED visits. These themes encompass broader perspectives of ED patients, extending beyond the technological intervention ideas provided. While describing these needs, we explain how participants addressed or wished to address them, along with their intervention ideas. Table 4 summarizes our key findings on patient needs and participants' suggestions for improving the intervention ideas.

Table 4. A summary of the 4 emerging themes based on patient needs during their ED[a] visit, discovered across all participatory design sessions, includes a description of each patient need and participant suggestions on how to address or improve the need.

Patient needs	Description	Patients suggestions for improvement
The feeling of inclusion in the ED care process	Patients feel that their voices are not heard by providers and want to be more involved in their care process	Create more communication channels with providers (eg, Help button)
Resources to improve patient comprehension of medical information	Patients feel inadequately literate about their health and want to have reliable ways to easily understand their condition, care plan, and treatment options	Use personalized information in intervention technology to enhance patients' understanding of their medical history
Relief of anxiety experienced by the patient over information overload and privacy concerns	Patients feel anxious because of privacy concerns and information overload in the ED. They strongly value emotional support from providers	Assist human-mediated emotional support, rather than replacing providers
Continuity in care and information	Patients perceive gaps in provider-to-provider communication that compromise care quality, emphasizing the need for improved technical infrastructure and procedures to ensure effective information delivery	Consider the accessibility of technology for different patient contexts

[a]ED: emergency department.

The Feeling of Inclusion in the ED Care Process

Throughout the 8 sessions, it is more than evident that participants wanted to be more included in the ED process. Participants often felt unheard, dismissed, and uninformed in their interactions with providers:

> *[The nurses] almost like, didn't believe me or didn't care, until they saw like, I have to pass out in front of them before they're going to do anything. And the second I passed out, I got moved to an actual room in the ED when I've been sitting in the hallway for hours.* [P13]

As described in the quote, participants sometimes felt that providers dismissed patient concerns unless they physically showed symptoms.

Their feeling of exclusion was caused mainly by the lack of timely and transparent communication with ED providers as they experienced providers not necessarily available to check on worsening patient conditions. Thus, it was challenging to inform providers of their health conditions:

> *Because there's a lot of times if you're on a two-hour wait and you're in the process of having a stroke, you're not able to get up and go back to the waiting room or the front desk, but it'd be good if something flagged that your condition was deteriorating.* [P24]

Because of the concern about potentially worsening conditions during the wait time, most participants favored I-4 (ie, wearable device) to track their conditions in real time and alert providers if necessary. In addition to the illustrated functionality, many participants suggested adding features such as pain-tracking and fall detection to the intervention, borrowing functionalities from other less popular intervention ideas, such as I-8 (ie, pain expression device) and I-5 (ie assistant robot). Some participants also thought that real-time tracking devices could be used to update patient priority in the posttriage stage:

> *If your condition is staying fine, or steady and stable and somebody else with a higher pain or priority*

XSL•FO
RenderX

> *could go first and you're still okay, you know something like that I could see being helpful.* [P3]

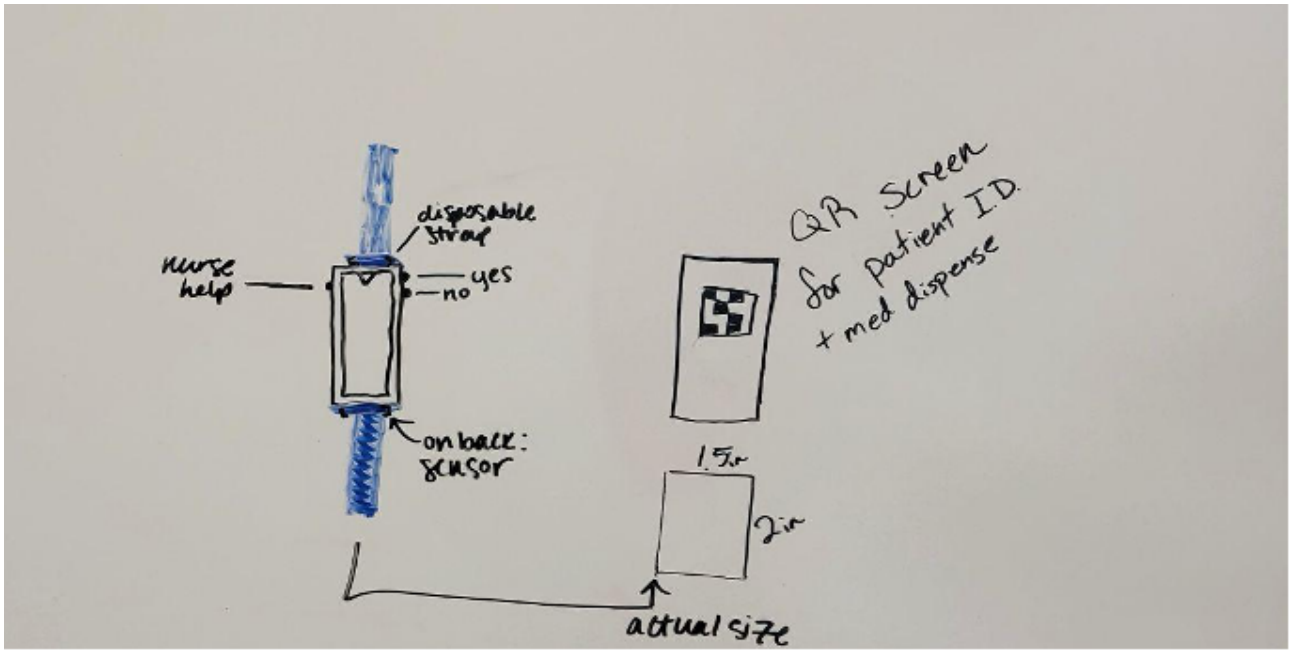
Participants felt that by using real-time tracking data, patients in critical condition could receive medical attention as soon as needed, ultimately improving the accuracy of the diagnosis.

Despite their general support for real-time tracking of symptoms, some participants expressed concerns about how providers could use such interventions. For instance, some participants were worried that providers would ignore tracking device emergency alerts in an already hectic environment. Others also expressed concern about providers becoming overreliant on tracking devices and potentially disregarding verbal patient feedback about worsening conditions if the device's data did not corroborate with patient concerns. Regardless of these worries, some participants pointed out that capturing real-time health data alerts could increase provider accountability in cases where patients felt ignored or misdiagnosed:

> *...if you're getting this information in real-time, then...people can be held to accountability. Because you go back and you look at their chart, you look at their record, "Hey, this person was in distress, and you neglected them," so I think in some ways, the AI could really assist in that because, again, there needs to be accountability to doctors and nurses and stuff that are just blowing you off.* [P18]

Participants also advocated for additional communication channels to create a more inclusive ED care process between patients and providers. A popular idea was incorporating a "help" button that participants could use to receive immediate medical attention from a provider, which they suggested could be implemented into one or many interventions (I-2: interactive display; I-3: table-top robot; I-4: wearable device; and I-5: assistant robot). Figure 4 demonstrates a sample drawing of an improved I-4 (ie, wearable device) with a "nurse help" button on the side. Moreover, many participants indicated that the I-4 (ie, wearable device) included a way to update their condition and pain levels as they changed during their visit, and some patients mentioned that concrete pain descriptors (eg, tactile or word based) could help them better describe their pain. Participants believed that having direct communication channels, such as a "help" button, with providers could lead to better care outcomes and higher trust in their care.

Figure 4. A sample drawing by a participant during a co-design activity. The drawing shows how I-4 (ie, a wearable device) could be improved with additional features such as the Nurse Help button.



As our participant pool included caregivers and patients, the role of the caregiver in patient-provider communication was also emphasized. Caregivers are especially crucial as patients advocate for children and older adults, as they can significantly help patients who cannot correctly remember or communicate their condition in the ED. Participants were receptive to the idea of incorporating I-1 (ie, virtual caregiver) into the ED process when an in-person caregiver was unavailable, and some emphasized that a caregiver list could be incorporated into patient records for cases when the patient was too incoherent to call a caregiver.

Resources to Improve Patient Comprehension of Medical Information

Another theme we found was patients needing more information resources to improve their comprehension of care-related information. Participants expressed that their understanding of medical information required more alignment with their providers. Despite having access to their patient portal, most patients found it hard to familiarize themselves with the terminologies providers used in the ED when communicating about their condition or test results:

> *When I see...I think it's [a] diagnosis or something or like conditions I have...in my chart or portal or*

XSL•FO
RenderX

> *whatever. Like it says Latin...something just saying that my knees are bad...So I wish in parentheses it could say in English...the layman's term for that.* [P25]

A similar concern around patient health literacy arose when clarifying personal medical information. Participants said they often used Google (Google LLC) and other web resources to help understand their condition or diagnosis better while waiting for a provider's consultation. One participant expressed that it could be more accessible to conduct a Google search because patients can have a baseline understanding to prepare follow-up questions for providers. However, participants also noted that when seeking medical information on the internet in this way, they could be misled into thinking their condition had severe side effects or symptoms that were not identified or considered relevant to the provider's diagnosis.

To address the concern of potentially misleading diagnoses, participants frequently mentioned the necessity of incorporating personalized information into intervention technology to enhance patients' understanding of their medical history. Depending on the patient's level of health literacy, they might struggle with formulating questions or initiating conversations when communicating with their provider. In many design sessions, participants suggested that hospital-specific information can be curated based on the patient's medical history to provide relevant and reliable search results sourced directly from the ED that might otherwise be retrieved from a Google search. Participants specifically valued the features of the I-6 (ie, ED chatbot) and I-4 (ie, wearable device) for providing personalized patient information and helping effectively facilitate conversations from the patient to provider and with the patient themselves—such technology can break down complex medical terminology so that patients can follow up with providers in a more efficient way and aid in making informed decisions about their condition:

> *I think that it [I-6: ED chatbot] can be helpful to the emergency room staff or doctors because they could input the information that they want you [patients] to know...so they can control the information that's being given. Versus you [patients] googling it.* [P3]

Along with the difficulty in understanding medical information regarding their care, participants expressed the challenge of explaining their pain to providers. Issues with patient-provider communication arose when patients felt unable to properly convey pain levels, struggling with arbitrary scales from 1 to 10. Several participants mentioned that current pain scales do not provide adequate ways to express different types of pain:

> *Whenever they say on a scale of one to 10, I don't feel pain like other people.* [P15]

Some participants who struggled with chronic pain also recalled that current pain scales do not take into account that people might consider different levels of pain to be "normal." In such cases, the lack of medical information comprehension hinders patients from accurately describing their medical conditions (eg, pain) to providers. While many EDs ask about pain on a scale of 1 to 10, it is difficult to pinpoint or trace a patient's condition over time based on this abstract measurement:

> *The description of the pain varies for me.... And, I do live with pain.... All the rest of the pain is still there.* [P16]

Some participants found the intervention idea of an I-8 (ie, pain expression device) valuable in addressing this issue, particularly for patients who cannot communicate verbally with a provider or have a language barrier. Participants particularly liked the device's tactile capabilities as this feature could help anchor patients to a basic description or visualization of their pain to communicate with providers effectively. At the same time, some participants were also concerned about the ability of a patient to understand how a complex device works when in pain, further delaying their time in the ED and getting a proper diagnosis from providers:

> *If they [ED patients] are in pain already and we [the participants and researchers] have to stop and give an explanation of how to work it [a new intervention], I think it might present some difficulties, and there might actually end up being more of a barrier in terms of getting the information.* [P20]

Relief of Patient Anxiety Over Information Overload and Privacy Concerns

Many of our participants mentioned patient anxiety during their ED visit as one of their top concerns. On top of being admitted to the ED, participants experienced anxiety due to various factors beyond diagnosis, including long wait times, fear of contagious diseases (especially during and after the COVID-19 pandemic), and uncertainty about the subsequent steps in the ED process. They also noted that being overwhelmed with information as a patient in the ED aggravated their anxiety. Whether from a Google search or looking at test results, patients might not be able to understand all the information presented to them. Participants thought it was essential to consider the types of information that would be accessible—such as vitals or pain levels—to patients, providers, or both:

> *Do [patients] know how to read it? Do they know how to interpret the information that's given to them? Too much information for a patient can be bad. Then their mind wanders because they don't know.* [P30]

Potential anxiety induced by overwhelming information brought different perspectives regarding how much information should be displayed in the intervention ideas. Participants wanted to be informed but felt too much information could be unnecessary for patients who did not understand medical terminology:

> *I think that's a hard one. Because some people—it doesn't stress them out to know, and they like to know, other people can like to say their heart rates [are] high, but it can make it worse because then they're anxious about it.* [P13]

Participants across sessions discussed different scopes of information access available for each intervention idea. For example, patients might be presented with a fundamental view that only shows whether their vitals were in a normal range. At the same time, providers might have more detailed access to numeric values and other data that could help track patient conditions over time. Participants suggested this could be paired

XSL•FO
RenderX

with alerts on the I-4 (ie, wearable device) for the patient and provider, or providers only, to notify hospital staff in an emergency and better support care priority in the waiting room. Some participants even considered extending this idea for the entire duration of a patient's ED visit. Overall, participants suggested informing providers of a patient's condition would help mitigate patient anxiety in an ED setting.

The other factor that contributed to patient anxiety was related to patient privacy. Some participants felt that their health information was exposed to others in the waiting room when providers periodically checked in with them. Participants did not feel their privacy was protected during these discussions and thus felt their rights were being violated. Participants also attributed this feeling to more functional ED issues such as the waiting room's size, layout, structure, and overall ED design. With a small waiting room and open, shared space for beds in the ED, participants felt uncomfortable about exchanging confidential health information between patients and providers in open environments and public areas:

> *I heard everything that was going on. I could have told you what was going on....Because of our health and our health being we have rights for nobody to know.* [P16]

P20 also had a similar experience:

> *I'm hearing people's whole names being called out in this huge waiting room. I felt like the confidentiality was being breached because I knew that person was there getting care.*

To address this issue, many participants advocated for existing ED processes to be remediated at the structural level. This included the addition of more ED beds with closed walls or partitions, private spaces in the waiting room, and priority-based interventions. Overcrowded EDs induce long wait times and quickly fill up space in the waiting room where patient information is highly vulnerable to others. Some participants also suggested interventions, such as I-2 (ie, interactive display) and I-5 (ie, assistant robot), should avoid displaying personal information to protect patient privacy, primarily because these interventions would be concentrated in the waiting room.

To alleviate anxiety induced in general by multiple factors during ED visits, many participants stressed the importance of having human providers as an aspect of emotional support for patients in the ED care process. When presented with ideas that limited patient-to-provider interactions (eg, I-3: table-top robot and I-10: AI caller), participants expressed how technological interventions should support human providers rather than replace them. For this reason, interventions involving AI or robot technology were not popular with participants because they could have limitations in understanding patient requests or evaluating a patient's condition because of the lack of sophisticated technology in this space. Participants further highlighted empathy as another critical factor to consider because technology in the form of a robot would not convey as much sympathy and support as a human would in a stressful ED situation. Many participants thus preferred to have human support for alleviating their anxiety along with technological interventions because humans can communicate medical information and reassure patients in a manner that helps in coping with an anxiety-inducing environment:

> *I don't think people would have a lot of confidence in [I-5: assistant robot] either. You know what I'm saying—it needs to be people who still like to interact.* [P18]

Continuity in Care and Information

During group discussions, many participants shared their experiences with discontinuity in care and information. On the basis of their experience, participants perceived that discontinuity might have occurred due to the handoff process within the care team. They noted that the involvement of multiple providers in their care team meant that information would often get lost during handoffs. Because of such missing information, some participants experienced discrepancies in provider-to-provider communication that caused patient confusion:

> *Even a stay in the hospital, the doctors just rotate and roll over.... The first doctor said they wouldn't consider [discharging] until it [the protein level] goes below 10,000. Then the next doctor said, "Oh no, I'm not going to even think about it unless it's below 500." So, each time the shift changed—they told us something different.* [P10]

Such discrepancies in care also occurred between hospitals. For instance, P27 shared her experience when she received a different diagnosis in the second hospital:

> *So if he [P27's son] would have asked for a second opinion, at that point, you're leaving me in a lot of pain, and I still don't know what's going on. If he could have asked for that second opinion at that point, maybe he would have been diagnosed with a real issue five days earlier than he was.* [P27]

Reflecting on those situations where information discontinuity had occurred, most participants favored the I-7 (ie, virtual specialist) to improve the breakdowns. They insisted that for interventions such as I-7 (ie, virtual specialist), an ED physician should be present when the patient talks to the specialist so that the ED care team could have updated information. They envisioned that bringing specialists directly to the ED care process would improve diagnosis and documentation in patient care records with the second opinion. Participants also expressed the need to allow for secure but easy information sharing with providers outside the hospital when external providers were invited to the ED care process. For instance, a global access health care platform could be used across hospitals for patient medical data transfer. Therefore, they thought that I-7 (ie, virtual specialist) could be expanded to include medical professionals beyond certain specialists:

> *I think it depends on the context, the question, but for certain questions, it could be any nurse, it could be a PA [physician assistant], or someone who had familiarity with the patient. Based on the level of question, it may need to be a provider. Or a pharmacist.* [P27]

XSL•FO
RenderX

To promote continuity of care, participants commented that the technological interventions should be designed in an accessible way for different patient situations. For example, some patients wanted to use their existing wearable devices like smartwatches or fitness trackers instead of wearables provided by the ED to integrate live vitals tracking with historical data accurately. They also suggested that the I-6 (ie, ED chatbot) could be offered as an app for mobile devices that patients and caregivers could download when they arrived at the ED and use throughout their ED stay and after their ED care. These participants expected that using interventions with patients' devices would ensure that patient-collected health data would be used to its fullest extent during care in the ED and would allow providers to glimpse the patient's complete health profile over time rather than only view current condition. On the other hand, some other participants emphasized the need for care continuity for those who may not have the same access to technology (eg, smartphones and smartwatches) or may have different amounts of previously collected health data, for instance, patients with chronic illnesses may have more personal health data tracked than those without. Thus, it is important that intervention use did not become a privilege reserved only for a subset of patients who had access to certain technological and data resources:

> *I think that one, look at the person that comes in without any communications, no phone or no computer, you know, nothing.... I often go into situations and emergencies without any kind of feature.... I carry a phone—an old-fashioned flip phone with me. I can't access some of those things.* [P20]

Participants' desire for continuity of care was also shown during their co-design activity when they worked to improve the intervention designs. They envisioned an integrative, connected system where multiple intervention ideas were integrated with the current health portal or text notification system to flow health-related information continuously. For example, they envisioned a tablet that could be used for I-7 (ie, virtual specialist), displaying vitals synchronized with I-4 (ie, wearable device) or the patient's health device (Figure 4). Through such integration of features, participants envisioned inventions to enhance the continuity of care and information and to address their complex challenges in the ED care process. Although a few participants felt concerned that such an integrated intervention could delay ED stay time or contribute to patient information overload, many still thought our suggested interventions could be integrated with the existing patient portal and offered in an app format on various devices.

Discussion

Design Implications for Patient-Centered Interventions in ED

On the basis of the PD sessions, we identified ED patients' needs in the ED care process and expectations for future technological interventions. This section discusses how these findings may lead to design guidelines for more patient-centered interventions in EDs.

First, technology-based *interventions should address patients' dynamic needs to promote continuity in care*. Our participants found that their needs could be met by more than one intervention idea or multiple specific features because their complex needs may change depending on where they stand in the ED care process. For instance, many participants suggested integrating critical functions into one system in the co-design activity so that the system could provide necessary and relevant information to them, from arrival at the ED to postvisit follow-up. This desire was evident when some participants envisioned an integrated system of 2 different intervention ideas (I-7: virtual specialist and I-4: wearable device) that could track patients' status and share data with specialists during the diagnosis process. Such integrated interventions may also be used to support issues in patient-provider communication. Our findings demonstrated that patients lacked sufficient information and interactions with ED providers. Thus, our participants shared how integrated interventions may help fill this information gap by providing relevant communication channels or resources based on patients' dynamic needs during ED visits. However, mere integration of systems may lead to additional challenges for patients. As some of our participants highlighted, numerous resources may cause information overload that can lead to patient anxiety. In another case, inviting outside providers (ie, specialists) would require extra documentation, potentially resulting in additional patient burdens. To incorporate a patient-centered integrated system in the ED, we suggest that the system be capable of adapting to patients' needs throughout the ED care timeline rather than simply adding features into one system. Existing technologies are primarily designed to provide specific information to patients in a certain part of the ED process (eg, a mobile app for an interactive report [7] to inform patients about their real-time progress and care plans during their visits). Such technologies can be integrated with other features, such as I-10 (ie, AI caller) for after-visit support, to address the dynamically evolving needs of patients that extend beyond their stay in the ED. An integrated system that supports patients starting from triage to beyond their discharge would provide timely, relevant information to patients based on their status within the ED care process. This integration approach to promoting continuity in care resonates with prior studies that designed a multicomponent app for patients and caregivers, the Bone Marrow Transplant Roadmap [19]. The app included components that reflect the information needs of patients and caregivers during their clinic visits, such as an overview of the criteria needed for discharge, real-time laboratory results, personalized medication lists, and educational materials. This app was further expanded to include outpatient settings to better support caregivers [20]. Similar to these multicomponent approaches, we suggest the integration of various features to better meet ED patients' dynamic needs during or even after their ED visits.

Second, *interventions should account for both the amount and timing of information provided to patients.* Our participants shared their challenges in accessing the appropriate amount of information during their ED visits. They felt excluded from the care process when they received insufficient information, while overwhelming information heightened their anxiety. Thus, participants suggested that the patients should have the decision

XSL•FO
RenderX

to choose the timing and extent of information they receive for applicable interventions, such as opting for partial, complete, or no information access. One example of this suggestion was using a chatbot to provide patients with personalized care information. On the basis of the participants' ideas, I-6 (ie, ED chatbot) could be further developed to offer relevant information depending on the steps patients undergo in the care process and their current health condition. Prior studies have shown that chatbots can provide more accessible and relevant information to patients [21]. In cancer contexts, a chatbot provided answers to patients with breast cancer with similar satisfaction as physicians [22], while another chatbot collected patient-reported symptoms during chemotherapy and reduced ED visits and unscheduled hospitalizations [23]. While these studies presented the opportunities for chatbots to provide information to patients, our findings suggest more careful considerations on the timing of providing information to patients. The need for timely and appropriate information resonates with the findings of a prior study [24], which describes how temporal dependency can delay the information that ED patients need. As ED medical activities often depend on previous activities, the information patients need or want is usually delayed. Without having relevant information during the ED process, patients may experience communication challenges with ED providers, such as different perceptions of what information is critical to ED care [25]. Therefore, as our participants suggested, interventions should provide ED patients with optimized information relevant to the current step of the care process or options to control the amount and timing of information they expect to receive.

Third, *interventions should empower patients to be more active in their care process for better patient safety and care quality*. Our findings highlighted participants' need for patient autonomy in ED care. Unlike other patient care settings, the ED context is particularly intimidating for patients because ED visits are unexpected and urgent. Each visit may be different from the previous one. However, we found that participants in our study were eager to take a more active role in the ED care process. In the PD sessions, the need for empowerment was evident when most participants suggested features that allowed them to communicate assertively with providers about their symptoms. For instance, although I-4 (ie, wearable device) was initially designed to track patients' condition so they would not feel forgotten while waiting, many participants wanted to have another function, similar to a "help" button, to request support more actively from providers when their conditions worsened or changed. This suggestion indicates how ED patients want to be more active stakeholders in the ED care process rather than be considered passive stakeholders who are just being assessed for their symptoms. Patient-empowering interventions are implemented on the basis of the prior studies' insights from existing interventions, such as an inpatient portal [26] and a smartphone app [27]. For instance, a smartphone app called MySurgery [27] was developed to provide information about surgery risks and practical step-by-step advice for each risk. The app demonstrated a significant ability to enable patients to engage actively in discussions regarding their care and adopt behaviors related to their safety. Similarly, ED patients can be empowered with relevant information and advice, such as explaining medical terms or notices about the next step in the ED care process. This need for patient empowerment resonates with previous studies on shared decision-making in the ED [28,29]. These studies identified that ED patients want some degree of involvement in medical decision-making. Extending this line of study, we found contextual nuances and more concrete examples of ED interventions to promote patient involvement. For instance, a review study showed the opportunity for wearable devices to monitor patient vital signs and provide the collected data to providers to improve decision-making [30]. Our findings suggest how ED patients perceive such wearable devices as an opportunity to be more actively involved in the care process rather than letting the devices passively collect data. Therefore, we suggest that future ED interventions be designed to empower patients to be included in shared decision-making.

Fourth, *interventions should optimize human resources, depending on patients' needs*. We identified that some issues cannot be addressed through technological interventions alone. These issues may be related to distress or emotional needs of ED patients. For instance, our participants preferred human support rather than interactions with technical tools in alleviating their feeling of anxiety during ED visits. This preference implies that ED patients can manage their stress through human interactions, although the suggested interventions, such as I-5 (ie, assistant robot), can still provide medical information. Moreover, human resources for ED patients can include those who are not ED providers. Many participants expressed their need to connect with informal caregivers (eg, family members), specialists, or other providers during their ED visit. Their needs were particularly evident when patients visited the ED once because they did not know about the ED care process. On the basis of their previous visits, patients learned that some external resources (eg, family members and specialists) could provide more information to help assess their conditions or, in the case of family members, provide emotional support during the ED visit. This finding indicates that ED patients leverage resources they are more familiar with in the intimidating and unfamiliar ED settings. Prior studies in the medical literature have shown the opportunities and challenges of telehealth for patients [31-33]. However, those identified challenges (eg, unsteady or poorly framed video [33]) and strategies to overcome them may not be implemented in the fast-paced ED care process. Depending on the patient's current conditions, the right timing (eg, when a specialist can be contacted) and method (eg, simple video call or additional camera to show the patient's status) for providing external resources to patients should be determined. Thus, designing patient-centered interventions should optimize human resources to address ED-specific challenges, as we identified in our study. It will be critical to identify when and how interventions should provide external resources to patients. For instance, some of our participants wanted to be connected with specialists when they met ED providers after examinations so that they could have second opinions on their symptoms.

Finally, *interventions should be designed considering patients' perspectives on implementation*. Through co-design activities, we identified how our participants expected intervention ideas to be implemented to address their needs better. However, their expectations could differ from actual implementations,

XSL•FO
RenderX

potentially conflicting with what providers or the hospital expect. In such cases, it is critical to mitigate those expectations to implement patient-centered interventions. For instance, one of the potential reasons for our participants' preference for I-7 (ie, virtual specialist) was the assumption that specialists are available to have on-demand consultations. In contrast, in actual implementations, it would require other resources (eg, cost, time, and human resources) to implement a system to become connected with specialists. As a prior study shows, ED providers must modify their work practices to adopt the new interventions [34]. Moreover, some of our participants pointed out the accessibility of devices that ED patients may not have. While mobile device ownership is high among ED patients [35], our suggested intervention ideas involved devices that patients may not have, such as wearable trackers. Thus, it is essential to consider how ED patients perceive the implementation of technological interventions in the ED.

Limitations

This study has some limitations. Given their qualitative nature, the findings are specific to the context we examined. Providers and patients from other EDs may face different challenges. However, our findings will contribute to identifying potential interventions that address such challenges through a PD approach. In addition, we recruited participants with varying backgrounds from 2 ED units (pediatric and adult EDs), which could help with the study's generalizability. Another limitation is that we did not distinguish participants' previous experiences as ED patients or caregivers. While the ED experiences of patients and caregivers do not have noticeable differences, future work can explore caregiver-specific perspectives on interventions designed to address their challenges in the ED care process. Finally, we only presented concepts, not interactive system prototypes. Therefore, participants may not be quite familiar with some intervention ideas. They could only imagine how an intervention might work but could not see the real system or product. However, storyboards are a common Human-Computer Interaction methodology for eliciting user needs in the early system design stage. They are useful for validating user needs and expectations before implementing a system in a complex environment similar to an ED.

Conclusions

In this PD study, we invited ED patients to design sessions to assess their needs in the ED care process and co-design technical interventions. On the basis of the analysis of collected data, we identified 4 themes related to ED patients' needs: the feeling of exclusion from the ED care process, limited resources for patient comprehension of medical information, anxiety about overwhelming information and privacy, and discontinuity in care and information. Concerns were also expressed about the accuracy of the diagnosis. These findings also informed us to develop guidance for designing future technology-based patient-centered interventions to improve the diagnostic process in the ED. To refine the identified design guidelines, we aim to conduct a PD study with ED health care providers. By identifying similarities and differences between ED patients and providers, we expect to present concrete design guidelines for technological interventions to support the ED diagnostic process and patient safety.

Acknowledgments

This project was funded under grant number R18HS026622 from the Agency for Healthcare Research and Quality (AHRQ), U.S. Department of Health and Human Services (HHS). HS is supported in part by the Houston Veterans Administration (VA) Health Systems Research and Development (HSR&D) Center for Innovations in Quality, Effectiveness and Safety (CIN13–413), the VA National Center for Patient Safety, and the Agency for Healthcare Research and Quality (R01HS028595, R18HS029347, and R01HS029318). The authors are solely responsible for this document's contents, findings, and conclusions, which do not necessarily represent the views of AHRQ. Readers should not interpret any statement in this report as the official position or policy of the US government, AHRQ, HHS, or VA. None of the authors have any affiliation or financial involvement that conflicts with the material presented in this report.

Data Availability

The datasets generated and analyzed during this study are not publicly available because the consent form for the participants stated that the data could only be accessible to the research team.

Conflicts of Interest

None declared.

Multimedia Appendix 1
A screen capture of our affinity diagramming. We used the Miro board [18] for affinity diagramming.
[PDF File (Adobe PDF File), 161 KB - formative_v9i1e63610_app1.pdf]

References

1. Newman-Toker DE, Peterson SM, Badihian S, Hassoon A, Nassery N, Parizadeh D, et al. Diagnostic Errors in the Emergency Department: A Systematic Review. Rockville, MD: Agency for Healthcare Research and Quality (US); Dec 2022.

XSL•FO
RenderX

2. Ehrler F, Tuor C, Rey R, Siebert JN. A mobile app to improve patient management in emergency departments: caregiver needs analysis, design and early technology acceptance assessment. Stud Health Technol Inform 2021 Oct 27;285:233-238. [doi: 10.3233/SHTI210605] [Medline: 34734879]
3. Scheder-Bieschin J, Blümke B, de Buijzer E, Cotte F, Echterdiek F, Nacsa J, et al. Improving emergency department patient-physician conversation through an artificial intelligence symptom-taking tool: mixed methods pilot observational study. JMIR Form Res 2022 Feb 07;6(2):e28199 [FREE Full text] [doi: 10.2196/28199] [Medline: 35129452]
4. Murray L, Gopinath D, Agrawal M, Horng S, Sontag D, Karger DR. MedKnowts: unified documentation and information retrieval for electronic health records. In: Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology. 2021 Presented at: UIST '21; October 10-14, 2021; Virtual Event. [doi: 10.1145/3472749.3474814]
5. Greenbaum NR, Jernite Y, Halpern Y, Calder S, Nathanson LA, Sontag DA, et al. Improving documentation of presenting problems in the emergency department using a domain-specific ontology and machine learning-driven user interfaces. Int J Med Inform 2019 Dec;132:103981. [doi: 10.1016/j.ijmedinf.2019.103981] [Medline: 31605881]
6. Ho TW, Tsai CJ, Hsu CC, Chang YT, Lai F. Indoor navigation and physician-patient communication in emergency department. In: Proceedings of the 3rd International Conference on Communication and Information Processing. 2017 Presented at: ICCIP '17; November 24-26, 2017; Tokyo, Japan. [doi: 10.1145/3162957.3162971]
7. Vardoulakis LP, Karlson A, Morris D, Smith G, Gatewood J, Tan D. Using mobile phones to present medical information to hospital patients. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI '12; May 5-10, 2012; Austin, TX. [doi: 10.1145/2207676.2208601]
8. Chai PR, Schwartz E, Hasdianda MA, Azizoddin DR, Kikut A, Jambaulikar GD, et al. A brief music app to address pain in the emergency department: prospective study. J Med Internet Res 2020 May 20;22(5):e18537 [FREE Full text] [doi: 10.2196/18537] [Medline: 32432550]
9. Sikka N, Shu L, Ritchie B, Amdur RL, Pourmand A. Virtual reality-assisted pain, anxiety, and anger management in the emergency department. Telemed J E Health 2019 Dec 01;25(12):1207-1215. [doi: 10.1089/tmj.2018.0273] [Medline: 30785860]
10. Delshad S, Dontaraju VS, Chengat V. Artificial intelligence-based application provides accurate medical triage advice when compared to consensus decisions of healthcare providers. Cureus 2021 Aug;13(8):e16956 [FREE Full text] [doi: 10.7759/cureus.16956] [Medline: 34405077]
11. Muller MJ, Kuhn S. Participatory design. Commun ACM 1993 Jun;36(6):24-28. [doi: 10.1145/153571.255960]
12. Kusunoki D, Sarcevic A, Zhang Z, Yala M. Sketching awareness: a participatory study to elicit designs for supporting ad hoc emergency medical teamwork. Comput Support Coop Work 2015 Feb 01;24(1):1-38 [FREE Full text] [doi: 10.1007/s10606-014-9210-5] [Medline: 25870498]
13. Pollack AH, Miller A, Mishra SR, Pratt W. PD-atricians: leveraging physicians and participatory design to develop novel clinical information tools. AMIA Annu Symp Proc 2016;2016:1030-1039 [FREE Full text] [Medline: 28269900]
14. Østervang C, Lassen A, Schmidt T, Coyne E, Dieperink KB, Jensen CM. Development of a health information system to promote emergency care pathways: a participatory design study. Digit Health 2022 Dec 26;8:20552076221145856 [FREE Full text] [doi: 10.1177/20552076221145856] [Medline: 36601282]
15. Daniel M, Park S, Seifert CM, Chandanabhumma PP, Fetters MD, Wilson E, et al. Understanding diagnostic processes in emergency departments: a mixed methods case study protocol. BMJ Open 2021 Sep 24;11(9):e044194 [FREE Full text] [doi: 10.1136/bmjopen-2020-044194] [Medline: 34561251]
16. Mahajan P, Mollen C, Alpern ER, Baird-Cox K, Boothman RC, Chamberlain JM, et al. An operational framework to study diagnostic errors in emergency departments: findings from a consensus panel. J Patient Saf 2021 Dec 01;17(8):570-575. [doi: 10.1097/PTS.0000000000000624] [Medline: 31790012]
17. Scupin R. The KJ method: a technique for analyzing data derived from Japanese ethnology. Hum Org 2008 Jan 22;56(2):233-237. [doi: 10.17730/humo.56.2.x335923511444655]
18. Miro: The Innovation Workspace. Miro. URL: https://miro.com/ [accessed 2025-02-04]
19. Runaas L, Hanauer D, Maher M, Bischoff E, Fauer A, Hoang T, et al. BMT roadmap: a user-centered design health information technology tool to promote patient-centered care in pediatric hematopoietic cell transplantation. Biol Blood Marrow Transplant 2017 May;23(5):813-819 [FREE Full text] [doi: 10.1016/j.bbmt.2017.01.080] [Medline: 28132870]
20. Chaar D, Shin JY, Mazzoli A, Vue R, Kedroske J, Chappell G, et al. A mobile health app (Roadmap 2.0) for patients undergoing hematopoietic stem cell transplant: qualitative study on family caregivers' perspectives and design considerations. JMIR Mhealth Uhealth 2019 Oct 24;7(10):e15775 [FREE Full text] [doi: 10.2196/15775] [Medline: 31651402]
21. Clark M, Bailey S. Chatbots in health care: connecting patients to information. Can J Health Technol 2024 Jan 22;4(1). [doi: 10.51731/cjht.2024.818]
22. Bibault JE, Chaix B, Guillemassé A, Cousin S, Escande A, Perrin M, et al. A chatbot versus physicians to provide information for patients with breast cancer: blind, randomized controlled noninferiority trial. J Med Internet Res 2019 Nov 27;21(11):e15787 [FREE Full text] [doi: 10.2196/15787] [Medline: 31774408]
23. Huang MY, Weng CS, Kuo HL, Su YC. Using a chatbot to reduce emergency department visits and unscheduled hospitalizations among patients with gynecologic malignancies during chemotherapy: a retrospective cohort study. Heliyon 2023 May;9(5):e15798 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e15798] [Medline: 37206031]

XSL•FO
RenderX

24. Park SY, Chen Y. Patient strategies as active adaptation: understanding patient behaviors during an emergency visit. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17; May 6-11, 2017; Denver, CO. [doi: 10.1145/3025453.3025978]
25. Park SY, Chen Y, Raj S. Beyond health literacy: supporting patient-provider communication during an emergency visit. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: CSCW '17; February 25-March 1, 2017; Portland, OR. [doi: 10.1145/2998181.2998357]
26. McAlearney AS, Fareed N, Gaughan A, MacEwan SR, Volney J, Sieck CJ. Empowering patients during hospitalization: perspectives on inpatient portal use. Appl Clin Inform 2019 Jan 13;10(1):103-112 [FREE Full text] [doi: 10.1055/s-0039-1677722] [Medline: 30759491]
27. Russ S, Latif Z, Hazell AL, Ogunmuyiwa H, Tapper J, Wachuku-King S, et al. A smartphone app designed to empower patients to contribute toward safer surgical care: community-based evaluation using a participatory approach. JMIR Mhealth Uhealth 2020 Jan 20;8(1):e12859 [FREE Full text] [doi: 10.2196/12859] [Medline: 31958067]
28. Schoenfeld EM, Kanzaria HK, Quigley DD, Marie PS, Nayyar N, Sabbagh SH, et al. Patient preferences regarding shared decision making in the emergency department: findings from a multisite survey. Acad Emerg Med 2018 Oct 19;25(10):1118-1128 [FREE Full text] [doi: 10.1111/acem.13499] [Medline: 29897639]
29. Schoenfeld EM, Goff SL, Downs G, Wenger RJ, Lindenauer PK, Mazor KM. A qualitative analysis of patients' perceptions of shared decision making in the emergency department: "let me know I have a choice". Acad Emerg Med 2018 Jul 07;25(7):716-727 [FREE Full text] [doi: 10.1111/acem.13416] [Medline: 29577490]
30. Nino V, Claudio D, Schiel C, Bellows B. Coupling wearable devices and decision theory in the United States emergency department triage process: a narrative review. Int J Environ Res Public Health 2020 Dec 21;17(24):9561 [FREE Full text] [doi: 10.3390/ijerph17249561] [Medline: 33371223]
31. Donelan K, Barreto EA, Sossong S, Michael C, Estrada JJ, Cohen AB, et al. Patient and clinician experiences with telehealth for patient follow-up care. Am J Manag Care 2019 Jan;25(1):40-44 [FREE Full text] [Medline: 30667610]
32. Powell RE, Henstenburg JM, Cooper G, Hollander JE, Rising KL. Patient perceptions of telehealth primary care video visits. Ann Fam Med 2017 May 08;15(3):225-229 [FREE Full text] [doi: 10.1370/afm.2095] [Medline: 28483887]
33. Wootton AR, McCuistian C, Legnitto Packard DA, Gruber VA, Saberi P. Overcoming technological challenges: lessons learned from a telehealth counseling study. Telemed J E Health 2020 Oct 01;26(10):1278-1283. [doi: 10.1089/tmj.2019.0191] [Medline: 31800368]
34. Park SY, Chen Y. Adaptation as design: learning from an EMR deployment study. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI '12; May 5-10, 2012; Austin, TX. [doi: 10.1145/2207676.2208361]
35. Kim E, Torous J, Horng S, Grossestreuer AV, Rodriguez J, Lee T, et al. Mobile device ownership among emergency department patients. Int J Med Inform 2019 Jun;126:114-117. [doi: 10.1016/j.ijmedinf.2019.03.020] [Medline: 31029252]

Abbreviations

AI: artificial intelligence
ED: emergency department
PD: participatory design

Edited by A Mavragani; submitted 25.06.24; peer-reviewed by K Fuji, A Londral; comments to author 30.08.24; revised version received 31.10.24; accepted 21.12.24; published 12.02.25.

Please cite as:
Seo W, Jain S, Le V, Li J, Zhang Z, Singh H, Pasupathy K, Mahajan P, Park SY
Designing Patient-Centered Interventions for Emergency Care: Participatory Design Study
JMIR Form Res 2025;9:e63610
URL: https://formative.jmir.org/2025/1/e63610
doi:10.2196/63610
PMID:

©Woosuk Seo, Shruti Jain, Vivian Le, Jiaqi Li, Zhan Zhang, Hardeep Singh, Kalyan Pasupathy, Prashant Mahajan, Sun Young Park. Originally published in JMIR Formative Research (https://formative.jmir.org), 12.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Effect of the Virtual Reality–Based Biofeedback Intervention DEEP on Stress, Emotional Tension, and Anger in Forensic Psychiatric Inpatients: Mixed Methods Single-Case Experimental Design

Lisa Klein Haneveld[1,2], MSc; Tessa Dekkers[1], PhD; Yvonne H A Bouman[2], PhD; Hanneke Scholten[3], PhD; Joanneke Weerdmeester[4], PhD; Saskia M Kelders[1], PhD; Hanneke Kip[1,2], PhD

[1]Centre for eHealth and Wellbeing Research, Department of Psychology, Health and Technology, University of Twente, Enschede, The Netherlands

[2]Department of Research, Transfore, Deventer, The Netherlands

[3]Department of Child and Adolescent Psychiatry and Psychology, Erasmus Medical Centre, Rotterdam, The Netherlands

[4]Research Group Creative Making Processes & Technology, HKU University of the Arts, Utrecht, The Netherlands

Corresponding Author:
Lisa Klein Haneveld, MSc
Centre for eHealth and Wellbeing Research
Department of Psychology, Health and Technology
University of Twente
Drienerlolaan 5
Enschede, 7522 NB
The Netherlands
Phone: 31 619210156
Email: l.kleinhaneveld@transfore.nl

Abstract

Background: Decreasing aggression through stress reduction is an important part of forensic psychiatric treatment. DEEP is an experience-based virtual reality intervention that uses biofeedback to train diaphragmatic breathing and increase relaxation. Although DEEP has shown promising results in reducing stress and anxiety in students and adolescents in special education, it has not been examined in forensic psychiatric populations.

Objective: This study aimed to evaluate DEEP's potential to reduce stress, emotional tension, and anger in forensic psychiatric inpatients.

Methods: A mixed methods, alternating treatment, single-case experimental design was conducted with 6 Dutch forensic inpatients. For 20 days, participants engaged in 4 DEEP sessions. Experience sampling was used for continuous monitoring of stress, emotional tension, and anger twice daily. A repeated linear mixed model was used as a primary statistical approach for analyzing the experience sampling data as well as visual analyses. Finally, semistructured interviews were conducted with participants and health care professionals to compare quantitative with qualitative results.

Results: Of the 6 participants, 3 (50%) completed all 4 DEEP sessions, while the other 3 (50%) missed one session due to technical difficulties or absence from the inpatient clinic. P1 showed a significant reduction of stress after session 2 (β=–.865; *P*=.005). No significant changes over time were found, although an experienced effect was reported during the interviews. P2 showed no significant results. They reported the sessions as being repetitive, with no experienced effect. P3 showed a momentary increase of emotional tension after the first session (β=–.053; *P*=.002), but no changes were observed over time. No experienced effects were reported in the interview. P4 did not show significant results over time, and was hesitant to report clear experienced effects. P5 showed a significant decline of emotional tension (β=–.012; *P*=.006), stress (β=–.014; *P*=.007), and anger (β=–.007; *P*=.02) over time. They also reported short-term experienced effects in the interview. P6 showed a significant decline of stress over time (β=–.029; *P*<.001) and reported experiencing substantial effects. Finally, health care professionals reported a relaxing effect of DEEP in their patients but did not expect many long-term effects because no clear behavioral changes were observed.

Conclusions: DEEP shows promise in teaching deep breathing techniques to forensic psychiatric inpatients, potentially decreasing stress, emotional tension, and anger in some patients. However, DEEP is not a one-size-fits-all intervention that supports every patient because the effectiveness on the outcome measures varied among participants. To increase effectiveness, emphasis should

XSL•FO
RenderX

be put on supporting patients to transfer deep breathing skills into their daily lives. This highlights the importance for the structural integration of DEEP into current treatment protocols.

(JMIR Form Res 2025;9:e65206) doi:10.2196/65206

KEYWORDS

virtual reality; VR; diaphragmatic breathing; biofeedback; DEEP; forensic psychiatry; mental health; stress; single-case experimental design; experience sampling method

Introduction

Background

Forensic mental health care is focused on the treatment of individuals who exhibit imminent aggressive or sexually inappropriate behavior due to 1 or more psychiatric illnesses [1,2]. Reducing aggression in treatment is complex because patients often show difficulties with reflection, reading, and writing due to low literacy and treatment motivation [3,4]. However, treatment is mostly focused on talking and thinking about behavior, guided by treatment frameworks such as cognitive behavioral therapy, with less attention for more experience-based or experimental approaches [5]. This shows a need for interventions that can support patients in preventing aggressive behavior in an experiential and engaging way.

Earlier research has shown the promise of virtual reality (VR) and serious games to better involve patients with lower cognitive skills and treatment motivation [5,6]. Previous research has also shown that the VR game DEEP fits the skills and needs of the forensic population [7]. In DEEP, the player moves through an immersive world where they have to use deep diaphragmatic breathing to navigate through and explore the world [8]. The user receives feedback on how they are breathing, which helps them to adjust their breathing and learn to apply diaphragmatic breathing. This technique of providing direct feedback on bodily signals is called biofeedback [9]. Diaphragmatic breathing has been shown to support relaxation, well-being, and anger regulation [10,11].

Previous research has shown the potential of DEEP to reduce negative emotions. First, a randomized controlled trial with DEEP in undergraduate students showed a decrease in anxiety symptoms after using DEEP 4 times [12]. However, these results cannot be generalized, as the study population consists of a relatively healthy, highly educated, and homogeneous target group, which is not the case for the forensic psychiatric population. Second, a single-case experimental design (SCED) study with adolescents in special education demonstrated DEEP's ability to decrease state anxiety and stress [11]. By using an SCED, this study was able to provide insights into both short-term and potential learning effects of DEEP. Moreover, their level of disruptive behavior in class seemed to decrease [11]. Although this population and the forensic target group have overlapping characteristics, the results are not entirely generalizable to a forensic setting. Namely, this study did not include state anger or aggression, which are important treatment outcomes in forensic mental health. There seems to be a need for more specific studies that investigate whether DEEP is fitting the characteristics of the forensic psychiatric setting, focusing on those specific outcomes measures. On the basis of these earlier studies and qualitative research in which DEEP was discussed with forensic patients and health care professionals, DEEP could be a fitting intervention for the hard-to-involve forensic psychiatric patient group. First, because of its appealing and immersive design in VR, DEEP might be able to engage populations with lower motivation, which is not always possible with regular breathing or relaxation interventions [7]. Second, DEEP uses diaphragmatic breathing, where an individual learns to breathe deeply using the diaphragm muscle, as opposed to shallowly through the rib cage and chest [13-15]. In addition to direct relaxation, the use of deep breathing during stressful situations can enhance emotion regulation and reduce stress [16]. By reducing stress, applying deep breathing may also reduce anger and aggression, because these constructs are important factors that may underlie outbursts of aggression [17-19]. By letting patients experience and adjust their own way of breathing in an engaging way, DEEP might be able to support them in increasing relaxation and well-being and decrease emotional tension, stress, and anger.

Objectives

This study aimed to investigate whether DEEP can support forensic inpatients in reducing stress, emotional tension, and anger.

In this mixed methods SCED study, the following research questions were formulated:

1. To what extent is there a significant improvement in self-reported stress, emotional tension, and anger in forensic psychiatric inpatients on days when DEEP is being used compared to days when it is not used?
2. To what extent is there a significant improvement in self-reported stress, emotional tension, and anger in forensic inpatients over time during the period when DEEP was used?
3. To what extent do patients and health care professionals observe or experience effects on emotional tension, stress, and anger that they attribute to DEEP?

Methods

Study Design

To answer the research questions, a mixed methods SCED was applied. Data on stress, emotional tension, and anger were collected quantitatively via experience sampling (ES) as well as qualitatively via interviews. In SCEDs, participants serve as both the intervention group and their own control group. Only a small number of participants are needed to provide sufficient statistical power (3-6), with a minimum of 3 data points per phase, including a baseline and follow-up [20]. Therefore, an

XSL•FO
RenderX

SCED is a fitting study design to use within the forensic psychiatric population, as they form a heterogeneous and complex-to-involve target group with high comorbidity. To study the effectiveness of DEEP, an alternating treatment SCED was chosen [20]. In this type of design, an intervention is not applied on certain days (A phase) and is implemented on other days (B phase), which is then repeated often in a brief period [21,22].

Setting and Participants

The study was conducted at Dutch inpatient clinics with various security levels. Five health care professionals (ie professionals providing social and behavioral support in mental health settings) were included. Of these, 4 (80%) identified as female and 1 (20%) as male, with ages ranging between 27 and 39 (mean 31.2, SD 7.1) years. Patients were included via convenience sampling, which meant that they were included via their health care professional, who identified them as suitable candidates and asked them to participate. A patient was included if (1) they were aged ≥18 years, (2) they were admitted to a forensic psychiatric hospital, (3) they were expected to reside in the same clinic for at least 5 weeks, (4) they were treated for aggression regulation problems, and (5) their health care professional deemed them physically and mentally capable to participate. A patient was excluded if (1) they experienced psychotic episodes or any other form of psychiatric crisis that could negatively impact participation, or (2) they experienced epileptic episodes. In total, 6 forensic inpatients were included. All (6/6, 100%) identified as male, with ages ranging between 23 and 45 (mean 33.4, SD 9.3) years. The data collection commenced 1 week after inclusion, allowing patients sufficient time to reconsider and withdraw from the study if they wished. The inclusion of participants was done in line with the guidelines from the Declaration of Helsinki and from the Medical Research Involving Human Subjects Act (wet medisch-wetenschappelijk onderzoek met mensen in Dutch).

Materials

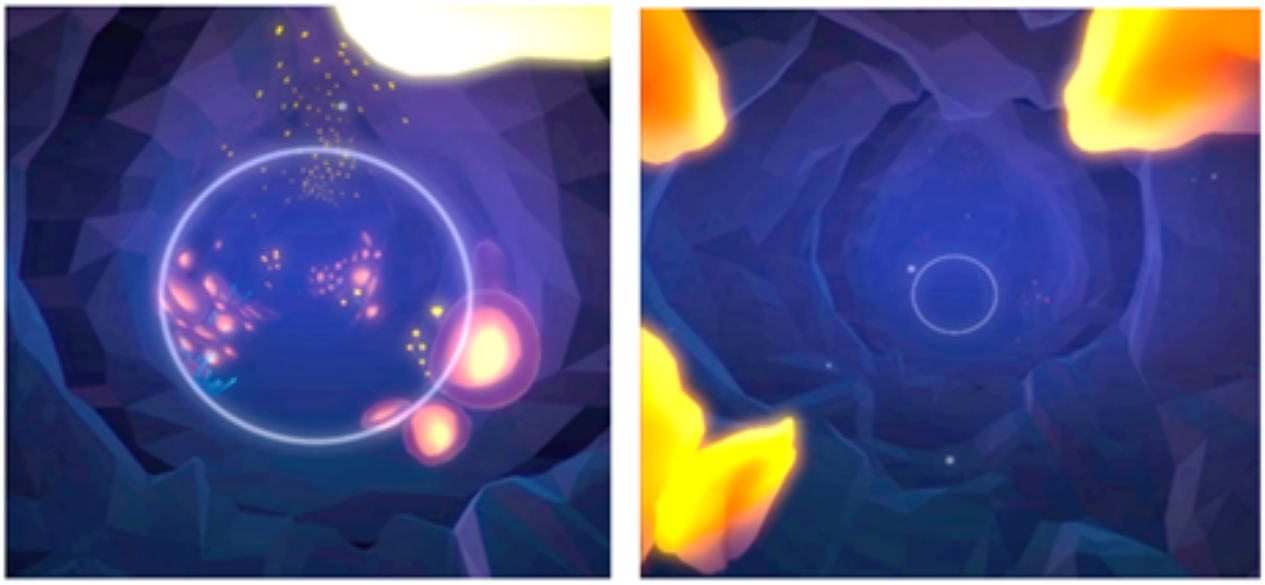
Avicenna App

During the SCED, the participants were monitored through ES, providing insight into outcome measures, which were measured multiple times per day [23]. ES increases the ecological validity, possibilities, and reliability of frequent data collection [24]. The participants received questions regarding their experienced stress, emotional tension, and anger by use of the Ethica app (formerly known as the Avicenna app), which assists researchers in collecting real-time data [25]. During the study period, participants received 3 notifications per day (at 10:00 AM, 2:00 PM, and 9:00 PM) through the installed app on their mobile phones with a set of 3 statements about the study outcomes: “At this moment I feel [emotionally tense, stressed or angry].” The participants scored these statements on a 5-point Likert scale, where 1 was “not the case at all” and 5 was “very much the case.”

The DEEP Intervention

DEEP is a VR game based on scientific knowledge about anxiety and stress regulation [8]. The player finds themselves in an immersive virtual underwater world where they navigate using their diaphragmatic breathing, which is measured by an abdominal belt [8,26]. The player’s breathing is visualized in the VR environment through a breathing circle, as well as by the illumination and growth of plants and corals, which mirror the player’s breath (Figure 1). By using deep breathing, the player is able to increase their progress in the game, while shallow breathing, which often occurs while experiencing negative emotions, will lead to in-game stagnation. During this study, a beta-version of DEEP was used on the Oculus Rift headset. Oculus (currently known as Meta Quest) is a well-known provider of VR technology and is Conformité Européene certified [27]. In this study, the participants would sit on a spinning desk chair with the VR headset on their head and the abdominal belt around their diaphragm. Both the headset and the belt were attached to a laptop by wires.

Figure 1. A visual representation of a deep breathing exercise in DEEP, as measured with an abdominal belt, with the left image showing a breathing circle during the measured inhalation and right during exhalation.



XSL•FO
RenderX

Procedure

Informing and Including Participants

First, the health care professional would ask the patient for approval to be approached by the researchers and to provide them a flyer and information letter. If the patient expressed interest in participating, an appointment was scheduled with the patient to discuss the study. During this meeting, the researcher informed the patient about the time commitment and measurement methods. The health care professional was not present at this meeting to avoid role conflicts that could potentially hinder the voluntary decision to participate due to the dependency relationship with their health care professional. If the patient felt adequately informed and decided to participate in the study, they signed an informed consent. Next, the researcher helped the patient in installing the Avicenna app and assisted them in completing the first set of questions regarding stress, emotional tension, and anger on Avicenna, for instance, by reading them aloud.

Study Procedure

The SCED was conducted between May 2022 and October 2022 by one researcher (LKH). An overview of the SCED procedure is provided in Table 1.

Table 1. Overview of the alternating treatment design conducted with forensic inpatients from a Dutch forensic psychiatric organization (N=6).

	3 days	1 day	3 days	1 day	3 days	1 day	3 days	1 day	3 days
P1	Baseline (A)	DEEP (B)	No DEEP (A)	DEEP (B)	No DEEP (A)	DEEP (B)	No DEEP (A)	DEEP (B)	No DEEP (A)
P2	Baseline (A)	DEEP (B)	No DEEP (A)	DEEP (B)	No DEEP (A)	DEEP (B)	No DEEP (A)	DEEP (B)	No DEEP (A)
P3	Baseline (A)	DEEP (B)	No DEEP (A)	DEEP (B)	No DEEP (A)	DEEP (B)	No DEEP (A)	DEEP (B)	No DEEP (A)

In the B phase, DEEP was introduced to the participant. During a period of 20 days, DEEP was introduced 4 times, based on a previous study on DEEP [11]. The researcher visited the location to prepare the DEEP session and supported the patient in its use. DEEP was started via the laptop, which meant that the researcher could watch along with what the participant was seeing through the headset. Before the first session, the researcher provided participants with a brief explanation of the DEEP concept, deep breathing, and what they would be doing during the DEEP sessions. During the first session, the researcher also provided some help if a participant would get stuck or felt lost in the game. The other 3 sessions were done by the patient independently, while the researcher was there to observe.

Semistructured Interviews

A week after the completion of the SCED, 2 separate semistructured interviews (mean 38, SD 4.23 min) were conducted with participating patients and their daily health care professionals to assess whether they experienced an effect of DEEP on the level of stress, emotional tension, and anger of the participant and if any other benefits or disadvantages for the participants and their behavior had been observed. This was further expanded by asking the participants whether DEEP had an effect during the session, the day of the session, and in their daily lives in general. The health care professionals were asked the same types of questions about the patient's experienced stress, emotional tension, and anger, and they were also asked about the behavior of the participant during the SCED.

Data Analysis

Visual Analyses

The responses to the ES questions resulted in a maximum of 60 data points per participant (3 times per day for 20 days). Time series graphs were constructed per participant in R (version 4.0.0; R Foundation for Statistical Computing) using the *scplot* package (version 0.3.8), which is an add-on package to the *scan* package (version 0.60.0) for visualizing single-case data [28]. Each graph illustrated stress, emotional tension, and anger over time, with days on which DEEP was used marked. Visual analysis was primarily used to gain an overall impression of the data [29,30]. Of the 6 aspects of visual analysis to consider in single case research [29], the level, variability, and trend of each phase were inspected. Level refers to the difference in mean values between phases. Variability refers to the fluctuation of the data, in this case reflected by the data's range. Trend refers to the slope of the best-fitting line for data within a phase and is used to examine whether the outcomes decreased or increased over time and whether the introduction of the intervention changed this trajectory [29].

Repeated Linear Mixed Model Analysis

The data were statistically analyzed using linear mixed models (LMMs) embedded in the MultiSCED web application [31]. MultiSCED is an interactive web application that provides R functionality. LMMs compute 4 parameters—intercept, time effect, intervention effect, and interaction of intervention over time—for each participant, enabling phase-wise comparisons at the participant level. Using MultiSCED, individual linear models were constructed for each participant to determine the individual effects of DEEP on stress, emotional tension, and anger. The additional use of SCED-specific fit indices to statistically examine the nonoverlap between data points across phases A and B such as nonoverlap of all pairs [32], treatment as usual–uncontrolled [33], and typicality of level change [34] were considered. However, this approach was abandoned as half of the participants did not reach the sufficient number of 5 data points in the B phase required for reliable effect sizes [34].

Semistructured Interviews

A deductive approach was used to analyze the interview data [35]. This way, potential findings from both the ES method (ESM) data and interviews complement each other. The interviews were recorded and transcribed verbatim. Afterward, the transcriptions were coded deductively for the outcome measures stress, emotional tension, and anger. Any relevant fragments regarding other outcomes measured were inductively coded. On the basis of these fragments, an initial version of the

XSL•FO
RenderX

coding scheme with main code and subcodes was established [36]. This coding scheme was then independently applied to the first 3 interviews by the first coder (LKH), which then was sent to the second coder (HK). An agreement rate of 81% was found, which is acceptable in the coding process [37].

Ethical Considerations

The Single-Case Reporting Guideline in Behavioral Interventions (SCRIBE) was used to design and report this study [38]. The checklist can be found in Multimedia Appendix 1 [38]. Ethics approval was given by the Ethics Committee of the University of Twente (Behavioral, Management and Social Sciences, no. 220190). Each patient was thoroughly informed about the study through information folders, information letters, and clarification from the researcher. Before participating, every patient signed an informed consent form. As compensation for their participation, all patients received 2 vouchers worth €10 (US $10) halfway through and at the end of the study. To ensure privacy and confidentiality, personal data regarding age, gender, and psychiatric disorder were stored on the secured research drive of the University of Twente. Moreover, names of patients and therapists were omitted from the transcripts, and only participant codes were used to connect quantitative and qualitative data, resulting in a pseudoanonymized dataset.

Results

Overview

The results of the ES and interviews are reported for each participant individually. The interview results per participant are reported using the codes regarding the outcomes measures, that is, stress, emotional tension, and anger. The coding schemes on group level can be found in Multimedia Appendix 1.

Descriptive Statistics

The mean scores on stress, emotional tension, and anger per phase are depicted in Table 2. Half of the participants (3/6, 50%) showed lower average levels of emotional tension (P4-P6), and most lower levels of anger (P1 and P4-P6) and stress (P1 and P3-P6). However, other participants reported more stress (P2), emotional tension (P2 and P3), and anger (P2) on DEEP days.

XSL•FO
RenderX

Table 2. Descriptive statistics of experience sampling data on emotional tension (ET), stress, and anger per participant per phase (scored with a 1-5–point Likert scale).

Participant and outcome	Reported range in Likert scale (1-5)	A phase, mean (SD)	B phase, mean (SD)	Total missing datapoints, for scoring stress, anger and ET combined, n (%)
P1				
Stress	3-5	4.00 (0.24)	3.75 (0.5)	—[a]
ET	3-4	3.92 (0.28)	4 (0)	20 (33)
Anger	2-4	3.39 (0.60)	3 (0.82)	—
P2				
Stress	2-5	2.28 (0.75)	3 (1.41)	—
ET	2-5	2.30 (0.76)	3 (1.41)	12 (20)
Anger	1-5	2.39 (0.93)	3 (1.41)	—
P3				
Stress	2-5	2.08 (0.49)	2 (0)	—
ET	1-4	2.03 (0.37)	2.5 (1)	18 (30)
Anger	2-4	2.05 (0.32)	2 (0)	—
P4				
Stress	2-4	2.12 (0.45)	2 (0)	—
ET	2-4	2.26 (0.66)	2 (0)	15 (25)
Anger	2-5	2.23 (0.68)	2 (0)	—
P5				
Stress	2-4	2.59 (0.69)	2.50 (0.71)	—
ET	2-4	2.42 (0.60)	2 (0)	4 (7)
Anger	1-3	2.13 (0.39)	2 (0)	—
P6				
Stress	1-4	2.64 (1.01)	2.25 (1.26)	—
ET	1-4	2.98 (0.93)	2.25 (1.26)	1 (2)
Anger	1-5	1.82 (1.14)	1.50 (0.58)	—

[a]Not applicable.

Results of ES and Interviews

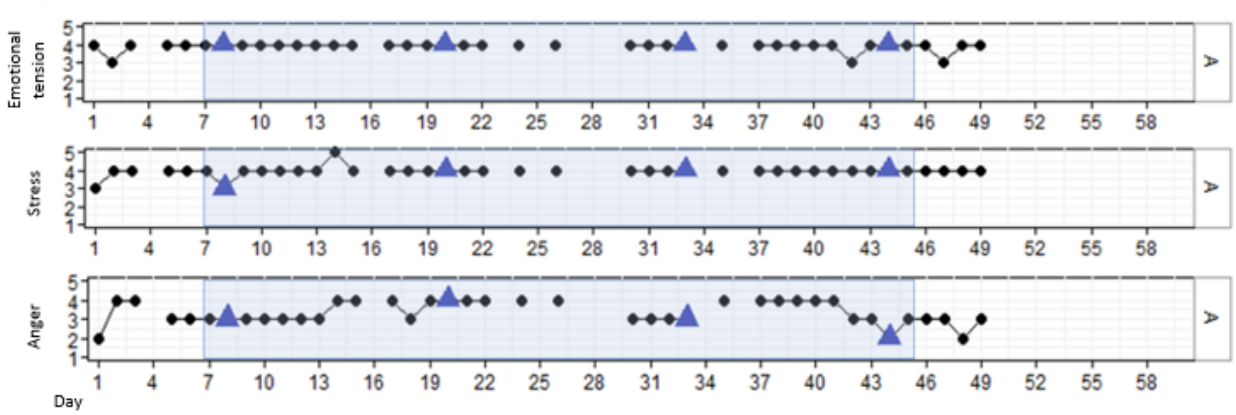
The results of the visual analyses, LMMs, and interviews are described per participant. At the end of this section, a synthesis of the qualitative and quantitative results is presented. An overview of the LMM outcomes can be found in Multimedia Appendix 1.

Results for P1

P1 was a man aged 44 years who, according to himself, showed difficulty controlling his temper when others irritated or did not agree with him. During the study, P1 was enthusiastic about working with DEEP and was interested in how breathing in a VR environment would help him regulate his emotions. P1 completed all 4 DEEP sessions with no difficulty. Figure 2 illustrates that P1 showed consistent levels of high emotional tension and stress independent of DEEP use. In addition, a significant momentary reduction in stress was found after session 2 (β=–.865; *P*=.005). However, this effect disappeared over time (β=.023; *P*=.02). Although no significant changes over time were found, anger (β=–.001; *P*=.93) seemed to variate more than stress (β=–.002; *P*=.57) or emotional tension (β=–.002; *P*=.52).

XSL•FO
RenderX

Figure 2. Stress, emotional tension, and anger in P1 over time, measured with experience sampling. The light blue marking represents the period DEEP was introduced and withdrawn. The blue triangles indicate the first data point after a DEEP session.



The interview outcomes and examples of quotes regarding those outcomes are shown in Table 3. During the interview, P1 indicated that he found DEEP a pleasant experience. P1 felt that DEEP was more of a helpful reference to understand deep breathing than a direct intervention in increasing emotion regulation. His advice was to not only offer DEEP as a stand-alone intervention but also provide patients with assignments, which can help to better apply deep breathing into daily life situations. Because P1 was trying out for assisted living, he had little contact with health care professionals during this study. Therefore, his health care professionals had no clear view on his daily emotions and behavior during this study.

Table 3. Results of the semistructured interview with P1 regarding the experienced effects on emotional tension, stress, and anger.

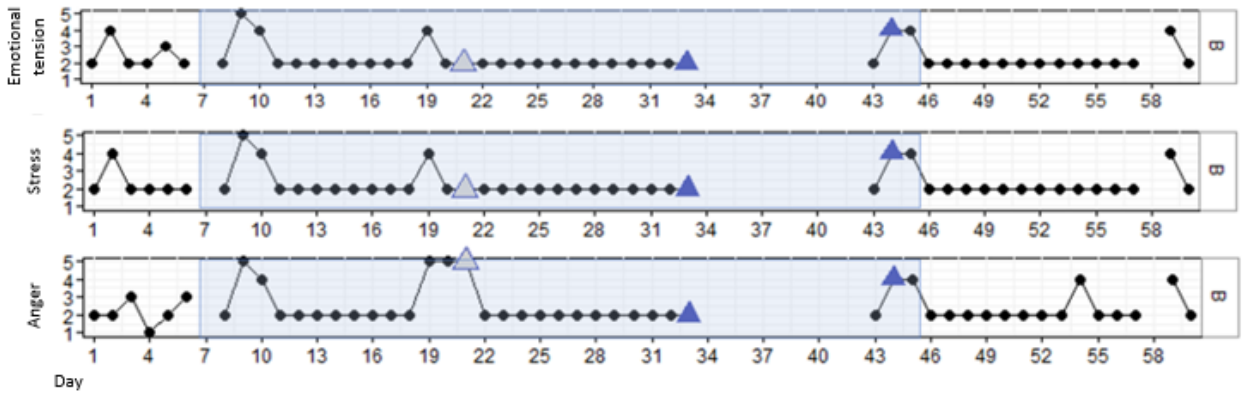
Outcome measure	Times mentioned	Example quotes from P1
Emotional tension	5	"DEEP became a reference for me. I do the session and think back on it. Like oh, I feel tense, let's pretend that I am doing DEEP. Just in my head, feeling like a little fish."
Stress	2	"And I've actually had quite a bit of trouble with panic attacks and stress. But it's all pretty much gone now. And when I start feeling a bit off, I just focus on my breathing."
Anger	1	"And it helped just as well with anger. Absolutely. Simply by breathing. Just going back. Back to basics, back to swimming and just breathing calmly."

Results for P2

P2 was a male patient aged 23 years who, according to his health care professionals, was enthusiastic about his treatment but also impulsive and manipulative in nature. He was glad to participate in this study and was curious to find out whether the deep breathing skills he learned in therapy could support him during a DEEP session. P2 completed 3 out of 4 DEEP sessions. The first session was canceled due to a conflict between him and another inpatient. Figure 3 shows that P2 reported some variation in stress, emotional tension, and anger at the start of the study, which stabilized after data point 10 (emotional tension and stress) and data point 22 (anger). In addition, all outcomes temporarily increased after his last use of DEEP. However, no significant results in stress (β=.002; *P*=.57), emotional tension (β=–.006; *P*=.29), or anger (β=–.002; *P*=.78) were found over time or directly after DEEP sessions.

XSL•FO
RenderX

Figure 3. Stress, emotional tension, and anger in P2 over time, measured with experience sampling. The light blue marking represents the period DEEP was introduced and withdrawn. The blue triangles indicate the first data point after a DEEP session. The gray triangles indicate a DEEP session in which no measurements were taken.



During the interview, P2 indicated that he really liked the first DEEP session but that the last session felt repetitive as he was doing the same thing each time. Moreover, he found the DEEP sessions relaxing, but did not find it to be helpful during moments of conflict. The health care professional of P2 did not really notice any changes in behavior after the DEEP sessions. She did acknowledge that it might have had a positive effect on his well-being, as he was more aware of his breathing. However, during a conflict, she did not notice that he was less angry or that he was able to regulate his emotions through deep breathing (Table 4).

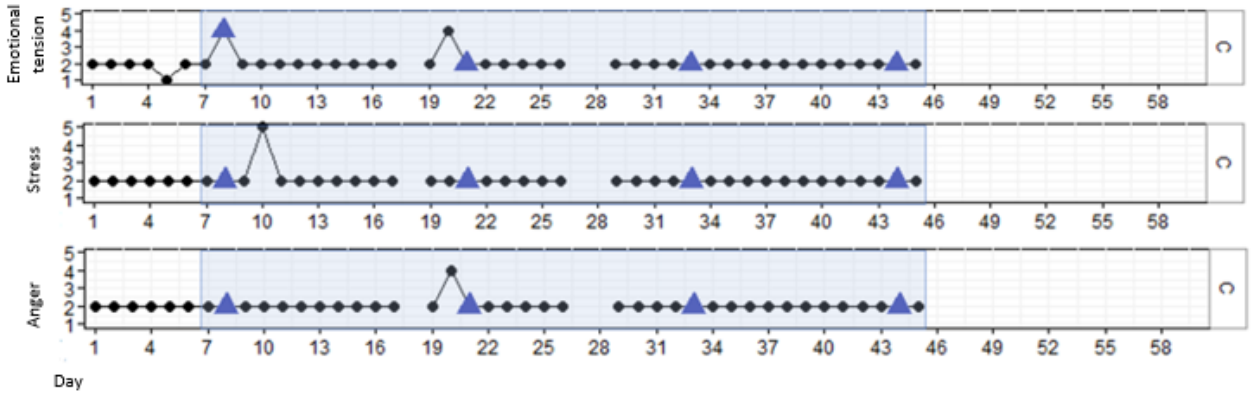
Table 4. Results of the semistructured interview with P2 regarding the experienced effects on emotional tension, stress, and anger.

Outcome measure	Times mentioned	Example quotes from P2
Emotional tension	5	"I would recommend DEEP to other patients. Not everyone would maybe benefit, but it could help most of them with developing breathing skills and a better way to relax and become less tense."
Stress	1	"I did find the DEEP sessions to be effective in reducing my stress. Because I found that when I really focused on my breathing I became more calm and mindful."
Anger	1	"But I do notice that with DEEP, whenever I get stressed or angry or things like that, I do pay attention to my breathing. I don't realize it. It's already ingrained in my system [due to earlier treatment that used breathing exercises] that I do that."

Results for P3

P3 was a man aged 31 years. His health care professionals indicated that he sometimes struggles to access his emotions and talk about them. P3 was interested in participating because he has worked in the IT field and was intrigued by the technology behind the VR game and how the biofeedback in DEEP operates. P3 completed all 4 DEEP sessions. Figure 4 shows that P3 reported low levels and little fluctuation of stress, emotional tension, and anger throughout the study. In addition, his reported emotional tension was significantly increased after the first session (β=1.875; $P<.001$). However, this effect did not remain over time (β=–.053; $P=.002$), indicating only a momentary increase in tension with first-time use of DEEP. Finally, no changes over time in stress (β=–.006; $P=.34$), emotional tension (β=.002; $P=.71$), and anger (β=–.001; $P=.85$) were found.

Figure 4. Stress, emotional tension, and anger in P3 over time, measured with experience sampling. The light blue marking represents the period DEEP was introduced and withdrawn. The blue triangles indicate the first data point after a DEEP session.



XSL•FO
RenderX

During the interview, P3 stated that he expected more from the DEEP sessions. He felt that the VR showed some technical defects and that it did not help him in reducing negative emotions. According to his health care professional, P3 was not entirely engaged in DEEP and what it might teach him. He was more involved in the technical part, such as the VR and abdominal belt, and expressed annoyance that it did not always work correctly (Table 5).

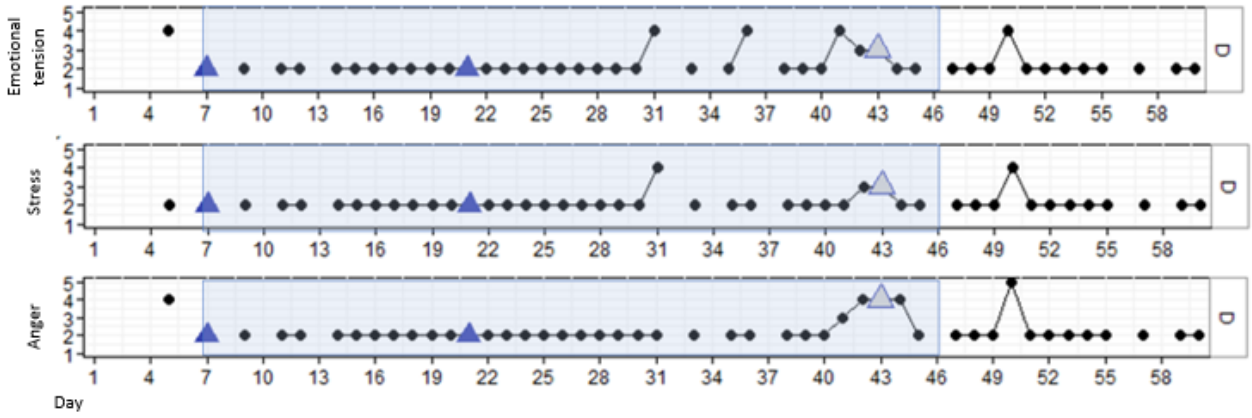
Table 5. Results of the semistructured interview with P3 regarding the experienced effects on emotional tension, stress, and anger.

Outcome measure	Times mentioned	Example quotes from P3
Emotional tension	2	"I was able to focus on my breathing during DEEP, which made me feel relaxed. But I already knew this from going to the gym and deep breath during my exercises."
Stress	1	"I don't know if DEEP helps with reducing stress. I would say that you would need more triggers to test that. Like that things get a bit scary in the game and that you have to keep breathing to calm yourself down."
Anger	2	"Something just came to mind. I don't even remember what it was about, but I was irritated recently. But expressed it calmly, and was immediately vocal about it, you know. But then, did I take a breath? No, that feels pointless."

Results for P4

P4 was a man aged 24 years. According to his health care professionals, he was a quiet man who had difficulty recognizing and articulating his own emotions. P4 completed 3 of 4 sessions. The third session was canceled due to the researcher being unavailable unexpectedly. As is shown in Figure 5, P4 reported low stress, emotional tension, and anger during the first half of the study. More variety in these outcomes was observed in the second half of the study. However, no significant results in stress (β=.004; *P*=.44), emotional tension (β=.000; *P*=.97), and anger (β=.004; *P*=.61) were found over time, nor was a direct effect observed after a DEEP session.

Figure 5. Stress, emotional tension, and anger in P4 over time, measured with experience sampling. The light blue marking represents the period DEEP was introduced and withdrawn. The blue triangles indicate the first data point after a DEEP session. The gray triangles indicate a DEEP session in which no measurements were taken.



During the interview, P4 stated DEEP sessions to be a fun and relaxing experience. He also felt that after a DEEP session, he was able to stay calm the rest of the day, even when conflict occurred. However, P5 was hesitant to state that DEEP helped him using his deep breathing in daily life situations. Except for during a DEEP session, he did not use his deep breathing to cope with negative emotions, such as anger. His health care professional was also hesitant to ascribe a direct effect of DEEP on P5's behavior. She did find P5 to be calm and tired after a DEEP session and indicated that it provided him with more insight on deep breathing. However, she did not find him less tense, stressed, or angry during the SCED, as he was already behaving well in the clinic (Table 6).

Table 6. Results of the semistructured interview with P4 regarding the experienced effects on emotional tension, stress, and anger.

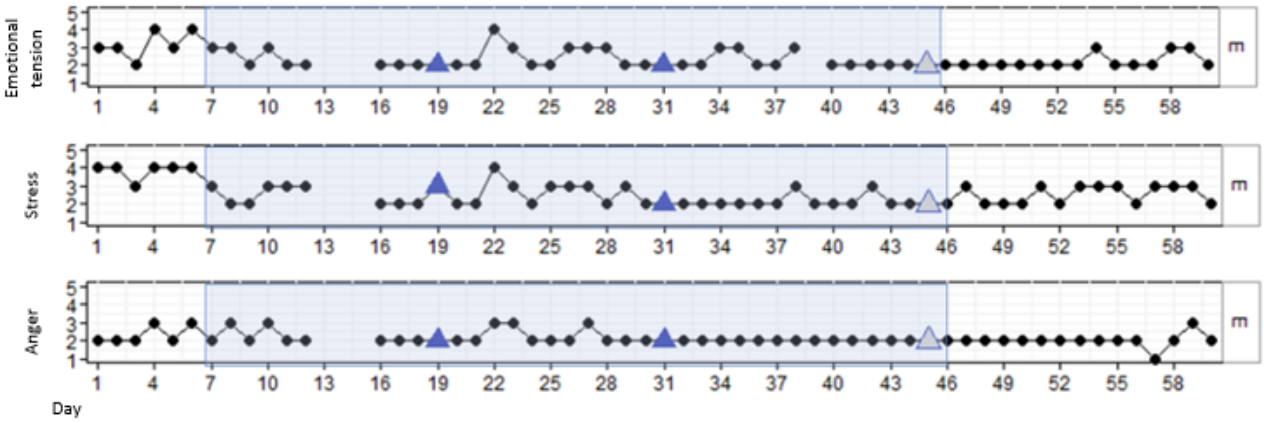
Outcome measure	Times mentioned	Example quotes from P4
Emotional tension	6	"I loved the DEEP sessions. It was nice and really relaxing. Just to be in a calm environment made me less tense already."
Stress	4	"DEEP gave me an empty mind. I had nothing to think or stress about. The mornings you [researcher] came around with DEEP were nice days. I could really relax."
Anger	4	"After the session we were cooking, and there was someone bossing people around, and normally I would say: do something yourself for once! Normally, I would react with irritation, but now I was very calm and could respond more calmly."

XSL•FO
RenderX

Results for P5

P5 was a man aged 37 years. According to both him and his health care professional, he did not always recognize when his anger was increasing and could suddenly explode. He was curious about how DEEP could help him in recognizing his increasing emotional tension and using deep breathing techniques in a timely manner. P5 completed 3 of 4 DEEP sessions. The first session was canceled due to a technical problem with the VR set. As is shown in Figure 6, P5 reported fluctuations on the 3 outcome measures. The DEEP sessions were consistently associated with the lowest scores on emotional tension, stress, and, to a lesser extent, anger. In addition, a significant decline over time in emotional tension (β=–.012; *P*=.006), stress (β=–.014; *P*=.007), and anger (β=–.007; *P*=.02) was found, with high to moderate levels of stress until day 9 and moderate to low levels of stress after that.

Figure 6. Stress, emotional tension, and anger in P5 over time, measured with experience sampling. The light blue marking represents the period DEEP was introduced and withdrawn. The blue triangles indicate the first data point after a DEEP session. The gray triangles indicate a DEEP session in which no measurements were taken.



During the interview, P5 stated that DEEP was a helpful intervention for him to relax and focus on his own body. However, P5 was not able to use his deep breathing techniques and reduce his anger during conflicts or stressful situations. He stated that during those moments, he is so engaged in the conflict that he is not thinking about his bodily signals or what he should do to prevent aggression. His health care professional agreed with these statements. She acknowledged that P5 really felt happy and relaxed after each DEEP session and that he was able to keep calm during most conflicts. However, the health care professional did not find DEEP to be effective in supporting P5 during stressful moments or aggressive outbursts (Table 7).

Table 7. Results of the semistructured interview with P5 regarding the experienced effects on emotional tension, stress, and anger

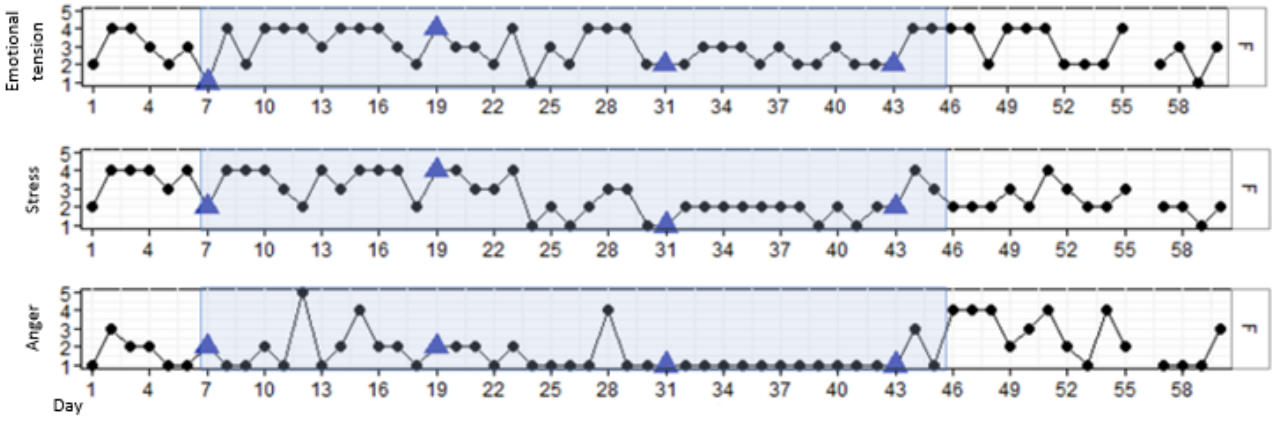
Outcome measure	Times mentioned	Example quotes from P5
Emotional tension	5	“You know, one thing I’ve become more aware of is my breathing. Even before getting out of bed, I sometimes pay attention to my abdominal breathing. So, I’m conscious that breathing can be a momentary relief of tension to start and get through the day.”
Stress	2	“I do feel less stressed, but I am not entirely sure if it can be linked to doing DEEP. I think I have to do more DEEP sessions to discover that.”
Anger	2	“During DEEP I was in a calm environment to chill and get out of my head. But I don’t know if those breathing exercises work if I am in a fight with a family member. I don’t think so to be honest.”

Results for P6

P6 was a man aged 33 years. According to his health care professional, he often struggled to regulate his negative emotions in a healthy way. DEEP appealed to P6 because he wanted to find distraction through it and calm his thoughts. P6 completed all 4 DEEP sessions. However, the second session for P6 was interrupted due to a signal disruption between the VR headset and the VR sensor after 8 minutes. The session was eventually completed but the participant did experience more discomfort than usual, because his vision kept shifting, causing difficulty to navigate properly. As is shown in Figure 7, P6 reported both highly fluctuating levels of stress, emotional tension, and anger. His findings are mixed, with DEEP sessions being associated with both the least (session 1; day 2) and highest (session 2; day 6) emotional tension. However, anger levels did seem to decrease with the use of DEEP, notably after session 3 (day 10), which was followed by a longer period of stability. In addition, a significant decline of stress was found over time (β=–.029; *P*<.001) but not in emotional tension (β=–.009; *P*=.22) or anger (β=.009; *P*=.33). In addition, as neither of the interaction effects are significant, these declines cannot be attributed directly to a momentary effect of DEEP.

XSL•FO
RenderX

Figure 7. Stress, emotional tension, and anger in P6 over time, measured with experience sampling. The light blue marking represents the period DEEP was introduced and withdrawn. The blue triangles indicate the first data point after a DEEP session.



During the interview, P6 expressed being pleasantly surprised by how much he liked the DEEP sessions. He found the environment and music of DEEP to be a great distraction from his negative thoughts and felt that it helped him to cope with his negative emotions. According to his health care professional, P6 really experienced that he can use his deep breathing to reduce emotional tension and stress. She stated that, although there were some conflicts during the data collection period, P6 clearly was less angry and explosive toward the health care professionals (Table 8).

Table 8. Results of the semistructured interview with P6 regarding the experienced effects on emotional tension, stress, and anger.

Outcome measure	Times mentioned	Example quotes from P6
Emotional tension	6	"What really stuck with me was that I started every DEEP session tense and that I came out of the session calm and happy."
Stress	2	"I do remember that, during the DEEP sessions, I was able to breath in a deep and calm way which felt really nice and made my body relax as well."
Anger	3	"I've had quite a rough couple of weeks. And after the DEEP sessions, I noticed that I didn't immediately react to conflicts, but found a way to cope. I didn't explode right away but managed to keep myself together. Also, because I emerged from such a session feeling so relaxed."
Self-efficacy[a]	2	"I know that if I explode, I bring a lot of tension to the staff and the group. DEEP helped me to respond more calmly. But now the study has ended, and I don't know how I will do without the help of DEEP."

[a]This was not a predetermined outcome measure but additionally mentioned by P6.

Synthesis of the Results

In this study, we used multiple research methods to assess the effectiveness of DEEP on stress, emotional tension, and anger in forensic psychiatric patients. In Table 9, we juxtaposed the quantitative and qualitative results per participant. This provides an overview on group level and an individualized examination of whether they paint a consistent picture or present contradictory results. Some discrepancies were observed between quantitative and qualitative results. Two patients showed no significant changes quantitatively, but reported experiencing effects of DEEP in the interviews. Two other patients displayed short-term significant changes in emotional tension or stress, with one patient reporting matching experiences and the other reporting no effects. The final 2 patients showed significant changes over time, aligning with their reported experiences.

XSL•FO
RenderX

Table 9. Overview of quantitative and qualitative results per participant (N=6).

ID	Quantitative outcomes	Qualitative outcomes
P1	• Variation in stress, ET[a], and anger • Momentary decrease of stress • No significant changes over time	• Experienced short-term effect on ET and stress • No experienced long-term effect
P2	• Stabilization of stress, ET, and anger in second half SCED[b] • No significant changes over time	• Experienced short-term effects on ET • No experienced long-term effects
P3	• Low stress, ET, and anger throughout SCED • Momentary increase of ET • No significant changes over time	• No experienced short-term effects • No experienced long-term effects
P4	• More fluctuation of stress, ET, and anger in second half SCED • No significant changes over time	• Experienced short-term effect on ET, stress, and anger • Experienced long-term effect on ET
P5	• Fluctuation, but lower levels in stress, ET, and anger after DEEP sessions. • Significant reduction of stress, ET, and anger over time	• Experienced short-term effect on ET and stress • Experienced long-term effect on ET and stress
P6	• Elevated levels and fluctuation of stress, ET, and anger throughout SCED • Significant reduction of stress over time	• Experienced short-term effect on ET, stress, and anger • Experienced long-term effect on stress and anger

[a]ET: emotional tension.

[b]SCED: single-case experimental design.

Discussion

Principal Findings

This study aimed to investigate whether DEEP, through its immersive focus on relaxation and diaphragmatic breathing techniques, is able to decrease stress, emotional tension, and anger in forensic psychiatric inpatients. First, much variation in effectiveness was found between each participant. In 2 participants, a momentary effect on emotional tension and stress right after a DEEP session was found. Two other participants showed a decrease in stress, emotional tension, and anger over time, with no direct effect of DEEP. The last 2 participants did not show any significant effects. These results underline that DEEP is no one-size-fits-all solution and that users might benefit from DEEP in diverse ways or not at all. Second, most participants felt enthusiastic about DEEP and found it to be a helpful intervention to enhance short-term relaxation. Some participants also found DEEP to be able to teach them diaphragmatic breathing, which, in their experience, helped them reduce aggressive outbursts or conflicts. However, most patients were not yet able to put what they had learned in DEEP into practice, which was also indicated by their health care professionals. Finally, some discrepancies were found in the synthesis of the ES and qualitative results; although most patients experienced a positive effect of DEEP on their well-being, 67% (4/6) of them showed no significant change in emotional tension, stress, or anger.

Variation in Effectiveness

An important finding of this study is that DEEP seems able to reduce momentary emotional tension, stress, and anger and induce short-term relaxation in some but not all forensic inpatients. One possible explanation for this is that the participants in this study exhibit heterogeneity in their history of offenses, treatment objectives, cognitive abilities, and psychosocial characteristics [4,39], which might explain the differences in their experiences during the DEEP sessions and variation in effectiveness. This underlines earlier research, which states that there is no one-size-fits-all approach to support all psychiatric patients in their treatment and that a personalized approach is essential [40], which is in particular the case for forensic patients [3]. Furthermore, DEEP was provided as a stand-alone intervention within a research context without a more personalized instruction. The participants stated that they had trouble applying what they learned in DEEP into daily life situations. This shows that a structural integration of DEEP into existing treatment could support patients in generalizing deep breathing skills, making the intervention more effective as is supported by earlier studies. Finally, some technical difficulties with the abdominal waistband occurred during the study, which limited the impact biofeedback could have on participants. Future research could further investigate these reasons for variation in effectiveness, for example, by focusing on patient-, contextual-, and intervention-related factors [41-43].

Generalization of Deep Breathing Skills

Another finding of this study is that DEEP was able to induce short-term effects on stress, emotional tension, and anger. However, a long-term effect that could directly be ascribed to DEEP was not found in any participants. Similar to earlier evaluation studies of DEEP, the DEEP sessions were conducted as a “stand-alone” intervention, without an extensive introduction or debriefing from a health care professional [8,11,26]. In the interviews, patients and health care professionals stated that more information about diaphragmatic breathing could help them to actively use this as a coping

XSL•FO
RenderX

strategy to prevent aggression in their daily lives. Therefore, DEEP might require some form of psychoeducation and support with integration in daily life to generalize the skills that are taught in DEEP [7]. In addition, other studies recommended that technological interventions, such as VR, should be embedded within existing treatment as opposed to being used as a separate tool [6,44,45]. Therefore, future research should focus on determining the most optimal way to integrate DEEP in a treatment context where health care professionals are conducting the DEEP sessions.

Using Deep Breathing During Conflicts

A third finding of this study is that most participants often forgot to apply diaphragmatic breathing during conflicts because they were already too involved or angry. An explanation for this is that patients were often not aware of their increasing emotional tension or anger, which is a common problem in forensic psychiatric patients [46]. Consequently, necessary precondition for applying diaphragmatic breathing to prevent aggression seems to be that patients recognize their increasing tension early on [18,47,48]. This degree of recognition and trust in one's own bodily signals—which is called interoceptive awareness—is indeed found to be related to the improvement of emotion regulation skills [48]. Therefore, it would be interesting to determine whether the effect of DEEP or other emotion regulation interventions can be enhanced if the user's interoceptive awareness increases. Consequently, a recommendation for future research is to combine DEEP with other biofeedback devices such as wearables that can be used to increase interoceptive awareness in daily life [46,47].

Strengths and Limitations

First, by using a mixed methods single-case approach, this study was able to differentiate individual responses to a novel intervention. However, the data quality of the current SCED study limited opportunities to use SCED-specific effect sizes such as treatment as usual–uncontrolled, nonoverlap of all pairs, and typicality of level change and, as such, to be able to draw robust conclusions about the effectiveness of DEEP. To meet standards in future studies, at least 5 data points in each phase should be produced. This could be successfully reached by extending the DEEP phase by embedding the VR sessions in treatment as usual. This way more data points are easily collected and long-term changes in outcome measures might be reached. Second, by using convenience sampling as a sampling method, researchers and health care professionals were able to recruit and include participants, while ensuring a time- and cost-effective procedure [49]. However, it could have led to selection bias as health care professionals may have only recruited patients that they deemed motivated to participate to ensure that they were able to complete the study [49,50]. Therefore, it is important to remain careful when generalizing study outcomes, particularly as the vulnerable forensic psychiatric patients could not be involved in this study. Third, attention should be paid to the impact of technical issues, as some DEEP sessions were canceled or interrupted due to signal loss between the abdominal belt and the VR equipment. It should be noted that these issues could have inhibited the possible effectiveness of DEEP. Fourth, this study used ES, which is a suitable way to study real-time changes in mental states [51,52]. As a significant decrease in stress, emotional tension, and anger over time—without a direct effect of DEEP—was found in 2 patients, ES could have provided an unforeseen effect on their well-being, just by asking them to reflect on their emotions thrice a day. However, by using a mixed methods design, these possible unforeseen effects have been partly addressed in semistructured interviews because quantitative ESM data were supplemented by experienced effectiveness. Finally, emotional tension, stress, and anger are complex constructs to measure, particularly using ES [53]. Some participants mentioned effects of DEEP that were not reflected in the ESM data, raising questions about the extent to which the items were fully valid. Finally, participants scored their stress, emotional tension, and anger by using a 5-point Likert scale, an often-used scoring system in validated questionnaires [54,55]. However, this resulted in less detailed information about the emotional state of the participants; although an increase from 1 to 3 may seem significant, it simply indicates that the participant selected "(3) neutral/do not know," which does not directly reflect an improvement. To encounter this, a visual analog scale of 0 to 10 or 0 to 100 might have been a better way to assess the outcomes in a more specific, sensitive way.

Conclusions

This study has shown that DEEP seems a promising intervention for some, but not all, forensic psychiatric inpatients to practice deep breathing techniques in a novel and engaging way. DEEP is not a one-size-fits-all intervention that works for every patient, as the effectiveness in decreasing emotional tension, stress, and anger varied among the participants. Despite variation in quantitative results, both inpatients and the health care professionals found DEEP to be of added value in providing short-term relaxation. In addition, DEEP seems to be an engaging way to acquire deep breathing as a coping strategy. However, generalization to daily lives was difficult, which can be explained by problems with interoceptive awareness and lack of integration in treatment protocols. This underlines the need for a structural integration of interventions such as DEEP in existing treatment programs or psychoeducation. Future research should therefore focus not only on whether DEEP is effective, but also on why and for whom it is effective, by identifying patient-, context-, and intervention-related factors.

Acknowledgments

This study was supported by a grant from Kwaliteit Forensische Zorg to HK (KFZ 2020-146). The authors would like to thank the managers and health care professionals of the participating clinics for their time and effort to help conduct this study. The authors would also like to thank the patients for participating in this study.

XSL•FO
RenderX

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

HK and LKH contributed to the design and planning of the study. LKH contributed to data collection. All authors contributed to the interpretation and reporting of the results and approved the manuscript.

Conflicts of Interest

JW works with Explore DEEP Ltd as a scientific advisor. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Overview of Single-Case Reporting Guideline in Behavioral Interventions (SCRIBE) checklist, linear regression models, and coding schemes.
[DOCX File , 32 KB - formative_v9i1e65206_app1.docx]

References

1. Meynen G. Walls and laws: structural barriers to forensic psychiatric research. Eur Psychiatry 2017 Jul 23;44:208-209. [doi: 10.1016/j.eurpsy.2017.04.010] [Medline: 28646733]
2. Sygel K, Wallinius M. Immersive virtual reality simulation in forensic psychiatry and adjacent clinical fields: a review of current assessment and treatment methods for practitioners. Front Psychiatry 2021 May 28;12:673089 [FREE Full text] [doi: 10.3389/fpsyt.2021.673089] [Medline: 34122189]
3. Kip H, Bouman YH. Van denken en praten naar doen en ervaren: op zoek naar de toegevoegde waarde van technologie voor de forensische psychiatrie. Sancties 2020;1:249-260.
4. Kip H, Bouman YH. A perspective on the integration of eHealth in treatment of offenders: combining technology and the risk-need-responsivity model. Front Psychiatry 2021 Sep 1;12:703043 [FREE Full text] [doi: 10.3389/fpsyt.2021.703043] [Medline: 34539462]
5. Kip H, Oberschmidt K, Bierbooms J, Dijkslag D. Technologie in de forensische zorg - Crossing borders. Utrecht: Kwaliteit Forensische Zorg (KFZ). 2019. URL: https://www.researchgate.net/publication/334825702_Technologie_in_de_forensische_zorg_-_Crossing_borders [accessed 2024-04-29]
6. Klein Haneveld L, Kip H, Bouman YH, Weerdmeester J, Scholten H, Kelders SM. Een sprong in het diepe: de evaluatie en implementatie van DEEP, een virtual reality biofeedback game, in de forensische zorg. Utrecht: Kwaliteit Forensische Zorg. URL: https://kfz.nl/projecten/een-sprong-in-het-diepe-de-evaluatie-en-implementatie-van-deep [accessed 2024-04-29]
7. Klein Haneveld L, Kip H, Bouman YH, Weerdmeester J, Scholten H, Kelders SM. Exploring the added value of virtual reality biofeedback game DEEP in forensic psychiatric inpatient care-a qualitative study. Front Psychol 2023 Nov 9;14:1201485 [FREE Full text] [doi: 10.3389/fpsyg.2023.1201485] [Medline: 38023054]
8. Weerdmeester JW. Take a DEEP breath: exploring the potential of game-based biofeedback interventions for anxiety regulation. Radboud University. URL: https://repository.ubn.ru.nl/bitstream/handle/2066/239933/239933.pdf?sequence=2&isAllowed=y [accessed 2024-04-29]
9. Schwartz MS, Andrasik F. S. In: Biofeedback: A Practitioner's Guide. 4th edition. New York, NY: The Guilford Press; 2017.
10. Fitry AP, Ramdani VH, Suhanda S, Rosmiati R. Implementation of deep breathing relaxation techniques to control anger at violent behavior. KIAN J 2023 Mar 30;2(1):26-32. [doi: 10.56359/kian.v2i1.237]
11. Bossenbroek R, Wols A, Weerdmeester J, Lichtwarck-Aschoff A, Granic I, van Rooij MM. Efficacy of a virtual reality biofeedback game (DEEP) to reduce anxiety and disruptive classroom behavior: single-case study. JMIR Ment Health 2020 Mar 24;7(3):e16066 [FREE Full text] [doi: 10.2196/16066] [Medline: 32207697]
12. Weerdmeester JW, van Rooij MM, Maciejewski DF, Engels RC, Granic I. A randomized controlled trial assessing the efficacy of a virtual reality biofeedback video game: anxiety outcomes and appraisal processes. Technol Mind Behav 2021 Aug 05;2(2):23. [doi: 10.1037/tmb0000028]
13. Agrawal V, Naik V, Duggirala M, Athavale S. Calm a mobile based deep breathing game with biofeedback. In: Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play. 2020 Presented at: CHI PLAY '20; November 2-4, 2020; Virtual Event p. 153-157. [doi: 10.1145/3383668.3419876]
14. Chen YF, Huang XY, Chien CH, Cheng JF. The effectiveness of diaphragmatic breathing relaxation training for reducing anxiety. Perspect Psychiatr Care 2017 Oct 23;53(4):329-336. [doi: 10.1111/ppc.12184] [Medline: 27553981]
15. Hopper SI, Murray SL, Ferrara LR, Singleton JK. Effectiveness of diaphragmatic breathing for reducing physiological and psychological stress in adults: a quantitative systematic review. JBI Database System Rev Implement Rep 2019 Sep;17(9):1855-1876. [doi: 10.11124/JBISRIR-2017-003848] [Medline: 31436595]

XSL•FO
RenderX

16. Brzozowski A, White RG, Mitchell IJ, Beech AR, Gillespie SM. A feasibility trial of an instructed breathing course in prison to improve emotion regulation in people with substance use difficulties. J Forens Psychiatry Psychol 2020 Dec 10;32(2):308-325. [doi: 10.1080/14789949.2020.1859589]
17. Asberg K. Hostility/anger as a mediator between college students' emotion regulation abilities and symptoms of depression, social anxiety, and generalized anxiety. J Psychol 2013 Sep;147(5):469-490. [doi: 10.1080/00223980.2012.715601] [Medline: 24003591]
18. Pérez-Fuentes MD, Molero Jurado MD, Barragán Martín AB, Gázquez Linares JJ. Family functioning, emotional intelligence, and values: analysis of the relationship with aggressive behavior in adolescents. Int J Environ Res Public Health 2019 Feb 06;16(3):478 [FREE Full text] [doi: 10.3390/ijerph16030478] [Medline: 30736326]
19. Sprague J, Verona E, Kalkhoff W, Kilmer A. Moderators and mediators of the stress-aggression relationship: executive function and state anger. Emotion 2011 Feb;11(1):61-73. [doi: 10.1037/a0021788] [Medline: 21401226]
20. Krasny-Pacini A, Evans J. Single-case experimental designs to assess intervention effectiveness in rehabilitation: a practical guide. Ann Phys Rehabil Med 2018 May;61(3):164-179 [FREE Full text] [doi: 10.1016/j.rehab.2017.12.002] [Medline: 29253607]
21. Cariveau T, Fetzner D. Experimental control in the adapted alternating treatments design: a review of procedures and outcomes. Behav Interv 2022 Jan 25;37(3):805-818. [doi: 10.1002/bin.1865]
22. Sindelar PT, Rosenberg MS, Wilson RJ. An adapted alternating treatments design for instructional research. Educ Treat Children 1985;8(1):67-76.
23. Smith JD. Single-case experimental designs: a systematic review of published research and current standards. Psychol Methods 2012 Dec;17(4):510-550 [FREE Full text] [doi: 10.1037/a0029312] [Medline: 22845874]
24. van Berkel N, Ferreira D, Kostakos V. The experience sampling method on mobile devices. ACM Comput Surv 2017 Dec 06;50(6):1-40. [doi: 10.1145/3123988]
25. A reliable and featureful outcome assessment platform. Avicenna Research. URL: https://avicennaresearch.com/features/outcome-assessment [accessed 2024-04-29]
26. Van Rooij M, Lobel A, Harris O, Smit N, Granic I. DEEP: a biofeedback virtual reality game for children at-risk for anxiety. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2016 Presented at: CHI EA '16; May 7-12, 2016; San Jose, CA p. 1989-1997 URL: https://dl.acm.org/doi/10.1145/2851581.2892452 [doi: 10.1145/2851581.2892452]
27. Aan de slag met de Oculus Rift. Meta Quest. URL: https://www.oculus.com/rift-s/ [accessed 2024-04-29]
28. Wilbert J. Analyzing single-case data with R and scan. zenodo. URL: https://jazznbass.github.io/scan-Book/ [accessed 2024-04-29]
29. Kratochwill TR, Hitchcock J, Horner RH, Levin JR, Odom SL, Rindskopf DM, et al. Single-case design technical documentation. Institute of Education Sciences, U.S. Department of Education. 2010. URL: https://ies.ed.gov/ncee/WWC/Docs/ReferenceResources/wwc_scd.pdf [accessed 2024-04-29]
30. Kratochwill TR, Hitchcock JH, Horner RH, Levin JR, Odom SL, Rindskopf DM, et al. Single-case intervention research design standards. Remedial Spec Educ 2012 Aug 15;34(1):26-38. [doi: 10.1177/0741932512452794]
31. Declercq L, Cools W, Beretvas SN, Moeyaert M, Ferron JM, Van den Noortgate W. MultiSCED: a tool for (meta-)analyzing single-case experimental data with multilevel modeling. Behav Res Methods 2020 Feb 10;52(1):177-192. [doi: 10.3758/s13428-019-01216-2] [Medline: 30972557]
32. Parker RI, Vannest K. An improved effect size for single-case research: nonoverlap of all pairs. Behav Ther 2009 Dec;40(4):357-367. [doi: 10.1016/j.beth.2008.10.006] [Medline: 19892081]
33. Parker RI, Vannest KJ, Davis JL, Sauber SB. Combining nonoverlap and trend for single-case research: Tau-U. Behav Ther 2011 Jun;42(2):284-299. [doi: 10.1016/j.beth.2010.08.006] [Medline: 21496513]
34. Landman W, Bogaerts S, Spreen M. Typicality of level change (TLC) as an additional effect measure to NAP and Tau-U in single case research. Behav Modif 2023 Aug 31;48(1):51-74. [doi: 10.1177/01454455231190741]
35. Plano Clark VL, Huddleston-Casas CA, Churchill SL, O'Neil Green D, Garrett AL. Mixed methods approaches in family science research. J Fam Issues 2008 Apr 01;29(11):1543-1566. [doi: 10.1177/0192513x08318251]
36. Bree RT, Gallagher G. Using Microsoft Excel to code and thematically analyse qualitative data: a simple, cost-effective approach. All Ireland J High Educ 2016;8(2):2811-2824 [FREE Full text]
37. Miles MB, Huberman AM. Qualitative Data Analysis: An Expanded Sourcebook. Thousand Oaks, CA: Sage Publications; 1994.
38. Tate RL, Perdices M, Rosenkoetter U, Shadish W, Vohra S, Barlow DH, et al. The single-case reporting guideline in BEhavioural interventions (SCRIBE) 2016 statement. Remedial Spec Educ 2016 Aug 02;37(6):370-380. [doi: 10.1177/0741932516652893]
39. van Nieuwenhuizen C, Bogaerts S, de Ruijter EA, Bongers IL, Coppens M, Meijers RA. TBS-behandeling geprofileerd een gestructureerde casussenanalyse. Wetenschappelijk Onderzoek- en Datacentrum. 2011. URL: https://repository.wodc.nl/bitstream/handle/20.500.12832/1837/samenvatting_tcm28-71170.pdf?sequence=1&isAllowed=y [accessed 2021-02-06]

XSL•FO
RenderX

40. Pieterse M, Kip H, Martínez RR. The complexity of eHealth implementation: a theoretical and practical perspective. In: Kip H, Beerlage-de Jong N, van Gemert-Pijnen L, Sanderman R, Kelders SM, editors. eHealth Research, Theory and Development: A Multi-Disciplinary Approach. New York, NY: Routledge; 2018.
41. Nielsen SB, Lemire S, Tangsig S. Unpacking context in realist evaluations: findings from a comprehensive review. Evaluation 2021 Oct 25;28(1):91-112. [doi: 10.1177/13563890211053032]
42. Nielsen K, Miraglia M. What works for whom in which circumstances? On the need to move beyond the 'what works?' question in organizational intervention research. Hum Relat 2016 Nov 12;70(1):40-62. [doi: 10.1177/0018726716670226]
43. van Hooft SM, Been-Dahmen JM, Ista E, van Staa A, Boeije HR. A realist review: what do nurse-led self-management interventions achieve for outpatients with a chronic condition? J Adv Nurs 2017 Jun 09;73(6):1255-1271. [doi: 10.1111/jan.13189] [Medline: 27754557]
44. Kip H, Bouman YH, Kelders SM, van Gemert-Pijnen LJ. eHealth in treatment of offenders in forensic mental health: a review of the current state. Front Psychiatry 2018 Feb 21;9:42 [FREE Full text] [doi: 10.3389/fpsyt.2018.00042] [Medline: 29515468]
45. Maples-Keller JL, Bunnell BE, Kim S, Rothbaum BO. The use of virtual reality technology in the treatment of anxiety and other psychiatric disorders. Harv Rev Psychiatry 2017;25(3):103-113 [FREE Full text] [doi: 10.1097/HRP.0000000000000138] [Medline: 28475502]
46. Ter Harmsel JF, Noordzij ML, van der Pol TM, Swinkels LT, Goudriaan AE, Popma A. Exploring the effects of a wearable biocueing app (Sense-IT) as an addition to aggression regulation therapy in forensic psychiatric outpatients. Front Psychol 2023 Mar 10;14:983286 [FREE Full text] [doi: 10.3389/fpsyg.2023.983286] [Medline: 36968738]
47. de Looff PC, Noordzij ML, Nijman HL, Goedhard L, Bogaerts S, Didden R. Putting the usability of wearable technology in forensic psychiatry to the test: a randomized crossover trial. Front Psychiatry 2024 Jun 14;15:1330993 [FREE Full text] [doi: 10.3389/fpsyt.2024.1330993] [Medline: 38947186]
48. Klocek A, Řiháček T. The associations between interoceptive awareness, emotion regulation, acceptance, and well-being in patients receiving multicomponent treatment: a dynamic panel network model. Res Psychother 2023 Jul 26;26(2):e25 [FREE Full text] [doi: 10.4081/ripppo.2023.659] [Medline: 37503659]
49. Golzar J, Tajik O, Noor S. Convenience sampling. Int J Lang Stud 2022;1(2):72. [doi: 10.22034/ijels.2022.162981]
50. Emerson RW. Convenience sampling revisited: embracing its limitations through thoughtful study design. J Vis Impair Blind 2021 Jan 15;115(1):76-77. [doi: 10.1177/0145482x20987707]
51. Kramer I, Simons CJ, Hartmann JA, Menne-Lothmann C, Viechtbauer W, Peeters F, et al. A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial. World Psychiatry 2014 Feb 04;13(1):68-77 [FREE Full text] [doi: 10.1002/wps.20090] [Medline: 24497255]
52. von Klipstein L, Servaas MN, Schoevers RA, van der Veen DC, Riese H. Integrating personalized experience sampling in psychotherapy: a case illustration of the Therap-i module. Heliyon 2023 Mar;9(3):e14507 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e14507] [Medline: 36967959]
53. Palmier-Claus JE, Myin-Germeys I, Barkus E, Bentley L, Udachina A, Delespaul PA, et al. Experience sampling research in individuals with mental illness: reflections and guidance. Acta Psychiatr Scand 2011 Jan;123(1):12-20. [doi: 10.1111/j.1600-0447.2010.01596.x] [Medline: 20712828]
54. Joshi A, Kale S, Chandel S, Pal D. Likert scale: explored and explained. Br J Appl Sci Technol 2015 Jan 10;7(4):396-403. [doi: 10.9734/bjast/2015/14975]
55. Lobsy J, Wetmore A. CDC coffee break: using Likert scales in evaluation survey work. Centers for Disease Control and Prevention. 2014. URL: https://stacks.cdc.gov/view/cdc/136903 [accessed 2024-04-29]

Abbreviations

ES: experience sampling
ESM: experience sampling method
LMM: linear mixed model
SCED: single-case experimental design
SCRIBE: Single-Case Reporting Guideline in Behavioral Interventions
VR: virtual reality

XSL•FO
RenderX

Edited by A Mavragani; submitted 08.08.24; peer-reviewed by S Bogaerts; comments to author 29.10.24; revised version received 18.11.24; accepted 11.12.24; published 12.02.25.

Please cite as:
Klein Haneveld L, Dekkers T, Bouman YHA, Scholten H, Weerdmeester J, Kelders SM, Kip H
The Effect of the Virtual Reality–Based Biofeedback Intervention DEEP on Stress, Emotional Tension, and Anger in Forensic Psychiatric Inpatients: Mixed Methods Single-Case Experimental Design
JMIR Form Res 2025;9:e65206
URL: https://formative.jmir.org/2025/1/e65206
doi:10.2196/65206
PMID:

©Lisa Klein Haneveld, Tessa Dekkers, Yvonne H A Bouman, Hanneke Scholten, Joanneke Weerdmeester, Saskia M Kelders, Hanneke Kip. Originally published in JMIR Formative Research (https://formative.jmir.org), 12.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Machine Learning Approach to Identifying Wrong-Site Surgeries Using Centers for Medicare and Medicaid Services Dataset: Development and Validation Study

Yuan-Hsin Chen[1*], MPH, MD; Ching-Hsuan Lin[2*], MPH, MD; Chiao-Hsin Fan[3], MS; An Jim Long[3], PhD; Jeremiah Scholl[3], PhD; Yen-Pin Kao[4], MD; Usman Iqbal[5], MBA, PharmD, PhD; Yu-Chuan Jack Li[4,6,7,8], MD, PhD

[1]Department of Surgery, Massachusetts General Hospital, Boston, MA, United States

[2]Center for the Evaluation of Value and Risk in Health, Tufts Medical Center, Boston, MA, United States

[3]AESOP Technology Inc, Taipei, Taiwan

[4]Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, New Taipei City, Taiwan

[5]School of Population Health, Faculty of Medicine and Health, University of New South Wales, Sydney, Australia

[6]International Center for Health Information and Technology, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

[7]Department of Dermatology, Wan Fang Hospital, Taipei, Taiwan

[8]Research Center for Artificial Intelligence in Medicine, Taipei Medical University, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Yu-Chuan Jack Li, MD, PhD
Graduate Institute of Biomedical Informatics
College of Medical Science and Technology
Taipei Medical University
No. 301, Yuantong Rd. Zhonghe Dist
New Taipei City, 235
Taiwan
Phone: 886 266382736
Email: jack@tmu.edu.tw

Abstract

Background: Wrong-site surgery (WSS) is a critical but preventable medical error, often resulting in severe patient harm and substantial financial costs. While protocols exist to reduce wrong-site surgery, underreporting and inconsistent documentation continue to contribute to its persistence. Machine learning (ML) models, which have shown success in detecting medication errors, may offer a solution by identifying unusual procedure-diagnosis combinations. This study investigated whether an ML approach can effectively adapt to detect surgical errors.

Objective: This study aimed to evaluate the transferability and effectiveness of an ML-based model for detecting inconsistencies within surgical documentation, particularly focusing on laterality discrepancies.

Methods: We used claims data from the Centers for Medicare and Medicaid Services Limited Data Set (CMS-LDS) from 2017 to 2020, focusing on surgical procedures with documented laterality. We developed an adapted Association Outlier Pattern (AOP) ML model to identify uncommon procedure-diagnosis combinations, specifically targeting discrepancies in laterality. The model was trained on data from 2017 to 2019 and tested on 2020 orthopedic procedures, using *ICD-10-PCS* (*International Classification of Diseases, Tenth Revision, Procedure Coding System*) codes to distinguish body part and laterality. Test cases were classified based on alignment between procedural and diagnostic laterality, with 2 key subgroups (right-left and left-right mismatches) identified for evaluation. Model performance was assessed by comparing precision-recall curves and accuracy against rule-based methods.

Results: The findings here included 346,382 claims, of which 2170 claims demonstrated with significant laterality discrepancies between procedures and diagnoses. Among patients with left-side procedures and right-side diagnoses (603/1106), 54.5% were confirmed as errors after clinical review. For right-side procedures with left-side diagnoses (541/1064), 50.8% were classified as errors. The AOP model identified 697 and 655 potentially unusual combinations in the left-right and right-left subgroups, respectively, with over 80% of these cases confirmed as errors following clinical review. Most confirmed errors involved

XSL•FO
RenderX

discrepancies in laterality for the same body part, while nonerror cases typically involved general diagnoses without specified laterality.

Conclusions: This investigation showed that the AOP model effectively detects inconsistencies between surgical procedures and diagnoses using CMS-LDS data. The AOP model outperformed traditional rule-based methods, offering higher accuracy in identifying errors. Moreover, the model's transferability from medication-disease associations to procedure-diagnosis verification highlights its broad applicability. By improving the precision of identifying laterality discrepancies, the AOP model can reduce surgical errors, particularly in orthopedic care. These findings suggest that the model enhances patient safety and has the potential to improve clinical decision-making and outcomes.

(JMIR Form Res 2025;9:e68436) doi:10.2196/68436

KEYWORDS

patient safety; wrong site surgery; medical errors; machine learning; claim data

Introduction

Wrong-site surgery (WSS) is recognized as the fifth most severe medical error in the United States [1]. A recent study revealed that one-third of inpatient WSS incidents resulted in high-severity injuries, with 7.4% of these cases leading to death [2]. This preventable mistake not only causes a big impact on patients but also imposes considerable financial burdens on the health care system [3]. According to a review in the United States National Practitioner Data Bank, a total of US $1.3 billion was recorded in payouts for surgical never events for 2 decades (1990-2010) [4]. Despite efforts aimed at reducing surgical errors through preprocedure verification, site marking, and a preoperative timeout, WSS events continue to occur. Specifically, the services most frequently responsible for those cases were orthopedic, including spine and intervertebral disc surgery, arthroscopy, and surgery on muscles or tendons [2,5].

The underreporting of WSS further complicates efforts to address this issue. In the United States, only 2% of sentinel events are reported to The Joint Commission, contributing to the persistence of WSS [1]. Moreover, the World Health Organization's 2024 Global Patient Safety Report highlights that only 38% of countries have implemented systems to report preventable and catastrophic medical errors, leading to the widespread underestimation of their true frequency [6]. Among the factors to WSS, inconsistent documentation and inappropriate history and examination information contribute significantly to the occurrence of WSS [2,3]. For example, a patient may have a radiology report indicating a right renal mass from an outside hospital, while a referral document refers to a recent ultrasound report of a left renal mass, leading to a left nephrectomy [2]. Due to the inescapable nature of human error, the ability to automatically identify such discrepancies with the emerging technology could substantially reduce the incidence of WSS.

To address issues of inconsistent documentation and inappropriate history and examination information, a machine learning (ML) model that can automatically detect uncommon associations between diagnosis and surgical procedures may offer a viable solution for identifying and preventing WSS. Recent studies have shown promising results using ML algorithms to detect medication errors, using unsupervised association rule learning to identify unreasonable medication-disease combinations [7,8]. We hypothesized that the similar characteristics between procedure-diagnosis and medication-disease associations could be used to train a model capable of detecting surgical errors. Therefore, this study aims to assess the transferability of an adapted ML model and validate its performance in detecting inconsistencies in surgical procedure documentation.

Methods

Data Source

The Centers for Medicare and Medicaid Services Limited Data Set (CMS-LDS) files provide detailed information on health care services given to beneficiaries. This database includes claims data that record the medical services provided, along with related diagnosis and procedure codes [9]. This dataset is crucial for health care research and quality assessments, helping to measure health care outcomes in various settings [10-12]. We included claims with complete records of both medical and surgical procedures.

Model Construction

In this study, we constructed the adapted version of the Association Outlier Pattern (AOP) model using the ML algorithms, with methods described in a previous paper [13]. We used CMS-LDS data from 2017 to 2019 as a training set to develop the model. This AOP model is designed to automatically detect uncommon or rare associations between specific diseases and surgical procedures, with a particular focus on discrepancies between the diagnosed side (left or right) and the side operated on. The model determined a procedure to be substantiated if the index procedure could be explained by a relevant diagnosis. However, if there was a procedure that could not be explained by any of the diagnoses, then the procedure-diagnosis combination would be viewed as unsubstantiated. All data collection and analysis procedures in this study fully comply with the principles outlined in the "Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research" [14].

Test Set Development

The study used inpatient CMS-LDS data from 2020 as a test set. We limited our testing data to orthopedic procedures, identified by the Attending Physician Specialty Code for Orthopedic Surgery. The study population was selected only if the *ICD-10-PCS* (*International Classification of Diseases, Tenth*

XSL•FO
RenderX

Revision, Procedure Coding System) and *ICD-10-CM* (*International Classification of Diseases, Tenth Revision, Clinical Modification*) codes included identifiable body parts and laterality information. The fourth character of the *ICD-10-PCS* code was used to differentiate body parts and laterality. Cases in which the only procedure performed was the insertion of an infusion device were excluded, as the rationale for the insertion site is typically clinician-determined and often lacks a documented record.

Procedures and diagnoses were categorized by laterality: right, left, or both. "Right-side procedures" indicated that all procedures performed on the patient were on the right side, and "both-side procedures" indicated that the patient received both right- and left-side procedures. The same definitions applied to the laterality of diagnoses. Patients with complete laterality information for both procedures and diagnoses were classified into 9 subgroups based on the combination of procedural and diagnostic laterality, such as right-right, right-left, right-both, left-right, left-left, left-both, both-right, both-left, and both-both. The analysis focused on 2 subgroups with significant discrepancies between procedure and diagnosis laterality: right-side surgery with left-side diagnosis (right-left) and left-side surgery with right-side diagnosis (left-right).

Test Set Evaluation

The accuracy of these 2 subgroups was evaluated through a rigorous review process that examined the alignment between documented procedures and diagnoses. Two authors (YHC and CHL) independently reviewed the consistency between procedures and diagnoses. In cases of disagreement, both authors conducted a thorough secondary review of all available medical records and reached a consensus through discussion. A confirmed error case was defined as a scenario in which no direct association existed between the surgery performed and any recorded diagnosis. Clinical criteria for defining an error case included: (1) all diagnoses did not sufficiently justify the need for the surgery; (2) the anatomical areas mentioned in the diagnoses and procedures were either not the same, not in proximity (eg, shoulder joint and humerus are in proximity), or not appropriately related (eg, upper leg relates to femur bone); and (3) the anatomical areas were similar, but the laterality was opposite. These criteria ensured that discrepancies in laterality between diagnoses and surgical procedures were systematically identified.

Confirmed error cases were further categorized based on whether the procedure and diagnosis involved the same or different body parts. Confirmed nonerror cases were classified into five categories based on potential rationale: (1) same body part between procedure and diagnosis, but with the unspecified side in the diagnosis; (2) same body part between procedure and diagnosis, but with no laterality specified; (3) diagnosis without a specific body part that can explain the index procedure; (4) anatomical areas in the diagnoses and procedures were in proximity; and (5) the index procedure could be explained by a combination of diagnoses and concurrent procedures.

Model Performance Metrics

To assess the generalization and extrapolation performance of the model, we applied the trained AOP model to identify potentially unusual combinations of procedures and diagnoses among the 2 subgroups with significant discrepancies (right-left and left-right) in the test set. We plotted the precision-recall curves to obtain the optimal probability threshold to maximize the predictive capability of the AOP model. Precision-recall curves have been considered an effective metric for accessing the model, especially the data is an unbalanced dataset [15,16]. The optimal probability threshold is where a point can achieve high precision score while only sacrificing minimal recall (Figure S1 in Multimedia Appendix 1). By calculating the accurate prediction rate of the AOP model and the rule-based method (right-left and left-right), we can compare the performance between these 2 methods.

Ethical Considerations

This study used the CMS-LDS files, which provide beneficiary-level information for research aimed at improving the US health care system. Access to the data was granted through the standard Centers for Medicare and Medicaid Services (CMS) approval process, including entering into a Data Use Agreement with CMS. The study was exempt from human subject research ethics review, as the data used were deidentified in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. Since the study involved secondary analyses of deidentified data, informed consent was not required, and no direct interaction with human participants occurred.

To ensure privacy and confidentiality, all analyses strictly adhered to CMS policies, including the cell suppression rule, which prohibits the publication of results with cell sizes of 10 or fewer (eg, patients or services). Furthermore, no attempts were made to reidentify individuals from the data, maintaining compliance with ethical and privacy standards.

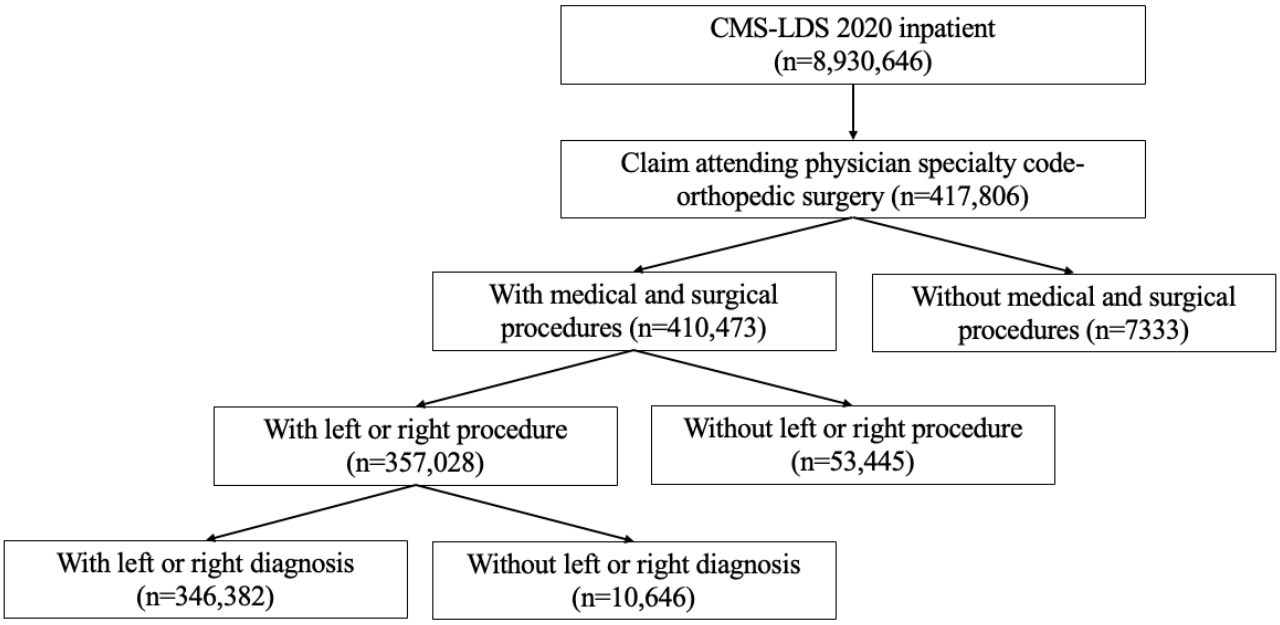
Results

Baseline Study Population

Data from a total of 346,382 claims were included in the study. Among them, the data indicated that 1106 patients had a left-side procedure with right-side diagnoses, and 1064 patients had a right-side procedure with left-side diagnoses (Figure 1).

XSL•FO
RenderX

Figure 1. Selection of test set study population from Centers for Medicare and Medicaid Services Limited Data Set (CMS-LDS).



Subgroups based on the combination of procedural and diagnostic laterality.

Laterality	Left-side procedures	Right-side procedures	Both-side procedures
Left-side diagnoses	109,545	1064	1075
Right-side diagnoses	1106	128,101	871
Both-side diagnoses	48,147	50,013	6420

The study used inpatient CMS-LDS data from 2020 as the test set. Testing data was restricted to orthopedic procedures, and identified using the Attending Physician Specialty Code for Orthopedic Surgery. The study population included only cases where the *ICD-10* (*International Statistical Classification of Diseases, Tenth Revision*) procedure and diagnosis codes provided clear information about the body part and laterality.

Test Set Population Characteristics

Among the 1106 patients with left-side procedures with right-side diagnoses, 603/1106 (54.5%) cases were confirmed as errors, while 503/1106 (45.5%) were deemed reasonable combinations after applying the clinical review criteria. Similarly, among the 1064 patients with right-side procedures with left-side diagnoses, 541/1064 (50.8%) cases were confirmed as errors, while 523/1064 (49.2%) were deemed reasonable combinations (Table 1). Most confirmed errors involved the same body part in both procedure and diagnosis but with opposite laterality. Most nonerror cases were considered reasonable due to the presence of a relatively general diagnosis code with an unspecified body part, or a diagnosis with a specified body part but no laterality.

XSL•FO
RenderX

Table 1. Performance of rule-based method and association outlier pattern (AOP) model by laterality.

	Rule-based method[a]	AOP[b,c] model
Left-side procedure with right-side diagnoses		
Confirmed error	n=603 (54.5%)	n=576 (82.6%)
Same body part	n=421	n=404
Not same body part	n=182	n=172
Confirmed nonerror	n=503 (45.5%)	n=121 (17.4%)
Unspecified side of diagnosis[d]	n=39	n=17
No laterality for diagnosis[e]	n=131	n=34
No specific body part nor laterality[f]	n=251	n=58
Proximity area[g]	n=4	n=2
Others[h]	n=78	n=10
Right-side procedure with left-side diagnoses		
Confirmed error	n=541 (50.8%)	n=524 (80.0%)
Same body part	n=390	n=376
Not same body part	n=151	n=148
Confirmed nonerror	n=523 (49.2%)	n=131 (20.0%)
Unspecified side of diagnosis	n=50	n=17
No laterality for diagnosis	n=271	n=75
No specific body part nor laterality	n=129	n=25
Proximity area	n=4	n=4
Others	n=69	n=10

[a]Left-side procedure with right-side diagnoses: n=1106; right-side procedure with left-side diagnoses: n=1064.

[b]Left-side procedure with right-side diagnoses: n=697; right-side procedure with left-side diagnoses: n=655.

[c]AOP: association outlier pattern.

[d]Same body part between procedure and diagnosis, but with the unspecified side in the diagnosis.

[e]Same body part between procedure and diagnosis, but with no laterality specified.

[f]Diagnosis without a specific body part that can explain the index procedure.

[g]Anatomical areas in the diagnoses and procedures were in proximity.

[h]The index procedure could be explained by a combination of diagnoses and concurrent procedures.

This table summarizes the performance of the rule-based method and the AOP machine learning model in identifying surgical procedure-diagnosis inconsistencies in the CMS-LDS from 2020. The study focused on patients undergoing orthopedic procedures with discrepancies in laterality between procedures and diagnoses. Results are presented for 2 subgroups: left-side procedures with right-side diagnoses (n=1106) and right-side procedures with left-side diagnoses (n=1064). Clinical review outcomes are categorized into confirmed errors and nonerrors, with further classification based on the nature of the diagnosis and its alignment with the index procedure.

We focused on 2 subgroups with significant discrepancies regarding laterality between procedure and diagnosis, where combinations could be reasonable or unreasonable. For instance, a patient who underwent right knee joint arthroplasty but was diagnosed only with left knee osteoarthritis was deemed to have an error, as no diagnoses could justify the need for right knee joint replacement. Conversely, a nonerror case might involve a patient who underwent right lower leg detachment surgery with a diagnosis of diabetes-related gangrene (Table 2). Although the diagnosis did not specify the body part or laterality, it could explain the reason for the surgery and did not conflict with other diagnoses.

XSL•FO
RenderX

Table 2. Test set evaluation examples for clinical review of procedure-diagnosis discrepancies.

	Index procedures	Diagnoses
Confirmed error		
Same body part	• Right knee joint • 0SRC0J9-Replacement of right knee joint with synthetic substitute, cemented, open approach	• M1712-Unilateral primary osteoarthritis, left knee • F17210-Nicotine dependence, cigarettes, uncomplicated
Not same body part	• Left upper arm tendon, left shoulder joint • 0LS40ZZ-Reposition left upper arm tendon, open approach • 0RRK0JZ-Replacement of left shoulder joint with synthetic substitute, open approach	• M19021-Primary osteoarthritis and right elbow • E785-Hyperlipidemia, unspecified • I10-Essential (primary) hypertension • E7800-Pure hypercholesterolemia, unspecified
Confirmed non-error		
Unspecified body part and unspecified laterality	• Left hip joint • 0SPB0JZ-Removal of synthetic substitute from left hip joint, open approach • 0SRB049-Replacement of left hip joint with ceramic on polyethylene synthetic substitute, cemented, open approach	• T8484XA-Pain due to internal orthopedic prosthetic devices, implants and grafts, initial encounter • Y792-Prosthetic and other implants, materials and accessory orthopedic devices associated with adverse incidents • S0501XA-Injury of conjunctiva and corneal abrasion without foreign body, right eye, initial encounter • H10501-Unspecified blepharoconjunctivitis and right eye
Unspecified body part and unspecified laterality	• Right lower leg • 0Y6H0Z2-Detachment at right lower leg, mid, open approach	• E1152-Type 2 diabetes mellitus with diabetic peripheral angiopathy with gangrene • I96-Gangrene, not elsewhere classified • N186-End stage renal disease • I69354-Hemiplegia and hemiparesis following cerebral infarction affecting left nondominant side • E1122-Type 2 diabetes mellitus with diabetic chronic kidney disease • E785-Hyperlipidemia, unspecified • I252-Old myocardial infarction • Z85528-Personal history of other malignant neoplasm of kidney
Unspecified body part and unspecified laterality	• Right femoral shaft and right upper femur • 0QB80ZX-Excision of right femoral shaft, open approach, and diagnostic • 0QH606Z-Insertion of intramedullary internal fixation device into right upper femur, open approach	• C7951-Secondary malignant neoplasm of bone • C3402-Malignant neoplasm of left main bronchus • R200-Anesthesia of skin • R591-Generalized enlarged lymph nodes
No laterality for diagnosis	• Left knee • 0SPD08Z-Removal of spacer from left knee joint, open approach • 0SRD0J9-Replacement of left knee joint with synthetic substitute, cemented, open approach	• Z4733-Aftercare following explanation of knee joint prosthesis • E785-Hyperlipidemia, unspecified • I4510-Unspecified right bundle-branch block • I10-Essential (primary) hypertension

This table provides representative examples of confirmed errors and nonerrors from the test set of orthopedic procedures in the CMS-LDS data from 2020. Confirmed errors include cases where procedural laterality (eg, left or right) did not align with the diagnosis, while nonerrors involve reasonable justifications for the procedure based on provided diagnoses. Each example lists the index procedures, associated diagnoses, and rationale for classification based on clinical review.

AOP Model Performance Evaluation

Among the 1106 patients with left-side procedures with right-side diagnoses, the AOP model identified 697 potentially unusual combinations. After applying the clinical review criteria, 576 out of 697 (82.6%) cases were confirmed as errors, while 121 out of 697 (17.4%) were deemed reasonable. Similarly, among the 1064 patients with right-side procedures with left-side diagnoses, the AOP model identified 655 potentially unusual combinations. Of these, 524 out of 655 (80%) cases were confirmed as errors, and 131 out of 655 (20%) were reasonable combinations (Table 1). Most confirmed errors involved the

XSL•FO
RenderX

same body part in both procedure and diagnosis but with opposite laterality. Most nonerror cases were considered reasonable due to a general diagnosis with an unspecified body part or a diagnosis with a specified body part but no laterality.

An example of a confirmed error among AOP model-identified cases was a patient who received right shoulder joint arthroplasty, while the diagnoses only included a left humerus fracture. Since a left arm fracture could not justify right shoulder joint replacement, this case was deemed an error. Conversely, an example of a nonerror case was a patient who underwent left glenoid cavity replacement, with a diagnosis of unspecified side rotator cuff tear or rupture (Table 3). As an unspecified side shoulder injury could justify left shoulder surgery, this case was deemed a nonerror.

Table 3. Association outlier pattern (AOP) model evaluation examples for identifying unusual procedure-diagnosis combinations.

	Index procedures	Diagnoses
Confirmed error		
Same body part	• Left hip • 0SRB02A-Replacement of Left Hip Joint with Metal on Polyethylene Synthetic Substitute, Uncemented, Open Approach	• M1611 - Unilateral primary osteoarthritis, right hip • I10-Essential (primary) hypertension • F329-Major depressive disorder, single episode, unspecified
Not same body part	• Right shoulder joint • 0RRJ0J6-Replacement of right shoulder joint with synthetic substitute, humeral surface, open approach	• S42242A-4-part fracture of surgical neck of left humerus, initial encounter for closed fracture • Z880-Allergy status to penicillin • Z79899-Other long term (current) drug therapy
Confirmed non-error		
Unspecified body part and unspecified laterality	• Right hip bursa • 0MBL0ZZ-Excision of right hip bursa and ligament, open approach	• T8131XA-Disruption of external operation (surgical) wound, not elsewhere classified, initial encounter • G40909-Epilepsy, unspecified, not intractable, and without status epilepticus • I110-Hypertensive heart disease with heart failure • E669-Obesity, unspecified • G2581-Restless legs syndrome • Z96642-Presence of left artificial hip joint • B964-Proteus (mirabilis) (morganii) as the cause of diseases classified elsewhere • T8141XA-Infection following a procedure, superficial incisional surgical site, initial encounter • I428-Other cardiomyopathies • Y838-Other surgical procedures as the cause of abnormal reaction of the patient, or of later complication, without mention of misadventure at the time of the procedure
No laterality for diagnosis	• Left glenoid cavity • 0PR80JZ-Replacement of left glenoid cavity with synthetic substitute, open approach	• M75120-Complete rotator cuff tear or rupture of unspecified shoulder, not specified as traumatic • G8918-Other acute postprocedural pain • M109-Gout and unspecified • Z89511-Acquired absence of right leg below the knee

This table illustrates specific examples of confirmed errors and nonerrors identified by the AOP machine learning model in the CMS-LDS 2020 test set. Confirmed errors are characterized by laterality discrepancies between the procedure and diagnosis, while nonerrors reflect plausible procedure-diagnosis relationships despite general or unspecified diagnoses. Each case includes the index procedures, associated diagnoses, and the rationale for classification.

Discussion

Principal Findings

In this study, we evaluated the effectiveness of the AOP model in detecting whether surgical procedures are substantiated with diagnoses records using CMS-LDS data. We found the AOP model is more accurate in identifying inconsistencies in surgical procedure documentation when compared with the traditional rule-based method. Meanwhile, the adapted AOP model also showed the transferability from detecting the medication-disease association and expanding to verifying the procedure-diagnosis association.

Our study adds to the existing literature that ML has penetrated the medical field with great success. ML allows for more detail to be mined from the data, allowing for the development of better diagnostic and prognostic tools than traditional approaches [17,18]. Similarly, the AOP model can detect associations between specific diseases and surgical procedures by learning patterns from a large database, while the rule-based method depends on the laterality of procedures and diagnoses. For example, among the 2 subgroups with significant discrepancies

XSL•FO
RenderX

regarding laterality between procedure and diagnosis (left-right and right-left), a rule-based method classified all 2170 cases as inappropriate combinations due to the conflict in laterality. In contrast, the AOP model classified only 1352 of these cases as inappropriate. In terms of identifying inappropriate procedure-diagnosis combinations, the AOP model identified most errors similarly to the rule-based method (1100 for AOP vs 1144 for rule-based). In addition, the AOP performed with much higher accuracy (1100/1352, 81.4%) compared with the rule-based method (1144/2170, 52.7%).

The AOP model outperformed the traditional rule-based model, demonstrating a lower rate of false signals. Comparing the confirmed nonerror percentage between the rule-based method and the AOP model, there was a significantly lower rate of confirmed nonerror cases identified by the AOP model. In these cases, most confirmed nonerrors identified by the rule-based method could be explained by other diagnoses without mentioning body part or laterality. For example, cancer metastasis to bone may explain the reason for pelvic surgery, even without a specific pelvic metastasis diagnosis. This is because some diagnoses may not provide detailed information about the patient's condition. Similarly, a patient who underwent lower leg detachment surgery with a diagnosis of diabetes-related gangrene could have the surgery justified, even if the gangrene diagnosis did not specify the body part. The rule-based method focuses solely on the laterality of the diagnosis, whereas the AOP model uses ML methods to explain the association between procedures and diagnoses. Therefore, when the association is deemed normal, the AOP model does not label them as inappropriate.

This study also showed several noteworthy findings. First, cases with fewer diagnoses were more likely to be labeled as inappropriate procedure-diagnosis combinations. For example, a patient who underwent surgeries for the left upper arm and left shoulder had only 4 diagnoses, including a right elbow problem and other chronic diseases. This may indicate incomplete diagnosis coding or an error in the surgical site. Second, the AOP model effectively identified unusual procedure-diagnosis combinations within subgroups where traditional rule-based methods failed to do so. Specifically, procedures that aligned with their corresponding diagnoses were deemed appropriate by conventional methods, such as left-side procedures with left-side diagnoses (left-left) and right-side procedures with right-side diagnoses (right-right). However, the AOP model continued to uncover atypical combinations within these groups. For instance, the model identified a patient who underwent left hip arthroplasty despite having a diagnosis of left knee osteoarthritis only.

The strengths of our study include its being the first to explore whether a ML model could play a surveillance role in detecting wrong site surgery. The effectiveness and fast pace to review the inconsistencies of surgical documentation help increase reporting mistakes in a timely manner without putting too much workforce on this issue. The findings from our study raise important questions regarding whether these errors stem from documentation mistakes or represent actual incidents involving patients. Therefore, in addition to retrospective reviews, it is essential to leverage modern technology to facilitate real-time error prevention. The potential of our model could add value on the current technology to encompass proactive prediction and analysis of diagnoses and clinical evidence to inform accurate surgical decisions and improve documentation practices [19]. In addition, the potential adaptability of the AOP model extends beyond orthopedic surgery to other specialties where procedural and diagnostic codes require logical alignment. For example, in ophthalmic surgery, the model could identify mismatches such as a diagnosis of left-sided cataract paired with a procedure for right-eye cataract surgery. Similarly, otolaryngology and other surgical fields with specific laterality could benefit. However, certain specialties may pose unique challenges due to differences in coding systems or the absence of laterality indicators, requiring further validation and specialty-specific refinements.

The AOP model has limitations. There were still nonerror cases identified by the AOP model. For instance, a patient who underwent left glenoid cavity replacement surgery with a diagnosis of unspecified rotator cuff tear was marked as inappropriate by the AOP model. However, the diagnosis could explain the necessity of the surgery. The reason the AOP model labeled this as an inappropriate combination may be due to the low frequency of this procedure-diagnosis combination in the CMS dataset. If the training dataset has a limited number of surgeries or diagnoses, then the association may be viewed as low and therefore marked as an inappropriate combination.

Another limitation is that the AOP model failed to identify some error cases labeled by the rule-based method. For example, a patient received an internal fixation surgery of the right upper femur with diagnoses of secondary malignant neoplasm of bone and left femur osteonecrosis. While the diagnosis of secondary malignant neoplasm could justify the surgical necessity, the diagnosis of left femur osteonecrosis conflicted with the right femur surgery. In such cases, the AOP model uses the maximum association within all procedure-diagnosis combinations, meaning that if the association between the surgery and a diagnosis is high, the model will view the combination as appropriate, even if the other diagnosis is in conflict. Therefore, in cases where one diagnosis can justify surgery, but another is conflicting, the AOP model may not accurately identify the combination. However, this aspect of the AOP model presents an opportunity for optimization, and future adjustments incorporating more detailed information through advanced ML techniques could yield a more accurate model, especially in cases where diagnoses are vague, incomplete, or lack clear specifications of body parts or sides.

Building on these strengths and limitations, future steps for implementing the AOP model in electronic health record systems could proceed along 2 directions. First, the model could support retrospective reviews by efficiently identifying WSS events, reducing review times and workforce burden, and enabling more timely reporting. Second, it could facilitate real-time consistency checks when clinicians input procedure codes. For example, if a surgeon enters a right-sided knee arthroplasty for a patient diagnosed with left knee osteoarthritis, the AOP model could flag the discrepancy and issue an alert. Such real-time notifications may prevent WSS before it occurs. However, to achieve successful implementation, the model must

XSL•FO
RenderX

be optimized to address the limitations, and its performance must be validated across diverse electronic health record systems and datasets across different facilities.

Conclusion

Overall, the AOP model enhances patient safety by systematically analyzing the cooccurrence frequency of diseases and surgical procedures on opposite sides, thereby identifying potentially unusual procedure-diagnosis combinations. Our analysis demonstrated a significant improvement in accuracy rates when the AOP model was used, compared with scenarios where it was not applied. By improving precision in identifying and validating cases with discrepancies between the diagnosed and operated sides, the AOP model has the potential to reduce the risk of errors in orthopedic treatment planning. This enhanced accuracy enables health care providers to make more informed decisions, reducing the risk of WSS and ultimately contributing to improved patient outcomes in orthopedic health care settings.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors.

Generative artificial intelligence was not used in any portion of the manuscript writing.

Data Availability

This study used Centers for Medicare and Medicaid Services Limited Data Set (CMS-LDS) files. Access to these data is restricted and governed by CMS. Researchers interested in accessing the data must follow CMS's application process and enter into a Data Use Agreement. Due to these restrictions, the data cannot be deposited in publicly available repositories. Requests for data access should be directed to CMS.

Authors' Contributions

YHC, CHL, and YPK were contributed to conceptualization. YHC, CHF, AJL, and JS were handled methodology. YHC and CHL were writing–original draft preparation, formal analysis, and visualization. CHF and AJL were performed data curation. JS, YPK, and UI were writing–review and editing. YCJL managed supervision

Conflicts of Interest

JS, AJL, and Y-CJL are significant shareholders in, and continue to work with, AESOP Technology Inc, the company that developed and owns IP related to the AOP model in the study. C-HF is an employee at AESOP Technology Inc.

Multimedia Appendix 1
Supplement Figure 1. Precision-Recall curve of Association Outlier Pattern (AOP) Model in Test Set.
[PNG File , 87 KB - formative_v9i1e68436_app1.png]

References

1. Robinson TP, Bilimoria KY, Yang AD. Understanding a surgeon's worst nightmare: wrong site surgery. Jt Comm J Qual Patient Saf 2023;49(5):237-238. [doi: 10.1016/j.jcjq.2023.03.006] [Medline: 36990955]
2. Tan J, Ross JM, Wright D, Pimentel MPT, Urman RD. A contemporary analysis of closed claims related to wrong-site surgery. Jt Comm J Qual Patient Saf 2023;49(5):265-273. [doi: 10.1016/j.jcjq.2023.02.002] [Medline: 36925434]
3. Zil-E-Ali A, Laubscher L, Kourampi I, Tsagkaris C. Is surgery on the right track? The burden of wrong-site surgery. Proc (Bayl Univ Med Cent) 2023;36(5):657-660 [FREE Full text] [doi: 10.1080/08998280.2023.2231714] [Medline: 37614864]
4. Mehtsun WT, Ibrahim AM, Diener-West M, Pronovost PJ, Makary MA. Surgical never events in the United States. Surgery 2013;153(4):465-472. [doi: 10.1016/j.surg.2012.10.005] [Medline: 23257079]
5. Neily J, Mills PD, Eldridge N, Dunn EJ, Samples C, Turner JR, et al. Incorrect surgical procedures within and outside of the operating room. Arch Surg 2009;144(11):1028-1034. [doi: 10.1001/archsurg.2009.126] [Medline: 19917939]
6. Global patient safety report 2024, Geneva: World Health Organization. URL: https://iris.who.int/bitstream/handle/10665/376928/9789240095458-eng.pdf?sequence=1 [accessed 2024-05-30]
7. Chin YPH, Song W, Lien CE, Yoon CH, Wang WC, Liu J, et al. Assessing the international transferability of a machine learning model for detecting medication error in the general internal medicine clinic: multicenter preliminary validation study. JMIR Med Inform 2021;9(1):e23454 [FREE Full text] [doi: 10.2196/23454] [Medline: 33502331]
8. Huang CY, Nguyen PA, Yang HC, Islam MM, Liang CW, Lee FP, et al. A probabilistic model for reducing medication errors: a sensitivity analysis using electronic health records data. Comput Methods Programs Biomed 2019;170:31-38. [doi: 10.1016/j.cmpb.2018.12.033] [Medline: 30712602]
9. Limited Data Set (LDS) Files | CMS. URL: https://www.cms.gov/data-research/files-for-order/limited-data-set-lds-files [accessed 2025-10-25]

XSL•FO
RenderX

10. Hakkarainen TW, Arbabi S, Willis MM, Davidson GH, Flum DR. Outcomes of patients discharged to skilled nursing facilities after acute care hospitalizations. Ann Surg 2016;263(2):280-285 [FREE Full text] [doi: 10.1097/SLA.0000000000001367] [Medline: 26445466]
11. Avila A, Sato EH, Acuña AJ, Vasavada K, Kamath AF. Gender differences in reimbursement among orthopaedic surgeons: a cross-sectional analysis of medicare claims. J Am Acad Orthop Surg 2023;31(15):e570-e578. [doi: 10.5435/JAAOS-D-22-00823] [Medline: 37253194]
12. Adashi EY, Gruppuso PA. The centers for medicare and medicaid services (CMS) COVID-19 brief: unsettling racial and ethnic health disparities. J Am Board Fam Med 2021;34(Supplement):S13-S15. [doi: 10.3122/jabfm.2021.s1.200450]
13. Nguyen PA, Syed-Abdul S, Iqbal U, Hsu MH, Huang CL, Li HC, et al. A probabilistic model for reducing medication errors. PLoS One 2013;8(12):e82401 [FREE Full text] [doi: 10.1371/journal.pone.0082401] [Medline: 24312659]
14. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016;18(12):e323 [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]
15. Muschelli J. ROC and AUC with a binary predictor: a potentially misleading metric. J Classif 2020;37(3):696-708 [FREE Full text] [doi: 10.1007/s00357-019-09345-1] [Medline: 33250548]
16. Zhou QM, Zhe L, Brooke RJ, Hudson MM, Yuan Y. A relationship between the incremental values of area under the ROC curve and of area under the precision-recall curve. Diagn Progn Res 2021;5(1):13 [FREE Full text] [doi: 10.1186/s41512-021-00102-w] [Medline: 34261544]
17. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 2018;19(6):1236-1246 [FREE Full text] [doi: 10.1093/bib/bbx044] [Medline: 28481991]
18. Wang K, Shi Q, Sun C, Liu W, Yau V, Xu C, et al. Corrigendum: a machine learning model for visualization and dynamic clinical prediction of stroke recurrence in acute ischemic stroke patients: a real-world retrospective study. Front Neurosci 2023;17:1235340 [FREE Full text] [doi: 10.3389/fnins.2023.1235340] [Medline: 37492408]
19. Gloystein DM, Heiges BA, Schwartz DG, DeVine JG, Spratt D. Innovative technology system to prevent wrong site surgery and capture near misses: a multi-center review of 487 cases. Front Surg 2020;7:563337 [FREE Full text] [doi: 10.3389/fsurg.2020.563337] [Medline: 33195386]

Abbreviations

AOP: association Outlier Pattern
CMS: Centers for Medicare and Medicaid Services
CMS-LDS: Centers for Medicare and Medicaid Services Limited Data Set
HIPAA: Health Insurance Portability and Accountability Act
ICD-10: International Statistical Classification of Diseases, Tenth Revision
ICD-10-CM: International Classification of Diseases, Tenth Revision, Clinical Modification
ICD-10–PCS: International Classification of Diseases, Tenth Revision, Procedure Coding System
ML: machine learning
WSS: Wrong-site surgery

Edited by A Mavragani; submitted 06.11.24; peer-reviewed by S Bharati; comments to author 06.12.24; revised version received 20.12.24; accepted 12.01.25; published 13.02.25.

Please cite as:
Chen YH, Lin CH, Fan CH, Long AJ, Scholl J, Kao YP, Iqbal U, Li YCJ
Machine Learning Approach to Identifying Wrong-Site Surgeries Using Centers for Medicare and Medicaid Services Dataset: Development and Validation Study
JMIR Form Res 2025;9:e68436
URL: https://formative.jmir.org/2025/1/e68436
doi:10.2196/68436
PMID:

©Yuan-Hsin Chen, Ching-Hsuan Lin, Chiao-Hsin Fan, An Jim Long, Jeremiah Scholl, Yen-Pin Kao, Usman Iqbal, Yu-Chuan Jack Li. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring the Acceptance of Just-in-Time Adaptive Lifestyle Support for People With Type 2 Diabetes: Qualitative Acceptability Study

Eclaire A G Hietbrink[1,2], RN, MSc; Anouk Middelweerd[1], PhD; Wendy d'Hollosy[1], PhD; Laura K Schrijver[1,2], MSc; Gozewijn D Laverman[1,2], MD, PhD; Miriam M R Vollenbroek-Hutten[1,3], PhD

[1]Department of Biomedical Signals and Systems, University of Twente, Enschede, The Netherlands

[2]Department of Internal Medicine/Nephrology, Ziekenhuisgroep Twente, Almelo, The Netherlands

[3]Medisch Spectrum Twente, Enschede, The Netherlands

Corresponding Author:
Eclaire A G Hietbrink, RN, MSc
Department of Biomedical Signals and Systems
University of Twente
Drienerlolaan 5
Enschede, 7500 AE
The Netherlands
Phone: 31 53 4892760
Email: e.a.g.hietbrink@utwente.nl

Abstract

Background: The management of type 2 diabetes (T2D) requires individuals to adopt and maintain a healthy lifestyle. Personalized eHealth interventions can help individuals change their lifestyle behavior. Specifically, just-in-time adaptive interventions (JITAIs) offer a promising approach to provide tailored support to encourage healthy behaviors. Low-effort self-reporting via ecological momentary assessment (EMA) can provide insights into individuals' experiences and environmental factors and thus improve JITAI support, particularly for conditions that cannot be measured by sensors. We developed an EMA-driven JITAI to offer tailored support for various personal and environmental factors influencing healthy behavior in individuals with T2D.

Objective: This study aimed to assess the acceptability of EMA-driven, just-in-time adaptive lifestyle support in individuals with T2D.

Methods: In total, 8 individuals with T2D used the JITAI for 2 weeks. Participants completed daily EMAs about their activity, location, mood, overall condition, weather, and cravings and received tailored support via SMS text messaging. The acceptability of the JITAI was assessed through telephone-conducted, semistructured interviews. Interview topics included the acceptability of the EMA content and prompts, the intervention options, and the overall use of the JITAI. Data were analyzed using a hybrid approach of thematic analysis.

Results: Participants with a mean age of 70.5 (SD 9) years, BMI of 32.1 (SD 5.3) kg/m², and T2D duration of 15.6 (SD 7.7) years had high self-efficacy scores in physical activity (ie, 32) and nutrition (ie, 29) and were mainly initiating or maintaining behavior changes. The identified themes were related to the intervention design, decision points, tailoring variables, intervention options, and mechanisms underlying adherence and retention. Participants provided positive feedback on several aspects of the JITAI, such as the motivating and enjoyable messages that appeared well tailored to some individuals. However, there were notable differences in individual experiences with the JITAI, particularly regarding intervention intensity and the perceived personalization of the EMA and messages. The EMA was perceived as easy to use and low in burden, but participants felt it provided too much of a snapshot and too little context, reducing the perceived tailoring of the intervention options. Challenges with the timing and frequency of prompts and the relevance of some tailoring variables were also observed. While some participants found the support relevant and motivating, others were less inclined to follow the advice. Participants expressed the need for even more personalized support tailored to their specific characteristics and circumstances.

Conclusions: This study showed that an EMA-driven JITAI can provide motivating and tailored support, but more personalization is needed to ensure that the lifestyle support more closely fits each individual's unique needs. Key areas for improvement include developing more individually tailored interventions, improving assessment methods to balance active and passive data collection, and integrating JITAIs within comprehensive lifestyle interventions.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e65026) doi:10.2196/65026

KEYWORDS

eHealth; just-in-time adaptive intervention; ecological momentary assessment; type 2 diabetes; behavior change; physical activity; nutrition; acceptability; formative evaluation; mobile phone

Introduction

Background

A healthy lifestyle has a pivotal role in effective management of type 2 diabetes (T2D) [1,2]. This approach includes addressing important contributors to T2D, such as insufficient physical activity, unhealthy dietary habits, inadequate sleep, and elevated stress levels. Substantial scientific evidence supports the substantial health benefits of lifestyle interventions [3-8], including studies indicating that lifestyle modifications can potentially reverse or achieve remission in T2D. In addition, a 10-year study demonstrated that lifestyle interventions could result in economic benefits in the form of reduced health care costs [9].

However, individuals with T2D often struggle to adhere to lifestyle guidelines [10-13]. One explanation is that social and environmental changes, such as the prevalence of sedentary jobs and the widespread availability of highly processed foods, have made it increasingly challenging for individuals to adopt and maintain a healthy lifestyle [14]. Moreover, managing a chronic condition poses additional challenges, including the time and effort required to manage symptoms, adhere to treatment plans, and cope with physical and psychosocial consequences [15,16]. Individuals need knowledge and self-management skills to implement sustainable lifestyle changes [2]. Due to variations in personal, environmental, and disease-related factors, self-management skills can differ substantially from person to person [17]. Therefore, a personalized approach is essential to guide each individual toward a healthier lifestyle, providing tailored support that meets their unique needs.

eHealth, which involves using technology to promote health, well-being, and health care [18], holds considerable potential in delivering personalized lifestyle support. Research has demonstrated that eHealth interventions can be both acceptable and effective for improving physical activity, diet, and health outcomes [19-23]. A promising approach to provide personalized support is just-in-time adaptive interventions (JITAIs). JITAIs are designed to dynamically provide the right support at the right time [24]. A key element of JITAIs is the need for insights into behavior and context to adapt the support accordingly. Today, wearables, smartphone sensors, and remote monitoring devices, such as activity trackers, GPS, and smartphone calendars, can be used for real-time monitoring and to inform intervention decisions [24,25]. In addition, low-effort self-reporting via ecological momentary assessment (EMA) can provide further insights into an individual's situation and emotional state, which may improve the support offered by JITAIs [26]. EMA involves repeated sampling of an individual's behaviors, cognitions, emotions, and environmental factors in their natural environment through short questionnaires that are prompted multiple times within a certain time frame [27]. EMAs allow for adapting support to individual needs, particularly for conditions that cannot be measured by sensors. Unlike typical eHealth interventions, JITAIs can provide near–real-time and contextually relevant support, tailored to an individual's needs and preferences [24].

JITAIs have been studied to determine their added value in supporting several lifestyle behaviors, such as physical activity, diet, and substance use. Studies have examined their acceptability; feasibility; and, to a lesser extent, effectiveness [28-32]. Overall, JITAIs appear to be acceptable, although some feasibility challenges were identified, including issues with smartphone battery life, sensor reliability, the punctuality of JITAI messages, and the incomplete automation of JITAI delivery [28,31]. People were more likely to follow JITAI messages and to find them more relevant than randomly selected messages [33,34]. Evidence on the effectiveness of JITAIs for behavior change is mixed, and there is insufficient data to draw well-founded conclusions [28,29,32]. JITAIs that use algorithms to tailor interventions based on past behavior and current needs, along with guidance from a human, are more effective [30].

Objectives

Our research team previously developed the E-Supporter, an evidence-based and tailored digital coach for people with chronic conditions, primarily T2D [35]. A previous feasibility study on the E-Supporter revealed that individuals with T2D need more tailored support to better align with their personal circumstances [36]. This finding highlights the need for interventions that can adapt to individual needs. JITAIs offer the opportunity to provide personalized, contextually relevant support when it is needed. In many JITAIs, tailoring is often based on objective parameters related to healthy lifestyles, such as wearable data [37-39]. However, these eHealth interventions do not frequently address the more nuanced aspects of individuals' experiences, such as disease-related challenges or environmental factors [28]. While a few JITAIs assess these conditions via EMA [26,32], none currently tailor support to the unique characteristics of individuals with T2D. Therefore, we developed an EMA-driven JITAI designed to offer tailored assistance for several personal and environmental factors that influence healthy behavior. Given its novelty, assessing acceptability is crucial for refinements and future development [18,40]. Hence, we conducted an acceptability assessment for the EMA-driven JITAI designed to support a healthy lifestyle for individuals with T2D.

Methods

Study Design

We performed a qualitative study to assess the acceptability of the JITAI over 2 weeks among individuals with T2D. The JITAI was designed to support physical activity and healthy nutrition in individuals with T2D. The study was conducted during a

XSL•FO
RenderX

national lockdown due to COVID-19, requiring all contact with participants to take place remotely.

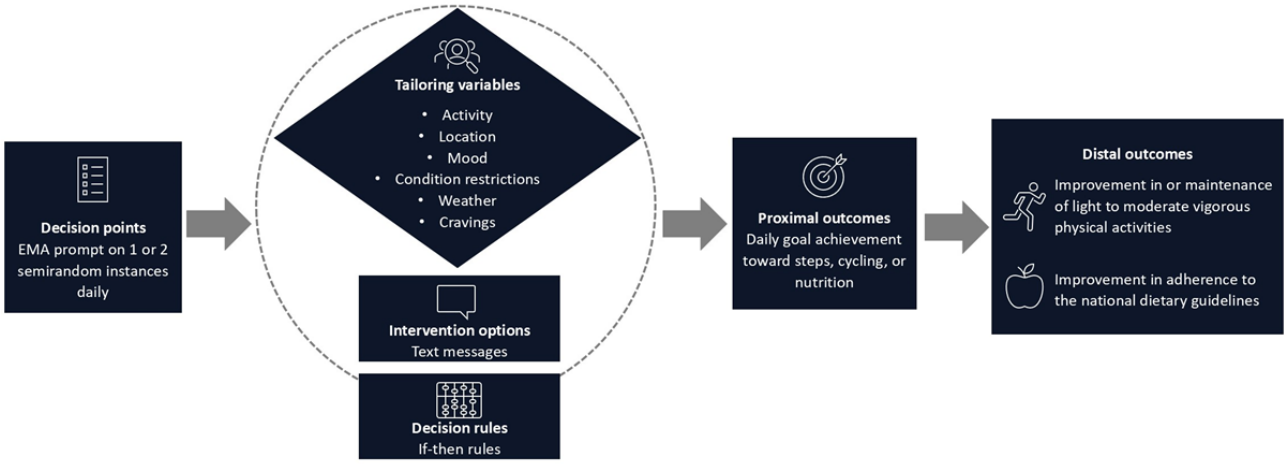
Intervention Description

Overview

The foundation of the developed JITAI is rooted in the E-Supporter 1.0 [35], designed to promote physical activity and a healthy diet among individuals with chronic diseases, initially focusing on T2D. The E-Supporter 1.0 was participatory developed with input from both patients with T2D and health care professionals. It is tailored to several variables, including behavioral objectives, the phase of behavior change, type of chronic disease, time of day, and goal achievement. The behavior change strategies are based on the *Health Action Process Approach* [41] and theories to support behavior maintenance [42,43]. An in-depth description of the E-Supporter 1.0 can be found elsewhere [35].

We used a systematic approach to design the JITAI according to the principles outlined in the framework by Nahum-Shani et al [24]. This process involved developing the 6 components of a JITAI: distal outcome, proximal outcome, tailoring variables, intervention options, decision rules, and decision points (Figure 1 [24]). The development process involved iterative feedback on all components from a group of multidisciplinary experts, including a health scientist, sports scientist, computer scientist, and diabetologist.

Figure 1. Overview of the intervention components based on the conceptual model of just-in-time adaptive interventions. EMA: ecological momentary assessment.



Distal Outcome

The distal outcome represents *the ultimate long-term goal* of the intervention, indicating the desired outcome that is expected to be achieved [24]. The distal outcome regarding the physical activity module is improvement in, or maintenance of, light to moderate vigorous physical activities in people with T2D. Regarding the nutritional module, the distal outcome is improvement in adherence to the national dietary guidelines [44]. A substantiation for the choice of the distal outcomes can be found in the study by Hietbrink et al [35].

Proximal Outcome

The proximal outcome represents *the short-term goal* that the intervention aims to achieve and indicates the progress to the distal outcome [24]. For this JITAI, we opted for a proximal outcome centered on actual behavior. The E-Supporter 1.0 allows users to set specific goals for steps, cycling, or nutrition, contributing to the attainment of distal outcomes [35]. Consequently, we focused on *daily goal achievement* as our proximal outcome, which is considered a meaningful indicator of short-term behavioral progress toward the distal outcome [24].

Tailoring Variables

Tailoring variables refer to individual-specific information used for personalization, determining when and how to intervene [24]. To identify tailoring variables, we conducted a survey assessing facilitators and barriers for a healthy lifestyle among individuals with a lifestyle-related chronic condition (L K Schrijver, MSc, unpublished data, March 2020). In the domain of physical activity, important dynamic factors included weather, mood, daily activities, fatigue, disease symptoms, pain, and social support. Regarding dietary choices, mood, location (eg, eating out in a restaurant), social occasions, and the presence of unhealthy foods emerged as dynamic factors. The research group clustered these factors into the tailoring variables for the JITAI. The following 6 tailoring variables were identified: activity, location, mood, physical or conditional restrictions, weather, and cravings. In this study, we used EMA as an active assessment method (ie, input from the user is required) to evaluate all tailoring variables. Each variable was addressed through a specific EMA question. The research team designed, evaluated, and refined the EMA questions and corresponding response options. Table 1 presents an overview of the tailored EMA questions designed to assess the identified variables.

XSL•FO
RenderX

Table 1. Ecological momentary assessment questions and corresponding response options designed to assess the tailoring variables during the study period.

Tailoring variable	Question	Response options	Coaching domain (ie, physical activity or nutrition)
Activity	What are you currently doing?	• At work or studying • Leisure activity • Social occasion • Traveling • Eating (ie, other than a social occasion) • Being physically active • Otherwise	Both
Location	What is your current location?	• At home • At someone else's house • At work • In a vehicle • In a restaurant or café or on a terrace • At a sports club or other sports location • Outside, not otherwise specified • Otherwise	Both
Mood	How are you feeling at the moment?	• I feel positive emotions (eg, happy, cheerful, loved, confident, or relaxed) • I feel negative emotions (eg, sad, angry, stressed, ashamed, or anxious) • I feel neutral (neither positive nor negative)	Both
Restrictions in physical or mental condition	Are you currently experiencing restrictions in your overall condition?	• I experience no restrictions • I experience discomfort due to my physical limitations • I experience pain complaints • I experience diabetes-specific symptoms (eg, hyperglycemia or hypoglycemia) • I experience fatigue • Otherwise	Physical activity
Weather	How do you feel about the weather for doing physical activity outdoors?	• The weather is nice to be physically active outside • The weather is not suitable (ie, too hot, cold, or wet) to be physically active outside	Physical activity
Cravings	Are you currently craving food?	• No, I do not feel like eating • Yes, I feel like eating healthy • Yes, I feel like eating unhealthy • Yes, but I do not crave a specific food	Nutrition

Intervention Options

Intervention options represent the diverse behavior change actions that can be taken in response to decision points [24]. The content of the intervention options was rooted in the theoretical framework of the E-Supporter 1.0 [35], structured around 3 behavioral change phases: initiation, action, and maintenance. For each phase, relevant determinants of behavior were identified, and corresponding behavior change techniques (BCTs) were integrated to influence behavior effectively. BCTs are active components included in interventions designed to change behavior through various mechanisms [45].

The JITAI contained a set of 215 short SMS text messages, each tailored to address the tailoring variables (Table 2). These messages aimed to build skills, motivate, prompt reflection, and provide encouragement. Input to develop the intervention options was gathered from the following sources: the literature, the Dutch Diabetes Fund, the Dutch Nutrition Center, and the Dutch guidelines for physical activity and healthy diet [44,46]. The determinants and BCTs from the E-Supporter 1.0 were used as basis to develop the JITAI intervention options [35]. In addition, new determinants and BCTs were added to align with the selected tailoring variables, such as activity and location, which use environmental cues and context to prompt and guide individuals in various settings (eg, work, home, or during specific activities) [47]. For a comprehensive overview of determinants and BCTs, please refer to Multimedia Appendix 1.

XSL•FO
RenderX

Table 2. Examples of intervention options provided during the study, featuring tailored text messages customized for each tailoring variable.

Tailoring variable	Answer option	Example intervention option
Activity	At work or studying	"Do you have a sedentary job? Scheduling a daily walk during your break can help. Maybe your colleagues would like to join you. It is not only good for physical activity but also a fun way to spend time together!"
Location	In a restaurant or café or on a terrace	"Hi! Eating out at a restaurant? Try to choose the healthier option on the menu. Here are some examples of fairly healthy main courses: lean meat (eg, chicken or turkey fillet) or fatty fish (eg, salmon), served with vegetables, potatoes, or whole grain rice."
Weather	The weather is nice to be physically active outside	"Hello *[name]*, with this beautiful weather, everything around you looks even nicer. Take a walk around the neighborhood and enjoy it!"
Mood	I feel negative emotions (think sad, angry, stressed, angry, ashamed, or anxious)	"Hello *[name]*, I am sorry you are not feeling great. Even when you are feeling down, a walk might help. Give it a try, it could do you some good!"
Physical restrictions	I experience fatigue	"Hi, Feeling tired? You can break up your daily physical activity into shorter sessions. That way, you will have time to rest in between."
Food cravings	Yes, I feel like eating unhealthy food	"If the temptation to snack is strong, doing something else can help, like going for a walk or talking to someone. A little distraction can make the craving fade!"

Decision Rules

The decision rules serve as the link between tailoring variables and intervention options, enabling personalized adaptation of the intervention based on when, for whom, and which intervention option to offer [24]. All intervention options were systematically coded according to a behavioral goal, phase of behavior change, determinants of behavior, BCTs, tailoring variable from the EMA, and time of day. In our intervention, we used rule-based decisions with if-then statements—if certain conditions were met, then a specific intervention option was provided. A fitting intervention option was randomly selected. In instances where none of the assessed tailoring variables influenced health behavior or the EMA was not filled out, a text message was sent from the message set derived from the previously developed E-Supporter 1.0.

Decision Points

A decision point refers to a specific moment in time when the delivery of interventions to the user is determined [24]. In this JITAI, intervention decisions were triggered by each EMA prompt. EMAs were initiated through time-based sampling, occurring at 1 or 2 semirandom instances daily—once in the morning and again in the afternoon or evening.

Participants and Recruitment

We aimed to recruit 8 to 10 individuals diagnosed with T2D for this study, as this group size typically yields valuable insights into the acceptability of technologies [48]. To be eligible for participation, individuals needed to have a diagnosis of T2D, be aged ≥18 years, own a mobile phone, have the ability to understand Dutch, and possess the capacity to understand the informed consent procedure. Participants were recruited from the observational Diabetes and Lifestyle Cohort Twente study, a longitudinal study to obtain lifestyle, glucose, and disease-related parameters in patients with T2D [49], from Ziekenhuisgroep Twente (ZGT), a regional hospital in the east of the Netherlands. ZGT has a diabetes outpatient clinic where internists, nurse practitioners, and diabetes nurses provide daily care for a large population of patients with T2D. This hospital is also actively engaged in extensive scientific research related to diabetes. Individuals from the Diabetes and Lifestyle Cohort Twente cohort who had previously expressed willingness to participate in additional scientific research were contacted by telephone by the researcher (EAGH) to determine their interest in participating in this study. Detailed study information was provided orally, and interested individuals were sent the participant information sheet via email. After a week, potential participants were contacted again by telephone to inquire about their decision to participate. Those who chose to participate provided their informed consent by signing a digital or paper consent form before study enrollment. Although participants were recruited from ZGT hospital, participants used the JITAI in their home situation.

Procedures

The researcher introduced herself and the aims of the study before the start of the intervention. Participants were instructed to use the intervention for a period of 2 weeks, with the option to receive coaching for physical activity, healthy nutrition, or both lifestyle domains. Before starting the intervention, participants completed a web-based survey, providing questionnaires related to self-efficacy and phase of behavior change (refer to the Data Collection and Analysis section). The information on participants' self-efficacy and phase of behavior change was used to identify the most suitable self-efficacy messages and phase of behavior change messages for delivering intervention options (eg, people in the action phase received fewer messages related to behavior initiation and more messages regarding the action and maintenance phase) [35].

Throughout the intervention, participants were prompted to fill in the EMAs at 1 or 2 semirandom time intervals each day—one in the morning and another in the afternoon or evening [26]. At the prompt time, participants received a text message containing a link to access the EMA questionnaire. Intervention options were delivered via texting, with a maximum of 2 messages per day.

Following the 2-week intervention period, semistructured interviews were conducted via telephone and audio recorded

XSL•FO
RenderX

by EAGH, a female PhD candidate in health sciences, to gather in-depth insights and feedback from participants regarding their experiences with the intervention. The interviewer followed a preestablished interview guide, had prior experience in conducting interviews, and had no prior relationship with the study participants. Each interview lasted between 20 and 40 minutes.

Data Collection and Analysis

Acceptability

After the 2-week intervention period, participants engaged in a semistructured interview to elicit their thoughts, experiences, and suggestions for improvement. We examined the comprehensiveness and ease of use regarding the use of the JITAI [18]. Furthermore, the focus was primarily on satisfaction with the experience and the perceived usefulness of the intervention [40]. The interview schedule (Multimedia Appendix 2) covered three core parts: (1) acceptability of the EMA content and prompts, (2) acceptability of the intervention options (ie, the text messages), and (3) use of the JITAI (ie, the combinations of EMA and the messages).

The interview questions regarding the EMA questionnaire section were based on recommendations to assess the acceptability of EMAs [50], ensuring questions regarding item relevance, comprehensiveness, and instrument comprehensibility. In addition, participants' perceptions of EMA burden in terms of length, intensity, and motivation for completion were addressed. To evaluate the intervention options and the use of the JITAI, the interview schedule was based on the *Unified Theory of Acceptance and Use of Technology 2* framework [51]. Questions aligned with the *Unified Theory of Acceptance and Use of Technology 2* determinants, such as performance expectancy, effort expectancy, use behavior, and behavioral intention.

The interviews were analyzed based on a hybrid approach of thematic analysis, as outlined by Bingham [52] using ATLAS.ti 23 (Lumivero, LLC). This method combines both inductive and deductive strategies, enabling the development of new insights from the data while grounding the findings in an existing theoretical framework. The inductive phase began with reviewing the transcripts to familiarize with participants' responses, followed by an open inductive coding approach. Each text segment offering a notable perspective on the research questions received a descriptive open code. EAGH coded all transcripts, while AM independently coded 25% (2/8) and cross-checked the remainder. Open codes were iteratively refined, and discrepancies between EAGH and AM were resolved through discussions until consensus was achieved. The process of axial coding followed, involving clustering and combining open codes to identify patterns within and across interviews. Themes were developed by incorporating axial codes into theme statements that captured the observed data patterns. In the deductive phase, the *conceptual model of JITAI components* proposed by Nahum-Shani et al [24] was used to refine and classify the theme statements based on JITAI's main intervention components. Initially, themes were broadly assigned to the JITAI component where the definitions showed the greatest overlap. This classification was iteratively revised, with themes being reassigned until all themes were grouped under the most appropriate intervention component within the model. Throughout these phases, the research team continually discussed and refined the findings until consensus was reached.

Self-Efficacy Level

Self-efficacy was measured using the Exercise Self-Efficacy Scale (ESES) [53], consisting of 10 items that assess self-confidence in performing physical activity. In addition, a modified version of the ESES was used to determine self-efficacy related to a healthy diet [54]. The assessment used a 4-point Likert scale, ranging from *1 (ie, not at all true)* to *4 (ie, always true)*, yielding sum scores between 10 and 40, with higher scores indicating higher self-efficacy levels. Self-efficacy scores were treated as a continuous variable, obtained by summing the scores from the 10 items of the ESES questionnaire. Mean (SD) was used to present self-efficacy scores.

Phase of Behavior Change

To assess the phase of behavior change, single-question self-assessment stages of change scale [55] was used. This scale uses a single question to inquire about an individual's engagement in healthy behavior, aligning with the 5 phases of behavior change described in the transtheoretical model [56]. The phase of behavior change was categorized as an ordinal variable, ranging from 1 to 5, representing the following phases: (1) maintenance, (2) action, (3) preparation, (4) contemplation, and (5) precontemplation. These 5 answer options were aligned with the 3 behavior change phases targeted in the E-Supporter 1.0 intervention to allow for tailoring of intervention content. Participants who selected answer options 3, 4, or 5 were in the initiation phase, participants who chose answer option 2 were classified in the action phase, and participants who selected answer option 1 were categorized in the maintenance phase.

Demographics

Information on age (years), sex (ie, male, female, or other), BMI (kg/m^2), duration of T2D diagnosis (years), medication use (ie, none, oral, or insulin), and diabetes-related complications (ie, none, retinopathy, nephropathy, or neuropathy) was obtained from the hospital's electronic patient record. Having a job (ie, yes or no) and highest level of education (ie, no formal education, elementary education, lower secondary education, upper secondary education, higher vocational education, preuniversity education, university education, or postgraduate education) were questioned. Demographic characteristics were reported as mean (SD) for continuous variables or frequency (%) for categorical variables.

Ethical Considerations

The study protocol was reviewed and approved by the Medical Ethics Research Committee Twente, Enschede, the Netherlands (K20-05). In addition, the local advisory committee of practical feasibility in ZGT hospital approved to conduct this study (ZGT17-39). We obtained informed consent from all study participants, either digitally or on paper. Participants did not receive financial compensation for their participation due to the short duration of the study. Audio recordings of the interviews

XSL•FO
RenderX

were deleted after transcription. All participant-identifiable data were removed or pseudonymized. The data are stored on a secure, encrypted server at the University of Twente, accessible only to the researchers.

Results

Participant Characteristics

In total, 12 individuals were invited to participate in the study. Three people declined participation due to personal circumstances or perceiving the study burden as too high. One participant initially agreed but withdrew prematurely due to family circumstances. The characteristics of the 8 participants who completed the study are presented in Table 3. The participants had a mean age of 70.5 (SD 9.0) years, a mean BMI of 32.1 (SD 5.3) kg/m^2, and a mean T2D diagnosis of 15.6 (SD 7.7) years. Microvascular complications included neuropathy in 50% (4/8) of the participants and nephropathy in 25% (2/8) of the participants. In addition, of the 8 participants, 5 (63%) were using insulin and 6 (75%) remained active in work or volunteer activities. The most common educational level was higher vocational education (4/8, 50%).

A total of 6 participants opted for coaching on both physical activity and nutrition, while 1 participant preferred coaching exclusively on physical activity and another participant preferred coaching solely on nutrition. The results indicate that participants reported higher self-efficacy levels in the domain of nutrition compared to physical activity, with mean scores of 32 (SD 4) and 29 (SD 7), respectively (Table 4). Regarding the phase of behavior change, a large proportion of participants were in the maintenance phase for both physical activity (3/7, 43%) and nutrition (5/7, 71%).

Table 3. Participant characteristics (n=8).

Characteristic	Participants
Age (y), mean (SD)	70.5 (9)
Sex, n (%)	
Male	6 (75)
Female	2 (25)
BMI (kg/m^2), mean (SD)	32.1 (5.3)
Type 2 diabetes duration (y), mean (SD)	15.6 (7.7)
Complications, n (%)	
Retinopathy	0 (0)
Neuropathy	4 (50)
Nephropathy	2 (25)
Insulin treatment, n (%)	5 (63)
Employed, n (%)	6 (75)
Highest level of education, n (%)	
No formal education	1 (13)
Elementary education	0 (0)
Lower secondary vocational education	1 (13)
Secondary vocational education	1 (13)
Upper secondary vocational education	0 (0)
Higher vocational education	4 (50)
Preuniversity education	0 (0)
University education	0 (0)
Postgraduate education	0 (0)
Other	1 (13)

XSL•FO
RenderX

Table 4. Self-efficacy levels and phase of behavior change of participants for the physical activity and nutrition domains.

	Physical activity (n=7)	Nutrition (n=7)
Self-efficacy level, mean (SD)	29 (7)	32 (4)
Phase of behavior change, n (%)		
Initiation	3 (43)	2 (29)
Action	1 (14)	0 (0)
Maintenance	3 (43)	5 (71)

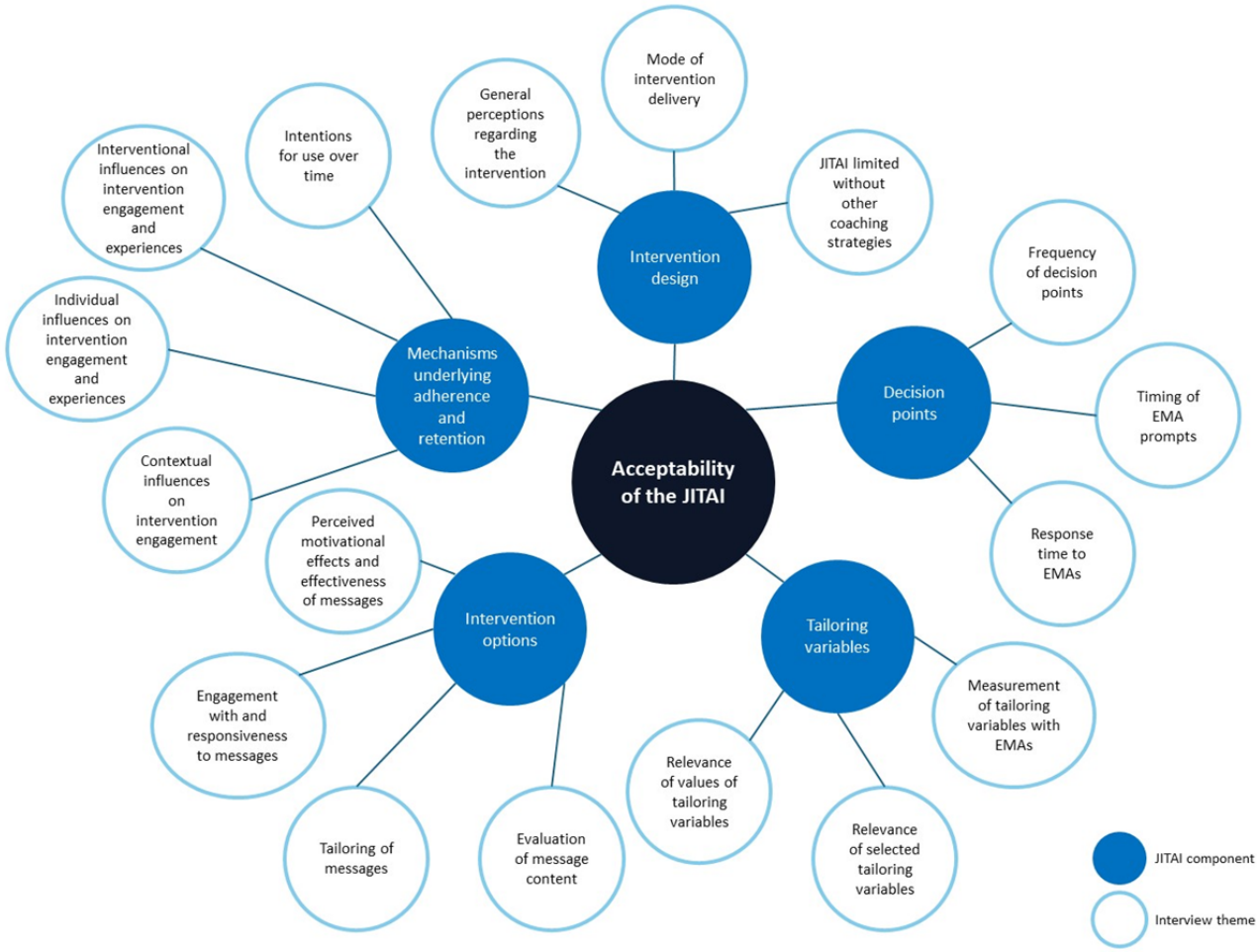
Acceptability

Overview

Figure 2 [24] provides an overview of various JITAI components from the conceptual model of JITAIs by Nahum-Shani et al [24], with the themes identified from the interviews.

Table 5 presents the results of the interviews. The bold terms represent the various components from the conceptual model of JITAIs by Nahum-Shani et al [24], including the identified themes and their associated definitions that were identified from the interviews.

Figure 2. Overview of the just-in-time adaptive intervention (JITAI) components and associated themes identified from the interviews with individuals with type 2 diabetes. EMA: ecological momentary assessment.



XSL•FO
RenderX

Table 5. Summary of the interview themes classified under JITAI[a] components related to the acceptability of the JITAI.

JITAI components and identified themes	Definitions
Intervention design	
General perceptions regarding the intervention	General perceptions about the intervention and associated components.
Mode of intervention delivery	Preferences regarding the method through which the intervention is delivered.
JITAI limited without other coaching strategies	Views on the need for additional support and the integration of other behavior change strategies.
Decision points	
Frequency of decision points	Perceptions and preferences regarding the frequency of EMA[b] prompts and intervention options.
Timing of EMA prompts	Perceptions and experiences regarding the appropriateness and predictability of the timing of EMA prompts.
Response time to EMAs	Behaviors and perceptions regarding the timeliness of their responses to EMA prompts.
Tailoring variables	
Measurement of tailoring variables with EMAs	Perspectives on the usefulness and constraints associated with using EMAs to measure tailoring variables.
Relevance of selected tailoring variables	Perceptions of the appropriateness and usefulness of the tailoring variables included in the EMAs.
Relevance of values of tailoring variables	Perceptions of the suitability and adequacy of the response options provided for the tailoring variables in the EMAs.
Intervention options	
Evaluation of message content	Perceptions and assessments of the content of the messages.
Tailoring of messages	Perceptions regarding the personalization and relevance of the messages they received.
Engagement with and responsiveness to messages	Interactions with and reactions to the messages.
Perceived motivational effects and effectiveness of messages	Perceptions of how the messages participants received influenced their motivation and behavior change efforts.
Mechanisms underlying adherence and retention	
Contextual influences on intervention engagement	External factors that impact participants' engagement with the intervention.
Individual influences on intervention engagement and experiences	Personal factors that affect participants' engagement and experiences with the intervention.
Interventional influences on intervention engagement and experiences	Various factors related to the intervention itself that influence participants' engagement and experiences.
Intentions for use over time	Intentions and willingness to continue using the JITAI system over a certain duration.

[a]JITAI: just-in-time adaptive intervention.

[b]EMA: ecological momentary assessment.

Intervention Design

All intervention components received *average* to *good* ratings, ranging from 7 to 9, with all components scoring approximately equally.

Overall, participants found this JITAI to be an easy-to-use and generally useful concept for supporting a healthy lifestyle. In addition to providing advice on a healthy lifestyle, participants felt that it also made them more aware of their lifestyle choices for the day.

In addition, most participants appreciated receiving technology-supported lifestyle guidance. In this study, SMS text messaging was used, which was perceived as an easy medium. Most participants preferred using an app, although 1 participant found using an app more complicated than SMS text messaging. Participant 8 had a clear preference for using an app; he stated the following:

> *I do everything via my phone. I find using a phone or tablet excellent. ... I would actually enjoy that [using an app], it is the easiest thing there is. I do everything on my mobile. It is convenient if personal goals and medical records are also available on it.* [Male, aged 59 years]

However, some participants preferred in-person support or guidance, such as support from a spouse as social support or professional guidance from a health care provider. The preference for in-person guidance mainly stemmed from the

XSL•FO
RenderX

ability to provide more context for certain events or feelings. Participant 3 illustrated this as follows:

> *For example, negative [mood] I will never indicate [in an EMA]. If I disagree with something, I will never put it in writing like that. If I have you on the phone personally, I will say what I feel, but on paper, that is not how I am. If I have to check on paper whether you are positive or negative, I will always indicate positive because if I indicate negative, I cannot provide any context, and then it is just how you want to read it and how you want to hear it. If I can provide an explanation to you, then I will more likely indicate that something is negative.* [Male, aged 77 years]

Some participants felt that relying solely on EMA-based messages to promote lifestyle was limited. They believed that coaching in this form alone lacked the pressure to actually change behavior. Examples to address this lack of pressure included setting goals and self-monitoring. Participant 1 had previously worked with goal setting and mentioned the following:

> *For example, the goal was: achieve 10,000 steps this week. And if I am missing 100, then I think: I will take a few more steps, I still want to reach those 10,000. It is an extra movement and pressure to do the exercise that I would not normally do, just to meet the set goal.* [Male, aged 78 years]

Other options mentioned were to make coaching more appealing by linking coaching to medical information about the participant or by providing healthy eating advice based on offers at the local supermarket.

Decision Points

For most participants, the number of prompts per day was acceptable, partly due to the low burden of completing the EMAs. However, some participants preferred a different prompt frequency, with options for both more and fewer EMA prompts mentioned. One participant suggested increasing the frequency from 3 to 4 EMA prompts per day, as it triggered engagement with a healthy lifestyle. Another participant preferred receiving the EMA prompts every other day, given their retired lifestyle with fewer varied activities.

The timing of the EMA prompts and messages was considered suboptimal by most participants. While they indicated that the timing was not disruptive, some prompts were deemed inconvenient, particularly when they occurred at times when participants could not respond, such as during work. Furthermore, some participants found the timing of the prompts predictable, as they often received them at approximately the same times each day, suggesting that more variation in timing would be preferred. Participant 5 elaborated on this, as follows:

> *For example, feeling hungry, that was often in the morning. But I had just eaten breakfast, so then I am not hungry and not craving anything. If you were to receive the questionnaire more in the evenings, that might be different, because then you're more likely watching TV and, well, then you might have a different answer.* [Female, aged 77 years]

The occasionally unsuitable timing of the prompts and the fact that not everyone consistently had their mobile phone with them affected the time frame used to complete the EMAs. Some participants deliberately made time to respond to the prompts upon receiving them, while others waited for a suitable moment. None of the participants intentionally delayed completing the EMAs.

Tailoring Variables

One participant misunderstood the purpose of the EMA and did not grasp why they had to fill out a questionnaire so frequently, but overall, participants were satisfied with using an EMA to gain insight into their situation. However, two main points of criticism were that (1) EMA responses provide only a snapshot and (2) there was insufficient opportunity to provide context or additional explanations for the response options in the EMA.

Regarding the first point, participants noted that while an EMA is a useful tool to assess the situation at a specific moment, it lacks insight into what happens for the rest of the day and therefore does not adequately account for rapid changes that may occur in one's situation, as explained by participant 3:

> *For example, I can be very negative in the morning, but after noon, I can be in a very good mood. I can bump my head and start swearing, and that can change at any moment. Therefore, it is a bit of a snapshot. That is why I often give standard answers, and I do not think you will gain much insight from that.* [Male, aged 77 years]

With regard to the second point, some participants felt there was insufficient opportunity to provide explanations for multiple-choice question answers. They would prefer an open field where they could freely elaborate on the multiple-choice questions.

In addition, participants felt a slight obligation to complete the EMAs because they felt researchers were waiting for their responses. Furthermore, some participants mentioned that they could not always answer completely freely due to feeling the need to provide socially desirable responses and being reluctant to give *negative* answers. However, all participants were comfortable sharing personal information, such as location or activity, via an EMA. The reasons for this included generally having no objection to sharing personal information, having given informed consent for this, and feeling a sense of autonomy by only providing information if they wished to do so.

Regarding the relevance of the tailoring variables, participants felt that the chosen variables were generally suitable for providing insights into their personal and environmental situations. Some participants mentioned that there could be room to include additional questions in the EMA without substantially increasing the burden. One participant suggested adding a specific question about experiencing hyperglycemia and hypoglycemia, so that this could also be taken into account in lifestyle advice. Furthermore, in some cases, certain questions were not relevant to a person, such as the mood question, which some participants mentioned always being in a good mood.

XSL•FO
RenderX

Most participants found the values of the tailoring variables, meaning the answer options in the EMA, sufficiently complete and clear. However, some improvement points were mentioned, such as making more distinctions in the answers for certain tailoring variables, for example, by including more types of weather conditions or also adding a neutral option in the *weather* tailoring variable. In addition, participant 5 indicated that the activity *eating* could not provide the researchers with enough information on the current activity. She made the following suggestion:

> *You could also fill in "eating," I have done that before. Those things, it is clear, but it could be a bit clearer... For example, what am I eating? If you are eating your regular meal or if you are eating a snack. It would be nice if you can provide more information about that.* [Female, aged 77 years]

Intervention Options

Most participants found the content of the messages enjoyable, funny, and varied, with clear tips for improving lifestyle. Although there was some variation in how enjoyable participants found the messages, the content was not perceived as intrusive. Most feedback was related to the content of the recipes referenced in the messages. For example, 1 participant mentioned that the recipes were quite complex, requiring many ingredients that are not commonly used, leading to items expiring. Furthermore, some minor suggestions were made to improve the content of the messages, such as adding reminders.

Participants were generally satisfied with the degree of tailoring in the messages and could see that their EMA responses were reflected in the messages. However, participants also felt that there was variation in how well the messages aligned with individual circumstances. Participants noticed tailoring the most when they were addressed by their first names, when messages matched their personal preferences, and when messages matched their location or activity at that time. Participant 6 illustrated this with the following example:

> *Yes, I noticed from the little tips you got, for example asking a colleague for a walk during the break or traveling by bike to work instead of by car when you are at work.* [Female, aged 55 years]

Some participants had a more neutral opinion on the tailoring of the messages, such as that the messages sometimes still contained too much general knowledge or did not sufficiently align with their specific situation. Participant 4 believed that his preferences were considered, but he found that the messages did not sufficiently address his physical limitations:

> *Here it says at one point that it is good to move. I completely agree with you, but for me, it is currently difficult to move.... We used to do a lot of cycling and walking. I have had trouble with my right leg for five years now.* [Male, aged 71 years]

Participants indicated that they always read the messages, and some of them were even curious about the message they would receive based on the EMA, as indicated by participant 6:

> *It was fun. I was always waiting for the response I got. When I was at work, I got a message like: ask your colleague or do something with your colleague. When I had been to the store, I got different answers again, that was fun.* [Female, aged 55 years]

It was not always possible to read the messages immediately, but participants did so as soon as they could. Participants also showed a preference for interactivity, as several participants responded to the messages via SMS text messaging and expressed enjoyment in giving a reaction to the messages.

The responses regarding the motivating effect and perceived effectiveness of the messages varied. Almost all participants found receiving messages motivating to work on their lifestyle. Some participants indicated that they did not always act on the advice, partly because the advice was not always applicable to them. Nevertheless, it helped some participants make healthier choices, as participant 8 indicated with the following examples:

> *It did trigger me. Sometimes in the evening I thought: let's get some fries. Then I thought: no, let's just peel two potatoes, with a piece of meat and vegetables. I did that twice.* [Male, aged 59 years]

> *For example, that church service. Are you going to church? How? In that questionnaire, you indicate that you are going to do an activity. Then you get a message with: how could you improve this activity? Do you think about this? Are you going by bike, car, or walking? I started cycling, otherwise I would have taken the car.* [Male, aged 59 years]

Mechanisms Underlying Adherence and Retention

A contextual factor influencing engagement with the intervention was the COVID-19 pandemic and the lockdown period in the Netherlands during which the study was conducted. For instance, 1 participant mentioned that daily activities had substantially reduced due to the COVID-19 pandemic, referring to limited opening hours of stores and restaurants during the lockdown. As a result, the participant reported providing nearly identical answers every day, which lowered his engagement with the intervention. Another participant mentioned that they had become less active during the COVID-19 pandemic due to disruptions in their regular routine.

In addition, participants indicated various individual factors that influenced the use of the intervention. Some participants reported making many lifestyle adjustments or losing weight before participating in the study, making the intervention information less useful for them. Other participants expressed that they felt no need to improve their lifestyle because they were satisfied with their current health and lifestyle. Finally, some participants felt a slight resistance to actually adjusting their lifestyle, which seemed to be independent from the indicated phase of behavior change (Table 4). This feeling partly stemmed from their unwillingness to make certain lifestyle changes (eg, a participant who did not want to take the stairs) and from a desire to also be allowed to make unhealthy choices. Participant 2 already made many improvements to his lifestyle in recent years but indicated that he sometimes wanted to make unhealthy choices:

XSL•FO
RenderX

> *I am very careful with my eating. In the past 2.5 years, I have lost 18-20 kilos. I try to maintain my current weight now. That is certainly better for my diabetes too. But I do still snack occasionally.* [Male, aged 78 years]

The intervention itself also influenced engagement and experiences with the intervention. An important factor was the intervention burden. All participants found the burden of completing the EMA minimal. Participants found the EMA instructions and questions clear and the questionnaire brief, and it took little time to complete the EMA. They also found the link to the EMA questionnaire in the text message easy to use. As a result, participants remained motivated. However, the lack of variation in the EMA questionnaire reduced engagement over time for some participants. Because participants received the exact same questions every day, they found it monotonous. Several participants mentioned that they started filling out the questions on *autopilot*, as participant 5 described:

> *Overall, there was enough motivation, but because it was always the same questionnaire, you start thinking about it less. Only for activities like drinking coffee, I sometimes gave a different answer, but otherwise, it was always the same answer. You do not have to think about it anymore.* [Female, aged 77 years]

Participants suggested rearranging the questions or asking different questions on different days to make the EMA more engaging and encourage careful consideration when completing it.

In addition, participants provided feedback on the continued use of the intervention, with most participants willing to use the intervention for >2 weeks. Some participants were willing to fill out EMAs and receive messages daily without indicating a maximum period of use, but others saw more value in a maximum period of several weeks to months.

Discussion

Principal Findings

With this study, we aimed to gain insight into the acceptability of our initial JITAI version among individuals with T2D. This preliminary phase is pivotal for gaining insights into the target group's experiences with the intervention, allowing us to assess and enhance its potential value based on their feedback [18]. Overall, participants expressed satisfaction with the intervention, as supported by their average to high ratings across its components. This opinion was evident from the positive feedback on the motivating and varied text messages, as well as some examples of personalized support that participants experienced.

It became evident that experiences varied substantially and were often influenced by individual and contextual factors that were not sufficiently captured with the EMAs. This variability was observed across all aspects of the intervention design. On the basis of the participants' feedback, we have identified three key areas for improvement: (1) optimizing tailoring, (2) refining assessment methods, and (3) emphasizing integration with other behavioral intervention strategies.

Regarding the optimization of tailoring, our study highlighted the need for further refinement in this area. Although tailoring variables were selected based on target group input, their relevance varied among participants. For instance, mood- or disease-related symptoms were relevant to some but not all individuals. The decision to offer 2 prompts per day was informed by previous E-Supporter studies [35,36], yet participant feedback on prompt frequency varied. Similarly, we noted nonoptimal timing of prompts, aligning with challenges observed in related studies [57]. This heterogeneity among individuals' experiences requires a shift toward individually tailored intervention designs. Realizing such individual tailoring can be approached in several ways. A straightforward method is to allow individuals to specify their preferences for tailoring variables, intervention options, as well as the frequency and timing of prompts before or during the intervention period [58,59]. Applying machine learning, such as reinforcement learning (RL), is a more advanced method for improving tailoring. RL learns and adapts from interaction and feedback. RL allows to tailor support strategies to individual needs and preferences regarding timing, frequency, and intervention types [60-63].

With respect to the assessment methods, participants mentioned the potential of EMAs to personalize intervention support, but they also identified limitations. These limitations included the repetitive nature of the EMA questionnaire and its snapshot insight into the current situation. Some participants reported reduced engagement over time, possibly attributed to the lack of variation in questions or the impact of the COVID-19 pandemic and associated restrictions during the study period [64]. Lockdown measures reduced social interactions and outdoor activities, leading to participants spending more time at home and engaging less in social or work-related activities. Consequently, responses to EMAs became predictable, and the participant's interest diminished over time. EMAs rely on user input to deliver personalized support [27]. Decreased user engagement may lead to either skipping questions or responding without due consideration. This may influence the reliability and validity of measurements. As Nahum-Shani et al [24] emphasized, balancing response burden with the need for frequent assessment for capturing dynamic changes in certain variables is critical. Exploring variability in measured constructs across EMAs and considering alternative measurement methods for tailoring variables, such as GPS tracking for location and weather application programming interface, could simplify the data collection [57,61,65,66]. Nevertheless, issues such as privacy concerns with GPS and individuals' perceptions of weather suitability for outdoor activities require consideration. While optimal measurement strategies for tailoring variables remain uncertain, we aim to find the right balance between passive and active methods for accommodating their variability in future interventions.

To realize sustainable behavior change, our study revealed that many participants found the intervention motivating, but some expressed no need, unwillingness, or resistance to follow the advice. The participants had T2D for an average of 15.6 (SD

XSL•FO
RenderX

7.7) years, which suggests they were already familiar with making lifestyle changes. This is supported by their relatively high self-efficacy scores, as individuals with high self-efficacy tend to feel confident in their ability to make behavioral changes. Consequently, their perceived need for external support from the intervention may have been reduced. Furthermore, many participants were already in the maintenance phase of behavior change, suggesting that they had already implemented some lifestyle changes. These individuals might require specific support for relapse prevention or, in some cases, no additional support at all [43]. A smaller group was in the initiation phase, indicating that they were just beginning to consider lifestyle changes. This group might require a stronger focus on motivational strategies to build commitment [1,2]. These differences highlight the need for interventions that can adapt to the wide variety of individual needs. However, JITAIs are generally designed to offer brief, just-in-time suggestions for immediate behavior adjustments [26,28,57]. While the JITAI may be effective for action-related changes, this design may fall short of addressing the more comprehensive strategies required. Therefore, we emphasize the importance of situating JITAIs within broader behavior change interventions. Integrating a JITAI within a comprehensive lifestyle intervention that includes extensive behavioral change strategies, as envisioned in our integration with the E-Supporter content, could provide the necessary support to facilitate sustainable behavior change.

Strengths and Limitations

A strength of this study was the application of a hybrid approach to thematic analysis for analyzing the interviews. An advantage of this approach is that we were able to combine the strengths of both inductive and deductive approaches. The hybrid approach allowed us to use the flexibility of inductive coding to create data-driven insights and also use the deductive method's ability to build upon existing knowledge, ensuring that the findings were both data driven and theoretically grounded.

A limitation of our study was the inability to measure the impact on proximal outcomes. Ideally, we would have supported the intervention with wearable activity monitoring and an app for dietary monitoring. However, due to COVID-19 restrictions preventing physical contact with participants, we were unable to distribute wearables or assist with app installations. In the future, our aim is to incorporate goal setting and self-monitoring into the intervention, as done in the E-Supporter. This approach can provide additional support in the behavior change process and enable examination of the intervention effects on proximal outcomes. As it was not the aim of this study to measure these outcomes, it does not affect the answer to the research question. Another limitation is the limited evaluation of intervention options for certain tailoring variables, such as negative mood or poor weather. As reported by participants, negative mood was seldom indicated in the EMA, resulting in no corresponding intervention options being sent. Furthermore, the evaluation was conducted during sunny weather, leading to minimal use of intervention options for poor weather conditions. Consequently, we lacked insight into the acceptability of these intervention options. However, the content and BCTs used in these intervention options are similar to others, so we anticipate similar perceptions of these messages compared to others. Finally, our sample size was relatively small, which may affect the robustness of the results. While this study provided valuable insights into the acceptability of the JITAI, our study population was relatively homogeneous, consisting predominantly of older individuals with a longer duration of diabetes. These characteristics appear to be a relatively good representation of the target group, characterized by an older average age, a slightly higher proportion of men than women, and including individuals with lower education levels, which are all sociodemographic factors associated with T2D [67,68]. However, the homogeneity limits the strength of our conclusions. Future research should focus on recruiting a more diverse population, including younger individuals and those with a shorter diabetes duration, to improve the generalizability of our findings.

Conclusions

This study presents the first version of an EMA-driven JITAI designed to support a healthy lifestyle in individuals with T2D. The intervention received positive feedback, particularly for its motivating messages that felt tailored. However, individual differences in experiences highlight the need for improvement in tailoring, assessment methods, and the integration of JITAI with other behavioral intervention strategies.

Acknowledgments

The authors would like to express their gratitude to all participants who were willing to participate in this study, especially because the COVID-19 pandemic made it necessary to conduct all communications remotely. The authors used the assistance of the generative artificial intelligence tool ChatGPT by OpenAI [69] to check English grammar and refine language fluency but not for data collection, data analysis, or writing the manuscript.

Data Availability

The data generated during this study are not publicly available because of the personal information shared during the interviews but are available from the corresponding author on reasonable request.

Authors' Contributions

EAGH designed the intervention, performed the data collection, and drafted the initial version of the manuscript. AM contributed to the data analysis and ongoing refinement of the manuscript. WH, LKS, GDL, and MMRV-H provided ongoing feedback related

XSL•FO
RenderX

to the intervention design, methods, results, and discussion of this study. All the authors critically evaluated and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
E-Supporter determinants and behavior change techniques.
[PDF File (Adobe PDF File), 178 KB - formative_v9i1e65026_app1.pdf]

Multimedia Appendix 2
Translated interview schedule.
[PDF File (Adobe PDF File), 206 KB - formative_v9i1e65026_app2.pdf]

References

1. Golovaty I, Ritchie ND, Tuomilehto J, Mohan V, Ali MK, Gregg EW, et al. Two decades of diabetes prevention efforts: a call to innovate and revitalize our approach to lifestyle change. Diabetes Res Clin Pract 2023 Apr;198:110195 [FREE Full text] [doi: 10.1016/j.diabres.2022.110195] [Medline: 36470316]
2. van Ommen B, Wopereis S, van Empelen P, van Keulen HM, Otten W, Kasteleyn M, et al. From diabetes care to diabetes cure-the integration of systems biology, eHealth, and behavioral change. Front Endocrinol (Lausanne) 2017;8:381 [FREE Full text] [doi: 10.3389/fendo.2017.00381] [Medline: 29403436]
3. Lean MEJ, Leslie WS, Barnes AC, Brosnahan N, Thom G, McCombie L, et al. Durability of a primary care-led weight-management intervention for remission of type 2 diabetes: 2-year results of the DiRECT open-label, cluster-randomised trial. Lancet Diabetes Endocrinol 2019 May;7(5):344-355 [FREE Full text] [doi: 10.1016/S2213-8587(19)30068-3] [Medline: 30852132]
4. Ried-Larsen M, Johansen MY, MacDonald CS, Hansen KB, Christensen R, Wedell-Neergaard AS, et al. Type 2 diabetes remission 1 year after an intensive lifestyle intervention: a secondary analysis of a randomized clinical trial. Diabetes Obes Metab 2019 Oct;21(10):2257-2266 [FREE Full text] [doi: 10.1111/dom.13802] [Medline: 31168922]
5. Pot GK, Battjes-Fries MC, Patijn ON, van der Zijl N, Pijl H, Voshol P. Lifestyle medicine for type 2 diabetes: practice-based evidence for long-term efficacy of a multicomponent lifestyle intervention (Reverse Diabetes2 Now). BMJ Nutr Prev Health 2020 Dec 18;3(2):188-195 [FREE Full text] [doi: 10.1136/bmjnph-2020-000081] [Medline: 33521528]
6. Chen L, Pei JH, Kuang J, Chen HM, Chen Z, Li ZW, et al. Effect of lifestyle intervention in patients with type 2 diabetes: a meta-analysis. Metabolism 2015 Feb;64(2):338-347 [FREE Full text] [doi: 10.1016/j.metabol.2014.10.018] [Medline: 25467842]
7. García-Molina L, Lewis-Mikhael AM, Riquelme-Gallego B, Cano-Ibáñez N, Oliveras-López MJ, Bueno-Cavanillas A. Improving type 2 diabetes mellitus glycaemic control through lifestyle modification implementing diet intervention: a systematic review and meta-analysis. Eur J Nutr 2020 Jun;59(4):1313-1328. [doi: 10.1007/s00394-019-02147-6] [Medline: 31781857]
8. Gostoli S, Raimondi G, Popa AP, Giovannini M, Benasi G, Rafanelli C. Behavioral lifestyle interventions for weight loss in overweight or obese patients with type 2 diabetes: a systematic review of the literature. Curr Obes Rep 2024 Jun;13(2):224-241 [FREE Full text] [doi: 10.1007/s13679-024-00552-5] [Medline: 38436787]
9. Espeland MA, Glick HA, Bertoni A, Brancati FL, Bray GA, Clark JM, et al. Impact of an intensive lifestyle intervention on use and cost of medical services among overweight and obese adults with type 2 diabetes: the action for health in diabetes. Diabetes Care 2014 Sep;37(9):2548-2556 [FREE Full text] [doi: 10.2337/dc14-0093] [Medline: 25147253]
10. Mirahmadizadeh A, Khorshidsavar H, Seif M, Sharifi MH. Adherence to medication, diet and physical activity and the associated factors amongst patients with type 2 diabetes. Diabetes Ther 2020 Feb 08;11(2):479-494 [FREE Full text] [doi: 10.1007/s13300-019-00750-8] [Medline: 31916213]
11. Hagedoorn IJM, den Braber N, Oosterwijk MM, Gant CM, Navis G, Vollenbroek-Hutten MMR, et al. Low physical activity in patients with complicated type 2 diabetes mellitus is associated with low muscle mass and low protein intake. J Clin Med 2020 Sep 25;9(10):3104 [FREE Full text] [doi: 10.3390/jcm9103104] [Medline: 32992990]
12. Jalving AC, Gant CM, Binnenmars SH, Soedamah-Muthu SS, Bakker SJL, Navis G, et al. Glycaemic control in the diabetes and Lifestyle Cohort Twente: a cross-sectional assessment of lifestyle and pharmacological management on Hba1c target achievement. Diabetes Obes Metab 2018 Oct;20(10):2494-2499 [FREE Full text] [doi: 10.1111/dom.13399] [Medline: 29862616]
13. Jarvie JL, Pandey A, Ayers CR, McGavock JM, Sénéchal M, Berry JD, et al. Aerobic fitness and adherence to guideline-recommended minimum physical activity among ambulatory patients with type 2 diabetes mellitus. Diabetes Care 2019 Jul;42(7):1333-1339 [FREE Full text] [doi: 10.2337/dc18-2634] [Medline: 31221698]

XSL•FO
RenderX

14. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2224-2260 [FREE Full text] [doi: 10.1016/S0140-6736(12)61766-8] [Medline: 23245609]
15. Schmidt SK, Hemmestad L, MacDonald CS, Langberg H, Valentiner LS. Motivation and barriers to maintaining lifestyle changes in patients with type 2 diabetes after an intensive lifestyle intervention (the U-TURN trial): a longitudinal qualitative study. Int J Environ Res Public Health 2020 Oct 13;17(20):7454 [FREE Full text] [doi: 10.3390/ijerph17207454] [Medline: 33066239]
16. Ahola AJ, Groop PH. Barriers to self-management of diabetes. Diabet Med 2013 Apr;30(4):413-420. [doi: 10.1111/dme.12105] [Medline: 23278342]
17. Ahmad F, Joshi SH. Self-care practices and their role in the control of diabetes: a narrative review. Cureus 2023 Jul;15(7):e41409 [FREE Full text] [doi: 10.7759/cureus.41409] [Medline: 37546053]
18. Kip H, van Gemert-Pijnen L, Jong NB, Sanderman R, Kelders SM. eHealth Research Theory and Development: A Multidisciplinary Approach. Milton Park, UK: Taylor & Francis; 2018.
19. Milne-Ives M, Lam C, de Cock C, van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
20. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can smartphone apps increase physical activity? Systematic review and meta-analysis. J Med Internet Res 2019 Mar 19;21(3):e12053 [FREE Full text] [doi: 10.2196/12053] [Medline: 30888321]
21. Schoeppe S, Alley S, van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
22. Villinger K, Wahl DR, Boeing H, Schupp HT, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: a systematic review and meta-analysis. Obes Rev 2019 Oct;20(10):1465-1484 [FREE Full text] [doi: 10.1111/obr.12903] [Medline: 31353783]
23. Singh B, Ahmed M, Staiano AE, Gough C, Petersen J, Vandelanotte C, et al. A systematic umbrella review and meta-meta-analysis of eHealth and mHealth interventions for improving lifestyle behaviours. NPJ Digit Med 2024 Jul 05;7(1):179 [FREE Full text] [doi: 10.1038/s41746-024-01172-y] [Medline: 38969775]
24. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
25. Spruijt-Metz D, Nilsen W. Dynamic models of behavior for just-in-time adaptive interventions. IEEE Pervasive Comput 2014 Jul 09;13(3):13-17. [doi: 10.1109/MPRV.2014.46]
26. Goldstein SP, Evans BC, Flack D, Juarascio A, Manasse S, Zhang F, et al. Return of the JITAI: applying a just-in-time adaptive intervention framework to the development of m-health solutions for addictive behaviors. Int J Behav Med 2017 Oct;24(5):673-682 [FREE Full text] [doi: 10.1007/s12529-016-9627-y] [Medline: 28083725]
27. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
28. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr 03;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
29. Perski O, Hébert ET, Naughton F, Hekler EB, Brown J, Businelle MS. Technology-mediated just-in-time adaptive interventions (JITAIs) to reduce harmful substance use: a systematic review. Addiction 2022 May;117(5):1220-1241 [FREE Full text] [doi: 10.1111/add.15687] [Medline: 34514668]
30. Wang L, Miller LC. Just-in-the-moment adaptive interventions (JITAI): a meta-analytical review. Health Commun 2020 Nov;35(12):1531-1544. [doi: 10.1080/10410236.2019.1652388] [Medline: 31488002]
31. Oikonomidi T, Ravaud P, LeBeau J, Tran VT. A systematic scoping review of just-in-time, adaptive interventions finds limited automation and incomplete reporting. J Clin Epidemiol 2023 Feb;154:108-116. [doi: 10.1016/j.jclinepi.2022.12.006] [Medline: 36521653]
32. Forman EM, Goldstein SP, Zhang F, Evans BC, Manasse SM, Butryn ML, et al. OnTrack: development and feasibility of a smartphone app designed to predict and prevent dietary lapses. Transl Behav Med 2019 Mar 01;9(2):236-245 [FREE Full text] [doi: 10.1093/tbm/iby016] [Medline: 29617911]
33. Rabbi M, Pfammatter A, Zhang M, Spring B, Choudhury T. Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. JMIR Mhealth Uhealth 2015 May 14;3(2):e42 [FREE Full text] [doi: 10.2196/mhealth.4160] [Medline: 25977197]
34. Ding X, Xu J, Wang H, Chen G, Thind H, Zhang Y. WalkMore: promoting walking with just-in-time context-aware prompts. In: Proceedings of the IEEE Wireless Health. 2016 Presented at: WH 2016; October 25-27, 2016; Bethesda, MD. [doi: 10.1109/wh.2016.7764558]

XSL•FO
RenderX

35. Hietbrink EAG, Middelweerd A, van Empelen P, Preuhs K, Konijnendijk AAJ, Oude Nijeweme-d'Hollosy W, et al. A digital lifestyle coach (E-Supporter 1.0) to support people with type 2 diabetes: participatory development study. JMIR Hum Factors 2023 Jan 12;10:e40017 [FREE Full text] [doi: 10.2196/40017] [Medline: 36633898]
36. Hietbrink EAG, Oude Nijeweme-d'Hollosy W, Middelweerd A, Konijnendijk AAJ, Schrijver LK, Ten Voorde AS, et al. A digital coach (e-supporter 1.0) to support physical activity and a healthy diet in people with type 2 diabetes: acceptability and limited efficacy testing. JMIR Form Res 2023 Jul 28;7:e45294 [FREE Full text] [doi: 10.2196/45294] [Medline: 37505804]
37. Op den Akker H, Cabrita M, Op den Akker R, Jones VM, Hermens HJ. Tailored motivational message generation: a model and practical framework for real-time physical activity coaching. J Biomed Inform 2015 Jun;55:104-115 [FREE Full text] [doi: 10.1016/j.jbi.2015.03.005] [Medline: 25843359]
38. Lustria ML, Noar SM, Cortese J, van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
39. Chen J, Cade JE, Allman-Farinelli M. The most popular smartphone apps for weight loss: a quality assessment. JMIR Mhealth Uhealth 2015 Dec 16;3(4):e104 [FREE Full text] [doi: 10.2196/mhealth.4334] [Medline: 26678569]
40. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
41. Schwarzer R. Health action process approach (HAPA) as a theoretical framework to understand behavior change. Actualidades Psicología 2016 Dec 05;30(121):119-130. [doi: 10.15517/ap.v30i121.23458]
42. Rothman AJ. Toward a theory-based analysis of behavioral maintenance. Health Psychol 2000 Jan;19(1S):64-69. [doi: 10.1037/0278-6133.19.suppl1.64] [Medline: 10709949]
43. Marlatt GA, Donovan DM. Relapse Prevention: Maintenance Strategies in the Treatment of Addictive Behaviors, Second Edition. New York, NY: The Guilford Press; 2005.
44. Kromhout D, Spaaij CJ, de Goede J, Weggemans RM. The 2015 Dutch food-based dietary guidelines. Eur J Clin Nutr 2016 Aug;70(8):869-878 [FREE Full text] [doi: 10.1038/ejcn.2016.52] [Medline: 27049034]
45. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
46. Weggemans RM, Backx FJG, Borghouts L, Chinapaw M, Hopman MTE, Koster A, et al. The 2017 Dutch physical activity guidelines. Int J Behav Nutr Phys Act 2018 Jun 25;15(1):58 [FREE Full text] [doi: 10.1186/s12966-018-0661-9] [Medline: 29940977]
47. Johnston M, Carey RN, Connell Bohlen LE, Johnston DW, Rothman AJ, de Bruin M, et al. Development of an online tool for linking behavior change techniques and mechanisms of action based on triangulation of findings from literature synthesis and expert consensus. Transl Behav Med 2021 May 25;11(5):1049-1065 [FREE Full text] [doi: 10.1093/tbm/ibaa050] [Medline: 32749460]
48. Jaspers MWM. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009 May;78(5):340-353. [doi: 10.1016/j.ijmedinf.2008.10.002] [Medline: 19046928]
49. Gant CM, Mensink I, Binnenmars SH, van der Palen JAM, Bakker SJL, Navis G, et al. Body weight course in the DIAbetes and LifEstyle Cohort Twente (DIALECT-1)-a 20-year observational study. PLoS One 2019 Jun 19;14(6):e0218400 [FREE Full text] [doi: 10.1371/journal.pone.0218400] [Medline: 31216324]
50. Degroote L, DeSmet A, de Bourdeaudhuij I, van Dyck D, Crombez G. Content validity and methodological considerations in ecological momentary assessment studies on physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2020 Mar 10;17(1):35 [FREE Full text] [doi: 10.1186/s12966-020-00932-9] [Medline: 32151251]
51. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012 Mar;36(1):157-178. [doi: 10.2307/41410412]
52. Bingham AJ. From data management to actionable findings: a five-phase process of qualitative data analysis. Int J Qual Methods 2023 Aug;22:1-11. [doi: 10.1177/16094069231183620]
53. Kroll T, Kehn M, Ho PS, Groah S. The SCI Exercise Self-Efficacy Scale (ESES): development and psychometric properties. Int J Behav Nutr Phys Act 2007 Aug 30;4:34 [FREE Full text] [doi: 10.1186/1479-5868-4-34] [Medline: 17760999]
54. Fokkema EMS. The development and evaluation of an eCoaching module to improve healthy nutrition in type 2 diabetes patients. University of Twente. 2019. URL: https://essay.utwente.nl/79927/ [accessed 2023-07-28]
55. Marttila J, Nupponen R. Assessing stage of change for physical activity: how congruent are parallel methods? Health Educ Res 2003 Aug;18(4):419-428. [doi: 10.1093/her/cyf034] [Medline: 12939124]
56. Prochaska JO, Diclemente CC, Norcross JC. In search of how people change: applications to addictive behaviors. In: Marlatt GA, VandenBos GR, editors. Addictive Behaviors: Readings on Etiology, Prevention, and Treatment. Washington, DC: American Psychological Association; 1997:671-696.

XSL•FO
RenderX

57. Mair JL, Hayes LD, Campbell AK, Buchan DS, Easton C, Sculthorpe N. A personalized smartphone-delivered just-in-time adaptive intervention (JitaBug) to increase physical activity in older adults: mixed methods feasibility study. JMIR Form Res 2022 Apr 07;6(4):e34662 [FREE Full text] [doi: 10.2196/34662] [Medline: 35389348]
58. Kandola A, Edwards K, Straatman J, Dührkoop B, Hein B, Hayes J. Digital self-management platform for adult asthma: randomized attention-placebo controlled trial. J Med Internet Res 2024 Apr 29;26:e50855 [FREE Full text] [doi: 10.2196/50855] [Medline: 38684084]
59. Chen YP, Woodward J, Bista D, Zhang X, Singh I, Obajemu O, et al. Investigating contextual notifications to drive self-monitoring in mHealth apps for weight maintenance. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. 2024 Presented at: CHI '24; May 11-16, 2024; Honolulu, HI. [doi: 10.1145/3613904.3641993]
60. Gönül S, Namlı T, Coşar A, Toroslu İ. A reinforcement learning based algorithm for personalization of digital, just-in-time, adaptive interventions. Artif Intell Med 2021 May;115:102062. [doi: 10.1016/j.artmed.2021.102062] [Medline: 34001322]
61. Vandelanotte C, Trost S, Hodgetts D, Imam T, Rashid M, To QG, et al. Increasing physical activity using an just-in-time adaptive digital assistant supported by machine learning: a novel approach for hyper-personalised mHealth interventions. J Biomed Inform 2023 Aug;144:104435 [FREE Full text] [doi: 10.1016/j.jbi.2023.104435] [Medline: 37394024]
62. Liao P, Greenewald K, Klasnja P, Murphy S. Personalized HeartSteps: a reinforcement learning algorithm for optimizing physical activity. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Mar;4(1):18 [FREE Full text] [doi: 10.1145/3381007] [Medline: 34527853]
63. Grua EM, de Sanctis M, Malavolta I, Hoogendoorn M, Lago P. An evaluation of the effectiveness of personalization and self-adaptation for e-Health apps. Inf Softw Technol 2022 Jun;146:106841. [doi: 10.1016/j.infsof.2022.106841]
64. Ontwikkelingen coronavirus in 2020. Rijksoverheid. 2020. URL: https://www.rijksoverheid.nl/onderwerpen/coronavirus-tijdlijn/2020 [accessed 2024-05-17]
65. Thomas Craig KJ, Morgan LC, Chen CH, Michie S, Fusco N, Snowdon JL, et al. Systematic review of context-aware digital behavior change interventions to improve health. Transl Behav Med 2021 May 25;11(5):1037-1048 [FREE Full text] [doi: 10.1093/tbm/ibaa099] [Medline: 33085767]
66. Rajanna V, Lara-Garduno R, Behera DJ, Madanagopal K, Goldberg D, Hammond T. Step up life: a context aware health assistant. In: Proceedings of the Third ACM SIGSPATIAL International Workshop on the Use of GIS in Public Health. 2014 Presented at: HealthGIS '14; November 4, 2014; Dallas, TX. [doi: 10.1145/2676629.2676636]
67. Kontochristopoulou AM, Karatzi K, Karaglani E, Cardon G, Kivelä J, Wikström K, et al. Sociodemographic, anthropometric, and lifestyle correlates of prediabetes and type 2 diabetes in Europe: the Feel4Diabetes study. Nutr Metab Cardiovasc Dis 2022 Aug;32(8):1851-1862. [doi: 10.1016/j.numecd.2022.03.021] [Medline: 35610084]
68. Kyrou I, Tsigos C, Mavrogianni C, Cardon G, van Stappen V, Latomme J, et al. Sociodemographic and lifestyle-related risk factors for identifying vulnerable groups for type 2 diabetes: a narrative review with emphasis on data from Europe. BMC Endocr Disord 2020 Mar 12;20(Suppl 1):134 [FREE Full text] [doi: 10.1186/s12902-019-0463-3] [Medline: 32164656]
69. What can I help with? ChatGPT. URL: https://chatgpt.com/?ref=dotcom [accessed 2024-04-29]

Abbreviations

BCT: behavior change technique
EMA: ecological momentary assessment
ESES: Exercise Self-Efficacy Scale
JITAI: just-in-time adaptive intervention
RL: reinforcement learning
T2D: type 2 diabetes
ZGT: Ziekenhuisgroep Twente

Edited by A Mavragani; submitted 02.08.24; peer-reviewed by A Hornsten; comments to author 18.11.24; revised version received 06.12.24; accepted 18.12.24; published 19.02.25.

Please cite as:
Hietbrink EAG, Middelweerd A, d'Hollosy W, Schrijver LK, Laverman GD, Vollenbroek-Hutten MMR
Exploring the Acceptance of Just-in-Time Adaptive Lifestyle Support for People With Type 2 Diabetes: Qualitative Acceptability Study
JMIR Form Res 2025;9:e65026
URL: https://formative.jmir.org/2025/1/e65026
doi:10.2196/65026
PMID:39969969

XSL•FO
RenderX

©Eclaire A G Hietbrink, Anouk Middelweerd, Wendy d'Hollosy, Laura K Schrijver, Gozewijn D Laverman, Miriam M R Vollenbroek-Hutten. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Nomophobia, Psychopathology, and Smartphone-Inferred Behaviors in Youth With Depression: Longitudinal Study

Tianyi Zhang[1], MIT; Andres Camargo[2], MCR; Lianne Schmaal[2], PhD; Vassilis Kostakos[1], PhD; Simon D'Alfonso[1], PhD

[1]School of Computing and Information Systems, University of Melbourne, Melbourne, Australia

[2]Centre for Youth Mental Health, University of Melbourne, Melbourne, Australia

Corresponding Author:
Tianyi Zhang, MIT
School of Computing and Information Systems
University of Melbourne
Level 4, Melbourne Connect (Building 290)
Melbourne, 3010
Australia
Phone: 61 449930128
Email: tianyiz4@student.unimelb.edu.au

Abstract

Background: Smartphones have become an indispensable part of people's lives, and the fear of being without them, what has been termed "no mobile phone phobia" (nomophobia), is a growing phenomenon. The rise of problematic smartphone use highlights the urgent need to explore the intricate relationship between smartphones and human behavior. However, the connections between nomophobia, mental health indicators, smartphone use patterns, and daily activities remain largely underexplored.

Objective: This study aimed to explore the relationship between young adults with depression and smartphones and investigate nomophobia by analyzing data obtained from a pilot study of depression in a youth cohort. Exploring nomophobia can enhance our understanding of the dynamics between young adults and smartphone use, potentially empowering them to manage and regulate their smartphone use more effectively.

Methods: During an 8-week period, data collected via smartphone sensors, such as locations and screen status, were gathered from a cohort of 41 individuals diagnosed with major depressive disorder. In addition to passive-sensing smartphone data, the study collected ecological momentary assessments and psychometric measures, including the Nomophobia Questionnaire, which formed the basis of our investigation. We explored statistical associations among smartphone-derived behavioral features, psychometric indicators, and nomophobia. In addition, we used behavioral and psychometric data to develop regression models demonstrating the prediction of nomophobia levels.

Results: Our findings revealed that the level of nomophobia was positively associated with depression and negative affect, lower geolocation movements, and higher comfort with smartphone sensing. The exploratory predictive linear regression models demonstrated the feasibility of predicting an individual's Nomophobia Questionnaire score based on their smartphone sensing data. These models effectively used input features derived from both a combination of smartphone sensing data and psychometric measures and from smartphone sensing data alone.

Conclusions: Our work is the first to explore the relationship between nomophobia and smartphone sensor data. It provides valuable insights into the predictors of nomophobia level, contributing to the understanding of the relationship between smartphones and human behavior and paving the way for future studies.

(JMIR Form Res 2025;9:e57512) doi:10.2196/57512

KEYWORDS

mobile sensing; nomophobia; digital phenotyping; depression; mental health; smartphone use; personal sensing; behavior analysis; machine learning; mobile health; mobile phone

XSL•FO
RenderX

Introduction

Background

Nomophobia is a portmanteau of "no mobile phone phobia" and describes the fear of being without a smartphone. The concept of nomophobia has been proposed and researched in several early works [1,2], and a subsequent paper [3] presented a validated quantitative scale to measure it, known as the Nomophobia Questionnaire (NMP-Q). The rapid proliferation of smartphones has made nomophobia a timely phenomenon to investigate.

By July 2022, more than two-thirds of the world's population used mobile phones [4]. However, the popularity of smartphones has raised concerns about their impact on individuals' well-being. Smartphone addiction, especially among youth populations, has become a prominent issue. Approximately 95% of teenagers (aged 13-18 years) use cell phones every day in the United States [5]. It is claimed that overattachment to smartphones can lead to actual harmful outcomes such as depression [6], anxiety [7], and reduction in physical activity in the young population.

The notions of problematic smartphone use and smartphone addiction remain unsettled and need further exploration to determine their status as genuine diagnostic phenomena. The last decade has seen various attempts to quantify these notions using self-report scales [8]; however, more work is needed to better understand and predict nomophobia. One promising avenue is the analysis of nomophobia in terms of *digital footprints*, a term used to denote the data trails generated from human interactions with digital devices (eg, smartphones) and the internet (eg, social media). The idea of computationally mining digital footprints to infer an individual's (or a group's) behavior, preferences, and mental states has given rise to a field of research associated with the terms *digital phenotyping* [9], *personal sensing* [10], and *psychoinformatics* [11].

In this study, we investigated the relationship among nomophobia, smartphone sensor data, several mental health measures, and several debriefing questions concerning participant comfort with personal data sensing. This investigation was based on the collection of passive smartphone sensing data, ecological momentary assessment (EMA) responses, and various psychometric measures collected at baseline and the 8-week follow-up of a mental health digital phenotyping study of 41 young people diagnosed with major depressive disorder (MDD). We studied this information in relation to accompanying results on the NMP-Q.

Related Work

Research suggests that nomophobia has been a prevalent phenomenon across youth populations. According to previous investigations, up to 99% of students or young people have some degree of nomophobia [12-16], and a quarter of university students experience severe nomophobia [17,18]. This, in turn, adversely affects their mental or physical health and performance [19], even to the extent that there have been proposals to include nomophobia in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* [20].

As a nascent field, investigation into nomophobia is still emerging [19,21]. One study [22] used structural equation modeling and semantic network analysis to investigate the cognitive factors associated with nomophobia, showing that people with higher nomophobia tend to "perceive smartphones as their extended selves" [22] due to personal memories. Examining the relationships between nomophobia and psychological assessments, researchers have found various associations between nomophobia and social media use [23], loneliness [23-25], smartphone addiction [26], stress [25], problematic phone use [27], personality traits (Big Five personality traits) [28] and emotional stability [28], depression [29], anxiety [24,29], hyperactivity [29], social phobia [30], and oppositional symptoms [29]. One study [31] found that vertical collectivism, which involves sacrificing personal goals and submitting to authorities, is positively associated with nomophobia. Another study [32] took a qualitative approach and found a strong correlation between smartphone addiction and personal feelings such as social anxiety, antisocial behaviors, and loneliness.

In terms of nomophobia measurement, the basis of our work is the NMP-Q scale [3]. This validated scale consists of 4 dimensions: communication, connection, informativeness, and convenience. The 140-point total on the scale represents the sum of the responses to 20 questions, each rated on a Likert scale from 1 to 7. As per the scale developers, the total score can be categorized into 4 nomophobia severity levels: none and mild, moderate, and severe nomophobia. Since its release, this scale has been used in a handful of studies that affirm the reliance on mobile phones as an emerging issue for young people [16,32-34].

A recent study [35] investigated the NMP-Q scale in terms of various independent variables, particularly certain psychopathological measures, smartphone use, and lifestyle measurements. The study identified interpersonal sensitivity, obsession-compulsion, and the number of hours of smartphone use per day as predictors of nomophobia. While these results are of interest, the study was limited in terms of smartphone sensing. Only a few smartphone use–related inputs were obtained, such as the number of years using a smartphone, the number of daily hours using a smartphone, and the most used apps. Furthermore, the method to measure smartphone use was based on simply asking participants to subjectively provide estimates of these figures. While these results are of interest, the investigation of nomophobia against a greater variety of objectively collected smartphone sensing data is called for.

Thus, building upon this earlier work, in this study, we explored nomophobia outcomes against a range of objectively collected smartphone data inputs and psychometric scale results. Unlike work in digital phenotyping that uses smartphone data to infer psychopathology, our research on nomophobia uses objective phone data to directly characterize patterns of smartphone use.

XSL•FO
RenderX

Methods

Study Design

The data for this study were obtained from the SMARTSENSE-D project [36]. Participants were recruited from primary and specialized youth mental health services in Melbourne, Victoria, Australia, over a 2-year period between 2020 and 2021. Recruitment took longer than anticipated due to the COVID-19 pandemic situation in Melbourne. In total, 41 young people (aged 16-25 years) with a diagnosis of MDD consented to participate in data collection for 8 weeks.

With consent from their clinicians, we contacted potential participants who met the diagnosis criteria for MDD based on results from the Mini-International Neuropsychiatric Interview [37] and Patient Health Questionnaire–9 [38]. We included participants with an MDD diagnosis based on the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* [39], who regularly used a smartphone, understood smartphone data tracking, and were willing to participate. We excluded those at high risk of suicide, actively manic or actively psychotic, not familiar enough with smartphones to complete daily EMA surveys, or with insufficient ability to communicate in English.

Each participant provided demographic information, psychometric assessments (before and after the 8-week period), continuous smartphone sensing data (Android phones only), and data from a wrist actigraphy device during the 8-week trial period. We included in our investigation score results from the following measures and questionnaires that were administered in the study: (1) a demographic questionnaire (baseline assessment only) that captured demographic information such as age, gender, educational level, and other relevant background information; (2) the Depression, Anxiety, and Stress Scale–21 (DASS-21) [40,41], a set of 3 self-reported scales designed to measure the emotional states of depression, anxiety, and stress; (3) the Quick Inventory of Depressive Symptomatology (QIDS) [42,43], a brief self-report tool that assesses the severity of depressive symptoms; (4) the University of California, Los Angeles (UCLA), Loneliness Scale [44,45], which is designed to measure subjective feelings of loneliness and social isolation; (5) the Penn State Worry Questionnaire (PSWQ) [46,47], a self-report measure that assesses the tendency to worry excessively; (6) the Rumination Response Scale (RRS) [48,49], which assesses the extent to which individuals focus on their depressive symptoms and the possible causes and consequences of these symptoms; (7) the NMP-Q [3,50], which measures the severity of nomophobia, which is the fear of being without a mobile phone; and (8) a debriefing questionnaire (follow-up assessment only), which used a Likert scale from 1 (*strongly disagree*) to 7 (*strongly agree*) to capture participant attitudes toward such smartphone sensing and their perspectives on its feasibility and acceptability and also captured participant attitudes regarding each of the sensors being activated on their phones using a Likert scale from 1 (*extremely uncomfortable*) to 7 (*extremely comfortable*).

The AWARE-Light smartphone sensing app [51] was installed on each participant's Android phone and continuously collected geolocation, keyboard, communication (through calls and messages), screen status, app use, and touch operation data using sensors [52,53]. In addition, the AWARE-Light app administered EMA surveys twice daily at noon and 8 PM. The debriefing questionnaire items and EMA questions can be found in Multimedia Appendices 1 and 2, respectively.

Ethical Considerations

This study involved the collection of primary data directly from participants. All participants provided informed consent, which included detailed information about the purpose of the study, the types of data collected, how the data would be stored securely, and how they would be used. Participants were informed that their data would remain confidential and anonymized, and they had the option to withdraw from the study at any time without consequence. The study protocol was reviewed and approved by the University of Melbourne Human Research Ethics Committee (ID: 1955691.4). This study was carried out in accordance with the principles contained in the Declaration of Helsinki, the Australian National Health and Medical Research Council National Statement on Ethical Conduct in Human Research, and the National Health and Medical Research Council Australian Code for the Responsible Conduct of Research.

Participants were reimbursed Aus $100 (US $62.32) upon completion of the study, as well as an additional Aus $1.5 (US $0.93) for completion of each EMA survey (maximum of Aus $126 [US $78.53] if all surveys were completed).

Data Processing

Preprocessing

As participants installed the AWARE-Light sensing app at different times on their first study day, we discarded the data collected during the initial day of the 56-day data collection period. Therefore, the analysis was based on 55 full 24-hour days for each participant starting from the second day of installation.

We handled missing values in 2 steps. First, we only retained the sensor data for a participant if the data were recorded for at least 27 out of the 55 days; if a participant's sensor captured data for ≤26 days, then the data were excluded. The phone call and SMS text message sensors were exempt from this requirement as they may not be used every day. As such, we had different numbers of participants with valid data for each sensor dataset. Given this filtering, we ended up focusing on the smartphone sensor data of participants whose location and screen sensors met the threshold, resulting in a total of 27 valid sets of participant data.

Second, we used imputation to handle missing values in active data, including EMA questions and psychological assessments. In sum, <2% of the psychometric data were missing and imputed, and approximately 30% of EMA data were missing values. The days with missing EMA answers were excluded from feature calculations, such as the average and SD.

To resolve the missing data in the QIDS, we left the missing field as null to represent the missing data. Because the questions in these scales vary significantly, we believe that participants' attitude toward certain questions cannot be represented by the

XSL•FO
RenderX

ratings for other questions in the same scale. Note that the total value of the QIDS remained unchanged after resolving the missing data.

To resolve the missing data in assessments other than the QIDS, we first calculated the average ratings of nonmissing questions for each participant and then used the average score to replace the missing data. In this way, we assumed that participants had no special preference toward the questions they skipped.

For psychometric assessments, missing data arose from participants selecting "Prefer not to say," which we treated as missing data. This suggests that the missingness might be related to the participants' unwillingness to disclose certain information, implying that the data were missing not at random. Given that <2% of the data were missing, we opted for a straightforward imputation method. While this approach does not explicitly account for the potential missing-not-at-random nature of the data, the small proportion of missing data minimized the impact on our overall analysis. Nevertheless, we acknowledge that more sophisticated methods such as pattern-mixture models or sensitivity analysis could be considered in future studies to address any potential biases.

We verified the correlation and root mean square error (RMSE) values for the psychometric data and found that the lowest correlation between the original and imputed datasets was 0.89, whereas the highest correlation corresponded to an RMSE of 0, indicating a near-perfect match. Specifically, the RMSE values for the baseline DASS-21 total, baseline UCLA Loneliness Scale total, baseline PSWQ total, baseline RRS total, and follow-up RRS total were 0.61, 0.94, 2.15, 0.78, and 2.56, respectively. This indicates that the overall patterns remained robust and intact after imputation.

Feature Generation

The smartphone sensor data collected were used to generate a variety of information features. Message, keyboard, and screen status features were generated using the RAPIDS feature generation tool (AWARE Framework) [54], and the library by Doryab et al [55] was used to generate geolocation features such as average speed, maximum length staying at clusters, and number of frequent places [56]. Specifically, the Hierarchical Density-Based Spatial Clustering of Applications With Noise (HDBSCAN) [57] algorithm was used for cluster generation, effectively handling noise and clusters of varying densities.

We first generated daily smartphone sensor data and then averaged the data across the study period. For instance, to calculate the "maximum time spent at significant locations" feature, we first determined the maximum length of time that each participant stayed in various clusters per day, and then we computed the average duration over the 56 days for each participant.

In addition, we created our own Python scripts to work with certain sensors or generate features that were not available using the aforementioned tools, including touch operations, screen status, app use, and daily EMA questions. The fundamental features are the counts and average values of the interaction with the phone, such as counts of clicks and scrolling from touch operations, average and SD of EMA values, and the number of screen unlocks during each hour of the day.

The features derived from passively collected smartphone sensor data were preselected based on their relevance and potential to provide meaningful insights into user behavior and nomophobia. Location-based features offer detailed insights into the daily movement and habits of participants, encompassing various aspects of mobility, such as variability, speed, distance, and location diversity. Message traces and keyboard use for information retrieval or exchange indicate the extent of users' reliance on their smartphones. Screen use time and frequency, count of touch actions, and app use for particular app types not only reflect smartphone engagement but also serve as potential indicators of problematic smartphone use [27].

Regarding the psychometric measures investigated (DASS-21, QIDS, UCLA Loneliness Scale, PSWQ, RRS, and NMP-Q), apart from using the aggregate score of an assessment scale, in certain cases, we also investigated the subscores based on the dimensions or domains in each scale. For example, the four dimensions in the nomophobia scale [27] are (1) not being able to communicate (belonging and connectedness), (2) losing connectedness, (3) not being able to access information, and (4) giving up convenience. Similarly, several of the other psychometric assessments, such as the DASS-21, QIDS, and RRS (see the Models section), were also composed of different dimensions.

It is worth noting that we focused on analyzing the psychometric measures from the follow-up assessment for several reasons. First, the follow-up assessment results aligned better with the phone sensor data in terms of chronological order. The follow-up assessments were chronologically recorded after the passive sensing and could reveal more information about the entire study period. Second, the follow-up assessments from various questionnaires aligned better with each other as they were taken at the same time point. Third, the psychometric measures changed little throughout the 56-day study duration. The ratio of the absolute scale difference (calculated using the formula [after − before]/before) for the DASS-21, PSWQ, RRS, and QIDS was approximately 0.1, whereas the ratio for the UCLA Loneliness Scale was <0.05. This subtle difference between the baseline and follow-up absolute values indicates only minor variations across these scales, suggesting stability in the measured psychometric attributes over the study period.

Apart from the total scores for each psychometric measure, we explored the subfactors because they offered a more detailed and sensitive view of the data. While the total score provides a holistic perspective, the subfactors reveal detailed insights into different aspects of the psychometric constructs. These subfactors can highlight specific areas of change that may not be apparent in the overall scores, offering a more granular understanding. Therefore, we were interested in examining the correlation and predictive power of both the total scores and the subfactor scores of the psychometric measures.

Models

We used the Spearman rank correlation [58] to investigate the statistical associations between nomophobia scores and other

XSL•FO
RenderX

variables of interest. The Spearman correlation was chosen due to the nonnormal distribution of ordinal Likert-type questionnaire measures and continuous smartphone-collected features. We adjusted all *P* values using the original Benjamini-Hochberg method [59] to control the false discovery rate at 20%. This adjustment allows up to 20% of the rejected null hypotheses to be false positives and was chosen to ensure a reasonable balance between sensitivity and specificity in our exploratory analysis.

To test the correlations between smartphone sensor features and nomophobia, we only included a participant's dataset for a sensor if it had at least 27 out of the 55 days of availability (location and screen sensors were selected after this filtering process as they were the 2 sensors with sufficient data). Correlations between psychometric and debriefing questionnaire features and the NMP-Q were explored using valid data for 41 participants. Correlations between EMA features and nomophobia were explored using valid data for 85% (35/41) of the participants.

In total, we selected 48 correlation pairs to investigate, as shown in Textbox 1. The significant feature pairs are presented in the Results section.

We generated the smartphone sensor and psychometric features (Multimedia Appendix 3) shown in Textbox 2.

Textbox 1. Selected pairs of features for correlation investigation.

- A total of 9 location sensor features were each paired with the Nomophobia Questionnaire (NMP-Q) total score (Multimedia Appendix 3).
- A total of 3 screen sensor features were each paired with the NMP-Q total score (Multimedia Appendix 3).
- A total of 8 ecological momentary assessment (EMA) features were each paired with the NMP-Q total score. Each feature was obtained by averaging the morning and evening scores of an EMA affect question (Multimedia Appendix 3).
- A total of 2 debriefing questionnaire features—one based on the question concerning overall comfort with using AWARE-Light (statement: "I felt comfortable using the app") and the other based on the average of questions concerning comfort with the accelerometer, app, communication, location, light, keyboard, network, and screen sensors—were paired with the NMP-Q total score as well as the NMP-Q subdimension scores for a total combination of 10 pairs (Multimedia Appendix 3).
- A total of 4 Depression, Anxiety, and Stress Scale–21 (DASS-21) features were each paired with the NMP-Q total score (the DASS-21 total score, as well as the scores for the 3 subdimensions of depression, stress, and anxiety). Each subdimension or factor was added to the exploration given that they each represented separate mental health conditions. Furthermore, this investigation had a particular focus on depression given the cohort of young adults with depression (Multimedia Appendix 3).
- A total of 11 Quick Inventory of Depressive Symptomatology (QIDS) features were each paired with the NMP-Q total score (the QIDS total score, as well as the scores for the 10 subdimensions). As with the DASS-21, given the focus on depression, it was decided to investigate the subdimensions of this scale as well (Multimedia Appendix 3).
- One Rumination Response Scale total score feature was paired with the NMP-Q total score (Multimedia Appendix 3).
- One Penn State Worry Questionnaire total score feature was paired with the NMP-Q total score (Multimedia Appendix 3).
- One University of California, Los Angeles, Loneliness Scale total score feature was paired with the NMP-Q total (Multimedia Appendix 3).

XSL•FO
RenderX

Textbox 2. Smartphone sensor and psychometric features generated.

Location

- The RAPIDS tool (AWARE Framework) used Density-Based Spatial Clustering of Applications With Noise algorithms (HDBSCAN) to calculate location points into clusters. On the basis of the time spent in each location cluster, 8 features derived from location clusters were explored in correlation with the Nomophobia Questionnaire (NMP-Q) scale. In total, 5 of these features describe significant places (the minimum, maximum, average, and SD of the time spent per day at significant locations), as well as the number of visited significant places per day. The other 3 features include the average speed (indicating how fast a person moves) per day, total traveled distance per day, and time spent at home per day. There may be some overlap of information if the home is considered the most significant location for the person. The maximum period of staying at significant locations refers to the longest duration that a user stays at any identified significant location over a given period, identifying key locations that are important to the user, such as their home, workplace, or frequently visited places. Extended stays at certain locations can provide insights into routines and lifestyle patterns. The SD of the time spent in a cluster reflects the degree of variability in the time participants spend at different location clusters, with a higher SD indicating greater variation.

Screen

- A screen use episode is defined as the period from when the phone is unlocked until the screen is turned off. A total of 3 screen features were explored for correlation: the total duration of unlock episodes, considered as the time when smartphones are actively used; the longest duration of any unlock episode, representing the longest use episode per day; and the number of all unlock episodes, indicating how many times the person unlocked their smartphone in a day. For predictive models, 3 additional features were considered: the minimum, average, and SD of unlock episode duration in a day.

Messaging

- In total, 3 pairs of messaging features were explored for predictive models, including the total number of messages received from or sent to the contact who sent messages most frequently per day, the total number of messages received and sent by the participant per day, and the number of different contacts from whom the participant received and sent messages per day.

App use

- A total of 4 app use features were included in the candidate pool for the predictive model: the total number of times that social media apps were used by the participant per day, the total number of times that communication apps (eg, messaging and email) were used per day, the total number of times that entertainment apps (eg, games and videos) were used per day, and the total number of times that music or audio apps were used per day.

Touch actions

- One feature derived from collected smartphone touch data was the average number of items scrolled. This metric is calculated by tracking the number of distinct scroll actions a user performs on their smartphone screen during a specified period.

Psychometric subfactors

- Our study included both the total scores and subfactor scores of the following psychometrics: Depression, Anxiety, and Stress Scale–21 (DASS-21); Quick Inventory of Depressive Symptomatology (QIDS); University of California, Los Angeles (UCLA), Loneliness Scale; Penn State Worry Questionnaire (PSWQ); Rumination Response Scale (RRS); and NMP-Q. However, not all combinations across the subfactors were explored. We primarily focused on the total score on the NMP-Q for exploring correlations with smartphone sensor features and other psychometrics. In addition, for specific debriefing questions, we examined subfactors of participants' NMP-Q scores to understand why they felt comfortable with sensitive data being collected via smartphones. The DASS-21 is designed to be interpreted using 3 subscales: depression, anxiety, and stress. The QIDS also includes subfactors even though some subfactors are represented by a single question. In contrast, the RRS, PSWQ, and UCLA Loneliness Scale do not have subfactor representations and, thus, were not explored in this context.

In general, we investigated the correlation between each of these features and the NMP-Q total score. The one exception to this was the debriefing questionnaire features, which we also tested against the NMP-Q subdimensions because of our preassumption that using AWARE-Light and sensor activation might naturally correlate with certain aspects of nomophobia, such as the need for communication or convenience.

Apart from the correlation analysis, we also experimented with constructing some supervised machine learning models with the follow-up nomophobia scores as the outcome to be predicted. In our supervised learning models, we only included data from participants whose location and screen sensor data met the threshold of at least 27 out of the 55 days of availability (which resulted in 27/41, 66% of the participants included in the analysis). If a participant met this requirement for both sensors, then data for all their sensors, their psychometric results, and their EMA data were retained for regression.

After data cleaning, linear regression was applied to 47 extracted features (Multimedia Appendix 4) from 66% (27/41) of the participants, with filtering based on the combination of location and screen sensors. We focused on exploring two general models: (1) those whose input was only passive smartphone sensor features and (2) those whose input combined passive and active data, including EMA, phone sensors, and psychometrics. We implemented both wrapper methods and filter methods to test various combinations of features using R^2 to evaluate the models. Of the 47 preselected features, we included those that showed a potential improvement in model performance of >0.01 in terms of R^2 results. To predict NMP-Q scores, we used leave-one-out cross-validation for model

XSL•FO
RenderX

evaluation due to the small sample size. In addition to the R^2 score, we used mean absolute error (MAE), mean absolute percentage error (MAPE), and RMSE to assess the regression errors.

Results

Data Description

In total, 41 individuals took part in the study. Their demographic characteristics are presented in Table 1. The means and SDs of the psychometrics are shown in Table 2. The NMP-Q total scores of the participants only slightly changed between baseline and follow-up, as shown in Figure 1, with a Spearman test-retest reliability of 0.87 and a Cronbach α of 0.93. A total of 4 nomophobia severity levels as defined in the study by Yildirim and Correia [3] were used: no nomophobia (score of 0-20), mild nomophobia (score of 21-59), moderate nomophobia (score of 60-99), and severe nomophobia (score of 100-140). All participants (41/41, 100%) had an NMP-Q score of >20, with most participants (18/41, 44%) reporting a moderate level of nomophobia. A similar percentage of participants reported a mild (11/41, 27%) or severe (12/41, 29%) level of nomophobia.

Table 1. Demographic information of the young adults with depression involved in the 8-week longitudinal study (N=41).

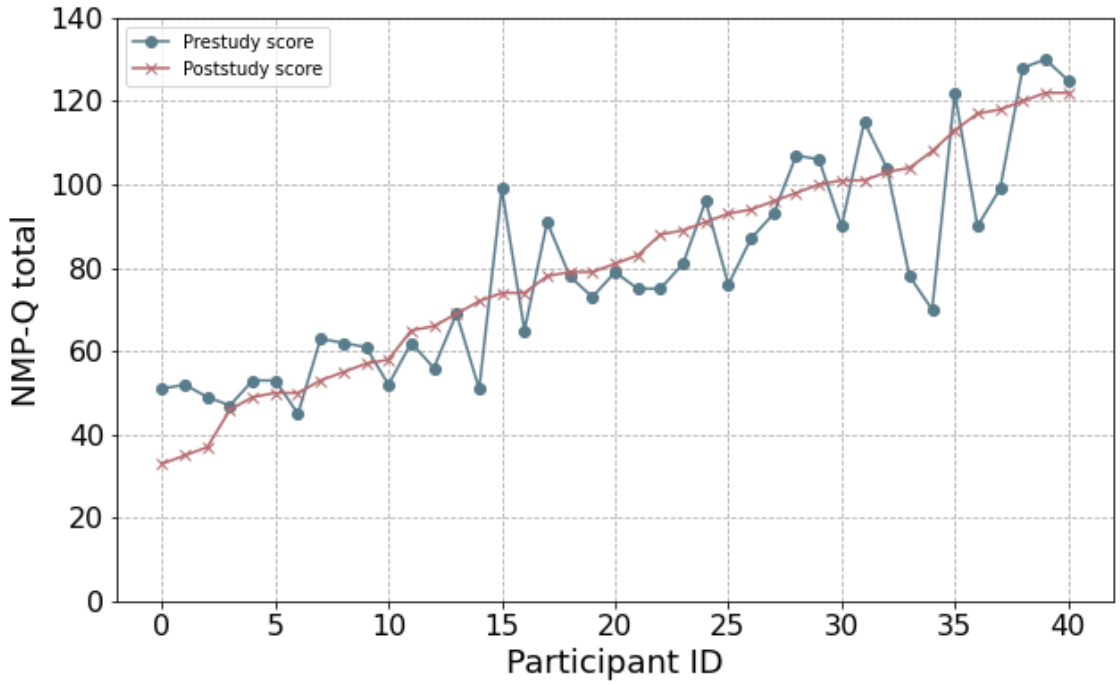
Characteristic and categories	Participants, n (%)
Sex at birth	
Female	29 (71)
Male	12 (29)
Age group (y)	
16-18	11 (27)
19-22	17 (41)
23-25	13 (32)
Country of birth	
Australia	35 (85)
Overseas (South America, Asia, Europe, and New Zealand)	6 (15)

Table 2. Means, SDs, and score ranges of the psychometric measurements taken at the end of the 8-week study among 41 participants for the Depression, Anxiety, and Stress Scale–21 (DASS-21); Penn State Worry Questionnaire (PSWQ); Rumination Response Scale (RRS); University of California, Los Angeles (UCLA), Loneliness Scale; and Quick Inventory of Depressive Symptomatology (QIDS).

Psychometric scale	Scores, mean (SD; range)
DASS-21	30.6 (12.9; 0-126)
PSWQ	59.1 (12.9; 16-80)
RRS	26.9 (5.5; 4-40)
UCLA Loneliness Scale	57.0 (5.0; 20-80)
QIDS	15.0 (5.5; 0-27)

XSL•FO
RenderX

Figure 1. Comparison of Nomophobia Questionnaire (NMP-Q) total scores among 41 participants with major depressive disorder at baseline and at the end of the 8-week longitudinal study period, revealing minimal differences.



Statistical Analysis

Exploratory data analysis was conducted using the Spearman rank correlation analyzing follow-up NMP-Q results against the psychometric, questionnaire, passive smartphone sensing, and EMA data. As detailed previously, we examined 48 feature pairs, with each pair involving the NMP-Q total score or that of one of its subfactors. Of these 48 tested combinations, we found 11 (23%) significant results that are presented and analyzed in this section, with the remaining 37 (77%) tests yielding nonsignificant results (Multimedia Appendix 3).

To qualify for inclusion as significant, the results needed to have $P \leq .05$ and a correlation coefficient of >0.3 or <−0.3, in line with the standard by Cohen [60], according to which results of >0.3 are of moderate effect.

Overall, significant correlations with the NMP-Q were found for certain EMA items, certain geolocation features, certain debriefing questions, and certain psychometric scores. The following is a descriptive summary of these correlation results.

Regarding location data (Table 3), there was a positive correlation between nomophobia (NMP-Q total score) and time spent at home, time spent at the most significant location, the maximum time spent in a cluster (at a daily level, this feature captures the maximum time that a participant spent at any of their significant locations [as determined by the geolocation clustering algorithm] for the day), and the SD of the time spent in a cluster (significant location; at a daily level, this feature captures the SD of all the durations of time that a participant spent at their significant locations for the day).

Regarding debriefing questions (Table 4), a positive correlation was found between the comfort level of a participant using the sensing app and their nomophobia in terms of the NMP-Q factor of giving up convenience, and a positive correlation was found between the comfort level of a participant having certain sensors tracking their personal information and their nomophobia in terms of the NMP-Q factor of giving up convenience.

Regarding the DASS-21 (Table 5), there was a positive correlation between the depression score and nomophobia (NMP-Q total score).

Regarding the QIDS (Table 5), mood and fatigue in the depressive cohort were positively correlated with nomophobia (NMP-Q total score).

Regarding EMA (Table 6), participants who felt more relaxed and happier daily reported lower nomophobia (NMP-Q total score).

XSL•FO
RenderX

Table 3. Significant correlations between Nomophobia Questionnaire total scores and smartphone-based location sensor features among 27 participants with major depressive disorder during the 8-week study period.

Location feature	*P* value	Spearman correlation coefficient
The maximum period staying at significant locations	.02	0.35
The SD of the time spent in a cluster	.04	0.32
Time spent at home	.02	0.35
Time spent at the most significant location	.03	0.35

Table 4. Significant correlations between the "giving up convenience" subfactor of the Nomophobia Questionnaire and participants' responses to debriefing questions at the conclusion of the 8-week study period observed among 41 participants with major depressive disorder.

Feature	*P* value	Spearman correlation coefficient
Participant comfort with using AWARE-Light (item 4 in Multimedia Appendix 1)	.02	0.37
Participant comfort with sensor activation (an average of items 21-28 in Multimedia Appendix 1)	.01	0.40

Table 5. Significant correlations between Nomophobia Questionnaire total scores and various psychometric measures assessed among 41 participants with major depressive disorder at the conclusion of the 8-week study period.

Psychometric measure	*P* value	Spearman correlation coefficient
DASS-21[a] factor—depression	.02	0.38
QIDS[b] factor—mood	.046	0.31
QIDS factor—energy and fatigue	.02	0.37

[a]DASS-21: Depression, Anxiety, and Stress Scale–21.

[b]QIDS: Quick Inventory of Depressive Symptomatology.

Table 6. Significant correlations between Nomophobia Questionnaire total scores and responses to ecological momentary assessment (EMA) questions captured over the 8-week study period among 41 participants with major depressive disorder.

EMA measure	*P* value	Spearman correlation coefficient
Average of morning and evening EMA score for how happy participants felt	.03	–0.34
Average of morning and evening EMA score for how relaxed participants felt	.03	–0.337

Predicting Nomophobia

Tables 7 and 8 show the attributes of our best models for passively plus actively collected data and passively collected data only, respectively. Our best model achieved an R^2 value of 0.779 (MAE=9.42; MAPE=13.56%; RMSE=12.09). Figure 2 shows the distribution of predicted and actual NMP-Q scores based on both phone sensor features and psychometric measures input into a linear regression model. We also developed a model to predict individuals' NMP-Q scores using only smartphone sensor data. Our best model achieved an R^2 of 0.732 (MAE=10.88; MAPE=14.44%; RMSE=13.32), as illustrated in Figure 3.

Table 7. Overview of smartphone sensor data and survey features included in the linear regression model to predict Nomophobia Questionnaire total scores among 27 participants with depressive symptoms over the 8-week study period, reaching an R^2 value of 0.779.

Rank	Feature	Sensor or questionnaire	Attribute coefficient
1	Count of social media app use	App	−17.55
2	Sleep disturbance	QIDS[a]	−4.40
3	Count of sent messages to the most frequent contact	Message	−1.5
4	The total traveled distance	Location	−1.32
5	SD of unlock episode duration	Screen	0.81
6	The maximum period staying at significant locations	Location	0.03

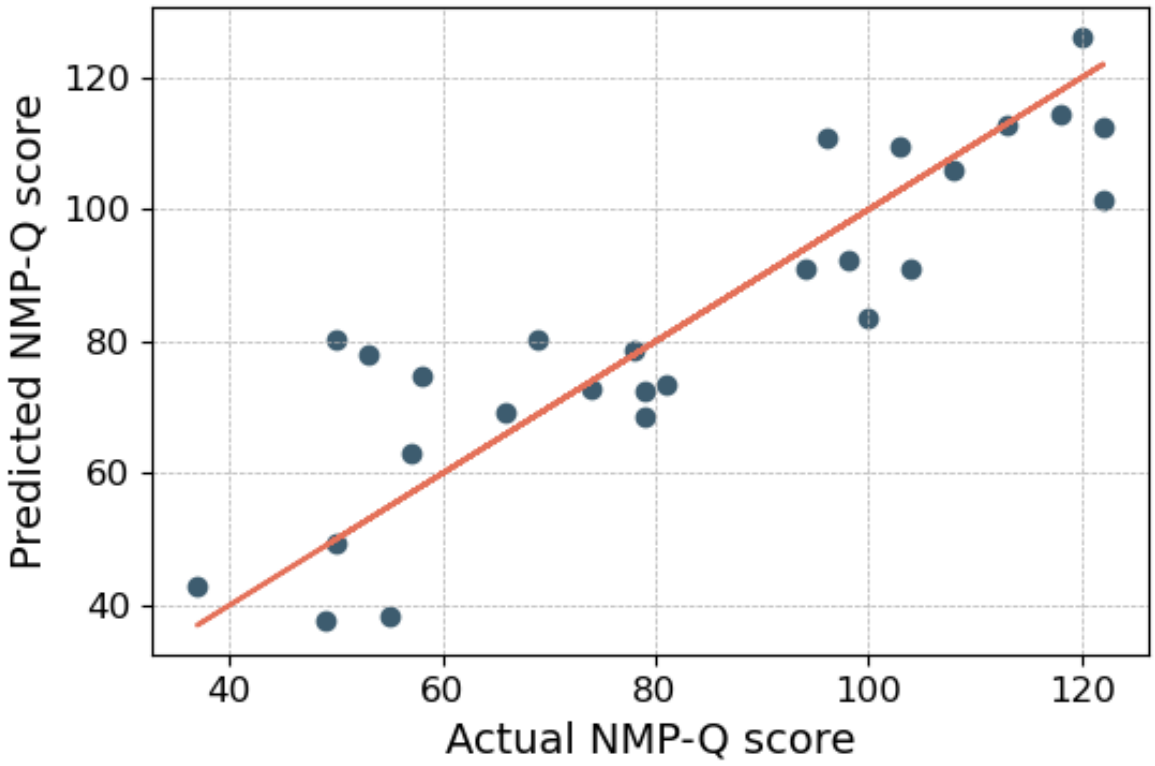
[a]QIDS: Quick Inventory of Depressive Symptomatology.

XSL•FO
RenderX

Table 8. Overview of smartphone sensor features included in the linear regression model to predict Nomophobia Questionnaire total scores among 27 participants with depressive symptoms over the 8-week study period, reaching an R^2 value of 0.732.

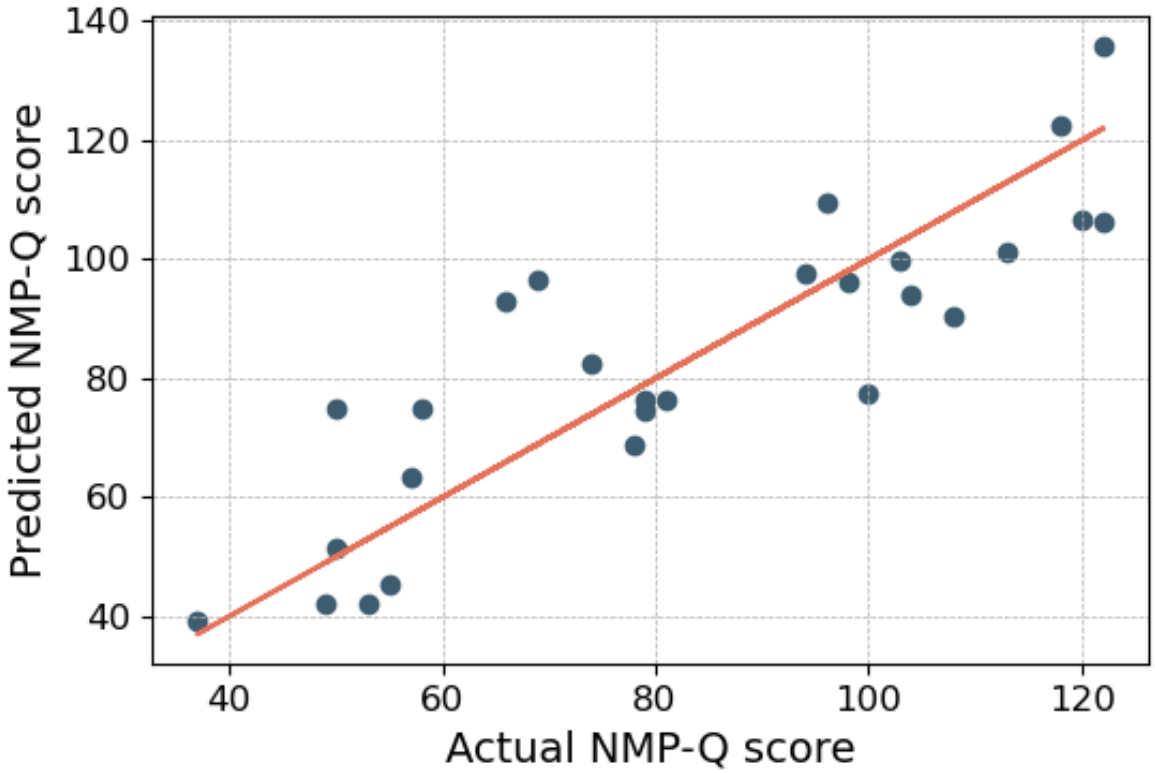
Rank	Feature	Sensor or questionnaire	Attribute coefficient
1	Count of social media app use	App	−20.59
2	Count of sent messages to the most frequent contact	Message	−2.23
3	SD of unlock episode duration	Screen	0.81
4	The maximum period of staying at significant locations	Location	0.02
5	The area covered by a participant	Location	–0.0005
6	The total traveled distance	Location	–0.00000218
7	The average amount of items scrolled	Touch	0.00000158

Figure 2. Comparison of actual versus predicted Nomophobia Questionnaire (NMP-Q) total scores using a combination of psychometric measures and smartphone sensor data from 27 participants with depressive symptoms who had valid data during the 8-week study period. This highlights the potential of machine learning for accurate predictions.



XSL•FO
RenderX

Figure 3. Comparison of actual versus predicted Nomophobia Questionnaire (NMP-Q) total scores using only smartphone sensor data from 27 participants with depressive symptoms who had valid data during the 8-week study period without incorporating psychometric measures. This demonstrates the feasibility of machine learning–based predictions.



Both models' performance was mostly related to the frequency of social media app use. The count of sent messages to the most frequent contact, the duration of staying at clusters, and the SD of the duration of screen unlock episodes were also important features revealing smartphone use and location information in both models.

Discussion

Themes

The results from our exploratory study are suggestive of several themes concerning nomophobia and its associations with both psychopathology and smartphone use.

Depression and Negative or Positive Affect

Given that the SMARTSENSE-D study focused on a cohort of young people diagnosed with clinical depression and varying degrees of symptom severity, the data provided a good opportunity to explore associations between nomophobia and depression. The 2 psychometric scales concerning depression that were used in the study, the DASS-21 and QIDS, both showed significant associations with nomophobia. The DASS-21 depression scale and 2 QIDS factors positively correlated with total nomophobia scores.

Our EMA results across the study indicate that the more positive affect one has in general, the lower their nomophobia. The most salient of these results concern happiness and relaxation, for which the EMA results negatively correlated with total nomophobia scores.

These findings regarding associations among depression, affect, and nomophobia are in line with previous findings [61-67]. Given the co-occurrence of depression and nomophobia, there is a possibility that depressive symptoms and negative affect are influenced by nomophobia, or vice versa. The connection and causality among depression, negative feelings, and nomophobia need to be further validated through additional research.

Location

Our results suggest that participants who spent more time at home or at their otherwise most visited location had greater nomophobia. Relatedly, one's maximum duration at a significant location also positively correlated with nomophobia. This implies that the more an individual is confined to their home or the more concentrated their geolocation to a particular location, the greater their sense of dependency on their smartphone.

Our results also showed a positive correlation between nomophobia and the SD of all the durations of time that a participant spent at their significant locations for the day. This is a result that is perhaps less straightforward to interpret; we can say that, if one has a lower SD, then there is a more evenly spread diversity in the time they spend at their various significant locations. This contrasts with a higher SD, which implies a concentration in fewer places. In this case, the SD

XSL•FO
RenderX

result is in line with the other geolocation results discussed previously.

The fact that nomophobia was associated both with geolocation and depressive symptoms is in line with previous research that has demonstrated a significant link between movement patterns and depression [56,68-70]. Thus, a triangle emerges that connects geolocation, depression, and nomophobia. Ultimately, limited or confined geolocation activity is associated with nomophobia and can be an indicator of depression.

Nomophobia-Privacy Trade-Off

A trend of positive correlations emerged between nomophobia and comfort with being tracked via a smartphone sensing app and sharing data. The NMP-Q "giving up convenience" dimension positively correlated with comfort levels with using the AWARE-Light sensing app and activating sensors within it.

Thus, a trade-off seems to emerge whereby individuals with nomophobic tendencies, particularly their concern of giving up the convenience afforded by smartphones, are less concerned about being smartphone sensed. Alternatively, one could perhaps frame this as being about privacy concerns outweighing certain smartphone benefits and that those more concerned about privacy are less inclined to develop nomophobia.

At any rate, these findings relate to those of previous studies that have shown how privacy concerns can influence smartphone use behaviors as a dependent indicator [71,72]. Our results from the self-reported evaluations of both participants' nomophobia levels and attitudes toward privacy further support this correlation between smartphone use and privacy concerns.

Smartphone Use

Previous studies [27] have indicated a positive correlation between nomophobia and problematic smartphone use, including screen status (eg, screen locks and unlocks and turning the screen on or off). While none of our formal results involving screen sensor features were significant, we did observe using a scatter plot that, when limited to participants in the mild and severe NMP-Q categories (moderate category excluded), there was a noticeable positive correlation between the number of phone unlocks and NMP-Q scores. However, further work is required to formally strengthen the case that nomophobia is associated with greater screen use, such as unlock instances.

Predictive Models

The predictive results shown by our linear regression models indicate the feasibility of predicting the NMP-Q score based on passive human behavioral features. The implications of these results can be valuable for the understanding, detection, and control of nomophobia.

The most important feature in both models, the count of social media app use, showed a negative relationship with the predicted ground-truth NMP-Q score, indicating that engagement with social media apps co-occurs with a lower level of nomophobia. Meanwhile, the count of sent messages to the most frequent contact and the total traveled distance demonstrated different levels of negative association with the NMP-Q. These findings should be highlighted for further exploration regarding the role of social media apps in managing the relationship between users and smartphones.

Limitations

Our primary goal was to understand the relationship between the features and the NMP-Q, demonstrating the feasibility of predicting NMP-Q scores based on smartphone sensor data and psychometrics or even using only passively collected data. However, we acknowledge that more complex models such as Extreme Gradient Boosting or support vector machines might be valuable in future research with more data to explore potential nonlinear relationships within the data. These advanced models could offer deeper insights and improve prediction accuracy by capturing intricate patterns that simpler linear models might miss.

This study was exploratory and aimed to gain preliminary insights into associations between smartphone-inferred behavioral features and nomophobia (NMP-Q score), as well as the potential of machine learning models to predict NMP-Q scores from such features. The primary intent was not to provide a model ready for practical application at this stage but to demonstrate the feasibility of model use in future stages with more extensive data over a longer time frame. Predictive models for NMP-Q score could assist researchers, patients with depression, and mental health professionals in understanding the impact of smartphone use and daily behaviors on mental health from a technological perspective.

Our inferential statistical results provide a starting point for more dedicated studies with larger samples sizes to strengthen the claims suggested by our results. The use of a 20% false discovery rate may have introduced potential false positives, especially in a small cohort of participants. This threshold aimed to balance false positives and false negatives. However, to achieve more robust and precise conclusions, the findings need further confirmation with larger sample sizes. The larger datasets would also incorporate more advanced machine learning techniques and be less prone to issues such as overfitting. With more data and refined models, clinicians could better understand depression and its relationship with smartphone use, aiding those with depressive symptoms. This could offer valuable insights into how these digital interactions influence their lives and aid in better mental health management.

In addition to the small sample size, another limitation was the imperfection of our data collection given technical issues encountered, issues with AWARE-Light in particular, and the general difficulty of getting smartphone sensing data collection to work across various sensors and various participant phones. Thus, we had to implement a data quality filtering procedure to ensure that we only used a phone's data for a given sensor if those data met certain threshold requirements. Consequently, not all participants were included for every sensor.

Conclusions

Our study explored associations among nomophobia, clinical measures, EMAs, smartphone sensor data, and smartphone use questionnaires. We discovered that greater nomophobia is linked to more depressive symptoms, limited geolocation movement

XSL•FO
RenderX

and (domestic) stationariness, comfort with being smartphone sensed, and negative emotions. In doing so, we demonstrated that smartphone sensing can effectively quantify user behaviors and perceptions underlying nomophobia and the phenomena that its measures are designed to capture.

Nomophobia is a nascent notion that merits further investigation. Beyond conceptualizations and measurement scales, use of the technology in question (ie, smartphones) to provide objective quantification of nomophobic-related use and behavioral phenomena requires further exploration through larger and more in-depth purpose-built studies. It is hoped that these initial exploratory findings will contribute to paving the way for such future work.

Acknowledgments

The authors would like to thank the headspace National Youth Mental Health Foundation for supporting the recruitment process and those who volunteered to participate in the SMARTSENSE-D study. They would also like to thank the University of Melbourne, the Cisco University Research Program Fund, and the Silicon Valley Community Foundation for their funding support. This study was funded by the University of Melbourne (Faculty of Medicine, Dentistry, and Health Sciences establishment grant awarded to LS) and by a grant from the Cisco University Research Program Fund and the Silicon Valley Community Foundation awarded to SD (ID 2021-232746 3696). The authors acknowledge the use of generative artificial intelligence solely for language refinement, with no involvement in the research, methodologies, or coding presented in this work.

Data Availability

The datasets generated or analyzed during this study are not publicly available because not all participants in this study provided consent for their data to be openly shared. For the remaining participants, the data are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questions presented in the debriefing questionnaire administered at the end of the 8-week study. Items 1 to 20 are provided in scales from 1 to 7 indicating the level of agreement. Items 21 to 28 are provided in scales from 1 to 7 indicating the comfort level of the participants with the sensor being activated on their smartphones. Items 29 and 30 are provided as Yes or No options.
[DOCX File , 23 KB - formative_v9i1e57512_app1.docx]

Multimedia Appendix 2
Daily ecological momentary assessment items for 41 participants with major depressive disorder. The 21-question survey, administered at noon and 8 PM, used 7-point Likert-scale items (questions 1-14 and 18), binary (yes or no) items (questions 15 and 19), and open-text responses (questions 16, 17, 20, and 21).
[DOCX File , 23 KB - formative_v9i1e57512_app2.docx]

Multimedia Appendix 3
Comprehensive correlation analysis results between Nomophobia Questionnaire total scores and various study variables, including psychometric measures assessed at the end of the study and smartphone sensing features recorded throughout the study. The analysis includes 41 participants for the psychometric data and 27 participants for variables involving smartphone sensor features, including both significant and nonsignificant associations.
[DOCX File , 26 KB - formative_v9i1e57512_app3.docx]

Multimedia Appendix 4
Preselected features used in predictive models to estimate Nomophobia Questionnaire total scores for 27 participants with major depressive disorder. These features include psychometric measures collected at the end of the 8-week study and smartphone-derived variables recorded throughout the study.
[DOCX File , 26 KB - formative_v9i1e57512_app4.docx]

References

1. Nomophobia is the fear of being out of mobile phone contact - and it's the plague of our 24/7 age. The Standard. 2012 Apr 13. URL: https://www.standard.co.uk/hp/front/nomophobia-is-the-fear-of-being-out-of-mobile-phone-contact-and-its-the-plague-of-our-24-7-age-6634478.html [accessed 2025-02-07]

XSL•FO
RenderX

2. King AL, Valença AM, Silva AC, Baczynski T, Carvalho MR, Nardi AE. Nomophobia: dependency on virtual environments or social phobia? Comput Hum Behav 2013 Jan;29(1):140-144. [doi: 10.1016/j.chb.2012.07.025]
3. Yildirim C, Correia AP. Exploring the dimensions of nomophobia: development and validation of a self-reported questionnaire. Comput Hum Behav 2015 Aug;49:130-137. [doi: 10.1016/j.chb.2015.02.059]
4. Digital around the world. Data Reportal. URL: https://datareportal.com/global-digital-overview [accessed 2025-02-07]
5. Rideout V, Peebles A, Mann S, Robb MB. Common sense census: media use by tweens and teens, 2021. Common Sense. 2022. URL: https://calio.dspacedirect.org/items/e760d928-dea3-4b81-901b-5bd48cb25f7d [accessed 2025-02-07]
6. Xie H, Tao S, Zhang Y, Tao F, Wu X. Impact of problematic mobile phone use and insufficient physical activity on depression symptoms: a college-based follow-up study. BMC Public Health 2019 Dec 05;19(1):1640. [doi: 10.1186/s12889-019-7873-z] [Medline: 31805915]
7. Coyne SM, Stockdale L, Summers K. Problematic cell phone use, depression, anxiety, and self-regulation: evidence from a three year longitudinal study from adolescence to emerging adulthood. Comput Hum Behav 2019 Jul;96:78-84. [doi: 10.1016/j.chb.2019.02.014]
8. Harris B, Regan T, Schueler J, Fields SA. Problematic mobile phone and smartphone use scales: a systematic review. Front Psychol 2020 May 5;11:672 [FREE Full text] [doi: 10.3389/fpsyg.2020.00672] [Medline: 32431636]
9. Torous J, Kiang MV, Lorme J, Onnela JP. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Ment Health 2016 May 05;3(2):e5165 [FREE Full text] [doi: 10.2196/mental.5165] [Medline: 27150677]
10. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 08;13:23-47 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
11. Montag C, Duke É, Markowetz A. Toward psychoinformatics: computer science meets psychology. Comput Math Methods Med 2016;2016:2983685 [FREE Full text] [doi: 10.1155/2016/2983685] [Medline: 27403204]
12. Qutishat M, Rathinasamy Lazarus E, Razmy AM, Packianathan S. University students' nomophobia prevalence, sociodemographic factors and relationship with academic performance at a University in Oman. Int J Afr Nurs Sci 2020;13:100206. [doi: 10.1016/j.ijans.2020.100206]
13. Dalbudak I, Yilmaz T, Yigit S. Nomophobia levels and personalities of university students. J Educ Learn 2020;9(2):166-177. [doi: 10.5539/jel.v9n2p166]
14. Gurbuz IB, Ozkan G. What is your level of nomophobia? An investigation of prevalence and level of nomophobia among young people in Turkey. Community Ment Health J 2020 Jul;56(5):814-822. [doi: 10.1007/s10597-019-00541-2] [Medline: 31902048]
15. Bragazzi NL, Re TS, Zerbetto R. The relationship between nomophobia and maladaptive coping styles in a sample of Italian young adults: insights and implications from a cross-sectional study. JMIR Ment Health 2019 Apr 24;6(4):e13154 [FREE Full text] [doi: 10.2196/13154] [Medline: 30763254]
16. Farooqui IA, Pore P, Gothankar J. Nomophobia: an emerging issue in medical institutions? J Ment Health 2018 Oct;27(5):438-441. [doi: 10.1080/09638237.2017.1417564] [Medline: 29271270]
17. Humood A, Altooq N, Altamimi A, Almoosawi H, Alzafiri M, Bragazzi NL, et al. The prevalence of nomophobia by population and by research tool: a systematic review, meta-analysis, and meta-regression. Psych 2021 Jun 21;3(2):249-258. [doi: 10.3390/psych3020019]
18. Jilisha G, Venkatachalam J, Menon V, Olickal JJ. Nomophobia: a mixed-methods study on prevalence, associated factors, and perception among college students in Puducherry, India. Indian J Psychol Med 2019 Nov 01;41(6):541-548. [doi: 10.4103/ijpsym.ijpsym_130_19]
19. Rodríguez-García AM, Moreno-Guerrero AJ, López Belmonte J. Nomophobia: an individual's growing fear of being without a smartphone-a systematic literature review. Int J Environ Res Public Health 2020 Jan 16;17(2):580 [FREE Full text] [doi: 10.3390/ijerph17020580] [Medline: 31963208]
20. Bragazzi NL, Del Puente G. A proposal for including nomophobia in the new DSM-V. Psychol Res Behav Manag 2014 May 16;7:155-160 [FREE Full text] [doi: 10.2147/PRBM.S41386] [Medline: 24876797]
21. Bhattacharya S, Bashar MA, Srivastava A, Singh A. NOMOPHOBIA: NO MObile PHone PhoBIA. J Family Med Prim Care 2019 Apr;8(4):1297-1300 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_71_19] [Medline: 31143710]
22. Han S, Kim KJ, Kim JH. Understanding nomophobia: structural equation modeling and semantic network analysis of smartphone separation anxiety. Cyberpsychol Behav Soc Netw 2017 Jul;20(7):419-427. [doi: 10.1089/cyber.2017.0113] [Medline: 28650222]
23. Yıldız Durak H. What would you do without your smartphone? Adolescents' social media usage, locus of control, and loneliness as a predictor of nomophobia. Addicta 2018 Sep 1;5(3):543-557. [doi: 10.15805/addicta.2018.5.2.0025]
24. Kara M, Baytemir K, Inceman-Kara F. Duration of daily smartphone usage as an antecedent of nomophobia: exploring multiple mediation of loneliness and anxiety. Behav Inf Technol 2019 Oct 09;40(1):85-98. [doi: 10.1080/0144929x.2019.1673485]
25. Dahiya P. Nomophobia and its effects on stress and loneliness among young adults. Int J Indian Psychol 2021;9(3):993-999. [doi: 10.25215/0903.091]

XSL•FO
RenderX

26. Yildiz Durak H. Investigation of nomophobia and smartphone addiction predictors among adolescents in Turkey: demographic variables and academic performance. Soc Sci J 2019 Dec 23;56(4):492-517. [doi: 10.1016/j.soscij.2018.09.003]
27. Kaviani F, Robards B, Young KL, Koppel S. Nomophobia: is the fear of being without a smartphone associated with problematic use? Int J Environ Res Public Health 2020 Aug 19;17(17):6024 [FREE Full text] [doi: 10.3390/ijerph17176024] [Medline: 32824979]
28. Argumosa-Villar L, Boada-Grau J, Vigil-Colet A. Exploratory investigation of theoretical predictors of nomophobia using the Mobile Phone Involvement Questionnaire (MPIQ). J Adolesc 2017 Apr;56:127-135. [doi: 10.1016/j.adolescence.2017.02.003] [Medline: 28231484]
29. Moreno-Guerrero AJ, Aznar-Díaz I, Cáceres-Reche P, Rodríguez-García AM. Do age, gender and poor diet influence the higher prevalence of nomophobia among young people? Int J Environ Res Public Health 2020 May 24;17(10):3697 [FREE Full text] [doi: 10.3390/ijerph17103697] [Medline: 32456304]
30. Uysal S, Özen H, Madenoğlu C. Social phobia in higher education: the influence of nomophobia on social phobia. Global eLearn J 2016;5(2):1-8.
31. Arpaci I. Culture and nomophobia: the role of vertical versus horizontal collectivism in predicting nomophobia. Inf Dev 2017 Sep 05;35(1):96-106. [doi: 10.1177/0266666917730119]
32. Anshari M, Alas Y, Sulaiman E. Smartphone addictions and nomophobia among youth. Vulnerable Child Youth Stud 2019 May 06;14(3):242-247. [doi: 10.1080/17450128.2019.1614709]
33. Yildirim C, Sumuer E, Adnan M, Yildirim S. A growing fear: prevalence of nomophobia among Turkish college students. Inf Dev 2016 Jul 09;32(5):1322-1331. [doi: 10.1177/0266666915599025]
34. MB P, Madhukumar S, TS MM. A study on nomophobia - mobile phone dependence, among students of a medical college in Bangalore. Nat J Community Med 2015 Sep 30;6(03):340-344.
35. Gonçalves S, Dias P, Correia AP. Nomophobia and lifestyle: smartphone use and its relationship to psychopathologies. Comput Hum Behav Rep 2020;2:100025. [doi: 10.1016/j.chbr.2020.100025]
36. D'Alfonso S, Zhang T, Camargo A, Tagliaferri S, Davies P, Kostakos V, et al. SmartSense-D. Open Science Framework Home. 2024. URL: https://osf.io/z4mj8 [accessed 2025-02-07]
37. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33;quiz 34-57. [Medline: 9881538]
38. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
39. Guha M. Diagnostic and statistical manual of mental disorders: DSM-5 (5th edition). Ref Rev 2014;28(3):36-37 [FREE Full text] [doi: 10.1108/RR-10-2013-0256]
40. Henry JD, Crawford JR. The short‐form version of the Depression Anxiety Stress Scales (DASS‐21): construct validity and normative data in a large non‐clinical sample. Br J Clinic Psychol 2011 Jan 13;44(2):227-239. [doi: 10.1348/014466505x29657]
41. Osman A, Wong JL, Bagge CL, Freedenthal S, Gutierrez PM, Lozano G. The Depression Anxiety Stress Scales-21 (DASS-21): further examination of dimensions, scale reliability, and correlates. J Clin Psychol 2012 Dec;68(12):1322-1338. [doi: 10.1002/jclp.21908] [Medline: 22930477]
42. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol Psychiatry 2003 Sep 01;54(5):573-583. [doi: 10.1016/s0006-3223(02)01866-8] [Medline: 12946886]
43. Reilly TJ, MacGillivray SA, Reid IC, Cameron IM. Psychometric properties of the 16-item Quick Inventory of Depressive Symptomatology: a systematic review and meta-analysis. J Psychiatr Res 2015 Jan;60:132-140. [doi: 10.1016/j.jpsychires.2014.09.008] [Medline: 25300442]
44. Russell DW. UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure. J Pers Assess 1996 Feb;66(1):20-40. [doi: 10.1207/s15327752jpa6601_2] [Medline: 8576833]
45. Russell D, Peplau LA, Cutrona CE. The revised UCLA Loneliness Scale: concurrent and discriminant validity evidence. J Pers Soc Psychol 1980 Sep;39(3):472-480. [doi: 10.1037//0022-3514.39.3.472] [Medline: 7431205]
46. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State Worry Questionnaire. Behav Res Ther 1990;28(6):487-495. [doi: 10.1016/0005-7967(90)90135-6] [Medline: 2076086]
47. Davey GC, Wells A. The Penn State Worry Questionnaire (PSWQ). In: Worry and its Psychological Disorders: Theory, Assessment and Treatment. Hoboken, NJ: John Wiley & Sons; Jan 01, 2006:99-119.
48. Treynor W, Gonzalez R, Nolen-Hoeksema S. Rumination reconsidered: a psychometric analysis. Cogn Ther Res 2003;27:247-259. [doi: 10.1023/A:1023910315561]
49. Quick BL, Stephenson MT. The Reactance Restoration Scale (RRS): a measure of direct and indirect restoration. Commun Res Rep 2007 May 11;24(2):131-138. [doi: 10.1080/08824090701304840]

XSL•FO
RenderX

50. Jahrami H, Saif Z, Trabelsi K, Bragazzi NL, Vitiello MV. Internal consistency and structural validity of the nomophobia questionnaire (NMP-Q) and its translations: a systematic review with meta-analysis. Heliyon 2023 Apr 14;9(4):e15464 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e15464] [Medline: 37128316]
51. van Berkel N, D'Alfonso S, Kurnia Susanto R, Ferreira D, Kostakos V. AWARE-Light: a smartphone tool for experience sampling and digital phenotyping. Pers Ubiquitous Comput 2022 Nov 05;27:435-445. [doi: 10.1007/s00779-022-01697-7]
52. AWARE-Light homepage. AWARE-Light. URL: https://www.aware-light.org/ [accessed 2025-02-09]
53. Sensors - AWARE. AWARE. URL: https://awareframework.com/sensors/ [accessed 2025-02-09]
54. Vega J, Li M, Aguillera K, Goel N, Joshi E, Khandekar K, et al. Reproducible analysis pipeline for data streams: open-source software to process data collected with mobile devices. Front Digit Health 2021 Nov 18;3:769823 [FREE Full text] [doi: 10.3389/fdgth.2021.769823] [Medline: 34870271]
55. Doryab A, Chikarsel P, Liu X, Dey AK. Extraction of behavioral features from smartphone and wearable data. arXiv Preprint posted online on December 18, 2018 [FREE Full text] [doi: 10.48550/arXiv.1812.10394]
56. Canzian L, Musolesi M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp '15; September 7-11, 2015; Osaka, Japan. [doi: 10.1145/2750858.2805845]
57. McInnes L, Healy J, Astels S. hdbscan: hierarchical density based clustering. J Open Source Softw 2017 Mar 21;2(11):205. [doi: 10.21105/joss.00205]
58. Zar JH. Spearman rank correlation. In: Encyclopedia of Biostatistics. Hoboken, NJ: John Wiley & Sons; Jul 15, 2005.
59. Thissen D, Steinberg L, Kuang D. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. J Educ Behav Stat 2002;27(1):77-83. [doi: 10.3102/10769986027001077]
60. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York, NY: Routledge; May 13, 2013.
61. Sharma M, Amandeep, Mathur DM, Jeenger J. Nomophobia and its relationship with depression, anxiety, and quality of life in adolescents. Ind Psychiatry J 2019;28(2):231-236 [FREE Full text] [doi: 10.4103/ipj.ipj_60_18] [Medline: 33223716]
62. Kubrusly M, de Barros Silva PG, de Vasconcelos GV, Leite ED, de Almeida Santos P, Rocha HA. Nomophobia among medical students and its association with depression, anxiety, stress and academic performance. Rev Bras Educ Med 2021 Jul;45(3):e162. [doi: 10.1590/1981-5271v45.3-20200493.ing]
63. V S, Subramanian A, Satyanarayan S, Subramanian S, S B, N B, et al. Prevalence of nomophobia and its association with stress, anxiety and depression among students. Biomedicine 2021 Jan;40(4):522-525. [doi: 10.51248/.v40i4.333]
64. Bano N, Khan MA, Asif U, Beer J, Rawass H. Effects of nomophobia on anxiety, stress and depression among Saudi medical students in Jeddah, Saudi Arabia. J Pak Med Assoc 2021 Mar;71(3):854-858 [FREE Full text] [doi: 10.47391/JPMA.983] [Medline: 34057935]
65. Kuscu TD, Gumustas F, Rodopman Arman A, Goksu M. The relationship between nomophobia and psychiatric symptoms in adolescents. Int J Psychiatry Clin Pract 2021 Mar;25(1):56-61. [doi: 10.1080/13651501.2020.1819334] [Medline: 32940094]
66. Çakmak Tolan Ö, Karahan S. The relationship between nomophobia and depression, anxiety and stress levels of university students. Int J Psychol Educ Stud 2022 Mar 06;9(1):115-129. [doi: 10.52380/ijpes.2022.9.1.559]
67. Kateb SA. The prevalence and psychological symptoms of nomophobia among university students. J Res Curr Instruct Educ Technol 2017 Jul;3(3):155-182. [doi: 10.21608/JRCIET.2017.24454]
68. Raugh IM, James SH, Gonzalez CM, Chapman HC, Cohen AS, Kirkpatrick B, et al. Geolocation as a digital phenotyping measure of negative symptoms and functional outcome. Schizophr Bull 2020 Dec 01;46(6):1596-1607 [FREE Full text] [doi: 10.1093/schbul/sbaa121] [Medline: 32851401]
69. Gong J, Huang Y, Chow PI, Fua K, Gerber MS, Teachman BA, et al. Understanding behavioral dynamics of social anxiety among college students through smartphone sensors. Inf Fusion 2019 Sep;49:57-68. [doi: 10.1016/j.inffus.2018.09.002]
70. Beekman AT, de Beurs E, van Balkom AJ, Deeg DJ, van Dyck R, van Tilburg W. Anxiety and depression in later life: co-occurrence and communality of risk factors. Am J Psychiatry 2000 Jan;157(1):89-95. [doi: 10.1176/ajp.157.1.89] [Medline: 10618018]
71. Jang C, Sung W. Beyond the privacy paradox: the moderating effect of online privacy concerns on online service use behavior. Telemat Inform 2021 Dec;65:101715. [doi: 10.1016/j.tele.2021.101715]
72. Debatin B, Lovejoy JP, Horn AK, Hughes BN. Facebook and online privacy: attitudes, behaviors, and unintended consequences. J Comput Mediat Commun 2009 Oct;15(1):83-108. [doi: 10.1111/j.1083-6101.2009.01494.x]

Abbreviations

DASS-21: Depression, Anxiety, and Stress Scale–21
EMA: ecological momentary assessment
HDBSCAN: Hierarchical Density-Based Spatial Clustering of Applications With Noise
MAE: mean absolute error
MAPE: mean absolute percentage error
MDD: major depressive disorder

XSL•FO
RenderX

NMP-Q: Nomophobia Questionnaire
PSWQ: Penn State Worry Questionnaire
QIDS: Quick Inventory of Depressive Symptomatology
RMSE: root mean square error
RRS: Rumination Response Scale
UCLA: University of California, Los Angeles

Edited by A Mavragani; submitted 19.02.24; peer-reviewed by A Hassan, DL Smith; comments to author 13.06.24; revised version received 03.07.24; accepted 02.01.25; published 19.02.25.

Please cite as:
Zhang T, Camargo A, Schmaal L, Kostakos V, D'Alfonso S
Nomophobia, Psychopathology, and Smartphone-Inferred Behaviors in Youth With Depression: Longitudinal Study
JMIR Form Res 2025;9:e57512
URL: https://formative.jmir.org/2025/1/e57512
doi:10.2196/57512
PMID:39969982

©Tianyi Zhang, Andres Camargo, Lianne Schmaal, Vassilis Kostakos, Simon D'Alfonso. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Adapting the Germ Defence Web-Based Intervention to Improve Infection Prevention and Control in Care Homes: Interview Study Among Care Home Staff

Alex Hall[1,2], PhD; Johanna Aguilera-Muñoz[1,2,3], PhD; Lisa McGarrigle[1,2], PhD; Charlotte Eost-Telling[1,2,3], PhD; James Denison-Day[4], PhD; Christie Cabral[5], PhD; Merlin Willcox[6], DPhil; Chris Todd[1,2,3], PhD

[1]School of Health Sciences, University of Manchester, Manchester, United Kingdom

[2]Manchester Academic Health Science Centre, Manchester, United Kingdom

[3]National Institute for Health and Care Research Applied Research Collaboration Greater Manchester, Manchester, United Kingdom

[4]Centre for Clinical and Community Applications of Health Psychology, University of Southampton, Southampton, United Kingdom

[5]Centre for Academic Primary Care, University of Bristol, Bristol, United Kingdom

[6]Department of Primary Care and Population Sciences, University of Southampton, Southampton, United Kingdom

Corresponding Author:
Alex Hall, PhD
School of Health Sciences
University of Manchester
Jean McFarlane Building
Oxford Road
Manchester, M13 9PL
United Kingdom
Phone: 44 01613068705
Email: alex.hall@manchester.ac.uk

Abstract

Background: Infection prevention and control (IPC) is vital in care homes as it can reduce morbidity and mortality by 30%. Ensuring good IPC practice is a perennial challenge in the varied and complex context of care homes. Behavior change interventions delivered via digital technology may be effective in improving IPC among care home staff.

Objective: This study aimed to evaluate how an evidence-based, digital behavior change intervention called Germ Defence can be rapidly adapted to meet the needs of care homes.

Methods: This study applied the person-based approach, which emphasizes iterative approaches to optimizing interventions via individual user feedback. Phase 1 involved initial edits to the website by the research team to create Germ Defence for Care Homes (GDCH) version 1. Phase 2 consisted of stakeholder consultation on GDCH version 1 followed by edits to create GDCH version 2. The formal research (phases 3 and 4) involved individual think-aloud interviews with 21 staff members from management, care, and ancillary positions in 4 care homes providing real-time feedback as they worked through GDCH. Edits were made to create GDCH version 3 between phases 3 and 4. During the development of GDCH versions 2 and 3, it became clear that the intervention would need more fundamental changes beyond the pragmatic, incremental changes that would be possible within the scope of this study. Analysis was completed via a rapid, qualitative descriptive approach to develop a high-level summary of key findings from the interview data.

Results: There were mixed results about the attractiveness of GDCH and its suitability to the care home context. Participants felt that the images needed to be aligned much more closely with the meaning of adjacent text. Many participants felt that they would not have time to read a text-based website, and some suggested that more engaging content, including audio and video, may be preferable. Most participants felt that the overall concept of Germ Defence was clearly relevant to their context. Some felt that it might be a useful introduction for new staff members or a refresher for current staff, but others felt that it did not add anything to their existing IPC training. There were mixed opinions about the level of detail provided in the information offered by the site. While the goal-setting behavior change mechanism may have potential, the findings suggested that it may be unsuitable for care homes and more work is needed to refine it.

XSL•FO
RenderX

Conclusions: Much more work needs to be done to make Germ Defence more engaging, accessible, and relevant to the care home workforce. Our study highlights the challenges of rapidly adapting an existing intervention to a new context. Future research in this area will require a pragmatic methodological approach with a focus on implementation.

(JMIR Form Res 2025;9:e66706) doi:10.2196/66706

KEYWORDS

care homes; long-term care; nursing homes; infection prevention and control; behavioral intervention development; person-based approach; qualitative

Introduction

Background

Infection prevention and control (IPC) is vital in long-term care facilities as it can reduce morbidity and mortality by 30% and may reduce unplanned hospital admissions and overuse of antibiotics [1]. In England (the setting of this study), IPC is a central focus of the Care Quality Commission, the independent health and care regulator that inspects long-term care facilities [2]. International terminology on long-term care facilities varies; in this paper, we use the term *care home* to refer to a long-term care facility providing 24-hour care, either with nursing (ie, in which residents with complex needs receive care provided by registered nurses on-site) or without nursing (ie, in which residents with less complex needs receive support with personal care from nonregistered staff, with nursing and health care input provided as required by community practitioners). In England and Wales, there are approximately 278,000 care home residents aged ≥65 years (approximately three-quarters of whom are aged ≥80 years), approximately half of whom live in homes with nursing and half of whom live in homes without nursing and over two-thirds of whom experience substantial limitations in their ability to carry out day-to-day activities [3].

The COVID-19 pandemic of 2020 to 2022 brought both the importance of and challenges in ensuring good IPC to the forefront of public consciousness [4-6]. It also raised the profile of care homes, which are often marginalized and forgotten by the public and policy makers and which were hit particularly hard [7]. Many countries struggled with ensuring good IPC in care homes during the pandemic, but although international comparisons are difficult [8], the United Kingdom is perceived to have fared particularly badly. In England, the first 5 months of the pandemic saw approximately 20,000 excess deaths among care home residents that were directly attributable to COVID-19 [9]. Before the pandemic, care home residents in England aged ≥65 years had a 10-fold higher mortality than older people in private homes, but in the first wave of the pandemic, this rose to an 18-fold difference [10]. At least for the first months of the pandemic, care homes experienced substantial, widely reported challenges in accessing COVID-19 testing and in isolating residents and a lack of availability of personal protective equipment (PPE; eg, gloves, face masks, and aprons) [11,12]. Other systemic problems linked to a higher risk of COVID-19 outbreaks included a lack of statutory sick pay for staff (thereby discouraging staff to take time off work) and reliance on agency workers (who work across multiple sites) to address workforce shortages [13]. Care homes were locked down with only essential staff access, and government guidance on visiting care homes was not issued until July 2020, which was 4 months into the pandemic, and, thereafter, proceeded to change frequently. Care home managers struggled with late notice about changes (often being informed after national televised briefings), gaps regarding risks and harms to residents, a lack of clarity on how to interpret guidance coupled with uncoordinated support from local regulators, and lack of acknowledgment of implementation challenges [14]. Restrictive IPC measures in care homes have been the subject of intense debates regarding their psychosocial impacts on residents and potential breaches of mental capacity and human rights law [15,16].

COVID-19 stands as the contemporary paradigmatic case study of (failures in) IPC in care homes, but many of the challenges it highlighted long predate the pandemic. Other common viral infections in care homes include seasonal influenza (another respiratory virus) and norovirus (the main cause of gastroenteritis, which is spread through physical contact); bacterial infections such as *Salmonella* sp and group A *Streptococcus* sp; and infections caused by nonviral and nonbacterial agents, including scabies [17-19]. Care homes are particularly vulnerable to all these pathogens because of the physical proximity between people and the fact that most residents are of advanced age and living with chronic health conditions and require hands-on work with physical contact. In England, all care homes are required to have IPC plans, and all staff are required to complete IPC training [2]. A Cochrane rapid review of qualitative evidence (published between 2005 and 2020) of barriers to and facilitators of health care workers' adherence to IPC guidelines showed numerous enduring challenges that resounded during the COVID-19 pandemic, including (but not limited to) lengthy and changing guidance, increased workload in following guidance, lack of easy access to handwashing facilities, lack of training about infections and PPE use, lack of good-quality PPE, and difficulties in using PPE if patients felt frightened or stigmatized by it [20]. Some of these challenges are particularly pertinent to care homes as they function as both people's homes and places of care and are staffed by a workforce that may lack training and skills compared to those in other health care sites, including key ancillary staff (eg, cooks and cleaners) who are crucial to IPC [21]. Hand hygiene among care home staff may have been poor for many years [22], and a systematic review of evidence published before the pandemic found that poor hand hygiene was the most common cause of transmission of infections in care homes [23]. The COVID-19 pandemic reinforced the importance of PPE and behavioral measures such as handwashing to reduce transmission of infection [24] but also strikingly highlighted unintended consequences where use of PPE such as gloves may actually reduce compliance with

XSL•FO
RenderX

handwashing [25]. A recent meta-ethnography of qualitative studies of IPC in care homes found that IPC can be considered by staff to be "outside" the control of the care home both in terms of the source of infection and in actions to control it—the tension of implementing IPC in what is meant to be a homely environment appears to contribute to a sense that outbreaks are inevitable, with care home staff ambivalent about the benefits of applying IPC and perceiving a lack of ownership over IPC practice [26]. Ensuring good IPC practice is a perennial challenge in the varied and complex context of care homes. It remains of primary importance to consider how to improve IPC interventions and their implementation to help protect care home residents and staff from outbreaks of infectious diseases.

A recent systematic review of randomized controlled trials of interventions to prevent transmission of infections in care homes found only 4 trials of interventions to improve hand hygiene and IPC, with variable results [27]. Only 1 study showed a reduction in pneumonia incidence, 1 showed a reduction in influenza outbreaks, and 1 showed a reduction in incidence of influenzalike illness. Achieving high levels of adherence to infection control is challenging, and the interventions were all different; 3 contained in-person presentations, and only 1 contained an e-learning component. The review recommended developing more impactful behavior change interventions for improving IPC in care homes using behavioral science and the person-based approach.

Germ Defence: A Web-Based Behavior Change IPC Intervention

The World Health Organization outlines 8 core components of IPC programs in health care, including emphasis on multimodal strategies to change individual behavior [28]. A systematic review of IPC interventions in care homes found that those incorporating behavior change strategies via education, monitoring, and feedback were the most successful [29]. There is increasing recognition of the potential role of digital health technologies in IPC [30-33], which may be associated with their increased use to facilitate behavior change across a range of health domains, particularly via goal setting and self-management [34]. The intervention in this study is Germ Defence, a mobile-friendly website providing targeted, tailored advice on how and why people should adopt IPC behaviors, supplementing public health guidance with evidence-based and theory-based behavior change techniques [35]. It consists of introductory content about the importance of IPC followed by a series of pages arranged in parallel pathways where users can select from 2 components of interest (handwashing and reducing illness) and follow sequential pages of content streams that result in tailored advice and goal setting to improve or increase IPC behaviors.

Germ Defence was originally developed in response to the swine influenza (H1N1) and H5N1 influenza outbreaks of 2009 to 2010 in the form of an intervention then called PRIMIT (Primary Care Trial of a Website-Based Infection Control Intervention to Modify Influenza-like Illness and Respiratory Infection Transmission) to encourage handwashing delivered in 4 weekly sessions [36,37]. A subsequent large randomized controlled trial of >20,000 people living in shared households conducted across 3 winters (2011-2013) showed that the PRIMIT intervention increased handwashing and reduced respiratory tract infections [38]. The intervention was developed using the person-based approach, which involved conducting in-depth qualitative research to develop detailed understandings of how to overcome the behavioral barriers that users might encounter to engaging with the target behaviors [39]. It also used behavior change approaches to incorporate education, personalized goal setting, environmental prompting, and emotional motivation [40]. A process evaluation showed that the intervention was effective for both men and women, older and younger people, and people with both lower and higher levels of education [40]. Importantly, the biggest change in handwashing behavior occurred after the first of the 4 weekly sessions, so the intervention structure was subsequently adapted for wider dissemination to comprise 1 stand-alone session, with further optional content available immediately afterward [40]. The adapted intervention—now known as Germ Defence—was disseminated "in the wild" to help reduce seasonal colds and influenza.

In 2020, Germ Defence was rapidly adapted for the COVID-19 pandemic by a team of medical, public health and behavior change experts, and public contributors [35], with cross-sectional work highlighting its potential to improve IPC behavior [41]. It was endorsed by the chief medical officer for England as a priority via a large cluster-randomized controlled trial in primary care rolled out across all general practices in England (N=6579) [42]. At the same time, work with schoolchildren, parents, and school staff to use Germ Defence to mitigate the spread of COVID-19 in schools suggested that it is possible to adapt the intervention for use in this institutional context [43,44]. The sense of collective responsibility promoted by the intervention was seen to be persuasive in adopting the suggested behaviors, although lack of physical space and resources within schools may be barriers to adoption [43].

Study Aim

IPC is a perennial challenge in care homes. Behavior change interventions delivered via digital technology may be effective in improving IPC. The existing evidence base for Germ Defence, its positioning as a national priority response to COVID-19, and its potential for adaptation to institutional contexts highlighted that consideration of adapting it to suit care homes was merited. The aim of this study was to evaluate whether and how Germ Defence might be adapted to meet the needs of care home staff.

Methods

Design

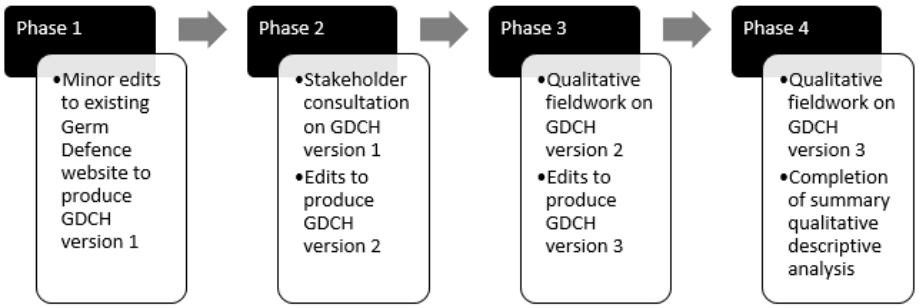
This study followed Medical Research Council guidance on intervention development by drawing on the intervention development aspects of the person-based approach, which emphasizes iterative approaches to optimizing interventions [39,45,46]. It involved modification of the existing Germ Defence for Schools website [44] across 4 phases (Figure 1). We chose the school version as our starting point because of its potential greater initial applicability to an institutional context than the original version intended for use only in private

XSL•FO
RenderX

residences. After its initial general introduction pages, the school version contains 2 pathways of advice that users can choose to work through, one focusing on how to protect themselves in their own homes and one focusing on how to protect themselves in school. We intended to mimic this structure and change the organizational pathway from schools to care homes. Given the limited scope of our project, we focused our attention on the general introduction pages and the organizational pathway only.

Phase 1 involved preliminary research team edits to the website to create Germ Defence for Care Homes (GDCH) version 1. Phase 2 consisted of stakeholder consultation followed by further edits to create GDCH version 2. Phase 3 was the first half of the formal research, involving an initial round of qualitative fieldwork with care home staff followed by edits to create GDCH version 3. Finally, phase 4 was a second round of qualitative fieldwork exploring care home staff perceptions on GDCH version 3.

Figure 1. Four phases of the study.



Preliminary Work: Phases 1 and 2

In phase 1, we made some initial small changes to the existing website to ensure that it referred to care homes rather than schools and to viruses more generally rather than to SARS-CoV-2 specifically. This became GDCH version 1. In phase 2, we held brief individual consultations with 8 staff members from 3 care homes (n=3, 38% managers; n=3, 38% carers; and n=2, 25% nurses) to gauge their initial responses to GDCH version 1. The changes they suggested were incorporated into GDCH version 2, which was then used in the qualitative fieldwork. Details of key changes are presented in the Results section.

Formal Research: Phases 3 and 4

Recruitment

The formal research involved 2 phases of qualitative fieldwork with iterative intervention development. We conducted fieldwork with care home staff in Greater Manchester. We aimed to purposively sample between 4 and 6 care homes representing a range of characteristics known to be related to infection rates (large and small size, modern purpose-built and retrofitted houses, with and without nursing care, not for profit and for profit, and inspection quality rating) [9]. Within each care home, we aimed to recruit 5 to 6 staff members purposively sampled to represent a wide range of roles, including managers, carers, and ancillary staff. Therefore, the total estimated sample was between 20 and 36 staff members, which aligns with previous work on qualitative sampling [47]. Care homes were recruited via a range of existing networks and contacts in the region, including the National Institute for Health and Care Research Applied Research Collaboration Greater Manchester Care Home Research Collaboration and the Enabling Research in Care Homes network. Within each care home, we provided promotional materials about the study for display and sought guidance from managers as to suitable staff members to invite.

Data Collection

The person-based approach to intervention development involves using "think-aloud" interviews to understand users' contexts and views on every aspect of the intervention and modifying the intervention accordingly [46]. Think-aloud interviews have roots in cognitive psychology and are commonly used to collect real-time feedback and information about people's thought processes as they engage with a particular phenomenon of interest [48]. All staff members in phases 3 and 4 participated in individual think-aloud interviews. Participants were shown the contemporaneous version of the GDCH website and were asked to work through the intervention in real time. The think-aloud element drew on an interview guide that included neutral prompts and follow-up probes to encourage participants to elaborate on their actions and verbalizations in the moment (eg, "Can you tell me a bit more about that?"). This approach offered participants the opportunity to express immediate reactions to all aspects of intervention content and functionality as they used them, page by page. Details of key changes after each phase are presented in the Results section.

All interviews were conducted individually face-to-face on care home premises, lasting 28 to 54 minutes. They were audio recorded and transcribed verbatim. We also collected demographic data on participants' age, ethnicity, role, and length of time in the post. Data collection was undertaken by one researcher (JA-M) in May 2023 to June 2023. This researcher had a background in qualitative research within organizational psychology and, before data collection, attended training on the person-based approach and think-aloud interviews.

Data Analysis

Our initial approach to data analysis was to adopt the "table of changes" analysis commonly used with the person-based approach [39,45,46]. This analysis offers a rapid but rigorous method for extracting quotes and feedback about each component of the intervention from the think-aloud interviews and considering the relative importance of possible changes to

XSL•FO
RenderX

the intervention. One researcher (JA-M) read the interview transcripts and extracted quotes from them into a Microsoft Excel (Microsoft Corp) document detailing positive and negative comments by website page or section. This document was shared with the wider research team, and we discussed possible changes and their potential importance. Following phase 3, we made changes to GDCH version 2 that we considered important yet practicable within the resources of the study. The revised GDCH version 3 was then used in phase 4 with the same intention. However, during the development of GDCH versions 2 and 3, it became apparent that the website would need more fundamental changes than any of the pragmatic, incremental changes identified via the table of changes approach, which would be beyond the limited scope of this study. Therefore, we conducted a rapid qualitative descriptive analysis [49] to present a high-level summary of the key findings from the interview data. This involved taking the data that we had arranged in our table-of-changes Microsoft Excel documents, coding them, and grouping these codes into topic summaries. One researcher (AH) developed a set of codes that were discussed and agreed upon with the team. These codes (n=28) were applied to the remaining data by another researcher (JA-M). The team agreed on the grouping of these codes into 3 higher-level topics.

Ethical Considerations

Research participants in phases 3 and 4 provided written informed consent. Ethics approval was granted by the University of Manchester Proportionate University Research Ethics Committee (November 8, 2022; reference 2022-15380-25722). Participants were informed about the study's purpose and procedures and their right to withdraw at any time without detriment. To protect participants' privacy, all data were fully anonymized. Research team members (JA-M and AH) reviewed the transcripts and removed any identifying information. Care homes were offered reimbursement of up to £25 (US $24.76) per staff member released to participate for up to an hour [50]. Payments were offered to the organizations rather than to individual staff members to contribute to backfilling of time if necessary and to avoid any perception of unfairness among nonparticipating staff.

Results

Care Homes and Participants

In total, we recruited 4 care homes. All homes were for-profit organizations, with variation in type of home, building design, size, and quality rating (Table 1).

Overall, we recruited 21 members of staff (Table 2). Participants occupied all major positions commonly found within care homes, including caring, senior caring, nursing, leadership and management, directorship, cleaning, and kitchen roles.

Table 1. Care homes in the study.

Care home number	Care home type	Building type	Size	Ownership	Inspection rating[a]
1	With nursing	Purpose built	26 beds	For profit	Good
2	Without nursing	Purpose built	40 beds	For profit	Outstanding
3	With nursing	Purpose built	31 beds	For profit	Requires improvement
4	Without nursing	Retrofitted house	17 beds	For profit	Inadequate (under the previous provider; was being reinspected at the time of this study)

[a]Care homes in England are subject to regular, usually unannounced inspection by the Care Quality Commission, the sector regulator. The inspection focuses on 5 domains assessing the extent to which a home is safe, effective, caring, responsive, and well led. The inspection results in 1 of 4 overall ratings: "outstanding," "good," "requires improvement," or "inadequate." Homes with the 2 lowest ratings are likely to be reinspected sooner than homes with the 2 highest ratings.

XSL•FO
RenderX

Table 2. Participant details[a].

ID	Care home	Role	Study phase
P1	Home 1	Housekeeper	3
P2	Home 1	Senior carer	3
P3	Home 2	Health care assistant (enhanced)	3
P4	Home 2	Registered manager and director	3
P5	Home 2	Care assistant (plus domestic team lead)	3
P6	Home 2	Health and development lead	3
P7	Home 1	Senior carer	3
P8	Home 1	Deputy manager	3
P9	Home 2	Quality and compliance lead	3
P10	Home 2	Care team lead (plus acting care manager)	3
P11	Home 3	Carer and health care assistant	4
P12	Home 3	Carer and health care assistant	4
P13	Home 3	Domestic cleaner	4
P14	Home 3	Head chef	4
P15	Home 3	Team leader	4
P16	Home 3	Staff nurse (mental health)	4
P17	Home 4	Domestic assistant	4
P18	Home 4	Chef	4
P19	Home 4	Senior care assistant	4
P20	Home 4	Care assistant	4
P21	Home 4	Registered manager	4

[a]Time in the post ranged from 4 months to 21 years. The mean age was 43.38 (SD 13.27) years. In total, 90% (19/21) were female, and 67% (14/21) were of White British ethnicity.

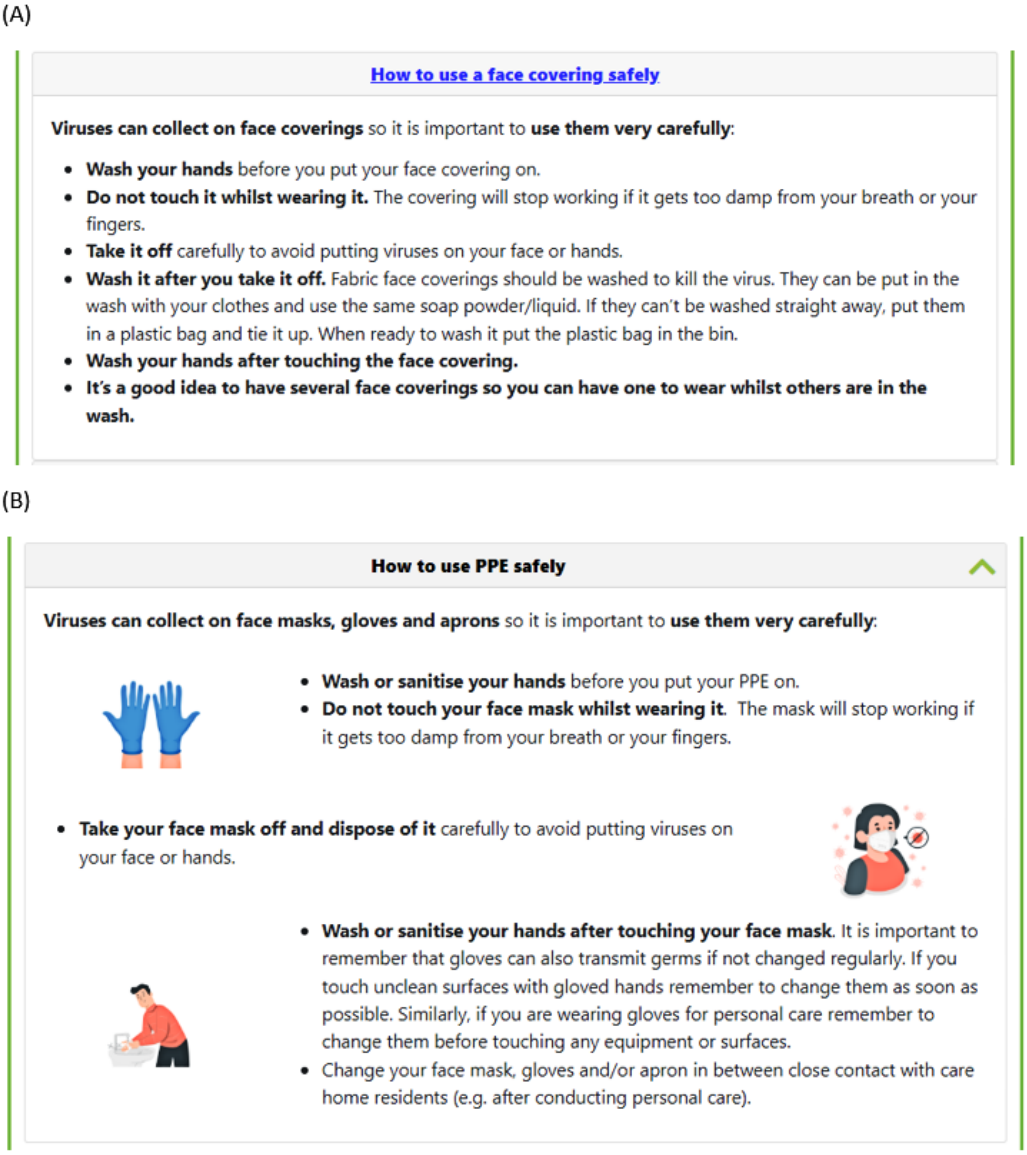
Phases 1 and 2: Minor Edits

In phase 1, we made 2 types of minor changes to some of the wording of the website to create GDCH version 1. The first was to change references from "school" to "care home." The second was to broaden the focus from COVID-19 to talk about infectious diseases more generally, including changing references to "COVID-19" to "respiratory viruses." A major comment by the stakeholders consulted in phase 2 highlighted that a lot of the content in the section about mask wearing was very much dated to the COVID-19 pandemic, which may be unsuitable because permanent mask wearing was now not part of mandatory practice in care homes. When developing GDCH version 2, we amended this section to refer to using PPE more generally, which included some mention of masks but also other PPE commonly used in care homes, such as gloves and aprons, and delivering personal care (Figure 2).

In phase 1, GDCH version 1 retained the original Germ Defence behavior change component that asked participants to indicate how often they had performed certain IPC actions (eg, handwashing) per day over the previous week on an ordinal scale from 1 to 10 and then to make a plan for how frequently they intended to perform this behavior over the following week. In phase 2, some stakeholders felt that the scale from 1 to 10 would be unsuitable in the care home context as care home staff would wash their hands much more frequently. Therefore, in GDCH version 2, we removed this scale altogether. GDCH version 1 also contained a component that invited staff to think about points at which they had performed certain IPC actions in a typical working day on a 5-point scale ("almost never," "sometimes," "quite often," "very often," and "always"). We revised it to include more specific care home scenarios (eg, before coming into close contact with a resident and when donning PPE). We retained a sixth option of "does not apply to me" on certain questions (eg, those about PPE) because we were aware that some staff members, such as kitchen staff, would not perform certain activities, such as administering personal care.

XSL•FO
RenderX

Figure 2. (A) Germ Defence for Care Homes version 1 face covering and (B) Germ Defence for Care Homes version 2 personal protective equipment (PPE) sections following stakeholder feedback in phase 2.



Phases 3 and 4: Key Findings and Further Intervention Modifications

Overview

Findings from phases 3 and 4 were grouped into 3 broad topic areas presented in the following sections along with key iterative changes made during the different project phases. *Attractiveness and accessibility of Germ Defence* reflects the overall appeal of the website and how accessible and engaging it was perceived to be. *Usefulness of Germ Defence* reflects the extent to which participants felt that Germ Defence might be useful in the care home context. *Individual behavior change in a complex environment* reflects findings relating to the goal-setting behavior change mechanism underpinning the intervention.

Attractiveness, Accessibility, and Usability of the Germ Defence Website

Participants in phases 3 and 4 made a wealth of suggestions for revising the images and the text of the website. One example of this was in relation to a cartoon showing how germs may spread between people (Figure 3A). In phase 3, some participants found this cartoon confusing:

> *You're looking at that thinking, I don't know what's going on. You're reading that and then you've got to work out who's who and which way it's all going.* [P6; health and development lead; phase 3]

However, others were positive about the cartoon and its accompanying text:

> *...it just shows how easily it's passed on to multiple other people. So, I think the text is enough and the picture is enough and I think they work quite well together.* [P4; registered director and manager; phase 3]

> *I think that tells the story quite well. Obviously, if it was in a care home setting...if you change the name Chloe to, like...Maureen or Beryl or Joyce, these are, sort of, the names that we have that are, sort of, older names for people...But, yeah, I mean, in terms of the*

XSL•FO
RenderX

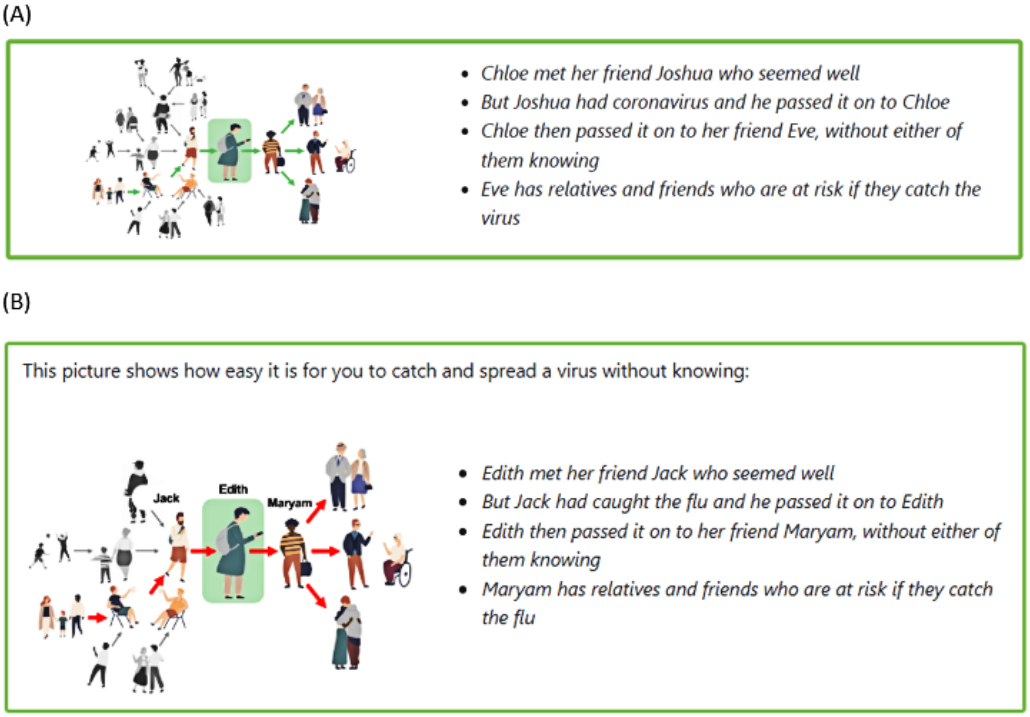
image, I think it works. [P9; quality and compliance lead; phase 3]

As an initial change following phase 3, we edited the names of some of the characters to be more reflective of care home resident age profiles (Figure 3B). Feedback in phase 4 suggested that the accompanying text was still necessary as the image alone did not clearly convey the message:

Once you read all this, you know what it is. But without the reading and without seeing anything, I can't tell with this picture what it is. [P15; team leader; phase 4]

Some participants suggested incorporating an image of a germ within the cartoon showing how it might move between people. Others suggested using color coding, such as red hands on one person and red shoulders on someone else they had touched. Another suggested adding Venn diagram–like circles around the images to show overlapping relationships. These suggestions highlighted that even revising this one cartoon may require substantial work, likely conducted over multiple phases of iterative development.

Figure 3. Cartoon showing how germs may spread between people (A) in Germ Defence for Care Homes (GDCH) version 2, used in phase 3, and (B) in GDCH version 3, used in phase 4, with character names changed.



In both phase 3 and phase 4, some participants felt that they would not have time to sit and read through a text-based website. Some suggested that more engaging content, including via audio and video delivery, would be preferable (and identified some existing resources) and might address concerns surrounding English-language reading ability:

It's a good way of explaining it, but with it being a website...There's a video that we watched in training...I think it's a hospital setting and he's walking through, and it just shows how easy it is once you're touching it and then that's infected. I feel like that, from my experience, seeing that explains it a lot better than the picture there. [P5; care assistant and domestic team lead; phase 3]

Or one thing you could add, I don't know, is you could add an audio option where people could just click the audio and listen instead. You know, some people can't understand English or they don't...they can't read so that's another thing you can add to it. [P11; carer and health care assistant; phase 4]

Regarding engagement with the website, some participants in phases 3 and 4 expressed a preference for practical training over reading material:

So, I think having that training face-to-face really enables people to see the risks that they're putting themselves at if they're not following the correct procedures and risks to our residents. [P4; registered manager and director; phase 3]

I think practical's better for me, personally, because academically, I don't enjoy that side of it, I'm more practical, 'cause I'm a chef, I'm creative and that side. [P18; chef; phase 4]

These findings highlighted that any adaptation of Germ Defence may need to include a practical element of applied action.

In phase 4, some participants suggested that, overall, the images on the web pages needed to stand out from the text more and cautioned that they needed to convey the message without depending on accompanying text as the level of English-language ability among the care home workforce is highly variable:

XSL•FO
RenderX

> *...a lot of the girls, we do it on iPads, we do it on phones, things like that. So that picture's going to be minute...So I think you're best putting that picture on a page on its own, if you want it to stand out.* [P18; chef; phase 4]

> *...we have people that might not be able to understand English as well as other people but they can understand what images are. So, that's it, so this picture is equally as important as the contents within the text, yes.* [[P11; carer and health care assistant; phase 4]]

In summary, participants made a wealth of suggestions for changes that were not always mutually compatible. Incorporating these suggestions would require many phases of iterative development that were beyond the scope of our study.

Usefulness of Germ Defence Within the Care Home Context

The revised content about PPE (Figure 2) was received positively by some participants in phases 3 and 4. Some participants suggested that the content could better reflect requirements in the care home context, including how to properly dispose of PPE rather than simply saying, "dispose of it carefully." Some acknowledged that mask wearing was not mandatory practice and suggested that masks were not used at all in their care homes:

> *I think it's really good...I suppose, so mask wearing is becoming less prevalent in homes, so we wouldn't...it wouldn't be the sort of, something we'd push now, because we don't wear masks in our home.* [P9; quality and compliance lead; phase 3]

However, we avoided making any further changes to references to masks because we were aware that homes may still use them in the event of another disease outbreak. This was exemplified in phase 4:

> *I think that's, yes, giving you the right information. Not necessarily right now. I wouldn't say that you need to wear a face mask in a care home now, but if there was a virus in the care home which was really contagious as coronavirus was, then that is a good point to have there.* [P20; care assistant; phase 4]

Overall, there were mixed views on the potential utility of Germ Defence. Most participants felt that the overall concept of Germ Defence was clearly relevant to their context, and some felt that it might be useful as an introduction for new staff members or as a refresher for current staff:

> *We wouldn't use it that much...The team have enough to do. They're really busy. I could imagine that I might use that on orientation, so when I'm getting people in and first starting then I think this is a good introduction to what IPC is all about and then when they're doing the competencies out on the floor or that's when they're actually learning all the practical sides of it.* [P4; registered manager and director; phase 3]

> *We know that infection prevention and control can be difficult, sometimes, especially in a busy care home. The Germ Defence website will give you a hand, a refresher, on how to prevent and control the spread of infections. Yeah.* [P16; staff nurse; phase 4]

However, others felt that they were already doing what Germ Defence suggested and that it did not seem to add anything to their existing IPC training, which was felt to be adequate because inspections had been passed:

> *I don't think they're helpful because I'm already doing it...We've had Infection Control round twice and they've not pulled anything that I've done, so I'm doing my job well...So I wouldn't...* [P17; domestic assistant; phase 4]

> *And if there's an...like on my phone, probably this morning, I've got an update to say that one of the policies has been updated. So, then we will read through it quickly...as things change, we're kept very much up to date, and I think, I don't know, I don't think we need this additionally, really.* [P18; chef; phase 4]

There were also mixed opinions about the level of detail provided in the information offered by the site. For example, in phase 3, some participants felt that, at certain points, there needed to be comprehensive information to remind them about the transmission of germs and viruses as it is a complex subject:

> *You need a lot of information to let you know how the infection starts, and how to stop it, and control it, and stuff. If you didn't have the information, you wouldn't know.* [P1; housekeeper; phase 3]

However, at other points, some participants felt that brevity was preferable:

> *It's nice and...I like...personally I like a bullet point, so have, like, long...Because of my attention, and other people's, I think, it needs to be bullet pointed [so it] sticks.* [P3; health care assistant (enhanced); phase 3]

Participants in phase 3 expressed mixed opinions about the level at which Germ Defence was pitched. Some felt that it was appropriate as it balanced recognition of previous knowledge with a need to be accessible to care staff from a wide range of educational backgrounds:

> *I think sometimes some other professional bodies tend to dismiss carers as bottom of the rung, which we're not, it's a highly professional career now. So, I like that it recognises that we will have some knowledge around this, the infographic is clear, it sets it out very clearly how infection is spread. I think it's pitched at just the right level really because there are some people who work in care, whose educational background is not as good as others.* [P4; registered manager and director; phase 3]

However, others felt that it was too basic for experienced staff:

> *I realise that new staff are coming into an environment and they don't know things, but I think*

XSL•FO
RenderX

> *when you've got experienced staff, it sounds a bit, as I say, condescending.* [P8; deputy manager; phase 3]

As a first attempt to acknowledge this ambivalence, we added a self-rated confidence scale (1=least confident; 10=very confident) to the third page of the website in GDCH version 3 that acknowledged that staff may already be familiar with IPC measures. The page would display 1 of 2 messages depending on the self-rating selection made by the staff member: for lower confidence (1-6), the message would say that Germ Defence would give more advice and guidance about IPC and, for higher confidence (7-10), the message would say that Germ Defence would provide a handy refresher on IPC. In phase 4, feedback on this scale was that some participants felt that staff who rated themselves as very confident about IPC practice would not continue to look through the website:

> *So that's just informing us that if we want to look further we can go to that website...Personally, I don't think many people would look at it...I would say...half the people, they wouldn't bother.* [P18; chef; phase 4]

Individual Behavior Change in a Complex Environment

These findings relate to the goal-setting mechanism of the intervention, which we modified slightly following stakeholder feedback in phase 2. In phases 3 and 4, participants' feedback on the goal-setting pages suggested that, although they were easy to read and it was clear how to interact with the pages to answer the questions, it was not clear whether this might lead to any sustained change in behavior. We were initially concerned about the potential for social desirability bias in responses and that participants would simply choose "always" for every behavior. However, this did not appear to be the case as participants appeared to acknowledge shortfalls, at least to some extent:

> *[Regarding washing hands] touching something that other people have touched, like door handles, and furniture. I'd say, "very often," but sometimes we might forget, you know what I mean.* [P1; housekeeper; phase 3]

Some participants appeared to be unconvinced by the "making a plan" goal-setting activity, citing externally driven IPC requirements as having the ultimate influence over individual behavior and highlighting that noncompliance could result in disciplinary action:

> *Making a plan, yeah, I mean, essentially, we're wearing an apron and a mask because that is a requirement, so to make the plan of wearing them is what we're doing anyway because it's a requirement to do it. So if the aprons came back into being a requirement from IPC, then we wouldn't be planning on doing it, we'd have to do it, otherwise you wouldn't be coming to work, you'd have to do it. And the PPE for the personal care is a requirement, and if you don't do it you'd have to go down the conversations and supervisions and that type of thing.* [P6; health and development lead; phase 3]

This participant went on to say that the "making a plan" activity may be more beneficial for new, inexperienced colleagues rather than for experienced staff for whom IPC behavior was "embedded."

Some of the statements regarding the behavior change component specifically related to wearing face masks, which were not mandated at the time of data collection. There were examples to suggest that, if an aspect of IPC was not mandatory, it may be seen as not applicable by some participants:

> *It doesn't apply to me but obviously it is optional when you're giving personal care. But personally, I don't wear them when it's not mandatory. So, I'd say "does not apply to me."* [P5; care assistant and domestic team lead; phase 3]

There were other examples of how participants' answers might change depending on the type of activity and area of the home they were thinking about when reading the page:

> *Face mask, not all the time, these days, but sometimes if we want to get a resident's room, we are choosing, but in the kitchen, we don't use it all the time. "Quite often," maybe.* [P14; head chef; phase 4]

Therefore, it was unclear whether the questions and possible responses were sensitive enough to the breadth of activities that might arise in the care home setting. Other participants (eg, P12, carer and health care assistant) indicated that they would choose "always" for most questions about their behavior the previous week and would not make any changes to their future plan of behavior. It was unclear whether this response was influenced by social desirability bias. However, some participants appeared more reflective over the exercise, such as one manager who suggested that it made them question their behavior:

> *It makes you think. So when you have ticked your answers, if they're not "always" on both sets of the questions, it makes you think about your actions...I clicked "always," because that's what I do. But I don't do it on close contact with a resident, so that was "very often," because most times I do. Whereas on the second part, and then reading and thinking, that's maybe one of the places where I should always wash my hands, in close contact with the resident.* [P21; registered manager; phase 4]

If participants did not indicate substantial improvements to their plan for future activity (ie, if they continued to select answers indicating a low frequency of behavior), they would then see a page asking them to reconsider. Some suggested that they felt indifferent about this page—there was nothing "wrong with it," but it did not appear to be very motivating:

> *Interviewer: So, now you have this page saying "are you happy with your plan?" What do you think of this?*
>
> *Participant: I don't really think anything of it. I don't know...sanitise your hands sometimes before coming into close contact with a care home resident, why not try wash and sanitising them quite often from now on. Yes, I won't say anything's wrong with it.* [P20; care assistant; phase 4]

XSL•FO
RenderX

There were also some comments about the utility of conceiving the goal-setting exercise across a weekly time frame, how the intervention might be implemented alongside the remit of infection control leads, and whether there needed to be some output that captured the plan that a staff member had made as "evidence" of their proposed change in behavior:

> *[It would be useful if it was] downloadable or we can save it in some way to build it into the audit for us, as an evidence-base...we wouldn't do it weekly, they get so many competency checks and supervisories and everything else that this would be yet another thing we would add in. But it's certainly something we could add in as part of that auditing process that we do which is done at a minimum of twice yearly...Then if we find somebody isn't following procedure or isn't changing the PPE effectively, then we can go back and go, "look, you said you were going to do it, so do it," and just have that evidence-base there a little bit. I think but because we keep on top of it and because we've got infection leads, it's not something we would do weekly by any stretch. I think you would be hard pressed to find a home that would do it weekly.* [P4; registered manager and director; phase 3]

Therefore, overall, our findings suggest that, while the goal-setting behavior change mechanism may have potential, more work would be needed to refine it to facilitate successful individual behavior change when there is organizational-level regulation about what IPC actions should be taken by staff and how often.

Discussion

Principal Findings

This paper reports on qualitative work to evaluate whether and how an existing web-based, evidence-based behavior change IPC intervention called Germ Defence might be adapted to meet the needs of care home staff. Overall, our findings showed that adapting Germ Defence to this context is extremely challenging. Participants had mixed views about the appeal of Germ Defence and its potential suitability to the care home context and suggested that much more work would need to be done to make Germ Defence more engaging, accessible, and relevant to the care home workforce. The original Germ Defence intervention was designed for a very different population with little IPC knowledge. The elements that explained the need for IPC tended to duplicate existing training, and the goal setting for IPC actions did not appear to make sense to care home staff, whose primary concern may have been to follow mandatory IPC requirements. Following completion of our care home study, the results of the large cluster-randomized controlled trial of the COVID-19 adaption of Germ Defence in English general practice underway at the same time [42] showed no evidence that the intervention affected rates of respiratory tract infections or other health outcomes [51]. As the actual use of the intervention was below the 25% assumed in the sample size calculations, it is difficult to draw robust conclusions [51]. However, the results of this trial are important to note in the context of our study. Our findings suggest that substantial further work would be required to adapt Germ Defence to the care home context.

As qualitative descriptive approaches do not aim to provide in-depth theoretical analysis [52], it is useful to offer some brief discussion of our findings by drawing upon the widely used theoretical model of the Unified Theory of Acceptance and Use of Technology (UTAUT) [53]. The UTAUT's unit of analysis is the individual user, which is applicable for our study because our aim was to work with individual members of care home staff to try to adapt Germ Defence to their occupational context [54]. Although most commonly used quantitatively, the UTAUT has been used as a qualitative analytical and interpretive lens (eg, a recent qualitative meta-synthesis of digital interventions for antimicrobial prescribing and monitoring [55]). The UTAUT comprises 4 main constructs; our data speak mostly to the first 2 that focus on perceptions of benefit and ease of use.

Performance expectancy refers to the degree to which participants believe that using the intervention would be beneficial. Our findings within this construct appear to be mixed. Some respondents felt that Germ Defence could be potentially useful to support induction of new staff or provide refresher training, but others suggested that it may be too simplistic. The degree to which participants thought that it would offer a genuine benefit was unclear. Although we did not explicitly investigate the current IPC training provided in each care home in depth, some participants' responses suggested that more work would need to be done for Germ Defence to be seen as a useful addition to their existing training and practice.

Effort expectancy refers to the degree of ease of use of the intervention. Again, our findings within this construct were mixed, with some participants talking positively about the simple layout but others finding it confusing. Many participants suggested that the images would need substantial refinement as they did not always clearly convey the message of the text. Feedback also suggested that a text-heavy website would not be suitable because staff may be too busy to read through it all, may prefer audio or video content, or may prefer more practical training. Although the goal-setting exercises appeared to be easy to understand in principle, an investigation into their effectiveness for behavior change was beyond the scope of this study. We did not have the resources to develop multimedia content (eg, video or audio options) to assess whether it would be engaging enough to result in repeated, realistic practice of the desired behaviors.

Although the reasons for this low take-up are unclear, we note that Germ Defence was initially developed approximately 15 years before this work, and in the context of care homes, it is unclear from our study whether further context-specific adaptations would be sufficient to deliver engaging content that meets contemporary expectations. For instance, the potential of virtual reality (VR) and augmented reality to deliver engaging and realistic training in technical medical skills has long been recognized, and there is increasing recognition of their potential to support training in nonclinical skills [56-58]. Recent work exploring the development and use of VR and augmented reality hand hygiene training with care home staff has found that staff were extremely positive about immersive VR delivery,

XSL•FO
RenderX

perceiving it to be more engaging and enjoyable than previous training and valuing its inherent repeated practice, task realism, and feedback inherent to the intervention [59]. However, there are some challenges with this mode of delivery (eg, some side effects of nausea from using VR that render it unsuitable for all and up-front cost implications) [59].

The other 2 constructs of the UTAUT refer to external factors influencing implementation of the intervention. The focus of our study meant that we had less data relevant to these constructs. *Social influence* captures the degree to which participants perceive that important others believe that they should use the intervention. Although our interviews did not explicitly focus on this construct, some participants suggested that external directives regarding IPC from authoritative bodies would be a stronger driver of behavior than individual goal setting. However, several studies suggest that adherence to external directives is suboptimal [21,22,26]. Therefore, Germ Defence could be designed to improve adherence to IPC rules and guidance for care homes and explain when certain IPC measures are useful even if they are not mandatory (eg, explaining why mask wearing reduces transmission of respiratory infections when a staff member or resident has a cough). Finally, *facilitating conditions* highlights the degree to which participants feel that their organization's infrastructure would support the intervention's use. Although these kinds of implementation questions were not addressed directly in our study, some participants suggested that Germ Defence would need to be seen as compatible with their routine use of smartphones and tablets. Some managers suggested that Germ Defence could be used alongside IPC audits, suggesting that there is potential support from care homes' administrative infrastructure. These issues warrant further consideration.

Strengths and Limitations

The main strength of this study is its diverse sample of organizations and participants. We recruited care homes across the spectrum of quality rating according to the sector regulator and homes providing both residential only and nursing care. We did not observe any strong differences in views across types of homes, suggesting that our findings are likely to apply across the care home sector. Within the homes, we recruited participants working in a wide range of roles, including senior management, direct care provision, and ancillary positions, and with a wide range of experience, from a few months to >2 decades. Some of our findings highlight possible differences in views from staff in different roles (eg, the questionable relevance of some of the current content to ancillary staff and managers' views on the potential of the intervention to be used to demonstrate compliance), which would warrant further consideration in intervention development. Therefore, we believe that our findings are likely to have a high degree of transferability to the care home sector more widely. In addition, we provided a transparent description of our study methods, including the iterative phases of the study, to promote credibility and dependability.

The main limitation of our study is its relatively rapid nature. It was conceived in response to the crisis faced in the care home sector from the COVID-19 pandemic, with the aim to consider how an existing evidence-based intervention might be adapted to meet the needs of care homes. The think-aloud interviews to capture participants' real-time reactions to using Germ Defence were not designed to elicit detailed perspectives on potential facilitators of and challenges to implementation of the intervention that may contribute to intervention design. Our data lend themselves to a descriptive summary analysis rather than to a more nuanced exploration via a more interpretive approach, such as thematic analysis [60]. Despite these limitations, our study highlights the challenges of making rapid adaptations to Germ Defence in this context. Any such work would require a more sophisticated methodological approach, drawing more comprehensively on implementation theory. This could involve a greater focus on more fundamental aspects of intervention design advocated at an earlier stage of the person-based approach methodology [46], including a detailed investigation of current IPC training and practice and more detailed, theoretically informed exploration of participants' understandings of the intervention (eg, via a greater focus on the constructs of the UTAUT or of normalization process theory [61]). Such an approach was beyond the scope of this study. However, it is also unclear whether such work would be merited. Finally, it is worth noting that the ongoing independent public inquiry into the UK government's decision-making during the pandemic includes scrutiny on the extent to which care homes were protected and highlights that, in addition to single interventions, there is an urgent need for well-prepared, coherent, strategic IPC responses to such health emergencies across health and care systems [62,63].

Conclusions

Our study highlights the challenges of rapidly adapting an existing intervention to a new context. Future research in this area would require a pragmatic methodological approach with a focus on implementation. In the context of the GDCH intervention, it is unclear whether such work is warranted.

Acknowledgments

This study is funded by the National Institute for Health and Care Research (NIHR) School for Primary Care Research program (grant reference 582) and supported by the NIHR Applied Research Collaboration Greater Manchester (NIHR200174) and by Senior Investigator awards to CT (NIHR200299 and NIHR205156). MW's salary was partly funded by the NIHR (NIHR302412). CC acknowledges support from the NIHR Health Protection Research Unit in Behavioural Science and Evaluation at the University of Bristol. The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, the Department of Health and Social Care, or its partner organizations. The authors would like to thank all the care homes and staff members who took part in the study. They acknowledge the support of the NIHR Applied Research Collaboration Greater Manchester and the Enabling Research in Care Homes network with the recruitment of care homes. They thank Professor Lucy

XSL•FO
RenderX

Yardley, Professor Mike Moore, Professor Paul Little, Professor Alistair Hay, Dr Sarah Denham, and members of the original Germ Defence development team for their support of this project.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

AH, LM, MW, CC, and CT developed the initial study design and acquired funding. CC conducted the stakeholder consultations in phase 2. AH and JA-M recruited care homes and participants in phases 3 and 4. JA-M collected the interview data in phases 3 and 4. AH, JA-M, LM, CE-T, CC, and CT conducted analyses across each phase and agreed upon changes to the Germ Defence for Care Homes website in discussion with JD-D, who edited the intervention website coding. AH led the drafting of the manuscript with support from LM and CE-T. All authors provided comments and edits to drafts and approved the final version. CT is the principal investigator.

Conflicts of Interest

JD-D wrote the code for Germ Defence, and JD-D and MW have been involved in previous work developing and evaluating Germ Defence.

References

1. Health care without avoidable infections: the critical role of infection prevention and control. World Health Organization. 2016. URL: https://www.who.int/publications/i/item/WHO-HIS-SDS-2016.10 [accessed 2025-02-04]
2. Infection prevention and control in care homes. Care Quality Commission. 2022. URL: https://tinyurl.com/mtbyvexe [accessed 2024-02-22]
3. Older people living in care homes in 2021 and changes since 2011. Office for National Statistics. 2023. URL: https://tinyurl.com/2ee6muk2 [accessed 2024-02-22]
4. Aw J, Seng JJ, Seah SS, Low LL. COVID-19 vaccine hesitancy-a scoping review of literature in high-income countries. Vaccines 2021 Aug 13;9(8):900 [FREE Full text] [doi: 10.3390/vaccines9080900] [Medline: 34452026]
5. Greenhalgh T, Schmid MB, Czypionka T, Bassler D, Gruer L. Face masks for the public during the covid-19 crisis. BMJ 2020 Apr 09;369:m1435 [FREE Full text] [doi: 10.1136/bmj.m1435] [Medline: 32273267]
6. Martin GP, Hanna E, McCartney M, Dingwall R. Science, society, and policy in the face of uncertainty: reflections on the debate around face coverings for the public during COVID-19. Critical Public Health 2020 Aug 10;30(5):501-508. [doi: 10.1080/09581596.2020.1797997]
7. Oliver D. David Oliver: let's not forget care homes when covid-19 is over. BMJ 2020 Apr 24;369:m1629. [doi: 10.1136/bmj.m1629] [Medline: 32332026]
8. Comas-Herrera A, Marczak J, Byrd W, Lorenz-Dant K, Patel D, Pharoah D. LTCcovid International living report on COVID-19 and Long-Term Care. LTCcovid, Care Policy & Evaluation Centre, London School of Economics and Political Science. 2022. URL: https://ltccovid.org/international-living-report-covid-ltc/ [accessed 2025-02-04]
9. Morciano M, Stokes J, Kontopantelis E, Hall I, Turner AJ. Excess mortality for care home residents during the first 23 weeks of the COVID-19 pandemic in England: a national cohort study. BMC Med 2021 Mar 05;19(1):71 [FREE Full text] [doi: 10.1186/s12916-021-01945-2] [Medline: 33663498]
10. Schultze A, Nightingale E, Evans D, Hulme W, Rosello A, Bates C, et al. Mortality among Care Home Residents in England during the first and second waves of the COVID-19 pandemic: an observational study of 4.3 million adults over the age of 65. Lancet Reg Health Eur 2022 Mar;14:100295 [FREE Full text] [doi: 10.1016/j.lanepe.2021.100295] [Medline: 35036983]
11. Rajan S, Comas-Herrera A, Mckee M. Did the UK government really throw a protective ring around care homes in the COVID-19 pandemic? J Long Term Care 2020 Nov 11:185-195 [FREE Full text] [doi: 10.31389/jltc.53]
12. Gordon AL, Goodman C, Achterberg W, Barker RO, Burns E, Hanratty B, et al. Commentary: COVID in care homes-challenges and dilemmas in healthcare delivery. Age Ageing 2020 Aug 24;49(5):701-705 [FREE Full text] [doi: 10.1093/ageing/afaa113] [Medline: 32402088]
13. Shallcross L, Burke D, Abbott O, Donaldson A, Hallatt G, Hayward A, et al. Factors associated with SARS-CoV-2 infection and outbreaks in long-term care facilities in England: a national cross-sectional survey. Lancet Healthy Longevity 2021 Mar;2(3):e129-e142. [doi: 10.1016/s2666-7568(20)30065-9]
14. Dixon J, Lorenz-Dant K, Stubbs E, Harrison Dening K, Mostaghim M, Casson D. Registered care home managers' experiences of responding to the national care home visiting guidance in England during the Covid-19 pandemic; a multi-method qualitative study. BMC Geriatr 2023 Apr 19;23(1):236 [FREE Full text] [doi: 10.1186/s12877-023-03935-w] [Medline: 37076808]

XSL•FO
RenderX

15. Loveday H. Infection prevention and control should be applied as a force for good: minimising restrictions in residential and nursing care. J Infect Prev 2020 Nov 14;21(6):210-211 [FREE Full text] [doi: 10.1177/1757177420977224] [Medline: 33408757]
16. Kuylen M, Wyllie A, Bhatt V, Fitton E, Michalowski S, Martin W. COVID-19 and the Mental Capacity Act in care homes: perspectives from capacity professionals. Health Soc Care Community 2022 Sep 09;30(5):e3018-e3028 [FREE Full text] [doi: 10.1111/hsc.13747] [Medline: 35138011]
17. Cassell JA, Middleton J, Nalabanda A, Lanza S, Head MG, Bostock J, et al. Scabies outbreaks in ten care homes for elderly people: a prospective study of clinical features, epidemiology, and treatment outcomes. Lancet Infect Dis 2018 Aug;18(8):894-902. [doi: 10.1016/s1473-3099(18)30347-5]
18. Utsumi M, Makimoto K, Quroshi N, Ashida N. Types of infectious outbreaks and their impact in elderly care facilities: a review of the literature. Age Ageing 2010 May 23;39(3):299-305. [doi: 10.1093/ageing/afq029] [Medline: 20332371]
19. Petrignani M, van Beek J, Borsboom G, Richardus JH, Koopmans M. Norovirus introduction routes into nursing homes and risk factors for spread: a systematic review and meta-analysis of observational studies. J Hosp Infect 2015 Mar;89(3):163-178. [doi: 10.1016/j.jhin.2014.11.015] [Medline: 25601744]
20. Houghton C, Meskell P, Delaney H, Smalle M, Glenton C, Booth A, et al. Barriers and facilitators to healthcare workers' adherence with infection prevention and control (IPC) guidelines for respiratory infectious diseases: a rapid qualitative evidence synthesis. Cochrane Database Syst Rev 2020 Apr 21;4(4):CD013582 [FREE Full text] [doi: 10.1002/14651858.CD013582] [Medline: 32315451]
21. Luijnenburg O, Samsi K, Kessler I, Norrie C, Martineau S, Manthorpe J. 'I wasn't on the front line per se, but I was part of health care': contributions and experiences of ancillary staff in care homes in England during the COVID-19 pandemic. J Health Serv Res Policy 2024 Jul 20;29(3):143-152 [FREE Full text] [doi: 10.1177/13558196241246178] [Medline: 38642016]
22. Smith A, Carusone SC, Loeb M. Hand hygiene practices of health care workers in long-term care facilities. Am J Infect Control 2008 Sep;36(7):492-494. [doi: 10.1016/j.ajic.2007.11.003] [Medline: 18786453]
23. Lee MH, Lee GA, Lee SH, Park YH. A systematic review on the causes of the transmission and control measures of outbreaks in long-term care facilities: back to basics of infection control. PLoS One 2020 Mar 10;15(3):e0229911 [FREE Full text] [doi: 10.1371/journal.pone.0229911] [Medline: 32155208]
24. NIHR: the effectiveness of PPE in reducing the transmission of COVID-19 in health and social care settings: December 2021 update, 12 December 2021. United Kingdom Government. 2021 Dec 23. URL: https://tinyurl.com/y2c9emvd [accessed 2025-02-04]
25. Sandbekken IH, Hermansen Å, Utne I, Grov EK, Løyland B. Students' observations of hand hygiene adherence in 20 nursing home wards, during the COVID-19 pandemic. BMC Infect Dis 2022 Feb 14;22(1):156 [FREE Full text] [doi: 10.1186/s12879-022-07143-6] [Medline: 35164685]
26. Daker-White G, Panagioti M, Giles S, Blakeman T, Moore V, Hall A, et al. Beyond the control of the care home: a meta-ethnography of qualitative studies of infection prevention and control in residential and nursing homes for older people. Health Expect 2022 Oct 21;25(5):2095-2106 [FREE Full text] [doi: 10.1111/hex.13349] [Medline: 34420254]
27. Willcox ML, Lavu D, Yousaf U, Dalton S, Roberts N, Plüddemann A. The effectiveness of interventions to reduce the transmission of acute respiratory infections in care homes: a systematic review. J Public Health 2024 Dec 01;46(4):551-563. [doi: 10.1093/pubmed/fdae178] [Medline: 39136085]
28. Improving infection prevention and control at the health facility: an interim practical manual. World Health Organization. 2018 Jan 3. URL: https://www.who.int/publications/i/item/WHO-HIS-SDS-2018.10 [accessed 2025-02-04]
29. Lee MH, Lee GA, Lee SH, Park YH. Effectiveness and core components of infection prevention and control programmes in long-term care facilities: a systematic review. J Hosp Infect 2019 Aug;102(4):377-393 [FREE Full text] [doi: 10.1016/j.jhin.2019.02.008] [Medline: 30794854]
30. Gray J. Infection control: beyond the horizon. J Hosp Infect 2015 Apr;89(4):237-240. [doi: 10.1016/j.jhin.2014.11.004] [Medline: 25586986]
31. Huang F, Brouqui P, Boudjema S. How does innovative technology impact nursing in infectious diseases and infection control? A scoping review. Nurs Open 2021 Sep 25;8(5):2369-2384 [FREE Full text] [doi: 10.1002/nop2.863] [Medline: 33765353]
32. Dykgraaf SH, Matenge S, Desborough J, Sturgiss E, Dut G, Roberts L, et al. Protecting nursing homes and long-term care facilities from COVID-19: a rapid review of international evidence. J Am Med Dir Assoc 2021 Oct;22(10):1969-1988 [FREE Full text] [doi: 10.1016/j.jamda.2021.07.027] [Medline: 34428466]
33. Mahmood S, Hasan K, Colder Carras M, Labrique A. Global preparedness against COVID-19: we must leverage the power of digital health. JMIR Public Health Surveill 2020 Apr 16;6(2):e18980 [FREE Full text] [doi: 10.2196/18980] [Medline: 32297868]
34. Taj F, Klein MC, van Halteren A. Digital health behavior change technology: bibliometric and scoping review of two decades of research. JMIR Mhealth Uhealth 2019 Dec 13;7(12):e13311 [FREE Full text] [doi: 10.2196/13311] [Medline: 31833836]
35. Morton K, Ainsworth B, Miller S, Rice C, Bostock J, Denison-Day J, et al. Adapting behavioral interventions for a changing public health context: a worked example of implementing a digital intervention during a global pandemic using rapid

XSL•FO
RenderX

optimisation methods. Front Public Health 2021 Apr 26;9:668197 [FREE Full text] [doi: 10.3389/fpubh.2021.668197] [Medline: 33981669]
36. Yardley L, Miller S, Teasdale E, Little P, Primit Team. Using mixed methods to design a web-based behavioural intervention to reduce transmission of colds and flu. J Health Psychol 2011 Mar;16(2):353-364. [doi: 10.1177/1359105310377538] [Medline: 20929941]
37. Yardley L, Miller S, Schlotz W, Little P. Evaluation of a web-based intervention to promote hand hygiene: exploratory randomized controlled trial. J Med Internet Res 2011 Dec 09;13(4):e107 [FREE Full text] [doi: 10.2196/jmir.1963] [Medline: 22155673]
38. Little P, Stuart B, Hobbs FD, Moore M, Barnett J, Popoola D, et al. An internet-delivered handwashing intervention to modify influenza-like illness and respiratory infection transmission (PRIMIT): a primary care randomised trial. The Lancet 2015 Oct;386(10004):1631-1639. [doi: 10.1016/s0140-6736(15)60127-1]
39. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
40. Ainsworth B, Steele M, Stuart B, Joseph J, Miller S, Morrison L, et al. Using an analysis of behavior change to inform effective digital intervention design: how did the PRIMIT website change hand hygiene behavior across 8993 users? Ann Behav Med 2017 Jun 1;51(3):423-431 [FREE Full text] [doi: 10.1007/s12160-016-9866-9] [Medline: 27909944]
41. Ainsworth B, Miller S, Denison-Day J, Stuart B, Groot J, Rice C, et al. Infection control behavior at home during the COVID-19 pandemic: observational study of a web-based behavioral intervention (Germ Defence). J Med Internet Res 2021 Feb 25;23(2):e22197 [FREE Full text] [doi: 10.2196/22197] [Medline: 33566791]
42. Horwood J, Chalder M, Ainsworth B, Denison-Day J, de Vocht F, Elwenspoek MM, et al. Primary care implementation of Germ Defence, a digital behaviour change intervention to improve household infection control during the COVID-19 pandemic: a structured summary of a study protocol for a randomised controlled trial. Trials 2021 Apr 09;22(1):263 [FREE Full text] [doi: 10.1186/s13063-021-05188-7] [Medline: 33836825]
43. Treneman-Evans G, Ali B, Denison-Day J, Clegg T, Yardley L, Denford S, et al. The rapid adaptation and optimisation of a digital behaviour-change intervention to reduce the spread of COVID-19 in schools. Int J Environ Res Public Health 2022 May 31;19(11):6731 [FREE Full text] [doi: 10.3390/ijerph19116731] [Medline: 35682312]
44. Germ Defence. University of Southampton. URL: https://germdefenceschools.bss.design/ [accessed 2025-02-04]
45. Bradbury K, Morton K, Band R, van Woezik A, Grist R, McManus RJ, et al. Using the person-based approach to optimise a digital intervention for the management of hypertension. PLoS One 2018;13(5):e0196868 [FREE Full text] [doi: 10.1371/journal.pone.0196868] [Medline: 29723262]
46. Yardley L, Ainsworth B, Arden-Close E, Muller I. The person-based approach to enhancing the acceptability and feasibility of interventions. Pilot Feasibility Stud 2015 Oct 26;1(1):37 [FREE Full text] [doi: 10.1186/s40814-015-0033-z] [Medline: 27965815]
47. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022 Jan;292:114523 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
48. Pan Z, Cui Y, Leighton JP, Cutumisu M. Insights into computational thinking from think‐aloud interviews: a systematic review. Appl Cogn Psychol 2022 Dec 23;37(1):71-95. [doi: 10.1002/acp.4029]
49. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
50. Payment guidance for researchers and professionals. National Institute for Health and Care Research. 2024 Dec 5. URL: https://www.nihr.ac.uk/payment-guidance-researchers-and-professionals [accessed 2025-02-04]
51. Ainsworth B, Horwood J, Walter SR, Miller S, Chalder M, De Vocht F, OpenSAFELY Collaborative, et al. Implementing Germ Defence digital behaviour change intervention via all primary care practices in England to reduce respiratory infections during the COVID-19 pandemic: an efficient cluster randomised controlled trial using the OpenSAFELY platform. Implement Sci 2023 Dec 04;18(1):67 [FREE Full text] [doi: 10.1186/s13012-023-01321-z] [Medline: 38049846]
52. Neergaard MA, Olesen F, Andersen RS, Sondergaard J. Qualitative description - the poor cousin of health research? BMC Med Res Methodol 2009 Jul 16;9:52 [FREE Full text] [doi: 10.1186/1471-2288-9-52] [Medline: 19607668]
53. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
54. Shachak A, Kuziemsky C, Petersen C. Beyond TAM and UTAUT: future directions for HIT implementation research. J Biomed Inform 2019 Dec;100:103315 [FREE Full text] [doi: 10.1016/j.jbi.2019.103315] [Medline: 31629923]
55. Van Dort BA, Carland JE, Penm J, Ritchie A, Baysari MT. Digital interventions for antimicrobial prescribing and monitoring: a qualitative meta-synthesis of factors influencing user acceptance. J Am Med Inform Assoc 2022 Sep 12;29(10):1786-1796 [FREE Full text] [doi: 10.1093/jamia/ocac125] [Medline: 35897157]
56. Gasteiger N, van der Veer SN, Wilson P, Dowding D. How, for whom, and in which contexts or conditions augmented and virtual reality training works in upskilling health care workers: realist synthesis. JMIR Serious Games 2022 Feb 14;10(1):e31644 [FREE Full text] [doi: 10.2196/31644] [Medline: 35156931]

XSL•FO
RenderX

57. Mäkinen H, Haavisto E, Havola S, Koivisto JM. User experiences of virtual reality technologies for healthcare in learning: an integrative review. Behav Inf Technol 2020 Jul 01;41(1):1-17. [doi: 10.1080/0144929x.2020.1788162]
58. Ruthenbeck GS, Reynolds KJ. Virtual reality for medical training: the state-of-the-art. J Simul 2017 Dec 19;9(1):16-26. [doi: 10.1057/jos.2014.14]
59. Gasteiger N, van der Veer SN, Wilson P, Dowding D. Virtual reality and augmented reality smartphone applications for upskilling care home workers in hand hygiene: a realist multi-site feasibility, usability, acceptability, and efficacy study. J Am Med Inform Assoc 2023 Dec 22;31(1):45-60 [FREE Full text] [doi: 10.1093/jamia/ocad200] [Medline: 37846147]
60. Braun V, Clarke V. Thematic Analysis: A Practical Guide. Thousand Oaks, CA: SAGE Publications; 2021.
61. May C, Finch T. Implementing, embedding, and integrating practices: an outline of normalization process theory. Sociology 2009 Jun 15;43(3):535-554. [doi: 10.1177/0038038509103208]
62. What is the UK Covid-19 inquiry? UK Covid-19 Inquiry. URL: https://covid19.public-inquiry.uk/ [accessed 2024-02-22]
63. Module 1 report: the resilience and preparedness of the United Kingdom: a report by The Rt Hon the Baroness Hallett DBE Chair of the UK Covid-19 Inquiry. UK Covid-19 Inquiry. 2024. URL: https://tinyurl.com/ycx4vfnr [accessed 2025-02-04]

Abbreviations

GDCH: Germ Defence for Care Homes
IPC: infection prevention and control
PPE: personal protective equipment
PRIMIT: Primary Care Trial of a Website-Based Infection Control Intervention to Modify Influenza-like Illness and Respiratory Infection Transmission
UTAUT: Unified Theory of Acceptance and Use of Technology
VR: virtual reality

Edited by A Mavragani; submitted 23.09.24; peer-reviewed by J Dixon, I Wilson; comments to author 08.11.24; revised version received 28.11.24; accepted 29.11.24; published 19.02.25.

Please cite as:
Hall A, Aguilera-Muñoz J, McGarrigle L, Eost-Telling C, Denison-Day J, Cabral C, Willcox M, Todd C
Adapting the Germ Defence Web-Based Intervention to Improve Infection Prevention and Control in Care Homes: Interview Study Among Care Home Staff
JMIR Form Res 2025;9:e66706
URL: https://formative.jmir.org/2025/1/e66706
doi:10.2196/66706
PMID:

©Alex Hall, Johanna Aguilera-Muñoz, Lisa McGarrigle, Charlotte Eost-Telling, James Denison-Day, Christie Cabral, Merlin Willcox, Chris Todd. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Designing Values Elicitation Technologies for Mental Health and Chronic Care Integration: User-Centered Design Approach

Isabel R Rooper[1,2*], BA; William W Liem[1,3*], MSW; Martha Burla[1], MPH; Jacob Gordon[4], MSW, PhD; Lara M Baez[2,5], PhD; Rachel Kornfield[2,5], PhD; Andrew B L Berry[1,2], PhD

[1]Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[2]Center for Behavioral Intervention Technologies, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[3]Institute for Sexual and Gender Minority Health and Wellbeing, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[4]College of Nursing, University of Cincinnati, Cincinnati, OH, United States

[5]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[*]these authors contributed equally

Corresponding Author:
Andrew B L Berry, PhD
Department of Medical Social Sciences
Northwestern University Feinberg School of Medicine
625 N Michigan Ave
Ste 2100
Chicago, IL, 60611
United States
Phone: 1 312 503 4910
Email: andrew.berry@northwestern.edu

Abstract

Background: Individuals with multiple chronic conditions (MCCs) and mental health challenges such as depression or anxiety have complex health needs and experience significant challenges with care coordination. Approaches to enhance care for patients with MCCs typically focus on eliciting patients' values to identify and align treatment priorities across patients and providers. However, these efforts are often hindered by both systems- and patient-level barriers, which are exacerbated for patients with co-occurring mental health symptoms. Technology-enabled services (TES) offer a promising avenue to facilitate values elicitation and promote patient-centered care for these patients, though TES have not yet been tailored to their unique needs.

Objective: This study aimed to identify design and implementation considerations for TES that facilitate values elicitation among patients with MCCs and depression or anxiety. We sought to understand the preferences of both clinicians and patients for TES that could bridge the gap between mental and physical health care.

Methods: Using human-centered design methods, we conducted 7 co-design workshops with 18 participants, including primary care clinicians, mental health clinicians, and patients with MCCs and depression or anxiety. Participants were introduced to TES prototypes that used various formats (eg, worksheets and artificial intelligence chatbots) to elicit and communicate patients' values. Prototypes were iteratively refined based on participant feedback. Data from these sessions were analyzed using reflexive thematic analysis to uncover themes related to service, technology, and implementation considerations.

Results: Three primary themes were identified. (1) Service considerations: TES should help patients translate elicited values into actionable treatment plans and include low-burden, flexible activities to accommodate fluctuations in their mental health symptoms. Both patients and clinicians indicated that TES could be valuable for improving appointment preparation and patient-provider communication through interpersonal skill-building. (2) Technology considerations: Patients expressed openness to TES prototypes that used artificial intelligence, particularly those that provided concise summaries of appointment priorities. Visual aids and simplified language were highlighted as essential features to support accessibility for neurodiverse patients. (3) Implementation considerations: Clinicians and patients favored situating values elicitation in mental health care settings over primary care and preferred self-guided TES that patients could complete independently before appointments.

Conclusions: Findings indicate that TES can address the unique needs of patients with MCCs and mental health challenges by facilitating values-based care. Key design considerations include ensuring TES flexibility to account for fluctuating mental health symptoms, facilitating skill-building for effective communication, and creating user-friendly technology interfaces. Future research

XSL•FO
RenderX

should explore how TES can be integrated into health care settings to enhance care coordination and support patient-centered treatment planning. By aligning TES design with patient and clinician preferences, there is potential to bridge gaps in care for this complex patient population.

(JMIR Form Res 2025;9:e68419) doi:10.2196/68419

KEYWORDS

chronic care management; anxiety; depression; values elicitation; eHealth; patient-centered care; technology-enabled services; human-centered design; multimorbidity; user-centered design; multiple chronic conditions

Introduction

Approximately 42% of adults in the United States have multiple chronic conditions (MCCs), commonly including diabetes, heart disease, and hypertension [1]. Managing MCCs is a persistent challenge for both patients and their health care providers due to the complex health care needs of these patients. Care approaches that target one condition at a time have been deemed ineffective in addressing the complexities of competing health priorities [2,3], which leaves patients susceptible to gaps in care [4]. Specifically, recommended treatments and clinical workflows for each condition may conflict, causing patients to feel overwhelmed and unsupported in managing their complex health conditions. Patients with MCCs are also burdened with tracking and sharing medical information between providers, which complicates their ability to effectively communicate their health needs and priorities across their various care team members [5]. These challenges are intensified when patients' chronic conditions include mental health challenges such as depression or anxiety. Symptoms such as amotivation, low mood, and concentration difficulties associated with depression or anxiety can impair patients' decision-making capacity [6], making it challenging for them to identify and articulate their health priorities during values elicitation. These symptoms can lead to cognitive and emotional barriers, such as limited working memory capacity that can impair patients' ability to reflect and communicate their values, which are critical steps for informed, patient-centered care planning [7] that centralizes what patients consider most important for their health and well-being [8,9]. The traditional separation between physical and mental health care providers further complicates coordinated care, exacerbating the challenges these patients face in aligning their care with their personal values [10].

Approaches such as the Patient Priorities Care model [11] have been developed to enhance care for patients with MCCs. Such approaches focus on identifying and aligning health priorities between patients and providers to better address their respective needs [12]. Common strategies include eliciting patients' values to facilitate patient-provider discussions about priorities and inform patient-centered care management. Structured communication tools have shown promise in supporting patients with MCCs in incorporating their values into their treatment plans, including technology-enabled acceptance and commitment therapy and self-efficacy–enhancing interviewing techniques [13]. Despite this progress, current interventions still fail to meet the unique requirements of patients who also experience depression or anxiety. For instance, existing interventions do not adequately account for the ways in which mental health symptoms may impact patients' capacity to engage with values elicitation activities [14]. Prior research in eliciting patients' values among primary care physicians has also revealed significant communication barriers due to patient-level constraints (eg, overlapping symptoms, capacity to engage) and the perceived irrelevance of values in health care [15]. Structural barriers, such as time constraints and competing clinical priorities, have also led to low intervention uptake among providers [16], which signals the need for values elicitation interventions that are designed for implementation.

These barriers underscore the need for efficient interventions that (1) can be implemented within clinicians' workflows, (2) are tailored to patients with MCCs and depression or anxiety, and (3) establish the relevance of values to health care. Recent work that explored the challenges faced by patients with MCCs and depression or anxiety in communicating their values to providers highlighted the need for user-friendly, flexible technology-enabled services (TES)—digital tools and platforms that enhance health care delivery by integrating technology into care—that facilitate values elicitation across both mental health and primary care settings [14]. Research is needed to further clarify patients' and providers' design preferences for values elicitation TES tailored to this population. Building on insights from past work focusing on patients' preferences and needs in using TES for values elicitation [14], this study expands the scope to incorporate providers' preferences and needs with respect to TES for values elicitation.

Human-centered design processes iteratively engage key parties (eg, patients and providers) through interactive methods such as co-design workshops to understand their goals, challenges, and motivations, and produce TES that adequately address key parties' needs [17]. Given that implementation challenges have often limited the real-world impact of TES [18], implementation considerations should also be centered throughout design processes. The Accelerated Creation-to-Sustainment (ACTS) model [18] is a framework for expediting and improving the development and implementation of digital health solutions by integrating insights from human-computer interaction, implementation science, and clinical trial methodologies. The ACTS model aggregates design considerations into 3 groups: service, technology, and implementation. "Service" encompasses the behavioral strategies facilitated by digital tools and the expected roles of providers and patients; "technology" refers to the technologies that enable service delivery; and "implementation" involves methods to integrate digital tools into clinical practice, as well as broader contextual factors. This study used a human-centered design approach, informed by the ACTS model, to define design and implementation

XSL•FO
RenderX

considerations for TES that help patients articulate their values and support providers in identifying and acting upon those values collaboratively.

Methods

Recruitment

Participants included both providers and patients. English-speaking providers with experience in primary care or mental health settings were recruited from a large medical-academic center in the Midwest United States via study advertisements and emails. Additional providers were reached through snowball sampling via personal contacts. Demographic information relevant to the foci of this study, such as role and department, was collected from providers. However, other demographic data were not gathered, as this analysis did not intend to assess demographic characteristics and their relationship to clinicians' preferences regarding values elicitation TES. Patient participants were English-speaking adults with at least 2 self-reported chronic medical conditions and a history of depression or anxiety. The study prioritized engaging individuals with current or prior lived experience balancing mental and physical health challenges in order to capture nuanced insights into how these conditions intersect with chronic illness management and values elicitation. This approach aligns with human-centered design principles, emphasizing the value of incorporating a wide range of perspectives to inform participatory design [19]. Eligible patients were identified via institutional research recruitment registries and invited to participate in the study via email. Trained study personnel then scheduled interested individuals for workshops. Twelve providers and 6 patients verbally granted informed consent and participated in workshops.

Ethical Considerations

This study was approved by the Northwestern University institutional review board (STU00212476). Informed consent was obtained from all participants prior to participation in the study. Participants were provided with detailed information about the study's purpose, procedures, potential risks, and benefits, as well as assurances of confidentiality and the voluntary nature of their involvement. Participants were compensated with a US $50 gift card for their participation. Study data were deidentified to ensure participant privacy and confidentiality. Each participant was assigned a unique identification number for analysis and reporting purposes, and all personally identifying information was removed before sharing data outside the research team. Study data were securely stored in Health Insurance Portability and Accountability Act–compliant electronic locations.

Workshops and Prototypes

Overview

Five provider workshops (workshops 1-5) and 2 patient workshops (workshops 6-7) were conducted via Zoom (Zoom Video Communications Inc) between February 2023 and March 2024 (Table 1). Researchers can identify themes and evaluate the most and least promising design directions with as few as 5-6 qualitative interviews [20]. Each workshop involved discussing and critiquing prototypes to explore specific aspects of TES design and implementation. Workshops lasted 60 minutes and included 4-6 participants (primary care providers [PCPs], mental health providers [MHPs], or patients) and 2 researchers (ABLB, JG, or WWL). At the beginning of each workshop, the researchers introduced themselves and established the workshop goals, confidentiality procedures, and prototype rationale. Prototypes shown in each workshop were tailored to address specific design questions, representing various values elicitation methods that were iteratively refined based on feedback from earlier workshops. Prototypes were intended as "thinking tools" to elicit design considerations, rather than high-fidelity prototypes to be optimized (see Table 2 for an overview of prototypes and Multimedia Appendix 1 for example prototypes) [14].

XSL•FO
RenderX

Table 1. Summary of workshop phases, objectives, activities, and prototypes used to explore clinicians' and patients' preferences for technology-enabled services for values elicitation, with prototypes tailored to specific goals in each phase.

Phase	Objectives	Activities	Prototypes	Purpose of prototypes
Phase 1 (N=4): Building buy-in for values elicitation (MHPs[a] and PCPs[b])	Explore how to communicate values elicitation rationale to patients.	• Introduced rationale and workflow for values elicitation. • Participants designed invitations to patients to engage in values elicitation.	Values bull's eye worksheet [21]	Provide an example of a structured reflection tool on values and health priorities.
Phase 2 (N=4): Facilitating values elicitation (MHPs)	Understand how to facilitate discussions about patients' values and their application in care planning.	• Reviewed prototypes for eliciting and reflecting on values. • Facilitated discussions on how to implement values elicitation in patient interactions. • Drafted messages for MHPs to share values with PCPs.	Values bull's eye worksheet	Provide an example of a structured reflection tool on values and health priorities.
			Premeeting with an MHP	Elicit reaction to completing a values exercise *before* meeting with an MHP.
			Story-sharing	Elicit reaction to completing a values exercise *while* meeting with an MHP.
Phase 3 (N=4): Shared care planning (PCPs)	Specify how patient values should be communicated to clinicians. Evaluate strategies and tools to enhance the effectiveness of shared care planning.	• Critiqued patient scenarios and prototypes. • Identified barriers to implementing shared care planning. • Brainstormed templates and workflows for sharing values with PCPs.	Values bull's eye worksheet	Provide an example of a structured reflection tool on values and health priorities.
			Inbox storyboard	Evaluate actionability of patients' values in PCP visits.
			Previsit summary	Understand whether preappointment preparation could center patient values without adding significant workload burden.
			Workflow diagram	Illustrates care coordination between primary and mental health to explore proposed integrations.
Phase 3 (N=6): Shared care planning (patients)	Explore prototype methods to address challenges in identifying and communicating values, such as prioritizing health concerns, aligning values with treatment.	• Explored patient-centered approaches for sharing values. • Discussed patient feedback on interactive prototypes.	Story-sharing	Elicit reaction to completing a values exercise *while* meeting with an MHP.
			PCP simulator	Understand what types of communication-building support would be needed to communicate values to PCPs.

[a]MHPs: mental health providers.

[b]PCPs: primary care providers.

XSL•FO
RenderX

Table 2. Values elicitation technology–enabled service (TES) prototypes were shown to clinicians and patients to elicit their preferences for TES design and implementation[a].

Workshop displayed	Prototype name and summary
W1-W3	Values bull's eye worksheet [21]: Established tool from acceptance and commitment therapy to help patients reflect on their values via a mobile interface.
W2	Inbox storyboard: Workflow prototype in which patients complete values elicitation activities preappointment and mental health providers receive an inbox message with the results.
W4	Previsit summary: Document available in patients' electronic health records listing patients' personal values and appointment-related questions.
W4	Workflow diagram: Graphic displaying workflows to facilitate collaborative care management across mental and primary health care.
W5	Premeeting with an MHP[b]: Patients review self-care behaviors and assess their alignment with their values before meeting with an MHP; results electronically transmit to MHPs.
W5-W7	Story-sharing: Patients are emailed an exercise to complete preappointment and then meet with an MHP to reflect on their values and appointment priorities.
W6-W7	PCP[c] simulator: Patients use an AI[d] chatbot, playing the role of their PCP, to practice communicating their values to a dismissive PCP and receive takeaways (eg, communication tips and appointment priorities).

[a]The prototypes were used as "thinking tools" during the workshops.

[b]MHP: mental health provider.

[c]PCP: primary care provider.

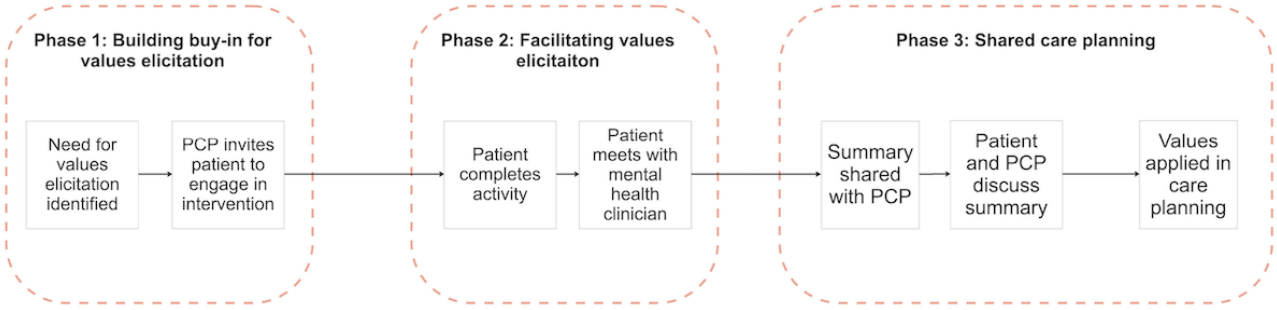
[d]AI: artificial intelligence.

Prototype Development

Building on past work exploring the design requirements to support patients with MCC and anxiety or depression to identify and share their values [14], the team identified three key principles needed to support this population: (1) Credibility: the person eliciting the patient's values has the training and experience to do so effectively and with beneficence. (2) Trustworthiness: sharing the patient's values with the PCP will only benefit—and not negatively affect—the patient's care. (3) Time worthiness: engaging in a reflection exercise to identify and communicate one's values should offer meaningful benefits for their care, given competing demands on patients' time and capacity (eg, appointments with multiple specialists).

The team generated a high-level service flow illustrating how values could be elicited and then applied in shared care planning (Figure 1). The service flow included 3 phases for conducting values elicitation with MHPs and PCPs.

Figure 1. High-level overview of values elicitation service design, which guided the formulation of the technology-enabled prototypes that were shown to clinicians and patients as part of this qualitative study. PCP: primary care provider.



The first phase, building buy-in for values elicitation, focused on strategies to communicate the rationale for values elicitation to patients, as well as the credibility, trustworthiness, and timeworthiness of various approaches to values elicitation. Perspectives from both PCPs and MHPs were elicited to understand how to build trust and demonstrate the value of values elicitation from both physical and mental health care perspectives.

The second phase, facilitating values elicitation, aimed to refine the activities and technologies through which values could be elicited. Since previous research indicated that patients in this population often preferred engaging in values elicitation with MHPs [14], this phase centered on gathering MHPs' perspectives on how to structure and instrument values elicitation activities.

Finally, the third phase, shared care planning, focused on applying elicited patient values in shared care planning between patients and PCPs, including possible implementation barriers and facilitators. PCPs' insights were sought in this phase to identify workflow opportunities and challenges, given their central role in coordinating care across diverse clinical settings.

XSL•FO
RenderX

Prototypes were then developed to illustrate alternatives within and across these phases. For example, in phase 2, prototypes explored MHP-driven values elicitation versus patient-driven, self-guided values elicitation. By seeking confirmatory and refuting evidence from multiple participant groups in response to a range of possible design directions, this approach increased the rigor of the design research process.

Iterative Design Workshops

While workshops were being conducted, the research team met weekly to discuss and enumerate the insights garnered through each workshop, align them with the research questions and workshop objectives, and use these insights to guide prototype development for pursuant workshops. Accordingly, prototypes shown in later workshops responded to the input and critique from participants in earlier workshops.

Following the provider workshops, the research team refined 2 prototypes that were then presented to patients in 2 workshops to understand how their preferences aligned with providers' preferences. These 2 prototypes integrated insights from provider workshops and previous research with this patient population [14], which found that patients were open to values elicitation activities facilitated by MHPs, as well as self-guided values elicitation. This prior work additionally found that while patients generally know what they want to discuss with providers, they often lack the communication skills to convey these priorities effectively. Accordingly, two prototype options were created, listed in Multimedia Appendix 1: (1) a story-sharing prototype with an MHP and (2) a "PCP simulator" focused on building patients' communication skills. The artificial intelligence (AI) tool within the PCP simulator was well suited to this task, as it generated text tailored to different patient-provider scenarios, helping patients practice communicating their values and priorities with a range of provider responses.

Analyses

This analysis combined 2 complementary approaches: iterative prototype analysis to refine specific aspects of a TES, and thematic analysis to generate insights across workshops. Together, these approaches provided a robust framework for exploring values elicitation and shared care planning in diverse health care settings. Grounded in reflection-in-action [22] and research through design [23], this dual focus allowed us to rigorously address the complexities of designing and implementing values-based health care.

Iterative Refinement of Tools

Throughout this iterative process, prototypes served as "thinking tools" to elicit participants' feedback and address emerging questions about the design space. This approach aligns with research through design methodologies, where artifacts are both products of research and vehicles for generating insights. Prototype analysis was informed by reflection-in-action principles, emphasizing adaptive learning and real-time evaluation. Each phase of the workshops was analyzed iteratively to refine prototypes and activities.

For example, feedback was gathered on initial concepts to identify areas where tools could better address user needs. This feedback informed the iterative development of prototypes, with adjustments made to enhance their usability and relevance. Later, barriers to integrating prototypes into existing workflows were examined. Insights from this process guided further refinements to ensure that the tools aligned with practical implementation requirements.

Thematic Analysis

Workshops were recorded via Zoom and transcribed. Transcripts were then deidentified and edited for accuracy. Finalized transcripts were qualitatively analyzed for themes using reflexive thematic analysis [24,25] to synthesize insights across workshops. This analysis involved the following steps:

1. Codebook development: Lead coauthors IRR and WL reviewed transcripts and workshop insights to create an initial codebook. Five coders then applied this codebook to a sample transcript, refining it collaboratively to resolve discrepancies.
2. Coding and memoing: Coders independently applied the refined codebook to transcripts using Dedoose (SocioCultural Research Consultants LLC) and wrote memos to capture emerging themes.
3. Collaborative synthesis: Weekly team meetings facilitated the development of themes, which were subsequently grouped using the ACTS model into design considerations for service, technology, and implementation.

This thematic analysis triangulated stakeholder perspectives by engaging participants across clinical roles and settings to validate findings and highlight role-specific nuances. Analysts' reflexivity, shaped by expertise in health education, technology design, and lived experience managing chronic conditions enhanced these findings. By combining iterative prototyping with reflexive thematic analysis, this methodology was intentionally designed to bridge theory and practice. Prototypes generated actionable insights for refining design elements, while thematic analysis captured the broader social and systemic contexts in which these tools would be implemented. This dual approach ensured that these findings were both grounded in stakeholder needs and positioned to contribute to future research and practice in values-based health care.

Results

Participants

From the primary care domain, participants included 4 physicians, 1 internal medicine physician's assistant, and 1 nurse practitioner, hereafter referred to as "PCPs." From the mental health domain, participants ("MHPs") included 2 clinical psychologists, 3 behavioral care coordinators, and 1 medical social worker. This grouping reflects the participants' primary practice contexts, with PCPs focused on managing physical health and chronic conditions, and MHPs primarily engaged in addressing mental health care needs. Full sample characteristics are shown in Tables 3 and 4.

XSL•FO
RenderX

Table 3. Sample characteristics of the clinicians who participated in workshops that explored the design and implementation of technology-enabled services for values elicitation.

ID	Domain	Role	Workshop
1	Primary care	Primary care physician	W1
2	Mental health	Patient care coordinator	W1
3	Mental health	Patient care coordinator	W1
4	Mental health	Patient care coordinator	W1
5	Primary care	Internal medicine physician assistant	W2
6	Primary care	Nurse practitioner	W2
7	Mental health	Psychiatrist	W3
8	Mental health	Medical social worker	W3
9	Primary care	Primary care physician	W4
10	Primary care	Primary care physician	W4
11	Mental health	Clinical psychologist	W5
12	Mental health	Clinical psychologist	W5

Table 4. Sample characteristics of the patients who participated in workshops that explored the design and implementation of technology-enabled services for values elicitation.

ID	Gender	Race	Hispanic	Age (years)	Chronic conditions	Workshop
13	Man	White	No	66	Cancer, glaucoma, and hyperlipidemia	W6
14	Woman	White	No	82	Arthritis, kidney or bladder problems, macular degeneration, and hypertension	W6
15	Woman	White	No	66	Hyperlipidemia, restless leg syndrome, IBS[a], and arthritis	W6
18	Woman	White	No	48	Arthritis, asthma, cancer, and Hashimoto disease	W7
16	Man	African American	No	61	Diabetes, arthritis, hypertension, and chronic pain	W7
17	Woman	White	No	38	Diabetes, hyperlipidemia, asthma, autism spectrum disorder, and ADHD[b]	W7

[a]IBS: irritable bowel syndrome.

[b]ADHD: attention-deficit/hyperactivity disorder.

Service Considerations

Service considerations reflect the unique experiences and goals of patients with MCCs and depression or anxiety. Patients and providers preferred service models for values elicitation that addressed concerns about anxiety, agenda setting, and time pressure during clinical visits. For instance, patients described how managing MCCs leads to anxiety before appointments with their PCP because they feel pressure to maximize their limited time together and had poor previous experiences with dismissive providers. To alleviate this anxiety, patients wanted support preparing for appointments. Specifically, patients said that values elicitation TES should produce a tangible outcome (eg, a list of questions) supporting preparedness and serving as memory aids. This goal reflected that their anxiety could interfere with their ability to remember information during appointments, as one participant described:

> *A tool like this [the PCP simulator] would help me, especially if I forgot to put some stuff in. It'll make sure I don't miss anything that I need to share with my primary care physician.* [P16, Patient]

Likewise, providers said that balancing MCCs makes agenda-setting challenging for some patients, so TES should be designed to help patients clarify their agenda for the appointment by reflecting on their health concerns and establishing priorities to address with providers.

Communication challenges were cited by providers (eg, due to patients' mental health symptoms) and patients, who said that values elicitation TES could offer specific language to use during appointments, allowing patients to refer to pregenerated

XSL•FO
RenderX

language instead of finding the right words in the moment, which can be challenging when experiencing anxiety. MHPs and patients said that TES that facilitate interpersonal skill-building (eg, practicing effective communication strategies) could improve preappointment anxiety and preparation, as well as provider-patient communication during appointments.

Participants reacted to TES prototypes using service models that were self-guided (ie, completed independently by a patient) versus provider-facilitated (ie, collaboratively completed by a patient and a facilitator). Both patients and providers preferred self-guided over facilitated TES due to appointment time constraints. They said that patients needed sufficient time and capacity to engage with values elicitation. Patients appreciated how self-guided TES could be completed on their own schedules.

> *Best case scenario, I would prepare about a week before a physician visit. My schedule means that may or may not happen. So, having the ability to do it at whatever time, but save that information would be incredibly important for me.* [P18, Patient]

In addition, patients described how it would be challenging for them to engage with values elicitation when their mental health symptoms are high, indicating that self-guided TES, provided well in advance of appointments, may address patients' fluctuating needs and abilities.

> *My days are very different in that my level of anxiety and depression in some days dictate what I can and can't do. So, on good days, [values elicitation] is something that I would definitely be interested in doing. But on not-so-good days, probably not so much.* [P13, Patient]

Providers agreed that, if facilitated, TES should be led by MHPs due to the relevance of values to mental health care and perceived constraints of PCPs, extrapolated in the section "Implementation Considerations." One MHP said that in addition to mental health, clinicians facilitating values elicitation should understand *physical* health, in order to best support patients managing MCCs. They recalled an instance where lack of knowledge about a patient's physical health condition was a barrier to their care:

> *The Crohn's Disease was super important to improving the depression. But I had very little insight into what her treatment was for the Crohn's Disease, [which] stunts my ability to work with her.* [P12, MHP]

Across these findings, patients and providers suggested a flexible service model that accommodates patient-directed or clinician-facilitated values elicitation and emphasized the need to support previsit communication skill-building.

Technology Considerations

The majority of providers' feedback focused on service considerations ("Service Considerations" section). In terms of technology, PCPs emphasized constraints of short appointment times and competing clinical priorities, meaning that any provider-facing TES interfaces must present concise, at-a-glance information about patients' values to be acceptable. The primary requirement is to enable PCPs to identify and address the highest priority health issues in a limited time. When viewing the previsit summary prototype, one provider commented:

> *I would say there's too much information there...Because remember, this guy has diabetes, but also is obese, and also...we need to talk about medication for the depression. So, I'm grateful I have a 40-minute appointment, but how can we address this in a 20-minute?* [P9, PCP]

PCPs also said that the interface should provide clear instructions for interpreting the "results" of values elicitation, such that those results could be applied to patient care.

Providers said that the technology should integrate into a patient's electronic health record and facilitate information sharing between providers. For instance, an MHP said that they would not have time to translate a patient's values into a message to the patient's PCP, so the technology would need to facilitate that information-sharing process.

> *I don't know that I would look at a screening tool, assess it, and then filter that information for the PCP. I wouldn't have that time. So, if it's available, I would think they can look at it or not.* [P8, MHP]

Patients were shown prototypes that differed in the inclusion of AI. Patients expressed varying levels of openness to the use of AI in values elicitation TES. Participants found the use of conversational AI for skill-building and appointment preparation (as in the "PCP simulator" prototype) helpful and were generally unconcerned with the AI delivery method, although 1 participant expressed mistrust of AI and said that they would prefer that a person facilitate the same process. Another participant described a lack of familiarity with the technology.

> *I have not had a lot of personal experience with AI. But conceptually, I'm not opposed to it. And I would not be opposed to it in this particular scenario.* [P13, Patient]

Across designs, patients expressed concerns that technologies enabling values elicitation must be accessible, particularly for neurodivergent patients and those with conditions such as depression and anxiety, which can affect cognitive processing and communication. For example, incorporating emoticons was suggested to simplify complex emotions and support users in expressing their values more easily. When asked whether they preferred the story-sharing prototype or the AI chatbot, all patient participants chose the "PCP simulator" for its skill-building and appointment preparation functionalities.

Implementation Considerations

Providers said that values elicitation should be implemented in mental health care (rather than primary care) and preferred that patients use TES previsit to avoid wasting valuable appointment time. They identified specific barriers, unique to the primary care domain, that reduce the feasibility of implementing a values elicitation TES therein.

PCPs said that patients' values were more relevant to MHPs than PCPs. Due to their training and focus on physical health,

XSL•FO
RenderX

PCPs did not feel equipped or well-positioned to elicit patients' values. Generally, PCPs also perceived values as nonactionable, meaning they did not view patients' values as information they could apply in patient care. Accordingly, they preferred not seeing their patients' values enumerated unless those values were explicitly connected to a patient's primary physical health concern.

> *To me, it's just too nebulous and complex for me to use that information...or know what it actually means. And how it relates to their current chief complaint. And how I'm going to take their value and their chief complaint and validate it in a way that I wouldn't otherwise do.* [P10, PCP]

Compared with MHPs, who defined values in alignment with values elicitation literature (ie, what patients consider most important for their health and well-being [8,9]), PCPs conceptualized values in nonspecific and variable terms, at times conflating values with goals and social history.

> *I actually have done this [values elicitation] for the last 25 years; I've always been big on social history..."where did you go to high school?" Because immediately that opens up so many conversations and goes in different directions that you start to understand people's values.* [P1, PCP]

While this comment shows the provider's commitment to building relationships with patients, their misunderstanding of values elicitation complicates its implementation. It creates a false impression that PCPs are effectively identifying patients' care priorities, when in fact, the potential of values elicitation to enhance patient care has not yet been fully realized.

Conflating values and goals similarly undermines the purpose of values elicitation, as goals are typically specific and shorter-term, while values represent the deeper, longer-term priorities guiding patients' overall care decisions. For example, a goal might be "to lower blood pressure," while the underlying value could be "to maintain independence and live an active life." By focusing on clinical goals without addressing these broader values, providers risk delivering care that aligns with immediate clinical targets but neglects what truly matters to patients. As a result, patients may feel that their overall well-being is overlooked, compromising the effectiveness of patient priorities-aligned care [26].

Regarding implementation processes, providers highlighted the many moving pieces that must be accounted for when implementing TES within existing health care systems. For example, a PCP cited workflow integration concerns and said that it is difficult to implement requisite workflows to refer patients to TES and review resulting information when TES are new and infrequently used.

> *If it's only one in 12 patients, the doctor's workflow isn't going to be used to that kind of a process. And it might be less accepted by the doctor.* [P10, PCP]

PCPs and MHPs also said that structural barriers, such as short appointments, meant that implementing a new TES during an appointment would be infeasible. However, some providers were open to reviewing patients' values before an appointment, so long as the patients completed the TES independently outside their appointment time and the presentation of those values matched providers' other requirements (eg, actionability).

Discussion

Principal Findings

This study used iterative prototyping of values elicitation TES to identify the needs and preferences of providers and patients managing multiple mental and physical health conditions. By applying the ACTS model, this study identified service, technology, and implementation considerations for TES design that help patients articulate their values and support providers in identifying and acting upon those values collaboratively. These findings largely supported the Patient Priorities Care approach [11], in that participants generally perceived identifying and aligning health priorities across clinicians and patients as useful. Findings also extended the Patient Priorities Care model to apply to patients with comorbid mental and physical health conditions. In particular, these patients' mental health symptoms and related needs have not yet been well represented in values elicitation research. This population can benefit from clarifying and communicating their values to promote patient-centered care. Per these results, design recommendations are organized in accordance with the ACTS model and summarized in Table 5.

XSL•FO
RenderX

Table 5. Design recommendations for values elicitation technology–enabled services for patients with multiple chronic conditions and depression or anxiety, organized by the ACTS model's 3 domains: service, technology, and implementation.

Design recommendation	Challenge addressed	Rationale
Service		
Incorporate clear, structured reflection prompts that elicit values actionable for health treatment planning.	Difficulty translating elicited values into actionable health treatment plans.	Structured prompts help patients articulate and connect their values to their treatment planning, making their values actionable for providers.
Offer various services, including low-burden activities when mental health symptoms are high.	Mental health symptoms pose a barrier to engaging with values elicitation.	Providing a suite of tools can address patients' diverse needs and allow flexibility when patients have low capacity for engagement.
Incorporate self-guided skill-building features to handle difficult conversations.	Difficulties with communication and preappointment anxiety.	Supports self-reflection and skill-building to enhance patient communication and manage preappointment anxiety.
When relevant, involve facilitators experienced in physical and mental health.	Lack of understanding between mental and physical health care.	Facilitators with joint expertise can holistically offer values elicitation to support patients with MCCs[a].
Technology		
Create summaries of health priorities for patients to bring to appointments.	Patients experience anxiety and sometimes forget information during appointments.	Providing tangible outputs helps patients organize their priorities, reducing anxiety and helping avoid forgotten concerns.
Engage patients with intellectual disabilities in design work.	Intellectual disabilities can pose barriers to TES[b] engagement.	Centering neurodiverse patients' perspectives will improve tool accessibility.
When used, explain the clinical intent of TES that use AI[c].	Some mistrust in the AI prototype.	Despite some mistrust, patients accepted the AI prototype due to its perceived use for skill-building.
Implementation		
Establish concordance among providers about how to conceptualize values.	Discordance poses barriers, as providers must understand a TES to implement it.	Aligning providers' perceptions of the purpose and value of a TES will help facilitate TES implementation.
Design for real-world structural barriers (eg, communication constraints).	Lack of interprofessional communication infrastructure.	Designing for present care coordination challenges helps TES add value for patients and providers.
Center implementation considerations (eg, via the ACTS model) throughout TES design processes.	Difficulty fitting new TES into existing health care systems and workflows.	Engaging with implementers (eg, clinicians and staff) yields contextual considerations that could become implementation determinants.

[a]MCCs: multiple chronic conditions.

[b]TES: technology-enabled services.

[c]AI: artificial intelligence.

Service Considerations

Results indicate that TES should assist patients in preparing for difficult conversations during appointments and provide specific language to use during consultations, both of which help alleviate their preappointment anxiety. This study identified an opportunity to enhance Patient Priorities Care approaches by addressing a critical skill gap: participants expressed a need for support not only in generating values-aligned priorities but also in developing self-advocacy skills to communicate those priorities effectively during appointments. This finding highlights that TES outputs (eg, a list of questions) alone do not guarantee effective communication; rather, participants suggested that TES should actively support patient-provider communication by strengthening patients' self-advocacy skills. This finding aligns with previous work showing that mental health symptoms can make patient-provider communication particularly difficult among patients with MCCs [27]. Yet, these findings contrast with earlier work that found that patients with MCCs and depression or anxiety preferred values elicitation facilitated by an MHP, given their credentials and the potential therapeutic benefits [14]. In this study, despite presenting an MHP-directed TES prototype, patients preferred a self-directed service model, which addressed their need to prepare for difficult conversations with providers during appointments by accounting for interpersonal dynamics with providers, including dismissiveness. This preference aligns with previous research indicating that interactions with dismissive PCPs can heighten anxiety and complicate communication about health-related values and priorities [28]. Participants may therefore have prioritized this issue over deeper therapeutic conversations about values with MHPs. Given that this patient population already

XSL•FO
RenderX

experiences unique barriers to communicating their health priorities (eg, due to their overlapping symptoms), addressing these communication challenges through skill-building is a promising area for future design work and a strategy that can empower patients to engage more actively in their appointments, leading to better care coordination and collaboration [29].

The Serious Illness Conversation Guide (SICG) provides a patient-tested framework for addressing communication challenges and self-advocacy skill gaps [30], making it highly relevant for values elicitation. By emphasizing structured communication tools, the SICG guides clinicians through sensitive discussions about patient values, offering clear language and actionable steps to ensure that these conversations are patient-centered. This structured approach could inform the design of TES by incorporating tools that allow patients to practice articulating their values before appointments. Such features align with patients' preference for self-directed models, enabling them to prepare for discussions in a low-pressure setting while retaining the SICG's focus on clarity and structure. In addition, TES inspired by the SICG could include clinician training modules to improve active listening, equipping providers to better respond to patients' expressed values. These combined strategies address key interpersonal dynamics, such as the need to counter dismissiveness, while empowering patients to effectively advocate for their priorities.

Findings illustrated the need for values elicitation service models to incorporate expertise from clinicians who understand the interplay between mental and physical health. Many MHPs lack access to PCPs' medical notes that detail patients' ongoing physical health issues, which are important for understanding how symptoms may overlap [31]. This gap places the burden on patients with MCCs to clearly articulate their complex health needs to both PCPs and MHPs, a task that is particularly challenging for those experiencing depression or anxiety. Literature suggests that patients with MCCs and mental health conditions often face cognitive and emotional barriers that make it difficult to convey the full picture of their health to providers [32]. This communication gap can hinder care coordination and increase the risk of fragmented or inadequate care. To address this, it is essential for TES to not only support values elicitation but also empower patients with tools that help them communicate their physical and mental health concerns across care teams and settings. Future research should explore these tools, particularly in promoting better care collaboration and interprofessional communication between providers in managing complex care needs.

Technology Considerations

Results from this study underscore the diverse technology preferences among patients and providers, highlighting the need for a multifaceted approach to designing TES for values elicitation. Patients and providers emphasized that the technology interface must be concise and user-friendly, accommodating the limited time available during appointments. PCPs also said that they needed clear instructions for interpreting values elicitation results. These findings align with Patient Priorities Care approaches, which emphasize the need for values to be actionable to providers. Furthermore, these findings are consistent with literature showing that technology interfaces in health care must present clear, actionable information to be effective [33], as providers often face information overload, where the volume and complexity of available information exceed their capacity to process and use it [34]. In the context of values elicitation, interfaces that provide at-a-glance summaries of values and care priorities, and straightforward instructions for responding to those, can help alleviate the challenges of managing complex patient information within constrained appointment times. This finding aligns with design recommendations for eliciting patient values in clinical conversations, which emphasize that providers must be able to easily access and interpret patient values for effective care planning [35].

Patient feedback revealed that technology design must consider the cognitive and sensory needs of users, particularly those with mental health challenges or intellectual disabilities. Previous work found that cognitively and emotionally demanding values elicitation activities can be particularly challenging for individuals with depression or anxiety, suggesting that designs should balance requisite effort against potential benefits by making activities shorter, more enjoyable, and less taxing to encourage patient engagement [14]. These findings expand upon this to include accommodating diverse communication needs among patients with intellectual disabilities, as highlighted by suggestions to include visual aids such as emoticons in the interface. Existing literature on eHealth interventions for people with intellectual disabilities calls for more intentional participatory development and iteration with end users [36], emphasizing the importance of seeing the "whole picture" of both mental and physical health when designing values elicitation TES. As individuals age with MCCs, the complications of these conditions—combined with cognitive challenges such as depression, anxiety, and intellectual disabilities—may contribute to increased difficulty in engaging with health technologies. Future design work should therefore involve patients with intellectual disabilities and those managing MCCs to cocreate inclusive technologies that address the interconnected nature of physical and mental health, promoting accessibility for this historically underrepresented group.

While patients expressed general openness to using a generative AI tool to build communication skills, responses were mixed when it came to identifying and effectively communicating their values, particularly when dealing with dismissive providers. Some participants expressed uncertainty or mistrust of AI, preferring human facilitation. This hesitation is consistent with studies indicating that while AI has potential in health care, its acceptance is often tempered by concerns about trust and transparency [37,38], emphasizing the need for AI-based tools to be carefully designed and transparently integrated into patient care in order to build trust and acceptance [37]. This finding underscores the importance of clearly communicating the rationale and benefits of values elicitation TES, as patients prefer assurance about how their values will be used [14]. In the context of AI-based tools, transparency about the elicitation and application of these values becomes even more critical.

Given these varied preferences, a suite of tools may be indicated that address relevant behavioral targets (eg, appointment

XSL•FO
RenderX

preparation and preappointment anxiety) using varied technology-based approaches (eg, skill-building, virtual worksheets, and meeting with a values elicitation facilitator). Future design work should continue mapping this array of preferences and iterate TES to accommodate them.

Implementation Considerations

The approach used in this study to implement values elicitation TES reflects a broader perspective than typical Patient Priorities Care conceptions. Specifically, the research team sought to understand the various touchpoints that patients encounter across their care journeys, given their engagement with multiple providers across different settings. By expanding the perspective of inquiry beyond single conditions or discrete interactions, this study identified new design directions, particularly in understanding how to bridge the conceptualizations of values held by both patients and providers. Establishing a common understanding of values is essential to successfully implement values elicitation TES; otherwise, challenges may arise due to knowledge gaps regarding how and why to engage patients in TES for values elicitation. Future work is therefore needed to help PCPs identify the relevance of patients' values to patients' health, and potentially, to reframe values elicitation to more clearly resonate with their practice. Indeed, PCPs in this study highlighted the importance of linking patients' values to patients' health concerns to make those values actionable. This finding aligns with patients' perspectives as well, in that patients may withhold values from providers that they perceive as irrelevant to their health care [15]. Because these varied conceptualizations of values and their perceived relevance to health care pose an implementation barrier, designing TES for values elicitation requires navigating and aligning these conceptualizations. Co-designing values elicitation educational materials with providers may help improve their perceptions of the relevance of values elicitation to this patient population, as well as prepare them to articulate its purpose to patients.

Furthermore, TES must reflect real-world challenges, such as the lack of care coordination infrastructure in many health care systems. Providers in this study said that values elicitation TES should integrate seamlessly with existing electronic health records and facilitate efficient information sharing between providers, which aligns with literature indicating the importance of interoperability and streamlined communication in improving care coordination. Yet, they also cited structural barriers that prevent interprofessional communication, which means that in the current system, patients must directly communicate their values to their PCP. Patient Priorities Care services—a common, evidence-based approach to values elicitation for patients with MCCs [11]—often assume that staff, such as medical assistants, are available and trained to support PCPs in this process. However, not all health systems are equipped with the resources or staffing models to integrate these roles effectively, limiting the feasibility of this approach in certain settings. In practice, the absence of such interprofessional communication infrastructure means that patients bear the burden of conveying their values themselves, a task made complicated when their mental health symptoms limit their ability to engage with values elicitation. The applicability of Patient Priorities Care tools to patients with MCCs and depression or anxiety could therefore be improved by accounting for these implementation challenges (eg, via self-guided instead of facilitated approaches that require staff resources), as well as this patient population's unique service requirements (eg, tool responsiveness to high mental health symptoms).

Building on these challenges, the SICG emphasizes the need for system-level changes, such as integrating structured prompts into electronic health records and developing standardized documentation templates that allow clinicians to efficiently reference and act on patient values. By embedding values elicitation into routine workflows, these tools reduce reliance on additional staffing resources, making them more feasible for implementation across diverse health care settings. Furthermore, the SICG's approach to training clinicians in communication strategies could be adapted to help providers articulate the relevance of values elicitation to patients and to foster greater patient engagement. These strategies align with findings from this study, particularly the need to link values to actionable health priorities and to ensure that TES design reflects real-world structural constraints. Incorporating these SICG-inspired interventions could enhance the scalability and effectiveness of TES by streamlining workflows, linking patient values to actionable health priorities, and addressing structural constraints across diverse care environments.

Reflections on the ACTS Model

The ACTS model provided a vital framework for generating these findings, particularly in addressing the complex needs of patients with multiple mental and physical health conditions. By focusing on service, technology, and implementation, the model helped identify the design and implementation needs of values elicitation TES. It emphasized the importance of empowering patients in their health care interactions, creating user-friendly tools that accommodate diverse needs, and addressing the structural barriers that hinder care coordination. The ACTS model's structured approach ensured that TES design is both practical and effective, facilitating seamless integration into the existing health care landscape, and ultimately, meeting the needs of this complex patient population.

Limitations

Because this work principally engaged providers from a large medical-academic center, the implementation considerations identified (eg, lack of interprofessional communication and workflow integration concerns) may not apply to other clinical contexts. Per the ACTS model, future work should continue designing TES with institution-specific implementation considerations in mind, which other designers can identify by using similar methods to this study. In addition, the providers who participated in these workshops were interested in values elicitation; their preferences may therefore not be transferable to other contexts.

Due to the small sample size and highly iterative prototypes, consensus was not achieved regarding preferred TES design features or approaches. Despite this, findings offer distinct, novel, and significant insights to inform future TES design.

Sample characteristics pose another limitation. Demographic data were not collected from providers. Future work must collect

XSL•FO
RenderX

providers' demographics to address the potential influence of misalignments between patient and clinician identities on care delivery, as these differences may affect how values are perceived and prioritized in clinical interactions. In addition, most patient participants self-identified as White and non-Hispanic. The lack of racial and ethnic diversity among patients limits the generalizability of these results and must be addressed in subsequent research. Future studies would also benefit from assessing patients' current levels of depression or anxiety to learn how symptom severity may shape patients' preferences.

Implications and Future Directions

This study found opportunities to address the unique needs of patients with MCCs and depression or anxiety by designing TES that account for patients' fluctuating mental health symptoms, build patients' communication skills, and facilitate care collaboration across mental and physical health care. Next steps from this study include conducting continued design work with providers, such as service journey mapping, to uncover additional barriers and facilitators to integrating values elicitation TES into real-world practice. Building on this engagement with patients and providers, future efforts should also expand to include stakeholders who shape systems-level decisions around adoption and implementation. Engaging hospital administrators, policymakers, and other system influencers can help address the structural barriers identified in this study, particularly in care coordination. This broader stakeholder involvement will be essential for integrating TES into health care systems effectively, ensuring that these tools are supported by the necessary resources and workflows.

Conclusions

This study offers novel insights into the design and implementation of TES for values elicitation among patients with MCCs and depression or anxiety, who have been underrepresented in values elicitation research to date. Findings demonstrate a clear patient and provider preference for self-guided TES that support communication skill-building, enabling patients to articulate and advocate for their values, particularly when engaging with dismissive providers. This study extends prior values elicitation research by identifying and addressing important barriers faced by this patient population and offers design recommendations for other TES designers to apply these insights in practice. In addition, applying the ACTS model offered a novel approach that enabled the research team to identify and centralize implementation considerations. Indeed, this study underscores the practical challenges of integrating TES into existing health care workflows, highlighting the need for tools that work within time constraints and respond to structural barriers, such as limited care coordination resources. These insights provide a foundation for future TES development aimed at improving care coordination and patient-centered care for this complex patient population, emphasizing the importance of designing tools that are both scalable and sensitive to mental health symptoms. Future research should further explore co-design processes and strategies for overcoming real-world implementation challenges.

Acknowledgments

ABLB, JG, RK, and WWL conceptualized and designed the study; ABLB, JG, and WWL facilitated the workshops; data were analyzed by IRR, WWL, MB, LMB, and JG; IRR and WWL drafted the manuscript; and MB, LMB, JG, ABLB, and RK critically reviewed and improved the manuscript with intellectual content. All authors read and approved the final manuscript. This project was supported by the National Institutes of Health (grant numbers P30AG059988, P50MH119029, T32 MH115882, and T32 MH130325).

Data Availability

Deidentified data may be made available upon reasonable request to the corresponding author, provided that appropriate data use agreements are established to ensure the continued protection of participant confidentiality.

Conflicts of Interest

RK has accepted consulting fees and honoraria from CHESS Health and Jimini Health. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Workshop prototypes.
[DOCX File , 5958 KB - formative_v9i1e68419_app1.docx]

References

1. Benavidez GA, Zahnd WE, Hung P, Eberth JM. Chronic disease prevalence in the US: sociodemographic and geographic variations by zip code tabulation area. Prev Chronic Dis 2024;21:E14 [FREE Full text] [doi: 10.5888/pcd21.230267] [Medline: 38426538]
2. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press; 2001.

XSL•FO
RenderX

3. Redelmeier DA, Tan SH, Booth GL. The treatment of unrelated disorders in patients with chronic medical diseases. N Engl J Med 1998;338(21):1516-1520. [doi: 10.1056/NEJM199805213382106] [Medline: 9593791]
4. Albreht T, Dyakova M, Schellevis FG, Van den Broucke S. Many diseases, one model of care? J Comorb 2016;6(1):12-20 [FREE Full text] [doi: 10.15256/joc.2016.6.73] [Medline: 29090167]
5. Murphy E, Doyle J, Hannigan C, Smith S, Kuiper J, Jacobs A, et al. Perceptions and use of technology to support older adults with multimorbidity. Stud Health Technol Inform 2017;242:160-167. [Medline: 28873794]
6. Hindmarch T, Hotopf M, Owen GS. Depression and decision-making capacity for treatment or research: a systematic review. BMC Med Ethics 2013;14:54 [FREE Full text] [doi: 10.1186/1472-6939-14-54] [Medline: 24330745]
7. Moran TP. Anxiety and working memory capacity: a meta-analysis and narrative review. Psychol Bull 2016;142(8):831-864. [doi: 10.1037/bul0000051] [Medline: 26963369]
8. Friedman B, Hendry DG. Value Sensitive Design: Shaping Technology With Moral Imagination. Cambridge, MA: MIT Press; 2019.
9. Lim CY, Berry ABL, Hirsch T, Hartzler AL, Wagner EH, Ludman EJ, et al. Understanding what is most important to individuals with multiple chronic conditions: a qualitative study of patients' perspectives. J Gen Intern Med 2017;32(12):1278-1284 [FREE Full text] [doi: 10.1007/s11606-017-4154-3] [Medline: 28849368]
10. Ee C, Lake J, Firth J, Hargraves F, de Manincor M, Meade T, et al. An integrative collaborative care model for people with mental illness and physical comorbidities. Int J Ment Health Syst 2020;14(1):83 [FREE Full text] [doi: 10.1186/s13033-020-00410-6] [Medline: 33292354]
11. Tinetti ME, Naik AD, Dindo L, Costello DM, Esterson J, Geda M, et al. Association of patient priorities-aligned decision-making with patient outcomes and ambulatory health care burden among older adults with multiple chronic conditions: a nonrandomized clinical trial. JAMA Intern Med 2019;179(12):1688-1697 [FREE Full text] [doi: 10.1001/jamainternmed.2019.4235] [Medline: 31589281]
12. Ongwere T, Cantor G, Clawson J, Shih P, Connelly K. Design and care for discordant chronic comorbidities: a comparison of healthcare providers' perspectives. : ACM; 2021 Presented at: PervasiveHealth '20: 14th EAI International Conference on Pervasive Computing Technologies for Healthcare; 2020 May 18-20; Atlanta, GA p. 133-145. [doi: 10.1145/3421937.3422013]
13. Mosler F, Packer K, Jerome L, Bird V. Structured communication methods for mental health consultations in primary care: a scoping review. BMC Prim Care 2023;24(1):175 [FREE Full text] [doi: 10.1186/s12875-023-02129-y] [Medline: 37661251]
14. Liem WW, Lattie EG, Taple BJ, Stamatis CA, Gordon J, Kornfield R, et al. Improving collaborative management of multiple mental and physical health conditions: a qualitative inquiry into designing technology-enabled services for eliciting patients' values. Proc ACM Hum Comput Interact 2024;8(CSCW2):1-30 [FREE Full text] [doi: 10.1145/3687000] [Medline: 39822336]
15. Lim CY, Berry ABL, Hirsch T, Hartzler A, Wagner EH, Ludman E, et al. 'It just seems outside my health': how patients with chronic conditions perceive communication boundaries with providers. : ACM; 2016 Presented at: DIS '16: Proceedings of the 2016 ACM Conference on Designing Interactive Systems; 2016 June 4-8; Brisbane, QLD, Australia p. 1172-1184. [doi: 10.1145/2901790.2901866]
16. Overbeck G, Davidsen AS, Kousgaard MB. Enablers and barriers to implementing collaborative care for anxiety and depression: a systematic qualitative review. Implement Sci 2016;11(1):165 [FREE Full text] [doi: 10.1186/s13012-016-0519-y] [Medline: 28031028]
17. Ratwani RM, Fairbanks RJ, Hettinger AZ, Benda NC. Electronic health record usability: analysis of the user-centered design processes of eleven electronic health record vendors. J Am Med Inform Assoc 2015;22(6):1179-1182. [doi: 10.1093/jamia/ocv050] [Medline: 26049532]
18. Mohr DC, Lyon AR, Lattie EG, Reddy M, Schueller SM. Accelerating digital mental health research from early design and creation to successful implementation and sustainment. J Med Internet Res 2017;19(5):e153 [FREE Full text] [doi: 10.2196/jmir.7725] [Medline: 28490417]
19. Maguire M. Methods to support human-centred design. Int J Hum Comput Stud 2001;55(4):587-634. [doi: 10.1006/ijhc.2001.0503]
20. Guest G, Namey E, Chen M. A simple method to assess and report thematic saturation in qualitative research. PLoS One 2020;15(5):e0232076 [FREE Full text] [doi: 10.1371/journal.pone.0232076] [Medline: 32369511]
21. Lundgren T, Luoma JB, Dahl J, Strosahl K, Melin L. The bull's-eye values survey: a psychometric evaluation. Cogn Behav Pract 2012;19(4):518-526. [doi: 10.1016/j.cbpra.2012.01.004]
22. Schön DA. Educating the Reflective Practitioner: Toward a New Design for Teaching and Learning in the Professions. Hoboken, NJ: Jossey-Bass; 1987:81.
23. Zimmerman J, Forlizzi J, Evenson S. Research through design as a method for interaction design research in HCI. : ACM; 2007 Presented at: CHI07: CHI Conference on Human Factors in Computing Systems; 2007 April 28 - May 3; San Jose, CA p. 493-502. [doi: 10.1145/1240624.1240704]
24. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

XSL•FO
RenderX

25. Braun V, Clarke V. Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. Couns Psychother Res 2020;21(1):37-47. [doi: 10.1002/capr.12360]
26. Tinetti ME, Hashmi A, Ng H, Doyle M, Goto T, Esterson J, et al. Patient priorities-aligned care for older adults with multiple conditions: a nonrandomized controlled trial. JAMA Netw Open 2024;7(1):e2352666 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.52666] [Medline: 38261319]
27. Choi BM, Obeng-Kusi M, Axon DR. Association between patient-provider communication and self-perceived mental health in US adults with cancer: real-world evidence through medical expenditure panel survey. Diseases 2022;10(4):88 [FREE Full text] [doi: 10.3390/diseases10040088] [Medline: 36278587]
28. Hildenbrand GM, Perrault EK, Rnoh RH. Patients' perceptions of health care providers' dismissive communication. Health Promot Pract 2022;23(5):777-784. [doi: 10.1177/15248399211027540] [Medline: 34344222]
29. Menear M, Dugas M, Careau E, Chouinard M, Dogba MJ, Gagnon M, et al. Strategies for engaging patients and families in collaborative care programs for depression and anxiety disorders: a systematic review. J Affect Disord 2020;263:528-539 [FREE Full text] [doi: 10.1016/j.jad.2019.11.008] [Medline: 31744737]
30. Serious illness conversation guide. Ariadne Labs. 2023. URL: https://www.ariadnelabs.org/wp-content/uploads/2023/05/Serious-Illness-Conversation-Guide.2023-05-18.pdf [accessed 2025-01-28]
31. Torab-Miandoab A, Samad-Soltani T, Jodati A, Rezaei-Hachesu P. Interoperability of heterogeneous health information systems: a systematic literature review. BMC Med Inform Decis Mak 2023;23(1):18 [FREE Full text] [doi: 10.1186/s12911-023-02115-5] [Medline: 36694161]
32. Maizes V, Rakel D, Niemiec C. Integrative medicine and patient-centered care. Explore (NY) 2009;5(5):277-289. [doi: 10.1016/j.explore.2009.06.008] [Medline: 19733814]
33. AlQudah AA, Al-Emran M, Shaalan K. Technology acceptance in healthcare: a systematic review. Appl Sci 2021;11(22):10537. [doi: 10.3390/app112210537]
34. Hall A, Walton G. Information overload within the health care system: a literature review. Health Info Libr J 2004;21(2):102-108 [FREE Full text] [doi: 10.1111/j.1471-1842.2004.00506.x] [Medline: 15191601]
35. Berry A, Lim C, Hartzler A, Hirsch T, Ludman E, Wagner EH, et al. Creating conditions for patients' values to emerge in clinical conversations: perspectives of health care team members. : ACM; 2017 Presented at: DIS '17: Designing Interactive Systems Conference 2017; 2017 June 10-14; Edinburgh, United Kingdom p. 1165-1174 URL: https://europepmc.org/abstract/MED/28890950 [doi: 10.1145/3064663.3064669]
36. van Calis JFE, Bevelander KE, van der Cruijsen AWC, Leusink GL, Naaldenberg J. Toward inclusive approaches in the design, development, and implementation of eHealth in the intellectual disability sector: scoping review. J Med Internet Res 2023;25:e45819 [FREE Full text] [doi: 10.2196/45819] [Medline: 37252756]
37. Megaro A. Transparency in AI systems for value co-creation in healthcare. In: Visvizi A, Troisi O, Grimaldi M, editors. Big Data and Decision-Making: Applications and Uses in the Public and Private Sector. Emerald Studies in Politics and Technology. England: Emerald Publishing Limited; 2023:93-105.
38. Panigutti C, Beretta A, Fadda D, Giannotti F, Pedreschi D, Perotti A, et al. Co-design of human-centered, explainable AI for clinical decision support. ACM Trans Interact Intell Syst 2023;13(4):1-35. [doi: 10.1145/3587271]

Abbreviations

ACTS: Accelerated Creation-to-Sustainment
AI: artificial intelligence
MCCs: multiple chronic conditions
MHP: mental health provider
PCP: primary care provider
SICG: Serious Illness Conversation Guide
TES: technology-enabled service

Edited by A Mavragani; submitted 11.11.24; peer-reviewed by PS Foong, V Gejjegondanahalli Yogeshappa; comments to author 18.11.24; revised version received 06.12.24; accepted 15.12.24; published 19.02.25.

Please cite as:
Rooper IR, Liem WW, Burla M, Gordon J, Baez LM, Kornfield R, Berry ABL
Designing Values Elicitation Technologies for Mental Health and Chronic Care Integration: User-Centered Design Approach
JMIR Form Res 2025;9:e68419
URL: https://formative.jmir.org/2025/1/e68419
doi:10.2196/68419
PMID:

XSL•FO
RenderX

©Isabel R Rooper, William W Liem, Martha Burla, Jacob Gordon, Lara M Baez, Rachel Kornfield, Andrew B L Berry. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Classroom-Based Intervention for Reducing Sedentary Behavior and Improving Spinal Health: Pragmatic Stepped-Wedge Feasibility Randomized Controlled Trial

Dominic Fisher[1,2], PhD; Rentia Maart[2], PhD; Lehana Thabane[3], PhD; Quinette Louw[2], PhD

[1]School of Health Professions, Faculty of Health, University of Plymouth, Plymouth, United Kingdom

[2]Division of Physiotherapy, Department of Health and Rehabilitation Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

[3]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada

Corresponding Author:
Dominic Fisher, PhD
School of Health Professions
Faculty of Health
University of Plymouth
North Road East
InterCity Place
Plymouth, PL4 6AB
United Kingdom
Phone: 44 01752587580
Email: dominic.fisher@plymouth.ac.uk

Abstract

Background: Noncommunicable diseases (NCDs) resulting from sedentary behavior (SB) are adding a further strain on the South African health system, which is already struggling to manage infectious diseases. Some countries have enabled children to reduce SB at school by substituting traditional furniture with sit-stand classroom furniture, allowing learners to interrupt prolonged bouts of sitting with standing without interrupting their school work. Alternating between sitting and standing also benefits spinal health by interrupting prolonged periods of high spinal loading, but no such intervention has been trialed in South Africa. The potential to reduce strain on the health system by reducing the incidence of NCDs and improving spinal health requires further consideration. Before embarking on a large classroom-based trial, it is essential to determine the acceptability of the intervention, its impact on teachers' practices, and the logistical and pragmatic considerations of data collection.

Objective: This study aimed to assess the feasibility of implementing a classroom-based intervention to reduce SB and improve spinal health in primary school learners, to assess the pragmatics of delivering and adherence to the intervention, and assess the pragmatics of measuring physical activity and postural dynamism data with wearable sensors.

Methods: We used a stratified, closed-cohort, randomized, 2-cluster, stepped-wedge design with a pragmatic approach. One grade 5 and grade 6 class each was recruited from contrasting socioeconomically categorized, state-funded primary schools in the Western Cape province, South Africa. Classroom furniture was substituted with sit-stand desks, and health education and movement videos (HEMVs) were shown during class time. Skin-mounted activPAL physical activity monitors were used to measure SB and postural topography and Noraxon myoMOTION inertial measurement units (IMUs) to measure spinal movement. The study was evaluated for feasibility by tracking school retention, successful delivery of the HEMVs, the use of sit-stand desks, compliance with the wearable sensors, and data accuracy. We deductively analyzed teachers' interviews and learners' focus groups using Atlas.ti 9 software. Descriptive analysis of quantitative data was performed using Microsoft Excel.

Results: Cluster 1 withdrew from the study before follow-up SB, postural topography, and spinal movements were measured. All feasibility outcomes, namely (1) classroom retention, (2) delivery of HEMVs, (3) learner and teacher acceptance and usage of sit-stand classroom furniture, (4) 100% compliance with wearing skin-mounted sensors for the duration of the intended measurement period, and (5) minimum 80% eligibility of sensor data gathered included in data analysis, were met in cluster 2. The study found that it is feasible to conduct a larger trial with minor modifications to the methodology.

XSL•FO
RenderX

Conclusions: We recommend a whole-school approach to support the intervention and a monitoring strategy to track the impact of the intervention on the classroom. Furthermore, we recommend contextualized teacher training on how sit-stand desks and HEMVs can be used as classroom management tools.

Trial Registration: Pan African Trials Registry PACTR201811799476016; https://tinyurl.com/y4upoys8

International Registered Report Identifier (IRRID): RR2-10.2196/18522

(JMIR Form Res 2025;9:e65169) doi:10.2196/65169

KEYWORDS

sedentary behavior; spinal health; classroom-based intervention; sit-stand desks; spinal; African; Africa; primary school; child; youth; randomized controlled trial; RCT; infectious disease; acceptability; data collection; teacher; classroom-based interventions; primary school learners; physical activity; closed-cohort; interview; quantitative data; wearable; wearable sensor; spine

Introduction

The prevalence of preventable, long-term noncommunicable diseases (NCDs) presents a serious strain on the South African health care system, already under pressure from the burden of infectious disease management [1]. Although the causes of NCDs are multifactorial, there is consensus that lifestyle, including sedentary behavior (SB), is a major contributor [2]. Any waking behavior with energy expenditure less than or equal to 1.5 metabolic equivalent units while sitting, lying down, or reclining is considered sedentary [1]. Although initial SB interventions were mainly focused on adults, the understanding that SB in adulthood tracks through from childhood and adolescence [3] expanded SB research to include younger populations. Interventions aimed at reducing SB in children have shown promise, with several studies demonstrating good efficacy in reducing SB by educating parents, teachers, and children about the harmful effects of SB [4]. Interventions to address SB initially focused on substituting discretionary screen time with periods of physical activity. However, understanding that the beneficial effects of physical activity, even at the prescribed World Health Organization (WHO) dose, do not undo the detrimental physiological effects of SB [5] has led to the development of interventions to prevent the accumulation of SB.

School-based SB studies have demonstrated high levels of accumulated periods of nondiscretionary sedentary time during class [6]. These high levels of SB are attributable to teachers' preference that children remain seated during class [7] and environmental barriers to interrupting SB when teachers encourage movement [8]. The conventional classroom environment is not conducive to increasing children's movement without compromising teaching and learning. As a result, classroom-based SB interventions trialed dynamic sit-stand furniture to overcome the environmental barriers posed by traditional classroom furniture.

Interrupting prolonged sitting bouts by alternating between sitting and standing may have the additional benefit of improving spinal health. Prolonged sitting has been shown to increase axial loading of the spine and increase back muscle activation, which could lead to back pain [9]. Given that spinal pain tracks to adulthood from childhood [7,10] and given the reported high levels of nondiscretionary classroom sedentary time, alternating children's postural topography may have dual long-term benefits on children's cardiometabolic and spinal health. It is proposed that alternating between sitting and standing (alternating postural topography) mitigates the effects of prolonged loading in one position by providing periods of relative rest. This approach aligns with the theory of postural dynamism, which encourages regular spinal movement as opposed to maintaining a single preferred spinal position. Altering postural topography may promote spinal health by mitigating axial load on spinal posture and encouraging regular spinal movement. The dual benefit of interrupting prolonged sitting and encouraging transitioning between sitting and standing during class to mitigate the long-term risks to cardiometabolic and spinal health holds significant public health potential in resource-limited contexts, such as South Africa.

Classroom-based interventions comprising dynamic sit-stand furniture, health education, and movement integration that have been conducted in the United States, Europe, Australia, and New Zealand have shown good efficacy in reducing SB without disturbing teaching and learning activities [4]. To the best of our knowledge, no such studies have been conducted in South Africa. Conducting a trial of a classroom-based intervention in South Africa would have to consider a range of contextual factors related to the educational system and socioeconomic inequality, not to mention logistic and pragmatic considerations. Before implementing such a trial, it is important to establish the contextual viability of a previously untested intervention. Given the likely increase in children's movement in response to a classroom-based intervention and the potential impact on teacher practice, determining the acceptability of the intervention was considered [11,12]. In addition, a pragmatic methodology to measure postural topography, physical activity, and spinal movement needed to be established before being implemented in a large trial. Furthermore, the integrity of data collected in a new environment needed to be assessed. These are all key aspects needed to establish the feasibility of conducting a large trial to assess the beneficial effects of a novel SB intervention in a South African classroom context.

The aim of this study was, thus, to assess the feasibility of implementing a classroom-based intervention to reduce classroom SB and promote spinal health in primary school learners. The study objectives were to assess the pragmatics of delivering and adherence to a classroom-based intervention and assess the pragmatics of measuring physical activity and postural dynamism data with activPAL and Noraxon myoMOTION inertial measurement units (IMUs), respectively. This

XSL•FO
RenderX

manuscript focused only on the primary feasibility outcomes. The preliminary findings of the effects will be published subsequently.

Methods

Ethical Considerations

This study was granted ethical approval by the Stellenbosch University Health Research Ethics Committee (reference number S17/08/130) and institutional permission by the Western Cape Education Department (reference number 20170525-1279). It was also registered with the Pan African Trials Registry (PACTR201811799476016; International Registered Report Identifier [IRRID] RR1-10.2196/18522).

Prospective participants who expressed interest in participating in the project were provided with study information documents and consent forms. Learners were provided with child assent and parent/guardian consent forms, as well as study information for parents and guardians, which were then collected by the research team upon completion. Before the study began, learners were informed about the intervention and measurement methods. Study data were anonymized by assigning a unique participant identifier to data obtained from participants. As compensation, participants were offered retention of the intervention materials at the conclusion of the project.

Study Design

A school-based, stratified, closed-cohort, randomized, 2-cluster, stepped-wedge design with a pragmatic approach was used for this study. Participant clusters (school classrooms) were the unit of randomization to determine the order of implementing the intervention. This study report followed the CONSORT (Consolidated Standards of Reporting Trials) extension statement for pilot and feasibility trials (Multimedia Appendix 1) [13]. The protocol for this feasibility trial has been published elsewhere [14]. The CONSORT stepped-wedge flow diagram (Table 1) for cluster randomized trials [15] shows the timing of crossover from control to intervention conditions.

Table 1. Stepped-wedge cluster RCT[a] diagram of the study sequence.

Cluster	Randomized transition from usual classroom conditions to intervention condition					
	Weeks 1-2 (April 23-May 10, 2018)	Week 3 (May 11-16, 2018)	Weeks 4-5 (April 23-May 10, 2018)	Weeks 6-14 (May 14-August 10, 2018)	Weeks 15-16 (August 13-24, 2018)	July 31-August 15, 2019
Cluster 1, Q5[b] classroom	Usual conditions	Baseline measurement and introduction of intervention	Intervention	Intervention	Withdrawal from study and qualitative interviews	Usual conditions: withdrawal from study
Cluster 2, Q3[c] classroom	Usual conditions	Usual conditions	Baseline measurement and introduction of intervention	Intervention	Follow-up measurement and qualitative interviews	Follow-up measurement and qualitative interviews

[a]RCT: randomized controlled trial.

[b]Q5: quintile 5.

[c]Q3 quintile 3.

The stepped-wedge design was used to allow the intervention to be evaluated within the bounds of the logistical constraints and context of each cluster [16]. Given the limited availability of measurement equipment, the stepped-wedge design allowed sequential baseline and follow-up measurements. Furthermore, given the evidence of the effectiveness of the mode of intervention in other contexts [4], and other classroom-based interventions, the stepped-wedged design allowed both clusters to receive the potential benefit of the intervention.

Study Setting

The study was conducted in the Western Cape province of South Africa, a region of broad language, cultural, and socioeconomic diversity. In South Africa, publicly funded (state schools) are categorized into quintiles according to socioeconomic factors, including income, literacy, and employment rates of the surrounding community [17]. Lower-quintile schools are more resource limited than upper-quintile schools and thus receive greater state funding per registered learner. In the metro central district, there are fewer lower-quintile schools, and most schools fall under the quintile 5 (Q5) category.

Sampling and Recruitment

To ensure a diverse range of contextual factors, primary schools from quintile 3 (Q3) and Q5 were selected exclusively from the central metro district of the Western Cape Education Department. The rationale was that if it were deemed feasible to conduct a future trial in the more challenging, resource-strained context, it would be feasible in less challenging, adequately resourced contexts. A downloaded list of publicly funded schools in the central metro district from the Western Cape Education Department website was delimited to school categories Q3 and Q5. School principals, stratified according to Q3 and Q5 schools, were randomly contacted by the principal investigator via telephone or email for recruitment. Afterward, the principals nominated grade 5 or 6 teachers to attend an information session.

Sample Size

A sample size calculation was not performed as per the CONSORT extension statement for feasibility studies [18]. The sample target was to include a wide range of contextual factors relevant to the feasibility outcomes. The diversity between the clusters (namely classrooms from Q3 and Q5 schools) was considered sufficient to provide a range of considerations to

XSL•FO
RenderX

assess the feasibility of the intervention and the pragmatics of data collection. The individual participants selected for measuring physiological outcomes (SB and postural dynamism) were randomly chosen from the class list provided by the class teacher. An equal number of male and female learners were sampled.

Intervention

The aim of the intervention is to improve the long-term cardiometabolic and spinal health of the public by reducing the accumulation of prolonged SB that has been shown to be pervasive in schools. Providing learners with an opportunity to transition between sitting and standing during class time helps arrest sedentary physiology and provides respite from prolonged axial spinal loading. The study intervention comprised a novel height-adjustable sit-stand desk and a playlist of 9 health education and movement videos (HEMVs). The study authors (DF and QL) contributed to the development of the intervention. The design concept was informed in part by a systematic review of the efficacy of classroom-based interventions to improve spinal health and reduce the SB of schoolchildren [4], as well as a qualitative study of educators' perceptions of learners' movement during class time conducted in the Western Cape province, South Africa [7]. The systematic review provided compelling evidence for the efficacy of classroom-based interventions that include alternative, sit-stand classroom furniture and health education to improve spinal health and SB outcomes.

Sit-Stand Desks

The research team considered the sit-stand desks available in South Africa prohibitively expensive and thus unsuitable for implementation in the study. An innovation team developed a novel, multifunctional, height-adjustable, sit-stand classroom desk called the KUZE (Multimedia Appendix 2) for this study. Learners and teachers were shown how to select the correct height for sitting and standing. All the usual classroom chairs were removed and replaced with KUZE sit-stand desks during the study period.

Health Education and Movement Videos

The innovation team comprising the authors, professional video content creators, and a teacher who advocates for using body movement in teaching mathematics developed a series of HEMVs for this study. Learners were instructed to follow the videos, which included an interactive component requiring them to solve simple arithmetic problems using corresponding body movements (Multimedia Appendix 3). The teachers were handed the HEMVs on a mobile external hard drive and, after discussions with the researchers, were given the freedom to develop their own strategies for playing the videos during class time. The videos are available to view on YouTube [14].

Control Conditions

The control conditions were the usual classroom conditions. Cluster 1 and cluster 2 classrooms had similar dimensions. The cluster 1 classroom furniture comprised metal framed, wooden, all-in-one, tandem desks that fit 2 learners abreast (Multimedia Appendix 4). The cluster 2 classroom furniture had a combination of single and double metal-framed wooden tables with accompanying plastic-molded chairs (Multimedia Appendix 4). The HEMVs were not played during control conditions.

Pragmatics of Physiological Data Collection Using Wearable Sensors

Physical Activity Monitoring: activPAL

Participants' classroom physical activity and postural topography (sitting, standing, stepping. and sit-to-stand transitioning) were measured using activPAL3 microsensors (PAL Technologies). In the past decade, activPAL sensors have been widely used to objectively measure physical activity [19-21]. Sensors were applied to participants before the start of lessons and removed after the end of lessons on physiological data collection days. The sensors were attached to the anterior right thigh with a waterproof nitrile sleeve and Opsite dressing, as prescribed by the user manual. Data logged on the sensors were downloaded into a secure file at the end of each day.

Spinal Movement: Noraxon myoMOTION Inertial Measurement Units

Postural dynamism was measured using Noraxon myoMOTION IMUs. Wearable IMUs allow for the assessment of postural dynamism in an ecologically valid setting of the classroom. The IMUs combine on-board triaxial gyroscopes, accelerometers, and magnetometers for accurate sensor orientation tracking [22]. IMUs were attached to the head with an elasticated Velcro belt and to the neck, thorax, and sacrum directly to the skin using double-sided tape.

Description and Measurement of Feasibility Outcomes

The feasibility outcomes include retention of the study, fidelity to the intervention (acceptance, and usage of the KUZE sit-stand desks, delivery of the HEMVs), and integrity of the physiological data collected (compliance with wearing the activPAL and IMU sensors). The success indicators (set a priori) and methods of measuring the 5 feasibility outcomes are outlined in Table 2.

XSL•FO
RenderX

Table 2. Summary of feasibility outcomes, success indicators, and measurement methods.

Feasibility outcome	Success indicator	Measurement method
Cluster retention	Both clusters remained in the study until follow-up measurements were obtained.	• Retention rate recorded on the project management record • Qualitative feedback obtained from teachers regarding retention
Delivery of HEMVs[a]	Teachers developed a routine of playing the HEMVs and adhered to the routine.	• Qualitative feedback obtained from teachers and learners at the exit interview
Acceptance and usage of the KUZE desk in the classroom	Learners and teachers accepted and used the KUZE as classroom furniture for the entire study period.	• Retention rate recorded on the project management record • Qualitative feedback obtained from teachers and learners regarding acceptance of the KUZE
Compliance with wearing activPAL and IMU[b] sensors	There was 100% compliance with wearing activPAL sensors and IMUs for the duration of the intended measurement period.	• Recorded on the project management record
Integrity of physical activity and postural dynamism data	Of all data captured by activPAL and IMU sensors, 80% were eligible for inclusion in the analysis.	• Recorded on the project management record

[a]HEMV: health education and movement video.

[b]IMU: inertial measurement unit.

Feasibility outcomes were measured using both qualitative and quantitative methods. For the qualitative measurement of the feasibility outcomes, individual depth interviews (IDIs) with the class teachers and focus group discussions (FGDs) with a subgroup of learner participants from each cluster were conducted at the end of the intervention period. The success criteria of the feasibility outcomes were used as the framework for the IDIs and FGDs. The quantitative methods encompassed the physiological data obtained from the activPAL and IMU sensors, as well as estimates of retention and compliance, which were monitored by the researcher.

Interpretation of Feasibility Criteria

The success criteria set a priori were:

- Continue with a large pilot/trial if all 5 success criteria are met in both clusters;
- Make minor modifications to the protocol if 3 or more criteria are met in both clusters before continuing with a pilot/trial; or
- Make significant protocol modifications if 2 or less criteria are met in both clusters [18].

Data Analysis

Qualitative Analysis

A deductive analysis of data from IDIs with teachers and FGDs with learners was conducted to determine the feasibility outcomes and success indicators. The qualitative feedback sessions were recorded and transcribed in full and then analyzed using Atlas.ti 9, a computer-assisted qualitative data analysis software program that facilitates the organization, coding, analysis, and visualization of data [23]. Learner and teacher responses were grouped into corresponding themes related to classroom retention, intervention delivery, acceptance, and usage of the KUZE. At the end of the intervention, 2 FGDs per cluster were conducted. The focus group comprised both individuals who provided physiological data and those who were only exposed to the intervention but did not provide physiological data. Verbatim representative quotes of participant responses were provided.

Quantitative Analysis

Categorical data were analyzed descriptively and presented as percentages. The number of participant classes (clusters) that remained in the study for the duration of the study was documented and presented as a percentage of the total participant classes that enrolled in the study. The proportion of data used in the analysis compared to data collected was presented as a percentage. All analyses were performed using Microsoft Excel.

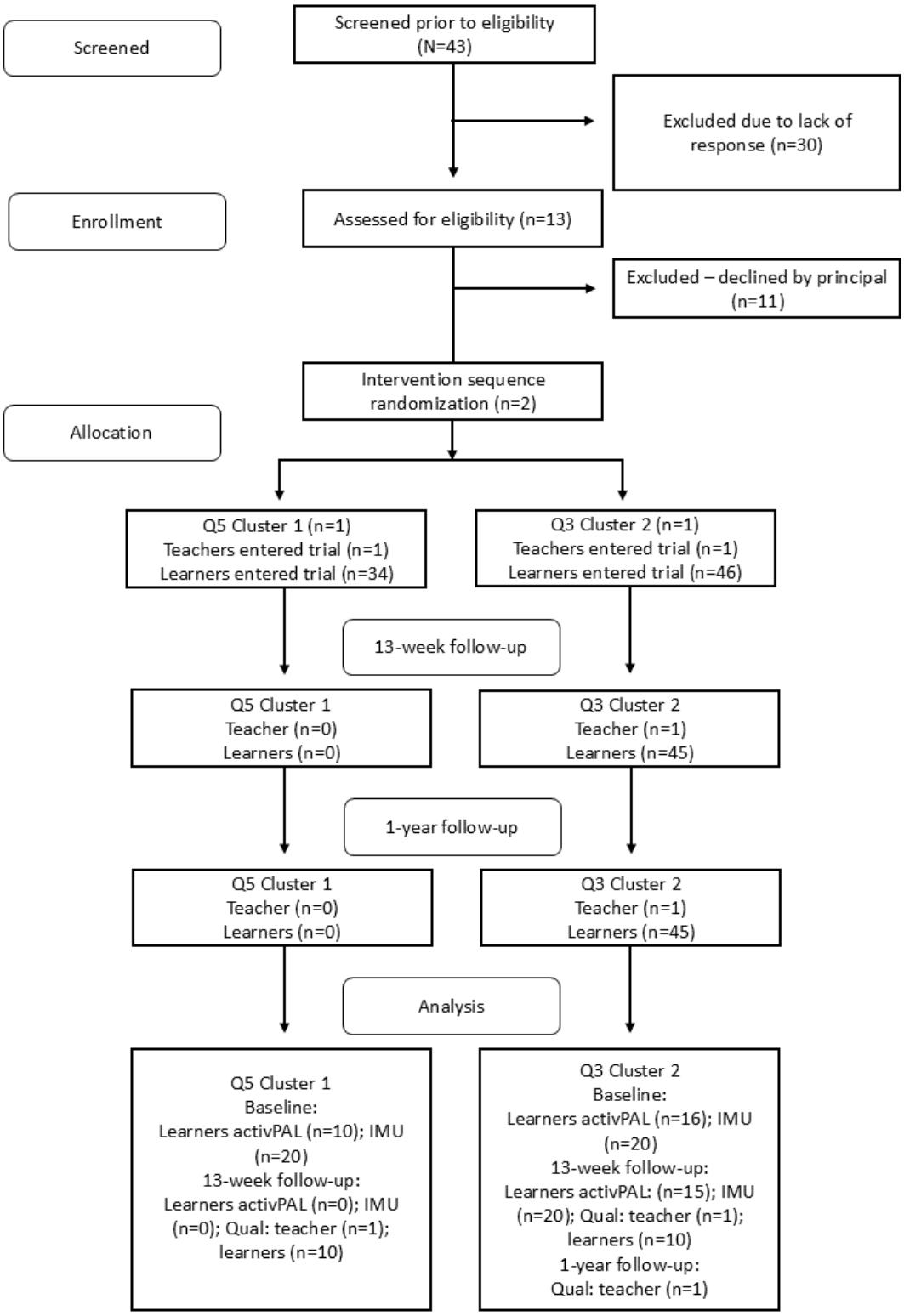
Results

Demographic Characteristics

The CONSORT flow diagram of the study is illustrated in Figure 1. The baseline demographic characteristics of participants were similar across clusters. The cluster 1 (Q5) school comprised a grade 6 class, while cluster 2 (Q3) was a grade 5 class. Thus, the ages of the learners in this study were 10-11 years. An equal number of male and female learners made up the study sample.

XSL•FO
RenderX

Figure 1. CONSORT flow diagram of the feasibility RCT. CONSORT: Consolidated Standards of Reporting Trials; IMU: inertial measurement unit; Q3: quintile 3; Q5: quintile 5; Qual: qualitative; RCT: randomized controlled trial.



Feasibility Outcome 1: Cluster Retention

The cluster 1 (Q5) teacher withdrew her consent to participate in the study after the 13-week intervention and before physiological follow-up measurements were taken. However, the teacher and learners participated in the IDI and FGDs, respectively. Cluster 2 (Q3) remained in the study throughout the study period. Table 3 demonstrates at which stage the cluster 1 teacher withdrew.

XSL•FO
RenderX

Table 3. Components of retention for clusters 1 and 2.

Cluster	Recruitment	Baseline measurements	Intervention	Follow-up measurements	Qualitative interviews
1	✓	✓	✓	Withdrew	✓
2	✓	✓	✓	✓	✓

Qualitative Findings

Based on the qualitative findings, the reasons for withdrawal from the study were analyzed into 3 themes: the way access to the classroom was negotiated; the teacher's preference for not being video-recorded during data collection; and the perceived disruption to teaching and learning by the study.

Access Negotiation

The same strategy was used to recruit both clusters. Negotiation of access to the classroom was conducted via the school principal and the original teacher (cluster 1). However, a last-minute substitute teacher was employed immediately before the commencement of the study. The school principal discussed the project and its requirements with the substitute teacher. However, the substitute teacher felt that she was not adequately informed of the full extent of the study, despite receiving the study information material.

> *So, I didn't really know. So, I was like, okay, I'll wait for you (the researcher) to come; maybe you'll tell me, and then suddenly, it was a whole setup in my class, and I was, like, what is going on?* [Teacher, cluster 1]

The teacher in cluster 2 was satisfied with the recruitment strategy.

> *Yes. It was the right way. Starting with the principal of course…the principal would never decide on her own. She had to go to the teachers and bring this to the teachers and hear what they say.* [Teacher, cluster 2]

Preference for Not Being Video-Recorded

The cluster 1 teacher was uncomfortable about being recorded on video. She reported that she was self-conscious about being on camera, which influenced her interactions with the learners. The teacher in cluster 2 did not have objections to the use of video for postural dynamism data collection.

> *Well, I mean, having someone basically watching you (laughing). You don't really have freedom with your kids because you say something, but you get recorded on video.* [Teacher, cluster 1]

Perception of Disruption to Teaching and Learning

The cluster 1 teacher felt that the intervention (KUZE) interfered with her preferred way of arranging the classroom furniture.

> *And I found that because the way my class is set out, I put my class when it was rows, so before, I had groups, but then when the desk came, I only could have rows because that was the only thing that could fit in my class.* [Teacher, cluster 1]

To standardize the orientation of learners for postural dynamism measurement, the learners were required to sit facing the front of the classroom. The teacher reported that the learners often had to be moved from their usual desks, and this impacted interpersonal dynamics with the person next to them.

> *Maybe that, but also, as you know my class, I had groups. So, you needed your kids to face the front. So, that also caused moving around and disruptions, like this child don't want to sit next to that child, but yet, they had to because...to accommodate the project.* [Teacher, cluster 1]

The teacher was concerned that disruptions to teaching and learning would impact the learners' academic performance and decided it was best to withdraw from the study.

> *I don't think my kids can afford to have all that interruptions, especially during the second term, when we have exams, and the third term, when we have systemics. So, I needed to be focused all the time. So, having the disruption was just a lot to manage.* [Teacher, cluster 1]

Feasibility Outcome 2: Delivery of HEMVs

Both teachers reported having developed a routine for playing the HEMVs during class. The cluster 1 teacher played the HEMVs intermittently during the day, while the cluster 2 teacher used the videos as a classroom management tool to his advantage. He played the videos at the start of the periods to engage learners and prepare them for the lesson.

> *So, what I did is the morning I played them one, in the beginning of the day and before first break. Then, second break, I don't have them, so...they go for 2 hours…they go somewhere else (to a different classroom), and then maybe a video at the end of the day.* [Teacher, cluster 1]

> *Because of the start of the period, I want them to bring their minds before doing maths…it's like an exercise where everybody would feel right for the lesson…We didn't do that a lot every day. Most of the time, I did the videos when it's my first period in my class.* [Teacher, cluster 2]

Feasibility Outcome 3: Acceptance and Usage of the KUZE Desk in the Classroom

Teachers and learners in both clusters accepted the KUZE as the classroom furniture for the duration of the intervention period. The cluster 1 teacher allowed the learners to stand during groupwork activities for specific subjects and during times where independent work was required.

> *I think the desk was helpful in that situation, yes, where they had their own work to do without having extra work on the board or looking at the board. So,*

XSL•FO
RenderX

> *when they had independent tasks like that, then I think it was okay to be standing, but also not for a long period of time because they tend to move around and they're free to walk around.* [Teacher, cluster 1]
>
> *I think it was actually amazing for me not to be sitting...just sitting there like a stick insect the whole time, and sometimes, you can't even concentrate in class because your head is down. But you're active from those desks you brought us (KUZE).* [Learner, cluster 1]

Learners in cluster 2 were allowed to use their discretion in whether to sit or stand while completing their work and to break up periods of sitting. At the end of the intervention period, the KUZE desks were removed from cluster 1 and the usual classroom furniture was returned. The cluster 2 teacher requested that the KUZE remain as the standard classroom furniture in his class beyond the completion of the study. When asked about whether they wanted to return to using their usual classroom furniture, learners from cluster 2 chose to keep the KUZE furniture.

> *I think I was going to use it for the new learners maybe next year. Because as I can see now, the other classes, they see these desks. Now they are also curious...* [Teacher, cluster 2]
>
> *I want you to keep the cosi (KUZE) chairs for us…and our teacher said the cosi (KUZE) chairs are good for us so that we can stand.* [Learner, cluster 2]

The teacher in cluster 2 reported that the KUZE desks provided a valuable advantage of more space to access all the areas of the classroom.

> *The advantage they (KUZE desks) give me is that in that classroom there is space; it's spacious, you see. I can now move in between them…I am able to move around, looking at their books.* [Teacher, cluster 2]

Feasibility Outcome 4: Compliance With Wearing activPAL and IMU Sensors

There was 100% compliance with wearing activPAL sensors and IMUs. No learners reported any adverse skin reactions to wearing the activPAL sensors at any stage of the study. During the 13-week follow-up measurement, 10 participants reported mild head discomfort from the elasticated Velcro strap used to secure the head IMU. In all cases, discomfort was resolved by readjusting the strap. Thereafter, the learners were able to return to class and resume learning activities. The mild discomfort that learners reported from the head strap had resolved before they were allowed to return to the class to complete the period of data collection. All learners selected for physiological measurements complied with wearing the activPAL sensors and IMUs throughout the study.

Feasibility Outcome 5: Integrity of Physical Activity and Postural Dynamism Data

A low rate of data loss was recorded for physical activity and postural dynamism (Table 4). An average of 94.64% of recorded SB and spinal movement data had sound integrity and was considered eligible for inclusion in the analysis for the secondary objectives. The rate of data integrity met the a priori target of 80%. Data loss from wearable sensors was attributable to technical problems during downloading to the project computer or due to batteries running out of charge.

Table 4. Data integrity from wearable sensors.

Measurement period	Cluster 1		Cluster 2	
	activPAL (%)	IMU[a] (%)	activPAL (%)	IMU (%)
Baseline	87.23	92.86	98.48	100.00
13 weeks	Withdrew	Withdrew	98.48	95.00
1-year follow up	Withdrew	Withdrew	83.75	Not measured

[a]IMU: inertial measurement unit.

Summary of Feasibility Outcomes

Three of the feasibility outcomes were achieved in 3 criteria in both clusters. As cluster 1 was not retained in the study, the acceptance and usage of the KUZE could only be assessed partially through qualitative feedback and was thus categorized as "partially met." All 5 feasibility outcomes were achieved in cluster 2 (Table 5).

XSL•FO
RenderX

Table 5. Achievement of success criteria.

Feasibility outcome	Success indicator	Outcome	
		Cluster 1	Cluster 2
Classroom retention	Both clusters remained in the study until follow-up measurements were obtained.	Unmet	Met
Delivery of HEMVs[a]	The teacher developed a routine of playing the videos and adhered to the routine.	Met	Met
Acceptance and usage of the KUZE desk in the classroom	Learners and teachers accepted and used the KUZE as classroom furniture for the entire study period.	Partially met	Met
Compliance with wearing activPAL and IMU[b] sensors	There was 100% compliance with wearing activPAL sensors and IMUs for the duration of the intended measurement period.	Met	Met
Integrity of physical activity and postural dynamism data	Of all data captured by activPAL and IMU sensors, 80% were eligible for inclusion in the analysis.	Met	Met

[a]HEMV: health education and movement video.

[b]IMU: inertial measurement unit.

Discussion

Principal Findings

This study effectively assessed the feasibility of implementing a classroom-based intervention to reduce classroom SB and promote spinal health in primary school learners in the Western Cape province of South Africa. Our findings showed that the a priori success criteria to decide to continue with a large trial were met. Having met 3 feasibility outcome success criteria (delivery of HEMVs, compliance with wearing physical activity and postural dynamism sensors, and integrity of the data gathered) in both clusters indicates that with minor modifications to the study methodology, it is feasible to progress to a larger trial. Despite this positive finding, reduced cluster retention may be a risk of a larger trial. Therefore, strategies to enhance retention are an important consideration for a larger trial.

Individual teacher-related aspects should be considered in retention strategies. The last-minute substitution of the teacher meant that the research team was unable to establish a satisfactory level of engagement, rapport, and support before the study started. This may have led to a negative attitude toward the study, which was compounded by her perceived loss of autonomy due to the study-related change to the classroom furniture and configuration. Holistic support of teachers is important in classroom-based movement integration interventions [24] to maintain adequate levels of engagement and support throughout the process of the study. Such a holistic whole-school approach that includes senior school management and other school stakeholders has gathered traction in interventions to increase physical activity in UK primary schools [25]. However, such a whole-school approach will need to be contextualized for the South African school system, considering the scarce track record of implementation of similar interventions. Given the importance of teachers in implementing the study and the dynamic nature of school settings, the design of a future trial should incorporate pragmatism to ensure that timely and appropriate adaptation is possible throughout the study [26].

Although not retaining cluster 1 (Q5) was disappointing, it provides a valuable learning opportunity to inform a larger trial. It is, therefore, recommended that a customized school context–specific program of engagement, awareness, and involvement in the study precede a future trial to improve participants' experience of participation [27] and to optimize its acceptability and efficacy [28]. Prior teacher involvement and program monitoring may also provide an opportunity to codevelop cluster retention strategies for implementation in a future trial; aid the development, implementation, and adaptation of the intervention; and track cluster progress against study goals [29]. Cocreation of school-based health interventions has been shown to improve the feasibility of interventions in the complex environment of the school context [30]. Including teachers in the development and implementation of an intervention can leverage their expertise of the classroom environment and enable them to find solutions to concerns about the potential classroom disruptions and reconfiguration of classroom furniture. Raising awareness about the study with a view to including teachers to codevelop the intervention may be important considerations for a future trial, given the sizeable investment requirement in this emerging area of research in South Africa.

Despite no previous classroom-based interventions including sit-stand desks in South Africa, the KUZE desk was well accepted by teachers and learners. Teachers recognized the positive impact of allowing learners to stand on their ability to work independently and on their concentration after prolonged periods of sitting. However, some concerns were raised about disruption to teaching and learning and loss of control of the classroom, which presents a potential threat to the rollout of a future trial. Similar concerns were voiced by teachers in a previous UK-based trial [11]. These concerns are reasonable, given the longstanding tradition and practice of enforcing learners to remain seated during class time. Interrupting periods of sitting with standing tasks challenges teachers' practice and propensity to enforce sitting during teaching and learning activities. In a study of teachers' perspectives of learners' movement during class, participants reported that they believed that sitting enhances learners' concentration [7] and were

XSL•FO
RenderX

unaware of the evidence that interrupting prolonged sitting may enhance their concentration and engagement with cognitive tasks [31,32]. Teachers in the study by Fisher and Louw [7] associated enforcing that learners remain seated with teachers' maintaining control of the classroom and their level of classroom management skill and teaching pedagogy. It may be prudent to emphasize the benefits of interrupting prolonged sitting for learners in future teacher training and support as part of a future trial of the intervention.

The activPAL sensor wear protocol used in this study did not appear to impact the logistics or integrity of the physical activity data collected. The decision to apply and remove the activPAL sensors daily was mainly motivated by not wanting to risk losing the expensive devises. The research team also considered the ethics of the potential safety risks study participants would bear by wearing expensive devices in their communities without any mitigating measures in place. Most studies of free-living physical activity using activPAL adopt a 24-hour, 7-day wear protocol [33]. However, this wear protocol has drawn some criticism, with evidence of dose–response noncompliance with wear protocols approaching the 7-day period [34]. Reports of skin irritation caused by perspiration, particularly in warm, humid conditions, have threatened study participants' compliance with wearing the sensors [35]. The fact that no skin irritation was reported by study participants justifies the rationale for the wear protocol. If this activPAL wear protocol was adopted in a future study, it may be useful to determine the validity of the single wear protocol compared to the 24-hour, 7-day wear protocol.

Study Limitations

The study was limited to schools in a specific geographical and socioeconomic context. Therefore, the findings may not be broadly generalizable. The interpretation of some findings was limited by the suboptimal retention of 1 cluster. The study was also limited to specific preset feasibility criteria and important implementation factors; for example, teacher and school dynamics were not considered. The feasibility study's data are also arguably too limited for sample size considerations for a larger trial, and further research is required.

Conclusion

This study addressed the uncertainty of the acceptability of a classroom-based SB and spinal health intervention, its impact on teachers' practice, and the logistical and pragmatic considerations of data collection in a previously unresearched setting. Having established the feasibility of some aspects of the intervention in the local context, the acceptance of sit-stand desks remains questionable and presents a potential risk to the success of a larger trial. Although teachers and learners from both clusters accepted and used the intervention classroom furniture for the duration of the intervention period, the feedback from the cluster 1 teacher raises some uncertainty. However, these concerns can be addressed, as described in the recommendations. The results of this study indicate that minor revisions to the current methodology are required to conduct a large cluster randomized controlled trial (RCT) of an intervention aimed at reducing SB and improving spinal health in primary schools in the Western Cape province.

Recommended changes to the methodology include:

- A holistic whole-school approach to the intervention, including all grade teachers and school management, to support implementation of the intervention, particularly playing HEMVs
- A monitoring strategy to track the impact of the intervention on the classroom, teacher, and school's environment, as well as the ability to adapt and implement pragmatic strategies to enhance retention
- Inclusion of teachers in the cocreation, monitoring, and implementation of the intervention to improve the feasibility of a future trial
- Contextualized teacher training on SB and how sit-stand desks and HEMVs can be used as a classroom management tool

Acknowledgments

This project was funded by the National Research Foundation (CPRR150720128324). The authors thank the innovation team, Jaco Langenhoven and Sjan-Mari van Niekerk (sit-stand desk design and production) and Kurt Minnaar and D'ave Meyer (health education and movement video production), for contributing to the study.

Data Availability

The datasets used and analyzed during the study are available from the corresponding author upon reasonable request.

Authors' Contributions

DF was the major contributor to drafting and writing the manuscript. Conceptualization, formal analysis, investigation, data curation, and visualization were managed by DF and QL; methodology by DF, QL, and LT; validation and writing—review and editing by QL, LT, and RM; and funding acquisition by QL. All authors have read and agreed to the published version of the manuscript. No generative artificial intelligence was used in the production of this manuscript.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
CONSORT (Consolidated Standards of Reporting Trials) extension statement for feasibility of sedentary behavior (SB) and spinal health.
[PDF File (Adobe PDF File), 359 KB - formative_v9i1e65169_app1.pdf]

Multimedia Appendix 2
Learner using KUZE for sitting and standing.
[PNG File , 551 KB - formative_v9i1e65169_app2.png]

Multimedia Appendix 3
Still from a health education and movement video (HEMV).
[PNG File , 294 KB - formative_v9i1e65169_app3.png]

Multimedia Appendix 4
Cluster classroom furniture.
[PNG File , 296 KB - formative_v9i1e65169_app4.png]

References

1. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, SBRN Terminology Consensus Project Participants. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. Int J Behav Nutr Phys Act 2017 Jul 10;14(1):75 [FREE Full text] [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]
2. Poses-Ferrer E, Parisi R, Gonzalez-Viana A, Castell C, Arias de la Torre J, Jones A, et al. Daily sitting time and its association with non-communicable diseases and multimorbidity in Catalonia. Eur J Public Health 2022 Feb 01;32(1):105-111 [FREE Full text] [doi: 10.1093/eurpub/ckab201] [Medline: 34850878]
3. Biddle SJ, Pearson N, Ross GM, Braithwaite R. Tracking of sedentary behaviours of young people: a systematic review. Prev Med 2010 Nov;51(5):345-351. [doi: 10.1016/j.ypmed.2010.07.018] [Medline: 20682330]
4. Fisher D, Louw Q. The effect of classroom-based interventions on sedentary behavior and spinal health in schoolchildren: systematic review. Interact J Med Res 2022 Oct 26;11(2):e39006 [FREE Full text] [doi: 10.2196/39006] [Medline: 36287590]
5. Ryan DJ, Stebbings GK, Onambele GL. The emergence of sedentary behaviour physiology and its effects on the cardiometabolic profile in young and older adults. Age (Dordr) 2015 Oct 28;37(5):89 [FREE Full text] [doi: 10.1007/s11357-015-9832-7] [Medline: 26315694]
6. Ridgers ND, Salmon J, Ridley K, O'Connell E, Arundell L, Timperio A. Agreement between activPAL and ActiGraph for assessing children's sedentary time. Int J Behav Nutr Phys Act 2012 Feb 19;9(1):15 [FREE Full text] [doi: 10.1186/1479-5868-9-15] [Medline: 22340137]
7. Fisher D, Louw Q. Primary school learners' movement during class time: perceptions of educators in the Western Cape, South Africa. BMC Public Health 2023 Dec 13;23(1):2501 [FREE Full text] [doi: 10.1186/s12889-023-17428-3] [Medline: 38093224]
8. Gibson CA, Smith BK, DuBose KD, Greene JL, Bailey BW, Williams SL, et al. Physical activity across the curriculum: year one process evaluation results. Int J Behav Nutr Phys Act 2008;5(1):36. [doi: 10.1186/1479-5868-5-36]
9. O'Sullivan K, O'Sullivan P, O'Keeffe M, O'Sullivan L, Dankaerts W. The effect of dynamic sitting on trunk muscle activation: a systematic review. Appl Ergon 2013 Jul;44(4):628-635. [doi: 10.1016/j.apergo.2012.12.006] [Medline: 23369370]
10. Hoy D, Brooks P, Blyth F, Buchbinder R. The epidemiology of low back pain. Best Pract Res Clin Rheumatol 2010 Dec;24(6):769-781. [doi: 10.1016/j.berh.2010.10.002]
11. Clemes SA, Bingham DD, Pearson N, Chen Y, Edwardson CL, McEachan RRC, et al. Stand out in class: restructuring the classroom environment to reduce sitting time - findings from a pilot cluster randomised controlled trial. Int J Behav Nutr Phys Act 2020 Apr 29;17(1):55 [FREE Full text] [doi: 10.1186/s12966-020-00958-z] [Medline: 32349761]
12. Hinckson EA, Aminian S, Ikeda E, Stewart T, Oliver M, Duncan S, et al. Acceptability of standing workstations in elementary schools: a pilot study. Prev Med 2013 Jan;56(1):82-85. [doi: 10.1016/j.ypmed.2012.10.014] [Medline: 23103223]
13. Campbell MK, Piaggio G, Elbourne DR, Altman DG, CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. BMJ 2012 Sep 04;345(sep04 1):e5661-e5661. [doi: 10.1136/bmj.e5661] [Medline: 22951546]
14. Fisher D, Louw Q, Thabane L. Sedentariness and back health in Western Cape primary school students: protocol for a pragmatic stepped-wedge feasibility randomized controlled trial. JMIR Res Protoc 2020 Nov 30;9(11):e18522 [FREE Full text] [doi: 10.2196/18522] [Medline: 33089783]
15. Hemming K, Taljaard M, Grimshaw J. Introducing the new CONSORT extension for stepped-wedge cluster randomised trials. Trials 2019 Jan 18;20(1):68 [FREE Full text] [doi: 10.1186/s13063-018-3116-3] [Medline: 30658677]
16. Copas AJ, Lewis JJ, Thompson JA, Davey C, Baio G, Hargreaves JR. Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. Trials 2015 Aug 17;16(1):352 [FREE Full text] [doi: 10.1186/s13063-015-0842-7] [Medline: 26279154]

XSL•FO
RenderX

17. van Dyk H, White C. Theory and practice of the quintile ranking of schools in South Africa: a financial management perspective. SAJE 2019 Sep 30;39(Supplement 1):S1-S9. [doi: 10.15700/saje.v39ns1a1820]
18. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot Feasibility Stud 2016 Oct 21;2(1):64. [doi: 10.1186/s40814-016-0105-8]
19. Ridley K, Ridgers ND, Salmon J. Criterion validity of the activPAL™ and ActiGraph for assessing children's sitting and standing time in a school classroom setting. Int J Behav Nutr Phys Act 2016 Jul 07;13(1):75 [FREE Full text] [doi: 10.1186/s12966-016-0402-x] [Medline: 27387031]
20. Contardo Ayala A, Salmon J, Timperio A, Sudholz B, Ridgers N, Sethi P, et al. Impact of an 8-month trial using height-adjustable desks on children's classroom sitting patterns and markers of cardio-metabolic and musculoskeletal health. Int J Environ Res Public Health 2016 Dec 10;13(12):1227 [FREE Full text] [doi: 10.3390/ijerph13121227] [Medline: 27973414]
21. Sherry AP, Pearson N, Ridgers ND, Johnson W, Barber SE, Bingham DD, et al. Impacts of a standing desk intervention within an English primary school classroom: a pilot controlled trial. Int J Environ Res Public Health 2020 Sep 26;17(19):7048 [FREE Full text] [doi: 10.3390/ijerph17197048] [Medline: 32993142]
22. Balasubramania S. Comparison of Angle Measurements between Vicon and Myomotion Systems. AZ: Center for Adaptive Neural Systems (ANS), Arizona State University; 2013.
23. Friese S. Qualitative Data Analysis with ATLAS.ti. Third edition. Thousand Oaks, CA: SAGE Publications; 2019:xxxiii.
24. Routen A, Johnston J, Glazebrook C, Sherar L. Teacher perceptions on the delivery and implementation of movement integration strategies: the CLASS PAL (Physically Active Learning) Programme. Int J Educ Res 2018 Mar;88:48-59. [doi: 10.1016/j.ijer.2018.01.003]
25. Jones G, Longbon K, Williams S. Exploring the acceptability and feasibility of a whole school approach to physical activity in UK primary schools: a qualitative approach. BMC Public Health 2022 Nov 30;22(1):2236 [FREE Full text] [doi: 10.1186/s12889-022-14647-y] [Medline: 36451192]
26. Pearson M, Chilton R, Wyatt K, Abraham C, Ford T, Woods H, et al. Implementing health promotion programmes in schools: a realist systematic review of research and experience in the United Kingdom. Implement Sci 2015 Oct 28;10(1):149. [doi: 10.1186/s13012-015-0338-6]
27. Wilson P, Mathie E, Keenan J, McNeilly E, Goodman C, Howe A, et al. ReseArch with Patient and Public invOlvement: a RealisT evaluation – the RAPPORT study. Health Serv Deliv Res 2015 Sep;3(38):1-176. [doi: 10.3310/hsdr03380] [Medline: 26378332]
28. Sanders EB, Stappers PJ. Probes, toolkits and prototypes: three approaches to making in codesigning. CoDesign 2014 Mar 06;10(1):5-14. [doi: 10.1080/15710882.2014.888183]
29. Schultz L, Ruel-Bergeron J. Considerations for monitoring school health and nutrition programs. Front Public Health 2021 Jul 16;9:645711 [FREE Full text] [doi: 10.3389/fpubh.2021.645711] [Medline: 34336757]
30. Clifford L, Tyler R, Knowles Z, Ashworth E, Boddy L, Foweather L, et al. Co-creation of a school-based motor comeptence and mental health intervention: move well, feel good. Children (Basel) 2023 Aug 17;10(8):1403 [FREE Full text] [doi: 10.3390/children10081403] [Medline: 37628403]
31. Tobia V, Sacchi S, Cerina V, Manca S, Fornara F. The influence of classroom seating arrangement on children's cognitive processes in primary school: the role of individual variables. Curr Psychol 2022 Oct 31;41(9):6522-6533 [FREE Full text] [doi: 10.1007/s12144-020-01154-9] [Medline: 33162725]
32. Wannarka R, Ruhl K. Seating arrangements that promote positive academic and behavioural outcomes: a review of empirical research. Support Learn 2008 May 06;23(2):89-93. [doi: 10.1111/j.1467-9604.2008.00375.x]
33. Edwardson CL, Winkler EA, Bodicoat DH, Yates T, Davies MJ, Dunstan DW, et al. Considerations when using the activPAL monitor in field-based research with adult populations. J Sport Health Sci 2017 Jun;6(2):162-178 [FREE Full text] [doi: 10.1016/j.jshs.2016.02.002] [Medline: 30356601]
34. Shi Y, Huang W, Yu J, Sheridan S, Sit CHP, Wong SHS. Compliance and practical utility of continuous wearing of activPAL™ in adolescents. Pediatr Exerc Sci 2019 Aug 01;31(3):363-369. [doi: 10.1123/pes.2018-0170] [Medline: 30605620]
35. Maddocks M, Byrne A, Johnson CD, Wilson RH, Fearon KCH, Wilcock A. Physical activity level as an outcome measure for use in cancer cachexia trials: a feasibility study. Support Care Cancer 2010 Dec 3;18(12):1539-1544. [doi: 10.1007/s00520-009-0776-2] [Medline: 19956982]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
FGD: focus group discussion
HEMV: health education and movement video
IDI: individual depth interview
IMU: inertial measurement unit
NCD: noncommunicable disease

XSL•FO
RenderX

RCT: randomized controlled trial
SB: sedentary behavior

Edited by A Mavragani; submitted 07.08.24; peer-reviewed by F Yang; comments to author 29.10.24; revised version received 18.11.24; accepted 28.11.24; published 24.02.25.

Please cite as:
Fisher D, Maart R, Thabane L, Louw Q
A Classroom-Based Intervention for Reducing Sedentary Behavior and Improving Spinal Health: Pragmatic Stepped-Wedge Feasibility Randomized Controlled Trial
JMIR Form Res 2025;9:e65169
URL: https://formative.jmir.org/2025/1/e65169
doi:10.2196/65169
PMID:

©Dominic Fisher, Rentia Maart, Lehana Thabane, Quinette Louw. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Formative Evaluation of Suicide Prevention Websites for Men: Qualitative Study with Men at Risk of Suicide and with Potential Gatekeepers

Doreen Reifegerste[1], Prof Dr; Anna J M Wagner[2], Prof Dr; Lisa Huber[3], MA; Manuel Fastuca[3]

[1]School of Public Health, Bielefeld University, Bielefeld, Germany

[2]Centre for Language Studies, Radboud University, Nijmegen, The Netherlands

[3]GIM Gesellschaft für innovative Marktforschung mbH, Heidelberg, Germany

Corresponding Author:
Doreen Reifegerste, Prof Dr
School of Public Health
Bielefeld University
Universitaetsstrasse 25
Bielefeld, 33615
Germany
Phone: 49 521 10667 ext 807
Email: doreen.reifegerste@uni-bielefeld.de

Abstract

Background: The suicide rate among men exceeds that of women worldwide. One important measure in suicide prevention for men is digital communication interventions, as they enable easy and anonymous access to information resources. This is especially important for men who might not be reached by traditional, in-person prevention methods. Thus, as part of an interdisciplinary project on suicide prevention for men, two specific digital communication prevention measures were developed: (1) a website to inform men at risk about suicide prevention, and (2) a website to educate potential gatekeepers who are in contact with men at risk of suicide about appropriate life-saving measures. Both websites needed evaluation to explore how they are perceived by (1) men and by (2) potential gatekeepers of men at risk of suicide. This is crucial, as existing research lacks formative evaluation that informs the development of intervention communication materials.

Objective: This study aimed to analyze whether these websites were perceived as (1) comprehensible and engaging, (2) authentic and trustworthy, as well as (3) useful by (potential) users. Furthermore, we examined (4) additional ideas for effective communication about suicide prevention.

Methods: We conducted (1) individual videoconference interviews with 24 men to evaluate the website and (2) four focus groups with 8 gatekeepers in each group (32 participants) to evaluate the online education program. The focus group sample was equally distributed regarding gender and age. Recruitment was conducted together with a field research partner who posted adverts on Facebook and Instagram (Meta) to reach as many potential participants as possible in an efficient way. All participants were asked to evaluate the intervention materials using a fictitious scenario of a man experiencing a mental health crisis before the interviews or focus groups took place.

Results: The videos were perceived as (1) catchy, comprehensible, and empathetic, but too long for a short introduction. A balanced mix of emotional and informative content was considered appropriate and helpful. The health information provided was perceived as (2) serious and trustworthy due to citing scientific institutions and video material of men who had experienced suicidal ideation. (3) The intervention's applicability for men experiencing acute crisis was critiqued, but it was regarded as very useful for comprehensive information. (4) Further communication channels and addressing other male subgroups or gender identities were presented as possible extensions of the program.

Conclusions: Effective suicide prevention research should address both the groups at risk and their support network. Digital communication interventions can provide low-threshold access. Videos with personalized examples are important to give men someone to identify with, which validates their emotional responses and supports their self-esteem, while videos with experts provide relevant and credible information.

(JMIR Form Res 2025;9:e59829) doi:10.2196/59829

XSL•FO
RenderX

KEYWORDS

mental health; suicide prevention; men's health; evaluation; website; gatekeeper; suicide; male; suicide risk; digital communication; intervention; suicidal behavior; digital intervention; digital media

Introduction

Suicide rates among men surpass those among women worldwide. For instance, in Germany, approximately 75% of all suicides are carried out by men [1]. Even though recent reviews have emphasized advances in our understanding of risk factors and countermeasures for suicidal behavior in men and women [2,3], it remains a reality that men are at substantially higher risk. Preventive measures are hence needed that specifically address men and consider the particular needs of men at risk and their suicide-related communication [4].

One important measure in suicide prevention is digital communication interventions. Digital media such as websites are used as information sources on suicide methods but could also be used as easily accessible sources of suicide prevention support [5]. As they enable easy and anonymous access to information resources, digital media can be beneficial tools for a variety of different population groups [6]. Men in particular might profit from digital interventions, as they are less likely to be reached by traditional, in-person prevention methods that might not be tailored to their needs [7]. Recent systematic reviews [8,9] show that communication-based digital interventions appear to be effective measures for male suicide prevention. While the interventions demonstrate high overall effectiveness on suicide-related outcomes (eg, knowledge, attitudes, and cognitions), one of the reviews [9] also found no adverse effects linked to these interventions among men. This means that the risk associated with digital interventions for men at risk seems to be particularly low.

In addition, interventions for so-called "gatekeepers" (people who are in contact with individuals at risk of suicide, eg, family members or friends) play a central role in suicide prevention [10]. They can be considered to fulfill an important bridging function, extending intervention reach to this group of men at risk and being well-placed to initiate difficult conversations [11]. Sharing distressing personal experiences with trusted social contacts can be an important first step in seeking psychological support, for example in the case of clinical depression [12]. Men seem to be less likely to seek professional or informal help and less likely to communicate directly to their closest contacts about suicidal intentions; however, they still show signs of change (eg, further self-isolation, sudden anger, or hopelessness) before the suicide, which makes gatekeeper interventions all the more important [13,14].

Training gatekeepers on how to approach men at risk, and react appropriately in case of disclosure, is of paramount importance. Based on evidence from a systematic review, gatekeepers can be effective in their bridging function for suicide prevention [15], particularly where training programs succeed in sensitizing gatekeepers and building their competencies for engaging with men at risk [16]. Digital training programs for adolescents have been shown to be successful in the development of gatekeepers' competencies, such as for teachers and peers [17-19].

To date, however, further evaluations for digital interventions for adult men and their (potential) gatekeepers are required [20]. This especially holds true for qualitative research evaluations, such as with interviews and focus groups. While randomized controlled trials (RCTs) and quantitative evaluations are necessary for testing programs' effectiveness and have also been completed within this project [11], qualitative methods offer the chance to not only test the usability of the developed programs but can also generate insights extending beyond the specific intervention [21].

We therefore conducted two qualitative studies evaluating (1) a website addressing particularly men at risk about suicide prevention, and (2) a web-based program to educate potential gatekeepers about appropriate and critical life-saving measures. The communication strategies selection for the websites was based on the findings of recent systematic reviews, which focused on effective message strategies and communication campaigns for suicide prevention [8,9]. Here, it was recommended to include testimonials and educational materials. The content was also informed by research on gender-specific suicide-related communication [14] as well as the typical communication behavior of men in the days before a suicide [13]. For more details on the structure, components, and materials of both websites [11,22].

The website is specifically aimed at men experiencing acute (suicidal) crises and provides visitors with information about suicidal experiences and behavior, offers of help, and signposts ways to access health provision and mental health support. Using short videos, men at risk talk about their experiences with suicidal crises and how they found their way out of the crisis. In addition, the videos address images of masculinity, stigmatization, shame, and overcoming obstacles to seeking help. Furthermore, experts from various professions within the health and care sectors provide more detailed information. These experts include a male professor and psychiatrist as well as a female professor and psychologist, all of whom primarily conduct research on mental health and suicide.

The web-based program for gatekeepers is aimed at people who are worried about a loved one and consists of four modules. The first module provides information about suicidal experiences and behavior, detailing warning signs. The second module explores communication and conversation strategies. The third module focuses on support services for signposting. The last module addresses the burden on relatives themselves, and their options for self-care, and gives further recommendations for stress-management strategies.

The web-based program (similar to the website) consists of videos featuring men at risk and experts from the health and care sectors. Participants also learn about communication strategies and conversation skills using exemplar audio recordings of people with suicidal experiences and behavior.

XSL•FO
RenderX

There were multiple focuses when evaluating both programs, given their differing objectives and content. On a more superficial level, usability evaluations were key. These have the objective of analyzing whether the digital interface is easy to understand and appealing to potential users [21]. Usability, in general, "is the ability of a product to be understood, learned, used, and attractive to the user" [21], which emphasizes a user-centered approach and focuses on their satisfaction. In health-related contexts, the credibility and authenticity of information is another important determinant of user satisfaction [23]. Because there is an excess of information available online (of varying quality) and because people at risk are very vulnerable, trust is key to their health information-seeking behaviors [24]. Often, expert sources as well as individuals with personal experience can be convincing in this regard.

Finally, the aspect of perceived utility should be considered, because this also influences the intention to keep using the sites and recommendations to others.

Thus, in our evaluations, we aim to analyze whether the website and the web-based gatekeeper program are perceived as (1) comprehensible and engaging, (2) trustworthy, and (3) useful by the two different audiences. Furthermore, we were interested in (4) additional ideas for the communication of suicide prevention beyond the developed digital interventions for men.

Methods

Participants

To pursue these 4 research goals, a 2-step study with both men and gatekeepers was conducted.

For the evaluation of the websites for males, 24 videoconference interviews with men in 3 age groups (18 to 30 years old, 31 to 50 years old, 50 years and older; 8 participants per age group) were conducted.

We included all persons who identified as men and were older than 18 years. We did not explicitly search for men with a history of suicidal behavior nor relatives of men with a history of suicidal experience and behavior given ethical considerations. However, based on the recruitment information (which disclosed the aim of the website as one for suicide prevention), many participants reported that they themselves were affected by suicidal ideation and behavior, or had vicarious experiences through friends, acquaintances, or relatives. We did not exclude these participants from the sample.

For the evaluation of the gatekeeper program, four videoconference focus groups with gatekeepers with 8 people each (32 participants in total) were conducted. The focus group sample was equally distributed regarding gender and age (50%, 16/32 female and 50%, 16/32 male; 50%, 16/32 individuals from younger, ie, under 40 years old, and 50%, 16/32 from older age groups, ie, over 40 years old).

For an overview of the specific age and gender of each participant refer to Table 1. A videoconference format for the interviews and the focus groups was chosen to maximize ease of attendance and the potential for anonymous participation.

XSL•FO
RenderX

Table 1. Age and gender of participants in interviews (24 men) and focus groups (16 men and 16 women in 4 groups) for the evaluation of websites for suicide prevention.

Method and gender	Age (years)	Interviewer or moderator
Interviews		
Male	48	Female
Male	63	Female
Male	39	Female
Male	24	Male
Male	48	Male
Male	67	Female
Male	56	Female
Male	66	Female
Male	62	Female
Male	49	Female
Male	30	Female
Male	33	Female
Male	30	Female
Male	29	Female
Male	35	Male
Male	36	Male
Male	40	Male
Male	65	Male
Male	25	Male
Male	24	Male
Male	54	Female
Male	51	Female
Male	20	Female
Male	21	Female
Focus group 1		Male, Female
Female	60	
Female	58	
Female	55	
Female	47	
Female	41	
Female	72	
Female	41	
Female	51	
Focus group 2		Male, Female
Male	35	
Male	18	
Male	36	
Male	27	
Male	33	
Male	34	

XSL•FO
RenderX

Method and gender	Age (years)	Interviewer or moderator
Male	27	
Male	40	
Focus group 3		Male, Female
Female	22	
Female	34	
Female	32	
Female	24	
Female	29	
Female	30	
Female	37	
Female	30	
Focus group 4		Male, Female
Male	71	
Male	44	
Male	65	
Male	58	
Male	61	
Male	75	
Male	46	
Male	41	

Sample sizes were therefore consistent with recommendations for sample sizes in interviews and focus groups to reach sufficient saturation [25]. As it happens, data saturation in the interviews was already reached after about 20 interviews. The number of people refusing to take part in the study or not showing up at the scheduled time was low to nonexistent (especially in comparison to market research projects).

Recruitment

Recruitment was conducted together with the field research partner Quotapoint Pharma and Health in Germany, which specializes in recruiting participants for qualitative studies on health-related topics. They used an announcement in their newsletter (refer to Multimedia Appendix 1 for an English version of the recruitment material). They also posted adverts on Facebook (to reach potential participants older than 35 years) as well as Instagram (to reach potential participants younger than 35 years). This approach to recruitment was chosen in order to reach as many potential participants as possible in an efficient way. After the initial contact was made, a screening process was carried out using a screening tool, with which suitable participants were filtered according to the sample criteria. During the screening process, a brief sociodemographic check was conducted due to the all-male target group for the interviews and the gender-mixed sample for the focus groups. Following this, a trigger warning was given that the subsequent screening interview would be about suicide prevention to minimize the psychological stress on the potential participants. Participant recruitment was documented with regard to the filtering criteria, for example, age or gender, and checked before the participants were finally invited to the interviews and focus groups.

Ethical Considerations

Ethical approval for this study was obtained from the Ethics Committee of Bielefeld University (number 2023-153). All participants were provided with an information form describing the study (refer to Multimedia Appendix 2 for an English translation of this consent form). Participants in both studies were required to return signed consent before they could participate in the study, via a web-based form and a data protection agreement. As the interviews and focus groups took place as a videoconference, the participants were once again reminded about their data and confidentiality rights at the start of the interviews. The interviews were recorded only once active consent to recording had been given by the participants.

As stated above, we avoided explicitly searching for men with a history of suicidal behavior consideration due to ethical considerations.

All data were collected anonymously. All participants were provided with financial compensation in recognition of time spent and their contribution to the study (€90 [US $94] per participant in interviews; €140 [US $146] per participant in focus groups).

Procedure

After giving their consent, all participants were asked to extensively study the intervention materials and briefly evaluate them via a quantitative survey. This was both for the website

XSL•FO
RenderX

for men at risk and the gatekeeper online program (step 1), and was completed before the interviews and focus groups took place (step 2). In the first step, called the pretask, the participants were asked to engage with and reflect on the content in a home setting. Initial findings on the attractiveness, intuitiveness, and comprehensibility of the content were collected using short questions via online survey software. Surveys took approximately 5-10 minutes to complete. Responses from the surveys were included as context information in the qualitative data analysis and were used during the interviews as conversation starters.

In the second step, the interviews and focus groups were conducted in July 2023 using the ZoomX videoconferencing platform (Zoom and Telekom), which is hosted in Germany. After a brief summary of the project (and in the case of the focus group a round of introduction), participants were informed about the research goals, their ability to end participation at every step of the data collection, and about the applicable data protection rules. Moreover, they were given the opportunity to ask questions before the interviews and focus groups began.

All interviews were conducted by one of two authors (LH and MF; one male and one female) and ran for 45 minutes. All focus groups were conducted jointly by one moderator and one assistant and ran for approximately 90 minutes (refer to Table 1 for details on role allocation). These were recorded, using both audio and video. During the interviews, the interviewer took notes on the content of participant discussions. Meanwhile, during the focus groups, the assistant listened to the discussion of the moderator and participants, while taking handwritten notes on all discussion points.

Materials

The website for men with an increased risk of suicide entitled “Männer-stärken” (empowering or strengthening men) offers information on suicide prevention. The website contains videos, information summaries, and lists of helpful recommendations. Complex content is conveyed in simple language and with the help of visual illustrations. Long text sections are avoided in favor of concise information summaries and lists. Graphic illustrations feature a line art and watercolor design. The website is primarily designed in yellow-themed color.

The web-based program “Help for Relatives” provides information on the topic of suicide or suicide prevention in men, with the help of text content on the website and an accompanying manual, as well as four videos approximately 20 minutes long each. It is structured into four modules: (1) suicidal ideation and behavior, (2) communication, (3) offers of assistance, and (4) self-care. Here, visual elements are additionally used to convey complex information.

The short, standardized web-based surveys conducted next assessed a range of topics relevant for website evaluation (confirming participant identity, first impressions of the web-based platform, liking or disliking different aspects as open text fields and grading items as well as optimization potential) comprising closed- and open-ended questions. Their results were used as conversation prompts to provide an easy entrance into the interview situation and generate conversation.

The interviews and focus groups were structured, with a series of guiding questions and prompts exploring the views of men and gatekeepers on the potential design, content, functioning, and user experience (eg, optional phone and email support) of a web-based navigation tool (eg, liking or disliking the online platform for different aspects such as understandability, authenticity, the trustworthiness of given information, and the usability in everyday life and in acute crisis). Questions and prompts were developed based on mental health research as well as usability research. Participants were encouraged to share their perspectives verbally with the group or the interviewer.

Several techniques were used to deal as sensitively as possible with the topic of suicide. One technique was the scenario technique or vignette methodology [26], which was used to keep the interviewees’ psychological stress as low as possible and enable them to evaluate the digital interventions on the basis of a standardized crisis situation: a fictitious scenario of a man in a psychological crisis was presented, which served as the basis for the interviews. In addition (and also to minimize emotional stress), we first asked the men indirectly about the assumed perceptions of others, before they were asked to present their own views. This so-called circular questioning is a technique often used in therapy to investigate sensitive topics [27]. All materials were pretested in advance with persons matching the sample structure (refer to Multimedia Appendix 3 for the guides and short protocols of the interviews and the focus groups).

Data Analysis

Regarding personal reflections that were relevant to our analysis, there were methodological as well as gender aspects to consider during the data analysis phase. One part of the team comprises established scientific researchers, who are rooted in theory-based quantitative and experimental as well as qualitative research, while the other part has long experience in market research for new websites and consumer products. We tried to link these 2 worlds and systemize the content of the interviews and focus groups against the background of theories like the unified theory of acceptance and use of technology and similar theories [28] as well as known consumer needs for products in digital environments (Multimedia Appendices 4 and 5). Such products, like cars or digital devices also often address a primarily male audience.

At the same time the three researchers in the team who identify as female tried to be especially empathetic to male perspectives, whereby we for instance explicitly asked for input from the male researcher in our team.

Taking into account the reflexivity of the researchers, it is important to reflect on the different perspectives and life experiences of the researchers who did the coding (a man over 50 and a woman under 30) in terms of how this could influence the analysis of the data. In the sense of Clark and Braun [29], this serves to make the subjective influences of the researchers transparent and to consider them as a research resource rather than a bias. On the one hand, it is therefore important to emphasize age and generational perspectives. The male researcher over the age of 50 years has a more experienced perspective than the female researcher under the age of 30 years,

XSL•FO
RenderX

which proved to be a suitable age structure with regard to the sample of interviewees aged 18-60 years, as all age splits could be covered within the researcher perspectives. At the same time, generationally different views on suicide, suicide prevention, and mental health can be countered from an age perspective. Furthermore, by choosing a man and a woman to analyze and collect the data, gender-specific needs of the participants and insights can be gained from the data. As a man, the male researcher is more attuned to the lifeworld issues and needs of the male target group, whereas the female researcher is more sensitive to gender-specific differences. Finally, the degree of digitality can also be reflected in the analysis and collection of data. It can be assumed here that the younger female researcher has a high level of familiarity with digital technologies and communication styles, which could make a difference in the consideration and interpretation of what the participants say with regard to the user-friendliness and address of the websites. Due to the selected age range of the sample, it is also possible to take into account any degree of digitality of the participants.

Qualitative data were managed using transcripts, an Excel table, and protocols of the interviews and the focus groups. A thematic analysis [29,30] was conducted by 2 authors on interview and focus group data notes, chat conversations, and survey responses to identify and summarize the key topics and preferences raised by the interview and focus group participants within each area of the research questions while preserving the breadth and diversity of perspectives presented. Analyses and developing categories were regularly discussed with all members of the research team.

A combination of deductive and inductive approaches to coding was hence applied to the research. Deductive is here defined as linking the answers to the theoretical and empirical background. More specifically, deductive themes were derived from literature and applied to the material. In contrast, inductive approaches entail being open for topic-specific feedback and including themes that arise from the material and which had not been previously assumed by the researchers.

With regard to the methodological explanations by Braun and Clarke [29] regarding established coding practices, the decision was therefore made to use “mashups,” that is, a mixture of existing literature-based codes and immanent interpretation. This is due to the special feature of the study that, in addition to the content-related exploration of the relevance of such websites for the target group, aspects of user experience were also examined, which sometimes include merely technical aspects.

Before starting the data analysis, an initial broad coding framework was developed based on the key areas of comprehensibility, appeal, credibility, utility, and ideas for further communication measures in relation to the presentation of the website content. The data was coded deductively within each category. As the coding progressed, subcategories that were not included in the deductive framework were developed inductively, such as the relevance of suicide prevention services as a website per se.

Results

Comprehensibility and Appeal

With regard to our first aim of analyzing the comprehensibility and appeal of the videos, findings showed that they were perceived as catchy, understandable, and empathetic by both men and gatekeepers, but too lengthy. Men of all 3 age groups evaluated the website as looking serious, clear, and professional. They especially noted the yellow colors as being positive, friendly, and informative. In contrast, they would perceive the color black as being too masculine, sad, and aggressive.

The content was perceived as easy to receive due to simple and understandable language, and plain typography in combination with images and videos that make complex topics accessible to the general public. The combination of images, video, and text was considered appropriate and varied. Especially the graphic illustrations, short sentences, and comprehensible statistics were perceived as more effective than “text deserts” and allow easy understanding:

> *Numbers also make the whole thing a bit more tangible, I'm perhaps more of a numbers person, so it makes sense in that direction, but I always find it much easier to understand, otherwise you're in kind of an empty space where you talk around [the topic] without having any numbers.* [male, 24 years, interview]

For some videos in the gatekeeper program, the language from featured experts was perceived as being too scientific that is, using too many scientific terms.

Highly modern or fancy websites were not expected by most viewers given the seriousness of the topic; rather, they expected timeless design and a clear presentation of facts.

> *If I make the thing colorful and pretty, then it loses a bit of its meaning;* [male, >41 years, focus groups]

> *It just has to be good and clear and, not like a thousand other sites, I always knew how to navigate [the site] and found guidance.* [male, 62 years, interview]

However, participants also rated the website as too unmodern due to long texts and videos. Especially the videos (on average 7-8 minutes and up to 20 minutes in length) were perceived as too lengthy. Feedback indicated that these demanded too much attention and increased the risk of participants discontinuing the program. The same applies to searching and scrolling on the website, which was also perceived as not user-friendly. Respondents suggested making the aim of the website for men at risk and the gatekeepers’ program clearer at first glance. Specifically, it was suggested to offer specific content sections and “trailer-length” videos for a first impression. The option of more detailed formats should be an option for cases where more information is required. Some even wanted more facts and information in this regard.

Some of the men perceived the title of the website for men (which means empowering or strengthening men) as nonintuitive

XSL•FO
RenderX

and misleading because they had associations with potency substances or coaching about masculinity.

> *If I didn't know the context, then I would think of, I'll say, potency or something like that.* [male, 66 years, interview]

> *well, strengthening can be in many ways, so that could also be some sexual enhancers or some power food supplement stores for bodybuilders.* [male, 51 years, interview]

In addition, they found the claim of "strengthening men" rather counterproductive for the destigmatization of mental distress, which they felt should not be connected to weakness.

> *For me, it's more of an illness that can be treated like any other illness and has nothing to do with strength or weakness. So, I would tend to think that it's the supposedly strong men, who are strong in the classic sense, who can get it because they're not so good at dealing with problems.* [male, 67 years, interview]

Credibility and Authenticity

In terms of authenticity and trustworthiness, the information on the website and in the web-based program was perceived as serious and trustworthy due to referencing universities, health insurance companies, and scientific institutions as sources as well as facts and figures. In addition, the perceived clarity, color choice, and timelessness were attributed as professional and serious.

> *What naturally contributes to credibility is that the logos of the participating universities were right at the front of the home page. Then there were these experts, which of course also increased credibility to a certain extent.* [male, 33 years, interview]

> *I found the interview with the professor very, very good, especially because you can see how extensive the problem is, where he had his problems, and to understand that there is a big process behind it.* [male, <40 years, focus group]

The same perception applies to facts and statistics as an integral part of evidence-based information.

Several men in the interviews stated that it is good that the programs' developers are based in universities. This lent credibility to the websites, which were perceived as trustworthy. The men were sure that the websites were based on the latest research rather than being driven by advertising or a pharmaceutical company for reasons of profit, even when they did not recognize all the universities involved.

In addition, including links to other sources increased the perception of content presented as authoritative.

Participants also rated the videos featuring men at risk of suicidal behavior as authentic and easy to understand. Including these videos offered visitors faces to identify with, and validation for experiences. It also created the feeling for visitors who self-identified as at risk that they are not alone with problems (even without having spoken to others personally about the topic), making it possible to still feel a sense of belonging.

> *Especially with the videos, it brings a personal closeness and makes it easier to accept support. Because you realize that there are also real men who have the same problems.* [male, 36 years, interview]

At the same time, using and engaging with these vignettes can offer a supported form of self-diagnosis which helps men at risk to classify their own feelings and symptoms.

> *That you can draw parallels from yourself to a description or identify with a person who also describes [these experiences].* [male, 56 years, interview]

That the men at risk featured in the videos were of various ages was also perceived as helpful to recognize that this is a problem for all ages.

A video featuring individuals sharing their own experiences can also serve a more inspiring purpose. Sharing these experiences and also how that person also managed to recover can demonstrate how this is possible, changing perceptions of how good the chances are of a recovering person also depicts an example of how good the chances of recovery can be:

> *There were also patients in it, I think several times, I think the same person was in several videos. (Which then takes you by the hand, along the lines of) Look, I'm already here, why don't you come too, you can see I'm not doing so badly. Or it makes me feel better.* [male, 30 years, interview]

In contrast, the men at risk found archive photos of emotional states on the website for the gatekeepers as noncredible, perceiving these instead as too dramatic and inauthentic. This underlines how important the credibility of the material is, as where material otherwise seems staged, inauthentic, and prompts an adverse reaction:

> *So, when I'm told what I'm supposed to feel in such a way, I am out… and that's what these stock photos do, because in my opinion they want to create something very specific and emphasize what I as a consumer am supposed to think.* [male, <40 years, focus group]

Utility

In terms of utility, the websites are perceived as helpful overall. The participants think that the websites are a good place to find information, especially because men at risk also shared their experiences. Because of this inclusion, the information shared was perceived to be helpful as realistic and practical advice by users. Behavioral recommendations were considered to be helpful tips, which were highly relevant to everyday life.

With regards to the website, one man (21 years) reported in an interview:

> *I particularly liked the "Glossary of concerns" and the "Addressing distress" pages, because these are the most practical in my opinion. I also really liked the fact that all the texts always included a link to*

XSL•FO
RenderX

> *further help or resources. These resources were very well structured based on different needs. Contact details were also given everywhere.*
>
> *I mean, you get a lot of information so that you can perhaps recognize the signs [...] when you see the symptoms... irritability, listlessness, and so on.* [male, 36 years, interview]

The gatekeepers reported similar reactions to their program:

> *It was very much about.... how do I try to understand a person and how do I communicate with that person so that the person appreciates it, that's what you're allowed to do?* [male, <40 years, focus group]

The participants found digital anonymity on the internet and the possibility of searching for information without being obliged to do something helpful. According to them, it provides men at risk and concerned family members support with low barriers to access because help is provided in anonymous internet spaces for this very sensitive and stigmatized topic. Several men in the interviews and focus groups stated that they experienced less shame when talking in a videoconference about mental health problems and more potential to engage with the idea of getting help, because of the anonymity.

The help site for relatives was considered to provide significant added value emotionally, as it can support the family members to minimize difficult feelings related to powerlessness.

> *If the people notice something in their surroundings that they can then check with this program as a first step, ok, are these worries that I'm having justified, is this really going in the direction that I'm thinking?* [female, <40 years, focus group]

At the same time, however, the applicability in crises is questioned as to whether men at risk of suicide in an acute crisis would still visit websites to get help because they might not have the energy to do so:

> *So within this situation, he might not want to read something for hours or anything like that, but he's overstimulated, so he doesn't know what to do next, he just needs someone with whom he can somehow communicate verbally, to one.* [male, 66 years, interview]

Some were even concerned that the "unregulated internet space" could promote a problematic approach to the topic of suicide risk and counteract the actual aim of the service. They were worried that such a website could increase the risk of seeking out further information and then receiving "bad advice" or "getting the wrong person" in other online spaces, which would then hinder them from seeking help in "real life."

Participants suggested prefiltering the offers of help for those at risk in an acute crisis, by providing emergency numbers and direct contact persons or a helpline who can offer help in a crisis. Given that websites provide information asynchronously, the hope for immediate, quick help for relatives in emotional distress (contact details, chats, and direct contact options for immediate personal offers of help) remains unfulfilled.

Table 2 summarizes the answers to RQ1 to RQ3.

Table 2. Summary of Results of the Interviews and the Focus Groups and their Feedback on Comprehensibility, Credibility, and Utility of the Websites for Suicide Prevention.

Theme	Interviews with men (24 participants)	Focus groups with Gatekeepers (32 participants)
Comprehensibility and appeal	• Videos were perceived as catchy, understandable, and empathetic • But too lengthy and too fixed • Website is perceived as authoritative, easily understandable • Yellow color as positive, friendly, and informative • The website title was discussed as it seemed ambiguous	• Videos were perceived as catchy, understandable, and empathetic • But too lengthy and too fixed • Too many scientific terms in the expert videos • A modern or fancy design is not expected
Credibility and authenticity	• Websites are perceived as trustworthy due to references to scientific institutions as well as facts and figures • Generic stock photos provoke adverse responses	• Websites are perceived as trustworthy due to references to scientific institutions as well as facts and figures • Generic stock photos provoke adverse responses
Utility	• A good place to find information • But might not be appropriate or helpful in emergency cases	• A good place to find information • But might not be appropriate or helpful in emergency cases

Ideas Beyond the Intervention Material: Other Communication Channels

Both the topic, the desire for education about the topic, and its destigmatization were evaluated as critically important, especially from the group of men younger than 30 years. As a consequence, the respondents also recommended using more communication channels for promotion in both the health and public sector to make the information more accessible to everyone. They suggested using all conceivable communication channels to draw attention to this offer of help, to destigmatize the topic, and to widen the reach for education purposes.

> *I think this should be widely distributed now, because ultimately, of course, I have to reach the men everywhere* [man, 51 years, interview]

XSL•FO
RenderX

This could involve promoting the websites more prominently on social media channels like Instagram and TikTok (ByteDance), in magazines, outdoor media, TV health shows, cinema and radio spots, at train stations, and in sports stadiums.

They also suggested connecting the website for men at risk more visibly with the online program for gatekeepers. This would mean that where the website for men is mentioned in a private conversation with a friend (gatekeeper) who has already screened for information and helpful content from the website. Conversely, a man in acute crisis could recommend the program for the gatekeepers to his relatives.

While it was considered valuable that topics around masculinity and suicidal ideation are being addressed, suicide prevention services specifically for men were questioned as a concept by the participants. Despite the information about the statistically higher suicide rate among men, some of the content produced on the sites was considered too conservative for the way that gender roles were presented, because they offered no identification potential for other gender identities, such as nonbinary, intersex, or transgender persons.

Discussion

Main Results

The goals of the study were to collect information on the comprehensibility, attractiveness, authenticity, and usability of websites for suicide prevention for men in crisis as well as their gatekeepers. Both websites present psychoeducational resources from experts and patients for both this (vulnerable) group at risk and for their support network. Overall, the findings of this study provide valuable insights into the design and usability of suicide prevention websites addressing men in crisis and their gatekeepers and may provide important implications for future digital prevention interventions.

Potential users of the websites evaluated the balanced mix of emotional content by persons at risk and informative content by scientists as adequate, credible, and helpful. This highlights the potential of this particular combination to address stigmatized topics such as suicidality and mental health and aligns with previous research suggesting that personal narratives combined with information may enhance mental health outcomes and decrease stigmatizing attitudes [31]. However, there were also challenges noted by interview participants, who indicated that they felt overwhelmed with video lengths and the multitude of options to navigate on the websites. Studies on video length have shown that videos with a length of more than 6 minutes can reduce interaction and understanding of learning significantly, thereby acting as a barrier to usage [32]. Based on cognitive load theory and taking into account the limited capacity of active working memory, they recommend using short videos that segment information into small blocks [33]. In addition, based on theories of cognitive capacity, studies on website design show how simplified navigation reduces barriers and prolongs usage [34].

Another key area identified was the need for additional communication channels to signpost to and promote the websites. Users expressed the importance of outreach to raise awareness about the websites and options for individualized and interpersonal communication after using the websites, especially in crisis situations. This aligns with research advocating for multiple communication channels in mental health interventions [35].

In sum, our study shows that both websites are perceived as trustworthy and reliable sources for male suicide prevention. Based on the perceived shortcomings of the programs and participants' recommendations, the study also provides important implications and improvement strategies for future interventions. This includes communicating brief and concise information as well as integrating options for more immediate emergency support. Overall, our findings underline the need for comprehensible, well-structured digital information which provides a balance between information richness and clarity. For both gatekeepers and the men at risk themselves, practical advice should be incorporated into communication interventions, and shorter content formats should be preferred, particularly considering video length. While digital interventions are undoubtedly crucial, our study also reveals the need to embed them in offline contexts and accompany them using other communication channels.

Limitations

Some limitations of the studies must be noted.

Due to the recruitment strategy used, we had a sample of very open, highly involved, and topic-aware participants who occasionally saw themselves as ambassadors and rated the social relevance of suicide prevention as very important. This means that emphasis on education and destigmatization related to this topic was already important for them. Thus, our results cannot be held to be transferable to people who may have internal barriers and fears about engaging with the topic. Furthermore, our cross-sectional studies focused on short-term effects and website evaluations. A longitudinal study would be needed to evaluate whether the websites are a helpful tool in the long run [36].

With regard to male gender and identification itself, we made broad generalizations, as it was our goal to create information platforms that were as universal and accessible as possible by as many men as possible. However, this also meant that specific challenges and burdens of people identifying as male and related subgroups (such as homosexual men, transgender men, and men from minority ethnic groups) were not specifically taken into account. Here it is pertinent that members of the LGBTQIA+ (lesbian, gay, bisexual, transgender, queer/questioning, intersex, asexual, and others) community are much more prone to suicidal thoughts and mental health problems than the population at large [37]. Developing prevention campaigns and information tools specifically tailored for them is thus of paramount importance. The same applies to other minorities, such as immigrants, non-traditional German-born individuals, older individuals, living alone, or those with wider mental health issues. Future research is needed that examines interventions for those vulnerable groups more closely and works with more diverse samples and communities.

XSL•FO
RenderX

Finally, our studies did not focus on creating an intervention directed at changing masculine tropes. This could be worthwhile, as studies suggest that social media campaigns can trigger conversations about masculinity [38,39], potentially questioning behavior driven by traditional masculine norms and value sets that aggravate the risk of (mental health) distress and negative experiences.

Conclusion

Both an RCT of the gatekeeper website [11] as well as this qualitative evaluation indicated that the program was positively evaluated by the participants, in the sense that it was seen as comprehensible and informative. The RCT [11] reported that the training of gatekeepers through a relatively short digital training program is effective based on the knowledge gained. In addition, the qualitative research demonstrates that both types of videos included in the program seem to be relevant for knowledge gain. While the videos featuring real individuals constitute important figures for personal identification and validation, supporting emotional health and self-esteem, the expert videos provide relevant and credible information. Such websites can, therefore, support gatekeepers in suicide prevention. However, it should not be overlooked that such interventions could be perceived to function as "tasking" the gatekeeper to manage the situation, thereby placing a specific burden on gatekeepers, who might feel even more responsible for the suicidal ideation and behavior.

Future studies in this area should also develop guidelines for population-specific communication with men at risk as well as gatekeepers.

Findings from brainstorming ideas beyond the tested interventions highlighted the need to integrate such low-barrier digital offers as one step in a strategy for more comprehensive gatekeeper support. Websites for men and gatekeepers can only fulfill one role in this strategy, with regard to national suicide prevention measures, and must be supplemented by other measures, such as a regional care network for mental health, architectural measures at high-risk sites such as overpasses, and helplines provision [4].

Acknowledgments

Funding for the study was provided by the research project "MEN-ACCESS: Suicide prevention for men" grant (95F12003D) from the "GKV Bündnis für Gesundheit" in Germany.

Data Availability

The datasets generated or analyzed during this study, that is, the transcripts of interviews and focus groups are not publicly available due to data protection reasons but specific portions of the data that are sufficiently anonymized may be obtained from the corresponding author on reasonable request.

Authors' Contributions

DR wrote the manuscript, acquired the funding, and administered the project and the methodology. AW reviewed and edited the manuscript. LH and MF conducted and analyzed the interviews. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Recruitment material.
[PDF File (Adobe PDF File), 165 KB - formative_v9i1e59829_app1.pdf]

Multimedia Appendix 2
Consent form.
[PDF File (Adobe PDF File), 564 KB - formative_v9i1e59829_app2.pdf]

Multimedia Appendix 3
Protocols for interviews and focus groups.
[PDF File (Adobe PDF File), 132 KB - formative_v9i1e59829_app3.pdf]

Multimedia Appendix 4
Guide focus groups.
[PDF File (Adobe PDF File), 288 KB - formative_v9i1e59829_app4.pdf]

Multimedia Appendix 5
Guide interviews.

XSL•FO
RenderX

[PDF File (Adobe PDF File), 252 KB - formative_v9i1e59829_app5.pdf]

References

1. Schelhase T. [Suicides in Germany: results from the official cause of death statistics]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2022 Jan;65(1):3-10 [FREE Full text] [doi: 10.1007/s00103-021-03470-2] [Medline: 34962579]
2. Hamilton E, Klimes-Dougan B. Gender differences in suicide prevention responses: implications for adolescents based on an illustrative review of the literature. Int J Environ Res Public Health 2015;12(3):2359-2372. [doi: 10.3390/ijerph120302359] [Medline: 25711358]
3. Krysinska K, Batterham PJ, Christensen H. Differences in the effectiveness of psychosocial interventions for suicidal ideation and behaviour in women and men: a systematic review of randomised controlled trials. Arch Suicide Res 2017;21(1):12-32. [doi: 10.1080/13811118.2016.1162246] [Medline: 26983580]
4. Struszczyk S, Galdas PM, Tiffin PA. Men and suicide prevention: a scoping review. J Ment Health 2019;28(1):80-88. [doi: 10.1080/09638237.2017.1370638] [Medline: 28871841]
5. Till B, Niederkrotenthaler T. Surfing for suicide methods and help: content analysis of websites retrieved with search engines in austria and the United States. J Clin Psychiatry 2014;75(8):886-892. [doi: 10.4088/JCP.13m08861] [Medline: 25099284]
6. Franco-Martín MA, Muñoz-Sánchez JL, Sainz-de-Abajo B, Castillo-Sánchez G, Hamrioui S, de la Torre-Díez I. A systematic literature review of technologies for suicidal behavior prevention. J Med Syst 2018;42(4):71. [doi: 10.1007/s10916-018-0926-5] [Medline: 29508152]
7. Balcombe L, De Leo D. The potential impact of adjunct digital tools and technology to help distressed and suicidal men: an integrative review. Front Psychol 2021;12:796371 [FREE Full text] [doi: 10.3389/fpsyg.2021.796371] [Medline: 35058855]
8. Pirkis J, Rossetto A, Nicholas A, Ftanou M, Robinson J, Reavley N. Suicide prevention media campaigns: a systematic literature review. Health Commun 2019;34(4):402-414. [doi: 10.1080/10410236.2017.1405484] [Medline: 29190128]
9. Wagner A, Reifegerste D. Desirable and adverse effects of communicative suicide prevention interventions among men. Crisis 2024;45(5):365-375. [doi: 10.1027/0227-5910/a000965] [Medline: 39136159]
10. Yonemoto N, Kawashima Y, Endo K, Yamada M. Gatekeeper training for suicidal behaviors: a systematic review. J Affect Disord 2019;246:506-514. [doi: 10.1016/j.jad.2018.12.052] [Medline: 30599375]
11. Hofmann L, Wagner B. Efficacy of an online gatekeeper program for relatives of men at risk of suicide - a randomized controlled trial. BMC Public Health 2024;24(1):2693 [FREE Full text] [doi: 10.1186/s12889-024-20193-6] [Medline: 39358752]
12. Vogel DL, Wade NG, Wester SR, Larson L, Hackler AH. Seeking help from a mental health professional: the influence of one's social network. J Clin Psychol 2007;63(3):233-245. [doi: 10.1002/jclp.20345] [Medline: 17211874]
13. Hofmann L, Wagner B. Men's behavior and communication in the days prior to a suicide-a psychological autopsy study. Int J Environ Res Public Health 2023;20(17):6668 [FREE Full text] [doi: 10.3390/ijerph20176668] [Medline: 37681808]
14. Spahn C, Kranz S, Teismann T, Forkmann T, Schönfelder A, Rath D, et al. Gender-specific aspects of suicide-related communication in a high risk sample of psychiatric inpatients. JOMH 2022;18(9):188 [FREE Full text] [doi: 10.31083/j.jomh1809188]
15. Zalsman G, Hawton K, Wasserman D, van Heeringen K, Arensman E, Sarchiapone M, et al. Suicide prevention strategies revisited: 10-year systematic review. Lancet Psychiatry 2016;3(7):646-659. [doi: 10.1016/S2215-0366(16)30030-X] [Medline: 27289303]
16. Gould MS, Cross W, Pisani AR, Munfakh JL, Kleinman M. Impact of applied suicide intervention skills training on the national suicide prevention lifeline. Suicide Life Threat Behav 2013;43(6):676-691 [FREE Full text] [doi: 10.1111/sltb.12049] [Medline: 23889494]
17. Ghoncheh R, Gould MS, Twisk JW, Kerkhof AJ, Koot HM. Efficacy of adolescent suicide prevention e-learning modules for gatekeepers: a randomized controlled trial. JMIR Ment Health 2016;3(1):e8 [FREE Full text] [doi: 10.2196/mental.4614] [Medline: 26825006]
18. Bartgis J, Albright G. Online role-play simulations with emotionally responsive avatars for the early detection of native youth psychological distress, including depression and suicidal ideation. Am Indian Alsk Native Ment Health Res 2016;23(2):1-27. [doi: 10.5820/aian.2302.2016.1] [Medline: 27115130]
19. Lamis DA, Underwood M, D'Amore N. Outcomes of a suicide prevention gatekeeper training program among school personnel. Crisis 2017;38(2):89-99. [doi: 10.1027/0227-5910/a000414] [Medline: 27561223]
20. Mann JJ, Michel CA, Auerbach RP. Improving suicide prevention through evidence-based strategies: a systematic review. Am J Psychiatry 2021;178(7):611-624 [FREE Full text] [doi: 10.1176/appi.ajp.2020.20060864] [Medline: 33596680]
21. Silva AG, Caravau H, Martins A, Almeida AMP, Silva T, Ribeiro Ó, et al. Procedures of user-centered usability assessment for digital solutions: scoping review of reviews reporting on digital solutions relevant for older adults. JMIR Hum Factors 2021;8(1):e22774 [FREE Full text] [doi: 10.2196/22774] [Medline: 33439128]

XSL•FO
RenderX

22. Hofmann L, Glaesmer H, Wagner B, Spahn C. Männer in suizidalen krisen: [Men in suicidal crises: Approaches to gender-specific suicide prevention and psychosocial care]. Psychotherapeutenjournal 2023:271-278.
23. Jenkins EL, Ilicic J, Barklamb AM, McCaffrey TA. Assessing the credibility and authenticity of social media content for applications in health communication: scoping review. J Med Internet Res 2020;22(7):e17296 [FREE Full text] [doi: 10.2196/17296] [Medline: 32706675]
24. Kim Y. Trust in health information websites: a systematic literature review on the antecedents of trust. Health Informatics J 2016;22(2):355-369 [FREE Full text] [doi: 10.1177/1460458214559432] [Medline: 25518944]
25. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022;292:114523 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
26. Evans SC, Roberts MC, Keeley JW, Blossom JB, Amaro CM, Garcia AM, et al. Vignette methodologies for studying clinicians' decision-making: validity, utility, and application in ICD-11 field studies. Int J Clin Health Psychol 2015;15(2):160-170 [FREE Full text] [doi: 10.1016/j.ijchp.2014.12.001] [Medline: 30487833]
27. Evans N, Whitcombe S. Using circular questions as a tool in qualitative research. Nurse Res 2016;23(3):26-29. [doi: 10.7748/nr.23.3.26.s6] [Medline: 26793984]
28. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. J Med Internet Res 2019;21(5):e13117 [FREE Full text] [doi: 10.2196/13117] [Medline: 31115340]
29. Braun V, Clarke V. Toward good practice in thematic analysis: avoiding common problems and be(com)ing a researcher. Int J Transgend Health 2023;24(1):1-6 [FREE Full text] [doi: 10.1080/26895269.2022.2129597] [Medline: 36713144]
30. Boyatzis RE. Transforming Qualitative Information: Thematic Analysis and Code Development. Thousand Oaks: Sage Publications, Inc; 1998.
31. Rennick-Egglestone S, Morgan K, Llewellyn-Beardsley J, Ramsay A, McGranahan R, Gillard S, et al. Mental health recovery narratives and their impact on recipients: systematic review and narrative synthesis. Can J Psychiatry 2019;64(10):669-679 [FREE Full text] [doi: 10.1177/0706743719846108] [Medline: 31046432]
32. Afify MK. Effect of interactive video length within e-learning environments on cognitive load, cognitive achievement and retention of learning. Turkish Online Journal of Distance Education 2020:68-89. [doi: 10.17718/tojde.803360]
33. Slemmons K, Anyanwu K, Hames J, Grabski D, Mlsna J, Simkins E, et al. The impact of video length on learning in a middle-level flipped science setting: implications for diversity inclusion. J Sci Educ Technol 2018;27(5):469-479. [doi: 10.1007/s10956-018-9736-2]
34. Rotondi AJ, Spring MR, Hanusa BH, Eack SM, Haas GL. Designing eHealth applications to reduce cognitive effort for persons with severe mental illness: page complexity, navigation simplicity, and comprehensibility. JMIR Hum Factors 2017;4(1):e1 [FREE Full text] [doi: 10.2196/humanfactors.6221] [Medline: 28057610]
35. Sampogna G, Bakolis I, Evans-Lacko S, Robinson E, Thornicroft G, Henderson C. The impact of social marketing campaigns on reducing mental health stigma: results from the 2009-2014 time to change programme. Eur Psychiatry 2017;40:116-122 [FREE Full text] [doi: 10.1016/j.eurpsy.2016.08.008] [Medline: 27997875]
36. Holmes G, Clacy A, Hermens DF, Lagopoulos J. The long-term efficacy of suicide prevention gatekeeper training: a systematic review. Arch Suicide Res 2021;25(2):177-207. [doi: 10.1080/13811118.2019.1690608] [Medline: 31809659]
37. Haas AP, Eliason M, Mays VM, Mathy RM, Cochran SD, D'Augelli AR, et al. Suicide and suicide risk in lesbian, gay, bisexual, and transgender populations: review and recommendations. J Homosex 2011;58(1):10-51 [FREE Full text] [doi: 10.1080/00918369.2011.534038] [Medline: 21213174]
38. Scotti Requena S, Pirkis J, Currier D, Conway M, Lee S, Turnure J, et al. An evaluation of the boys do cry suicide prevention media campaign on twitter: mixed methods approach. JMIR Form Res 2023;7:e49325 [FREE Full text] [doi: 10.2196/49325] [Medline: 37676723]
39. Schlichthorst M. Engaging men in conversations about masculinity and suicide? An evaluation of the Man Up social media campaign. 2020. URL: https://minerva-access.unimelb.edu.au/bitstream/handle/11343/265962/316ab37c-af07-eb11-94cb-0050568d7800_mphil%20thesis-mschlichthorst-final.pdf [accessed 2025-01-23]

Abbreviations

LGBTQIA+: lesbian, gay, bisexual, transgender, queer/questioning, intersex, asexual, and others
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by A Mavragani; submitted 23.04.24; peer-reviewed by J Arroyave-Cabrera, V Carli, P Stas; comments to author 16.10.24; revised version received 25.11.24; accepted 21.12.24; published 26.02.25.

Please cite as:
Reifegerste D, Wagner AJM, Huber L, Fastuca M
Formative Evaluation of Suicide Prevention Websites for Men: Qualitative Study with Men at Risk of Suicide and with Potential Gatekeepers
JMIR Form Res 2025;9:e59829
URL: https://formative.jmir.org/2025/1/e59829
doi:10.2196/59829
PMID:

©Doreen Reifegerste, Anna J M Wagner, Lisa Huber, Manuel Fastuca. Originally published in JMIR Formative Research (https://formative.jmir.org), 26.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Automated Digital Safety Planning Interventions for Young Adults: Qualitative Study Using Online Co-design Methods

Jonah Meyerhoff[1], PhD; Sarah A Popowski[1], BA; Tanvi Lakhtakia[1], BA; Emily Tack[1], BA; Rachel Kornfield[1], PhD; Kaylee P Kruzan[1], LCSW, PhD; Charles J Krause[1], MPH; Theresa Nguyen[2], LCSW; Kevin Rushton[2], BA; Anthony R Pisani[3,4], PhD; Madhu Reddy[5], PhD; Kimberly A Van Orden[3], PhD; David C Mohr[1], PhD

[1]Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[2]Mental Health America, Alexandria, VA, United States

[3]Department of Psychiatry, School of Medicine and Dentistry, University of Rochester, Rochester, NY, United States

[4]Department of Pediatrics, School of Medicine and Dentistry, University of Rochester, Rochester, NY, United States

[5]Department of Informatics, Donald Bren School of Information and Computer Sciences, University of California, Irvine, CA, United States

Corresponding Author:
Jonah Meyerhoff, PhD
Department of Preventive Medicine
Feinberg School of Medicine
Northwestern University
750 North Lake Shore Drive
Chicago, IL, 60611
United States
Phone: 1 3125031232
Email: jonah.meyerhoff@northwestern.edu

Abstract

Background: Young adults in the United States are experiencing accelerating rates of suicidal thoughts and behaviors but have the lowest rates of formal mental health care. Digital suicide prevention interventions have the potential to increase access to suicide prevention care by circumventing attitudinal and structural barriers that prevent access to formal mental health care. These tools should be designed in collaboration with young adults who have lived experience of suicide-related thoughts and behaviors to optimize acceptability and use.

Objective: This study aims to identify the needs, preferences, and features for an automated SMS text messaging–based safety planning service to support the self-management of suicide-related thoughts and behaviors among young adults.

Methods: We enrolled 30 young adults (age 18-24 years) with recent suicide-related thoughts and behaviors to participate in asynchronous remote focus groups via an online private forum. Participants responded to researcher-posted prompts and were encouraged to reply to fellow participants—creating a threaded digital conversation. Researcher-posted prompts centered on participants' experiences with suicide-related thought and behavior-related coping, safety planning, and technologies for suicide-related thought and behavior self-management. Focus group transcripts were analyzed using thematic analysis to extract key needs, preferences, and feature considerations for an automated SMS text messaging–based safety planning tool.

Results: Young adult participants indicated that an automated digital SMS text message–based safety planning intervention must meet their needs in 2 ways. First, by empowering them to manage their symptoms on their own and support acquiring and using effective coping skills. Second, by leveraging young adults' existing social connections. Young adult participants also shared 3 key technological needs of an automated intervention: (1) transparency about how the intervention functions, the kinds of actions it does and does not take, the limits of confidentiality, and the role of human oversight within the program; (2) strong privacy practices—data security around how content within the intervention and how private data created by the intervention would be maintained and used was extremely important to young adult participants given the sensitive nature of suicide-related data; and (3) usability, convenience, and accessibility were particularly important to participants—this includes having an approachable and engaging message tone, customizable message delivery options (eg, length, number, content focus), and straightforward menu navigation. Young adult participants also highlighted specific features that could support core coping skill acquisition (eg, self-tracking, coping skill idea generation, reminders).

XSL•FO
RenderX

Conclusions: Engaging young adults in the design process of a digital suicide prevention tool revealed critical considerations that must be addressed if the tool is to effectively expand access to evidence-based care to reach young people at risk for suicide-related thoughts and behaviors. Specifically, automated digital safety planning interventions must support building skillfulness to cope effectively with suicidal crises, deepening interpersonal connections, system transparency, and data privacy.

(JMIR Form Res 2025;9:e69602) doi:10.2196/69602

KEYWORDS

mental health services; technology; therapy; computer assisted; SMS text messaging

Introduction

Background

The prevalence of suicidal thoughts among young adults, those between the ages of 18 and 25 years, in the United States has increased by approximately 160% since 2013 [1,2]. While most young adults who experience suicidal thoughts do not go on to make a suicide attempt, suicidal thoughts are indicators of risk that warrant attention and intervention [3]. Among adults who experienced mental health symptoms in the last year, young adults have the lowest rates of formal mental health care treatment [2]. Structural (eg, cost, accessibility, and provider availability) and attitudinal (eg, stigma, strong preferences for self-management) barriers are particularly important for determining help-seeking behavior [4] and contribute to low rates of mental health treatment engagement [4-6].

Current approaches to suicide prevention underserve young adults facing structural or attitudinal barriers to mental health services. Most prevention efforts are delivered in the context of mental health interventions via formal outpatient, inpatient, emergency department, and other clinical care settings. Prevention is also delivered in informal settings such as crisis helplines, targeting individuals before suicidal crises arise (eg, in schools or workplaces) in an effort to bring suicide prevention approaches to those who need it in places they are most likely to access it [7]. As most young adults with suicide-related thoughts and behaviors do not engage in formal mental health treatment, where most evidence-based suicide prevention is delivered [8-15], and they experience strong preferences for self-management, young adults are less likely to encounter evidence-based suicide prevention interventions.

The high need for preventive interventions and low formal mental health treatment engagement among young adults experiencing suicidal thoughts and behaviors [1,2] signal a critical gap in suicide prevention infrastructure. Digital suicide prevention interventions are an important avenue by which access to evidence-based suicide prevention care can be increased. Digital suicide prevention interventions offer the benefit of being able to circumvent many of the attitudinal and structural barriers that prevent young adults from accessing evidence-based preventive interventions [4-6]. They can be accessed on young adults' devices (eg, smartphones) and used privately.

The safety planning intervention (SPI) [16], a brief intervention aimed at interrupting the progression from suicidal ideation to suicidal behavior by increasing coping ability, has typically been administered collaboratively by clinicians in clinical care settings across ages [17-19] and is considered standard of care [20,21]. The SPI typically includes 6 areas of focus: warning signs, coping skills, people for distraction, people for help, professional organizations and resources, and environmental safety [16]. The product of the SPI is often a brief document with personalized lists of resources and coping activities for each of the 6 focus areas. When delivered with fidelity, clinicians ask patients targeted questions and follow-up on patients' responses to elicit highly personalized information that can enhance the efficacy of an individual's safety plan by making it as relevant as possible to the individual's life circumstances [22]. However, safety planning is frequently delivered by individuals facing barriers to implementation including insufficient training and time [23,24], which poses a risk to safety plan quality and efficacy.

The SPI is one of the most common digital suicide prevention intervention elements [25]. There is an emerging literature on automated digital SPIs that guide users through the creation of a highly personalized safety plan [25]. Several have been shown to be acceptable [26-29] and even demonstrate that resultant safety plan quality can be comparable to collaboratively developed safety plans [30,31]. Extant literature on digital SPIs primarily focuses on broad adult samples and narrow adolescent populations [25]. Boudreaux et al [26] piloted an automated, computer-based SPI that enabled adults who experienced suicide-related thoughts and behaviors to complete a safety plan and offered brief guidance on how to complete the safety plan. The resultant safety plan was a 1-page PDF that users could download, access on different devices, and share with trusted others. Melvin et al [27] and Rainbow et al [29] developed a mobile app–based approach called BeyondNow that includes a customizable SPI with prepopulated suggestions for each of the 6 safety planning steps and offers the ability to edit or share the completed safety plan and a quick access button to a crisis helpline. The app is publicly accessible and has strong acceptability metrics and is associated with promising improvements in suicide-related coping. Finally, Methi et al [28] created an automated web-based e-SPI specifically designed for adolescents that includes age-appropriate instructions, examples, and illustrations that assist users in completing the online safety plan via text-fill boxes. The content adolescents add is converted to an online safety card that users can save. These efforts highlight that automated safety planning is feasible to implement and that there are opportunities to engage individuals in new ways, expanding beyond single-session safety planning and crafting safety plans via mediums that are more conversational in nature.

XSL•FO
RenderX

Among this emerging literature, several key needs and preferences have surfaced. First, having quick access to information such as one's documented safety plan, crisis resources, or personal contact information can be extremely valuable during a crisis as crisis states make it difficult to navigate complex user interfaces [32,33]. Second, there is a favorable view of digital artifacts (eg, photos, songs) that inspire hope and increase motivation for living [27]. Third, across the literature, users of digital SPIs have ambivalence about whether or how they want to involve others in their personal networks in the creation or use of their safety plans [32,33]. One way early digital SPIs have addressed this issue is through the use of optional prescripted messages users can generate when outside of crisis states and send to trusted others to elicit social or instrumental support [32]. Fourth, extant literature highlights that strong privacy practices emerge as critical priorities for users of digital safety plans [33,34]. Fifth, the ability to download and edit created safety plans after their initial creation is critical for ensuring safety plans remain relevant and accessible over time [27,34,35]. Finally, using validating language that minimizes assumptions about a user (eg, that they have a network of others that can provide support during a crisis) and ensuring the digital SPI can provide examples or menus of items if a user has difficulty generating safety planning content independently can prove beneficial [35].

The vast majority of digital SPIs are mobile app or browser based [25,36], meaning that individuals must visit a specific app or website in order to interact with or create their safety plan. In this study, we aimed to understand the potential of an SMS text messaging–based safety plan creation approach. Texting has the technological affordance of being a conversational medium, potentially addressing some of the shortcomings of automated digital safety plans that are documentation only. By eliciting information relevant to safety planning in a conversational manner, an automated texting–based SPI may be able to increase the personalization and relevance of digital safety plans. SMS text messaging is the most commonly used communication medium and young adults are the age group sending the greatest number of SMS text messages [37,38]. It is more widely used than social media apps [39]. Therefore, an SMS text messaging–based SPI is capable of meeting young adults in a space where they are already spending a significant amount of time. Indeed, prior work has shown that texting interventions for suicide prevention have been acceptable to users [40-45].

The relatively sparse existing literature on the needs and preferences of digital SPI users is focused on the app- or web-based designs. Further, it lacks focus on the needs of the adult age group with the highest rates of suicide-related thoughts and behaviors, but the lowest rates of treatment-seeking: young adults. Moreover, there is a paucity of literature regarding automated, interactive, texting interventions to support suicide prevention, broadly, and safety planning, more specifically.

Objective

In this study, we aimed to understand the needs, preferences, constraints, and design considerations for an SMS text message–based automated safety planning tool for young adults who have lived experience of suicide-related thoughts and behaviors. Informed by prior literature, our work was guided by the following research questions: (1) What are the broad needs and preferences of young adults who experience suicide-related thoughts and behaviors for suicide prevention? (2) What are the features that can support young adults with suicide-related thoughts and behaviors to create an effective safety plan?

Methods

Participants

A total of 30 participants between the ages of 18 and 24 years were enrolled in this study (Table 1). Individuals were eligible to participate if they resided in the United States, were between the ages of 18 and 24 years, owned a smartphone, were proficient in English, and endorsed past 2-week suicidal ideation. Suicidal ideation was assessed using a modified item from the DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition) Cross-Cutting Symptom Inventory (DSM-5-CCSI) [46], "During the past 2 weeks, how often have you been bothered by thoughts of actually hurting yourself [or ending your life]?" Disambiguation of suicidal thoughts from nonsuicidal self-injurious thoughts was assessed via the Depression Severity Index—Suicidality Subscale (DSI-SS) [47], a 4-item self-report measure assessing the frequency and intensity of suicidal thinking. Participants were not eligible to participate if they endorsed symptoms consistent with mania or psychosis in the last 2 weeks (assessed via the DSM-5-CCSI) or the presence of strong suicidal desire assessed via item 9 on the Beck Depression Inventory—Second Edition [48] and a subsequent score of 2 or 3 (range 0-3).

XSL•FO
RenderX

Table 1. Participant characteristics.

Characteristics	Values
Age, mean (SD)	21.1 (1.9)
Gender, n (%)	
Male	3 (10)
Female	22 (73)
Nonbinary	5 (17)
Transgender, n (%)	4 (13)
Race, n (%)	
Black/African American	3 (10)
American Indian/Alaska Native	1 (3)
Asian	5 (17)
Native Hawaiian/Pacific Islander	0 (0)
White	18 (60)
More than one race	2 (7)
Prefer not to answer	1 (3)
Hispanic/Latino, n (%)	3 (10)
Total DSI-SS[a] score (4 items, with each coded between 0 and 3; range 0-12), mean (SD)	4.8 (1.6)
Ever seen a therapist or mental health professional, n (%)	25 (83)
Currently seeing a therapist or mental health professional, n (%)	11 (37)
Plan to see a therapist or mental health professional in the next 8 weeks, n (%)	17 (57)
Never seen a therapist or mental health professional and does not plan to see one in the next 8 weeks, n (%)	1 (3)

[a]DSI-SS: Depression Severity Index—Suicidality Subscale.

Procedures

Individuals were recruited via an advertisement embedded on Mental Health America's (MHA) website. The advertisement was shown to individuals who visited a resource page about suicidal thoughts and behaviors, endorsed suicidal or self-harm ideation in the past 2 weeks on MHA's online screening website, or endorsed past-month suicidal ideation on MHA's nonsuicidal self-injury survey. Individuals who clicked the ad were directed to a Northwestern University–hosted informational web page about the study. Interested individuals could elect to provide informed consent for eligibility screening via Northwestern University's instance of REDCap [49,50] and were subsequently routed to the eligibility screening surveys.

If eligible, interested individuals could review the REDCap-based study consent documentation and the code of conduct for engaging with research staff and fellow participants on the private research forum. The code of conduct detailed guidelines such as refraining from posting identifying information or detailed descriptions of self-harm, suicide methods, or behaviors. The code of conduct also overviewed harassment and moderation policies. To be enrolled in the study, individuals needed to provide affirmative informed consent and agree to abide by the code of conduct.

Participants engaged in 1 of 2 asynchronous remote community (ARC) [51-53] groups—akin to an online focus group. We used ARC to increase the engagement of underrepresented populations by decreasing barriers related to synchronous or in-laboratory research tasks [54]. Additionally, the group format was chosen to facilitate conversation between adults with shared experiences. ARC methods are used in the field of human-computer interaction to facilitate conversation, elicit several perspectives simultaneously, and provide the environment for participants to build on each other's creative contributions [51]. ARC groups were held between November and December 2022, or February and March 2023. Two ARC groups were held to increase the sample size and diversity of responses to several lines of questioning and protect against group effects as well as to gather data on lines of inquiry that were not covered in the first group. ARC groups involved participants registering for a private online forum using FocusGroupIt [55] in which researchers and participants could create posts visible for all in the study and with which participants could engage or respond. Each group comprised 15 participants. Past work demonstrates similar sample sizes are sufficient to elicit diverse perspectives [56]. Participants were provided with automatically generated anonymous usernames (eg, participant 1, participant 2). Each ARC group consisted of 8 total prompts, 1 every 3 days. Prompts focused on terminology young adults used for suicidal ideation, prompting events for suicide-related thoughts and behaviors, safety plan use, and technology to support both safety planning and help-seeking.

XSL•FO
RenderX

Research staff monitored forums daily for posts violating the posted code of conduct. Offending posts were deleted, and the original poster was contacted by research staff with an explanation of why the post was removed, and, if needed, to conduct a risk assessment. Only 1 such post was removed and there was no indication of suicide-risk elevation for the poster.

Data Analysis

The corpus for this study included 3 prompts from each group containing relevant data to answering the primary research questions (see Multimedia Appendix 1). These prompts comprising the corpus covered perceptions of safety planning via SMS text messages (group 1, prompt 4; group 2, prompt 3), the role of automated messages in supporting safety planning (group 1, prompt 5; group 2, prompt 4), use of a safety plan after it is created (group 1, prompt 7), and the role of crisis services in suicide prevention care (group 2, prompt 7).

Forum transcripts were analyzed using Thematic Analysis [57,58] to extract key needs, preferences, and design considerations for a digital safety planning tool. Thematic analysis was facilitated using Dedoose software (Version 9.0.9) [59]. One coder (JM) read through the corpus and generated an initial codebook via open coding responses to each of the 3 prompts from each ARC group. Coding occurred at the sentence level within a participant's post. Two additional coders (SAP and TL) read through the entire corpus to familiarize themselves with the data and then applied the initial codebook to overlapping transcripts. All 3 coders then met to revise the initial codebook. Discrepancies were discussed until a consensus was reached among all coders. Importantly, the analysis software, Dedoose, facilitates the attribution of code applications to the research staff who coded the data, enabling the ability to identify and resolve coding discrepancies. All coders then divided the remaining data and applied the revised codebook so that the entire corpus was coded by at least two coders. Coders met 3 more times to discuss and resolve discrepancies. The saturation of codes was determined to be reached once all discrepancies in codes were resolved and no additional codes were identified in the overlapping transcripts. Coders then met to organize codes into axial themes iteratively.

Positionality Statement

At the time of this study, JM was a clinical psychologist and faculty member with a background in both clinical and research work in suicide prevention. He held funded grants focused on increasing access to care through the design, development, and evaluation of digital mental health interventions. SAP was a research assistant who supported the administration of the ARC methods described in this paper. Both JM and SAP interacted with participants whose data comprise the study corpus through study procedures such as email and, in some cases, interviews. To maximize objectivity during coding and analysis, all participants were assigned a coded ID number. TL was a research assistant external to the project and had no prior contact with any study participants.

Ethical Considerations

Before beginning any screening or study procedures, all interested individuals and participants in this study provided affirmative informed consent, both for initial eligibility screening and subsequently for participation in the study. Participants were compensated up to US $80 for responding to the study prompts and for commenting on other participants' posts.

Throughout the study, participants' data were, and continue to be, considered confidential and protected. Only research personnel and individuals securing appropriate credentials and data use agreements were able to access study data and complete transcripts. Study transcripts were deidentified and maintained on secure encrypted servers at Northwestern University.

All study procedures were administered at Northwestern University and all study procedures were approved by the Northwestern University Institutional Review Board (project number STU00217191).

Results

Young Adults' Perspectives on Automated Digital SPIs: Treatment Mechanisms, Core Elements, and Skill-Building Features

Across participants in our ARC groups, young adults expressed their need for automated digital SPIs. Young adults discussed (1) the treatment mechanisms an automated digital SPI must address, (2) core elements of the technology that affect the perceived usefulness and acceptability of an automated digital SPI, and (3) key skill-building features they would like to see in such an intervention.

Meeting Young Adults' Needs

Key Needs for SPI Support: Self-Management and Social Connection

We identified 2 key needs that young adults thought an SPI should support. The first need was for self-management and the second was for social connection.

Empowering Young Adults to Manage Their Own Symptoms by Supporting Coping Skill Acquisition and Use

Young adults wanted an automated digital SPI to support the self-management of their suicide-related thoughts and behaviors through coping skills that foster a sense of self-empowerment. Self-empowerment approaches enable young adults to rely on their own skillfulness to effectively cope with their suicide-related thoughts and behaviors, and to avoid the stigma they associated with help-seeking. Some participants already possessed coping skills that facilitated self-empowerment and shared how these coping skills allowed them to navigate future suicidal crises:

> *I have never thought about using one of the crisis services, as I never felt as though I have needed to use one. I think it would be a bit embarrassing for me, as I strongly dislike talking about depression and/or suicide with others. I have also always been able to talk myself down without the need for a crisis service intervention.* [G2-P2]

XSL•FO
RenderX

G2-P2's statement suggests that through practice in talking themselves down, they have been able to mitigate escalating suicidal crises and this experience provides confidence in their own ability. While participants, in general, noted the potential value of a tool to support coping skill use and acquisition, others such as G2-P3 highlighted the potential benefits of an approach that involved other people:

> *I ultimately chose not to use these [crisis helpline] services because I was able to eventually re-ground myself, however, in retrospect, I think it would have been more beneficial to talk to someone about this in the moment rather than going through it alone.*

Young adult participants described the impact they believed an automated SMS text messaging–based safety planning tool could have on their self-empowerment in moments of need. G1-P4 shared that by creating a safety plan with the assistance of a digital tool, an individual is taking the lead in planning for their own coping. They described that one of the "benefits would be reassurance of the feeling of not being alone from something you created." G1-P15 also highlighted the benefits of a self-created safety plan, commenting, "the fact that the individual will be able to change it, and have the power over their safety would probably make a lot of people who use the plan a lot more secure." Overall, young adult participants were clear that an automated safety planning service would need to be able to provide them with the ability to choose how they cope with their symptoms.

Young adult participants shared that one of the motivations for using self-management tools was prior experiences of invalidation when discussing their suicide-related thoughts and behaviors with others. As G2-P12 recalled,

> *After reaching out to a hotline, I had to speak with a campus therapist and she laughed at me for considering suicide because my problems seemed "frivolous" or "overexaggerated". I haven't been able to trust a hotline since, and after that one incident, I've cut myself off to dozens of resources that may help me with the fear they may joke about me in the same manner.*

This experience of stigma, wherein the participant expected help and was instead met with judgment-worsened symptoms sowed distrust in systems and made future help-seeking less appealing, ultimately leading to decreased support and increased social isolation. As a result, young adults noted that it was important to have resources that enabled them to self-manage their symptoms and that were offered in a welcoming and nonjudgmental tone, mitigating the need to disclose suicide-related thoughts and behaviors to others to obtain support. For example, G1-P13 shared how messages could support effective coping, but would need to strike the right tone "I think automated messages like this would be extremely helpful in providing a structured way to create a safety plan. [...] these messages should be welcoming, even if they aren't actually run by a person, and include greetings or transitions."

Participants described how an automated digital SPI delivered via SMS text messaging could foster self-empowerment by providing actionable coping resources in an easily accessible format. G2-P14 shared, "[the intervention] would need to be automated, where it texts my questions and I respond and it can provide me with the appropriate part of my personalized safety plan." G2-P14's comment depicts a scenario in which a user of an automated safety planning service is asked targeted questions that elicit personalized coping resources and then these personalized coping resources are sent back to the user at opportune times, facilitating self-management of suicidal crises and self-efficacy by enabling an individual to use coping resources effectively. This approach was echoed by others, for example, G1-P8 noted "Having the automated messages send me reminders for self-care and other things based on what I write in response to the messages (plus what I already have in my safety plan) would be great as well."

Some participants shared how a digital SMS text messaging–based approach must accommodate change over time to fully support self-management of suicide-related thoughts and behaviors. As an individual's circumstances, resources, and skillfulness change over time, what works for them at the outset will need to be updated to be more relevant and effective. G1-P6 shared that, "So much can change in just a second, and what worked for me yesterday might not work for me today." Editing and maintaining one's safety plan was seen as a critical component of self-empowerment in that it enables a constant cycle of being able to evaluate a safety plan's efficacy against real-world suicide-related thought and behavior and crisis mitigation and management attempts. The ability to edit and maintain a safety plan allows for modification and adaptation in the event of actual or anticipated underperformance so that it can perform better in the future.

Drawing on Social Connections

The second core element many young adults in our study wanted was an automated digital SPI to support self-management of suicide-related thoughts and behaviors by supporting users in connecting with others. Participants had different ideas about how this could be accomplished, but the vast majority agreed that interpersonal connections were critical for any suicide-related thought and behavior intervention. The findings that follow highlight how components that support interpersonal connection may be critical targets in an automated intervention aimed at addressing suicide-related thoughts and behaviors, and could potentially work by creating accountability structures, creating openness to connection with others, fostering a feeling of being cared for, eliciting outside perspectives, and recruiting help when needed.

Creating social accountability was a core function participants said was important to consider for the self-management of suicide-related thoughts and behaviors. Participants noted that it was difficult to commit to using certain coping skills when in a suicidal crisis, but accountability to others would make it more likely for them to use the planned coping skills. One participant highlighted the utility of others to maintain accountability structures:

> *I do feel like a lot of safety plans are made with that individual having to keep themselves accountable in mind. But, part of a safety plan can include other*

XSL•FO
RenderX

> *people/ resources that can help in those times and in some ways keep you accountable and safe.* [G2-P16]

As G2-P16 articulates, the system could encourage the engagement of trusted others in the digital intervention to enhance the efficacy of digital safety planning.

Participants also wanted a digital intervention to facilitate connections with trusted others in their network. G2-P7 put it clearly, "I need human interaction more than anything right now." This participant's perspective highlights that many young adults are aware of the need for connection, especially in the context of crisis states, and that it is necessary to, first, build and, second, reach out to those support networks when needed. G2-P3 stated, "when experiencing feelings of loneliness/isolation alongside suicidal thoughts, it is especially beneficial to have social support mechanisms in place." Young adults were open to an automated digital SPI but emphasized that it needed to also facilitate connection and strengthen ties to trusted others.

Participants imagined that an automated intervention could connect individuals to their support networks and allow young adults to get outside perspectives from mental health professionals or trusted others. When experiencing suicidal thoughts, young adults shared that it is challenging to access perspectives outside of their own, but acknowledged that it can be helpful for diverting or avoiding an escalating suicidal crisis. As a result, participants raised the importance of obtaining outside perspectives:

> *Sometimes it is really hard to notice signs of crisis in yourself, and having something point it out to you could be really helpful. I've noticed in myself that sometimes when I get really worked up I don't notice how dangerous my behavior has become until a loved one points it out to me and expressed their concerns. It's helpful to get an outside perspective.* [G1-P6]

While many participants wanted other people involved in the perspective-taking process, some did not. G2-P7 shared how experiences with past treatment-seeking cast doubt on the utility of involving others to offer outside perspectives, "The best therapy, in my opinion, is taking time to yourself and getting to the root of the issue alone. It was the best thing I could have ever done for myself." However, participants discussed that an automated system should nonetheless prompt an individual to consider different ways of looking at one's circumstances, experiences, and values providing an outside perspective that could serve a similar function without integrating a human in the loop by default. As G2-P3 states, "these messages could serve as gentle reminders of values and the things that are important to people, [...] I would be more likely to respond to these messages if they contained thought-provoking questions that probed my values."

There was a minority of participants who shared that a potential drawback to an automated SMS text messaging–based SPI is that it had the potential to maintain or exacerbate social isolation by reducing a point of potential human contact such as calling or texting with a crisis counselor. For example, G2-P2 shared, "even if it designed for you, the digital aspect of the text is a bit isolating." An SMS text messaging suicide-related thought and behavior intervention may be acceptable for many, but participants cautioned that if the interpersonal considerations were not sufficiently designed for, this type of intervention could cause harm. Specifically noting that automated suicide prevention interventions could be misused in place of a point of human contact (eg, crisis counselor support), potentially worsening subjective experiences of isolation.

Technology Needs

Technological Needs for Automated Digital SPIs: Transparency, Privacy, and Usability

The young adult participants in our study also discussed needs related to the technological aspects of an automated digital SPI. Specifically, they highlighted the importance of (1) transparency or clarity regarding expectations, assumptions, and limitations of automated digital SPI; (2) privacy or the practices, policies, and technologies that limit exposure of personal information; and (3) usability, convenience, and accessibility of the intervention.

The Need for Transparency

Participants detailed the important ways in which being transparent in the context of an intervention for suicide-related thoughts and behaviors is critical to uptake and efficacy. One participant shared how transparency about crisis helpline practices helped them feel more comfortable using a particular crisis helpline:

> *Even just having some ideas of what may be asked of my or what to expect, vocabulary to express my feelings, etc. may help make the experience of asking for help much less scary and more accessible.* [G2-P13]

G2-P13 highlighted how, in the context of crisis helplines, having information about the scripts and procedures counselors follow before making the call can help put callers at ease by facilitating mental preparation. This in turn can reduce uncertainty during a crisis call, reduce fear about disclosing one's suicide-related thoughts and behaviors to a stranger, and increase how readily accessible a helpline is to a prospective caller.

Similarly, in the context of an automated SMS text messaging–based SPI, young adult participants described that information about how the program functioned (eg, how many messages the system sent) and guidelines on the limits of confidentiality would be important when considering whether to use such a service. As G1-P3 articulated, "I have really bad anxiety so I am immediately thinking of well what if I do/say something "wrong" will they have a list of numbers and contact people and tell them." Given the potential real-life consequences of mandated reporting laws, even if relatively rare (eg, involuntary hospitalization, removal from one's home, or physical harm from emergency responders), participants in our study wanted transparency about the kind of information that would result in the automated intervention activating a mandated report. As G1-P3 alludes to, knowing in advance how the system responds to user input and operationalizes mandated reporting

XSL•FO
RenderX

can help ameliorate symptoms of anxiety while being instructive as to how to get the most benefit from the service.

The Role of Strong Privacy Practices

In addition to transparent information about how a suicide-related thought and behavior prevention intervention functions, the underlying assumptions it makes about users, and the procedures to which it adheres, young adult participants also expressed the importance of privacy including (1) autonomy over data and data use, (2) anonymity, and (3) discretion.

Participants wanted to ensure they maintained control over their data. In the context of an SMS text messaging–based SPI, participants were concerned that their messages could be read by another individual, the information they share could be reported to third parties, and that the messages they sent would be identifiable. As G2-P13 shared, "I would be concerned about privacy though. I would want to know that what I shared over text would not be read by another person without my consent, and that what I shared would not be traced back to me or reported to someone." By contrast, others saw the ability to create a safety plan via SMS text messages as a way to maintain control over how, when, and what the safety plan contained which they viewed as a form of data autonomy.

Another important technical specification for suicide-related thought and behavior-focused interventions was the ability to remain anonymous. As G2-P9 noted, "there is a fear of being traced and how it is not completely anonymous because the phone number is linked to you." Anonymity was one of the factors that influenced whether or not participants would use a crisis service. For example, despite one participant's misinterpretation of the National Suicide and Crisis Lifeline's processes, they noted that a key factor in making the Lifeline an attractive intervention was the ability to remain anonymous (The National Suicide and Crisis Lifeline maintains confidentiality except in cases of imminent risk and they do not enable completely anonymous contacts. Their Information Technology infrastructure and counselors collect callers' and texters' personally identifiable information such as phone numbers, IP addresses, or volunteered other information that may be linked back to an individual.).

Relatedly, participants wanted digital or remote suicide-related thought and behavior-related interventions to be discreet. This was especially true when considering the use of a texting–based SPI. Participants were particularly concerned that people in their lives with physical access to their devices, but no or partial knowledge of their suicide-related thoughts and behaviors, might be able to access material that would result in an unwanted disclosure of an individual's suicide-related thoughts or behaviors. G2-P5 noted, "I relate to the fear of being found out. Even with the promise of anonymity I'm still worried about someone in my life finding out and reporting me." Another participant, G2-P12, contextualized this well-represented fear around a digital intervention, "I definitely would also be nervous someone could see it on accident, because that's definitely happened to me before where I opened my phone and people I was eating lunch with saw what I had recently seen." While participants expressed general concerns around access to SMS text message data (eg, G1-P3 "I think the privacy is a big concern.", G1-P2 "I think many people feel vulnerable about their mental state and want it treated with as much privacy as possible."), most of the specific concerns about discretion and privacy focused on people in one's day-to-day environment gaining access to information and how that disclosure would negatively affect an individual's relationships. As G1-P6 explains, "most texting apps on phones don't have locks or passwords as far as I know, and if someone got a hold of the person's phone they would have immediate access to intimate and private details about the texter's mental health."

Usability, Convenience, and Accessibility

Participants shared a number of technical considerations related to an automated digital SPI. Among them was the importance of having a highly functional user interface. Participants were already familiar with the texting medium but noted that text-based menu navigation, relative to graphical user interfaces, could become tedious and frustrating. For example, G1-P8 said, "Text messages have a very specific, not customizable format. For something like a safety plan, I'd much prefer an app or website where I could work through my plan on a page or multiple pages, however that would look and be designed." However, participants largely agreed that texting–based safety planning could offer an important level of convenience and accessibility that may be especially important for self-management of suicide-related thoughts and behaviors. For example, G2-P9 noted, "The benefits is that it is easily accessible though and your phone is everywhere with you so if you need to refer back to it you will have it with you."

Participants also noted that an SMS text messaging–based service can make safety planning accessible to individuals that experience anxiety speaking to others about their concerns, in ways that synchronous voice call support cannot. G1-P1 shared, "I think the idea of texting gives me much less anxiety than having to call." Therefore, specific communication channels may provide different levels of comfort, based on participants' associations and past experiences.

Participants preferred tools that were efficient to use and did not require significant time or effort. This was especially true for participants who imagined using an automated digital SPI while in crisis. Further, if a program requires significant effort on the part of the individual, it could be demotivating and unhelpful. As G1-P13 shared, "I would be more likely to respond to these messages if they didn't take a lot of effort to respond to--if they involved a long, inconvenient chain of messages, for example, I wouldn't feel motivated to respond." It is critical that messages are focused and direct, but that they also have some semblance of personality. Participants like G1-P11 shared how they wanted messages styled, "The messaging dynamic should be concise but not totally devoid of feeling, even if it is totally automated." Participants wanted messages that made interacting with a tool interesting without being overly demanding.

Skill Building Features

Beyond intervention and technological requirements, participants in our study shared numerous preferences and feature suggestions for an automated digital SPI that helped

XSL•FO
RenderX

support skill building, not simply skill documentation—extending the role and function of a safety plan. For example, participants shared that they wanted self-tracking features so that they could identify patterns in their suicide-related thoughts and behaviors and build insights over time. G1-P4 shared one way this feature could be integrated into the larger automated digital SPI:

> *it could work in conjunction with tools where you can mark your feelings and behaviors, and once you know how you feel prior to an episode, it could help identify when you may potentially be feeling depressed or experiencing suicidal thoughts or behaviors.*

Others wanted suggestions for coping skills. This augments traditional safety planning by making automated digital SPIs a coping idea generation and refinement system, hosting a repository for useful coping skills that might spark inspiration for a user, especially if that user is having difficulty generating coping skills on their own. One participant shared how this might work:

> *I think that using specific questions to personalize someone's safety plan could be really useful. One big problem I have with the general concept of safety plans is that I feel that I wouldn't know what specific things to add to it, this would definitely solve that issue. I think specificity would definitely help me answer or respond to those messages, I feel less motivated to answer messages from automated services where the prompt is super vague or not very engaging. Other text messages that would be helpful to receive could be resources to distract or help cope with difficult situations. Another way the service could ask for things to incorporate into the plan without directly asking could be to ask for the user's interests or things that might help them calm down.* [G1-P11]

Relatedly, some participants wanted reminders to use coping skills. While safety plans have traditionally been a static document, the translation to an automated digital SPI enables the system to prompt the user to assess the need to use coping skills through periodic reminders. As G1-P2 shared, "I personally feel like getting reminders to relax and check in with myself would be helpful, since suicidal thoughts can be brought on by stress."

Taken together, participants shared that through reminders, skill suggestions, and self-tracking features, an automated digital SPI presents an opportunity to build skillfulness and insight for a potential user, rather than simply documenting effective coping skills.

Discussion

Principal Findings and Design Implications

This study aimed to understand the needs, preferences, and design requirements and features necessary for an automated SMS text messaging–based suicide prevention intervention. Our results suggest that young adults who experience suicide-related thoughts and behaviors are open to embedding suicide prevention interventions within SMS text messages, but these interventions must be designed to meet their unique needs and preferences.

One of the important needs arising from our ARC groups was the importance of social connection and accountability in automated digital SPI design. Young adults who experienced suicide-related thoughts and behaviors were interested in interventions that strengthen existing protective factors and social connections. This finding aligns with the Interpersonal Psychological Theory of Suicide [60,61] and the Three-Step Theory of Suicide [62,63], which each identify social disconnection as a key driver of suicidal desire and ideation. An automated digital SPI, such as the one proposed in this study, may offer psychoeducation to users on methods of leveraging existing social connections to build social accountability for symptom self-management. It may prompt users to consider available coping activities that foster a sense of connection (eg, letter writing, watching a movie with others). Additionally, an automated digital SPI may be able to support users in crafting messages to trusted others to solicit support around safety plan use and offer resources that users can send to trusted others to educate them on how to respond to someone experiencing suicidal thoughts (eg, tips to mitigate unhelpful reactions and methods of engaging in nonjudgmental listening). Importantly, our results also show that connection to trusted individuals within participants' personal networks was particularly valued relative to simulated connection or companionship achieved exclusively through texting with an automated chatbot. Designers of automated SPIs should consider designs that encourage and support real-world interpersonal connections with people in users' personal networks; otherwise, they could unintentionally exacerbate a sense of isolation.

Our finding that young adults believed that having social accountability may support their use of a digital safety plan, as well as their application of coping skills, also sits within the context of digital health literature that suggests, for some, framework-based human support and accountability [64,65] can facilitate longitudinal engagement in digital interventions [66]. Although human support aimed at engaging users in a digital health intervention is conceptually distinct from the social accountability young adults saw as valuable in facilitating the use of a safety plan—not necessarily a digital health intervention—there is conceptual overlap that signals the importance of others for facilitating the use of a therapeutic tool that may, at times, be difficult to engage with.

Participants in our study wanted to be highly informed about how an automated digital SPI would handle personal data. This finding has implications for the design of digital suicide prevention interventions broadly. Intervention systems must proactively communicate meta-information about how the intervention operates, as this has the potential to increase confidence and comfort using the intervention. Prior work demonstrates that individuals are more likely to make use of crisis helpline resources when they are provided alongside barrier-reduction interventions that target informational gaps such as misunderstandings of crisis helpline practices with regard to active rescues (ie, emergency responders being called to make contact with an individual) before receiving a referral to crisis helplines [67]. Moreover, U.S. Suicide and Crisis

XSL•FO
RenderX

helpline (ie, 988) data suggest that several of the top barriers to contacting the Suicide and Crisis helpline are due to fears about law enforcement being called and about privacy and confidentiality [68]. These top concerns are addressable through robust transparency practices that frontload information about the decision rules and processes used within the intervention and by its administrators. For instance, having clear and simple language before a user engages in the intervention that addresses (1) how and when confidentiality is broken and emergency services are called; (2) how data inputs are used such as contact information for individuals a user can reach out to for help during a crisis; (3) if, when, and how trusted others are contacted, and (4) proactive psychoeducational information about how privacy and confidentiality are maintained within the program (ie, data security measures). Finally, it is important to offer resources that empower users to ask questions about privacy, confidentiality, and clinical decision-making in other contexts such as formal treatment settings and crisis helpline calls. Thus, while our findings suggest the value of self-management to reduce the need for crisis services, they also suggest important ways that crisis services can improve.

Themes raised across our ARC groups highlight the importance of ensuring that an automated SPI is usable and convenient. In the context of SMS text messaging–based safety planning, this means the system design must incorporate bidirectional or interactive messaging. To date, SMS text messaging interventions for suicide prevention [40-44] have primarily used unidirectional messaging (originating from the intervention system and being sent to the user), leaving individuals with limited opportunity to engage with intervention material in ways that reflect their lived experiences. To ensure an SMS text messaging–based automated SPI is convenient and usable, the system must use the conversational medium to be interactive. Eliciting safety plan entries by prompting users to consider new or different aspects of coping and prompting users to be highly specific are all unique affordances offered by texting. To make an SMS text messaging system usable and convenient, a system must also leverage a text-based navigation menu. This enables users to access settings, jump to different parts of their safety plan, and modify elements of a safety plan, if needed. To this end, participants highlighted that while texting may be useful for safety plan creation, having multiple methods of editing a safety plan may be necessary if the navigation menu becomes cumbersome. Thus, offering multiple modes of interacting with one's safety plan (ie, via SMS text messages and via a web-based editing tool) may be required to make an automated digital SPI maximally usable.

Our results reveal an overlap between young adults' stated needs and preferences and innovations in the interactive text–based market. Our results highlighted that young adults had several needs and preferences that may lend themselves to innovation using emerging technologies. Specifically, young adults expressed a desire for a digital SPI service to support, among other needs, (1) coping skill acquisition and use, (2) quick and simple menu navigation and easy access to desired information, (3) interpersonal connection to trusted others, and (4) clear privacy policies. The recent commercial availability of highly customizable large language models (LLMs), a class of machine learning model trained on massive linguistic data sets and designed to process and generate text-based data, has increasingly been integrated into text-based mental health interventions in the form of chatbots, content generators, and as classifier algorithms that code or categorize data so that an intervention becomes more responsive to user inputs [69]. LLMs integrated into digital chatbots can tailor responses based on context, previous interactions, and user inputs. This emerging technology is beginning to be integrated into digital suicide prevention tools, especially as it relates to the detection of suicide-related user-generated content [70], and comes with significant risks in addition to substantial promise [71]. While our study did not explicitly seek to understand the role of LLMs in an automated digital SPI, participants in our study offered important insights that are relevant to intervention development in the context of a market where commercial LLMs are available and actively being incorporated into different digital mental health technologies.

First, as it relates to supporting coping skill acquisition and use, LLM-based messages may be launched to support users in ideating coping skills when they have difficulty generating their own. Moreover, the ability of LLMs to personalize content based on individuals' contexts and inputs creates opportunities to increase the relevance of a user's safety plan at the moment. With a user's existing safety plan as context, LLMs may be able to support idea generation about how to overcome barriers to coping (eg, if an individual's safety plan contains going for a walk outside as a coping skill, but weather is prohibitive, LLM-based messages may be able to support a user identifying relevant solutions, such as going for a walk indoors, taking protective gear, using other coping skills on one's safety plan). Second, LLMs may be able to improve the experience of navigating rigid text-based menus embedded in an SMS text messaging intervention by enabling the system to accept user responses that do not exactly match preprogrammed keywords that direct a user down different branching logic trees and instead allowing users to respond using natural language. Third, young adults were clear that they wanted a digital SPI to support interpersonal connections to other humans. While they were open to an automated digital SPI, connection to real people was crucial. LLM-based chatbots have often been used as conversational agents [70], some of which have personified the conversational agent. This may exacerbate a suicide-specific vulnerability: isolation and loneliness. Past work highlights that young adults with depression were concerned that using a robot for companionship could worsen their symptoms by making the absence of real human companions more apparent [72]. More research is needed to understand whether LLM-based suicide prevention tools should present themselves as personified agents or as self-service tools. Finally, given participants' emphasis on the importance of privacy and transparency, if an LLM was to be integrated into an automated digital SPI to enhance messages, it would need to (1) proactively make transparent ways in which user inputs would be used to train models and (2) disclose the potential risk of confidentiality breaches through the LLM itself, as has been documented in other contexts [73].

XSL•FO
RenderX

By engaging young adults with lived experience of suicide-related thoughts and behaviors to understand their needs, preferences, constraints, and the context of existing technological solutions, we can identify design requirements that are able to be applied across digital suicide prevention tools.

Limitations and Future Directions

This study is limited in several important ways. First, while we engaged in rigorous user-centered design practices to understand the needs and preferences of young people who experience suicide-related thoughts and behaviors, we were only able to engage a relatively small number of individuals, and our results may not be generalizable to broader young adult populations. In our sample of 30 participants, 11 (37%) were in formal mental health treatment, and 17 (57%) planned to seek formal treatment in the coming months. These treatment-engaged numbers are greater than the approximately 12% of US adults who have experienced suicidal thoughts and sought treatment in the past year [2]. Moreover, we recruited exclusively from MHA's website which may inherently limit the perspectives of individuals included in the study to those who are open to technology-assisted interventions and services, and potentially more actively engaged in self-management of mental health symptoms. While we aimed to cover a broad range of topics in our ARC groups, our discussions were guided by researcher-posted prompts and centered on eliciting needs and preferences as they related to safety planning or texting. To this end, in this study we did not systematically examine needs and preferences for risk management in self-guided digital safety planning tools; however, future work must aim to better understand the needs and preferences across interested parties (eg, consumers, research participants, clinicians, organizations) as it relates to risk management procedures. Finally, individuals at the highest levels of suicide risk (ie, individuals who may have had elevated levels of suicidal desire and intent) were not eligible to participate in this study and thus their voices were systematically excluded as the study format lacked appropriate support for these individuals. Future work should focus on applying these design considerations and assessing their value within the context of deployable digital suicide prevention interventions.

Conclusion

Engaging young adults in the design process of a digital suicide prevention tool revealed critical considerations that must be addressed if the tool is to effectively expand access to evidence-based care to reach young people at risk for suicide-related thoughts and behaviors. Specifically, automated digital SPIs must support building skillfulness to cope effectively with suicidal crises, deepening interpersonal connections, system transparency, and data privacy.

Acknowledgments

We thank Khendra Lucas of Mental Health America for input on, and support of, this work. JM acknowledges support from the National Institute of Mental Health (grant K08 MH128640). KAVO acknowledges support from the National Institute on Aging (grant K24AG084885).

Data Availability

Because of the sensitive nature of text-based data and the potential risk of reidentification, data are not publicly available. However, deidentified data are available upon reasonable request with the appropriate data use agreements in place.

Conflicts of Interest

JM has accepted consulting fees from Boehringer Ingelheim and Shirley Ryan Ability Lab. DCM has accepted honoraria and consulting fees from Boehringer-Ingelheim, Optum Behavioral Health, Centerstone Research Institute, Woebot Health, and the One Mind Foundation; royalties from Oxford Press; and has an ownership interest in Adaptive Health, Inc. RK has accepted consulting fees and honoraria from CHESS Health and Jimini Health. ARP has an ownership interest in SafeSide Prevention, which receives fees for suicide prevention education and consultation. The University of Rochester receives royalties from SafeSide Prevention and declares this interest. SafeSide Prevention had no role in and received no fees in relation to this work. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
Initial research prompts used to generate threaded conversations during asynchronous remote community online focus groups.
[DOCX File , 16 KB - formative_v9i1e69602_app1.docx]

References

1. Substance Abuse and Mental Health Services Administration (SAMHSA). National Survey on Drug Use and Health 2013 (NSDUH-2013). U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. URL: https://www.samhsa.gov/data/report/results-2013-national-survey-drug-use-and-health-mental-health-detailed-tables [accessed 2024-09-25]
2. Substance Abuse and Mental Health Services Administration (SAMHSA). National Survey on Drug Use and Health 2023 (NSDUH-2023). U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services

XSL•FO
RenderX

Administration, Center for Behavioral Health Statistics and Quality. URL: https://www.samhsa.gov/data/release/2023-national-survey-drug-use-and-health-nsduh-releases [accessed 2024-09-25]
3. Jobes DA, Joiner TE. Reflections on suicidal ideation. Crisis 2019 Jul;40(4):227-230 [FREE Full text] [doi: 10.1027/0227-5910/a000615] [Medline: 31274031]
4. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. Psychol Med 2011 Aug;41(8):1751-1761 [FREE Full text] [doi: 10.1017/S0033291710002291] [Medline: 21134315]
5. Czyz EK, Horwitz AG, Eisenberg D, Kramer A, King CA. Self-reported barriers to professional help seeking among college students at elevated risk for suicide. J Am Coll Health 2013;61(7):398-406 [FREE Full text] [doi: 10.1080/07448481.2013.820731] [Medline: 24010494]
6. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
7. Caine ED. Forging an agenda for suicide prevention in the United States. Am J Public Health 2013 May;103(5):822-829. [doi: 10.2105/AJPH.2012.301078] [Medline: 23488515]
8. Stanley B, Mann JJ. The need for innovation in health care systems to improve suicide prevention. JAMA Psychiatry 2020 Jan 01;77(1):96-98. [doi: 10.1001/jamapsychiatry.2019.2769] [Medline: 31577340]
9. Hom MA, Stanley IH, Joiner TE. Evaluating factors and interventions that influence help-seeking and mental health service utilization among suicidal individuals: a review of the literature. Clin Psychol Rev 2015 Aug;40:28-39. [doi: 10.1016/j.cpr.2015.05.006] [Medline: 26048165]
10. Michelmore L, Hindley P. Help-seeking for suicidal thoughts and self-harm in young people: a systematic review. Suicide Life Threat Behav 2012 Oct;42(5):507-524. [doi: 10.1111/j.1943-278X.2012.00108.x] [Medline: 22889130]
11. Abar B, Greener E, DeRienzo V, Botelho S, Wiegand T, Conner K. Perceived access to care among adult patients who attempted suicide. Psychol Serv 2018 Aug;15(3):305-308. [doi: 10.1037/ser0000196] [Medline: 30080088]
12. Zuromski KL, Dempsey CL, Ng THH, Riggs-Donovan CA, Brent DA, Heeringa SG, et al. Utilization of and barriers to treatment among suicide decedents: results from the Army Study to Assess Risk and Resilience Among Servicemembers (Army STARRS). J Consult Clin Psychol 2019 Aug;87(8):671-683 [FREE Full text] [doi: 10.1037/ccp0000400] [Medline: 31008631]
13. O'Reilly LM, Wiggs KK, Quinn PD. Suicide risk before mental health treatment initiation: implications for screening and access to care. Psychiatr Serv 2023 Jan 01;74(1):70-72. [doi: 10.1176/appi.ps.202100657] [Medline: 35707857]
14. Yakunina ES, Rogers JR, Waehler CA, Werth JL. College students' intentions to seek help for suicidal ideation: accounting for the help-negation effect. Suicide Life Threat Behav 2010 Oct;40(5):438-450. [doi: 10.1521/suli.2010.40.5.438] [Medline: 21034207]
15. Carlton PA, Deane FP. Impact of attitudes and suicidal ideation on adolescents' intentions to seek professional psychological help. J Adolesc 2000 Feb;23(1):35-45. [doi: 10.1006/jado.1999.0299] [Medline: 10700370]
16. Stanley B, Brown GK. Safety planning intervention: a brief intervention to mitigate suicide risk. Cognitive and Behavioral Practice 2012 May;19(2):256-264. [doi: 10.1016/j.cbpra.2011.01.001]
17. Abbott-Smith S, Ring N, Dougall N, Davey J. Suicide prevention: what does the evidence show for the effectiveness of safety planning for children and young people? - A systematic scoping review. J Psychiatr Ment Health Nurs 2023 Oct;30(5):899-910 [FREE Full text] [doi: 10.1111/jpm.12928] [Medline: 37052321]
18. Ferguson M, Rhodes K, Loughhead M, McIntyre H, Procter N. The effectiveness of the safety planning intervention for adults experiencing suicide-related distress: a systematic review. Arch Suicide Res 2021 Apr 29:1-24. [doi: 10.1080/13811118.2021.1915217] [Medline: 33913799]
19. Nuij C, van Ballegooijen W, de Beurs D, Juniar D, Erlangsen A, Portzky G, et al. Safety planning-type interventions for suicide prevention: meta-analysis. Br J Psychiatry 2021 Aug;219(2):419-426. [doi: 10.1192/bjp.2021.50] [Medline: 35048835]
20. National Action Alliance for Suicide Prevention: Transforming Health Systems Initiative Work Group. Recommended standard care for people with suicide risk: making health care suicide safe. National Action Alliance for Suicide Prevention. Washington, DC: Education Development Center, Inc; 2018. URL: https://theactionalliance.org/sites/default/files/action_alliance_recommended_standard_care_final.pdf [accessed 2025-02-24]
21. Toolkit: Engage. Zero Suicide. 2024. URL: https://zerosuicide.edc.org/toolkit/engage#collaborative-safety-plan [accessed 2024-10-31]
22. Gamarra JM, Luciano MT, Gradus JL, Wiltsey Stirman S. Assessing variability and implementation fidelity of suicide prevention safety planning in a regional VA healthcare system. Crisis 2015;36(6):433-439. [doi: 10.1027/0227-5910/a000345] [Medline: 26648231]
23. Davis M, Siegel J, Becker-Haimes EM, Jager-Hyman S, Beidas RS, Young JF, et al. Identifying common and unique barriers and facilitators to implementing evidence-based practices for suicide prevention across primary care and specialty mental health settings. Arch Suicide Res 2023;27(2):192-214 [FREE Full text] [doi: 10.1080/13811118.2021.1982094] [Medline: 34651544]

XSL•FO
RenderX

24. Haroz EE, Bajaj MA, Nestadt PS, Campo JV, Wilcox HC. Clinician perspectives on suicide safety planning and its implementation. Arch Suicide Res 2024 Jul 01:1-11. [doi: 10.1080/13811118.2024.2370852] [Medline: 38949292]
25. Gryglewicz K, Orr VL, McNeil MJ, Taliaferro LA, Hines S, Duffy TL, et al. Translating suicide safety planning components into the design of mHealth app features: systematic review. JMIR Ment Health 2024 Mar 28;11:e52763 [FREE Full text] [doi: 10.2196/52763] [Medline: 38546711]
26. Boudreaux ED, Brown GK, Stanley B, Sadasivam RS, Camargo CA, Miller IW. Computer administered safety planning for individuals at risk for suicide: development and usability testing. J Med Internet Res 2017 May 15;19(5):e149 [FREE Full text] [doi: 10.2196/jmir.6816] [Medline: 28506957]
27. Melvin GA, Gresham D, Beaton S, Coles J, Tonge BJ, Gordon MS, et al. Evaluating the feasibility and effectiveness of an Australian safety planning smartphone application: a pilot study within a tertiary mental health service. Suicide Life Threat Behav 2019 Jun;49(3):846-858 [FREE Full text] [doi: 10.1111/sltb.12490] [Medline: 29999193]
28. Methi N, Weeks I, Hunt R, McGuire TC, Rubin A, Decker M, et al. Accessibility and utility of an electronic self-guided safety plan for adolescents. J Clin Child Adolesc Psychol 2024;53(6):935-943. [doi: 10.1080/15374416.2024.2395271] [Medline: 39312731]
29. Rainbow C, Tatnell R, Blashki G, Fuller-Tyszkiewicz M, Melvin GA. Digital safety plan effectiveness and use: findings from a three-month longitudinal study. Psychiatry Res 2024 Mar;333:115748. [doi: 10.1016/j.psychres.2024.115748] [Medline: 38277811]
30. Larkin C, Tulu B, Djamasbi S, Garner R, Varzgani F, Siddique M, et al. Comparing the acceptability and quality of intervention modalities for suicidality in the emergency department: randomized feasibility trial. JMIR Ment Health 2023 Oct 24;10:e49783 [FREE Full text] [doi: 10.2196/49783] [Medline: 37874619]
31. Rainbow C, Tatnell R, Blashki G, Melvin GA. Perceived usefulness of self-guided versus collaborative suicide safety plans in online help-seekers. Crisis 2024 Jul;45(4):294-300. [doi: 10.1027/0227-5910/a000940] [Medline: 38319612]
32. Buus N, Erlangsen A, River J, Andreasson K, Frandsen H, Larsen JLS, et al. Stakeholder perspectives on using and developing the MYPLAN suicide prevention mobile phone application: a focus group study. Arch Suicide Res 2020;24(1):48-63. [doi: 10.1080/13811118.2018.1489319] [Medline: 29985770]
33. Kennard BD, Biernesser C, Wolfe KL, Foxwell AA, Craddock LSJ, Rial KV, et al. Developing a brief suicide prevention intervention and mobile phone application: a qualitative report. J Technol Hum Serv 2015 Oct 01;33(4):345-357 [FREE Full text] [doi: 10.1080/15228835.2015.1106384] [Medline: 26977137]
34. O'Grady C, Melia R, Bogue J, O'Sullivan M, Young K, Duggan J. A mobile health approach for improving outcomes in suicide prevention (SafePlan). J Med Internet Res 2020 Jul 30;22(7):e17481 [FREE Full text] [doi: 10.2196/17481]
35. Larkin C, Djamasbi S, Boudreaux ED, Varzgani F, Garner R, Siddique M, et al. ReachCare Mobile Apps for Patients Experiencing Suicidality in the Emergency Department: Development and Usability Testing Using Mixed Methods. JMIR Form Res 2023 Jan 27;7:e41422 [FREE Full text] [doi: 10.2196/41422] [Medline: 36705961]
36. Sutori S, Hadlaczky G, Eliasson E, Wasserman D, Carli V. Systematic review and meta-analysis: effectiveness of stand-alone digital suicide preventive interventions for the self-management of suicidality. J Technol Behav Sci 2023 Dec 30;9(3):440-451. [doi: 10.1007/s41347-023-00374-7]
37. Newport F. The new era of communication among Americans. Gallup. 2014. URL: https://news.gallup.com/poll/179288/new-era-communication-americans.aspx [accessed 2020-08-30]
38. U.S. smartphone use in 2015. Pew Research Center. 2015. URL: https://www.pewresearch.org/internet/2015/04/01/us-smartphone-use-in-2015/ [accessed 2024-09-24]
39. Griffioen N, Scholten H, Lichtwarck-Aschoff A, van Rooij M, Granic I. Everyone does it—differently: a window into emerging adults' smartphone use. Humanit Soc Sci Commun 2021 Jul 20;8(1):1-11. [doi: 10.1057/s41599-021-00863-1]
40. Czyz EK, Arango A, Healy N, King CA, Walton M. Augmenting safety planning with text messaging support for adolescents at elevated suicide risk: development and acceptability study. JMIR Ment Health 2020 May 25;7(5):e17345 [FREE Full text] [doi: 10.2196/17345] [Medline: 32160150]
41. Comtois KA, Kerbrat AH, DeCou CR, Atkins DC, Majeres JJ, Baker JC, et al. Effect of augmenting standard care for military personnel with brief caring text messages for suicide prevention: a randomized clinical trial. JAMA Psychiatry 2019 May 01;76(5):474-483. [doi: 10.1001/jamapsychiatry.2018.4530] [Medline: 30758491]
42. Berrouiguet S, Larsen ME, Mesmeur C, Gravey M, Billot R, Walter M, HUGOPSY Network, et al. Toward mHealth brief contact interventions in suicide prevention: case series from the suicide intervention assisted by messages (SIAM) randomized controlled trial. JMIR Mhealth Uhealth 2018 Jan 10;6(1):e8 [FREE Full text] [doi: 10.2196/mhealth.7780] [Medline: 29321126]
43. Pisani AR, Wyman PA, Petrova M, Judd E, Schmeelk-Cone K, Thiha P, et al. Framework for supporting adolescent peer leaders: a pilot using text messaging in a school-based substance use prevention program. J Prim Prev 2019 Apr;40(2):243-254 [FREE Full text] [doi: 10.1007/s10935-019-00545-4] [Medline: 30827007]
44. Thiha P, Pisani AR, Gurditta K, Cherry E, Peterson DR, Kautz H, et al. Efficacy of web-based collection of strength-based testimonials for text message extension of youth suicide prevention program: randomized controlled experiment. JMIR Public Health Surveill 2016 Nov 09;2(2):e164 [FREE Full text] [doi: 10.2196/publichealth.6207] [Medline: 27829575]

XSL•FO
RenderX

45. Czyz EK, King CA, Prouty D, Micol VJ, Walton M, Nahum-Shani I. Adaptive intervention for prevention of adolescent suicidal behavior after hospitalization: a pilot sequential multiple assignment randomized trial. J Child Psychol Psychiatry 2021 Aug;62(8):1019-1031 [FREE Full text] [doi: 10.1111/jcpp.13383] [Medline: 33590475]
46. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition. Washington, DC: American Psychiatric Association; 2013.
47. Stanley IH, Hom MA, Christensen K, Keane TM, Marx BP, Björgvinsson T. Psychometric properties of the Depressive Symptom Index-Suicidality Subscale (DSI-SS) in an adult psychiatric sample. Psychol Assess 2021 Oct;33(10):987-997. [doi: 10.1037/pas0001043] [Medline: 34180693]
48. Beck A, Steer R, Brown G. Manual for the Beck Depression Inventory-II. San Antonio, TX: Psychological Corporation; 1996.
49. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
50. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
51. MacLeod H, Jelen B, Prabhakar A, Oehlberg L, Siek K, Connelly K. A guide to using asynchronous remote communities (ARC) for researching distributed populations. EAI Endorsed Trans Perv Health Tech 2017 Jul 18;3(11):e4. [doi: 10.4108/eai.18-7-2017.152898]
52. Mastre J, MacLeod H, Connelly C, Dunbar J, Beck J, Siek K, et al. Defining through expansion: conducting asynchronous remote communities (ARC) research with stigmatized groups. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Apr 21 Presented at: The 2018 CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal QC, Canada p. 1-13. [doi: 10.1145/3173574.3174131]
53. Prabhakar A, Guerra-Reyes L, Kleinschmidt V, Jelen B, MacLeod H, Connelly K, et al. Investigating the suitability of the asynchronous, remote, community-based method for pregnant and new mothers. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. New York, NY: ACM Press; 2017 Presented at: The 2017 CHI Conference on Human Factors in Computing Systems; May 6-11, 2017; Denver, CO p. 4924-4934. [doi: 10.1145/3025453.3025546]
54. Hinshaw SP, Hoagwood K, Jensen PS, Kratochvil C, Bickman L, Clarke G, et al. AACAP 2001 research forum: challenges and recommendations regarding recruitment and retention of participants in research investigations. J Am Acad Child Adolesc Psychiatry 2004 Aug;43(8):1037-1045. [doi: 10.1097/01.chi.0000129222.89433.3d] [Medline: 15266200]
55. FocusGroupIt. FocusGroupIt. URL: https://www.focusgroupit.com/ [accessed 2023-11-24]
56. Meyerhoff J, Kornfield R, Mohr DC, Reddy M. Meeting young adults' social support needs across the health behavior change journey: implications for digital mental health tools. Proc ACM Hum Comput Interact 2022 Nov 11;6(CSCW2):1-33 [FREE Full text] [doi: 10.1145/3555203] [Medline: 36387059]
57. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
58. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qualitative Research in Sport, Exercise and Health 2019 Jun 13;11(4):589-597. [doi: 10.1080/2159676x.2019.1628806]
59. Dedoose. Los Angeles, CA: SocioCultural Research Consultants, LLC; 2018. URL: https://www.dedoose.com/ [accessed 2024-09-24]
60. Joiner T. Why People Die by Suicide. Cambridge, MA: Harvard University Press; 2005.
61. Van Orden KA, Witte TK, Cukrowicz KC, Braithwaite SR, Selby EA, Joiner TE. The interpersonal theory of suicide. Psychol Rev 2010 Apr;117(2):575-600 [FREE Full text] [doi: 10.1037/a0018697] [Medline: 20438238]
62. Klonsky ED, Pachkowski MC, Shahnaz A, May AM. The three-step theory of suicide: description, evidence, and some useful points of clarification. Prev Med 2021 Nov;152(Pt 1):106549. [doi: 10.1016/j.ypmed.2021.106549] [Medline: 34538372]
63. Klonsky ED, May AM. The three-step theory (3ST): a new theory of suicide rooted in the “ideation-to-action” framework. International Journal of Cognitive Therapy 2015 Jun;8(2):114-129. [doi: 10.1521/ijct.2015.8.2.114]
64. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
65. Schueller SM, Tomasino KN, Mohr DC. Integrating human support into behavioral intervention technologies: the efficiency model of support. Clin Psychol Sci Pract 2016 Nov 17;24(1):27-45. [doi: 10.1111/cpsp.12173]
66. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
67. Jaroszewski AC, Morris RR, Nock MK. Randomized controlled trial of an online machine learning-driven risk assessment and intervention platform for increasing the use of crisis services. J Consult Clin Psychol 2019 Apr;87(4):370-379. [doi: 10.1037/ccp0000389] [Medline: 30883164]

XSL•FO
RenderX

68. Ad Council Research Institute, National Action Alliance for Suicide Prevention, Suicide Prevention Resource Center. Framework for Successful Messaging. URL: https://suicidepreventionmessaging.org/sites/default/files/2024-05/988%20FR_Trusted%20Messengers_Report_Final-508.pdf [accessed 2024-07-12]
69. Stade EC, Stirman SW, Ungar LH, Boland CL, Schwartz HA, Yaden DB, et al. Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. Npj Ment Health Res 2024 Apr 02;3(1):12. [doi: 10.1038/s44184-024-00056-z] [Medline: 38609507]
70. Guo Z, Lai A, Thygesen JH, Farrington J, Keen T, Li K. Large language models for mental health applications: systematic review. JMIR Ment Health 2024 Oct 18;11:e57400 [FREE Full text] [doi: 10.2196/57400] [Medline: 39423368]
71. Choudhury M, Pendse S, Kumar N. Benefits and harms of large language models in digital mental health. arXiv. Preprint posted online on November 7, 2023 2023:1 [FREE Full text] [doi: 10.31234/osf.io/y8ax9]
72. Kornfield R, Meyerhoff J, Studd H, Bhattacharjee A, Williams JJ, Reddy M, et al. Meeting users where they are: user-centered design of an automated text messaging tool to support the mental health of young adults. Proc SIGCHI Conf Hum Factor Comput Syst 2022 Apr;2022:329. [doi: 10.1145/3491102.3502046] [Medline: 35574512]
73. White J. How strangers got my email address from ChatGPT's model. The New York Times. 2023 Dec 22. URL: https://www.nytimes.com/interactive/2023/12/22/technology/openai-chatgpt-privacy-exploit.html [accessed 2024-11-19]

Abbreviations

ARC: asynchronous remote community
CCSI: Cross-Cutting Symptom Inventory
DSI-SS: Depression Severity Index—Suicidality Subscale
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
LLM: large language model
MHA: Mental Health America
SPI: safety planning intervention

Edited by A Mavragani; submitted 03.12.24; peer-reviewed by C Wijaya, C Rainbow, A AL-Asadi; comments to author 30.12.24; revised version received 04.02.25; accepted 11.02.25; published 26.02.25.

Please cite as:
Meyerhoff J, Popowski SA, Lakhtakia T, Tack E, Kornfield R, Kruzan KP, Krause CJ, Nguyen T, Rushton K, Pisani AR, Reddy M, Van Orden KA, Mohr DC
Automated Digital Safety Planning Interventions for Young Adults: Qualitative Study Using Online Co-design Methods
JMIR Form Res 2025;9:e69602
URL: https://formative.jmir.org/2025/1/e69602
doi:10.2196/69602
PMID:

©Jonah Meyerhoff, Sarah A Popowski, Tanvi Lakhtakia, Emily Tack, Rachel Kornfield, Kaylee P Kruzan, Charles J Krause, Theresa Nguyen, Kevin Rushton, Anthony R Pisani, Madhu Reddy, Kimberly A Van Orden, David C Mohr. Originally published in JMIR Formative Research (https://formative.jmir.org), 26.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluation of the Clear Fear Smartphone App for Young People Experiencing Anxiety: Uncontrolled Pre– and Post–Follow-Up Study

Chiara Samele[1], MSc, PhD; Norman Urquia[1], MSc, PhD; Rachel Edwards[2], MSc; Katie Donnell[2], MSc; Nihara Krause[2], MSc, PsychD

[1]Informed Thinking, London, United Kingdom

[2]stem4, London, United Kingdom

Corresponding Author:
Nihara Krause, MSc, PsychD
stem4
Connect House, 133-137 Alexandra Road
Wimbledon
London, SW19 7JY
United Kingdom
Phone: 44 7956396375
Email: dr.nihara.krause@stem4.org.uk

Abstract

Background: Mobile health apps are proving to be an important tool for increasing access to psychological therapies early on, particularly with rising rates of anxiety and depression in young people.

Objective: We aimed to assess the usability, acceptability, safety, and effectiveness of a new app, Clear Fear, developed to help young people manage symptoms of anxiety using the principles of cognitive behavioral therapy.

Methods: The Clear Fear app was developed to provide cognitive behavioral strategies to suit anxiety disorders. An uncontrolled pre– and post–follow-up design over a 9-week period was used to assess the app and its effects. This study comprised 3 phases: baseline (stage 1), post–app familiarization phase (stage 2), and follow-up (stage 3). Eligible participants were aged between 16 and 25 years with mild to moderate anxiety but not currently receiving treatment or in contact with specialist mental health services or using other interventions or apps to help monitor or manage their mental health. A community sample was recruited via advertisements, relevant websites, and social media networks. Eligible participants completed standardized self-report tools and questionnaires at each study stage. These measured probable symptoms of anxiety (7-item Generalized Anxiety Disorder scale) and depression (Mood and Feelings Questionnaire); emotional and behavioral difficulties (Strengths and Difficulties Questionnaire); and feedback on the usability, accessibility, and safety of the app. Mean scores at baseline and follow-up were compared using paired 2-tailed *t* tests or Wilcoxon signed rank tests. Qualitative data derived from open-ended questions were coded and entered into NVivo (version 10) for analysis.

Results: A total of 48 young people entered the study at baseline, with 37 (77%) completing all outcome measures at follow-up. The sample was mostly female (37/48, 77%). The mean age was 20.1 (SD 2.1) years. In total, 48% (23/48) of the participants reached the threshold for probable anxiety disorder, 56% (27/48) had positive scores for probable depression, and 75% (36/48) obtained a total score of “very high” on the Strengths and Difficulties Questionnaire for emotional and behavioral difficulties. The app was well received, offering reassurance, practical and immediate help to manage symptoms, and encouragement to seek help, and was generally found easy to use. A small minority (3/48, 6%) found the app difficult to navigate. The Clear Fear app resulted in statistically significant reductions in probable symptoms of anxiety (t_{36}=2.6, 95% CI 0.41-3.53; *P*=.01) and depression (*z*=2.3; *P*=.02) and behavioral and emotional difficulties (t_{47}=4.5, 95% CI 3.67-9.65; *P*<.001), representing mostly medium to large standardized effect sizes.

Conclusions: The Clear Fear app was found to be usable, acceptable, safe, and effective in helping manage symptoms of anxiety and depression and emotional and behavioral difficulties.

(JMIR Form Res 2025;9:e55603) doi:10.2196/55603

XSL•FO
RenderX

KEYWORDS

mental health; anxiety; depression; emotional and behavioral difficulties; mobile phone app; cognitive behavioral therapy; digital tool; young people; mobile phone

Introduction

Background

The rates of probable mental health problems in the United Kingdom among young people have steadily increased since 2017. According to National Health Service Digital data (2022), for example, the rates for young people aged 17 to 19 years with mental health problems rose by 7.6% in 2020 and reached 25.7% in 2022 [1]. In 2017 to 2018, depression and anxiety among young women aged 16 to 24 years rose by 26% compared to the previous year [2].

The high volume of young people with experience of symptoms associated with moderate depression and anxiety has prompted a rise in referrals to already overstretched specialist mental health services. For example, in November 2023, a total of 445,000 people were in contact with children and young people's mental health services compared to 362,000 in November 2021 [3]. New referrals to these services reached a peak of 87,000 per month in 2021, the highest since 2019 [4]. While referrals have increased, there remains a huge shortfall in those who need support and receive it. In 2020, the proportion of children and young people needing and receiving support was 27% [4].

Therefore, it is important to find ways to ensure that young people have timely access to support. One growing avenue is support through mobile health apps.

Approximately 96% of those aged between 16 and 24 years own a smartphone [5]. Mobile health apps now constitute an important tool for increasing access to psychological therapies, particularly for young people. Several published systematic reviews have assessed the usability, feasibility, acceptability, functionality, and effectiveness of specific mobile mental health apps for young people [6,7]. Despite the published literature on how mobile app technology can increase access to psychological therapy, relatively little research exists on its effectiveness and whether it improves symptoms of mental health problems. However, there is some evidence suggesting that certain mobile apps for mental health can help improve the monitoring and management of mental health [8]. A handful of studies confirm that web-based cognitive behavioral therapy (CBT) is as effective as face-to-face treatment for anxiety and depression [9]. Among young people, mobile mental health apps also offer an important opportunity to prevent mental health problems or intervene early to reduce the risk of a long-term mental health condition into adulthood [10].

Objectives

Given the limited available evidence on the effectiveness of mental health apps for young people, this pilot study aimed to assess whether the Clear Fear app, developed to help young people manage symptoms of anxiety, could reduce these symptoms.

The main aims of this study were to assess the usability, acceptability, safety, and effectiveness of the Clear Fear app, specifically (1) what particular features of the app young people found most helpful or beneficial; (2) the ease of use, acceptability, and frequency with which young people used the app, as well as whether they found it safe and any improvements that could be made; and (3) whether the Clear Fear app was able to reduce or improve symptoms of anxiety primarily (and any adverse behavior) following its use during the study period.

Thus, this study was designed to test the following hypothesis: a self-guided mobile cognitive behavioral app can improve anxiety symptoms and be acceptable to young people over a period of 9 weeks.

Methods

Clear Fear App

Aims of the App

The Clear Fear app aims to support young people in managing symptoms of anxiety early in their stage of development. It does this by (1) providing cognitive behavioral strategies to suit anxiety disorders; (2) encouraging self-monitoring and self-management and building resilience; and (3) teaching alternative, adaptive ways to deal with anxiety.

Users manage their anxiety and "face their fear" in a gradual way while learning to manage panic, challenge worries, and regulate their emotions.

The app provides a first step for young people looking for immediate techniques to help manage anxious thoughts and behaviors for longer-term change. It was also designed to be easily accessible for those who do not meet the criteria for mental health services or are currently on a waiting list to receive appropriate treatment for anxiety disorder. The app does not replace specialist treatment but can be used in addition to it.

Development, Structure, and App Function

The app was developed by NK at stem4, a charity supporting teenage mental health, in collaboration with young people. The development process began with reviewing available mainstream apps that aimed to manage anxiety through mindfulness. Integral to the process was the coproduction of the Clear Fear app with a group of 30 young people (aged between 14 and 17 years) who took part in a series of user engagement sessions to explore desired outputs, user journeys, visual concepts, colors, imagery, security and privacy considerations, user experience, and tone of voice. The app provides a low-intensity CBT treatment to monitor and change thoughts, emotions, and behaviors associated with mild to moderate symptoms of anxiety. It seeks to calm fear responses through a cognitive behavioral framework along with breathing, relaxation, and mindfulness exercises.

The Clear Your Fear section of the app provides activities to help users with the 4 key areas—dealing with emotions,

XSL•FO
RenderX

managing worries, reacting to worries, and managing physical responses to anxiety—and is described in Textbox 1 and Figure 1. Each activity provides some brief introductory information to explain why the activity can help, as well as examples where necessary throughout.

The multiple areas for goal setting provided in the app facilitate the following: identifying specific goals, tailoring each goal to suit each individual, identifying barriers, problem-solving, breaking long-term goals into short-term challenges, accomplishing easier challenges first, prompts, cueing, reinforcing success, social support, and tracking personal information to enhance self-directed learning.

Textbox 1. Description of the Clear Your Fear section of the Clear Fear app.

Dealing With Your Emotions

This consists of activities for users to learn to identify, express, and regulate emotions, which can help reduce anxiety.

- Express Yourself activities provide opportunities for note keeping and finding alternative ways of emotional regulation.
- Stay Calm activities help users learn to calm the body and mind by doing either the guided breathing activity or mindfulness exercises.
- Laugh & Smile activities aid in substituting negative emotions for positive ones through the use of humor by viewing jokes or gifs.

Managing Your Worries

This focuses on learning how thinking can contribute to anxiety and how to make a change by learning to understand, identify, and change fearful thoughts (Worry Warriors); have a break from anxious thoughts (Worry Box); or put a problem-solving strategy into practice (Worry Ladder).

- Worry Warriors guides the user through a process of cognitive restructuring to identify and combat 3 common types of automatic negative thoughts (those that have been magnified by anxiety or taken to the extreme or that are about something negative that is expected to happen), together with self-monitoring, seeking evidence, and restructuring thoughts. The resulting positive thoughts that are created can then be revisited in the Grit Box, located in another area of the app. Therapeutic gains from this process can include symptom relief and an increased sense of control over anxiety through its objectification.
- Worry Box is an activity in which the user types in their worries to be stored in this section of the app, which can then be emptied to metaphorically clear them.
- Worry Ladder guides the user through dealing with their worries one step at a time (and creating a visual ladder of these steps in the process) by first stating a worry and then typing in possible solutions; choosing one to try; and, finally, noting what happened and what worked. Both the Worry Ladder and Worry Box, together with problem-solving, help with prioritizing and containing anxiety.

Reacting to Worries

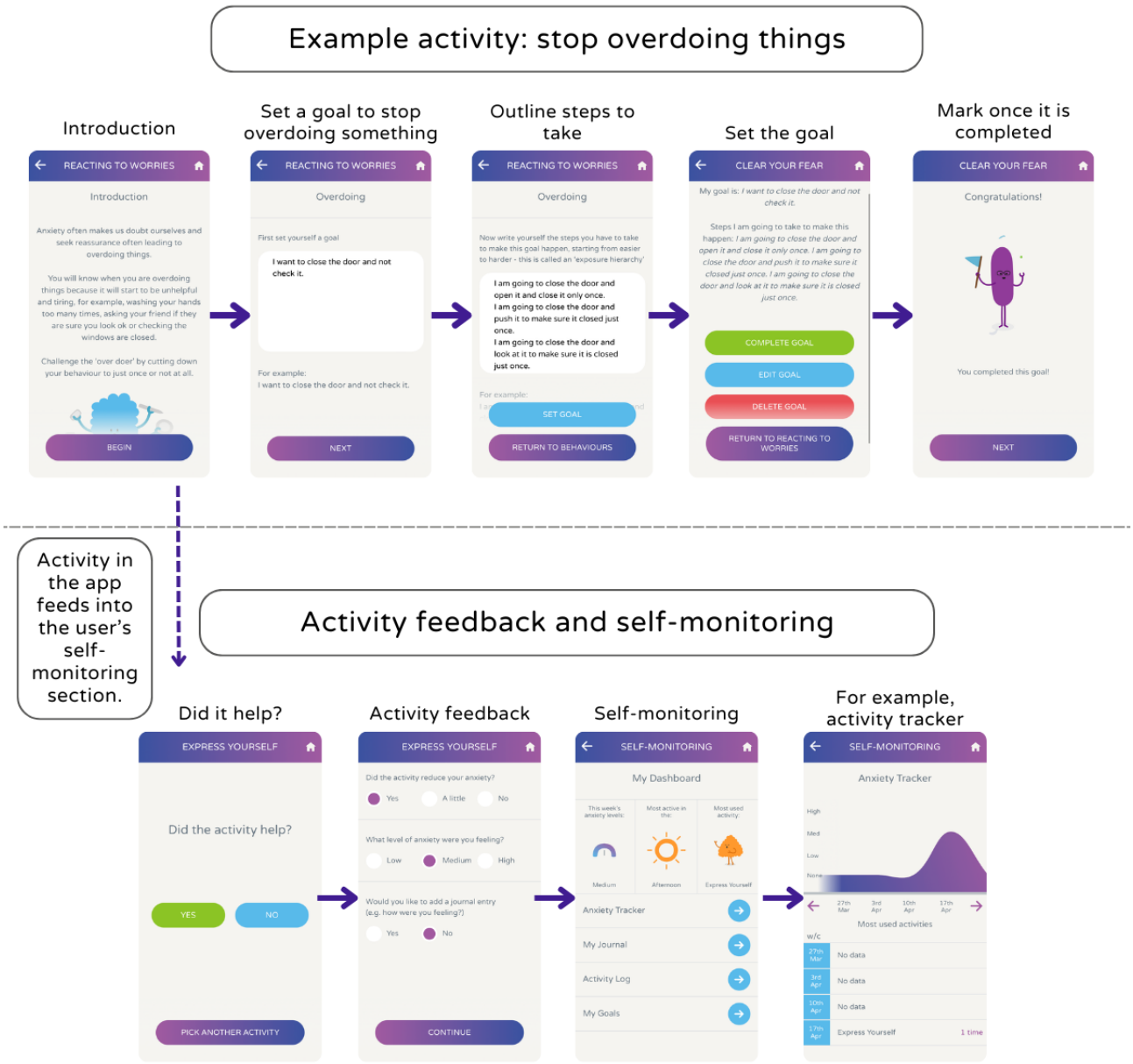
This involves the user identifying situations in which they want to experience less anxiety and learn to face their worries with the help of activities to lessen overdoing or avoiding things. In each case, the user is guided through setting a goal and breaking it down into an exposure hierarchy of manageable steps and can also mark when a goal has been achieved.

Managing Physical Responses to Anxiety

This consists of activities for users to cope with and lessen the physical effects of anxiety. In total, 4 categories of goals are presented (exercise, diet, relaxation, and sleep), within which the user is given brief information as to why the activity can help and suggestions of goals to set, which can be input into the app. The user is also given the option of setting push notifications (at daily, weekly, or monthly intervals) for the goal to help remind them.

XSL•FO
RenderX

Figure 1. Flowchart and screenshots of the Clear Fear app, the digital intervention used in this uncontrolled pre– and post–follow-up study to evaluate its usability, acceptability, safety, and effectiveness in helping young people manage symptoms of anxiety using the principles of cognitive behavioral therapy over a 9-week period.



Activity Feedback and Self-Monitoring

Upon completion of an activity, the user is asked to provide the following feedback: whether the activity helped (Yes or No), whether it reduced their anxiety (Yes, A little, or No), and the level of anxiety they were feeling (Low, Medium, or High). This information feeds into that user's self-monitoring section on the app, which includes an anxiety profile and anxiety tracker to help the user self-monitor across every task.

Grit Box

The Grit Box is based on the concept that grit is a necessary factor in resilience. It holds the user's own Worry Warrior Positive Thoughts, which are generated from their activity on the app. It also contains a range of motivational statements, positive quotes, and information on resilient and inspirational people as suggested during the coproduction process with young people.

Other Areas of the App

Psychoeducation is provided in the information section and in optional notifications sent to the users to support their momentum in reaching personal goals and in gaining valuable psychological insights (eg, setting realistic goals, overcoming fears, and increasing motivation). On achieving a goal, users receive a plan on how to maintain the new skills and avoid falling back into old habits.

In addition, information about anxiety is provided via messages sent to the user to help normalize their experience, as well as giving information on the value and functioning of anxiety. This includes information on specific types of anxiety (namely,

XSL•FO
RenderX

generalized anxiety, social anxiety, performance anxiety, obsessive-compulsive disorder and health anxiety, separation anxiety, exam anxiety, phobias, and fear of missing out).

Finally, users can receive physiological support and help in dealing with panic and panic attacks based on mindfulness exercises in both visual and auditory formats to improve low self-esteem, emotional reactivity, and functioning [11], as well as signposts to further support.

The Clear Fear App's Contribution to Digital Mental Health

The Clear Fear app is a clinician-developed, evidence-based, and comprehensive CBT app that differs from those using mindfulness and meditation as their primary interventions, artificial intelligence chatbots, and other available apps for anxiety management developed *without* clinician input or comprehensive evidence-based approaches [12,13].

The Clear Fear app is designed primarily for managing anxiety symptoms in young people. It provides a comprehensive evidence-based approach that is both structured and personalizable, incorporating CBT exercises, psychoeducation, mindfulness, and self-monitoring. The content, language, visuals, tone of voice, and user journeys are age appropriate. For younger users aged between 11 and 15 years, the Clear Fear app includes age-appropriate design and language, ease of navigation, age-relevant gamification, less abstract cognitive concepts, tasks that cater to shorter attention spans, and increased socialization.

It also prioritizes user privacy and data safety as it does not collect identifiable information or require user accounts, can be passcode protected, is regularly reviewed and updated, and is compliant with regulatory standards such as the National Health Service clinical safety standard DCB0129.

Unlike many mental health apps, Clear Fear is free to download in the United Kingdom, with no subscriptions, in-app purchases, or advertisements, and can be used without an internet connection, making it highly accessible.

Study Design

This study used an uncontrolled pre– and post–follow-up design over a 9-week period to assess the app and its effects. The study comprised three main phases: (1) baseline and app introduction (stage 1), (2) post–app familiarization phase (stage 2), and (3) follow-up (stage 3).

Participants and Eligibility Criteria

Participants were a community sample recruited via stem4's website and their existing social media networks.

Potential participants were eligible to take part in the study if they (1) were aged between 16 and 25 years and (2) had mild to moderate anxiety.

Participants were excluded if they were (1) in contact with mental health services or taking medication for a mental health problem; (2) using other interventions or apps to monitor their mental health; or (3) at risk of suicide, diagnosed with a severe and enduring mental health problem such as psychosis or major depression, or experiencing substance use problems.

Potential participants were screened using a 10-item questionnaire developed for this study on suicide ideation, any severe mental health problems, and substance use issues (drugs and alcohol).

Recruitment

The main method for recruiting participants was through advertisements on the stem4 website and via stem4 social media channels inviting young people who experienced anxiety and had not previously used the Clear Fear app to take part in the study. stem4 also provided details of the study to schools within their local area. Eligible participants were offered a web-based shopping voucher as an incentive to take part in the study.

A web-based screening questionnaire (developed for this study) was sent to all young people interested in taking part in the study to assess their study eligibility and level of risk. This included a 10- to 15-item screening tool to identify any current mental health issues or history of and contact with mental health services. Those identified with an increased risk at initial screening were excluded from the study to prioritize their safety. These individuals were provided with an information pack containing emergency contact numbers and guidance on further support, and they were encouraged to notify the research team if they accessed additional support.

Eligible participants were provided with an information sheet detailing the study and asked to complete an informed consent form before taking part. Participants could leave the study at any time during the study without giving a reason if they wished. Participants were also made aware that the Clear Fear app does not replace traditional treatment and were appropriately signposted to sources of help where necessary. Participants also received an introductory leaflet with information on symptoms and suggested strategies for mitigating risk and were directed to sections on the app that provided immediate signposts and access to services.

Furthermore, at the start of the study, all participants completed the Safety Net section of the app, which is a vital safety feature that is always available to the user and can be shared with responsible adults. The users populated what they could think and do when they needed help and the numbers of people they could contact to go along with the Safety Net section's prepopulated emergency numbers.

The app also prompted users to seek help from a trusted adult or general practitioner if they noticed that their anxiety levels, from the tracker in the Self-Monitoring section, were consistently high.

Sample Size

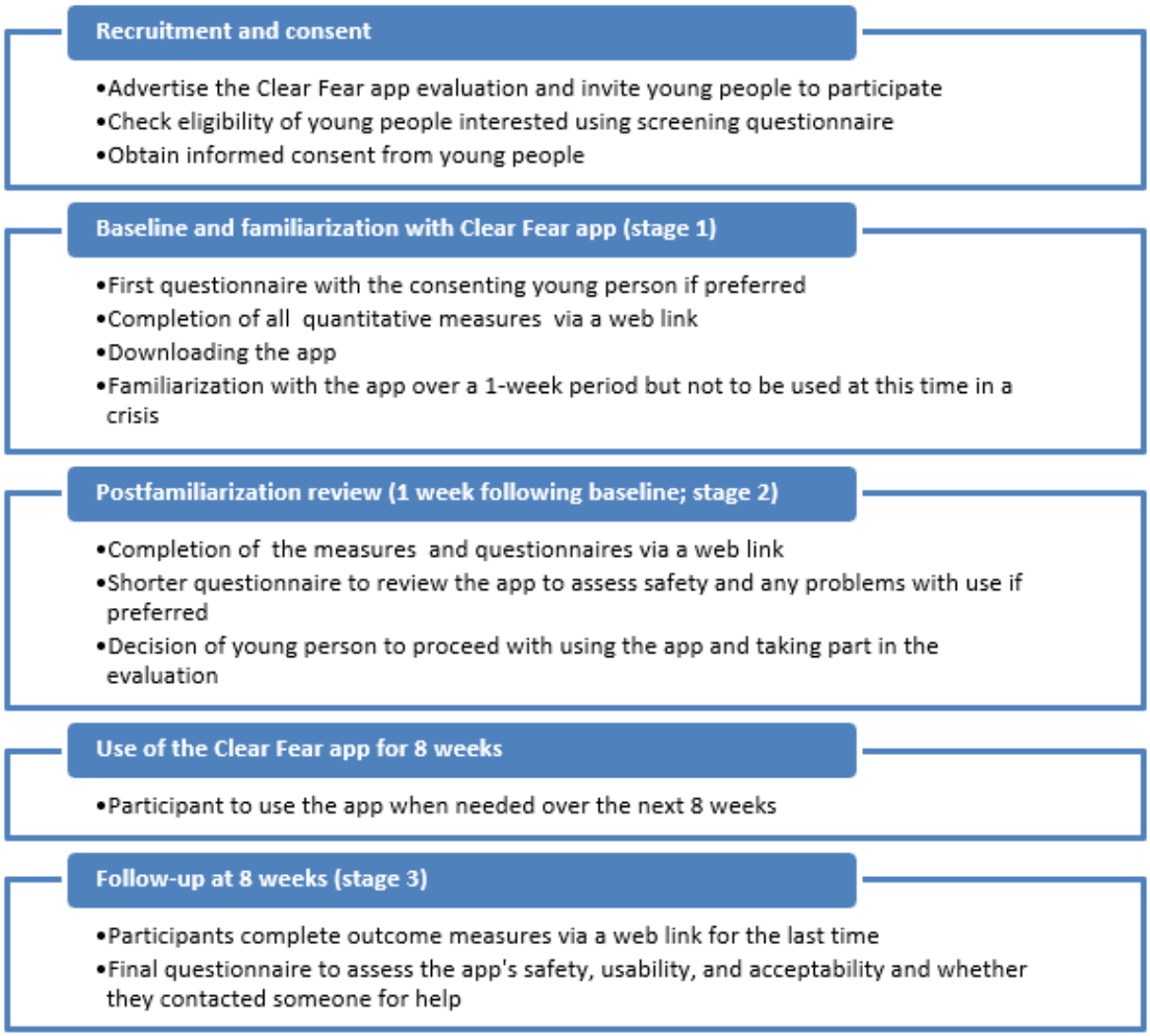
The aim was to recruit approximately 45 young people to address the main study aims. No formal sample size calculation was made given the preliminary nature of the study and piloting of the Clear Fear app. We allowed for attrition rates of up to 15% at the postfamiliarization and follow-up time points.

XSL•FO
RenderX

Study Procedure

Figure 2 shows the study procedure, including the 3 main study stages and the recruitment process. We also consulted with 6 people (aged between 14 and 17 years) to provide feedback on the study proposal and procedure regarding their preferences concerning web-based data collection.

Figure 2. Diagram of the procedure of this uncontrolled pre– and post–follow-up study to evaluate the Clear Fear app's usability, acceptability, safety, and effectiveness in helping young people manage symptoms of anxiety using the principles of cognitive behavioral therapy over a 9-week period.



Baseline (Stage 1)

Upon entry to the study, participants were provided with web-based instructions on how to download and use the Clear Fear app and the study procedure. Participants were then allocated a study ID number to maintain confidentiality and anonymity and emailed a link to the web-based baseline questionnaires. The link was created by the researchers using SurveyMonkey (SurveyMonkey Inc).

At this stage, participants were asked only to familiarize themselves with the Clear Fear app over the coming week and not to use it during a time of crisis. This was to ensure the person's safety until they became fully versed in using the Clear Fear app and to obtain baseline measures of the frequency and severity of anxiety they experienced and their existing coping mechanisms.

Post–App Familiarization Phase (Stage 2)

One week after baseline, web-based questionnaires were sent again to participants, which included a user questionnaire to gauge their assessment of the Clear Fear app and any problems they may have had. If the participant wished to continue taking part in the study, they were asked to complete the outcome measures again. Participants were informed that they could now use the Clear Fear app as often as they wanted over the following 8 weeks.

Follow-Up (Stage 3)

After an 8-week period of using the Clear Fear app, participants were sent the web-based questionnaires for a final time to complete, again including the user questionnaire for their specific feedback on using the app. In total, 2 to 3 attempts were made to follow up on participants if no response was received.

XSL•FO
RenderX

If no response was received, it was assumed that the participant had withdrawn from the study.

Ongoing Monitoring and Risk Assessment

Throughout the study, participants self-reported their anxiety levels at regular intervals. If they reported high levels of anxiety or emerging new symptoms, they were encouraged to consult a general practitioner or, depending on the severity, seek more urgent help. After the study, participants were given resources on maintaining well-being.

stem4, the owners of the app, had a clinical safety protocol and a hazard mitigation list that was managed and reviewed regularly with clear timelines of response. Their clinical safety officer was available to respond to inquiries daily.

Data Collection

A standardized questionnaire was developed to gather the eligible participants' demographic characteristics. To assess the Clear Fear app and its effects, participants were asked to complete 3 outcome measures to assess their emotional state or mood and behavior and the Clear Fear user questionnaire. Participants completed these 4 measures at stages 2 (post–app familiarization phase) and 3 (follow-up).

Outcome Assessments

In total, 3 self-report outcome measures were administered to participants to assess their symptoms of anxiety (main outcome) and depression and their behavior.

For anxiety, the 7-item Generalized Anxiety Disorder scale (GAD-7) was used, covering various areas, including social phobia, anxiety, and obsessive-compulsive disorder [14]. Total scores of 5, 10, and 15 indicate probable mild, moderate, and severe anxiety, respectively.

For depression, the 13-item self-report Mood and Feelings Questionnaire (MFQ) was used, with item ratings of *true* (score of 2), *sometimes true* (score of 1), or *not true* (score of 0) [15]. A total score of ≥12 may indicate depression.

For emotional and behavioral difficulties, the Strengths and Difficulties Questionnaire (SDQ) was used. This is a 25-item questionnaire to measure behavioral and emotional difficulties and assess psychological adjustment and mental health problems in children and young people [16].

The child and adolescent versions of the MFQ and the SDQ were used for participants aged between 16 and 17 years, and the adult version was used for those aged ≥18 years.

Clear Fear App Questionnaire

The Clear Fear app questionnaire, developed for the purposes of this study, was used to assess the Clear Fear app's safety, acceptability, satisfaction, and usability and included 21 items (of which 7 were closed ended, 9 were open ended for free-text responses, and 5 had Likert scales from 1-5).

For acceptability and user satisfaction, 10 items, including 8 open-ended questions, asked about what parts of the Clear Fear app were used, whether the content was useful, what features were liked or disliked and why, whether the app met the participants' expectations, and whether the app was unloaded.

For safety, 4 items asked whether the app helped reassure, reduce, or increase feelings of anxiety and whether the participants had to contact someone.

For usability, 7 items were related to how easy the Clear Fear app was to use, whether any technical issues were encountered, frequency of use, and at what time of the day it was used.

Data Analysis

Statistical analyses for all quantitative data were conducted using SPSS Statistics for Windows (version 29; IBM Corp). The demographic profile of participants and other quantitative data were described using frequencies, percentages, and means. The Shapiro-Wilk test of normality and visual assessment of *Q-Q* plots were used to check whether continuous variables for the outcome scores were normally distributed.

Mean scores derived from the outcome measures (GAD-7, MFQ, and SDQ) were compared between baseline and follow-up. Tests of significance were conducted using 2-tailed *t* tests or Wilcoxon signed rank tests (for continuous variables not distributed normally) to compare baseline and follow-up scores for all outcome and clinical measures. Standardized effect sizes were calculated using the Cohen *d* based on the mean change in outcome scores between baseline (stage 1) and follow-up (stage 3).

Qualitative data derived from open-ended questions were coded and entered into NVivo (version 10; QSR International) and analyzed using a thematic analysis of the text to identify themes and patterns in the data and how they addressed the main aims of the study [17]. Additional analyses (means and SDs) were conducted to examine any differences in total outcome scores for the GAD-7 and MFQ among age, gender, and ethnic groups. Differences among these groups were also explored through qualitative feedback on the Clear Fear app's acceptability using thematic analysis and categorization according to age, gender, and ethnic group.

Ethical Considerations

Ethics approval for this study was obtained from the London – City and East Research Ethics Committee (United Kingdom; 20/PR/0268), and all practices and procedures throughout the research process were in line with the stipulations of the Declaration of Helsinki.

This included that eligible participants be suitably informed of the study details, risks, and benefits via an information sheet before being asked to consent to the study. Only those who thus provided their informed consent went on to enter the study.

Participants' data were anonymized, collected and stored securely, and managed solely by the researchers. Participants were allocated a study ID number by a project administrator separate from the researchers.

Finally, participants were given a £20 (US $24.44) Amazon voucher as an honorarium upon completion of the final stage.

XSL•FO
RenderX

Results

Overview

Recruitment for the study took place between February 2021 and June 2022. A total of 48 young people completed the baseline measures. At the postfamiliarization phase (stage 2), 92% (44/48) of the participants completed all measures and the Clear Fear app user questionnaire. At the 8-week follow-up (stage 3), 77% (37/48) of the participants submitted all required responses, constituting a dropout rate of 23% (11/48).

Demographic Characteristics

The demographic characteristics of the sample are listed in Table 1. The mean age of the sample was 20.1 (SD 2.1) years, with 62% (30/48) aged between 18 and 25 years. The sample was predominantly female (37/48, 77%). Most were students, with a third (16/48, 33%) attending secondary school and 29% (14/48) at university. In terms of ethnicity, 67% (32/48) were White British, and 10% (5/48) were Black British.

Table 1. Characteristics of the participants at baseline (stage 1) of this uncontrolled pre– and post–follow-up study to evaluate the Clear Fear app's usability, acceptability, safety, and effectiveness in helping young people manage symptoms of anxiety using the principles of cognitive behavioral therapy over a 9-week period (N=48).

Characteristics	Values
Age (y), mean (SD; range)	20.1 (2.1; 17-24)
16-17, n (%)	18 (38)
18-25, n (%)	30 (62)
Sex, n (%)	
Female	37 (77)
Male	11 (23)
Marital status, n (%)	
Single	41 (85)
Married or cohabiting	4 (8)
Other long-term relationship	3 (6)
Ethnicity, n (%)	
Asian or Asian British	2 (4)
Black British	5 (10)
Mixed heritage	3 (6)
White British	32 (67)
White other	3 (6)
Other	3 (6)
Student status, n (%)	
Attending secondary school	16 (33)
Attending university	14 (29)
Attending sixth form or college	10 (21)
Not currently a student	8 (17)
Employment status, n (%)	
Student	33 (69)
Employed part time	6 (12)
Employed full time	5 (10)
Unemployed	2 (4)
Other	2 (4)

Baseline Scores

Table 2 lists the numbers and percentages of participants with probable symptoms of anxiety and depression and emotional and behavioral difficulties at baseline that would require further assessment. In total, 48% (23/48) of the participants reached the threshold for probable anxiety disorder as measured using the GAD-7, with 29% (14/48) finding it *very difficult* or *extremely difficult* to go about their daily routine or get along with others.

XSL•FO
RenderX

On the basis of the short MFQ measure, 56% (27/48) of the participants obtained a positive score for probable depression. A total of 75% (36/48) obtained *very high* on the SDQ total score, with 52% (25/48) of participants reporting their problems to be considerably distressing (*quite a lot* or *a great deal*). The ability to learn (23/48, 48%) and get along with others close to them (21/48, 44%) was also impacted for many.

Table 2. Numbers and percentages of participants meeting the threshold for probable symptoms of anxiety and depression and emotional and behavioral difficulties at baseline (stage 1) and follow-up (stage 3) following use of the Clear Fear app over a 9-week period to evaluate the app's usability, acceptability, safety, and effectiveness in helping young people manage symptoms of anxiety using the principles of cognitive behavioral therapy.

	Baseline (stage 1; n=48), n (%)	Follow-up (stage 3; n=37), n (%)
Generalized anxiety disorder (GAD-7[a])[b]	23 (48)	8 (22)
Mood and feelings (short MFQ[c])[d]	27 (56)	8 (22)
Emotional and behavioral difficulties (SDQ[e])[f]	36 (75)	23 (62)
Definite or severe problem	19 (40)	10 (27)
Problem lasting >12 months	19 (40)	14 (38)
Problem causing distress[g]	25 (52)	12 (32)
Difficulties getting along with close family or friends	21 (44)	14 (38)
Difficulties making and keeping friends	14 (29)	12 (32)
Difficulties with learning, studying, or work	23 (48)	17 (46)
Difficulties with hobbies and leisure	14 (29)	13 (35)
Difficulties impacting on others	19 (40)	6 (16)

[a]GAD-7: 7-item Generalized Anxiety Disorder scale.

[b]Total score of ≥10.

[c]MFQ: Mood and Feelings Questionnaire.

[d]Total score of ≥12.

[e]SQD: Strengths and Difficulties Questionnaire.

[f]Total score of 20 to 40 (*very high*).

[g]Those reporting quite a lot to a great deal.

Postfamiliarization Phase (Stage 2): Clear Fear App Use

A total of 92% (44/48) of the participants completed the self-report measures and questionnaires at stage 2. In total, 77% (34/44) of the participants used iOS (iPhone) as their operating system, with the remainder (10/44, 23%) using an Android phone. Only 5% (2/44) of the participants encountered technical issues when using the Clear Fear app (eg, poor display or phone freezing).

Participants used the Clear Fear app quite frequently; for many, it was several times a week (18/44, 41%) or every day (11/44, 25%). The evening was the preferred time for over half (24/44, 55%) of the participants.

Commonly used parts of the Clear Fear app included Anxiety Types (25/44, 57%), Clear Your Fear (23/44, 52%), and Self-Monitoring (20/44, 45%; Table 3). Parts of the Clear Fear app that were used less included Express Yourself (6/44, 14%), Laugh And Smile (6/44, 14%), information (6/44, 14%), Stop Overdoing Things (4/44, 9%), and Set an Exercise Goal (4/44, 9%).

XSL•FO
RenderX

Table 3. Participants' use of the Clear Fear app at the postfamiliarization phase (stage 2; 1 week following baseline) of this uncontrolled pre– and post–follow-up study to evaluate the app's usability, acceptability, safety, and effectiveness in helping young people manage symptoms of anxiety using the principles of cognitive behavioral therapy over a 9-week period (N=44).

	Participants, n (%)
Frequency of use	
More than once a day	2 (5)
Every day	11 (25)
Several times a week	18 (41)
Weekly	10 (23)
Less than once a week	3 (7)
Parts of the app used most	
Anxiety Types	25 (57)
Clear Your Fear	23 (52)
Self-Monitoring	20 (45)
Stay Calm	14 (32)
Dealing With Your Emotions	13 (30)
Managing Your Worries	12 (27)
Safety Net	12 (27)
Combat Negative Thoughts	10 (23)
Immediate Help	8 (18)
Managing Physical Responses to Anxiety	8 (18)

Follow-Up (Stage 3)

Usability and Acceptability

Overview

At follow-up (stage 3), 77% (37/48) of the participants completed the Clear Fear app user questionnaire. A total of 70% (26/37) of the participants reported using the Clear Fear app to address and manage feelings of anxiety, panic attacks, fear, stress, and worry. Participants were asked what they did before using the Clear Fear app to manage these symptoms. Most (23/37, 62%) used positive distractions, such as talking to family members, exercising, or other activities. The remaining participants either did nothing or used negative distractions such as "be in a daze."

The use pattern of the Clear Fear app was similar to that in stage 2 as the most used parts continued to be Anxiety Types (20/37, 54%), Clear Your Fear (24/37, 65%), and Self-Monitoring (16/37, 43%) and the least used parts were Express Yourself (4/37, 11%) and Stop Overdoing Things (4/37, 11%). Participants again mostly used the Clear Fear app either several times a week (17/37, 46%) or weekly (14/37, 38%) and in the evening (15/37, 41%). Fewer participants reported using the Clear Fear app every day at stage 3 compared to stage 2 (4/37, 11% vs 11/44, 25%, respectively). Most participants (31/37, 84%) found the Clear Fear app easy or very easy to use.

In total, 4 key themes were derived from the qualitative data (open-ended text responses). These included the Clear Fear app's interactive approach, active exercises, navigation, and app look and content.

Interactive Approach

When asked what features of the Clear Fear app were liked the most and why, a common theme appeared to include its interactive and direct approach while simultaneously providing information and guidance on how to manage feelings of anxiety and fear:

> *I liked the way the app asked what you were afraid of and then what you could do to battle that. I liked the way the questions made you face your fears and really think about it and not hide them.* [P21; female; student; aged 20 years]

This interactive approach gave one participant a sense of being more in control:

> *I like things where I actively did something with my anxiety, such as putting worries away, breathing exercises, clear fear, etc. The active involvement made me feel more in control.* [P59; female; volunteer; aged 24 years]

Active Exercises

The range of different activities, including breathing and writing exercises, to guide participants through their negative thoughts and feelings was well received. The immediate help feature and the available activities were considered helpful:

> *I liked using the immediate help feature and doing the breathing activities and panic attack feature because it helped to calm me down while having a panic attack. I also liked the safety net feature to*

XSL•FO
RenderX

> *remind myself what help I could get.* [P65; female; student; aged 18 years]

Encouraging participants to seek help when they needed it was another important feature of the Clear Fear app, together with offering practical and immediate strategies to manage their anxiety.

Navigation

> *The app is generally great and easy to use, and very helpful.* [P61; male; working part time; aged 21 years]

While most participants found the Clear Fear app easy to use and follow, a small minority (5/37, 14%) found the Clear Fear app difficult to use. A total of 60% (3/5) of these participants found the app difficult to navigate and asked for clearer instructions for finding activities:

> *The app was a little confusing to navigate on the home page, I found it confusing to go to the activities because there was no instruction as to how to do that.* [P32; female; student; aged 17 years]

> *Some features take a long time to find and I forgot which buttons I needed to press to get to them.* [P35; female; student; aged 17 years]

Another 40% (2/5) mentioned that they wanted to be able to make entries to the journal part of the app without having to carry out the activities first:

> *I couldn't find a way to always access the journal. My entries were after activities where it gives you the option to write. I'd love there to be a more accessible journal.* [P59; female; other; aged 22 years]

In total, 2 features of Clear Fear, Grit Box and Laugh and Smile, were considered less helpful by a small minority (4/37, 11%).

App Look and Content

While the Clear Fear app had content that participants found useful, some preferred not to have to read so much. One participant described expecting the Clear Fear app to have a better layout and look, with a less clinical appearance:

> *I expected the physical layout and colour schemes and fonts to look more calming and relaxing. The white backgrounds and lack of obvious clickable buttons made it look very clinical. To some extent the confusing layout made me feel more anxious when I used it.* [P29; female; student; aged 18 years]

Another participant thought that the Clear Fear app would have a means through which to communicate with someone or something:

> *I thought that there would be some sort of communication with something or someone like an area where you can talk to those cute creature things and have them say nice messages.* [P65; female; working full time; aged 21 years]

There was also a suggestion to have the Clear Fear app speak, perhaps to offer reassurance or include games:

> *I wanted more options of help than just writing everything down and having to read the help. E.g. videos to watch on the topic, music to calm thoughts and someone speaking the words.* [P30; female; student; aged 18 years]

Safety

Table 4 lists the numbers and percentages of participants concerning their responses to questions about the Clear Fear app's safety. Many participants found that the Clear Fear app reassured them either *a lot* (11/37, 30%) or *somewhat* (16/37, 43%). Similarly, the Clear Fear app was said to help reduce feelings of anxiety by *a lot* (13/37, 35%) or *somewhat* (14/37, 38%).

Table 4. Participants' responses regarding the safety of the Clear Fear app following 8 weeks of use (at stage 3) in this uncontrolled pre– and post–follow-up study to evaluate the app's usability, acceptability, safety, and effectiveness in helping young people manage symptoms of anxiety using the principles of cognitive behavioral therapy (N=37).

Questions and responses	Participants, n (%)
Did the Clear Fear app help reassure you when feeling anxious?	
Yes, a lot	11 (30)
Yes, somewhat	16 (43)
Yes, a little	5 (14)
No, not at all	2 (5)
Did the Clear Fear app reduce your feelings of anxiety?	
Yes, a lot	13 (35)
Yes, somewhat	14 (38)
Yes, a little	7 (19)
No, not at all	2 (5)
Did the Clear Fear app help you contact someone to speak to about your anxiety?	
Yes	21 (57)
No	16 (43)

XSL•FO
RenderX

Just over half (21/37, 57%) of the participants who responded to this question reported that the Clear Fear app helped them speak to someone about their anxiety. However, 43% (16/37) said that the Clear Fear app did not help them contact someone.

Qualitative User Feedback and Differences by Age, Gender, and Ethnicity

Some differences were found among age, gender, and ethnic groups from the qualitative data gathered from 3 user feedback questions concerning the acceptability of the Clear Fear app—which parts were most liked and disliked and what users were expecting from the app. Younger users (aged 16-18 years) preferred guided, calming features such as Clear Your Fear, whereas older users (aged 19-25 years) appreciated action-oriented, self-reflective tools such as the worry box and anxiety management features. The small sample of men focused on more immediate relief tools, such as breathing exercises and the Laugh and Smile feature, whereas female participants opted more for features that allow for emotional processing (eg, writing and emotional guidance). Regarding ethnicity, Black British users highlighted a focus on privacy and control, and Asian British participants valued ease of use and humor.

There were few differences between demographic groups for the small number of participants who noted parts of the app that they disliked (5/37, 14%). Younger users (aged 16-18 years) appeared less satisfied with navigating the app and features such as Safety Net or Laugh and Smile. Male respondents were less dissatisfied with the Clear Fear app, although they provided very brief responses. Black British users found the app's interface and navigation difficult and preferred customizable features, whereas Asian British participants provided limited negative feedback.

Managing emotions, anxiety, and coping strategies were the main expectations from the app across each of the age, gender, and ethnic groups.

Effectiveness of the Clear Fear App: Outcomes at Follow-Up (Stage 3)

Tables 5 and 6 list mean scores (and subscores for the SDQ), mean differences, standardized effect sizes, 95% CIs, and *P* values for each outcome measure at baseline and follow-up. At follow-up, 8 weeks after stage 2, statistically significant reductions were found in all mean total scores for symptoms of anxiety (GAD-7; t_{36}=2.6, 95% CI 0.41-3.53; *P*=.01) and depression (short MFQ; *z*=2.3; *P*=.02) and emotional and behavioral difficulties (SDQ; t_{47}=4.5, 95% CI 3.67-9.65; *P*<.001). The mean change in mean symptom scores revealed mostly medium to large standardized effect sizes.

Subscores for the SDQ also revealed significant decreases in emotional problems (t_{47}=4.5, 95% CI 1.14-2.99; *P*<.001), conduct problems (t_{47}=4.2, 95% CI 0.64-1.82; *P*<.001), hyperactivity (t_{47}=2.3, 95% CI 0.11-1.59; *P*=.02), and peer problems (t_{47}=1.6, 95% CI 0.39-1.61; *P*=.002).

Table 5. Baseline and follow-up (9 weeks later) mean comparisons, mean differences, and standardized effect sizes for participants' anxiety, depression, and behavioral and emotional symptom scores following their use of the Clear Fear app in this uncontrolled pre– and post–follow-up study to evaluate the app's usability, acceptability, safety, and effectiveness in helping young people manage symptoms of anxiety using the principles of cognitive behavioral therapy.

Outcome measure	Baseline score (stage 1; n=48), mean (SD)	Follow-up score (stage 3; n=37), mean (SD)	Mean difference (SD)	Standardized effect size, Cohen d_z
Probable anxiety disorder (GAD-7[a] total)	9.4 (4.5)	7.5 (3.7)	–1.9 (4.1)	0.46
Probable depression (short MFQ[b] total)	11.3 (5.4)	8.7 (5.1)	–2.6 (5.2)	0.50
Behavioral and emotional difficulties	23.3 (4.8)	16.6 (9.9)	–6.7 (8.3)	0.80
Emotional problems	5.9 (2.1)	3.9 (2.9)	–2.0 (2.5)	0.80
Conduct problems	2.7 (1.8)	1.5 (1.2)	–1.2 (1.5)	0.79
Hyperactivity	4.1 (1.9)	3.3 (2.7)	–0.8 (2.3)	0.34
Peer problems	3.1 (1.8)	2.1 (1.9)	–1.0 (1.8)	0.54

[a]GAD-7: 7-item Generalized Anxiety Disorder scale.

[b]MFQ: Mood and Feelings Questionnaire.

XSL•FO
RenderX

Table 6. Baseline and follow-up (9 weeks later) test of significance of the results for participants' anxiety, depression, and behavioral and emotional symptom scores following their use of the Clear Fear app in this uncontrolled pre– and post–follow-up study to evaluate the app's usability, acceptability, safety, and effectiveness in helping young people manage symptoms of anxiety using the principles of cognitive behavioral therapy.

Outcome measure	Test of significance	*P* value
Probable anxiety disorder (GAD-7[a] total)	t_{36}=2.6 (95% CI 0.41-3.53)	.01
Probable depression (short MFQ[b] total)	z=2.3[c]	.02
Behavioral and emotional difficulties (SDQ[d] total)	t_{47}=4.5 (95% CI 3.67-9.65)	<.001
Emotional problems	t_{47}=4.5 (95% CI 1.14-2.99)	<.001
Conduct problems	t_{47}=4.2 (95% CI 0.64-1.82)	<.001
Hyperactivity	t_{47}=2.3 (95% CI 0.11-1.59)	.02
Peer problems	t_{47}=1.6 (95% CI 0.39-1.61)	.002

[a]GAD-7: 7-item Generalized Anxiety Disorder scale.

[b]MFQ: Mood and Feelings Questionnaire.

[c]*Z* score from Wilcoxon signed rank test.

[d]SDQ: Strengths and Difficulties Questionnaire.

Differences in Anxiety and Depression Symptoms by Age, Gender, and Ethnicity

Few differences in the mean total outcome scores were found for the GAD-7 and MFQ when examined by age, gender, and ethnic group. Almost all groups showed improved anxiety and depression symptom scores at stage 3. The exceptions were Black British (5/37, 14%) and mixed-heritage (3/37, 8%) ethnic groups, whose GAD-7 mean scores increased slightly between stages 1 and 3 (from 13.0 to 13.3 and 11.3 to 13.0, respectively). Black British participants also had a slightly increased MFQ mean score at stage 3 (from 13.0 to 13.3). The greatest improvements in mean scores were a mean reduction of –5.0 (SD 9.8) in the GAD-7 for White other individuals (3/37, 8%) and mean reductions in the MFQ of –5.3 (SD 5.3) for male individuals (11/37, 30%), –6.0 (SD 11.3) for White other individuals (3/37, 8%), and –4.4 (SD 7.8) for those aged 18 to 25 years (25/37, 68%).

Discussion

Principal Findings

This paper details the development and evaluation of the Clear Fear app. The main findings on usability, acceptability, safety, and effectiveness were positive overall. The app was effective in reducing probable symptoms of anxiety and depression and emotional and behavioral difficulties. Therefore, this appears to support the hypothesis that a self-guided mobile cognitive behavioral app can improve anxiety symptoms and be acceptable to young people over a period of 9 weeks. However, the limitations of this study (outlined in a later section) should be noted when interpreting these findings.

A post hoc power analysis was conducted using G*Power (version 3.1.9.7) to check whether the sample size used in this study was large enough to facilitate sufficient statistical power for the analyses conducted [18]. The sample of 48 participants at baseline (with a medium effect size of 0.5 and an α of .05) and the sample at stage 3 (follow-up; n=37) provided 90% power and 80% power, respectively. Therefore, the study was adequately powered for the statistical analyses conducted.

Comparison With Prior Work

Usability

Certain parts of the Clear Fear app were frequently used both at stage 2 and stage 3, including Anxiety Types, Clear Your Fear, and Self-Monitoring, features that perhaps targeted participants' specific needs and symptoms. Over three-quarters of participants at stage 3 (31/37, 84%) reported the Clear Fear app to be easy to use. Patterns of Clear Fear app use remained broadly similar at stages 2 and 3, demonstrating consistent use throughout the study. Other similar studies have reported high use of mental health apps among young people during the first week, which often declines over time [19,20].

A very small minority of participants (5/48, 10%) found the Clear Fear app difficult to use. The app contains many different features and activities, perhaps too many, which, for a minority of participants, made it confusing to navigate and, subsequently, less useful and helpful when trying to find or retrieve parts. The information given and some of the requirements, such as writing down goals and ways to achieve them, may be overwhelming for some. Striking the right balance between ease of use and informativeness is important [21]. Therefore, based on the feedback from a small minority of study participants, future updates of the Clear Fear app could ensure easier navigation and direct access to features if required.

Popular and frequently used features of the Clear Fear app included Anxiety Types, Clear Your Fear, and Self-Monitoring. This appeared to show some participants' preference for dealing head-on with their feelings of anxiety and fear. The self-monitoring feature provides important emotional awareness to track and reflect on feelings and emotions to identify which promote positive or negative moods [19]. Previous research has also shown that self-monitoring mood is commonly included in many mobile mental health apps for young people and is the most liked and engaged with section [22,23].

XSL•FO
RenderX

Qualitative Feedback

The key themes derived from the open-ended responses to the Clear Fear app user questionnaire revealed which aspects participants found useful and what they liked and disliked. The Clear Fear app's approach to tackling feelings of anxiety and fear appeared beneficial. The exercises (breathing, meditation, and journal writing) within the Clear Fear app were also beneficial in helping some participants calm down from a panic attack. These exercises and other distraction strategies are important for managing anxiety. Some participants also suggested adding videos to watch, calming music, games, and someone speaking reassuring words, which could potentially enhance the Clear Fear app. Video games have been shown to be a useful therapeutic distraction, which helps address and treat symptoms of anxiety [24].

When selecting mobile mental health apps, young people have been found to prefer those that are accessible, secure, and evidence based [25]. The Clear Fear app meets these important criteria.

Safety

The Clear Fear app also appeared safe to use, with nearly three-quarters of participants (31/37, 84%) reporting that it offered a lot or some reassurance when feeling anxious or fearful. These reports were confirmed by the pre– and post–follow-up scores for the outcome measures. The "immediate help" feature was an important reminder for some participants to seek help when they needed it while in the meantime providing information and practical strategies for dealing with their anxiety and fear. Encouragingly, over half (21/37, 57%) of the participants reported that the Clear Fear app helped them speak to someone about their anxiety.

Effectiveness

At the beginning of this study, just under half (23/48, 48%) of the participants met the threshold for probable symptoms of anxiety disorder, over half (27/48, 56%) had probable depression, and three-quarters (36/48, 75%) obtained a score of *very high* for emotional and behavioral difficulties. By the end of the study, this had decreased to just under a quarter (8/37, 22%) of participants meeting the threshold for probable anxiety disorder, the same proportion (8/37, 22%) for probable depression, and just under two-thirds (23/37, 62%) scoring *very high* for emotional and behavioral difficulties. Consequently, the Clear Fear app appeared to significantly reduce scores for probable anxiety disorder (P=.01), probable depression (P=.02), and behavioral and emotional difficulties (P<.001) after 8 weeks of use. Moreover, at stage 3 of this study, far fewer participants met the threshold for symptoms of probable anxiety disorder and depression in particular.

These results add to the limited available evidence of the effectiveness of app interventions targeting anxiety and depression in young people and confirm previous findings that such apps, according to one review, can produce small to medium positive effect sizes [26]. Despite being distinct conditions, there is a strong relationship between anxiety and depression, which can coexist in young people [27]. This may explain the improvement in depression scores found in this study even though the Clear Fear app is designed for the management of anxiety symptoms. This improvement may be due to several reasons. Anxiety and depression often share common risk and maintaining factors, and addressing anxiety may help with some of the factors that are common to depression. Lowered anxiety may also help improve a person's self-esteem and confidence; enhance social connection; and help with focus, purpose, and application to work, all contributing to enhanced mood.

Differences by Age, Gender, and Ethnic Group

Generally, the Clear Fear app revealed few differences across the age, gender, and ethnic groups examined. However, the differences found highlighted the need to improve the app's navigation and refine less popular features for Black British and younger users. Interestingly, Black British users had slightly increased scores at follow-up for anxiety and depression symptoms, which underlines the need for more culturally appropriate content.

Limitations

The small sample of participants prevented further statistical tests to compare differences among age, gender, and ethnic groups. However, such differences were further examined in terms of mean outcomes and qualitative feedback (a strength of this study) to understand the user experiences of the Clear Fear app according to age, gender, and ethnicity.

The Clear Fear app is suitable for ages of ≥11 years; however, participants in the study were aged between 16 and 25 years, making the study findings generalizable only to this age group. Children and adolescents may respond differently to the Clear Fear app layouts, interaction styles, and content than those aged 16 to 25 years.

It is difficult to know what biases may have been introduced during recruitment, for example, how participants self-selected to take part in the study, the different avenues of information (eg, how participants found out about the study and the Clear Fear app), and their motivation to take part. The offer of an incentive (a shopping voucher) could have led to some form of selection bias regarding those who took part in the study. However, the attrition rates were low, with over three-quarters of participants (37/48, 77%) remaining in the study at follow-up, preventing selective biases associated with high dropout rates.

The follow-up of 8 weeks was a relatively short period, and a lack of a control group also presents further limitations in the study design. However, this exploratory uncontrolled pretest-posttest design provides promising initial findings; some understanding of how the Clear Fear app was used; its acceptability; and potential effects on reducing symptoms of anxiety, depression, and emotional and behavioral difficulties. This study was also adequately powered to conduct significance tests on the main outcome after carrying out a post hoc G*Power calculation.

Future Research

Future studies could benefit from a more robust evaluation design with a larger and more diverse sample. A randomized controlled trial with one or more comparison groups (including

XSL•FO
RenderX

a control group) would help determine causality and control for potential confounders and other external influencing factors. This combined with recruiting participants from a wider representation, including children and adolescents (between the ages of 11 and 15 years), could extend the generalizability of this study's findings. Incorporating clinician-administered outcome measures and extending the follow-up period beyond 8 weeks would enhance understanding of the Clear Fear app's effectiveness and user engagement over time (eg, over 3, 6, and ≥12 months). Future studies could also help determine how and why the Clear Fear app helped with symptoms of depression.

Future studies could also examine the Clear Fear app's effectiveness across different cultural contexts to reveal how cultural norms and societal pressures impact user engagement and outcomes. This could help identify any necessary adaptations to the app's content or functionality to make it culturally sensitive and relevant.

Similarly, future studies could explore how the Clear Fear app can be integrated with specialist mental health services. This could involve studying the app's role in supporting young adults on treatment waiting lists through enhancing engagement, complementing face-to-face therapy, being administered as a stand-alone intervention, and helping reduce increased workloads for health care providers. Given the need for ongoing support following discharge and in preventing relapse, the Clear Fear app may also have a function in helping young adults stay well.

In addition, further research could focus on understanding the perceptions and acceptance of digital health tools among both patients generally and clinicians in various health care settings, which may vary widely based on local practices and resources. This research will help refine the development of digital interventions, ensuring that they are both accessible and effective for diverse populations and integrated within existing health care systems.

Conclusions

The findings indicate that the Clear Fear app is an accessible, easy-to-use, safe, and effective early intervention for young people experiencing mild to moderate anxiety and depression. Therefore, this study appears to support the hypothesis that a self-guided mobile cognitive behavioral app can improve anxiety symptoms and be acceptable to young people over a period of 9 weeks. However, it is important to note this study's limitations, which may restrict the validity and generalizability of these findings, particularly until further research is conducted with larger samples of young people and specifically with children aged between 11 and 15 years to continue to assess the Clear Fear app's functioning and use. It is also important to examine its effectiveness over a longer period, using more robust and possibly alternative methodologies to better understand its use and compliance. The quantitative and qualitative feedback from participants also provide important information for further improvements to the Clear Fear app through future updates.

Acknowledgments

The authors warmly thank all the young people who participated in this study and provided their feedback on the Clear Fear app. They are also deeply grateful to the young people who gave their time to provide feedback on the Clear Fear app during its development. The authors are immensely grateful for the support from Comic Relief and the Paul Hamlyn Foundation's Tech for Good fund for funding the development of the Clear Fear app and to the Nominet #RESET mental health program for funding its evaluation.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to conditions outlined as part of the participants' informed consent, as well as the data compliance regulations of stem4; however, they may be available from the corresponding author on reasonable request.

Authors' Contributions

Conception, app development, and coproduction were carried out by NK. Study recruitment and coordination were conducted by KD and RE. Study planning, design, and data collection and analysis were carried out by CS and NU. Manuscript preparation was carried out by all authors.

Conflicts of Interest

None declared.

References

1. Newlove-Delgado T, Marcheselli F, Williams T, Mandalia D, Davis J, McManus S, et al. Mental health of children and young people in England, 2022. NHS Digital. 2022. URL: https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2022-follow-up-to-the-2017-survey [accessed 2023-12-12]
2. Young people's well-being in the UK: 2020. Office for National Statistics. 2020 Oct 02. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/bulletins/youngpeopleswellbeingintheuk/2020 [accessed 2024-04-29]

XSL•FO
RenderX

3. Garratt K, Kirk-Wade E, Long R. Children and young people's mental health: policy and services (England). House of Commons Library. 2024 Jan. URL: https://commonslibrary.parliament.uk/research-briefings/cbp-7196/ [accessed 2024-11-15]
4. Grimm F, Alcock B, Butler JE, Crespo RF, Davies A, Peytrignetet S, et al. Briefing: Improving children and young people's mental health services. The Health Foundation. 2022 Jul. URL: https://www.health.org.uk/sites/default/files/upload/publications/2022/CYPMHbriefing_Web_Final.pdf [accessed 2023-12-12]
5. Hiley C. UK mobile phone statistics, 2023. Uswitch. 2023 Feb 01. URL: https://www.uswitch.com/mobiles/studies/mobile-statistics/ [accessed 2023-12-12]
6. Leech T, Dorstyn D, Taylor A, Wenjing L. Corrigendum to "mental health apps for adolescents and young adults: a systematic review of randomised controlled trials". Child Youth Serv Rev 2021 Sep;128:106184. [doi: 10.1016/j.childyouth.2021.106184]
7. Bear HA, Ayala Nunes L, Ramos G, Manchanda T, Fernandes B, Chabursky S, et al. The acceptability, engagement, and feasibility of mental health apps for marginalized and underserved young people: systematic review and qualitative study. J Med Internet Res 2024 Jul 30;26:e48964 [FREE Full text] [doi: 10.2196/48964] [Medline: 39078699]
8. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018 Dec;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]
9. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
10. Bergin AD, Perez Vallejos E, Davies EB, Daley D, Ford T, Harold G, et al. Preventive digital mental health interventions for children and young people: a review of the design and reporting of research. NPJ Digit Med 2020;3:133-020 [FREE Full text] [doi: 10.1038/s41746-020-00339-7] [Medline: 33083568]
11. Goldin PR, Gross JJ. Effects of mindfulness-based stress reduction (MBSR) on emotion regulation in social anxiety disorder. Emotion 2010 Feb;10(1):83-91 [FREE Full text] [doi: 10.1037/a0018441] [Medline: 20141305]
12. Sucala M, Cuijpers P, Muench F, Cardo R, Soflau R, Dobrean A, et al. Anxiety: there is an app for that. A systematic review of anxiety apps. Depress Anxiety 2017 Jun;34(6):518-525. [doi: 10.1002/da.22654] [Medline: 28504859]
13. Bry LJ, Chou T, Miguel E, Comer JS. Consumer smartphone apps marketed for child and adolescent anxiety: a systematic review and content analysis. Behav Ther 2018 Mar;49(2):249-261 [FREE Full text] [doi: 10.1016/j.beth.2017.07.008] [Medline: 29530263]
14. Chorpita BF, Moffitt CE, Gray J. Psychometric properties of the revised child anxiety and depression scale in a clinical sample. Behav Res Ther 2005 Mar;43(3):309-322. [doi: 10.1016/j.brat.2004.02.004] [Medline: 15680928]
15. Daviss WB, Birmaher B, Melhem NA, Axelson DA, Michaels SM, Brent DA. Criterion validity of the mood and feelings questionnaire for depressive episodes in clinic and non-clinic subjects. J Child Psychol Psychiatry 2006 Sep;47(9):927-934. [doi: 10.1111/j.1469-7610.2006.01646.x] [Medline: 16930387]
16. Goodman R. The strengths and difficulties questionnaire: a research note. J Child Psychol Psychiatry 2006;38(5):581-586. [doi: 10.1111/j.1469-7610.1997.tb01545.x] [Medline: 9255702]
17. Braun V, Clarke V. Thematic Analysis: A Practical Guide. Thousand Oaks, CA: Sage Publications; 2021.
18. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
19. Pramana G, Parmanto B, Kendall PC, Silk JS. The SmartCAT: an m-health platform for ecological momentary intervention in child anxiety treatment. Telemed J E Health 2014 May;20(5):419-427 [FREE Full text] [doi: 10.1089/tmj.2013.0214] [Medline: 24579913]
20. Reid SC, Kauer SD, Khor AS, Hearps SJ, Sanci LA, Kennedy AD, et al. Using a mobile phone application in youth mental health - an evaluation study. Aust Fam Physician 2012 Sep;41(9):711-714. [Medline: 22962650]
21. Garrido S, Cheers D, Boydell K, Nguyen QV, Schubert E, Dunne L, et al. Young people's response to six smartphone apps for anxiety and depression: focus group study. JMIR Ment Health 2019 Oct 02;6(10):e14385 [FREE Full text] [doi: 10.2196/14385] [Medline: 31579023]
22. Kenny R, Dooley B, Fitzgerald A. Feasibility of "copesmart": a telemental health app for adolescents. JMIR Ment Health 2015 Aug 10;2(3):e22 [FREE Full text] [doi: 10.2196/mental.4370] [Medline: 26552425]
23. Melbye S, Kessing LV, Bardram JE, Faurholt-Jepsen M. Smartphone-based self-monitoring, treatment, and automatically generated data in children, adolescents, and young adults with psychiatric disorders: systematic review. JMIR Ment Health 2020 Oct 29;7(10):e17453 [FREE Full text] [doi: 10.2196/17453] [Medline: 33118950]
24. Kowal M, Conroy E, Ramsbottom N, Smithies T, Toth A, Campbell M. Gaming your mental health: a narrative review on mitigating symptoms of depression and anxiety using commercial video games. JMIR Serious Games 2021 Jun 16;9(2):e26575 [FREE Full text] [doi: 10.2196/26575] [Medline: 34132648]
25. Kabacińska K, McLeod K, MacKenzie A, Vu K, Cianfrone M, Tugwell A, et al. What criteria are young people using to select mobile mental health applications? A nominal group study. Digit Health 2022 May 29;8:20552076221102775 [FREE Full text] [doi: 10.1177/20552076221102775] [Medline: 35663239]

XSL•FO
RenderX

26. Lecomte T, Potvin S, Corbière M, Guay S, Samson C, Cloutier B, et al. Mobile apps for mental health issues: meta-review of meta-analyses. JMIR Mhealth Uhealth 2020 May 29;8(5):e17458 [FREE Full text] [doi: 10.2196/17458] [Medline: 32348289]
27. Axelson DA, Birmaher B. Relation between anxiety and depressive disorders in childhood and adolescence. Depress Anxiety 2001;14(2):67-78. [doi: 10.1002/da.1048] [Medline: 11668659]

Abbreviations

CBT: cognitive behavioral therapy
GAD-7: 7-item Generalized Anxiety Disorder scale
MFQ: Mood and Feelings Questionnaire
SDQ: Strengths and Difficulties Questionnaire

Edited by A Mavragani; submitted 18.12.23; peer-reviewed by S Mitra, J Abbas, A Futterman Collier; comments to author 07.10.24; revised version received 30.11.24; accepted 24.12.24; published 28.02.25.

Please cite as:
Samele C, Urquia N, Edwards R, Donnell K, Krause N
Evaluation of the Clear Fear Smartphone App for Young People Experiencing Anxiety: Uncontrolled Pre– and Post–Follow-Up Study
JMIR Form Res 2025;9:e55603
URL: https://formative.jmir.org/2025/1/e55603
doi:10.2196/55603
PMID:40053764

©Chiara Samele, Norman Urquia, Rachel Edwards, Katie Donnell, Nihara Krause. Originally published in JMIR Formative Research (https://formative.jmir.org), 28.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Digital Program for Daily Life Management With Endometriosis: Pilot Cohort Study on Symptoms and Quality of Life Among Participants

Zélia Breton[1,2]; Emilie Stern[3,4], MSc; Mathilde Pinault[1], PhD; Delphine Lhuillery[5,6], MD; Erick Petit[7,8], MD; Pierre Panel[9], MD; Maïa Alexaline[1], PhD

[1]Lyv Healthcare, Nantes, France
[2]Université Paris-Saclay, Gustave Roussy, Inserm, 94805, Villejuif, France
[3]GHU Paris, Psychiatry & Neurosciences, Paris, France
[4]Laboratoire de Psychopathologie et Processus de Santé, Université Paris Cité, Boulogne-Billancourt, France
[5]Paris Saint-Joseph Hospital Group, EndoSud Ile-de-France, Paris, France
[6]Clinique Oudinot, Paris, France
[7]Department of Radiology, Fondation Hôpital St Joseph, Paris, France
[8]Centre de l'Endométriose, RESENDO (Réseau Ville-Hopital Endometriose), Paris, France
[9]Department of Gynecology-Obstetrics, Centre Hospitalier de Versailles, Le Chesnay-Rocquencourt, France

Corresponding Author:
Maïa Alexaline, PhD
Lyv Healthcare
6 rue Edouard Nignon
Nantes, 44300
France
Phone: 33 650208680
Email: maia@lyv.app

Abstract

Background: After experiencing symptoms for an average of 7 years before diagnosis, patients with endometriosis are usually left with more questions than answers about managing their symptoms in the absence of a cure. To help women with endometriosis after their diagnosis, we developed a digital program combining user research, evidence-based medicine, and clinical expertise. Structured around cognitive behavioral therapy and the quality of life metrics from the Endometriosis Health Profile score, the program was designed to guide participants for 3 months.

Objective: This cohort study was designed to measure the impact of a digital health program on the symptoms and quality of life levels of women with endometriosis.

Methods: In total, 63% (92/146) of the participants were included in the pilot study, recruited either free of charge through employer health insurance or via individual direct access. A control group of 404 women with endometriosis who did not follow the program, recruited through social media and mailing campaigns, was sampled (n=149, 36.9%) according to initial pain levels to ensure a similar pain profile to participants. Questionnaires assessing quality of life and symptom levels were emailed to both groups at baseline and 3 months. Descriptive statistics and statistical tests were used to analyze intragroup and intergroup differences, with Cohen *d* measuring effect sizes for significant results.

Results: Over 3 months, participants showed substantial improvements in global symptom burden, general pain level, anxiety, depression, dysmenorrhea, dysuria, chronic fatigue, neuropathic pain, and endo belly. These improvements were significantly different from the control group for global symptom burden (participants: mean –0.7, SD 1.6; controls: mean –0.3, SD 1.3; *P*=.048; small effect size), anxiety (participants: mean –1.1, SD 2.8; controls: mean 0.2, SD 2.5; *P*<.001; medium effect size), depression (participants: mean –0.9, SD 2.5; controls: mean 0.0, SD 3.1; *P*=.04; small effect size), neuropathic pain (participants: mean –1.0, SD 2.7; controls: mean –0.1, SD 2.6; *P*=.004; small effect size), and endo belly (participants: mean –0.9, SD 2.5; controls: mean –0.3, SD 2.4; *P*=.03; small effect size). Participants' quality of life improved between baseline and 3 months and significantly differed from that of the control group for the core part of the Endometriosis Health Profile-5 (participants: mean –5.9, SD 21.0; controls: mean 1.0, SD 14.8; *P*=.03; small effect size) and the EQ-5D (participants: mean 0.1, SD 0.1; controls: mean –0.0, SD

XSL•FO
RenderX

0.1; *P*=.001; medium effect size). Perceived knowledge of endometriosis was significantly greater at 3 months among participants compared to the control group (*P*<.001).

Conclusions: This study's results suggest that a digital health program providing medical and scientific information about endometriosis and multidisciplinary self-management tools may be useful to reduce global symptom burden, anxiety, depression, neuropathic pain, and endo belly while improving knowledge on endometriosis and quality of life among participants.

(JMIR Form Res 2025;9:e58262) doi:10.2196/58262

KEYWORDS

digital program; endometriosis; integrative therapies; quality of life; nonpharmacological intervention; daily life; adenomyosis; lesions; women; digital; France; case control; digital health; pilot study; control group; global symptom; burden; depression; neuropathic pain; chronic; multidisciplinary; web based; mobile health; mHealth; intervention

Introduction

Background

Endometriosis is a chronic, inflammatory disease affecting an estimated 1 in 10 women of childbearing age. It is defined by the presence of endometrial-like tissue outside the uterine cavity [1]. Manifestations of the disease are mainly painful (eg, dysmenorrhea, dyspareunia, dyschezia, and dysuria), with symptoms substantially impacting the quality of life (QOL) of those affected [2]. There is no definitive cure, and available solutions aim to reduce symptoms. The initial solutions are generally hormonal, which are not always well tolerated by patients [1], consistent with their fertility goals, or effective in relieving pain [3]. While surgery was once widely recommended to patients with endometriosis, given the frequent disease recurrence [4], it is now increasingly restricted to selected patients for whom it can be relevant in European countries [5]. Thus, living with endometriosis means learning to deal with the symptoms, using tools other than or in complement to hormonal therapy. After experiencing symptoms for an average of 7 years before diagnosis [6], the lack of an adapted care pathway for symptom management is a crucial unmet need of the patients [7]. Various nonpharmacological interventions have been studied to help with symptom management, such as dietary changes, physical activity, sex therapy, and mind-body interventions, which seem pertinent for the daily management of endometriosis [8-12]. Digital tools may offer a solution to enhance accessibility to and observance of nonpharmacological interventions toward endometriosis symptom reduction.

Prior Work

Developing a program to improve QOL requires the construction of content corresponding to what can modulate the QOL of women with endometriosis. Numerous studies to assess changes in patients' QOL have been using the Endometriosis Health Profile (EHP) score as a reference for almost 20 years [13]. The EHP is a standardized tool for measuring QOL. Initially developed with 30 items (EHP-30) [14], a shortened version with 5 items (ie, EHP-5) was also validated [15]. The following EHP components were taken into account in developing the programs: pain, control and powerlessness, emotional well-being, social support, self-image, work life, infertility and children management, sexual relationships, and relation to the medical profession.

Cognitive behavioral therapy (CBT) is a well-supported treatment for various disorders [16]. CBT's goal is to help individuals better understand the relationship and interaction between their emotions, cognitions (received ideas), and behavior to be able to manage their symptoms and quit the vicious cycle that contributing factors may create through behavior change (develop coping mechanisms), cognitive restructuring (restructure maladaptive thoughts), and emotion regulation (relaxation and acceptance) [17-19]. It was found to improve the QOL for patients with chronic pain conditions [20-23], especially by encouraging a problem-solving attitude in the patients receiving the treatment [24]. Internet-based CBT programs have been proven effective for a variety of disorders [25-29]. When considering CBT-based programs for chronic pain [18,30], the following components are found: psychoeducation on pain mechanisms and biopsychosocial model, goal settings, CBT-based behavior skills for managing pain (eg, pleasant activities and sleep hygiene), emotion regulation and mind-body interventions (eg, relaxation methods and mindfulness), cognitive restructuring, and long-term action plan for maintenance and anticipation of obstacles.

Goal of This Study

A digital solution for endometriosis symptom management was developed using a CBT approach based on EHP items. The digital support solution focused on the scientific and medical state of the art around 5 nonpharmacological interventions (disease education including pain mechanism, diet, adapted physical activity, well-being and mental health, and sexual health). The *School of Endo* digital program was developed and made available in France to help women with endometriosis after their diagnosis.

This cohort study was designed to measure the impact of a digital health program on the symptoms and QOL levels of women with endometriosis. The study also aimed to provide insights into the value of a digital program for the day-to-day management of endometriosis and to address the lack of real-life data studies for digital support in endometriosis.

Methods

Development of a Digital Health Program

To develop a digital health program that adequately addresses the needs of patients with endometriosis, several steps were followed, where product development and scientific research

XSL•FO
RenderX

complemented each other [31]. As the first step, 120 semistructured interviews and >100 web-based questionnaires with patients and health professionals were conducted to understand the needs of women with endometriosis. According to this user research, the primary need expressed by women with endometriosis was to find help with the daily management of symptoms, especially pain. A small exploratory preliminary program was developed and tested by women to complement the user research. A scientific rationale and advice from medical experts completed the user research. User research, evidence-based medicine, and clinical expertise were combined to enable the development of the program [32].

The structure of the program's content was based on CBT and the QOL items of the EHP score, and participants were advised to follow it for 3 months. The program was accessible from November 2022 for women with endometriosis aged >18 years, either at no fee through employer health insurance (*Mutuelle Générale de l'Education Nationale*) or individual direct payment. Program participants were invited to join the pilot study on a voluntary basis.

This program sought to provide evidence-based information and tools to empower patients on symptom management through a wide range of content: videos, exercises, written content, live sessions, quizzes, and a community-based platform. A total of 12 endometriosis experts were involved in the construction of the program, constituting a multidisciplinary team.

Scientific rationale highlighted some factors of interest for the program. In an Australian study on 484 women with endometriosis, dietary changes were reported among the most effective strategies in terms of reducing self-reported pain associated with the disease [8]. Optimizing nutrition is a nonpharmacological intervention that can be used to manage the symptoms of endometriosis. The positive effects of physical activity have been documented for various health conditions, including mental health (improved cognitive function, enhanced QOL, improved sleep, and reduced signs of anxiety and depression) and chronic pain [12,33,34]. According to current research, psychological interventions and mind-body approaches are promising for improving QOL and alleviating pain, anxiety, depression, stress, and fatigue in women with endometriosis [9,10]. The impact of endometriosis on sex life harms patients' self-esteem. Sex therapy and psychological interventions can be very beneficial during treatment to improve QOL [11].

The program addressed every aspect of the disease through 5 sections: disease education including pain mechanism, diet, adapted physical activity, well-being and mental health, and sexual health. It started with a section on knowledge of the disease, with an introduction to the program and the multidisciplinary approach. Psychoeducation was focused on defining the disorder, its causes, and its impacts. It helped the participant generate a model of their disorder and pain to obtain a representation of a functional analysis of the factors perpetuating the symptoms. The first part of the program concluded with educating the participants on setting goals and initiating change and provided tools to help them set their goals.

CBT content is included throughout the program. Each section mentioned subsequently contains education, advice, and tools created by health care professionals in the field.

The diet section covered step-by-step behavior change through good mealtime habits, from breakfast to dinner, with the help of calendars to fill in. This section provided information on why and how to make dietary changes with endometriosis, particularly on the role of diet in inflammation and digestive symptoms. A section dedicated to identifying inflammatory foods was also available to participants.

The sections dedicated to physical activity presented different ways to stay active while experiencing symptoms and suggested various methods, including yoga, Pilates, physiotherapy exercises, and daily habits, to move the body. A 30-minute yoga session and a Pilates session were available each week. These sessions focused mainly on pelvic mobility. Physiotherapy complemented the physical activity with ventral breathing exercises and a focus on the perineum and abdominal muscles.

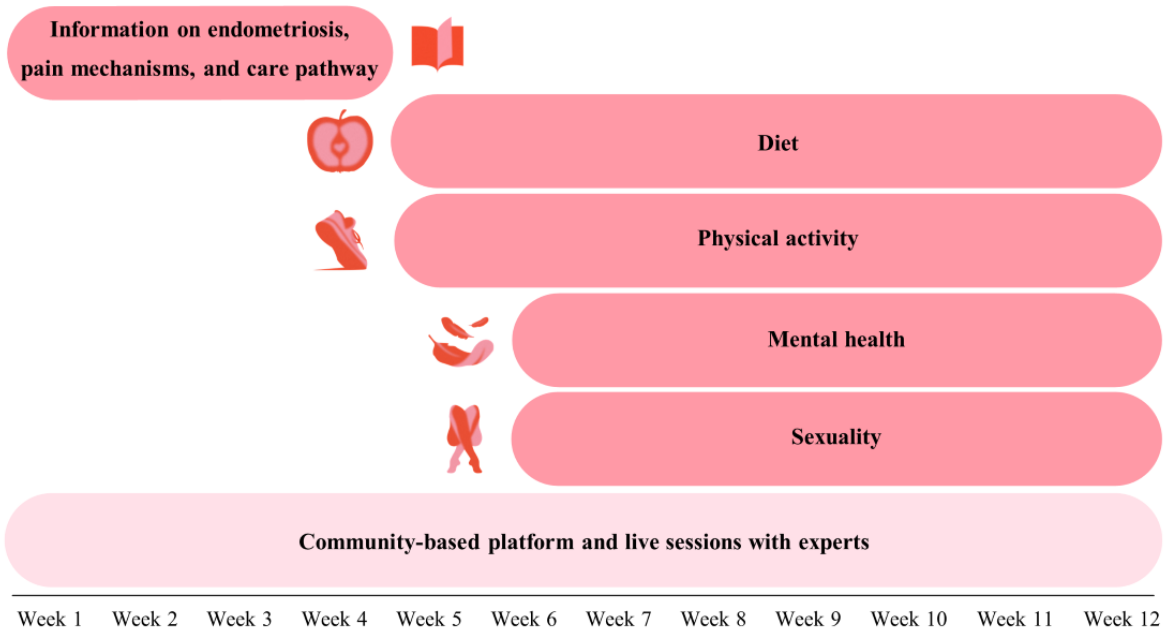
The mental health section focused mainly on education and prevention to help with emotional regulation. Relaxation practices, sophrology exercises, Beck columns, and a weekly diary to initiate behavior changes were proposed.

The sexuality and intimacy section featured physiotherapy exercises, sex therapy tools, and advice to rediscover pleasure and self-confidence in sexual relations.

The program allowed participants to create their toolbox and action plan. Section contents were delivered weekly for 3 months and were available for a further 3 months (Figure 1).

XSL•FO
RenderX

Figure 1. Sections contained in the 3-month digital program offered to participants with endometriosis on which this cohort study is based.



Ethical Considerations

A study protocol was established in compliance with French regulations, reviewed by the Lyv Healthcare company's ethics committee, and filed with the French public structure for studies on health data: Health Data Hub (N° F20221114165253). Participants answered the questionnaires on a voluntary basis after being informed about the research project objectives and were provided with contact information to exercise their rights. At any time, participants had the possibility to opt out. All collected data were pseudonymized and stored in compliance with health data security standards (Health Data Hosting certification, h*ébergeurs de données de santé*). No personal information was shared with third parties, and all data analysis was performed on deidentified datasets. No compensation was provided to either program participants or the control group for their participation in the study.

Impact of the Digital Health Program

Questionnaires

Our study compared 2 groups, distinct in terms of exposure to the digital program, both having participants with endometriosis. The study is considered a cohort study because the allocation of respondents to the exposure criterion was not controlled for the research participant, and there was no prospective assignment. Two web-based questionnaires were sent to program participants via email links and reminders on the program's community-based platform. The first questionnaire was sent at the time of inclusion in the program (time point 0 [T0]) and the second one after 3 months (T0+3 mo), with 2 weeks to submit their answers. Participants completed the questionnaires on their own, requiring approximately 15 minutes to complete. Questionnaires were structured into 2 parts, including a main part (53 questions) and a bonus part (24 questions), explaining differences in the sample size for some outcomes. Both questionnaires were electronically tested before diffusion. A back button allowed respondents to review and modify their answers, while IP addresses were used to ensure that each participant was a unique visitor, with only the first submission used for analysis.

Population Sample

Women were free to choose whether to take part in the program, with no obligation to complete it in full. Program participants were given the freedom to answer the questionnaires. This pilot study included participants who completed the main part entirely at baseline and 3 months.

Only women who reported being diagnosed with endometriosis could answer the questionnaires. The diagnosis could be clinical, imagery based, or surgical. To record the effect of active use of the program on participants, the threshold of following at least half of the proposed content types was retained. As a result, among the participants, only women who declared having tested >50% of the proposed content type in the program were included.

The terms woman and women are used in this paper. It should be noted that people with a uterus may or may not identify with these terms and experience endometriosis. Sex and gender were not criteria for inclusion in the study.

Control Group

A control group of women with endometriosis who did not follow the program was recruited via social media and a health insurance provider email campaign (*Mutuelle Générale de l'Education Nationale*). The women in the control group were volunteers, motivated solely by the desire to contribute to advancing research, as no incentives, prizes, or rewards were offered for their participation. The control group received the

XSL•FO
RenderX

same questionnaires as program participants during the same periods to eliminate external effects on measured parameters.

The control group was sampled on the initial pain level to have a similar profile between them and program participants. Women were selected from the control group by stratifying the sample to reproduce the distribution of pain levels among program participants. QOL answers were not selected for this stratification as these questions were part of bonus questionnaires. A random sample of women from the control group was selected so that the same proportion was represented for each modality of the participants' general pain level. The proportional sample was thus selected, reducing the number of women in the control group to 149.

Study Outcomes Measures

Numeric Rating Scales

Numeric Rating Scales (NRSs) are 11-point scales, ranging from 0 (no pain or symptom) to 10 (intense pain or symptom), primarily used to measure symptom intensity. NRS evaluated the level of overall pain, anxiety, depression, dysmenorrhea, dyspareunia, dyschezia, dysuria, chronic pelvic pain, gastrointestinal disorders, chronic fatigue, neuropathic pain, and endo belly. Endo belly is a common term used by patients with endometriosis to describe the cyclic bloating of the abdomen [35] as "uncomfortable or painful, often accompanied by a sensation of abdominal fullness," and which "often forces women with endometriosis to wear loose clothing" [36]. Endo belly affects >80% of women with gastrointestinal symptoms due to endometriosis [37].

Women without menstruation did not respond to the dysmenorrhea NRS, and women without sexual activities did not respond to the dyspareunia NRS, which explains differences in sample size for these outcomes.

An average of each of these NRSs was used to determine a global symptom burden score, ranging from 0 to 10.

EHP-5 Scale

EHP-5 is a QOL scale specific to endometriosis. A main score is calculated out of 100 (100=poor QOL and 0=good QOL), based on 5 items with 5 modalities (scored as 0, 25, 50, 75, and 100). The main score is the average of the scores for the 5 items. A modular score is calculated out of 100 (100=worst QOL and 0=best QOL), based on 6 optional 5-modality items (scored as 0, 25, 50, 75, and 100). The score is the average of the scores for the 6 items.

EQ-5D Scale

EQ-5D is a QOL scale nonspecific to endometriosis. It comprises 5 items with 5 modalities scored from 1 to 5. The concatenation of the 5 scores gives a combination. Each combination ranging from 11111 to 55555 has a numerical equivalent between –0.53 and 1. EQ-5D score ranges from –0.53 (poor QOL) to 1 (good QOL).

Statistical Analysis

Symptoms were considered to have evolved (improved or deteriorated) if scores had changed by >2 points compared with T0 [38]. QOL was considered to have evolved (improved or deteriorated) if the score had changed by >15% compared with T0 [39]. Otherwise, outcomes were considered stable.

The minimum sample size was calculated through a feasibility study power analysis for outcomes between program participants and the control group. This feasibility study was carried out on a preliminary program among 55 participants to identify ways to improve the program and observe symptoms and QOL levels. Global pain on the NRS was assumed by a mean (SD) of 5.9 (2.3). Assuming a difference of at least 2 points, a power (1–β) of 0.80, α=.05, and using a 2-sided 2-sample test, 11 program participants and 11 participants from the control group would be required to detect group differences. QOL was measured on the EHP-5, assumed by a mean (SD) of 53.6 (19.0). Assuming a change in score of at least 15%, a power of 0.80, α=.05, and using a 2-sided 2-sample test, 13 program participants and 13 participants from the control group would be required for detecting between-group differences. On the basis of power analyses, it was planned to include answers from at least 13 program participants and 13 from the control group.

Descriptive statistics were carried out (number of participants, percentage, mean, and SD). Statistical tests used to measure the significance of intragroup evolution were the chi-square test and the Fisher exact test for discrete variables, the Wilcoxon signed rank test for paired and nonparametric continuous variables, and the paired 2-tailed Student *t* test for parametric continuous variables. Statistical tests used to measure between-group differences (participants vs controls) were the chi-square test and the Fisher exact test for discrete variables, the Mann-Whitney *U* test for unpaired and nonparametric continuous variables, and the unpaired 2-tailed Student *t* test for parametric continuous variables. Furthermore, we applied the Benjamini-Hochberg method to adjust for multiple comparisons. This method controls the false discovery rate, offering a balance between minimizing false positives and maintaining statistical power. Cohen *d* tests were performed to measure effect size when the result was statistically significant. If Cohen $d>|0.2|$, the effect size was considered small, medium if Cohen $d>|0.5|$, and large if Cohen $d>|0.8|$. Each outcome was studied separately. A sensitivity analysis was conducted to note the impact of program follow-up according to baseline QOL level. For this, linear regression models and interaction tests were used with the EHP-5 core QOL level (good QOL: 0 to 32, medium QOL: 33 to 65, and low QOL: 66 to 100). Statistical analysis was performed with SAS software (version 9.4; SAS Institute).

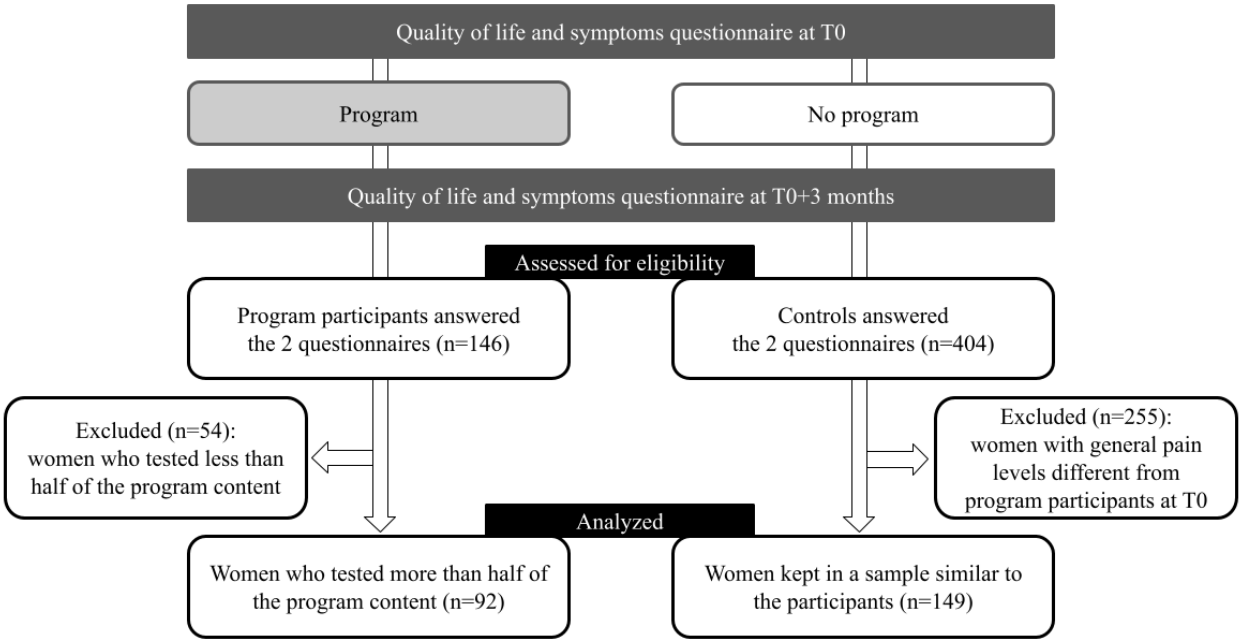
Results

Population Description

The pilot study was based on the responses to the main questionnaire part from 92 program participants diagnosed with endometriosis and 149 women diagnosed with endometriosis who did not follow the program, constituting the control group (Figure 2).

XSL•FO
RenderX

Figure 2. Cohort study flow diagram: cohort study responders were divided according to their participation in the 3-month digital program for women with endometriosis. T0: time point 0.



Respondents to the questionnaires took part in this study on a voluntary basis. Table 1 presents the characteristics of the program participants and those of the control group at baseline. On average, participants were aged 36.7 (SD 6.8) years and were diagnosed with the disease when they were aged 32.5 (SD 6.7) years. In total, 26% (24/92) of the participants were experiencing or had experienced infertility problems. There were no substantial differences in socioeconomic characteristics or pain levels at the time of inclusion compared to the control group. Participants had a slightly poorer QOL at the time of inclusion than those in the control group (P=.04) and were more often on a continuous hormonal pill (P=.046). The baseline symptom levels and QOL of program participants are provided in Multimedia Appendix 1.

Analyses were based on the answers given by participants to the initial questionnaire and the one at 3 months.

XSL•FO
RenderX

Table 1. Sociodemographic, clinical, pain-related, and quality of life baseline characteristics of program participants and those in the control group with endometriosis (N=241).

	Program participants (n=92)	Control group (n=149)	*P* value
Sociodemographic characteristics			
Age (y), mean (SD)	36.7 (6.8)	36.6 (7.2)	.91
Single status, n (%)	23 (25)	29 (19.5)	.64
Does not feel surrounded and feels very alone, n (%)	6 (6.5)	11 (7.4)	.92
Perceived financial situation (0=very insecure to 10=very comfortable), mean (SD)	5.4 (1.7)	5.4 (1.9)	.73
Number of children, mean (SD)	0.7 (1.0)	0.8 (1.0)	.34
Number of pregnancies, mean (SD)	1.1 (1.4)	1.3 (1.5)	.11
Diagnosis of endometriosis and treatment			
Experience of the pathway to diagnosis (0=very poor to 10=very good), mean (SD)	2.7 (2.4)	3.4 (2.7)	.09
Age at diagnosis (y), mean (SD)	32.5 (6.7)	31.6 (6.8)	.21
Age at first symptoms (y), mean (SD)	18.7 (7.3)	19.9 (8.7)	.56
Time between first set of symptoms and diagnosis, mean (SD)	13.5 (8.9)	11.5 (8.6)	.10
Hormonal treatment, n (%)	53 (57.6)	65 (43.6)	.046
Symptoms and quality of life			
Experienced infertility problems, n (%)	24 (26.1)	47 (31.5)	.37
Dysmenorrhea level (0=no pain to 10=worst pain), mean (SD)	7.3 (1.8)	6.5 (2.3)	.12
Global symptom burden (0=no symptom to 10=worst symptom intensity), mean (SD)	5.3 (1.6)	5.0 (1.6)	.17
Proportion of global symptom burden ≥8, n (%)	5 (5.4)	4 (2.7)	.31
Quality of life: EHP-5[a] (0=best to 100=worst), mean (SD)	53.8 (20.4)[b]	45.2 (20.3)[c]	.04
Proportion of quality of life: EHP-5 ≥80, n (%)	11 (12)	9 (6)	.15

[a]EHP-5: Endometriosis Health Profile-5.

[b]n=39 program participants.

[c]n=65 individuals from the control group.

Impact on Endometriosis Symptoms

Initially, the participants' most severe symptom was chronic fatigue (mean 7.5, SD 2.1), followed by dysmenorrhea (mean 7.3, SD 1.8) and anxiety (mean 6.5, SD 2.2; Table 2).

Overall, there was a tendency for all symptoms to improve during the 3 months of following the program (Table 3). For some symptoms, the evolution was not significant compared with the control group: dyspareunia, dyschezia, chronic pelvic pain, gastrointestinal disorders, overall pain, dysuria, and chronic fatigue. Over 3 months, the global symptom burden, general level of pain, anxiety, depression, dysmenorrhea, dysuria, chronic fatigue, neuropathic pain, and endo belly improved significantly for the program participants. These improvements were significantly different compared to the control group for global symptom burden (*P*=.048; small effect size), anxiety (*P*<.001; medium effect size), depression (*P*=.04; small effect size), neuropathic pain (*P*=.004; small effect size), and endo belly (*P*=.03; small effect size).

XSL•FO
RenderX

Table 2. Outcome evaluation at baseline and 3 months for program participants versus those in the control group with endometriosis.

Health outcome	Program participants (n=92), n (%)	Control group (n=149), n (%)	T0[a]				T0+3 mo			
			Program participants, mean (SD)	Control group, mean (SD)	Participants vs control group		Program participants, mean (SD)	Control group, mean (SD)	Participants vs control group	
					P value	Cohen *d*			*P* value	Cohen *d*
Global symptom burden	90 (97.8)	147 (98.7)	5.3 (1.6)	5.0 (1.6)	.17	0.1[b]	4.6 (1.6)	4.7 (1.8)	.75	–0.0[b]
Overall pain	90 (97.8)	147 (98.7)	6.2 (2.5)	6.0 (2.4)	.68	0.1[b]	5.5 (2.6)	5.3 (2.6)	.66	0.1[b]
Anxiety	90 (97.8)	147 (98.7)	6.5 (2.2)	5.9 (2.7)	.12	0.2[c]	5.5 (2.9)	6.1 (2.7)	.12	–0.2[c]
Depression	89 (96.7)	147 (98.7)	4.0 (3.0)	4.2 (3.2)	.87	–0.0[b]	3.3 (3.1)	4.2 (3.3)	.05	–0.3[c]
Dysmenorrhea	35 (38)	74 (49.7)	7.3 (1.8)	6.5 (2.3)	.12	0.3[c]	5.6 (2.6)	5.8 (2.6)	.54	–0.1[b]
Dyspareunia	43 (46.7)	84 (56.4)	3.9 (2.4)	4.1 (2.8)	.90	–0.0[b]	3.1 (2.5)	3.5 (2.7)	.47	–0.2[c]
Dyschezia	90 (97.8)	147 (98.7)	3.7 (3.2)	4.2 (2.8)	.19	–0.2[c]	3.3 (2.7)	3.6 (2.9)	.60	–0.1[b]
Dysuria	90 (97.8)	147 (98.7)	2.2 (2.9)	1.8 (2.5)	.36	0.2[c]	1.6 (2.0)	1.6 (2.4)	.35	0.0[b]
Chronic pelvic pain	90 (97.8)	147 (98.7)	5.2 (2.8)	4.8 (2.7)	.22	0.2[b]	4.8 (2.8)	4.5 (2.7)	.48	0.1[b]
Gastrointestinal disorders	90 (97.8)	147 (98.7)	5.9 (2.7)	5.3 (2.7)	.10	0.2[c]	5.7 (2.3)	5.1 (2.8)	.13	0.2[c]
Chronic fatigue	90 (97.8)	147 (98.7)	7.5 (2.1)	7.0 (2.2)	.09	0.2[c]	6.8 (2.3)	6.8 (2.3)	.77	–0.0[b]
Neuropathic pain	90 (97.8)	147 (98.7)	5.4 (3.1)	4.6 (3.2)	.09	0.2[c]	4.3 (2.9)	4.5 (3.0)	.47	–0.1[b]
Endo belly	90 (97.8)	147 (98.7)	6.2 (2.5)	5.7 (2.7)	.09	0.2[c]	5.4 (2.7)	5.3 (2.7)	.92	0.0[b]
QOL[d]: EHP-5[e] core part	39 (42.4)	65 (43.6)	53.8 (20.4)	45.2 (20.3)	.04	0.0[b]	47.9 (19.4)	46.2 (21.5)	.88	0.0[b]
QOL: EHP-5, modular part	39 (42.4)	64 (43)	34.9 (13.0)	28.7 (18.4)	.05	0.4[c]	26.7 (15.4)	26.4 (18.8)	.59	0.0[b]
QOL: EQ-5D score	33 (35.9)	80 (53.7)	0.8 (0.1)	0.8 (0.2)	.06	–0.2[c]	0.9 (0.1)	0.8 (0.2)	.44	0.0[b]

[a]T0: time point 0.

[b]No effect size.

[c]Small effect size.

[d]QOL: quality of life.

[e]EHP-5: Endometriosis Health Profile-5.

XSL•FO
RenderX

Table 3. Continuous outcomes evolution between baseline and 3 months for program participants versus the control group with endometriosis.

Health outcome	Program participants, mean (SD)	Within program participants		Control group, mean (SD)	Within control group		Program participants vs control group		Benjamini-Hochberg threshold
		P value	Cohen *d*[a]		*P* value	Cohen *d*	*P* value	Cohen *d*	
Global symptom burden	–0.7 (1.6)	*<.001*[b]	*0.4*[c]	–0.3 (1.3)	*.004*	*0.2*[c]	*.048*	*–0.3*[c]	0.022
Overall pain	–0.7 (2.9)	*.02*	*0.3*[c]	–0.7 (2.5)	*<.001*	*0.3*[c]	.84	—[d]	0.050
Anxiety	–1.1 (2.8)	*<.001*	*0.4*[c]	0.2 (2.5)	.51	—	*<.001*	*–0.5*[e]	*0.003*
Depression	–0.9 (2.5)	*.002*	*0.3*[c]	0.0 (3.1)	.92	—	*.04*	*–0.3*[c]	0.019
Dysmenorrhea	–1.9 (2.8)	*<.001*	*0.7*[e]	–0.7 (2.4)	*.01*	*0.3*[c]	.05	—	0.025
Dyspareunia	–0.5 (2.8)	.28	—	–0.7 (2.2)	*.007*	*0.3*[c]	.69	—	0.044
Dyschezia	–0.4 (2.3)	.16	—	–0.6 (2.7)	*.004*	*0.2*[c]	.47	—	0.038
Dysuria	–0.5 (2.3)	*.045*	*0.2*[c]	–0.1 (2.3)	.41	—	.22	—	0.034
Chronic pelvic pain	–0.5 (3.0)	.15	—	–0.3 (2.6)	.29	—	.53	—	0.041
Gastrointestinal disorders	–0.2 (2.7)	.36	—	–0.2 (2.3)	.31	—	.76	—	0.047
Chronic fatigue	–0.7 (2.5)	*.02*	*0.3*[c]	–0.2 (2.3)	.29	—	.14	—	0.031
Neuropathic pain	–1.0 (2.7)	*<.001*	*0.4*[c]	–0.1 (2.6)	.58	—	*.004*	*–0.4*[c]	*0.009*
Endo belly	–0.9 (2.5)	*.002*	*0.3*[c]	–0.3 (2.4)	.21	—	*.03*	*–0.2*[c]	0.013
QOL[f]: EHP-5[g] core part	–5.9 (21.0)	.09	—	1. (14.8)	.59	—	*.03*	*–0.4*[c]	0.016
QOL: EHP-5, modular part	–8.2 (16.9)	*.004*	*–0.5*[e]	–2.5 (11.1)	.06	—	.10	—	0.028
QOL: EQ-5D score	0.1 (0.1)	*.01*	*0.5*[e]	–0.0 (0.1)	.08	—	*.001*	*0.7*[e]	*0.006*

[a]The effect size was considered small if Cohen *d* >|0.2|, medium if Cohen *d* >|0.5|, and large if Cohen *d* >|0.8|.

[b]Values in italics indicate significant *P* values.

[c]Small effect size.

[d]Not applicable.

[e]Medium effect size.

[f]QOL: quality of life.

[g]EHP-5: Endometriosis Health Profile-5.

When looking at symptom evolution (improvement, stability, or deterioration), the distribution was not significantly different between the 2 groups for dyspareunia, dyschezia, dysuria, chronic pelvic pain, gastrointestinal disorders, and chronic fatigue (Table 4). Global symptom burden improved in 20% (18/90) of the program participants versus 6.1% (9/147) in the control group (*P*=.003). Anxiety levels improved in 42% (38/90) and deteriorated in 11% (10/90) of the program participants versus 26.5% and 28.6%, respectively, in the control group (*P*=.002). Depression deteriorated less in program participants (9/89, 10%) than in the control group (41/147, 27.9%; *P*=.003). Neuropathic pain improved more in program participants (37/90, 41%) than in the control group (34/147, 23.1%; *P*=.02), as did endo belly (37/90, 41% vs 36/147, 24.5%; *P*=.03).

Actively following the digital program for 3 months was associated with a significant improvement in global symptom burden, anxiety, depression, neuropathic pain, and endo belly perception among program participants when compared to the control group.

XSL•FO
RenderX

Table 4. Evolution of symptoms between baseline and 3 months for program participants versus the control group with endometriosis.

Health outcome	Program participants			Control group			Program participants vs control group, *P* value	Benjamini-Hochberg threshold
	Improvement	Stable	Deterioration	Improvement	Stable	Deterioration		
Global symptom burden (n=90 participants; n=147 controls)							*.003* [a]	*0.009*
Participants, n (%)	18 (20)	68 (75.6)	4 (4.4)	9 (6.1)	133 (90.5)	5 (3.4)		
Points, mean (SD)	–3.1 (0.9)	–0.3 (1.0)	2.6 (0.7)	–2.9 (0.9)	–0.2 (1.0)	2.5 (0.5)		
Overall pain (n=90 participants; n=147 controls)							*.09*	0.029
Participants, n (%)	34 (37.8)	39 (43.3)	17 (18.9)	43 (29.3)	85 (57.8)	19 (12.9)		
Points, mean (SD)	–3.8 (1.7)	0.1 (0.8)	3.3 (1.3)	–3.7 (1.6)	–0.1 (0.7)	3.2 (1.5)		
Anxiety (n=90 participants; n=147 controls)							*.002*	*0.006*
Participants, n (%)	38 (42.2)	42 (46.7)	10 (11.1)	39 (26.5)	66 (44.9)	42 (28.6)		
Points, mean (SD)	–3.6 (1.8)	–0.1 (0.7)	4.2 (1.9)	–2.9 (1.2)	0.2 (0.7)	3.0 (1.4)		
Depression (n=89 participants; n=147 controls)							*.003*	*0.012*
Participants, n (%)	28 (31.5)	52 (58.4)	9 (10.1)	42 (28.6)	64 (43.5)	41 (27.9)		
Points, mean (SD)	–3.7 (1.7)	0.0 (0.8)	3.0 (1.5)	–3.5 (1.6)	0.0 (0.7)	3.7 (2.1)		
Dysmenorrhea (n=35 participants; n=74 controls)							.07	0.024
Participants, n (%)	19 (54.3)	14 (40)	2 (5.7)	24 (32.4)	39 (52.7)	11 (14.9)		
Points, mean (SD)	–4.0 (2.1)	0.1 (0.7)	3.0 (0.0)	–3.3 (1.6)	–0.2 (0.8)	2.9 (1.1)		
Dyspareunia (n=43 participants; n=84 controls)							.49	0.038
Participants, n (%)	13 (30.2)	21 (48.8)	9 (20.9)	25 (29.8)	48 (57.1)	11 (13.1)		
Points, mean (SD)	–3.9 (1.6)	0.0 (0.8)	3.1 (1.5)	–3.4 (1.6)	0.0 (0.7)	2.3 (0.6)		
Dyschezia (n=90 participants; n=147 controls)							.63	0.047
Participants, n (%)	23 (25.6)	51 (56.7)	16 (17.8)	46 (31.3)	75 (51)	26 (17.7)		
Points, mean (SD)	–3.3 (1.5)	–0.1 (0.7)	3.1 (1.1)	–3.5 (1.7)	–0.1 (0.8)	3.3 (2.1)		
Dysuria (n=90 participants; n=147 controls)							.38	0.035
Participants, n (%)	24 (26.7)	55 (61.1)	11 (12.2)	28 (19.1)	97 (66)	22 (15)		
Points, mean (SD)	–3.3 (1.7)	–0.1 (0.5)	3.5 (1.4)	–3.4 (1.5)	0.0 (0.5)	3.6 (1.6)		
Chronic pelvic pain (n=90 participants; n=147 controls)							.30	0.032
Participants, n (%)	33 (36.7)	33 (36.7)	24 (26.7)	44 (29.9)	69 (46.9)	34 (23.1)		
Points, mean (SD)	–3.7 (1.6)	0.0 (0.7)	3.3 (1.2)	–3.2 (1.6)	0.0 (0.8)	3.1 (1.2)		
Gastrointestinal disorders (n=90 participants; n=147 controls)							.90	0.050
Participants, n (%)	27 (30)	42 (46.7)	21 (23.3)	42 (28.6)	73 (49.7)	32 (21.8)		
Points, mean (SD)	–3.2 (1.3)	–0.1 (0.8)	3.4 (1.6)	–2.8 (1.1)	0.0 (0.8)	3.0 (1.3)		
Chronic fatigue (n=90 participants; n=147 controls)							.49	0.041
Participants, n (%)	29 (32.2)	46 (51.1)	15 (16.7)	37 (25.2)	81 (55.1)	29 (19.7)		
Points, mean (SD)	–3.4 (1.6)	–0.1 (0.8)	2.9 (1.4)	–3.1 (1.4)	0.0 (0.7)	2.9 (1.5)		
Neuropathic pain (n=90 participants; n=147 controls)							*.02*	0.015
Participants, n (%)	37 (41.1)	39 (43.3)	14 (15.6)	34 (23.1)	84 (57.1)	29 (19.7)		
Points, mean (SD)	–3.5 (1.6)	–0.2 (0.7)	3.1 (1.7)	–3.4 (1.6)	0.0 (0.7)	3.6 (1.8)		
Endo belly (n=90 participants; n=147 controls)							*.03*	0.021
Participants, n (%)	37 (41.1)	37 (41.1)	16 (17.8)	36 (24.5)	81 (55.1)	30 (20.4)		
Points, mean (SD)	–3.4 (1.4)	–0.1 (0.9)	3.3 (1.8)	–3.5 (1.8)	0.0 (0.7)	2.7 (1.1)		

XSL•FO
RenderX

[a]Italicization indicates significant *P* values.

Impact on QOL

With all 3 scores used (EHP-5 core part, EHP-5 modular part, and EQ-5D), active program participants showed an improvement in their QOL after 3 months on the program (Table 2). This evolution observed between 0 and 3 months significantly differed from the control group for the core part of the EHP-5 (*P*=.03; small effect size) and the EQ-5D (*P*=.001; medium effect size; Table 3). Therefore, the use of the program was associated with an improvement of QOL in participants with endometriosis.

When looking at the types of evolution (improvement, stability, or deterioration) in QOL, an improvement of the EQ-5D and the EHP-5 core part was observed in 2 and 3 times more women in the program participants than in the control group, respectively (Table 5). However, the difference in the types of evolution was not significant between the 2 groups for EQ-5D (*P*=.58) and the core part of EHP-5 (*P*=.07). A significant difference in the types of evolution between the 2 groups was observed for the modular part of the EHP-5 (*P*=.02), with an improvement in 31% (12/39) of the program participants versus 11% (7/64) of the controls.

Table 5. Evolution of quality of life (QOL) score items and perceived knowledge between baseline and 3 months for program participants versus the control group with endometriosis.

Health outcome	Program participants			Control group			Program participants vs control group, *P* value	Benjamini-Hochberg threshold
	Improvement	Stable	Deterioration	Improvement	Stable	Deterioration		
QOL: EHP-5[a] core part (n=39 participants; n=65 controls)							.07	0.026
Participants, n (%)	16 (41)	17 (43.6)	6 (15.4)	13 (20)	40 (61.5)	12 (18.5)		
Points, mean (SD)	–25.0 (8.2)	–0.9 (6.9)	30.8 (11.6)	–18.1 (3.8)	0.5 (6.6)	23.3 (11.3)		
QOL: EHP-5 modular part (n=39 participants; n=64 controls)							*.02* [b]	0.018
Participants, n (%)	12 (30.8)	24 (61.5)	3 (7.7)	7 (10.9)	54 (84.4)	3 (4.7)		
Points, mean (SD)	–28.1 (9.3)	–2.4 (6.1)	25.0 (8.3)	–23.2 (3.3)	–1.3 (6.9)	23.6 (2.4)		
QOL: EQ-5D score (n=33 participants; n=80 controls)							.58	0.044
Participants, n (%)	2 (6.1)	31 (93.9)	0 (0)	2 (2.5)	78 (97.5)	0 (0)		
Points, mean (SD)	0.4 (0.1)	0.0 (0.1)	__[c]	0.4 (0.1)	–0.0 (0.1)	—		
Perceived knowledge (n=90 participants; n=148 controls), n (%)	47 (52.2)	41 (45.6)	2 (2.2)	21 (14.2)	119 (80.4)	8 (5.4)	*<.001*	*0.003*

[a]EHP-5: Endometriosis Health Profile-5.

[b]Values in italics indicate significant *P* values.

[c]Not applicable.

To better understand these evolutions in QOL, we focused on the items that make up the QOL scores (Table 6). There was a significant difference at baseline (*P*=.03; small effect size) in the core part of the EHP-5 score on the item concerning control and powerlessness. Program participants were more affected than those in the control group. This significant difference disappeared at 3 months. Besides, after 3 months, there was a significant improvement in the sexual relationship item (*P*=.02; small effect size) in the modular part of the EHP-5 score, which was not observed in the control group. Participants were less often worried about pain during sexual intercourse after 3 months on the program than before starting the program. Finally, the usual activity item in the EQ-5D score improved nearly significantly for the program participants (*P*=.05; small effect size) after 3 months, which was not the case in the control group, with participants feeling less limited in their usual activities after 3 months on the program.

The benefits of a digital program also depend on how it is used; in this study, it was found that the greater the frequency of use or the greater the diversity of sections consulted, the greater the improvement in symptoms or QOL (Multimedia Appendices 2 and 3). The results did not show any statistically significant correlations between the viewing of action-oriented content and the program's effectiveness (Multimedia Appendix 3).

XSL•FO
RenderX

Table 6. Quality of life (QOL) score items evaluation at baseline and 3 months for program participants versus the control group with endometriosis.

Health outcome	Program participants				Control group				Program participants vs control group			
	T0[a], mean (SD)	T0 +3 mo, mean (SD)	*P* value	Cohen *d*	T0, mean (SD)	T0 +3 mo, mean (SD)	*P* value	Cohen *d*	T0, *P* value	T0, Cohen *d*	T0+3 mo, *P* value	T0+3 mo, Cohen *d*
QOL: EHP-5[b] core part												
Pain	1.4 (1.1)	1.2 (1.0)	.17	—[c]	1.2 (1.0)	1.2 (1.0)	.95	—	.25	—	.99	—
Control and powerlessness	2.5 (1.4)	2.1 (1.2)	.06	—	1.9 (1.2)	1.9 (1.2)	.93	—	.03	0.4[d]	.62	—
Emotional well-being	2.5 (1.0)	2.3 (0.9)	.39	—	2.3 (1.0)	2.3 (1.1)	.59	—	.79	—	.90	—
Social support	2.3 (1.0)	2.1 (0.9)	.19	—	1.9 (1.2)	2.0 (1.1)	.64	—	.15	—	.87	—
Self-image	2.2 (1.3)	2.0 (1.4)	.46	—	1.8 (1.2)	1.9 (1.4)	.48	—	.13	—	.80	—
QOL: EHP-5 modular part												
Work life	1.2 (1.1)	0.9 (1.1)	.37	—	0.9 (1.2)	0.8 (1.3)	.46	—	.13	—	.21	—
Taking care of children	2.2 (1.2)	2.1 (1.1)	.19	—	1.4 (1.1)	1.3 (1.2)	.99	—	.03	0.7[e]	.04	0.7[e]
Sexual relationships	2.2 (1.5)	1.7 (1.4)	.02	–0.4[d]	1.7 (1.4)	1.7 (1.4)	.16	—	.15	—	.99	—
Medical profession	1.3 (1.2)	0.6 (0.8)	.06	—	1.2 (1.3)	1.1 (1.4)	.66	—	.81	—	.21	—
Treatment	2.4 (1.0)	2.2 (1.2)	.25	—	2.1 (1.3)	1.9 (1.4)	.24	—	.32	—	.36	—
Infertility	2.2 (1.6)	2.5 (1.4)	.66	—	1.9 (1.7)	2.1 (1.7)	.68	—	.45	—	.52	—
QOL: EQ-5D score												
Mobility	1.4 (0.7)	1.2 (0.5)	.11	—	1.5 (0.8)	1.5 (0.8)	.62	—	.87	—	.06	—
Self-care	1.1 (0.2)	1.0 (0.2)	.99	—	1.1 (0.4)	1.1 (0.4)	.99	—	.61	—	.17	—
Usual activities	2.2 (0.9)	1.9 (0.9)	.05	–0.4[d]	1.9 (0.9)	1.9 (0.8)	.63	—	.05	—	.92	—
Pain or discomfort	2.9 (0.8)	2.5 (0.7)	.06	—	2.5 (1.0)	2.5 (1.0)	.64	—	.06	—	.47	—
Anxiety or depression	2.6 (0.9)	2.4 (1.0)	.19	—	2.4 (0.9)	2.6 (1.0)	.04	0.2[d]	.28	—	.52	—
VAS[f] (0=worst to 100=best QOL)	58.8 (17.1)	63.0 (15.7)	.23	—	63.4 (20.7)	62.8 (19.2)	.79	—	.13	—	.99	—

[a]T0: time point 0.

[b]EHP-5: Endometriosis Health Profile-5.

[c]Not applicable.

[d]Small size effect.

[e]Medium size effect.

[f]VAS: visual analog scale.

Impact on Perceived Knowledge of Endometriosis

Initially, there was no significant difference in perceived knowledge between program participants and the control group (*P*=.74; Table 7). In total, 22% (20/92) of the program participants felt that they knew little about the disease, 71% (65/92) had good knowledge, and 8% (7/92) considered themselves experts (vs 34/149, 22.8%; 99/149, 66.4%; and 16/149, 10.7%, respectively for the control group). At 3 months, none of the program participants considered themselves as knowing little about the disease (0/92, 0%); 64% (59/92) considered themselves as having good knowledge, and 36% (33/92) considered themselves experts (vs 25/149, 16.8%; 104/149, 69.8%, and 20/149, 13.4%, respectively, for the control group). The perceived knowledge of endometriosis was significantly different at 3 months between the 2 groups (*P*<.001; Table 5). As expected, the program seemed to improve perceived knowledge of endometriosis.

XSL•FO
RenderX

Table 7. Perceived knowledge on endometriosis at baseline and 3 months for program participants versus the control group. The significance for program participants versus control group is P=.74 at T0 and P<.001 at T0+3 months.

Perceived knowledge on endometriosis	Program participants (n=92), n (%)		Control group (n=149), n (%)	
	T0[a]	T0+3 mo	T0	T0+3 mo
I know little	20 (21.7)	0 (0)	34 (22.8)	25 (16.8)
I have good knowledge, but I am not an expert	65 (70.7)	59 (64.1)	99 (66.4)	104 (69.8)
I consider myself an expert	7 (7.6)	33 (35.9)	16 (10.7)	20 (13.4)

[a]T0: time point 0.

Multiple Comparisons and Sensitivity Analysis

To address the potential for type I errors due to multiple testing in this pilot analysis, we applied the Benjamini-Hochberg method to adjust significance thresholds while maintaining statistical power. This approach controls the false discovery rate and is particularly suited for exploratory analyses.

Table 1 illustrates that the QOL levels, as assessed by the EHP-5 core part, significantly differed between the 2 groups at baseline (*P*=.04). However, the proportion of participants reporting poor QOL was similar in both groups (*P*=.15). To explore the potential influence of these baseline differences on the evolution of outcomes, we conducted interaction analyses using linear regression models, as detailed in Multimedia Appendix 4. These analyses revealed that baseline QOL differences primarily affected the evolution of dyschezia. Specifically, dyschezia showed more significant improvement in program participants compared to the control group, particularly among women with good baseline QOL. In addition, baseline QOL levels influenced changes in the EHP core part score at 3 months. This score improved significantly among program participants with a distinct pattern compared to the control group, especially for women with good QOL at baseline. Conversely, improvements were more pronounced in the control group than in the program participants for women with moderate or poor QOL at baseline.

Discussion

Principal Findings

This pilot study of 92 program participants and 149 individuals in the control group shows that following a 3-month digital health program for the self-management of endometriosis symptoms is associated with a significant reduction in endometriosis-related symptoms (anxiety, depression, neuropathic pain, and endo belly perception), a reduction of the global symptom burden, and a significant improvement in the participants' knowledge of endometriosis. Furthermore, following the program seemed to be associated with an improvement in QOL. Combining CBT approaches and EHP-focused programs in a digital tool has proved useful in inducing symptom relief and a better QOL.

Comparison With Prior Work

Previous studies have shown that patients often use nonpharmacological interventions to cope with their symptoms [8,40]. However, only one short-term study has examined the value of combining these nonpharmacological interventions as part of a multidisciplinary approach for QOL with endometriosis [41], and a recent publication tested the application over 3 months [42]. Instead, most published studies analyzed the benefits of such interventions individually. These studies, among other things, have underpinned the development of our proposed program, which offers patients various tools and solutions to manage their symptoms.

As confirmed by our results, CBT has been found to be helpful in managing pain and improving the scores of depression and QOL in patients with endometriosis [43-45]. New studies are also beginning to be conducted on the benefits of psychological and relaxation interventions for patients with endometriosis, thus far demonstrating improvement in pain [46], stress [47], anxiety, depression, and QOL [48,49]. The improvement of mental health–related symptoms in our results might be in part due to the psychological and relaxation approaches included in our program. Given the link between mental health and QOL in women with endometriosis [50], future research should explore the use of more comprehensive scales for anxiety and depression. Furthermore, nutrition and physical activity have been studied separately and proved to have an impact on endometriosis symptoms, stress, anxiety, and QOL [51-56]. Finally, although sexual issues are a prevalent symptom, sex education for women with endometriosis has only been assessed in one study, where an improvement in sexual function and quality of sexual life was observed [57]. Interestingly, in our study, the EHP-5 score for the sexual intercourse apprehension item decreased for participants, notwithstanding persisting dyspareunia. This highlights a change in the perception of sexual anxiety after 3 months, despite the persistence of pain. Digital programs for the management of endometriosis should study the long-term effect of sex education content on dyspareunia, both in terms of symptoms and perception.

While the efficacy and relevance of each intervention for the management of endometriosis symptoms have thus been confirmed by previous research, a multidisciplinary approach is in line with current guidelines; therefore, it should be encouraged for women with endometriosis and more generally for those who experience chronic pain [58,59]. Future studies should seek to determine the optimal combination of nonpharmacological interventions according to patient profiles to find the right balance between diet, physical activity, stress management, sleep, and environmental enrichment on top of their medical care [60,61].

In terms of methodology, our study assessed each symptom using numerical scales to simplify user responses and limit attrition rates. For QOL, we used 2 different scales: the EQ-5D, which is nonspecific to endometriosis, and the EHP-5, which

XSL•FO
RenderX

is specific to endometriosis. These were previously compared in a French study [39], which concluded that while both scales were appropriate for assessing the QOL of women with endometriosis, the EHP-5 was better at assessing health-related QOL, especially regarding medical treatment and the intensity of dysmenorrhea. According to our results, the EHP-5 does seem to better discriminate with regard to endometriosis, which echoes the conclusions of the French study [39]. However, regardless of the scale used, following our digital program was associated with an improved QOL. This was true for 2 to 3 times more participants than individuals in the control group. Our study used a 15% threshold to define a change in QOL, allowing stringent evolutions to be highlighted. Program participants had a poorer QOL level at baseline compared to the control group (despite a similar pain profile). However, after 3 months of following the program, program participants caught up with the QOL level of the control group.

Digital tools are a way to bring multidisciplinary interventions to patients, enhancing their accessibility. Studies on digital interventions for various conditions have shown positive results, such as improvements in anxiety and depression through CBT-based approaches [62], symptom management and QOL enhancements via mindfulness-based stress-reduction programs [63], or better health outcomes through digital personalized diets [64]. Similarly, multidisciplinary digital programs have demonstrated significant benefits across diverse chronic conditions, including improved QOL and symptom management. For instance, internet-based programs have been shown to enhance knowledge, increase exercise habits, and reduce heart failure symptoms [65], while personalized digital care programs have improved the QOL in adults with autoimmune diseases and post–COVID-19 condition [66]. Digital support programs have also improved self-reported QOL in patients with rheumatoid arthritis [67] and managing symptoms in patients with cancer [68]. In our study, we observed that a digital tool using a CBT approach is effective in helping patients manage a condition, in this case, endometriosis. This may have to do with the fact that digital technology can bring about a different perception and thus reduce the stigma surrounding intimate issues associated with endometriosis, such as psychological or sexuality aspects [28,69-71]. These findings underscore the potential of multidisciplinary digital programs to relieve symptoms, improve QOL, and enhance disease-related knowledge across various chronic conditions.

For digital therapeutics to achieve their purpose, they must meet patient needs in terms of content while optimizing patient adherence and fulfilling prescriber requirements. The content of our program was based on EHP components to address subjects that were relevant to patients with endometriosis and was approved by health care specialists. Other studies have looked at digital technology for managing endometriosis or pelvic pain and highlighted the need for educational content, psychological and social support, and patient empowerment in particular [40]. As for patient adherence, to enhance digital therapeutic efficiency, it is crucial that future studies analyze specific factors influencing patient engagement in a digital program, such as trust, interactions, and consideration [72].

Keeping patients engaged to increase their frequency of use and guiding them through nonpharmacological, multidisciplinary interventions are therefore key to a program's efficiency. Guidelines and studies encourage the multidisciplinary management of endometriosis. However, only 2 studies on the same intervention, evaluating the QOL after 2 and 12 weeks following a multimodal program, have been conducted in the context of digital intervention [41,42]. Our study contributes to answering the need for new research on the multidisciplinary management of endometriosis and the development of programs to make such management accessible [73,74].

Limitations and Strengths

Our study is subject to several limitations. It is based on voluntary participation only, with no compulsory questionnaires and no obligation to complete the program. It also includes a selection bias, with no information as to whether the profiles of the women who answered the questionnaires were similar to the overall profiles of the women who took part in the program. Although the participants in this study was not randomized, differences between program participants and the control group were minimized by selecting a control group sample with pain levels similar to those of program participants (despite a different initial QOL level, the effect of which was analyzed in sensitivity analyses). Using the Benjamini-Hochberg method, we were able to control the false discovery rate. Nevertheless, we interpreted results with borderline *P* values cautiously, acknowledging the increased risk of type I errors inherent in multiple testing.

Furthermore, participants self-reported their diagnosis of endometriosis, which may be clinical, imaging based, or surgical. However, one can assume that women who do not have endometriosis would not attend the program and would not spend time filling in research questionnaires.

A strength of the program is that it was designed based on user research, scientific rationale, and medical expertise. Our approach, targeting EHP components and using CBT, allowed the development of an effective digital tool for the self-management of endometriosis symptoms. In addition, the use of 2 different questionnaires on QOL showed that following the program was associated with an improvement of the health-related QOL (EQ-5D) and the QOL specific to endometriosis (EHP-5). Very little amount of data exists on the reliability and efficiency of a digital program for the management of endometriosis symptoms. Our study is the first step in identifying key factors to be considered for developing a digital health program for the daily management of endometriosis and in demonstrating its positive impact on patient symptoms and QOL.

Conclusions

Our results suggest that a digital health program providing medical and scientific information about endometriosis as well as multidisciplinary self-management tools may be a helpful and effective resource for women to manage life with endometriosis alongside their medical care. Hence, a digital program for endometriosis that combines integrative solutions,

XSL•FO
RenderX

focuses on EHP components, and uses a CBT approach can enhance patient care for those with endometriosis.

Acknowledgments

The authors would like to thank Laura Matteo for her help in proofreading the manuscript, and they would like to thank the entire team involved in developing this program. The authors would also like to graciously thank this pilot study's program participants for their involvement and time. This study was funded by the French National Association for Research and Technology and by Lyv Healthcare.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to the consent agreements made with participants, which did not include provisions for public data release but are available from the corresponding author on reasonable request.

Authors' Contributions

ZB and MA were responsible for conceptualizing the study. ZB was involved in the development of the program, study design, data collection and extraction, statistical analysis, data curation, and formal analysis. ES was involved in the development of the program and study design. MA was involved in the development of the program, study design, data collection, statistical analysis, and editing of the manuscript and supervised the entire research process. PP, EP, and DL were involved in the development of the program. The original draft of the manuscript was written collaboratively by ZB, MA, and ES. In addition, ZB, ES, MP, DL, EP, PP, and MA contributed to the writing, reviewing, and editing of the manuscript. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

The PhD of ZB was partly funded by the French National Association for Research and Technology and Lyv Healthcare. MA, MP, and ZB are employed at the Lyv Healthcare. ES, PP, DL, and EP are on the scientific committee of Lyv Healthcare and are paid speakers in the School of Endo program.

Multimedia Appendix 1
Distribution of symptom levels and quality of life for endometriosis program participants at baseline.
[PNG File , 243 KB - formative_v9i1e58262_app1.png]

Multimedia Appendix 2
The evolution of outcomes between baseline and 3 months for all endometriosis program participants versus the control group according to the content thresholds of the program tested.
[DOCX File , 19 KB - formative_v9i1e58262_app2.docx]

Multimedia Appendix 3
The evolution of outcomes between baseline and 3 months for all endometriosis program participants according to the sections thresholds of the program tested.
[DOCX File , 15 KB - formative_v9i1e58262_app3.docx]

Multimedia Appendix 4
Beta coefficients and 95% CIs of nonadjusted linear regression models evaluating associations between endometriosis program participation and health outcomes and interaction tests according to baseline quality of life levels (Endometriosis Health Profile-5 core part).
[DOCX File , 30 KB - formative_v9i1e58262_app4.docx]

References

1. Zondervan KT, Becker CM, Missmer SA. Endometriosis. N Engl J Med 2020 Mar 26;382(13):1244-1256. [doi: 10.1056/NEJMra1810764] [Medline: 32212520]
2. Della Corte L, Di Filippo C, Gabrielli O, Reppuccia S, La Rosa VL, Ragusa R, et al. The burden of endometriosis on women's lifespan: a narrative overview on quality of life and psychosocial wellbeing. Int J Environ Res Public Health 2020 Jun 29;17(13):4683 [FREE Full text] [doi: 10.3390/ijerph17134683] [Medline: 32610665]

XSL•FO
RenderX

3. O'Hara R, Rowe H, Fisher J, O'Hara R, Rowe H, Fisher J. Managing endometriosis: a cross-sectional survey of women in Australia. J Psychosom Obstet Gynaecol 2022 Sep;43(3):265-272 [FREE Full text] [doi: 10.1080/0167482X.2020.1825374] [Medline: 33050751]
4. Ceccaroni M, Bounous VE, Clarizia R, Mautone D, Mabrouk M. Recurrent endometriosis: a battle against an unknown enemy. Eur J Contracept Reprod Health Care 2019 Dec 25;24(6):464-474. [doi: 10.1080/13625187.2019.1662391] [Medline: 31550940]
5. Becker CM, Bokor A, Heikinheimo O, Horne A, Jansen F, Kiesel L, ESHRE Endometriosis Guideline Group. ESHRE guideline: endometriosis. Hum Reprod Open 2022;2022(2):hoac009 [FREE Full text] [doi: 10.1093/hropen/hoac009] [Medline: 35350465]
6. Nnoaham KE, Hummelshoj L, Webster P, d'Hooghe T, de Cicco Nardone F, de Cicco Nardone C, World Endometriosis Research Foundation Global Study of Women's Health consortium. Impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries. Fertil Steril 2011 Aug;96(2):366-73.e8 [FREE Full text] [doi: 10.1016/j.fertnstert.2011.05.090] [Medline: 21718982]
7. Giudice LC, Horne AW, Missmer SA. Time for global health policy and research leaders to prioritize endometriosis. Nat Commun 2023 Dec 04;14(1):8028 [FREE Full text] [doi: 10.1038/s41467-023-43913-9] [Medline: 38049392]
8. Armour M, Sinclair J, Chalmers KJ, Smith CA. Self-management strategies amongst Australian women with endometriosis: a national online survey. BMC Complement Altern Med 2019 Jan 15;19(1):17 [FREE Full text] [doi: 10.1186/s12906-019-2431-x] [Medline: 30646891]
9. Evans S, Fernandez S, Olive L, Payne LA, Mikocka-Walus A. Psychological and mind-body interventions for endometriosis: a systematic review. J Psychosom Res 2019 Sep;124:109756. [doi: 10.1016/j.jpsychores.2019.109756] [Medline: 31443810]
10. Hansen KE, Brandsborg B, Kesmodel US, Forman A, Kold M, Pristed R, et al. Psychological interventions improve quality of life despite persistent pain in endometriosis: results of a 3-armed randomized controlled trial. Qual Life Res 2023 Jun 17;32(6):1727-1744 [FREE Full text] [doi: 10.1007/s11136-023-03346-9] [Medline: 36797461]
11. Facchin F, Buggio L, Dridi D, Barbara G, Vercellini P. The subjective experience of dyspareunia in women with endometriosis: a systematic review with narrative synthesis of qualitative research. Int J Environ Res Public Health 2021 Nov 18;18(22):12112 [FREE Full text] [doi: 10.3390/ijerph182212112] [Medline: 34831868]
12. Mińko A, Turoń-Skrzypińska A, Rył A, Bargiel P, Hilicka Z, Michalczyk K, et al. Endometriosis-a multifaceted problem of a modern woman. Int J Environ Res Public Health 2021 Aug 02;18(15):8177 [FREE Full text] [doi: 10.3390/ijerph18158177] [Medline: 34360470]
13. Jones GL, Budds K, Taylor F, Musson D, Raymer J, Churchman D, et al. A systematic review to determine use of the Endometriosis Health Profiles to measure quality of life outcomes in women with endometriosis. Hum Reprod Update 2024 Mar 01;30(2):186-214 [FREE Full text] [doi: 10.1093/humupd/dmad029] [Medline: 38007607]
14. Jones G, Jenkinson C, Kennedy S. Evaluating the responsiveness of the endometriosis health profile questionnaire: the EHP-30. Qual Life Res 2004 Apr;13(3):705-713. [doi: 10.1023/b:qure.0000021316.79349.af]
15. Jones G, Jenkinson C, Kennedy S. Development of the short form endometriosis health profile questionnaire: the EHP-5. Qual Life Res 2004 Apr;13(3):695-704. [doi: 10.1023/b:qure.0000021321.48041.0e]
16. Hofmann SG, Asnaani A, Vonk IJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. Cognit Ther Res 2012 Oct 01;36(5):427-440 [FREE Full text] [doi: 10.1007/s10608-012-9476-1] [Medline: 23459093]
17. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. Clin Psychol Rev 2006 Jan;26(1):17-31. [doi: 10.1016/j.cpr.2005.07.003] [Medline: 16199119]
18. Schubert K, Lohse J, Kalder M, Ziller V, Weise C. Internet-based cognitive behavioral therapy for improving health-related quality of life in patients with endometriosis: study protocol for a randomized controlled trial. Trials 2022 Apr 12;23(1):300 [FREE Full text] [doi: 10.1186/s13063-022-06204-0] [Medline: 35414092]
19. DiGiuseppe R, Venezia R, Gotterbarn R. What is cognitive behavior therapy? In: Vernon A, Doyle KA, editors. Cognitive Behavior Therapies: A Guidebook for Practitioners. New York, NY: American Counseling Association; 2018:1-35.
20. Ehde DM, Dillworth TM, Turner JA. Cognitive-behavioral therapy for individuals with chronic pain: efficacy, innovations, and directions for research. Am Psychol 2014;69(2):153-166. [doi: 10.1037/a0035747] [Medline: 24547801]
21. Mikocka-Walus A, Druitt M, O'Shea M, Skvarc D, Watts JJ, Esterman A, et al. Yoga, cognitive-behavioural therapy versus education to improve quality of life and reduce healthcare costs in people with endometriosis: a randomised controlled trial. BMJ Open 2021 Aug 09;11(8):e046603 [FREE Full text] [doi: 10.1136/bmjopen-2020-046603] [Medline: 34373298]
22. Morley S, Eccleston C, Williams A. Systematic review and meta-analysis of randomized controlled trials of cognitive behaviour therapy and behaviour therapy for chronic pain in adults, excluding headache. Pain 1999 Mar;80(1-2):1-13. [doi: 10.1016/s0304-3959(98)00255-3] [Medline: 10204712]
23. de C Williams AC, Eccleston C, Morley S. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database Syst Rev 2012 Nov 14;11(11):CD007407 [FREE Full text] [doi: 10.1002/14651858.CD007407.pub3] [Medline: 23152245]

XSL•FO
RenderX

24. Burns JW, Nielson WR, Jensen MP, Heapy A, Czlapinski R, Kerns RD. Specific and general therapeutic mechanisms in cognitive behavioral treatment of chronic pain. J Consult Clin Psychol 2015 Feb;83(1):1-11. [doi: 10.1037/a0037208] [Medline: 24979313]
25. Andersson G, Carlbring P. Internet-assisted cognitive behavioral therapy. Psychiatr Clin North Am 2017 Dec;40(4):689-700. [doi: 10.1016/j.psc.2017.08.004] [Medline: 29080594]
26. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18 [FREE Full text] [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
27. Luik AI, van der Zweerde T, van Straten A, Lancee J. Digital delivery of cognitive behavioral therapy for insomnia. Curr Psychiatry Rep 2019 Jun 04;21(7):50 [FREE Full text] [doi: 10.1007/s11920-019-1041-0] [Medline: 31161406]
28. Stern E, Micoulaud Franchi JA, Dumas G, Moreira J, Mouchabac S, Maruani J, et al. How can digital mental health enhance psychiatry? Neuroscientist 2023 Dec 04;29(6):681-693. [doi: 10.1177/10738584221098603] [Medline: 35658666]
29. Gold SM, Friede T, Meyer B, Moss-Morris R, Hudson J, Asseyer S, et al. Internet-delivered cognitive behavioural therapy programme to reduce depressive symptoms in patients with multiple sclerosis: a multicentre, randomised, controlled, phase 3 trial. Lancet Digit Health 2023 Oct;5(10):e668-e678 [FREE Full text] [doi: 10.1016/S2589-7500(23)00109-7] [Medline: 37775187]
30. Murphy JL, Cordova MJ, Dedert EA. Cognitive behavioral therapy for chronic pain in veterans: evidence for clinical effectiveness in a model program. Psychol Serv 2022 Feb;19(1):95-102 [FREE Full text] [doi: 10.1037/ser0000506] [Medline: 32986454]
31. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337(sep29 1):a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
32. Koninckx PR, Ussia A, Gordts S, Keckstein J, Saridogan E, Malzoni M, et al. The 10 "cardinal sins" in the clinical diagnosis and treatment of endometriosis: a Bayesian approach. J Clin Med 2023 Jul 07;12(13):4547 [FREE Full text] [doi: 10.3390/jcm12134547] [Medline: 37445589]
33. Ambrose KR, Golightly YM. Physical exercise as non-pharmacological treatment of chronic pain: why and when. Best Pract Res Clin Rheumatol 2015 Feb;29(1):120-130 [FREE Full text] [doi: 10.1016/j.berh.2015.04.022] [Medline: 26267006]
34. Borisovskaya A, Chmelik E, Karnik A. Exercise and chronic pain. Adv Exp Med Biol 2020;1228:233-253. [doi: 10.1007/978-981-15-1792-1_16] [Medline: 32342462]
35. Velho RV, Werner F, Mechsner S. Endo belly: what is it and why does it happen?: a narrative review. J Clin Med 2023 Nov 19;12(22):7176 [FREE Full text] [doi: 10.3390/jcm12227176] [Medline: 38002788]
36. Luscombe GM, Markham R, Judio M, Grigoriu A, Fraser IS. Abdominal bloating: an under-recognized endometriosis symptom. J Obstet Gynaecol Can 2009 Dec;31(12):1159-1171. [doi: 10.1016/s1701-2163(16)34377-8] [Medline: 20085682]
37. Maroun P, Cooper MJ, Reid GD, Keirse MJ. Relevance of gastrointestinal symptoms in endometriosis. Aust N Z J Obstet Gynaecol 2009 Aug;49(4):411-414. [doi: 10.1111/j.1479-828X.2009.01030.x] [Medline: 19694698]
38. Pokrzywinski RM, Soliman AM, Snabes MC, Chen J, Taylor HS, Coyne KS. Responsiveness and thresholds for clinically meaningful changes in worst pain numerical rating scale for dysmenorrhea and nonmenstrual pelvic pain in women with moderate to severe endometriosis. Fertil Steril 2021 Feb;115(2):423-430 [FREE Full text] [doi: 10.1016/j.fertnstert.2020.07.013] [Medline: 33066973]
39. Aubry G, Panel P, Thiollier G, Huchon C, Fauconnier A. Measuring health-related quality of life in women with endometriosis: comparing the clinimetric properties of the Endometriosis Health Profile-5 (EHP-5) and the EuroQol-5D (EQ-5D). Hum Reprod 2017 Jun 01;32(6):1258-1269. [doi: 10.1093/humrep/dex057] [Medline: 28383700]
40. Gagnon MM, Brilz AR, Alberts NM, Gordon JL, Risling TL, Stinson JN. Understanding adolescents' experiences with menstrual pain to inform the user-centered design of a mindfulness-based app: mixed methods investigation study. JMIR Pediatr Parent 2024 Apr 08;7:e54658 [FREE Full text] [doi: 10.2196/54658] [Medline: 38587886]
41. Rohloff N, Rothenhöfer M, Götz T, Schäfer SD. Observational pilot study on the influence of an app-based self-management program on the quality of life of women with endometriosis. Arch Gynecol Obstet 2024 Aug 13;310(2):1157-1170. [doi: 10.1007/s00404-024-07468-4] [Medline: 38869629]
42. Rohloff N, Götz T, Kortekamp SS, Heinze NR, Weber C, Schäfer SD. Influence of app-based self-management on the quality of life of women with endometriosis. Cureus 2024 Aug;16(8):e67655. [doi: 10.7759/cureus.67655] [Medline: 39314601]
43. Donatti L, Malvezzi H, de Azevedo BC, Baracat EC, Podgaec S. Cognitive behavioral therapy in endometriosis, psychological based intervention: a systematic review. Rev Bras Ginecol Obstet 2022 Mar 16;44(3):295-303 [FREE Full text] [doi: 10.1055/s-0042-1742406] [Medline: 35576938]
44. Donatti L, Podgaec S, Baracat EC. Efficacy of cognitive behavioral therapy in treating women with endometriosis and chronic pelvic pain: a randomized trial. J Health Psychol 2024 Apr 02:13591053241240198. [doi: 10.1177/13591053241240198] [Medline: 38566501]

XSL•FO
RenderX

45. Wu S, Wang X, Liu H, Zheng W. Efficacy of cognitive behavioral therapy after the surgical treatment of women with endometriosis: a preliminary case-control study. Medicine (Baltimore) 2022 Dec 23;101(51):e32433 [FREE Full text] [doi: 10.1097/MD.0000000000032433] [Medline: 36595829]
46. Jafari H, Courtois I, Van den Bergh O, Vlaeyen JW, Van Diest I. Pain and respiration: a systematic review. Pain 2017 Jun;158(6):995-1006. [doi: 10.1097/j.pain.0000000000000865] [Medline: 28240995]
47. Petit E, Lhuillery D, Loriau J, Sauvanet E. Endométriose: Diagnostic et Prise en Charge. Paris: Elsevier-Masson, Pratique en Gynécologie-Obstétrique; 2020.
48. Zhao L, Wu H, Zhou X, Wang Q, Zhu W, Chen J. Effects of progressive muscular relaxation training on anxiety, depression and quality of life of endometriosis patients under gonadotrophin-releasing hormone agonist therapy. Eur J Obstet Gynecol Reprod Biol 2012 Jun;162(2):211-215. [doi: 10.1016/j.ejogrb.2012.02.029] [Medline: 22455972]
49. Samami E, Shahhosseini Z, Khani S, Elyasi F. Pain-focused psychological interventions in women with endometriosis: a systematic review. Neuropsychopharmacol Rep 2023 Sep 27;43(3):310-319 [FREE Full text] [doi: 10.1002/npr2.12348] [Medline: 37366616]
50. Pontoppidan K, Olovsson M, Grundström H. Clinical factors associated with quality of life among women with endometriosis: a cross-sectional study. BMC Womens Health 2023 Oct 24;23(1):551 [FREE Full text] [doi: 10.1186/s12905-023-02694-5] [Medline: 37875883]
51. Cirillo M, Argento FR, Becatti M, Fiorillo C, Coccia ME, Fatini C. Mediterranean diet and oxidative stress: a relationship with pain perception in endometriosis. Int J Mol Sci 2023 Sep 27;24(19):14601 [FREE Full text] [doi: 10.3390/ijms241914601] [Medline: 37834048]
52. Moore JS, Gibson PR, Perry RE, Burgell RE. Endometriosis in patients with irritable bowel syndrome: specific symptomatic and demographic profile, and response to the low FODMAP diet. Aust N Z J Obstet Gynaecol 2017 Apr 17;57(2):201-205. [doi: 10.1111/ajo.12594] [Medline: 28303579]
53. Borghini R, Porpora MG, Casale R, Marino M, Palmieri E, Greco N, et al. Irritable bowel syndrome-like disorders in endometriosis: prevalence of nickel sensitivity and effects of a low-nickel diet. An open-label pilot study. Nutrients 2020 Jan 28;12(2):341 [FREE Full text] [doi: 10.3390/nu12020341] [Medline: 32012984]
54. Oddsson SJ, Gunnarsdottir T, Johannsdottir LG, Amundadottir ML, Frimannsdottir A, Molander P, et al. A new digital health program for patients with inflammatory bowel disease: preliminary program evaluation. JMIR Form Res 2023 Apr 28;7:e39331 [FREE Full text] [doi: 10.2196/39331] [Medline: 37115598]
55. Gonçalves AV, Barros NF, Bahamondes L. The practice of hatha yoga for the treatment of pain associated with endometriosis. J Altern Complement Med 2017 Jan;23(1):45-52. [doi: 10.1089/acm.2015.0343] [Medline: 27869485]
56. Posadzki P, Ernst E, Terry R, Lee MS. Is yoga effective for pain? A systematic review of randomized clinical trials. Complement Ther Med 2011 Oct;19(5):281-287. [doi: 10.1016/j.ctim.2011.07.004] [Medline: 21944658]
57. Matloobi M, Amini L, Shahali S, Haghani H, Tahermanesh K, Hassanlouei B, et al. Effect of sex education on sexual function and sexual quality of life in women with endometriosis: a quasi-experimental study. Int J Gynaecol Obstet 2022 Dec 25;159(3):702-710. [doi: 10.1002/ijgo.14254] [Medline: 35514234]
58. Armour M, Avery J, Leonardi M, Van Niekerk L, Druitt ML, Parker MA, et al. Lessons from implementing the Australian National Action Plan for Endometriosis. Reprod Fertil 2022 Jul 01;3(3):C29-C39 [FREE Full text] [doi: 10.1530/RAF-22-0003] [Medline: 35928674]
59. Endometriosis: diagnosis and management. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/ng73 [accessed 2024-04-29]
60. Wattier JM. Antalgiques et alternatives thérapeutiques non médicamenteuses pluridisciplinaires. Gynecol Obstet Fertil Senol 2018 Mar;46(3):248-255. [doi: 10.1016/j.gofs.2018.02.002] [Medline: 29510963]
61. De Hoyos G, Ramos-Sostre D, Torres-Reverón A, Barros-Cartagena B, López-Rodríguez V, Nieves-Vázquez C, et al. Efficacy of an environmental enrichment intervention for endometriosis: a pilot study. Front Psychol 2023 Oct 10;14:1225790 [FREE Full text] [doi: 10.3389/fpsyg.2023.1225790] [Medline: 37885745]
62. Verkleij M, Georgiopoulos AM, Barendrecht H, Friedman D. Pilot of a therapist-guided digital mental health intervention (eHealth CF-CBT) for adults with cystic fibrosis. Pediatr Pulmonol 2023 Jul 05;58(7):2094-2103. [doi: 10.1002/ppul.26438] [Medline: 37144856]
63. Miazga E, Starkman H, Schroeder N, Nensi A, McCaffrey C. Virtual mindfulness-based therapy for the management of endometriosis chronic pelvic pain: a novel delivery platform to increase access to care. J Obstet Gynaecol Can 2024 Jun;46(6):102457 [FREE Full text] [doi: 10.1016/j.jogc.2024.102457] [Medline: 38614242]
64. Jactel SN, Olson JM, Wolin KY, Brown J, Pathipati MP, Jagiella VJ, et al. Efficacy of a digital personalized elimination diet for the self-management of irritable bowel syndrome and comorbid irritable bowel syndrome and inflammatory bowel disease. Clin Transl Gastroenterol 2023 Jan 01;14(1):e00545 [FREE Full text] [doi: 10.14309/ctg.0000000000000545] [Medline: 36322404]
65. Tomita MR, Tsai BM, Fisher NM, Kumar NA, Wilding G, Stanton K, et al. Effects of multidisciplinary internet-based program on management of heart failure. J Multidiscip Healthc 2008 Dec 01;2009(2):13-21 [FREE Full text] [doi: 10.2147/jmdh.s4355] [Medline: 20505786]

XSL•FO
RenderX

66. Bundy N, De Jesus M, Lytle M, Calabrese L, Gobin C, Dyhrberg M. Self-evidence-based digital care programme improves health-related quality of life in adults with a variety of autoimmune diseases and long COVID: a retrospective study. RMD Open 2023 May 16;9(2):e003061 [FREE Full text] [doi: 10.1136/rmdopen-2023-003061] [Medline: 37192812]
67. Dobies B, White A, Isberg A, Gudmundsson S, Oddsson S. Digital health program improves quality of life in rheumatoid arthritis: a retrospective analysis of real-world data. Clin Exp Rheumatol 2024 Jan;42(1):10-14 [FREE Full text] [doi: 10.55563/clinexprheumatol/rng5n9] [Medline: 38306018]
68. Gudmundsson GH, Mészáros J, Björnsdóttir Á, Ámundadóttir ML, Thorvardardottir GE, Magnusdottir E, et al. Evaluating the feasibility of a digital therapeutic program for patients with cancer during active treatment: pre-post interventional study. JMIR Form Res 2022 Oct 13;6(10):e39764 [FREE Full text] [doi: 10.2196/39764] [Medline: 36227639]
69. Hollis C, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr;206(4):263-265 [FREE Full text] [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
70. Sims OT, Gupta J, Missmer SA, Aninye IO. Stigma and endometriosis: a brief overview and recommendations to improve psychosocial well-being and diagnostic delay. Int J Environ Res Public Health 2021 Aug 03;18(15):8210 [FREE Full text] [doi: 10.3390/ijerph18158210] [Medline: 34360501]
71. Abdulai AF, Howard AF, Yong PJ, Noga H, Parmar G, Currie LM. Developing an educational website for women with endometriosis-associated dyspareunia: usability and stigma analysis. JMIR Hum Factors 2022 Mar 03;9(1):e31317 [FREE Full text] [doi: 10.2196/31317] [Medline: 35238789]
72. Stern E, Breton Z, Alexaline M, Geoffroy PA, Bungener C. Redefining the relationship in digital care: a qualitative study of the Digital Therapeutic Alliance. Encephale (Forthcoming) 2024 May 08 [FREE Full text] [doi: 10.1016/j.encep.2024.02.011] [Medline: 38724431]
73. Agarwal SK, Antunez-Flores O, Foster WG, Hermes A, Golshan S, Soliman AM, et al. Real-world characteristics of women with endometriosis-related pain entering a multidisciplinary endometriosis program. BMC Womens Health 2021 Jan 07;21(1):19 [FREE Full text] [doi: 10.1186/s12905-020-01139-7] [Medline: 33413295]
74. Agarwal SK, Foster WG, Groessl EJ. Rethinking endometriosis care: applying the chronic care model via a multidisciplinary program for the care of women with endometriosis. Int J Womens Health 2019;11:405-410 [FREE Full text] [doi: 10.2147/IJWH.S207373] [Medline: 31413643]

Abbreviations

CBT: cognitive behavioral therapy
EHP: Endometriosis Health Profile
NRS: Numeric Rating Scale
QOL: quality of life
T0: time point 0

Edited by A Mavragani; submitted 11.03.24; peer-reviewed by M Armour, A AL-Asadi, G Joseph; comments to author 12.05.24; revised version received 24.06.24; accepted 08.01.25; published 28.02.25.

Please cite as:
Breton Z, Stern E, Pinault M, Lhuillery D, Petit E, Panel P, Alexaline M
A Digital Program for Daily Life Management With Endometriosis: Pilot Cohort Study on Symptoms and Quality of Life Among Participants
JMIR Form Res 2025;9:e58262
URL: https://formative.jmir.org/2025/1/e58262
doi:10.2196/58262
PMID:39791286

©Zélia Breton, Emilie Stern, Mathilde Pinault, Delphine Lhuillery, Erick Petit, Pierre Panel, Maïa Alexaline. Originally published in JMIR Formative Research (https://formative.jmir.org), 28.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Daily Treatment Monitoring for Patients Receiving Home-Based Peritoneal Dialysis and Prediction of Heart Failure Risk: mHealth Tool Development and Modeling Study

Jia Wu[1*], MM; Youjia Zeng[1*], PhD; Jun Yang[1*], MM; Yutong Yao[1], BMed; Xiuling Xu[1], BMed; Gaofeng Song[1], PhD; Wuyong Yi[1], BMed; Taifen Wang[1], BMed; Yihou Zheng[1], BMed; Zhongwei Jia[2*], PhD; Xiangyu Yan[3,4*], PhD

[1]Shenzhen Hospital of Traditional Chinese Medicine, Shenzen, China

[2]Department of Global Health, School of Public Health, Peking University, Beijing, China

[3]School of Disaster and Emergency Medicine, Tianjin University, Tianjin, China

[4]Key Laboratory of Medical Rescue Key Technology and Equipment, Ministry of Emergency Management, Tianjin, China

[*]these authors contributed equally

Corresponding Author:
Xiangyu Yan, PhD
School of Disaster and Emergency Medicine
Tianjin University
No 92 Weijin Road
Tianjin, 300072
China
Phone: 86 02287370177 ext 307
Email: yanxiangyu1123@163.com

Abstract

Background: Peritoneal dialysis is one of the major renal replacement modalities for patients with end-stage renal disease. Heart failure is a common adverse event among patients who undergo peritoneal dialysis treatment, especially for those who undergo continuous ambulatory peritoneal dialysis at home, because of the lack of professional input-output volume monitoring and management during treatment.

Objective: This study aims to develop novel mobile health (mHealth) tools to improve the quality of home-based continuous ambulatory peritoneal dialysis treatment and to build a prediction model of heart failure based on the system's daily treatment monitoring data.

Methods: The mHealth tools with a 4-layer system were designed and developed using Spring Boot, MyBatis Plus, MySQL, and Redis as backend technology stack, and Vue, Element User Interface, and WeChat Mini Program as front-end technology stack. Patients were recruited to use the tool during daily peritoneal dialysis treatment from January 1, 2017, to April 20, 2023. Logistic regression models based on real-time treatment monitoring data were used for heart failure prediction. The sensitivity, specificity, accuracy, and Youden index were calculated to evaluate the performance of the prediction model. In the sensitivity analysis, the ratio of patients with and without heart failure was set to 1:4 and 1:10, respectively, to better evaluate the stability of the prediction model.

Results: A WeChat Mini Program named Futou Bao for patients and a patient data management platform for doctors was developed. Futou Bao included an intelligent data upload function module and an auxiliary function module. The doctor's data management platform consisted of 4 function modules, that is, patient management, data visualization and marking, data statistics, and system management. During the study period, the records of 6635 patients who received peritoneal dialysis treatment were uploaded in Futou Bao, with 0.71% (47/6635) of them experiencing heart failure. The prediction model that included sex, age, and diastolic blood pressure was considered as the optimal model, wherein the sensitivity, specificity, accuracy, and Youden index were 0.75, 0.91, 0.89, and 0.66, respectively, with an area under the curve value of 0.879 (95% CI 0.772-0.986) using the validation dataset. The sensitivity analysis showed stable results.

Conclusions: This study provides a new home-based peritoneal dialysis management paradigm that enables the daily monitoring and early warning of heart failure risk. This novel paradigm is of great value for improving the efficiency, security, and personalization of peritoneal dialysis.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e56254) doi:10.2196/56254

KEYWORDS

peritoneal dialysis; mHealth; patient management; heart failure; prediction model

Introduction

Peritoneal dialysis is one of the 3 major renal replacement modalities (ie, hemodialysis, peritoneal dialysis, and kidney transplantation) for end-stage renal disease (ESRD) [1-3]. Continuous ambulatory peritoneal dialysis (CAPD) gradually becomes the mainstream mode for peritoneal dialysis [4]. Patients can receive the treatment at home because the treatment site environment needs less treatment resources and is flexible [5,6]. Compared with hemodialysis, peritoneal dialysis as a home-based, cost-saving renal replacement therapy has several advantages, including protection of residual kidney function, stable hemodynamics, and less interference with normal family, work, and social life. Peritoneal dialysis therapy has been recognized and accepted by patients with ESRD.

One common and serious adverse event in patients undergoing CAPD is capacity-overloaded heart failure. Previous studies have shown a high prevalence of congestive heart failure among patients who receive peritoneal dialysis treatment. A multicenter cohort study in mainland China conducted in 2012 reported that heart failure occurred in 47.3% of the patients who underwent peritoneal dialysis [7]. Contemporaneously, Wang et al's [8] study conducted in Hong Kong showed a heart failure prevalence of 39% in patients treated with peritoneal dialysis. Although the prevalence reported by subsequent studies showed a decrease to 29% in Wang et al's study [9] and to 19.6% in Sun et al's study [10], the prevalence remained at a high level. Therefore, the prevention and early detection of heart failure in patients undergoing peritoneal dialysis treatment is a vital issue that cannot be ignored.

Although operating peritoneal dialysis treatment at home has many advantages, there are still several treatment bottlenecks in the traditional home-based peritoneal dialysis approach. First, because patients manage the treatment at home, doctors cannot provide timely and accurate follow-up. Second, the effectiveness of peritoneal dialysis, the occurrence of adverse events represented by heart failure, and the quality of life of patients are closely related to the training and management level of the medical institutions and the self-management levels of the patients. Third, it is difficult to implement good remote whole-process quality management of patients and prevent adverse events in a timely and effective manner for patients receiving CAPD at home.

Given the treatment bottleneck in the traditional home-based peritoneal dialysis method, it is essential to develop a user-friendly remote intelligent treatment management platform. The internet, big data, and other mobile health (mHealth) technologies can enable seamless integration of mobile equipment, systems, and home-based CAPD treatment operations to conduct timely detection and early intervention of adverse events represented by heart failure through analyzing the automatically collected data by the system.

The aim of this study was to develop mHealth tools and systems to improve the efficiency, accuracy, and personalization of the home-based peritoneal dialysis treatment and to build a prediction model of heart failure in patients with peritoneal dialysis based on the systems' daily treatment monitoring data.

Methods

mHealth Tool Design and Development Procedures

The mHealth tools developed in this study were intended to promote the intelligence of the home peritoneal dialysis treatment in the following 4 aspects.

1. Automated data collection: Through the Bluetooth connection device, the system enables the automatic collection of peritoneal dialysis data, including peritoneal dialysis solution weight and patient's weight data, which can reduce manual operation and the possibility of data input error.
2. Daily data analysis: The system timely processes and analyzes the uploaded data and provides intuitive data visualization to help doctors better understand the patient's treatment status.
3. Remote treatment monitoring: In order to improve the efficiency and convenience of treatment, patients do not need to go to the hospital frequently for treatment, and doctors can remotely monitor and manage the treatment process of patients.
4. Personalized treatment plan: Through the real-time data analysis and treatment monitoring, doctors can develop personalized treatment plans according to the specific conditions of patients and improve the treatment effect.

In order to improve the mHealth system's usage convenience, this study plans to use Spring Boot, MyBatis Plus, MySQL, and Redis as backend technology stack, and Vue, Element User Interface, and WeChat Mini Program as the front-end technology stack. The whole system is divided into 4 layers: data storage layer, core service layer, system application layer, and user interface layer. This architecture could make the system efficient, reliable, and extensible, and meet the functional and performance requirements of peritoneal dialysis treatment management. The details of each layer are as follows (Figure 1):

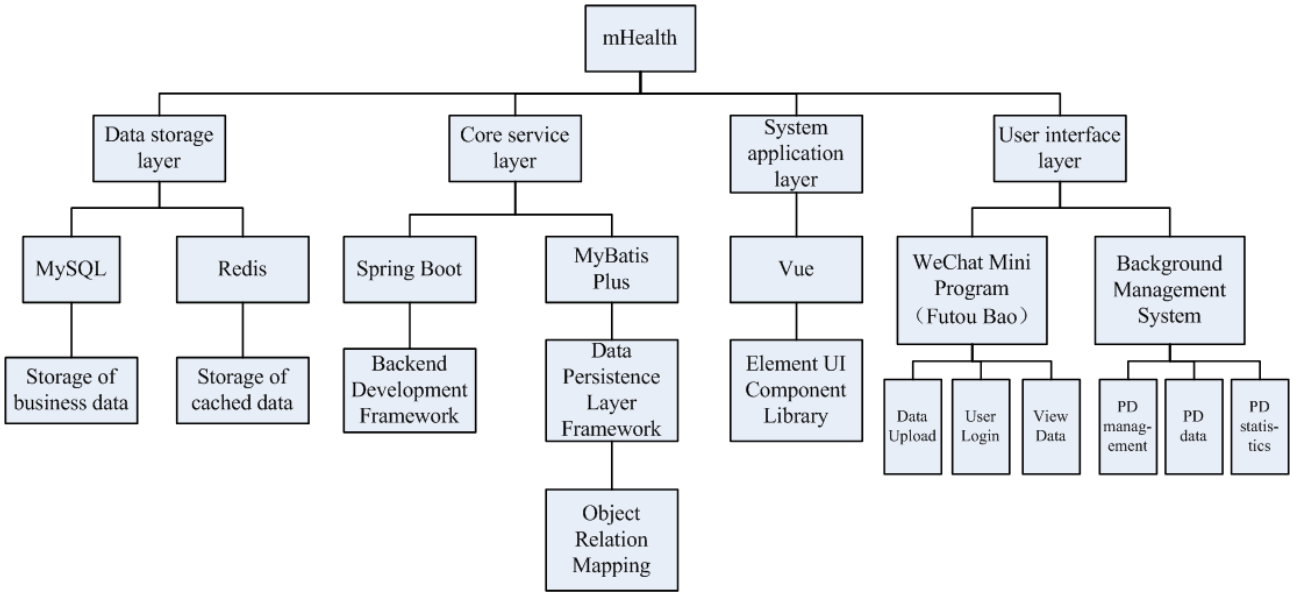
1. Data storage layer: This layer is responsible for data storage and management. MySQL database is used to store various data, including user information and medical records. Redis is used as a nonrelational database to store login token and treatment time point information as a cache to improve the ability to query data.
2. Core service layer: This layer is the core operation processing layer of the system, responsible for realizing the core functions and business logic of the system. Spring Boot framework was adopted as the backend development framework, combined with MyBatis Plus as the persistence

XSL•FO
RenderX

framework. Each function module of the system was developed through the annotation method of Spring Boot and used the Object Relational Mapping provided by MyBatis Plus to interact with the MySQL database.

3. System application layer: This layer is responsible for receiving and handling user requests. A single-page application based on Vue framework was used as the front-end technology of the background management system, and Element User Interface was used as the user interface component library to build the web pages and interactive interface of the system. For patient users, the development framework of the WeChat Mini Program was adopted to realize the user's access and use of the system in WeChat (the most popular mobile app in China).
4. User interface layer: This is the interface layer for the system to interact with users and is responsible for displaying the data and functions of the system to users in an intuitive and friendly way.

Figure 1. Architecture of the mobile health system. mHealth: mobile health; PD: peritoneal dialysis; UI: user interface.



Design of the Longitudinal Study and Model Prediction

From January 1, 2017, to April 20, 2023, a longitudinal study was conducted based on the usage of the Futou Bao system among outpatients of Shenzhen Hospital of Traditional Chinese Medicine. Patients who met the following criteria were recruited and included in the longitudinal study. The inclusion criteria for the participants were as follows.

1. Met the clinical standard of home peritoneal dialysis, including (1) patients with ESRD who needed continuous renal replacement therapy and chose peritoneal dialysis as the kidney replacement therapy; (2) patients underwent peritoneal dialysis catheterization, and the peritoneal dialysis catheter was inserted, which met the requirements of the peritoneal dialysis treatment; and (3) patients accepted the training of home peritoneal dialysis operation and obtained the qualification of self-completion of peritoneal dialysis operation after the training.
2. Intended to receive medical follow-up in Shenzhen Hospital of Traditional Chinese Medicine.
3. Had no difficulty in using smartphones and WeChat.
4. Willing to follow the procedure of the mHealth tool for operate peritoneal dialysis at home and sign the informed consent document.

The exclusion criteria included (1) patients diagnosed with heart failure and met either of the following 2 conditions, namely, acute heart failure classified as New York Heart Association class III or above and nonacute heart failure classified as New York Heart Association class II or below but combining with peripheral edema; and (2) patients had other serious physical and mental diseases that they could not complete the peritoneal dialysis at home with the help of the mHealth tool that we developed. In addition, in the process of using the mHealth tool, when clinical evaluation indicated that the patient did not meet the conditions of home peritoneal dialysis and needed hospitalization, the use of the mHealth tool would be stopped immediately.

Prior to patient recruitment, a specialized patient home peritoneal dialysis management team was selected, including professional doctors and nurses from the Department of Nephrology. Training for the patient management team was conducted, including home peritoneal dialysis operation and system use methods, patient recruitment methods, patient follow-up methods, and management methods. Thereafter, the above management team recruited the patients in their daily work and conducted one-on-one training for the patients who met the enrollment criteria. Patients were asked to use the mHealth tool during each peritoneal dialysis at home. Designated doctors and nurses were assigned to patients for treatment monitoring and follow-up.

During the usage of Futou Bao, patients' daily peritoneal dialysis treatment data were recorded in the system, which were obtained to create the dataset of the modeling process. In the dataset, identifiable information of the patients, such as names, were replaced with a string of encrypted ID numbers to blind the data analysts during analysis. Sex, age, time duration of peritoneal

XSL•FO
RenderX

dialysis, weight, urine volume, diastolic blood pressure, systolic blood pressure, and pulse pressure difference were obtained and included in the dataset as prediction variables to predict the onset of heart failure. Because the dialysate recommended in the home-based peritoneal dialysis was 1.5% glucose concentrations, the type of dialysate was not considered as a prediction variable in this study. For the heart failure outcome, because the occurrence of heart failure was always a progressive process among patients who received peritoneal dialysis, the 15 days before the date of hospitalization for heart failure was defined as the time duration of progression of heart failure. The data collected in that time duration were considered to be related to heart failure. Due to the careful and close patient management, the exact time of heart failure hospitalization could be obtained so as to avoid information bias. Because heart failure occurred infrequently, we randomly selected 6 times the number of patients with heart failure from the data of patients without heart failure to form the study dataset (number of patients with heart failure vs number of patients without heart failure = 1:6). After that, 80% of the data in the study dataset were randomly selected as the training dataset, while 20% were selected as the validation dataset of the prediction model. The sampling method without replacement was adopted in the above process. In the sensitivity analysis, we set the ratio of patients with and without heart failure to 1:4 and 1:10, respectively, to better evaluate the stability of the prediction model.

Statistical Analysis

The therapeutic effect and process of patients undergoing peritoneal dialysis treatment is self-specific, which should be considered in conjunction with the patient's own treatment monitoring data. For improving the adaptability of prediction model, it is more important and valuable to observe patients' own indicator value changes, including weight, urine volume, and blood pressure. Considering the different baseline data of each patient, the least squares method was used to fit the data of each patient, and discrete points were defined as the actual data that were not in the confidence intervals of the fitted data calculated by least squares method [11]. With the help of this method, the urine volume, diastolic blood pressure, systolic blood pressure, and pulse pressure difference were converted from numerical variables to categorical variables (ie, normal, abnormal). In this study, both α=.05 and α=.01 were used as the CI criterion (ie, 95% CI and 99% CI) to divide the categories of variables, so as to choose the optimal model.

Logistic regression models were used as prediction models. Four models were built using the training dataset. Models 1 and 2 used the variables based on the 95% CI criterion, while models 3 and 4 used the variables based on the 99% CI criterion. In models 1 and 3, all of the variables were included, while models 2 and 4 used the backward method and only included variables that were statistically significant. Receiver operating characteristic curves were presented, and the corresponding area under the curve (AUC) values were calculated. The AUC value and Akaike's information criterion (AIC) value were considered comprehensively to select the optimal model, which showed the highest AUC value and the lowest AIC value. Using the validation dataset, the sensitivity, specificity, accuracy, and Youden index of the optimal model were calculated, and the receiver operating characteristic curve and decision curve were presented to evaluate the performance of the prediction model comprehensively. The specific calculation formulas were as follows (Figure 2):

Sensitivity = True Positive / (True Positive + False Negative) (Formula 1)

Specificity = True Negative / (False Positive + True Negative) (Formula 2)

Accuracy = (True Positive + True Negative) / (True Positive + False Positive + False Negative + True Negative) (Formula 3)

Youden index = (Sensitivity + Specificity) – 1 (Formula 4)

A 2-sided *P* value of .05 was used for null hypothesis tests. Statistical analyses were conducted using SPSS version 19.0 (IBM Corp) and R version 4.3.0 (R Core Team).

Figure 2. Definitions of true positive, false positive, false negative, and true negative.

	Actual heart failure	Actual non-heart failure
Predicted heart failure	True Positive (TP)	False Positive (FP)
Predicted non-heart failure	False Negative (FN)	True Negative (TN)

Ethics Approval

This study was approved by the medical ethics committee of Shenzhen Hospital of Traditional Chinese Medicine (B2019043 and K2021-036-01). Written informed consent was obtained from the participants. In the analysis process, identifiable information of the patients (ie, names) was replaced with a string of encrypted ID numbers to protect patients' privacy. There was no compensation (money or gift) provided to the participants during the implementation of the study. However, participants received careful and fair health services and management of home-based peritoneal dialysis treatment.

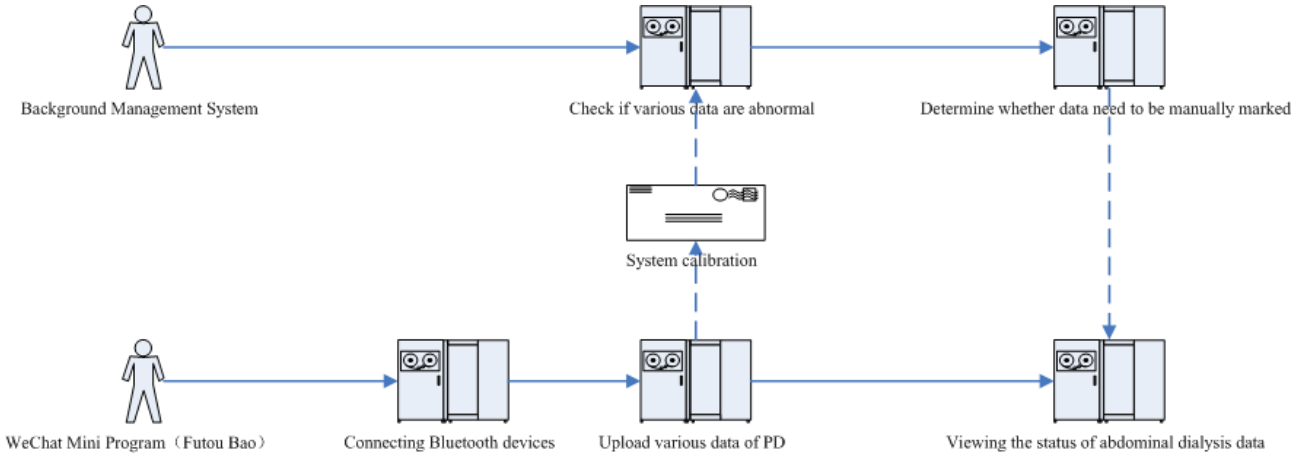
XSL•FO
RenderX

Results

mHealth Tools Development: Modules and Functions

In order to realize safe and convenient peritoneal dialysis for patients at home and to make it convenient for doctors to know the treatment and the physical condition of the patients in real time, an mHealth system was developed, comprising a WeChat Mini Program for patients named Futou Bao (“腹透宝” in Chinese) and a patient data management platform used by doctors (Figure 3).

Figure 3. Functional flow diagram. PD: peritoneal dialysis.



WeChat Mini Program (Futou Bao) for Patients

Futou Bao is a WeChat Mini Program that patients can use easily on WeChat, which is one of the most popular mobile social apps in China by scanning the quick response code or searching the name (“腹透宝” in Chinese) through WeChat. The main function of Futou Bao is the intelligent data upload function for patients receiving peritoneal dialysis treatment at home. Futou Bao has a series of auxiliary functions to complement this main function module.

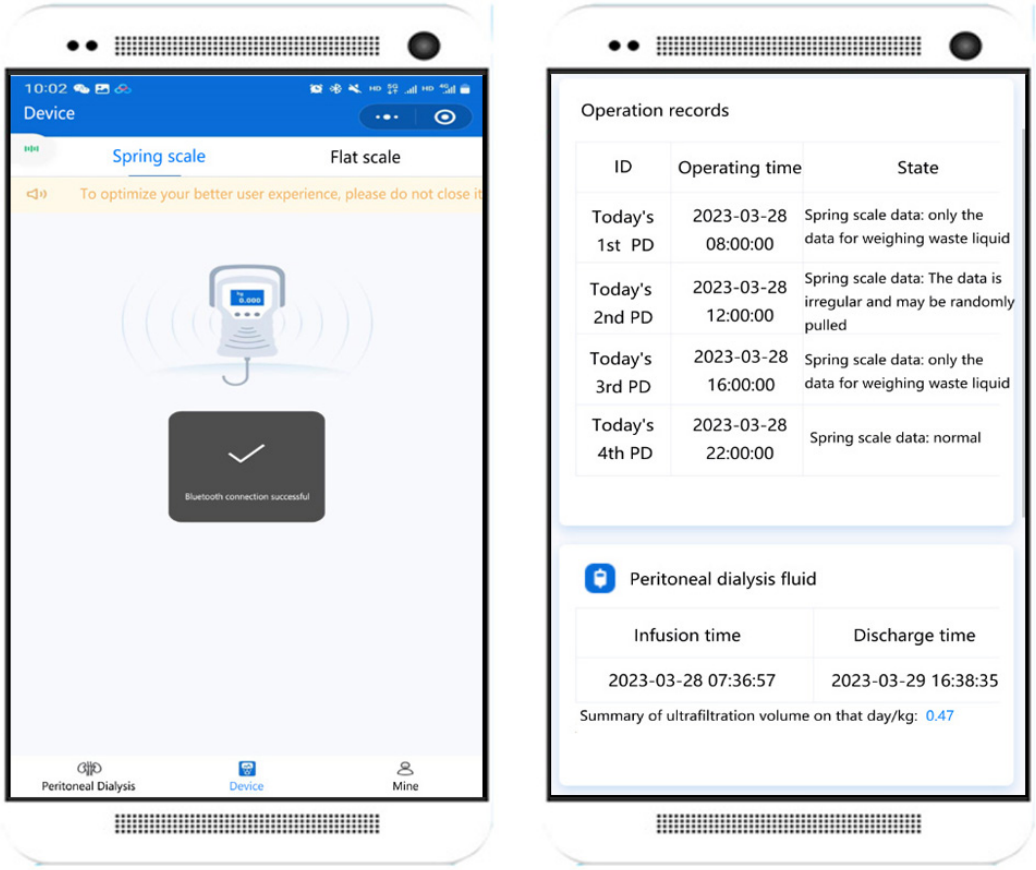
Intelligent Data Upload Function Module

In this module, patients’ data during the treatment can be uploaded to the system in real time, including blood pressure, urine volume, weight, and peritoneal dialysis solution weight at different time points during peritoneal dialysis. Data can be uploaded at patients’ homes in the following 2 ways: by the patients filling in the interface of the module and by the devices automatically uploading through Bluetooth. The detailed data uploading process of the patients during peritoneal dialysis treatment is as follows (Figure 4):

1. Patients measure their own blood pressure and upload it through the module interface, including systolic and diastolic blood pressure, on the day of peritoneal dialysis.
2. Patients need to measure the urine volume per urination according to the training method and calculate the urine volume on the day of receiving peritoneal dialysis treatment and upload the data through the module interface.
3. After peritoneal dialysis, patients need to weigh themselves on the Bluetooth-enabled weighing scale, after which the weight data can be uploaded directly to the system through Bluetooth (Figure 4).
4. For the peritoneal dialysis solution weight, Futou Bao adopts the method of intelligent data collection in the whole process of peritoneal dialysis treatment. Patients need to place the peritoneal dialysis solution on a Bluetooth-enabled spring scale or electronic flat scale before the start of peritoneal dialysis treatment, and the device can upload the peritoneal dialysis solution weight data through Bluetooth in real time (Figure 4).

XSL•FO
RenderX

Figure 4. User interface of Futou Bao. (A) Uploading data through Bluetooth. (B) Peritoneal dialysis operation records. PD: peritoneal dialysis.



(A) (B)

After uploading peritoneal dialysis data, patients can also check whether the peritoneal dialysis operation is correct or not in this module, including abnormal problems automatically identified by the system and manually marked by doctors so that patients can improve their peritoneal dialysis operation at home (Figure 4).

Auxiliary Function Module

There are 2 auxiliary functions to support the normal operation of the data upload function, as follows.

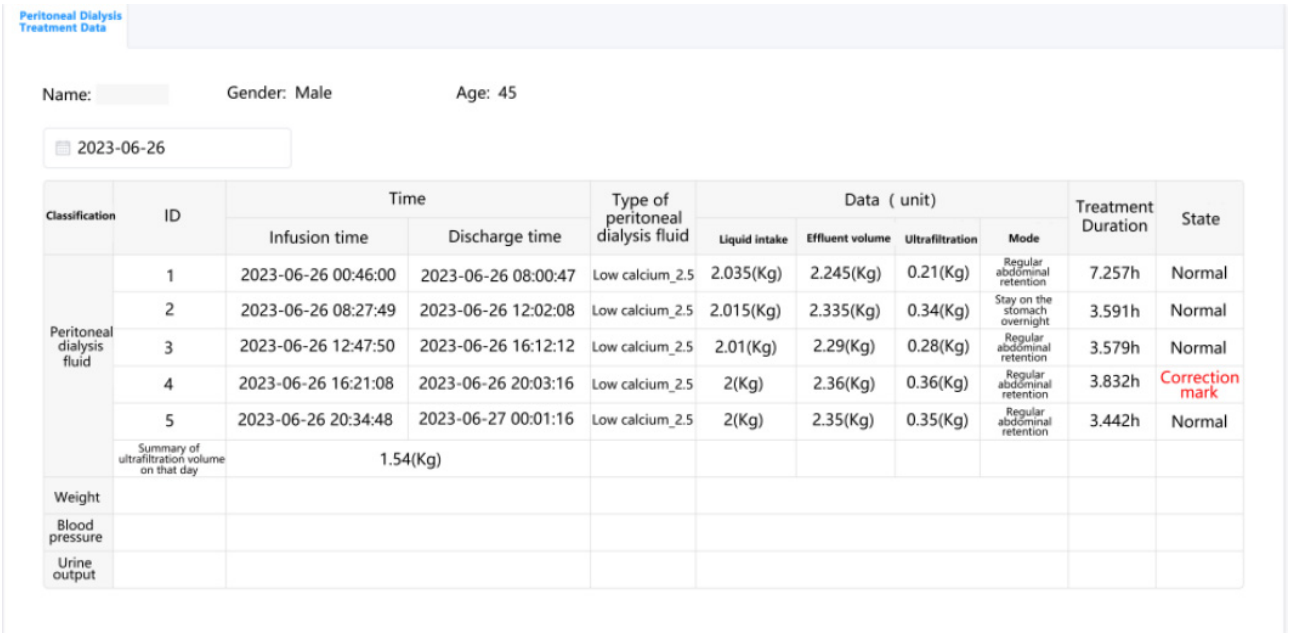
1. Password change and retrieve function: newly enrolled patients will have a default password, but patients can change the password according to personal needs. If patients forget the password, they can request password retrieval.
2. Bluetooth device connection management function: The patient can see the connection status of the Bluetooth-enabled scales on Futou Bao. When the patients encounter problems when connecting the Bluetooth device, the system will prompt the corresponding reason such as Bluetooth not enabled or connection timeout to help the patient solve the connection problem.

Patients' Data Management Platform for Doctors

The management platform for doctors contains 4 function modules: patient management module, data visualization and marking module, data statistics module, and system management module (Figures 5-9).

XSL•FO
RenderX

Figure 5. Management of patients' daily peritoneal dialysis records, including real-time data of peritoneal dialysis fluid, weight, blood pressure, and urine output.



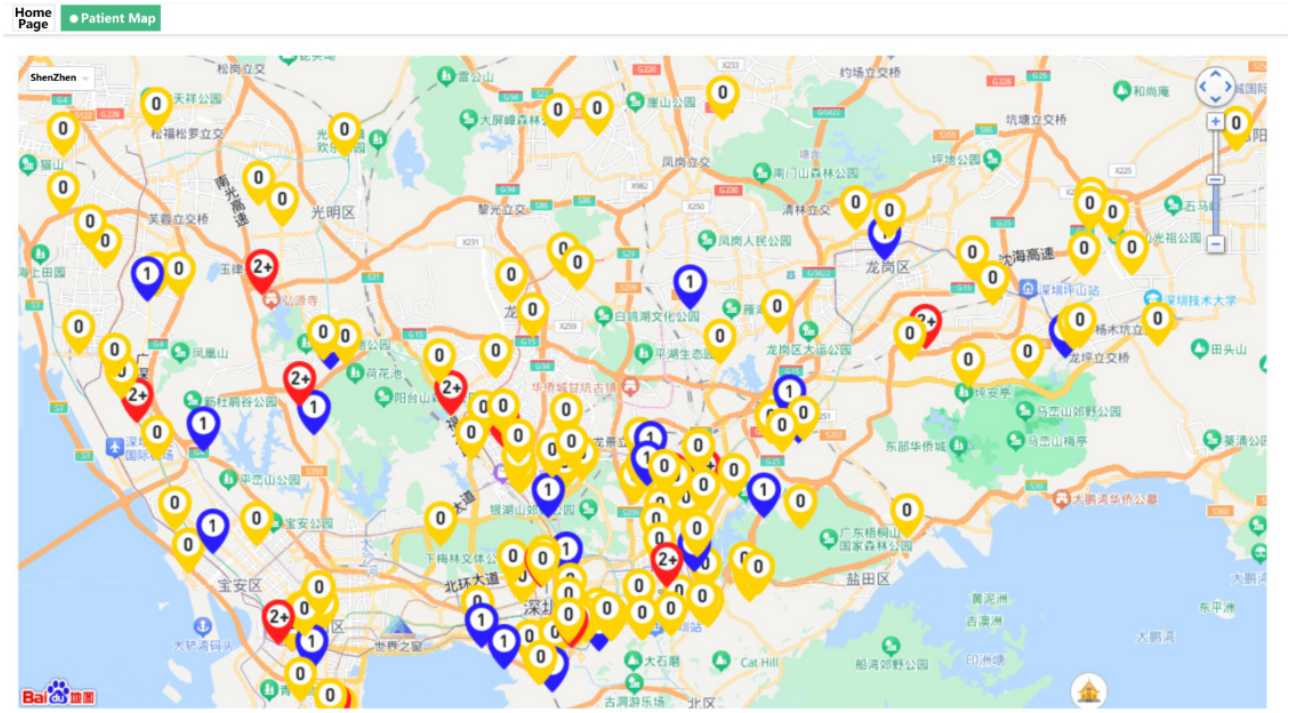
Peritoneal Dialysis Treatment Data

Name: Gender: Male Age: 45

2023-06-26

Classification	ID	Time		Type of peritoneal dialysis fluid	Data (unit)				Treatment Duration	State
		Infusion time	Discharge time		Liquid intake	Effluent volume	Ultrafiltration	Mode		
Peritoneal dialysis fluid	1	2023-06-26 00:46:00	2023-06-26 08:00:47	Low calcium_2.5	2.035(Kg)	2.245(Kg)	0.21(Kg)	Regular abdominal retention	7.257h	Normal
	2	2023-06-26 08:27:49	2023-06-26 12:02:08	Low calcium_2.5	2.015(Kg)	2.335(Kg)	0.34(Kg)	Stay on the stomach overnight	3.591h	Normal
	3	2023-06-26 12:47:50	2023-06-26 16:12:12	Low calcium_2.5	2.01(Kg)	2.29(Kg)	0.28(Kg)	Regular abdominal retention	3.579h	Normal
	4	2023-06-26 16:21:08	2023-06-26 20:03:16	Low calcium_2.5	2(Kg)	2.36(Kg)	0.36(Kg)	Regular abdominal retention	3.832h	Correction mark
	5	2023-06-26 20:34:48	2023-06-27 00:01:16	Low calcium_2.5	2(Kg)	2.35(Kg)	0.35(Kg)	Regular abdominal retention	3.442h	Normal
	Summary of ultrafiltration volume on that day	1.54(Kg)								
Weight										
Blood pressure										
Urine output										

Figure 6. Map of patients.



XSL•FO
RenderX

Figure 7. Scatterplot of the peritoneal dialysis solution weight during the treatment.

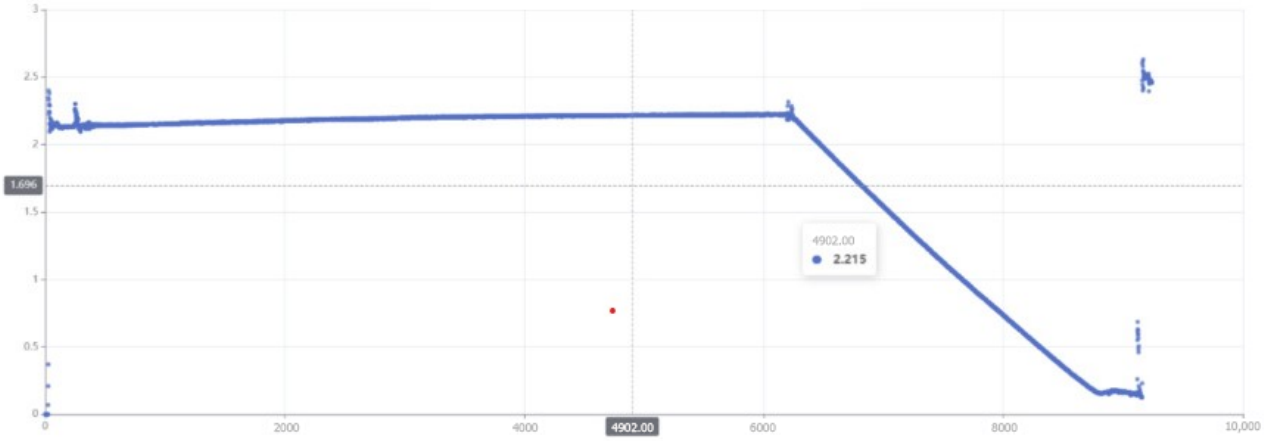


Figure 8. Peritoneal dialysis solution weight data at each treatment time point.

Data point

State	ID	Time	Weight	Manual marking ID	Manual marking Weight(kg)	Edit
Start point	30	2023-06-27 08:10:59	2.225			Manual location
Flushing point	32	2023-06-27 08:11:00	2.38			Manual location
Starting point of liquid entry	6217	2023-06-27 08:36:25	2.21			Manual location
End point of liquid entry	8769	2023-06-27 08:46:54	0.165			Manual location
Sealing starting point	8769	2023-06-27 08:46:54	0.165			Manual location
Sealing end point	9146	2023-06-27 08:48:27	0.165			Manual location
Hanging waste liquid point	9177	2023-06-27 08:48:35	2.505			Manual location

Figure 9. Statistics of daily and weekly peritoneal dialysis frequency and proportion of effective operations.

Name	Daily peritoneal dialysis frequency	Daily operational efficiency	Number of peritoneal dialysis attempts this week	Effective operation this week	Number of peritoneal dialysis attempts last week	Last week's operation efficiency
[redacted]	5.0	60.0%	7.0	85.714%	33.0	93.939%
[redacted]	3.0	0.0%	4.0	0.0%	21.0	0.0%
[redacted]	3.0	100.0%				
[redacted]	5.0	100.0%				
[redacted]	3.0	100.0%				

Patient Management Module

There are 3 main functions in this module. The first function is that doctors can view and manage the patients' basic information, including age, sex, date of first dialysis, and time duration of peritoneal dialysis treatment. The second function is the management of patients' daily peritoneal dialysis records, in which doctors can query and view patients' specific treatment records according to the treatment date (Figure 5). Based on the treatment records, doctors can know the regularity and operation of patients' peritoneal dialysis at home. The third function in this module can provide a map of the patients' addresses, which can help doctors to identify and help patients with operational problems during peritoneal dialysis in a timely manner (Figure 6).

Peritoneal Dialysis Data Visualization and Marking Module

In this module, the data visualization function can generate scatterplots of peritoneal dialysis solution weights based on the data uploaded by the spring scale or flat scale during the treatment (Figure 7). In addition to the scatterplots, doctors can also view the peritoneal dialysis solution weight data of each treatment time point, including liquid inlet starting point, liquid inlet end point, tube sealing starting point, tube sealing end point, and waste liquid hanging point (Figure 8). By viewing patient data and charts, doctors can observe patients' peritoneal dialysis operation habits, manually mark patients' abnormal operations and causes, and display corresponding prompts to patients through Futou Bao to correct possible wrong operations. Common marks for abnormal data include forgetting to weigh the waste liquid, only recording liquid output phase data, lack

XSL•FO
RenderX

of liquid output phase data, and irregular data (which may be caused by randomly pulling the spring scale).

Data Statistics Module

The data statistics module consists of 4 main functions: follow-up statistics, effective operation statistics, operation time statistics of each treatment point, and comparative statistics of treatment data in recent days. First, follow-up statistics can check the number of patients' follow-up visits, the number of new Futou Bao users, and the number of withdrawals in each year. Second, effective operation statistics function can be used to check the number of peritoneal dialysis treatments and effective operations on a given date. The effective operation rate is calculated to evaluate the overall situation of operational capability and correctness. This function makes it convenient for doctors to correct operational errors for patients (Figure 9). Third, the operation time statistics of each treatment point can be effectively viewed together with the above function. The total operation time, hand hygiene time, infusion time, and tube sealing time of the patients can be viewed. Fourth, comparative statistics of treatment data in recent days function can provide doctors with recent days' important process data during peritoneal dialysis and mark the trends of data (ie, rising or declining), which can make it easier for doctors to know the effect of peritoneal dialysis treatment.

System Management Module

The system provides permission control menus based on the role-based access control method. The super administrator assigns management rights to different roles among medical staff. This process can fully ensure the meticulous management of patients by medical staff and protect the privacy of patients.

Prediction Model of Heart Failure

General Information of the Study Participants

During the time period of this study, records of 6635 patients who underwent peritoneal dialysis treatment were uploaded in Futou Bao with 0.71% (47/6635) showing heart failure. Based on the 1:6 random sampling, treatment data of 329 patients were included in the modeling analysis, that is, 47 (14.3%) patients with heart failure and 282 (85.7%) patients without heart failure. Table 1 shows the detailed characteristics of the included patients. Most of them were females (228/329, 69.3%), with a median age of 46.55 years and a median peritoneal dialysis time duration of 0.94 years. According to the ratio of 8:2, the study dataset was randomly divided into the training dataset and the verification dataset, which included 264 patients and 65 patients, respectively (Figure 10).

XSL•FO
RenderX

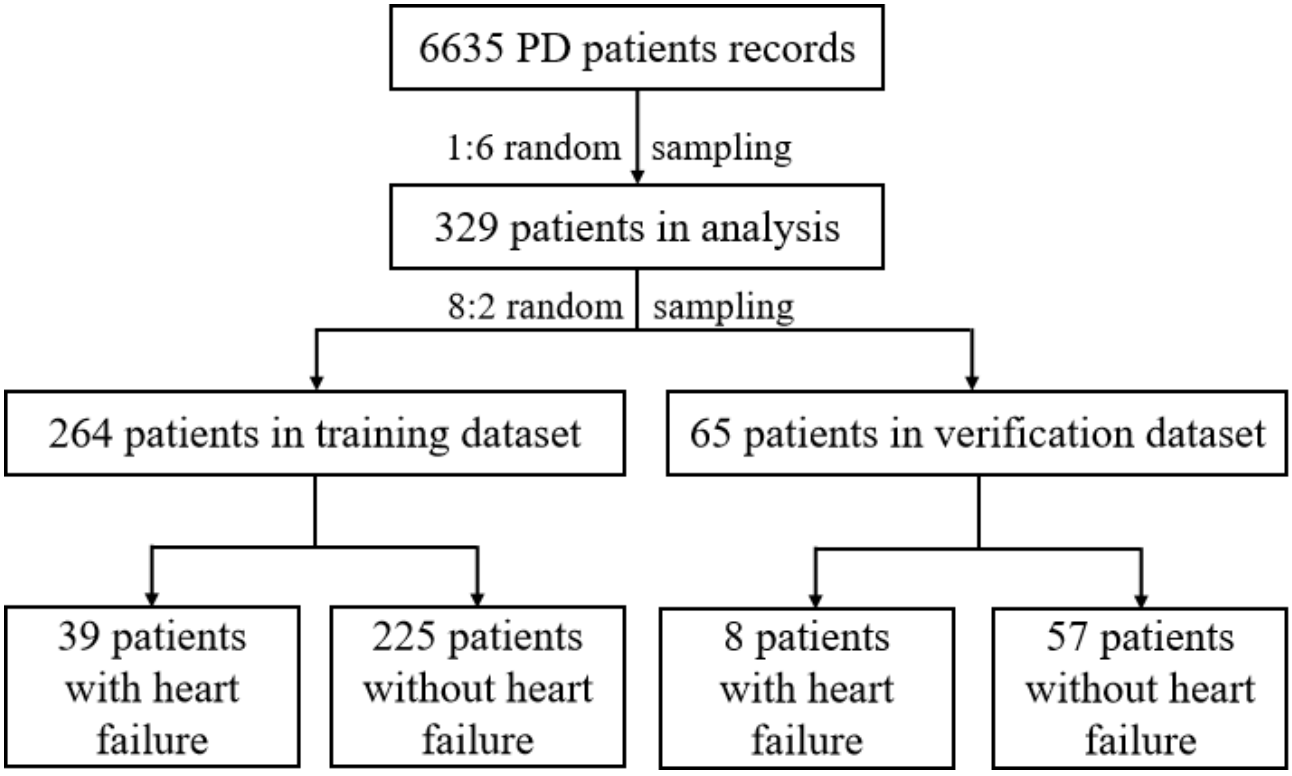
Table 1. Basic characteristics of the patients using Futou Bao.

Characteristics	All patients (N=6635)	Patients included in the analysis		
		Total (n=329)	Training set (n=264)	Validation set (n=65)
Sex, n (%)				
Male	2578 (38.85)	101 (30.70)	79 (29.92)	22 (33.85)
Female	4057 (61.15)	228 (69.30)	185 (70.08)	43 (66.15)
Age (years), median (IQR)	46.90 (44.68, 50.59)	46.55 (43.96, 51.19)	46.48 (43.94, 50.82)	46.91 (44.04, 52.68)
Time duration of peritoneal dialysis (years), median (IQR)	1.00 (0.53, 1.64)	0.94 (0.49, 1.78)	0.91 (0.46, 1.77)	1.08 (0.53, 1.87)
Weight, n (%)				
Based on 95% CI				
Normal	6115 (92.16)	299 (90.88)	239 (90.53)	60 (92.31)
Abnormal	520 (7.84)	30 (9.12)	25 (9.47)	5 (7.69)
Based on 99% CI				
Normal	6458 (97.33)	320 (97.26)	257 (97.35)	63 (96.92)
Abnormal	177 (2.67)	9 (2.74)	7 (2.65)	2 (3.08)
Urine volume, n (%)				
Based on 95% CI				
Normal	6089 (91.77)	298 (90.58)	241 (91.29)	57 (87.69)
Abnormal	546 (8.23)	31 (9.42)	23 (8.71)	8 (12.31)
Based on 99% CI				
Normal	6384 (96.22)	313 (95.14)	252 (95.45)	61 (93.85)
Abnormal	251 (3.78)	16 (4.86)	12 (4.55)	4 (6.15)
Systolic blood pressure, n (%)				
Based on 95% CI				
Normal	6189 (93.28)	299 (90.88)	239 (90.53)	60 (92.31)
Abnormal	446 (6.72)	30 (9.12)	25 (9.47)	5 (7.69)
Based on 99% CI				
Normal	6458 (97.33)	314 (95.44)	251 (95.08)	63 (96.92)
Abnormal	177 (2.67)	15 (4.56)	13 (4.92)	2 (3.08)
Diastolic blood pressure, n (%)				
Based on 95% CI				
Normal	6326 (95.34)	313 (95.14)	251 (95.08)	62 (95.38)
Abnormal	309 (4.66)	16 (4.86)	13 (4.92)	3 (4.62)
Based on 99% CI				
Normal	6515 (98.19)	319 (96.96)	256 (96.97)	63 (96.92)
Abnormal	120 (1.81)	10 (3.04)	8 (3.03)	2 (3.08)
Pulse pressure difference, n (%)				
Based on 95% CI				
Normal	6232 (93.93)	299 (90.88)	238 (90.15)	61 (93.85)
Abnormal	403 (6.07)	30 (9.12)	26 (9.85)	4 (6.15)
Based on 99% CI				
Normal	6478 (97.63)	316 (96.05)	254 (96.21)	62 (95.38)
Abnormal	157 (2.37)	13 (3.95)	10 (3.79)	3 (4.62)

XSL•FO
RenderX

Characteristics	All patients (N=6635)	Patients included in the analysis		
		Total (n=329)	Training set (n=264)	Validation set (n=65)
Heart failure, n (%)				
Yes	47 (0.71)	47 (14.29)	39 (14.77)	8 (12.31)
No	6588 (99.29)	282 (85.71)	225 (85.23)	57 (87.69)

Figure 10. Data processing flowchart. PD: peritoneal dialysis.



Optimal Prediction Model Selection

Four prediction models based on logistic regression were constructed using the training dataset (Tables 2-3). The AIC values of the 4 models were 147.23, 138.62, 145.18, and 136.50, respectively. The AUC value of model 4 (the lowest AIC value) was higher than that of model 1 (the highest AIC value) ($AUC_{Model\ 4}$ vs $AUC_{Model\ 1}$: 0.896 vs 0.893); however, the difference was not statistically significant (*P*=.78) (Figure 11).

Based on the above comparison, model 4 was considered as the optimal prediction model. Sex, age, and diastolic blood pressure were included in model 4. Females were less likely to experience heart failure than males (adjusted odds ratio [AOR] 0.06, 95% CI 0.01-0.22), while older age (AOR 1.19, 95% CI 1.13-1.25) and abnormal diastolic blood pressure during treatment (AOR 38.79, 95% CI 5.08-291.63) were the risk factors of heart failure (Table 3).

XSL•FO
RenderX

Table 2. Impact factors of heart failure in patients receiving peritoneal dialysis treatment (abnormal value based on 95% CI).

Characteristics	Univariable model		Multivariable model 1		Multivariable model 2	
	Odds ratio (95% CI)	*P* value	AOR[a] (95% CI)	*P* value	AOR (95% CI)	*P* value
Sex						
Male	1.00	N/A[b]	1.00	N/A	N/A	N/A
Female	0.10 (0.02-0.35)	.002	0.05 (0.01-0.22)	<.001	0.05 (0.01-0.26)	<.001
Age	1.17 (1.12-1.23)	<.001	1.18 (1.11-1.28)	<.001	1.17 (1.12-1.23)	<.001
Time duration of peritoneal dialysis	2.73 (1.82-4.32)	<.001	0.92 (0.42-1.80)	.83	N/A	N/A
Weight						
Normal	1.00	N/A	1.00	N/A	N/A	N/A
Abnormal	0.48 (0.07-1.70)	.33	1.24 (0.15-6.22)	.81	N/A	N/A
Urine volume						
Normal	1.00	N/A	1.00	N/A	N/A	N/A
Abnormal	3.61 (1.36-9.06)	.007	1.54 (0.31-7.84)	.59	N/A	N/A
Systolic blood pressure						
Normal	1.00	N/A	1.00	N/A	N/A	N/A
Abnormal	4.83 (1.94-11.68)	.001	4.49 (0.62-45.50)	.17	8.02 (2.09-30.77)	.002
Diastolic blood pressure						
Normal	1.00	N/A	1.00	N/A	N/A	N/A
Abnormal	2.74 (0.71-8.93)	.11	2.31 (0.22-17.93)	.44	N/A	N/A
Pulse pressure difference						
Normal	1.00	N/A	1.00	N/A	N/A	N/A
Abnormal	2.37 (0.87-5.59)	.07	1.05 (0.11-6.83)	.96	N/A	N/A

[a]AOR: adjusted odds ratio.

[b]N/A: not applicable.

XSL•FO
RenderX

Table 3. Impact factors of heart failure in patients receiving peritoneal dialysis treatment (abnormal value based on 99% CI).

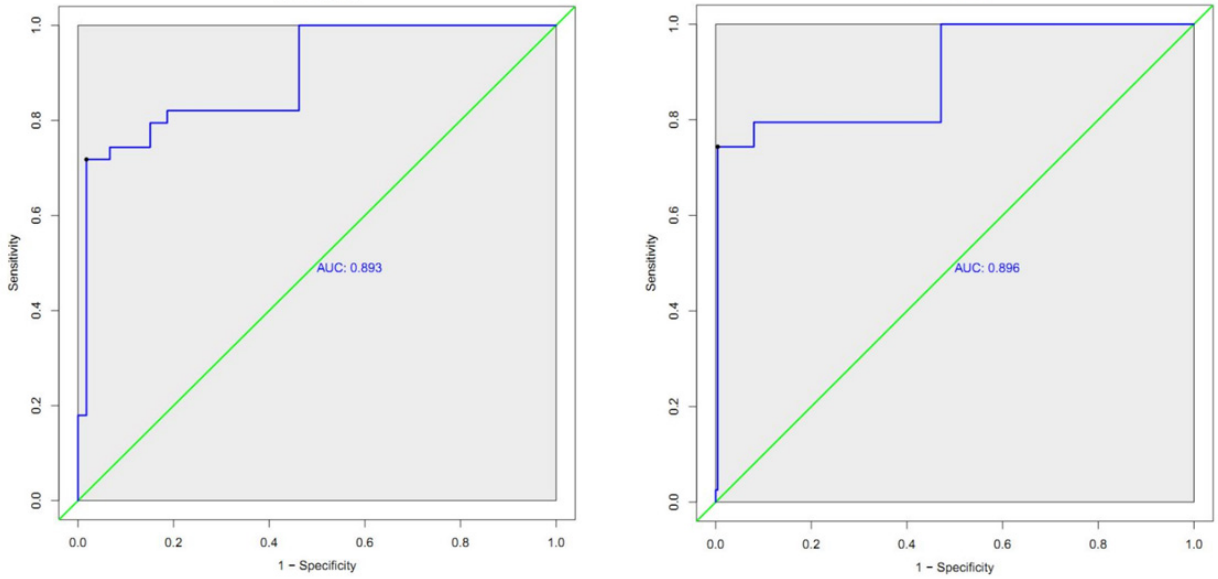
Characteristics	Univariable model		Multivariable model 3		Multivariable model 4	
	Odds ratio (95% CI)	*P* value	AOR[a] (95% CI)	*P* value	AOR (95% CI)	*P* value
Sex						
Male	1.00	N/A[b]	1.00	N/A	1.00	N/A
Female	0.10 (0.02-0.35)	.002	0.04 (0.01-0.22)	.001	0.06 (0.01-0.22)	<.001
Age	1.17 (1.12-1.23)	<.001	1.20 (1.12-1.30)	<.001	1.19 (1.13-1.25)	<.001
Time duration of peritoneal dialysis	2.73 (1.82-4.32)	<.001	0.95 (0.41-1.89)	.89	N/A	N/A
Weight						
Normal	1.00	N/A	1.00	N/A	N/A	N/A
Abnormal	2.38 (0.33-11.49)	.31	4.34 (0.23-44.51)	.25	N/A	N/A
Urine volume						
Normal	1.00	N/A	1.00	N/A	N/A	N/A
Abnormal	3.10 (0.79-10.41)	.08	1.19 (0.11-12.44)	.88	N/A	N/A
Systolic blood pressure						
Normal	1.00	N/A	1.00	N/A	N/A	N/A
Abnormal	3.99 (1.15-12.69)	.02	1.17 (0.04-18.27)	.91	N/A	N/A
Diastolic blood pressure						
Normal	1.00	N/A	1.00	N/A	1.00	N/A
Abnormal	6.31 (1.44-27.83)	.01	9.87 (0.07-290.03)	.41	38.79 (5.08-291.63)	<.001
Pulse pressure difference						
Normal	1.00	N/A	1.00	N/A	N/A	N/A
Abnormal	4.17 (1.02-15.35)	.03	1.35 (0.01-74.38)	.90	N/A	N/A

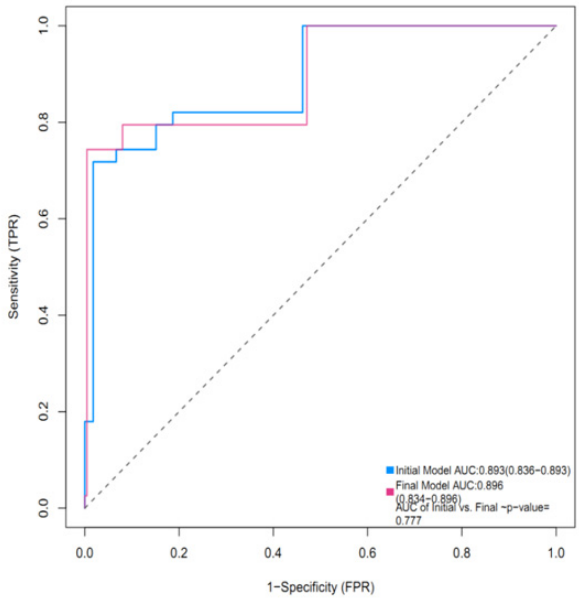
[a]AOR: adjusted odds ratio.

[b]N/A: not applicable.

XSL•FO
RenderX

Figure 11. Receiver operating characteristic curves of 2 heart failure prediction models in the training dataset. (A) Initial model (model 1). (B) Final model (model 4). (C) Model comparison. AUC: area under the curve; FPR: false positive rate; TPR: true positive rate.





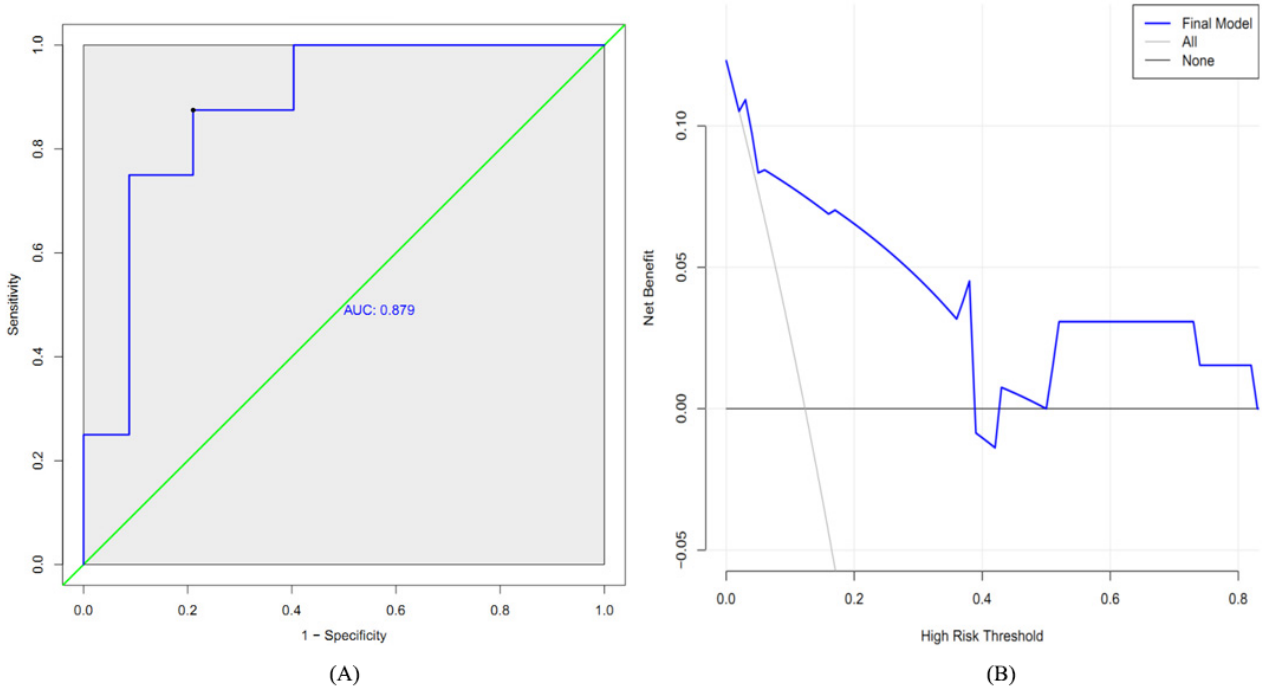
(C)

Validation of the Optimal Prediction Model

Based on the validation dataset, the sensitivity, specificity, accuracy, and Youden index of the optimal model were 0.75, 0.91, 0.89, and 0.66, respectively. Figure 12A shows the ROC curve of the model, of which the AUC value was 0.879 (95% CI 0.772-0.986). The decision curve also showed the acceptable performance of the model (Figure 12B).

XSL•FO
RenderX

Figure 12. Performance of the final model in the testing dataset. (A) Receiver operating characteristic curve. (B) Decision curve. AUC: area under the curve.



Sensitivity Analysis

In the sensitivity analysis, after changing the ratio of patients with and without heart failure (1:4 and 1:10), the analysis showed stable results. The variables included in the optimal model were constant, including sex, age, and diastolic blood pressure during treatment (Tables S1-S2 of Multimedia Appendix 1). In validation, the models also showed good performance. The AUC values were 0.877 (95% CI 0.774-0.980) in the 1:4 ratio dataset and 0.913 (95% CI 0.825-0.999) in the 1:10 ratio datasets, respectively (Figures S1-S2 in Multimedia Appendix 1). The Youden indexes were 0.76 and 0.68 in the 2 datasets, respectively.

Discussion

Principal Findings

In this study, a practical and user-friendly mHealth tool named Futou Bao and its supporting management system were developed for patients with ESRD who managed their peritoneal dialysis treatment at home in China. With its comprehensive functions that provided intelligent data upload, visualization, management, and doctor-patient interaction, the mHealth tool and system could effectively solve the difficulties in managing peritoneal dialysis treatment at home, increase treatment effectiveness, and minimize complication risk. Compared with the CKD-PD (chronic kidney disease-peritoneal dialysis) app [12] developed by researchers in the Khon Kaen University, Thailand, in the same period as our study, Futou Bao had several additional advantages. First, automatic data upload via Bluetooth could reduce data recording errors. Second, the recording of peritoneal dialysis solution weight data was more detailed, which included the weight data at each time point during the whole treatment process, and it could generate data visualization plots timely. Third, Futou Bao could provide better interactions between doctors and patients. The operation problems marked by doctors can be notified to the patients through the platform in a timely manner to help patients improve the peritoneal dialysis operation quality. Fourth, the Futou Bao system was able to locate patients' addresses to detect serious health risks and provide timely medical assistance. In addition, a previous study conducted among Thai patients using the CKD-PD app mainly focused on the optimization of the app's functionality and usefulness with only 10 participants involved [12]. By contrast, in this study, a larger number of patients were included to explore a better way for assisting clinical decision-making with the help of constructing prediction models based on the usage records of Futou Bao. We found that most of the patients with ESRD using Futou Bao, as an eHealth management and promoting tool, in our study were females, which was different from the real-world clinical experiences, wherein most patients treated for ESRD with any form of dialysis were males. However, in the process of recruitment, we only relied on patients' voluntary participation in the study and did not induce them in any form or select any of them based on their sex. This interesting phenomenon can be explained by the higher acceptance of the eHealth approaches among females. A study focusing on drug use treatment in China using a mobile app also showed that female users were more willing to receive health care messages and communicate with health workers through the mobile app and were more likely to have positive health changes than male users, which could support our finding [13].

Given the importance of early detection of heart failure in patients receiving home-based peritoneal dialysis, it is of great value to make full use of the data recorded in the process of

XSL•FO
RenderX

daily peritoneal dialysis treatment to perform real-time prediction and for early warning of heart failure risk. However, the variables recorded during the daily home-based peritoneal dialysis were limited to some extent. In previous studies, the prediction models of heart diseases among patients with CKD and ESRD mainly relied on clinical test indicators and biomarkers that could not be tested at home. Kessler et al [14] built a risk prediction model of cardiovascular events for patients with ESRD based on the FOSIDIAL trial and found that C-reactive protein and left ventricular mass index were important prognostic factors of cardiovascular events. Weekley and Peralta [15] indicated that combining creatinine, cystatin C, and albuminuria could improve the detection and risk stratification for heart failure and other cardiovascular events. Other biomarkers, including estimated glomerular filtration rate, urine albumin-to-creatinine ratio, and β_2-microglobulin, were also proved to be associated with heart failure in patients with CKD [16-20]. Based on evidences from the Chronic Renal Insufficiency Cohort study, a Pooled Cohort Equations to Prevent Heart Failure in the CKD population was constructed to predict incident heart failure hospitalizations [21]. A series of clinical factors were included in this model, such as blood pressure, fasting glucose, total cholesterol, high-density lipoprotein cholesterol, QRS duration, estimated glomerular filtration rate, and albuminuria [21]. A recent research based on the data of a Chronic Renal Insufficiency Cohort study indicated that adding N-terminal brain natriuretic peptide and high sensitivity troponin-T to the prediction model could improve the discrimination of heart failure [22].

Considering the unavailability of laboratory test results in daily home-based peritoneal dialysis and the lack of heart failure prediction model in patients receiving peritoneal dialysis treatment, our study innovatively establishes a practical prediction model of heart failure in patients undergoing peritoneal dialysis based on the monitoring data of daily peritoneal dialysis treatment at home. The factors in the prediction model included sex, age, time duration of peritoneal dialysis, weight, urine volume, and blood pressure, which were accessible and real-time data monitored and recorded by the Futou Bao. In addition, the effectiveness of the prediction model based on routine monitoring variables was acceptable. In the model validation, the accuracy of the prediction model could reach 89%, with sensitivity of 75% and higher specificity of 91%. Therefore, the mHealth tools and the predictive model developed in this study enabled monitoring and early warning of heart failure risk, which has formed a new paradigm for the management of patients receiving home-based peritoneal dialysis treatment. From another perspective, during the 5-year follow-up, the proportion of patients experiencing heart failure in our study was less than 1%, which was much lower than that reported in previous studies [7-10]. This result could indicate that our management paradigm was practical and effective.

Limitations

There were several limitations in this study. First, the number of patients with heart failure during the observation period of this study was small, which resulted in a relatively small number of patients in the dataset for the prediction model study. Volunteer bias may enhance the management effectiveness of the system. In future studies, we will conduct randomized controlled trials to better rule out confounding factors and reduce the occurrence of bias. Second, the mHealth tools were only used for daily data collection and monitoring, and the exclusion criteria were relatively strict that patients with previous heart failure or coronary heart disease history were not included in this study. To better expand the coverage of patients included and managed in the future, more variables in their medical history should be added to better predict the risk of heart failure. Third, the mHealth system was only used in the scenario of large cities represented by Shenzhen. In future studies, the regional scope of the system should be further expanded. In addition, the sex and age composition of the users was different from that in clinical practice experiences. In the future, the performance of Futou Bao should be promoted in the following 3 ways: (1) further investigation is needed to explore the obstacles in male users' acceptance of Futou Bao, and the functional design should consider the needs of both male and female users; (2) the improvement of Futou Bao should enhance the usage experience of older adults to help them use the tool more conveniently; and (3) on the basis of ensuring patient privacy, family caregivers can be authorized to use the patient platform so that more older adults who do not have the ability to use smartphones can benefit from using this home-based system.

Conclusions

This is a pioneering study that describes the development of an intelligent home-based peritoneal dialysis management and monitoring system based on mHealth technology in China. This system consists of a WeChat Mini Program named Futou Bao for patients, which includes an intelligent data upload function module and an auxiliary function module; and a patient data management platform for doctors, which includes 4 function modules: patient management, data visualization and marking, data statistics, and system management. Based on the daily peritoneal dialysis treatment data recorded in the system, a real-time prediction model of heart failure in patients receiving peritoneal dialysis treatment was built and verified. This accuracy of the prediction model was accepted, with the AUC value of 0.879 (95% CI 0.772-0.986). The new home-based peritoneal dialysis management paradigm that we developed enabled real-time monitoring and early warning of heart failure risk. In the future, it is essential to evaluate the effectiveness of mHealth tools through multicenter studies, which should include regions with different economic levels and medical and health service levels to improve the universality of the research results.

Acknowledgments

We thank the staff of Shenzhen Hospital of Traditional Chinese Medicine for their contributions in patient management. This work was supported by the Health and Family Planning Commission of Shenzhen Municipality (SZLY2018004); the Guangdong

XSL•FO
RenderX

Medical Science and Technology Research Fund Project, China (B2019043); the Key Laboratory of Medical Rescue Key Technology and Equipment, Ministry of Emergency Management (Open Fund Project YJBKFKT202415); the Natural Science Foundation of Tianjin (23JCQNJC01430); and the National Natural Science Foundation of China (72274210).

Data Availability

The datasets in this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: JW, XY, Y Zeng, JY, and ZJ
Data curation: YY, XX, GS, WY, TW, and Y Zheng
Formal analysis: JW, Y Zeng, XY
Funding acquisition: JW, Y Zeng, XY, ZJ
Methodology: JW, Y Zeng, XY
Project administration: Y Zeng, XY, ZJ
Resources: Y Zeng, JW
Supervision: XY, ZJ
Validation: XY, Y Zeng, JY
Visualization: JW, Y Zeng, XY
Writing – original draft: JW, XY, Y Zeng, JY
Writing – review & editing: XY, ZJ

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary data.
[DOCX File , 150 KB - formative_v9i1e56254_app1.docx]

References

1. GBD Chronic Kidney Disease Collaboration. Global, regional, and national burden of chronic kidney disease, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2020 Feb 29;395(10225):709-733 [FREE Full text] [doi: 10.1016/S0140-6736(20)30045-3] [Medline: 32061315]
2. Vanholder R, Annemans L, Brown E, Gansevoort R, Gout-Zwart JJ, Lameire N, European Kidney Health Alliance. Reducing the costs of chronic kidney disease while delivering quality health care: a call to action. Nat Rev Nephrol 2017 Jul;13(7):393-409. [doi: 10.1038/nrneph.2017.63] [Medline: 28555652]
3. Liyanage T, Ninomiya T, Jha V, Neal B, Patrice HM, Okpechi I, et al. Worldwide access to treatment for end-stage kidney disease: a systematic review. The Lancet 2015 May;385(9981):1975-1982. [doi: 10.1016/s0140-6736(14)61601-9]
4. Moncrief J. The birth and development of continuous ambulatory peritoneal dialysis. Contrib Nephrol 2017;189:85-90. [doi: 10.1159/000450689] [Medline: 27951554]
5. Li PK, Chow KM, Van de Luijtgaarden MW, Johnson DW, Jager KJ, Mehrotra R, et al. Changes in the worldwide epidemiology of peritoneal dialysis. Nat Rev Nephrol 2017 Feb;13(2):90-103. [doi: 10.1038/nrneph.2016.181] [Medline: 28029154]
6. Teitelbaum I. Peritoneal dialysis. N Engl J Med 2021 Nov 04;385(19):1786-1795. [doi: 10.1056/nejmra2100152]
7. Hou F, Jiang J, Chen J, Yu X, Zhou Q, Chen P, et al. China collaborative study on dialysis: a multi-centers cohort study on cardiovascular diseases in patients on maintenance dialysis. BMC Nephrol 2012 Aug 30;13:94 [FREE Full text] [doi: 10.1186/1471-2369-13-94] [Medline: 22935444]
8. Wang AY, Wang M, Lam CW, Chan IH, Lui S, Sanderson JE. Heart failure with preserved or reduced ejection fraction in patients treated with peritoneal dialysis. Am J Kidney Dis 2013 Jun;61(6):975-983. [doi: 10.1053/j.ajkd.2012.12.030] [Medline: 23474006]
9. Wang I, Lu C, Lin C, Liang C, Yen T, Liu Y, et al. Comparison of the risk of de novo cardiovascular disease between hemodialysis and peritoneal dialysis in patients with end-stage renal disease. Int J Cardiol 2016 Sep 01;218:219-224. [doi: 10.1016/j.ijcard.2016.05.036] [Medline: 27236118]
10. Sun C, Sung J, Wang J, Li C, Kuo Y, Lee C, et al. A comparison of the risk of congestive heart failure-related hospitalizations in patients receiving hemodialysis and peritoneal dialysis - A retrospective propensity score-matched study. PLoS One 2019;14(10):e0223336 [FREE Full text] [doi: 10.1371/journal.pone.0223336] [Medline: 31574134]
11. J W, Z J, M P, T Y. Real-time monitoring and reminding of remote peritoneal dialysis system based on the principle of least squares. Int J Nephrol Kidney Fail 2021;7(3):1-6. [doi: 10.16966/2380-5498.214]

XSL•FO
RenderX

12. Lukkanalikitkul E, Kongpetch S, Chotmongkol W, Morley MG, Anutrakulchai S, Srichan C, et al. Optimization of the chronic kidney disease-peritoneal dialysis app to improve care for patients on peritoneal dialysis in Northeast Thailand: user-centered design study. JMIR Form Res 2022 Jul 06;6(7):e37291 [FREE Full text] [doi: 10.2196/37291] [Medline: 35793137]
13. Li Y, Yan X, Wang Z, Zhang B, Jia Z. Clear the fog of negative emotions: A new challenge for intervention towards drug users. J Affect Disord 2021 Nov 01;294:305-313. [doi: 10.1016/j.jad.2021.07.029] [Medline: 34311330]
14. Kessler M, Zannad F, Lehert P, Grünfeld JP, Thuilliez C, Leizorovicz A, FOSIDIAL Investigators. Predictors of cardiovascular events in patients with end-stage renal disease: an analysis from the Fosinopril in dialysis study. Nephrol Dial Transplant 2007 Dec;22(12):3573-3579. [doi: 10.1093/ndt/gfm417] [Medline: 17611249]
15. Weekley C, Peralta C. Advances in the use of multimarker panels for renal risk stratification. Curr Opin Nephrol Hypertens 2012 May;21(3):301-308 [FREE Full text] [doi: 10.1097/MNH.0b013e328352132d] [Medline: 22449944]
16. Matsushita K, Ballew SH, Coresh J. Cardiovascular risk prediction in people with chronic kidney disease. Current Opinion in Nephrology and Hypertension 2016;25(6):518-523. [doi: 10.1097/mnh.0000000000000265]
17. Matsushita K, Coresh J, Sang Y, Chalmers J, Fox C, Guallar E, et al. Estimated glomerular filtration rate and albuminuria for prediction of cardiovascular outcomes: a collaborative meta-analysis of individual participant data. The Lancet Diabetes & Endocrinology 2015 Jul;3(7):514-525. [doi: 10.1016/s2213-8587(15)00040-6]
18. Foster MC, Coresh J, Hsu CY, et al. Erratum regarding "serum β-trace protein and β2-microglobulin as predictors of ESRD, mortality, and cardiovascular disease in adults with CKD in the Chronic Renal Insufficiency Cohort (CRIC) study. Am J Kidney Dis 2024 Jun;83(6):840. [doi: 10.1053/j.ajkd.2024.04.002] [Medline: 38782534]
19. Matsushita K, Sang Y, Ballew SH, Astor BC, Hoogeveen RC, Solomon SD, et al. Cardiac and kidney markers for cardiovascular prediction in individuals with chronic kidney disease: the Atherosclerosis Risk in Communities study. ATVB 2014 Aug;34(8):1770-1777. [doi: 10.1161/atvbaha.114.303465]
20. Liabeuf S, Lenglet A, Desjardins L, Neirynck N, Glorieux G, Lemke H, European Uremic Toxin Work Group (EUTox). Plasma beta-2 microglobulin is associated with cardiovascular disease in uremic patients. Kidney Int 2012 Dec;82(12):1297-1303 [FREE Full text] [doi: 10.1038/ki.2012.301] [Medline: 22895515]
21. Mehta R, Ning H, Bansal N, Cohen J, Srivastava A, Dobre M, et al. Ten-year risk-prediction equations for incident heart failure hospitalizations in chronic kidney disease: findings from the chronic renal insufficiency cohort study and the multi-ethnic study of atherosclerosis. J Card Fail 2022 Apr;28(4):540-550 [FREE Full text] [doi: 10.1016/j.cardfail.2021.10.007] [Medline: 34763078]
22. Zelnick LR, Shlipak MG, Soliman EZ, Anderson A, Christenson R, Kansal M, CRIC Study Investigators. Prediction of incident heart failure in CKD: The CRIC study. Kidney Int Rep 2022 Apr;7(4):708-719 [FREE Full text] [doi: 10.1016/j.ekir.2022.01.1067] [Medline: 35497796]

Abbreviations

AIC: Akaike's information criterion
AOR: adjusted odds ratio
AUC: area under the curve
CAPD: continuous ambulatory peritoneal dialysis
CKD-PD: chronic kidney disease-peritoneal dialysis
ESRD: end-stage renal disease
mHealth: mobile health

Edited by A Mavragani; submitted 11.01.24; peer-reviewed by M Papasotiriou, X Zhuang; comments to author 31.05.24; revised version received 24.07.24; accepted 03.02.25; published 03.03.25.

Please cite as:
Wu J, Zeng Y, Yang J, Yao Y, Xu X, Song G, Yi W, Wang T, Zheng Y, Jia Z, Yan X
Daily Treatment Monitoring for Patients Receiving Home-Based Peritoneal Dialysis and Prediction of Heart Failure Risk: mHealth Tool Development and Modeling Study
JMIR Form Res 2025;9:e56254
URL: https://formative.jmir.org/2025/1/e56254
doi:10.2196/56254
PMID:40053710

©Jia Wu, Youjia Zeng, Jun Yang, Yutong Yao, Xiuling Xu, Gaofeng Song, Wuyong Yi, Taifen Wang, Yihou Zheng, Zhongwei Jia, Xiangyu Yan. Originally published in JMIR Formative Research (https://formative.jmir.org), 03.03.2025. This is an open-access

XSL•FO
RenderX

article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Experiences With an In-Bed Real-Time Motion Monitoring System on a Geriatric Ward: Mixed Methods Study

Stefan Walzer[1], MSc; Isabel Schön[2], MA; Johanna Pfeil[2], BA; Nicola Merz[2], MA; Helga Marx[3], MSc; Sven Ziegler[3], MSc; Christophe Kunze[1], Prof Dr

[1]Care and Technology Lab, Furtwangen University, Furtwangen im Schwarzwald, Germany

[2]AGP Social Research, FIVE e.V., Freiburg, Germany

[3]Center of Implementing Nursing Care Innovations Freiburg, Nursing Direction, Medical Center, University of Freiburg, Freiburg, Germany

Corresponding Author:
Stefan Walzer, MSc
Care and Technology Lab
Furtwangen University
Robert-Gerwig-Platz 1
Furtwangen im Schwarzwald, 78120
Germany
Phone: 49 7723920295
Email: stefan.walzer@hs-furtwangen.de

Abstract

Background: Older adults now make up about two-thirds of hospital admissions, with up to 50% experiencing cognitive impairments such as dementia. These patients often struggle with adherence to care plans and maintaining regular day or night cycles, presenting challenges for nurses. Hospitals are typically unprepared to manage this patient population, resulting in increased nurse workload and challenges like managing motor agitation, which can lead to falls or accidental removal of medical devices.

Objective: This study aimed to (1) assess how an in-bed real-time motion monitoring system (IRMS) impacts nurses' perceptions of physical and mental stress, (2) evaluate the IRMS's effect on the care process, (3) explore ethical implications like patient autonomy and privacy, and (4) understand how nurses acquire knowledge about the technology and how this affects their assessment of the IRMS.

Methods: The IRMS, which provides real-time motion monitoring and bed edge or exit information, was implemented in the geriatric ward of a university medical center. The study followed a monocentric, explorative evaluation design using a mixed methods approach. It lasted 24 weeks and had two phases. In Phase 0 (6 weeks), patients received standard care. In Phase 1 (18 weeks), the IRMS was introduced. Initial data were gathered through focus groups and participant observations during manufacturer training sessions. At the end of the intervention, a survey, a second focus group, and an interview were conducted to capture nurses' experiences. The study follows the Good Reporting of a Mixed Method Study (GRAMMS) checklist for reporting.

Results: Initial training sessions with 12 participants (10 nurses and 2 physiotherapists) showed varying levels of engagement, with the second session demonstrating more optimism and interprofessional collaboration. A total of 10 questionnaires were completed (10/21, 48%). Survey results showed that 80% (8/10) of nurses found the IRMS valuable for assessing the quality of work, and 90% (9/10) were willing to continue using it. The system was regarded as reliable for monitoring bed edge and exit events. Usability was positively rated, with minimal concerns about documentation burden. Focus group discussions (n=3 per session) indicated that nurses viewed the system as reliable and appreciated its role in reducing anxiety related to fall prevention. However, concerns about patient privacy and monitoring were raised. Nurses expressed a willingness to continue using the IRMS but reaffirmed their ability to care for patients without it.

Conclusions: Nurses had a generally positive attitude toward the IRMS, recognizing its benefits, particularly for nighttime monitoring. Although its effectiveness in preventing falls remains inconclusive, the system helps reduce nurses' fear of falls and enhances their responsiveness. The study highlights the broader impact of the IRMS beyond fall prevention and stresses the importance of thoughtful integration into health care practice.

(JMIR Form Res 2025;9:e63572) doi:10.2196/63572

XSL•FO
RenderX

KEYWORDS

nurses; geriatric patients; cognitive impairment; technology; fall prevention; hospital; mixed methods; patient; learning process; assessment; autonomy; impairment; real-time motion; university; geriatric ward; survey; anxiety; willingness; patient privacy; effectiveness; monitoring system; health care practice

Introduction

Background

The global increase in emergency hospital admissions of older individuals is expected to continue, driven by ongoing demographic shifts. Currently, older adults make up approximately two-thirds of hospital inpatients, with up to 50% of this population experiencing some degree of cognitive impairment, including dementia-related conditions [1-3]. Patients with cognitive impairment (hereafter referred to as patients) often struggle to adhere to care plans in the hospital setting, and their day or night cycles are frequently disrupted. This presents additional challenges for nurses, as their workload increases, especially since hospitals are often not adequately equipped to care for patients with cognitive impairment [4]. A particular challenge for hospital nurses is motor agitation in and around the patient's bed, which can lead to incidents such as falls or accidental removal of catheters or vascular access devices [5]. In light of these challenges, there is ongoing discussion about the potential role of technological innovations in supporting nurses and caregivers [6,7].

Bed exit information systems (BES) are well-known in this context [8,9]. In a previous study by the research team [10], the focus was widened to the complex interplay of factors from the nurses' perspective. The study addressed the perception of nurses on several aspects as adverse events and the implementation in the clinical setting, as well as their assessment of usefulness. Therefore, nurses on regular wards at a university medical center used a bed exit system to better understand its potential in hospitals [10]. The system, which was placed under the mattress in the patient's bed, informed nurses via the nurse call system whether patients were mobilizing to the edge of the bed or leaving the bed unassisted. An advantage of this system was its ease of use, which provided benefits to the nursing staff. However, due to the complex and sometimes unpredictable course of symptoms, it was only possible to use it in a targeted way for a limited number of patients (patients with cognitive impairments who also have a tendency to walk and a risk of falls or may become disoriented). That technological limitations and special characteristics strongly affect the effect of a BES becomes also evident in the study of Considine et al [11], in which alarms at the bedside caused specific challenges for the nurses, including issues related to alarm fatigue, disruptions in patient care, and difficulties in prioritizing responses to alarms.

Therefore, the research team looked for another technology with a higher degree of functionality (allowing for the setting of either bed edge or bed exit notifications) and an additional function (real-time motion monitoring), with the aim of evaluating its implementation. Here as well, the research question focused on possible benefits for the nurses in patient care and the integration of the technology in the working routines. To answer this question, the nursing staff of a geriatric ward in a university hospital used the Mobility Monitor, an in-bed real-time motion monitoring system (IRMS) from the company Compliant Concept AG, as an additional aid in the standard care of patients [12]. In addition to bed edge and bed exit information, the system provides nurses with real-time monitoring to observe and analyze the patient's movement patterns in bed. In a previous evaluation project [13], the IRMS was found to have potential benefits for the nightly monitoring of patients on standard wards in general and the care of patients with delirium in particular.

Overall, previous studies have predominantly addressed single aspects of the technology, such as the efficacy and effectiveness of BES in terms of reducing falls or fall rates [14,15], and there has been limited research into nurses' perspectives and experiences of using BES [16,17].

Aim and Research Question

Guided by the results of previous studies, the purpose of this study was to better understand the positive effects and disadvantages of the use of an IRMS from a nursing point of view.

The study was designed to address the following research questions:

- How does the use of IRMS affect nurses' perception of physical and mental stress?
- How is the care process affected by the use of the IRMS?
- What kind of ethical implications could arise from the nursing perspective, such as perceptions of autonomy and privacy for patients?
- How do nurses acquire knowledge about the technology, and how does this learning process influence their assessment of it?

Methods

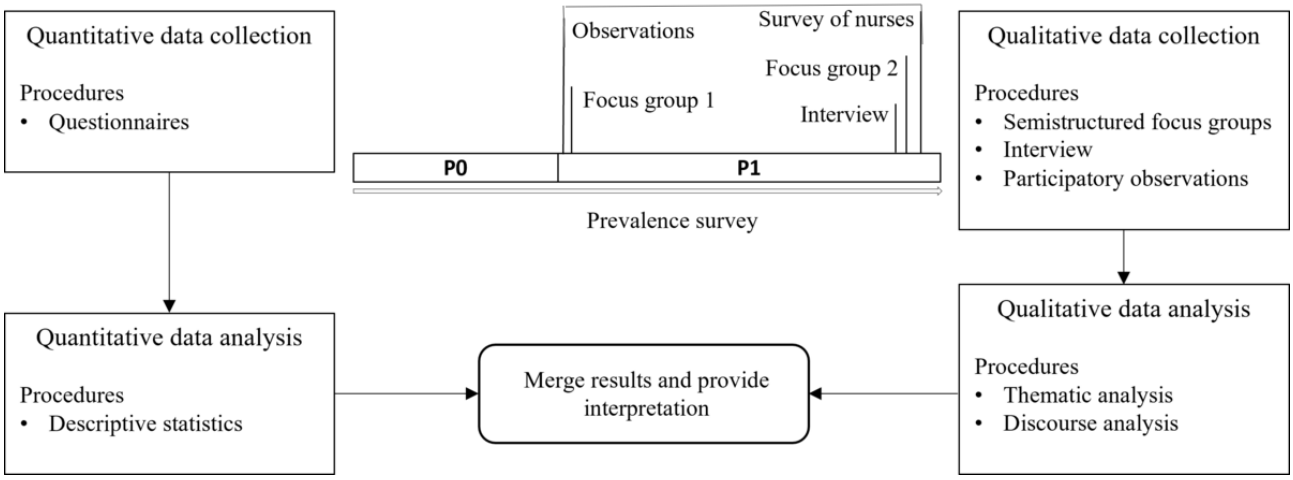
Reporting was guided by the Good Reporting of a Mixed Method Study (GRAMMS) criteria proposed by O'Cathain et al [18].

Design

This study was designed using a parallel mixed methods triangulation approach [19], incorporating a quantitative survey, qualitative focus groups, and one interview, as well as participant observations to obtain a more comprehensive understanding from multiple perspectives (Figure 1). A key principle of mixed methods research is that combining quantitative and qualitative data provides a richer understanding of the research question compared to using either data source in isolation [20]. The integration of results is realized in the Discussion section of this paper.

XSL•FO
RenderX

Figure 1. Parallel mixed methods triangulation study design.



The study was conducted in a geriatric ward at a university medical center in the south of Germany. The ward has 20 beds with an average occupancy of 18 patients at any one time. Staffing typically includes three registered nurses on the day shift and usually three on the evening shift, but occasionally two registered nurses and two nursing assistants, depending on availability. During the night shift, there are usually two staff members on duty. Three physician assistants and one supervising physician are responsible for the ward.

Spanning 24 weeks, the study period was divided into two distinct phases. Phase 0 (baseline survey P0) lasted 6 weeks. During this phase, patients were cared for according to the university medical center care standard. Phase 1 (intervention phase P1) lasted 18 weeks. For this intervention phase, the geriatric ward was equipped with the IRMS, and data about the experience of nurses were collected. The IRMS were available to nurses as an aid to patient care, assisting nurses in assessing the patients' cognitive status and tendency to get out of bed. Nurses were able to make their own decisions about which patients they would use the IRMS with. Both the (1) real-time monitoring of the patient's movement patterns in bed and (2) bed edge and bed exit information could be turned on or off by the nurses.

During the study, 22 devices were made available to the ward from the first day of the intervention phase (Phase 1), ensuring full coverage. The research team integrated the devices into the beds on that day, which also marked the first training session, with a second session conducted the following week. Preparatory work, including software installation and setup of the screens in the nursing station, was carried out in the weeks prior (Phase 0). The devices were customized by the supplier (Compliant Concept) in coordination with the University Medical Center's technical team to ensure compatibility with the call system.

The training consisted of a 2-hour session, which was held twice by a company representative for the nursing staff. While the focus was primarily on nurses, other professionals working on the ward, such as physicians and physiotherapists, were also invited to participate. The training included a presentation for the IRMS introduction, as well as practical exercises to reinforce learning. Additionally, research team members provided further assistance as needed, ensuring ongoing support for any questions or challenges the nursing team faced. Informational materials were also available to the staff, including a PowerPoint (Microsoft Corp) presentation, a practical written guideline for patient entry into the software, and the company's manual. Regarding communication with patients about the IRMS, nurses were not specifically trained to discuss its use, but it ultimately depended on their professionalism in interacting with patients.

Participants

The survey and focus group sessions involved nurses who met the predetermined criteria outlined in Textbox 1. A total of 21 nurses met the inclusion criteria at P1. Potential participants were informed about the survey, focus groups, participant observations, and interview through ward meetings and internal institutional mailing lists, supplemented by email communications. They were then formally invited to participate. Nurses who expressed an interest were given full written information about the study and the nature of their participation. The survey, focus groups, participant observation, and interview were conducted during working hours with the authorization of the respective organization and no additional incentives or rewards were provided to participants.

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- ≥18 years
- Nurses with at least three years of apprenticeship or equivalent international training with professional recognition in Germany
- Employees working on the included ward during the study period

Exclusion criteria

- Employees of the included ward who belong to other professional groups.
- Employees of the reserve pool.
- Nurses from other wards helping out
- Trainees

Data Collection

The study followed a parallel mixed methods approach with the following elements (see Figure 1 for an overview of the study design):

- Survey of nurses at time point P1.
- Focus groups with nurses at time points P0 and P1, before and after IRMS use.
- Interview with one nurse at time point P1, after IRMS use.
- Participant observations during the initial introduction of the technology on the ward by the manufacturer.

Nurse Survey

The survey was developed in REDCap (Research Electronic Data Capture; Vanderbilt University) [21]. A comprehensive review of the literature did not identify a suitable assessment tool for collecting data on this topic in the German language. The web-based survey, which consisted of 48 items, was based on the technology attitude or acceptance, burden, and expectations or experiences survey used in a previous study [10]. It primarily used items from Spagnolli et al's [22] survey on user acceptance of wearable technologies, which were translated into German by our team, and a subset of Isfort et al's [23] survey on nurses' perceived burden in patient care, which was available in German. The instrument was further supplemented with custom questions and sociodemographic data, including age, gender, and years of professional experience of the nursing staff. However, due to the small sample size, these sociodemographic data were deemed insufficient for meaningful subgroup analysis or generalizable insights. Presenting these variables could risk overinterpretation or imply trends unsupported by the data. Therefore, we chose to focus on the main findings directly related to our research questions. In previous projects using paper and pencil surveys, it was found that the nursing team preferred a digital format due to its environmental benefits, such as the elimination of paper waste. A total of 21 nurses were invited to participate via email with a link to the web-based survey, allowing them to complete the survey at their convenience. The survey began immediately after the intervention phase, starting 11 days after the IRMS had been in use for 20 weeks. The survey is included in Multimedia Appendix 1.

Focus Groups and Interview

Each focus group was conducted during the first survey phase (P0) and at the end of the second survey phase (P1) by two researchers (IS and SW) to capture the context of technology use and to further explore the issues identified in the staff survey. The initial focus group guide addressed expectations regarding the potential for alleviation, usability, and the influence on social interaction with patients (Multimedia Appendix 2). In contrast, the guide for the discussion after the intervention centered on the experiences related to these same themes. Topics included the challenges of caring for patients with cognitive impairment, the tendency of patients to get out of bed unsupervised, and the expectations, experiences, and evaluations related to the use of the IRMS (Multimedia Appendix 3). This approach aimed to derive insights into nurses' perceptions of physical and mental stress, as well as any changes resulting from the implementation of the technology. Inferences about potential alterations in the care process, patient outcomes, and the perceptions of autonomy or privacy by patients can be deduced from nurses' comments on processes, perceptions, and evaluations before, during, and after the use of the IRMS.

The two researchers facilitating the discussions both have prior experience in conducting focus groups, with one having a background in nursing science (SW) and the other in sociology (IS). A semistructured interview guide was used, focusing on the concepts of interest while allowing for flexibility. The structure and content of the guide were informed by the core findings of previous studies [10,13] and the literature reviewed (see Background under Introduction section). Since one participant from the first focus group could not participate in the second, a follow-up interview was conducted using the same guideline at P1. Both the focus group and the interview were audio-recorded and transcribed using a denaturalized approach, omitting idiosyncratic speech elements [24]. Transcripts were pseudonymized prior to analysis.

Participant Observations

The analysis of the focus groups in a previous study [10] showed that the initial introduction of the technology on the ward by the manufacturer is an important factor for the further course of implementation. Therefore, this factor was investigated as a decisive condition for the implementation of IRMS. For this purpose, observations in the sense of a focused ethnography

XSL•FO
RenderX

[25] were carried out by two researchers (SW and IS) during the training sessions. The observations were open with the consent of the participants and the trainers. Field notes taken during the observations were collected, converted to a digital text format, and pseudonymized prior to analysis.

Analysis

Qualitative data (focus groups, participant observation, and interview) were analyzed by SW, IS, and NM. We systematically organized and structured the data according to the principles of Kuckartz's content analysis approach [26]. The main categories were shaped by the guiding questions and supplemented with inductive aspects derived from analysis discussions. Additionally, the investigation of specific aspects of the discussion dynamics in the focus groups was inspired by elements of the documentary method, particularly as outlined by Bohnsack [27] and Bohnsack and Schäffer [28]. This methodological combination aimed to provide a robust foundation for analyzing both the explicit content structure and the underlying social dynamics and interpretations of meaning, thereby promoting a comprehensive understanding of the phenomena under study. Descriptive quantitative data analysis by SW, IS, and HM was performed using SPSS Statistics (IBM Corp) for Windows. Frequencies and percentages were calculated for all quantitative variables.

Setting

The study was conducted at the University Medical Center Freiburg.

Ethical Considerations

The conceptualization and implementation of this study were based on the principles of the Declaration of Helsinki. By German law, survey studies with a focus on employees must be approved by the Employee Council at the respective institution. The responsible committee of the Employee Council of the University Medical Center Freiburg provided approval for this study in written format (approval 106/20 [MPG §23b]). Participation in the focus groups was voluntary, no personal data were collected, and anonymity was always maintained. All potential participants received written information on the study (reason for the study, objective, processes, and data protection), were informed about the decision of the Employee Council and its subcommittees, and had the opportunity to contact the investigators in case of questions at any time during the study. Informed consent to participate was assumed if individuals completed the survey and were confirmed (by ticking a box) at the beginning of the survey.

Results

Participant Observations

The two initial training sessions were diverging. There were 12 participants in total—10 nurses and 2 physiotherapists. They took place in two very different places. The first one was in a lecture hall where the participants were spread around the room. They seemed quite tired after the morning shift but were eager to understand the technology. The amount of information seemed rather overwhelming in this setting. However, the nurses seemed to be able to follow very well. The second introduction took place on the ward with two participants from the physiotherapy profession. The atmosphere was livelier and more optimistic regarding the use of the technology. There was an interprofessional exchange, and strategies for using the IRMS were discussed. Once again, the participants' questions demonstrated their engagement and ability to quickly grasp the basic concepts of the technology. The presenter began the session by mentioning the number of miles nurses walk every day. However, this topic was not explored further in the discussion, and we conclude that the participants did not prioritize it as a goal for using the technology. Additionally, the presenter emphasized the expertise of the nurses in making decisions about the use of the IRMS. The presentation of case scenarios primarily focused on night-time use. Although preventing falls and other adverse events was framed as a key interest of the research team, the representative highlighted both fall prevention and motion analysis, with a particular focus on the latter. One participant's question suggested that targeting specific patients might be a promising strategy.

Survey

At the time of the survey, participants had been using the IRMS on the ward for 20 weeks and had gained initial experience of its functionality. A total of 11 of a possible 21 individuals participated in the survey, 10 of which were completed (10/21, 48%). Nurses' responses to the survey confirmed that caring for patients with cognitive impairment is a significant burden (items 8-19 in Multimedia Appendix 1). Notably, participants demonstrated a general openness and minimal fear of technology (items 1-7 in Multimedia Appendix 1).

When asked about potential improvements from the IRMS, most respondents were optimistic and agreed that it would improve outcomes in several areas. The only areas where they did not agree were preventing the removal of peripheral venous catheters and drains, improving privacy, and disturbing other patients (items 23, 25, and 26 in Multimedia Appendix 1). However, nurses also see many opportunities to improve care and work processes through the use of technology, including feeling safer due to the use of the IRMS, receiving support in care planning, and assisting in coordinating their work schedule (items 28-30 in Multimedia Appendix 1). The impact on patient safety is also seen as positive (item 20 in Multimedia Appendix 1). The survey responses from the nurses indicate a clearly positive view regarding the IRMS's effectiveness in preventing falls and preventing patients from leaving the ward unnoticed (Figure 2).

In terms of organizational aspects, nurses found the IRMS to be a valuable tool for care planning and time management (Figure 2). In addition, 80% (n=8) of respondents (statements from "tend to agree" to "fully agree") confirmed that the technology enabled them to assess the quality of their work (Figure 2). For nurses to find the IRMS helpful, they must find it reliable—which they consistently confirmed in the survey, both for bed edge and bed exit information and for motion monitoring (items 40-42 in Multimedia Appendix 1). In terms of usability, respondents generally agreed that using the technology was not annoying and did not increase the

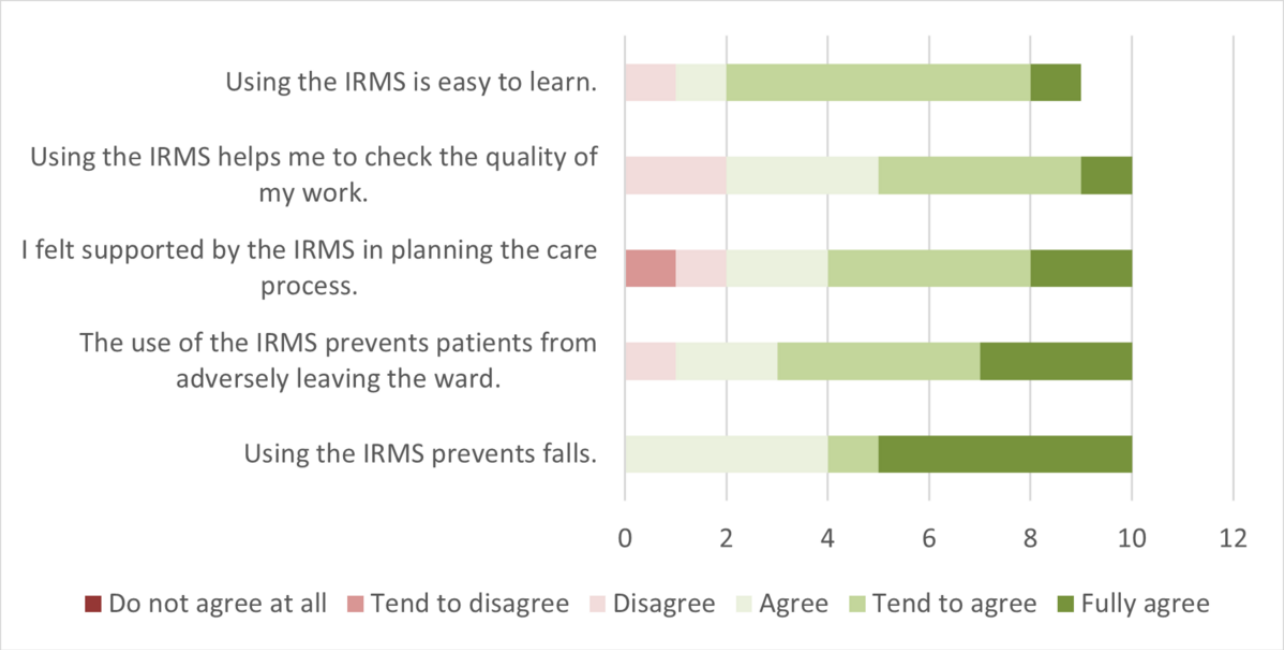
XSL•FO
RenderX

documentation burden. However, the initial setup was considered by some to be labor-intensive. Opinions varied between respondents, possibly influenced by the actual involvement in the setup process. (items 32, 33, and 37 in Multimedia Appendix 1). Overall, nurses found it easy to learn how to use the device (Figure 2).

Responses regarding the potential reduction of work-related movements, such as checking on patients at risk of falling, were mixed, with differing opinions between participants. Further insight into this variation can be gained by analyzing the focus groups, as described below.

Ethical considerations were explored by assessing whether nurses felt that the IRMS helped to maintain patient privacy. In general, respondents tended to think that it did not, a nuance that may be clarified by analysis of the focus groups. When asked if they felt monitored by the technology, 8 out of 10 nurses said that this was not the case. Almost all of the nurses surveyed (n=9; statements from agree to fully agree) indicated that they would continue to use the IRMS in the future.

Figure 2. Survey results on the effectiveness and organizational aspects of the in-bed real-time motion monitoring system (IRMS).



Focus Groups and Interview

Overview

The focus groups and the interview largely confirm the results of the survey and can shed more light on some details. Each focus group had a total of three participants, with one person attending both sessions. The issues of benefits, privacy concerns, and usability were discussed more controversially than the survey suggests. This is illustrated by the following analyses. A summary of the themes, including their descriptions and primary insights, is provided in Table 1. The originals of the quotes used can be found in Multimedia Appendix 4.

Table 1. Summary of themes and key insights from focus groups and the interview.

Theme	Description	Key insights
Benefits	Discussion on the benefits of the IRMS[a], particularly regarding fall prevention and night shift management.	The IRMS was seen as helpful in improving patient safety and reducing anxiety through quick responses. It was especially valuable during night shifts to manage patient movements.
Privacy concerns	Concerns regarding patient privacy in relation to monitoring through the IRMS.	Some patients felt monitored or “controlled.” However, this was considered an exception. A clear explanation of the system helped alleviate some concerns.
Usability	Experiences of nurses with the usability and reliability of the system.	The IRMS was considered increasingly reliable. Suggestions for improvement mainly focused on integration with the documentation system and reducing “false alarms.”
Overall assessment	Overall assessment of the system and nurses’ willingness to continue using the IRMS.	Nurses found the system helpful, but not essential. They emphasized that care could still be provided without it, though some additional visual checks would be necessary.

[a]IRMS: in-bed real-time motion monitoring system.

XSL•FO
RenderX

Benefits

Nurses participating in both focus groups reported that they found the technology helpful in preventing or responding to falls. However, reports of specific responses to information about bed exit or bed edge information remained largely vague despite repeated questioning. The meaning of the "usefulness" of the IRMS becomes clearer when considering that the additional information allows nurses to better plan their next steps, potentially allowing them to respond more quickly to falls.

> *And when it rings, we know, this woman is sitting at the edge of the bed. Then we know...We have ideas. Then we know how we have to organize the process or how to do the plan for the patient.* [Focus Group 1, P3]

In addition, the absence of narratives about specific fall-prevention situations emphasizes that it was not the prevention of falls themselves that played a crucial and relieving role, but rather a sense of safety in relation to falls. Simply knowing that they would be alerted if the patient got up allowed nurses to perform their other activities with less anxiety.

> *Well. Especially regarding falls there was a certain safety, because the notification actually always sets off and you don't have to permanently do visual controls and go to the rooms. This certainly creates a feeling of safety, to know "Okay, the patient is not in danger of falls or similar right now." Well and for coverage of the positioning and in order to see: "Okay, was there sufficient positioning? Have I supported the patient sufficiently?...it gives a certain feeling of safety when you can check this on the monitor."* [Focus Group 2, P3]

In addition to an increased sense of security, another clearly identified benefit is improved information about patient self-movement. This is particularly important at night, when patients may need to be repositioned at defined intervals according to nursing standards. Right at the beginning of the interview, the IRMS is described as a technology that goes beyond and is superior to the existing technical possibilities (eg, a mat on the floor in front of the bed that sends a signal to the call system when touched by the patient) and is superior to them. According to the interviewee, the technology is particularly valuable during night shifts because it constantly monitors whether patients are moving or not. Previously, patients had to be awakened and repositioned every 2 to 3 hours to relieve pressure.

> *In nightshifts [with the IRMS] you immediately notice: Do people move, don't they move. And before, okay you walk there every 2 to 3 hours, turn the people from left to right, have to wake them up and this is bad, when people are fast asleep and you see: Oh, he's moving, moving sufficiently, perfect. AND, it's very good that when somebody really has developed a decubitus here, we can prove: Okay, every 2 hours, the monitor has recorded it and there were immediate reactions, there were positionings and so we can demonstrate this.* [Interview]

The IRMS provides an indication of whether there has been sufficient spontaneous movement, allowing nurses to avoid unnecessary repositioning. This helps prevent unnecessary disruption to the patient's sleep and supports nurses in re-evaluating their own judgments about the need for repositioning. If a patient develops a pressure ulcer, the recorded data can be used by nursing staff to demonstrate that regular pressure-relieving positioning has been carried out correctly (interview and focus groups). The interviewee also mentioned the recording of bed exits, especially for patients with hyperdelirium. The ability to detect physical tension in patients through specific movement patterns is highlighted as another, initially unintended, "function" of the device.

Privacy Concerns

The issue of privacy was not directly raised in the focus group discussions. However, the analysis provides insights into the different responses in the survey. The difference may be due to different perceptions or reference points of "privacy." In the second focus group discussion, it was reported that with the use of the IRMS, room doors are occasionally closed more often, as monitoring can take place without direct visual contact. However, according to nurses' reports, some patients occasionally expressed feeling monitored and controlled. For example, the interviewee reported a situation where the technology was seen by a patient as a "spy in the bed."

> *This one patient, who always said: "Ha, the spy betrayed me again, actually I wanted to try to go to the toilet by myself." There you stand next to them and say: "You shouldn't walk alone; you are still quite weak-kneed with the rollator."* [Interview]

Due to the design of the study, it remains uncertain how patients experience these reactions, whether positively or negatively. It is worth noting that the term "spying" typically carries a negative connotation in German. This anecdote, involving a patient at risk of falling who was attempting to use the toilet independently, illustrates the effectiveness of the IRMS in preventing potentially dangerous situations. The system alerted the nursing staff, enabling them to assist the patient immediately. The nurses perceive their immediate presence when the system is activated as a positive aspect, as it allows for a quick response to the patient's needs. However, the extent to which patients are comfortable with this form of monitoring seems to be a critical factor.

> *P3: I think they were not really aware, that there is a special mat in their bed. They didn't really notice it, most of them. The others quickly mentioned it. But they didn't mind either, whether there was something like that or not. So.*
>
> *P1: Except for the ones who ranted, because they were worried, well...one once even said: "Am I being controlled?"*
>
> *P3: Okay no, I've never come across that.*
>
> *P1: And then there were some, who worried [about] the space for their cell phone plug.* [Focus Group 2, P1 and P3]

XSL•FO
RenderX

This highlights an interesting ethical conflict. The nurses in the focus group at P1 also report patients' complaints or confusion about the sense of being controlled, but they regard these as special cases, similar to minor concerns like charging their mobile phones. However, the primary cause of confusion appears to be the patient's general state of confusion. Explaining the technology seemed effective in addressing this confusion. However, the nurses did not report exploring deeper complaints, such as the feeling of being controlled or monitored. This assessment by the nurses regarding patient privacy may correlate with their own perceptions of control associated with IRMS. In the first focus group at P0 and the interview (P1), some team members initially viewed the IRMS as a potential "control device," referring to the supervision of nurses and their activities, for example, by line managers. Although this perception dissipated relatively quickly as the IRMS became more established, it highlights nurses' concerns about potential surveillance, which were also discussed in the focus groups. After a longer period of use (according to reports from the second focus group), nurses no longer feared being monitored themselves. While in the first focus group, they were suspicious of the potential monitoring of their work performance, in the second focus group, they expressed hope that the IRMS would demonstrate that they were making correct judgments about patient positioning. The issue of surveillance was eventually dismissed as irrelevant.

Usability

With minor reservations, the nurses consider the system to be reliable, especially in the second focus group and the interview at P1, more so than at the beginning. They have no fundamental objections to the usability of the system, but they do have some suggestions for improvement. Most of these suggestions focus on improving the documentation process. For example, if there were an interface to the documentation system and a simplified or automated transfer of recorded data to mandatory documentation, the IRMS could streamline the documentation process overall. This could also help nurses to provide a plausible explanation for justified deviations from standards of care, thereby encouraging the use of their own professional knowledge and experience.

In addition to the survey's assessment that the technology is easy to learn, focus group participants and the interviewee emphasize that its use requires the development of routines. They point out that a lack of familiarity with the system can lead to disruptions in ward operations, particularly among other professional groups who do not work exclusively on the participating ward and have not developed routines with the IRMS. This often results in "false alerts" during the day, such as physiotherapists forgetting to turn off alerts before their interventions.

In line with the inconsistent assessment in the survey regarding the reduction of walking distances, different perspectives emerged in the focus group discussions. When using the IRMS, nurses do not have to repeatedly check on patients at risk of falling by visiting them. This is especially an improvement at night. However, this process rarely exists in isolation, as nurses often combine these checks with other tasks. The implementation of the system does not alter these interrelated activities, which continue as usual.

> *Those who are hyperactive at night, you pick that up anyways, because then you are continuously close to the bed, because they're ringing all the time, calling or doing anything else.* [Focus Group 2, P1]

Overall Assessment

When asked whether they would like to continue using the technology, the nurses in the second focus group were slightly more ambivalent than in the survey. On the one hand, they emphasize that care can be provided without such technology, but on the other hand, they generally find the IRMS helpful.

> *Moderator: And if there was no IRMS starting from tomorrow on. How would you like that?*
>
> *Person 3: Well...*
>
> *Several: [laughing]*
>
> *Person 2: I would miss it. I would like to*
>
> *Person 3: Yes, it would be missed. But the work still could be done. That's not the way it is.*
>
> *Person 2: Well maybe a thing or two would be more complicated again. Because we would have to do more visual controls at the patients' beds, for patients who could wander or patients with dementia.*
>
> *Person 3: Yes.*
>
> *Person 1: Exactly, probably one fall or two would happen* [2.5 seconds]
>
> *Person 3: but those could not be prevented with IRMS, either.*
>
> *Person 1: Yes, sure! We would certainly go back to text-book, which means positioning according to the clock, but-*
>
> *Person 2: This would not make our work impossible*
>
> *Person 1: (laughing) Exactly. Nursing can do that* [Focus Group 2, P1, P2, and P3]

Overall, the nurses expressed a willingness to continue using the IRMS but highlighted that care is not entirely dependent on such a technology. In the following discussion, possible backgrounds for this evaluation will be examined further.

Discussion

Principal Findings

The two initial training sessions showed differences in both effectiveness and atmosphere. The second introductory session, held on the ward with additional therapeutic staff, fostered a much livelier and more optimistic atmosphere toward the technology. Interprofessional exchange was seamless, with active discussion of strategies for implementation and use of the IRMS. These observations are consistent with the research of Koukourikos et al [29], which emphasizes the effectiveness of hands-on training in familiar environments over formal settings. Moreover, Kahn et al [30] highlight the importance of interprofessional collaboration from the outset of technology implementation in hospital care settings. Building on these

XSL•FO
RenderX

findings, a recent study using a design-based research approach highlights the importance of structured training and support programs to help nurses effectively integrate innovative technologies into practice [31].

The survey, focus groups, and the interview showed that nurses benefit most from motion monitoring, especially at night. The increased sense of security can reduce psychological distress. This finding is consistent with the systematic review by Mileski et al [32], which included 28 studies. Whether the technology is effective in preventing falls could not be determined as this study was not designed to measure its effectiveness and the nurses' statements in this regard remain ambiguous. However, the technology does appear to reduce the nurses' fear of falls and make them feel better equipped to respond in a timely and appropriate manner should a fall occur.

Interestingly, the focus on fall prevention was primarily driven by the research question, but it emerged that other benefits were more relevant. One reason for the limited focus on fall prevention might be that nurses' responses to the information provided by the system were constrained by existing routines. Since there was no significant adaptation of processes beyond the introduction of the technology, nurses could not change their workflows significantly. More comprehensive changes—such as ongoing training programs and organizational policies supporting flexibility in nursing practices—would be needed for nurses to react more promptly when desired. Future studies would benefit from exploring how such ongoing support influences technology adoption and the potential long-term impact on nursing workflows.

The literature also suggests that alarms, particularly when they are frequent or audible to patients, may not be effective as standalone interventions [33-35]. A high number of alerts or alarms (especially if patients can hear them) may even have adverse effects, as suggested by the study by Considine et al [11]. Excessive alarms can increase anxiety and stress levels, disturb sleep, and desensitize both patients and staff, leading to delayed responses to critical situations. This may explain why the system was primarily used at night when nurses could respond more directly to the system's alerts. Goals such as reducing physical strain by reducing the need to walk to check on patients are negated because such rounds often include other essential nursing activities. For example, during these rounds, nurses are often checking on multiple patients, assessing their comfort, administering medication, or performing other routine tasks. This underlines that technology can only reach its full potential if it is integrated into comprehensive, targeted changes and is part of a broader, well-structured care plan [32].

Another factor contributing to the mixed evaluation of the IRMS could be debates about the professionalization of nursing in Germany [36]. Notably, in the second focus group, it was emphasized that nursing care could still be provided effectively without the use of technology.

Concerns about surveillance disappeared over time. In the survey and the second focus group, nurses no longer expressed privacy concerns and even minimized patients' worries. These findings are consistent with a previous study on a BES used across several wards at the University Medical Center [10]. On one hand, it could be argued that patient privacy is actually enhanced compared to personal observation, as physical movements can be monitored discreetly, without patients' awareness. However, patients still voiced concerns about being monitored. While the focus group participants did not consider this a serious issue, it remains essential to address patient privacy. Nurses could explore ways to make the technology more understandable, ensuring that patients are informed and that their consent is respected. Situations may arise where a patient's rejection of the technology—due to significant stress or a violation of personal values—should lead to discontinuation of the system's use. Future research could explore effective communication strategies to address privacy concerns and promote patient acceptance of monitoring technologies.

Finally, integration with documentation standards and systems remains a major concern contributing to nurses' workload, as discussed in the reviews by Mileski et al [32] and in studies on the implementation of digital nursing technologies [37-39]. Considering these aspects in the development of new technologies could significantly enhance the benefits of nursing. By supporting conditions that allow nurses to work in a self-directed and knowledge-based way, both their empowerment and professional standing could be strengthened.

Limitations

When discussing the use of the IRMS, it is important to clarify that the study's aim was not to assess its effectiveness or similar outcomes but rather to explore nurses' perceptions of its usefulness.

One limitation of the study is the small sample size. Ideally, multiple focus group discussions would have been conducted and analyzed at each time point to identify specific discursive patterns through comparison. Unfortunately, this was not feasible due to a limited number of eligible participants and a lack of volunteers. For practical reasons, an interview was conducted instead of a focus group. While an interview does not capture the same level of social consensus on issues as a focus group, the results were notably similar. The survey achieved a response rate representing 48% of the nursing staff in the geriatric ward. Although the sample size was relatively small, it provided valuable insights into the perceptions of nearly half of the relevant nursing staff. Given the limited number of participants, it is possible that the findings could be influenced by selection bias, especially if the individuals who participated were more likely to be early adopters or held leadership positions within their wards. Early adopters, including those in leadership roles, may exhibit a more positive perception of new technologies, potentially leading to an overrepresentation of optimistic perspectives in our results.

However, it is important to note that this study does not aim to present representative findings. Instead, it seeks to provide an in-depth insight into the implementation and use of the technology within a specific context. In qualitative research, some degree of bias is inherent, as findings are inevitably tied to the specific social situations and perspectives of the participants involved.

XSL•FO
RenderX

Conclusions

Based on the findings, the participating nurses have a positive attitude toward the use of a real-time in-bed movement monitoring system, highlighting its benefits for nurses and patients, particularly during nighttime monitoring. While its effectiveness in preventing falls remains inconclusive, the system appears to reduce nurses' fear of falls and improve their ability to respond effectively. The results highlight the wider impact of the system beyond fall prevention and underline the importance of carefully integrating such technologies into health care practice. The study acknowledges the challenges of embedding technology into care, including the limitations of nurses' responses to the IRMS information and ongoing discussions about professional dynamics in Germany. It also identifies opportunities for future research and development, particularly in relation to integration with documentation systems. Overall, the study argues for a deliberate, holistic approach to health care technology that considers professional relationships, patient needs, and system improvements.

Acknowledgments

We thank all the participants for their time and willingness to share their experiences with us. This study was conducted at the Center of Implementing Nursing Care Innovations, Freiburg, Germany. This joint project on the development and research of new care technologies in Germany is funded by the Federal Ministry of Education and Research (grants 16SV7889, 16SV7888, and 16SV7886K).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

The presentation of the authors' contributions followed the recommendations of the Contributor Roles Taxonomy (CRediT) [40]. SW contributed to the conceptualization and methodology of the study, conducted the investigation, performed formal analysis, curated the data, and wrote the original draft of the manuscript. IS was involved in the conceptualization and methodology of the research, carried out the investigation, conducted formal analysis, curated the data, and contributed to the writing of the original draft. JP performed the formal analysis, curated the data, and contributed to the writing and editing of the manuscript. NM assisted with formal analysis and data curation. HM was involved in the conceptualization and methodology. SZ contributed to the conceptualization and methodology. CK contributed to the conceptualization, supervision, resource review, and editing. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey of the nursing staff.
[DOCX File , 20 KB - formative_v9i1e63572_app1.docx]

Multimedia Appendix 2
Focus group guide I.
[DOCX File , 29 KB - formative_v9i1e63572_app2.docx]

Multimedia Appendix 3
Focus group guide II.
[DOCX File , 31 KB - formative_v9i1e63572_app3.docx]

Multimedia Appendix 4
Translated quotes.
[DOCX File , 23 KB - formative_v9i1e63572_app4.docx]

References

1. Mudge AM, Lee-Steere K, Treleaven E, Cahill M, Finnigan S, McRae P. Cognitive impairment in older hospital inpatients: prevalence, care needs and carer perceptions. Aust Health Rev 2021;46(2):244-250. [doi: 10.1071/ah20286]
2. Jackson TA, MacLullich AMJ, Gladman JRF, Lord JM, Sheehan B. Undiagnosed long-term cognitive impairment in acutely hospitalised older medical patients with delirium: a prospective cohort study. Age Ageing 2016;45(4):493-499. [doi: 10.1093/ageing/afw064] [Medline: 27076525]

XSL•FO
RenderX

3. George J, Long S, Vincent C. How can we keep patients with dementia safe in our acute hospitals? A review of challenges and solutions. J R Soc Med 2013;106(9):355-361 [FREE Full text] [doi: 10.1177/0141076813476497] [Medline: 23759885]
4. Fogg C, Meredith P, Bridges J, Gould GP, Griffiths P. The relationship between cognitive impairment, mortality and discharge characteristics in a large cohort of older adults with unscheduled admissions to an acute hospital: a retrospective observational study. Age Ageing 2017;46(5):794-801 [FREE Full text] [doi: 10.1093/ageing/afx022] [Medline: 28338808]
5. Choi SSW, Cajita MI, Gitlin LN. A review of measures of three common dementia-related behaviors: rejection of care, aggression, and agitation. Geriatr Nurs 2020;41(6):692-708. [doi: 10.1016/j.gerinurse.2020.04.003] [Medline: 32402574]
6. Husebo BS, Heintz HL, Berge LI, Owoyemi P, Rahman AT, Vahia IV. Sensing technology to monitor behavioral and psychological symptoms and to assess treatment response in people with dementia. a systematic review. Front Pharmacol 2019;10:1699 [FREE Full text] [doi: 10.3389/fphar.2019.01699] [Medline: 32116687]
7. Astell AJ, Bouranis N, Hoey J, Lindauer A, Mihailidis A, Nugent C, et al. Technology and dementia: the future is now. Dement Geriatr Cogn Disord 2019;47(3):131-139 [FREE Full text] [doi: 10.1159/000497800] [Medline: 31247624]
8. Oh-Park M, Doan T, Dohle C, Vermiglio-Kohn V, Abdou A. Technology utilization in fall prevention. Am J Phys Med Rehabil 2021;100(1):92-99. [doi: 10.1097/PHM.0000000000001554] [Medline: 32740053]
9. Komagata H, Hattori T, Ohshima R, Kakinuma E, Ishikawa M, Shinoda K, et al. Development of human behavior monitoring system around bed using infrared depth sensor. : IEEE; 2019 Presented at: IEEE 1st Global Conference on Life Sciences and Technologies (LifeTech); March 12-14, 2019; Japan. [doi: 10.1109/lifetech.2019.8884068]
10. Walzer S, Schön I, Pfeil J, Klemm S, Ziegler S, Schmoor C, et al. Nurses' perspectives and experiences of using a bed-exit information system in an acute hospital setting: mixed methods study. JMIR Form Res 2025 Feb 05;9:e64444 [FREE Full text] [doi: 10.2196/64444] [Medline: 39908092]
11. Considine J, Berry D, Mullen M, Chisango E, Webb-St Mart M, Michell P, et al. Nurses' experiences of using falls alarms in subacute care: a qualitative study. PLoS One 2023;18(6):e0287537 [FREE Full text] [doi: 10.1371/journal.pone.0287537] [Medline: 37347774]
12. Compliant Concept AG 2023. URL: https://www.compliant-concept.ch/produkte/spital/#produkt-mobility-monitor [accessed 2025-01-28]
13. Ziegler S, Schmoor C, Schöler LM, Schepputat S, Takem E, Grotejohann B, et al. Potential for reducing immobility times of a mobility monitor in-bed sensor system—a stepped-wedge cluster-randomised trial. BMC Nurs 2023;22(1):478 [FREE Full text] [doi: 10.1186/s12912-023-01658-2] [Medline: 38104112]
14. Røyset B, Talseth-Palmer BA, Lydersen S, Farup PG. Effects of a fall prevention program in elderly: a pragmatic observational study in two orthopedic departments. Clin Interv Aging 2019;14:145-154 [FREE Full text] [doi: 10.2147/CIA.S191832] [Medline: 30697039]
15. Cuttler SJ, Barr-Walker J, Cuttler L. Reducing medical-surgical inpatient falls and injuries with videos, icons and alarms. BMJ Open Qual 2017;6(2):e000119 [FREE Full text] [doi: 10.1136/bmjoq-2017-000119] [Medline: 29450285]
16. Hubbartt B, Davis SG, Kautz DD. Nurses' experiences with bed exit alarms may lead to ambivalence about their effectiveness. Rehabil Nurs 2011;36(5):196-199. [doi: 10.1002/j.2048-7940.2011.tb00195.x] [Medline: 21882797]
17. Shee AW, Phillips B, Hill K, Dodd K. Feasibility, acceptability, and effectiveness of an electronic sensor bed/chair alarm in reducing falls in patients with cognitive impairment in a subacute ward. J Nurs Care Qual 2014;29(3):253-262. [doi: 10.1097/NCQ.0000000000000054] [Medline: 24487696]
18. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008;13(2):92-98. [doi: 10.1258/jhsrp.2007.007074] [Medline: 18416914]
19. Ingham-Broomfield R. A nurses' guide to mixed methods research. Aust J Adv Nurs 2016;33(4):46-52. [doi: 10.37464/2016.334.1559]
20. Creswell JW. A Concise Introduction to Mixed Methods Research. United States: SAGE publications; 2021.
21. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
22. Spagnolli A, Guardigli E, Orso V, Varotto A, Gamberini L. Measuring User Acceptance Of Wearable Symbiotic Devices: Validation Study Across Application Scenarios. Cham: Springer; 2014.
23. Isfort M, Klostermann J, Gehlen D, Siegling B. Eine bundesweite befragung von leitenden pflegekräften zur pflege und patientenversorgung von menschen mit demenz im krankenhaus. Köln: deutsches institut für angewandte pflegeforschung 2014 [Website in German]. Pflege-Thermometer. 2014. URL: https://www.dip.de/fileadmin/data/pdf/material/Pflege-Thermometer_2014.pdf [accessed 2014-07-26]
24. Azevedo V, Carvalho M, Costa F, Mesquita S, Soares J, Teixeira F, Maia. Interview transcription: conceptual issues, practical guidelines, and challenges. Rev Enf Ref 2017;IV Série(Nº14):159-168. [doi: 10.12707/riv17018]
25. Cruz EV, Higginbottom G. The use of focused ethnography in nursing research. Nurse Res 2013;20(4):36-43. [doi: 10.7748/nr2013.03.20.4.36.e305] [Medline: 23520711]
26. Kuckartz U. Qualitative text analysis: a systematic approach. In: Compendium for Early Career Researchers in Mathematics Education. Cham: Springer; 2019:181-197.

XSL•FO
RenderX

27. Bohnsack R. Documentary method. In: The SAGE Handbook of Qualitative Data Analysis. United States: Sage Publications; 2014:217-233.
28. Bohnsack R, Schäffer B. Exemplarische Textinterpretation: Diskursorganisation und dokumentarische Methode [Book in German]. New York: Springer; 2013.
29. Koukourikos K, Tsaloglidou A, Kourkouta L, Papathanasiou IV, Iliadis C, Fratzana A, et al. Simulation in clinical nursing education. Acta Inform Med 2021;29(1):15-20 [FREE Full text] [doi: 10.5455/aim.2021.29.15-20] [Medline: 34012208]
30. Khan AI, Barnsley J, Harris JK, Wodchis WP. Examining the extent and factors associated with interprofessional teamwork in primary care settings. J Interprof Care 2022;36(1):52-63 [FREE Full text] [doi: 10.1080/13561820.2021.1874896] [Medline: 33870838]
31. Walzer S, Barthel C, Pazouki R, Marx H, Ziegler S, Koenig P, et al. Teaching in the digital age-developing a support program for nursing education providers: design-based research. JMIR Form Res 2025 Jan 15;9:e66109 [FREE Full text] [doi: 10.2196/66109] [Medline: 39813674]
32. Mileski M, Brooks M, Topinka JB, Hamilton G, Land C, Mitchell T, et al. Alarming and/or alerting device effectiveness in reducing falls in long-term care (LTC) facilities? A systematic review. Healthcare 2019;7(1):51. [doi: 10.3390/healthcare7010051]
33. Barker AL, Morello RT, Ayton DR, Hill KD, Brand CA, Livingston PM, et al. Acceptability of the 6-PACK falls prevention program: a pre-implementation study in hospitals participating in a cluster randomized controlled trial. PLoS One 2017;12(2):e0172005 [FREE Full text] [doi: 10.1371/journal.pone.0172005] [Medline: 28199376]
34. Chu RZ. Preventing in-patient falls: the nurse's pivotal role. Nursing 2017;47(3):24-30. [doi: 10.1097/01.nurse.0000512872.83762.69]
35. Hamm J, Money AG, Atwal A, Paraskevopoulos I. Fall prevention intervention technologies: a conceptual framework and survey of the state of the art. J Biomed Inform 2016;59:319-345 [FREE Full text] [doi: 10.1016/j.jbi.2015.12.013] [Medline: 26773345]
36. Roth C, Wensing M, Breckner A, Mahler C, Krug K, Berger S. Keeping nurses in nursing: a qualitative study of German nurses' perceptions of push and pull factors to leave or stay in the profession. BMC Nurs 2022;21(1):48 [FREE Full text] [doi: 10.1186/s12912-022-00822-4] [Medline: 35193561]
37. Bail K, Gibson D, Acharya P, Blackburn J, Kaak V, Kozlovskaia M, et al. Using health information technology in residential aged care homes: an integrative review to identify service and quality outcomes. Int J Med Inform 2022;165:104824 [FREE Full text] [doi: 10.1016/j.ijmedinf.2022.104824] [Medline: 35792376]
38. Abell B, Naicker S, Rodwell D, Donovan T, Tariq A, Baysari M, et al. Identifying barriers and facilitators to successful implementation of computerized clinical decision support systems in hospitals: a NASSS framework-informed scoping review. Implement Sci 2023;18(1):32 [FREE Full text] [doi: 10.1186/s13012-023-01287-y] [Medline: 37495997]
39. Abdolkhani R, Petersen S, Walter R, Zhao L, Butler-Henderson K, Livesay K. The impact of digital health transformation driven by COVID-19 on nursing practice: systematic literature review. JMIR Nurs 2022;5(1):e40348 [FREE Full text] [doi: 10.2196/40348] [Medline: 35867838]
40. Brand A, Allen L, Altman M, Hlava M, Scott J. Beyond authorship: attribution, contribution, collaboration, and credit. Learned Publ 2015;28(2):151-155. [doi: 10.1087/20150211]

Abbreviations

BES: bed exit information system
GRAMMS: Good Reporting of a Mixed Method Study
IRMS: in-bed real-time motion monitoring system
REDCap: Research Electronic Data Capture

Edited by A Mavragani; submitted 26.06.24; peer-reviewed by H Hatamabadi, S Kling; comments to author 15.10.24; revised version received 28.11.24; accepted 19.01.25; published 04.03.25.

Please cite as:
Walzer S, Schön I, Pfeil J, Merz N, Marx H, Ziegler S, Kunze C
Experiences With an In-Bed Real-Time Motion Monitoring System on a Geriatric Ward: Mixed Methods Study
JMIR Form Res 2025;9:e63572
URL: https://formative.jmir.org/2025/1/e63572
doi:10.2196/63572
PMID:40053780

XSL•FO
RenderX

©Stefan Walzer, Isabel Schön, Johanna Pfeil, Nicola Merz, Helga Marx, Sven Ziegler, Christophe Kunze. Originally published in JMIR Formative Research (https://formative.jmir.org), 04.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Patient Adoption of Digital Use Cases in Family Medicine and a Nuanced Implementation Approach for Family Doctors: Quantitative Web-Based Survey Study

Julian Beerbaum[1], MSc; Sibylle Robens[2], PhD; Leonard Fehring[3,4], MD; Achim Mortsiefer[5], Prof Dr; Sven Meister[1,6], Prof Dr

[1]Health Care Informatics, Faculty of Health, School of Medicine, Witten/Herdecke University, Witten, Germany

[2]Department of Psychology and Psychotherapy, Witten/Herdecke University, Witten, Germany

[3]Department of Gastroenterology, Helios University Hospital Wuppertal, Witten/Herdecke University, Wuppertal, Germany

[4]Faculty of Health, School of Medicine, Witten/Herdecke University, Witten, Germany

[5]General Practice II and Patient-Centeredness in Primary Care, Institute of General Practice and Primary Care, Faculty of Health, School of Medicine, Witten/Herdecke University, Witten, Germany

[6]Department Healthcare, Fraunhofer Institute for Software and Systems Engineering, Dortmund, Germany

Corresponding Author:
Sven Meister, Prof Dr
Health Care Informatics, Faculty of Health
School of Medicine
Witten/Herdecke University
Pferdebachstr. 11
Witten, 58448
Germany
Phone: 49 230292678629
Email: sven.meister@uni-wh.de

Abstract

Background: Digital use cases describe the application of technology to achieve specific outcomes. Several studies in health care have examined patients' overall attitudes toward digitalization and specific use cases. However, these studies have failed to provide a comparison of patient acceptance criteria between inherently different digital use cases in family medicine.

Objective: To address this research gap, this paper aimed to assist family doctors in selecting digital use cases by comparing the underlying patient adoption factors and in driving usage of these use cases by presenting a differentiated implementation approach.

Methods: Adapting an established Unified Theory of Acceptance and Use of Technology (UTAUT) questionnaire to 4 digital use cases in family medicine, we surveyed a large cross-sectional sample of adults living in Germany. The results of the web-based survey were then analyzed via descriptive statistics, ANOVA, and hierarchical regression models to compare the effects of sociodemographic and technology acceptance factors on the intention to use a specific use case.

Results: Our web-based survey included 1880 participants. Of these 1880 participants, only 304 (16.2%) agreed that the degree of digitalization is important when selecting a family practice. However, more digitally literate participants attributed greater importance to this criterion (B=0.226, SE 0.023; β=.223; P<.001), and digital literacy was found to be dependent on age (Welch $F_{3,968.29}$=53.441; P<.001). Regarding sociodemographic characteristics, only digital literacy demonstrated a significant effect on the intention to use for all use cases, particularly scheduling doctor appointments online (B=0.322, SE 0.033; β=.408; P<.001). Furthermore, *performance expectancy* was the strongest predictor of the intention to use for all use cases, while further effects of technology acceptance factors depended on the use case (receiving medical consultations via video: B=0.603, SE 0.049; β=.527; P<.001; scheduling doctor appointments online: B=0.566, SE 0.043; β=.513; P<.001; storing personal medical information via electronic health records: B=0.405, SE 0.047; β=.348; P<.001; and providing personal information before consultation digitally [digital anamnesis]: B=0.434, SE 0.048; β=.410; P<.001). To illustrate, *perceived privacy and security* had an effect on the intention to use electronic health records (B=0.284, SE 0.040; β=.243; P<.001) but no effect on the intention to use video consultations (B=0.068, SE 0.042; β=.053; P=.10).

XSL•FO
RenderX

Conclusions: In the selection and implementation of digital use cases, family doctors should always prioritize the perceived value of the digital use case for the patient, and further criteria might depend on the digital use case. Practice owners should therefore always harmonize the introduction of digital use cases with their own patient care strategies. Not every digital innovation fits every strategy and therefore every practice.

(JMIR Form Res 2025;9:e58867) doi:10.2196/58867

KEYWORDS

technology acceptance; UTAUT; family doctor; digital health; eHealth; video consultation; electronic health records; digital anamnesis; online appointment scheduling

Introduction

Background

The benefits of applying digital use cases, which we understand as the application of digital technologies to achieve a specific outcome, are well documented for the health care sector [1,2]. For example, in family medicine in Germany, where the family doctor is the first contact point to the health system and thus particularly important to ensure adequate treatment [3], digital use cases could address overarching challenges such as the shortage of family doctors [4]. After years of sporadic offering and use of digital tools in family medicine, COVID-19 accelerated their use [2,5]. Furthermore, this trend toward digitalization increased the diversity of digitally enabled use cases (eg, from online appointment systems to digital anamnesis tools and video consultation offerings) [5]. To illustrate this, the American Medical Association showed in 2022 in their longitudinal survey that the average number of digital applications used by a doctor increased by 72% between 2016 and 2022 [6].

It is apparent that, today more than ever, family doctors can decide which of the various offerings to implement. However, overall adoption rates of digital health care use cases by both patients and family doctors are still at low levels in some regions. For example, in Germany in 2023, only about 1% of all statutorily insured patients stored personal medical information via electronic health records [7,8]. A similar picture has emerged in other countries (eg, only about 3% of medical appointments in November 2023 in general practice in the United Kingdom were conducted via video) [9]. To stimulate broader adoption of digital use cases, legislative bodies in countries, such as the United Kingdom and Germany, revised their regulatory frameworks, such as directing the default use of electronic health records in Germany [10,11]. Nevertheless, to translate these incentives into actual patient usage of digital use cases, we argue that it is vital to assess the underlying factors determining patient adoption behavior. This evaluation is particularly important as family doctors consider patient expectations when deciding to offer patient-facing use cases [12,13].

In assessing technology acceptance, we premise that differentiation and comparison between use cases might offer fruitful insights, as we presume inherent structural differences between digital use cases (eg, degree of voluntariness of use for patients or degree of dealing with sensitive medical information). Thereby, we add to the academic discourse since there are several studies in health care examining patients' overall digital attitudes toward digitalization [14] as well as for specific use cases [15-18]. However, these studies have failed to provide a nuanced assessment and comparison of patient acceptance criteria between distinct digital use cases in family medicine. This differentiation is particularly vital as potentially diverse patient technology acceptance criteria per use case might emerge, and therefore, a differentiated approach to drive adoption might be needed.

Objectives

This survey-based study provides a nuanced comparison of patients' technology acceptance of key digital use cases in family medicine (ie, receiving medical consultations via video, scheduling doctor appointments online, storing personal medical information via electronic health records, and providing personal information before consultation digitally [digital anamnesis]).

Accordingly, this paper aims to support family doctors in selecting digital use cases by showcasing their relative importance to patients and in driving their use. In doing so, it highlights the need for a differentiated implementation approach based on different patient acceptance criteria for each digital use case.

For the scope of this research paper, we focused on family doctors as they act as gatekeepers for access to digital technologies [19]. Additionally, we hypothesized that family doctors could be important for driving adoption with an already established and trusted patient-doctor relationship.

Methods

Research Approach

Building on the insights of a rapid literature review on patient acceptance of 4 digital use cases, we leveraged an existing questionnaire assessing the acceptance of electronic health records in family medicine based on the Unified Theory of Acceptance and Use of Technology (UTAUT) model [20] and adapted it to the other use cases. The objective of our convenience sample approach was to gather a cross-sectional sample of adult patients living in Germany.

Digital Use Cases in Family Medicine

Family medicine typically provides the first medical contact within the health care system [21], and substantial evidence supports the application of digital technologies (digital use cases) in this setting [1,2]. Due to the high pace of innovation and great variety of digital use cases in family medicine, we limited the number of use cases in scope. The final selection of

XSL•FO
RenderX

digital use cases was a joint decision of a multi-professional team (comprising 4 researchers with diverse backgrounds in the digitalization of family medicine and 1 practicing family doctor). The decision was based on (1) the applicability of these use cases for patients in family medicine and (2) their current perceived practical relevance for both patients and family doctors. Thereby, we adhered to the following use case definitions:

- Receiving medical consultations via video: We followed the definition by Keuper et al [22] that "a video consultation is considered as a real-time visual and audio (digital) contact moment between the patient and health care provider." This excludes, for example, telehealth applications that do not facilitate a contact moment (eg, telemonitoring [23]).
- Scheduling doctor appointments online: We understand online appointment scheduling as the use case of booking an appointment via the internet in real-time [24].
- Storing personal medical information via electronic health records: We followed the definition of the National Coordinator for Health Information Technology of the United States that "EHRs are real-time, patient-centered records that make information available instantly and securely to authorized users" [25].
- Providing personal information before consultation digitally (digital anamnesis): Patient anamnesis refers to the data that patients provide about their personal medical history [26]. Digital anamnesis then applies this concept to the digital space as the information-gathering process occurs via a digital application.

Nevertheless, the full scale of adoption of the various digital use cases is yet to be realized [27]. Overall barriers can be pinpointed to a fragmentation of the digital health landscape, inadequate legislative frameworks, and limited clarity on information exchange, highlighting the necessity to consider the patients' perspectives even more [28].

Research Model

Our research attempts to compare different use cases in family medicine and highlight substantial differences between factors in patient technology acceptance.

Technology acceptance describes the positive decision of an individual toward using technology and explains the various factors and contexts incentivizing individuals to use a particular technology [29,30]. This range of factors in the decision-making process to use a tool in a health care setting is reflected in the various theories used to describe this technology acceptance. AlQudah et al [30] and Heinsch et al [31] demonstrated that the Technology Acceptance Model (TAM) and the UTAUT are the most applied theories in health care settings. Nevertheless, many studies combine different theories to explain technology acceptance with various factors [32,33].

The original UTAUT model consists of 4 constructs. The definitions of these UTAUT constructs by Venkatesh et al [34] are as follows. *Performance expectancy* is defined as "the degree to which an individual believes that using the system will help him or her to attain gains in job performance." In the family medicine setting, this translates to the usefulness of the use case to manage one's health or disease in a family practice setting. *Effort expectancy* is defined as "the degree of ease associated with the use of the system" [34], which, in our setting, relates to how easily patients can use the digital use case. *Social influence* is defined as "the degree to which an individual perceives that important others believe he or she should use the new system" [34]. In family medicine, these important others could range from physicians to public opinions to friends and family members. *Facilitating condition* is defined as "the degree to which an individual believes that an organizational and technical infrastructure exists to support use of the system" [34], which, in our context, is associated with having the right resources, compatible technologies, and knowledge to use the digital use case. The additionally included dimension in our study *perceived privacy and security* refers to "the degree to which patients believe that the respective use case is safe from intrusion and personal information is protected" [20,35]. We support the justification of Abd-Alrazaq et al [20] for applying the UTAUT model in the family medicine setting as the model has shown explanatory power in comparable contexts [31].

In line with our objective to assist family doctors by demonstrating the relative importance of digital use cases for patients and by highlighting different patient adoption factors, we tested 3 key hypotheses (hypothesis A, hypothesis B, and hypothesis C).

First, to illustrate the relative importance of each use case for patients, we hypothesized that awareness, availability, and actual usage behavior differ across use cases. For example, commercial studies have shown that 78% of surveyed patients wish for online appointment scheduling at their family doctor, while only 32% wish for an electronic health record and 27% wish for a video consultation [36].

Hypothesis A is as follows: Different use cases have differences in awareness, availability, and actual usage by patients.

Second, to determine the underlying adoption factors of digital use cases in family medicine, we formulated hypotheses B and C. Our rapid literature review, which aligned with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) extension for Rapid Reviews (PRISMA-RR) and recommendations by Tricco et al [37] and King et al [38], showed that different use cases have different factors driving technology acceptance [39,40]. The detailed approach and results can be found in the checklist in Multimedia Appendix 1.

Hypothesis B is as follows: Different use cases have different significant technology acceptance factors driving behavioral intention to use.

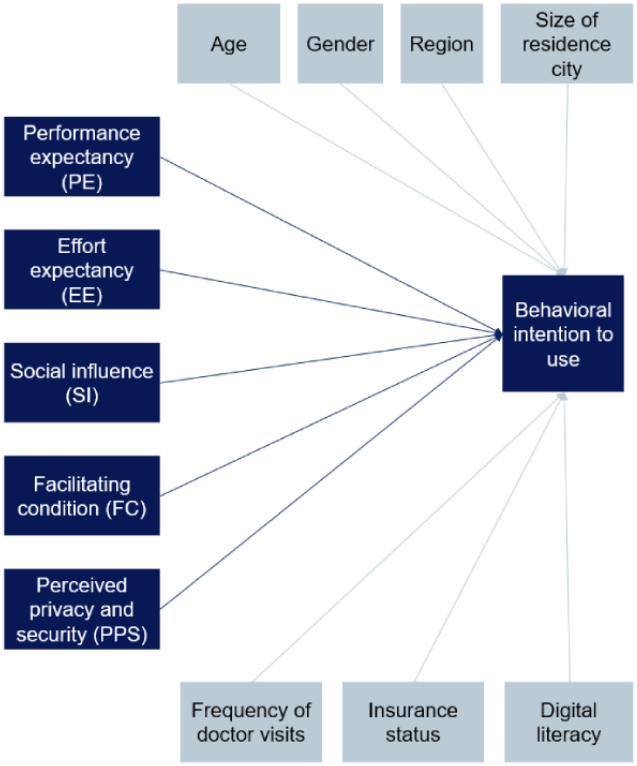
Various studies have indicated that sociodemographic factors affect technology acceptance. Mueller et al [4] argued that living in rural areas with potentially less doctor density might increase the likelihood of using digital offerings such as video consultation as it "saves the trip to the physician's practice." Thus, we hypothesized that the degree of the intended use of digital offerings varies (eg, by region and size of a city).

XSL•FO
RenderX

Hypothesis C is as follows: Different use cases have different sociodemographic characteristics impacting behavioral intention to use.

While hypothesis A pertains to indicated usage behavior, hypotheses B and C test the effect of a respective technology acceptance factor on the behavioral intention to use a specific digital use case in family medicine. The research model is depicted in Figure 1.

Figure 1. Research model. The effects of sociodemographic and technology acceptance factors on patients' intention to use a digital use case in family medicine in Germany in 2023 [20]. Dark blue indicates areas to collect evidence for hypothesis B, while gray indicates areas to test hypothesis C.



In these superordinate hypotheses, we replicated the same research model for each digital use case, with each testing the same subordinate hypotheses. Therefore, all subordinate hypotheses have the same components.

- Component 1 (illustrated in the research model): Respective technology acceptance factor [B]/respective sociodemographic characteristic [C] positively influences behavioral intention to use …
- Component 2: … a respective digital use case [online-booking platforms; digital anamnesis tools; video consultation platforms; electronic health records]

Survey Construction and Design

The questionnaire was constructed based on a rapid literature review on technology acceptance of digital use cases in family medicine (Multimedia Appendix 1) and our research model. Given its demonstrated validity and reliability in a family medicine setting (eg, average Cronbach α of .95; average "average variance contracted" [AVE] of 0.95; and average composite reliability [CR] value of 0.87), we leveraged an existing questionnaire assessing a comparable use case in a family medicine setting in the United Kingdom [20]. The questionnaire extends the UTAUT model [34] with additional items on *perceived privacy and security* [35], which also emerged in our literature review as an additional element to understand technology acceptance in health care settings [18,40]. Dropped items with lower factor loadings in the original model of Abd-Alrazaq et al [20] were excluded from our questionnaire. For data restriction purposes, our study did not look at the actual use behavior.

To ensure construct comparability in different use cases, we replaced "patient portal" in the original questionnaire with the names of the other use cases. The questionnaire was then translated into German based on existing validated translations of the UTAUT construct [41] and pretested with 21 participants from the sample target group. Minor adjustments to answer options, the order of questions, and the wording of selected items were made (the translated survey can be found in Multimedia Appendix 2). From a methodological viewpoint, the survey study was aligned with CHERRIES (Checklist for Reporting Results of Internet E-Surveys) (Multimedia Appendix 3). The web-based survey was finally conducted between May 30, 2023, and November 22, 2023, using the software LimeSurvey (LimeSurvey GmbH).

Participants were shown brief descriptions of all the digital use cases in the scope of this paper to minimize any misunderstanding. Participants were then randomly assigned to 2 of the digital use cases for which they answered the technology acceptance items on a Likert scale. The number of use cases per participant was later reduced from 2 to 1 due to higher-than-expected dropout rates. This adjustment was appropriate as the order of questions was not changed.

XSL•FO
RenderX

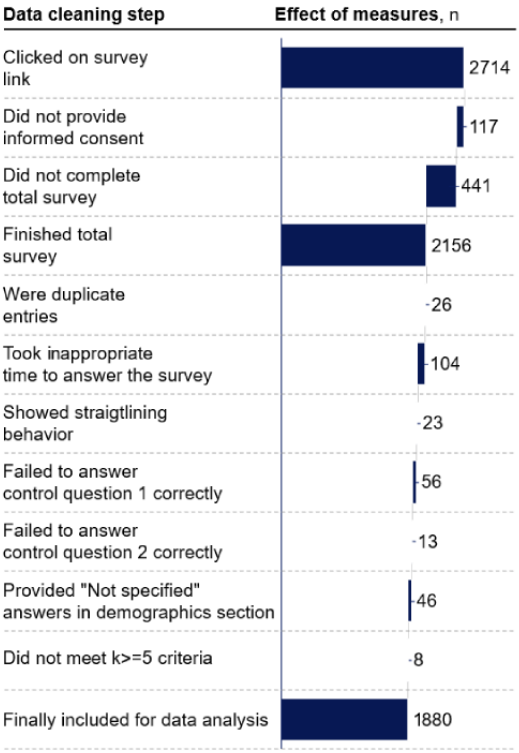
Recruitment and Sample Composition

Our convenience sampling approach aimed to recruit a large cross-sectional sample of adults living in Germany. Since the only limiting factor to participate in the study was reaching legal age, our sample was broad, which is reflected in the variety of recruitment channels. Participants were mainly recruited via the Social Sciences' Panel, a convenience pool project for scientific research, as well as social media platforms such as LinkedIn, Instagram, and Facebook; online (research) forums; and flyers in the waiting rooms of family practices and other facilities. The questionnaire was only posted online in the open-access mode, and no responses from other sources (eg, physical questionnaires) were manually added. Further details, such as the wording of the advertisement of the survey, can be found in Multimedia Appendix 3.

Applied data cleaning mechanisms included removing duplicate entries, excluding the fastest 5% responses, eliminating entries with no reasonable answer behavior (straight lining or incorrectly answering 2 control questions), and deleting entries with no specified demographics and those not meeting the k≥5 criteria [42,43]. We believe that these data-cleaning measures contributed to the validity and reliability of our study. First, we avoided overrepresentation of multiple entries by single participants. Second, we enhanced response quality by minimizing obvious nondiligent answer behavior. Finally, we ensured meaningful analysis by eliminating "not specified" responses while at the same time adhering to our privacy policy by deleting responses that failed the k≥5 criteria.

Details of data cleaning measures can be found in Figure 2.

Figure 2. Included responses for data analysis. Impact of dropout and data cleaning measures.



Empirical Analysis

Validation of the research model and statistical analysis via IBM SPSS 29 were performed separately for each use case in 3 steps.

First, Cronbach α values were calculated to evaluate the internal consistency of the study. The acceptable cutoff point in line with our exploratory research design was set at .60 [44]. Second, descriptive statistics were applied to test hypothesis A that different use cases have differences in awareness, availability, and actual usage among patients. Third, to test our hypotheses B and C regarding whether different use cases have different significant factors influencing behavioral intention to use, each research model was tested separately via multiple hierarchical regression. Thus, we tested the subordinated hypotheses for each digital use case.

The hierarchical regression was adequate to differentiate between the effects of technology acceptance and sociodemographic factors on intention to use. Therefore, we collected evidence for hypotheses B and C with a consistent analytical approach. Additionally, given the 4 different research models and a subsequent high number of subordinate hypotheses, a robust yet straightforward analytical approach was deemed necessary. The statistical significance level was set at $P<.05$ [45].

The first block entered into each model included only sociodemographic variables (age, gender, region of residency, size of the residence city, frequency of doctor visits, insurance status, and digital literacy). This allowed us to isolate the effects of different sociodemographic factors on intention to use, validating hypothesis C. The second block added the adapted UTAUT factors into the model (*performance expectancy*, *effort expectancy*, *social influence*, *facilitating condition*, and

XSL•FO
RenderX

perceived privacy and security). The final model addressed hypothesis B, testing the effect of technology acceptance factors on the intention to use a specific use case.

Four requirements for multiple regression models have been tested and confirmed for each of the use cases. First, multicollinearity can be rejected as all variance inflation factor (VIF) values were below 10 [46]. Second, autocorrelation can be rejected as the Durbin-Watson value for each use case was close to 2 [47]. Third, the normal distribution of residuals can be confirmed after visual inspection of the histogram of the standardized residuals. Lastly, homoscedasticity can be assumed as the regression of the standardized residual versus standardized predicted value scatter plot did not yield any considerable pattern.

By assessing usage rates, testing the technology acceptance model for each digital use case, and comparing our results across use cases, we collected evidence to assess our hypotheses A, B, and C. We accepted or rejected these superordinated hypotheses based on varying usage rates or different factors impacting the behavioral intention to use in different use cases (the decision was further validated with our research team).

Ethical Considerations

Our human subject research study was approved by the Ethics Committee of Witten/Herdecke University (Nr. S-245/2022). All participants provided informed consent in line with the GDPR (General Data Protection Regulation) guidelines prior to the start of the survey and had the ability to opt-out. No personal data were gathered, and sociodemographic data points were anonymized in accordance with k≥5 anonymity. All participants filled out the questionnaire voluntarily, and no compensation was offered.

Results

Internal Consistency and Sample Size

We calculated Cronbach α values for all 5 technology acceptance factors and the intention to use a specific digital use case. Therefore, we validated 24 variables across all 4 models. Of the 24 variables across all use cases, 23 met the threshold of >.60 [48], while 1 variable showed a Cronbach α of .57 (scheduling doctor appointments online: facilitating condition). The Cronbach α values for all 5 technology acceptance factors and intention to use are shown in Table 1.

Additionally, we leveraged the convenience sampling strategy to gather a large cross-sectional sample of German-speaking adults. Our final sample included 1880 adult respondents living in Germany. The demographic characteristics of the sample are shown in Table 2.

As the number of assigned use case assessments was reduced during the study, 175 out of the 1880 participants initially assessed the technology acceptance of 2 digital use cases, resulting in a total of 2055 technology acceptance assessments, with limited overlap between the groups (receiving medical consultations via video: n=494 [number of respondents who also assessed other use cases: n=80]; scheduling doctor appointments online: n=542 [number of respondents who also assessed other use cases: n=75]; storing personal medical information via electronic health records: n=528 [number of respondents who also assessed other use cases: n=103]; providing personal information before consultation digitally [digital anamnesis]: n=491 [number of respondents who also assessed other use cases: n=92]).

Table 1. Construct reliability (Cronbach α) in each digital use case.

Variable	Number of items	Digital use case			
		Receiving medical consultations via video	Scheduling doctor appointments online	Storing personal medical information via electronic health records	Providing personal information before consultation digitally (digital anamnesis)
Behavioral intention to use	3	.91	.90	.92	.90
Performance expectancy	3	.88	.85	.85	.87
Effort expectancy	4	.75	.71	.74	.76
Social influence	3	.91	.92	.91	.92
Facilitating condition	3	.70	.57	.60	.62
Perceived privacy and security	3	.71	.61	.74	.73

XSL•FO
RenderX

Table 2. Characteristics of the adult respondents living in Germany in 2023.

Variable	Value (N=1880)
Gender, n (%)	
Male	821 (43.7)
Female	1059 (56.3)
Age group (years), n (%)	
18-30 years	295 (15.7)
31-50 years	519 (27.6)
51-65 years	608 (32.3)
>65 years	458 (24.4)
Frequency group (visits in the last 12 months), n (%)	
0	132 (7.0)
1	362 (19.3)
2-3	705 (37.5)
>3	681 (36.2)
Insurance status, n (%)	
Statutorily	1459 (77.6)
Privately	421 (22.4)
Population size of residence (inhabitants), n (%)	
≤20,000	525 (27.9)
20,001-200,000	581 (30.9)
>200,000	774 (41.2)
Region, n (%)	
East	437 (23.2)
North	326 (17.3)
South	365 (19.4)
West	752 (40.0)
Self-assessed digital literacy, n (%)	
Low (1-4)	103 (5.5)
Medium (5-7)	455 (24.2)
High (8-10)	1322 (70.3)
Digital literacy score, mean (SD)	8.0 (1.8)

Current Patient Adoption Rates of Digital Use Cases in Family Medicine

There was support for hypothesis A that different use cases have differences in awareness, availability, and actual usage by patients.

Overall, low offering and usage rates of digital use cases were evident. For example, nearly all respondents (1750/1880, 93.1%) were aware of online booking platforms; however, only approximately one-third (617/1880, 32.8%) stated that this service is offered in their family practice. Nevertheless, the vast majority (537/617, 87.0%) of those respondents who were offered online booking services in their family practice reported using them. This usage-to-availability ratio was higher for scheduling doctor appointments online than for video consultations where only less than half of the participants (94/225, 41.8%) stated usage. The results are presented in Table 3. The share of respondents using digital use cases when offered at their family doctor was as follows: receiving medical consultations via video, 94/225 (41.8%); scheduling doctor appointments online, 537/617 (87.0%); storing personal medical information via electronic health records, 108/166 (65.1%); providing personal information before consultation digitally (digital anamnesis), 53/78 (68.0%).

XSL•FO
RenderX

Table 3. Patient awareness, availability, and patient usage of selected digital use cases in family medicine in Germany in 2023.

Statement and digital use case	Response option, n (%)				
	Yes	No	I don't know	Not offered in my family practice	Not specified
"I am aware of this use case"					
VC[a]	1302 (69.3)	571 (30.4)	—[b]	—	7 (0.4)
OS[c]	1750 (93.1)	130 (6.9)	—	—	0 (0.0)
EHR[d]	1220 (64.9)	638 (33.9)	—	—	22 (1.2)
DA[e]	784 (41.7)	1075 (57.2)	—	—	21 (1.1)
"The use case is offered at my family doctors "					
VC	225 (12.0)	1092 (58.1)	—	559 (29.7)	4 (0.2)
OS	617 (32.8)	1031 (54.8)	—	229 (12.2)	3 (0.2)
EHR	166 (8.8)	690 (36.7)	—	1016 (54.0)	8 (0.5)
DA	78 (4.0)	956 (50.9)	—	835 (44.4)	11 (1.0)
"I use this digital use case at my family doctors"					
VC	94 (5.0)	909 (48.4)	—	793 (42.2)	84 (5.0)
OS	537 (28.6)	583 (31.0)	—	712 (37.9)	48 (3.0)
EHR	108 (5.7)	907 (48.2)	—	622 (33.1)	243 (12.9)
DA	53 (3.0)	877 (46.6)	—	759 (40.4)	191 (10.2)

[a]VC: receiving medical consultations via video.

[b]Not applicable.

[c]OS: scheduling doctor appointments online.

[d]EHR: storing personal medical information via electronic health records.

[e]DA: providing personal information before consultation digitally (digital anamnesis).

There was a low usage rate of digital use cases as only 16.2% (304/1880) of respondents agreed or fully agreed that the degree of digitalization is key when selecting a family practice. Furthermore, the willingness to change the family practice when a particular digital use case is unavailable was found to depend on the use case (69/1880, 3.7% for receiving medical consultations via video; 200/1880, 10.6% for scheduling doctor appointments online; 131/1880, 7.0% for storing personal medical information via electronic medical records; and 82/1880, 4.4% for providing personal information before consultation digitally [digital anamnesis]). Nevertheless, regression analysis showed that when selecting a family practice, participants with higher digital literacy tended to assign a higher value to the degree of digitalization (B=0.226, SE 0.023; β=.223; $P<.001$).

Factors Explaining Technology Acceptance in Family Medicine

To better understand the predictors of technology acceptance, we conducted a multiple hierarchical linear regression analysis on the intention to use each use case. Block 1 collects evidence to validate hypothesis C, which involves testing the sole effect of sociodemographic variables on intention to use, while block 2 refers to hypothesis B, which involves validating the effect of technology acceptance factors. In doing so, we did not moderate prior experience with a particular use case to reduce the complexity of each model. The results are illustrated in Table 4. All details of the respective models can be found in Multimedia Appendix 4.

XSL•FO
RenderX

Table 4. Standardized coefficients for hierarchical linear regression models explaining patients' intention to use a specific digital use case in family medicine in Germany in 2023.

Block and variable	Receiving medical consultations via video (n=494)		Scheduling doctor appointments online (n=542)		Storing personal medical information via electronic health records (n=528)		Providing personal information before consultation digitally (digital anamnesis) (n=491)	
	β	*P* value	β	*P* value	β	*P* value	β	*P* value
Block 1: Sociodemographic variables								
Frequency of doctor visits (vs no doctor visit in the last 12 months)								
1 visit	.032	.69	.213	.01[a]	–.008	.92	.025	.74
2-3 visits	.134	.15	.233	.01[a]	.142	.10	.089	.27
>3 visits	.137	.14	.224	.01[a]	.066	.45	.136	.10
Gender (vs male)								
Female	.900	.04[a]	.084	.04[a]	–.101	.02[a]	–.037	.41
Age (vs 18-30 years)								
31-50 years	.087	.13	–.011	.84	–.017	.75	.013	.82
51-65 years	.107	.09	–.023	.69	–.021	.71	–.125	.04[a]
>65 years	.032	.60	–.083	.16	.020	.74	–.158	.01[a]
Size of the residence city (vs ≤20,000)								
20,001-200,000	–.025	.62	.022	.66	.056	.29	.007	.90
>200,000	–.057	.30	.032	.53	.023	.67	–.005	.92
Region (vs East)								
South	–.061	.22	.005	.93	–.028	.57	–.038	.48
West	–.061	.25	–.035	.48	–.031	.53	.022	.69
North	–.026	.64	.029	.58	.052	.33	.020	.74
Type of insurance	–.044	.32	–.010	.81	–.022	.60	.067	.13
Digital literacy	.335	<.001[a]	.408	<.001[a]	.329	<.001[a]	.250	<.001[a]
Block 2[b]: Sociodemographic variables + technology acceptance factors								
Performance expectancy	.527	<.001[a]	.513	<.001[a]	.348	<.001[a]	.410	<.001[a]
Effort expectancy	.135	.005[a]	.218	<.001[a]	.096	.03[a]	.236	<.001[a]
Social influence	.206	<.001[a]	.054	.07	.101	<.001[a]	.173	<.001[a]
Facilitating condition	.039	.36	.028	.44	.198	<.001[a]	.085	.045[a]
Perceived privacy and security	.053	.10	.094	<.001[a]	.243	<.001[a]	.104	.003[a]

[a]Statistically significant (P<.05).

[b]The effect of sociodemographic variables is not displayed as the focus of the analysis is on the effect of technology acceptance factors on intention to use.

All final models (including steps 1 and 2) significantly predicted technology acceptance, whereas the addition of the UTAUT variables in the second step increased the explained variance significantly in all models for all use cases. In the first step, sociodemographic variables alone explained on average 14.6% of the intention to use a specific digital use case (receiving medical consultations via video: R^2=0.124; $F_{14,479}$=4.849; P<.001; scheduling doctor appointments online: R^2=0.189; $F_{14,527}$=8.780; P<.001; storing personal medical information via electronic health records: R^2=0.150; $F_{14,513}$=6.484; P<.001; providing personal information before consultation digitally [digital anamnesis]: R^2=0.121; $F_{14,476}$=4.698; P<.001). The final models explained on average 62.8% of the variance in intention to use (receiving medical consultations via video: R^2=0.624; $F_{19,474}$=41.356; P<.001; scheduling doctor appointments online: R^2=0.624; $F_{19,522}$=45.617; P<.001; storing personal medical

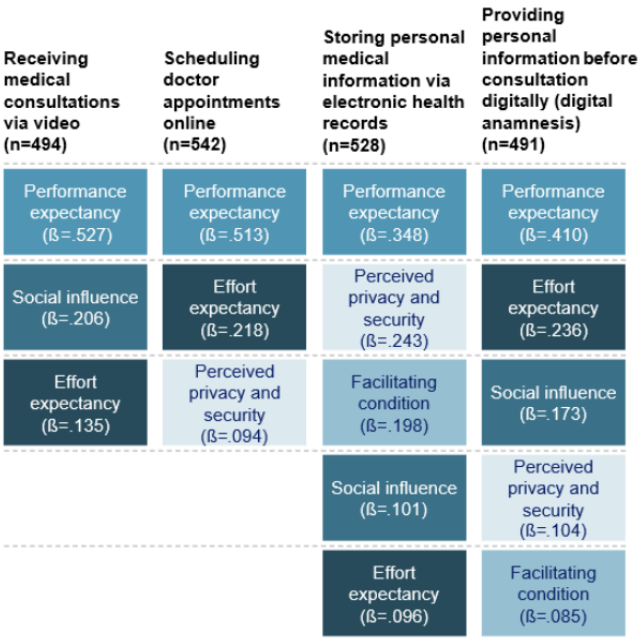
XSL•FO
RenderX

information via electronic health records: R^2=0.631; $F_{19,508}$=45.655; P<.001; providing personal information before consultation digitally [digital anamnesis]: R^2=0.631; $F_{19,471}$=42.364; P<.001).

There was support for hypothesis B that different use cases have different significant technology acceptance factors. Only *performance expectancy* and *effort expectancy* were significant predictors of technology acceptance for all use cases. Additionally, the magnitude of the effects of technology acceptance factors differed between use cases (*performance expectancy*: receiving medical consultations via video: B=0.603, SE 0.049; β=.527; P<.001; scheduling doctor appointments online: B=0.566, SE 0.043; β=.513; P<.001; storing personal medical information via electronic health records: B=0.405, SE 0.047; β=.348; P<.001; providing personal information before consultation digitally [digital anamnesis]: B=0.434, SE 0.048; β=.410; P<.001; *effort expectancy*: receiving medical consultations via video: B=0.236, SE 0.084; β=.135; P=.01; scheduling doctor appointments online: B=0.391, SE 0.075; β=.218; P<.001; storing personal medical information via electronic health records: B=0.153, SE 0.070; β=.096; P=.03; providing personal information before consultation digitally [digital anamnesis]: B=0.366, SE 0.073; β=.236; P<.001).

We found that *social influence* was significant for all use cases, except for scheduling doctor appointments online (receiving medical consultations via video: B=0.227, SE 0.036; β=.206; P<.001; storing personal medical information via electronic health records: B=0.108, SE 0.033; β=.101; P=.001; providing personal information before consultation digitally [digital anamnesis]: B=0.186, SE 0.036; β=.173; P<.001). Moreover, *perceived privacy and security* was significant for all use cases, except for receiving medical consultations via video (scheduling doctor appointments online: B=0.117, SE 0.037; β=.094; P=.002; storing personal medical information via electronic health records: B=0.284, SE 0.040; β=.243; P<.001; providing personal information before consultation digitally [digital anamnesis]: B=0.121, SE 0.040; β=.104; P=.003). *Facilitating condition* was only significant for storing personal medical information via electronic health records (B=0.299, SE 0.061; β=.198; P<.001) and providing personal information before consultation digitally (digital anamnesis) (B=0.127, SE 0.063; β=.085; P=.045). A simplified illustration of the different magnitudes of effects is presented in Figure 3.

Figure 3. Magnitudes of the effects of technology acceptance factors on the behavioral intention to use a digital use case in family medicine in Germany in 2023 (ranking based on β).



There was partial support for hypothesis C that different use cases have different sociodemographic factors impacting the behavioral intention, as self-assessed digital literacy had a positive effect for all use cases (receiving medical consultations via video: B=0.341, SE 0.049; β=.335; P<.001; scheduling doctor appointments online: B=0.322, SE 0.033; β=.408; P<.001; storing personal medical information via electronic health records: B=0.308, SE 0.041; β=.329; P<.001; providing personal information before consultation digitally [digital anamnesis]: B=0.239, SE 0.044; β=.250; P<.001).

Further, significant effects of sociodemographic variables on intention to use depended on the use case. Gender (female vs male) had an effect on 3 digital use cases, but the direction of that effect varied (receiving medical consultations via video: B=0.321, SE 0.159; β=.090; P=.04; scheduling doctor appointments online: B=0.247, SE 0.122; β=.084; P=.04; storing personal medical information via electronic health records: B=–0.356, SE 0.149; β=–.101; P=.02). Additionally, the frequency of doctor visits had a positive effect on the intention to use online appointment systems and age had a negative effect on digital anamnesis tools. Compared with respondents who had not visited a doctor in the last 12 months, all frequency groups showed a positive effect (1 visit in the last 12 months: B=0.783, SE 0.285; β=0.213; P=.006; 2-3 visits: B=0.698, SE 0.267; β=.233; P=.009; >3 visits: B=0.690, SE 0.270; β=.224;

XSL•FO
RenderX

P=.01). Compared with the age group of 18-30 years, there were negative effects for the age groups of 51-65 years (B=–0.476, SE 0.235; β=–.125; *P*=.04) and >65 years (B=–0.614, SE 0.239; β=–.158; *P*=.01).

The size of the residence city, region, and type of insurance did not affect the intention to use any digital use cases. Remarkably, Welch ANOVA found that digital literacy was dependent on age (Welch $F_{3,968.29}$=53.441; *P*<.001).

Discussion

Principal Findings

This research paper aimed to compare the different technology acceptance factors of patients for key digital use cases in family medicine and to support family doctors in selecting and implementing digital use cases. Our results indicated that the degree of digitalization of family practice is not critically important for patients. Additionally, our results fully supported hypothesis A, highlighting that different use cases have differences in awareness, availability, and actual usage among patients. Further, we validated hypothesis B, showing that different use cases are affected by different technology acceptance factors. Nevertheless, our results only partially supported hypothesis C as digital literacy had a significant effect on the intention to use for all use cases.

Thus, the varying relative importance of technology acceptance factors for each digital use case emphasizes the need for a nuanced selection and implementation approach of digital use cases for family practice. Simultaneously, the partial rejection of different effects of sociodemographic variables could imply that our recommendations can be applied in different sociodemographic contexts.

Comparison With Prior Work

We introduced a systematic assessment of technology acceptance for online booking platforms and digital anamnesis tools as existing studies concentrated only on video consultation platforms and electronic health records. Hence, we could only compare our results to prior work on these 2 use cases.

Receiving Medical Consultations via Video

The strongest effect of *performance expectancy* has been validated in various studies [18,49]. However, the literature indicates mixed results for the effect of *effort expectancy* and *social influence* on the intention to use video consultation platforms. For example, Schmitz et al [18] rejected the effect of *social influence*, claiming that medical appointments are highly personal, and thus, the fact of undergoing a medical treatment is not subjective to the opinions of others, regardless of a virtual or physical setting. However, our findings showed that the mode of the consultation might be affected by other opinions. This is supported by Esber et al [17], who confirmed the significance of *social influence* on the acceptance of video consultations, and Mueller et al [4], who stressed the importance of social cues to drive the technology acceptance of preusers.

Additionally, we found support for our rejection of the effect of *facilitating condition* on the intention to use video consultations in previous work [18,39,49,50]. Schmitz et al [18] and Brooks [51] mention that video consultation technologies are comparable to already widespread technologies, such as FaceTime, and thus, people are already well equipped for their use in a medical setting.

Surprisingly, we rejected the effect of *perceived privacy and security* on the intention to use video consultations but accepted this effect for other use cases. This contrasts with the findings of Zobair et al [52] and Schmitz et al [18] who highlighted the effect of perceived security on the intention to use telemedicine or telehealth in Bangladesh and Germany. Viana Pereira et al [39] postulated that doctors could also provide advice on general health topics via video and thus would not cause confidentiality concerns among patients. Nevertheless, we argued that the neglect of the importance of *perceived privacy and security* for patients requires further validation.

Storing Personal Medical Information via Electronic Health Records

Performance expectancy was constantly confirmed as a significant predictor of intention to use [16,20,40,53-63]. Additionally, we found evidence for positive effects of *effort expectancy* [16,20,53,55-57,62] and *facilitating condition* [20,60] on intention to use. Nevertheless, our findings contradict those of Tavares and Olivera [59] who suggested that early adopters do not perceive obstacles in using electronic health records due to their higher cognitive ability and experience with both technological complexities and IT infrastructure. This is further supported by Aydin and Kumru [40] who rejected *facilitating condition* as a predictor of the intention to use electronic medical records for Gen-Z university students. We explain this discrepancy with our more diverse sample, which we believe is less biased toward high cognitive ability and technological experience.

Our findings of a positive effect of *social influence* confirm the findings of Tavares et al [16] who illustrated inadequate promotion activities as a possible reason for the failed implementation of a comparable health record technology [64]. Nevertheless, Abd-Alrazaq et al [20] identified *social influence* as a nonsignificant predictor for behavioral intention, explaining that usage is voluntary and thus less objective to *social influence*. We hypothesized that the ongoing discussion on the transition from an opt-in to an opt-out model in Germany (default mandatory use of electronic health records) could partially explain the significance of *social influence* in our research [10].

Further studies that assessed comparable constructs of *perceived security and privacy* confirmed the positive effect on the intention to use electronic health records [20,40,53,55,56].

Effect of Sociodemographic Variables on the Intention to Use Digital Use Cases

Except for digital literacy, we found no clear pattern of the effect of sociodemographic variables on the intention to use a specific digital use case. The vital role of digital literacy in technology acceptance is in line with research that postulates a digital divide of patient groups in engaging in digital use cases [65]. However, this role of digital literacy necessitates further differentiation. First, the participants in our study might be rather digitally savvy, and thus, this effect might be overgeneralized. Second,

XSL•FO
RenderX

we acknowledge different magnitudes of this effect across use cases, suggesting that inherent characteristics of use cases (eg, technical complexity) might impact the role of digital literacy in technology acceptance. In identifying digitally literate patients, our findings showed a negative effect of age on digital literacy, with further evidence in other studies [66].

Our study found no effect of the residence city size on patients' intention to use video consultations. This is particularly interesting as, for example, receiving video consultations is seen as a means to ensure adequate treatment opportunities in rural areas, with Kane and Gillis [67] highlighting that the use of telemedicine by physicians is higher in nonmetropolitan areas than in metropolitan areas in the United States. On the other hand, McGrail et al [68] noted no difference in the offering of telemedicine between rural and urban settings in Canada. This contradiction indicates that the effect of the residence city size on intention to use or actual usage depends on further factors (eg, the potential difference in the digital literacy of people living in rural vs metropolitan areas or the lack of availability of video consultation offerings in some regions).

Measures for Family Doctors to Increase Patient Usage of Digital Use Cases

Importance of the Digital Maturity of Family Practices for Patients

Family doctors should carefully assess whether investing in digital use cases is adequate to attract and sustain patients. Our results showed that the level of digitalization of family practices was not a vital criterion for patients when choosing a doctor. This confirms the findings of Kuruoglu et al [69] and Khatami et al [70] who did not mention the degree of digitalization as a selection criterion for patients.

Nevertheless, our results showed that patients with higher digital literacy had greater importance for the degree of digitalization when selecting a family practice. Thus, digital investment may be more vital for family doctors treating more digitally literate patients. Hence, family doctors who refuse to invest in digital technology could lose these patient groups to family doctors who invest in digital technology.

Selection of Digital Use Cases

Family doctors should consider patient demand in the selection of digital use cases. Our study highlights the attractiveness of use cases for patients by comparing offering and actual usage rates. According to this approach, scheduling doctor appointments online appeared to be the most attractive digital use case in Germany in 2023, followed by providing personal information before consultation digitally (digital anamnesis). Video consultation was the least favorable use case for patients. Interestingly, a 2023 study on the digital offerings of family doctors in Germany indicated that video consultations were nearly 4 times more often offered than digital anamnesis tools [71]. Nevertheless, we are aware that patient demand is only 1 factor for offering use cases [13], and further patient-specific and use case–specific barriers might bias this attractivity assessment. We excluded electronic health records as family doctors in Germany must document patient treatment within the respective electronic health record of the patient ("Elektronische Patientenakte" [10]). Hence, family doctors do not select this digital use case voluntarily.

Implementation of Digital Use Cases

To incentivize usage among patients, family doctors should adopt an implementation approach for each use case, given that we showed the varying importance of technology acceptance factors in each use case. Nevertheless, we identified 3 factors that affected the intention to use for all use cases, and these should always be considered by family doctors.

First, doctors should prioritize the value addition of the digital use case to the patient. This is especially important for video consultations. The focus on value addition to drive adoption aligns with other academic papers [18,49] that stress the ability of doctors to communicate the added value to the patient. Hence, family doctors should receive adequate training or communication material from providers to explain the added value to patients effectively.

Second, family doctors should assess the importance of an intuitive interface for solutions as *effort expectancy* had a significant effect across all use cases. Nevertheless, the relative importance of this factor varied by use case. Since we believe that *effort expectancy* is dependent on the design and user experience of solutions and family doctors have limited control over these, the correct initial selection of a solution provider is vital.

Third, family practices should consider the digital literacy of their patients as it is a predictor of the intention to use across all use cases. We acknowledge that the digital literacy of family doctors affects their own technology adoption [72]. Hence, we believe that not only family doctors themselves should ensure that patients with low digital literacy are not excluded from using digital use cases.

Further considerations for family doctors depend on the respective digital use case. To increase the use of video consultations, family doctors should additionally consider endorsements from testimonials or other influencing bodies. Surprisingly, we showed no effect of *perceived privacy and security* on the intention to use video consultations. This could indicate that this factor might be neglectable for family doctors; however, further validation is needed. On the contrary, *perceived privacy and security* is important for online booking platforms. Therefore, family doctors should articulate how online booking platforms ensure the privacy of end users and how platforms communicate about this. Additionally, *perceived privacy and security* is the second most important factor to drive the use of electronic health records. Therefore, family doctors should focus on whether patients feel their data are protected. While this factor is also to be considered for digital anamnesis tools, having an easy-to-use solution here is more important than for electronic health records.

Limitations and Further Research

To the best of our knowledge, our web-based survey with 1880 included participants is one of the largest recent studies assessing technology acceptance in a health care setting. By comparing 4 distinct use cases, we provided a new nuanced assessment

XSL•FO
RenderX

and avoided attitude generalization. Contrary to studies that only included users or nonusers of a digital use case [20,54], our research included both users and nonusers in a family medicine context.

Nevertheless, we acknowledge that our study had some limitations. First, due to our voluntary and web-based survey approach, our sample might be biased toward participants with higher digital literacy and higher interest in digital health topics. This excluded patients with no access to the internet. However, we believe our sample is appropriate as digital use cases require internet access and our sample includes these potential users. Additionally, we tried to minimize bias by distributing physical questionnaires in family practices; however, responses to distributed flyers were limited. Second, replicating the same construct with the exact wordings for each use case potentially did not address use case specifics and thus increased the risk of ambiguity of items. For example, an electronic health record is not offered by the family doctor but rather by the insurance provider. In addition, the questionnaire was translated into German for the first time, which might explain the presence of a few low Cronbach α values. Dropping items would have significantly increased the Cronbach α values; however, to ensure comparability with the original questionnaire, we refrained from doing so. For example, for the worst Cronbach α (scheduling doctor appointments online: *facilitating condition*), only 1 item showed a low correlation with the other items. Furthermore, we did not have access to data validating the offer rates and usage rates of digital use cases by the participants but rather relied on the subjective inputs of the respondents. Nevertheless, as we offered the response options "I don't know" and "Not specified," we hypothesized high validity in the remaining responses with "Yes" or "No" answer entries. In addition, we did not account for prior experience, potentially not differentiating the distinct effects of expectations or past use on technology acceptance. Finally, we only looked at behavioral intention to use; however, various studies have highlighted a gap between intention and usage behavior, which is referred to as the intention-behavior gap [73]. Thus, our research did not predict the actual usage of digital use cases.

There are 3 areas for future research. First, it is vital to understand the technology acceptance of people with limited access to the internet (eg, people living in elderly homes) or low digital literacy and discuss measures on how to minimize the digital divide. Second, the effects of the "Digital Act (Digital-Gesetz)" of Germany [10] on usage by patients should be examined, particularly the mandatory implementation of electronic health records. Third, our research focused on Germany, and thus, the applicability of our findings to other countries could be discussed.

Conclusion

Patients assign lower importance to the degree of digitalization of their family practices. Additionally, the usage of digital use cases remains low. Therefore, family doctors should carefully assess whether investments in digital use cases are adequate. Nevertheless, as the benefits of these solutions are established in the literature, we see future potential for higher adoption. For capturing this adoption potential, we believe in the importance of the role of family doctors owing to the established and trusted patient-doctor relationship. To leverage this role and increase patient usage, family doctors should align their selection and implementation criteria with our findings of the varying technology acceptance factors. We showed that *performance expectancy*, *effort expectancy*, and *digital literacy* have positive effects on the intention to use across all use cases, but further acceptance factors depend on the use case. To illustrate, family doctors should always emphasize the perceived added value for patients, whereas the neglect of privacy and security aspects for video consultation platforms requires further validation. Family doctors should always prioritize *performance expectancy*, while guides for the selection and implementation of digital tools might emerge for each digital use case.

Acknowledgments

We thank the Social Sciences' Panel (a convenience pool project of the Institute for Communication Science and Media Research [LMU Munich] and the German Communication Association [DGPuK]) for their support in the recruitment of participants for our survey.

Data Availability

The datasets incorporated and examined during this research project are available from the corresponding author upon reasonable request and on consideration of restrictions associated with the ethics approval.

Authors' Contributions

JB, SM, and LF conceptualized the main research question and derived the appropriate research design. JB conducted the literature review, created the survey in an online tool, and analyzed the gathered information with adequate statistical measures. SR provided feedback regarding the selection of appropriate statistical tools and reviewed the statistical data analysis. JB authored the initial manuscript for which SM, LF, AM, and SR contributed revisions. SM orchestrated the research project. All authors reviewed the manuscript for submission.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Literature review (including the PRISMA-ScR [Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews] checklist).
[DOCX File , 2575 KB - formative_v9i1e58867_app1.docx]

Multimedia Appendix 2
Translated survey questionnaire.
[DOCX File , 50 KB - formative_v9i1e58867_app2.docx]

Multimedia Appendix 3
CHERRIES (Checklist for Reporting Results of Internet E-Surveys).
[DOCX File , 53 KB - formative_v9i1e58867_app3.docx]

Multimedia Appendix 4
Overview of regression coefficients for the regression models predicting the intention to use digital use cases.
[DOCX File , 92 KB - formative_v9i1e58867_app4.docx]

References

1. Bashshur RL, Howell JD, Krupinski EA, Harms KM, Bashshur N, Doarn CR. The empirical foundations of telemedicine interventions in primary care. Telemed J E Health 2016 May;22(5):342-375 [FREE Full text] [doi: 10.1089/tmj.2016.0045] [Medline: 27128779]
2. Wanderås MR, Abildsnes E, Thygesen E, Martinez SG. Video consultation in general practice: a scoping review on use, experiences, and clinical decisions. BMC Health Serv Res 2023 Mar 30;23(1):316 [FREE Full text] [doi: 10.1186/s12913-023-09309-7] [Medline: 36997997]
3. Wangler J, Jansky M. How can primary care be secured in the long term? - a qualitative study from the perspective of general practitioners in Germany. Eur J Gen Pract 2023 Dec;29(1):2223928 [FREE Full text] [doi: 10.1080/13814788.2023.2223928] [Medline: 37387362]
4. Mueller M, Knop M, Niehaves B, Adarkwah CC. Investigating the acceptance of video consultation by patients in rural primary care: empirical comparison of preusers and actual users. JMIR Med Inform 2020 Oct 22;8(10):e20813 [FREE Full text] [doi: 10.2196/20813] [Medline: 32969339]
5. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of digital technologies in health care during the COVID-19 pandemic: systematic review of early scientific literature. J Med Internet Res 2020 Nov 06;22(11):e22280 [FREE Full text] [doi: 10.2196/22280] [Medline: 33079693]
6. AMA digital health care 2022 study findings. American Medical Association. URL: https://www.ama-assn.org/about/research/ama-digital-health-care-2022-study-findings [accessed 2023-12-06]
7. TI-Dashboard. Gematik. URL: https://www.gematik.de/telematikinfrastruktur/ti-dashboard [accessed 2024-01-31]
8. Versicherte in der gesetzlichen Krankenversicherung. Bundesministerium für Gesundheit. URL: https://www.bundesgesundheitsministerium.de/gesetzlich-versicherte [accessed 2024-01-31]
9. Appointments in General Practice. NHS England. URL: https://digital.nhs.uk/data-and-information/publications/statistical/appointments-in-general-practice#about-this-publication [accessed 2024-02-13]
10. Gesetz zur Beschleunigung der Digitalisierung des Gesundheitswesens (Digital-Gesetz – DigiG). Bundesministerium für Gesundheit. URL: https://www.bundesgesundheitsministerium.de/ministerium/gesetze-und-verordnungen/guv-20-lp/digig.html [accessed 2024-01-31]
11. Securing Excellence in Primary Care (GP) Digital Services: The Primary Care (GP) Digital Services Operating Model 2021-2023. NHS England. URL: https://www.england.nhs.uk/digitaltechnology/digital-primary-care/gp-digital-services-operating-model-21-23/ [accessed 2024-02-29]
12. Breedvelt JJ, Zamperoni V, Kessler D, Riper H, Kleiboer AM, Elliott I, et al. GPs' attitudes towards digital technologies for depression: an online survey in primary care. Br J Gen Pract 2019 Mar;69(680):e164-e170 [FREE Full text] [doi: 10.3399/bjgp18X700721] [Medline: 30559111]
13. Weik L, Fehring L, Mortsiefer A, Meister S. Understanding inherent influencing factors to digital health adoption in general practices through a mixed-methods analysis. NPJ Digit Med 2024 Feb 27;7(1):47 [FREE Full text] [doi: 10.1038/s41746-024-01049-0] [Medline: 38413767]
14. De Santis KK, Jahnel T, Sina E, Wienert J, Zeeb H. Digitization and health in Germany: cross-sectional nationwide survey. JMIR Public Health Surveill 2021 Nov 22;7(11):e32951 [FREE Full text] [doi: 10.2196/32951] [Medline: 34813493]
15. Napitupulu D, Yacub R, Putra A. Factor influencing of telehealth acceptance during COVID-19 outbreak: extending UTAUT model. IJIES 2021 Jun 30;14(3):267-281. [doi: 10.22266/ijies2021.0630.23]
16. Tavares J, Goulão A, Oliveira T. Electronic Health Record Portals adoption: Empirical model based on UTAUT2. Inform Health Soc Care 2018 Mar;43(2):109-125. [doi: 10.1080/17538157.2017.1363759] [Medline: 29035646]

XSL•FO
RenderX

17. Esber A, Teufel M, Jahre L, In der Schmitten J, Skoda E, Bäuerle A. Predictors of patients' acceptance of video consultation in general practice during the coronavirus disease 2019 pandemic applying the unified theory of acceptance and use of technology model. Digit Health 2023;9:20552076221149317 [FREE Full text] [doi: 10.1177/20552076221149317] [Medline: 36815005]
18. Schmitz A, Díaz-Martín A, Yagüe Guillén MJ. Modifying UTAUT2 for a cross-country comparison of telemedicine adoption. Comput Human Behav 2022 May;130:107183 [FREE Full text] [doi: 10.1016/j.chb.2022.107183] [Medline: 35017788]
19. Dahlhausen F, Zinner M, Bieske L, Ehlers JP, Boehme P, Fehring L. There's an app for that, but nobody's using it: Insights on improving patient access and adherence to digital therapeutics in Germany. Digit Health 2022;8:20552076221104672 [FREE Full text] [doi: 10.1177/20552076221104672] [Medline: 35811758]
20. Abd-Alrazaq A, Bewick BM, Farragher T, Gardner P. Factors affecting patients' use of electronic personal health records in England: cross-sectional study. J Med Internet Res 2019 Jul 31;21(7):e12373 [FREE Full text] [doi: 10.2196/12373] [Medline: 31368442]
21. The European Definition of General Practice / Family Medicine. WONCA. URL: https://www.woncaeurope.org/file/41f61fb9-47d5-4721-884e-603f4afa6588/WONCA_European_Definitions_2_v7.pdf [accessed 2024-01-12]
22. Keuper J, Batenburg R, Verheij R, van Tuyl L. Use of e-health in Dutch general practice during the COVID-19 pandemic. Int J Environ Res Public Health 2021 Nov 26;18(23):12479 [FREE Full text] [doi: 10.3390/ijerph182312479] [Medline: 34886204]
23. Farias FACD, Dagostini CM, Bicca YA, Falavigna VF, Falavigna A. Remote patient monitoring: a systematic review. Telemed J E Health 2020 May;26(5):576-583. [doi: 10.1089/tmj.2019.0066] [Medline: 31314689]
24. Zhao P, Yoo I, Lavoie J, Lavoie BJ, Simoes E. Web-based medical appointment systems: a systematic review. J Med Internet Res 2017 Apr 26;19(4):e134 [FREE Full text] [doi: 10.2196/jmir.6747] [Medline: 28446422]
25. What is an electronic health record (EHR)? HealthIT. URL: https://www.healthit.gov/faq/what-electronic-health-record-ehr [accessed 2023-12-04]
26. Kopp M, Wetzl M, Geissler F, Roth JP, Wallner R, Hoefler D, et al. Structured digital self-assessment of patient anamnesis prior to computed tomography: performance evaluation and added value. J Med Syst 2021 Jan 28;45(3):30 [FREE Full text] [doi: 10.1007/s10916-020-01690-8] [Medline: 33511485]
27. Robinson SK, Meisnere M, Phillips RL, McCauley L, editors. Implementing High-Quality Primary Care: Rebuilding the Foundation of Health Care. Washington, DC: National Academies Press; 2021.
28. Popa V, Geissler J, Vermeulen R, Priest E, Capperella K, Susuzlu G, et al. Delivering digital health solutions that patients need: a call to action. Ther Innov Regul Sci 2024 Mar;58(2):236-241 [FREE Full text] [doi: 10.1007/s43441-023-00592-4] [Medline: 38082118]
29. Taherdoost H. A review of technology acceptance and adoption models and theories. Procedia Manufacturing 2018;22:960-967. [doi: 10.1016/j.promfg.2018.03.137]
30. AlQudah AA, Al-Emran M, Shaalan K. Technology acceptance in healthcare: a systematic review. Applied Sciences 2021 Nov 09;11(22):10537. [doi: 10.3390/app112210537]
31. Heinsch M, Wyllie J, Carlson J, Wells H, Tickner C, Kay-Lambkin F. Theories informing eHealth implementation: systematic review and typology classification. J Med Internet Res 2021 May 31;23(5):e18500 [FREE Full text] [doi: 10.2196/18500] [Medline: 34057427]
32. Rahi S, Khan MM, Alghizzawi M. Factors influencing the adoption of telemedicine health services during COVID-19 pandemic crisis: an integrative research model. Enterprise Information Systems 2020 Dec 08;15(6):769-793. [doi: 10.1080/17517575.2020.1850872]
33. Hsieh H, Lai J, Chuang B, Tsai C. Determinants of telehealth continuance intention: a multi-perspective framework. Healthcare (Basel) 2022 Oct 15;10(10):2038 [FREE Full text] [doi: 10.3390/healthcare10102038] [Medline: 36292485]
34. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
35. Whetstone M, Goldsmith R. Factors influencing intention to use personal health records. Intl J of Pharm & Health Mrkt 2009 Apr 03;3(1):8-25. [doi: 10.1108/17506120910948485]
36. Doctolib Digital Health Report 2023. Doctolib. URL: https://about.doctolib.de/news/doctolib-digital-health-report-2023/ [accessed 2024-01-29]
37. Tricco AC, Zarin W, Antony J, Hutton B, Moher D, Sherifali D, et al. An international survey and modified Delphi approach revealed numerous rapid review methods. J Clin Epidemiol 2016 Feb;70:61-67. [doi: 10.1016/j.jclinepi.2015.08.012] [Medline: 26327490]
38. King VJ, Stevens A, Nussbaumer-Streit B, Kamel C, Garritty C. Paper 2: Performing rapid reviews. Syst Rev 2022 Jul 30;11(1):151 [FREE Full text] [doi: 10.1186/s13643-022-02011-5] [Medline: 35906677]
39. Viana Pereira F, Tavares J, Oliveira T. Adoption of video consultations during the COVID-19 pandemic. Internet Interv 2023 Mar;31:100602 [FREE Full text] [doi: 10.1016/j.invent.2023.100602] [Medline: 36694630]
40. Aydin G, Kumru S. Paving the way for increased e-health record use: elaborating intentions of Gen-Z. Health Syst (Basingstoke) 2023 Oct 03;12(3):281-298 [FREE Full text] [doi: 10.1080/20476965.2022.2129471] [Medline: 37860594]

XSL•FO
RenderX

41. Harborth D, Pape S. German translation of the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) questionnaire. SSRN Journal 2018:1-12. [doi: 10.2139/ssrn.3147708]
42. Bais F, Schouten B, Toepoel V. Investigating response patterns across surveys: do respondents show consistency in undesirable answer behaviour over multiple surveys? Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique 2020 Aug 26;147-148(1-2):150-168. [doi: 10.1177/0759106320939891]
43. Arevalo M, Brownstein NC, Whiting J, Meade CD, Gwede CK, Vadaparampil ST, et al. Strategies and lessons learned during cleaning of data from research panel participants: cross-sectional web-based health behavior survey study. JMIR Form Res 2022 Jun 23;6(6):e35797 [FREE Full text] [doi: 10.2196/35797] [Medline: 35737436]
44. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. Res Sci Educ 2017 Jun 7;48(6):1273-1296. [doi: 10.1007/s11165-016-9602-2]
45. Andrade C. The value and statistical significance: misunderstandings, explanations, challenges, and alternatives. Indian J Psychol Med 2019;41(3):210-215 [FREE Full text] [doi: 10.4103/IJPSYM.IJPSYM_193_19] [Medline: 31142921]
46. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning: With applications in R. New York, NY: Springer; 2017.
47. Turner P. Critical values for the Durbin-Watson test in large samples. Applied Economics Letters 2019 Nov 12;27(18):1495-1499. [doi: 10.1080/13504851.2019.1691711]
48. Hair J, Black W, Babin B, Anderson R. Multivariate Data Analysis: A Global Perspective. London, UK: Pearson Education; 2010.
49. Amin R, Hossain MA, Uddin MM, Jony MTI, Kim M. Stimuli influencing engagement, satisfaction, and intention to use telemedicine services: an integrative model. Healthcare (Basel) 2022 Jul 18;10(7):1327 [FREE Full text] [doi: 10.3390/healthcare10071327] [Medline: 35885854]
50. Alexandra S, Handayani PW, Azzahro F. Indonesian hospital telemedicine acceptance model: the influence of user behavior and technological dimensions. Heliyon 2021 Dec;7(12):e08599 [FREE Full text] [doi: 10.1016/j.heliyon.2021.e08599] [Medline: 34977414]
51. Brooks NP. Telemedicine is here. World Neurosurg 2016 Nov;95:603-604. [doi: 10.1016/j.wneu.2016.02.113] [Medline: 26960280]
52. Zobair KM, Sanzogni L, Houghton L, Sandhu K, Islam MJ. Health seekers' acceptance and adoption determinants of telemedicine in emerging economies. AJIS 2021 Dec 26;25:3071. [doi: 10.3127/ajis.v25i0.3071]
53. Abd-Alrazaq A, Alalwan AA, McMillan B, Bewick BM, Househ M, Al-Zyadat A. Patients' adoption of electronic personal health records in England: secondary data analysis. J Med Internet Res 2020 Oct 07;22(10):e17499 [FREE Full text] [doi: 10.2196/17499] [Medline: 33026353]
54. Agrawal L, Ndabu T, Mulgund P, Sharman R. Factors affecting the extent of patients' electronic medical record use: an empirical study focusing on system and patient characteristics. J Med Internet Res 2021 Oct 28;23(10):e30637 [FREE Full text] [doi: 10.2196/30637] [Medline: 34709181]
55. Alsyouf A, Lutfi A, Alsubahi N, Alhazmi FN, Al-Mugheed K, Anshasi RJ, et al. The Use of a Technology Acceptance Model (TAM) to predict patients' usage of a personal health record system: the role of security, privacy, and usability. Int J Environ Res Public Health 2023 Jan 11;20(2):1347 [FREE Full text] [doi: 10.3390/ijerph20021347] [Medline: 36674105]
56. Bile Hassan I, Murad MAA, El-Shekeil I, Liu J. Extending the UTAUT2 model with a privacy calculus model to enhance the adoption of a health information application in Malaysia. Informatics 2022 Mar 28;9(2):31. [doi: 10.3390/informatics9020031]
57. Liu J, Gong X, Weal M, Dai W, Hou S, Ma J. Attitudes and associated factors of patients' adoption of patient accessible electronic health records in China - A mixed methods study. Digit Health 2023;9:20552076231174101 [FREE Full text] [doi: 10.1177/20552076231174101] [Medline: 37188077]
58. Niemöller S, Hübner U, Egbert N, Babitsch B. How to access personal health records? measuring the intention to use and the perceived usefulness of two different technologies: a randomised controlled study. Stud Health Technol Inform 2019 Sep 03;267:197-204. [doi: 10.3233/SHTI190827] [Medline: 31483273]
59. Tavares J, Oliveira T. New integrated model approach to understand the factors that drive electronic health record portal adoption: cross-sectional national survey. J Med Internet Res 2018 Nov 19;20(11):e11032 [FREE Full text] [doi: 10.2196/11032] [Medline: 30455169]
60. Keung K, Lee C, Ng K, Leung S, Choy K. An empirical study on patients' acceptance and resistance towards electronic health record sharing system: a case study of Hong Kong. International Journal of Knowledge and Systems Science 2018;9(2):1-27. [doi: 10.4018/ijkss.2018040101]
61. Alharbey R. Integrative adoption model for personal health records: A structural equation modeling approach. Int. j. adv. appl. sci 2022 Jan;9(1):117-127. [doi: 10.21833/ijaas.2022.01.014]
62. Hoogenbosch B, Postma J, de Man-van Ginkel JM, Tiemessen NA, van Delden JJ, van Os-Medendorp H. Use and the users of a patient portal: cross-sectional study. J Med Internet Res 2018 Sep 17;20(9):e262 [FREE Full text] [doi: 10.2196/jmir.9418] [Medline: 30224334]
63. Razmak J, Bélanger C. Using the technology acceptance model to predict patient attitude toward personal health records in regional communities. ITP 2018 Apr 03;31(2):306-326. [doi: 10.1108/ITP-07-2016-0160]

XSL•FO
RenderX

64. Yasnoff WA, Shortliffe EH. Lessons learned from a health record bank start-up. Methods Inf Med 2014;53(2):66-72. [doi: 10.3414/ME13-02-0030] [Medline: 24477917]
65. Saeed SA, Masters RM. Disparities in health care and the digital divide. Curr Psychiatry Rep 2021 Jul 23;23(9):61 [FREE Full text] [doi: 10.1007/s11920-021-01274-4] [Medline: 34297202]
66. Chang JE, Lai AY, Gupta A, Nguyen AM, Berry CA, Shelley DR. Rapid transition to telehealth and the digital divide: implications for primary care access and equity in a post-COVID era. Milbank Q 2021 Jun;99(2):340-368 [FREE Full text] [doi: 10.1111/1468-0009.12509] [Medline: 34075622]
67. Kane CK, Gillis K. The use of telemedicine by physicians: still the exception rather than the rule. Health Aff (Millwood) 2018 Dec;37(12):1923-1930. [doi: 10.1377/hlthaff.2018.05077] [Medline: 30633670]
68. McGrail KM, Ahuja MA, Leaver CA. Virtual visits and patient-centered care: results of a patient survey and observational study. J Med Internet Res 2017 May 26;19(5):e177 [FREE Full text] [doi: 10.2196/jmir.7374] [Medline: 28550006]
69. Kuruoglu E, Guldal D, Mevsim V, Gunvar T. Which family physician should I choose? The analytic hierarchy process approach for ranking of criteria in the selection of a family physician. BMC Med Inform Decis Mak 2015 Aug 05;15:63 [FREE Full text] [doi: 10.1186/s12911-015-0183-1] [Medline: 26242399]
70. Khatami F, Shariati M, Khedmat L, Bahmani M. Patients' preferences in selecting family physician in primary health centers: a qualitative-quantitative approach. BMC Fam Pract 2020 Jun 11;21(1):107 [FREE Full text] [doi: 10.1186/s12875-020-01181-2] [Medline: 32527224]
71. KBV - PraxisBarometer 2023: Praxen nutzen deutlich öfter digitale Anwendungen. Kassenärztliche Bundesvereinigung (KBV). URL: https://www.kbv.de/html/1150_67527.php [accessed 2024-02-14]
72. Weik L, Fehring L, Mortsiefer A, Meister S. Big 5 personality traits and individual- and practice-related characteristics as influencing factors of digital maturity in general practices: quantitative web-based survey study. J Med Internet Res 2024 Jan 22;26:e52085 [FREE Full text] [doi: 10.2196/52085] [Medline: 38252468]
73. Faries MD. Why we don't "just do it": understanding the intention-behavior gap in lifestyle medicine. Am J Lifestyle Med 2016;10(5):322-329 [FREE Full text] [doi: 10.1177/1559827616638017] [Medline: 30202289]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
PRISMA-RR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Rapid Reviews
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by A Mavragani; submitted 27.03.24; peer-reviewed by DG Goldberg, P Rasche, T Hüer, AL Brecher; comments to author 01.10.24; revised version received 13.10.24; accepted 12.01.25; published 05.03.25.

Please cite as:
Beerbaum J, Robens S, Fehring L, Mortsiefer A, Meister S
Patient Adoption of Digital Use Cases in Family Medicine and a Nuanced Implementation Approach for Family Doctors: Quantitative Web-Based Survey Study
JMIR Form Res 2025;9:e58867
URL: https://formative.jmir.org/2025/1/e58867
doi:10.2196/58867
PMID:40053731

©Julian Beerbaum, Sibylle Robens, Leonard Fehring, Achim Mortsiefer, Sven Meister. Originally published in JMIR Formative Research (https://formative.jmir.org), 05.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Feasibility and Links Between Emotions, Physical States, and Eating Behavior in Patients After Metabolic Bariatric Surgery: Experience Sampling Study

Ellen A M Kuipers[1,2], MD; Josien G Timmerman[3], PhD; Marc J van Det[1], MD, PhD; Miriam M R Vollenbroek-Hutten[2,4], PhD

[1]Department of Surgery, Hospital Group Twente, Almelo, The Netherlands

[2]Biomedical Signals and Systems, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands

[3]ZGT Academy, Hospital Group Twente, Almelo, The Netherlands

[4]Board of Directors, Medisch Spectrum Twente, Enschede, The Netherlands

Corresponding Author:
Ellen A M Kuipers, MD
Department of Surgery
Hospital Group Twente
Zilvermeeuw 1
Almelo, 7609 PP
The Netherlands
Phone: 31 620024496
Email: el.kuipers@zgt.nl

Abstract

Background: Lifestyle modification is essential to achieve and maintain successful outcomes after metabolic bariatric surgery (MBS). Emotions, physical states, and contextual factors are considered important determinants of maladaptive eating behavior, emphasizing their significance in understanding and addressing weight management. In this context, experience sampling methodology (ESM) offers promise for measuring lifestyle and behavior in the patient's natural environment. Nevertheless, there is limited research on its feasibility and association among emotions and problematic eating behavior within the population after MBS.

Objective: This study aimed to examine the feasibility of ESM in the population after MBS regarding emotions, physical states, contextual factors, and problematic eating behavior, and to explore the temporal association among these variables.

Methods: An experience sampling study was conducted in which participants rated their current affect (positive and negative), physical states (disgust, boredom, fatigue, and hunger), contextual factors (where, with whom, and doing what), and problematic eating behavior (ie, grazing, dietary relapse, craving, and binge eating) via smartphone-based ESM questionnaires at 6 semirandom times daily for 14 consecutive days. Feasibility was operationalized as the study's participation rate and completion rate, compliance in answering ESM questionnaires, and response rates per day. At the end of the study period, patients reflected on the feasibility of ESM in semistructured interviews. Generalized estimation equations were conducted to examine the temporal association between emotions, physical states, contextual factors, and problematic eating behavior.

Results: In total, 25 out of 242 participants consented to participate, resulting in a study participation rate of 10.3%. The completion rate was 83%. Overall compliance was 57.4% (1072/1868), varying from 13% (11/84) to 89% (75/84) per participant. Total response rates per day decreased from 65% (90/138) to 52% (67/130) over the 14-day study period. According to the interviews, ESM was considered feasible and of added value. Temporal associations were found for hunger and craving (odds ratio 1.04, 95% CI 1.00-1.07; *P*=.03), and for positive affect and grazing (odds ratio 1.61, 95% CI 1.03-2.51; *P*=.04).

Conclusions: In this exploratory study, patients after MBS were not amenable to participate. Only a small number of patients were willing to participate. However, those who participated found it feasible and expressed satisfaction with it. Temporal associations were identified between hunger and craving, as well as between positive affect and grazing. However, no clear patterns were observed among emotions, physical states, context, and problematic eating behaviors.

(JMIR Form Res 2025;9:e60486) doi:10.2196/60486

XSL•FO
RenderX

KEYWORDS

feasibility; experience sampling methodology; metabolic bariatric surgery; eating behavior; positive and negative affect; physical states; contextual factors; mobile phone

Introduction

Metabolic bariatric surgery (MBS) is the most effective long-term strategy in the management of severe obesity and its related comorbidities [1]. Lifestyle modification, including nutritional compliance and physical activity, is essential to achieve and maintain optimal weight loss results and to improve obesity-related comorbidities after MBS [2-4]. Poor dietary adherence is significantly associated with recurrent weight gain after MBS [1,5,6]. Furthermore, maladaptive eating behaviors may cause nutritional noncompliance with the inability to remain consequently adherent to an intended diet (ie, dietary relapse) and thereby negatively impact postsurgical weight loss outcomes [7-10]. Most theoretical models of eating disorders posit that negative emotions and numerous physical states (ie, hunger and fatigue) contribute to the development and maintenance of maladaptive eating behaviors [11-17]. However, little is known whether these theoretical constructs apply to individuals who have undergone MBS. To our best knowledge, only 1 study has explored the relationship between negative affect and loss of control (LOC) eating in patients who underwent MBS, finding that higher momentary negative affect predicted more severe LOC eating both before and after MBS [18]. Since this study focused on 1 aspect of disordered eating behavior, it is important to enhance our understanding of how emotions, physical states, and other maladaptive eating behaviors are interconnected in this specific population. This is particularly important given the high prevalence of problematic eating behaviors in the population with obesity, including grazing, food cravings, and binge eating [7,19-22]. Grazing, defined as unplanned and repetitive eating of small amounts of food between meals and snacks [23], is reported to occur in up to 26% of patients seeking MBS [7]. Postoperatively, the prevalence is even higher, with rates reported as high as 47% [8]. Food cravings, characterized by an intense desire to consume specific foods, are experienced by nearly all patients after MBS, with only about 10% reporting no cravings [24]. Binge eating involves the consumption of excessive amounts of food within a discrete period of time, accompanied by an LOC feeling during the binge eating episode [9]. Binge eating, which is present in 24% of patients before MBS [25], is strongly associated with obesity and negatively impacts weight outcomes after MBS [8].

Since lifestyle and lifestyle modification after MBS takes place primarily outside of medical visits and clinical settings, there is an urgent need to have ecologically valid methods that capture daily life to better understand and change behaviors. Experience sampling methodology (ESM), also known as ecological momentary assessment, addresses this need by collecting real-time data on individuals' behavior, thoughts, emotions, and environmental contexts as they go about their daily lives [26]. Unlike retrospective questionnaires, ESM prompts participants at random or scheduled times throughout the day to record their current state or activities. This approach provides a more accurate representation of behavior within the individual's natural environment, capturing their typical routines in daily life [27], while traditional self-reports may miss crucial details about how and why a participant's experiences or behaviors change over time and context.

In the past years, an increasing number of experience sampling studies have been conducted in various health domains, representing a wide range of diseases and populations [17,28-37]. However, the use of ESM to evaluate (disordered) eating behaviors following MBS remains limited [18,38,39]. Previous research has focused on specific aspects, such as 1 type of problematic eating behavior [18], eating practices that do not align with postoperative dietary or eating behavior recommendations (eg, drinking liquids with meals and not starting meals with protein) [39], or elements of eating regulation (eg, dietary restraint and eating in the absence of hunger) [38]. Our goal is to conduct a more comprehensive assessment by examining a broader range of problematic eating behaviors and their influencing factors—an important objective given the high prevalence of disordered eating behavior in the population after MBS [20], as outlined earlier in this section.

The feasibility of ESM has been previously evaluated in various populations, including its use to monitor dietary intake and physical activity among Dutch vocational education students [28], to study depressive symptoms among caregivers of patients with cancer [31], and to examine daily life experiences and activities in patients with acute coronary syndrome [33]. These studies show varying results regarding compliance, participation rates, and completion rates, indicating that feasibility for ESM differs between populations. However, the feasibility of ESM has scarcely been studied in the population that has undergone or is undergoing MBS. To our knowledge, only 1 experience sampling feasibility study was conducted in adults undergoing MBS, focusing on the use of a smartphone app for lifestyle education before surgery [40]. This study reported high satisfaction and perceived helpfulness, but the study was carried out in patients before MBS. After MBS, patients are dealing with physical and mental changes, along with the need to modify their behavior. This may impact the feasibility of ESM, as patients are likely to prioritize managing these challenges over using an experience sampling app for monitoring. Although some studies have used ESM to evaluate (disordered) eating behaviors after MBS [18,38,39,41], they did not specifically focus on evaluating feasibility. While these studies reported compliance, they did not assess other feasibility metrics. Notably, participants in these studies received monetary compensation [38,39,41], which may have positively influenced participation and compliance, potentially affecting feasibility outcomes [42]. This raises concerns about the generalizability of these findings to the real-world clinical setting.

Therefore, the primary aim of this exploratory study is to establish the feasibility of ESM among patients shortly after MBS, focusing on maladaptive eating behaviors, emotions, physical states, and context. The secondary aim is to explore

XSL•FO
RenderX

associations between emotions, physical states, context, and maladaptive eating behaviors in the population after MBS using experience sampling measurements.

Methods

Study Design and Participants

A prospective observational cohort study was performed at the obesity center of Hospital Group Twente (ZGT) Almelo/Hengelo, the Netherlands. All patients who underwent MBS at ZGT during the inclusion period (from February 2022 to July 2022) were considered for participation in this study. We included patients who had recently undergone MBS, as lifestyle modification mostly occurs after surgery, making this period particularly relevant for monitoring to enable timely intervention. The inclusion and exclusion criteria are detailed in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Internet access at home
- Good comprehension of the Dutch language

Exclusion criteria

- No possession of a smartphone that supports the experience sampling methodology app
- Participation in an intervention program or behavioral treatment that is not included in the regular care pathway after metabolic bariatric surgery
- People who worked on night shifts (as it is impossible to complete the experience sampling questionnaires during the day)

Study Procedure

Recruitment

Patients were initially recruited through flyers sent by post and by email 5 weeks after surgery. Due to a minimal response, the recruitment strategy was expanded; patients were approached following a postoperative group meeting or patients were contacted by telephone. Recruitment proceeded as follows:

1. Via flyer by post or email: Participant information was provided to the patients by scanning the QR code on the flyer or clicking on the link. If the patient was interested in participating after reading the information, the patient could send an email to the first author (EAMK). Later, individuals were contacted via telephone by the first author (EAMK) or a student (Charlotte van den Berg) to provide verbal information about the study.
2. During a postoperative group meeting: A student (Charlotte van den Berg) provided verbal and written information about the study. If interested, patients shared their contact details for inclusion. Furthermore, a subset of these patients was contacted by phone to encourage participation.
3. By telephone: Postoperative patients were contacted by telephone by a student (Charlotte van den Berg) or by the first author (EAMK) to inform patients about the study and motivate them to participate.

After inclusion, participants received verbal instructions on installing and using the ESM app. A user manual for the ESM app was sent via email. The participants used their own email addresses to create an account in the ESM app Ethica Data. After entering a unique registration code to join the study, participants were briefed on the study, and at this point, participants signed informed consent in the Ethica Data app. Patients who consented were asked if they would like to participate in an interview at the end of the study period to discuss the feasibility of the ESM app.

Questionnaires and Interviews

Participants were instructed to answer ESM questionnaires for 14 consecutive days, to capture both weekdays and weekends. The app was programmed to send auditory and visual notifications at 6 semirandom times throughout the day (1 prompt every 135 min, between 8:00 AM and 9:30 PM), signaling the availability of a questionnaire (signal contingent). At each prompt, participants answered questions regarding current positive and negative affect, physical states, context, and problematic eating behavior. A delayed response triggered a 1-time reminder after 5 minutes, and each questionnaire expired after 15 minutes to minimize recall bias. Participants were asked to carry their smartphones throughout the day and continue their daily lives as usual. Participants who consented to the interview were contacted by phone to discuss the feasibility of the ESM app.

Data Collection

Baseline Measures

Demographics and surgical data were obtained from electronic patient records. Collected patient characteristics included age; gender; type of surgery; time since surgery; and obesity-related comorbidities, such as hypertension, diabetes mellitus, dyslipidemia, osteoarthrosis, gastroesophageal reflux disease, and obstructive sleep apnea syndrome. At the start of the study, each participant was asked to enter their current weight and height into the Ethica Data app.

Feasibility

Feasibility was operationalized as participation rate, completion rate, compliance, and response rate per day.

- The participation rate was calculated as the percentage of approached patients willing to participate in the study.
- The completion rate represented the percentage of participants who filled out at least 1 ESM questionnaire on the 14th day of the study period.

XSL•FO
RenderX

- Compliance with experience sampling was calculated as the percentage of answered ESM questionnaires out of all issued ESM questionnaires.
- The response rate per day was the percentage of ESM questionnaires answered per day, calculated for each day of the study period.

At the end of the 2-week period, participants were invited to participate in a semistructured interview to share their experiences. The interviews were conducted by the first author (EAMK) and a master's student in health science (Charlotte van den Berg) via telephone in November and December 2022. The interviews, guided by a semistructured interview guide (Multimedia Appendix 1), lasted between 15 and 30 minutes. This guide covered the feasibility themes to be discussed, including user-friendliness, technical performance, and content, along with optional questions. With participants' permission, the interviews were audio-recorded, transcribed verbatim, and anonymized.

ESM Measures

ESM data were collected using the Ethica Data app. All ESM questions were provided in the patients' native language (Dutch). ESM data included the following items:

1. Emotion: Participants rated their feelings as "angry or annoyed," "anxious or scared," "relaxed or calm," "cheerful or happy," "sad or gloomy," and "tensed or stressed" [43,44]. These items reflect both positive and negative affect and were scored on a visual analog scale from 0 (not at all or very bad) to 10 (extremely or very good).
2. Physical states: Disgust, boredom, fatigue, and hunger were rated. The first ESM questionnaire in the morning also asked participants to rate their sleep quality. Physical state items were scored on a visual analog scale from 0 to 10.
3. Context and activities: Questions about the context (where, with whom) and activities (doing what) had multiple-choice response options.
4. Problematic eating behavior: Participants answered 4 questions about problematic eating behavior. The questions and answer options are presented in Table 1.

Table 1. Experience sampling questions about problematic eating behavior and their answer options.

Behavior	Question	Answer option
Craving	"Have you had a craving for a specific food in the past 30 minutes? If yes, how strong was this desire?"	VAS[a]: 0-10
Dietary relapse	"In the past 60 minutes, did you eat and/or drink anything that is likely to negatively impact your weight?"	Yes or no
Binge eating	"Have you eaten unusually large amounts of unhealthy food in the past 60 minutes?"	Yes or no
Grazing	"In the past 60 minutes, have you eaten unplanned and repetitive small amounts of food outside of planned meals and snacks?"	Yes or no

[a]VAS: visual analog scale.

Statistical Analysis

Statistical analysis was performed using SPSS (version 24; IBM Corp). Patient characteristics were expressed as means with SDs for parametric continuous data or frequencies and percentages for categorical data. For nonparametric continuous data, median and IQR were reported. Feasibility metrics were expressed as percentages.

To investigate the relationship between emotions, physical states, context, and problematic eating behaviors, generalized estimation equations (GEEs) were used. GEEs are an extension of generalized linear models designed to account for the correlation between repeated measures within participants, as data collected through ESM involve multiple observations per individual [45]. This makes GEEs the appropriate choice for our analysis, as they allow us to model repeated measures data while adjusting for within-subject correlations without assuming that these correlations are the same across all participants.

We used a negative binomial distribution with a binary logistic link function, as our dependent variables (eg, dietary relapse and grazing) were dichotomous (yes or no) and did not follow a normal distribution. The negative binomial distribution is particularly suitable for overdispersed count data, which are often encountered in studies involving behavioral data. An exchangeable correlation structure was selected, assuming a constant correlation between repeated measures within individuals over time. While centering is customary in ESM studies, in this study, not centering the variables aligns better with the interpretability of the results. Retaining the raw values ensures that the intercept is directly relatable to real-world states (eg, no negative or positive emotion), which facilitates reader comprehension without altering the estimated effects.

Momentary assessments (level 1) were nested within participants (level 2). Momentary levels of positive affect were calculated by averaging the scores of the "relaxed or calm" and "cheerful or happy" items, while momentary levels of negative affect were calculated by averaging the "angry or annoyed", "anxious or scared", "sad or gloomy" and "tensed or stressed" items. Contextual factors were recoded as dichotomous variables such as "being at home' (yes or no), "being alone" (yes or no), and "relaxing" (yes or no), where "relaxing" encompassed resting or doing nothing. Due to nonnormality, craving was dichotomized (yes or no).

To account for potential temporal dependencies, we lagged the independent variables (emotions, physical states, and contextual factors) by 1 time point, meaning that the predictors at time t–1 were used to examine their influence on problematic eating behavior at time t. The first assessment of each day was

XSL•FO
RenderX

excluded from the analysis to avoid carry-over effects from the previous day.

A minimum response rate of 33% for ESM data was applied, which is the generally accepted threshold for ESM studies [46]. Univariate GEE analyses were conducted for each predictor variable, and given the exploratory nature of the study, we did not adjust for multiple comparisons. The significance level was set at $P<.05$.

Thematic Analysis

Using the software program ATLAS.ti (version 22; Lumivero), the interview transcripts were reviewed and analyzed thematically by the first author. Initially, each transcript was read multiple times without conducting a thematic analysis. Next, comments related to feasibility issues were extracted from the transcripts without interpretation or categorization. These feasibility issues were then classified deductively according to the 7 categories of the eHealth Usability Benchmarking Instrument (HUBBI), namely, basic system performance, task-technology fit, interface design, navigation and structure, information and terminology, guidance and support, and satisfaction [47]. Finally, transcripts were reread to ensure that all relevant codes were identified.

Ethical Considerations

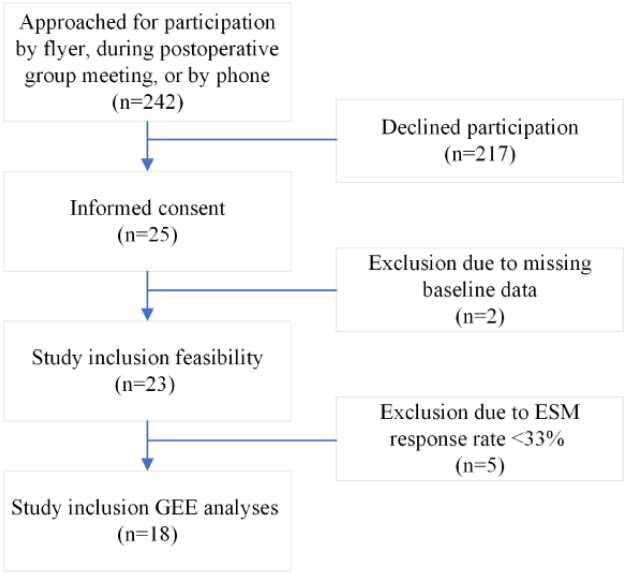
The study protocol was reviewed and approved by the local medical ethics committee (local registration 2020-48), and all patients provided informed consent. Data collected via the Ethica Data app were securely stored on Ethica Data's servers. Ethica Data employs 2 separate databases located on 2 different servers. Participant registration data (ie, name and email) are stored on one server, while study data (ie, questionnaire responses) are stored on another. Study data are stored in a coded format and cannot be traced back to individual participants. The study data stored on Ethica Data's server were accessible only to researchers with a password-protected account and could be imported by the researchers to ZGT's server. The dataset contains only participant numbers, and the combination of participant numbers and email addresses was securely stored in a password-protected file within the research department. Only authorized persons had access to these data. Audio recordings of interviews were used to generate deidentified transcripts, after which the recordings were immediately destroyed. Both the audio recordings and transcripts were securely stored on a protected hard drive. Participants with a minimum response rate of 80% on the ESM questionnaires had the chance to win a €50 voucher (US $52.32), which was randomly awarded to 1 participant. No additional monetary compensation was provided.

Results

Participants

From February to July 2022, a total of 242 patients were informed about the study. Out of the 242 patients who were approached to participate in the ESM study, 25 patients consented to participate, resulting in a participation rate of 10.3%. Figure 1 presents the study flowchart.

Figure 1. Study flowchart of patients approached for participation between February and July 2022. ESM: experience sampling methodology; GEE: generalized estimation equation.



Two participants were excluded from analyses due to missing baseline data. The remaining 23 participants were included to answer the primary aim of the study. Although 2 participants withdrew due to a high perceived burden, they were included in the feasibility analysis. For the secondary aim of this study, an additional 5 participants were excluded due to a response rate of less than 33% for all ESM questionnaires. Thus, a total of 18 participants were included in this analysis.

Demographics

The 23 participants had a mean age of 47 (SD 11) years, and 17 (74%) were female. The median time since surgery was 36 (IQR 31-88) days. At inclusion, median BMI was 36 (IQR 33-39) kg/m^2. The most performed bariatric procedure was one-anastomosis gastric bypass (n=16, 70%). The most prevalent obesity-related comorbidities were dyslipidemia (n=21, 91%)

XSL•FO
RenderX

and osteoarthrosis (n=13, 57%). Descriptive statistics of the participants are shown in Table 2.

Table 2. Participant characteristics for the total study population, as well as for those with a response rate ≥33% and <33% on the experience sampling questionnaires.

Participant characteristic	Total (n=23)	Response rate ≥33% (n=18)	Response rate <33% (n=5)
Age (years), mean (SD)	47 (11)	49 (12)	42 (9)
Sex, n (%)			
Female	17 (74)	13 (72)	4 (80)
Male	6 (26)	5 (28)	1 (20)
BMI (kg/m^2), median (IQR)	36 (33-39)	36 (33-38)	39 (32-42)
Days between surgery and study start, median (IQR)	36 (31-88)	34 (28-79)	88 (60-103)
Comorbidities, n (%)			
T2DM[a]	3 (13)	2 (11)	1 (20)
Hypertension	10 (44)	9 (50)	1 (20)
OSAS[b]	8 (35)	7 (39)	1 (20)
Osteoarthrosis	13 (57)	10 (56)	3 (60)
GERD[c]	10 (44)	8 (44)	2 (40)
Dyslipidemia	21 (91)	17 (94)	4 (80)
Type of surgery, n (%)			
SG[d]	1 (4)	0 (0)	1 (20)
RYGB[e]	3 (13)	3 (17)	0 (0)
OAGB[f]	16 (70)	13 (72)	3 (60)
OAGB+correction HH[g]	2 (9)	1 (6)	1 (20)
OAGB+cholecystectomy	1 (4)	1 (6)	0 (0)

[a]T2DM: type 2 diabetes mellitus.
[b]OSAS: obstructive sleep apnea syndrome.
[c]GERD: gastroesophageal reflux disease.
[d]SG: sleeve gastrectomy.
[e]RYGB: Roux-en-Y gastric bypass.
[f]OAGB: one-anastomosis gastric bypass.
[g]HH: Hiatal hernia.

Feasibility

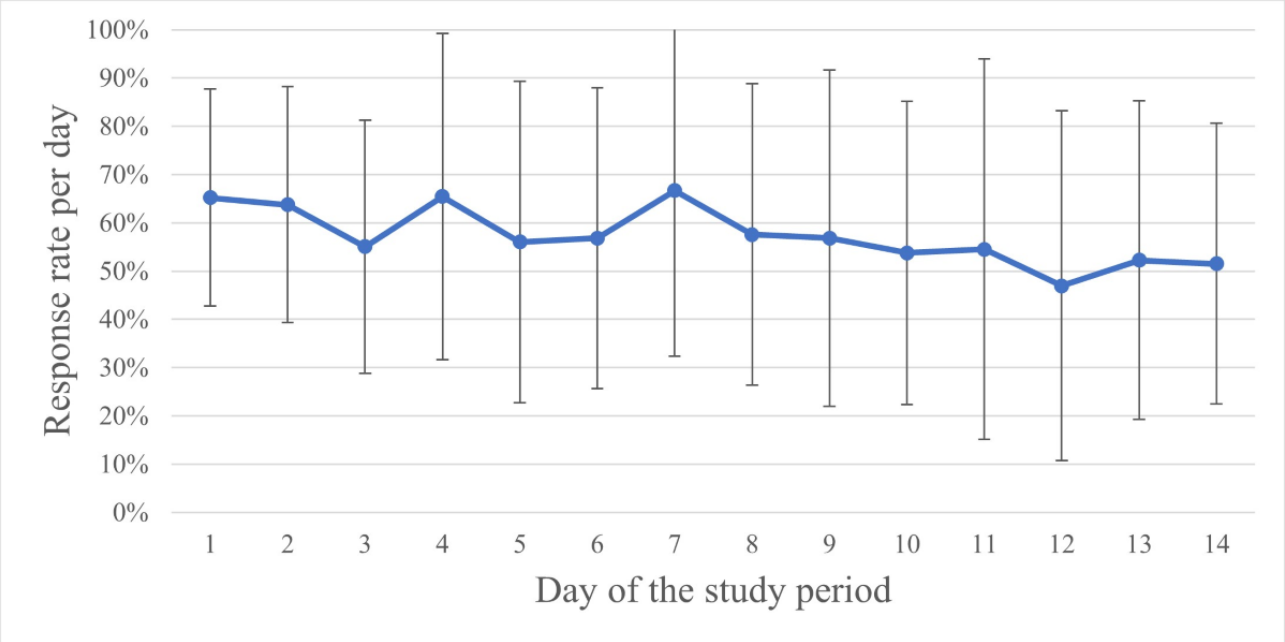
Overview

The compliance per participant ranged from 13% (11/84 questionnaires completed) to 89% (75/84 questionnaires completed). Overall, participants responded to 1072 out of 1868 prompts, resulting in an overall compliance of 57.4%. The compliance in answering experience sampling questionnaires per participant is presented in Multimedia Appendix 2.

The mean number of answered ESM questionnaires was 47 (SD 21). Total response rates were calculated from day 1 to day 14, as illustrated in Figure 2. Over time, the total response rate declined by 14%. The study completion rate was 83% (19/23).

XSL•FO
RenderX

Figure 2. Mean response rates per day and SD for the total study population (n=23) over 14 days of study participation.



Semistructured Interviews

Overview

In total, 10 patients participated in the semistructured interviews, of whom 7 (70%) were female. The mean age was 46 (range 28-64) years, and the mean BMI was 38 (SD 6) kg/m^2. The median time between surgery and participation in the ESM study was 34 (IQR 28-87) days.

The categories that were indicated in the interviews, based on the HUBBI, are basic system performance, task-technology fit, information and terminology, guidance and support, and satisfaction [47].

Basis System Performance

The participants reported that the app was easy to use and straightforward to install on their mobile phones, with no technical issues during installation. When asked about technical performance in general, 1 participant reported repeated technical problems that prevented her from completing several questionnaires. In addition, she reported not receiving auditory notifications and sometimes inconsistent visual notifications.

Task-Technology Fit

Most participants reported they could not complete all questionnaires within the 15-minute time frame due to work or travel commitments. Participants perceived the number of questions per questionnaire as appropriate and found them easy to answer since the questionnaires were nearly identical. The daily number of questionnaires per day was also considered manageable, with 1 participant stating:

> *Not too much, not too little [the number of questionnaires].* [Participant 6]

Another participant felt that 6 questionnaires per day was the maximum, saying:

> *Otherwise, you spend so much time working on it [the questionnaires], I don't think that's healthy.* [Participant 7]

Furthermore, 3 participants shared their views on the 2-week survey duration, and 2 of them indicated that a longer duration would have been acceptable, with 1 noting:

> *If it's not going well immediately [after MBS], it [filling in the questionnaires] can provide insights, because it's quite astonishing how you change after bariatric surgery.* [Participant 6]

In addition, 7 participants indicated that using the app encouraged them to reflect on their behavior, providing insights and increasing awareness. One participant found the questions negatively framed, which led to his reluctance to use the app, ultimately causing him to withdraw from the study. Another participant felt she was doing very well at that moment. Therefore, she experienced no additional support by using the app.

Information and Terminology

The participants found the questions clear, though 5 participants noted that the questions were repetitive. One participant commented:

> *Maybe I expected different questions, as not much has changed in such a short time [time since previous questionnaire].* [Participant 10]

Another participant suggested occasionally changing the order of the questions:

> *This way, you have to think carefully about which question to answer.* [Participant 2]

Furthermore, 2 participants would have preferred open questions as well:

> *Imagine you experience a certain emotion or that you ate something 'wrong', what was that about? If you*

XSL•FO
RenderX

> *could explain a little more, it would provide greater depth.* [Participant 10]

As mentioned above, 1 participant indicated that the questions were negatively framed, feeling that the negative aspects overshadowed questions about positive emotions.

Guidance and Support

The participants were not bothered about lacking insight into their own answers, understanding it was for scientific research. As one stated:

> *It is especially important for you as researchers, I don't necessarily have to see it myself [the answers].* [Participant 3]

However, 5 participants felt that having access to their answers could be valuable. One of them mentioned:

> *It might be interesting to see if there were any changes or patterns.* [Participant 10]

Another remarked:

> *On one hand, I see the benefit, on the other, it could raise many questions. Ultimately, if things aren't going well, you're the one who must seek help.* [Participant 6]

Overall, participants were positive about using the app in bariatric care, as it could help individuals become more aware of their actions:

> *I can imagine that some people have had surgery and then feel a little lost. Of course, you have group meetings with the dietician, but the application allows for personal reflection in a quieter moment.* [Participant 5]

In total, 4 participants indicated that using the app before MBS could be interesting:

> *Are people aware of their emotions and their eating behavior before surgery?* [Participant 2]

Participants expressed mixed opinions on how an app should be integrated into follow-up care after MBS. One suggested that questions about changes in eating patterns, the creeping in of old habits, physical activity, and mental well-being may be relevant 1 or 2 years after MBS. Another participant expressed interest in questions about nutrition and related uncertainties. One participant preferred an app where she could ask questions directly to a health care professional. Participants recognized the need to tailor the app to their specific stage after surgery. While most participants preferred using the app in the short term (within a few months) after MBS, 1 participant preferred using the app on a longer-term (at least 6 months) after MBS. He indicated that, in the short term after surgery, he would have benefited more from an app that provided meal or snack suggestions, and calculated protein intake based on those suggestions.

Satisfaction

In general, participants were positive about the app, finding it helpful in the follow-up after MBS. Most participants experienced personal benefits from using the app, noting that it encouraged more conscious eating by prompting questions at random times:

> *What did I actually eat or drink and how do I feel about that?* [Participant 2]

One participant experienced a personal disadvantage by using the app:

> *I remember thinking that if I continue this [the questionnaires], it will even make me depressed. I don't think that's the intention.* [Participant 9]

For improvement, participants suggested extending the time a questionnaire remains open and adding the question:

> *What do all these questions do to you?* [Participant 2]

If the answer is negative, it was recommended that the app alert the health care professional:

> *I think you should contact this person, because he can't figure it out on his own.* [Participant 2]

Association Between Positive and Negative Affect, Contextual Factors, and Problematic Eating Behavior

Frequencies and Percentages of Problematic Eating Behaviors

Frequencies and percentages of the different types of problematic eating behavior are illustrated in Multimedia Appendix 3.

- Dietary relapse: In total, 6 participants reported no dietary relapse over 14 days. The remaining participants reported a total of 45 dietary relapses, with individual counts ranging from 1 to 8.
- Grazing: Grazing occurred in total 13 times among 7 participants (range per participant: 1-3), while 11 participants reported no grazing during the study period.
- Binge eating: A total of 7 binge eating episodes were reported by 3 participants (range per participant: 2-3), while 15 participants reported no binge eating episodes during the study period.
- Craving: Craving was absent (score 0) in 65% (643/994) of all completed questionnaires. In addition, 3 participants consistently reported no craving throughout the study, while the remaining 15 participants experienced a total of 351 craving episodes (range per participant: 3-53).

GEE Analyses

Data from the ESM questionnaires were nested within 18 individuals. The number of observations per predictor or outcome variable ranged from 987 to 1005. After lagging each predictor and setting the first prompt of each day to missing, the number of observations available for the GEE analyses ranged from 594 to 605. Table 3 summarizes the GEE results, showing the temporal associations between lagged predictors (ie, positive and negative affect, physical states, and contextual factors) and problematic eating behaviors (craving, binge eating, dietary relapse, and grazing). Significant associations are described below, while nonsignificant trends are noted for exploratory purposes.

XSL•FO
RenderX

- Craving: Hunger was significantly associated with craving (OR 1.04, 95% CI 1.00-1.07; *P*=.03).
- Binge eating: The odds of binge eating increased by 1.61 times with each unit increase in negative affect (95% CI 0.77-3.34; *P*=.21) and decreased by 0.51 times for each unit increase in disgust (95% CI 0.20-1.28; *P*=.15). When at home, the odds of binge eating decreased by 0.27 times compared with not being at home (95% CI 0.03-2.67; *P*=.26).
- Dietary relapse: The odds of dietary relapse decreased by 0.40 times for being at home (95% CI 0.14-1.17; *P*=.09) and by 0.67 times for each unit increase in boredom (95% CI 0.40-1.14; *P*=.14).
- Grazing: A significant temporal association was found for positive affect and grazing, with the odds of grazing increasing by 1.61 times for each unit increase in positive affect (95% CI 1.03-2.51; *P*=.04). The odds of grazing decreased by 0.68 times with each unit increase in negative affect (95% CI 0.40-1.15; *P*=.15). When at home, the odds of grazing decreased by 0.16 times compared with not being at home (95% CI 0.02-1.07), with this result approaching statistical significance (*P*=.06).

XSL•FO
RenderX

Table 3. Generalized estimation equation analyses showing the temporal associations between lagged predictors (positive and negative affect, physical states, and contextual factors) and subsequent problematic eating behaviors (craving, binge eating, dietary relapse, and grazing) in patients shortly after metabolic bariatric surgery.

Eating behavior and predictor	Odds ratio (95% CI)	*P* value
Craving		
Negative affect	1.06 (0.79-1.42)	.69
Positive affect	1.02 (0.91-1.15)	.70
Boredom	1.03 (0.89-1.20)	.70
Disgust	1.01 (0.82-1.25)	.91
Fatigue	1.07 (0.96-1.18)	.21
Hunger	1.04 (1.00-1.07)	.03
Being alone	1.01 (0.87-1.17)	.92
Being at home	1.15 (0.93-1.42)	.21
Relaxing	0.93 (0.78-1.09)	.36
Binge eating		
Negative affect	1.61 (0.77-3.34)	.21
Positive affect	0.95 (0.53-1.69)	.86
Boredom	0.93 (0.57-1.50)	.76
Disgust	0.51 (0.20-1.28)	.15
Fatigue	1.10 (0.81-1.50)	.55
Hunger	1.00 (0.68-1.48)	≥.99
Being alone	0.80 (0.11-5.81)	.83
Being at home	0.27 (0.03-2.67)	.26
Relaxing	0.80 (0.17-3.91)	.79
Dietary relapse		
Negative affect	0.84 (0.46-1.52)	.56
Positive affect	1.15 (0.91-1.46)	.23
Boredom	0.67 (0.40-1.14)	.14
Disgust	0.89 (0.59-1.36)	.59
Fatigue	1.14 (0.90-1.45)	.28
Hunger	0.89 (0.72-1.10)	.28
Being alone	0.82 (0.35-1.91)	.65
Being at home	0.40 (0.14-1.17)	.09
Relaxing	0.97 (0.30-3.20)	.97
Grazing		
Negative affect	0.68 (0.40-1.15)	.15
Positive affect	1.61 (1.03-2.51)	.04
Boredom	0.81 (0.43-1.55)	.53
Disgust	0.88 (0.57-1.38)	.59
Fatigue	1.23 (1.00-1.51)	.06
Hunger	0.77 (0.55-1.09)	.15
Being alone	1.14 (0.33-3.94)	.83
Being at home	0.16 (0.02-1.07)	.06
Relaxing	0.90 (0.14-5.97)	.91

XSL•FO
RenderX

Discussion

Principal Results

This exploratory study describes the feasibility of ESM in patients after MBS, focusing on negative and positive affect, physical states, contextual factors, and problematic eating behaviors. The results indicate that patients shortly after MBS may be less inclined to participate in experience sampling studies. Nonetheless, actual usage was considered feasible, given the moderate compliance and high completion rate upon patient participation. In addition, results from the semistructured interviews indicated that the ESM app used in this study showed to be usable and of added value from the patient's perspective. This indicates that patients who chose to participate were motivated to complete the ESM questionnaires. Concerning the second research aim, temporal associations were found, but only between hunger and craving, and between positive affect and grazing.

Comparison With Previous Work

Our study faced a low participation rate of 10.3%. Several participants who were approached face-to-face and declined participation, mentioned the high expected burden, or indicated that study participation would likely not fit into their daily life or work schedule. Ratcliff et al [48] performed an experience sampling feasibility study to assess the adherence to postoperative diet and activity patterns in adolescents who underwent MBS. Their participation rate was 62%, which is remarkably higher than ours. However, they included participants at least 12 months after MBS, while we approached patients eligible for participation a few weeks after MBS. This timing difference may have influenced the willingness to participate, considering the overwhelming changes patients are dealing with in the first weeks after surgery. Furthermore, in the study by Ratcliff et al [48], participants completed only 3 daily phone diaries (a form of ESM) instead of 14 days of measurements. As we were interested in the feasibility of ESM, we opted for an intensive sampling strategy to gain insight into the daily response rates throughout the study period. This may have negatively affected our participation rate because of the high expected burden shortly following MBS. In addition, to obtain more realistic feasibility metrics, patients were not incentivized with monetary remuneration, except for 1 voucher randomly awarded among participants with a minimum response rate of 80%. This approach offers insights into the effectiveness of ESM in clinical practice without the use of financial incentives but might have further contributed to a low participation rate. Finally, we recruited patients through flyers sent by post and email, as well as during group meetings. An individualized or personalized recruitment strategy could potentially increase participation rates.

Our compliance rate was 56%, while previous experience sampling studies regarding affect and problematic eating behaviors showed higher compliance rates ranging from 62% to 81% [17,44,49,50]. Interestingly, compliance rates were higher (62%) in the studies by Stevenson et al [50] and Boh et al [44] (80% for participants with healthy weight and 81% for participants who were overweight), despite both studies administering more daily ESM questionnaires over a similar study duration compared with ours. First, differences in study populations may account for the discrepancy in compliance rates. For instance, Boh et al [44] included participants who were both healthy-weight and overweight, but none of these participants had recently undergone MBS. MBS entails associated changes and challenges, particularly in the short term following surgery. This might have played a role in the lower compliance rates observed in our study. Participants in the study by Stevenson et al [50] were younger compared with our participants (age 24 years versus 47 years in our study). This could potentially explain their higher compliance rates, as mobile apps for self-management are known to be used less frequently by older adults compared with younger individuals [51,52]. Second, to minimize recall bias, we restricted the time window for participants to respond to ESM questionnaires to 15 minutes. Several participants mentioned in the semistructured interviews that this short-time window hindered their ability to respond, possibly contributing to our lower compliance rate. However, other ESM studies with longer response windows have reported compliance rates that are either lower or similar to ours [28,31], whereas others with similar time windows achieved higher compliance rates [33]. This indicates that extending the time window or increasing reminder frequency, is unlikely to positively impact compliance within our study population.

This study is one of the few to assess problematic eating behaviors using ESM in the population after MBS [18,38,39]. While different methods exist for assessing problematic eating behaviors, we used terminology consistent with previous research in populations with overweight or obesity [53,54]. Our study identified a significant temporal association between hunger and craving (P=.03). However, an OR close to 1 (OR 1.04) indicates a minimal effect of hunger on craving. Previous studies also found that food cravings closely follow hunger [55,56]. Given that food craving is defined as an intense desire to consume a specific food, and hunger as an urgent need for food or a specific nutrient, it is unsurprising that both constructs are closely intertwined. This overlap might create difficulties for participants to differentiate between both constructs and therefore, an association between hunger and craving could have arisen. Our results also revealed a significant temporal association between positive affect and grazing (OR 1.61, 95% CI 1.03-2.51; P=.04), where grazing behaviors increased with higher levels of positive affect. As grazing behaviors were not specifically assessed in previous studies, direct comparisons with existing literature are not possible. However, practical implications arise from the identified associations. For example, recognizing a positive mood might help patients in acknowledging the potential risk of grazing behavior.

Overall, our study yielded inconclusive results regarding the association between negative emotions, positive emotions, and problematic eating behaviors, as no clear patterns emerged from the GEE analyses. ORs regarding problematic eating behaviors for both positive and negative emotions (including boredom and disgust) were notably ambiguous. This is unexpected, as most theoretical models of eating disorders propose that negative emotions trigger episodes of problematic eating behavior [13].

XSL•FO
RenderX

It is likely that the impact of MBS, which encompasses both mental and physical adjustments, might have influenced the inconsistencies in positive and negative affect on problematic eating behaviors in our study. Future studies focusing on the longer term after MBS are needed to assess the association between both positive and negative affect and problematic eating behaviors. Furthermore, participants in our study exhibited fewer episodes of problematic eating behaviors compared with existing literature [57], resulting in a low number of events, which likely impacted the robustness of our analyses. Consequently, the limited frequency of problematic eating episodes observed in our sample, combined with potential influences of post-MBS adaptations, may have contributed to the inconclusive findings regarding the role of emotions in problematic eating behaviors.

Study Strengths and Limitations

As we were interested in the feasibility of ESM in the population after MBS, we used an intensive sampling strategy over 14 consecutive days with 6 prompts per day, which is a strength of our study. Furthermore, we collected both quantitative and qualitative data to answer our primary research aim, which increased the validity of the study.

There are a few limitations to this study, so the results should be interpreted with caution. First, the relatively small sample size is a limitation of our study. However, the small sample size was justifiable, due to the large amount of collected data by intensive sampling, and to the exploratory nature of the study. Despite the small sample size, we add to the limited knowledge about the feasibility of ESM in the population after MBS. Second, we conducted multiple tests without correction, which may increase the risk of type I statistical errors, that is, finding significant results while there are no true underlying associations. In exploratory studies, the primary goal is to identify potential relationships, making it more appropriate to forego corrections. This approach allows for the identification of possible associations that can be further examined and validated in follow-up confirmatory studies. Third, the overall frequency of problematic eating behaviors in our study population was extremely low. To minimize recall bias, we asked participants about problematic eating behaviors occurring within the past 30 or 60 minutes, which may have missed behaviors occurring outside this time frame. Wider time windows could probably overcome this limitation, but this goes hand in hand with recall bias. Furthermore, behavior can change when monitored, even without an intended change, a phenomenon known as “measurement reactivity,” which may have increased participants’ awareness of their eating behaviors and led them to adjust their responses or actions accordingly. Together, this may have resulted in underreporting of problematic eating behaviors in our study. Finally, our assessment of binge eating focused solely on the quantity of food consumed, without evaluating LOC. Previous research highlights that LOC may be a stronger indicator of binge eating episodes, particularly in patients with obesity and patients undergoing MBS [58,59]. By focusing on quantity alone, we may have misclassified certain episodes or overlooked critical associations related to LOC. However, it is important to note that our primary aim was to assess the feasibility of ESM, which is not directly impacted by this limitation. For our secondary aim, the small number of binge eating episodes (n=7 across 3 participants) already posed challenges in detecting meaningful associations. Including the LOC aspect could have further reduced this number, potentially making statistical analyses even more difficult or infeasible. Nonetheless, we recognize that future studies should include LOC as a key criterion for binge eating to enhance content validity and better capture associations with momentary predictors.

Future Perspectives

This study provides insights into the feasibility of ESM in the population after MBS. This information could be used to further refine the development of ESM to monitor patients during their postoperative trajectory. Several participants mentioned that the questionnaires were repetitive and monotonous. Compliance might be increased if the questions are asked in a different order. Furthermore, patients would appreciate a free-input text area to explain their answers about emotion and problematic eating behavior. Feasibility for health care personnel should be considered and evaluated in future studies, especially when the objective of ESM includes remote monitoring.

Monitoring efforts should be particularly directed toward individuals exhibiting problematic eating behaviors. Such results could help patients and health care professionals to identify problematic eating behavior episodes by recommending them to focus on specific emotions, physical states, or contextual factors.

Conclusion

From this exploratory study to describe the feasibility of ESM regarding negative and positive affect, physical states, contextual factors, and problematic eating behaviors, findings suggest that ESM can be feasible in the population after MBS. Despite the low participation rate, compliance with experience sampling was moderate, and the completion rate was high. According to the interviews, patients considered ESM feasible and of added value, as awareness arose by using the ESM app. This offers a perspective for the use of ESM in the follow-up after MBS. Temporal associations were found between hunger and craving, and between positive affect and grazing. Nevertheless, no distinct patterns emerged among emotions, physical states, context, and problematic eating behaviors.

Acknowledgments

The study was performed as part of the Exceptional Deep and Intelligent Coach (EDIC) project, grant 628.011.021, funded by the Netherlands Organization for Scientific Research (NWO). Artificial intelligence was not used in writing this manuscript.

XSL•FO
RenderX

Data Availability

The data generated and analyzed during this study are not publicly available since participants were not asked for consent to make their data publicly available. The data are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Semistructured interview guide.
[DOCX File , 16 KB - formative_v9i1e60486_app1.docx]

Multimedia Appendix 2
Compliance in answering experience sampling questionnaires per participant during the 14-day study period.
[PNG File , 110 KB - formative_v9i1e60486_app2.png]

Multimedia Appendix 3
Descriptive data for the different types of problematic eating behavior.
[DOCX File , 19 KB - formative_v9i1e60486_app3.docx]

References

1. Sjöström L, Lindroos A, Peltonen M, Torgerson J, Bouchard C, Carlsson B, et al. Lifestyle, diabetes, and cardiovascular risk factors 10 years after bariatric surgery. N Engl J Med 2004;351(26):2683-2693. [doi: 10.1056/NEJMoa035622] [Medline: 15616203]
2. Karmali S, Brar B, Shi X, Sharma AM, de Gara C, Birch DW. Weight recidivism post-bariatric surgery: a systematic review. Obes Surg 2013;23(11):1922-1933. [doi: 10.1007/s11695-013-1070-4] [Medline: 23996349]
3. Amundsen T, Strømmen M, Martins C. Suboptimal weight loss and weight regain after gastric bypass surgery-postoperative status of energy intake, eating behavior, physical activity, and psychometrics. Obes Surg 2017;27(5):1316-1323 [FREE Full text] [doi: 10.1007/s11695-016-2475-7] [Medline: 27914028]
4. Kaouk L, Hsu AT, Tanuseputro P, Jessri M. Modifiable factors associated with weight regain after bariatric surgery: a scoping review. F1000Res 2019;8:615 [FREE Full text] [doi: 10.12688/f1000research.18787.2] [Medline: 32983412]
5. Freire RH, Borges MC, Alvarez-Leite JI, Toulson Davisson Correia MI. Food quality, physical activity, and nutritional follow-up as determinant of weight regain after Roux-en-Y gastric bypass. Nutrition 2012;28(1):53-58. [doi: 10.1016/j.nut.2011.01.011] [Medline: 21885246]
6. Reid RER, Oparina E, Plourde H, Andersen RE. Energy intake and food habits between weight maintainers and regainers, five years after Roux-en-Y gastric bypass. Can J Diet Pract Res 2016;77(4):195-198. [doi: 10.3148/cjdpr-2016-013] [Medline: 27744735]
7. Colles SL, Dixon JB, O'Brien PE. Grazing and loss of control related to eating: two high-risk factors following bariatric surgery. Obesity (Silver Spring) 2008;16(3):615-622 [FREE Full text] [doi: 10.1038/oby.2007.101] [Medline: 18239603]
8. Kofman MD, Lent MR, Swencionis C. Maladaptive eating patterns, quality of life, and weight outcomes following gastric bypass: results of an internet survey. Obesity (Silver Spring) 2010;18(10):1938-1943 [FREE Full text] [doi: 10.1038/oby.2010.27] [Medline: 20168309]
9. Meany G, Conceição E, Mitchell JE. Binge eating, binge eating disorder and loss of control eating: effects on weight outcomes after bariatric surgery. Eur Eat Disord Rev 2014;22(2):87-91 [FREE Full text] [doi: 10.1002/erv.2273] [Medline: 24347539]
10. Pizato N, Botelho P, Gonçalves VSS, Dutra E, de Carvalho K. Effect of grazing behavior on weight regain post-bariatric surgery: a systematic review. Nutrients 2017;9(12):1322 [FREE Full text] [doi: 10.3390/nu9121322] [Medline: 29206132]
11. Fairburn CG, Cooper Z, Shafran R. Cognitive behaviour therapy for eating disorders: a "transdiagnostic" theory and treatment. Behav Res Ther 2003;41(5):509-528. [doi: 10.1016/s0005-7967(02)00088-8] [Medline: 12711261]
12. Polivy J, Herman CP. Causes of eating disorders. Annu Rev Psychol 2002;53:187-213. [doi: 10.1146/annurev.psych.53.100901.135103] [Medline: 11752484]
13. Dingemans A, Danner U, Parks M. Emotion regulation in binge eating disorder: a review. Nutrients 2017;9(11):1274 [FREE Full text] [doi: 10.3390/nu9111274] [Medline: 29165348]
14. Ahlich E, Rancourt D. Boredom proneness, interoception, and emotional eating. Appetite 2022;178:106167. [doi: 10.1016/j.appet.2022.106167] [Medline: 35843373]
15. van Strien T, Herman CP, Anschutz DJ, Engels RC, de Weerth C. Moderation of distress-induced eating by emotional eating scores. Appetite 2012;58(1):277-284. [doi: 10.1016/j.appet.2011.10.005] [Medline: 22037008]

XSL•FO
RenderX

16. Koball AM, Meers MR, Storfer-Isser A, Domoff SE, Musher-Eizenman DR. Eating when bored: revision of the emotional eating scale with a focus on boredom. Health Psychol 2012;31(4):521-524. [doi: 10.1037/a0025893] [Medline: 22004466]
17. Schaefer LM, Engel SG, Wonderlich SA. Ecological momentary assessment in eating disorders research: recent findings and promising new directions. Curr Opin Psychiatry 2020;33(6):528-533 [FREE Full text] [doi: 10.1097/YCO.0000000000000639] [Medline: 32740204]
18. Williams-Kerver GA, Steffen KJ, Smith KE, Cao L, Crosby RD, Engel SG. Negative affect and loss of control eating among bariatric surgery patients: an ecological momentary assessment pilot investigation. Obes Surg 2020;30(6):2382-2387 [FREE Full text] [doi: 10.1007/s11695-020-04503-6] [Medline: 32125646]
19. Hudson JI, Hiripi E, Pope HG, Kessler RC. The prevalence and correlates of eating disorders in the national comorbidity survey replication. Biol Psychiatry 2007;61(3):348-358 [FREE Full text] [doi: 10.1016/j.biopsych.2006.03.040] [Medline: 16815322]
20. Opolski M, Chur-Hansen A, Wittert G. The eating-related behaviours, disorders and expectations of candidates for bariatric surgery. Clin Obes 2015;5(4):165-197. [doi: 10.1111/cob.12104] [Medline: 26173752]
21. de Zwaan M, Hilbert A, Swan-Kremeier L, Simonich H, Lancaster K, Howell LM, et al. Comprehensive interview assessment of eating behavior 18-35 months after gastric bypass surgery for morbid obesity. Surg Obes Relat Dis 2010;6(1):79-85. [doi: 10.1016/j.soard.2009.08.011] [Medline: 19837012]
22. Barbuti M, Brancati GE, Calderone A, Fierabracci P, Salvetti G, Weiss F, et al. Prevalence of mood, panic and eating disorders in obese patients referred to bariatric surgery: patterns of comorbidity and relationship with body mass index. Eat Weight Disord 2022;27(3):1021-1027 [FREE Full text] [doi: 10.1007/s40519-021-01236-y] [Medline: 34137006]
23. Heriseanu AI, Hay P, Corbit L, Touyz S. Grazing in adults with obesity and eating disorders: a systematic review of associated clinical features and meta-analysis of prevalence. Clin Psychol Rev 2017;58:16-32. [doi: 10.1016/j.cpr.2017.09.004] [Medline: 28988855]
24. Guthrie H, Tetley D, Hill AJ. Quasi-prospective, real-life monitoring of food craving post-bariatric surgery: comparison with overweight and normal weight women. Clin Obes 2014;4(3):136-142 [FREE Full text] [doi: 10.1111/cob.12054] [Medline: 25826768]
25. Erden Aki SO, Yıldız MI, Kamış GZ, Aytulun A, Karakaya J, Işıklı S. Relationship of depression, impulsivity, distress intolerance and coping styles with maladaptive eating patterns in bariatric candidates. Eat Weight Disord 2022;27(8):3351-3366. [doi: 10.1007/s40519-022-01465-9] [Medline: 35999437]
26. Mehl MR, Conner TS. Handbook of Research Methods for Studying Daily Life. New York, NY: The Guilford Press; 2012.
27. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
28. Spook JE, Paulussen T, Kok G, van Empelen P. Monitoring dietary intake and physical activity electronically: feasibility, usability, and ecological validity of a mobile-based Ecological Momentary Assessment tool. J Med Internet Res 2013;15(9):e214 [FREE Full text] [doi: 10.2196/jmir.2617] [Medline: 24067298]
29. Timmerman JG, Dekker-van Weering MGH, Stuiver MM, Groen WG, Wouters MWJM, Tönis TM, et al. Ambulant monitoring and web-accessible home-based exercise program during outpatient follow-up for resected lung cancer survivors: actual use and feasibility in clinical practice. J Cancer Surviv 2017;11(6):720-731 [FREE Full text] [doi: 10.1007/s11764-017-0611-6] [Medline: 28396993]
30. Mackesy-Amiti ME, Boodram B. Feasibility of ecological momentary assessment to study mood and risk behavior among young people who inject drugs. Drug Alcohol Depend 2018;187:227-235 [FREE Full text] [doi: 10.1016/j.drugalcdep.2018.03.016] [Medline: 29684890]
31. Shaffer KM, Chow PI, Glazer JV, Le T, Reilley MJ, Jameson MJ, et al. Feasibility of ecological momentary assessment to study depressive symptoms among cancer caregivers. Psychooncology 2021;30(5):756-764 [FREE Full text] [doi: 10.1002/pon.5627] [Medline: 33432717]
32. Porras-Segovia A, Molina-Madueño RM, Berrouiguet S, López-Castroman J, Barrigón ML, Pérez-Rodríguez MS, et al. Smartphone-based ecological momentary assessment (EMA) in psychiatric patients and student controls: a real-world feasibility study. J Affect Disord 2020;274:733-741. [doi: 10.1016/j.jad.2020.05.067] [Medline: 32664009]
33. Jean FAM, Sibon I, Husky M, Couffinhal T, Swendsen J. Feasibility and validity of ecological momentary assessment in patients with acute coronary syndrome. BMC Cardiovasc Disord 2020;20(1):499 [FREE Full text] [doi: 10.1186/s12872-020-01774-w] [Medline: 33246420]
34. Maatoug R, Peiffer-Smadja N, Delval G, Brochu T, Pitrat B, Millet B. Ecological momentary assessment using smartphones in patients with depression: feasibility study. JMIR Form Res 2021;5(2):e14179 [FREE Full text] [doi: 10.2196/14179] [Medline: 33625367]
35. Wooldridge JS, Soriano EC, Harris DE, Afari N. Feasibility and acceptability of ecological momentary assessment of psychosocial factors and self-management behaviors among veterans with type 2 diabetes. Diabetes Spectr 2022;35(1):76-85 [FREE Full text] [doi: 10.2337/ds21-0020] [Medline: 35308149]
36. Glenn CR, Kleiman EM, Kearns JC, Santee AC, Esposito EC, Conwell Y, et al. Feasibility and acceptability of ecological momentary assessment with high-risk suicidal adolescents following acute psychiatric care. J Clin Child Adolesc Psychol 2022;51(1):32-48. [doi: 10.1080/15374416.2020.1741377] [Medline: 32239986]

XSL•FO
RenderX

37. Arend AK, Kaiser T, Pannicke B, Reichenberger J, Naab S, Voderholzer U, et al. Toward individualized prediction of binge-eating episodes based on ecological momentary assessment data: item development and pilot study in patients with bulimia nervosa and binge-eating disorder. JMIR Med Inform 2023;11:e41513 [FREE Full text] [doi: 10.2196/41513] [Medline: 36821359]
38. Thomas JG, Schumacher LM, Vithiananthan S, Jones DB, Smith KE, Chou T, et al. Ecological momentary assessment of changes in eating behaviors, appetite, and other aspects of eating regulation in roux-en-Y gastric bypass and sleeve gastrectomy patients. Appetite 2023;183:106465 [FREE Full text] [doi: 10.1016/j.appet.2023.106465] [Medline: 36701847]
39. Bond DS, Thomas JG, Jones DB, Schumacher LM, Webster J, Evans EW, et al. Ecological momentary assessment of gastrointestinal symptoms and risky eating behaviors in Roux-en-Y gastric bypass and sleeve gastrectomy patients. Surg Obes Relat Dis 2021;17(3):475-483 [FREE Full text] [doi: 10.1016/j.soard.2020.11.017] [Medline: 33353862]
40. Mundi MS, Lorentz PA, Grothe K, Kellogg TA, Collazo-Clavell ML. Feasibility of smartphone-based education modules and ecological momentary assessment/intervention in pre-bariatric surgery patients. Obes Surg 2015;25(10):1875-1881. [doi: 10.1007/s11695-015-1617-7] [Medline: 25702141]
41. Thomas JG, Bond DS, Ryder BA, Leahey TM, Vithiananthan S, Roye GD, et al. Ecological momentary assessment of recommended postoperative eating and activity behaviors. Surg Obes Relat Dis 2011;7(2):206-212. [doi: 10.1016/j.soard.2010.10.007] [Medline: 21130703]
42. Wrzus C, Neubauer AB. Ecological momentary assessment: a meta-analysis on designs, samples, and compliance across research fields. Assessment 2023;30(3):825-846 [FREE Full text] [doi: 10.1177/10731911211067538] [Medline: 35016567]
43. Macht M, Haupt C, Salewsky A. Emotions and eating in everyday life: application of the experience-sampling method. Ecology of Food and Nutrition 2004;43(4):11-21. [doi: 10.1080/03670240490454723]
44. Boh B, Jansen A, Clijsters I, Nederkoorn C, Lemmens LH, Spanakis G, et al. Indulgent thinking? Ecological momentary assessment of overweight and healthy-weight participants' cognitions and emotions. Behav Res Ther 2016;87:196-206. [doi: 10.1016/j.brat.2016.10.001] [Medline: 27736671]
45. Goldstein H. Multilevel Statistical Models. 4th edition. Chichester, UK: John Wiley & Sons Inc; 2011.
46. Myin-Germeys I, Kuppens P. The Open Handbook of Experience Sampling Methodology: A Step-by-Step to Designing, Conducting, and Aalyzing ESM Studies. 2nd edition. Leuven, Belgium: Center for Research on Experience Sampling and Ambulatory Methods Leuven; 2022.
47. Broekhuis M, van Velsen L. Improving usability benchmarking for the eHealth domain: the development of the eHealth UsaBility Benchmarking instrument (HUBBI). PLoS One 2022;17(2):e0262036 [FREE Full text] [doi: 10.1371/journal.pone.0262036] [Medline: 35176033]
48. Ratcliff MB, Zeller MH, Inge TH, Hrovat KB, Modi AC. Feasibility of ecological momentary assessment to characterize adolescent postoperative diet and activity patterns after weight loss surgery. Surg Obes Relat Dis 2014;10(4):705-710 [FREE Full text] [doi: 10.1016/j.soard.2014.01.034] [Medline: 24837564]
49. Manasse SM, Schumacher LM, Goldstein SP, Martin GJ, Crosby RD, Juarascio AS, et al. Are individuals with loss-of-control eating more prone to dietary lapse in behavioural weight loss treatment? An ecological momentary assessment study. Eur Eat Disord Rev 2018;26(3):259-264 [FREE Full text] [doi: 10.1002/erv.2583] [Medline: 29484774]
50. Stevenson BL, Dvorak RD, Wonderlich SA, Crosby RD, Gordon KH. Emotions before and after loss of control eating. Eat Disord 2018;26(6):505-522. [doi: 10.1080/10640266.2018.1453634] [Medline: 29565734]
51. Bhuyan SS, Lu N, Chandak A, Kim H, Wyant D, Bhatt J, et al. Use of mobile health applications for health-seeking behavior among US adults. J Med Syst 2016;40(6):153. [doi: 10.1007/s10916-016-0492-7] [Medline: 27147516]
52. Jiang Y, West BT, Barton DL, Harris MR. Acceptance and use of eHealth/mHealth applications for self-management among cancer survivors. Stud Health Technol Inform 2017;245:131-135 [FREE Full text] [doi: 10.3233/978-1-61499-830-3-131] [Medline: 29295067]
53. Conceição E, Mitchell JE, Vaz AR, Bastos AP, Ramalho S, Silva C, et al. The presence of maladaptive eating behaviors after bariatric surgery in a cross sectional study: importance of picking or nibbling on weight regain. Eat Behav 2014;15(4):558-562. [doi: 10.1016/j.eatbeh.2014.08.010] [Medline: 25213792]
54. Roefs A, Boh B, Spanakis G, Nederkoorn C, Lemmens LHJM, Jansen A. Food craving in daily life: comparison of overweight and normal-weight participants with ecological momentary assessment. J Hum Nutr Diet 2019;32(6):765-774 [FREE Full text] [doi: 10.1111/jhn.12693] [Medline: 31430000]
55. Reichenberger J, Richard A, Smyth JM, Fischer D, Pollatos O, Blechert J. It's craving time: time of day effects on momentary hunger and food craving in daily life. Nutrition 2018;55-56:15-20. [doi: 10.1016/j.nut.2018.03.048] [Medline: 29960151]
56. Dicker-Oren SD, Gelkopf M, Greene T. The dynamic network associations of food craving, restrained eating, hunger and negative emotions. Appetite 2022;175:106019. [doi: 10.1016/j.appet.2022.106019] [Medline: 35500722]
57. Goldschmidt AB, Crosby RD, Cao L, Wonderlich SA, Mitchell JE, Engel SG, et al. A preliminary study of momentary, naturalistic indicators of binge-eating episodes in adults with obesity. Int J Eat Disord 2018;51(1):87-91 [FREE Full text] [doi: 10.1002/eat.22795] [Medline: 29112288]
58. Goldschmidt AB, Engel SG, Wonderlich SA, Crosby RD, Peterson CB, Le Grange D, et al. Momentary affect surrounding loss of control and overeating in obese adults with and without binge eating disorder. Obesity (Silver Spring) 2012;20(6):1206-1211 [FREE Full text] [doi: 10.1038/oby.2011.286] [Medline: 21938073]

XSL•FO
RenderX

59. Ivezaj V, Lydecker JA, Wiedemann AA, Duffy AJ, Grilo CM. Does bariatric binge-eating size matter? Conceptual model and empirical support. Obesity (Silver Spring) 2020;28(9):1645-1651 [FREE Full text] [doi: 10.1002/oby.22876] [Medline: 32729221]

Abbreviations

ESM: experience sampling methodology
GEE: generalized estimation equation
HUBBI: eHealth Usability Benchmarking Instrument
LOC: loss of control
MBS: metabolic bariatric surgery
OR: odds ratio
ZGT: Hospital Group Twente

Edited by A Mavragani; submitted 07.06.24; peer-reviewed by M Ben Amar, G Williams - Kerver; comments to author 12.08.24; revised version received 11.09.24; accepted 07.01.25; published 05.03.25.

Please cite as:
Kuipers EAM, Timmerman JG, van Det MJ, Vollenbroek-Hutten MMR
Feasibility and Links Between Emotions, Physical States, and Eating Behavior in Patients After Metabolic Bariatric Surgery: Experience Sampling Study
JMIR Form Res 2025;9:e60486
URL: https://formative.jmir.org/2025/1/e60486
doi:10.2196/60486
PMID:40053719

©Ellen A M Kuipers, Josien G Timmerman, Marc J van Det, Miriam M R Vollenbroek-Hutten. Originally published in JMIR Formative Research (https://formative.jmir.org), 05.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

FRAILSURVEY—an mHealth App for Self-Assessment of Frailty Based on the Portuguese Version of the Groningen Frailty Indicator: Validation and Reliability Study

Luis Midao[1,2,3], PhD; Mafalda Duarte[4,5], Prof Dr; Rute Sampaio[6], Prof Dr; Marta Almada[1,2], PhD; Cláudia Camila Dias[7,8], Prof Dr; Constança Paúl[3,4], Prof Dr; Elísio Costa[1,2], Prof Dr

[1]RISE-Health, Biochemistry Lab, Faculty of Pharmacy, University of Porto, Porto, Portugal

[2]Porto4Ageing—Competences Centre on Active and Healthy Ageing, Faculty of Pharmacy, University of Porto, Porto, Portugal

[3]ICBAS—School of Medicine and Biomedical Sciences, University of Porto, Porto, Portugal

[4]RISE-Health, Department of Behavioural Sciences, School of Medicine and Biomedical Sciences, University of Porto, Porto, Portugal

[5]ISAVE—Superior Institute of Health, Amares, Braga, Portugal

[6]RISE-Health, Department of Biomedicine, Faculty of Medicine of the University of Porto, Porto, Portugal

[7]Knowledge Management Unit, Faculty of Medicine of the University of Porto (FMUP), Porto, Portugal

[8]CINTESIS@RISE, Department of Community Medicine, Information and Health Decision Sciences (MEDCIDS), Faculty of Medicine of the University of Porto (FMUP), Porto, Portugal

Corresponding Author:
Luis Midao, PhD
RISE-Health, Biochemistry Lab, Faculty of Pharmacy, University of Porto
R. Jorge de Viterbo Ferreira 228
Porto, 4050-313
Portugal
Phone: 351 22 042 8500
Email: luismidao@gmail.com

Abstract

Background: Portugal is facing the challenge of population ageing, with a notable increase in the proportion of older individuals. This has positioned the country among those in Europe with a high prevalence of frailty. Frailty, a geriatric syndrome characterized by diminished physiological reserve and heightened vulnerability to stressors, imposes a substantial burden on public health.

Objective: This study seeks to address two primary objectives: (1) translation and psychometric evaluation of the European Portuguese version of the Groningen Frailty Indicator (GFI); and (2) development and evaluation of the FRAILSURVEY app, a novel assessment tool for frailty based on the GFI. By achieving these objectives, the study aims to enhance the accuracy and reliability of frailty assessment in the Portuguese context, ultimately contributing to improved health care outcomes for older individuals in the region.

Methods: To accomplish the objectives of the study, a comprehensive research methodology was used. The study comprised 2 major phases: the initial translation and validation of the GFI into European Portuguese and the development of the FRAILSURVEY app. Following this, an extensive examination of the app's validity and reliability was conducted compared with the conventional paper version of the GFI. A randomized repeated crossover design was used to ensure rigorous evaluation of both assessment methods, using both the paper form of the GFI and the smartphone-based app FRAILSURVEY.

Results: The findings of the study revealed promising outcomes in line with the research objectives. The meticulous translation process yielded a final version of the GFI with robust psychometric properties, ensuring clarity and comprehensibility for participants. The study included 522 participants, predominantly women (367/522, 70.3%), with a mean age of 73.7 (SD 6.7) years. Psychometric evaluation of the European Portuguese GFI in paper form demonstrates good reliability (internal consistency: Cronbach a value of 0.759; temporal stability: intraclass correlation coefficient=0.974) and construct validity (revealing a 4D structure explaining 56% of variance). Evaluation of the app-based European Portuguese GFI indicates good reliability (interinstrument reliability: Cohen k=0.790; temporal stability: intraclass correlation coefficient=0.800) and concurrent validity (r=0.694; P<.001).

XSL•FO
RenderX

Conclusions: Both the smartphone-based app and the paper version of the GFI were feasible and acceptable for use. The findings supported that FRAILSURVEY exhibited comparable validity and reliability to its paper counterpart. FRAILSURVEY uses a standardized and validated assessment tool, offering objective and consistent measurements while eliminating subjective biases, enhancing accuracy, and ensuring reliability. This app holds promising potential for aiding health care professionals in identifying frailty in older individuals, enabling early intervention, and improving the management of adverse health outcomes associated with this syndrome. Its integration with electronic health records and other data may lead to personalized interventions, improving frailty management and health outcomes for at-risk individuals.

(JMIR Form Res 2025;9:e51975) doi:10.2196/51975

KEYWORDS

frailty; mHealth; assessment; validation; GFI; reliability; self-assessment; Groningen Frailty Indicator; FRAILSURVEY; mobile phone

Introduction

Background

Portugal is currently confronting a significant population ageing phenomenon, placing it among the European countries with one of the highest percentages of older individuals. As of 2021, the ageing index reached 182, meaning there were 182 people aged 65 years or older for every 100 individuals aged 0-14 years. This ageing trend results from various contributing factors, such as low birth rates, extended life expectancy, and emigration of young people. According to the 2011 census, 19% of the Portuguese population were aged 65 years or older. A substantial increase was denoted to 23.4% (2,423,639/10,343,066) in the 2021 census. Projections indicate that this percentage will continue to rise in the upcoming years. Estimations forecast that by 2080, the proportion of people aged 65 years and older will escalate to 37.5% of the total population [1-4]. This demographic change has significant implications for the country's economy, social security, and health care systems. The increasing number of older people will require greater resources to provide for their care and support, including health care services, long-term care facilities, and home care services. At the same time, the working-age population will be under pressure to support the social security system as the number of retirees increases and the number of workers decreases. As the proportion of older people grows, so does the prevalence of age-related diseases, comorbidities, and geriatric syndromes. Portugal ranks among the European countries with the highest prevalence of frailty (20.7%) [5,6].

Frailty is characterized by a reduced physiologic reserve and increased susceptibility to stressors, leading to an elevated risk of adverse health outcomes such as falls, hospitalization, disability, and death. Typically associated with ageing, frailty may encompass a blend of physical, cognitive, and psychosocial factors, making it a complex and multifaceted condition, and, therefore, identifying and addressing frailty is essential for preventing comorbidities, optimizing health outcomes, and improving the quality of life in older adults [7-11]. However, evaluating frailty poses challenges due to its multifactorial nature, resulting in various tools and methods commonly used for its assessment. It is crucial to recognize that every assessment tool is flawed, lacks standardization, and different approaches may be more suitable for specific populations or settings [10,12].

The Groningen Frailty Indicator (GFI) stands out for its comprehensive assessment for evaluating frailty in older adults. With its rapid administration and validated reliability in diverse populations, the GFI serves as a predictive tool for adverse outcomes such as hospitalization and mortality. Its multidimensional approach offers a nuanced understanding of frailty, making it invaluable for routine screening in clinical practice, empowering health care professionals to craft suitable care plans for their patients [13-16]. Covering physical, psychological, social, and cognitive aspects, the paper version of the GFI has undergone validation in diverse settings, such as among Romanian community-dwelling older adults [17] and Chinese nursing home residents [18], and different populations, such as patients with rheumatoid arthritis [19], patients with end-stage hip and knee osteoarthritis [20], and head and neck cancer surgery patients [21], confirming its accuracy, consistency, and predictive value. Distinguishing itself from other instruments, the GFI boasts a straightforward administration process, requiring no more than 10 minutes to complete. Furthermore, as a noninvasive instrument, individuals can complete it through self-reporting or seek guidance from a trained health professional, if necessary [15,16]. There are already some versions in European Portuguese of frailty scales, such as the Edmonton Frailty Scale [22], Program on Research for Integrating Services for the Maintenance of Autonomy questionnaire (PRISMA-7) [23], FRAIL Scale [24], Clinical Frailty Scale [25], SUNFRAIL [26], and Tilburg Frailty Indicator [27]; however, some are not yet validated for the Portuguese population. To our knowledge, there is still no study on the translation and validation of the GFI into European Portuguese.

Despite being commonly used, paper-based questionnaires have several disadvantages: manual data entry is time-consuming and leads to potential errors; accessibility is limited as they must be physically distributed and collected; data security may be compromised; and incomplete or illegible responses and reduced data analysis capabilities are common challenges. Furthermore, they lack interactive features that enhance engagement and data quality, such as in digital formats [28]. Smartphone-based assessments aim to develop a user-friendly tool that streamlines screening and monitoring by scoring data [29,30]. Such an approach has the potential to increase health literacy resulting in cost savings. Nevertheless, it is crucial to validate the smartphone-based apps to ensure their accuracy and reliability

XSL•FO
RenderX

compared with the original paper questionnaires, as there may be possible response bias between the 2 versions [31].

Aim

Given the high burden of frailty in Portugal and to contribute to improving frailty assessment, enhancing health care delivery, and addressing the specific needs of older adults in Portugal, this study aimed to translate and validate the GFI into European Portuguese. Then, this study also aimed to examine the validity and reliability of the digital version of the GFI, the app FRAILSURVEY.

Methods

This is a quantitative, exploratory research that involves analysis of psychometric data and that was developed in 2 major phases: initially, the GFI was translated and validated into European Portuguese, and the application FRAILSURVEY was developed and its validity and reliability studied as a mobile health (mHealth) tool for frailty assessment.

Translation and Validation of the GFI

The GFI is a short, easy-to-use instrument that has proven to be a good alternative for assessing frailty in the population aged 65 years or older [32]. Textbox 1 indicates the domains assessed in the GFI scale, as well as scoring items. The self-assessment version of the GFI published in 2013 was used, composed of 15 items and divided into 4 domains [33]. The physical domain has 9 questions related to toileting, shopping, mobility functions, dressing, physical fitness, vision, hearing, unexplained weight loss, and medication use. The cognitive domain includes 1 item related to cognitive concerns. The social domain consists of 3 questions about emotional isolation, social relations, and social support. The psychological domain comprises 2 questions related to mood. Each item has a score of 0 or 1 attributed, with 1 indicating a dependency problem. For example, if the participant cannot dress or undress without help, he would score 1. The total score ranged from 0 to 15, and higher scores indicate greater levels of frailty. The originally proposed cutoff point of 4 or higher represents frailty [16].

Textbox 1. Self-assessment version of the Groningen Frailty Indicator (GFI): with 15 items divided by 4 main domains (physical, cognitive, social, and psychological), the GFI scores between 0 and 15. People with scores of 4 and above are considered frail [33].

Physical domain

Are you able to carry out these tasks single-handedly and without any help? (The use of help resources such as a walking stick, walking frame, or wheelchair is considered independent.)

1. Shopping
2. Walking around outside (around the house or to the neighbors)
3. Dressing and undressing
4. Going to the toilet
5. What mark do you give yourself for physical fitness? (Scales 0-10)
6. Do you experience problems in daily life due to poor vision?
7. Do you experience problems in daily life due to being hard of hearing?
8. During the last 6 months have you lost a lot of weight unwillingly? (3 kg in 1 month or 6 kg in 2 months)
9. Do you take 4 or more different types of medicine?

Cognitive domain

10. Do you have any complaints about your memory?

Social domain

11. Do you sometimes experience emptiness around yourself?
12. Do you sometimes miss people around yourself?
13. Do you sometimes feel abandoned?

Psychological domain

14. Have you recently felt downhearted or sad?
15. Have you recently felt nervous or anxious?

Scoring

Questions 1-4: Yes=0; No=1

Question 5: 0-6=1; 7-10=0

Questions 6-15: No=0; Yes=1

For the development of the European Portuguese version of the GFI, the self-report instrument guidelines defined by the ISPOR Task Force for Translation and Cultural Adaptation were followed [34]. Briefly, the translation was done by 2 technicians,

XSL•FO
RenderX

with experience in this type of task, and the final version of each item (of the 15) always resulted from complete agreement between them. After that an English technician carried out the back-translation, and a consensus was reached as the back-translation version was in line with the original version.

Developing the Smartphone-Based App of the GFI

The smartphone-based app was developed for both Android (the operating system created by Google) and iOS (created by Apple) systems. FRAILSURVEY is a questionnaire that comprises 2 sets of questions: 19 questions about sociodemographic data; social resources; and self-perception of health, nutrition, medication, psychosocial and cognitive status, and time occupation, plus a set of 15 questions used to assess frailty status, the European Portuguese version of the GFI. FRAILSURVEY also uses the originally proposed cutoff point of 4 or higher representing frailty [16].

FRAILSURVEY development was based on an accessibility principle for older adults to guarantee greater autonomy to use it. Therefore, considerations such as vision, hearing, motor control, and cognition were considered. Considering the vision of the older population, the app features a typographic font designed for low-vision people, APHont, developed by the American Printing House for the Blind. It also presents a way to increase or decrease the font size, as well as the possibility of listening to the audio information.

The results obtained classify the person as robust or frail. The app displays tailored recommendations if the person is classified as frail. The recommendations revolve around aspects such as physical exercise, a healthy diet, and social networks, among others (Figure 1). The app generates an anonymous database for research purposes.

Figure 1. Printscreens of the FRAILSURVEY app.

XSL•FO
RenderX

Study Design

A randomized repeated crossover design was used with the paper form of the GFI and the smartphone-based app FRAILSURVEY, as used in other similar studies [35]. This design entailed the dual administration of both assessment methods to each participant on distinct occasions, separated by a 1-week interval. The order of presentation for the assessments was randomized to mitigate any potential sequence effects. In evaluating test-retest reliability, a subset of participants was randomly chosen from the larger cohort to participate in retest assessments approximately 2 weeks after their initial interaction. These retest assessments encompassed both the paper-based form and the application [36].

Ethical Considerations

Ethical clearance for the research was granted by the ethical review board of the faculty of pharmacy, University of Porto (33-11-2017). Before their involvement, the rights and responsibilities of participants were comprehensively elucidated, and written consent was obtained from all individuals. Participation was voluntary and no compensation was provided. Throughout the process of data collection and entry, strict confidentiality measures were upheld and all data were anonymized.

Participants

Between February 2019 and February 2020, older adults living in the district of Porto, Portugal, were invited in cultural and sports associations, day centers, and nursing homes to take part in this study. A total of 522 participants were recruited for this study, using a convenience sampling method. The sample included a diverse population of community-dwelling older individuals, residents of nursing homes, and older adults living in their own homes but availing daily care facilities. The inclusion criteria encompassed individuals aged 65 years and older, regardless of gender, who could independently complete the assessments. The sample exceeded the sample size required to be 10 times larger than the number of GFI items [37]. For test-retest reliability, 10% (52/522) of the participants were recruited randomly, but only 24 participants accepted to perform the retests and, therefore, were included in the test-retest analysis [36].

Statistical Analysis

Categorical variables were summarized as absolute and relative frequencies. Descriptive statistics were used to report the subject characteristics of the study sample. Differences between different subgroups were studied, that is, gender, age, marital status, education, and financial status. GFI scores (from 0 to 15) were presented as mean (SD), and differences between subgroups were tested through the Kruskal-Wallis test. GFI classification (nonfrail and frail) was presented as frequencies, and differences between different subgroups were tested through the Mann-Whitney *U* test. To address the issue of multiple comparisons, adjustments were made, and the Bonferroni correction was applied to all statistical tests involving associations across multiple demographic variables.

The psychometric properties of the European Portuguese version of the GFI were assessed as follows: internal consistency referred to the extent to which the items within a scale or questionnaire consistently measure the same construct and was assessed by Cronbach a. The retest was conducted 2 weeks after the initial contact with the participant for assessing temporal stability. A 2-week interval allows for a sufficient gap between the 2 administrations to minimize potential memory effects and recall bias while maintaining a relatively short time frame. The most commonly used statistical measure to evaluate temporal stability is the intraclass correlation coefficient (ICC), and therefore it was used in this study. The ICC value ranges between 0 and 1, with higher values indicating greater temporal stability. Values above 0.9 suggest excellent temporal stability, while values between 0.75 and 0.9 indicate good temporal stability and values between 0.5 and 0.75 show moderate temporal stability [38]. Construct validity was assessed by performing exploratory factor analysis with principal components and varimax rotation methods [39].

The psychometric properties of the FRAILSURVEY were assessed as follows: interinstrument reliability refers to the consistency of GFI scores obtained by the same person between the paper and smartphone versions. Cohen k was used to assess it. Cohen suggested the k result be interpreted as follows: values ≤0 as indicating no agreement, 0.01-0.20 as none to slight (light), 0.21-0.40 as fair (fair), 0.41-0.60 as moderate (moderate), 0.61-0.80 as substantial (considerable), and 0.81-1.00 as almost perfect agreement [40]; temporal stability was evaluated by the ICC, as described previously; and the concurrent validity was studied through Pearson correlation coefficient that was used to quantify the relationship between the scores obtained in the app and in paper and with the app, the scores from the established criterion measure, and the scores from the instrument we intended to validate. Pearson correlation coefficient assesses the strength and direction of the correlation, indicating the degree to which the 2 measures agree or disagree. The validity assessment in this study was based on demonstrating the absence of significant differences in the total score of the GFI and each item of the GFI. Two-tailed *t* tests were used, and a *P* value of <.05 was considered statistically significant. The significance level used was .05. Data were analyzed using SPSS Statistics (version 28.0; IBM Corp).

Results

European Portuguese Version of the GFI

The English version of the GFI tool was translated into European Portuguese by 2 native Portuguese speakers. The back-translation was done by a professional service provider who knew nothing about the purpose of the research. A panel of experts assessed the similarity of the versions and reviewed and resolved all discrepancies and ambiguities by consensus. To assess the comprehensibility of the language and wording, the revised preliminary European Portuguese version was used with a pilot sample of 10 native Portuguese-speaking older adults. The final European Portuguese version of the GFI was defined and used in the validation process. No changes were made to the version used, as the items were judged to be clear and easy to understand. In the end, the final version was produced (Table S1 in Multimedia Appendix 1).

XSL•FO
RenderX

Participants

In this study, 522 participants were included, with a mean average age of 73.7 (SD 6.7) years and consisted mainly of women (367/522, 70.3%). Almost two-thirds of the sample were aged between 65 and 74 years, most participants were married (311/522, 59.6%), the majority had only primary school completed (283/522, 54.2%), and 59.2% (309/522) reported having moderate money to meet their needs (Table 1). Using the paper version of the European Portuguese version of the GFI, the average score was 4.57, significantly higher in women (5.05) than in men (3.43; *P*<.001). The score increases significantly with age, with higher scores being more present in older people (GFI=3.73 for participants aged between 65 and 74 years; GFI=5.58 for participants aged between 75 and 84 years, and; GFI=7.77 for participants aged 85 years or older; *P*<.001). There are also significant differences in the mean GFI scores according to marital status, with married (3.67) and divorced people (4.62) having lower values while widowed (6.02) or single people (6.13) having higher values (*P*<.001). Frailty is also related to education. We reported that higher frailty scores are associated with lower levels of education, and with the economic level, with greater scores of frailty in people with lower economic levels.

Table 1. Participants' sociodemographic characteristics according to their frailty status: 522 participants were included in this randomized repeated crossover design study.

	Variables, n (%)	Portuguese GFI[a], mean (SD)	*P* value (Kruskal-Wallis test)	Nonfrailty, n (%)	Frailty, n (%)	*P* value (Mann-Whitney *U* test)
Sex			<.001			<.001
Female	367 (70.3)	5.05 (3.04)		127 (34.6)	240 (65.4)	
Male	155 (29.7)	3.43 (2.702)		89 (57.4)	66 (42.6)	
Age (years)			<.001			<.001
65-74	341 (65.3)	3.73 (2.68)		181 (53.1)	160 (46.9)	
75-84	134 (25.7)	5.58 (2.91)		31 (23.1)	103 (76.9)	
85+	47 (9.0)	7.77 (2.86)		4 (8.5)	43 (91.5)	
Marital status			<.001			.003
Single	32 (6.1)	6.13 (3.24)		6 (18.8)	26 (81.3)	
Married	311 (59.6)	3.67 (2.47)		161 (51.8)	150 (48.2)	
Widowed	158 (30.3)	6.02 (3.34)		38 (24.1)	120 (75.9)	
Divorced	21 (4.0)	4.62 (2.94)		11 (52.4)	10 (47.6)	
Education			<.001			<.001
Illiterate	36 (6.9)	7.03 (3.11)		5 (13.9)	31 (86.1)	
Unfinished primary school	80 (15.3)	5.90 (2.95)		19 (23.8)	61 (76.3)	
Primary school	283 (54.2)	4.29 (2.97)		128 (45.2)	155 (54.8)	
High school	92 (17.6)	3.52 (2.32)		47 (51.1)	45 (48.9)	
Higher education	31 (5.9)	4.00 (3.24)		17 (54.8)	14 (45.2)	
Do you think you have enough money to meet your needs?			<.001			<.001
Nothing	12 (2.3)	5.00 (3.77)		6 (50.0)	6 (50.0)	
Little	190 (36.4)	5.89 (3.20)		47 (24.7)	143 (75.3)	
Moderately	309 (59.2)	3.78 (2.63)		158 (51.1)	151 (48.9)	
Quite enough	11 (2.1)	(1.96)		4 (45.5)	6 (54.5)	

[a]GFI: Groningen Frailty Indicator.

According to the GFI frailty classification, using the cutoff point of 4 [16], 58.6% of the people included in this study are frail. Frailty is significantly greater in women (65.4% vs 42.6% in men; *P*<.001), in older people (91.5% in people aged 85 years and older vs 76.9% in people aged 75-84 years vs 46.9% in people aged 65 and 74 years; *P*<.001), and in single people and widows (81.3% and 75.9%, respectively, vs 48.2% in married people and 47.6% in divorced people; *P*<.001). The proportion of frail people is also significantly higher for people with the lowest level of education and at lower economic levels.

Item Descriptive Analysis and Missing Data

Of the 15 indicators, the most prevalent among all people are the items that assess nervousness and anxiety (present in

XSL•FO
RenderX

320/521, 61.4% of the participants), problems in daily life due to poor vision reported by 56% (291/520) of the participants, and polypharmacy reported by 53.6% (279/522; Table 2). The least prevalent items are mobility issues, more specifically, the inability to dress and undress alone (51/521, 9.8%) and to use the bathroom without help (41/522, 7.9%), and the feeling of abandonment (37/522, 7.1%).

Table 2. Groningen Frailty Indicator: missing values and criteria fulfillments: of the 522 participants, questions 3, 4, 14, and 15 had 1 missing value; questions 6 and 9 had 2 missing values; and question 5 had 10 missing values.

Question	Fulfills criteria, n (%)	Missing values, n
1. Shopping	522 (18.4)	0
2. Walking around outside (around the house or to the neighbors)	522 (13.0)	0
3. Dressing and undressing	521 (9.8)	1
4. Going to the toilet	521 (7.9)	1
5. How do you rate your physical fitness? (scale 0-10)	512 (35.5)	10
6. Do you experience problems in daily life due to poor vision?	520 (56.0)	2
7. Do you experience problems in daily life due to being hard of hearing?	522 (35.1)	0
8. During the last 6 months, did you lose a lot of weight unwillingly? (3 kg in 1 month or 6 kg in 2 months)	522 (11.5)	0
9. Do you take 4 or more different types of medicine?	520 (53.6)	2
10. Do you have any complaints about your memory?	522 (35.2)	0
11. Do you sometimes experience an emptiness around you?	522 (39.1)	0
12. Do you sometimes miss people around you?	522 (28.0)	0
13. Do you sometimes feel abandoned?	522 (7.1)	0
14. Did you feel downhearted or sad recently?	521 (46.6)	1
15. Did you feel nervous or anxious recently?	521 (61.4)	1

Of the 522 participants, there are only 18 missing values in 7 of the 15 GFI questions. The average of missing values per participant is 0.03. Only 1 participant has missing values for 2 GFI questions, while the other 17 missing values come from 17 different participants. The question that has the most missing values (1.9%) is the question that asks the participant how he rates his physical fitness (from 0 to 10). The small proportion of missing data, constituting only 0.23% of the total values, suggests that it is improbable for these missing values to significantly impact the overall patterns or conclusions of the analysis, while the large sample size (522 participants) ensures that the study maintains sufficient statistical power to detect meaningful effects or relationships despite the negligible reduction in sample size attributable to missing data.

Psychometric Evaluation of the European Portuguese Version of the GFI (Paper)

Internal Consistency

The obtained Cronbach a value of 0.759 suggests a moderate level of internal consistency and no item was dropped (Table 3) [33].

Temporal Stability

The values varied between 0.504 and 1.000, being an overall value of 0.974 (Table 3).

XSL•FO
RenderX

Table 3. Psychometric Evaluation of the European Portuguese version of the GFI[a]—reliability analysis (internal consistency and temporal stability): no item was dropped and the Cronbach a value of 0.759 suggests a moderate level of internal consistency.

Question	Internal Consistency			Temporal stability, ICC[b]
	Scale average if the item is removed	Scale variation if the item is removed	Cronbach a if the item is removed	
1. Shopping	4.40	7.876	0.729	1.000
2. Walking around outside (around the house or to the neighbors)	4.46	8.189	0.737	0.947
3. Dressing and undressing	4.49	8.362	0.741	1.000
4. Going to the toilet	4.51	8.504	0.744	0.909
5. How do you rate your physical fitness? (scale 0-10)	4.23	7.882	0.740	0.940
6. Do you experience problems in daily life due to poor vision?	4.02	8.438	0.764	1.000
7. Do you experience problems in daily life due to being hard of hearing?	4.24	8.381	0.759	1.000
8. During the last 6 months, did you lose a lot of weight unwillingly? (3 kg in 1 month or 6 kg in 2 months)	4.47	8.952	0.765	1.000
9. Do you take 4 or more different types of medicine?	4.05	8.087	0.751	1.000
10. Do you have any complaints about your memory?	4.24	8.257	0.755	0.911
11. Do you sometimes experience an emptiness around you?	4.20	7.795	0.737	0.958
12. Do you sometimes miss people around you?	4.30	8.014	0.742	0.727
13. Do you sometimes feel abandoned?	4.51	8.683	0.751	0.880
14. Did you feel downhearted or sad recently?	4.13	7.547	0.727	1.000
15. Did you feel nervous or anxious recently?	3.97	8.092	0.749	0.504
Total			0.759	0.974

[a]GFI: Groningen Frailty Indicator.

[b]ICC: intraclass correlation coefficient.

Construct Validity

The suitability of the factor analysis by checking the existence of significant correlations between the items was confirmed by Kaiser-Meyer-Olkin test (0.799) and by Bartlett's Test of Sphericity (χ^2_{105}=2074.7; P<.001). The Kaiser-Meyer-Olkin measure assesses the suitability of data for factor analysis by evaluating proportions of variance in variables potentially linked to underlying factors, with higher values indicating better fit, typically above 0.6 or 0.7. Its advantage lies in providing a simple index for researchers to interpret and make decisions. Bartlett's Test of Sphericity validates the interrelationship of variables necessary for factor analysis, testing the null hypothesis of no correlation between variables, with a significant result (P<.05) confirming the suitability of the dataset for analysis factorial [41].

Evaluation of the scree plot and the size of the eigenvalues strongly suggest that the GFI has a 4D structure, explaining 56.0% of the variance (Table 4). This analysis produced four subscales: (1) Daily Activities and Physical Fitness (items 1-5), (2) Nourishment, (3) Health Problems (items 6, 7, 9, and 10), and (4) Psychosocial Functioning (items 11-15).

XSL•FO
RenderX

Table 4. Psychometric Evaluation of the European Portuguese version of the GFI[a]—validity analysis (construct validity): exploratory factor analysis structure of the Portuguese version of the GFI: loading for each factor and each item in the model with 4 factors after an orthogonal varimax rotation and factor extraction using principal components (n=502): GFI has a 4D structure, explaining 56.0% of the variance.

Items	Factors[b]			
	1	2	3	4
Daily Activities and Physical Fitness				
1. Shopping	0.822	0.167	0.165	–0.009
2. Walking around outside (around the house or to the neighbors)	0.863	0.084	0.087	–0.038
3. Dressing and undressing	0.880	0.076	0.002	0.018
4. Going to the toilet	0.872	0.028	–0.028	0.061
5. How do you rate your physical fitness? (scale 0-10)	0.505	0.179	0.279	–0.034
Health Problems				
6. Do you experience problems in daily life due to poor vision?	–0.008	0.033	0.511	0.325
7. Do you experience problems in daily life due to being hard of hearing?	0.061	–0.073	0.705	0.094
9. Do you take 4 or more different types of medicine?	0.183	0.134	0.563	–0.172
10. Do you have any complaints about your memory?	0.062	0.206	0.465	0.073
Nourishment				
8. During the last 6 months, did you lose a lot of weight unwillingly? (3 kg in 1 month or 6 kg in 2 months)	–0.053	0.041	0.153	0.832
Psychosocial Functioning				
11. Do you sometimes experience an emptiness around you?	0.092	0.817	0.058	–0.007
12. Do you sometimes miss people around you?	0.029	0.778	0.042	0.152
13. Do you sometimes feel abandoned?	0.270	0.432	–0.092	0.368
14. Did you feel downhearted or sad recently?	0.258	0.676	0.190	–0.005
15. Did you feel nervous or anxious recently?	–0.008	0.497	0.389	–0.214
Percentage of total variance explained for each factor	N/A[c]	13.50	8.66	6.67

[a]GFI: Groningen Frailty Indicator.

[b]Percentage of total variance explained for the 4 extracted factors=56%; Kaiser-Meyer-Olkin=0.799 and Bartlett's Test of Sphericity <0.001.

[c]N/A: not applicable.

Psychometric Evaluation of the FRAILSURVEY—an mHealth Tool for Frailty Assessment

Interinstrument Reliability

The values varied between 0.100 and 0.634, being an overall value of 0.790 (Table 5). From the 15 items, 2 items had a substantial agreement, 5 had a moderate agreement, 6 had a fair agreement, 2 had none to a slight agreement, and the overall score had an almost perfect agreement.

XSL•FO
RenderX

Table 5. Psychometric Evaluation of the FRAILSURVEY—reliability analysis (interinstrument reliability and temporal stability)a.

Question	Interinstrument reliability				Temporal stability, ICC[b]
	SE values	Observed k values	Lower limit (95% CI)	Upper limit (95% CI)	
1. Shopping	0.045	0.634	0.546	0.722	0.793
2. Walking around outside (around the house or to the neighbors)	0.053	0.608	0.503	0.712	0.867
3. Dressing and undressing	0.069	0.522	0.387	0.657	0.958
4. Going to the toilet	0.079	0.468	0.314	0.622	0.875
5. How do you rate your physical fitness? (scale 0-10)	0.043	0.419	0.335	0.503	0.788
6. Do you experience problems in daily life due to poor vision?	0.042	0.266	0.182	0.349	0.958
7. Do you experience problems in daily life due to being hard of hearing?	0.045	0.289	0.201	0.378	0.798
8. During the last 6 months, did you lose a lot of weight unwillingly? (3 kg in 1 month or 6 kg in 2 months)	0.086	0.100	0.000	0.270	0.958
9. Do you take 4 or more different types of medicine?	0.042	0.337	0.256	0.419	0.913
10. Do you have any complaints about your memory?	0.045	0.236	0.147	0.324	0.724
11. Do you sometimes experience an emptiness around you?	0.041	0.454	0.374	0.533	0.857
12. Do you sometimes miss people around you?	0.047	0.336	0.245	0.428	0.667
13. Do you sometimes feel abandoned?	0.099	0.188	0.000	0.381	0.793
14. Did you feel downhearted or sad recently?	0.039	0.443	0.366	0.520	0.718
15. Did you feel nervous or anxious recently?	0.043	0.328	0.243	0.413	0.589
GFI[c] score	0.027	0.790	0.737	0.844	0.800

[a]Evaluation of the app-based European Portuguese GFI indicates good reliability (interinstrument reliability: Cohen k=0.790; temporal stability: ICC=0.800).

[b]ICC: intraclass correlation coefficient.

[c]GFI: Groningen Frailty Indicator.

Temporal Stability

The retest was administered 2 weeks after the initial contact with the participant to evaluate the temporal stability of the measurements. The ICC values varied between 0.589 and 0.958, being an overall value of 0.800 (Table 5).

Concurrent Validity

All values were positive, ranging from 0.101 to 0.634, and the total score was 0.694 (Table 6). These results demonstrate a positive correlation between responses in both formats, paper, and app.

XSL•FO
RenderX

Table 6. Psychometric evaluation of the FRAILSURVEY—validity analysis (concurrent validity): a positive correlation between responses in both formats, paper, and app was observed.

Question	N	FRAILSURVEY, mean (SD)	Paper, mean (SD)	*r* value	*P* value
1. Shopping	522	0.20 (0.40)	0.18 (0.39)	0.634	<.001
2. Walking around outside (around the house or to the neighbors)	522	0.15 (0.36)	0.13 (0.34)	0.609	<.001
3. Dressing and undressing	521	0.10 (0.30)	0.10 (0.30)	0.522	<.001
4. Going to the toilet	521	0.09 (0.28)	0.08 (0,27)	0.468	<.001
5. How do you rate your physical fitness? (scale 0-10)	512	0.35 (0.48)	0.36 (0.48)	0.419	<.001
6. Do you experience problems in daily life due to poor vision?	520	0.52 (0.50)	0.56 (0.50)	0.266	<.001
7. Do you experience problems in daily life due to being hard of hearing?	522	0.33 (0.47)	0.35 (0.48)	0.290	<.001
8. During the last 6 months, did you lose a lot of weight unwillingly? (3 kg in 1 month or 6 kg in 2 months)	522	0.10 (0.30)	0.11 (0.32)	0.101	.02
9. Do you take 4 or more different types of medicine?	520	0.55 (0.50)	0.54 (0.50)	0.337	<.001
10. Do you have any complaints about your memory?	522	0.38 (0.49)	0.35 (0.48)	0.236	<.001
11. Do you sometimes experience an emptiness around you?	522	0.38 (0.49)	0.39 (0.49)	0.454	<.001
12. Do you sometimes miss people around you?	522	0.32 (0.47)	0.28 (0.45)	0.338	<.001
13. Do you sometimes feel abandoned?	522	0.08 (0.28)	0.07 (0.26)	0.189	<.001
14. Did you feel downhearted or sad recently?	521	0.50 (0.50)	0.47 (0.50)	0.444	<.001
15. Did you feel nervous or anxious recently?	521	0.64 (0.48)	0.61 (0.49)	0.329	<.001
Total score	522	4.67 (3.01)	4.57 (3.04)	0.694	<.001

Discussion

Principal Findings

Translating and validating frailty assessment tools are vital for cross-cultural research, ensuring inclusivity and diversity, and allowing researchers and health care professionals to identify frailty in older adults independently or irrespective of linguistic and cultural backgrounds [42]. Translated versions enhance measurement accuracy and consistency by retaining the psychometric properties and measurement validity. This facilitates the development of comprehensive public health policies and targeted interventions, improving care and quality of life for ageing populations worldwide [43]. Overall, translating and validating frailty assessment tools are crucial in advancing research, clinical practice, and public health initiatives related to frailty [44]. Frailty assessment tools are valuable for identifying at-risk individuals and tailoring interventions to their needs. By identifying frailty early, health care practitioners can develop strategies to prevent functional decline and mitigate adverse health outcomes [45].

In this study, the translation of the GFI into European Portuguese underwent rigorous processes, resulting in a final version with strong psychometric properties. The translated version proved to be clear and easily understandable.

Among the 522 participants in the study, the majority were women, which aligns with typical demographic patterns in this type of research. In addition, most participants were aged 65-74 years, were either married or widowed, had completed only primary education, and had moderate money to meet their needs.

The mean GFI score was 4.57 in this study, whereas validation studies conducted in China, Romania, and Lebanon were 4.60, 5.68, and 6.80, respectively [17,18,46]. Consistent with previous research, the average score was significantly higher in women than in men [5,12,17,46]. As anticipated, there was a significant positive correlation between age and GFI score and, additionally, marital status showed significant differences, with singles and widowers having higher scores [47]. Frailty also demonstrated significant associations with lower education levels and lower economic statuses, as already shown [48].

The GFI comprises 15 frailty indicators and the most prevalent were polypharmacy, vision problems, and feelings of anxiety or nervousness, all reported by more than 50% of the participants. Indeed, the link between polypharmacy and frailty has been established through a study using SHARE data, which explored the connection between mortality, frailty, and polypharmacy across 18 countries. According to these findings, 59.4% of the identified frail population was found to be taking multiple medications, a percentage that closely aligns with the results obtained in this study, where 53.6% of the population were under polypharmacy [5]. Frailty and poor vision are interconnected conditions with a bidirectional relationship, as already described [49]. Poor vision can contribute to frailty by impairing balance, mobility, and daily functioning, increasing the risk of falls and social isolation. It has already been described that people with visual impairment have 3.7 greater odds of being frail [49]. Anxiety symptoms, reported by 61.4% (320/521) of the participants of this study, were reported to be notably more prevalent among older adults with frailty than among those who are robust, which has also been reported elsewhere [50].

XSL•FO
RenderX

The psychometric evaluation of the European Portuguese version of the GFI demonstrated good reliability and validity. In previous psychometric studies assessing the internal consistency of the GFI, Cronbach α values ranged from 0.680 to 0.810. Nevertheless, in this study, the obtained value was 0.759, which is similar to Cronbach α value of the original scale [32]. Internal consistency maintained the same value even if any item was deleted. The assessment of reliability over time involved a retest conducted 2 weeks after the initial contact, and the ICC was used for calculation. The ICC values ranged from 0.504 to 1.000, with an overall value of 0.974. This value surpassed the reliability values obtained in 2 comparable studies conducted on the Chinese population (0.950 and 0.939) [18,51]. These results demonstrate that the European Portuguese version of the GFI exhibits excellent reliability over time, confirming its consistency as a measurement tool. The factor analysis yielded the same number of factors as the original scale, indicating that the translated version maintains a similar underlying factor structure to the original scale, suggesting that it measures the same fundamental constructs or dimensions as the items in the original version [52]. This encouraging finding indicates that the translation process effectively retained the core structure of the scale. Moreover, this confirmation of reliability and validity further demonstrates that the translated version of the GFI effectively preserves the essential elements of the original instrument.

The general hypothesis of this study was that a smartphone GFI app would show the same validity and reliability as the paper version, and it was proved in this study. The interinstrument reliability was studied to investigate the agreement and consistency between the 2 instruments, thereby ensuring the robustness of the results. High interinstrument reliability strengthens the credibility and accuracy of research outcomes, contributing to the overall quality and validity of assessments and evaluations. In this study, high levels of agreement were recorded between the instruments, proving that the digital format of the questionnaire is as reliable as the paper version. The reliability assessment over time was done in the same way as for the paper version of the GFI. The ICC values ranged from 0.589 to 0.958, with an overall value of 0.800, evidencing that, just like on paper, the digital version of the GFI remained reliable over time. The validity of the application was assessed by comparing the answers and total scores of the same individuals in the paper format (conventional) and the application (the new format to be validated). The results revealed a robust, positive, and statistically significant relationship between the responses in both forms. These findings not only establish the reliability of the application but also demonstrate its validation.

Health apps are vital in modern health care, providing easy access to health information, monitoring tools, and personalized advice [53]. The importance of mHealth becomes clear when examining its impact on health care accessibility. Many people, especially those living in disadvantaged or remote regions, face significant challenges in accessing health services. mHealth serves as a powerful equalizer, enabling individuals to connect with health care professionals, get immediate medical advice, and access health-related information, regardless of their geographic location [54]. This level of accessibility empowers patients and fosters early intervention and preventative care, ultimately resulting in improved health outcomes [55]. However, the proliferation of health apps on smartphone-based app stores is still on the rise due to the ease of app development, the absence of regulation, and the direct marketing to users, leading to an overcrowded scenario, which lacks reliability and accuracy. However, the ability of health stakeholders to discern the quality of these apps still lags significantly. This challenge largely stems from the lack of validation, which hinders app reliability and validity assurance [56]. Some potential benefits of validating health apps include improved accuracy and reliability of the app, increased user confidence in the app, and improved safety and effectiveness [57,58].

Assessing frailty is crucial as it helps identify individuals at higher risk of adverse health outcomes and enables timely interventions to mitigate these risks. Moreover, frailty assessments may guide treatment decisions, leading to tailored health care approaches focused on maintaining independence and managing chronic conditions. The impact of frailty in individuals and communities is profound, increasing health care usage, decreasing quality of life, and adding strain on families and caregivers. Understanding these effects allows for prioritized interventions, effective resource allocation, and appropriate support, ultimately improving the well-being and outcomes of frail individuals. The advantages of using apps for frailty assessment include their ease of use, ability to provide real-time feedback to the user, and their potential for improving the accuracy and reliability of frailty assessment [59,60]. An app designed to assess frailty offers numerous advantages and can revolutionize how frailty is identified. Its accessibility on smartphones and digital devices ensures easy and convenient frailty assessment, even in remote areas, while enabling early detection of an individual's physical and cognitive abilities, allowing for timely interventions.

However, a key challenge in implementing mHealth solutions for frailty is the digital divide. Frail individuals, especially those from disadvantaged backgrounds, may lack access to necessary technology. This gap exacerbates health inequalities, as those in need of mHealth interventions the most may have the least access to them [61]. Ensuring fair access to technology and mHealth tools is a critical ethical issue [62]. Creating mHealth apps and devices that are easy to use and accessible to frail individuals is essential. This involves overcoming challenges related to age-related impairments in vision, hearing, and dexterity, which were considered in the development of the FRAILSURVEY [63].

The advancements in mHealth for frailty care hold the potential for deeply personalized interventions. Using cutting-edge data analytics and artificial intelligence, mHealth platforms will be adept at customizing interventions according to the distinct needs and characteristics of each frail individual [64]. The integration of mHealth platforms with electronic health records holds the potential to enhance care coordination and facilitate seamless data sharing. This interoperability will streamline processes, improve communication, and, ultimately, raise the standard of service. Furthermore, the evolving landscape of mHealth in frailty care will require ongoing ethical dialogues.

XSL•FO
RenderX

As technological capabilities progress, ethical dilemmas around privacy, data security, informed consent, and fairness in algorithms will persist as paramount concerns. Navigating the delicate balance between innovation and ethical protection will present a constant challenge [65,66].

This study has some limitations that should be highlighted. Despite using a randomized, repeated crossover design and recruiting 522 participants from diverse settings, including community-dwelling older adults and nursing home residents, limitations persisted, only a fraction of participants underwent test-retest analysis, reducing the robustness of the results. In addition, the digital literacy barrier among older adults may have impacted the accessibility of the smartphone-based app, FRAILSURVEY. Furthermore, biases inherent in translating and validating scales into different languages may have influenced the study results. However, the similarity of psychometric properties to previous validation studies and the pioneering demonstration of the validity of the smartphone version offer promising insights into frailty assessment methods. This study marks a significant contribution to the field by establishing the scientific validity of the smartphone version of the GFI.

Conclusions

Overall, FRAILSURVEY uses a standardized and validated assessment tool, providing objective and consistent measurements, eliminating subjective biases, and enhancing accuracy and reliability. Compared with traditional in-person assessments, its cost-effectiveness can reduce health care expenses and increase accessibility. The future integration with electronic health records and other health-related data could enable personalized interventions tailored to an individual's specific needs and medical history, ultimately leading to more effective frailty management and improved overall health outcomes for at-risk individuals.

Acknowledgments

This work is financed by national funds from FCT-Fundação para a Ciência e a Tecnologia, I.P., in the scope of 2 projects (UIDP/04378/2020 and UIDB/04378/2020) of the Research Unit on Applied Molecular Biosciences-UCIBIO and a project (LA/P/0140/2020) of the Associate Laboratory Institute for Health and Bioeconomy-i4HB, and under the grant attributed to Luís Midão (SFRH/BD/137090/2018). Also, the authors want to thank all participants of the study.

Data Availability

The dataset generated during this study is not publicly accessible due to ethical considerations. However, both the dataset and the study metadata can be obtained from the corresponding author upon request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Final version of the European Portuguese version of the Groningen Frailty Indicator.
[DOCX File , 33 KB - formative_v9i1e51975_app1.docx]

References

1. Costa A, Henriques J, Alarcão V, Henriques A, Madeira T, Virgolino A, et al. Active aging awareness and well-being among older adults in portugal. Front Public Health 2023;11:1149731 [FREE Full text] [doi: 10.3389/fpubh.2023.1149731] [Medline: 37124788]
2. Instituto Nacional de Estatística. Censos de 2011. 2012. 2012. URL: https://censos.ine.pt/xportal/xmain?xpid=CENSOS&xpgid=censos2011_apresentacao [accessed 2025-01-25]
3. Instituto Nacional de EstatísticaI. Projeções de População Residente. 2020. URL: https://www.cpidoso.pt/populacao/ [accessed 2020-03-20]
4. Instituto Nacional de Estatística, I. Censos de 2021. 2022. URL: https://tinyurl.com/6b4b6usy [accessed 2025-01-25]
5. Midão L, Brochado P, Almada M, Duarte M, Paúl C, Costa E. Frailty status and polypharmacy predict all-cause mortality in community dwelling older adults in Europe. Int J Environ Res Public Health 2021;18(7):3580 [FREE Full text] [doi: 10.3390/ijerph18073580] [Medline: 33808273]
6. Tessier L, Wulf N, Momose Y. Long-term care in the context of population ageing: a rights-based approach to universal coverage. ILO Working Paper 2023 [FREE Full text] [doi: 10.54394/rhjd4529]
7. Cesari M, Prince M, Thiyagarajan JA, De Carvalho IA, Bernabei R, Chan P, et al. Frailty: an emerging public health priority. J Am Med Dir Assoc 2016;17(3):188-192. [doi: 10.1016/j.jamda.2015.12.016] [Medline: 26805753]
8. de Souto Barreto P, Morley JE, Chodzko-Zajko W, H Pitkala K, Weening-Djiksterhuis E, Rodriguez-Mañas L, International Association of GerontologyGeriatrics—Global Aging Research Network (IAGG-GARN)the IAGG European Region Clinical Section. Recommendations on physical activity and exercise for older adults living in long-term care facilities: a taskforce report. J Am Med Dir Assoc 2016;17(5):381-392. [doi: 10.1016/j.jamda.2016.01.021] [Medline: 27012368]

XSL•FO
RenderX

9. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001;56(3):M146-M156. [doi: 10.1093/gerona/56.3.m146] [Medline: 11253156]
10. Kojima G. Frailty as a predictor of future falls among community-dwelling older people: a systematic review and meta-analysis. J Am Med Dir Assoc 2015;16(12):1027-1033. [doi: 10.1016/j.jamda.2015.06.018] [Medline: 26255098]
11. Vermeiren S, Vella-Azzopardi R, Beckwée D, Habbig A, Scafoglieri A, Jansen B, Gerontopole Brussels Study group. Frailty and the prediction of negative health outcomes: a meta-analysis. J Am Med Dir Assoc 2016;17(12):1163.e1-1163.e17. [doi: 10.1016/j.jamda.2016.09.010] [Medline: 27886869]
12. Manfredi G, Midão L, Paúl C, Cena C, Duarte M, Costa E. Prevalence of frailty status among the European elderly population: findings from the survey of health, aging and retirement in Europe. Geriatrics Gerontology Int 2019;19(8):723-729. [doi: 10.1111/ggi.13689]
13. Daniels R, van Rossum E, de Witte L, Kempen GI, van den Heuvel W. Interventions to prevent disability in frail community-dwelling elderly: a systematic review. BMC Health Serv Res 2008;8:278 [FREE Full text] [doi: 10.1186/1472-6963-8-278] [Medline: 19115992]
14. de Vries NM, Staal J, van Ravensberg C, Hobbelen J, Olde Rikkert M, Nijhuis-van der Sanden M. Outcome instruments to measure frailty: a systematic review. Ageing Res Rev 2011;10(1):104-114. [doi: 10.1016/j.arr.2010.09.001] [Medline: 20850567]
15. Peters LL, Boter H, Buskens E, Slaets JP. Measurement properties of the groningen frailty indicator in home-dwelling and institutionalized elderly people. J Am Med Dir Assoc 2012;13(6):546-551 [FREE Full text] [doi: 10.1016/j.jamda.2012.04.007] [Medline: 22579590]
16. Steverink N, Slaets J, Schuurmans H, Van Lis M. Measuring frailty: developing and testing the GFI (Groningen Frailty Indicator). Gerontologist 2001;41(1):236-237.
17. Olaroiu M, Ghinescu M, Naumov V, Brinza I, Heuvel WVD. The psychometric qualities of the Groningen Frailty Indicator in Romanian community-dwelling old citizens. Fam Pract 2014;31(4):490-495. [doi: 10.1093/fampra/cmu022] [Medline: 24846305]
18. Xiang W, Cheng Y, Li Z, Han J, Li K. Cross-cultural adaptation and validation of the Groningen Frailty Indicator in Chinese nursing home residents. Aging Clin Exp Res 2020;32(6):1035-1042. [doi: 10.1007/s40520-019-01178-7] [Medline: 30900215]
19. Oetsma S, Boonen A, Starmans M, Peeters R, van Onna M. Validation of two frailty questionnaires in older patients with rheumatoid arthritis: a cross-sectional study. Clin Exp Rheumatol 2020;38(3):523-528 [FREE Full text] [Medline: 31694746]
20. Meessen JMTA, Leichtenberg CS, Tilbury C, Kaptein BL, Koster LA, Slagboom PE, et al. Frailty in end-stage hip or knee osteoarthritis: validation of the Groningen Frailty Indicator (GFI) questionnaire. Rheumatol Int 2018;38(5):917-924 [FREE Full text] [doi: 10.1007/s00296-017-3868-1] [Medline: 29147763]
21. Bras L, Peters T, Wedman J, Plaat B, Witjes M, van Leeuwen B, et al. Predictive value of the Groningen Frailty Indicator for treatment outcomes in elderly patients after head and neck, or skin cancer surgery in a retrospective cohort. Clin Otolaryngol 2015;40(5):474-482. [doi: 10.1111/coa.12409] [Medline: 25754107]
22. Castro MDL, Alves M, Martins A, Luísa Papoila A, Botelho M, Fragata J. Reproducibility and validity of the Portuguese Edmonton Frail Scale version in cardiac surgery patients. Rev Port Cardiol 2023;42(4):295-304 [FREE Full text] [doi: 10.1016/j.repc.2022.03.009] [Medline: 36738965]
23. Santiago LM, Silva R, Velho D, Rosendo I, Simões J. Cross-cultural adaptation and validation of the PRISMA-7 Scale for European Portuguese. Fam Med Prim Care Rev 2020;22(1):59-66. [doi: 10.5114/fmpcr.2020.92507]
24. Faria A, Sousa-Santos AR, Mendes J, Limas de Sousa AS, Amaral TF. Development of the Portuguese versions of FRAIL Scale and SARC-F: screening tools for physical frailty and sarcopenia (Desenvolvimento das versões portuguesas dos questionários FRAIL Scale e SARC-F: ferramentas de rastreio para a fragilidade física e sarcopenia). Acta Port Nutr 2021;26:90-94. [doi: 10.21011/apn.2021.2614]
25. Pereira Pinto M, Martins S, Mesquita E, Fernandes L. European Portuguese version of the Clinical Frailty Scale: translation, cultural adaptation and validation study. Acta Med Port 2021;34(11):749-760. [doi: 10.20344/amp.14543] [Medline: 33565417]
26. Cardoso AF, Bobrowicz-Campos E, Teixeira-Santos L, Cardoso D, Couto F, Apóstolo J. Validation and screening capacity of the European Portuguese version of the SUNFRAIL tool for community-dwelling older adults. Int J Environ Res Public Health 2021;18(4):1394 [FREE Full text] [doi: 10.3390/ijerph18041394] [Medline: 33546251]
27. Coelho T, Santos R, Paúl C, Gobbens RJ, Fernandes L. Portuguese version of the Tilburg Frailty Indicator: transcultural adaptation and psychometric validation. Geriatr Gerontol Int 2015;15(8):951-960. [doi: 10.1111/ggi.12373] [Medline: 25255891]
28. Smith MJ, Reiter MJ, Crist BD, Schultz LG, Choma TJ. Improving patient satisfaction through computer-based questionnaires. Orthopedics 2016;39(1):e31-e35. [doi: 10.3928/01477447-20151218-07] [Medline: 26709557]
29. Ahn DT. Benefits and risks of apps for patients. Curr Opin Endocrinol Diabetes Obes 2022;29(1):17-22. [doi: 10.1097/MED.0000000000000701] [Medline: 34878413]

XSL•FO
RenderX

30. Lecomte T, Potvin S, Corbière M, Guay S, Samson C, Cloutier B, et al. Mobile apps for mental health issues: meta-review of meta-analyses. JMIR Mhealth Uhealth 2020;8(5):e17458 [FREE Full text] [doi: 10.2196/17458] [Medline: 32348289]
31. Pagliari C, Sloan D, Gregor P, Sullivan F, Detmer D, Kahan JP, et al. What is eHealth (4): a scoping exercise to map the field. J Med Internet Res 2005;7(1):e9 [FREE Full text] [doi: 10.2196/jmir.7.1.e9] [Medline: 15829481]
32. Schuurmans H, Steverink N, Lindenberg S, Frieswijk N, Slaets JPJ. Old or frail: what tells us more? J Gerontol A Biol Sci Med Sci 2004;59(9):M962-M965. [doi: 10.1093/gerona/59.9.m962] [Medline: 15472162]
33. Bielderman A, van der Schans CP, van Lieshout MJ, de Greef MH, Boersma F, Krijnen WP, et al. Multidimensional structure of the Groningen Frailty Indicator in community-dwelling older people. BMC Geriatr 2013;13:86 [FREE Full text] [doi: 10.1186/1471-2318-13-86] [Medline: 23968433]
34. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, ISPOR Task Force for Translation and Cultural Adaptation. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health 2005;8(2):94-104 [FREE Full text] [doi: 10.1111/j.1524-4733.2005.04054.x] [Medline: 15804318]
35. Kim JH, Kwon S, Shim SR, Sun HY, Ko YM, Chun D, et al. Validation and reliability of a smartphone application for the International Prostate Symptom Score questionnaire: a randomized repeated measures crossover study. J Med Internet Res 2014;16(2):e38. [doi: 10.2196/jmir.3042] [Medline: 24513507]
36. Streiner DL, Norman GR, Cairney. Health Measurement Scales: A Practical Guide to Their Development and Use. 5th ed. United Kingdom: Oxford University Press; 2015.
37. Cappelleri JC, Zou KH, Bushmakin AG, J. Alvir JM, Alemayehu D, Symonds T. Patient-Reported Outcomes: Measurement, Implementation and Interpretation. 1st ed. New York, NY: Chapman and Hall/CRC; 2014.
38. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
39. Kaiser HF. Computer program for varimax rotation in factor analysis. Educ Psychol Meas 1959;19(3):413-420. [doi: 10.1177/001316445901900314]
40. Marston L. Introductory Statistics for Health and Nursing Using SPSS. London: Sage; 2009.
41. Shrestha N. Factor analysis as a tool for survey analysis. Am J Appl Math Stat 2021;9(1):4-11. [doi: 10.12691/ajams-9-1-2]
42. Sperber AD. Translation and validation of study instruments for cross-cultural research. Gastroenterology 2004;126(1 Suppl 1):S124-S128. [doi: 10.1053/j.gastro.2003.10.016] [Medline: 14978648]
43. Dambi JM, Corten L, Chiwaridzo M, Jack H, Mlambo T, Jelsma J. A systematic review of the psychometric properties of the cross-cultural translations and adaptations of the Multidimensional Perceived Social Support Scale (MSPSS). Health Qual Life Outcomes 2018;16(1):80 [FREE Full text] [doi: 10.1186/s12955-018-0912-0] [Medline: 29716589]
44. Donoghue OA, O'Shea M, McGarrigle CA, Lang D, Hoey C, Whitty H, et al. Translating frailty assessment methodologies and research-based evidence to clinical education and practice. J Aging Soc Policy 2022;34(4):537-551. [doi: 10.1080/08959420.2020.1777825] [Medline: 32634336]
45. Abbasi M, Rolfson D, Khera AS, Dabravolskaj J, Dent E, Xia L. Identification and management of frailty in the primary care setting. CMAJ 2018;190(38):E1134-E1140 [FREE Full text] [doi: 10.1503/cmaj.171509] [Medline: 30249759]
46. Khamis R, Sabbah H, Sabbah S, Peters L, Droubi N, Sabbah I. Evaluating the psychometric properties of the Arabic version of the Groningen Frailty Indicator among Lebanese elderly people. J Egypt Public Health Assoc 2019;94(1):28 [FREE Full text] [doi: 10.1186/s42506-019-0028-3] [Medline: 32813111]
47. Garre-Olmo J, Calvó-Perxas L, López-Pousa S, Gracia Blanco, MD, Vilalta-Franch J. Prevalence of frailty phenotypes and risk of mortality in a community-dwelling elderly cohort. Age Ageing 2013;42(1):46-51. [doi: 10.1093/ageing/afs047]
48. Wang J, Hulme C. Frailty and socioeconomic status: a systematic review. J Public Health Res 2021;10(3):2036 [FREE Full text] [doi: 10.4081/jphr.2021.2036] [Medline: 33942603]
49. Varadaraj V, Lee MJ, Tian J, Ramulu PY, Bandeen-Roche K, Swenor BK. Near vision impairment and frailty: evidence of an association. Am J Ophthalmol 2019;208:234-241 [FREE Full text] [doi: 10.1016/j.ajo.2019.08.009] [Medline: 31465753]
50. Tan M, Bhanu C, Frost R. The association between frailty and anxiety: a systematic review. Int J Geriatr Psychiatry 2023;38(5):e5918. [doi: 10.1002/gps.5918] [Medline: 37157226]
51. Luh RL, Yu, SY, Yang CH. Chinese version of Groningen Frailty Index: validation in a Chinese elderly population. Neuropsychiatry 2018;8(6):1949-1956.
52. Krabbe PF. Chapter 7—Validity. In: The Measurement of Health and Health Status. San Diego, CA: Academic Press; 2017:113-134.
53. Maaß L, Freye M, Pan C, Dassow H, Niess J, Jahnel T. The definitions of health apps and medical apps from the perspective of public health and law: qualitative analysis of an interdisciplinary literature overview. JMIR Mhealth Uhealth 2022;10(10):e37980 [FREE Full text] [doi: 10.2196/37980] [Medline: 36315221]
54. Wilson K. Mobile cell phone technology puts the future of health care in our hands. CMAJ 2018;190(13):E378-E379 [FREE Full text] [doi: 10.1503/cmaj.180269] [Medline: 29615420]
55. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3:4 [FREE Full text] [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]

XSL•FO
RenderX

56. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020;27(2):330-340 [FREE Full text] [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
57. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3:14 [FREE Full text] [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
58. Lagan S, Sandler L, Torous J. Evaluating evaluation frameworks: a scoping review of frameworks for assessing health apps. BMJ Open 2021;11(3):e047001 [FREE Full text] [doi: 10.1136/bmjopen-2020-047001] [Medline: 33741674]
59. Chang R, Low H, McDonald A, Park G, Song X. Web-based software applications for frailty assessment in older adults: a scoping review of current status with insights into future development. BMC Geriatr 2021;21(1):723 [FREE Full text] [doi: 10.1186/s12877-021-02660-6] [Medline: 34922466]
60. Morley JE, Vellas B, van Kan A, Anker SD, Bauer JM, Bernabei R, et al. Frailty consensus: a call to action. J Am Med Dir Assoc 2013;14(6):392-397 [FREE Full text] [doi: 10.1016/j.jamda.2013.03.022] [Medline: 23764209]
61. Mugueta-Aguinaga I, Garcia-Zapirain B. Is technology present in frailty? Technology a back-up tool for dealing with frailty in the elderly: a systematic review. Aging Dis 2017;8(2):176-195 [FREE Full text] [doi: 10.14336/AD.2016.0901] [Medline: 28400984]
62. Brall CP, Schröder-Bäck P, Maeckelberghe E. Ethical aspects of digital health from a justice point of view. Eur J Public Health 2019;29(Suppl 3):18-22 [FREE Full text] [doi: 10.1093/eurpub/ckz167] [Medline: 31738439]
63. Tajudeen FP, Bahar N, Tan MP, Peer Mustafa MB, Saedon NI, Jesudass J. Understanding user requirements for a senior-friendly mobile health application. Geriatrics (Basel) 2022;7(5):110 [FREE Full text] [doi: 10.3390/geriatrics7050110] [Medline: 36286212]
64. Bhatt P, Liu J, Gong Y, Wang J, Guo Y. Emerging artificial intelligence-empowered mHealth: scoping review. JMIR Mhealth Uhealth 2022;10(6):e35053 [FREE Full text] [doi: 10.2196/35053] [Medline: 35679107]
65. Jacobson NC, Bentley KH, Walton A, Wang SB, Fortgang RG, Millner AJ, et al. Ethical dilemmas posed by mobile health and machine learning in psychiatry research. Bull World Health Organ 2020;98(4):270-276 [FREE Full text] [doi: 10.2471/BLT.19.237107] [Medline: 32284651]
66. Zhang X, Saltman R. Impact of electronic health record interoperability on telehealth service outcomes. JMIR Med Inform 2022;10(1):e31837 [FREE Full text] [doi: 10.2196/31837] [Medline: 34890347]

Abbreviations

GFI: Groningen Frailty Indicator
ICC: intraclass correlation coefficient
mHealth: mobile health
PRISMA-7: Program on Research for Integrating Services for the Maintenance of Autonomy questionnaire

Edited by A Mavragani; submitted 21.08.23; peer-reviewed by S Pandey, H Veldandi, D Portela, E Prochaska, R Nooripour; comments to author 21.03.24; revised version received 09.04.24; accepted 11.04.24; published 07.03.25.

Please cite as:
Midao L, Duarte M, Sampaio R, Almada M, Dias CC, Paúl C, Costa E
FRAILSURVEY—an mHealth App for Self-Assessment of Frailty Based on the Portuguese Version of the Groningen Frailty Indicator: Validation and Reliability Study
JMIR Form Res 2025;9:e51975
URL: https://formative.jmir.org/2025/1/e51975
doi:10.2196/51975
PMID:40053720

©Luis Midao, Mafalda Duarte, Rute Sampaio, Marta Almada, Cláudia Camila Dias, Constança Paúl, Elísio Costa. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluation of the MyFertiCoach Lifestyle App for Subfertile Couples: Single-Center Evaluation of Augmented Standard Care

Jesper Smeenk[1], MD, PhD; Ellen Smit[2], BBA; Marc Jacobs[3], PhD; Ilse van Rooij[1], MD, PhD

[1]Elisabeth TweeSteden Ziekenhuis, Tilburg, The Netherlands

[2]Ferring BV, Hoofddorp, The Netherlands

[3]MSJ Advies, Helmond, The Netherlands

Corresponding Author:
Jesper Smeenk, MD, PhD
Elisabeth TweeSteden Ziekenhuis
Doctor Deelenlaan 5
Tilburg, 5042 AD
The Netherlands
Phone: 31 (013) 221 00 00
Email: j.smeenk@etz.nl

Abstract

Background: Many couples undergoing fertility treatment face multiple lifestyle risk factors that lower their chances of achieving pregnancy. The MyFertiCoach (MFC) app was designed as an integrated lifestyle program featuring modules on healthy weight management, nutrition, exercise, quitting smoking, reducing alcohol and drug use, and managing stress. We hypothesized that supplementing standard care with the MFC app would improve lifestyle outcomes.

Objective: This study aims to assess the impact of the MFC app on changing multiple lifestyle habits in women seeking fertility treatment. The primary outcome is the change in the total risk score (TRS) at 3- and six-month follow-ups. The TRS is calculated for each individual as the sum of all risk scores per behavior (eg, vegetable/fruit/folic acid intake, smoking, and alcohol use) at 3 and 6 months. A higher TRS indicates unhealthier nutrition and lifestyle habits and a lower likelihood of achieving pregnancy. The secondary endpoints include changes in BMI, activity score, preconception dietary risk score, distress score (eg, perceived burden), smoking habits, alcohol intake, and program adherence.

Methods: This retrospective, observational, single-center evaluation included patients between January 1, 2022, and December 31, 2023. Subfertile female patients aged 18-43 years and their partners, who were referred to a gynecologist, were invited to participate in online lifestyle coaching via the MFC app. The gynecologist selected relevant lifestyle modules based on the results of integrated screening questionnaires. We used (hierarchical) linear mixed models (LMMs) to estimate changes in outcomes. For missing data patterns deemed missing not at random, joint modeling was applied. Statistical significance was set at $P \leq .05$, with methods in place to maintain the same false-positive rate.

Results: A total of 1805 patients were invited to participate in the evaluation, with an average of 737 (40.83%) completing the screening questionnaire at baseline. For the TRS, 798 (44.21%) patients were included at baseline, of whom 517 (64.8%) involved their partner. On average, 282 of 744 (37.9%) patients submitted at least one follow-up questionnaire. Patients rated the app above average (n=137, median score of 7 on a 1-10 scale) on days 7 and 14. The TRS decreased by an average of 1.5 points ($P<.001$) at T3 and T6 compared with baseline, a clinically meaningful improvement. All secondary outcomes showed statistically significant positive changes for patients who used a relevant lifestyle module ($P<.001$). Most improvements were achieved by 3 months and remained significant at 6 months ($P<.001$), except for alcohol intake ($P<.53$). These findings were consistent across both LMMs and joint models.

Conclusions: Our evaluation of a mobile health app integrated into standard care demonstrates immediate and clinically meaningful improvements in key lifestyle parameters among women seeking to become pregnant. Additional scientific research is needed to identify the causal pathways leading to sustained effectiveness. To maintain and enhance these outcomes, further tailoring of patient-specific programs is essential.

(JMIR Form Res 2025;9:e64239) doi:10.2196/64239

XSL•FO
RenderX

KEYWORDS

fertility; mHealth; pregnancy; lifestyle; app; smartphone

Introduction

Many couples undergoing fertility treatment have multiple lifestyle risk factors, such as obesity, unbalanced nutritional habits, low physical activity, and alcohol consumption or smoking. These habits not only increase the risk of cardiovascular and metabolic diseases but also reduce the chance of becoming pregnant [1-4]. Reducing any or all of these lifestyle factors before pregnancy or assisted reproductive technology treatment can improve reproductive health [5-9]. In summary, the more negative lifestyle factors present, the lower the chance of becoming pregnant and the longer the time until pregnancy [10].

For instance, improving dietary patterns during the preconception period reduces the risk of several adverse birth outcomes, such as fetal growth restriction and babies born small or large for gestational age. Maternal complications, including gestational diabetes, hypertensive disorders, and premature delivery, also appear to decrease [11,12].

Moreover, cognitive behavioral therapy and psychological support during fertility treatment have been suggested to result in significantly more viable pregnancies compared with routine care [13-15].

While the evidence supporting the importance of improving and maintaining a healthy lifestyle is abundantly clear, changing habits and behaviors remains difficult [16]. Health-compromising behaviors are hard to stop, and health-promoting behaviors are challenging to adopt, making behavior change particularly difficult. As health care budgets face increasing pressure due to noncommunicable diseases associated with an aging population, workforce shortages, and rising costs of novel medical technologies, it is crucial that lifestyle interventions are cost-effective—for instance, through the use of online programs and apps [17].

Several lifestyle intervention programs for women actively seeking medical support to achieve pregnancy have shown high attrition rates and limited effects [18]. However, mobile health (mHealth) apps have the potential to overcome these obstacles by providing individualized, tailored, and repeated information [19,20]. Indeed, growing evidence suggests that mobile technology can effectively improve inadequate nutrition, lifestyle, and medication adherence [21,22].

Many lifestyle apps are currently available on the market, but very few are specifically designed for couples wishing to have children. To provide guidance for both patients and health care providers (HCPs), the MyFertiCoach (MFC) app was developed as a 6-month integrated lifestyle program with modules on healthy weight; nutrition; exercise; and quitting smoking, alcohol, drugs, and stress.

The MFC app supports couples trying to conceive and pregnant women in modifying unhealthy dietary and lifestyle habits. HCPs can use the app to help patients improve their lifestyle. Based on demographic parameters and an assessment of nutrition and lifestyle habits, a screening process is conducted, and a personalized coaching program consisting of lifestyle modules is created. Interactive elements, visual support, and motivational interviewing are incorporated as much as possible. Additionally, at various points during the program, we will assess whether lifestyle improvements have been achieved across different domains.

This retrospective, observational, single-center evaluation will be the first to assess the impact of the MFC app on changing 7 lifestyle habits in women seeking fertility treatment: healthy weight, diet, exercise, smoking, alcohol, drugs, and stress. Given the established influence of these lifestyle factors on subfertility, we expect that positive changes in these habits will also enhance reproductive health.

Methods

Overview

This is a retrospective, single-center evaluation conducted at the Elisabeth-TweeSteden Hospital in Tilburg, the Netherlands. This work does not assess the effectiveness of a new intervention before implementation but rather evaluates an addition to standard care after its integration.

Ethical Considerations

This study did not undergo formal approval by a medical ethics board for several reasons. First, the app was not purchased for scientific purposes but rather to augment standard care. As such, it received explicit approval from the board of directors for purchase and subsequently became part of standard care.

Additionally, all patients were free to use the app or not; their treatment remained unchanged if they chose not to use it. App usage was entirely voluntary, and data storage complied with national and European GDPR (General Data Protection Regulation) guidelines. The data were anonymized, and none of the researchers had direct access to them. Thus, the hospital purchased an app that was offered to patients as an optional resource. Each patient indicated whether they wanted to use the app and, if so, whether their anonymized data could be used for scientific purposes. As a result, the data set was generated with full patient consent and in compliance with privacy regulations.

Furthermore, our data analysis was conducted retrospectively, long after data collection had been completed. One of the authors (MJ) has previously used a similar approach, analyzing distress thermometer data stored in the electronic patient system [23]. For that study, the institutional review board was consulted, and an exemption was granted as it was classified as file research.

Adding a research paper is not the same as legislation, but the Dutch Central Committee on Research Involving Human Subjects states that “Retrospective research/research with patient files does not fall within the scope of the WMO, as the research subject is not physically involved in the research. Nor have the data that are to be researched been gathered for the sake of the

XSL•FO
RenderX

research. The research subjects do not have to change their behaviour for the sake of the research." [24]. We believe the last sentence is the most relevant: although patients filled out questionnaires and could use the app to support their care and contact their physician, the app did not require them to change their behavior for the sake of care or research. In fact, many patients we approached chose not to participate, and many who did eventually dropped out. They were free to do so. As the data were analyzed retrospectively, we consider our evaluation comparable to file research [25]. Following the decision tree for scientific medical research, our study falls under non-WMO research: it is medical-scientific research, but no patient was required to change procedures or behavior. The app was as freely available as any app that can be purchased from the Apple Store (Apple Inc.) or Google Play (Alphabet Inc.).

The MyFertiCoach App

The MFC app is a coaching program specifically designed to support lifestyle changes. It was developed by Ferring Pharmaceuticals B.V. in collaboration with a multidisciplinary team of health care professionals, with input from the patients' association Freya and patients who participated in the test phase.

HCPs can use the app to support patients and their partners while actively monitoring the (pre)active treatment phase. While the patient and their partner are motivated within the domains relevant to them, progress and results are visualized on a dashboard. Data essential for the next steps in treatment can also be used for evaluations and scientific research.

The MFC app promotes lifestyle changes over 6 months through motivation (using motivational interviewing techniques), information (on various domains), and relaxation (by incorporating game elements). It focuses on 7 key domains: achieving a healthy weight, maintaining a healthy diet, engaging in regular exercise, quitting smoking, stopping alcohol consumption, discontinuing drug/anabolic steroid use, and reducing stress through mindfulness. These domains were determined by the evaluation group and are supported by literature demonstrating their influence on the likelihood of a healthy pregnancy.

During the registration process, baseline demographic parameters and questionnaires on dietary and lifestyle habits are completed to screen patients on various factors, including BMI, smoking, nutrition, alcohol use, and stress. These assessments provide insights into patients' habits and help determine which domains require attention or intervention. The identified domains are then presented to the patient in the order of importance.

Patients were free to choose whether to complete the questionnaires, which took an average of 10 minutes in total. No financial compensation was provided for completing them.

To assess patients' knowledge, skills, and confidence in managing their own health or illness, both the patient and their partner are asked to complete the Patient Activation Measure questionnaire [26]. This provides HCPs with deeper insights into patients' perceived ability to improve their health, allowing for adjustments in the level of coaching and motivational interviewing techniques used.

At the first consultation, based on the screening outcomes, the HCP and patient jointly decide which aspects of their lifestyle to focus on. The relevant modules—healthy weight, healthy eating, healthy activity, smoking cessation, alcohol cessation, drug cessation, and stress reduction with mindfulness—are then activated. Once activated, patients receive module-specific questionnaires and general patient-reported outcome measures on a monthly basis. Patients are free to complete these at their convenience, but a baseline screening measure is required for inclusion in the evaluation. The program duration is set at 6 months.

Using the diary page, each patient can access all functionalities that allow them to record their daily mood and receive motivating messages. They can also send messages to their HCP with questions (eg, about nutrition or other lifestyle factors). The HCP can respond at their discretion. Each patient's progress is visible to the HCP via the MFC platform through domain-specific visualizations. Patients can track their progress using the app.

Recruitment and Data Collection

All new women aged 18-45 years who visited the fertility clinic were signed up for the MFC app after providing written consent and before their first appointment with the HCP. The MFC program, its modules, and the questionnaires were designed as a screening tool for lifestyle factors and were offered to all new patients as part of standard care.

The data extraction took place on February 1, 2024, and included all patients from January 1, 2022, to December 31, 2023, who completed the screening and had at least one activated module. Patients received a push notification and an email 1 month after the previous questionnaire. Patients who did not provide a baseline score and at least one follow-up score for an outcome (regardless of time) were excluded from further analysis of that particular outcome.

In the case of pregnancy, patients could check a box in the app indicating that they were pregnant (positive test). These patients would then receive additional questionnaires related to their pregnancy. As pregnancy is a defined endpoint, the formal intervention by the app was stopped, meaning these patients were no longer included in the evaluation. However, because some women indicated that they wanted to continue using the app, all modules—except the healthy weight module—remained active for 6 months. The app was never formally stopped by the physician, as it is part of standard care.

Outcomes

Assessing the Impact of the MFC App: Primary and Secondary Measures of Lifestyle Program Benefits

To properly evaluate the MFC app, several outcome measures have been defined to assess any additional benefits of our lifestyle program. The primary outcome was the change in the total risk score (TRS) [10] at the 3- and 6-month follow-ups. The secondary endpoints are changes in BMI, activity score, preconception dietary risk (PDR) score [27], distress score [28], smoking habits, alcohol intake, and program adherence.

XSL•FO
RenderX

Total Risk Score

The TRS is based on the Rotterdam Reproduction Risk Score (R3 score), the PDR score, and other existing evidence of associations with reproductive and pregnancy outcomes. An individual's TRS is defined as the sum of all risk scores per behavior (ie, based on vegetable, fruit, and folic acid intake, as well as smoking and alcohol use after 3 and 6 months) [10].

Vegetable and fruit intake were each subdivided into a risk score of 0, 1, 2, or 3. A score of 0 represents an adequate daily intake (≥200 g of vegetables per day and ≥2 pieces of fruit per day, respectively). Scores of 1 and 2 both indicate a "nearly adequate" intake (vegetable intake of 150 to <200 g and fruit intake of 1.5 to <2 pieces per day), with score 1 reflecting the participant's intention to change this risk factor and score 2 indicating no such intention. A score of 3 represents an inadequate daily intake (vegetable intake <150 g and fruit intake <1.5 pieces).

If a participant received a score of 1 or 2, an additional question regarding their intrinsic motivation was asked to determine whether they intended to improve their behavior related to this risk factor.

Folic acid supplement use was considered adequate (score 0) or inadequate (score 3) based on whether a participant met the recommendation of taking a 400-mg folic acid supplement daily during the periconceptional period. There is no evidence or recommendation for folic acid supplementation beyond 12 weeks of pregnancy, although this does not mean patients are not allowed to continue its use. Therefore, if a patient became pregnant while using the app, took folic acid for the first 12 weeks of pregnancy, and then stopped, she received a score of 0 for folic acid supplement use. This approach ensured consistency in scoring across all patients, allowing for a more conservative estimate of the app's benefits.

Risk scores for smoking and alcohol consumption were based on average daily use: no smoking (score 0), smoking 1-5 cigarettes (score 1), 6-14 cigarettes (score 3), or ≥15 cigarettes (score 6); and no drinking (score 0), drinking <1 alcoholic beverage per day (score 1), 1-2 beverages per day (score 2), or ≥2 beverages per day (score 3).

An important distinction is that the TRS is calculated separately for females and their partners (male/female). A higher TRS indicates unhealthier nutrition and lifestyle habits, which lowers the chance of becoming pregnant [10]. Based on previous findings, where a 1-point increase in the TRS was associated with a 21% reduction in pregnancy likelihood, we considered a 1-point decrease in the TRS clinically relevant.

Other Scores

We considered a BMI (calculated as weight divided by height squared) between 18.5 and 25 kg/m^2 to represent a "normal" weight [29]. A BMI ≥29 kg/m^2 is associated with a lower chance of becoming pregnant, with each unit increase above 29 linked to a 4% decrease in fertility [2]. Therefore, we considered a 1-point drop in the high BMI group clinically relevant.

To determine the activity score, we followed the Dutch physical activity guidelines, which state that adults should be physically active for at least 30 minutes a day on at least five days per week. This activity can include sports, walking, cycling, or strenuous household chores.

To compute a score, we devised 2 sets of 3 questions. The first set assessed the number of active minutes per week (ie, <75, 75 to <150, and ≥150 minutes), while the second set assessed muscle- and bone-enhancing activities (ie, never, once a week, or at least twice a week). Scores ranged from 1 to 3 for each set, with a total possible score between 2 and 6.

The level of distress was measured using the distress score, which provides a general impression of the problems couples experience in 5 areas: practical problems, family/social problems, emotional problems, religious/spiritual problems, and physical problems [30]. Additionally, couples were separately asked to rate their level of distress using a "thermometer" ranging from 0 to 10. The physician could then explore with the couple where their greatest concerns lay and discuss available options to address or reduce any symptoms or concerns as much as possible. A score of 5 or higher on the thermometer indicated a serious "burden." For our analysis, we only used the thermometer scores.

The PDR score was based on 6 nutritional questions covering the intake of key food groups, in line with dietary recommendations from the Netherlands Nutrition Centre. According to current guidelines, individuals should consume at least four slices of whole wheat bread daily (or an equivalent amount of cereals); use monounsaturated or polyunsaturated oils; and consume ≥200 g of vegetables daily, ≥2 pieces of fruit daily, ≥3 servings of meat or meat substitutes weekly, and ≥1 serving of fish weekly. Patients received 1 point for each recommendation met, with a maximum PDR score of 6, representing highly adequate nutrition according to the Netherlands Nutrition Centre's guidelines.

The smoking and alcohol scores were based on the number of cigarettes or alcoholic beverages consumed per day. Adherence was measured by the number of active patients over time.

Statistical Analysis

Evaluation Plan

This evaluation was conducted in accordance with a Statistical Analysis Plan (see Multimedia Appendix 1).

Descriptives and Transformations

We used descriptive statistics for each outcome of interest: the number of patients and observations for frequency data, and the mean and 95% CI for continuous data. For each outcome, we also calculated the number and percentage of responses per time point, as well as the mean and SD of response time. This allowed us to track response levels at each time point and assess the appropriateness of the predefined time-point categories (eg, 3 months, 6 months).

Some outcomes are known composites of aggregated parameters (eg, BMI), while others are transformations based on the specific manual of the questionnaire. We followed each manual's

XSL•FO
RenderX

guidelines on handling missing data when transforming or aggregating specific questionnaire items.

Missing Data

There is no official cut-off for what constitutes an acceptable or unacceptable level of missingness. Additionally, the reason for missing data is often unclear, except in cases where a participant has died. Therefore, we conducted a missing pattern analysis to determine the amount of missing data for each parameter separately and to identify how often multiple parameters are missing simultaneously.

We examined missing patterns across all items in the data set and aimed to classify them into 3 recognized types of missing data patterns: missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR) [31]. Of these 3 patterns, MNAR is the most problematic [32,33], as the missingness occurs for a specific reason related to the actual score a participant could have provided but did not.

As there are no formal statistical tests for assessing the size and nature of missing data, we relied heavily on visual inspection using bar charts, box plots, kernel density plots, and matrix plots. For example, if a questionnaire is missing at a particular time point and continues to be missing at subsequent time points, this would appear as a linear increase in missing data. Moreover, if factors such as age or having a partner are likely to influence responses, we would observe different response distributions across groups.

Power

As this evaluation was exploratory in nature, we did not conduct a formal power analysis to estimate the sample size required to detect a specific effect size with a given level of statistical significance. Statistical significance was determined by checking whether the 95% CI excluded 0.

Statistical Models Used

The main objective of this evaluation is to investigate the effect of following the MFC app program on changes in the TRS after 3 and 6 months. We analyzed the secondary outcomes (ie, BMI, activity score, PDR score, distress score, smoking habits, and alcohol intake) in the same way as the primary outcome, except for program adherence, which was described using descriptive statistics.

To estimate change, our primary analysis focused on patients for whom the intended module was made available. Accordingly, we used the following combinations: activity score for users of the activity module; alcohol score for users of the alcohol module; BMI for users of the weight module; last scores for users of the mindfulness module; the PDR score for users of the nutrition module; and smoking habits for users of the smoking module. The TRSs were analyzed for all participants who responded during the evaluation.

Using a maximum of 7 measurements per patient (1 baseline measurement and 6 follow-up measurements), and in the absence of a clinically relevant cut-off value, we estimated change using a (hierarchical) linear mixed model (LMM) [34]. LMMs are regression models designed to analyze clustered (longitudinal) data and can accommodate both MCAR and MAR data [35].

An important note is that we did not build separate models for the 3- and 6-month time points. As longitudinal data are correlated (ie, multiple measurements from the same patient), we constructed a single model from which the statistical significance of change can be derived at any time point up to the last observed measurement. Using an LMM, we can include time as either a categorical variable (eg, 1 month, 3 months, or 6 months) or as a continuous variable, represented by the number of months or the actual time in days. This approach allows us to evaluate the app's effect using days rather than months.

Because of the correlated nature of the data, we developed several LMMs with different random effects structures: random intercept, random slope, and random intercept with random slope [36]. This modeling strategy allows us to account for different assumptions, such as patients having the same starting point but different trajectories (random slope), the same trajectory but different starting points (random intercept), or both different starting points and different trajectories (random intercept and random slope).

Model Selection

The primary explanatory factors included in the models are time, partner status, age, the number of active modules, and the specific module associated with an outcome (eg, the nutrition module for the PDR score). To appropriately model time as a continuous variable, we will use linear functions, polynomial terms, or natural cubic splines [37].

To evaluate the relevance of these factors, we will iteratively remove parameters from a fully specified fixed-effects model using the chi-square test, considering both main effects and interaction effects. Model accuracy and parsimony will be assessed using the *F* test (analysis of variance) and the Akaike information criterion (AIC), with lower AIC values indicating better model fit. Model assumptions will be evaluated through graphical diagnostics and visualizations.

It is important to note that we will not build separate models for the 3- and 6-month time points. Instead, we will construct a single model from which the statistical significance of change can be derived at any time point up to the last observed measurement.

The analysis of the primary and secondary outcomes will result in several different models. This raises concerns about spurious results, as each model carries a 5% false-positive rate when determining statistical significance. To safeguard against such statistically spurious findings, it is important to assess whether the 7 outcomes are correlated at any given time point.

To address this, we will first create a correlation matrix using Spearman rank correlation, as many of the outcomes are ordinal in nature. If correlations are consistently higher than 0.50 between outcomes and across time points, we will include the correlated outcomes as predictors in the original LMM models. For example, if there is a correlation between distress and BMI, the distress outcome model will include both time and BMI as predictors.

XSL•FO
RenderX

As all variables have a clinical connection to the outcomes of interest, we allowed this form of statistical reasoning to guide variable inclusion. However, the final decision was based on model estimates for each variable, taking into account that multicollinearity can inflate SEs and affect variance estimates in mixed models. In summary, we included variables deemed both clinically important and statistically necessary.

Multiple Imputations in the Case of MNAR

Although patients may exhibit intermittent missingness—commonly associated with MAR situations and addressed using MAR models or multiple imputations—a sudden stop (ie, monotone missingness or drop out) strongly indicates an MNAR response. This is problematic for our analysis if it occurs before the 3- or 6-month follow-up. To address this, we will create a joint model, which combines an LMM with a survival model [38,39]. This approach is recognized as a viable alternative to older methods such as pattern-mixture modeling [40]. One of the authors has previously applied this method successfully to analyze quality-of-life changes in a population of patients with palliative pancreatic cancer [41].

The joint model yields 2 key outputs: (1) a survival model on program adherence, which incorporates information from the LMM; and (2) an LMM that adjusts its estimates based on the survival model. For each joint model, we used Markov chain Monte Carlo with 4 chains and 100,000 iterations, of which 10,000 were designated as burn-in. Model fit was assessed using trace and density plots.

Sensitivity Analyses

Adjusting for Timing Discrepancies in Longitudinal Data: Realignment of Assessment Periods

When analyzing longitudinal data, there is often a discrepancy between the intended timing of questionnaire returns and the actual return dates. This means that a 3-month assessment of an intervention may, in practice, represent a time frame that is either longer or shorter than planned. To address this, we created new time categories that better align with the intended schedule of questionnaire returns. For example, T3 is defined as 90 days plus or minus 14 days, and T6 as 180 days plus or minus 14 days.

Statistical Package Used

All analyses were performed using R Statistical Software (v4.1.2; R Foundation), a widely recognized and accepted program for statistical analysis. For the LMMs, we used the lme4 package. For the joint models, we used the JMbayes2 package in conjunction with the nlme package for constructing the LMM component. The survival package was used to build the Cox proportional hazards component of the joint models.

Results

Patient Characteristics

Table 1 presents the response rates during the evaluation period, along with response times, which varied by questionnaire (Figure S4 in Multimedia Appendix 2) and over time (Figure S2 in Multimedia Appendix 2). In total, 1815 patients were invited to participate, and all of them agreed to join, achieving a 100% recruitment rate. Of these, 818 patients (45.07%) agreed to participate. The mean age of the invited population was 34.1 (SD 5.46) years. A comparison between responders and nonresponders showed similar age distributions: 33.0 (SD 5.03) years for responders versus 35.1 (SD 5.63) years for nonresponders (Figure S1 in Multimedia Appendix 2).

XSL•FO
RenderX

Table 1. Number and percentage of responses, and the mean and SD of the response time.

Questionnaire and condition	Measurement moments							
	Screening	T0	T1	T2	T3	T4	T5	T6
Total risk score								
Regardless of the active module								
n (%)	1815 (100)	546 (30.1)	270 (14.9)	148 (8.2)	91 (5.0)	50 (2.8)	36 (2.0)	21 (1.2)
Mean (SD)	N/A[a]	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Activity								
Regardless of the active module								
n (%)	1815 (100)	758 (41.8)	235 (12.9)	139 (7.7)	85 (4.7)	50 (2.7)	34 (1.9)	20 (1.1)
Mean (SD)	N/A	0 (0)	55.6 (40.9)	89.6 (25.1)	120.6 (23.5)	154.1 (26.3)	186.3 (24.0)	220.3 (27.8)
Only active module exercise								
n (%)	444 (100)	438 (98.6)	233 (52.5)	138 (31.1)	85 (19.1)	50 (11.3)	34 (7.7)	20 (4.5)
Mean (SD)	N/A	0 (0)	55.8 (41.0)	89.7 (25.1)	120.6 (23.5)	154.1 (26.3)	186.3 (24.0)	220.6 (27.8)
Alcohol								
Regardless of the active module								
n (%)	1815 (100)	761 (41.9)	210 (11.6)	113 (6.2)	61 (3.4)	38 (2.1)	23 (1.3)	7 (0.4)
Mean (SD)	N/A	0 (0)	34.4 (23.7)	62.0 (24.0)	92.2 (26.2)	130.9 (31.9)	158.6 (31.2)	196.3 (45.6)
Only active module alcohol								
n (%)	209 (100)	199 (95.2)	95 (45.5)	42 (20.1)	26 (12.4)	14 (6.7)	11 (5.3)	5 (2.4)
Mean (SD)	N/A	0 (0)	46.2 27.2)	76.2 (27.1)	104.3 (28.4)	154.5 (38.4)	177.4 (35.2)	211.8 (45.5)
BMI (kg/m ^{2})								
Regardless of the active module								
n (%)	1815 (100)	792 (43.6)	209 (11.5)	115 (6.3)	73 (4.0)	45 (2.5)	32 (1.8)	19 (1.0)
Mean (SD)	N/A	0 (0)	57.9 (43.5)	93.3 (30.1)	125.9 (27.7)	154 (25.2)	187.1 (25.3)	225.8 (27.6)
Only active module weight								
n (%)	420 (100)	416 (99)	207 (49.3)	115 (27.4)	73 (17.4)	45 (10.7)	32 (7.6)	19 (4.5)
Mean (SD)	N/A	0 (0)	57.9 (43.5)	93.3 (30.1)	125.9 (27.7)	154.0 (25.2)	187.1 (25.3)	225.8 (27.6)
Distress								
Regardless of the active module								
n (%)	1815 (100)	764 (42.1)	84 (4.6)	42 (2.3)	27 (1.5)	17 (0.9)	13 (0.7)	7 (0.4)
Mean (SD)	N/A	0 (0)	65.7 (64.5)	115.5 (89.9)	160.8 (110.3)	203.8 (135.5)	220.0 (108.2)	223.9 (28.7)
Only active module mindfulness								
n (%)	157 (100)	153 (97.5)	83 (52.9)	42 (26.8)	27 (17.2)	17 (10.8)	13 (8.3)	7 (4.5)
Mean (SD)	N/A	0 (0)	66.4 (64.5)	115.5 (89.9)	160.8 (110.3)	203.8 (135.5)	220.0 (108.2)	223.9 (28.7)
Preconception dietary risk								

XSL•FO
RenderX

Questionnaire and condition	Measurement moments							
	Screening	T0	T1	T2	T3	T4	T5	T6
Regardless of the active module								
n (%)	1815 (100)	783 (43.1)	380 (20.9)	218 (12.0)	131 (7.2)	80 (4.4)	56 (3.1)	32 (1.8)
Mean (SD)	N/A	0 (0)	55.2 (42.3)	87.8 (34.6)	123 (42.5)	155.7 (50.2)	192.6 (58.1)	222.9 (26.9)
Only active module nutrition								
n (%)	693 (100)	680 (98.1)	379 (54.7)	218 (31.5)	131 (18.9)	80 (11.5)	56 (8.1)	32 (4.6)
Mean (SD)	N/A	0 (0)	55.3 (42.3)	87.8 (34.6)	123 (42.5)	155.7 (50.2)	192.6 (58.1)	222.9 (26.9)
Smoking								
Regardless of the active module								
n (%)	1815 (100)	761 (41.9)	579 (31.9)	375 (20.7)	262 (14.4)	191 (10.5)	131 (7.2)	94 (5.2)
Mean (SD)	N/A	0 (0)	20 (20.4)	44.4 (24.5)	71.7 (32.6)	88.6 (40.0)	114.1 (41.9)	129.6 (51.6)
Only active module smoking								
n (%)	130 (100)	119 (91.5)	82 (63.1)	53 (40.8)	42 (32.3)	28 (21.5)	18 (13.8)	11 (8.5)
Mean (SD)	N/A	0 (0)	16.4 (16.5)	41.4 (20.5)	63.9 (29.6)	85.0 (34.8)	105.8 (33.2)	126.2 (40.5)

[a]N/A: not applicable.

Among the 818 participants, 532 (65%) also included their partner. The partner-to-no-partner ratio differed significantly between those who accepted the invitation (532/818, 65.0% vs 286/818, 35.0%) and those who did not (323/997, 32.4% vs 674/997,67.6%; χ^2_1=190.81; *P*<.001).

When age distribution was further stratified by partner status, the results remained consistent. For nonparticipants, the mean ages were 34.0 (SD 4.59) years with a partner and 35.6 (SD 6.02) years without a partner. For participants, the mean ages were 32.1 (SD 4.22) years with a partner and 34.7 (SD 4.22) years without a partner.

Ultimately, data from 798 of 1815 patients (43.97% of those invited) were included in the analysis. These participants had a mean age of 33.0 (SD 5.04) years, with 517 (64.8%) including their partner. Of the 784 (98.2%) participants who reported their educational background, the majority (423/784, 53.9%) held a higher professional or university degree. Notably, higher education levels were associated with greater follow-up rates (see Figure S5 in Multimedia Appendix 2).

MFC Programs and Grading by Patients

Patients engaged with a diverse range of personal programs. Specifically, 693 patients included a nutrition module, while 130 patients included a smoking module. In total, the distribution of active modules per patient was as follows: we included 113 patients with 1 active module, 271 patients with 2 active modules, and 289 with 3 active modules; 4, 5, and 6 active modules were attached to 89, 33, and 3 patients, respectively. When asked to grade the MFC app on the seventh and fourteenth day of using the app (Figure S6 in Multimedia Appendix 2), 115 and 68 patients, respectively, gave a median score of 7 (IQR 3).

Patient Response and Response Time

The percentage of patients who filled out a questionnaire dropped rapidly across time (Figure S9 in Multimedia Appendix 2). The percentage of patients who had both a baseline score and at least one follow-up score (Figure S11 in Multimedia Appendix 2) ranged from 84 out of 764 (11.0%) patients for the distress score to 579 out of 761 (76.1%) patients for the smoking score, regardless of the module included. Across questionnaires, we received an average follow-up response rate of 38.3% (average of 282 patients/average of 737 patients). Compared with patients who did not return a follow-up questionnaire (see Figures S12-S14 and S16 in Multimedia Appendix 2), those who did respond had different median baseline scores for BMI (29.4 vs 24.3 kg/m^2), exercise (3 vs 5), distress score (6 vs 2), and alcohol (2 vs 0.1). However, the median scores for the TRS (3), PDR (4), and smoking (0) did not differ at follow-up. The follow-up percentage did differ much between those who included a partner and those who did not (see Figure S8 in Multimedia Appendix 2). When we accounted for the number of active modules, the graphs showed a slightly different pattern (see Figures S7, S8, and S10 in Multimedia Appendix 2), but the data became too sparse for additional statistical analysis. At baseline, the response rate ranged from 41.76% (758/1815) to 43.64% (792/1815), regardless of the active module, and from 91.5% (119/130) to 99.0% (416/420) per active module. These percentages declined over time, reaching 8.5% (11/130) for the smoking module and 2.4% (5/209) for the alcohol module at the last time point (see Figure S15 in Multimedia Appendix 2). The average response time did not align with the expected response time. While all patients completed their baseline questionnaire on time, the average response time for the first follow-up, regardless of the module, ranged from 20 days (smoking score) to 65.7 days (distress score). When considering only patients with active

XSL•FO
RenderX

modules, the response times were 16.4 and 66.4 days for the smoking and distress scores, respectively.

As time progressed, the gap between actual and expected response times widened. At the fifth time point—approximately 5 months after the follow-up—response times, regardless of the module, ranged from 220 days for the distress score to 114 days for the smoking score. The longest response time was observed for BMI, with an average of 226 days for the 6-month follow-up.

Sensitivity Analyses

Although secondary outcomes showed correlations at each time point, none consistently exceeded 0.5 (see Figures S24 and S25 in Multimedia Appendix 2). Therefore, we did not use any secondary outcomes to explain each other.

When we created new time categories that better aligned with the expected return of questionnaires (eg, T3 was defined as 90 days plus minus 14 days, T6 as 180 days plus minus 14 days), we reduced our effective sample size (see Figure S3 in Multimedia Appendix 2). However, this did not alter the distribution of values across the outcomes (see Figure S14 in Multimedia Appendix 2). Consequently, we used the Tx categories for our LMM models (main analysis) while using the actual time until questionnaire submission for our joint models, which also accounted for missing data.

Missing Data Pattern Analysis

The missing data pattern analysis revealed a monotone pattern (see Figures S31-S36 in Multimedia Appendix 2), meaning that once a patient stopped returning questionnaires, they dropped out entirely. This suggests a potential MNAR pattern. Consequently, we successfully fitted joint models for the TRS (see Figure S19 in Multimedia Appendix 2), activity (see Figure S26 in Multimedia Appendix 2), alcohol intake (see Figure S27 in Multimedia Appendix 2), BMI (see Figure S29 in Multimedia Appendix 2), PDR (see Figure S30 in Multimedia Appendix 2), and smoking habits (see Figure S28 in Multimedia Appendix 2). The results from the joint models align with the outcomes of the LMM models, although they are somewhat harder to interpret due to the inclusion of splines. Therefore, we reported the output of the LMM models, as they better followed the original questionnaire sampling. A joint model for distress did not converge. In summary, the MNAR analysis via joint modeling did not warrant further investigation.

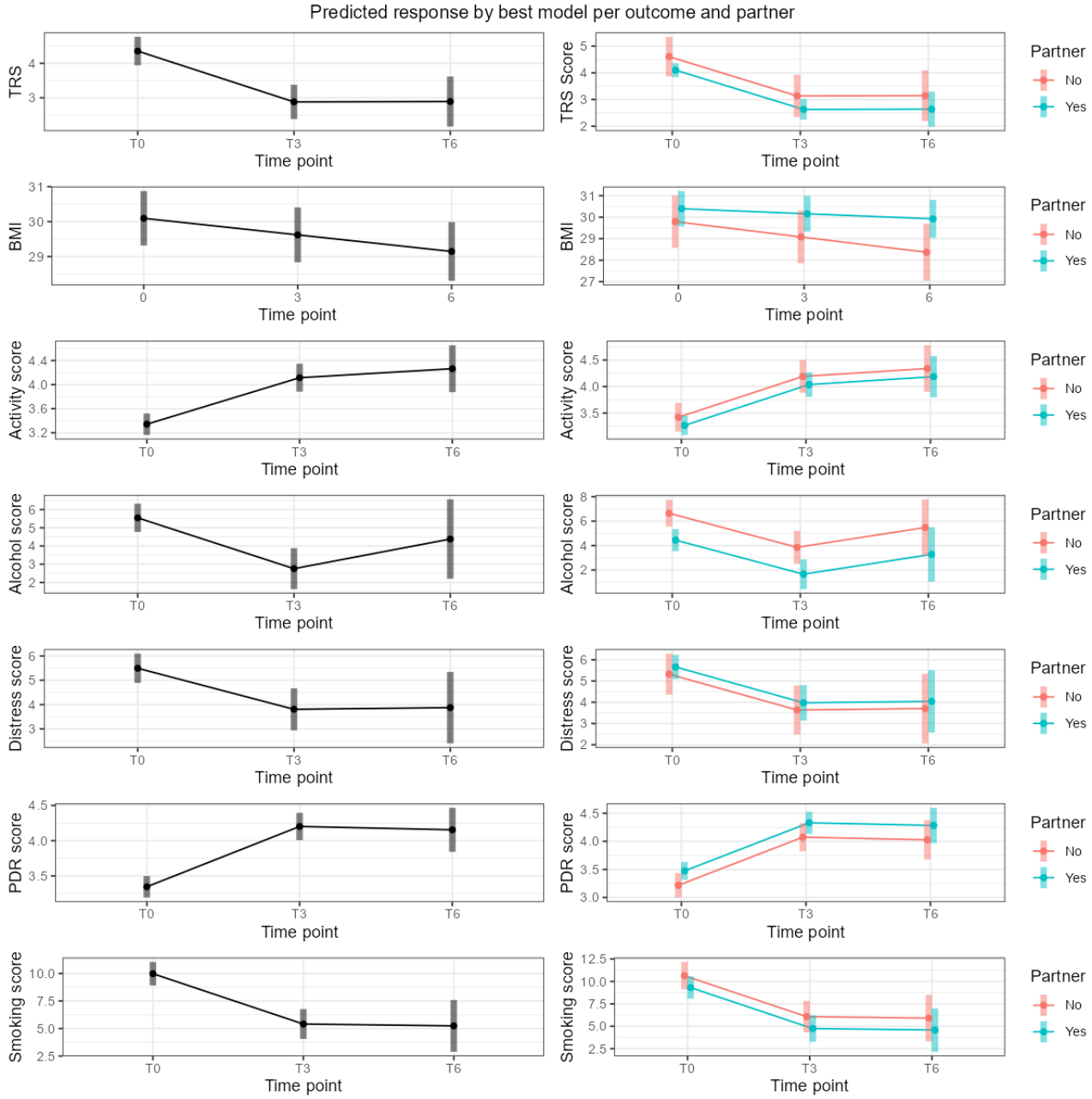
Results for the Primary and Secondary Outcomes

The model selection results (see Figures S17-S23 in Multimedia Appendix 2) show 14 models for alcohol intake, BMI, activity, and the distress score. For PDR and smoking habits, we created 15 models, while for the TRS, we included 13 different models. In almost all selection procedures, the AIC was lowest for models that included a main effect of time (as a dummy variable), partner, age, and the number of active modules.

Figure 1 and Table 2 show the results for primary and secondary outcomes at baseline, 3-month, and 6-month follow-ups. The number of patients and observations included ranged from 424 patients (1040 observations) for the TRS to 83 patients (272 observations) for the distress score.

XSL•FO
RenderX

Figure 1. Predicted response by best model per outcome and partner.



XSL•FO
RenderX

Table 2. Results following the MFC[a] app at the 3- and 6-month follow-up.[b]

Questionnaire	Values		Measurement moments									
			T0		T3				T6			
	N	Observed	Mean	95% CI	Mean	95% CI	Δ[c]	*P* value	Mean	95% CI	Δ	*P* value
Total risk score	424	1040	3.81	3.43-4.19	2.34	1.87-2.81	–1.47	<.001	2.35	1.65-3.06	–1.46	<.001
Activity	230	786	3.41	3.25-3.57	4.18	3.97-4.40	0.77	<.001	4.33	3.95-4.7	0.92	<.001
Alcohol	94	286	5.43	4.71-6.14	2.64	1.55-3.72	–2.79	<.001	4.26	2.10-6.42	–1.16	.53
BMI (kg/m^2)	207	698	30.1	29.4-30.8	29.6	28.9-30.3	–0.47	<.001	29.1	28.4-29.9	–0.95	<.001
Distress	83	272	5.53	4.94-6.12	3.84	2.99-4.70	–1.69	<.001	3.91	2.45-5.37	–1.62	.07
Preconception dietary risk	265	375	3.42	3.30-3.55	4.28	4.11-4.45	0.86	<.001	4.23	3.93-4.53	0.81	<.001
Smoking	313	81	9.95	8.89-11.02	5.38	4.03-6.75	–4.75	<.001	5.22	2.87-7.56	–4.74	<.001

[a]MFC: MyFertiCoach.

[b]Results are based on the most optimal model following an extensive model selection procedure.

[c]Δ: difference between categories.

For all outcomes, the change in score compared with baseline was statistically significant at 3 months (at the 5% level or *P*<.001) and considered positive (eg, lower BMI, reduced smoking, or improved nutrition). These changes remained statistically significant (*P*<.001)at 6 months, except for alcohol intake (*P*<.53) and distress levels (*P*<.07). For alcohol, part of the initial positive response (Δ –2.79; *P*<.001) diminished by 6 months (Δ –1.16; *P*<.531). For distress, the initial statistically significant change (Δ –1.69; *P*<.001) shifted to a positive trend (Δ –1.62; *P*<.073). BMI was the only outcome that continued improving over 6 months (30.1 vs 29.6 vs 29.1 kg/m^2; *P*<.001). For the TRS, the changes were clinically relevant at both 3 and 6 months. For BMI, clinical relevance was only observed at 6 months.

When we stratified the longitudinal analysis by partner presence (Figure 1 and Table 3), statistical significance was maintained at 3 months (*P*<.001), except for BMI, which was no longer significant in the group that included a partner (30.4 vs 30.2 kg/m^2 vs 29.9 kg/m^2; *P*<.07). However, the group without a partner showed an even greater decrease in BMI (29.8 vs 29.1 vs 28.4 kg/m^2; *P*<.001). Regardless of partner status, the change in alcohol intake remained nonsignificant (*P*=.53).

XSL•FO
RenderX

Table 3. Results following the MFC[a] app at the 3- and 6-month follow-up.[b]

Questionnaire	Partner	Values		Measurement moments									
				T0		T3				T6			
		N	Observed	Mean	95% CI	Mean	95% CI	Δ[c]	*P* value	Mean	95% CI	Δ	*P* value
Total risk score		424	1040										
	Yes			3.56	3.35-3.77	2.1	1.74-2.44	1.47	<.001	2.1	1.47-2.73	1.46	<.001
	No			4.07	3.34-4.79	2.59	1.82-3.37	1.47	<.001	2.6	1.67-3.54	1.46	<.001
Activity		230	786										
	Yes			3.33	3.17-3.50	4.1	3.89-4.33	–0.77	<.001	4.26	3.87-4.64	–0.9	<.001
	No			3.49	3.23-3.75	4.26	3.96-4.56	–0.77	<.001	4.41	3.98-4.84	–0.9	<.001
Alcohol		94	286										
	Yes			4.33	3.44-5.21	1.5	0.32-2.75	2.79	<.001	3.2	0.93-5.39	1.2	.53
	No			6.52	5.51-7.54	3.7	2.44-5.03	2.79	<.001	5.4	3.09-7.63	1.2	.53
BMI (kg/m^2)		207	698										
	Yes			30.4	29.6-31.2	30.2	29.3-31.0	–0.23	.07	29.9	29.0-31.0	–0.5	.07
	No			29.8	28.6-31.0	29.1	27.9-30.3	–0.71	<.001	28.4	27.0-29.7	–1.4	<.001
Distress		83	272										
	Yes			5.70	5.14-6.26	4.01	3.18-4.84	1.69	<.001	4.08	2.62-5.54	1.62	.07
	No			5.36	4.40-6.32	3.67	2.53-4.82	1.69	<.001	3.74	2.11-5.37	1.62	.07
Preconception dietary risk		1265	375										
	Yes			3.55	3.42-3.69	4.41	4.23-4.59	–0.86	<.001	4.36	4.06-4.66	–0.81	<.001
	No			3.30	3.10-3.49	4.15	3.92-4.38	–0.86	<.001	4.10	3.77-4.44	–0.81	<.001
Smoking		313	81										
	Yes			9.29	8.03-10.54	4.72	3.23-6.20	–4.75	<.001	4.55	2.13-6.97	–4.74	<.001
	No			10.64	9.09-12.14	6.04	4.29-7.80	–4.75	<.001	5.88	3.28-8.47	–4.74	<.001

[a]MFC: MyFertiCoach.

[b]Results are based on the most optimal model following an extensive model selection procedure.

[c]Δ: difference between categories.

When we further analyzed outcomes by partner presence and time point (Table 4), we observed lower alcohol scores (Δ –2.2; *P*<.001) and higher PDR scores (Δ 0.256; *P*=.03) in patients who included their partner. These were both main effects. However, for BMI, we identified an interaction effect, showing a more pronounced decline at 6 months (Δ –1.555; *P*=.04) in patients who did not include their partner.

XSL•FO
RenderX

Table 4. Results following the MFC[a] app at the 3- and 6-month follow-up[b].

Questionnaire and partner	Value		Measurement moments								
			T0			T3			T6		
	N	Observed	Mean	Δ[c] 95% CI	Δ *P* value	Mean	Δ 95% CI	Δ *P* value	Mean	Δ 95% CI	*P* value
Total risk score	424	1040									
Yes			3.56	0.505 (–0.244 to 1.25)	.19	2.1	0.505 (–0.244 to 1.25)	.19	2.1	0.505 (–0.244 to 1.25)	.19
No			4.07	N/A[d]	N/A	2.59	N/A	N/A	2.6	N/A	N/A
Activity	230	786									
Yes			3.33	0.154 (–0.135 to 1.051)	.29	4.1	0.154 (–0.135 to 1.051)	.29	4.3	0.154 (–0.135 to 1.051)	.29
No			3.49	N/A	N/A	4.3	N/A	N/A	4.4	N/A	N/A
Alcohol	94	286									
Yes			4.33	2.2 (0.922 to 3.47)	<.001	1.5	2.2 (0.922 to 3.47)	<.001	3.2	2.2 (0.922 to 3.47)	<.001
No			6.52	N/A	N/A	3.7	N/A	N/A	5.4	N/A	N/A
BMI (kg/m^2)	207	698									
Yes			30.4	–0.599 (–1.96 to 0.767)	.39	30.2	–1.077 (–2.46 to 0.3021)	.12	29.9	–1.555 (–3.06 to –0.0518)	.04
No			29.8	N/A	N/A	29.1	N/A	N/A	28.4	N/A	N/A
Distress	83	272									
Yes			5.70	–0.339 (–1.37 to 0.691)	.51	4.01	–0.339 (–1.37 to 0.691)	.51	4.08	–0.339 (–1.37 to 0.691)	.51
No			5.36	N/A	N/A	3.67	N/A	N/A	3.74	N/A	N/A
Preconception dietary risk	1265	375									
Yes			3.55	–0.256 (–0.484 to –0.028)	.03	4.41	–0.256 (–0.484 to –0.028)	.03	4.36	–0.256 (–0.484 to –0.028)	.03
No			3.30	N/A	N/A	4.15	N/A	N/A	4.10	N/A	N/A
Smoking	313	81									
Yes			9.29	–0.909 (–0.485 to 1.458)	.15	4.72	–0.909 (–0.485 to 1.458)	.15	4.55	–0.909 (–0.485 to 1.458)	.15
No			10.64	N/A	N/A	6.04	N/A	N/A	5.88	N/A	N/A

[a]MFC: MyFertiCoach.

[b]Results are based on the most optimal model following an extensive model selection procedure.

[c]Δ: difference between categories.

[d]N/A: not applicable.

Discussion

Principal Findings

Our findings indicate that introducing a lifestyle app can immediately improve several lifestyle factors associated with subfertility and the success of fertility treatment. Specifically, the assessment of the TRS shows that app usage leads to a clinically relevant improvement. The peak effect of the intervention appears to occur around the third month, after which improvements plateau—except for BMI, which remained clinically relevant at 6 months. The MFC app was designed with input from both patients and health care professionals to ensure it targeted lifestyle domains deemed most useful. As a result, the app was intended to be more than just a digital

XSL•FO
RenderX

summary of suboptimal lifestyle behaviors. Through shared decision-making, both patients and HCPs contributed to the final selection of active modules, integrating the app into the broader treatment plan. From a preventive medicine perspective, the app provided a valuable starting point and contributed to more efficient care, offering an inexpensive way to influence lifestyle. This impact is also reflected in the data, where most effects were observed immediately after app use began.

However, it appears that not everyone is eager to start an intervention of this type. While we cannot determine with certainty why some individuals chose not to participate, we observed lower participation rates among those with lower levels of education and those without a partner. It remains unclear whether the presence of a partner played a significant role in the acceptance of this evaluation. However, even when better follow-up led to a more favorable effect of the app, partner participation did not consistently enhance its effectiveness.

On an individual level, the reasons why some patients stopped using the app—or never started—remain uncertain. One plausible explanation is that certain users did not identify with the app, despite extensive screening and discussions. Another possibility is that aspects of the app's design or usability did not appeal to everyone, although we do not have specific data to confirm this.

In summary, we believe the MFC app is a cost-effective addition to standard care, helping subfertile women and their partners implement necessary lifestyle changes. Financial costs are minimal, and health care professionals require little time investment. While the app primarily attracts a specific group—those with higher education and a partner—it does lead to meaningful improvements in key lifestyle factors.

Limitations

Because of the nature of our evaluation, we were unable to account for all known and unknown confounders. Despite conducting several sensitivity analyses, we cannot rule out the possibility that the temporal differences—or lack thereof—observed between patients, including those with a partner, may be influenced by factors beyond patient characteristics or the app itself. Additionally, the pattern of missing data cannot be formally assessed in a quantitative manner, meaning we do not know why patients dropped out when they did.

Furthermore, we did not include a formal comparative group, so we cannot determine how outcomes would have developed in patients who did not receive guidance through the app. These limitations—along with others, such as the lack of randomization or blinding—prevent us from drawing causal inferences.

Many uncertainties remain. For example, we do not know why patients dropped out, nor do we have full transparency regarding the exact selection process. Although modules were chosen through shared decision-making, it is unclear whether the final choice was driven solely by motivation. However, this does not imply that a lack of improvement is due to low motivation or suboptimal care. It is possible that the interventions themselves did not sufficiently support patients in improving suboptimal lifestyle factors.

Special consideration should be given to folic acid use and pregnancy. While we did not extract these data, it is possible that some patients became pregnant while using the app and continued to use it. As Dutch guidelines recommend folic acid supplementation during the first 3 months of pregnancy, some patients may not have used folic acid, leading to a higher TRS. If this is the case, the app's beneficial effect could be even greater than our current estimates suggest.

The consistency of findings across LMMs and joint models does not eliminate the potential limitations of our study. While LMMs can handle data that are MAR and joint models can address missing-not-at-random (MNAR) data by incorporating estimates from survival analysis, no statistical safeguard can fully account for the high amount of missing data in our evaluation. Ultimately, there is no way to account for what we do not know.

As a result, we cannot claim that the high probability of MNAR data was entirely mitigated by the joint models. However, we can state that incorporating survival models alongside mixed models produced estimates that were very similar to those obtained using mixed models alone. This strengthens our confidence in the robustness of the model estimates given the available data. Nonetheless, our findings should be interpreted with caution due to the high level of missing data—an inherent limitation when evaluating interventions in standard care settings.

Additionally, our findings are based on a select group of patients from a single center in the Netherlands, where we observed a high rate of attrition during implementation. As we cannot fully determine the reasons for dropout, there is a potential risk of bias, as those who continued participating may differ systematically from those who dropped out.

Comparison With Prior Work

In the Netherlands, 2 other groups are actively developing apps to support treatment: the international LIFESTYLE study, in which the Medical University of Groningen participates [6,7,42], and the Smarter Pregnancy platform at Erasmus Medical Center [43]. However, these initiatives primarily involve hands-on interventions, making them financially expensive and time-intensive, although they tend to yield better and more stable responses. This suggests that maintaining patient motivation may require a similar level of investment from health care professionals.

Lifestyle changes are notoriously difficult to achieve and sustain. Eating, drinking, and smoking are habits that provide immediate gratification, while their negative effects manifest only in the long term. Psychological stressors are also challenging to address and vary depending on their nature—for example, losing a loved one differs from facing persistent financial difficulties. Consequently, our findings of immediate but not long-lasting changes align with what has been extensively documented in previous research.

Patients varied significantly in their personal trajectories. Some showed substantial improvement, others experienced only marginal benefits, and some saw no measurable change despite their efforts. Analyzing heterogeneous data often results in

XSL•FO
RenderX

marginal main effects which, even when statistically significant, may not translate into meaningful clinical change. Attempts to include interaction effects—such as partner presence, number of active modules, and age—did not provide clearer insights into why some patients responded well while others did not.

However, given the nature of the app and the study setting, the observed dropout rates were not unexpected. We aimed to facilitate lifestyle changes in a predominantly working-class population with limited health literacy. This does not imply that these patients cannot benefit from or are unwilling to engage with such interventions, but rather that additional efforts are necessary to achieve long-term change. Therefore, evaluations of mHealth apps should not rely solely on questionnaires but should also incorporate in-depth follow-up interviews, which we plan to conduct in future development stages.

Conclusions

Our evaluation of an mHealth app in standard care, designed to support women seeking pregnancy, demonstrated an immediate and clinically relevant improvement in key lifestyle parameters. Further research is needed to identify causal pathways leading to sustained effectiveness. To maintain and enhance these results, patient-specific programs must be further tailored.

Acknowledgments

We thank Thomas Bachet, Naomi Tielen, the patient association group Freya, and the nurses at the Elisabeth TweeSteden Hospital for their useful insights and support. We also thank our “thinktank”: Prof. dr. A Hoek (University Medical Center Groningen), Prof. dr. J Laven (Erasmus Medical Center), Dr. M. Schoonenberg (Nij Geertgen) and Drs. G Van den Dool (Nij Linge).

Data Availability

The data sets generated during or analyzed during this evaluation are not publicly available due to privacy reasons. However, because the patients give explicit consent that their data can be anonymously used for scientific research, the data are available from the corresponding author upon reasonable request.

Conflicts of Interest

ES works at Ferring B.V. The remaining authors have declared no conflicts of interest.

Multimedia Appendix 1
Supplementary figures.
[ZIP File (Zip Archive), 2713 KB - formative_v9i1e64239_app1.zip]

Multimedia Appendix 2
Statistical analysis plan.
[DOCX File , 600 KB - formative_v9i1e64239_app2.docx]

References

1. Godfrey KM, Gluckman PD, Hanson MA. Developmental origins of metabolic disease: life course and intergenerational perspectives. Trends Endocrinol Metab 2010 Apr;21(4):199-205. [doi: 10.1016/j.tem.2009.12.008] [Medline: 20080045]
2. van der Steeg JW, Steures P, Eijkemans MJ, Habbema JDF, Hompes PG, Burggraaff JM, et al. Obesity affects spontaneous pregnancy chances in subfertile, ovulatory women. Hum Reprod 2008 Feb 14;23(2):324-328. [doi: 10.1093/humrep/dem371] [Medline: 18077317]
3. Fitzsimons KJ, Modder J, Greer IA. Obesity in pregnancy: risks and management. Obstet Med 2009 Jun 22;2(2):52-62. [doi: 10.1258/om.2009.090009] [Medline: 27582812]
4. Schneider E, Hamer O, Smith C, Hill J. Evaluating the association of female obesity with the risk of live birth following IVF: implications for clinical practice. Pract Midwife 2024 Jan 16;27(4):20240101-20240451. [doi: 10.1093/humupd/dmz011] [Medline: 39045260]
5. Hassan MA, Killick SR. Negative lifestyle is associated with a significant reduction in fecundity. Fertil Steril 2004 Feb;81(2):384-392 [FREE Full text] [doi: 10.1016/j.fertnstert.2003.06.027] [Medline: 14967378]
6. Einarsson S, Bergh C, Friberg B, Pinborg A, Klajnbard A, Karlström PO, et al. Weight reduction intervention for obese infertile women prior to IVF: a randomized controlled trial. Hum Reprod 2017 Aug 01;32(8):1621-1630. [doi: 10.1093/humrep/dex235] [Medline: 28854592]
7. Mutsaerts MA, van Oers AM, Groen H, Burggraaff JM, Kuchenbecker WKH, Perquin DAM, et al. Randomized trial of a lifestyle program in obese infertile women. N Engl J Med 2016 May 19;374(20):1942-1953. [doi: 10.1056/NEJMoa1505297] [Medline: 27192672]
8. Hoek A, Wang Z, van Oers AM, Groen H, Cantineau AE. Effects of preconception weight loss after lifestyle intervention on fertility outcomes and pregnancy complications. Fertil Steril 2022 Sep;118(3):456-462 [FREE Full text] [doi: 10.1016/j.fertnstert.2022.07.020] [Medline: 36116799]

XSL•FO
RenderX

9. Oostingh EC, Hall J, Koster MP, Grace B, Jauniaux E, Steegers-Theunissen RP. The impact of maternal lifestyle factors on periconception outcomes: a systematic review of observational studies. Reprod Biomed Online 2019 Jan;38(1):77-94. [doi: 10.1016/j.rbmo.2018.09.015] [Medline: 30424937]
10. van Dijk MR, Koster MP, Willemsen SP, Huijgen NA, Laven JS, Steegers-Theunissen RP. Healthy preconception nutrition and lifestyle using personalized mobile health coaching is associated with enhanced pregnancy chance. Reprod Biomed Online 2017 Oct;35(4):453-460. [doi: 10.1016/j.rbmo.2017.06.014] [Medline: 28688924]
11. van Oers AM, Mutsaerts MAQ, Burggraaff JM, Kuchenbecker WKH, Perquin DAM, Koks CAM, et al. Association between periconceptional weight loss and maternal and neonatal outcomes in obese infertile women. PLoS ONE 2018 Mar 28;13(3):e0192670. [doi: 10.1371/journal.pone.0192670]
12. Sauder KA, Starling AP, Shapiro AL, Kaar JL, Ringham BM, Glueck DH, et al. Diet, physical activity and mental health status are associated with dysglycaemia in pregnancy: the Healthy Start Study. Diabet Med 2016 Mar 03;33(5):663-667 [FREE Full text] [doi: 10.1111/dme.13093] [Medline: 26872289]
13. Steegers-Theunissen R, Hoek A, Groen H, Bos A, van den Dool G, Schoonenberg M, et al. Pre-Conception Interventions for Subfertile Couples Undergoing Assisted Reproductive Technology Treatment: Modeling Analysis. JMIR Mhealth Uhealth 2020 Nov 23;8(11):e19570. [doi: 10.2196/19570]
14. Verkuijlen J, Verhaak C, Nelen WLDM, Wilkinson J, Farquhar C. Psychological and educational interventions for subfertile men and women. Cochrane Database Syst Rev 2016 Mar 31;3(3):CD011034 [FREE Full text] [doi: 10.1002/14651858.CD011034.pub2] [Medline: 27031818]
15. Gaitzsch H, Benard J, Hugon-Rodin J, Benzakour L, Streuli I. The effect of mind-body interventions on psychological and pregnancy outcomes in infertile women: a systematic review. Arch Womens Ment Health 2020 Jan 02;23(4):479-491. [doi: 10.1007/s00737-019-01009-8]
16. Blüher M. Obesity: global epidemiology and pathogenesis. Nat Rev Endocrinol 2019 May 27;15(5):288-298. [doi: 10.1038/s41574-019-0176-8] [Medline: 30814686]
17. de Jong MJ, van der Meulen-de Jong AE, Romberg-Camps MJ, Becx MC, Maljaars JP, Cilissen M, et al. Telemedicine for management of inflammatory bowel disease (myIBDcoach): a pragmatic, multicentre, randomised controlled trial. The Lancet 2017 Sep;390(10098):959-968. [doi: 10.1016/s0140-6736(17)31327-2]
18. Hiller R, Griesinger G. How effective are lifestyle interventions for overweight women trying to conceive? Curr Opin Obstet Gynecol 2023 Jun 01;35(3):230-237. [doi: 10.1097/GCO.0000000000000874] [Medline: 37185354]
19. Meyers AJ, Domar AD. Research-supported mobile applications and internet-based technologies to mediate the psychological effects of infertility: a review. Reprod Biomed Online 2021 Mar;42(3):679-685. [doi: 10.1016/j.rbmo.2020.12.004] [Medline: 33487558]
20. Timmers T, Keijsers M, Kremer JAM, Janssen L, Smeenk J. Supporting women undergoing IVF treatment with timely patient information through an app: randomized controlled trial. JMIR Mhealth Uhealth 2021 Aug 27;9(8):e28104 [FREE Full text] [doi: 10.2196/28104] [Medline: 34448725]
21. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013 Jan 15;10(1):e1001363. [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
22. Wei HX, Yang YL, Luo TY, Chen WQ. Effectiveness of mobile health interventions for pregnant women with gestational diabetes mellitus: a systematic review and meta-analysis. Journal of Obstetrics and Gynaecology 2023 Aug 22;43(2):1-13. [doi: 10.1080/01443615.2023.2245906] [Medline: 37605977]
23. Jacobs M, Anderegg MCJ, Schoorlemmer A, Nieboer D, Steyerberg EW, Smets EMA, et al. Patients with oesophageal cancer report elevated distress and problems yet do not have an explicit wish for referral prior to receiving their medical treatment plan. Psycho-Oncology 2016 Jun;26(4):452-460. [doi: 10.1002/pon.4161]
24. The Central Committee on Research Involving Human Subjects. Non-WMO research. The Central Committee on Research Involving Human Subjects. 2024. URL: https://english.ccmo.nl/investigators/additional-requirements-for-certain-types-of-research/non-wmo-research [accessed 2024-11-20]
25. The Central Committee on Research Involving Human Subjects. File research. The Central Committee on Research Involving Human Subjects. 2024. URL: https://english.ccmo.nl/investigators/additional-requirements-for-certain-types-of-research/non-wmo-research/file-research [accessed 2024-11-20]
26. Rademakers J, Nijman J, van der Hoek L, Heijmans M, Rijken M. Measuring patient activation in the Netherlands: translation and validation of the American short form Patient Activation Measure (PAM13). BMC Public Health 2012 Jul 31;12(1):1-7. [doi: 10.1186/1471-2458-12-577]
27. Twigt JM, Bolhuis MEC, Steegers EAP, Hammiche F, van Inzen WG, Laven JSE, et al. The preconception diet is associated with the chance of ongoing pregnancy in women undergoing IVF/ICSI treatment. Hum Reprod 2012 Aug 15;27(8):2526-2531. [doi: 10.1093/humrep/des157] [Medline: 22593431]
28. van Nuenen FM, Donofrio SM, Tuinman MA, van de Wiel HBM, Hoekstra-Weebers JEHM. Feasibility of implementing the 'Screening for Distress and Referral Need' process in 23 Dutch hospitals. Support Care Cancer 2016 Aug 26;25(1):103-110. [doi: 10.1007/s00520-016-3387-8]

XSL•FO
RenderX

29. Silvestris E, Lovero D, Palmirotta R. Nutrition and female fertility: an interdependent correlation. Front Endocrinol (Lausanne) 2019 Jun 7;10:346-313. [doi: 10.3389/fendo.2019.00346] [Medline: 31231310]
30. Ownby KK. Use of the distress thermometer in clinical practice. J Adv Pract Oncol 2019 Mar;10(2):175-179 [FREE Full text] [Medline: 31538028]
31. Carpenter JR, Smuk M. Missing data: a statistical framework for practice. Biom J 2021 Jun 24;63(5):915-947. [doi: 10.1002/bimj.202000196] [Medline: 33624862]
32. Eekhout I, de Boer RM, Twisk JWR, de Vet HCW, Heymans MW. Missing data: a systematic review of how they are reported and handled. Epidemiology 2012 Sep;23(5):729-732. [doi: 10.1097/EDE.0b013e3182576cdb] [Medline: 22584299]
33. Fielding S, Fayers PM, McDonald A, McPherson G, Campbell MK, RECORD Study Group. Simple imputation methods were inadequate for missing not at random (MNAR) quality of life data. Health Qual Life Outcomes 2008 Aug 04;6(1):57 [FREE Full text] [doi: 10.1186/1477-7525-6-57] [Medline: 18680574]
34. Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, et al. Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol 2009 Mar;24(3):127-135. [doi: 10.1016/j.tree.2008.10.008] [Medline: 19185386]
35. Gabrio A, Plumpton C, Banerjee S, Leurent B. Linear mixed models to handle missing at random data in trial-based economic evaluations. Health Econ 2022 Jun 02;31(6):1276-1287 [FREE Full text] [doi: 10.1002/hec.4510] [Medline: 35368119]
36. Murphy J, Weaver NE, Hendricks AE. Accessible analysis of longitudinal data with linear mixed effects models. Dis Model Mech 2022 May 01;15(5):1-8. [doi: 10.1242/dmm.048025] [Medline: 35521689]
37. Elhakeem A, Hughes RA, Tilling K, Cousminer DL, Jackowski SA, Cole TJ, et al. Using linear and natural cubic splines, SITAR, and latent trajectory models to characterise nonlinear longitudinal growth trajectories in cohort studies. BMC Med Res Methodol 2022 Mar 15;22(1):1-20. [doi: 10.1186/s12874-022-01542-8]
38. Ibrahim JG, Chu H, Chen LM. Basic concepts and methods for joint models of longitudinal and survival data. JCO 2010 Jun 01;28(16):2796-2801. [doi: 10.1200/jco.2009.25.0654]
39. Asar O, Ritchie J, Kalra PA, Diggle PJ. Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. Int J Epidemiol 2015 Feb;44(1):334-344. [doi: 10.1093/ije/dyu262] [Medline: 25604450]
40. Shen C. Application of pattern-mixture models to outcomes that are potentially missing not at random using pseudo maximum likelihood estimation. Biostatistics 2005 Apr 01;6(2):333-347. [doi: 10.1093/biostatistics/kxi013]
41. Tello Valverde CP, Ebrahimi G, Sprangers MA, Pateras K, Bruynzeel AM, Jacobs M, et al. Impact of short-course palliative radiation therapy on pancreatic cancer-related pain: prospective phase 2 nonrandomized PAINPANC trial. Int J Radiat Oncol Biol Phys 2024 Feb 01;118(2):352-361. [doi: 10.1016/j.ijrobp.2023.08.055] [Medline: 37647972]
42. van Oers AM, Mutsaerts MAQ, Burggraaff JM, Kuchenbecker WKH, Perquin DAM, Koks CAM, LIFEstyle study group. Cost-effectiveness analysis of lifestyle intervention in obese infertile women. Hum Reprod 2017 Jul 01;32(7):1418-1426. [doi: 10.1093/humrep/dex092] [Medline: 28486704]
43. Van Dijk MR, Huijgen NA, Willemsen SP, Laven JS, Steegers EA, Steegers-Theunissen RP. Impact of an mHealth platform for pregnancy on nutrition and lifestyle of the reproductive population: a survey. JMIR Mhealth Uhealth 2016 May 27;4(2):e53. [doi: 10.2196/mhealth.5197] [Medline: 27234926]

Abbreviations

AIC: Akaike information criterion
GDPR: General Data Protection Regulation
HCP: health care provider
LMM: linear mixed model
MAR: missing at random
MCAR: missing completely at random
MFC: MyFertiCoach
mHealth: mobile health
MNAR: missing not at random
PDR: preconception dietary risk
TRS: total risk score

XSL•FO
RenderX

Edited by A Mavragani; submitted 12.07.24; peer-reviewed by J Grieger, T Wang; comments to author 06.09.24; revised version received 25.09.24; accepted 17.01.25; published 10.03.25.

Please cite as:
Smeenk J, Smit E, Jacobs M, van Rooij I
Evaluation of the MyFertiCoach Lifestyle App for Subfertile Couples: Single-Center Evaluation of Augmented Standard Care
JMIR Form Res 2025;9:e64239
URL: https://formative.jmir.org/2025/1/e64239
doi:10.2196/64239
PMID:

©Jesper Smeenk, Ellen Smit, Marc Jacobs, Ilse van Rooij. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development of a Voice-Activated Virtual Assistant to Improve Insomnia Among Young Adult Cancer Survivors: Mixed Methods Feasibility and Acceptability Study

Hunter Groninger[1,2,3], MD; Hannah Arem[3,4], PhD; Lylian Ayangma[3], BA; Lisa Gong[3], BA; Eric Zhou[5], PhD; Daniel Greenberg[6], BA

[1]MedStar Health Research Institute, Washington, DC, United States
[2]MedStar Washington Hospital Center, Washington, DC, United States
[3]Georgetown University School of Medicine, Washington, DC, United States
[4]MedStar Health Research Institute, Hyattsville, MD, United States
[5]Dana Farber Cancer Institute, Boston, MA, United States
[6]MediaRez LLC, Washington, DC, United States

Corresponding Author:
Hunter Groninger, MD
MedStar Health Research Institute
c/o 110 Irving Street NW
Room 2A68
Washington, DC, 20010
United States
Phone: 1 202 877 7445
Email: hunter.groninger@medstar.net

Abstract

Background: Up to 75% of young adult cancer survivors (YACS) experience chronic insomnia, negatively affecting physical and emotional health and overall quality of life. Cognitive behavioral therapy for insomnia (CBT-I) is a gold-standard intervention to address insomnia. To improve CBT-I access and treatment adherence, screen-based digital CBT-I platforms have been developed. However, even with these digital products, widespread uptake of CBT-I remains limited, and new strategies for CBT-I delivery are warranted.

Objective: The objective of this study is to understand how YACS experience insomnia and how they might incorporate technology-delivered CBT-I into a daily routine and test the feasibility and acceptability of a novel screen-free voice-activated virtual assistant–delivered CBT-I prototype.

Methods: Eligible participants—ages 18-39, living with a history of cancer (any type, any stage), self-reporting on average less sleep than National Sleep Foundation recommendations, and English-speaking—were recruited from a major urban cancer center, 2 regional oncology clinics, and 2 cancer survivorship support groups. We conducted 4 focus groups to understand the YACS experience of insomnia, their routine use of technology at home, particularly voice-activated virtual assistants such as Amazon Alexa, and input on how CBT-I might be delivered at home through a smart speaker system. We developed a prototype device to deliver key elements of CBT-I at home along with circadian lighting and monitoring of post-bedtime device use, collected YACS user perspectives on this prototype, and then conducted a single-arm feasibility and acceptability study.

Results: In total, 26 YACS (6-7 participants per group) experiencing insomnia participated in focus groups to share experiences of insomnia during cancer survivorship and to provide input regarding a CBT-I prototype. Common triggers of insomnia included worry about disease management and progression, disease-related pain and other symptoms, choices regarding personal device use, and worry about the impact of poor sleep on daily functioning. In total, 12 participants completed device prototype testing, engaging with the prototype 94% of the assigned times (twice daily for 14 days; meeting predetermined feasibility cutoff of engagement ≥70% of assigned times) and rating the prototype with an overall mean score of 5.43 on the Satisfaction subscale of the Usability, Satisfaction, and Ease of Use scale (range 4.42-7; exceeding the predetermined cutoff score for acceptability of 5.0). All participants completing the study reported they would be interested in using the prototype again and would recommend it to someone else with insomnia.

XSL•FO
RenderX

Conclusions: YACS were highly engaged with our voice-activated virtual assistant–delivered CBT-I prototype and found it acceptable to use. Following final device development, future studies should evaluate the efficacy of this intervention among YACS.

Trial Registration: ClinicalTrials.gov NCT05875129; https://clinicaltrials.gov/study/NCT05875129

(JMIR Form Res 2025;9:e64869) doi:10.2196/64869

KEYWORDS

cancer; survivor; insomnia; cognitive behavioral therapy; technology; app; oncology; mobile health; artificial intelligence; young adults; sleep; mHealth; mobile health; CBT; voice-activated virtual assistant; virtual assistants; focus group; qualitative research

Introduction

Insomnia, characterized by the struggle to initiate or maintain sleep or the experience of nonrestorative sleep, brings along symptoms that hinder daily functioning, such as weariness, concentration difficulties, and mood disruptions [1]. Among the estimated 650,000 young adult cancer survivors (YACS, ages 18-39 years old) in the United States, up to 75% are grappling with insomnia symptoms, with nearly 30% likely to experience insomnia disorder [2-8]. Untreated insomnia is associated with a myriad of psychological and medical consequences, exacerbating the long-term impact of cancer diagnosis and treatment [3,4,9-11].

Insomnia adversely affects cardiometabolic and immune system health, neurobehavioral function, depression, fatigue, and overall quality of life [9,12-14]. These repercussions can also result in reduced physical activity and increased sedentary behavior, both linked to negative health outcomes in YACS [4,15]. In comparison to those with regular sleep patterns, cancer survivors dealing with insomnia tend to have more medical visits, double the hospitalization rate, and a higher reliance on medication [16]. Furthermore, insomnia serves as a risk factor for various mental illnesses, with the greatest risk ratios observed in depressive disorders. Despite its persistent nature during and after cancer treatment, discussions about insomnia with health care providers remain infrequent [17]. Addressing the need for accessible and effective insomnia treatment targeted to support this population with distinct cancer trajectories due to diagnosis at a young age is imperative.

Cognitive behavioral therapy for insomnia (CBT-I) is the gold standard intervention recommended by the National Comprehensive Cancer Network (NCCN) to address insomnia in cancer survivor populations [18-20]. Previous studies have demonstrated the effectiveness of CBT-I among YACSs [21,22]. However, widespread accessibility to CBT-I remains challenging due to an insufficient number of trained providers, a lack of insurance coverage, and even language barriers [23,24]. The administration of CBT-I demands specialized training, and its delivery, typically over 5-6 in-person sessions, can be time-consuming and expensive, creating multiple obstacles to care. Though CBT-I remains the gold standard intervention, adherence can be challenging due to a variety of patient and treatment-related factors as well as multiple changes to bedtime required by the CBT-I component called sleep restriction [25,26]. Even with shorter treatment plans of 3-4 sessions, dropout rates in trials have reached 20%-30%, suggesting potential challenges in real-world adherence [27,28].

In response to this, internet-based or app-delivered digital automated therapies have been developed and proven easier to access by patients, though many of these programs are more instructional than interactive, which may be less engaging for participants [29-31]. Such platforms can make therapy more flexible and accessible, particularly for individuals without access to in-person therapy or who prefer to engage with CBT-I at home. Several meta-analyses and reviews indicate that electronic delivery of CBT-I is just as effective as traditional, face-to-face approaches [29,30]. These programs are grounded in evidence-based practices and typically incorporate components like those found in traditional CBT-I [32,33]. However, there are concerns that the self-directed format of these programs might lead to lower treatment adherence [34]. Some critics also point to a lack of comprehensive quality control over these digital therapeutics as a potential issue [34,35]. Furthermore, these internet-delivered therapies require user engagement with digital screens right at bedtime, potentially confounding the physiologic process of melatonin secretion that enhances sleep [36].

Building on prior work [37], we hypothesized that we could address these issues—improving access to CBT-I, greater technology usability, and reducing screen-time use—at once by integrating 3 distinct technologies running custom software and that YACS would find this technology appealing to use. First, we aimed to deliver the most important therapeutic components of CBT-I interactively via an AI-assisted smart speaker system and without the use of mentally and emotionally activating, melatonin-suppressing screens at night. Second, we deployed smart lighting controls to facilitate CBT-I suggested changes to the room environment. Finally, we used a customized router that provided feedback to each participant about post-bedtime screen use. We supported the ability of YACS to make different choices about screen use at night. We believe these 3 technologies could work together in a customized package to meet the best elements of conversation CBT-I with new ways to support adherence through complex and lengthy treatment protocols. Thus, using an approach involving active stakeholder input and an iterative design process, we aimed to develop and test a novel prototype to increase sleep duration for YACS.

Methods

Participants and Recruitment

Recruitment occurred via patient referrals from 11 oncologists at a large urban academic medical center in Washington, DC,

XSL•FO
RenderX

from 2 regional oncology practices focused on YACS populations, 2 regional cancer survivor support groups, and Facebook social media advertising. The Washington, DC-based academic medical center annually provides oncology care to over 2000 adult patients living with cancer, including approximately 360 YACS. Regional young adult oncology practices included a large outpatient clinic in Northern Virginia and a similarly large clinic in Charlotte, NC. Regional cancer survivor support groups were based in Washington, DC, and suburban Maryland in the Baltimore-Washington, DC region. In all settings, approved recruitment flyers were placed strategically in clinic waiting areas or were shared directly with potential participants by providers (oncology clinics) or support staff (cancer survivor support groups). Members of the study team were available by phone or email to answer questions. Potential participants were screened for eligibility by the research team.

Participants were included if between 18 and 39 years old, living with a history of cancer (any type, any stage), self-reported to average less sleep than recommended for their age by the National Sleep Foundation for ≥3 months (7-9 hours per night) [1], could engage with the study prototype and complete surveys and interviews in English, and had Wi-Fi access at home. Individuals were excluded if they were enrolled in hospice care or lived outside the continental United States.

Focus Groups

Our research team followed the consolidated criteria for reporting qualitative research (COREQ) process through this qualitative study [38]. Qualitative data was gathered through 4 formative focus groups led by study authors (HG and DG), held virtually between December 2022 and March 2023. During recruitment, potential participants were informed that the team sought to learn more about the experience of chronic insomnia and gather perspectives on the use of mobile technology to improve sleep at home. Prior to the study, no study team member had any relationship with any participant. We included 26 participants (6 in focus group 1; 7 in group 2; 7 in group 3; and 6 in group 4). Prior to joining the virtual meeting, participants were asked to turn off their video cameras and deidentify their names (eg, “Guest” or “Participant”) making participation anonymous. Focus group questions and prompts—outlined in Table 1—were developed by members of the research team based on prior work with cancer survivors and tested with 2 YACS who were not part of the focus groups. Each focus group consisted of a brief introduction to the project followed by a facilitated conversation regarding participant experiences of insomnia and sleep disruption in cancer survivorship. Our inquiries explored participant insomnia experiences, elucidating perceived causes and symptoms. Participants were also asked about the current use of technology at home, particularly smart technology such as smart speakers and other voice-activated virtual assistants, screen use habits, and concerns about privacy related to smart technology use. Each focus group lasted approximately 90 minutes. Members of the research team continually reviewed focus group content and, after the fourth focus group, determined that content saturation had been achieved.

XSL•FO
RenderX

Table 1. Focus group questions aimed at understanding chronic insomnia of young adult cancer survivors and portable technology habits.

Theme	Questions
Perceptions and experiences of insomnia	• What is your insomnia experience like? How long has it been occurring? • How has insomnia affected your cancer journey?
Mitigation strategies for insomnia	• What steps have you taken to improve your sleep? What have you found to be helpful, if anything?
Device use and habits	• Which connected devices do you use every day?
Associated between device use and sleep habits	• To what extent do you think that the use of connected devices affects your sleep? • Have you taken any efforts to try and change your connected device use, specifically as it relates to how you sleep? • If yes, what have you found to be helpful? • If not, what would interest you to make changes?
VAVA[a] technology use and habits	• Do you use VAVAs at home or elsewhere? Examples include Apple Siri, Amazon Echo/Dot, or Google Assistant. • If you have not used VAVAs, why not?
Comfort using VAVA-delivered strategies to improve insomnia	• What are your thoughts on a smart speaker system giving you instructions like this about your sleep? • What other ways do you think that technology might be able to help make better choices to improve your sleep?
Familiarity with smart lighting and integration into sleep-wake experience	• What is your familiarity with smart lighting? • What are your thoughts on having smart lights installed in your bedroom that would automatically change in the evening to set up an optimal bedtime for you? • What if the smart lights also automatically changed prior to an optimal wake time?
Screen time usage and device prototype feedback about habits	• What are your thoughts on seeing weekly or monthly reports on how often your devices impact your sleep, and which devices have the most significant effect on your sleep? • Would you have privacy concerns about a system that provides feedback about your internet usage to you? • If yes, how do these privacy concerns compare to your privacy concerns about other apps you use and websites you visit? • Would those privacy concerns change your interest in using a system like this for better sleep? • What system would make you feel confident about privacy protection on a system like this?

[a]VAVA: voice-activated virtual assistant.

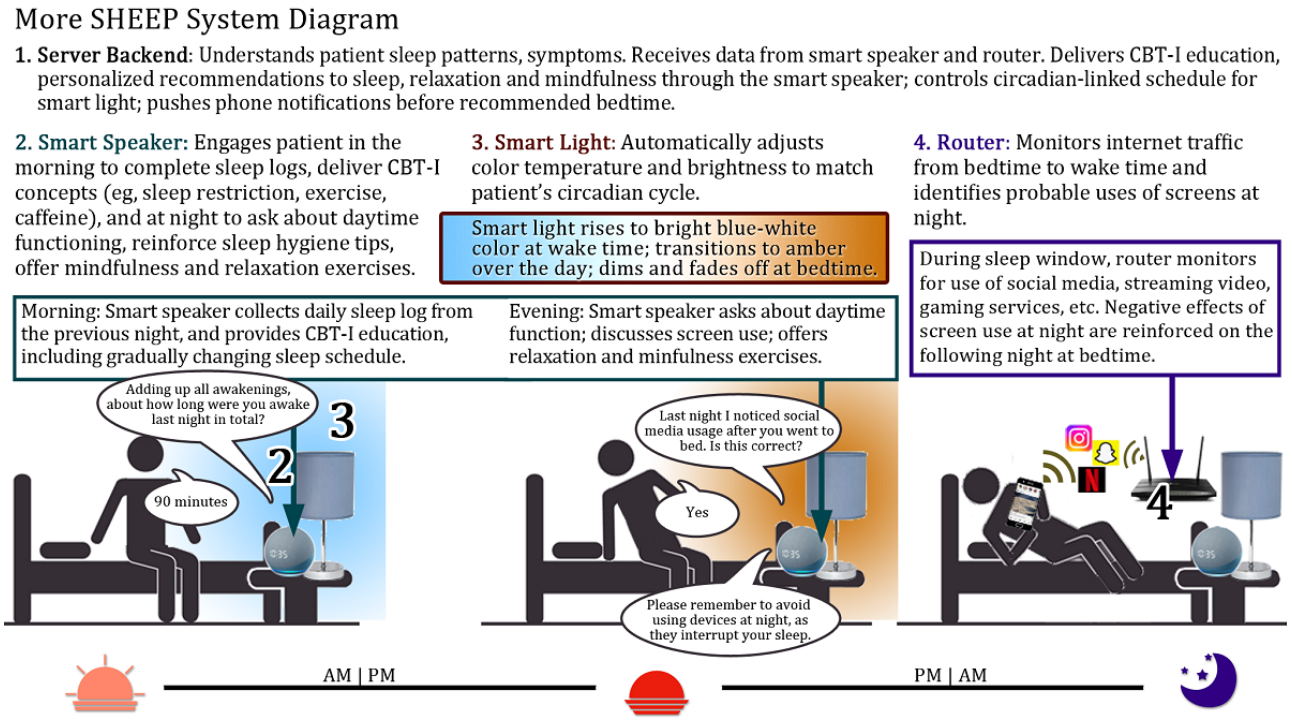
Device Prototype Development

The More SHEEP (More Sleep Hours Electronic Education Program) system includes proprietary Media Rez technology for delivering elements of CBT-I via natural voice interactions for more engaging patient education and patients can complete sleep logs without bedtime screen use (Figure 1). The team used the first 2 focus groups with YACS to design the systems and 2 follow-up focus groups to iteratively revise the software. The system was built around an AI-driven smart speaker using an Amazon Alexa device (Echo Dot 3rd Gen). In preparation for the 2-week trial, the technical team reduced the 6-session course of CBT-I to only 2 weeks by concentrating material from later weeks into the earlier sessions, with the goal of providing a comprehensive experience of virtual CBT-I sessions in only 2 weeks. The team also provided a lamp with a color-changing Philips smart bulb controlled by our custom software, which changed its on and off times based on each patient's unique sleep restriction and sleep consolidation recommendations according to the rules of CBT-I. As a result, the smart lamp bulbs automatically changed color temperature throughout the day. It began with a bright white with blue wavelengths in the morning to suppress melatonin production and reduce sleepiness and shifted to a warm amber in the evenings, so melatonin is not suppressed, and the patient can feel naturally sleepy [39,40]. The technical team also developed router control software to record and categorize the web-based traffic (social media, streaming video, and gaming) by the participant after bedtime (but not other members of the household, by having only the participant use the WiFi from the included custom router), so this screen use information could be presented to each patient to help them make better decisions about evening device use. For example, the system compares IP addresses visited by the patient with known sites like Netflix, Call of Duty, and Instagram. The system was also built to intelligently filter out other sources of internet use, like automatic updates and other forms of device telemetry. In this system, patient privacy and confidentiality are ensured since the system immediately

XSL•FO
RenderX

discards all identifiable information about post-bedtime website usage once that usage is shared with the patient. No information is collected about websites visited at any other time. Because this IP information is not stored, it cannot be summarized to participants or accessed at a future time, either by the participant or by a third party such as an internet hacker. No information is collected from other people in the household who are using conventional WiFi and not using this system. Together, these 3 elements combine to deliver a comprehensive, innovative sleep solution.

Figure 1. The More SHEEP (More Sleep Hours Electronic Education Program) system connects a smart speaker, smart lighting, and unique internet router to provide tailored cognitive behavioral therapy for insomnia (CBT-I) and individualized screen use data to individuals with chronic insomnia.



Device Testing

Following prototype development and refinement, we conducted a single-arm study to evaluate the feasibility and acceptability of the system. While the research team remained knowledgeable about each participant, the technology company Media Rez was blinded to participants to preserve anonymity. For this phase, we screened 27 potential participants for participation, 6 of whom expressed interest and then did not respond to further communication, 5 changed their minds about participation after hearing more information about the study, 1 person wished to proceed with consent but was about to leave the country for an extended travel period, 1 person decided not to participate due to an upcoming medical procedure, and 1 person was initially eligible and then became ineligible due to a hospice referral. We thus recruited and enrolled 15 eligible YACS who were not involved in the focus groups.

After completing written consent, participants received (either by mail or picked up in person) a study materials box containing the prototype smart speaker, smart lamp, router, and setup instructions. A member of the study team assisted with device setup via a brief phone call with each participant. Participants were instructed to engage with the prototype system twice daily for 14 days, including entering sleep logs, receiving smart speaker–delivered education about relevant content (eg, sleep hygiene, relaxation strategies), and reviewing personalized data regarding screen time use. If challenges occurred during system setup or use during the 14-day study period, a research team member contacted the tech team directly for troubleshooting advice to preserve the anonymity of the participants to the tech team.

Study feasibility was determined by user engagement with the prototype as assessed by: (1) the number of days over the study period that participants interacted with the system (≥70% interaction days used as cutoff) and (2) the average number of daily interactions with the prototype (≥2 daily interactions used as cutoff). Study acceptability was assessed by an average score ≥5 ("somewhat agree") on the Satisfaction subscale of the Usability, Satisfaction, and Ease of Use scale; this subscale is composed of 7 items measuring satisfaction with program content and the method of program delivery [41]. Additionally, all participants completed a semi-structured interview to assess overall user experience, with a focus on overall impressions of the system as well as each specific component, enthusiasm for longer-term use, and recommendations for device improvements.

We used a grounded theory approach for qualitative data analysis. Focus groups (formative phase) and semi-structured interviews (following device testing) were audio recorded and transcribed. Coding was led by HG (trained in qualitative analysis) with assistance from other team members (LA, LG) and reviewed by members of the research team. Initial coding of qualitative data was descriptive in nature, based on the interview guide structure. From the focus group data, emerging themes and exemplary quotes were extracted to develop the prototype prior to testing. Semi-structured interview data will

XSL•FO
RenderX

be used to further develop the device prior to future efficacy testing. All quantitative data (following device testing) were summarized using descriptive statistics.

Ethical Considerations

Per institutional guidelines at the primary study site, this study was first reviewed and approved by the Georgetown University Lombardi Comprehensive Cancer Center Scientific Review Committee prior to review and approval by the MedStar Health Research Institute Institutional Review Board (number 00005548). Focus group participants provided verbal consent to participate at the beginning of each focus group conducted. Participants for device feasibility testing provided written consent prior to the device prototype being shipped to the home. Per protocol approved by the IRB, all identifiable participant data was stored on password-protected, encrypted servers. Focus group participants each received a $30 gift card. Device testing participants were offered a choice to either (1) keep the study device after a factory reset or (2) receive a US $50 gift card.

Results

Focus Groups

In total, 26 YACS experiencing insomnia participated in focus groups to share experiences of insomnia during cancer survivorship and to assess the practicality of using this system for delivering components of CBT-I (Table 1). All 4 focus groups echoed similar experiences with insomnia, describing cancer survivorship as characterized by insufficient sleep and diminished energy, whether struggling to initiate sleep or stay asleep, and often finding themselves awake in the middle of the night. Common threads in perceived triggers and symptoms encompassed anxiety and worry about disease management and progression, disease-related pain and other symptoms, choices regarding personal device use, and worry about the impact of poor sleep on daily functioning (Table 2).

XSL•FO
RenderX

Table 2. Sample scripts for MORE Sheep modules

Module name	CBT-I component
Education	
• *Another thing I will do is automatically adjust the provided smart lamp. Tonight, beginning an hour before your bedtime of 10:00 P.M., the lamp will begin to fade out, and turn itself off just before bed. In the morning, it will turn on at its lowest brightness and slowly get brighter until it is fully bright at your wake-up time of 7:00 A.M. As I learn more about your sleep patterns, I may suggest changing when you sleep, and the lamp will act as a visual cue of your sleep schedule. In the coming days, I'll explain the effects of light, your circadian rhythm, and how your body uses a hormone called melatonin to regulate sleep.*	Psychoeducation
• *Congratulations! You left the bed four times this week when you couldn't sleep and listened to three relaxation scripts until you were tired enough to go back to the bedroom to sleep. This is good, because it helps you associate the bed only with sleep, and not with lying awake. In this case, sleep restriction could be helpful, making your total time in bed closer to your actual sleep time. Instead of a 10:00 pm bedtime, would you like to try going to bed at 10:45 for the next few nights?* • If Yes: *Good. I'll reset your alarms and lights for 10:45, and we'll see how that works out.*	Psychoeducation and stimulus control
Morning	
• *We worked together to establish a bedtime of 10:45 and wake time of 7 A.M. The hours I set for your sleep are not a lifelong sentence. It was merely the necessary step to get you to try to improve your sleep. As you may be starting to see, sleep restriction and stimulus control are meant to help consolidate your sleep to improve your ability to fall asleep and stay asleep.*	Sleep restriction
• *Rate the quality of your sleep last night from 1-5 (with one being the best and five being the worst). About how many times did you wake up last night?*	Sleep hygiene
Night	
• *Being active each day is important. Even a 10 minute walk is a great start! Would you like me to remind you to get moving during the day?* • If Yes: *Great. What time can I send you a push notification on the app to be active?* • User Time Input: *Ok, I'll send you a notification to be active each day at 1:30 P.M.*	Psychoeducation
• *Please rate your tiredness today on a scale of 1 to 5, with 1 being the least tired and 5 being the most tired.*	Sleep Hygiene
• *Last night I detected internet usage during your sleep window. Specifically, it appears you used devices late at night. This activity appears to include streaming video. Is this correct?* • If Yes: *Remember that this will hurt your sleep. Let me give you a tip for tonight: Establish a bedtime routine where you give yourself time to check devices one more time and then commit to keeping them off until morning. It should ideally be 30 minutes or more before bed. Give yourself the time before bed to view social media or watch a show, but promise yourself that once it is bedtime, your devices are off limits. Consider turning on a "do not disturb" mode, or fully power off these devices, and keep them away from the bed.*	Psychoeducation, Sleep Hygiene
Relaxation:	
• *For this first relaxation exercise, I'd like to play a few samples of each type. In the future, I'll help you find which type you enjoy most.*	Relaxation

> *I think most times I get [insomnia] it was a little bit combination of stress and having emotional traumas that led to sleeplessness.*
>
> *What causes me to stay awake is when I am having some level of depression and anxiety.*
>
> *At the end of the day, [if] you sleep one or two hours and wake up, the experience is bad. A lot of people are looking for how you can get yourself calm. There is a lot of anxiety and depression [living with cancer].*

Use of personal technologies was frequently cited as a trigger for insomnia:

> *Sometimes these [screen] devices become a problem for me in terms of sleeping. I put them in Airplane mode so that I can get better sleep… [but if] I forget to put it on Airplane mode and if my phone disturbs me, I find it difficult to fall asleep again.*
>
> *In order not to get distracted, sometimes [I find myself] setting a limited time to spend with my device and try to adhere to it. And when it is time [to stop phone use] I know it's time to quit time to go to bed, though it is really not easy.*

Common strategies to mitigate sleep disturbances included web- or app-based meditation programs to calm anxiety, worry, or distraction through screen devices. Participants frequently noted a reluctance to seek pharmacotherapies for insomnia, voicing a preference for nonpharmacologic interventions.

XSL•FO
RenderX

> *I try to use [apps to] wind down stuff on my phone. I try to make sure that [I] turn off any screen usage and make sure that I am on my bed getting ready to sleep.*
>
> *I think about a smart speaker for therapy, it is something very unique. It [makes] you feel like [you are in] a relationship, like talking to someone you know. So, for me personally I think it is going to help positively to help [instead of] the stress of going to see therapist from day to day, time to time, consider appointments, traffic, and out of pockets, financially.*
>
> *You know when you have such a device around you, it can literally put you back to sleep. You do not need pills.*

Participants were familiar with smart speakers or voice-activated devices. All participants reported using smartphone-based voice functions, like Apple Siri, for straightforward tasks (eg, creating a calendar reminder) or information gathering (eg, local weather). A minority used in-home smart speakers, primarily to play music or deliver news. All participants expressed interest in the concept of AI-delivered CBT-I and agreed that a smart speaker delivery system could promote frequent engagement with this therapy in the convenience of home. One participant was familiar with smart lighting but did not use it regularly. The remaining participants expressed interest in incorporating smart lighting into a home-based insomnia program.

> *You know when you are in a room where the lighting systems feels comfortable for sleep, you realize you can sleep more better and easier than compared to when we are in a room whereby you get you know more blue light. So that tends to be a bit of a problem to sleep and also affects the entire sleeping pattern.*
>
> *My thought here is that having a smart lighting in the bedroom is what will improve our sleep, our sleeping pattern. Because a lot of people have some type of insomnia because of depression, anxiety, or one thing or the other. But I believe with the installation of smart lighting, the color changes, and the variation of the lighting system, patients might be forced to sleep at a particular time.*

Regarding the additional internet router function to monitor and provide feedback about personal device use, all agreed such data could promote making different choices about screen time use. A minority of participants expressed some concern about privacy with router monitoring of internet usage but also expressed acceptance with additional information about prototype privacy protection.

> *Anything that has to do with privacy, you have to be very careful because you don't want [personal information] to go out there and fall into the wrong hands.*
>
> *At least every device connected to a router must have its own IP address. So, I don't know because each of the IP addresses must have an IP address which can be used to access each of the devices that is connected to a router. I don't know how [you are] going to secure that in terms of privacy of information or data.*

No participants expressed worry about privacy due to the smart speaker itself. A few participants appreciated the possibility of being more open with AI technology than with an actual in-person therapist.

> *It is like meeting a therapist, it's more like talking to a human, it actually saves us the stress of having to access any health care and I'm talking about having discussion about insomnia or sleep disorder. And it is safe for me having it around me in my [home] environment…goes a long way to have a kind of communication level that will ease the tension of having to say whatever you are feeling.*
>
> *The most interesting part of it to me is that the therapist's [voice] in the device can make you feel comfortable when you are trying to make sure you get good sleep.*

Device Testing

In total, 15 participants had an average age of 30.6 (SD 6.8); 10 identified as women; 8 identified as Caucasian, 7 identified as Black or African American, and 1 identified as Hispanic. Of 15 enrolled participants, 2 became lost to follow-up immediately after enrolling (sent the device prototype but never initiated the program), and 1 completed 1 day of participation and then dropped out from concerns about using home Wi-Fi for the project as they were also using high-security home Wi-Fi features for work. The remaining 12 participants completed the study protocol, including a survey and a semi-structured interview.

Participants were able to adhere to the More SHEEP program directions, engaging with the prototype 94% of the assigned times (twice daily for 14 days) and meeting the predetermined feasibility cutoff of engagement ≥70% of assigned times. Participants also found the More SHEEP prototype acceptable, with an overall mean score of 5.43 on the Satisfaction subscale of the Usability, Satisfaction, and Ease of Use scale (range 4.42-7), exceeding the predetermined cutoff score for acceptability of 5.0. All participants completing the program reported they would be interested in using the prototype again, would recommend it to someone else with insomnia, and would be interested in participating in future studies evaluating new iterations of the prototype system.

Following participation, interview themes informed individual experiences with each of the More SHEEP system components (smart speaker, smart lighting, router, and More SHEEP smartphone app) (Table 3).

XSL•FO
RenderX

Table 3. The authors enrolled 15 young adult cancer survivors to participate in a single-arm feasibility-acceptability study of the More SHEEP prototype that included qualitative data collection to inform future prototype development. Participant-generated themes inform prototype feasibility, acceptability, and areas for potential improvement.

Prompt	Content
Familiarity with smart device use	• *I have another [smart speaker] device and my grandfather has one too because he's blind. This one definitely had a more pleasant voice (Participant #1)*
Feasibility of use	• *The app reminded me twice a day on my phone and it was easy to spend a few minutes with the program before bed or after getting up in the morning. No problem at all (#3)* • *For the most part I found [the smart speaker] easy other than when I asked for something that it didn't know how to give me. It understood what I was talking about and was easy to use (#12)*
Acceptability of use	• *I thought it was all pretty easy to use and get used to. I even tried taking it on a visit to family and it worked there for me too. (#6)* • *I do the check-ins each morning and evening. I think I remember to do it almost every day, so it's pretty easy to do, particularly with it being just audio, rather than having to go to a website and read [material]*
Challenges with smart lamp	• *I ended up not using the lamp at all because my partner goes to bed at different times so it was disruptive. The rest was easy to use (#4)* • *I did sometimes modify my answers to make the smart lamp go off a little earlier…I would find myself throwing a blanket over the top (#1)*
Relevance of CBT-I[a] language and setting	• *Occasionally, the [software] voice mentioned something related to cancer survivorship and I really felt that this program was designed for someone like me (#2)* • *I think the awareness [of my sleep choices] was one thing I wasn't expecting to just pick up on…being more aware of how I felt [around bedtime] (#1)* • *I feel like it had surprisingly little survivorship information. For the most part, I felt like I was learning general sleep habits (#12)* • *I did think that this [home system] is very convenient as somebody who has three small kids and cancer and a husband so I don't have to go anywhere and I feel like I'm getting something from these sessions (#11)*
Preference for more customization	• *It would be really cool if there was a way to make the lamp different colors maybe depending on the education happening or on my mood. Maybe combine the speaker and the lamp into one, that would be good too (#7)* • *I like being able to use sounds at night to sleep. I put the thunderstorm [track] on, or the ocean, and I'm able to sleep (#14)* • *It would definitely be nice if [the system] could say, if you could ask how much time you have or something like that*
Need to design for varying spaces or needs	• *I'm in college and have roommates so I wasn't sure if I should try this at school or at home. I'm participating from home but if I used this at school it would have to be different (#10)* • *I have a 4-year-old and she can't sleep by herself so she's coming into my room; then I think, can I activate Alexa without waking her up?*

[a]CBT-I: cognitive behavioral therapy for insomnia.

Discussion

Principal Findings

We discovered a keen interest among participants in this technology in this preliminary investigation into the experiences of YACS dealing with insomnia and smart home devices, particularly when tailored for user-friendly personalization. Most participants expressed significant interest in using this technology. Participants generally reported an enhanced understanding of CBT-I and its potential impact on mitigating chronic sleep disruption and rated the program as both feasible and acceptable. Participants highlighted the high feasibility of daily sleep logs during our demonstration, a crucial component of successful engagement with CBT-I. This observation implies that the data gathered by the smart home device can be leveraged by AI programming to generate personalized recommendations and schedules, surpassing the scope of conventional sleep hygiene education.

Feedback from participants also informed areas for prototype improvement. Some participants felt the More SHEEP content could speak more to the YACS experience of insomnia and could include more information and recommendations directed specifically to cancer survivorship; in collaboration with young adult cancer survivorship experts, future device iterations could include more language targeted to YACS. Participants often noted that a smart speaker or smart lighting may not be conducive to sharing sleep space. For example, the smart light might remain on to address sleep restrictions per CBT-I protocol for a participant while a bed partner wants to sleep without lighting. Similarly, a participant might not want to awaken others using a smart speaker at bedtime—this was particularly important to 1 participant living in a university dormitory. Finally, some participants expressed preferences for more sleep-related content, like sleep music or white noise, allowing

XSL•FO
RenderX

more customizability. Formative testing processes of future prototype iterations may address these concerns, ultimately enhancing participant engagement and treatment adherence.

Developments in home technology play an increasing role in improving insomnia and related symptoms by providing tools and solutions to address sleep disturbances. Examples include but are not limited to mobile sleep-tracking devices, light therapy, smart thermostats, biometric feedback devices, and relaxation or meditation apps [42-46]. As a gold standard for the management of chronic insomnia, implementation of CBT-I continues to be explored in this space, primarily through mobile device applications [47-51]. Nevertheless, even through the use of mobile technologies, maximizing treatment adherence remains challenging. In total, 1 study specifically focused on YACS, delivering an internet-based CBT-I intervention over 6-9 weeks and finding that, while outcomes of insomnia, daytime sleepiness, fatigue, and quality of life improved at the study's 2 timepoints, most participants did not complete the assigned protocol [52]. Compared to this and other device-driven insomnia mitigation strategies, our prototype is unique in its CBT-I delivery through a smart speaker, limiting device screen exposure whose blue wavelength light can impede nighttime melatonin secretion, as well as smart lighting use that can provide important diurnal visual cues (and block blue wavelength light at night). While most of our study participants had never used these technologies before, acceptability was high, and participants were able to easily engage with the program structure. While we believe that our prototype's inherent ease of use through voice activation in the bedroom will improve long-term CBT-I treatment adherence, we did not design this early study to evaluate this.

Several limitations of this project deserve mention. First, our insights come from a small number of participants, primarily due to the constraints of the study's limited timeframe. Nevertheless, this sample size appropriately aligns with stage I (intervention refinement) of the NIH stage model for behavioral intervention development. Additionally, as we reached the conclusion of data collection, the absence of emerging additional themes in our qualitative data suggested a saturation of feedback. Second, we note selection bias, as some eligible YACS learned of the study but chose not to participate. We also were not able to complete data collection for 2 enrolled participants. Furthermore, although some information was revealed during post-participation interviews, we did not collect data regarding other factors that might influence the choice to enroll or study adherence, such as income, living situation, or education that might influence outcomes. Given the nature of the intervention, it is possible these same factors could also influence clinical outcomes related to insomnia—thus, future efficacy studies of our prototype should, at a minimum, collect such patient-level data and potentially ensure diversity of participants through purposive recruiting strategies, increasing generalizability of results. By design, this study intentionally did not evaluate clinical outcomes such as improvement in insomnia, mood, daytime fatigue, etc, focusing specifically on device feasibility of use and acceptability over 14 days. While this time frame facilitated participant experience with the More SHEEP prototype, a full CBT-I treatment typically requires approximately 6 weeks; therefore, future prototype refinement may benefit from piloting a longer period of use at home. Nonetheless, this early development phase focused on evaluating device feasibility and acceptability and we believe we have achieved our primary goal. Subsequent studies will delve deeper, refining and testing the prototype for practical feasibility and effectiveness.

Conclusions

This study combined end-user participant input toward the development of a novel smart speaker, smart lighting, or a router system to deliver key elements of CBT-I for YACS to be used in the home setting. Now that the system has demonstrated feasibility and met acceptability benchmarks, the prototype can continue to be developed to incorporate suggestions from our participants and subsequently be tested for efficacy in the YACS population. Ultimately, this system has the potential to increase much-needed access to CBT-I among cancer survivors.

Acknowledgments

The authors thank the participants for generously sharing their time, interest, and input. No generative artificial intelligence was used in the preparation of this manuscript.

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

Media Rez is a for-profit company that seeks to commercialize an eventual product based on this technology, consistent with the National Institutes of Health Small Business Innovation Research program.

References

1. The National Sleep Foundation. URL: https://www.thensf.org [accessed 2024-11-14]
2. Armenian SH, Chao C. Burden of morbidity and mortality in adolescent and young adult cancer survivors. J Clin Oncol 2024;42(6):735-742. [doi: 10.1200/jco.23.01751]

XSL•FO
RenderX

3. Daniel LC, Aggarwal R, Schwartz LA. Sleep in adolescents and young adults in the year after cancer treatment. J Adolesc Young Adult Oncol 2017;6(4):560-567. [doi: 10.1089/jayao.2017.0006] [Medline: 28628351]
4. Zhou ES, Recklitis CJ. Insomnia in adult survivors of childhood cancer: a report from project REACH. Support Care Cancer 2014;22(11):3061-3069. [doi: 10.1007/s00520-014-2316-y] [Medline: 24935649]
5. Hinds PS, Hockenberry MJ, Gattuso JS, Srivastava DK, Tong X, Jones H, et al. Dexamethasone alters sleep and fatigue in pediatric patients with acute lymphoblastic leukemia. Cancer 2007;110(10):2321-2330 [FREE Full text] [doi: 10.1002/cncr.23039] [Medline: 17926333]
6. Zupanec S, Jones H, Stremler R. Sleep habits and fatigue of children receiving maintenance chemotherapy for ALL and their parents. J Pediatr Oncol Nurs 2010;27(4):217-228. [doi: 10.1177/1043454209358890] [Medline: 20562390]
7. Daniel LC, Wang M, Mulrooney DA, Srivastava DK, Schwartz LA, Edelstein K, et al. Sleep, emotional distress, and physical health in survivors of childhood cancer: a report from the childhood cancer survivor study. Psychooncology 2019;28(4):903-912 [FREE Full text] [doi: 10.1002/pon.5040] [Medline: 30817058]
8. Ameringer S, Erickson JM, Macpherson CF, Stegenga K, Linder LA. Symptoms and symptom clusters identified by adolescents and young adults with cancer using a symptom heuristics app. Res Nurs Health 2015;38(6):436-448. [doi: 10.1002/nur.21697] [Medline: 26455729]
9. Olson K. Sleep-related disturbances among adolescents with cancer: a systematic review. Sleep Med 2014;15(5):496-501. [doi: 10.1016/j.sleep.2014.01.006] [Medline: 24726572]
10. Mulrooney DA, Ness KK, Neglia JP, Whitton JA, Green DM, Zeltzer LK, et al. Fatigue and sleep disturbance in adult survivors of childhood cancer: a report from the childhood cancer survivor study (CCSS). Sleep 2008;31(2):271-281 [FREE Full text] [doi: 10.1093/sleep/31.2.271] [Medline: 18274275]
11. Clanton NR, Klosky JL, Li C, Jain N, Srivastava DK, Mulrooney D, et al. Fatigue, vitality, sleep, and neurocognitive functioning in adult survivors of childhood cancer: a report from the childhood cancer survivor study. Cancer 2011;117(11):2559-2568 [FREE Full text] [doi: 10.1002/cncr.25797] [Medline: 21484777]
12. Hsu CY, Chen YT, Chen MH, Huang CC, Chiang CH, Huang PH, et al. The association between insomnia and increased future cardiovascular events: a nationwide population-based study. Psychosom Med 2015;77(7):743-751. [doi: 10.1097/PSY.0000000000000199] [Medline: 26355726]
13. Hall MH, Smagula SF, Boudreau RM, Ayonayon HN, Goldman SE, Harris TB, et al. Association between sleep duration and mortality is mediated by markers of inflammation and health in older adults: the health, aging and body composition study. Sleep 2015;38(2):189-195 [FREE Full text] [doi: 10.5665/sleep.4394] [Medline: 25348127]
14. Pigeon WR, Bishop TM, Krueger KM. Insomnia as a precipitating factor in new onset mental illness: a systematic review of recent findings. Curr Psychiatry Rep 2017;19(8):44. [doi: 10.1007/s11920-017-0802-x] [Medline: 28616860]
15. Zhou ES, Manley PE, Marcus KJ, Recklitis CJ. Medical and psychosocial correlates of insomnia symptoms in adult survivors of pediatric brain tumors. J Pediatr Psychol 2016;41(6):623-630. [doi: 10.1093/jpepsy/jsv071] [Medline: 26275977]
16. Ancoli-Israel S. Recognition and treatment of sleep disturbances in cancer. J Clin Oncol 2009;27(35):5864-5866. [doi: 10.1200/jco.2009.24.5993]
17. Khater W, Masha'al D, Al-Sayaheen A. Sleep assessment and interventions for patients living with cancer from the patients' and nurses' perspective. Int J Palliat Nurs 2019;25(7):316-324. [doi: 10.12968/ijpn.2019.25.7.316] [Medline: 31339821]
18. Management of Chronic Insomnia Disorder in Adults. A clinical practice guideline from the American College of Physicians. Ann Intern Med 2016;165(2):33-125. [doi: 10.7326/p16-9016]
19. Sanft T, Denlinger CS, Armenian S, Baker KS, Broderick G, Demark-Wahnefried W, et al. NCCN guidelines insights: survivorship, version 2.2019. J Natl Compr Canc Netw 2019;17(7):784-794 [FREE Full text] [doi: 10.6004/jnccn.2019.0034] [Medline: 31319383]
20. Blumenstein KG, Brose A, Kemp C, Meister D, Walling E, DuVall AS, et al. Effectiveness of cognitive behavioral therapy in improving functional health in cancer survivors: a systematic review and meta-analysis. Crit Rev Oncol Hematol 2022;175:103709. [doi: 10.1016/j.critrevonc.2022.103709] [Medline: 35580765]
21. Reynolds SA, Ebben MR. The cost of insomnia and the benefit of increased access to evidence-based treatment: cognitive behavioral therapy for insomnia. Sleep Med Clin 2017;12(1):39-46. [doi: 10.1016/j.jsmc.2016.10.011] [Medline: 28159096]
22. Koffel E, Bramoweth AD, Ulmer CS. Increasing access to and utilization of cognitive behavioral therapy for insomnia (CBT-I): a narrative review. J Gen Intern Med 2018;33(6):955-962 [FREE Full text] [doi: 10.1007/s11606-018-4390-1] [Medline: 29619651]
23. Thomas A, Grandner M, Nowakowski S, Nesom G, Corbitt C, Perlis ML. Where are the behavioral sleep medicine providers and where are they needed? A geographic assessment. Behav Sleep Med 2016;14(6):687-698 [FREE Full text] [doi: 10.1080/15402002.2016.1173551] [Medline: 27159249]
24. Vargas I, Egeler M, Walker J, Benitez DD. Examining the barriers and recommendations for integrating more equitable insomnia treatment options in primary care. Front Sleep 2023;2. [doi: 10.3389/frsle.2023.1279903] [Medline: 39210962]
25. Matthews EE, Arnedt JT, McCarthy MS, Cuddihy LJ, Aloia MS. Adherence to cognitive behavioral therapy for insomnia: a systematic review. Sleep Med Rev 2013;17(6):453-464 [FREE Full text] [doi: 10.1016/j.smrv.2013.01.001] [Medline: 23602124]

XSL•FO
RenderX

26. Mellor A, Kavaliotis E, Mascaro L, Drummond SPA. Approaches to the assessment of adherence to CBT-I, predictors of adherence, and the association of adherence to outcomes: a systematic review. Sleep Med Rev 2022;63:101620. [doi: 10.1016/j.smrv.2022.101620] [Medline: 35398650]
27. Zhou ES, Partridge AH, Recklitis CJ. A pilot trial of brief group cognitive-behavioral treatment for insomnia in an adult cancer survivorship program. Psychooncology 2017;26(6):843-848. [doi: 10.1002/pon.4096] [Medline: 26872123]
28. Buysse DJ, Germain A, Moul DE, Franzen PL, Brar LK, Fletcher ME, et al. Efficacy of brief behavioral treatment for chronic insomnia in older adults. Arch Intern Med 2011;171(10):887-895. [doi: 10.1001/archinternmed.2010.535] [Medline: 21263078]
29. Zachariae R, Lyby MS, Ritterband LM, O'Toole MS. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia - a systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev 2016;30:1-10. [doi: 10.1016/j.smrv.2015.10.004] [Medline: 26615572]
30. Seyffert M, Lagisetty P, Landgraf J, Chopra V, Pfeiffer PN, Conte ML, et al. Internet-delivered cognitive behavioral therapy to treat insomnia: a systematic review and meta-analysis. PLoS One 2016;11(2):e0149139. [doi: 10.1371/journal.pone.0149139] [Medline: 26867139]
31. Hall DL, Arditte Hall KA, Gorman MJ, Comander A, Goldstein MR, Cunningham TJ, et al. The survivorship sleep program (SSP): a synchronous, virtual cognitive behavioral therapy for insomnia pilot program among cancer survivors. Cancer 2022;128(7):1532-1544 [FREE Full text] [doi: 10.1002/cncr.34066] [Medline: 34914845]
32. Rowan AB, Magnante AT, Urh N, Figueroa L. Cognitive behavioral therapy for insomnia self-management mobile apps: a review of efficacy and quality. J Clin Psychol Med Settings 2024;31(3):537-549. [doi: 10.1007/s10880-023-09992-y]
33. Bai N, Cao J, Zhang H, Liu X, Yin M. Digital cognitive behavioural therapy for patients with insomnia and depression: a systematic review and meta-analysis. J Psychiatr Ment Health Nurs 2024;31(4):654-667. [doi: 10.1111/jpm.13024] [Medline: 38226714]
34. Simon L, Steinmetz L, Feige B, Benz F, Spiegelhalder K, Baumeister H. Comparative efficacy of onsite, digital, and other settings for cognitive behavioral therapy for insomnia: a systematic review and network meta-analysis. Sci Rep 2023;13(1):1929 [FREE Full text] [doi: 10.1038/s41598-023-28853-0] [Medline: 36732610]
35. Thorndike FP, Morin CM, Ojile J, Edington S, Gerwien R, Ong JC, et al. Effect of a prescription digital therapeutic for chronic insomnia on post-treatment insomnia severity, depression, and anxiety symptoms: results from the real-world DREAM study. Front Psychiatry 2024;15:1450615 [FREE Full text] [doi: 10.3389/fpsyt.2024.1450615] [Medline: 39319356]
36. Figueiro MG, Wood B, Plitnick B, Rea MS. The impact of light from computer monitors on melatonin levels in college students. Neuro Endocrinol Lett 2011;32(2):158-163. [Medline: 21552190]
37. Arem H, Scott R, Greenberg D, Kaltman R, Lieberman D, Lewin D. Assessing breast cancer survivors' perceptions of using voice-activated technology to address insomnia: feasibility study featuring focus groups and in-depth interviews. JMIR Cancer 2020;6(1):e15859 [FREE Full text] [doi: 10.2196/15859] [Medline: 32348274]
38. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
39. Lockley SW, Brainard GC, Czeisler CA. High sensitivity of the human circadian melatonin rhythm to resetting by short wavelength light. J Clin Endocrinol Metab 2003;88(9):4502-4505. [doi: 10.1210/jc.2003-030570] [Medline: 12970330]
40. Bourgin P, Hubbard J. Alerting or somnogenic light: pick your color. PLoS Biol 2016;14(8):e2000111 [FREE Full text] [doi: 10.1371/journal.pbio.2000111] [Medline: 27525420]
41. Gao M, Kortum P, Oswald F. Psychometric Evaluation of the USE (Usefulness, Satisfaction, and Ease of use) Questionnaire for Reliability and Validity. 2018 Presented at: Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 2018 September 27; United States. [doi: 10.1177/1541931218621322]
42. Lin LY, Tam KW, Huang TW. Effect of bright light therapy on cancer-related fatigue and related symptoms: a systematic review and meta-analysis of randomized controlled trials. J Psychosom Res 2023;174:111501. [doi: 10.1016/j.jpsychores.2023.111501] [Medline: 37797569]
43. Jalali N, Sahu KS, Oetomo A, Morita PP. Usability of smart home thermostat to evaluate the impact of weekdays and seasons on sleep patterns and indoor stay: observational study. JMIR mHealth uHealth 2022;10(4):e28811 [FREE Full text] [doi: 10.2196/28811] [Medline: 35363147]
44. Nguyen A, Pogoncheff G, Dong BX, Bui N, Truong H, Pham N, et al. A comprehensive study on the efficacy of a wearable sleep aid device featuring closed-loop real-time acoustic stimulation. Sci Rep 2023;13(1):17515 [FREE Full text] [doi: 10.1038/s41598-023-43975-1] [Medline: 37845236]
45. Huberty J, Bhuiyan N, Eckert R, Larkey L, Petrov M, Todd M, et al. Insomnia as an unmet need in patients with chronic hematological cancer: protocol for a randomized controlled trial evaluating a consumer-based meditation app for treatment of sleep disturbance. JMIR Res Protoc 2022;11(7):e39007 [FREE Full text] [doi: 10.2196/39007] [Medline: 35776489]
46. Park S, Lee SW, Han S, Cha M. Clustering insomnia patterns by data from wearable devices: algorithm development and validation study. JMIR mHealth uHealth 2019;7(12):e14473 [FREE Full text] [doi: 10.2196/14473] [Medline: 31804187]
47. Subotic-Kerry M, Werner-Seidler A, Corkish B, Batterham PJ, Sicouri G, Hudson J, et al. Protocol for a randomised controlled trial evaluating the effect of a CBT-I smartphone application (Sleep Ninja) on insomnia symptoms in children. BMC Psychiatry 2023;23(1):684 [FREE Full text] [doi: 10.1186/s12888-023-05185-x] [Medline: 37730577]

XSL•FO
RenderX

48. Jeon JM, Ma J, Kwak P, Dang B, Buleje I, Ancoli-Israel S, et al. Developing a novel mobile application for cognitive behavioral therapy for insomnia for people with schizophrenia: integration of wearable and environmental sleep sensors. Sleep Breath 2024;28(3):1491-1498 [FREE Full text] [doi: 10.1007/s11325-023-02980-4] [Medline: 38177830]
49. Chan CS, Wong CYF, Yu BYM, Hui VKY, Ho FYY, Cuijpers P. Treating depression with a smartphone-delivered self-help cognitive behavioral therapy for insomnia: a parallel-group randomized controlled trial. Psychol Med 2021;53(5):1799-1813. [doi: 10.1017/s0033291721003421]
50. Watanabe Y, Kuroki T, Ichikawa D, Ozone M, Uchimura N, Ueno T. Effect of smartphone-based cognitive behavioral therapy app on insomnia: a randomized, double-blind study. Sleep 2023;46(3):zsac270. [doi: 10.1093/sleep/zsac270] [Medline: 36355920]
51. Luyster FS, Ritterband LM, Sereika SM, Buysse DJ, Wenzel SE, Strollo PJ. Internet-based cognitive-behavioral therapy for insomnia in adults with asthma: a pilot study. Behav Sleep Med 2020;18(1):10-22. [doi: 10.1080/15402002.2018.1518229] [Medline: 30252506]
52. Zhou ES, Recklitis CJ. Internet-delivered insomnia intervention improves sleep and quality of life for adolescent and young adult cancer survivors. Pediatr Blood Cancer 2020;67(9):e28506. [doi: 10.1002/pbc.28506] [Medline: 32568460]

Abbreviations

CBT-I: cognitive behavioral therapy for insomnia
More SHEEP: More Sleep Hours Electronic Education Program
YACS: young adult cancer survivor

Edited by A Mavragani; submitted 29.07.24; peer-reviewed by S Joshi; comments to author 04.10.24; revised version received 22.10.24; accepted 30.10.24; published 10.03.25.

Please cite as:
Groninger H, Arem H, Ayangma L, Gong L, Zhou E, Greenberg D
Development of a Voice-Activated Virtual Assistant to Improve Insomnia Among Young Adult Cancer Survivors: Mixed Methods Feasibility and Acceptability Study
JMIR Form Res 2025;9:e64869
URL: https://formative.jmir.org/2025/1/e64869
doi:10.2196/64869
PMID:

©Hunter Groninger, Hannah Arem, Lylian Ayangma, Lisa Gong, Eric Zhou, Daniel Greenberg. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Zoom-Delivered Empowered Relief for Chronic Pain: Observational Longitudinal Pilot Study Exploring Feasibility and Pain-Related Outcomes in Patients on Long-Term Opioids

Karlyn A Edwards[1], PhD; Troy C Dildine[2], PhD; Dokyoung S You[2], PhD; Ashley M Herrick[2], BS; Beth D Darnall[2], PhD; Sean C Mackey[2], MD, PhD; Maisa S Ziadni[2], PhD

[1]Division of General Internal Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[2]Division of Pain Medicine, Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University, Stanford, CA, United States

Corresponding Author:
Maisa S Ziadni, PhD
Division of Pain Medicine
Department of Anesthesiology, Perioperative and Pain Medicine
Stanford University
1070 Arastradero Rd, Suite 200
Stanford, CA, 94304
United States
Phone: 1 650 497 7671
Email: mziadni@stanford.edu

Abstract

Background: Patients with chronic pain on long-term opioid therapy often face barriers to accessing effective nonpharmacological treatments, including the burden of multiple sessions, lack of trained clinicians, and travel time. Empowered Relief (ER), a 2-hour, single-session pain relief skills class, can improve pain and quality of life among patients with chronic pain when delivered in person or virtually.

Objective: This study examined the impact of Zoom-delivered ER among people with chronic pain on long-term opioid therapy. We assessed (1) the feasibility and acceptability of Zoom-delivered ER; (2) changes in pain and opioid use outcomes at 3 and 6 months after treatment; and (3) daily associations among pain, opioid dose, and the Pain Catastrophizing Scale (PCS) before and after treatment.

Methods: During the early COVID-19 pandemic, we conducted an uncontrolled pilot study of a Zoom-delivered ER among 60 adults (n=45, 76% female participants; n=52, 88% White participants) experiencing chronic pain who were receiving daily prescribed opioids (≥10 morphine-equivalent daily dose). Participants completed assessments at enrollment, before class, after class, 3 months after treatment, and 6 months after treatment. Furthermore, participants completed 2 daily assessment periods (spanning 14 consecutive days) before and after the class. We used a multilevel modeling approach to examine (1) the raw changes in PCS, average pain intensity, pain interference, and self-reported opioid dose at 3 and 6 months after treatment and (2) daily-level changes in average pain intensity and opioid dose before and after the class.

Results: Of the 60 participants enrolled, 41 (68%) attended the class and 24 (59% of the 41 class attendees) reported satisfaction with the Zoom-delivered class. PCS score was significantly reduced at 3 months (β=–3.49, *P*=.01; Cohen *d*=0.35) and 6 months after treatment (β=–3.61, *P*=.01; Cohen *d*=0.37), and pain intensity was significantly reduced at 3 months (β=–0.56, *P*=.01; Cohen *d*=0.39) compared to enrollment. There were no significant reductions in pain interference or opioid dose. Across daily assessments, higher daily pain catastrophizing was associated with worse daily pain (β=.42, *P*<.001) and higher self-reported opioid use (β=3.14, *P*<.001); daily pain intensity significantly reduced after the class (β=–.50, *P*<.001). People taking prescribed opioids as needed trended toward decreasing their daily opioid use after the class (β=–9.31, *P*=.02), although this result did not survive correction for multiplicity.

Conclusions: Improvements to future Zoom-delivered ER iterations are needed to improve feasibility and acceptability among people with chronic pain and daily prescribed opioid use. Despite this, findings show a promising preliminary impact of the intervention on pain outcomes. A larger randomized controlled trial of Zoom-delivered ER among this patient population is currently under way.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e68292) doi:10.2196/68292

KEYWORDS

Empowered Relief; single session; chronic pain; prescription opioids; telehealth; daily data; pain intensity; pain catastrophizing

Introduction

Background

One in 5 Americans has chronic pain [1], a condition associated with substantial personal and societal costs. Annual expenditures related to chronic pain, including lost productivity, exceed half a trillion dollars [2]. Individuals with chronic pain frequently report impacts on physical, psychological, and social functioning [3]. Over 100,000 Americans receive daily prescribed opioids [4], and while they can be an effective pain treatment for some [5], this treatment approach can also present several health risks [6], including overdose and death [7]. Furthermore, prolonged opioid use can contribute to mental health challenges, such as depression [8]. Thus, the gold standard for chronic pain management is a multimodal approach that integrates medical, behavioral, and rehabilitative treatments to minimize opioid use and maximize patient functioning [9,10].

Behavioral pain treatments are low-risk approaches that may reduce opioid-related risks and mitigate the broader impacts of chronic pain [11]. For example, cognitive behavioral therapy (CBT) alone [12], and when combined with posttreatment therapeutic interactive voice response telephonic support [13], has been shown to reduce opioid misuse. Moreover, acceptance and commitment therapy was shown to aid in opioid cessation [14] in patients flagged as being at risk for problematic opioid use after surgery. In addition, acceptance and commitment therapy was shown to aid in opioid reduction combined with mindfulness-based relapse prevention [15] in patients with prescription opioid misuse. In addition, a technique named Mindfulness-Orientated Recovery Enhancement (MORE) can lower pain interference and severity, as well as opioid misuse among people with comorbid chronic pain and opioid misuse [16]. Few interventions have been tested among patients who do not specifically misuse opioids, and the findings are mixed [13,17]. While studies have demonstrated the effectiveness of behavioral interventions, few patients can access them due to low availability and high treatment time burden (typically requiring 8-16 hours of in-person or online treatment) [18]. Poor insurance coverage and availability are primary barriers [19], in addition to travel time and the length of the interventions [20]. Treatment barriers were exacerbated by the COVID-19 pandemic, which caused further delays that contributed to worsened pain [21], function, and psychological distress [22,23]. Furthermore, treatment burdens limit both the accessibility and scalability of behavioral pain interventions, and in part, opioid prescribing may be the result of fewer available alternatives, particularly for patients in low-resource settings [24].

A single-session behavioral pain relief skills class, Empowered Relief (ER), was developed to address these barriers. ER synthesizes pain neuroscience education, cognitive behavioral skills, and mindfulness-based skills into a 2-hour, single-session pain relief skills class that enhances self-regulation of pain, including pain catastrophizing [25]. Pain catastrophizing, defined as a pattern of thinking that involves rumination, magnification of pain, and helplessness, is a robust predictor of worse pain and opioid use outcomes [26,27], including greater opioid use [28], opioid misuse [29], pain-related disability [30], and negative affect [31]. Previous work has demonstrated the efficacy of ER when delivered in person [32] and in synchronous [33] and asynchronous [34,35] web-based formats. Notably, when delivered via Zoom to patients with mixed chronic pain conditions (*not taking opioids*; N=101), it was superior to a waitlist control in reducing pain catastrophizing, pain intensity, and pain interference at 3 months after treatment [33]. A larger comparative efficacy trial in people with chronic low back pain (N=263) found in-person ER to be noninferior to 8 sessions of CBT for reducing pain catastrophizing, pain intensity, and pain interference (2 hours vs 16 hours of treatment time) at 3 [32] and 6 months [36]. In addition, 2 randomized controlled trials (RCTs) examining postsurgical outcomes showed that a web-based, automated version of ER tailored to surgical patients had opioid-sparing effects [34] and reduced pain [35] up to 3 months after surgery.

Critical gaps remain in disseminating high-quality digital pain treatments and using digital data collection methods, which can reduce barriers to participation in treatment and research studies [37,38]. Few studies have used digital daily assessments among people with chronic pain taking daily prescribed opioids [31,39-43], despite support for their feasibility and acceptability for measuring daily opioid use among patients with chronic pain [44]. Of those few studies, results consistently show heightened daily-level associations among pain catastrophizing, pain, and opioid craving and use [39,41-43]. However, only 2 studies have examined these associations before and after a behavioral pain intervention among people with chronic pain on long-term opioid therapy. Both studies found significant reductions in daily reports of pain, stress, and craving during and 1 month after a mindfulness-based intervention (MORE) [45,46] involving 16 hours of treatment. Further research of briefer behavioral pain interventions and their impact on daily-level pain and opioid use is needed.

Objectives

We aimed to address these gaps by conducting an uncontrolled pilot study to assess the feasibility, acceptability, and impact of Zoom-delivered ER on pain and opioid use outcomes among patients with chronic pain receiving daily prescribed opioids. We hypothesized that at least 80% of enrolled patients would engage in the treatment. In addition, we hypothesized that ≥80% of patients would rate treatment acceptability, satisfaction, usefulness, and ease of understanding at ≥8 out of 10 and rate ease, comfort, and overall satisfaction of using the Zoom platform to be ≥5 out of 7 (higher scores indicating greater satisfaction, comfort, and ease). Second, we examined changes in pain and opioid use outcomes from enrollment to 3 months (primary end point) and 6 months after treatment. We hypothesized that there would be a significant reduction in pain

XSL•FO
RenderX

intensity, pain interference, and pain catastrophizing at 3 and 6 months after treatment compared to enrollment. We examined changes in opioid dose as an exploratory outcome. Finally, we used 2 rounds of digital daily assessments before and after the class to characterize changes in daily associations between pain catastrophizing, pain intensity, and opioid use before and after the ER class. We hypothesized that higher daily pain catastrophizing would be significantly associated with higher pain intensity and self-reported opioid use. We also hypothesized that daily use of behavioral skills would be associated with a reduction in pain intensity and opioid use after class.

Methods

Study Design

The Stanford University School of Medicine coordinated this uncontrolled, prospective trial during the COVID-19 pandemic. We designed this study to assess the feasibility and acceptability of delivering ER via Zoom (Zoom Communications) to inform the transitioning of a larger RCT (ClinicalTrials.gov: NCT03950791) to the Zoom platform [47]. Thus, this study was not powered to detect differences in outcomes; however, findings will be used to determine adequate power and sample size for future studies. Reporting of the current trial is in concordance with the CONSORT (Consolidated Standards of Reporting Trials) 2010 extension for randomized pilot and feasibility trials. Participants with mixed-etiology chronic pain who were taking daily prescribed opioid medication and were between the ages of 18 and 80 years were enrolled. Full data collection, including all posttreatment assessments, was conducted between June 2020 and June 2021. The 2 ER classes occurred between June 2020 and November 2020.

Ethical Considerations

This study protocol was submitted as a protocol modification to the larger clinical trial and was approved by the Stanford University Institutional Review Board (48784). To enroll in the study, electronic informed consent was obtained from all participants in concordance with the institutional review board protocol. Study data were deidentified and anonymized after study completion. Participants were compensated US $135 for completing the baseline, preclass, postclass, 3-month, and 6-month assessments. The class was provided at no charge. Participants who completed at least 10 of the 14 daily surveys were entered into a raffle for an iPad. Participants could be entered up to twice if they met this threshold for both the baseline and follow-up daily surveys.

Recruitment and Enrollment

Participants were recruited remotely through targeted emails to lists of patients who agreed to be contacted for research purposes through the following three sources: (1) Stanford's Learning Health System, comprising patients who have received care at the Stanford Pain Management Center (a referral-based tertiary outpatient pain clinic in Redwood City, California, that offers multidisciplinary pain care); (2) the Stanford Systems Neuroscience and Pain Lab database of community members with chronic pain who expressed interest in participating in research; and (3) study advertisements on the Stanford Pain Division website. Approximately 17% (10/60) of the study sample were clinic patients and 83% (50/60) were from the larger community. Patients who were in 1 of the 2 databases (Stanford's Learning Health System or Stanford Systems Neuroscience and Pain Lab) were sent an email asking their interest in participating in a chronic pain treatment research study, and if interested, were directed to click on a link to a web-based screening form. Community members who saw the study advertisement on web were also directed to click on a link to fill out a web-based screening form. Those who met the initial eligibility criteria were contacted by research staff by phone to confirm their eligibility. Inclusion criteria were (1) individuals aged between 18 and 80 years; (2) individuals with noncancer chronic pain (defined as pain duration ≥6 months); (3) individuals who self-reported current use of prescription opioids ≥10 morphine-equivalent daily dose (MEDD) for ≥3 months; (4) individuals with fluency in the English language; and (5) individuals who were able to attend a one-time web-based class and complete web-based surveys. Exclusion criteria were (1) ongoing legal action or disability claim, (2) active participation in a CBT-based treatment, (3) inability to provide informed consent and complete study procedures, and (4) active suicidality. Eligible participants were asked to complete an electronic informed consent form and were immediately scheduled for the next available class and directed to fill out the web-based baseline assessment. Recruitment occurred in 2 waves, with recruitment windows open for the 6 weeks preceding each scheduled class.

Assessments

All surveys were collected through REDCap (Research Electronic Data Capture; Vanderbilt University) [48], a web-based electronic data capture platform, which is a secure (password-protected), Health Insurance Portability and Accountability Act–compliant platform hosted by the Stanford University School of Medicine. Survey links were sent via SMS text message or email to participants. Participants completed their first survey at enrollment, followed by baseline daily surveys (14 consecutive days of daily surveys), which were initiated 2 to 4 weeks before the scheduled class. Three days before the class, participants completed a preclass survey, and immediately following the class, participants completed a postclass satisfaction survey. The follow-up daily surveys (14 consecutive days) were initiated 1 month after the class. All daily assessments were deployed at the same time each evening. Furthermore participants completed posttreatment surveys at 3 and 6 months after treatment (Multimedia Appendix 1 illustrates the study flow). There was no in-person contact with a study therapist or research staff. SMS text messages and phone call reminders were deployed by a research assistant to encourage completion of all assessments.

Study Intervention: ER

ER was developed in 2013 [25] and compresses pain neuroscience education and key elements and skills from CBT-based pain and mindfulness interventions into a 2-hour single-session class [32,33]. Each class was led by 1 trained doctoral-level clinical psychologist using a PowerPoint (Microsoft Corporation) slide deck via Zoom to participant

XSL•FO
RenderX

cohorts. Didactic content included psychoeducation about mind-body science as it relates to pain and negative mindset, as well as opioid safety. Participants learned skills on how to identify a negative mindset and self-treat it. During the class, participants practiced skills to decrease physiological hyperarousal, improved regulation of cognition and emotion, and engaged in adaptive behaviors to modulate attention and counteract helplessness. Skills included relaxation, cognitive reframing and thought restructuring, and identifying pleasant activities. Participants self-tailored the information described during the class and developed their own comprehensive self-treatment plan to stop and prevent catastrophizing. Participants left the class with a self-written personalized plan and a 20-minute binaural relaxation response app and were provided verbal and written instructions on how to download the app onto their mobile phone or portable electronic device.

Class Platform

One week before the class, patients were emailed a Zoom link to join the class, along with information on when and how to attend the class. In addition, participants were offered a one-on-one Zoom training session with a research assistant before the intervention, if needed. Participant cohorts consisted of 19 to 22 participants per class. The Zoom platform was used to deliver the ER classes, and instructors screen-shared their PowerPoint presentation slides throughout the class. Class participants were encouraged to ask questions at any time through the Zoom chat box and unmute themselves to participate. Zoom classes were password-protected and hosted within the firewalled Stanford University School of Medicine and Stanford Healthcare systems.

Training and Monitoring of Instructors

Instructors were doctoral-level clinical psychologists trained and certified in ER delivery before administering treatment. Existing treatment manuals and highly structured and standardized class content assured treatment fidelity. A research coordinator served as a fidelity rater and directly observed all classes. Cohort effects were minimized due to the highly structured single-session format and limited participant interaction.

Measures

Feasibility and Acceptability

The assessment of the feasibility and acceptability of the ER Zoom-delivered intervention replicated published methods and was assessed immediately after the class [33]. We assessed feasibility by the proportion of participants that attended the class. Acceptability was assessed by participant ratings of treatment satisfaction, relevance, usefulness of information presented, ease of understanding, and likelihood to use the skills learned. Responses were measured on an 11-point Likert scale from 0 to 10, with higher scores indicating higher acceptability and satisfaction across respective items.

In addition, we assessed satisfaction with using the Zoom platform, such as ease of operating Zoom, engagement with the class material during the Zoom class, and ability to find a private location. We assessed comfort with the Zoom instructor and other class participants, connectedness to the Zoom instructor, as well as overall satisfaction with the Zoom platform. Ratings were assessed using a 7-point Likert scale from 1 “strongly disagree” to 7 “strongly agree,” with higher scores indicating greater satisfaction. Finally, we asked participants for open text responses regarding general feedback on the class and what they found helpful or challenging.

Pain and Opioid Use Measures

The following measures were collected at enrollment, before class (3 days before the class), and 3 and 6 months after treatment.

Pain Intensity

Respondents rated their average pain intensity over the previous 7 days on a numerical pain rating scale of 0 “no pain” to 10 “worst pain imaginable” [49,50].

Patient Reported Outcome Measurement Information System Pain Interference

PROMIS (Patient Reported Outcome Measurement Information System) pain interference short form (8 item) [51,52] was used to assess self-reported consequences of pain, including engagement with daily activities. Higher scores indicate higher interference with activities. The National Institutes of Health PROMIS short-form measures have been applied in pain research [51], and selected domains were identified by the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials as core outcomes [53]. The web-based PROMIS assessment center software [54] was used to calculate the short-form T scores. A standardized T score for each PROMIS measure is generated for each patient. A score of 50 reflects the mean of the US general population, with an SD of 10. Internal consistency of the pain interference measure at baseline was high (α=.96).

Pain Catastrophizing Scale

The 13-item Pain Catastrophizing Scale (PCS) [26] was used to measure negative thoughts and emotional responses to pain. PCS includes 3 subscales: helplessness, magnification, and rumination. The response scale ranges from 0 “not at all” to 4 “all the time”; total sum scores range from 0 to 52. It is scored by summing all items, with higher scores indicating greater catastrophizing. The PCS has demonstrated validity and reliability in mixed-etiology chronic pain and is a psychometrically trusted instrument [55]. The internal consistency of the PCS in our sample at baseline was high (α=.93).

Opioid Dose

Overview

Self-reported opioid type and dose was converted to MEDD using the US Department of Health & Human Services Opioid Oral Morphine Milligram Equivalent Conversion Factors table [56] and the State of Ohio Board of Pharmacy Oral Morphine Milligram Equivalent Conversion Table [57]. Our methods reflected self-reported daily opioid use captured in other trials [34,35,58]. A board-certified anesthesiology and pain medicine physician verified the calculations. In addition, patients were categorized into one of four opioid cohorts based on their opioid

XSL•FO
RenderX

prescription type at enrollment: (1) short-acting prescribed opioids as needed (pro re nata; PRN); (2) prescribed consistent, long-acting opioids; (3) prescribed buprenorphine; and (4) intrathecal pump.

The following measures were collected during the baseline (2-4 weeks before the class) and follow-up (1 month after the class) daily survey periods. Daily behavioral skill use was only assessed during the follow-up daily survey period.

Daily PCS

The daily PCS is a 3-item version of the PCS developed for use in daily diary studies to facilitate research on mechanisms of catastrophizing treatment. It has demonstrated validity and reliability in patients with chronic pain [59]. Scores were calculated by summing all items with a range from 0 to 12. Items assessed patterns of unhelpful thoughts about pain within the last 24 hours and were rated on a 5-point Likert scale from 0 “not at all” to 4 “all the time.” Example items included, “During the past 24 hours I kept thinking about how much I hurt.” The internal consistency of the follow-up daily PCS in our sample was high (α=.91).

Daily Pain Intensity

Daily average pain ratings were assessed on a 0 to 10 numerical rating scale (0=no pain and 10=worst pain imaginable) based on pain experienced in the past 24 hours [49,60].

Daily Opioid Dose

Daily opioid dose, as reported by participants, was converted to MEDD using the US Department of Health & Human Services Opioid Oral Morphine Milligram Equivalent Conversion Factors table [56] and the State of Ohio Board of Pharmacy Oral Morphine Milligram Equivalent Conversion Table [57]. Our methods reflected self-reported daily opioid use captured in other trials [34,35,58]. A board-certified anesthesiology and pain medicine physician verified the calculations.

Daily Skills Use

Skills use was assessed for 14 days during the follow-up daily assessments only. Three items were asked to assess the frequency of cognitive, behavioral, and relaxation techniques used over the past 24 hours using a 6-point Likert scale ranging from 0 “No use” to 5 “Five or more times.” The items were “How many times did you engage in deep breathing or relaxation today?” “How many times did you use distraction or reframing today?” and “How many times did you use pain relief actions today?”

Data Analysis Plan

Analyses were carried out using SPSS (IBM Corp) [61] and RStudio (RStudio, Inc) [62], a per-protocol approach was used. Descriptive data were calculated, including means and SDs across all study variables and time points. To examine attrition, differences in baseline demographic, pain, and opioid characteristics were assessed between those with complete and missing data at each study time point using chi-square tests and ANOVAs. To examine treatment satisfaction, average participant ratings across each satisfaction item were calculated, as well as the proportion of participants who rated each item ≥8 out of 10. To examine satisfaction with the Zoom platform, average participant ratings across each satisfaction item were calculated as well as the proportion of participants who rated each item strongly agree, agree, or somewhat agree (≥5 out of 7).

To examine raw changes in pain intensity, pain catastrophizing, pain interference, and opioid dose from enrollment to 3 and 6 months after treatment, multilevel models with a subject-level random intercept and a fixed effect of time were used. In total, 4 time points (enrollment, before class, 3 months, and 6 months) were included in each model. Linear multilevel models were used to assess changes in pain intensity, pain interference, and pain catastrophizing, and a negative binomial multilevel model was used to assess changes in opioid dose (MEDD), given the highly skewed distribution. Negative binomial models use a natural logarithm link function to account for overdispersion and skewed data. The raw coefficients from this model were converted to a log scale and interpreted as incidence rate ratios, which describe the proportional change in the outcome associated with a 1-unit increase of the predictor [63]. Furthermore, we ran sensitivity analyses for each model, removing tramadol, buprenorphine, and methadone users from the analysis (due to difficulty in equating MEDD). Finally, we calculated Cohen *d* effect sizes (0.20=small, 0.50=medium, and 0.80=large) [64] to aid interpretation.

Across the daily assessments, a similar multilevel modeling approach was used to take advantage of the repeated-measures design. All multilevel models included random intercepts at the participant level and fixed effects of each predictor. To examine predictors of change in daily average pain intensity across the baseline and follow-up daily assessments (relative to enrollment), a linear multilevel model was used. Predictors included daily pain catastrophizing; time (dichotomized factor of pre- vs postclass assessment); dummy-coded opioid cohort (PRN, long-acting, buprenorphine, and intrathecal pump cohorts); and time’s interactions with each factor. Change in opioid dose (MEDD) was transformed into percentage change to approximate a normal distribution. Outliers at the upper bound were windsorized [65] to the 95th percentile (Multimedia Appendix 2 provides more details). A linear multilevel model was used to examine predictors of percentage change in opioid dose across the baseline and follow-up daily assessments (relative to enrollment). Predictors included daily pain catastrophizing, daily average pain intensity, time (before vs after class), opioid cohort, and time’s interaction with each factor. For both the multilevel models assessing the time and daily predictors, we ran sensitivity analyses for each model, removing tramadol and methadone users from the analysis (due to difficulty in equating MEDD).

Finally, to examine the role of behavioral skill use on changes in pain and opioid dose, 2 additional linear multilevel models were conducted across the follow-up daily assessments. Each behavioral skill was dichotomized into no use and ≥1 use in the past 24 hours and added as predictors in each model. Only dummy-coded PRN and long-acting opioid cohorts were examined due to small sample sizes within intrathecal pump and buprenorphine cohorts and rank deficiency errors with maximal models. To examine changes in average pain intensity,

XSL•FO
RenderX

predictors included daily pain catastrophizing, time (before vs after class), PRN versus long-acting cohort, behavioral skills (relax, reframe, and actions), and time's interaction with every factor. To examine the percentage change in opioid dose, predictors included daily pain catastrophizing, daily average pain intensity, time (before vs after class), PRN versus long-acting cohort, behavioral skills (relax, reframe, and actions), and time's interaction with every factor.

Multimedia Appendix 2 gives a description of pain and opioid dose outcome calculations and all model specifications. Across all models, data were inspected to ensure model assumptions were met (ie, homoscedasticity in residuals), and *P* values were corrected for multiple comparisons using the Benjamini-Hochberg false discovery rate (BH-FDR) procedure, an alternative measurement to the family-wise error rate [66]. The BH-FDR procedure has been shown to have greater power than classical family-wise error rate procedures (eg, Bonferroni), while still adequately controlling for type 1 errors. The desired α level (α=.05), number of tests, and raw *P* value for each test were used to calculate the BH-FDR [66].

Results

Overview

Participants were primarily female (45/59, 76%), White (52/59, 88%), non-Hispanic or Latino (57/59, 97%), married (28/59, 48%), and not working due to pain (28/59, 47%). The most prevalent pain condition was chronic low back pain (35/59, 58%); most had pain >5 years (54/59, 92%; Table 1). Oxycodone was the most commonly prescribed opioid (14/59, 24%; Multimedia Appendix 3 for all opioid types). The use of PRN opioids was most common among the sample (32/59, 78%), followed by long-acting prescribed opioids taken consistently (16/59, 39%), buprenorphine (10/59, 24%), and intrathecal pump (2/59, 5%). Table 2 provides descriptive data for all monthly and daily-level variables.

XSL•FO
RenderX

Table 1. The enrolled sample demographic characteristics of people with chronic pain prescribed daily opioids[a].

Measure	Values
Age (y), mean (SD)	59.13 (14.54)
Gender, n (%)	
Men	13 (22)
Women	45 (76)
Other	1 (2)
Race, n (%)	
American Indian or Alaska Native	2 (3)
Asian	1 (2)
Black or African American	2 (3)
White	52 (87)
>1 race	1 (2)
Other	1 (2)
Ethnicity, n (%)	
Hispanic or Latino	2 (3)
Non-Hispanic or Latino	57 (95)
Marital status, n (%)	
Never married	8 (13)
Partnered	8 (13)
Married	28 (47)
Separated	1 (2)
Widowed	4 (7)
Divorced	10 (17)
Education, n (%)	
Some high school	1 (2)
High school diploma or GED[b]	1 (2)
Some college, no degree	17 (28)
Associates or vocational degree	10 (17)
Bachelor's degree	18 (30)
Master's, professional, or doctoral degree	12 (20)
Annual household income (US $), n (%)	
<24,999	19 (32)
25,000-64,999	9 (15)
65,000-84,999	6 (10)
85,000-124,999	11 (18)
>125,000	14 (23)
Employment[c], n (%)	
Full time or part time	9 (15)
Homemaker or not employed	5 (8)
Not working due to pain	28 (47)
Retired	20 (33)
Receiving disability, n (%)	

XSL•FO
RenderX

Measure	Values
Yes	19 (32)
No	40 (67)
Chronic pain condition[c], n (%)	
Postsurgical recovery	13 (22)
Chronic low back pain	35 (58)
Complex regional pain syndrome	7 (12)
Pelvic pain	7 (12)
Migraines	15 (25)
Headaches	9 (15)
Neuropathic pain	33 (55)
Abdominal pain	11 (18)
Myofascial pain	23 (38)
Fibromyalgia	14 (23)
Other	22 (37)
Pain duration (y), n (%)	
1-5	5 (8)
>5	54 (90)

[a]In total, 60 participants enrolled but 1 withdrew and did not provide demographic information.

[b]GED: General Educational Development.

[c]The participants could select all that apply.

Table 2. Descriptive data of monthly and daily-level study variables across all study time points (enrollment through 6 months) among people with chronic pain prescribed daily opioids[a].

	Enrollment	Baseline daily dairies	Before class	Follow-up daily dairies	3 mo	6 mo
Total, n (%)	60 (100)	56 (93)	53 (88)	40 (67)	38 (63)	36 (60)
Pain intensity, mean (SD)	5.76 (1.71)	5.56 (2.10)	5.58 (1.70)	4.64 (2.30)	5.07 (1.80)	5.34 (2.03)
Pain catastrophizing, mean (SD)	17.29 (10.48)[b]	3.90 (2.90)[c]	15.30 (9.83)[b]	3.10 (2.92)[c]	13.32 (12.12)[b]	13.14 (11.61)[b]
Pain interference, mean (SD)[d]	63.49 (7.45)	—[e]	62.38 (7.28)	—	61.26 (9.10)	61.00 (10.38)
Opioid dose[f], mean (SD)	92.11 (147.13)	96.01 (166.62)	102.84 (171.06)	74.10 (120.79)	84.03 (136.30)	87.42 (136.30)

[a]Daily dairies were completed for 14 days approximately 2 to 4 weeks before the class (baseline daily dairies) and for 14 days approximately 4 weeks after the class (follow-up daily dairies). Means and SDs for this period were calculated across all 14 days.

[b]13-item Pain Catastrophizing Scale.

[c]3-item daily Pain Catastrophizing Scale.

[d]Patient Reported Outcome Measurement Information System pain interference scale *t* score.

[e]Not available.

[f]Opioid dose as measured by morphine-equivalent daily dose.

A total of 60 participants enrolled in the study, and 7% (n=4) withdrew from the study before completing the baseline survey (7% pretreatment attrition). Of the 56 (N=60, 93%) participants who completed the baseline survey, 41 (73%) participants attended the class (68% of the enrolled sample), 38 (63% of the enrolled sample) completed the 3-month follow-up survey, and 36 (60% of the enrolled sample) completed the 6-month follow-up survey (Multimedia Appendix 1 illustrates the study flowchart). Attrition analyses indicated that those who did not attend the class or complete the follow-up daily surveys were more likely to be female participants (χ^2_{40}=8.72, *P*=.01). This pattern held at the 3-month posttreatment assessment (gender: χ^2_{37}=6.51, *P*=.04). No other differences were observed in study variables or demographic characteristics between those who had complete or missing data at any other study time points.

XSL•FO
RenderX

Feasibility and Acceptability

Some (41/60, 68% of the enrolled sample) attended the class and completed the postclass survey. Of those who attended, 59% (24/41) of the sample rated Zoom-delivered ER to be satisfactory (ie, ≥8 out of 10; mean 8.02, SD 2.08), and 73% (30/41) rated the class easy to understand (mean 7.93, SD 2.1). Some (26/41, 63%) rated the class as useful (mean 7.83, SD 2.27), 63% (26/41) rated the class as relevant (mean 8.10, SD 2.21), and 77% (31/41) stated they were likely to use the skills (mean 8.41, SD 1.94). Open feedback was collected about the class, and 63% (26/41) provided comments. The most common feedback was that the class reduced travel barriers (15/41, 37%), increased comfort by being able to participate at home (13/41, 32%), and reduced concerns about the COVID-19 pandemic (6/41, 15%). Some (7/41, 17%) participants reported difficulty engaging in the relaxation exercise or the class due to pain or fatigue, 14% (6/41) reported a preference for an in-person class, and 12% (5/41) had been exposed to similar content in the past or were unsure if it would be helpful for their specific pain condition.

Nearly all participants (39/41, 95%) found the Zoom platform easy to operate (≥5 out of 7; mean 6.44, SD 0.92); 97% (40/41) easily found a private location to attend the class (mean 6.71, SD 0.68), and 93% (38/41) reported ease in engaging in the class material via Zoom (mean 6.05, SD 1.24). Some (26/41, 63%) participants reported feeling connected to the instructor (mean 5.22, SD 1.64). Overall, 95% (39/41) of participants reported being somewhat satisfied (n=3, 7%), satisfied (n=17, 42%), or completely satisfied (n=19, 46%) with the Zoom platform (mean 6.24, SD 0.97). A majority (33/41, 80%) of participants reported no problems with the Zoom platform. Of those who reported a problem (n=8, 20%), the most frequently cited problem was unstable internet connection and poor audio quality (n=5, 12%).

Pain and Opioid Use Outcomes

Findings from the 4 multilevel models examining the effect of time on raw changes in pain and opioid outcomes are outlined in Table 3. The corrected significance thresholds for each model were pain intensity (P=.008), pain interference (P=.001), pain catastrophizing (P=.01), and opioid dose (P=.05). Results showed a significant reduction in pain catastrophizing from enrollment to 3 months (mean –2.89, SD 10.89), and from enrollment to 6 months (mean –2.83, SD 10.89) both with a small effect size (Cohen d=0.35 and 0.37, respectively). The average pain intensity significantly reduced from enrollment to 3 months posttreatment (mean –0.52, SD 1.40), indicating a small effect size (Cohen d=0.39). However, average pain intensity did not maintain a significant change from enrollment to 3 months in our sensitivity analysis (Multimedia Appendix 4) nor from enrollment to 6 months posttreatment (mean –0.24, SD 1.42). Reductions in pain interference from enrollment to 3 months (mean –2.03, SD 4.94) and 6 months (mean –2.14, SD 6.77) did not survive correction, although they demonstrated small effect sizes (Cohen d=0.27 and 0.28, respectively). There were no significant reductions in opioid dose across the monthly study time points.

XSL•FO
RenderX

Table 3. Multilevel regressions examining changes in pain and opioid use outcomes from baseline through 6-month time points among people with chronic pain prescribed daily opioids.

Outcomes and variables	Coefficient[a] (SE)	Test statistic[a]	*P* value	Cohen *d*[b]
Linear				
Pain intensity				
Intercept	5.76 (0.23)	24.68	<.001[c]	—[d]
Before class[e]	–0.16 (0.19)	–0.84	.40	0.10
3 mo[e]	–0.56 (0.21)	–2.67	.008[c]	0.39
6 mo[e]	–0.29 (0.21)	–1.37	.17	0.22
Pain interference				
Intercept	63.49 (1.09)	58.02	<.001[c]	—
Before class[e]	–1.06 (0.77)	–1.36	.18	0.15
3 mo[e]	–1.98 (0.88)	–2.25	.03	0.27
6 mo[e]	–1.95 (0.89)	–2.18	.03	0.28
Pain c atastrophizing				
Intercept	17.29 (1.44)	12.05	<.001[c]	—
Before class[e]	–2.07 (1.22)	–1.70	.09	.20
3 mo[e]	–3.49 (1.38)	–2.53	.01[c]	.35
6 mo[e]	–3.61 (1.40)	–2.57	.01[c]	.37
Negative binomial				
Opioid dose (MEDD[f])				
Intercept	46.50 (7.13)	25.00	<.001[c]	—
Before class[e]	1.02 (0.07)	0.29	.77	–0.07
3 mo[e]	0.89 (0.07)	–1.52	.13	0.06
6 mo[e]	0.96 (0.07)	–.48	.63	0.03

[a]The linear multilevel model reports an unstandardizedβ coefficient and *t* value statistics. The negative binomial multilevel model, via glmer, reports incidence rate ratios and *z*-value statistics.

[b]The effect size *d* calculated as baseline minus follow-up divided by pooled SD. Positive *d* values indicate improvements.

[c]Significant values based on the corrected *P* value.

[d]Not available.

[e]Reference group is the baseline.

[f]MEDD: morphine-equivalent daily dose.

Across the 2-week baseline daily assessment period (before class), 93% (56/60) of participants completed 95.5% (749/784) of deployed daily surveys. Across the 2-week follow-up daily assessment period (after class), 68% (41/60) of participants completed 95.6% (549/574) of deployed daily surveys. From the 41 participants who attended class, 1136 baseline and follow-up daily surveys were used to estimate changes in pain and opioid dose using linear multilevel models (Table 4). The corrected significance thresholds for each model were pain intensity (*P*=.001) and opioid dose (*P*=.003). Results showed that higher daily pain catastrophizing was associated with increases in average pain intensity across the daily assessments relative to enrollment (β=0.42, SE 0.02, *P*<.001), and there were greater reductions in pain across the follow-up daily surveys (after class) compared to the baseline daily surveys (before class; β=–0.51, SE 0.12, *P*<.001) although this effect did not survive sensitivity analyses (β=–0.41, SE 0.13, *P*=.002; Multimedia Appendix 5). Opioid cohort and all interaction factors were not significantly associated with change in average pain intensity. When examining changes in opioid dosage, higher daily pain catastrophizing (β=3.14, SE 0.74, *P*<.001) and daily average pain (β=3.47, SE 1.16, *P*=.003) were associated with increases in opioid dose across the daily assessments relative to enrollment, although the effect of pain did not survive our sensitivity analyses (β=3.58, SE 1.40, *P*=.01). Those taking short-acting opioids as needed (PRN)

XSL•FO
RenderX

showed a trend toward reduced opioid dose across the daily assessments (β=–29.88, SE 10.68, *P*=.03), although this association did not survive correction. Furthermore, the time × PRN cohort interaction did not survive correction, although it indicated a trend toward greater opioid dose reductions for the PRN cohort across the follow-up daily surveys (after class) compared to the baseline daily surveys (before class; β=–9.31, SE 3.94, *P*=.02). Figure 1 depicts the raw percentage change in opioid dose between the long-acting and PRN cohorts.

XSL•FO
RenderX

Table 4. Multilevel linear regressions predicting change in pain intensity and opioid dose from enrollment across the baseline (before class) and follow-up (after class) daily diaries among people with chronic pain prescribed daily opioids.

Outcome and variables	Coefficient[a] (SE)	Test statistic[a]	*P* value
Change in pain intensity[b]			
Intercept	–0.42 (0.46)	–0.92	.36
PCS[c,d]	0.42 (0.02)	18.94	<.001[e]
Pre-post[f]	–0.51 (0.12)	–4.20	<.001[e]
PRN[g] cohort	–0.55 (0.58)	–0.96	.34
Long-acting cohort	0.23 (0.65)	0.36	.72
Buprenorphine cohort	–0.82 (0.69)	–1.18	.25
ITP[h] cohort	0.77 (1.45)	0.53	.60
Pre-post x PCS	0.03 (0.03)	0.99	.32
Pre-post x PRN	0.21 (0.15)	1.39	.16
Pre-post x LA[i]	–0.27 (0.17)	–1.58	.12
Pre-post x Buprenorphine	–0.04 (0.18)	–0.20	.84
Pre-post x ITP	–0.32 (0.39)	–0.82	.41
Percentage change in opioid dose[j]			
Intercept	5.77 (9.66)	0.60	.55
PCS[d]	3.14 (0.74)	4.23	<.001[e]
Pre-post[e]	1.40 (3.11)	0.45	.65
Average pain	3.47 (1.16)	2.99	.003[e]
PRN cohort	–26.88 (12.21)	–2.20	.03
LA cohort	–17.36 (13.65)	–1.27	.21
Buprenorphine cohort	–6.97 (14.59)	–0.48	.64
ITP cohort	–17.64 (30.51)	–0.58	.56
Pre-post × PCS	–1.72 (1.13)	–1.52	.13
Pre-post × pain	–1.55 (1.78)	–0.87	.38
Pre-post × PRN	–9.31 (3.94)	–2.37	.02
Pre-post × LA	3.08 (4.39)	0.70	.48
Pre-post × buprenorphine	–8.93 (4.67)	–1.92	.06
Pre-post × ITP	16.98 (9.90)	1.72	.09

[a]The linear multilevel model reports an unstandardized β coefficient and *t* value statistics. [b]The outcome is change in average pain intensity in the past 24 hours and depicts a change at each daily assessment compared to the enrollment.

[c]PCS: Pain Catastrophizing Scale.

[d]Daily PCS.

[e]Significant values based on the corrected *P* value.

[f]0: preclass daily assessments; 1: postclass daily assessments.

[g]PRN: pro re nata.

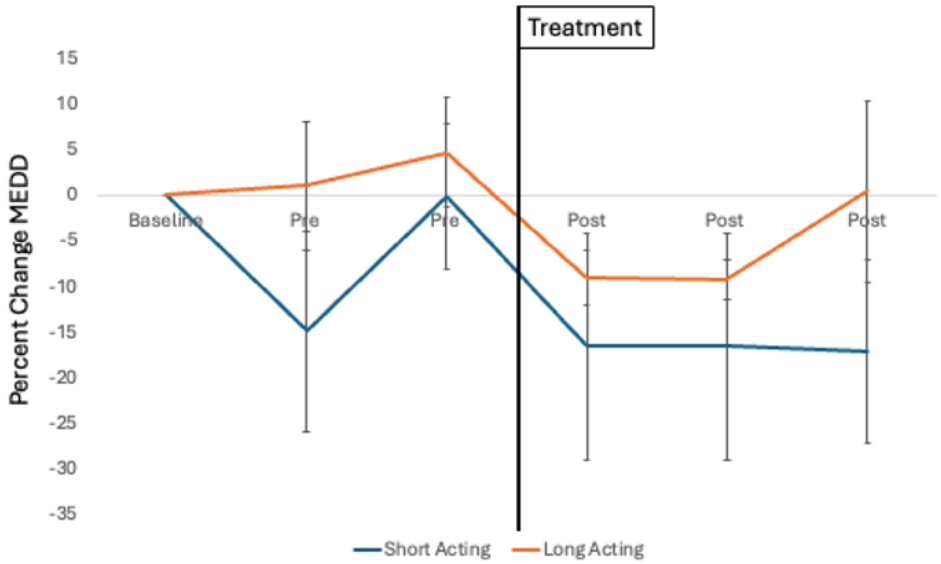
[h]ITP: intrathecal pump.

[i]LA: long acting.

[j]The percentage change in opioid dose is measured using morphine-equivalent daily dose and depicts a change at each daily assessment compared to the enrollment.

XSL•FO
RenderX

Figure 1. The raw percentage change in opioid dose (morphine-equivalent daily dose; MEDD) from enrollment across study time points between participants with chronic pain taking long-acting opioids and short-acting (prescribed opioids as needed) opioids.



Across the follow-up daily surveys (1 month after the class; n=549), 446 (81.2%) of the daily surveys endorsed using relaxation at least once, 403 (73.4%) endorsed using reframe at least once, and 410 (74.7%) endorsed using pain relief actions at least once. Findings from the multilevel models are provided in Multimedia Appendix 6. The corrected significance thresholds for each model were pain intensity (P=.001) and opioid dose (P=.05). Results showed that higher daily pain catastrophizing was associated with increase in average pain intensity relative to enrollment (β=0.48, SE 0.04, P<.001). The use of behavioral skills, time, and interactions was not a significant predictor of change in average pain intensity. Furthermore, daily pain catastrophizing, daily pain, use of the behavioral skills, time, or their interactions were not significant predictors of percentage change in opioid dose across the follow-up daily surveys.

Discussion

Principal Findings

We performed an uncontrolled pilot study to investigate the initial feasibility, acceptability, and preliminary impact of a Zoom-delivered ER class among people with chronic pain taking daily opioid medication. Our results showed significant reductions in pain catastrophizing and average pain intensity that lasted 3 months after treatment, although with small effects. Pain catastrophizing, a target of the ER class, continued to exhibit significant reductions at 6 months after treatment. Daily-level results also showed a trend toward opioid dose reduction for people taking PRN opioids, though did not survive correction (P=.02). The findings from this pilot study provide justification for a larger, appropriately powered RCT.

Feasibility and Acceptability

We conducted this study during the first year of the COVID-19 pandemic (2020 to 2021), while access to traditional, in-person pain psychology interventions was extremely limited. Although we observed lower-than-hypothesized engagement rates (≥80%), with 68% (41/60) of the enrolled patients attending the class, we noted that our engagement was relatively high when compared to the substantially reduced health care use that occurred during the early stages of the COVID-19 pandemic [21]. For those who participated in the class, 63% (26/41) to 77% (31/41) of the patients rated the class as satisfactory, useful, and easy to understand. Furthermore, they reported high satisfaction with web-based delivery, which reduced travel barriers and concerns about the COVID-19 pandemic and increased comfort by attending the intervention from home. In addition, ease and comfort with digital assessments was evident, as >93% (749/784) of the offered daily surveys were completed (ie, at least 26 of 28 days per person). Future work that involves follow-up interviews with participants is needed to identify potential adaptations to improve the intervention's engagement and acceptability. On the basis of our initial findings, implementing strategies to address digital literacy and poor internet access may improve engagement, along with further tailoring of the relaxation exercise to better engage people with more severe pain and fatigue. In addition, increased reminders and incentives may increase engagement in the intervention.

Pain and Opioid Use Outcomes

Following the ER class, individuals reported significant reductions in pain catastrophizing at 3 and 6 months after treatment and significant reductions in pain intensity at 3 months after treatment. Furthermore, results showed a trend toward decreased pain interference at 3 and 6 months after treatment, although these did not survive multiplicity correction (P=.03). This study of people taking daily opioids yielded smaller effect sizes (Cohen d=0.35-0.39) as compared to prior RCTs of in-person ER [32] (Cohen d=0.43-0.89) and Zoom-delivered ER [33] (Cohen d=0.54-0.76) in patients with chronic pain with no enrollment criterion applied for daily opioid use. While there are notable methodological differences between studies (eg, different pain conditions, enrollment criteria, and control conditions), the smaller effect sizes may also be due to higher rates of disability, which were 32% in this sample as compared to 5% to 10% in the previous 2 studies. Patients on long-term opioid therapy may require more intensive treatment, a different type of treatment, or booster sessions, especially in the setting

XSL•FO
RenderX

of greater disability. Regardless, this study's findings suggest that Zoom-delivered ER may impart small benefits in pain intensity and pain catastrophizing, and trending toward improvement for pain interference (though did not survive correction, *P*=.03), in people with chronic pain on long-term opioid therapy, warranting further testing in an RCT.

Our study is among the few to use daily digital assessments among patients with mixed-etiology chronic pain and taking long-term opioid medication. Our findings replicate previous work showing that higher daily reports of pain catastrophizing are associated with worse pain intensity, and higher daily pain intensity and pain catastrophizing are associated with taking greater doses of prescribed opioids [31,39,59]. In addition, daily ratings of pain decreased from pre-to-post class, similar to previous work that showed a mindfulness intervention (MORE) reduced daily-level pain, negative affect, opioid craving, and opioid use [45,46]. Furthermore, we examined the differences between patients prescribed PRN opioids and long-acting opioids. Results showed a trend that people prescribed PRN opioids were more likely to reduce their opioid dose from before to after class as compared to patients on long-acting opioids, although this association did not survive our rigorous multiple comparison correction (*P*=.02; Figure 1 illustrates a raw percentage change across both groups). Our findings provide a signal that a brief, Zoom-delivered behavioral pain treatment may help patients taking PRN opioids rely less on their medication, as has been found with other multisession behavioral treatments [15,16]. For those on long-acting opioids, patients are likely to need additional support and treatments, with active clinician-patient engagement targeting a slow, gradual weaning of their opioid dose over several weeks or months [67]. Behavioral treatments and psychoeducation are important treatment components as they offer patients additional pain coping skills and may help reduce potential pain flares during the weaning process [17,68,69]. Future RCTs are warranted to examine Zoom-delivered ER alone in reducing PRN opioid use and in the context of an opioid weaning protocol in conjunction with prescribers.

Limitations

We noted several limitations in this study that are important to consider when interpreting the findings. First, this study was an uncontrolled pilot trial. Therefore, inferences about intervention causation cannot be assumed. We did use a rigorous statistical approach to control for potential type 1 errors. Furthermore, the actual therapeutic effect size cannot be ascertained and separated from potential regression to the mean effects. A larger RCT (NCT03950791) is currently underway to further assess the efficacy of Zoom-delivered ER among patients with chronic pain receiving daily prescribed opioids [58]. Second, our sample consisted of largely White individuals, non-Hispanic or Latino individuals, and educated women, although our sample did report higher rates of disability (19/59, 32%) than in our previous studies. Future RCTs are needed to assess the efficacy of the intervention in diverse and underserved patient populations who experience greater barriers to treatment [70]. Third, this study evidenced significant attrition (only 41/60, 60% of the enrolled sample completed the final 6-month assessment). While we did not find significant differences between those with complete versus missing data on our study variables, stressors related to the COVID-19 pandemic likely impacted study attrition. Importantly, we had exceptional compliance rates with digital daily assessments (93%-95%). Future research may consider using additional engagement strategies, such as increased digital assessments or phone calls in-between study time points. Fourth, we suspect that we did not see any significant associations between daily behavioral skill use and daily pain or opioid use because of the once-daily assessments that were completed at the same time each day and asked participants to recall information from the last 24 hours. Future research should use stratified random sampling, prompts with a shorter recall window, and other objective measures (eg, electronic pill caps) to reduce measurement error [71]. Finally, there are other variables that we did not measure that, in previous work, have moderated associations between pain and opioid use, such as opioid craving [72] and positive affect [45]. Future research using larger samples would benefit from modeling both known risk factors (eg, opioid craving, negative affect, and pain catastrophizing) [39,72] and protective factors (eg, positive affect and use of behavioral skills) [45] to further elucidate their impact on daily reports of pain and opioid use as well as their change following behavioral pain treatments.

Conclusions

Improvements to future Zoom-delivered ER iterations are needed to improve feasibility and acceptability among people with chronic pain and daily prescribed opioid use. However, despite this, we also found a promising preliminary impact of the intervention on pain outcomes. There was high compliance with digital daily assessments, and results showed a promising trend of opioid dose reduction among people taking PRN opioids (though did not survive correction, *P*=.02). Our results indicate significant reductions in pain catastrophizing and daily pain intensity, suggesting that even a brief, digitally delivered behavioral intervention can positively influence pain outcomes in this population. Although reductions in opioid use did not reach statistical significance (*P*=.02), the observed trend among patients prescribed PRN opioids highlights ER's potential role in supporting reduced reliance on medication.

We developed this study to inform a larger RCT that aims to rigorously assess the efficacy of Zoom-delivered ER using digital daily and monthly level assessments currently in data collection [58]. Our initial findings suggest that Zoom-delivered ER may be a promising intervention that can improve pain-related outcomes and reduce common barriers to accessing pain psychology interventions, such as travel and time burdens [20]. Dissemination of brief, accessible, and scalable pain psychology interventions, such as ER, stands to improve patient outcomes by increasing equitable access to effective pain psychology treatments and may be useful tools to help reduce pain and opioid use. ER offers a promising approach to expanding the reach of behavioral pain treatments, particularly for populations with limited access to in-person care.

XSL•FO
RenderX

Acknowledgments

Research reported in this work was supported by the National Institute on Drug Abuse of the National Institutes of Health under award numbers K12DA050607 (KAE), T32DA035165 (KAE and TCD), K23DA047473 (MSZ), K23DA048972 (DSY), and K24DA053564 (BDD). Additional funding support from the National Institute on Neurological Diseases and Stroke was received under the award numbers K24NS126781 and R61NS11865 (SCM). Funders were not involved in the study design, data collection, analysis or interpretation of results, or manuscript writing.

Data Availability

The data generated are available upon reasonable request to the corresponding author (MSZ).

Conflicts of Interest

Stanford University receives fees for continuing medical education Empowered Relief (ER) clinician certification workshops; BDD receives no personal monies for the ER certification workshops nor for the dissemination or adoption of ER. BDD is the chief science adviser at AppliedVR, and her consulting role with this company (personal fees) is unrelated to the current research. BDD receives royalties for 4 pain treatment books she has authored or coauthored. She is the principal investigator for 2 pain research awards from the Patient-Centered Research Outcomes Research Institute, one of which involves the study of ER. BDD is the principal investigator for 2 National Institutes of Health (NIH) grants that are investigating the efficacy of ER (one is the Mobile and Online-Based Interventions to Lessen Pain [MOBILE] Relief study). BDD serves on the board of directors for the Institute for Brain Potential, which is licensed by Stanford University to offer ER clinician certification workshops, and is on the medical advisory board for the Facial Pain Association. BDD is a scientific member of the NIH Interagency Pain Research Coordinating Committee, a former member of the Centers for Disease Control and Prevention Opioid Workgroup (2020 to 2021), and a current member of the Pain Advisory Group of the American Psychological Association. SCM receives research funding from the NIH, Food and Drug Administration, and Patient-Centered Outcomes Research Institute (administered through Stanford University). He is an unpaid adviser to both ACTTION (Analgesic, Anesthetic, and Addiction Clinical Trial Translations and Innovations, Opportunities, and Networks) on their oversight committee and the American Chronic Pain Association for their scientific oversight. All other authors report no other disclosures or conflicts of interest.

Multimedia Appendix 1
Participant flowchart.
[DOCX File , 52 KB - formative_v9i1e68292_app1.docx]

Multimedia Appendix 2
Supplementary methods.
[DOCX File , 22 KB - formative_v9i1e68292_app2.docx]

Multimedia Appendix 3
Type of prescribed opioid used at enrollment.
[DOCX File , 19 KB - formative_v9i1e68292_app3.docx]

Multimedia Appendix 4
A sensitivity analysis of multilevel regressions examining changes in pain and opioid outcomes across monthly assessments.
[DOCX File , 20 KB - formative_v9i1e68292_app4.docx]

Multimedia Appendix 5
A sensitivity analysis of multilevel linear regressions predicting change in pain intensity and opioid dose from baseline across the baseline (before class) and follow-up (after class) daily dairies.
[DOCX File , 20 KB - formative_v9i1e68292_app5.docx]

Multimedia Appendix 6
Multilevel linear regressions predicting change in pain intensity and opioid dose from enrollment across the follow-up (after class) daily dairies.
[DOCX File , 25 KB - formative_v9i1e68292_app6.docx]

References

1. Yong RJ, Mullins PM, Bhattacharyya N. Prevalence of chronic pain among adults in the United States. Pain 2022 Feb 01;163(2):e328-e332. [doi: 10.1097/j.pain.0000000000002291] [Medline: 33990113]

XSL•FO
RenderX

2. Institute of Medicine (US) Committee on Advancing Pain Research, Care, and Education. Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research. Vol. xvii. Washington, DC: National Academies Press (US); 2011.
3. Dominick CH, Blyth FM, Nicholas MK. Unpacking the burden: understanding the relationships between chronic pain and comorbidity in the general population. Pain 2012 Feb;153(2):293-304. [doi: 10.1016/j.pain.2011.09.018] [Medline: 22071318]
4. Aroke HA, Buchanan AL, Bratberg J, Hogan JW, Rich JD, Wen X, et al. Initial patterns of prescription opioid supply and risk of mortality among insured adults in the United States. J Addict Med 2021 Apr 01;15(2):99-108 [FREE Full text] [doi: 10.1097/ADM.0000000000000701] [Medline: 32769771]
5. Farrar JT, Bilker WB, Cochetti PT, Argoff CE, Haythornthwaite J, Katz NP, et al. Evaluating the stability of opioid efficacy over 12 months in patients with chronic noncancer pain who initially demonstrate benefit from extended release oxycodone or hydrocodone: harmonization of Food and Drug Administration patient-level drug safety study data. Pain 2022 Jan 01;163(1):47-57 [FREE Full text] [doi: 10.1097/j.pain.0000000000002331] [Medline: 34261978]
6. Annual surveillance report of drug-related risks and outcomes. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/overdose-prevention/media/pdfs/pubs/2019-cdc-drug-surveillance-report.pdf [accessed 2024-04-29]
7. Dasgupta N, Funk MJ, Proescholdbell S, Hirsch A, Ribisl KM, Marshall S. Cohort study of the impact of high-dose opioid analgesics on overdose mortality. Pain Med 2016 Jan;17(1):85-98. [doi: 10.1111/pme.12907] [Medline: 26333030]
8. Gomes T, Juurlink D, Moineddin R, Gozdyra P, Dhalla I, Paterson J, et al. Geographical variation in opioid prescribing and opioid-related mortality in Ontario. Healthc Q 2011;14(1):22-24 [FREE Full text] [doi: 10.12927/hcq.2011.22153] [Medline: 21301233]
9. Gatchel RJ, Peng YB, Peters ML, Fuchs PN, Turk DC. The biopsychosocial approach to chronic pain: scientific advances and future directions. Psychol Bull 2007 Jul;133(4):581-624. [doi: 10.1037/0033-2909.133.4.581] [Medline: 17592957]
10. Bruns D, Mueller K, Warren PA. Biopsychosocial law, health care reform, and the control of medical inflation in Colorado. Rehabil Psychol 2012 May;57(2):81-97. [doi: 10.1037/a0028623] [Medline: 22686548]
11. Driscoll MA, Edwards RR, Becker WC, Kaptchuk TJ, Kerns RD. Psychological interventions for the treatment of chronic pain in adults. Psychol Sci Public Interest 2021 Sep;22(2):52-95. [doi: 10.1177/15291006211008157] [Medline: 34541967]
12. Jamison RN, Ross EL, Michna E, Chen LQ, Holcomb C, Wasan AD. Substance misuse treatment for high-risk chronic pain patients on opioid therapy: a randomized trial. Pain 2010 Sep;150(3):390-400 [FREE Full text] [doi: 10.1016/j.pain.2010.02.033] [Medline: 20334973]
13. Naylor MR, Naud S, Keefe FJ, Helzer JE. Therapeutic Interactive Voice Response (TIVR) to reduce analgesic medication use for chronic pain management. J Pain 2010 Dec;11(12):1410-1419 [FREE Full text] [doi: 10.1016/j.jpain.2010.03.019] [Medline: 20620119]
14. Dindo L, Zimmerman MB, Hadlandsmyth K, StMarie B, Embree J, Marchman J, et al. Acceptance and commitment therapy for prevention of chronic postsurgical pain and opioid use in at-risk veterans: a pilot randomized controlled study. J Pain 2018 Oct;19(10):1211-1221 [FREE Full text] [doi: 10.1016/j.jpain.2018.04.016] [Medline: 29777950]
15. Vowles KE, Witkiewitz K, Cusack KJ, Gilliam WP, Cardon KE, Bowen S, et al. Integrated behavioral treatment for veterans with co-morbid chronic pain and hazardous opioid use: a randomized controlled pilot trial. J Pain 2020;21(7-8):798-807 [FREE Full text] [doi: 10.1016/j.jpain.2019.11.007] [Medline: 31760109]
16. Garland EL, Hanley AW, Nakamura Y, Barrett JW, Baker AK, Reese SE, et al. Mindfulness-oriented recovery enhancement vs supportive group therapy for co-occurring opioid misuse and chronic pain in primary care: a randomized clinical trial. JAMA Intern Med 2022 Apr 01;182(4):407-417 [FREE Full text] [doi: 10.1001/jamainternmed.2022.0033] [Medline: 35226053]
17. Sullivan MD, Turner JA, DiLodovico C, D'Appollonio A, Stephens K, Chan Y. Prescription opioid taper support for outpatients with chronic pain: a randomized controlled trial. J Pain 2017 Mar;18(3):308-318 [FREE Full text] [doi: 10.1016/j.jpain.2016.11.003] [Medline: 27908840]
18. Rini C, Williams DA, Broderick JE, Keefe FJ. Meeting them where they are: using the internet to deliver behavioral medicine interventions for pain. Transl Behav Med 2012 Mar 28;2(1):82-92 [FREE Full text] [doi: 10.1007/s13142-011-0107-2] [Medline: 22924084]
19. Spears CA, Hodges SE, Kiyani M, Yang Z, Edwards RM, Musick A, et al. Health care resource utilization and management of chronic, refractory low back pain in the United States. Spine (Phila Pa 1976) 2020 Oct 15;45(20):E1333-E1341 [FREE Full text] [doi: 10.1097/BRS.0000000000003572] [Medline: 32453242]
20. Darnall BD, Scheman J, Davin S, Burns JW, Murphy JL, Wilson AC, et al. Pain psychology: a global needs assessment and national call to action. Pain Med 2016 Feb;17(2):250-263 [FREE Full text] [doi: 10.1093/pm/pnv095] [Medline: 26803844]
21. Lacasse A, Pagé MG, Dassieu L, Sourial N, Janelle-Montcalm A, Dorais M, et al. Impact of the COVID-19 pandemic on the pharmacological, physical, and psychological treatments of pain: findings from the chronic pain and COVID-19 Pan-Canadian study. Pain Rep 2021;6(1):e891 [FREE Full text] [doi: 10.1097/PR9.0000000000000891] [Medline: 33598594]
22. Serrano-Ibáñez ER, Esteve R, Ramírez-Maestre C, Ruiz-Párraga GT, López-Martínez AE. Chronic pain in the time of COVID-19: stress aftermath and central sensitization. Br J Health Psychol 2021 May;26(2):544-552. [doi: 10.1111/bjhp.12483] [Medline: 33099793]

XSL•FO
RenderX

23. Shanthanna H, Nelson AM, Kissoon N, Narouze S. The COVID-19 pandemic and its consequences for chronic pain: a narrative review. Anaesthesia 2022 Sep;77(9):1039-1050 [FREE Full text] [doi: 10.1111/anae.15801] [Medline: 35848380]
24. Carey EP, Nolan C, Kerns RD, Ho PM, Frank JW. Association between facility-level utilization of non-pharmacologic chronic pain treatment and subsequent initiation of long-term opioid therapy. J Gen Intern Med 2018 May;33(Suppl 1):38-45 [FREE Full text] [doi: 10.1007/s11606-018-4324-y] [Medline: 29633137]
25. Darnall BD, Sturgeon JA, Kao MC, Hah JM, Mackey SC. From catastrophizing to recovery: a pilot study of a single-session treatment for pain catastrophizing. J Pain Res 2014;7:219-226 [FREE Full text] [doi: 10.2147/JPR.S62329] [Medline: 24851056]
26. Sullivan MJ, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. Psychol Assess 1995 Dec;7(4):524-532. [doi: 10.1037/1040-3590.7.4.524]
27. Martinez-Calderon J, Jensen MP, Morales-Asencio JM, Luque-Suarez A. Pain catastrophizing and function in individuals with chronic musculoskeletal pain: a systematic review and meta-analysis. Clin J Pain 2019 Mar;35(3):279-293. [doi: 10.1097/AJP.0000000000000676] [Medline: 30664551]
28. Sharifzadeh Y, Kao M, Sturgeon JA, Rico TJ, Mackey S, Darnall BD. Pain catastrophizing moderates relationships between pain intensity and opioid prescription: nonlinear sex differences revealed using a learning health system. Anesthesiology 2017 Jul;127(1):136-146 [FREE Full text] [doi: 10.1097/ALN.0000000000001656] [Medline: 28614083]
29. Martel MO, Edwards RR, Jamison RN. The relative contribution of pain and psychological factors to opioid misuse: a 6-month observational study. Am Psychol 2020 Sep;75(6):772-783. [doi: 10.1037/amp0000632] [Medline: 32915022]
30. Severeijns R, Vlaeyen JW, van den Hout MA, Weber WE. Pain catastrophizing predicts pain intensity, disability, and psychological distress independent of the level of physical impairment. Clin J Pain 2001 Jun;17(2):165-172. [doi: 10.1097/00002508-200106000-00009] [Medline: 11444718]
31. Carpenter RW, Lane SP, Bruehl S, Trull TJ. Concurrent and lagged associations of prescription opioid use with pain and negative affect in the daily lives of chronic pain patients. J Consult Clin Psychol 2019 Oct;87(10):872-886 [FREE Full text] [doi: 10.1037/ccp0000402] [Medline: 31556664]
32. Darnall BD, Roy A, Chen AL, Ziadni MS, Keane RT, You DS, et al. Comparison of a single-session pain management skills intervention with a single-session health education intervention and 8 sessions of cognitive behavioral therapy in adults with chronic low back pain: a randomized clinical trial. JAMA Netw Open 2021 Aug 02;4(8):e2113401 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.13401] [Medline: 34398206]
33. Ziadni MS, Gonzalez-Castro L, Anderson S, Krishnamurthy P, Darnall BD. Efficacy of a single-session "empowered relief" zoom-delivered group intervention for chronic pain: randomized controlled trial conducted during the COVID-19 pandemic. J Med Internet Res 2021 Sep 10;23(9):e29672 [FREE Full text] [doi: 10.2196/29672] [Medline: 34505832]
34. Darnall BD, Ziadni MS, Krishnamurthy P, Flood P, Heathcote LC, Mackey IG, et al. "My surgical success": effect of a digital behavioral pain medicine intervention on time to opioid cessation after breast cancer surgery-a pilot randomized controlled clinical trial. Pain Med 2019 Nov 01;20(11):2228-2237 [FREE Full text] [doi: 10.1093/pm/pnz094] [Medline: 31087093]
35. Ziadni MS, You DS, Keane R, Salazar B, Jaros S, Ram J, et al. "My surgical success": feasibility and impact of a single-session digital behavioral pain medicine intervention on pain intensity, pain catastrophizing, and time to opioid cessation after orthopedic trauma surgery-a randomized trial. Anesth Analg 2022 Aug 01;135(2):394-405. [doi: 10.1213/ANE.0000000000006088] [Medline: 35696706]
36. Darnall BD, Burns JW, Hong J, Roy A, Slater K, Poupore-King H, et al. Empowered relief, cognitive behavioral therapy, and health education for people with chronic pain: a comparison of outcomes at 6-month follow-up for a randomized controlled trial. Pain Rep 2024 Jan;9(1):e1116 [FREE Full text] [doi: 10.1097/PR9.0000000000001116] [Medline: 38288134]
37. Eccleston C, Blyth FM, Dear BF, Fisher EA, Keefe FJ, Lynch ME, et al. Managing patients with chronic pain during the COVID-19 outbreak: considerations for the rapid introduction of remotely supported (eHealth) pain management services. Pain 2020 May;161(5):889-893 [FREE Full text] [doi: 10.1097/j.pain.0000000000001885] [Medline: 32251203]
38. Hadjiat Y, Arendt-Nielsen L. Digital health in pain assessment, diagnosis, and management: overview and perspectives. Front Pain Res (Lausanne) 2023;4:1097379 [FREE Full text] [doi: 10.3389/fpain.2023.1097379] [Medline: 37139342]
39. Finan PH, Carroll CP, Moscou-Jackson G, Martel MO, Campbell CM, Pressman A, et al. Daily opioid use fluctuates as a function of pain, catastrophizing, and affect in patients with sickle cell disease: an electronic daily diary analysis. J Pain 2018 Jan;19(1):46-56 [FREE Full text] [doi: 10.1016/j.jpain.2017.08.010] [Medline: 28943232]
40. Miller MB, Curtis AF, Chan WS, Deroche CB, McCrae CS. Daily associations between sleep and opioid use among adults with comorbid symptoms of insomnia and fibromyalgia. J Clin Sleep Med 2021 Apr 01;17(4):729-737 [FREE Full text] [doi: 10.5664/jcsm.9002] [Medline: 33226334]
41. Bruneau A, Frimerman L, Verner M, Sirois A, Fournier C, Scott K, et al. Day-to-day opioid withdrawal symptoms, psychological distress, and opioid craving in patients with chronic pain prescribed opioid therapy. Drug Alcohol Depend 2021 Aug 01;225:108787. [doi: 10.1016/j.drugalcdep.2021.108787] [Medline: 34091157]
42. Verner M, Sirois A, Coutu-Nadeau E, Fournier C, Carriere J, Martel MO. The association between perceived injustice and opioid craving in patients with chronic pain: the mediating role of daily pain intensity, negative affect, and catastrophizing. J Addict Med 2023;17(1):35-41. [doi: 10.1097/ADM.0000000000001022] [Medline: 35861344]

XSL•FO
RenderX

43. Frimerman L, Verner M, Sirois A, Scott K, Bruneau A, Perez J, et al. Day-to-day hedonic and calming effects of opioids, opioid craving, and opioid misuse among patients with chronic pain prescribed long-term opioid therapy. Pain 2021 Aug 01;162(8):2214-2224. [doi: 10.1097/j.pain.0000000000002220] [Medline: 33729213]
44. Anderson Goodell EM, Nordeck C, Finan PH, Vandrey R, Dunn KE, Thrul J. Feasibility and acceptability of using smartphone-based EMA to assess patterns of prescription opioid and medical cannabis use among individuals with chronic pain. Internet Interv 2021 Dec;26:100460 [FREE Full text] [doi: 10.1016/j.invent.2021.100460] [Medline: 34646751]
45. Garland EL, Bryan CJ, Finan PH, Thomas EA, Priddy SE, Riquino MR, et al. Pain, hedonic regulation, and opioid misuse: modulation of momentary experience by Mindfulness-Oriented Recovery Enhancement in opioid-treated chronic pain patients. Drug Alcohol Depend 2017 Apr 01;173 Suppl 1:S65-S72 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.07.033] [Medline: 28363322]
46. Garland EL, Gullapalli BT, Prince KC, Hanley AW, Sanyer M, Tuomenoksa M, et al. Zoom-based mindfulness-oriented recovery enhancement plus just-in-time mindfulness practice triggered by wearable sensors for opioid craving and chronic pain. Mindfulness (N Y) 2023 May 24;14(6):1-17 [FREE Full text] [doi: 10.1007/s12671-023-02137-0] [Medline: 37362184]
47. Ziadni MS, Anderson SR, Gonzalez-Castro L, Darnall BD. Comparative efficacy of a single-session "Empowered Relief" videoconference-delivered group intervention for chronic pain: study protocol for a randomized controlled trial. Trials 2021 May 22;22(1):358. [doi: 10.1186/s13063-021-05303-8] [Medline: 34022930]
48. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
49. Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001 Nov;94(2):149-158. [doi: 10.1016/S0304-3959(01)00349-9] [Medline: 11690728]
50. Cook KF, Dunn W, Griffith JW, Morrison MT, Tanquary J, Sabata D, et al. Pain assessment using the NIH toolbox. Neurology 2013 Mar 12;80(11_supplement_3):23 [FREE Full text] [doi: 10.1212/wnl.0b013e3182872e80]
51. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, PROMIS Cooperative Group. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010 Nov;63(11):1179-1194 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.04.011] [Medline: 20685078]
52. Cook KF, Bamer AM, Amtmann D, Molton IR, Jensen MP. Six patient-reported outcome measurement information system short form measures have negligible age- or diagnosis-related differential item functioning in individuals with disabilities. Arch Phys Med Rehabil 2012 Jul;93(7):1289-1291. [doi: 10.1016/j.apmr.2011.11.022] [Medline: 22386213]
53. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, IMMPACT. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005 Jan;113(1-2):9-19. [doi: 10.1016/j.pain.2004.09.012] [Medline: 15621359]
54. HealthMeasures consists of PROMIS, Neuro-QoL, ASCQ-Me, and NIH toolbox. HealthMeasures. URL: https://www.healthmeasures.net/ [accessed 2024-04-29]
55. Osman A, Barrios FX, Gutierrez PM, Kopper BA, Merrifield T, Grittmann L. The Pain Catastrophizing Scale: further psychometric evaluation with adult samples. J Behav Med 2000 Aug;23(4):351-365. [doi: 10.1023/a:1005548801037] [Medline: 10984864]
56. Opioid oral morphine milligram equivalent (MME) conversion factors table for prescription drug coverage. US Department of Health & Human Services. URL: https://www.hhs.gov/guidance/document/opioid-oral-morphine-milligram-equivalent-mme-conversion-factors-0 [accessed 2024-04-29]
57. Oral morphine milligram equivalent conversion table. State of Ohio Board of Pharmacy. 2017. URL: https://www.ohiopmp.gov/Documents/MorphineEquivalentDailyDoseConversionTable.pdf [accessed 2022-10-30]
58. Ziadni MS, Chen AL, Winslow T, Mackey SC, Darnall BD. Efficacy and mechanisms of a single-session behavioral medicine class among patients with chronic pain taking prescription opioids: study protocol for a randomized controlled trial. Trials 2020 Jun 12;21(1):521 [FREE Full text] [doi: 10.1186/s13063-020-04415-x] [Medline: 32532346]
59. Darnall BD, Sturgeon JA, Cook KF, Taub CJ, Roy A, Burns JW, et al. Development and validation of a daily pain catastrophizing scale. J Pain 2017 Sep;18(9):1139-1149 [FREE Full text] [doi: 10.1016/j.jpain.2017.05.003] [Medline: 28528981]
60. Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. J Clin Nurs 2005 Aug 30;14(7):798-804. [doi: 10.1111/j.1365-2702.2005.01121.x] [Medline: 16000093]
61. IBM SPSS statistics for Windows. IBM Corp. 2019. URL: https://www.ibm.com/products/spss-statistics [accessed 2024-04-29]
62. RStudio Team. Integrated development for R. RStudio. URL: http://www.rstudio.com/ [accessed 2022-06-06]
63. Atkins DC, Baldwin SA, Zheng C, Gallop RJ, Neighbors C. A tutorial on count regression and zero-altered count models for longitudinal substance use data. Psychol Addict Behav 2013 Mar;27(1):166-177 [FREE Full text] [doi: 10.1037/a0029508] [Medline: 22905895]
64. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd edition. New York, NY: Routledge; Dec 1989.

XSL•FO
RenderX

65. Reifman A, Keyton K. Winsorize. In: Salkind NJ, editor. Encyclopedia of Research Design. Thousand Oaks, CA: Sage Publications; 2010:1636-1638.
66. Benjamini Y, Hochberg Y. On the adaptive control of the false discovery rate in multiple testing with independent statistics. J Educ Behav Stat 2000 Mar 01;25(1):60-83 [FREE Full text] [doi: 10.3102/10769986025001060]
67. Dowell D, Ragan KR, Jones CM, Baldwin GT, Chou R. CDC clinical practice guideline for prescribing opioids for pain - United States, 2022. MMWR Recomm Rep 2022 Nov 04;71(3):1-95 [FREE Full text] [doi: 10.15585/mmwr.rr7103a1] [Medline: 36327391]
68. Ashton-James CE, Glare P, Darnall BD. Out of office hours: scalable, on-demand, digital support for patients tapering prescription opioids. Pain 2020 Oct 3;161(10):2252-2254. [doi: 10.1097/j.pain.0000000000001947] [Medline: 32501860]
69. Darnall BD, Ziadni MS, Stieg RL, Mackey IG, Kao M, Flood P. Patient-centered prescription opioid tapering in community outpatients with chronic pain. JAMA Intern Med 2018 May 01;178(5):707-708 [FREE Full text] [doi: 10.1001/jamainternmed.2017.8709] [Medline: 29459978]
70. Janevic MR, McLaughlin SJ, Heapy AA, Thacker C, Piette JD. Racial and socioeconomic disparities in disabling chronic pain: findings from the health and retirement study. J Pain 2017 Dec;18(12):1459-1467 [FREE Full text] [doi: 10.1016/j.jpain.2017.07.005] [Medline: 28760648]
71. Stone AA, Obbarius A, Junghaenel DU, Wen CK, Schneider S. High-resolution, field approaches for assessing pain: ecological momentary assessment. Pain 2021 Jan;162(1):4-9 [FREE Full text] [doi: 10.1097/j.pain.0000000000002049] [Medline: 32833794]
72. Mun CJ, Finan PH, Epstein DH, Kowalczyk WJ, Agage D, Letzen JE, et al. Craving mediates the association between momentary pain and illicit opioid use during treatment for opioid-use disorder: an ecological momentary assessment study. Addiction 2021 Jul;116(7):1794-1804 [FREE Full text] [doi: 10.1111/add.15344] [Medline: 33220102]

Abbreviations

BH-FDR: Benjamini-Hochberg false discovery rate
CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
ER: Empowered Relief
MEDD: morphine-equivalent daily dose
MORE: Mindfulness-Orientated Recovery Enhancement
PCS: Pain Catastrophizing Scale
PRN: pro re nata
PROMIS: Patient Reported Outcome Measurement Information System
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture

Edited by A Mavragani; submitted 07.11.24; peer-reviewed by MA Bushey, H Rafi; comments to author 16.12.24; revised version received 06.01.25; accepted 30.01.25; published 11.03.25.

Please cite as:
Edwards KA, Dildine TC, You DS, Herrick AM, Darnall BD, Mackey SC, Ziadni MS
Zoom-Delivered Empowered Relief for Chronic Pain: Observational Longitudinal Pilot Study Exploring Feasibility and Pain-Related Outcomes in Patients on Long-Term Opioids
JMIR Form Res 2025;9:e68292
URL: https://formative.jmir.org/2025/1/e68292
doi:10.2196/68292
PMID:

©Karlyn A Edwards, Troy C Dildine, Dokyoung S You, Ashley M Herrick, Beth D Darnall, Sean C Mackey, Maisa S Ziadni. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Assessment of Cognitive Health in Outpatient Primary Care: Usability Study

Adam J Doerr[1*], MD; Taylor A Orwig[1*], MS; Matthew McNulty[1], BS; Stephanie Denise M Sison[2], MD; David R Paquette[3], PhD; Robert Leung[1], MD; Huitong Ding[4], PhD; Stephen B Erban[5], MD; Bruce R Weinstein[2], MD; Yurima Guilarte-Walker[6], MS; Adrian H Zai[6], MD, PhD, MPH; Allan J Walkey[5], MD; Apurv Soni[1,5], MD, PhD; David D McManus[1,5], MD; Honghuang Lin[1,5], PhD

[1]Program in Digital Medicine, Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States

[2]Division of General Internal Medicine, Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States

[3]Massachusetts AI and Technology Center, Manning College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, MA, United States

[4]Department of Anatomy and Neurobiology, Chobanian & Avedisian School of Medicine, Boston University, Boston, MA, United States

[5]Division of Health System Science, Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States

[6]Division of Health Informatics and Implementation Science, Department of Population and Quantitative Health Sciences, University of Massachusetts Chan Medical School, Worcester, MA, United States

[*]these authors contributed equally

Corresponding Author:
Honghuang Lin, PhD
Program in Digital Medicine
Department of Medicine
University of Massachusetts Chan Medical School
S6-755
55 Lake Avenue North
Worcester, MA, 01655
United States
Phone: 1 7744554881
Email: honghuang.lin@umassmed.edu

Abstract

Background: Screening for cognitive impairment in primary care is important, yet primary care physicians (PCPs) report conducting routine cognitive assessments for less than half of patients older than 60 years of age. Linus Health's Core Cognitive Evaluation (CCE), a tablet-based digital cognitive assessment, has been used for the detection of cognitive impairment, but its application in primary care is not yet studied.

Objective: This study aimed to explore the integration of CCE implementation in a primary care setting.

Methods: A cohort of participants was recruited from the upcoming schedules of participating PCPs at UMass Memorial Medical Center. Eligibility criteria included individuals aged ≥65 years; ability to read, write, and speak in English or Spanish; no previous diagnosis of cognitive impairment; and no known untreated hearing or vision impairment. Research coordinators collected consent from participants and facilitated the screening process. PCPs reviewed reports in real time, immediately before the scheduled visits, and shared results at their discretion. A report was uploaded to each participant's REDCap (Research Electronic Data Capture; Vanderbilt University) record and linked to the encounter in the electronic health record. Feedback from patients and their caregivers (if applicable) was collected by a tablet-based survey in the clinic before and after screening. Participating PCPs were interviewed following the completion of the study.

Results: The screened cohort included 150 patients with a mean age of 74 (SD 7) years, of whom 65% (97/150) were female. The CCE identified 40 patients as borderline and 7 as positive for cognitive impairment. A total of 84 orders were placed for select laboratory tests or referrals to neurology and neuropsychology within 20 days of CCE administration. Before the assessment, 95% (143/150) of patients and all 15 caregivers expressed a desire to know if their or their loved one's brain health was declining. All except one patient also completed the postassessment survey. Among them, 96% (143/149) of patients reported finding the CCE easy to complete, and 70% (105/149) felt that the experience was beneficial. In addition, 87% (130/149) of patients agreed

XSL•FO
RenderX

or strongly agreed that they wanted to know their CCE results. Among the 7 participating PCPs, 6 stated that the CCE results influenced their patient care management, and all 7 indicated they would continue using the CCE if it were made available after the study.

Conclusions: We explored the integration of the CCE into primary care visits, which showed minimal disruption to the practice workflow. Future studies will be warranted to further validate the implementation of digital cognitive impairment screening tools within primary care settings in the real world.

(JMIR Form Res 2025;9:e66695) doi:10.2196/66695

KEYWORDS

cognitive assessment; primary care; digital; cognitive impairment; digital assessment; assessment; cognitive health; cognition; primary care; cognitive evaluation; Core Cognitive Evaluation; CCE; cohort; impairment; cognitive; outpatient

Introduction

Dementia is a neurodegenerative disorder characterized by progressive deficits in one or more cognitive domains severe enough to interfere with an individual's functional status [1]. Mild cognitive impairment (MCI) is a precursor of dementia, characterized by more subtle cognitive deficits that do not interfere with functional status [1]. Although definitional variability and reporting bias complicate the estimation of prevalence [2], MCI likely afflicts 12% to 18% of people aged 60 years and older [3,4]. The rate of progression from MCI to dementia is estimated to be 10% to 15% yearly [5].

Only about 18% of Americans are familiar with MCI [1]. However, when MCI is described, 85% report they would want to know if they had either MCI or early dementia, with reasons cited including to plan for the future, to enable earlier treatment of symptoms, to take steps to prevent further deterioration, and to better understand their condition [1]. Similar sentiments are held by family members, for whom early detection enables a more gradual adaptation to the caregiver role, resulting in an improved sense of competence and less psychological distress [6,7]. On a broader level, early detection can delay institutionalization and inform the allocation of human resources within a health care system already marred by shortages in the dementia care workforce [8-10].

Most primary care physicians (PCPs) believe screening for MCI and dementia in older patients is important [1]. Following the detection of MCI or early dementia, PCPs commonly intervene by assessing for reversible causes, recommending lifestyle changes, and referring to specialists [1]. However, most PCPs feel uncomfortable in this domain and report conducting cognitive assessments for less than half of patients older than 60 years of age [1]. Reported barriers to screening in primary care settings include low provider confidence (particularly in distinguishing MCI from normal aging), time constraints, competing priorities, poorly defined protocols for screening, and lack of financial incentives [1,11,12]. Only 40% of PCPs report being familiar with existing cerebrospinal fluid and imaging biomarker tests that can aid in the diagnosis of dementia, while only 18% of patients are referred for such testing when MCI is detected [1,11,12]. An unfortunate consequence is that more than 90% of patients with MCI are not identified [13], resulting in delayed diagnosis when advanced stages of dementia have developed, disability is higher, treatment options are limited [14], and caregiver stress is high [15].

Alzheimer's disease is the most common neurodegenerative cause of MCI and dementia in individuals aged ≥65 years. Biomarkers for AD span multiple modalities, including molecular profiling of bodily fluids, brain imaging, electroencephalography, and neuropsychological testing [16]. With increased computational power, particularly in the form of artificial intelligence, digital biomarkers have emerged as promising new noninvasive approaches for early detection of MCI and early dementia due to AD [17]. One example is the digitization of the traditional clock drawing test (CDT). Although thoroughly validated, the CDT is a classically analog test that requires manual supervision and skilled interpretation, limiting its use in busy clinical practice and making it susceptible to interobserver variability [18]. Linus Health's Core Cognitive Evaluation (CCE) is a tablet-based cognitive assessment tool consisting of an immediate recall task, a digital CDT administered on an Apple iPad using an Apple Pencil, and a delayed recall task (collectively termed Digital Clock and Recall [DCR]) [19,20]. Performance is analyzed by a machine learning algorithm to yield a score (ranging from 0 to 5), which is indicative of cognitive impairment (0-1), borderline for cognitive impairment (2-3), or not indicative of cognitive impairment (4-5). The DCR is paired with a Life and Health Questionnaire (LHQ), which screens for modifiable lifestyle and psychosocial risk and protective factors related to brain health, enabling prediction of future dementia risk and personalized recommendations for patients. The DCR has demonstrated greater sensitivity and accuracy than existing cognitive assessment tests such as the Mini-Mental State Examination (MMSE) and the Mini-Cog [21] and takes substantially less time to complete (approximately 3 min).

These qualities make the CCE well suited for the busy primary care setting. However, the impact of the CCE in primary care with respect to patient and provider satisfaction, downstream clinical decision-making, and reimbursement remains unclear. The objective of this study was to evaluate the feasibility and impact of screening for cognitive impairment in an outpatient, primary care setting using the Linus Health CCE.

XSL•FO
RenderX

Methods

Screening and Recruitment

Participants for this study were recruited from the Benedict Adult Primary Care Clinic at UMass Memorial Medical Center. Eligible patients were identified through a list of upcoming visits generated from the electronic health record system (Epic Systems). Eligibility criteria included being aged 65 years or older; having the ability to read, write, and speak in English or Spanish; having no previous diagnosis of cognitive impairment; and having no untreated hearing or vision impairment that would hinder audible conversation or the use of a tablet. Recruitment letters were mailed to eligible patients, inviting them to participate in the study. These letters included an opt-out phone number and email address. If no response was received, a research coordinator followed up with a phone call 3-5 days after the letter was sent to assess interest. Patients who expressed interest received another phone call one day before their scheduled visit to confirm their participation, remind them to arrive an hour early, and advise them to bring their reading glasses or hearing aids, if applicable. All contacted patients were logged in a recruitment tracker, which recorded details such as patient name, participation status, PCP name, recruitment letter mail date, follow-up call date, visit date and type, and participation reminder call date.

Assessment

Participants were greeted in the clinic by research coordinators, escorted to a quiet and private room without an analog clock present, and completed the consent process. They then completed preassessment surveys on study tablets. If the patients were accompanied by their caregivers, their caregivers were also provided with a fact sheet and completed a survey. Simultaneously, research coordinators registered participants on a separate tablet equipped with the Linus Health CCE. Participants then completed the CCE on a study tablet. Participants were first administered the DCR, in the following sequence: (1) during the immediate recall condition, 3 words were voiced by the tablet, and the patient was asked to repeat them immediately; (2) during the command clock condition, the tablet voiced the following command: “draw the face of a clock, put in all of the numbers, and set the hands to ‘10 after 11”’; (3) during the copy clock condition, the tablet displayed a picture of an analog clock set to 11:10 and asked the patient to copy it; and (4) during the delayed recall condition, the patient was asked to repeat the 3 words initially voiced by the tablet. The DCR was not adjusted for demographics or education similar to the original Mini-Cog score. Patients then completed the LHQ and a postassessment survey on the study tablet before their visit with the PCP. A total of 7 PCPs who consented to participate were identified and given a fact sheet. They went through a training session before the study. After assessments were completed for the entire cohort, all PCPs involved in the study completed a provider feedback survey sent by email and completed an interview assessing their impressions of the CCE.

Results Sharing

After completion of the CCE, results were immediately uploaded into the Linus Health portal. This generated a provider-facing report consisting of DCR scores; interpretations; playbacks of clock drawing and verbal recall; clinical decision support; and a personalized patient-facing Brain Health Action Plan, which identifies modifiable risk factors for cognitive impairment and provides evidence-based suggestions for intervention. These results were printed and given to PCPs immediately before patient visits for review in real time. DCR results were shared with patients at the discretion of PCPs, while the Brain Health Action Plans were printed and universally shared with each patient.

Data Management and Analysis

Study data were collected and managed using REDCap (Research Electronic Data Capture; Vanderbilt University) tools hosted at UMass Chan Medical School [22,23]. Informed consent was also obtained using REDCap within the eConsent framework on a study tablet [24]. REDCap is a secure, web-based software platform designed to support data capture for research studies, providing (1) an intuitive interface for validated data capture; (2) audit trails for tracking data manipulation and export procedures; (3) automated export procedures for seamless data downloads to common statistical packages; and (4) procedures for data integration and interoperability with external sources. Identifiable patient health information was concealed from Linus Health throughout the study and thereafter.

Ethical Considerations

The study protocol was approved by the Western-Copernicus Group Institutional Review Board (approval number 1341324), and informed consent was obtained from all patients. Caregivers and PCPs were provided with a fact sheet to guide them in completing the survey. No compensation for participation was provided. For data analysis, only deidentified data were exported from REDCap to ensure confidentiality and compliance with ethical standards.

Results

Screened Cohort

A flowchart of participant recruitment is shown in Figure S1 in Multimedia Appendix 1. During the screening process, 1469 patients were identified as meeting the eligibility criteria. Among them, 160 accepted the invitation to participate. Of these, 9 either did not show up, changed their minds, or canceled their appointments, and one participant was found ineligible to provide consent. Ultimately, 150 patients from the panels of 7 PCPs were successfully enrolled and completed the CCE between March and December 2023. The baseline demographic and clinical characteristics of studied patients are listed in Table 1. The participants had a mean age of 74 (SD 7) years; 65% (97/150) were female, and 74% (111/150) held a college degree or higher. The CCE identified 40 patients as borderline and 7 patients as positive for cognitive impairment. The number of patients recruited from each PCP panel ranged from 9 to 38.

XSL•FO
RenderX

Table 1. Patient demographic and clinical characteristics. Mean (SD) values for continuous variables, and counts (%) for categorical variables have been reported.

Characteristics	Values (n=150)
Age (years), mean (SD)	74 (7)
Female, n (%)	97 (65)
Race, n (%)	
White	147 (98)
Black	2 (1)
Not reported or unknown	1 (1)
Ethnicity, n (%)	
Hispanic or Latino	2 (1)
Not Hispanic or Latino	136 (91)
Not recorded or unknown	12 (8)
Education, n (%)	
High school	11 (7)
Some college	28 (19)
College graduate and above	111 (74)
Hypertension, n (%)	111 (74)
Diabetes mellitus, n (%)	32 (21)
Hyperlipidemia, n (%)	135 (90)
Smoking, n (%)	35 (23)
Ischemic heart disease, n (%)	27 (18)
Atrial fibrillation or flutter, n (%)	24 (16)
Heart failure, n (%)	10 (7)
Peripheral vascular disease, n (%)	7 (5)
Obstructive sleep apnea, n (%)	62 (41)
Parkinson disease, n (%)	1 (1)
Depression, n (%)	37 (25)
Stroke, n (%)	9 (6)
Transient ischemic attack, n (%)	10 (7)
Epilepsy, n (%)	4 (3)
Hearing loss, n (%)	66 (44)
B_{12} deficiency, n (%)	3 (2)

Provider Response to the CCE

Among the 7 PCPs participating in the study, 5 had more than 10 years of clinical practice, one had 5-10 years of clinical practice, and one was a junior clinician. All PCPs stated that it was important to detect undiagnosed cognitive impairment, and most noted time or workflow constraints as a barrier to cognitive testing. All PCPs reviewed test results before meeting with their patient, with an average reported review time of 2 minutes, and more time is needed if the results indicated a cognitive issue. A majority of the PCPs stated that the CCE did not interfere with patient visits (5/7, 71%) and that it did not lead to any stress or challenges for their patient (6/7, 86%). Most PCPs (4/7, 57%) reported that the CCE did not cause them any stresses or challenges; among the 3 providers who answered affirmatively, one discussed the stress of breaking the news of poor CCE scores to patients, another cited the challenge of finding the time to discuss results, and the last voiced concerns about false positive scores that might result from extraneous factors like anxiety, comorbidities, and language barriers. All PCPs stated that they generally agreed with the CCE results, and a majority (6/7, 86%) of PCPs stated that the results of the CCE changed their patient care management. All PCPs reported they would continue using the CCE if made available following the study.

XSL•FO
RenderX

Downstream Clinical Action Following the CCE

Referrals to neurology and neuropsychology, select laboratory tests possibly related to the workup of dementia, and neuroimaging studies ordered after administration of the CCE are reported in Table 2. There were 84 such orders placed within 20 days of CCE administration, and an additional 58 orders were placed within the subsequent 160 days thereafter. Moreover, among the 7 patients with DCR scores of 0 and 1, a total of 4 (57%) were later diagnosed with cognitive impairment. In contrast, only 16 (11%) of the 143 patients with scores of 2 or above were later diagnosed with cognitive impairment.

Table 2. Near-term and delayed cognition-relevant orders placed following the administration of the Core Cognitive Evaluation (CCE).

Cognition-relevant orders	Days since the CCE	
	0-20	21-180
Referral, n		
Neurology	3	2
Neuropsychology	0	1
Imaging test, n		
CT[a] head	2	5
MRI[b] brain	4	11
PET[c] brain	0	2
Laboratory test, n		
B_{12}	26	7
Folate	11	4
Rapid plasma reagin	3	0
Thyroid-stimulating hormone	35	26
Total tests, n	84	58

[a]CT: computed tomography.

[b]MRI: magnetic resonance imaging.

[c]PET: positron emission tomography.

Patient and Caregiver Perspectives on Brain Health Prior to the CCE

The results of the preassessment survey for patients and caregivers are displayed in Table 3, and the stratified results are shown in Table S1 in Multimedia Appendix 1. Before the screening, 58% (87/150) of patients agreed or strongly agreed they worry about their brain health, while 95% (143/150) of patients and all caregivers reported they would want to know if their or their loved one's brain health was declining. Approximately two-thirds of both patients (100/150, 67%) and caregivers (10/15, 67%) agreed or strongly agreed their plans for the next 5-10 years would change in response to a decline in brain health; "How we spend our time" was the most frequently selected decision that would be impacted. About two-thirds (95/150, 63%) of patients reported they go to their PCPs for help with taking care of their brain health.

XSL•FO
RenderX

Table 3. Patient and caregiver perspectives on brain health before the administration of the Core Cognitive Evaluation.

Survey questions	Strongly agree, n (%)	Agree, n (%)	Neither agree nor disagree, n (%)	Disagree, n (%)	Strongly disagree, n (%)
Patients (n=150)					
I worry about my memory or brain health.	15 (10)	72 (48)	28 (19)	17 (11)	18 (12)
I understand how to take care of my brain health.	48 (32)	69 (46)	29 (19)	3 (2)	1 (1)
I go to my primary care physician to help me take care of my brain health.	45 (30)	50 (33)	43 (29)	9 (6)	3 (2)
If my brain health was declining, I would want to know.	81 (54)	62 (41)	3 (2)	1 (1)	3 (2)
If I found out that my brain health was declining, it would change my plans for the next 5-10 years.	37 (25)	63 (42)	28 (19)	14 (9)	8 (5)
Caregivers (n=15)					
I worry about my relative's memory or brain health.	3 (20)	5 (33)	4 (27)	2 (13)	1 (7)
I understand how to help my relatives take care of their brain health.	5 (33)	4 (27)	6 (40)	0 (0)	0 (0)
I would want to take my relative to their primary care physician to help me take care of their brain health.	7 (47)	6 (40)	1 (7)	1 (7)	0 (0)
If my relative's brain health was declining, I would want to know.	8 (53)	7 (47)	0 (0)	0 (0)	0 (0)
If I knew my relative's brain health was declining, it would change our plans for the next 5-10 years.	4 (27)	6 (40)	4 (27)	0 (0)	1 (7)

Patient Satisfaction With the Core Cognitive Evaluation

The results of the postassessment survey for patients are displayed in Table 4, and the stratified results are shown in Table S2 in Multimedia Appendix 1. One patient did not complete the assessment. Among participants who completed the postassessment survey, 96% (143/149) of patients reported being able to easily complete the CCE, while 70% (105/149) reported that completing the CCE was beneficial. Among those who felt the CCE was beneficial, all felt it enabled them to be productive about their brain health, about two-thirds (68/105, 65%) felt it gave them more control over their long-term brain health, and about one-third (37/105, 35%) felt it addressed a current specific concern about their brain health. As a result of taking the CCE, only one patient reported feeling sad or depressed, while 5% (8/149) reported feeling worried or anxious. Most patients (130/149, 87%) agreed or strongly agreed they would like to know their CCE results.

Table 4. Patient satisfaction after the administration of the Core Cognitive Evaluation. One participant did not complete the postassessment survey (n=149).

Survey questions	Strongly agree, n (%)	Agree, n (%)	Neither agree nor disagree, n (%)	Disagree, n (%)	Strongly disagree, n (%)
I was able to complete this assessment easily	107 (71)	36 (24)	3 (2)	0 (0)	3 (2)
Taking this assessment was beneficial	46 (31)	59 (39)	35 (23)	6 (4)	3 (2)
Taking this assessment was stressful	0 (0)	8 (5)	16 (11)	44 (29)	81 (54)
This assessment made me feel anxious or worried	0 (0)	8 (5)	12 (8)	39 (26)	90 (60)
This assessment made me feel sad or depressed	0 (0)	1 (1)	6 (4)	32 (21)	110 (73)
I was to know my results from this assessment	84 (56)	46 (31)	8 (5)	3 (2)	8 (5)

Discussion

Principal Findings

Cognitive impairment is a formidable public health threat, with significant burdens posed to individual patients, their caregivers, the health care system, and the economy. With two disease-modifying therapies now approved by the US Food and Drug Administration (FDA) [25,26] and commercially available and covered by the Centers for Medicare and Medicaid Services (CMS) [27], early detection of cognitive impairment has assumed new relevance. In this study, a cohort of older outpatients without a previous diagnosis of cognitive impairment was screened in a primary care setting using the CCE.

The preassessment survey overall comported with the existing body of literature suggesting patients are concerned about their brain health and would strongly prefer to detect cognitive impairment as early as possible. Overall, participant satisfaction with the screening process was high, with a low incidence of patient-reported negative psychological impact. A majority of PCPs stated that the CCE results changed their care management, and all PCPs stated they would continue using the CCE if made available following the study.

XSL•FO
RenderX

The protocol used in this study prioritized minimizing disruption in an already busy clinical environment by aligning cognitive assessments with patients' scheduled appointments. Central to this workflow was a dedicated team member who facilitated consenting, screening, and delivery of results to providers. Although this role was filled by a research coordinator in the study, it could feasibly be assigned to clinical staff, such as medical assistants, in real-world practice given the simplicity of the screening platform and the minimal training required to administer the survey. Meanwhile, the timing of assessments before visits and the provision of concise results with easily digestible interpretations to PCPs obviate the need for manual grading and interpretation, reducing both the time burden and expertise requirements associated with conventional assessments. This design therefore serves as a pragmatic framework for the implementation of a digital cognitive impairment screening program in a real-world primary care setting.

There are noteworthy limitations inherent to this workflow and its evaluation. While potentially more streamlined for PCPs, the study required patients to arrive early for their scheduled appointments. This represents an added medical burden for the older population and may compete with other previsit agenda items. This extra time requirement was necessary to ensure that cognitive assessments could be completed without disrupting the scheduled clinical visits, allowing PCPs to review the results in real time. However, it might also impact a spouse or family member helping the patient to spend more time onsite in the office. This burden could be mitigated in future studies with improved electronic health record integration and automated clinical decision support implementation, or by exploring alternative workflows, such as integrating the CCE into regular appointment times or offering remote assessments. Furthermore, while patients were surveyed before their scheduled visits and were blind to their CCE results, PCPs were surveyed following enrollment completion, introducing the potential for recall bias in their evaluation of the screening process. This retrospective feedback collection allowed PCPs to reflect on their experiences with multiple patients but may have affected the accuracy of their assessments. In addition, the study's design, which involved a convenience sample from a single primary care clinic, limits the generalizability of the findings. While this approach facilitated close monitoring of the implementation process, it may not fully represent the broader population. Future studies should expand the sample size and include multiple sites across diverse geographic and health care settings. In addition, most of the participants in this study are non-Hispanic White, which greatly limits its generalizability to other racial and ethnic populations. More efforts have to be made to increase the diversity of participants in the future [28]. Finally, irrespective of feasibility, the efficacy of a widely adopted digital cognitive screening program using the CCE remains speculative until tested in a randomized trial. It would be essential to further investigate strategies for sustainable and scalable integration of the CCE in diverse clinical environments, including enhanced electronic health record integration.

Conclusion

In summary, we explored the integration of the CCE into primary care visits, which showed minimal disruption to the practice workflow. Future studies will be warranted to further validate the implementation of digital cognitive impairment screening tools within primary care settings in the real world.

Acknowledgments

We thank patients and their families, whose help and participation made this work possible. The investigators' time was partly supported by grants from P30AG073107, U01AG068221, R01AG080670, R01AG083735, R21HL175584, R01HL155343, U01HL146382, U54HL143541, and R33HL158541. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Institutes of Health or the US Department of Health and Human Services. The funding agencies had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

DDM has received consultancy fees from Heart Rhythm Society, Fitbit, Flexcon, Pfizer, Avania, NAMSA, and Bristol Myers Squibb. DDM reports receiving research support from Fitbit, Apple, Care Evolution, Boeringher Ingelheim, Pfizer, and Bristol Myers Squibb. The authors have no further interests to declare.

Multimedia Appendix 1
Flowchart illustrating the participant screening and enrollment, patient perspectives on brain health prior to the administration of the CCE stratified by DCR score groups, and patient satisfaction after the administration of the CCE stratified by DCR score groups. CCE: Core Cognitive Evaluation; DCR: Digital Clock and Recall.
[DOCX File , 53 KB - formative_v9i1e66695_app1.docx]

References

XSL•FO
RenderX

1. Alzheimer's Association. 2022 Alzheimer's disease facts and figures. Alzheimers Dement 2022 Apr;18(4):700-789. [doi: 10.1002/alz.12638] [Medline: 35289055]
2. Casagrande M, Marselli G, Agostini F, Forte G, Favieri F, Guarino A. The complex burden of determining prevalence rates of mild cognitive impairment: a systematic review. Front Psychiatry 2022;13:960648 [FREE Full text] [doi: 10.3389/fpsyt.2022.960648] [Medline: 36213927]
3. Petersen RC, Lopez O, Armstrong MJ, Getchius TSD, Ganguli M, Gloss D, et al. Practice guideline update summary: mild cognitive impairment: report of the guideline development, dissemination, and implementation subcommittee of the American Academy of Neurology. Neurology 2018;90(3):126-135 [FREE Full text] [doi: 10.1212/WNL.0000000000004826] [Medline: 29282327]
4. Rajan KB, Weuve J, Barnes LL, McAninch EA, Wilson RS, Evans DA. Population estimate of people with clinical Alzheimer's disease and mild cognitive impairment in the United States (2020-2060). Alzheimers Dement 2021;17(12):1966-1975 [FREE Full text] [doi: 10.1002/alz.12362] [Medline: 34043283]
5. Mitchell AJ, Shiri-Feshki M. Rate of progression of mild cognitive impairment to dementia--meta-analysis of 41 robust inception cohort studies. Acta Psychiatr Scand 2009;119(4):252-265. [doi: 10.1111/j.1600-0447.2008.01326.x] [Medline: 19236314]
6. de Vugt ME, Verhey FRJ. The impact of early dementia diagnosis and intervention on informal caregivers. Prog Neurobiol 2013;110:54-62. [doi: 10.1016/j.pneurobio.2013.04.005] [Medline: 23689068]
7. de Vugt ME, Stevens F, Aalten P, Lousberg R, Jaspers N, Winkens I, et al. Do caregiver management strategies influence patient behaviour in dementia? Int J Geriatr Psychiatry 2004;19(1):85-92. [doi: 10.1002/gps.1044] [Medline: 14716704]
8. National Academies of Sciences, Division of Behavioral and Social Sciences and Education, Board on Behavioral, Cognitive, and Sensory Sciences, Committee on the Decadal Survey of Behavioral and Social Science Research on Alzheimer's Disease and Alzheimer's Disease-Related Dementias. Reducing the Impact of Dementia in America: A Decadal Survey of the Behavioral and Social Sciences. Washington, DC: National Academies Press; 2021.
9. Banerjee S, Wittenberg R. Clinical and cost effectiveness of services for early diagnosis and intervention in dementia. Int J Geriatr Psychiatry 2009;24(7):748-754. [doi: 10.1002/gps.2191] [Medline: 19206079]
10. Dubois B, Padovani A, Scheltens P, Rossi A, Dell'Agnello G. Timely diagnosis for Alzheimer's Disease: a literature review on benefits and challenges. J Alzheimers Dis 2016;49(3):617-631 [FREE Full text] [doi: 10.3233/JAD-150692] [Medline: 26484931]
11. Gong N, Yang D, Zou J, He Q, Hu L, Chen W, et al. Exploring barriers to dementia screening and management services by general practitioners in China: a qualitative study using the COM-B model. BMC Geriatr 2023;23(1):55 [FREE Full text] [doi: 10.1186/s12877-023-03756-x] [Medline: 36717818]
12. Bernstein Sideman A, Al-Rousan T, Tsoy E, Piña Escudero SD, Pintado-Caipa M, Kanjanapong S, et al. Facilitators and barriers to dementia assessment and diagnosis: perspectives from dementia experts within a global health context. Front Neurol 2022;13:769360 [FREE Full text] [doi: 10.3389/fneur.2022.769360] [Medline: 35418934]
13. Mattke S, Jun H, Chen E, Liu Y, Becker A, Wallick C. Expected and diagnosed rates of mild cognitive impairment and dementia in the U.S. medicare population: observational analysis. Alzheimers Res Ther 2023;15(1):128 [FREE Full text] [doi: 10.1186/s13195-023-01272-z] [Medline: 37481563]
14. Mattke S, Jun H, Chen E, Liu Y, Becker A, Wallick C. Expected and diagnosed rates of mild cognitive impairment and dementia in the U.S. medicare population: observational analysis. Alzheimers Res Ther 2023;15(1):128 [FREE Full text] [doi: 10.1186/s13195-023-01272-z] [Medline: 37481563]
15. Richardson TJ, Lee SJ, Berg-Weger M, Grossberg GT. Caregiver health: health of caregivers of Alzheimer's and other dementia patients. Curr Psychiatry Rep 2013;15(7):367. [doi: 10.1007/s11920-013-0367-2] [Medline: 23712718]
16. Gunes S, Aizawa Y, Sugashi T, Sugimoto M, Rodrigues PP. Biomarkers for Alzheimer's disease in the current state: a narrative review. Int J Mol Sci 2022;23(9):4962 [FREE Full text] [doi: 10.3390/ijms23094962] [Medline: 35563350]
17. Vrahatis AG, Skolariki K, Krokidis MG, Lazaros K, Exarchos TP, Vlamos P. Revolutionizing the early detection of Alzheimer's disease through non-invasive biomarkers: the role of artificial intelligence and deep learning. Sensors (Basel) 2023;23(9):4184 [FREE Full text] [doi: 10.3390/s23094184] [Medline: 37177386]
18. Agrell B, Dehlin O. The clock-drawing test. 1998. Age Ageing 2012;41 Suppl 3:iii41-iii45. [doi: 10.1093/ageing/afs149] [Medline: 23144287]
19. Jannati A, Toro-Serey C, Gomes-Osman J, Banks R, Ciesla M, Showalter J, et al. Digital clock and recall is superior to the mini-mental state examination for the detection of mild cognitive impairment and mild dementia. Alzheimers Res Ther 2024;16(1):2 [FREE Full text] [doi: 10.1186/s13195-023-01367-7] [Medline: 38167251]
20. Banks R, Higgins C, Greene BR, Jannati A, Gomes-Osman J, Tobyne S, et al. Clinical classification of memory and cognitive impairment with multimodal digital biomarkers. Alzheimers Dement (Amst) 2024;16(1):e12557 [FREE Full text] [doi: 10.1002/dad2.12557] [Medline: 38406610]
21. Gomes-Osman J, Borson S, Toro-Serey C, Banks R, Ciesla M, Jannati A, et al. Digital clock and recall: a digital, process-driven evolution of the mini-cog. Front Hum Neurosci 2024;18:1337851 [FREE Full text] [doi: 10.3389/fnhum.2024.1337851] [Medline: 39253069]

XSL•FO
RenderX

22. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
23. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
24. Lawrence CE, Dunkel L, McEver M, Israel T, Taylor R, Chiriboga G, et al. A REDCap-based model for electronic consent (eConsent): moving toward a more personalized consent. J Clin Transl Sci 2020;4(4):345-353 [FREE Full text] [doi: 10.1017/cts.2020.30] [Medline: 33244416]
25. FDA converts novel Alzheimer's disease treatment to traditional approval. U.S. Food and Drug Administration. 2023. URL: https://www.fda.gov/news-events/press-announcements/fda-converts-novel-alzheimers-disease-treatment-traditional-approval [accessed 2025-01-14]
26. FDA approves treatment for adults with Alzheimer's disease. U.S. Food and Drug Administration. 2024. URL: https://www.fda.gov/drugs/news-events-human-drugs/fda-approves-treatment-adults-alzheimers-disease [accessed 2025-01-14]
27. Statement: broader medicare coverage of leqembi available following FDA traditional approval. Centers for Medicare and Medicaid Services. 2023. URL: https://www.cms.gov/newsroom/press-releases/statement-broader-medicare-coverage-leqembi-available-following-fda-traditional-approval [accessed 2025-01-14]
28. Fisher JA, Kalbaugh CA. Challenging assumptions about minority participation in US clinical research. Am J Public Health 2011;101(12):2217-2222. [doi: 10.2105/AJPH.2011.300279] [Medline: 22021285]

Abbreviations

CCE: Core Cognitive Evaluation
CDT: clock drawing test
CMS: Centers for Medicare and Medicaid Services
DCR: Digital Clock and Recall
FDA: US Food and Drug Administration
LHQ: Life and Health Questionnaire
MCI: mild cognitive impairment
MMSE: Mini-Mental State Examination
PCP: primary care physicians
REDCap: Research Electronic Data Capture

Edited by A Mavragani; submitted 20.09.24; peer-reviewed by K Possin; comments to author 29.10.24; revised version received 08.12.24; accepted 27.12.24; published 12.03.25.

Please cite as:
Doerr AJ, Orwig TA, McNulty M, Sison SDM, Paquette DR, Leung R, Ding H, Erban SB, Weinstein BR, Guilarte-Walker Y, Zai AH, Walkey AJ, Soni A, McManus DD, Lin H
Digital Assessment of Cognitive Health in Outpatient Primary Care: Usability Study
JMIR Form Res 2025;9:e66695
URL: https://formative.jmir.org/2025/1/e66695
doi:10.2196/66695
PMID:

©Adam J Doerr, Taylor A Orwig, Matthew McNulty, Stephanie Denise M Sison, David R Paquette, Robert Leung, Huitong Ding, Stephen B Erban, Bruce R Weinstein, Yurima Guilarte-Walker, Adrian H Zai, Allan J Walkey, Apurv Soni, David D McManus, Honghuang Lin. Originally published in JMIR Formative Research (https://formative.jmir.org), 12.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Teen Perspectives on Integrating Digital Mental Health Programs for Teens Into Public Libraries ("I Was Always at the Library"): Qualitative Interview Study

Ashley A Knapp[1], PhD; Katherine Cohen[2], MA; Kaylee P Kruzan[3], LCSW, PhD; Rachel Kornfield[3], PhD; Miguel Herrera[1], MEd; Aderonke B Pederson[4], MD; Sydney Lee[1]; Kathryn Macapagal[2], PhD; Chantelle A Roulston[2], BA; Kaleigh Clarke[5], BS; Clarisa Wijaya[6], MA; Robert Simmons[7], MSW; Latonia Jackson[7], BA; Simrandeep Kour[8], MA; Sandra Franco[9], MS, MEd; David C Mohr[3], PhD

[1]Department of Psychiatry and Behavioral Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[2]Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[3]Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[4]Department of Psychiatry, Harvard Medical School, Massachusetts General Hospital, Boston, MA, United States

[5]Bradley University, Peoria, IL, United States

[6]Department of Psychology and Psychiatry, University of Illinois Chicago, Chicago, IL, United States

[7]Oak Park Public Library, Oak Park, IL, United States

[8]Chitkara University, Chandigarh, India

[9]Chicago School of Education, Loyola University, Chicago, IL, United States

Corresponding Author:
Ashley A Knapp, PhD
Department of Psychiatry and Behavioral Sciences
Feinberg School of Medicine
Northwestern University
750 North Lake Shore Drive
Chicago, IL, 60611
United States
Phone: 1 3125033751
Email: ashley.knapp@northwestern.edu

Abstract

Background: Rising rates of anxiety among teens necessitate innovative approaches for implementing evidence-based mental health support. Public libraries, seen as safe spaces for patrons with marginalized identities, offer free public services such as broadband internet access. Many teens spend significant amounts of time in their local libraries due to the safety of this space as well as the trusted adults working there. The American Library Association has shifted its priorities to focus more on mental health through employing social workers and providing mental health programs. As such, public libraries may be promising sites for the implementation of digital mental health (DMH) programs for teens.

Objective: This study aimed to examine how teens who attended their local public library experienced and managed their anxiety, what mental health supports they were interested in receiving, and how DMH programs and public libraries can meet their needs.

Methods: We interviewed 16 teens aged 12-18 (mean 15.2, SD 2.0) years who used the library frequently at the time of the interviews. Of these teen patrons, 56% (9/16) identified as female, 31% (5/16) identified as male, and 12% (2/16) identified as nonbinary. Most (11/16, 69%) identified as either White or Black or African American individuals, with the remainder (5/16, 31%) identifying as Hispanic or Latino or Chinese American individuals or with ≥2 races. The interviews were individual and semistructured, designed to elicit recommendations for designing and implementing digital tools in libraries to improve teen mental health. Interview transcripts were coded by multiple coders using thematic analysis to synthesize key themes.

Results: Teens reported experiencing uncontrollability, unpredictability, and anger related to their anxiety, which they managed using strategies such as guided breathing, distress tolerance, and social connection. They also talked about other helpful management techniques (eg, progressive muscle relaxation, journaling, and mood tracking). Teens underscored the importance of pairing mood

XSL•FO
RenderX

tracking with daily activities to reveal patterns. They also stressed the significance of context and anxiety severity when choosing anxiety management strategies. Teens underscored the centrality of the public library in their lives and their view of it as a safe space where they can easily access resources and connect with friends and trusted adults. When considering the design of a DMH program implemented in libraries, they suggested including personalization for different identities, gamification, and simple navigation. Teens emphasized the importance of protecting their privacy within digital programs and that their end goal was to use the skills learned in the DMH program *offline*.

Conclusions: Teens who frequently used their local public library expressed interest in receiving digital tools via libraries to help them manage anxiety. Their recommendations will help inform future research on the adaptation and implementation of DMH programs for teens in public libraries.

(JMIR Form Res 2025;9:e67454) doi:10.2196/67454

KEYWORDS

public libraries; digital mental health; teens; youth; adolescents; anxiety; mental health; implementation; safe spaces; mobile phone; smartphone

Introduction

Background

Researchers, clinicians, and policy makers in the United States have declared a teen mental health crisis due to increasing rates of internalizing distress combined with limited access to mental health programs [1-3]. The Centers for Disease Control and Prevention found that over one-third of teens in 2019 experienced persistent feelings of hopelessness, a transdiagnostic predictor of anxiety and depression [4,5]. Anxiety is particularly common among teens, with rates increasing roughly 27% between 2016 and 2019 [6] and continuing to rise since the pandemic [7]. Prevention and treatment efforts for teens have the potential to reduce the likelihood that internalizing disorders will persist or worsen in adulthood [8], but demand far exceeds the available supply of teen mental health programs [9,10]. Increasing the number of providers would not fully ameliorate the issue as many teens face additional barriers such as high treatment costs, lack of transportation to clinics, and complex regulations regarding parental consent [11]. Consequently, advocates call for researchers and clinicians to meet teens where they are by implementing mental health interventions in settings frequently used by teens [12-14]. In this paper, we focus on public libraries as a promising yet understudied setting for delivering teen mental health programs, particularly programs targeting anxiety.

Many teens, especially those with marginalized identities, spend significant amounts of time in libraries [15]. Public libraries are a *third place*, where patrons can spend time without being expected to spend money [16,17]. The commercialization of space, particularly in urban areas, has led to libraries being one of the few indoor spaces teens can inhabit for free without facing accusations of *loitering*. As a result, libraries have transitioned in recent years to become beacons of public services, offering health services, professional development classes, art workshops, and more [18]. An increasing number of libraries now offer teen-specific programming [19-21]. Libraries are also seen as a *safe space* for teens with marginalized identities, including historically underserved racial and ethnic minority (HURE) teens; teens who identify as lesbian, gay, bisexual, transgender, queer, intersex, or asexual (LGBTQIA+); teens living in lower-resourced and disinvested areas; and teens living with mental health conditions [15,16,18,22-27]. Unsurprisingly, the teen mental health crisis is visible in public libraries. Within the last decade, the American Library Association has shifted its priorities to focus more on mental health through actions such as employing social workers within public libraries and providing patrons with mental health programs [25,26,28]. At the same time, librarians report wanting more mental health support for teen patrons that is accessible and on demand and requires little additional training for already overtaxed staff [15]. Accordingly, interventions that leverage digital technologies, or digital mental health (DMH) programs, may be well suited for use in libraries.

DMH programs are an ideal way to reach teens for several reasons. Teens commonly use the internet for mental health support and psychoeducation [29]. DMH programs can be implemented in practically any location at relatively low cost. Libraries offer free broadband internet access and computer access to all patrons, allowing DMH programs to be easily accessed in this setting. Self-administered DMH programs provide availability on demand when teens need support, eliminating the need to involve a facilitator or caregiver and also satisfying the needs of teens for independence and autonomy [11]. Despite the promise of DMH programs, many face problems with user engagement and sustainment by organizations outside the research context [30,31]. Researchers have suggested that this may reflect, in part, that DMH programs are typically designed for research contexts rather than for implementation in real-world settings [32,33]. On average, it takes 17 years [34] for research evidence to reach real-world settings and change practice. These statistics have not significantly changed in decades, highlighting the need for innovative solutions to address this research-practice gap.

To ensure that DMH programs are implemented in real-world settings, researchers may look toward approaches that (1) center the community who will use the program, (2) design the program according to the community's needs, and (3) evaluate implementation metrics outside of effectiveness. Community-based participatory research (CBPR) is an approach to research that prioritizes partnership and shared decision-making with community members impacted by the research [35]. The application of CBPR principles in the development of DMH programs ensures that DMH programs

XSL•FO
RenderX

can positively impact the community of interest. Human-centered design (HCD) is defined as a set of methods that require that "the human perspective is considered from the initial conception to the eventual design, and that the people who will use and/or be affected by the designed product are involved in the process" [32]. Similarly to CBPR, HCD elevates the end users' voices in intervention design and implementation, increasing the likelihood that design outputs will be acceptable and sustainable. Finally, implementation research (IR) involves studying how evidence-based practices can best be implemented in real-world settings outside of research contexts (eg, examining determinants of and strategies for implementation and evaluating and designing for the deployment context) [36]. In combination, CBPR stipulates principles for partnering with a community that may use a DMH program, HCD provides methods for effectively creating the DMH program such that end users will be engaged, and IR plans for and evaluates the deployment of the DMH program outside of a research setting. The frameworks improve the likelihood that DMH programs will have high uptake and a meaningful and sustained impact beyond research.

Across CBPR, HCD, and IR frameworks, the first step in creating interventions that are engaging and sustainable is to understand and center the needs, goals, and preferences of end users, their communities, and eventual deployment settings [31,33]. For example, previous research suggests that teens often experience anxiety through somatic symptoms and excessive worrying. However, it is unknown whether these findings fully generalize to the experiences of teen library patrons or whether there are additional factors to consider that are unique to the community they are in [37,38].

Objectives

The purpose of this study was to use CBPR, HCD, and IR methods to understand the lived experiences of teen library patrons related to anxiety, elicit recommendations on what teen patrons would like to see in a DMH program for anxiety and anxiety management, and understand contextual factors and teens' perspectives of the public library as a deployment setting. Our goal was to use this information to inform the adaptation of DMH programs that can be implemented in public libraries to meet teens where they already are and help alleviate the teen mental health crisis [15].

Methods

Setting and Participants

Building on our previous needs assessment with library workers [15], this study examined teen patrons' needs and preferences related to the implementation of a DMH program for anxiety within a public library that serves communities in the west side and west of Chicago, Illinois. We will use *teens* or *teen patrons* to describe the teens who use the library frequently as this is the term the library community and teens use. The library serves community members from diverse racial and economic backgrounds, prioritizing the needs of marginalized patrons and those most affected by health inequities in their programming and the services they offer the community. AAK interviewed 16 teens aged 12-18 (mean 15.2, SD 2.0) years who used the library frequently at the time of the interviews. The interviewer held a master's degree and doctoral degree in experimental psychology, was an assistant professor, and identified as female at the time of the interview. We define *frequently* as those teens who either attended weekly programming at the library or visited the library during the week. Of these 16 teen patrons, 56% (n=9) identified as female, 31% (n=5) identified as male, and 12% (n=2) identified as nonbinary, and 44% (n=7) identified as White individuals, 25% (n=4) identified as Black or African American individuals, 12% (n=2) identified as Hispanic or Latino individuals, 12% (n=2) identified with ≥2 races, and 6% (n=1) identified as Chinese American individuals.

Procedures

All participants were recruited from a local Chicagoland public library between 2020 and 2021. The library's Teen Services staff coordinated with the research team to post study flyers in the library common areas, promote the study in the library's Teen Services programming, post in the library's newsletter, and receive referrals from other teen patrons. Recruitment occurred until saturation was achieved in the interviews. No participant dropped out or refused to be part of the study; no repeat interviews were conducted. During the first year of the COVID-19 pandemic, the physical library was shut down; however, programming for teens continued via Zoom (audio only). Our research team built rapport with teens in person before the pandemic onset (and then once the physical library opened again), as well as attending the internet-based weekly programming events led by teens when the library was shut down. A total of 16 teen patrons participated in a 1-hour, individual needs assessment interview with members of the research study team.

Due to COVID-19 restrictions, all interviews were conducted over the phone or using audio-only Zoom calls (Zoom Video Communications). The semistructured interviews covered how participants experience and manage anxiety, current use of technology, and current or future library programming focused on teen mental health (Multimedia Appendix 1). Participants were also asked about what resources or features within a digital program may help them better manage anxiety. All interviews were audio recorded, transcribed, and uploaded to Dedoose (SocioCultural Research Consultants) [39] for qualitative data analysis. If teens expressed thoughts of suicide or self-harm behaviors during the interview, a predetermined safety procedure was in place to assess risk of imminent harm and take appropriate action (including safety planning for those deemed low risk and connecting the teens with emergency services for those at high risk). While some teens expressed experiencing self-harm thoughts in the past, no teen expressed current suicide or self-harm thoughts or behaviors, so the safety protocol was not enacted.

Data Analysis

In total, 3 researchers (AAK, MH, and CW) conducted a thematic qualitative data analysis of the transcribed interviews. These coders all had training in qualitative analysis, all held master's degrees, and one also held a doctoral degree. Coders used an iterative approach to create and refine the codebook. No field notes were used in the analysis. First, coders read and annotated 3 interview transcripts each, labeling relevant words,

XSL•FO
RenderX

phrases, or sections with codes. Then, coders met to discuss and conceptualize the annotations, identify highly salient codes, and group codes into broader categories. The coders met regularly to discuss the application of codes, resolve discrepancies in code application, and refine the codebook accordingly. Next, coders were randomly assigned to apply the updated codebook to 5 or 6 additional interview transcripts each, after which they met to resolve discrepancies and further refine the codebook by removing codes that were underused or not present within the transcripts. After these iterations, the final codebook was applied to all transcripts, encompassing the following key themes: (1) public libraries and mental health, (2) teen anxiety experiences and management, and (3) DMH program design for teens and implementation considerations (see Multimedia Appendix 2 for the COREQ [Consolidated Criteria for Reporting Qualitative Research] checklist). The community members were primarily involved in member checking [40-43] in that they were most interested in reading the first drafts of the paper and verifying that the main concepts identified by researchers in the qualitative data analysis (and, thus, highlighted in the *Results* section) were an accurate representation of the teen patrons' feedback and also of the library itself. The summary of the key themes is shown in Textbox 1.

XSL•FO
RenderX

Textbox 1. Summary of key themes and subthemes from the qualitative analysis conducted on the transcripts of the individual semistructured interviews with teens who frequently used a local Chicagoland library between 2020 and 2021. The interviews covered how teens experienced and managed anxiety; current use of technology; and current or future library programming focused on teen mental health, including digital mental health (DMH) programs for anxiety (Multimedia Appendix 1).

Public libraries and mental health

- Most teens liked the centrality and easy access of the library.
- The library provides teens easy access to reliable mental health resources to connect with other teens on this subject.
- Teens from historically underserved racial and ethnic minority backgrounds and other marginalized identities perceived libraries as safe spaces.

Teen lived experiences of anxiety

- Uncontrollability: teens identified catastrophizing, overthinking, and a lack of control over thoughts, emotions, and actions as feelings of anxiousness. For some, anxiety led to feelings of sadness and depression and thoughts of self-harm. The cyclical pattern resulting from the co-occurrence of anxiety and depression symptoms severely impacted teens' drive and motivation. The uncontrollability of anxiety led teens to feel burned out and tired, with a sense of dissociation from reality and derealization.
- Unpredictability: new experiences were a major source of anxiety related to unpredictability. Experiencing an unpredictable situation prompted urges to escape the situation or avoid future anxiety-provoking situations.
- Anger: anxiety can lead to teens getting easily frustrated, having a short fuse, and feeling angry. Sources of anger ranged from irritation over mistakes to frustration and bitterness toward personal anxiety. External annoyance and bitterness usually ended in arguments with family and friends.

How teens manage anxiety

- Breathing and progressive muscle relaxation: teens used apps with guided breathing, in-the-moment breathing exercises prompted by devices tracking physical changes, and progressive muscle relaxation to help regulate high anxiety levels.
- Grounding and distress tolerance: grounding techniques were used to disrupt uncomfortable thoughts and feelings. One method was the 5-4-3-2-1 technique, where one names things related to each sense. Another anxiety-managing technique used the sense of taste, namely, salty and sour tastes.
- Social connection: engaging in various activities with loved ones, namely, talking to and connecting with family and friends through conversation, helped with easing anxiety. Teens with anxiety felt validated and less isolated when sharing experiences with other teens or mental health professionals. Teens suggested an anonymous social component in a DMH program serving as a safe space for them to share experiences and advice with one another. Teens also wanted access to a mental health professional to provide advice in a safe space.
- Journaling and mood tracking: teens used journaling to work through and release their anxiety, helping them gain confidence and motivation to address stressors. Teens appreciated when mood tracking was used to reveal patterns such as relationships between their anxiety and daily activities. Some teens desired a pop-up mood tracker when first opening the program that could tailor mood management techniques to their input responses (eg, if high anxiety is reported, the program directs teens to in-the-moment anxiety relief techniques).
- Psychoeducation and crisis resources: teens found psychoeducation and crisis resources to be useful in contextualizing and learning the patterns and origins of their anxiety. Teens emphasized the importance of warm handoffs and easy access to reliable resources when in a crisis.
- Strategy choice and context: teens wanted a wide range of anxiety-mitigating techniques. Having variety allows teens to experiment and figure out what works best for them. Teens noted that context and anxiety levels played important roles in choosing a technique. The main goal of the resources should be to allow for practicing skills to eventually be used without technological assistance.

Desired DMH program features

- Personalization: teens liked app algorithms that curated or suggested content based on feedback, use, and data entry, as well as reports on patterns. They emphasized the need to understand who the program is targeting and provide resources specific to that population.
- Aesthetics and layout: teens preferred simple and concise app layouts that made navigation of the resources easy. Teens recommended that apps should be low commitment as longer materials were hard to keep up with and became overwhelming. They favored content delivered in teen-friendly language to keep the program more relatable and up-to-date.
- Gamification and animations: videos, animations and graphics, and mini games were the best modalities for information acquisition and engagement for teens. Content would be even more compelling if it featured teen celebrities or popular social media influencers. Teens liked interactive digital features, such as haptic and touch screen features, while practicing skills.
- Privacy and accessibility: protecting teens' privacy was a top priority—teens worried about parental access to and surveillance of their phones and parents potentially finding out that they were struggling with mental health issues. Teens noted school rules and caregivers restricting phone use as access barriers. Paid subscriptions, an overt mental health focus in case of onlookers, and phone limits (eg, low storage and limited or inconsistent broadband access) were access barriers to resources on apps.

XSL•FO
RenderX

Ethical Considerations

All research protocols were approved by the Northwestern University Institutional Review Board (STU00212314) before study enrollment. Both the interested teens and their caregivers provided permission to take part in the study. In the process of transcribing interviews, all potential identifying information was removed to protect the privacy of the participants, and each transcription was labeled as *P1*, *P2*, etc. Transcripts were not returned to participants. Each adolescent was given a US $25 Visa gift card for their time and participation in the interviews.

Results

Public Libraries and Teen Mental Health

Teens discussed resources they enjoyed at the library, the impact the library had on them, and their opinion on how libraries could support teen mental health. Most teens talked about the centrality and easy access of the library. P13 stated the following:

> *Because of where I get off at the train on the metro, I get off right across the street from the library so I can go to the library and just chill there.*

Other teens described the library as a community access point. They noted that the library has many connections throughout the community and serves as a hub for resources.

When reflecting on the role public libraries should play in teen mental health, most teens agreed that the library is a great place for teens to access reliable mental health resources and connect with other teens about mental health via library programming. While some teens had specific ideas for new mental health offerings, such as creating a website or smartphone app to connect teens to resources to manage their mental health, they also spoke about resources that the library already had in place. For example, P8 described the following:

> *Like obviously books are a resource—and kind of just having that level of flexibility whether it's mentorship or whether it's grounding yourself in a story or whether it's a pamphlet...having those different opportunities for how people have different or how different people interact with their own mental health and coping mechanisms.*

Teens emphasized how the library was a safe space for them because of the trusted adults that worked there, friends they made at the library, and the library's intentional teen programming. P12 summarized this nicely:

> *I was always at the library, and I was always doing these groups and these fun summer clubs. I had a lot of friends from the library...I do think that the library is a great resource place. There's a lot of good, trusted adults. It's quiet. And sometimes in a hectic household or a hectic friend fight, you might just wanna go to a place and feel welcomed and feel quiet and feel calm.*

Another teen, P2, underscored their admiration for the trusted adults available via their teen programming at the library:

> *Even though you're talking to an adult, you're just talking to people you know that care about you.*

The library was perceived as a safe space for many HURE and LGBTQIA+ teens in our sample, as well as those teens most experiencing inequities. Several of these teens mentioned that a social justice program within the library led by a local activist helped create a safe space in the library. In addition to weekly meetings for this program, teens researched and led town hall meetings educating the public on issues such as climate change's disproportionate effect on lower-resourced communities. When asking P3 what they liked about this program, they responded the following:

> *For well me as like you know like a black, queer individual in every sense of the word, obviously because we [the group] love each other and stuff it's great, but it's also like I can just talk about this stuff...about stuff that I usually wouldn't talk to normal people you know your average Joe or whatever about—talking about the oppression and the prison system and mass incarceration and all that stuff...But it definitely just helps so I can get all of those problems or frustrations out and also just for mental health just seeing these people [in the group]...It's just like we're a family and that's really nice.*

As can be seen in this quote, one HURE teen described finding solace in the shared experiences and similar backgrounds within this group, which allowed them to safely discuss topics including structural racism and oppression. Another teen, P13, described the following:

> *It's just nice to have a space where I can be Black and annoyed and then there's other people who get it.*

Teens who attended this particular library program also emphasized their appreciation for the trusted adult who led the group, highlighting that the local activist empowered them and elevated their voices, challenging them to think critically while researching and reporting on current issues.

Teens' Lived Experiences of Anxiety

Overview

Teens shared about their experiences with anxiety, which impacted them in terms of physical sensations (eg, racing heart, difficulty breathing, and feeling shaky or flushed) and socially (eg, being easily embarrassed in public and worrying about disappointing or being judged or rejected by others). They also shared experiences less addressed in the teen anxiety literature; specifically, they spoke about their feelings of uncontrollability and unpredictability and the role of anger in their anxiety. As there is already a robust literature on physical sensations and social anxiety, we focused on these less explored aspects of teens' experiences of anxiety.

Uncontrollability

Teens spoke about catastrophizing and overthinking as part of their anxiety, as well as feeling like their thoughts, emotions,

XSL•FO
RenderX

and actions were out of their control. For example, P15 described the following:

> *My thought train is just going a million miles an hour. It's like this thought leads to this other thought and it turns into a big fireball of negative thoughts...like out of control.*

In addition to feeling a lack of control, some teens' anxiety led to feelings of sadness and depression and thoughts of self-harm. P13 described their experience as follows:

> *For me, it's usually like the anxiety comes and then the sadness once I cool down. So, I guess comparing it to a storm, it's like the anxiety is the hurricane and then the sadness is kind of like the eye of the storm and then the anxiety comes back and then there's a hurricane that follows it.*

This teen highlighted the cyclical nature of symptoms, an experience shared by many others. Teens went on to describe how the co-occurrence of anxiety and symptoms of depression had a severe impact on their motivation, ultimately taking a toll on their grades and relationships with friends and family.

Finally, the uncontrollability of anxiety led some teens to feel burned out and tired, causing them to zone out, operate on autopilot, and shut down. They described feeling a sense of dissociation and derealization because of their uncontrollable anxiety. P4 described the following:

> *Dissociating and not feeling like I'm really present. Just kind of I become less aware of my surroundings.*

They went on to describe an escalation in which they felt completely detached from reality.

Unpredictability

In addition to the uncontrollability of their anxiety, several teens spoke about the unpredictable nature of their anxiety, especially when it came to new experiences. For example, P7 explained the following:

> *If I go somewhere new, I think like, "Where is everything?" because I don't know where anything is and then like, "What's gonna happen? What am I gonna do? What are people gonna be like? Are people gonna judge me?"*

Teens commonly described that experiencing an unpredictable situation led to urges to escape it or avoid future anxiety-provoking situations, such as new experiences. P9 described their reaction to new social situations:

> *I feel like people are staring at me. And all I wanna do is just hide usually. So, I try to get out of the situation, like wherever I am in public.*

Anger

Several teens described getting easily frustrated, having a short fuse, and feeling very angry when anxious. Some internalized their anger, becoming irritated at themselves for making a mistake or experiencing frustration and bitterness toward their own anxiety and anxiety-provoking situations. Other teens described feeling annoyed and bitter externally, like they "needed to go punch something" (P16). This usually ended in arguments with family and friends, adding to their anxiety and feelings of isolation.

Teen Anxiety Management Strategies

Overview

Teens described using multiple strategies to manage their anxiety that included internally focused strategies for self-management as well as externally focused strategies for connecting with others and seeking resources. Teens reported diverse experiences with mental health treatment and resources—some described experiences with formal therapy, others used smartphone apps for mental health, some had experiences with both, and others had no previous experience engaging with these resources. We describe teens' experiences in the following sections to inform future interventions that can promote and reinforce strategies that teens already feel comfortable using and that are already integrated into their daily lives. These methods of coping are well aligned with evidence-based strategies.

Breathing and Progressive Muscle Relaxation

Most teens mentioned that breathing exercises helped them manage anxiety in the moment. For example, P7 described using breathing exercises while having catastrophic and overwhelming thoughts:

> *Because I focus on my breathing instead of my thoughts. It kind of clears my mind too.*

Teens also described using apps with guided breathing exercises (eg, facilitating box breathing and deep breathing) alongside interactive visualizations (eg, haptic and touch screen features). P6 described the following:

> *I really like the ones that are like circles that tell you to hold it in—that says to breathe in and hold it, then breathe out and hold it.*

Other teens enjoyed apps that used inputs from other devices, such as smartwatches or exercise wristbands, to trigger breathing exercises when their heart rate increased. Other teens found utility in using breathing exercises as a supplemental strategy that could help once they had regulated their high anxiety levels using other strategies described in this section. Another anxiety-reducing technique that teens reported using in apps was progressive muscle relaxation. When describing a smartphone app that included this technique, P7 stated the following:

> *My favorite is where you tighten your muscles, which is like you ball your fists up and then relax them and stuff...I just kind of like it because it's like you have all your stress and all your feelings and then you just let them go.*

Grounding and Distress Tolerance

Several teens spoke about grounding and distress tolerance techniques, which were used to disrupt uncomfortable thoughts and feelings and then return to a more manageable level of anxiety. Some mentioned using the 5-4-3-2-1 technique when at their highest levels of anxiety—or, as P7 called it, "the five

XSL•FO
RenderX

senses thingy"—where they named things related to each sense (eg, name 5 things they saw, 4 things they felt, and so on). When paired with breathing exercises, this technique made their anxiety manageable. Another teen, P6, described a unique technique:

> *When I feel really anxious, what I do is put salt on my tongue. So, that kinda just it distracts me from everything that I'm thinking about...It just disconnects me from reality for a bit because of the strong taste of salt messes up everything else that I was thinking about.*

This teen mentioned doing the same exercise and achieving the same results with sour-tasting things such as lemons, limes, or vinegar.

Social Connection

All teens spoke about the importance of social connection in dealing with their anxiety even if social contexts brought on anxiety. P8 described this phenomenon as follows:

> *As much as I do have social anxiety, I'm a very social person.*

The activities with friends and family that helped ease their anxiety ranged from low-key, in-person activities (eg, watching movies or listening to music together) to internet-based activities (eg, where teens would play video games together). The simple act of talking to and connecting with friends and loved ones was the most mentioned social activity that helped in easing anxiety, with teens emphasizing how these types of "conversations are really key in those [high-anxiety] moments" (P2). P8 described how conversations such as these helped:

> *Whether it's just completely rambling you know like distracting me—or guiding me through breathing exercises or going on a walk with me and make sure I keep moving and that I'm paying attention to my surroundings like, "Oh, do you see that over there?" and that kind of thing and just helping to ground me and stuff.*

Teens were asked to brainstorm how a DMH program could provide social connection to help with anxiety. Teens reflected that they felt validated and less isolated when reading about other teens' experiences of anxiety and how they dealt with it. Accordingly, most teens suggested including an anonymous social component within a DMH program that would serve as "a safe space to talk to each other and see what people are going through" (P5). In addition to these shared experiences and advice, some teens wanted connection with a mental health professional or monitors to provide professional advice while also keeping it a safe space. P8 spoke to the benefits of this kind of platform:

> *Being able to ground yourself in other people's experiences so that you can ground yourself in your own and feel more validated or feel making anxiety more approachable.*

Journaling and Mood Tracking

Teens shared how journaling and mood tracking helped them alleviate or better understand their anxiety. P3 described how they used journaling to work through their anxiety:

> *After I'm done being upset in the moment, I'll just write about that moment and it just kind of just helps me release those emotions—get them down so I remember them, remember what happened.*

This cathartic nature of journaling was shared by most teens who used it as an anxiety management strategy. P14 had a unique technique where they would create characters and storylines based on a hardship or anxious situation they were experiencing. This teen garnered confidence and motivation to address stressors through these stories. P14 described how their story creation helped:

> *Yeah, it definitely takes me out of the moment and then it also it just feels like if I can beat this big milestone and do that and achieve that and then have it, homework's nothing. I can finish that homework assignment because I can do this. I can do this. I'm more powerful than this homework. I will squash it, you know? Like a bug...It's just a confidence thing.*

Other teens saw value in reviewing patterns in their mood-tracking data, such as relationships between their anxiety and daily activities. P4 described the following:

> *It's nice to see a daily tracker so you can see the bigger trend of what causes your stress and what periods and times you're more stressed.*

Teens desired a mood tracker that pops up when first opening the program so that their answers could tailor their interaction with the program. For example, if they entered a high anxiety rating, the program might direct the teens to in-the-moment anxiety relief techniques.

Psychoeducation and Crisis Resources

Several teens described the benefits of including psychoeducation and crisis resources in a DMH program embedded within the library. Teens found psychoeducation helpful in contextualizing and learning the patterns and origins of their anxiety. P1 expanded on their idea for integrating psychoeducation and resources into a DMH program:

> *I think it would be helpful to have a resources page...articles on anxiety, quick advice, I guess, and then maybe also a hotline page.*

Teens also mentioned the importance of warm handoffs to reliable resources and the need for easy access to these resources when in a crisis. For example, P16 suggested that the future program should "definitely start off with, definitely a phone number to a social worker that if someone is feeling anxious, or they wanna hurt themselves, you can talk to them."

Strategy Choice and Context

Overall, teens wanted a variety of anxiety-mitigating strategies to choose from as different strategies could work better or worse for different teens and as variety allows them to experiment to find what works best for them. Context and anxiety level played

XSL•FO
RenderX

important roles in choosing a technique. For example, some teens described that they typically had their phones with them in public and would use social media or other apps to deal with their stress in the moment. P7 described the following:

> *It depends on where I am, but if I'm at home I'll play my guitar, listen to music, or I'll do meditation sometimes and sometimes I'll draw...But if I'm in a public place, then I'll do deep breathing.*

Other teens expressed that their approach depended on how anxious they were, with some strategies (eg, physical activity) being better suited for high-anxiety moments and others (eg, psychoeducation and listening to music) being better suited for lower-anxiety moments. Regardless of the coping skill or how it was taught, the consensus from teens was that the main goal should be practicing skills to then use them without technological assistance when in-the-moment anxiety occurs. One participant described that "it's not just about the technology or about the app" (P8). Most teens appreciated the supportive technologies they had previously used that helped them practice exercises, which in turn made it easier to use the acquired skills outside the app or in real life.

Desired DMH Program Features

In addition to specific anxiety management strategies, teens discussed the top features they would like in a DMH program embedded within the library. These desired features fell under 4 broad themes: personalization, aesthetics and layout, gamification and animations, and privacy and accessibility.

Personalization

Teens discussed the importance of being able to both personalize and receive tailored information from programs. For example, P15 suggested the following:

> *You know how you can type something into WebMD and it will show what the problem is and also how to fix it. I think that's what this app should look like. If you have anxiety over a bad grade, you can do like, "What can I do to overcome my anxiety about that or stress?"*

As this example suggests, teens liked when apps used algorithms to curate and suggest tailored content based on their feedback, use, and entry of data into that app. Teens also appreciated reports on their particular use patterns (eg, features most used) and shortcuts to view these data. In addition, it was important that the mental health program offer resources specific to the population it was designed for. For example, P13 described the following:

> *One thing that I notice a lot with resources where people put resources for like, "Did you know about this and blah, blah, blah?" I'd be going through it, and it would be like, "This is clearly meant to educate white people."*

Several HURE teens underscored that current mental health resources were catered to a small group of people, White teens in particular, and did not reflect them or their mental health needs.

Aesthetics and Layout

Many teens preferred simple digital program layouts that were easy to navigate. They described being overwhelmed if the home screen of the program had too many options. For example, P11 said that they did not want the following:

> *A lot of folders or that you have to click a lot of things to get to something. I think it should be the important things, and then from there it can branch into more specific ways to cope with types of stress or recommendations or things like that.*

Other teens emphasized that apps should be low commitment (ie, relatively quick and easy to navigate and use on a daily basis). Apps with higher commitment, or those that required significant time or energy to use, and those with longer materials were difficult to keep up with and quickly became overwhelming. Furthermore, teens shared that it would be important to have regularly updated content in teen-friendly language. P8 said the following:

> *Teens can be turned off by aesthetics that just feel kind of like dated or they feel like they were made by people much older than them.*

Gamification and Animations

Several teens underscored that the best modalities for information acquisition and engagement were videos, animations and graphics, and mini games. P5 said the following:

> *Like a video lesson or something that the user could interact with or participate along with because there are many resources out there for a number of things...words don't help as much as seeing the person or listening.*

They went on to describe that using teen actors in videos could become cheesy but videos could be compelling if they featured teen celebrities or popular social media influencers. Teens also enjoyed interactive smartphone or tablet features (eg, haptic and touch screen features) where they could touch the screen while practicing skills. Teens gave examples that ranged from using the notification vibrations on the phone to guide deep breathing exercises to dragging their finger across the screen until an image was revealed.

Privacy and Accessibility

Protecting the teens' privacy was among the top priorities. P6 emphasized the following:

> *Making sure that my information's safe and only to me. It's not shared with anyone else, especially, random people.*

Other teens expressed worry regarding their parents' access to and surveillance of their smartphones and them potentially finding out that they were struggling with mental health issues. P16, for example, stated the following:

> *It's just my parents, every week, they checked my phone. And if they saw I was searching, "How to overcome anxiety," I wouldn't want them to make me the center of attention. And make them drop everything that they're doing.*

XSL•FO
RenderX

Some teens had their smartphone use restricted by caregivers or school rules. Policies ranged from not using smartphones during school hours to caregivers taking away teens' phones as punishment (eg, when they said "something bad") or at specific times (eg, being unable to use anything other than the music app on their smartphones after 8 PM on weeknights). Teens were also concerned about paid subscriptions, onlookers being able to see that the program was for mental health, and smartphone-specific barriers (eg, low battery, storage, and data of phone and inconsistent access to broadband). Many teens spoke about liking a particular app, such as a mood-tracking app, but reported that they stopped using it because they had to use their cellular data or pay to use the full version of the app, further limiting their use.

Discussion

Principal Findings

This study aimed to understand the needs and perspectives of teen library patrons in an effort to support the creation of a teen-focused DMH tool that can be disseminated in libraries. The results of the needs assessment interviews with teens who frequently use the library provided critical information on their perspectives of the library's function and programming. Teens underscored the centrality of and, thus, easy access to their public library. They also referred to the library as a hub for connecting patrons to community resources and services. This is consistent with our previous research, in which library workers identified a main function of the library to be a conduit to guide patrons to desired resources [15].

Teens talked about the teen-specific programming they attended at their library, indicating that this is where they met a lot of their friends and fostered relationships with the trusted adults leading those programs. These factors led teens to view the library as a *safe space*, which is consistent with several other reports by library workers and researchers underscoring that communities see public libraries as safe spaces [27,28]. HURE teens and teens with marginalized identities (eg, LGBTQIA+ teens, teens living in lower-resourced and disinvested areas, and teens living with mental health conditions) found belongingness, companionship, and solidarity in library programming that spoke to different aspects of their identity. Given that there are more unsafe spaces for these groups due to marginalization, it is critically important to invest in and maintain safe spaces that bring individuals with shared experiences and similar backgrounds together and that celebrate the different aspects of their identities, especially within the teen population.

Teens shared the nuanced experiences they had regarding their anxiety. They spoke to the uncontrollable and unpredictable feelings associated with their anxiety, which led some teens to catastrophize, feel depressed, or have thoughts of self-harm, which then cyclically fed back into their anxiety. Other teens described dissociating to cope with adverse feelings or feeling the need to avoid or flee anxiety-provoking situations. There is evidence of the uncontrollable and unpredictable nature of anxiety in the adult anxiety literature, especially in relation to panic disorder, yet those variables have not been well studied in the teen anxiety literature and are worthy of future research [44-47]. More research is needed to better understand these 2 variables and their potential role as risk or maintenance factors, which could then be used as targets for DMH programs for anxiety and other mental health conditions in teens [48-52].

Other teens talked about their anger in relation to their anxiety. Some described the internal aspects of anger that may be expressed as frustration toward their own anxiety or frustration about making mistakes. Others described the external aspects of anger as experiencing frustrations that often led to conflicts with friends and loved ones. This finding is aligned with those of extant literature showing that anxiety and depression commonly manifest as anger and irritability in teens [4,48-53]. In addition, there is evidence in the literature suggesting that anger can be a prominent symptom co-occurring with several internalizing disorders [54]. Stress and anger as a mood symptom may be connected to experiences of racism among teens [54]. Similarly to other studies [54-56], our study shows that anger and racism-based stress play a role in the emotions of teens who may feel disenfranchised by the current status quo in society. Anger is often seen as a maladaptive response to stress; however, with regard to racism-triggered stress, anger can be an expected expression, and sometimes this expression allows for healthy processing of emotional stressors. Given that most of the teens in this study identified as HURE or multiracial, along with the literature indicating that anger is a known response to stress related to structural racism and discrimination, additional research is needed to better understand the relationship between anger (as a common stress response to racism) and anxiety among teens who are Black or Indigenous or hold other racial and ethnic identities. While the most potent intervention would be to address the ultimate problem of structural racism and discrimination, interventions that provide support and validation to HURE teens in the face of systemic racism and oppression are critical. For example, these programs could be offered as part of library services and provide teens with acknowledgment of factors such as privilege, oppression, and discrimination and ways to curate social safety, as well as coping strategies to alleviate the mood symptoms caused by racism [51,57,58].

Teens described a variety of coping strategies (eg, guided breathing, muscle relaxation, social connection, distress tolerance, and mood and behavior tracking) that they used to help feel more in control of their anxiety, lessen their symptoms, increase knowledge of anxiety patterns, and increase their motivation to approach rather than avoid anxiety-provoking experiences. For example, teens gained knowledge about their anxiety when mood tracking was used to reveal patterns such as relationships between their anxiety and daily activities. They liked this because it emphasized behaviors that they could change or control that led to observable effects on their mood. When applied to the design of a DMH program for teens, it may be beneficial to include some of these strategies as on-demand activities. Leveraging haptic and touch screen features to actively engage teens in these strategies, especially strategies that they are familiar with and are already integrated into their daily lives, may also be worthwhile [59,60].

XSL•FO
RenderX

In terms of overall goals of a DMH program, teens wanted to learn a variety of anxiety management strategies and, ultimately, be able to choose the most effective strategy in particular contexts (eg, exercise for high-anxiety moments, listening to music for low-anxiety moments, playing guitar at home, and deep breathing exercises in public). Interestingly, while they wanted the DMH program to facilitate learning of skills, they ultimately wanted to be able to practice these skills without technological assistance when in-the-moment anxiety occurred. This use goal is consistent with other reports from the DMH literature; for example, Latino and Latina [61] teens indicated that their main goals for using DMH programs were to practice skills within the program and then, ultimately, be able to use the learned skills *offline*. If teens are using DMH programs with this goal in mind, this could have implications for how we interpret engagement metrics with teen DMH programs [62]. Teens' low, sporadic, or drop-off in engagement with a DMH program may not necessarily reflect a poor outcome or poor retention; rather, it could be that teens used the program as much as they needed to incorporate the skills learned into their daily lives. Mixed methods research is needed to better understand the patterns of teen engagement with DMH programs; how those patterns map onto teens' actual use of the skills acquired in real life; and, ultimately, which patterns are most helpful in mental health management.

Teens shared their feedback on additional features that they would like to see in a DMH program for anxiety. It was important to teens that the aesthetics and function of the DMH program had the characteristics of clean design and easy navigation. Teens also stressed the importance of protecting their privacy, making the program accessible (eg, free and aligned with caregiver and school restrictions on devices), and designing discreet branding, all of which are consistent with several previous reports describing teens' or therapists' design recommendations regarding a teen DMH program [63,64]. In addition, teens reflected on the characteristics that increased their engagement with a digital program, which included gamification, videos and animations, and buy-in from teen celebrities or social media influencers. Other pediatric fields have found partnering with teen social media influencers to be a helpful strategy in enhancing knowledge and even outcomes of health care delivery [65]. In addition to the other features mentioned, partnering with popular teen social media influencers who are known for speaking about their mental health on social media outlets may be a promising strategy for increasing awareness and potentially intervention effectiveness of teen DMH programs. However, as teen culture and preferences rapidly and constantly shift (particularly regarding social media), DMH developers must balance ensuring relevance and timeliness of interventions with ensuring that the content still addresses theoretical mechanisms of change [66].

As for the additional feature of personalization, HURE teens underscored that the available teen mental health resources were White centric and they did not feel that *their* lived experiences were represented in current mental health resources. This suggests that we as scientists and designers need to intentionally improve the inclusion of populations who are underrepresented in teen mental health research, in particular HURE and LGBTQIA+ teens as well as those from lower-resourced and disinvested areas [67-76]. Teens advised that developers get to know the audience for the program and work with that population to personalize the program to their needs and preferences. This feedback raises an important design question: how do we equitably design DMH programs in community spaces that serve people with different identities as well as with different mental health needs and preferences? There are calls in the literature to move away from a *one-size-fits-all* model when developing DMH services [77-80]; however, more discussion and research are needed to better understand the best practices of DMH service development that is intended to meet the mental health needs and preferences of heterogeneous audiences (eg, a program encompassing tailored versions for subgroups within the heterogenous audience or different programs for different subpopulations) [81].

Limitations

We would be remiss to not include a discussion on study limitations. The public library we worked with was well resourced with existing social services (eg, employed social workers, had a mental health and teen service budget, and offered free mental health assessments). Accordingly, the results of this study may not be generalizable to other public libraries that have different infrastructures and resources. More research is needed to better understand the teen resources offered by different public libraries and their branches across the United States. Second, we interviewed teen participants residing in a large, socially and politically progressive, and racially and ethnically diverse urban area. HURE teens living in areas with different sociopolitical environments and sociodemographic compositions may have different perspectives on whether and how they would like to engage in library-based DMH services. Third, we did not screen for anxiety before the interviews; however, it was apparent from teen interview responses that there was high variability in severity levels (level of interference) and treatment experience (some talked about their experiences with therapists, and some disclosed no experience with formal treatment). Finally, while the racial and ethnic makeup of the coders was diverse in representation, it is possible that our views as researchers led us to de-emphasize, overlook, or misinterpret aspects of the qualitative data, especially on topics such as equity, racism, and racial justice. The researchers took measures to reduce bias by engaging in researcher consensus throughout data analysis and member checking with teen and library staff from the partnering local library, some of whom are coauthors who reviewed and edited this report. We also included coauthors who are experts in the field of creating and testing DMH programs for Black, Indigenous, and other racial and ethnic minority populations.

Conclusions

Disseminating and implementing DMH programs in public libraries is a promising strategy to increase access to care among teens in need of support for their anxiety. By understanding teens' perspectives and designing tools with their needs centered, the promise of this approach may be realized. To the best of our knowledge, this is the first report elevating teen voices and opinions on implementing DMH programs in public libraries.

XSL•FO
RenderX

Future research is now needed to test whether DMH programs adapted in accordance with the suggestions presented in this paper are acceptable, useful, and sustainable.

Acknowledgments

This research was supported by the National Institute of Mental Health (K01 MH121854, R34MH128410, K01MH131898, NIMH K23MH128535, P50MH119029, K01MH125172k, and P50MH119029).

Data Availability

Due to the sensitive nature of the qualitative data with minors and the potential risk of identifying variables within the individual transcripts that would potentially impact the privacy of the sample, the individual interviews are not publicly available, nor will requests for data sharing be granted. In addition, public sharing of the individual interviews was not included in the teen assent and parent consent forms.

Conflicts of Interest

DCM has accepted honoraria and consulting fees from Apple, Inc; Otsuka Pharmaceutical; Pear Therapeutics; and the One Mind Foundation, as well as royalties from Oxford University Press, and has an ownership interest in Adaptive Health, Inc. RS and LJ are employed by Oak Park Public Library. KC has received consulting fees from Kooth LLC and Koko. KC is a part-time employee at Lyra Health, Inc. RK has accepted consulting fees and honoraria from CHESS Health and Jimini Health. ABP has served as a consultant to Otsuka on developing clinical educational materials. None of the other authors have conflicts to declare.

Multimedia Appendix 1
Interview script.
[DOCX File , 15 KB - formative_v9i1e67454_app1.docx]

Multimedia Appendix 2
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[PDF File (Adobe PDF File), 422 KB - formative_v9i1e67454_app2.pdf]

References

1. Gruber J, Prinstein MJ, Clark LA, Rottenberg J, Abramowitz JS, Albano AM, et al. Mental health and clinical psychological science in the time of COVID-19: challenges, opportunities, and a call to action. Am Psychol 2021 Apr;76(3):409-426 [FREE Full text] [doi: 10.1037/amp0000707] [Medline: 32772538]
2. AAP-AACAP-CHA declaration of a national emergency in child and adolescent mental health. American Academy of Pediatrics. URL: https://www.aap.org/en/advocacy/child-and-adolescent-healthy-mental-development/aap-aacap-cha-declaration-of-a-national-emergency-in-child-and-adolescent-mental-health/?srsltid=AfmBOopEGS8BeQBnlkoPjbVrYJyDife8fPC3b5pf9h74RcpB5y_Qdxfr [accessed 2024-10-10]
3. Social media and youth mental health: the U.S. surgeon general's advisory. U.S. Department of Health and Human Services. 2023. URL: https://www.hhs.gov/surgeongeneral/reports-and-publications/youth-mental-health/social-media/index.html [accessed 2024-10-10]
4. Youth risk behavior survey data summary and trends report: 2009-2019. Center for Disease Control and Prevention. 2020 Oct 21. URL: https://www.cdc.gov/yrbs/dstr/pdf/YRBSDataSummaryTrendsReport2019-508.pdf [accessed 2024-10-10]
5. Castellanos-Ryan N, Brière FN, O'Leary-Barrett M, Banaschewski T, Bokde A, Bromberg U, et al. The structure of psychopathology in adolescence and its common personality and cognitive correlates. J Abnorm Psychol 2016 Nov;125(8):1039-1052 [FREE Full text] [doi: 10.1037/abn0000193] [Medline: 27819466]
6. Lebrun-Harris LA, Ghandour RM, Kogan MD, Warren MD. Five-year trends in US children's health and well-being, 2016-2020. JAMA Pediatr 2022 Jul 01;176(7):e220056 [FREE Full text] [doi: 10.1001/jamapediatrics.2022.0056] [Medline: 35285883]
7. Fortuna LR, Brown IC, Lewis Woods GG, Porche MV. The impact of COVID-19 on anxiety disorders in youth: coping with stress, worry, and recovering from a pandemic. Child Adolesc Psychiatr Clin N Am 2023 Jul;32(3):531-542 [FREE Full text] [doi: 10.1016/j.chc.2023.02.002] [Medline: 37201965]
8. Uhlhaas PJ, Davey CG, Mehta UM, Shah J, Torous J, Allen NB, et al. Towards a youth mental health paradigm: a perspective and roadmap. Mol Psychiatry 2023 Aug;28(8):3171-3181 [FREE Full text] [doi: 10.1038/s41380-023-02202-z] [Medline: 37580524]
9. National projections of supply and demand for selected behavioral health practitioners: 2013-2025. National Center for Health Workforce Analysis. 2016. URL: https://tinyurl.com/ycyp2f8e [accessed 2024-10-10]

XSL•FO
RenderX

10. Shortages in school psychology: challenges to meeting the growing needs of U.S. students and schools. National Association of School Psychologists. URL: https://ewscripps.brightspotcdn.com/ab/02/d7d25fd547f9b94feabaf81118dd/school-psychology-shortage-2017.pdf [accessed 2024-10-10]
11. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113. [doi: 10.1186/1471-244x-10-113]
12. Benningfield MM. Meeting youth where they are: substance use disorder treatment in schools. Child Adolesc Psychiatr Clin N Am 2016 Oct;25(4):661-668. [doi: 10.1016/j.chc.2016.05.003] [Medline: 27613344]
13. Hardy BJ, Lesperance A, Foote I, Native Youth Sexual Health Network (NYSHN), Firestone M, Smylie J. Meeting indigenous youth where they are at: knowing and doing with 2SLGBTTQQIA and gender non-conforming Indigenous youth: a qualitative case study. BMC Public Health 2020 Dec 07;20(1):1871 [FREE Full text] [doi: 10.1186/s12889-020-09863-3] [Medline: 33287787]
14. Kazdin AE. Addressing the treatment gap: a key challenge for extending evidence-based psychosocial interventions. Behav Res Ther 2017 Jan;88:7-18. [doi: 10.1016/j.brat.2016.06.004] [Medline: 28110678]
15. Knapp AA, Hersch E, Wijaya C, Herrera MA, Kruzan KP, Carroll AJ, et al. "The library is so much more than books": considerations for the design and implementation of teen digital mental health services in public libraries. Front Digit Health 2023 Jul 25;5:1183319 [FREE Full text] [doi: 10.3389/fdgth.2023.1183319] [Medline: 37560198]
16. Kosmicki LR. Equal access as social equity: a literature review of programming and services in public libraries. Progress Libr 2019(47):51-80 [FREE Full text]
17. Elmborg JK. Libraries as the spaces between us recognizing and valuing the third space. Ref User Serv Q 2011 Jun;50(4):338-350. [doi: 10.5860/rusq.50n4.338]
18. Morgan AU, Dupuis R, D'Alonzo B, Johnson A, Graves A, Brooks KL, et al. Beyond books: public libraries as partners for population health. Health Aff (Millwood) 2016 Nov 01;35(11):2030-2036. [doi: 10.1377/hlthaff.2016.0724] [Medline: 27834243]
19. Li X, Ha Y, Aristeguieta S. Teens' vision of an ideal library space: insights from a small rural public library in the United States. Evid Based Libr Inf Pract 2023 Dec 15;18(4):52-67. [doi: 10.18438/eblip30410]
20. Ornstein E, Reid PH. 'Talk to them like they're people': a cross-cultural comparison of teen-centred approaches in public library services. J Librariansh Inf Sci 2021 Jun 12;54(3):451-468. [doi: 10.1177/09610006211020090]
21. Agosto DE, Bell JP, Bernier A, Kuhlmann M. "This is our library, and it's a pretty cool place": a user-centered study of public library YA spaces. Public Libr Q 2015 Mar 03;34(1):23-43. [doi: 10.1080/01616846.2015.1000777]
22. Barone M, Fifelski S, Stocker A, Worm T, Patterson A, Braamse J, et al. Developing a model policy for public libraries to address homelessness and mental illness. Dela J Public Health 2020 Sep 28;6(4):46-51 [FREE Full text] [doi: 10.32481/djph.2020.09.014] [Medline: 34467152]
23. Simon MA, O'Brian CA, Nava M, Dahdouh R, Wafford QE, Mack S, et al. Public libraries as key partners for advancing health equity. Am J Public Health 2021 Jan;111(1):40-42. [doi: 10.2105/AJPH.2020.306004] [Medline: 33326277]
24. Pressley T. Public libraries, serious mental illness, and homelessness: a survey of public librarians. Public Libr Q 2017 Mar 02;36(1):61-76. [doi: 10.1080/01616846.2017.1275772]
25. Hall K, McAlister S. Library services and resources in support of mental health: a survey of initiatives in public and academic libraries. J Libr Adm 2021 Sep 30;61(8):936-946. [doi: 10.1080/01930826.2021.1984137]
26. Wahler EA, Provence MA, Johnson SC, Helling J, Williams M. Library patrons' psychosocial needs: perceptions of need and reasons for accessing social work services. Soc Work 2021 Oct 02;66(4):297-305. [doi: 10.1093/sw/swab032] [Medline: 34333662]
27. Libraries as safer spaces. American Library Association. URL: https://www.ala.org/aboutala/libraries-safer-spaces [accessed 2024-10-10]
28. Stringer H. Libraries as mental health hubs. American Psychological Association. 2020 Apr 01. URL: https://www.apa.org/monitor/2020/04/libraries-health-hubs [accessed 2023-02-15]
29. Pretorius C, Chambers D, Cowan B, Coyle D. Young people seeking help online for mental health: cross-sectional survey study. JMIR Ment Health 2019 Aug 26;6(8):e13524 [FREE Full text] [doi: 10.2196/13524] [Medline: 31452519]
30. Cohen KA, Schleider JL. Adolescent dropout from brief digital mental health interventions within and beyond randomized trials. Internet Interv 2022 Jan 17;27:100496 [FREE Full text] [doi: 10.1016/j.invent.2022.100496] [Medline: 35257001]
31. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
32. Lyon AR, Dopp AR, Brewer SK, Kientz JA, Munson SA. Designing the future of children's mental health services. Adm Policy Ment Health 2020 Sep;47(5):735-751 [FREE Full text] [doi: 10.1007/s10488-020-01038-x] [Medline: 32253634]
33. Mohr DC, Lyon AR, Lattie EG, Reddy M, Schueller SM. Accelerating digital mental health research from early design and creation to successful implementation and sustainment. J Med Internet Res 2017 May 10;19(5):e153 [FREE Full text] [doi: 10.2196/jmir.7725] [Medline: 28490417]
34. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. J R Soc Med 2011 Dec;104(12):510-520 [FREE Full text] [doi: 10.1258/jrsm.2011.110180] [Medline: 22179294]

XSL•FO
RenderX

35. Israel BA, Eng E, Schulz AJ, Parker EA. Introduction to methods in community-based participatory research for health. In: Methods in Community-Based Participatory Research for Health. Hoboken, NJ: John Wiley & Sons; 2005.
36. Bauer MS, Damschroder L, Hagedorn H, Smith J, Kilbourne AM. An introduction to implementation science for the non-specialist. BMC Psychol 2015 Sep 16;3(1):32 [FREE Full text] [doi: 10.1186/s40359-015-0089-9] [Medline: 26376626]
37. Crawley SA, Caporino NE, Birmaher B, Ginsburg G, Piacentini J, Albano AM, et al. Somatic complaints in anxious youth. Child Psychiatry Hum Dev 2014 Aug;45(4):398-407 [FREE Full text] [doi: 10.1007/s10578-013-0410-x] [Medline: 24129543]
38. Weems CF, Silverman WK, La Greca AM. What do youth referred for anxiety problems worry about? Worry and its relation to anxiety and anxiety disorders in children and adolescents. J Abnorm Child Psychol 2000 Feb;28(1):63-72. [doi: 10.1023/a:1005122101885] [Medline: 10772350]
39. Dedoose (Version 9.0.17), cloud application for managing, analyzing, and presenting qualitative and mixed method research data. SocioCultural Research Consultants. 2021. URL: https://www.dedoose.com/ [accessed 2024-10-10]
40. Doyle S. Member checking with older women: a framework for negotiating meaning. Health Care Women Int 2007;28(10):888-908. [doi: 10.1080/07399330701615325] [Medline: 17987459]
41. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
42. Russell CK, Gregory DM. Evaluation of qualitative research studies. Evid Based Nurs 2003 Apr;6(2):36-40 [FREE Full text] [doi: 10.1136/ebn.6.2.36] [Medline: 12710415]
43. Mays N, Pope C. Assessing quality in qualitative research. BMJ 2000 Jan 01;320(7226):50-52 [FREE Full text] [doi: 10.1136/bmj.320.7226.50] [Medline: 10617534]
44. Zvolensky MJ, Eifert GH, Lejuez C, Hopko DR, Forsyth JP. Assessing the perceived predictability of anxiety-related events: a report on the perceived predictability index. J Behav Ther Exp Psychiatry 2000;31(3-4):201-218. [doi: 10.1016/s0005-7916(01)00006-4] [Medline: 11494957]
45. Lejuez CW, Eifert GH, Zvolensky MJ, Richards JB. Preference between onset predictable and unpredictable administrations of 20% carbon-dioxide-enriched air: implications for better understanding the etiology and treatment of panic disorder. J Exp Psychol Appl 2000 Dec;6(4):349-358. [doi: 10.1037//1076-898x.6.4.349] [Medline: 11218343]
46. Zvolensky MJ, Lejuez CW, Eifert GH. Prediction and control: operational definitions for the experimental analysis of anxiety. Behav Res Ther 2000 Jul;38(7):653-663. [doi: 10.1016/s0005-7967(99)00090-x] [Medline: 10875188]
47. Gregor KL, Zvolensky MJ. Anxiety sensitivity and perceived control over anxiety-related events: evaluating the singular and interactive effects in the prediction of anxious and fearful responding to bodily sensations. Behav Res Ther 2008 Sep;46(9):1017-1025. [doi: 10.1016/j.brat.2008.06.003] [Medline: 18675954]
48. Ream GL. Minority stress and intersectionality in LGBTQIA+ youth mental health disparities. Am J Public Health 2024 Jul;114(7):669-671. [doi: 10.2105/AJPH.2024.307676] [Medline: 38662974]
49. Fish JN. Future directions in understanding and addressing mental health among LGBTQ youth. J Clin Child Adolesc Psychol 2020;49(6):943-956 [FREE Full text] [doi: 10.1080/15374416.2020.1815207] [Medline: 33074740]
50. Fish JN, Kapostasy S, Russell ST. Be YOU!: a collaborative effort to address minority stress for LGBTQ+ youth in school settings. J Gay Lesbian Soc Serv 2023;35(4):434-455 [FREE Full text] [doi: 10.1080/10538720.2022.2148036] [Medline: 38322581]
51. Diamond LM, Alley J. Rethinking minority stress: a social safety perspective on the health effects of stigma in sexually-diverse and gender-diverse populations. Neurosci Biobehav Rev 2022 Jul;138:104720. [doi: 10.1016/j.neubiorev.2022.104720] [Medline: 35662651]
52. Mahon NE, Yarcheski A, Yarcheski TJ, Hanks MM. A meta-analytic study of predictors of anger in adolescents. Nurs Res 2010;59(3):178-184. [doi: 10.1097/NNR.0b013e3181dbba04] [Medline: 20404774]
53. Walsh LM, Wolk CB, Haimes EM, Jensen-Doss A, Beidas RS. The relationship between anger and anxiety symptoms in youth with anxiety disorders. J Child Adolesc Couns 2018;4(2):117-133 [FREE Full text] [doi: 10.1080/23727810.2017.1381930] [Medline: 30828611]
54. Osman M, Williams MT, Alftieh A. State of the science on racial microaggressions. Behav Ther 2024 Nov;55(6):1172-1188. [doi: 10.1016/j.beth.2024.05.005] [Medline: 39443060]
55. Adkins-Jackson PB, Joseph VA, Ford TN, Avila-Rieger JF, Gobaud AN, Keyes K. State-level structural racism and adolescent mental health in the United States. Am J Epidemiol 2024 Jul 03:kwae164 (forthcoming). [doi: 10.1093/aje/kwae164] [Medline: 38960643]
56. Oshri A, Reck AJ, Carter SE, Uddin LQ, Geier CF, Beach SR, et al. Racial discrimination and risk for internalizing and externalizing symptoms among Black youths. JAMA Netw Open 2024 Jun 03;7(6):e2416491 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.16491] [Medline: 38865126]
57. Shen J, Rubin A, Cohen K, Hart EA, Sung J, McDanal R, et al. Randomized evaluation of an online single-session intervention for minority stress in LGBTQ+ adolescents. Internet Interv 2023 Jun 07;33:100633 [FREE Full text] [doi: 10.1016/j.invent.2023.100633] [Medline: 37635950]

XSL•FO
RenderX

58. Heard-Garris NJ, Hope E. Our voices matter: racial justice activism intervention to address structural racism and prevent depression in Black and Latinx youth. Lurie Children's Hospital of Chicago. URL: https://www.paperdigest.org/grant/?grant_id=nih_10879839 [accessed 2024-10-10]
59. Roquet CD, Theofanopoulou N, Freeman JL, Schleider J, Gross JJ, Davis K, et al. Exploring situated and embodied support for youth's mental health: design opportunities for interactive tangible device. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI '22; April 29-May 5, 2022; New Orleans, LA. [doi: 10.1145/3491102.3502135]
60. MacLean KE. Designing affective haptic experience for wellness and social communication: where designers need affective neuroscience and psychology. Curr Opin Behav Sci 2022 Jun;45:101113. [doi: 10.1016/j.cobeha.2022.101113]
61. Agapie E, Chang K, Patrachari S, Neary M, Schueller SM. Understanding mental health apps for youth: focus group study with Latinx youth. JMIR Form Res 2022 Oct 18;6(10):e40726 [FREE Full text] [doi: 10.2196/40726] [Medline: 36256835]
62. Liverpool S, Mota CP, Sales CM, Čuš A, Carletto S, Hancheva C, et al. Engaging children and young people in digital mental health interventions: systematic review of modes of delivery, facilitators, and barriers. J Med Internet Res 2020 Jun 23;22(6):e16317 [FREE Full text] [doi: 10.2196/16317] [Medline: 32442160]
63. Matthews M, Doherty G. In the mood: engaging teenagers in psychotherapy using mobile phones. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI '11; May 7-12, 2011; Vancouver, BC. [doi: 10.1145/1978942.1979379]
64. Knapp AA, Cohen K, Nicholas J, Mohr DC, Carlo AD, Skerl JJ, et al. Integration of digital tools into community mental health care settings that serve young people: focus group study. JMIR Ment Health 2021 Aug 19;8(8):e27379 [FREE Full text] [doi: 10.2196/27379] [Medline: 34420928]
65. Psihogios AM, Roth M, Gomez C, Hekimian-Brogan E, McQueen C, Yanez B. Partnering with social media influencers to equitably improve adolescent and young adult cancer outcomes: a novel strategy to support cancer care delivery. JCO Oncol Pract 2024 Jun;20(6):755-760. [doi: 10.1200/OP.23.00449] [Medline: 38408282]
66. Fitzpatrick S, Crenshaw AO, Donkin V, Collins A, Xiang A, Earle EA, et al. We have spent time, money, and effort making self-help digital mental health interventions: is anyone going to come to the party? J Med Internet Res 2024 Sep 19;26:e58198 [FREE Full text] [doi: 10.2196/58198] [Medline: 39298760]
67. McCall T, Foster M, Tomlin HR, Schwartz TA. Black American women's attitudes toward seeking mental health services and use of mobile technology to support the management of anxiety. JAMIA Open 2023 Oct 17;6(4):ooad088 [FREE Full text] [doi: 10.1093/jamiaopen/ooad088] [Medline: 37860603]
68. McCall T, Foster M, Schwartz TA. Attitudes toward seeking mental health services and mobile technology to support the management of depression among Black American women: cross-sectional survey study. J Med Internet Res 2023 Jul 19;25:e45766 [FREE Full text] [doi: 10.2196/45766] [Medline: 37467027]
69. Williams KD, Wijaya C, Stamatis CA, Abbott G, Lattie EG. Insights into needs and preferences for mental health support on social media and through mobile apps among Black male university students: exploratory qualitative study. JMIR Form Res 2022 Aug 31;6(8):e38716 [FREE Full text] [doi: 10.2196/38716] [Medline: 36044261]
70. Ramos G, Aguilera A, Montoya A, Lau A, Wen CY, Cruz Torres V, et al. App-based mindfulness meditation for people of color who experience race-related stress: protocol for a randomized controlled trial. JMIR Res Protoc 2022 Apr 14;11(4):e35196 [FREE Full text] [doi: 10.2196/35196] [Medline: 35436228]
71. Aguilera A, Arévalo Avalos MR, Rosales K, Reyes Y, Hernandez-Ramos R, Ramos G, et al. Effectiveness-implementation hybrid trial of Spanish language, digital cognitive-behavioral therapy (dCBT) intervention for depression and anxiety - protocol for the SUPERA (SUpport from PEeRs to expand Access) study. Contemp Clin Trials 2024 Feb;137:107422 [FREE Full text] [doi: 10.1016/j.cct.2023.107422] [Medline: 38145715]
72. Eustis EH, LoPresti J, Aguilera A, Schueller SM. Cultural responsivity in technology-enabled services: integrating culture into technology and service components. J Med Internet Res 2023 Oct 03;25:e45409 [FREE Full text] [doi: 10.2196/45409] [Medline: 37788050]
73. Pederson AB. Management of depression in Black people: effects of cultural issues. Psychiatr Ann 2023 Mar;53(3):122-125 [FREE Full text] [doi: 10.3928/00485713-20230215-01] [Medline: 37396678]
74. Coyne CA, Wongsomboon V, Korpak AK, Macapagal K. "We have to figure it out ourselves": transfeminine adolescents' online sexual experiences and recommendations for supporting their sexual health and wellbeing. Front Reprod Health 2022 Jan 16;4:1034747 [FREE Full text] [doi: 10.3389/frph.2022.1034747] [Medline: 36726593]
75. Jozsa K, Kraus A, Korpak AK, Birnholtz J, Moskowitz DA, Macapagal K. "Safe behind my screen": adolescent sexual minority males' perceptions of safety and trustworthiness on geosocial and social networking apps. Arch Sex Behav 2021 Oct;50(7):2965-2980. [doi: 10.1007/s10508-021-01962-5] [Medline: 34581948]
76. Katapally TR. Smart indigenous youth: the smart platform policy solution for systems integration to address indigenous youth mental health. JMIR Pediatr Parent 2020 Sep 25;3(2):e21155 [FREE Full text] [doi: 10.2196/21155] [Medline: 32975527]
77. Vis C, Schuurmans J, Aouizerate B, Atipei Craggs M, Batterham P, Bührmann L, et al. Effectiveness of self-guided tailored implementation strategies in integrating and embedding internet-based cognitive behavioral therapy in routine mental health

XSL•FO
RenderX

care: results of a multicenter stepped-wedge cluster randomized trial. J Med Internet Res 2023 Feb 03;25:e41532 [FREE Full text] [doi: 10.2196/41532] [Medline: 36735287]
78. McCall T, Threats M, Pillai M, Lakdawala A, Bolton CS3. Recommendations for design of a mobile application to support management of anxiety and depression among Black American women. Front Digit Health 2022 Dec 23;4:1028408 [FREE Full text] [doi: 10.3389/fdgth.2022.1028408] [Medline: 36620185]
79. Ellis DM, Draheim AA, Anderson PL. Culturally adapted digital mental health interventions for ethnic/racial minorities: a systematic review and meta-analysis. J Consult Clin Psychol 2022 Oct;90(10):717-733. [doi: 10.1037/ccp0000759] [Medline: 36227330]
80. Spadaro B, Martin-Key NA, Bahn S. Building the digital mental health ecosystem: opportunities and challenges for mobile health innovators. J Med Internet Res 2021 Oct 13;23(10):e27507 [FREE Full text] [doi: 10.2196/27507] [Medline: 34643537]
81. Lyles CR, Aguilera A, Nguyen O, Sarkar U. Bridging the digital health divide: how designers can create more inclusive digital health tools. California Health Care Foundation. 2022. URL: https://www.chcf.org/wp-content/uploads/2022/02/BridgingDigitalDivideDesigners.pdf [accessed 2024-10-09]

Abbreviations

CBPR: community-based participatory research
COREQ: Consolidated Criteria for Reporting Qualitative Research
DMH: digital mental health
HCD: human-centered design
HURE: historically underserved racial and ethnic minority
IR: implementation research
LGBTQIA+: lesbian, gay, bisexual, transgender, queer, intersex, and asexual

Edited by A Mavragani; submitted 11.10.24; peer-reviewed by E Rubenstein, A Jolliff; comments to author 19.11.24; revised version received 17.12.24; accepted 22.12.24; published 13.03.25.

Please cite as:
Knapp AA, Cohen K, Kruzan KP, Kornfield R, Herrera M, Pederson AB, Lee S, Macapagal K, Roulston CA, Clarke K, Wijaya C, Simmons R, Jackson L, Kour S, Franco S, Mohr DC
Teen Perspectives on Integrating Digital Mental Health Programs for Teens Into Public Libraries ("I Was Always at the Library"): Qualitative Interview Study
JMIR Form Res 2025;9:e67454
URL: https://formative.jmir.org/2025/1/e67454
doi:10.2196/67454
PMID:

©Ashley A Knapp, Katherine Cohen, Kaylee P Kruzan, Rachel Kornfield, Miguel Herrera, Aderonke B Pederson, Sydney Lee, Kathryn Macapagal, Chantelle A Roulston, Kaleigh Clarke, Clarisa Wijaya, Robert Simmons, Latonia Jackson, Simrandeep Kour, Sandra Franco, David C Mohr. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Use of Video Consultation Between 2017 and 2020 in Outpatient Medical Care in Germany and Characteristics of Their User Groups: Analysis of Claims Data

Theresa Hüer[1], Dr rer pol; Anke Walendzik[1], Dr rer medic; Lara Kleinschmidt[1], MSc; Klemens Höfer[1], MA; Beatrice Nauendorf[2], MSc; Juliane Malsch[2], MSc; Matthias Brittner[3], MA; Paul Brandenburg[4], MA; André Aeustergerling[5], MA; Udo Schneider[6], PD Dr rer pol; Anja Wadeck[6], MA; Sebastian Liersch[7], Dr PH; Stephanie Sehlen[7], Dr rer pol; Katharina Schwarze[8], MA; Jürgen Wasem[1], Prof Dr

[1]Institute for Health Care Management and Research, University of Duisburg-Essen, Essen, Germany

[2]Kassenärztliche Vereinigung Berlin, Berlin, Germany

[3]Kassenärztliche Vereinigung Westfalen-Lippe, Dortmund, Germany

[4]Kassenärztliche Vereinigung Schleswig-Holstein, Bad Segeberg, Germany

[5]Kassenärztliche Vereinigung Mecklenburg-Vorpommern, Schwerin, Germany

[6]Techniker Krankenkasse, Hamburg, Germany

[7]AOK Nordost, Potsdam, Germany

[8]AOK NordWest, Dortmund, Germany

Corresponding Author:
Theresa Hüer, Dr rer pol
Institute for Health Care Management and Research
University of Duisburg-Essen
Thea-Leymann-Straße 9
Essen, 45127
Germany
Phone: 49 201 183 2592
Email: theresa.hueer@medman.uni-due.de

Abstract

Background: Supplementing outpatient medical care with the use of video consultations could, among other benefits, improve access, especially in structurally disadvantaged areas.

Objective: This claims data analysis, carried out as part of the German research project "Preference-based use of video consultation in urban and rural regions," aimed to analyze the use of video consultations and the characteristics of its user groups.

Methods: Claims data from 3 Statutory Health Insurance Funds (SHIFs) and 4 Associations of Statutory Health Insurance Physicians (ASHIPs) from the period April 2017 to the end of 2020 were used. Data from a sample of about 6.1 million insured and 33,100 physicians and psychotherapists were analyzed. In addition to data on the use of video consultations, patient data on sociodemographic characteristics, diagnoses, and place of residence were included. To analyze the physicians' perspectives, specialty groups, demographic characteristics, and the type of practice location were also included. In consideration of the principles of data economy and the fact that data analysis represents merely a preliminary phase within the broader project, the SHIFs and ASHIPs transmitted aggregated data (cross-tabulations per subgroup analysis) to the evaluator. For this reason, the analyses were constrained to a comparison of video consultation users versus nonusers, differentiated according to the aforementioned subgroups. Furthermore, the association between place of residence or type of region of the practice location and the use of video consultation was examined. A significance level of $P<.05$ was set for chi-square tests.

Results: From 2017 to 2019, almost no video consultations were used in outpatient care in the German health care system. Although this changed considerably in relative terms with the start of the COVID-19 pandemic (but still at a very low absolute level), there was also a clear decline in the use of video consultations as the number of infections flattened out. Physicians working in psychotherapy and psychological psychotherapists used video consultations with around 16% (44,808/282,530) of their treatment cases in the second quarter of 2020, followed by psychotherapists using video consultations for children (10,828/113,293, 10%). Although the absolute number of treatment cases with video consultations among general practitioners was very high compared

XSL•FO
RenderX

with other specialist groups, their share of video consultations in all treatment cases was very low at 0.3% (29,600/9,837,118). Younger age groups and those located in urban areas used video consultations more frequently; this applies to both patients (age groups: χ^2_7=9903.2, *P*<.001; region types: χ^2_2=3746.2, *P*<.001) and service providers (age groups: χ^2_3=11,338.2, *P*<.001; region types: χ^2_2=8474.1, *P*<.001).

Conclusions: The current use of video consultations is below its potential in terms of scope and user groups. The widespread and lasting use of video consultations will only succeed if the potential user groups accept this form of service provision and recognize its advantages. Further analyses (both qualitative, such as focus group discussions, and quantitative, such as preference surveys) should therefore investigate the preferences of user groups for the use of video consultations.

International Registered Report Identifier (IRRID): RR2-10.2196/50932

(JMIR Form Res 2025;9:e60170) doi:10.2196/60170

KEYWORDS

video consultation; outpatient medical care; user groups; claims data analysis; Germany; physician; psychotherapist; sociodemographic; healthcare; digital health; digital consultation; telehealth; telemonitoring; telemedicine

Introduction

The implementation of video consultations in the German Social Health Insurance (SHI) system in April 2017 was intended to establish a basis for providing outpatient medical services regardless of location and thus also in structurally disadvantaged regions in Germany. Older people and those with limited mobility were to be given easier access to medical services. Video consultations can be used in a variety of ways and therefore offer potential benefits for different patient groups. Particularly in the case of long travel distances to the nearest (specialist) physician, for follow-up appointments, or for the provision of repeated prescriptions, video consultations can be a useful tool, and the insured do not have to visit a doctor's office for every appointment. In the field of psychotherapy, therapeutic or certain diagnostic sessions can also take place in the form of video consultations [1-3].

In the beginning, video consultations in Germany could only be provided by a few specialist groups and only for selected indications and follow-up appointments. In the second quarter of 2019, these restrictions were mostly abolished, and the assessment of the appropriateness of a diagnosis or treatment via video consultation was the responsibility of the respective physician. The remuneration for video consultations in outpatient medical care within the German SHI system is primarily based on age-differentiated quarterly flat rates. These rates cover the entire medical care of a patient, not just video consultations, provided by the attending physician during the specified period. If a patient is treated exclusively via video in a quarter and there are no other personal physician-patient contacts, the quarterly flat rate is reduced by a discount specific to the specialist group. In addition to the quarterly flat rates per patient, there are supplementary fee schedule positions (FSPs) that compensate for the additional technical effort (€4.60 [US $4.77] per video consultation, FSP 01450) and the digital authentication of an unknown patient in video consultations (€1.15 [US $1.19] per video consultation with a previously unknown patient, FSP 01444). From October 2019 to the end of September 2021, there was also an incentive grant (at that time approximately €10 [US $10.36] per video consultation for a maximum of 50 video consultations per physician) to stimulate the use of video consultations [1].

Despite the potential of video consultations for different user groups (patients as well as physicians and psychiatrists), financing by the German SHI system since 2017, and support measures such as incentive payments, video consultations do not yet appear to have established themselves as an integral part of outpatient medical care. In order to investigate the underlying obstacles, it is important to examine the actual use of video consultations on a solid data basis in detail in order to identify suitable groups and occasions for the development of measures to promote video consultations.

The German committee for negotiating outpatient medical remuneration in the SHI system (“Bewertungsausschuss Ärzte”) published a report on the use of video consultations in 2018 and in an updated form in 2022. These reports show the development of the billing frequency of video consultations in outpatient medical care within the SHI system for the data period April 2017 to the end of 2021, differentiated according to the specialist groups as well as the age groups and gender of the insured [1,4]. In 2022, Gensorowsky et al [5] also published an article on the use of video consultations in Germany based on data from a nationwide health insurance fund, which also covered the billing frequency of video consultations for the period April 2017, but only up to the end of 2018. The analysis was differentiated according to the specialist groups of the physicians and the age groups and type of region in which insured persons who used video consultations live [5]. However, Gensorowsky et al [5] only examined the period up to the end of 2018 and were therefore unable to account for the effects of the COVID-19 pandemic. A more recent 2023 study by Petrick and Kreuzenbeck [6] considered data through mid-2020 but was based on data from only one Social Health Insurance Fund (SHIF) and therefore did not include information from the perspective of physicians as well as psychotherapists. Thus, recent studies are needed to investigate the characteristics of insured persons using the service. Additionally, it would be interesting to explore the specialist groups that offer video consultations. It is not possible to analyze the characteristics of physicians and psychotherapists (eg, provision of video consultations by age group, gender, region of practice location)

XSL•FO
RenderX

using data from SHIFs. This requires data from the Associations of Statutory Health Insurance Physicians (ASHIPs).

Against this background, this article dealt in particular with closing the aforementioned gaps in the evidence by analyzing not only claims data from SHIFs but also data from ASHIPs. The data period April 2017 to the end of 2020 was included in order to take the effects of the COVID-19 pandemic into consideration, a relevant extension of the observation period compared with the paper by Gensorowsky et al [5], which ended in 2018. In addition to the analysis of the frequency of video consultations, detailed characteristics of their user groups (users and insured persons and providers and physicians and psychotherapists) will also be examined. The age and gender distributions, employment status, and residence or type of region will be compared between users and nonusers of video consultations. In the case of service providers, specialty groups as well as their age and gender distributions, type of operation and employment, and the type of region of the practice location were considered. In particular, a detailed analysis of the association between place of residence or type of region of the practice location and the use of video consultations represents a significant contribution to the literature.

This claims data analysis was conducted as part of the German research project "Preference-based use of video consultations in urban and rural regions," which is funded by the Innovation Fund of the German Federal Joint Committee (funding number 01VSF20011). The research project addresses the question of how video consultations should be designed to enhance the quality of outpatient medical care while considering the preferences of insured individuals, physicians, and psychotherapists. A study protocol has been published elsewhere [7].

The objective of this data analysis was to examine the use of video consultations in Germany, with a particular emphasis on the impact of the COVID-19 pandemic, and to discern salient disparities between patient and physician subgroups. Our objective was to ascertain the prevalence of dependencies according to the following sociodemographic characteristics of insured persons: age group, gender, employment status, and place of residence. In addition, we wanted to investigate how video consultations were integrated into the overall treatment of a patient during the quarter; was this done exclusively via video, or was the video consultation only used as a supplement? For physicians and psychotherapists, the aim was to identify dependencies between video consultation use and the following variables: age group, gender, specialist group, practice type, and practice location.

Methods

Data

In this analysis, anonymized, aggregated data from 3 SHIFs (Techniker Krankenkasse, AOK Nordost, AOK NordWest) as well as 4 ASHIPs were used to reflect both the perspective of the insured or patients and that of the physicians and psychotherapists. Data from April 2017 to the end of 2020 from insured and physicians as well as psychotherapists who lived or practiced in 1 of 4 (of a total of 17) German regions ("Westphalia-Lippe," "Mecklenburg-Western Pomerania," "Schleswig-Holstein," and "Berlin") were analyzed. These regions were selected because they represent both rural and urban districts. They were classified as rural or urban districts using the interactive web application "INKAR - Indicators and Maps of Spatial and Urban Development" from the German Federal Institute for Research on Building, Urban Affairs and Spatial Development. Among other things, the settlement structure of the district is considered a combination of population density and the proportion of settlement area.

Aggregated data from about 6.1 million insured living in the 4 selected regions were analyzed. This corresponds to a share of approximately 8.4% of all SHI-insured persons in Germany (annual average in 2020=73,274,131) [8]. In addition to data on the use of video consultations, patient data on sociodemographic characteristics and diagnostic information were included. Of the 6.1 million insured persons included in the analysis, approximately 4.6 million (approximately 74%) had at least one outpatient treatment case on average across all included quarters ("users of outpatient services"). The group of outpatient service users is an important group in the analysis, as only they were faced with the actual decision of whether to contact a physician in person or via video. Table 1 summarizes the details of the SHIF data used and the characteristics of the insured who were included.

Table 2 shows the key information from the ASHIP data and the characteristics of the approximate 33,000 physicians and psychotherapists included. This corresponds to 18.4% (33,000/180,581 in 2020) of all SHI-accredited physicians and psychotherapists in Germany [9].

XSL•FO
RenderX

Table 1. Social Health Insurance Fund (SHIF) data.

Data type	Data
Source	3 SHIFs: Techniker Krankenkasse, AOK Nordost, AOK NordWest
Regions	4 of 17 regions in Germany: Berlin, Schleswig-Holstein, Mecklenburg-Western Pomerania, Westphalia-Lippe
Time period	Q2[a] 2017 until the end of 2020
Exclusion criteria	1. Insured living abroad, 2. Insured in employee health insurance in the respective SHIF, 3. Insured who refused the use of their data for research purposes
Insured persons (Q2 2017-Q4[b] 2020)	
Minimum (Q4 2017), n	5,915,199
Maximum (Q2 2020), n	6,318,851
Average	6,131,573
Insured persons with at least one treatment case (Q2 2017-Q4 2020), n (%)	
Minimum (Q2 2020)	4,277,149 (67.7)
Maximum (Q1[c] 2020)	4,761,242 (76.7)
Average	4,550,963 (74.2)
Gender (Q2 2020; n=4,277,148), n (%)	
Female (including diverse)	2,369,554 (55.4)
Male	1,907,594 (44.6)
Age (years; Q2 2020; n=4,277,148), n (%)	
≤20	742,967 (17.4)
21-30	447,669 (10.5)
31-40	511,062 (11.9)
41-50	478,659 (11.2)
51-60	717,920 (16.8)
61-70	576,561 (13.5)
71-80	429,612 (10)
≥81	372,698 (8.7)
Employment status (Q2 2020; n=4,229,406), n (%)	
Employed persons (employees/workers, self-employed persons)	2,047,604 (48.4)
Not employed persons (unemployed, social welfare recipients, emigrants, refugees, students)	967,533 (22.9)
Retired persons	1,214,269 (28.7)
Region type of the place of residence (Q2 2020; n=4,277,139), n (%)	
Urban regions	2,295,715 (53.7)
Mixed regions	1,379,654 (32.2)
Rural regions	601,770 (14.1)

[a]Q2: second quarter.

[b]Q4: fourth quarter.

[c]Q1: first quarter.

XSL•FO
RenderX

Table 2. Association of Statutory Health Insurance Physicians (ASHIP) data.

Data type	Data
Source	All SHI[a] physicians and psychotherapists in 4 ASHIP regions: Berlin, Schleswig-Holstein, Mecklenburg-Western Pomerania, Westphalia-Lippe
Regions	4 of 17 regions in Germany: Berlin, Schleswig-Holstein, Mecklenburg-Western Pomerania, Westphalia-Lippe
Time period	Q2[b] 2017 until the end of 2020
Exclusion criteria	Specialist groups that are not allowed to provide video consultations (eg, laboratory physicians)
Physicians (Q2 2017-Q4[c] 2020)	
Minimum (Q2 2017), n	32,128
Maximum (Q4 2020), n	34,127
Q2 2020, n	33,758
Average	33,231
Gender (Q2 2020; n=33,758), n (%)	
Female (including diverse)	14,930 (44.2)
Male	18,828 (55.8)
Age (years; Q2 2020; n=33,758), n (%)	
≤40	2516 (7.5)
41-50	7940 (23.5)
51-60	12,840 (38)
≥61	10,462 (31)
Aggregated specialist groups (Q2 2020; n=33,765), n (%)	
Primary care	11,988 (35.5)
Psychotherapy, psychiatry, neurology	8663 (25.7)
Gynecology	2492 (7.4)
Orthopedics	1459 (4.3)
Surgery	1391 (4.1)
Dermatology	791 (2.3)
Otorhinolaryngology and phoniatrics	947 (2.8)
Other specialist groups	6034 (17.9)
Type of practice (Q2 2020; n=33,782[d]), n (%)	
Individual practice	17,904 (53)
Joint practices (“Berufsausübungsgemeinschaft” and “Praxisgemeinschaft“)	10,738 (31.8)
Medical care center (“Medizinisches Versorgungszentrum [MVZ]“)	3585 (10.6)
Hospital physicians authorized for outpatient care	1555 (4.6)
Region type of the location of the physician’s office (Q2 2020; n=33,757), n (%)	
Urban regions	19,128 (56.7)
Mixed regions	10,086 (29.9)
Rural regions	4543 (13.5)

[a]SHI: Social Health Insurance.

[b]Q2: second quarter.

[c]Q4: fourth quarter.

[d]The sample size is higher than that of Q2 2020 overall, because physicians may be assigned to multiple practice types.

XSL•FO
RenderX

Finally, it is important to emphasize that, depending on which data set is used for analysis (SHIF data or ASHIP data), either the entire outpatient medical care of all SHI-insured persons with residence in the 4 regions is taken into account (ASHIP data) or only the medical care of those SHI-insured persons who are insured with 1 of the 3 SHIFs and live in the 4 regions (SHIF data) is taken into account.

Data Analysis

The German Federal Data Protection Act defines the principle of data avoidance and data minimization. Accordingly, the aim of the evaluation must always be taken into account when selecting and designing data processing systems. In light of this, a 2-step evaluation was carried out in this analysis: in the first step, a preliminary analysis of the individual data was carried out by the SHIFs and ASHIPs, and, in the second step, the aggregated data were merged and finally evaluated. This was to ensure that the evaluating body (University of Duisburg-Essen) only received the information necessary to answer the research questions.

Step 1: Preparatory Work and Preliminary Analysis of Personal Data at the SHIFs

In preparation for the data analysis, an evaluation concept and a data protection concept were first developed. Data forwarding agreements were also concluded bilaterally between the evaluating body and the 7 data providers for the forwarding of the aggregated data.

The developed evaluation concept comprised tables to be completed by all data providers on the use of video consultations, differentiated according to various characteristics of the insured (age, gender, employment status, residence or type of region) and of the physicians and psychotherapists (age, gender, specialist group, type of operation and employment, type of region of the practice location). This information was collected for users of video consultations as well as nonusers of video consultations. The frequency of video consultation use can be identified using the FSP 01450 (technical surcharge for video consultations), as this can be billed for every video contact. A limitation in the use of this FSP is the fact that there is a maximum billing quota for this FSP. In this respect, the use of this FSP requires the assumption that physicians will continue to use FSP 01450 even if they have already exhausted their billing quota per physician and quarter.

Due to the aforementioned (partial) flat-rate billing in outpatient medical care in Germany, which also includes video consultations, an evaluation of the specific services performed in video consultations is only possible to a limited extent. However, it is possible to evaluate certain individual services that are provided during video consultations that can be billed separately. This is possible because the fee schedule items are provided with a standardized national label if they are performed by video. A list of these services has been published by the National Association of Statutory Health Insurance Physicians [10].

Based on this, the preliminary analyses of the individual data were carried out by the SHIFs and ASHIPs.

On the basis of the data forwarding agreements, these were transmitted securely by each data supplier to the evaluating body.

Step 2: Merging and Final Evaluation of the Anonymized and Aggregated Data at the Evaluating Body

The aggregated data for each data supplier were initially merged by the evaluating body so that all further evaluations were based on the data from all SHIFs or all ASHIPs. Since the data were not provided at the individual level, the analysis of the aggregated data (only available as cross tables) was performed using Excel (version 16.0; Microsoft Corp).

The analyses included a comparison of video consultation users (VC users) versus nonusers, differentiated according to subgroups (insured: age group, gender, employment situation; physicians and psychotherapists: specialist group, age group, gender, type of operation, and employment). Furthermore, the association between the place of residence or type of region of the practice location and the use of video consultation was examined. To identify significant differences between subgroups, a significance level of $P<.05$ was set. Since the analysis was based on aggregated data (in the form of cross-tabulations, nominally scaled data), chi-square tests were used. Analyses based on individual and microdata (eg, regression analyses) were therefore excluded. However, this approach was chosen for reasons of data economy and because the data analysis was only a preparatory step in the overall project.

Ethical Considerations

The study was granted ethical approval by the Ethics Committee of the Medical Faculty of the University of Duisburg-Essen on September 27, 2022 (reference: 21-10283-BO). As the University of Duisburg-Essen only had access to aggregated data, there was no possibility of participant identification.

Results

Development of the Use of Video Consultations Over Time

From 2017 to 2019, video consultations played almost no role in outpatient medical care in Germany (quarters 2 and 3 [Q2 and Q3] 2017: 0; Q4 2017-Q3 2019: <300 video consultations; Q4 2019: 718 video consultations). Since the start of the COVID-19 pandemic, the use of video consultations has increased substantially. The billing frequency of FSP 01450 showed a huge increase in the first quarter of 2020 in all regions considered; this continued in the second quarter of 2020. This means that the number of video consultations carried out within 1 quarter increased approximately 6-fold from 44,680 (Q1 2020) to 271,483 (Q2 2020). The second quarter of 2020 covers the largest period of the first lockdown during the COVID-19 pandemic. In the summer months (Q3 2020), the use of video consultations decreased (109,683, which corresponds to only 40.4% [109,683/271,483] of the amount in Q2 2020), although the number of video consultations carried out then rose sharply again in the fourth quarter (192,405, which corresponds to

XSL•FO
RenderX

70.9% [192,405/271,483] of the amount in Q2 2020). This clearly shows that the use of video consultations not only increased sharply during the COVID-19 pandemic but also largely reflected the development of COVID-19 infections. In the period from April 2017 to the end of 2020, a total of 620,970 video consultations were held in the relevant regions in Germany.

When looking at the absolute figures, however, it should be noted that physician contacts with video consultations still only accounted for a fraction of all contacts of this type. If we look at the SHIF data and examine which of the insured persons included had at least one physician contact ("users"; Q2 2020: 4,277,149/6,318,851, 67.7%), the proportion of those who had a video consultation was extremely low, with a maximum of 1% (43,139/4,227,149) in the second quarter of 2020, 0.3% (13,543/4,227,149) in the first quarter of 2020, 0.4% (19,989/4,227,149) in the third quarter of 2020, and 0.8% (34,557/4,227,149) in the fourth quarter of 2020.

By examining the number of video consultations used per VC user, we see that, in the second quarter of 2020, 50% of VC users (21,569/43,139) had a maximum of 1 video consultation. The mean value per VC user was significantly higher, at 2.5 video consultations, which is due to the fact that some VC users made use of a large number of video consultations; the maximum here was 38 video consultations in the second quarter of 2020.

Characteristics of VC User Groups (Patients)

Overview of Analysis of VC User Groups

In the following sections, the results of the comparison between VC users and nonusers in terms of sociodemographic information are presented in order to investigate whether they differed systematically from each other. In addition, the association between the patient's place of residence and the use of video consultations was also analyzed. The underlying population for all analyses was the insured who had at least one outpatient medical contact in the second quarter of 2020, because only they were faced with the decision of a video consultation or face-to-face contact. The second quarter of 2020 was selected for analysis because, as previously described, the number of video consultations was highest during this period, allowing for subgroup analyses despite the absolute small number of VC users.

Differences by Age Group

Table 3 shows, based on the example of the second quarter of 2020, that video consultations were mainly used by young to middle-aged insured persons (20-50 years old). The largest group was those aged 31 years to 40 years, with a share of 1.7% (9378/563,715) of all insured in this age group, with at least one outpatient contact with a doctor (users). The correlation was statistically significant (χ^2_7=9903.2, *P*<.001).

Table 3. Insured with at least one video consultation (VC) by age group (second quarter 2020), based on Social Health Insurance Fund data (χ27=9903.2, *P*<.001).

≥1 VC	Age (years), n (%)							
	≤20 (n=840,922)	21-30 (n=506,117)	31-40 (n=563,715)	41-50 (n=505,860)	51-60 (n=718,517)	61-70 (n=551,065)	71-80 (n=397,565)	≥81 (n=334,520)
Yes	6600 (0.8)	7287 (1.4)	9378 (1.7)	6304 (1.3)	7140 (1)	3036 (0.6)	1589 (0.4)	1805 (0.5)
No	834,322 (99.2)	498,830 (98.6)	554,337 (98.3)	499,556 (98.7)	711,377 (99)	548,029 (99.4)	395,976 (99.6)	332,715 (99.5)

Differences by Gender

Table 4 shows that video consultations were used disproportionately more often by women in the second quarter of 2020. The correlation was statistically significant (χ^2_1=1533.8, *P*<.001).

Table 4. Insured with at least one video consultation (VC) by gender (second quarter 2020), based on Social Health Insurance Fund data (χ21=1533.8, *P*<.001).

≥1 VC	Women (n=2,369,554), n (%)	Men (n=1,907,594), n (%)
Yes	27,922 (1.2)	15,217 (0.8)
No	2,341,632 (98.8)	1,892,377 (99.2)

Employment Status

Table 5 shows that, for the second quarter of 2020, video consultations were used disproportionately often by employed people than people not in employment and retired persons (χ^2_2=3566.2, *P*<.001).

XSL•FO
RenderX

Table 5. Insured with at least one video consultation (VC) by employment status (second quarter 2020), based on Social Health Insurance Fund data (χ22=3566.2, *P*<.001).

≥1 VC	Employed persons (employees/workers, self-employed persons; n=2,150,544), n (%)	Not employed persons (unemployed, social welfare recipients, emigrants, refugees, students; n=1,096,458), n (%)	Retired persons (n=1,123,538), n (%)
Yes	25,729 (1.2)	8737 (0.8)	7157 (0.6)
No	2,124,815 (98.8)	1,087,721 (99.2)	1,116,381 (99.4)

Place of Residence of the Insured

Table 6 shows that video consultations were used disproportionately more often by the insured living in urban regions in the second quarter of 2020 (χ^2_2=3746.2, *P*<.001). Considering that a central reason for the introduction of video consultations was to improve access in rural areas, this is striking.

Table 6. Insured with at least one video consultation (VC) by place of residence (second quarter 2020), based on Social Health Insurance Fund data (χ22=3746.2, *P*<.001).

≥1 VC	Urban regions (n=2,380,202), n (%)	Mixed regions (n=1,428,651), n (%)	Rural regions (n=609,420), n (%)
Yes	29,184 (1.2)	10,878 (0.8)	3077 (0.5)
No	2,351,018 (98.8)	1,417,773 (99.2)	606,343 (99.5)

In addition, we investigated how video consultations were integrated into the overall treatment of a patient during the quarter; was this done exclusively via video, or was the video consultation only used as a supplement? The exclusive use of video consultations for the treatment of a patient by a physician in a quarter can be identified by a mandatory flag, which leads to a deduction from the physician's quarterly flat rate. Overall, video consultations were used as a supplement in 80% (34,525/43,139) of cases. If this is broken down by region, major differences become apparent. VC users in rural regions used them disproportionately more often exclusively (2001/3077, 65%) compared with those living in urban (4508/29,284, 15.4%) or mixed (2105/10,878, 19.4%) regions (see Table 7).

Table 7. Exclusive or only supplementary use of video consultations (VCs) by place of residence (second quarter 2020), based on Social Health Insurance Fund data (χ22=4285.8, *P*<.001).

Type of VC use	Urban regions (n=29,184), n (%)	Mixed regions (n=10,878), n (%)	Rural regions (n=3077), n (%)	Total (n=43,139), n (%)
VC exclusively	4508 (15.4)	2105 (19.4)	2001 (65)	8614 (20)
VC supplementary	24,676 (84.6)	8773 (80.6)	1076 (35)	34,525 (80)

Video consultations were therefore used less frequently in rural regions, but when they are used in these regions, in 65% (2001/3077) of cases, there was no further personal doctor-patient contact during the quarter.

Characteristics of the Provider Groups (Physicians and Psychotherapists)

Overview of Analysis of the Provider Groups

In the following sections, the results of the comparisons between video consultation providers (VC providers) and nonproviders are presented based on ASHIP data with regard to demographic information, information on their specialist group and practice, and employment type. In addition, we analyzed whether there was an association between the location of the medical practice and the availability of video consultations. The comparison was made on the basis of treatment cases provided by the respective physicians and psychotherapists, with or without the use of video consultations.

Differences by Age Group

Table 8 shows that, for the second quarter of 2020, physicians and psychotherapists in younger age groups had more treatment cases with video consultations than older physicians and psychotherapists. The largest group in relative numbers was those younger than 40 years (12,680/1,235,393, 1%), while only 0.4% (22,766/6,374,630) of treatment cases with video consultations were performed among those older than 60 years. The correlation was statistically significant (χ^2_3=11,338.2, *P*<.001).

Table 8. Treatment cases with and without video consultations (VCs) by physicians or psychotherapists and their age (second quarter 2020), based on Association of Statutory Health Insurance Physicians data (χ23=11,338.2, *P*<.001).

Use of VC	Age (years), n (%)			
	≤40 (n=1,235,393)	41-50 (n=5,480,427)	51-60 (n=10,390,058)	≥61 (n=6,374,630)
With VC	12,680 (1)	28,287 (0.5)	42,208 (0.4)	22,766 (0.4)
Without VC	1,222,713 (99)	5,452,140 (99.5)	10,347,850 (99.6)	6,351,864 (99.6)

XSL•FO
RenderX

Differences by Gender

The comparison of service providers by gender showed that female physicians or psychotherapists used video consultations with treatment cases slightly more frequently than men (see Table 9).

Table 9. Treatment cases with or without video consultations (VCs) by physicians or psychotherapists and their gender (second quarter 2020), based on Association of Statutory Health Insurance Physicians data (χ21=2492.6, *P*<.001).

Use of VC	Women (n=9,053,887), n (%)	Men (n=13,274,371), n (%)
With VC	51,677 (0.6)	55,979 (0.4)
Without VC	9,002,210 (99.4)	13,218,392 (99.6)

Information on the Specialist Group

A comparative analysis of the professional groups showed that physicians and psychotherapists working in the field of psychotherapy (for both adults and children) were the most likely to use video consultations. Physicians working in psychotherapy and psychological psychotherapists used video consultations with 15.9% (44,808/282,530) of their treatment cases in the second quarter of 2020, followed by psychotherapists for children, who used video consultations with approximately 10% (10,828/113,293, 9.6%) of their treatment cases. Although the absolute number of treatment cases with video consultations among general practitioners was very high compared with other specialist groups (29,600 compared with 10,828 among psychotherapists for children), their share of video consultations with all treatment cases in this specialist group was very low, at 0.3% (29,600/9,837,118). This is caused by the high number of general practitioners and treatment cases in this group.

Type of Practice

Table 10 shows that physicians or psychotherapists in individual practices (73,494/11,089,465, 0.7% of treatment cases) used video consultations 2.4 times more frequently than those in joint practices (23,251/8,453,284, 0.3% of cases); this includes different forms of joint practices.

Table 10. Treatment cases with or without video consultations (VCs) by physicians or psychotherapists and their type of practice (second quarter 2020), based on Association of Statutory Health Insurance Physicians data (χ22=14,990.5, *P*<.001).

Use of VC	Individual practice (n=11,162,959), n (%)	Joint practices (n=8,476,535), n (%)	Medical care center (n=2,485,753), n (%)
With VC	73,494 (0.7)	23,251 (0.3)	10,423 (0.4)
Without VC	11,089,465 (99.3)	8,453,284 (99.7)	2,475,330 (99.6)

Location of the Physician's Office

As with the situation for insured persons, Table 11 shows that physicians and psychotherapists in urban regions used video consultations 2.3 times more frequently with treatment cases than those in rural regions.

Table 11. Treatment cases with or without video consultations (VCs) by physicians or psychotherapists and their practice location (second quarter 2020), based on Social Health Insurance Fund data (χ22=8474.1, *P*<.001).

Use of VC	Urban regions (n=12,153,573)	Mixed regions (n=6,870,939)	Rural regions (n=3,303,069)
With VC	72,982 (0.6)	26,136 (0.4)	8538 (0.3)
Without VC	12,080,591 (99.4)	6,844,803 (99.6)	3,294,531 (99.7)

Services and Diagnoses Associated With Video Consultations

In 2020, a total of 715,104 services were identified as having been provided and billed as part of video consultations. Approximately 55% (392,380/715,104, 54.9%) of these were claimable individual psychotherapy services, of which 54% (211,788/392,380) were behavioral psychotherapy, 30.3% (118,986/392,380) were depth psychology psychotherapy, and 15.7% (61,606/392,380) were analytic psychotherapy. Of the 715,104 services provided via video consultation, 30% (214,411/715,104) were counseling services, and about 15% (106,580/715,104, 14.9%) were other psychotherapeutic services. A more detailed analysis showed that the majority of counseling services were psychotherapeutic consultations (168,388/214,411, 78.5%) and only a few were primary care consultations (40,418/214,411, 18.9%).

The analysis of diagnoses (according to the International Statistical Classification of Diseases and Related Health Problems, 10th revision, German Modification) in treatment cases (quarterly reference) with contacts that took place exclusively via video consultation showed a similar picture. The top 10 diagnoses in 2020 included, in particular, diagnoses related to mental and behavioral disorders: reactions to severe stress and adjustment disorders (ICD code F43), depressive episode (F32), other anxiety disorders (F41), recurrent depressive disorder (F33), somatoform disorders (F45), and phobic disorders (F40). In addition, the top 10 most frequently coded diagnoses included back pain (M54), essential (primary) hypertension (I10), and COVID-19 pandemic–related indications such as acute upper respiratory tract infection (J06).

XSL•FO
RenderX

Discussion

Principal Findings and Interpretation of Underlying Reasons

In the years 2017 to 2019, almost no video consultations were used in outpatient medical care in the German health care system. This changed significantly with the onset of the COVID-19 pandemic: In the first and second quarters of 2020, there was a sharp increase in the use of video consultations in all regions analyzed. However, even in the second quarter, the proportion of the insured with at least one video consultation was extremely low, at a maximum of 1%. There was a decline from the second to the third quarter, with a further significant increase in the fourth quarter—a trend in use that is in line with the COVID-19 infection rate.

Video consultations were used significantly more often by younger insured persons, women, employed persons, and the insured living in urban areas. These findings are in line with the results of other studies stating that the younger and employed generation as well as people living in urban environments are more interested in using video consultations than older age groups or people residing in rural areas [3]. Another reason could be that physicians act as gatekeepers and preselect patients, assuming that younger people and people from urban areas with presumably better internet access are more suitable for video consultations [11]. The picture is similar for physicians and psychotherapists, with younger age groups and physicians and psychotherapists with an urban-based practice also using video consultations more frequently. In order to promote the sustainable use of video consultations, it is important to identify the key barriers for those groups. A lack of technical requirements has often been cited as a key barrier for patients as well as physicians and psychotherapists [12-14]. Physicians were also concerned about negative effects on the doctor-patient relationship. In addition, some believed that the lack of physical examination methods could more easily lead to treatment errors or misdiagnosis. Concerns were also expressed about privacy and data security [11,15].

There was no evidence of increased use in structurally disadvantaged rural areas, an often-cited potential for video consultations [3]. A number of hypotheses may be posited to explain the underlying reasons for the observed phenomenon. For instance, the comparatively poorer broadband infrastructure in rural areas or a different age structure could be identified as potential explanatory factors. The population in rural regions is, on average, older and therefore potentially less digitally proficient than that in urban regions. In addition, people from urban regions tend to use psychotherapy, which is the predominant use case for video consultations, more often than people from rural regions [5].

Psychotherapists and physicians working in a psychotherapy specialty (both for adults and children) were, by far, the most frequent users of video consultations. This was also reflected in the analysis of services and diagnoses in video consultations. The reason for this may be the nature of the services in this area, which are primarily what is known as “talking medicine.” Somatic specialists generally rely on physical examinations, blood samples, for example, as a supplement.

Video consultations are primarily used as an adjunct to care and rarely occur without further doctor-patient contact. However, in rural areas, where video consultations are less common, the opposite is true: There is often no face-to-face contact in addition to video consultations. One reason for this could be that these patients used a video consultation with a central platform rather than with their usual physician. On these platforms, personal follow-up contact is usually not possible at all, as physicians from completely different regions of Germany can also offer video consultations there. This may be because their usual physician does not offer video consultations or because there are no physicians in the area with vacancies in the specialty they are looking for. In any case, this is an indicator of a potential lack of care in rural areas.

Strengths and Limitations

This data analysis was based on data from 3 large German SHIFs with a wide range of enrollees (about 6.1 million insured or approximately 8.4% of all SHI-insured) and 4 ASHIPs (about 33,000 physicians and psychotherapists or 18.4% of all SHI-accredited physicians and psychotherapists in Germany). It is also worth noting that insured persons from different types of SHIFs were considered, so that they do not represent only one type of SHIF (eg, in terms of age and gender). The selection of the 4 regions ensured that both rural and urban enrollees and providers were included.

A first limitation could be that only those with statutory health insurance were included, not those with private health insurance, which accounts for only about 10% of all insured persons in Germany.

The time period included allowed for an analysis of the impact of the COVID-19 pandemic on the use of video consultations. The second limitation is that this period could also have had a biasing effect, for example in the analysis of diagnoses coded in relation to video consultations, given the extremely high coding of respiratory diseases in the second quarter of 2020. A slightly longer data period would also have been desirable to analyze the effects after the pandemic. However, the data period used, in particular the year 2020, was considered sufficient for the analysis of the characteristics of the user groups.

Third, the statistical evaluation was hindered by a number of evident challenges. One principal reason for this was that, for reasons of data minimization, only aggregated data were made available to the evaluator, which significantly constrained the range of analysis options. Consequently, only chi-square analysis and, for example, no regression analyses could be conducted. It should be noted, however, that even minor discrepancies can become statistically significant due to the large sample size. To ensure transparency, the absolute and relative data were consistently provided in the cross tables, allowing readers to assess them independently. Additionally, the study faced a challenge with balancing the proportions, with approximately 1% of the sample using a video consultation and the remaining 99% not using it. This imbalance can complicate the analysis.

XSL•FO
RenderX

Fourth, not all variables are available in health insurance data; for example, information on educational status would have been useful. However, in this mixed methods project, a survey was conducted to ask, for example, about the use of video consultations and the level of education [8]. The findings will be published separately.

Fifth, the generalizability of the results for Germany to other countries may be limited. Aspects such as the slow progress of digitization in Germany, as well as the existence of statutory health insurance and a relatively high density of physicians compared with other countries, could explain the hesitant acceptance of video consultations and may not apply to other countries.

Conclusion and Potential for Future Research

The current extent of video consultation use remains well below its potential. Additionally, it is notable that the groups for whom it was anticipated that video consultations would offer substantial benefits, such as individuals with limited mobility and those residing in rural areas, currently use video consultations at a significantly lower rate than groups who are currently considered to be well served, including individuals in urban areas and young people.

Unfortunately, initial analyses of a period at the end of the pandemic are concerning. According to these data, the already low uptake rate declined further toward the end of the pandemic. According to a report by BARMER, a German SHIF, monthly uptake in the second quarter of 2022 was more than 50% lower than during the pandemic waves in the second quarter of 2020 and the first quarter of 2021. It appears to have stabilized at a relatively low level [16].

The widespread and lasting use of video consultations will only succeed if the potential user groups accept this form of service provision and recognize its advantages. Further analyses should therefore investigate the preferences of user groups for the use of video consultations. Both qualitative (interviews, focus group discussions) and quantitative research approaches (surveys [eg, with preference elicitation methods such as discrete choice experiments]) are suitable for this purpose. As part of this study, a survey was conducted among the insured, physicians, and psychotherapists to identify the primary barriers to the use of video consultations [7]. The findings will be released in a separate publication.

Acknowledgments

This study was entirely funded by the Innovation Fund of the German Federal Joint Committee (funding number 01VSF20011). The funder did not influence the design of the study, the collection, analysis, and interpretation of data or the writing process of this manuscript. We acknowledge support by the Open Access Publication Fund of the University of Duisburg-Essen.

Data Availability

The data sets on which the routine data analysis is based are not publicly available for privacy reasons.

Authors' Contributions

All authors contributed to the manuscript at various stages and read and approved the final version. TH was the main contributor in all 14 steps according to the Credit taxonomy.

Conflicts of Interest

None declared.

References

1. Bericht des Bewertungsausschusses und des ergänzenden Bewertungsausschusses zur telemedizinischen Leistungserbringung im Einheitlichen Bewertungsmaßstab (BT-Drucksache 20/4982). Deutscher Bundestag. 2022. URL: https://dserver.bundestag.de/btd/20/049/2004982.pdf [accessed 2024-12-16]
2. Consolidated telemedicine implementation guide. World Health Organization. 2022. URL: https://iris.who.int/bitstream/handle/10665/364221/9789240059184-eng.pdf [accessed 2024-12-16]
3. Waschkau A, Traulsen P, Steinhäuser J. Evaluation of synchronous and asynchronous telemedical applications in primary care in rural regions of northern Germany-results and lessons learned from a pilot study. Int J Environ Res Public Health 2022 Nov 11;19(22):14860 [FREE Full text] [doi: 10.3390/ijerph192214860] [Medline: 36429577]
4. Unterrichtung durch die Bundesregierung - Bericht des Bewertungsausschusses zur Überprüfung des Einheitlichen Bewertungsmaßstabes auf die Möglichkeit zur ambulanten telemedizinischen Leistungserbringung (BT-Drucksache 19/0620). Deutscher Bundestag. 2018. URL: http://dip21.bundestag.de/dip21/btd/19/060/1906020.pdf [accessed 2024-12-16]
5. Gensorowsky D, Surmann B, Schmidt J, Greiner W. [Level of use and user groups of online video consultations in outpatient medical care: analysis of claims data]. Gesundheitswesen 2022 Feb;84(2):110-116. [doi: 10.1055/a-1312-6439] [Medline: 33412593]

XSL•FO
RenderX

6. Petrick N, Kreuzenbeck CCJ. [Effects of the Covid-19 pandemic on online use of video consultation by general practitioners in Germany - a secondary data analysis of German health insurance data]. Gesundheitswesen 2023 Mar;85(3):188-192. [doi: 10.1055/a-1999-7390] [Medline: 36693392]
7. Kleinschmidt L, Walendzik A, Wasem J, Höfer K, Nauendorf B, Brittner M, et al. Preference-based implementation of video consultations in urban and rural regions in outpatient care in Germany: protocol for a mixed methods study. JMIR Res Protoc 2024 Apr 11;13:e50932 [FREE Full text] [doi: 10.2196/50932] [Medline: 38602749]
8. Gesetzliche Krankenversicherung. Mitglieder, mitversicherte Angehörige und Krankenstand. Jahresdurchschnitt 2020. Bundesministerium für Gesundheit. 2020. URL: https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/Statistiken/GKV/Mitglieder_Versicherte/KM1_JD_2020_CPS_bf.pdf [accessed 2024-12-16]
9. Federal Medical Register. Kassenärztliche Bundesvereinigung. 2020. URL: https://www.kbv.de/html/bundesarztregister.php#content6889 [accessed 2024-12-16]
10. Übersicht Vergütung für Videosprechstunde. Kassenärztliche Bundesvereinigung. 2024. URL: https://www.kbv.de/media/sp/Videosprechstunde__uebersicht_Verguetung.pdf [accessed 2024-12-16]
11. Cowan KE, McKean AJ, Gentry MT, Hilty DM. Barriers to use of telepsychiatry: clinicians as gatekeepers. Mayo Clin Proc 2019 Dec;94(12):2510-2523. [doi: 10.1016/j.mayocp.2019.04.018] [Medline: 31806104]
12. Thiyagarajan A, Grant C, Griffiths F, Atherton H. Exploring patients' and clinicians' experiences of video consultations in primary care: a systematic scoping review. BJGP Open 2020 Mar 17;4(1):bjgpopen20X101020. [doi: 10.3399/bjgpopen20x101020]
13. Donaghy E, Atherton H, Hammersley V, McNeilly H, Bikker A, Robbins L, et al. Acceptability, benefits, and challenges of video consulting: a qualitative study in primary care. Br J Gen Pract 2019 Sep;69(686):e586-e594 [FREE Full text] [doi: 10.3399/bjgp19X704141] [Medline: 31160368]
14. Christensen LF, Moller AM, Hansen JP, Nielsen CT, Gildberg FA. Patients' and providers' experiences with video consultations used in the treatment of older patients with unipolar depression: a systematic review. J Psychiatr Ment Health Nurs 2020 Jun 26;27(3):258-271. [doi: 10.1111/jpm.12574] [Medline: 31677331]
15. Mueller M, Knop M, Niehaves B, Adarkwah CC. Investigating the acceptance of video consultation by patients in rural primary care: empirical comparison of preusers and actual users. JMIR Med Inform 2020 Oct 22;8(10):e20813 [FREE Full text] [doi: 10.2196/20813] [Medline: 32969339]
16. Mangiapane S. Five years of video consultations in statutory health care: Has the pandemic led to a breakthrough? In: Repschläger U, editor. BARMER Gesundheitswesen. Berlin, Germany: BARMER Institute for Health Systems Research; 2023:82-99.

Abbreviations

ASHIP: Association of Statutory Health Insurance Physicians
FSP: fee schedule position
Q: quarter
SHI: Social Health Insurance
SHIF: Social Health Insurance Fund
VC: video consultation

Edited by A Schwartz; submitted 03.05.24; peer-reviewed by D Gensorowsky, K Moon, K De Santis, S Muellmann, E van der Velde, D Singh; comments to author 02.10.24; revised version received 20.10.24; accepted 27.01.25; published 14.03.25.

Please cite as:
Hüer T, Walendzik A, Kleinschmidt L, Höfer K, Nauendorf B, Malsch J, Brittner M, Brandenburg P, Aeustergerling A, Schneider U, Wadeck A, Liersch S, Sehlen S, Schwarze K, Wasem J
Use of Video Consultation Between 2017 and 2020 in Outpatient Medical Care in Germany and Characteristics of Their User Groups: Analysis of Claims Data
JMIR Form Res 2025;9:e60170
URL: https://formative.jmir.org/2025/1/e60170
doi:10.2196/60170
PMID:40085136

©Theresa Hüer, Anke Walendzik, Lara Kleinschmidt, Klemens Höfer, Beatrice Nauendorf, Juliane Malsch, Matthias Brittner, Paul Brandenburg, André Aeustergerling, Udo Schneider, Anja Wadeck, Sebastian Liersch, Stephanie Sehlen, Katharina Schwarze, Jürgen Wasem. Originally published in JMIR Formative Research (https://formative.jmir.org), 14.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/),

XSL•FO
RenderX

which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development of a Clinical Decision Support Tool to Implement Asthma Management Guidelines in Pediatric Primary Care: Qualitative Study

David A Fedele[1], PhD; Jessica M Ray[2], PhD; Jaya L Mallela[3], MS; Jiang Bian[2], PhD; Aokun Chen[2], PhD; Xiao Qin[2], MS; Ramzi G Salloum[2], PhD; Maria Kelly[4], MD; Matthew J Gurka[5], PhD; Jessica Hollenbach[6,7], PhD

[1]Center for Healthcare Delivery Science, Nemours Children's Health, Jacksonville, FL, United States

[2]Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, FL, United States

[3]Department of Clinical & Health Psychology, University of Florida, Gainesville, FL, United States

[4]Department of Pediatrics, University of Florida, Gainesville, FL, United States

[5]Department of Public Health Sciences, University of Virginia, Charlottesville, VA, United States

[6]Asthma Center, Connecticut Children's Medical Center, Hartford, CT, United States

[7]Department of Pediatrics, UConn Health, Farmington, United States

Corresponding Author:
David A Fedele, PhD
Center for Healthcare Delivery Science
Nemours Children's Health
807 Children's Way
Jacksonville, FL, 32610
United States
Phone: 1 904 697 3600
Email: david.fedele@nemours.org

Abstract

Background: There is a longstanding gap between national asthma guidelines and their implementation in primary care. Primary care providers (PCPs) endorse numerous provider and practice or clinic-related barriers to providing guidelines-based asthma care. To reduce asthma morbidity in primary care, PCPs need access to tools that facilitate adherence to national guidelines, which can be delivered at the point of care, are minimally burdensome, and fit within the clinic workflow. Clinical decision support (CDS) tools are health IT systems that can be housed in the electronic health record (EHR) system.

Objective: This study aimed to follow user-centered design principles and describe the formative qualitative work with target stakeholders (ie, PCPs and IT professionals) to inform our design of an EHR-embedded CDS tool that adheres to recent, significant changes in asthma management guidelines.

Methods: Purposive sampling was used to recruit three separate subgroups of professionals (n=15) between (1) PCPs with previous experience using a paper-based CDS tool for asthma management, (2) PCPs without previous experience using CDS tools for asthma management, and (3) health care IT professionals. The PCP interview guide focused on their practice, familiarity with national asthma guidelines, and how a CDS tool embedded in the EHR might help them provide guideline-based care. The health care IT professional guide included questions on the design and implementation processes of CDS tools into the EHR. Qualitative data were audio-recorded, transcribed, and then analyzed using an inductive approach to develop themes.

Results: Themes were organized into 2 domains, current practice and CDS tool development. The themes that emerged from PCPs included descriptions of assessments conducted to make an asthma diagnosis, previous attempts or opportunities to implement updated national asthma guidelines, and how a CDS tool could be implemented using the EHR and fit into the current asthma management workflow. The themes that emerged from health care IT professionals included processes used to design CDS tools and strategies to collect evidence that indicated a tool's value to a practice and the broader health system.

Conclusions: In this study, user-centered design principles were used to guide a qualitative study on perceived barriers and facilitators to a primary care–based, EHR-integrated asthma CDS tool. PCPs expressed their interest in adopting an asthma CDS tool that was low burden and efficient but could help them adhere to national asthma guidelines and improve clinic workflow. Similarly, health care IT professionals perceived an asthma CDS tool to be useful, if it adhered to EHR design standards.

XSL•FO
RenderX

Implementation of a CDS tool to improve adherence of PCPs to recently updated national asthma guidelines could be beneficial in reducing pediatric asthma morbidity.

(JMIR Form Res 2025;9:e65794) doi:10.2196/65794

KEYWORDS

clinical decision support; asthma; primary care; guidelines; pediatric; asthma care; morbidity; health information technology; electronic health record; EHR; user-centered design; inductive approach; digital health; health technology

Introduction

Asthma is the most common childhood chronic illness in the United States, affecting >5 million children [1], and is a leading cause of emergency department and urgent care visits [2], school absences [3,4], and reduced quality of life among youth [5,6]. The National Asthma Education and Prevention Program (NAEPP) guidelines are designed to prevent and reduce pediatric asthma morbidity via disseminating state of the science guidance to health care providers on asthma diagnosis and management [7]. NAEPP guidelines call for health care providers to regularly adhere to 4 pillars of asthma management—assessment of asthma control, provision of asthma education to families (eg, written asthma treatment plans), assessment of environmental triggers, and following pharmacological therapy recommendations [4]. Notably, NAEPP guidelines were recently updated in 2020 to include single maintenance and reliever therapy (SMART), a paradigm-shifting change to recommendations for the management of youth with persistent asthma [7].

Use of NAEPP guidelines improves asthma control, reduces the likelihood of exacerbations, and decreases health care usage [4,7-9]. However, there is a persistent gap between NAEPP guidelines and their implementation in practice, particularly among primary care providers (PCPs) [10,11]. PCPs are the frontline health care providers for youth with asthma and provide most ambulatory care [12,13]. Suboptimal adherence among PCPs to national asthma guidelines is documented in numerous studies [10,12,14,15]. For example, a chart review study of a large cohort of primary care practices found that only 9% of PCPs used a validated asthma control tool, just 6% provided an asthma treatment plan, and 30% failed to prescribe anti-inflammatory medication when indicated [16]. This pattern of findings is mirrored in recent results from the National Asthma Survey of Physicians [12].

The considerable gap between NAEPP guidelines and their implementation in primary care are driven by provider and practice or clinic-related factors [12]. PCPs encounter numerous barriers to providing guidelines-based asthma care, including time constraints, lack of familiarity and knowledge of the guidelines, and low self-efficacy surrounding guideline implementation [17,18]. Furthermore, primary care clinics are commonly characterized as suffering themselves from "information chaos" with limited administrative bandwidth for comprehensive, evidenced-based interventions [19]. To reduce asthma morbidity in primary care, PCPs need access to tools that facilitate adherence to national guidelines, which can be delivered at the point of care, are minimally burdensome, and fit within the clinic workflow [15,20,21].

Clinical decision support (CDS) tools are health IT systems that can be housed in the electronic health record (EHR) system and be effective in improving provider adherence to guidelines and patient outcomes [22,23]. CDS tools assist providers in patient-level clinical decision-making (eg, asthma diagnosis and severity designation) at the point of care via collecting relevant data, synthesizing information, and providing real-time recommendations, thereby reducing provider burden and complexity surrounding adherence and implementation of guidelines [24]. CDS tools were shown to be the most promising method for modifying provider behavior in a systematic review of 68 intervention studies on health care provider adherence to national asthma guidelines [14]. There is a critical need for the creation of a CDS tool for pediatric asthma that attends to the needs of PCPs, especially in the current era of recently updated NAEPP guidelines [21].

While paper-based CDS tools have been shown to increase PCP adherence to NAEPP guidelines and reduce asthma-related morbidity including the number of outpatient visits, hospitalizations, and emergency department visits [25-27], the ubiquity of EHRs in delivery and documentation of primary care, including for asthma management [28], provides the opportunity to provide real-time decision support within the electronic workflow. In addition, electronically presented CDS may support better adherence due to automatic provision of guidance and patient-specific suggestions to support decision-making [29]. While recent work provides key guidance to support implementation of the recent guidelines into electronic CDS, further work is necessary to understand local contexts and clinician behaviors that will support successful intervention development and implementation [30]. This study describes our formative qualitative work with target stakeholders (ie, PCPs and IT professionals) to inform our design of an EHR-embedded CDS tool that adheres to the 4 pillars of asthma management and is congruent with recent, significant changes in NAEPP guidelines [7].

Methods

Overview

We used purposive sampling to recruit three separate subgroups of professionals between January and March 2023, that are (1) PCPs with previous experience using a paper-based CDS tool for asthma management, (2) PCPs without previous experience using CDS tools for asthma management, and (3) health care IT professionals. Participants were recruited through study announcement emails to department listservs and brief presentations to raise awareness of the study during department meetings. In total, 15 individuals agreed to participate. We conducted 2 focus groups with PCPs (n=5 and n=4 participants,

XSL•FO
RenderX

respectively), 1 health care IT professional focus group (n=2), 3 PCP individual interviews, and 1 health care IT professional individual interview. Some participants had a previous professional relationship with the study team (eg, had participated in previous studies). The first focus group was with PCPs (n=5) who had previous experience using Easy Breathing, a validated paper-based CDS tool for asthma management that translates NAEPP guidelines into a usable, efficient, and effective format [10,11,25-27,31]. The second focus group (n=4) and 3 individual interviews were with PCPs who did not have previous experience with CDS tools for pediatric asthma. The third focus group (n=2) and 1 individual interview were with health care IT professionals who had previous experience developing ambulatory CDS tools. Focus groups and interviews were conducted over Zoom (Zoom Video Communications) or in person in clinic workrooms. Focus groups and interviews were facilitated by a PhD-level, certified asthma educator with expertise in clinic- and school-based asthma management intervention (JH [female]) or a PhD-level clinical psychologist with expertise in mobile and eHealth asthma interventions (DAF [male]). Both study authors have experience in conducting qualitative interviews with families of children with asthma and medical providers. Participants were made aware of the goals of the study during the informed consent process and the interviewers briefly summarized their professional expertise in pediatric asthma. Participants were compensated US $50 for their time. In addition, a brief demographics survey was distributed during each focus group.

Focus groups and interviews were conducted in a semistructured format, with separate guides developed for PCPs and health care IT professionals. Sessions ranged in length from 25 to 65 minutes, with an average length of 38 minutes. The PCP interview guide included open-ended questions on how the provider currently diagnoses and treats asthma, if they use any validated tools for diagnosis or treatment planning, familiarity with NAEPP guidelines, how a CDS tool embedded in the EHR might help them provide guidelines-based care, and what functions could be useful in such a tool. The health care IT professional guide included questions on design and implementation processes of CDS tools into the EHR, best practices for collaboration with research and clinical teams, and strategies for evaluating and maintaining CDS tools across multiple clinics.

Focus groups and interviews were audio-recorded, transcribed, and then analyzed by 3 coders (DAF, JH, and JLM) using NVivo 14 (Lumivero). Field notes were taken during data collection. We used an inductive approach by coding data based on similar topics and content across transcripts, clustered by PCPs and health care IT professionals, to develop themes. Themes were derived from overlapping content between PCPs' data and included descriptions of assessments conducted to make an asthma diagnosis, previous attempts or opportunities to implement updated NAEPP guidelines, and how a CDS tool could be implemented using the EHR and fit into current asthma management workflow. Themes from overlapping content between health care IT professionals' data included discussion of processes used to design CDS tools (eg, initial assessment of required functionality, testing mock designs to evaluate alignment with PCP workflow) and strategies to collect evidence that indicated a tool's value to a practice and the broader health system. We then collapsed themes into 2 domains, Current Practice and CDS Tool Development. The 3 coders then used a final codebook that described the themes in each domain to analyze all transcripts (Table 1) independently. Codes for each transcript were then discussed and modified collectively until consensus was reached.

XSL•FO
RenderX

Table 1. Qualitative codebook for interviews with primary care providers and information (for technology health professionals).

Domain and code	Definition
Current practice	
PCPs[a] describe their asthma diagnosis process and workflow	Description of assessments PCPs conduct to make an asthma diagnosis.
PCPs describe asthma follow-up care	Descriptions of symptom monitoring at follow-up visits.
Barriers to asthma management	Mentions of barriers to making asthma diagnosis and to monitoring symptoms at follow-up visits.
Asthma action and treatment plans	Statements about how asthma treatment regimen is communicated to families.
Opportunities to implement new guidelines	Statements about barriers to implementing new guidelines and explanations of previous attempts to implement them (individual PCP or practice-wide).
CDS[b] tool development	
How can a CDS tool fit into provider workflow	Recommendations on components and functionality of ideal tool, including patient surveys, how responses are displayed for the provider, how treatment or prescription options are displayed for the provider.
How can a CDS tool be best designed	Recommendations on design strategies and examples of tools that were successful or unsuccessful in having the components or functionality that providers wanted and used.
Strategies to evaluate impact of a CDS tool	Explanations of how to gather evidence needed to demonstrate tool's value to a practice and the institution or health system. Mentions of pipeline through which projects move forward.

[a]PCP: primary care provider.

[b]CDS: clinical decision support.

Ethical Considerations

The institutional review board at the University of Florida approved all study procedures and measures (IRB 202202480). All individuals provided informed consent before participating in the current study and were reimbursed US $50 for their time. All study data presented below are deidentified.

Results

Participant Demographics

Across the 3 subgroups, 11 participants reported their gender as women and 4 reported as men. Participants reported their race as White (n=11), Black or African American (n=1), or Asian (n=3), and 2 reported their ethnicity as Hispanic or Latino-a-x. To account for providers with multiple racial or ethnic identities, participants could select more than 1 race or ethnicity response such that the total n and percentages summed to greater than 15 and 100%, respectively. PCPs using a CDS tool had an average of 14.6 (SD 13.9) years in practice, while those not using a CDS tool had an average of 16.1 (7.6) years (Tables 2 and 3).

Table 2. Demographics of study participants.

Characteristic	Participants, n (%)
Gender	
Men	4 (27)
Women	11 (73)
Race or ethnicity[a]	
White	11 (73)
Black or African American	1 (7)
Asian	3 (20)
Hispanic or Latino-a-x	2 (13)

[a]Participants could select more than one race or ethnicity response.

XSL•FO
RenderX

Table 3. Demographics of study participant subgroup.

Participant subgroup	Years in practice		
	n	Mean (SD)	Range
PCPs[a] with asthma CDS[b] tool experience	5	14.6 (13.9)	3-36
PCPs without asthma CDS tool experience	7	16.1 (7.6)	3.5-27
IT professionals	3	—[c]	—

[a]PCP: primary care provider.

[b]CDS: clinical decision support.

[c]Not available.

Current Practice for Asthma Management

Within the Current Practice domain, five themes were developed, that are (1) description of asthma diagnosis processes, (2) asthma follow-up care, (3) barriers to asthma management, (4) asthma action or treatment plans, (5) opportunities to implement updated NAEPP guidelines (refer to Table 1 for full codebook and definitions of themes). Quotations from participants in each of the three subgroups are identified as (1) PCPs A through E for those with asthma CDS tool experience, (2) PCPs 1 through 10 for those without asthma CDS tool experience, and (3) IT professionals 1 through 3.

Diagnosis Processes and Follow-Up Care

PCPs currently implementing a paper-based asthma CDS tool described a patient-reported survey to guide them in making an asthma diagnosis. They described assessing asthma control with the Asthma Control Test (ACT) during visits for specific patient ages (eg, 4 and 8 years) or on designated days of the week. Importantly, these PCPs discussed variable CDS tool implementation and workflows across clinics. Providers expressed the ease of conducting follow-up care using the ACT, which is one of the validated measures included in the CDS tool. The ACT is a validated measure of asthma control and gives providers an objective measure, making overall workflow more standardized and efficient.

> *It's much easier every time [to use a CDS tool]. Implementation science tells you that you know just when it's called for. That's part of why we do the Asthma Control [Test] at every visit. Instead of somebody having to figure out whether the patient should be diagnosed with asthma or not, and the symptom severity level.* [PCP C (with CDS tool experience)]

Alternatively, PCPs without experience using an asthma CDS tool described their diagnosis and follow-up processes as based primarily on symptom presentation**,** including recurrent cough, nighttime cough, chest tightness, shortness of breath, and wheezing, with some using structured observation to assess treatment response for acute exacerbations (ie, monitoring symptoms in clinic after administering albuterol) or level of control at follow-up visits. In total, 3 PCPs mentioned using recommendations, such as those published yearly by GINA (Global Initiative for Asthma) to inform their diagnostics.

> *The GINA guidelines…Yeah, it's what I know, and it's quite updated, and I like it, and I'm comfortable. But I will tell you it's not the easiest thing to do, especially in the clinic setup. Usually in my previous practice, we had a lot of things that we're clicking on, and you know it sort of generates for you at least a good estimate. And the reason why I say this because I only have 20 min to do, you know…Even if I suspect that this kid is asthma, after 10 minutes registering and getting roomed in, I only have 10 minutes to figure this out so.* [PCP 7 (without CDS tool experience)]

It was also apparent that PCPs without experience using an asthma CDS tool had knowledge of national and global asthma care recommendations but relied on a more general familiarity with recommendations rather than using a specific tool to aid in their adherence.

> *I probably should do something, but I usually just ask questions to kinda get an idea of how poorly they are controlled, based on how many times they are using it [quick-relief medication]. Basically going through the guidelines, but not as complete.* [PCP 10 (without CDS tool experience)]

Barriers to Asthma Management

All PCPs endorsed time and workflow management as barriers to managing asthma care and adhering to guidelines consistently and thoroughly.

> *I would say the main limitation for this and any guideline or anything else we do in clinic is just time. Cause there's so much to cover in such a short period of time, and trying to prioritize the most important things and spend the limited amount of time that we have on those. Because the most important thing is, albuterol whenever you're wheezing, and then for the kids that are on steroids, to use them consistently, to be able to emphasize that in a way that it will happen whenever they leave. So that's a hard thing to do.* [PCP 4 (without CDS tool experience)]

Among PCPs using a paper-based CDS tool, needing to transfer information from paper documentation into multiple EHR modules (eg, progress note, prescriptions, after-visit summary, and patient instructions) was reported as a barrier.

> *The scoring, writing your impression, especially the [patient]survey. And then you're gonna, you know, basically do the same thing again, or some similar variation of it in the note or in the chart for sure it's*

XSL•FO
RenderX

> *like it's definitely work.* [PCP A (with CDS tool experience)]

Potentially missing asthma diagnoses due to the lack of a CDS tool at the point of care was also identified as a barrier to management.

> *Missing the possible, not the ones that are established. For the ones you're considering a new diagnosis. It's a little bit out of sight out of mind. So people aren't thinking about [the CDS tool] to administer the survey, and that's a little bit cumbersome to go and try to find it.* [PCP A (with CDS tool experience)]

Although PCPs not using a CDS tool described prioritizing the most important aspects of management in their verbal communication with families, they also reported gaps potentially caused by a lack of standardized and accessible treatment plans. In addition, 3 of these PCPs noted insurance coverage of medication costs as influential on their prescribing decisions.

> *We typically just tell them the medications, and which one is which. You know it'd be nice if I had some kind of chart with all of them on there. The parents, and they usually do get confused. Is it the orange one? Is it this one? So, but I try to go over it with them, which one is rescue, which one is the daily medication. But yeah, we don't have examples of them in the clinic.* [PCP 10 (without CDS tool experience)]

> *So, the biggest barrier that I've had is like people that get insurance denies like Flovent versus QVAR…I'm not even sure what the insurance wants. Then, the weirdest things is they want to replicate for kids, like that's the only one they'll approve, but it's for a kid that can't even use it. Like the barrier, I would say it would be cost, and then sometimes instead of doing like a low [dose]… sometimes the insurance will only cover like a high dose. And so you're doing it [a higher dose] instead of like a lower dose twice a day. But you're just hoping that the overall works out to be okay. You have to base on cost. And so I can think of one [patient] in particular that we've had to change it based on cost.* [PCP 5 (without CDS tool experience)]

Asthma Action Plans

All PCPs reported using treatment plans to communicate with families but used variable methods to accomplish this task. The use of asthma action plans that are printed or sent to patients via a patient portal was reported by the PCPs using a paper-based CDS tool. Sending additional copies of the treatment plan was described as a useful function.

> *The majority of families that I have that are actually following their treatment plan. Aware of it adherent usually have a paper copy, that is probably on the fridge or something like that…I do appreciate that if it [the treatment plan] was not done, or if they need another copy or a nurse needs another copy, I can very quickly send it to them again [via patient portal].* [PCP A (with CDS tool experience)]

Furthermore, 3 PCPs commented that confirming if a treatment plan documented on paper was up-to-date was difficult and may be easier with an EHR-embedded tool that could track treatment plan documentation over multiple visits.

> *You don't know when it [the treatment plan] was updated you have to search and dig, to find out when it's updated…And then there's the prescribing piece where you're writing all this out, and then you get to do it again out in the prescription module.* [PCP C (with CDS tool experience)]

PCPs not using a paper-based CDS tool indicated using different strategies to communicate treatment plans, for example, with inhaler labels, verbal explanations of which medications to take when, visual aids, and treatment plan templates from previous training. Some clinicians also used EHR-based tools, including personally developed or borrowed dot phrases (ie, preformatted content that can be pulled into a clinical note using an abbreviation or word), but the type of tool and use differed between PCPs.

> *I have a dot phrase for an asthma action plan, which I feel like I do more with ones that have controller medicine.* [PCP 2 (without CDS tool experience)]

> *I do mostly verbal. Sometimes we'll look up pictures of the different inhalers on the computer just to make sure that they know which one is red, which one is orange, and how can they use it. And then I have like a PDF of an asthma action plan, but I don't use it as often as I probably should.* [PCP 4 (without CDS tool experience)]

> *Yeah, it's a smartphrase from a colleague of mine who she got from her attending when she was a resident… I really like the colored ones, because it's more, if it looks good, the parents will read it…So, it's just a lot of typing.* [PCP 7 (without CDS tool experience)]

Opportunities to Implement Guidelines

Several PCPs not using a CDS tool referenced awareness of national guidelines and intention to use in practice but noted relying primarily on how they were trained. More experienced PCPs acknowledged that they may be most familiar with previous versions of guidelines while both newer and more experienced providers expressed a desire for training on more recent guidelines.

> *You know, it's probably the guidelines that I learned … when I finished residency. It is what was ingrained in me. So, I probably, I know the guidelines changed a few years ago where you can use controller meds more intermittently. I still struggle to wrap my head around that concept.* [PCP 1 (without CDS tool experience)]

> *We do have grand rounds every week. That would be a good avenue that I think of. Or you know, to make it a bit more personal level, like someone can go to the clinic. We do have provider meetings every second Thursday. It will be great if someone could, because you know we're general. We're the gatekeepers. With*

XSL•FO
RenderX

> *all the guidelines you can barely keep up, so it may be a bit of a spoon-feeding, but I mean it's part of you know provider support.* [PCP 7 (without CDS tool experience)]

> *I know GINA just had a recent update, so I looked through it, but I wouldn't say that I remember everything that's on it. I would definitely look over it to see what the updates were. The biggest thing I got out of it, because there's a lot of stuff that's beyond what I would do as a primary care physician in peds. But I remember the biggest change I think, was using an inhaled corticosteroid as a rescue inhaler. But I don't know how many people are actually following that, at this point. To be honest, I haven't really even started doing that, either. It's hard to change your ways.* [PCP 10 (without CDS tool experience)]

CDS Tool Development

Within the CDS Tool Development domain, 3 themes were developed based on health care IT professionals' and PCPs' responses. First, how CDS tools can fit into provider workflow; second, best practices for CDS tool design; and third, strategies to evaluate the impact of CDS tools.

CDS Tool Fit With Provider Workflow

Health care IT professionals emphasized the importance of provider perspectives in design processes to promote implementation that aligns with realistic clinic workflow. The discrepancy between providers' initial conceptualization of a CDS tool and the clinical outcomes they seek to improve was highlighted:

> *What we find often, I think, is that providers will come to us saying they want something but, when we actually talk with them, have a conversation with them about what they're trying to accomplish, it turns out that what we can do for them is completely different than what they thought they originally wanted.* [IT professional]

PCPs using a paper-based CDS tool provided examples of how they might translate their processes into the EHR, noting opportunities to reduce providers' time and documentation burden. They also suggested sending asthma questionnaires to patients via patient portal before their appointment to inform diagnosis and treatment planning but commented that patient response rates would be a limitation of this strategy.

> *Going from your script to the form… it's a lot of effort to sort of click your way through. If doing all that work to make a treatment plan could then just generate my script at the same time.* [PCP A (with CDS tool experience)]

> *This paper ACT [Asthma Control Test], I generally actually think it's distributed pretty consistently. What I find is it's often done during the visit, so I'm not necessarily looking at it to the very end, or I'm asking them to finish it. So we can keep our conversation. If there were a way that they could, have the ACT done prior. We have some of our pre-visit questionnaire and then surveys that we that we do. They can access ahead of the visit. Not that a lot of them take that. But I think if there were people that were more willing to take that up, it would certainly be very helpful to them.* [PCP A (with CDS tool experience)]

> *And then it [the questionnaire] pops into Admin. It says you know they have asthma and you're pulled right in with your smart phrase, and you can review it right there, the scores and everything.* [PCP D (with CDS tool experience)]

Importantly, PCPs indicated a willingness to modify their workflow to include a CDS tool if embedded in the EHR. Specifically, PCPs mentioned EHR functions they believed could be useful, including reminders of established diagnoses, treatment options, progress note templates, and handouts for families.

> *If they are known asthmatics, it would be good if there would be an automatic bullet, and I mean I don't have to look for it, I don't have to type it, it'll remind me that hey this kid is asthmatic. In terms of diagnosing, it will be great to have something that we can just click, click, click, and maybe at some point something that they can fill out as well. Especially if they come for wheezing or cough. Then it [the EHR] can all summarize it for us. So that you know it's a more realistic, doable approach. Everything is on our shoulders.* [PCP 7 (without CDS tool experience)]

I like a good order set that kind of spell things out a little bit. I think especially for these updated, like possible treatment options to be reminded of that, of what the options are, and what these new guidelines are.Cause I think otherwise without that reminder, I'm just going to fall back on what I was trained to do the last three years. [PCP 2 (without CDS tool experience)]

> *It would be cool if it was embedded in a note template. Everybody tends to use the same well child check templates, and then we all make little tweaks on it. But if there is a standard asthma template, and I guess if you clicked like the certain things, and then it embedded where they were, and if they needed to step up. It could define. It's not too much thinking about whether they're mild, moderate, or severe right. There's like definitions for this, and so it could quickly like auto populate that sort of stuff and have like something to show the parents.* [PCP 5 (without CDS tool experience)]

Best Design Practices

Health care IT professionals discussed the importance of using tools the way they are already designed within the EHR, which often requires first understanding current workflows and then guiding clinicians from their usual practice toward a modified workflow that accomplishes clinical goals while being compatible with existing functionality. Compromise between individual PCPs' nuances in how they work and new workflows can be challenging. As such, they indicated that regular communication with a designated physician champion for the

XSL•FO
RenderX

implementation of a new CDS tool is paramount to an effective design process and subsequent clinician buy-in.

> *…we ask them, how are you doing this today?... Sometimes they'll ask for something that is totally different than the process they're following today. They may be asking for this because they want to simplify things, but they may be asking it because they're in that mindset of oh, I'm electronic now, and the way we did it before isn't applicable at all. But that's not necessarily the case.* [IT professional 2]

> *We are almost always asking them to do something different. So even if it is, for instance, sending out some sort of inventory, filling out a form, they still have to navigate to a place to get to that form and open it. It is that sort of moving them from what they're doing now to something completely different, even though it potentially has a lot of benefit for them.* [IT professional 1]

The health care IT professionals also provided an example of mechanisms for different CDS tools, such as pulling data from multiple records into a health maintenance summary to track chronic care (eg, routine screenings), drug dosing calculators within the medication ordering processes, and best practice alerts (BPAs) using in conjunction with questionnaires to guide decision-making. Importantly, they highlighted that even though clinicians may envision a new tool as a BPA pop-up, the risk of alert fatigue may compromise the potential utility of the tool in changing clinician behavior as well as negatively impact other CDS tools.

> *We're trying to really integrate our data analytics group generating reports to say, "How are we doing with this clinical decision support tool depending upon the nature of the tool?" We want to see that at least meeting expectations in a changing behavior because if it's not, and we have alert fatigue, then, of course, that also has a deleterious effect on all the other local decision support systems we put out there.* [IT professional 3]

> *We don't just want to encourage people to make BPAs for everything, because ultimately they get ignored, and so there might be some other tool. Understanding exactly what they are wanting to do and where they envision this working in their workflow.* [IT professional 1]

Finally, the health care IT professionals emphasized consideration of how EHR changes may affect a broader health system outside of the clinic implementing a new tool.

> *Sometimes an ask will be something that we need to set at what we call the system level. Which means it would impact users at [other clinic sites]…And because of the way they practice, which sometimes can be completely different, we have to come up with either work around where we don't have to set something at the system level, which those aren't always available.* [IT professional 2]

Evaluation of Impact

Working closely with a physician champion to share a comprehensive understanding of existing workflows and potential changes informs evaluation of a new CDS tool. One health care IT professional described how workflow events (ie, discrete steps a provider is taking in the EHR during an encounter) must be captured in a reporting system to reflect whether the tool is supporting physician decision-making behavior in an intended manner. Demonstrating impact to key stakeholders, for example, health system financial leadership, physicians being asked to implement a new CDS tool, and other clinical professionals working alongside them, was discussed as essential to willingness to change usual practices. Considering impacts on the broader health system was again highlighted as critical for regulatory reporting, a standard aspect of CDS tool implementation.

> *I ask the question about how do you want to demonstrate to your stakeholders that you've made an impact. So how are we going to be able to report on this and show, what are you going to measure, to demonstrate that it's efficacious, so that they can then sell this, so they can get people on board to follow. And use the tools we've implemented.* [IT professional 1]

> *We must know that workflow, because at certain points those key workflow events can be captured in a reporting system that way. We know that whatever report we have is reflective of the workflow to make sure that we're indeed doing the right thing.* [IT professional 3]

> *There's also differences within the various departments, various workflows, not just among clinicians among departments that we may be wanting to implement a tool that crosses, so that from a from a maintenance standpoint we have a single tool that we can use to collect data for across the organization and report out in one place. We must make sure that tool is going to work for everyone, and they still yet all those results and information still funnels out to the same place.* [IT professional 1]

Discussion

Principal Findings

The goal of the current study was to inform the design of an EHR-embedded CDS tool that adheres to the 4 pillars of NAEPP guidelines by gathering qualitative data from target stakeholders (ie, PCPs and health care IT professionals). Overall, PCPs described a variety of barriers to consistent guideline adherence during their care of children with asthma. Commonly endorsed barriers included having limited time during encounters, which was linked to inconsistency in the assessment of asthma symptoms and inconsistent provision of asthma treatment plans. Furthermore, PCPs described current difficulties with integrating guidelines into their workflow and that searching for documentation tools or printed resources was often challenging. These data are consistent with previous research showing that PCPs regularly encounter extrinsic barriers that limit their ability

XSL•FO
RenderX

to adhere to asthma guidelines [20]. Table 4 summarizes the key themes and conclusions that EHR-embedded CDS tool can improve care and guideline adherence.

Table 4. Summary of key themes and conclusions from qualitative interviews.

Theme	Conclusion
Limited time leads to inconsistent assessment of asthma symptoms and variable communication about treatment plans depending on provider	EHR[a]-embedded CDS[b] tool can improve care through efficient, standardized symptom assessment and diagnosis at the point of care and templates and resources to optimize provision and documentation of consistent treatment plans
Providers are aware of national guidelines but lack supports to implement them in practice	EHR-embedded CDS tool can improve guidelines adherence through structured support throughout clinical encounter to promote guideline implementation and fit within existing clinical procedures

[a]EHR: electronic health record.

[b]CDS: clinical decision support.

Data from this study suggest that CDS tools were positively viewed by PCPs and may be beneficial for overcoming noted barriers to improve provider adherence to NAEPP guidelines. PCPs who have previously used a paper-based CDS tool described that it helps them more readily identify new asthma diagnoses, improves their adherence to several NAEPP guidelines (eg, assessment of asthma control and provision of treatment plan), and were enthusiastic about its implementation in the EHR to improve asthma management. Similarly, PCPs who had not used an asthma-focused CDS tool identified several areas where such a tool could be beneficial to improve adherence to NAEPP guidelines (eg, improving consistency in provision of treatment plans) and expressed interest in incorporating a CDS tool into the EHR. These findings align with previous studies demonstrating that CDS tools improve provider adherence to national asthma guidelines [20] and are viewed favorably by PCPs [21,32], suggesting that CDS tools may be an especially viable mechanism to increase PCP adherence to NAEPP guidelines.

A notable change in recent guidelines is the inclusion of SMART [7]. SMART shows promise in reducing asthma exacerbations among those individuals with suboptimal asthma control [33]. However, although PCPs who did not currently use a CDS tool for asthma management indicated a general awareness of new guidelines, they acknowledged often relying on how they were originally trained to treat pediatric asthma, being unfamiliar with key changes to guidelines, and having limited professional development opportunities to familiarize themselves with updated guidelines. Thus, an important takeaway from the current study was that integration of an asthma management CDS tool into the EHR may facilitate adoption of SMART and be especially timely [30].

A unique aspect of this study is that we purposely included health care IT professionals to gather their perspective on how to best integrate a CDS tool into the EHR. Overall, their feedback echoed best practices for the development of CDS tools, which include designing for the end user's context in mind and providing the right information at the right time in the PCPs' workflow [34]. Feedback from PCPs on the development of our CDS tool aligned with that of that health care IT professionals and largely focused on ways the tool could improve efficiency and consistency. Specifically, PCPs described a desire to have guidance across the encounter to assist with diagnosis, severity determination, assessment of asthma control, and prescribing and ordering in a manner that is concordant with NAEPP guidelines. Furthermore, there was strong interest in support in the EHR for automating the note writing and treatment plan components of the encounter to reduce time-related barriers, improve documentation of relevant asthma-related information, and increase the likelihood that families would be provided an asthma treatment plan.

Strengths and Limitations of This Study

Our study has several notable strengths. First, consistent with strengths of this study include following best practices for the formative development of EHR-embedded CDS tools, we purposely engaged the end users of our planned CDS tool to provide their perspectives on its need at the point of care, the fit within their workflow, and its potential features. A unique strength of this study is the inclusion of the perspectives of 2 groups of PCPs, those with a longstanding history of using an effective paper-based CDS tool for pediatric asthma and those who have not used CDS in asthma care, across 2 health care systems in geographically distinct areas. This allowed us to gather diverse perspectives on our key research questions. Finally, we also gathered important data from health care IT professionals to understand the opportunities and challenges that scientific teams interested in developing EHR-based CDS tools should consider.

There are several study limitations that should be considered. First, we acknowledge that this study included a modest number of health care providers, all of which represent clinicians from large academic medical centers and use the EHR in their daily practice. Thus, although it is notable that health care providers were recruited from 2 distinct academic medical centers, the findings from this study may not generalize to health care providers with less experience using an EHR or who practice outside of academic medicine. Second, the responses that we received were consistent across individuals, however, we acknowledge that the current study included a small number of health care IT professionals. It is possible that recruitment of additional health care IT professionals would have yielded different opinions or data. Third, the health care IT professionals were from a single academic medical center that uses an enterprise-wide EHR system. Therefore, it is possible that their opinions may not be generalizable to other health care systems or EHR platforms. Finally, participants were largely recruited via convenience sampling–based, introducing some concern for

XSL•FO
RenderX

selection bias. We attempted to limit this concern by directly encouraging open and honest responses and ensuring that the study investigator did not have a professional relationship with the participant whenever feasible.

Conclusions and Future Directions

In this study, we used user-centered design principles to guide a qualitative study on perceived barriers and facilitators to a primary care–based, EHR-integrated asthma CDS tool. Participants in this mixed-methods approach expressed their interest in adopting an efficient asthma management CDS tool that could help them adhere to NAEPP guidelines and improve clinic workflow. Similarly, health care IT professionals perceived having an asthma CDS tool to be useful, as long as it adhered to EHR design standards. Additional recommendations from IT professionals included limiting BPA-alert fatigue and building automated reports to track use of the CDS tool. Thus, future directions will be to design an asthma CDS tool that (1) prompts PCPs in a low-burden approach that limits alert fatigue, (2) guides PCPs in diagnosing asthma and assigning a severity using native EHR tools, (3) provides PCPs with guidelines- and severity-concordant therapies to create an asthma treatment plan, and (4) streamlines follow-up care using EHR-embedded asthma control assessment tools.

Acknowledgments

We would like to thank the health care providers who took time away from their busy work schedules to provide their professional opinions and guidance on this work. Research reported in this publication was supported by the University of Florida Clinical and Translational Science Institute, which is supported in part by the National Center for Advancing Translational Sciences of the National Institutes of Health (award UL1TR001427). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

The datasets generated during and/or analyzed during this study are not publicly available due to institutional regulations but may be available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Most recent national asthma data. Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/asthma/most_recent_national_asthma_data.htm [accessed 2024-01-14]
2. Masoli M, Fabian D, Holt S, Beasley R, Global Initiative for Asthma (GINA) Program. The global burden of asthma: executive summary of the GINA dissemination committee report. Allergy 2004;59(5):469-478. [doi: 10.1111/j.1398-9995.2004.00526.x] [Medline: 15080825]
3. Most recent national asthma data. CDC. URL: https://www.cdc.gov/asthma/most_recent_national_asthma_data.htm [accessed 2025-01-14]
4. Guidelines for the Diagnosis and Management of Asthma 2007 (EPR-3). Bethesda, MD: National Heart, Lung, and Blood Institute; 2007.
5. Cui W, Zack MM, Zahran HS. Health-related quality of life and asthma among United States adolescents. J Pediatr 2015;166(2):358-364 [FREE Full text] [doi: 10.1016/j.jpeds.2014.10.005] [Medline: 25454936]
6. Bruzzese JM, Bonner S, Vincent EJ, Sheares BJ, Mellins RB, Levison MJ, et al. Asthma education: the adolescent experience. Patient Educ Couns 2004;55(3):396-406. [doi: 10.1016/j.pec.2003.04.009] [Medline: 15582346]
7. Expert Panel Working Group of the National Heart, Lung,Blood Institute (NHLBI) administeredcoordinated National Asthma EducationPrevention Program Coordinating Committee (NAEPPCC), Cloutier MM, Baptist AP, Blake KV, Brooks EG, Bryant-Stephens T, et al. 2020 focused updates to the asthma management guidelines: a report from the national asthma education and prevention program coordinating committee expert panel working group. J Allergy Clin Immunol 2020;146(6):1217-1270 [FREE Full text] [doi: 10.1016/j.jaci.2020.10.003] [Medline: 33280709]
8. Gibson PG, Powell H. Written action plans for asthma: an evidence-based review of the key components. Thorax 2004;59(2):94-99 [FREE Full text] [doi: 10.1136/thorax.2003.011858] [Medline: 14760143]
9. Janson SL, McGrath KW, Covington JK, Cheng SC, Boushey HA. Individualized asthma self-management improves medication adherence and markers of asthma control. J Allergy Clin Immunol 2009;123(4):840-846 [FREE Full text] [doi: 10.1016/j.jaci.2009.01.053] [Medline: 19348923]
10. Cloutier MM, Wakefield DB, Sangeloty-Higgins P, Delaronde S, Hall CB. Asthma guideline use by pediatricians in private practices and asthma morbidity. Pediatrics 2006;118(5):1880-1887. [doi: 10.1542/peds.2006-1019] [Medline: 17079558]
11. Cloutier MM, Grosse SD, Wakefield DB, Nurmagambetov TA, Brown CM. The economic impact of an urban asthma management program. Am J Manag Care 2009;15(6):345-351 [FREE Full text] [Medline: 19514800]

XSL•FO
RenderX

12. Akinbami LJ, Salo PM, Cloutier MM, Wilkerson JC, Elward KS, Mazurek JM, et al. Primary care clinician adherence with asthma guidelines: the national asthma Survey of physicians. J Asthma 2020;57(5):543-555 [FREE Full text] [doi: 10.1080/02770903.2019.1579831] [Medline: 30821526]
13. Wisnivesky JP, Lorenzo J, Lyn-Cook R, Newman T, Aponte A, Kiefer E, et al. Barriers to adherence to asthma management guidelines among inner-city primary care providers. Ann Allergy Asthma Immunol 2008;101(3):264-270. [doi: 10.1016/S1081-1206(10)60491-7] [Medline: 18814449]
14. Okelo SO, Butz AM, Sharma R, Diette GB, Pitts SI, King TM, et al. Interventions to modify health care provider adherence to asthma guidelines: a systematic review. Pediatrics 2013;132(3):517-534 [FREE Full text] [doi: 10.1542/peds.2013-0779] [Medline: 23979092]
15. Cabana MD, Rand CS, Becher OJ, Rubin HR. Reasons for pediatrician nonadherence to asthma guidelines. Arch Pediatr Adolesc Med 2001;155(9):1057-1062. [doi: 10.1001/archpedi.155.9.1057] [Medline: 11529809]
16. Yawn BP, Rank MA, Cabana MD, Wollan PC, Juhn YJ. Adherence to asthma guidelines in children, tweens, and adults in primary care settings: a practice-based network assessment. Mayo Clin Proc 2016;91(4):411-421 [FREE Full text] [doi: 10.1016/j.mayocp.2016.01.010] [Medline: 26944837]
17. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly: Management Information Systems 1989;13(3):319-339. [doi: 10.2307/249008]
18. Wood BL, Miller BD, Lehman HK. Review of family relational stress and pediatric asthma: the value of biopsychosocial systemic models. Fam Process 2015;54(2):376-389. [doi: 10.1111/famp.12139] [Medline: 25683472]
19. Beasley JW, Wetterneck TB, Temte J, Lapin JA, Smith P, Rivera-Rodriguez AJ, et al. Information chaos in primary care: implications for physician performance and patient safety. J Am Board Fam Med 2011;24(6):745-751 [FREE Full text] [doi: 10.3122/jabfm.2011.06.100255] [Medline: 22086819]
20. Cloutier MM. Asthma management programs for primary care providers: increasing adherence to asthma guidelines. Curr Opin Allergy Clin Immunol 2016;16(2):142-147. [doi: 10.1097/ACI.0000000000000242] [Medline: 26849166]
21. Bousquet J. Electronic clinical decision support system (eCDSS) in the management of asthma: from theory to practice. Eur Respir J 2019;53(4):1900339 [FREE Full text] [doi: 10.1183/13993003.00339-2019] [Medline: 31023866]
22. Jaspers MWM, Smeulers M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. J Am Med Inform Assoc 2011;18(3):327-334 [FREE Full text] [doi: 10.1136/amiajnl-2011-000094] [Medline: 21422100]
23. Kruse CS, Ehrbar N. Effects of computerized decision support systems on practitioner performance and patient outcomes: systematic review. JMIR Med Inform 2020;8(8):e17283. [doi: 10.2196/17283] [Medline: 32780714]
24. Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med 2012;157(1):29-43 [FREE Full text] [doi: 10.7326/0003-4819-157-1-201207030-00450] [Medline: 22751758]
25. Cloutier MM, Hall CB, Wakefield DB, Bailit H. Use of asthma guidelines by primary care providers to reduce hospitalizations and emergency department visits in poor, minority, urban children. The Journal of Pediatrics 2005;146(5):591-597. [doi: 10.1016/j.jpeds.2004.12.017]
26. Cloutier MM, Wakefield DB. Translation of a pediatric asthma-management program into a community in connecticut. Pediatrics 2011;127(1):11-18 [FREE Full text] [doi: 10.1542/peds.2010-1943] [Medline: 21135006]
27. Cloutier MM, Jones GA, Hinckson V, Wakefield DB. Effectiveness of an asthma management program in reducing disparities in care in urban children. Ann Allergy Asthma Immunol 2008;100(6):545-550. [doi: 10.1016/S1081-1206(10)60058-0] [Medline: 18592817]
28. Hollenbach JP, Collins MS, Wasser C, Fedele D. Implementation of standardized asthma management programs in outpatient settings. Ann Allergy Asthma Immunol 2023;130(5):571-576. [doi: 10.1016/j.anai.2023.01.016] [Medline: 36702245]
29. Van de Velde S, Heselmans A, Delvaux N, Brandt L, Marco-Ruiz L, Spitaels D, et al. A systematic review of trials evaluating success factors of interventions with computerised clinical decision support. Implement Sci 2018;13(1):114 [FREE Full text] [doi: 10.1186/s13012-018-0790-1] [Medline: 30126421]
30. Shero ST, Ammary-Risch NJ, Lomotan EA, Mardon RE, Michaels M. Creating implementable clinical practice guidelines: the 2020 focused updates to the national heart, lung, and blood institute's asthma management guidelines. Implement Sci Commun 2023;4(1):36 [FREE Full text] [doi: 10.1186/s43058-023-00417-3] [Medline: 37003961]
31. Cloutier MM, Wakefield DB, Carlisle PS, Bailit HL, Hall CB. The effect of easy breathing on asthma management and knowledge. Arch Pediatr Adolesc Med 2002;156(10):1045-1051. [doi: 10.1001/archpedi.156.10.1045] [Medline: 12361453]
32. Matui P, Wyatt JC, Pinnock H, Sheikh A, McLean S. Computer decision support systems for asthma: a systematic review. NPJ Prim Care Respir Med 2014;24:14005 [FREE Full text] [doi: 10.1038/npjpcrm.2014.5] [Medline: 24841952]
33. Beasley R, Harrison T, Peterson S, Gustafson P, Hamblin A, Bengtsson T, et al. Evaluation of budesonide-formoterol for maintenance and reliever therapy among patients with poorly controlled asthma: a systematic review and meta-analysis. JAMA Netw Open 2022;5(3):e220615 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.0615] [Medline: 35230437]
34. Bates DW, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. J Am Med Inform Assoc 2003;10(6):523-530 [FREE Full text] [doi: 10.1197/jamia.M1370] [Medline: 12925543]

XSL•FO
RenderX

Abbreviations

ACT: Asthma Control Test
BPA: best practice alert
CDS: clinical decision support
GINA: Global Initiative for Asthma
EHR: electronic health record
NAEPP: National Asthma Education and Prevention Program
PCP: primary care provider
SMART: single maintenance and reliever therapy

Edited by A Mavragani; submitted 28.08.24; peer-reviewed by J Walsh, D Drummond; comments to author 01.12.24; revised version received 16.12.24; accepted 02.01.25; published 18.03.25.

Please cite as:
Fedele DA, Ray JM, Mallela JL, Bian J, Chen A, Qin X, Salloum RG, Kelly M, Gurka MJ, Hollenbach J
Development of a Clinical Decision Support Tool to Implement Asthma Management Guidelines in Pediatric Primary Care: Qualitative Study
JMIR Form Res 2025;9:e65794
URL: https://formative.jmir.org/2025/1/e65794
doi:10.2196/65794
PMID:

©David A Fedele, Jessica M Ray, Jaya L Mallela, Jiang Bian, Aokun Chen, Xiao Qin, Ramzi G Salloum, Maria Kelly, Matthew J Gurka, Jessica Hollenbach. Originally published in JMIR Formative Research (https://formative.jmir.org), 18.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Designing a Digital Intervention to Increase Human Milk Feeding Among Black Mothers: Qualitative Study of Acceptability and Preferences

Loral Patchen[1], CNM, IBCLC, MA, MSN, PhD; Jeannette Tsuei[2], MPhil; Donna Sherard[3], MPH; Patricia Moriarty[1], CCRC, MSHS; Zoe Mungai-Barris[4], BA; Tony Ma[5], MS; Elina Bajracharya[5], MS; Katie Chang[5], MS; William Douglas Evans[6], PHD

[1]Healthcare Delivery Research, MedStar Health Research Institute, Columbia, MD, United States

[2]RAND, Santa Monica, CA, United States

[3]Changeable, Silver Spring, MD, United States

[4]Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

[5]Benten Technologies, Manassas, VA, United States

[6]Department of Prevention and Community Health & Global Health, Milken Institute School of Public Health, George Washington University, Washington, DC, United States

Corresponding Author:
Loral Patchen, CNM, IBCLC, MA, MSN, PhD
Healthcare Delivery Research
MedStar Health Research Institute
10980 Grantchester Way, 7th Floor
Columbia, MD, 21044
United States
Phone: 1 7205845076
Email: loral.patchen@medstar.net

Abstract

Background: Breastfeeding rates among US mothers, particularly Black or African American mothers, fall short of recommended guidelines. Despite the benefits of human milk, only 24.9% of all infants receive human milk exclusively at 6 months.

Objective: Our team previously explored the key content areas a mobile health intervention should address and the usability of an initial prototype of the Knowledge and Usage of Lactation using Education and Advice from Support Network (KULEA-NET), an evidence-based mobile breastfeeding app guided by preferences of Black or African American parents. This study aimed to identify the preferences and acceptability of additional features, content, and delivery methods for an expanded KULEA-NET app. Key social branding elements were defined to guide app development as a trusted adviser. The study also aimed to validate previous findings regarding approaches to supporting breastfeeding goals and cultural tailoring.

Methods: We conducted a qualitative study using in-depth interviews and focus groups with potential KULEA-NET users. A health branding approach provided a theoretical framework. We recruited 24 participants across 12 interviews and 2 focus groups, each with 6 participants. The Data methods aligned with qualitative research principles and concluded once saturation was reached. Given the focus on cultural tailoring, team members who shared social identities with study participants completed data collection and coding. Two additional team members, 1 with expertise in social branding and 1 certified in lactation, participated in the thematic analysis.

Results: All participants identified as Black or African American mothers, and most interview participants (7/12, 58%) engaged in exclusive breastfeeding. In total, 4 themes were recognized. First, participants identified desired content, specifying peer support, facilitated access to experts, geolocation to identify resources, and tracking functions. Second, delivery of content differentiated platforms and messaging modality. Third, functionality and features were identified as key factors, highlighting content diversity, ease of use, credibility, and interactivity. Finally, appealing aspects of messaging to shape a social brand highlighted support and affirmation, inclusivity and body positivity, maternal inspiration, maternal identity, social norms, and barriers to alignment with aspirational maternal behaviors as essential qualities. Crosscutting elements of themes included a desire to communicate with other mothers in web-based forums and internet-based or in-person support groups to help balance the ideal

XSL•FO
RenderX

medical recommendations for infant feeding with the contextual realities and motivations of mothers. Participants assigned high value to personalization and emphasized a need to achieve both social and factual credibility.

Conclusions: This formative research suggested additional elements for an expanded KULEA-NET app that would be beneficial and desired. The health branding approach to establish KULEA-NET as a trusted adviser is appealing and acceptable to users. Next steps include developing full app functionality that reflects these findings and then testing the updated KULEA-NET edition in a randomized controlled trial.

(JMIR Form Res 2025;9:e67284) doi:10.2196/67284

KEYWORDS

health equity; breastfeeding; qualitative; mobile health; black mothers; preferences; cultural tailoring; mobile phone

Introduction

Background

Breastfeeding has been shown to have numerous health benefits for infants and mothers, yet rates of breastfeeding among US mothers, particularly Black or African American mothers, fall short of recommended guidelines. The American Academy of Pediatrics [1], American College of Obstetricians and Gynecologists [2], and the World Health Organization [3] all recommend exclusive breastfeeding or exclusive human milk feeding (EHMF) for approximately 6 months after birth. Breastfeeding has been shown to be effective in reducing overall infant mortality, gastrointestinal tract and respiratory infections, sudden infant death syndrome, otitis media, and childhood obesity [4]. For mothers, breastfeeding has been associated with postpartum weight loss and a reduction in the risk of diabetes, cardiovascular disease, reproductive cancers, and postpartum depression [5,6].

Despite these benefits, only 24.9% of infants receive breast milk exclusively at 6 months [7]. Racial and ethnic disparities also exist in breastfeeding rates—non-Hispanic Black women are less likely to initiate breastfeeding and breastfeed for at least 6 months compared to the US national rate [8]. The American College of Obstetricians and Gynecologists emphasizes the importance of breastfeeding for underserved women due to the increased prevalence of adverse health outcomes among underserved groups, such as obesity, diabetes, and cardiovascular disease. Improving breastfeeding among Black or African American women not only improves short-term outcomes but also has the potential to reduce other long-term health disparities.

Barriers to initiation and continuation of breastfeeding can include structural, cultural, and societal factors, such as the need to return to work, lack of community support, or stigma around breastfeeding in public. For Black or African American women, reasons for low breastfeeding rates include not wanting to breastfeed, lack of breastfeeding information and education, having to return to work or school, lack of familial support, and being located in neighborhoods with hospitals that are less likely to implement practices that support breastfeeding [9]. Quintero et al [8] found that racial and ethnic disparities exist related to the role of breastfeeding information on breastfeeding initiation and duration. Black or African American women were less likely to receive information from their friends and family than other racial or ethnic groups.

Technology-based interventions have been shown to increase breastfeeding initiation and duration. A global meta-analysis of 15 randomized controlled trials found that technology-based interventions substantially increased the exclusive breastfeeding rate 6 months after delivery compared to usual care [10]. Mobile health (mHealth) technologies allow for greater access and dissemination of information at the exact time of need [11-13]. mHealth interventions have included a variety of modalities, including telephone support, televideo lactation consultations, SMS text messaging support, breastfeeding monitoring, and education. Several mHealth interventions have focused on the important role of peer support in increasing positive breastfeeding experiences, social connections, and a sense of belonging and self-efficacy [14]. Other mHealth technologies include location-based features, such as FeedFinder, a smartphone app using a GPS to support women in finding, reviewing, and sharing public breastfeeding places with others [15]. The My Baby Now app, which provides evidence-based information and support on infant feeding and active play from pregnancy to 18 months, found an increase in breastfeeding knowledge and EHMF intentions among Australian mothers. The authors also noted that mothers with lower levels of education found the app to be more helpful and rated it more highly [16].

Few studies have focused on the use of mHealth for breastfeeding in underserved populations. The Lactation Advice Through Texting Can Help trial connected low-income participants to breastfeeding peer counselors via SMS text messaging but did not find an impact on EHMF rates [17]. Notably, Lactation Advice Through Texting Can Help relied on text messages only and did not include additional features. Furthermore, the authors did not describe specific efforts to tailor messages to participants. A study of Breastfeeding Friend, an app with interactive advice, breastfeeding strategies and recommendations, educational content, and links to resources, also found no difference in breastfeeding rates among low-income, first-time mothers compared to a control app with digitized handouts. The authors found that women were receptive to receiving educational content through apps, but they also preferred in-person breastfeeding support, which could be provided through real-time chat functionality or SMS text messaging with experts [18]. While this intervention was designed with the intent to serve low-resource users and it delivered content at the fifth-grade reading level, it did not specify attention to cultural alignment or other aspects of social identity. Additional research is needed on culturally tailored

XSL•FO
RenderX

mHealth interventions that address the barriers to breastfeeding of Black or African American women.

In our previous study, our team explored key content areas and usability of an initial prototype of the Knowledge and Usage of Lactation using Education and Advice from Support Network (KULEA-NET), an evidence-based mHealth breastfeeding intervention guided by the preferences of Black or African American parents [19]. Participants appreciated tips on returning to work, feed and diaper logs, information on infant growth and development, SMS text messaging with a lactation consultant, and a GPS map of breastfeeding-friendly locations. After the initial study was completed, the development team incorporated additional features based on these preliminary findings and supported by other research. Additional features included *social support* (eg, videos with relatable Black or African American women and family members being supportive of breastfeeding), *emotional support* (eg, personalized text messages, communication between the mother and her support individual, and educational entertainment and testimonial videos), *informational support* (eg, knowledge base), *instrumental support* (eg, breastfeeding-friendly locations, breastfeeding classes, and access to lactation support), and *appraisal* (eg, feeding and diaper logs with personalized feedback) [20-23].

Objectives

The study aimed to assess the feasibility and acceptability of additional content and its delivery in the refined KULEA-NET app and to validate previous findings in supporting breastfeeding goals. The authors have adopted the more inclusive language of human milk feeding to report study findings.

Methods

Design

We conducted a qualitative study using in-depth interviews and focus groups to understand the preferences of potential users and assess the acceptability of KULEA-NET. A health branding approach provided a theoretical framework for this study [24]. This approach appeals to an individual's self-interest and emphasizes the cost and benefit tradeoffs of a health-promoting behavior in the hope that the individual will adopt that behavior [25]. We conceptualized the KULEA-NET brand to represent mental schema that help new Black or African American mothers encompass the benefits of breastfeeding. We designed the app to offer trustworthy advice and engaging social support, which would lead to adoption and continued use. This schema captures the benefit that outweighs the costs of breastfeeding and helps new mothers succeed in consistent breastfeeding. This approach was based on the work of Evans et al [26] and others that suggest "[h]ealth branding determines behavioral choice by building consumer relationships and identification with health behaviors and their benefits." As the nature of this study was exploratory, we used qualitative methods to gain insight into content and features that would appeal to potential users, preferred content delivery methods, and feedback on app content and prototypes. The combination of in-depth interviews and focus groups provided not only depth and breadth on our study topics but also identification of individual and contextual circumstances [27] surrounding breastfeeding.

Ethical Considerations

This study was reviewed by the Institutional Review Board of the MedStar Health Research Institute and approved as human participant research (IRB#00005929) on March 9, 2023. Informed consent was obtained from all study participants before data collection. All study data are deidentified, and no participant identifiers are included. Study participants were compensated with an electronic US $50 gift card for the in-depth interviews and an electronic US $75 gift card for the focus groups.

Setting and Relevant Context

This study was conducted in the United States at an economically and culturally diverse urban hospital with an academic research center and medical training programs. The hospital supports the largest number of births among all health systems in the city, >3000 births in 2022. The hospital primarily serves patients from its surrounding urban community, which was predominately White (36.3%) and Black or African American (43.5%) in 2022 [28]. In the United States, substantial racial or ethnic disparities exist in breastfeeding initiation rates. Compared to the national average, non-Hispanic Black mothers are less likely to initiate breastfeeding (73.6% vs 84.1%) [29]. In terms of resources, the Center for Disease Control and Prevention's 2022 Breastfeeding Report Card rated this urban area as having enacted paid family and medical leave legislation and having state licensing regulations for childcare centers that were fully aligned with the standard of encouraging and supporting breastfeeding by planning for mothers to feed on-site [7]. Despite these positive indicators, other barriers to breastfeeding remain, such as breastfeeding education, familial and social support, intergenerational and historical trauma, and misinformation [8].

Recruitment Procedures

Participants were invited to participate in the study using both active and passive recruitment approaches. A study research coordinator for women's health studies approached patients of ambulatory health care offices specializing in obstetrics and gynecology regarding the opportunity to participate in ongoing research initiatives. This location provides a comprehensive maternal-infant program that includes a full scope of services from health care to legal, behavioral health, and social services [30]. If interested in this study, eligibility was established and enrollment completed as appropriate. In addition, lactation educators were aware of the study and referred interested patients to the study research coordinator, who determined interest, eligibility, and completed enrollment as appropriate. Finally, information about the study was shared electronically with patients through a pregnancy and postpartum care app provided to them by the health care setting. Interested individuals contacted the study research coordinator directly regarding eligibility and enrollment.

Sample

For both in-depth interviews and focus groups, our desired population was Black or African American parents aged between 18 and 34 years who were English speaking, intended to continue breastfeeding after hospital discharge, owned a

XSL•FO
RenderX

smartphone with internet access, and were willing to provide verbal informed consent. We excluded parents of infants with conditions that required admission to the neonatal intensive care unit or interfered with breastfeeding (specific conditions are listed in Textbox 1).

Textbox 1. Exclusion criteria for this qualitative study on support for successful human milk feeding among Black or African American mothers, highlighting medical contraindications to breastfeeding.

Study exclusion criteria

- Infants born with cleft lip or palate, congenital heart defects, Down syndrome, neural tube defects, or other conditions that require admission to a neonatal intensive care unit or interfere with breastfeeding
- HIV
- Untreated, active tuberculosis
- T-cell lymphotropic virus type I or type II
- Illicit drug use
- Receiving radiation therapy
- Exposed to anthrax
- Undergone breast surgery
- Active hepatitis B or C
- Prescription drug use incompatible with lactation

We used a purposive criterion sampling method to recruit eligible patients who did not have contraindications for breastfeeding. Potential participants were contacted by the study team during a prenatal or postpartum appointment. Recruitment flyers were also posted at the center to recruit participants for the focus groups and electronically delivered to patients through a push notification associated with a remote blood pressure monitoring app used by patients receiving care at this center. In addition, snowball sampling was used to identify other potential participants that met the target population criteria. Participants were compensated with an electronic US $50 gift card for the in-depth interviews and a US $75 gift card for the focus groups.

We aimed to recruit 30 participants across the interviews and focus groups based on previous literature that recommends at least 12 participants to reach thematic saturation [31,32]. Recruitment continued until data saturation was reached on nearly all topics and no new themes emerged. In total, we recruited 24 participants across 12 in-depth interviews and 2 focus groups with 6 participants each. Due to this research's exploratory nature and the relatively homogenous nature of our target population, our sample size provided considerable data and was found to be adequate following qualitative research principles [33].

Data Collection

All in-depth interviews and focus groups took place on the internet via Microsoft Teams from June 1, 2023, to March 26, 2024. Recruitment occurred between June 1, 2023, and March 25, 2024. All recordings were transcribed and deleted once transcription was completed. Confidentiality was protected by assigning each participant an unidentifiable code and reporting only aggregated sample information and anonymized quotations. The study team consisted of the principal investigators (ie, multiple-principal investigator team), coinvestigators, app development technologist, study coordinator, research assistant, research nurse, and the community leadership council. Interviews lasted between 30 minutes and 1 hour, while focus groups lasted 1 hour. Team members who shared racially concordant social identities conducted data collection.

Verbal consent was obtained from every participant at the start of the interview or focus group. Before the interview or focus group, the study team provided participants with the opportunity to have any questions answered before obtaining consent. We ensured the consenting process occurred in a private location selected by the participant where other people were not present and the study team's private office. It was recorded via Microsoft Teams.

A semi-structured interview guide and discussion guide were used for the interviews and focus groups, respectively. For the interviews, topics focused on media and information sources used during pregnancy and motherhood and potential app content and features. The topics of the focus groups centered on identity as a mother, preferences for receiving support, and desired content and features for the app.

Data Analysis

Overview

The study team used deductive thematic analysis [34] to identify and interpret patterns or themes in the interview and focus group transcripts. The primary goal of the analysis was to identify themes and use them to inform the development of the content and functionality of KULEA-NET. The research team reviewed each transcript and developed a list of initial codes relevant to the research questions using the topics from the discussion guides as a starting point. An iterative process was used where codes were compared and discussed at regular team meetings during data collection. The coding dictionary was developed with each code having a distinct and mutually exclusive definition. Two team members coded the transcripts, compared codes, discussed any discrepancies, and modified the codes after

XSL•FO
RenderX

obtaining agreement. To develop the themes, 2 additional research team members collaborated with those who coded the data, reviewing the codes and grouping them into meaning or a pattern that answered the research questions. A summary of the thematic analysis is presented in Table 1. All coding and analysis were performed by the coders, and no data analysis software was used.

Table 1. Thematic analysis of responses from Black or African American mothers who participated in focus groups and individual in-depth interviews identified 4 themes: desired content, preferred modes of content delivery, requested functionality and features, and appealing aspects of messaging to shape a social brand for the app.

Theme, theme definition, and category	Category definition
Desired content and features	
Resources, information, and features participants sought and found helpful	
Peer support	Connection and communication with other mothers
Policy and advocacy support	Knowledge of mother's legal rights and guidance on tools to advocate for oneself
Access to experts	Access to lactation consultants, doulas, or health care providers through the app
Location-based resources	Geolocation functionality to identify helpful resources
Tracking and planning	Documenting the progress of pregnancy or breastfeeding and planning for exclusive breastfeeding
Content delivery	
Method in which content is delivered to users	
Device type	Preferred devices or platforms
Messaging	Preferred messaging modality—text messaging versus notifications
Desired app characteristics	
Preferences, requirements, or functionality participants requested	
Content diversity	Variety and range of content presented to app users
Ease of use	Capacity of an app to be used easily and effectively
Credibility	Trustworthy, reliable, and accurate information or sources in the app
Interactivity	Interaction between the user and the app that allows for real-time engagement, user input, and responsive feedback
Messaging	
Communication style, tone, and content	
Support and affirmation	Emotional support, tangible support, or validation from partners, family, peers, and community
Inclusivity and body positivity	Acceptance of all feeding modalities, breastfeeding decisions, and comfort with the body in relation to breastfeeding
Maternal inspiration	Positive qualities and characteristics related to motherhood that women aim to embody
Maternal identity	The significance mothers attribute to motherhood and how they define their role as mothers
Social norms	Perceptions and attitudes on breastfeeding among family, friends, or community
Barriers to alignment with aspirational maternal behaviors	Challenges to fulfilling maternal behaviors that are part of a maternal identity

Reflexivity

Our team participated in a reflexive process by taking notes throughout coding and consideration of thematic analysis. We used these to guide reflection on experiences related to infant feeding and considered how lived experiences informed coding and interpretation. Given the focus on cultural alignment, team members with racially concordant social identities completed data collection, coding, as well as thematic analyses related to cultural tailoring. They reflected on how their own support systems and cultural diversity within racial and ethnic communities are connected to the process of information gathering and interpretation. Team members with social branding and lactation expertise joined analysis of social branding and clinical themes only. As a group, we considered how to share these narratives with authenticity and sensitivity. A debriefing of coding and analysis was conducted with the study's community leadership group to critically assess findings and interpretations. This review assisted the research team with identifying potential biases and validating interpretations.

XSL•FO
RenderX

Results

Characteristics of the Sample

Characteristics of the participants (N=24) are detailed in Table 2. Fewer identifying details were asked of focus group participants, given the group setting.

Table 2. All study participants identified as Black or African American women. Details of breastfeeding history were asked during individual interviews, and approximately one-third shared a history of practicing only human milk feeding with their infants.

Sample characteristics	Interviews (n=12), n (%)	Focus groups (n=12), n (%)
Identified as Black	12 (100)	12 (100)
Identified as women	12 (100)	12 (100)
Pregnancy status		
Currently pregnant	6 (50)	0 (0)
Not currently pregnant	6 (50)	12 (100)
Primiparous	2 (33)	N/A[a]
Multiparous	4 (67)	N/A
EHMF[b]	7 (58)	N/A
Did not practice EHMF with first child but started EHMF with recent child	4 (36)	N/A
Currently, EHMF	2 (18)	N/A
Plan to EHMF	5 (45)	N/A
Has EHMF for >6 months	4 (36)	N/A
Attempted EHMF in past but unsuccessful	3 (25)	N/A

[a]N/A: not applicable.

[b]EHMF: exclusive human milk feeding. Study participants were able to select more than one option, percents do not sum to 100%.

Themes

Among the 12 interviews and 2 focus groups, four themes emerged that were related to KULEA-NET: (1) desired content and features, (2) content delivery, (3) desired characteristics, and (4) messaging.

Desired Content and Features

Peer Support

Participants described a desire to communicate with other mothers in web-based forums and internet-based or in-person support groups. They wanted to interact with mothers at different stages, such as first-time and more experienced moms, to learn about their experiences and ask questions:

> *I like seeing, you know, like a thread, a question and people giving their opinion their advice, their tips, how they dealt with that similar situation.*

Participants referenced their appreciation of the peer support features of existing apps (eg, Peanut, What to Expect, and My Pregnancy) and websites (eg, Postpartum Support International and The Feeding Flock) they have used. Another participant described the benefits of community and not feeling alone:

> *I like group settings as well because I feel like when [...] another person has asked a question that you have then it's almost like a sense of belonging because you don't feel like you're going crazy trying to figure something out. [...] You feel like a sense of like, ohh yes, we're all in this together.*

Some participants also mentioned the ability of peer support to fill in gaps in support from providers:

> *I feel like I've been gaining more information and support from like these other outside support groups more than I've been feeling supported by my provider, which is kind of sad.*

Policy and Advocacy Support

Some participants described a desire to know their rights as mothers and for tools to advocate those rights to employers. Participants wanted to know the legality of breastfeeding in public, their rights for pumping and milk storage at work, and their rights for work breaks as pregnant mothers. One participant envisioned links to credible state-based resources:

> *So if they have an app that tells them "ok, you are pregnant, this is your right." Not just say this is your right, give them an external link to like resources they can use if they don't want to rub their boss the wrong way or whatever. When you know it is your legal right, you won't have an issue with it and you will have something to back it up because you read it and you can go talk to HR.*

Access to Experts

Among existing apps, participants expressed that they liked the ability to connect with experts, such as lactation consultants, doulas, or providers. Multiple participants cited the brief time they had with a lactation consultant in the hospital after birth as insufficient to meet their needs. Others discussed how they

XSL•FO
RenderX

did not feel listened to by their providers and wanted access to expert-based information:

> *I didn't get a chance to ask all of [my] questions, or if I did, they were kind of rushed or the answer wasn't like a real answer.*

Participants found it valuable to connect with lactation specialists during extended hours, such as when they are the only ones awake with the baby, via video calls or SMS text messaging. One participant mentioned her experience using the video chat feature of an existing app:

> *I used it like at 2:00 in the morning because I was, you know, desperate because I had an issue with my breast. But anyway, so it was helpful during the video call because the lactation consultant she was able to like, you know, see my breast and see like, what was the issue and provide like, you know, real time suggestions like what I can do in my case.*

Location-Based Resources

Geolocation-based features that shared breastfeeding-friendly locations were well received by participants who evaluated the KULEA-NET prototypes:

> *I like it. Instead of trying to go on Google trying to figure out places where to go, at least you have, you know, the place pointed out to you where you can go.*

Participants suggested adding women, infants, and children locations and places that offer free baby supplies to this feature:

> *I was gonna say locations where you can go that gives free diapers.*

Tracking and Planning

Participants wanted the ability to track pregnancy and breastfeeding indicators (eg, size of baby, amount of milk pumped, and feeding times). In addition to tracking, planning for EHMF was a critical contributor to EHMF behavior. One participant discussed switching to EHMF with her second child due to planning:

> *I had looked into getting an IBCLC [International Board Certified Lactation Consultant] to help me and things like that - resources I didn't know the first time, I was seeking them out in advance. Just to make sure I had an arsenal of things to help me so I wouldn't quit.*

Participants envisioned the app as a tool to help them create a plan, integrate it into their daily lives, and receive guidance throughout the planning process.

Content Delivery

Device Type

Preference for an app that could be accessed via a smartphone or tablet was preferred over a laptop, as participants found those easier to use when holding a baby, more lightweight, and accessible while on the go:

> *When you're about to be a mom, your phone is your best friend.*

Type, Frequency, and Choice of Messages and Notifications

Participants were divided regarding how they preferred to receive information from KULEA-NET. Some preferred texts, particularly if they came from the same number, while others preferred notifications. The appropriate amount of text messages or notifications ranged from 1 to 3 per week. Daily messages or notifications were thought to be too intrusive and overwhelming:

> *Too many texts can be annoying, aggressive, and spam-like.*

Several participants also wanted a choice in either opting in or out of messages and notifications or selecting the frequency.

Desired Characteristics

In addition to the desired content and features, participants provided input on the characteristics KULEA-NET should exhibit.

Content Diversity

While KULEA-NET is designed for Black or African American women, participants recommended that the app approach cultural specificity with sensitivity to avoid the potential othering of social identities:

> *Because I feel like within the media, it's always portrayed as though, umm, it's more of a Caucasian thing to breastfeed, and maybe it's more popular, so to speak, in that culture, I guess, but I feel like there's plenty of black women that breastfeed celebrities that breastfeed.*

Participants suggest materials include images with a diverse range of women and not only Black and White women:

> *I like it overall, I'm but one thing I would mention is it needs to be more diversity.*

Moreover, participants did not want an app that only addressed breastfeeding but rather provided support throughout the prenatal to postnatal phases:

> *We have app fatigue—we need one that can transition through different stages of motherhood.*

One participant discussed how a comprehensive app would reduce app fatigue:

> *So it can also be like a lot to continuously hop from app to app to app and so the more information that you can have in one place I think is the better. the more that you can get in one app the better.*

Ease of Use

Information in the app should be easy to navigate, easy to read, and presented in various formats (ie, text, images, video, and audio) to appeal to various learning styles:

> *I'm just a visual person, so I like being able to see the person, it seems more real and personal.*

One participant explained that information-heavy apps were overwhelming:

> *How do I say, sometimes there are apps with too much stuff and then it doesn't feel like it's helpful.*

XSL•FO
RenderX

Credibility

Many participants commented on the need to emphasize the credibility of the content in the app by detailing sources, citations, or academic affiliations. Participants mentioned past experiences of receiving conflicting information from web-based sources, making them cautious of internet-based information:

> *A lot of the apps contradict themselves like "oh don't give into your cravings" and then in the third trimester they're like "eat whatever you want." What am I supposed to be listening to?*

Emphasis on credibility would make it easier and less time-consuming for participants to discern the validity of the information. While forums with peers were desirable, participants also expressed the need for a provider to moderate forums to reduce the potential of misinformation.

Interactivity

Video-based visits with lactation consultants supplemented by video examples of breastfeeding techniques were suggested. Participants wanted to share a live, visual display of the breastfeeding process, such as latching, with a lactation consultant to obtain effective support:

> *Phone or video will be most helpful so the consultant can see what you're doing and give real time advice, or at least a phone, because you might be not hands free, but at least on the phone you can put on speaker and speak that way.*

Messaging

Support and Affirmation

Participants appreciated empathetic support from their partners, parents, and children, which took the pressure off being a "perfect mom." They also appreciated partners who took up responsibilities when mothers needed rest:

> *We [my partner and I] have expectations that we don't set the bar too high. We know that in life there are various struggles and imperfections and he keeps me grounded when I feel like I have to be perfect as like a perfect mom and the perfect life.*

Some participants sought comfort in seeing other mothers struggling in similar ways or found breastfeeding inspiration on social media. One participant described the encouraging influence social media had on her decision to exclusively breastfeed. She said that seeing other people doing it made her believe that she could do it too. Referring to her Facebook group, 1 participant shared the following:

> *Yeah, because honestly, that is one of the things that got me through was the other members, you know, somebody who's been there, done that not just on uh educated level, but you know, somebody in my shoes.*

Many participants emphasized that the app should have positive reinforcement and words of affirmation, rather than a punitive tone. Pregnancy affirmations delivered by the app helped prepare participants mentally:

> *I like the daily motivational quote. Basically saying you're doing a great job, which we often need to hear as moms, even if we feel like we're not doing a good job. So I think that's great.*

Participants did not want to be pressured or judged for their breastfeeding decisions and behaviors:

> *I don't want it to make it seem like oh if you breastfeed, you are successful, but if you don't breastfeed that you're not successful....So just not to limit just to exclusive breastfeed[ing], like anyone could be successful.*

One participant described that she does not want her partner's opinions about what she is doing with her body. She considers breastfeeding a decision that she makes about her body and wants his support whether or not she breastfeeds:

> *I mean that's great [information for partners], but I don't want to hear my partner's opinion about what I'm doing to my body. It's not a driving factor to know how my partner feels about what I'm choosing to do.*

Inclusivity and Body Positivity

Many participants expressed the need for the app to include messaging that was inclusive of all feeding modalities. The app should indicate respect for all infant feeding choices, as some mothers are unable to breastfeed due to medical issues:

> *...I don't want it to seem like it's very exclusive to like breastfeed[ing] I feel like it should be open to anyone because there are different, you know, different circumstances and situations in which a mother can't provide breastfeeding, you know?*

A couple of participants expressed regret or guilt for not breastfeeding and wanted to avoid those negative emotions:

> *I wanted to [breastfeed] because I figured it's holistic, you know. That's how, you know, my son would have been able to fight off like sicknesses and things. So, for me not to do it, it was kind of hurtful.*

Body positivity was also found to be an important element of breastfeeding for participants. One participant explained how she feels more comfortable with her body with her second baby compared to her first because her breasts feel less "sexualized" and now symbolize a food source for her baby:

> *Definitely feeling more comfortable with my skin as far as I look as my boobs as nutrients you know. I look at it as milk, as food for the baby. I don't look at it as sexualized you know.*

Maternal Inspiration

When asked for examples of inspiration or influence, participants highlighted public figures who were able to balance their personal lives, careers, and other activities outside of motherhood:

> *Just how she balanced both working and taking care of three kids as well as, you know, still being a wife and still having her individual life going on.*

XSL•FO
RenderX

Others expressed admiration for celebrity or influencer mothers who are physically active and prioritize nutrition as mothers:

> *Ciara is very much active, and she was very active during her pregnancy, so that was a lot of inspiration for me to be active throughout my pregnancy.*

Single mothers in participants' personal lives and in popular culture were seen as displaying strength and resilience:

> *She's going through some things that an average person, you know that maybe somebody else would have broken down, but she continues to push through for her girls.*

Maternal Identity

Participants highlighted providing emotional safety, a better relative childhood, and trust in their motherly instinct to be key qualities of a good, capable mother. Creating a safe space and promoting confidence were important goals:

> *I could be a good, safe space for the child, emotionally help them grow into being a productive member of society. But then also help them see that they're uniquely gifted and uniquely made. And to help like, bring and call those things out so that they can feel like a confident person.*

Participants had a desire to provide a better childhood for their children than their own:

> *You know, being mindful of choices that we've made or that were made for us that we didn't necessarily agree with as we got older, you know, it takes that intentionality [and] awareness to say, you know, I'm gonna do this differently.*

Focusing on their child's needs rather than parenting norms and trusting their own "motherly instinct" sometimes over medical advice was valued:

> *The doctors may say one thing, right? You as a mom know what's best for your baby.*

Social Norms

Decisions on breastfeeding were often informed by practices among peers, family members, or a participant's broader community. Seeing sisters, mothers, and aunts breastfeed provided examples of normative behavior that influenced participants' breastfeeding decisions. One participant described her decision as "more so what I saw from my community." Breastfeeding was also associated with challenges to traditional norms and the expectations of support systems:

> *Family members were a barrier for breastfeeding in the sense that it is almost not possible for family to see her the way they want to because of the breastfeeding...They're used to bottle feeding with formula, so they're like, give me your baby, go sleep overnight. Just trying to not discourage me intentionally from breastfeeding, but just in the sense that they wanted to help. So they're like, "Why, it's so, so difficult when you're breastfeeding. Don't [you] wanna try formula?" [It's] not in a bad way. I know they were just trying to help, but that definitely was a hurdle for me.*

Barriers to Alignment With Aspirational Maternal Behaviors

Mental health challenges, including stress and anxiety, were described as important barriers to being a "good mom." One participant described how exhaustion can interfere with her intentions:

> *Mental exhaustion, like just being mentally tired and not wanting to do anything, just wanting to relax...When I may have other things that I could be doing or need to be doing.*

Discussion

Multiplicity of Maternal Identity

Using a health branding framework, we initially hypothesized that a targeted intervention leveraging audience segmentation based on sociodemographic characteristics would encourage intended behavior change [24,25]. However, our previous study revealed mixed responses to a culturally tailored intervention featuring Black or African American imagery. Participants informed us that providing for one's child was a universal desire that should not require race- or ethnicity-specific content [19]. In this study, we delved into participants' perceptions of maternal identity to better inform our communication approach. Our findings reiterated commonalities in maternal identity, particularly shared motivations of fostering emotional safety and providing a better childhood. Furthermore, sources of maternal inspiration highlighted were not exclusively Black or African American and were mentioned for a variety of desirable characteristics, such as success in balancing the multiple components of mothers' lives. Racial identity is merely one aspect of maternal identity among Black or African American mothers. Effective role models must speak to the multidimensional nature of motherhood, particularly the skillful managing of these diverse identities.

Previous research has shown the importance of avoiding the uniform treatment of sociodemographic groups, instead recognizing the multiple identities and intersecting factors that exist within them. A systematic review of mHealth apps among populations considered vulnerable, specifically low-socioeconomic status and historically marginalized racial and ethnic groups, found few improvements in health outcomes. The review underscored the need for further research that differentiates between broad categories of low-socioeconomic status and racial and ethnic minority groups to understand how mHealth tools can effectively address a range of needs [35]. Our mHealth intervention must appeal to the varied needs and lived experiences of mothers by incorporating features that consider sociocultural norms, environmental factors, household composition, resources, and availability of breastfeeding role models. KULEA-NET includes features informed by the barriers Black or African American mothers may face, such as breastfeeding education and tips, a GPS-based map of breastfeeding-friendly spaces, and feeding and diaper change logs. The delivery of this content will address the multiplicity

XSL•FO
RenderX

of identities among Black or African American mothers, which includes using imagery and messaging that speaks to a range of current and aspirational identities and helps them feel represented.

Nonoppositional Messaging

While EHMF for 6 months is recommended, our findings revealed that mothers wanted a more inclusive approach to breastfeeding—one that supports more options to accommodate the realities of a mother's circumstances and available resources. A recent study evaluated the "breast is best" narrative and noted a similar finding where mothers sought a more nuanced view toward breastfeeding and formula feeding to recognize a diversity of lived experiences and avoid inducing feelings of shame or failure when mothers were unable to EHMF [36].

In social marketing and health branding, oppositional and nonoppositional strategies are commonly used. Nonoppositional strategies aim to promote behavior change without directly challenging existing beliefs or practices (eg, eating 5 servings of fruits and vegetables), while oppositional strategies confront or compete with unhealthy messages (eg, antitobacco or antismoking messages) [25]. Our participants discussed workplace challenges to breastfeeding and the desire to know their rights or have tools to advocate for breastfeeding at work. Although mothers may intend to exclusively breastfeed, they may lack the means to achieve this goal. Employment before 6 months without adequate breaks and space for pumping has been shown to be a key barrier to EHMF [37,38]. Asiodu et al [39] emphasized the need to discuss combination feeding openly and alleviate some of the emotional burden and stress associated with not meeting expectations or societal pressures of EHMF. The tone and messaging of KULEA-NET must take a nonoppositional approach and acknowledge the circumstances that influence the ability of women to breastfeed. It will aim to provide positive reinforcement and encouragement to support mothers on their breastfeeding journey, rather than generate guilt or shame.

Critical Role of Support

Various forms of tangible and emotional support emerged as key features participants desired in KULEA-NET. Tangible support included peer support, partner support, access to experts, and assistance in navigating breastfeeding rights with employers. In addition, mothers expressed a strong need for emotional support to cope with the pressures of motherhood. Participants voiced the importance of nonjudgmental acceptance and encouragement of a mother's choices.

Access to breastfeeding support has been identified as an important mediator, accounting for two-thirds of the Black-White disparity in breastfeeding outcomes [9]. Influential supportive figures include not only providers and community-based experts but also family members, partners, friends, peers, church members, and even strangers. Moreover, Black or African American women are most likely to initiate and continue breastfeeding if they receive adequate support (ie, assistance with lactation, domestic activities, childcare, and groceries) from these sources during the first 2 weeks post partum [40]. Support from these various sources can help mothers overcome barriers to EHMF.

In this study, participants discussed examples of tangible and emotional support. Intimate relationships, such as partners, mothers, aunts, and grandmothers, not only shared some of the burdens of motherhood but also served as role models and fostered a normative culture of breastfeeding. Peers allowed mothers to learn from others' experiences, ask questions, and fill in when providers did not provide sufficient breastfeeding support. Peer support could also be internet-based, through forums or social media, in providing a sense of belonging, connection, and empowerment. Gyamfi et al [40] stressed the importance of exploring all avenues, including internet-based platforms, to establish culturally sensitive breastfeeding groups among peers in the Black or African American community to promote breastfeeding. Peer support features within an mHealth app can provide emotional support and community engagement and address a variety of unique needs in the Black or African American community.

KULEA-NET aims to integrate these findings by offering an app that recognizes the diverse lived experiences of Black or African American women, addresses both EHMF and combination feeding, and provides social support to improve breastfeeding outcomes. Future research will evaluate the efficacy of the fully developed version of this mHealth intervention in increasing breastfeeding rates through a randomized controlled trial. Findings of that trial will include usability and user engagement metrics and compare outcomes of exclusive human milk feeding success rates.

Future research should examine the conceptual model underlying KULEA-NET and its focus on cultural tailoring. This would involve evaluating the pathway of change and mediating factors, such as social support and app features, and their effect on breastfeeding outcomes. In addition, further research should explore the optimal timing for introducing the app (eg, enrolling participants in a KULEA-NET promotion study before vs after delivery) as well as the dosage and frequency of use for specific features. Studies should seek to identify the mechanisms of change through which KULEA-NET works.

Limitations

While our qualitative methodology provided a range of positive and negative perspectives on the KULEA-NET app and detailed information about the experiences of Black or African American mothers, there are some limitations to note. Our inclusion criteria included a requirement that participants were willing or intended to breastfeed, which meant that our sample may have been predisposed to breastfeed and may not have been fully representative of Black or African American mothers. In addition, multiple participants worked in health care, and our sample was informed of the benefits of EHMF compared to other nursing methods, which limited the external validity of our findings. Another limitation was inconsistency in the interview questions asked of participants. Some new questions were introduced in later interviews that were not included in earlier interviews, mainly questions concerning participant knowledge of their rights as mothers. For these questions, we had a smaller sample size, resulting in less detail on this topic.

XSL•FO
RenderX

Conclusions

Formative research suggests that the KULEA-NET app is acceptable and beneficial to participants. The personalization, support, and maternal identification features stood out as key benefits provided by the app. Implementation and promotion of app use represent important next steps. Future study developments will include full app functionality and testing in a randomized controlled trial.

Acknowledgments

The authors wish to acknowledge members of the project community leadership group for their contributions to this work: Tiara Caldwell, Jashia Pinkney, Cassietta Pringle, Wendy Post, and Kymberly Sharpe. This research was funded through a Small Business Innovation Research phase I and phase II application with the goal of commercializing the Knowledge and Usage of Lactation using Education and Advice from Support Network platform.

Data Availability

Deidentified study transcripts are available upon reasonable request to the corresponding author, with an established data transfer agreement.

Conflicts of Interest

LP, WDE, JT, ZMB, DS, and PM have nothing to disclose. TM, KC, and EB are employed at Benten Technologies and declare the following potential conflicts of interest and relevant relationships to promote transparency. Benten Technologies has a commercial interest in the outcomes of this research. While Benten Technologies is involved in the commercialization of the developed technology, the qualitative study and planned future randomized controlled trial are conducted independently at clinical and academic sites. These sites have no financial ties to Benten Technologies, ensuring that the study's design, data collection, and analysis remain unbiased and solely focused on scientific integrity and patient outcomes regarding the National Institutes of Health funding and product development under its Small Business Innovation Research award.

Multimedia Appendix 1
COREQ checklist.
[PDF File (Adobe PDF File), 144 KB - formative_v9i1e67284_app1.pdf]

References

1. Meek JY, Noble L, Section on Breastfeeding. Policy statement: breastfeeding and the use of human milk. Pediatrics 2022 Jul 01;150(1):e2022057988. [doi: 10.1542/peds.2022-057988] [Medline: 35921640]
2. . Committee Opinion No. 570: breastfeeding in underserved women: increasing initiation and continuation of breastfeeding. Obstet Gynecol 2013 Aug;122(2 Pt 1):423-427. [doi: 10.1097/01.AOG.0000433008.93971.6a] [Medline: 23969829]
3. Exclusive breastfeeding for optimal growth, development and health of infants. World Health Organization. 2023 Aug 9. URL: https://www.who.int/tools/elena/interventions/exclusive-breastfeeding [accessed 2025-02-19]
4. Dieterich CM, Felice JP, O'Sullivan E, Rasmussen KM. Breastfeeding and health outcomes for the mother-infant dyad. Pediatr Clin North Am 2013 Feb;60(1):31-48 [FREE Full text] [doi: 10.1016/j.pcl.2012.09.010] [Medline: 23178059]
5. Patnode CD, Henninger ML, Senger CA, Perdue LA, Whitlock EP. Primary care interventions to support breastfeeding: updated evidence report and systematic review for the US preventive services task force. JAMA 2016 Oct 25;316(16):1694-1705. [doi: 10.1001/jama.2016.8882] [Medline: 27784101]
6. Pope CJ, Mazmanian D. Breastfeeding and postpartum depression: an overview and methodological recommendations for future research. Depress Res Treat 2016;2016:4765310. [doi: 10.1155/2016/4765310] [Medline: 27148457]
7. Breastfeeding report card. Centers for Disease Control and Prevention. 2024 Jun 11. URL: https://www.cdc.gov/breastfeeding-data/breastfeeding-report-card/index.html [accessed 2025-02-19]
8. Quintero SM, Strassle PD, Londoño Tobón A, Ponce S, Alhomsi A, Maldonado AI, et al. Race/ethnicity-specific associations between breastfeeding information source and breastfeeding rates among U.S. women. BMC Public Health 2023 Mar 17;23(1):520 [FREE Full text] [doi: 10.1186/s12889-023-15447-8] [Medline: 36932332]
9. Jiang L, Nianogo RA, Prelip M, Wang MC. The role of breastfeeding support in racial/ethnic disparities in breastfeeding practices. Prev Med 2024 Feb;179:107857. [doi: 10.1016/j.ypmed.2024.107857] [Medline: 38224744]
10. Qian J, Wu T, Lv M, Fang Z, Chen M, Zeng Z, et al. The value of mobile health in improving breastfeeding outcomes among perinatal or postpartum women: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021 Jul 16;9(7):e26098 [FREE Full text] [doi: 10.2196/26098] [Medline: 34269681]
11. Chen H, Chai Y, Dong L, Niu W, Zhang P. Effectiveness and appropriateness of mHealth interventions for maternal and child health: systematic review. JMIR Mhealth Uhealth 2018 Jan 09;6(1):e7 [FREE Full text] [doi: 10.2196/mhealth.8998] [Medline: 29317380]

XSL•FO
RenderX

12. Guerra-Reyes L, Christie VM, Prabhakar A, Harris AL, Siek KA. Postpartum health information seeking using mobile phones: experiences of low-income mothers. Matern Child Health J 2016 Nov 17;20(Suppl 1):13-21 [FREE Full text] [doi: 10.1007/s10995-016-2185-8] [Medline: 27639571]
13. Seyyedi N, Rahmatnezhad L, Mesgarzadeh M, Khalkhali H, Seyyedi N, Rahimi B. Effectiveness of a smartphone-based educational intervention to improve breastfeeding. Int Breastfeed J 2021 Sep 20;16(1):70 [FREE Full text] [doi: 10.1186/s13006-021-00417-w] [Medline: 34544429]
14. Morse H, Brown A. The benefits, challenges and impacts of accessing social media group support for breastfeeding: a systematic review. Matern Child Nutr 2022 Oct 12;18(4):e13399 [FREE Full text] [doi: 10.1111/mcn.13399] [Medline: 35821651]
15. Simpson E, Garbett A, Comber R, Balaam M. Factors important for women who breastfeed in public: a content analysis of review data from FeedFinder. BMJ Open 2016 Oct 24;6(10):e011762 [FREE Full text] [doi: 10.1136/bmjopen-2016-011762] [Medline: 27797996]
16. Laws RA, Cheng H, Rossiter C, Kuswara K, Markides BR, Size D, et al. Perinatal support for breastfeeding using mHealth: a mixed methods feasibility study of the My Baby Now app. Matern Child Nutr 2023 Apr;19(2):e13482 [FREE Full text] [doi: 10.1111/mcn.13482] [Medline: 36725007]
17. Martinez-Brockman JL, Harari N, Segura-Pérez S, Goeschel L, Bozzi V, Pérez-Escamilla R. Impact of the lactation advice through texting can help (LATCH) trial on time to first contact and exclusive breastfeeding among WIC participants. J Nutr Educ Behav 2018 Jan;50(1):33-42.e1. [doi: 10.1016/j.jneb.2017.09.001] [Medline: 29325660]
18. Lewkowitz AK, López JD, Carter EB, Duckham H, Strickland T, Macones GA, et al. Impact of a novel smartphone application on low-income, first-time mothers' breastfeeding rates: a randomized controlled trial. Am J Obstet Gynecol MFM 2020 Aug;2(3):100143. [doi: 10.1016/j.ajogmf.2020.100143] [Medline: 33345878]
19. Patchen L, Ellis L, Harrington CB, Ma T, Mohanraj R, Andrews V, et al. Engaging African American parents to develop a mobile health technology for breastfeeding: KULEA-NET. J Hum Lact 2020 Aug 11;36(3):448-460. [doi: 10.1177/0890334420930208] [Medline: 32525434]
20. Mieso B, Neudecker M, Furman L. Mobile phone applications to support breastfeeding among African-American women: a scoping review. J Racial Ethn Health Disparities 2022 Feb 20;9(1):32-51. [doi: 10.1007/s40615-020-00927-z] [Medline: 33219430]
21. Schindler-Ruwisch JM, Roess A, Robert RC, Napolitano MA, Chiang S. Social support for breastfeeding in the era of mHealth: a content analysis. J Hum Lact 2018 Aug 22;34(3):543-555. [doi: 10.1177/0890334418773302] [Medline: 29787686]
22. Coughlin SS. The need for research-tested smartphone applications for promoting breastfeeding. Mhealth 2016 May 04;2(5):18 [FREE Full text] [doi: 10.21037/mhealth.2016.04.03] [Medline: 27231691]
23. Lee Y, Moon M. Utilization and content evaluation of mobile applications for pregnancy, birth, and child care. Healthc Inform Res 2016 Apr;22(2):73-80 [FREE Full text] [doi: 10.4258/hir.2016.22.2.73] [Medline: 27200216]
24. Ling PM, Lisha NE, Neilands TB, Jordan JW. Join the commune: a controlled study of social branding influencers to decrease smoking among young adult hipsters. Am J Health Promot 2020 Sep 20;34(7):754-761 [FREE Full text] [doi: 10.1177/0890117120904917] [Medline: 32077305]
25. Evans WD, McCormack L. Applying social marketing in health care: communicating evidence to change consumer behavior. Med Decis Making 2008 Jun 30;28(5):781-792. [doi: 10.1177/0272989x08318464]
26. Evans WD, Blitstein J, Vallone D, Post S, Nielsen W. Systematic review of health branding: growth of a promising practice. Transl Behav Med 2015 Mar 12;5(1):24-36 [FREE Full text] [doi: 10.1007/s13142-014-0272-1] [Medline: 25729450]
27. Verhoef MJ, Casebeer AL. Broadening horizons: integrating quantitative and qualitative research. Can J Infect Dis 1997 Mar;8(2):65-66 [FREE Full text] [doi: 10.1155/1997/349145] [Medline: 22514478]
28. District of Columbia. Data USA. URL: https://datausa.io/profile/geo/district-of-columbia#:~:text=by%20Nicolas%20Raymond-,Population%20%26%20Diversity,country%20(89.2k%20people) [accessed 2025-02-19]
29. Chiang KV, Li R, Anstey EH, Perrine CG. Racial and ethnic disparities in breastfeeding initiation — United States, 2019. MMWR Morb Mortal Wkly Rep 2021 May 28;70(21):769-774 [FREE Full text] [doi: 10.15585/mmwr.mm7021a1] [Medline: 34043611]
30. Patchen L, McCullers A, Beach C, Browning M, Porter S, Danielson A, et al. Safe babies, safe moms: a multifaceted, trauma informed care initiative. Matern Child Health J 2024 Jan 20;28(1):31-37 [FREE Full text] [doi: 10.1007/s10995-023-03840-z] [Medline: 37982953]
31. Guest G, Bunce A, Johnson L. How many interviews are enough?: an experiment with data saturation and variability. Field Methods 2006 Feb 01;18(1):59-82. [doi: 10.1177/1525822x05279903]
32. Fugard AJ, Potts HW. Supporting thinking on sample sizes for thematic analyses: a quantitative tool. Int J Soc Res Methodol 2015 Feb 10;18(6):669-684. [doi: 10.1080/13645579.2015.1005453]
33. Hamilton AB, Finley EP. Qualitative methods in implementation research: an introduction. Psychiatry Res 2019 Oct;280:112516 [FREE Full text] [doi: 10.1016/j.psychres.2019.112516] [Medline: 31437661]

XSL•FO
RenderX

34. Braun V, Clarke V. One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qual Res Psychol 2020 Aug 12;18(3):328-352. [doi: 10.1080/14780887.2020.1769238]
35. Stowell E, Lyson MC, Saksono H, Wurth RC, Jimison H, Pavel M, et al. Designing and evaluating mHealth interventions for vulnerable populations: a systematic review. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18; April 21-26, 2018; Montreal, QC. [doi: 10.1145/3173574.3173589]
36. Scott SF, Bute JJ. "It doesn't have to be all or nothing": how individuals who use infant formula understand the breastfeeding master narrative. Health Commun 2025 Feb 21;40(2):321-331. [doi: 10.1080/10410236.2024.2344284] [Medline: 38644673]
37. Tran V, Reese Masterson A, Frieson T, Douglass F, Pérez-Escamilla R, O'Connor Duffany K. Barriers and facilitators to exclusive breastfeeding among Black mothers: a qualitative study utilizing a modified Barrier Analysis approach. Matern Child Nutr 2023 Jan 13;19(1):e13428 [FREE Full text] [doi: 10.1111/mcn.13428] [Medline: 36098279]
38. DeVane-Johnson S, Woods-Giscombé C, Thoyre S, Fogel C, Williams R2. Integrative literature review of factors related to breastfeeding in African American women: evidence for a potential paradigm shift. J Hum Lact 2017 May 05;33(2):435-447. [doi: 10.1177/0890334417693209] [Medline: 28380305]
39. Asiodu IV, Waters CM, Dailey DE, Lyndon A. Infant feeding decision-making and the influences of social support persons among first-time African American mothers. Matern Child Health J 2017 Apr 27;21(4):863-872 [FREE Full text] [doi: 10.1007/s10995-016-2167-x] [Medline: 27565664]
40. Gyamfi A, Spatz DL, Jefferson UT, Lucas R, O'Neill B, Henderson WA. Breastfeeding social support among African American women in the United States: a meta-ethnography. Adv Neonatal Care 2023 Feb 01;23(1):72-80 [FREE Full text] [doi: 10.1097/ANC.0000000000001021] [Medline: 35939758]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Studies
EHMF: exclusive human milk feeding
KULEA-NET: Knowledge and Usage of Lactation using Education and Advice From Support Network
mHealth: mobile health

Edited by A Mavragani; submitted 09.10.24; peer-reviewed by L Ou; comments to author 11.01.25; revised version received 27.01.25; accepted 29.01.25; published 19.03.25.

Please cite as:
Patchen L, Tsuei J, Sherard D, Moriarty P, Mungai-Barris Z, Ma T, Bajracharya E, Chang K, Evans WD
Designing a Digital Intervention to Increase Human Milk Feeding Among Black Mothers: Qualitative Study of Acceptability and Preferences
JMIR Form Res 2025;9:e67284
URL: https://formative.jmir.org/2025/1/e67284
doi:10.2196/67284
PMID:

©Loral Patchen, Jeannette Tsuei, Donna Sherard, Patricia Moriarty, Zoe Mungai-Barris, Tony Ma, Elina Bajracharya, Katie Chang, William Douglas Evans. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring Laypersons' Experiences With a Mobile Symptom Checker App as an Interface Between eHealth Literacy, Health Literacy, and Health-Related Behavior: Qualitative Interview Study

Roland Koch[1], Dr med; Marie-Theres Steffen[1], Dr med; Anna-Jasmin Wetzel[1], MSc; Christine Preiser[2], Dr phil; Malte Klemmt[3,4], PhD; Hans-Jörg Ehni[5], Prof Dr; Regina Mueller[6], MA, Dr rer med; Stefanie Joos[1], Prof Dr

[1]Institute for General Practice and Interprofessional Care, Tübingen University Hospital, Tübingen, Germany

[2]Institute of Occupational and Social Medicine and Health Services Research, University Hospital Tübingen, Tübingen, Germany

[3]Institute of Applied Social Science, University of Applied Science Würzburg-Schweinfurt, Würzburg, Germany

[4]Institute for General Practice and Palliative Care, Medizinische Hochschule Hannover, Hanover, Germany

[5]Institute of Ethics and History of Medicine, University Hospital Tübingen, Tübingen, Germany

[6]Institute of Philosophy, University of Bremen, Bremen, Germany

Corresponding Author:
Roland Koch, Dr med
Institute for General Practice and Interprofessional Care
Tübingen University Hospital
Tübingen
Germany
Phone: 49 1758065961
Email: roland.koch@med.uni-tuebingen.de

Abstract

Background: Symptom checkers aim to help users recognize medical symptoms and recommend actions. However, they are not yet reliable for self-triage or diagnostics. Health literacy plays a role in their use, but the process from symptom recognition to health care consultation remains unclear.

Objective: This qualitative observatory study explored how laypersons use symptom checkers, focusing on the process of use, entry points and outcomes, and the role of health literacy. Laypersons are defined as individuals who are neither medical professionals nor developers of such apps. Three research questions were addressed: (1) How do such users describe the process of using symptom checkers? (2) What are entry points and possible outcomes of symptom checker app use? (3) How are health literacy and eHealth literacy expressed during the use of symptom checker apps?

Methods: As part of the Ethical, Legal, and Social Implications of Symptom Checker Apps in Primary Health Care project, 15 laypersons (n=9, 60% female and n=6, 40% male; mean age 30.7, SD 13.6 years) were interviewed about their experiences with the symptom checker Ada. The interviews were analyzed using an integrative approach combining social positioning, agency, and the Rubicon model as a heuristic framework.

Results: App use follows a cyclic process comprising 4 steps: motivation (influenced by biography and context), intention formation (assigning a purpose), intention implementation (recruiting resources), and evaluation (transforming interactions into health-related insights). Biographical, social, and contextual factors shape process initiation. Users use symptom checkers for 3 main purposes: understanding their condition, receiving recommendations for action, and documenting or communicating health-related information. Each purpose requires specific planning and integration into health-related behaviors drawing on personal, social, and technological resources. Evaluation depends on contextual factors, app outputs, and the outcomes of users' health-related actions. Users assess whether the app aligns with their expectations, condition severity, and previous experiences, with health literacy playing a critical role in validation processes.

Conclusions: Symptom checker use is a complex, cyclic process shaped by context, biography, and health literacy. Users are motivated by health concerns influenced by personal, social, and contextual factors, with trust and attitudes impacting initial engagement. Intention formation reflects a balance between user skills and context, where app outputs inform decisions but may not always lead to action, especially in ambiguous situations. Users rely on personal resources and social networks to integrate app use into health-related behaviors, highlighting the limitations of symptom checkers in providing social or empathetic support.

XSL•FO
RenderX

Symptom checkers have the potential to serve as an interface between users and health care, but future development must address the complexity of their use to unlock this potential.

International Registered Report Identifier (IRRID): RR2-10.2196/34026

(JMIR Form Res 2025;9:e60647) doi:10.2196/60647

KEYWORDS

symptom checker apps; health literacy; eHealth literacy; qualitative research; interview study; artificial intelligence; AI

Introduction

Background

This qualitative study investigated how eHealth literacy (eHL) [1,2], health literacy (HL) [3-6], and health-related behavior (HRB) influence how laypersons, defined as individuals who are neither medical professionals nor developers of symptom checker apps, engage with a specific mobile symptom checker app. Symptom checker apps offer readily accessible tools for self-triage by prompting users to enter health-related information [3,7]. This study focused on the Ada app by Ada Health GmbH.

Research shows that symptom checker apps are often used out of curiosity or anxiety or to confirm the severity of symptoms [4,7-9]. Factors such as users' age, confidence in self-assessment, trust in the app and technology, technology affinity, past experiences with the health care system, and proficiency in navigating health information play critical roles in shaping user experiences and engagement with symptom checker apps [3-7,9-16].

There is limited and partly conflicting evidence to conclude whether symptom checker app use directly prompts HRB, such as seeking professional health care [3,9,13,16,17]. Emerging literature suggests that symptom checker app use may foster self-care but could also lead to unnecessary health care use, particularly when the apps err on the side of caution in nonacute settings (eg, feeling tired over several days) [18-22]. Symptom checker app use can influence users' anxiety about seeking medical help [3,4,17,23], with many preferring "wait and see" recommendations [7,17], although satisfaction with the app does not always correlate with adherence to its advice [13,24].

In addition, previous studies highlight that eHL and HL significantly impact users' interaction with digital health tools and health outcomes, yet there is a knowledge gap in understanding these literacy concepts in the context of symptom checker app use [25-31]. Another gap in the literature is how the social context influences the use of digital health applications such as symptom checker apps, especially considering that eHL and HL are distributed unequally across regions and social groups [32-36].

How the occurrence of a symptom leads layperson users to engage in symptom checker app use and, consecutively, act on the app's output has not yet been described as a process [3,37]. There are several challenges to researching symptom checker apps, such as different user profiles, the context dependency of app use, and the rapid rate of development in the field [3,14,38].

Objectives

This study aimed to fill that gap by describing the process through which layperson users engage with symptom checker apps for health concerns, examining the influence of eHL and HRB on their interactions with the app and possible outcomes. Specifically, this study addressed (1) the steps that layperson users follow in using symptom checker apps, (2) key entry points and outcomes of symptom checker app use, and (3) the roles of eHL and HL in this process.

Methods

Study Design

This observational, exploratory qualitative study was embedded in the larger context of the joint Ethical, Legal, and Social Implications of Symptom Checker Apps in Primary Health Care (CHECK.APP) project [39]. CHECK.APP represents a collaboration among the Institute for General Practice and Interprofessional Care, the Institute of Ethics and History of Medicine, the Institute for Occupational and Social Medicine and Health Services Research at Tübingen University Hospital, and the Institute of Applied Social Science. The interdisciplinary research team consists of persons with professional training and research experience in medicine, family medicine, psychology, sociology, law, medical ethics, and philosophy. The project is conducted in the context of the German health care system with a focus on primary health care.

The CHECK.APP project is based on a mixed methods design and has four foci: (1) ethical, legal, and social issues [3]; (2) epidemiology of symptom checker app use and predictors of use [9]; (3) patterns of symptom checker app use and impact on individuals [16]; and (4) impact of symptom checker app use on the health care system and health care workers [40]. This study was part of focus 3. Focus 3 consisted of a diary study that was complemented with a follow-up qualitative interview study. The 6-week self-monitoring diary study aimed to generate mixed methods data on symptom checker app use patterns [16].

This study was conducted following the diary study and was designed to complement the quantitative analysis of the use patterns observed in the diary study by providing an in-depth exploration of the motivations and experiences of symptom checker app users. The reporting of this study followed the Standards for Reporting Qualitative Research guidelines statement [41]. The reporting checklist is included as Multimedia Appendix 1.

How Ada Works

On the Ada app, users enter their symptoms by responding to a series of questions. On the basis of their answers, Ada

XSL•FO
RenderX

generates follow-up questions to refine the symptom description. For example, if a user reports a rash, the app displays various images of rashes, prompting the user to select the one that best matches their condition. The app then yields a list of possible causes, recommendations on care urgency, and self-care advice (eg, “Eight of ten users with these symptoms had an acute infection—it is recommended that you go see a doctor”).

Sample

For sample size estimation, we applied the 5D model of information power by Malterud et al [42], which considers study aim, sample specificity, use of established theory, quality of dialogue, and analysis strategy. According to the 5D model, the more specific each dimension, the smaller the required sample size in a qualitative study. Given the study aim of an in-depth exploration of symptom checker app use by laypersons, the specific focus on users of a single app, the analysis approach grounded in established theory (see the Data Analysis section), the expertise of the researchers involved with an expected high dialogue quality, and the planned techniques to enhance trustworthiness (see the Data Analysis section), a sample of 15 interviews was estimated to generate sufficient information power to explore the phenomenon under study. We reserved the option to recruit additional participants if the estimated sample size proved insufficient. To determine this, the analysis began with an initial reading of the transcripts even before the interviews were completed (see the Data Analysis section).

Participants for this study were recruited from a pool of 48 active symptom checker app users who had previously completed the 6-week self-observation diary study. All participants in the diary study were active users of Ada (Ada Health GmbH) and were originally recruited through the survey conducted as part of the CHECK.APP project. Recruitment was limited to existing Ada users to comply with ethical requirements prohibiting the initiation of symptom checker app use for the study. Participants used Ada within the context of their daily lives. Incentives were provided for maintaining the diary and participating in the interviews.

These diary study participants were approached and asked whether they were interested in undergoing additional in-depth interviews based on their respective diary experiences. All 48 persons who took part in the diary study were laypersons, meaning that they were neither health care professionals nor symptom checker app developers [16]. A chronic condition was reported by 21% (10/48) of the potential interview partners. In total, 35% (17/48) of the potential participants used the symptom checker app infrequently (less than once per week). Mean age was 27 (SD 9.1; range 19-64) years. A total of 65% (31/48) of the potential participants identified as female, and 35% (17/48) identified as male. No person identified as nonbinary. Racial characteristics were not asked about in the diary study [16]. Of the 48 persons who took part in the diary study, 28 (58%) expressed interest in an interview. The participants of the diary study from whom the interviewees were selected primarily consisted of well-educated younger female adults, a limitation discussed later in the manuscript.

Among these 28 persons, interviewees were stratified based on age, gender, use frequency, and experiences addressed in the diaries to achieve the best heterogeneity within the rather homogeneous sample. The relevant characteristics of the 15 selected interview participants can be found in Table 1. Mean age was 30.7 (SD 13.6) years.

XSL•FO
RenderX

Table 1. Study sample for the symptom checker app user interviews (N=15).

Relative dimension and features	Participants, n (%)
Age (y)	
20-29	10 (67)
30-39	2 (13)
40-49	1 (7)
50-59	1 (7)
60-69	1 (7)
Gender	
Women	9 (60)
Men	6 (40)
Diverse or nonbinary	0 (0)[a]
Symptom checker app use frequency	
Sparse (less than once per week)	6 (40)
Frequent (more than once per week)	9 (60)
Educational level	
Upper secondary school leading to university entrance qualification	12 (80)
Intermediate secondary school	2 (13)
Basic secondary school	1 (7)
Place of residence	
Urban	13 (87)
Rural	2 (13)
Chronic condition	
Present	5 (33)
Not present	10 (67)

[a]Despite our efforts, no nonbinary or diverse persons were included in the study.

Data Collection

The interviews were conducted by a pair of interviewers using an interview guide (Multimedia Appendix 2). The interview guide was created by the research team for this study and piloted with 2 interview partners who were not included in the analysis. The test interviews were used to review the guide and interview styles in a group feedback discussion between the interviewers and their supervisors. No changes were made to the interview guide during data collection. The interviews were conducted by a pair of interviewers comprising 1 representative from the Institute for General Practice and Interprofessional Care (AJW, RK, or MTS) and 1 representative from either the Institute of Ethics and History of Medicine (RM) or the Institute of Applied Social Science (MK). After consent was obtained, the interviews were conducted using the Zoom (Zoom Video Communications) video client and audio recorded on local, secure computers at Tübingen University Hospital. The interviews were conducted from January 18, 2022, to March 15, 2022, with an average duration of 46 (SD 10.4; range 35-76) minutes.

The interviews were transcribed pseudonymously by a certified office (Amanu). The transcripts were stored on protected servers at Tübingen University Hospital. For analysis, Microsoft Excel and Microsoft Word (Microsoft Corp) were used.

Data Analysis

This qualitative interview study followed a constructivist research paradigm—users construct their social reality by integrating experiences, self-image, daily interactions, and bodily perceptions into their biography [43]. They learn from experience [44]. These integrated factors are expressed through the social act of narration during the interviews.

The integrative basic method by Kruse [45] is an interpretative-reconstructive qualitative approach to reveal manifest and latent meanings and the way in which users “make sense” of the app and their experience with it. It represents an in-depth analysis of self-perception, biographical connections, self-image, and the perception of social roles within interview material [45]. The analysis was conducted by RK and MTS using strategies to enhance trustworthiness and rigor (see the Data Analysis section). The process is visualized in Figure 1.

Initial transcript reading commenced on March 1, 2022, while 3 interviews were still pending. This revealed thematic complexity; distinct patterns; and notable contrasts,

XSL•FO
RenderX

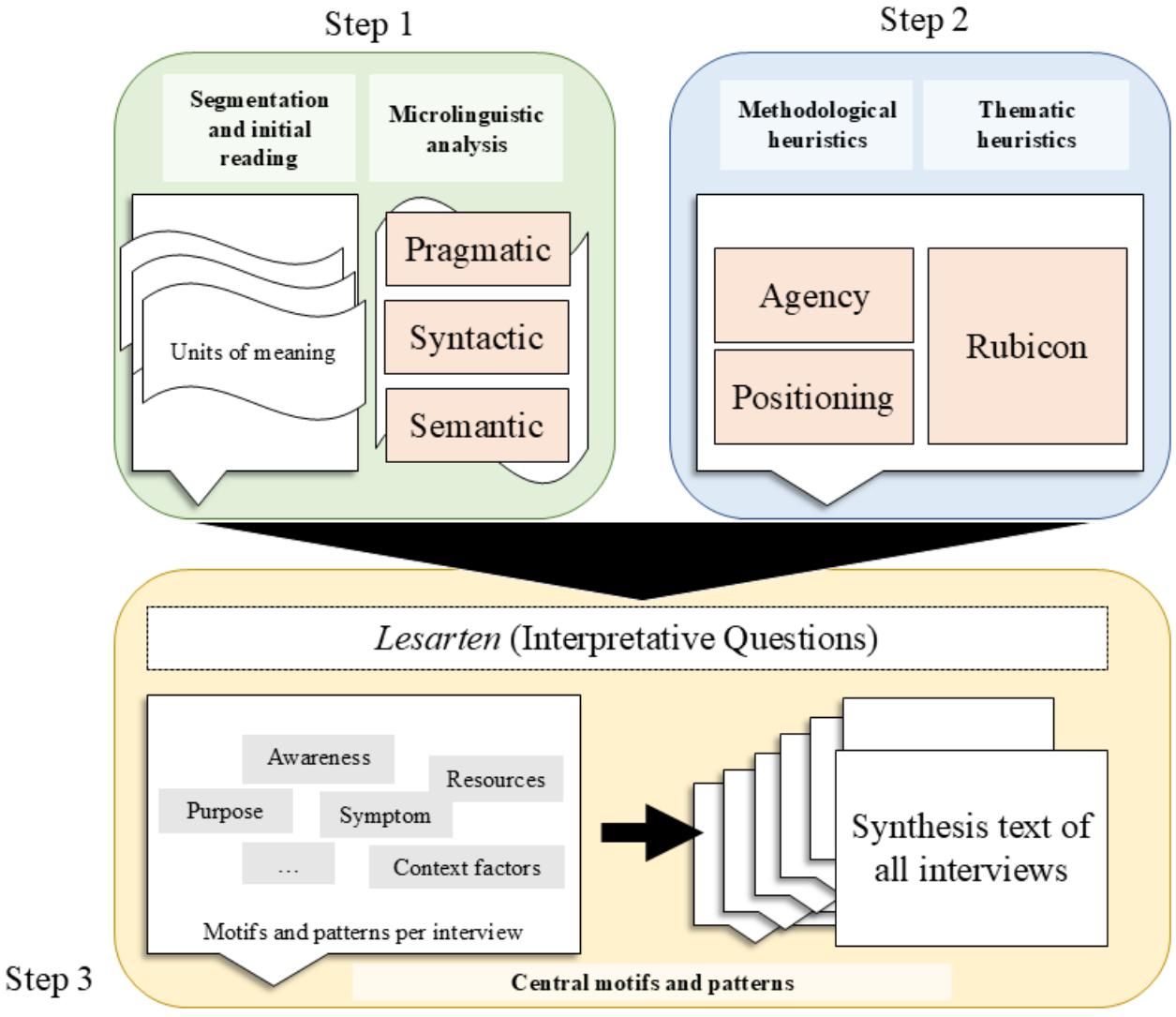
contradictions, and differences within the data. As a result, recruitment of additional participants beyond the planned sample size of 15 interviews was deemed unnecessary.

The first step in the integrative basic method is a detailed microlinguistic analysis in which transcripts are segmented into discrete units of meaning guided by cues such as pauses, interviewer questions, and topic changes. Each unit is then analyzed on three distinct linguistic levels capturing different aspects of the speaker's attention: (1) pragmatic (interactional aspects, including the relationship between interviewer and participant), (2) syntactic (grammatical choices that reflect cognitive structures), and (3) semantic (specific word choices, metaphors, and other lexical elements). This descriptive in-depth analysis of interaction, syntax, and semantics lays the groundwork for the following analysis of concepts such as agency, positioning, and the Rubicon model. Microsoft Excel was used for this step.

The transcripts were then analyzed using methodological and thematic heuristics. In our case, an analysis of social positioning and agency was used to understand how layperson users perceive the symptom checker app themselves and their scope for action embedded in a network of participants [46,47]. In terms of thematic heuristics, we used the Rubicon model by Achtziger and Gollwitzer [48] as a framework for motivational processes and intention building. It served both as a thematic lens to help explore the intrapersonal motivational process and as a macrostructure for the final synthesis text. The analyses were conducted using Microsoft Word.

In the final stage, the preceding microlinguistic analysis and the methodological and thematic analyses were condensed and compared against the background of the research questions to identify core motifs and modes of thematization within each interview and between interviews. This step began with formulating interpretative questions ("Lesarten" in German). The questions formed critical bridges between microanalysis (step 1) and interpretative analysis (step 2) to ensure that the interpretations remained grounded in the data. For instance, a question might be the following: "What role does Ada play in the progression from symptom onset to the utilization of health care?"

Figure 1. Visual representation of the integrative basic method. The process begins with 2 analysis steps: segmentation into units of meaning and microlinguistic analysis (step 1) and methodological and thematic heuristics, which are applied to the units of meaning (step 2). Both steps are then unified through interpretative questions ("Lesarten") into a final step (step 3), identifying central motifs and patterns within and across the interviews. Speech bubbles symbolize interviews, whereas flags denote units of meaning.



XSL•FO
RenderX

By answering the questions, central motifs and patterns specific to each interview were identified inductively. These motifs and patterns were then compared across all interviews to identify recurring themes (commonalities) and deviations within and across transcripts. While commonalities were used to generalize the findings, deviations were used to refine them. These insights were then used to reveal and formulate central, overarching motifs and patterns that were categorized using headlines (eg, "Context of app use") and structured according to the steps of the Rubicon model. Microsoft Excel was used for this step.

To enhance trustworthiness, RK and MTS performed each of the aforementioned 3 analysis steps independently on the same transcript. They then compared their results after each step. Differences were discussed, and if possible, a mutual understanding of the texts and the method was established. When disagreements arose, they were discussed in monthly multidisciplinary methodological workshops within the CHECK.APP project research team.

Relying on the method of peer checking, the interim results were presented and discussed several times in a multidisciplinary research workshop on qualitative research methods. For a member check with most of the interview partners, the interim results were discussed and validated during a 3-hour session. Furthermore, in a multidisciplinary validation workshop, intermediate states of the analysis as well as open questions were discussed and illuminated under consideration of the perspectives of law, ethics, social medicine, and general medicine within the CHECK.APP project team. The results of these discussions were used to inform interpretative analysis (eg, by considering more perspectives) and the formulation of the final results text.

Ethical Considerations

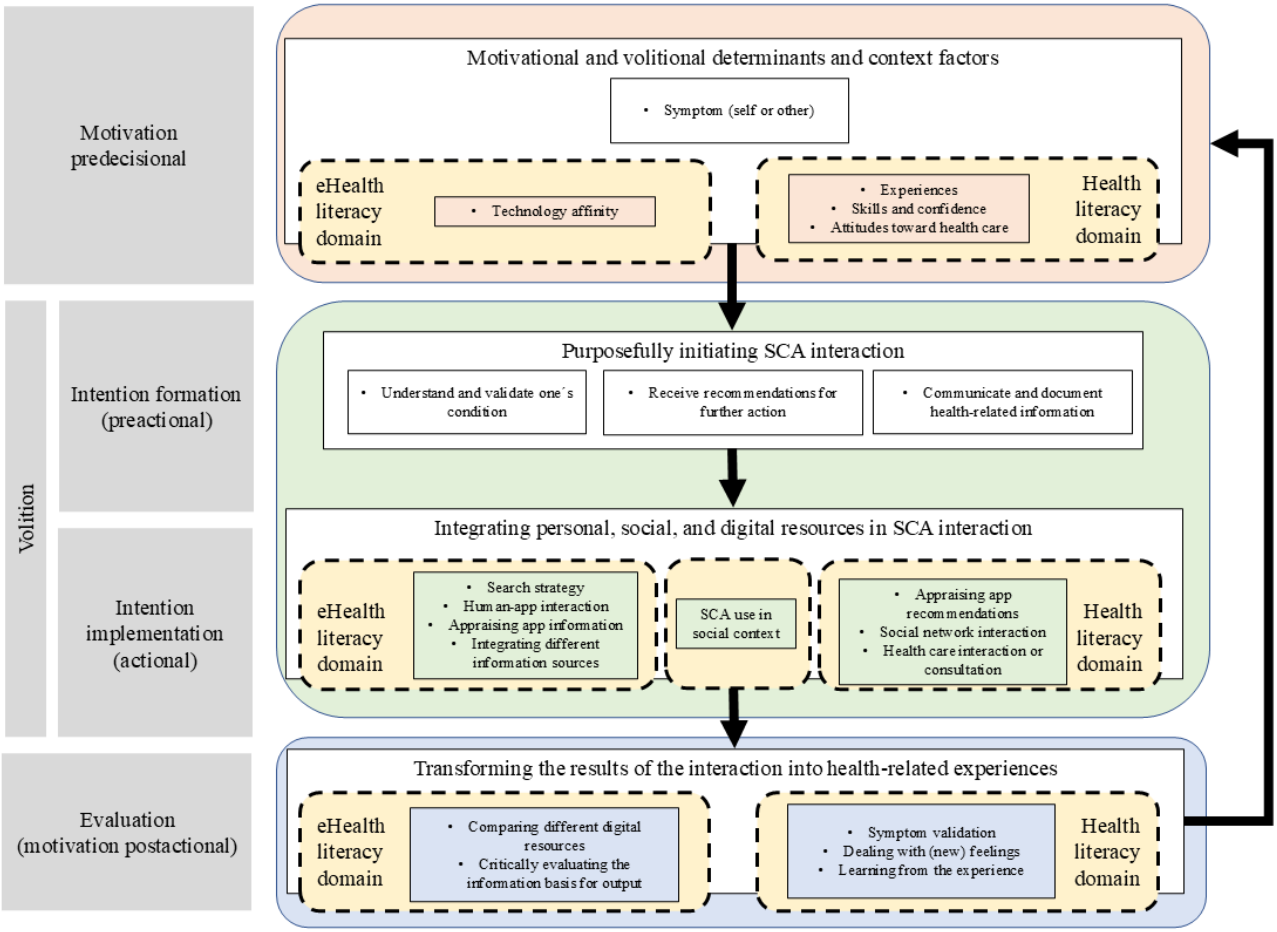
This study received a positive vote from the ethics committee of the University of Tübingen (464/2020BO) and was conducted following the Declaration of Helsinki. All participants gave their informed consent before the interviews began. The participants received monetary compensation (€50, approximately US $50) for taking part in the interviews. All data were stored on encrypted servers.

Results

Visualization of the Symptom Checker App Use Process

On the basis of the 4 stages of the Rubicon model, a cyclic model of the symptom checker app use process according to the interview partners was developed. It is presented in Figure 2.

Figure 2. Symptom checker app (SCA) use process based on 15 interviews with SCA users. The figure is based on the Rubicon model. It describes the path from an entry point (such as noticing a health-related concern) via intention formation (influenced by motivational and volitional determinants) and volition to health-related behavior as an outcome. The latter can take place in different contexts. The behavior, in turn, leads to new experiences that are evaluated by the individual and integrated into the biography.



XSL•FO
RenderX

In the following sections, the synthesized text produced through the integrative basic method is presented. Interview excerpts are provided to illustrate the data basis for the central motifs and patterns. Unless stated otherwise, the excerpts are from the interview partners. If the interviewer asked a question, it is marked with "interviewer."

Motivation: Predecisional

Motivational Stage

Users assessed the severity of the issue by comparing it to their own preexisting experiences. The interview partners did not reflect on the relevance or validity of their perceived symptoms and experiences—they took for granted that both were real and valid. If the symptoms were perceived as less severe and comparable experiences were available in the users' biography, the sensation was described as manageable:

> *Yes. I have that once in a while. I don't even know where that comes from. I also think that it's possibly one of those things that kind of comes from stress, I definitely feel like that. And because I just never knew what it was, but I also knew that it would go away again, so because I had it before, as I said, I just looked it up in the app once.* [NT45]

At times, when individuals could not find relatable experiences from their own health history or perceived the issue as more severe, they expressed negative emotions such as fear or shock:

> *...In the end, it was a migraine attack. But I didn't think about it at that moment, I was shocked at first.* [NT39]

Trust in the app also played a role in the decision to use it, as the following excerpt illustrates. The same interview partner continued to elaborate on how trust in the app competed with trust in health care professionals:

> *But, yes, because I couldn't describe my problem in detail—I didn't trust the app so much at that moment—it was important for me to go to the doctor quickly, because I knew, okay, there are important things that I couldn't tell the app, and then that would explain the diagnosis, so to speak.* [NT41]

Context of App Use

Interview partners reported on different contexts in which they considered using the app. The most commonly reported context was the occurrence of a health-related concern. They first reflected on whether the symptom was even "real"—meaning whether they had just imagined it or whether the perceptions really represented an issue worth dealing with. They compared the symptom with their health-related experiences and biography and tried to make sense of what was happening to them:

> *Yes, I also clearly had chest pains. I mean, at that point I was already asking myself: Okay, to what extent is this happening now? But I mean, then I lay there and I had the feeling and thought: Yes, come on—so it was already real for me at that moment.* [NT47]

Users also reported being motivated to use the app by the health-related concerns of others. One user reported that she used the app for her family. She said that she was more resistant to possible anxiety than her family, whom she described as prone to hypochondria:

> *Interviewer: And what made you decide to [use the app] for [your daughter] and not with her?*
>
> *Interviewee: That I wanted to check it out first...My husband is a bit of a hypochondriac and so is my child. So when I sort of poke them, or it could be something dramatic, then I always don't know what will come of it. That's why I did a pre-check for myself first.* [NT21]

Health system accessibility or formal requirements because of their work further influenced users' motivation to take health-related actions:

> *Interview partner: ...It...is also one of the reasons why you don't like going to the doctor so much, is that it is hard to get an appointment with such specific doctors somehow. Especially a dermatologist or something like that, you have to wait two or three months and then you usually let it drop.*
>
> *Interviewer: So access to a doctor or to care plays a role for you?*
>
> *Interview partner: Yes. Yes. I think if I knew, okay, I'm going to get an appointment in one, two, three weeks, then I'd be more likely to have something like that checked out.* [NT41]

Volition: Intention Formation (Preactional)

The motivational process described previously led to a planning stage for HRB during which users purposefully engaged with symptom checker apps with 3 distinct goals in mind. They sought to (1) understand and validate their condition, (2) receive recommendations for further action, and (3) communicate and document health-related issues.

Understanding and Validating One's Condition

Users were interested in understanding their condition. The app provided an opportunity to find more information. Information seeking was described as one way to deal with the concerns:

> *...I see [the app] as a first aid, first source of information, where you can think about...: What do I do now? Then I can take a look or maybe I think again: What do I do now?...so as...again to update the normal knowledge that you have as a non-medical person.* [NT37]

Finding a clear cause also drove the users and induced HRB:

> *I woke up with 39 and something degrees. And of course I opened the app again and entered my symptoms, just to see...what it was saying now...Yes, then I knew...Okay, these are clearly the symptoms. Which of these is now Corona, which is maybe something else. Yes, I can still remember...that I was lying in bed suffering and had the app open.* [NT05]

XSL•FO
RenderX

Sometimes, users described that an element of surprise was needed to find an explanation for their concerns. They admitted that they were not aware of all possible causes of their conditions. The ability to be surprised was attributed to both physicians and the app as external information sources:

> *Maybe it's (what is needed to accept a diagnosis) a "click" moment that you didn't realize before. As an example, maybe a doctor or a diagnosis app asks a question [like: Does it get worse when it's cold] and then you think to yourself: Ah, yes, when I think about it, it does get worse when it's cold, or whatever.* [NT24]

Other times, users already had a suspicion of their own, which could be verified by the symptom checker app:

> *...I also have migraines and sometimes I can't quite identify whether it's a migraine or a headache. And the app often helps me and then I can practically distinguish: OK, do I treat it like a headache now or do I treat it like a migraine now.* [NT41]

In summary, the purpose of gathering information was to add to one's own knowledge about the condition and actively seek out validation of one's subjective impression.

Receiving Recommendations for Further Action

In addition to collecting more information about one's condition, users stated wanting to receive options for further health-related actions to improve their condition:

> *What were the first experiences?...I used the feature best or most, simply these suggestions about which therapies you can take for which symptoms. And so I always looked here again to see what the app actually recommends.* [NT44]

Even if the app was primarily used for information gathering, the recommendations were received and considered:

> *...I looked to see what information was available if I somehow...well, with children there's always something. And then I just thought the app was so good, because sometimes they had a few other tips. And then I just looked to see what one says, what the other says...It was always a push in the direction: "I should go to the doctor."* [NT21]

Users also typed in symptom combinations several times to see whether anything changed in the app's output:

> *Well, of course I also had accompanying symptoms during the heart rhythm disturbances. So sometimes trembling or high blood pressure, but some of it was only temporary. I then sort of started to play with that. I added the symptom once and indicated it and left it out once to see what it recommended. Because the option of selecting that a symptom only occurs temporarily or is concomitant, only temporarily, is not available in the app. And then I wanted to play around with what would come out in the end if I indicated this and didn't indicate it. But apart from that, the reason was...I gave everything I had and always answered the questions truthfully.* [NT19]

If the sense of medical urgency was very high or life-threatening, app use was explicitly avoided. One participant described such a hypothetical scenario:

> *If I somehow had symptoms of a heart attack or something like that, then I would probably call the ambulance service directly and not look on the app to see if I really had a heart attack or if I was bleeding to death or something like that.* [NT49]

Communicating and Documenting Health-Related Issues

Users documented their own symptoms on the app:

> *It was recommended to me a long time ago to track symptoms, just to keep a diary of symptoms...* [NT31]

The app was also used to prepare for a visit to health care professionals. In that context, the documentation contained on the app was primarily used for the orderly and rapid presentation of concerns:

> *So I can already describe my state of health more specifically than if I hadn't inquired beforehand.* [NT06]

Volition: Intention Implementation

Personal Resources

To implement the purpose of the initiated action, users integrated personal, social, and digital resources into their behavior. One example leading to a health care visit was provided by one of the interview partners:

> *My doctor didn't tell me that [oral iron supplement pills] could have side effects. And then I had discolored feces and abdominal pain and all sorts of things for a week or two...So I used the app and then I googled it and then I think I also talked to a friend about it. (00:04:27) She was like: Hey, maybe it's from the iron tablets. And then I realized that it was the iron tablets. And then I stopped taking them and then I talked to my internal medicine physician and she was like: Yes...And prescribed me other iron tablets.* [NT41]

Personal resources were provided by biographical reference points such as previous experiences or special knowledge from one's own professional activity or skills. As soon as the purpose of the action was determined in the preactional phase, personal resources were factored into the user's action. Users described certain skills that they used in the process but also a sense of self-efficacy. If enough personal resources were available, users could actively decide against the app's recommendation, as the following example shows:

> *And then I entered that [the symptoms]. The app said it could be a torn ligament or a fracture and that I should go to the emergency room. And then I thought: ...I don't know how I could have broken any bone in the way I was walking. For me, the ligaments were more plausible...I've had it before, so I already had a bit of experience of what it's like. And then I thought: OK, the app says one thing, but I'll just look*

XSL•FO
RenderX

at it again in two or three days and see how it develops. [NT42]

Social Resources

Users also intentionally and purposefully sought out social resources. Friends and family were assigned a role in understanding and responding to concerns:

> *...I once suspected a stroke and then I was at home and had a reduced field of vision and thought: What is this? But that was before I knew about these apps. And of course I immediately called a friend and said: What could this be? What is that? And she said: Make sure you have it checked out the next day.* [NT21]

Health care professionals such as physicians were assigned a special role—a visit to a health care professional was seen as mandatory to obtain a final validated diagnosis or to obtain health care services such as an authorization for sick leave or a prescription. The interactions with physicians were mostly intentionally initiated by the users if other resources were depleted or did not improve their situation:

> *And I had then already contacted my family doctor anyway, just from the fact that I also need a sick note...Of course, if this [the symptoms] were to last longer, the app would be of no use to me in the end, then I would have to go back to the neurologist or...the family doctor.* [NT39]

The only social resource with physical access to the users' body were physicians as they could conduct physical examinations:

> *And then I went there [to the General Practitioner], described it again and showed it. Then...[she] turned the [painful arm] in different directions and said that no part, neither the forearm nor the upper arm muscle, is affected in any way. That comes in any case, exactly, so from the tendons...And then I just said: yes, [the concern comes] from climbing...And then she said: how much do you climb? So if you don't climb four hours a day, it's very unlikely that it comes from that.* [NT31]

Reviewing the input and results together with health care professionals (eg, a general practitioner [GP]) was mostly described hypothetically. However, the prospect of discussing app results with their GP was seen as beneficial by the users:

> *...what I would like is when you come to the doctor and say: I assume that I have this because I googled it or found out on an app. That he then says: Yes, what symptoms did you enter? Or: How did you then come up with this result?...So that he then, I don't know, doesn't have to start again with Adam and Eve, but that he can already inquire more intensively.* [NT06]

Mostly, users made a conscious decision to leave the app out of the social context of health care. They expected it to have a negative impact on patient-physician relationships:

> *Interviewer: Did you also mention there (at the GP) that you used the app?*
>
> *Interview partner: No, because I have the impression that doctors tend to react badly to this. Because then, they always like to say Dr. Google...and the topic has a bit of a bad reputation.* [NT07]

Digital and Technological Resources

The app and search engines were seen as competitors in the same category with distinct advantages and challenges. Search engines were mainly used as a supplement to obtain information that was missing on the app:

> *...when you're constantly googling, sometimes very worrying answers come up and you can also enter a lot of specific things in the app and it's just more practical and you somehow feel, how should I put it, the diagnosis somehow feels more trustworthy than through Google.* [NT41]
>
> *...I missed in the symptom checkers, that you somehow get such a tip, such old home recipes...belly compress or something, and then such a guide to it...I then thought: ...now I'm googling, what can one do, yes, perhaps with a homeopathic approach. Because that's what I was missing.* [NT21]

In contrast to app use, users found search engines overwhelming when they tried to check their symptoms there:

> *[Referring to searching concerns with internet search engines] In principle I never really figured it out, because there were 10, 15, 20 possibilities, and I couldn't really assess it at all. Really.* [NT19]

There are other technological resources and devices such as users' wearable devices or imaging procedures at the hospital in addition to the app. Devices able to monitor and measure bodily functions were described as determining and important factors that can contribute to a decision:

> *I...always measured the blood pressure. And even my [smartwatch], which is so clever, then recognized that there is a heart rhythm disturbance, at times. That time, it showed a rhythm disturbance...Then I thought: Okay, now I have to go to the doctor.* [NT19]

The user experience and usability features of the app were discussed. For example, the chatbot-based approach with questions asked to be answered by the user was perceived as both helpful and limited. Users missed the opportunity to clarify their input, ask questions, or address their own uncertainty and ambiguity:

> *...I would have put it (a surprising app-result) down to the fact that the app didn't really understand me...because I somehow couldn't really convey exactly what I wanted to say on the basis of the questions and the selection options...I would have said: OK, it just misunderstood me.* [NT37]

According to the users, affinity for technology played a role in the use of the app:

> *Yes, for my mum, for example, I use it (the app) quite often, because she's not that into the internet and apps and stuff.* [NT41]

XSL•FO
RenderX

Evaluation

The outcomes of the interaction between users and the symptom checker app were integrated into their subsequent experiences and interpreted in the context of their past health-related experiences. Once the HRB had been executed, a comparison was made as to whether the intention of the action had been fulfilled. Whether the app's results represented an actual diagnosis was negotiated individually by each user. Users described having difficulty accepting the app results depending on the topic and their own interpretation of the results. If the 2 diverged too drastically, they were at a loss and saw potential problems in understanding or misinterpreting the given recommendations:

> *I think the app always told me: stress and mild depression or something like that. I didn't feel that stressed, but yes, I think that was the thing...I read through it, but then I thought: Well, I still don't know exactly what's going on.* [NT37]

Concerning recommendations especially, users critically evaluated the app's output. How they reacted to it depended on the intended purpose of the action, the user, and the context in which the assessment took place:

> *I was a bit shocked that [the app] now sends someone directly to the emergency room (with the app result tonsillitis). Because I work for a health insurance company...and I find it a bit exaggerated that patients are sent so quickly to the emergency room. They are so overloaded at the moment.* [NT39]

Users reported that they were reassured by validated external information about their condition. The app sometimes could provide such reassurance:

> *It's a bit like this: Is it true now? Of course, we are aware, especially with this vaccination, that the side effects come from the vaccination. But somehow it's still like that: Yes, you feel safer and somehow more confirmed, even if you only have it in an app, but, yes, you are somehow more reassured.* [NT41]

On the other hand, users also found the app results unsettling and misleading:

> *So it could really only be overstretched ligaments (user's suspicion), this fracture (app result) made me a bit...A bit scared at that moment.* [NT42]

Users acknowledged that having received information contributed to a learning process that affected how they reacted to future symptoms:

> *I still have it on my mobile phone, I still like to look at it once in a while, even if it's just to advance my knowledge a bit. And, yes, it has become...I have found that I look at it a little less, but also because I find that when things repeat themselves in some way, you already know how you could react.* [NT44]

Discussion

Principal Findings

Layperson Perspectives on Using Symptom Checkers for Health-Related Concerns

This study's findings indicate that the use of a symptom checker app by layperson users is a multifaceted and iterative learning process that involves individual motivational factors; contextual elements; and interaction with digital, social, and personal resources. Both eHL and HL play a significant role in this process. The subsequent Discussion section addresses the 3 research questions.

The use of symptom checkers can be mapped onto the 4 distinct steps of the Rubicon model: motivation, intention formation, intention implementation, and evaluation [48]. These stages can be viewed as an iterative learning cycle that generates health-related experiences according to the constructivist paradigm [44]. During the motivational stage, users negotiate app use depending on the context. If users choose to use the app, they do so intentionally for 3 distinct purposes: obtaining information and health information–seeking behavior (HISB), receiving recommendations, and documenting and communicating health-related issues. The purpose determines the user's strategy and the resources used to achieve it, including personal, social, and digital or technological resources. Users' experiences with the symptom checker app can influence their motivation to use it in the future and how they incorporate it into their HRB. Making meaning plays a crucial role in this as users try to understand their bodily perceptions and symptoms [43]. This process aligns with the model proposed by von Wagner et al [29], which links HL and HRB and emphasizes the role of learning in HL. Our study highlights that the use and evaluation of symptom checker apps in line with the dynamic concept of HL varies over time and is dependent on the context and purpose of use [25,27,28,49]. This poses a challenge when researching symptom checker apps [4,7,38].

Entry Points and Possible Outcomes of the Symptom Checker App Use Process

Entry Points

Entry points were identified during the motivational or predecisional phase. They were influenced by the context in which symptom checker app use was considered [29]—in addition to the user's own health needs, situations involving the health needs of others were also relevant. Motivational determinants included attentiveness toward symptoms, past experiences (biographical factors), confidence in one's health, trust in the app, and one's ability to respond to symptoms. Some of these factors for app use were also identified by Aboueid et al [4,6] and Meyer et al [7]. The accessibility of health care and availability of social networks were described as external factors. Aboueid et al [4] identified the lack of accessibility of health care as an enabler of app use, whereas social influences were identified as a barrier. The diary study from the CHECK.APP project found that the first occurrence of a symptom and certain symptoms such as heart-, skin-, or eye-related concerns increased the likelihood of symptom checker app use [16]. In

XSL•FO
RenderX

the interview sample, interview partners considered the aforementioned factors in their decision to use the app but did not specify whether they were facilitators or enablers. Although many of the listed factors have been identified in previous research, their role in symptom checker app use is ambiguous as they can act as both barriers and enablers [3,10,21]. As both situational and individual factors shape when and for what purpose symptom checker apps are used, the term "context factors" seems much more appropriate than "barriers and enablers" [3,10,27,28].

Outcomes

Meaningful Learning Experiences

The most common outcome of using symptom checker apps as a digital information source and HISB is the act of learning, which is difficult to measure [1]. This outcome is expressed through the purpose of understanding one's condition based on users' need for more information. Users already have predefined concepts about their condition in mind when planning their course of action. Therefore, they also want to validate these concepts by gathering information. In a substudy of the CHECK.APP project by Wetzel et al [9], it was found that symptom checker apps were the least important for HISB in the general population. The subsample of active symptom checker app users reported that internet use and consulting a physician were more important for HISB than symptom checker apps [9]. In the study by Meyer et al [7], most symptom checker app users stated that they used symptom checker apps to understand the cause of their condition and found the information they received useful. During the evaluation stage, users consider the extent to which the information is meaningful. They base their evaluation on their experiences with symptom checker app use and the health-related actions and determine whether the intended purpose was fulfilled and whether meaningful learning experiences were generated [24,44].

HRB Outcome

Another type of outcome is HRB, which is planned in the intention formation stage and implemented in the following stage. The actions described in the intention implementation stage relate to measurable HRB, such as visiting a health care professional or seeking advice from friends or family [29]. Depending on the purpose of app use, 2 types of behavior can be distinguished in this category.

As a result of using symptom checker apps for the purpose of receiving recommendations for further action, such recommendations are generated. This occurs frequently—in the sample of Meyer et al [7], this purpose ranked second among symptom checker app users.

Our interview partners considered whether to follow app recommendations thoroughly. If recommendations do not align with users' concepts and expectations, they are sometimes discarded as unrealistic, effectively preventing health care visits. This mismatch between concept and recommendation can also cause anxiety, creating the need for more information or confirmation by health care professionals [10,11,16]. Turner et al [50] also observed the rejection of app recommendations, especially if they were given in a context in which the action could not be implemented (eg, out of hours). Verzantvoort et al [24] found that users intended to follow app advice if the advice was to contact their GP during the daytime (75%) and practice self-care (67%). If the concept and recommendation align or if users are very anxious or surprised about the result, a visit to health care professionals is a possible outcome.

To better understand this behavior, a dialogue option could be implemented in symptom checker apps asking the following—"Have you followed my recommendation?"—and whether further resources such as health care professionals have been contacted. That way, not only symptoms but also user behavior and impact of the symptom checker app on health care use could be traced.

Communicating and documenting health-related information serves users' need to structure their experiences and prepare for contact with health care services. This sets up symptom checker apps as a possible interface between the user and health care services. This purpose has not been described yet apart from the CHECK.APP project [10]. Symptom checker apps can increase patient autonomy, as other research has postulated [17]. However, users specifically omit telling their physicians about symptom checker app use.

According to our sample, direct communication with health care professionals over app results and app use was avoided. Aboueid et al [4] found that attitudes toward physicians and the health care system may impact app use. Meyer et al [7] reported that only approximately 50% of patients using symptom checker apps considered informing their physicians about their symptom checker app use. The study indicates that users may share the output of the app, including given recommendations or probable diagnoses. However, they are less likely to inform their physicians that they used a symptom checker app to generate these results. Further research is needed to explore ways to facilitate open communication between health care professionals and symptom checker app users. The quality of the communication and its contents are suggested as measures for this outcome.

Satisfaction With App Use

A third outcome is user satisfaction with the app, which emerges during the evaluation stage. This outcome is closely linked to the app's intended purpose and the 2 previously mentioned outcomes: HRB and learning. As such, satisfaction with symptom checker apps represents a multifaceted and complex construct [13,24].

Satisfaction resulting from app use depends on how well the users perceive the symptom checker app's alignment with their needs, expectations, and intended purpose. Users become insecure and sometimes dissatisfied with the app when their concept does not match the output. This association with user satisfaction indicates that users' expectations also play a role in how they validate the app [24]. In a substudy of the CHECK.APP project, Müller et al [10] distinguished between user expectations and motivations for app use and found that these factors play a role in symptom checker app use. In fact, according to the sample in the study Kopka et al [51], 33% of symptom checker app users reported that the symptom checker

XSL•FO
RenderX

app's usefulness varied. Turner et al [50] conducted a study on an online symptom checker and found that users were dissatisfied with it because too little information on what to do themselves was provided. From their point of view, the outcome did not fit their intended purpose. This should be critical for first-time users but can also influence how users validate the app over time [51].

This indicates that simply measuring a single outcome of app use may be a shortcoming of previous research as it ignores the context of use and the intended purpose. Symptom checker apps provide a wide range of potential services that are integrated into a complex network of resources [4,5,7,13,18,20]. Therefore, researchers and developers must consider addressing the intended purpose of use and expectations and specify the desired outcome.

HL and eHL in Symptom Checker Use

Motivational Stage

The main finding of this study was that both eHL and HL are present throughout every stage of the complex process of symptom checker app use. However, different aspects of these concepts are relevant for different stages. Therefore, we will discuss the expression of eHL and HL in relation to the stages of the symptom checker app use process in the following sections.

In the motivational stage, users are driven primarily by the presence of a health-related concern and the context in which it occurs. Personal resources such as the ability to process health information and access services (as defined by the concept of HL [25]) or to navigate electronic, digital, and mobile health information (as defined by eHL [1]) play a crucial role. However, the application of these skills is influenced and shaped by the surrounding context [27,28]. The description of these skills by the interview partners can also be conceptualized as self-efficacy and self-care, which are interlinked with HL [52]. In a substudy of the CHECK.APP project, Wetzel et al [8] found that, unlike eHL, self-efficacy could be a determinant of app use—however, a sensitivity analysis revealed that the initial correlation was not viable. Kopka et al [51] found that individuals who perceived symptom checker apps as useful had higher self-efficacy. This ambiguity can be resolved by considering the context of use as an additional factor. Some symptoms may be more alarming than others and may exceed the available skills of even those with high HL [16].

Trust in apps, or the lack thereof, plays a role at the motivational stage. Regarding eHL, the interview partners in our sample positioned themselves as having high technical knowledge and the ability to critically appraise app output (eHL). They trusted the app but also their HL skills. This is consistent with the observation made by Kopka et al [11]: users with high eHL are more likely to trust symptom checker apps, possibly because they have higher confidence in their ability to critically evaluate app output. Conversely, lack of trust has been identified as a factor in intentional nonuse [9]. Neither technology affinity nor eHL are predictors that differentiate symptom checker app users from nonusers [8]. Aboueid et al [14] identified different technology affinity profiles as predictors of future app use for self-triage. Both technology affinity and trust represent attitudes rather than skills. They shape expectations and motivation to use apps but are not predictors [10].

In our sample, interview participants did not link their attitudes toward the app with the frequency of its use. Instead, their motivation was shaped by specific and meaningful experiences, such as situations in which the app provided valuable assistance with a critical symptom, subsequently confirmed by a health care professional. Research suggests that factors such as hypochondria [8] and the presence of new or unfamiliar symptoms [16] can drive app use, indicating that eHL and HL influence not only the frequency of app use but also the purposes for which symptom checker apps are used and the manner in which they are used. On the basis of these findings, we conclude that frequency of app use is an oversimplified parameter and should be interpreted with caution in future analyses.

In summary, the skill sets defined in HL and eHL, as well as context factors, influence how the motivational stage unfolds [27,28].

Intention Formation

Symptom checker apps can serve as both an information source and a tool for shaping one's HRB. Navigating digital tools for information gathering and evaluating this information is related to eHL. Applying the information to one's behavior and learning from it is more closely related to HL [24,29,30]. Therefore, planning to use symptom checker apps relates to both eHL and HL [1,25].

Users with high HL may anticipate the app's output and choose not to use it or only use it for information gathering and HISB. In addition, users may experiment with different inputs to observe changes in the output. This demonstrates the interconnection between HL and eHL during the intention formation stage. Users understand the inner workings of the app (eHL) and plan to use it to satisfy their information needs but do not intend to actually follow its advice (HL) [1,53]. According to Kopka et al [11], users are more easily persuaded to follow app recommendations when they are ambivalent or unsure about their own decision. This aligns with interview partners' self-reports that they were less inclined to follow the app's recommendations if they felt that their condition was critical or if the recommendation significantly differed from their own concept (eg, recommendation to seek emergency help when users felt perfectly healthy). In addition, using apps to communicate about symptoms connects the concepts of eHL and HL with a social context in health care [2,26].

Intention Implementation

Users use a range of resources associated with eHL and HL to achieve their intended goals. Key personal resources include the ability to evaluate one's health and navigate the health care system (HL), critically assess information from various platforms, and effectively use and understand digital applications (eHL). These resources are pivotal in shaping how symptom checker apps are integrated into users' actions [1,25,28,49]. In our sample, users expressed confidence in their eHL skills, including media literacy, information literacy, and computer literacy. However, the extent to which they used these skills

XSL•FO
RenderX

depended on their purpose and the availability of other resources.

The demand for applying interactive HL is not met by symptom checker apps because they do not allow for direct communication with others in a social context. This is particularly evident when users seek to validate their own ideas about their condition or receive recommendations for action. In such cases, users may seek out social resources and discuss their condition with friends and family. Alternatively, they may present the app results to a health care professional without disclosing their use of the app. This demonstrates that users extend app use to other contexts, integrating those contexts into their overall user experience [29,48]. This study demonstrated that HRB can transition from digital interaction to social interaction in the context of symptom checker app use. This finding aligns with the concept of distributed HL by Edwards et al [34], highlighting the importance of social networks in realizing the benefits of symptom checker apps. Real-life interaction can provide what the app interaction lacks. The ability to actively listen, notice, and respond to emotional nuances; provide support; and express understanding for the person's situation is also important. According to the observation by Wetzel et al [9], symptom checker app users still refer to friends, family, and physicians for HISB, albeit to a lesser extent than nonusers. Future research in ethics and social sciences will reveal the impact of large language models that mimic empathic responses and conversations on the use of symptom checker apps. Preliminary data suggest that such models may enhance the symptom checker app experience [54].

The interview partners considered symptom checker apps as one of the many digital resources available today. According to users, the app asked dichotomous questions, leaving them with a feeling of not being able to tell their whole story, address ambiguity, or ask questions themselves [24]. However, symptom checker apps offer a more personalized experience than internet searches [4], which remain a core component of HISB even among symptom checker app users [9]. The personal resources that enable users to use the app and apply its results also impact the integration of other digital resources into their HRB [1]. Therefore, there is potential in enhancing interoperability between different technologies such as wearables and apps.

Evaluation

The app's results, whether they provide information or recommendations, may challenge users' preexisting concepts regarding health-related issues. The extent to which users follow these recommendations depends on their preconceived notions about their condition; their expectations for its resolution; and their ability to cope with uncertainty, which is related to HL [10,25].

Users evaluate whether the app results align with the severity of their condition, their expectations of the appropriate course of action, and past experiences. This evaluation requires critical HL [25]. If users feel uncertain due to app use, they may need to contact others, such as health care professionals or their social network, for additional validation outside the app. Individuals with high levels of hypochondria [8] or anxiety [51] may be considered vulnerable groups in this context, particularly if they have low self-efficacy and HL and lack access to social or health care resources.

For the interview partners, symptom checker apps served as an interface among information gathering [4,9], learning and sensemaking [7,43], and interaction with the health care system [9,22]. In summary, the skill set necessary for meaningful interaction with symptom checker apps in the context of health care services is best described using eHL and HL. Together with contextual factors, both models can provide a deeper understanding of the dynamic mechanisms underlying symptom checker app use.

Implications for Health Care and Research With Regard to Symptom Checker App Use

Overreliance on users' HL and eHL results in wasted potential [6]. The findings of this study support the idea that, for the implementation of symptom checker apps in health care, transparency is crucial [7,12]. The lack of transparency and reliance on users' HL and eHL are design flaws in symptom checker apps, which, from a legal standpoint, only present information to be interpreted by the user who provided the input. In a worst-case scenario, users may feel isolated and insecure [3,21]. If the app was designed as a communication tool between physicians and users, such as in teleconsultations, their potential as symptom trackers and decision aids for both user groups could be realized in a safe setting. Users' insecurities could be discussed with a health professional, and possible harm, such as health anxiety, could be reduced.

The skill sets provided by eHL and HL might be the reason why users are able to use symptom checker apps to their health benefit at all despite symptom checker apps' questionable diagnostic accuracy [18,38,55]. Concrete and meaningful experiences with the app shaped users' attitudes toward and evaluation of the app. A pure cause-and-effect relationship, as suggested by its user interface principle (input–algorithm or artificial intelligence–output), is prevented by users' HL, self-efficacy, and self-care. Satisfaction with app use is influenced by users' HL, context, and experiences [10,13,24,29]. Symptom checker app development and research should take into account user expectations, context of use, and the need for social interaction.

Our study uncovered phenomena that require further investigation in the future. These include symptom checker app use in various social settings [24]. Our results show that users frequently engage with symptom checker apps not only for their own health concerns but also to assess symptoms of family members and friends. This social use—such as evaluating a partner's symptoms or a parent using the app for their child—occurs more often than previously reported in the literature. Different technological resources such as wearables, other apps, and internet searches were perceived as possible contributors to the 3 purposes by the interview partners. Thus, research on the topic should consider the entire landscape of digital resources and their interoperability.

Limitations

Experience-based research is limited to the information that interview participants are willing to share about themselves. It

XSL•FO
RenderX

is not possible to directly observe the psychological processes that occur in someone's mind. However, by collecting and synthesizing users' stories and perspectives, we were able to describe a general process. The Rubicon model helped us organize the stories coherently and with a focus on motivation.

It is important to critically evaluate the method and sample used. It should be noted that, even in our sample of predominantly younger, White, well-educated female users, each individual had their own perspective on how the app fulfilled its purpose in the described contexts. While a general process could be derived, the outcomes were highly dependent on the user. In addition, the results are primarily applicable to the context of the German health care system and, therefore, can only be transferred to similar health care systems, which is a common challenge in symptom checker app research [38].

Our sample consisted of predominantly very reflective individuals with presumably high HL. Our sampling strategy to approach users based on diverse characteristics could only partly be implemented because the sample from which we recruited the interview partners represented the current major user group of symptom checker apps: young, female, and well educated individuals [5,9,17]. Thus, the sample reflects the most typical characteristics of symptom checker app users, indicating that the results are likely applicable to this core group [3,9].

While the mixed methods approach of the CHECK.APP project allowed for the description of this population from different perspectives using different methods, it limits the applicability of our results to individuals with less HL. We can neither make assumptions about nor extend our findings to marginalized groups excluded from app use [15,56].

The integrative basic method builds on the tradition of interpretative-reconstructive qualitative methods such as ethnomethodological conversation analysis and narrative analysis. It proved to be suitable to analyze how medical laypersons make sense of the symptom checker app. While we could show in another study of the CHECK.APP project how GPs position symptom checker apps in the black box of the "unorganized stage" of patients' reflections on their symptoms [40], our study succeeded in shedding light on this exact black box.

Another limitation of this study is the focus on just 1 app. While we believe to have extrapolated generalizable information, it is possible that other apps may yield different user experiences and human-app interactions. In addition, due to the pandemic, the interviews were conducted on the web via videoconferencing software. It is possible that conducting in-person interviews would have attracted different interviewees and revealed additional information due to the different setting.

While software-supported tools such as MAXQDA or f4analyse exist to support qualitative analysis and the authors and analysts are well acquainted with them, we deliberately chose not to use these tools especially for the interpretative parts of the analysis because their workflow is optimized for categorization of qualitative data, not so much for in-depth interpretative analyses. To ensure that our operationalization remained understandable, we provided data examples in the results text.

Conclusions

In our qualitative interview study, we could demonstrate that, from symptom checker app users' perspective, a simple cause-and-effect relationship between symptoms and symptom checker app use is unlikely, at least in individuals with high eHL and HL. Rather, symptom checker app use is described as a complex, cyclic process. Context-dependent and biographical factors, as well as the dynamic concepts of eHL and HL, were expressed in users' descriptions.

According to our limited sample, symptom checker apps are used for 3 distinct purposes: understanding one's condition, receiving recommendations for action, and communicating on and documenting health-related information. Each purpose warrants its own planning, implementation, and evaluation in HRB. Each purpose may be implemented using different personal, social, and technological resources.

Symptom checker apps have shortcomings with respect needs related to interactive HL. Users seek external validation of app findings in their social networks and professional health care services. However, symptom checker apps have the potential to become an interface between users and health care services. This potential has not been realized, which should be a design goal for their continued development.

Acknowledgments

This study was funded by the Federal Ministry of Education and Research (grant 01GP1907A). The sponsor had no influence on the study design, data collection, or its analysis or interpretation. The authors thank several of their interview partners (who will remain anonymous for privacy reasons) for their contribution to the member check workshop in which the aforementioned results were presented to and discussed with the participants. The generative artificial intelligence tool ChatGPT by OpenAI was used to revise the language and style of the manuscript, which were further reviewed and revised by the study group. The authors acknowledge support from the open access publication fund of the University of Tübingen.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request. Due to data protection regulations, interview data can only be made available if the study participant associated with the data approves the request.

XSL•FO
RenderX

Authors' Contributions

RK drafted the manuscript together with MTS. MK, RM, RK, and AJW conducted the interviews. CP provided vital methodological input for analysis. MTS, RK, AJW, MK, and CP met regularly in multidisciplinary methodological workshops, which were led by CP. All listed authors revised the manuscript. This process was coordinated by RK. All authors approved the final manuscript draft before submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Standards for Reporting Qualitative Research statement checklist.
[PDF File (Adobe PDF File), 226 KB - formative_v9i1e60647_app1.pdf]

Multimedia Appendix 2
Interview guide.
[DOCX File , 24 KB - formative_v9i1e60647_app2.docx]

References

1. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
2. Levin-Zamir D, Bertschi I. Media health literacy, eHealth literacy, and the role of the social environment in context. Int J Environ Res Public Health 2018 Aug 03;15(8):1643 [FREE Full text] [doi: 10.3390/ijerph15081643] [Medline: 30081465]
3. Müller R, Klemmt M, Ehni HJ, Henking T, Kuhnmünch A, Preiser C, et al. Ethical, legal, and social aspects of symptom checker applications: a scoping review. Med Health Care Philos 2022 Dec;25(4):737-755 [FREE Full text] [doi: 10.1007/s11019-022-10114-y] [Medline: 36181620]
4. Aboueid S, Meyer S, Wallace JR, Mahajan S, Chaurasia A. Young adults' perspectives on the use of symptom checkers for self-triage and self-diagnosis: qualitative study. JMIR Public Health Surveill 2021 Jan 06;7(1):e22637 [FREE Full text] [doi: 10.2196/22637] [Medline: 33404515]
5. Aboueid S, Liu RH, Desta BN, Chaurasia A, Ebrahim S. The use of artificially intelligent self-diagnosing digital platforms by the general public: scoping review. JMIR Med Inform 2019 May 01;7(2):e13445 [FREE Full text] [doi: 10.2196/13445] [Medline: 31042151]
6. Aboueid S, Meyer SB, Wallace JR, Mahajan S, Nur T, Chaurasia A. Use of symptom checkers for COVID-19-related symptoms among university students: a qualitative study. BMJ Innov 2021 Apr;7(2):253-260. [doi: 10.1136/bmjinnov-2020-000498] [Medline: 34192014]
7. Meyer AN, Giardina TD, Spitzmueller C, Shahid U, Scott TM, Singh H. Patient perspectives on the usefulness of an artificial intelligence-assisted symptom checker: cross-sectional survey study. J Med Internet Res 2020 Jan 30;22(1):e14679 [FREE Full text] [doi: 10.2196/14679] [Medline: 32012052]
8. Wetzel AJ, Klemmt M, Müller R, Rieger MA, Joos S, Koch R. Only the anxious ones? Identifying characteristics of symptom checker app users: a cross-sectional survey. BMC Med Inform Decis Mak 2024 Jan 23;24(1):21 [FREE Full text] [doi: 10.1186/s12911-024-02430-5] [Medline: 38262993]
9. Wetzel AJ, Koch R, Koch N, Klemmt M, Müller R, Preiser C, et al. 'Better see a doctor?' Status quo of symptom checker apps in Germany: a cross-sectional survey with a mixed-methods design (CHECK.APP). Digit Health 2024;10:20552076241231555 [FREE Full text] [doi: 10.1177/20552076241231555] [Medline: 38434790]
10. Müller R, Klemmt M, Koch R, Ehni HJ, Henking T, Langmann E, et al. "That's just Future Medicine" - a qualitative study on users' experiences of symptom checker apps. BMC Med Ethics 2024 Feb 16;25(1):17 [FREE Full text] [doi: 10.1186/s12910-024-01011-5] [Medline: 38365749]
11. Kopka M, Schmieding ML, Rieger T, Roesler E, Balzer F, Feufel MA. Determinants of laypersons' trust in medical decision aids: randomized controlled trial. JMIR Hum Factors 2022 May 03;9(2):e35219 [FREE Full text] [doi: 10.2196/35219] [Medline: 35503248]
12. Miller S, Gilbert S, Virani V, Wicks P. Patients' utilization and perception of an artificial intelligence-based symptom assessment and advice technology in a British primary care waiting room: exploratory pilot study. JMIR Hum Factors 2020 Jul 10;7(3):e19713 [FREE Full text] [doi: 10.2196/19713] [Medline: 32540836]
13. Chambers D, Cantrell AJ, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open 2019 Aug 01;9(8):e027743 [FREE Full text] [doi: 10.1136/bmjopen-2018-027743] [Medline: 31375610]
14. Aboueid S, Meyer SB, Wallace J, Chaurasia A. Latent classes associated with the intention to use a symptom checker for self-triage. PLoS One 2021;16(11):e0259547 [FREE Full text] [doi: 10.1371/journal.pone.0259547] [Medline: 34731217]

XSL•FO
RenderX

15. Luger TM, Houston TK, Suls J. Older adult experience of online diagnosis: results from a scenario-based think-aloud protocol. J Med Internet Res 2014 Jan 16;16(1):e16 [FREE Full text] [doi: 10.2196/jmir.2924] [Medline: 24434479]
16. Wetzel AJ, Preiser C, Müller R, Joos S, Koch R, Henking T, et al. Unveiling usage patterns and explaining usage of symptom checker apps: explorative longitudinal mixed methods study. J Med Internet Res 2024 Dec 09;26:e55161 [FREE Full text] [doi: 10.2196/55161] [Medline: 39652845]
17. Arellano Carmona K, Chittamuru D, Kravitz RL, Ramondt S, Ramírez AS. Health information seeking from an intelligent web-based symptom checker: cross-sectional questionnaire study. J Med Internet Res 2022 Aug 19;24(8):e36322 [FREE Full text] [doi: 10.2196/36322] [Medline: 35984690]
18. Wallace W, Chan C, Chidambaram S, Hanna L, Iqbal FM, Acharya A, et al. The diagnostic and triage accuracy of digital and online symptom checker tools: a systematic review. NPJ Digit Med 2022 Aug 17;5(1):118 [FREE Full text] [doi: 10.1038/s41746-022-00667-w] [Medline: 35977992]
19. Schmieding ML, Kopka M, Schmidt K, Schulz-Niethammer S, Balzer F, Feufel MA. Triage accuracy of symptom checker apps: 5-year follow-up evaluation. J Med Internet Res 2022 May 10;24(5):e31810 [FREE Full text] [doi: 10.2196/31810] [Medline: 35536633]
20. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 08;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]
21. Pairon A, Philips H, Verhoeven V. A scoping review on the use and usefulness of online symptom checkers and triage systems: how to proceed? Front Med (Lausanne) 2022 Jan 6;9:1040926 [FREE Full text] [doi: 10.3389/fmed.2022.1040926] [Medline: 36687416]
22. Ceney A, Tolond S, Glowinski A, Marks B, Swift S, Palser T. Accuracy of online symptom checkers and the potential impact on service utilisation. PLoS One 2021;16(7):e0254088 [FREE Full text] [doi: 10.1371/journal.pone.0254088] [Medline: 34265845]
23. Knitza J, Muehlensiepen F, Ignatyev Y, Fuchs F, Mohn J, Simon D, et al. Patient's perception of digital symptom assessment technologies in rheumatology: results from a multicentre study. Front Public Health 2022;10:844669 [FREE Full text] [doi: 10.3389/fpubh.2022.844669] [Medline: 35273944]
24. Verzantvoort NC, Teunis T, Verheij TJ, van der Velden AW. Self-triage for acute primary care via a smartphone application: practical, safe and efficient? PLoS One 2018;13(6):e0199284 [FREE Full text] [doi: 10.1371/journal.pone.0199284] [Medline: 29944708]
25. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Promot Int 2000;15(3):259-267. [doi: 10.1093/heapro/15.3.259]
26. Nutbeam D, Lloyd JE. Understanding and responding to health literacy as a social determinant of health. Annu Rev Public Health 2021 Apr 01;42:159-173 [FREE Full text] [doi: 10.1146/annurev-publhealth-090419-102529] [Medline: 33035427]
27. Parker RM, Ratzan S. Re-enforce, not re-define health literacy-moving forward with health literacy 2.0. J Health Commun 2019 Nov 21;24(12):923-925. [doi: 10.1080/10810730.2019.1691292] [Medline: 31752603]
28. Parker R, Ratzan SC. Health literacy: a second decade of distinction for Americans. J Health Commun 2010 Aug 31;15 Suppl 2(sup2):20-33. [doi: 10.1080/10810730.2010.501094] [Medline: 20845190]
29. von Wagner C, Steptoe A, Wolf MS, Wardle J. Health literacy and health actions: a review and a framework from health psychology. Health Educ Behav 2009 Oct;36(5):860-877. [doi: 10.1177/1090198108322819] [Medline: 18728119]
30. Neter E, Brainin E. Association between health literacy, eHealth literacy, and health outcomes among patients with long-term conditions. Eur Psychol 2019 Jan;24(1):68-81. [doi: 10.1027/1016-9040/a000350]
31. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. Am J Health Behav 2007;31 Suppl 1:S19-S26. [doi: 10.5555/ajhb.2007.31.supp.S19] [Medline: 17931132]
32. Stormacq C, Van den Broucke S, Wosinski J. Does health literacy mediate the relationship between socioeconomic status and health disparities? Integrative review. Health Promot Int 2019 Oct 01;34(5):e1-17. [doi: 10.1093/heapro/day062] [Medline: 30107564]
33. Stormacq C, Wosinski J, Boillat E, Van den Broucke S. Effects of health literacy interventions on health-related outcomes in socioeconomically disadvantaged adults living in the community: a systematic review. JBI Evid Synth 2020 Jul;18(7):1389-1469. [doi: 10.11124/JBISRIR-D-18-00023] [Medline: 32813388]
34. Edwards M, Wood F, Davies M, Edwards A. 'Distributed health literacy': longitudinal qualitative analysis of the roles of health literacy mediators and social networks of people living with a long-term health condition. Health Expect 2015 Oct;18(5):1180-1193 [FREE Full text] [doi: 10.1111/hex.12093] [Medline: 23773311]
35. Sørensen K, Pelikan JM, Röthlin F, Ganahl K, Slonska Z, Doyle G, et al. Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). Eur J Public Health 2015 Dec 05;25(6):1053-1058 [FREE Full text] [doi: 10.1093/eurpub/ckv043] [Medline: 25843827]
36. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
37. Fiske A, Buyx A, Prainsack B. The double-edged sword of digital self-care: physician perspectives from Northern Germany. Soc Sci Med 2020 Sep;260:113174 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113174] [Medline: 32659512]

XSL•FO
RenderX

38. Ilicki J. Challenges in evaluating the accuracy of AI-containing digital triage systems: a systematic review. PLoS One 2022 Dec 27;17(12):e0279636 [FREE Full text] [doi: 10.1371/journal.pone.0279636] [Medline: 36574438]
39. Wetzel AJ, Koch R, Preiser C, Müller R, Klemmt M, Ranisch R, et al. Ethical, legal, and social implications of symptom checker apps in primary health care (CHECK.APP): protocol for an interdisciplinary mixed methods study. JMIR Res Protoc 2022 May 16;11(5):e34026. [doi: 10.2196/34026]
40. Preiser C, Radionova N, Ög E, Koch R, Klemmt M, Müller R, et al. The doctors, their patients, and the symptom checker app: qualitative interview study with general practitioners in Germany. JMIR Hum Factors 2024 Nov 18;11:e57360 [FREE Full text] [doi: 10.2196/57360] [Medline: 39556813]
41. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
42. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov 10;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
43. Bruner J. Acts of Meaning. Cambridge, MA: Harvard University Press; 1990.
44. Kolb DA. Experiential Learning: Experience as the Source of Learning and Development. Hoboken, NJ: Prentice-Hall; 1984.
45. Kruse J. Qualitative Interviewforschung: Ein integrativer Ansatz. Weinheim, Germany: Beltz Juventa; 2014.
46. Davies B, Harré R. Positioning: the discursive production of selves. J Theory Soc Behav 2007 Jul;20(1):43-63. [doi: 10.1111/j.1468-5914.1990.tb00174.x]
47. McAdams DP. Narrative identity. In: Schwartz SJ, Luyckx K, Vignoles VL, editors. Handbook of Identity Theory and Research. Berlin, Germany: Springer Science + Business Media; 2011:99-115.
48. Achtziger A, Gollwitzer PM. Motivation and volition in the course of action. In: Heckhausen J, Heckhausen H, editors. Motivation and Action. Cambridge, UK: Cambridge University Press; 2008:272-295.
49. Parker R. Measuring health literacy: what? So what? Now what? In: Measures of Health Literacy: Workshop Summary. Washington, DC: National Academies Press (US); 2009:91-102.
50. Turner J, Knowles E, Simpson R, Sampson F, Dixon S, Long J, et al. Impact of NHS 111 Online on the NHS 111 telephone service and urgent care system: a mixed-methods study. Health Serv Deliv Res 2021 Nov;9(21):1-148. [doi: 10.3310/hsdr09210] [Medline: 34780129]
51. Kopka M, Scatturin L, Napierala H, Fürstenau D, Feufel MA, Balzer F, et al. Characteristics of users and nonusers of symptom checkers in Germany: cross-sectional survey study. J Med Internet Res 2023 Jun 20;25:e46231 [FREE Full text] [doi: 10.2196/46231] [Medline: 37338970]
52. Lee EH, Lee YW, Moon SH. A structural equation model linking health literacy to self-efficacy, self-care activities, and health-related quality of life in patients with type 2 diabetes. Asian Nurs Res (Korean Soc Nurs Sci) 2016 Mar;10(1):82-87 [FREE Full text] [doi: 10.1016/j.anr.2016.01.005] [Medline: 27021840]
53. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
54. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023 Jun 01;183(6):589-596 [FREE Full text] [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
55. Schmieding ML, Mörgeli R, Schmieding MA, Feufel MA, Balzer F. Benchmarking triage capability of symptom checkers against that of medical laypersons: survey study. J Med Internet Res 2021 Mar 10;23(3):e24475 [FREE Full text] [doi: 10.2196/24475] [Medline: 33688845]
56. Wang X, Luan W. Research progress on digital health literacy of older adults: a scoping review. Front Public Health 2022;10:906089 [FREE Full text] [doi: 10.3389/fpubh.2022.906089] [Medline: 35991040]

Abbreviations

CHECK.APP: Ethical, Legal, and Social Implications of Symptom Checker Apps in Primary Health Care
eHL: eHealth literacy
GP: general practitioner
HISB: health information–seeking behavior
HL: health literacy
HRB: health-related behavior

XSL•FO
RenderX

Edited by A Mavragani; submitted 17.05.24; peer-reviewed by S Saur, SC Chong; comments to author 25.08.24; revised version received 28.11.24; accepted 08.02.25; published 21.03.25.

Please cite as:
Koch R, Steffen MT, Wetzel AJ, Preiser C, Klemmt M, Ehni HJ, Mueller R, Joos S
Exploring Laypersons' Experiences With a Mobile Symptom Checker App as an Interface Between eHealth Literacy, Health Literacy, and Health-Related Behavior: Qualitative Interview Study
JMIR Form Res 2025;9:e60647
URL: https://formative.jmir.org/2025/1/e60647
doi:10.2196/60647
PMID:

©Roland Koch, Marie-Theres Steffen, Anna-Jasmin Wetzel, Christine Preiser, Malte Klemmt, Hans-Jörg Ehni, Regina Mueller, Stefanie Joos. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring Physicians' Dual Perspectives on the Transition From Free Text to Structured and Standardized Documentation Practices: Interview and Participant Observational Study

Olga Golburean[1], MPhil; Rune Pedersen[2], PhD; Line Melby[3], PhD; Arild Faxvaag[1], PhD

[1]Department of Neuromedicine and Movement Science, Faculty of Medicine and Health Sciences, NTNU-Norwegian University of Science and Technology, Trondheim, Norway

[2]Norwegian Centre for E-Health Research, University Hospital of North Norway, Tromsø, Norway

[3]Department of Health Research, SINTEF Digital, Trondheim, Norway

Corresponding Author:
Olga Golburean, MPhil
Department of Neuromedicine and Movement Science
Faculty of Medicine and Health Sciences
NTNU-Norwegian University of Science and Technology
Elgeseter Gate 28
Trondheim, 7030
Norway
Phone: 47 91787389
Email: olga.golburean@ntnu.no

Abstract

Background: Clinical documentation plays a crucial role in providing and coordinating care. Despite the widespread adoption of electronic health record (EHR) systems, many end users still document clinical data in a manner similar to traditional paper-based records. To fully leverage the benefits of EHR systems, it is necessary to adopt new documentation approaches that facilitate easy access to information at the point of care and seamless exchange of information across health care facilities.

Objective: We aimed to evaluate how the transition from an older EHR system to a cross-institutional EHR system impacts physicians' documentation practices and gain a deeper understanding of the factors influencing their choice between free text and structured and standardized documentation methods.

Methods: A qualitative study was conducted between September 2023 and January 2024. It involved participant observations and individual semistructured interviews with physicians at a university hospital in Norway. Data were analyzed using reflexive thematic analysis.

Results: The analysis revealed 3 main themes. First, physicians encountered challenges during the implementation phase of the new EHR system due to its complexity and their unfamiliarity with its use. However, with time, physicians gradually adopted new documentation processes. This integration or adoption primarily occurred by learning through practical experience and collaborative knowledge exchange with their peers. Second, although the implementation of the new EHR system had increased structured and standardized clinical documentation, free text remained the preferred method, with some exceptions. In addition, the fact that many physicians still relied on free-text documentation created a sense of distrust among them toward some of the standardized clinical data. Finally, the informants had mixed perceptions of Systematized Nomenclature of Medicine–Clinical Terms. Some viewed it as a more nuanced terminology system, while others found it more complex. Most informants found using templates for routine procedures beneficial as it saved time in the documentation process and ensured that all necessary parameters and documentation requirements were met.

Conclusions: The study findings revealed that physicians' acceptance of new documentation processes is influenced by various social and technological factors. These factors include previous documentation experiences, perceived benefits, familiarity with the EHR system, time constraints, and user-friendliness of the system. While physicians generally have a positive attitude toward using templates for routine procedures, they often create their own templates, and data within these templates are documented in a free-text format. To address this, health care organizations should consider implementing common standardized or semistandardized templates to reduce disparities in documentation, enhance data recording, and ensure adherence to guidelines.

XSL•FO
RenderX

Furthermore, to facilitate the transition to the new documentation processes, we recommend providing physicians with customized training programs and platforms for tacit knowledge exchange.

(JMIR Form Res 2025;9:e63902) doi:10.2196/63902

KEYWORDS

documentation; documentation practice; standardized documentation; structured documentation; secondary use of data; interoperability; electronic health record; EHR

Introduction

Background

Electronic health record (EHR) systems have been advocated as means to enhance the efficiency of health care services, facilitating improved coordination and delivery of safe and evidence-based care [1,2]. Data from an EHR system can also provide health professionals with the information they need to make informed decisions at the point of care and the opportunity to reflect on their clinical work and learn from past experiences [3]. In addition, an EHR system facilitates access to data collected during clinical care, which may be used for secondary purposes (eg, research and augmenting the capabilities of clinical decision support systems) [4-6]. Nevertheless, clinical data need to be captured in a fixed structure in order to facilitate their use [4]. Health professionals play a crucial role in the documentation process, facilitating the accumulation of data in the EHR system. However, the implementation of the EHR system often results in end users still recording clinical data in a manner reminiscent of traditional paper-based records, although in a digital format, using computers and keyboards [7]. Entering structured and standardized data into an EHR system can be time-consuming, and the level of acceptance among different end users can vary greatly. Health professionals, frequently constrained by time limitations, might be reluctant to assume the data entry task unless they see significant benefits [8,9].

Clinical Documentation in EHR Systems

Health care systems are bound by stringent laws and regulations. In Norway, the use of electronic health data for the provision of health care services (ie, the primary use of health data) is regulated by the Patient Record Act [10]. This act aims to ensure that patients receive high-quality health care by making health data available for health professionals in a quick, effective, and secure manner. Health professionals, on their side, have a duty to meet documentation requirements as outlined in the Health Personnel Act [11]. This act specifically mandates that patient journals must be maintained in accordance with high professional standards, ensuring that they include pertinent and essential information about the patient and the health care services rendered. Furthermore, the journals must be easily comprehensible for other qualified health care personnel. The use of electronic health data for purposes such as statistics, public health, health surveillance, research and product development, education, or teaching (ie, the secondary use of health data) is also strictly regulated [10,12].

The adoption of the EHR system has significantly expanded the volume of clinical data available. However, a considerable portion of these data remains underused, largely due to the dominance of unstructured documentation, such as free-text notes and narrative reports [7]. It is estimated that 80% of the EHR data exist in unstructured documents [13,14]. Consequently, there is increasing pressure on health professionals to adopt structured and standardized documentation methods [6]. This involves the use of standardized coding systems for data entry, the incorporation of models and terminologies such as Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT), and the use of structured forms [6,15]. Structured and standardized data encompass various patient details, including demographics, laboratory tests, height, weight, blood pressure, medications, allergies, and more [16,17]. Since structured and standardized data types are organized within a fixed structure, they can be easily analyzed using statistical or machine learning methods [16]. Furthermore, data from EHR systems also play a crucial role in the realization of learning health systems, where data are captured as a by-product of care. These data are subsequently analyzed to generate new knowledge, which, in turn, informs continuous improvement and innovation in health care service delivery [18,19]. While on the one hand, structured and standardized documentation practices are associated with enhanced quality of clinical notes [20,21], on the other hand, research highlights that the adoption of the EHR system has led to increased documentation time and reduced attention to patient care [6]. Moreover, structured and standardized documentation can impede expressivity in notes. Free-text documentation is crucial in health care delivery as it allows health professionals to record detailed and nuanced patient narratives. Thus, Rosenbloom et al [22] advocated that health professionals should have the flexibility to decide their documentation style according to workflow needs and content specifics. Their study highlighted that structured documentation is ideal for scenarios requiring data reuse, whereas narrative documentation is suitable when data reuse is not a priority. Furthermore, Levy et al [23] emphasized that health professionals encounter an excessive documentation burden when the usability of the documentation systems fails to adequately support patient care delivery. This highlights the crucial role of systems' usability and the need to assess their impact on documentation practices.

The aim of this study is to evaluate how the transition from an older EHR system to a newer cross-institutional EHR system, which enables structured and standardized documentation, impacts physicians' documentation practices. It seeks to gain insights into the factors that impact physicians' choice of documentation methods, such as free text or structured and standardized documentation.

XSL•FO
RenderX

Theory

Standardizing work processes in health care, such as implementing standardized guidelines for procedures and treatments, is regarded as a useful approach to improving the quality of health care due to its ability to improve outcomes, reduce the likelihood of errors, and eliminate unnecessary variations [24-26]. The increased adoption of EHR systems has led to increased interest in integrating standardized clinical tools within these systems. This includes, among others, standards pertaining to clinical documentation, encompassing what, where, and how information should be documented (eg, structured text, use of templates, and narrative text) [27-30]. Standardizing the use of the EHR system contributes to achieving interoperable EHR system, enabling the seamless exchange of clinical data among various health care organizations [31-33]. Nevertheless, the pursuit of standardization in health care, particularly in the realm of EHR systems, has proven to be challenging [31,34]. This is due to the highly heterogeneous nature of health care, marked by the constant occurrence of unforeseen circumstances. A significant portion of health care professionals' responsibilities cannot be strictly defined by standardized procedures alone, as a substantial amount of their work is tailored to address the unique requirements of individual patients [35,36]. Thus, it has been suggested that providing flexibility is essential in health care work, as it enables health care professionals to customize their work to factors such as patient needs, good judgment, workflow, and prevailing circumstances [24]. Cherns [37] further emphasized the importance of incorporating sufficient flexibility within technical systems. This entails the ability to tailor the system according to specific local needs and adapt to evolving requirements [33,37]. This is justifiable, as Orlikowski [38] highlighted the dynamic nature of technologies, which are continually shaped by human responses to a multitude of influences, such as environmental, political, and cultural factors. Nevertheless, it is important to note that people have their own distinct viewpoints, "practice lenses" that impact how they use the systems [33,38], and may also rely on their own meanings and interpretations of the technology's functions, known as "interpretative flexibility," which may deviate from the designer's original intentions [39,40]. Considering the aforementioned information, it is notable that the implementation of EHR systems and the standardized processes they encompass (eg, clinical documentation) extend beyond being mere technological tools. Instead, EHR systems are integral components of a sociotechnical system, which encompasses a dynamic interaction between technology, health professionals, organizational structures, and the social environment in which they are embedded [41-43]. Thus, literature has previously emphasized the significance of using the sociotechnical system theory [44] in the design, development, implementation, use, and evaluation of health information technology [45] and applied it to examine its impact on documentation practices [46].

The Context

Although all health care organizations in Norway use an EHR system, many of these systems are not adequately designed for structured and standardized documentation [47]. The daily documentation primarily comprises unstructured notes written in free text, which makes it difficult to use for secondary purposes, such as quality improvements. Moreover, in many cases, medications are manually recorded and signed on paper-based charts before being scanned as images into the EHR system [47]. According to Norwegian health authorities, health professionals lack proper access to patient health information when needed at the point of care [48]. This challenge is primarily attributed to the fragmented nature of the health care systems, coupled with the insufficient integration of various information and communication technology systems [48]. Thus, in 2012, the Norwegian government launched the "One citizen—one health record" initiative. Its goal is to facilitate the seamless integration and accessibility of essential health information at every stage of patient care, regardless of when or where it is required [48]. As part of this initiative, the Health Platform (Helseplattformen) project was established [49], with the aim to replace the existing fragmented EHR systems with an integrated solution for both primary and specialist health professionals in the entire central Norway region [50,51]. The project selected Epic Systems Corporation as the vendor for the EHR system [49].

Health Platform aimed to shift toward more structured and standardized documentation of clinical data in order to achieve optimal system functionality and make data machine readable. A pivotal change within the Health Platform project is the transition from the *International Classification of Diseases, 10th Revision* (*ICD-10*) to the more comprehensive and nuanced terminology system, SNOMED CT [52]. The complexity of SNOMED CT arises from a dense hierarchical structure and "is-a" relationships that link broader and more specific concepts, as explored by Abeysinghe et al [53]. This allows SNOMED CT to capture nuances that provide better clinical coverage, making it more suitable for the EHR system and supporting patient care, compared to *ICD*, which is mainly used for diagnoses and billing [54]. The lack of detail in *ICD* codes can lead to a loss of specificity in a clinical context, which means that rare conditions can be grouped under a single code, thus impacting the EHR system's ability to provide accurate clinical decision support [54].

Methods

Study Design and Settings

We used a descriptive qualitative research method to conduct this case study based on participant observations and individual semistructured interviews with physicians. Qualitative research methods enable researchers to delve deeply into the subject matter, providing comprehensive insights and detailed descriptions from the participants' perspectives. Unlike quantitative methods, qualitative approaches, such as interviews and observations, excel at offering detailed explanations, particularly when investigating questions pertaining to *why* and *how* [35,55].

We conducted this study in 1 surgical department at a large university hospital in the central Norway region, which transitioned from a facility-centered EHR system to Health Platform, the new cross-institutional EHR system (Epic Systems Corporation), in November 2022. The university hospital houses

XSL•FO
RenderX

approximately 1000 beds and employs around 11,000 staff members. The surgical department includes an outpatient clinic, inpatient ward, surgical unit, and emergency room, complemented by a specialized division for pediatrics.

Recruitment of Participants

Participants for this study were recruited through verbal and written approaches. First, an email was sent to all physicians working within the surgical department, requesting their participation in an observational study. Second, recognizing the challenges of recruiting busy physicians, particularly during the transition to the new EHR system, the first author (OG) attended the physicians' morning meeting to provide additional information about the project and give them the opportunity to ask direct questions about the study. The study aimed to recruit participants with varying levels of professional experience, including experienced physicians and interns, to ensure a diverse range of perspectives. After completing the observational study phase, a subsequent email invitation was sent to the same group of individuals (ie, physicians in the surgical department), inviting them to participate in individual interviews. The first author (OG) reached out to physicians interested in participating in the study to schedule a convenient time and location for their meeting. In total, 14 physicians participated in the study, and the characteristics of the participants are summarized in Table 1.

Table 1. Demographic characteristics of the study participants (N=14).

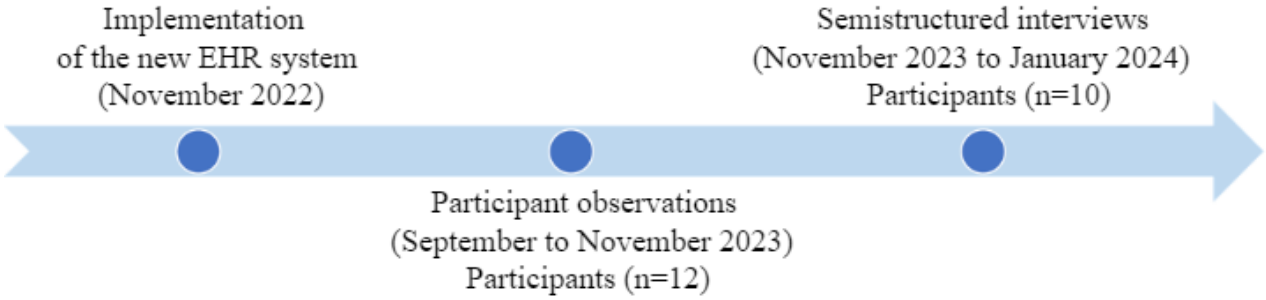
Characteristics	Participants, n (%)
Age group (y)	
25-39	9 (64)
≥40	5 (36)
Sex	
Female	6 (43)
Male	8 (57)
Practice (y)	
<5	3 (21)
5-15	7 (50)
>15	4 (29)

Data Collection

Observations

Participant observations occurred from September 2023 to November 2023. In total, 12 physicians were observed for a combined duration of approximately 44 hours. Figure 1 depicts the information regarding data collection activities. The observation sessions ranged from 1 to 3 times, with each session varying in duration from 1 hour to 3 hours. The observation was conducted at the surgical department at several health care facilities where the physicians were working, including an outpatient clinic, an inpatient ward, and an emergency room. Before these observations, a taxonomy was developed to guide the observational process. It was developed based on the research question, a review of existing literature [56-58], and collaborative inputs from field experts (AF, LM, and RP). The primary focus during observations was on physicians' interactions with the new EHR system, with particular attention given to their documentation practices. Throughout each observational session, the observer (OG) recorded notes, capturing both the physicians' activities and, when feasible, engaging them in dialogue about their EHR use. Following observations, the observer supplemented these notes with additional remarks and personal reflections.

Figure 1. Data collection activities. EHR: electronic health record.



Interviews

After the completion of the observations, individual, semistructured interviews were conducted face-to-face with the participants by the first author (OG). In total, 10 physicians were interviewed, 8 (80%) of whom had also been part of previous observations. The interviews were conducted between November 2023 and January 2024 (Figure 1). The interviews

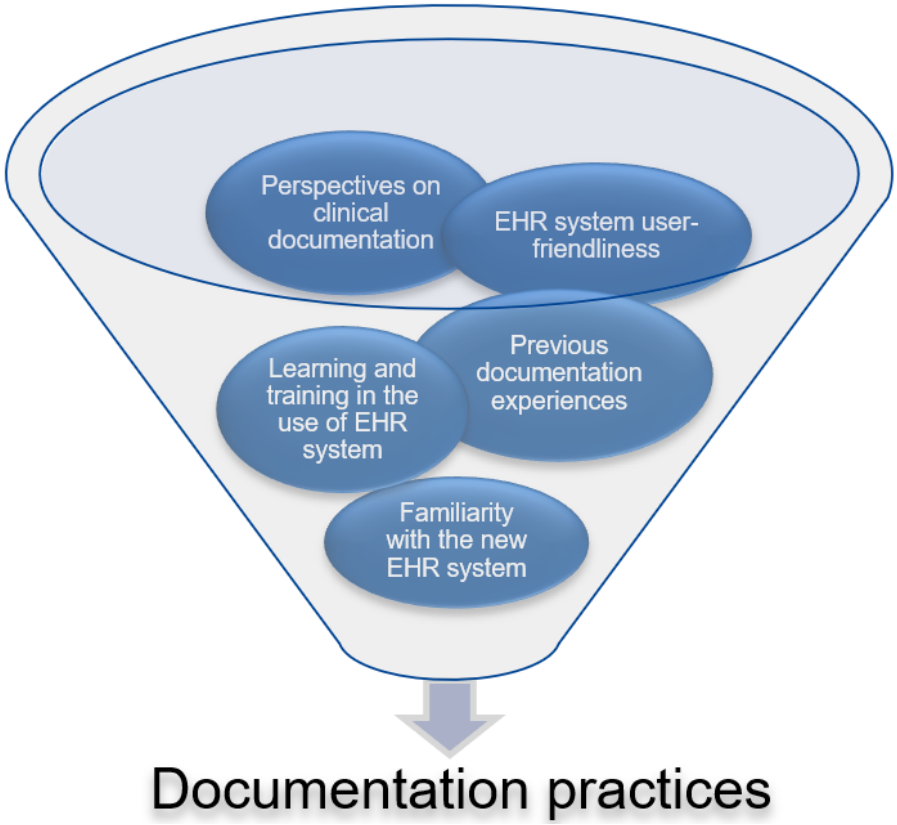
XSL•FO
RenderX

lasted between 34 and 60 minutes, with an average duration of 44 (SD 9.6) minutes. The interview guide, developed based on the existing literature [58-60] and expert collaboration in the field, was initially outlined before the observational phase. However, following the observations, which provided in-depth insights into the physicians' work practices with the new EHR system, the interview guide was revised to incorporate emerging, relevant questions. Multimedia Appendix 1 provides the interview guide. Each interview was audio recorded and transcribed verbatim by the first author (OG). The process of transcription began immediately after the first interview and was integral to the preliminary data analysis phase, which involved a thorough familiarization with the collected data. The interview phase concluded once data saturation was achieved. This happened when further interviews no longer provided any new insights or themes.

Data Analysis

Data were organized using NVivo (Lumivero) software [61] and analyzed through a reflexive thematic analysis approach [62]. The transcripts of the interviews and field notes from the observation sessions were thoroughly examined to familiarize them with the data. Then, specific codes were developed using an inductive approach, meaning that the codes were derived from the data. The initial codes were created using a semantic coding approach. These initial codes were reviewed, leading to the formation of initial themes. Throughout this process, we continuously reviewed and revised the raw data, codes, and initial themes to ensure consistency and coherence. Moreover, we conducted a 2-day workshop where authors collaboratively reviewed the preliminary themes and their associated codes. During this workshop, we used latent analyses to gain a deeper understanding of the data within the themes. Specifically, we delved into the factors that influenced physicians' documentation practices in the new EHR system using the sociotechnical system theory [44]. These factors encompassed various aspects, such as implications of learning and training in the use of EHR system, physicians' perspectives and previous experiences with clinical documentation, familiarity with the new EHR system, and the influence of technology (new EHR system) and its user-friendliness on physicians' documentation practices (Figure 2). The research team involved in the data analysis possessed diverse backgrounds that enriched the interpretation of the data. The researchers included experts in health informatics, medicine, qualitative research, and implementation science. This multidisciplinary perspective facilitated a comprehensive understanding of the complexities inherent in the data collected. Using the workshop notes, the first author (OG) further refined the themes. This phase served as the basis for drafting the initial report. The authors adhered to the SRQR (Standards for Reporting Qualitative Research) checklist [63]. Multimedia Appendix 2 provides the complete SRQR checklist.

Figure 2. Sociotechnical factors impacting physicians' documentation practices. EHR: electronic health record.



Ethical Considerations

This study was registered and approved by Sikt, Norwegian Agency for Shared Services in Education and Research (project number 540255). Before conducting observations and interviews, all participants were provided with written and verbal information about the study. Written consent was individually obtained from each participant. No personally identifiable data were collected, and in cases where indirectly identifiable information was mentioned during interviews, such details were

XSL•FO
RenderX

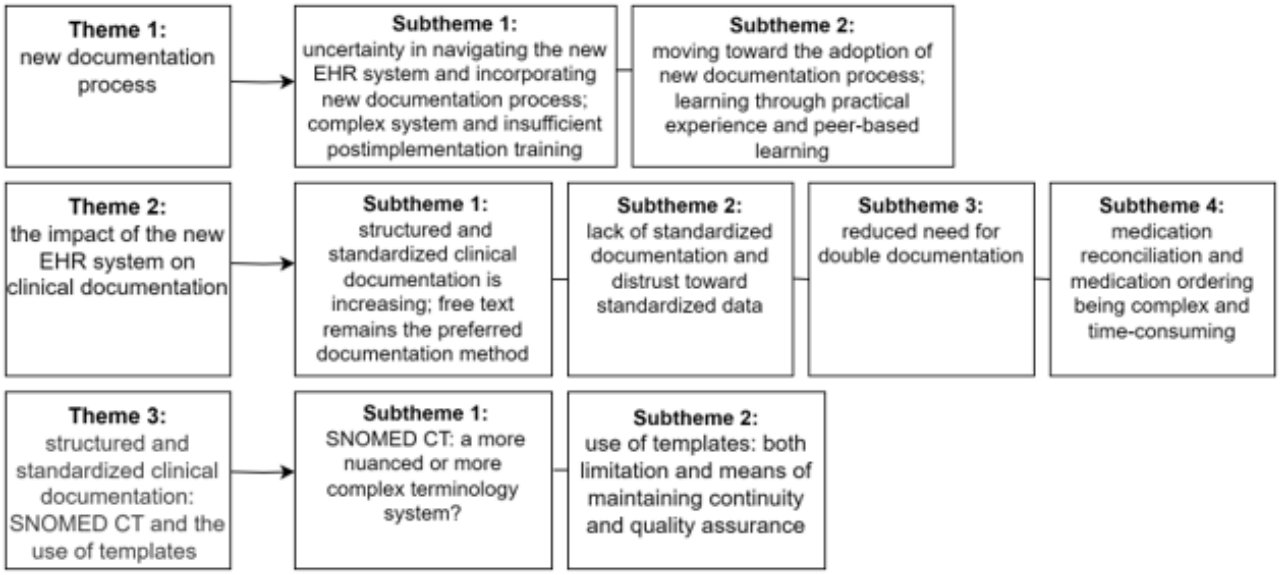
omitted from all reports, papers, and presentations. Participants in this study did not receive any form of compensation.

Results

Overview

Through observations and interviews, 3 main themes, with corresponding subthemes related to documentation practices, were identified (Figure 3). The first theme shed light on the initial challenges faced by physicians during the implementation phase, specifically their uncertainty in navigating the new EHR system and its complexity. It also showed the gradual integration of physicians into the new processes of the EHR system, primarily by learning through practical experience and collaborative knowledge exchange with their peers (peer-based learning). The second theme contemplated the influence of the new EHR system on physicians' documentation practices. It delved into the preferred method of documentation and provided insights into the challenges and benefits associated with the implementation of the new documentation processes. In addition, it highlighted the issue of nonstandardized documentation, which in some cases leads to a sense of distrust among physicians toward standardized clinical data. The third theme delved into the perspectives of physicians regarding the adoption of the new terminology system, SNOMED CT, and their opinions on using templates for clinical data documentation.

Figure 3. Themes and subthemes regarding clinical documentation practices emerged from the interviews and observations. EHR: electronic health record; SNOMED CT: Systematized Nomenclature of Medicine–Clinical Terms.



New Documentation Processes

Uncertainty in Navigating the New EHR System and Incorporating New Documentation Processe; Complexity of the System and Insufficient Postimplementation Training

Physicians encountered a learning curve while transitioning to the new EHR system, necessitating adjustments to specific documentation practices. The complexity of certain processes and the additional administrative responsibilities linked to the new system left physicians feeling unsure about how to efficiently navigate the system and, in certain cases, how to accurately document clinical data:

> *... you spend more time searching for the right function that you need to document correctly, and to search for information, and then maybe you are a little bit unsure about some things, you wonder if you have included everything...there is a lot of clicking to move forward; it wasn't like that before; it used to be more like create note, write note, done, so there are quite a lot of steps to move forward [in the new EHR]; the system controls you a bit, sort of.* [Informant 10]

Initially, many physicians were unfamiliar with the specific functionalities of the new EHR system and attempted to use and document clinical data in a similar manner as they did in the previous system. Consequently, certain informants encountered difficulties when using the new system due to its complexity and lack of familiarity with effective navigation:

> *I think that what we need now is to go through each workflow and make sure that everyone does it in the same way so that we can coordinate more, also that you can improve the shortcuts that exist and the ways of working efficiently so that everyone learns it because there are some who use the system very cumbersomely.* [Informant 9]

Furthermore, informants highlighted that the training they received before the implementation of the new EHR system was not sufficient. This was mainly due to substantial modifications made to the EHR system since the training and because the training did not address their specific role. In addition, some informants mentioned that differentiating between essential and less significant aspects addressed in the training became difficult if they had not yet used the system. Most informants also believed that there was insufficient personalized training provided after the implementation, and

XSL•FO
RenderX

they felt that more specific and customized training was necessary to effectively use the system's advanced features, particularly once physicians had acquired a basic understanding of the system:

> *We received a lot of tips and tricks in the beginning, but then it was more about doing, surviving, staying afloat, and stabilizing ourselves a bit...having new training now would have been a much bigger help at this time than it was before...* [Informant 1]

Moving Toward the Adoption of New Documentation Processes and Learning Through Practical Experience and Peer-Based Learning

On the basis of the observation sessions and interview data, it was evident that physicians were gradually adapting to new processes in the new EHR system and making changes to specific documentation practices:

> *When I used to admit patients, I used to write a continuous text with previous illnesses, family, and social, such as whether they lived at home or in a nursing home. I used to just list all previous illnesses, which were often taken from a previous admission journal. I have partly done it that way now as well, but there is more and more emphasis on not doing it and instead entering the information in problem lists of medical and surgical history. So, it's a transition for me...* [Informant 10]

This adaptation and learning occurred both through practice (actual use of the system) and knowledge sharing among colleagues (peer-based learning). The informants found it valuable to learn from experienced colleagues. Occasionally, they also organized small group meetings where these colleagues provided guidance on navigating the system effectively. However, most physicians expressed concerns about the complexity of the newly implemented EHR system, which hindered their learning process and effective use. It was emphasized that simplifying the system and eliminating unnecessary features would greatly enhance its usability.

The Impact of the New EHR System on Clinical Documentation

Structured and Standardized Clinical Documentation Is Increasing, and Free Text Remains the Preferred Documentation Method

Compared to the previous EHR system, a greater proportion of patient information was being entered in a structured and standardized format. In certain cases, the system enforced a hard stop, making it mandatory for users to input the required standardized information. For instance, in the old EHR system, the findings and diagnosis code were described in a free-text format along with codes such as *ICD-10*, whereas in the new EHR system, these aspects were documented in a standardized format using SNOMED CT. Another example was the transition from paper-based to digital documentation for ordering surgical procedures. As a result, numerous tasks were performed electronically, eliminating the reliance on paper, and more patient health information was documented in a structured and standardized format:

> *Well, then [previous EHR] there was nothing standardized, so...no, no, then I manually entered the procedure code and diagnosis code, and the medications were mostly just copied from one note to the next, and then you hoped that what you had copied over was correct.* [Informant 6]

Although structured and standardized documentation has become more prevalent, many physicians still preferred using free text whenever possible. This preference was primarily due to the flexibility and time-saving nature of free-text documentation. It was also easier for physicians to provide a comprehensive overview of a patient's medical condition through free-text notes. Furthermore, some informants highlighted their preference for reading high-quality free-text notes as they provided a better understanding of the patient's condition compared to information presented in templates or diagnosis codes:

> *The only way I manage to get an overview of the medical history of a patient is basically if someone has written something reasonable in a journal note.* [Informant 4]

Moreover, physicians were accustomed to free-text documentation, and transitioning to structured and standardized formats required additional time and training to ensure accuracy.

Structured and standardized documentation, on the other hand, was often perceived as time-consuming, more challenging, and sometimes misaligned with physicians' workflows. In addition, one physician added that he did not see the benefits of adopting a more intricate approach, such as standardized documentation:

> *Yes, no, I need to see the benefit of doing it in a more complicated way, then I might have chosen a different way of documentation than free text. So far, I haven't been able to see the major benefit, so that's why I document in free text...* [Informant 2]

Nonetheless, some informants acknowledged that reading and finding free-text documentation could be challenging in certain situations, particularly when there were numerous notes present. In such cases, important information might become obscured or go unnoticed, whereas standardized information automatically appeared when reviewing the patient journal.

Lack of Standardized Documentation and Distrust Toward Standardized Data

On the basis of our findings, the preference for free-text documentation method was often accompanied by a culture of distrust among physicians toward standardized data available in the EHR system. For instance, the new EHR system integrated several processes related to clinical documentation that were not present in the previous EHR system. These included processes such as the documentation of the "problem list," which is an overview of medical conditions that might impact the health care given to the patient, and "medical history," which lists the conditions that have been resolved and no longer impact patients' health or treatment. These conditions, if documented,

XSL•FO
RenderX

were automatically displayed when reviewing patient journals without the need to access journal notes. However, physicians were uncertain about the accuracy of these standardized data and sometimes could not rely on it, as they were aware that many of their colleagues did not adhere to standardized documentation:

> *It depends on how many people enter standardized data, so I feel that I can't rely on what is stated there. One can enter previous illnesses and medical history, but it is rarely done. So, then, the patient looks healthy on paper, but there are many illnesses, as you can find in the old journal system.* [Informant 4]

Reduced Need for Double Documentation

In certain situations, physicians favored structured and standardized documentation; for instance, in the new EHR system, they transitioned from paper-based to digital documentation when ordering surgical procedures. By adhering to structured and standardized documentation during the preoperative phase, they could save time in the postoperative phase, as some of the information was automatically filled in the postoperative template, eliminating redundant documentation. As a result, physicians recognized the benefits of using structured and standardized documentation for frequently performed routine procedures:

> *It is to some extent a bit easier [postoperative documentation] because all the information I enter beforehand also appears in the postoperative description afterwards, so before, I used to manually write in codes and bleeding amount, for example, but all of this is now automatically included in the postoperative description, so actually, writing the postoperative description afterwards is easier.* [Informant 8]

Medication Reconciliation and Medication Ordering Being Complex and Time-Consuming

With the implementation of the new EHR system, physicians transitioned from manually recording medications on paper-based charts to electronically documenting them in a standardized format. The informants expressed their frustration and difficulties with the current medication reconciliation process, ordering and administering medicines, and inaccuracies in the medication list. This frustration was particularly evident during patient admission and discharge, as there were numerous medications to reconcile. However, while some informants acknowledged the challenges of entering and managing medication reconciliation in the new EHR system compared to the previous system, they also emphasized the potential for mitigating patient risk. Moreover, some informants expressed contentment with the shift from paper charts to electronic charts for medication lists. Conversely, certain individuals contended that this process was demanding and could potentially increase patient risk:

> *I find the medication lists to be a bit tangled and unclear, especially when you have to discharge the patient and know what they are actually on, I always feel like there is something that gets overlooked, some medications that are missing or others that have been included but shouldn't have been, so, the medication lists are challenging.* [Informant 1]

Furthermore, it was observed that, in some cases, physicians who encountered difficulties with medication ordering or the medication reconciliation process sought assistance from colleagues who were more proficient in using the system.

Structured and Standardized Clinical Documentation: SNOMED CT and the Use of Templates

SNOMED CT: More Nuanced or More Complex Terminology System?

The transition from *ICD-10* to SNOMED CT for documenting patient health conditions was perceived by some informants as challenging and by others as a more nuanced diagnostic system. Challenges usually arose due to difficulties associated with finding the correct diagnosis, as physicians were presented with multiple alternatives when starting to write the name of the diagnosis. In addition, sometimes physicians could not find the diagnosis they needed. When diagnosing with SNOMED CT, physicians could specify the location of a medical condition using the SNOMED CT code:

> *...in ICD, there is no difference between right and left, or bilaterally, whereas in SNOMED it is easier to...there, the diagnosis code can already indicate differences between right and left, and bilaterally, so it is potentially that I write right and left in the diagnosis code, which I did not do before...* [Informant 7]

However, in some cases, it was not possible to do this for certain conditions. For instance, during one of the observations, it was noted that when the physician was documenting the diagnosis of deep subfascial lipoma, the physician could not specify the location of the lipoma in SNOMED CT. As a solution, the physician noted the location in the daily notes. The physician pointed out that this information could be essential, especially for surgeons. In addition, the physician mentioned that they could specify the location for a "general" lipoma but not for a subfascial lipoma.

Nonetheless, some informants perceived SNOMED CT as a more nuanced terminology system. Furthermore, SNOMED CT eliminated the need for physicians to commit to memorizing the precise code linked to a specific diagnosis. Instead, they could search for a broader term related to the diagnosis they intended to document. Subsequently, they were presented with multiple alternative options, allowing them to choose the most fitting one.

It is important to note that, in numerous cases, most physicians relied on *ICD-10* codes to identify the corresponding SNOMED CT diagnosis.

Use of Templates: Both Limitation and Means of Maintaining Continuity and Quality Assurance

Most informants emphasized the benefits of using templates to document routine procedures, such as the most frequently performed surgeries. They believed that this approach saved

XSL•FO
RenderX

time, and, for some informants, it helped them remember all the necessary parameters and documentation requirements:

> *It goes faster because there is already a lot of text that is already there, and then you just need to fill in the right and left, how long are the screws, what type of plate was used, what kind of things...and the rehabilitation afterward, is it six weeks, or eight weeks...* [Informant 7]

In addition, it could serve as a checklist for each step involved in a procedure. Many informants had created their own templates, with some developing up to 20 templates for frequently executed operations. However, some of the informants also mentioned that it took time to create these templates properly, so they relied on templates designed by their colleagues. It was also noted that, despite being in a template, these data were not searchable because they were written in free-text format.

Conversely, the informants highlighted that the use of structured and standardized templates for nonstandardized procedures was viewed as both time-consuming and ineffective. It was mentioned that by using a standardized template, physicians might end up documenting unnecessary information, thereby complicating the clinical workflow, particularly during hectic days.

Discussion

Principal Findings

This study provides insights into physicians' documentation practices after they transitioned from an older EHR system to a cross-institutional EHR system, which enables standardized and structured documentation. This transition necessitated physicians to change their documentation habits to some extent, resulting in more clinical data being recorded in a structured and standardized format. However, the degree of acceptance toward new documentation processes is influenced by a range of social and technological factors.

First, according to the study participants, training in system use and ongoing technical support were of utmost importance, particularly after implementing the EHR:

> *We received a lot of tips and tricks in the beginning, but then it was more about doing, surviving, staying afloat, and stabilizing ourselves a bit...having new training now would have been a much bigger help at this time than it was before...* [Informant 1]

The importance of training has been well-documented in previous research [64-67]. However, most of the attention and resources allocated for training tend to be centered on the pre–go-live phase and the initial implementation stage of the EHR system, with limited support available for an extended period after the go-live phase and for system updates [65,68,69]. In addition, it has been noted that users may feel overwhelmed during initial training and prioritize familiarizing themselves with the fundamental aspects of the system rather than striving to use it efficiently [68]. This concern was also emphasized by the informants in our study. Moreover, peer-based learning, where physicians are taught by their expert peers in the use of the EHR system, was deemed beneficial. This suggests that acquiring knowledge from fellow experts could prove more advantageous, as these peers possess a deep understanding of the physicians' workflows and the specific challenges they face when using the system. Peer-based learning has also been observed when physicians use the templates created by their colleagues. Hence, it can be advantageous to create platforms where physicians can share tacit knowledge, enabling them to mutually benefit from their expertise in effectively using the system and implementing best practices for documentation. Providing additional training in documentation practices should be considered, particularly during the transition to the new EHR system, which necessitates changes in physicians' documentation practices. This training becomes even more significant considering the alarming statistics from a study conducted in Norway. The study, which encompassed 14 hospitals, revealed that in 2003, an incorrect main diagnosis code was assigned in 37.8% of hospital stays, and in 2008, in 36.2% of hospital stays [47]. According to physicians working in these hospitals, the main factors contributing to this issue were challenging code systems (59%) and inadequate training (50%) [70]. The significance of education and ongoing training in good documentation practices as well as its contribution to achieving a learning health system, has been emphasized in numerous research studies [71-75].

Second, in conjunction with the aforementioned first factor, physicians found the transition to the new EHR system to be challenging. The challenge stems from the complex navigation needed to find and execute specific tasks, the overwhelming amount of information in the EHR system, and the inefficient user interface. Furthermore, physicians' past recording habits in older EHR systems can impact their documentation practices in the new EHR system. As pointed out by Orlikowski [38], when people use technology, "people also draw on their skills, power, knowledge, assumptions, and expectations about the technology and its use, influenced typically by training, communication, and previous experiences" [76]. Given that physicians have established workflows in the previous system, adjusting to new processes in the new EHR system might hinder their adaptability and work efficiency. This finding aligns with previous research conducted by Joukes et al [77], who identified that the preimplementation circumstances played a significant role in shaping the perceived advantages of implementing a new EHR system with structured and standardized clinical data recording. Specifically, the authors emphasized that the perceived impact of the implementation varied depending on whether users were transitioning from older EHR systems or paper-based records. In a paper-based record center, the implementation yielded positive perceived effects on EHR use, data quality, and data reuse. However, in a center that had previously used an EHR system, the perceived effects were predominantly negative or neutral. The authors further highlighted that individuals with previous exposure to EHR systems tended to draw comparisons between the new system and their old one, thereby potentially recognizing that certain tasks were accomplished more efficiently in the former system [77].

XSL•FO
RenderX

While the adoption of the new EHR system has led to an increase in structured and standardized data recording, it is often not a decision made by physicians but rather an imposition by the new system and health organization. In certain cases, the structured and standardized documentation does not align with physicians' traditional workflow; therefore, it takes more time compared to free-text documentation. Moreover, physicians occasionally experience uncertainty and difficulties regarding the correct way to document structured and standardized data or question their necessity. This may potentially lead to inconsistencies in data recording. Data that could be standardized may instead be documented in a free-text format, and when faced with time constraints, physicians may use work-arounds, such as quickly clicking through to proceed further. This practice might also explain physicians' distrust toward some of the standardized clinical data available in the EHR system, such as problem lists and medical history. One advantage of the problem list and medical history is that these remain visible to physicians whenever they access patient records, without the need to open journal notes. While this provides better information at the point of care, the problem lists and medical history lose their advantage if physicians fail to document or update it or if they lack trust in the accuracy and reliability of this information. Moreover, the prevalent reliance of physicians on the *ICD-10* codes to identify the corresponding SNOMED CT codes may undermine the advantages of using a more comprehensive terminology system, SNOMED CT. This reliance can lead to loss of clinical detail and mapping inaccuracies. This practice indicates a discrepancy between the physicians' mental models (understanding and reasoning about systems) and the designer's conceptual model (how the system is intended to be used) [78,79]. Mental models, derived from a range of factors, including experiences, knowledge, and perception, have been widely applied in the field of human-computer interaction research [79,80]. Gaining an understanding of physicians' mental models can be important for effectively optimizing the system and tailoring training in documentation practices in the new EHR system. The greater the disparity between the designer's conceptual model and the user's mental model, the more challenging it becomes for individuals to effectively use the system [78,79]. Furthermore, this emphasizes the need to distinguish between technologies as artifacts and technologies in practice (how technology is actually used rather than solely focusing on the technology itself), as it is only through effective use that technology can impact productivity [38]. Consequently, when enhancing technology, it becomes vital to consider sociotechnical factors, which, in turn, necessitates the technology to possess sufficient flexibility in order to be customized according to specific local needs and evolving requirements [33,37]. Moreover, Rosenbloom et al [22] have underscored the significance of offering flexibility in the documentation processes, as they are highly valued and essential for health care professionals. By using free-text documentation, health care professionals can document intricate and nuanced details about a patient, which can also prove to be more valuable for subsequent health care professionals taking over the patient's care [22]. Some of our study participants have also emphasized that reading high-quality free-text journal notes can provide them with more crucial information compared to solely reviewing standardized or structured data:

> *The only way I manage to get an overview of the medical history of a patient is basically if someone has written something reasonable in a journal note.* [Informant 4]

Having the right balance between standardization and flexibility in documentation processes is crucial, particularly in situations where unforeseen clinical findings or unexpected circumstances arise [22].

According to the study participants, most physicians have a positive attitude toward using templates for documenting routine procedures:

> *It goes faster [using templates] because there is already a lot of text that is already there, and then you just need to fill in the right and left, how long are the screws, what type of plate was used, what kind of things...and the rehabilitation afterward, is it six weeks, or eight weeks...* [Informant 7]

While certain physicians have created their own templates for frequently conducted surgeries, others depend on templates designed by their peers. However, the information contained within these templates is recorded in a free-text format. As a result, these data are not conducive to analysis using statistical or machine learning techniques. Moreover, when physicians develop their own templates for commonly performed surgeries, it can result in inconsistencies in documentation and potential omissions of critical data that need to be documented. On the other hand, this behavior might be elucidated through the analysis of physician-patient dialogues in Norwegian general practice by Nessa [81,82]. His research emphasized that these conversations were a fundamental aspect of a physician's work, encompassing rich narratives and certain elements such as context and emotional tone [81,82]. Such complexities may be challenging to capture within a standardized template and encoded data. Thus, adhering strictly to encoded data and standardized templates implied by the organization may impose a semiotic burden on physicians, requiring them to exert considerable mental effort to fit complex narratives into standardized templates. The attitudes of health care professionals toward structured and standardized documentation may also depend on their perception of the benefits it provides [21,83]. Our study participants also highlighted this aspect (eg, saving time in postoperative documentation), suggesting that physicians may view the implementation of standardized or semistandardized templates for commonly performed procedures, developed in collaboration with physicians and following established guidelines, as a positive step. Moreover, this approach could potentially minimize variations in clinical documentation. This approach might also enable more data to be consistently recorded in a standardized format and enhance guideline compliance. Several studies have shown that the use of templates improves the accuracy of documentation [84-86]. For instance, Thomson et al [85] demonstrated the advantages of using a standardized operative note template adhering to established guidelines. This implementation led to enhanced documentation quality and a decrease in the missing data.

XSL•FO
RenderX

Limitations and Future Research

One limitation of our study is that it involved a limited number of physicians, and data were collected from only 1 hospital. Therefore, the findings of this study may only reflect physicians' perspectives working within this specific context. Nonetheless, we have made an effort to gather information on documentation practices that can be applied more broadly. Specifically, our focus was on understanding documentation practices when physicians transition from an older EHR system to a new EHR system that facilitates structured and standardized documentation. In addition, we aimed to identify the factors influencing physicians' choice of documentation methods, such as free text and standardized documentation, regardless of the EHR system they use. Thus, our findings may have relevance beyond the specific organization and the EHR system used in this study, providing insights that can be applied to a wider context.

Moreover, although it is evident that the implementation of the new EHR system has increased the amount of clinical data recorded in a standardized format, our study did not objectively evaluate either the quality of these data (eg, accuracy and completeness) or their effectiveness for secondary uses. Consequently, future research should aim to objectively assess the quality of the data. Furthermore, additional investigations could delve into the perceived data quality from the perspectives of individuals using the data from the new EHR system for secondary purposes, such as quality assurance. This would help determine whether the transition to a more standardized EHR system has positively impacted the use of these data for secondary purposes.

Conclusions

The findings of this study identified that physicians' acceptance of new documentation processes is influenced by a range of social and technological factors. These factors include physicians' past experiences with documentation, perceived benefits, familiarity with the EHR system, and time constraints, all of which can impact their choice of documentation methods. Furthermore, it is important to consider the user-friendliness of the EHR system and how well the documentation processes align with physicians' workflows. To facilitate the transition to a new EHR system, health care organizations should provide tailored training programs on documentation practices. This is particularly crucial after the implementation of the new system. In addition, it would be advantageous to establish platforms that allow physicians to exchange tacit knowledge and expertise on effectively using the system. Moreover, physicians highly value flexibility in clinical documentation. Therefore, health care organizations should consider implementing standardized documentation processes that also align with physicians' preferences and allow them some flexibility in the documentation process. For instance, introducing common standardized or semistandardized templates for routine procedures, developed in collaboration with physicians, could potentially minimize disparities in clinical documentation, ensure the consistent recording of data, and improve adherence to guidelines.

Acknowledgments

The authors would like to extend their gratitude to the physicians for their time commitment in sharing their experiences and providing insightful perspectives on the use of the new electronic health record system and documentation practices. This study was funded by the Liaison Committee between the Central Norway Regional Health Authority and the Norwegian University of Science and Technology.

Data Availability

The datasets generated or analyzed during this study are not publicly available to protect participant privacy but are available from the corresponding author on reasonable request.

Authors' Contributions

OG, AF, LM, and RP conceptualized and designed the study. OG conducted data collection and performed the initial data analyses. AF, LM, and RP contributed to the analyses and interpretation of the data. OG wrote the initial draft of the manuscript. All authors performed critical revisions of the paper and contributed to the final version of the manuscript. All authors have reviewed and approved the submission of the paper in its current form.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Semistructured interview guide.
[DOCX File , 19 KB - formative_v9i1e63902_app1.docx]

Multimedia Appendix 2
SRQR (Standards for Reporting Qualitative Research) checklist.

XSL•FO
RenderX

[DOCX File , 27 KB - formative_v9i1e63902_app2.docx]

References

1. Gatiti P, Ndirangu E, Mwangi J, Mwanzu A, Ramadhani T. Enhancing healthcare quality in hospitals through electronic health records: a systematic review. J Health Inform Dev Ctries 2021 Nov 02;15(2):1-25.
2. Kruse CS, Mileski M, Vijaykumar AG, Viswanathan SV, Suskandla U, Chidambaram Y. Impact of electronic health records on long-term care facilities: systematic review. JMIR Med Inform 2017 Sep 29;5(3):e35 [FREE Full text] [doi: 10.2196/medinform.7958] [Medline: 28963091]
3. Janssen A, Kay J, Talic S, Pusic M, Birnbaum RJ, Cavalcanti R, et al. Electronic health records that support health professional reflective practice: a missed opportunity in digital health. J Healthc Inform Res 2023 Jan 20;6(4):375-384 [FREE Full text] [doi: 10.1007/s41666-022-00123-0] [Medline: 36744083]
4. Capurro D, Yetisgen M, van Eaton E, Black R, Tarczy-Hornoch P. Availability of structured and unstructured clinical data for comparative effectiveness research and quality improvement: a multisite assessment. EGEMS (Wash DC) 2014 Jul 11;2(1):1079 [FREE Full text] [doi: 10.13063/2327-9214.1079] [Medline: 25848594]
5. Olsen L, Aisner D, McGinnis J, Institute of Medicine (US) Roundtable on Evidence-Based Medicine. The Learning Healthcare System: Workshop Summary. Washington, DC: National Academies Press; 2007.
6. Joukes E, Abu-Hanna A, Cornet R, de Keizer NF. Time spent on dedicated patient care and documentation tasks before and after the introduction of a structured and standardized electronic health record. Appl Clin Inform 2018 Jan;9(1):46-53 [FREE Full text] [doi: 10.1055/s-0037-1615747] [Medline: 29342479]
7. Ebbers T, Takes RP, Smeele LE, Kool RB, van den Broek GB, Dirven R. The implementation of a multidisciplinary, electronic health record embedded care pathway to improve structured data recording and decrease electronic health record burden. Int J Med Inform 2024 Apr;184:105344 [FREE Full text] [doi: 10.1016/j.ijmedinf.2024.105344] [Medline: 38310755]
8. Bush RA, Kuelbs C, Ryu J, Jiang W, Chiang G. Structured data entry in the electronic medical record: perspectives of pediatric specialty physicians and surgeons. J Med Syst 2017 May;41(5):75 [FREE Full text] [doi: 10.1007/s10916-017-0716-5] [Medline: 28324321]
9. Gilbert JA. Physician data entry: providing options is essential. Health Data Manag 1998 Sep;6(9):84-6, 88, 90-2. [Medline: 10185232]
10. Lov om behandling av helseopplysninger ved ytelse av helsehjelp (pasientjournalloven). Lovdata. 2015. URL: https://lovdata.no/dokument/NL/lov/2014-06-20-42 [accessed 2024-06-17]
11. Lov om helsepersonell m.v. (helsepersonelloven). Lovdata. URL: https://lovdata.no/dokument/NL/lov/1999-07-02-64/KAPITTEL_8#KAPITTEL_8 [accessed 2024-06-17]
12. Lov om behandling av personopplysninger (personopplysningsloven). Lovdata. 2018. URL: https://lovdata.no/dokument/NL/lov/2018-06-15-38 [accessed 2024-06-17]
13. Post AR, Burningham Z, Halwani AS. Electronic health record data in cancer learning health systems: challenges and opportunities. JCO Clin Cancer Inform 2022 Mar;6:e2100158 [FREE Full text] [doi: 10.1200/CCI.21.00158] [Medline: 35353547]
14. Kong HJ. Managing unstructured big data in healthcare system. Healthc Inform Res 2019 Jan;25(1):1-2 [FREE Full text] [doi: 10.4258/hir.2019.25.1.1] [Medline: 30788175]
15. Joukes E, Cornet R, de Bruijne MC, de Keizer NF, Abu-Hanna A. Development and validation of a model for the adoption of structured and standardised data recording among healthcare professionals. BMC Med Inform Decis Mak 2018 Jun 28;18(1):54 [FREE Full text] [doi: 10.1186/s12911-018-0640-8] [Medline: 29954388]
16. Sun W, Cai Z, Liu F, Fang S, Wang G. A survey of data mining technology on electronic medical records. In: Proceedings of the IEEE 19th International Conference on e-Health Networking, Applications and Services. 2017 Presented at: HealthCom 2017; October 12-15, 2017; Dalian, China. [doi: 10.1109/HealthCom.2017.8210774]
17. Tayefi M, Ngo P, Chomutare T, Dalianis H, Salvi E, Budrionis A, et al. Challenges and opportunities beyond structured data in analysis of electronic health records. WIREs Comput Stat 2021 Feb 14;13(6):e1549. [doi: 10.1002/wics.1549]
18. Seid M, Hartley DM, Margolis PA. A science of collaborative learning health systems. Learn Health Syst 2021 Jun 01;5(3):e10278 [FREE Full text] [doi: 10.1002/lrh2.10278] [Medline: 34277944]
19. Zafar MA, Beck AF, Chirumbolo C, Wilson KD, Haralson K, Rich K, et al. A regional learning health system of congregate care facilities for COVID-19 response. Learn Health Syst 2024 Jan 24;8(3):e10407 [FREE Full text] [doi: 10.1002/lrh2.10407] [Medline: 39036529]
20. Ebbers T, Kool RB, Smeele LE, Dirven R, den Besten CA, Karssemakers LH, et al. The impact of structured and standardized documentation on documentation quality; a multicenter, retrospective study. J Med Syst 2022 May 27;46(7):46 [FREE Full text] [doi: 10.1007/s10916-022-01837-9] [Medline: 35618978]
21. Vuokko R, Mäkelä-Bengs P, Hyppönen H, Lindqvist M, Doupi P. Impacts of structuring the electronic health record: results of a systematic literature review from the perspective of secondary use of patient data. Int J Med Inform 2017 Jan;97:293-303. [doi: 10.1016/j.ijmedinf.2016.10.004] [Medline: 27919387]

XSL•FO
RenderX

22. Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. J Am Med Inform Assoc 2011;18(2):181-186 [FREE Full text] [doi: 10.1136/jamia.2010.007237] [Medline: 21233086]
23. Levy DR, Withall JB, Mishuris RG, Tiase V, Diamond C, Douthit B, et al. Defining documentation burden (DocBurden) and excessive DocBurden for all health professionals: a scoping review. Appl Clin Inform 2024 Oct;15(5):898-913 [FREE Full text] [doi: 10.1055/a-2385-1654] [Medline: 39137903]
24. Sinsky CA, Bavafa H, Roberts RG, Beasley JW. Standardization vs customization: finding the right balance. Ann Fam Med 2021;19(2):171-177 [FREE Full text] [doi: 10.1370/afm.2654] [Medline: 33685879]
25. Kriznik NM, Lamé G, Dixon-Woods M. Challenges in making standardisation work in healthcare: lessons from a qualitative interview study of a line-labelling policy in a UK region. BMJ Open 2019 Nov 27;9(11):e031771 [FREE Full text] [doi: 10.1136/bmjopen-2019-031771] [Medline: 31780591]
26. Wears RL. Standardisation and its discontents. Cogn Technol Work 2015 Feb;17(1):89-94 [FREE Full text] [doi: 10.1007/s10111-014-0299-6] [Medline: 25667566]
27. Brunner J, Cannedy S, McCoy M, Hamilton AB, Shelton J. Software is policy: electronic health record governance and the implications of clinical standardization. J Gen Intern Med 2023 Oct;38(Suppl 4):949-955 [FREE Full text] [doi: 10.1007/s11606-023-08280-7] [Medline: 37798574]
28. Bansler JP. Challenges in user-driven optimization of EHR: a case study of a large epic implementation in Denmark. Int J Med Inform 2021 Apr;148:104394. [doi: 10.1016/j.ijmedinf.2021.104394] [Medline: 33485217]
29. Koh J, Ahmed M. Improving clinical documentation: introduction of electronic health records in paediatrics. BMJ Open Qual 2021 Feb;10(1):e000918 [FREE Full text] [doi: 10.1136/bmjoq-2020-000918] [Medline: 33589503]
30. Electronic health records and the importance of how to document. Pan American Health Organization. 2021. URL: https://iris.paho.org/handle/10665.2/55410 [accessed 2024-05-10]
31. Ulriksen GH, Pedersen R, Ellingsen G. Infrastructuring in healthcare through the OpenEHR architecture. Comput Supp Coop Work 2017 Feb 13;26:33-69. [doi: 10.1007/s10606-017-9269-x]
32. Meum T, Ellingsen G, Monteiro E, Wangensteen G, Igesund H. The interplay between global standards and local practice in nursing. Int J Med Inform 2013 Dec;82(12):e364-e374. [doi: 10.1016/j.ijmedinf.2013.02.005] [Medline: 23688846]
33. Eason K. Local sociotechnical system development in the NHS national programme for information technology. J Inf Technol 2007 Sep;22(3):257-264. [doi: 10.1057/palgrave.jit.2000101]
34. Pedersen R, Ellingsen G, Monteiro E. Standardized nursing work: works in practice but not in theory? Stud Health Technol Inform 2010;157:91-96. [Medline: 20543373]
35. Pedersen R. Standardizing work in healthcare through architecture, routines and technologies. In: Proceedings of the 10th International Conference on the Design of Cooperative Systems. 2012 Presented at: COOP 2012; May 30-June 1, 2012; Marseille, France. [doi: 10.1007/978-1-4471-4093-1_2]
36. Schaper LK, Pervan GP. ICT and OTs: a model of information and communication technology acceptance and utilisation by occupational therapists. Int J Med Inform 2007 Jun;76 Suppl 1:S212-S221. [doi: 10.1016/j.ijmedinf.2006.05.028] [Medline: 16828335]
37. Cherns A. The principles of sociotechnical design. Hum Relat 1976 Aug;29(8):783-792 [FREE Full text] [doi: 10.1177/001872677602900806]
38. Orlikowski WJ. Using technology and constituting structures: a practice lens for studying technology in organizations. Organ Sci 2000 Aug;11(4):404-428. [doi: 10.1287/orsc.11.4.404.14600]
39. Pedersen R. Standardization strategies in healthcare practices? In: Proceedings of the Fifteenth International Symposium for Health Information Management Research. 2011 Presented at: ISHIMR 2011; September 8-9, 2011; Zurich, Switzerland.
40. Suchman LA. Plans and Situated Actions: The Problem of Human-Machine Communication. Cambridge, UK: Cambridge University Press; 1987.
41. Bostrom RP, Heinen JS. MIS problems and failures: a socio-technical perspective. Part I: the causes. MIS Q 1977 Sep;1(3):17-32. [doi: 10.2307/248710]
42. Orlikowski WJ. The duality of technology: rethinking the concept of technology in organizations. Organ Sci 1992 Aug 01;3(3):398-427. [doi: 10.1287/orsc.3.3.398]
43. Sittig DF, Singh H. A sociotechnical approach to electronic health record related safety. In: Key Advances in Clinical Informatics: Transforming Health Care Through Health Information Technology. Cambridge, MA: Academic Press; 2017.
44. Abbas R, Michael K. In: Papagiannidis S, editor. Socio-Technical Theory: A Review. Newcastle upon Tyne, UK: TheoryHub Book; 2023.
45. Sittig DF, Singh H. A new socio-technical model for studying health information technology in complex adaptive healthcare systems. In: Kaufman DR, Kannampallil TG, Patel VL, editors. Cognitive Informatics for Biomedicine: Human Computer Interaction in Healthcare. Cham, Switzerland: Springer International Publishing; 2015.
46. Laukvik LB, Lyngstad M, Rotegård AK, Fossum M. Utilizing nursing standards in electronic health records: a descriptive qualitative study. Int J Med Inform 2024 Apr;184:105350 [FREE Full text] [doi: 10.1016/j.ijmedinf.2024.105350] [Medline: 38306850]

XSL•FO
RenderX

47. Utredning av «en innbygger – én journal». Helsedirektoratet. 2014 Sep. URL: https://www.regjeringen.no/contentassets/355890dd2872413b838066702dcdad88/ikt_utfordringsbilde_helse_omsorgssektoren.pdf [accessed 2024-04-12]
48. Meld. St. 9 (2012–2013). Regjeringen.no. 2012. URL: https://www.regjeringen.no/no/dokumenter/meld-st-9-20122013/id708609/ [accessed 2024-04-17]
49. Bakgrunn for Helseplattformen. Helseplattformen. URL: https://www.helseplattformen.no/bakgrunn-for-helseplattformen [accessed 2024-04-17]
50. Kartlegging av elektroniske pasientjournalsystemer og deres støtte for nasjonale e-helseløsninger. Helsedirektoratet. 2023. URL: https://tinyurl.com/y7mupvvv [accessed 2025-03-15]
51. Årlig melding 2017 for Helse Midt-Norge RHF til Helse- og omsorgsdepartementet. Regjeringen.no. 2017. URL: https://www.regjeringen.no/contentassets/18f93a5faa81417b8ac28f74a02cc0ee/arlig-melding-2017-fra-helse-midt-norge-rhf.pdf [accessed 2024-03-10]
52. Stedenfeldt M, Wekre LJ, Tørhaug T, Gotaas E, Toch-Marquardt M. Fra fritekst til struktur og felles helsespråk. Michael 2023;20(Supplement 31):106-114 [FREE Full text]
53. Abeysinghe R, Zheng F, Bernstam EV, Shi J, Bodenreider O, Cui L. A deep learning approach to identify missing is-a relations in SNOMED CT. J Am Med Inform Assoc 2023 Feb 16;30(3):475-484 [FREE Full text] [doi: 10.1093/jamia/ocac248] [Medline: 36539234]
54. Fung KW, Xu J. An exploration of the properties of the CORE problem list subset and how it facilitates the implementation of SNOMED CT. J Am Med Inform Assoc 2015 May;22(3):649-658 [FREE Full text] [doi: 10.1093/jamia/ocu022] [Medline: 25725003]
55. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc 2004;11(2):104-112 [FREE Full text] [doi: 10.1197/jamia.M1471] [Medline: 14633936]
56. Mamykina L, Vawdrey DK, Stetson PD, Zheng K, Hripcsak G. Clinical documentation: composition or synthesis? J Am Med Inform Assoc 2012 Nov 01;19(6):1025-1031 [FREE Full text] [doi: 10.1136/amiajnl-2012-000901] [Medline: 22813762]
57. Pizziferri L, Kittler AF, Volk LA, Honour MM, Gupta S, Wang S, et al. Primary care physician time utilization before and after implementation of an electronic health record: a time-motion study. J Biomed Inform 2005 Jun;38(3):176-188 [FREE Full text] [doi: 10.1016/j.jbi.2004.11.009] [Medline: 15896691]
58. Rowlands S, Tariq A, Coverdale S, Walker S, Wood M. A qualitative investigation into clinical documentation: why do clinicians document the way they do? Health Inf Manag 2022 Sep;51(3):126-134. [doi: 10.1177/1833358320929776] [Medline: 32643428]
59. Ho YX, Gadd CS, Kohorst KL, Rosenbloom ST. A qualitative analysis evaluating the purposes and practices of clinical documentation. Appl Clin Inform 2014;5(1):153-168 [FREE Full text] [doi: 10.4338/ACI-2013-10-RA-0081] [Medline: 24734130]
60. Pankhurst T, Evison F, Atia J, Gallier S, Coleman J, Ball S, et al. Introduction of systematized nomenclature of medicine-clinical terms coding into an electronic health record and evaluation of its impact: qualitative and quantitative study. JMIR Med Inform 2021 Nov 23;9(11):e29532 [FREE Full text] [doi: 10.2196/29532] [Medline: 34817387]
61. AlYahmady HH, Al Abri SS. Using Nvivo for data analysis in qualitative research. Int Interdiscip J Educ 2013 Feb;2(2):181-186. [doi: 10.12816/0002914]
62. Braun V, Clarke V. One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qual Res Psychol 2020 Aug 12;18(3):328-352. [doi: 10.1080/14780887.2020.1769238]
63. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014;89(9):1245-1251. [doi: 10.1097/acm.0000000000000388]
64. Ratwani R, Fairbanks T, Savage E, Adams K, Wittie M, Boone E, et al. Mind the gap. A systematic review to identify usability and safety challenges and practices during electronic health record implementation. Appl Clin Inform 2016 Nov 16;7(4):1069-1087 [FREE Full text] [doi: 10.4338/ACI-2016-06-R-0105] [Medline: 27847961]
65. Blavin F, Ramos C, Shah A, Devers K. Lessons from the literature on electronic health record implementation: a study funded by the office of the national coordinator for health information technology of the U.S. Department of Health and Human Services. The Urban Institute. 2013 Aug 1. URL: https://www.healthit.gov/sites/default/files/hit_lessons_learned_lit_review_final_08-01-2013.pdf [accessed 2024-05-13]
66. Samadbeik M, Fatehi F, Braunstein M, Barry B, Saremian M, Kalhor F, et al. Education and training on electronic medical records (EMRs) for health care professionals and students: a scoping review. Int J Med Inform 2020 Oct;142:104238. [doi: 10.1016/j.ijmedinf.2020.104238] [Medline: 32828034]
67. Heponiemi T, Gluschkoff K, Vehko T, Kaihlanen AM, Saranto K, Nissinen S, et al. Electronic health record implementations and insufficient training endanger nurses' well-being: cross-sectional survey study. J Med Internet Res 2021 Dec 23;23(12):e27096 [FREE Full text] [doi: 10.2196/27096] [Medline: 34941546]
68. Bredfeldt CE, Awad EB, Joseph K, Snyder MH. Training providers: beyond the basics of electronic health records. BMC Health Serv Res 2013 Dec 02;13:503 [FREE Full text] [doi: 10.1186/1472-6963-13-503] [Medline: 24295150]

XSL•FO
RenderX

69. Sachs J, Yonek JC, Mullins CD, Hamilton A, Rachman F, Kaleba E, et al. Impact of health IT on lab order use in community health centers. Agency for Healthcare Research and Quality. 2011 Nov. URL: https://digital.ahrq.gov/file/18903/download?token=9XtjSeyh [accessed 2024-05-13]
70. Riksrevisjonens undersøkelse av kodekvaliteten ved helseforetakene. Riksrevisjonen. 2006. URL: https://www.stortinget.no/globalassets/pdf/dokumentserien/2005-2006/dok_3_7_2005_2006.pdf [accessed 2024-06-11]
71. Forrest CB, Crandall WV, Bailey LC, Zhang P, Joffe MM, Colletti RB, et al. Effectiveness of anti-TNFα for Crohn disease: research in a pediatric learning health system. Pediatrics 2014 Jul;134(1):37-44 [FREE Full text] [doi: 10.1542/peds.2013-4103] [Medline: 24935993]
72. Rattray NA, Damush TM, Miech EJ, Homoya B, Myers LJ, Penney LS, et al. Empowering implementation teams with a learning health system approach: leveraging data to improve quality of care for transient ischemic attack. J Gen Intern Med 2020 Nov 01;35(Suppl 2):823-831 [FREE Full text] [doi: 10.1007/s11606-020-06160-y] [Medline: 32875510]
73. Mungmode A, Noor N, Weinstock RS, Izquierdo R, Indyk JA, DeSalvo DJ, et al. Making diabetes electronic medical record data actionable: promoting benchmarking and population health improvement using the T1D exchange quality improvement portal. Clin Diabetes 2022;41(1):45-55 [FREE Full text] [doi: 10.2337/cd22-0072] [Medline: 36714251]
74. Alonso GT, Corathers S, Shah A, Clements M, Kamboj M, Sonabend R, et al. Establishment of the T1D exchange quality improvement collaborative (T1DX-QI). Clin Diabetes 2020 Apr;38(2):141-151 [FREE Full text] [doi: 10.2337/cd19-0032] [Medline: 32327886]
75. Vandenberg AE, Kegler M, Hastings SN, Hwang U, Wu D, Stevens MB, et al. Sequential implementation of the EQUIPPED geriatric medication safety program as a learning health system. Int J Qual Health Care 2020 Sep 23;32(7):470-476. [doi: 10.1093/intqhc/mzaa077] [Medline: 32671390]
76. Orlikowski WJ, Gash DC. Technological frames: making sense of information technology in organizations. ACM Trans Inf Syst 1994 Apr;12(2):174-207. [doi: 10.1145/196734.196745]
77. Joukes E, de Keizer NF, de Bruijne MC, Abu-Hanna A, Cornet R. Impact of electronic versus paper-based recording before EHR implementation on health care professionals' perceptions of EHR use, data quality, and data reuse. Appl Clin Inform 2019 Mar;10(2):199-209 [FREE Full text] [doi: 10.1055/s-0039-1681054] [Medline: 30895574]
78. Staggers N, Norcio AF. Mental models: concepts for human-computer interaction research. Int J Man Mach Stud 1993 Apr;38(4):587-605. [doi: 10.1006/imms.1993.1028]
79. Kaufman DR, Kannampallil TG, Patel VL. Cognition and human computer interaction in health and biomedicine. In: Patel VL, Kannampallil TG, Kaufman DR, editors. Cognitive Informatics for Biomedicine. Cham, Switzerland: Springer; 2015.
80. Holtrop JS, Scherer LD, Matlock DD, Glasgow RE, Green LA. The importance of mental models in implementation science. Front Public Health 2021 Jul 6;9:680316 [FREE Full text] [doi: 10.3389/fpubh.2021.680316] [Medline: 34295871]
81. Nessa J. From a medical consultation to a written text. 1. Transcribing the doctor-patient dialogue. Scand J Prim Health Care 1995 Jun 12;13(2):83-88 [FREE Full text] [doi: 10.3109/02813439508996741] [Medline: 7569483]
82. Nessa J. From a medical consultation to a written text. 2. Pragmatics and textlinguistics applied to medicine. Scand J Prim Health Care 1995 Jun 12;13(2):89-92 [FREE Full text] [doi: 10.3109/02813439508996742] [Medline: 7569484]
83. Ludwick DA, Doucette J. Adopting electronic medical records in primary care: lessons learned from health information systems implementation experience in seven countries. Int J Med Inform 2009 Jan;78(1):22-31. [doi: 10.1016/j.ijmedinf.2008.06.005] [Medline: 18644745]
84. Mahapatra P, Ieong E. Improving documentation and communication using operative note proformas. BMJ Qual Improv Rep 2016 Feb 08;5(1):u209122.w3712 [FREE Full text] [doi: 10.1136/bmjquality.u209122.w3712] [Medline: 26893892]
85. Thomson DR, Baldwin MJ, Bellini MI, Silva MA. Improving the quality of operative notes for laparoscopic cholecystectomy: assessing the impact of a standardized operation note proforma. Int J Surg 2016 Mar;27:17-20 [FREE Full text] [doi: 10.1016/j.ijsu.2016.01.037] [Medline: 26804350]
86. Kahn D, Stewart E, Duncan M, Lee E, Simon W, Lee C, et al. A prescription for note bloat: an effective progress note template. J Hosp Med 2018 Jun 01;13(6):378-382. [doi: 10.12788/jhm.2898] [Medline: 29350222]

Abbreviations

EHR: electronic health record
ICD-10: International Classification of Diseases, 10th Revision
SNOMED CT: Systematized Nomenclature of Medicine–Clinical Terms
SRQR: Standards for Reporting Qualitative Research

XSL•FO
RenderX

Edited by A Mavragani; submitted 04.07.24; peer-reviewed by D Chartash; comments to author 09.01.25; revised version received 17.01.25; accepted 19.01.25; published 21.03.25.

Please cite as:
Golburean O, Pedersen R, Melby L, Faxvaag A
Exploring Physicians' Dual Perspectives on the Transition From Free Text to Structured and Standardized Documentation Practices: Interview and Participant Observational Study
JMIR Form Res 2025;9:e63902
URL: https://formative.jmir.org/2025/1/e63902
doi:10.2196/63902
PMID:

©Olga Golburean, Rune Pedersen, Line Melby, Arild Faxvaag. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Co-Design of a Depression Self-Management Tool for Adolescent and Young Adult Cancer Survivors: User-Centered Design Approach

Karly M Murphy[1], PhD; Rachel Glock[1], BA; David Victorson[2], PhD; Madhu Reddy[3], PhD; Sarah A Birken[4], PhD; John M Salsman[5], PhD

[1]Department of Psychology, East Carolina University, Greenville, NC, United States

[2]Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[3]Department of Informatics, University of California Irvine, Irvine, CA, United States

[4]Department of Implementation Science, Wake Forest University School of Medicine, Winston-Salem, NC, United States

[5]Department of Social Sciences and Health Policy, Wake Forest University School of Medicine, Winston-Salem, NC, United States

Corresponding Author:
Karly M Murphy, PhD
Department of Psychology
East Carolina University
1000 E 5th St
Greenville, NC, 27858
United States
Phone: 1 2527374638
Fax: 1 2523286283
Email: murphykar22@ecu.edu

Abstract

Background: Adolescent and young adult (AYA) cancer survivors are more likely to experience elevated depressive symptoms than older survivors and healthy age-matched peers. Despite the elevated risk of depressive symptoms in AYA cancer survivors and the existence of evidence-based interventions to address depression, it is unclear whether AYA cancer survivors can access support services. Digital tools are a potential solution to overcoming barriers to AYA cancer survivors' unmet needs for psychosocial support, but they have not been tailored to the needs and preferences of this unique population.

Objective: This study engaged AYA cancer survivors and their providers in the concept generation and ideation step of the user-centered design process through online co-design workshops. The goal was to generate concepts and ideas for a digital depression self-management tool tailored to AYA cancer survivors.

Methods: We conducted 5 co-design workshops—4 with AYA cancer survivors and 1 with providers who serve them. Participants were asked to provide feedback on an existing digital mindfulness course using an "I like, I wish, I wonder" framework. Then, participants were asked "How might we..." questions focused on brainstorming ideas for how the digital tool might work. Participants brainstormed responses independently and then worked as a group to categorize and expand on their ideas. Co-design workshops were autotranscribed using Webex (Cisco) software. Transcripts underwent thematic analysis with additional context provided by the products created during the workshop.

Results: Eight AYA cancer survivors (aged 15-37 years) and 4 providers (2 oncologists and 2 social workers) participated in co-design workshops. We identified 6 themes: barriers to engagement, desired content, preferences for content delivery, preferences for interface, features, and aspects to avoid. Each theme had 2-7 subthemes that we relied upon when making design decisions for the prototype.

Conclusions: Co-design workshops provided critical insights that informed the prototype development of a digital depression self-management tool tailored to AYA cancer survivors. Key takeaways that were integrated into prototype design include (1) using stories from other AYA cancer survivors to demonstrate concepts; (2) delivering content in brief lessons; and (3) using encouraging notifications, organizational tools, and reward systems to keep AYA cancer survivors engaged with the tool. Some of the themes identified in this study (eg, desired content and features) are consistent with known strategies for promoting user engagement and co-design work in other cancer survivors. However, this study extended previous research by identifying uniquely

XSL•FO
RenderX

relevant strategies for tailoring to AYA cancer survivors, such as delivering content in brief sessions to overcome the time constraints AYA cancer survivors experience, providing opportunities for private expression, and maintaining an encouraging tone throughout the tool. These data were used to inform the prototype development of a digital depression self-management tool tailored to AYA cancer survivors.

(JMIR Form Res 2025;9:e67175) doi:10.2196/67175

KEYWORDS

adolescent and young adult; cancer survivorship; depressive symptoms; self-management; co-design workshops; user-centered design; thematic analysis; intervention tailoring; digital mental health; evidence-based intervention; digital tools; psychosocial support

Introduction

Depressive Symptoms Are Prevalent and Problematic in Adolescent and Young Adult Cancer Survivors

Annually, approximately 90,000 adolescents and young adults (AYAs) between the ages of 15 and 39 years are diagnosed with cancer in the United States, with over 2.1 million AYA cancer survivors living in the United States as of 2020. These numbers are expected to rise as the incidence of cancer in this age group increases [1,2]. Cancer survivors are up to 5 times more likely to experience depression than members of the general population [3]. Furthermore, a number of studies have demonstrated that younger cancer survivors are more likely to experience elevated depressive symptoms when compared with both older survivors and healthy age-matched peers [4-6]. Among AYA cancer survivors, estimates of the prevalence of depression range from 25% to 32% [7]. Elevated rates of depressive symptoms in AYA cancer survivors are likely due to disruptions or delays in the achievement of developmental milestones (eg, completion of education and becoming a parent) [8]. Depressive symptoms in AYA cancer survivors are problematic because depressive symptoms are associated with diminished health-related quality of life, increased health care use and costs, and elevated mortality [9].

Psychosocial Interventions Are Efficacious in Treating Depressive Symptoms

The American Psychological Association's clinical practice guidelines recommend cognitive behavioral therapy (CBT) for the treatment of depressive disorders in AYAs [10]. CBT focuses on helping patients recognize and change negative automatic thoughts (by cognitive restructuring) and maladaptive behaviors (by behavioral activation) that play a role in the development and maintenance of depressive symptoms [11]. Among adults, mindfulness-based cognitive therapy (MBCT) has also been shown to be efficacious. MBCT integrates elements of CBT with mindfulness-based stress reduction to improve one's awareness of and relationship to unwanted thoughts, feelings, and bodily sensations [10,12]. More recently, there has been substantial interest in the use of positive psychology interventions to manage psychological symptoms in patients with severe and chronic illnesses. Systematic reviews and meta-analyses of trials of these interventions suggest they have small to moderate effects on depression and may be more effective than other depression interventions [13,14]. Thus, there are several topline evidence-based interventions (EBIs) for depression. Despite the elevated risk of depression in AYA cancer survivors and the existence of EBIs to address depression, it is unclear whether AYA cancer survivors can access such services. Indeed, the AYA Health Outcomes and Patient Experience study found that between 56% and 75% of AYA cancer survivors reported an unmet need for psychosocial support 12 months postdiagnosis [15].

Digital Tools Can Increase Psychosocial Intervention Reach

Digital tools are a potential solution to overcoming barriers to AYA cancer survivors' unmet needs for psychosocial support and a preferred mode of delivery for this population [16]. They allow interventions to be delivered remotely using smartphone technology that the vast majority of AYA cancer survivors already own, making them highly scalable [17]. Several studies have demonstrated the feasibility of digital tools for the self-management of physical symptoms among AYA cancer survivors [18]. Furthermore, research suggests psychosocial interventions for depression can be adapted for digital delivery with evidence for efficacy among healthy AYAs [19-25]. However, few existing digital interventions targeting depression have been evaluated in studies focused on AYA cancer survivors.

Evidence for Digital Interventions Targeting Depression in AYA Cancer Survivors Has Been Mixed

In a 2015 meta-analysis, Richter et al [26] concluded there was limited evidence for the efficacy of existing interventions targeting psychosocial distress in AYA cancer survivors, including depressive symptoms. Since that time, Campo et al [27] published the outcomes of a mindful self-compassion videoconference intervention with 25 AYA cancer survivors and demonstrated improvements in depression following the intervention (Cohen d=0.99). Zhang et al [28] recently demonstrated the feasibility and preliminary efficacy of technology-assisted CBT for reducing depression in a pilot randomized controlled trial of 17 AYA cancer survivors (Hedges g=1.12). Although these findings are promising, others have been less so. For example, Sansom-Daily et al [29] found that an online group-based CBT intervention for AYA cancer survivors early in cancer survivorship was associated with a higher level of depressive symptoms compared with a control group (P=.04). As such, it remains unclear whether CBT, mindfulness, or other interventions will be most efficacious for improving depressive symptoms in this population. Furthermore, the interventions tested by both Campo et al [27] and

XSL•FO
RenderX

Sansom-Daily et al [29] were delivered by interventionists via videoconference; while this strategy addresses some barriers to AYA cancer survivors' participation in psychosocial interventions (eg, distance or travel-related expenses), they are not available on demand as AYA cancer survivors prefer [16].

Content Tailoring and User-Centered Design Increase Efficacy and Engagement

While there is a growing body of available general mental health apps, these tools do not address the specific needs and concerns of AYA cancer survivors, such as negative thoughts about cancer and feelings of isolation from peers [30-32]. A significant consequence of this is low user engagement [33]. In the general population, only 1 out of 5 apps downloaded are used more than once [34]. The risk of nonuse of mobile health apps can be reduced with tailoring and effective design [35]. Tailoring involves developing individualized interventions based on key individual difference variables such as age and cancer history [36]. Data suggest that tailoring intervention content to the population of interest facilitates greater content relevance, which may result in more efficacious interventions and higher rates of long-term use [37]. User-centered design (UCD) processes allow for the integration of end-user input and feedback regarding features designed to increase user engagement [38]. For example, McCurdie et al [39] proposed a UCD process that consists of three iterative steps: (1) concept generation and ideation, (2) prototype design and system development, and (3) evaluation [39]. Following the final evaluation phase, the product is deployed. Users are at the center of this process, providing input at the concept generation and ideation step as well as the evaluation step. In this study, we aimed to engage AYA cancer survivors in the concept generation and ideation step of this process and design an initial prototype.

Research Context and Guiding Strategy

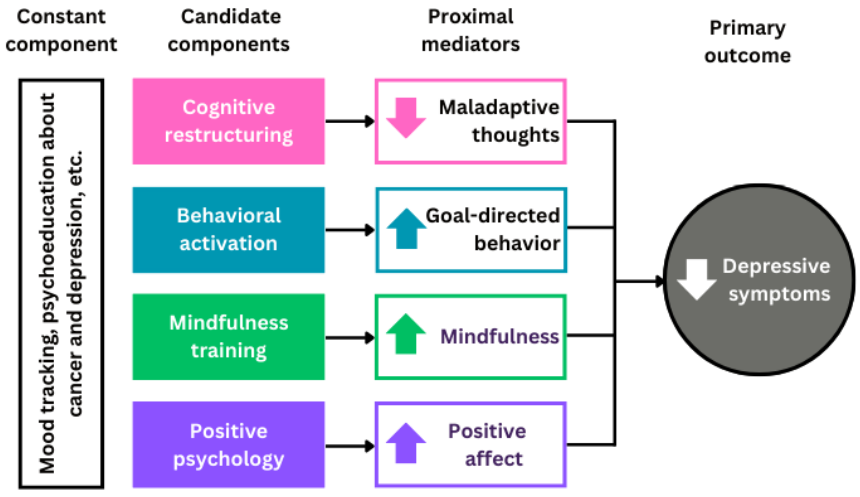
The work presented in this paper forms part of a wider study informed by the multiphase optimization strategy (MOST) that aims to prepare and optimize a digital depression self-management tool for AYA cancer survivors. MOST is a comprehensive framework that can be used to optimize an intervention by systematically identifying the most promising components [40,41]. MOST includes three phases: (1) preparation, which lays the groundwork for optimization (eg, identification of a guiding conceptual model, choosing candidate components, formative work, and pilot-testing); (2) optimization, in which a trial (ie, a factorial experiment) is used to identify the components that meet the optimization criterion; and (3) evaluation, in which the effectiveness of the optimized intervention is confirmed. The work of this study was informed by the tasks to be completed during the preparation phase of MOST [42].

Methods

Overview

In alignment with guidance from MOST, we devised a theoretically and empirically supported model that informed our selection of candidate intervention components (Figure 1). We then engaged in an iterative UCD process to adapt these candidate intervention components for delivery through an existing self-guided digital tool and to address the unique needs and concerns of AYA cancer survivors to ensure user engagement. We began this process by conducting co-design workshops with AYA cancer survivors and oncology providers that serve this population to generate a concept for the adapted tool. At the same time, the end users of the digital tool will be AYA cancer survivors. Input from providers supplemented AYA cancer survivors' suggestions with expertise regarding how this tool might overcome existing barriers to the management of depressive symptoms among AYA cancer survivors. Of note, the positive psychology component had already been modified for use with AYA cancer survivors and thus was not included in the co-design process [43]. Based on workshop findings, KM designed the prototype of the tool.

Figure 1. The guiding conceptual model for the preparation of a digital depression self-management tool for adolescent and young adult cancer survivors.



Participants

Inclusion criteria for AYA cancer survivors were as follows: (1) diagnosed with cancer between the ages of 12 and 39 years; (2) currently between the ages of 15 and 39 years; (3) at least 1 month after completion of primary treatment; (4) fluent in spoken and written English; (5) have access to a smartphone with a data plan; and (6) have access to a computer with internet,

XSL•FO
RenderX

a webcam, and a microphone that could be used to participate in a Webex (Cisco) videoconference. Notably, the National Cancer Institute (NCI) typically defines AYA cancer survivors as individuals diagnosed between the ages of 15 and 39 years and currently aged 15-39 years [1]. We chose to expand the age of diagnosis for adolescents as they are frequently diagnosed with hematologic malignancies, which require lengthy treatments such as stem cell transplants. Thus, by the time these patients reach posttreatment survivorship, they are considered emerging adults. Decreasing the age of diagnosis to 12 years for this subset of our sample ensured that adolescents currently aged 15-17 years were represented.

Inclusion criteria for oncology providers that serve AYAs were as follows: (1) licensed provider who treats AYA cancer survivors; (2) fluent in spoken and written English; (3) have access to a smartphone with a data plan; and (4) have access to a computer with internet, a webcam, and a microphone that could be used to participate in a Webex videoconference.

Potentially eligible participants were excluded if they reported a current diagnosis of a severe or persistent mental illness or reported severe suicidal ideation, including plan and intent. A safety protocol was in place to ensure appropriate care was provided if a potentially eligible participant reported suicidal ideation.

Recruitment and Enrollment

AYA cancer survivors were recruited from an NCI-designated comprehensive cancer center in the Southeastern region of the United States. Potentially eligible AYA cancer survivors were identified through an electronic medical record review. First, an automated Epic (Epic Systems) report was used to identify patients in the correct age range who had been diagnosed with cancer. Study staff then conducted additional chart review to evaluate eligibility with regard to age at diagnosis, type of diagnosis, time since completion of primary treatment, and history of severe or persistent mental illness. Potentially eligible providers were identified based on their affiliation with an AYA oncology program within North Carolina.

Study staff contacted potentially eligible participants or (in the case of adolescents) their parents or guardians by email and (if screening was not completed and no response was received from the participant after 1 week) phone call. Interested individuals or (in the case of adolescents) their parents or guardians were directed to a website with more information about the study and asked to complete a REDCap (Research Electronic Data Capture; Vanderbilt University) [44] screening form to determine their eligibility. This screening form requested that prospective participants (or, in the case of adolescents, their parents on their behalf) provide their current age, age at cancer diagnosis, the type of cancer they had been diagnosed with, how long it had been since they completed cancer treatment, whether they had ever been diagnosed with a severe or persistent mental illness, whether they had access to a smartphone and computer, and their ability to speak and read English. Eligible and interested participants and (in the case of adolescents) their parents or guardians were automatically directed to the appropriate assent and consent forms (see the *Ethical Considerations* subsection for a detailed explanation of the assent and consent process).

Upon provision of assent and consent, AYA cancer survivors and providers were asked to provide their demographic information and availability for participation in study procedures. Co-design workshop participants were scheduled in groups of at least 3 participants.

Co-Design Workshops

Overview

KMM conducted 4 online co-design workshops with AYA cancer survivors and 1 co-design workshop with providers that serve this population using Webex [45]. AYA cancer survivors and providers completed workshop activities from their homes or offices. No one else was present during workshops other than the participants and the researcher. KMM is a female individual with a PhD in Clinical Psychology and, at the time of the interview, was an instructor at Wake Forest University School of Medicine. KMM has extensive training in qualitative data collection methods and completed a certificate in UCD. She did not have a relationship with AYA cancer survivor participants before the study commencement but had reviewed their demographic and clinical characteristics. Regarding providers, KMM had previously interacted with some of the participants in professional settings but did not have any formal working relationship with them. As the principal investigator of this study, KMM's primary interest in the research topic was informing the development of a digital tool to help AYA cancer survivors better manage symptoms of depression, and this was communicated to all participants. Providers and AYA cancer survivors attended separate workshops. Each workshop lasted approximately 1 hour. Workshops were audio and video recorded and autotranscribed.

Workshop Activities

To begin each workshop, participants were provided with information about the goal of the project, namely, to develop a tool that helps AYA cancer survivors manage symptoms of depression. In addition, participants were introduced to EBIs for depression (eg, behavioral activation and mindfulness training) that we aimed to adapt and tailor for digital delivery to AYA cancer survivors. After providing some general guidelines for brainstorming and an introduction to Figma (Figma, Inc) [46] (a digital whiteboard tool), participants were asked to provide feedback on an existing digital mindfulness course (Wakeful) using an "I like, I wish, I wonder" prompts within Figma (Figure 2). In this context, wakefulness was presented to participants as an example to spur ideas. Then, participants were asked "How might we..." questions focused on brainstorming ideas for how the digital tool might work (Figure 3). Participants first brainstormed responses to each question independently and then worked as a group to categorize and expand on their ideas. Slides presented during workshops are included as Multimedia Appendix 1 (AYA cancer survivor workshops) and Multimedia Appendix 2 (provider workshop).

XSL•FO
RenderX

Figure 2. Sample responses to the "I like, I wish, I wonder" activity from the adolescent and young adult cancer survivor workshop.

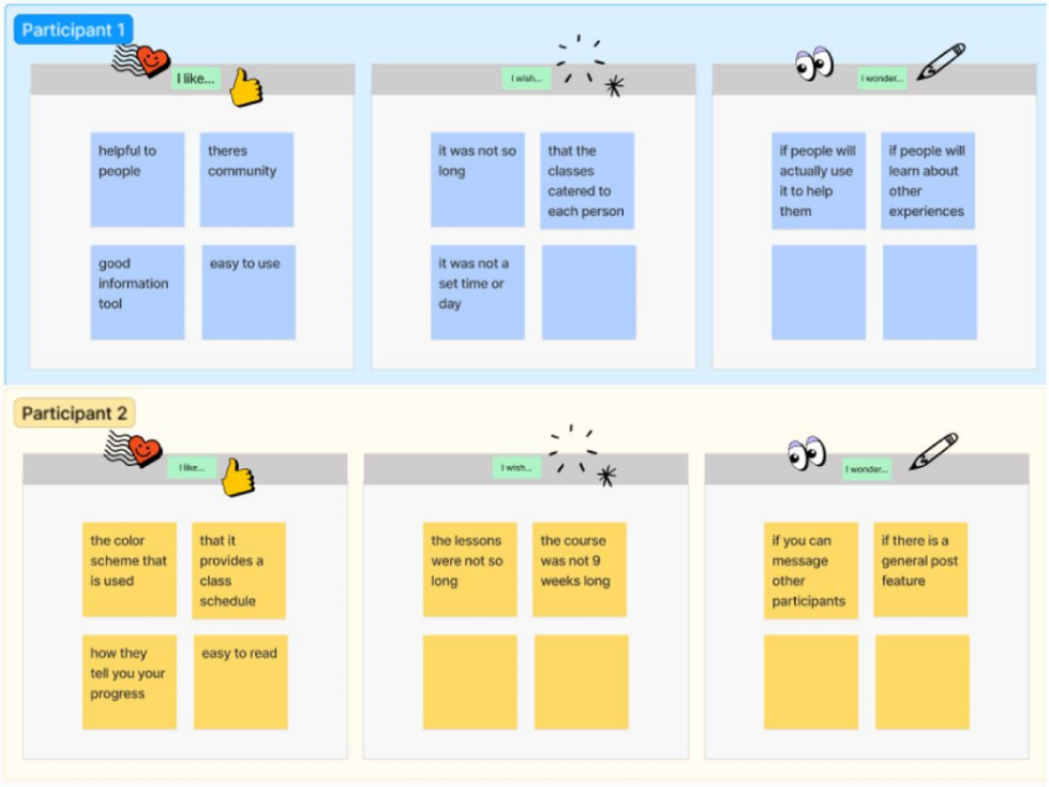
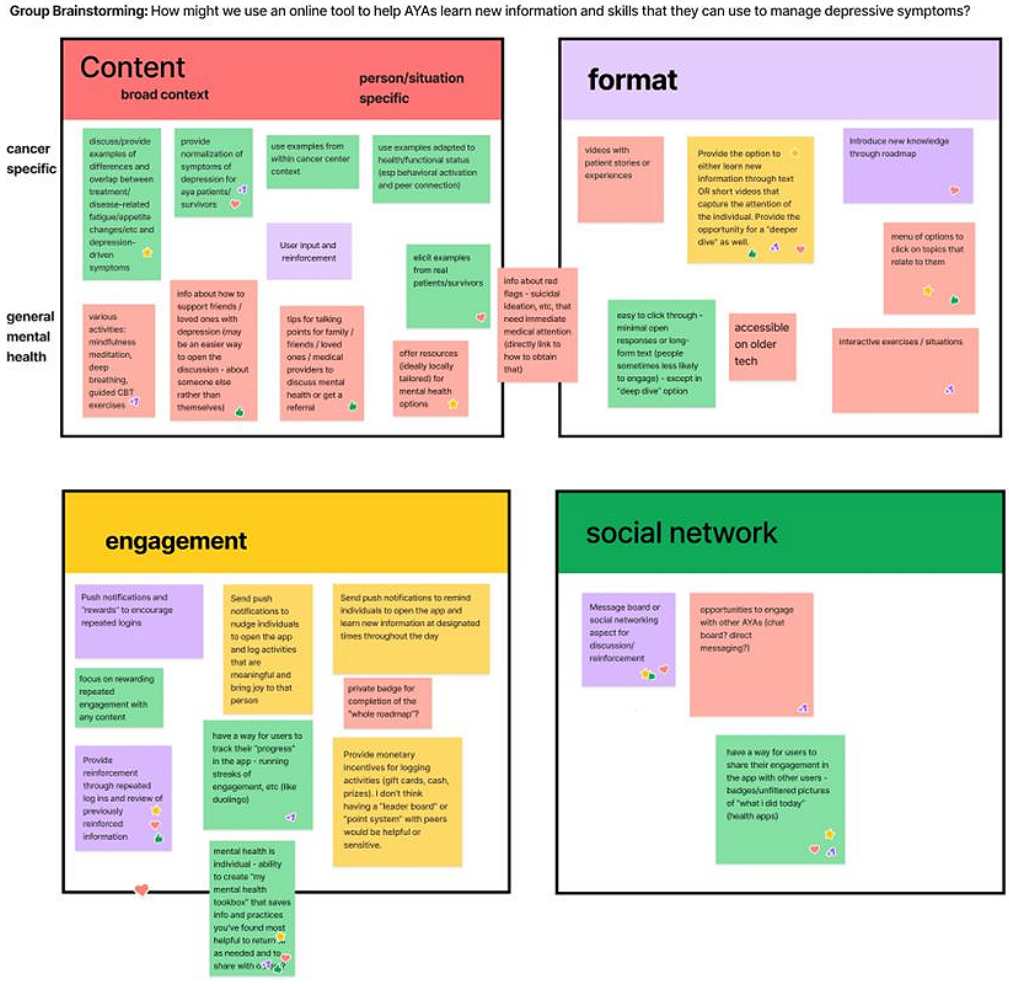


Figure 3. Sample responses to the group brainstorming activity from the provider workshop. AYA: adolescent and young adult.



XSL•FO
RenderX

Data Processing and Analysis

Field notes were made by the researcher following each workshop. Co-design workshops were autotranscribed using Webex software. Transcripts of co-design workshops were cleaned by study team members with additional context provided by workshop products (ie, sticky notes). Reference was made to the original audio and video recordings as needed for clarification. Transcripts were not returned to participants for correction, given the availability of audio and video recordings.

RG, with oversight from KMM, conducted a thematic analysis of workshop transcripts guided by the methodological framework by Braun and Clarke [47]; the analysis was characterized by an inductive approach, where themes were directly derived from the transcripts rather than predefined before the analysis. The analysis began with a familiarization phase of understanding the content, which provided initial insights into overarching themes. Following familiarization, the coding phase included identifying and assigning a descriptive code that succinctly encapsulated the content. These codes systematically organized the data and identified patterns that later informed the identification of themes. The thematic analysis provided a comprehensive understanding of the data, highlighting the participant interactions within the co-design workshops. Participants did not provide feedback on the findings; rather, findings were used to develop a prototype that was then presented to additional participants to gather feedback.

Ethical Considerations

All study procedures were reviewed and approved by the Wake Forest University Health Sciences Institutional Review Board (IRB00075887). This included an informed consent and assent process, which primarily occurred through REDCap. Prospective participants and (in the case of adolescents) their parents were provided with a summary of the study, information about why the study was being conducted, the procedures involved in participating in the study, the duration of the study, the designation of the study as no more than minimal risk, a description of the minimal risks inherent in participating in the study (eg, potential for breach of confidentiality or psychological discomfort), the benefits to participating in the study (eg, helping others in the future), alternatives to participating in the study, compensation for participating in the study, and contact information for the principal investigator and Institutional Review Board.

In the case of young adults, eligible participants were asked to review this information about the study and required to give their consent to participate before providing their demographic and clinical information and participating in the workshop. In the case of adolescents younger than the age of 18 years, study staff initially contacted the parent of the patient. Parents were asked to review information about the study and provide their permission for the adolescent to participate. Upon provision of parental permission, study staff contacted the adolescent directly. Adolescents were provided with age-appropriate information about the study through REDCap and were required to provide their assent to participate before providing their demographic and clinical information and participating in the workshop.

All remote data collection occurred through Health Insurance Portability and Accountability Act (HIPAA)–compliant platforms, and all identifiable data were stored on secure, password-protected servers (eg, REDCap and institutional servers) only accessible by authorized study staff. Recordings of workshops were transcribed and deidentified for analysis. All information identifying participants will be deleted upon publication of the primary study outcomes.

Participants were compensated US $40 for participating in a workshop. Compensation was provided in the form of a digital gift card that was emailed to the participant following the workshop.

Results

Participants

In total, 4 providers and 8 AYA cancer survivors participated in 5 co-design workshops, consistent with the recommended number of focus group discussions needed to achieve saturation [48]. There were 2 additional AYA cancer survivors who expressed interest in the study but did not ultimately participate in a workshop due to scheduling conflicts. The provider workshop had 4 participants, while each of the AYA cancer survivors' workshops had 2 participants. All workshops lasted approximately 1 hour.

The majority of providers identified as non-Hispanic White (4/4, 100%) and heterosexual (3/4, 75%). Disciplines represented by providers included medical oncology and social work. AYA cancer survivors ranged in age from 15 to 37 years (mean 23.5, SD 8.85 years). The majority of AYA cancer survivors identified as non-Hispanic White (6/8, 75%), female (6/8, 75%), and heterosexual (7/8, 87.5%). Each AYA cancer survivor had been diagnosed with cancer between the ages of 12 and 36 years. All AYA cancer survivors had completed treatment between 1 month and 5 years before participating in the workshop.

Thematic Analysis

Thematic analysis revealed 6 overarching themes: barriers to engagement, preferences for content delivery, desired content, preferences for interface, features, and aspects to avoid. Each of these themes and their subthemes are described in detail below, with exemplar quotes. Of note, some quotes were spoken by participants, and this was pulled from the transcript; other quotes were never spoken by participants but instead typed out on the Figma board as part of engaging with the digital workshop materials. As such, we have denoted whether quotes are from the "Transcript" or "Workshop Materials," respectively.

Theme 1: Barriers to Engagement

This theme represents the challenges participants might face that hinder their ability to engage effectively with the tool. These barriers are crucial in understanding the obstacles that must be addressed to enhance participation and involvement. The 2 primary barriers were time constraints and lack of usability.

XSL•FO
RenderX

Subtheme 1.1: Time Constraints

Overall, the participants emphasized challenges with time constraints due to balancing work, school, and family responsibilities. They highlighted that these time constraints often lead to difficulties in consistently engaging in activities to support their needs, as highlighted by the quote below.

> *Yeah, probably, um, lesser time periods, maybe spread them out a little bit more. Um, For the busy momma's out there, 'cause I tell ya when you work 80+ hours a week and go to school full time and have a family.* [24-Year-old leukemia survivor, transcript]

Subtheme 1.2: Lack of Usability

A significant barrier noted by both groups of participants, particularly older AYA cancer survivor users and providers, is a lack of usability. This includes challenges related to outdated technology and user interfaces that may not be intuitive for this demographic. Addressing these usability issues is essential to enhance the overall user experience and to ensure the tool is accessible.

> *Okay, I'm like the older AYA cancer survivors, and I'm struggling with my brain. It's trying to, like, learn the new technology before it can think...so, I almost feel like, I can't think.* [37-Year-old breast cancer survivor, transcript]

> *There would be attention paid, and that the design would be such that it would be accessible on, like, multiple, even maybe older operating systems. Because especially if these people are dealing with the financial toxicity of cancer treatment. Their tech may be a couple of versions back. So, I guess paying attention to accessibility and ease of use.* [Provider 3, medical oncologist, transcript]

Theme 2: Desired Content

This theme represents the types of content that participants find most valuable and engaging. Understanding these preferences helps ensure that the content is relevant, engaging, and aligned with AYA cancer survivors' interests and needs. We identified 4 specific types of content: information about mental health, resources, AYA cancer survivors' stories, and cancer-specific content.

Subtheme 2.1: User Guidance

Participants expressed a desire for the tool to include a comprehensive guide for the tool. They indicated that having a structured guide would help them navigate the tool more effectively, understand its features, and maximize its functionalities.

> *I think, yeah, there could be like a guide like, when you first open the app or like, uh, not really a tutorial, but like sort of have it go through the app by its features and stuff. I think that would be pretty helpful.* [15-Year-old gastrointestinal stromal tumor (GIST) survivor, transcript]

Subtheme 2.2: Information About Mental Health

Participants and providers expressed a strong desire for mental health information to be integrated throughout the tool. They emphasized the importance of incorporating relevant mental health content, such as information about depression, to address the specific needs and concerns of AYA cancer survivors.

> *How depression can start, what depression can lead to, and what I need to do if I am depressed.* [19-Year-old brain tumor survivor, workshop materials]

Subtheme 2.3: Resources

Participants and providers expressed a strong desire for the tool to include a variety of resources that address different aspects of the AYA cancer survivors' cancer experiences. This includes a desire to have information about how to cope with mental health difficulties; rediscovering enjoyable activities after cancer; and external resources, particularly for emergent mental health concerns. Incorporating these elements will ensure that the tool offers comprehensive support and valuable information tailored to the needs and interests of AYA cancer survivors.

> *Information about immediate medical attention for suicidal ideation.* [Provider 1, medical oncologist, workshop materials]

> *Have helpful tools outside of this for people to manage their depression on a daily basis.* [20-Year-old thyroid cancer survivor, workshop materials]

> *This is actually a tool to help me figure out who am I now after going through this...Like, okay, I tried soccer again and on a scale of 1 to 10, I enjoyed, you know, like. More or less some sort of, I think, it would feel more. Um, I don't know what the word is like, less like, you feel a little bit, it would feel a little more empowering, I think, and a little bit less like, there's something else cancer has taken from you.* [37-Year-old breast cancer survivor, transcript]

> *Mental health is individual—ability to create "my mental health toolbox" that saves info and practice you've found most helpful to return to as needed and to share with others.* [Provider 3, medical oncologist, workshop materials]

Subtheme 2.4: AYA Cancer Survivors' Stories

Both participants and providers highlighted the importance of including stories from AYA cancer survivors within the tool. They emphasized that featuring personal experiences from peers can provide valuable insights, foster a sense of connection, and offer relatable examples that resonate with the AYA cancer survivors' community.

> *I think you could give, like, ask people if they would volunteer to give like their stories you could have certain like videos or articles from them that showed how they got through it sort of like a motivation.* [15-Year-old GIST survivor, transcript]

XSL•FO
RenderX

> *Include videos with patient stories or experiences.* [Provider 1, medical oncologist, workshop materials]

Subtheme 2.5: Cancer-Specific Content

Providers noted that incorporating cancer-specific content could be highly beneficial. This specialized content can enhance the tool's relevance and effectiveness by offering targeted guidance and support for those affected by cancer.

> *Discuss/provide examples of differences and overlap between treatment/disease-related symptoms and depression-driven symptoms.* [Provider 2, social worker, workshop materials]

Theme 3: Preferences for Content Delivery

This theme represents how participants prefer to receive content, which is essential for tailoring to meet the needs of users who are AYA cancer survivors. Understanding these preferences helps ensure that content is delivered in a way that aligns with participants' schedules and learning styles. We identified 3 important aspects of content delivery: intervention duration, session duration, and learning modalities.

Subtheme 3.1: Intervention Duration

Participants preferred shorter intervention durations, reflecting the multiple responsibilities that AYA cancer survivors juggle. They indicated a reduced timeframe would better accommodate their busy schedules and enhance their ability to engage consistently.

> *I just wish it wasn't as long as a nine-week course because, like, with me, I'm trying to go to school and, like, work at the same time. So, like it'd be nice if like, this was something that I can do, like, after school, or like between school and work. So that way it wouldn't take so long.* [20-Year-old thyroid cancer survivor, transcript]

Subtheme 3.2: Session Duration

Participants expressed a desire for shorter session durations to better accommodate AYA cancer survivor users' busy schedules. They highlighted that reduced session lengths would make it easier for them to fit the sessions into their diverse and demanding routines.

> *Uh, probably just to make it shorter, I don't want to like be on here for a long time.* [19-Year-old brain tumor survivor, transcript]

> *I guess if I could change one thing about it, it would be lessening the class time.* [24-Year-old leukemia survivor, transcript]

Subtheme 3.3: Learning Modalities

Both participants and providers expressed a desire for varied learning and interaction modalities. They emphasized the need for diverse content delivery methods to cater to different learning styles and preferences.

> *I liked how like there were scripts for the videos so, in case someone doesn't want to, like, watch a video. They can just read through it.* [15-Year-old GIST survivor, transcript]

> *Provide the option to either learn new information through text or short videos.* [38-Year-old breast cancer survivor, workshop materials]

Theme 4: Preferences for Interface

This theme represents participants' preferences regarding the design and functionality of the tool's interface. Understanding these preferences helps ensure that the interface is intuitive, user-friendly, and visually engaging. We identified 2 preferences for interface design: ease of use and color scheme.

Subtheme 4.1: Ease of Use

Both participants and providers emphasized the importance of a user-friendly interface, highlighting the need for intuitive navigation and accessibility to ensure AYA cancer survivor users can interact with the tool efficiently and comfortably. Participants prefer that the tool resumes from where they left off, allowing for a seamless and uninterrupted user experience.

> *I like that it [a sample interface] is very easy to read. It looks like it looks very user friendly as well.* [24-Year-old leukemia survivor, transcript]

> *It is important that the tool is easy to click through—minimal open responses.* [Provider 3, medical oncologist, workshop materials]

Subtheme 4.2: Color Scheme

Participants had varied opinions on the tool prototype's color scheme. Some expressed enjoyment of the current design, appreciating its visual appeal. However, others felt that the color scheme was more suited to a younger audience, suggesting a need for a more age-appropriate or versatile design to better align with the preferences of all AYA cancer survivor users.

> *Personally, I like it [a sample interface] a lot. Like, I like, how all the colors kind of like I don't know, I don't know how to put it. Um, blend well, I guess.* [15-Year-old thyroid cancer survivor, transcript]

> *I feel like I like the colors too, but I think it [a sample interface] sort of it's more of a younger audience all like the bright colors and stuff. I think if it's designed for all ages, I think there should be, like, uh, three colors that you use, and not just like all over the color spectrum.* [15-Year-old GIST survivor, transcript]

Theme 5: Features

This theme represents the essential functionalities and characteristics that participants consider important for an effective tool. Identifying and integrating these features ensures that the tool meets the diverse needs of AYA cancer survivor users, making it more engaging, user-friendly, and supportive. We identified 7 functionalities participants indicated would be important to include in the tool: daily mood tracking, encouraging notifications, organizational tools, reward system, interactive exercises, connection, and private expression.

Subtheme 5.1: Daily Mood Tracking

Participants preferred daily mood tracking instead of hourly updates. They indicated that daily tracking is more manageable, less intrusive, and aligns better with their routines. Participants

XSL•FO
RenderX

thought that such "check-ins" would be helpful as they provide valuable opportunities for self-reflection.

> *15-Year-old GIST survivor (transcript): I think, yeah, I like how it would just be like, for the whole day, instead of each hour.*
>
> *Like um, mostly I'm thinking of a message that would ask how were doing.*
>
> *KMM: Yeah, yeah, so kind of like a check-in.*
>
> *21-Year-old large B cell lymphoma survivor (transcript): Yes, exactly.*
>
> *15-Year-old GIST survivor (transcript): Yeah, yeah, it [mood tracking by the hour] makes some might feel overwhelmed.*

Subtheme 5.2: Encouraging Notifications

Participants and providers noted the effectiveness of using notifications within the tool to promote user engagement. They highlighted that encouraging notifications may facilitate consistent interaction with the tool and aid in maintaining a positive outlook.

> *Yeah, I think just to get a reminder, like some days you might forget to add like your thoughts or whatever, I think just having some sort of notifications, maybe just like, once a day.* [15-Year-old GIST survivor, transcript]
>
> *Send push notifications to nudge individuals to open the app and log activities that are meaningful and bring joy to that person.* [Provider 1, medical oncologist, workshop materials]

Subtheme 5.3: Organizational Tools

Both participants and providers valued the integration of organizational elements into the tool, such as a to-do list, distinct weekly sessions, and goal or progress tracking. These features provide users with a structured framework, helping them manage tasks, stay organized, and track their progress. In addition, participants enjoyed the session schedule and progress monitoring features of an exemplar interface.

> *Include a To-do list kind of option.* [15-Year-old thyroid cancer survivor, workshop materials]
>
> *Introduce new knowledge through roadmap.* [Provider 3, medical oncologist, workshop materials]

Subtheme 5.4: Reward System

Both participants and providers recommended incorporating reward systems to enhance user engagement. These reward systems can provide positive reinforcement, recognize achievements, and encourage consistent participation.

> *I also like when you have sort of point or reward system. So, you can, I guess it gives you more motivation to what your thoughts and ideas on there.* [15-Year-old GIST survivor, transcript]
>
> *Provide monetary incentives for logging activities (gift cards, cash, prizes).* [Provider 1, medical oncologist, workshop materials]

Subtheme 5.5: Interactive Exercises

Providers recommended incorporating interactive elements into the exercises and situations within the tool.

> *Include interactive exercises/situations.* [Provider 1, medical oncologist, workshop materials]

Subtheme 5.6: Connection

Participants and providers wanted users to have the option to connect with other AYA cancer survivors through the tool as was shown in an exemplar interface. They suggested that this feature could facilitate meaningful connections and encourage mutual support among users.

> *I like how you can chat with the other people to experience the app together and talk with each other about your experience.* [20-Year-old thyroid cancer survivor, transcript]
>
> *I like that you have other people to chat with and can experience the course together.* [38-Year-old breast cancer survivor, transcript]
>
> *Message board or social networking aspect for discussion/reinforcement.* [Provider 2, social worker, workshop materials]

Subtheme 5.7: Private Expression

Participants expressed a desire for a feature within the tool that ensures the confidentiality and security of their emotions and thoughts. They emphasized the importance of having a safe space to express themselves without concerns about privacy.

> *There's some type of place where you don't have to worry at all about who's gonna see it or what's gonna see it, and you can just put it, and it's either completely just for you or you can give, like, only up to 3 people access, like, some sort of like the vault of all of my horrible feelings can go to this place. And I can just sort of dump in there, even if it's like, even if it is just for the person and just for them to see later, I think there's some value in doing that.* [38-Year-old breast cancer survivor, transcript]

Theme 6: Aspects to Avoid

Both participants and providers identified undesirable or problematic strategies for engagement. Addressing these concerns is crucial to ensure that the tool is user-friendly, engaging, and aligned with the needs of AYA cancer survivor users. We identified competition and negative tone as challenges to engagement.

Subtheme 6.1: Competition

This theme highlights the importance of avoiding features that may foster competition between participants. Providers suggested that ensuring the tool promotes a positive, non-competitive experience is essential for maintaining user engagement and fostering a sense of community.

> *Avoid having a "leaderboard" or "point system" with peers.* [Provider 1, medical oncologist, workshop materials]

XSL•FO
RenderX

Subtheme 6.2: Negative Tone

Participants emphasized the need to avoid negative or potentially discouraging content. It is important to ensure that all content is supportive, uplifting, and encouraging.

> *People who have a lot of self-hate towards themselves that might wanna want to put negative responses towards themselves that could possibly make them feel worse.* [21-Year-old survivor of multiple cancers, transcript]

The Prototype Depression Self-Management Tool

Based on feedback and recommendations from workshop participants, the initial design of the tool was determined by the study team (see Table 1 for a summary of how each theme and subtheme informed design decisions). There were a few conflicts between the providers' and AYA cancer survivors' suggestions; however, in the case of conflicting suggestions, AYA cancer survivors' preferences were prioritized as they are the target end users of the digital tool.

KM used Canva (Canva Inc) [49] to design a prototype of a depression self-management tool tailored to the needs and preferences of AYA cancer survivors. The tool was designed with the aim of providing mood tracking, psychoeducation about depression and cancer, and 4 adapted evidence-based interventions (aEBIs) for depression to this population via a web-based, app-optimized smartphone delivery (Figure 4). Each aEBI was designed to be housed in its own module, which would consist of a series of brief lessons. Each lesson would include educational content regarding the lesson topic (presented via video and transcript), a story of an AYA cancer survivor that illustrated the lesson topic, app questions that asked the user to apply the educational information to the AYA cancer survivors' story, an activity in which they reflected on what they learned related to their experience, and a practice activity that asked users to apply what they learned in their day-to-day life (Figure 5). The tool also featured a discussion board section that would allow users to connect with one another about the content presented via the tool.

XSL•FO
RenderX

Table 1. Summary of how each theme and subtheme informed design decisions.

Theme and subtheme	Design decision
Barriers to engagement	
Time constraints	Each aEBI[a] will be delivered via a series of brief, on-demand lessons that users can complete when it is convenient for them.
Lack of usability	A web-based application with a responsive design that has established usability among AYA[b] cancer survivors will be adapted for content delivery.
Desired content	
User guidance	During research studies, participants will complete an orientation session that includes instructions on how to use the digital tool. Once the tool has been optimized, a user guide will be developed and provided to users.
Information about mental health	Relevant mental health information will be included in the constant component so that it is provided to all participants. External resources for emergent mental health concerns will be provided to users via the constant component.
Resources	Each aEBI will include relevant practice activities to help users apply content in their daily lives. Future research will inform how to best integrate these practice activities into the digital tool.
AYA cancer survivors stories	The tool will use AYA cancer survivors' stories to illustrate educational content.
Cancer-specific content	The constant component and aEBIs will include information that is tailored to AYA cancer survivors.
Preferences for content delivery	
Intervention duration	Content will be delivered over a duration shorter than 9 weeks. Additional research will inform intervention duration.
Session duration	The constant component and each aEBI will be delivered via a series of brief lessons. Additional research will inform session duration.
Learning modalities	Educational content will be presented via multiple modalities, such as a brief video with a transcript provided. Additional research will inform what content should be presented via each modality.
Preferences for interface	
Ease of use	A web-based application with a responsive design that has established usability among AYA cancer survivors will be adapted for content delivery.
Color scheme	The color scheme will be updated to reflect the age/maturity level of the target user population.
Features	
Daily mood tracking	Users will be asked to track their mood daily rather than hourly.
Encouraging notifications	Users will receive encouraging text message reminders to engage with the tool. Additional research will inform the frequency of these messages.
Organizational tools	The tool will display users' progress toward completion of content and provide them with clear next steps.
Reward system	Additional research will be conducted to determine how to reward users for engagement in a way that is feasible and scalable.
Interactive exercise	Users will be asked to apply the information they learn to the stories of other AYA cancer survivors as well as their own experiences. Future research will explore how users prefer to input such responses in the tool.
Connection	The tool will include discussion boards that allow users to connect with other AYA cancer survivors.
Private expression	Users will be able to choose whether their responses to reflection questions are shared with other users.
Aspects to avoid	
Competition	Users will not be able to see the progress of other users.
Negative tone	A positive and encouraging tone will be used throughout the tool.

[a]aEBI: adapted evidence-based intervention.

[b]AYA: adolescent and young adult.

XSL•FO
RenderX

Figure 4. Dashboard and learning modules.

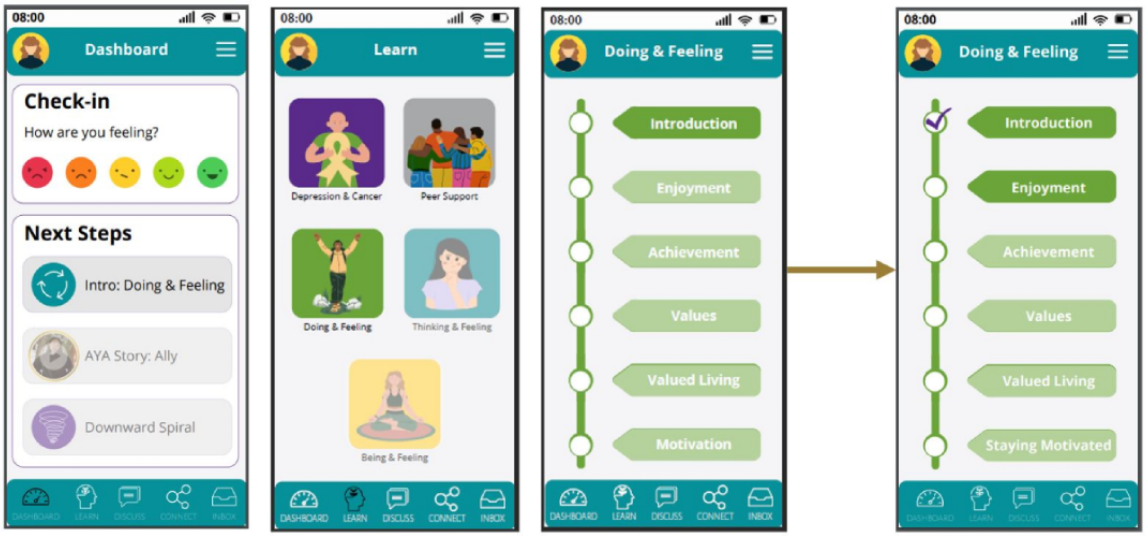
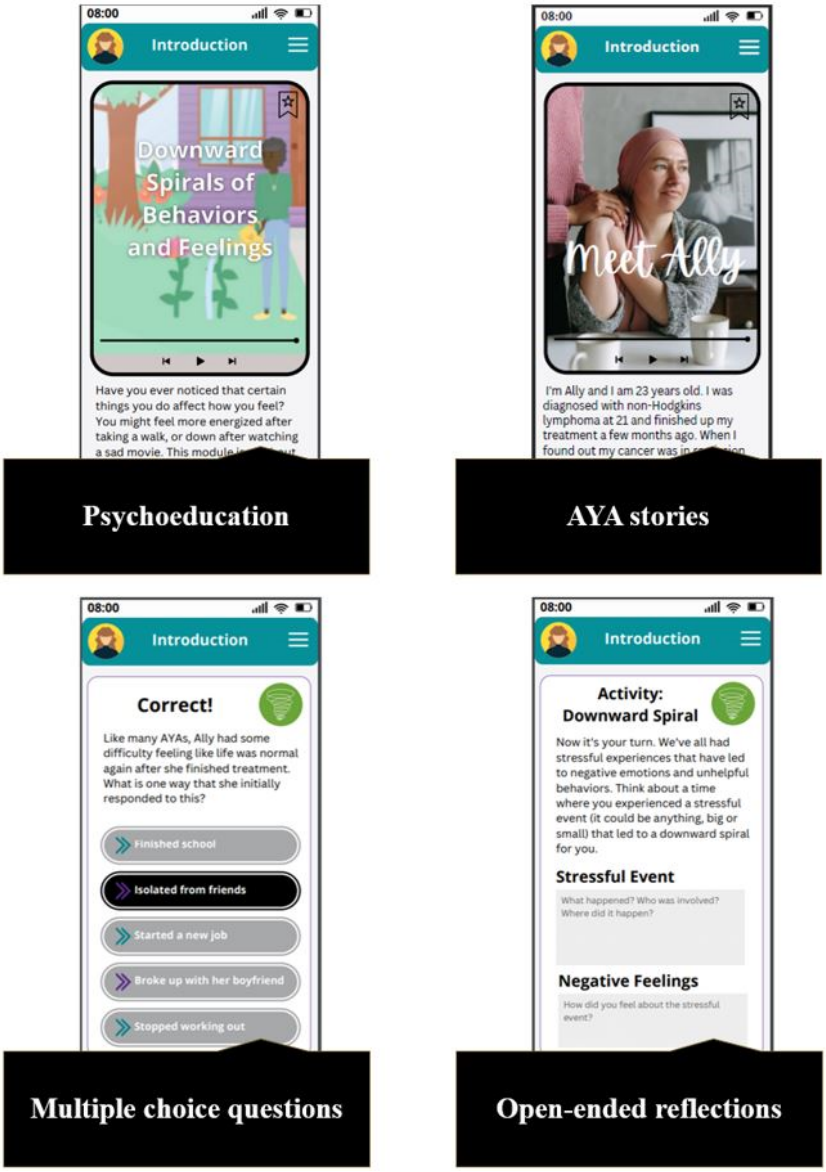


Figure 5. Example lesson structure and design. AYA: adolescent and young adult.



XSL•FO
RenderX

Discussion

Principal Findings

AYA cancer survivors often experience depressive symptoms that go unaddressed. Digital mental health tools are a promising strategy to ensure that patients have the skills and tools necessary to manage symptoms of depression. This study described the initial co-design of a depression self-management intervention to ensure that the final product is tailored to the needs and preferences of AYA cancer survivors, fulfilling the concept generation and ideation step of the UCD process set forth by McCurdie et al [39] and contributing to the tasks of the preparation phase of the MOST framework [42]. Feedback from AYA cancer survivors during the co-design workshops provided critical insights into ways in which a digital tool could be tailored to engage this population in evidence-based strategies for depression self-management. Specifically, users emphasized the importance of overcoming barriers to engagement, such as time constraints; provided suggestions for content to be included in the tool and how it should be delivered; and recommended features to enhance engagement with the tool. Using these findings, we designed a midfidelity prototype of a digital depression self-management tool tailored to the needs and preferences of AYA cancer survivors.

Comparison With Previous Work

A primary goal of the UCD process in this project was to ensure the tool we develop is engaging for AYA cancer survivors. As such, many of the themes identified in this study are consistent with previously identified strategies for promoting user engagement. For example, a systematic review of technology-supported strategies in digital mental health interventions identified four categories of engagement strategies used in existing tools: (1) reminders to use the tool, (2) tailored information, (3) peer support, and (4) coaching techniques [50]. In our study, the first 3 of these categories were emphasized by participants as important to include in the digital tool we are designing.

The themes identified in this study also broadly align with findings from co-design work in other groups of cancer survivors. For example, Adler et al [51] engaged cancer survivors with physical disabilities in a co-design process of a self-management tool. Their team identified themes of “engaging” and “socialization” that have significant overlap with our “features” theme. Participants in both studies suggested the use of participative features, progress tracking, and opportunities to connect with others who have had similar experiences as strategies to help users engage with a self-management tool. Similarly, in a co-design study focused on a tool to help cancer survivors engage in healthy behaviors, Leske et al [52] identified relatability and relevance as a theme, which has considerable overlap with the AYA stories and cancer-specific content subthemes of our desired content theme [52]. In both studies, it became evident that providing cancer-specific information and stories from peers would enhance the relevance of the tool for the target population.

Finally, this study extended previous research by engaging a novel population of AYA cancer survivors and their providers in the co-design process. This allowed us to elucidate strategies for tailoring that are uniquely relevant to this population, such as the importance of ensuring that the content is delivered in brief sessions to overcome the time constraints AYA cancer survivors experience, providing opportunities for private expression of their thoughts and feelings, and using a positive and encouraging tone throughout the tool.

Limitations

Our findings are limited by the size and characteristics of our research sample and the use of web-based rather than in-person activities. Our total sample size of 12 individuals across 5 workshops is consistent with recommendations for reaching saturation and sample sizes in design research with similar populations [48,53]. However, this small group of participants inherently limits the generalizability of this study. In addition, despite efforts to schedule as many AYA cancer survivors for the same workshop as possible, difficulties in scheduling resulted in groups of 2. These group brainstorming sessions would have benefited from larger groups with more people to generate ideas. Furthermore, despite efforts to recruit a diverse group of AYA cancer survivors and providers, we did not achieve a representative research sample with regard to race, ethnicity, gender, and sexual orientation. However, our AYA cancer survivors sample was representative of the vast majority of the AYA cancer survivors’ age range and common cancer diagnoses. Finally, the activities we were able to engage AYA cancer survivors in were limited to the capabilities of existing digital tools; we were not able to use the materials typically relied upon during in-person, co-design sessions, such as markers, Post-it notes, and posters. At times, we also encountered technical difficulties with the tools that limited user engagement in co-design activities. For example, one user was not able to load Figma on her device during the workshop, so she was limited to adding her insights verbally or through the chat feature of Webex. However, the use of online tools for co-design ultimately provided invaluable flexibility in terms of location and scheduling to ensure the engagement of busy AYA cancer survivors in the process.

Conclusions

This co-design study improved our understanding of the needs and preferences of AYA cancer survivors for a digital tool that may help them better manage symptoms of depression. This process allowed us to develop a shared vision of what such a digital tool would look like and how it would work. As such, our findings informed the development of a prototype of the tool. In addition, the findings from this work provide insights into the preferences of AYA cancer survivors for digital mental health tools more broadly. For example, AYA cancer survivors broadly agreed on the need for brief intervention models. However, it remains to be seen whether delivering interventions in a brief manner negatively impacts efficacy. As such, findings from this study are likely informative for future research aimed at designing engaging digital interventions for AYA cancer survivors. The next phase of this study is to have AYA cancer survivors evaluate the prototype through semistructured individual interviews to inform the design of a working prototype. This formative work will continue to contribute to

XSL•FO
RenderX

the tasks set forth by the preparation phase of MOST, as will a subsequent pilot study of the tool.

Acknowledgments

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under award K99CA248701. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We would like to thank Ms Nyla Parks for her assistance with cleaning transcriptions of the co-design workshops and Ms Kamah Woelfel for her assistance in recruiting participants for this study. The authors attest that there was no use of generative artificial intelligence technology in the generation of text, figures, or other informational content of this paper.

Data Availability

The datasets generated or analyzed during this study are available in the Open Science Framework repository [54].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Adolescent and young adult cancer survivor workshop slides.
[PPTX File , 6092 KB - formative_v9i1e67175_app1.pptx]

Multimedia Appendix 2
Provider workshop slides.
[PPTX File , 5096 KB - formative_v9i1e67175_app2.pptx]

References

1. Miller KD, Fidler-Benaoudia M, Keegan TH, Hipp HS, Jemal A, Siegel RL. Cancer statistics for adolescents and young adults, 2020. CA Cancer J Clin 2020;70(6):443-459 [FREE Full text] [doi: 10.3322/caac.21637] [Medline: 32940362]
2. Page LL, Devasia TP, Mariotto A, Gallicchio L, Mollica MA, Tonorezos E. Prevalence of cancer survivors diagnosed during adolescence and young adulthood in the United States. J Natl Cancer Inst 2024. [doi: 10.1093/jnci/djae250] [Medline: 39383200]
3. Hartung TJ, Brähler E, Faller H, Härter M, Hinz A, Johansen C, et al. The risk of being depressed is significantly higher in cancer patients than in the general population: Prevalence and severity of depressive symptoms across major cancer types. Eur J Cancer 2017;72:46-53. [doi: 10.1016/j.ejca.2016.11.017] [Medline: 28024266]
4. Salsman JM, Garcia SF, Yanez B, Sanford SD, Snyder MA, Victorson D. Physical, emotional, and social health differences between posttreatment young adults with cancer and matched healthy controls. Cancer 2014;120(15):2247-2254 [FREE Full text] [doi: 10.1002/cncr.28739] [Medline: 24888335]
5. Park EM, Rosenstein DL. Depression in adolescents and young adults with cancer. Dialogues Clin Neurosci 2022;17(2):171-180. [doi: 10.31887/dcns.2015.17.2/epark]
6. Tai E, Buchanan N, Townsend J, Fairley T, Moore A, Richardson LC. Health status of adolescent and young adult cancer survivors. Cancer 2012;118(19):4884-4891 [FREE Full text] [doi: 10.1002/cncr.27445] [Medline: 22688896]
7. Lang MJ, David V, Giese-Davis J. The age conundrum: a scoping review of younger age or adolescent and young adult as a risk factor for clinical distress, depression, or anxiety in cancer. J Adolesc Young Adult Oncol 2015;4(4):157-173 [FREE Full text] [doi: 10.1089/jayao.2015.0005] [Medline: 26697266]
8. Bellizzi KM, Smith A, Schmidt S, Keegan THM, Zebrack B, Lynch CF, et al. Positive and negative psychosocial impact of being diagnosed with cancer as an adolescent or young adult. Cancer 2012;118(20):5155-5162 [FREE Full text] [doi: 10.1002/cncr.27512] [Medline: 22415815]
9. Satin JR, Linden W, Phillips MJ. Depression as a predictor of disease progression and mortality in cancer patients: a meta-analysis. Cancer 2009;115(22):5349-5361 [FREE Full text] [doi: 10.1002/cncr.24561] [Medline: 19753617]
10. McQuaid JR, Lin EH, Barber JP, Breland-Noble AM, Cuijpers P, Greenberg LS. Clinical Practice Guideline for the Treatment of Depression Across Three Age Cohorts. Washington, DC: American Psychological Association; 2019.
11. Hollon SD, Beck AT. Cognitive therapy of depression. In: Cognitive-Behavioral Interventions: Theory, Research, and Procedures. New York, NY: Academic Press; 1979:153-203.
12. Segal Z, Williams M, Teasdale J. Mindfulness-Based Cognitive Therapy for Depression. New York, NY: Guilford Press; 2012.

XSL•FO
RenderX

13. Hendriks T, Schotanus-Dijkstra M, Hassankhan A, de Jong J, Bohlmeijer E. The efficacy of multi-component positive psychology interventions: a systematic review and meta-analysis of randomized controlled trials. J Happiness Stud 2019;21(1):357-390. [doi: 10.1007/s10902-019-00082-1]
14. Lim WL, Tierney S. The effectiveness of positive psychology interventions for promoting well-being of adults experiencing depression compared to other active psychological treatments: a systematic review and meta-analysis. J Happiness Stud 2023;24(1):249-273 [FREE Full text] [doi: 10.1007/s10902-022-00598-z] [Medline: 36373089]
15. Smith AW, Keegan T, Hamilton A, Lynch C, Wu X, Schwartz SM, et al. Understanding care and outcomes in adolescents and young adult with cancer: a review of the AYA HOPE study. Pediatr Blood Cancer 2019;66(1):e27486 [FREE Full text] [doi: 10.1002/pbc.27486] [Medline: 30294882]
16. Rabin C, Simpson N, Morrow K, Pinto B. Intervention format and delivery preferences among young adult cancer survivors. Int J Behav Med 2013;20(2):304-310. [doi: 10.1007/s12529-012-9227-4] [Medline: 22328444]
17. Mobile fact sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2024-09-06]
18. Devine KA, Viola AS, Coups EJ, Wu YP. Digital health interventions for adolescent and young adult cancer survivors. JCO Clin Cancer Inform 2018;2:1-15 [FREE Full text] [doi: 10.1200/CCI.17.00138] [Medline: 30652583]
19. Rathbone AL, Clarry L, Prescott J. Assessing the efficacy of mobile health apps using the basic principles of cognitive behavioral therapy: systematic review. J Med Internet Res 2017;19(11):e399 [FREE Full text] [doi: 10.2196/jmir.8598] [Medline: 29187342]
20. Huguet A, Rao S, McGrath PJ, Wozney L, Wheaton M, Conrod J, et al. A systematic review of cognitive behavioral therapy and behavioral activation apps for depression. PLoS One 2016;11(5):e0154248 [FREE Full text] [doi: 10.1371/journal.pone.0154248] [Medline: 27135410]
21. Dahne J, Lejuez CW, Diaz VA, Player MS, Kustanowitz J, Felton JW, et al. Pilot randomized trial of a self-help behavioral activation mobile app for utilization in primary care. Behav Ther 2019;50(4):817-827 [FREE Full text] [doi: 10.1016/j.beth.2018.12.003] [Medline: 31208690]
22. Ebert DD, Zarski AC, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
23. Ye X, Bapuji SB, Winters SE, Struthers A, Raynard M, Metge C, et al. Effectiveness of internet-based interventions for children, youth, and young adults with anxiety and/or depression: a systematic review and meta-analysis. BMC Health Serv Res 2014;14:313 [FREE Full text] [doi: 10.1186/1472-6963-14-313] [Medline: 25037951]
24. Clarke G, Kelleher C, Hornbrook M, Debar L, Dickerson J, Gullion C. Randomized effectiveness trial of an internet, pure self-help, cognitive behavioral intervention for depressive symptoms in young adults. Cogn Behav Ther 2009;38(4):222-234 [FREE Full text] [doi: 10.1080/16506070802675353] [Medline: 19440896]
25. Clarke AM, Kuosmanen T, Barry MM. A systematic review of online youth mental health promotion and prevention interventions. J Youth Adolesc 2015;44(1):90-113. [doi: 10.1007/s10964-014-0165-0] [Medline: 25115460]
26. Richter D, Koehler M, Friedrich M, Hilgendorf I, Mehnert A, Weißflog G. Psychosocial interventions for adolescents and young adult cancer patients: a systematic review and meta-analysis. Crit Rev Oncol Hematol 2015;95(3):370-386. [doi: 10.1016/j.critrevonc.2015.04.003] [Medline: 25922217]
27. Campo RA, Bluth K, Santacroce SJ, Knapik S, Tan J, Gold S, et al. A mindful self-compassion videoconference intervention for nationally recruited posttreatment young adult cancer survivors: feasibility, acceptability, and psychosocial outcomes. Support Care Cancer 2017;25(6):1759-1768. [doi: 10.1007/s00520-017-3586-y] [Medline: 28105523]
28. Zhang A, Zebrack B, Acquati C, Roth M, Levin NJ, Wang K, et al. Technology-assisted psychosocial interventions for childhood, adolescent, and young adult cancer survivors: a systematic review and meta-analysis. J Adolesc Young Adult Oncol 2022;11(1):6-16 [FREE Full text] [doi: 10.1089/jayao.2021.0012] [Medline: 33960845]
29. Sansom-Daly U, Wakefield C, Ellis S, McGill B, Donoghoe M, Butow P, et al. Online, group-based psychological support for adolescent and young adult cancer survivors: results from the recapture life randomized trial. Cancers (Basel) 2021;13(10):2460 [FREE Full text] [doi: 10.3390/cancers13102460] [Medline: 34070134]
30. Smith AW, Seibel NL, Lewis DR, Albritton KH, Blair DF, Blanke CD, et al. Next steps for adolescent and young adult oncology workshop: an update on progress and recommendations for the future. Cancer 2016;122(7):988-999 [FREE Full text] [doi: 10.1002/cncr.29870] [Medline: 26849003]
31. Zebrack B, Bleyer A, Albritton K, Medearis S, Tang J. Assessing the health care needs of adolescent and young adult cancer patients and survivors. Cancer 2006;107(12):2915-2923 [FREE Full text] [doi: 10.1002/cncr.22338] [Medline: 17103383]
32. Zebrack B, Isaacson S. Psychosocial care of adolescent and young adult patients with cancer and survivors. JCO 2012;30(11):1221-1226. [doi: 10.1200/jco.2011.39.5467]
33. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018;21(3):116-119 [FREE Full text] [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
34. App Uninstall Statistics 2018: 21% of users abandon an app after one use. Localytics. URL: https://uplandsoftware.com/localytics/resources/blog/21-percent-of-users-abandon-apps-after-one-use/ [accessed 2024-09-06]

XSL•FO
RenderX

35. Lyon AR, Koerner K. User-centered design for psychosocial intervention development and implementation. Clin Psychol (New York) 2016;23(2):180-200 [FREE Full text] [doi: 10.1111/cpsp.12154] [Medline: 29456295]
36. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008;23(3):454-466. [doi: 10.1093/her/cyn004] [Medline: 18349033]
37. Lustria MLA, Noar SM, Cortese J, van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
38. Schueller SM, Muñoz R, Mohr DC. Realizing the potential of behavioral intervention technologies. Curr Dir Psychol Sci 2013;22(6):478-483. [doi: 10.1177/0963721413495872]
39. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
40. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
41. Collins LM, Nahum-Shani I, Guastaferro K, Strayhorn JC, Vanness DJ, Murphy SA. Intervention optimization: a paradigm shift and its potential implications for clinical psychology. Annu Rev Clin Psychol 2024;20(1):21-47 [FREE Full text] [doi: 10.1146/annurev-clinpsy-080822-051119] [Medline: 38316143]
42. Landoll RR, Vargas SE, Samardzic KB, Clark MF, Guastaferro K. The preparation phase in the multiphase optimization strategy (MOST): a systematic review and introduction of a reporting checklist. Transl Behav Med 2022;12(2):291-303. [doi: 10.1093/tbm/ibab146] [Medline: 34850214]
43. Salsman JM, McLouth LE, Cohn M, Tooze JA, Sorkin M, Moskowitz JT. A web-based, positive emotion skills intervention for enhancing posttreatment psychological well-being in young adult cancer survivors (EMPOWER): protocol for a single-arm feasibility trial. JMIR Res Protoc 2020;9(5):e17078 [FREE Full text] [doi: 10.2196/17078] [Medline: 32463014]
44. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
45. Webex. URL: https://web.webex.com/sign-in [accessed 2024-09-12]
46. Figma: The Collaborative Interface Design Tool. URL: https://www.figma.com/ [accessed 2024-09-06]
47. Braun V, Clarke V. Successful Qualitative Research: A Practical Guide for Beginners. Thousand Oaks, CA: Sage Publications Ltd; 2013.
48. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022;292:114523 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
49. Canva. URL: https://www.canva.com/ [accessed 2024-09-06]
50. Gan DZQ, McGillivray L, Larsen ME, Christensen H, Torok M. Technology-supported strategies for promoting user engagement with digital mental health interventions: a systematic review. Digit Health 2022;8:20552076221098268 [FREE Full text] [doi: 10.1177/20552076221098268] [Medline: 35677785]
51. Adler RF, Morales P, Sotelo J, Magasi S. Developing an mHealth app for empowering cancer survivors with disabilities: co-design study. JMIR Form Res 2022;6(7):e37706 [FREE Full text] [doi: 10.2196/37706] [Medline: 35881439]
52. Leske M, Koczwara B, Blunt J, Morris J, Eakin E, Short CE, et al. Co-designing healthy living after cancer online: an online nutrition, physical activity, and psychosocial intervention for post-treatment cancer survivors. J Cancer Surviv 2024;18(2):606-616 [FREE Full text] [doi: 10.1007/s11764-022-01284-y] [Medline: 36374435]
53. Schwartz LA, Psihogios AM, Henry-Moss D, Daniel LC, Ver Hoeve ES, Velazquez-Martin B, et al. Iterative development of a tailored mHealth intervention for adolescent and young adult survivors of childhood cancer. CPPP 2019 Mar;7(1):31-43. [doi: 10.1037/cpp0000272]
54. Co-design workshop transcripts. Open Science Framework. URL: https://osf.io/8txsu/files/osfstorage?view_only=a51f736aedd44dff9ef59793f7bb7a59 [accessed 2025-03-04]

Abbreviations

aEBI: adapted evidence-based intervention
AYA: adolescent and young adult
CBT: cognitive behavioral therapy
EBI: evidence-based intervention
GIST: gastrointestinal stromal tumor
HIPAA: Health Insurance Portability and Accountability Act
MBCT: mindfulness-based cognitive therapy
MOST: multiphase optimization strategy
NCI: National Cancer Institute
REDCap: Research Electronic Data Capture

XSL•FO
RenderX

UCD: user-centered design

Edited by A Mavragani; submitted 10.10.24; peer-reviewed by K Devine; comments to author 10.01.25; revised version received 28.01.25; accepted 11.02.25; published 24.03.25.

Please cite as:
Murphy KM, Glock R, Victorson D, Reddy M, Birken SA, Salsman JM
Co-Design of a Depression Self-Management Tool for Adolescent and Young Adult Cancer Survivors: User-Centered Design Approach
JMIR Form Res 2025;9:e67175
URL: https://formative.jmir.org/2025/1/e67175
doi:10.2196/67175
PMID:40126551

©Karly M Murphy, Rachel Glock, David Victorson, Madhu Reddy, Sarah A Birken, John M Salsman. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Safe Listening Beliefs, Attitudes, and Practices Among Gamers and Esports Participants: International Web-Based Survey

Nicola Diviani[1,2,3], PhD; Shelly Chadha[4], PhD; Peter Mulas[4], MDeSc, MClinAud; Sara Rubinelli[1,2,3], PD, PhD

[1]Swiss Paraplegic Research, Nottwil, Switzerland

[2]Faculty of Health Sciences and Medicine, University of Lucerne, Lucerne, Switzerland

[3]Center for Rehabilitation in Global Health Systems, University of Lucerne, Lucerne, Switzerland

[4]Department of Noncommunicable Diseases, World Health Organization, Geneva, Switzerland

Corresponding Author:
Nicola Diviani, PhD
Swiss Paraplegic Research
Guido A Zäch Strasse 4
Nottwil, 6207
Switzerland
Phone: 41 419396583
Email: nicola.diviani@paraplegie.ch

Abstract

Background: The global rise of video gaming and esports has raised significant concerns about hearing loss due to loud sound exposure. While these activities provide entertainment and have applications in health care, the auditory health risks and behavioral factors influencing listening habits among gamers remain underexplored. Research is needed to develop tailored interventions that address the unique barriers, attitudes, and beliefs of gamers and esports participants, promoting safer listening practices and minimizing auditory health risks.

Objective: This study aimed to explore listening behaviors, attitudes, and awareness regarding hearing health risks among video gamers and esports participants. The findings are intended to guide the design and implementation of technological features that encourage safer listening practices, in alignment with the World Health Organization's Safe Listening initiative.

Methods: An open web-based survey was conducted from September 2022 to January 2023, targeting video gamers and esports enthusiasts. Participants were recruited via World Health Organization social media platforms and outreach to stakeholders. The survey assessed gaming behaviors, listening habits, awareness about hearing health, beliefs, readiness to change listening behaviors, and communication preferences. Data were analyzed using descriptive statistics and multinomial logistic regression.

Results: A total of 488 responses were collected, with 67.2% (n=328) of participants identifying as male, and 56.4% (n=275) having a college degree or higher. Of the respondents, 90.8% (n=443) were actively engaged in video gaming, while 54.9% (n=268) viewed esports, and 13.9% (n=68) participated in esports events. Notably, 24.8% (n=110) of gamers, 18.3% (n=49) of esports viewers, and 37.1% (n=23) of esports players reported using high or very high volume settings. Despite around half of the participants experiencing symptoms indicative of hearing damage (eg, ringing in the ears), only 34.3% (n=152) of gamers, 35.8% (n=92) of esports players, and 39.7% (n=27) of esports viewers reported taking sound breaks every hour. The study identified a balanced distribution across readiness-to-change stages, with 30.3% (n=148) in the precontemplation stage, 35.3% (n=173) in the contemplation stage, and 34.2% (n=167) in the action stage. Factors such as perceived susceptibility to hearing loss, perceived benefits of preventive action, and self-efficacy significantly influenced readiness to change. Communication preferences indicated that 51% (n=249) of participants were interested in receiving more information on hearing health, with health care professionals and governmental agencies being the most trusted sources.

Conclusions: The findings highlight an urgent need for interventions to promote safe listening practices among gamers, emphasizing a gap between awareness and preventive action. The integration of safe listening features into video games and esports platforms, along with targeted communication strategies, can enhance auditory health awareness and protective behaviors. Future research should evaluate the effectiveness of these interventions to ensure comprehensive auditory health protection in the digital entertainment sector.

(JMIR Form Res 2025;9:e60476) doi:10.2196/60476

XSL•FO
RenderX

KEYWORDS

video gaming; esports; hearing health; safe listening; auditory risk; health behavior; intervention design; games; listening; auditory; survey; college; data analysis; awareness; listening behavior; gaming

Introduction

Video gaming and esports have undergone an exponential rise in recent years, evolving into a major entertainment force that captivates individuals across different age groups [1]. With projections suggesting that the global video game market will grow at an annual rate of 12.9% from 2022 to 2030—reaching an estimated US $399.6 billion by 2028—and the global gamer population increasing from 2.7 billion in 2021 to an expected 3.04 billion by 2027, the influence of these digital platforms is significant [2,3]. Concurrently, esports has experienced a surge in popularity, with its global audience forecasted to reach 577 million by 2025, and the market is anticipated to grow at a rate of 16.7% annually, resulting in a value of US $4.28 billion by 2030 [4,5].

Beyond serving as engaging and entertaining mediums, video games and esports are increasingly acknowledged for their potential in health care, particularly through serious gaming applications aimed at education, rehabilitation, and therapy, thus highlighting their versatile utility beyond mere entertainment [6-8]. Despite these benefits, concerns about the potential health risks associated with video gaming and esports have emerged. The documented impacts on mental and physical health are compounded by the risks of hearing loss, tinnitus, and other auditory disorders resulting from exposure to loud and persistent sound effects inherent in these activities [9-11]. Echoing the World Health Organization's (WHO) warning, over 1 billion young people are at an elevated risk of hearing loss due to recreational exposure to loud sounds, including the use of personal audio devices and attendance at loud public events [12].

The public health realm has seen significant attention directed toward the issue of hearing loss from recreational sound exposure, leading to a wealth of research and the development of global standards aimed at mitigating this risk. The pervasive nature of this issue, covering activities ranging from personal listening device use to attendance at live events with high decibel levels, has been illuminated through extensive research. Such studies have consistently underscored the risk of hearing damage from prolonged exposure to loud music via earphones or headphones and attendance at entertainment venues where volumes exceed safe listening thresholds, potentially leading to noise-induced hearing loss [13-16].

In response, the WHO and the International Telecommunication Union have collaborated to launch the "Make Listening Safe" initiative, which recommends standards for incorporating safe listening features into personal audio devices. This includes software that monitors volume levels and exposure time, aiming to regulate sound exposure in recreational settings and promote a balance between enjoyment and health [17,18]. These standards not only aim to raise awareness among the public and stakeholders about the risks associated with loud sound exposure but also to provide concrete guidelines for safe listening practices, such as sound level limits and user warnings about unsafe listening volumes.

Research focused on hearing health within the context of video gaming and esports, on the other hand, remains practically nonexistent, particularly concerning unsafe listening habits and their contributing factors [19,20]. Existing research from different fields, such as physical activity promotion [21], and related areas like safe listening using personal audio devices [13,14], as well as behavioral models such as the health belief model (HBM) [22], can be useful starting points to better understand the factors influencing behavior and the barriers that need to be addressed. However, we know that interventions aimed at changing health behavior must be tailored to the specific characteristics and preferences of the population being targeted. A factor that plays a significant role in one context might indeed not have the same impact in another. For example, self-efficacy may be a major determinant for behaviors requiring significant skill, such as learning new physical activities [23], but may be less relevant for behaviors like adjusting volume settings, which require minimal effort. Similarly, barriers to safe listening, such as the worry about not enjoying music in noisy environments, could be a major factor for users of personal audio devices outdoors [13] but not for gamers who play indoors and often in solitude. In the gaming context, other factors, such as immersion or competitive performance, may be more significant barriers. The large body of existing evidence from other fields may therefore not directly apply to the gaming and esports environments, as behavioral models used to understand determinants of behavior must be specifically adapted to this context. Only by understanding how these models apply to video gamers and esports participants, we can effectively tailor interventions and communication strategies.

Our study specifically aims to address this gap by (1) gaining insights into the listening habits and safe listening behaviors of video gamers and esports players, exploring the determinants of these behaviors; (2) understanding the link between attitudes, beliefs, awareness, and stages of change regarding auditory health; and (3) identifying the most effective sources of information that can be leveraged to promote safer listening practices. These findings will support the development of interventions that can be integrated into industry standards, contributing to the design of technological features that encourage safer listening within gaming and esports platforms. By aligning with the WHO Safe Listening initiative, our work aims to ensure that auditory health considerations are integrated into the rapidly evolving landscape of video gaming and esports, fostering a more holistic approach to hearing health across all recreational sound exposure settings.

Methods

Sample

From September 2022 to January 2023, an open, web-based survey was shared on the WHO's social media platforms,

XSL•FO
RenderX

including posts on X (formerly Twitter) and Facebook. These posts were designed to reach a broad convenience sample of video gamers and esports enthusiasts. Additionally, key stakeholders were contacted via email and encouraged to share the survey link within their local networks to increase participation. No paid advertising or boosted posts were used for recruitment. Participation in the survey was completely voluntary, and no incentives or compensation were offered to participants. To participate in the survey, individuals had to have a good command of English, Spanish, French, or Chinese and be either video game players or involved in esports. No further specific eligibility criteria for access were specified, and participants could freely choose to participate. Of the 1043 individuals who submitted the first survey page, 532 (51%) completed the survey and were retained to be included in the analyses. To ensure that participants responded to the survey with adequate attention, response time was evaluated, and 35 (6.7%) individuals who completed the survey in less than 5 minutes were excluded, resulting in a total of 488 responses retained for the analyses.

Instrument and Measures

Overview

To develop the web-based questionnaire, a combination of established theories and empirical studies was used. Specifically, the HMB [22], the transtheoretical model (TTM) of change [24], and previous research on music- and noise-induced hearing loss were drawn upon [25]. The HBM suggests that individuals are more likely to adopt preventive measures if they perceive themselves to be at risk of developing a severe health condition (perceived susceptibility and severity), believe that taking action can prevent it (perceived benefits), and have the confidence to perform the necessary actions (self-efficacy). The TTM suggests that individuals can be categorized into different groups based on their readiness to adopt preventive behaviors.

The questionnaire (Multimedia Appendix 1) consisted of seven main sections designed to assess various aspects of participants' listening habits when gaming or engaging in esports. These sections included questions on (1) their gaming or esports behaviors and related listening habits (respondents were redirected to different sections of the survey depending on whether they played videogames, watched esports, participated in esports, or a combination), (2) the consequences of their listening behavior, (3) their knowledge and awareness of hearing loss, (4) their beliefs about hearing loss and safe listening, (5) their readiness to change their listening behavior, (6) their desire for information on hearing loss and safe listening and their preferred sources of information, and (7) their sociodemographic information.

The survey consisted of 10 pages, each containing 3 to 7 items. Respondents were required to answer basic and sociodemographic questions, which included a nonresponse option. All other items were optional and could be left unanswered. Completeness checks were performed after submission, highlighting mandatory items if incomplete. A "Back" button allowed respondents to review and modify their answers before final submission.

The survey was pretested with 15 individuals who had experience in video gaming or esports. Their feedback helped refine the questions for clarity and relevance, ensuring that the final version was suitable for the target population. The survey was hosted on the Qualtrics web-based platform, a widely used platform that ensures data security through end-to-end encryption and complies with General Data Protection Regulation guidelines to protect participant anonymity. Several accessibility features, such as adjustable text sizes and screen reader compatibility, were included to accommodate participants with disabilities.

To prevent multiple submissions from the same participant, cookies were used on the introduction page and were valid for 24 months. IP address checks were not used to identify duplicate entries. Data collected through the survey were stored in secure cloud storage provided by Qualtrics, with all responses anonymized, and participants assigned unique identifiers to ensure confidentiality. The data will be securely stored for 5 years and then deleted in compliance with data protection regulations. We adhered to the CHERRIES (Checklist for Reporting Results of Internet E-Surveys; Multimedia Appendix 2) for reporting the results.

Gaming and Esports Behaviors and Related Listening Habits

To assess the frequency of video game or esports viewing or playing and volume-level preferences, a series of single-item indicators were used. Participants were asked to rate the frequency of their gaming or esports activities, as well as the volume level they typically set on their devices, using a scale ranging from 1=minimum volume to 10=maximum volume. These questions were adapted from measures used in prior research on the topic, such as the Listening Habits Questionnaire (LHQ) [25].

Preventive Behaviors and Consequences of Listening

Participants were asked to indicate whether they take a break from sound every hour and whether they check the sound dose. The frequency of experiencing a ringing or fuzziness in the ears after playing video games or engaging in esports was measured on scales ranging from 1=never to 5=always. Participants were asked to rate how often they experienced these symptoms using questions from the LHQ [25].

Awareness and Knowledge

Participants were presented with 3 statements related to hearing loss in the context of gaming and esports, which were collaboratively developed with a team of international experts in the field of hearing. The statements included "listening to sounds above 85dB over a period of time can cause permanent damage to your hearing," "the amount of time you listen to a sound affects how much damage it will cause," and "listening to video games or esports sounds at loud listening levels may damage hearing." Participants were asked to indicate whether they believed each statement to be true or false and were scored accordingly. A knowledge score was then computed by summing the scores for all 3 statements, resulting in a range of 0 to 3. The average knowledge score was 2.77 (SD 0.61), with an α coefficient of .690.

XSL•FO
RenderX

Beliefs

An adapted version of a section of the LHQ [25] was used to assess the constructs of the HBM. The adaptation involved removing the distinction between quiet and loud environments after internal pretesting. Participants were then asked to rate their agreement on a 7-point scale (with higher scores indicating greater agreement) on a total of 22 statements about hearing loss susceptibility related to gaming and esports. These statements covered perceived susceptibility to hearing loss (4 items, α=.776), perceived severity of hearing loss (6 items, α=.865), benefits of preventing hearing loss (5 items, α=.830), barriers to preventing hearing loss (3 items, α=.638), and self-efficacy for taking preventative action (4 items, α=.845). The scores for each construct were obtained by averaging the individual scores on items within each subscale.

Readiness to Change

To assess readiness to change listening behaviors, a set of 12 statements was adapted from a survey by Rollnick et al [26] for excessive drinkers and a previous study on safe listening in the context of personal listening devices [13]. Each statement represented one aspect of the 3 stages of the TTM (precontemplation, contemplation, and action). Respondents answered each statement on a 5-point scale (ranging from 2=strongly disagree to +2=strongly agree), as per the authors' instructions. The subscales showed acceptable to good internal consistency (precontemplation: α=.626; contemplation: α=.740; and action: α=.784). Scores for items related to the same stage were added, resulting in a stage score ranging from –8 to +8. The respondent's stage of change was determined by the stage with the highest score. If 2 stages had equally high scores, the stage furthest along the change process was chosen (eg, if precontemplation and contemplation had the highest scores, contemplation was designated as the stage of change).

Communication Needs and Preferences

Participants were asked if they desired more information about hearing loss and safe listening, with response options of yes or no. Those who answered affirmatively were then asked to indicate their preferred channels for receiving this information on a 7-point scale, ranging from 1=definitely no to 7=definitely yes, which included options such as public health campaigns, gaming device interfaces, and social media. Additionally, all participants were asked to rate how much they trust in different sources of information about hearing loss and safe listening, such as health care professionals, governmental and international agencies, and family and friends, using a 7-point scale ranging from 1=not at all to 7=completely.

Sociodemographics

Participants were requested to provide information about their sex, age, the highest level of education completed, and country of residence. For analytical purposes, sex (male and female), age (16-35 and >35 years), and educational level (low and high) were dichotomized.

Data Analysis

The data were analyzed using SPSS Statistics (version 21.0; IBM Corp). Descriptive statistics, including basic frequencies, were calculated to provide an overview of the sample characteristics across all variables of interest, such as demographics, knowledge, and beliefs related to hearing health. To assess group differences in the descriptive analyses, we used chi-square tests for categorical variables and 1-way ANOVAs for continuous variables, supplemented by effect size measures such as Cramer *V* or φ for chi-square tests and partial eta squared (η_p^2) for ANOVAs. These tests and effect size measures allowed us to compare key differences between sex, age, and education subgroups within the sample, providing a more detailed understanding of both the statistical significance and the magnitude of group differences across the variables of interest. To assess the relationship between predictors (knowledge and beliefs about hearing health) and the outcome variable (readiness to change listening behaviors), we performed a multinomial logistic regression analysis. We chose this method because the outcome variable (readiness to change) is categorical with multiple levels, and it allowed us to evaluate how each predictor influenced the likelihood of participants being in different stages of readiness to change while controlling for other variables. The use of multinomial logistic regression was particularly suited to account for the complexity of the data and to handle multiple predictors influencing a nonbinary categorical outcome. All assumptions for the chi-square tests, ANOVAs, and multinomial logistic regression (such as normality, independence, and multicollinearity) were tested and met, ensuring the robustness of our findings.

Ethical Considerations

Ethics approval for research in Switzerland is governed by the Federal Act on Research Involving Human Beings (Human Research Act, HRA, SR 810.30 [27]). According to chapter 1, article 2, paragraph 2 of this act, ethics approval is required for research involving human participants. However, our study, which involves an anonymous web-based survey, does not fall within the scope of the HRA. Specifically, article 2, paragraph 2 states: "It [the HRA] does not apply to research which involves: a. IVF embryos in accordance with the Stem Cell Research Act of 19 December 2003; b. anonymized biological material; c. anonymously collected or anonymized health-related data." Therefore, our project does not require ethics approval (see also the Swissethics decision tree [28]). This web-based anonymous survey was conducted in strict adherence to the principles outlined in the Declaration of Helsinki. Informed consent was obtained from all participants prior to starting the survey. Participants were presented with a web-based informed consent form, which included information about the voluntary nature of the study, data anonymity, and their right to withdraw at any time without any consequences. Consent was confirmed through a mandatory checkbox before proceeding to the survey questions. No compensation was provided to study participants.

Results

Overview

Of the survey participants, the majority were male (n=355, 67.7%), with an average age of 28.4 (SD 10.8) years. The highest level of education attained by most respondents was a 4-year college degree (n=107, 21.9%) or a high school diploma

XSL•FO
RenderX

or General Educational Development (n=101, 20.7%). Although respondents hailed from 92 different countries, the largest number of respondents came from the United States (n=70, 14.3%), followed by the United Kingdom (n=51, 10.5%) and India (n=40, 8.2%). More details about the participants are shown in Table 1.

Table 1. Sample characteristics (N=488).

	Values, n (%)
Sex	
Male	328 (67.2)
Female	144 (29.5)
No answer	16 (3.3)
Age (years)	
16-35	386 (79.1)
>35	102 (20.9)
Education	
Low education (less than college)	196 (40.2)
Middle education (college)	133 (27.3)
High education (more than college)	142 (29.1)
No answer	17 (3.5)
Country of origin (top 5)	
United States	70 (14.3)
United Kingdom	51 (10.5)
India	40 (8.2)
Belgium	34 (7)
Australia	21 (4.3)
Involved in (more than 1 answer possible)	
Video games	443 (90.8)
Esports viewing	268 (54.9)
Esports participation	68 (13.9)

Gaming and Esports Behaviors and Related Listening Habits

Details of the overall gaming and esports behaviors and the related listening habits are shown in Table 2 (video game players) and Table 3 (esports viewers and esports players).

XSL•FO
RenderX

Table 2. Behaviors and related listening habits of video game players (n=443).

	Values
Time spent playing (hours per week), n (%)	
0-1	29 (6.5)
1-2	30 (6.8)
3-4	50 (11.3)
4-6	49 (11.1)
6-10	73 (16.5)
10-15	74 (16.7)
15-30	82 (18.5)
>30	56 (12.6)
Volume setting, mean (SD) (scale 1-100)	47.6 (21.3)
Volume setting, n (%)	
Very low (0%-20%)	49 (11.1)
Low (20%-40%)	134 (30.2)
Medium (40%-60%)	150 (33.9)
High (60%-80%)	85 (19.2)
Very high (80%-100%)	25 (5.6)
Importance of sounds, n (%)	
Not at all important	14 (3.2)
Not so important	25 (5.6)
Somewhat important	87 (19.6)
Very important	151 (34.1)
Extremely important	166 (37.5)
Sound break every hour, n (%)	
Yes	152 (34.3)
Checks use statistics, n (%)	
Yes	45 (10.2)
Experience ringing in ears, n (%)	
Never	253 (57.1)
Rarely	125 (28.2)
Sometimes	51 (11.5)
Often	6 (1.4)
Always	8 (1.8)
Experience fuzziness in ears, n (%)	
Never	206 (46.5)
Rarely	148 (33.4)
Sometimes	71 (16)
Often	11 (2.5)
Always	7 (1.6)

XSL•FO
RenderX

Table 3. Behaviors and related listening habits of esports viewers and players.

	Viewers (n=287)	Players (n=62)
Frequency of viewing, n (%)		
Daily	14 (5.2)	15 (24.2)
Weekly	75 (28)	23 (37.1)
Monthly	79 (29.5)	11 (17.7)
Yearly	68 (25.4)	6 (9.7)
Less than yearly	32 (11.9)	7 (11.3)
Volume setting, mean (SD) (scale 1-100)	45.6 (20.4)	55.7 (22.5)
Volume setting, n (%)		
Very low (0%-20%)	34 (12.7)	5 (8.1)
Low (20%-40%)	80 (29.9)	10 (16.1)
Medium (40%-60%)	105 (39.2)	24 (38.7)
High (60%-80%)	36 (13.4)	14 (22.6)
Very high (80%-100%)	13 (4.9)	9 (14.5)
Experience ringing in ears, n (%)		
Never	186 (69.4)	36 (53.7)
Rarely	46 (17.2)	20 (29.9)
Sometimes	29 (10.8)	6 (9)
Often	2 (0.7)	0 (0)
Always	5 (1.9)	5 (7.5)
Experience fuzziness in ears, n (%)		
Never	187 (69.8)	33 (49.3)
Rarely	49 (18.3)	18 (26.9)
Sometimes	23 (8.6)	8 (11.9)
Often	4 (1.5)	3 (4.5)
Always	5 (1.9)	5 (7.5)
Importance of sounds, n (%)		
Not at all important	10 (3.7)	1 (1.5)
Not so important	33 (12.3)	8 (11.8)
Somewhat important	79 (29.5)	11 (16.2)
Very important	80 (29.9)	20 (29.4)
Extremely important	66 (24.6)	28 (41.2)
Break from sound every hour, n (%)		
Yes	96 (35.8)	27 (39.7)
Checks use statistics, n (%)		
Yes	37 (13.8)	12 (17.6)

Video Game Players

Of the surveyed participants, 90.8% (n=443) reported engaging in video games (Table 2). Around one-third of gamers (n=158, 35.7%) reported playing for up to 6 hours per week, followed by a second group (n=147, 33.2%) who played between 6 and 15 hours per week, and a third group (n=138, 31.1%) who played more than 15 hours per week. Male respondents reported significantly more time spent on video gaming compared to female respondents (P=.008; φ=0.210), while those with a higher educational level spent significantly less time playing video games compared to those in the lower and medium educational groups (P=.03; Cramer V=0.175). The average volume setting reported by participants was just below 50% (mean 47.6, SD 21.3), with almost 1 in 4 respondents listening at a high or very high volume (n=110, 24.8%). No significant differences were

XSL•FO
RenderX

found across sexes, ages, or educational levels. Finally, most respondents indicated that sounds are either extremely (n=166, 37.5%) or very (n=151, 34.1%) important when playing video games. Male respondents, on average, placed greater importance on video game sounds than female respondents (P=.03; Cramer V=0.161).

Esports Viewers and Players

More than half of the study participants (n=281, 57.6%) reported engaging in esports, with most of them only viewing (n=213, 75.8%) and approximately one-fifth (n=55, 19.6%) both viewing and participating (Table 3). A small percentage (n=13, 4.6%) reported only participation. Among viewers (n=268), the majority viewed on a monthly basis (n=79, 29.5%). Male participants were found to watch esports events more often than female participants (P=.03; Cramer V=0.202). The average volume setting reported was slightly below 50% (mean 45.6, SD 20.4), with almost 1 in 5 respondents listening at a high or very high volume (n=49, 18.3%). Participants with lower educational levels reported significantly higher volume settings when watching esports events ($F_{2,254}$=3.139; P=.04), while no significant differences were observed across sex or age groups.

Among esports participants (n=68), the majority took part in esports events on a weekly basis (n=23, 37.1%). No significant differences were observed regarding sex, age, or education. The average volume setting for esports participation was higher than for viewing at 55.7 (SD 22.5), with more than 1 in 5 respondents listening at a high or very high volume (n=23, 21%). No significant differences were observed across sexes, ages, educational levels, or frequency of participation.

Overall, esports viewers and players consider sounds in esports as either somewhat (n=82, 29.2%), very (n=85, 30.2%), or extremely (n=68, 24.2%) important. Among esports players, in particular, sound was considered extremely important by a majority (n=28, 41.2%). Participants with lower educational levels placed significantly greater importance on sounds in esports events compared to those with higher educational levels (P=.02; Cramer V=0.183).

Protective Behavior and Consequences of Listening

Only about one-third of video game players (n=152, 34.3%) reported taking a break from sound every hour. Just 1 in 10 (n=45, 10.2%) reported checking the sound dosage information provided by the gaming device. Male and female respondents differed significantly in whether they checked information on how much time they spend listening and their sound dosages through their gaming device (P=.03; Cramer V=0.144). A similar picture emerged for esports players and participants. Over one-third of them (n=106, 37.7%) reported taking an hourly break, and only a few reported checking the sound dosage (n=43, 15.3%). No group differences were observed.

Regarding the consequences of listening, almost half of the video game players reported having experienced at least once a ringing in their ears after playing (n=190, 42.9%), and more than half have experienced fullness or fuzziness in their ears (n=237, 53.5%). Female respondents reported feeling their ears full or fuzzy after playing video games more often than male respondents (P=.006; Cramer V=0.184). Regarding esports, we observed 2 different pictures for viewers and players. Among esports viewers, less than one-third reported having experienced a ringing in their ears (n=82, 30.6%) or a feeling of fullness or fuzziness (n=81, 30.2%) after watching esports. Among esports players, however, this percentage increased to 46.3% (n=31) for the ringing in their ears and to 50.7% (n=34) for the feeling of fullness or fuzziness. Participants' characteristics were not associated with the frequency of experiencing consequences.

Awareness and Knowledge

Overall, the respondents were very knowledgeable about the risks for hearing related to exposure to video games, scoring on average 2.78 (SD 0.62) of a maximum of 3 points indicating a correct answer to all questions. No significant differences among groups were observed.

Beliefs

The majority of participants rated the perceived severity of hearing loss as high (mean 5.84, SD 1.31), while their perceived susceptibility to it was relatively low (mean 3.70, SD 1.54). Participants recognized the benefits of preventing hearing loss (mean 5.84, SD 1.22) but also faced significant barriers to doing so (mean 4.53, SD 1.56), such as the belief that turning down the volume would limit their enjoyment of video games and esports. Despite this, respondents demonstrated high confidence in their ability to modify their listening behavior (mean 5.24, SD 1.57). We observed few group differences in beliefs, except for higher self-efficacy among male respondents ($F_{1,470}$=4.13; P=.04; η_p^2=0.009) and lower perceived susceptibility among younger respondents ($F_{1,486}$=12.33; P<.001; η_p^2=0.025). In addition, respondents with higher education reported perceiving more benefits of prevention compared to the other 2 groups ($F_{2,468}$=3.65; P=.03; η_p^2=0.015).

Readiness to Change

Overview

We observed a balanced distribution of participants into the 3 readiness-to-change groups. Most of them (n=173, 35.3%) could be categorized as belonging to the "contemplation" stage, followed by the "action" stage (n=167, 34.2%) and the "precontemplation" stage (n=148, 30.3%). A chi-square test revealed a significant association between educational level and stage of change (P=.009; Cramer V=0.120), with individuals in the high education group more likely to be in the action stage (n=60, 42.3%) compared to those in the medium (n=50, 37.6%) and low education groups (n=49, 25%). Table 4 shows the multinomial logistic regression results for factors associated with being in one of the 3 readiness-to-change stages. Those who were in the contemplation stage were the reference category.

XSL•FO
RenderX

Table 4. Associations of knowledge and beliefs with readiness to change (N=488)[a].

	Readiness to change			
	Precontemplation		Action	
	OR[b] (95% CI)	*P* value	OR (95% CI)	*P* value
Knowledge	1.293 (0.769-0.174)	.33	0.998 (0.628-1.584)	.99
Perceived susceptibility	0.557 (0.452-0.687)	<.001	1.031 (0.851-1.249)	.76
Perceived severity	1.344 (1.056-1.712)	.02	1.082 (0.868-1.349)	.48
Benefits of preventive action	0.611 (0.461-0.809)	.001	0.968 (0.732-1.281)	.82
Barriers to preventive action	0.985 (0.818-1.186)	.54	0.808 (0.680-0.959)	.01
Self-efficacy in prevention	1.518 (1.237-1.863)	<.001	1.431 (1.196-1.711)	<.001

[a]Multinomial logistic regression analysis with contemplation as the reference category. The model is adjusted for sex, age, and education. Pseudo R^2: Cox and Snell=0.274; Nagelkerke=0.308.

[b]OR: odds ratio.

Multivariate Results: Contemplation Versus Precontemplation Stage

The second column of Table 4 shows the odds ratio (OR) for being in the precontemplation stage versus being in the contemplation stage for each variable listed. The odds of being in the precontemplation versus the contemplation group decrease the more one perceives to be susceptible to hearing loss (OR 0.557, 95% CI 0.452-0.687) and the more one perceives the benefits of preventive actions (OR 0.611, 95% CI 0.461-0.809). On the other hand, the odds of being in the precontemplation group increase the more one perceives hearing loss as severe (OR 1.344, 95% CI 1.056-1.712) and the more one has self-efficacy in the prevention of hearing loss (OR 1.518, 95% CI 1.237-1.863).

Multivariate Results: Contemplation Versus Action Stage

The fourth column of Table 4 shows the OR for being in the action stage versus being in the contemplation stage for each variable listed. The odds of being in the action versus the contemplation group decrease the more barriers to preventive action for hearing loss one perceives (OR 0.808, 95% CI 0.680-0.959) and increase the more one has self-efficacy in the prevention of hearing loss (OR 1.431, 95% CI 1.196-1.711).

Communication Needs and Preferences

More than half of the participants (n=249, 51%) expressed an interest in expanding their knowledge on the topics of hearing loss and safe listening in the context of video gaming and esports. While investigating the preferred channels for obtaining such information, it was found that no single mode was unequivocally favored. Nonetheless, noteworthy options included instruction manuals for gaming devices, dedicated websites, and user interfaces integrated into gaming devices. No significant differences were observed in preferences for sex, age, or education. Regarding trusted sources of information, respondents showed a clear preference for health care professionals and governmental or international agencies (eg, WHO), while traditional media and religious leaders were among the least trusted sources. Female respondents reported significantly higher trust in government or international health agencies (eg, WHO) compared to male respondents ($F_{1,470}$=8.76; P=.003; η_p^2=0.018). Younger respondents reported higher trust in government or international health agencies ($F_{1,486}$=4.98; P=.03; η_p^2=0.010) as well as higher trust in producers of gaming or esports devices ($F_{1,486}$=15.12; P<.001; η_p^2=0.030) and doctors or pharmacists ($F_{1,486}$=12.59; P<.001; η_p^2=0.025). Significant differences in trust were observed across educational levels, with respondents holding lower education reporting higher trust in producers of gaming or esports devices ($F_{2,468}$=4.88; P=.008; η_p^2=0.020) as well as higher trust in newspapers or magazines ($F_{2,468}$=4.00; P=.02; η_p^2=0.017). More details about communication needs and preferences can be found in Table 5.

XSL•FO
RenderX

Table 5. Communication needs and preferences (N=488).

	Values
Interest in receiving more information, n (%)	
Yes	249 (51)
Preferred channels[a], mean (SD)	
Instructions of gaming device	5.91 (1.58)
User interface of gaming device	5.84 (1.79)
Dedicated website	5.82 (1.65)
Social media	5.70 (1.78)
Mass media	5.43 (1.89)
Public health campaign	5.29 (1.95)
Written information	4.90 (2.14)
Interpersonal communication	4.86 (2.11)
Public events	4.20 (2.31)
Trust in sources of information[b], mean (SD)	
Health care professionals	6.22 (1.32)
Governmental or international agencies	6.01 (1.46)
Charitable organizations	4.65 (1.76)
Producers of gaming devices	4.48 (1.90)
Family or friends	4.21 (1.66)
Internet	4.19 (1.65)
Radio	3.87 (1.80)
Newspapers or magazines	3.85 (1.80)
Television	3.78 (1.75)
Religious organizations or leaders	2.74 (1.98)

[a]Scores ranging from 1 to 7, with higher scores indicating a higher willingness to receive information from the source.

[b]Scores ranging from 1 to 7, with higher scores indicating higher trust in the source.

Discussion

Overview

In this study, we aimed to explore the gaming and esports behaviors, related listening habits, and awareness about hearing health risks among gamers and esports participants. This investigation was motivated by the increasing concerns over potential hearing damage due to prolonged exposure to high sound levels in video gaming environments and the need for designing and implementing specifically designed features that make video games safer.

Principal Findings

Our first specific aim was to gain insights into the listening habits and safe listening behaviors of video gamers and esports players, exploring the determinants of these behaviors. Our results show that an important part of players set the volume at high or very high levels, with younger participants reporting significantly higher volume settings than older respondents. Furthermore, we observed that most viewers and players consider sounds in esports to be (very) important, and the volume level for esports participation was higher than for viewing. Although most respondents were aware of the risks of loud sound exposure and recognized the benefits of prevention, they still engaged in risky listening behaviors. For instance, only about one-third of video game and esports players reported taking a break from sound every hour, and only a few reported checking the sound dosage. This lack of protective behavior is particularly concerning, given that around half of the video game and esports players reported having experienced at least once a ringing or feelings of fullness or fuzziness in their ears after playing. These findings suggest that many video gamers and esports players are experiencing symptoms of hearing damage, yet are not taking the necessary steps to protect their hearing.

Second, we aimed to understand the link between attitudes, beliefs, awareness, and stages of change regarding auditory health. Applying the constructs of the HBM, we were able to determine that, despite an overall recognition of the risks associated with loud sound exposure, several factors get in the way of a positive behavior change. For instance, many believed that turning down the volume would limit their enjoyment of

XSL•FO
RenderX

video games and esports. In addition, while many rated the perceived severity of hearing loss as high, their perceived susceptibility to it was relatively low. Finally, low self-efficacy also seems to play a role in the decision to adopt safer listening practices. These results are in line with what is suggested by current theories of behavior change and evidence in the field of health behaviors and of safe listening in other contexts, such as findings in physical activity promotion, where individuals are often aware of the benefits of exercise but are deterred by perceived barriers like time constraints or lack of motivation [29]. Similarly, in safe listening research for personal audio devices, users often understand the risk of hearing damage but continue to listen at high volumes, believing that lowering the volume would reduce sound quality or enjoyment, especially in noisy environments [13]. Concretely, these findings suggest a need for education programs specifically designed to show how prevention does not require particular skills and does not necessarily limit enjoyment (eg, through testimonials [30]) and to increase the perceived susceptibility [31] to hearing loss among video gamers and esports players.

We also observed a balanced distribution of participants into the 3 readiness-to-change groups, as defined by the TTM [24], with most of them belonging to the "contemplation" stage, followed by the "action" stage and the "precontemplation" stage. While many video gamers and esports players are aware of the risks of loud sound exposure and are considering taking action, they may need additional support to move from contemplation to action. The multivariate analysis revealed several factors that were associated with participants' readiness to change their listening behavior in the context of video gaming and esports. Participants who perceived themselves as more susceptible to hearing loss and recognized the benefits of preventive actions were less likely to be in the precontemplation stage and more likely to be in the contemplation stage. On the other hand, participants who perceived hearing loss as severe and had more self-efficacy in preventing hearing loss were more likely to be in the precontemplation stage. Participants who had more self-efficacy in preventing hearing loss were more likely to be in the action stage than in the contemplation stage. Finally, participants who perceived more barriers to preventive actions were less likely to be in the action stage. These findings suggest that interventions aimed at promoting safe listening habits among video gamers and esports viewers and players should be tailored to participants' beliefs and attitudes about hearing loss and preventive actions [32]. In particular, our findings suggest that interventions to increase self-efficacy [33] may be effective in promoting safe listening habits among video gamers and esports players independent of the stage they are in. At the same time, interventions for those in the precontemplation stage should focus on increasing the perceived susceptibility and on showing the benefits of prevention, while those targeted at those in the contemplation stage should include information aimed at reducing barriers, for instance, by showing how listening to lower volumes does not reduce enjoyment.

Finally, we aimed to identify the most effective sources of information that can be leveraged to promote safer listening practices. Approximately half of the participants expressed a keen interest in expanding their knowledge on hearing loss and safe listening in the context of video gaming and esports, suggesting a good degree of openness to information. Regarding trusted sources of information, respondents showed a clear preference for health care professionals and governmental or international agencies. Female respondents showed significantly more trust in governmental or international agencies, while younger respondents showed more trust in health care professionals, governmental or international agencies, and producers of gaming devices. This information is important, as it can inform future communication efforts especially tailored to video gamers and esports players.

Overall, our findings not only highlight the importance of promoting safe listening behaviors among video gamers and esports players but also contribute to the broader literature on behavior change. By identifying key barriers and beliefs unique to this population—such as prioritizing immersion and performance over auditory health—our research offers new insights into how behavior change models, like the HBM, can be tailored to different contexts. This contributes to a more nuanced understanding of how interventions should be tailored to address context-specific motivations and barriers. In doing so, our study advances the development of targeted interventions aimed at reducing unsafe listening habits, with the potential to inform both gaming platform design and broader behavior change strategies. These findings are significant for the ongoing development of health promotion strategies that encourage long-term behavioral change, reinforcing the importance of adapting existing models to the specific needs of the different communities.

Practice Implications

Our findings support the need for the design and implementation of safe listening features in video games and also provide useful insights into how these features could look like. First, the substantial engagement in video gaming and the pronounced importance of sound highlight the necessity for interventions that do not compromise sound quality. Features enabling the tracking of the sound dose, such as dosimetry [34] or dynamic range compression [35], have emerged as crucial innovations, which could ensure that gamers can remain engaged and fully immersed without risking their hearing health.

Second, the observation of gamers frequently listening at high volumes and not engaging in protective practices, such as regular sound breaks, motivates the introduction of features enabling an automatic reduction of volume [36] and in-game notifications [37] related to measured sound exposure. These interventions would directly encourage gamers to adjust their volume settings and take necessary breaks, promoting safer listening habits organically within the gaming environment.

Third, the reported adverse effects, including ringing in the ears and feelings of fullness, highlight the need for features allowing users to remove particularly harmful sounds [38] and, considering the detrimental impact of tinnitus in terms of psychological or emotional effects, sleep disturbance, auditory issues, and overall health [39], support the introduction of general safe listening warnings in video games. These features aim to provide a more comfortable gaming experience for individuals experiencing tinnitus, reduce the experiences of

XSL•FO
RenderX

tinnitus, and raise awareness about the potential adverse effects of unsafe listening practices.

Fourth, despite participants being somewhat knowledgeable about the risks of hearing damage, the application of this knowledge was limited, indicating a gap between awareness and protective action. This gap underscores the importance of user guides, detailing what can be done at different levels, and adaptive headphone safety modes [40,41] that automatically optimize listening settings for safety when switching from a loudspeaker to headphones, directly addressing the need to bridge knowledge with actionable safe listening practices.

Fifth, the significant interest among gamers in expanding their knowledge about hearing health and safe listening practices supports the creation of a safer listening mode. Besides directly contributing to protecting users' hearing, the very existence of such a mode could educate them on safer listening practices, addressing the community's desire for information and tools to protect their hearing health.

Finally, our findings show that gamers often face barriers to adopting safer listening practices due to concerns about compromising their gaming experience. This suggests a need for customizable features that balance individual preferences with hearing safety. Features like sound category controls [42,43] and dynamic range tests [44], tailored to specific gaming contexts, could help ensure safe listening without detracting from the immersive experience.

Limitations

This study has several limitations to consider when interpreting the results. First, relying on self-reported data may introduce bias and may not accurately reflect actual behaviors. Second, the cross-sectional nature of the study prevents causal inferences. Third, the lack of assessment of noise-canceling headphone use, which could mitigate hearing damage, limits the findings. Finally, the open web-based survey recruitment strategy, shared via social media and stakeholder networks, may have led to selection bias, as those more interested in hearing health or gaming were more likely to participate. Despite this, the consistency of our findings with previous studies on related topics [13,14] supports their validity, allowing us to draw meaningful conclusions. Future research should address these limitations and evaluate the effectiveness of educational interventions in promoting safe listening habits among video gamers and esports players.

Conclusions

This study underscores the urgency of developing interventions to foster safe listening habits in video gamers and esports enthusiasts, pointing out their general lack of awareness and preventive actions against hearing damage. Besides stressing the need for effective communication strategies tailored to this group's specific preferences, the findings support creating targeted interventions to encourage safe listening and minimize hearing loss risk among these audiences. Incorporating safe listening features directly into video games could offer a strategic method to address prolonged exposure to high sound levels without detracting from the gaming experience. These personalized, gamer-friendly solutions could significantly boost auditory health awareness and practices. The insights into gamer behavior and preferences provided by this study not only enrich discussions on digital health interventions but also pave the way for enhancing hearing health in the digital entertainment sector, contributing to the broader goal of leveraging technology to improve health outcomes.

Acknowledgments

The authors would like to express their sincere gratitude to all the participants who took the time to complete the survey, without whom this study would not have been possible. The authors also extend their appreciation to the organizations who helped with the recruitment of study participants. The authors are thankful for their support and cooperation in disseminating information about the study to their members and networks. This work was supported by the World Health Organization, Geneva, Switzerland. The views reported are solely those of the authors and do not necessarily represent the views of the World Health Organization. The authors used ChatGPT by OpenAI to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors contributed to the conceptualization of the study and the investigation. ND and SR developed the study methodology, with ND additionally overseeing recruitment, data collection, and formal analysis. ND also took on project administration and supervision responsibilities. ND and SR wrote the original draft of the manuscript, while all authors participated in reviewing and editing. Funding acquisition was managed by SC.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Survey questionnaire.
[PDF File (Adobe PDF File), 234 KB - formative_v9i1e60476_app1.pdf]

Multimedia Appendix 2
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[PDF File (Adobe PDF File), 52 KB - formative_v9i1e60476_app2.pdf]

References

1. Franks RR, King D, Bodine W, Chisari E, Heller A, Jamal F, et al. AOASM position statement on esports, active video gaming, and the role of the sports medicine physician. Clin J Sport Med 2022;32(3):e221-e229 [FREE Full text] [doi: 10.1097/JSM.0000000000001034] [Medline: 35470342]
2. Global video games market report and forecast 2023-2028. Claight Corp. URL: https://www.expertmarketresearch.com/reports/video-games-market [accessed 2025-03-03]
3. Games - Worldwide | Statista Market Forecast. Statista. URL: https://www.statista.com/outlook/dmo/digital-media/video-games/worldwide [accessed 2025-03-03]
4. Gilbert N. 75 Significant esports statistics: 2024 market, viewership & investment growth. FinancesOnline. 2023. URL: https://financesonline.com/esports-statistics/ [accessed 2025-03-03]
5. Esports market size, share, trends and analysis by region, revenue stream, gaming genre and segment forecast to 2030. Mark Res Rep Consult Glob UK Ltd. URL: https://www.globaldata.com/store/report/esports-market-analysis/ [accessed 2025-03-03]
6. Bossen D, Broekema A, Visser B, Brons A, Timmerman A, van Etten-Jamaludin F, et al. Effectiveness of serious games to increase physical activity in children with a chronic disease: systematic review with meta-analysis. J Med Internet Res 2020;22(4):e14549 [FREE Full text] [doi: 10.2196/14549] [Medline: 32234697]
7. Gentry SV, Gauthier A, L'Estrade Ehrstrom B, Wortley D, Lilienthal A, Tudor Car L, et al. Serious gaming and gamification education in health professions: systematic review. J Med Internet Res 2019;21(3):e12994 [FREE Full text] [doi: 10.2196/12994] [Medline: 30920375]
8. Kim Y, Hong S, Choi M. Effects of serious games on depression in older adults: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022;24(9):e37753 [FREE Full text] [doi: 10.2196/37753] [Medline: 36066964]
9. Kelly S, Leung J. The new frontier of esports and gaming: a scoping meta-review of health impacts and research agenda. Front Sports Act Living 2021;3:640362 [FREE Full text] [doi: 10.3389/fspor.2021.640362] [Medline: 33796849]
10. Chan G, Huo Y, Kelly S, Leung J, Tisdale C, Gullo M. The impact of eSports and online video gaming on lifestyle behaviours in youth: a systematic review. Comput Hum Behav 2022;126:106974. [doi: 10.1016/j.chb.2021.106974]
11. Yin K, Zi Y, Zhuang W, Gao Y, Tong Y, Song L, et al. Linking esports to health risks and benefits: current knowledge and future research needs. J Sport Health Sci 2020;9(6):485-488 [FREE Full text] [doi: 10.1016/j.jshs.2020.04.006] [Medline: 32417192]
12. Hearing loss due to recreational exposure to loud sounds: a review. World Health Organization. 2015. URL: https://iris.who.int/handle/10665/154589 [accessed 2025-02-22]
13. Diviani N, Zanini C, Amann J, Chadha S, Cieza A, Rubinelli S. Awareness, attitudes, and beliefs about music-induced hearing loss: towards the development of a health communication strategy to promote safe listening. Patient Educ Couns 2019;102(8):1506-1512. [doi: 10.1016/j.pec.2019.03.013] [Medline: 30935699]
14. Diviani N, Chadha S, Arunda M, Rubinelli S. Attitudes towards safe listening measures in entertainment venues: results from an international survey among young venue-goers. Int J Environ Res Public Health 2021;18(23):12860 [FREE Full text] [doi: 10.3390/ijerph182312860] [Medline: 34886583]
15. Portnuff CD. Reducing the risk of music-induced hearing loss from overuse of portable listening devices: understanding the problems and establishing strategies for improving awareness in adolescents. Adolesc Health Med Ther 2016;7:27-35 [FREE Full text] [doi: 10.2147/AHMT.S74103] [Medline: 26929674]
16. Jiang W, Zhao F, Guderley N, Manchaiah V. Daily music exposure dose and hearing problems using personal listening devices in adolescents and young adults: a systematic review. Int J Audiol 2016;55(4):197-205. [doi: 10.3109/14992027.2015.1122237] [Medline: 26768911]
17. Guidelines for safe listening devices/systems (Recommendation ITU-T H.870). International Telecommunication Union. 2018. URL: https://www.itu.int/rec/T-REC-H.870-201808-I/en [accessed 2018-12-06]
18. WHO global standard for safe listening venues and events. World Health Organization. 2022. URL: https://www.who.int/publications/i/item/9789240043114 [accessed 2025-02-22]
19. Iannace G, Ciaburro G, Trematerra A. Video games noise exposure in teenagers and young adults. Noise Vib Worldw 2019;51(1-2):3-11. [doi: 10.1177/0957456519889956]
20. Dillard LK, Mulas P, Der C, Fu X, Chadha S. Risk of sound-induced hearing loss from exposure to video gaming or esports: a systematic scoping review. BMC Public Health 2024;2(1):e000253. [doi: 10.1136/bmjph-2023-000253]

XSL•FO
RenderX

21. Müller-Riemenschneider F, Reinhold T, Nocon M, Willich S. Long-term effectiveness of interventions promoting physical activity: a systematic review. Prev Med 2008;47(4):354-368. [doi: 10.1016/j.ypmed.2008.07.006] [Medline: 18675845]
22. Rosenstock IM. The health belief model and preventive health behavior. Health Educ Monogr 1974;2(4):354-386. [doi: 10.1177/109019817400200405]
23. Pauline J. Physical activity behaviors, motivation, and self-efficacy among college students. Coll Stud J 2013;47(1):64-74 [FREE Full text]
24. DiClemente CC, Prochaska JO. Toward a comprehensive, transtheoretical model of change: stages of change and addictive behaviors. In: Treating Addictive Behaviors. 2nd Edition. New York, NY: Plenum Press; 1998:3-24.
25. Portnuff CDF, Fligor BJ, Arehart KH. Teenage use of portable listening devices: a hazard to hearing? J Am Acad Audiol 2011;22(10):663-677. [doi: 10.3766/jaaa.22.10.5] [Medline: 22212766]
26. Rollnick S, Heather N, Gold R, Hall W. Development of a short 'readiness to change' questionnaire for use in brief, opportunistic interventions among excessive drinkers. Br J Addict 1992;87(5):743-754. [doi: 10.1111/j.1360-0443.1992.tb02720.x] [Medline: 1591525]
27. Federal Act on Research Involving Human Beings (Human Research Act, HRA, SR 810.30). URL: https://www.fedlex.admin.ch/eli/cc/2013/617/en [accessed 2024-10-21]
28. Swissethics Decision Tree. URL: https://swissethics.ch/assets/other_study_documents/decision-tree_v2.1_06.02.25_en.pdf [accessed 2025-03-03]
29. Koh YS, Asharani PV, Devi F, Roystonn K, Wang P, Vaingankar JA, et al. A cross-sectional study on the perceived barriers to physical activity and their associations with domain-specific physical activity and sedentary behaviour. BMC Public Health 2022;22(1):1051 [FREE Full text] [doi: 10.1186/s12889-022-13431-2] [Medline: 35614402]
30. Miller-Day M, Hecht ML. Narrative means to preventative ends: a narrative engagement framework for designing prevention interventions. Health Commun 2013;28(7):657-670 [FREE Full text] [doi: 10.1080/10410236.2012.762861] [Medline: 23980613]
31. Greene K, Brinn LS. Messages influencing college women's tanning bed use: statistical versus narrative evidence format and a self-assessment to increase perceived susceptibility. J Health Commun 2003;8(5):443-461. [doi: 10.1080/713852118] [Medline: 14530147]
32. Noar SM, Harrington NG, Aldrich RS. The role of message tailoring in the development of persuasive health communication messages. Ann Int Commun Assoc 2016;33(1):73-133 [FREE Full text] [doi: 10.1080/23808985.2009.11679085]
33. van de Laar KE, van der Bijl JJ. Strategies enhancing self-efficacy. In: Self-Efficacy in Nursing: Research and Measurement Perspectives. New York, NY: Springer Publishing Company; 2002:63.
34. Pietrzak A. Musicians' daily sound exposure assessed by full-day dosimetry. Vib Phys Syst 2019;30(1):1-8 [FREE Full text]
35. Kates JM. Principles of digital dynamic-range compression. Trends Amplif 2005;9(2):45-76 [FREE Full text] [doi: 10.1177/108471380500900202] [Medline: 16012704]
36. Kim H, Ryu D, Park J. Implementation of a volume controller for considering hearing loss in Bluetooth headset. In: Kim T, Adeli H, Robles RJ, Balitanas M, editors. Advanced Communication and Networking. Berlin, Heidelberg: Springer; 2011:430-438.
37. Pretolesi D, Motnikar L, Till B, Uhl J. Notifying users: customisation preferences for notifications in health and well-being applications. In: Meschtscherjakov A, Midden C, Ham J, editors. Persuasive Technology. Cham: Springer Nature Switzerland; 2023:410-424.
38. Horizon Forbidden West: what are tinnitus sounds. Game Rant. URL: https://gamerant.com/horizon-forbidden-west-tinnitus-sounds-option/ [accessed 2024-05-13]
39. Swain SK. Impact of tinnitus on quality of life: a review. Int J Adv Med 2021;8(7):1006. [doi: 10.18203/2349-3933.ijam20212416]
40. How to raise or lower the maximum headphone volume. Nintendo Eur AG. URL: https://www.nintendo.com/en-gb/Support/Nintendo-Switch/How-to-Raise-or-Lower-the-Maximum-Headphone-Volume-1516285.html [accessed 2024-05-13]
41. Adaptive audio for mobile. Game Developer. URL: https://www.gamedeveloper.com/audio/adaptive-audio-for-mobile [accessed 2024-05-13]
42. Marks A. The Complete Guide to Game Audio: For Composers, Musicians, Sound Designers, Game Developers. New York, NY: Routledge; 2012.
43. Audio Settings | Game Dev Handbook. URL: https://gamedev.catalystsoftworks.com/ui/settings/audio/ [accessed 2024-05-13]
44. Audio dynamic range calibration system. Martin Bussy-Pâris. 2016. URL: https://martinbussy.com/audio-dynamic-range-calibration-system/ [accessed 2024-05-13]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
HBM: health belief model
HRA: Human Research Act

XSL•FO
RenderX

LHQ: Listening Habits Questionnaire
OR: odds ratio
TTM: transtheoretical model
WHO: World Health Organization

Edited by A Mavragani; submitted 13.05.24; peer-reviewed by RE Ferdig, K Song; comments to author 06.09.24; revised version received 27.09.24; accepted 28.10.24; published 25.03.25.

Please cite as:
Diviani N, Chadha S, Mulas P, Rubinelli S
Safe Listening Beliefs, Attitudes, and Practices Among Gamers and Esports Participants: International Web-Based Survey
JMIR Form Res 2025;9:e60476
URL: https://formative.jmir.org/2025/1/e60476
doi:10.2196/60476
PMID:40131338

©Nicola Diviani, Shelly Chadha, Peter Mulas, Sara Rubinelli. Originally published in JMIR Formative Research (https://formative.jmir.org), 25.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Perceived Trust and Professional Identity Threat in AI-Based Clinical Decision Support Systems: Scenario-Based Experimental Study on AI Process Design Features

Sophia Ackerhans[1], MSc; Kai Wehkamp[2], Prof Dr Med; Rainer Petzina[2], Dr phil, Prof Dr Med; Daniel Dumitrescu[3], Dr med; Carsten Schultz[1], Prof Dr

[1]Kiel Institute of Responsible Innovation, University of Kiel, Kiel, Germany

[2]Medical School Hamburg, Hamburg, Germany

[3]Clinic for General and Interventional Cardiology/Angiology, Heart and Diabetes Center North Rhine-Westphalia, Ruhr-University Bochum, Medical Faculty OWL (University Bielefeld), Bad Oeynhausen, Germany

Corresponding Author:
Sophia Ackerhans, MSc
Kiel Institute of Responsible Innovation
University of Kiel
Westring 425
Kiel, 24118
Germany
Phone: 49 431880479
Email: ackerhans@bwl.uni-kiel.de

Abstract

Background: Artificial intelligence (AI)–based systems in medicine like clinical decision support systems (CDSSs) have shown promising results in health care, sometimes outperforming human specialists. However, the integration of AI may challenge medical professionals' identities and lead to limited trust in technology, resulting in health care professionals rejecting AI-based systems.

Objective: This study aims to explore the impact of AI process design features on physicians' trust in the AI solution and on perceived threats to their professional identity. These design features involve the explainability of AI-based CDSS decision outcomes, the integration depth of the AI-generated advice into the clinical workflow, and the physician's accountability for the AI system-induced medical decisions.

Methods: We conducted a 3-factorial web-based between-subject scenario-based experiment with 292 medical students in their medical training and experienced physicians across different specialties. The participants were presented with an AI-based CDSS for sepsis prediction and prevention for use in a hospital. Each participant was given a scenario in which the 3 design features of the AI-based CDSS were manipulated in a 2×2×2 factorial design. SPSS PROCESS (IBM Corp) macro was used for hypothesis testing.

Results: The results suggest that the explainability of the AI-based CDSS was positively associated with both trust in the AI system (β=.508; $P<.001$) and professional identity threat perceptions (β=.351; $P=.02$). Trust in the AI system was found to be negatively related to professional identity threat perceptions (β=–.138; $P=.047$), indicating a partially mediated effect on professional identity threat through trust. Deep integration of AI-generated advice into the clinical workflow was positively associated with trust in the system (β=.262; $P=.009$). The accountability of the AI-based decisions, that is, the system required a signature, was found to be positively associated with professional identity threat perceptions among the respondents (β=.339; $P=.004$).

Conclusions: Our research highlights the role of process design features of AI systems used in medicine in shaping professional identity perceptions, mediated through increased trust in AI. An explainable AI-based CDSS and an AI-generated system advice, which is deeply integrated into the clinical workflow, reinforce trust, thereby mitigating perceived professional identity threats. However, explainable AI and individual accountability of the system directly exacerbate threat perceptions. Our findings illustrate the complex nature of the behavioral patterns of AI in health care and have broader implications for supporting the implementation of AI-based CDSSs in a context where AI systems may impact professional identity.

(JMIR Form Res 2025;9:e64266) doi:10.2196/64266

XSL•FO
RenderX

KEYWORDS

artificial intelligence; clinical decision support systems; explainable artificial intelligence; professional identity threat; health care; physicians; perceptions; professional identity

Introduction

Background

Research in the field of health care innovation has begun to examine the impact of artificial intelligence (AI) on human-cognitive structures and processes of medical decision-making and problem-solving [1,2]. AI enables machines to process and analyze vast datasets independently [3], improving clinical outcomes in areas such as sepsis detection and error reduction [4,5]. AI-based clinical decision support systems (CDSSs) can enhance patient care while reducing costs and errors [6]. However, successful implementation requires acceptance by health care professionals, which is often hindered by skepticism, distrust, and perceived threats to professional identity [7,8].

Process design features, such as explainability and reliability, play a crucial role in shaping trust in AI-based CDSSs and may influence physicians' professional identity [8-11]. While prior research has explored the relationship between trust and technology acceptance [12,13], the impact of these design features on professional identity threats remains underexplored. This study investigates how process design features influence physicians' trust in AI-based CDSSs and how this trust relates to professional identity threats, providing insights to support AI implementation in health care.

Conceptual Framework

Early research on technology adoption in organizations primarily focused on how users perceive and respond to technical aspects. For instance, the Technology Acceptance Model [14] delineates the key determinants shaping users' attitudes and intentions toward adopting new workplace technology. Recent studies have scrutinized technology adoption from the perspective of trust formation [12,15,16], revealing that a lack of trust in technology can lead to its misuse and nonadoption of technology [15]. For example, initial trust in technology positively mediates the relationship between technology-related factors (such as performance expectancy, effort expectancy, and task complexity) and behavioral intentions [16,17]. Trust denotes the degree to which an individual is willing to believe in or be vulnerable to the actions of another party [18]. This is particularly relevant in situations characterized by risk or uncertainty, where individual behavior can lead to loss or harm [19]. Importantly, this description extends trust beyond human interaction to encompass technology, including AI [20]. This study focuses on the user's trust in the AI-based technology, rather than interpersonal trust (trust between individuals) or social trust (trust in an institution) [21]. Trust is a crucial prerequisite for physicians in adopting AI [22], as AI is perceived as risky due to the complexity and unpredictability of its behavior [11]. Thus, the formation of trust in an AI-based system among physicians is influenced by the AI system's representation and tangibility, that is how the underlying rationale of AI tools' decision outcomes are presented to the user [12]. Consequently, the representation of AI functionalities is expected to play a pivotal role in fostering trust in AI among physicians.

IT systems that disrupt traditional work routines often meet with particular resistance. This defensiveness originates, among other things, from the potential threat to autonomous work processes and traditional routines [23]. Studies have emphasized the relevance of a perceived threat or enhancement posed by new technologies to professional identities, underscoring the importance of sociocognitive elements related to human expertise, including hierarchy, status, autonomy, control, and legitimacy, particularly in knowledge-intensive organizations [23-26].

Professional identity pertains to how individuals define and distinguish themselves in their professional roles [27]. Medical professionals exhibit a particularly strong attachment to their social group and professional identity, which evolves through rigorous socialization during education and the acquisition of extensive expertise through practical experience [28,29]. Health care professionals often have the autonomy to make treatment decisions, intensifying their motivation and dedication and reinforcing their identification with the medical field [30]. Disruptions to existing workflows will be perceived as threats to health professionals' identity if they restrict autonomy in terms of decision-making and sovereignty over medical knowledge [8].

A recent literature review on the implementation challenges of CDSSs showed that a CDSS transparency in elucidating decision outcomes (the "explainable" AI), integration depth of the AI-generated system advice into the clinical workflow, and security- and privacy-related mechanisms such as system-induced individual accountability frame professional identity threats and enhancements among physicians [10]. These 3 process design features of AI-based systems may have positive and negative behavioral consequences.

The AI's ability to analyze vast amounts of data and transparently present the outcome of the analysis can complement physicians' expertise and strengthen their professional identity, by facilitating more informed decision-making and affirming physicians in their perceived efficiency and effectiveness. However, as the medical environment in which AI is used is usually complex and the behavior of AI is not deterministic, the opaque, multilayered process of AI decision-making is difficult to predict, and the underlying logic of the decision may be poorly understood [11,30]. The risk of overreliance on AI may lead to a perceived erosion of traditional medical expertise. Physicians may associate AI recommendations with a devaluation of their professional judgment [7].

AI-based systems that integrate seamlessly into clinical workflows can enhance the efficiency and productivity of clinicians, allowing them to focus more on patient care. This

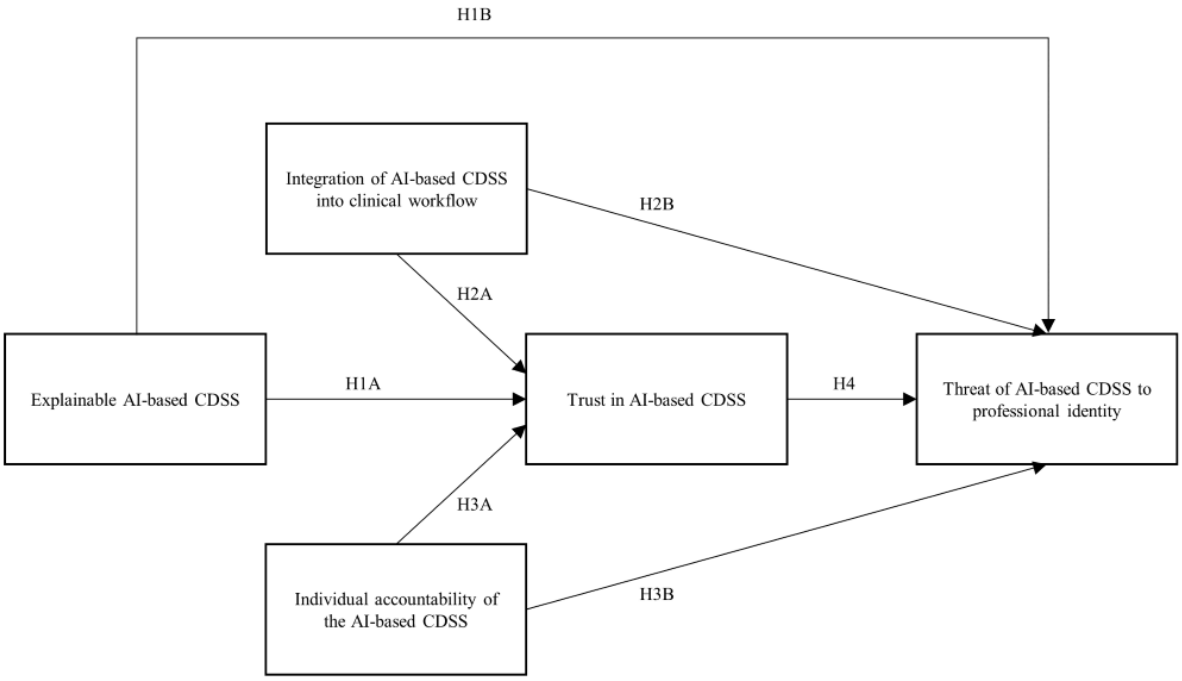
XSL•FO
RenderX

can strengthen their trust in the IT system [31]. However, AI systems that provide detailed behavioral advice, which closely resemble physicians' established decision-making parameters, can disrupt the intrinsic rhythm of medical practice. Such disturbances may foster a perception of diminished autonomy among physicians that could potentially compromise the physicians' professional identity [10].

In addition, the use of digital signatures and the monitoring of user credentials play a critical role in safeguarding patient information and maintaining trust in the health care system [10,32]. The requirement for user signatures on treatment decisions ensures that actions affected by the AI-based CDSS are traceable to specific, authorized users, validating the technologically derived information, and reinforcing the confidentiality and integrity of patient data. The tracking of user credentials and actions within AI-based CDSS underscores the responsibility of physicians in making informed, ethical treatment decisions, which reinforces their professional identity as accountable caretakers of patient health [32]. However, these systems may on the other hand negatively influence the professional identity and role of physicians by embedding accountability and increasing monitoring of diagnostic and therapeutic decisions by clinical managers and external agents such as government bodies and health insurance [8]. Electronic monitoring facilitates the regulation and control of both processes and outcomes. This technological oversight grants clinical managers and regulators the capability to impose a form of direct governance over health care professionals. It effectively binds physicians to a self-regulatory framework, which is less predicated on the active enforcement of control than on the perceived omnipresence of surveillance [33]. Furthermore, detailed tracking of decisions and actions can be used in legal situations, increasing the liability risks for medical professionals. This might lead to more defensive behavior, where professionals take extra precautions, not out of clinical necessity, but to protect against potential legal repercussions [7].

In the subsequent section, we formulate hypotheses regarding the role of trust in AI-based CDSSs and the potential threat or enhancement to the professional identity of health care professionals stemming from the use of AI-based CDSSs. In this context, we investigate the impact of AI explainability, depth of integration of the AI-generated system advice into the clinical workflow, and system-induced individual accountability. Figure 1 provides a visual representation of our proposed research model.

Figure 1. The proposed research framework for analyzing the impact of AI process design features on perceived professional identity threat. AI: artificial intelligence; CDSS: clinical decision support system; H1A: hypothesis 1A; H1B: hypothesis 1B; H2A: hypothesis 2A; H2B: hypothesis 2B; H3A: hypothesis 3A; H3B: hypothesis 3B; H4, hypothesis 4.



Hypothesis 1A: The Explainability of an AI-Based CDSS Has a Positive Effect on Trust for Health Care Professionals in the Decision Outcomes of Such Systems

Explainable AI represents a crucial component in building trust in AI-based technologies by ensuring transparency in the system's operational principles [12,34]. Such a system empowers users to understand the system's internal operations and decision rationale. For instance, it can illustrate how inputs are mathematically transformed into outputs [35]. Various levels of explainability can be implemented, including the presentation of varying weightings of data leading to AI decisions [36]. Numerous studies highlight the importance of explainable models in health care, enabling health care professionals to trust AI-generated outcomes [37,38]. Consequently, we propose that the transparent display of the data used and the mathematical basis of decision outcomes in an explainable AI-based CDSS fosters higher levels of trust.

XSL•FO
RenderX

Hypothesis 1B: The Explainability of an AI-Based CDSS Has a Positive Effect on Physicians' Professional Identity Threat

Nevertheless, explainable AI-based CDSSs may, in contrast, pose a potential threat to medical authority and the role of physicians [7,8]. As the potential of the AI system becomes transparent to health care professionals, they are more likely to believe in a shift in hierarchy and an erosion of physicians' autonomy and status as a result of AI systems [8,26,39]. In scenarios where AI partially supersedes a physician's expertise, particularly in high-risk situations, it exacerbates their sense of loss of control and status [7], which can induce stress as their professional decision-making control and autonomy may be undermined [40]. An explainable AI-based CDSS may as such be perceived as a threat to their established roles and professional standing.

Hypothesis 2A: An AI-Based CDSS That Is Deeply Integrated Into the Clinical Workflow Has a Positive Effect on Physicians' Trust in the System

Evaluating the AI-based CDSSs' fit into the clinical workflow is crucial because users' trust in AI varies with the expected efficacy of the technical system [41]. For instance, physicians often disregard AI recommendations that are considered unnecessary, potentially missing out on relevant decision-enhancing benefits [7,42]. Detailed clinical treatment recommendations, especially for high-risk events or complex tasks, can reinforce confidence in the system by reducing uncertainty through the clinical information provision [30,34,43]. Consequently, the degree of trust in the technology's effectiveness is a result of users' perception of its ability to offer sufficient, efficient, and prompt assistance.

Hypothesis 2B: An AI-Based CDSS That Is Deeply Integrated Into the Clinical Workflow Has a Positive Effect on Physicians' Perceived Professional Identity Threat

Physicians' strong identification with their profession [28] may make them skeptical of AI-driven system advice, especially in situations where the AI task undermines their own clinical workflow and tasks [30,44]. Physicians may fear losing control over medical decisions if they are compelled to follow the system's advice [7]. Integrated AI-based systems with automated advice, which closely mirror established human tasks, may disrupt current established processes. Moreover, explicit advice on expected behavior may diminish professional authority and autonomy, eroding individual control. When systems control actions, the scope for human professional judgment is limited, impacting the ability to provide proficient recommendations. This situation can lead to a sense of redundancy [44].

Hypothesis 3A: An AI-Based CDSS That Induces Individual Accountability Has a Negative Effect on Physicians' Trust in the AI-Based CDSS

A mandatory use of AI-generated feedback, aimed at enhancing human-AI collaboration, can inadvertently bring attention to managerial monitoring and tracking, potentially leading to concerns about excessive control and distrust. For instance, demanding health care professionals to approve AI system decisions can be seen as invasive and indicative of a lack of managerial discretion, which may undermine both trust in the system and compliance [45]. Similar issues have arisen in contexts like Uber (Uber Technologies Inc), where continuous tracking eroded drivers' autonomy and trust [46]. Thomas et al [47] investigated barriers to adopting e-prescribing and found that requiring health care professional signatures on prescriptions initially caused frustration and a preference for paper-based prescriptions. However, over time, the use of signatures became less of a barrier. Still, constant signature requirements can increase ambiguity about treatment responsibility, leading to reduced trust in the system. Physician signatures for validating the AI system decisions make them cocreators of the AI-based advice, increasing risk awareness and responsibility for the IT system's actions, and as such decreasing system trust.

Hypothesis 3B: An AI-Based CDSS That Induces Individual Accountability Has a Positive Effect on the Physicians' Perceived Threat to Professional Identity

Active endorsement of AI-based CDSS decisions can lead to an increased perceived reliance of health care professionals on the system. This perceived dependency diminishes their sense of autonomous decision-making and intensifies the pressure on physicians to integrate AI results into their professional judgment [30]. This pressure to endorse and validate AI-based decisions raises questions about the distribution of responsibility and accountability, reducing an individual's ability to influence a decision [48]. This kind of AI can infringe physicians' autonomy and control as they have to justify AI-based CDSS decisions, despite a lack of knowledge about the data quality and accuracy of the system [44]. The traceability provided by signatures raises concerns about the benefits and risks of AI systems, potentially jeopardizing a physician's reputation as a medical services provider [49].

Hypothesis 4: Trust in an AI-Based CDSS Mediates the Relationship Between an Explainable AI-Based CDSS and Physicians' Perception of Threat to Their Professional Identity

Finally, we anticipate a mediated relationship between the analyzed process design features of AI systems and their influence on the perception of threats to professional identity via trust. Empirical research has nuanced human trust in AI into the dimensions of cognitive and affective trust [50]. Cognitive trust in AI, rooted in rational evaluation and reliability, and affective trust, driven by emotional and interpersonal connections, both emerge as key facets that influence the acceptance and use of AI [11]. For cognitive trust, AI reliability and explainability have been identified as pivotal, as the degree to which the functionality and decision-making processes of the AI are understandable to users is crucial for building trust. Affective trust, on the other hand, is based on affective bonds and feelings toward AI technology rather than rational evaluations of its capabilities. The degree to which AI demonstrates behaviors that promote perceived closeness to actual medical tasks can foster emotional bonds between users and AI, enhancing trust on an emotional level [11]. Given the relationship between the use of an AI-based CDSS and threat

XSL•FO
RenderX

to professional identity, it is reasonable to assume that physicians' cognitive and affective trust in explainable AI-based CDSSs can reduce their perception of threat to their professional identity.

Methods

Study Design

Scenario-based experimental surveys present a valuable approach to quantitative research by integrating the strengths of traditional surveys and experimental designs. This method enhances external validity through the representativeness and multivariate measurement capabilities of surveys, while simultaneously improving internal validity through controlled, experimental interventions. By addressing the limitations of both approaches—namely, the multicollinearity and passive measurement inherent in surveys, and the lack of representativeness and simplified settings typical of experiments—scenario-based studies offer a comprehensive and rigorous method for investigating respondents' beliefs, attitudes, and judgments [51].

Sample and Data Collection

The study was conducted with German medical students, physicians, and nursing professionals. The study sample was composed primarily of German medical students in the second (clinical) part of the course study program, as they represent future users of AI-based risk prediction systems and are at a critical stage of developing their professional identity. During their education, medical students gain professional experience by interacting directly with patients under supervision, allowing them to apply theoretical knowledge to real-world clinical scenarios. This hands-on training is essential for developing diagnostic skills, clinical decision-making, and effective patient communication. Medical students undergo continuous practical training in university clinics, teaching hospitals, or other medical centers. In addition, they are required to complete a 3-month internship in a hospital before starting their medical study program or during nonteaching periods of the preclinical stage, which familiarizes them with hospital operations and clinical procedures. Furthermore, a 4-month clinical traineeship is mandatory during the transition between the preclinical and clinical stages of medical education. These structured experiences progressively immerse students in hospital environments, allowing them to interact with diverse physicians and develop their professional identities. Research supports the idea that medical students' professional identity evolves rapidly during their education. Niemi [52] notes that critical moments for identity development occur at the start of professional training, as students begin to conceptualize themselves as future physicians. Similarly, Pitkala and Mantyranta [53] found that while medical students initially lacked confidence, their interactions with patients during clinical placements increased their sense of credibility and comfort by the end of their first year. Madill and Latchford [54] further observed that students' identity progression during their first year of training was marked by growing competence and dedication. Advanced medical students, therefore, are well-positioned to recognize and articulate perceptions of professional identity threats.

For the study, 8 different scenarios of a fictitious AI-based CDSS were developed, which showed the user the risk of developing sepsis within the next 48 hours. Sepsis is a complex infectious disease that can develop from any focus of infection. It can be caused by viruses, bacteria, fungi, or parasites and can affect patients of all ages [55]. According to an analysis of Germany-wide Diagnosis Related Groups statistics, there was an incidence of 158 patients with sepsis per 100,000 inhabitants in Germany in 2015. The proportion of sepsis patients among all hospital patients was 0.7%. 53.8% of patients with hospital-acquired sepsis were treated in intensive care units and 41.7% died in hospital [56]. In addition to the lack of improvement in treatment standards and measures to reduce hospital-acquired infections, this high mortality rate is attributed to a lack of quality initiatives for early detection [57], which illustrates the high relevance of AI-supported CDSSs for sepsis risk prediction.

Five physicians from a university hospital in northern Germany evaluated the scenario-based web-based survey during a pretest to ensure that all 3 factors studied were successfully manipulated and that participants could distinguish between them. Participants were familiar with the topic of sepsis, as this condition is considered one of the most common preventable causes of death and requires timely antibiotic therapy, which largely depends on the physician's judgment if an AI-based CDSS is not used. Manipulation checks were conducted to ensure that participants understood the manipulation of the scenario. It was also checked whether the respondents answered the survey seriously (seriousness check).

Participants were recruited via paper and digital flyers distributed in senior medical lectures at medical faculties in northern Germany. In addition, participants were recruited via social media and newsletters from medical faculties at German universities. Participation in the study was voluntary, and participants could access the web-based survey via a link and QR code provided on the recruitment flyers. Before the participants could access the questionnaire, they had to give their informed consent to participate. Participants in the study were able to view information about the duration of the survey and data collection (location and duration of data storage, investigator, and purpose of the study) via a link to a pop-up on the first page of the survey before giving their consent. The data was collected between the end of 2022 and summer 2023. Of the 673 people who clicked on the QR code or link for the survey, 333 (49.48%) participants accessed the survey. We received 292 completed responses, which we used for analysis.

Design and Manipulations

In a 2×2×2 factorial design, 3 factors were manipulated: AI-based CDSS decision explainability (explained vs not explained), integration depth of AI-based CDSS advice into the clinical workflow (detailed vs no detailed clinical treatment advice), and system-induced accountability of a physician in relation to the AI-based CDSS advice (system required a signature vs no signature). These manipulations were designed to produce 8 different treatments (Table 1). All participants were given a scenario of AI Med Predict, a fictitious AI-based CDSS. We modified the wording of the scenarios while keeping

XSL•FO
RenderX

the content consistent to exclude other interpretations. The participants were randomly divided into 8 groups, each exposed to one of the treatments (between-subject design). A minimum size of 23 in each group was calculated with an effect size of 0.85, an α of .05, and a statistical power of 0.8 using the statistical software G*Power. Thus, the cell sizes for the 8 treatments were large enough for this exploratory study (n_1=29; n_2=44; n_3=41; n_4=45; n_5=31; n_6=36; n_7=42; and n_8=24) [58,59].

We developed a hypothetical AI-based CDSS tool for sepsis risk prediction to operationalize AI-based CDSS decision explainability, depth of integration of AI-based CDSS advice into the clinical workflow, and system-induced individual accountability. The simulated CDSS, set in a fictional ward view, displayed the patient's name, age, case number, primary diagnosis, room and bed, progress remark, and estimated sepsis risk score in percent. Subsequently, the respondent has seen a scenario involving a hypothetical patient with a high-risk prognosis for sepsis within 48 hours.

The explainability of the AI-based CDSS decision was altered through 2 treatments. If the user received a detailed explanation of the AI-based CDSS decision that included the patient data used in the risk analysis calculation, such as preexisting conditions, recent procedures, vital signs, laboratory values, or age, we coded the variable as 1. If the user did not receive an explanation of the estimated sepsis risk score, we coded the variable as 0. The integration depth of AI-generated advice into the clinical workflow was coded as 1 if the system specified preventive treatment procedures, such as catheter change and microbiological investigations. The respondent was also provided with a link to the German sepsis prevention, diagnostic, treatment, and aftercare decision guidelines [60]. If the system made an unspecific call for preventive activities, the variable was coded as 0. System-induced individual accountability was coded as 1 if the system hypothetically sought a signature from the respondent to acknowledge the risk prediction and to confirm the planned preventative actions. If the system instructed the user to examine another subject without requiring a signature, the variable was coded as 0.

To ensure treatment efficacy, we used manipulation checks. Specifically, the respondents were asked to evaluate the explainability of the AI-based CDSS decision (eg, whether AI Med Predict for risk prediction of sepsis shows them the composition or explanation of the risk value), the depth of integration of AI-based CDSS advice into the clinical workflow (eg, whether AI Med Predict for risk prediction of sepsis gives them an exact recommended course of action including a link to treatment guideline) regarding the measures they should initiate for the patient), and system-induced individual accountability for the decision outcome of the AI-based CDSS (eg, whether AI Med Predict for risk prediction of sepsis finally prompts them for their signature). Only participants were included in the dataset, which have successfully passed the manipulation check. On average, the survey completion time was 15 minutes 16 seconds.

Table 1. Structure of the 2×2×2 scenario-based experiment design with 8 scenarios.

Scenario	Explainability of AI[a]-based CDSS[b] decision		Integration depth of AI-based CDSS advice into the clinical workflow		Individual accountability induced by AI-based CDSS	
	Explained	Not explained	Detailed treatment advice	No detailed treatment advice	Signature required	No signature required
1		✓		✓		✓
2	✓			✓		✓
3		✓	✓			✓
4	✓		✓			✓
5		✓		✓	✓	
6	✓			✓	✓	
7		✓	✓		✓	
8	✓		✓		✓	

[a]AI: artificial intelligence.

[b]CDSS: clinical decision support system.

Measurement Scales

All question items were measured using a 5-point Likert scale from "strongly disagree" to "strongly agree" (Multimedia Appendix 1 lists the measurement items in detail). Professional identity threat was measured by a 6-item scale adopted from Walter and Lopez [8]. Sample items include "Using AI Med Predict to predict risk of sepsis reduces my control over clinical decisions." Trust in AI-based CDSSs was measured by adapting the 8-item scale developed by Hoffman et al [34]. An example item is "I have confidence in the AI Med Predict system for risk prediction of sepsis. I have the feeling that it works well." Because the 2 scales represent relatively recent additions and include self-developed items, we conducted an exploratory factor analysis (EFA) to assess their construct validity [61]. For the scales we used, a principal component analysis with varimax rotation was performed. The EFA enabled all items to load on any factor, with the number of factors being defined by an eigenvalue larger than 1.00 [62]. As reported in Multimedia Appendix 2, this analysis yielded 2 factors. The first is composed of items 1, 2, 3, and 5 of the professional identity threat scale adapted from [8]. The second factor includes the

XSL•FO
RenderX

first 4 items of the AI trust scale proposed by [34]. Any remaining items that failed to load on their respective scales were subsequently eliminated. Despite this, the scales appear cohesive enough for analysis. The scales for professional identity threat and trust in technology demonstrated reliability, with Cronbach α values of 0.83 and 0.81, respectively [63].

Personal innovativeness with technology was examined as a covariate using a scale, titled the same as Agarwal and Prasad [64], which measures the respondent's willingness to explore new information technology. EFA was used to evaluate the personal innovativeness scale's component structure [61]. We found a single-factor solution for personal innovativeness with a technology scale and a satisfactory Cronbach α (0.86), indicating excellent construct validity and reliability of these measures [63,64].

Ethical Considerations

Ethical approval was granted by the Central Ethics Committee of Kiel University (ZEK-19/22). Informed consent was obtained from all participants prior to participation in the scenario-based experiment. The confidentiality of the participants was maintained at all times. The participants received no compensation for their participation. In return for participating in the survey, we promised an earmarked donation per completed response to a medical charity in the survey.

Results

Descriptive Statistics

A total of 333 responses were collected. Respondents who indicated that their education or employment was not related to the health care field were excluded, as their experiences may not be relevant to the research question (n=16, 4.8%). Manipulation checks were included to verify whether participants properly understood and engaged in experimental conditions. Participants who failed these checks were excluded to maintain data integrity (n=16, 4.8%). In addition, seriousness checks were used to ensure participants provided thoughtful responses. Those who failed these checks, typically by answering in a way that indicated random or inattentive participation, were excluded (n=9, 2.7%). This resulted in a final sample of 292 (87.7%) participants for data analysis. 278 (95.2%) were medical students, 8 (2.7%) participants were physicians across different specialties, and 6 (2.1%) were nursing professionals (eg, intensive care nurses). The participants' average age was assessed using a categorical scale, where 1=18 years or younger and 7=60 years or older. The mean score for age was 2.53 (SD 0.75), indicating that the typical participant in the sample was approximately 24-25 years of age, corresponding to the mid-20s age range. 70% of the participants were female (Table 2). On average, the medical students attended their seventh semester and 74.1% had 3 or more years of bedside experience. 75% of the physicians and 83.3% of the nursing professionals were younger than 30 years and were still active in medical specialist training. 57.2% of the respondents stated that they lived in Schleswig-Holstein or Hamburg, 9.6% in Lower Saxony, 9.3% in Baden-Wuerttemberg, and 8.9% in Bavaria. The complete demographic characteristics of the sample are present in Multimedia Appendix 3.

The mean values, SDs, and Pearson correlations for the study variables are displayed in Table 2. The mediator and dependent variables are weakly correlated (trust and professional identity threat: r=–0.119; P=.042). Explainability of the AI-based CDSS' decision outcome and trust in an AI-based CDSS are moderately correlated (r=0.367; P<.001). AI-based system advice (r=–0.115; P=.049) and accountability of the AI-based CDSS via a system-required signature (r=0.122; P=.037) are weakly correlated with professional identity threat.

XSL•FO
RenderX

Table 2. Mean values, SDs, and correlation matrix (N=292).

	Mean	SD	1	2	3	4	5	6	7	8
Explainable AI[a]-based CDSS[b]	0.49	0.501	1.00	—[c]	—	—	—	—	—	—
AI-based CDSS system advice	0.48	0.500	–0.117[d]	1.00	—	—	—	—	—	—
AI-based CDSS tracking	0.54	0.499	–0.108	–0.045	1.00	—	—	—	—	—
Trust	3.278	0.655	0.367[e]	0.114	–0.022	1.00	—	—	—	—
Professional identity threat	1.907	0.708	0.034	–0.115[f]	0.122[g]	–0.119[h]	1.00	—	—	—
Personal innovativeness	3.463	0.799	0.069	–0.016	–0.005	0.074	–0.103	1.00	—	—
Age (1-7 scale)[i]	2.50	0.780	–0.009	0.137[j]	–0.036	–0.104	–0.027	–0.077	1.00	—
Sex (1: female)	0.70	0.58	0.039	–0.064	0.006	0.133[k]	–0.030	–0.300[l]	–0.089	1.00

[a] AI: artificial intelligence.

[b]CDSS: clinical decision support system.

[c]Not applicable.

[d]P=.045.

[e]P<.001.

[f]P=.049.

[g]P=.04.

[h]P=.042.

[i]Participants reported their age using a categorical scale: 1=18 years or younger, 2=19-24 years, 3=25-29 years, 4=30-39 years, 5=40-49 years, 6=50-59 years, and 7=older than 60 years.

[j]P=.02.

[k]P=.02.

[l]P<.001.

Hypothesis Testing

First, we conducted a 1-way ANOVA to assess the effects of the 8 differently manipulated scenarios on trust in an AI-based CDSS and the threat of AI-based CDSS to professional identity. The level of trust differed statistically significant for the different scenarios, ($F_{7,284}$=2.596; P<.01). The level of perceived professional identity threat differed statistically significant on a 10% level for the different scenarios ($F_{7,284}$=1.818; P<.08). Figure 2 and Figure 3 show the mean values of the 2 variables across the 8 scenarios, along with pairwise comparisons using a post hoc least significant difference test to control for type I error.

The 8 scenarios given in Figures 2 and 3 are as follows. In scenario 1, the AI-based CDSS decision is not explained, the treatment advice is not deeply integrated into the clinical workflow, and there is no individual accountability induced by the system via signature. In scenario 2, the AI-based CDSS decision is explainable; however, the treatment advice remains not deeply integrated into the clinical workflow, and there is no individual accountability induced by the system via signature. Scenario 3 features an AI-based CDSS decision that is not explained, but the treatment advice is deeply integrated into the clinical workflow, while individual accountability is not induced by the system via signature. In scenario 4, the AI-based CDSS decision is explainable, the treatment advice is deeply integrated into the clinical workflow, and no individual accountability is induced by the system via signature. Scenario 5 involves a not explained AI-based CDSS decision with treatment advice that is not deeply integrated into the clinical workflow; however, individual accountability is induced by the system via signature. In scenario 6, the decision is explainable, the treatment advice is not deeply integrated into the clinical workflow, and individual accountability is induced by the system via signature. Scenario 7 presents a not explained AI-based CDSS decision combined with deeply integrated treatment advice in the clinical workflow, and individual accountability is induced by the system via signature. Finally, scenario 8 includes an explainable AI-based CDSS decision, deeply integrated treatment advice within the clinical workflow, and individual accountability is induced by the system via signature. In all scenarios, the error bars represent an SE of 1.

XSL•FO
RenderX

Figure 2. Mean values of variable trust in AI-based CDSS, with pairwise comparisons (post hoc results). AI: artificial intelligence; CDSS: clinical decision support system. *P<.05, **P<.01, ***P<.001.

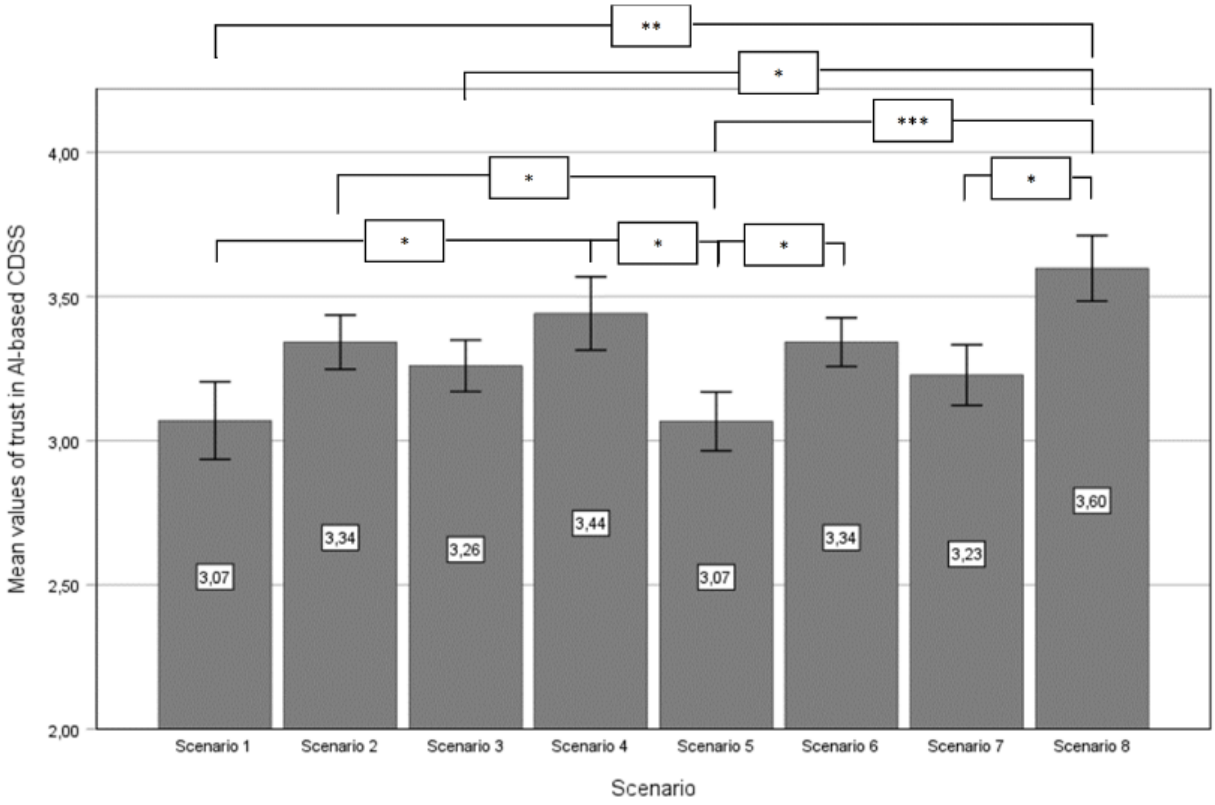
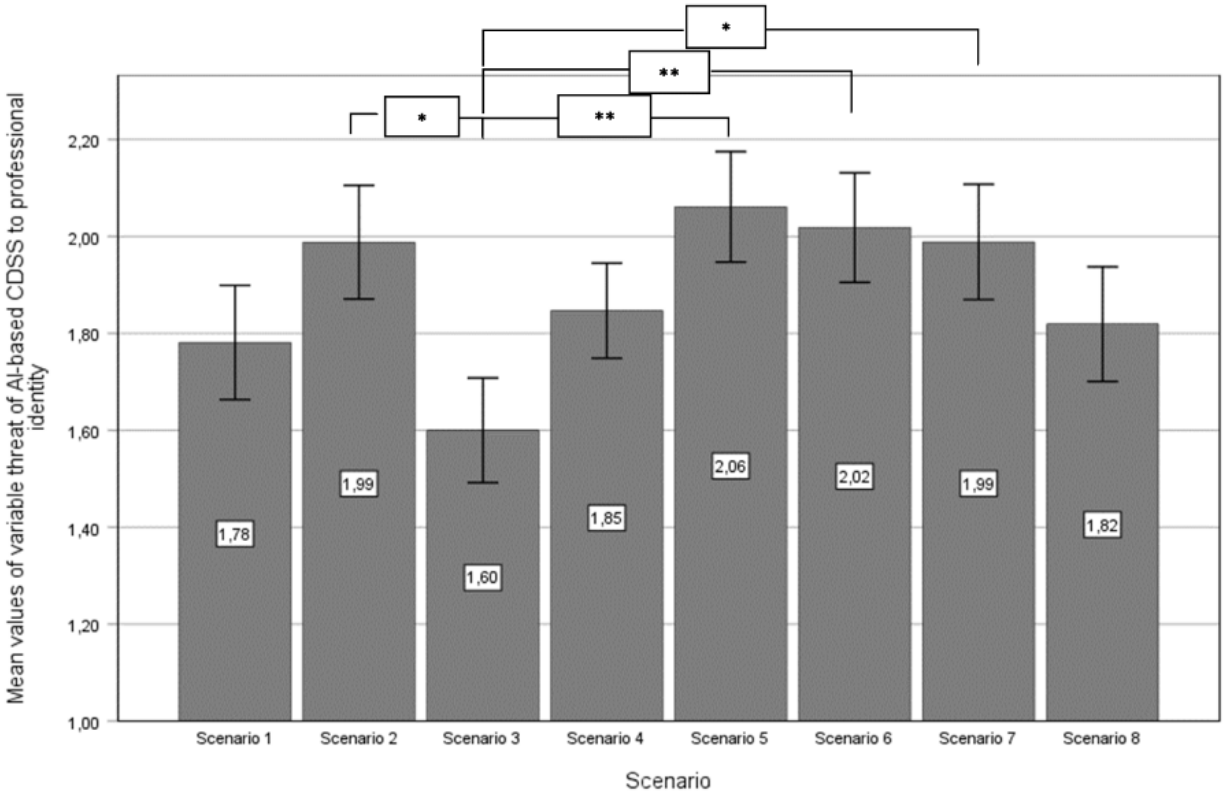


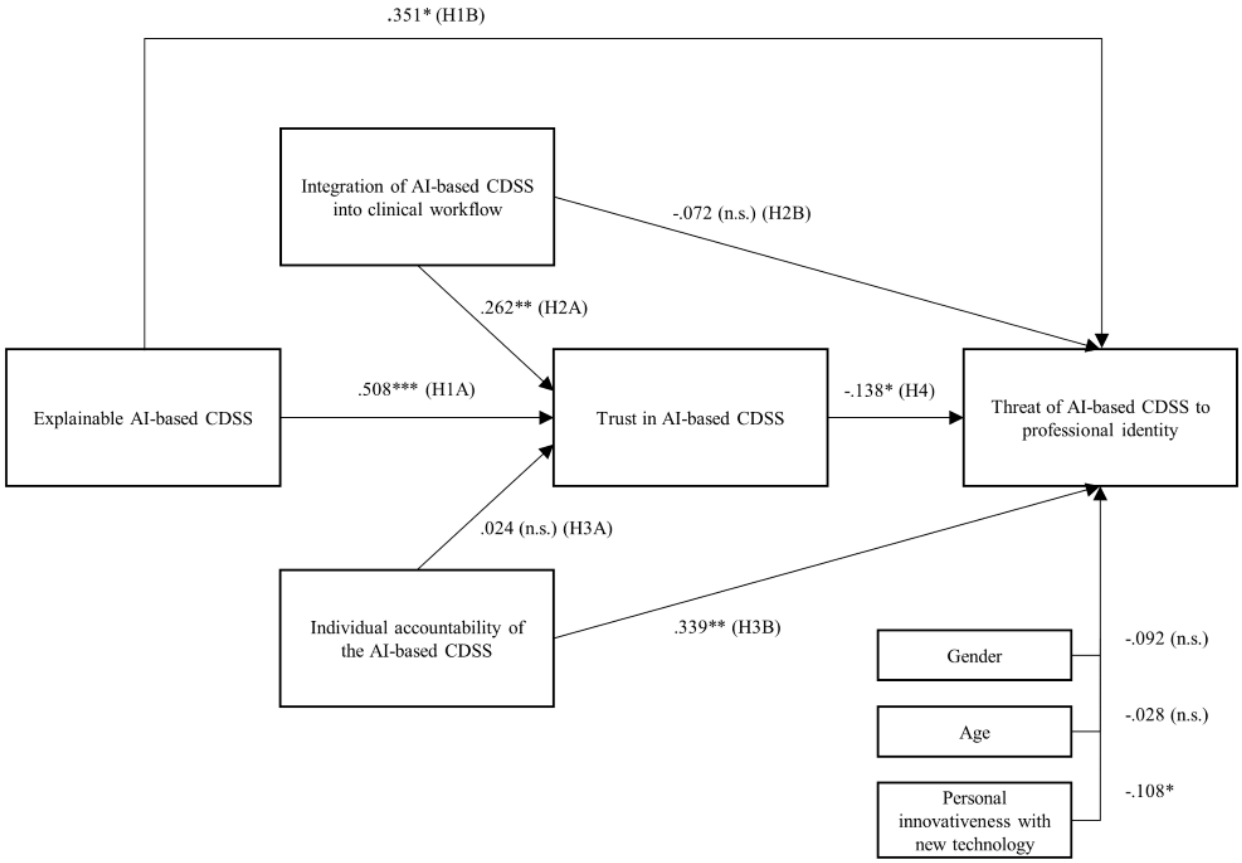
Figure 3. Mean values of variable threat of AI-based CDSS to professional identity, with pairwise comparisons (post hoc results). AI: artificial intelligence; CDSS: clinical decision support system. *P<.05, **P<.01.



XSL•FO
RenderX

Further, we performed SPSS PROCESS (IBM Corp) macro model 10 to test the proposed hypotheses (Figure 4) [65]. PROCESS is an observed-variable modeling tool that relies on ordinary least squares regression, mediation, or moderation analysis [66]. We carried out bootstrapping to test the indirect effects with a sample size of 5000 and a 95% CI [67]. As suggested by Preacher and Hayes [67], we examined the significance of the effects by checking whether the CIs included 0. As control variables, we considered participants' age, sex, and personal innovativeness in relation to new technologies. The model explains 7.2% of the variance of the dependent variable, professional identity threat (adjusted R^2=0.072; $F_{9,282}$=2.439; P=.01). We found a significant positive association between AI-based CDSS explainability and trust in such a system (β=.508; P<.001). This finding provides support for hypothesis 1A. Furthermore, it was shown that the level of trust in an AI-based CDSS has a negative relationship with a perceived threat to professional identity (β=–.138; P=.047). The results further demonstrated that the indirect effect of the explainability of an AI-based CDSS on professional identity threat via trust in an AI-based CDSS was significant in all treatment manipulations. While holding the other 2 process design features constant, the indirect effect of the explainability of AI via trust was a×b=–.073 (95% CI –0.145 to –0.009). In addition, we found a significant positive direct association between the explainability of an AI-based CDSS and professional identity threat (β=.351; P=.02), providing support for hypothesis 1B. Thus, the impact of explainable AI-driven CDSS on the perception of professional identity threat is found to be partially mediated by the level of trust placed in AI, providing support for hypothesis 4. We found a significant positive association between an AI-based CDSS that is strongly integrated into the clinical workflow and trust in such a system (β=.262; P=.009), providing support for hypothesis 2A. Further, the direct effect of an AI-based CDSS inducing individual accountability on the perceived threat to professional identity was significant (β=.339; P=.004), which supports hypothesis 3B. Finally, sex (1: female) significantly influences trust in an AI-based CDSS (β=.208; P=.01) while personal innovativeness significantly influences the professional identity threat posed by an AI-based CDSS (β=–.108; P=.046; Figure 4). At the same time, the direct influence of a strongly integrated AI-based CDSS into the clinical workflow on a perceived threat to professional identity is not significant. Similarly, a system inducing a physician's accountability does not significantly affect trust in an AI-based CDSS (Table 3).

Figure 4. Regression results (N=292). AI: artificial intelligence; CDSS: clinical decision support system; H1A: hypothesis 1A; H1B: hypothesis 1B; H2A: hypothesis 2A; H2B: hypothesis 2B; H3A: hypothesis 3A; H3B: hypothesis 3B; H4, hypothesis 4.



XSL•FO
RenderX

Table 3. Path analysis results with coefficients, *P* values, and 95% CIs (N=292).

Hypothesis and path	Coefficients (β)	*P* value	95% CI
H1A[a]: AI[b]-based CDSS[c] explainability → trust in the AI-based CDSS	.508	<.001	0.258 to 0.758
H1B[d]: AI-based CDSS explainability → perceived threat of AI-based CDSS to professional identity	.351	.02	0.052 to 0.651
H2A[e]: Integration of AI-based CDSS advice into the clinical workflow → trust in the AI-based CDSS	.262	.009	0.067 to 0456
H2B[f]: Integration of AI-based CDSS advice into the clinical workflow → perceived threat of AI-based CDSS to professional identity	–.072	.538	–0.301 to 0.158
H3A[g]: Individual accountability of the AI-based CDSS → trust in the AI-based CDSS	.024	.81	–0.173 to 0.222
H3B[h]: Individual accountability of the AI-based CDSS → perceived threat of AI-based CDSS to professional identity	.339	.004	0.110 to 0.569
H4[i]: Trust in the AI-based CDSS → perceived threat of AI-based CDSS to professional identity	–.138	.047	–0.275 to –0.002
AI-based CDSS explainability → trust in the AI-based CDSS → perceived threat of AI-based CDSS to professional identity	–.073	—[j]	–0.145 to –0.009

[a]H1A: hypothesis 1A.

[b]AI: artificial intelligence.

[c]CDSS: clinical decision support system.

[d]H1B: hypothesis 1B.

[e]H2A: hypothesis 2A.

[f]H2B: hypothesis 2B

[g]H3A: hypothesis 3A.

[h]H3B: hypothesis 3B.

[i]H4: hypothesis 4.

[j]Not applicable.

A formal test of mediation was performed to analyze, independently of the other variables of influence, whether explainable AI-based CDSS predicts a perceived professional identity threat, and whether this direct path would be mediated by trust in the system. A total effect of an explainable AI-based CDSS on professional identity threat perceptions without involving trust was not observed (β=.065; *P*=.43). However, after incorporating the mediator trust in AI into the model, the explainability of an AI-based CDSS significantly predicted the mediator (β=.465; *P*<.001), which in turn significantly influenced perceptions of professional identity threat (β=–.156; *P*<.05). The direct effect of an explainable AI-based CDSS on professional identity threat perceptions via trust was not significant (β=.138; *P*=.12). Thus, our findings suggest that the relationship between AI-based CDSS explainability and professional identity threat perceptions is partially mediated by the trust in the system (indirect effect a×b=–.073 (95% CI –0.143 to –0.009).

Moreover, we have refined the estimation of the model by including other control variables, namely 5 personality dimensions: conscientiousness, extraversion, agreeableness, emotional stability, and openness to experience [68], as well as health care job tenure, and surgical specialties. Here we found that none of the variables had a significant effect on the threat to professional identity or on the change in path coefficients.

Discussion

Principal Results

The present study examines how AI-based CDSS process design features affect health care professionals' trust in AI and their perceived threat to professional identity. The findings show that decision outcome representations and human-AI interaction mechanisms shape health care professionals' trust and professional identity threat perceptions. Our study revealed that trust in AI partially mediates the influence of an explainable AI-based CDSS and strongly integrated AI-based system on the perceived identity threats among health care professionals. This supports earlier studies on AI explainability and clinical workflow integration, establishing trust in technology [10,34]. While AI decision-making may be challenging to understand in certain instances, providing clarity about the algorithm and AI advisory behavior may help align consumers' expectations with the performance of AI [11]. Our study also found a direct

XSL•FO
RenderX

and significant positive effect of explainable AI on the perceived threat to professional identity, a finding that aligns with research on the threat posed by AI in professional work [8]. In this context, AI's enhanced capabilities may raise questions not only about its compatibility with physicians' expertise and experience but also about its individual consequences. Future research needs to ascertain whether explainable AI undermines the perceived professional identity, whether trust in technology stabilizes through emergent dependability in real-life settings, or if this finding is favored by the controlled laboratory setting.

Comparison With Prior Work

In research on AI-based CDSSs, the effect of system integration into the clinical workflow on the emergence of professional identity threats has been investigated and considered as evidence of high absorption of individual medical competence [7,26,30]. Our study does not find any significant effect of AI-based system advice that is deeply integrated into the clinical workflow on individual professional identity threat assessments. This might be attributed to the different interpretations individuals have with high information content and medical guidelines. Future research should investigate human biases associated with AI-based advice, such as medical guidelines, not only to gain a deeper understanding of their functioning role in shaping professional identity threat perceptions among health care professionals but also to prevent their misuse or overuse. Furthermore, future AI research should focus on the influence of AI-based systems' reliability, that is, predicted consistent performance [12], on technological trustworthiness [11]. Another potential area of investigation may be the influence of various representations of AI reliability on threats to professional identity. Furthermore, the ongoing discussion about the ethical responsibility of AI-based CDSSs plays a significant role in this context [12]. Physicians must be able to adhere to the system's decision without fear of biased data composition.

Previous research on monitoring the behavior of health care professionals has identified it as an important factor for distrust in technology [45]. However, as our research indicates, the relationship between AI system monitoring and trust in AI could be more complex, as a system that induces accountability of the physician does not necessarily result in low trust in the AI system and rejection among health care professionals. Health care professionals may perceive a signing function as micromanagement and distrust it, for example, by misusing it to avoid responsibility and traceability of the AI tool's judgment. Our findings imply that such system functionalities interfere profoundly with personal physician identity perceptions. The significant positive relationship between an AI-assisted system that is inducing individual accountability and a perceived threat to professional identity demonstrates that agreement to an AI-assisted treatment decision is viewed as invasive and antagonistic to professional competence. Nevertheless, decision support can be seen as a functionality that enhances professional identity as it is followed rather than relying on individual professional competence. Especially in situations where direct consultation with a more experienced physician is not available, the reassurance of an AI-assisted decision and its validation may lead to increased trust in the technology [30]. Future research should explore how incorporating a human decision into an AI-assisted CDSS promotes trust in the technology and the socio-cognitive aspects of such technology adoption in the medical context.

The evaluative mechanisms used in this investigation to determine trust in AI outcomes predominantly address cognitive trust. Yet, empirical evidence underscores the significant role of emotional dimensions, such as affective trust, in the efficacious deployment of AI within the health care sector. It is imperative that forthcoming research delves more deeply into the determinants influencing the emotional aspects of trust in AI and additional factors such as cooperation and reliance in physician-AI relations [11].

Adopting such a human-centric perspective presents a substantial opportunity for interdisciplinary collaboration among scholars interested in the evolution and assimilation of AI technologies. While the bulk of existing studies on the interplay between humans and AI are mainly conducted by experts in cognitive engineering and information systems, the field of organizational research offers a valuable vantage point.

This study offers valuable insights for health care managers as well as developers and implementers of AI-based CDSSs, which should approach AI integration with an understanding of the health care professionals' perspective, factoring in aspects that foster human trust in AI and the relevance and value of specific tasks to the workforce. Consequently, in determining the tasks to delegate to AI, it is critical to evaluate not only the technological capabilities at their disposal but also to consider the human element, including perceived threats or enhancements to their professional identity, the interests and motivations of the employees, and the potential for enhancing their trust and productivity through AI collaboration. Specifically, our study shows that the explanation of AI-based CDSS decision outcomes proves beneficial for enhancing trust in AI among health care professionals. AI implementers and health care executives should focus on presenting AI decisions as recommendations that align with existing clinical workflows and which can be overruled by a human decision, rather than disrupting work processes. Furthermore, whenever feasible, AI-based CDSSs should refrain from querying accountability that exposes personal information about the operator. To resolve liability issues, established standards and legal frameworks should be developed, freeing individual practitioners of the duty of actively validating this information.

Limitations

This study has several limitations that warrant careful consideration. First, the small sample size constrained the ability to apply certain filtering criteria, thereby limiting the robustness of findings related to the influence of specific variables. A larger sample would facilitate more precise filtering, particularly the exclusion of preclinical medical students whose limited professional experience and socialization may not accurately represent perceptions of threats to professional identity. Moreover, the sample's composition, with 95% comprising medical students, presents a potential selection bias, as their experiences and perceptions may differ significantly from those of practicing physicians, nurses, or other health care professionals. Consequently, the generalizability of the findings

XSL•FO
RenderX

to a broader health care population is limited. Future research should aim to include a more heterogeneous sample of health care professionals to enhance the validity and applicability of the results.

Additionally, the study's use of scenario-based surveys, while methodologically advantageous in combining elements of both traditional surveys and experimental designs, presents inherent limitations. Despite attempts to balance internal and external validity, scenario-based studies may still fail to fully capture the complexity of real-world clinical settings, thus constraining the external validity and broader applicability of the findings [51]. For instance, laboratory studies indicate that users initially exhibit high levels of trust in AI-based CDSS, which diminishes over time as inconsistencies in risk prediction arise [69]. In contrast, field studies suggest that lower initial trust may increase through sustained interaction [70]. Such discrepancies highlight the difficulty in generalizing findings from controlled experimental settings to real-world environments. Furthermore, variability in respondents' interpretations of the hypothetical scenarios presented may introduce bias, potentially affecting the reliability of the results [51].

Finally, the study's focus on an AI-based risk prediction system for sepsis may limit the generalizability of the findings to other clinical conditions or AI applications in different contexts. Future research should examine whether the observed effects are consistent across AI systems designed for other clinical functions to provide a more comprehensive understanding of the factors influencing the adoption and trust of AI in health care settings.

Conclusions

In this study, we demonstrated how distinct AI process design features such as the explainability of AI-based CDSS decision outcomes, the integration depth of AI-based CDSS advice into the clinical workflow, and higher accountability through enforced signatures shape trust in AI and professional identity among health care professionals. Our research demonstrated the critical role that trust in AI played in fostering professional identity among health care workers. While explainable AI-based CDSS and systems highly integrated into clinical workflows enhanced trust, which supported the compatibility of the AI system with professional identity among health care professionals, explainable AI and enforced accountability led to a perceived professional identity threat. The findings of this study can be applicable to a variety of scenarios where the potential threat to professional identity posed by the rising deployment of AI-based systems played a role, especially in knowledge-intensive settings, where well-established and taken-for-granted behaviors, attitudes, and beliefs are questioned by AI solutions.

Acknowledgments

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors. The authors attest that there was no use of generative AI in the generation of text, figures, or other informational content of this paper.

Data Availability

Data are available from the corresponding author on reasonable request.

Authors' Contributions

SA conceived the study, developed the literature search, conducted the web-based survey, analyzed, and interpreted the data, conceptualized and wrote the sections of the manuscript (original draft, review, and editing), and created the tables and figures of the manuscript. KW conceived the study, interpreted the data, and edited the manuscript. RP and DD reviewed the manuscript. CS coordinated and conceived the study, interpreted the data, and edited the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Final list of items used for the hypothesis testing.
[DOCX File , 25 KB - formative_v9i1e64266_app1.docx]

Multimedia Appendix 2
Exploratory factor analysis.
[DOCX File , 25 KB - formative_v9i1e64266_app2.docx]

Multimedia Appendix 3
Sample properties of medical students, physicians, and nursing professionals.
[DOCX File , 38 KB - formative_v9i1e64266_app3.docx]

XSL•FO
RenderX

References

1. Kellogg KC. Subordinate activation tactics: Semi-professionals and micro-level institutional change in professional organizations. Adm Sci Q 2019;64(4):928-975. [doi: 10.1177/0001839218804527]
2. Huang MH, Rust RT. Artificial intelligence in service. J Serv Res 2018;21(2):155-172. [doi: 10.1177/1094670517752459]
3. Brynjolfsson E, McAfee A. The business of artificial intelligence: What it can—and cannot—do for your organization. Harvard Business Review. 2017. URL: https://hbr.org/2017/07/the-business-of-artificial-intelligence [accessed 2025-02-25]
4. Goh KH, Wang L, Yeow AYK, Poh H, Li K, Yeow JJL, et al. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. Nat Commun 2021;12(1):711 [FREE Full text] [doi: 10.1038/s41467-021-20910-4] [Medline: 33514699]
5. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542(7639):115-118 [FREE Full text] [doi: 10.1038/nature21056] [Medline: 28117445]
6. Fogel AL, Kvedar JC. Artificial intelligence powers digital medicine. NPJ Digit Med 2018;1(1):5 [FREE Full text] [doi: 10.1038/s41746-017-0012-2] [Medline: 31304291]
7. Liberati EG, Ruggiero F, Galuppo L, Gorli M, González-Lorenzo M, Maraldi M, et al. What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. Implement Sci 2017;12:1-13 [FREE Full text] [doi: 10.1186/s13012-017-0644-2] [Medline: 28915822]
8. Walter Z, Lopez MS. Physician acceptance of information technologies: role of perceived threat to professional autonomy. Decis Support Syst 2008;46(1):206-215. [doi: 10.1016/j.dss.2008.06.004]
9. Shin D. The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI. Int J Hum Comput Stud 2021;146:102551. [doi: 10.1016/j.ijhcs.2020.102551]
10. Ackerhans S, Huynh T, Kaiser C, Schultz C. Exploring the role of professional identity in the implementation of clinical decision support systems—a narrative review. Implement Sci 2024;19(1):11 [FREE Full text] [doi: 10.1186/s13012-024-01339-x] [Medline: 38347525]
11. Glikson E, Woolley AW. Human trust in artificial intelligence: review of empirical research. Acad Manag Ann 2020;14(2):627-660. [doi: 10.5465/annals.2018.0057]
12. Hoff KA, Bashir M. Trust in automation: integrating empirical evidence on factors that influence trust. Hum Factors 2015;57(3):407-434. [doi: 10.1177/0018720814547570] [Medline: 25875432]
13. Pavlou PA. Consumer acceptance of electronic commerce: integrating trust and risk with the technology acceptance model. Int J Electronic Commer 2003;7(3):101-134. [doi: 10.1080/10864415.2003.11044275]
14. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989;13(3):319-340. [doi: 10.2307/249008]
15. Panicker RO, Soman B, Gangadharan KV, Sobhana NV. An adoption model describing clinician's acceptance of automated diagnostic system for tuberculosis. Health Technol 2016;6:247-257. [doi: 10.1007/s12553-016-0136-4]
16. Fan W, Liu J, Zhu S, Pardalos PM. Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS). Ann Oper Res 2020;294(1):567-592. [doi: 10.1007/s10479-018-2818-y]
17. Choudhury A, Shamszare H. Investigating the impact of user trust on the adoption and use of ChatGPT: survey analysis. J Med Internet Res 2023;25:e47184 [FREE Full text] [doi: 10.2196/47184] [Medline: 37314848]
18. Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. Acad Manag Rev 1995;20(3):709-734. [doi: 10.5465/AMR.1995.9508080335]
19. Mcknight DH, Carter M, Thatcher JB, Clay PF. Trust in a specific technology: An investigation of its components and measures. ACM Trans Manage Inf Syst 2011;2(2):1-25. [doi: 10.1145/1985347.1985353]
20. Wang W, Qiu L, Kim D, Benbasat I. Effects of rational and social appeals of online recommendation agents on cognition- and affect-based trust. Decis Support Syst 2016;86:48-60. [doi: 10.1016/j.dss.2016.03.007]
21. Montague E. Validation of a trust in medical technology instrument. Appl Ergon 2010;41(6):812-821 [FREE Full text] [doi: 10.1016/j.apergo.2010.01.009] [Medline: 20189163]
22. Paul DL, McDaniel Jr RR. A field study of the effect of interpersonal trust on virtual collaborative relationship performance. MIS Q 2004;28(2):183-227. [doi: 10.2307/25148633]
23. Barrett M, Oborn E, Orlikowski WJ, Yates J. Reconfiguring boundary relations: robotic innovations in pharmacy work. Organ Sci 2012;23(5):1448-1466. [doi: 10.1287/orsc.1100.0639]
24. Pollock TG, Lashley K, Rindova VP, Han JH. Which of these things are not like the others? Comparing the rational, emotional, and moral aspects of reputation, status, celebrity, and stigma. Acad Manag Ann 2019;13(2):444-478. [doi: 10.5465/annals.2017.0086]
25. Ferlie E, Fitzgerald L, Wood M, Hawkins C. The nonspread of innovations: the mediating role of professionals. Acad Manag J 2005;48(1):117-134. [doi: 10.5465/AMJ.2005.15993150]
26. Jussupow E, Spohrer K, Heinzl A. Identity threats as a reason for resistance to artificial intelligence: survey study with medical students and professionals. JMIR Form Res 2022;6(3):e28750 [FREE Full text] [doi: 10.2196/28750] [Medline: 35319465]

XSL•FO
RenderX

27. Chreim S, Williams B, Hinings CR. Interlevel influences on the reconstruction of professional role identity. Acad Manag J 2007;50(6):1515-1539. [doi: 10.5465/amj.2007.28226248]
28. Pratt MG, Rockmann KW, Kaufmann JB. Constructing professional identity: the role of work and identity learning cycles in the customization of identity among medical residents. Acad Manag J 2006;49(2):235-262. [doi: 10.5465/amj.2006.20786060]
29. Weaver R, Peters K, Koch J, Wilson I. 'Part of the team': professional identity and social exclusivity in medical students. Med Educ 2011;45(12):1220-1229. [doi: 10.1111/j.1365-2923.2011.04046.x] [Medline: 21999250]
30. Lebovitz S, Lifshitz-Assaf H, Levina N. To engage or not to engage with AI for critical judgments: how professionals deal with opacity when using AI for medical diagnosis. Organ Sci 2022;33(1):126-148. [doi: 10.1287/orsc.2021.1549]
31. Van Biesen W, Van Cauwenberge D, Decruyenaere J, Leune T, Sterckx S. An exploration of expectations and perceptions of practicing physicians on the implementation of computerized clinical decision support systems using a Qsort approach. BMC Med Inform Decis Mak 2022;22(1):185 [FREE Full text] [doi: 10.1186/s12911-022-01933-3] [Medline: 35842722]
32. Fernández-Alemán JL, Señor IC, Lozoya P, Toval A. Security and privacy in electronic health records: a systematic literature review. J Biomed Inform 2013;46(3):541-562 [FREE Full text] [doi: 10.1016/j.jbi.2012.12.003] [Medline: 23305810]
33. Orlikowski WJ. Integrated information environment or matrix of control? The contradictory implications of information technology. Account Manag Inf Technol 1991;1(1):9-42. [doi: 10.1016/0959-8022(91)90011-3]
34. Hoffman RR, Mueller ST, Klein G, Litman J. Metrics for explainable AI: challenges and prospects. arXiv:1812.04608 2018. [doi: 10.48550/arXiv.1812.04608]
35. Doran D, Schulz S, Besold TR. What does explainable AI really mean? A new conceptualization of perspectives. arXiv preprint arXiv:1710.00794 2017. [doi: 10.48550/arXiv.1710.00794]
36. Yang M, Liu C, Wang X, Li Y, Gao H, Liu X, et al. An explainable artificial intelligence predictor for early detection of sepsis. Crit Care Med 2020;48(11):e1091-e1096. [doi: 10.1097/CCM.0000000000004550] [Medline: 32885937]
37. Zheng Q, Delingette H, Ayache N. Explainable cardiac pathology classification on cine MRI with motion characterization by semi-supervised learning of apparent flow. Med Image Anal 2019;56:80-95. [doi: 10.1016/j.media.2019.06.001] [Medline: 31200290]
38. Yan F, Chen Y, Xia Y, Wang Z, Xiao R. An explainable brain tumor detection framework for MRI analysis. Appl Sci 2023;13(6):3438. [doi: 10.3390/app13063438]
39. Lamb R, Davidson E. Information and communication technology challenges to scientific professional identity. Inf Soc 2005;21(1):1-24. [doi: 10.1080/01972240590895883]
40. Kaplan A, Haenlein M. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Bus Horiz 2019;62(1):15-25 [FREE Full text] [doi: 10.1016/j.bushor.2018.08.004]
41. Shibl R, Lawley M, Debuse J. Factors influencing decision support system acceptance. Decis Support Syst 2013;54(2):953-961. [doi: 10.1016/j.dss.2012.09.018]
42. Russ AL, Zillich AJ, McManus MS, Doebbeling BN, Saleem JJ. Prescribers' interactions with medication alerts at the point of prescribing: a multi-method, in situ investigation of the human-computer interaction. Int J Med Inform 2012;81(4):232-243. [doi: 10.1016/j.ijmedinf.2012.01.002] [Medline: 22296761]
43. de Visser E, Parasuraman R. Adaptive aiding of human-robot teaming: effects of imperfect automation on performance, trust, and workload. J Cogn Eng Decis Mak 2011;5(2):209-231. [doi: 10.1177/1555343411410160]
44. Jussupow E, Spohrer K, Heinzl A, Gawlitza J. Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence. Inf Syst Res 2021;32(3):713-735. [doi: 10.1287/isre.2020.0980]
45. Bernstein ES. The transparency paradox: A role for privacy in organizational learning and operational control. Adm Sci Q 2012;57(2):181-216. [doi: 10.1177/0001839212453028]
46. Möhlmann M, Zalmanson L. Hands on the wheel: Navigating algorithmic management and uber drivers' autonomy. 2017 Presented at: International Conference on Information Systems (ICIS 2017); December 10-13, 2017; Seoul, South Korea.
47. Thomas CP, Kim M, McDonald A, Kreiner P, Kelleher SJ, Blackman MB, et al. Prescribers' expectations and barriers to electronic prescribing of controlled substances. J Am Med Inform Assoc 2012;19(3):375-381 [FREE Full text] [doi: 10.1136/amiajnl-2011-000209] [Medline: 21946239]
48. Ananny M, Crawford K. Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. New Media Soc 2018;20(3):973-989. [doi: 10.1177/1461444816676645]
49. de Vries AE, van der Wal MH, Nieuwenhuis MM, de Jong RM, van Dijk RB, Jaarsma T, et al. Perceived barriers of heart failure nurses and cardiologists in using clinical decision support systems in the treatment of heart failure patients. BMC Med Inform Decis Mak 2013;13:54 [FREE Full text] [doi: 10.1186/1472-6947-13-54] [Medline: 23622342]
50. McAllister DJ. Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. Acad Manag J 1995;38(1):24-59. [doi: 10.2307/256727]
51. Atzmüller C, Steiner PM. Experimental vignette studies in survey research. Methodology 2010;6(3):128-138. [doi: 10.1027/1614-2241/a000014]
52. Niemi PM. Medical students' professional identity: Self-reflection during the preclinical years. Med Educ 1997;31(6):408-415. [doi: 10.1046/j.1365-2923.1997.00697.x] [Medline: 9463642]

XSL•FO
RenderX

53. Pitkala KH, Mantyranta T. Professional socialization revised: medical students' own conceptions related to adoption of the future physician's role—a qualitative study. Med Teach 2003;25(2):155-160. [doi: 10.1080/0142159031000092544] [Medline: 12745523]
54. Madill A, Latchford G. Identity change and the human dissection experience over the first year of medical training. Soc Sci Med 2005;60(7):1637-1647 [FREE Full text] [doi: 10.1016/j.socscimed.2004.08.035] [Medline: 15652694]
55. Improving the prevention, diagnosis and clinical management of sepsis. World Health Organisation. 2017. URL: https://www.who.int/activities/improving-the-prevention-diagnosis-and-clinical-management-of-sepsis [accessed 2025-02-20]
56. Fleischmann-Struzek C, Mikolajetz A, Schwarzkopf D, Cohen J, Hartog CS, Pletz M, et al. Challenges in assessing the burden of sepsis and understanding the inequalities of sepsis outcomes between National Health Systems: secular trends in sepsis and infection incidence and mortality in Germany. Intensive Care Med 2018;44(11):1826-1835 [FREE Full text] [doi: 10.1007/s00134-018-5377-4] [Medline: 30284637]
57. Bauer M, Groesdonk HV, Preissing F, Dickmann P, Vogelmann T, Gerlach H. [Mortality in sepsis and septic shock in Germany. Results of a systematic review and meta-analysis]. Anaesthesist 2021;70(8):673-680 [FREE Full text] [doi: 10.1007/s00101-021-00917-8] [Medline: 33559687]
58. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
59. Cohen J. Statistical power analysis. Curr Dir Psychol Sci 1992;1(3):98-101. [doi: 10.1111/1467-8721.ep10768783]
60. Brunkhorst FM, Weigand MA, Pletz M, Gastmeier P, Lemmen SW, Meier-Hellmann A, et al. [S3 Guideline Sepsis-prevention, diagnosis, therapy, and aftercare: long version]. Med Klin Intensivmed Notfmed 2020;115:37-109. [doi: 10.1007/s00063-020-00685-0] [Medline: 32356041]
61. Thompson B, Daniel LG. Factor analytic evidence for the construct validity of scores: a historical overview and some guidelines. Educ Psychol Meas 1996;56(2):197-208. [doi: 10.1177/0013164496056002001]
62. Kaiser HF. A second generation little jiffy. Psychometrika 2025;35(4):401-415. [doi: 10.1007/BF02291817]
63. Nunnally JC, Bernstein IH. Psychometric Theory. New York, NY: McGraw-Hill; 1994.
64. Agarwal R, Prasad J. A conceptual and operational definition of personal innovativeness in the domain of information technology. Inf Syst Res 1998;9(2):204-215. [doi: 10.1287/isre.9.2.204]
65. The PROCESS macro for SPSS, SAS, and R. processmacro.org. 2023. URL: https://processmacro.org/version-history.html [accessed 2025-02-20]
66. Hayes AF, Montoya AK, Rockwood NJ. The analysis of mechanisms and their contingencies: PROCESS versus structural equation modeling. Australas Mark J 2017;25(1):76-81. [doi: 10.1016/j.ausmj.2017.02.001]
67. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods 2008;40(3):879-891. [doi: 10.3758/brm.40.3.879] [Medline: 18697684]
68. Gosling SD, Rentfrow PJ, Swann WB. A very brief measure of the Big-Five personality domains. J Res Personal 2003;37(6):504-528. [doi: 10.1016/S0092-6566(03)00046-1]
69. Madhavan P, Wiegmann DA. Similarities and differences between human-human and human-automation trust: an integrative review. Theor Issues Ergon Sci 2007;8(4):277-301. [doi: 10.1080/14639220500337708]
70. Ullman D, Malle BF. Human-robot trust: Just a button press away. 2017 Presented at: HRI '17: Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction; March 6-9, 2017; Vienna, Austria p. 309-310 URL: https://dl.acm.org/doi/pdf/10.1145/3029798.3038423 [doi: 10.1145/3029798.3038423]

Abbreviations

AI: artificial intelligence
CDSS: clinical decision support system
EFA: exploratory factor analysis

Edited by A Mavragani, G Tsafnat; submitted 12.07.24; peer-reviewed by J Odumegwu, M Ji; comments to author 07.09.24; revised version received 01.11.24; accepted 05.01.25; published 26.03.25.

Please cite as:
Ackerhans S, Wehkamp K, Petzina R, Dumitrescu D, Schultz C
Perceived Trust and Professional Identity Threat in AI-Based Clinical Decision Support Systems: Scenario-Based Experimental Study on AI Process Design Features
JMIR Form Res 2025;9:e64266
URL: https://formative.jmir.org/2025/1/e64266
doi:10.2196/64266
PMID:

XSL•FO
RenderX

©Sophia Ackerhans, Kai Wehkamp, Rainer Petzina, Daniel Dumitrescu, Carsten Schultz. Originally published in JMIR Formative Research (https://formative.jmir.org), 26.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Stigma and Behavior Change Techniques in Substance Use Recovery: Qualitative Study of Social Media Narratives

Annie T Chen[1], MSIS, PhD; Lexie C Wang[2], MS; Shana Johnny[3], RN, MN, DNP; Sharon H Wong[1], MS; Rahul K Chaliparambil[4], BS; Mike Conway[5], PhD; Joseph E Glass[6], MSW, PhD

[1]Department of Biomedical Informatics and Medical Education, University of Washington School of Medicine, Seattle, WA, United States

[2]Department of Linguistics, University of Washington, Seattle, WA, United States

[3]School of Nursing, University of Washington, Seattle, WA, United States

[4]Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[5]School of Computing and Information Systems, University of Melbourne, Melbourne, Australia

[6]Kaiser Permanente Washington Health Research Institute, Seattle, WA, United States

Corresponding Author:
Annie T Chen, MSIS, PhD
Department of Biomedical Informatics and Medical Education
University of Washington School of Medicine
Box 358047
Seattle, WA, 98109
United States
Phone: 1 206 221 3369
Email: atchen@uw.edu

Abstract

Background: Existing literature shows that persons with substance use disorder (SUD) experience different stages of readiness to reduce or abstain from substance use, and tailoring intervention change strategies to these stages may facilitate recovery. Moreover, stigma may serve as a barrier to recovery by preventing persons with SUDs from seeking treatment. In recent years, the behavior change technique (BCT) taxonomy has increasingly become useful for identifying potential efficacious intervention components; however, prior literature has not addressed the extent to which these techniques may naturally be used to recover from substance use, and knowledge of this may be useful in the design of future interventions.

Objective: We take a three-step approach to identifying strategies to facilitate substance use recovery: (1) characterizing the extent to which stages of change are expressed in social media data, (2) identifying BCTs used by persons at different stages of change, and (3) exploring the role that stigma plays in recovery journeys.

Methods: We collected discussion posts from Reddit, a popular social networking site, and identified subreddits or discussion forums about 3 substances (alcohol, cannabis, and opioids). We then performed qualitative data analysis using a hybrid inductive-deductive method to identify the stages of change in social media authors' recovery journeys, the techniques that social media content authors used as they sought to quit substance use, and the role that stigma played in social media authors' recovery journeys.

Results: We examined 748 posts pertaining to 3 substances: alcohol (n=316, 42.2%), cannabis (n=335, 44.8%), and opioids (n=135, 18%). Social media content representing the different stages of change was observed, with the majority (472/748, 63.1%) of narratives representing the *action* stage. In total, 11 categories of BCTs were identified. There were similarities in BCT use across *precontemplation*, *contemplation*, and *preparation* stages, with social support seeking and awareness of natural consequences being the most common. As people sought to quit or reduce their use of substances (*action* stage), we observed a variety of BCTs, such as the repetition and substitution of healthful behaviors and monitoring and receiving feedback on their own behavior. In the *maintenance* stage, reports of diverse BCTs continue to be frequent, but offers of social support also become more common than in previous stages. Stigma was present throughout all stages. We present 5 major themes pertaining to the manifestation of stigma.

Conclusions: Patterns of BCT use and stigmatizing experiences are frequently discussed in social media, which can be leveraged to better understand the natural course of recovery from SUD and how interventions might facilitate recovery from substance use. It may be important to incorporate stigma reduction across all stages of the recovery journey.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e57468) doi:10.2196/57468

KEYWORDS

stigma; substance use; transtheoretical model; behavior change techniques; social media

Introduction

Background

As of 2022, among individuals aged ≥12 years in the United States, 17.3%, or 48.7 million people, met the diagnostic criteria for substance use disorder (SUD) in the last year [1]. Considering individuals needing treatment for SUD in the past year, it is estimated that only approximately one-fourth received treatment [2].

SUD is a highly stigmatized condition [3], and there is considerable evidence that individuals experiencing SUD are faced with stigmatizing experiences in their communities, at work, and in health care contexts [4]. Because stigma substantially worsens outcomes for individuals with SUD [5], mitigating stigma's effects is important in improving outcomes for individuals experiencing SUD.

A number of models have been developed to understand health behavior change [6]. In the context of SUD and addiction recovery, the most widely used theory is the transtheoretical model (TTM) [7]. The TTM was originally developed in the 1980s in the context of designing smoking cessation interventions and remains a dominant model for conceptualizing recovery phases [8]. The theory posits that individuals experiencing SUD can vary in their readiness to change a given health-related behavior and that determining a person's "change readiness" can facilitate the choice of an appropriate intervention. As such, successful behavior change is a *process* that takes place over time, rather than a single decision at a specific point in time. According to the TTM, an individual experiencing SUD can be located in 1 of 6 stages: *precontemplation* (ie, not intending to make a behavior change within the next 6 months), *contemplation* (ie, intending to make a change within the next 6 months), *preparation* (ie, intending to take action within the next 30 days), *action* (ie, making active, sustained attempts at behavior change for less than 6 months), *maintenance* (ie, keeping up a sustained behavior change for more than 6 months), and *termination* (ie, completed behavior change, with no desire to return to the previous behavior). In addition to the stages of change, the TTM argues that particular processes of change are effective for individuals at different stages, thus serving as recommendations for intervention design.

However, in recent years, there has been increased awareness of the need to better understand behavioral interventions, particularly due to the rapid evolution of technologies that are often used to deliver interventions [9]. The behavior change technique (BCT) taxonomy [10] is a resource that has been developed to serve as a method for specifying interventions and the components that account for their effectiveness. The taxonomy includes a standardized set of 93 BCTs proposed to be "active ingredients" of interventions (eg, personalized feedback about health behaviors, self-monitoring of symptoms and behaviors, and reinforcement of positive change).

Social media and other forms of consumer-generated textual data are a widely used resource for understanding health behaviors and serve as a useful complement to more established data collection methods (eg, surveys and qualitative interviews) [11,12]. A key advantage of social media as a source of data is that they provide spontaneous, naturalistic, unmediated, and first-person accounts of individual user's feelings and attitudes, which can provide insights that more hypothesis-driven data collection strategies may not necessarily capture [13].

Reddit is a large, predominantly text-based social media site organized into subreddits (ie, forums centered on a common theme or topic) in which users pseudonymously create posts and respond to other users' posts. Compared to short text microblog services such as Twitter (rebranded to X), video-based services (eg, YouTube and TikTok), and image-based services (eg, Pinterest and Instagram), Reddit encourages a more verbose, discursive style of interaction, often including rich, detailed narratives that share users' experiences, foregrounding the voices and experiences of individuals living with SUD. As such, Reddit has been extensively used as a resource for investigating substance use in the context of smoking cessation and vaping [14-16], cannabis use [16-18], alcohol use [19,20], and opioid use [21,22].

Objectives

With this paper, we aimed to build on existing work that uses Reddit data to explore social media users' experiences of using (and ceasing to use) alcohol, cannabis, and opioids. We selected these 3 substances because public perception differs for different substances [23]. The stigma associated with the use of a substance can vary due to factors such as perceived dangerousness, social acceptance, and legality [24], and these substances may complement one another in terms of the insights offered with respect to people's use experiences and recovery strategies. Moreover, alcohol and cannabis are among the most widely used substances in the United States [25], and there was widespread concern about the opioid crisis during the study period [26,27]. We used qualitative data analysis methods to characterize the behavior change strategies that may be used at different stages of the TTM and explored a challenge that Reddit authors report experiencing throughout the stages—stigma relating to substance use. Our investigation was framed around 3 research questions (RQs):

- RQ1: What are the manifestations of the stages of the TTM among Reddit authors?
- RQ2: Which BCTs emerge organically at different stages of the TTM?
- RQ3: How do participants experience stigma, and how do these experiences of stigma affect recovery?

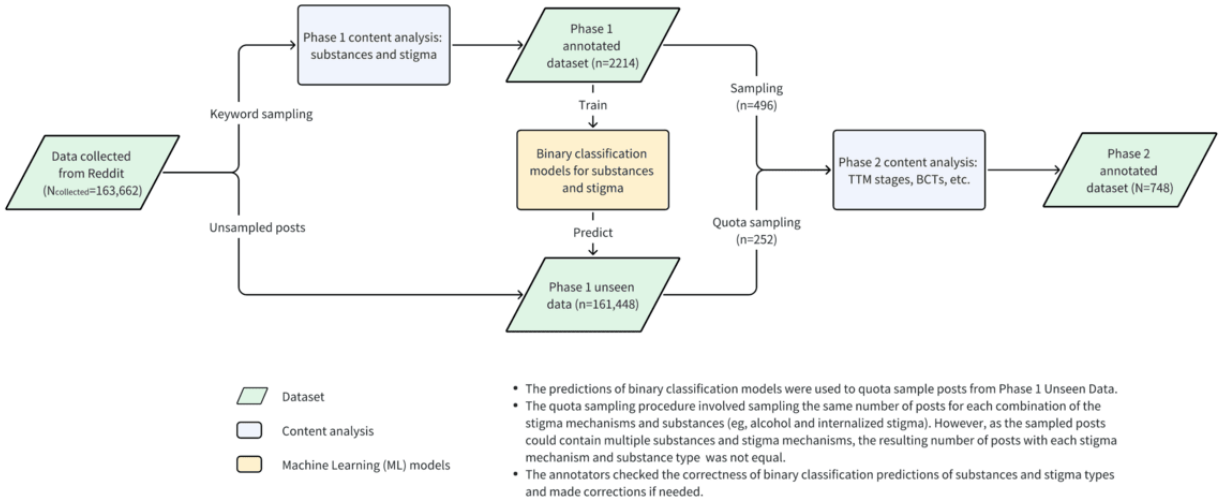
XSL•FO
RenderX

Methods

Data Collection and Preparation

The data collection and preparation process included multiple phases (Figure 1). In phase 1, as described in a previous study [28], we collected data from different subreddits relating to alcohol, cannabis, or opioid use (N=163,662 posts) using the pushshift.io application programming interface [29]. In the selection of subreddits, we sought to be inclusive of orientation toward use, including those more oriented toward recovery (eg, r/stopdrinking, r/leaves, and r/OpiatesRecovery) and those adopted various other positions regarding consumption and use (eg, r/alcohol, r/cripplingalcoholism, r/Marijuana, r/Petioles, r/trees, r/cannabis, r/opiates). The posts were authored between January 1, 2013, and December 31, 2019. Data collected from subreddits only included "initiating" posts, as previous work has reported that these posts tend to contain richer narratives than replies to the initiating post [30].

Figure 1. Data preparation flowchart illustrating the process of identifying subreddit posts containing narratives of stigma and recovery related to alcohol, opioid, and cannabis use. BCT: behavior change technique; TTM: transtheoretical model.



We then performed content analysis on a subset of the corpus (2214/163,662, 1.35%) sampled using keywords. Given the salience of stigma in substance use recovery, posts were eligible if they contained keywords suggesting that stigma might be present, including adjectives that are highly indicative of stigma (eg, "shame," "untrustworthy," and "disappoint"), pejorative labels (eg, "crackhead," "junkie," and "alcoholic"), and references to persons who might be involved in stigma-related experiences (eg, partners, parents, and coworkers). In phase 1 content analysis, we annotated the substances used and stigma mechanisms present (ie, internalized, anticipated, and enacted stigma). These procedures are described in detail in a prior study [31].

Building on this foundation, we developed a set of binary classification models using the phase 1 annotated dataset as training data. The models use machine learning (ML) algorithms to differentiate between the presence versus absence of substance use and stigma mechanisms. Architecturally, the models combine an autoencoder RoBERTa framework [32] and a multilayer perceptron that takes features that may be informative of stigma as input. The features include Term Frequency-Inverse Document Frequency weighted n-grams (ie, the n-grams are weighted based on their importance in a given post, relative to the corpus), affective and psychological lexical features, and additional handcrafted features. A detailed description of the development and evaluation of the classifiers is available in a prior study [33]. These models were then used to generate ML predictions of substances and stigma on the unseen data (161,448/163,662, 98.65%).

In phase 2, we performed qualitative data analysis to delve deeper into Reddit authors' experiences of stigma and the behaviors they engaged in while seeking to recover from substance use. To enrich the concentration of stigma in our sample, we sampled the first 496 posts in the phase 1 annotated dataset consisting of posts with and without stigma and combined this with another sample of 252 posts selected from the unseen data. Post selection was performed using the previously mentioned ML predictions of stigma and substances, with the goal of sampling equal proportions of the 3 stigma mechanisms and the 3 substances. Using this procedure, we were able to ensure that the resulting dataset for this study (N=748) contained sufficient posts relating to the 3 substances and a varied distribution of posts with and without stigma for subsequent qualitative analyses.

Data Analysis

We performed a qualitative analysis of the posts using a hybrid inductive-deductive method, which seeks to leverage the strengths of each respective approach [34]. The deductive or "top-down" part of our analysis was based on 2 theoretical frameworks and a taxonomy: the stigma framework [31,35], the stages of change from the TTM [7], and BCTs as defined in the behavior change taxonomy [10]. We also included the substances of interest.

As noted by Proudfoot [34], using a hybrid inductive-deductive approach to analysis "helps to ensure that the voices of the participants are valued, while simultaneously allowing for more theory-led analysis." Given the sensitive nature of our research

XSL•FO
RenderX

focus and the richness of our dataset, adopting an inductive component was particularly important for ensuring that these voices are heard. Thus, our codebook development included an inductive or "bottom-up" component in which we adapted the definitions to our data to account for the specifics of the health behavior context and data type. For example, though the stage of *action* has been defined as having modified their lifestyles within the past 6 months [7], often there is no explicit declaration of time in the social media data. Thus, to differentiate *action* from the next stage in the TTM, *maintenance*, our coding guide specifies that situations where social media authors appeared to be actively seeking to quit substance use are indicative of the *action* stage, whereas situations where users maintain routinized strategies are characteristic of the *maintenance* stage.

In developing codes for BCTs, we focused on the high-level groupings of the taxonomy [10], such as "antecedents," which include more specific techniques, such as "restructuring the physical environment" and "restructuring the social environment." Though the BCT taxonomy does not differentiate between social support seeking and offering, we chose to do so due to the diverse ways that social support has been studied in the literature [36] and the potential importance of differentiating support sought and received in terms of health outcomes [37]. The taxonomy includes 16 high-level groups; we only included a subset of these because the others did not appear salient in our data.

We also coded the substances of interest (eg, alcohol, cannabis, and opioids) and the stigma mechanisms encountered, identified either through qualitative coding or using the models and checked by the annotators, as described previously. The stigma mechanisms—enacted, anticipated, and internalized stigma—were adapted from the stigma framework [31,35]. Enacted stigma describes experiences of stereotyping, prejudice, and discrimination due to a stigmatized attribute; anticipated stigma is the expectation that one might experience stereotyping, prejudice, and discrimination in the future due to a stigmatized attribute; and internalized stigma is the application of negative feelings and beliefs about a stigmatized group to oneself (Table 1) [38]. For the stigma mechanisms, we also identified the actors involved (eg, family, friends, partners, coworkers, others, society, and self).

Our annotation process involved 2 primary coders (SJ and RKC), who independently coded the discussion posts; an adjudicator (SHW), who reviewed the codes and facilitated reconciliation; and 1 additional investigator (ATC), who provided additional feedback as needed. All disagreements were reconciled by majority or consensus. Interrater reliability for the substance and stigma codes are reported in a prior study [33]. For the stages of the TTM and BCTs, we assessed interrater reliability using Cohen κ [39]; overall interrater reliability was 0.71 between the 2 coders. Dedoose qualitative data analysis software was used to perform the coding [40]. For each research question, we report our findings using descriptive statistics as well as narrative descriptions of each theme.

Table 1. The stigma framework: definitions and examples.

	Definition	Example
Enacted stigma	Experiences of stereotyping, prejudice, and discrimination from others due to a stigmatized attribute.	"My husband called me an addict and said I'd never become clean."
Anticipated stigma	The expectation that one might experience stereotyping, prejudice, and discrimination from others due to a stigmatized attribute.	"My co-workers are giving me looks now at work after I went in drunk last week. I don't know how I'm going to face them day after day."
Internalized stigma	Individuals internalize negative stereotypes about the stigmatized group that they belong to and apply these to themselves, affecting their sense of self-worth.	"I'm a stoner. I am an awful person..."

Ethical Considerations

Social media can offer naturalistic, first-person accounts of individual user's feelings and attitudes, allowing for understandings that hypothesis-driven data collection strategies may not necessarily capture [13]. Moreover, prior literature has reported that social media authors often believe social media content is public domain and can be freely used in research without consent as long as individual information is not reported [41].

However, particular care must be taken to protect the identities and ensure the anonymity of content authors [42]. In this study, we analyzed the collected data without attempting to contact or connect social media authors to their posts; then, we used synthetic or paraphrased quotations based on the original social media content to illustrate the themes we present. As there are situations in which search engines might be used to retrieve social media posts [42], we constructed synthetic quotations to protect the identities of content authors.

In the development of these synthetic quotations, we seek to faithfully represent the voices of social media authors while protecting their identities by selecting original quotations that are representative of the data [43], faithful representation of concepts in the synthesized quotations, and using language that resembles language in the source medium to maximize authenticity. As we did not seek to connect social media authors to their posts, we also did not compensate them. This study has received ethics approval from both the University of Washington Human Subjects Division (STUDY00015737) and the University of Melbourne Human Research Ethics Committee (2023-25512-48127).

Reflexivity Statement

Our analysis incorporates many recommended practices for establishing rigor and trustworthiness in qualitative research

XSL•FO
RenderX

[44]. We ensured thick, rich descriptions using a sampling strategy that leveraged the large volume of available data while maintaining our specific research focus, and through a method of synthetic quotation creation. We performed *triangulation* during the data analysis and reporting process through the use of multiple coders of different backgrounds, including those with clinical experience (SJ and RKC) and informatics experience (ATC and SHW) and an investigator with significant experience in qualitative research, analysis of social media data, and the lived experiences of persons with chronic conditions (ATC). An important part of the annotation and synthetic quotation creation process involved discussion among the research team to ensure consistent coding over time, faithfulness to the data, and contextualization of the data considering the complementary backgrounds of the researchers. The discussion had several purposes, including consideration of different perspectives and mitigating the potential bias of any 1 researcher. Last but not least, our procedures facilitate *auditability* through detailed notes on our data preparation process and annotation process.

Results

Sample

Our sample comprised 748 posts pertaining to 3 substances: alcohol (n=316, 42.2%), cannabis (n=335, 44.8%), and opioids (n=135, 18%). Each post relates to at least 1 substance. On average, posts were 348 (SD 315) words long. Internalized stigma (369/748, 49.3%) was the most commonly expressed across all stages, but anticipated (131/748, 17.5%) and enacted stigma (111/748, 14.8%) were also present.

RQ1: How and to What Extent Are the Posts Relating to Each Stage Represented in the Sample?

The posts reflected different stages of change (Table 2). A subset of the posts was in the *precontemplation* stage, when social media authors do not express an intent to quit. These could include posts in which people are simply talking about experiences that they have had relating to substance use, which could include legal consequences, trouble at work or home, and more, but do not demonstrate an intention to quit. Often in the *precontemplation* stage, internalized stigma manifested as hopelessness or existential questions:

> *I feel horrible about my life every day. I feel bad about not taking care of my kids but I just keep picking up the bottle.... My life is hopeless.*
>
> *Am I depressed because I am a drunk or a drunk because I am depressed?*

In the *contemplation* stage, social media authors begin thinking about quitting but may identify barriers to doing so:

> *I'm thinking about quitting again, but I can't get time off of work to go into a detox program...*

They may also reach out to the Reddit community to ask their opinion about whether to quit, how to do so, and other related questions.

In the *preparation* stage, social media authors report taking different steps to quit, including deleting the phone numbers of dealers, planning to enroll in rehabilitation programs, telling friends and family about their use to seek support in the quitting process, and posting on Reddit to keep themselves "accountable." Many also mention that it is time to stop. At times, there is an event that may trigger the *preparation* stage, such as such as running out of the substance being used or being discovered by parents. Throughout the stages, feelings of depression, despair, and hopelessness are often expressed.

In the data that we annotated, the most commonly represented stage was *action*. In these posts, Reddit authors often shared their experiences with their recovery journeys. These posts often included rich detail concerning challenges that they faced:

> *I have been thinking about attending meetings but am also worried that the experiences of those worse off will cause me to slip.*
>
> *The holidays are the worst. I keep thinking that I'll just have a drink but I know that won't happen.*

In the *maintenance* stage, social media authors often share how long they have been in recovery and seek to encourage others by sharing their experiences:

> *This morning, I played with my son. I went to the park with him and enjoyed the moment with a clarity that I would have thought impossible a year ago. I'm so grateful that I'm now here. If you are struggling, take it one day at a time. It will get better.*

However, not all the narratives in the *maintenance* stage express positive sentiment. Some express continued cravings, feeling like they might relapse, and problems experienced while sober. While some say that they do not feel tempted at all, others say that they do not feel that they will relapse but sometimes reminisce about the effect of their substance of choice:

> *While I don't think I'm in any danger of relapsing, sometimes I feel the call of sweet mary jane and look back fondly.*

According to the TTM, people often slip back to prior stages. In the social media content, people often reported situations in which they had been sober for some time and then relapsed or are on the point of relapse:

> *Five months sober. Last month I decided I could have a drink while I was out with my coworkers, and it's been downhill from there. I have to stop!*
>
> *Feeling like I might relapse. I have been sober for a year and two months, but my life is as messed up now as before. I don't feel any better, and now I have no one to talk to. I don't want to start over again and feel the high heart rate, paranoia, guilt, and lack of motivation all over again, but the desire to smoke is overwhelming.*

As presented in Table 2, not all the data were representative of a stage of change. This could happen for various reasons. For example, a person might write a post providing advice but not clearly indicate that they were in recovery or provide commentary on relevant material, such as policy, a movie, or book on how to quit.

XSL•FO
RenderX

Table 2. Distribution of the stages of change in the dataset (N=748).

Stage	Posts, n (%)
Precontemplation	95 (12.7)
Contemplation	68 (9.1)
Preparation	43 (5.7)
Action	472 (63.1)
Maintenance	45 (6)
Unknown	25 (3.3)

RQ2: What BCTs Do People Naturally Use as They Seek to Recover From Substance Use in the Different Stages?

In Textbox 1, we present the BCTs observed, along with their definitions and examples. In Figure 2, we depict the prevalence of BCTs observed in the TTM stages, with the length of each bar reflecting the number of posts in which a given BCT was observed. The frequency of BCTs varied for the different stages of change. There were similarities in the *precontemplation*, *contemplation*, and *preparation* stages, with social support seeking and awareness of natural consequences being the most common:

> *Reddit, help! I have lost everything because of my habit... dropped out of college and am living at home with my parents. I feel like a failure. I need help but I don't know how to start.* [Social support seeking and natural consequences]

Posters reported that they sought social support online, in online support groups and other digital and web-based formats such as subreddits, as well as from family, friends, and meetings such as Alcoholics Anonymous.

In the *contemplation* stage, compared to the *precontemplation* stage, we observed a greater proportion of posts including some element of *shaping knowledge* or the acquisition of knowledge and information related to the antecedents of the behavior and ways to change it. At this time, social media authors expressed increased awareness of reasons or triggers for use, including loneliness, boredom, negative thoughts, and daily routines or life situations that make post authors more prone to use:

> *I want to quit and am looking for any advice you all have. I've been smoking for years and recently I've noticed all these changes in me that make me want to quit. First, I am anxious all of the time now and I can't sleep. I have no money and weed is draining my finances. I am also using weed as a coping mechanism for stress, which only makes me more stressed.* [Shaping knowledge and natural consequences]

In the *preparation* stage, we observed a greater proportion of goals and planning, such as seeking out accountability partners and making accountability posts. Post authors also express commitment to quitting or plans to enroll in treatment or detox programs.

In the *action* stage, as people sought to quit their use of substances, we observed a variety of BCTs. For example, by recognizing that merely having access to a substance would facilitate use, social media authors threw away alcohol and equipment that they used to engage in use, such as pipes. They also enlisted the help of others to ensure that they did not have access to substances and, at times, relocated (antecedents):

> *Gave away all of the wine in the house.* [Antecedents]
>
> *Put what little I had in my Civic and drove as far as I could south, that was 2 months ago and I haven't used since, aside from a sub or two for withdrawals. I lost everything when I was using—my friends, my job, my savings, my dignity. No more. Applying to jobs in this new town and planning to play pickup this evening at the local gym. It's time to rebuild.* [Antecedents, natural consequences, and goals and planning]

They also sought to engage in repetition and substitution of healthful behaviors, such as exercising while monitoring their own progress (feedback and monitoring):

> *I've started working out regularly and it has helped with the anxiety...* [Repetition and substitution]
>
> *Six days down! I hate what weed has done to me. I am always in a fog and can't remember what happened yesterday. My heart races. But I am slowly feeling better and clearer. I'm learning what withdrawal is like, trying to give space for my feelings, and gradually getting through it.* [Feedback and monitoring]
>
> *It has been a month since I last smoked. At first, I had a hard time sleeping and was tired all the time. But each day gets better. I have a little more energy. I can think more clearly.* [Feedback and monitoring]

XSL•FO
RenderX

Textbox 1. Behavior change technique (BCT) groupings, definitions, and examples (Adapted from the behavior change taxonomy [10]).

Natural consequences

- Definition: actions related to the identification of consequences of a behavior to be used as a facilitator of behavior change.
- Identified consequences should not be future, anticipated consequences. Only consequences that were experienced in the past or experienced in the present should be coded.
- Examples: becoming aware of health, social, environmental, and emotional consequences.

Antecedents

- Definition: actions that are related to reducing or mitigating antecedents (preconditions) that might lead to the undesired behavior.
- Examples: restructuring the physical and social environment (eg, moving to a new place) and avoidance or reducing exposure to behavioral cues, such as others who engage in substance use.

Shaping knowledge

- Definition: actions related to the accumulation of knowledge and information related to the behavior and ways to change it.
- Examples: information about antecedents and instructions on how to perform a behavior.

Repetition and substitution

- Definition: actions related to substituting the unwanted behavior with another and rehearsing the performance of a desired behavior.
- Examples: behavioral practice or rehearsal, behavioral substitution, and habit reversal.

Regulation

- Definition: actions related to emotional regulation and managing mental and cognitive stress.
- Examples: reducing negative emotions.

Goals and planning

- Definition: the identification of behaviors to be changed, creation of goals, and development of plans to facilitate change.
- Examples: reviewing goals and finding accountability partners.

Social support

- Definition: actions related to seeking, receiving, or providing social support to assist with behavior change. In this study, we used a broader conceptualization of social support, encompassing both emotional and informational support.
- Examples: general (unspecified), seeking, received, and providing and offering.

Comparison of outcomes

- Definition: the identification of potential or realized outcomes as a way to encourage or motivate behavior change.
- Examples: listing pros and cons and persuasive argument.

Identity

- Definition: shifting one's perspective on their past or current identity or actions to encourage behavior change.
- Examples: framing and reframing.

Self-belief

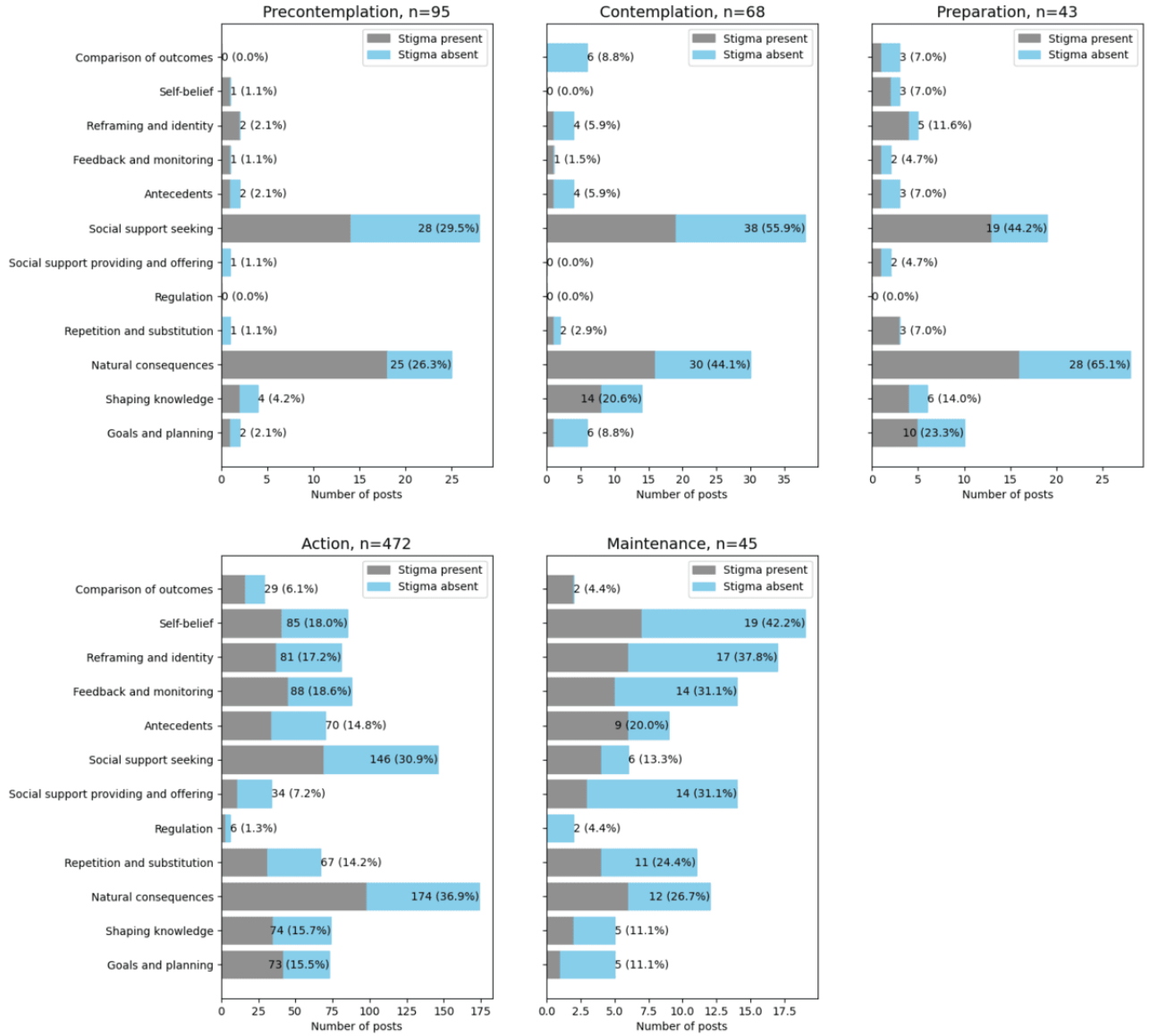
- Definition: increasing one's confidence in successfully performing the desired behavior through reinforcing one's belief in oneself.
- Examples: mental rehearsal of successful performance, focus on past success, and self-talk.

Feedback and monitoring

- Definition: monitoring one's behavior and the effects of one's behavior to facilitate the behavior change.
- Examples: self-monitoring of behavior and its outcomes.

XSL•FO
RenderX

Figure 2. The use of behavior change techniques (BCTs) and the presence of stigma mechanisms in individual stages of change.



Social support is an important part of the quitting process. Social media authors can experience challenges in removing antecedents (ie, triggers) from their environment if others are not supportive:

> *I have been trying to quit drinking, but a close friend insists on going to the bar whenever we hang out. Every time we go I have a hard time stopping myself from ordering a drink. When I bring this up, he says that I am just weak. What can I do?* [Shaping knowledge, antecedents, and social support seeking]

Social media authors also sought information from others as to how to maintain the new habits that they tried to adopt, while addressing symptoms that they experience in withdrawal:

> *The cravings are bad, but what is worse is not being able to sleep. Does that happen to you all? How long do the symptoms last?* [Social support seeking]

We also observed examples of positive self-talk and reframing, in which participants adopted or attempted to adopt new perspectives:

> *I've reached the two week mark. It hasn't been easy, but I'm proud of myself, and hopeful that I can do it.* [Self-belief]

> *I went to a party and for once, I didn't make a fool of myself! I had a great time, and was able to focus on every person I talked to...* [Self-belief]

In the *maintenance* stage, reports of diverse BCTs continue to be frequent. Notably, expressions of self-belief, altered perspective (ie, reframing and identity), and providing support and encouragement to others (ie, social support providing and offering) also become common:

> *Today I live without shame and look forward to the future. To everyone still struggling out there—you can do it!* [Social support providing and offering]

Many BCTs that became prominent in the *action* stage, such as self-belief, reframing and identity, feedback and monitoring, and repetition and substitution, remained common or were even observed more often in the *maintenance* stage. While this may partially be due to the nature of the venue—being an online

XSL•FO
RenderX

community in which people may strive to be positive to encourage others—this may also underscore the importance of continual reaffirmation and efforts at abstinence:

> *When I think of my past, I still feel the guilt and shame. But I think of them as a good reminder of what I don't want to go back to. I also tell myself that my past doesn't have to define me.* [Reframing and identity]

As we reflect back on the BCTs used, we observed that there was diversity in the strategies used. For many, it was a combination of BCTs and social support:

> *I told my partner last night about everything. She was incredibly supportive, and we talked about a plan. I am going to restart therapy. I am worried that I'll start using again, but for the first time in a long time, I feel hopeful.* [Goals and planning, social support seeking, and social support received]

As the previous excerpt demonstrates, this process can be a gradual one. While this poster states that others are supportive of their efforts, this is not always the case:

> *I am in rehab and on subs. I am trying hard to stay clean, but I still smoke weed. I am trying to be a good mother to my kids, but I feel so alone. I have no friends left. I had a friend that I kept up with, but it was too hard to stop using while continuing to spend time with them. I don't know what to do and would love to have some support.* [Antecedents and social support seeking]

Social media authors may experience challenges due to their life contexts, including feelings of guilt and other pressures due to family obligations. They may also face challenges dealing with exposure cues in their day-to-day lives. In addition, people may struggle with stress, self-doubt, and loneliness, leading them to ask for others' opinions on how to deal with these hardships.

RQ 3: How Do Participants Experience Stigma, and How Do These Experiences Affect Recovery?

Overview

In this research question, we aimed to examine how the presence of stigma may affect social media authors' recovery journeys. In Figure 2, we observed that substantive amounts of stigma (ie, in blue) are present in all stages. In this section, we review key themes involving stigma, which were often observed in multiple stages. We also present stage-specific characterizations as appropriate.

Theme 1: Experiencing Enacted Stigma From Others

Social media authors reported experiences of enacted stigma involving different persons that they encounter. Of the different actors that we examined, enacted stigma from family and those who were not part of a person's social circle was most common. Authors mentioned experiences with pharmacists and physicians discounting health issues, attributing health issues to substance use, and being reticent to provide services when they were aware that participants engaged in substance use. In addition, Reddit authors often spoke of how their use took a toll on their relationships with family, friends, and partners, resulting in broken trust and interpersonal conflict:

> *I got into a huge fight with my parents. They called me a drug addict and made me feel worthless.* [Enacted stigma]

One scenario that occasionally arose as social media authors sought to quit substance use was resistance from their social circle. For example, there were examples of those who told their family they wanted to quit drinking, and their family would say that they were not an "alcoholic" or "You're not that far gone." Social media authors often had to find new support networks, such as Alcoholics Anonymous, as their family and friends were not always supportive of their efforts to quit.

Theme 2: Experiencing Anticipated Stigma

Social media authors also experienced anticipated stigma. Many feared disappointing others, particularly family and friends, or feared repercussions at work, and thus hid their substance use. As many keep their substance use a secret, they may reach out to the Reddit community to ask questions that they are not able to ask elsewhere, often using temporary (ie, "throwaway") accounts:

> *This is a throwaway account. I have been addicted to oxy for years and am worried I will lose my job if anyone finds out. I am thinking of taking suboxone but am wondering if it would be better to quit cold turkey or would the withdrawals be worse?* [Anticipated stigma]

Some authors reach out on Reddit asking whether they should disclose their use to others around them. There are also times when they do choose to do so and are surprised to find that others are empathetic and supportive:

> *I have been feeling so ashamed and guilty, but I finally told my parents about my drinking and they were surprisingly supportive. I am beginning a rehab program tomorrow.* [Goals and planning, internalized stigma, and anticipated stigma]

Similar to enacted stigma, Reddit authors most often anticipated stigma from family and others outside of their social circle.

Theme 3: The Challenges of Internalized Stigma

There were posts that expressed internalized stigma, including internal struggles that people experienced due to regret, guilt, shame, and disappointment:

> *I have been able to avoid drinking when others are, but my biggest challenge is when I am alone and start to have negative thoughts and sink into this pool of shame and disappointment. Then I think about what a failure I am and how I have let everyone down.* [Internalized stigma]

Social media authors often shared challenges that they experienced adjunctive to substance use, such as financial difficulties, poor mental health, and increased social isolation. In the *contemplation* stage, social media authors often speak of "hating themselves" and hesitate to take steps forward due to

XSL•FO
RenderX

lack of confidence in their own ability to succeed. In contrast, in the *maintenance* stage, social media authors share that they feel less guilt and have more compassion for themselves:

> *I always felt this deep self-loathing each time I slipped. I didn't think I would ever make it. But little by little, I am learning to have compassion for myself. Life is still hard, but it gets a little easier one day at a time.* [Internalized stigma]

The process of recovery was gradual, and there could be setbacks:

> *I started out the year in constant shame. I was taking a quick sip whenever I had the chance, and then worrying that I was messing up my child. My mom, who is an alcoholic, came to visit, setting off a three-day bender. Now I am back on track. I haven't had a drop for a week and I haven't even been tempted. But I have some stressful stuff coming up at work and am worried. Thanks for listening.* [Feedback and monitoring, natural consequences, and internalized stigma]

As these posts illustrate, there can be triggers due to work and interpersonal interactions, and people may try for some time, intermittently experiencing setbacks, before being able to sustain abstinence.

Theme 4: Multiple Forms of Stigma

There were also examples of multiple forms of stigma in the same post. Shame, guilt from letting others down, and broken trust leading to tolls on relationships with others were common challenges:

> *I have lost family and friends due to my drug use. My relatives hide their wallets whenever I go over. I feel terrible that no one trusts me, and the worst failure ever.* [Enacted stigma and internalized stigma]

In earlier stages, people often report hiding their substance use from others and experiencing feelings of negative self-worth from doing so:

> *I hide my stash from my husband. The other day my five-year old accidentally found it and I was mortified. I feel horrible about hiding this but I don't think I can stop...* [Anticipated stigma and internalized stigma]

As illustrated by the quotation above, people may feel low self-worth. As a result, they may also avoid others or be excluded by others (enacted stigma), resulting in social isolation.

Theme 5: Enacted Stigma Directed Toward Others

Social media authors also demonstrated enacted stigma toward others. One common situation was when they referred to family members using pejorative labels such as "addicts" or "alcoholics," which occurred throughout the stages. Another example was a lack of empathy for others among some who felt that they were succeeding in their recovery journeys (ie, often occurring in the *action* or *maintenance* stages):

> *As I stopped smoking, I came to look down upon the behavior of my stoner friends...* [Enacted stigma]

In moments of introspection, some individuals reflected on their own behavior, seeking input from fellow Redditors while contemplating their motivations. For example, 1 social media author mentioned that they wanted to help others who were struggling but also felt it is impossible, and so instead chose to distance themselves.

Discussion

Principal Findings

"In this study, we performed qualitative data analysis to identify the stages of change and BCTs used by social media authors in various stages of readiness to reduce their use of the following substances: alcohol, cannabis, and opioids. There were some similarities in the use of BCTs in the earlier stages of change (ie, *precontemplation*, *contemplation*, and *preparation*). In posts with content that reflected the *contemplation* stage, authors described an awareness of the negative impacts that substance use had on their lives (ie, natural consequences), and as they prepared to quit, they increasingly took concrete steps to do so by sharing their decision with their social network and committing to it (ie, social support seeking, goals, and planning) and seeking help in other ways.

Posts in the *action* stage reflected that individuals leveraged other BCTs, such as modifying their daily routines (ie, repetition and substitution), modifying their physical and social environment (ie, antecedents), self-affirmation (ie, self-belief), and viewing their situation through new lenses (ie, reframing and identity). Substantive use of BCTs are observed, and in some cases more prominently, in the *maintenance* stage. This may partly stem from the nature of online discussions, where many individuals aim to give back to the community by sharing their experiences with successful health behavior change, offering encouragement to those trying to quit.

We also examined the role of stigma in social media authors' recovery journeys. We observed substantive amounts of stigma across all stages of change. The influence of social interactions was also salient. Social media authors shared situations in which they experienced or anticipated stigma from strangers, acquaintances, and loved ones. When people receive support from others, this can be affirming and facilitate recovery; however, social isolation can also result from isolating oneself or from enacted stigma. Some social media authors reported that the fear of disappointing others prevented them from reaching out for support. In contrast, other authors sought support from both online and offline social networks (eg, family and friends) and reported that this support was instrumental in successful recovery journeys. Previous research has found that perceived stigma is associated with poorer mental health, while perceived social support is linked to better mental health [45].

Comparison With Prior Work

This study extends the existing literature concerning recovery mechanisms by identifying situations in which commonly used BCTs appeared in social media narratives, thus affording insight into how people may seek to address the challenges of recovery in everyday life. Consistent with existing literature, which has identified controlling one's environment and social support as

XSL•FO
RenderX

commonly used strategies for behavioral change in the context of substance use [46,47], we observed examples of persons seeking the support of others (*social support seeking*) and using *antecedents*. These including discarding equipment necessary for substance use behavior and physically removing themselves from temptations. However, posters also lamented the difficulties of encountering "triggers" or cues in their day-to-day life that provoked relapse. The use of the antecedents BCT is also common in evidence-based substance use treatments [48], and a prior systematic review found that remotely delivered substance use interventions [49] incorporating intervention components related to "avoidance/reducing exposure to cues for behavior," "pros and cons," and "self-monitoring of behavior" may be effective. Social media authors also sought to restructure their lives to mitigate exposures and observed the effects of behavioral change.

In this study, we also observed that self-belief and cognitive reframing (changing one's perspective), were more common in the *action* and *maintenance* stages. As recovery trajectories progress, there is a trend toward individualistic and agentic identity [50]; BCTs such as self-belief and cognitive reframing may facilitate this process.

Our findings also illustrate the importance of temporality and flux in the experience of the person in recovery, who is not static but rather evolving due to their experiences over time. At times, people may struggle with low self-worth due to internalized and anticipated stigma, impeding their path to recovery; at other times, one might also observe enacted stigma from them toward someone else, often when the person is in recovery. There is also a need for more research concerning how family and romantic partners may also enact stigma [51].

Practice Implications

These findings have implications for the design of stigma reduction interventions to promote substance use recovery. Textbox 2 summarizes the key themes identified, along with implications for subsequent intervention design. For instance, to counter stigma and feelings of isolation, interventions could provide education to improve self-advocacy and emotion regulation. In particular, treatments such as acceptance and commitment therapy could potentially combat internalized stigma by assisting individuals with acceptance and compassion toward the self [52-54].

Textbox 2. Practice implications considering the role of stigma in recovery journeys.

Experiencing enacted stigma from others

- Provide persons with substance use disorder (SUDs) training in navigating challenging situations involving enacted stigma (eg, in self-advocacy with providers and emotion regulation in everyday situations).
- Assist persons with SUDs to develop alternative support networks and other mechanisms to nurture resilience.

Experiencing anticipated stigma

- Promote the use of behavior change techniques (BCTs) encouraging persons with SUDs to develop and strengthen support mechanisms.
- Teach strategies for disclosure of topics related to one's SUD.

The challenges of internalized stigma

- Emphasize self-compassion.
- The internal struggles and gradual process were significant obstacles. Intervention components that serve as a toolkit for participants to combat these challenges, such as feedback and monitoring and comparison of outcomes, can perhaps bolster natural mechanisms.

Multiple forms of stigma

- Use approaches that address the different stigma mechanisms individually but also emphasize that the process of recovery involves rebuilding on multiple fronts.

Enacted stigma directed toward others

- Assist persons with SUDs to consider the situations of others with an open countenance.

In addition, interventions could encourage the development of new support networks or teach strategies to intentionally seek support from others. Doing so could lessen the impact of the stigma while developing a social network that is supportive of one's recovery. Given that family interactions were a common source of stigma, interventions could teach strategies to encourage positive interactions with family members or facilitate development of additional support networks outside the family as needed. These study findings show that when family members were supportive, participants valued the support of those that they were close to.

Reddit authors also struggled with internalized stigma and challenges involving multiple forms of stigma, in which authors felt bad about their interactions with others. Interventions that assist persons experiencing self-stigma to disclose and seek support from others could be helpful [55].

Given that this study focused on occurrences of stigma expressed by people who were engaged in online forums, it is also worthwhile to consider how these findings have implications for the design of internet-based interventions. Reddit authors often mentioned the improvements that they observed over time. These improvements seemed to increase their confidence and

XSL•FO
RenderX

hope that they would continue to improve. There may be an opportunity here to leverage mobile health apps to facilitate self-monitoring and tracking of incremental improvements. Previous research has also recognized the value of using wearables to monitor biologic data, such as blood alcohol concentration sensors [56], and self-monitoring surveys can help participants reflect on associations between self-monitoring domains and adhere to drug use reduction goals [57]. These are potentially fruitful directions for future research.

The findings from this study could also improve the degree of personalization of content, such as delivering stage-matched intervention content. For example, the results suggest that *shaping knowledge* (information about changing a behavior) is more prominent in the *contemplation* stage as opposed to *precontemplation*, and that people use diverse BCTs as they actively seek to quit substance use. Thus, it may be beneficial to present content that encourages thoughts about antecedents or other methods of shaping knowledge to individuals in the *contemplation* stage, while offering suggestions for different BCTs to those in the *preparation* or *action* stages. In the *maintenance* stage, we observed that some individuals faced the potential for relapse, and other perceived no improvement in their lives. Thus, there may be a need for intervention content tailored to particular challenges of maintenance, such as the position of abstinence as opposed to moderation or maintaining recovery in the context of other aspects of life.

Finally, prior literature has observed that those who engage in social media value the information and support from the online communities [28,36,58,59]. In this study, we observed that social media authors experienced different information and support needs in different phases of their recovery journey. During the *action* phase, social media authors were likely to have questions about the quitting experience, including strategies, withdrawal symptoms, and timing of those symptoms. At all phases, people may experience questions about social interactions, such as “protecting their quit” and changes in how they may view and relate to others as they engage in maintenance. Thus, insights from these study findings can inform the provision of appropriate information at different stages of the recovery journey.

Limitations

This study has various limitations. First, as this study was based on social media data, there is bias as to who is represented due to differences in people’s decisions about which media to use [60]. However, as indicated by the sample posts, social media authors shared their challenges with substance use in diverse life contexts, including home, work, and school contexts. For example, some social media authors experienced concern with the effect that their use might have on their children, while others described their fears and anxieties with respect to work repercussions. Thus, the study findings still have implications for the diverse life contexts in which people may confront and seek to apply strategies for recovery.

In addition, this study used sampling approaches that sought to increase the proportion of posts containing stigma in the sample. As such, the posts in our sample may not represent the natural distribution of stages of change in substance use–related content on Reddit. There was less data available on stages other than *action*, and there is a need for more research concerning information and support needs of persons with SUDs in the earlier phases. Finally, in this study, we used content analysis to identify different stages of change to characterize BCT use in the different stages. The study does not examine longitudinal changes in the same person over time, and this could be an area for future exploration.

Conclusions

This study used social media as a naturalistic data source to explore how people may use BCTs to recover from substance use, including how this use may vary depending on a person’s readiness to quit or reduce their substance use. We observed that there were differences in the BCTs that were most commonly used in the different stages. Stigma was present in a substantive portion of the narratives at all stages, suggesting that there is a need to incorporate intervention components to assist participants address stigma-related challenges. Consideration of these patterns might be used to personalize and deliver content in future interventions promoting substance use recovery.

Acknowledgments

This study was supported by the National Institute on Drug Abuse of the National Institutes of Health (R21DA056684). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the potential sensitivity of the social media posts included in our study, to protect social media authors’ identities and anonymity.

Authors' Contributions

ATC and JEG conceptualized the study. SJ, SHW, and RKC performed the content analysis. LCW performed the data processing and rendered the data visualizations. ATC, MC, and SHW authored the initial draft. ATC, LCW, and JEG revised the manuscript. All authors provided feedback and approved the submission.

XSL•FO
RenderX

Conflicts of Interest

Pear Therapeutics provided digital therapeutic prescriptions at no cost to Kaiser Permanente Washington during a quality improvement pilot study led by JEG.

References

1. Highlights for the 2022 national survey on drug use and health. Substance Abuse and Mental Health Services Administration. URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt42731/2022-nsduh-main-highlights.pdf [accessed 2024-02-11]
2. SAMHSA announces National Survey on Drug Use and Health (NSDUH) results detailing mental illness and substance use levels in 2021. U.S. Department of Health and Human Services. 2023 Jan 04. URL: https://www.hhs.gov/about/news/2023/01/04/samhsa-announces-national-survey-drug-use-health-results-detailing-mental-illness-substance-use-levels-2021.html [accessed 2023-11-22]
3. Corrigan P, Schomerus G, Shuman V, Kraus D, Perlick D, Harnish A, et al. Developing a research agenda for understanding the stigma of addictions part I: lessons from the mental health stigma literature. Am J Addict 2017 Jan;26(1):59-66. [doi: 10.1111/ajad.12458] [Medline: 27779803]
4. Yang LH, Wong LY, Grivel MM, Hasin DS. Stigma and substance use disorders: an international phenomenon. Curr Opin Psychiatry 2017 Sep;30(5):378-388 [FREE Full text] [doi: 10.1097/YCO.0000000000000351] [Medline: 28700360]
5. van Boekel LC, Brouwers EP, van Weeghel J, Garretsen HF. Stigma among health professionals towards patients with substance use disorders and its consequences for healthcare delivery: systematic review. Drug Alcohol Depend 2013 Jul 01;131(1-2):23-35. [doi: 10.1016/j.drugalcdep.2013.02.018] [Medline: 23490450]
6. Davis R, Campbell R, Hildon Z, Hobbs L, Michie S. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. Health Psychol Rev 2015;9(3):323-344 [FREE Full text] [doi: 10.1080/17437199.2014.941722] [Medline: 25104107]
7. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
8. Enhancing Motivation for Change in Substance Use Disorder Treatment: Updated 2019. Rockville, MD: Substance Abuse and Mental Health Services Administration (US); 2019.
9. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
10. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
11. Paul M, Dredze M. You are what you tweet: analyzing Twitter for public health. Proc Int AAAI Conf Web Soc Media 2021 Aug 03;5(1):265-272. [doi: 10.1609/icwsm.v5i1.14137]
12. Amaya A, Bach R, Keusch F, Kreuter F. New data sources in social science research: things to know before working with Reddit data. Soc Sci Comput Rev 2019 Dec 18;39(5):943-960. [doi: 10.1177/0894439319893305]
13. Conway M, O'Connor D. Social media, big data, and mental health: current advances and ethical implications. Curr Opin Psychol 2016 Jun;9:77-82 [FREE Full text] [doi: 10.1016/j.copsyc.2016.01.004] [Medline: 27042689]
14. Struik L, Yang Y. e-cigarette cessation: content analysis of a quit vaping community on Reddit. J Med Internet Res 2021 Oct 25;23(10):e28303 [FREE Full text] [doi: 10.2196/28303] [Medline: 34694229]
15. Hu M, Benson R, Chen AT, Zhu SH, Conway M. Determining the prevalence of cannabis, tobacco, and vaping device mentions in online communities using natural language processing. Drug Alcohol Depend 2021 Nov 01;228:109016 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.109016] [Medline: 34560332]
16. Benson R, Hu M, Chen AT, Zhu SH, Conway M. Examining cannabis, tobacco, and vaping discourse on Reddit: an exploratory approach using natural language processing. Front Public Health 2022 Jan 05;9:738513 [FREE Full text] [doi: 10.3389/fpubh.2021.738513] [Medline: 35071153]
17. Sowles SJ, Krauss MJ, Gebremedhn L, Cavazos-Rehg PA. "I feel like I've hit the bottom and have no idea what to do": supportive social networking on reddit for individuals with a desire to quit cannabis use. Subst Abus 2017;38(4):477-482 [FREE Full text] [doi: 10.1080/08897077.2017.1354956] [Medline: 28704167]
18. Rhidenour KB, Blackburn K, Barrett AK, Taylor S. Mediating medical marijuana: exploring how veterans discuss their stigmatized substance use on Reddit. Health Commun 2022 Sep;37(10):1305-1315. [doi: 10.1080/10410236.2021.1886411] [Medline: 33602000]
19. Ricard BJ, Hassanpour S. Deep learning for identification of alcohol-related content on social media (Reddit and Twitter): exploratory analysis of alcohol-related outcomes. J Med Internet Res 2021 Sep 15;23(9):e27314 [FREE Full text] [doi: 10.2196/27314] [Medline: 34524095]

XSL•FO
RenderX

20. Gauthier RP, Costello MJ, Wallace JR. “I will not drink with you today”: a topic-guided thematic analysis of addiction recovery on Reddit. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI '22; April 29-May 5, 2022; New Orleans, LA. [doi: 10.1145/3491102.3502076]
21. Balsamo D, Bajardi P, De Francisci Morales G, Monti C, Schifanella R. The pursuit of peer support for opioid use recovery on Reddit. Proc Int AAAI Conf Web Soc Media 2023 Jun 02;17(1):12-23 [FREE Full text] [doi: 10.1609/icwsm.v17i1.22122]
22. Krawczyk N, Bunting AM, Frank D, Arshonsky J, Gu Y, Friedman SR, et al. "How will I get my next week's script?" Reactions of Reddit opioid forum users to changes in treatment access in the early months of the coronavirus pandemic. Int J Drug Policy 2021 Jun;92:103140 [FREE Full text] [doi: 10.1016/j.drugpo.2021.103140] [Medline: 33558165]
23. Cunningham JA, Sobell LC, Chow VM. What's in a label? The effects of substance types and labels on treatment considerations and stigma. J Stud Alcohol 1993 Nov;54(6):693-699. [doi: 10.15288/jsa.1993.54.693] [Medline: 8271805]
24. Janulis P, Ferrari JR, Fowler P. Understanding public stigma toward substance dependence. J Appl Soc Psychol 2013 Apr 18;43(5):1065-1072. [doi: 10.1111/jasp.12070]
25. Key substance use and mental health indicators in the United States: results from the 2022 National Survey on Drug Use and Health. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. 2023. URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt42731/2022-nsduh-nnr.pdf [accessed 2025-03-07]
26. Skolnick P. The opioid epidemic: crisis and solutions. Annu Rev Pharmacol Toxicol 2018 Jan 06;58:143-159. [doi: 10.1146/annurev-pharmtox-010617-052534] [Medline: 28968188]
27. Saloner B, McGinty EE, Beletsky L, Bluthenthal R, Beyrer C, Botticelli M, et al. A public health strategy for the opioid crisis. Public Health Rep 2018;133(1_suppl):24S-34S [FREE Full text] [doi: 10.1177/0033354918793627] [Medline: 30426871]
28. Chen AT, Johnny S, Conway M. Examining stigma relating to substance use and contextual factors in social media discussions. Drug Alcohol Depend Rep 2022 May 05;3:100061 [FREE Full text] [doi: 10.1016/j.dadr.2022.100061] [Medline: 36845987]
29. Baumgartner J, Zannettou S, Keegan B, Squire M, Blackburn J. The Pushshift Reddit dataset. Proc Int AAAI Conf Web Soc Media 2020 May 26;14(1):830-839. [doi: 10.1609/icwsm.v14i1.7347]
30. MacLean D, Gupta S, Lembke A, Manning C, Heer J. Forum77: an analysis of an online health forum dedicated to addiction recovery. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing. 2015 Presented at: CSCW '15; March 14-18, 2015; Vancouver, BC. [doi: 10.1145/2675133.2675146]
31. Earnshaw VA, Smith LR, Chaudoir SR, Amico KR, Copenhaver MM. HIV stigma mechanisms and well-being among PLWH: a test of the HIV stigma framework. AIDS Behav 2013 Jun;17(5):1785-1795 [FREE Full text] [doi: 10.1007/s10461-013-0437-9] [Medline: 23456594]
32. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: a robustly optimized BERT pretraining approach. arXiv. Preprint posted online on July 26, 2019 2025 [FREE Full text] [doi: 10.48550/arXiv.1907.11692]
33. Roesler D, Johnny S, Conway M, Chen AT. A theory-informed deep learning approach to extracting and characterizing substance use-related stigma in social media. BMC Digit Health 2024 Aug 16;2:60. [doi: 10.1186/s44247-024-00065-0]
34. Proudfoot K. Inductive/deductive hybrid thematic analysis in mixed methods research. J Mix Methods Res 2022 Sep 20;17(3):308-326. [doi: 10.1177/15586898221126816]
35. Livingston JD, Milne T, Fang ML, Amari E. The effectiveness of interventions for reducing stigma related to substance use disorders: a systematic review. Addiction 2012 Jan;107(1):39-50 [FREE Full text] [doi: 10.1111/j.1360-0443.2011.03601.x] [Medline: 21815959]
36. Chen AT. What's in a virtual hug? A transdisciplinary review of methods in online health discussion forum research. Libr Inf Sci Res 2014 Apr;36(2):120-130. [doi: 10.1016/j.lisr.2013.10.007]
37. Wang YC, Kraut R, Levine JM. To stay or leave?: the relationship of emotional and informational support to commitment in online health support groups. In: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. 2012 Presented at: CSCW '12; February 11-15, 2012; Washington, DC. [doi: 10.1145/2145204.2145329]
38. Smith LR, Earnshaw VA, Copenhaver MM, Cunningham CO. Substance use stigma: reliability and validity of a theory-based scale for substance-using populations. Drug Alcohol Depend 2016 May 01;162:34-43 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.02.019] [Medline: 26972790]
39. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960 Apr;20(1):37-46. [doi: 10.1177/001316446002000104]
40. Salmona M, Lieber E, Kaczynski D. Qualitative and Mixed Methods Data Analysis Using Dedoose: A Practical Approach for Research Across the Social Sciences. Thousand Oaks, CA: SAGE Publications; 2019.
41. Bond CS, Ahmed OH, Hind M, Thomas B, Hewitt-Taylor J. The conceptual and practical ethical dilemmas of using health discussion board posts as research data. J Med Internet Res 2013 Jun 07;15(6):e112 [FREE Full text] [doi: 10.2196/jmir.2435] [Medline: 23748147]
42. Ford E, Shepherd S, Jones K, Hassan L. Toward an ethical framework for the text mining of social media for health research: a systematic review. Front Digit Health 2020;2:592237 [FREE Full text] [doi: 10.3389/fdgth.2020.592237] [Medline: 34713062]

XSL•FO
RenderX

43. Lingard L. Beyond the default colon: effective use of quotes in qualitative research. Perspect Med Educ 2019 Dec;8(6):360-364 [FREE Full text] [doi: 10.1007/s40037-019-00550-7] [Medline: 31758490]
44. Morse JM. Critical analysis of strategies for determining rigor in qualitative inquiry. Qual Health Res 2015 Sep;25(9):1212-1222. [doi: 10.1177/1049732315588501] [Medline: 26184336]
45. Birtel MD, Wood L, Kempa NJ. Stigma and social support in substance abuse: implications for mental health and well-being. Psychiatry Res 2017 Jun;252:1-8. [doi: 10.1016/j.psychres.2017.01.097] [Medline: 28237758]
46. Snoek A, Levy N, Kennett J. Strong-willed but not successful: the importance of strategies in recovery from addiction. Addict Behav Rep 2016 Dec;4:102-107 [FREE Full text] [doi: 10.1016/j.abrep.2016.09.002] [Medline: 27975079]
47. Stokes M, Schultz P, Alpaslan A. Narrating the journey of sustained recovery from substance use disorder. Subst Abuse Treat Prev Policy 2018 Sep 27;13(1):35 [FREE Full text] [doi: 10.1186/s13011-018-0167-0] [Medline: 30261888]
48. De Crescenzo F, Ciabattini M, D'Alò GL, De Giorgi R, Del Giovane C, Cassar C, et al. Comparative efficacy and acceptability of psychosocial interventions for individuals with cocaine and amphetamine addiction: a systematic review and network meta-analysis. PLoS Med 2018 Dec 26;15(12):e1002715 [FREE Full text] [doi: 10.1371/journal.pmed.1002715] [Medline: 30586362]
49. Howlett N, García-Iglesias J, Bontoft C, Breslin G, Bartington S, Freethy I, et al. A systematic review and behaviour change technique analysis of remotely delivered alcohol and/or substance misuse interventions for adults. Drug Alcohol Depend 2022 Oct 01;239:109597 [FREE Full text] [doi: 10.1016/j.drugalcdep.2022.109597] [Medline: 35963209]
50. Webb L, Clayson A, Duda-Mikulin E, Cox N. ' I' m getting the balls to say no': trajectories in long-term recovery from problem substance use. J Health Psychol 2022 Jan;27(1):69-80 [FREE Full text] [doi: 10.1177/1359105320941248] [Medline: 32693631]
51. Gunn AJ, Sacks TK, Jemal A. "That's not me anymore": resistance strategies for managing intersectional stigmas for women with substance use and incarceration histories. Qual Soc Work 2018 Jul;17(4):490-508 [FREE Full text] [doi: 10.1177/1473325016680282] [Medline: 30116159]
52. Luoma JB, Platt MG. Shame, self-criticism, self-stigma, and compassion in acceptance and commitment therapy. Curr Opin Psychol 2015 Apr;2:97-101 [FREE Full text] [doi: 10.1016/j.copsyc.2014.12.016]
53. Luoma JB, Kohlenberg BS, Hayes SC, Bunting K, Rye AK. Reducing self-stigma in substance abuse through acceptance and commitment therapy: model, manual development, and pilot outcomes. Addict Res Theory 2008;16(2):149-165 [FREE Full text] [doi: 10.1080/16066350701850295] [Medline: 27746709]
54. Luoma JB, Kohlenberg BS, Hayes SC, Fletcher L. Slow and steady wins the race: a randomized clinical trial of acceptance and commitment therapy targeting shame in substance use disorders. J Consult Clin Psychol 2012 Feb;80(1):43-53 [FREE Full text] [doi: 10.1037/a0026070] [Medline: 22040285]
55. Corrigan PW, Rao D. On the self-stigma of mental illness: stages, disclosure, and strategies for change. Can J Psychiatry 2012 Aug;57(8):464-469 [FREE Full text] [doi: 10.1177/070674371205700804] [Medline: 22854028]
56. Courtney JB, Russell MA, Conroy DE. Acceptability and validity of using the BACtrack skyn wrist-worn transdermal alcohol concentration sensor to capture alcohol use across 28 days under naturalistic conditions - a pilot study. Alcohol 2023 May;108:30-43 [FREE Full text] [doi: 10.1016/j.alcohol.2022.11.004] [Medline: 36473634]
57. Swendeman D, Sumstine S, Aguilar E, Gorbach PM, Comulada WS, Gelberg L. Feasibility and acceptability of mobile phone self-monitoring and automated feedback to enhance telephone coaching for people with risky substance use: the QUIT-mobile pilot study. J Addict Med 2021 Apr 01;15(2):120-129 [FREE Full text] [doi: 10.1097/ADM.0000000000000707] [Medline: 32732684]
58. Park A, Conway M, Chen AT. Examining thematic similarity, difference, and membership in three online mental health communities from Reddit: a text mining and visualization approach. Comput Human Behav 2018 Jan;78:98-112 [FREE Full text] [doi: 10.1016/j.chb.2017.09.001] [Medline: 29456286]
59. Chen AT. Exploring online support spaces: using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups. Patient Educ Couns 2012 May;87(2):250-257. [doi: 10.1016/j.pec.2011.08.017] [Medline: 21930359]
60. Cesare N, Grant C, Nsoesie EO. Understanding demographic bias and representation in social media health data. In: Companion Publication of the 10th ACM Conference on Web Science. 2019 Presented at: WebSci '19; June 30-July 3, 2019; Boston, MA. [doi: 10.1145/3328413.3328415]

Abbreviations

BCT: behavior change technique
ML: machine learning
RQ: research question
SUD: substance use disorder
TTM: transtheoretical model

XSL•FO
RenderX

Edited by A Mavragani; submitted 21.02.24; peer-reviewed by A Davies, D Podder; comments to author 13.08.24; revised version received 22.09.24; accepted 18.12.24; published 26.03.25.

Please cite as:
Chen AT, Wang LC, Johnny S, Wong SH, Chaliparambil RK, Conway M, Glass JE
Stigma and Behavior Change Techniques in Substance Use Recovery: Qualitative Study of Social Media Narratives
JMIR Form Res 2025;9:e57468
URL: https://formative.jmir.org/2025/1/e57468
doi:10.2196/57468
PMID:

©Annie T Chen, Lexie C Wang, Shana Johnny, Sharon H Wong, Rahul K Chaliparambil, Mike Conway, Joseph E Glass. Originally published in JMIR Formative Research (https://formative.jmir.org), 26.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

User Experience With a Personalized mHealth Service for Physical Activity Promotion in University Students: Mixed Methods Study

Silke Wittmar[1], MSc; Tom Frankenstein[1], MSc; Vincent Timm[2], MA; Peter Frei[3], PhD; Nicolas Kurpiers[3], PhD; Stefan Wölwer[2], MA; Axel Georg Meender Schäfer[1], PhD

[1]Faculty of Social Work and Health, HAWK University of Applied Sciences and Arts Hildesheim/Holzminden/Göttingen, Hildesheim, Germany
[2]Faculty of Design, HAWK University of Applied Sciences and Arts Hildesheim/Holzminden/Göttingen, Hildesheim, Germany
[3]Institute for Sport Science, University of Hildesheim, Hildesheim, Germany

Corresponding Author:
Silke Wittmar, MSc
Faculty of Social Work and Health
HAWK University of Applied Sciences and Arts Hildesheim/Holzminden/Göttingen
Goschentor 1
Hildesheim, 31134
Germany
Phone: 49 5121881369
Email: silke.wittmar@hawk.de

Abstract

Background: Regular physical activity (PA) is known to offer substantial health benefits, including improved physical fitness, reduced risk of disease, enhanced psychological well-being, and better cognitive performance. Despite these benefits, many university students fail to meet recommended PA levels, risking long-term health consequences.

Objective: This study evaluated the user experience (UX) of futur.move, a digital intervention aimed at promoting PA among university students. The service delivers personalized, evidence-based content to foster sustained engagement in PA.

Methods: A mixed methods approach was used to evaluate the prototype of futur.move. UX assessments included on-site and online user tests, standardized questionnaires, and online focus groups. A total of 142 university students participated, with 23 joining additional focus groups. Each participant tested the service for 30 minutes. Quantitative data were collected using the User Experience Questionnaire and analyzed descriptively, followed by correlation analysis with variables such as PA level, age, gender, and experience with PA apps. Qualitative insights were gathered from transcribed focus group discussions and analyzed using content-structuring, qualitative content analysis. Quantitative findings were cross-validated with qualitative data.

Results: The UX received positive ratings across 4 User Experience Questionnaire scales (range –3 to +3; higher numbers indicate positive UX): attractiveness (median 1.67, IQR 1.04-2.17), perspicuity (median 1.5, IQR 0.5-2), stimulation (median 1.5, IQR 1-2), and novelty (median 1.25, IQR 0.5-2). Weak correlations were found between adherence to World Health Organization guidelines for PA and the perspicuity subscale (η=0.232, *P*=.04), and between age and the perspicuity (Kendall τb=0.132, *P*=.03) and stimulation subscales (Kendall τb=0.144, *P*=.02), and a moderate correlation was found between gender and the novelty subscale (η=0.363, *P*=.004). Critical feedback from focus group discussions highlighted issues with manual data entry. Qualitative findings aligned with the quantitative results, emphasizing students' appreciation for the personalized, diverse content and social networking features of futur.move.

Conclusions: futur.move demonstrates favorable UX and aligns with student needs, particularly through its personalized content and social features. Improvements should focus on reducing manual data entry and enhancing feature clarity, particularly for the features "your condition" and "goal setting." While correlations between UX ratings and demographic variables were weak to moderate, they warrant further investigation to better address the diverse target audience. The feedback from the students serves as a basis for further adapting the service to their needs and expectations. Future work will involve coding an advanced prototype and conducting a longitudinal study to assess its impact on PA behavior and sustained engagement.

(JMIR Form Res 2025;9:e64384) doi:10.2196/64384

XSL•FO
RenderX

KEYWORDS

usability testing; health promotion; exercise; smartphone app; mHealth; physical activity; user experience; user; university student; undergraduate; college; student; mixed methods; physical fitness; digital intervention; mobile health; promote; engagement; mobile phone

Introduction

The words commonly attributed to Hippocrates, "walking is humanity's best medicine," are more relevant than ever. In an era dominated by sedentary lifestyles, university students could greatly benefit from incorporating more physical activity (PA) into their daily routines.

Regular PA offers many important health benefits, including improved muscular and cardiorespiratory fitness, reduced risk of diseases, and enhanced psychological well-being [1-3]. PA also positively affects sleep and cognitive performance, including academic achievements and attention [1,3-5]. PA can also improve social skills and self-confidence [6].

The World Health Organization (WHO) recommends at least 150 minutes of moderate- to vigorous-intensity aerobic PA or at least 75 minutes of vigorous-intensity aerobic PA and muscle-strengthening activities of moderate to vigorous intensity for all major muscle groups on 2 or more days a week for adults [3]. Despite the benefits of regular PA, only a minority of young adults, including university students, meet the recommended PA levels, with 81% of adolescents [7] and 28% of adults worldwide and 42% in high-income Western countries [8] failing to adhere to these guidelines. Furthermore, studies show that students' PA levels drop significantly during examination periods and that they are more active during weekdays than on weekends [9-11]. As students undergo significant life transitions, their lifestyle habits, including PA patterns, can persist into adulthood if not addressed [12,13]. Therefore, universities face the challenge of supporting students in adopting more active lifestyles [14].

Promoting PA through mobile health (mHealth) services holds great potential, especially as students are frequent users of technology [15]. mHealth interventions are preferred by young people over traditional face-to-face approaches [16]. However, many current apps often fail to incorporate evidence-based recommendations for PA promotion, with developers prioritizing marketing strategies over addressing real user needs [17,18].

Positive user experience (UX) influences user satisfaction, promoting sustained product use and increasing recommendations [19]. "To create successful products or services it is necessary to ensure that the product has a sufficiently high user experience" [20]. UX is defined as "...a person's perceptions and responses that result from the use or anticipated use of a product, system or service. Users' perceptions and responses include the users' emotions, beliefs, preferences, perceptions, comfort, behaviors, and accomplishments that occur before, during and after use" [21]. Evaluating UX early in the prototyping phase allows developers to incorporate user feedback and refine features to meet diverse user needs [20,21]. In the context of PA services, successful interventions include behavior change techniques, such as self-monitoring, goal setting, behavioral feedback, social support, and social comparison, to effectively increase PA [22-26]. The challenge remains to maintain long-term user engagement with mHealth interventions [16]. Numerous studies have identified poor functionality and technical issues as barriers to engagement with mobile interventions, while design features such as personalization, social interaction, and customized feedback enhance user engagement [16]. Personalization is crucial, as students' PA needs vary widely [6], and one-size-fits-all approaches are inadequate [27-29].

This study aims to comprehensively evaluate the UX of the futur.move prototype, an mHealth service designed to enhance PA among university students. Specifically, this study will focus on the following UX aspects: (1) attractiveness—whether users like the service, (2) perspicuity—whether the service is easy to learn, (3) stimulation—whether using the service is exciting and motivating, and (4) novelty—whether the service is innovative and creative. The central research question is: "How do students evaluate the UX of the futur.move prototype?" Of particular interest are the specific improvements students suggest to optimize implementation.

Methods

Study Design

This study design follows a sequential approach with a series of cross-sectional surveys using structured online questionnaires with qualitative focus group discussions to comprehensively evaluate the futur.move prototype. On-site tests were conducted with a large participant group at the university, while smaller online sessions were held via Zoom (Zoom Video Communications Inc).

This study report follows the CHERRIES (Checklist for Reporting Results of Internet e-Surveys) reporting guideline for reporting web-based surveys [30] (Multimedia Appendix 1).

futur.move

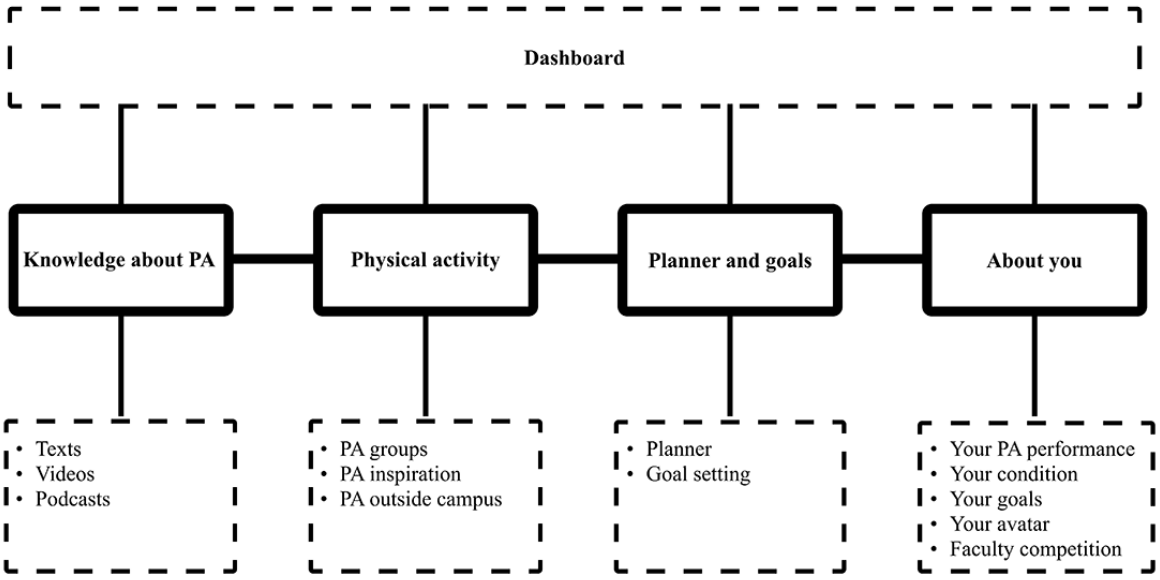
futur.move is a digital service that aims to improve students' PA through a smartphone app and website in the form of a progressive web application. According to a definition of service design, a service should be useful, usable, and desirable for the customer, while being effective, efficient, and unique for the provider. Services include interactions and interfaces between the customer and the provider [31].

The development of futur.move follows the holistic framework for the development of eHealth technologies [32] and includes 5 phases: context analysis, design, UX evaluation, programming, and long-term implementation. Currently, the prototype is in the UX evaluation phase (phase 3). The design is informed by a literature review, 2 student creative workshops (n=13 participants), and social cognitive theory [33]. Studies report that PA interventions are more successful when they are theory

XSL•FO
RenderX

based [34,35]. Components of social cognitive theory such as health-related knowledge, individual goal setting, and social support are transferred into the functions of the app to support users in successfully and sustainably increasing their PA. An interdisciplinary team of UX designers and health scientists developed the app iteratively in the collaborative design tool Figma (Figma Inc, 2022) with active involvement from the target audience. Figure 1 shows the content structure of futur.move features.

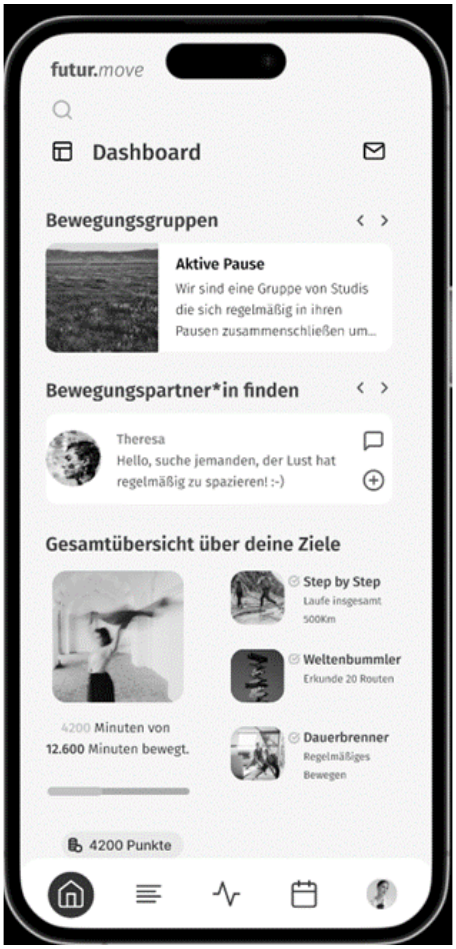
Figure 1. Structural overview of the features and content offered by the futur.move mHealth service. PA: physical activity. mHealth: mobile health.



After registration and reporting PA habits via the European Health Interview Survey–Physical Activity Questionnaire (EHIS-PAQ) [36], users access personalized content across five areas: (1) "dashboard," (2) "knowledge," (3) "physical activity," (4) "planner and goals," and (5) "about you" (Figure 1). A messaging function is also available. In the "knowledge" section, users access PA-related information through various formats such as text, videos, and podcasts. The "physical activity" section enables users to find inspiration for PA, join or create PA groups, and use links to find PAs outside the campus. "Planner and goals" offers a calendar view and goal-setting features, while "about you" manually tracks PA habits and, goals, and provides a faculty competition. By participating in PA, users can earn points for their faculties and win rewards such as a faculty party. Figure 2 shows the personalizable dashboard of futur.move.

XSL•FO
RenderX

Figure 2. Screenshot of the customizable dashboard of the futur.move mHealth service as viewed on a smartphone. mHealth: mobile health. An English version of the dashboard is available in Multimedia Appendix 2.



Participants and Recruitment

User testing was conducted from October 27 to December 21, 2023, at University of Applied Sciences and Arts Hildesheim/Holzminden/Göttingen (HAWK; N=6300 students in winter 2022/2023).

The convenience sample for this study was recruited from the entire student body of HAWK. For the on-site sessions, a high number of participants was targeted to enable valid statistical analyses. Therefore, researchers reached out directly to professors from all 3 university campuses and various faculties, requesting their support in implementing this study within their seminars. On-site sessions were finally held in 4 seminars across different faculties and locations. Participation in this study was integrated into the respective seminars; however, it remained entirely voluntary for all participants, ensuring that their consent was informed and given freely without any obligation.

For the online sessions with focus group discussions, all students at HAWK were invited via email, flyers, and the university's social media channels. Flyers were posted in high-traffic areas on the university campuses, such as the cafeteria and the library. Additionally, first-year students from the Faculty of Social Work and Health were personally approached at a welcome event for new students. Due to low participation, students were also approached at a campus café. The latter method of reaching out to students proved to be the most effective. A link distributed through the mentioned media allowed registration via HAWK's online learning tool, followed by contact with the project team, provision of information, and signing of the consent form. Only those who returned the signed form participated. To participate in both the on-site and online data collection, inclusion criteria were enrollment and no prior involvement in futur.move.

Recommended sample sizes for quantitative UX testing to estimate key performance parameters range from 10 to 1689 participants, depending on the margin of error. "A sample size of 115 at 90% confidence can detect a 10% significant difference" [37]. Therefore, we aimed to recruit a minimum of 65 participants for the survey to achieve 90% confidence. For qualitative UX testing, it is recommended to conduct focus groups with 4 to 37 participants to identify usability issues. A sample size of 18 is needed to detect a 10% insight occurrence [37]. Consequently, we aimed to recruit at least 18 participants for the focus group discussions. Given the maximum group size of 6 to 8 participants per focus group [38], we planned for a minimum of 3 groups.

Ethical Considerations

Ethical approval was obtained from the HAWK ethics committee on September 29, 2023. All participants gave written informed consent before participation, agreeing to the use of deidentified data for secondary analyses and the publication of results. Data were collected anonymously, and no personally identifiable information was included in the final dataset.

XSL•FO
RenderX

Participation was voluntary. Participants could withdraw at any time without justification. Data were stored on the university's server for secure keeping until January 10, 2033, and will then be deleted by the researchers. Throughout this study, only deidentified data were used, and access was restricted solely to the research team. Participants in the online sessions received a €10 (US $10.85) voucher for an online retailer.

Procedures

On-Site Data Collection

Quantitative data were collected on-site during the seminars with a project team member present. Participants assumed the role of "Kim," a fictitious user of futur.move, to understand prestored data and experience all app features. After an introduction and onboarding, they received a task sheet with 6 tasks to independently test the service for approximately 30 minutes (eg, finding exercise information, Multimedia Appendix 3). They accessed the prototype via a device (eg, laptop or tablet).

After completing each task, participants provided initial feedback on the task sheet and then completed the online questionnaire. This included sociodemographic data, PA behavior (EHIS-PAQ) [36], and usability ratings using the User Experience Questionnaire (UEQ) [39]. The survey, hosted via SoSci Survey (version 3.5.02), presented the items to each participant in the same order, allowed participants to change their answers using a back button, and omitted irrelevant questions. There was no check for completeness before submission. No cookies were used. The online survey consisted of 22 pages, including the introduction and conclusion. The number of items on each page varied. Each participant used a unique ID to merge task sheet responses with the online survey for analysis. The ID was randomly assigned on each task sheet. Each code could only be used once. On-site sessions lasted approximately 60 minutes per session.

A pretest was conducted to ensure clarity and feasibility. Four students tested the task sheets, while 2 project members and 2 individuals from the target group tested the online survey. Feedback led to revisions of both instruments.

Online Data Collection

Online data collection was conducted via the Zoom videoconferencing tool with 2 project team moderators. Participants registered for user tests, and sessions were scheduled based on availability. Participation was voluntary, and participants could withdraw at any time.

The first part of the online session mirrored the on-site data collection. The moderators reviewed task sheets and survey feedback to refine focus group discussions, gathering targeted qualitative insights beyond the prepared questions.

Measures

Demographics

As part of the online survey, participants provided their age, gender, height, weight, faculty affiliation, and whether they had used apps to improve PA before. Additionally, the EHIS-PAQ [36] was used to collect standardized data on students' PA. The EHIS-PAQ is a domain-specific instrument for measuring (1) work-related PA, (2) transport-related PA, and (3) leisure-time PA. It distinguishes between "aerobic" and "muscle-strengthening" PA and allows the assessment of compliance with health-promoting PA recommendations [36].

About UX

UX was assessed using the UEQ [39]. The UEQ is a quantitative instrument for assessing the UX of interactive products. It originally consisted of 6 subscales with a total of 26 items with classic usability aspects (ie, efficiency, perspicuity, or dependability) and UX aspects (ie, originality or stimulation). The authors state that entire subscales can be omitted if they are irrelevant to the investigation [40]. Consequently, the "efficiency" and "dependability" scales from the original questionnaire were not included in the analysis, as no valid results could be expected for these scales at the stage of development of the service. In this study the items of the four subscales (1) attractiveness (do the users like the service or not?), (2) perspicuity (is it easy to learn how to use the service?), (3) stimulation (is the use of the service exciting and motivating?), and (4) novelty (is the service innovative and creative?) were used. The attractiveness scale is represented by 6 items, while the other scales are represented by 4 items each. The items are semantic differentials, for example, attractive and unattractive. They can be rated on a 7-point scale from –3 (horribly bad) to +3 (extremely good). Scale scores are average item ratings across participants. Scores between –0.8 and 0.8 represent a neutral evaluation of the scale, scores>0.8 represent a positive evaluation, and scores<0.8 represent a negative evaluation [40].

Evaluation of Individual Features

On the task sheet for testing futur.move, students had the opportunity to indicate how much they liked each feature of futur.move. They could tick a value from very good (1) to poor (5) on a 5-point Likert scale. Within this framework, the students rated the five features: (1) faculty competition, (2) your condition, (3) knowledge, (4) PA group (hiking), (5) PA inspirations, and (6) goal setting. Additionally, the online questionnaire asked students how they would rate futur.move overall on a 5-point Likert scale from 1 (very good) to 5 (poor).

Focus Groups

The focus groups were conducted to qualitatively summarize the quantitative feedback from the students. The questions in the focus groups were, therefore, related to the quantitative part of the respective user tests. Follow-up questions were individually prepared by the moderators by reviewing the anonymized data from the user tests. The focus group structure followed the recommended procedure in empirical social science [41]. The duration of the focus groups was approximately 120 minutes, and they were conducted by VT and SW, who moderated the discussion. The discussion was recorded using the recording feature of Zoom.

The first question in the focus groups was what participants remembered most about futur.move. Following this, an overview of the UEQ scales was briefly presented. Based on the UEQ scores, questions about divergences in the scoring of individual

XSL•FO
RenderX

scales were discussed. Students were asked to describe their quantitative assessments in more detail, and discrepancies in the service ratings from the task sheet were examined. Finally, there was an opportunity for participants to report any further requirements. Discussions were recorded with participants' consent for later transcription and stored on the university's server.

Data Analyses

Quantitative Data Analyses

The quantitative data from the questionnaires were analyzed using IBM SPSS (version 27; IBM Corp).

The UEQ Data Analysis Tool (version 12) transformed the UEQ data, which were then merged in SPSS using participant IDs. Meta-information was added, and variable labeling was adjusted, including scale levels [42]. Data cleansing involved checking for incorrect or implausible information and marking missing values with the code 999. Questionnaires with missing values were included in the analysis.

For the UEQ data, items on the subscales were analyzed for divergent ratings. Cases were recommended for exclusion if at least 3 subscales were conspicuous or if more than 15 of 26 items had identical ratings. No records were excluded based on plausibility, but 8 records from the task sheets were excluded due to missing online survey data.

As the UEQ data were not normally distributed, the median and IQR were calculated. Nonetheless, scale means, variances, and Cronbach α were calculated using the UEQ Data Analysis Tool, to use the provided benchmark for categorizing scale scores against data from 468 studies on various products, rating them from "poor" to "excellent" [40].

In addition to the descriptive data analysis, a correlation analysis was conducted to explore the relationships between the UX evaluation of the futur.move, as measured by the UEQ, and various demographic and usage-related factors. These factors included compliance with WHO guidelines for PA, age, gender, and prior experience with apps aimed at promoting PA.

Associations were analyzed using the η and Pearson chi-square tests for metric and nominal-scaled data. For associations between 2 metric variables, Kendall τb was used. The significance level for all tests was set at $\alpha \leq .05$.

Qualitative Data Analyses

The focus group recordings were automatically transcribed verbatim using the tool noScribe (version 0.4.1). The transcripts were then manually corrected and anonymized, with participant names replaced by the IDs matching those from the quantitative survey, allowing data linkage.

Two project members analyzed the transcripts using MAXQDA 2022 (VERBI-Software. Consult. Sozialforschung GmbH 2021) following content-structuring, qualitative content analysis [43]. The data were analyzed based on the question: "How do students evaluate the futur.move prototype?" with the 4 UEQ scales (attractiveness, perspicuity, stimulation, and novelty) used as deductive categories. Further inductive categories could be formed based on the material, but no additional categories emerged. Subcategories of the main categories were formed inductively during the analysis.

Results

Quantitative Findings

Demographics

A total of 142 people (mean age 24, SD 4 years; female: n=105, 74%; male: n=33, 23%) participated in the user tests. Change to: A total of 142 people (mean age 24, SD 4 years; female: n=105, 75%; male: n=33, 23%) participated in the user tests. Of these, 23 also took part in the focus groups. Most participants were students from the Faculty of Social Work and Health (n=60, 42%). Approximately a third of the students had prior experience using an app to promote PA (n=49, 35%). The majority of students (n=87, 61%) met the WHO guideline of practicing more than 150 minutes of moderate aerobic activity per week, while 49% (n=69) met the guideline of doing muscle-strengthening exercises at least 2 days a week. Overall, 40% (n=57) of participants met the WHO guidelines for PA, fulfilling both criteria. The mean BMI of the students was 23 (SD 4), which falls at the upper end of the normal weight range. Table 1 summarizes the information provided by the students.

XSL•FO
RenderX

Table 1. Demographics of university students participating in user tests conducted from October 27 to December 21, 2024.

Characteristics	Participants
Age (years), mean (SD)	24 (4)
Gender, n (%)	
Female	105 (74)
Male	33 (23)
Diverse	1 (1)
Not specified	3 (2)
Faculty, n (%)	
Social work and health	60 (42)
Design	26 (18)
Engineering and health	22 (16)
Architecture, engineering, and conservation	20 (14)
Management, social work, and construction	13 (9)
Not specified	1 (1)
"Do you already use digital services to promote physical activity?" n (%)	
Yes	49 (35)
No	74 (52)
Not specified	19 (13)
BMI (kg/m^2), mean (SD)	23 (4)
Fulfillment WHO[a] guideline >150 minutes of moderate aerobic activity, n (%)	
Yes	87 (61)
No	55 (39)
Fulfillment WHO guideline ≥2 days/week muscle strengthening, n (%)	
Yes	69 (49)
No	73 (51)
Fulfillment WHO guideline for physical activity overall, n (%)	
Yes	57 (40)
No	85 (60)

[a]WHO: World Health Organization.

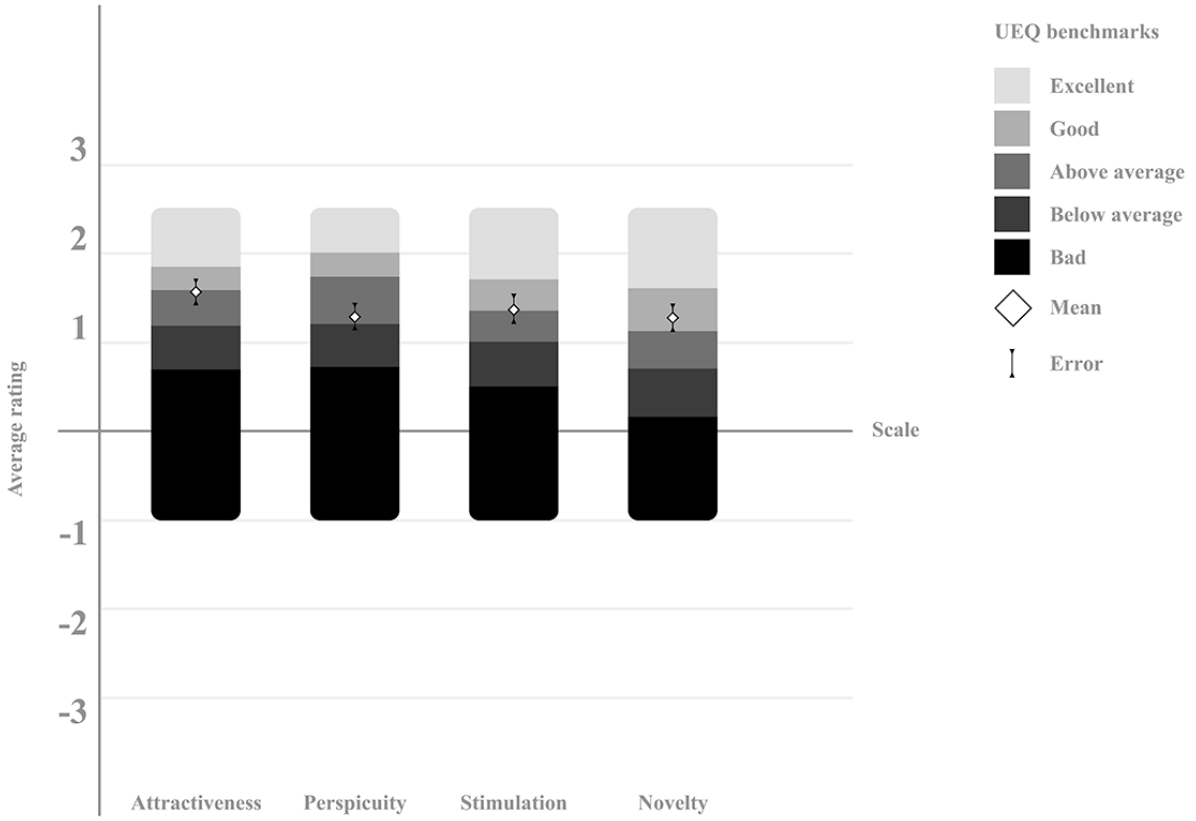
futur.move UX

The UX of futur.move is indicated by the average score on each respective scale. As the data are not normally distributed, the median, IQR, and the minimum and maximum range are also reported. Attractiveness received a median score of 1.67 (IQR 1.04-2.17) and a mean score of 1.55 (SD 0.91) on a scale from –3 to +3. Perspicuity was rated with a median of 1.5 (IQR 0.5-2) and a mean of 1.28 (SD 0.97), stimulation with a median of 1.5 (IQR 1-2) and a mean of 1.37 (SD 1.03), and novelty with a median of 1.25 (IQR 0.5-2) and a mean of 1.27 (SD 0.98). Positive ratings are indicated by scores greater than 0.8 (UEQ manual).

The UEQ benchmark refers to the mean scores and SD. According to the UEQ benchmark results shown in Figure 3, futur.move falls into the "good" category for hedonic quality (stimulation and novelty scales). The attractiveness and perspicuity scales are in the "above average" category. A score is considered "good" if it is among the top 25% in the benchmark dataset and "above average" if 25% of other products score better. Therefore, the overall quality of futur.move can be classified as good according to the UEQ. However, none of the scale scores fall into the "excellent" category, which includes the top 10% of results, indicating room for improvement.

XSL•FO
RenderX

Figure 3. UX evaluation results of the futur.move mHealth service by study participants using the UEQ, compared to a dataset of 468 product reviews, analyzed with the UEQ Data Analysis Tool Version 12. Error bars represent the 95% CI. mHealth: mobile health; UEQ: User Experience Questionnaire; UX: user experience.

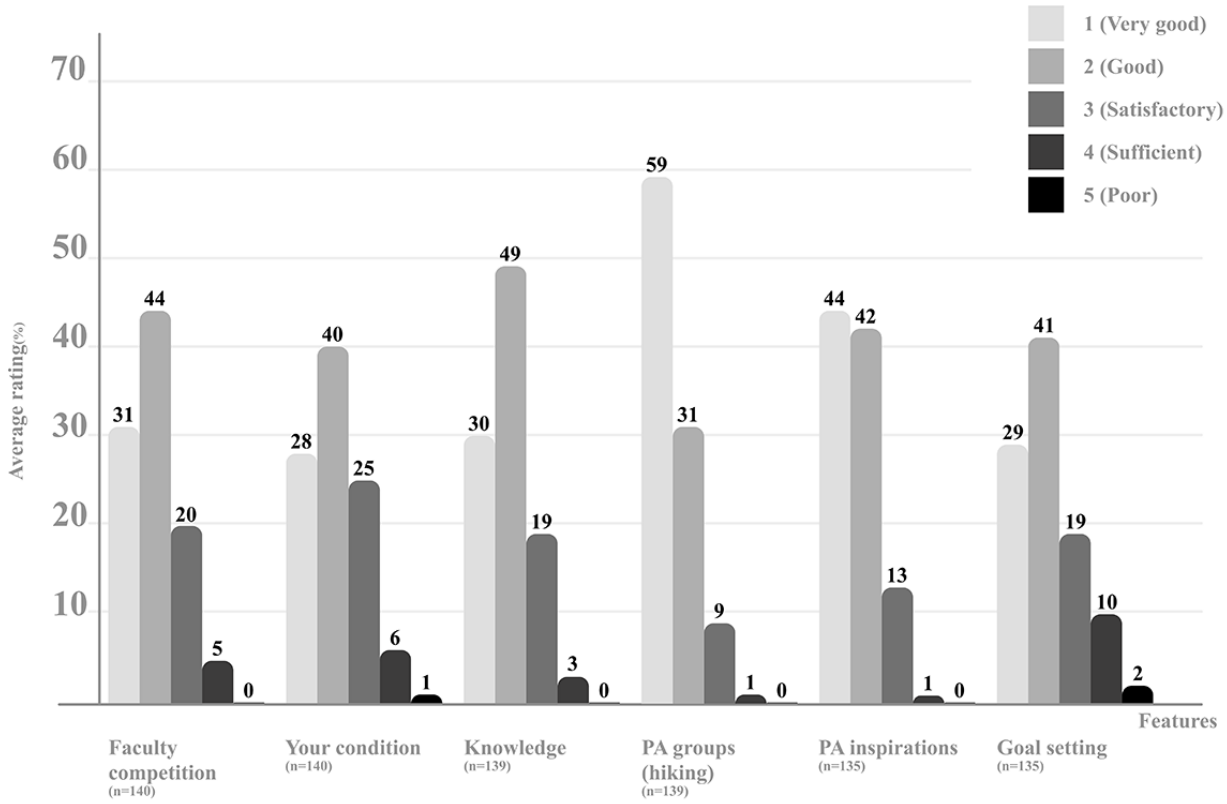


Evaluation of Individual Features

When rating the individual features of futur.move, at least 68% (95/142) of participants rated all features with a score of 2 or better (Figure 4). The PA features (PA groups and PA inspirations) received the highest scores, with 90% (126/139) and 86% (116/135) of participants rating them with a score of 2 or better, respectively. Only a small number of respondents rated 2 features with a score of 5 (very poor): “your condition” (1/142, 0.7%) and “goal setting” (3/142, 2%).

XSL•FO
RenderX

Figure 4. Average ratings of individual features of the futur.move mHealth service by study participants. Each feature was evaluated on a scale from 1 (very good) to 5 (poor). mHealth: mobile health; PA: physical activity.



Overall Grade

Overall, 105 (74%) of 142 participants gave futur.move an overall grade of 2 (n=97, 68%) or better (n=8, 6%). No grade lower than 4 was awarded (grade of 3: n=29, 20%; grade of 4: n=8, 6%). The average grade for the overall evaluation of futur.move was 2.3 (German grading scale, where 1 is the best grade and 6 the worst); a grade of 2 corresponds to a "good" rating.

Correlation Analysis

A summary table of the correlation analysis results can be found in Multimedia Appendix 4. To analyze the direction of correlations calculated with η, corresponding bar charts were used. These are also included in Multimedia Appendix 4.

The correlation analysis reveals a weak to moderate relationship between adherence to WHO guidelines for PA and the UX evaluation using the UEQ subscale perspicuity (η=0.232, *P*=.04). Adherence to WHO guidelines is associated with less favorable UX ratings.

Additionally, 2 weak but significant relationships were identified between age and the UEQ subscales perspicuity (Kendall τb=0.132, *P*=.03) and stimulation (Kendall τb=0.144, *P*=.02). Higher age was associated with more favorable evaluations of futur.move on both subscales.

A significant moderate relationship was also found between gender and the UEQ subscale novelty (η=0.363, *P*=.004), with women providing higher ratings for novelty compared to men.

No significant influence was observed for the variable prior experience with apps for PA on the UX evaluation using the UEQ.

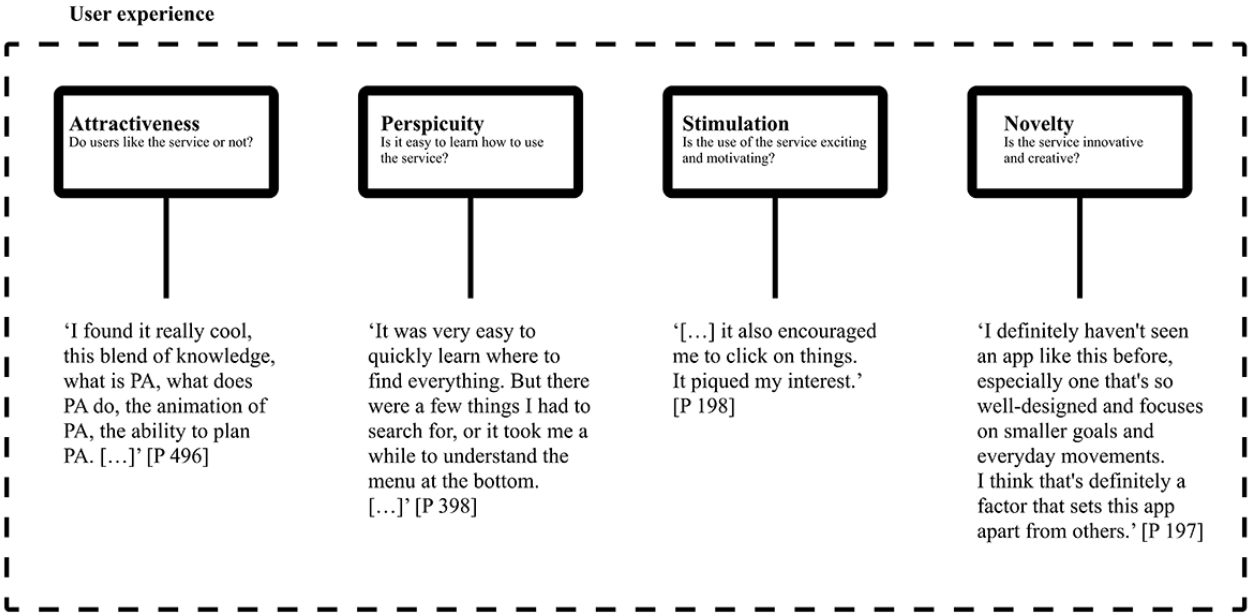
Qualitative Findings

Overview

Four focus group discussions were conducted with a total of 23 participants. The main categories corresponded to the four scales of the UEQ: (1) attractiveness, (2) perspicuity, (3) stimulation, and (4) novelty. Figure 5 shows an overview of the UX dimensions with an exemplary anchor quote representing each category. The quotes mentioned in the text can be found in Multimedia Appendix 5.

XSL•FO
RenderX

Figure 5. Overview of user experience categories for the futur.move mHealth service from the qualitative content analysis of 4 focus group discussions, structured according to UEQ subscales with illustrative anchor quotations. mHealth: mobile health; P: participant; UEQ: User Experience Questionnaire.



Attractiveness

Most participants liked futur.move overall and had fun when using it (quote 1 in Multimedia Appendix 5). Some participants were particularly attracted by the variety of features (quotes 2 and 3 in Multimedia Appendix 5). However, few individuals found the platform to be less attractive. According to the participants, this was due to enduring a recent injury, living away from the university (quote 4 in Multimedia Appendix 5), or undertaking part-time studies (quote 5 in Multimedia Appendix 5). Many participants highlighted the networking opportunities with other students and the links to PA opportunities in the region (quotes 3 and 6 in Multimedia Appendix 5).

Most participants thought the "physical activity inspirations" were particularly attractive because of their diversity (quote 7 in Multimedia Appendix 5), perceived as a variety of offers for different needs. One participant thought it was particularly positive that the PA options were low-threshold; that is, they did not always entail sport but general PA such as a walk around the city (quote 8 in Multimedia Appendix 5). Most participants found the group activities attractive, even though other perspectives were also considered (quotes 9 and 10 in Multimedia Appendix 5). One person emphasized that the group recommendations were also a good way for first-year students to make contacts upon arrival in a new city (quote 10 in Multimedia Appendix 5). When using a digital service, the physical presence was also discussed. One participant particularly liked that futur.move offers the opportunity to meet in person (quote 11 in Multimedia Appendix 5). Most participants thought the "knowledge" feature was attractive. In particular, the variety of media was mentioned as appealing (quote 12 in Multimedia Appendix 5). However, details such as the structure of the texts also contributed to the perceived attractiveness (quote 13 in Multimedia Appendix 5). Many of the participants liked the "goal setting" feature in general. One of the reasons given was the possibility to set small, realistic goals (quote 14 in Multimedia Appendix 5). Several focus groups discussed appropriate target periods (quote 15 in Multimedia Appendix 5) and manual data entry (quote 16 in Multimedia Appendix 5).

Some of the participants thought it was attractive to compete with others in a faculty competition (quote 17 in Multimedia Appendix 5), while others would rather avoid such competition (quote 18 in Multimedia Appendix 5). One person cited a lack of connection to their own faculty as a reason (quote 19 in Multimedia Appendix 5).

In terms of the attractiveness of individual features, the option to design one's own avatar elicited mixed feedback from participants. While 1 person thought it was appealing (quote 20 in Multimedia Appendix 5), another did not find the choice appealing at all (quote 21 in Multimedia Appendix 5).

Most participants were very positive about the design of futur.move (quote 22 in Multimedia Appendix 5). One participant particularly emphasized the design of the "knowledge" feature, which was perceived as clean (quote 23 in Multimedia Appendix 5). One person had a good feeling about the overall service but perceived a design break in the "your condition" feature (quote 24 in Multimedia Appendix 5). The color scheme of the service was rarely discussed, and apart from 1 comment on the color scheme in the "your condition" feature, was perceived as positive (quote 25 in Multimedia Appendix 5). The choice of images was particularly appealing to some respondents (quote 26 in Multimedia Appendix 5).

Perspicuity

Most participants described the platform as clear, tidy, and not too packed. The font, large headings, and organized layout contributed to this (quote 27 in Multimedia Appendix 5). Most

XSL•FO
RenderX

features were also found to be easy to understand. Most participants reported that they found their way around quickly, although it took some time to get used to the system in the beginning (quote 28 in Multimedia Appendix 5). While certain aspects of the tasks required exploration, participants found the navigation easy to learn and were able to use it quickly (quote 29 in Multimedia Appendix 5).

In contrast, not all participants understood the diagrams in the "your condition" section or the points system for the "faculty competition" (quotes 30 and 31 in Multimedia Appendix 5). In 1 focus group, the comprehensibility of the navigation buttons in particular was discussed and perceived as somewhat difficult to comprehend (quote 32 in Multimedia Appendix 5).

Stimulation

Overall, the design and content of the service caught the interest of most participants and stimulated them to explore the content (quote 33 in Multimedia Appendix 5). In this context, individual feedback raised the question about the long-term motivation to use the service (quote 34 in Multimedia Appendix 5).

The "physical activity" feature received a lot of positive feedback, with the variety of activities seeming to be a key advantage. Not all activities appealed to everyone (quote 35 in Multimedia Appendix 5), but there was something for everyone (quote 7 in Multimedia Appendix 5). Most participants were particularly positive about the possibility of using futur.move to become physically active in the first place or even to initiate groups themselves (quote 36 in Multimedia Appendix 5). The "knowledge" feature also appealed to many participants, partly because of the variety of media and the design of the contributions (quote 37 in Multimedia Appendix 5). Many participants thought the "goal setting" feature was motivating, especially because small goals could be set and big goals could be broken down (quotes 38 and 39 in Multimedia Appendix 5). Opinions differed on the period length to achieve a target and whether the tool would be used during implementation (quotes 40 and 41 in Multimedia Appendix 5).

The "faculty competition" feature was discussed by participants. While some felt motivated by the competitive nature of the event (quote 42 in Multimedia Appendix 5), some participants did not find it appealing (quote 18 in Multimedia Appendix 5). Furthermore, the focus groups discussed the possibility of providing information about one's own mood or having PA data displayed in the "your condition" feature: while some found this feature helpful, others thought it was uninteresting or even problematic to confront them with their own performance (quotes 43, 44, and 45 in Multimedia Appendix 5). Designing an avatar was a less motivating feature of the service to use for some participants (quote 46 in Multimedia Appendix 5).

Novelty

Participants indicated that they were particularly attracted by the creative design of the service (quote 47 in Multimedia Appendix 5) and the "fresh" headings (quote 48 in Multimedia Appendix 5). Some participants thought the service was new (quote 49 in Multimedia Appendix 5). It was not the content itself that was described as new, but its combination as an infrastructure service (quote 50 in Multimedia Appendix 5). The fact that the content was already familiar to some participants from other contexts was also a reason for some to perceive the service as less new (quote 51 in Multimedia Appendix 5).

Discussion

Principal Findings

The evaluation of the UX of futur.move using the UEQ revealed generally positive results, with attractiveness receiving the highest ratings. Compared to the UEQ benchmark, all scores were at least "above average." However, all ratings still leave room for improvement. It has to be taken into account that despite rating options between –3 and +3, real-world applications typically produce lower values, with averages unlikely to exceed +2 or go below –2 [40]. "Your condition" and "goal setting" were less favorably rated, suggesting these areas are less engaging or useful for users. This feedback provides valuable insights for targeted enhancements to increase user satisfaction and engagement.

Qualitative results complemented the quantitative findings, reflecting the positive feedback on the service's content and design. Participants particularly valued the personalization features, as well as the social and the variety of PA options, which enhanced their overall enjoyment. However, critical feedback identified areas for improvement, notably the manual data entry process and difficulties in comparing progress with other users.

The correlation analysis provides additional insights into potential areas for improving UX, though the findings should be interpreted with caution. A weak to moderate correlation was found between adherence to WHO guidelines for PA and the UEQ subscale perspicuity, contradicting that individuals who meet WHO guidelines for PA provide higher UX ratings for futur.move using the UEQ. As none of the other subscales showed any associations with adherence to the WHO guideline, this is probably a chance finding, as the relationship cannot be explained conceptually.

Interestingly, older students rated the UX higher on perspicuity and stimulation compared to younger students. These findings point to a need for better understanding what motivates the core target group of students aged 20-30 years to engage with the service and how to make it more intuitive for them. Although age appears to have a weak influence overall, these insights can guide refinements to improve engagement with this age group.

The observation that women perceive futur.move as more novel than men suggests that women rate UX higher, specifically in the subscale novelty. This finding emphasizes the importance of tailoring innovative features to better engage men and other genders, ensuring that the service effectively appeals to the diverse target audience. Addressing these issues could significantly enhance usability and overall UX. The analysis of the qualitative data by PA, gender, and age showed no additional insights.

XSL•FO
RenderX

Comparison With Prior Work

Our study expands the existing literature by being to our knowledge the first to investigate the UX of an mHealth prototype in an early development phase using a combined qualitative and quantitative approach. It provides valuable insights into the strengths and weaknesses of the prototype, along with practical recommendations for further development.

Previous studies on UX in mHealth services aimed at promoting PA primarily focused on fully programmed apps in long-term studies, with a predominant emphasis on usability aspects [44-47]. For example, 1 study evaluated UX aspects using the UEQ without incorporating qualitative methods [48]. In contrast, our study integrates qualitative and quantitative UX methods in an early development phase to derive key insights into participatory development and its impact on UX.

The evaluation of **futur.move** revealed strengths in usability, personalization, and design, with the ease of use being particularly notable after a brief familiarization period. At the same time, weaknesses such as the need for manual data entry were identified, which could negatively affect long-term usage [16,29]. These findings align with studies emphasizing the importance of automated tracking features, intuitive navigation, and the critical role of personalization options [16,29,49].

The investigation of social comparison features revealed individual preferences. While some participants found the "faculty competition" motivating, others expressed a negative attitude toward competition, highlighting the diversity of user values. A systematic review recommends using competitive elements sparingly and in combination with other features [16]. The results of other studies suggest that intergroup competition approaches are beneficial [50,51].

Gender and age correlate with UX ratings. Women tended to evaluate the aspect of novelty more positively than men. One explanation could be that men are generally more frequent users of technology [52] and tend to score higher on measures of technology acceptance and competence [53,54]. This could explain why men perceive futur.move as less novel, as they may already be familiar with a wider range of similar offerings. Conversely, women often place greater emphasis on the design and user-friendliness of an application [55], which could positively influence their evaluation of its novelty. These findings highlight the importance of considering gender-specific preferences in UX design. For example, future development could focus on creating a balance between functionality and aesthetic appeal to meet the expectations of both genders. Additionally, efforts to better understand how men and women perceive and engage with novel features in mHealth apps could further optimize the UX of futur.move.

Research indicates that younger people are more likely to use smartphone apps for health-related purposes. For instance, as early as 2016, one-third of students surveyed in a study were already using mHealth apps [56]. Empirical evidence also shows that interest in technology tends to decrease with age among older adults [57]. Interestingly, in our study, older students rated the perspicuity and stimulation provided by futur.move more favorably. This could be attributed to differences in preferences, as older individuals may value ease of navigation and straightforward information over other features. While younger users generally engage more frequently with health apps, older adults may respond more positively to specific UX elements when tailored to their needs. Identifying these aspects should be a focus in future development.

Limitations

In this study, a convenience sample of students was used. Participants were not selected based on their PA habits or need for PA in the university context. Due to initial reluctance from students to participate, they were engaged through personal networks within the student community. Not all courses of study and faculties of the university were covered. This recruitment process has led to a strong representation of health students in the sample. It can be assumed that some participants are already aware of health-related issues and the importance of PA for health. Therefore, the generalizability of the results to the entire student community may be limited.

In this study, 74% of the participants were female, similar to other mHealth intervention studies aimed at improving PA among students [49,58,59]. A study found that female students are also more likely to engage in health promotion activities on campus than males [14]. The participants' age and BMI were similar to other studies [58-60]. Nevertheless, BMI has limitations as it does not account for body composition and can lead to health risk misestimation [61]. Self-reported height and weight can also cause BMI underestimation [62]. For this reason, BMI is not used in this study to refer to the students' physical health. The service does not provide recommendations to students based on BMI. Future evaluation of the service should consider possible underestimations of self-reported data, especially concerning the effectiveness of the service, and include objective measurement methods.

Due to the presentation of the prototype in Figma, not all of the developed content of the service could be fully experienced. As a click dummy, a clickable prototype with limited functionality, the ability to make all activities of the service tangible is limited. For example, push notifications, which will be part of the service, could not be evaluated in this format. The evaluation of activities within the service was indirect and could not be directly tested in vivo. As futur.move was evaluated in a cross-sectional survey, the results provide insights into how students perceived its use but do not give information on the actual use of the service. A longitudinal evaluation of the programmed service is needed to collect data on user behavior and the effects of the service on PA behavior and motivation in long-term use. However, this evaluation requires not only a programmed service but also its implementation in everyday campus life, even if only on a trial basis.

Implications

Future research should include analyses focused on the values and needs of younger students and men. These findings should then be incorporated into the further development of the service to better address and meet these groups' specific requirements.

The service's clarity, noted by most participants, confirms the successful layout of the application on which the usability of

XSL•FO
RenderX

the service was based. Nonetheless, feedback on comprehensibility, such as navigation buttons, should be addressed for usability. The manual data entry process requires further review with the target group, as it was negatively received and could impede long-term use. Data tracking through devices might offer a solution here. Further investigation is also needed for competitive elements such as the faculty challenge.

Despite some criticism, the overall evaluation of futur.move is positive. futur.move meets a substantial proportion of the diverse needs of the target group, underscoring the importance of adaptability.

The next step is to assess whether using futur.move actually improves students' PA. Thus, a longitudinal study should be carried out to assess real-world usability, long-term adherence, effectiveness in regards to PA, and other health-related outcomes and correlations with academic success.

Motivating students to participate in the development process remains challenging. Although 60% of surveyed students expressed interest in health programs [14], recruiting participants has been difficult. This suggests that a desire for the service does not necessarily lead to active participation in its design. Therefore, future studies should place greater emphasis on understanding the incentives and motivations that drive involvement.

Based on our findings, the UX evaluation methods used in this study could be adapted for other populations, such as individuals with chronic pain, obesity, or cardiovascular disease, to align with their needs and values in designing digital services or mobile apps that promote PA.

Conclusions

The results of this study provide relevant information on how to further develop futur.move as a multidimensional, tailor-made, and participatory mHealth service to enhance PA among university students in hybrid spaces. By emphasizing personalization—a crucial factor also identified in prior research that showed the limitations of one-size-fits-all approaches [49]—futur.move aligns well with the diverse needs of its target audience.

Overall, the UX evaluation yielded positive results for the prototype. Students particularly appreciated its personalized content, engaging social features, and thoughtful design elements. However, the UX evaluation, complemented by qualitative feedback from focus group discussions, identified 2 specific features that need significant improvement to better meet user expectations. Additionally, reducing manual data entry emerged as a critical area for optimization, as its current implementation may hinder adherence and long-term user engagement.

While correlations between UX ratings and demographic variables were found to be weak to moderate, they warrant further investigation to better address the diverse characteristics of the student population. These findings may provide essential guidance for future development steps.

Building on these results, the next phase will refine the prototype by addressing the identified areas for improvement and implementing it as a smartphone app. This app will be tested and evaluated within the university context to ensure it effectively supports students in increasing PA while maintaining a high standard of UX.

Acknowledgments

This study was funded by the "Stiftung Innovation in der Hochschullehre" (Innovation in Higher Education Foundation) under the funding line "Strengthening Higher Education through Digitalization" (grant FBM2020-VA-212-5-05380). The funders had no role in this study's design, collection, management, analysis and interpretation of data, writing of the report, and decision to submit the report for publication, and had no final authority on any of these activities. The authors would like to thank Roger Skarsten, PhD, for his support in the linguistic revision of this paper. We used ChatGPT-4o [63], developed by OpenAI (2025), to review the linguistic revisions of this paper.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

S Wittmar drafted the original version of this paper. TF carried out the quantitative analysis. S Wittmar and TF carried out the qualitative analysis. The graphics were created by VT and S Wittmar. S Wittmar, VT, PF, NK, S Wölwer, and AGMS were involved in the design of this study. All authors contributed to the drafting of the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CHERRIES for the futur.move mHealth service. mHealth: mobile health.
[DOCX File , 25 KB - formative_v9i1e64384_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Screenshot of the customizable dashboard of the futur.move mHealth service as viewed on a smartphone in English. mHealth: mobile health.
[PNG File , 828 KB - formative_v9i1e64384_app2.png]

Multimedia Appendix 3
Task sheet for user testing to evaluate the user experience of the futur.move mHealth service in German and English. mHealth: mobile health.
[PDF File (Adobe PDF File), 185 KB - formative_v9i1e64384_app3.pdf]

Multimedia Appendix 4
Table of correlation analysis results for user experience evaluation of futur.move mHealth service using UEQ subscales, including graphical interpretation. mHealth: mobile health; UEQ: User Experience Questionnaire.
[DOCX File , 85 KB - formative_v9i1e64384_app4.docx]

Multimedia Appendix 5
Illustrative quotations from focus group discussions on the user experience of the futur.move mHealth service to support qualitative findings, presented in German and English. mHealth: mobile health.
[DOCX File , 32 KB - formative_v9i1e64384_app5.docx]

References

1. Miko HC, Zillmann N, Ring-Dimitriou S, Dorner TE, Titze S, Bauer R. Effects of physical activity on health. Gesundheitswesen 2020;82(S 03):S184-S195 [FREE Full text] [doi: 10.1055/a-1217-0549] [Medline: 32984942]
2. Warburton DE, Glendhill N, Quinney A. The effects of changes in musculoskeletal fitness on health. Can J Appl Physiol 2001;26(2):161-216. [doi: 10.1139/h01-012] [Medline: 11312416]
3. WHO guidelines on physical activity and sedentary behaviour. World Health Organization. 2020. URL: https://iris.who.int/bitstream/handle/10665/336656/9789240015128-eng.pdf?sequence=1 [accessed 2025-03-07]
4. Meng F, Xie C, Qiu F, Geng J, Li F. Effects of physical activity level on attentional networks in young adults. Int J Environ Res Public Health 2022;19(9):5374 [FREE Full text] [doi: 10.3390/ijerph19095374] [Medline: 35564768]
5. Redondo-Flórez L, Ramos-Campo DJ, Clemente-Suárez VJ. Relationship between physical fitness and academic performance in university students. Int J Environ Res Public Health 2022;19(22):14750 [FREE Full text] [doi: 10.3390/ijerph192214750] [Medline: 36429470]
6. Faselt F, Geuter G. Grundlagen zur Bewegungsförderung in Lebenswelten. Bewegungsförderung in Lebenswelten. Bielefeld: LIGA.NRW; 2011. URL: https://www.lzg.nrw.de/_media/pdf/liga-fokus/LIGA_Fokus_14.pdf#page=11 [accessed 2025-03-13]
7. Guthold R, Stevens GA, Riley LM, Bull FC. Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1·6 million participants. Lancet Child Adolesc Health 2020;4(1):23-35 [FREE Full text] [doi: 10.1016/S2352-4642(19)30323-2] [Medline: 31761562]
8. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Health 2018;6(10):e1077-e1086 [FREE Full text] [doi: 10.1016/S2214-109X(18)30357-7] [Medline: 30193830]
9. Clemente FM, Nikolaidis PT, Martins FML, Mendes RS. Physical activity patterns in university students: do they follow the public health guidelines? PLoS One 2016;11(3):e0152516 [FREE Full text] [doi: 10.1371/journal.pone.0152516] [Medline: 27022993]
10. Ráthonyi G, Takács V, Szilágyi R, Bácsné Bába É, Müller A, Bács Z, et al. Your physical activity is in your hand-objective activity tracking among university students in Hungary, one of the most obese countries in Europe. Front Public Health 2021;9:661471 [FREE Full text] [doi: 10.3389/fpubh.2021.661471] [Medline: 34604150]
11. Arias-Palencia NM, Solera-Martínez M, Gracia-Marco L, Silva P, Martínez-Vizcaíno V, Cañete-García-Prieto J, et al. Levels and patterns of objectively assessed physical activity and compliance with different public health guidelines in university students. PLoS One 2015;10(11):e0141977 [FREE Full text] [doi: 10.1371/journal.pone.0141977] [Medline: 26536605]
12. Sweileh WM. Visualized analysis of research publications on physical activity among higher education students. Am J Health Educ 2023;54(5):364-375. [doi: 10.1080/19325037.2023.2232427]
13. Hirvensalo M, Lintunen T. Life-course perspective for physical activity and sports participation. Eur Rev Aging Phys Act 2011;8(1):13-22. [doi: 10.1007/s11556-010-0076-3]
14. Meyer B. Wie geht's Deutschlands Studierenden? Gesundheitsreport. 2023. URL: https://www.tk.de/resource/blob/2149886/e5bb2564c786aedb3979588fe64a8f39/2023-tk-gesundheitsreport-data.pdf [accessed 2025-03-07]

XSL•FO
RenderX

15. Kettunen E, Kari T, Makkonen M, Critchley W, Sell A. Digital coaching among university students with low levels of physical activity: a quantitative intervention study on exercise self-efficacy. 2019 Presented at: Humanizing technology for a sustainable society: Conference proceedings; June 16-19, 2019; Bled, Slovenia p. 861-880. [doi: 10.18690/978-961-286-280-0.45]
16. Schwarz A, Winkens LHH, de Vet E, Ossendrijver D, Bouwsema K, Simons M. Design features associated with engagement in mobile health physical activity interventions among youth: systematic review of qualitative and quantitative studies. JMIR mHealth uHealth 2023;11:e40898 [FREE Full text] [doi: 10.2196/40898] [Medline: 36877551]
17. Guo Y, Bian J, Leavitt T, Vincent HK, Vander Zalm L, Teurlings TL, et al. Assessing the quality of mobile exercise apps based on the American college of sports medicine guidelines: a reliable and valid scoring instrument. J Med Internet Res 2017;19(3):e67 [FREE Full text] [doi: 10.2196/jmir.6976] [Medline: 28270378]
18. Knight E, Stuckey MI, Prapavessis H, Petrella RJ. Public health guidelines for physical activity: is there an app for that? A review of Android and Apple app stores. JMIR mHealth uHealth 2015;3(2):e43 [FREE Full text] [doi: 10.2196/mhealth.4003] [Medline: 25998158]
19. Kujala S, Roto V, Väänänen-Vainio-Mattila K, Karapanos E, Sinnelä A. UX Curve: a method for evaluating long-term user experience. Interact Comput 2011;23(5):473-483. [doi: 10.1016/j.intcom.2011.06.005]
20. Schrepp M, Hinderks A, Thomaschewski J. Applying the User Experience Questionnaire (UEQ) in different evaluation scenarios. In: Hutchison D, Kanade T, Kittler J, Kleinberg JM, Kobsa A, Mattern F, et al, editors. Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience. Vol. 8517. Cham, Switzerland: Springer International Publishing; 2014:383-392.
21. Ergonomics of human-system interaction, part 210: human-centred design for interactive systems. ISO. 2024. URL: https://www.iso.org/standard/77520.html [accessed 2025-03-07]
22. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]
23. Laranjo L, Ding D, Heleno B, Kocaballi B, Quiroz JC, Tong HL, et al. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. Br J Sports Med 2021;55(8):422-432. [doi: 10.1136/bjsports-2020-102892] [Medline: 33355160]
24. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
25. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
26. Zhang J, Brackbill D, Yang S, Becker J, Herbert N, Centola D. Support or competition? How online social networks increase physical activity: a randomized controlled trial. Prev Med Rep 2016;4:453-458 [FREE Full text] [doi: 10.1016/j.pmedr.2016.08.008] [Medline: 27617191]
27. Arigo D, Jake-Schoffman DE, Wolin K, Beckjord E, Hekler EB, Pagoto SL. The history and future of digital health in the field of behavioral medicine. J Behav Med 2019;42(1):67-83 [FREE Full text] [doi: 10.1007/s10865-018-9966-z] [Medline: 30825090]
28. Tong HL, Quiroz JC, Kocaballi AB, Fat SCM, Dao KP, Gehringer H, et al. Personalized mobile technologies for lifestyle behavior change: a systematic review, meta-analysis, and meta-regression. Prev Med 2021;148:106532. [doi: 10.1016/j.ypmed.2021.106532] [Medline: 33774008]
29. Zhou M, Mintz Y, Fukuoka Y, Goldberg K, Flowers E, Kaminsky P, et al. Personalizing mobile fitness apps using reinforcement learning. CEUR Workshop Proc 2018;2068 [FREE Full text] [Medline: 32405286]
30. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet E-surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
31. Mager B. Service Design. In: Erlhoff M, Marshall T, editors. Wörterbuch Design: Begriffliche Perspektiven des Design. Basel, Germany: Birkhäuser; 2008:362.
32. van Gemert-Pijnen JEWC, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
33. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
34. Gür F, Gür GC, Ayan V. The effect of the ERVE smartphone app on physical activity, quality of life, self-efficacy, and exercise motivation for inactive people: a randomized controlled trial. Eur J Integr Med 2020;39:101198. [doi: 10.1016/j.eujim.2020.101198]
35. Messing S, Rütten A. Quality criteria for the conception, implementation and evaluation of interventions for physical activity promotion: a state-of-the-art review. Gesundheitswesen 2017;79(S 01):S60-S65. [doi: 10.1055/s-0042-123378] [Medline: 28399588]

XSL•FO
RenderX

36. Finger JD, Tafforeau J, Gisle L, Oja L, Ziese T, Thelen J, et al. Development of the European Health Interview Survey-Physical Activity Questionnaire (EHIS-PAQ) to monitor physical activity in the European Union. Arch Public Health 2015;73:59 [FREE Full text] [doi: 10.1186/s13690-015-0110-z] [Medline: 26634120]
37. What sample size do you really need for UX research? UserTesting. 2023. URL: https://www.usertesting.com/blog/ux-research-sample-size [accessed 2025-03-07]
38. Hennink MM. Focus Group Discussions. New York, NY: Oxford University Press; 2013.
39. Laugwitz B, Held T, Schrepp M. Construction and evaluation of a user experience questionnaire. In: HCI and usability for education and work.: Springer; 2008 Presented at: 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008; November 20-21, 2008; Graz, Austria p. 63-76. [doi: 10.1007/978-3-540-89350-9]
40. Schrepp M. User Experience Questionnaire Handbook. 2023. URL: https://www.ueq-online.org/ [accessed 2024-05-03]
41. Schulz M, Mack B, Renn O. Fokusgruppen in der empirischen Sozialwissenschaft: Von der Konzeption bis zur Auswertung. Wiesbaden, Germany: Springer VS; 2012.
42. Doering N, Bortz J. Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften. Berlin, Germany: Springer; 2016.
43. Kuckartz U. Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. 4., überarbeitete Aufl. Weinheim, Germany: Beltz; 2018.
44. Boerema S, van Velsen L, Hermens H. An intervention study to assess potential effect and user experience of an mHealth intervention to reduce sedentary behaviour among older office workers. BMJ Health Care Inform 2019;26(1):e100014 [FREE Full text] [doi: 10.1136/bmjhci-2019-100014] [Medline: 31744844]
45. Moore R, Al-Tamimi AK, Freeman E. Investigating the potential of a conversational agent (Phyllis) to support adolescent health and overcome barriers to physical activity: co-design study. JMIR Form Res 2024;8:e51571 [FREE Full text] [doi: 10.2196/51571] [Medline: 38294857]
46. Signorelli GR, Monteiro-Guerra F, Rivera-Romero O, Núñez-Benjumea FJ, Fernández-Luque L. Breast cancer physical activity mobile intervention: early findings from a user experience and acceptability mixed methods study. JMIR Form Res 2022;6(6):e32354 [FREE Full text] [doi: 10.2196/32354] [Medline: 35731554]
47. Tong HL, Quiroz JC, Kocaballi AB, Ijaz K, Coiera E, Chow CK, et al. A personalized mobile app for physical activity: an experimental mixed-methods study. Digit Health 2022;8:20552076221115017 [FREE Full text] [doi: 10.1177/20552076221115017] [Medline: 35898287]
48. Kushendriawan MA, Santoso HB, Putra POH, Schrepp M. Evaluating user experience of a mobile health application 'Halodoc' using user experience questionnaire and usability testing. J Sist Inf (J Inf Sys) 2021;17(1):58-71. [doi: 10.21609/jsi.v17i1.1063]
49. Figueroa CA, Gomez-Pathak L, Khan I, Williams JJ, Lyles CR, Aguilera A. Ratings and experiences in using a mobile application to increase physical activity among university students: implications for future design. Univers Access Inf Soc 2024;23(2):821-830 [FREE Full text] [doi: 10.1007/s10209-022-00962-z] [Medline: 36624825]
50. Buis LR, Poulton TA, Holleman RG, Sen A, Resnick PJ, Goodrich DE, et al. Evaluating active U: an internet-mediated physical activity program. BMC Public Health 2009;9:331 [FREE Full text] [doi: 10.1186/1471-2458-9-331] [Medline: 19744311]
51. Nuijten RCY, van Gorp P, Borghouts T, Le Blanc P, van den Berg P, Kemperman A, et al. Preadolescent students' engagement with an mHealth intervention fostering social comparison for health behavior change: crossover experimental study. J Med Internet Res 2021;23(7):e21202 [FREE Full text] [doi: 10.2196/21202] [Medline: 34326041]
52. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
53. Hargittai E, Piper AM, Morris MR. From internet access to internet skills: digital inequality among older adults. Univ Access Inf Soc 2018;18(4):881-890. [doi: 10.1007/s10209-018-0617-5]
54. Neyer FJ, Felber J, Gebhardt C. Entwicklung und Validierung einer Kurzskala zur Erfassung von technikbereitschaft. Diagnostica 2012;58(2):87-99. [doi: 10.1026/0012-1924/a000067]
55. Venkatesh V, Morris MG. Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. MIS Q 2000;24(1):115. [doi: 10.2307/3250981]
56. Dockweiler C, Boketta R, Schnecke J, Hornberg C. Use and acceptance of mHealth- applications among young adults in Germany / nutzungsverhalten und akzeptanz von mHealth-applikationen bei jungen erwachsenen in Deutschland. 2016 Presented at: TELEMED 2015 - Erfahrungen, Stand und Perspektiven von Gesundheitstelematik und Telemedizin in Deutschland; March 09, 2015; Berlin, Germany URL: https://tinyurl.com/y2te4mn2
57. Ellis D, Allaire JC. Modeling computer interest in older adults: the role of age, education, computer knowledge, and computer anxiety. Hum Factors 1999;41(3):345-355. [doi: 10.1518/001872099779610996] [Medline: 10665203]
58. Al-Nawaiseh HK, McIntosh WA, McKyer LJ. An-m-health intervention using smartphone app to improve physical activity in college students: a randomized controlled trial. Int J Environ Res Public Health 2022;19(12):7228 [FREE Full text] [doi: 10.3390/ijerph19127228] [Medline: 35742477]

XSL•FO
RenderX

59. Davis A, Ellis R. A quasi-experimental investigation of college students' ratings of two physical activity mobile apps with varied behavior change technique quantity. Digit Health 2019;5:2055207619891347 [FREE Full text] [doi: 10.1177/2055207619891347] [Medline: 31827878]
60. Pentakota N, Ramaswamy G, Thekkur P, Nair D, Chinnakali P, Kumar Saya G. Is a smartphone application effective in improving physical activity among medical school students? Results from a quasi-experimental study. Int J Adolesc Med Health 2021;33(4):20180192. [doi: 10.1515/ijamh-2018-0192] [Medline: 30920953]
61. Aussagekraft des Body-Mass-Index ist begrenzt. Deutsches Ärzteblatt. 2023. URL: https://www.aerzteblatt.de/nachrichten/140448/Aussagekraft-des-Body-Mass-Index-ist-begrenzt [accessed 2025-03-07]
62. Draeger A, Obi N, Jagodzinski A, Becher H. Unterschätzung des BMI bei Eigenangaben zu Größe und Gewicht. Gemeinsame Jahrestagung -der Deutschen Gesellschaft für Epidemiologie e.V. (DGEpi), -der Deutschen Gesellschaft für Medizinische Soziologie e.V. (DGMS) und -der Deutschen Gesellschaft für Sozialmedizin und Prävention e.V. (DGSMP): Georg Thieme Verlag KG 2017;79(08/09):656-804. [doi: 10.1055/s-0037-1605952]
63. Hello GPT-4o. OpenAI. 2024 May 13. URL: https://openai.com/index/hello-gpt-4o/ [accessed 2025-03-14]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet e-Surveys
EHIS-PAQ: European Health Interview Survey–Physical Activity Questionnaire
HAWK: University of Applied Sciences and Arts Hildesheim/Holzminden/Göttingen
mHealth: mobile health
PA: physical activity
UEQ: User Experience Questionnaire
UX: user experience
WHO: World Health Organization

Edited by A Mavragani; submitted 17.07.24; peer-reviewed by J Fiedler, K Kammler-Sückler; comments to author 26.11.24; revised version received 14.01.25; accepted 19.01.25; published 28.03.25.

Please cite as:
Wittmar S, Frankenstein T, Timm V, Frei P, Kurpiers N, Wölwer S, Schäfer AGM
User Experience With a Personalized mHealth Service for Physical Activity Promotion in University Students: Mixed Methods Study
JMIR Form Res 2025;9:e64384
URL: https://formative.jmir.org/2025/1/e64384
doi:10.2196/64384
PMID:

©Silke Wittmar, Tom Frankenstein, Vincent Timm, Peter Frei, Nicolas Kurpiers, Stefan Wölwer, Axel Georg Meender Schäfer. Originally published in JMIR Formative Research (https://formative.jmir.org), 28.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Convergent and Known-Groups Validity and Sensitivity to Change of the Virtual Performance Measure in Patients With Hip and Knee Osteoarthritis: Longitudinal Study

Helen Razmjou[1], PhD; Suzanne Denis[1], MSc, PT; Susan Robarts[1], BSc, BHSCPT, MSc; Amy Wainwright[1], MSc, PT; Patricia Dickson[1], MSc, OT; John Murnaghan[1], MD, MSc, MA

Sunnybrook Health Science Centre, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Helen Razmjou, PhD
Sunnybrook Health Science Centre
University of Toronto
43 Wellesley Street East
Toronto, ON, M4Y 1H1
Canada
Email: helen.razmjou@sunnybrook.ca

Abstract

Background: Subsequent to the COVID-19 pandemic in 2020, a different approach to health care utilization was required to improve safety and efficiency. In the postpandemic era, virtual care and remote assessment of musculoskeletal conditions has become more common, and examining the accuracy of these remote encounters remains vital. In 2023, an innovative, video-based tool—the Virtual Performance Measure (VPM)—was introduced to assess the functional difficulties of patients with osteoarthritis of the knee joint. Further validation of this tool is warranted to expand its application longitudinally and in more diverse populations.

Objective: This study examined the longitudinal validity of the VPM, a digitally based outcome tool, in patients with osteoarthritis of the hip and knee joints who had undergone arthroplasty.

Methods: Patients completed a web-based survey after watching 40 videos that demonstrated 10 functional tasks with increasing difficulty, prior to and at approximately 3-5 months following surgery. The Lower Extremity Functional Scale (LEFS) was used as the reference measure. Longitudinal convergent and known-groups validity as well as sensitivity to change were assessed.

Results: The data of 120 patients (n=80, 67% female; mean age 67, SD 9 years; n=58, 48% with hip osteoarthritis and n=62, 52% with knee osteoarthritis) were examined. There was a statistically significant improvement in both LEFS (t_{119}=16.04, P<.001) and VPM total scores (t_{119}=13.92, P<.001) over time. The correlation between the postoperative LEFS and VPM scores was higher (r=0.66; P<.001) than the correlation between the change scores of these measures (r=0.51; P<.001). The area under the curve value for the VPM's ability to differentiate between urgent and nonurgent candidates for surgery was 0.71 (95% CI 0.57-0.84). Sensitivity to change as measured by the standardized response mean was 1.27 (95% CI 1.09-1.45), indicating good ability to detect change over time.

Conclusions: The VPM demonstrated sufficient longitudinal convergent and known-groups validity as well as sensitivity to change in patients with hip and knee osteoarthritis following arthroplasty. This tool has a potential to improve the delivery of care by increasing access, reducing the frequency of in-person visits, and improving the overall efficiency of the health care system following a major surgery.

(JMIR Form Res 2025;9:e69001) doi:10.2196/69001

KEYWORDS

virtual; video-based outcome; longitudinal validity; sensitivity to change; osteoarthritis

Introduction

Osteoarthritis of the hip and knee joint is a disabling condition in older adults with significant physical, social, and economic burdens. Traditionally, two types of outcome measurement have been used in the assessment of patients with lower-extremity osteoarthritis. Patient-reported outcome measures document perceived physical difficulty within a specific timeframe [1].

XSL•FO
RenderX

Performance-based measures reflect the actual ability of the patients to execute a specific physical activity at the time of assessment [2]. The first video-based outcome measure that permitted a remote assessment of functional difficulties was introduced by a team of Dutch investigators in 2014 using videos with animated figures [3-5]. Subsequent to the COVID-19 pandemic in 2020, a different approach to health care utilization was required to improve safety and efficiency. In 2023, a next-generation, innovative, virtual assessment tool—the Virtual Performance Measure (VPM)—was introduced [6]. The VPM replaced animated figures with human models and combined features of self-report and performance-based tools, creating an outcome measure that could be used remotely. The original studies by the developers have shown cross-sectional validity of the VPM in patients with hip and knee joint osteoarthritis [6,7]. As clinicians continue to face challenges in care delivery and digital technology is being adopted more often by older persons in the postpandemic era, further validation of this tool is warranted to expand its application longitudinally and in more diverse populations.

The objective of this study was to examine the longitudinal convergent construct and known-groups validity as well as sensitivity to change of the VPM total score in patients with osteoarthritis of the hip and knee joints who had undergone joint replacement surgery. We explored the role of age, sex, and severity score on the postoperative VPM score after accounting for the preoperative VPM score. We hypothesized that (1) the postoperative and change scores of the VPM would correlate moderately with those of a traditional, validated, self-report functional tool; (2) the VPM would differentiate between patients who were considered urgent versus nonurgent for joint arthroplasty; and (3) the VPM score would show a large sensitivity to change following surgery.

Methods

Participants

This was a longitudinal study. All patients were recruited at a rapid-access clinic of a large arthroplasty center of an academic hospital in Canada. Inclusion criteria were failed nonsurgical management and candidacy for primary elective arthroplasty. Exclusion criteria included inability to read English, presence of cognitive conditions, lack of access to internet, or inability to use web-based surveys.

Ethical Considerations

The study protocol was approved by the Human Ethics Research Board of the Sunnybrook Health Sciences Centre (Project# 3703). All patients provided informed consent for participation in the study. The web-based survey used a participant number, no personal identifiers were collected, and study data were anonymous. Patients were not compensated for their participation in the study.

VPM Procedure

Patients who agreed to participate in the study received a link from a research assistant to a web-based survey. After entering their participant number, they completed the survey after watching 40 videos of 10 tasks on a smart phone, computer, or tablet twice—prior to surgery and 3-5 months following surgery. Activities varied from simple tasks such as sitting on a chair to physically demanding tasks such as navigating stairs, to provide a wide variety of activities and avoid ceiling and floor effects. The activities and their variations were selected based on input from physical therapists, occupational therapists, and patient representatives. The final set of the videos used an age-appropriate model.

The videos represented 10 commonly executed functional tasks with 4 levels of difficulty (normal to severe). Patients had an option of choosing “unable to perform” if the task was considered impossible to perform. The daily tasks included sitting and rising from a chair, putting on or taking off socks, getting into the shower, picking up an item from the floor, sitting down and getting up from the floor, walking on even ground, and ascending and descending stairs. The survey results were sent to the researchers via a secure platform. Further details on the development and cross-sectional validity of the tool have been provided elsewhere [6,7]. The videos are available to public (patients and researchers) [8].

Preoperative Assessment

Initial assessment was conducted by advanced practice providers, who assigned a score to each candidate using the Severity Scoring System (SSS). The SSS is a validated scoring system and acts as a guide to interpreting findings across three main elements of the standardized assessment: (1) patient’s clinical presentation, (2) self-report pain and functional difficulties, and (3) severity of osteoarthritis based on imaging. The score helps improve access and outcomes in candidates for joint arthroplasty by determining an appropriate management plan, including assigning urgency and priority levels for surgery [6,7,9,10]. The SSS has been aligned with the Wait Time Information System priority levels and classifications in Canada [11]. The clinical component score of the SSS ranges from 0-3, the functional component score varies from 0-3, and the radiological examination score varies from 0-2. The total severity score ranges between 0-8, with higher numbers indicating a higher severity. Patients with a score of 7 or higher are considered urgent for surgical management.

Reference Outcome Measure

The reference measure used for convergent validity was the Lower Extremity Functional Scale (LEFS) [12], which was completed prior to surgery and after surgery along with the VPM. The LEFS is a validated, patient-reported outcome measure with 20 questions, with a score ranging from 0-80, where higher numbers indicate better function. The LEFS has shown validity and reliability in patients with lower-extremity osteoarthritis and after hip and knee arthroplasty [6,7,9,10,12,13].

Statistical Analysis

Descriptive analyses of patient characteristics were conducted. The preoperative LEFS and VPM scores were compared with postoperative scores, and the magnitude of change over time was examined using paired, 1-tailed *t* test statistics. The analysis of covariance was used to examine the role of age, sex, and severity score on postoperative VPM scores after adjusting for

XSL•FO
RenderX

the preoperative level of functional difficulty. Interactions among independent variables were explored.

Longitudinal construct convergent validity was used to examine the correlation between postoperative scores and the amount of change (*postoperative score – preoperative score*) between the VPM and LEFS, using Spearman correlations. We hypothesized moderate correlations (r>0.30 and <0.50) [14] between the VPM total score and the LEFS score due to using different modes of measurement (self-report survey vs video-based survey).

Known-groups validity in a longitudinal context is examined by the assessment of average change in subgroups that are expected to have experienced a different amount of change in function after surgery. The receive operating characteristic curve and the area under the curve (AUC) value (area under the receiver operating characteristic curve) were used to assess the diagnostic ability of the VPM in detecting a difference in change over time between a binary predictor that indicated priority for surgery (eg, severity score ≥7 and <7). An AUC of above 0.70 is considered sufficient for diagnostic validity in clinical practice [15].

Sensitivity to change refers to the capacity of a tool to measure the magnitude of observed change statistically. We therefore measured the magnitude of change between the scores at baseline and following surgery using the standardized response mean (SRM). The SRM was calculated as *(mean change score / SD of the change score)* [14]. The Cohen criteria were used to interpret the magnitude of the SRM value [16]. An SRM of 0.80 is considered large in clinical practice.

Results

The data of 120 patients (n=80, 67% female and n=40, 33% male; mean age 67, SD 9 years; age range 40-90 years; n=58, 48% hip osteoarthritis and n=62, 52% knee osteoarthritis) were examined. Table 1 shows the demographic characteristics of the sample included. There was a statistically significant improvement in both LEFS (t_{119}=16.04, P<.001) and VPM total scores (t_{119}=13.92, P<.001). Younger age ($F_{4,115}$=9.93; P=.002) and male sex ($F_{4,115}$=4.68; P=.03) were predictive of better VPM scores at follow-up after adjusting for the preoperative VPM score and severity of the condition.

The results of the longitudinal convergent validation showed strong correlations between the postoperative LEFS and VPM scores (r=0.66; P<.001) and moderate correlation between the change scores of these measures (r=0.51; P<.001).

The logistic regression that calculated the overall diagnostic value of the VPM change score for the differentiation between high priority versus moderate or low priority (based on the SSS) showed acceptable predictive validity, with an AUC of 0.71 (95% CI 0.57-0.84; Figure 1).

The SRM was 1.27 (95% CI 1.09-1.45), indicating good sensitivity to change and ability of the VPM tool to detect a true effect of surgical management.

XSL•FO
RenderX

Table 1. Demographics, disease characteristics, and preoperative and postoperative outcomes (N=120).

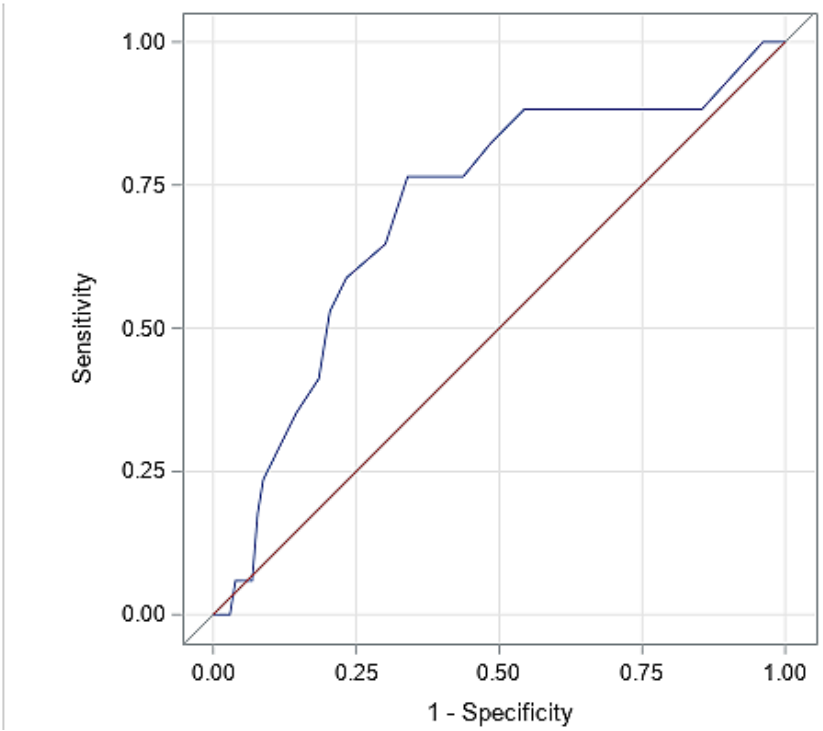
Variables	Values
Age (years), mean (SD; range)	67 (9; 40-92)
Sex, n (%)	
Female	80 (67)
Male	40 (37)
Affected joint, n (%)	
Hip	58 (48)
Knee	62 (52)
Walking devices, n (%)	
Yes	35 (29)
No	85 (71)
Gait, n (%)	
Normal	10 (8)
Abnormal	103 (86)
Missing	7 (6)
Severity score, mean (SD)	
Clinical (0-3)	2 (0.48)
Functional (0-3)	1.90 (0.50)
Radiological (0-2)	1.85 (0.34)
Total score (0-8)	5.8 (10.96)
Outcome measures, mean (SD)	
Preoperative LEFS[a] score	32 (14)
Postoperative LEFS score	59 (13)
Preoperative VPM[b] score	60 (18)
Postoperative VPM score	81 (13)

[a]LEFS: Lower Extremity Functional Scale.

[b]VPM: Virtual Performance Measure.

XSL•FO
RenderX

Figure 1. ROC analysis of the VPM change score in relation to surgical urgency. ROC: receiver operating characteristic; VPM: Virtual Performance Measure. Model (0.7079), ROC1 (0.5000).



Discussion

Principal Findings

This study demonstrated longitudinal convergent construct and known-groups validity as well as high sensitivity to change in patients with osteoarthritis of the hip and knee joints who had undergone joint replacement surgery. The postoperative and change scores of the VPM correlated strongly with those of the LEFS, a self-report functional tool. The VPM score was able to differentiate between patients who were considered urgent versus nonurgent for surgery (high priority vs moderate or low priority) based on the severity score. Younger patients and male patients showed a better postoperative VPM score after accounting for the preoperative VPM score and severity of the condition.

The VPM has shown cross-sectional convergent and known-groups validity in patients with knee and hip osteoarthritis [6,7]. The cross-sectional validity, however, is not sufficient for using a tool after an intervention, and longitudinal studies are required to detect change by extending beyond a single point in time and establishing the impact of surgery on improving function. To our knowledge, this is the first study that demonstrates the longitudinal validity of a digitally based outcome measure that uses a human model. In the cross-sectional study of patients with knee osteoarthritis [6] and hip joint osteoarthritis [7], the correlations between the preoperative LEFS and VPM score were 0.66 (P<.001) and 0.68 (P<.001), respectively. In this study, there was a similar correlation between the postoperative scores (r=0.66; P<.001). The correlation between the change scores of the LEFS and VPM was lower (r=0.51; P<.001). In terms of known-groups validity, AUC values of 0.90 and 0.72 were reported in relation to candidacy for total knee and hip arthroplasty [6,7]. The lower AUC value of 0.71 of this study may be related to a higher variability in the patient sample and the stricter criteria of the predictor variable (urgency vs candidacy).

The sudden rise in popularity and demand for virtual care was a consequence of the 2020 COVID-19 pandemic, which required a different approach to health care utilization and access across the world. In the postpandemic era, clinicians continue to rely on virtual modes of examination and management to maintain efficiency and access. Thus, establishing the diagnostic validity of digital tools ensures the accuracy of clinical information obtained through virtual care. Improved access to broadband internet in rural and remote areas and higher use of new technologies by older individuals will lessen the burden of commute for families who often have to drive their older family members to arthroplasty centers, which are often located in larger cities. As technology evolves, alternative approaches to virtual functional assessment that provide accurate data on range of motion and function should be explored in osteoarthritis care. A potential tool for future virtual care is the sensorless motion tracking technology that uses cameras and computer vision software to track the movements of a patient on personal devices such as a phone, tablet, or laptop. With the use of artificial intelligence, sensorless tracking systems may provide accurate and reliable results with no additional hardware; this avoids the complexity of sensor-based systems that require wearing multiple high-cost sensors and devices and have limited ability to track related body parts, thus affecting calibration and scaling.

The implication of this study in the era of fast developing new technologies is establishing confidence in the ability of a digital tool to provide valid results, predict true change, and differentiate between groups with different characteristics and patterns of recovery, while improving access. These results will

XSL•FO
RenderX

inform further research in more diverse populations, shape policy, and help to find solutions to today's health care challenges.

This study has some limitations. Patients who participated in this study had advanced osteoarthritis that required joint arthroplasty. The results may not be applicable to patients with less pathology and higher functional levels. The model used in the videos may not represent younger patients, female patients, or patients with greater levels of obesity, and these potential differences warrant investigation in future studies.

Conclusions

The VPM demonstrated sufficient longitudinal convergent and known-group validity as well as sensitivity to change in patients with hip and knee osteoarthritis following arthroplasty. This tool has the potential to improve the delivery of care by increasing access, reducing the frequency of in-person visits, and improving the overall efficiency of the health care system following a major surgery.

Data Availability

Due to ongoing nature of the study, partial data are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Gandhi R, Tsvetkov D, Davey JR, Syed KA, Mahomed NN. Relationship between self-reported and performance-based tests in a hip and knee joint replacement population. Clin Rheumatol 2009 Mar 8;28(3):253-257. [doi: 10.1007/s10067-008-1021-y] [Medline: 18853222]
2. Stratford PW, Kennedy DM, Woodhouse LJ. Performance measures provide assessments of pain and function in people with advanced osteoarthritis of the hip or knee. Phys Ther 2006 Nov;86(11):1489-1496. [doi: 10.2522/ptj.20060002] [Medline: 17079748]
3. Terwee CB, Coopmans C, Peter WF, Roorda LD, Poolman RW, Scholtes VA, et al. Development and validation of the computer-administered animated activity questionnaire to measure physical functioning of patients with hip or knee osteoarthritis. Phys Ther 2014 Feb;94(2):251-261. [doi: 10.2522/ptj.20120472] [Medline: 24029297]
4. Peter WF, de Vet HCW, Terwee CB. Reliability of the Animated Activity Questionnaire for assessing activity limitations of patients with hip and knee osteoarthritis. Musculoskeletal Care 2018 Sep 17;16(3):363-369. [doi: 10.1002/msc.1349] [Medline: 29770556]
5. Peter WF, Poolman RW, Scholtes VA, de Vet HCW, Terwee CB. Responsiveness and interpretability of the Animated Activity Questionnaire for assessing activity limitations of patients with hip or knee osteoarthritis. Musculoskeletal Care 2019 Dec 12;17(4):327-334. [doi: 10.1002/msc.1418] [Medline: 31402572]
6. Razmjou H, Denis S, Robarts S, Falconer J, Anunciacion M, Nunn A, et al. Virtual performance measure in osteoarthritis: an innovative transformation of patient care. Osteoarthr Cartil Open 2023 Dec;5(4):100410 [FREE Full text] [doi: 10.1016/j.ocarto.2023.100410] [Medline: 37829082]
7. Razmjou H, Denis S, Robarts S, Wainwright A, Dickson P, Roszkowski A, et al. Comparative assessment of a virtual performance measure with self-report and performance-based outcomes in patients with hip osteoarthritis. BMC Musculoskelet Disord 2025 Jan 08;26(1):30 [FREE Full text] [doi: 10.1186/s12891-025-08282-y] [Medline: 39780103]
8. Examining your functional difficulties related to your hip or knee osteoarthritis by the Virtual Performance Measure. Sunnybrook Health Sciences Centre. URL: https://sunnybrook.ca/content/?page=virtual-performance-measures [accessed 2025-03-25]
9. Robarts S, Palinkas V, Boljanovic D, Razmjou H. Validity of the virtual severity scoring system in patients with knee osteoarthritis. 2023 Apr 11 Presented at: The International Combined Orthopaedic Research Societies (ICORS), World Congress of Orthopaedic Research; September 7-9, 2022; Edinburgh, Scotland p. 51. [doi: 10.1302/1358-992x.2023.8.051]
10. Robarts S, Denis S, Kennedy D, Dickson P, Juma S, Palinkas V, et al. Patient gender does not influence referral to an orthopaedic surgeon by advanced practice orthopaedic providers: a prospective observational study in Canada. BMC Health Serv Res 2021 Sep 11;21(1):952 [FREE Full text] [doi: 10.1186/s12913-021-06965-5] [Medline: 34511124]
11. Trypuc J, Hudson A, MacLeod H. Ontario's wait time strategy: part 1. Healthc Q 2006 Mar 15;9(2):44-51, 2. [doi: 10.12927/hcq.2006.18101] [Medline: 16640132]
12. Binkley JM, Stratford PW, Lott SA, Riddle DL. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. North American Orthopaedic Rehabilitation Research Network. Phys Ther 1999 Apr;79(4):371-383. [doi: 10.1037/t35109-000] [Medline: 10201543]
13. Mehta SP, Fulton A, Quach C, Thistle M, Toledo C, Evans NA. Measurement properties of the Lower Extremity Functional Scale: a systematic review. J Orthop Sports Phys Ther 2016 Mar;46(3):200-216. [doi: 10.2519/jospt.2016.6165] [Medline: 26813750]

XSL•FO
RenderX

14. Liang MH, Larson MG, Cullen KE, Schwartz JA. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. Arthritis Rheum 1985 May 11;28(5):542-547. [doi: 10.1002/art.1780280513] [Medline: 4004963]
15. Hosmer DW, Lemeshow S. Chapter 5: assessing the fit of the model. In: Applied Logistic Regression, 2nd ed. New York, NY: John Wiley and Sons; 2000:160-164.
16. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.

Abbreviations

AUC: area under the curve
LEFS: Lower Extremity Functional Score
SRM: standardized response mean
SSS: Severity Scoring System
VPM: Virtual Performance Measure

Edited by A Mavragani; submitted 19.11.24; peer-reviewed by M O'Connor; comments to author 03.01.25; revised version received 10.01.25; accepted 11.03.25; published 28.03.25.

Please cite as:
Razmjou H, Denis S, Robarts S, Wainwright A, Dickson P, Murnaghan J
Convergent and Known-Groups Validity and Sensitivity to Change of the Virtual Performance Measure in Patients With Hip and Knee Osteoarthritis: Longitudinal Study
JMIR Form Res 2025;9:e69001
URL: https://formative.jmir.org/2025/1/e69001
doi:10.2196/69001
PMID:

©Helen Razmjou, Suzanne Denis, Susan Robarts, Amy Wainwright, Patricia Dickson, John Murnaghan. Originally published in JMIR Formative Research (https://formative.jmir.org), 28.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Ambulance Commanders' Reluctance to Enter Road Tunnels in Simulated Incidents and the Effects of a Tunnel-Specific e-Learning Course on Decision-Making: Web-Based Randomized Controlled Trial

Johan Hylander[1], RN, MD, PhD; Lina Gyllencreutz[2], RN, PhD; Michael Haney[3], MD, PhD; Anton Westman[1,4], MD, PhD

[1]Department of Diagnostics and Intervention, Surgery, Umeå University, Umeå, Sweden

[2]Department of Nursing, Umeå University, Umeå, Sweden

[3]Department of Diagnostics and Intervention, Anesthesiology and Intensive Care Medicine, Umeå University, Umeå, Sweden

[4]Department of Anaesthesia and Intensive Care Medicine, Karolinska University Hospital, Huddinge, Stockholm, Sweden

Corresponding Author:
Johan Hylander, RN, MD, PhD
Department of Diagnostics and Intervention, Surgery
Umeå University
Linnaeus väg 6
Umeå, 90736
Sweden
Phone: 46 90 786 50 00
Email: johan.hylander@umu.se

Abstract

Background: The optimal response to a major incident in a road tunnel involves efficient decision-making among the responding emergency services (fire and rescue services, police, and ambulances). The infrequent occurrence of road tunnel incidents may entail unfamiliarity with the tunnel environment and lead to uncertain and inefficient decision-making among emergency services commanders. Ambulance commanders have requested tunnel-specific learning materials to improve their preparedness.

Objective: We aimed to assess decision-making among ambulance commanders in simulated road tunnel incidents after they had participated in a tunnel-specific e-learning course designed to support timely and correct decisions in this context.

Methods: We conducted a web-based intervention study involving 20 participants from emergency medical services in Sweden who were randomly allocated to a test or control group. The control group (n=10, 50%) received a lecture on general incident management, while the intervention group (n=10, 50%) completed an e-learning course consisting of 5 modules focused on tunnel structure, safety, and collaboration in response. The participants took part in 2 simulation-based assessments for ambulance commander decision-making in major road tunnel incidents 1 month and 6 months after their allocated study intervention. In each simulation, the participants decided on the best course of action at 15 independent decision points, designed as multiple-choice questions. The primary outcome was the correct response to the question regarding how to appropriately enter the road tunnel. The secondary outcome measurements were correct or incorrect responses and the time taken to decide for each of the 15 decisions. Limited in-depth follow-up interviews were conducted with participants (n=5, 25%), and collected data were analyzed using qualitative content analysis.

Results: All 20 participants completed the first simulation, and 16 (80%) completed the second. The main finding was that none (0/20, 0%) of the participants correctly answered the question on entering the tunnel system in the 1-month assessment. There were no significant differences between the groups (P=.59; 2-sample test of proportions) in the second assessment. The e-learning course was not associated with more correct answers at the first assessment, including accounting for participant factors (mean difference between groups: –0.58 points, 95% CI –1.88 to 0.73; P=.36). The e-learning course was also not associated with a shorter time to completion compared to the nonintervention group in either assessment. Interviews identified 3 categories linked to the main outcome: *information* (lack of), *risk* (limited knowledge and equipment), and *mitigation* (access to maps and aide-mémoire).

XSL•FO
RenderX

Conclusions: Participation in a tunnel-specific e-learning course did not result in a measurable change in ambulance commanders' decision-making behavior during simulated road tunnel incidents. The observed hesitation to enter the road tunnel system may have several plausible causes, such as the lack of actionable intelligence and tunnel-specific plans. This novel approach to assessing commander decision-making may be transferable to other educational settings.

(JMIR Form Res 2025;9:e58542) doi:10.2196/58542

KEYWORDS

e-learning; major incident; incident management; disaster medicine; road tunnels

Introduction

Background

The United Nations projects that 68% of the world's population will live in urban areas by 2050, with an estimated increase from 4.2 billion to 6.7 billion people [1]. Increased urbanization and growing cities will require reliable infrastructure [2]. Traffic congestion is expensive for both communities and individuals. The analytical company INRIX reported that traffic congestion costs the average American 97 hours and US $1348 annually [3]. To limit traffic congestion and exhaust emissions, road tunnels are key to future urban development [2,4]. Traffic incidents in road tunnels may result in significant societal costs and individual health impairment.

Moreover, increasingly complex road tunnels are being constructed. For example, the Norwegian Rogfast ferry replacement project, which is 26.7 km long and 392 m deep, will become the world's longest and deepest subsea road tunnel upon completion in the early 2030s [5]. Furthermore, road tunnels can have different designs, including single or separate twin-tunnel tubes. Some tunnel systems include subsea passages and multiple on-ramp and off-ramp systems. In addition, networks of different tunnels may include emergency exits and rooms for technical equipment, shelters, and service tunnels [6]. Understanding the layout of a specific tunnel may benefit emergency services and maintenance personnel responding to a road tunnel incident.

When a major incident occurs in a road tunnel, such as a fire or a traffic crash, the evacuation of tunnel users is the first step in limiting casualties. A complex tunnel structure may hamper self-evacuation and complicate the accessibility of the responding emergency services (police, fire and rescue, and emergency medical services [EMSs]) to the incident site and injured individuals. Moreover, the compact designs of modern road tunnels contribute to unique medical challenges concerning specific injury patterns, specific trauma mechanisms, and increased extraction times [7-9]. Hence, accessibility issues may result in considerable delays in vital treatment, potentially leading to preventable loss of life.

Timely and adequate decision-making by emergency service commanders to facilitate correct task priority and minimize time to treatment is crucial and difficult. This is illustrated by the 2013 Norwegian Gudvanga monotube tunnel fire, where a decision was made to ventilate smoke to aid firefighter access, which, in turn, engulfed 67 evacuating persons in smoke, resulting in acute smoke inhalation injuries [10].

A study by Kristiansen et al [11] exploring the collaboration between responding emergency service commanders during a Norwegian road tunnel exercise concluded that the ambulance commanders were treated as outsiders in the decision-making process owing to uncertainty regarding their management roles. Similarly, in Sweden, an imbalance between the incident commanders and ambulance or medical commanders concerning specific road tunnel knowledge has been identified as a collaborative obstacle [12]. As important decisions need to be taken within the initial 20 minutes of the incident, this obstacle may lead to further delay [13]. The current prehospital medical management system in Sweden uses a general all-hazard approach. This means that the same management method is used regardless of the incident type or environment [14,15]. In contrast, Norwegian fire and rescue services include educational material on decision-making, task allocation, risks, human behavior, and coping with uncertainties in their training curriculum [16]. Swedish ambulance staff with road tunnels in their catchment areas are expected to initiate incident management in the event of a road tunnel incident, even without tunnel-specific knowledge.

Knowledge of the tunnel system and the ability to understand emerging situations are important for fire and rescue service commanders and personnel when responding to road tunnel fires [17]. Studies concerning EMS incident management in road tunnel incidents in Norway [18] and Sweden [19] have found that there is a perceived need for tunnel-specific and readily accessible learning materials and training courses. These are needed to be better prepared to optimally assess response needs in a complex environment and make timely decisions to facilitate early access to the injured persons.

Objective and Hypotheses

The main aim of this study was to evaluate whether participation in a tunnel-specific e-learning course affected ambulance commanders' abilities to make appropriate decisions in simulated road tunnel incidents. The secondary aim was to explore participant confidence in decision-making through structured questions and responses.

Hypothesis 1

Hypothesis 1 is that the tunnel-specific e-learning course will lead to more participants correctly choosing to enter the road tunnel system in the 1-month postcourse simulation-based assessment compared to those who have not participated in the course. Furthermore, the same is hypothesized for the second assessment, 6 months after the e-learning course intervention.

XSL•FO
RenderX

Hypothesis 2

Hypothesis 2 is that the tunnel-specific e-learning course will lead to participants giving more correct answers across the 15 decision points for the simulated tunnel incidents (fire or car crash) compared to those who have not participated in the course, taking into account relevant participant factors, including educational level and previous practical experience with tunnel incidents.

Secondary Hypothesis

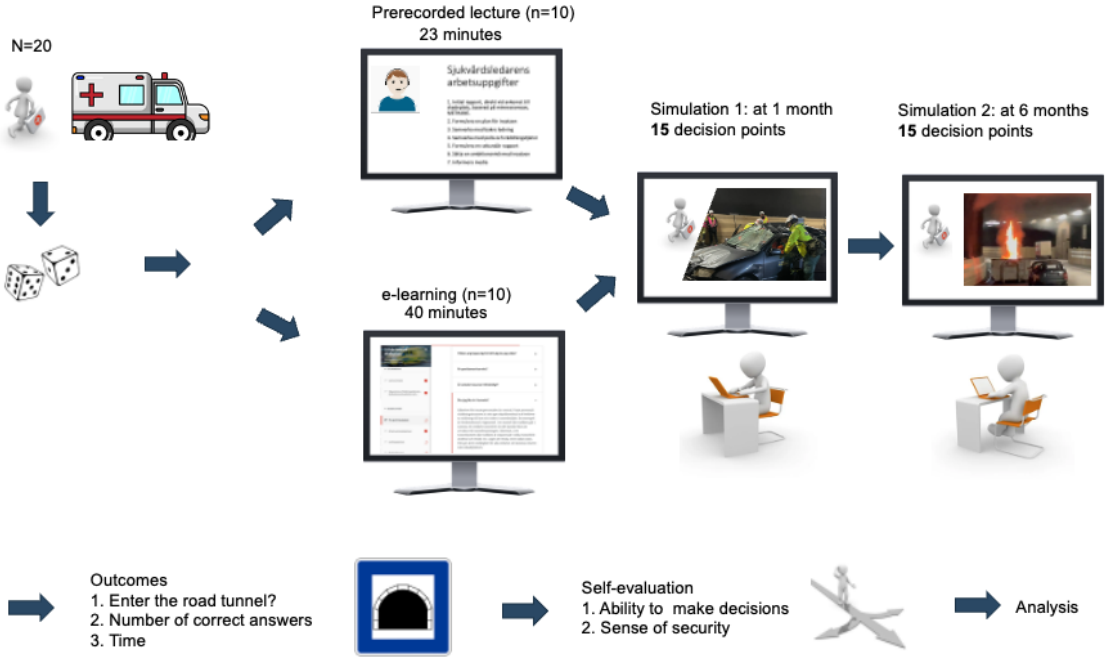
The secondary hypothesis is that tunnel-specific e-learning course participation will lead to a shorter time to make decisions in the simulation-based assessments compared with no e-learning course participation.

Methods

Study Design

The difficulty in evaluating the effectiveness of educational efforts in the medical management of major incidents has been described in the literature [20]. This study has taken an innovative approach to this issue. A schematic overview of the study design is presented in Figure 1. The design of this study is a 2-arm, assessor-blinded, randomized, prospectively controlled trial. The intervention was web based and accessed by individual participants via a purpose-built online platform. This study adhered to the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) reporting guidelines [21,22].

Figure 1. Schematic overview of the study design. This study is a 2-arm, assessor-blinded, randomized, prospectively controlled trial. Participants (N=20) were enrolled and randomized to either a prerecorded lecture or the tunnel-specific e-learning course. Both groups completed follow-up simulations at 1 and 6 months, which included 15 decision points (each designed as a multiple-choice question). Outcomes included the decision to correctly enter the tunnel, the number of correct decisions, and time. Participants also self-evaluated their decision-making ability and sense of security in handling road tunnel incidents.



Ethical Considerations

This study was approved by the Swedish Ethical Review Authority (registration numbers 2021-04810 and 2022-05388-02) and conducted in accordance with the Declaration of Helsinki [23]. As this study examined the effect of an educational course on health care provider performance in a simulation, trial registration was not deemed necessary, as described by the International Committee of Medical Journal Editors recommendations [24]. Potential participants were provided with written information, including informed consent descriptions, regarding the aim of the study and the voluntary nature of participation. They were also informed that deidentified data (eg, demographic data, such as sex and educational level) would be included in a scientific article. In addition, they were given the contact information of the study's data protection officer if, after enrollment, they wished to have their personal information removed. Potential participants were also informed that they would not receive any compensation for participation. Informed consent was obtained from the participants before they were given access to their allotted course material (refer to the Study Sequence section and Multimedia Appendix 1 for further details).

Setting

This study was conducted between March and December 2022, with participants from a large city in southern Sweden that has several twin-tube road tunnels in its catchment area. In the studied city, the EMS comprised emergency ambulances, physician-staffed vehicles, a dedicated incident management

XSL•FO
RenderX

unit, and a helicopter. Only staff from the dedicated incident management unit and emergency ambulances were included as study participants. These units were staffed by either a registered nurse and an emergency medical technician or 2 nurses. The educational level of those who staff emergency ambulances in Sweden varies from emergency medical technicians to registered nurses with a 1-year subspecialization, such as prehospital care. The EMS personnel are trained in advanced life support and prehospital management leadership. The prehospital management leadership course provided for Swedish EMS staff is a mandatory 2-day course covering the basics of incident management and draws from the Major Incident Medical Management and Support (MIMMS) course [14,25]. A separate 2-day leadership course focusing on stress resilience and leadership traits is available, although this course is not mandatory or nationally implemented [15].

In Sweden, including the studied region, the standard operating procedure (SOP) concerning major incidents is that the first arriving ambulance crew at an incident site initiates "command and control" procedures by assuming the roles of ambulance and medical commanders [25]. Consequently, all personnel who staff emergency ambulances need to be able to initiate command and control.

Sample Size Calculation, Participants, and Randomization

Sample Size Calculation

The sample size calculation was based on findings from local pretesting and recent literature [18,19], where ambulance commanders highlighted concerns about tunnel risks that could lead to hesitation in entering the road tunnel system. On the basis of this, the control group was expected to respond with close to 0 (10% response rate) correct answers to the critical categorical question (decision to correctly enter the road tunnel) and, similarly, a close to 0 (10%) correct response rate to the 15 tunnel-specific decision points (given their lack of tunnel-specific education and unfamiliarity with the tunnel environment). The intervention group was expected to correctly answer the critical single categorical question and to have a general correct response rate of 75% for all questions. Using a power of 0.8 and a 2-tailed α of .05 with the Pearson chi-square test, the estimated sample size was 8 participants per group. We aimed to include a minimum of 10 participants in each group owing to anticipated dropouts after inclusion.

Inclusion and Exclusion Criteria

The inclusion criteria were participants who were currently employed by the regional EMS and assigned to the ambulance service or dedicated incident management unit as either registered nurses or emergency medical technicians. The participants could also assume the role of ambulance commander in road tunnel incidents, in accordance with the SOP. The exclusion criteria were individuals who were involved in the development or validation of the tunnel-specific e-learning course for this study and those who were diagnosed with or had self-perceived posttraumatic stress disorder or anxiety due to road tunnel incidents.

Recruitment Process

The ambulance station commander assisted in informing potential participants of the regional study. A letter of invitation to participate in the study was distributed via "a weekly newsletter" in the organization. Furthermore, short presentations concerning the aim of and participation in the study were conducted by the first author during the organization's morning briefings. In total, 23 participants were recruited during a 3-month period (March to May 2022), of whom 20 (87%) participants were included in the study, while 3 (13%) were in reserve. Before the intervention was initiated, 3 (13%) planned participants withdrew, and the 3 (13%) individuals in reserve were included as participants.

Randomization

The first author collected the participants' email addresses. The 20 participants' email addresses were sorted in alphabetical order and assigned a case number. The participants were randomly allocated to the intervention or control arm using a random number generator. A randomization key containing case identification details was stored in a locked place to which only the first author had access.

Course Material and Study Sequence

Online Platform

An online platform was created for this study using Moodle (Moodle HQ) open-source software [26]. The online platform contained both the e-learning course and the control (nonintervention) "course." Participants were given individual access to one of the courses after a randomized allocation. Furthermore, both simulations were stored on the platform (hidden from the participants) and became available during the assessment periods (1 and 6 months after the educational intervention). Once started, neither simulation could be aborted. Thus, the participants had 1 opportunity to complete each simulation. The participants were informed of this multiple times in writing.

Prerecorded Lecture (Control Arm)

The control course consisted of a 23-minute prerecorded lecture on general incident management. The lecture consisted of a brief historical overview of major incidents where incident management was highlighted as important, including examples from the 2011 bombing in Oslo, Utøya island shooting in Norway, and the 2017 truck attack in Stockholm, Sweden [27,28]. The lecture emphasized the roles of the first ambulance at the incident site, focusing on the ambulance commander–specific tasks according to the MIMMS concept, with emphasis on the mnemonic METHANE (major incident, exact location, type of incident, hazards, access and egress, number of casualties, and emergency services) [25]. The use of the mnemonic METHANE has been an SOP in Swedish EMS prehospital incident management for the past 20 years [29]. As such, the general lecture did not include new learning or tunnel-specific elements. The purpose of the control course was to refresh the participants' general knowledge of incident management in major incidents. The control course was validated by 5 EMS personnel who had extensive prehospital experience from parts of the country outside the studied region.

XSL•FO
RenderX

The personnel were asked to provide feedback on the content of the course and perceived usefulness to ambulance commanders. Feedback included comments, for example, "Please describe MIMMS a bit more, otherwise the lecture contains much that I already knew." Changes were made according to their feedback.

The Tunnel-Specific e-Learning Course (Intervention Arm)

The intervention in this study consists of a tunnel-specific and gamified e-learning course created using Articulate Global software [30]. The e-learning course was developed in collaboration with the Faculty of Humanities at Umeå University, Sweden.

The aim of the e-learning course was to increase ambulance commanders' ability to make the proper decision at the right time by increasing knowledge of the specific tunnel environment. A particular concern was giving participants the confidence to enter the road tunnel in a safe manner, for example, using the unaffected tunnel tube during a major incident, which, according to discussions among stakeholders, is an effective method for gaining access to the scene and those injured.

The e-learning course (content outline is provided in Multimedia Appendix 2) consisted of materials concerning collaboration in a tunnel-specific context and organizations' different tasks from dispatch to arrival at the incident scene [18,19,31], methods of identifying injured individuals (eg, triage), principles for treating injuries (eg, smoke inhalation), and specific risks (eg, explosions) [32-34]. The e-learning course content was validated by 8 stakeholders from organizations involved in road tunnel rescue efforts, including fire and rescue services, EMS, police, emergency dispatch centers, and infrastructure owners. One or 2 participants from each organization participated. In addition, the course content was reviewed by 2 individuals with in-depth knowledge of prehospital education and extensive experience in acting as ambulance commanders in a variety of scenarios. Changes were made according to their feedback.

The e-learning course consisted of 5 distinct modules with learning materials on different aspects of road tunnel rescue efforts, including the road tunnel structure, potential injury patterns, safety, and collaboration, which were determined after the validation process. The modules were, in collaboration with an expert in pedagogy, incorporated into a gamified e-learning setting. This design was deemed suitable as an element of gameplay that may stimulate learning [35]. The learning materials consisted of text, images (eg, illustrating how to correctly position an ambulance in a major road tunnel incident), and short videos (eg, showing an example of fire development in a road tunnel fire) combined with module-specific quizzes (respondents needed to score 80% correct answers to pass each module) to facilitate learning [36] (refer to Multimedia Appendix 3 for an example).

Road tunnel layouts, different injury patterns, collaboration, and aspects of personal and scene safety were integrated into the e-learning course modules. Using the revised version of the taxonomy of learning by Bloom, which was proposed by Krathwohl [37], study participants needed to *understand* (eg, the layout of the tunnel, tunnel-specific risks, and probable injury patterns), *analyze* (how to collaborate with other organizations to get a clearer picture of the incident and incident-specific risks), and *evaluate* (the tunnel layout, risks, and specific injury patterns) to be able to make an informed decision regarding where the injured person should be transported for definitive care. As decision-making is a complex cognitive process, different aspects were covered in the e-learning course. For example, risk management included the identification of environmental risks (smoke and leakage of flammable or corrosive fluids) and personal risks (inhalation of toxic fumes and limited visibility in a smoke-filled environment). Furthermore, risk-mitigation measures (eg, how to position ambulances in the tunnel system) and tunnel safety features (fire suppression systems and jet fans) were described. Under the course segment "Should I enter the road tunnel?" participants were informed of the importance of conducting a risk assessment before entering the road tunnel system and that using the unaffected tunnel tube (in case of a fire or major incident) was considered safe.

The completed e-learning course was validated by 5 EMS staff members (who also validated the control course) based on language, logical progression, design, and overall usability. The overall impression of the course design was described as "appealing." The course was described as "easy to navigate" and "fun," although "not designed to be used via smartphone." Changes were made according to validating EMS staff members feedback; for example, those included in the study were advised not to use smartphones while taking their allotted courses or simulations.

Prospectively Controlled Trial

The trial consisted of a knowledge test based on the tunnel-specific intervention's course content, which was tested in 2 digital simulations (assessment points) conducted 1 month and 6 months after the intervention. The simulations also included a time measurement for decision-making. Time was measured individually for the 15 separate decisions in each simulation. Between the first and second knowledge tests (assessment points), there was no control over for what the participants read, studied, or discussed in relation to tunnel incident responses.

Knowledge Test

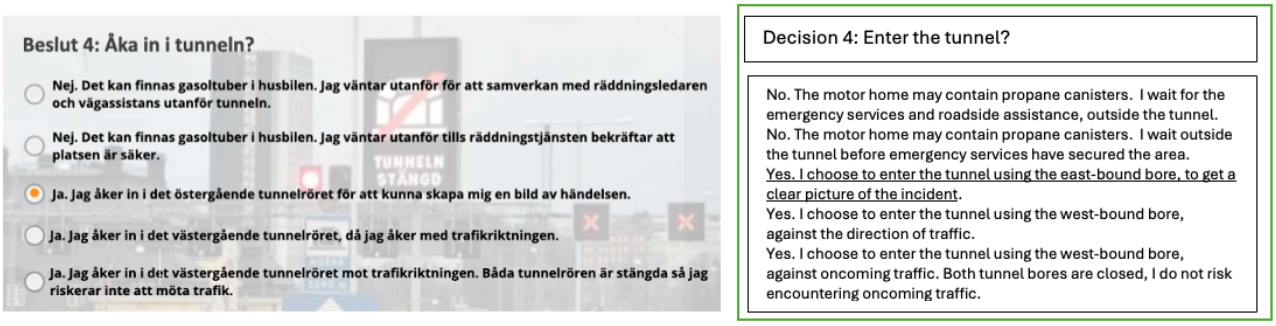
The knowledge test was developed based on the literature on performance indicators [38-40], interview findings [18,19,31], and focus group discussions among stakeholders [13]. A total of 15 independent multiple-choice questions (MCQs), each with 1 correct option out of 5 alternatives, were created. This design was chosen because MCQs are commonly used in medical education to assess higher-order thinking, so the participating personnel should be accustomed to the method [41]. In total, the participants could get 15 points per simulation. The questions were formulated as decision points (Multimedia Appendix 4) and were directly related to the content of the intervention (e-learning course). The questions were evaluated during the pilot testing of the digital simulations.

XSL•FO
RenderX

Simulations

The digital simulations were created in collaboration with an expert in pedagogy and by integrating learning materials into an e-learning environment. The content was created using Microsoft PowerPoint (version 16.71) and modified using Storyline 360 (Articulate Global, LLC). The 2 simulations of major road tunnel incidents were visualized through a combination of text, images, and videos (eg, exercises); audio (dispatch calls and alarm signals); and animations to add a sense of realism. To mimic a real tunnel incident, the simulations followed the same logical, decision-based pathway, starting from dispatch and continuing through to the conclusion of a rescue effort. One example of a decision-based question is "Do you enter the road tunnel"? (Figure 2).

Figure 2. Decision from one of the simulations on how to enter the road tunnel, designed as a multiple-choice question with 5 different choices, one of which is correct. The text is provided in Swedish, with the corresponding English translation shown to the right (the correct answer is underlined).



The simulations were designed as forward-moving only. The aforementioned kind of modified essay question is commonly used in medical education, where students are presented with a case and gradually presented with more information as the case progresses, similar to unfolding an incident where initial information may be scarce [42]. The scenario unfolds regardless of what the participant answers, and the participant cannot go back and change answers to the previous questions. During the simulation, participants received updated information via an audio clip (eg, prerecorded dispatch call) or an image containing information on certain decisions made by others, for example, "the road traffic center decided to close the road tunnel." Actual footage or video sequences from road tunnel interiors were also included.

Two different simulations (1 month and 6 months after the educational intervention) were created. In the first assessment (at 1 month), the simulated road tunnel incident was a collision between a bus and a passenger car, with numerous casualties. The scenario did not include fire; smoke; or added risks, such as leakage of fuels or dangerous cargo. In the second assessment (at 6 months), the scenario was a collision between a mobile home and multiple vehicles, again with numerous casualties. The mobile home contained a liquefied petroleum gas cylinder as an added risk element. This scenario developed into a road tunnel fire at a later stage. These 2 scenarios were chosen based on the experience that most fatalities in tunnel incidents result from traffic crashes and fire [43].

The simulations were validated separately by 5 EMS personnel (external validation) and 2 stakeholders (internal validation) who had in-depth education in prehospital management and extensive experience in acting as ambulance commanders. Their feedback included comments such as "The time pressure was noticeable, I chose to read instead of listen...I lost focus...But the same goes for a real incident." After the feedback was considered and changes were made to the simulations, a final version was established by the coauthors. No changes were made to either course once data collection from study participants was started.

Besides the choice of decision in each simulation, the time taken to make each decision was also recorded. This was a feature available in the software used in the development of the simulations (Storyline 360). When the participant was presented with the MCQ options, a hidden timer started, and the timer stopped when the participant chose an alternative and clicked on "next" to advance to the next question. The timing for each question was validated separately by the first author and the e-learning expert using stopwatches. Time was also measured during the evaluation of the simulations.

Study Sequence

Individual emails containing information about the study, together with a unique link to either the control or intervention course, were sent to the participants. When the participants clicked on the link, they were directed to a website (where the course platform was located) and prompted to log in. After logging in and before gaining access to their allotted course, the participants were asked to answer a short survey regarding their demographics: sex, level of education (registered nurse or emergency medical technician), additional 2-day leadership course, the number of years in EMS, and experience in managing tunnel incidents. Furthermore, the participant had to give consent by ticking a checkbox before completing the survey. Upon completion of the survey, participants could open their respective courses and had 2 weeks to access the content and complete the course.

Approximately 1 month after the participants completed their respective courses, they were sent an email with an invitation to log onto the e-learning website to complete a simulation and assessment, again with the expectation of participating within 2 weeks. As described earlier, the scenario contained 15 decision points where the participant could decide on the best course of action, choosing from multiple choices, with 1 correct answer.

XSL•FO
RenderX

The participants' responses and the time taken to make each decision were recorded.

Six months after the participants completed their respective courses, they were sent an email prompting them to log in to the e-learning website to complete the second simulation and assessment (scenario details given earlier) and again within 2 weeks. Participants who had provided their phone numbers were sent an SMS text message reminder. The participants' decision responses were recorded in the same manner.

Evaluation Form

Upon completion of the second simulation, the participants were redirected to an evaluation form and asked to subjectively assess (decreased, unchanged, or increased) whether the course (intervention or control) influenced their ability to make decisions and affected their sense of security as ambulance commanders in road tunnel incidents. In total, 2 email reminders of participation were sent after 5 days and 10 days. The data collection period was extended by 1 week because of a limited response rate, and the time extension yielded 5 additional responses.

Follow-Up Interviews

A limited number of in-depth follow-up interviews were conducted with participants (4/10, 40% from the intervention group and 1/10, 10% from the control group) within 2 months from the second simulation. The interviews were conducted using a semistructured interview guide and focused on participants describing their reasoning behind the decision to enter or not enter the road tunnel in both scenarios (Multimedia Appendix 5). The collected data were analyzed using the method of qualitative content analysis [44], where the transcribed data were broken down (deconstructed) into meaning units. Next, each meaning unit was labeled with a code, and similar codes were grouped into subcategories and categories.

Outcomes and Analysis Plan

Outcomes

The outcomes were defined as choosing the correct decision when presented with MCQs and the time spent making each decision. The first primary outcome was the correctness of responses to the question concerning the ambulance commander's decision to enter the road tunnel. The second primary outcome was the number of correct responses across the 15 different decision points. The secondary outcome was the time taken to register each decision.

After the conclusion of the simulation-based assessments, participants were asked to self-evaluate whether they felt that the e-learning course or control course influenced their ability to make decisions in general in road tunnel incidents (decreased, unchanged, or increased) as well as influenced their sense of safety in acting as ambulance commanders in road tunnel incidents (decreased, unchanged, or increased).

Data Curation and Availability

Pseudonymized participant test scores and response times were electronically recorded. The dataset was stored on an encrypted hard drive (2 TB; ADATA Technology Co, Ltd).

Statistical Analysis

Statistical analysis was performed using SPSS Statistics software (Macintosh version 28; IBM Corp). A 2-sample proportion test was used to compare the grouped averages of proportions. $P<.05$ was considered significant for identifying the difference between groups. Multivariate logistic regression was used to assess the associations between the primary outcome and participant factors as well as between secondary outcomes and participant factors.

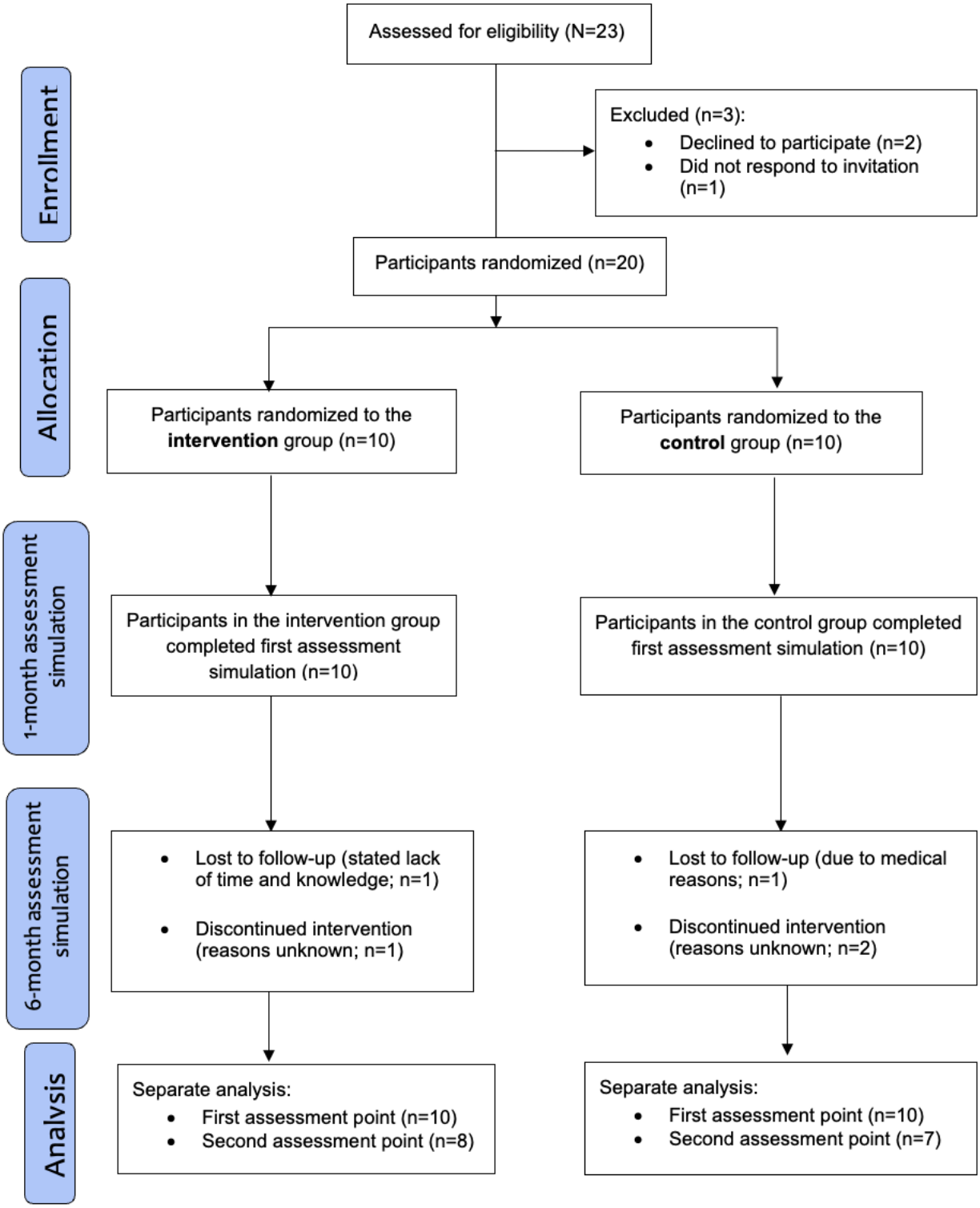
Results

Participants

A total of 20 participants participated in the course to which they were assigned and the first simulation-based assessment. Of these, 16 (80%) participants, with 8 (40%) in each group, participated in the second simulation-based assessment at 6 months. However, 1 (5%) participant in the control group did not complete the assessment, and data from this participant were not included in the grouped responses for the second assessment (Figure 3). Participant demographics, such as sex, years of experience, and educational level, are provided in Table 1.

XSL•FO
RenderX

Figure 3. Study flowchart. Description of the number of participants enrolled and lost to follow-up during the 6-month study period.



XSL•FO
RenderX

Table 1. Participant demographics, including sex, years of experience, level of education, education in prehospital management, whether they have attended a 2-day separate leadership course, and whether they have experience in managing road tunnel incidents (N=20).

Descriptive data	Intervention group (n=10), n (%)	Control group (n=10), n (%)
Sex		
Female	2 (20)	2 (20)
Male	8 (80)	8 (80)
Experience (y)		
5-10	4 (40)	6 (60)
10-15	3 (30)	1 (10)
15-20	2 (20)	0 (0)
>20	1 (10)	3 (30)
Level of education		
Registered nurse	1 (10)	0 (0)
Specialist nurse	8 (80)	8 (80)
EMT[a]	1 (10)	2 (20)
Education in prehospital management	9 (90)	10 (100)
Attended a 2-day separate leadership course	2 (20)	2 (20)
Have experience in managing road tunnel incidents	3 (30)	3 (30)

[a]EMT: emergency medical technician.

Main Results

The main finding was that none (0/10, 0%) of the participants in the control group (without the tunnel-specific e-learning course) and the intervention group (with the tunnel-specific e-learning course) responded correctly to the single categorical question to enter the road tunnel. The second main finding was that the average number of correct answers among the 15 decision points was 7.0 (SD 1.8) for the control group at the first assessment point (1 month) and 6.5 (SD 1.8) for the intervention group, with no difference between the groups (*P*=.86, 2-sample test of proportions).

Secondary Results

Tunnel-specific e-learning course exposure was not associated with more correct answers in the first or second simulation assessments when participant factors were considered in the multivariable logistic regression analysis (Table 2). Participant factors included sex, work experience, educational level, participation in a separate 2-day leadership course, and having practical experience in tunnel management. The findings showed that there were associations between more correct answers and having participated in a separate 2-day leadership course, female sex, and practical experience, at least at the first assessment point. The total response time at the first assessment point was not associated with course exposure (*P*=.80), even when considering other participant factors.

At the 6-month assessment point, 25% (2/8) of the participants in the intervention group and 14% (1/7) in the control group chose to correctly enter the road tunnel, with no difference between the groups (*P*=.59). Multivariable logistic regression analysis identified no factors associated with the total number of correct responses for either group.

Of the 20 participants, the evaluation form was completed by 11 (55%) participants (control group: n=6, 30% and intervention group: n=5, 25%). In the intervention group with 5 participants, confidence in acting as ambulance commanders was reported to be unchanged (n=3, 60%) or to have increased (n=2, 40%). In the control group with 6 participants, the confidence was reported to be unchanged (4/6, 67%) or to have increased (n=2, 33%). Regarding the course's impact on decision-making, 3 (60%) participants in the intervention group reported it as unchanged, and 2 (40%) reported it as increased. By contrast 2 (33%) participants in the control group reported it as unchanged, and 4 (67%) reported it as increased. None of the participants reported decreased ability to make decisions or confidence in handling road tunnel incidents.

The follow-up interviews identified 3 categories linked to the decision not to enter the road tunnel correctly: *information* (lack of updated information, risk of misunderstandings, and unavailability of information on the actual tunnel system), *risks* (limited knowledge of site risks, fear of being questioned by EMS personnel if they decided to enter the tunnel, and lack of proper equipment), and *mitigation* (need of site-specific maps, access to updated course material annually, and access to aide-mémoire).

XSL•FO
RenderX

Table 2. Multivariable logistic regression (including demographic data, such as sex, work experience, educational level, additional leadership course, and experience with tunnel incidents), number of correct answers, and total response time at both assessment points.

Assessment point	Unstandardized β	B, 95% CI	*P* value
Assessment point 1			
Total correct answers			
Course exposure	–.58	–1.88 to 0.73	.36
Sex	2.10	0.39 to 3.79	.20
Work experience	–1.25	–2.18 to –3.25	.01
Educational level	.75	–0.23 to 1.73	.12
Attended a 2-day separate leadership course	3.05	0.55 to 5.55	.02
Having practical experience in tunnel management	1.09	–0.46 to 2.64	.15
Total response time (s)			
Course exposure	44.18	–320.59 to 408.96	.80
Sex	–216.16	–692.90 to 260.59	.35
Work experience	–42.72	–302.35 to 216.91	.73
Educational level	–71.07	–346.09 to 665.95	.59
Attended a 2-day separate leadership course	–35.22	–736.39 to 665.96	.92
Having practical experience in tunnel management	308.37	–125.75 to 742.48	.15
Assessment point 2			
Total correct answers			
Course exposure	–.28	–2.91 to 2.36	.82
Sex	1.18	–3.20 to 5.57	.55
Work experience	.30	–3.03 to 3.09	.98
Educational level	–.57	–2.52 to 1.38	.42
2-day separate leadership course	–3.21	–11.87 to 5.45	.52
Practical experience in tunnel management	.98	–2.60 to 4.56	.55
Total response time (s)			
Course exposure	151.07	–236.08 to 538.21	.39
Sex	42.21	–603.22 to 687.63	.88
Work experience	–420.39	–870.51 to 29.74	.06
Educational level	162.94	–123.93 to 449.80	.23
2-day separate leadership course	1255.37	–19.22 to 2529.95	.05
Practical experience in tunnel management	531.25	4.42 to 1058.08	.05

Discussion

Principal Findings

The main finding of this study was that none (0/20, 0%) of the participants correctly answered the question regarding entering the road tunnel at the first assessment point. This was unexpected.

Despite the tunnel-specific e-learning course specifically addressing how to use the unaffected tunnel tube as one of the important course learning goals, it did not succeed in supporting this expected behavior. There are several possible and different explanations for this result. One could be that most (14/20, 70%) participants lacked any experience with real tunnel incidents, and actual experience may be required to decide to enter the road tunnel properly. This is supported by other reports [45,46], which indicate that decisions often need to be based on experience. Inexperienced decision makers may depend on mnemonics and guidelines [47], and inexperienced commanders may become more hesitant in their decision-making if the guidelines or mnemonics are not adjusted to the situation or environment. Easily accessible and adequate guidelines may facilitate decision-making in this particular setting. This is supported by the participant interviews, indicating a need for adequate aide-mémoire tailored to the specific tunnel environment. Adding useful mnemonics to the tunnel-specific e-learning course may facilitate participant decision-making.

XSL•FO
RenderX

Another possible explanation for participants not actively choosing to enter the road tunnel may be concerns about personal safety. This is in line with the findings from tunnel incident reports [13,18] where incident scene safety is an important concern for the ambulance commander. In addition, it can be perceived as correct to initially wait when there is limited information about site risks. This was illustrated by the aforementioned 2013 Gudvanga tunnel fire where a large portion of the tunnel became filled with smoke due to reversed jet fans [10]. However, in the event of a fire in a twin-tube road tunnel, the unaffected tube should be regarded as a safe environment, as fans ventilate air in the same direction in both tubes, which is expected to prevent smoke spillover to the opposite tube. The unaffected tube as a safe environment was emphasized multiple times during the tunnel-specific e-learning course, which may have influenced the participants to adopt a more cautious approach. Using the opposite tube would also result in bringing the EMS personnel closer to the incident, which could be beneficial for the treatment of critical injuries, such as internal bleeding or smoke inhalation injuries. It seems that this approach needs to be addressed further in future educational efforts. For example, on-site visits could be an important element of education and provide the EMS personnel with a sense of security, which has been indicated by others [18]. Other cost-effective solutions, such as "visiting" road tunnels in a virtual environment, may create a sense of understanding of the tunnel layout.

Other possible explanations for participants choosing not to enter the road tunnel in the simulations could be related to the technical interface of the simulation medium, simulated scenarios, or the technical aspects of the assessment, which were based on MCQs. Because of this design, we could not confirm what the study participants understood in a granular manner. Participants may have considered the correct answer to the question concerning tunnel entry. There is a 20% chance of answering correctly by simply guessing. However, in this case, 0% (0/20) did, which may indicate a conscious choice to not enter the road tunnel. However, there might be challenges regarding responder clarity with MCQ formats, where the responder simply failed to convert their reasoning to an intended MCQ response. MCQs can often be improved for clarity [48,49], for example, by gathering experience with responses and modifying questions to improve precision in assessing specific reasoning [50]. Ideally, a more detailed and interactive verbal assessment could provide a more in-depth understanding of participants' reasoning to assess their competence [51]. In this first evaluation of the simulations, MCQs were chosen to limit subjective assessor bias and facilitate participant accessibility.

Another possible explanation for not choosing to enter the road tunnel early is that the tunnel-specific e-learning course was not sufficiently clear or effective in generating specific knowledge among the participants on when to enter the road tunnel system. Although the course content was reviewed before this study, there was general agreement among the convenience-sampled stakeholders that the learning objectives were clear. It is possible that there was some knowledge decay over the weeks before the 1-month assessment. During the course, learning objectives were reinforced at multiple points. However, it is possible that the learning objectives were not adequately reinforced during the course. This result is important for the further development of this type of e-learning course, which is potentially going to be part of the formal training within the EMS system as commissioned by the Swedish Transport Administration.

Both the tunnel-specific e-learning course and the assessment were conducted in a digital environment, and the practical details of how the participants connected to the course and assessment could have influenced their learning and performance. The e-learning course and simulations used in this study were designed for use on PCs; however, how the participants accessed them was not controlled or reported. Participants' experiences with e-learning formats, materials, or assessment methods are likely to influence learning and performance, where less familiarity and comfort with e-learning media could impede learning. A Cochrane review concluded that e-learning was no more (or less) effective than traditional learning in improving health professionals' skills [52]. Developing this type of course or preparing for responses in complex and dynamic settings is challenging [53]. The applicability of a tunnel-specific e-learning course or SOP in a real road tunnel incident in Sweden needs to be studied further.

Secondary Results

There was poor performance at the 6-month assessment point in both groups for the primary outcome, and the total number of correct responses did not differ between groups or between the same group at 1 month. Furthermore, secondary findings showed that participant factors, such as sex, occupational experience, and previous leadership course experience, were associated with somewhat better performance in general for the total number of correct decisions made. This can be relevant to how receptive the participants are to the tunnel-specific e-learning content. One notable secondary finding was that an additional course in leadership was associated with poorer performance in the assessment, although this was a very limited sample and may not be a true finding. It may be relevant to try to determine precourse receptivity to specific learning goals at some depth to inform course content and logistics.

Despite not observing an intervention effect in this study, participants reported that their ability to make decisions or act as ambulance commanders in road tunnel incidents increased or remained unchanged after the trial, indicating boosted or unaffected professional confidence, respectively. This personal trait (confidence), together with work experience, has been described as important for ambulance commanders in the Swedish EMS [54]. These aspects need to be included when designing this type of highly specialized e-learning course so that both experienced and inexperienced staff may benefit.

The in-depth follow-up interviews offered some insights into the reasoning behind the main outcome. In addition to the aide-mémoire, lack of information (including risks) has been identified as an issue. This is concurrent with recent findings from interviews with other emergency services than the EMS responding to road tunnel incidents [31]. A possible support tool could be an application for joint information sharing, similar to those currently used in the United Kingdom [55]. Timely

XSL•FO
RenderX

access to adequate information may aid in the decision-making process and result in an informed (and quicker) decision to enter the road tunnel.

Ambulance commander performance has been assessed in different ways and reported [56] in web surveys [54] and interviews [42], with performance indicators for completing certain tasks, such as sending the first report to dispatch or formulating guidelines for response within a predetermined time frame [38-40]. Assessing performance indicators has been used as a method to evaluate incident management in Sweden [15]. This study presents a new method for evaluating ambulance commander performance using simulation-based assessment, which measures both the time and choice of decisions. With this method, each decision can be evaluated independently, potentially allowing the identification of particularly time-sensitive and difficult decisions. This type of detailed information can help to identify specific issues or points that might benefit from extra focus and resources.

Limitations

In the context of this newly designed and tested e-learning course and the limited achievement of the learning outcomes, several study design limitations can be discussed. Despite substantial effort in validating the course content using external (EMS personnel outside the studied region) and internal (EMS personnel with in-depth training and experience in decision-making) methods, limitations in the validity and reliability of the simulations and assessment methods may have introduced systematic errors into the results. Regarding the primary outcome, the assessment method may not have detected course learning with sufficient granularity.

From a pedagogical standpoint, different methods of how study participants were presented with learning materials were used. The control group received a short lecture (passive learning), and the intervention group was active and participated in a gamified learning environment. This difference in how learning materials were provided has not been studied, as the content of the tunnel-specific e-learning course and its influence on decision-making was the main focus. However, research findings indicate that students participating in active learning environments learn more but have a lesser feeling of learning compared to students exposed to passive learning [57]. A feeling of learning less in an educational setting could have affected the participants' decision-making abilities. In future revisions of this tunnel-specific e-learning course, the method of delivery needs to be considered. For example, using a blended learning model (including elements of both passive and active learning) may have a positive impact on learning and critical thinking [58]. In addition, more formative assessment steps in the e-learning–focused tunnel response could provide more reinforcement of primary learning goals, even though it was thought that this aspect was adequate before the study.

Using knowledge tests (MCQs) after 1 and 6 months could have influenced the result. It is known that knowledge and clinical skills decay over time [59,60]. If the first simulation had followed directly after the completion of the allotted courses, a different outcome may have been observed. This study design was chosen because there may be considerable time between an educational course and a real incident, leading to some knowledge decay. For example, major road tunnel incidents are uncommon in Sweden. During the years 2003 to 2013, a total of 926 incidents occurred in 10 of Sweden's approximately 20 road tunnels (length of >300 m). In these incidents, 525 injuries were reported. Most (n=523) were minor injuries (injury severity score of 1-8), and no deaths were reported [61]. Hence, it may be some time before participants can apply their knowledge in a real-life setting, and retaining skills for nonroutine situations is difficult [62]. Still, none (0/20, 0%) of the participants correctly entered the road tunnel, which indicates a conscious choice. Participants were not given the correct answers after the first simulation, as this may have influenced them to change their answers in the second simulation. Further development of the e-learning course is needed to facilitate the decision to correctly enter the road tunnel in case of a major incident.

The evaluation process of the tunnel-specific e-learning course and simulation could have benefited from a more granular assessment, such as using the 4-level (reaction, learning, behavior, and results) model of assessing the effectiveness of training programs proposed by Kirkpatrick [63]. In this study, level 1 was assessed during the development phase of the tunnel-specific e-learning course. Level 3 of the Kirkpatrick evaluation model has been addressed through participants' self-assessed responses on how their ability to make decisions or sense of security had changed. Level 2 of the Kirkpatrick evaluation model (learning) could have been addressed properly by adding a third question to the evaluation form, asking whether the participants' knowledge of managing road tunnel incidents decreased, remained unchanged, or increased. Some aspects concerning level 2 of the Kirkpatrick evaluation model were discussed by the participants in the follow-up interviews. However, a more structured evaluation could have been used to assess learning. This element should be incorporated in future revisions of the tunnel-specific e-learning course.

Other aspects that could potentially disturb learning goal assessment in this study include external participant-specific factors or human factors, such as whether participants were unmotivated, tired, distracted, or stressed when attending the course or the simulations. Adding images, videos, and sounds to the simulation as part of realism may have also affected the participants' focus or stress levels, for example, as information overload can degrade performance. The National Board of Medical Examiners addresses this aspect of combining media and MCQs and recommends choosing video clips with fewer distracting features [64]. Furthermore, a real road tunnel incident may include noxious sensory inputs, such as smoke and high noise levels. There must be a balance in designing training or education to incorporate enough realism without overwhelming the participant with excessive input.

Ambiguity or difficulty in interpreting the MCQs could have contributed to systematic errors in the results. More extensive work in formulating and evaluating MCQs may be an additional step in future validation processes to support both formative and normative assessment in this type of context. MCQs have the advantage of standardization and efficient data collection but, at the same time, must be carefully constructed to ensure high validity concerning specific learning goals [50]. MCQs

XSL•FO
RenderX

can be useful for simple and rapid performance assessments, although other forms of assessment can provide more granular data concerning learning phenomena [51]. However, in this serious incident context, where immediate binary decisions may be needed in the setting of complex and incomplete information, MCQ assessment was judged to be similar to the rapid decision requirements in situations faced by ambulance commanders.

Applicability

This is the first iteration of a course intended for further development based on this and other validation work for learning effectiveness in this specific, complex environment and professional response. A final e-learning course could be used to give EMS personnel a basic understanding of the tunnel environment and various risks. The final version of the e-learning course could include tunnel-specific plans and aide-mémoire and be used as a time- and cost-efficient complement to other training methods, such as tabletop or full-scale exercises. The findings in this study may be seen as a proof of concept or demonstration of feasibility for course delivery and assessment. Further evaluation and reconstruction are warranted. For example, by incorporating a simulation as the first step (and providing the user feedback on the result), the EMS personnel could be able to identify individual areas of improvement (eg, a lack of risk awareness). This feature may be useful so that personnel with different levels of experience can focus on specific sections of the course. After completing the course, EMS personnel could take a second simulation. Results from both simulations can be used to identify individual progression (or regression). Using this method, a partial individual tailoring for this kind of educational course could be constructed. Thus, the findings indicate that both the gamified e-learning content and assessments can serve as good learning alternatives. The limitations presented in this study may provide some support for other researchers considering the development of e-learning systems for content areas involving complex interventions in professional circumstances.

Conclusions

Participation in a tunnel-specific e-learning course did not provide a measurable change in ambulance commanders' decision-making behavior in simulated road tunnel incidents. The unwarranted reluctance of ambulance commanders in these simulations to enter the described twin-tube road tunnel system could lead to unnecessary delays in the treatment of time-sensitive injuries, with preventable morbidity and mortality, if this were to occur in a real-life situation. This type of assessment of commander performance could be applicable in a similar educational context.

Acknowledgments

The authors would like to express appreciation for the participants who contributed to this study. The authors would also like to thank Satish Strömberg for digitalizing the simulations and adapting the learning materials to an e-learning environment, Britt-Inger Saveman for contributing to the study design and constructive discussions on the finalization of the manuscript, and Magnus Hultin for contributing to constructive discussions on the validity and reliability of the results. This study was funded by the Swedish Transport Administration and Region Västerbotten, Sweden. The funding agencies had no role in or impact on the study design, data collection, data analysis, or writing of the manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

JH, LG, and MH contributed to the planning of the study, data collection, data analysis, and writing the manuscript. AW contributed to supervision, planning of the study, data collection, data analysis, and writing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Written information (in Swedish) sent to potential participants before enrollment.
[PDF File (Adobe PDF File), 116 KB - formative_v9i1e58542_app1.pdf]

Multimedia Appendix 2
Content outline of the tunnel-specific e-learning course.
[DOCX File , 19 KB - formative_v9i1e58542_app2.docx]

Multimedia Appendix 3
Screenshots from the course module 3 (in Swedish).
[DOCX File , 994 KB - formative_v9i1e58542_app3.docx]

XSL•FO
RenderX

Multimedia Appendix 4
Decisions in the simulations.
[DOCX File , 14 KB - formative_v9i1e58542_app4.docx]

Multimedia Appendix 5
Semistructured interview guide (in Swedish).
[DOCX File , 18 KB - formative_v9i1e58542_app5.docx]

Multimedia Appendix 6
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 481 KB - formative_v9i1e58542_app6.pdf]

References

1. 68% of the world population projected to live in urban areas by 2050. United Nations. URL: https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html [accessed 2024-01-30]
2. Broere W. Urban underground space: solving the problems of today's cities. Tunn Undergr Space Technol 2016 May;55:245-248. [doi: 10.1016/j.tust.2015.11.012]
3. INRIX: congestion costs each American 97 hours, $1,348 a year. INRIX. URL: https://inrix.com/press-releases/scorecard-2018-us/ [accessed 2024-01-30]
4. Cui J, Nelson JD. Underground transport: an overview. Tunn Undergr Space Technol 2019 May;87:122-126. [doi: 10.1016/j.tust.2019.01.003]
5. E39 Rogfast. Statens Vegvesen. URL: https://www.vegvesen.no/vegprosjekter/europaveg/e39rogfast/ [accessed 2024-01-30]
6. Gandit M, Kouabenan DR, Caroly S. Road-tunnel fires: risk perception and management strategies among users. Saf Sci 2009 Jan;47(1):105-114. [doi: 10.1016/j.ssci.2008.01.001]
7. Lyon RM, Sanders J. The Swiss bus accident on 13 March 2012: lessons for pre-hospital care. Crit Care 2012 Jul 11;16(4):138 [FREE Full text] [doi: 10.1186/cc11370] [Medline: 22784360]
8. Li YZ. Study of fire and explosion hazards of alternative fuel vehicles in tunnels. Fire Saf J 2019 Dec;110:102871. [doi: 10.1016/j.firesaf.2019.102871]
9. Ouyang Z, Carlton A, Guo Q, Naito C, Quiel S. Blast vulnerability of drop ceilings in roadway tunnels. J Perform Constr Facil 2020 Dec;34(6):04020121. [doi: 10.1061/(asce)cf.1943-5509.0001526]
10. Report road 2015/02 report on fire in a heavy goods vehicle in the Gudvanga tunnel on the E16 road in Aurland on 5 August 2013. Accident Investigation Board Norway. 2015. URL: https://havarikommisjonen.no/Vei/Avgitte-rapporter/2015-02-eng?pid=SHT-Report-ReportFile&attach=1 [accessed 2024-04-29]
11. Kristiansen E, Johansen FH, Carlström E. When it matters most: collaboration between first responders in incidents and exercises. J Contingencies Crisis Manag 2018 Jul 04;27(1):72-78. [doi: 10.1111/1468-5973.12235]
12. Doohan IS, Saveman BI, Gyllencreutz L. Limited medical perspective at a strategic level in relation to mass casualty incidents in Swedish tunnels. Int J Emerg Manag 2019;15(4):360. [doi: 10.1504/ijem.2019.104205]
13. Eklund A, Karlsson S, Gyllencreutz L. Building "common knowledge" when responding to major road tunnel incidents: an inter-organisational focus group study. Int J Emerg Serv 2022 Oct 24;12(2):145-160. [doi: 10.1108/ijes-02-2022-0006]
14. Rüter A, Nilsson H, Vikström T. Medical Command and Control at Incidents and Disasters: From the Scene of the Incident to the Hospital Ward. Stockholm, Sweden: Studentlitteratur; 2006.
15. Nilsson H, Kristansson T. Katastrofmedicinsk beredskap - att leda sjukvård på ett särskilt sätt. Stockholm, Sweden: Studentlitteratur; 2015.
16. Bjørnsen G, Dettweiler U, Njå O, Knudsen K. Towards an understanding of learning within the Norwegian fire and rescue services – focusing on tunnel fire safety. J Workplace Learn 2022 Sep 22;35(1):112-128. [doi: 10.1108/jwl-04-2022-0051]
17. Bjørnsen G, Billett S, Njå O. First responders' perceived and actual competence in tunnel fire safety. Fire Saf J 2023 Apr;136:103758. [doi: 10.1016/j.firesaf.2023.103758]
18. Hylander J, Saveman BI, Björnstig U, Gyllencreutz L. Prehospital management provided by medical on-scene commanders in tunnel incidents in Oslo, Norway - an interview study. Scand J Trauma Resusc Emerg Med 2019 Aug 20;27(1):78 [FREE Full text] [doi: 10.1186/s13049-019-0649-8] [Medline: 31429788]
19. Hylander J, Saveman BI, Björnstig U, Gyllencreutz L. Senior ambulance officers in Swedish emergency medical services: a qualitative study of perceptions and experiences of a new management role in challenging incidents. BMJ Open 2020 Dec 07;10(12):e042072 [FREE Full text] [doi: 10.1136/bmjopen-2020-042072] [Medline: 33293325]
20. Baetzner AS, Wespi R, Hill Y, Gyllencreutz L, Sauter TC, Saveman BI, et al. Preparing medical first responders for crises: a systematic literature review of disaster training programs and their effectiveness. Scand J Trauma Resusc Emerg Med 2022 Dec 24;30(1):76 [FREE Full text] [doi: 10.1186/s13049-022-01056-8] [Medline: 36566227]
21. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]

XSL•FO
RenderX

22. Eysenbach G, CONSORT-eHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
23. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
24. Clinical trials registration. ICMJE. URL: https://www.icmje.org/about-icmje/faqs/clinical-trials-registration/ [accessed 2024-01-30]
25. Advanced Life Support Group (ALSG), Gleeson T, Mackway-Jones K. Major Incident Medical Management and Support: The Practical Approach at the Scene, 4th edition. Hoboken, NJ: Wiley-Blackwell; 2022.
26. About moodle. MoodleDocs. URL: https://docs.moodle.org/400/en/About_Moodle [accessed 2022-11-28]
27. Sollid SJ, Rimstad R, Rehn M, Nakstad AR, Tomlinson AE, Strand T, Collaborating group. Oslo government district bombing and Utøya island shooting July 22, 2011: the immediate prehospital emergency medical service response. Scand J Trauma Resusc Emerg Med 2012 Jan 26;20(1):3 [FREE Full text] [doi: 10.1186/1757-7241-20-3] [Medline: 22280935]
28. Utvärdering av hanteringen av attentatet i Stockholm 7 April 2017: redovisning av regeringsuppdrag Ju2017/05643/SSK. The Swedish Civil Contingencies Agency. 2018. URL: https://www.msb.se/sv/publikationer/utvardering-av-hanteringen-av-attentatet-i-stockholm-7-april-2017--redovisning-av-regeringsuppdrag-ju201705643ssk/ [accessed 2024-04-29]
29. Nilsson H. PS- Prehospital Sjukvårdsledning Ett nationellt koncept. Kunskapscentrum för Katastrofmedicin (KcKm) Linköpings Universitet. URL: https://psconcept.se/wp-content/uploads/PS-koncept.pdf [accessed 2025-01-26]
30. Gamification. Oxford Advanced Learner's Dictionary. URL: https://www.oxfordlearnersdictionaries.com/definition/english/gamification [accessed 2024-01-30]
31. Hylander J, Saveman BI, Björnstig U, Gyllencreutz L, Westman A. Time-efficiency factors in road tunnel rescue as perceived by Swedish operative personnel – an interview study. Int J Emerg Serv 2022 Feb 02;11(2):312-324. [doi: 10.1108/ijes-03-2021-0011]
32. Bazyar J, Farrokhi M, Khankeh H. Triage systems in mass casualty incidents and disasters: a review study with a worldwide approach. Open Access Maced J Med Sci 2019 Feb 15;7(3):482-494 [FREE Full text] [doi: 10.3889/oamjms.2019.119] [Medline: 30834023]
33. Anseeuw K, Delvau N, Burillo-Putze G, De Iaco F, Geldner G, Holmström P, et al. Cyanide poisoning by fire smoke inhalation: a European expert consensus. Eur J Emerg Med 2013;20(1):2-9. [doi: 10.1097/mej.0b013e328357170b]
34. Mathews ZR, Koyfman A. Blast injuries. J Emerg Med 2015 Oct;49(4):573-587. [doi: 10.1016/j.jemermed.2015.03.013] [Medline: 26072319]
35. van Gaalen AE, Brouwer J, Schönrock-Adema J, Bouwkamp-Timmer T, Jaarsma AD, Georgiadis JR. Gamification of health professions education: a systematic review. Adv Health Sci Educ Theory Pract 2021 May 31;26(2):683-711 [FREE Full text] [doi: 10.1007/s10459-020-10000-3] [Medline: 33128662]
36. Mayer RE. Using multimedia for e‐learning. Comput Assist Learn 2017 Jun 13;33(5):403-423. [doi: 10.1111/jcal.12197]
37. Krathwohl DR. A revision of Bloom's taxonomy: an overview. Theory Pract 2010 Jun 24;41(4):212-218. [doi: 10.1207/s15430421tip4104_2]
38. Gryth D, Rådestad M, Nilsson H, Nerf O, Svensson L, Castrén M, et al. Evaluation of medical command and control using performance indicators in a full-scale, major aircraft accident exercise. Prehosp Disaster Med 2010 Jun 28;25(2):118-123. [doi: 10.1017/s1049023x00007834] [Medline: 20467989]
39. Rådestad M, Nilsson H, Castrén M, Svensson L, Rüter A, Gryth D. Combining performance and outcome indicators can be used in a standardized way: a pilot study of two multidisciplinary, full-scale major aircraft exercises. Scand J Trauma Resusc Emerg Med 2012 Aug 28;20(1):58 [FREE Full text] [doi: 10.1186/1757-7241-20-58] [Medline: 22929479]
40. Rådestad M, Jirwe M, Castrén M, Svensson L, Gryth D, Rüter A. Essential key indicators for disaster medical response suggested to be included in a national uniform protocol for documentation of major incidents: a Delphi study. Scand J Trauma Resusc Emerg Med 2013 Sep 11;21(1):68 [FREE Full text] [doi: 10.1186/1757-7241-21-68] [Medline: 24025366]
41. Javaeed A. Assessment of higher ordered thinking in medical education: multiple choice questions and modified essay questions. MedEdPublish (2016) 2018 Jun 12;7:128 [FREE Full text] [doi: 10.15694/mep.2018.0000128.1] [Medline: 38074575]
42. Hugelius K, Edelbring S, Blomberg K. Prehospital major incident management: how do training and real-life situations relate? A qualitative study. BMJ Open 2021 Sep 23;11(9):e048792 [FREE Full text] [doi: 10.1136/bmjopen-2021-048792] [Medline: 34556511]
43. Bassan S. Overview of traffic safety aspects and design in road tunnels. IATSS Res 2016 Jul;40(1):35-46. [doi: 10.1016/j.iatssr.2016.02.002]
44. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
45. Rimstad R, Sollid SJ. A retrospective observational study of medical incident command and decision-making in the 2011 Oslo bombing. Int J Emerg Med 2015 Mar 4;8(1):4 [FREE Full text] [doi: 10.1186/s12245-015-0052-9] [Medline: 25852774]

XSL•FO
RenderX

46. Gonzalez C, Brunstein A. Training for emergencies. J Trauma 2009 Aug;67(2 Suppl):S100-S105. [doi: 10.1097/TA.0b013e3181adc0d5] [Medline: 19667839]
47. Penney G, Launder D, Cuthbertson J, Thompson MB. Threat assessment, sense making, and critical decision-making in police, military, ambulance, and fire services. Cogn Tech Work 2022 Feb 04;24(3):423-439. [doi: 10.1007/s10111-022-00694-3]
48. Jaap A, Dewar A, Duncan C, Fairhurst K, Hope D, Kluth D. Effect of remote online exam delivery on student experience and performance in applied knowledge tests. BMC Med Educ 2021 Feb 02;21(1):86 [FREE Full text] [doi: 10.1186/s12909-021-02521-1] [Medline: 33530962]
49. Bustraan J, Henny W, Kortbeek JB, Brasel KJ, Hofmann M, Schipper IB. MCQ tests in Advanced Trauma Life Support (ATLS©): development and revision. Injury 2016 Mar;47(3):665-668. [doi: 10.1016/j.injury.2015.11.024] [Medline: 26686592]
50. American Educational Research Association. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association; 2011.
51. Ten Cate O. How can entrustable professional activities serve the quality of health care provision through licensing and certification? Can Med Educ J 2022 Aug 15;13(4):8-14 [FREE Full text] [doi: 10.36834/cmej.73974] [Medline: 36091739]
52. Vaona A, Banzi R, Kwag KH, Rigon G, Cereda D, Pecoraro V, et al. E-learning for health professionals. Cochrane Database Syst Rev 2018 Jan 21;1(1):CD011736 [FREE Full text] [doi: 10.1002/14651858.CD011736.pub2] [Medline: 29355907]
53. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
54. Rosander M, Jonson C. Professional confidence in the roles as ambulance and medical incident commander. J Conting Crisis Manag 2017 Jan 11;25(4):289-300. [doi: 10.1111/1468-5973.12156]
55. Slater A. How ambulance commanders manage a mass casualty incident. Int J Emerg Serv 2022 Apr 27;11(2):235-247. [doi: 10.1108/IJES-03-2021-0012]
56. Rimstad R, Braut GS. Literature review on medical incident command. Prehosp Disaster Med 2015 Apr;30(2):205-215. [doi: 10.1017/S1049023X15000035] [Medline: 25659266]
57. Deslauriers L, McCarty LS, Miller K, Callaghan K, Kestin G. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. Proc Natl Acad Sci U S A 2019 Sep 24;116(39):19251-19257 [FREE Full text] [doi: 10.1073/pnas.1821936116] [Medline: 31484770]
58. Niu Y, Xi H, Liu J, Sui X, Li F, Xu H, et al. Effects of blended learning on undergraduate nursing students' knowledge, skills, critical thinking ability and mental health: a systematic review and meta-analysis. Nurse Educ Pract 2023 Oct;72:103786. [doi: 10.1016/j.nepr.2023.103786] [Medline: 37793247]
59. Srivilaithon W, Amnuaypattanapon K, Limjindaporn C, Diskumpon N, Dasanadeba I, Daorattanachai K. Retention of basic-life-support knowledge and skills in second-year medical students. Open Access Emerg Med 2020;12:211-217 [FREE Full text] [doi: 10.2147/OAEM.S241598] [Medline: 33061682]
60. Offiah G, Ekpotu LP, Murphy S, Kane D, Gordon A, O'Sullivan M, et al. Evaluation of medical student retention of clinical skills following simulation training. BMC Med Educ 2019 Jul 16;19(1):263 [FREE Full text] [doi: 10.1186/s12909-019-1663-2] [Medline: 31311546]
61. Strömgren P, Berg S, Ekman A, Bergström K. Slutrapport - Tunnlars framkomlighet och säkerhet. Stockholm, Sweden: Trafikverket; 2014.
62. Klostermann M, Conein S, Felkl T, Kluge A. Factors influencing attenuating skill decay in high-risk industries: a scoping review. Safety 2022 Mar 28;8(2):22. [doi: 10.3390/safety8020022]
63. Kirkpatrick D. Great ideas revisited: techniques for evaluating training programs. Train Dev 1996;50(1):54-57 [FREE Full text]
64. NBME Item-writing guide: constructing written test questions for health sciences. 6th edition. National Board of Medical Examiners. URL: https://www.studocu.com/row/document/jamaa%D8%A9-alkahr%D8%A9/medical-education/nbme-item-writing-guide-2020/60196327 [accessed 2024-04-29]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
EMS: emergency medical services
MCQ: multiple-choice question
METHANE: major incident, exact location, type of incident, hazards, access and egress, number of casualties, and emergency services
MIMMS: Major Incident Medical Management and Support
SOP: standard operating procedure

XSL•FO
RenderX

Edited by A Mavragani; submitted 18.03.24; peer-reviewed by T Macnow, Y Asada; comments to author 17.09.24; revised version received 27.01.25; accepted 02.03.25; published 28.03.25.

Please cite as:
Hylander J, Gyllencreutz L, Haney M, Westman A
Ambulance Commanders' Reluctance to Enter Road Tunnels in Simulated Incidents and the Effects of a Tunnel-Specific e-Learning Course on Decision-Making: Web-Based Randomized Controlled Trial
JMIR Form Res 2025;9:e58542
URL: https://formative.jmir.org/2025/1/e58542
doi:10.2196/58542
PMID:

©Johan Hylander, Lina Gyllencreutz, Michael Haney, Anton Westman. Originally published in JMIR Formative Research (https://formative.jmir.org), 28.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Videoconference-Delivered Acceptance and Commitment Therapy for Family Caregivers of People With Dementia: Pilot Randomized Controlled Trial

Areum Han[1], PhD; Robert Oster[2], PhD; Hon Yuen[1], PhD; Jeremy Jenkins[3], MS; Jessica Hawkins[1], BA; Lauren Edwards[4], BA

[1]Department of Occupational Therapy, University of Alabama at Birmingham, Birmingham, AL, United States

[2]Division of Preventive Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

[3]Telehealth Private Practice, Billings, MT, United States

[4]Arts in Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

Corresponding Author:
Areum Han, PhD
Department of Occupational Therapy
University of Alabama at Birmingham
SHPB 339
1720 2nd Avenue South
Birmingham, AL, 35294
United States
Phone: 1 2059752882
Email: ahan@uab.edu

Abstract

Background: Family caregivers of individuals with dementia face significant mental health challenges. Acceptance and commitment therapy (ACT) has emerged as a promising intervention for improving these caregivers' mental health. While various delivery modes of ACT have been explored, there is a need for evidence on the efficacy of videoconference-delivered ACT programs for this population.

Objective: This pilot randomized controlled trial, conducted in the United States, aims to assess the effects of a videoconference-delivered, therapist-guided ACT program on reducing depressive symptoms and improving other mental health outcomes among family caregivers with depression who give care to individuals with dementia, compared to a control group that received psychoeducation materials only.

Methods: This 2-arm, parallel-group pilot randomized controlled trial randomly assigned 33 family caregivers to either a 10-week videoconference-delivered ACT program (n=16, 48%) or a control group that received psychoeducation materials alone (n=17, 52%). Depressive symptoms (primary outcome) were measured using the Patient Health Questionnaire-9. Secondary outcomes included anxiety, stress, psychological quality of life (QoL), caregiver burden, predeath grief, guilt, and ACT process measures. Outcomes were assessed in the pretest, posttest (10-12 weeks after pretest), and a 3-month follow-up (3 months after posttest, approximately 5-6 months after pretest). An intent-to-treat approach was used for all outcome analyses. Linear mixed-effects models for repeated measures were used to analyze outcomes.

Results: The ACT group reported significantly greater improvements in stress (P=.043) and psychological QoL (P=.014) in the posttest compared to the control group. Within the ACT group, participants experienced a significant decrease in depressive symptoms, with a mean (SE) change of –6.09 (1.16) points (95% CI –8.42 to –3.76; P<.001) in the posttest and –6.71 (1.45) points (95% CI –9.63 to –3.81; P<.001) in the 3-month follow-up. These changes exceed the estimated minimal clinically important difference on the Patient Health Questionnaire-9. In addition, the ACT group reported significant improvements in anxiety, stress, psychological QoL, caregiver burden, predeath grief, guilt, values-driven action, and experiential avoidance at both posttest and 3-month follow-up. A sensitivity analysis, excluding 1 participant with near-outlier data, revealed statistically significant between-group differences in depressive symptoms at posttest (P=.037); stress at posttest (P<.001) and in 3-month follow-up (P=.001); psychological QoL at posttest (P<.001); caregiver burden at posttest (P=.003) and in 3-month follow-up (P=.003); predeath grief in 3-month follow-up (P=.031); and values-driven action at posttest (P=.032).

XSL•FO
RenderX

Conclusions: The videoconference-delivered ACT program showed promise in improving mental health outcomes and ACT processes among family caregivers with depression who give care to individuals with dementia. Future studies should aim to replicate these findings with larger, more diverse caregiver populations and explore the long-term efficacy of videoconference-delivered ACT programs.

Trial Registration: ClinicalTrials.gov NCT05043441; https://clinicaltrials.gov/study/NCT05043441

(JMIR Form Res 2025;9:e67545) doi:10.2196/67545

KEYWORDS

acceptance and commitment therapy; Alzheimer disease; caregivers; dementia; depression; web-based intervention; quality of life; randomized controlled trial; stress; videoconferencing

Introduction

In the United States, around 11 million individuals offer unpaid care to individuals with Alzheimer disease and related dementias (ADRD), with approximately two-thirds identifying as women [1]. Most of these caregivers are adult children caring for their parents, with spouses making up the next largest group [1]. These family caregivers provide critical support for individuals with ADRD, helping their relatives age in place [2,3]. Therefore, maintaining these caregivers' health is imperative; however, they face significant unmet mental health needs [4]. Recent meta-analyses estimate that approximately 34% of these family caregivers experience depressive symptoms, and 32.1% experience anxiety [5,6]. The demands of caring for individuals with ADRD often prevent caregivers from engaging in self-care and other personally valued activities and can lead to significant caregiver burden, stress, and reduced quality of life (QoL) [5,7-9]. Psychoeducation and skill-building interventions have been the predominant approaches to addressing these caregivers' needs, but meta-analyses indicate that these interventions alone do not significantly improve caregivers' mental health, particularly in alleviating depressive symptoms [10]. In contrast, psychoeducational interventions with psychotherapeutic components, as well as psychotherapy and counseling, have shown small effects in reducing depressive symptoms of these caregivers [10].

Acceptance and commitment therapy (ACT) is an evidence-based, transdiagnostic approach that combines acceptance and mindfulness with behavior change processes to support living in alignment with personal values [11,12]. ACT is grounded in the psychological flexibility model, which encompasses six core processes: (1) acceptance of unwanted thoughts and emotions; (2) cognitive defusion, or detachment from unhelpful thoughts and emotions; (3) present-moment awareness; (4) self-as-context, or seeing oneself as the observer of thoughts and feelings; (5) values clarification; and (6) establishing patterns of behavior that align with values [12]. ACT skills training can help family caregivers manage the difficult thoughts and emotions arising from their caregiving roles, such as depression, anxiety, stress, guilt, and grief while managing caregiving responsibilities alongside other important areas of life [13,14].

To our knowledge, there have been 11 intervention studies investigating ACT for family caregivers of individuals with ADRD, most of which were published within the last 4 years [14-24]. These studies were conducted in Spain [14-17], the United States [18-21], the United Kingdom [22], Germany [23], and the Netherlands [24]. They demonstrated the acceptability and benefits of ACT in improving depressive symptoms, anxiety, stress, caregiver burden, guilt, grief, and ACT processes (eg, experiential avoidance and values-driven action) [14-24]. However, most of these studies were preliminary, with half using one-group pretest-posttest designs as feasibility and pilot studies [17,18,20-22,24], and the rest using randomized controlled trials (RCTs) conducted on relatively smaller scales or resulting in relatively smaller sample sizes at follow-ups [14-16,19-23] compared to typical RCTs. For example, an RCT conducted in Spain randomized 135 caregivers into in-person ACT, in-person cognitive behavioral therapy, and a control condition, and 6-month follow-up data were analyzed for 55 participants in total (25 in the ACT group) [15]. Most importantly, delivery modes varied across studies, including in-person programs [14-17], telephone-based programs [21,23], and internet-delivered programs, such as coach-guided videoconferences conducted by our study team [18,19] and self-help web modules [20,22,24]). More research is needed to understand the effects of ACT delivered in various delivery modes and across different countries for family caregivers of individuals with ADRD. Intervention effects may vary due to cultural influences on caregiving experiences, coping styles, acceptability, and the suitability of different delivery modes based on caregivers' characteristics, including race or ethnicity, preferences, and needs [25].

Web-based delivery of ACT could be an effective alternative to in-person services, enhancing accessibility and scheduling flexibility for caregivers who cannot leave their relatives unattended [26-28]. In particular, meta-analyses of ACT across different populations suggest that therapist-guided, web-based programs have greater effects on depressive symptoms and ACT process measures than unguided programs [26,28]. In addition, the use of a self-help approach alone is limited to individuals with mild depressive symptoms, as it is not recommended for those with more severe symptoms [29]. Individually and remotely delivered ACT programs with therapist guidance can provide higher engagement for these caregivers compared to other internet delivery modes [30]. Therefore, videoconference-delivered, therapist-guided programs may be particularly beneficial for family caregivers who face significant mental health challenges, including depressive symptoms, and require one-on-one real-time interaction with a therapist but struggle to attend in-person therapy sessions [18,19].

XSL•FO
RenderX

This study aimed to assess the effects of a videoconference-delivered, therapist-guided ACT program on mental health outcomes and ACT processes in family caregivers with depression who give care to individuals with ADRD, compared to a control group that received psychoeducation materials alone conducted in the United States. The study builds on insights gained and findings from 2 smaller-scale videoconference-delivered ACT studies conducted by our team [18,19], which informed the decision to focus on caregivers with depressive symptoms and to modify the outcome measures.

Methods

Study Design and Overview

This study used a 2-arm, parallel-group RCT design. Family caregivers of individuals with ADRD were randomly assigned in a 1:1 ratio to either the videoconference-delivered ACT group or the control group that received psychoeducation materials alone. The study was conducted from late March 2022 to mid-June 2024, with data collection and entry spanning approximately 2.2 years, and a recruitment period lasting about 1.7 years.

Ethical Considerations

This study was reviewed and approved by the institutional review board at the University of Alabama, Birmingham (IRB-300008123) and registered at ClinicalTrials.gov (NCT05043441) before participant enrollment. All procedures adhered to ethical guidelines for human participant research. All participants provided informed consent prior to their involvement in the study. To protect participants' privacy and confidentiality, all study data were deidentified before analysis. Data were securely maintained and accessible only to authorized research personnel. Participants received a reloadable ClinCard as compensation for their time—up to US $150 for the intervention group for completing sessions and questionnaires, and up to US $90 for the control group for completing questionnaires only.

Eligibility Criteria

Participants were recruited based on the following inclusion criteria: (1) adults (aged 18 years or older) primarily responsible for the care of a relative with ADRD, (2) reporting at least mild depressive symptoms (Patient Health Questionnaire-9 [PHQ-9] scores≥5 [31]), (3) access to a computer or smartphone with internet connectivity, and (4) ability to provide informed consent. Exclusion criteria included cognitive, physical, or sensory deficits; being a non-English speaker; psychiatric hospitalizations or mental illness diagnoses in the past 2 years; use of antipsychotic or anticonvulsant medications; or plans to place the relative with ADRD in a nursing home within 6 months.

Target Sample Size

The target sample size was adjusted from 64 to 32 due to challenges in recruiting family caregivers with depression within the funding period. The following power calculation was performed to justify the revised sample size; this was done before study completion and before the statistical analyses were performed, and as such, is considered a priori. This calculation represents the possible between-group effect and its corresponding effect size (Cohen *d*). Assuming that we will compare repeated measurements for group means, that there are 2 groups, 3 time points (ie, that there are 3 measurements obtained for each participant), an SD for depressive symptoms (PHQ-9 score of 5.20) [32], a moderate correlation between measurements on the same participant of 0.50, and 13 participants per group (26 for the study), we will have 80% power to detect a statistically significant difference in change scores for PHQ-9 of 4.67 and greater between the 2 groups at a 0.05 significance level (2-sided). This is equivalent to being able to detect a statistically significant effect size of 0.90 and greater. This result compares favorably with a previous ACT intervention study for family caregivers with depression who give care to individuals with ADRD in Spain, which reported an effect size of 1.17 for reducing depressive symptoms in the in-person ACT group compared to a control group at posttest and included participants with posttest data in the analyses [15]. Our power calculation was performed using nQuery (version 9.4; Statistical Solutions, Ltd). Considering an estimated 16% dropout rate based on a meta-analysis of dropout rates in ACT RCTs [33], the study aimed to enroll at least 32 participants in total (16 per group). This sample size is also consistent with the recommendation of a minimum of 10 participants per group for pilot RCTs [34].

Procedures

A study flyer containing the project coordinator's contact information and a link to the eligibility screening survey created in Qualtrics was distributed to directors of Area Agencies on Aging and centers or groups supporting individuals with ADRD and their caregivers (eg, adult day care centers and support groups) across various geographic locations within the United States, including the south (eg, Alabama, Tennessee, Texas), the northeast (eg, New York, New Jersey, Massachusetts), the Midwest (eg, Michigan, Wisconsin, Indiana), and the west (eg, California, Arizona). Potential participants contacted the project coordinator for more information and completed the eligibility screening survey. Printed consent forms and pretest questionnaires were sent by mail to eligible participants who expressed interest in taking part, along with a preaddressed, prestamped envelope for their return.

A co-investigator, who had no direct contact with participants, used a computer-generated randomization sequence to create a list, which was provided to the project coordinator in sequentially numbered, opaque sealed envelopes. After the project coordinator received the signed consent forms and completed the pretest questionnaires, the envelope associated with the participant's ID number was opened and informed them of their group assignment. Participants were asked to complete follow-up questionnaires roughly 3 and 6 months after completing their pretest questionnaires.

The ACT and Control Groups

Participants assigned to the ACT group received individual ACT sessions, led by a trained ACT coach (a licensed professional counselor), via Zoom (Zoom Video Communications) videoconferencing for 1 hour per week over

XSL•FO
RenderX

10 weeks. Additionally, a 1-time 1-hour booster session was provided 1.5 months after the completion of the 10th weekly session to facilitate the maintenance of treatment effects [16]. Participants were provided with additional worksheets and web-based ACT resources for continued use until the booster session. The overview of the 10 weekly ACT sessions is presented in Multimedia Appendix 1. The ACT protocol was developed specifically for family caregivers of individuals with ADRD [18,19].

Printed copies of the ACT session worksheets were sent by mail to participants in the ACT group prior to the first session. The program aimed to actively involve participants through in-session learning activities such as worksheets, metaphors, and case scenarios, along with homework assignments. After each session, the coach shared additional web-based ACT resources, like guided video exercises, and provided a summary of the strategies discussed to support practice between sessions.

Regardless of group assignment, supplemental psychoeducation materials were sent to all participants, including video links and hard copies of resources to better meet these caregivers' needs. These materials aimed to enhance their understanding of dementia (eg, types, stages, and symptoms), care strategies for different stages of dementia, strategies for managing behavioral and psychological symptoms of dementia, and communication techniques. Participants in the control group continued with care as usual and received the same psychoeducation materials provided by the study team. After their participation in the study ended, the control group was given access to web-based ACT resources and offered a free optional 1-hour session with the ACT coach.

Feasibility Outcomes

The feasibility of recruitment, eligibility, and retention was assessed by tracking the number of potential participants who completed the eligibility screening survey, those who were excluded, those who consented, the number of dropouts, and the time required to recruit the target sample size. Participant adherence to the program was evaluated through session attendance in the ACT group. The ACT coach maintained session logs, which were reviewed by the principal investigator (the first author) to ensure the coach followed the protocol consistently across participants.

Data Collection and Entry

Overview

Data were collected using self-reported questionnaires in the pretest (baseline), posttest (approximately 10-12 weeks after the pretest), and a 3-month follow-up (approximately 3 months after the posttest, or 5-6 months after the pretest) for both groups. The pretest questionnaire included demographic and caregiving-related questions, as presented in Multimedia Appendix 2. The project coordinator periodically contacted participants regarding the completion and mailing of their questionnaires. When the project coordinator received the completed paper questionnaire, she checked for any missing responses. If any were found, the project coordinator called the participant to obtain the missing information. Two undergraduate research assistants, who were unaware of group assignments, handled data entry using Qualtrics. Data entered by each assistant was compared by a co-investigator who was not involved in data analysis to identify and resolve any potential data entry errors and to finalize the dataset before analysis.

Primary Outcome Measure

The primary outcome of this study was the PHQ-9, a 9-item self-report tool that assesses the severity of depressive symptoms over the past 2 weeks, rated on a scale from 0 to 3 [31]. Scores range from 0 to 27, with higher scores indicating more severe depressive symptoms. The PHQ-9 categorizes depressive symptom severity as follows: 0-4=no depressive symptoms, 5-9=mild, 10-14=moderate, 15-19=moderately severe, and 20-27=severe [31]. It has demonstrated good internal consistency and test-retest reliability, as well as criterion validity, construct validity, diagnostic validity for major depressive disorder, and sensitivity to change [31]. The minimal clinically important difference (MCID) is estimated to be 5 points on the PHQ-9 scale [35].

Secondary Outcome Measures

The Generalized Anxiety Disorder-7 Scale (GAD-7) is a 7-item self-report tool that assesses the severity of anxiety symptoms over the past 2 weeks, rated on a scale from 0 to 3 [36]. Scores range from 0 to 21, with higher scores indicating more severe anxiety symptoms. Cutoff points of 5, 10, and 15 denote mild, moderate, and severe anxieties, respectively [36]. The GAD-7 has shown good internal consistency and test-retest reliability, as well as criterion validity, convergent validity, construct validity, and sensitivity to change [36-39]. The MCID for the GAD-7 is estimated to be 3 or 4 points [39,40].

The Perceived Stress Scale-10 (PSS-10) is a 10-item self-report questionnaire that assesses how unpredictable, uncontrollable, and overloaded individuals perceived their lives to be over the past month, using a scale of 0 to 4 [41,42]. Scores range from 0 to 40, with higher scores indicating greater stress [42]. The scale has shown good internal consistency and test-retest reliability, as well as convergent validity, construct validity, and concurrent validity [43].

The World Health Organization Quality of Life Instrument–Psychological Health component (WHOQOL-BREF–Psych) has 6 items that assess psychological QoL over the past 2 weeks, rated on a scale of 1 to 5 [44]. Scores range from 6 to 30, with higher scores denoting better psychological QoL. The WHOQOL-BREF–Psych component has demonstrated excellent intrarater reliability, interrater reliability, internal consistency, and adequate convergent validity [44].

The Short Version of the Zarit Burden Interview (ZBI) is a 12-item self-report tool that assesses caregivers' perceived burden, rated on a scale from 0 to 4 [45]. Scores range from 0 to 48, with higher scores indicating higher levels of caregiver burden. Suggested cutoff points are as follows: 0-9=no or little burden, 10-19=mild to moderate burden, and 20-48=high burden [45]. The ZBI has good internal consistency as well as convergent and discriminant validity in caregivers of individuals with ADRD [45,46].

XSL•FO
RenderX

The Marwit-Meuser Caregiver Grief Inventory-Brief Form (MM-CGI-BF) is a 6-item self-report questionnaire that assesses predeath grief in caregivers of individuals with ADRD (ie, losses, including anticipation of future losses), rated on a scale from 1 to 5 [47]. Scores range from 6 to 30, with higher scores indicating greater grief. The MM-CGI-BF has demonstrated good internal consistency, test-retest reliability, and construct validity in caregivers of individuals with ADRD [47].

The Caregiver Guilt Questionnaire (CGQ) is a 22-item self-report measure that assesses feelings of guilt in caregivers, rated on a scale from 0 to 4 [48]. Scores range from 0 to 88, with higher scores indicating greater levels of guilt. The CGQ has good internal consistency and convergent validity in caregivers of individuals with ADRD [48].

Measures Related to ACT Processes

The Acceptance and Action Questionnaire-II is a 7-item self-report measure that assesses experiential avoidance (or efforts to control or suppress unwanted thoughts and emotions, even when doing so is ineffective or has negative consequences) and the broader measure of psychological inflexibility, rated on a scale from 1 to 7 [49]. Scores range from 7 to 49, with higher scores indicating greater experiential avoidance. The Acceptance and Action Questionnaire-II has demonstrated good internal consistency and test-retest reliability, as well as concurrent, predictive, discriminant, and incremental validity [49,50].

The Cognitive Fusion Questionnaire-7 is a 7-item self-report tool that assesses cognitive fusion (ie, being so entangled with thoughts that they dictate behavior in inflexible ways), rated on a scale from 1 to 7 [51]. Scores range from 7 to 49, with higher scores indicating greater cognitive fusion. The Cognitive Fusion Questionnaire-7 has shown good internal consistency and test-retest reliability, as well as concurrent and discriminant validity [51,52].

The Engaged Living Scale-9 is a 9-item self-report questionnaire that assesses values-driven action (ie, clarity and engagement with personal values), rated on a scale from 1 to 5 [53]. Scores range from 9 to 45, with higher scores indicating greater values-driven action. The Engaged Living Scale-9 has demonstrated good internal consistency and construct validity [53,54].

The Self-Compassion Scale-Short Form (SCS-SF) is a 12-item self-report questionnaire that assesses self-compassion, rated on a scale from 1 to 5 [55]. Scores range from 12 to 60, with higher scores indicating higher levels of self-compassion. The SCS-SF has shown good internal consistency and test-retest reliability, as well as convergent validity [56].

Data Analysis

Composite scores, with appropriate reverse coding of the responses for the outcome measures, were computed for analysis. Descriptive statistics were used to summarize participants' characteristics. Normality checks, including boxplots, stem-and-leaf plots, normal probability plots, and the Kolmogorov-Smirnov test, were performed on continuous variables. All continuous variables were found to be at least approximately normally distributed. All statistical tests were 2-sided, with a significance level of α=.05. Statistical analyses were conducted using SAS (version 9.4; SAS Institute, Inc).

To test the equivalence of groups (ACT and control groups) in participants' demographic and caregiving-related characteristics, as well as outcome measures in the pretest, the chi-square test (or Fisher exact test when the assumptions for the chi-square were not satisfied) was used for categorical variables, and the 2-sample *t* test was used for continuous variables. An intent-to-treat approach was used for all outcome analyses (ie, data from all participants were analyzed as randomized). Three participants in the control group and 2 participants in the ACT group who did not complete the posttest and 3-month follow-up evaluations were included in these analyses.

Outcomes were explored using linear mixed-effects models for repeated measures [57-60]. These models included group (1=control group and 2=ACT group), which is treated as a fixed effect, time point (1=pretest, 2=posttest, and 3=3-month follow-up), which is treated as fixed but where the slopes of the time points are treated as random, and the group by time point interaction (including components for the group × posttest interaction and the group by 3-month follow-up interaction), with a random intercept for participants. All 3 time points were included in our linear mixed effects models. These models were estimated using restricted maximum likelihood, and a first-order autoregressive covariance matrix was assumed. All linear mixed effects models were run using PROC MIXED of SAS, invoking the REPEATED statement. This statement allows us to incorporate all aspects of random variation into a covariance structure (for our data, a first-order autoregressive covariance matrix) associated with the individual participants. Overall *P* values and their associated *F* values are first reported for group, time point, and the group×time point interaction. Next, the estimated mean change from pretest to posttest within each group, the absolute difference in mean change between groups at posttest, the estimated mean change from pretest to 3-month follow-up within each group, and the absolute difference in mean change between groups in 3-month follow-up, along with corresponding SEs and 95% CIs, were reported; for a given variable, all of these results were obtained from the same linear mixed effects model.

Linear mixed effects models (including those for repeated measures) are robust against missing data, as they can accommodate unbalanced data patterns [57-60]. Therefore, all available observations were included in the analyses. For each outcome variable, participants with at least 1 nonmissing outcome measure were included in the analysis. A Cohen *d* type effect size [61] was calculated for between-group effects using the absolute difference of the means between the 2 groups divided by the common SD and for within-group effects using the estimated change in the means between the 2 time points divided by the SD of change. Due to the exploratory and preliminary nature of this pilot RCT study, no adjustment of *P* values was conducted for multiple statistical comparisons of the outcome measures [62].

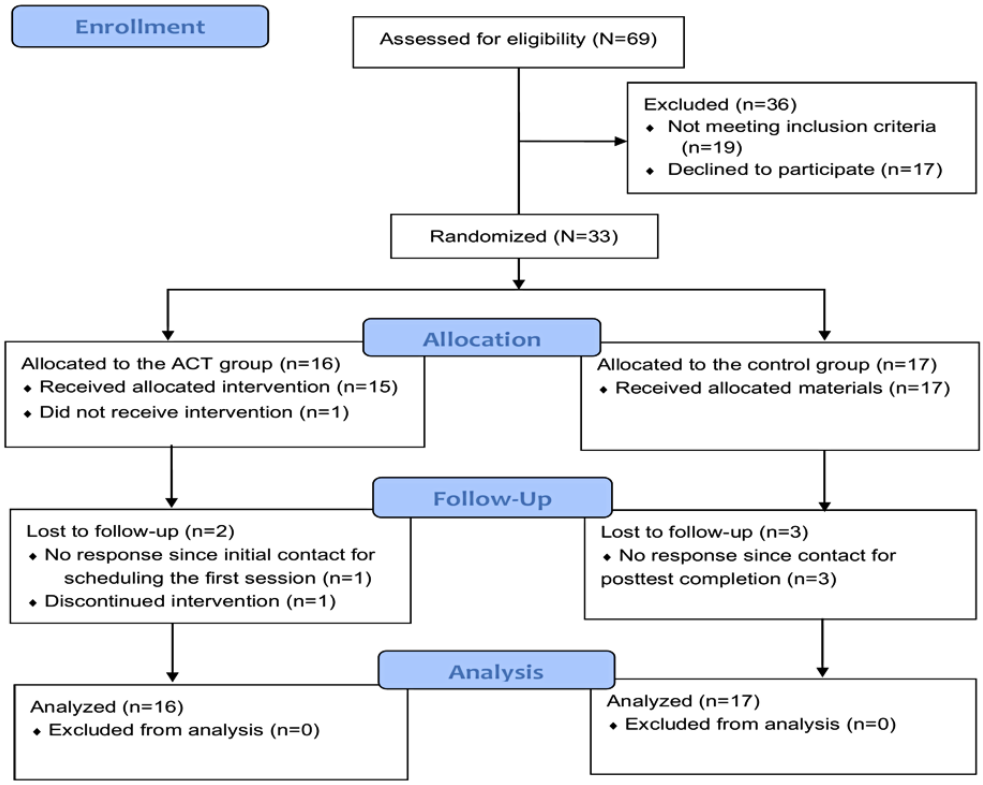
XSL•FO
RenderX

Results

Participant Flow and Feasibility Outcomes

Figure 1 shows the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) flow diagram for participant recruitment (CONSORT-EHEALTH checklist provided in Multimedia Appendix 3). A total of 69 potential participants completed the study eligibility screening survey over the 1.7-year recruitment period. Of these, 19 did not meet the eligibility criteria, mostly due to not meeting the PHQ-9 screening criteria. Seventeen eligible participants chose not to participate in the study, mostly for unknown reasons, and a few due to concerns about the required language used in the consent form.

Figure 1. Participant flow diagram of a pilot randomized controlled trial of a videoconference-delivered ACT group and a control group receiving psychoeducation materials, with pretest, posttest, and 3-month follow-up evaluations for depressed family caregivers of individuals with dementia in the United States. ACT: acceptance and commitment therapy.



The remaining 33 participants gave informed consent, completed pretest questionnaires, and were randomly assigned to either the ACT group or the control group. All participants except 5 (3 in the control group and 2 in the ACT group) completed the posttest and 3-month follow-up questionnaires. Three participants (2 from the control group and 1 from the ACT group) were lost to follow-up after completing the pretest questionnaire and group assignment (no response). One participant in the control group requested withdrawal from the study after the group assignment, and 1 participant in the ACT group requested withdrawal after completing 6 sessions due to a lack of preference for mindfulness practices. All remaining participants in the ACT group completed all 10 weekly ACT sessions and the booster session 1.5 months after the 10th session. The coach adhered to the protocol across participants.

Baseline Characteristics of Participants

Table 1 presents the characteristics of participants (N=33) and the results of between-group differences in characteristics. Multimedia Appendix 4 presents between-group differences in outcomes at pretest. There were no statistically significant differences between the ACT and control groups in characteristics and outcome measures at pretest. The majority of participants were female (n=29, 88%), non-Hispanic White (n=24, 73%), and daughters of individuals with ADRD (n=21, 64%). Participants' ages ranged from 23 to 78 (mean 55.2, SD 12.7) years. The time since the relative's diagnosis of ADRD ranged from 0.4 to 14.2 (mean 4.8, SD 3.8) years.

XSL•FO
RenderX

Table 1. Characteristics of participants at pretest in a pilot randomized controlled trial of a videoconference-delivered ACT[a] group and a control group receiving psychoeducation materials for family caregivers with depression who give care to individuals with dementia in the United States.

	All (N=33)	ACT group (n=16)	Control group (n=17)	Difference between groups, *P* value
Age (years), mean (SD; range)	55.2 (12.7; 23-78)	54.1 (13.4; 23-78)	56.4 (12.2; 31-78)	.543
Sex, n (%)				≥.99
Female	29 (88)	14 (88)	15 (88)	
Male	4 (12)	2 (12)	2 (12)	
Education, n (%)				.61
Postgraduate degree	18 (55)	8 (50)	10 (59)	
Bachelor's degree or less	15 (45)	8 (50)	7 (41)	
Race, n (%)				≥.99
Black or Hispanic	9 (27)	4 (25)	5 (29)	
Non-Hispanic White	24 (73)	12 (75)	12 (71)	
Marital status, n (%)				.83
Married or living with a partner	20 (61)	10 (63)	10 (59)	
Never, divorced, or widowed	13 (39)	6 (37)	7 (41)	
Employment status, n (%)				.12
Retired or unemployed	17 (52)	6 (37)	11 (65)	
Employed	16 (48)	10 (63)	6 (35)	
Relationship to the relative with dementia, n (%)				.44
Non-spousal family members	25 (76)	11 (69)	14 (82)	
Spousal family members	8 (24)	5 (31)	3 (18)	
Living with the relative with dementia, n (%)				≥.99
Living together	27 (82)	13 (81)	14 (82)	
Not living together	6 (18)	3 (19)	3 (18)	
Caregiving hours per week, n (%)				.08
40 hours or more	22 (67)	13 (81)	9 (53)	
Less than 40 hours	11 (33)	3 (19)	8 (47)	
Other family members helping with care, n (%)				.85
No	18 (55)	9 (56)	9 (53)	
Yes	15 (45)	7 (44)	8 (47)	
Experience in support groups, n (%)				.88
Never attended	19 (58)	9 (56)	10 (59)	
Attending or attended before	14 (42)	7 (44)	7 (41)	
Years since the relative's diagnosis of dementia, mean (SD; range)	4.8 (3.8; 0.4-14.2)	3.9 (2.9; 0.6-9.6)	5.6 (4.4; 0.4-14.2)	.19
Type of dementia in the relative, n (%)				.56
Alzheimer disease	13 (40)	5 (31)	8 (47)	
Other dementias	11 (33)	7 (44)	4 (24)	
Not sure	9 (27)	4 (25)	5 (29)	
Stage of dementia in the relative, n (%)				.89
Early or mild	5 (15)	2 (13)	3 (18)	
Middle or moderate	16 (49)	8 (50)	8 (47)	
Late or severe	9 (27)	5 (31)	4 (23)	

XSL•FO
RenderX

	All (N=33)	ACT group (n=16)	Control group (n=17)	Difference between groups, *P* value
Not sure	3 (9)	1 (6)	2 (12)	

[a]ACT: acceptance and commitment therapy.

Effects of the Videoconference-Delivered ACT Program on Outcomes

Multimedia Appendix 5 presents the overall *P* values for group, timepoint, and group×timepoint, along with their associated *F* values. The group×time interaction term was statistically significant for only the variable WHOQOL-BREF–Psych ($F_{2,52}$=4.24; *P*=.020). All variables displayed statistically significant overall timepoint (within-group) effects (*F* values range from 3.32 to 20.73 and *P* values range from <.001 to .044), except for MM-CGI-BF ($F_{2,52}$=2.60; *P*=.084).

Multimedia Appendix 6 presents the outcome measure scores in the pretest, posttest, and 3-month follow-up for each group, change scores in the posttest and 3-month follow-up from the pretest for each group, and findings from within-group and between-group comparisons in a single table. For a given variable in Multimedia Appendices 4 and 5, all inferential statistical results were obtained from the same linear mixed effects model. Within-group and between-group differences in outcomes in posttest and 3-month follow-up are presented in Table 2 and Table 3, respectively.

Two significant between-group differences were observed in the posttest. The ACT group reported significantly greater improvements than the control group in stress and psychological QoL in the posttest. For stress, the estimated within-group change in score for the ACT group (mean –7.75, SE 1.62 points; 95% CI –10.99 to –4.51; *P*<.001) compared to the control group (mean –3.03, SE 1.61 points; 95% CI –6.25 to 0.19; *P*=.065) resulted in an absolute between-group difference of mean –4.72 (SE 2.28) points (95% CI –9.29 to –0.15; *P*=.043). For psychological QoL, the estimated within-group change in score for the ACT group (mean 4.03, SE 0.96 points; 95% CI 2.10-5.96; *P*<.001) compared to the control group (mean 0.56, SE 0.96 points; 95% CI –1.37 to 2.48; *P*=.56) resulted in an absolute between-group difference of mean 3.48 (SE 1.36) points (95% CI 0.75-6.20; *P*=.014). No significant between-group differences were observed for other outcome measures in the posttest, and none were observed for any outcome measure in the 3-month follow-up (Tables 2 and 3).

For within-group comparisons in the ACT group, participants experienced a significant decrease in depressive symptoms, with a mean change of –6.09 (SE 1.16) points (95% CI –8.42 to –3.76; *P*<.001) in the posttest and –6.71 (SE 1.45) points (95% CI –9.63 to –3.81; *P*<.001) in the 3-month follow-up. These changes exceed the estimated MCID on the PHQ-9 scale [35]. In addition, the ACT group reported significant improvements in anxiety, stress, psychological QoL, caregiver burden, predeath grief, guilt, values-driven action, and experiential avoidance in both posttest and 3-month follow-up (Tables 2 and 3). For example, the ACT group experienced a significant decrease in anxiety symptoms, with a mean change of –4.41 (SE 1.14) points (95% CI –6.71 to –2.12; *P*<.001) at posttest and –4.09 (SE 1.47) points (95% CI –7.04 to –1.14; *P*=.008) in the 3-month follow-up. These changes exceed the estimated MCID on the GAD-7 [39,40]. No significant within-group differences in change scores for cognitive fusion and self-compassion in the ACT group were observed (Tables 2 and 3).

For within-group comparisons in the control group that received psychoeducation materials alone, participants experienced a significant decrease in depressive symptoms in the posttest and 3-month follow-up (Tables 2 and 3), but these changes were smaller than the estimated MCID on the PHQ-9 scale [35]. Additionally, the control group reported significant improvements in anxiety in 3-month follow-up, self-compassion at posttest, values-driven action in the posttest and 3-month follow-up, and experiential avoidance in the posttest (Tables 2 and 3).

XSL•FO
RenderX

Table 2. Within-group and between-group differences in outcomes at posttest in a pilot randomized controlled trial of a videoconference-delivered ACT[a] group and a control group receiving psychoeducation materials for family caregivers with depression who give care to individuals with dementia in the United States.

	Estimated within-group Change Δ (posttest – pretest)[b]				Absolute between-group difference in Δ ACT versus Δ control at posttest[b]	
	ACT group		Control group		Mean (SE; 95% CI)	*P* value; ES
	Mean (SE; 95% CI)	*P* value; ES[c]	Mean (SE; 95% CI)	*P* value; ES		
PHQ-9[d] (–)[e]	–6.09 (1.16; –8.42 to –3.76)	<.001; 1.31	–3.55 (1.15); (–5.87 to –1.24)	.003; 0.75	–2.54 (1.64; –5.82 to 0.75)	.13; 0.27
GAD-7[f] (–)[e]	–4.41 (1.14; –6.71 to –2.12)	<.001; 0.96	–2.14 (1.14; –4.42 to 0.15)	.067; 0.46	–2.27 (1.61; –5.51 to 0.97)	.17; 0.25
PSS-10[g] (–)[e]	–7.75 (1.62; –10.99 to –4.51)	<.001; 1.20	–3.03 (1.61; –6.25, 0.19)	.065; 0.46	–4.72 (2.28; –9.29 to –0.15)	.043; 0.36
WHOQOL-BREF–Psych[h]	4.03 (0.96; 2.10-5.96)	<.001; 1.05	0.56 (0.96; –1.37 to 2.48)	.56; 0.14	3.48 (1.36; 0.75-6.20)	.014; 0.45
ZBI-12[i] (–)[e]	–7.11 (1.92; –10.96 to –3.27)	<.001; 0.93	–2.34 (1.91; –6.16 to 1.49)	.23; 0.30	–4.78 (2.70; –10.21 to 0.65)	.083; 0.31
MM-CGI-BF[j] (–)[e]	–2.99 (1.45; –5.91 to –0.08)	.04; 0.51	–1.53 (1.44; –4.43 to 1.36)	.29; 0.26	–1.46 (2.05; –5.57 to 2.65)	.48; 0.12
CGQ[k] (–)[e]	–8.64 (3.45; –15.55 to –1.72)	.015; 0.63	–5.89 (3.43; –12.78 to 0.99)	.092; 0.42	–2.74 (4.86; –12.50 to 7.02)	.58; 0.10
SCS-SF[l]	2.66 (2.01; –1.38 to 6.69)	.19; 0.33	5.79 (2.00; 1.79-9.80)	.005; 0.70	–3.14 (2.83; –8.82 to 2.54)	.27; 0.19
ELS-9[m]	6.24 (1.42; 3.40-9.08)	<.001; 1.10	3.53 (1.41; 0.70-6.36)	.016; 0.61	2.71 (2.00; –1.30 to 6.72)	.18; 0.24
AAQ-II[n] (–)[e]	–5.16 (1.90; –8.97 to –1.34)	.009; 0.68	–4.14 (1.90; –7.95 to –0.34)	.033; 0.53	–1.01 (2.69; –6.40 to 4.38)	.71; 0.07
CFQ-7[o] (–)[e]	–1.29 (2.07; –5.44 to 2.86)	.54; 0.16	–0.28 (2.06; –4.42 to 3.86)	.89; 0.03	–1.01 (2.92; (–6.88 to 4.85)	.73; 0.06

[a]ACT: Acceptance and commitment therapy.

[b]Established from linear mixed effects models.

[c]ES: effect size (Cohen *d*).

[d]PHQ-9: Patient Health Questionnaire-9.

[e]A minus sign in parentheses indicates that a decline in each variable means positive outcomes.

[f]GAD-7: Generalized Anxiety Disorder-7 Scale.

[g]PSS-10: Perceived Stress Scale -10.

[h]WHOQOL-BREF-Psych: World Health Organization Quality of Life Assessment-BREF-Psychological Health Component.

[i]ZBI-12: Zarit Burden Interview-12.

[j]MM-CGI-BF: Marwit–Meuser Caregiver Grief Inventory-Brief-Form.

[k]CGQ: Caregiver Guilt Questionnaire.

[l]SCS-SF: Self-Compassion Scale-Short Form.

[m]ELS-9: Engaged Living Scale-9.

[n]AAQ-II: Acceptance and Action Questionnaire-II.

[o]CFQ-7: Cognitive Fusion Questionnaire-7.

XSL•FO
RenderX

Table 3. Within-group and between-group differences in outcomes in 3-month follow-up in a pilot randomized controlled trial of a videoconference-delivered ACT[a] group and a control group receiving psychoeducation materials for family caregivers with depression who give care to individuals with dementia in the United States.

	Estimated within-group change Δ (3-month follow-up – pretest)[b]				Absolute between-group difference in Δ ACT versus Δ control in 3-month follow-up[b]	
	ACT[a] group		Control group		Mean (SE; 95% CI)	*P* value; ES
	Mean (SE; 95% CI)	*P* value; ES[c]	Mean (SE; 95% CI)	*P* value; ES		
PHQ-9[d] (-)[e]	6.71 (1.45; –9.63 to –3.81)	<.001; 1.00	–4.89 (1.44; –7.78 to –2.00)	.001; 0.82	–1.83 (2.05; –5.93 to 2.28)	.38; 0.16
GAD-7[f] (–)[e]	–4.09 (1.47; –7.04 to –1.14)	.008; 0.69	–3.23 (1.46; –6.16 to –0.30)	.032; 0.54	–0.86 (2.07; –5.03 to 3.30)	.68; 0.07
PSS-10[g] (–)[e]	–7.04 (2.02; –11.09 to –2.99)	.001; 0.87	–1.95 (2.00; –5.96 to 2.07)	.34; 0.24	–5.10 (2.84; –10.80 to 0.60)	.079; 0.31
WHOQOL-BREF–Psych[h]	2.92 (1.19; 0.52-5.31)	.018; 0.61	2.16 (1.18; –0.21 to 4.53)	.073; 0.44	0.76 (1.68; –2.61 to 4.12)	.65; 0.08
ZBI-12[i] (–)[e]	–9.55 (2.47; –14.50 to –4.60)	<.001; 0.97	–3.60 (2.45; –8.51 to 1.31)	.15; 0.36	–5.95 (3.47; –12.92 to 1.02)	.093; 0.30
MM-CGI-BF[j] (–)[e]	–4.11 (1.82; –7.76 to –0.47)	.028; 0.57	–0.51 (1.80; –4.12 to 3.10)	.78; 0.07	–3.60 (2.56; –8.73 to 1.53)	.16; 0.25
CGQ[k] (–)[e]	–11.22 (4.39; –20.03 to –2.42)	.013; 0.64	–7.34 (4.35; –16.07 to 1.39)	.098; 0.41	–3.88 (6.18; –16.28 to 8.51)	.53; 0.11
SCS-SF[l]	4.63 (2.48; –0.35 to 9.61)	.068; 0.47	3.55 (2.46; –1.38 to 8.48)	.15; 0.35	1.08 (3.49; –5.93 to 8.09)	.76; 0.05
ELS-9[m]	5.33 (1.81; 1.69-8.97)	.005; 0.73	4.96 (1.80; 1.34-8.57)	.008; 0.67	0.37 (2.56; –4.76 to 5.50)	.89; 0.03
AAQ-II[n] (–)[e]	–6.48 (2.52; –11.55 to –1.42)	.013; 0.64	–3.32 (2.51; –8.35 to 1.72)	.19; 0.32	–3.17 (3.56; –10.31 to 3.98)	.38; 0.15
CFQ-7[o] (–)[e]	–5.10 (2.74; –10.59 to 0.39)	.068; 0.47	–3.59 (2.72; –9.05 to 1.87)	.19; 0.32	–1.51 (3.86; (–9.25 to 6.24)	.70; 0.07

[a]ACT: acceptance and commitment therapy.

[a]Established from linear mixed effects models.

[c]ES: effect size (Cohen *d*).

[d]PHQ-9: Patient Health Questionnaire-9.

[e]A minus sign in parentheses indicates that a decline in each variable means positive outcomes.

[f]GAD-7: Generalized Anxiety Disorder-7 Scale.

[g]PSS-10: Perceived Stress Scale-10.

[h]WHOQOL-BREF-Psych: World Health Organization Quality of Life Assessment-BREF-Psychological Health Component.

[i]ZBI-12: Zarit Burden Interview-12.

[j]MM-CGI-BF: Marwit–Meuser Caregiver Grief Inventory-Brief-Form.

[k]CGQ: Caregiver Guilt Questionnaire.

[l]SCS-SF: Self-Compassion Scale-Short Form.

[m]ELS-9: Engaged Living Scale-9.

[n]AAQ-II: Acceptance and Action Questionnaire-II.

[o]CFQ-7: Cognitive Fusion Questionnaire-7.

Discussion

Principal Findings

This study, along with our previous work [18,19], represents the first efforts to evaluate videoconference-delivered ACT for family caregivers of individuals with ADRD. In particular, this pilot RCT targeted caregivers with at least mild depressive symptoms, differing from our earlier studies that included those with depressive, anxiety, or stress symptoms [18,19]. This decision was informed by previous findings that showed more positive outcomes when focusing specifically on caregivers with depressive symptoms [19], as well as meta-analyses indicating that web-based ACT interventions have larger effects on depressive symptoms when the population primarily has depressive symptoms [27].

XSL•FO
RenderX

Despite the small sample size, the videoconference-delivered ACT group demonstrated significantly greater improvements in stress and psychological QoL compared to the control group at posttest. The structured nature of ACT, delivered via videoconferencing with a coach, likely provided caregivers with opportunities to engage in real-time emotional processing and skill-building, enabling better responses to day-to-day stressful caregiving demands and situations [13,18]. This likely contributed to the greater improvements in stress and psychological QoL. Cohen *d* effect sizes for these between-group comparisons in stress and psychological QoL were 0.36 and 0.45, respectively, indicating small effect sizes. These findings align with meta-analyses on the effects of internet-delivered ACT compared to control conditions in RCTs across diverse populations [27], as well as the effects of ACT on family caregivers of individuals with chronic conditions [63]. Importantly, this is the first study to show significantly greater improvements in stress and psychological QoL for family caregivers of individuals with ADRD in an ACT program compared to a control group, providing valuable new evidence for this population.

However, no significant between-group differences were found for the other outcome measures in the posttest, nor for any outcome measure at the 3-month follow-up. This may be because this small-scale RCT was underpowered to detect between-group differences in most outcomes. Also, the control group received psychoeducation materials, which may have provided some benefits, particularly since more than half of the participants had not previously attended support groups. In addition, 1 participant in the control group, who found the study information through the clinical trial registration, had near-outlier data, which may have influenced some of the results. This participant's data approached outlier status (ie, outside of 3 SDs from the mean). A sensitivity analysis was conducted excluding this participant's data. The sensitivity analysis revealed statistically significant between-group differences in depressive symptoms in the posttest (absolute between-group difference=–3.25 points; *P*=.037); stress at posttest (absolute between-group difference=–6.45 points; *P*<.001) and in 3-month follow-up (absolute between-group difference=–7.65 points; *P*=.001); psychological QoL in the posttest (absolute between-group difference=4.35 points; *P*<.001); caregiver burden in the posttest (absolute between-group difference=–6.7 points; *P*=.003) and in the 3-month follow-up (absolute between-group difference=–8.83 points; *P*=.003); predeath grief in 3-month follow-up (absolute between-group difference=–5.31 points; *P*=.031); and values-driven action at posttest (absolute between-group difference=3.67 points; *P*=.032). The effect sizes for the differences between groups were larger in the posttest for all outcomes except for the SCS-SF and in the 3-month follow-up for all outcomes when excluding this participant's data than when including it. Notably, the effect sizes shifted from small to medium in stress and caregiver burden in both the posttest and 3-month follow-up, as well as in psychological QoL in the posttest. A table detailing the results from this sensitivity analysis can be found in Multimedia Appendix 7.

For within-group comparisons, the videoconference-delivered ACT group experienced significant improvements in depressive symptoms, anxiety, stress, psychological QoL, caregiver burden, predeath grief, guilt, values-driven action, and experiential avoidance in both posttest and 3-month follow-up. The mean changes in the 3-month follow-up were similar to those in the posttest, indicating that the treatment effects were maintained over time. Additionally, Cohen *d* effect sizes for these within-group comparisons indicated moderate to large effect sizes in both posttest and 3-month follow-up for these outcomes. These findings align with within-group improvements observed in prior ACT studies delivered through various modes for this population [14-23]. While the control group, which received psychoeducational materials alone, showed some within-group improvements, these changes were smaller compared to the ACT group and were not statistically significant in most cases.

Limitations

This study had several limitations. The small sample size may have limited the statistical power to detect small to moderate significant between-group differences in outcomes, especially at the 3-month follow-up. We adjusted our target sample size due to challenges in recruiting family caregivers with depression who give care to individuals with ADRD for an RCT within our funding period. Recruitment relied on sharing our study flyer with community sites, but we lacked access to the population through clinics, and some potential participants had already been recruited for previous studies, which further constrained our recruitment capacity. Another limitation of this pilot RCT is that, due to its exploratory and preliminary nature, no adjustment of *P* values was conducted for multiple statistical comparisons of the outcome measures. As a result, the findings should be interpreted with caution, and further studies with larger sample sizes and appropriate adjustments for multiple comparisons are needed to confirm the reliability and robustness of the results. This study did not collect data on participants' income levels or the types of private caregiving support services they were receiving at the time of recruitment. Therefore, these factors were not considered in the data analysis when comparing group differences or controlling for these factors if between-group differences were identified. Given that financial resources and access to caregiving support services may influence caregivers' mental health and engagement in interventions [64], future research should include these variables to better understand how socioeconomic status and caregiving resources impact mental health outcomes and intervention effects among family caregivers.

Conclusions

The videoconference-delivered, therapist-guided ACT program demonstrated feasibility and positive effects on mental health outcomes among family caregivers with depression who give care to individuals with ADRD. Given the barriers many caregivers face in accessing in-person mental health services, this mode of delivery represents a promising alternative for improving caregivers' mental health. Future studies are needed to replicate these findings with larger, more diverse caregiver populations and explore the long-term efficacy of videoconference-delivered ACT programs.

XSL•FO
RenderX

Acknowledgments

The authors sincerely appreciate the study participants for generously giving their time and contributing to this research. The authors appreciate the directors and staff members of the Area Agencies on Aging and adult day service programs for individuals with Alzheimer disease and related dementias for their assistance in sharing the study flyer with potential participants, and Bella Fancher and Mattie White for their help in entering data. In addition, the authors thank Donna Burnett from the Department of Occupational Therapy and Faye Suggs from the Office of Sponsored Programs at the University of Alabama at Birmingham for their tremendous administrative support in managing the research grant. This work was supported by an Alzheimer's Association Research Grant (AARG-21-846868). The sponsor had no role in the design, execution, analysis, or interpretation of the research findings.

At the time of the study, Areum Han was affiliated with the University of Alabama at Birmingham. She is now affiliated with the College of Nursing, University of South Florida, Tampa, FL, USA.

Data Availability

The findings from the data analysis are presented in the main document as Tables and Multimedia Appendices. Deidentified participant data are available from the corresponding author upon reasonable request.

Authors' Contributions

AH, RO, and HY conceptualized the study, secured the funding, and served as investigators on the project. AH supervised the overall project, RO conducted the data analysis, and HY oversaw data entry by 2 undergraduate students. JJ served as a coach delivering sessions and JH and LE served as project coordinators. AH and RO wrote the first draft of the manuscript, and all authors contributed to the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of sessions provided to the videoconference-delivered acceptance and commitment therapy group for depressed family caregivers of individuals with dementia in the United States.
[DOCX File , 13 KB - formative_v9i1e67545_app1.docx]

Multimedia Appendix 2
Participant demographic and caregiving-related questions administered to all participants at pretest in a pilot randomized controlled trial for depressed family caregivers of individuals with dementia in the United States.
[DOCX File , 13 KB - formative_v9i1e67545_app2.docx]

Multimedia Appendix 3
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist.
[PDF File (Adobe PDF File), 1180 KB - formative_v9i1e67545_app3.pdf]

Multimedia Appendix 4
Between-group differences in outcomes at pretest in a pilot randomized controlled trial of a videoconference-delivered acceptance and commitment therapy group versus a control group receiving psychoeducation materials for depressed family caregivers of individuals with dementia in the United States.
[DOCX File , 14 KB - formative_v9i1e67545_app4.docx]

Multimedia Appendix 5
Results of overall between-group and within-group comparisons in a pilot randomized controlled trial of a videoconference-delivered acceptance and commitment therapy group versus a control group receiving psychoeducation materials for depressed family caregivers of individuals with dementia in the United States.
[DOCX File , 15 KB - formative_v9i1e67545_app5.docx]

Multimedia Appendix 6

XSL•FO
RenderX

Scores in outcome measures with between-group and within-group comparisons in a pilot randomized controlled trial of a videoconference-delivered acceptance and commitment therapy group versus a control group receiving psychoeducation materials for depressed family caregivers of individuals with dementia in the United States.
[DOCX File , 18 KB - formative_v9i1e67545_app6.docx]

Multimedia Appendix 7
Results of a sensitivity analysis excluding one participant with near-outlier data in a pilot randomized controlled trial of a videoconference-delivered acceptance and commitment therapy group versus a control group receiving psychoeducation materials for depressed family caregivers of individuals with dementia in the United States.
[DOCX File , 16 KB - formative_v9i1e67545_app7.docx]

References

1. Alzheimer's Association. 2023 Alzheimer's disease facts and figures. Alzheimers Dement 2023;19(4):1598-1695. [doi: 10.1002/alz.13016] [Medline: 36918389]
2. Coehlo DP, Hooker K, Bowman S. Institutional placement of persons with dementia: What predicts occurrence and timing? J Fam Nurs 2007;13(2):253-277. [doi: 10.1177/1074840707300947] [Medline: 17452605]
3. Guterman EL, Allen IE, Josephson SA, Merrilees JJ, Dulaney S, Chiong W, et al. Association between caregiver depression and emergency department use among patients with dementia. JAMA Neurol 2019;76(10):1166-1173 [FREE Full text] [doi: 10.1001/jamaneurol.2019.1820] [Medline: 31282955]
4. Zwingmann I, Michalowsky B, Esser A, Kaczynski A, Monsees J, Keller A, et al. Identifying unmet needs of family dementia caregivers: results of the baseline assessment of a cluster-randomized controlled intervention trial. J Alzheimers Dis 2019;67(2):527-539 [FREE Full text] [doi: 10.3233/JAD-180244] [Medline: 30584136]
5. Collins RN, Kishita N. Prevalence of depression and burden among informal care-givers of people with dementia: a meta-analysis. Ageing Soc 2019;40(11):2355-2392. [doi: 10.1017/s0144686x19000527]
6. Kaddour L, Kishita N. Anxiety in informal dementia carers: a meta-analysis of prevalence. J Geriatr Psychiatry Neurol 2020;33(3):161-172. [doi: 10.1177/0891988719868313] [Medline: 31409196]
7. Gonzalo-Ciria L, Gascón-Catalán A, Laborda-Soriano A, Cambra-Aliaga A, Ruiz-Garrós MC, Perez-de-Heredia-Torres M. Difficulties fulfilling self-care needs among family caregivers: an observational study. Am J Occup Ther 2024;78(3):7803205020. [doi: 10.5014/ajot.2024.050528] [Medline: 38640086]
8. Karg N, Graessel E, Randzio O, Pendergrass A. Dementia as a predictor of care-related quality of life in informal caregivers: a cross-sectional study to investigate differences in health-related outcomes between dementia and non-dementia caregivers. BMC Geriatr 2018;18(1):189 [FREE Full text] [doi: 10.1186/s12877-018-0885-1] [Medline: 30139354]
9. Lin X, Moxley JH, Czaja SJ. Caring for dementia caregivers: psychosocial factors related to engagement in self-care activities. Behav Sci 2023;13(10):851 [FREE Full text] [doi: 10.3390/bs13100851] [Medline: 37887501]
10. Cheng ST, Li KK, Losada A, Zhang F, Au A, Thompson LW, et al. The effectiveness of nonpharmacological interventions for informal dementia caregivers: an updated systematic review and meta-analysis. Psychol Aging 2020;35(1):55-77. [doi: 10.1037/pag0000401] [Medline: 31985249]
11. Dindo L, Van Liew JR, Arch JJ. Acceptance and commitment therapy: a transdiagnostic behavioral intervention for mental health and medical conditions. Neurotherapeutics 2017;14(3):546-553 [FREE Full text] [doi: 10.1007/s13311-017-0521-3] [Medline: 28271287]
12. Hayes SC, Pistorello J, Levin ME. Acceptance and commitment therapy as a unified model of behavior change. Couns Psychol 2012;40(7):976-1002. [doi: 10.1177/0011000012460836]
13. Han A, Yuen HK, Jenkins J, Edwards L. A qualitative study exploring experiences of female family caregivers of people with dementia in acceptance and commitment therapy. Clin Psychol 2023;28(1):26-36. [doi: 10.1080/13284207.2023.2285803]
14. Romero-Moreno R, Márquez-González M, Gallego-Alberto L, Cabrera I, Vara-García C, Pedroso-Chaparro MDS, et al. Guilt focused intervention for family caregivers. Preliminary results of a randomized clinical trial. Clin Gerontol 2022;45(5):1304-1316. [doi: 10.1080/07317115.2022.2048287] [Medline: 35286236]
15. Losada A, Márquez-González M, Romero-Moreno R, Mausbach BT, López J, Fernández-Fernández V, et al. Cognitive-behavioral therapy (CBT) versus acceptance and commitment therapy (ACT) for dementia family caregivers with significant depressive symptoms: results of a randomized clinical trial. J Consult Clin Psychol 2015;83(4):760-772. [doi: 10.1037/ccp0000028] [Medline: 26075381]
16. Márquez-González M, Romero-Moreno R, Cabrera I, Olmos R, Pérez-Miguel A, Losada A. Tailored versus manualized interventions for dementia caregivers: the functional analysis-guided modular intervention. Psychol Aging 2020;35(1):41-54. [doi: 10.1037/pag0000412] [Medline: 31985248]
17. Gallego-Alberto L, Márquez-González M, Romero-Moreno R, Cabrera I, Losada A. Pilot study of a psychotherapeutic intervention for reducing guilt feelings in highly distressed dementia family caregivers (innovative practice). Dementia 2021;20(2):759-769. [doi: 10.1177/1471301219886761] [Medline: 31707841]

XSL•FO
RenderX

18. Han A, Yuen HK, Jenkins J, Yun Lee H. Acceptance and commitment therapy (ACT) guided online for distressed caregivers of persons living with dementia. Clin Gerontol 2022;45(4):927-938. [doi: 10.1080/07317115.2021.1908475] [Medline: 33794127]
19. Han A, Yuen HK, Jenkins J. The feasibility and preliminary effects of a pilot randomized controlled trial: videoconferencing acceptance and commitment therapy in distressed family caregivers of people with dementia. J Health Psychol 2023;28(6):554-567 [FREE Full text] [doi: 10.1177/13591053221141131] [Medline: 36591636]
20. Fauth EB, Novak JR, Levin ME. Outcomes from a pilot online acceptance and commitment therapy program for dementia family caregivers. Aging Ment Health 2022;26(8):1620-1629. [doi: 10.1080/13607863.2021.1942432] [Medline: 34233133]
21. Fowler NR, Judge KS, Lucas K, Gowan T, Stutz P, Shan M, et al. Feasibility and acceptability of an acceptance and commitment therapy intervention for caregivers of adults with Alzheimer's disease and related dementias. BMC Geriatr 2021;21(1):127 [FREE Full text] [doi: 10.1186/s12877-021-02078-0] [Medline: 33593296]
22. Kishita N, Gould RL, Farquhar M, Contreras M, Van Hout E, Losada A, et al. Internet-delivered guided self-help acceptance and commitment therapy for family carers of people with dementia (iACT4CARERS): a feasibility study. Aging Ment Health 2022;26(10):1933-1941 [FREE Full text] [doi: 10.1080/13607863.2021.1985966] [Medline: 34617489]
23. Risch AK, Lechner-Meichsner F, Wilz G. Telephone-based acceptance and commitment therapy for caregivers of persons with dementia: results of a randomized controlled trial. Clin Gerontol 2024:1-19 [FREE Full text] [doi: 10.1080/07317115.2024.2393307] [Medline: 39169671]
24. Atefi GL, van Knippenberg RJM, Bartels SL, Losada-Baltar A, Márquez-González M, Verhey FRJ, et al. A web-based intervention based on acceptance and commitment therapy for family caregivers of people with dementia: mixed methods feasibility study. JMIR Aging 2024;7:e53489 [FREE Full text] [doi: 10.2196/53489] [Medline: 38574360]
25. Knight BG, Sayegh P. Cultural values and caregiving: the updated sociocultural stress and coping model. J Gerontol B Psychol Sci Soc Sci 2010;65B(1):5-13. [doi: 10.1093/geronb/gbp096] [Medline: 19934166]
26. Thompson EM, Destree L, Albertella L, Fontenelle LF. Internet-based acceptance and commitment therapy: a transdiagnostic systematic review and meta-analysis for mental health outcomes. Behav Ther 2021;52(2):492-507. [doi: 10.1016/j.beth.2020.07.002] [Medline: 33622516]
27. Han A, Kim TH. Efficacy of internet-based acceptance and commitment therapy for depressive symptoms, anxiety, stress, psychological distress, and quality of life: systematic review and meta-analysis. J Med Internet Res 2022;24(12):e39727 [FREE Full text] [doi: 10.2196/39727] [Medline: 36485030]
28. Han A, Kim TH. The effects of internet-based acceptance and commitment therapy on process measures: systematic review and meta-analysis. J Med Internet Res 2022;24(8):e39182 [FREE Full text] [doi: 10.2196/39182] [Medline: 36040783]
29. Common mental health problems: identification and pathways to care. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/cg123 [accessed 2011-05-25]
30. Durepos P, MacLean R, Ricketts N, Boamah SA, Witherspoon R, Gould O, et al. Engaging care partners of persons living with dementia in acceptance and commitment therapy (ACT) programs: a scoping review. Aging Ment Health 2024;28(5):725-737. [doi: 10.1080/13607863.2023.2288864] [Medline: 38100551]
31. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
32. Molander P, Hesser H, Weineland S, Bergwall K, Buck S, Jäder Malmlöf J, et al. Internet-based acceptance and commitment therapy for psychological distress experienced by people with hearing problems: a pilot randomized controlled trial. Cogn Behav Ther 2018;47(2):169-184. [doi: 10.1080/16506073.2017.1365929] [Medline: 28895787]
33. Ong CW, Lee EB, Twohig MP. A meta-analysis of dropout rates in acceptance and commitment therapy. Behav Res Ther 2018;104:14-33. [doi: 10.1016/j.brat.2018.02.004] [Medline: 29477890]
34. Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. Stat Methods Med Res 2016;25(3):1057-1073 [FREE Full text] [doi: 10.1177/0962280215588241] [Medline: 26092476]
35. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the Patient Health Questionnaire-9. Med Care 2004;42(12):1194-1201. [doi: 10.1097/00005650-200412000-00006] [Medline: 15550799]
36. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
37. Kertz S, Bigda-Peyton J, Bjorgvinsson T. Validity of the Generalized Anxiety Disorder-7 scale in an acute psychiatric sample. Clin Psychol Psychother 2013;20(5):456-464. [doi: 10.1002/cpp.1802] [Medline: 22593009]
38. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
39. Toussaint A, Hüsing P, Gumz A, Wingenfeld K, Härter M, Schramm E, et al. Sensitivity to change and minimal clinically important difference of the 7-item Generalized Anxiety Disorder Questionnaire (GAD-7). J Affect Disord 2020;265:395-401. [doi: 10.1016/j.jad.2020.01.032] [Medline: 32090765]
40. Kroenke K, Baye F, Lourens SG. Comparative responsiveness and minimally important difference of common anxiety measures. Med Care 2019;57(11):890-897. [doi: 10.1097/MLR.0000000000001185] [Medline: 31415337]

XSL•FO
RenderX

41. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983;24(4):385. [doi: 10.2307/2136404]
42. Cohen S. Perceived stress in a probability sample of the United States. In: Spacapan S, Oskamp S, editors. The Social Psychology of Health: The Claremont Symposium on Applied Social Psychology. Thousand Oaks, CA: Sage Publications; 1988:31-67.
43. Ezzati A, Jiang J, Katz MJ, Sliwinski MJ, Zimmerman ME, Lipton RB. Validation of the perceived stress scale in a community sample of older adults. Int J Geriatr Psychiatry 2014;29(6):645-652 [FREE Full text] [doi: 10.1002/gps.4049] [Medline: 24302253]
44. Skevington S, Lotfy M, O'Connell K, WHOQOL Group. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. Qual Life Res 2004;13(2):299-310. [doi: 10.1023/B:QURE.0000018486.91360.00] [Medline: 15085902]
45. Bédard M, Molloy DW, Squire L, Dubois S, Lever J, O'Donnell M. The Zarit Burden Interview: a new short version and screening version. Gerontologist 2001;41(5):652-657. [doi: 10.1093/geront/41.5.652] [Medline: 11574710]
46. Pinyopornpanish K, Pinyopornpanish M, Wongpakaran N, Wongpakaran T, Soontornpun A, Kuntawong P. Investigating psychometric properties of the thai version of the Zarit Burden Interview using Rasch model and confirmatory factor analysis. BMC Res Notes 2020;13(1):120 [FREE Full text] [doi: 10.1186/s13104-020-04967-w] [Medline: 32122404]
47. Liew T, Yap P. A brief, 6-item scale for caregiver grief in dementia caregiving. Gerontologist 2020;60(1):e1-e10. [doi: 10.1093/geront/gny161] [Medline: 30544229]
48. Losada A, Márquez-González M, Peñacoba C, Romero-Moreno R. Development and validation of the caregiver guilt questionnaire. Int Psychogeriatr 2010;22(4):650-660 [FREE Full text] [doi: 10.1017/S1041610210000074] [Medline: 20170587]
49. Bond FW, Hayes SC, Baer RA, Carpenter KM, Guenole N, Orcutt HK, et al. Preliminary psychometric properties of the Acceptance and Action Questionnaire-II: a revised measure of psychological inflexibility and experiential avoidance. Behav Ther 2011;42(4):676-688. [doi: 10.1016/j.beth.2011.03.007] [Medline: 22035996]
50. Fledderus M, Oude Voshaar MAH, Ten Klooster PM, Bohlmeijer ET. Further evaluation of the psychometric properties of the Acceptance and Action Questionnaire-II. Psychol Assess 2012;24(4):925-936. [doi: 10.1037/a0028200] [Medline: 22545700]
51. Gillanders DT, Bolderston H, Bond FW, Dempster M, Flaxman PE, Campbell L, et al. The development and initial validation of the Cognitive Fusion Questionnaire. Behav Ther 2014;45(1):83-101 [FREE Full text] [doi: 10.1016/j.beth.2013.09.001] [Medline: 24411117]
52. Donati MA, Berrocal C, Bernini O, Gori C, Primi C. Measuring cognitive fusion through the Cognitive Fusion Questionnaire-7: measurement invariance across non-clinical and clinical psychological samples. PLoS One 2021;16(2):e0246434 [FREE Full text] [doi: 10.1371/journal.pone.0246434] [Medline: 33534868]
53. Trindade IA, Ferreira C, Pinto-Gouveia J, Nooren L. Clarity of personal values and committed action: development of a shorter Engaged Living Scale. J Psychopathol Behav Assess 2015;38(2):258-265. [doi: 10.1007/s10862-015-9509-7]
54. Reilly ED, Ritzert TR, Scoglio AA, Mote J, Fukuda SD, Ahern ME, et al. A systematic review of values measures in acceptance and commitment therapy research. J Contextual Behav Sci 2019;12:290-304. [doi: 10.1016/j.jcbs.2018.10.004]
55. Raes F, Pommier E, Neff KD, Van Gucht D. Construction and factorial validation of a short form of the Self-Compassion Scale. Clin Psychol Psychother 2011;18(3):250-255. [doi: 10.1002/cpp.702] [Medline: 21584907]
56. Hayes JA, Lockard AJ, Janis RA, Locke BD. Construct validity of the Self-Compassion Scale-Short Form among psychotherapy clients. Couns Psychol Q 2016;29(4):405-422. [doi: 10.1080/09515070.2016.1138397]
57. Detry MA, Ma Y. Analyzing repeated measurements using mixed models. JAMA 2016;315(4):407-408. [doi: 10.1001/jama.2015.19394] [Medline: 26813213]
58. Diggle P, Heagerty P, Liang K, Zeger S. Analysis of Longitudinal Data. Oxford, UK: Oxford University Press; 2013.
59. Hayat MJ, Hedlin H. Modern statistical modeling approaches for analyzing repeated-measures data. Nurs Res 2012;61(3):188-194. [doi: 10.1097/NNR.0b013e31824f5f58] [Medline: 22551993]
60. Muhammad LN. Guidelines for repeated measures statistical analysis approaches with basic science research considerations. J Clin Invest 2023;133(11):e171058 [FREE Full text] [doi: 10.1172/JCI171058] [Medline: 37259921]
61. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates; 1989.
62. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology 1990;1(1):43-46. [Medline: 2081237]
63. Han A, Yuen HK, Jenkins J. Acceptance and commitment therapy for family caregivers: a systematic review and meta-analysis. J Health Psychol 2021;26(1):82-102. [doi: 10.1177/1359105320941217] [Medline: 32659142]
64. Scheuermann J, Pendergrass A, Diehl K, Herr RM. The roles of employment status and income in the mental health of informal caregivers in Germany. BMC Public Health 2024;24(1):2802 [FREE Full text] [doi: 10.1186/s12889-024-20252-y] [Medline: 39396955]

Abbreviations

ACT: acceptance and commitment therapy

XSL•FO
RenderX

ADRD: Alzheimer's disease and related dementias
CGQ: Caregiver Guilt Questionnaire
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
GAD-7: Generalized Anxiety Disorder-7 Scale
MCID: minimal clinically important difference
MM-CGI-BF: Marwit–Meuser Caregiver Grief Inventory-Brief-Form
PHQ-9: Patient Health Questionnaire-9
PSS-10: Perceived Stress Scale-10
QoL: quality of life
RCT: randomized controlled trial
SCS-SF: Self-Compassion Scale-Short Form
WHOQOL-BREF-Psych: World Health Organization Quality of Life Assessment-BREF-Psychological Health Component
ZBI-12: Zarit Burden Interview-12

Edited by A Mavragani; submitted 17.10.24; peer-reviewed by ST Cheng, N Kiselev; comments to author 06.01.25; revised version received 23.01.25; accepted 05.03.25; published 31.03.25.

Please cite as:
Han A, Oster R, Yuen H, Jenkins J, Hawkins J, Edwards L
Videoconference-Delivered Acceptance and Commitment Therapy for Family Caregivers of People With Dementia: Pilot Randomized Controlled Trial
JMIR Form Res 2025;9:e67545
URL: https://formative.jmir.org/2025/1/e67545
doi:10.2196/67545
PMID:

©Areum Han, Robert Oster, Hon Yuen, Jeremy Jenkins, Jessica Hawkins, Lauren Edwards. Originally published in JMIR Formative Research (https://formative.jmir.org), 31.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Personalized Physician-Assisted Sleep Advice for Shift Workers: Algorithm Development and Validation Study

Yufei Shen[1,2]; Alicia Choto Olivier[1]; Han Yu[1]; Asami Ito-Masui[3]; Ryota Sakamoto[3]; Motomu Shimaoka[3]; Akane Sano[1]

[1]Rice University, Houston, TX, United States

[2]The University of Texas at Austin, Austin, TX, United States

[3]Mie University, Tsu, Japan

Corresponding Author:
Akane Sano
Rice University
6100 Main St.
Houston, TX, 77005
United States
Phone: 1 7133483821
Email: akane.sano@rice.edu

Abstract

Background: In the modern economy, shift work is prevalent in numerous occupations. However, it often disrupts workers' circadian rhythms and can result in shift work sleep disorder. Proper management of shift work sleep disorder involves comprehensive and patient-specific strategies, some of which are similar to cognitive behavioral therapy for insomnia.

Objective: Our goal was to develop and evaluate machine learning algorithms that predict physicians' sleep advice using wearable and survey data. We developed a web- and app-based system to provide individualized sleep and behavior advice based on cognitive behavioral therapy for insomnia for shift workers.

Methods: Data were collected for 5 weeks from shift workers (N=61) in the intensive care unit at 2 hospitals in Japan. The data comprised 3 modalities: Fitbit data, survey data, and sleep advice. After the first week of enrollment, physicians reviewed Fitbit and survey data to provide sleep advice and selected 1 to 5 messages from a list of 23 options. We handcrafted physiological and behavioral features from the raw data and identified clusters of participants with similar characteristics using hierarchical clustering. We explored 3 models (random forest, light gradient-boosting machine, and CatBoost) and 3 data-balancing approaches (no balancing, random oversampling, and synthetic minority oversampling technique) to predict selections for the 7 most frequent advice messages related to bedroom brightness, smartphone use, and nap and sleep duration. We tested our predictions under participant-dependent and participant-independent settings and analyzed the most important features for prediction using permutation importance and Shapley additive explanations.

Results: We found that the clusters were distinguished by work shifts and behavioral patterns. For example, one cluster had days with low sleep duration and the lowest sleep quality when there was a day shift on the day before and a midnight shift on the current day. Our advice prediction models achieved a higher area under the precision-recall curve than the baseline in all settings. The performance differences were statistically significant (P<.001 for 13 tests and P=.003 for 1 test). Sensitivity ranged from 0.50 to 1.00, and specificity varied between 0.44 and 0.93 across all advice messages and dataset split settings. Feature importance analysis of our models found several important features that matched the corresponding advice messages sent. For instance, for message 7 (*darken the bedroom when you go to bed*), the models primarily examined the average brightness of the sleep environment to make predictions.

Conclusions: Although our current system requires physician input, an accurate machine learning algorithm shows promise for automatic advice without compromising the trustworthiness of the selected recommendations. Despite its decent performance, the algorithm is currently limited to the 7 most popular messages. Further studies are needed to enable predictions for less frequent advice labels.

(JMIR Form Res 2025;9:e65000) doi:10.2196/65000

XSL•FO
RenderX

KEYWORDS

cognitive behavioral therapy; CBT; health care workers; machine learning; medical safety; web-based intervention; app-based intervention; shift work; shift work sleep disorders; shift workers; sleep disorder; wearable sensors; well-being

Introduction

Background

In the modern economy, shift work is prevalent in numerous occupations. Data from 2010 estimated that half of the workers in food preparation and serving, >45% of workers in protective service, and >35% of health care practitioners had alternative shifts apart from regular day shifts in the United States [1]. However, shift work often conflicts with the daily rhythm of sleep and wakefulness, which increases the risk of shift work sleep disorder (SWSD), a circadian rhythm disorder characterized by shift work–related sleepiness and insomnia [2]. Multiple studies have shown the substantial prevalence of SWSD among shift workers. A survey of nurses in 3 federal hospitals in Ethiopia found that 25.6% (102/399) of participants had SWSD [3]. Another study of nurses in Norway reported prevalence rates of 44.2% (580/1313) and 23.6% (146/619) for SWSD-indicative symptoms among night workers and day workers, respectively [4]. A random population sample of 1163 participants in Australia revealed that 10.1% (91/898) of day workers and 32.1% (85/265) of night workers fulfilled the study's criteria for SWSD and 1.3% (12/898) of day workers and 9.1% (24/265) of night workers were further classified under severe SWSD based on the extent of negative SWSD impacts on their life [5]. Shift workers' sleep problems are also associated with risks of mental health issues [6-8], work errors [8], burnout [6], and turnover intentions [9].

Proper management of SWSD emphasizes comprehensive and patient-specific strategies such as circadian adaptation, maintaining good sleep hygiene, strategic napping, clockwise shift rotation, and wakefulness promotions for night shifts [10]. Some of these strategies are analogous to the cognitive behavioral therapy for insomnia (CBTI), a psychological therapy recommended by the American College of Physicians as initial treatment for chronic insomnia [11], which consists of multiple approaches to improve sleep, including education about normal sleep, sleep hygiene practice, sleep restriction to consolidate high-quality sleep, and relaxation [12].

Related Work

Researchers have experimented with CBTI on shift workers with sleep problems and high risk of SWSD. A study by Järnefelt et al [13] on 19 shift workers with insomnia, more than half of whom exhibited SWSD characteristics, showed that group-based CBTI treatment improved their insomnia severity and sleep over a 6-month postintervention period, and a follow-up study further demonstrated long-lasting benefits of CBTI for up to 24 months after the intervention [14]. Järnefelt et al [15] also conducted a randomized controlled trial of group-based CBTI and self-help CBTI against a sleep hygiene education control on 83 shift workers with insomnia more recently. They found that group-based CBTI improved participants' mental health significantly, although both CBTI interventions enhanced their sleep and reduced their insomnia symptoms without significant differences from the control group, and all interventions were more beneficial for participants without SWSD characteristics [15]. Peter et al [16] used web-based CBTI for shift workers with sleep problems (n=21) and found it to be as effective as face-to-face SWSD outpatient CBTI treatment (n=12) for improving sleep efficiency and insomnia symptoms. Their research group also proposed a randomized controlled trial in 2021 to further assess their web-based CBTI intervention for shift workers against a waitlist control and face-to-face CBTI [17]. Although this existing research on CBTI for shift workers has shown promising improvement in participants' sleep, group-based and face-to-face CBTI required participants to attend several weekly sessions for up to 120 minutes led by specialists trained for >10 hours, and the long time commitment likely caused many participants to miss some sessions, which might have weakened the intervention effects [13,15,16]. Self-help, web-based CBTI was relatively more accessible to shift workers, but time commitment for weekly sessions was still necessary and likely contributed to the considerable dropout rates, and the interventions either required human input for professional feedback or only resembled a web-based learning module without any personalization [15,16]. Therefore, a system that automatically provides personalized CBTI interventions and sleep recommendations for shift workers and requires less time commitment could benefit them substantially and yield low dropout rates.

In recent years, several web-based and mobile CBTI systems have been developed to deliver automated and individualized CBTI interventions. For example, Sleep Healthy Using the Internet [18] and Sleepio [19] provide automated web-based CBTI modules equipped with interactive web design elements and user-tailored recommendations, and multiple randomized controlled trials have demonstrated their effectiveness against insomnia and related mental health problems among various demographic groups [18,20-23]. Beun et al [24,25] developed Sleepcare, a mobile app equipped with an automated digital coach interacting with users to provide CBTI interventions, adapt interventions to user characteristics, and promote adherence. Its efficacy has been demonstrated in a randomized controlled trial of 151 participants by Horsch et al [26]. CBT-i Coach is another phone app developed as a supplemental tool for face-to-face CBTI in which users can receive sleep education, customize sleep hygiene advice and relaxation activities, and practice personalized sleep restrictions [27]. In a pilot study, the app did not impair CBTI effectiveness and potentially improved patient adherence [28]. A subsequent variant of the app called Insomnia Coach was applied as a self-help CBTI intervention on 25 veterans with insomnia and showed significant treatment effects compared to the waitlist control [29].

In addition to web-based and mobile CBTI, multiple automated sleep recommender systems (RSs) have also been developed to provide personalized sleep suggestions. SleepCoacher collects

XSL•FO
RenderX

data from smartphone sensors and user self-reports to compute sleep variables and includes recommendation templates reviewed by clinicians for each variable combination [30]. To deliver a recommendation, the system chooses the template of the most correlated combination, further personalizes it according to the user's sleep status, and also evaluates the effectiveness of the recommendation. In total, 2 studies on 43 undergraduate students showed the benefits of this system on sleep, which were greater when users adhered more to the system. Daskalova et al [31] explored cohort-based sleep recommendations in which their system used demographic, exercise, and sleep data collected from 1 million users of a wearable device to assign each participant to a user group in which all members had similar profiles and chose a recommendation that showed the greatest improvement among the group in some aspects of sleep behaviors on which the participant performed poorly compared to these similar users. Compared to general sleep suggestions, cohort-based recommendations improved sleep duration more, and many participants were motivated to follow these recommendations. Pandey et al [32] used event mining to discover optimal causal relationships between a user's lifestyle and sleep and provided recommendations to match their current behaviors and environment with these relationships for better sleep. On the other hand, some systems do not directly provide advice to users but, instead, help them self-adjust their sleep habits. For example, Lullaby uses multiple sensors to track users' sleep and environment and helps them discover environmental factors with negative impacts on their sleep [33]. ShutEye serves as a peripheral display on mobile phones to remind users of recommended and inadvisable time windows for various sleep-related activities [34]. SleepBandits enables users to perform numerous self-experiments to adjust their behaviors and uses data from mobile phone sensors to evaluate experiment effects on their sleep [35].

Although existing automated and individualized CBTI and sleep RSs have demonstrated their great potential in improving sleep, they have not been modified to accommodate shift workers' irregular schedules of shift and wakefulness [13] and, thus, do not include strategies specifically beneficial to shift workers. Moreover, these CBTI interventions are mostly delivered through weekly learning modules and tasks requiring active involvement and a regular time commitment, and their personalization options are limited and mainly focus on sleep restrictions. Unlike CBTI, sleep RSs directly send behavior change recommendations to users, but their advice-tailoring processes often lack clinical support and have raised doubts about their credibility among users [31].

Objectives

We aimed to develop and evaluate a machine learning algorithm that predicts physicians' sleep advice using wearable and survey data. We also examined the model's potential in assisting physicians and even replacing them in providing sleep advice in the system.

We characterized shift workers' behaviors using clustering analyses, estimated physicians' advice selection strategies by interpreting the developed advice prediction models, and assessed the effectiveness of the provided sleep recommendations to gain more insights into the system.

Methods

Data Collection

The data were collected from shift workers, including physicians and nurses, in the intensive care unit (ICU) at Mie University Hospital (n=47) and Suzuka Chuo General Hospital (n=14) in Japan [36,37]. Eligible participants must have 8 hours of work per shift and poor sleep quality demonstrated through a Pittsburgh Sleep Quality Index score of ≥5. The detailed inclusion and exclusion criteria have been described in the work by Ito-Masui et al [36,37]. Each participant enrolled in a 5-week trial, with the first week as a baseline without an intervention (preintervention time point) and the following 4 weeks with interventions. During the 5 weeks, participants filled out daily surveys and wore a Fitbit (Google) to collect physiological and behavioral data. Physicians reviewed the data and provided sleep advice to participants as interventions 3 to 4 times a week. In total, 2123 days of multimodal data were collected from these participants for the entire study.

Ethical Considerations

This study was approved by the Clinical Research Ethics Review Committee of Mie University Hospital, Tsu, Japan (H2020-083). This study was approved by the ethics committee of Suzuka General Hospital (review number 254). All procedures were performed in accordance with the ethical standards of the institutional research committee and the 2013 Declaration of Helsinki. This study was registered in the University Hospital Medical Information Network Clinical Trials Registry on May 1, 2020 (ID: UMIN000040547). Participants provided written consent after receiving sufficient information about the study. They retained the right to opt out of the study at any time. To protect privacy, all collected data were deidentified. No compensation was offered for participation.

Data Modalities

The collected data included 3 main modalities—Fitbit data, survey data, and sleep advice. We followed the methods by Yu et al [38] to extract some features from the same Fitbit and survey data, and engineered extra features to accommodate our objectives.

Fitbit Data

Raw data from Fitbit included minute-by-minute steps and heart rate as well as the start time, end time, duration, and efficiency of each sleep period. From raw heart rate data, we computed their average, SD, and sample entropy. Sample entropy has been used for cardiovascular time series, and a low entropy value suggests greater self-similarity of the series [39]. We also computed features from step count per minute to account for the variability in participants' daily activities. Specifically, we computed and stored the duration of each continuous segment with and without steps in minutes. We then computed information entropy for the stationary and active segments. A higher entropy corresponded to more variability in participants' moving or nonmoving behaviors. Finally, we retrieved

XSL•FO
RenderX

information about the main sleep period with the longest duration from raw sleep data.

Survey Data

Participants received a morning survey and an evening survey every day with questions about their daily behaviors and well-being. Features were then extracted from their answers.

Morning Survey

Features from the morning surveys were divided into 3 categories—sleep, well-being, and miscellany.

Sleep

In the morning surveys, participants indicated whether they had slept in the previous 24 hours and provided the number of naps between 8 AM and 8 AM on the following day. They also described how they woke up (naturally, via an alarm, or via other means) and reported the time it took for them to fall asleep, the duration of their phone use before sleep, the brightness level of their sleep environment, and sleep quality. Specifically, sleep quality was quantified through the answers to the following statements using 5-point Likert scales (*strongly disagree*=1; *strongly agree*=5): (1) *I slept soundly*, (2) *I fell asleep immediately*, (3) *I was able to recover from fatigue*, (4) *I didn't wake up in the middle of sleep*, and (5) *I was satisfied with sleep*.

In addition to the aforementioned features retrieved directly from surveys, we constructed some sleep features using available survey data. In the surveys, participants reported the start time and end time of their main sleep and nap periods. According to the time entries, we constructed a binary sequence with 1-minute resolution for every day, in which a bit of 1 meant that the participants were sleeping or napping at that minute of the day. If time entries of the main sleep periods were missing from the surveys, the entries from Fitbit were used. From the binary sequences, we computed the duration of the main sleep and nap periods. We also calculated the sleep regularity index (SRI) with sliding windows of 3, 5, and 7 days. The SRI represents the probability of any 2 time points that are 24 hours apart having the same sleep or awake state averaged over a specified time window and is scaled to range from −100 to 100, where a value of 100 indicates the same sleep schedule across all days and a value of −100 means that an individual has completely flipped their sleep schedule between any 2 consecutive days [40]. Previous studies have shown that the SRI is positively correlated with the academic performance of college students [40] and a lower SRI is associated with increased stress, depression, and risk of cardiovascular diseases in older adults [41].

Well-Being

Participants reported 6 well-being metrics in the morning surveys. A total of 5 metrics—alertness, happiness, energy, health, and calmness—were reported using a continuous scale from 1 to 100, with 100 as the most positive. The sixth metric, current sleepiness, was reported via a 9-point Likert scale from *strongly awake* (1) to *strongly sleepy* (9).

Behavior

The morning surveys also recorded participants' daily behaviors, including the amount of caffeine intake and last intake time, amount of alcohol intake, and bath time. From the time entries, we computed the durations between their last caffeine intake and sleep and between bath and sleep.

Evening Survey

Features from the evening surveys were divided into 2 categories—work and well-being.

Work

In the evening surveys, participants reported their work schedules for the current day, the day before, and the day before the day before in binary sequences with 30-minute resolution, where a bit of 1 meant work during that 30-minute interval. From the binary sequences, we computed work hours for the entire day as well as for 3 different periods of the day (1: midnight-8 AM, 2: 8 AM-4 PM, and 3: 4 PM-midnight). The periods corresponded to midnight shifts, day shifts, and afternoon shifts of the participants. Moreover, by comparing the binary work sequences and binary sleep sequences constructed from the morning surveys, we calculated participants' nap durations during work and minimum durations between the last sleep periods and the start of shifts and between the end of shifts and following sleep periods. A binary indicator of whether participants did any activities other than work outside their homes was also retrieved from the surveys.

Well-Being

Similarly, as in the morning surveys, participants also answered several questions about their well-being in the evening surveys. In addition to the 6 metrics reported in the morning surveys (alertness, happiness, energy, health, calmness, and current sleepiness), they provided 3 other metrics: sleepiness during the day, stress, and tiredness, which were all reported via 5-point Likert scales (5=*strongly awake*, *stressful*, or *tired*).

Sleep Advice

Starting from the second week of enrollment, physicians reviewed Fitbit and survey data to provide sleep advice. One physician provided advice once a week to all participants, and 2 physicians gave advice 2 to 3 times a week to participants from each of the 2 hospitals. Over the 4 weeks with interventions, participants received advice approximately 13 times, and a total of 786 pieces of advice were sent to the 61 participants who completed their 5 weeks of the study. Every time physicians provided sleep advice, they chose 1 to 5 messages from 23 options, listed in Table 1. To better summarize the options, we divided them into 6 categories according to their descriptions: (1) dietary intake (messages 1, 2, and 3), (2) activity (messages 4, 5, 6, 16, 17, and 18), (3) sleep (messages 7, 8, 9, 10, 20, 21, and 22), (4) shift (message 11), (5) nap (messages 12, 13, 14, 15, and 23), and (6) mentality (messages 19). After participants received the sleep advice, they were able to respond to the advice as *eager to follow* or *difficult to follow*.

XSL•FO
RenderX

Table 1. Content of the 23 sleep advice messages that physicians sent to the participants.

Message category and ID	Message content
Dietary intake	
1	"Refrain from consuming alcohol before sleep."
2	"Stop consuming caffeine 3 hours before lights-out."
3	"Refrain from eating midnight snacks."
Activity	
4	"Refrain from using a smartphone in the bedroom."
5	"Take a bath a little earlier than usual."
6	"Refrain from exercising 3 hours before lights-out."
16	"Relax. 1) Make a habit of relaxing before sleep, including taking a bath in warm water, light stretching, using aroma, and drinking herbal tea."
17	"Relax. 2) Use the Fitbit breathing program called Relax."
18	"Do moderate exercise regularly."
Sleep	
7	"Darken the bedroom when you go to bed."
8	"When you do not fall asleep within 15 minutes, leave the bedroom and stay out of the bedroom until you get sleepy."
9	"Set your wake-up time according to each work shift."
10	"Do not try too hard to fall asleep quickly."
20	"Continue current sleep habits."
21	"Try to make enough time for sleep."
22	"Get up at the same time, whether you are working or on holiday."
Shift	
11	"Management of work shift: Consult with your superior so that shift rotation will be clockwise in general."
Nap	
12	"Take a nap. 1) If possible, take a nap for approximately 90 minutes before the night shift."
13	"Take a nap. 2) If possible, take a nap for 15–20 minutes during rest time while you are working."
14	"Take a nap. 3) If possible, take a nap earlier after a late-night shift to refresh."
15	"Create an environment for taking a nap: Ask your family for cooperation to create a quiet environment during a nap in the daytime or evening at home. Sharing your work shift schedule with your family by placing it where every member of your family can see it might be a good idea. Avoid strong lights for several hours before a nap and darken the room during a nap."
23	"Do not nap for too long."
Mentality	
19	"Be broad-minded and try to approach things positively."

Participant Characteristic Analyses

We investigated the characteristics of the participants' daily activities to find participant subgroups who had similar behaviors and understand the relationship between their behaviors and well-being. We used agglomerative hierarchical clustering to find the groupings. The algorithm initiates all samples as individual clusters and merges clusters using Ward linkage to minimize the within-cluster variances at every step [42,43]. After all the samples are merged, a dendrogram is constructed to show the merged path for each sample. By setting thresholds on the height of the dendrogram, clusters and their samples can be identified.

We used daily features from Fitbit and the surveys, except for the participants' well-being and sleep quality metrics, to find the clusters. To include as many samples as possible during clustering, we did not use the following features as they were not available for some days: SRI; nap duration during work; and the time between caffeine intake and sleep, between bath and sleep, between last sleep and the start of shift, and between the end of shift and following sleep (eg, participants did not work or take a bath). All features were standardized across all participants for clustering.

We then performed principal-component analysis on the standardized features and overlaid the cluster assignments on a 2D t-distributed stochastic neighbor embedding (t-SNE) plot

XSL•FO
RenderX

generated from the principal components that explain 95% of the total variance to observe how clusters interacted with each other. t-SNE creates a low-dimensional mapping of high-dimensional data by matching the conditional probabilities in both spaces that are proportional to the similarity of the data points [44]. To characterize participants' behaviors and well-being, we summarized the clusters' profiles by observing feature averages and examined distributions of the well-being metrics for each cluster.

Advice Prediction Models

As mentioned in the *Sleep Advice* section, physicians reviewed the Fitbit and survey data to provide sleep advice to participants. We developed machine learning models to predict their message selections and evaluated the performance of the advice prediction models.

Prediction Approach

The advice prediction task can be formulated as a binary classification, with the positive class representing the selection of an advice message. We found substantial class imbalances for the selected messages. Specifically, of 786 advice pieces, the number of selections ranged from 64 (8.1%) to 390 (49.6%) for the 7 most frequent advice messages (messages 4, 7, 12, 14, 15, 20, and 21). Consequently, we decided to build individual binary classifiers only for these messages as there would be too few positive samples to develop models for less frequent advice.

To construct samples for the machine learning models, we computed the average and SD of daily Fitbit and survey features across the previous 4 days for each date on which physicians provided advice. Samples with missing values were excluded. According to the physicians, they often considered the participants' responses to previous advice, especially when they found it difficult to choose the messages. Therefore, we incorporated advice responses into our models and enabled the models to output class probabilities to indicate certainty about a message selection. We used 12 Fitbit features (steps, sleep, and heart rate), 53 morning survey features, 50 evening survey features, and 23 advice response features. Table S1 in Multimedia Appendix 1 summarizes the features used by the classifier.

Dataset Split

To evaluate the advice prediction models, the data needed to be split into a training set and a test set. We designed 2 ways to split the data: a user-dependent split and a user-independent split. Under the user-dependent setting, the data were split chronologically, where the training set contained the first 70% of each participant's data and the testing set had the remaining 30%. Under the user-independent setting, we split the data according to the order of participant enrollment such that the models were trained on data from the first 70% of participants enrolled and tested on the 30% of participants who enrolled later in the study. We chose these train-test split methods to ensure that, when predicting sleep advice for a date of a participant, no data from later days of this participant's enrollment would be used in dependent models and only data from participants enrolled earlier would be used in independent models. In particular, the dependent split examined the effects of personalization by leveraging participants' past data to predict their future sleep advice, whereas the independent split assessed model generalizability by predicting sleep messages for unseen participants.

Classifiers

Considering the abundant features but relatively small size of our data, we focused on ensemble classifiers for their feature selection capabilities and robustness to overfitting. In particular, we chose random forest (RF), a bagging-based model, and 2 gradient-boosting models, namely, light gradient-boosting machine (LGBM) [45] and CatBoost [46], as our candidates. To find the optimal sets of model hyperparameters, we tuned the classifiers on the training set using 5-fold stratified cross-validation and 100 iterations of random search over a grid of parameters. Random search was used instead of grid search because it can span a larger parameter space and find models as good as or better than those found using grid search under the same computation budget [47]. When tuning the classifiers under the user-independent setting, we made sure that the training data were divided into subsets with nonoverlapping participants. Finally, classifiers with optimal hyperparameters were fitted on the entire training set.

Data Balancing

To address class imbalances that could hinder the development of effective advice prediction models, we experimented with 2 data-balancing methods: random oversampling and synthetic minority oversampling technique (SMOTE) [48]. Random oversampling increases the representation of the minority class by duplicating existing samples through sampling with replacement. In contrast, SMOTE enhances the minority class by generating synthetic data points through interpolation between existing samples and their nearest neighbors. Data balancing was only applied to the training data during classifier tuning and training. To better understand the effectiveness of data balancing, we also trained models on the original imbalanced data for comparison.

Prediction Evaluation

By combining 3 classifiers with 3 data-balancing approaches, we trained 9 model configurations to predict the selection of a particular advice message under user-dependent and user-independent settings. On the basis of 2 key considerations, we chose the area under the precision-recall curve (AUPRC) as the main metric to evaluate and compare the performance of the different models. First, given the relatively few occurrences of sleep advice, it was crucial for the model to provide advice when a participant needed it. The precision-recall curve is particularly well suited for this scenario because it focuses on the model's ability to correctly predict the positive class, which is actionable sleep advice. Second, due to class imbalances, the default decision threshold of 0.5 may not be optimal for converting the model's probability outputs into predicted classes. As the precision-recall curve reflects the trade-off between false alarms and missed advice predictions across all decision thresholds, the AUPRC provides a comprehensive evaluation of the model's ability to predict sleep advice without requiring a specific decision threshold.

XSL•FO
RenderX

For each model configuration, we repeated the tuning and training process 10 times with different random seeds and computed the average AUPRC on the test set. To ensure a fair comparison, the same set of seeds was used across all configurations. In addition, we performed 10 random guessing trials by sampling probability predictions from a standard uniform distribution and used a 1-sided Welch *t* test to assess whether the model performance significantly exceeded random guessing. This procedure was repeated for each of the 7 most frequent advice messages.

Feature Importance

After training the advice prediction models, we analyzed feature importance using 2 approaches to identify the key contributors to predictions on the test set. The first approach was permutation importance, which evaluates feature importance by measuring the decrease in prediction performance after randomly permuting a specific feature [49]. Greater performance degradation indicates higher feature importance. Using the AUPRC as the performance metric, we permuted each feature 10 times per classifier run and summarized the results on the test set. The second method was Shapley additive explanations (SHAP) [50,51], which leverages game theory to explain outputs of machine learning models at both local and global levels. In particular, we computed the mean absolute SHAP values across all test samples per classifier run to determine global feature importance. Both permutation and SHAP importance values were reported as averages across all repeated runs of each model configuration.

Results

Participant Characteristics

Tables 2 and 3 summarize the participants' daily features and well-being scores, respectively. Their average step count was 7545 (SD 4419), their average sleep duration was 6.77 (SD 2.87) hours, their average nap duration was 53 (SD 108.2) minutes, their average sleep regularity was 52 (SD 19.3), and their average work duration was 5.62 (SD 4.19) hours.

We used 77.01% (1635/2123) of the samples with all selected features available for hierarchical clustering. Figure 1 shows the resulting dendrogram. The default distance threshold of 52.24 (purple dashed line) identified 6 clusters color coded and labeled with numbers from 0 to 5 in the dendrogram. In addition, a few smaller clusters merged at approximately 30, so we used a lower threshold of 31 (red dashed line) to divide the 6 clusters into 13 subclusters (*a*-*m*) and further investigated the behavior variability within each larger cluster. Specifically, cluster 0 had 5 subclusters: *c*, *f*, *g*, *e*, and *m*. Cluster 1 had 2 subclusters: *a* and *b*. Cluster 2 had 2 subclusters: *k* and *d*. Clusters 3 and 4 did not have any subclusters and themselves were labeled as *j* and *h*, respectively. Cluster 5 had 2 subclusters: *l* and *i*. Figure S1 in Multimedia Appendix 1 illustrates the clusters and subclusters overlaid on t-SNE plots obtained from the same data used for clustering. Cluster and subcluster structures were clearly observable in both plots. For example, cluster 0 in the plot of 6 clusters has an upper branch and a lower branch on either side of cluster 1, which are further divided into subclusters *c* and *g*, respectively, in the plot of 13 subclusters.

Table S2 in Multimedia Appendix 1 summarizes cluster and subcluster profiles for different aspects of participants' behaviors and well-being. Major differences among the 6 clusters were observed in the well-being metrics, whereas subclusters can also be distinguished by work duration and shift types. The profiles will be discussed in detail later in the *Discussion* section.

XSL•FO
RenderX

Table 2. Participants' daily physiological and behavioral features.

Feature	Values
hrmean[a] (bpm), mean (SD)	74.87 (8.6)
hrstd[b] (bpm), mean (SD)	13.42 (3.45)
hrentropy[c], mean (SD)	0.63 (0.21)
steps, mean (SD)	7545 (4419.36)
duration_entropy_non_step, mean (SD)	2.34 (0.34)
duration_entropy_step, mean (SD)	1.69 (0.32)
nap_count, mean (SD)	0.56 (0.69)
brightness_sleep (scale of 1-100), mean (SD)	18.63 (23.09)
sleep_duration (min), mean (SD)	406.46 (172.01)
nap_duration (min), mean (SD)	52.99 (108.23)
sleep_regularity_3 days (scale of –100 to 100), mean (SD)	52.25 (28.3)
sleep_regularity_5 days (scale of –100 to 100), mean (SD)	52.1 (22.6)
sleep_regularity_7 days (scale of –100 to 100), mean (SD)	52.02 (19.26)
deep_sleep (scale of 1-5), mean (SD)	3.2 (1.14)
immediate_sleep (scale of 1-5), mean (SD)	3.55 (1.14)
fatigue_recover (scale of 1-5), mean (SD)	2.84 (1.05)
mid_awake (scale of 1-5), mean (SD)	3.2 (1.31)
sleep_satisfy (scale of 1-5), mean (SD)	2.85 (1.02)
alcohol_amount (cups), mean (SD)	0.3 (0.63)
caffeine_amount (cups), mean (SD)	1.53 (1.46)
worktime_today_duration (h), mean (SD)	5.62 (4.19)
worktime_today_duration (1: midnight-8 AM; h), mean (SD)	1.52 (2.85)
worktime_today_duration (2: 8 AM-4 PM; h), mean (SD)	2.15 (3.25)
worktime_today_duration (3: 4 PM-midnight; h), mean (SD)	1.96 (3.08)
extrawork_activities (n=1865), n (%)	924 (49.54)
sleep_prev_24[d] (n=2092), n (%)	2088 (99.81)
wake_natural (n=1890), n (%)	793 (41.96)
wake_alarm (n=1890), n (%)	882 (46.67)
wake_other (n=1890), n (%)	198 (10.48)

[a]hrmean: average daily heart rate.

[b]hrstd: SD of daily heart rate.

[c]hrentropy: entropy of daily heart rate.

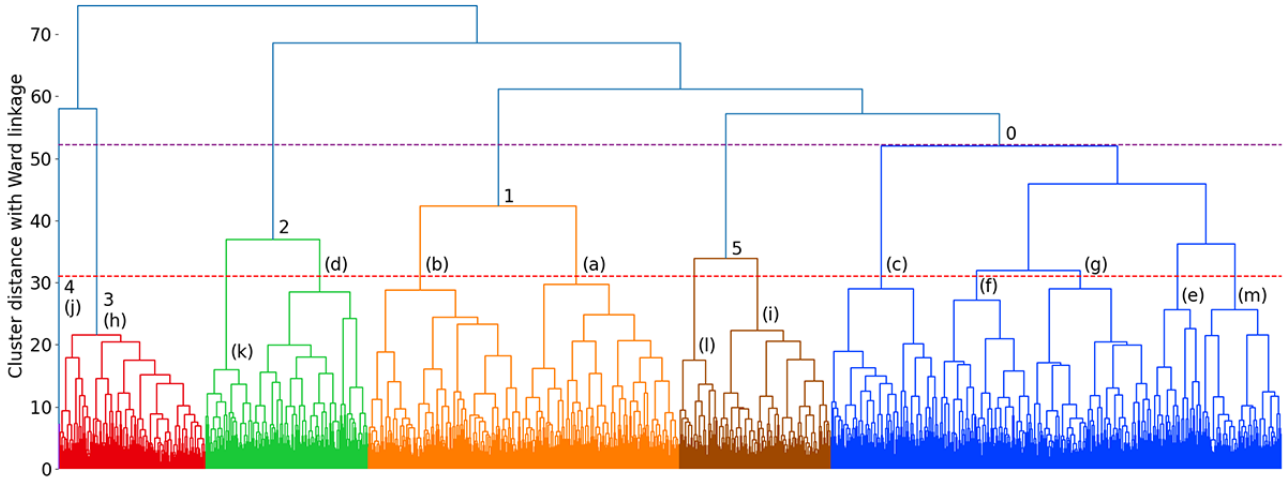
[d]sleep_prev_24: binary indicator of sleep in the previous 24 hours.

XSL•FO
RenderX

Table 3. Mean and SD of participants' daily well-being scores. Morn indicates morning and eve indicates evening.

Well-being metric	Values, mean (SD)
sleepiness_morn (scale of 1-9)	5.49 (1.99)
sleepiness_now_eve (scale of 1-9)	5.7 (1.78)
sleepiness_daytime (scale of 1-5)	2.59 (1.23)
stress (scale of 1-5)	2.93 (1.1)
tiredness (scale of 1-5)	2.51 (1.01)
alertness_morn (scale of 1-100)	41 (23.02)
happiness_morn (scale of 1-100)	52.99 (17.34)
energy_morn (scale of 1-100)	44.02 (21.27)
health_morn (scale of 1-100)	51.71 (18.12)
calmness_morn (scale of 1-100)	55.97 (18.1)
alertness_eve (scale of 1-100)	41.08 (21.1)
happiness_eve (scale of 1-100)	54.27 (18.18)
energy_eve (scale of 1-100)	45.88 (21.08)
health_eve (scale of 1-100)	52.4 (18.19)
calmness_eve (scale of 1-100)	56.85 (19)

Figure 1. Dendrogram from hierarchical clustering of daily features using Ward linkage. The purple dashed line corresponds to a distance threshold of 52.24 (0.7 times the maximum cluster distance as implemented by SciPy [52]) and identifies 6 clusters labeled and color coded (cluster 0=blue, cluster 1=orange, cluster 2=green, cluster 3=red, cluster 4=purple, and cluster 5=brown). The red dashed line represents a distance threshold of 31 and identifies 13 subclusters (a-m).



Advice Statistics

Figure 2 shows the frequency of advice provided as well as the message categories in a pie chart. As physicians often selected multiple messages for each day, 1332 advice messages were selected in 786 advice pieces. The outer ring of the pie chart indicates the proportions of each category with respect to the 1332 total occurrences of all message options, and the numbers in parentheses after each category name correspond to the frequency of the category. The inner ring shows the proportion of each message with respect to the total occurrences. The legend lists the corresponding color of each message, and each parenthesis includes the proportion of all advice pieces containing that message, which is computed by dividing the frequency of the message by 786 total advice counts. Messages in the sleep and nap categories were selected more frequently than messages from other categories. Messages in the sleep category were selected the greatest number of times (817/1332, 61.34%), and messages in the nap category were selected 21.7% (289/1332) of the time. Moreover, only a few messages were frequently selected by physicians. For example, message 20 (*continue current sleep habits*) was selected in 49.6% (390/786) of advice pieces, and message 21 (*try to make enough time for sleep*) was selected in 28.1% (221/786) of advice pieces. On the other hand, 70% (16/23) of the messages were selected in <50 advice pieces (50/786, 6.4%), and 7 messages were selected <10 times (10/786, 1.3%).

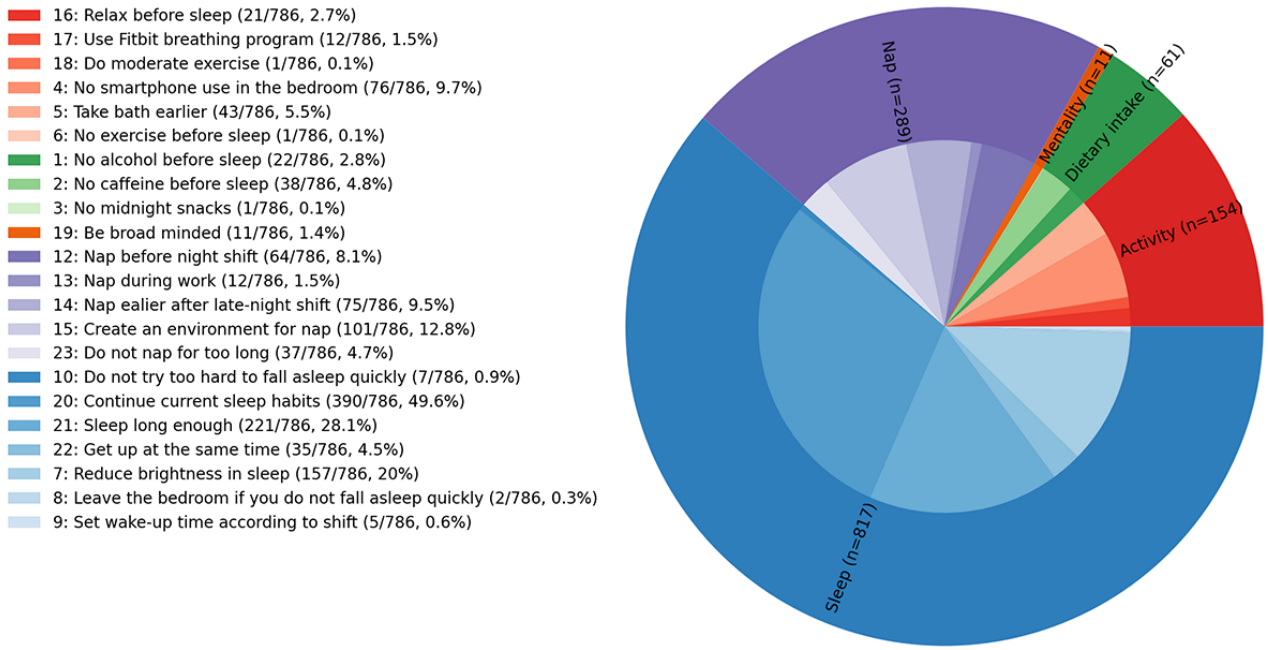
Figure 3 summarizes the frequency and proportion of responses to advice pieces containing each message. Sometimes,

XSL•FO
RenderX

participants did not respond to the advice. Although multiple messages might be selected for 1 advice piece, participants responded to the advice but not to individual messages. Participants showed different preferences for messages. For example, participants expressed difficulty following message 20 (*continue current sleep habits*) only 0.5% (2/390) of the time, whereas they experienced difficulty following message 4 (*refrain from using a smartphone in the bedroom*) 20% (15/76) of the time and following message 2 (*stop consuming caffeine 3 hours before lights-out*) 26% (10/38) of the time. The discrepancies might be explained by the message design. Message 20 suggests that participants continue their current sleep habits, which participants might feel is easier to follow as it would not change their current lifestyle greatly. On the other hand, message 2 and message 4 address caffeine consumption and phone use, which might be challenging to follow as it is suggested that participants change some essential aspects of their routine.

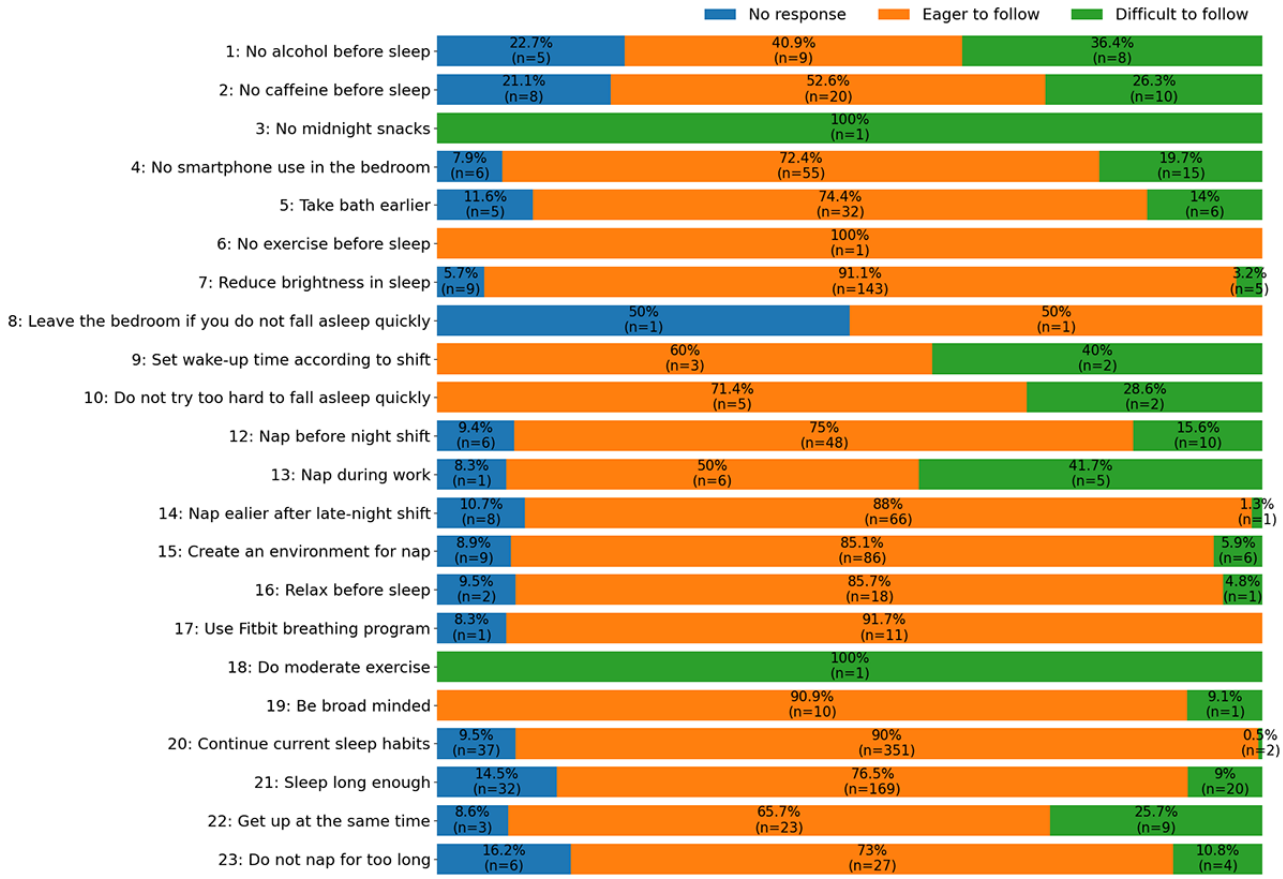
Table S3 in Multimedia Appendix 1 summarizes the distributions of positive and negative advice labels for both the training and test data under user-dependent and user-independent settings. In addition, it includes the ratios of negative-to-positive sample counts. Class imbalances were evident for all messages except message 20 (*continue current sleep habits*). In both data split settings, messages 21 (*try to make enough time for sleep*) and 7 (*darken the bedroom when you go to bed*) showed moderate imbalances, with negative-to-positive ratios of <5. For the remaining 4 messages, class imbalances were more pronounced, with negative-to-positive ratios mostly ranging from 5 to 10. Furthermore, we observed substantial discrepancies in class imbalance between the user-independent training and test sets of messages 12 (*take a nap for approximately 90 minutes before the night shift*), 14 (*take a nap earlier after a late-night shift to refresh*), and 15 (*create an environment for taking a nap*). While negative-to-positive ratios remained <10 in the training sets, the ratios exceeded 15 in the test sets, indicating distinct preferences among physicians when providing sleep advice to different shift workers.

Figure 2. The pie chart shows the selection frequency of each sleep advice message. All 23 message options are divided into 6 categories: dietary intake (messages 1, 2, and 3), activity (messages 4, 5, 6, 16, 17, and 18), sleep (messages 7, 8, 9, 10, 20, 21, and 22), shift (message 11), nap (messages 12, 13, 14, 15, and 23), and mentality (message 19). Message 11 is excluded from the chart because of zero frequency.



XSL•FO
RenderX

Figure 3. Frequency and proportion of message responses in a stacked bar plot. Each bar includes the frequency and proportion of participants' responses (no response, eager to follow, and difficult to follow) to advice pieces containing each message. Although multiple messages might be selected for 1 advice piece, participants responded to the advice but not to individual messages.



Advice Prediction Results

Following the approach described in the *Methods* section, we obtained the AUPRC of 9 different configurations of classifiers and data-balancing methods on the test set under dependent and independent settings. Table 4 summarizes the performance of the proposed models, where each model configuration was tuned and trained 10 times with different random seeds, as well as results from 10 random-guessing trials. The model performance was compared with random guessing using a 1-sided Welch *t* test.

As shown in Table 4, under the user-dependent setting, RF performed the best for predicting 6 advice messages. The only exception was advice message 7 (*darken the bedroom when you go to bed*), where CatBoost achieved the highest AUPRC. However, RF's AUPRC was only 0.02 lower than that of CatBoost. Under the user-independent setting, RF remained competitive, achieving the best performance for predicting 3 advice messages. Its performance on other messages was also close to that of the top performers. LGBM yielded the highest prediction AUPRC for messages 12 (*take a nap for approximately 90 minutes before the night shift*) and 14 (*take a nap earlier after a late-night shift to refresh*). CatBoost was the top performer for messages 7 (*darken the bedroom when you go to bed*) and 21 (*try to make enough time for sleep*). These results highlight the effectiveness of RF in mitigating overfitting when working with limited data. While models trained on balanced data achieved the best performance on 4 advice messages in both dependent and independent settings, data-balancing techniques overall did not result in a substantial improvement in AUPRC and even degraded the performance for some advice labels. Statistical tests revealed that the best prediction models for each advice message and data split setting performed significantly better than random guessing, with $P<.001$ for 13 models and $P=.003$ for the remaining one.

XSL•FO
RenderX

Table 4. Advice prediction area under the precision-recall curve (AUPRC) of the proposed models (random forest [RF], light gradient-boosting machine [LGBM], and CatBoost) and random guessing baseline obtained under different data-balancing approaches (none, random oversampling, and synthetic minority oversampling technique [SMOTE]) and dataset split settings (user dependent and user independent). A 1-sided Welch *t* test was used to compare the models to random guessing.

Setting, message ID, and balancing method	RF (AUPRC)		LGBM (AUPRC)		CatBoost (AUPRC)		Random guessing (AUPRC), mean (SD)
	Values, mean (SD)	*P* value	Values, mean (SD)	*P* value	Values, mean (SD)	*P* value	
Dependent							
4							0.12 (0.02)
None	*0.33 (0.07)* [a]	*<.001*	0.28 (0.06)	<.001	0.24 (0.04)	<.001	
Random oversampling	0.25 (0.05)	<.001	0.21 (0.04)	<.001	0.20 (0.03)	<.001	
SMOTE	0.24 (0.03)	<.001	0.18 (0.02)	<.001	0.18 (0.02)	<.001	
7							0.17 (0.02)
None	0.56 (0.04)	<.001	0.54 (0.01)	<.001	0.57 (0.02)	<.001	
Random oversampling	0.54 (0.03)	<.001	0.52 (0.03)	<.001	0.54 (0.03)	<.001	
SMOTE	0.52 (0.03)	<.001	0.56 (0.02)	<.001	*0.58 (0.02)*	*<.001*	
12							0.12 (0.02)
None	*0.32 (0.03)*	*<.001*	0.23 (0.03)	<.001	0.27 (0.06)	<.001	
Random oversampling	0.31 (0.04)	<.001	0.21 (0.04)	<.001	0.26 (0.07)	<.001	
SMOTE	0.19 (0.03)	<.001	0.17 (0.01)	<.001	0.22 (0.05)	<.001	
14							0.10 (0.02)
None	0.26 (0.05)	<.001	0.24 (0.04)	<.001	0.23 (0.05)	<.001	
Random oversampling	0.25 (0.04)	<.001	0.22 (0.04)	<.001	0.23 (0.05)	<.001	
SMOTE	*0.29 (0.05)*	*<.001*	0.22 (0.05)	<.001	0.22 (0.04)	<.001	
15							0.19 (0.03)
None	*0.31 (0.03)*	*<.001*	0.26 (0.05)	.001	0.25 (0.05)	.007	
Random oversampling	0.28 (0.02)	<.001	0.27 (0.05)	<.001	0.22 (0.04)	.06	
SMOTE	0.25 (0.01)	<.001	0.24 (0.02)	<.001	0.24 (0.02)	.002	
20							0.57 (0.02)
None	0.80 (0.02)	<.001	0.76 (0.02)	<.001	0.79 (0.02)	<.001	
Random oversampling	0.80 (0.01)	<.001	0.77 (0.02)	<.001	0.80 (0.02)	<.001	
SMOTE	*0.81 (0.02)*	*<.001*	0.77 (0.02)	<.001	0.79 (0.01)	<.001	
21							0.23 (0.02)
None	0.38 (0.03)	<.001	0.38 (0.03)	<.001	0.41 (0.01)	<.001	
Random oversampling	0.38 (0.03)	<.001	0.36 (0.02)	<.001	0.40 (0.03)	<.001	
SMOTE	*0.44 (0.04)*	*<.001*	0.40 (0.04)	<.001	0.44 (0.06)	<.001	
Independent							
4							0.11 (0.05)
None	*0.25 (0.03)*	*<.001*	0.17 (0.04)	.005	0.20 (0.03)	<.001	
Random oversampling	0.19 (0.04)	.001	0.16 (0.04)	.01	0.15 (0.03)	.02	
SMOTE	0.18 (0.06)	.007	0.12 (0.04)	.25	0.11 (0.02)	.58	
7							0.16 (0.02)
None	0.67 (0.05)	<.001	0.61 (0.03)	<.001	0.66 (0.03)	<.001	
Random oversampling	0.60 (0.08)	<.001	0.58 (0.06)	<.001	0.62 (0.04)	<.001	

XSL•FO
RenderX

Setting, message ID, and balancing method	RF (AUPRC)		LGBM (AUPRC)		CatBoost (AUPRC)		Random guessing (AUPRC), mean (SD)
	Values, mean (SD)	*P* value	Values, mean (SD)	*P* value	Values, mean (SD)	*P* value	
SMOTE	0.66 (0.06)	<.001	0.64 (0.04)	<.001	*0.70 (0.04)*	*<.001*	
12							0.03 (0.01)
None	0.07 (0.01)	<.001	*0.09 (0.01)*	*<.001*	0.06 (0.01)	<.001	
Random oversampling	0.06 (0.01)	<.001	0.08 (0.02)	<.001	0.07 (0.02)	<.001	
SMOTE	0.07 (0.01)	<.001	0.06 (0.01)	<.001	0.06 (0.01)	<.001	
14							0.06 (0.02)
None	0.09 (0.01)	.001	0.07 (0.01)	.18	0.10 (0.02)	<.001	
Random oversampling	0.10 (0.01)	<.001	0.07 (0.02)	.11	0.09 (0.01)	.001	
SMOTE	0.09 (0.02)	.002	*0.11 (0.04)*	*.003*	0.09 (0.03)	.004	
15							0.06 (0.01)
None	0.11 (0.02)	<.001	0.09 (0.03)	.003	0.09 (0.02)	.002	
Random oversampling	*0.11 (0.03)*	*<.001*	0.10 (0.02)	<.001	0.10 (0.02)	<.001	
SMOTE	0.08 (0.03)	.04	0.07 (0.01)	.07	0.08 (0.02)	.03	
20							0.53 (0.02)
None	0.64 (0.01)	<.001	0.64 (0.01)	<.001	0.62 (0.01)	<.001	
Random oversampling	0.64 (0.01)	<.001	0.64 (0.02)	<.001	0.63 (0.01)	<.001	
SMOTE	*0.64 (0.01)*	*<.001*	0.64 (0.02)	<.001	0.63 (0.01)	<.001	
21							0.24 (0.03)
None	0.36 (0.01)	<.001	0.38 (0.02)	<.001	0.34 (0.02)	<.001	
Random oversampling	0.35 (0.02)	<.001	0.36 (0.03)	<.001	0.35 (0.03)	<.001	
SMOTE	0.38 (0.02)	<.001	0.36 (0.03)	<.001	*0.39 (0.02)*	*<.001*	

[a]The best-performing configuration for each sleep advice message is italicized.

Important Features

We extracted permutation and SHAP feature importance for the best model configuration when predicting each advice message under both user-dependent and user-independent settings. Table 5 shows the features that ranked among the top 5 most important using both permutation importance and SHAP, with average AUPRC decrease and mean absolute SHAP values of at least 0.01. All the features represented the average and SD of daily Fitbit and survey features across the previous 4 days for each date on which physicians provided advice. As expected, several important features closely matched their corresponding advice messages under both data split settings. The average amount of phone use before sleep was a key factor in predicting message 4 (*refrain from using a smartphone in the bedroom*). Similarly, the average brightness of the sleep environment was the most significant feature for predicting message 7 (*darken the bedroom when you go to bed*). The model examined average sleep duration for predictions of message 21 (*try to make enough time for sleep*). When predicting message 12 (*take a nap for approximately 90 minutes before the night shift*), the SD of work hours between 4 PM and midnight on the current day across the previous 4 days was an important factor. This suggests a potential relationship between the frequency of night shifts and the triggering of this sleep advice. In addition, the response to the previous selection of message 20 (*continue current sleep habits*) contributed to the prediction of messages 7, 20, and 21. This aligns with physicians' insights that they often considered participants' responses to previous advice when providing sleep recommendations. These findings demonstrate the models' ability to capture meaningful patterns that resonate with physicians' decision-making processes.

XSL•FO
RenderX

Table 5. Important features identified using permutation importance and Shapley Additive Explanations (SHAP) for the best model configuration when predicting each advice message under both user-dependent and user-independent settings. Features are listed if they ranked among the top 5 most important using both permutation importance and SHAP, with average area under the precision-recall curve (AUPRC) decrease and mean absolute SHAP values of at least 0.01.

Setting, message ID, and feature	AUPRC decrease, mean (SD)	Absolute SHAP value, mean (SD)
Dependent		
4		
Average amount of phone use before sleep	0.19 (0.06)	0.05 (0.01)
7		
Average brightness of sleep environment	0.33 (0.05)	1.14 (0.25)
Response to previous selection of message 20	0.04 (0.03)	0.29 (0.06)
SD of brightness of sleep environment	0.03 (0.02)	0.18 (0.06)
SD of the binary indicator of activities other than work outside the homes	0.02 (0.01)	0.09 (0.04)
12		
SD of work hours on the current day for 4 PM-midnight	0.05 (0.07)	0.02 (0.01)
14		
SD of daily caffeine intake	0.03 (0.02)	0.02 (0.01)
20		
Response to previous selection of message 20	0.12 (0.03)	0.10 (0.04)
Average sleep satisfaction	0.01 (0.01)	0.02 (0.01)
Average fatigue recovery level	0.01 (0.01)	0.03 (0.01)
21		
Response to previous selection of message 20	0.11 (0.05)	0.07 (0.03)
Average sleep duration	0.03 (0.02)	0.02 (0.005)
SD of work hours on the current day for midnight-8 AM	0.03 (0.03)	0.01 (0.01)
Independent		
4		
Average amount of phone use before sleep	0.13 (0.04)	0.09 (0.02)
7		
Average brightness of sleep environment	0.47 (0.05)	1.35 (0.39)
Response to previous selection of message 20	0.03 (0.02)	0.18 (0.06)
12		
SD of work hours on the current day for 4 PM-midnight	0.03 (0.02)	0.28 (0.10)
14		
Response to previous selection of message 20	0.03 (0.04)	0.25 (0.10)
15		
Average sleep satisfaction	0.02 (0.03)	0.05 (0.04)
Average fatigue recovery level	0.02 (0.02)	0.02 (0.01)
20		
Response to previous selection of message 20	0.11 (0.02)	0.16 (0.03)
Average sleep satisfaction	0.02 (0.01)	0.03 (0.01)
21		
Response to previous selection of message 20	0.09 (0.04)	0.45 (0.07)
Average sleep duration	0.02 (0.03)	0.17 (0.08)

XSL•FO
RenderX

Discussion

Principal Findings

In this study, we used daily surveys and Fitbit to obtain physiological and behavioral data of ICU shift workers. Physicians reviewed the data to provide sleep advice from a list of 23 messages to the workers. We extracted features from the collected data and (1) conducted analyses to characterize participants' behaviors and advice and (2) constructed and evaluated 3 models (RF, LGBM, and CatBoost) and 3 data-balancing approaches (no balancing, random oversampling, and SMOTE) to predict sleep advice selections for the 7 most frequent messages sent by the physicians. The advice prediction models achieved higher AUPRC than the random guessing baseline for predicting all the 7 advice messages under both user-dependent and user-independent settings. The performance differences were statistically significant, with $P<.001$ for 13 tests and $P=.003$ for 1 test. Feature importance analysis revealed several important features that closely matched their corresponding advice messages sent to participants. For instance, for message 7 (*darken the bedroom when you go to bed*), the models primarily examined the average brightness of the sleep environment to make predictions. Similarly, the average amount of phone use before sleep was a key factor in predicting message 4 (*refrain from using a smartphone in the bedroom*). The analysis also supported physicians' insights that they often considered participants' responses to previous advice when providing sleep recommendations.

Participant Characteristics

Hierarchical clustering revealed different work shifts and behavior patterns among clusters and subclusters, and we investigated their relationships with their well-being scores. First, work shifts on the current day contributed to low well-being scores for clusters 3 and 5, and several well-being metrics were further distinguished by their different shift patterns. For example, cluster 3 had a much shorter average sleep duration of approximately 220 minutes and the lowest sleep quality. This was potentially caused by day shifts the day before and midnight shifts on the current day for almost all samples. In this cluster, participants finished their work at approximately 4 PM the day before; they then started the next shift period at approximately midnight and finished at approximately 8 AM on the following day. After work, they filled out the morning surveys and reported their well-being, sleep duration, and sleep quality. Low scores on these metrics reflected their tiredness shortly after work and lack of rest between work shifts.

On the other hand, most samples in cluster 5 did not have shifts the day before but had day shifts on the current day. As a result, participants could have more time for sleep, reflected by a longer average sleep duration of 400 minutes, better sleep quality, and higher well-being scores in the morning compared with samples in cluster 3. Such discrepancies in shift patterns and well-being metrics could even be observed between subclusters *d* and *k* of the same larger cluster 2. Although most samples of both subclusters had midnight shifts the day before, subcluster *d* majorly did not have shifts on the current day and achieved an average level of well-being scores. The profile of subcluster *k* was more similar to that of cluster 3, with midnight shifts on the current day and much worse well-being and sleep quality than in subcluster *d*, which is also supported by the t-SNE plot (right plot in Figure S1 in Multimedia Appendix 1), where subcluster *k* is closer to subcluster *h* (equivalence of cluster 3) than to subcluster *d*. Furthermore, most of cluster 1 did not have a work shift on the current day, which led to its high well-being scores as 1 day off probably helped participants relax. Nevertheless, shifts on the current day were not always negatively correlated with well-being. Specifically, cluster 0 had above-average well-being scores although most samples had work shifts on the current day.

Furthermore, several subclusters were characterized by some distinct features of their profiles. For example, subclusters *i* and *l* of cluster 5 had similar work patterns, but they differed significantly in average daily step count, where subcluster *i* had the highest count of >10,000 among all subclusters, whereas subcluster *l* had the lowest count of <2000. Although multiple clusters and subclusters contained samples with overtime shifts of >8 hours, subcluster *e* was the only one with >9 hours of average work duration for 3 consecutive days (10.6 hours the day before the day before, 9.4 hours the day before, and 11.5 hours on the current day). The long work duration might also contribute to other characteristics of this subcluster, including the shortest time to fall asleep and the highest daily caffeine consumption. Despite the existence of some outliers, the clusters and subclusters generally grouped well, as shown in Figure S1 in Multimedia Appendix 1, and their profiles summarized from average feature values provided many insights into participants' daily behaviors, shifts, and well-being and their interrelationships.

Advice Predictions

While we evaluated the advice prediction models using the AUPRC computed from predicted class probabilities, it is also important to understand the model performance in terms of actual predicted classes. Given the class imbalances, adjusting the decision thresholds was necessary to optimize prediction performance. Specifically, for each classifier initialized with its optimal hyperparameters, we performed 5-fold stratified cross-validation on the training set to determine the optimal threshold that maximized its F_1-score. The F_1-score was chosen as the performance metric to obtain a balance between missed advice predictions and false alarms. The optimal threshold was then applied to convert prediction probabilities into class labels on the test set. As we repeated the tuning and training process 10 times with different random seeds for each model configuration, we created an ensemble of these classifiers to mitigate potential overfitting. Final predictions were determined through majority voting across the ensemble. To resolve ties during voting, we used only 9 of the 10 classifiers generated for each model configuration. We report the sensitivity, specificity, precision, and F_1-scores of the ensembles from the RF, LGBM, and CatBoost classifiers with different data-balancing approaches in Tables S4, S5, and S6 in Multimedia Appendix 1, respectively.

XSL•FO
RenderX

Table 6 presents the performance of the model configurations with the highest F_1-score for predicting each sleep advice message under user-dependent and user-independent settings. Baseline F_1-scores, calculated by always predicting the positive class, are also provided for comparison. All models outperformed the baseline. Under the user-dependent setting, the models achieved sensitivity and specificity within the range of 0.5 to 0.9. Sensitivity tended to decrease for less frequent advice messages, whereas specificity showed the opposite trend. Under the user-independent setting, the models consistently demonstrated decent sensitivity, exceeding 0.7 across all advice messages, whereas specificity varied between 0.4 and 0.9. The higher sensitivity observed in this setting may suggest that the models were more effective at identifying universal decision rules shared across participants than at accounting for individual behavioral variability.

Table 6. Advice prediction sensitivity, specificity, precision, and F1-score of the best-performing ensemble classifier for each advice message under both user-dependent and user-independent settings. Ensemble classifiers were constructed through majority voting across 9 random forest (RF), light gradient-boosting machine (LGBM), or CatBoost classifiers trained using different random seeds. The decision threshold for each classifier was optimized to maximize the F1-score. The baseline F1-score from always predicting the positive class is also listed.

Setting and message ID	Model	Balancing method	Sensitivity	Specificity	Precision	F_1-score	Baseline F_1-score (always positive)
Dependent							
4	LGBM	None	0.56	0.86	0.34	0.42	0.20
7	CatBoost	Random oversampling	0.69	0.88	0.52	0.59	0.28
12	RF	None	0.50	0.83	0.27	0.35	0.19
14	RF	Random oversampling	0.58	0.81	0.27	0.37	0.19
15	RF	Random oversampling	0.55	0.72	0.29	0.38	0.30
20	RF	SMOTE[a]	0.84	0.50	0.67	0.75	0.71
21	CatBoost	None	0.81	0.58	0.34	0.48	0.35
Independent							
4	RF	None	0.95	0.54	0.17	0.29	0.17
7	CatBoost	SMOTE	0.73	0.93	0.63	0.68	0.25
12	LGBM	Random oversampling	1.00	0.60	0.07	0.14	0.06
14	CatBoost	None	0.91	0.68	0.12	0.22	0.09
15	RF	SMOTE	0.77	0.48	0.08	0.15	0.11
20	RF	SMOTE	0.82	0.46	0.61	0.70	0.67
21	RF	Random oversampling	0.81	0.44	0.31	0.45	0.38

[a]SMOTE: synthetic minority oversampling technique.

Single-Label Versus Multilabel Classification

As physicians could select multiple messages when providing sleep advice, it is feasible to formulate sleep advice prediction as a multilabel classification task, which outputs predictions for all message selections simultaneously. Compared to the current approach of training separate single-label classifiers for each message, multilabel classification has the advantage of leveraging information about message coselections during training, potentially improving prediction performance. However, this approach presents several challenges. First, some models do not inherently support multilabel classification and, instead, combine separate single-label classifiers for multilabel outputs. This strategy fails to capture and use the coselection information effectively. Second, the multilabel setting requires careful handling of data imbalances both within individual labels and across their co-occurrence patterns. Finally, multilabel classification complicates feature importance analysis, which aims to identify key contributors for predicting individual message selections. In fact, we trained a user-dependent multilabel RF classifier without any data balancing and did not observe substantial performance improvement compared to the single-label counterparts.

Comparison With Prior Work

Our system provided sleep and behavior recommendations for shift workers with poor sleep quality. Although our system did not follow the CBTI protocol, we borrowed elements from it, such as sleep hygiene advice and relaxation. We designed sleep recommendations specifically targeted for shift workers, including strategic napping and clockwise shift rotation. Our system offered several advantages over existing CBTI and sleep RSs. First, while most mobile and web-based CBTI systems require users to review sleep-related learning modules, our system reduced users' time commitment by sending them simple pieces of advice to follow, which could be beneficial as shift workers often have a poor work-life balance [53]. Consequently, our study yielded a low dropout rate of 5% (3/64), as opposed to the rates of other CBTI systems that range from 4% to 61% [15,16,18-23,26,29,54].

XSL•FO
RenderX

Moreover, our system received the most support from sleep specialists compared to existing sleep RSs, which only have limited or even no clinician support. For example, clinicians only provided recommendation templates for further customization by the automated SleepCoacher system [30]. The cohort-based sleep recommender by Daskalova et al [31] relied on profiles of similar users as the only recommendation source. The system developed by Pandey et al [32] constructed its recommendations based on discovered relationships between the user's lifestyle and sleep. The lack of clinician support could impair the credibility of the provided sleep recommendations among users. For instance, users of SleepCoacher suggested adding justifications for the advice [30], and several users of the cohort-based sleep recommender by Daskalova et al [31] thought that the recommendations were implausible and even untrustworthy as they contradicted the users' own beliefs. However, our system only delivered clinician-reviewed recommendations, which greatly enhanced their trustworthiness.

Furthermore, only our study has evaluated the accuracy of an algorithm for automatic sleep advice provision. Many existing CBTI systems have been able to provide sleep restriction and sleep hygiene recommendations automatically, but no analysis is available of how close these generated recommendations are to actual clinicians' advice. Similarly, sleep RSs have not compared their recommendations with professionals' suggestions. Specifically, clinicians prepared all the recommendations in the preliminary study of SleepCoacher, and the provisions were automated by a correlation-based algorithm in the final study, but the study did not evaluate the accuracy of the algorithm for substituting clinicians [30]. Despite the effectiveness of these systems, an analysis of algorithm accuracy could further justify provided recommendations and resolve users' doubts about their credibility. On the other hand, although our system is not automated at its current stage, we developed an algorithm and assessed its accuracy for automatic advice provision, and it showed decent performance and great potential in providing clinical-level sleep recommendations.

Limitations and Future Work

False Predictions

To investigate the possible causes of false-positive and false-negative predictions, we examined the relationship between feature values and advice predictions. Specifically, we chose the feature *brightness_sleep* (brightness level of the sleep environment) and 1 repetition of the dependent advice prediction model for message 7 (*darken the bedroom when you go to bed*). We selected this combination because the feature had a dominant effect on predictions of the message label, as shown in the *Important Features* section, and could demonstrate the limitations of the models more clearly with the existence of an explicit decision rule.

For each participant, we plotted the trend of the feature across all days of their enrollment and marked each predicted and the actual selection of the message separately. Figure S2 in Multimedia Appendix 1 includes such plots for 6 participants to illustrate incorrect predictions and limitations of the advice prediction models. First, the models did not have memories about previous message selections and might produce false positives as duplicate responses to previous feature values. For example, the model made 3 positive predictions on days 29, 31, and 33 in response to a peak of brightness on day 29 for participant 1102, yet only the prediction on day 29 was an actual positive. Similarly, 2 positive predictions were made for participant 1154 on days 26 and 29 for a brightness surge on day 26, but the latter was a false-positive. Shortening the feature averaging window before each message selection might eliminate some of these false positives, but it would increase the risk of producing false negatives when a day with a favorable value for message selection is excluded from the window.

Another scenario of false positives occurred for participants 1315 and 1181. Both participants received 4 false positives although their brightness values only fluctuated within a small range of approximately 30 (scale of 1-100). A closer look at individual decision trees of the trained RF showed that some trees considered SDs of brightness levels but most trees made predictions by comparing average brightness with a threshold. Thus, false-positive predictions were made because average brightness exceeded the thresholds of most decision trees despite the lack of brightness peaks. However, it could be tricky to make correct predictions under such a scenario. For example, physicians selected the message twice on days 9 and 16 for participant 1315 although brightness remained stable. A plausible explanation could be that physicians initially suggested that the participant lower the brightness because it was constantly above the threshold, but later, they stopped as neither lower levels nor high peaks of brightness were observed after the suggestions. Nevertheless, it would be very difficult for the model to capture such change.

Unlike false positives, most false negatives occurred only under 1 scenario: when brightness peaks were not high enough and adjacent days had very low feature values, which caused averages to fall below the threshold, for example, on day 33 for participant 1164 and on days 29, 31, and 33 for participant 1167. Measuring the relative height of a peak (difference between maximum and minimum values of the feature averaging window) instead of averaging might be helpful to eliminate these false negatives, but it may produce other false negatives when feature values remain high and stable. From illustrative plots of feature values and advice predictions, we demonstrated several causes of false-positive and false-negative predictions, but further investigations and experiments were necessary to properly improve the advice prediction models as some false positives might benefit participants and strategies to eliminate certain false predictions might introduce others.

System Usability

Although our developed algorithm achieved a decent performance in predicting sleep advice under both dependent and independent settings, the algorithm is limited to the 7 most popular advice messages among 23 choices due to infrequent occurrences of the remaining options. Therefore, further studies are necessary to gather enough data to enable predictions for less frequent advice labels.

We gathered feedback on the sleep advice models from clinicians (n=2) through questionnaires. They commented that they decided their suggestions mainly based on morning survey,

XSL•FO
RenderX

work shift, and wearable data. After they selected the suggestions, they referred to the output of the models and made the final decisions.

Future studies can compare the intervention effects of a clinician-only input system, an algorithm-assisted system, and a fully automated system to demonstrate the usefulness of the machine learning algorithm. It will also be beneficial to validate the generalizability of our advice prediction models with more diverse participant populations.

Given the effectiveness of existing sleep RSs with limited clinical support, we believe that our automated system has the advantage of being built on the knowledge of sleep specialists even if our algorithm still requires improvement to further boost performance. We chose to evaluate our models in user-dependent and user-independent settings instead of using leave-one-out cross-validation because the advice frequencies were extremely imbalanced across participants. However, such a scheme will fit real-world scenarios the best when models use all available data from current users to predict sleep advice for a new user. It is also worth experimenting with personalization beyond leave-one-out cross-validation by continuously incorporating data accumulated from new users as they progress to better adapt models to specific user characteristics.

Conclusions

We developed a web- and app-based system to provide individualized sleep and behavior advice with CBTI elements for shift workers. We collected data from shift workers (N=61) in the ICU at 2 hospitals in Japan for 5 weeks, which included Fitbit, survey, and sleep advice data. We used hierarchical clustering to characterize shift workers' physiological and behavioral features and identified clusters with distinct patterns. For example, days with low sleep duration and the lowest sleep quality were associated with a day shift on the day before and a midnight shift on the current day. Furthermore, we constructed and evaluated 3 machine learning models (RF, LGBM, and CatBoost) and 3 data-balancing approaches (no balancing, random oversampling, and SMOTE) to predict sleep advice selections for the 7 most frequent advice messages sent by the physicians. Our advice prediction models achieved higher AUPRC and F_1-scores compared to the baseline. The analysis of the feature importance of our models found several important features that closely matched their corresponding advice messages sent to participants. Although our current system requires physician input, an accurate machine learning algorithm would be promising for automation without hurting the trustworthiness of the selected recommendations. Despite a decent performance, the algorithms developed in this study are limited to the 7 most popular advice messages among 23 choices due to infrequent occurrences of the remaining advice options. Therefore, further studies are necessary to gather enough data to enable predictions for less frequent advice labels.

Acknowledgments

This research was funded by the Japan Agency for Medical Research and Development (19217687) and the National Science Foundation (1840167).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

AIM, RS, MS, and AS contributed to the conceptualization and methodology of the study. Data curation was performed by AIM, RS, and MS. YS, ACO, and HY conducted formal analyses. YS developed the software and performed validation and visualization. YS and ACO wrote the original draft, whereas AS reviewed and edited the manuscript. AS provided supervision and project administration. Both MS and AS were responsible for funding acquisition.

Conflicts of Interest

AS provided consulting services for Suntory Global Innovation Center and received honoraria from Oak Ridge Associated Universities, the Nara Institute of Science and Technology, the Taiwanese Society for Nutritional Psychiatry Research, the Korea Advanced Institute of Science and Technology, Amrita Vishwa Vidyapeetham, the European Science Foundation, and the National Science Foundation. AS also received travel support from Apple and the Taiwanese Society for Nutritional Psychiatry Research, as well as research funding from Meta Platforms, General Motors Company, POLA, and Nippon Electric Company.

Multimedia Appendix 1
Additional tables and figures for participant characteristic clustering and advice prediction models.
[DOCX File , 2765 KB - formative_v9i1e65000_app1.docx]

References

XSL•FO
RenderX

1. Alterman T, Luckhaupt SE, Dahlhamer JM, Ward BW, Calvert GM. Prevalence rates of work organization characteristics among workers in the U.S.: data from the 2010 National Health Interview Survey. Am J Ind Med 2013 Jun;56(6):647-659 [FREE Full text] [doi: 10.1002/ajim.22108] [Medline: 22911666]
2. Wickwire EM, Geiger-Brown J, Scharf SM, Drake CL. Shift work and shift work sleep disorder: clinical and organizational perspectives. Chest 2017 May;151(5):1156-1172 [FREE Full text] [doi: 10.1016/j.chest.2016.12.007] [Medline: 28012806]
3. Haile KK, Asnakew S, Waja T, Kerbih HB. Shift work sleep disorders and associated factors among nurses at federal government hospitals in Ethiopia: a cross-sectional study. BMJ Open 2019 Aug 27;9(8):e029802 [FREE Full text] [doi: 10.1136/bmjopen-2019-029802] [Medline: 31462478]
4. Flo E, Pallesen S, Magerøy N, Moen BE, Grønli J, Hilde Nordhus I, et al. Shift work disorder in nurses--assessment, prevalence and related health problems. PLoS One 2012 Apr 2;7(4):e33981 [FREE Full text] [doi: 10.1371/journal.pone.0033981] [Medline: 22485153]
5. Di Milia L, Waage S, Pallesen S, Bjorvatn B. Shift work disorder in a random population sample--prevalence and comorbidities. PLoS One 2013;8(1):e55306 [FREE Full text] [doi: 10.1371/journal.pone.0055306] [Medline: 23372847]
6. Cheng WJ, Cheng Y. Night shift and rotating shift in association with sleep problems, burnout and minor mental disorder in male and female employees. Occup Environ Med 2017 Jul;74(7):483-488. [doi: 10.1136/oemed-2016-103898] [Medline: 27810939]
7. Khan WA, Jackson ML, Kennedy GA, Conduit R. A field investigation of the relationship between rotating shifts, sleep, mental health and physical activity of Australian paramedics. Sci Rep 2021 Jan 13;11(1):866 [FREE Full text] [doi: 10.1038/s41598-020-79093-5] [Medline: 33441601]
8. Saleh AM, Awadalla NJ, El-masri YM, Sleem WF. Impacts of nurses' circadian rhythm sleep disorders, fatigue, and depression on medication administration errors. Egypt J Chest Dis Tuberc 2014 Jan;63(1):145-153. [doi: 10.1016/j.ejcdt.2013.10.001]
9. Blytt KM, Bjorvatn B, Moen BE, Pallesen S, Harris A, Waage S. The association between shift work disorder and turnover intention among nurses. BMC Nurs 2022 Jun 06;21(1):143 [FREE Full text] [doi: 10.1186/s12912-022-00928-9] [Medline: 35668393]
10. Wright KP, Bogan RK, Wyatt JK. Shift work and the assessment and management of shift work disorder (SWD). Sleep Med Rev 2013 Feb;17(1):41-54. [doi: 10.1016/j.smrv.2012.02.002] [Medline: 22560640]
11. Qaseem A, Kansagara D, Forciea MA, Cooke M, Denberg TD, Clinical Guidelines Committee of the American College of Physicians. Management of chronic insomnia disorder in adults: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2016 Jul 19;165(2):125-133 [FREE Full text] [doi: 10.7326/M15-2175] [Medline: 27136449]
12. Williams J, Roth A, Vatthauer K, McCrae CS. Cognitive behavioral treatment of insomnia. Chest 2013 Feb 01;143(2):554-565 [FREE Full text] [doi: 10.1378/chest.12-0731] [Medline: 23381322]
13. Järnefelt H, Lagerstedt R, Kajaste S, Sallinen M, Savolainen A, Hublin C. Cognitive behavioral therapy for shift workers with chronic insomnia. Sleep Med 2012 Dec;13(10):1238-1246. [doi: 10.1016/j.sleep.2012.10.003] [Medline: 23168269]
14. Järnefelt H, Sallinen M, Luukkonen R, Kajaste S, Savolainen A, Hublin C. Cognitive behavioral therapy for chronic insomnia in occupational health services: analyses of outcomes up to 24 months post-treatment. Behav Res Ther 2014 May;56:16-21. [doi: 10.1016/j.brat.2014.02.007] [Medline: 24632111]
15. Järnefelt H, Härmä M, Sallinen M, Virkkala J, Paajanen T, Martimo KP, et al. Cognitive behavioural therapy interventions for insomnia among shift workers: RCT in an occupational health setting. Int Arch Occup Environ Health 2020 Jul 18;93(5):535-550 [FREE Full text] [doi: 10.1007/s00420-019-01504-6] [Medline: 31853633]
16. Peter L, Reindl R, Zauter S, Hillemacher T, Richter K. Effectiveness of an online CBT-I intervention and a face-to-face treatment for shift work sleep disorder: a comparison of sleep diary data. Int J Environ Res Public Health 2019 Aug 24;16(17):3081 [FREE Full text] [doi: 10.3390/ijerph16173081] [Medline: 31450619]
17. Retzer L, Feil M, Reindl R, Richter K, Lehmann R, Stemmler M, et al. Anonymous online cognitive behavioral therapy for sleep disorders in shift workers-a study protocol for a randomized controlled trial. Trials 2021 Aug 16;22(1):539 [FREE Full text] [doi: 10.1186/s13063-021-05437-9] [Medline: 34399824]
18. Ritterband LM, Thorndike FP, Gonder-Frederick LA, Magee JC, Bailey ET, Saylor DK, et al. Efficacy of an internet-based behavioral intervention for adults with insomnia. Arch Gen Psychiatry 2009 Jul;66(7):692-698 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2009.66] [Medline: 19581560]
19. Espie CA, Kyle SD, Williams C, Ong JC, Douglas NJ, Hames P, et al. A randomized, placebo-controlled trial of online cognitive behavioral therapy for chronic insomnia disorder delivered via an automated media-rich web application. Sleep 2012 Jun 01;35(6):769-781 [FREE Full text] [doi: 10.5665/sleep.1872] [Medline: 22654196]
20. Batterham PJ, Christensen H, Mackinnon AJ, Gosling JA, Thorndike FP, Ritterband LM, et al. Trajectories of change and long-term outcomes in a randomised controlled trial of internet-based insomnia treatment to prevent depression. BJPsych Open 2017 Sep;3(5):228-235 [FREE Full text] [doi: 10.1192/bjpo.bp.117.005231] [Medline: 28959453]
21. Hagatun S, Vedaa Ø, Nordgreen T, Smith OR, Pallesen S, Havik OE, et al. The short-term efficacy of an unguided internet-based cognitive-behavioral therapy for insomnia: a randomized controlled trial with a six-month nonrandomized follow-up. Behav Sleep Med 2019;17(2):137-155. [doi: 10.1080/15402002.2017.1301941] [Medline: 28345961]

XSL•FO
RenderX

22. Cheng P, Luik AI, Fellman-Couture C, Peterson E, Joseph CL, Tallent G, et al. Efficacy of digital CBT for insomnia to reduce depression across demographic groups: a randomized trial. Psychol Med 2019 Feb 24;49(3):491-500 [FREE Full text] [doi: 10.1017/S0033291718001113] [Medline: 29792241]
23. Freeman D, Sheaves B, Goodwin GM, Yu LM, Nickless A, Harrison PJ, et al. The effects of improving sleep on mental health (OASIS): a randomised controlled trial with mediation analysis. Lancet Psychiatry 2017 Oct;4(10):749-758 [FREE Full text] [doi: 10.1016/S2215-0366(17)30328-0] [Medline: 28888927]
24. Beun RJ, Brinkman WP, Fitrianie S, Griffioen-Both F, Horsch C, Lancee J, et al. Improving adherence in automated e-coaching. In: Proceedings of the 11th International Conference on Persuasive Technology. 2016 Presented at: PERSUASIVE '16; April 5-7, 2016; Salzburg, Austria p. 276-287 URL: https://link.springer.com/chapter/10.1007/978-3-319-31510-2_24 [doi: 10.1007/978-3-319-31510-2_24]
25. Beun RJ, Fitrianie S, Griffioen-Both F, Spruit S, Horsch C, Lancee J, et al. Talk and Tools: the best of both worlds in mobile user interfaces for e-coaching. Pers Ubiquit Comput 2017 May 19;21(4):661-674. [doi: 10.1007/S00779-017-1021-5]
26. Horsch CH, Lancee J, Griffioen-Both F, Spruit S, Fitrianie S, Neerincx MA, et al. Mobile phone-delivered cognitive behavioral therapy for insomnia: a randomized waitlist controlled trial. J Med Internet Res 2017 Apr 11;19(4):e70 [FREE Full text] [doi: 10.2196/jmir.6524] [Medline: 28400355]
27. Kuhn E, Weiss BJ, Taylor KL, Hoffman JE, Ramsey KM, Manber R, et al. CBT-I coach: a description and clinician perceptions of a mobile app for cognitive behavioral therapy for insomnia. J Clin Sleep Med 2016 Apr 15;12(4):597-606 [FREE Full text] [doi: 10.5664/jcsm.5700] [Medline: 26888586]
28. Koffel E, Kuhn E, Petsoulis N, Erbes CR, Anders S, Hoffman JE, et al. A randomized controlled pilot study of CBT-I coach: feasibility, acceptability, and potential impact of a mobile phone application for patients in cognitive behavioral therapy for insomnia. Health Informatics J 2018 Mar;24(1):3-13 [FREE Full text] [doi: 10.1177/1460458216656472] [Medline: 27354394]
29. Kuhn E, Miller KE, Puran D, Wielgosz J, YorkWilliams SL, Owen JE, et al. A pilot randomized controlled trial of the insomnia coach mobile app to assess its feasibility, acceptability, and potential efficacy. Behav Ther 2022 May;53(3):440-457. [doi: 10.1016/j.beth.2021.11.003] [Medline: 35473648]
30. Daskalova N, Metaxa-Kakavouli D, Tran A, Nugent N, Boergers J, McGeary J, et al. SleepCoacher: a personalized automated self-experimentation system for sleep recommendations. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology. 2016 Presented at: UIST '16; October 16-19, 2016; Tokyo, Japan p. 347-358 URL: https://dl.acm.org/doi/10.1145/2984511.2984534 [doi: 10.1145/2984511.2984534]
31. Daskalova N, Lee B, Huang J, Ni C, Lundin J. Investigating the effectiveness of cohort-based sleep recommendations. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Sep 18;2(3):1-19. [doi: 10.1145/3264911]
32. Pandey V, Upadhyay D, Nag N, Jain R. Personalized user modelling for context-aware lifestyle recommendations to improve sleep. In: Proceedings of the 5th International Workshop on Health Recommender Systems. 2020 Presented at: HealthRecSys '20; September 26, 2020; Virtual Event p. 2020 URL: https://www.researchgate.net/publication/344345398_Personalized_User_Modelling_for_Context-Aware_Lifestyle_Recommendations_to_Improve_Sleep
33. Kay M, Choe EK, Shepherd J, Greenstein B, Watson N, Consolvo S, et al. Lullaby: a capture and access system for understanding the sleep environment. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing. 2012 Presented at: UbiComp '12; September 5-8, 2012; Pittsburgh, PA p. 226-234 URL: https://dl.acm.org/doi/10.1145/2370216.2370253 [doi: 10.1145/2370216.2370253]
34. Bauer J, Consolvo S, Greenstein B, Schooler J, Wu E, Watson N, et al. ShutEye: encouraging awareness of healthy sleep recommendations with a mobile, peripheral display. In: Proceedings of the 2012 SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI '12; May 5-10, 2012; Austin, TX p. 1401-1410 URL: https://dl.acm.org/doi/10.1145/2207676.2208600 [doi: 10.1145/2207676.2208600]
35. Daskalova N, Yoon J, Wang Y, Araujo C, Beltran G, Nugent N, et al. SleepBandits: guided flexible self-experiments for sleep. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20; April 25-30, 2020; Honolulu, HI p. 1-13 URL: https://dl.acm.org/doi/10.1145/3313831.3376584 [doi: 10.1145/3313831.3376584]
36. Ito-Masui A, Kawamoto E, Sakamoto R, Yu H, Sano A, Motomura E, et al. Internet-based individualized cognitive behavioral therapy for shift work sleep disorder empowered by well-being prediction: protocol for a pilot study. JMIR Res Protoc 2021 Mar 18;10(3):e24799 [FREE Full text] [doi: 10.2196/24799] [Medline: 33626497]
37. Ito-Masui A, Sakamoto R, Matsuo E, Kawamoto E, Motomura E, Tanii H, et al. Effect of an internet-delivered cognitive behavioral therapy-based sleep improvement app for shift workers at high risk of sleep disorder: single-arm, nonrandomized trial. J Med Internet Res 2023 Aug 22;25:e45834 [FREE Full text] [doi: 10.2196/45834] [Medline: 37606971]
38. Yu H, Itoh A, Sakamoto R, Shimaoka M, Sano A. Forecasting health and wellbeing for shift workers using job-role based deep neural network. In: Proceedings of the 9th EAI International Conference on Wireless Mobile Communication and Healthcare. 2020 Presented at: MobiHealth' 20; November 19, 2020; Virtual Event p. 89-103 URL: https://link.springer.com/chapter/10.1007/978-3-030-70569-5_6 [doi: 10.1007/978-3-030-70569-5_6]

XSL•FO
RenderX

39. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol 2000 Jun;278(6):H2039-H2049 [FREE Full text] [doi: 10.1152/ajpheart.2000.278.6.H2039] [Medline: 10843903]
40. Phillips AJ, Clerx WM, O'Brien CS, Sano A, Barger LK, Picard RW, et al. Irregular sleep/wake patterns are associated with poorer academic performance and delayed circadian and sleep/wake timing. Sci Rep 2017 Jun 12;7(1):3216 [FREE Full text] [doi: 10.1038/s41598-017-03171-4] [Medline: 28607474]
41. Lunsford-Avery JR, Engelhard MM, Navar AM, Kollins SH. Validation of the sleep regularity index in older adults and associations with cardiometabolic risk. Sci Rep 2018 Sep 21;8(1):14158 [FREE Full text] [doi: 10.1038/s41598-018-32402-5] [Medline: 30242174]
42. Murtagh F, Legendre P. Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? J Classif 2014 Oct 18;31(3):274-295. [doi: 10.1007/s00357-014-9161-z]
43. Nielsen F. Hierarchical clustering. In: Nielsen F, editor. Introduction to HPC with MPI for Data Science. Cham, Switzerland: Springer; 2016:195-211. [doi: 10.1007/978-3-319-21903-5]
44. van der Maaten L, Hinton G. Viualizing data using t-SNE. J Mach Learn Res 2008:2579-2605 [FREE Full text]
45. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: NIPS '17; December 4-9, 2017; Long Beach, CA p. 3149-3157 URL: https://dl.acm.org/doi/10.5555/3294996.3295074
46. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. 2018 Presented at: NIPS '18; December 3-8, 2018; Montréal, QC p. 6639-6649 URL: https://dl.acm.org/doi/abs/10.5555/3327757.3327770
47. Bergstra J, Bengio Y. Random Search for Hyper-Parameter Optimization. J Mach Learn Res 2012;13:281-305 [FREE Full text]
48. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. J Artif Intell 2002 Jun 01;16:321-357. [doi: 10.1613/jair.953]
49. Breiman L. Random forests. Mach Learn 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]
50. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: NIPS '17; December 4-9, 2017; Long Beach, CA p. 4768-4777 URL: https://dl.acm.org/doi/10.5555/3295222.3295230
51. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020 Jan 17;2(1):56-67 [FREE Full text] [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]
52. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020 Mar;17(3):261-272 [FREE Full text] [doi: 10.1038/s41592-019-0686-2] [Medline: 32015543]
53. Puttonen S, Härmä M, Hublin C. Shift work and cardiovascular disease - pathways from circadian stress to morbidity. Scand J Work Environ Health 2010 Mar 20;36(2):96-108 [FREE Full text] [doi: 10.5271/sjweh.2894] [Medline: 20087536]
54. Ritterband LM, Thorndike FP, Ingersoll KS, Lord HR, Gonder-Frederick L, Frederick C, et al. Effect of a web-based cognitive behavior therapy for insomnia intervention with 1-year follow-up: a randomized clinical trial. JAMA Psychiatry 2017 Jan 01;74(1):68-75. [doi: 10.1001/jamapsychiatry.2016.3249] [Medline: 27902836]

Abbreviations

AUPRC: area under the precision-recall curve
CBTI: cognitive behavioral therapy for insomnia
ICU: intensive care unit
LGBM: light gradient-boosting machine
RF: random forest
RS: recommender system
SHAP: Shapley additive explanations
SMOTE: synthetic minority oversampling technique
SRI: sleep regularity index
SWSD: shift work sleep disorder
t-SNE: t-distributed stochastic neighbor embedding

XSL•FO
RenderX

Edited by A Mavragani; submitted 02.08.24; peer-reviewed by A Kalluchi, Y Chen, D Wang; comments to author 07.10.24; revised version received 15.01.25; accepted 12.02.25; published 01.04.25.

Please cite as:
Shen Y, Choto Olivier A, Yu H, Ito-Masui A, Sakamoto R, Shimaoka M, Sano A
Personalized Physician-Assisted Sleep Advice for Shift Workers: Algorithm Development and Validation Study
JMIR Form Res 2025;9:e65000
URL: https://formative.jmir.org/2025/1/e65000
doi:10.2196/65000
PMID:

©Yufei Shen, Alicia Choto Olivier, Han Yu, Asami Ito-Masui, Ryota Sakamoto, Motomu Shimaoka, Akane Sano. Originally published in JMIR Formative Research (https://formative.jmir.org), 01.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Examining the Client Experience of Digital Tools in Blended Care Therapy: Qualitative Interview Study

Emily G Lattie[1], PhD; Miranda Beltzer[1], PhD; Alethea Varra[1], PhD; Connie E Chen[1], MD; Anita Lungu[1], PhD

Lyra Health, Burlingame, CA, United States

Corresponding Author:
Emily G Lattie, PhD
Lyra Health
270 East Lane
Burlingame, CA, 94010
United States
Phone: 1 (877) 505 7147
Email: elattie@lyrahealth.com

Abstract

Background: Lyra Health's short-term blended care therapy model, Lyra Care Therapy (LCT), has demonstrated effectiveness at scale. In LCT, clients participate in synchronous telehealth sessions and asynchronous guided practice sessions, in which they are provided with digital tools to reinforce key concepts and skills. These digital tools include animated video lessons that use storytelling to show characters learning and implementing new skills from therapy, written psychoeducational materials, interactive exercises that prompt reflection and skills practice, symptom assessments, and messaging with therapists. Past research on LCT found that time spent in therapy sessions and viewing digital video lessons predicts improvements in depression and anxiety symptoms.

Objective: This study aims to explore the client experience of LCT digital tools and to understand clients' perceived benefits and challenges of using digital tools while in LCT.

Methods: In total, 12 ethnically and racially diverse adults (5 male, 6 female, and 1 pangender) who had graduated from LCT in the previous 4 months participated in semistructured interviews. Interviews focused on experiences with the digital components of LCT (ie, video lessons, symptom assessments, and written exercises). Transcripts were analyzed using thematic analytic methods to determine the benefits and challenges associated with components of the LCT model.

Results: In total, 3 primary themes were generated through thematic analysis. These themes centered around supporting knowledge and skill development, the benefits and challenges associated with the range of digital tools, and the combination of flexibility and accountability promoting positive change. First, we summarize the ways in which guided practice sessions allow clients to develop additional knowledge and learn skills related to their mental health and well-being. Then, we describe how the range of digital tools included in LCT presents different benefits and challenges for clients. Finally, we discuss how flexibility inherent in having both synchronous and asynchronous sessions, along with the accountability from a provider, encourages clients to continue to practice skills related to their mental health and well-being.

Conclusions: Results provide insights into the unique contributions of different components of the LCT model on therapeutic gains. While perceived time constraints and content preferences can impact engagement with digital tools, overall the digital tools were perceived as carrying significant value for participants in the LCT program.

(JMIR Form Res 2025;9:e68249) doi:10.2196/68249

KEYWORDS

anxiety; depression; blended care therapy; mental health care; digital mental health; digital tools; qualitative interviews; Lyra Care Therapy; video lesson; symptom assessment; written exercise; thematic analysis; LCT model; therapeutic value; client experience

XSL•FO
RenderX

Introduction

Blended care therapy programs, which combine synchronous face-to-face therapy sessions with asynchronous therapeutic digital content, can offer the benefits of both traditional face-to-face and more novel internet-based mental health services [1-3]. Blended care therapy programs can differ in the degree to which they focus on the synchronous and asynchronous components. Some programs have focused primarily on adding face-to-face contact to digital interventions [4-6], while other programs have more deeply integrated the synchronous and asynchronous components of therapy to create new models of care delivery [7,8]. These integrated blended care therapy models typically address many of the barriers to engaging with traditionally delivered in-person care, including transportation, geography, mobility, and time constraints [9,10]. These blended care therapy models also address many of the barriers to engaging with entirely asynchronous internet-based mental health interventions, such as low perceived relevance of the intervention tools to one's own life, a lack of motivation to consistently use digital tools, and a poor user experience with digital tools [11,12].

Lyra Care Therapy (LCT) is Lyra Health's short-term blended care therapy model that is offered as an employee mental health benefit. LCT is an integrated blended care therapy model that has demonstrated effectiveness at scale for treating both depression and anxiety across diverse racial and ethnic groups [13,14]. This is a flexible model of blended care therapy and treats a wide variety of clinical presentations. In a sample of over 33,000 clients who received LCT in the United States, 86.6% of clients met criteria for clinical improvement by the end of care. Further, this recent research demonstrates that completing digital lessons and exercises and receiving provider feedback messages was associated with greater improvements in client depression and anxiety symptoms, beyond the effect of therapy sessions [14].

In LCT, clients participate in synchronous telehealth therapy sessions and in asynchronous guided practice sessions. The average length of an LCT episode is 6 synchronous therapy sessions over 8 weeks [14]. Consistent with best practices in mental health care, Lyra Care therapists are trained in and receive ongoing supervision and consultation on evidence-based mental health care [15,16]. Synchronous telehealth therapy sessions, in which therapists meet with clients on a secure video platform that is compliant with the Health Insurance Portability and Accountability Act, last for 50 minutes and typically occur weekly or biweekly. During these synchronous therapy sessions, therapists and clients establish goals for treatment, work together on developing and practicing strategies and skills, and monitor progress toward those treatment goals. Therapists also use time in these sessions to introduce the content that they are recommending the client to use during the guided practice sessions. As a flexible provider-led model of blended care, the recommended digital tools are selected in a collaborative way based on the content of the session, the client's goals, case conceptualization, and anticipated challenges during the following week; then the provider gets commitment from the client to do the practice.

Guided practice sessions take place on the digital therapy platform, which clients can access on a web browser or via the Lyra Health app. The "guided" aspect of guided practice sessions refers to therapists guiding clients to asynchronous tools relevant to their care, sending messages in between sessions, and providing asynchronous feedback on the client's progress as the client completes between-session activities. The "practice sessions" aspect refers to the asynchronous skills building and practice that happen when clients engage with these tools between sessions. This content is accessible asynchronously at any time of day or night. These digital tools were developed by Lyra Health based on transdiagnostic treatment approaches, including the Unified Protocol [17], acceptance and commitment therapy [18], and dialectical behavior therapy [19]. After each synchronous therapy session, therapists assign digital tools to their clients to reinforce key concepts and skills related to what is being worked on in treatment. These digital tools include animated video lessons that use storytelling to show characters learning and implementing new skills from therapy, written psychoeducational materials, interactive exercises that prompt reflection and skills practice, symptom assessments, and messaging with therapists. The digital tool library has grown over time, with new video lessons, exercises, and guides added on an ongoing basis. As the client engages with the digital tools, the therapist receives alerts and is able to provide asynchronous support through in-platform commenting and messaging. In a recent analysis of over 33,000 clients who completed LCT, participants viewed a median of 6.0 (IQR 3.0-9.0) digital lessons, completed 6.0 (IQR 2.0-12.0) exercises, and exchanged 16.0 (IQR 9.0-28.0) direct messages with their providers during guided practice sessions [14]. This model supports the practice of skills outside of synchronous telehealth sessions and helps support therapist fidelity to evidence-based practices.

Past research on LCT found that time spent in therapy sessions, time spent viewing digital video lessons, time spent completing exercises, and time spent messaging with the therapist all predict improvements in depression and anxiety symptoms [14,20]. These findings indicate unique statistically significant and clinically significant benefits for clients engaging in both the synchronous telehealth sessions and in the asynchronous guided practice sessions.

However, little is known about the client experience of participating in LCT. For clients to fully engage in mental health services, the services should be optimized to have a strong user experience. Qualitative research on a slightly different model of blended care therapy, which included a greater average number of telehealth sessions (an average of 21.96 sessions relative to our average of 6 sessions) and a different model of between-session practice assignments (consisting of text-based lessons, exercises, and testimonials, with small interactive elements such as audio and video clips), found that clients were generally satisfied with all components of the blended care therapy model but sometimes felt overwhelmed by the digital tools [7]. Identifying the client-perceived value of and challenges associated with specific components of blended care services is essential to understanding and designing more usable, enjoyable, and effective methods of delivering mental health

XSL•FO
RenderX

care. This study explores the client experience of digital tools while in LCT and aims to understand the perceived benefits and challenges of using digital tools while participating in LCT.

Methods

Overview

We conducted a semistructured interview study to explore the client experience of using digital tools while in LCT and to understand our client's perceived benefits and challenges for engaging with these tools.

Participants

To ensure that potential participants had familiarity with LCT digital tools and would be able to reflect on their experiences in LCT, we sent a recruitment email to Lyra clients who had graduated from LCT in the prior 4 months, were at least 18 years of age, and had completed at least 50% of digital assignments (ie, video lessons, exercises, and guides) during their course of therapy. LCT is available to individuals as an employer-sponsored mental health benefit, Lyra Health, offered by Lyra Clinical Associates. Potential participants were invited to complete a short screening survey to confirm their interest in participating and to provide demographic information. Invitations to schedule an interview were sent to participants, along with a copy of the institutional review board–approved consent document.

Ethical Considerations

All study procedures were approved by the Western-Copernicus Group Institutional Review Board (approval number: WCG IRB Protocol #20232618) prior to enrolling participants. Participants were sent a copy of the institutional review board–approved consent document prior to their scheduled interview. At the beginning of the interview appointment time, the interviewer reviewed the consent form, and participants provided verbal consent. After the interviews, participants were compensated for their time with a US $75 gift card. Interviews were recorded for transcription purposes, and the recordings were deleted after transcription was completed. All quotations presented in this manuscript have been deidentified.

Study Procedures

We interviewed a total of 12 participants. Interviews lasted approximately 60 minutes and were conducted over Zoom by the lead author. After verbal consent was obtained, the interview began and was recorded for transcription purposes. During these semistructured interviews, participants were prompted to reflect on their experiences in LCT, including the impact that the digital tools had on their lives and how different types of digital tools may have helped or hindered their progress in psychotherapy. Example questions from the interview guide include "How would you describe the experience of having digital content to work on in between sessions? What was helpful? What was unhelpful?" and "What components of the platform were most helpful to you in making progress in therapy?"

Positionality Statement

The authors all have training as clinical psychologists or physicians and have experience in the design, development, and implementation of technology-enabled mental health services. The majority of the authors have experience in delivering evidence-based psychotherapy to individuals across the lifespan and drew on these professional experiences while designing the study and interpreting the data. Throughout the analysis, the authors took their own positionality and identities into account while analyzing study data.

Analysis

Transcripts were analyzed in Dedoose [21] using a codebook thematic analytic approach [22]. The 6 steps of analysis, outlined by Braun and Clarke [23], were followed, in which the lead author familiarized themself with the data prior to generating a preliminary codebook used to organize the initial analysis of the data. Reflexive practices were used to iteratively develop themes, and the themes were reviewed, defined, and presented in this paper. Analytic decisions were documented throughout the research process. Because interrater reliability rests on the assumption that meaning is fixed within the data, this analytic approach does not recommend using formal interrater reliability metrics [22].

Results

Overview

A total of 12 individuals participated in these interviews. As seen in Table 1, participants held diverse social and cultural identities.

In total, 3 primary themes were generated through thematic analysis. These themes centered around supporting knowledge and skill development, the benefits and challenges associated with the range of digital tools, and the combination of flexibility and accountability promoting positive change. First, we summarize the ways in which guided practice sessions allowed clients to develop knowledge and learn skills related to their mental health and well-being. Then, we describe how the range of digital tools included in LCT presents different benefits and challenges for clients. Finally, we discuss how flexibility inherent in having both synchronous and asynchronous sessions, along with the accountability from a provider, encourages clients to continue to practice skills related to their mental health and well-being.

XSL•FO
RenderX

Table 1. Participant demographics.

Gender	Participants, n
Male	5
Female	6
Pangender	1
Age (years)	
18-24	1
25-34	4
35-44	2
45-54	2
≥55	2
Race or ethnicity	
Black	3
South Asian	2
East Asian	1
Hispanic or Latino	1
White	5
Education	
Vocational school	2
Some college	2
Bachelor’s degree	6
Master’s degree	2
Household annual income (in US $)	
Less than $20,000	1
$20,000-$34,000	1
$35,000-$49,000	2
$50,000-$74,000	2
$75,000-$99,000	2
More than $100,000	3

The Addition of Asynchronous Guided Practice Sessions Allows for the Development of Both Knowledge and Skills

Gaining knowledge and developing skills are central to making gains in psychotherapy. The combination of synchronous therapy sessions with a mental health provider and asynchronous guided practice sessions was viewed by many participants as essential to their progress in LCT. While many of the participants were familiar with synchronous therapy sessions, either from their own past treatment experiences or general knowledge of what “therapy” typically entails, the concept of having asynchronous guided practice sessions was new to most participants and appeared key in supporting the development of knowledge and skills. Overall, these asynchronous sessions supported clients in using the synchronous sessions more efficiently and often allowed clients and therapists to communicate more effectively about issues related to the care episode.

Participants shared ways in which the asynchronous guided practice sessions were integrated into their lives and how they helped to identify areas for growth and to provide support for both knowledge acquisition and skills practice. Pt 6 reflected:

> *I got a video to learn about the approach with cognitive behavioral therapy with my particular therapist and I was sold on that completely. The tools you're learning you then apply in your daily life. I liked the hybrid learning or the flip learning where you do work outside the classroom so when you come to therapy, you can talk about it. And for me, it's important to go back to them outside of that 45 min because if not, we would waste time during my session.*

This participant had a background in education and drew these favorable parallels to the increasingly popular flipped classroom pedagogy in which active class time is used for working through exercises and discussing concepts that were learned outside of

XSL•FO
RenderX

the classroom. By completing exercises asynchronously, the participant felt that subsequent sessions were more efficient because they could use the time to talk through their progress and any challenges that arose between sessions.

Many participants in the sample had past experiences with receiving mental health care in which knowledge and skill development felt less supported. Pt 5 reflected:

> *Previously, I had another therapist on a different platform. I had an opportunity to meet with her very regularly and while it was helpful, I did not get the same benefits that I think I did from using Lyra because my therapist with Lyra would issue me little assignments, like videos, that often spoke to situations that I was in personally.*

Among participants who reflected on their past experiences in therapy, there were particularly favorable attitudes towards the model of blended care therapy in which asynchronous guided practice sessions supported the clients in making specific treatment gains, often in a more consolidated amount of time. As Pt 12 commented:

> *Whenever I've done therapy without tools like this, it's taken me a couple of iterations to get to the point. With the tools, I get to think faster.*

This notion of "thinking faster," supported by guided practice session digital tools, appears to translate into more active and engaged practice of therapeutic skills and a quicker time to symptom relief.

Participants commented both on how the use of digital tools in guided practice sessions helped them reflect more deeply and on how therapist review of their digital tools assignments allowed for the therapist and client to communicate more effectively about presenting problems and concerns. As Pt 1 reflected:

> *So I ended up being assigned the journal… I started going in and just writing a little more and I didn't realize I hadn't truly opened up to the therapist assigned to me. When I started to journal, she learned more. She was like, "we never talked about this. You've never talked about this in sessions"*

Here, we see that the guided practice sessions allow for clients to practice skills being taught in therapy and also allow for clients to communicate with their therapists in an enhanced way. Through this enhanced communication, therapists can comment and reinforce or explore new topic areas with their clients. This leads to a more deeply shared understanding and to more personalized approaches to structuring subsequent synchronous therapy sessions.

Pt 5 went on to comment:

> *It was probably the most effective thing that I've ever had. I appreciated it a lot like, yes, my therapist was amazing. She did a really great job with helping me a lot, but having those videos along with her support and her guidance, I feel really helped me to make the progress that I have today.*

Multiple participants echoed this belief that the provider-led selection of the digital tools was critical to their treatment success and that the personalization offered the advantage of digital tools being perceived as highly relevant to the participant and their presenting problems and goals for therapy. It appears that digital tools alone are unlikely to support the reported level of knowledge and skills development. Rather, the combination of synchronous telehealth sessions and the digital tools embedded within asynchronous guided practice sessions was found to be of benefit here.

The Range of Available Digital Tools Presents Different Benefits and Challenges for Clients

Overall, digital tools appeared to support the acquisition of knowledge and support the practice of skills, while also helping to build a common language between clients and their therapist. In line with what is known about how different individuals will respond to different treatment components in different ways [24], we observed that the various digital tools embedded in LCT each presented their own benefits and challenges for clients. These benefits and challenges were discussed in the context of having the therapist select digital tools based on intended relevancy to the client. Within this theme, we identify factors that contribute to both the strengths and weaknesses of different types of digital tools within blended care therapy.

Video Lessons

Video lessons (ie, animated videos featuring stories of characters as they face mental health challenges and work with a therapist to learn skills to address those challenges) were the most popular type of digital tool with participants in this study. The majority of participants described favorable ways in which the lesson content was relatable. In addition to videos teaching therapeutic concepts and skills, they served a valuable function in normalizing the experience of a range of mental health concerns. Multiple participants reflected on how video lessons helped deepen their understanding of skills and strategies that were discussed in therapy sessions. For example, Pt 6 noted:

> *The videos allowed you to see glimpses of what was happening in [the character's] life, but then they would use strategies to work through in the moment just like what I was doing with [provider name]. It was not just teaching, but it was also practicing the strategies which was helpful because it creates the framework right of how to actually implement it, think about it and process it, which creates a roadmap.*

Storytelling, as evident in these video lessons, appears to be a vital component for supporting the learning acquisition process.

Similarly, Pt 10 highlighted:

> *I think the videos helped a lot. It puts things on a more personal and a less straight, clinical level… I remember actually watching some of the first videos that were an introduction to depression and anxiety and it walks you through the different effects that it can have on someone and the different ways it can manifest. That part really resonated with me and I really understood it. It made me feel less crazy and*

XSL•FO
RenderX

> *instead of having something wrong, I understood this disconnect between my thoughts and my emotional reactions and how it is something that I can work through. It's absolutely something that other people also deal with and that definitely helped a lot.*

The ability of lessons to normalize common mental health issues appeared to help this participant overcome self-stigma related to his symptoms and become more motivated to make changes.

Providing video lessons without support or guidance may be less useful than when integrated into a blended care therapy model. Pt 5 commented:

> *I appreciated that in tandem with me venting everything under the sun to her, she was also able to assign me things where if I brought up a very particular problem, she's like, "Oh, this sounds a lot like what you're going through. I want you to take a look at this and to make sure that you're actually paying attention.*"

This highlighted how the guidance therapists provide around video lessons supports therapy clients in learning new practices that are in line with what is being discussed in therapy sessions. The support of asynchronous learning and skill practice allowed for the client to work toward their therapeutic goals within the context of a supportive relationship with an expert clinician.

However, video lessons presented some challenges. Some participants reported that it was challenging to find time to view the videos, which are typically 8-10 minutes long. Although many participants found that the video lessons fit well into their lives, a small portion of study participants reported that it was challenging to keep up with the video lessons (and other digital tools) due to the time commitment required outside of synchronous therapy sessions. For example, Pt 9 described having some physical health issues emerge and needing to spend significant time and energy navigating the health care system. They reflected that:

> *medical stuff was just tacked on top of it, so remembering to do the check-ins and watch all the videos was derailed and I didn't always keep up with it.*

For some people, dedicating 50 minutes every week or two to synchronous therapy sessions feels like the maximum amount of time that they can dedicate to mental health care.

Participants varied in how much they related to characters presented in the video lessons. For example, a college-aged man reflected that a specific lesson featuring a middle-aged woman character, who was overwhelmed with messes that her husband and child were leaving around the home, was particularly relatable because:

> *this kind of thing is absolutely something that happens in most people's lives whether it's a mom dealing with a mess or whether it's me dealing with roommates and a mess. I definitely think it can create some trouble in thinking clearly and can create some very angry and tense emotions. Having that reminder that I can walk through it using the Lyra platform is very helpful.*

However, a middle-aged woman participant reflected on the same digital lesson as less engaging, noting:

> *I didn't relate to it because I don't have kids. There were other videos featuring kids, a husband, and work stress and my situation is kind of different because I have a fiance, we don't have kids, and my work doesn't necessarily stress me out.*

Having relatable characters featured in the video lessons appears to increase clients' ability to empathize with and learn from the lessons, and further examination of factors that increase and decrease relatability will be fruitful for the continued development of characters and storylines for video lessons.

Digital Exercises

Some participants found it challenging to write about their thoughts, feelings, and experiences in digital exercises, though with encouragement from their providers, they found these exercises helped them feel better. While a few of the participants reflected sentiments such as "writing isn't my thing," participants also described ways in which their therapists were able to tailor the written assignments to meet their needs. For example, Pt 6 reflected:

> *I get caught in needing to write it perfectly… [provider name] made it really clear that it doesn't matter and to make bullet points in a section and we can work with it. She used it as an opportunity to say, "what do you think it's about writing or putting your things down?" So it was good to be able to name it and process it. But it is also attached to some of the challenges.*

This discomfort with writing down thoughts or experiences was echoed by other clients who did not initially find written exercises to be helpful. Pt 4 continued to reflect on this point by noting:

> *When I did write the little bits that I did, I did feel better. It wasn't too pushy. It was just like you said, just a little bit of encouragement.*

For participants like these ones, the guidance and support of the therapist appeared essential for them to be able to engage with, and subsequently benefit from, the written exercises.

Symptom Assessments

While in LCT, clients are prompted to complete symptom assessments prior to each session. This practice is in line with measurement-based care principles and allows both the therapist and the client to track changes in symptoms and adjust treatment as necessary. Some participants found the weekly symptom assessments to be helpful and motivating from the beginning of therapy. For example, participants who had past experiences in therapy without routine symptom assessments, such as Pt 9, stated:

> *The most useful to me was the check-ins to see where I'm at and see where the progress was over time. I*

XSL•FO
RenderX

> *liked checking in to see where I was at and to see progress.*

However, some participants did not initially value these assessments but grew to appreciate them over time. As Pt 6 highlighted:

> *I was answering the same questions over and over again. But then, all of the sudden my responses were starting to change. It felt affirming because at first, I was just producing the same answers and then they changed! It showed progress for myself.*

While symptom assessments felt repetitive and tedious for some, the vast majority of participants appreciated the feedback that symptom assessment provided in terms of progress tracking.

Flexibility and Accountability Encourage Continued Skills Practice

Evidence-based short-term therapy models, such as the Lyra Care model, typically focus on developing and then practicing new skills or strategies to reduce suffering and improve well-being. Clients are prompted to set goals for treatment at the onset of a course of therapy, and the client and therapist work together to support progress toward those goals. As clients make progress on their goals, they work toward graduating or taking a break from therapy. This is marked by a client being ready to stop having regularly scheduled therapy sessions and to manage any remaining symptoms independently or with the support of their social network. In these short-term therapy models, therapy is conceptualized as a time to learn and apply skills and to begin generalizing those skills. When someone graduates from therapy, they are encouraged to continue doing the things that work well for them and to have a plan for how they will address new challenges moving forward. The combination of flexibility and accountability woven into this model of blended care therapy, and supported by the available digital tools, appears to promote continued practice of therapeutic skills.

Through these interviews, we observed that the digital content available through the guided practice sessions provided durable opportunities for reflection and discovery. Digital tools could prompt clients to reflect on their current experiences, identify their current emotions and thoughts, and discover both factors contributing to their current emotional state and strategies that they could use to feel better. Convenience matters when examining how a digital tool can be used while in therapy and how it might be used after graduation. Several participants reported that the written exercises were important for practicing skills and keeping the momentum going in between sessions. As Pt 10 commented:

> *I got a lot more comfortable with taking out my phone when I started having an issue or even just making a mental note of what I was going to put in a thought log or something of that nature. It's extremely helpful as an in the moment tool while also being something that is discrete enough that I could take out my phone any time of the day, anywhere I was, and just seem like I was texting where in reality I was, I was working through whatever stress or anxious thoughts I was having, so that definitely became a very helpful tool.*

Having supportive tools almost always available to use can be key for integrating and applying the skills taught in LCT to diverse situations in one's life.

While a plethora of digital mental health tools are available, many individuals report difficulty in finding high-quality digital mental health tools and are quick to disengage from using such tools [25,26]. Here, we observed that the guidance from and accountability provided by the blended care therapist helped participants learn how specific tools could be useful to and usable for them. Tools were typically introduced during synchronous telehealth sessions, and then follow-up support and recommendations were given during guided practice sessions. Pt 8 noted:

> *my therapist did a good job of marketing the tools to me of this being something you can do if you're feeling like this, and giving me examples of when even would be good times to even use it. So, when I would be experiencing a certain feeling that we discussed in therapy, I would be reminded of my therapist telling me that I should do this when I experience this, and that would lead me to go onto the platform and try an activity.*

Participants also spoke of how automated prompts from the system prompted greater accountability to engage. Pt 5 reflected:

> *During my time with my therapist, I would get, like, little email reminders or text message reminders that were prompting me to make sure that I get my assignments or my weekly health assessments done prior to my next session. I found that really helpful.*

This support received from both the therapist and the automated system prompted clients to experiment with tools during an episode of care and learn how the tools can benefit them, which can then facilitate ongoing habitual practices.

While video lessons were particularly popular for clients while they were in care with their therapist and supported the learning of new skills, written practices were often carried forward after graduating from a course of blended care therapy as a means of continued skill practice. For example, Pt 8 noted:

> *Now, I have a journal and even though it's not necessarily the Lyra exercise itself, it had components that I remember from it. When I was really anxious the other day about some random thing, I just opened up a note on my phone and started challenging my thoughts.*

This participant described one of the primary goals of short-term CBT—that they were able to better recognize their emotional experiences and quickly apply an evidence-based strategy (cognitive restructuring) to proactively manage their experience.

XSL•FO
RenderX

Discussion

Principal Findings

Results of this study support the previously observed therapeutic value of digital video lessons in LCT and provide insights into the unique contributions of different components of the LCT model on therapeutic gains. While perceived time constraints and content preferences can impact engagement with digital tools, overall the digital tools were perceived as carrying significant value while participants were clients in LCT. The combination of synchronous telehealth sessions and asynchronous guided practice sessions both helped clients learn new skills that were personalized to their current needs and helped clients apply these skills in their lives through between-session practice and feedback from therapists.

Asynchronous and synchronous components of this blended care model (ie, LCT) worked together to produce client-reported benefits. The clinically important elements of this treatment model seem to be (1) the therapist selecting relevant skills, (2) the therapist introducing skills in session, (3) the client completing between-session practice, (4) the therapist reviewing and commenting on between-sessions practices, and (5) reflection in the subsequent synchronous session. Guided practice sessions appear to enhance synchronous therapy sessions by helping facilitate ongoing communication between therapist and client and supporting skills acquisition and practice both in and out of sessions. Together, this model appears to allow clients to make clinical progress more quickly than they have seen in traditional models of psychotherapy.

Past quantitative analyses have demonstrated the value of video lesson viewing on clinical outcomes. These qualitative analyses further support the value of video lessons. Storytelling, embedded in these lessons, appears to be important in normalizing a range of experiences. Storytelling in digital mental health interventions is increasingly being recognized as clinically valuable. Narrative-based mental lessons support clients in reflecting on their own patterns and help them to feel a sense of connection to similar others as they apply the concepts to their own lives [27]. Past research has found that satisfaction with client stories embedded in internet-delivered cognitive behavioral therapy to be related to both treatment satisfaction and decreases in anxiety symptom severity [28]. Many clients in this study found the video lessons normalized experiences and that the characters and storylines were relatable, sometimes in surprising ways. However, some clients had more difficulty relating to the characters. Given that relatability appears to support engagement in the guided practice session process, this opens future directions for improving the relatability of digital lesson content. Additional research should be conducted to better understand which characteristics are most important for relatability and how relatability of characters relates to effectiveness of between-session practice. Additional training for therapists on how to frame the between-session practice may also be warranted to better support clients in developing their ability to relate to and engage with the important elements of the lesson.

At its core, LCT is an evidence-based short-term therapy model. While short-term models of therapy have strong research evidence for reducing symptoms of mental ill-health [29,30], not all individuals are initially comfortable with or confident in the approach of short-term therapy. Between-session practice helps clients build structure and self-efficacy in using skills that prepares them for a confident graduation. A large body of literature demonstrates that completing therapy homework relates to more positive treatment outcomes [31,32]. Here, we observed that the structure of skill practice outside of synchronous telehealth sessions allowed clients to figure out what skills work for them and in what situations those skills can be used. Participants in this study reflected on ways in which they figured out how to practice skills in real time, such as on their smartphones or by writing in a notebook. These skill-learning experiences are supported both with guided practice sessions and troubleshooting during synchronous sessions with the therapist. As digital tools were identified as being a helpful component of the LCT process with specific benefits of supporting work toward therapy graduation, additional work should be done to determine how to best support the transition to self-care with digital tools after graduation. Barriers to engaging in between-session practice included both time constraints and content preferences (eg, video length, preferences for reading or listening). Future work should continue to explore how clients would respond to different formats of between-session digital content.

Limitations and Additional Future Directions

The qualitative nature of this study is, in many ways, a strength. Qualitative research allows for deeper understanding of our client experience by honing in on how research participants experienced different components of LCT. However, the methods employed in this study do not support the quantification of experiences [33], and qualitative research typically does not aim to be generalizable across populations. Thus, the experiences detailed in this manuscript are unlikely to be representative of all experiences that one may have in LCT. Historically, research on technology-enabled mental health interventions has been conducted on largely homogenous samples as White individuals, and particularly White young to middle-age adult women, are typically over-represented in research trials [34,35]. While our study sample was diverse in regards to race, ethnicity, gender, age, and income, the participants may not be representative of all clients who engage with blended care services. We intentionally selected a sample of clients who had been engaged with digital content while in therapy (by having completed at least 50% of assignments)—this decision was made to maximize the likelihood that participants would be able to meaningfully reflect on the utility of the different treatment components and the full model of LCT. We cannot report on the exact digital tools each participant in the study used during their course of therapy as their engagement in therapy was separate from their engagement with this qualitative research study. Clients in LCT are not “required” to complete digital content to progress in therapy, and this sample does not represent individuals who did not deeply engage with the digital tools. Additionally, therapy is a dyadic process involving both a client and a therapist, and both sides of this

XSL•FO
RenderX

process, as well as the interactions between the client and therapist, should be examined [36]. This study is reflective only of the client experience of digital tools while in LCT and the client-perceived benefits and challenges of using digital tools while participating in LCT. Future work will examine the therapist's experience of delivering this type of blended care.

Conclusions

The between-session digital content embedded in short-term blended care therapy programs, such as LCT, can keep therapy clients engaged in skill development outside of sessions and can drive therapeutic gains. This model of care has demonstrated clinical effectiveness at scale in past quantitative studies and can support clients in developing mental health and coping skills in a relatively limited treatment period (average of 6 sessions over 8 weeks [1,14]). These qualitative results support the previously observed therapeutic value of digital video lessons in LCT and provide insights into the unique contributions of different components of the LCT model on therapeutic gains. As we continue to work toward increasing access to evidence-based mental health care, blended care therapy programs can be considered as a viable method for meeting client desires to work synchronously with a mental health provider while also offering the benefits of asynchronous care focused on skill development.

Acknowledgments

The study was funded by Lyra Health. The funder of the study was involved in study design, data interpretation, and writing of the report, and was informed during data collection and data analysis. The corresponding author had final responsibility for the decision to submit for publication. The authors declare that there was no use of generative artificial intelligence technology in the generation of texts, figures, or other informational content in this paper.

Data Availability

Given the small dataset, interview data are not publicly available to ensure participants' privacy. For questions about the dataset, contact the corresponding author.

Conflicts of Interest

EGL, MB, CEC, AV, and AL are employed by Lyra Health and Lyra Clinical Associates, receive income from Lyra Health and Lyra Clinical Associates, and have been granted equity in Lyra Health.

References

1. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO world mental health surveys. Psychol Med 2014;44(6):1303-1317 [FREE Full text] [doi: 10.1017/S0033291713001943] [Medline: 23931656]
2. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JEWC. Mixing online and face-to-face therapy: how to benefit from blended care in mental health care. JMIR Ment Health 2016;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]
3. Kemmeren L, van Schaik A, Draisma S, Kleiboer A, Riper H, Smit J. Effectiveness of blended cognitive behavioral therapy versus treatment as usual for depression in routine specialized mental healthcare: E-COMPARED trial in the Netherlands. Cogn Ther Res 2023 Mar 24;47(3):386-398. [doi: 10.1007/s10608-023-10363-y]
4. Kay-Lambkin F, Baker A, Lewin T, Carr V. Acceptability of a clinician-assisted computerized psychological intervention for comorbid mental health and substance use problems: treatment adherence data from a randomized controlled trial. J Med Internet Res 2011;13(1):e11 [FREE Full text] [doi: 10.2196/jmir.1522] [Medline: 21273184]
5. Høifødt RS, Lillevoll KR, Griffiths KM, Wilsgaard T, Eisemann M, Waterloo K, et al. The clinical effectiveness of web-based cognitive behavioral therapy with face-to-face therapist support for depressed primary care patients: randomized controlled trial. J Med Internet Res 2013;15(8):e153 [FREE Full text] [doi: 10.2196/jmir.2714] [Medline: 23916965]
6. Pier C, Austin DW, Klein B, Mitchell J, Schattner P, Ciechomski L, et al. A controlled trial of internet-based cognitive-behavioural therapy for panic disorder with face-to-face support from a general practitioner or email support from a psychologist. Ment Health Fam Med 2008;5(1):29-39 [FREE Full text] [Medline: 22477844]
7. Etzelmueller A, Radkovsky A, Hannig W, Berking M, Ebert DD. Patient's experience with blended video- and internet based cognitive behavioural therapy service in routine care. Internet Interv 2018;12:165-175 [FREE Full text] [doi: 10.1016/j.invent.2018.01.003] [Medline: 30135780]
8. Lungu A, Jun JJ, Azarmanesh O, Leykin Y, Chen CE. Blended care-cognitive behavioral therapy for depression and anxiety in real-world settings: pragmatic retrospective study. J Med Internet Res 2020;22(7):e18723 [FREE Full text] [doi: 10.2196/18723] [Medline: 32628120]
9. Andrade LH, Alonso JE, Mneimneh Z, Wells J, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol Med 2014;44(6):1303-1317 [FREE Full text] [doi: 10.1017/S0033291713001943] [Medline: 23931656]

XSL•FO
RenderX

10. Mohr DC, Hart SL, Howard I, Julian L, Vella L, Catledge C, et al. Barriers to psychotherapy among depressed and nondepressed primary care patients. Ann Behav Med 2006;32(3):254-258. [doi: 10.1207/s15324796abm3203_12] [Medline: 17107299]
11. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
12. Moskalenko MY, Hadjistavropoulos HD, Katapally TR. Barriers to patient interest in internet-based cognitive behavioral therapy: informing e-health policies through quantitative analysis. Health Policy Technol 2020;9(2):139-145. [doi: 10.1016/j.hlpt.2020.04.004]
13. Owusu JT, Wang P, Wickham R, Varra AA, Chen C, Lungu A. Real-world evaluation of a large-scale blended care-cognitive behavioral therapy program for symptoms of anxiety and depression. Telemed J E Health 2022;28(10):1412-1420 [FREE Full text] [doi: 10.1089/tmj.2021.0590] [Medline: 35263185]
14. Espel-Huynh HM, Wang L, Lattie EG, Wickham RE, Varra A, Chen CE, et al. Clinical effects of asynchronous provider-guided practice sessions during blended care therapy for anxiety and depression: pragmatic retrospective cohort study. J Med Internet Res 2024;26:e60502 [FREE Full text] [doi: 10.2196/60502] [Medline: 39422996]
15. Borders LD. Best practices in clinical supervision: another step in delineating effective supervision practice. Am J Psychother 2014;68(2):151-162. [doi: 10.1176/appi.psychotherapy.2014.68.2.151] [Medline: 25122982]
16. Watkins CE. Psychotherapy supervision developments and innovations for the new millennium: contributions from the cutting edge. Am J Psychother 2014;68(2):141-149. [doi: 10.1176/appi.psychotherapy.2014.68.2.141] [Medline: 25122981]
17. Barlow D, Farchione T, Sauer-Zavala S, Murray-Latin H, Bullis J, Ellard K. Unified Protocol for Transdiagnostic Treatment of Emotional Disorders: Therapist Guide. New York, NY: Oxford University Press; 2017.
18. Hayes S, Strosahl K, Wilson K. Acceptance and Commitment Therapy: The Process and Practice of Mindful Change. New York, NY: Guilford Press; 2011.
19. Linehan MM, Korslund KE, Harned MS, Gallop RJ, Lungu A, Neacsiu AD, et al. Dialectical behavior therapy for high suicide risk in individuals with borderline personality disorder: a randomized clinical trial and component analysis. JAMA Psychiatry 2015;72(5):475-482. [doi: 10.1001/jamapsychiatry.2014.3039] [Medline: 25806661]
20. Lungu A, Wickham RE, Chen S, Jun JJ, Leykin Y, Chen CEJ. Component analysis of a synchronous and asynchronous blended care CBT intervention for symptoms of depression and anxiety: pragmatic retrospective study. Internet Interv 2022;28:100536 [FREE Full text] [doi: 10.1016/j.invent.2022.100536] [Medline: 35433277]
21. Dedoose. 2020. URL: https://www.dedoose.com/ [accessed 2025-03-08]
22. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019;11(1):1-9. [doi: 10.1080/2159676x.2019.1628806]
23. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
24. Lutz W, Gómez PJ, Prinz J, Rubel J, Moggia D. Do certain patients respond better to specific forms of psychotherapy? In: APA Handbook of Psychotherapy: Evidence-Based Practice, Practice-Based Evidence, and Contextual Participant-Driven Practice. Washington, DC: American Psychological Association; 2024:63-81.
25. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
26. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
27. Bhattacharjee A, Williams JJ, Chou K, Tomlinson J, Meyerhoff J, Mariakakis A, et al. "I Kind of Bounce off It": translating mental health principles into real life through story-based text messages. Proc ACM Hum Comput Interact 2022;6(CSCW2):398 [FREE Full text] [doi: 10.1145/3555123] [Medline: 36816014]
28. Hadjistavropoulos HD, Hill TG, Philipp-Muller AE, Dear B, Titov N. Perceptions of client stories in internet-delivered cognitive behaviour therapy: a mixed-methods evaluation. Internet Interv 2023;34:100692. [doi: 10.1016/j.invent.2023.100692]
29. Robinson L, Kellett S, Delgadillo J. Dose-response patterns in low and high intensity cognitive behavioral therapy for common mental health problems. Depress Anxiety 2020;37(3):285-294. [doi: 10.1002/da.22999] [Medline: 32027435]
30. Driessen E, Hegelmaier LM, Abbass AA, Barber JP, Dekker JJ, Van HL, et al. The efficacy of short-term psychodynamic psychotherapy for depression: a meta-analysis update. Clin Psychol Rev 2015;42:1-15. [doi: 10.1016/j.cpr.2015.07.004] [Medline: 26281018]
31. Lebeau RT, Davies CD, Culver NC, Craske MG. Homework compliance counts in cognitive-behavioral therapy. Cogn Behav Ther 2013;42(3):171-179. [doi: 10.1080/16506073.2013.763286] [Medline: 23419077]
32. Kazantzis N, Whittington C, Zelencich L, Kyrios M, Norton PJ, Hofmann SG. Quantity and quality of homework compliance: a meta-analysis of relations with outcome in cognitive behavior therapy. Behav Ther 2016;47(5):755-772. [doi: 10.1016/j.beth.2016.05.002] [Medline: 27816086]

XSL•FO
RenderX

33. Sale JEM, Lohfeld LH, Brazil K. Revisiting the quantitative-qualitative debate: implications for mixed-methods research. Qual Quant 2002;36(1):43-53 [FREE Full text] [doi: 10.1023/A:1014301607592] [Medline: 26523073]
34. Titov N, Dear B, Nielssen O, Staples L, Hadjistavropoulos H, Nugent M, et al. ICBT in routine care: a descriptive analysis of successful clinics in five countries. Internet Interv 2018;13:108-115 [FREE Full text] [doi: 10.1016/j.invent.2018.07.006] [Medline: 30206525]
35. Alegría M, Alvarez K, Ishikawa RZ, DiMarzio K, McPeck S. Removing obstacles to eliminating racial and ethnic disparities in behavioral health care. Health Aff (Millwood) 2016;35(6):991-999 [FREE Full text] [doi: 10.1377/hlthaff.2016.0029] [Medline: 27269014]
36. Atzil-Slonim D, Tschacher W. Dynamic dyadic processes in psychotherapy: introduction to a special section. Psychother Res 2020;30(5):555-557. [doi: 10.1080/10503307.2019.1662509] [Medline: 31476979]

Abbreviations

LCT: Lyra Care Therapy

Edited by A Mavragani; submitted 04.11.24; peer-reviewed by JC Perez; comments to author 06.12.24; revised version received 14.01.25; accepted 04.03.25; published 02.04.25.

Please cite as:
Lattie EG, Beltzer M, Varra A, Chen CE, Lungu A
Examining the Client Experience of Digital Tools in Blended Care Therapy: Qualitative Interview Study
JMIR Form Res 2025;9:e68249
URL: https://formative.jmir.org/2025/1/e68249
doi:10.2196/68249
PMID:

©Emily G Lattie, Miranda Beltzer, Alethea Varra, Connie E Chen, Anita Lungu. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Remote Intervention Based on mHealth and Community Health Workers for Antiretroviral Therapy Adherence in People With HIV: Pilot Randomized Controlled Trial

Shivesh Shourya[1], BS; Jianfang Liu[2], BE, MAS, MPA, PhD; Sophia McInerney[2], BS; Trinity Casimir[2], BA; James Kenniff[2], BS; Trace Kershaw[1], PhD; David Batey[3], MSW, PhD; Rebecca Schnall[2], MPH, PhD

[1]Department of Social and Behavioral Sciences, Yale School of Public Health, New Haven, CT, United States

[2]Columbia University School of Nursing, New York, NY, United States

[3]Tulane University School of Social Work, New Orleans, LA, United States

Corresponding Author:
Shivesh Shourya, BS
Department of Social and Behavioral Sciences
Yale School of Public Health
60 College Street
New Haven, CT, 06510
United States
Phone: 1 6503346176
Email: shivesh.shourya@yale.edu

Abstract

Background: Despite the availability of antiretroviral therapy (ART), only 66% of people with HIV in the United States achieve viral suppression, largely due to suboptimal ART adherence. Barriers such as limited access to care and forgetfulness impact adherence rates, which must be maintained at ≥95% to prevent viral load rebound. Combination interventions leveraging community health worker (CHW) support and mobile health (mHealth) technologies have the potential to overcome previously identified barriers and provide cost-effective support for improving adherence and viral suppression outcomes in people with HIV.

Objective: This pilot study aimed to evaluate the feasibility, acceptability, and preliminary efficacy of remote delivery of the Community Health Worker and mHealth to Improve Viral Suppression (CHAMPS) intervention, combining the WiseApp, CHW support, and the CleverCap smart pill bottle. A secondary aim was to gather participants' feedback on the usability of the app and pill bottle as well as to better understand their experiences with remote study procedures.

Methods: This mixed methods pilot study involved 40 participants with HIV, who were randomly assigned to a control group (n=20, 50%) or the CHAMPS intervention (n=20, 50%) over 3 months. The intervention group participated in up to 12 sessions with CHWs and used the WiseApp, paired with a CleverCap smart pill bottle, to support ART adherence. Remote baseline and follow-up visits were conducted via Zoom and included surveys measuring adherence, self-efficacy, and usability (measured by Health Information Technology Usability Evaluation Scale [Health-ITUES] and Poststudy System Usability Questionnaire [PSSUQ]). Semistructured interviews explored participants' experiences with the intervention. Thematic analysis was used to identify key facilitators and barriers based on the Mobile Health Technology Acceptance Model.

Results: Remote delivery of the CHAMPS intervention was feasible, with high usability ratings for both the WiseApp and CleverCap (overall scores on Health-ITUES: mean 4.35, SD 0.58 and PSSUQ: mean 2.04, SD 1.03). In the intervention group, there were nonsignificant improvements in self-reported adherence scores (P=.29) and in self-efficacy scores (P=.07). The adjusted odds ratio for achieving undetectable viral load in the intervention group compared to the control group was 3.01 (95% CI –1.59 to 4.12), indicating a medium effect size in favor of the intervention. Overall study retention was 75% (30/40), with higher retention in the control group. Participants valued the flexibility of remote study procedures, particularly Zoom-based study visits and mailed blood sample kits. Qualitative feedback highlighted the intervention's acceptability and ability to overcome logistical barriers.

Conclusions: The remote CHAMPS pilot study demonstrated the feasibility and acceptability of combining mHealth tools with CHW support to promote medication adherence among people with HIV. While further optimization is needed to enhance its impact, this intervention shows potential for improving health outcomes in diverse underserved populations.

XSL•FO
RenderX

Trial Registration: ClinicalTrials.gov NCT05938413; https://clinicaltrials.gov/study/NCT05938413

(JMIR Form Res 2025;9:e67997) doi:10.2196/67997

KEYWORDS

HIV; antiretroviral therapy adherence; ART; ART adherence; mobile health technology; mHealth technology; community health workers; CHWs; smart pill bottle; feasibility evaluation; mobile health technology acceptance model

Introduction

Background

With the advent of antiretroviral therapy (ART) for HIV, people with HIV can now achieve near-normal life expectancies [1-4]. At the end of 2021, >50% of people with HIV in the United States were aged ≥50 years, primarily due to the effectiveness of ART [5]. However, the rates of viral suppression among the 1.2 million people with HIV in the United States continue to remain low, with current estimates of approximately 66% among people with HIV, despite the wider availability of ART [5]. The suboptimal rates of viral suppression may be attributed to poor ART adherence [6] and occur during a period of increasing health care fragmentation, leading to higher costs of care and lower health outcomes for many patients with chronic illnesses, such as people with HIV [7].

The Ending the HIV Epidemic plan identifies viral suppression as a cornerstone to preventing and treating HIV infections [8-10]. HIV viral suppression is dependent on adherence to ART, and studies demonstrate that high rates of ART adherence (approximately 95%) are attributed to higher rates of viral suppression (approximately 78%) [11]. However, even moderate decreases in adherence to ART (from 95% to 80%) can lead to drastic reductions in viral suppression rates (from 78% to 20%), underscoring that ART adherence is maintained at ≥95% [11]. Poor adherence to ART can lead to poor outcomes for the patient, the emergence of drug-resistant HIV strains [12], and increased rates of HIV transmission [13]. Suboptimal levels of ART adherence, along with low levels of engagement in the HIV care continuum, have also been attributed to the progression of HIV disease and premature deaths among people with HIV [14]. Therefore, there is an urgent need to develop and evaluate interventions that enhance adherence to ART.

Developing interventions to increase ART adherence must address prevalent barriers to optimal ART adherence by championing a comprehensive approach to understanding and addressing the unmet needs of people with HIV. Previous studies have identified numerous barriers to sustained adherence to ART; engagement with medical care; and, consequently, viral suppression, including HIV stigma or fear of HIV status disclosure [15-19], geographic barriers or limited time resources for receiving or accessing HIV care [16,17,19], negative experiences within medical institutions or HIV care centers [15,16,19], and forgetfulness [18,19]. Similar studies have outlined strong networks of social support, HIV case managers, telemedicine appointments, and reminder tools as facilitators of ART adherence and medical care engagement [16-19]. Recently, there has been an increased focus on the utility of community health workers (CHWs), who can take on roles as outreach workers, patient navigators, health advisers, or peer leaders in helping patients access and manage HIV in primary care settings [20]. While the profile, skill level, and job scope of CHWs can vary widely based on context, they often serve as patient navigators and peer educators in the United States, focusing on HIV care retention, ART adherence, and viral suppression support, particularly in underresourced communities [21-23]. Several studies have found that integrating CHWs was associated with improved care retention, ART adherence, and viral suppression outcomes [20,24,25]; nonetheless, other studies have found no significant difference in viral suppression outcomes between people with HIV with assigned CHWs and those without CHW interventionists [26,27]. While some studies failed to show statistical significance in viral suppression outcomes between intervention and control groups, they emphasized the utility of a combined intervention, including CHW components and other interventions, to address barriers to ART adherence [26,27]. Mobile health (mHealth) interventions offer themselves as novel interventions in HIV care settings, given their ubiquity, ability to overcome geographic barriers to care, and low associated costs [28]. However, previous studies have provided mixed results, with some finding increased viral suppression and retention in care rates among mHealth intervention participants [29,30] and others finding no statistically significant difference in viral suppression and retention rates between intervention and control participants [31]. Previous study limitations include poor integration of mHealth technologies within HIV primary care settings, and researchers have advocated for a combination approach with patient navigation services to help prioritize human communications and achieve desired outcomes related to viral suppression [31]. Thus, interventions that combine the use of CHWs and mHealth technologies hold promise for increasing rates of ART adherence and viral suppression within the United States [24,32].

Objectives

In this pilot study of the Community Health Worker and Mobile Health to Improve Viral Suppression (CHAMPS) remote intervention, conducted by our team at the Columbia University School of Nursing, we tested the feasibility of a remotely delivered CHW intervention through the mHealth app WiseApp. Participants were recruited from across the United States and provided with a smart pill bottle, CleverCap, which interfaced with the WiseApp for personalized medication reminders [33,34]. We also gathered feedback on the acceptability of remotely conducted study procedures.

XSL•FO
RenderX

Methods

Study Design

Overview

The pilot study was a 2-arm randomized controlled trial among 40 people with HIV who were followed over a 3-month period. Participants were randomly assigned to the CHAMPS intervention (20/40, 50%) or the standard of care control (20/40, 50%) arm. Recruitment for the study occurred over a 7-month period, beginning in July 2023 and closing at the end of January 2024. The study was initiated on July 26, 2023, with the enrollment of the first participant, and concluded in May 2024, with the final participant completing their last study visit on May 7, 2024.

Intervention Arm and Description of the Planned Intervention

The CHAMPS intervention was a 3-month intervention guided by the study team's previous work on CHW and mHealth interventions. Specifically, the Birmingham Access to Care study (NCT03205982) adapted the Anti-Retroviral Treatment and Access to Services intervention, an evidence-based intervention designated by the Centers for Disease Control and Prevention, to support re-engagement in care for people with HIV who had dropped out of care [34,35]. This approach emphasized strength-based case management and motivational interviewing, fostering a close, supportive relationship between CHWs and participants, and was used to design the content structure of the CHW sessions. In addition, the WiseApp study leveraged end user feedback to develop a self-management app for people with HIV, incorporating features such as medication trackers, push notification reminders, and linkage to the CleverCap smart pill bottle [34,36].

Participants in the intervention arm were assigned a CHW at the end of their baseline visit. CHWs were study team members trained on the intervention, including the content of each session, motivational interviewing, strength-based case management, Anti-Retroviral Treatment and Access to Services [37], HIV and substance use, the WiseApp and associated mHealth technology, and field safety. CHWs administered at least 10 or up to 12 individual sessions to the participants throughout the intervention. Table 1 provides an outline and description of CHW session contents and structure. All sessions were conducted remotely via secure Zoom (Zoom Video Communications, Inc) calls, in compliance with the Health Insurance Portability and Accountability Act (HIPAA); phone calls; or the chat feature of the WiseApp, depending on the participant's preference.

XSL•FO
RenderX

Table 1. Structure and content of the community health worker (CHW) sessions delivered to participants in the Community Health Worker and Mobile Health to Improve Viral Suppression (CHAMPS) intervention arm.

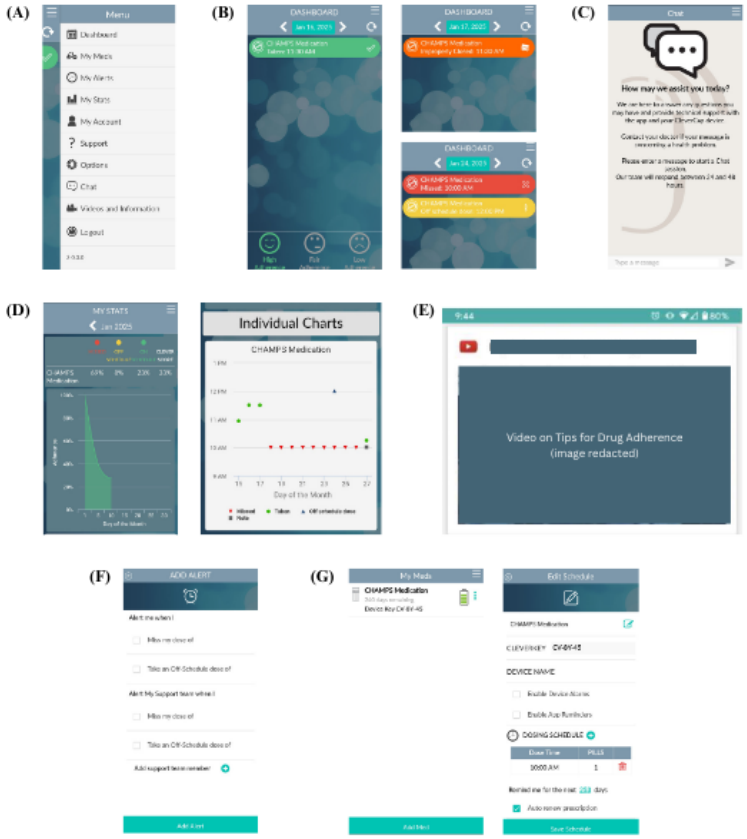
Session number and title	Session content	Window for session
(1) Building the relationship	Introduce the goals of the CHAMPS intervention to the participant (approximately 15 to 20 minutes).	At baseline visit or within 1 to 2 days of baseline visit
(2) Introduction to the WiseApp	Discuss how the WiseApp can be used to facilitate communication between the CHW and the participant. Review the medication tracking function and how this can be used by the CHW and the participant (approximately 30 to 45 minutes).	At baseline visit or within 1 to 2 days of baseline visit
(3) Emphasizing personal strengths	Help the participant self-identify personal strengths, abilities, and skills. Check in on any technical issues and study logistics with the participant (approximately 30 to 45 minutes).	Week 1
(4) Learning to make contact	Assist the participant in preparing a list of questions to ask their care provider. Check in on any technical issues and study logistics with the participant (approximately 30 to 45 minutes).	Week 2
(5) Check-in call 1	Check in with the participant on their upcoming appointments and any needed documents; address any potential barriers to care through offering local resources. Check in on any technical issues and study logistics with the participant (approximately 15 to 20 minutes).	Week 3
(6) Primary care provider appointment 1	Support the participant's efforts during their primary care provider remote visit, on the basis of the participant's comfort level (approximately 15 to 20 minutes).	On the basis of any upcoming appointments
(7) Debriefing the provider visit 1	Solicit the participant's input on what went well during their primary care provider visit. Elicit from the participant what was learned from the care visit and what strengths they demonstrated. Check in on any technical issues and study logistics with the participant (approximately 15 to 20 minutes).	At the time of or within 1 to 2 days of session 6
(8) Reviewing progress	Plan for and create an action plan for the transition process as the participant nears the end of the study period and prepares to engage with standard-of-care case manager and related clinical professionals (eg, social worker). Check in with the participant on their upcoming appointments and any needed documents. Check in on any technical issues and study logistics with the participant (approximately 30 to 45 minutes).	Anytime during weeks 4 to 7
(9) Check in call 2	Check in with the participant on their upcoming appointments and any needed documents; address any potential barriers to care through offering local resources. Check in on any technical issues and study logistics with the participant (approximately 15 to 20 minutes).	Anytime during weeks 8 to 10
(10) Completing the work	Review the transition process action plan with the participant. Begin transition to standard-of-care case manager and related clinical professionals (eg, social worker) (approximately 30 to 45 minutes).	Anytime during weeks 11 to 12
(11) Optional check in session 1	Check in on any technical issues and study logistics with the participant (approximately 15 to 20 minutes).	Anytime during the study and as needed by the participant
(12) Optional check in session 2	Check in on any technical issues and study logistics with the participant (approximately 15 to 20 minutes).	Anytime during the study and as needed by the participant

Intervention participants received 1 daily reminder through CleverCap's programmed alarm to take their medication at their chosen time. Additional alarms could be programmed by participants, or mobile or app alerts could be set for missed or off-schedule doses, in which case, participants received >1 daily notification. Furthermore, participants were notified if they reopened the pill bottle after an alarm had already been triggered (flagged as an off-schedule dose) or if the cap was improperly closed. However, no reminders were sent, and no alarms were triggered if the CleverCap pill bottle lost power, was deactivated by the study team, was damaged, or lacked a set dosing schedule.

Figure 1 illustrates the WiseApp user interface, with screenshots of a demonstration account. The main menu (Figure 1A) allows navigation to various tabs. The dashboard (Figure 1B) provides an overview of medication adherence, including statuses such as taken (green), missed (red), off-schedule dose (yellow), and improperly closed cap (orange). It also displays adherence feedback through emoji indicators (eg, high adherence: green smiley face). The chat interface (Figure 1C) enables communication with CHWs, while the My Stats tab (Figure 1D) presents a percentage breakdown of adherence metrics, dose timing, and adherence scores. The Videos and Information tab (Figure 1E) includes testimonial videos with adherence tips. The My Alerts tab (Figure 1F) allows participants and CHWs to set reminders for missed or off-schedule doses. Finally, the My Meds tab (Figure 1G) enables participants and CHWs to manage medication details and adjust dosing schedules.

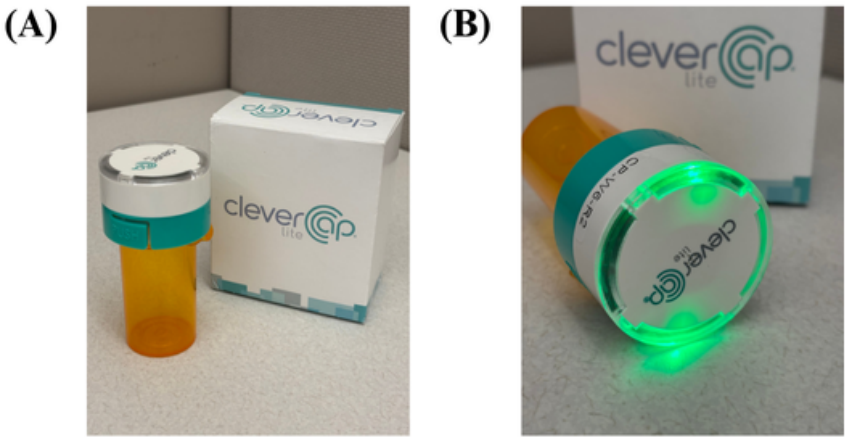
XSL•FO
RenderX

Figure 1. User interface of WiseApp showcasing demo account data: (A) main menu, (B) dashboard with medication adherence status, (C) chat interface, (D) My Stats tab, (E) Videos and Information tab, (F) My Alerts tab, and (G) My Meds tab and Edit Schedule interface.



The CleverCap pill bottle (Figure 2) tracked medication adherence by recording when the bottle was opened, with a dose marked as "taken" only if the cap remained off for at least 5 seconds. Figure 2A shows the CleverCap and its packaging, which includes a micro-USB charger. Figure 2B demonstrates CleverCap's built-in reminder system, which provided a visual and auditory alarm when it was time to take a scheduled dose. The bottle flashed green lights for 2 minutes and emitted a loud alarm to ensure that the reminder was noticeable. These cues automatically ceased once the cap was unscrewed, signaling a recorded dose. In addition, if no action was taken within the 2-minute reminder window, the visual and auditory cues stopped, and the dose was recorded as "missed." No visual or auditory cues were provided for off-schedule doses. If the alarm failed to activate due to incorrect or missing programmed alerts on the WiseApp or due to the pill bottle losing power, participants may have missed their medication without documentation in the app.

Figure 2. The CleverCap pill bottle used for medication adherence monitoring: (A) device and packaging and (B) active CleverCap alarm for dosing reminder.



XSL•FO
RenderX

Standard of Care Control Arm

The control arm did not receive any additional interventions beyond the usual standard of care, which consisted of routine clinical services as needed, including referrals to mental health or other ancillary services if clinically indicated.

Recruitment and Eligibility

Participants were eligible to participate if they (1) were able to speak, read, and write in English; (2) were aged ≥18 years; (3) were willing to provide a valid form of identification for verification; (4) were willing to participate in any assigned arm of the intervention; (5) were diagnosed with HIV ≥6 months ago; (6) had an HIV-1 RNA level of >50 copies/ml or self-reported being virally unsuppressed in the past 12 months (defined as a viral load of ≥200 copies/ml in alignment with the clinical definition of viral suppression in the United States); (7) were owners of a smartphone; (8) were capable and willing to provide informed consent for study participation and consent for access to medical records; and (9) lived in the United States. The use of different viral load thresholds aligns with established clinical guidelines and reflects advances in monitoring technology. A threshold of ≥200 copies/ml is used to determine whether someone is detectable for HIV, as it accounts for small fluctuations in viral levels and testing variations, providing a practical measure of treatment effectiveness [38]. However, with improvements in viral load assay sensitivity, a threshold of 50 copies/ml is increasingly used to identify low levels of HIV in the blood, as achieving a viral load below this threshold is considered optimal for reducing transmission risk and improving long-term health outcomes [39]. Viral load data were obtained through one of the following methods: (1) participant-uploaded viral load test results from their electronic health record or laboratory reports, (2) dried blood spot (DBS) collection analyzed by a certified laboratory, or (3) retrieval through electronic health record access via a HIPAA electronic release of information (eROI) authorization). To ensure consistency in timing, only uploaded viral load results dated within 4 to 6 weeks before the study visit were accepted. DBS collection was offered to all participants as an alternative to laboratory-based testing. Participants were not eligible if they met any exclusion criteria, including (1) residing in a nursing home, prison, or receiving in-patient psychiatric care at the time of enrollment; (2) terminal illness with a life expectancy <3 months; (3) planning to move out of the United States in the next 3 months; or (4) participating in a study that targets viral suppression for people with HIV. The study team carefully considered including newly diagnosed participants (diagnosed <6 months) and chose not to include this subset of people with HIV because newly diagnosed individuals are often treatment naive; thus, their adherence behaviors are unknown and may change frequently as they begin their ART regimen.

Participants were recruited via online advertisements posted on POZ.com, Craigslist, as well as Facebook and Instagram (Meta Platforms, Inc). In addition, the study team emailed community-based organizations serving people with HIV across the country with requests to hang study flyers. To determine preliminary eligibility, participants either filled out a web-based screener or were screened by phone. Participants who filled out web-based screeners were followed up with phone calls and were given detailed instructions regarding the study before scheduling study visits. To finalize eligibility, participants were requested to provide confirmation documentation of viral load either via filling out an eROI, by uploading results obtained within the past 4 to 6 weeks in their electronic health record to a secure study platform, or by participating in a screening visit where trained study staff instructed participants on how to self-collect blood samples via a finger-prick method.

Study Visits and Visit Structure

Overview

The study consisted of 2 required visits: a baseline visit and a 3-month follow-up visit. The study also consisted of an optional screening visit for participants who could not provide confirmation of viral load for eligibility. All study visits were conducted using secure HIPAA-compliant Zoom technology. At the start of each study visit, participants were required to show a valid form of identification for verification and to avert fraud.

Screening Visit

Screening visits lasted approximately 1 hour. Participants provided informed consent electronically through the REDCap (Research Electronic Data Capture; Vanderbilt University) system, a secure platform for data collection [40,41]. Before consenting, participants had the opportunity to ask study staff any questions about the process. Before the screening visit, participants were sent a DBS collection kit along with detailed instructions. Trained study staff guided participants through the DBS collection process, which included the use of a Becton Dickinson Microtainer contact-activated lancet blade (1.5 mm × 2.0 mm; blue, high flow) to collect the sample. Care instructions for the puncture site were provided to ensure proper aftercare. Participants were instructed on how to package and ship their samples back to the laboratory at the Miriam Hospital in Rhode Island for processing. Upon receipt of laboratory results, study staff followed up with the participants to inform them of their eligibility for the study. Compensation was not provided for screening visits.

Baseline Visit

Baseline visits lasted 2 hours. Participants provided e-consent through REDCap, with the opportunity to ask study staff any questions before participating in the baseline visit. Upon enrollment, participants were randomized to either the intervention or control arm through a randomization module in REDCap. The randomization module was generated by the study team's data manager, ensuring an unbiased allocation process, while study staff, who were separate from the data manager and did not have access to the randomization module, were responsible for assigning participants to their respective study groups. Intervention participants were informed that they would be sent a CleverCap bottle and would have up to 12 study sessions over the duration of the study with an assigned CHW. Participants were required to complete a comprehensive survey via Qualtrics (Qualtrics International Inc). The survey gathered information on demographics, a single-item ART adherence self-reported scale item (SRSI) [42], and self-efficacy as a

XSL•FO
RenderX

mediator of HIV treatment adherence (HIV Treatment Adherence Self-Efficacy Scale [HIV-ASES]) [43]. Baseline viral load and CD4 counts were recorded from information obtained via eROI, participant self-uploaded laboratory results, or the screening visit. Participants were compensated US $40 in the form of an Amazon gift code for their time.

3-Month Follow-Up Visit

Follow-up visits lasted 1 hour. Participants completed a similar survey as the one administered at baseline with the addition of measures for usability displayed to intervention participants, including the Health Information Technology Usability Evaluation Scale (Health-ITUES) [44] and the Poststudy System Usability Questionnaire (PSSUQ) [45]. Participants also provided viral load and CD4 counts obtained via eROI, participant self-uploaded laboratory results, or DBS collection. Real-time adherence data from the CleverCap pill bottles were downloaded from the CleverCap website. At the end of the follow-up visit, all participants were given the option to participate in an optional qualitative interview regarding their experience in the study and the perceived utility of the remotely delivered CHAMPS intervention. Follow-up interviews were conducted exclusively using secure HIPAA-compliant Zoom technology and were audio recorded for transcription. Consent for follow-up interviews was included in the study consent offered to participants at the baseline visit. Participants were compensated US $50 for their time in the follow-up visit and an additional US $35 if they opted to participate in the interview. Compensation was provided in the form of Amazon gift codes.

Quantitative Data Analysis

Descriptive statistics for demographic variables were reported as counts and percentages. Statistical significance for differences between the control and intervention groups was assessed using the Mann-Whitney *U* test for continuous variables, such as age, and the chi-square test for the remaining categorical variables. A 2-sample *z* test for equality of proportions, with a continuity correction, was used to assess whether the differences in retention rates between the intervention and control groups were statistically significant.

The primary outcome, viral load (virally unsuppressed vs suppressed), was analyzed using a 2-way frequency table by time. The adjusted odds ratio (aOR), adjusting for baseline viral load status, was calculated for the CHAMPS group (intervention vs control).

Self-rated adherence scores using the SRSI [42], a 6-point Likert scale ranging from 1 (“very poor”) to 6 (“excellent”), were reported as mean and SD. Independent 2-tailed *t* tests were used to compare self-rated adherence scores between the intervention and control groups, while paired 2-tailed *t* tests were used to compare scores within groups over time. Missing data for the paired 2-tailed *t* test were handled under the assumption that they was missing completely at random, as participants were lost to follow-up due to external factors unrelated to study data collection.

For adherence data collected through CleverCap, the average percentage of doses taken and missed was calculated, with SD reported for the participants in the analyzed intervention group. The correlation between follow-up SRSI scores and the average percentage of doses taken was calculated using Pearson correlation.

Usability measures, including the PSSUQ [45] and the Health-ITUES [44], were analyzed by calculating the mean and SD for the overall scales and their subscales. The PSSUQ (18 items and 3 subscales) was measured using a 7-point Likert scale, ranging from 1 (“strongly agree”) to 7 (“strongly disagree”). The Health-ITUES (20 items and 4 subscales) was measured using a 5-point Likert scale, ranging from 1 (“strongly disagree”) to 5 (“strongly agree”).

Self-efficacy for HIV treatment adherence was measured using the HIV-ASES (12 items and 2 subscales) [43], a 10-point Likert scale ranging from 1 (“cannot do at all”) to 10 (“certain can do it”). Mean and SD were computed for self-efficacy scores. Independent 2-tailed *t* tests were used to compare HIV-ASES scores between groups, while paired 2-tailed *t* tests were used to compare scores within groups over time, with missing data handled as missing completely at random due to external factors unrelated to the study.

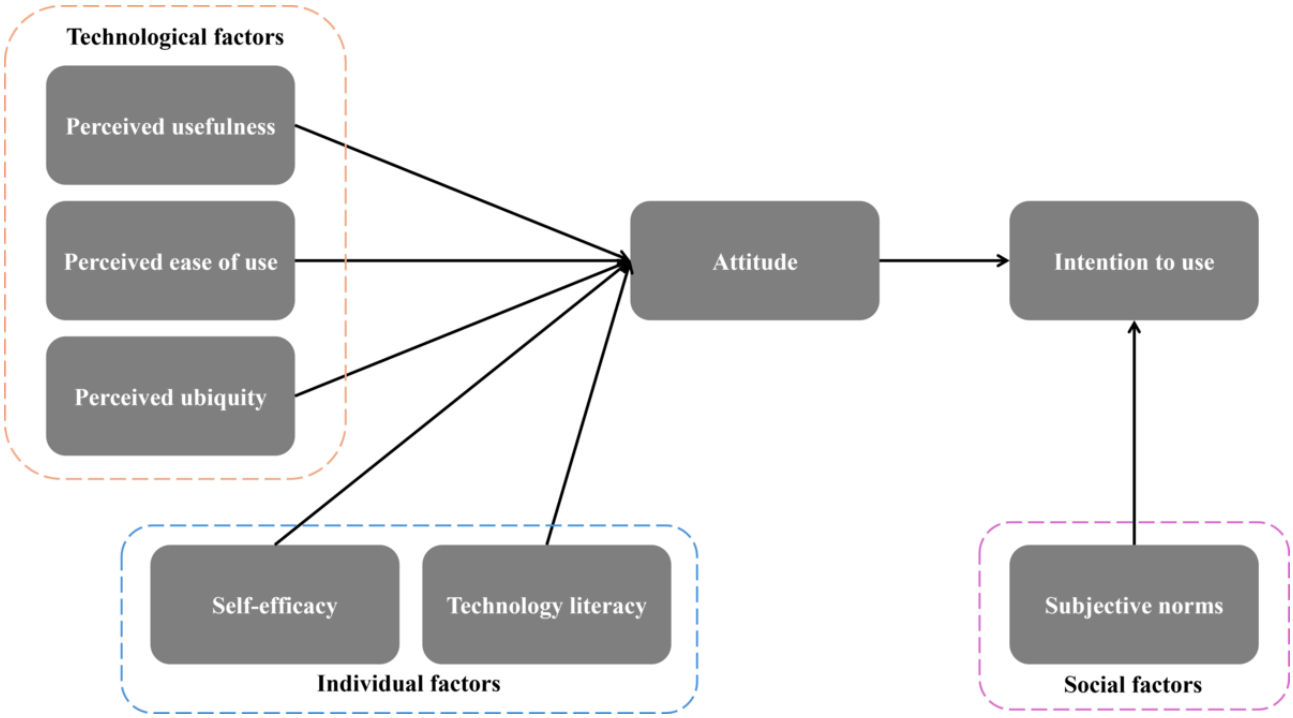
Cronbach α was calculated for scales such as SRSI, Health-ITUES, PSSUQ, and HIV-ASES to measure internal reliability. Statistical analyses were conducted using SAS (SAS Institute) [46] and R software (R Foundation for Statistical Computing) [47], and significance was determined at a *P* value of .05.

Qualitative Data Analysis

Following transcription, 3 members of the research team (SS, SM, and TC) conducted a double-coding process on the transcripts from 52% (12/23) intervention group participants and 48% (11/23) control group participants. An initial codebook was developed based on the Mobile Health Technology Acceptance Model (MHTAM), which includes technological, individual, and social factors that influence the acceptance of health care technology [22]. This model has been adapted for this study to examine an individual’s intention to use mHealth interventions for HIV treatment adherence or participate in the CHAMPS pilot study procedures, as it examines key factors such as perceived usefulness, ease of use, and ubiquity, alongside individual beliefs and social influences (Figure 3). Data analysis was conducted using Excel (Microsoft Corporation), which has been demonstrated as a practical tool for qualitative analysis due to its capabilities for organizing, coding, and analyzing textual data [48]. The methodological approach was guided by Ose [49], who outlined systematic steps for using Excel to manage and code qualitative data efficiently.

XSL•FO
RenderX

Figure 3. The Mobile Health Technology Acceptance Model adapted to analyze participant perceptions of the Community Health Worker and Mobile Health to Improve Viral Suppression (CHAMPS) intervention and experiences participating in the CHAMPS pilot study.



The codebook was created to align with the themes from the interview guide, resulting in the following finalized codes: (1) perceived usefulness, (2) perceived ease of use, (3) perceived ubiquity (ie, the perception of seamless access to health care networks and services at any time and place), (4) self-efficacy, (5) technology literacy, and (6) subjective norms. In addition, the codebook was applied deductively to analyze barriers and facilitators related to both the CHAMPS intervention and study procedures.

The coding process involved all coders coding 2 transcripts for initial consensus with the codebook and double coding 81% (17/21) of the remaining transcripts to enhance reliability, with regular meetings to compare and synthesize definitions for consistency. In certain cases, a consensus coder resolved discrepancies to ensure agreement. A small number of transcripts (4/21, 19%) were single coded due to availability constraints. Overall, double coding and consensus coding were applied to improve intercoder reliability and ensure a rigorous analytic process [50,51].

Ethical Considerations

All study procedures were reviewed and approved by the Columbia University Institutional Review Board (protocol number AAAU2064) before the recruitment or enrollment of participants. Participants provided electronic written consent, including screening consent for those requiring a screening visit and study consent for those proceeding directly to baseline. The baseline consent included language informing participants about an optional follow-up interview, with participants indicating their willingness to participate by initialing the consent form. Study data, including interview transcripts, were anonymized and deidentified to ensure confidentiality. Participants were compensated as follows: no compensation for the screening visit, US $40 for the baseline visit, US $50 for the follow-up visit, and US $35 for the optional qualitative interview at the follow-up visit. All images and data presented in this manuscript are deidentified to ensure the privacy of our participants.

Results

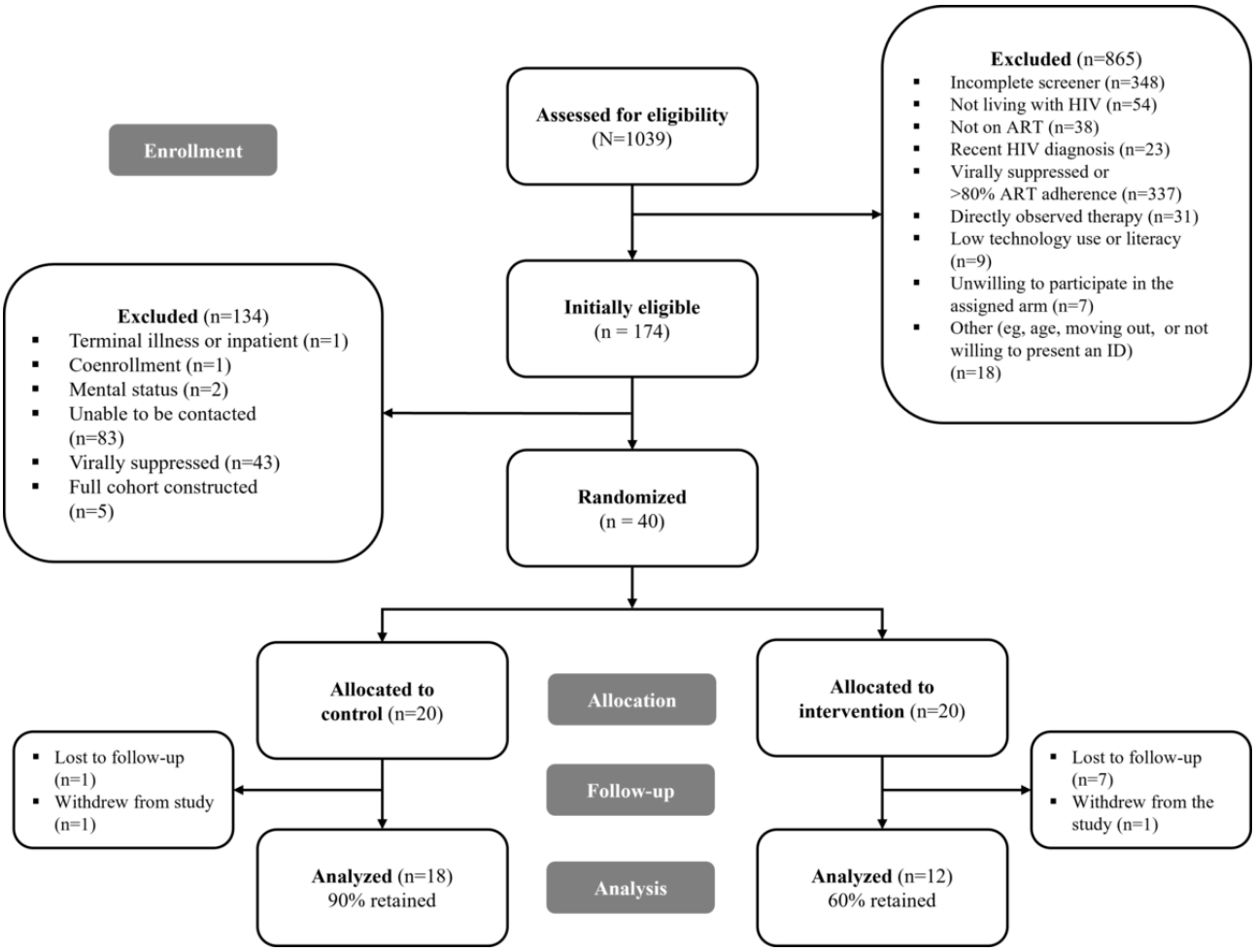
Participant Characteristics

Overview

A total of 40 participants were enrolled in the study, with 20 (50%) participants each randomly assigned to the intervention and control groups. Retention rates were lower in the intervention group (12/20, 60%) compared to the control group (18/20, 90%). Although retention was higher in the control group, this difference was not statistically significant (P=.07; Figure 4). Overall, the study retained 75% (30/40) of the participants, reflecting strong follow-up rates and participation in the study.

XSL•FO
RenderX

Figure 4. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of participant enrollment, allocation, follow-up, and analysis within the Community Health Worker and Mobile Health to Improve Viral Suppression (CHAMPS) pilot study.



Participant Demographics

At baseline, participant demographics were well balanced between the intervention and control groups, indicating that randomization was effective. There were no statistically significant differences in key sociodemographic variables, including age, gender, race, or income level, between the 2 groups (all *P*>.05). The median age of the participants in the control group was 51.5 (IQR 22.0) years, compared to 45.0 (IQR 16.8) years in the intervention group. Gender distribution was similar, with 60% (12/20) of participants in each group identifying as male and 40% (8/20) of participants in the intervention group and 35% (7/20) of the participants in the control group identifying as female. 1 participant in the control group (1/20, 5%) identified as “something else” and were provided the opportunity to elaborate more on how they conceptualized their gender. The racial composition was predominantly African American or Black (10/20, 50% in the control group and 14/20, 70% in the intervention group.)

Educational attainment, employment status, and health insurance coverage were also comparable between the groups, supporting the conclusion that randomization successfully minimized baseline differences between the intervention and control arms. Notably, most participants (25/40, 63%) in both intervention and control groups reported annual incomes <US $20,000, potentially highlighting shared socioeconomic barriers that might limit their ability to fully engage with HIV care and maximize its potential benefits (Table 2).

XSL•FO
RenderX

Table 2. Baseline sociodemographic characteristics of participants, comparing the control and intervention groups (N=40).

Category	Control (n=20)	Intervention (n=20)	*P* value
Age (y), median (IQR)	51.5 (22.0)	45.0 (16.8)	.53[a]
Gender, n (%)			
Male	12 (60)	12 (60)	.99
Female	7 (35)	8 (40)	.78
Something else	1 (5)	0 (0)	.32
Sex, n (%)			
Male	13 (65)	12 (60)	.84
Female	7 (35)	8 (40)	.80
Sexual orientation, n (%)			
Homosexual, gay, or lesbian	8 (40)	10 (50)	.64
Heterosexual or straight	9 (45)	7 (35)	.62
Bisexual	2 (10)	2 (10)	.99
Asexual	1 (5)	0 (0)	.32
Something else (please specify)	0 (0)	1 (5)	.32
Race, n (%)			
American Indian, Alaska Native, Native Hawaiian, or other Pacific Islander	0 (0)	0 (0)	.99
Asian or Asian American	1 (5)	0 (0)	.32
Black or African American	10 (50)	14 (70)	.41
White	9 (45)	5 (25)	.29
Something else	0 (0)	1 (5)	.32
Ethnicity, n (%)			
Hispanic or Latino	2 (10)	2 (10)	.99
Not Hispanic or Latino	18 (90)	18 (90)	.99
Relationship status, n (%)			
Single	10 (50)	14 (70)	.41
In a relationship with a man	6 (30)	3 (15)	.32
In a relationship with a woman	1 (5)	0 (0)	.32
Divorced or separated from a man	1 (5)	0 (0)	.32
Divorced or separated from a woman	1 (5)	0 (0)	.32
Widowed	0 (0)	1 (5)	.32
Other	1 (5)	2 (10)	.56
Children, n (%)			
No	12 (60)	15 (75)	.56
Yes	8 (40)	5 (25)	.41
Education, n (%)			
None	0 (0)	0 (0)	.99
Elementary school	0 (0)	0 (0)	.99
Some high school or no diploma	2 (10)	3 (15)	.66
High-school diploma or equivalent (eg, GED[b])	5 (25)	5 (25)	.99
Some college	6 (30)	4 (20)	.53
Associate degree or technical degree	4 (20)	3 (15)	.71

XSL•FO
RenderX

Category	Control (n=20)	Intervention (n=20)	*P* value
Bachelor or college degree	2 (10)	3 (15)	.66
Professional or graduate degree	1 (5)	2 (10)	.56
Employment status, n (%)			
Working full time	3 (15)	4 (20)	.71
Working part time (eg, seasonal and work-study)	6 (30)	3 (15)	.32
Unemployed and looking for work	3 (15)	5 (25)	.48
Unemployed and not looking for work	0 (0)	2 (10)	.16
Retired	5 (25)	2 (10)	.26
Disabled	4 (20)	6 (30)	.53
Annual income (US $), n (%)			
<10,000	7 (35)	8 (40)	.80
10,000-19,999	5 (25)	5 (25)	.99
20,000-39,999	4 (20)	2 (10)	.41
40,000-59,999	3 (15)	1 (5)	.32
60,000-79,999	0 (0)	1 (5)	.32
80,000-99,999	0 (0)	1 (5)	.32
100,000-149,999	0 (0)	1 (5)	.32
≥150,000	1 (5)	0 (0)	.32
I do not know	0 (0)	1 (5)	.32
Health insurance status, n (%)			
No	1 (5)	0 (0)	.32
Yes, through my job	0 (0)	1 (5)	.32
Yes, through a health exchange (Affordable Care Act)	1 (5)	2 (10)	.56
Yes, Medicaid or Medicare	17 (85)	17 (85)	.99
Yes, AIDS Drug Assistance Program	3 (15)	3 (15)	.99
Other	1 (5)	1 (5)	.99
I do not know	0 (0)	0 (0)	.99

[a]The *P* value for age was calculated using the Mann-Whitney *U* test to compare the distribution of ages between the control and intervention groups based on the full dataset of individual ages.

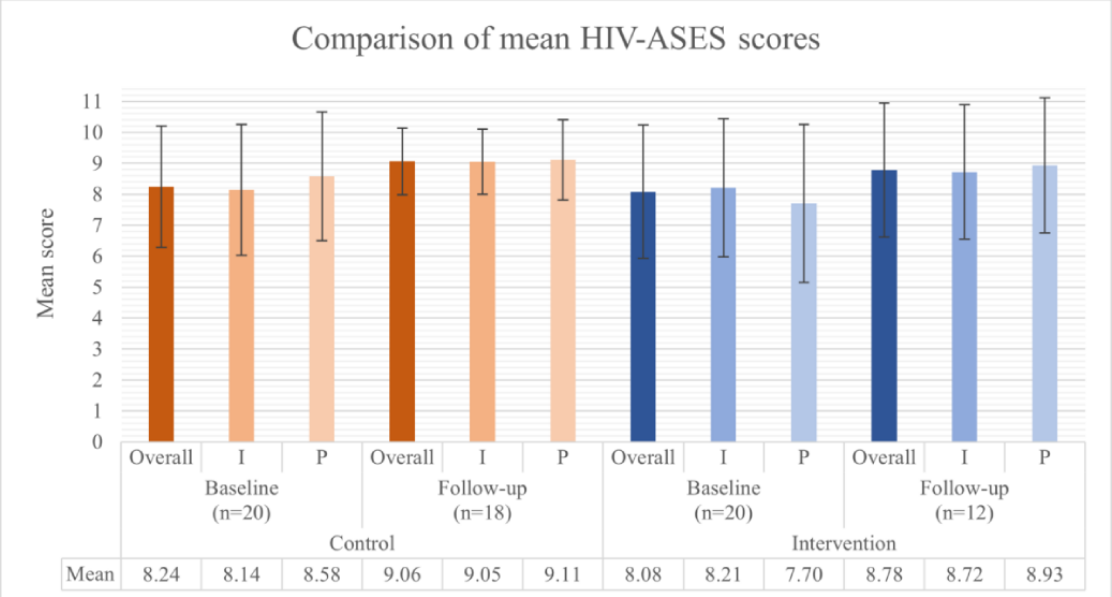
[b]GED: General Educational Development.

Self-Efficacy

Self-efficacy has been recognized as a key factor influencing medication adherence in HIV treatment and other medical conditions [43]. As such, it was quantitatively measured in this study using the HIV-ASES, which has a maximum mean score of 10. At baseline, the control group (20/20, 100%) had a mean HIV-ASES overall score of 8.24 (SD 1.96), and the intervention group (20/20, 100%) had a mean score of 8.08 (SD 2.25). At follow-up, the control group's (18/20, 90%) mean score increased to 9.06 (SD 1.08), while the intervention group's (12/20, 60%) mean score was 8.78 (SD 2.16). Although unexpected, the intervention group's lower self-efficacy score compared to the control group may reflect increased awareness of their adherence habits through engagement with the CHAMPS intervention, leading to more accurate and self-critical reporting. Conversely, social desirability bias may have inflated the control group's self-reported scores. No significant differences were observed between the groups at baseline (*P*=.81) or follow-up (*P*=.64), and no significant within-group changes were observed over time (control: *P*=.52 and intervention: *P*=.07; Figure 5). The overall scale had excellent internal consistency, with Cronbach α=0.95 at baseline and Cronbach α=0.98 at follow-up [52].

XSL•FO
RenderX

Figure 5. Self-efficacy for antiretroviral therapy adherence among control and intervention group participants, as measured by the HIV Treatment Adherence Self-Efficacy Scale (HIV-ASES) at baseline and follow-up. Error bars represent SDs. No statistically significant differences were observed between groups or within groups over time. I: Integration subscale; P: Perseverance subscale.



For the integration subscale, which includes the first 9 questions on the HIV-ASES and measures participants' ability to incorporate treatment into their daily lives [43], mean baseline scores were 8.14 (SD 2.11) for the control group and 8.21 (SD 2.23) for the intervention group. At follow-up, the mean scores increased to 9.05 (SD 1.05) and 8.72 (SD 2.17), respectively. No significant between-group differences were found at baseline (P=.92) or follow-up (P=.58), and there were no significant within-group changes (control: P=.26 and intervention: P=.89; Figure 5). The integration subscale had excellent reliability, with Cronbach α=0.95 at baseline and Cronbach α=0.97 at follow-up [52].

For the perseverance subscale, which includes the final 3 questions on the HIV-ASES and measures participants' confidence in maintaining adherence to their treatment regimens despite any challenges [43], the control group's baseline mean score was 8.58 (SD 2.08), and the intervention group's mean score was 7.70 (SD 2.55). Follow-up mean scores were 9.11 (SD 1.30) for the control group and 8.93 (SD 2.18) for the intervention group. No significant between-group differences were found at baseline (P=.24) or follow-up (P=.78), and no significant within-group changes were observed (control: P=.81 and intervention: P=.32; Figure 5). The perseverance subscale had acceptable and good reliability, with Cronbach α=0.87 at baseline and Cronbach α=0.93 at follow-up [52].

The quantitative findings, which indicated high self-efficacy in both the control and intervention groups, were further supported by qualitative feedback from participants, who described their personal commitment to staying on top of their medication routines. 1 participant shared how self-efficacy played a role in maintaining adherence:

> *Because I just like to have a good track record of me taking my medication in a timely fashion. So, when I saw that face wasn't a smile, and it was a frown, it made me stay on top of my, you know, on my game in terms of me being more consistent. Well, maintaining my consistency with taking my medication.* [CHP 13]

These findings highlight the strong sense of personal responsibility and self-efficacy that participants expressed in their medication adherence, where they were motivated by a desire to maintain their own health. A detailed list of additional illustrative quotes organized by the MHTAM themes and their respective definitions is provided in Multimedia Appendix 1.

Delivered Intervention

A total of 10 CHW sessions were conducted as planned, although the optional 2 additional sessions were not used due to declining participant engagement and the proximity of some sessions to the participants' final follow-up visits. CHWs used HIPAA-compliant Zoom calls, phone calls, and the WiseApp chat feature based on participant preferences. Completion rates (Table 3) were initially high, with 17 (85%) out of the 20 intervention participants completing sessions 1 to 3, but this declined to 9 (45%) participants by the 10th session. Participants generally preferred phone calls for session delivery, with a consistent majority opting for this modality across all sessions. The variability in delivery modes during sessions 1 and 2 reflects logistical factors: some participants had not received their CleverCap devices at the time of their baseline visit, necessitating follow-up sessions via Zoom call or phone call for session 2. In addition, 3 participants disengaged from the study immediately after the baseline visit, which contributed to the drop in session completion rates early in the intervention. To re-engage participants who missed sessions, CHWs attempted up to 3 follow-ups using participants' disclosed communication preferences (calls, SMS text messaging, or emails). When no response was received after 3 attempts, re-engagement efforts were discontinued.

XSL•FO
RenderX

Table 3. Completion rates and delivery model for community health worker sessions among participants in the intervention arm (n=20).

Session number	Participants completing the session, n (%)	Mode of session delivery, n (%)		
		Zoom	Phone	WiseApp chat
1	17 (85)	9 (53)	8 (47)	0 (0)
2	17 (85)	3 (18)	14 (82)	0 (0)
3	17 (85)	1 (6)	15 (88)	1 (6)
4	13 (65)	0 ()	13 (100)	0 (0)
5	13 (65)	1 (8)	11 (85)	1 (8)
6	13 (65)	0 (0)	13 (100)	0 (0)
7	13 (65)	0 (0)	12 (92)	1 (8)
8	12 (60)	1 (8)	11 (92)	0 (0)
9	12 (60)	1 (8)	10 (83)	1 (8)
10	9 (45)	2 (22)	7 (78)	0 (0)

The CleverCap pill bottles issued 3151 notifications, comprising notifications for taking their medication and alerts for nonadherence or improper use (Table 4). Notifications sent per participant varied widely, with a mean value of 158.3 (SD 99.9). Participants receiving >200 notifications often had frequently missed or off-schedule doses, while those receiving <90 notifications likely disengaged from the intervention by deactivating, losing, or destroying their pill bottles. This 90-notification threshold was based on the intervention design, which included 1 scheduled daily reminder over the 3-month study period (3 months ×30 days=90 reminders per participant). Any participant receiving substantially <90 notifications likely did not use the device consistently, whereas those receiving >90 notifications frequently triggered additional alerts due to missed or off-schedule doses.

Table 4. Frequency and distribution of medication adherence notifications issued by the CleverCap pill bottle to participants in the intervention arm (N=3151).

Notification type	Notifications, n (%)
Dose taken	1061 (33.67)
Missed dose	1585 (50.3)
Off-schedule dose	418 (13.26)
Improperly closed pill bottle	87 (2.76)

Intervention Characteristics

Overview

The aOR for achieving undetectable viral load in the intervention group compared to the control group was 3.01 (95% CI –1.59 to 4.12), controlling for baseline viral load status (Table 5). This aOR represents a medium effect size, suggesting an improvement in viral suppression in the intervention group [53]. Baseline viral load status by follow-up status and group is presented in Multimedia Appendix 2.

Table 5. Baseline and follow-up viral load status of control and intervention group participants (N=40).

Viral load[a]	Baseline, n (%)			Follow-up, n (%)		
	Control (n=20)	Intervention (n=20)	Total (n=40)	Control (n=18)	Intervention (n=12)	Total (n=29)
Not detectable (<200 copies/ml)	17 (85)	11 (55)	28 (70)	15[b] (83)	11 (92)	26 (90)
Detectable (≥200 copies/ml)	3 (15)	9 (45)	12 (30)	2 (11)	1 (8)	3 (10)

[a]For consistency with US clinical guidelines, viral load status is categorized using a threshold of ≥200 copies/ml to define detectability. While participants with viral loads between 50 and 199 copies/ml were eligible for the study, they were classified as suppressed in this analysis. In addition, details on the methods used to obtain viral load measurements at baseline and follow-up are provided in Multimedia Appendix 3.

[b]Data missing (1/18, 6%) as an insufficient sample provided by 1 participant precluded analysis and quantification.

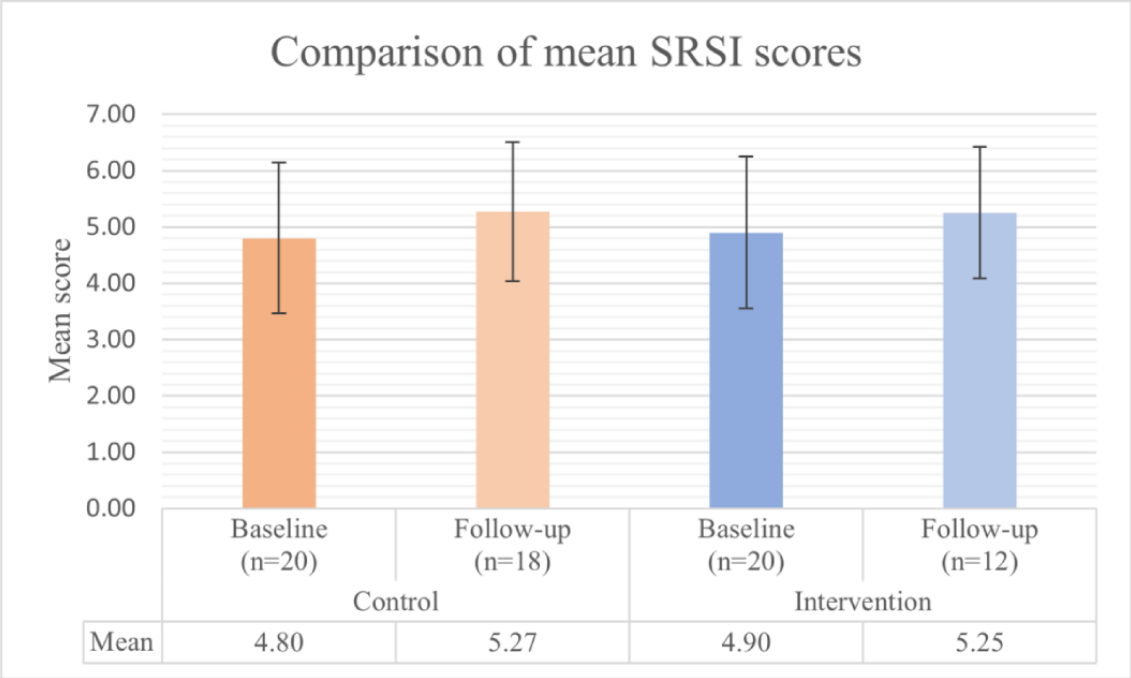
Notably, nearly three-fourths of the participants (28/40, 70%) had an undetectable viral load (<200 copies/ml) at baseline. Among those with undetectable viral load at baseline, viral suppression was largely maintained throughout the study period. Specifically, in the control group, the viral load of 1 participant became detectable at follow-up, whereas in the intervention group, all participants' viral loads that were initially undetectable remained undetectable. Among participants with a detectable

XSL•FO
RenderX

viral load (≥200 copies/ml) at baseline, 44% (4/9) of individuals in the intervention group achieved viral suppression by follow-up, compared to only 33% (1/3) in the control group.

At baseline, the control group (20/20, 100%) had a mean SRSI score of 4.80 (SD 1.34), while the intervention group (20/20, 100%) had a mean score of 4.90 (SD 1.35). At follow-up, the control group's (18/20, 90%) mean increased to 5.27 (SD 1.23), and the intervention group's (12/20, 60%) mean was 5.25 (SD 1.17). No significant differences were observed between the groups at baseline (P=.77) or follow-up (P=.95), and no significant within-group changes were observed over time (control: P=.16 and intervention: P=.29; Figure 6). The SRSI score showed excellent internal consistency, with Cronbach α=0.96 at baseline and Cronbach α=0.98 at follow-up [52].

Figure 6. Changes in self-reported adherence to antiretroviral therapy among control and intervention group participants, as measured using the self-reported scale item (SRSI) at baseline and follow-up. Error bars represent SDs. No statistically significant differences were observed between groups or within groups over time.



Adherence data from CleverCap showed that intervention group participants (12/20, 60%) took an average of 60.08% (SD 19.15%) of their prescribed doses, with 39.33% (SD 19.24%) of doses missed. A weak, nonsignificant correlation was found between the CleverCap adherence data (percentage taken) and follow-up SRSI scores (r=0.121; P=.71), suggesting that the high self-reported adherence scores did not strongly align with the electronic monitoring data.

Participant feedback can help explain the discrepancy between self-reported scores and actual adherence by highlighting individual factors, such as personal medication management preferences, that influence adherence. 1 participant noted the following, illustrating how participants' personal beliefs about the medication's dosing flexibility influenced their adherence behavior:

> *I know that Biktarvy has a 50-hour shelf-life. And so, I'm not really stressing over when I take it. I usually just make sure I take it before lunch.* [CHP 2]

Rather than relying on CleverCap's programmed alarms, participants adjusted their dosing schedules based on their own understanding of the medication's efficacy, which may have contributed to lower adherence rates recorded by the pill bottle.

Feasibility of the Remotely Delivered CleverCap Pill Bottle and the WiseApp Intervention

The feasibility of the intervention was assessed by supplementing usability data with feedback regarding participants' attitudes and intentions to use the intervention [54,55]. The usability of the intervention was assessed using the Health-ITUES and PSSUQ scores, which provided insights into participants' experiences and satisfaction with the intervention, respectively. The Health-ITUES had a maximum mean score of 5, with higher scores indicating better usability. Participants (12/20, 60%) rated the overall usability of the intervention positively, with an average overall score of 4.35 (SD 0.58). The scores ranged from 3.50 to 5, with an IQR of 0.84, reflecting consistently high usability ratings. Specific subscales, such as perceived usefulness (mean 4.42, SD 0.63) and perceived ease of use (mean 4.47, SD 0.71), received similarly high ratings, with IQRs of 1.11 and 0.50, respectively, indicating that participants generally found the pill bottle and app useful and easy to use (Table 6).

XSL•FO
RenderX

Table 6. Usability and acceptability of the Community Health Worker and Mobile Health to Improve Viral Suppression intervention, as assessed by the Health Information Technology Usability Evaluation Scale (Health-ITUES) and Poststudy System Usability Questionnaire (PSSUQ) among intervention group participants.

Scale	Scores, mean (SD)	Scores, median (IQR; range)	Cronbach α
Health-ITUES (n=12)			
Overall	4.35 (0.58)	4.48 (0.84; 3.50-5)	0.94
Impact	4.36 (0.66)	4.17 (1.08; 3.33-5)	0.92
Perceived usefulness	4.42 (0.63)	4.33 (1.11; 3.11-5)	0.83
Perceived ease of use	4.47 (0.71)	4.90 (0.50; 3.67-5)	0.63
User control	3.86 (0.95)	4.07 (1.75; 2.33-5)	0.79
PSSUQ (n=12)			
Overall	2.04 (1.03)	3.81 (1.31; 1-4.19)	0.96
System usefulness	1.87 (1.09)	1.79 (0.79; 1-4.86)	0.94
Information quality	2.13 (1.16)	1.67 (2.04; 1-3.67)	0.96
Interface quality	2.05 (0.98)	2.17 (1.58; 1-4.33)	0.86

These high ratings are echoed in participant feedback. The pill bottle's visual presence was perceived as useful for maintaining adherence. Participants also found the pill bottle easy to operate and intuitive, sharing positive experiences as follows:

> *Because I'm not really good with taking it at a certain time every day, at the same time every day. And so I feel like the CleverCap helped me with that...It does remind me, because I always see it, versus me keeping it in a white bottle. When I see the CleverCap on my dresser, it just reminds me. "Okay, it's time to take your medicine."...The functions of the CleverCap were easy to use. I didn't have to charge it many times. I charged it once. It's very easy. It reminded me to take my medicine. I really, really appreciated that. I just love it. It's very easy. And it definitely reminded me to stay on my meds.* [CHP 39]

However, the user control subscale received a lower average score (mean 3.86, SD 0.95) and a wider IQR of 1.75, suggesting more variability in responses and some limitations in how participants felt they could manage the intervention. These concerns were reflected in feedback about the pill bottle's fragility and usability during travel:

> *Because there were times I was out of town, and I don't like bringing the pill bottle with me...I didn't take the bottle. I took my medication, but I didn't take the reminder, the bottle...That container, with the alarm on, it's very sensitive. I dropped it one time; I was afraid it might break...So, that's why when I knew I was leaving to go out of town, I just felt it was more secure leaving it at home.* [CHP 13]

Despite this, some participants (3/12, 25%) were able to maintain adherence even when separated from the pill bottle, relying on their own strategies for medication management. 1 participant reflected on this experience, demonstrating that while CleverCap supported adherence, participants' self-efficacy and personal management strategies often played a crucial role in overcoming barriers related to the pill bottle's portability and user control:

> *When I went to D.C., I forgot the CleverCap...But I had my medicine. I just got the notifications like, oh, it's time to take your medicine. Didn't bring the cap.* [CHP 32]

The PSSUQ had a maximum mean score of 7, where lower scores indicate better usability. Participants reported generally high satisfaction with the system's usability, with an average overall score of 2.04 (SD 1.03). The scores ranged from 1 to 4.19, with an IQR of 1.31, indicating a favorable user experience with some variability. Ratings for system usefulness (mean 1.87, SD 1.09) and interface quality (mean 2.05, SD 0.98) subscales also indicate positive user experiences, with IQRs of 0.79 and 1.58, respectively, although information quality subscale (mean 2.13, SD 1.16) showed highest variability in participant responses, with an IQR of 2.04 (Table 5).

To assess the internal reliability of the scales used, Cronbach α was calculated for both the Health-ITUES and PSSUQ. The Health-ITUES demonstrated excellent overall reliability (Cronbach α=0.94), with variable consistency across the following subscales: perceived usefulness (Cronbach α=0.92, excellent), integration (Cronbach α=0.83, good), perceived ease of use (Cronbach α=0.63, acceptable), and user control (Cronbach α=0.76, acceptable). The PSSUQ exhibited excellent overall reliability (Cronbach α=0.96), with high consistency across the following subscales: system usefulness (Cronbach α=0.94, excellent), information quality (Cronbach α=0.96, excellent), and interface quality (Cronbach α=0.86, good) [52].

Our results suggest that, despite potential improvements in user control and information clarity, the remote delivery of the CleverCap pill bottle and WiseApp was generally well received, demonstrating feasibility and ease of use.

Perceptions of CHW Sessions

Participants shared a variety of experiences regarding their engagement with the CHW sessions, with responses generally falling into 3 key themes: ease of use and self-efficacy as a facilitator; positive social influence; and barriers to ubiquity, such as infrastructure or scheduling challenges.

XSL•FO
RenderX

Several participants (10/12, 83%) expressed that the flexibility and accessibility of the CHW sessions facilitated their engagement. The ability to communicate with CHWs through various channels, such as text, phone, and email, was highly appreciated. For instance, 1 participant highlighted the ease of communication with and persistence of the CHWs. In addition, participants with existing knowledge about HIV felt comfortable discussing their treatment with the CHWs, further facilitating their engagement:

> *The flexibility that you had was amazing. Like I said, you guys were available through a text message, through a phone call, through email...whatever it took. You guys didn't disappear. You guys didn't give up. You guys are like, hello, I am trying to talk to you, and kept on it. So, that was cool...To a person who already knows about HIV, I am completely comfortable talking about anything that has to do with it. So, that wasn't an issue for me.* [CHP 17]

Positive social influence, particularly through regular check-ins and motivational support from the CHWs, emerged as a key facilitator of engagement. Participants expressed appreciation for the ongoing support provided by the CHWs, with some (10/12, 83%) noting that the regular communication made them feel more accountable and supported in their adherence:

> *[The study] really was awesome. I mean the perk was having you call me once a week...Having your own personal somebody to motivate you and check up on you and make sure you're doing good and adhering to your medicine.* [CHP 39]

While many participants found the CHW sessions helpful, some (2/12, 16%) faced barriers to engagement due to external factors, such as infrastructure limitations or time constraints. Participants living in rural areas with limited internet access found it challenging to maintain consistent communication with CHWs. 1 participant from a rural area explained how poor internet connectivity hindered their participation:

> *Like I said, I'm busy. The issue with scheduling because, just like I said, I was always on the go. I'm always on the go...Travel, not having internet for interaction with you. That was sometimes challenging, I guess. Where I live, here in rural Alabama, I dare say that 75% of the county... [does] not have any kind of access to internet. It's scarce...the internet and phone service in a lot of spaces. And so that would be a challenge to speaking with you, as well as trying to look at any data on the phone.* [CHP 38]

In addition, some participants (4/12, 33%) mentioned that their busy schedules made it difficult to consistently engage in the CHW sessions (Multimedia Appendix 1).

Thus, while there were some challenges related to infrastructure and scheduling, participants' perceptions suggest that the CHW sessions were generally well received, particularly in terms of flexibility, ease of communication, and motivational support.

Acceptability of Remote Study Procedures

The acceptability of delivering the intervention and conducting study procedures remotely was assessed through participant feedback, which highlighted the overall usefulness of participating in the study; ease of remote interactions; and logistical considerations, such as shipping and blood sample collection.

Participants largely embraced the flexibility offered by remote study visits conducted via Zoom. Many (15/23, 65%) expressed comfort and appreciation for the convenience of participating from their own homes, along with increased privacy and adaptability. 1 participant noted the following:

> *I have used Zoom in the past for like work...But with Zoom, it's actually more...what's the word I'm looking for? This was more private...If I had to step out and go somewhere, I still could do it that way. But, I felt more comfortable doing it on my laptop, you know, at home. And, I liked the flexibility I had with you all, as well. So, it was great, because you all worked around my schedule.* [CHP 23]

Overall, participants found Zoom easy to use and appreciated the accessibility of conducting study visits remotely, especially in cases where mobility or health concerns made in-person visits difficult. 1 participant explained as follows:

> *Doesn't bother me. I'd rather [Zoom] than have to come out. It's difficult for me to walk right now...This is good because at least I can participate in things. And I don't have to make myself uncomfortable to do it, you know?* [CHP 30]

Remote collection of blood samples, facilitated through mailed kits, was another aspect of the study that participants generally found convenient. 1 participant described the process as efficient and accessible:

> *I think [the blood sample collection process is] very convenient. I'm nowhere near you guys and normally, you know, we'd have someone say, well, you're not able to come into the office and do the blood draws or whatever. So, this was very convenient.* [CHP 11]

Furthermore, participants expressed appreciation for the comprehensive kits and clear instructions provided (Multimedia Appendix 1). Although some challenges with blood sample collection were reported, including minor issues such as bruising and difficulties with blood flow, these were anticipated and communicated during the informed consent process, as such challenges are commonly associated with point-of-care testing [56]. Multimedia Appendix 3 provides additional numeric data supporting these findings.

The shipping process for study kits and the CleverCap pill bottle was largely seamless for participants, who noted the discreet packaging of materials as a positive aspect of the study. 1 participant mentioned how the packaging helped protect their privacy:

> *Just because of the delicateness of the subject matter, it's essentially discreet because nobody thinks a FedEx package is anything suspect or scary. It's just*

XSL•FO
RenderX

> *a FedEx package...it blended right in and it doesn't cause any red flags or make anybody feel uncomfortable in a public setting.* [CHP 17]

However, there were occasional logistical challenges with returning the packages. A few participants (2/23, 9%) noted issues with transportation and scheduling FedEx pickups, which delayed their ability to return the materials. 1 participant shared the following:

> *Because, I don't always have transportation, so it's been harder for me to get the package sent back in like a timely manner. I know you guys have tried setting up a pick-up with FedEx, but they never came to pick up the package. So, that's like been the only barrier I've had in participating in the program and getting the package to a FedEx location.* [CHP 27]

While logistical challenges, such as transportation for package returns, were noted, the study's design was practical and effective for most participants, supporting its acceptability and viability in remote settings.

Discussion

Principal Findings

This pilot study aimed to primarily assess the feasibility and usability of the remotely delivered CHAMPS intervention. While the preliminary results did not demonstrate significant improvements in adherence or self-efficacy in the intervention group compared to the control group, the findings provide important insights into how participant attitudes to use the intervention were shaped by technological, individual, and social factors. Thus, understanding these sociodemographic factors is essential for contextualizing the intervention's feasibility and the challenges participants faced in engaging with the CHAMPS intervention. The demographic characteristics of participants in this study partially align with key sociodemographic characteristics of people with HIV in the United States. These parallels suggest that our findings may be applicable to subsets of the broader population of people with HIV, particularly those disproportionately affected by intersecting socioeconomic and racial disparities. Most of our participants (24/40, 60%) identified as Black or African American individuals, consistent with national data showing that nearly half of people with HIV belong to this group [57]. The predominance of participants reporting annual incomes <US $20,000, which is nearly double the proportion of individuals living below the federal poverty level (US $15,060/y) compared to the national estimate of one-third of people with HIV, may reflect competing priorities that made it challenging to fully engage with the intervention and may help explain the nonsignificant results observed [57].

Baseline HIV-ASES scores showed that participants in both control and intervention groups had high self-efficacy, particularly in integrating HIV medication into their routines and maintaining long-term adherence. This is critical, as self-efficacy is a well-established predictor of successful health behavior change, particularly for chronic conditions such as HIV [58,59]. Although modest increases in scores were observed at follow-up, they were not statistically significant, and the control group reported higher self-efficacy than the intervention group, indicating that the intervention did not significantly enhance an already strong sense of self-efficacy. Qualitative feedback confirmed that participants relied on their preexisting knowledge and ability to adhere to medication, making additional tools unnecessary.

SRSI scores showed stable adherence in both groups from baseline to follow-up, with slight, nonsignificant increases. Although the intervention did not significantly improve adherence, qualitative data revealed that participants, especially those with higher self-efficacy, often prioritized their own understanding of medication efficacy over CleverCap reminders. High confidence in managing their treatment may have made these reminders less relevant, explaining the lack of significant differences between groups.

Health-ITUES results showed the intervention's success in usability, with an overall score of 4.35 exceeding the acceptable cut point of 4.32 [60]. High scores in perceived usefulness and ease of use subscale indicated that participants found WiseApp helpful and intuitive, supporting the MHTAM framework's emphasis on these factors for technology adoption [22]. However, lower scores in user control revealed challenges with CleverCap's portability, aligning with MHTAM's perceived ubiquity construct. PSSUQ scores also reflected positive usability, with an overall mean of 2.04 (SD 1.03). Participants rated system usefulness and interface quality highly.

Subjective norms emerged as an important factor for engagement with the CHW component of the intervention. The regular check-ins and motivational support from CHWs were highlighted as valuable by participants, reinforcing the social component of adherence. This reflects the broader role of CHWs in providing not only practical assistance but also emotional support, which has been shown to improve health outcomes in other studies [21-23].

This pilot study found high acceptability for remotely delivered procedures, with participants appreciating the convenience, privacy, and flexibility of remote interactions. Zoom visits were well received, allowing participation from home, which aligns with literature suggesting that remote health care improves patient satisfaction by reducing travel burden and enhancing accessibility [61-63]. These procedures were key in promoting participation and retention. Blood sample collection via mailed kits was convenient for most participants (27/30, 90%), demonstrating that, with clear guidance, it is a feasible alternative to in-person phlebotomy. Discreet packaging of study materials was also valued, as it addressed privacy needs and reduced HIV stigma, which is a critical factor in engaging people with HIV in care [64,65].

Our findings highlight the potential of mHealth interventions, such as CHAMPS, to support the "Treat" pillar of the Ending the HIV Epidemic initiative by enhancing accessibility and convenience for people with HIV. The usability and acceptability of remote delivery methods, such as mobile apps for medication management, CHW support, and mailed DBS collection, demonstrate a promising foundation for addressing barriers to care. These strategies could be further refined to promote viral suppression and improve engagement in HIV care

XSL•FO
RenderX

for populations facing challenges in accessing traditional health care services.

Limitations

This study has several limitations that should be considered. The small sample size of 40 participants limited the statistical power, making it difficult to detect significant differences between groups and reducing the generalizability of findings to the broader populations of people with HIV. Furthermore, the small sample size for positive outcome events, such as viral suppression at follow-up, may have affected the reliability of the effect size estimate, particularly within the intervention group. The lower retention rate in the intervention group compared to the control group, while not statistically significant, suggests challenges in effectively engaging participants assigned to the intervention. This highlights the need to address structural and social determinants of health when designing ART adherence interventions. Future research should explore strategies to reduce these barriers, such as more flexible engagement options or additional resources.

In addition, recruitment challenges, particularly among younger people with HIV (aged 18 to 29 y), were notable. Social media platforms used for advertising, such as POZ.com, primarily attracted older participants due to their demographic reach. Alternative platforms, such as TikTok, which are more commonly used by younger populations, should be prioritized in future campaigns [66]. Strategies such as peer-driven recruitment and culturally tailored SMS text messaging could help engage younger individuals who may be less trusting of traditional health care outreach methods. Addressing time constraints and providing more flexible scheduling options may also increase participation among this demographic [67,68]. Furthermore, it is important to note that participants who responded to advertisements on POZ.com may not be representative of the broader US population living with HIV, introducing potential selection bias. Financial incentives for study participation may have influenced responses, potentially skewing the participant pool toward individuals with greater financial needs.

Another significant limitation is the reliance on self-reported adherence data in the control group, which is prone to bias. Self-reported measures, particularly in the context of health behaviors, can be subject to overreporting due to recall bias or social desirability, where participants may provide responses that they believe are more acceptable to researchers [69,70]. Although the SRSI scale demonstrated high internal consistency, discrepancies between self-reported adherence and CleverCap data suggest that participants' actual behaviors differed. Some adjusted their adherence schedules based on personal beliefs, indicating that self-reported data may not fully reflect true adherence. In addition, the reliance on self-reported viral load data as an inclusion criterion may have introduced some inaccuracies, as nearly half of the sample reported being virally unsuppressed (threshold of ≥200 copies/ml) in the past 12 months despite being clinically undetectable at baseline. This difference suggests that participants may have misremembered or misreported their viral load history, highlighting a potential limitation in self-reported viral suppression measures.

Discrepancies between the planned and delivered intervention fidelity were observed, with several factors impeding full implementation. While the intervention was designed to provide consistent support through regular CHW sessions and timely CleverCap notifications, technical challenges, such as unreliable local telecommunication networks, affected the delivery of mHealth components, leading to missed or delayed notifications. In addition, participant engagement waned over time, with some individuals disengaging from CHW sessions and underusing the CleverCap device, a challenge commonly reported in mHealth interventions [71,72]. The older demographic of the study population introduced additional barriers, as many participants had difficulty programming medication alerts without CHW assistance, leading to missed doses. This aligns with previous research indicating that older adults frequently encounter usability issues with electronic adherence tools, which can negatively impact adherence [71]. These findings highlight the importance of tailoring adherence interventions to address technological literacy and providing structured training for older adults.

Finally, the lack of a universal data verification process for the DBS results posed a limitation. While DBS samples were collected and processed according to established protocols, most participants' (25/30, 83%) DBS results were not cross verified with other sources. Only a small subset of participants (5/30, 17%) who had requested a screening visit and subsequently provided electronic health record viral load data had their DBS results cross-checked. This reliance on unverified DBS data for most participants may introduce potential inaccuracies. Furthermore, the moderate sensitivity (80%-95%) and specificity (85%-90%) of DBS for viral load thresholds, such as 200 to 1000 copies/ml, could lead to misclassification, with some participants incorrectly identified as having high or low viral loads [73-75]. Future studies should incorporate systematic verification processes, such as routine cross-checking with laboratory results or electronic health records, to enhance the reliability of biomarker data.

Conclusions

The findings from this pilot study support the feasibility and acceptability of a remotely delivered mHealth intervention integrated with CHW support for people with HIV. Participants reported high levels of usability and satisfaction with both the WiseApp and CleverCap, highlighting the intervention's potential to promote medication adherence and engagement in care. Despite modest, nonsignificant improvements in self-reported adherence and self-efficacy, the flexibility of the intervention and its ability to address logistical challenges make it a promising tool for broader implementation. These results suggest that with further optimization, this intervention could be a valuable resource in enhancing health outcomes for people with HIV across diverse settings.

XSL•FO
RenderX

Acknowledgments

This study was supported by the National Institute of Nursing Research (award number R01NR019758). Dried blood spot kit processing at the Miriam Hospital (Providence, Rhode Island) was supported by the Providence-Boston Center for AIDS Research under award number P30AI042853. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors would like to extend their gratitude to the participants and community health workers involved in this study, whose contributions were invaluable. The authors would also like to thank the broader research team and their institutional partners for their continuous support and dedication to improving care for people with HIV.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to participant confidentiality and institutional data protection policies, but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Key themes, definitions, and illustrative quotes from control and intervention group participants, organized according to the Mobile Health Technology Acceptance Model.
[PDF File (Adobe PDF File), 174 KB - formative_v9i1e67997_app1.pdf]

Multimedia Appendix 2
Retention of participants at follow-up by baseline viral load status, assessing whether the loss to follow-up was differentially associated with baseline viral load in the intervention and control groups.
[PDF File (Adobe PDF File), 69 KB - formative_v9i1e67997_app2.pdf]

Multimedia Appendix 3
Methods of viral load and CD4 data collection at baseline and follow-up, including self-reported and laboratory measures.
[PDF File (Adobe PDF File), 68 KB - formative_v9i1e67997_app3.pdf]

Multimedia Appendix 4
CONSORT eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 17131 KB - formative_v9i1e67997_app4.pdf]

References

1. Deeks SG, Lewin SR, Havlir DV. The end of AIDS: HIV infection as a chronic disease. Lancet 2013 Nov 02;382(9903):1525-1533 [FREE Full text] [doi: 10.1016/S0140-6736(13)61809-7] [Medline: 24152939]
2. Marcus JL, Chao CR, Leyden WA, Xu L, Quesenberry CP, Klein DB, et al. Narrowing the gap in life expectancy between HIV-infected and HIV-uninfected individuals with access to care. J Acquir Immune Defic Syndr 2016 Sep 01;73(1):39-46 [FREE Full text] [doi: 10.1097/QAI.0000000000001014] [Medline: 27028501]
3. May MT, Gompels M, Delpech V, Porter K, Orkin C, Kegg S, UK Collaborative HIV Cohort (UK CHIC) Study. Impact on life expectancy of HIV-1 positive individuals of CD4+ cell count and viral load response to antiretroviral therapy. AIDS 2014 May 15;28(8):1193-1202 [FREE Full text] [doi: 10.1097/QAD.0000000000000243] [Medline: 24556869]
4. Sematlane NP, Knight L, Masquillier C, Wouters E. Adapting to, integrating and self-managing HIV as a chronic illness: a scoping review protocol. BMJ Open 2021 Jun 23;11(6):e047870 [FREE Full text] [doi: 10.1136/bmjopen-2020-047870] [Medline: 34162650]
5. HIV surveillance supplemental report: estimated HIV incidence and prevalence in the United States. Centers for Disease Control and Prevention. 2018. URL: https://stacks.cdc.gov/view/cdc/156513 [accessed 2024-04-29]
6. McComsey GA, Lingohr-Smith M, Rogers R, Lin J, Donga P. Real-world adherence to antiretroviral therapy among HIV-1 patients across the United States. Adv Ther 2021 Sep 14;38(9):4961-4974 [FREE Full text] [doi: 10.1007/s12325-021-01883-8] [Medline: 34390465]
7. Frandsen BR, Joynt KE, Rebitzer JB, Jha AK. Care fragmentation, quality, and costs among chronically ill patients. Am J Manag Care 2015 May;21(5):355-362 [FREE Full text] [Medline: 26167702]
8. Fauci AS, Redfield RR, Sigounas G, Weahkee MD, Giroir BP. Ending the HIV epidemic: a plan for the United States. JAMA 2019 Mar 05;321(9):844-845. [doi: 10.1001/jama.2019.1343] [Medline: 30730529]
9. Cohen MS, McCauley M, Gamble TR. HIV treatment as prevention and HPTN 052. Curr Opin HIV AIDS 2012;7(2):99-105. [doi: 10.1097/coh.0b013e32834f5cf2]

XSL•FO
RenderX

10. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, HPTN 052 Study Team. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med 2011 Aug 11;365(6):493-505 [FREE Full text] [doi: 10.1056/NEJMoa1105243] [Medline: 21767103]
11. Yu Y, Luo D, Chen X, Huang Z, Wang M, Xiao S. Medication adherence to antiretroviral therapy among newly treated people living with HIV. BMC Public Health 2018 Jul 04;18(1):825 [FREE Full text] [doi: 10.1186/s12889-018-5731-z] [Medline: 29973167]
12. Bandera A, Gori A, Clerici M, Sironi M. Phylogenies in ART: HIV reservoirs, HIV latency and drug resistance. Curr Opin Pharmacol 2019 Oct;48:24-32. [doi: 10.1016/j.coph.2019.03.003] [Medline: 31029861]
13. Martin F, Lee J, Thomson E, Tarrant N, Hale A, Lacey CJ. Two cases of possible transmitted drug-resistant HIV: likely HIV superinfection and unmasking of pre-existing resistance. Int J STD AIDS 2016 Jan 07;27(1):66-69 [FREE Full text] [doi: 10.1177/0956462415571671] [Medline: 25663247]
14. Mugavero MJ, Amico KR, Horn T, Thompson MA. The state of engagement in HIV care in the United States: from cascade to continuum to control. Clin Infect Dis 2013 Oct 23;57(8):1164-1171. [doi: 10.1093/cid/cit420] [Medline: 23797289]
15. Dasgupta S, Tie Y, Beer L, Fagan J, Weiser J. Barriers to HIV care by viral suppression status among US adults with HIV: findings from the Centers for Disease Control and Prevention medical monitoring project. J Assoc Nurses AIDS Care 2021;32(5):561-568 [FREE Full text] [doi: 10.1097/JNC.0000000000000249] [Medline: 33769329]
16. Kempf MC, McLeod J, Boehme AK, Walcott MW, Wright L, Seal P, et al. A qualitative study of the barriers and facilitators to retention-in-care among HIV-positive women in the rural southeastern United States: implications for targeted interventions. AIDS Patient Care STDS 2010 Aug;24(8):515-520. [doi: 10.1089/apc.2010.0065] [Medline: 20672971]
17. Marshall SA, Tao J, Horton HK, Montgomery MC, Baloh J, Gu M, et al. Barriers and facilitators to HIV care engagement: recommendations from people with HIV in rural Arkansas. J Rural Health 2023 Mar 06;39(2):459-468 [FREE Full text] [doi: 10.1111/jrh.12717] [Medline: 36203209]
18. Mills EJ, Nachega JB, Bangsberg DR, Singh S, Rachlis B, Wu P, et al. Adherence to HAART: a systematic review of developed and developing nation patient-reported barriers and facilitators. PLoS Med 2006 Nov 21;3(11):e438 [FREE Full text] [doi: 10.1371/journal.pmed.0030438] [Medline: 17121449]
19. Yehia BR, Stewart L, Momplaisir F, Mody A, Holtzman CW, Jacobs LM, et al. Barriers and facilitators to patient retention in HIV care. BMC Infect Dis 2015 Jun 28;15(1):246 [FREE Full text] [doi: 10.1186/s12879-015-0990-0] [Medline: 26123158]
20. Drainoni ML, Baughman AL, Bachman SS, Bowers-Sword R, Davoust M, Fortu K, et al. Integrating community health workers into HIV care teams: Impact on HIV care outcomes. J HIV/AIDS Soc Serv 2020 Jun 30;19(3):204-219. [doi: 10.1080/15381501.2020.1785364]
21. Greuel M, Sy F, Bärnighausen T, Adam M, Vandormael A, Gates J, et al. Community health worker use of smart devices for health promotion: scoping review. JMIR Mhealth Uhealth 2023 Feb 22;11:e42023 [FREE Full text] [doi: 10.2196/42023] [Medline: 36811947]
22. Hsiao CH, Tang KY. Examining a model of mobile healthcare technology acceptance by the elderly in Taiwan. J Glob Inf Tech Manag 2015 Dec 11;18(4):292-311. [doi: 10.1080/1097198x.2015.1108099]
23. Pollard R, Kennedy CE, Hutton HE, Mulamba J, Mbabali I, Anok A, et al. HIV prevention and treatment behavior change and the Situated Information Motivation Behavioral Skills (sIMB) model: a qualitative evaluation of a community health worker intervention in Rakai, Uganda. AIDS Behav 2022 Feb 30;26(2):375-384 [FREE Full text] [doi: 10.1007/s10461-021-03391-w] [Medline: 34327597]
24. Hammack AY, Bickham JN, Gilliard I, Robinson WT. A community health worker approach for ending the HIV epidemic. Am J Prev Med 2021 Nov;61(5 Suppl 1):S26-S31 [FREE Full text] [doi: 10.1016/j.amepre.2021.06.008] [Medline: 34686287]
25. Kenya S, Chida N, Symes S, Shor-Posner G. Can community health workers improve adherence to highly active antiretroviral therapy in the USA? A review of the literature. HIV Med 2011 Oct;12(9):525-534 [FREE Full text] [doi: 10.1111/j.1468-1293.2011.00921.x] [Medline: 21518221]
26. Cabral HJ, Davis-Plourde K, Sarango M, Fox J, Palmisano J, Rajabiun S. Peer support and the HIV continuum of care: results from a multi-site randomized clinical trial in three urban clinics in the United States. AIDS Behav 2018 Aug 06;22(8):2627-2639. [doi: 10.1007/s10461-017-1999-8] [Medline: 29306990]
27. Hoff E, Hansen L, Pulitzer Z, Campalans N, Salyards M, Muquith M, et al. A randomized control trial of a combined community health worker and re-entry intervention for people with HIV recently released from jail who use substances. J Subst Use Addict Treat 2023 Sep;152:209118. [doi: 10.1016/j.josat.2023.209118] [Medline: 37454733]
28. Senn TE, Braksmajer A, Coury-Doniger P, Urban MA, Rossi A, Carey MP. Development and preliminary pilot testing of a peer support text messaging intervention for HIV-infected black men who have sex with men. J Acquir Immune Defic Syndr 2017 Feb 01;74 Suppl 2(Suppl 2):S121-S127 [FREE Full text] [doi: 10.1097/QAI.0000000000001241] [Medline: 28079722]
29. Dillingham R, Ingersoll K, Flickinger TE, Waldman AL, Grabowski M, Laurence C, et al. PositiveLinks: a mobile health intervention for retention in HIV care and clinical outcomes with 12-month follow-up. AIDS Patient Care STDS 2018 Jun;32(6):241-250 [FREE Full text] [doi: 10.1089/apc.2017.0303] [Medline: 29851504]

XSL•FO
RenderX

30. Rana AI, van den Berg JJ, Lamy E, Beckwith CG. Using a mobile health intervention to support HIV treatment adherence and retention among patients at risk for disengaging with care. AIDS Patient Care STDS 2016 Apr;30(4):178-184 [FREE Full text] [doi: 10.1089/apc.2016.0025] [Medline: 27028183]
31. Ventuneac A, Kaplan-Lewis E, Buck J, Roy R, Aberg CE, Duah BA, et al. A mobile health intervention in HIV primary care: supporting patients at risk for ART non-adherence. HIV Res Clin Pract 2020 Oct 28;21(5):140-150 [FREE Full text] [doi: 10.1080/25787489.2020.1862972] [Medline: 33369547]
32. Grimsrud A, Bygrave H, Doherty M, Ehrenkranz P, Ellman T, Ferris R, et al. Reimagining HIV service delivery: the role of differentiated care from prevention to suppression. J Int AIDS Soc 2016 Dec;19(1):21484. [doi: 10.7448/ias.19.1.21484]
33. Haberer JE, Kahane J, Kigozi I, Emenyonu N, Hunt P, Martin J, et al. Real-time adherence monitoring for HIV antiretroviral therapy. AIDS Behav 2010 Dec 31;14(6):1340-1346 [FREE Full text] [doi: 10.1007/s10461-010-9799-4] [Medline: 20809380]
34. Wood OR, Schnall R, Kay ES, Jia H, Abua JA, Nichols TK, et al. A community health worker and mobile health app intervention to improve adherence to HIV medication among persons with HIV: the CHAMPS study protocol. BMC Public Health 2023 May 24;23(1):942 [FREE Full text] [doi: 10.1186/s12889-023-15616-9] [Medline: 37226141]
35. Batey DSR, Anderson WA, Joseph D, Musgrove KE, Mugavero MJ, Raper JL. Birmingham Access 2 Care: Community partnership to improve public health. 2015 Presented at: 2015 Engaged Scholarship Conference; Sept 29-30, 2015; State College, Pennsylvania.
36. Schnall R, Sanabria G, Jia H, Cho H, Bushover B, Reynolds NR, et al. Efficacy of an mHealth self-management intervention for persons living with HIV: the WiseApp randomized clinical trial. J Am Med Inform Assoc 2023 Feb 16;30(3):418-426 [FREE Full text] [doi: 10.1093/jamia/ocac233] [Medline: 36469808]
37. Anti-Retroviral Treatment and Access to Services (ARTAS): an individual-level, multi-session intervention for people who are recently diagnosed with HIV Implementation Manual. Academy for Educational Development Center on AIDS & Community Health (COACH). URL: https://aidsetc.org/sites/default/files/resources_files/artas_implementation_manual.pdf [accessed 2025-03-20]
38. Guidelines for the use of antiretroviral agents in adults and adolescents with HIV. Department of Health and Human Service. URL: https://clinicalinfo.hiv.gov/sites/default/files/guidelines/documents/adult-adolescent-arv/guidelines-adult-adolescent-arv.pdf [accessed 2024-04-29]
39. HIV viral load. Cleveland Clinic. URL: https://my.clevelandclinic.org/health/diagnostics/hiv-viral-load [accessed 2024-04-29]
40. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
41. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
42. Feldman BJ, Fredericksen RJ, Crane PK, Safren SA, Mugavero MJ, Willig JH, et al. Evaluation of the single-item self-rating adherence scale for use in routine clinical care of people living with HIV. AIDS Behav 2013 Jan 30;17(1):307-318 [FREE Full text] [doi: 10.1007/s10461-012-0326-7] [Medline: 23108721]
43. Johnson MO, Neilands TB, Dilworth SE, Morin SF, Remien RH, Chesney MA. The role of self-efficacy in HIV treatment adherence: validation of the HIV Treatment Adherence Self-Efficacy Scale (HIV-ASES). J Behav Med 2007 Oct 23;30(5):359-370 [FREE Full text] [doi: 10.1007/s10865-007-9118-3] [Medline: 17588200]
44. Schnall R, Cho H, Liu J. Health information technology usability evaluation scale (Health-ITUES) for usability assessment of mobile health technology: validation study. JMIR Mhealth Uhealth 2018 Jan 05;6(1):e4 [FREE Full text] [doi: 10.2196/mhealth.8851] [Medline: 29305343]
45. Lewis JR. Psychometric evaluation of the post-study system usability questionnaire: the PSSUQ. Proc Hum Factors Ergon Soc Annu Meet 1992 Oct 01;36(16):1259-1260. [doi: 10.1177/154193129203601617]
46. Improve performance, plan for productivity gains and accelerate innovation. SAS Communications. URL: https://www.sas.com/en_us/home.html [accessed 2025-03-20]
47. R Core Team. R: a language and environment for statistical computing. R Foundation. URL: https://www.r-project.org/ [accessed 2024-04-29]
48. Meyer DZ, Avery LM. Excel as a qualitative data analysis tool. Field Methods 2008 Sep 04;21(1):91-112. [doi: 10.1177/1525822x08323985]
49. Ose SO. Using excel and word to structure qualitative data. J Appl Soc Sci 2016 Aug 20;10(2):147-162. [doi: 10.1177/1936724416664948]
50. Morse JM. Critical analysis of strategies for determining rigor in qualitative inquiry. Qual Health Res 2015 Sep 16;25(9):1212-1222. [doi: 10.1177/1049732315588501] [Medline: 26184336]
51. O'Connor C, Joffe H. Intercoder reliability in qualitative research: debates and practical guidelines. Int J Qual Methods 2020 Jan 22;19:63. [doi: 10.1177/1609406919899220]
52. Konting MM, Kamaruddin N, Man NA. Quality assurance in higher education institutions: exist survey among Universiti Putra Malaysia graduating students. Int Educ Stud 2009 Feb 22;2(1):2009. [doi: 10.5539/ies.v2n1p25]

XSL•FO
RenderX

53. Sullivan GM, Feinn R. Using effect size-or why the P value is not enough. J Grad Med Educ 2012 Sep;4(3):279-282 [FREE Full text] [doi: 10.4300/JGME-D-12-00156.1] [Medline: 23997866]
54. Aschbrenner KA, Kruse G, Gallo JJ, Plano Clark VL. Applying mixed methods to pilot feasibility studies to inform intervention trials. Pilot Feasibility Stud 2022 Sep 26;8(1):217 [FREE Full text] [doi: 10.1186/s40814-022-01178-x] [Medline: 36163045]
55. O'Cathain A, Hoddinott P, Lewin S, Thomas K, Young B, Adamson J, et al. Maximising the impact of qualitative research in feasibility studies for randomised controlled trials: guidance for researchers. Pilot Feasibility Stud 2015;1:32 [FREE Full text] [doi: 10.1186/s40814-015-0026-y] [Medline: 27965810]
56. Operations manual for delivery of HIV prevention, care and treatment at primary health centres in high-prevalence, resource-constrained settings: edition 1 for field-testing and country adaptation. World Health Organization. 2008. URL: https://pubmed.ncbi.nlm.nih.gov/26290930/ [accessed 2024-04-29]
57. Dasgupta S, McManus T, Tie Y, Lin CY, Yuan X, Sharpe JD, et al. Comparison of demographic characteristics and social determinants of health between adults with diagnosed HIV and all adults in the U.S. AJPM Focus 2023 Sep;2(3):100115 [FREE Full text] [doi: 10.1016/j.focus.2023.100115] [Medline: 37790662]
58. Bennett B, Sharma M, Bennett R, Mawson AR, Buxbaum SG, Sung JH. Using social cognitive theory to predict medication compliance behavior in patients with depression in southern United States in 2016 in a cross-sectional study. J Caring Sci 2018 Mar 01;7(1):1-8 [FREE Full text] [doi: 10.15171/jcs.2018.001] [Medline: 29637050]
59. Sheeran P, Maki A, Montanaro E, Avishai-Yitshak A, Bryan A, Klein WM, et al. The impact of changing attitudes, norms, and self-efficacy on health-related intentions and behavior: a meta-analysis. Health Psychol 2016 Nov;35(11):1178-1188. [doi: 10.1037/hea0000387] [Medline: 27280365]
60. Loh KP, Liu J, Ganzhorn S, Sanabria G, Schnall R. Establishing a usability cut-point for the health information technology usability evaluation scale (Health-ITUES). Int J Med Inform 2022 Apr;160:104713 [FREE Full text] [doi: 10.1016/j.ijmedinf.2022.104713] [Medline: 35144102]
61. Banbury A, Nancarrow S, Dart J, Gray L, Parkinson L. Telehealth interventions delivering home-based support group videoconferencing: systematic review. J Med Internet Res 2018 Feb 02;20(2):e25 [FREE Full text] [doi: 10.2196/jmir.8090] [Medline: 29396387]
62. Cheng O. Utilization of telemedicine in addressing musculoskeletal care gap in long-term care patients. J Am Acad Orthop Surg Glob Res Rev 2020;4(4):e524. [doi: 10.1302/3114-210438]
63. Jong M, Kraishi M. A comparative study on the utility of telehealth in the provision of rheumatology services to rural and northern communities. Int J Circumpolar Health 2004 Dec;63(4):415-421 [FREE Full text] [doi: 10.3402/ijch.v63i4.17758] [Medline: 15709316]
64. Gillette E, Naanyu V, Nyandiko W, Chory A, Scanlon M, Aluoch J, et al. HIV-related stigma shapes research participation for youth living with HIV in Kenya. J Int Assoc Provid AIDS Care 2023 Apr 26;22:23259582231170732 [FREE Full text] [doi: 10.1177/23259582231170732] [Medline: 37101381]
65. Marellapudi A, Hussen SA, Brown DN, Fletcher MR, Henkhaus ME, Jones MD, et al. Understanding and addressing privacy and confidentiality concerns associated with the provision of mobile HIV care: a qualitative study. AIDS Care 2022 May 03;34(5):575-579 [FREE Full text] [doi: 10.1080/09540121.2021.1921104] [Medline: 33938335]
66. Taggart T, Grewe ME, Conserve DF, Gliwa C, Roman Isler M. Social media and HIV: a systematic review of uses of social media in HIV communication. J Med Internet Res 2015 Nov 02;17(11):e248 [FREE Full text] [doi: 10.2196/jmir.4387] [Medline: 26525289]
67. Audi C, Jahanpour O, Antelman G, Guay L, Rutaihwa M, van de Ven R, et al. Facilitators and barriers to antiretroviral therapy adherence among HIV-positive adolescents living in Tanzania. BMC Public Health 2021 Dec 13;21(1):2274 [FREE Full text] [doi: 10.1186/s12889-021-12323-1] [Medline: 34903209]
68. Zurbachew Y, Hiko D, Bacha G, Merga H. Adolescent's and youth's adherence to antiretroviral therapy for better treatment outcome and its determinants: multi-center study in public health facilities. AIDS Res Ther 2023 Dec 19;20(1):91 [FREE Full text] [doi: 10.1186/s12981-023-00588-y] [Medline: 38115098]
69. Hebel S, Kahn-Woods E, Malone-Thomas S, McNeese M, Thornton L, Sukhija-Cohen A, et al. Brief report: discrepancies between self-reported adherence and a biomarker of adherence in real-world settings. J Acquir Immune Defic Syndr 2020 Dec 01;85(4):454-457. [doi: 10.1097/QAI.0000000000002486] [Medline: 33136744]
70. Kagee A, Nel A. Assessing the association between self-report items for HIV pill adherence and biological measures. AIDS Care 2012 Jun 07;24(11):1448-1452. [doi: 10.1080/09540121.2012.687816] [Medline: 22670758]
71. Patel T, Ivo J, Faisal S, McDougall A, Carducci J, Pritchard S, et al. A prospective study of usability and workload of electronic medication adherence products by older adults, caregivers, and health care providers. J Med Internet Res 2020 Jun 02;22(6):e18073 [FREE Full text] [doi: 10.2196/18073] [Medline: 32348292]
72. Ping Y, Visaria A, Suppiah SD, Tan YW, Malhotra R. Prevalence and correlates of medication reminder app 'use and use intention' among older adults. Explor Res Clin Soc Pharm 2022 Jun;6:100150 [FREE Full text] [doi: 10.1016/j.rcsop.2022.100150] [Medline: 35755719]

XSL•FO
RenderX

73. Marconi A, Balestrieri M, Comastri G, Pulvirenti F, Gennari W, Tagliazucchi S, et al. Evaluation of the Abbott real-time HIV-1 quantitative assay with dried blood spot specimens. Clin Microbiol Infect 2009 Jan;15(1):93-97 [FREE Full text] [doi: 10.1111/j.1469-0691.2008.02116.x] [Medline: 19220340]
74. Pannus P, Claus M, Gonzalez MM, Ford N, Fransen K. Sensitivity and specificity of dried blood spots for HIV-1 viral load quantification: a laboratory assessment of 3 commercial assays. Medicine (Baltimore) 2016 Nov;95(48):e5475 [FREE Full text] [doi: 10.1097/MD.0000000000005475] [Medline: 27902602]
75. Vojnov L, Carmona S, Zeh C, Markby J, Boeras D, Prescott MR, DBS for VL Diagnostics Investigation Consortium. The performance of using dried blood spot specimens for HIV-1 viral load testing: a systematic review and meta-analysis. PLoS Med 2022 Aug 22;19(8):e1004076 [FREE Full text] [doi: 10.1371/journal.pmed.1004076] [Medline: 35994520]

Abbreviations

aOR: adjusted odds ratio
ART: antiretroviral therapy
CHAMPS: Community Health Worker and mHealth to Improve Viral Suppression
CHW: community health worker
DBS: dried blood spot
eROI: electronic release of information
Health-ITUES: Health Information Technology Usability Evaluation Scale
HIPAA: Health Insurance Portability and Accountability Act
HIV-ASES: HIV Treatment Adherence Self-Efficacy Scale
mHealth: mobile health
MHTAM: Mobile Health Technology Acceptance Model
PSSUQ: Poststudy System Usability Questionnaire
REDCap: Research Electronic Data Capture
SRSI: self-reported scale item

Edited by A Mavragani; submitted 25.10.24; peer-reviewed by AM Navarra, C O'Connor; comments to author 09.01.25; revised version received 03.02.25; accepted 05.03.25; published 02.04.25.

Please cite as:
Shourya S, Liu J, McInerney S, Casimir T, Kenniff J, Kershaw T, Batey D, Schnall R
A Remote Intervention Based on mHealth and Community Health Workers for Antiretroviral Therapy Adherence in People With HIV: Pilot Randomized Controlled Trial
JMIR Form Res 2025;9:e67997
URL: https://formative.jmir.org/2025/1/e67997
doi:10.2196/67997
PMID:

©Shivesh Shourya, Jianfang Liu, Sophia McInerney, Trinity Casimir, James Kenniff, Trace Kershaw, David Batey, Rebecca Schnall. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Assessing the Cultural Fit of a Digital Sleep Intervention for Refugees in Germany: Qualitative Study

Maja Blomenkamp[1], MSc; Andrea Kiesel[2], Prof Dr; Harald Baumeister[3], Prof Dr; Dirk Lehr[4], Prof Dr; Josef Unterrainer[5], Prof Dr; Lasse B Sander[5], PhD; Kerstin Spanhel[5], PhD

[1]Institute of Psychology, University of Freiburg, Freiburg, Germany

[2]Cognition, Action, and Sustainability Unit, Institute of Psychology, University of Freiburg, Freiburg, Germany

[3]Department of Clinical Psychology and Psychotherapy, Institute of Psychology and Education, Ulm University, Ulm, Germany

[4]Health Psychology and Applied Biological Psychology, Institute of Sustainability Psychology, Leuphana University Lüneburg, Lüneburg, Germany

[5]Institute of Medical Psychology and Medical Sociology, Faculty of Medicine, University of Freiburg, Freiburg, Germany

Corresponding Author:
Kerstin Spanhel, PhD
Institute of Medical Psychology and Medical Sociology
Faculty of Medicine
University of Freiburg
Hebelstr. 29
Freiburg, D-79104
Germany
Phone: 49 761 203 5530
Email: kerstin.spanhel@mps.uni-freiburg.de

Abstract

Background: Digital interventions have been suggested to facilitate access to mental health care for refugees, who experience structural, linguistic, and cultural barriers to mental health care. Sleep-e, a digital sleep intervention originally developed for German teachers, has been culturally adapted for refugees in Germany mainly coming from African and Middle East countries. With the increasing number of refugees from Ukraine and the associated diversity of cultural backgrounds among refugees in Germany, it is essential to assess whether existing digital interventions are culturally appropriate for this target group as well.

Objective: The study aimed to investigate the perceived cultural appropriateness of Sleep-e in both its original and culturally adapted versions among refugees in Germany, hereby exploring and possibly contrasting the needs of refugees coming from Ukraine and other countries of origin.

Methods: Overall, 13 refugees (6 from Ukraine, 23-66 years old; and 7 from other countries, 26-41 years old) participated in the study. Each participant went through parts of the original or culturally adapted version of the digital sleep intervention, with 5 participants going through both versions. A total of 17 semistructured interviews (11 for the adapted, 6 for the nonadapted intervention version) and 9 think-aloud sessions (6 for the adapted, 3 for the nonadapted intervention version) were conducted to assess cultural appropriateness, suggestions for adaptations, and perceived relevance. Data were transcribed, categorized, and analyzed using structured qualitative content analysis.

Results: The findings showed key differences in the perceived appropriateness and identification between the 2 refugee groups and the 2 intervention versions. Ukrainian participants expressed positive (n=70) and negative (n=56) feedback on the adapted intervention version, which revealed identity conflicts, as the adapted intervention version was targeted at a refugee population with whom they could not fully identify (18 negative feedback quotes concerning the refugee example characters). Whereas they identified with the European context in the original version, they found the problems described less relevant to their experiences. In contrast, participating refugees from other countries found the culturally adapted version more comprehensible and culturally appropriate (55 positive and 5 negative feedback quotes). No significant usability issues were reported, but several participants highlighted the need for an individualization of the intervention content.

Conclusions: Neither the original nor culturally adapted version of the digital sleep intervention fully met the needs of all refugee groups, highlighting the complexity of culturally adapting digital interventions for this population. Particularly, the identity conflict of participating Ukrainian refugees regarding the refugee context suggests that adaptation should go beyond regional considerations and consider the dynamics of social identity. These findings emphasize the relevance of including co-design

XSL•FO
RenderX

processes with different refugee populations to ensure broad identification and, herewith, cultural appropriateness of digital interventions.

Trial Registration: German Clinical Trials Register DRKS00036484; https://drks.de/search/de/trial/DRKS00036484

(JMIR Form Res 2025;9:e65412) doi:10.2196/65412

KEYWORDS

Ukraine; eHealth; sleep disturbances; low-threshold treatment; culturally sensitive treatment; refugee; digital sleep; Germany; digital intervention; interview; content analysis; qualitative study; mental burden; mental health care; electronic health; digital health

Introduction

The number of people forced to seek refuge in new countries is increasing worldwide [1]. These people, hereafter referred to as refugees, are exposed to multiple stressors before, during, and after leaving their home countries, and they therefore experience increased psychological vulnerability and prevalence of mental disorders [2,3], including sleep problems [4]. However, there are major gaps in mental health care for refugees in the receiving countries such as Germany [5,6], resulting from a lack of (culturally sensitive) services, language barriers, differences in help-seeking behavior and residence status, as well as stigmatization, among other barriers [7-9]. To overcome these barriers, culturally adapted digital interventions focusing on transdiagnostic or somatically perceived symptoms are being discussed and investigated as low threshold offers for refugees [10,11]. In the general population, digital interventions show good effectiveness for a wide range of mental health conditions [12,13]. Their usability appears to be essential in this context [14,15]. When offered to refugees, in addition to intervention usability, cultural adaptation is suggested to be essential for the acceptance and effectiveness of face-to-face and digital interventions, recognizing the specific burdens associated with the refugee context [16-18]. In this context, the question of the cultural specificity of the adaptations is becoming increasingly important [19]. The target group of refugees is characterized by a diversity of countries of origin that is constantly changing due to events in the global political situation, such as the start of the war in Ukraine in 2022. To increase refugees with diverse backgrounds using and profiting from digital interventions, these interventions need to be provided in a culturally sensitive (ie, with an openness to diverse backgrounds)—in contrast to a culturally specific (ie, separately for diverse backgrounds)—way [20]. Yet, so far, there are only a few digital interventions developed for refugees from diverse countries of origin [21-23].

Based on feedback from refugees in Germany, who at the time were mainly coming from African and Middle Eastern countries of origin, the digital sleep intervention Sleep-e was culturally adapted for this target group with the aim of making it accessible to people sharing the experience of flight. Thereby, (1) factors related to the flight (eg, separation from the family and the home country) and (2) factors related to the habits, values, and treatment concepts of their non-Western countries of origin were considered [24]. The culturally adapted version of the digital sleep intervention was accepted by refugees from African and Middle Eastern countries [22]. Based on the increasing number of refugees in Germany coming from Ukraine since the beginning of the war in Ukraine in 2022 [25], it seems essential to make such interventions available to this target group as well. Ukrainian refugees are shown to differ from refugees from other countries of origin, among others, in terms of sociodemographic aspects (age and gender), sociocultural aspects, differences in the flight route, legal rights, and welcoming culture upon arrival in Germany [26-29]. Building on the named 2 previous studies [22,24], this qualitative study aimed to investigate differences in the perceived cultural appropriateness and the cultural needs in regard to the digital sleep intervention among refugees from Ukraine and refugees from other, predominantly African and Middle Eastern countries of origin. A culturally adapted version and a nonadapted version of the intervention were used, and factors influencing the perceived cultural appropriateness were additionally investigated.

Methods

Recruitment and Participants

This study was conducted in 2 subsamples: refugees from Ukraine and refugees from countries other than Ukraine. Recruitment was carried out from February to April 2023 via various institutions and associations working in the refugee and migration context in Germany, mainly Freiburg and Karlsruhe. Many recruitment inquiries remained unanswered (eg, integration and language courses, migration offices, advice centers, and advertising via social media). Recruitment mainly succeeded via Refugium Freiburg (psychosocial advice for refugees), the German-Ukrainian Choir Karlsruhe, the University of Freiburg, the radio project “Our Voice” for refugees in Freiburg, and “Kontaktcafé Freiburg” (get-together café for refugees), as well as via personal contacts. Institutions were contacted consecutively face-to-face, via phone, or via email. They forwarded a flyer with information on the study to their clients face-to-face. People interested in study participation could contact the study team via email or phone and were then forwarded the detailed study information. Inclusion criteria were checked, which included the experience of flight, age of majority, and sufficient knowledge of German or English. No cutoff for insomnia severity was used as an inclusion criterion in order to also reach refugees with subclinical symptoms, who could be a key target group for the transdiagnostic and low-threshold digital sleep intervention. In addition, this allows information to be obtained on who might be interested in the intervention, regardless of existing clinical symptoms. Contacted participants who did not want to participate in the study mainly reported reasons such as language barriers or difficulties in

XSL•FO
RenderX

finding the time for study participation. None of the participants dropped out after signing the informed consent.

There were 6 participants from Ukraine (23-66 years old), and 7 participants coming from Afghanistan, Gambia, Syria, Madagascar, Cameroon, Ecuador, and Eritrea (26-41 years old), referred to below as refugees from other countries. On average, the participants revealed subthreshold sleep problems. Detailed sociodemographic data are shown in Table 1.

XSL•FO
RenderX

Table 1. Sociodemographic data and information on the conduction of the qualitative study in Germany in 2023 are illustrated for each participating refugee from Ukraine (ukr) or other (oth) countries of origin.

Subgroup and ID	Home country (gender)	Education	Age (years)	Years in Germany	Psychotherapy (current or past)	Insomnia severity[a]	Tested modules	Study mode (study language)	Duration (min)	Codes, n
Ukraine										
01_ukr	Ukraine (female)	High school	45	1	No	8	na[b]1: qtn[c], int[d] a[e]1: ThA[f], int	In person (German)	79	154
02_ukr	Ukraine (female)	High school	45	1	No	15	na2: qtn, ThA, int a2: int	In person (German)	53	57
03_ukr	Ukraine (female)	High school	66	1	No	7	na3: qtn, ThA, int a3: int	In person (English)	57	105
04_ukr	Ukraine (female)	High school	26	1	Yes	15	a2: qtn, ThA, int	Video call (English)	42	65
05_ukr	Ukraine/Nigeria (male)	High school	23	1	No	15	a3: qtn, ThA, int	In person (English)	63	108
06_ukr	Ukraine (female)	High school	33	0.5	Yes	19	a4: qtn, int	Video call (English)	15	42
Total, mean (SD)	—[g]	—	39.7 (15.9)	0.9 (0.2)	—	13.2 (4.7)	—	—	—	—
Other countries										
07_oth	Afghanistan (male)	High school	41	8	No	9	na1: qtn, int	Video call (German)	23	18
08_oth	Gambia (male)	No degree	27	3	Yes	18	na2: qtn, int a3: int	In person (English)	45	71
09_oth	Syria (male)	High school	26	7.5	Yes	8	na3: qtn, ThA, int a4: ThA, int	In person (German)	74	125
10_oth	Madagascar (male)	High school	30	10	Yes	18	a1: qtn, ThA, int	In person (French)	37	31
11_oth	Cameroon (female)	High school	33	1.5	No	6	a2: qtn, int	Video call (French)	35	13
12_oth	Ecuador (male)	High school	32	0.5	No	15	a3: qtn, ThA	In person (German)	33	18
13_oth	Eritrea (male)	No degree	31	8	Yes	12	a4: qtn, int	Video call (German)	11	23
Total, mean (SD)	—	—	31.4 (4.9)	5.5 (3.7)	—	12.3 (4.9)	—	—	—	—

[a]Insomnia Severity Scores [30]: 0-7, not clinically significant; 8-14, subthreshold; 15-21, clinical (moderate); 22-28, clinical (severe).

[b]na: nonadapted version of the digital sleep intervention (modules 1-3).

[c]qtn: questionnaire.

[d]int: interview.

[e]a: adapted version of the digital sleep intervention (modules 1-4).

[f]ThA: think-aloud.

[g]Not applicable.

XSL•FO
RenderX

Procedure

Data collection took place in one-to-one sessions (approximately 90 minutes of video-based or in-person, one-to-one meetings at the Institute of Medical Psychology and Medical Sociology, University of Freiburg, Germany) with MB, who was in the final year of her master's degree in clinical psychology at the time of study conduction and had previous experience in qualitative research (refer to Table 1 for further information). MB had previous contact with one participant (01_ukr); the other participants were unknown to her. The participants were informed that the research was part of MB's master thesis. Sessions were conducted in German, English, or French and were audio recorded. All participants completed a web-based questionnaire (sociodemographic data, Insomnia Severity Index with 7 Likert-scale items 0-4 [30], and Client Satisfaction Questionnaire adapted to internet-based interventions with 8 Likert-scale items 1-4 [31]). To assess the perceived cultural appropriateness of the respective interventions as part of the usability test, it was considered sufficient to work through individual modules of the adapted versus nonadapted intervention: many aspects relevant to the assessment are similar across the respective adapted versus nonadapted modules (eg, number of pictures, use of text versus video, and example characters with a German versus another cultural background). Following this assumption, the participants worked through one adapted or nonadapted module. They were then offered the opportunity to test a second (nonadapted or adapted) module. Three participants from Ukraine and 2 participants from other countries completed both an adapted and a nonadapted module. When assigning the modules, it was ensured that each module was seen at least once by participants from each subgroup. After each module completion, semistructured interviews were conducted in order to assess the participants' experiences with the content and design of the tested module, its appropriateness for themselves and refugees in general, and their own usage behavior (refer to Multimedia Appendix 1 for the interview guideline; as the interview guideline was shown to be appropriate in a first pilot interview, the pilot was included in the final results). The think-aloud method [32] was additionally used with 8 participants (5 participants from Ukraine, and 3 from other countries), depending on the language skills and time resources of the respective participants. Using this method, the participants were asked to speak out loud all of the thoughts they had while they went through the modules; for example, they could comment on the experienced design, functionality, or content of the intervention. Throughout data collection, the investigator made observational notes on verbal or nonverbal characteristics during the test. Table 1 shows in detail which modules were processed by which participants and in which format (interview or think-aloud) the data were collected.

Intervention

Two versions of the digital sleep intervention Sleep-e were used, which is a digital, brief cognitive behavioral intervention for sleep problems available in English and German. The intervention was provided on the eHealth platform eSano. The nonadapted version [24] is an unguided digital intervention consisting of 3 content-separated, consecutive modules based on the intervention GET.ON Recovery [33,34], which was developed and evaluated for German employees with sleep problems. Adaptations were conducted based on feedback from health care professionals working with refugees as well as refugees from Algeria, Eritrea, Iran, Iraq, and Syria [24], complemented by previously conducted cultural adaptations of digital interventions [35]. The resulting culturally adapted version consists of 4 content-separated, consecutive modules and was aimed at being suitable for refugees from diverse countries of origin in Germany, with the idea of a shared experience of pre-, peri-, and postmigration stressors. Its feasibility, acceptance, and preliminary effectiveness were investigated among refugees mainly coming from African and Middle East countries [22]. Table 2 provides information on the content of both versions of the intervention. Information on the adaptations suggested [24] and conducted [22] can be found in the respective articles.

XSL•FO
RenderX

Table 2. Overview of the modules of both an adapted and a nonadapted version of a digital Sleep-e intervention tested in the qualitative study with refugees in Germany.

Module	Content of the nonadapted version [24]	Content of the culturally adapted version [22]
Module 1	• Introduction of example characters from Germany ("text-based"[a] with photo) • Psychoeducation and exercise on sleep hygiene rules ("text-based", quiz) • Psychoeducation on sleep medication ("text-based") • Optional: sleep diary ("offline document")	• Introduction by a health expert and for example characters from Arabic-speaking countries ("video-based") • Psychoeducation on sleep problems, link to "migration" ("video-based") • Optional: relaxation exercise (audio-based) • Optional: sleep diary ("short web-based document")
Module 2	• Psychoeducation on the topic of rumination ("eg, explanatory video") • "3" exercises on the topic of rumination (audio- and text-based)	• Psychoeducation and "optional" exercise on sleep hygiene ("mainly video-based") • "Optional": psychoeducation on sleep medication ("text- and video-based") • "Optional": relaxation exercise ("audio-based") • "Optional": sleep diary ("short web-based document")
Module 3	• Reflection on the results, consolidation of what has been learned (text-based) • Planning of future exercises ("text-based") • Information on further offers of help ("text-based") • Psychoeducation on problems associated with poor sleep ("text-based")	• Psychoeducation on the topic of rumination, focus on migration topics (mainly video-based) • "Optional": 2 exercises on the topic of rumination (audio- and text-based) • "Optional": exercise on sleep hygiene rules • "Optional": relaxation exercise (audio-based) • "Optional": sleep diary
Module 4	—[b]	• "Optional": psychoeducation on problems associated with poor sleep (mainly video-based) • Information on further offers of "refugee-specific" help • "Optional": exercise on sleep hygiene rules • "Optional": relaxation exercise (audio-based) • "Optional": reflection on results • "Optional": sleep diary

[a]Differences between the adapted and the nonadapted version of the digital sleep intervention are in double quotes.

[b]Not applicable.

Data Analysis

Before analyzing data, the authors involved in the analysis (MB and KS) reflected on their own assumptions regarding the topic of the study. The following assumptions were identified: (1) refugees are often not well off and are grateful for any form of support and (2) Ukrainian refugees and refugees from other countries of origin differ in their needs. The participants were not involved in data transcription, analysis, or interpretation.

The collected data comprised 26 audio recordings (9 think-aloud sessions, including one participant who did 2 sessions; 17 interviews, including 5 participants who did 2 interviews). The recordings were automatically transcribed using Adobe Premiere (2021) and postprocessed by MB based on the transcription rules for computer-assisted analysis according to Kuckartz and Rädiker [36]. The data were analyzed using structured qualitative content analysis (primary analysis [36,37]) with a combination of content-related, evaluative, and contrasting categories [38]. MAXQDA 2022 [39] software was used to code a total of 830 text passages. The categories were created and continuously adapted by MB using an inductive and deductive hybrid approach [38], primarily based on the transcripts of the Ukrainian participants. This enabled the deductive inclusion and validation of theories and questions derived from the two previous studies, as well as the inductive inclusion of new aspects (eg, in regard to the war in Ukraine) introduced by the participants. After coding half of the transcripts (ie, 13), the new information could be integrated into the previously formed categories and no further changes were made to the structure of the category system, which indicates data saturation. In order to assess intercoder reliability and ensure the comprehensibility of the category definitions, 2 transcripts were additionally coded and discussed by KS (researcher at the time of study conduction). The different category formats were used to identify possible group differences (contrasting categories), evaluations (evaluative categories), and factors influencing the evaluations (content-related categories), as well as their interactions. As part of the secondary analysis, valence analyses [37] were used to compare the ratings of the intervention parts based on the evaluative categories (positive vs negative feedback regarding the adapted vs nonadapted intervention or regarding refugees from Ukraine vs from other countries of origin). In addition, the coding frequencies by categories, subgroups, and interventions were compared in frequency analyses, and overlaps between evaluative categories and contrasting or content-related categories were examined in group comparison. Such frequency analyses can provide insight into possible correlations but are of limited scalability due to the lack of objectivity of the number of codes. The qualitative results are shown along the developed category system, underpinned by quotes from the participants. The references of the quoted text passages comprise the number

XSL•FO
RenderX

of participants (01-13), the subgroup (coming from Ukraine [denoted as "ukr"] vs from other countries of origin [denote at "oth"]), the modules processed (modules 1-4 of the adapted version of the intervention [denoted from "a1" to "a4"], modules 1-3 of the nonadapted version of the intervention [denoted from "na1" to "na3"]), the type of assessment (think-aloud [denoted as "ThA"], and interview [denoted as "int"]) and the position in the transcript (denotes as "pos."). The quantitative data from the questionnaires was analyzed descriptively.

Ethical Considerations

The study was registered in the Freiburg Clinical Trials Registry (registration FRKS004288). Ethics approval was granted by the ethics committee of the University of Freiburg, Germany (22-1452-S2). The study was conducted and documented in accordance with the guidelines for research with refugees [40] and the COREQ (Consolidated Criteria for Reporting Qualitative Research) [41] (refer to Multimedia Appendix 2 for the checklist). Interested participants received detailed digital, written information on the study procedure, aim, and their rights. Afterward, they provided their digital informed consent before the start of the study or had the option of withdrawing study participation without any consequences. The protection of the privacy and confidentiality of the participants was ensured by pseudonymizing all collected data with an ID. Only the pseudonymized data were used during the study, and the data were only accessible to selected project members. A reference list, which linked personal information with the ID, was deleted at the end of the data collection. This means that all data collected (transcripts or questionnaires) are de facto anonymized, and it is not possible to draw any conclusions from these data about the individual study participants. The de facto anonymized data of the participants are stored by the researchers on secure servers for 10 years after publication and then deleted. The participants had the chance to win one of five €20 (€1=US $1.10) gift vouchers as compensation for their efforts.

Results

Overview

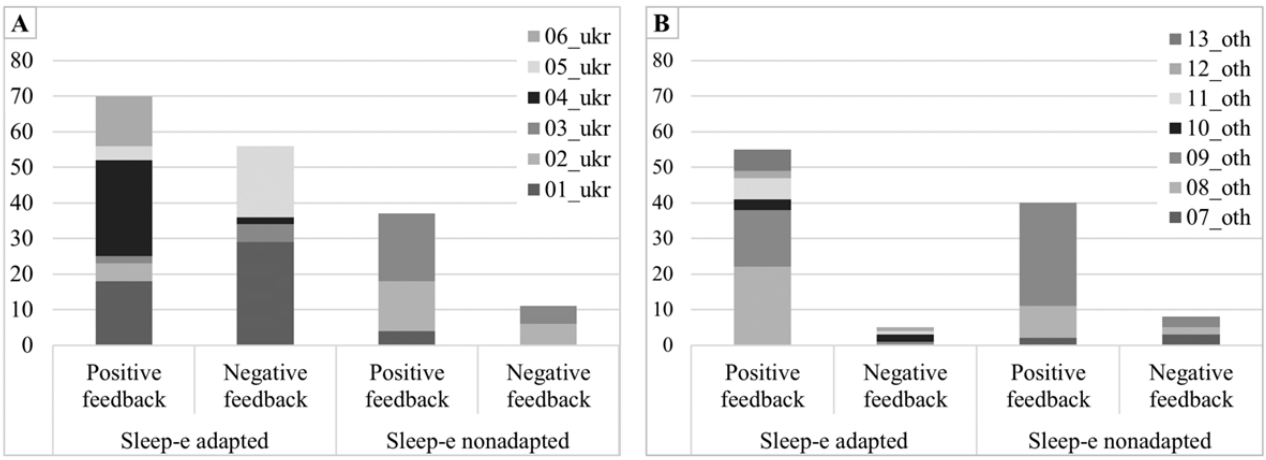
The developed category system (refer to Multimedia Appendix 3) resulted in 6 main and 14 subcategories. In total, 4 main categories informed the first research question and contained information on the perceived cultural appropriateness and the cultural needs among the participating refugees from Ukraine versus those from other countries of origin regarding the digital sleep intervention: (1) concept: adapted version, (2) concept: nonadapted version, (3) comparison: adapted and nonadapted version, and (4) use of the intervention. The developed category system furthermore resulted in 2 main categories informing about factors that influenced the perceived cultural appropriateness and needs of the participating refugees regarding the digital sleep intervention: (1) identity and (2) health care.

Perceived Cultural Appropriateness of the Digital Sleep Intervention

Overview

Across both versions of the intervention and both subgroups, there was more positive (n=202) than negative feedback (n=80; see Figure 1).

Figure 1. Frequency analyses of positive and negative feedback given by the refugees from (A) Ukraine and (B) other countries of origin participating in the qualitative study in Germany. The numbers of positive and negative feedback are illustrated separately for the adapted version and the nonadapted version of the digital intervention Sleep-e tested in the study. oth: other countries of origin; ukr: Ukraine.



> *You can take a lot from it also. And you can learn easily also.* [08_oth_a3(na2)_int, pos. 211]
>
> *To have the training..., it's much easier. Yeah, and it could solve your problem.* [06_ukr_a4_int, pos. 60]

An overall positive attitude also was revealed in the relatively high ratings in the client satisfaction questionnaire adapted to internet-based interventions, with an overall mean score of 26.0 (SD 4.7) and similar mean scores for the different versions or subgroups.

Adapted Intervention Version

The adapted version of the intervention was rated positively in both groups, but the feedback from the Ukrainian participants revealed a rather heterogeneous picture, with 70 positive and 56 negative feedback quotes, contrasting with 55 positive and 5 negative feedback quotes among participants from other countries. With regard to the conception of the intervention, linguistic simplifications (eg, subtitles), the video formats and quizzes, as well as good structuring and feasibility were

XSL•FO
RenderX

considered particularly helpful. Both groups criticized the quality of the videos as well as linguistic aspects such as the length of the texts and modules or language difficulties. As for the content of the intervention version, the experience reports from the example characters (refugees) in particular were described as motivating and interesting and the exercise content as helpful. Differences in the perceived appropriateness of the content between the 2 subgroups were found with regard to the identification with the example characters: whereas the participating refugees from other countries of origin were able to recognize and name similarities with the example characters, the participants from Ukraine did not identify with the example characters and described them as untrustworthy or boring (“I don't believe them that the exercises help.” [01_ukr_a1na1_int, pos. 102, translated]). In addition, the participants from Ukraine criticized the strong focus on the topic of flight, too negative content, and too little concrete advice. Suggestions for changes to the adapted version of the intervention included ideas for increasing identification, for example, through the opportunity to create a personalized profile with one’s own needs and issues (“Maybe bubbles and then it’s like: hit which bubble you feel speaks to you.” [05_ukr_a3_int, pos. 336]; “Create your own profile and then really see: voilà, that’s me.” [10_oth_a1_(ThA)int, pos. 147, translated]). According to the participants (especially those from Ukraine), this could also be achieved through a greater variety or abstractness (eg, by using graphics instead of real videos) of the example characters.

Nonadapted Intervention Version

The nonadapted version of the intervention was also rated positively in both groups (37 positive and 11 negative feedback quotes from the Ukrainian participants; 40 positive and 8 negative feedback quotes from participants from other countries). With regard to the conception of the intervention, its simple design and the clear, motivating, and comprehensible structure were mentioned. Both groups criticized the difficult language and an oversupply of text-based information. As for the content of the intervention version, the experience reports of the example characters (Germans) and the exercises and questions were described as helpful, as was the combination of theoretical and practical aspects (“So the first [nonadapted] version was technical for me. Going to see a therapist and answering questions I could fully understand and that was professional I would say.” [01_ukr_a1na1_int, pos. 6, translated]). However, the content of this version of the intervention was partly considered too scientific or insufficient for more serious problems by the participants from Ukraine (“I don’t see here any problems, serious problems. Just, time to bed, stop having alcohol, stop working after 9. But people who are under stress, need the more serious program.” [03_ukr_na3_int, pos. 123]), whereas the participants from other countries of origin mainly mentioned identification difficulties (“Except for alcohol and so on, because I'm a Muslim and I don’t drink alcohol. Yes, and the norm in the Arab-Islamic world is completely different. It’s practically a sin, it’s an absurdity, alcohol is for losers or sinners, or I don’t know, and not for normal people.” [09_oth_na3a4_int, pos. 57, translated]). Some statements of the Ukrainian participants did not include themselves when pointing out possible limitations of the cultural appropriateness for refugees (“But not all people can read so much text at once and the language is also a bit too scientific, like from a psychological book [...] For me it was okay.” [02_ukr_na2_ThA_int_a3_int, pos. 39, translated]). Suggestions for changes to the nonadapted version of the intervention included a more solution-oriented content and the presentation of more appropriate problems or situations (“I would yes, more different [example characters], for example in terms of age.” [07_oth_na1_int, pos. 111, translated]).

Factors Influencing the Perceived Cultural Appropriateness

As a main factor influencing the perceived cultural appropriateness of the intervention, identification was extracted. In this context, the participants emphasized the need to enhance identification with the intervention by including possible differences in the refugee contexts as well as cultural and linguistic differences in the intervention content.

Identification With the Refugee Context

Factors related to the refugee background and the identification with it were shown to influence the perceived appropriateness of the interventions among both groups of participating refugees. Some of these factors were related to their own flight experiences (“Especially when they came from Ukrainian zones after the bombing, after devastation, ruin, these bad things. They ask to get psychological treatment or to see a psychiatrist because they realize they need it.” [03_ukr_na3_int, pos. 72]; “That is another traumatic experience. Knowing that crossing the sea, some people die, some people survive.” [08_oth_a3(na2)_int, pos. 153]).

Yet, a separation of the participating Ukrainian refugees from being a refugee became apparent, revealing an underlying identity conflict (identity of being a refugee): “They don't identify as refugees, so in a way that [experience reports from the adapted version of the intervention] would even be a disconnect for them.” [05_ukr_a3_int, pos. 246]; “Just I lived in prosperity [...] I have everything. And I fled from Ukraine the second day of war and didn’t see all those bombings, devastations, and so on and so forth.” [03_ukr_a3_(ThA)int, pos. 11]. Along with this, the Ukrainian participants emphasized that they did not want to be confronted with the topic of flight (“If I were a refugee, I wouldn't want to focus on it...You have to forget that you are a refugee.” [01_ukr_a1na1_int, pos. 243, translated]).

Along with the identification as a refugee, the question of how far the identity of being a refugee is compatible with the Ukrainian identity was revealed, and whether refugees from Ukraine can identify with refugees from other countries of origin (same vs different identity). Thereby, statements about the same versus different identities were repeatedly raised (Textbox 1).

XSL•FO
RenderX

Textbox 1. Statements by the participating refugees in the qualitative study conducted in Germany about the same versus different identities of refugees from Ukraine (ukr) and refugees from other countries of origin (oth).

- "You have a world with Ukrainians inside who do not just view fleeing the same, you know, it's the same war, but at the same time, it is not the same (laughs)." [05_ukr_a3_int, pos. 252]
- "There are universal factors of flight and war." [09_oth_na3a4_int, pos. 306, translated] versus "It is never the same experience." [05_ukr_a3_int, pos. 300]
- "The people that came from the other places also, they face even more difficulties than the people that came from Ukraine." [08_oth_na2a3_ThA_int, pos. 353] versus "First-degree refugees, second-degree refugees...as if they were now downgraded, that is, with the refugees who are now less privileged." [09_oth_na3a4_int, pos. 342, translated]
- "They have a different pronunciation, different skin color, different way of speaking and different challenges than we do." [09_oth_na3a4_I, pos. 336, translated] versus "They are also refugees and they have had all these problems." [01_ukr_a1_T, pos. 432, translated]

The dilemma of same versus different identity also became evident in various statements regarding a possible identification of the participating refugees from Ukraine with the example characters in the adapted version of the intervention: "When I was listening to these people, I did not, I do not know their situation." [03_ukr_a3_(ThA)int, pos. 74] versus "I could relate. I understand his problem and I think that I have the same problem." [04_ukr_a2_ThA, pos. 127] versus "Oh, [for refugees] from Ukraine. No." [05_ukr_a3_int, pos. 240]. Various conflicting issues were found in their statements about the example characters (Table 3).

Table 3. Statements of the participating refugees from Ukraine can illustrate an ambivalent view on the example characters of the adapted version of the digital sleep intervention tested in the qualitative study in Germany.

Affirming perspective		Contrasting perspective
"To be honest, it is better for refugees." [01_ukr_a1na1_int, pos. 6, translated]	↔	"Not for me. [laughs] For me, I don't know, I understand that it's a completely different culture." [01_ukr_a1_ThA, pos. 432, translated]
"Of course, I don't have time to have this diary, because it's for people who are under stress and under problems...who have serious problems and they want to recover. They will do that." [03_ukr_a3_(ThA)int, pos. 116]	↔	"Hm no, for me, stress is the problem number one...Too much stress, for sure. It can have a negative impact on the sleeping quality. Yeah, you can't sleep because you think how to solve this problem." [03_ukr_na3_int, pos. 18 // 03_ukr_na3_ThA, pos. 189]
"You know, it [experience reports from refugees]'s something you can relate to, you know." [05_ukr_a3_int, pos. 234]	↔	"Even though I did not flee to Europe, like through the Mediterranean or the Sahara, fleeing the war in Ukraine, I don't really think it speaks to me." [05_ukr_a3_int, pos. 198]
"And besides: Being a refugee is difficult, it is always difficult." [01_ukr_a1_ThA, pos. 170, translated]	↔	"No, for me, no, for me, everything is easy to talk about. I don't have such or deep problems." [01_ukr_a1na1_int, pos. 165, translated]

Related to this, frequency analyses of feedback from the participating refugees from Ukraine revealed 9 positive feedback quotes and 18 negative feedback quotes on the example characters in the adapted intervention version. In contrast, the participating refugees from other countries of origin seemed to identify strongly with the example characters in the adapted version of the intervention and did not mention any negative criticism of them, but gave 16 positive feedback quotes. Feedback on the example characters in the nonadapted intervention version was more homogeneous, with both 4 positive and negative feedback quotes given by the participating refugees from Ukraine and 3 positive and 1 negative feedback quotes given by the participating refugees from other countries of origin.

Identification With the Country of Origin

Other factors related to the flight background that influenced the perceived appropriateness of the intervention versions were associated with the identification with the respective country of origin versus the identification with Germany. This goes back to the fact of coming from non-German countries of origin, which was linked to revealed cultural or linguistic differences.

> *Because I'm Muslim and I don't drink alcohol.* [09_oth_na3a4_int, pos. 57, translated]
>
> *Nope, so the psychologists work according to models and views and worldviews that don't match mine.* [09_oth_na3a4_int, pos. 243, translated]

Also, a better linguistic understanding (native language, linguistic facilitation such as subtitles or audios, etc) and the consideration of different reading comprehension levels seemed to increase the perceived cultural appropriateness: "And not every refugee also can read. But having access to audible information can be helpful." [08_oth_a3(na2)_int, pos. 137]. As such, the German language was perceived as stressful by one participant: "If I speak German, I'm tense. Yes, because it has to do with work, with problems, with documents and so on." [01_ukr_a1_ThA, pos. 359, translated]. Whereas the participants from other countries of origin named both cultural and linguistic differences and identification issues, the Ukrainian participants solely named linguistic differences.

Health care

Besides the refugee identity, the factor "health care" was identified to be linked to the perceived cultural appropriateness of the digital sleep intervention. Differences in the needs and

XSL•FO
RenderX

the use of mental healthcare became evident, revealing the importance of familiarity on the one hand and professionalism on the other hand.

> *I would opt for a European doctor.* [01_ukr_a1na1_int, pos. 237, translated]
>
> *That [nonadapted version] was professional I would say.* [01_ukr_a1na1_int, pos. 6, translated]
>
> *If you have problems, you'd rather go to mother, yes, [...] something like relatives, so your nation.* [01_ukr_a1_ThA, pos. 491, translated]

Although psychotherapy was partly described as "trendy" [02_ukr_na2_int, pos. 108, translated], stigmatization of mental health care became apparent: "Where I come from, it's very taboo and really inaccessible." [10_oth_a1_(ThA)int, pos. 129, translated]; this was also on a subconscious level (eg, observation from participant 11 [oth] when filling out the sociodemographic questionnaire: "Psychotherapy? Ah no no no!" [translated]).

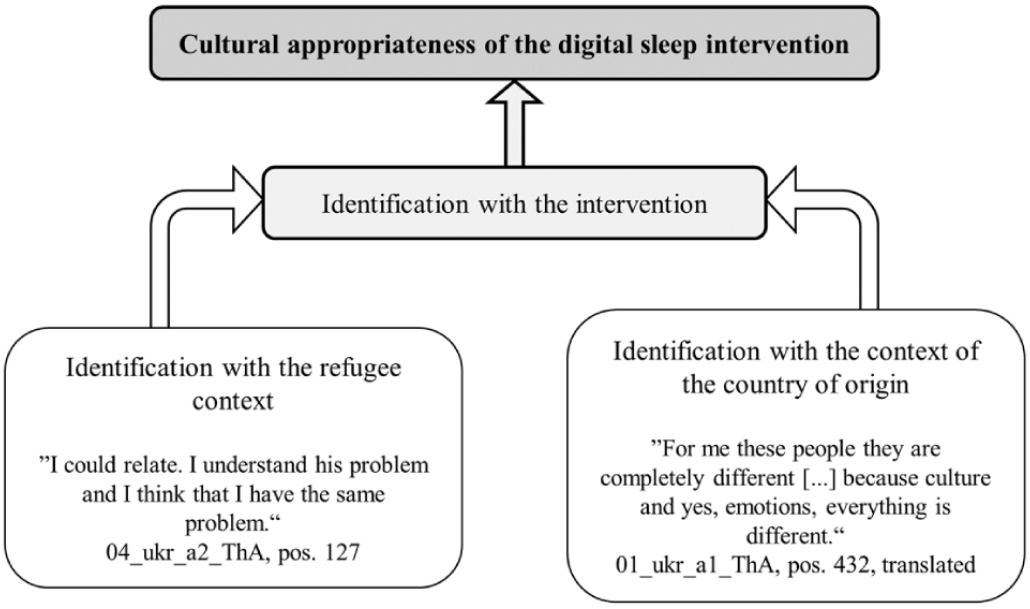
Discussion

Principal Findings

Overview

This study for the first time compared the cultural appropriateness perceived by refugees from Ukraine and other countries of origin regarding a digital intervention, specifically the original (developed for Germans) and the culturally adapted (adapted for refugees coming from African and Middle Eastern countries) versions of a digital sleep intervention. Overall, the participating refugees showed a positive attitude toward the intervention, yet the perceived cultural appropriateness regarding the 2 versions seemed to differ between the participants from Ukraine versus those from other countries of origin. The identification with the intervention content regarding the refugee context as well as the context of their countries of origin seemed to play a crucial role (refer to Figure 2). Identification with the refugee context refers to the extent to which participants see their own flight as a part of themselves and could relate to these issues in the intervention. Identification with the country of origin refers to the extent to which the participants identify with the behaviors, values, and norms of their country of origin and could relate to the behaviors, values, and norms illustrated in the intervention. Thereby, identity conflicts among the Ukrainian participants regarding the refugee context were identified as factors reducing the perceived cultural appropriateness of both intervention versions.

Figure 2. The role of identification in the cultural appropriateness of the digital intervention Sleep-e tested in the qualitative study in Germany, as perceived by the participating refugees from Ukraine and other countries of origin. a: adapted version of the intervention; pos: position in the transcript; ThA: think-aloud; ukr: Ukraine.



Differences in the Perceived Cultural Appropriateness of the Digital Sleep Intervention

Both versions of the digital sleep intervention received more positive than negative feedback, and feedback on less appropriate aspects mainly concerned design and language aspects (eg, including too much text and hard-to-understand language), which is rather general feedback on digital interventions than a refugee-specific point of view [42-44]. Regarding the participating refugees from countries other than Ukraine, the findings mainly corresponded to previous findings: whereas a lack of identification with the nonadapted version was shown [24], the adapted version seemed to be perceived as appropriate in terms of its content [22], and no negative feedback on the intervention content was given. Regarding the participating refugees from Ukraine, negative feedback on the intervention content was mainly given concerning the adapted version of the intervention and comprised a perceived overfocus

XSL•FO
RenderX

on the topic of flight, in particular regarding the example characters. Concerning the nonadapted version, less negative feedback indicated a perceived inappropriateness of the illustrated life situations and problems of the example characters, which related in particular to a high workload. Thus, both versions of the intervention appeared to be only partially appropriate for the participants from Ukraine.

Identification Influences the Perceived Cultural Appropriateness of the Digital Sleep Intervention

The lower perceived cultural appropriateness among the Ukrainian participants regarding the adapted version of the intervention was suggested to be linked to their lower identification with the intervention: Differences between the participants from Ukraine versus those from other countries of origin became apparent with regard to the identification with the refugee context and aspects related to the identification with the own country of origin. This resulted in differences in the identification with, for example, the example characters in the interventions. The revealed differences correspond to differences between refugees from Ukraine and refugees from other countries of origin suggested in the literature: Although Ukrainian refugees and refugees from other countries of origin share the experience of having had to flee, they differ in various other aspects, such as the welcome culture in Germany as well as sociodemographics, which was also revealed in this study [27,45].

Along with this, higher privileges of Ukrainian refugees and differences in refugee experience or arrival in Germany were discussed among the participants. Not only are there differences between refugees from Ukraine and from other countries, but there are also major differences in the experiences of war and flight among refugees from Ukraine, particularly at the time the study was conducted, which was only one year after the beginning of the war [27]: Some of the participating refugees from Ukraine had left the country right at the beginning of the war without having experienced it themselves, and also without having experienced the loss of family members or friends. Furthermore, the Ukrainian participants might still have hoped to return to their home country soon [27]. Different durations of their stay in Germany therefore seemed to be linked to different flight experiences and, herewith, with a different identification with the refugee context and associated problems described in the adapted intervention version. Depending on this, the adapted version of the intervention could therefore show a certain degree of perceived appropriateness for some of the participating refugees from Ukraine. A dilemma became apparent: Whereas the Ukrainian participants seemed to identify with the European culture of the example characters in the nonadapted version, they did not identify with their concerns, especially regarding a high workload; with regard to the example characters in the adapted version, they could possibly identify in parts with their refugee background, but not with their non-European cultural background. This was accompanied by identity conflicts regarding the refugee context in the statements of the Ukrainian participants, which were not found among the participants from other countries of origin. For example, one Ukrainian participant described the adapted version of the intervention as appropriate for refugees, but not for herself. The importance of identity issues and conflicts in the refugee context has been highlighted previously [46,47].

The found differences in the identity conflicts and the identification with the versions of the intervention emphasize the relevance of specifying a target group and their specific needs in the context of cultural adaptations [48,49]. The lack of including the living situation and the needs of Ukrainian refugees during the cultural adaptation of the present intervention [24] may have led to the identification difficulties and, herewith, the reduced perceived cultural appropriateness. In order to enhance the appropriateness of the intervention for refugees from Ukraine, the intervention could be further adapted, for example by providing the possibility for personalization of the content or of the example characters in order to improve the identification. This has already been applied, for example, as part of the Step-by-Step program in Lebanon [50,51].

Limitations

Several study limitations are to be named. First, the generalizability and robustness of the results are limited due to the qualitative and explorative character of the study. Furthermore, we could only include study participants with sufficient knowledge of German or English. Due to the open recruitment, this resulted in a rather educated and young sample, which possibly influenced the perceived cultural appropriateness of interventions [52]. Second, the majority of participants were recruited via private contacts. This may have biased the results in terms of social desirability, despite enquiring about perceived negative aspects of the interventions. Third, the duration and mode of conduct (video call vs in person) of the interviews with the study participants were not parallelized for the two subgroups. This might have led to an overrepresentation of the statements made by the Ukrainians, combined with the fact that the category system was initially developed based on the transcripts of Ukrainian participants. Fourth, the frequency analyses are not directly interpretable, as they were not related to the frequency of implementation of the respective intervention.

Conclusions

With the rising number of refugees worldwide, the need for low-threshold and culturally appropriate mental health services is increasing. In this qualitative study, the relevance of the identification of the participating refugees with the content of the digital intervention Sleep-e for their perceived cultural appropriateness was emphasized. There were substantial differences between the participating refugees from Ukraine and those from other countries of origin: The participants from Ukraine appeared to be in an identity conflict regarding their identification with the refugee context, which seemed to reduce the perceived appropriateness both of the version of the intervention adapted for refugees from African and Middle East countries and of the nonadapted version for Germans. Refugees from other countries of origin, on the other hand, could identify well with the adapted version of the intervention and perceived it as appropriate. The perceived cultural appropriateness for refugees from Ukraine should be improved by diversifying the problems addressed to consider current developments related to the war in Ukraine and other changes in world affairs. As

XSL•FO
RenderX

such, the cultural sensitivity of treatments should be considered as a dynamic condition related to time and context [53].

Acknowledgments

We would like to thank all study participants for their time and willingness to contribute. We further acknowledge support by the Open Access Publication Fund of the University of Freiburg.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

MB and KS initiated this study. MB, AK, LBS, and KS contributed to the design of this study. Data acquisition was facilitated by HB providing the platform for intervention access, and DL providing the original intervention GET.ON Recovery, and JU provide access to all further materials and facilities. MB largely contributed to the study preparation, recruitment, data collection, analyses, and interpretation. HB and DL provided expertise in digital interventions, and AK, LBS, JU, and KS supervised the whole project. MB and KS wrote the draft of the manuscript. All authors contributed to the further writing of the manuscript and approved the final version of it.

Conflicts of Interest

DL, LBS, and KS were involved in the development of the Sleep-e intervention or its predecessor versions.

Multimedia Appendix 1
Interview guideline.
[PDF File (Adobe PDF File), 90 KB - formative_v9i1e65412_app1.pdf]

Multimedia Appendix 2
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[PDF File (Adobe PDF File), 156 KB - formative_v9i1e65412_app2.pdf]

Multimedia Appendix 3
Category system.
[PDF File (Adobe PDF File), 127 KB - formative_v9i1e65412_app3.pdf]

References

1. Global trends: forced displacement in 2023. UNHCR. 2024. URL: https://www.unhcr.org/global-trends-report-2023 [accessed 2024-08-05]
2. Blackmore R, Boyle JA, Fazel M, Ranasinha S, Gray KM, Fitzgerald G, et al. The prevalence of mental illness in refugees and asylum seekers: a systematic review and meta-analysis. PLoS Med 2020;17(9):e1003337 [FREE Full text] [doi: 10.1371/journal.pmed.1003337] [Medline: 32956381]
3. Kien C, Sommer I, Faustmann A, Gibson L, Schneider M, Krczal E, et al. Prevalence of mental disorders in young refugees and asylum seekers in European countries: a systematic review. Eur Child Adolesc Psychiatry 2019;28(10):1295-1310 [FREE Full text] [doi: 10.1007/s00787-018-1215-z] [Medline: 30151800]
4. Richter K, Baumgärtner L, Niklewski G, Peter L, Köck M, Kellner S, et al. Sleep disorders in migrants and refugees: a systematic review with implications for personalized medical approach. EPMA J 2020;11(2):251-260 [FREE Full text] [doi: 10.1007/s13167-020-00205-2] [Medline: 32549917]
5. Uphoff E, Robertson L, Cabieses B, Villalón FJ, Purgato M, Churchill R, et al. An overview of systematic reviews on mental health promotion, prevention, and treatment of common mental disorders for refugees, asylum seekers, and internally displaced persons. Cochrane Database Syst Rev 2020;9(9):CD013458 [FREE Full text] [doi: 10.1002/14651858.CD013458.pub2] [Medline: 32885850]
6. Baggio S, Gonçalves L, Heeren A, Heller P, Gétaz L, Graf M, et al. The mental health burden of immigration detention: an updated systematic review and meta-analysis. Criminology - The Online Journal 2020;2(2):219-233. [doi: 10.18716/ojs/krimoj/2020.2.7]
7. Jannesari S, Lotito C, Turrini G, Oram S, Barbui C. How does context influence the delivery of mental health interventions for asylum seekers and refugees in low- and middle-income countries? a qualitative systematic review. Int J Ment Health Syst 2021;15(1):80 [FREE Full text] [doi: 10.1186/s13033-021-00501-y] [Medline: 34702334]

XSL•FO
RenderX

8. Satinsky E, Fuhr DC, Woodward A, Sondorp E, Roberts B. Mental health care utilisation and access among refugees and asylum seekers in Europe: a systematic review. Health Policy 2019;123(9):851-863 [FREE Full text] [doi: 10.1016/j.healthpol.2019.02.007] [Medline: 30850148]
9. Kaufman KR, Bhui K, Katona C. Mental health responses in countries hosting refugees from Ukraine. BJPsych Open 2022;8(3):e87 [FREE Full text] [doi: 10.1192/bjo.2022.55] [Medline: 35361299]
10. Liem A, Natari RB, Jimmy, Hall BJ. Digital health applications in mental health care for immigrants and refugees: a rapid review. Telemed J E Health 2021;27(1):3-16. [doi: 10.1089/tmj.2020.0012] [Medline: 32498658]
11. El-Haj-Mohamad R, Nohr L, Niemeyer H, Böttche M, Knaevelsrud C. Smartphone-delivered mental health care interventions for refugees: a systematic review of the literature. Glob Ment Health (Camb) 2023;10:e6 [FREE Full text] [doi: 10.1017/gmh.2022.61] [Medline: 36843879]
12. Moshe I, Terhorst Y, Philippi P, Domhardt M, Cuijpers P, Cristea I, et al. Digital interventions for the treatment of depression: a meta-analytic review. Psychol Bull 2021;147(8):749-786. [doi: 10.1037/bul0000334] [Medline: 34898233]
13. Pauley D, Cuijpers P, Papola D, Miguel C, Karyotaki E. Two decades of digital interventions for anxiety disorders: a systematic review and meta-analysis of treatment effectiveness. Psychol Med 2023;53(2):567-579 [FREE Full text] [doi: 10.1017/S0033291721001999] [Medline: 34047264]
14. Rismawan W, Marchira CR, Rahmat I. Usability, acceptability, and adherence rates of mobile application interventions for prevention or treatment of depression: a systematic review. J Psychosoc Nurs Ment Health Serv 2021;59(2):41-47. [doi: 10.3928/02793695-20201104-02] [Medline: 33180949]
15. Or CK, Chan AHS. Inspection methods for usability evaluation. In: User Experience Methods and Tools in Human-Computer Interaction. Boca Raton, FL: CRC Press; 2024:170-192.
16. Hall GCN, Ibaraki AY, Huang ER, Marti CN, Stice E. A meta-analysis of cultural adaptations of psychological interventions. Behav Ther 2016;47(6):993-1014. [doi: 10.1016/j.beth.2016.09.005] [Medline: 27993346]
17. Harper Shehadeh M, Heim E, Chowdhary N, Maercker A, Albanese E. Cultural adaptation of minimally guided interventions for common mental disorders: a systematic review and meta-analysis. JMIR Ment Health 2016;3(3):e44 [FREE Full text] [doi: 10.2196/mental.5776] [Medline: 27670598]
18. Soto A, Smith TB, Griner D, Domenech Rodríguez M, Bernal G. Cultural adaptations and therapist multicultural competence: two meta-analytic reviews. J Clin Psychol 2018;74(11):1907-1923. [doi: 10.1002/jclp.22679] [Medline: 30091201]
19. Heim E, Ramia JA, Hana RA, Burchert S, Carswell K, Cornelisz I, et al. Step-by-step: feasibility randomised controlled trial of a mobile-based intervention for depression among populations affected by adversity in Lebanon. Internet Interv 2021;24:100380 [FREE Full text] [doi: 10.1016/j.invent.2021.100380] [Medline: 33747798]
20. Migration and cultural diversity: cultural competence and cultural sensitivity. Finish Institute for Health and Welfare. 2023. URL: https://thl.fi/en/topics/migration-and-cultural-diversity/support-material/good-practices/cultural-competence-and-cultural-sensitivity [accessed 2024-08-08]
21. Böge K, Karnouk C, Hoell A, Tschorn M, Kamp-Becker I, Padberg F, et al. Effectiveness and cost-effectiveness for the treatment of depressive symptoms in refugees and asylum seekers: a multi-centred randomized controlled trial. Lancet Reg Health Eur 2022;19:100413 [FREE Full text] [doi: 10.1016/j.lanepe.2022.100413] [Medline: 35694653]
22. Spanhel K, Hovestadt E, Lehr D, Spiegelhalder K, Baumeister H, Bengel J, et al. Engaging refugees with a culturally adapted digital intervention to improve sleep: a randomized controlled pilot trial. Front Psychiatry 2022;13:832196 [FREE Full text] [doi: 10.3389/fpsyt.2022.832196] [Medline: 35280163]
23. El-Haj-Mohamad R, Böttche M, Vöhringer M, Specht F, Stammel N, Nesterko Y, et al. An internet-based cognitive behavioural intervention for adults with depression in Arabic-speaking countries: a randomized controlled trial. Stress Health 2024;40(5):e3432. [doi: 10.1002/smi.3432] [Medline: 38806256]
24. Spanhel K, Schweizer JS, Wirsching D, Lehr D, Baumeister H, Bengel J, et al. Cultural adaptation of internet interventions for refugees: results from a user experience study in Germany. Internet Interv 2019;18:100252 [FREE Full text] [doi: 10.1016/j.invent.2019.100252] [Medline: 31890608]
25. Ukraine situation flash update #71. UNHCR. 2024. URL: https://data.unhcr.org/en/documents/details/110114 [accessed 2024-08-05]
26. Weissbecker I, Khan O O, Kondakova N. N, Poole L L, Cohen JT. Mental health in transition: assessment and guidance for strengthening integration of mental health into primary health care and community-based service platforms in Ukraine. World Bank Group. 2017. URL: http://documents.worldbank.org/curated/en/310711509516280173 [accessed 2024-08-08]
27. Brücker H, Ette A, Grabka M, Kosyakova Y, Niehues W, Rother N, et al. Geflüchtete aus der Ukraine in Deutschland (Forschungsbericht 41). [Refugees from Ukraine in Germany (research report 41)] 2022 [FREE Full text] [doi: 10.12765/bro-2022-04]
28. Ukraine situation: regional refugee response plan. UNHCR. Geneva, Switzerland; 2023. URL: https://reporting.unhcr.org/ukraine-situation-regional-refugee-response-plan [accessed 2024-08-09]
29. Buchcik J, Kovach V, Adedeji A. Mental health outcomes and quality of life of Ukrainian refugees in Germany. Health Qual Life Outcomes 2023;21(1):23 [FREE Full text] [doi: 10.1186/s12955-023-02101-5] [Medline: 36894946]
30. Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001;2(4):297-307. [doi: 10.1016/s1389-9457(00)00065-4] [Medline: 11438246]

XSL•FO
RenderX

31. Boß L, Lehr D, Reis D, Vis C, Riper H, Berking M, et al. Reliability and validity of assessing user satisfaction with web-based health interventions. J Med Internet Res 2016;18(8):e234 [FREE Full text] [doi: 10.2196/jmir.5952] [Medline: 27582341]
32. Konrad K. Lautes Denken [Think Aloud]. In: Handbuch Qualitative Forschung in der Psychologie [Handbook Qualitative Research in Psychology]. VS Verlag für Sozialwissenschaften. Cham, Switzerland: Springer; 2010:476-490.
33. Thiart H, Lehr D, Ebert DD, Berking M, Riper H. Log in and breathe out: internet-based recovery training for sleepless employees with work-related strain - results of a randomized controlled trial. Scand J Work Environ Health 2015;41(2):164-174 [FREE Full text] [doi: 10.5271/sjweh.3478] [Medline: 25590336]
34. Ebert DD, Berking M, Thiart H, Riper H, Laferton JAC, Cuijpers P, et al. Restoring depleted resources: Efficacy and mechanisms of change of an internet-based unguided recovery training for better sleep and psychological detachment from work. Health Psychol 2015;34S:1240-1251. [doi: 10.1037/hea0000277] [Medline: 26651465]
35. Spanhel K, Balci S, Feldhahn F, Bengel J, Baumeister H, Sander LB. Cultural adaptation of internet- and mobile-based interventions for mental disorders: a systematic review. NPJ Digit Med 2021;4(1):128 [FREE Full text] [doi: 10.1038/s41746-021-00498-1] [Medline: 34433875]
36. Kuckartz U, Rädiker S. Qualitative Inhaltsanalyse [Qualitative content analysis]. In: Methoden, Praxis, Computerunterstützung [Methods, Practice, Computer Support]. Weinheim, Germany: Beltz Juventa; 2022.
37. Mayring P. Qualitative Inhaltsanalyse [Qualitative content analysis]. In: Grundlagen und Techniken [Basics and techniques]. Weinheim, Germany: Beltz Verlag; 2022.
38. Stamann C, Janssen M, Schreier M. Searching for the core: defining qualitative content analysis. Forum Qual Soc Res 2016;17(3):16. [doi: 10.17169/FQS-17.3.2581]
39. VERBI Software. MAXQDA. 2022. URL: https://www.maxqda.com/de/ [accessed 2024-08-08]
40. Guidance note - research on refugees, asylum seekers and migrants. European Commission. 2016. URL: https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/guide_research-refugees-migrants_en.pdf [accessed 2025-02-24]
41. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
42. Micklitz HM, Nagel Z, Jahn S, Oertelt-Prigione S, Andersson G, Sander LB. Digital self-help for people experiencing intimate partner violence: a qualitative study on user experiences and needs including people with lived experiences and services providers. BMC Public Health 2023;23(1):1471 [FREE Full text] [doi: 10.1186/s12889-023-16357-5] [Medline: 37533005]
43. Büscher R, Sander LB, Nuding M, Baumeister H, Teismann T. Blending video therapy and digital self-help for individuals with suicidal ideation: intervention design and a qualitative study within the development process. JMIR Form Res 2023;7:e49043 [FREE Full text] [doi: 10.2196/49043] [Medline: 37733414]
44. Seiferth C, Vogel L, Aas B, Brandhorst I, Carlbring P, Conzelmann A, et al. How to e-mental health: a guideline for researchers and practitioners using digital technology in the context of mental health. Nat. Mental Health 2023;1(8):542-554. [doi: 10.1038/s44220-023-00085-1]
45. Prati G. Attitudes toward the response of the EU to the Russian invasion of Ukraine: The role of empathic concern, identification with the EU, and the view that Ukraine is part of the European community. Peace and Conflict: Journal of Peace Psychology 2023;29(3):253-256. [doi: 10.1037/pac0000679]
46. Kizilhan JI. Kultursensitivität in der Psychotherapie [Culture sensitivity in psychotherapy]. In: Verhaltenstherapiemanual - Erwachsene [Behavioral therapy manual - adults]. Berlin/ Heidelberg, Germany: Springer; 2022:43-48.
47. Watzlawik M, Brescó de Luna I. The self in movement: being identified and identifying oneself in the process of migration and asylum seeking. Integr Psychol Behav Sci 2017;51(2):244-260. [doi: 10.1007/s12124-017-9386-6] [Medline: 28299535]
48. Heim E, Kohrt BA. Cultural adaptation of scalable psychological interventions: a new conceptual framework. CPE 2019;1(4):e37679. [doi: 10.32872/cpe.v1i4.37679]
49. Heim E, Burchert S, Wenger A. Mobile Angebote für Geflüchtete: Psychotherapeutische Interventionen im globalen Kontext [Mobile services for refugees: psychotherapeutic interventions in a global context]. PiD - Psychotherapie im Dialog 2018;19(04):61-65. [doi: 10.1055/a-0592-0339]
50. Abi Ramia J, Harper Shehadeh M, Kheir W, Zoghbi E, Watts S, Heim E, et al. Community cognitive interviewing to inform local adaptations of an e-mental health intervention in Lebanon. Glob Ment Health (Camb) 2018;5:e39 [FREE Full text] [doi: 10.1017/gmh.2018.29] [Medline: 30637112]
51. Harper Shehadeh MJ, Abi Ramia J, Cuijpers P, El Chammay R, Heim E, Kheir W, et al. Step-by-step, an E-mental health intervention for depression: a mixed methods pilot study from lebanon. Front Psychiatry 2019;10:986 [FREE Full text] [doi: 10.3389/fpsyt.2019.00986] [Medline: 32116815]
52. Heim E, Ademi B, Dacaj A, Hosny N, Burchert S, Cerga Pashoja A, et al. Why did our trial not work out? A qualitative analysis. Clin Psychol Eur 2024;6(2):e12887. [doi: 10.32872/cpe.12887] [Medline: 39119055]
53. Bernal G, Bonilla J, Bellido C. Ecological validity and cultural sensitivity for outcome research: issues for the cultural adaptation and development of psychosocial treatments with Hispanics. J Abnorm Child Psychol 1995;23(1):67-82. [doi: 10.1007/BF01447045] [Medline: 7759675]

XSL•FO
RenderX

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research

Edited by A Mavragani; submitted 07.10.24; peer-reviewed by C Or; comments to author 28.01.25; revised version received 20.02.25; accepted 21.02.25; published 03.04.25.

Please cite as:
Blomenkamp M, Kiesel A, Baumeister H, Lehr D, Unterrainer J, Sander LB, Spanhel K
Assessing the Cultural Fit of a Digital Sleep Intervention for Refugees in Germany: Qualitative Study
JMIR Form Res 2025;9:e65412
URL: https://formative.jmir.org/2025/1/e65412
doi:10.2196/65412
PMID:

©Maja Blomenkamp, Andrea Kiesel, Harald Baumeister, Dirk Lehr, Josef Unterrainer, Lasse B Sander, Kerstin Spanhel. Originally published in JMIR Formative Research (https://formative.jmir.org), 03.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Health Platform for Maternal Health: Design, Recruitment Strategies, and Lessons Learned From the PowerMom Observational Cohort Study

Toluwalase Ajayi[1,2], MD; Jacqueline Kueper[2], PhD; Lauren Ariniello[2], MBA; Diana Ho[2], MHA; Felipe Delgado[2], BS; Matthew Beal[3], MD; Jill Waalen[3], PhD; Katie Baca Motes[2], MBA; Edward Ramos[2], PhD

[1]Jacobs Center for Health Innovation, Department of Medicine and Pediatrics, University of California, San Diego, La Jolla, CA, United States

[2]Digital Trial Center, Scripps Research Translational Institute, La Jolla, CA, United States

[3]Division of Preventative Medicine, Department of Family Medicine, University of California, San Diego, La Jolla, United States

Corresponding Author:
Toluwalase Ajayi, MD
Jacobs Center for Health Innovation
Department of Medicine and Pediatrics
University of California, San Diego
9300 Campus Point Drive
MC 7196
La Jolla, CA, 92037
United States
Phone: 1 785 218 1643
Email: tajayi@health.ucsd.edu

Abstract

Background: Maternal health research faces challenges in participant recruitment, retention, and data collection, particularly among underrepresented populations. Digital health platforms like PowerMom (Scripps Research) offer scalable solutions, enabling decentralized, real-world data collection. Using innovative recruitment and multimodal techniques, PowerMom engages diverse cohorts to gather longitudinal and episodic data during pregnancy and post partum.

Objective: This study aimed to evaluate the design, implementation, and outcomes of the PowerMom research platform, with a focus on participant recruitment, engagement, and data collection across diverse populations. Secondary objectives included identifying challenges encountered during implementation and deriving lessons to inform future digital maternal health studies.

Methods: Participants were recruited via digital advertisements, pregnancy apps, and the PowerMom Consortium of more than 15 local and national organizations. Data collection included self-reported surveys, wearable devices, and electronic health records. Anomaly detection measures were implemented to address fraudulent enrollment activity. Recruitment trends and descriptive statistics from survey data were analyzed to summarize participant characteristics, assess engagement metrics, and quantify missing data to identify gaps.

Results: Overall, 5617 participants were enrolled from 2021 to 2024, with 69.8% (n=3922) providing demographic data. Of these, 48.5% (2723/5617) were younger than 35 years, 14% (788/5617) identified as Hispanic or Latina, and 13.7% (770/5617) identified as Black or African American. Geographic representation spanned all 50 US states, Puerto Rico, and Guam, with 58.3% (3276/5617) residing in areas with moderate access to maternity care and 16.4% (919/5617) in highly disadvantaged neighborhoods based on the Area Deprivation Index. Enrollment rates increased substantially over the study period, from 55 participants in late 2021 to 3310 in 2024, averaging 99.4 enrollments per week in 2024. Participants completed a total of 17,123 surveys, with 71.8% (4033/5617) completing the Intake Survey and 12.4% (697/5617) completing the Postpartum Survey. Wearable device data were shared by 1168 participants, providing more than 378,000 daily biometric measurements, including activity levels, sleep, and heart rate. Additionally, 96 participants connected their electronic health records, contributing 276 data points such as diagnoses, medications, and laboratory results. Among pregnancy-related characteristics, 28.1% (1578/5617) enrolled during the first trimester, while 15.1% (849/5617) reported information about the completion of their pregnancies during the study period. Among the 913 participants who shared delivery information, 56.1% (n=512) had spontaneous vaginal deliveries and 17.9% (n=163) underwent unplanned cesarean sections.

XSL•FO
RenderX

Conclusions: The PowerMom platform demonstrates the feasibility of using digital tools to recruit and engage diverse populations in maternal health research. Its ability to integrate multimodal data sources showcases its potential to provide comprehensive maternal-fetal health insights. Challenges with data completeness and survey attrition underscore the need for sustained participant engagement strategies. These findings offer valuable lessons for scaling digital health platforms and addressing disparities in maternal health research.

Trial Registration: ClinicalTrials.gov NCT03085875; https://clinicaltrials.gov/study/NCT03085875

(JMIR Form Res 2025;9:e70149) doi:10.2196/70149

KEYWORDS

maternal health research; digital health platforms; pregnancy monitoring; decentralized clinical trials; participant engagement; health disparities

Introduction

In 2020, there were approximately 24 pregnancy-related deaths per 100,000 births in the United States, which is more than double compared with other high-income countries [1]. Alarmingly, these rates have continued to rise, with disparities disproportionately affecting marginalized populations [2]. Non-Hispanic Black or African American pregnant individuals experience maternal mortality rates approximately 2.5 times higher than their White counterparts, with non-Hispanic American Indian or Alaska Native women facing rates three times higher [2,3]. Cardiovascular conditions, perinatal depression, and preexisting medical conditions are among the leading contributors to these disparities [4,5]. However, the root causes extend beyond medical risk factors, with social determinants of health and structural racism playing crucial roles [6,7].

Access to care, socioeconomic status, implicit bias, and environmental factors further exacerbate these inequities. Individuals from underrepresented communities are more likely to face barriers to health care, economic insecurity, and a lack of culturally competent care. These disparities are often intensified by distrust of health care systems, fueled by historical and ongoing discrimination in medical settings [8]. Addressing these disparities requires innovative research approaches that not only engage affected communities but also prioritize trust, accessibility, and cultural competence [9].

Traditional maternal health research has struggled to effectively recruit and engage underrepresented populations. Common recruitment barriers include mistrust of the health care system, concerns about data confidentiality, language differences, and logistical hurdles such as the lack of transportation or time constraints [10,11]. These challenges are compounded by the fact that traditional recruitment methods, such as clinic-based enrollment, often fail to reach the most vulnerable populations, leading to underrepresentation in research that is critical for health equity [12,13]. However, many of these traditional methods do not offer sufficient direct benefits to participants to offset their perceived and real costs of participation, particularly in communities with heightened mistrust [9].

The emergence of digital health platforms offers a potential solution to overcome these long-standing barriers, particularly in maternal health research [14]. Mobile and decentralized platforms provide an opportunity to recruit participants remotely, allowing for broad geographical reach and the ability to engage individuals in real-world settings [15]. Digital tools enable continuous, real-time data collection, which can provide a more holistic view of maternal health than episodic clinic visits. These platforms can also deliver tailored educational content and resources, building trust and empowering participants through transparency and engagement [16].

The PowerMom (Scripps Research) platform was designed to capitalize on these opportunities by implementing a bilingual (English and Spanish) digital research platform that facilitates remote, convenient participation. PowerMom's recruitment strategy was explicitly designed to harness the power of digital platforms to allow seamless, rolling self-enrollment and ensure wide-scale and inclusive participation across diverse communities. It leverages digital approaches such as push notifications and messaging to enhance participant engagement and retention [17-20], allowing for data collection that respects participants' time and cultural needs.

By providing a decentralized clinical trial environment, PowerMom enables participants to engage with the study on their own terms, enhancing adherence and yielding data that reflect a broad, comprehensive, and participant-centered perspective on how structural factors influence maternal health. The platform supports the collection of real-world, continuous data through app-based surveys, electronic health record (EHR) data linkage, and digital health technologies such as wearable sensors, capturing biometric data (eg, heart rate, sleep, and physical activity) that go beyond the traditional, episodic measures collected during prenatal clinic visits [21-26]. PowerMom's flexible architecture supports not only the collection of baseline data but also the integration of multiple substudies, which allows for more targeted research on specific maternal health issues while leveraging the broader cohort's data.

In this paper, we describe the design and methods of the PowerMom research platform, focusing on its recruitment strategies, data collection capabilities, and participant engagement efforts, and discuss lessons learned from its implementation in diverse populations. In addition, we provide a detailed description of the participant demographics recruited through PowerMom and examine their engagement with the various data collection opportunities throughout the pregnancy and postpartum stages.

XSL•FO
RenderX

Methods

Platform Design

The PowerMom platform, launched in March 2021, is built on CareEvolution's MyDataHelps secure digital infrastructure, which provides a flexible and scalable foundation for collecting and integrating a range of participant data, including biometric, survey, and EHR data. This cloud-based, Health Insurance Portability and Accountability Act (HIPAA)–compliant infrastructure provides secure data storage and management, leveraging CareEvolution's backend architecture to maintain participant confidentiality and data integrity. In addition, the platform's technical infrastructure adheres to industry-leading security protocols for data transfer, ensuring that biometric data from wearables, EHR data, and survey responses are securely encrypted and processed.

The PowerMom platform integrates data from various sources in real time, allowing for continuous monitoring of participants' health. This includes heart rate, activity levels, and sleep data collected from wearable devices such as Fitbit, Apple Watch, and other HealthKit or Google Fit–compatible technologies. By allowing participants to connect these devices, the platform facilitates a richer, more granular view of maternal health, capturing data beyond the traditional measures obtained during prenatal clinic visits.

To ensure that the platform reflects the needs of its target population, PowerMom's development involved collaboration with a participant advisory board composed of pregnant and postpartum individuals from underrepresented communities. The advisory board provided critical feedback on the platform's design, user interface, and engagement strategies, ensuring that the platform remains culturally relevant and user-friendly for diverse participants. This adheres to the CONSORT-EHEALTH (v1.6.1) guidelines for digital health intervention research to ensure comprehensive and standardized reporting. The checklist was used to structure study reporting, covering recruitment, platform usability, data integration, and participant engagement.

Participant Recruitment

Eligibility Criteria

Participants were eligible for PowerMom if they were pregnant or post partum (up to 8 weeks after delivery), aged 16 years and older, residing in the United States or its territories, and capable of providing consent in either English or Spanish.

eConsent

All eligible PowerMom participants were presented with an innovative electronic consent (eConsent) process, enabling them to remotely join the study. This eConsent process significantly broadened the potential participant pool by allowing individuals to complete consent forms at their convenience, reducing the need for in-person visits or transfer of paper forms. The eConsent interface was designed to be user-friendly and accessible across different devices, ensuring that participants could easily navigate the process (refer to Multimedia Appendix 1). This digital consent process not only facilitated seamless participation but also ensured compliance with ethical standards for informed consent in clinical research.

Study Setting

The study was conducted entirely remotely, using the PowerMom platform to engage participants across the United States, Puerto Rico, and Guam. The decentralized nature of the study eliminated geographic barriers, enabling recruitment from urban, suburban, and rural areas. Participants could enroll and participate using mobile devices or computers, providing flexibility for those in resource-limited settings.

Recruitment Strategy

PowerMom implemented a rolling, digital self-enrollment process to maximize reach and engagement. Recruitment took place from March 22, 2021, to July 30, 2024, and relied on a variety of digital and in-person outreach strategies. Participants in PowerMom are recruited into a baseline cohort but may also be invited to join additional substudies based on specific eligibility criteria. These substudies focus on more targeted research questions, allowing researchers to augment baseline data with additional detailed outcomes.

Digital Outreach

Recruitment relied heavily on digital advertising campaigns on social media platforms such as Facebook, Instagram, and Google Ads. In addition, partnerships with pregnancy-related apps such as Philips Pregnancy+ (Philips Digital) enabled in-app advertisements that reached individuals in their natural digital environments. These ads provided information about the study, emphasizing its accessibility and participant-centered design. Digital enrollment campaigns were designed to appear within the app, enabling participants to enroll in PowerMom seamlessly from their daily digital environment.

PowerMom Consortium

The PowerMom Consortium, established in December 2021, brought together more than 15 local and national organizations focused on improving maternal health. This consortium amplified recruitment efforts through joint campaigns, newsletters, and educational outreach, significantly increasing visibility among target demographics [27].

Community Engagement

One of the platform's key partners, Mae Health, is a culturally competent, digital health platform that connects Black expectant mothers with vital resources to improve pregnancy outcomes. The collaboration with Mae Health included educational campaigns and targeted outreach to Black and African American communities.

PowerMom also partnered with advocacy organizations such as the African American Wellness Center for Children and Families. These organizations provided on-the-ground recruitment efforts, ensuring that PowerMom reached communities that might otherwise be underrepresented in digital health research.

XSL•FO
RenderX

Data Collection

Self-Reported Outcomes

Participants were asked to complete several surveys throughout the study. An initial Intake Survey gathered demographic data, including race, ethnicity, education, socioeconomic status, and health care access. A Health History Survey collected information on pregnancy history, preexisting health conditions, and access to care.

Following the baseline surveys, participants received biweekly surveys to assess changes in their health during pregnancy. These surveys collected data on vital signs (eg, weight, blood pressure, and heart rate); pregnancy symptoms; and health behaviors, including tobacco, alcohol, and drug use. Biweekly surveys were released based on the completion of the Intake and Health History Surveys, and participants were notified about each new biweekly survey release via push notifications through the PowerMom app.

Participants also completed a Delivery Survey upon giving birth, capturing details about the delivery, birth outcomes, and complications. A Postpartum Survey, administered at 6-8 weeks post partum, assessed maternal mental health using the Edinburgh Postnatal Depression Scale and gathered information on postpartum complications such as gestational diabetes, pregnancy-induced hypertension, preeclampsia, and preterm birth (refer to Multimedia Appendix 1).

Wearable Devices

Participants were prompted to connect their wearable devices, including Fitbit, Apple Watch, and other compatible devices, for real-time monitoring of heart rate, physical activity, and sleep duration data. Participants had the option to upload up to 3 months of retrospective data (eg, prepregnancy data) and continue data collection prospectively throughout their pregnancy and the postpartum period.

EHR Integration

After completing the initial baseline surveys, participants were prompted to connect their pregnancy-related EHR data through the PowerMom app. The EHR integration can reach more than 85% of all US health care providers, ensuring comprehensive coverage of both historical and prospective health data. Participants could connect their EHR at any point during the study, providing a continuous flow of clinical information that complemented the self-reported and biometric data.

Study Engagement and Retention Strategies

The PowerMom platform was designed with a strong focus on participant engagement and retention. To keep participants engaged, the platform provided regular push notifications, educational content, and personalized data visualizations. For example, participants could track their own metrics—such as daily activity or sleep patterns—through visual feedback in the app, which helped build trust and maintain engagement.

In addition to digital strategies, participants were engaged through blog posts that provided some initial return of results based on aggregated data collected. The PowerMom advisory board also contributed to refining engagement strategies by providing input on how to make the platform more user-friendly and responsive to the needs of underrepresented populations. Engagement metrics were closely monitored, and iterative improvements were made to the platform based on participant feedback.

Data Integration

All data types (eg, survey data, wearable device data, and EHR data) were collected and integrated through the MyDataHelps PowerMom platform app. When discrepancies arose between data sources, such as activity metrics reported in surveys versus wearables, priority was given to wearable data due to their objective and continuous measurement capabilities. Survey data were used to provide context, particularly when participants reported conditions or limitations that could explain deviations. This approach ensured the accuracy and reliability of the data.

These data were securely transferred to the Scripps Research Institute for storage and analysis. The data management system leverages cloud-based infrastructure that is HIPAA compliant, ensuring the highest level of data integrity and security. The data were regularly monitored by the research team and cleaned to address any issues with synchronization between devices, missing data, or outliers, ensuring high-quality data for analysis.

Statistical Analysis

We evaluated the PowerMom platform using operational outcomes and participant characteristics. The analytic sample included all participants who met the baseline PowerMom eligibility criteria, assessed through an initial Screener Survey, and who provided eConsent to participation.

Participant Characteristics

Baseline sociodemographic and pregnancy-related characteristics of all enrolled participants were summarized using self-report data from the Screener, Intake, and Health History Surveys. Delivery outcomes were also described for participants who filled out the Delivery Survey. All eligible participants are included in table summaries of these characteristics, with missing data represented as “Unknown” responses for both single-question and survey-level nonresponse. Survey-level completion patterns, reported in the engagement metrics section, capture the number of eligible participants who completed at least one question on each survey.

Enrollment Trends

We analyzed enrollment trends beyond the three-year study recruitment period. Weekly and cumulative enrollment rates were calculated to demonstrate the impact of these strategies, culminating in the final total number of participants.

Summary statistics were calculated using frequencies and percentages for categorical variables and mean (SD) for continuous variables. All analyses were conducted using MyDataHelps Export Explorer (SQL) and Python (v3.8; Guido van Rossum).

Ethical Considerations

The study protocol for PowerMom was reviewed and approved by the Scripps Health Institutional Review Board (IRB 21-7738). The research adhered to ethical guidelines for human subject

XSL•FO
RenderX

research, ensuring compliance with all applicable regulations. All participants provided eConsent before enrollment. The eConsent process was designed to be accessible and user-friendly, enabling participants to complete the consent remotely via digital platforms. The consent process highlighted participants' rights, data confidentiality measures, and the voluntary nature of their participation. Participant data were anonymized or deidentified before analysis to ensure confidentiality. The PowerMom platform operates on a HIPAA-compliant, secure infrastructure, with additional protections provided by a Certificate of Confidentiality from the National Institutes of Health. This legal protection allows the platform to refuse the disclosure of participant data in civil, criminal, administrative, or legislative proceedings, adding an extra robust layer of privacy beyond standard HIPAA protections. Participants were not compensated for being part of the PowerMom study. However, participants recruited from the PowerMom study for one substudy were provided with monetary incentives. Participants in this substudy were eligible to receive up to US $175 in Amazon gift cards as compensation for completing specific milestones within the substudy. These incentives were structured to encourage engagement without coercion, according to institutional review board guidelines, and the details were explicitly communicated during the consent process.

Results

Demographic Data

Participants were drawn from diverse demographic backgrounds, with representation across different age groups, racial and ethnic categories, and geographic locations. Among the 5617 participants enrolled in the study, 3922 (69.8%) opted to share demographic information. Participants younger than 35 years comprised 48.5% (2723/5617) of the cohort, while 19.3% (1083/5617) were 35 years or older. Age data were missing for 32.2% (1811/5617) of participants.

Regarding ethnicity, 14% (788/5617) of participants reported identifying as Hispanic or Latina, while 55.8% (3134/5617) did not. Ethnicity data were missing for 30.2% (1695/5617) of participants. Racial backgrounds included 0.9% (52/5617) identifying as American Indian or Alaska Native, 1.9% (105/5617) as Asian, 13.7% (770/5617) as Black or African American or African, and 38.7% (2175/5617) as White. In addition, 8.5% (478/5617) identified as Hispanic, Latina, or Spanish without specifying another race, 0.2% (9/5617) as Native Hawaiian or Other Pacific Islander, and 5.9% (333/5617) as multiple races. Racial data were missing for 30.2% (1695/5617) of participants.

Geographic distribution based on zip code at the time of study enrollment showed that 11.3% (635/5617) of participants residing in the Northeast, 29% (1630/5617) in the South, 15.7% (880/5617) in the Midwest, and 15.2% (853/5617) in the West. Only 0.4% (22/5617) resided in the Pacific region, and less than 0.1% (6/5617) were from Puerto Rico. Geographic data were missing for 28.4% (1594/5617) of participants.

Education levels varied, with 12.5% (703/5617) of participants reporting a graduate or postgraduate degree, 18.9% (1063/5617) holding a college degree, and 6.1% (342/5617) reporting some college or technical school education. High-school graduates or those with General Educational Development attainment accounted for 27.5% (1545/5617), while 6% (337/5617) reported not completed high school. Education data were missing for 29% (1627/5617) of participants.

Access to maternity care was categorized based on geographic location using the March of Dimes Maternity Care Deserts Dashboard [28]. These data were reported by 62.6% (3520/5617) of participants, with 58.3% (3276/5617) residing in areas with moderate access and 4.3% (244/5617) in maternity care deserts or areas with low access. Access data were missing for 37.3% (2097/5617) of participants.

Neighborhood disadvantage was assessed using the Neighborhood Atlas' Area Deprivation Index (ADI). The higher the ADI value, the less resources the neighborhood has. Among all participants, 16.4% (919/5617) reported residence in neighborhoods with an ADI national rank of greater than or equal to 75, indicating higher levels of deprivation, while 24.6% (1381/5617), 19.7% (1108/5617), and 10.2% (574/5617) lived in areas with ADI national ranks between 50-75, 25-50, and less than 25, respectively. ADI data were missing for 29.1% (1635/5617) of participants.

Pregnancy and Delivery-Related Characteristics

Among the 5617 participants enrolled in the study, the pregnancy stage at enrollment varied. Approximately 28.1% (1578/5617) of participants enrolled during the first trimester, 26.6% (1493/5617) in the second trimester, and 16.2% (909/5617) in the third trimester. Postdelivery enrollment accounted for 13.1% (736/5617), while the pregnancy stage at enrollment was unknown for 16% (901/5617) of participants.

Most participants (3394/5617, 60.4%) reported being pregnant with one fetus, while 2.2% (121/5617) reported being pregnant with twins, 0.3% (16/5617) with triplets, and 0.3% (15/5617) with more than 3 fetuses. According to available data [29], around 2% to 3% of pregnancies in the United States result in the birth of triplets or higher-order multiples. The number of fetuses was not reported for 36.9% (2071/5617) of participants.

Regarding pregnancy history, 6.6% (371/5617) of participants reported no previous pregnancies, while 13.5% (759/5617), 12.2% (688/5617), 8% (447/5617), 5.8% (328/5617), and 3.2% (180/5617) reported 1, 2, 3, 4, and 5 previous pregnancies, respectively. In addition, 4.7% (266/5617) of participants reported more than 5 previous pregnancies, and 45.9% (2578/5617) did not provide information on past pregnancies.

When asked about the number of past births, 19.9% (1118/5617) reported having none, while 16.4% (923/5617), 10.2% (571/5617), 4.6% (260/5617), 2% (111/5617, 1% (58/5617), and 0.9% (53/5617) reported 1, 2, 3, 4, 5, and more than 5 births, respectively. Information on past births was missing for 44.9% (2523/5617) of participants. Of the 1106 (21.1%) participants who reported at least one previous pregnancy and a complication, the most commonly reported pregnancy

XSL•FO
RenderX

complications were miscarriage (n=637), preterm birth (n=250), and high blood pressure (n=234).

Participants were seeing a variety of prenatal care providers at baseline, with 37.5% (2104/5617) reported seeing a physician; 9.9% (555/5617) seeing a midwife; 2.3% (129/5617) seeing a doula; and smaller proportions seeing other providers such as nurse midwives (56/5617, 1%) or nurse practitioners (36/5617, 0.6%) or planning home births (9/5617, 0.2%). However, 8% (447/5617) reported not seeing any prenatal care provider, and 44.7% (2511/5617) did not report their provider.

Mental health treatment at baseline revealed that 3.4% (189/5617) of participants were being treated for anxiety, 1.8% (102/5617) for depression, and 9.8% (550/5617) for both anxiety and depression, while 40.9% (2297/5617) reported receiving no treatment. This information was missing for 44.2% (2480/5617) of participants. In addition, 51.4% (2889/5617) of participants reported taking medications or supplements at baseline, 4.6% (261/5617) reported taking none, and 44% (2469/5617) did not report this information. Among the 1449 (25.8%) participants who reported at least one prepregnancy condition, the three most common were COVID-19 (n=781), endometriosis (n=316), and a sexually transmitted illness (n=217).

Regarding family medical history, 9.2% (517/5617) of participants reported having a family member diagnosed with conditions or diseases during childhood, while 46.3% (2603/5617) did not, and 44.5% (2497/5617) did not provide this information. Of the 9.2% (517/5617) of participants that reported themselves, the baby's second biological parent, or anyone in their family having ever had any conditions or diseases diagnosed when they were a baby or in early childhood, the three most common diagnoses were heart defect (n=83), chromosome genetic disorder (n=58), and sickle cell disease or trait (n=21).

For delivery outcomes, 15.1% (849/5617) of participants reported giving birth, 0.9% (52/5617) reported a pregnancy loss, and 0.2% (12/5617) reported stillbirths. Delivery type information showed that 9.1% (512//5617) reported having spontaneous vaginal deliveries, 2.9% (163/5617) had unplanned cesarean sections, and 2.1% (117/5617) had planned cesarean sections. Forceps deliveries were reported by 0.1% (6/5617), vacuum extractions by 0.4% (20/5617), and vaginal births after cesarean by 0.4% (27/5617). Delivery type was not reported by 83.8% (4709/5617) of participants.

Finally, 14.4% (807/5617) of participants reported delivering in a hospital, 0.4% (22/5617) at home, and 0.3% (18/5617) in birthing centers. Delivery location data was unknown for 84.9% (4768/5617) of participants.

Engagement Trends and Survey Completion

The PowerMom platform collected a total of 17,123 individual surveys from 5617 pregnant and postpartum participants across all 50 US states, Puerto Rico, and Guam. Of the 5617 participants who completed the initial Screener Survey, 4031 (71.76%) completed the Intake Survey and 3148 (56.04%) completed the Health History Survey, which were the two baseline surveys administered. There were 1514 (26.95%) participants who completed at least one biweekly Health and Wellbeing Survey; 913 (16.25%) who completed the Delivery Survey; and 697 (12.41%) who completed the Postpartum Survey, which assessed mental health and was sent 6-8 weeks after the participant's reported delivery date.

Fraudulent Behavior

Fraud detection was an essential part of maintaining data integrity. Metrics such as unusually short completion times for eConsent or surveys, duplicate IP addresses, and repetitive data patterns were flagged. 580 of the flagged cases were determined to be fraud and excluded.

Enrollment Trends

Enrollment rates varied significantly from the platform's launch in 2021 through 2024. In the first year of PowerMom's launch, 70 participants were enrolled, with an average of 2.6 participants per week. As recruitment methods were refined and partnerships with pregnancy apps such as Philips Pregnancy+ and Mae Health were formed, the average enrollment rate increased substantially, reaching an average of 93 participants per week by 2024. Table 1 shows the enrollment trends by year, with a clear rise in participant engagement.

Table 1. Participant enrollment rates for the PowerMom study over time.

Year	Total participants (per year), n	Rate (per month)	Rate (per week)
2021 (starting on September 15, 2021)	55	15.7	3.6
2022	481	40.1	9.3
2023	1771	147.6	34.1
2024 (ending on August 22, 2024)	3310	427.1	99.4

Wearable Device Data

Of the 5617 participants, 1168 (20.8%) elected to share data from their wearable devices. The distribution of device use is as follows: 54.1% (632/1168) from Fitbit, 40.6% (474/1168) from Apple HealthKit, and 5.3% (62/1168 participants) from Google Fit. A total of 378,624 daily measurements were collected from participants, covering activity levels, sleep duration, and resting heart rate. These real-time biometric data points provide a detailed, continuous view of participants' health throughout their pregnancy journey.

EHR Data

Connecting EHR data was optional for PowerMom participants. A total of 96 (1.7%) participants chose to connect their EHR with the PowerMom platform, providing 276 distinct data points. These data points included diagnosis codes, lab results, medications, and procedures.

XSL•FO
RenderX

Table 2 presents the summary of the demographic, geographic, and access-to-care data for 5617 participants enrolled in the PowerMom study (2021-2024). The table includes age distribution, Hispanic or Latina identification, racial background, education levels, geographic regions (eg, Northeast and Midwest), and levels of access to maternity care based on the March of Dimes Maternity Care Deserts Dashboard. Data also include neighborhood deprivation as assessed by the ADI. Missing data are reported for each category.

Table 3 presents a summary of baseline health history and pregnancy-related characteristics for 5617 participants enrolled in the PowerMom study. The table includes pregnancy stage at enrollment, number of fetuses, previous pregnancies and births, prenatal care providers, and medication or supplement usage at baseline. In addition, the table details conditions such as prepregnancy and past pregnancy complications, mental health treatment (anxiety and depression), and family disease history. Delivery outcomes, types, and locations are also reported. Missing or unknown data percentages are indicated for each category to account for variability in reporting.

Table 1 presents participant enrollment rates for the PowerMom study, displayed by year. The table reports the total number of participants enrolled each year, along with calculated monthly and weekly enrollment rates. These trends highlight substantial increases in recruitment efficiency, particularly in 2023 and 2024, reflecting the impact of enhanced outreach strategies and platform optimization.

XSL•FO
RenderX

Table 2. Demographic and regional characteristics of PowerMom participants.

Characteristics	Values (N=5617), n (%)
Age category (years)	
<35	2723 (48.5)
≥35	1083 (19.3)
Missing	1811 (32.2)
Hispanic or Latina	
Yes	788 (14.0)
No	3134 (55.8)
Missing	1695 (30.2)
Racial background	
American Indian or Alaska Native	52 (0.9)
Asian	105 (1.9)
Black or African American or African	770 (13.7)
Hispanic, Latina, or Spanish (no other race and ethnicity specified)	478 (8.5)
Native Hawaiian or Other Pacific Islander	9 (0.2)
White	2175 (38.7)
Multiple races	333 (5.9)
Missing	1695 (30.2)
Region	
Northeast	635 (11.3)
South	1630 (29.0)
Midwest	880 (15.7)
West	853 (15.2)
Pacific	24 (0.4)
Puerto Rico	1 (<0.1)
Missing	1594 (28.4)
Education	
Graduate or postgraduate degree	703 (12.5)
College degree	1063 (18.9)
Some college or technical school	342 (6.1)
High school graduate or GED[a] (grade 12)	1545 (27.5)
Some high school (grades 9-11)	337 (6.0)
Missing	1627 (29.0)
Level of access to maternity care	
Moderate access to care/access to maternity care	3276 (58.3)
Maternity care desert/low access to care	244 (4.3)
Missing	2097 (37.3)
Neighborhood Atlas's ADI[b]	
≥75	919 (16.4)
≥50 and <75	1381 (24.6)
≥25 and <50	1108 (19.7)
<25	574 (10.2)

XSL•FO
RenderX

Characteristics	Values (N=5617), n (%)
Missing	1635 (29.1)

[a]GED: General Educational Development.
[b]ADI: Area Deprivation Index.

XSL•FO
RenderX

Table 3. Health history and pregnancy characteristics of PowerMom participants.

Characteristics	Values (N=5617), n (%)
Pregnancy stage at enrollment	
First trimester	1578 (28.1)
Second trimester	1493 (26.6)
Third trimester	909 (16.2)
Post delivery	736 (13.1)
Unknown	901 (16.0)
Number of fetuses	
1	3394 (60.4)
2	121 (2.2)
3	16 (0.3)
>3	15 (0.3)
Unknown	2071 (36.9)
Number of past pregnancies	
0	371 (6.6)
1	759 (13.5)
2	688 (12.2)
3	447 (8.0)
4	328 (5.8)
5	180 (3.2)
>5	266 (4.7)
Unknown	2578 (45.9)
Number of past births	
0	1118 (19.9)
1	923 (16.4)
2	571 (10.2)
3	260 (4.6)
4	111 (2)
5	58 (1)
>5	53 (0.9)
Unknown	2523 (44.9)
Prenatal care providers seen at baseline	
Doula	129 (2.3)
Homebirth	9 (0.2)
Physician	2104 (37.5)
Midwife	555 (9.9)
Nurse midwife	56 (1)
Nurse practitioner	36 (0.6)
Obstetrician	237 (4.2)
Other	143 (2.5)
None	447 (8.0)
Unknown	2511 (44.7)
Taking medications or supplements at baseline	

XSL•FO
RenderX

Characteristics	Values (N=5617), n (%)
No	261 (4.6)
Yes	2887 (51.4)
Unknown	2469 (44.0)
Receiving anxiety or depression treatment at baseline	
Anxiety and depression	550 (9.8)
Anxiety only	189 (3.4)
Depression only	102 (1.8)
Neither	2296 (40.9)
Unknown	2480 (44.2)
Parent or familial childhood condition or disease history	
No	2603 (46.3)
Yes	517 (9.2)
Unknown	2497 (44.5)
Prepregnancy condition	
No	1699 (30.2)
Yes	1449 (25.8)
Unknown	2469 (44.0)
Past pregnancy complication	
No	1575 (28.0)
Yes	1106 (19.7)
Not applicable	371 (6.6)
Unknown	2565 (45.7)
Pregnancy completion	
Gave birth	849 (15.1)
Pregnancy ended	52 (0.9)
Stillbirth	12 (0.2)
Unknown	4704 (83.7)
Delivery type	
Forceps delivery	5 (0.1)
Not applicable	64 (1.1)
Planned cesarean section	117 (2.1)
Spontaneous vaginal delivery	512 (9.1)
Unknown	4709 (83.8)
Unplanned C-section	163 (2.9)
VBAC[a]	27 (0.5)
Vacuum extraction	20 (0.4)
Delivery location	
Birthing center	18 (0.3)
Home	22 (0.4)
Hospital	807 (14.4)
Other	2 (0)

XSL•FO
RenderX

Characteristics	Values (N=5617), n (%)
Unknown	4768 (84.9)

[a]VBAC: vaginal births after cesarean.

Discussion

Overview

The PowerMom platform is a dynamic research tool designed to capture and store maternal health data, enabling the analysis of longitudinal and episodic data across all stages of pregnancy. This includes survey responses, EHR, and wearable sensor data from pregnant and postpartum individuals. PowerMom's ability to collect data from participants across all 50 US states, Puerto Rico, and Guam has facilitated the creation of a diverse dataset. The breadth and depth of these data allow for a detailed exploration of individual variations in pregnancies, helping researchers understand how these variations affect both maternal and fetal health. PowerMom's data collection supports advanced analytical approaches, including trend analysis, comparative studies, and machine learning techniques. This enables the platform to explore patterns and associations between different maternal health factors, including demographic variables, health behaviors, and outcomes, offering insights that can inform personalized approaches to maternal care.

Principal Findings

The PowerMom platform successfully engaged a diverse cohort of 5617 participants across all 50 US states, Puerto Rico, and Guam, demonstrating its effectiveness as a scalable tool for maternal health research. Of the 5617 participants, 3922 (69.8%) shared demographic data, revealing meaningful representation across age, racial and ethnic categories, and geographic regions. Nearly half (2723/5617, 48.5%) of participants were younger than 35 years, while 19.3% (1083/5617) were 35 years or older. Participants identifying as Black or African American accounted for 13.7% (770/5617) of the cohort, while Hispanic or Latina individuals represented 14.0% (788/5617). These data underscore the platform's success in reaching historically underrepresented populations in maternal health research.

Recruitment efforts improved significantly over time, with enrollment increasing from 55 participants in late 2021 to 3310 in 2024. By 2024, the platform reached an average enrollment rate of 99.4 participants per week, reflecting the impact of enhanced recruitment strategies and partnerships with apps such as Philips Pregnancy+ and Mae Health.

Comprehensive data collection was achieved through surveys, wearable devices, and EHR integration. A total of 17,123 surveys were completed across the cohort. The initial Screener Survey had a completion rate of 100%. However, subsequent surveys saw a decline in completion rates: 71.8% (4033/5617) for the Intake Survey, 56.0% (3145/5617) for the Health History Survey, and 12.4% (697/5617) for the Postpartum Survey. These trends highlight the need for strategies to sustain participant engagement over time.

Wearable device data were provided by 1168 participants, yielding more than 378,000 daily biometric measurements, including activity levels, sleep duration, and resting heart rate. These continuous data provide a granular, real-time view of maternal health throughout pregnancy and postpartum stages. In addition, 96 participants connected their EHRs to the platform, contributing 276 distinct data points, such as diagnoses, medications, and laboratory results. The rationale for the low EHR connection rate was not a metric of our study. However, we observed that the low willingness to share EHR data among participants (1.7%) is consistent with findings in the literature, such as the study by Kim et al [30]. Key factors cited in that study include privacy concerns, the lack of trust in data-sharing platforms, and technical challenges in linking EHRs. In addition, the demographic diversity within the cohort may have contributed to this hesitancy, as some populations with historical mistrust of health care systems might have been more hesitant to share sensitive information. Understanding these barriers is critical for future studies aiming to integrate EHR data more comprehensively.

The analysis of pregnancy and delivery-related characteristics further enriched the dataset. Approximately 28.1% (1578/5617) of participants enrolled during the first trimester, and 15.1% (849/5617) completed their pregnancies during the study period. Delivery data revealed that 9.1% (512/5617) of participants reported experiencing spontaneous vaginal births, while 2.1% (117/5617) and 2.9% (163/5617) underwent planned and unplanned cesarean deliveries, respectively. Mental health data indicated that 14.9% of participants reported receiving treatment for anxiety or depression at baseline.

Despite the platform's successes, notable data gaps were observed, including missing demographic information (eg, age data missing for 1811/5617, 32.2% of participants) and survey attrition. Due to attrition, only 15.1% (849/5617) of participants reported completing their pregnancies within the study period, limiting delivery-related analyses. These findings highlight both the potential and challenges of using digital platforms for maternal health research.

Comparison With Previous Work

The PowerMom platform builds upon lessons learned from our earlier digital pregnancy study, the Healthy Pregnancy Study [31]. The Healthy Pregnancy Study was embedded within a pregnancy app, allowing for seamless recruitment and the enrollment of more than 4000 participants. The study generated more than 14,000 individual surveys and more than 107,000 daily health measurements, including sleep, activity, and heart rate. The Healthy Pregnancy Study also highlighted key disparities in maternal health care access and treatment. For example, Black and rural participants reported lower use of prenatal vitamins, antiemetics, and antidepressants compared with their non-Black and urban counterparts [32]. PowerMom sought to build on these findings by enabling widespread participation from underrepresented communities and offering

XSL•FO
RenderX

valuable insights into how structural and health care factors affect maternal health outcomes.

Challenges Encountered

Substudy Enrollment Challenges

PowerMom's foundational protocol allows for the integration of additional substudies targeting specific research questions. However, when substudies have additional eligibility criteria, ensuring that participants can easily meet those criteria can be challenging. For example, incomplete baseline surveys or unmet eligibility criteria can hinder participant enrollment in substudies. Streamlining the baseline survey process and providing clear prompts for participants to join substudies have proven to be effective strategies in minimizing these challenges and boost substudy engagement.

Fraudulent Activity

One unexpected challenge during one of the PowerMom substudies was the fraudulent redemption of gift cards offered as participation incentives. The platform's automated system was exploited by 580 participants who enrolled fraudulently to claim gift cards, leading to a rapid spike in enrollment in both the baseline study and PowerMom Connect. To counter this, we introduced anomaly detection measures, including monitoring the time taken to complete eConsent and flagging participants with irregular patterns of completion. These measures successfully curtailed fraudulent activity. Future studies should incorporate automated fraud detection tools from the outset to prevent disruptions.

Attrition in Survey Completion

Participant attrition in completing surveys over time remains a key limitation of the PowerMom platform. While initial survey completion rates were high, subsequent surveys saw lower completion rates, potentially due to participant fatigue or competing priorities. Enhanced engagement strategies, such as personalized incentives, interactive content, and dynamic reminders, are needed to reduce participant burden and improve survey retention in future studies.

Limitations

While the PowerMom platform leverages decentralized and digital approaches to reach a large and diverse population, it is not without limitations. Despite the increasing penetration of smartphones in the United States (with 92% of adults owning a smartphone as of 2023 [33]), there remain disparities in broadband access and digital literacy, particularly in rural areas and among lower-income populations. These barriers may prevent certain individuals from participating fully in a purely digital, mobile-based platform. However, PowerMom's decentralized design mitigates the need for participants to live near academic medical centers. Therefore, increasing access to maternal health research for those in remote or underserved areas.

In addition, as noted in the *Attrition in Survey Completion* section, participant attrition in survey completion remains a challenge for longitudinal studies such as PowerMom. Ensuring sustained engagement throughout pregnancy and postpartum periods is critical for comprehensive data collection.

Lessons Learned

The relaunch of PowerMom in 2021 highlighted several key lessons for future digital health studies:

1. Scalability and adaptability: The platform's scalability and adaptability have proven essential in accommodating diverse participant needs. Culturally tailored outreach strategies and bilingual interfaces were particularly successful in engaging historically underrepresented populations.
2. Continuous data collection: The ability to collect real-world, continuous data from wearables and EHR systems provides a comprehensive view of maternal health that surpasses traditional, episodic clinic visits. This real-time data collection allows for more precise tracking of health changes over time.
3. Fraud detection and prevention: The experience of dealing with fraudulent activity emphasized the importance of implementing robust anomaly detection systems to protect the integrity of participant data and incentives. Future studies should incorporate automated fraud detection tools from the outset to avoid disruptions.
4. Participant retention: Attrition in survey completion is a challenge in longitudinal digital health studies. Future studies should incorporate strategies to maintain participant engagement, such as personalized incentives, interactive content, or dynamic reminders, which may help reduce participant fatigue and improve data completeness.

Conclusions

The PowerMom platform has demonstrated its potential as an innovative, scalable, and inclusive maternal health research tool. By leveraging decentralized, mobile-based recruitment and data collection, PowerMom successfully engaged a diverse cohort of 5617 participants across all 50 US states, Puerto Rico, and Guam. The platform achieved representation across critical demographic groups, including underrepresented racial and ethnic populations and individuals residing in areas with limited access to maternity care. Despite the platform's successes, challenges such as survey attrition, missing data, and fraudulent activity underscored the importance of implementing robust engagement and data quality strategies in future digital health studies.

Looking ahead, PowerMom's flexible and scalable architecture positions it as a transformative model for maternal health research. By continuing to enhance participant engagement strategies and expanding data collection capabilities to include genomic, environmental, and social determinants of health, PowerMom can generate actionable insights to address maternal health disparities. Ultimately, this platform represents a significant step toward improving maternal health outcomes, promoting health equity, and empowering diverse populations to actively participate in research that directly impacts their care.

XSL•FO
RenderX

Acknowledgments

This work received partial support from the National Center for Advancing Translational Sciences (UM1TR004407) and The Patrick J McGovern Foundation.

Data Availability

The datasets generated and analyzed during the PowerMom study are not publicly available due to privacy and confidentiality restrictions. The study collected sensitive health and demographic data from participants, which are deidentified but remain protected under institutional review board protocols. Researchers who wish to access the data for legitimate academic purposes can submit a formal request to PowerMom principal investigator, TA. Requests will be evaluated on a case-by-case basis to ensure compliance with ethical guidelines and data use agreements. The code used for data analysis is available upon request from the corresponding author.

Conflicts of Interest

KBM is a consultant for CareEvolution.

Multimedia Appendix 1

Participant flow and in-app surveys from the PowerMom study, conducted between 2021 and 2024. This figure illustrates the step-by-step flow of participants from recruitment through study completion, including survey engagement. Surveys captured demographic, health history, and pregnancy-related data at various stages, leveraging the PowerMom app's user-friendly interface to enhance participant engagement and data collection.

[DOCX File , 12642 KB - formative_v9i1e70149_app1.docx]

References

1. Hoyert DL. Maternal mortality rates in the United States, 2020. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nchs/data/hestat/maternal-mortality/2020/maternal-mortality-rates-2020.htm [accessed 2024-12-02]
2. Hirai AH, Owens PL, Reid LD, Vladutiu CJ, Main EK. Trends in severe maternal morbidity in the US across the transition to ICD-10-CM/PCS from 2012-2019. JAMA Netw Open 2022;5(7):e2222966 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.22966] [Medline: 35900764]
3. Petersen EE, Davis NL, Goodman D, Cox S, Syverson C, Seed K, et al. Racial/ethnic disparities in pregnancy-related deaths - United States, 2007-2016. MMWR Morb Mortal Wkly Rep 2019;68(35):762-765 [FREE Full text] [doi: 10.15585/mmwr.mm6835a3] [Medline: 31487273]
4. Creanga AA, Syverson C, Seed K, Callaghan WM. Pregnancy-related mortality in the United States, 2011-2013. Obstet Gynecol 2017;130(2):366-373 [FREE Full text] [doi: 10.1097/AOG.0000000000002114] [Medline: 28697109]
5. Iturralde E, Hsiao CA, Nkemere L, Kubo A, Sterling SA, Flanagan T, et al. Engagement in perinatal depression treatment: a qualitative study of barriers across and within racial/ethnic groups. BMC Pregnancy Childbirth 2021;21(1):512 [FREE Full text] [doi: 10.1186/s12884-021-03969-1] [Medline: 34271852]
6. Admon LK, Winkelman TNA, Zivin K, Terplan M, Mhyre JM, Dalton VK. Racial and ethnic disparities in the incidence of severe maternal morbidity in the United States, 2012-2015. Obstet Gynecol 2018;132(5):1158-1166. [doi: 10.1097/AOG.0000000000002937] [Medline: 30303912]
7. Bromley E, Nunes A, Phipps MG. Disparities in pregnancy healthcare utilization between Hispanic and non-Hispanic white women in Rhode Island. Matern Child Health J 2012;16(8):1576-1582. [doi: 10.1007/s10995-011-0850-5] [Medline: 21769588]
8. Chantler TEA, Lees A, Moxon ER, Mant D, Pollard AJ, Fiztpatrick R. The role familiarity with science and medicine plays in parents' decision making about enrolling a child in vaccine research. Qual Health Res 2007;17(3):311-322. [doi: 10.1177/1049732306298561] [Medline: 17301340]
9. Ayoub JJ, Abiad M, Forman MR, MINA Collaborators, Honein-AbouHaidar G, Naja F. The interaction of personal, contextual, and study characteristics and their effect on recruitment and participation of pregnant women in research: a qualitative study in Lebanon. BMC Med Res Methodol 2018;18(1):155 [FREE Full text] [doi: 10.1186/s12874-018-0616-5] [Medline: 30497391]
10. Chinn JJ, Eisenberg E, Artis Dickerson S, King RB, Chakhtoura N, Lim IAL, et al. Maternal mortality in the United States: research gaps, opportunities, and priorities. Am J Obstet Gynecol 2020;223(4):486-492.e6 [FREE Full text] [doi: 10.1016/j.ajog.2020.07.021] [Medline: 32682858]
11. Zullig LL, Fortune-Britt AG, Rao S, Tyree SD, Godley PA, Carpenter WR. Enrollment and racial disparities in cancer treatment clinical trials in North Carolina. N C Med J 2016;77(1):52-58 [FREE Full text] [doi: 10.18043/ncm.77.1.52] [Medline: 26763244]

XSL•FO
RenderX

12. Woodall A, Morgan C, Sloan C, Howard L. Barriers to participation in mental health research: are there specific gender, ethnicity and age related barriers? BMC Psychiatry 2010;10:103 [FREE Full text] [doi: 10.1186/1471-244X-10-103] [Medline: 21126334]
13. Williams B, Entwistle V, Haddow G, Wells M. Promoting research participation: why not advertise altruism? Soc Sci Med 2008;66(7):1451-1456. [doi: 10.1016/j.socscimed.2007.12.013] [Medline: 18222579]
14. Choudhury A, Nimbarte A. Editorial: mobile health interventions to address maternal health: ideas, concepts, and interventions. Front Digit Health 2024;6:1378416 [FREE Full text] [doi: 10.3389/fdgth.2024.1378416] [Medline: 38486918]
15. Stone JS, Hirshberg A. Telemedicine and digital health solutions in intrapartum and postpartum care. Curr Obstet Gynecol Rep 2024;13:188-194 [FREE Full text] [doi: 10.1007/s13669-024-00394-x]
16. Nadarzynski T, Knights N, Husbands D, Graham CA, Llewellyn CD, Buchanan T, et al. Achieving health equity through conversational AI: a roadmap for design and implementation of inclusive chatbots in healthcare. PLOS Digit Health 2024;3(5):e0000492 [FREE Full text] [doi: 10.1371/journal.pdig.0000492] [Medline: 38696359]
17. Leonard A, Hutchesson M, Patterson A, Chalmers K, Collins C. Recruitment and retention of young women into nutrition research studies: practical considerations. Trials 2014;15:23 [FREE Full text] [doi: 10.1186/1745-6215-15-23] [Medline: 24433229]
18. Frampton GK, Shepherd J, Pickett K, Griffiths G, Wyatt JC. Digital tools for the recruitment and retention of participants in randomised controlled trials: a systematic map. Trials 2020;21(1):478 [FREE Full text] [doi: 10.1186/s13063-020-04358-3] [Medline: 32498690]
19. Bower P, Brueton V, Gamble C, Treweek S, Smith CT, Young B, et al. Interventions to improve recruitment and retention in clinical trials: a survey and workshop to assess current practice and future priorities. Trials 2014;15:399 [FREE Full text] [doi: 10.1186/1745-6215-15-399] [Medline: 25322807]
20. Cummins KM, Brumback T, Chung T, Moore RC, Henthorn T, Eberson S, et al. Acceptability, validity, and engagement with a mobile app for frequent, continuous multiyear assessment of youth health behaviors (mNCANDA): mixed methods study. JMIR Mhealth Uhealth 2021;9(2):e24472 [FREE Full text] [doi: 10.2196/24472] [Medline: 33565988]
21. Quer G, Gouda P, Galarnyk M, Topol EJ, Steinhubl SR. Inter- and intraindividual variability in daily resting heart rate and its associations with age, sex, sleep, BMI, and time of year: Retrospective, longitudinal cohort study of 92,457 adults. PLoS One 2020;15(2):e0227709. [doi: 10.1371/journal.pone.0227709] [Medline: 32023264]
22. Jaiswal SJ, Quer G, Galarnyk M, Steinhubl SR, Topol EJ, Owens RL. Association of sleep duration and variability with body mass index: sleep measurements in a large US population of wearable sensor users. JAMA Intern Med 2020;180(12):1694-1696 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2834] [Medline: 32926073]
23. Quer G, Radin JM, Gadaleta M, Baca-Motes K, Ariniello L, Ramos E, et al. Wearable sensor data and self-reported symptoms for COVID-19 detection. Nat Med 2021;27(1):73-77. [doi: 10.1038/s41591-020-1123-x] [Medline: 33122860]
24. Gadaleta M, Radin JM, Baca-Motes K, Ramos E, Kheterpal V, Topol EJ, et al. Passive detection of COVID-19 with wearable sensors and explainable machine learning algorithms. NPJ Digit Med 2021;4(1):166 [FREE Full text] [doi: 10.1038/s41746-021-00533-1] [Medline: 34880366]
25. Radin JM, Quer G, Ramos E, Baca-Motes K, Gadaleta M, Topol EJ, et al. Assessment of prolonged physiological and behavioral changes associated with COVID-19 infection. JAMA Netw Open 2021;4(7):e2115959 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.15959] [Medline: 34232306]
26. Quer G, Gadaleta M, Radin JM, Andersen KG, Baca-Motes K, Ramos E, et al. Inter-individual variation in objective measure of reactogenicity following COVID-19 vaccination via smartwatches and fitness bands. NPJ Digit Med 2022;5(1):49 [FREE Full text] [doi: 10.1038/s41746-022-00591-z] [Medline: 35440684]
27. New research consortium to tackle maternal health disparities with digital technologies. Scripps Research. 2021. URL: https://www.scripps.edu/news-and-events/press-room/2021/20211207-ajayi-powermom-consortium.html [accessed 2024-12-02]
28. Access to maternity care. MODHealth. 2024. URL: https://mod.health360-insights.io [accessed 2024-12-02]
29. Multiple births. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nchs/fastats/multiple.htm [accessed 2024-12-02]
30. Kim KK, Sankar P, Wilson MD, Haynes SC. Factors affecting willingness to share electronic health data among California consumers. BMC Med Ethics 2017;18(1):25 [FREE Full text] [doi: 10.1186/s12910-017-0185-x] [Medline: 28376801]
31. Radin JM, Steinhubl SR, Su AI, Bhargava H, Greenberg B, Bot BM, et al. The healthy pregnancy research program: transforming pregnancy research through a ResearchKit app. NPJ Digit Med 2018;1:45 [FREE Full text] [doi: 10.1038/s41746-018-0052-2] [Medline: 31304325]
32. Ajayi T, Pawelek J, Bhargava H, Faksh A, Radin J. Self-reported medication use across racial and rural or urban subgroups of people who are pregnant in the United States: decentralized app-based cohort study. JMIR Form Res 2023;7:e50867 [FREE Full text] [doi: 10.2196/50867] [Medline: 38015604]
33. Mobile fact sheet. Pew Research Center. 2024. URL: https://tinyurl.com/y3ykwta4 [accessed 2024-09-10]

XSL•FO
RenderX

Abbreviations

ADI: Area Deprivation Index
eConsent: electronic informed consent
EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act

Edited by A Mavragani; submitted 16.12.24; peer-reviewed by B Senst; comments to author 20.01.25; revised version received 16.02.25; accepted 25.02.25; published 07.04.25.

Please cite as:
Ajayi T, Kueper J, Ariniello L, Ho D, Delgado F, Beal M, Waalen J, Baca Motes K, Ramos E
Digital Health Platform for Maternal Health: Design, Recruitment Strategies, and Lessons Learned From the PowerMom Observational Cohort Study
JMIR Form Res 2025;9:e70149
URL: https://formative.jmir.org/2025/1/e70149
doi:10.2196/70149
PMID:

©Toluwalase Ajayi, Jacqueline Kueper, Lauren Ariniello, Diana Ho, Felipe Delgado, Matthew Beal, Jill Waalen, Katie Baca Motes, Edward Ramos. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Use of Mobile Apps in Adolescent Psychotherapy: Assessment of Psychotherapists' Perspectives

Sarah Wüllner[1], Dr rer nat; Katharin Hermenau[1], Dr rer nat; Tobias Hecker[2,3], Dr rer nat; Michael Siniatchkin[1,4], Dr med

[1]University Clinic of Child and Adolescent Psychiatry and Psychotherapy, Medical School EWL, Protestant Hospital Bethel, Bielefeld University, Bielefeld, Germany

[2]Institute for Interdisciplinary Conflict and Violence Research, Bielefeld University, Bielefeld, Germany

[3]Department of Psychology, Faculty of Psychology and Sports Science, Bielefeld University, Bielefeld, Germany

[4]Department of Psychiatry, Psychotherapy and Psychosomatics of Children and Adolescents, University Hospital Rheinisch-Westfälische Technische Hochschule Aachen, Aachen, Germany

Corresponding Author:
Sarah Wüllner, Dr rer nat
University Clinic of Child and Adolescent Psychiatry and Psychotherapy
Medical School EWL, Protestant Hospital Bethel
Bielefeld University
Remterweg 13a
Bielefeld, 33617
Germany
Phone: 49 5211063166
Email: sarah.wuellner@uni-bielefeld.de

Abstract

Background: Therapy-accompanying mental health apps can play an important role in the psychotherapeutic treatment of adolescents. They can enhance adolescents' engagement and autonomy, provide immediate support in critical situations, and positively influence the therapeutic working alliance. Nevertheless, mental health apps are rarely used by psychotherapists. Furthermore, due to the limited or nonexistent use of apps in psychotherapy, little is known about the actual barriers and drivers affecting their integration into psychotherapists' daily routines. To better understand how mental health apps should be designed for practical use, it is essential to explore psychotherapists' perspectives on key app features and characteristics, as well as the factors influencing their integration into clinical practice.

Objective: This study aims to analyze which app features and characteristics are essential for psychotherapists to use a mobile app in psychotherapy with adolescents and to identify the key drivers and barriers influencing the integration of a psychotherapeutic app from the psychotherapists' perspectives.

Methods: We conducted 3 feasibility studies using Steps, a transdiagnostic, therapy-accompanying app for adolescents, across 3 different psychotherapeutic treatment contexts: inpatient treatment, treatment in psychiatric outpatient clinics, and outpatient treatment with psychotherapists in private practice. All studies followed a qualitative quasi-experimental design. Participants provided information on their age, occupation, years of work experience, media affinity, attitudes toward psychotherapeutic apps, perceived app quality and feasibility, and the implementation process of the therapy-accompanying app. Qualitative data were analyzed using deductive qualitative content analysis. A total of 40 mental health professionals participated across the 3 studies (study 1: n=18; study 2: n=13; study 3: n=9).

Results: Study participation and app usage rates were low across all studies. Six core features for a transdiagnostic, therapy-accompanying app were identified: mood checks, library, reminders, goals and tasks, emergency kit, and questionnaires. Additionally, the integration of mental health apps into daily routines was influenced by various drivers and barriers. The most significant barriers included technological issues and practical constraints, such as limited time and resources. The most important driver was the perceived improvement in treatment quality.

Conclusions: Overall, psychotherapists were generally open to using a therapy-accompanying mental health app. However, study participation and app usage remained low. As psychotherapists act as gatekeepers for patients' use of mental health apps, their needs should be prioritized in the development and implementation of such apps.

Trial Registration: German Clinical Trials Register DRKS00031258; https://drks.de/search/en/trial/DRKS00031258/details

XSL•FO
RenderX

(JMIR Form Res 2025;9:e65788) doi:10.2196/65788

KEYWORDS

mental health app; psychotherapy; adolescent; mHealth; youth; feasibility; implementation; app features; barriers; drivers

Introduction

Background

There is a high risk of developing mental disorders during adolescence, which may persist into adulthood [1-3]. According to the World Health Organization (WHO), one in seven 10- to 19-year-olds experiences a mental disorder [4]. In addition, the rates of treatment nonresponse and therapy dropout are high among adolescents [5-7]. A promising way to address these challenges in psychotherapeutic care is the use of therapy-accompanying mental health apps (MHAs) [8-10]. At the latest, the pandemic has demonstrated that MHAs can play an important role in the psychotherapeutic treatment of adolescents [11-13] and are widely accepted by them [11,14,15].

Therapy-accompanying MHAs are used as an adjunct to psychotherapy [16]. They can offer specific manualized treatments or therapy tools that can be individually incorporated into psychotherapy [17,18]. The majority of available MHAs focus on specific disorders rather than a transdiagnostic therapy approach [12,17-19]. However, most MHAs have demonstrated the potential to provide transdiagnostic support beyond their primarily targeted disorder, as each contains fundamentally transdiagnostic features [18,20,21]. Additionally, to date, all therapy-accompanying apps have been designed for outpatient treatment [16,18]. To the best of our knowledge, no MHA has been specifically designed for inpatient treatment.

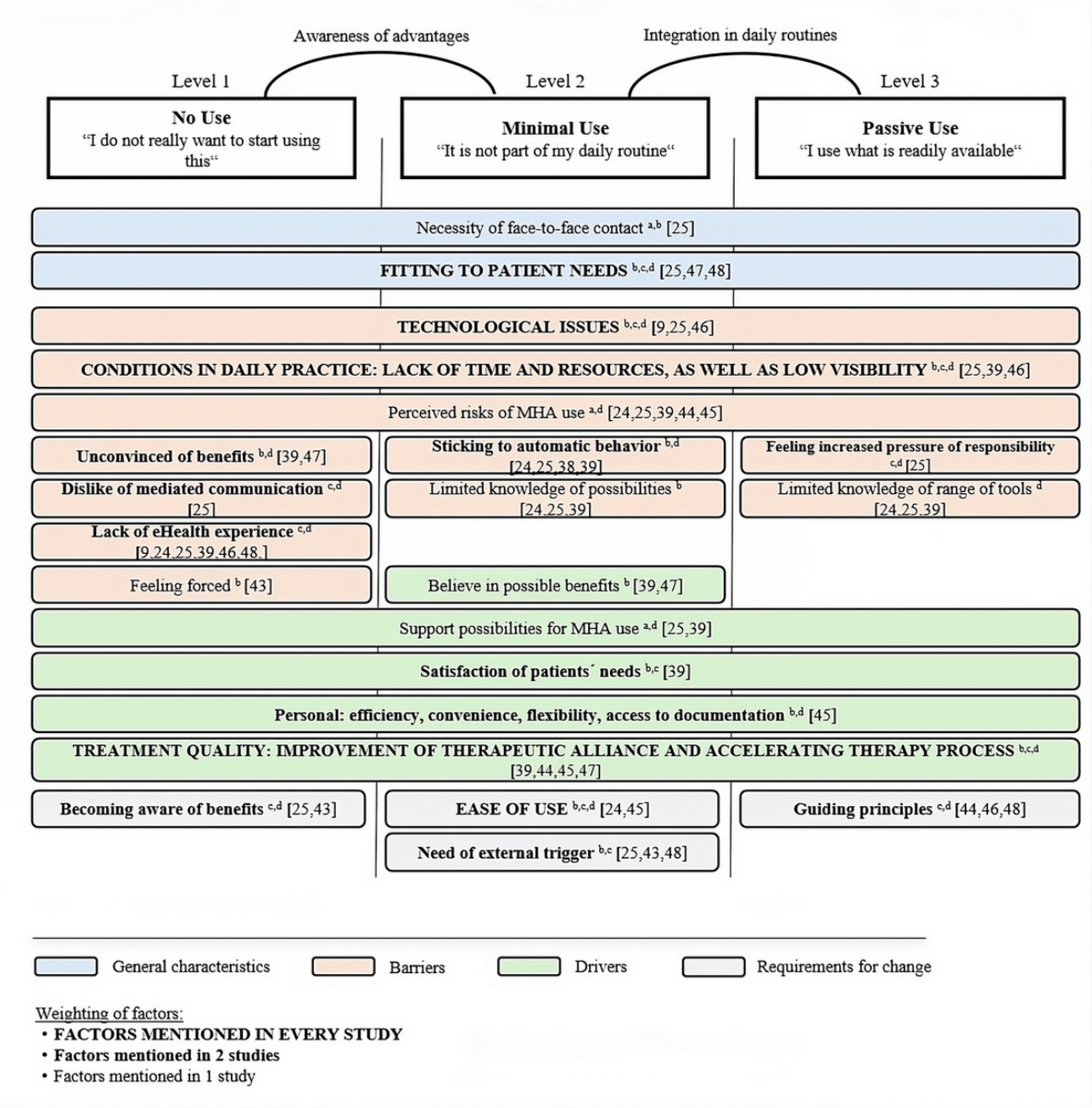
Therapy-accompanying MHAs offer several advantages in adolescent psychotherapy. They can enhance engagement [8-10] and promote greater autonomy [22,23]. Moreover, MHAs provide immediate support in critical situations, such as self-harm or acute suicidal ideation [16]. As a result, adolescents experience increased self-efficacy and improved therapeutic transfer to everyday life [8,16,17]. Furthermore, MHAs can have a beneficial influence on the therapeutic working alliance [8]. For psychotherapists, MHAs can facilitate documentation and monitoring of therapeutic progress, as well as communication with patients [24]. Another important aspect of integrating MHAs into psychotherapy is their effectiveness [24,25]. Previous research has shown that MHAs, in general, are effective in adolescent psychotherapy [18,26-35].

Although there are many reasons to use MHAs as an adjunct to psychotherapy with adolescents, they are rarely utilized by psychotherapists [24,36-41]. In 2019, only 9% of German health care professionals had tried an MHA in their psychotherapeutic practice [24]. The COVID-19 pandemic led to a slight increase in usage [37,39]; however, adoption remains low, with only 29% of German health care professionals reporting that they have used an app in their practice [39]. This raises the question: “Where does the discrepancy between the perceived benefits of MHAs and their limited use in psychotherapy originate?”

To better understand this discrepancy, it is essential to identify the factors associated with the integration of MHAs into psychotherapists’ working routines. Feijt et al [42] developed a model to explain the factors influencing psychotherapists’ use of e-mental health: the Levels of Adoption of e-Mental Health (LAMH) model. This model categorizes influencing factors into different domains, including general characteristics, perceived drivers and barriers, and requirements for change. All factors are categorized according to different levels of e-mental health usage. The authors distinguish 5 levels of use, ranging from “no use” (level 1) to “innovative use” (level 5). According to previous research [24,39], psychotherapists generally fall between level 1 (no use) and level 3 (passive users) in their use of MHAs [42]. Consequently, our study focuses on the lower levels of adoption (levels 1-3), excluding level 4 (active use) and level 5 (innovative use). The selection of general characteristics, perceived drivers and barriers, and requirements for change mentioned in Feijt et al’s [42] is based on qualitative research results. Nevertheless, these factors have also been identified in other studies [9,24,25,39,43-48]. The LAMH model, adapted for this study and supplemented by our findings, is presented in Figure 1. It also includes corresponding references for each factor in the adapted LAMH model.

XSL•FO
RenderX

Figure 1. Levels of Adoption of e-Mental Health (LAMH) adapted for mental health app (MHA) use (aLAMH). The model is adapted from the LAMH model of Feijt et al. Factors that are marked with an "a" were added to the original model based on the current study results. Factors mentioned in study 1 are marked with "b"; factors mentioned in study 2 are marked with "c"; and factors mentioned in study 3 are marked with "d." The factors in the model were also mentioned in other studies as relevant association factors with psychotherapists' mental health app use.



Like the development of the LAMH model, previous research on drivers and barriers has primarily focused on the theoretical perspectives of psychotherapists [9,24,44,45,47]. Furthermore, due to the limited or nonexistent use of apps in psychotherapy [37,39], little is known about the actual barriers and drivers influencing the integration of MHAs into psychotherapists' working routines. Therefore, gaining a more practical understanding of these factors is essential.

In addition to questions about potential drivers and barriers to MHA integration, the design and features of MHAs may play a crucial role in psychotherapists' use of these apps. What if psychotherapists do not use MHAs because the available apps do not sufficiently meet the needs of adolescent psychotherapeutic treatments? To better understand how MHAs should be designed for actual use in psychotherapists' working routines, it is essential to explore their perspectives on the key features and characteristics of therapy-accompanying MHAs.

Objectives

This study aims is to investigate how a therapy-accompanying MHA should be designed for successful integration into adolescent psychotherapy across different psychotherapeutic contexts. To assess psychotherapists' perspectives in a more practical way, we conducted a needs evaluation within the initial feasibility pilot testing of a transdiagnostic, therapy-accompanying app for adolescents in 3 different treatment settings: inpatient treatment, psychiatric outpatient

XSL•FO
RenderX

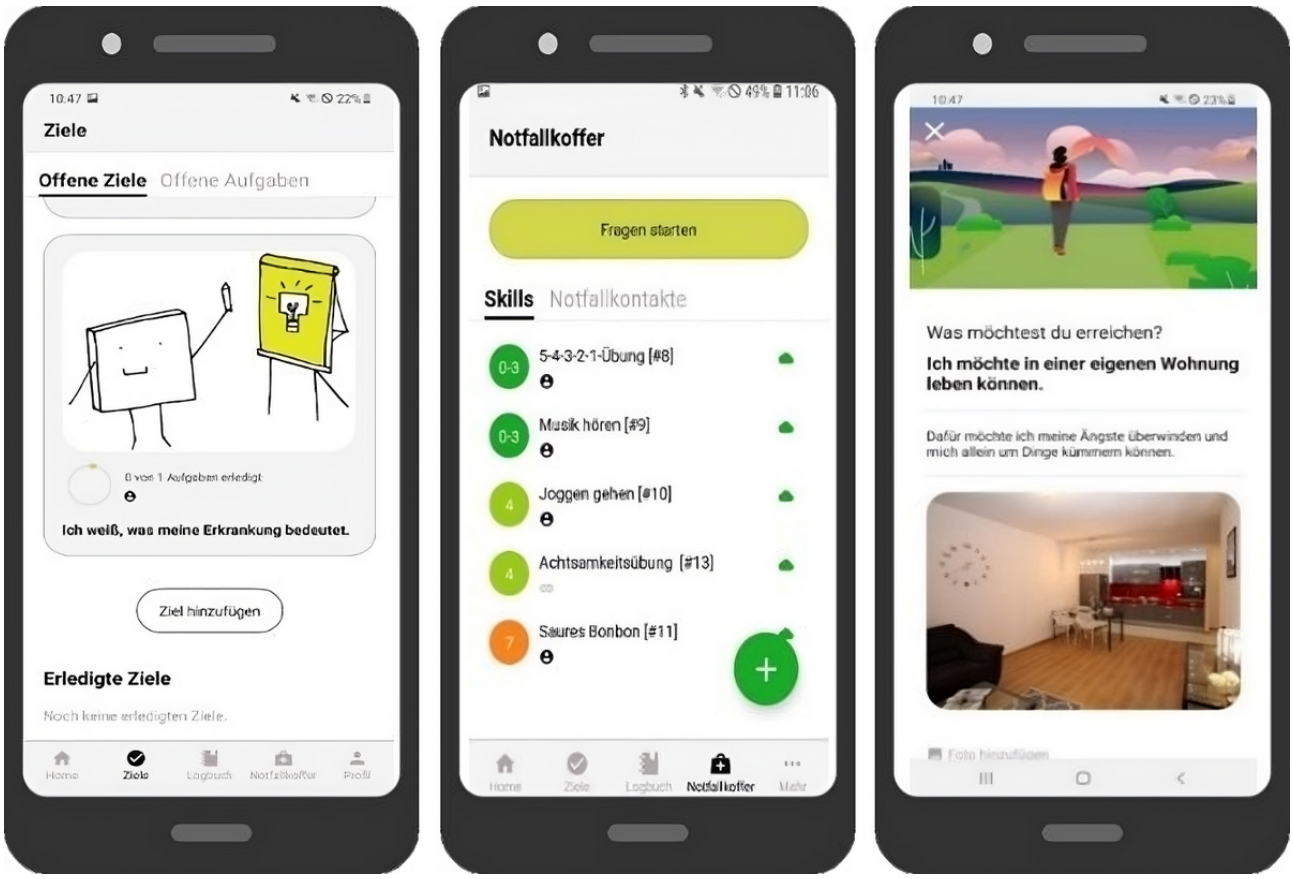
clinics, and outpatient treatment with private practice psychotherapists. We analyzed the essential app features and characteristics required for psychotherapists to use an MHA in adolescent psychotherapy, as well as the key drivers and barriers influencing its integration from the psychotherapists' perspectives.

Methods

Overview

Three feasibility studies were conducted in different settings using the transdiagnostic, therapy-accompanying app *Steps*. The app was designed as an adjunct to regular psychotherapeutic treatment for adolescents. It was developed through a participatory process by Circumradius GmbH and the Protestant Hospital Bethel in Bielefeld. Steps offered interconnected versions for both patients and their psychotherapists. While patients used the Steps app on their own Android-based smartphones, psychotherapists accessed a web-based version (see Figures 2 and 3). The app included various transdiagnostic features (Table 1). For implementation, psychotherapists were not required to follow specific usage rules or treatment manuals; they could independently decide to what extent they wanted to integrate the app into their psychotherapy sessions. All studies followed a qualitative quasi-experimental study design. In all studies, data were collected from both psychotherapists and patients. However, due to the objectives of this article and the comprehensiveness of the collected data, patient data were not included in the current analyses. Additionally, only methods and results relevant to this analysis are reported in this article.

Figure 2. Screenshots of different app features of the patient version of the app Steps: goals and tasks (left), emergency kit (center), and the main therapy goal (right).



XSL•FO
RenderX

Figure 3. Displays of different app features of the web-based psychotherapist version of the app Steps: list of skills in the emergency kit (left) and goals and tasks (right).

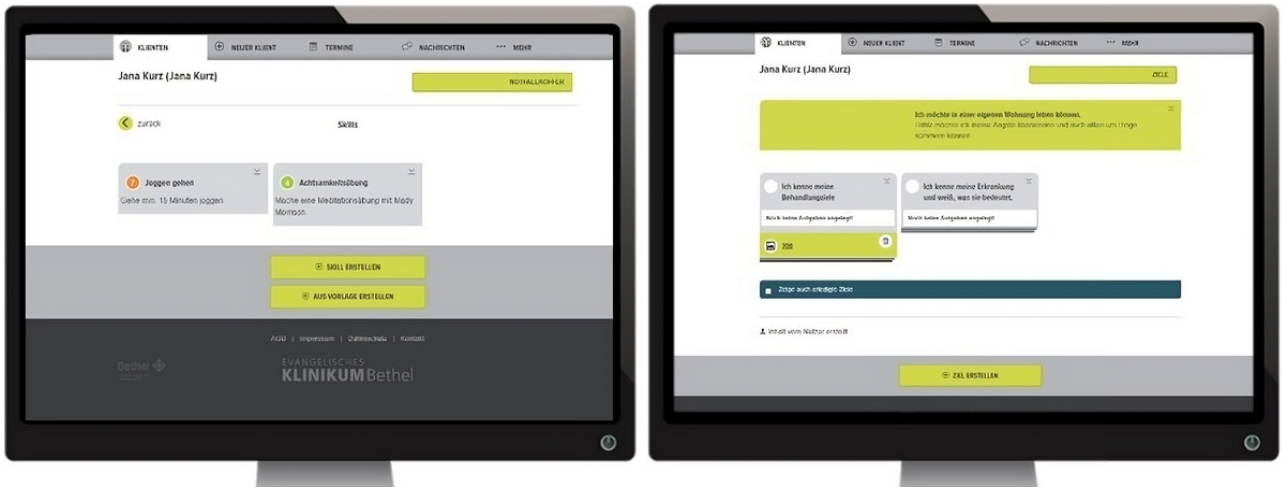


Table 1. Overview of the provided app features in the therapy-accompanying and transdiagnostic app Steps for adolescent psychotherapy.

App feature	Description
Mood checks	• Psychotherapists can customize mood checks for patients (type and number of items, response format, frequency of queries). • Patients are prompted and reminded by the app to complete the mood checks. • In the psychotherapist version, a graphical overview of the mood checks is available.
Main therapy goal	• Patients can record what they want to achieve through their psychotherapy. • Photos can be uploaded to visualize patients' therapy motivation.
Intermediate therapy goals	• Intermediate therapy goals can be set to structure or visualize therapy progress. • Patients can complete individual tasks to achieve intermediate goals.
Tasks	• Psychotherapists can set tasks for their patients, for example, to complete between therapy sessions. • When patients have completed a task, they receive an automated reflection query. Therapists can directly access the information entered and easily refer to the situation-related information in the next therapy session.
Logbook	• Mood checks done by patients are saved in the logbook. • Patients can also create free logbook entries to record important events, or to document the therapy progress.
Questionnaire manager	• Psychotherapists can set up questionnaires for their patients to complete directly in the app. • Questionnaires can be customized by psychotherapists (type and number of items, response format, and frequency of queries). • Answers to questionnaires can be downloaded directly in the psychotherapist version.
Emergency kit	• Skills and emergency contacts can be saved in the emergency kit. • Skills are sorted according to stress levels from 1 (low stress) to 10 (maximum stress). • Photos, home page links, or address details can be added to skill entries. • Optional: activation of queries for analysis of undesirable behavior (eg, nonsuicidal self-injuries).

Design and Procedure

The first study examined the feasibility of using Steps in inpatient treatment for adolescents. Adolescents admitted as inpatients who consented to participate in the study received a clinic-owned smartphone for their treatment. This smartphone included the Steps app along with other treatment-relevant apps and was available for use throughout their entire treatment period. The entire staff of the adolescent ward was trained in app usage and instructed to integrate Steps into the treatment process. All employees were invited to participate in the study on a voluntary basis. Psychotherapists completed a baseline survey at the start of their participation and a follow-up questionnaire at the end of the data collection period. For other employees, all questions were consolidated into a single survey conducted at the end of the project. Additionally, 4 employees participated in qualitative interviews assessing the app's quality and the implementation process of Steps.

The second study was conducted in psychiatric outpatient clinics. All staff members from the participating clinics were invited to take part. Based on the participation challenges

XSL•FO
RenderX

encountered in the first study, psychotherapists were offered 2 options for participation: (1) actively using the app in treatment with adolescents and providing feedback on their experiences, or (2) providing feedback on Steps after a brief introduction without directly using the program. Each participant received individual app training to familiarize themselves with the app and its application in therapy. If psychotherapists incorporated the app into their treatment, participating patients were required to install Steps on their own smartphones. If needed, adolescents could borrow a clinic-owned smartphone for study participation. Data were collected from participating employees at the start of their study involvement and again at the end of the data collection period using the open-source online survey tool LimeSurvey (LimeSurvey Team) [49]. The study was preregistered in the German Clinical Trials Register (DRKS00031258).

The third study was conducted with psychotherapists in private practice. Interested psychotherapists were recruited through the Association of Child and Adolescent Psychotherapists of East Westphalia-Lippe. The study procedure was consistent with that of the second study, with one difference: data for the second assessment were collected through qualitative interviews.

Ethical Considerations

All 3 studies received positive ethical approvals. Studies 1 and 2 were approved by the Ethics Committee of the Medical Association of Westphalia-Lippe (approval numbers 2020-878-f-S and 2022-528-f-S, respectively). Additionally, study 2 was approved by the Ethics Committee II of the University of Heidelberg (approval number 2022-643) and the Ethics Commission of the Faculty of Medicine at the University of Cologne (approval number 23-1026_1). Study 3 was approved by the Ethics Committee of Bielefeld University (approval number EUB-2023-050). In all 3 studies, participants received detailed study information sheets outlining the study procedures, research aims, and data protection guidelines. Participation was voluntary and confidential. Psychotherapists signed informed consent before their participation. To ensure data security, all participant data were collected in a pseudonymized manner. Personal data (eg, contact details or audiotapes) were stored separately from the pseudonymized data. Audiotapes of the qualitative interviews were transcribed. Once data analysis is complete, the audiotapes will be destroyed, and the data will be anonymized. Participants did not receive any compensation. This study followed the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist for study conceptualization, data analysis, and reporting.

Participants

A total of 40 mental health professionals participated in the 3 studies. Sample characteristics varied based on the treatment context. In study 1, participants included 18 employees from a psychiatric ward for adolescents at a German clinic specializing in child and adolescent psychiatry and psychotherapy. In study 2, 13 employees from 3 different psychiatric outpatient clinics in Germany took part. In study 3, 9 psychotherapists in private practice who treated adolescents aged 12-18 years were included. Detailed sample characteristics are presented in Table 2.

Table 2. Overview of the sample characteristics presented separately for each study context.

Sample characteristics	Study 1: Inpatient treatment (n=18 participants)	Study 2: Psychiatric outpatient clinics (n=13 participants)	Study 3: Psychotherapy in private practices (n=9 participants)
Included occupational groups	• Psychotherapists (n=3) • Specialist therapists (eg, music therapists, n=3) • Nursing professionals (n=11) • Youth volunteer (n=1)	• Psychotherapists (n=9) • Psychologist without further training (n=1) • Assistant doctors (n=3)	• Psychotherapists (n=9)
Age (years), mean (SD); range	• 32.50 (9.59); 22-62	• 35.15 (9.05); 27-61	• 54.44 (13.79); 33-72
Work experience (years), mean (SD); range	• 8.33 (9.46); 0.60-40.50	• 6.08 (5.12); 2.33-16.42	• 23.44 (14.34); 6-40

Measures

Survey Methodology and Questionnaire Design

In all studies, surveys were conducted using self-administered questionnaires, which included both quantitative and qualitative questions. Participants provided information on their age, occupation, years of work experience, media affinity, attitudes toward psychotherapeutic apps, perceived app quality and feasibility, and the implementation process of Steps. Most questions were purpose-built. Because of variations in study contexts and insights gained from previous studies, some questions and questionnaires differed between the studies. Detailed differences in the surveys are presented in Multimedia Appendix 1. In the inpatient treatment context (study 1), psychotherapists received different questionnaires than other employees to account for the distinct responsibilities and working areas of each occupational group.

Media Affinity and Attitudes Toward Psychotherapeutic Apps

In study 1, media affinity was assessed using the Affinity for Technology (TA-EG) questionnaire [50]. The TA-EG consists of 19 items across 4 subscales: enthusiasm for technology, competence in using technology, perceived positive consequences of technology, and perceived negative consequences of technology. Participants rated each item on a 5-point Likert scale from 1 (does not apply at all) to 5 (fully applies). Subscale scores were calculated as the mean of the associated items, ranging from 1 to 5 points. The internal

XSL•FO
RenderX

consistency values for the TA-EG questionnaire are presented in Table 3. Based on the experiences from study 1 and the differing working conditions in outpatient therapy settings, media affinity was assessed differently in studies 2 and 3. Participants answered self-created questions about the technical equipment available at their workplace, their use of media and smartphone apps, and the integration of media into their psychotherapies. Additionally, in study 3, participants completed the Affinity for Technology Interaction (ATI) questionnaire [51], a unidimensional measurement instrument consisting of 9 items. Participants rated each item on a 5-point Likert scale from 1 (not true at all) to 5 (completely true). The overall ATI score was calculated as the mean value of all items, ranging from 1 to 5. The internal consistency of the ATI is presented in Table 3.

Table 3. Means, SDs, ranges, and reliability coefficients of the questionnaires on media affinity, attitudes toward e-mental health, and app quality from psychotherapists' perspectives. The data are presented separately for each study context.

Questionnaire	n	Mean (SD)	Range	Actual scale range	Internal consistencies (Cronbach α)
Study 1					
TA-EG[a]					
Enthusiasm	18	2.93 (0.77)	1.60-4.40	1-5	0.90
Competencies	18	3.42 (0.62)	2.50-4.75	1-5	0.63
Positive consequences	18	4.01 (0.37)	3.40-4.80	1-5	0.57
Negative consequences	18	3.02 (0.63)	1.80-4.20	1-5	0.83
Study 2					
uMARS[b]					
Overall	10	3.75 (0.60)	2.77-4.52	1-5	0.87-0.90[c]
Engagement	11	3.76 (0.69)	2.70-4.80	1-5	0.76-0.82[c]
Functionality	10	3.77 (0.85)	2.38-5.00	1-5	0.70-0.88[c]
Aesthetics	10	3.76 (0.59)	3.00-4.58	1-5	0.37-0.47[c]
Study 3					
ATI[d]	9	2.98 (0.99)	1.56-4.67	1-5	0.92
MTPS[e]					
Potential to augment psychotherapy	9	3.56 (0.46)	3.00-4.25	1-5	0.42
Perceived risks	9	2.97 (0.85)	2.00-4.75	1-5	0.78
uMARS					
Overall	5	3.52 (0.56)	2.55-3.87	1-5	0.66-0.82[c]
Engagement	9	3.79 (0.66)	3.00-4.86	1-5	0.50-0.82[c]
Functionality	5	3.54 (0.90)	2.14-4.45	1-5	0.57-0.60[c]
Aesthetics	9	3.58 (0.86)	2.30-5.00	1-5	0.58-0.76[c]

[a]TA-EG: Affinity for Technology [50].

[b]uMARS: user version of the Mobile Application Rating Scale [52].

[c]Internal consistencies of the app *Steps* (summarized for the app version and the web version of the app).

[d]ATI: Affinity for Technology Interaction [51].

[e]MTPS: Psychotherapists' Attitudes Toward Using Modern Technologies in Psychotherapy and Counselling Scale [53].

Given the specific study context of inpatient treatment (study 1), attitudes toward psychotherapeutic apps were assessed using 12 self-created questions based on the Unified Theory of Acceptance and Usability of Technology (UTAUT) [54] and an adapted UTAUT questionnaire by Hennemann et al [14]. Participants responded to the items on a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). In study 2, the self-created questions were adapted to the outpatient study context. As a result, attitudes toward psychotherapeutic apps were measured using a 10-item questionnaire. In study 3, attitudes toward psychotherapeutic apps were measured using the Psychotherapists' Attitudes Toward Using Modern Technologies in Psychotherapy and Counselling Scale (MTPS) [52]. The MTPS is a multidimensional self-report questionnaire

XSL•FO
RenderX

with 16 items across 4 subscales: *potential to augment psychotherapy, psychoeducational value, perceived risks*, and *perceived relevance*. We included 2 subscales—*potential to augment psychotherapy* (eg, "Modern technologies can speed up the therapeutic process") and *perceived risks* (eg, "A stricter regulation of the content available via modern technologies in relation to psychotherapy and psychopathology should be enforced"). Both scales are considered central factors in practitioners' attitudes toward technology [52,55]. Psychotherapists rated their agreement with the statements on a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). Subscale scores were calculated as the mean values of the 4 associated items in each subscale, with scores ranging from 1 to 5. For this study, the MTPS was translated into German. The quality of the translation was ensured through multiple forward and backward translations. The internal consistencies of the subscales are presented in Table 3.

Perceived App Quality and Feasibility

As study 1 was the first evaluation study on Steps, perceived app quality and feasibility were assessed using purpose-built open questions, informed by previous evaluation studies [14,56-58]. The questions focused on app functionality, design, available features, and aspects of the implementation process (eg, "Steps makes my therapeutic work easier" or "The use of Steps fits well into everyday life on the ward"). To obtain a more distinct and structured assessment of perceived app quality and feasibility, 4 subscales of the user version of the Mobile Application Rating Scale (uMARS) [53] were used in studies 2 and 3: engagement, functionality, aesthetics, and subjective app quality. The uMARS is a self-report questionnaire that was translated into German for this study, with accuracy ensured through blind back-translation. Participants rated 17 items on a 5-point Likert scale (1=totally disagree and 5=fully agree). They evaluated both the patient version of the app and the web-based psychotherapist version of Steps. Subscale scores for *engagement*, *functionality*, and *aesthetics* were calculated as the mean value of the associated items for each subscale, ranging from 1 to 5. Five additional questions addressed *subjective app quality* (eg, "What is your overall [star] rating of the app?"). Overall app quality scores for both the app version and the web version were determined by the mean value of all items. The internal consistencies are presented in Table 3. In studies 2 and 3, the feasibility of Steps was assessed using purpose-built questions based on those used in study 1 but adapted to the contexts of psychiatric outpatient clinics and private practice psychotherapists.

Data Analysis

Qualitative interviews from studies 1 and 3 were transcribed. All qualitative data, including interview transcripts and free-text responses from the surveys, were analyzed using deductive qualitative content analysis [59]. An initial set of categories was developed based on the uMARS and LAMH model before data analysis began. In the first step, 1 researcher coded the qualitative data. Following the deductive qualitative content analysis process described by Cho and Lee [59], data that did not fit into the predetermined categories were assigned to newly created categories. The final set of categories for the content analysis is presented in Multimedia Appendix 2. All qualitative data were recoded using the complete set of categories. Following Mayring [60], a second rater coded a subset of the qualitative data (2 qualitative interviews) to ensure coding clarity. Any discrepancies between the 2 raters were thoroughly discussed. The frequency of each category mentioned by participants, as well as the number of participants who reported each category, was assessed. Categories were ranked and reported based on their occurrence. For this publication, participant quotes from the study were translated into English. Descriptive quantitative data were analyzed using SPSS version 29.0.0.0 (IBM Corp.).

Results

Study Participation and Attitudes Toward MHAs

Study participation was low across all studies (Table 2). Despite extensive recruitment efforts, only a few psychotherapists chose to test the transdiagnostic, therapy-accompanying app Steps as part of the studies. Additionally, patient participation was low, and dropout rates were high. Detailed information on study participation and dropout rates is provided in Multimedia Appendix 3. In all studies, mental health professionals reported moderate media affinity and mixed attitudes toward MHAs. In studies 2 (psychiatric outpatient clinics) and 3 (psychotherapists in private practice), most psychotherapists had never recommended MHAs to patients before (study 2: 9/12 participants; study 3: 8/9 participants). Frequencies of the self-created questions from studies 2 and 3 regarding the technical equipment available at work, the use of media and smartphone apps, and the integration of media into psychotherapies are provided in Multimedia Appendix 3. Descriptive statistics for the standardized questionnaires used in the studies are presented in Table 3.

Perceived App Quality and Feasibility of Steps

For inpatient treatment of adolescents, about half of the participants in study 1 rated the joy of use neutrally (n=10; fun to use: n=4, not fun to use: n=2) but found Steps helpful for patients (n=9; undecided: n=5, unhelpful: n=3). However, the app was rarely used for psychotherapeutic treatment. Almost all participants agreed that the app did not fit the inpatient context. Eight participants suggested that Steps might be more beneficial for outpatient treatments. Correspondingly, 1 participant mentioned using the app for follow-up care after patient discharge from the clinic.

> *I could imagine, [Steps] perhaps as a transitional app [...] to somehow bridge the time before therapy or as aftercare after therapy. [...]* [WI2]

By contrast, for outpatient treatment of adolescents, the majority of participants expressed interest in continuing to use the app for some patients after the project ended (study 2—yes, n=9; no, n=3; and missing, n=1; and study 3—yes, n=7 and no, n=2). Additionally, most psychotherapists indicated a willingness to recommend Steps to some colleagues (study 2—few, n=1; some, n=5; many, n=3; everyone, n=3; and missing, n=1; and study 3—few, n=1; some, n=5; many, n=2; and missing, n=1). However, while 5 psychotherapists in private practice were

XSL•FO
RenderX

willing to pay for app use (no, n=4), only 1 participant from the psychiatric outpatient clinics expressed willingness to pay (no, n=10). Furthermore, Steps received an average star rating of 3.55 (SD 0.65; range 2.00-4.00; n=11) in study 2 and 3.19 (SD 0.53; range 2.50-4.00; n=8) in study 3. These ratings were comparable to those of other MHAs [21]. Descriptive statistics for the uMARS are presented in Table 3.

Essential App Features

In the inpatient setting, the most frequently mentioned feature was setting therapy goals and tasks (n=4). While psychotherapists rated this feature as unhelpful and disliked the emphasis it placed on goals and tasks in the therapy process, nursing professionals and specialist therapists found it interesting to use.

> *I liked the visualization of the therapy process and successes [through the goals and tasks feature* [W18]

The second most frequently mentioned feature in the inpatient setting was the mood check (n=3). Additionally, setting reminders was identified as an important feature (n=4), with participants expressing a desire for greater flexibility.

> *It should be possible to set a reminder as flexible as in calendar apps [the current app version provided periods that were not detailed enough]* [W18]

The results from the outpatient therapy settings presented a different perspective. Employees of psychiatric outpatient clinics identified mood checks (mentioned 8 times by 5) and the emergency kit (mentioned 5 times by 5) as the most helpful features. Additionally, the questionnaire manager was considered useful for diagnostics if templates were available (mentioned 4 times by 4). However, participants were unwilling to spend time digitizing their own questionnaires.

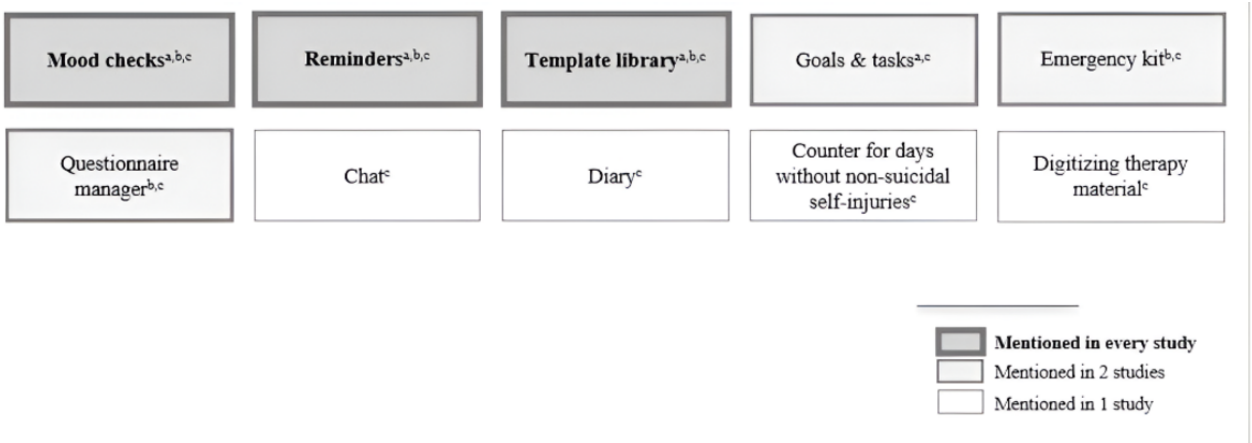
> *[Providing questionnaires was] far too laborious to enter them yourself.* [C14]

The availability of templates was also highlighted for other app features. Three participants from psychiatric outpatient clinics suggested a library feature to provide psychoeducational materials, tests, skills, or templates for mood checks. Additionally, psychotherapists in private practices identified mood checks (n=8), the library (n=8), and the questionnaire manager as the most helpful features of a therapy-accompanying app. Additionally, psychotherapists in private practices highlighted the option to digitize therapy materials (n=8) and facilitate communication with patients through an app as particularly helpful. However, the communication feature was also perceived as challenging. Three psychotherapists expressed concerns about legal provisions and responsibilities, particularly in cases where patients might communicate indicators of suicidality. Furthermore, a chat feature was seen as potentially leading to an “increased workload that is not practicable” (P1) for psychotherapists. Similarly, feedback on reminders was mixed. While the majority of psychotherapists found it a helpful feature (n=6), 1 psychotherapist noted that it could potentially contradict therapeutic processes:

> *I somehow also expect patients to remember appointments and perhaps manage this well via a calendar or whatever, and otherwise it’s actually a good indication of motivation for therapy for me.* [P2]

Other app features suggested by psychotherapists in private practices included a tool for tracking days without nonsuicidal self-injury, the provision of relaxation exercises, and support for organizational therapy tasks, such as sending documents or completing necessary forms and questionnaires. An overview of the key app features mentioned across the 3 studies is presented in Figure 4.

Figure 4. Essential app features for a transdiagnostic, therapy-accompanying app mentioned in the different studies. Features mentioned in study 1 are marked with "a"; features mentioned in study 2 are marked with "b"; and features mentioned in study 3 are marked with "c.".



Essential App Characteristics

Across all study contexts, ease of use and customizability were identified as crucial app characteristics. Ease of use was mentioned by several participants in each study (study 1, n=4; study 2, n=6; and study 3, n=9). Half of the interviewed participants from the inpatient treatment setting emphasized the importance of customizability (study 2, n=4 and study 3, n=8). However, some participants from study 1 highlighted the benefits of preinstalled therapy tools and materials, such as activity lists, meditation and relaxation exercises, and mindfulness practices. Psychotherapists in private practices particularly valued the ability to adapt the app to different psychotherapeutic approaches and individual workflows, considering it a key strength of Steps.

XSL•FO
RenderX

> *also open to different psychotherapeutic approaches...to offer psychotherapists a tool...that...yes, incorporates modern technologies into the treatment and makes it usable for both psychotherapists and patients at the same time, so I think that's an advantage of Steps.* [P6]

However, the ability to customize the app individually was also identified as a potential barrier by psychotherapists in private practices. They noted that digitizing materials and preparing personalized app usage require additional time. In line with this concern, 1 participant from the psychiatric outpatient clinics suggested implementing an easy copy function to facilitate the creation of custom app content (C1).

In addition, accessibility was highlighted as another crucial app characteristic. The absence of an iOS version was cited as the primary reason for not using Steps by 3 practitioners from psychiatric outpatient clinics and 3 psychotherapists in private practices. However, 1 participant from the psychiatric outpatient clinics and 8 out of 9 psychotherapists in private practices positively emphasized the benefit of both psychotherapists and patients having individual yet interrelated accounts within Steps.

> *I also think it's good that there's a psychotherapist app that I can use to control things to a certain extent. I think it's good to be able to provide the patient with materials. Yes, I think so, for now.* [P2]

Regarding a therapy-accompanying app for inpatient treatment, all participants in the qualitative interviews emphasized that individual access to each professional role is essential (n=4).

> *Yes, and then I might have had more of a connection to it [with my access]. So, because I only had access via the adolescents, it was forgotten a bit quicker* [WI3]

The design was another crucial factor, with 5 participants from the inpatient therapy setting directly mentioning its importance. Additionally, 3 out of 4 interviewed participants from study 1 emphasized that design serves as a motivator for usage. In the outpatient therapy settings, design was also highlighted as significant. Two practitioners from the psychiatric outpatient clinics reported feeling less motivated to use Steps due to the unattractive design of the psychotherapist version. Feedback on the patient version of the app was mixed. On the one hand, the "[...]individual design and yet reduced to the essentials[...]" (C17) was mentioned positively. On the other hand, C18 suggested that the "layout could be more colorful" to motivate adolescents to use the app. In the sample of psychotherapists in private practices, 3 psychotherapists suggested a more playful design, such as a customizable avatar or a reward system, to make the app more appealing to adolescents. Additionally, a participant proposed incorporating motivational features into the psychotherapist version to enhance engagement.

> *Maybe there's a nice seductiveness, so to speak. That a psychotherapist who doesn't like looking at smartphones likes to look at them, [...].* [P3]

Aspects for the Integration of MHAs Into Psychotherapists' Daily Routines

In all studies, one of the most common barriers was *conditions in daily practice* (study 1: mentioned 25 times by 12 in the surveys and several times by 4; study 2: mentioned by 1; and study 3: mentioned 23 times by 8). In the inpatient therapy setting, other common barriers included *sticking to automatic behavior* (mentioned 9 times by 6) in the surveys, as well as being *unconvinced of benefits* (mentioned 14 times by 3) and *sticking to automatic behavior* (mentioned 7 times by 3) in the qualitative interviews. Regarding the barrier "sticking to automatic behavior", WI1 of the inpatient therapy setting said:

> *How do I get what I have there now into this app, where I intuitively know how to proceed [without an app]? I didn't have time for that, and the adolescents didn't even think about it.*

In line with the barriers mentioned in the inpatient therapy setting, the barrier "technical issues" was also reported by 2 practitioners in the psychiatric outpatient clinics, while the barrier "unconvinced of benefits" was mentioned 16 times by 7 psychotherapists in private practice. Additionally, psychotherapists in private practice also identified "perceived risks" as one of the most common barriers to app usage (mentioned 12 times by 6). In the psychiatric outpatient clinics, "technical issues" was likewise mentioned as one of the most common barriers by 2 practitioners.

The most common drivers varied across study contexts. In the inpatient therapy setting, the most frequently mentioned driver in both surveys and interviews was *approving of treatment quality*. Other drivers mentioned once included *satisfaction of client's needs*,*personal efficiency*, and *belief in possible benefits*. By contrast, participants from the psychiatric outpatient clinics identified the potential to digitize aspects of psychotherapy as a driver for app use and an indicator of its suitability for client needs.

> *[...]because paper pencil questionnaires/mood checks etc. can be replaced. Young people are much more interested in this.* [C5]

Psychotherapists in private practices cited *personal efficiency* (15 times by 7), *approving of treatment quality* (14 times by 7), and *support offers for app use* (3 times by 3) as the 3 most common drivers. Regarding the general characteristics of app usage, all participants in the qualitative interviews of study 1 (n=4) and 8 out of 9 psychotherapists in private practices highlighted the app's fit to client needs as a crucial factor for its use. Additionally, 2 psychotherapists in private practices emphasized the necessity of face-to-face contact during the relationship-building phase at the beginning of therapy as a general characteristic. Regarding the requirements for change, all interviewed participants in the inpatient therapy setting mentioned the need for external triggers to integrate an app into their daily work routines.

> *[It needs someone who] can consistently accompany such a project and also pick up [everyone of] the colleagues.* [WI1]

XSL•FO
RenderX

An overview of all mentioned drivers, barriers, general characteristics, and requirements for change in all 3 studies is presented in Figure 1.

Discussion

Principal Findings

Psychotherapists play an important role in integrating MHAs into adolescent psychotherapies [38,61]. However, previous research has indicated that apps are rarely used in psychotherapeutic treatments, even though mental health professionals are generally interested in MHAs [36-41]. In line with this, participating psychotherapists reported positive attitudes toward MHAs, but app usage rates were low across all study contexts. Furthermore, the results of the 3 studies indicated that psychotherapists had little to no experience using MHAs for psychotherapeutic treatment. To highlight the discrepancy between positive attitudes toward MHAs and their low level of use in psychotherapeutic treatment, we assessed, on the one hand, psychotherapists' needs for a therapy-accompanying MHA and, on the other hand, potential factors influencing the integration of MHAs into their daily working routines.

Examining MHAs themselves, the current studies identified 6 core features for a transdiagnostic therapy-accompanying app: mood checks, template library, reminders, goals and tasks, emergency kit, and questionnaire manager. Considering patients' perspectives from previous research, they also reported similar core features for MHAs [10,47,62,63]. Psychotherapists highlighted accessibility, design, and customizability as important app characteristics. Consistent with previous research, accessibility remains a key characteristic of MHAs for both patients and psychotherapists [45,63,64]. Psychotherapists in the current studies highlighted the interrelatability of the app Steps. They found it helpful to have complete access to adolescents' entries, as well as the opportunity to provide additional content. Moreover, accessibility for adolescents was also mentioned. As Steps was available only for Android-based smartphones, the operating system was frequently cited as a key exclusion factor. Across all 3 studies, the impact of app design on both adolescents' and psychotherapists' engagement was particularly emphasized. MHAs need to be designed to be motivating and incentivizing to ensure long-term use for therapy support [62]. Another core app characteristic mentioned in every study was the customizability of MHAs. For example, psychotherapists in study 3 highlighted that digitizing their therapy materials was an important factor in their use of Steps. They appreciated the ability to continue working with their own materials rather than being restricted to content provided by the app. Additionally, a high degree of customizability allows psychotherapists to use an app independently of their specific psychotherapeutic approaches. Most available MHAs are based on cognitive behavioral theory [18], which creates higher barriers to adoption for psychotherapists who align with other therapeutic approaches [25,55,61,65].

Limited experience with MHAs is not only a result of psychotherapists' behavior but also a significant barrier to their integration into daily routines. Previous research suggests that limited experience with e-mental health is often accompanied by a lack of competency and only moderate knowledge of the advantages and possibilities of MHAs in psychotherapy. Consequently, limited experience with MHAs can lead to increased aversion toward their use [9,24,25,38,39,45,46]. In line with this, the majority of psychotherapists in studies 2 and 3 reported that they had never used or recommended MHAs before. Moreover, the stated requirements for change—*becoming aware of benefits* and *ease of use*—were reported multiple times in every study as key factors for MHA use. Consistent with the LAMH model of Feijt et al [42], aspects of ease of use were the most frequently mentioned in the current studies. All participants emphasized that MHAs need to be easy to understand and intuitive to use. Psychotherapists specifically highlighted the need for an app that is both easy to understand and simple to use, as daily routines leave little capacity for intensive familiarization with new methods such as MHAs. In Germany, the use of MHAs for psychotherapeutic treatment is rarely reimbursed by health insurance companies [66]. That may also contribute to the low willingness to engage with MHAs. Other important influencing factors were related to perceived benefits, such as "improvement of treatment quality" as a driver and "being unconvinced of benefits" as a barrier. In accordance with previous research, psychotherapists need to be convinced of the advantages of MHA use for both patients and their own workflow [8,24,39,45,47]. Psychotherapists reported being more likely to adopt MHAs if the app facilitated their working processes, particularly the organizational tasks of psychotherapy. Additionally, fitting the app to patients' needs and maintaining face-to-face contact were also identified as general characteristics in the current studies. From the psychotherapists' perspective, the usefulness of an app depends on the clients' needs and the phase of psychotherapy. Particularly during the initial stages, when establishing a therapeutic relationship, psychotherapists would not introduce a therapy-accompanying app. They believe that building a strong relationship first is essential for the successful implementation of MHAs. This aligns with previous studies indicating that psychotherapeutic support is a crucial factor in patients' engagement with apps [8].

Beyond various personal factors influencing psychotherapists' use of MHAs, the current studies also highlighted organizational factors. One significant and common barrier was technological issues. Consistent with previous research, adequate technical infrastructure—such as a stable internet connection and proper devices—is essential for the adoption of MHAs [9,25,46]. In studies 2 and 3, fewer than half of the psychotherapists reported using a smartphone or tablet at work. Limited access to technical equipment may contribute to the perceived difficulty of using MHAs. Furthermore, in line with previous research, psychotherapists expressed concerns about data security [24,39,44,45]. Their concerns extended beyond data protection within the apps to include transparent clinical standards and guiding principles as essential requirements for MHA use in psychotherapy [24,25,39,45,48]. Although apps are increasingly being implemented in health care and supported by policy makers [66], many psychotherapists still do not feel sufficiently confident in using MHAs [9,44,62]. Additionally, in our study, psychotherapists reported uncertainties regarding legal

XSL•FO
RenderX

provisions and responsibilities, such as those related to an optional communication feature or the provision of digitized diagnostic questionnaires. Additionally, the need for external triggers was mentioned as a requirement. In line with previous research, participating psychotherapists suggested features such as a reminder function in the therapist version of MHAs to enhance engagement [24,25]. Furthermore, the perceived commitment of the entire organization and management was highlighted as an important motivational factor for successfully and sustainably integrating MHAs into daily routines.

Practical Implications and Future Research

The study results provide insights into which app features and characteristics are essential for psychotherapists in different psychotherapeutic contexts. Future app development should build on these findings and ensure that essential app features and characteristics are incorporated into MHAs. Designing MHAs based on psychotherapists' needs will facilitate their integration into daily routines. App developers must strike a balance between maximum customizability and ease of use. For example, MHAs could offer a wide selection of templates and therapy materials through a library feature, along with the option to digitize psychotherapists' own materials to expand the existing library. However, future research and app development should not focus solely on psychotherapists' perspectives and needs but should also consider the perspectives and needs of adolescents. Moreover, the integration of MHAs into daily routines is influenced by various drivers and barriers. The most significant barriers to integrating MHAs were technological issues and conditions in daily practice, such as a lack of time and resources. The most important driver was the perceived improvement in treatment quality. To successfully integrate MHAs, implementation strategies need to be adjusted. The LAMH model of Feijt et al [42] provides a useful framework for understanding common drivers, barriers, general characteristics, and requirements for change, which were supported by our study results. These identified drivers and barriers must be directly addressed in implementation strategies. Psychotherapists who have not previously worked with MHAs require close support to implement app-based services into their daily treatment routines. Additionally, organizational changes are necessary to successfully integrate MHAs into the health care system. Psychotherapists need sufficient time and resources to familiarize themselves with MHAs and learn how to incorporate them into their practice. Moreover, adequate technical equipment and stable internet connections are essential for the effective integration of MHAs.

Strengths and Limitations

The current studies have several strengths and limitations. One of the most significant strengths is the exploration of psychotherapists' perspectives while they actively tested a transdiagnostic, therapy-accompanying app in their regular treatments. This approach allowed psychotherapists to go beyond theoretical considerations of potential drivers, barriers, or necessary app features. Instead, they gained firsthand experience with MHAs, identifying which features they found helpful and which were lacking in their work. Furthermore, to the best of our knowledge, this is the first study exploring the use of MHAs in the inpatient treatment of adolescents. However, the present studies also have some limitations. The small sample sizes in all 3 studies suggest that the app Steps was not well implemented in practice, limiting the generalizability of the findings to other settings. Additionally, the small sample sizes and the recruitment method pose a risk of selection bias. For instance, psychotherapists with a greater interest in technological developments may have been more likely to participate in the studies.

Conclusions

In sum, psychotherapists in this study were generally open to therapy-accompanying MHAs. However, both study participation and app usage were low. Psychotherapists identified various drivers and barriers influencing their use of MHAs, highlighting that adoption is shaped not only by personal factors but also by organizational challenges. To enable psychotherapists to integrate MHAs into their practice, changes in current working conditions are necessary. For example, they require adequate technical equipment and sufficient time to familiarize themselves with new therapy methods such as MHAs. Successfully integrating MHAs into daily psychotherapeutic routines is a challenge for the entire health care system, not just for individual psychotherapists.

Acknowledgments

The feasibility studies were funded by the v.Bodelschwinghsche Stiftungen Bethel (study 1), the Federal Ministry of Education and Research (BMBF; study 2), and the Bielefeld University (study 3). We acknowledge support for the publication costs by the Open Access Publication Fund of Bielefeld University and the Deutsche Forschungsgemeinschaft (DFG). We want to thank our project partners Circumradius for their great cooperation in app development and Dr.-Ing Birte Richter and the medical assistance working group of Bielefeld University for their great cooperation in study 3. Furthermore, we want to thank all project teams in Bielefeld, Mannheim, and Cologne for participation and support; KJP OWL e.V. for recruitment support; and all participants for sharing their experiences with the app *Steps*.

Data Availability

The data sets generated or analyzed during this study are available from the corresponding author upon reasonable request.

XSL•FO
RenderX

Disclaimer

No generative AI tools were used in the 3 studies and the manuscript writing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of surveys presented separately for each study.
[PDF File (Adobe PDF File), 773 KB - formative_v9i1e65788_app1.pdf]

Multimedia Appendix 2
The final set of categories for the deductive qualitative content analysis.
[PDF File (Adobe PDF File), 38 KB - formative_v9i1e65788_app2.pdf]

Multimedia Appendix 3
A flowchart of study participation is presented for each study, along with the frequencies of self-created questions regarding the technical equipment available at work, the use of media and smartphone apps, and the integration of media in psychotherapies for studies 2 and 3.
[PDF File (Adobe PDF File), 80 KB - formative_v9i1e65788_app3.pdf]

References

1. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun 01;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
2. Roza SJ, Hofstra MB, van der Ende J, Verhulst FC. Stable prediction of mood and anxiety disorders based on behavioral and emotional problems in childhood: a 14-year follow-up during childhood, adolescence, and young adulthood. Am J Psychiatry 2003 Dec;160(12):2116-2121. [doi: 10.1176/appi.ajp.160.12.2116] [Medline: 14638580]
3. Sacco R, Camilleri N, Eberhardt J, Umla-Runge K, Newbury-Birch D. A systematic review and meta-analysis on the prevalence of mental disorders among children and adolescents in Europe. Eur Child Adolesc Psychiatry 2022 Dec 30;33(9):2877-2894. [doi: 10.1007/s00787-022-02131-2] [Medline: 36581685]
4. World Health Organization. Mental health of adolescents. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health [accessed 2024-06-17]
5. de Soet R, Vermeiren RRJM, Bansema CH, van Ewijk H, Nijland L, Nooteboom LA. Drop-out and ineffective treatment in youth with severe and enduring mental health problems: a systematic review. Eur Child Adolesc Psychiatry 2024 Oct;33(10):3305-3319. [doi: 10.1007/s00787-023-02182-z] [Medline: 36882638]
6. de Haan AM, Boon AE, de Jong JTVM, Hoeve M, Vermeiren RRJM. A meta-analytic review on treatment dropout in child and adolescent outpatient mental health care. Clin Psychol Rev 2013 Jul;33(5):698-711. [doi: 10.1016/j.cpr.2013.04.005] [Medline: 23742782]
7. Gearing RE, Schwalbe CS, Short KD. Adolescent adherence to psychosocial treatment: mental health clinicians' perspectives on barriers and promoters. Psychother Res 2012 May;22(3):317-326. [doi: 10.1080/10503307.2011.653996] [Medline: 22313513]
8. Diano F, Sica LS, Ponticorvo M. A systematic review of mobile apps as an adjunct to psychological interventions for emotion dysregulation. Int J Environ Res Public Health 2023 Jan 12;20(2):20. [doi: 10.3390/ijerph20021431] [Medline: 36674189]
9. Feijt M, de Kort Y, Bongers I, Bierbooms J, Westerink J, IJsselsteijn W. Mental health care goes online: practitioners' experiences of providing mental health care during the COVID-19 pandemic. Cyberpsychol Behav Soc Netw 2020 Dec;23(12):860-864. [doi: 10.1089/cyber.2020.0370] [Medline: 32815742]
10. Gindidis S, Stewart SE, Roodenburg J. Adolescent experiences of app-integrated therapy. The Educational and Developmental Psychologist 2020 Oct 19;37(1):20-29. [doi: 10.1017/edp.2019.18]
11. Rauschenberg C, Schick A, Goetzl C, Roehr S, Riedel-Heller SG, Koppe G, et al. Social isolation, mental health, and use of digital interventions in youth during the COVID-19 pandemic: a nationally representative survey. Eur Psychiatry 2021 Mar 09;64(1):e20 [FREE Full text] [doi: 10.1192/j.eurpsy.2021.17] [Medline: 33686930]
12. Ellis LA, Meulenbroeks I, Churruca K, Pomare C, Hatem S, Harrison R, et al. The application of e-mental health in response to COVID-19: scoping review and bibliometric analysis. JMIR Ment Health 2021 Dec 06;8(12):e32948 [FREE Full text] [doi: 10.2196/32948] [Medline: 34666306]

XSL•FO
RenderX

13. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
14. Hennemann S, Beutel ME, Zwerenz R. Drivers and barriers to acceptance of web-based aftercare of patients in inpatient routine care: a cross-sectional survey. J Med Internet Res 2016 Dec 23;18(12):e337 [FREE Full text] [doi: 10.2196/jmir.6003] [Medline: 28011445]
15. Breil B, Dederichs M, Kremer L, Richter D, Angerer P, Apolinário-Hagen J. [awareness and use of digital health services in Germany: a cross-sectional study representative of the population]. Gesundheitswesen 2021 Dec;83(12):1019-1028. [doi: 10.1055/a-1335-4245] [Medline: 33862648]
16. Diano F, Ponticorvo M, Sica LS. Mental health mobile apps to empower psychotherapy: a narrative review. New York, NY: IEEE; 2022 Dec 05 Presented at: The IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE); October 26-28, 2022; Rome, Italy p. 306-311. [doi: 10.1109/MetroXRAINE54828.2022.9967663]
17. Melbye S, Kessing LV, Bardram JE, Faurholt-Jepsen M. Smartphone-based self-monitoring, treatment, and automatically generated data in children, adolescents, and young adults with psychiatric disorders: systematic review. JMIR Ment Health 2020 Oct 29;7(10):e17453 [FREE Full text] [doi: 10.2196/17453] [Medline: 33118950]
18. Wüllner S, Hermenau K, Krutkova M, Petras IK, Hecker T, Siniatchkin M. Mobile applications in adolescent psychotherapy during the COVID-19 pandemic: a systematic review. Front Public Health 2024;12:1345808 [FREE Full text] [doi: 10.3389/fpubh.2024.1345808] [Medline: 38420028]
19. Lui JHL, Marcus DK, Barry CT. Evidence-based apps? A review of mental health mobile applications in a psychotherapy context. Professional Psychology: Research and Practice 2017 Jun;48(3):199-210. [doi: 10.1037/pro0000122]
20. Gindidis S, Stewart S, Roodenburg J. A systematic scoping review of adolescent mental health treatment using mobile apps. Advances in Mental Health 2018 Sep 28;17(2):161-177. [doi: 10.1080/18387357.2018.1523680]
21. Domhardt M, Messner EM, Eder AS, Engler S, Sander LB, Baumeister H, et al. Mobile-based interventions for common mental disorders in youth: a systematic evaluation of pediatric health apps. Child Adolesc Psychiatry Ment Health 2021 Sep 13;15(1):49 [FREE Full text] [doi: 10.1186/s13034-021-00401-6] [Medline: 34517896]
22. d'Halluin A, Costa M, Morgiève M, Sebbane D. Attitudes of children, adolescents, and their parents toward digital health interventions: scoping review. J Med Internet Res 2023 May 02;25:e43102. [doi: 10.2196/43102] [Medline: 37129931]
23. Liverpool S, Mota CP, Sales CMD, Čuš A, Carletto S, Hancheva C, et al. Engaging children and young people in digital mental health interventions: systematic review of modes of delivery, facilitators, and barriers. J Med Internet Res 2020 Jun 23;22(6):e16317 [FREE Full text] [doi: 10.2196/16317] [Medline: 32442160]
24. Mayer G, Gronewold N, Alvarez S, Bruns B, Hilbel T, Schultz JH. Acceptance and expectations of medical experts, students, and patients toward electronic mental health apps: cross-sectional quantitative and qualitative survey study. JMIR Ment Health 2019 Nov 25;6(11):e14018 [FREE Full text] [doi: 10.2196/14018] [Medline: 31763990]
25. Davies F, Shepherd HL, Beatty L, Clark B, Butow P, Shaw J. Implementing web-based therapy in routine mental health care: systematic review of health professionals' perspectives. J Med Internet Res 2020 Jul 23;22(7):e17362 [FREE Full text] [doi: 10.2196/17362] [Medline: 32706713]
26. Zhou X, Edirippulige S, Bai X, Bambling M. Are online mental health interventions for youth effective? A systematic review. J Telemed Telecare 2021 Dec;27(10):638-666. [doi: 10.1177/1357633X211047285] [Medline: 34726992]
27. Ly KH, Topooco N, Cederlund H, Wallin A, Bergström J, Molander O, et al. Smartphone-supported versus full behavioural activation for depression: a randomised controlled trial. PLoS One 2015 May;10(5):e0126559 [FREE Full text] [doi: 10.1371/journal.pone.0126559] [Medline: 26010890]
28. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
29. Rauschenberg C, Schick A, Hirjak D, Seidler A, Paetzold I, Apfelbacher C, et al. Evidence synthesis of digital interventions to mitigate the negative impact of the COVID-19 pandemic on public mental health: rapid meta-review. J Med Internet Res 2021 Mar 10;23(3):e23365 [FREE Full text] [doi: 10.2196/23365] [Medline: 33606657]
30. Fischer-Grote L, Fössing V, Aigner M, Fehrmann E, Boeckle M. Effectiveness of online and remote interventions for mental health in children, adolescents, and young adults after the onset of the COVID-19 pandemic: systematic review and meta-analysis. JMIR Ment Health 2024 Mar 05;11:e46637 [FREE Full text] [doi: 10.2196/46637] [Medline: 38315524]
31. Haaf R, Vock P, Wächtershäuser N, Correll CU, Köhler S, Klein JP. [Efficacy of internet-based interventions for depression available in Germany-a systematic review and meta-analysis]. Nervenarzt 2024 Mar;95(3):206-215 [FREE Full text] [doi: 10.1007/s00115-023-01587-0] [Medline: 38260995]
32. Moshe I, Terhorst Y, Philippi P, Domhardt M, Cuijpers P, Cristea I, et al. Digital interventions for the treatment of depression: a meta-analytic review. Psychol Bull 2021 Aug;147(8):749-786. [doi: 10.1037/bul0000334] [Medline: 34898233]
33. Schröder D, Wrona KJ, Müller F, Heinemann S, Fischer F, Dockweiler C. Impact of virtual reality applications in the treatment of anxiety disorders: a systematic review and meta-analysis of randomized-controlled trials. J Behav Ther Exp Psychiatry 2023 Dec;81:101893 [FREE Full text] [doi: 10.1016/j.jbtep.2023.101893] [Medline: 37453405]

XSL•FO
RenderX

34. Hedman-Lagerlöf E, Carlbring P, Svärdman F, Riper H, Cuijpers P, Andersson G. Therapist-supported Internet-based cognitive behaviour therapy yields similar effects as face-to-face therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. World Psychiatry 2023 Jul;22(2):305-314. [doi: 10.1002/wps.21088] [Medline: 37159350]
35. Buttazzoni A, Brar K, Minaker L. Smartphone-based interventions and internalizing disorders in youth: systematic review and meta-analysis. J Med Internet Res 2021 Jan 11;23(1):e16490. [doi: 10.2196/16490] [Medline: 33427682]
36. Kaveladze BT, Wasil AR, Bunyi JB, Ramirez V, Schueller SM. User experience, engagement, and popularity in mental health apps: secondary analysis of app analytics and expert app reviews. JMIR Hum Factors 2022 Jan 31;9(1):e30766 [FREE Full text] [doi: 10.2196/30766] [Medline: 35099398]
37. Bell IH, Thompson A, Valentine L, Adams S, Alvarez-Jimenez M, Nicholas J. Ownership, use of, and interest in digital mental health technologies among clinicians and young people across a spectrum of clinical care needs: cross-sectional survey. JMIR Ment Health 2022 May 11;9(5):e30716. [doi: 10.2196/30716] [Medline: 35544295]
38. Kerst A, Zielasek J, Gaebel W. Smartphone applications for depression: a systematic literature review and a survey of health care professionals' attitudes towards their use in clinical practice. Eur Arch Psychiatry Clin Neurosci 2020 Mar;270(2):139-152. [doi: 10.1007/s00406-018-0974-3] [Medline: 30607530]
39. Weitzel EC, Schwenke M, Schomerus G, Schönknecht P, Bleckwenn M, Mehnert-Theuerkauf A, et al. E-mental health in Germany - what is the current use and what are experiences of different types of health care providers for patients with mental illnesses? Arch Public Health 2023 Jul 17;81(1):133. [doi: 10.1186/s13690-023-01150-y] [Medline: 37461064]
40. Linardon J. Rates of attrition and engagement in randomized controlled trials of mindfulness apps: systematic review and meta-analysis. Behav Res Ther 2023 Dec;170:104421 [FREE Full text] [doi: 10.1016/j.brat.2023.104421] [Medline: 37862854]
41. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
42. Feijt MA, de Kort YA, Bongers IM, IJsselsteijn WA. Perceived drivers and barriers to the adoption of emental health by psychologists: the construction of the levels of adoption of emental health model. J Med Internet Res 2018 Apr 24;20(4):e153 [FREE Full text] [doi: 10.2196/jmir.9485] [Medline: 29691215]
43. Békés V, Aafjes-van Doorn K. Psychotherapists' attitudes toward online therapy during the COVID-19 pandemic. Journal of Psychotherapy Integration 2020 Jun;30(2):238-247. [doi: 10.1037/int0000214]
44. Braun P, Drüge M, Hennemann S, Nitsch FJ, Staeck R, Apolinário-Hagen J. Acceptance of e-mental health services for different application purposes among psychotherapists in clinical training in Germany and Switzerland: secondary analysis of a cross-sectional survey. Front Digit Health 2022;4:840869. [doi: 10.3389/fdgth.2022.840869] [Medline: 35295621]
45. Bucci S, Berry N, Morris R, Berry K, Haddock G, Lewis S, et al. "They are not hard-to-reach clients. we have just got hard-to-reach services." staff views of digital health tools in specialist mental health services. Front Psychiatry 2019 May;10:344 [FREE Full text] [doi: 10.3389/fpsyt.2019.00344] [Medline: 31133906]
46. Ganapathy A, Clough BA, Casey LM. Organizational and Policy Barriers to the Use of Digital Mental Health by Mental Health Professionals. Telemed J E Health 2021 Dec;27(12):1332-1343. [doi: 10.1089/tmj.2020.0455] [Medline: 33646057]
47. Gindidis S, Stewart SE, Roodenburg J. Psychologists' motivations for integrating apps into therapy with secondary school-aged young people. J Psychol Couns Sch 2019 Oct 21;30(1):2-12. [doi: 10.1017/jgc.2019.22]
48. Mendes-Santos C, Nunes F, Weiderpass E, Santana R, Andersson G. Understanding mental health professionals' perspectives and practices regarding the implementation of digital mental health: qualitative study. JMIR Form Res 2022 Apr 12;6(4):e32558 [FREE Full text] [doi: 10.2196/32558] [Medline: 35412459]
49. LimeSurvey: an open source survey tool. Limesurvey GmbH. URL: https://www.limesurvey.org [accessed 2024-06-10]
50. Karrer K, Glaser C, Clemens C, Bruder C. Technikaffinität erfassen - der Fragebogen TA-EG. In: Lichtenstein A, Stößel C, Clemens C, editors. Der Mensch im Mittelpunkt technischer Systeme. Berlin, Germany: VDI Verlag GmbH; 2009:194-199.
51. Franke T, Attig C, Wessel D. A personal resource for technology interaction: development and validation of the Affinity for Technology Interaction (ATI) Scale. International Journal of Human–Computer Interaction 2018 Mar 30;35(6):456-467. [doi: 10.1080/10447318.2018.1456150]
52. Bagarić B, Jokić-Begić N. Measuring therapists' attitudes toward integration of technology in psychotherapy and predicting their use of technology. J Clin Psychol 2020 Jul;76(6):1151-1172. [doi: 10.1002/jclp.22967] [Medline: 32410312]
53. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jul 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
54. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425-478.
55. Schröder J, Berger T, Meyer B, Lutz W, Hautzinger M, Späth C, et al. Attitudes towards internet interventions among psychotherapists and individuals with mild to moderate depression symptoms. Cogn Ther Res 2017 Apr 22;41(5):745-756. [doi: 10.1007/s10608-017-9850-0]

XSL•FO
RenderX

56. O'Dea B, Han J, Batterham PJ, Achilles MR, Calear AL, Werner-Seidler A, et al. A randomised controlled trial of a relationship-focussed mobile phone application for improving adolescents' mental health. J Child Psychol Psychiatry 2020 Aug;61(8):899-913 [FREE Full text] [doi: 10.1111/jcpp.13294] [Medline: 32683737]
57. Kenny R, Dooley B, Fitzgerald A. Feasibility of "CopeSmart": a telemental health app for adolescents. JMIR Ment Health 2015 Aug 10;2(3):e22 [FREE Full text] [doi: 10.2196/mental.4370] [Medline: 26552425]
58. Gao M, Kortum P, Oswald F. Multi-language toolkit for the System Usability Scale. International Journal of Human–Computer Interaction 2020 Aug 19;36(20):1883-1901 [FREE Full text] [doi: 10.1080/10447318.2020.1801173] [Medline: 40101104]
59. Cho J, Lee E. Reducing confusion about grounded theory and qualitative content analysis: similarities and differences. The Qualitative Report 2014 Nov 8;19(32):1-20 [FREE Full text] [doi: 10.46743/2160-3715/2014.1028]
60. Mayring P. Qualitative inhaltsanalyse. In: Handbuch Qualitative Forschung in der Psychologie: Band 2: Designs und Verfahren. Wiesbaden, Germany: Springer Fachmedien Wiesbaden; 2020:495-511.
61. Schuster R, Topooco N, Keller A, Radvogin E, Laireiter AR. Advantages and disadvantages of online and blended therapy: replication and extension of findings on psychotherapists' appraisals. Internet Interv 2020 Oct;21:100326 [FREE Full text] [doi: 10.1016/j.invent.2020.100326] [Medline: 32477885]
62. Fausett C, Lee J, Godwin E, Palmer E, Fouquet S. Informing the design of mental health smartphone apps for adolescents. Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care 2020 Sep 16;9(1):218-222. [doi: 10.1177/2327857920091012]
63. Høgsdal H, Kyrrestad H, Rye M, Kaiser S. Exploring adolescents' attitudes toward mental health apps: concurrent mixed methods study. JMIR Form Res 2024 Jan 15;8:e50222 [FREE Full text] [doi: 10.2196/50222] [Medline: 38224474]
64. Kenny R, Dooley B, Fitzgerald A. Developing mental health mobile apps: exploring adolescents' perspectives. Health Informatics J 2016 Jun;22(2):265-275 [FREE Full text] [doi: 10.1177/1460458214555041] [Medline: 25385165]
65. Topooco N, Riper H, Araya R, Berking M, Brunn M, Chevreul K, E-COMPARED consortium. Attitudes towards digital treatment for depression: a European stakeholder survey. Internet Interv 2017 Jun;8:1-9 [FREE Full text] [doi: 10.1016/j.invent.2017.01.001] [Medline: 30135823]
66. Gesetz für eine bessere versorgung durch digitalisierung und innovation. Bundesministerium für Gesundheit. URL: https://www.bundesgesundheitsministerium.de/service/gesetze-und-verordnungen/detail/digitale-versorgung-gesetz-dvg.html [accessed 2025-03-29]

Abbreviations

ATI: Affinity for Technology Interaction
COREQ: Consolidated Criteria for Reporting Qualitative Research
LAMH: Levels of Adoption of e-Mental Health
MTPS: Psychotherapists' Attitudes Toward Using Modern Technologies in Psychotherapy and Counselling Scale
TA-EG: Affinity for Technology
uMARS: user version of the Mobile Application Rating Scale
UTAUT: Unified Theory of Acceptance and Usability of Technology
WHO: World Health Organization

Edited by A Mavragani; submitted 26.08.24; peer-reviewed by A Jolliff, N Döring; comments to author 18.12.24; revised version received 21.01.25; accepted 07.02.25; published 08.04.25.

Please cite as:
Wüllner S, Hermenau K, Hecker T, Siniatchkin M
The Use of Mobile Apps in Adolescent Psychotherapy: Assessment of Psychotherapists' Perspectives
JMIR Form Res 2025;9:e65788
URL: https://formative.jmir.org/2025/1/e65788
doi:10.2196/65788
PMID:

©Sarah Wüllner, Katharin Hermenau, Tobias Hecker, Michael Siniatchkin. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Developing an Internet-Based Cognitive Behavioral Therapy Intervention for Adolescents With Anxiety Disorders: Design, Usability, and Initial Evaluation of the CoolMinds Intervention

Nikita Marie Sørensen[1], MSc; Helene Skaarnes[2,3], MSc; Kim Mathiasen[1,2], MSc, PhD; Mikael Thastum[1], MSc, PhD; Johanne Jeppesen Lomholt[1], MSc, PhD

[1]Department of Psychology and Behavioral Sciences, Aarhus University, Aarhus C, Denmark

[2]Department of Clinical Research, University of Southern Denmark, Odense, Denmark

[3]Centre for Digital Psychiatry, Mental Health Services, Region of Southern Denmark, Odense, Denmark

Corresponding Author:
Nikita Marie Sørensen, MSc
Department of Psychology and Behavioral Sciences
Aarhus University
Bartholins Allé 11
Aarhus C, 8000
Denmark
Phone: 45 87159439
Email: nikita@psy.au.dk

Abstract

Background: Digital mental health interventions may help increase access to psychological treatment for adolescents with anxiety disorders. However, many clinical evaluations of digital treatments report low adherence and engagement and high dropout rates, which remain challenges when the interventions are implemented in routine care. Involving intended end users in the development process through user-centered design methods may help maximize user engagement and establish the validity of interventions for implementation.

Objective: This study aimed to describe the methods used to develop a new internet-based cognitive behavioral therapy intervention, CoolMinds, within a user-centered design framework.

Methods: The development of intervention content progressed in three iterative design phases: (1) identifying needs and design specifications, (2) designing and testing prototypes, and (3) running feasibility tests with end users. In phase 1, a total of 24 adolescents participated in a user involvement workshop exploring their preferences on graphic identity and communication styles as well as their help-seeking behavior. In phase 2, a total of 4 adolescents attended individual usability tests in which they were presented with a prototype of a psychoeducational session and asked to think aloud about their actions on the platform. In phase 3, a total of 7 families from the feasibility trial participated in a semistructured interview about their satisfaction with and initial impressions of the platform and intervention content while in treatment. Activities in all 3 phases were audio recorded, transcribed, and coded using thematic analysis and qualitative description design. The intervention was continuously revised after each phase based on the feedback.

Results: In phase 1, adolescent feedback guided the look and feel of the intervention content (ie, color scheme, animation style, and communication style). Participants generally liked content that was relatable and age appropriate and felt motivating. Animations that resembled “humans” received more votes as adolescents could better “identify” themselves with them. Communication should preferably be “supportive” and feel “like a friend” talking to them. Statements including praise—such as “You’re well on your way. How are you today?”—received the most votes (12 votes), whereas directive statements such as “Tell us how your day has been?” and “How is practicing your steps going?” received the least votes (2 and 0 votes, respectively). In phase 2, adolescents perceived the platform as intuitive and easy to navigate and the session content as easy to understand but lengthy. In phase 3, families were generally satisfied with the intervention content, emphasizing the helpfulness of graphic material to understand therapeutic content. Their feedback helped identify areas for further improvement, such as editing down the material and including more in-session breaks.

Conclusions: Using user involvement practices in the development of interventions helps ensure continued alignment of the intervention with end-user needs and may help establish the validity of the intervention for implementation in routine care practice.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e66966) doi:10.2196/66966

KEYWORDS

user-centered design; digital treatment; digital mental health; internet-based; cognitive behavioral therapy; anxiety; adolescents

Introduction

Background

Digital mental health interventions (DMHIs) may help increase access to psychological treatment as provision via the internet is presumed to counter some of the key treatment barriers, such as social stigma [1,2], preference for self-reliance [1,2], treatment costs [3], and waiting times [2,3]. It is also well documented that internet-based cognitive behavioral therapy (ICBT) with support from a therapist (eg, via telephone or email) is effective for a variety of mental health problems, including anxiety in children and adolescents [4-7].

However, many evaluations of ICBT programs for children and adolescents with anxiety report high rates of dropout and that participants often fail to complete assessments or treatment sessions [8-12]. In addition, when transitioning to routine care settings, these challenges with adherence and engagement remain [13].

It is still unclear whether participants drop out of treatment due to symptom improvement, symptom deterioration, or other reasons outside therapy. However, working experts in the field of ICBT for children and young people propose that the use of user-centered design methods may help maximize user engagement [14,15]. This methodology emphasizes the importance of (1) understanding adolescents as experts on their own preferences and (2) enabling adolescents to hold central positions as experts in all stages of the design process. As it was eloquently stated in a qualitative study of the Australian Momentum platform, this type of design process may be characterized by "a *designing with* as opposed to a *designing for* mindset" [16].

User-centered design processes typically involve three phases of development: (1) identifying needs and design specifications (with end users), (2) designing prototypes of the intervention and testing their usability, and (3) running feasibility tests with end users [17,18]. The design process must be understood as iterative, and each phase can be revisited if needed as the data collection and analyses progress.

To the best our knowledge, only a few studies describe the inclusion of end users in the design of ICBT interventions for children and adolescents with anxiety [16,19,20]. Hill et al [20] and Ludlow et al [16] both used co-design practices involving service users (parents and children) and service providers (clinicians from routine care) in the development of their ICBT anxiety interventions. In an initial evaluation of the clinical effectiveness, Hill et al [21] found that the dropout rate for their intervention was lower (13%) than those reported in other studies conducted in routine care (32% in the study by Moor et al [22], 21% in the study by Waite et al [13], and 17% in the study by Vigerland et al [23]). This indicates that using user-centered design practices may result in better adherence and engagement with the intervention when delivered in routine care settings.

The importance of including end users in the design of DMHIs has been increasingly acknowledged as imperative not only in developing DMHIs but also in the processes of implementing the interventions [14]. The project that this paper describes was initiated by the Region of Southern Denmark, a government organization that is responsible for the provision of health care services for the residents of the southern municipalities in Denmark. This was done with an intention of broadening the service provision of a DMHI to routine care settings at a national level. Transitioning from research to routine care settings can be tricky, and evidence from clinical trials alone does not guarantee the use of an intervention in routine care. Using user-centered design practices in the development of this intervention may facilitate a smoother transition to large-scale service provision.

Objectives

The objective of this study was to describe the methods used to develop a new ICBT intervention for adolescents aged 12 to 17 years with clinical anxiety disorders, CoolMinds, within a user-centered design framework.

Our hope is that the methods and results described in this paper will provide inspiration for others on how to include end users in the design processes of DMHIs. In this study, we specifically collected data focusing on how to enhance engagement with the DMHI guided by the following objectives:

1. What are adolescents' preferences on graphic identity and communication styles within a digital intervention?
2. How do adolescents prefer to seek help and be supported by their parents in this?
3. How do adolescents navigate a prototype of the digital platform and content?
4. What are the families' initial impressions of the platform and content while in treatment?

Methods

Setting

Intervention development was initiated in January 2022 as a collaborative project involving the Centre for the Psychological Treatment of Children and Adolescents and the Centre for Digital Psychiatry. The Centre for the Psychological Treatment of Children and Adolescents is a research center nested within the Department of Psychology and Behavioural Sciences at Aarhus University and is specialized in studying and delivering anxiety interventions to children and adolescents. The Centre for Digital Psychiatry is a research, development, and treatment facility specialized in the use of digital technologies for health promotion and psychiatric treatments and is located within the Mental Health Services in the Region of Southern Denmark.

XSL•FO
RenderX

The project includes intervention development, a feasibility trial, and a randomized controlled trial (RCT). In this paper, only design processes and results from intervention development and the evaluation of the first intervention prototype in the feasibility trial are presented.

Ethical Considerations

The feasibility trial and RCT were ethically approved by the Medical Research Ethics Committees in Denmark (ID: VMK 2211954), and the protocol for these studies was preregistered in ClinicalTrials.gov (NCT06076964) and published in *Trials* [24]. The workshops and usability tests were conducted as development projects, which do not require ethics approval. The reason for this is that these activities did not include any experimental manipulations, were not targeted at a vulnerable population, and the potential risks associated with participation in these activities were evaluated as very low. Participants were not compensated for taking part in any of the activities or studies.

For the workshops and usability tests, written informed consent was obtained regarding participation, audio recordings, and publication of results. For adolescents aged <15 years, consent was obtained from parents. For adolescents aged ≥15 years, consent was obtained from the adolescents themselves. However, parents were informed of the adolescents' participation and were urged to discuss this with their children before participation.

For the feasibility trial and the RCT, written and verbal informed consent were obtained from all parents regarding participation and publication. In addition, verbal informed consent was obtained for adolescents aged <15 years, and both verbal and written informed consent were obtained for adolescents aged ≥15 years.

An additional written consent form was used to obtain consent regarding participation in, recording of, and publication of information from feasibility interviews. For adolescents aged <15 years, written informed consent was obtained from parents. For adolescents aged ≥15 years, written informed consent was obtained from both the adolescents and parents.

The quotes reported in this paper were translated from Danish into English. Quotes containing sensitive or identifying information were excluded to protect participants' anonymity.

Procedure

The CoolMinds ICBT intervention was developed and continuously updated from January 2022, when the initial outlining of the project began, until June 2023, when the feasibility trial concluded. The development of the intervention followed the recommendations from Hill et al [14] and progressed in three iterations following the design phases: (1) identifying needs and design specifications, (2) designing and testing prototypes, and (3) running feasibility tests. Reporting of the study followed the Guidance for Reporting Involvement of Patients and the Public, and a Guidance for Reporting Involvement of Patients and the Public long-form checklist is available in Multimedia Appendix 1 [25].

Working Groups

Development processes were led by a working group consisting of junior researchers (authors NMS and HS); a psychologist with experience in producing written materials for adolescents (see the Acknowledgments section); and members of the development and implementation team, such as graphic designers, user involvement experts, and engineers (see the Acknowledgments section) and were supervised by senior researchers (authors JJL, KM, and MT).

During the development process, weekly meetings were held with all available team members. The meetings always commenced with a brief status from all involved teams. The purpose of these meetings was to facilitate the decision-making and implementation processes by assembling relevant team members [26]. This ensured a flow of information between teams when decision-making required the expertise of multiple disciplines (eg, the psychological content had to be informed by data from the user involvement practices and be aligned with the technical possibilities within the platform). When consensus on each topic was reached by all teams, smaller working groups were formed with relevant team members who implemented the decision-based outcomes.

Furthermore, a total of 3 team workshops were held with members representing each team in the working group. The workshops were held at different time points in the development process reflecting the writing of the intervention manuscript. The purpose of the workshops was to (1) map out the overall framework of the intervention content and (2) become aware of any missing information that would need to be researched or collected using user involvement practices. The topics of these workshops were (1) outlining the intervention content, (2) rethinking how to conduct exposure therapy in a digital format, and (3) parental involvement.

Outlining the Program

The outline of the program was drafted by the working group with expert supervision. This included decisions on therapeutic approach, key treatment components, and parental involvement. It was decided to develop a cognitive behavioral therapy (CBT)–based program as this is the most well-researched approach and has shown good results in previous evaluations using the internet-based format [27-31]. Decisions on which treatment components to include were based on either evidence from research and literature or clinician experience. Several reviews and meta-analyses suggest that key CBT components such as exposure therapy (in vivo and imaginal) and cognitive restructuring are some of the main components driving the treatment effects [27,32,33]. Evidence on the efficacy of coping strategies is scarce compared to that for the key components [34]. However, the group of clinicians involved in providing feedback on the program manuscript advocated for the inclusion of coping strategies as these, in their experience, were techniques that adolescents usually benefitted from and were easier for them to implement on a day-to-day basis. This is also in line with findings from qualitative evaluations of ICBT treatments, where adolescents preferred easy-to-use techniques [35,36]. It was also decided to create a separate parent program and offer the treatment as a family-based intervention as social support

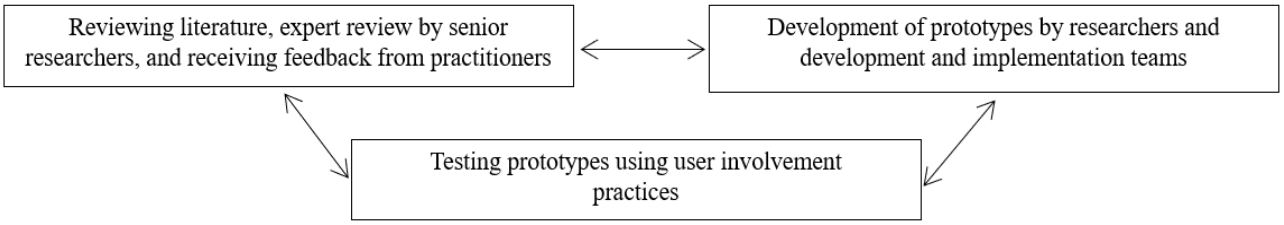
XSL•FO
RenderX

from parents may increase treatment adherence and response when treatment is delivered in a digital format [37]. However, the 2 programs were not connected on the platform, and thus, adolescents and parents were allowed to progress at their own pace within each of their programs.

In both programs, sessions were structured in a predefined order starting with psychoeducation followed by cognitive restructuring techniques and, finally, exposure therapy techniques. Sessions were opened automatically upon completion of the previous session unless the adolescent was to receive feedback from a therapist. In that case, the session was first opened upon receiving feedback.

Writing of the program manuscript was guided by experiences from previous and linked research activities, current literature in the field, and expertise of senior researchers and experienced clinical psychologists within and external to the project who continuously provided feedback on the manuscript. The literature and practitioner expertise were revisited during the development process to qualify and provide context to the incorporation of insights from the user involvement workshop and usability tests conducted using prototypes of the intervention. This iterative design process is illustrated in Figure 1.

Figure 1. Model of the iterative design process used across phases of development.



Phase 1: Identifying Design Specifications in a User Involvement Workshop

A total of 24 adolescents in grade 9 (aged >15 years) from a Danish public school participated in a user involvement workshop held in September 2022. The purpose of the user involvement workshop was to identify adolescents' preferences on graphic identity and communication styles within the intervention content (objective 1) and investigate adolescent help-seeking behavior (objective 2).

Activities included (1) voting for their favorite prototypes (out of a selection of 3-5 prototypes) using stickers, (2) discussing the pros and cons of each prototype in smaller groups and plenary, and (3) discussing how to reach out for help. Prototypes included different color schemes, styles of animation, expert videos, and types of feedback from the app (push notifications, mood ratings, and in-app rewards and praise). For polls including many prototypes (ie, color schemes, animation style, and notifications), adolescents had more than 1 vote. However, they were also allowed not to use all their votes. The topic of reaching out for help and receiving support was discussed in a column exercise in which adolescents were to give advice to another adolescent in mock-up scenarios about (1) how to handle parents being overinvolved, (2) how to involve parents, and (3) how to handle not having parental support and where to go for help then. Their advice was written down by an assistant in each group. Outcomes of the workshop were documented by audio recording group discussions and through photographs of votes and plenary summaries (both done on a whiteboard).

Members of the development team contacted 4 public schools via phone and, if they were interested, provided them with written information about the content and purpose of the workshop. One school agreed to participate, and the time of the workshop was planned in collaboration with the grade 9 teacher. The written information was then shared with parents via the school's intranet. Informed written consent was obtained from all adolescents at the beginning of the workshop using a printed consent form.

Phase 2: Testing Intervention Prototypes in Usability Tests

In total, 4 adolescents (all female) participated in individual usability tests during November 2022 and December 2022. A total of 2 tests were conducted in person with 50% (2/4) of the adolescents, who were aged 12 years, and 2 tests were conducted online with 50% (2/4) of the adolescents, who were aged 16 years. The purpose of the usability tests was to understand how adolescents would navigate the digital platform and perceive the intervention content within this platform (objective 3). Participants were instructed by an assistant to go through a psychoeducational session of the program and say out loud what they thought they were to do in the program and what they then did. They were also prompted to point out anything that confused them and anything that they were particularly fond of in the program. The assistant was present during the entire usability test and took extensive notes used for briefly interviewing the participants about their experience with the program content (eg, animation style and intelligibility of the text) and interface (eg, intuitiveness of navigating the platform), along with some standardized questions (Multimedia Appendix 2). The usability tests were also audio recorded.

Participants were recruited through word of mouth by members of the research group. Informed written consent was obtained from caregivers to children aged <15 years using a printed consent form. Informed written consent was obtained from adolescents aged >15 years using a printable digital consent form sent and received using secure emails (ie, encrypted and digitally signed emails).

Phase 3: Feasibility Tests With End Users

A total of 7 families (adolescents and their parents) from the feasibility trial participated in a 2-part online interview regarding (1) their satisfaction with trial procedures and (2) their

XSL•FO
RenderX

experiences using the program and with specific content components. We developed a semistructured interview guide with open-ended questions regarding general topics related to the trial and treatment procedures and questions regarding specific components of the program content (Multimedia Appendix 2). In this paper, only results regarding the program content are elaborated on. Results on satisfaction with trial procedures will be published elsewhere.

The purpose of these interviews was to explore the families' initial impressions of the platform and content while still in treatment (objective 4). The interviews were conducted with each family individually but at different time points during their treatment period. The families were picked at random during March 2023 and April 2023 depending on their progress (ie, whether they were at the beginning of, midway through, or at the end of their treatment period). This was done to ensure the collection of information about session-specific content while it was still novel for the participants. The interviews were conducted online and recorded.

For the participating adolescents (n=7), the mean age was 13.29 (SD 1.38; range 12-16) years. A total of 57% (4/7) were male, and 43% (3/7) were female. In total, 43% (6/14) of the parents had a high school or vocational education, 50% (7/14) had completed medium-cycle higher education, and 7% (1/14) had completed long-cycle higher education. All adolescents and their parents were born in Denmark.

Informed written consent was received from all participants aged >15 years and their caregivers before participation using an online consent form distributed using REDCap (Research Electronic Data Capture; Vanderbilt University), a secure, web-based software platform designed to support data capture for research studies [38,39] hosted at the Open Patient Data Exploratory Network within the Region of Southern Denmark.

Data Analysis

Phase 1: Identifying Design Specifications in a User Involvement Workshop

Votes were manually counted and ranked from the highest to lowest number of votes. The audio files from group discussions on pros and cons were transcribed, and the main arguments for and against each prototype were listed by prototype.

Audio files on help-seeking behavior were transcribed and analyzed using a qualitative description design by author NMS [40]. Data were analyzed using content analysis, and codes were sorted into themes.

Phase 2: Testing Intervention Prototypes in Usability Tests

Notes from the usability test were analyzed using a qualitative description design [40]. Data were coded into predefined categories reflecting each feature assessed in the usability test (eg, text, video, task, and animation).

Audio files were only used to expand the notes if anything was unclear.

Phase 3: Feasibility Tests With End Users

Due to technical issues, recordings from only 71% (5/7) of the interviews were available. These interviews were transcribed and analyzed using a qualitative description design by author NMS [40]. Data were coded using content analysis based on features (eg, text, video, animation, and audio) and preferences (ie, likes or dislikes). Data were then sorted to identify similar patterns, commonalities, and differences. The codes were then grouped into two main themes representing the general feedback from the participants on (1) satisfaction and acceptability and (2) parental involvement, as predefined in the interview guide.

Results

Phase 1: Identifying Design Specifications in a User Involvement Workshop

Overview

Results from the user involvement workshop are reported in Table 1.

XSL•FO
RenderX

Table 1. Overview of adolescent votes and comments from the user involvement workshop conducted at a Danish school, divided into pros and cons. Note that some statements are summaries of participants' comments and some are direct quotes (indicated by quotation marks).

	Number of votes	Pros	Cons
Color scheme (only the ones with the most votes are presented)			
Color scheme 7—pink, blue, green, yellow, and orange	7	—[a]	—
Color scheme 5—reds and grays	6	—	—
Animation style (only the ones with the most votes are presented)			
Animation 4—the pursuit of happiness	12	• Felt relatable as it represented "real" people. • "Felt like I could identify myself with the animations better because it used 'real' people."	—
Animation 1—letting go of stress	10	• It worked well because it did not contain that many things. • It felt very calming.	• The text could be presented a bit slower. • It was boring. • It was too fast.
Animation 2—changing perspective	10	• "I liked that everything was quite fast because it shows how chaotic your thoughts can be."	—
Expert videos			
Expert 3—female actress	8	• She seemed to be more like a psychologist. • It was nice that she also used her body while speaking. • "I liked the calm background and her voice." • It was nice and simple without any disruptive elements. • She seemed trustworthy and professional.	—
Expert 1—male professor	7	• Seemed smart. • "His way of talking appealed the most to me, but the books behind him didn't." • "He seems intelligent. The books only added to his professionalism."	• "A bit 'heavy.' He seems like a know-it-all." • "It is way too boring, you don't actually listen to what he says." • "The books in the background are disruptive. I can't really focus on him because of them." • The background is 'too noisy.'
Expert 2—2 young people	6	• It was nice that young people presented the information as it made it feel more relatable. • "It felt more age appropriate to have young people speak to me, but it also depends on what they're talking about."	• It would be nicer to only have one person talking. • It was confusing that they alternated talking. • They were a bit stiff.
Push notifications			
Notification 5—"You're well on your way. How are you today?"	12	• "It feels motivating to receive praise and a reminder in one." • "It is nice that it has both a motivating text and a question." • "It is nice that something positive is included." • "It feels like a pat on the back."	—
Notification 1—"It has been a while since we've heard from you. How are you?"	10	• "It feels like a friend is asking." • "It is good."	• "Can be understood both positively and negatively."

XSL•FO
RenderX

	Number of votes	Pros	Cons
Notification 2—"We know, you're working hard. Have you rewarded yourself?"	9	• "The thought of taking care of yourself is nice." • "I like the idea of rewarding myself."	• "Can maybe be a bit demanding."
Notification 3—"How has your day been?"	4	• "Covers a lot."	• "Is hard to answer as it is too broad."
Notification 6—"Tell us how your day has been?"	2	• "Is a simple, open-ended question."	• "Seems like an order." • "It sounds a bit negative."
Notification 4—"How is practicing your steps going?"	0	• More specific	• Not very deep • "Seems like a yes/no question." • "Can seem like a reprimand."
In-program feedback—ratings			
Rating 1—smiley faces	13	• "It is easier to understand faces as opposed to e.g., colors." • The intention and meaning were clear. • "The variety of options is a plus." • "It doesn't sugar coat anything, which is nice."	• "Maybe faces can be too specific."
Rating 2—weather	7	• "It does make good sense to compare your mood to the weather." • "The weather gives you more room for interpretation and ascribing emotions."	• "It is a bit difficult to understand."
Rating 3—waves	1	• The waves did leave more room for interpretation.	• "Needs more options." • "It is too unspecific. It can't relate to it at all." • "I don't understand it."
In-program feedback—completion			
Feedback 1—checkmark, praise, and quote	9	• It was simple and serious. • "I like the combination." • "It is good and easy to understand." • "I really like the use of quotes." • "It seems the most professional and the quote is like a reward."	—
Feedback 2—animated character receiving a trophy and confetti on a podium	8	• "The progress-bar is nice as you can track your progress and see how much is left of the program." • "It is nice that it has no gender." • "It is cute."	• "It is a bit too over the top." • "It needs to be used consequently otherwise you'd think you weren't doing good enough." • "It is a bit too 'cute.'"
Feedback 3—animated character waving as it walks along a progress bar	3	• "It is nice to have some movement in the animation." • "It is funny." • "It is good to know how far you've come."	• "It seems too childish."

[a]Not applicable; the adolescents did not have any comments on the specific prototype listed.

Color Schemes

Color scheme 7 received the most votes (7 votes), closely followed by color scheme 5 (6 votes).

Animation Style

Animation 4 received the most votes (12 votes), closely followed by animations 1 and 2 (tied; 10 votes each). The adolescents described animation 4 as the most "relatable" one. They thought that it "worked well" because it included "humans" and that

XSL•FO
RenderX

they were able to "identify better" with the animation because of this.

Expert Videos

Votes were almost tied between the expert videos. Expert 3 received the most votes (8 votes), followed by expert 1 (7 votes) and expert 2 (6 votes). The adolescents described how expert 3 seemed more like a psychologist in the way she looked and talked. They also preferred the background and simple setup of expert 3 (ie, plain background) compared to that of expert 1 (ie, shelves with books) as it was perceived as more "calm" and not "disrupting." However, some adolescents felt that the look and background of expert 1 added to the credibility of the expert and the information provided in the video.

Push Notifications

Notification 5 received the most votes (12 votes). The adolescents emphasized the motivating and acknowledging aspect of receiving praise and that it felt nice to combine praise and a question in contrast to just a question (ie, "You're well on your way. How are you today?"). The adolescents also described how the tone of the text felt more like "a friend" and "supportive." Notification 1 also received a large number of votes (10 votes; "It has been a while since we've heard from you. How are you?"). It was also described as feeling "like a friend." However, most adolescents also pointed to the potential "double meaning" as they thought that it could be "interpreted both as positive and negative." The same was the case for notification 2 (9 votes), which was described as both "pleasant" because of "the thought of caring for yourself" and "commanding."

Notifications 4 and 6 received the least votes and were described as "negative" and feeling like "being told what do to" (eg, "Tell us how your day has been?"). Notification 3 was perceived as too hard to answer because it was too "broad." However, some adolescents did like that it was not too specific and would "cover more."

Ratings

Rating 1, an array of different smiley faces, received the most votes (13 votes). The adolescents described how it was "the easiest to understand the meaning of" and that it was "an advantage that it had more options" to describe their mood more accurately. Ratings 2 (7 votes) and 3 (1 vote) were generally perceived as "hard to interpret," "difficult to understand," and too "unspecific." However, the metaphors of weather and waves could also "leave room for interpretation" and the "possibility to attribute more feelings to it" compared to the smiley faces in rating 1.

Completion

Feedback 1 and 2 received the most votes (9 and 8 votes, respectively). The adolescents preferred more "serious"-looking feedback and animations, such as simple checkmarks or written praise and quotes, and they wanted the animations to be the same "or else you could think that you've not done it well enough." The adolescents were very fond of the combination of praise and a quote as the quote then felt like "a reward" (eg, "Good job! As Walt Disney once said, the difference between winning and losing is not quitting"). Feedback 3 (3 votes) was perceived as "too immature," but the adolescents liked the visualization of how far they had come in the program.

Parental Involvement

When investigating adolescents' help-seeking behavior in the user involvement workshop, the general response on how to talk with parents about something difficult was to "just say it as it is" as "your parents want what's best for you anyway." However, all groups also mentioned seeking support from others, such as another family member (eg, grandparent), a teacher, friends, the school nurse, a physician, counseling services, or the internet, despite all scenarios being centered on parental involvement.

When looking at overarching themes across all scenarios, the themes *security* and *respect, boundaries, and mutual understanding* were identified (Multimedia Appendix 3). *Security* represents the feeling of relational safety ("they only want what's best for you") that enables the adolescent to feel comfortable speaking to someone else. *Respect, boundaries, and mutual understanding* represents the adolescents' need for autonomy ("she needs to establish boundaries for her parents") and having the defining voice in how to communicate about their difficulties ("Make agreements with your parents on what they can ask about and when").

Phase 2: Testing Intervention Prototypes in Usability Tests

The results from the usability tests can be found in Multimedia Appendix 3. Generally, the interface of the session was perceived as intuitive and easy to navigate by the adolescents. They expressed that "it is easy to understand, what I have to do" to move on in the program. However, most of the adolescents, especially the younger participants, felt that the session was "too long" and that there was "a bit too much text" and "too many questions." One of the youngest participants commented that she had "actually forgotten the video" she saw at the beginning of the session when she had to refer to it in an assignment at the end of the session.

The content was generally perceived as "understandable." Both graphics and videos were generally perceived as helpful to better understand the techniques and examples presented. However, for the youngest participants, some of the content was too complex at times. One participant explained that the animations were "good but at times confusing" if too much was happening at once. She noted that "you have to look at them for a while to understand them." Another participant commented that the graphics helped her understand the content better and that the order of presentation of content was important for her. She explained that it would be more helpful to see the graphics first and then the text-based explanation for her to better understand it (eg, seeing the feelings thermometer as a graphic and then reading about it).

Phase 3: Feasibility Tests With End Users

The results from the interviews conducted with families in treatment during the feasibility trial can be found in Multimedia Appendix 3.

XSL•FO
RenderX

Satisfaction and Acceptability

The participants were generally satisfied with the program content. Both adolescents and parents emphasized graphic elements such as animations, videos, and audio files as particularly beneficial for their understanding and acquiring of the treatment techniques. Both adolescents and parents noted how the audio files with stories from past patients with anxiety disorders were "relatable" and "helped them understand" their own or their children's anxiety. One adolescent noted that it helped him feel like he was "not the only one who has got something."

Both adolescents and parents also liked that the introductory sessions in each of their programs included videos on how to navigate the platform and functionalities within the program. One parent noted that "there were videos for everything, so it was impossible to do anything wrong." Another parent also noted that it was "helpful" that she "could always go back and watch it again if you needed to."

Parents particularly liked that they, in their program, were able to try out some of the same techniques and tasks as their children but directed at their own worries (ie, cognitive restructuring or doing an exposure task). One parent explained how this made her more capable of helping her child throughout the treatment. However, for some participants, it was still difficult to understand how to do the exposure tasks. Some of the parents pointed out that they would have liked to be presented with more disorder-specific examples of *stepladders* and *step planners*.

Almost all participants, both adolescents and parents, expressed that the amount of reading was too much. In addition, some participants found the number of questions and tasks to complete in each session overwhelming. One adolescent explained that he "needed to take breaks after doing tasks to stay motivated." One of the parents also said that the breaks within the sessions "felt like an acknowledgement and like it was okay to not complete everything in one go."

Parental Involvement

Similarly to the findings of the user involvement workshops, adolescents expressed how having their own treatment program separate from their parents gave them a sense of "freedom" and "confidentiality" regarding the program. Parents, on the other hand, expressed feeling a "lack of control" or being "excluded." However, some parents (of mainly older children) also noted that the lack of control or unawareness of the adolescent's progress was "maybe healthy" and "fine but took a little getting used to."

Both adolescents and parents did find it difficult to align their progress when completing separate programs. Some families noted that it was "weird" or "difficult to discuss the program content when they were not at the same session" and that it would sometimes lead to "misunderstandings." Another family noted that they sometimes had difficulties helping their adolescent with their tasks if they had not been introduced to the techniques yet.

Final Iteration

The first versions of the adolescent and parent programs were completed in December 2022 and evaluated in a feasibility trial from January 2023 to June 2023. Findings from the interviews conducted during the feasibility trial (described previously) were then used to update and expand existing material from the first version. The main updates made to the program were (1) editing down of material and converting text to graphics, (2) including in-session breaks to help adolescents stay motivated, and (3) aligning the adolescent and parent programs to be completed simultaneously. The second versions of the 2 programs are currently being evaluated in an RCT (ClinicalTrials.gov NCT06368557).

The adolescent and parent versions of the program content can be found in Multimedia Appendix 4.

Discussion

Principal Findings and Implications for Content Development

Overview

This study addresses a gap in the current research on developing DMHIs. First, this study provides examples of feasible methods of including end users in the initial stages of development processes. Second, the results of this study complement and expand the current knowledge on adolescent preferences with regard to DMHIs. This may ensure the development of relevant DMHIs and help establish the validity of these interventions for implementation in routine care settings.

Results from phase 1 guided the look and feel of the intervention content, such as the graphic identity (ie, colors and animations) and communication style. Generally, adolescents liked content that was relatable and age appropriate. Adolescents preferred communication that felt "like a friend" (ie, in a tone that was supportive). When discussing support and help-seeking behaviors, adolescents emphasized both themes of independence and the need for relational security. In addition, they pointed out how support may be provided not only by parents but also by others in their social circle. In phase 2, adolescents perceived the platform as intuitive and easy to navigate. Overall, the content was easy to understand but too lengthy. In phase 3, families were generally satisfied with the intervention content, but it was still too lengthy, and the 2 programs (ie, the adolescent and parent programs) needed to be better aligned to allow for adolescents and parents to progress at the same pace. Overall, feedback from end users helped identify areas for further improvement.

Objective 1: Preferences on Graphic Identity and Communication Style

Overview

Findings from the user involvement workshop helped clarify adolescent preferences regarding both graphics and text-based material. The prototypes with the most votes guided the development of all content types. Thus, it was important to

XSL•FO
RenderX

investigate this early in the development process while still outlining the program.

Graphic Identity

The color schemes and animation styles with the most votes were used when designing the project website and logo and in the design of graphic material for the intervention (Multimedia Appendix 5). Preferences on animation style and expert videos were also used when developing the graphical content of the intervention. Specifically, the somewhat equal number of votes regarding expert videos led to the inclusion of two types of experts in the program: (1) a psychologist explaining CBT techniques and psychoeducational material and (2) a professor explaining the research on anxiety and CBT. On the basis of adolescents' qualitative comments, we decided to keep the background simple (Multimedia Appendix 5).

Communication Style

The qualitative comments from the adolescent discussions of prototypes of notifications and feedback guided the writing style of the program. We used a sort of "friendly strictness," where the program should feel like a friend supporting but also pushing them forward. One way to do this was to formulate explanations and tasks as offers instead of directive statements while still emphasizing the importance of doing the tasks (ie, "A lot of people make the mistake of believing that their thoughts are truthful and factual. However, thoughts are just mere ideas that may or may not be true. You won't know if your thought is true unless you do a fact-check of it. So, one of the first things you should ask yourself, when you're feeling anxious is..."). Furthermore, the adolescents' comments also emphasized the positive impact of including quotes and praise. Thus, praise was included throughout the program at the conclusion of exercises and modules (ie, "Well done! Now you should have a bunch of ideas on how to reward yourself"). At module completion, the praise was also combined with quotes as they were perceived as rewarding by the adolescents (ie, "Congratulations! You've now completed Step 5. Walt Disney once said that the difference between winning and losing is most often not quitting"). We decided to include quotes as relevant in-program rewards to keep adolescents motivated to use the program.

We did not end up including push notifications in the feasibility version of the program due to technical limitations within the digital platform. Instead, they have been included in the RCT version that is currently being evaluated. Their potential impact on adherence will be investigated in that trial.

Relatedness

In the user involvement workshop, the topic of *being relatable* was evident in the qualitative comments across all preferred prototypes. Adolescents simply preferred content prototypes that were relatable in such a way that they could identify with them, such as using humanlike characters in animations, using appropriate metaphors, and using supportive and developmentally appropriate language. This is in line with previous research emphasizing the importance of including relatable and age-appropriate content [16,35,36,41] and, thus, including adolescents in the design process to clarify which types of content are seen as relevant.

This finding guided the writing of the intervention manuscript and development of graphic material. For example, we included recurring animated characters, iGuides (Multimedia Appendix 5), that guided the adolescents through the intervention content using a storytelling approach in a cartoonlike format. The purpose was to offer a more personal and age-appropriate dissemination of CBT concepts and techniques presented in the program. We also conducted interviews with past patients whose stories were then audio recorded by actors to provide the adolescents with a feeling of recognition and acknowledgment. These stories were included as part of psychoeducation (ie, examples of how it feels to have anxiety), cognitive restructuring (ie, examples of helpful thoughts), and exposure tasks (ie, examples of goals and use of the technique) and as motivational speeches (ie, how it was to receive treatment and overcome anxiety). In addition, we included "help a friend" tasks in which the adolescent had to help a friend (ie, a fictional case) using the presented CBT techniques to further the feeling of relatedness and mastery.

The theme of being "relatable" was also present in the interviews conducted with families from the feasibility trial in phase 3. In this case, families noted that the stories from past patients had helped them feel less alone ("I'm not the only one who has got something") and helped them understand their situation better. Thus, relatedness and relatability may be key themes when developing engaging internet-based interventions.

Objective 2: Help-Seeking Behavior

Insights on adolescent help-seeking behavior helped clarify how adolescents would prefer to be supported by parents and caregivers.

In the user involvement workshop, the overarching themes present in the adolescents' discussion of support and help-seeking behavior tapped into the duality of taking charge on their own (eg, by "establishing boundaries," "making agreements," and being "serious") and the need for relational security with others in this process (eg, by "just telling them" and "talking to someone you trust"). In addition, many adolescents also qualitatively emphasized how it could be helpful to speak to someone other than their parents (eg, a friend, a teacher, or another family member).

This duality may reflect how adolescents themselves initiate and engage in scaffolding efforts as learners. Scaffolding refers to the process "that enables a child or novice to solve a task or achieve a goal that would be beyond his unassisted efforts" [42]. Thus, it refers to the process of assisting a person to move through the zone of proximal development by "controlling those elements of the task that are initially beyond the learner's capability, thus permitting him to concentrate upon and complete only those elements that are within his range of competence" [42]. In this case, adolescents may themselves initiate this process by seeking out relevant "assistants" to support them in using the intervention and acquiring the skills and knowledge presented in the intervention.

In addition, there is some evidence suggesting that social support may be more important in ICBT treatments in which the level of therapist support is low compared to traditional CBT. In a

XSL•FO
RenderX

predictor analysis by Spence et al [37], social support from family, friends, or a special person significantly predicted higher levels of program adherence for adolescents completing the BRAVE-online self-help program. However, only family support predicted treatment outcome. This may underline the importance of parental support to increase treatment adherence in digital formats.

However, in a predictor analysis by Stjerneklar et al [43], the time spent by parents helping their adolescents complete treatment did not significantly predict treatment outcome. In this study, the level of therapist support was high (weekly phone calls set to a duration of approximately 20 minutes) compared to that in the study by Spence et al [37] (no support). Thus, the role of parental support may only be important in interventions with no or low levels of therapist support.

In broader terms, these themes may also reflect how adolescence as a developmental stage is a time when liberation from parents and the need for autonomy become more apparent. Turning to peers instead of parents for support may not be unusual for this age group, and thus, the program content should be inclusive of this. In addition, increasing autonomy may call for a communicative approach that emphasizes the adolescents' agency [44]. This meant that, instead of focusing solely on using parents for support during the ICBT treatment, adolescents were also prompted to create a list of their "support network" in the program's first session, identifying all people who would be able to support them and specifying how they would like to be supported by them.

In the feasibility trial conducted in phase 3 of the development process, autonomy and support were also themes that were discussed by the families. Adolescents generally enjoyed the autonomy related to having separate programs, whereas for parents, this was associated with mixed responses. Generally, parents of older adolescents were more positive toward their children working in the program on their own. This is in line with findings from a co-design study by Ludlow et al [45] with caregivers of young people aged 7 to 17 years. They described how the desired level of caregiver involvement depended on the young person's age, where involvement may be lower with older adolescents. Thus, allowing for different levels of caregiver involvement depending on the age of the adolescent could allow for more flexible and personalized use of the intervention.

Objective 3: Navigating the Platform and Content

The usability tests provided practical information on how end users would navigate the platform and content. On the basis of this, the intervention material was edited down significantly, and most text bodies were converted into graphics or short videos. On the basis of feedback from the youngest participants (aged 12 years), some of the existing graphics were also simplified and slowed down. Where possible, existing videos were edited down or split into sections to not last more than 2 to 3 minutes each. These findings were then used to guide the design of subsequent graphic elements.

Objective 4: First Impressions From Families in Treatment

Generally, families were satisfied with the program content and platform. Families emphasized the graphic elements of the content, such as animations, videos, and audio files, as particularly beneficial. On the basis of this, some text bodies were converted into graphics and videos, and other text-based elements such as checklists (eg, lists with examples of symptoms, behaviors, and rewards) were edited down. One session (on realistic thinking) was also split into 2 subsessions on identifying and working with anxious thoughts and practicing detective thinking.

However, some families did find it difficult to align their progress when completing separate programs. This led to 2 major revisions to the parent program related to enhancing the joint use of the 2 programs. First, guidelines on how to appropriately use the program were included in the introductory session. This included guidelines on how often to work with the program, how often to do exercises outside of the program, and tips on how to stay consistent with the help of parents, all with the purpose of better aligning the adolescent and parent progress when working separately. Second, the parent program was also restructured from 7 sessions to 10 sessions to better reflect the course of the adolescent program. Coping strategies presented in the adolescent program (in the "toolbox" session) were included in short in a "toolbox" session for the parents. In addition, the intended purpose and rationale of each task were elaborated on further in both the adolescent and parent programs to help families establish a mutual understanding of the program content.

Despite efforts to keep program content short, it was still too extensive, and the amount of reading was too much. In addition to converting text bodies to graphics and editing down the material, more in-session breaks were included, encouraging adolescents to take time to implement the strategies and then return to the platform.

Thus, findings from the feasibility interviews led to substantial revisions of the program, all related to enhancing its use. However, little is known about the role of including user-centered practices to enhance engagement with DMHIs. Some research conducted with adult populations indicates that using user-centered practices may help increase engagement [46,47], but more research is generally needed.

Limitations

This study has some limitations related to the representativeness of the samples and the extent of the involvement of end users in the design processes that may be of importance when interpreting the results.

Not all samples in this study may be representative of the actual end users, who are children with clinical anxiety. However, this argument assumes that there would be a difference in preferences between adolescents with and without anxiety, but perhaps it is feasible to first include adolescents as experts on *being young* as opposed to (but not exclusive of) *being anxious* when we want to gain knowledge on adolescent preferences in general.

XSL•FO
RenderX

Furthermore, demographic data on ethnicity and socioeconomic status were not collected for participants in phases 1 and 2, and it is unclear whether our sample was demographically representative. Future research could focus on collecting these data and recruiting participants from different locations (ie, schools in different areas, counseling services, and mental health services) to enhance diversity.

In addition, only 4 adolescents participated in the usability tests in phase 2. Although representing both the youngest and oldest adolescents in the intended age group (aged 12 and 16 years, respectively), the small number of participants limits the external validity of the results. It may be that the results obtained from the usability tests are specific to the platform and content viewed in this study. In the future, conducting usability tests online or "on-site" in places where adolescents regularly spend time (ie, in class or at youth centers) could aid the recruitment of more adolescents and increase sample size.

In this study, we emphasized the importance of systematically including adolescents as end users of an intervention. However, parents and clinicians may equally represent end users of this intervention and constitute valuable sources of information in a design and implementation context. We did include clinical psychologists working in routine care settings as part of the expert group to qualify the program manuscript but did not systematically collect data from them. This poses a limitation of this study, and it would be beneficial in the future to include all intended end users in the development processes.

It may also be that adolescents could have been involved more in the design of the intervention and at an earlier stage. In a study by Ludlow et al [16], the research team used co-design methods to understand preferences and perceptions of the design, functions, and engagement features within a new digital mental health platform. The aforementioned study used generative toolkits to facilitate the co-design processes through production of artifacts. These artifacts (and output from previous design phases) were then used to qualify the design of their new digital mental health platform for young people. This type of end-user involvement allows young people to hold a central position as experts at very early stages of the development process. However, it may also be extensive and require a lot of time and resources that may not be available.

When comparing this study to the one by Ludlow et al [16], it becomes apparent that, despite using user-centered methods, researchers have the power to define *what* end-user perspectives to include and *how* to investigate them. There may not be a way to solve this paradox, but it underlines how important it is to be aware of one's own impact as a researcher when engaging in these practices.

Conclusions

DMHIs have the potential to overcome treatment barriers such as stigma and costs and increase access to effective treatment at an early stage in symptom development. However, most DMHIs suffer from low treatment adherence. It has been proposed that the lack of end-user involvement in the development of intervention content and platforms may be of particular importance when interpreting the low adherence to DMHIs. In this study, we used user involvement practices at different stages of development to investigate adolescent and parent preferences on specific design components. Information from these practices was implemented when developing intervention content. This study emphasizes the importance of understanding the design process as iterative, where content prototypes are subjected to multiple phases of evaluation. This ensures the continued alignment of the intervention with end-user needs and may help establish the validity of the intervention for implementation in routine care settings. Future research may use user involvement practices such as those described in this paper to understand how to individualize intervention content and patient courses. In addition, user involvement practices could be used to investigate adolescent preferences on specific treatment components (eg, relaxation techniques and behavioral experiments) and the order in which they are presented (ie, in a specific order or as optional components).

Acknowledgments

The authors would like to thank all participating adolescents and families for their contributions to the development of the CoolMinds intervention. They greatly appreciate their interest and participation in the CoolMinds project. The authors would also like to thank psychologist Cecilie Holmen Frost for her role in writing and editing the CoolMinds intervention manuscript and the Development and Implementation Unit at the Centre for Digital Psychiatry for their contributions. Specifically, the authors would like to thank team leader in charge of user involvement Nicole Fronc Dumanski for her contribution to the planning and execution of the user involvement workshop, graphic designer Nicolai Drejer Walther for his extensive work on the graphic identity and content of the CoolMinds intervention, and consultant Alberthe Møller for setting up the intervention content on the digital platform. The development of CoolMinds is funded by a grant from the Mental Health Services of the Region of Southern Denmark and a grant from TrygFonden. Generative artificial intelligence was not used in the writing of this manuscript.

Data Availability

The data used in this manuscript contain sensitive and potentially identifiable information regarding both children and caregivers, and all participants were guaranteed full anonymity. As such, and in line with Danish Data Protection Agency regulations, the data used in this study are not publicly available. Unidentifiable data excerpts can be made available upon request to the corresponding author (nikita@psy.au.dk). Further data access requests may be sent to the Technology Transfer Office at Aarhus University (tto@au.dk).

XSL•FO
RenderX

Authors' Contributions

KM, JJL, MT, NMS, and HS contributed to conceptualization of the study. NMS conducted data curation. NMS completed the formal analysis. KM, MT, and JJL acquired funding. NMS, HS, and JJL collected data. KM, JJL, MT, NMS, and HS developed the methodology. NMS, and HS administered the project. NMS, JJL, and KM provided study materials. JJL, KM, and MT provided supervision. KM, JJL, and MT assisted with knowledge on previous and ongoing studies to validate the research conducted in this paper. NMS prepared data for presentation in the published work. NMS drafted the paper. JJL, KM, MT, and HS reviewed and edited the draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Guidance for Reporting Involvement of Patients and the Public checklist in long form.
[DOCX File , 19 KB - formative_v9i1e66966_app1.docx]

Multimedia Appendix 2
Interview guides.
[DOCX File , 22 KB - formative_v9i1e66966_app2.docx]

Multimedia Appendix 3
Results from the usability and feasibility tests.
[DOCX File , 39 KB - formative_v9i1e66966_app3.docx]

Multimedia Appendix 4
Overview of the content of the adolescent and parent program versions used when conducting the feasibility trial (version 1) and the randomized controlled trial (version 2).
[DOCX File , 21 KB - formative_v9i1e66966_app4.docx]

Multimedia Appendix 5
CoolMinds intervention content.
[DOCX File , 46164 KB - formative_v9i1e66966_app5.docx]

References

1. Nearchou FA, Bird N, Costello A, Duggan S, Gilroy J, Long R, et al. Personal and perceived public mental-health stigma as predictors of help-seeking intentions in adolescents. J Adolesc 2018 Jul;66:83-90. [doi: 10.1016/j.adolescence.2018.05.003] [Medline: 29800758]
2. Radez J, Reardon T, Creswell C, Orchard F, Waite P. Adolescents' perceived barriers and facilitators to seeking and accessing professional help for anxiety and depressive disorders: a qualitative interview study. Eur Child Adolesc Psychiatry 2022 Jun;31(6):891-907 [FREE Full text] [doi: 10.1007/s00787-020-01707-0] [Medline: 33502596]
3. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
4. Bennett SD, Cuijpers P, Ebert DD, McKenzie Smith M, Coughtrey AE, Heyman I, et al. Practitioner review: unguided and guided self-help interventions for common mental health disorders in children and adolescents: a systematic review and meta-analysis. J Child Psychol Psychiatry 2019 Aug;60(8):828-847. [doi: 10.1111/jcpp.13010] [Medline: 30775782]
5. Pennant ME, Loucas CE, Whittington C, Creswell C, Fonagy P, Fuggle P, et al. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
6. Vigerland S, Lenhard F, Bonnert M, Lalouni M, Hedman E, Ahlen J, et al. Internet-delivered cognitive behavior therapy for children and adolescents: a systematic review and meta-analysis. Clin Psychol Rev 2016 Dec;50:1-10 [FREE Full text] [doi: 10.1016/j.cpr.2016.09.005] [Medline: 27668988]
7. Wickersham A, Barack T, Cross L, Downs J. Computerized cognitive behavioral therapy for treatment of depression and anxiety in adolescents: systematic review and meta-analysis. J Med Internet Res 2022 Apr 11;24(4):e29842 [FREE Full text] [doi: 10.2196/29842] [Medline: 35404263]
8. Clarke AM, Kuosmanen T, Barry MM. A systematic review of online youth mental health promotion and prevention interventions. J Youth Adolesc 2015 Jan;44(1):90-113. [doi: 10.1007/s10964-014-0165-0] [Medline: 25115460]

XSL•FO
RenderX

9. March S, Spence SH, Donovan CL. The efficacy of an internet-based cognitive-behavioral therapy intervention for child anxiety disorders. J Pediatr Psychol 2009 Jun;34(5):474-487. [doi: 10.1093/jpepsy/jsn099] [Medline: 18794187]
10. March S, Spence SH, Donovan CL, Kenardy JA. Large-scale dissemination of internet-based cognitive behavioral therapy for youth anxiety: feasibility and acceptability study. J Med Internet Res 2018 Jul 04;20(7):e234 [FREE Full text] [doi: 10.2196/jmir.9211] [Medline: 29973338]
11. Spence SH, Donovan CL, March S, Kenardy JA, Hearn CS. Generic versus disorder specific cognitive behavior therapy for social anxiety disorder in youth: a randomized controlled trial using internet delivery. Behav Res Ther 2017 Mar;90:41-57. [doi: 10.1016/J.BRAT.2016.12.003]
12. Wuthrich VM, Rapee RM, Cunningham MJ, Lyneham HJ, Hudson JL, Schniering CA. A randomized controlled trial of the Cool Teens CD-ROM computerized program for adolescent anxiety. J Am Acad Child Adolesc Psychiatry 2012 Mar;51(3):261-270. [doi: 10.1016/j.jaac.2011.12.002] [Medline: 22365462]
13. Waite P, Marshall T, Creswell C. A randomized controlled trial of internet-delivered cognitive behaviour therapy for adolescent anxiety disorders in a routine clinical care setting with and without parent sessions. Child Adolesc Ment Health 2019 Sep;24(3):242-250. [doi: 10.1111/camh.12311] [Medline: 32677216]
14. Hill C, Creswell C, Vigerland S, Nauta MH, March S, Donovan C, et al. Navigating the development and dissemination of internet cognitive behavioral therapy (iCBT) for anxiety disorders in children and young people: a consensus statement with recommendations from the #iCBTLorentz Workshop Group. Internet Interv 2018 Feb 19;12:1-10 [FREE Full text] [doi: 10.1016/j.invent.2018.02.002] [Medline: 30135763]
15. Fleming T, Merry S, Stasiak K, Hopkins S, Patolo T, Ruru S, et al. The importance of user segmentation for designing digital therapy for adolescent mental health: findings from scoping processes. JMIR Ment Health 2019 May 08;6(5):e12656 [FREE Full text] [doi: 10.2196/12656] [Medline: 31066705]
16. Ludlow K, Russell JK, Ryan B, Brown RL, Joynt T, Uhlmann LR, et al. Co-designing a digital mental health platform, "momentum", with young people aged 7-17: a qualitative study. Digit Health 2023 Nov 28;9:20552076231216410 [FREE Full text] [doi: 10.1177/20552076231216410] [Medline: 38033517]
17. De Vito Dabbs A, Myers BA, Mc Curry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183. [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
18. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
19. Wozney L, Baxter P, Newton AS. Usability evaluation with mental health professionals and young people to develop an internet-based cognitive-behaviour therapy program for adolescents with anxiety disorders. BMC Pediatr 2015 Dec 16;15:213 [FREE Full text] [doi: 10.1186/s12887-015-0534-1] [Medline: 26675420]
20. Hill C, Chessell C, Percy R, Creswell C. Online Support and Intervention (OSI) for child anxiety: a case series within routine clinical practice. Behav Cogn Psychother 2022 May 04;50(4):429-445. [doi: 10.1017/s1352465822000157]
21. Hill C, Reardon T, Taylor L, Creswell C. Online support and intervention for child anxiety (OSI): development and usability testing. JMIR Form Res 2022 Apr 13;6(4):e29846 [FREE Full text] [doi: 10.2196/29846] [Medline: 35416781]
22. Moor S, Williman J, Drummond S, Fulton C, Mayes W, Ward N, et al. 'E' therapy in the community: examination of the uptake and effectiveness of BRAVE (a self-help computer programme for anxiety in children and adolescents) in primary care. Internet Interv 2019 Dec;18:100249. [doi: 10.1016/J.INVENT.2019.100249]
23. Vigerland S, Fredlander S, Aspvall K, Jolstedt M, Lenhard F, Mataix-Cols D, et al. Effectiveness of internet-delivered cognitive behavioural therapy for anxiety and obsessive-compulsive disorders within routine clinical care in rural Sweden. Internet Interv 2024 Jun;36:100738 [FREE Full text] [doi: 10.1016/j.invent.2024.100738] [Medline: 38617387]
24. Skaarnes H, Sørensen NM, Wisnewski AH, Lomholt JJ, Thastum M, McLellan L, et al. Internet-based cognitive behavioral intervention for adolescents with anxiety disorders: a study protocol for a parallel three armed randomized controlled trial. Trials 2024 Oct 13;25(1):674 [FREE Full text] [doi: 10.1186/s13063-024-08511-0] [Medline: 39396991]
25. Staniszewska S, Brett J, Simera I, Seers K, Mockford C, Goodlad S, et al. GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research. BMJ 2017 Aug 02;358:j3453 [FREE Full text] [doi: 10.1136/bmj.j3453] [Medline: 28768629]
26. Horvath KJ, Ecklund AM, Hunt SL, Nelson TF, Toomey TL. Developing internet-based health interventions: a guide for public health researchers and practitioners. J Med Internet Res 2015 Jan 23;17(1):e28 [FREE Full text] [doi: 10.2196/jmir.3770] [Medline: 25650702]
27. Whiteside SP, Sim LA, Morrow AS, Farah WH, Hilliker DR, Murad MH, et al. A meta-analysis to guide the enhancement of CBT for childhood anxiety: exposure over anxiety management. Clin Child Fam Psychol Rev 2020 Mar;23(1):102-121 [FREE Full text] [doi: 10.1007/s10567-019-00303-2] [Medline: 31628568]
28. Jolstedt M, Wahlund T, Lenhard F, Ljótsson B, Mataix-Cols D, Nord M, et al. Efficacy and cost-effectiveness of therapist-guided internet cognitive behavioural therapy for paediatric anxiety disorders: a single-centre, single-blind, randomised controlled trial. Lancet Child Adolesc Health 2018 Nov;2(11):792-801. [doi: 10.1016/s2352-4642(18)30275-x]

XSL•FO
RenderX

29. Nordh M, Wahlund T, Jolstedt M, Sahlin H, Bjureberg J, Ahlen J, et al. Therapist-guided internet-delivered cognitive behavioral therapy vs internet-delivered supportive therapy for children and adolescents with social anxiety disorder: a randomized clinical trial. JAMA Psychiatry 2021 Jul 01;78(7):705-713 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.0469] [Medline: 33978699]
30. Stjerneklar S, Hougaard E, McLellan LF, Thastum M. A randomized controlled trial examining the efficacy of an internet-based cognitive behavioral therapy program for adolescents with anxiety disorders. PLoS One 2019 Sep 18;14(9):e0222485 [FREE Full text] [doi: 10.1371/journal.pone.0222485] [Medline: 31532802]
31. Vigerland S, Ljótsson B, Thulin U, Öst LG, Andersson G, Serlachius E. Internet-delivered cognitive behavioural therapy for children with anxiety disorders: a randomised controlled trial. Behav Res Ther 2016 Jan;76:47-56 [FREE Full text] [doi: 10.1016/j.brat.2015.11.006] [Medline: 26649465]
32. Higa-McMillan CK, Francis SE, Rith-Najarian L, Chorpita BF. Evidence base update: 50 years of research on treatment for child and adolescent anxiety. J Clin Child Adolesc Psychol 2016;45(2):91-113. [doi: 10.1080/15374416.2015.1046177] [Medline: 26087438]
33. Luo A, McAloon J. Potential mechanisms of change in cognitive behavioral therapy for childhood anxiety: a meta-analysis. Depress Anxiety 2021 Feb;38(2):220-232. [doi: 10.1002/da.23116] [Medline: 33225527]
34. Bertie LA, Hudson JL. CBT for childhood anxiety: reviewing the state of personalised intervention research. Front Psychol 2021 Nov 26;12:722546 [FREE Full text] [doi: 10.3389/fpsyg.2021.722546] [Medline: 34899467]
35. Smart K, Smith L, Harvey K, Waite P. The acceptability of a therapist-assisted internet-delivered cognitive behaviour therapy program for the treatment of anxiety disorders in adolescents: a qualitative study. Eur Child Adolesc Psychiatry 2023 Apr;32(4):661-673 [FREE Full text] [doi: 10.1007/s00787-021-01903-6] [Medline: 34746976]
36. Radomski AD, Bagnell A, Curtis S, Hartling L, Newton AS. Examining the usage, user experience, and perceived impact of an internet-based cognitive behavioral therapy program for adolescents with anxiety: randomized controlled trial. JMIR Ment Health 2020 Feb 07;7(2):e15795 [FREE Full text] [doi: 10.2196/15795] [Medline: 32022692]
37. Spence SH, March S, Donovan CL. Social support as a predictor of treatment adherence and response in an open-access, self-help, internet-delivered cognitive behavior therapy program for child and adolescent anxiety. Internet Interv 2019 Aug 22;18:100268 [FREE Full text] [doi: 10.1016/j.invent.2019.100268] [Medline: 31890621]
38. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
39. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
40. Neergaard MA, Olesen F, Andersen RS, Sondergaard J. Qualitative description - the poor cousin of health research? BMC Med Res Methodol 2009 Jul 16;9:52 [FREE Full text] [doi: 10.1186/1471-2288-9-52] [Medline: 19607668]
41. Garrido S, Millington C, Cheers D, Boydell K, Schubert E, Meade T, et al. What works and what doesn't work? A systematic review of digital mental health interventions for depression and anxiety in young people. Front Psychiatry 2019 Nov 13;10:759. [doi: 10.3389/fpsyt.2019.00759] [Medline: 31798468]
42. Wood D, Bruner JS, Ross G. The role of tutoring in problem solving. J Child Psychol Psychiatry 1976 Apr;17(2):89-100. [doi: 10.1111/j.1469-7610.1976.tb00381.x] [Medline: 932126]
43. Stjerneklar S, Hougaard E, Thastum M. Guided internet-based cognitive behavioral therapy for adolescent anxiety: Predictors of treatment response. Internet Interv 2019 Mar;15:116-125 [FREE Full text] [doi: 10.1016/j.invent.2019.01.003] [Medline: 30792963]
44. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
45. Ludlow K, March S, Russell JK, Ryan B, Brown RL, Hides L, et al. Co-designing a new digital mental health platform, 'momentum', with caregivers of young people aged 7–17. Digit Health 2024 Aug 02;10. [doi: 10.1177/20552076241263693]
46. McCall HC, Hadjistavropoulos HD, Sundström CR. Exploring the role of persuasive design in unguided internet-delivered cognitive behavioral therapy for depression and anxiety among adults: systematic review, meta-analysis, and meta-regression. J Med Internet Res 2021 Apr 29;23(4):e26939 [FREE Full text] [doi: 10.2196/26939] [Medline: 33913811]
47. Patterson VC, Rossi MA, Pencer A, Wozney L. An internet-based cognitive behavioral therapy program for anxiety and depression (tranquility): adaptation co-design and fidelity evaluation study. JMIR Form Res 2022 Feb 02;6(2):e33374. [doi: 10.2196/33374] [Medline: 34910660]

Abbreviations

CBT: cognitive behavioral therapy
DMHI: digital mental health intervention
ICBT: internet-based cognitive behavioral therapy
RCT: randomized controlled trial

XSL•FO
RenderX

REDCap: Research Electronic Data Capture

Edited by A Mavragani; submitted 27.09.24; peer-reviewed by C Donovan, L Kagan; comments to author 15.01.25; revised version received 03.02.25; accepted 02.03.25; published 08.04.25.

Please cite as:
Sørensen NM, Skaarnes H, Mathiasen K, Thastum M, Lomholt JJ
Developing an Internet-Based Cognitive Behavioral Therapy Intervention for Adolescents With Anxiety Disorders: Design, Usability, and Initial Evaluation of the CoolMinds Intervention
JMIR Form Res 2025;9:e66966
URL: https://formative.jmir.org/2025/1/e66966
doi:10.2196/66966
PMID:

©Nikita Marie Sørensen, Helene Skaarnes, Kim Mathiasen, Mikael Thastum, Johanne Jeppesen Lomholt. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Opportunities and Challenges Surrounding the Use of Wearable Sensor Bracelets for Infectious Disease Detection During Hajj: Qualitative Interview Study

Noha Maddah[1,2], BSc, MSc, MPH; Arpana Verma[3], MBChB, MPH, PhD; John Ainsworth[1], MSc, PhD

[1]Division of Informatics Imaging and Data Sciences, School of Health Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, Centre for Health Informatics, The University of Manchester, Manchester, United Kingdom

[2]Department of Health Services and Hospitals Administration, Faculty of Economics and Administration, King Abdulaziz University, Jeddah, Saudi Arabia

[3]Division of Population Health, Health Services Research & Primary Care, Faculty of Biology, Medicine and Health, Manchester Academic Health Sciences Centre, The University of Manchester, Manchester, United Kingdom

Corresponding Author:
Noha Maddah, BSc, MSc, MPH
Division of Informatics Imaging and Data Sciences, School of Health Sciences, Faculty of Biology, Medicine and Health
Manchester Academic Health Science Centre, Centre for Health Informatics
The University of Manchester
Vaughn House
Manchester, M13 9GB
United Kingdom
Phone: 44 1612757675
Email: noha.maddah@postgrad.manchester.ac.uk

Abstract

Background: Wearable sensor bracelets have gained interest for their ability to detect symptomatic and presymptomatic infections through alterations in physiological indicators. Nevertheless, the use of these devices for public health surveillance among attendees of large-scale events such as hajj, the Islamic religious mass gathering held in Saudi Arabia, is currently in a nascent phase.

Objective: This study aimed to explore hajj stakeholders' perspectives on the use of wearable sensor bracelets for disease detection.

Methods: We conducted a qualitative, theoretically informed, interview-based study from March 2022 to October 2023 involving a diverse sample of hajj stakeholders, including technology experts, health care providers, and hajj service providers. The study was guided by the task-technology fit model and the unified theory of acceptance and use of technology to provide a comprehensive understanding of the factors influencing the acceptance and use of the technology. Semistructured in-depth interviews were used to capture perspectives on using wearable sensor bracelets for infectious disease detection during hajj. Thematic analysis of interview transcripts was conducted.

Results: A total of 14 individuals were interviewed. In total, 4 main themes and 13 subthemes emerged from the study, highlighting crucial challenges, considerations, recommendations, and opportunities in the use of wearable sensor bracelets for the presymptomatic detection of infectious diseases during hajj. Implementing wearable sensor bracelets for disease detection during hajj faces obstacles from multiple perspectives, encompassing users, implementing stakeholders, and technological factors. Hajj stakeholders were concerned about the substantial financial and operational barriers. The motivation of implementing stakeholders and users is essential for the acceptance and uptake of devices during hajj. Successful integration of wearables into the hajj surveillance system depends on several factors, including infrastructure, device features, suitable use cases, training, and a smooth organizational integration process.

Conclusions: This study provides valuable insights into the potential opportunities and challenges of adopting wearable sensor bracelets for disease detection during hajj. It offers essential factors to consider and important suggestions to enhance comprehension and ensure the effective implementation of this technology.

(JMIR Form Res 2025;9:e60484) doi:10.2196/60484

XSL•FO
RenderX

KEYWORDS

wearable sensor; unified theory of acceptance and use of technology; task-technology fit; hajj; presymptomatic detection; infectious diseases; artificial intelligence

Introduction

New Generation of Digital Tools for Public Health Surveillance

Understanding and reacting to health trends affecting communities is a core function of public health systems [1]. Public health surveillance is the tool used to respond to these health trends through "the ongoing, systematic collection, analysis, and interpretation of health-related data essential to planning, implementation, and evaluation of public health practice" [1]. The COVID-19 pandemic has brought to light the urgent need for real-time public health surveillance to enhance the use of evidence in decision-making [1]. The 2020 Riyadh Declaration was formulated as a response to the global pandemic, aiming to propose a series of suggestions that would effectively tackle the shortcomings observed in global public health response systems. The Riyadh Declaration emphasized the significance of implementing digital health solutions that are both scalable and sustainable, as well as the adoption of health intelligence [1].

Traditional public health initiatives targeted at preventing the spread of infectious diseases such as the recent COVID-19 have limitations that digital health technologies could address [2]. The World Health Organization has defined digital health technologies as "the field of knowledge and practice associated with the development and use of digital technologies to improve health" [3]. The use of digital health tools during the global pandemic has exhibited their efficacy in fighting, mitigating, and controlling the transmission of emerging infectious diseases [2]. There is a significant potential for adopting digital surveillance as a valuable tool in monitoring and controlling infectious diseases [4]. Internet of Things (IoT) technological advancement has proven its ability to augment and optimize conventional preventative public health strategies and methodologies [2]. These internet-based technologies possess certain characteristics that enable the expedited detection of infectious diseases [4]. The term IoT pertains to a network of interconnected objects, devices, and systems that engage in electronic communication and receive, process, and send digital data with minimal human involvement [5]. IoT is expected to reach its "plateau of productivity" during the next 5 to 10 years [5].

Detection is the initial and important aspect in preventing infectious diseases [5]. In the context of these diseases, the computation speed plays a crucial role in reducing the time required for their diagnosis, facilitating prompt implementation of preventive measures. The performance of system automation has been found to enhance surveillance's efficacy in widespread disease transmission. In addition, it has proven to be an effective means of reducing the workload of medical workers during pandemics [4].

Wearable Technologies and Body Sensors

Wearable technologies are integral to widespread IoT technical advancements [6]. Wearable technologies are intelligent electronic devices worn on the body to measure, analyze, and transfer information. The data may encompass many bodily indicators associated with essential physiological parameters and levels of physical exertion. These wearables have the capability to provide information to the user by means of display technology or vibrotactile feedback [7].

Wearable sensors have garnered considerable attention in the past 10 years, with a particular focus on their application within the health care sector [8,9]. Their ability to establish connections with other devices and the IoT, facilitating data exchange, has led to their designation as "smart" devices, setting them apart from traditional, nonconnected step counters, wristwatches, or clothing [10]. Wearable sensors have the capability to monitor, collect, and send data [6]. They offer precise, credible, and real-time data regarding an individual's behaviors and activities, hence aiding users in identifying potential issues. Wearables collect raw data from measurements using sensors to store them and use them for things such as performance evaluation and ongoing health monitoring [11]. These sensors can be applied in several areas, such as illness monitoring, diagnosis and treatment, and health management [12].

Various terms are used to refer to wearable sensor devices that capture and analyze real-time physical data in a nonintrusive manner. These terms encompass *smartwatches* [9,13], *smart sensors* [14], *wearable activity trackers* [15], *wearable sensors* [16], *wearable physiological sensors* [17], *integrated biosensors* [18], and *sensor bracelets* [19]. In the context of this research, the phrase "sensor bracelet" will be used.

Extensive research has been conducted on wearables to investigate their potential in health monitoring. However, there is a noticeable trend in current research toward applying wearables for the presymptomatic detection of infectious diseases [9,19]. This transition signifies the growing importance of early detection of diseases, especially within the context of public health and large-scale occurrences. The use of wearable technology has emerged as a significant factor in this pursuit, presenting the possibility of identifying infections before the manifestation of symptoms. This capability facilitates prompt interventions and preventive actions. The increasing emphasis on this matter highlights the significant contribution that wearables can make in augmenting health care and methods for disease control.

Wearable devices (sensor bracelets) have gained attention for their ability to anticipate both symptomatic and presymptomatic identification of viral respiratory infections by sensing changes in physiological markers, as highlighted in the recent literature [19,20]. Mitratza et al [16] conducted a systematic review to assess the value of a range of wearable sensor devices in detecting COVID-19. In their review, they noted that, while wearable devices show promising potential for detecting

XSL•FO
RenderX

COVID-19, the findings are still in the early stages. They emphasized the need to consider how differences in sensor methods, data processing, and algorithms affect the detection of infection-related changes in physiological measurements, as well as the importance of addressing sources of bias. Building on this, Risch et al [19] assessed the use of a smart bracelet in the early detection of changes in physiological parameters related to COVID-19. With a low false-positive rate, their wearable sensor bracelet algorithm was able to identify 71% of COVID-19 cases up to 2 days before symptoms manifested. This demonstrates the efficacy of using wearable technologies for the early detection of diseases. Mishra et al [9] conducted a study whereby they demonstrated the use of data from smartwatches to detect respiratory infections such as COVID-19 in real time within a large-scale scenario. The study identified 63% of COVID-19 infections with a low false-positive rate. Furthermore, Hirten et al [2] evaluated the longitudinal collection of heart rate variability (HRV) metrics from a wearable device for the prediction of COVID-19. Their study indicated that HRV data from the Apple Watch can predict COVID-19 and identify symptoms. In addition, alterations in HRV were observed before a positive polymerase chain reaction test, indicating its potential for early infection detection.

Hajj, the Largest Annual Religious Mass Gathering, and Disease Transmission

Safety and security have consistently been of utmost importance within the framework of hajj, a significant religious event that attracts participants worldwide. This annual massive religious gathering in Makkah, Saudi Arabia, attracts almost 3 million pilgrims from 180 countries. These pilgrims spend several weeks in the cities of Makkah and Madinah. The problem of overcrowding and the lack of compliance with government-mandated preventative measures for diseases have been highlighted in several previous studies [17,21-23]. This lack of compliance contributed to a long-standing high occurrence of respiratory infections before the COVID-19 outbreak. Low adherence among pilgrims has been attributed to a number of factors, including insufficient coverage and limited access to information [24].

Using intelligent and comprehensive monitoring tools for participants in this mass event is crucial to reducing the risk of future tragedies. Saudi Arabia uses several technical methods such as spatial computing, crowd simulation, mobile apps, and big data analytics for hajj activities [25].

In recent times, there has been an increasing focus on incorporating state-of-the-art technologies to tackle the distinctive difficulties linked to hajj. The hajj pilgrimage, characterized by its large-scale congregation of millions of pilgrims, poses significant logistical, health, and safety complexities. Disease surveillance has emerged as a crucial component of hajj management, progressively evolving over time by assimilating insights gained from previous operations.

The Kingdom of Saudi Arabia has been at the forefront of implementing digital solutions to improve the pilgrimage experience as part of the national strategy and the 2030 vision. One of the notable technological developments in this context involves the application of the smart bracelet initiative, which involved the collaboration of multiple governmental authorities. In 2021, the Pilgrims' Smart Bracelet initiative was introduced, offering 5000 pilgrims bracelets equipped with data storage and transmission capabilities, thus contributing to the IoT. The main purpose of the bracelet was to provide extensive information on the pilgrim, including their health condition. The device enabled the tracking of important health indicators such as blood oxygen saturation and pulse rate while also offering functionalities for requesting immediate medical or security support, guaranteeing swift reaction and help. In addition, pilgrims were given awareness messages via the bracelet they wore. The initiative's implementation entailed incorporating and enhancing business models and systems, spanning both operational and technological facets of the project [26].

The practice of postmortem detection for infectious diseases underscores the criticality of real-time data and analytics to avert their transmission, which has severe social and economic repercussions for citizens worldwide [5]. Within the framework of hajj, it is imperative to not only identify diseases before symptoms appear but also find and follow the individuals who have been in contact with infected individuals. This requires incorporating tracking elements into the device. The current state of affairs presents an opportune environment for scientific inquiry specifically aimed at recognizing and addressing nascent hazards and difficulties to predict and alleviate them. Hence, this study endeavored to investigate the opportunities and challenges surrounding using wearable sensor bracelets for presymptomatic detection of infectious diseases during hajj.

Methods

Theoretical Framework

Technology acceptance models have been criticized for their limited ability to evaluate the extent to which wearable technology offers the requisite functionalities to attain desired outcomes. This concern can be effectively resolved by integrating the task-technology fit (TTF) and technology acceptance models [27].

Studies have shown that integrating the TTF model and the unified theory of acceptance and use of technology (UTAUT) enhances the predictive strength for technology adoption. The TTF-UTAUT model integrates 2 key frameworks to predict technology adoption [28]. The TTF model is defined as the interdependence between individual (a technology user), technology (data, hardware, software tools, and the services they provide), and task (an activity carried out by individuals to produce the required output) characteristics [26]. The UTAUT suggests that the actual use of technology is determined by behavioral intention. The perceived likelihood of adopting the technology depends on the direct effect of 4 key constructs: performance expectancy, effort expectancy, social influence, and facilitating conditions [29].

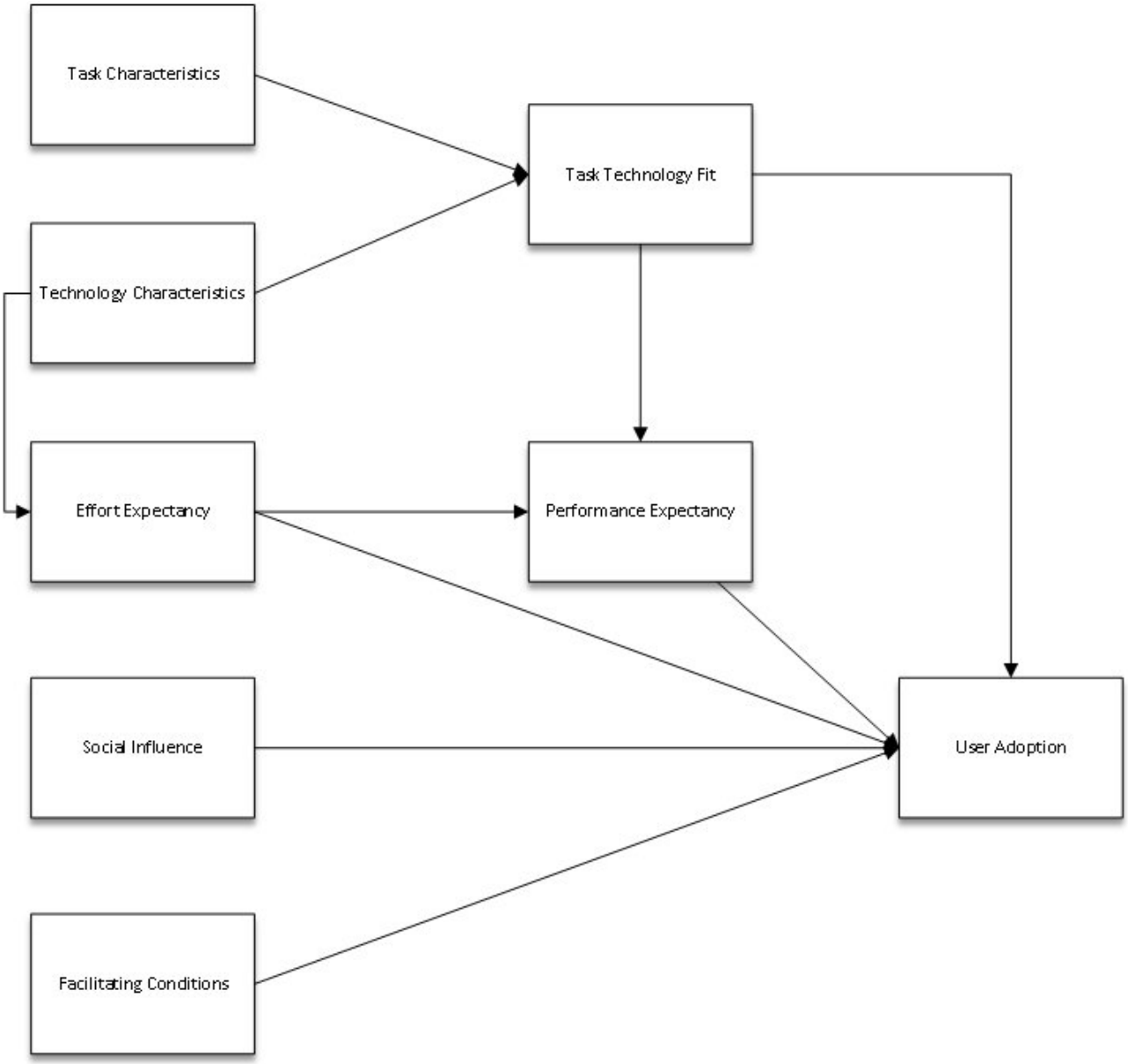
The UTAUT is considered beneficial for the complex nature of this research, which involves evaluating several interrelated factors in disciplines such as health informatics and public health during large-scale events. The UTAUT can provide insights into the determinants that affect the adoption and use of

XSL•FO
RenderX

wearable sensor bracelets among various stakeholders participating in the hajj pilgrimage. To provide a more thorough explanation, the TTF model was integrated with the UTAUT. The TTF model can be used to evaluate the degree to which the features of the technology align with the requirements and demands of its users, including both pilgrims and authorities. This integration improves understanding and focuses on particular elements inside the UTAUT.

To aid in effectively applying wearable sensor bracelets during the hajj pilgrimage, this integrated approach can offer a comprehensive understanding of stakeholders' viewpoints. Figure 1 [28] presents the model used.

Figure 1. Task-technology fit and unified theory of acceptance and use of technology.



Study Design

We conducted in-depth, semistructured, one-to-one qualitative interviews from March 2022 to October 2023 with 14 selected hajj stakeholders, organizers, and professionals working in the field to capture their perspectives on using wearable sensor bracelets for infectious disease detection during hajj.

Data Collection and Sampling Strategy

Overview

A nested cohort approach was used to explore stakeholders' perspectives [30]. Participants were identified through purposive sampling and snowball sampling techniques. Purposive sampling is a method used to gather information and gain a comprehensive understanding of a certain topic of interest. This methodology entails the selection of data cases based on their capacity to yield data that are rich in information for comprehensive analysis [31].

A purposive sampling strategy was used to identify participants who were directly involved in the smart bracelet initiative. This resulted in a core group of 4 participants who held senior positions within the organizational hierarchy. The key stakeholders were identified through the organizational website and were reached purposefully via email. These individuals

XSL•FO
RenderX

offered comprehensive perspectives on the initiative as a result of their direct expertise and authoritative decision-making positions. Concurrently, a more extensive cohort of 10 participants selected using snowball sampling was incorporated to gather insights into the general use of technology during hajj, specifically from individuals who directly interacted with pilgrims. Although these individuals were not directly involved in the smart bracelet initiative, their contact with pilgrims offered vital contextual information. We reached out to these participants directly via their contact numbers.

The inclusion and exclusion criteria for participants are detailed in Textbox 1. This study adhered to the guidelines outlined in the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist [32].

Textbox 1. Inclusion and exclusion criteria for participants in this study.

Inclusion criteria

- Individuals in senior positions within the organizational hierarchy who were directly involved in the smart bracelet initiative, including those with technological, medical, organizational, or logistical expertise
- Individuals who have experience and are present in service organizations during hajj (eg, *tawafa* organizations) and can offer their perspective on the technologies used and how pilgrims interact with or accept them

Exclusion criteria

- Individuals who are not directly involved in the smart bracelet initiative or do not occupy senior positions
- Individuals who lack pertinent experience or insight into the adoption and use of technology by pilgrims during hajj.

Interviews

We used a systematic approach to engage with our potential participants before the interviews started. The interview questions were then given to participants who replied favorably and indicated their desire to take part. This action guaranteed that participants had enough time to prepare for the subsequent interview, promoting a deeper and more thoughtful discussion. We then scheduled the interviews after allowing participants to receive the interview questions and prepare.

Interview Topic Guide

The interviews were guided by a semistructured topic guide exploring professional background, experience, and involvement with the 2021 Pilgrims' Smart Bracelet initiative. Questions were based on existing literature and organizations' online information to understand the current capabilities and existing tools used during hajj in addition to opportunities, gaps, and challenges associated with implementing wearable sensor bracelets for infectious disease detection during hajj. Questions were tailored to the specific characteristics of participants and evolved in light of emerging findings. We explored the advancements of existing tools and technologies used during hajj for public health surveillance, their limitations, and stakeholders' vision and perceptions regarding using wearable sensor bracelets for disease detection during hajj.

The interview questions for key stakeholders who were directly involved in the smart bracelet initiative primarily addressed the capabilities of the surveillance system, the technologies and tools used during hajj to support the event, their opinions on wearables for disease detection, and their insights into the smart bracelet initiative. In contrast, interviews conducted with participants who had minimal involvement in the smart bracelet initiative but were highly involved in assisting pilgrims focused on their experiences and viewpoints regarding the acceptability of the latest technology deployed during hajj by the pilgrims. A copy of the topic guide can be found in Multimedia Appendix 1.

Digital audio recorders with encryption were used to record the interviews. Interviews were transcribed verbatim, read numerous times, and annotated to fully immerse us in the facts.

Data Analysis

Thematic analysis of interview transcripts was conducted [31]. This study used an inductive thematic analysis to investigate participants' experiences and perceptions regarding the use of wearable sensor bracelets for disease detection during hajj. The UTAUT and TTF framework variables served as a guide for the interview questions, but themes were allowed to naturally arise from the data. In the subsequent analysis, a deductive approach was applied to identify relationships between the emergent themes and the theoretical constructs of the UTAUT and TTF model. This hybrid approach allowed for both the discovery of novel insights and the integration of findings with established theoretical frameworks.

The primary author, NM, led the data analysis by integrating the UTAUT [33] and TTF models [34] to create a comprehensive framework. This framework posits that the user's intention to use technology and their acceptance of it are influenced by many criteria derived from the UTAUT as well as the degree of compatibility between the task at hand and the technology being used [28].

In this analysis, we explored the use of wearable sensor bracelets for disease detection during hajj. We thoroughly evaluated several aspects such as the integration of technology, development of systems, augmentation of infrastructure, and acceptance among users and suppliers. This comprehensive exploration allowed us to understand the intricate interplay of factors influencing the successful deployment of wearable sensor technology within the unique context of hajj.

Upon completing data collection in our study, we promptly commenced the analysis process to identify any significant findings.

XSL•FO
RenderX

We eventually arrived at the realization that we had obtained enough data to attain data saturation when we noticed that the same subjects kept cropping up during our observations of the data-gathering process and no new ones were introduced. After completing the data collection and analysis, it became evident that no new themes were developing. This confirmed that we had achieved a state of saturation in our findings.

To improve the dependability and verification, synopses of discoveries were regularly communicated to other researchers engaged in the study subsequent to each interview with a participant. The completed transcripts and field notes were systematically arranged and categorized in NVivo (version 12; QSR International) using researcher-derived codes (latent codes) obtained from the interview guide. We followed the characteristics of an inductive coding strategy to guarantee thoroughness and complexity in our investigation [31].

Ethical Considerations

According to the ethical guidelines set forth by the University of Manchester, this research study did not require independent ethical review and was categorized as low-risk research ("Work with professionals; including interviews, focus groups, questionnaires, etc" [35]). All study data were fully anonymized, with no personal identifiers collected, and were securely stored to ensure privacy and confidentiality in accordance with the University of Manchester data protection guidelines. We initially sent a participant information sheet to each potential participant via digital communication tools informing them of the study's purpose and why they had been selected. Participation was voluntary, and no compensation was offered or required.

Results

Participant Characteristics

A total of 14 semistructured, in-depth interviews were carried out, with sessions having a duration ranging from 20 to 80 minutes. The mean duration of the interviews was 32.0 (SD 20.8) minutes. The variation in interview durations can be ascribed to the participants' varying levels of expertise and knowledge and the degree of their engagement with the smart bracelet initiative. Individuals that had greater direct engagement with the initiative generally offered more comprehensive perspectives, leading to lengthier interviews. In contrast, individuals with limited direct experience or engagement generally had shorter interviews.

The interview participants in our study encompassed a wide array of backgrounds, which included individuals involved in hajj, specialists in public health, and professionals from several disciplines (Table 1). A subset of the individuals involved in the study held prominent positions within governmental entities, whereas others held positions with varied levels of decision-making power.

Table 1. Study participants' expertise and sectors.

Participant number	Expertise	Sector
1	Systems engineering, business, and project management	Hajj and *umrah* services
2	Health care provider and public health specialist	Health care and public health
3	Health care provider and public health specialist	Health care and public health
4	Communication engineering and artificial intelligence	Technology or consulting (business)
5	Service provider	Accommodation
6	Service provider	Accommodation
7	Service provider	Accommodation
8	Service provider	Accommodation
9	Service provider	Guidance to holy sites
10	Service provider	Enrichment (educational and awareness)
11	Service provider	Arrivals and departures
12	Service provider	Arrivals and departures
13	Service provider	Transportation
14	Service provider	Transportation

In total, 2 distinct groups were formed based on the results of the conducted interviews. The initial cohort consisted of 4 participants who were directly involved in the 2021 initiative and made significant contributions to this study. The second cohort consisted of 10 persons who were not directly involved in the 2021 initiative and had no additional insights to contribute regarding the use of wearable devices during hajj. Nevertheless, these individuals were involved in the provision and coordination of hajj services as well as direct engagement with pilgrims.

Each participant had a minimum of 10 years of professional experience in their respective fields. The individuals who made these contributions were located in Saudi Arabia. A total of 71% (10/14) of the interviews were conducted via telephone, and the remaining 29% (4/14) were conducted using the Zoom platform (Zoom Video Communications). Audio recordings were gathered for all interviews.

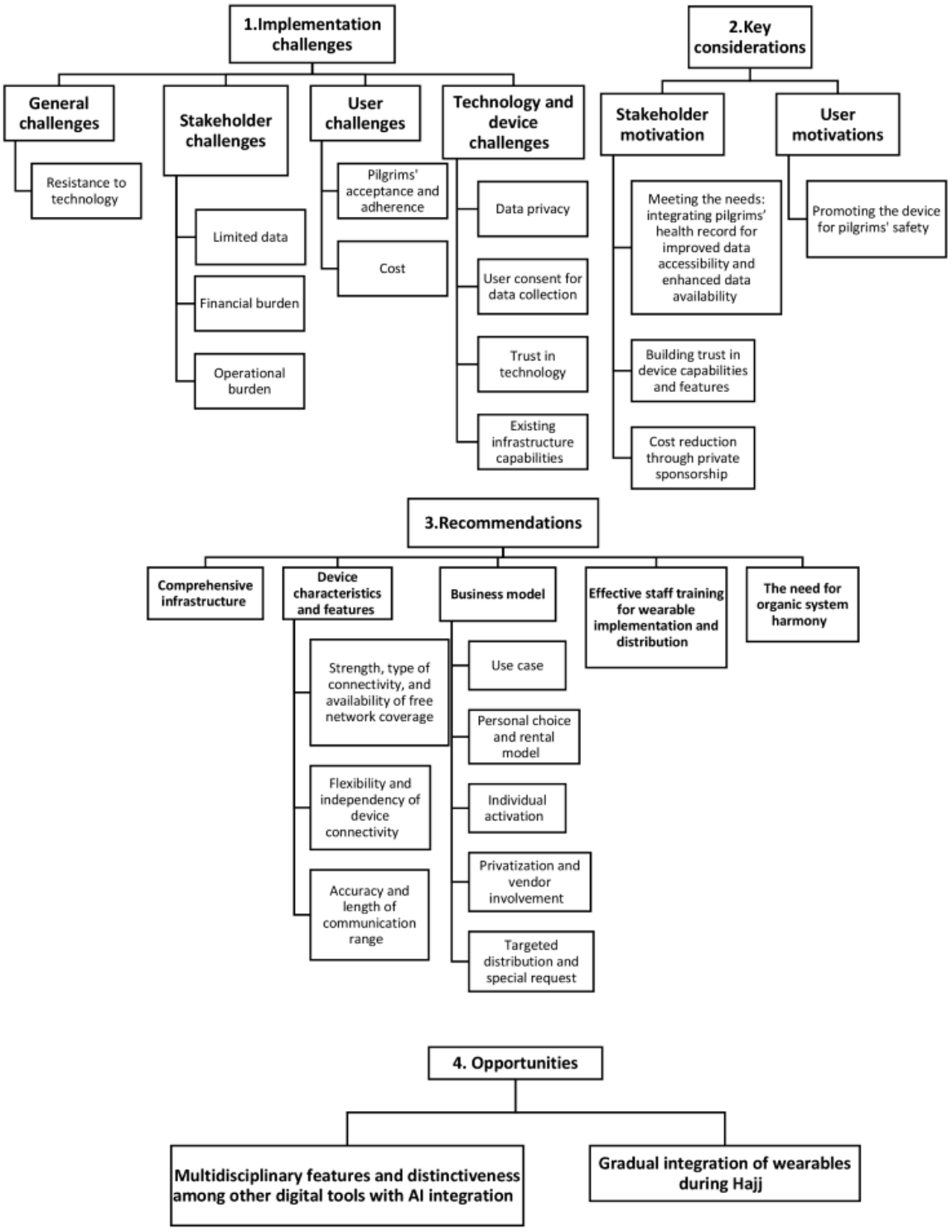
XSL•FO
RenderX

Generated Themes: An Overview

The analysis generated insights on 4 main themes and 13 subthemes (Figure 2), highlighting crucial challenges, considerations, recommendations, and opportunities in the use of wearable sensor bracelets for the presymptomatic detection of infectious diseases during hajj. A narrative synthesis was used to report the findings obtained from data collected from the first group of 29% (4/14) of the participants who played a role in the 2021 smart bracelet initiative. Table 2 highlights inputs from each of the 4 participants related to the identified themes.

For the second group, a brief summary of the findings is presented highlighting important observations without a detailed analysis.

Figure 2. Thematic analysis of the conceptual theme framework of the interviews. AI: artificial intelligence.



XSL•FO
RenderX

Table 2. Themes generated during data collection.

Theme and subthemes	Participant			
	1	2	3	4
Implementation challenges				
General challenges				
Resistance to technology		✓		
Stakeholder challenges				
Limited data		✓		
Financial burden			✓	✓
Operational burden	✓			
User challenges				
Pilgrims' acceptance and adherence		✓		
Cost				✓
Technology and device challenges				
Data privacy	✓	✓		✓
User consent for data collection	✓			✓
Trust in technology			✓	
Existing infrastructure capabilities			✓	
Key considerations				
Stakeholder motivations				
Meeting needs		✓		✓
Building trust in device features and capabilities	✓	✓		
Cost reduction through private sponsorship	✓	✓	✓	
Interest of entities involved	✓			✓
User motivations				
Promoting the device for pilgrims' safety	✓	✓		
Recommendations				
Comprehensive infrastructure			✓	✓
Device characteristics and features				✓
Business model	✓	✓		✓
Effective staff training for wearable implementation and distribution			✓	
The need for organic system harmony			✓	
Opportunities				
Multidisciplinary features and distinctiveness among other digital tools with AI[a] integration		✓		✓
Gradual integration of wearables during hajj		✓		

[a]AI: artificial intelligence.

The First Group's Perceptions of Using Wearable Sensor Bracelets

Overview

The idea of "smart hajj" was a recurring topic in the interviews as respondents emphasized their goals for the future and their current capabilities and constraints. The concept of smart hajj signifies the continuous endeavors of the hajj organization to adapt to technological advancements and improve the overall pilgrimage journey. The concept comprises various technical solutions designed to tackle difficulties and enhance the welfare and security of pilgrims.

However, the deployment of smart hajj and incorporating wearable devices into the technology used by hajj authorities present challenges and potential. Several important points were raised, and suggestions were made by the participants.

XSL•FO
RenderX

Implementation Challenges of Wearable Sensor Bracelets for Infectious Disease Detection During Hajj

In the context of using digital health technologies for public health surveillance, participants in our study highlighted a range of challenges in deploying wearable sensor bracelets for disease detection during hajj. The issues encompassed a range of areas, including general challenges, challenges connected to implementing stakeholders, challenges linked to users, and challenges related to devices and technology.

General Challenges: Resistance to Technology

Participants voiced apprehensions over the wider difficulties linked to integrating technology such as wearables within the public health domain. They emphasized a vital concern related to resistance from both users and implementing stakeholders due to factors such as data privacy, information security, and human rights:

> *We will face a challenge in using technology for public health due to data, information, and human rights. This challenge is present, and we will face resistance.* [Participant 2; health care provider and public health specialist]

Stakeholder Challenges: Limited Data, Financial Burden, and Operational Burden

Participants cited concerns regarding various factors from the hajj implementing stakeholders' perspective, including limited data availability, resistance to adoption, and the operational and financial burdens associated with wearable implementation. In addition, the conversation explored several aspects pertaining to the previous smart bracelet initiative, elucidating significant barriers that hindered its advancement.

The limited availability of health data can be attributed to the apprehensions of pilgrims regarding disclosing their medical records to hajj authorities. There was a concern among individuals that the act of revealing specific information pertaining to their health status and underlying medical disorders would potentially impede their ability to be granted permission to partake in the hajj pilgrimage or, alternatively, affect their eligibility to obtain hajj visas:

> *Pilgrims worry that if they mention any medical condition, that the Ministry of Hajj and Umrah will not issue them the visa. For example, they do not mention that they have Cancer fearing that they will not get issued the visa. I do not mind if the pilgrim has Cancer, I will treat you, but I need to know in advance to prepare and provide you with the service.* [Participant 2; health care provider and public health specialist]

One major issue brought up by participants, especially the hajj organizers, was the expense of integrating wearables throughout the pilgrimage. Considering the wide range of services they already offer, they voiced concerns regarding the cost load related to the adoption of wearables. Participants attributed the failure and incomplete implementation of the previous smart bracelet initiative (NUSUK) to the high cost of wearables and the significant budget needed for their deployment. Hajj organizers questioned whether they could cover the expense of integrating wearables for all pilgrims:

> *Let's return to the smart bracelet project and why it was not completed. Finance was an issue at that stage, and the projects couldn't proceed. The cost of these devices is considered high.* [Participant 4; communication engineering and artificial intelligence (AI)]

> *Can we afford to provide the wearable to all pilgrims?* [Participant 3; health care provider and public health specialist]

Participants highlighted the operational challenges associated with implementing ubiquitous sensor bracelets, including the need for staff training and the availability of expert teams for data analysis and decision-making. Concerns were also expressed over the substantial data flow generated by wearables. Despite the complete readiness of the expert teams, there were concerns that these efforts might not produce valuable outcomes owing to a lack of user engagement:

> *It will be a huge operational load.It needs a specialised team that deals with the data collected by the watch and decides whether the person needs a team to be dispatched. So, there was a big operational load with a minimal value, especially during Hajj.* [Participant 1; systems engineering, business, and project management]

User Challenges

Pilgrims' Acceptance and Adherence

Participants underlined the difficulty of obtaining user acceptance, particularly in light of the wide range of backgrounds, beliefs, and degrees of faith in technology exhibited by pilgrims hailing from >180 nations worldwide. Stakeholders observed that the implementation of the tracking and tracing feature may potentially discourage certain pilgrims from adopting or adhering to the technology:

> *Once you mention to the user that they can't take the bracelet off their wrist, or you can't give it to someone else, cut it, or I will notice what happens to the bracelet once you have done any of these. It will be like you are tracking me as a prisoner, and there will be high resistance.* [Participant 2; health care provider and public health specialist]

Cost

In addition to being a financial burden for the providers, wearable device costs may be seen as an obstacle by users if they are part of their hajj package. Participants voiced apprehension regarding the potential financial burden of including the cost of the bracelet in the pilgrims' bundles, deeming it to be costly:

> *If we estimate that the cost of the bracelet will be 100 Saudi Riyal, it will be considered expensive for pilgrims to add that amount to their package.* [Participant 4; communication engineering and AI]

XSL•FO
RenderX

Technology and Device Challenges: Data Privacy and Users' Consent, Trust in Technology, and Existing Infrastructure Capabilities

The issue of data privacy is intricate and delicate, particularly in relation to the acquisition of user data through diverse technological means. Privacy considerations are important in using wearable sensor bracelets for public health surveillance during events such as hajj. This is primarily due to the acquired data's sensitive nature and the wide range of users participating in these activities. When individuals' data pertaining to their health and movements are collected, it is natural for them to express concerns regarding the entities that possess access to these data and the potential applications they may be subjected to:

> *Well, the issue of wearables usually and all digital solutions are human rights.* [Participant 2; health care provider and public health specialist]

The interviewees emphasized the difficulties faced while implementing the hajj smart bracelet initiative (NUSUK), specifically focusing on data protection concerns. They underscored the significance of acquiring informed consent for data collection by technological means, such as health data and location information:

> *Regarding data privacy, no data can be collected without an individual consent. Tracking location and gathering medical or health data as you know are considered private data which fall under the General Data Protection Regulation (GDPR).* [Participant 4; communication engineering and AI]

Participants expressed certain reservations about the usefulness of wearable technology. One of the difficulties that participants expressed concerns about was the efficacy of wearable devices. This substantially enhances individuals' confidence in the deployment of wearable sensor bracelets for disease detection during the hajj pilgrimage. In alternative terms, participants emphasized the necessity for these technologies to possess a high level of efficacy to foster a sense of assurance in their use:

> *There are lots of challenges. First, the technology part. How effective is it, how does it capture data, and is it durable? Is it accurate? I think we still can't say that the wearables are ready from the technological side.* [Participant 3; health care provider and public health specialist]

Despite the notable advancements in the network infrastructure built in the Kingdom of Saudi Arabia, participants expressed concerns regarding the adequacy of connectivity strength necessary to effectively manage the vast number of devices used during hajj and the amount of data transmitted:

> *What if we have 2 million devices in an area with weak connections? Do you think we can manage it? Still early.* [Participant 3; health care provider and public health specialist]

Key Considerations in the Adoption of Wearable Sensor Bracelets for Disease Detection During Hajj

Overview

The motives of both consumers and providers significantly impact the successful integration of technology in any domain. In implementing wearable sensor bracelets during the hajj pilgrimage, it is imperative to thoroughly examine and comprehend the motives of all stakeholders and users concerning their acceptance and adoption of this technology. Participants identified several factors that should be considered to encourage stakeholders.

Stakeholder Motivation

Meeting Needs: Integrating Pilgrims' Health Records for Improved Data Accessibility and Enhanced Data Availability

Stakeholders are primarily driven to integrate technology into their strategic plans when it caters to their distinct requirements and eliminates the obstacles they have experienced. Furthermore, stakeholders tend to adopt a technology when it effectively mitigates current challenges and when the characteristics and capabilities of the device correspond to their specific needs. The motivation behind this behavior arises from a comprehensive understanding of how the technology may optimize their activities and contribute to their overarching goals.

The availability of health data, especially with the heterogeneity of languages of pilgrims attending hajj, is of tremendous value to health care professionals as it helps them determine the individual requirements of participants (hajj attendees) and make required preparations. Implementing a consolidated health data system would significantly streamline the identification and understanding of the medical background and requirements of individuals regardless of the health care facilities they visit within the holy sites and Makkah hospitals. Integrating systems would facilitate uninterrupted access to participants' records and optimize the delivery of critical health care services:

> *Ministry of Health needs to have a complete EHR for the pilgrim, they would prefer to have reliable and accurate health data, so data availability.* [Participant 4; communication engineering and AI]

> *Today, I need basic information. Wearables for public health are different from normal bracelets. For the normal bracelet, I need major clinical data. For example, I need to know in advance the conditions that the patient (pilgrim) has, such as if he has kidney failure, if she is pregnant and so on, so I get ready to serve them. I do not want to be surprised; I want to be ready.* [Participant 2; health care provider and public health specialist]

Building Trust in Device Capabilities and Features

From the perspective of health care service providers, participants emphasized the key features that must be incorporated into wearable sensor bracelets to instill faith in and reliance on these devices, hence minimizing the necessity for additional evaluation. The accuracy of the data collected by these devices is a crucial factor. Although health care providers

XSL•FO
RenderX

may choose to perform supplementary evaluations such as laboratory tests, it is imperative that the wearable device delivers data of utmost precision. The significance of accuracy cannot be overstated as it guarantees the reliability of the obtained information and establishes a robust basis for further evaluations:

> *If you look at Apple Fit watch or any other Fitbit watch, the heart rate increases if a person makes any physical effort. Sometimes even if you are sitting, your heart rate could be irregular.* [Participant 1; systems engineering, business, and project management]

In addition to recognizing the significance of trust and the attributes of the wearable sensor bracelet, participants underscored their preference for specific aspects of wearable technologies. A further motive emphasized by the study participants pertained to a particular attribute of the device, namely, its capacity to both transmit and receive data. Participants emphasized the significance of bidirectional data connection as it enables service providers to gather data from the wearables and monitor the bracelets themselves. This attribute was perceived as highly advantageous in terms of facilitating efficient surveillance and guaranteeing the integrity of the sensor bracelets, hence preventing any unauthorized tampering or substitution:

> *I am not a technology person, but I imagine that if we are using such bracelet, I would prefer that it send me data, and I could pull data from it as well. So, push and pull material for data. In addition, the activation and deactivation are central. The user shouldn't be able to activate or deactivate something. I need to know that someone has cut it, removed it, given to someone else or exchanged, so I can track people.* [Participant 2; health care provider and public health specialist]

Cost Reduction Through Private Sponsorship

Considering the cost of the bracelet is of utmost importance as participants emphasized the need to minimize the financial load on hajj authorities. The consensus among participants was that developing or deploying such a device within the context of hajj must strike a delicate equilibrium between cost-effectiveness and the essential functionalities it provides:

> *Honestly, the operation model for this whole business should be provided by the private sector, not the government.* [Participant 1; systems engineering, business, and project management]

Interest of Entities Involved

A significant aspect highlighted by participants was the identification of the interests of the entities involved in the project and assessing the value derived from integrating the technology into their road map development. Among the many organizations participating in the hajj season, such as the Ministry of Health, Ministry of Hajj and Umrah, Ministry of Interior, and others, a number of duties and responsibilities were highlighted:

> *The Ministry of Hajj and Umrah focuses on the quality of services, which is why we focus on measuring satisfaction with crowd management, movements, and safety. We do not focus on diseases and infectious diseases because it is different from the scope of work we are looking for. We are not wasting our effort and time on technologies of that interest. It is not our specialist as a ministry, but the Ministry of Health are working on this file.* [Participant 1; systems engineering, business, and project management]

User Motivation: Promoting the Device for Pilgrim Safety

The significance of encouraging pilgrims to adopt wearable sensor bracelets actively was emphasized by the participants. They stressed the importance of communicating a message that revolves around the person's interest in and reasons for using the technology, especially in light of their safety. It was widely agreed upon that presenting the use of wearables during hajj as a matter of personal preference closely linked to the individual's welfare enhanced the probability of acceptance among pilgrims. The participants expressed reservations about adopting methods that could be interpreted as invasive or limiting, emphasizing the importance of presenting wearables as instruments for individual safety and autonomy:

> *I think you can work it out with the Ministry of Hajj and Umrah, and the Hajj missions by agreeing that the wearable's objective is user health and user safety. They should propose it from the health and safety aspect, so people accept wearing it.* [Participant 2; health care provider and public health specialist]]

Recommendations for Wearable Sensor Bracelets for Disease Detection During Hajj

Interviewees underlined and provided valuable insights into many prerequisites that must be considered to successfully and efficiently implement wearable sensor bracelets for disease detection during hajj.

Comprehensive Infrastructure

The importance of establishing a resilient and all-encompassing infrastructure to effectively handle the large number of devices deployed in the vicinity of the holy sites of Makkah and enable seamless data transmission was emphasized as a crucial component. Moreover, the emphasis was placed on the many components and characteristics of the devices that could efficiently carry out the functions of disease detection and tracking throughout the hard period of hajj considering the large number of participants:

> *An important variable in the success of smart cities is first the infrastructure.* [Participant 4; communication engineering and AI]

Device Characteristics and Features

Strength, Type, and Independency of Connectivity

The current state of connection in holy places was underscored as being in its nascent phase notwithstanding the progress made

XSL•FO
RenderX

in 4G and 5G technologies and the deployment of IoT infrastructure:

> *In previous years, the infrastructure of internet connectivity for 4G and 5G improved remarkably. Connectivity needs to be strong around all holy sites and areas. Not all areas were 100% covered, some were covered under a certain company, and others were served under another coverage company. In addition, the large number of people in a certain area using that network could weaken the connectivity.* [Participant 4; communication engineering and AI]

A cost-effective communication infrastructure was suggested as a means to facilitate data transmission for devices and guarantee equitable service for the substantial number of pilgrims. The proposal entails offering complimentary access to this form of connectivity while considering the economic strain on both the organizers and the users:

> *There is the cost of the device itself and the cost of communication. What I mean by communication is your network package or availability of Wi-Fi services.* [Participant 4; communication engineering and AI]

Emphasizing the necessity for data transmission autonomy, it was critical that wearable sensor bracelets possess the capability to transmit data autonomously without dependence on an intermediary device such as a mobile phone. Autonomy is attained by the use of different forms of connectivity, such as ultrawide band:

> *We usually recommend that when using a smart watch not to connect it to a mobile phone, rather to a sim card or to any type of connectivity such as the UWB where data can be sent whenever it gets signal form a sensor, a device, or any connectivity.* [Participant 4; communication engineering and AI]

Accuracy and Length of Communication Range

Furthermore, the strong reliability of Bluetooth was highlighted, specifically in the context of short-range communication. Nevertheless, it was recognized that Bluetooth possesses certain restrictions, necessitating specific intervals for optimal functionality when operating over particular distances:

> *The advantage of Bluetooth is the strong accuracy. However, one of the limitations is that every 50-100 m you need an interval. The UWB you can reach up to 1000 meters, so less sensors.* [Participant 4; communication engineering and AI]

Apart from the various characteristics of the wearable sensor bracelet that were previously discussed, with a focus on accuracy and efficiency in wearable devices for disease detection during hajj, technologists emphasized the importance of battery life as a critical factor to be taken into account. The importance of selecting an appropriate connectivity option that aligns with the specific circumstances was underscored as it helps reduce excessive battery use and enables consistent and efficient use of these devices within the challenging pilgrimage setting:

> *Battery life as I mentioned is a factor you need to consider. How many days would you like the bracelet to last for. Nowadays companies manufacture devices based on your need, do you want it to last a day more, a week or what.* [Participant 4; communication engineering and AI]

Business Model: Selection of an Appropriate Use Case (a Methodology Used in System Analysis to Identify, Clarify, and Organize System Requirements)

Identifying the right and appropriate use case for the technology was highlighted based on user interest and needs and cost to facilitate the successful integration of new technology into the hajj system:

> *We need to know the appropriate and right use cases for technology usage and implementation. It needs to be clear.* [Participant 4; communication engineering and AI]

Participants emphasized the significance of rendering the use of wearables a personal decision, striking a balance between the device's advantages and the costs involved. The rental model, individual activation, and targeted distribution were identified as crucial elements in attaining a more customized and economically efficient strategy for deploying wearables during the hajj pilgrimage:

> *A person could personally benefit from this wearable if he/she feels they have a critical condition and needs this service; it can be requested as a special request. Anyone can activate it by connecting the watch with contact numbers in an emergency. The government does not need to provide this service, especially since the watch cost is very high. We are talking about 1000 Saudi Riyal [US $266.60] for one watch. Honestly, the value for the cost was not achieved in the project.* [Participant 1; systems engineering, business, and project management]

Effective Staff Training for Wearable Implementation and Distribution

Staff training was identified as a crucial component for successfully implementing wearables. The tasks encompassed managing a substantial volume of data and guaranteeing precise allocation of the wearable devices. The hajj event yields a substantial amount of data, necessitating the provision of sufficient training to staff to handle, analyze, and understand this information effectively:

> *Another thing we need to think about. When I receive a large amount of data, and I am not ready for it, how can I deal with it? Last Hajj, we suddenly received a large amount of data from other departments it was very detailed/360, and we had no idea how to deal with them. We didn't expect it.* [Participant 3; health care provider and public health specialist]

> *Another issue we need to consider is how we will guarantee the distribution of the device. You know the food itself; we have an issue with it with the campaigns. Think about this scenario. I am receiving pilgrims from various countries, and I have the*

XSL•FO
RenderX

> *bracelet to give to the campaign manager or staff that were not trained enough to match the bracelet with the right person. I can't guarantee this; most likely, there will be a mix-up.* [Participant 3; health care provider and public health specialist]

The Need for Organic System Harmony

Participants highlighted the significance of achieving a smooth and natural incorporation of wearable devices into the system to facilitate prospective expansion. This sentiment agrees with the perspective expressed by a participant who emphasized the importance of adopting an organic approach and considering integration as a complete system. The importance of careful deliberation was emphasized by the participant, stressing the need for alignment between decision makers and persons who wear the bracelets. The primary focus lies in establishing a cohesive, interrelated framework that promotes user-friendly functionality, comprehension, and expansion:

> *These things need to be organic; it is a system. We need to think about the person taking the decision, the person wearing the bracelet. Everything need to be harmonised and work together. We all need to know how to deal with it and there need to be built on organic growth.* [Participant 3; health care provider and public health specialist]

Opportunities and Advantages of Wearable Sensor Bracelets During Hajj

Multidisciplinary Features and Distinctiveness Among Other Digital Tools With AI Integration

Participants provided insights into the possible benefits of incorporating wearables, highlighting the numerous opportunities that arise from the varied range of features offered by these devices. They emphasized the potential for transformation that is enabled by improvements in AI capabilities, which play a vital role in developing and improving these devices. The perspectives shared by the participants shed light on a favorable view on the distinctive attributes of wearable devices and the potential synergies they offer with modern AI technologies. These observations indicate a wide range of opportunities for further advancements in this field:

> *Yes, a Wearable sensor bracelet is a good device. Can we use it? The answer is yes. AI will take a very huge platform in public health.* [Participant 2; health care provider and public health specialist]

Participants underscored the unique features of wearable sensing devices compared to other gadgets such as mobile phones. The participants emphasized that wearable devices can assess a wider array of physiological indicators in contrast to mobile devices. This distinctive capacity of wearable sensors makes them valuable instruments for gathering a complete array of health-related data, surpassing typical mobile devices' capabilities:

> *So, both devices; smart watches and mobile phone can use Bluetooth to track and trace infected individuals and people who have been in contact with. However, Smart watches have an advance of measuring vital signs.* [Participant 4; communication engineering and AI]

Gradual Integration of Wearables During Hajj

Participants emphasized the importance of gradually integrating wearable devices as public health tools for effective implementation into the hajj system:

> *...of course. Wearables started to get large space in healthcare. Yes, we can use wearables, but I wouldn't use them for public health at first. It could start with healthcare tackers, so I use the wearable, which can give signals to the healthcare tacker first. Each group of pilgrims have their own identifier (healthcare tacker) or medical mission, so it gives a warning to the medical mission, which gives him a summary of number of cases for example 4 cases have X without a diagnosis just based on the wearable. Then, as a public health department, I get notified and decide what actions need to be taken, such as checking A, B, & C to ensure it is X or Y.* [Participant 2; health care provider and public health specialist]

A Brief Summary of the Findings From the Second Group

The second group of participants cited their duty to serve the guests of God during their hajj journey to Makkah. They discussed the many types of *tawafa* organizations they belonged to (a service provider for external pilgrims) and the role of each one. *Tawafa* organizations are classified into 4 distinct groups. The first group consists of 6 organizations responsible for assisting pilgrims in Makkah and the holy sites that are segregated into subgroups according to the countries and continents from which they originate. These six organizations are for pilgrims coming from (1) Iran; (2) South Asia; (3) Turkey, Europe, Australia, and America; (4) Arabic countries; (5) African non-Arabic countries; and (6) Southeast Asia. The second group consists of 1 organization responsible for providing water to pilgrims at the Sacred House. The third group consists of an organization that is licensed to receive and dispatch pilgrims arriving via air, sea, and land ports, facilitating and arranging their departure protocols. The final group is an organization that is licensed to provide services to pilgrims visiting the sacred mosque located in Madinah city.

Several prevalent health concerns among the pilgrims they assist were brought up by the participants. The most common health problems associated with the hajj event include heat-related illnesses such as heat stroke, heart diseases, and infectious diseases. The importance of medical missions associated with each group of pilgrims was also underscored.

Participants commented on a common issue related to pilgrims' lack of adherence to specific preventive measures designed by hajj authorities to safeguard their well-being, such as hygiene and stampedes in Jamaraat Bridge. In addition, they highlighted that the most holy site among Mina, Muzdalifah, Mount Arafat, and the Holy Mosque where the crowd density becomes high is Mina.

XSL•FO
RenderX

Participants discussed the technological improvements introduced by hajj authorities, such as mobile apps and smart cards, which provide assistance and information to pilgrims from diverse backgrounds. Nevertheless, the participants also emphasized the difficulties of integrating these technologies on a broad scale during hajj. They stressed the importance of conducting more educational campaigns to assist pilgrims in comprehending and proficiently using these technologies as well as understanding the reasons behind their deployment.

Discussion

Principal Findings

Overview

In this study, we explored participant perceptions and experiences of using wearable sensor bracelets during the hajj pilgrimage. The primary objective was to investigate the potential advantages and obstacles associated with wearables monitoring pilgrims' well-being and detecting diseases. The main findings and themes that emerged from the interviews will be analyzed and summarized in this section to provide insights into the opportunities and challenges involved in incorporating wearable technology (sensor bracelets) into mass gatherings, including the hajj pilgrimage.

Wearable sensor bracelets have the capability to monitor the health of pilgrims and detect emerging infectious diseases during the hajj pilgrimage. However, achieving these capacities depends on properly tackling existing problems and meticulously evaluating multiple elements.

The diverse range of valuable information obtained from the interviews with participants reveals a variety of obstacles that hinder the seamless use of wearable sensor bracelets for detecting infectious diseases during the hajj pilgrimage. However, in the midst of these difficulties, a variety of factors for achieving successful implementation become apparent, uncovering hidden possibilities that these devices offer. Charting a path for future deployments in the dynamic landscape of mass gatherings such as hajj, we will unravel the multifaceted nature of these challenges, explore the promising vistas these wearables offer, and present recommendations highlighted by our interviewees' insightful perspectives as we navigate the discussion.

Challenges

Numerous challenges associated with implementing wearable sensor bracelets during hajj were highlighted. The primary obstacle underlined was the anticipated opposition to technology, which may potentially emerge from providers and users of wearable devices. Participants observed that trust in technology, data privacy, information security, and human rights concerns could influence this reluctance.

The opposition of providers is affected by both the operational and financial challenges associated with implementing wearable technology into the hajj system. Although data are an essential element in detecting diseases, staff training in data collection and analysis is required to handle the anticipated volume of data produced by the devices. Participants voiced reservations about the limited and inaccurate pre-hajj data on pilgrims' health. Furthermore, there was concern that, despite the dedicated efforts put into the implementation, the results may not correspond with the anticipated or intended value.

The issue of financial burden is not limited to providers; it affects pilgrims as well. Due to the financial obligations that pilgrims encounter during their hajj journey [36], the cost could constitute a barrier if the services were to be included in their hajj packages. The cost-related reluctance may affect users' willingness to use the device during their hajj pilgrimage. The use of tracking and tracing methods to monitor pilgrims throughout their journey may impede their compliance and reception due to variations in their backgrounds, beliefs, ages, approaches to conducting hajj rituals, level of trust in technology, and literacy levels [3,37].

The challenges brought up by stakeholders are consistent with those identified by Yeung et al [3], especially regarding the end users in our research, who are pilgrims. The stakeholders' worries about potential hurdles from users align with the findings of Yeung et al [3], suggesting that user resistance may stem from factors such as privacy concerns, lack of motivation, and technological fear.

The viewpoints conveyed by stakeholders and the difficulties they mentioned pertain to service providers and users and align with the concept of effort expectancy in the UTAUT model. This connection is particularly visible in the financial and operational efforts required to integrate wearables during hajj.

Data privacy is paramount while collecting user data through wearable devices in any domain. Collecting real-time data is essential for public health surveillance to diagnose infectious diseases promptly and implement appropriate interventions. Gathering sensitive data poses difficulties, especially when considering human rights and data privacy agreements. The use of wearable public health devices, particularly in mass gatherings, may be limited due to considerable challenges arising from privacy considerations.

The issue of data privacy, which is of utmost importance for wearable devices in the context of mass gatherings, is intricately linked to the difficulty of amassing data with user consent. Participants highlighted that addressing privacy concerns and securing user consent presented substantial obstacles, impeding the success of the previous smart bracelet initiative. Azodo et al [6], who discussed information privacy, data sharing, and autonomy, brought attention to this layer of complication brought about by using wearables in health care.

The stakeholders' willingness to adopt wearable sensor bracelets for disease detection and tracking during hajj was strongly linked to the idea of performance expectancy in the UTAUT model. The stakeholders' trust in the technology was greatly impacted by the device's performance and its capacity to meet fundamental functions. This is consistent with the concept of performance expectancy, which highlights the perceived utility of the technology. Furthermore, assessing the device's utility, including factors such as dependability, precision, and compatibility, highlights the significance of technology characteristics in the TTF model. These functionalities play a

XSL•FO
RenderX

vital role in shaping stakeholders' perceptions and adoption of the technology within the specific context of hajj.

Considerations for Promoting Acceptance and Adoption

Overcoming current obstacles that hinder the adoption and use of wearable sensor bracelets during hajj is vital. Participants emphasized important aspects that play a significant role in promoting the success of the implementation.

The interviews demonstrated that stakeholders believe that user motivation is crucial in shaping their acceptance of the device. The interviews covered several aspects that contribute to promoting acceptance. Improving data availability to enhance accessibility is a primary driver for addressing stakeholders' needs. Furthermore, considering the interests of the parties involved in managing hajj ensures the incorporation of technology into their plans for future development in line with their particular requirements and strategies.

The wearables' functionality and capacity to collect accurate, dependable, and real-time data were given thoughtful consideration, addressing issues brought up in the literature on the difficulty of guaranteeing wearable accuracy [11,27,38]. Increasing stakeholder trust in technology becomes increasingly challenging when it comes to the known difficulties in ensuring wearable accuracy. The need to ensure essential device functions was stressed by participants as a means of promoting wearable trust and dependence. These qualities are essential for overcoming wearable accuracy difficulties, building stakeholder trust, and influencing their choice to use technology for disease detection and hajj surveillance.

The emphasis on accuracy and different device characteristics is consistent with the concept of performance expectancy in the UTAUT. Stakeholders emphasized the necessity for features that qualify the device for successful disease diagnosis and saw performance as critical to the technology. In addition to being in line with performance expectations, addressing issues with accuracy and key features is crucial for shaping stakeholders' opinions of the technology.

There was a unanimous consensus on the necessity of incorporating private sponsorship as a business model for providing the device. By leveraging public-private partnerships, Aina et al [24] propose resolving the obstacles linked to smartification and digitalization in urban service financing. As suggested by stakeholders, this strategy reduces the financial and operational burden on hajj stakeholders related to implementing and controlling wearable devices for disease diagnosis and tracking. The fact of private sponsorship coincides with the UTAUT's *facilitating condition* construct, offering a favorable setting and support for the effective integration of wearable technology during hajj.

During the interviews, participants emphasized safety as a crucial factor that significantly affects the adherence to and acceptance, adoption, and use of wearable sensor bracelets by pilgrims. Having sufficient user motivation and willingness to acquire and use the wearable device is essential for acceptance [39]. Although Saudi Arabia has made significant efforts to guarantee the safety of pilgrims and control crowds to minimize the occurrence of stampedes and the spread of diseases, the difficulty remains [40].

It is widely agreed that presenting the use of wearables during hajj as a personal choice strongly linked to individual well-being greatly increases the chances of pilgrims accepting it. This is consistent with the participants' emphasis on the significance of selecting the appropriate business model for implementing such devices. Participants also voiced apprehensions over techniques that can be interpreted as intrusive or constraining, emphasizing the significance of portraying wearables as instruments for personal safety and independence.

Recommendations

Several crucial recommendations were proposed to tackle the limitations of incorporating wearable sensor bracelets for disease detection during hajj. A robust IoT infrastructure emerged as a crucial element in successfully deploying the device. The type of network connectivity and its strength were highlighted as key considerations, emphasizing the capacity to handle the volume of users and data transmission coupled with the availability of free connectivity for all pilgrims.

Participants emphasized the importance of implementing wearables for public health surveillance in a gradual and systematic manner. The gradual integration focuses on and enables smooth integration by initially guaranteeing the compatibility of the technology with the surveillance system used during hajj. Compatibility is a vital element within the TTF model that describes the features of the technology.

Participants with technology experience provided valuable comments that highlighted the need to address data availability difficulties faced by hajj stakeholders. They emphasized the importance of establishing a strong, accurate health data system to resolve these challenges. The proposed solution entails using blockchain technology to establish customized electronic health records specifically designed for pilgrims. This approach is in keeping with the recommendations presented in the research by Yeung et al [3], emphasizing the capacity of blockchain technology to tackle issues concerning the secure and decentralized storage and transmission of digital data in health care infrastructure.

Participants put forth a distinctive suggestion concerning carefully considering and selecting a suitable business model to ensure the seamless integration and use of wearable sensors during the hajj ritual. Walsh [35] states that smartwatch applications and use cases differ in many fields, offering various services and functions. The wide range of differences among pilgrims during hajj and *umrah* offers a chance to create applications that can improve their overall experience. A novel proposal entails implementing a rental system for individuals who wish to use such devices during their hajj pilgrimage either for overall well-being or for specific health concerns.

Opportunities

Implementing wearable sensor bracelets for infectious disease detection during hajj comes with difficulty, yet the opportunities for this technology's use are considerable.

XSL•FO
RenderX

The use of AI greatly enhances the potential of wearables in public health surveillance, representing a promising advancement. Wearable devices, which include communication modules and networking capabilities, possess modern hardware technologies. These devices can supply vital data that power AI procedures [41]. This technological advancement facilitates the uninterrupted transmission of up-to-date information without requiring direct engagement from health care staff, thereby establishing wearables as a potent and nonintrusive instrument for augmenting public health surveillance.

The capacity of wearable sensor bracelets to evaluate a wider range of physiological indications gives them a clear edge over other technologies such as mobile devices. Their ability to adapt and be used in various ways makes them very valuable in the health care industry.

Wearable sensor bracelets greatly enhance public health surveillance during hajj by carrying out functions such as disease identification, health monitoring, tracking, and tracing. They stand out from other devices because of their unique characteristics, which provide real-time health tracking, detection, and monitoring. This helps with the efficient administration of public health during hajj. This is consistent with the broader idea of the TTF model, in which wearables are customized to carry out certain duties and operations that are essential for effective public health management in the special circumstances of hajj.

The degree to which wearable sensor bracelets can fulfill the unique requirements of hajj stakeholders is one of the key elements determining the desire to use them for disease detection during hajj. The link between hajj stakeholders' desire to use wearables for disease detection and their perception of the devices' capabilities emphasizes the importance of performance expectancy as a critical element of the UTAUT, which affects providers' desire to embrace the technology. Furthermore, achieving successful implementation relies on facilitating conditions, underscoring the need for a supportive environment for efficient adoption.

Furthermore, the *technology characteristics* component, as defined by the TTF model, has a crucial influence on the intention to use. The characteristics and qualities of the technology have a substantial influence on providers' inclination to embrace it. The relationship between the UTAUT and TTF model highlights how users' intention to adopt and use wearable sensor bracelets for disease detection during hajj is influenced by their perception of the technology's performance, facilitating conditions, and characteristics.

Strengths, Limitations, and Future Work

Strengths

This study is the first to address the opportunities and limitations of wearable sensor bracelets for hajj. It provides a thorough understanding of these issues from the perspective of key stakeholders. The thoroughness of the investigation provides useful insights into the complex dynamics surrounding the implementation of these devices.

Moreover, the combined use of the UTAUT and TTF models proved advantageous in comprehending stakeholder viewpoints comprehensively. This method enabled a thorough investigation of the possible uses of wearable devices for detecting diseases, specifically during the hajj pilgrimage.

Limitations

One potential limitation is that the range of stakeholders from different disciplines that were included in the interviews was not very diverse. Broadening the pool of respondents would enhance our understanding of the system and the interconnectedness between its components.

In addition, a possible drawback arises from the involvement of a single researcher in the coding and analytic process. This limited worldview may lead to the potential for misjudging participants' perspectives. Adding numerous analysts or external assessments could have enhanced the coding process by incorporating varied perspectives and increasing its rigor. Considering the expansion of the analytical team in future studies may help limit any biases in interpreting participant replies.

Future Work

To optimize the efficient deployment of wearable sensor bracelets, further studies should prioritize obtaining perspectives from end users. Gaining insights into their perspectives and embracing their adoption of the device is essential for formulating tactics that align with user requirements and preferences. This would additionally enhance the successful integration and use of wearable sensor technology in pertinent contexts.

Conclusions

The significant contribution of this work lies in its thorough investigation of crucial factors in implementing wearable sensor bracelets for the detection of infectious diseases during the hajj event. Our study used the integration of the TTF and UTAUT frameworks to better understand the adoption of wearable sensors by stakeholders. The purpose of this integration was to investigate the perceived likelihood of technology adoption during hajj and the stakeholders' viewpoints on the compatibility between the task of disease detection and the technology used. Our aim was to integrate these theories to offer a thorough comprehension of the aspects that impact the acceptance and efficacy of wearable sensors in this particular setting. This study underscores the need to resolve crucial issues to guarantee the efficient implementation and use of wearable sensors in this specific case. Our results highlight multiple problems pertaining to suppliers, users, and technology, which impose a cost on hajj stakeholders. Significantly, the specific attributes and functionalities of the technology play a vital role in building confidence among providers when it comes to using wearables during hajj. This study highlights the importance of having the ability to access pre-hajj health data for health care providers to gain a more comprehensive understanding of pilgrims' requirements. Although there have been some breakthroughs, it is crucial to further develop the current infrastructure to support wearables fully. Wearables stand out from other devices due to their exceptional capacity to gather vital signs, and their

XSL•FO
RenderX

potential is further amplified by encouraging outcomes from AI algorithms. This study suggests choosing a suitable business model that is customized to meet the specific demands of the users to ensure a successful deployment. A crucial element of this implementation entails creating a complete infrastructure marked by a strong and reliable network capable of prompt data transfer and accommodating a large user population. Furthermore, the unique characteristics of the devices should be in line with cost-effective advantages to guarantee a sustainable implementation.

However, deploying wearable bracelets for detecting infectious diseases during hajj may face difficulties. Therefore, it is important to conduct pilot tests and thorough evaluations to ensure their usefulness before implementing them at a large scale. This prudent approach recognizes the intricacy of the hajj environment and highlights the importance of conducting evidence-based evaluations to determine the feasibility and effectiveness of the wearable technology. Notwithstanding these difficulties, there is a pervasive conviction regarding the innate capacity of these wearable devices, cultivating optimism regarding their ultimate and widespread use in the context of hajj.

Acknowledgments

This work was conducted as part of a doctoral study sponsored by the Ministry of Education, Saudi Arabia.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example topic guide for the interviews.
[DOCX File , 33 KB - formative_v9i1e60484_app1.docx]

References

1. Sahu KS, Majowicz SE, Dubin JA, Morita PP. NextGen public health surveillance and the internet of things (IoT). Front Public Health 2021 Dec 3;9:756675 [FREE Full text] [doi: 10.3389/fpubh.2021.756675] [Medline: 34926381]
2. Hirten RP, Danieletto M, Tomalin L, Choi KH, Zweig M, Golden E, et al. Use of physiological data from a wearable device to identify SARS-CoV-2 infection and symptoms and predict COVID-19 diagnosis: observational study. J Med Internet Res 2021 Feb 22;23(2):e26107 [FREE Full text] [doi: 10.2196/26107] [Medline: 33529156]
3. Yeung AW, Torkamani A, Butte AJ, Glicksberg BS, Schuller B, Rodriguez B, et al. The promise of digital healthcare technologies. Front Public Health 2023 Sep 26;11:1196596 [FREE Full text] [doi: 10.3389/fpubh.2023.1196596] [Medline: 37822534]
4. Hayati N, Ramli K, Suryanegara M, Salman M. An internet of things (IoT) reference model for an infectious disease active digital surveillance system. Int J Adv Comput Sci Appl 2021;12(9). [doi: 10.14569/ijacsa.2021.0120956]
5. Meraj M, Alvi SA, Quasim MT, Haidar SW. A critical review of detection and prediction of infectious disease using IOT sensors. In: Proceedings of the Second International Conference on Electronics and Sustainable Communication Systems. 2021 Presented at: ICESC 2021; August 04-06, 2021; Coimbatore, India. [doi: 10.1109/icesc51422.2021.9532992]
6. Azodo I, Williams R, Sheikh A, Cresswell K. Opportunities and challenges surrounding the use of data from wearable sensor devices in health care: qualitative interview study. J Med Internet Res 2020 Oct 22;22(10):e19542 [FREE Full text] [doi: 10.2196/19542] [Medline: 33090107]
7. Tavakoli M, Carriere J, Torabi A. Robotics, smart wearable technologies, and autonomous intelligent systems for healthcare during the COVID‐19 pandemic: an analysis of the state of the art and future vision. Adv Intell Syst 2020 May 05;2(7):2000071. [doi: 10.1002/aisy.202000071]
8. Mamdiwar SD, R A, Shakruwala Z, Chadha U, Srinivasan K, Chang CY. Recent advances on IoT-assisted wearable sensor systems for healthcare monitoring. Biosensors (Basel) 2021 Oct 04;11(10):372 [FREE Full text] [doi: 10.3390/bios11100372] [Medline: 34677328]
9. Mishra T, Wang M, Metwally AA, Bogu GK, Brooks AW, Bahmani A, et al. Pre-symptomatic detection of COVID-19 from smartwatch data. Nat Biomed Eng 2020 Dec 18;4(12):1208-1220 [FREE Full text] [doi: 10.1038/s41551-020-00640-6] [Medline: 33208926]
10. Hänsel K, Wilde N, Haddadi H, Alomainy A. Challenges with current wearable technology in monitoring health data and providing positive behavioural support. In: Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare. London, UK: ICST; 2015:158-161.

XSL•FO
RenderX

11. Kamišalić A, Fister IJ, Turkanović M, Karakatič S. Sensors and functionalities of non-invasive wrist-wearable devices: a review. Sensors (Basel) 2018 May 25;18(6):1714 [FREE Full text] [doi: 10.3390/s18061714] [Medline: 29799504]
12. Leu FY, Ko CY, You I, Choo KK, Ho CL. A smartphone-based wearable sensors for monitoring real-time physiological data. Comput Electr Eng 2018 Jan;65:376-392. [doi: 10.1016/j.compeleceng.2017.06.031]
13. Sawadi HK, Al-Zahrani IO, Al-Ghoraibi AF, Al-Ghamdi AM, Bati GF. Towards Hajj and Umrah digital transformation using smartwatches. Islamic Univ J Appl Sci 2023 Jul;V(I):91-107.
14. Mohammadzadeh N, Gholamzadeh M, Saeedi S, Rezayi S. The application of wearable smart sensors for monitoring the vital signs of patients in epidemics: a systematic literature review. J Ambient Intell Humaniz Comput 2023 Nov 13;14(5):6027-6041 [FREE Full text] [doi: 10.1007/s12652-020-02656-x] [Medline: 33224305]
15. Creaser AV, Hall J, Costa S, Bingham DD, Clemes SA. Exploring families' acceptance of wearable activity trackers: a mixed-methods study. Int J Environ Res Public Health 2022 Mar 15;19(6):3472 [FREE Full text] [doi: 10.3390/ijerph19063472] [Medline: 35329166]
16. Mitratza M, Goodale BM, Shagadatova A, Kovacevic V, van de Wijgert J, Brakenhoff TB, et al. The performance of wearable sensors in the detection of SARS-CoV-2 infection: a systematic review. Lancet Digit Health 2022 May;4(5):e370-e383 [FREE Full text] [doi: 10.1016/S2589-7500(22)00019-X] [Medline: 35461692]
17. Al-Shaery AM, Aljassmi H, Ahmed SG, Farooqi NS, Al-Hawsawi AN, Moussa M, et al. Real-time pilgrims management using wearable physiological sensors, mobile technology and artificial intelligence. IEEE Access 2022;10:120891-120900. [doi: 10.1109/access.2022.3221771]
18. de Lima LL, Ferreira AL, de Araujo WR. Sensing interfaces: materials for wearable sensors. Encycl Sens Biosens 2023;2:636-649 [FREE Full text] [doi: 10.1016/B978-0-12-822548-6.00127-8]
19. Risch M, Grossmann K, Aeschbacher S, Weideli OC, Kovac M, Pereira F, et al. Investigation of the use of a sensor bracelet for the presymptomatic detection of changes in physiological parameters related to COVID-19: an interim analysis of a prospective cohort study (COVI-GAPP). BMJ Open 2022 Jun 21;12(6):e058274 [FREE Full text] [doi: 10.1136/bmjopen-2021-058274] [Medline: 35728900]
20. Grzesiak E, Bent B, McClain MT, Woods CW, Tsalik EL, Nicholson BP, et al. Assessment of the feasibility of using noninvasive wearable biometric monitoring sensors to detect influenza and the common cold before symptom onset. JAMA Netw Open 2021 Sep 01;4(9):e2128534 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.28534] [Medline: 34586364]
21. Alqahtani AS, Althimiri NA, BinDhim NF. Saudi Hajj pilgrims' preparation and uptake of health preventive measures during Hajj 2017. J Infect Public Health 2019 Nov;12(6):772-776 [FREE Full text] [doi: 10.1016/j.jiph.2019.04.007] [Medline: 31023600]
22. Benkouiten S, Brouqui P, Gautret P. Non-pharmaceutical interventions for the prevention of respiratory tract infections during Hajj pilgrimage. Travel Med Infect Dis 2014 Sep;12(5):429-442 [FREE Full text] [doi: 10.1016/j.tmaid.2014.06.005] [Medline: 24999278]
23. Alqahtani AS, Wiley KE, Tashani M, Willaby HW, Heywood AE, BinDhim NF, et al. Exploring barriers to and facilitators of preventive measures against infectious diseases among Australian Hajj pilgrims: cross-sectional studies before and after Hajj. Int J Infect Dis 2016 Jun;47:53-59 [FREE Full text] [doi: 10.1016/j.ijid.2016.02.005] [Medline: 26875699]
24. Aina YA, Abubakar IR, Almulhim AI, Dano UL, Maghsoodi Tilaki MJ, Dawood SR. Digitalization and smartification of urban services to enhance urban resilience in the post-pandemic era: the case of the pilgrimage city of Makkah. Smart Cities 2023 Aug 11;6(4):1973-1995. [doi: 10.3390/smartcities6040092]
25. Felemban EA, Rehman FU, Biabani SA, Ahmad A, Naseer A, Majid AR, et al. Digital revolution for Hajj crowd management: a technology survey. IEEE Access 2020;8:208583-208609. [doi: 10.1109/access.2020.3037396]
26. Abalkhail AA, Al Amri SM. Saudi Arabia’s management of the Hajj season through artificial intelligence and sustainability. Sustainability 2022 Oct 29;14(21):14142. [doi: 10.3390/su142114142]
27. Mejia C, Ciarlante K, Chheda K. A wearable technology solution and research agenda for housekeeper safety and health. Int J Contemp Hosp Manag 2021 Aug 04;33(10):3223-3255. [doi: 10.1108/ijchm-01-2021-0102]
28. Papagiannidis S. TheoryHub Book. Newcastle upon Tyne, England: Newcastle University; 2022.
29. Marikyan M, Papagiannidis P. Unified theory of acceptance and use of technology. In: TheoryHub Book. Newcastle upon Tyne, UK: Newcastle University; 2022:148-165.
30. Onwuegbuzie AJ, Leech NL. Sampling designs in qualitative research: making the sampling process more public. Qual Report 2015 Jan 15;12(2):238-254 [FREE Full text] [doi: 10.46743/2160-3715/2007.1636]
31. Clarke V, Braun V. Successful Qualitative Research: A Practical Guide for Beginners. Thousand Oaks, CA: SAGE Publications; 2013.
32. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
33. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
34. Goodhue DL, Thompson RL. Task-technology fit and individual performance. MIS Q 1995 Jun;19(2):213-236. [doi: 10.2307/249689]

XSL•FO
RenderX

35. Walsh A. Policy on the ethical involvement of human participants in research. University of Manchester. 2021. URL: https://documents.manchester.ac.uk/display.aspx?DocID=28798 [accessed 2025-01-16]
36. Quaium A, Al-Nabhan NA, Rahaman M, Salim SI, Toha TR, Noor J, et al. Towards associating negative experiences and recommendations reported by Hajj pilgrims in a mass-scale survey. Heliyon 2023 May;9(5):e15486 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e15486] [Medline: 37144197]
37. Taibah H, Arlikatti S, Andrew SA, Maghelal P, DelGrosso B. Health information, attitudes and actions at religious venues: evidence from hajj pilgrims. Int J Disaster Risk Reduct 2020 Dec;51:101886 [FREE Full text] [doi: 10.1016/j.ijdrr.2020.101886] [Medline: 32995254]
38. Baker SB, Xiang W, Atkinson I. Internet of things for smart healthcare: technologies, challenges, and opportunities. IEEE Access 2017;5:26521-26544. [doi: 10.1109/access.2017.2775180]
39. Wen D, Zhang X, Lei J. Consumers' perceived attitudes to wearable devices in health monitoring in China: a survey study. Comput Methods Programs Biomed 2017 Mar;140:131-137. [doi: 10.1016/j.cmpb.2016.12.009] [Medline: 28254069]
40. Yamin M. Secure and healthy Hajj management: a technological overview. Am Acad Scholarly Res J 2015;7(3):195-202 [FREE Full text]
41. Nahavandi D, Alizadehsani R, Khosravi A, Acharya UR. Application of artificial intelligence in wearable devices: opportunities and challenges. Comput Methods Programs Biomed 2022 Jan 30;213:1-19 [FREE Full text] [doi: 10.1016/j.cmpb.2021.106541] [Medline: 34837860]

Abbreviations

AI: artificial intelligence
COREQ: Consolidated Criteria for Reporting Qualitative Research
HRV: heart rate variability
IoT: Internet of Things
TTF: task-technology fit
UTAUT: unified theory of acceptance and use of technology

Edited by A Mavragani; submitted 13.05.24; peer-reviewed by S Pesälä, S Chantziara, H Rashid; comments to author 06.08.24; revised version received 03.09.24; accepted 19.01.25; published 08.04.25.

Please cite as:
Maddah N, Verma A, Ainsworth J
Opportunities and Challenges Surrounding the Use of Wearable Sensor Bracelets for Infectious Disease Detection During Hajj: Qualitative Interview Study
JMIR Form Res 2025;9:e60484
URL: https://formative.jmir.org/2025/1/e60484
doi:10.2196/60484
PMID:

©Noha Maddah, Arpana Verma, John Ainsworth. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A New Mobile App to Train Attention Processes in People With Traumatic Brain Injury: Logical and Ecological Content Validation Study

Roxanne Laverdière[1], BA; Philip L Jackson[1], PhD; Frédéric Banville[2], PhD

[1]School of Psychology, Université Laval, Quebec, QC, Canada

[2]Department of Health Sciences, Université du Québec à Rimouski, Rimouski, QC, Canada

Corresponding Author:
Frédéric Banville, PhD
Department of Health Sciences
Université du Québec à Rimouski
300, allée des Ursulines
Rimouski, QC, G5L 3A1
Canada
Phone: 1 418 723 1986 ext 1931
Email: frederic_banville@uqar.ca

Abstract

Background: Attention is at the base of more complex cognitive processes, and its deficits can significantly impact safety and health. Attention can be impaired by neurodevelopmental and acquired disorders. One validated theoretical model to explain attention processes and their deficits is the hierarchical model of Sohlberg and Mateer. This model guides intervention development to improve attention following an acquired disorder. Another way to stimulate attention functions is to engage in the daily practice of mindfulness, a multicomponent concept that can be explained by the theoretical model of Baer and colleagues. Mobile apps offer great potential for practicing mindfulness daily as they can easily be used during daily routines, thus facilitating transfer. Laverdière and colleagues have developed such a mobile app called *Focusing*, which is aimed at attention training using mindfulness-inspired attentional exercises. However, this app has not been scientifically validated.

Objective: This research aims to analyze the logical content validity and ecological content validity of the *Focusing* app.

Methods: Logical content validation was performed by 7 experts in neuropsychology and mindfulness. Using an online questionnaire, they determined whether the content of the attention training app exercises is representative of selected constructs, namely the theoretical model of attention by Sohlberg and Mateer and the theoretical model of mindfulness by Baer and colleagues. A focus group was subsequently held with the experts to discuss items that did not reach consensus in order to change or remove them. Ecological content validation was performed with 10 healthy adults. Participants had to explore all sections of the app and assess the usability, relevance, satisfaction, quality, attractiveness, and cognitive load associated with each section of the app, using online questionnaires.

Results: Logical content validation results demonstrated a high content validity index (CVI) of the attention training app. Excellent scores (CVI ≥0.78) in both the attention and mindfulness models were obtained for all exercises in the app, except 2 exercises. One of these exercises was subsequently modified to include expert feedback, and one was removed. Regarding ecological content validation, the results showed that workload, quality, user experience, satisfaction, and relevance of the app were adequate. The Mobile Application Rating Scale questionnaire showed an average quality rating between 3.75/5 (SD 0.41) (objective quality) and 3.65/5 (SD 0.36) (subjective quality), indicating acceptable quality. The mean global attractiveness rating from the AttrakDiff questionnaire was 2.36/3 (SD 0.57), which represents one of the strengths of the app.

Conclusions: Logical and ecological content validation showed that *Focusing* is theoretically valid, with a high level of agreement among experts and healthy participants. This tool can be tested to train attention processes after a neurological insult such as traumatic brain injury.

(JMIR Form Res 2025;9:e64174) doi:10.2196/64174

XSL•FO
RenderX

KEYWORDS

mobile app; attention training; cognitive remediation; mindfulness; psychometric properties; content validation

Introduction

Attention Process Training

Attention is the ability to select relevant stimuli from the environment and inhibit those that are not relevant [1]. This cognitive process makes it possible for people to direct attention to the most salient elements to handle them appropriately. Indeed, attention is a factor of cognitive efficiency that supports memory, perception, and problem-solving [2]. Attention deficits are associated with significant difficulties in instrumental activities of daily living that influence the overall quality of life of individuals whose cognition is compromised [3,4].

Over the past decades, several experts have recommended the use of cognitive remediation, a nonpharmacological approach, for individuals with attention deficits [5].

People who live with attention deficits associated with a condition, such as traumatic brain injury (TBI) or attention-deficit/hyperactivity disorder, may benefit from attention training interventions [5]. This kind of intervention, particularly metacognitive strategy training (ie, anticipate, plan, seek, self-regulate, and evaluate), has been proven effective in neurological populations, such as people with TBI [6]. Cognitive remediation aims to improve cognitive functioning either by training to overcome deficits in functions or by allowing individuals to acquire strategies to best exploit their residual functions and help them engage in their daily activities [7]. Attention Process Training-III (APT-III) [8] is one of the most validated attention training programs for people with TBI. It includes two 45-minute sessions per week for 6 weeks. It includes two 45-minute sessions per week for 6 weeks, comprising visual and auditory exercises aimed at training the 5 attentional components of the model from Sohlberg and Mateer [9]. However, several of the studies addressing the effectiveness of this intervention have significant methodological limitations, including a lack of reliability and validity, and a low number of participants. Moreover, the results of the studies showed improvements among individuals in different attentional tests without any evidence of transfer to functional tasks [6].

According to the results of several studies [10], it is possible to stimulate and optimize attention functions using mindfulness practice. Indeed, mindfulness is a state of consciousness that emerges from deliberately focusing on an object in the present moment [11]. An analysis of the “cognitive mechanics” of mindfulness revealed that attention, working memory, and their different processes are key [1,10]. Indeed, Bishop et al [12] proposed 2 key components of mindfulness: self-regulation of attention (ie, the ability to maintain attention) and orientation to the present moment with curiosity, openness, and acceptance (ie, focusing on internal or external stimuli, inhibition, or activation of certain thoughts). Mindfulness interventions have shown promising results for different populations in terms of attentional capacities, particularly regarding sustained attention and attentional flexibility [13]. For instance, a previous study from our team [14] showed that a mindfulness intervention integrated into an ecological virtual environment was safe, feasible, and acceptable for people with mild TBI. The use of attentional focusing techniques of mindfulness could also be useful for the management of attention difficulties in people who have experienced moderate to severe TBI. Indeed, focusing on the present moment is a way of controlling one’s own attentional processes. As such, attention is considered central to the construct of mindfulness [10]. One limit of this type of intervention is access to a trained guide to accompany individuals, especially for people living in remote areas. It is therefore relevant to elaborate mindfulness interventions accompanied by a remote guide that is accessible from home, regardless of the region.

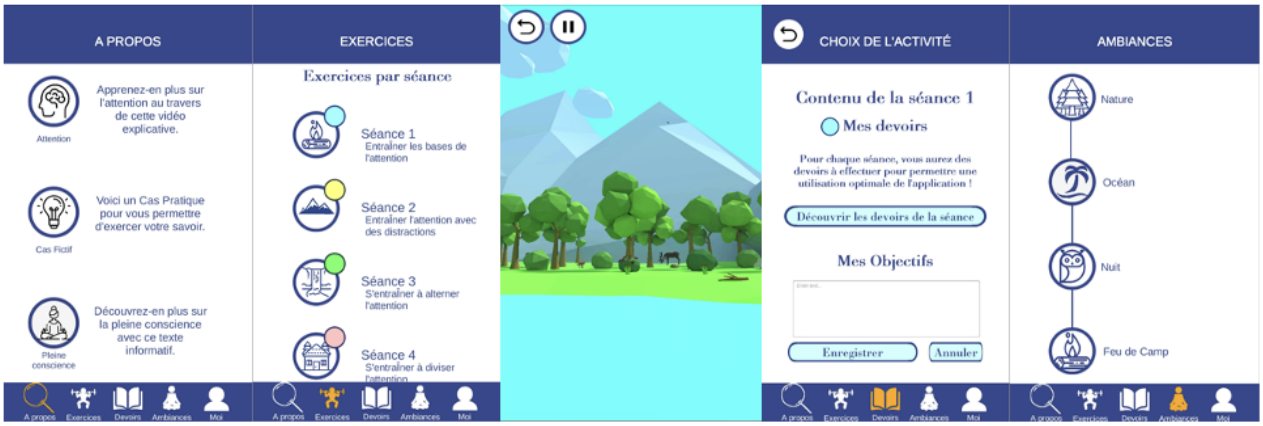
Although not everyone uses technology or is familiar with technology, mobile devices are ubiquitous in everyday life, and their apps are increasingly being used and perfected, particularly in health domains [15]. Countless mindfulness-based mobile apps have been developed; for instance, a 2021 systematic review identified 605 mindfulness-based mobile apps in European app stores [16]. Several meta-analyses confirmed the effectiveness of mindfulness interventions via mobile apps, particularly for mental health [17-19]. A study also showed that mindfulness interventions via a mobile app have effects comparable to mindfulness interventions in person [20]. These apps contain guided mindfulness exercises inspired by traditional approaches (mindfulness-based stress reduction [MBSR] [11] and mindfulness-based cognitive therapy [MBCT] [21]) that can be performed by individuals at home. However, even if studies show that these apps work, there is a lack of studies in the literature supporting the content and quality of these apps [22-25], and no studies have been conducted on the use of validated mindfulness apps in individuals with TBI. A systematic review in 2019 identified 53 mobile apps dedicated to people with TBI on the Apple App Store and Google Play Store [26]. Among these, there were compensatory measures, and some apps were aimed more at the remediation of cognitive functions, but none were aimed specifically at training attention using mindfulness. Findings from studies on the effectiveness of these apps in improving TBI symptoms are inconsistent. There is very limited research to validate the content of these apps, which poses a risk to users. Indeed, a specific validation is required to use these types of apps, considering that people who have experienced TBI (more specifically moderate or severe TBI) often have a variety of physical and cognitive deficits that can affect their ability to perform a mindfulness program via a mobile app. Physically, they may have motor disorders (paralysis, muscle weakness, and balance disorder), persistent fatigue, or sensory disorders (vision or hearing). Cognitively, deficits may involve attention, memory, language, and executive functions such as problem-solving and decision-making. These limitations may also be accompanied by emotional or behavioral disorders.

To explore the potential of mobile apps in the training of attention via a mindfulness focusing technique, we developed

XSL•FO
RenderX

an app, called *Focusing*, based on the Obesity-Related Behavioral Intervention Trials (ORBIT) intervention development model [27], which aimed at attention process training following the occurrence of neurological disorders such as TBI [28]. *Focusing* is currently available in French. The mobile app consists of mindfulness-inspired attentional focus exercises (Figure 1). The proposed attention training is based on 2 validated theoretical models: the mindfulness model of Baer et al [29] and the theoretical attention framework of Sohlberg and Mateer [9]. *Focusing* is composed of 5 sections (Multimedia Appendix 1). The first section "About" contains an interactive tutorial, a psychoeducational video about attention after TBI with a thought-provoking case example, and explanatory text about mindfulness. The second section "Exercises" contains the 4 attentional focus exercise sessions of the app. Each session includes 4 exercises (the first exercise is the same for each session to promote a state of calm conducive to the exercises of the session). Each session involves a component of attention, ranging from the simplest to the most complex. The third section "Homework" contains a list of homework assignments for each week. It also contains a "Goals" section, allowing the participant to set objectives and to consult with them throughout the week. The fourth section "Ambiences" contains ambient sounds representing nature and is aimed at freely practicing attentional focus in the visual and auditory modalities. The fifth section "Me" contains the participant's profile.

Figure 1. Screen captures of the different sections of the attention process training app (Focusing app) for people with traumatic brain injury.



Theoretical Framework

The development of the attention training app was based on the neuropsychological model of attention of Sohlberg and Mateer [9]. This model was developed based on data from populations with brain damage and includes the subcomponents of attention that are frequently affected by TBI. This model has been widely used to explain attention in general, regardless of the neurological disorder. It involves the processes of focused, sustained, selective, alternating, and divided attention. *Focused attention* represents the ability to respond specifically and in a targeted way to a sensory stimulus. *Sustained attention* is the ability to maintain a consistent behavioral response during continuous or repetitive activity. *Selective attention* represents the ability to maintain a given behavioral or cognitive response to stimuli exerting a distracting or competitive effect. *Alternating attention* shifts the focus between tasks with distinct cognitive requirements efficiently and repeatedly. Finally, *divided attention* represents the ability to perform multiple tasks simultaneously.

To integrate this theoretical model into the *Focusing* app, each of the 4 sessions trains one of the components of the model by Sohlberg and Mateer [9]: Session 1, sustained attention; Session 2, selective attention; Session 3, alternating attention; and Session 4, divided attention. Focused attention, being a basic process, is trained in all sessions. The order of the sessions was determined to train the simplest forms of attention (sustained and selective) before the most complex forms (alternating and divided).

The development of the *Focusing* app was also based on the 5-facet model of mindfulness of Baer et al [29]. This model describes mindfulness as a multidimensional construct. It includes the following 5 components: observation, description, acting mindful, nonjudgment of inner experience, and nonreactivity to inner experience. *Observation* refers to noticing or paying attention to internal and external experiences, such as sensations, cognitions, emotions, images, sounds, and smells. *Description* refers to labeling internal experiments with words. *Acting mindful* includes being attentive to your current activities and may be opposed to behaving mechanically while one's attention is focused elsewhere (often called autopilot). *Nonjudgment of inner experience* is about refraining from assessing thoughts and feelings. *Nonreactivity to inner experience* is about letting thoughts and feelings come and go, without being taken in or carried away by them.

To integrate the 5 mindfulness components of this model into *Focusing*, we selected and modified mindfulness exercises to fit the definitions of the components. The objective was for the intervention exercises to be broadly representative of all aspects of the model by Baer et al [29]. The components have therefore been integrated into the attentional exercises included in the 4 sessions.

XSL•FO
RenderX

Objectives

This study aimed to assess the content validity of the *Focusing* app. This research, inspired by the previous work of Vogt et al [30], had the specific goal of analyzing the logical content validity and ecological content validity of the attention training program delivered through the mobile app. Logical content validation aims to test whether the app and the proposed exercises are aligned with "an operational definition" of the theoretical framework. More specifically, experts were instructed to decide on the level of proximity between the proposed exercises and the theoretical constructs presented in this study. Ecological content validation (on the basis of user experience in everyday life) aims to determine, before the proof-of-concept stage with a clinical population, the ease of use of the mobile app and the comprehensibility of the exercises (including vocabulary and visual/auditory aspects) that are integrated in the app.

Methods

Early Version and Preliminary Work

An initial app was developed based on meditation exercises from the MBSR [11] and MBCT [21] approaches. Mindfulness and neuropsychology experts (N=7) were subsequently consulted with a Delphi approach to obtain their views on the exercises. In general, the experts recommended using a mindfulness intervention approach in its entirety. Indeed, it is not recommended to develop an intervention that modifies or combines 2 approaches. The experts strongly suggested starting over by doing something completely different from existing mindfulness approaches. In response to the recommendations of the first group of experts, a new version of the *Focusing* app for training attentional functions was developed by Laverdière et al [28] and was subjected to validation in this study.

Content Validation

Face validity represents the subjective feeling of the validity of an instrument [31,32]. It is not strictly about what the instrument measures, but what it seems to measure according to the research team. An instrument with good apparent validity will generate more positive attitudes, greater involvement in the task, and greater authenticity in the responses, which will increase the validity of the task performed [33]. For the development of the *Focusing* app, the research team implemented mindfulness-inspired attentional focus activities based on the understanding of the theoretical model of attention by Sohlberg and Mateer [9] and the model of mindfulness by Baer et al [29]. The construction of the intervention is based on the ORBIT model [28], which is considered valid by the research team.

Content validity assesses the extent to which the various items of an instrument are representative of the construct and its different facets [30,31]. Haynes et al [34] proposed 7 essential rules for validating the content of an instrument: Rule 1, rigorously define the domain and facets of the measured construct and validate this definition; Rule 2, use a sample of experts and members of the target population to create the items and define other aspects of the instrument; Rule 3, submit all aspects of the instrument to content validation; Rule 4, consult several experts to validate the content of the instrument and quantify their judgments using formalized scales; Rule 5, examine the proportional representation of items or tests relative to the different facets of the construct; Rule 6, present the results of content validation when publishing any new instrument; and Rule 7, consider all subsequent psychometric analyses to refine content validation.

To develop the *Focusing* app, validated theoretical models of attention and mindfulness were used (Rule 1) [9,29]. Mindfulness and attention training experts were first consulted when the intervention was created (Rule 2). Moreover, the entire attention process training was validated by experts and people from the general population (Rules 2, 3, and 4). The validation data allowed to refine the intervention and thus to have a development procedure and a valid intervention (Rule 5). Studies on the validation process are presented in this manuscript (Rule 6). All the data obtained will make it possible to reflect comprehensively on the modifications to be made to the app for improvement before testing the app in a clinical population (Rule 7).

Ethical Considerations

Ethics approval was obtained for this study from the *Centre Intégré Universitaire de Santé et de Services Sociaux de la Capitale-Nationale* (CIUSSS-CN) rehabilitation and social integration committee (approval number: MP-13-2022-2523). Written consent was obtained from all participants after they were informed about the study's objectives and potential benefits. Participation was completely voluntary, and participants retained the right to withdraw at any moment without prejudice. To protect their privacy, all data were kept strictly confidential and anonymized, ensuring that participants could not be identified. No financial compensation or incentives were offered for participation.

Study 1: Logical Content Validation

Participants

Logical content validation was conducted by a panel of 7 experts from Canada (n=5) and France (n=2) who participated in the validation process in February 2023. The selected experts had to meet the following inclusion criteria: (1) be a clinician, have at least 5 years of clinical experience with moderate or severe TBI clients, or have at least 5 years of experience in the mindfulness practice; (2) be a neuropsychologist, psychologist, or occupational therapist; and (3) have practiced or be familiar with physical rehabilitation. The experts were recruited via email from September 2022 to January 2023. A total of 21 experts were contacted. Of these 21 experts, 10 agreed to participate; however, 1 expert withdrew and 2 did not complete all the validation steps. The participants included 4 clinical neuropsychologists involved in teaching and research. They specialize in the development of techniques for evaluation or intervention in different clinical populations (TBI, autism spectrum disorder, and neurodegenerative diseases). The other 3 participants are clinical psychologists involved in teaching and research. They specialize in mindfulness interventions in pediatric and adult populations.

XSL•FO
RenderX

Procedure

Experts who agreed to participate in the study received an email containing information about the exercises and a link to download a desktop version of the *Focusing* app, which they were asked to explore over a 4-week period. The desktop version of the app corresponds to a preliminary version of the implementation on the Google Play Store and is visually identical to the currently available app on Android. The experts were invited to try the different sections of the app and to perform all the exercises of the program. They were asked to assess whether the exercises in the program are theoretically valid according to the theoretical model of attention by Sohlberg and Mateer [9] and the theoretical model of mindfulness by Baer et al [29], using a questionnaire on the Lime Survey platform (closed survey). In the questionnaire, experts had access to the definitions of the attentional components of the model by Sohlberg and Mateer [9] and the 5 components of the model of mindfulness by Baer et al [29] because their respective knowledge of the theoretical models was different. In addition, the component of attention trained in each exercise was identified. The experts had to assess whether each exercise involves the targeted component of attention, using a Likert scale ranging from 1 (totally disagree) to 4 (totally agree). Using the same Likert scale, they also had to assess the extent of training of each of the 5 components of mindfulness, in general, in the exercises. Indeed, the 5 components need not be present in all the exercises, but rather they should be in the mindfulness approach used for the construction of the exercises. They also had to comment on the exercises and suggest improvements when they considered that an item was not theoretically valid. Finally, the experts had to assess the relevance of the first exercise of each session, using a Likert scale ranging from 1 (totally irrelevant) to 4 (totally relevant). The questionnaire consisted of 35 items and 20 pages. The experts could see their progress in the questionnaire using a progress bar at the top of the page. They could also save their answers and resume the questionnaire later using their email and a password. They also had the option to change their answers and go back in the questionnaire. No cookies were used. The experts' responses were anonymous. Data and comments were then analyzed, and a discussion meeting with the experts (videoconference) was subsequently held to discuss the elements that did not reach consensus between the experts. The changes suggested by the experts were then applied to the mobile app. The experts concluded this meeting in a consensual manner by saying that they would consider the app to be valid once their suggestions were incorporated. After making the adjustments and incorporating the experts' suggestions, one of the experts reviewed the entire content of the intervention to validate it again. This expert specializes in adapting mindfulness interventions to different clinical populations. In addition, one of the mindfulness experts trained the person who lent her voice when recording voice tone and speed exercises. Once this step was completed, the usability of the new version of the mobile app was tested.

Data Analysis

For the evaluation of the content of the app and the theoretical model of Sohlberg and Mateer [9] regarding attention and that of Baer et al [29] regarding mindfulness by experts, a Likert-type scale was used with 4 possible responses: 1, I totally disagree; 2, I somewhat disagree; 3, I somewhat agree; and 4, I totally agree. For this analysis, the content validity index (CVI) was calculated from the count of the responses "I somewhat agree" and "I totally agree" divided by the total number of experts. According to the literature, the items evaluated must have a CVI greater than or equal to 0.78 (78%); otherwise, they need to be adjusted according to experts' suggestions [35].

Results

The experts assessed the theoretical validity of each exercise of the sessions in relation to the attentional model used [9] for a total of 16 exercises. They also assessed the theoretical validity of the attention training program in relation to the mindfulness model used [29].

Regarding the validity of the attention model (Table 1), the calculated CVI for 14 exercises (out of 16) ranged between 0.86 and 1.00, indicating that these exercises are valid [35]. One exercise with a CVI value of 0.71 required revision. The experts mentioned that this exercise was too focused on mindfulness concepts, which strayed away from the concept of selective attention. They also pointed out that more distractions should be present in the exercise. In addition, experts considered another exercise (Exercise 4.4) to be not valid, as it had a CVI value of 0.42, and thus, it was eliminated. According to the experts, Exercise 4.4, which consisted of a dual visual-auditory task, was too demanding for people with acquired brain injury. As mentioned above, the first exercise of each session is the same to allow participants to get into a state conducive to exercises aimed at training the different components of attention. The relevance of repeating this exercise was assessed by the experts. The calculated CVI was 0.86, indicating that the repetition of this exercise is relevant to the program.

For validation related to the mindfulness theoretical framework (Table 2), the calculated CVI for all 5 components of the model ranged between 0.86 and 1.00, indicating that the remediation program is valid with regard to its mindfulness component.

XSL•FO
RenderX

Table 1. Content validity index and data of each exercise based on the attention model by Sohlberg and Mateer [9], as evaluated by expert judges.

Instrument item[a]	CVI[b]	Mean (SD)	Modification
Session 1: focused and sustained attention			
Exercise 1.1	0.86	3.33 (0.47)	Minor
Exercise 1.2	0.86	3.29 (0.70)	Minor
Exercise 1.3	0.86	3.29 (0.70)	Minor
Exercise 1.4	0.86	3.14 (0.64)	Minor
Session 2: selective attention			
Exercise 2.2	0.71[c]	3.43 (0.90)	Major
Exercise 2.3	0.86	3.29 (0.70)	Minor
Exercise 2.4	0.86	3.14 (0.64)	Minor
Session 3: alternating attention			
Exercise 3.2	0.86	3.29 (0.70)	Minor
Exercise 3.3	1.00	3.71 (0.45)	No change
Exercise 3.4	0.86	3.14 (0.64)	Minor
Session 4: divided attention			
Exercise 4.2	0.86	3.57 (0.73)	Minor
Exercise 4.3	0.86	3.57 (0.73)	Minor
Exercise 4.4	0.42[c]	2.71 (0.88)	Item deleted

[a]Exercises 2.1, 3.1, and 4.1 are not included in the table as they have the same content as Exercise 1.1.

[b]CVI: content validity index.

[c]Items with a CVI below the acceptable level of 0.78.

Table 2. Content validity index and data of each component of the theoretical framework of mindfulness by Baer et al [29] for the whole training according to expert judges.

Instrument item	CVI[a]	Mean (SD)	Modification
Observing	1.00	3.86 (0.35)	No change
Describing	0.86	3.00 (0.53)	Minor
Acting with awareness	0.86	3.57 (0.73)	Minor
Nonjudging of inner experience	1.00	3.71 (0.45)	No change
Nonreactivity to inner experience	1.00	3.57 (0.49)	No change

[a]CVI: content validity index.

Synthesis of Expert Recommendations

During the logical content validation, 4 dimensions were explored by the experts: theoretical model of attention, theoretical model of mindfulness, technological aspects, and audio. Regarding the theoretical model of attention, the recommendations of the experts were to: (1) increase the duration of exercise in sustained attention; (2) increase distractions in selective attention; and (3) remind more frequently about the instructions in all exercises. Regarding the theoretical model of mindfulness, the recommendations of the experts were to: (1) simplify the vocabulary; (2) give more concrete examples to the participants; and (3) decrease the speed of the voice in the audio tracks of the exercises. In terms of technology, the experts reported some inconsistencies between the visual elements of the app and the audio content of the exercises.

Exercises were modified to reflect these recommendations. First, the sustained attention exercises (1.1 to 1.4) were extended by a few minutes each. Visual and auditory distractions were added in the selective attention exercises (2.1 to 2.4). Throughout the training exercises, the speech was adapted by adding reminders of instructions several times during a single exercise. The vocabulary was also adapted to be accessible to the general population, for example, by adding concrete examples and adapting the tone of voice (speaking less quickly). The technological inconsistencies raised by the experts were also corrected.

XSL•FO
RenderX

Study 2: Ecological Content Validation

Participants

For this study, a convenience sample of 10 healthy participants (7 female and 3 male participants) was recruited from the social networks of the research laboratory. Participants were required to meet the following criteria: (1) be over 18 years of age; (2) understand and speak French; (3) own an Android device; (4) have no diagnosis of TBI; (5) have no serious mental disorder including but not restricted to schizophrenia spectrum disorder, bipolar disorder, intellectual disability, and attention deficit disorder; (6) have no vestibular disorder; and (7) have no substance abuse.

An academic researcher also explored the mobile app. This expert is interested in the neuropsychology of rehabilitation, particularly among adults who have experienced TBI or stroke. This researcher's work also focuses on the use of technology in clinical neuropsychology (mobile devices, teleneuropsychology, and virtual reality). This expert did not complete the validation questionnaires, but this expert's comments were taken into consideration in the validation process.

Procedure

The 10 participants completed the entire study remotely. Interested participants were directed to an online eligibility questionnaire via a link in the recruitment advertisement. Eligible participants received the mobile app download procedure. They then had 2 weeks to explore by themselves the mobile app and fill out the questionnaires provided for the experimental protocol via Lime Survey (open survey). The self-administered questionnaires were: a sociodemographic questionnaire, the Simulation Task Load Index (SIM-TLX) questionnaire (mental load; [36]), the French version of the Mobile Application Rating Scale (MARS-F) questionnaire (quality of the app; [37,38]), the French version of the AttrakDiff questionnaire (user experience; [39,40]), and a questionnaire concerning the participants' satisfaction with the elements of the app as well as their perception of the relevance of these elements. The whole questionnaire set consisted of 107 items and 9 pages. Participants could see their progress in the questionnaire on a progress bar at the top of the page. They could also save their answers and resume the questionnaire later using their email and a password. They also had the option to change their answers and go back in the questionnaire. No cookies were used. The participants' responses were anonymous. Once all participants completed the exploration, questionnaire responses were analyzed and changes to the mobile app were made if problematic elements were identified.

Instruments

SIM-TLX Questionnaire

This questionnaire [36] provides a measure of participants' cognitive load. It contains 9 scales (mental demands, physical demands, temporal demands, frustration, task complexity, situational stress, distractions, perceptual strain, and task control). The responses range from 1 (very low) to 20 (very high).

MARS-F Questionnaire

This questionnaire [37,38] provides a measure of the app's quality. Items are rated on a scale of 1 to 5 (1=inadequate; 2=poor; 3=acceptable; 4=good; and 5=excellent). The MARS-F questionnaire comprises 4 main scales: (1) user engagement (5 items: entertainment, interest, customization, interactivity, and target group); (2) functionality (4 items: performance, usability, navigation, and gestural design); (3) esthetics (3 items: layout, graphics, and visual appeal); and (4) information quality (7 items: accuracy of app description, goals, quality of information, quantity of information, quality of visual information, credibility, and evidence base). For each scale, a mean score was calculated.

AttrakDiff Questionnaire

This questionnaire [39,40] provides a measure of the app user experience. It provides data on 3 quality scales (pragmatic, hedonic-stimulation, and hedonic-identity) and a scale of global attractiveness. Each item is rated on a scale ranging from –3 to +3. Values close to the average (between –1 and 1) are standard; they mean that the product meets its objective. Values outside this neutral zone are to be considered as positive (from 1 to 3) or negative (from –1 to –3) points of the app.

Satisfaction Questionnaire

In this questionnaire, participants rated on a scale of 1 (totally dissatisfied) to 5 (totally satisfied) their levels of satisfaction regarding the 5 sections (About, Exercises, Homework, Ambiences, and Me) and the functionalities of the app. Participants also rated the relevance of each section and functionality (technological issues, sound, and ability to navigate) on a scale from 1 (totally irrelevant) to 5 (totally relevant). They also documented the technical issues they encountered while exploring the app.

Data Analysis

To compute for all participants the variables surveyed by the questionnaires, central trend measurements (mean and SD) and frequency analyses were performed on the data. Exploratory correlations were made on the data.

Results

A total of 11 participants started the usability validation study. Only 10 participants (7 female and 3 male participants) completed the validation process of the app. Their mean age was 51 (SD 15) years, and the mean number of years of education was 13 (SD 4). Most participants (5/10, 50%) worked full-time, 40% (4/10) were retired, and 10% (1/10) were studying full-time. None worked in the technology industry. Moreover, 80% (8/10) of the participants regularly watched animated films and only 40% (4/10) regularly played video games. The smartphone was the most used technological tool by participants (9/10, 90% used it daily) followed by computers (4/10, 40% used it daily) and tablets (2/10, 20% used it daily). On a scale of 1 to 10, participants' comfort levels with these technological tools averaged 7.36 (SD 1.57) for smartphones, 5.55 (SD 2.58) for computers, and 5.45 (SD 2.62) for tablets.

The results of the different questionnaires are presented in Table 3.

XSL•FO
RenderX

Table 3. Scores of the scales of the questionnaires for usability validation of attention process training according to healthy participants.

Questionnaire scale	Score, mean (SD)	Range, minimum to maximum
SIM-TLX[a] (range: 1 [very low] to 20 [very high])		
Mental demands	4.73 (2.33)	1 to 8
Physical demands	1.45 (1.04)	1 to 4
Temporal demands	1.36 (0.81)	1 to 3
Frustration	1.45 (1.21)	1 to 5
Task complexity	3.82 (2.27)	1 to 7
Situational stress	1.18 (0.60)	1 to 3
Distractions	5.55 (3.59)	1 to 11
Perceptual strain	2.64 (4.20)	1 to 15
Task control	2.73 (2.41)	1 to 8
MARS-F[b] (range: 1 [low quality] to 5 [high quality])		
Engagement	3.98 (0.48)	1 to 5
Functionality	4.30 (0.45)	3 to 5
Esthetic	3.80 (0.45)	3 to 5
Information	2.89 (0.60)	1 to 5
Objective quality	3.74 (0.41)	1 to 5
Subjective quality	3.65 (0.36)	2 to 5
AttrakDiff (range: –3 to 3)		
Pragmatic quality	1.94 (0.61)	0 to 3
Hedonic quality: Simulation	1.79 (0.55)	–2 to 3
Hedonic quality: Identification	1.51 (0.57)	–2 to 3
Global attractivity	2.36 (0.57)	0 to 3
Satisfaction (range: 1 [totally dissatisfied/irrelevant] to 5 [totally satisfied/relevant])		
Satisfaction	4.30 (0.62)	2 to 5
Relevance	4.41 (0.47)	2 to 5

[a]SIM-TLX: Simulation Task Load Index.

[b]MARS-F: Mobile Application Rating Scale, French version.

The results showed that the app requires low physical and temporal demands according to the SIM-TLX questionnaire. All items scored less than 4. Even if the overall results on the 9 scales were low, mental demand, distractions, and task complexity were rated slightly higher than the others. The app generated a low level of frustration and stress (<1.5). Participants had a good sense of control when using the app (a small score indicates a good sense of control).

The results of the MARS-F questionnaire showed that, in general, the quality of the app was adequate according to the participants. The average quality rating for the app was between 3.75/5 (SD 0.41) (objective quality) and 3.65/5 (SD 0.36) (subjective quality), demonstrating moderate quality [37,38]. The results indicated that the app is functional and esthetic and provides an adequate level of engagement. The information scale had a lower score (mean 2.89, SD 0.60). This suggests that participants did not have enough information about the app development process to judge its credibility.

The result of AttrakDiff's global attractiveness scale indicated that the user experience was adequate, with a mean score of 2.36 (SD 0.57). The result of the pragmatic quality scale, which describes the usability of the app, indicated that using the app was easy. The result of the hedonic quality scale-stimulation indicated that the app appeared new, interesting, and stimulating. The result of the hedonic quality scale-identity indicated that participants liked to interact with the app (eg, voice). In fact, all results were between 1 and 3 on the –3 to +3 scale, which points out the positive aspects of the app.

The satisfaction questionnaire showed that participants were generally satisfied with the different sections and functionalities of the app. They also considered that the different sections and functionalities of the app were relevant.

Synthesis of Correlation Analyses

A positive and strong correlation was obtained between the participants' ease of using a tablet and their ease of using a

XSL•FO
RenderX

phone (r=0.832; P=.001). This suggests that *Focusing* could be used on a tablet as well.

A negative and moderate correlation was obtained between the result of the cognitive load of participants (SIM-TLX) and the functionality score of the app (r=–0.635; P=.049). Better functionality was therefore associated with a smaller mental load. Having obtained a high score in functionality (MARS-F), the app does not appear to lead to mental overload that could interfere with the attention processes of participants.

A positive and moderate correlation was obtained between pragmatics (ease of use; AttrakDiff) and engagement level (MARS-F) (r=0.657; P=.04). The ease of use of *Focusing* therefore increases participants' commitment to the intervention. Better engagement was positively and strongly correlated with the subjective quality (MARS-F) of the app perceived by the participant (r=0.821; P=004). The level of satisfaction of participants with the different sections of *Focusing* was also positively and moderately correlated with objective quality (MARS-F) (r=0.669; P=.04) and positively and strongly correlated with the overall level of attractiveness (AttrakDiff) of the app (r=0.919; P=.001).

Discussion

Main Results

Despite the growing popularity of mobile apps for cognitive training and mindfulness, there is still a lack of comprehensive analysis supporting the content and quality of these apps [22-25]. A systematic review identified 53 commercial cognitive training or rehabilitation apps in 2023 [23]. Among these apps, only 5 had an overall MARS score exceeding 4 points, and more than half had a score of less than 3 points. Considering that a MARS score of 3 points is acceptable, most of these apps did not even meet the acceptable criteria for quality [37,38]. Furthermore, to our knowledge, none of these apps were intended to address attention in people who have moderate or severe TBI.

Furthermore, no studies on the development of attention interventions identified in the literature used a validated development model. Due to the existing gaps in the methodology of studies on attention remediation interventions, it is essential to use a model like ORBIT in the development of future interventions to address these limitations. Promising behavioral interventions in development are often abandoned rather than refined or improved if they fail the initial tests. In some cases, successful interventions in early studies are not pushed toward more rigorous efficacy studies, while others are tested in efficacy studies prematurely.

This research aimed to validate the content and user experience of an attention training intervention developed using the ORBIT model via a mobile app called *Focusing*. The first step was to validate the logical content with experts to ensure that the exercises train the attention components of the model by Sohlberg and Mateer [9]. It also aimed to validate that the remediation program sessions incorporate the 5 components of the mindfulness model by Baer et al [29]. The results showed that experts (including experts in cognitive functions and experts in mindfulness) considered the intervention to be theoretically valid according to the 2 models. Regarding the theoretical model of attention, the experts validated that each session of the intervention involves a component of the attentional model. Following the expert logical validation process, 1 exercise was removed, 1 underwent major revisions, 10 were improved according to the recommendations, and 1 was not changed. The removed exercise, which consisted of a dual visual and auditory task, was deemed too demanding for people with acquired brain injury. Major changes included increasing the duration of exercise, increasing the level of distraction, and making the instructions more clear. Moreover, the experts confirmed that the 5 components of the mindfulness model are included throughout the exercises. Changes were made to the exercises as recommended by the experts: the vocabulary for the instructions was simplified, concrete and more practical examples were added, the speed of the voice track/instructions was reduced, and pauses were added to the voice track.

The second step was the ecological content validation of the *Focusing* app. The relevance, satisfaction, quality, attractiveness, and workload associated with each section of the app were evaluated by participants from the general population.

The results of the SIM-TLX questionnaire underlined the level of workload when using the app, which involves low physical and temporal demands. It also involves low stress and frustration. These results confirm that the *Focusing* app respects the principles of mindfulness. Indeed, mindfulness promotes being in the present moment and the acceptance of the experience. The scales of the SIM-TLX questionnaire associated with higher workload are mental demand, distraction, and task complexity. The results confirm that attentional training program exercises require some cognitive effort. In addition, distractors were deliberately integrated into the exercises to train selective attention.

MARS is one of the most widely used questionnaires in the literature for app quality assessment [37]. This questionnaire was also used by clinicians to assess the quality of 23 mindfulness apps available on the Apple App Store and Google Play Store [25]. The results showed that the app is of good quality and is esthetic, functional, and engaging. These features allow you to properly interact with the app. Resources can therefore be entirely devoted to the training of attention. The Quality Information scale had the worst result. To prevent misinformation and adverse effects from *Focusing* app use, information quality must be improved. In fact, wrong or misleading information could result in a decrease in user safety [41]. Informing participants of the involvement of experts in the development process (eg, psychotherapists and researchers in mindfulness or cognition training) might help to address this problem. Moreover, a better description of the app on the Play Store might improve information quality.

The results of the AttrakDiff questionnaire demonstrated good user experience of the *Focusing* app. The app was easy to use, interesting, and challenging. It was also pleasant to interact with the app. This also means that the voice (tone and speed) used for the exercises was adequate. The results showed that the recommendations of the experts during the logical content validation helped improve the user experience of the app.

XSL•FO
RenderX

In comparison with our study, the attention training application (ATA) developed by Hill et al [42,43] is an app aimed at training the attention of elderly people without cognitive impairment. To our knowledge, this app is the most recent in the literature to be validated with the target population. The ATA is an adaptation of the «dual n-back» computerized intervention [44]. This study did not use a validated intervention development model such as ORBIT. The app was, however, validated twice [42,43] with a small number of elderly people and using a mixed methods study design. This study corresponds to Phase I(b)-Refine of the ORBIT model. As in our study, participants were asked to complete questionnaires to capture familiarity with technology and to provide their general impressions of the ATA on a scale from 0 (negative) to 10 (positive in 4 categories: overall, easy, satisfying, and interesting). The User Interface Satisfaction questionnaire was also used [45]. Feedback had to be given on a 7-point Likert-type scale for different items. Interviews were conducted to allow participants to share their perceptions of the app. A limitation of this previous study is that the quality of the app was not assessed with the MARS questionnaire. Moreover, the cognitive load associated with using the app was not assessed.

Limitations

The content validation of the attention remediation program has some limitations. Although the logical content validation experts concluded that the program would be fully valid once their recommendations were incorporated, the revised program was only reviewed by 1 of these experts after changes were implemented. In addition, 1 of the experts could not be present at the focus group to discuss the recommendations.

The current validation of user experience was limited by the characteristics of the participants. First, the sample included more women (70%) than men (30%). However, only one-third of people with TBI in Canada are women [46,47]. In addition, most participants were in their 50s or 60s, while youth and young adults aged 12 to 19 years are 5.2 times more likely to have TBI [46]. Finally, participants in the sample considered themselves weakly to moderately comfortable with the use of electronic devices (computers, tablets, and smartphones), which may be related to the age of the participants. On the other hand, young people are more comfortable with the use of apps on smartphones. Different data about user experience could have been obtained with a sample that is very comfortable with technologies. However, by targeting people who are less skilled with technology, we ensured that this would not be an obstacle.

Future Research

Future studies should assess the safety, tolerability, and acceptability of this attention training intervention via a mobile app in a population with moderate or severe TBI. More specifically, as part of phases Ib (refine) and IIa (proof-of-concept) of the ORBIT model [27], 5 to 10 people with moderate or severe TBI should be recruited as part of a multiple single-case experimental design. The test phase should include a waiting period and an intervention period, both marked by pretests and posttests. At the end of this phase, it will be possible to determine whether the intervention is clinically acceptable and whether it can be made available to a larger sample, leading to phase III (efficacy) studies that will involve a randomized controlled design. At the end of phase III, we also want to translate the tool for the English-speaking population in order to increase the recruitment pool. This will involve translating the audio files and the menus of the different sections of the app. We also want to develop an iOS version of the app to make it more accessible and expand recruitment.

Conclusions

The *Focusing* app is one of the first mobile attention training apps scientifically validated with the ORBIT model and developed based on theoretical models of both attention and mindfulness. A systematic validation approach was used to ensure that the intervention was theoretically valid before clinical validation. In addition, this project provides a better understanding of the attentional processes modulated by attentional focus techniques based on mindfulness. In fact, this study attempts to dissect attentional processes from the angle of a theoretical model recognized in neuropsychology.

Acknowledgments

We thank Jonathan Caron-Roberge who helped with the programming of the mobile app. We also thank Jean-Luc Théberge who helped with the voice recordings of the sessions of the remediation program. Finally, we thank Marie-Josée Tremblay who provided her expertise to guide mindfulness exercises. This project received financial support from the *Fonds de Recherche du Québec, nature et technologies* (FRQ-NT) (reference number: 325564), the *Centre Interdisciplinaire de Recherche en Réadaptation et Intrégration Sociale* (Cirris), and the *Collectif de Recherche sur la Santé en Région* (CoRSeR).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Description of the attention training exercises in each session and the attention components and mindfulness components trained in the exercises.
[DOCX File , 18 KB - formative_v9i1e64174_app1.docx]

References

1. Buttle H. Attention and working memory in mindfulness-meditation practices. The Journal of Mind and Behavior 2011;32(2):123-134 [FREE Full text]
2. Cohen RA. The Neuropsychology of Attention. New York, NY: Plenum Press; 1993.
3. Azouvi P, Arnould A, Dromer E, Vallat-Azouvi C. Neuropsychology of traumatic brain injury: An expert overview. Rev Neurol (Paris) 2017;173(7-8):461-472 [FREE Full text] [doi: 10.1016/j.neurol.2017.07.006] [Medline: 28847474]
4. Silver JM, McAllister TW, Arciniegas DB. Textbook of Traumatic Brain Injury, Third Edition. Washington, DC: American Psychiatric Association Publishing; 2019.
5. Cicerone KD, Goldin Y, Ganci K, Rosenbaum A, Wethe JV, Langenbahn DM, et al. Evidence-based cognitive rehabilitation: systematic review of the literature from 2009 through 2014. Arch Phys Med Rehabil 2019 Aug;100(8):1515-1533 [FREE Full text] [doi: 10.1016/j.apmr.2019.02.011] [Medline: 30926291]
6. Ponsford J, Velikonja D, Janzen S, Harnett A, McIntyre A, Wiseman-Hakes C, et al. INCOG 2.0 guidelines for cognitive rehabilitation following traumatic brain injury, part II: attention and information processing speed. J Head Trauma Rehabil 2023;38(1):38-51. [doi: 10.1097/HTR.0000000000000839] [Medline: 36594858]
7. Franck N. Remédiation Cognitive, 2nd Edition. Issy Les Moulineaux, France: Elsevier Masson; 2017.
8. Sohlberg MM, Mateer CA. Attention Process Training APT3. Lash & Associates. URL: https://lapublishing.com/apt3-attention-process-training/ [accessed 2025-03-24]
9. Sohlberg MM, Mateer CA. Effectiveness of an attention-training program. J Clin Exp Neuropsychol 1987 Apr;9(2):117-130. [doi: 10.1080/01688638708405352] [Medline: 3558744]
10. Prakash RS. Mindfulness meditation: impact on attentional control and emotion dysregulation. Arch Clin Neuropsychol 2021 Oct 13;36(7):1283-1290 [FREE Full text] [doi: 10.1093/arclin/acab053] [Medline: 34651648]
11. Kabat-Zinn J. An outpatient program in behavioral medicine for chronic pain patients based on the practice of mindfulness meditation: theoretical considerations and preliminary results. Gen Hosp Psychiatry 1982 Apr;4(1):33-47. [doi: 10.1016/0163-8343(82)90026-3] [Medline: 7042457]
12. Bishop SR, Lau M, Shapiro S, Carlson L, Anderson ND, Carmody J, et al. Mindfulness: a proposed operational definition. Clinical Psychology: Science and Practice 2004 Aug 01;11(3):230-241. [doi: 10.1093/clipsy/bph077]
13. Bondolfi G, Jermann F, Zermatten A. Les approaches psychothérapeutiques basées sur la pleine conscience (mindfulness); Entre vogue médiatique et applications cliniques fondées sur des preuves. Psychothérapies 2011;3(31):167-174. [doi: 10.3917/psys.113.0167]
14. Landry L. Entraînement à la pleine conscience via la réalité virtuelle immersive auprès de personnes ayant des symptômes post-commotionnels persistants suite à un traumatisme craniocérébral léger : une étude de cas. Université de Montréal, Papyrus. 2020. URL: https://umontreal.scholaris.ca/items/0ed990d9-072c-4288-988e-705f70e16711 [accessed 2025-03-24]
15. Wang J, Cao B, Yu P, Sun L, Bao W, Zhu X. Deep Learning towards Mobile Applications. 2018 Presented at: 38th International Conference on Distributed Computing Systems (ICDCS); July 02-06, 2018; Vienna, Austria. [doi: 10.1109/icdcs.2018.00139]
16. Schultchen D, Terhorst Y, Holderied T, Stach M, Messner E, Baumeister H, et al. Stay present with your phone: a systematic review and standardized rating of mindfulness apps in European app stores. Int J Behav Med 2021 Oct;28(5):552-560 [FREE Full text] [doi: 10.1007/s12529-020-09944-y] [Medline: 33215348]
17. Spijkerman MPJ, Pots WTM, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: A review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
18. Jayewardene WP, Lohrmann DK, Erbe RG, Torabi MR. Effects of preventive online mindfulness interventions on stress and mindfulness: A meta-analysis of randomized controlled trials. Prev Med Rep 2017 Mar;5:150-159 [FREE Full text] [doi: 10.1016/j.pmedr.2016.11.013] [Medline: 28050336]
19. Gál É, tefan S, Cristea IA. The efficacy of mindfulness meditation apps in enhancing users' well-being and mental health related outcomes: a meta-analysis of randomized controlled trials. J Affect Disord 2021 Jan 15;279:131-142. [doi: 10.1016/j.jad.2020.09.134] [Medline: 33049431]
20. Compen F, Bisseling E, Schellekens M, Donders R, Carlson L, van der Lee M, et al. Face-to-face and internet-based mindfulness-based cognitive therapy compared with treatment as usual in reducing psychological distress in patients with cancer: a multicenter randomized controlled trial. JCO 2018 Aug 10;36(23):2413-2421. [doi: 10.1200/jco.2017.76.5669]
21. Segal ZV, Williams JMG, Teasdale JD. Mindfulness-based cognitive therapy for depression: A new approach to preventing relapse. New York, NY: Guilford Press; 2002.
22. East-Richard C, Laplante L, Vézina J, Cellard C. L'évaluation psychologique par le biais des applications mobiles. Canadian Psychology / Psychologie canadienne 2018 Aug;59(3):262-271. [doi: 10.1037/cap0000113]

XSL•FO
RenderX

23. Bang M, Jang CW, Kim HS, Park JH, Cho HE. Mobile applications for cognitive training: Content analysis and quality review. Internet Interv 2023 Sep;33:100632 [FREE Full text] [doi: 10.1016/j.invent.2023.100632] [Medline: 37312799]
24. Stec MA, Arbour MW, Hines HF. Client-centered mobile health care applications: using the mobile application rating scale instrument for evidence-based evaluation. J Midwifery Womens Health 2019 May;64(3):324-329. [doi: 10.1111/jmwh.12941] [Medline: 30887711]
25. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR Mhealth Uhealth 2015 Aug 19;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]
26. Christopher E, Alsaffarini KW, Jamjoom AA. Mobile health for traumatic brain injury: a systematic review of the literature and mobile application market. Cureus 2019 Jul 10;11(7):e5120 [FREE Full text] [doi: 10.7759/cureus.5120] [Medline: 31523551]
27. Czajkowski SM, Powell LH, Adler N, Naar-King S, Reynolds KD, Hunter CM, et al. From ideas to efficacy: The ORBIT model for developing behavioral treatments for chronic diseases. Health Psychol 2015 Oct;34(10):971-982 [FREE Full text] [doi: 10.1037/hea0000161] [Medline: 25642841]
28. Laverdière R, Banville F, Jackson PL. A mobile application to train attention processes after traumatic brain injury: Rational, design and feasibility evaluation. Annual Review of CyberTherapy and Telemedicine 2023;21:222-227.
29. Baer RA, Smith GT, Lykins E, Button D, Krietemeyer J, Sauer S, et al. Construct validity of the five facet mindfulness questionnaire in meditating and nonmeditating samples. Assessment 2008 Sep;15(3):329-342. [doi: 10.1177/1073191107313003] [Medline: 18310597]
30. Vogt DS, King DW, King LA. Focus groups in psychological assessment: enhancing content validity by consulting members of the target population. Psychol Assess 2004 Sep;16(3):231-243. [doi: 10.1037/1040-3590.16.3.231] [Medline: 15456379]
31. Anastasi A. Introduction à la psychométrie. Montréal: Guérin Universitaire; 1994.
32. Bernaud JL. Introduction à la psychométrie. Paris: Dunod; 2007.
33. Laveault D, Grégoire J. Introduction aux théories des tests en psychologie et en sciences de l'éducation. Bruxelles: de Boeck Université; 2014.
34. Haynes SN, Richard DCS, Kubany ES. Content validity in psychological assessment: A functional approach to concepts and methods. Psychological Assessment 1995 Sep;7(3):238-247. [doi: 10.1037/1040-3590.7.3.238]
35. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health 2007 Aug;30(4):459-467. [doi: 10.1002/nur.20199] [Medline: 17654487]
36. Harris D, Wilson M, Vine S. Development and validation of a simulation workload measure: the simulation task load index (SIM-TLX). Virtual Reality 2019 Dec 21;24(4):557-566. [doi: 10.1007/s10055-019-00422-9]
37. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
38. Saliasi I, Martinon P, Darlington E, Smentek C, Tardivo D, Bourgeois D, et al. Promoting health via mHealth applications using a French version of the mobile app rating scale: adaptation and validation study. JMIR Mhealth Uhealth 2021 Aug 31;9(8):e30480 [FREE Full text] [doi: 10.2196/30480] [Medline: 34463623]
39. Hassenzahl M, Burmester M, Koller F. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Szwillus G, Ziegler J, editors. Mensch & Computer 2003. Berichte des German Chapter of the ACM, vol 57. Stuttgart, Germany: Vieweg+Teubner Verlag; 2003:187-196.
40. Lallemand C, Koenig V, Gronier G, Martin R. Création et validation d'une version française du questionnaire AttrakDiff pour l'évaluation de l'expérience utilisateur des systèmes interactifs. European Review of Applied Psychology 2015 Sep;65(5):239-252. [doi: 10.1016/j.erap.2015.08.002]
41. Terhorst Y, Rathner E, Baumeister H, Sander L. «Hilfe aus dem App-Store?»: Eine systematische Übersichtsarbeit und Evaluation von Apps zur Anwendung bei Depressionen. Verhaltenstherapie 2018 May 8;28(2):101-112. [doi: 10.1159/000481692]
42. Hill NL, Mogle J, Colancecco E, Dick R, Hannan J, Lin FV. Feasibility study of an attention training application for older adults. Int J Older People Nurs 2015 Sep 15;10(3):241-249. [doi: 10.1111/opn.12092] [Medline: 26073545]
43. Hill NL, Mogle J, Wion R, Kitt-Lewis E, Hannan J, Dick R, et al. App-based attention training: Incorporating older adults' feedback to facilitate home-based use. Int J Older People Nurs 2018 Mar;13(1):12163 [FREE Full text] [doi: 10.1111/opn.12163] [Medline: 28752612]
44. Jaeggi SM, Buschkuehl M, Jonides J, Perrig WJ. Improving fluid intelligence with training on working memory. Proc Natl Acad Sci U S A 2008 May 13;105(19):6829-6833 [FREE Full text] [doi: 10.1073/pnas.0801268105] [Medline: 18443283]
45. Chin JP, Diehl VA, Norman KL. Development of an instrument measuring user satisfaction of the human-computer interface. In: CHI '88: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1988 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 15-19, 1988; Washington, DC, USA. [doi: 10.1145/57167.57203]
46. Gordon KE, Kuhle S. 'Reported concussion' time trends within two national health surveys over two decades. Brain Inj 2018;32(7):843-849. [doi: 10.1080/02699052.2018.1463105] [Medline: 29652525]

XSL•FO
RenderX

47. Champagne AS, Yao X, McFaull SR, Saxena S, Gordon KR, Babul S, et al. Commotions cérébrales autodéclarées au Canada : une étude transversale. Statistique Canada. 2023. URL: https://www150.statcan.gc.ca/n1/pub/82-003-x/2023006/article/00002-fra.htm [accessed 2025-03-24]

Abbreviations

ATA: attention training application
CVI: content validity index
MARS-F: Mobile Application Rating Scale, French version
MBCT: mindfulness-based cognitive therapy
MBSR: mindfulness-based stress reduction
ORBIT: Obesity-Related Behavioral Intervention Trials
SIM-TLX: Simulation Task Load Index
TBI: traumatic brain injury

Edited by A Mavragani; submitted 10.07.24; peer-reviewed by S Quaglini; comments to author 15.12.24; revised version received 29.01.25; accepted 04.02.25; published 09.04.25.

Please cite as:
Laverdière R, Jackson PL, Banville F
A New Mobile App to Train Attention Processes in People With Traumatic Brain Injury: Logical and Ecological Content Validation Study
JMIR Form Res 2025;9:e64174
URL: https://formative.jmir.org/2025/1/e64174
doi:10.2196/64174
PMID:40202789

©Roxanne Laverdière, Philip L Jackson, Frédéric Banville. Originally published in JMIR Formative Research (https://formative.jmir.org), 09.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Cost-Effectiveness Analysis of a Machine Learning–Based eHealth System to Predict and Reduce Emergency Department Visits and Unscheduled Hospitalizations of Older People Living at Home: Retrospective Study

Charlotte Havreng-Théry[1,2], MS, PhD; Arnaud Fouchard[3], MD; Fabrice Denis[4], MD, PhD; Jacques-Henri Veyron[2], MS; Joël Belmin[1,5], MD, PhD

[1]Laboratoire Informatique Médicale et Ingénierie des Connaissances en eSanté, Institut National de la Santé et de la Recherche Médicale and Sorbonne Université, Paris, France

[2]PRESAGE, Paris, France

[3]EIR Conseil Santé, Antony, France

[4]Institut Inter-Régional de Cancérologie Jean Bernard, Le Mans, France

[5]Hôpital Charles Foix, Assistance Publique-Hôpitaux de Paris, Ivry-sur-Seine, France

Corresponding Author:
Jacques-Henri Veyron, MS
PRESAGE
112-114 rue la Boétie
Paris, 75008
France
Phone: 33 622152004
Email: jhveyron@presage.care

Abstract

Background: Dependent older people or those losing their autonomy are at risk of emergency hospitalization. Digital systems that monitor health remotely could be useful in reducing these visits by detecting worsening health conditions earlier. However, few studies have assessed the medico-economic impact of these systems, particularly for older people.

Objective: The objective of this study was to compare the clinical and economic impacts of an eHealth device in real life compared with the usual monitoring of older people living at home.

Methods: This study was a comparative, retrospective, and controlled trial on data collected between May 31, 2021, and May 31, 2022, in one health care and home nursing center located in Brittany, France. Participants had to be aged >75 years, living at home, and receiving assistance from the home care service for at least 1 month. We implemented among the intervention group an eHealth system that produces an alert for a high risk of emergency department visits or hospitalizations. After each home visit, the home care aides completed a questionnaire on participants' functional status using a smartphone app, and the information was processed in real time by a previously developed machine learning algorithm that identifies patients at risk of an emergency visit within 7 to 14 days. In the case of predicted risk, the eHealth system alerted a coordinating nurse who could then inform the family carer and the patient's nurses or general practitioner.

Results: A total of 120 patients were included in the study, with 60 in the control group and 60 in the intervention group. Among the 726 visits from the intervention group that were not followed by an alert, only 4 (0.6%) resulted in hospitalizations (P<.001), confirming the relevance of the system's alerts. Over the course of the study, 37 hospitalizations were recorded for 25 (20.8%) of the 120 patients. Additionally, of the 120 patients, 9 (7.5%) were admitted to a nursing home, and 7 (5.8%) died. Patients in the intervention group (56/60, 93%) remained at home significantly more often than those in the control group (48/60, 80%; P=.03). The total cost of primary care and hospitalization during the study was €167,000 (€1=US $1.09), with €108,000 (64.81%) attributed to the intervention group (P=.20).

Conclusions: This study presents encouraging results on the impact of a remote medical monitoring system for older adults, demonstrating a reduction in both emergency department visits and hospitalization costs.

Trial Registration: ClinicalTrials.gov NCT05221697; https://clinicaltrials.gov/study/NCT05221697

XSL•FO
RenderX

(JMIR Form Res 2025;9:e63700) doi:10.2196/63700

KEYWORDS

monitoring; older adult; predictive tool; home care aide; emergency department visit; cost-effectiveness; artificial intelligence; electronic health; eHealth; emergency department; unscheduled hospitalization; aging; retrospective study; medico-economic; living at home; nursing home; emergency visit; Brittany; France; machine learning; remote monitoring; digital health; health informatics

Introduction

Older Adults and Emergency Hospitalizations

In France, over 13.4 million people are aged 65 years or older, representing 20% of the French population. This proportion has increased by 4% in 20 years [1]. With aging and frailty, the risk of unplanned hospitalizations is increased.

Studies show that in France, older adults living at home represent 13% of the population that is hospitalized at least once a year, compared to 6% for the general population [2,3]. According to medico-economic data, nearly 1.6 million people aged >80 years were hospitalized in 2017, twice the number in the general population. In addition, the average length of stay doubles between the ages of 60 and 80 years, reaching nearly 10 days for those aged >90 years [4].

Depending on the study, between 19% and 67% of hospitalizations of older people are considered avoidable [5,6].

Medical and social frailty and comorbidity of the older population are the main factors that increase the prevalence and time spent in hospital [3].

The emergency department (ED) is the main entry point for the older people. Half of the people who visit the ED are subsequently hospitalized. Moreover, older adults represent 41% of ED visits, and nearly 23% of them spend more than 8 hours there and 34% spend between 4 and 8 hours—more than twice as long as younger people. This longer stay in the ED is compounded by a higher rate of transfer to other departments or hospitals [3,7].

In people aged >80 years, main causes of admission to the ED are traumatic events (generally related to fall; 25%), cardiovascular events (17%), altered general condition or infection (12%), a respiratory symptom (12%), a gastrointestinal symptom (10%), and neurological symptoms (9%) [3,7].

Emergency hospitalization represents a significant medical and economic cost. Older adults' hospitalizations reduce the level of dependency by 30% to 60% (immobilization syndrome, confusion, malnutrition, urinary incontinence, undesirable effects related to treatments, etc). While people aged >75 years represent 8% of the population, they are responsible for one-fifth of health care expenditure [8,9]. In France, the main expenses to support the loss of autonomy of older adults are approximately €12.2 billion (€1=US $1.09) in care expenses and €10.7 billion in human and technical assistance. The challenge is not so much to reduce hospitalization rates or the average length of stay but rather to avoid preventable emergency hospitalizations [1,2,8,10,11].

Telehealth and Artificial Intelligence for the Prediction and Prevention of Emergency Situations in Older Adults

Therefore, it is crucial to anticipate adverse events at home. The use of eHealth systems, telemedicine, and connected objects represent a promising approach for keeping older adults at home and preventing the loss of autonomy. Patient-reported outcome measures benefit people with chronic diseases by improving their quality of life, reducing mortality, and reducing ED visits and hospitalizations [7,12].

We have implemented a machine learning–based eHealth device to predict and prevent ED visits and unscheduled hospitalizations for older people living at home [13,14]. In 2022, we conducted a multicenter study with people aged >65 years, living at home, and receiving regular visits of home care aides (HCAs) to evaluate this device in real life. After each home visit, HCAs completed a smartphone-based questionnaire on the functional status of the patients. The information was processed in real time by a machine learning algorithm developed beforehand in order to identify patients at risk of an ED visit within 7 to 14 days (with a predictive performance of 83% sensitivity and 86% specificity). This machine learning algorithm can also predict symptoms or events such as malnutrition, falls, swollen legs, or depressions [15]. In the case of predicted risk, the system alerted a nurse coordinator who could then inform the family caregiver and the patient's nurses or general practitioner. A total of 206 patients were included and followed for 10 months. Compliance was good. Among the 2656 home visits, 405 alerts were issued. The system significantly reduced the number of ED visits when an intervention was performed after an alert. The system was considered easy to use and useful, and it was well accepted by more than 90% of the HCAs and coordinating nurses [15].

This algorithm opens the possibility of mobilizing health professionals to intervene early in an acute illness or in the decompensation of a chronic illness before it leads to an emergency hospitalization.

Objectives

The main hypothesis of this study was as follows: using a simple tool to predict the risk can allow HCAs to anticipate the incidence of emergency hospitalizations and, thereby, reduce the cost and side effects of emergency hospitalization.

The objective of this study was to analyze the clinical and economic impacts of this eHealth device in real life compared to the usual monitoring of frail older people living at home. In France, as this medical device is the first to predict ED use, there are, to our knowledge, few medico-economic studies available.

XSL•FO
RenderX

Methods

Study Design

This study was a single-center, retrospective, and controlled trial on data collected between May 31, 2021, and May 31, 2022, in one health care and home nursing center located in Brittany, France.

Recruitment

Participants were recruited from among adults aged 75 years and older who were living at home and receiving the assistance from HCAs. These HCAs were not health care professionals and typically provided assistance with nonmedical tasks (eg, helping with meals, assisting with personal care, housekeeping, and running errands). The dependency levels of the persons were established according to the French national instrument, which stratifies the dependency level from *groupe iso-ressources* (GIR) 1 (very severe dependency) to GIR 6 (no dependency) [16].

Patients in palliative care situations (GIR 1) were not included. Patients were proposed to be included in the controlled group without being offered any incentives. Written consent was obtained from all participants of the intervention group to be included in the study.

The control group was composed of people receiving the usual support from the home care service. The adjustment between intervention and control groups (1:1 or 1:2 when possible) was based on age (within 2 years), gender, and level of dependency. Participants in the control group were informed of their enrollment in the study. Data were collected between May 30, 2021, and May 31, 2022.

Data Collection

Sociodemographic and pathway data—family situation, dependency level, ED visits (dates and causes), hospitalization (dates and causes), and death (date)—were collected via the tracking system (for the intervention arm) and the coordinating nurses (for the control arm).

Intervention

For participants in the intervention group, HCAs used a smartphone app, for which they received training on its use (a 1-hour session at the start of the study). Each week, they completed a simple and concise questionnaire for each patient via the app. This questionnaire comprised 25 yes-or-no questions covering functional and clinical autonomy (activities of daily living); medical symptoms such as fatigue, falls, pain, and undernutrition; as well as behavioral changes (cognitive disorders and aggression), communication with caregivers, and social life.

The collected data were transmitted in real time to a secure server, where an artificial intelligence algorithm analyzed them to assess risk levels and predict ED visits within 7 or 14 days. The risk assessment was displayed on a Conformité Européenne–marked, web-based secure medical platform called PRESAGE CARE. When the algorithm detected a high-risk level, an on-screen alert was sent to the nurse coordinator of the home care center. This alert included information on recent changes in the patient's functional status, along with decision-support insights to help the nurse coordinator take appropriate action, such as contacting the family caregiver or other health care professionals, or conducting a home reassessment.

No specific intervention protocol was imposed on health care professionals, allowing them full autonomy in their decision-making. This alert-based intervention model was presented to and approved by the *Agence du Numérique en Santé* (National Agency for eHealth). The study followed the guidelines and recommendations of the *Haute Autorité de Santé* (French Health Authority) regarding the cost-effectiveness evaluation of medical devices.

Outcomes

The primary outcome was the difference in the cumulative incidence of emergency hospitalizations between the two groups, with death and institutionalization considered competing events. Unscheduled hospitalizations were defined as those occurring after an ED visit or those explicitly recorded as such in the PRESAGE CARE console or the data extracted from the center's information system.

Secondary outcomes included a comparison of hospitalization rates following an alert that did not lead to an intervention versus those following an alert that triggered an intervention, as well as a cost comparison of hospitalizations between the two groups. Additionally, the probability of remaining at home was assessed, defined as the time elapsed from May 30, 2021 (study start date), to the first occurrence of one of the following events: unplanned hospitalization, death, or institutionalization.

Cost Analysis

For each group, the hospital costs attributable to the ED visits and hospitalizations (unplanned and planned) were calculated on the basis of the homogeneous group of patients, determined by the declared causes of the hospitalization (CIM-10 PMSI coding, which is the French adaptation of US diagnosis-related groups). Hospital costs have been determined on the basis of published costs by the type of illness and length of hospital stay (basic category of classification in medicine, obstetrics, or surgery). The costs of consultations with the attending physician related to the PRESAGE CARE system were also included. The average length of hospital stay was compared between the groups. The burden costs (gross salary and taxes) of nurses who managed alerts (intervention arm) and conducted regular activities (control arm) were defined according to the time needed to proceed their tasks and average gross salary for nurses [17].

The cost for data collection by HCAs was defined as training costs (1 hour of training time multiplied by the hourly burden costs for HCAs [17]). The collection of data into PRESAGE Care eHealth device occurred during regular visits so no extra time was counted. In this study, the medical device was deployed free of charge.

The analysis was performed from the point of view of the total hospital cost per event and per patient in relation to the medical

XSL•FO
RenderX

benefit determined by the number of deaths, the number of institutionalizations, and the data in the literature concerning the loss of functional independence following hospitalization (in the absence of data on the reassessment of the dependency level after hospitalization). The cost analysis was mainly carried out from the point of view of social security expenditure.

Coordinating nurses were asked to report time spend in order to check alerts and act according to the procedure. The number of ED visits, emergency hospitalizations, deaths, and nursing home admissions were compared between the groups. The times before the first event were compared between groups.

Statistical Analysis

Continuous variables were described in terms of means, medians, IQRs, and SDs, depending on the normality of the distribution. Categorical variables were described in terms of numbers and percentages. Groups were compared using the chi-square or Fisher test for categorical variables and the Wilcoxon or 2-tailed *t* test for quantitative variables. Comparisons of cumulative incidences were made by the Gray test (entry into an institution and death in competition). A value of $P<.05$ was considered significant. Statistical analyses were performed using Stata software (version 17; StataCorp LLC).

Ethical Considerations

The research protocol for this study was submitted to and approved by the French National Committee for Biomedical Research, the Committee for the Protection of Individuals, and the French Agency for Health Product Safety (registration: 2234275). The beneficiaries as well as the caregivers and professionals involved were informed of the nature of this study and gave their written consent. Privacy and data protection rules were presented to participants and are available upon request. No compensation was offered.

Results

Patients' Characteristics

A total of 120 patients were included in the study, with 60 (50%) in the control group and 60 (50%) in the intervention group. Among them, 86 (71.7%) were women, including 45 (75%) of the 60 patients in the intervention group and 41 (68%) of the 60 patients in the control group ($P=.76$). The mean age of participants was 83.77 years, with no significant difference between the groups ($P=.44$).

Dependency levels, assessed using the *autonomie gérontologique GIR* grid, showed an average dependency score of 3.79, with no significant differences between the groups (all $P>.05$; Table 1). Overall, the characteristics of participants in the intervention and control groups were comparable.

Table 1. Participants' characteristics.

Characteristics	Intervention group (n=60)	Control group (n=60)	*P* value
Age (years), mean (95% CI)	84.55 (81.78-87.32)	82.99 (80.08- 85.90)	.44
Female, n (%)	45 (75)	41 (68)	.42
Low dependency level (GIR[a] 5-6), n (%)	14 (23)	10 (17)	.36
Mild dependency level (GIR 3-4), n (%)	37 (62)	37 (62)	>.99
High dependency level (GIR 2), n (%)	9 (15)	8 (13)	.79

[a]GIR: *groupes iso-ressources*.

In the intervention arm, a total of 792 visits were monitored through the app. Among these visits, 66 alerts were issued and transmitted to a coordinating nurse, of which 21 (32%) led to a health intervention (such as a general practitioner visit or a home visit by a nurse).

Following an alert-triggered intervention, only 1 (5%) out of the 21 patients was hospitalized, compared to 10 (91%) out of the 11 patients when no intervention was performed after an alert ($P<.001$). Among the 726 visits that did not trigger an alert, 4 (0.6%) hospitalizations occurred ($P<.001$; Table 2).

Table 2. Alert-triggered interventions and hospitalizations.

	Total	No hospitalization	Hospitalizations	*P* value
Alerts (n=66)				<.001
No alert-triggered intervention, n (%)	45 (68)	35 (78)	10 (22)	
Alert-triggered intervention, n (%)	21 (33)	20 (95)	1 (5)	
No alerts (n=726), n (%)	726 (100)	722 (99.4)	4 (0.6)	<.001

Medical impact

Altered general condition (7/37, 19%) and falls with fractures (7/37, 19%) were the most frequently described cause, followed by dyspnea (6/37, 16%). The causes are relatively distributed (Table 3).

XSL•FO
RenderX

Table 3. Unplanned hospitalizations causes (n=37).

Causes	Total, n (%)	Intervention group, n (%)	Control group, n (%)
Altered general condition	7 (19)	1 (3)	6 (16)
Dyspnea	6 (16)	2 (5)	4 (11)
Fall with fracture	7 (19)	4 (10)	3 (8)
Emergency surgery	5 (14)	2 (5)	3 (8)
Acute pain	4 (11)	2 (5)	2 (5)
Psychological symptoms	2 (5)	0 (0)	2 (5)
Kidney failure	2 (5)	1 (3)	1 (3)
Heart failure	1 (3)	0 (0)	1 (3)
Stroke	1 (3)	1 (3)	0 (0)
Fall without fracture	1 (3)	1 (3)	0 (0)
Missing value	1 (3)	1 (3)	0 (0)
Total	37 (100)	15 (41)	22 (59)

Emergency Hospitalizations, Deaths, and Nursing Home Admissions

A total of 37 emergency hospitalizations were recorded during the study, involving 25 patients. The number of emergency hospitalizations was 32% lower in the intervention group (15/37, 41%) compared to the control group (22/37, 59%).

Of the 6 scheduled hospitalizations, 5 occurred in the intervention group. The cumulative incidence of emergency hospitalizations was 23.3% (95% CI 13.5%-34.7%) in the control group, compared to 16.7% (95% CI 8.5%-27.2%) in the intervention group (*P*=.37).

Among participants, 16 patients left their homes due to either admission to a nursing home (9/120, 7.5%) or death (7/120, 5.8%). There were 7 deaths: 1 (14%) in the intervention group and 6 (86%) in the control group. Nine patients were admitted to a nursing home, including 3 (33%) in the intervention group and 6 (67%) in the control group.

The proportion of patients who left home was significantly higher in the control group (12/16, 75%) than in the intervention group (4/16, 25%; *P*=.03; Table 4).

Table 4. Health events during the study.

Health events	Total (N=120)	Intervention group, (n=60)	Control group, (n=60)	Difference between groups (%)	*P* value
In terms of the number of events (n=37), n (%)					
Emergency hospitalizations	37 (100)	15 (40.5)	22 (59.5)	–32	.16
In terms of the number of patients, n (%)					
Emergency hospitalizations	25 (20.8)	10 (16.7)	15 (25)	–33	.26
Deaths	7 (5.8)	1 (1.7)	6 (10)	–67	.05
Entry into nursing homes or facilities	9 (7.5)	3 (5)	6 (10)	–50	.30
Patients who left home	16 (86.7)	4 (6.7)	12 (20)	–67	.03

Probability of Staying at Home

The survival analysis for the probability to stay at home was 68% (95% CI 54.5%-78.3%) for the control group versus 80% (95% CI 67.5%-88.1%) for the intervention group (*P*=.15).

Costs

Cost of Follow-Up

Eleven individuals benefited from a consultation with a general practitioner following an alert issued by the PRESAGE CARE system. These consultations were considered additional costs attributable to the use of the device. The time required for nurses to monitor patients in the intervention arm totaled 119 hours (equivalent to 2 minutes and 40 seconds per patient per week), resulting in a total burden cost of €3460 [16]. Training for HCAs was conducted in a group setting over 1 hour, representing a cost of €1334 for 30 HCAs. Training for the coordinating nurse lasted 5 hours, with an associated cost of €200. The total cost related to training time and device usage was estimated at €4994.

Hospital Costs

All causes of hospitalization and their durations were recorded and categorized according to the relevant homogeneous group of patients. The median length of stay per hospitalization was 9.8 (IQR 4.73-14.87) days in the intervention group, compared to 13.61 (IQR 6.70-20.52) days in the control group,

XSL•FO
RenderX

representing a 67% difference. Over the course of the year, the total number of hospitalization days was 6 times higher in the control group than in the intervention group, with 882 days versus 147 days, respectively. The annual cost attributed to emergency hospitalizations, based on data from the homogeneous group of patients, was €58,975 in the intervention group compared to €108,642 in the control group, reflecting a 45.72% reduction in the intervention group (Figure 1). The total cost for the intervention group was €64,300, compared to €108,642 in the control group, representing a 40.82% cost reduction (Table 5).

Figure 1. Emergency hospitalizations' costs evolution (cumutive cost, 1€ = US $ 1.09).

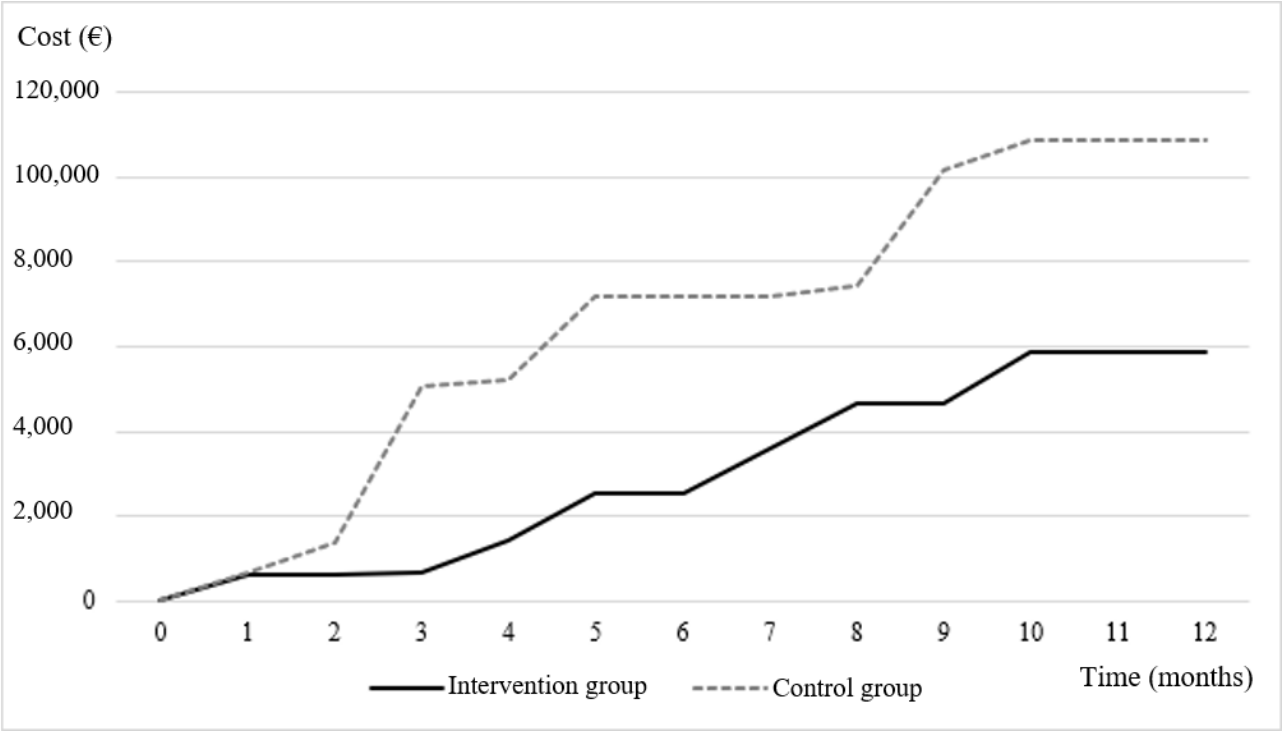


Table 5. Costs analyses.

	Intervention group	Control group	Difference between groups (%)	*P* value
Length of stay (days), median (IQR)	9.8 (4.73-14.87)	13.61 (6.70-20.52)	–28	.41
Total number of days of emergency hospitalizations, n	147	313	–53	.09
Annual emergency hospitalization costs (€)[a]	58,976	108,641	–45	.20
Cost for additional consultations with the attending physician (€)	330	—[b]	—	
Cost for data collection (training cost for HCAs[c]; €)	1334	—	—	
Cost for coordinating nurse (follow-up and alert management; €)	3460	—	—	
Cost for coordinating nurse (training cost; €)	200	—	—	
Total annual cost (€)	64,300	108,641	–41	.25
Mean cost by patient (€)	1071	1810	—	

[a]€1=US $1.09.

[b]Not applicable.

[c]HCA: home care aide.

Discussion

Principal Findings

The objective of this study was to assess the medico-economic impact of an eHealth device designed to predict and prevent emergency hospitalizations among frail older adults living at home.

The use of the PRESAGE CARE device led to a nearly 32% reduction in emergency hospitalizations compared to the control group. Overall, it reduced hospital-related costs by 45.72%. In

XSL•FO
RenderX

addition, this cost reduction must be considered in the context of additional expenses, including in-home care, posthospitalization consultations, and the increased risk of rehospitalization.

The median length of stay in the intervention group was consistent with findings in the literature, whereas it was higher in the control group, likely due to extreme values and the small cohort size.

Furthermore, over a 1-year follow-up of 60 participants, the system extended the mean time to failure in the care pathway by 38 days. This result is likely underestimated given the limited cohort size.

In the intervention arm, among the 726 visits not followed by an alert, only 4 (0.6%) hospitalizations followed the visit (P<.001), which confirm the relevance of the alerts issued by the system. Comparison of the postintervention hospitalization rate in the intervention group shows that an alert-triggered intervention significantly reduces hospitalization. These results are consistent with a previous study on this system [15]. Professionals can effectively anticipate a deteriorating health condition and act at the right time.

The multiplication of technological tools to promote home care rarely benefit from data on their real medico-economic benefits [17,18]. Moreover, few studies on the economic evaluation of telemedicine systems show a medico-economic gain [19-22]. In addition, to our knowledge, they have not evaluated the effectiveness of a predictive system for the prevention of unplanned hospitalizations of older people living at home.

Cost of the Device

The cost of the device has not been taken into account in this study as it is currently being evaluated in terms of its impact. The price of the health care sector (including hospital care) varies widely from one country to another. France remains 40% cheaper than the United States and 55% cheaper than Switzerland [23].

Increased Cost of Dependency and Quality of Life

According to the report of the Hospital and Older People Workshop published in 2018, the incidence of dependence related to hospitalization varies between 30% and 60% in people aged 70 years and older and increases to 50% in patients aged 85 years and older. After discharge from hospital, only 50% of patients recover their basic functional state, 33% within 6 months after discharge, and 14% at 1 year [10,19]. The prevalence of iatrogenic dependence is about 12%, and it is avoidable in 80% of cases. It is therefore possible to prevent it.

Moreover, after a visit to the ED, 56% of older patients are hospitalized, and 44% return home. The difficulty in finding suitable hospital places for older adults leads to an increase in the time spent by carers on calls to other departments. Older people spend more time in the ED, on average 4 hours more [10].

These data increase the medical and financial impact of devices aimed at predicting and preventing ED visits and emergency hospitalizations.

The increase in the loss of autonomy of older adults after hospitalization (bed rest, loss of mobility, and waiting time in the ED) leads to an increase in home care needs and increases the costs associated with the management of dependency.

This increase in posthospitalization dependence leads to a decrease in quality of life, which has been widely described in the literature [24].

This device represents a real gain in terms of medical benefits, quality of life, and hospital costs.

Limitations

Our study has several limitations. First, because the overall study design was not randomized at the outset of the study, there is a bias in the allocation of groups. Moreover, the number of deaths was higher in the control group, but the sample size did not allow us to identify a significant correlation with the intervention. The lack of significance of the different results can indeed be explained by the size of the cohort, the number of events, and the low statistical power. Some particularly long hospital stays imply a wide CI, impacting the significance of the results.

In view of the number of deaths and the lack of consensus on the relevance of using quality-adjusted life-year analyses to carry out medico-economic studies of dependent or frail older people (National Academy of Medicine) [21,24], it was not possible to carry out a quality-adjusted life-year study based on life expectancy.

Furthermore, the quality of life of the individuals could not be measured and compared. It is not possible to draw conclusions on the impact of PRESAGE CARE on the quality of life of beneficiaries. A prospective study will be conducted to evaluate this impact.

The size of the cohort is also a limitation and will require a larger-scale, multicenter, randomized controlled trial.

Conclusion

This first study evaluating the medico-economic impact in real life of the use of the PRESAGE CARE medical device shows a reduction of almost half of the hospital expenses among the beneficiaries of the device. In addition, PRESAGE CARE reduced unplanned hospitalizations by more than 30% and increased life expectancy in good health at home. A larger, randomized controlled trial is needed to confirm these already very encouraging results.

Acknowledgments

The authors would like to thank the care managers and staff of the home care facility (Le Connétable) for their contribution to the study, as well as GHT Rance-Emeraude and Regional Agency for Health (ARS) for funding this study.

XSL•FO
RenderX

Data Availability

The data that support the findings of this study are available from the corresponding author, JHV, upon reasonable request. Restrictions apply to the availability of the data collected by the home care aids and the algorithm that was used in this study, which is the subject of a patent.

Authors' Contributions

Each author met the International Committee of Medical Journal Editors criteria for authorship. JHV conceptualized the study; CHT conducted the acquisition of data; CHT and JHV contributed to the interpretation of the results; JHV and CHT were in charge of the formal analysis; and JB, JHV, CHT, FD, and AF drafted and revised the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

JHV is the chief executive officer of PRESAGE CARE. CH-T is employee of PRESAGE.

References

1. Larbi K, Roy D. 4 Millions de seniors seraient en perte d'autonomie en 2050. 4 Million seniors will be losing their autonomy in 2050 [Article in French]. Insee Première. 2019 Jul 25. URL: https://www.insee.fr/fr/statistiques/4196949 [accessed 2025-03-10]
2. Uscher-Pines L, Pines J, Kellermann A, Gillen E, Mehrotra A. Emergency department visits for nonurgent conditions: systematic literature review. Am J Manag Care 2013 Jan;19(1):47-59 [FREE Full text] [Medline: 23379744]
3. Boisguérin B, Mauro L. Les personnes âgées aux urgences : une patientèle au profil particulier. Older people in the emergency department: a patient base with a particular profile [Article in French]. Direction de la Recherches, des Etudes, de l'Evaluation des Statistiques (DREES). 2017 Mar 24. URL: https://drees.solidarites-sante.gouv.fr/publications/etudes-et-resultats/les-personnes-agees-aux-urgences-une-patientele-au-profil [accessed 2025-03-10]
4. Duong HV, Herrera LN, Moore JX, Donnelly J, Jacobson KE, Carlson JN, et al. National characteristics of emergency medical services responses for older adults in the United States. Prehosp Emerg Care 2018 Sep 01;22(1):7-14 [FREE Full text] [doi: 10.1080/10903127.2017.1347223] [Medline: 28862480]
5. Prince MJ, Wu F, Guo Y, Gutierrez Robledo LM, O'Donnell M, Sullivan R, et al. The burden of disease in older people and implications for health policy and practice. Lancet 2015 Feb 07;385(9967):549-562. [doi: 10.1016/S0140-6736(14)61347-7] [Medline: 25468153]
6. Analyse de l'activité hospitalière 2018 - Focus sur les patients âgés de 80 ans et +. Hospital activity analysis 2018 - focus on patients aged 80 and over [Article in French]. Agence technique de l'information sur l'hospitalisation. 2018. URL: https://www.atih.sante.fr/analyse-de-l-activite-hospitaliere-2018-focus-sur-les-patients-ages-de-80-ans-et [accessed 2025-03-10]
7. Boisguérin B, Mauro L. Les personnes âgées aux urgences : une santé plus fragile nécessitant une prise en charge plus longue. Elderly people in the emergency room: a more fragile health requiring a longer care [Article in French]. Direction de la Recherches, des Etudes, de l'Evaluation des Statistiques (DREES). 2017 Mar 24. URL: https://drees.solidarites-sante.gouv.fr/publications/etudes-et-resultats/les-personnes-agees-aux-urgences-une-sante-plus-fragile [accessed 2025-03-10]
8. The care of the older people in the health care system: progress too slow. Executive Summaries: Annual Public Report of the Cour des Comptes. Paris, France: Cour des Comptes; 2012. URL: https://www.ccomptes.fr/sites/default/files/EzPublish/Annual_public_report_2012_summaries.pdf [accessed 2025-03-10]
9. Roussel R. Personnes âgées dépendantes : les dépenses de prise en charge pourraient doubler en part de PIB d'ici à 2060. Dependency of older people: care expenditure could double as a share of GDP by 2060 [Article in French]. Direction de la Recherches, des Etudes, de l'Evaluation des Statistiques (DREES). 2017 Oct 12. URL: https://drees.solidarites-sante.gouv.fr/publications/etudes-et-resultats/personnes-agees-dependantes-les-depenses-de-prise-en-charge [accessed 2025-03-10]
10. Libault D. Rapport de la concertation grand âge et autonomie. Report on old age and autonomy [Article in French]. Ministry of Solidarity and Health (France). 2019 Mar 28. URL: https://tinyurl.com/3b4mekcn [accessed 2025-03-10]
11. Adams JG. Emergency department overuse: perceptions and solutions. JAMA 2013 Mar 20;309(11):1173-1174. [doi: 10.1001/jama.2013.2476] [Medline: 23512065]
12. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 2012 Jul 07;380(9836):37-43 [FREE Full text] [doi: 10.1016/S0140-6736(12)60240-2] [Medline: 22579043]
13. Denis F, Krakowski I. How should oncologists choose an electronic patient-reported outcome system for remote monitoring of patients with cancer? J Med Internet Res 2021 Sep 09;23(9):e30549 [FREE Full text] [doi: 10.2196/30549] [Medline: 34499046]
14. Veyron J, Friocourt P, Jeanjean O, Luquel L, Bonifas N, Denis F, et al. Home care aides' observations and machine learning algorithms for the prediction of visits to emergency departments by older community-dwelling individuals receiving home

XSL•FO
RenderX

care assistance: a proof of concept study. PLoS One 2019 Aug 13;14(8):e0220002 [FREE Full text] [doi: 10.1371/journal.pone.0220002] [Medline: 31408458]
15. Belmin J, Villani P, Gay M, Fabries S, Havreng-Théry C, Malvoisin S, et al. Real-world implementation of an eHealth system based on artificial intelligence designed to predict and reduce emergency department visits by older adults: pragmatic trial. J Med Internet Res 2022 Sep 08;24(9):e40387 [FREE Full text] [doi: 10.2196/40387] [Medline: 35921685]
16. Belmin J, Auffray J, Berbezier C, Boirin P, Mercier S, de Reviers B, et al. Level of dependency: a simple marker associated with mortality during the 2003 heatwave among French dependent elderly people living in the community or in institutions. Age Ageing 2007 May 24;36(3):298-303. [doi: 10.1093/ageing/afm026] [Medline: 17384419]
17. Les salaires dans les établissements de santé. Salaries in health establishments [Article in French]. Direction de la Recherches, des Etudes, de l'Evaluation des Statistiques (DREES). 2021. URL: https://tinyurl.com/4wfd2utv [accessed 2025-03-10]
18. Gentili A, Failla G, Melnyk A, Puleo V, Tanna GLD, Ricciardi W, et al. The cost-effectiveness of digital health interventions: a systematic review of the literature. Front Public Health 2022 Aug 11;10:787135 [FREE Full text] [doi: 10.3389/fpubh.2022.787135] [Medline: 36033812]
19. Rohrbach P, Dingemans A, Evers C, van Furth EF, Spinhoven P, Aardoom J, et al. Cost-effectiveness of internet interventions compared with treatment as usual for people with mental disorders: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2023 Jan 05;25:e38204 [FREE Full text] [doi: 10.2196/38204] [Medline: 36602854]
20. Sülz S, van Elten HJ, Askari M, Weggelaar-Jansen AM, Huijsman R. eHealth applications to support independent living of older persons: scoping review of costs and benefits identified in economic evaluations. J Med Internet Res 2021 Mar 09;23(3):e24363 [FREE Full text] [doi: 10.2196/24363] [Medline: 33687335]
21. Lizée T, Basch E, Trémolières P, Voog E, Domont J, Peyraga G, et al. Cost-effectiveness of web-based patient-reported outcome surveillance in patients with lung cancer. J Thorac Oncol 2019 Jun;14(6):1012-1020 [FREE Full text] [doi: 10.1016/j.jtho.2019.02.005] [Medline: 30776447]
22. Health care prices. Organisation for Economic Co-operation and Development (OECD). 2020 May. URL: https://www.oecd.org/content/dam/oecd/en/publications/reports/2020/05/health-care-prices_d4fbbd0a/bfc42344-en.pdf [accessed 2025-03-10]
23. Prévenir la dépendance iatrogène liée à l'hospitalisation chez les personnes âgée. Preventing iatrogenic dependency related to hospitalization in the elderlys [Article in French]. Haute Autorité de Santé. 2017 Sep. URL: https://www.has-sante.fr/jcms/c_2801190/fr/prevenir-la-dependance-iatrogene-liee-a-l-hospitalisation-chez-les-personnes-agees [accessed 2025-03-10]
24. Gonthier R. La qualité de vie chez le sujet âgé. Quality of life in old age [Article in French]. Bull Acad Natl Med 2007 Feb;191(2):237-244. [doi: 10.1016/s0001-4079(19)33071-7] [Medline: 17969545]

Abbreviations

ED: emergency department
GIR: groupe iso-ressources
HCA: home care aide

Edited by A Mavragani; submitted 27.06.24; peer-reviewed by LC Maclagan, A Gangadhara Rao, D Valvi, A Kalluchi; comments to author 16.09.24; revised version received 21.12.24; accepted 05.03.25; published 11.04.25.

Please cite as:
Havreng-Théry C, Fouchard A, Denis F, Veyron JH, Belmin J
Cost-Effectiveness Analysis of a Machine Learning–Based eHealth System to Predict and Reduce Emergency Department Visits and Unscheduled Hospitalizations of Older People Living at Home: Retrospective Study
JMIR Form Res 2025;9:e63700
URL: https://formative.jmir.org/2025/1/e63700
doi:10.2196/63700
PMID:40215100

©Charlotte Havreng-Théry, Arnaud Fouchard, Fabrice Denis, Jacques-Henri Veyron, Joël Belmin. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The EmpkinS-EKSpression Reappraisal Training Augmented With Kinesthesia in Depression: One-Armed Feasibility Study

Marie Keinert[1], MSc; Lena Schindler-Gmelch[1], Dr rer nat; Lydia Helene Rupp[1], MSc; Misha Sadeghi[2], MSc; Robert Richer[2], MSc; Klara Capito[1], MSc; Bjoern M Eskofier[2,3], Prof Dr; Matthias Berking[1], Prof Dr

[1]Department of Clinical Psychology and Psychotherapy, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany

[2]Machine Learning and Data Analytics Lab, Department of Artificial Intelligence in Biomedical Engineering (AIBE), Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany

[3]Translational Digital Health Group, Institute of AI for Health, Helmholtz Zentrum Muenchen, Neuherberg, Germany

Corresponding Author:
Marie Keinert, MSc
Department of Clinical Psychology and Psychotherapy
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
Nägelsbachstraße 25a
Erlangen, 91052
Germany
Phone: 49 9131 8567563
Email: marie.keinert@fau.de

Abstract

Background: Harboring dysfunctional depressogenic cognitions contributes to the development and maintenance of depression. A central goal of cognitive behavioral therapy (CBT) for depression is to invalidate such cognitions via cognitive reappraisal (CR). However, relatively low remission rates and high dropout rates in CBT demonstrate the need for further improvement. Potentially, the effects of CBT could be enhanced by addressing not only dysfunctional depressogenic cognitions but also body states associated with depression. This may be done, for example, by systematically pairing the invalidation of depressogenic cognitions with the performance of antidepressive kinesthesia.

Objective: This study aimed to examine the feasibility and clinical potential of a smartphone-based cognitive restructuring task that required users to deliberately perform antidepressive kinesthesia in conjunction with the rejection of depressogenic statements and the affirmation of antidepressive statements. This feasibility study was conducted as a precursor to a large-scale randomized controlled trial.

Methods: In total, 10 healthy participants engaged in a single 90-120-minute session of smartphone-based CR training. During the training, they completed 2 phases in which they were required to reject 20 depressogenic and affirm 20 antidepressive statements, respectively. Diagnostic assessments were conducted 1 week (T1) before and directly prior (T2) to the training, and again directly posttraining (T3) and at a 2-week follow-up posttraining (T4). Feasibility outcomes assessed at T3 included intervention safety recorded by study therapists, compliance, technical feasibility, usability assessed using the Short Version of the User Experience Questionnaire (UEQ-S), and acceptability assessed using the UEQ-S and self-developed items. Preliminary clinical potential was evaluated via single-item ratings of current depressed and positive mood assessed continuously during the training. Feasibility outcomes were analyzed descriptively, and clinical potential was examined using paired-sample *t* tests of pre and post ratings of mood at each training phase.

Results: Overall, the results indicated that the training was safe, feasible, and usable (UEQ-S pragmatic quality scale: mean 1.45, SD 0.71). However, acceptance was limited (UEQ-S hedonic quality scale: mean 1.05, SD 0.79). While 80% (8/10) of the participants were generally satisfied with the training, 80% (8/10) would recommend it to a friend, 90% (9/10) found it interesting, and 80% (8/10) rated it as “leading edge,” 40% (4/10) to 70% (7/10) did not consider it particularly helpful and 50% (5/10) found it repetitive. Preliminary results regarding clinical potential were promising, with significant increases in positive mood (rejection: Hedges g=0.63; affirmation: Hedges g=0.25), whereas changes in depressed mood were not significant.

Conclusions: This study evaluated the feasibility and acceptability of a smartphone-based CR training augmented with validating and invalidating kinesthesia. This provided valuable insights for further optimizing the intervention for the subsequent randomized controlled trial, but also potential similar interventions. If future studies confirm their clinical potential, such interventions offer a promising approach to enhancing CBT for depression.

XSL•FO
RenderX

Trial Registration: OSF Registries pw6ma; https://osf.io/pw6ma/

(JMIR Form Res 2025;9:e65357) doi:10.2196/65357

KEYWORDS

depression; cognitive reappraisal; facial expression; kinesthesia; smartphone-based intervention; mobile phone

Introduction

Depression is one of the most prevalent mental disorders [1] and severely impacts the quality of life and functioning of affected individuals [2]. Depression is also associated with an increased risk of cardiovascular disease [3] and mortality [4]. Although cognitive behavioral therapy (CBT) has been shown to be effective in depression treatment [5], dropout rates of up to 32 % [6], low remission rates of 42 % [7], and low rates of evidence-based treatment administration [8] highlight the need to further optimize CBT for depression.

CBT for depression is based on Beck's [9] cognitive model of depression, which assumes that dysfunctional beliefs (eg, "I am worthless") are a key factor in the etiology and maintenance of the disease. Therefore, one primary goal of CBT is the reappraisal of such dysfunctional beliefs. This is typically pursued with cognitive restructuring techniques, with which therapists guide patients to question the validity of their dysfunctional beliefs and to articulate more adaptive ones [9]. Theory and empirical evidence suggest that invalidating dysfunctional beliefs while strengthening functional beliefs has positive effects on depressed mood and other symptoms of depression [10,11]. To reduce the large treatment gap in depression, researchers have developed accessible and cost-effective CBT-based app interventions that include cognitive restructuring exercises, among others. Such interventions have proven to be effective in reducing symptoms of depression [12]. However, a meta-analysis evaluating single components of such interventions with individual participant data found no clear evidence for the efficacy of cognitive restructuring when used within digital interventions [13], highlighting the need to optimize digital cognitive restructuring interventions. Most of the mobile cognitive restructuring exercises in the context of depression take a similar approach as therapists in face-to-face therapy. They offer psychoeducational content about the relationship between dysfunctional beliefs and depression and provide Socratic questions to help individuals invalidate dysfunctional beliefs and develop more adaptive ones (eg, [14-19]). However, without the support of a therapist, this approach could be very challenging for patients, especially in the context of cognitive and motivational impairments in depression [20,21]. This is also supported by pilot results indicating lower adherence rates with a cognitive restructuring intervention than with an intervention focusing on behavioral activation [18]. Thus, in order to increase adherence, a structured cognitive restructuring exercise with preselected content might be of greater benefit in the context of app interventions without direct human interaction. To our knowledge, only 2 studies evaluate such an intervention. McCloud et al [22] propose a cognitive restructuring exercise based on the ABC technique [23] providing options for alternative interpretations for users to choose from. Stiles-Shields et al [24] propose a step-by-step procedure for identifying and changing maladaptive thoughts using Socratic questions, providing examples of possible thoughts and alternatives. However, the efficacy of McCloud et al [22] cognitive restructuring exercise cannot be determined as it was part of a multicomponent intervention. Stiles-Shields et al [24] conducted a pilot trial. Although they did find effects on symptoms of depression, the informative value is limited due to the small sample size. The efficacy of such structured cognitive restructuring exercises therefore still requires empirical support.

A potential problem of such exercises, particularly if they contain preselected content, could be that the invalidation of dysfunctional beliefs remains superficial and does not lead to emotional insight. This is also assumed by the interacting cognitive subsystems (ICS) theory [25], another seminal theoretical approach in the context of depression. The theory distinguishes intellectual and emotional beliefs and posits that cognitive restructuring only changes dysfunctional beliefs on an intellectual level, but not their "felt meaning." This is often experienced by patients who report, "I understand the belief is not true, but I still feel that it is."

ICS suggests that depressogenic thoughts and proprioceptive information interact to generate emotional states in general and depressed mood in particular [25]. Semantic concepts and specific meanings of depressogenic thoughts exist in the form of intellectual beliefs. To generate specific emotional states (eg, depressed mood) and thereby emotional beliefs, however, input from acoustic, visual, and proprioceptive channels (eg, facial expressions, gestures, and posture) is needed. Hence, according to ICS, if patients merely question dysfunctional beliefs on an intellectual level, they will still feel that they are true. To change the resulting emotional state (ie, depressed mood), proprioceptive information needs to additionally contradict the depressogenic schema, for example, assume an upright body posture and confident facial expression. Thus, the maintenance of depression is explained in terms of a coupling of cognitive and proprioceptive feedback loops that sustain each other, referred to as "cognitive interlock."

In light of these theoretical considerations, we assume that complementing cognitive reappraisal (CR, a key cognitive restructuring technique) with proprioceptions that both invalidate dysfunctional beliefs and validate functional beliefs offers a promising approach to improving CBT for depression. Kinesthesia (ie, proprioception of movement) and, more specifically, facial expressions might be particularly relevant in this context as facial muscles are very sensitive to variations in emotional experience [26].

XSL•FO
RenderX

Empirically, numerous studies provide evidence for a relationship between facial expressions on the one hand, and emotional experience [27] and depression on the other [28-30]. Patients with depression manifest less facial expression of positive affect [28]. Moreover, studies using electromyography (EMG) found greater activity of the corrugator supercilii muscle, a muscle that is activated during negative affective states [31], in response to affective content pictures [32,33], and determined clinical improvement to be associated with attenuated activity of the corrugator supercilii [34]. Consistently, an experimental manipulation of facial expression (ie, inducing a sad facial expression) correlated with negative interpretation bias in healthy adults [35]. The interconnection of depression and facial expression is also reflected in studies using botulinum toxin to block the corrugator supercilii in the treatment of depression, which led to significant and large reductions in symptoms of depression in placebo-controlled studies, as shown by Schulze et al [36] in a recent meta-analysis.

These findings clearly demonstrate the relationship between kinesthesia and depression. However, only very few studies investigated the relationship between kinesthesia and propositional (ie, reasoning) processes in depression. Moreover, to our knowledge, few studies so far have investigated the effects of augmenting cognitive interventions and, more specifically, CR interventions with validating and invalidating kinesthesia. Adaptations of the approach-avoidance modification training (AAMT), which originally required participants to push or pull joysticks to move disorder-relevant images away from or toward themselves [37], use disorder-relevant dysfunctional beliefs as stimuli [38-45]. Two of the studies following such an approach addressed depression (ie, with stimuli such as “I am a failure”), with promising effects on symptoms of depression [38,39]. Moreover, pilot findings from the stress context suggest that using the deliberate display of positive and negative emotions through facial expression, body posture, and a corresponding statement as responses within AAMT might enhance its efficacy [43-45]. Another, albeit not smartphone-based study by O’Toole and Michalak [46] followed an approach more closely aligned with cognitive restructuring in CBT. They evaluated a cognitive restructuring exercise delivered face-to-face combined with emotion-focused body postures and movements and found that it produced greater decreases in agreement with dysfunctional beliefs than isolated cognitive restructuring in healthy individuals.

Although these approaches are promising, evidence remains preliminary, particularly in the context of depression treatment and no study so far has evaluated the efficacy of a smartphone-based structured cognitive restructuring exercise enhanced with validating and invalidating kinesthesia. To advance research in this regard, we developed a structured smartphone-based and sensor-supported CR intervention, in which depressogenic verbalizations are paired with invalidating facial expressions and antidepressive verbalizations are paired with validating facial expressions. Our ultimate goal is to train machine-learning models in the automated assessment of depression, which could then be integrated into a fully automated, smartphone-based biofeedback training. The primary goal of this was to evaluate the feasibility and safety of this intervention in a limited healthy sample by testing its technical setup and procedures before implementation of a large-scale randomized controlled trial (RCT) involving participants with and without depression. The study protocol for the RCT has been published [47]. Second, we aimed to gain preliminary insights into the clinical potential of the intervention, expecting an overall decrease in depressed mood and an increase in positive mood throughout the training and beyond.

Methods

Design

In preparation for a large-scale RCT [47], we conducted an experimental feasibility study with a repeated-measures, within-subject design. The study is part of the Collaborative Research Center EmpkinS (Empatho-Kinaesthetic Sensor Technology – Sensor Techniques and Data Analysis Methods for Empatho-Kinaesthetic Modeling and Condition Monitoring) and was conducted at the EmpkinS Lab of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). Participants engaged in a 90-120-minute CR training session augmented with antidepressive facial expression (AFE) prompts on a study-provided smartphone (CR+AFE condition, ie, 1 of the 4 study conditions of the bigger RCT [47]). The study protocol did not involve blinding of either participants or investigators. The study was registered in the Open Science Framework [48] and we have adhered to the American Psychological Association journal article reporting standards [49] when writing the manuscript (Multimedia Appendix 1).

Participants

Participants were recruited between August 2022 and January 2023. Inclusion criteria were (1) a minimum age of 18 years and (2) sufficient German language skills (at least level B2 according to the Common European Framework of Reference for Languages). Exclusion criteria were (1) a Patient Health Questionnaire (PHQ)-8 score of ≥4, (2) sufficient criteria for any mental disorder (*International Classification of Diseases and Related Health Problems, Tenth Revision* codes F1-F9), (3) acute suicidality, (4) any impairment of facial expression, and (5) dyschromatopsia. We aimed to include N=10 participants, following heuristics from the literature concerning optimal sample sizes for feasibility studies [50,51] and internal considerations of practicality and time constraints. Sociodemographic and clinical baseline characteristics are summarized in Table 1.

XSL•FO
RenderX

Table 1. Sociodemographic and clinical baseline characteristics of participants (N=10) in the feasibility study investigating the single-session EmpkinS-EKSpression Reappraisal Training against depression in healthy individuals.

Categories	Values
Age (years; range 18-41), mean (SD)	24.00 (6.31)
Gender: female, n (%)	10 (100)
Education, n (%)	
> 10 years	4 (40)
University degree	6 (60)
Occupation, n (%)	
Student	9 (90)
Employed	1 (10)
Relationship status: single	10 (100)
Diagnosis of depression, n (%)	
Former	1 (10)
Current	0 (0)
Diagnosis of other mental disorders, n (%)	
Former	2 (20)
Current	0 (0)
Psychotherapeutic or psychiatric treatment	
Former, n (%)	4 (40)
Current, n (%)	0 (0)
PHQ-8[a], mean (SD)	2.10 (0.99)

[a]PHQ: Patient Health Questionnaire.

Procedures

Participants from the general population were recruited via social media posts targeting the Erlangen area, emails sent via student distribution lists of the FAU, and flyers posted in public places throughout the city of Erlangen. The recruitment material contained information about the conditions for participation, the duration of the study, and the compensation as well as a QR code and link to the study website for registration. The website hosted by Unipark provided information on study aims and procedures. Interested persons were directed to the study screening questionnaire assessing inclusion and exclusion criteria, as well as symptoms of depression with the Patient Health Questionnaire (PHQ)-8 [52] and contact information. Eligible participants were contacted via email or telephone and invited to a diagnostic session (T1). All on-site study sessions were held at the EmpkinS Lab of the FAU. The laboratory rooms were located in a scientific university building and set up specifically for this study. At the beginning of the diagnostic session participants provided written informed consent. We then conducted the Structured Clinical Interview for DSM-5 Disorders-Clinician version (SCID-5-CV; German version: [53]) to assess the diagnostic status of participants. After the interview, participants answered a set of questionnaires via Unipark.com (T1).

One week after the diagnostic session, the training session was held at the Lab. The entire session took about 3 hours. First, we conducted the GRID Hamilton Depression Rating Scale (GRID-HAMD; [54]) to assess depression. Next, participants answered a web-based preassessment (T2), also via Unipark.com and the physiological data acquisition was set up. Then, participants received a short psychoeducation explaining the interconnection of thoughts, mood, and the body. The subsequent smartphone-based training session took approximately 90-120 minutes. After completing the training, a short postassessment (T3) was completed, again via Unipark.com.

Two weeks after the training session, participants were invited to a follow-up assessment (T4). The GRID-HAMD was conducted via telephone and participants were asked to fill out an online questionnaire via Unipark.com. They received the link to the questionnaires via email one day before the telephone interview.

The interviews and study sessions were conducted by trained clinical psychologists and based on manuals outlining study procedures and instructions for participants.

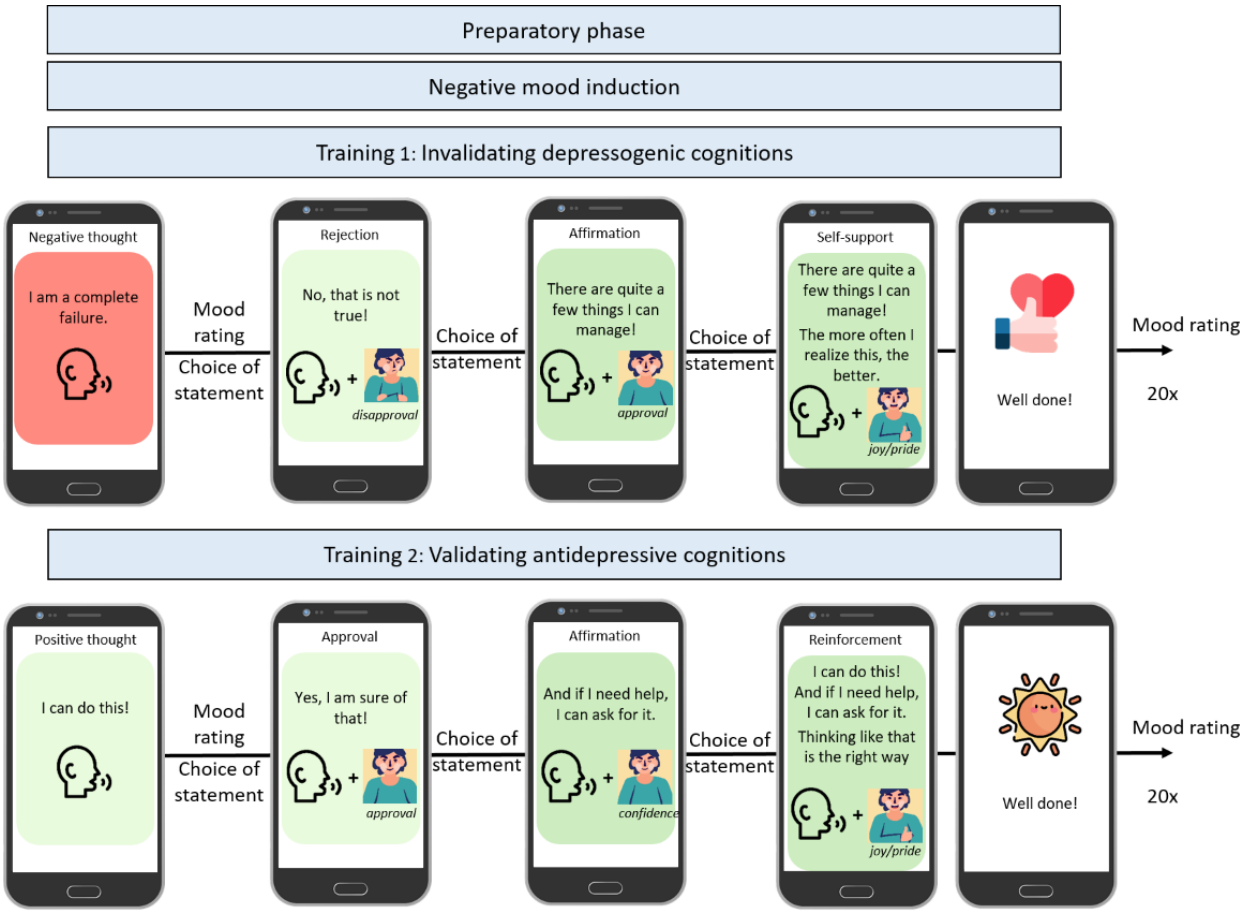
Smartphone Intervention

During the training, participants completed four phases (Figure 1): (1) the preparatory phase, (2) the depressed mood induction phase, (3) the first training phase, and (4) the second training phase. During the preparatory phase, participants were shown a fixation cross (3 sec) followed by different stimuli (8 sec) in randomized order on the smartphone. The stimuli consisted of

XSL•FO
RenderX

self-recordings of the participants, words congruent and incongruent with depressed mood (ie, "hopeless" and "happy"), and words representing the so-called cognitive triad of depression (ie, "me," "the world," and "the future" [55]), which were each shown 4 times, resulting in 24 stimuli in total. Subsequently, in order to allow working with current depressed mood during the training, participants received a validated mood induction beforehand. During this mood induction phase, participants read aloud ten depressogenic cognitions shown on the smartphone screen (eg, "My future is absolutely hopeless"). In addition, sadness-inducing music was played (an excerpt from "Adagio in G minor" by Tomaso Giovanni Albinoni). The induction procedure was previously validated [56,57].

Figure 1. Intervention procedure of the single-session EmpkinS-EKSpression Reappraisal Training against depression was examined in this feasibility study in healthy individuals.



Both training phases started with the study therapist providing instructions on how to engage in the training, followed by on-screen instructions with text and animations in the EmpkinS-EKSpression app. Then, participants completed 2 practice trials before starting the actual training. In the actual training, participants were confronted with 20 depressogenic statements (eg, "I am a complete failure"; first training phase) and antidepressive (eg, "I can do this;" second training phase) statements presented in randomized order on the smartphone screen. The statements were derived from validated questionnaires in the context of depression (Beck Hopelessness Scale, German version [58]; Cognitive Triad Inventory, German version [59]; Automatic Thoughts Questionnaire, German version [60]; and Rosenberg Self-Esteem Scale, German version [61]). A team of 3 experts in clinical psychology adapted the items to customize them for the intervention.

In the first training phase, participants were asked to reject the depressogenic statements in the following three invalidation steps: (1) reject the presented cognition, (2) affirm the rejection, and (3) self-support themselves. In each step, participants were able to choose from three statements (ie, 3 statements expressing rejection, affirmation, and self-support, respectively) which they had to manually select on the screen and also read aloud. Their verbalization was to be amplified by the explicit expression of congruent kinesthesia, including mimic, gesture, and posture. Thus, their rejection of each depressogenic statement was performed with an expression of disapproval, the affirmation with an expression of approval, and self-support with an expression of joy (trials 1-10) and pride (trials 11-20).

In the second training phase, participants received instructions to validate the antidepressive statements in the following sequence: (1) approve the positive cognition with an expression of approval, (2) affirm the approval with an expression of confidence, and (3) reinforce themselves with an expression of joy (trials 1-10) and pride (trials 11-20). Here, to conclude, participants read aloud all 3 statements they had selected.

After each trial, reinforcing statements and pictures were displayed on the screen. Throughout the sequence of training trials, study therapists rated the persuasiveness of participants'

XSL•FO
RenderX

emotional expression, provided feedback to participants, and maintained control of the training app using a tablet.

Measures

Feasibility of the Intervention

To determine the feasibility of the intervention, we assessed safety, technical feasibility, compliance, usability, and acceptability.

Safety of the Intervention

To determine the safety of the intervention, study therapists recorded whether participants experienced significant mood deterioration during the training. The intervention was deemed safe if no more than 10% (1/10) of the training sessions were prematurely terminated due to significant mood deterioration.

Compliance

The compliance rate was defined as the number of participants completing the intervention per protocol. The intervention was considered feasible if the compliance rate was at least 50% (5/10).

Technical Feasibility

Technical difficulties with the EmpkinS-EKSpression app during the training were to be reported by study therapists. We considered the intervention feasible if at least 80% (8/10) of the sessions were completed without major technical problems.

Usability

The usability of the training app was assessed with the pragmatic quality scale of the short version of the User Experience Questionnaire (UEQ-S [62]) and with one self-constructed item. In the UEQ-S, 8 pairs of opposite adjectives related to the user experience are rated on a 7-point Likert scale from –3 (fully agree with the negative term) to +3 (fully agree with the positive term). The pragmatic quality scale consists of 4 of the 8 items, with the other 4 items belonging to the hedonic quality scale. The self-constructed item asked participants how intelligible the app instructions were on a 4-point Likert scale from 1 (not at all) to 4 (very). We considered the intervention usable if >80% (8/10) of participants rated the app as positive (score ≥1), that is, supportive (rather than obstructive), easy (rather than complicated), efficient (rather than inefficient), clear (rather than confusing), and very intelligible.

Acceptability

Acceptability of the intervention was assessed with the "hedonic quality" scale of the UEQ-S. The intervention was considered acceptable if >80% of participants rated the training as positive (score ≥1) that is, interesting (vs not interesting), exciting (vs boring), inventive (vs conventional), and leading edge (vs usual). A total of 12 additional self-generated items were used to assess acceptability (Multimedia Appendix 2 provides the items). Except for 2 items, they were all rated on a 4-point Likert scale from 1 (not at all) to 4 (very). One item asking whether participants would recommend the training to a friend was answered with yes or no. Another item asking whether the training takes a reasonable amount of time was rated on a 5-point Likert scale from 1 (far too high) to 5 (far too low). The intervention was considered acceptable if >80% of participants indicated they would recommend the training to a friend, were rather content with the training, would continue the training given the opportunity, considered the training as rather helpful, and fun (score ≥3), as well as requiring a reasonably acceptable amount of time (score 2-4), and rate the training as only a little or not at all strenuous and difficult to focus on (score ≤2). Moreover, open-ended questions were asked by study therapists during the training and at the T3 assessment to gain additional, qualitative feedback for improvement. Acceptability of the overall study was assessed with one item asking participants to rate how burdensome study participation had been on a 4-point Likert scale from 1 (not at all) to 4 (very). The study was considered acceptable if >80% of participants rated participation as only a little or not at all burdensome.

Clinical Measures

The primary clinical outcome was current depressed mood and the secondary outcome was current positive mood.

Current depressed mood was assessed using an 11-point Likert scale from 0 (no depressed mood at all) to 10 (very strong depressed mood) during both diagnostic and training sessions. During training, participants submitted 108 ratings in total: 24 during the preparatory phase, one after the negative mood induction, one immediately before the first training item, one after each initial statement, and one after the completion of each training trial.

Current positive mood was also assessed using an 11-point Likert-scale from 0 (no positive mood at all) to 10 (very strong positive mood). Similar to the number of ratings of depressed mood, participants submitted a total of 108 ratings of positive mood.

Demographic and Health-Related Information

Demographic and health-related variables included age, date of birth, German language skills, nationality, gender, height, weight, academic degree, professional occupation, relationship status, scholarly fields of study (if any), current or previous diagnosis of a mental illness, psychotherapy, medication, color blindness, and Botox treatment. In addition, the PHQ-8 [52] was used to assess current depressive symptom severity during diagnostic screening. The PHQ-8 consists of 8 items assessing the frequency of 8 of the 9 *Diagnostic and Statistical Manual of Mental Disorders* criteria of depression during the previous 2 weeks on a 4-point Likert scale (0=not at all to 3=nearly every day). The items are added up to a sum maximum score of 24, with a score ≥10 indicating clinically relevant symptom severity [52].

Additional Measures

As one aim of our feasibility study was to test the technical setup of the study and protocol procedures, we included all measures to be assessed in the RCT. However, the additional measures listed below will not be analyzed in this feasibility evaluation. The clinical status of participants was assessed with the SCID-5-CV during diagnostic screening. Symptoms of depression were assessed with the GRID-HAMD [54] during the training session and at follow-up. In addition, self-reported symptoms of depression were assessed with the Centre for

XSL•FO
RenderX

Epidemiological Studies Depression Scale (German version [63]). Current suicidality was assessed with a single self-developed item. Dysfunctional attitudes were assessed with the German short version of the Dysfunctional Attitude Scale, form A [64], and automatic thoughts with the German version of the Automatic Thoughts Questionnaire-Revised [60]. Finally, the emotional state was assessed with the Self-Report Instrument for the Assessment of Emotion-Specific Regulation Skills parts A and B (SEK-ES [65]).

Two video cameras were used to assess kinesthesia (ie, facial expression and body posture): an Azure Kinect depth camera (Microsoft) and a high-resolution RGB camera (Sony SRG-300H), positioned 1.2 meters in front of participants. In addition, bipolar EMG of various facial muscles (ie, corrugator supercilii, zygomaticus major, masseter, orbicularis oculi) and of the shoulder muscle (trapezius) were assessed with the BioPac MP160 system (Biopac Inc). Further psychophysiological measures assessed with the BioPac MP160 system included electrocardiogram, respiration (using a respiration belt), and electrodermal activity. In addition, heart rate and respiration were assessed with radar-based technology [66]. The preparatory phase served as a baseline period for physiological data acquisition. Further, we assessed pupillometry data. Moreover, the smartphone app recorded videos of each trial and participants' digital activity on the app using the smartphone's front camera.

Statistical Analysis Plan

The safety and feasibility of the intervention and collection of kinesthetic and physiological data were determined based on descriptive analyses. Results were compared to predefined threshold values (see above). As a manipulation check, we analyzed whether the negative mood induction successfully elicited an increase in depressed mood and a decrease in positive mood. To this end, a paired-sample *t* test was computed to compare depressed or positive moods before and after the negative mood induction. To analyze the clinical potential of the intervention, we also computed paired-sample *t* tests to compare the mean rating of depressed (dependent variable 1) and positive mood (dependent variable 2) after the presentation of the statements, with the mean depressed and positive mood rating after completion of the trials. We analyzed the clinical potential separately for the 2 training phases. For effect sizes, we computed Hedges *g* as proposed by Lakens [67] for small samples, with Hedges *g* values of 0.2, 0.5, and 0.8 representing small, moderate, or large effects, respectively [68]. Moreover, we implemented 2 linear mixed-effects models with time-point as a fixed effect on level 1 and random intercepts and random slopes to analyze the course of depressed and positive mood ratings over the course of the training. The dependent variables were depressed mood and positive mood, respectively. However, as our data showed too little variance, particularly during the second training phase, the models appeared to not converge. Therefore, we merely analyzed the course of depressed and positive mood on a descriptive level.

The level of significance for all analyses was α=.05. The analyses were conducted with R Studio (version 4.3.2; Posit) [69].

Ethical Considerations

The study followed the Declaration of Helsinki's ethical guidelines and obtained ethical approval from FAU's Ethics Committee (20-443-B, 20-443_1-B, 20-443_2-B). All participants provided written informed consent at 2 time points during the study: (1) for the screening at the beginning of the online questionnaire via Unipark and (2) for the rest of the study at the beginning of T1. To deidentify questionnaire and app data, each participant received a personal code. A list connecting personal information with the code was stored on a secure server and accessible by study staff only. Video and audio data collected during the study that could not be completely anonymized were stored on an external hard drive that was locked away at all times.

Participants received compensation up to €40 (€10 for the diagnostic session, €20 for the training session, €5 for the follow-up assessment, and a €5 bonus for completing the whole study; a currency exchange rate of 1€=US $1.09 is applicable). Psychology students at FAU could alternatively receive course credits.

Results

Feasibility of the Intervention

The study flow is displayed in Figure 2. Regarding safety and compliance, none of the participants terminated the training prematurely. Technical difficulties with the EmpkinS-EKSpression app occurred in 2 training sessions: one of them had to be terminated immediately after the start due to an error in the app, leading to the drop-out of that participant (Figure 2). The other session was terminated after the first training phase also due to an error in the app. Here, the data recorded by the smartphone app (including mood ratings delivered throughout the training) were missing. Therefore, this participant was not included in the analysis of clinical potential. However, the data were included in the feasibility analyses.

With regards to usability as assessed with the pragmatic quality scale of the UEQ-S, favorable results emerged. Except for the instrument's "complicated – easy" item, all items were rated positively (Table 2). The total score of the pragmatic quality scale was a mean of 1.45 (SD 0.71). A total of 70% (7/10) of participants assessed the overall usability positively. Ratings for our self-developed item on the intelligibility of the app instructions ranged from 3 to 4 with a median of 4 (mean 3.9, SD 0.32). 90% (9/10) of participants rated the instructions as very intelligible.

XSL•FO
RenderX

Figure 2. Consolidated Standards of Reporting Trials (CONSORT) flow diagram of participants in the feasibility study investigating the single-session EmpkinS-EKSpression Reappraisal Training against depression in healthy individuals from screening to analysis.

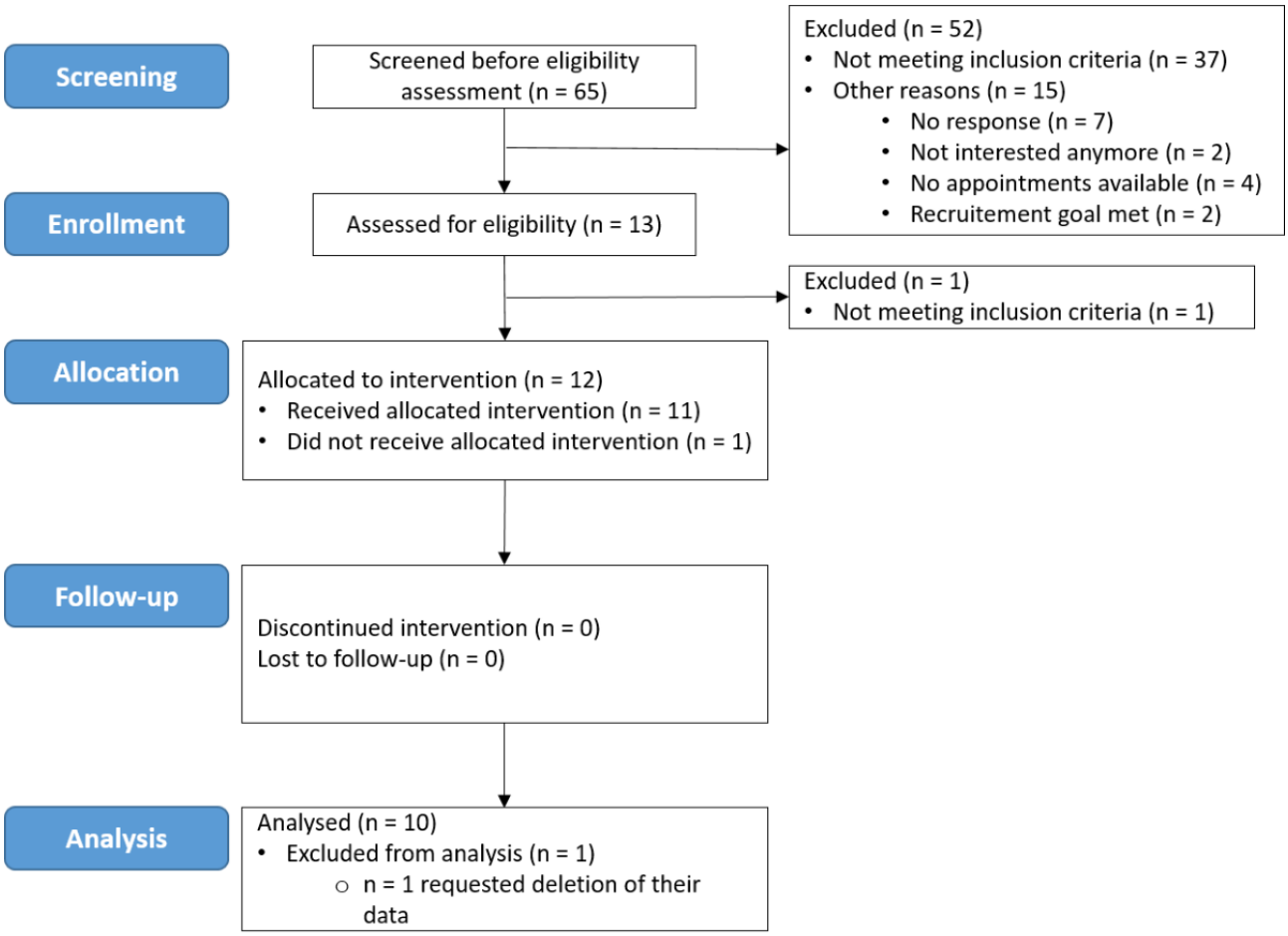


Table 2. Descriptive statistics of the short version of the User Experience Questionnaire assessing the usability of the single-session EmpkinS-EKSpression Reappraisal Training against depression examined in this feasibility study in healthy individuals.

	Median score	Score range	Scores, mean (SD)	Participants (N=10) with scores meeting benchmark[a], n (%)
Pragmatic quality scale				
Obstructive–supportive	2	0 to 3	1.8 (0.79)	9 (90)
Complicated–easy	2	–3 to 3	1 (2.11)	6 (60)
Inefficient–efficient	1	–1 to 2	1.1 (0.99)	8 (80)
Confusing–clear	2	–1 to 3	1.9 (1.29)	9 (90)
Hedonic quality scale				
Not interesting–interesting	1.5	0 to 3	1.5 (0.85)	9 (90)
Boring–exciting	0.5	–1 to 2	0.2 (1.14)	5 (50)
Conventional–inventive	1.5	–2 to 3	1.1 (1.45)	7 (70)
Usual–leading edge	2	–2 to 3	1.4 (1.51)	8 (80)

[a]The predefined benchmark for the items considered to be rated as positive was a score ≥1. The percentage of participants meeting this benchmark is displayed here.

For acceptability, as assessed with the hedonic quality scale of the UEQ-S, results were more mixed (Table 2). While the total score was a mean of 1.05 (SD 0.79), only 50% (5/10) of participants gave a positive total score for the hedonic quality scale. Regarding our self-developed items, 80% (8/10) of participants indicated they would recommend the training to a friend. Further descriptive statistics are displayed in Table 3. In the open evaluation of the training, 4 participants revealed that they found the training repetitive; 2 found it exhausting. A total of 5 participants found the emotional display, 3 pride in

XSL•FO
RenderX

particular, initially unfamiliar and difficult to engage in. Moreover, 5 participants were critical that the standardized statements were not better tailored for individuals, and 3 were dissatisfied with the study setting (ie, uncomfortable seating position, feeling observed by the study therapist, and awkward electrodes). In contrast, some positive feedback included that the training was fun and interesting (n=4), that it provided new ideas for dealing with negative thoughts (n=4), and that there were several statements to choose from (n=1). Regarding the acceptability of the study, participation burden ratings ranged from 1 to 2, with a median of 1 (mean 1.1, SD 0.2), hence, 100% (10/10) of participants rated the participation as only a little or not at all burdensome.

Table 3. Descriptive statistics of self-constructed items assessing the acceptability of the single-session EmpkinS-EKSpression Reappraisal Training against depression examined in this feasibility study in healthy individuals (N=10).

	Median score	Score range	Scores, mean (SD)	Participants (N=10) with scores meeting benchmark[a], n (%)
Likelihood of continuation	2	1 to 3	1.9 (0.74)	2 (20)
Satisfaction in general	3	2 to 4	2.9 (0.57)	8 (80)
Helpfulness				
Training part 1	2	1 to 4	2.4 (0.84)	4 (40)
Training part 2[b]	3	2 to 3	2.67 (0.5)	6 (60)
Training in total	3	2 to 3	2.6 (0.52)	6 (60)
Modifying negative thoughts	2	2 to 3	2.3 (0.48)	3 (30)
Coping with negative thoughts	2	2 to 4	2.5 (0.71)	4 (40)
Enjoyment	2	2 to 3	2.4 (0.52)	4 (40)
Strenuousness	2	2 to 3	2.2 (0.42)	8 (80)
Reasonable amount of time	2	2 to 4	2.4 (0.7)	3 (30)
Difficulty to focus	1	1 to 2	1.3 (0.48)	10 (100)

[a]The predefined benchmark for the items considered to be rated as positive was a score ≥3 for the likelihood of continuation, satisfaction in general, helpfulness, and enjoyment; a score between 2 and 4 for a reasonable amount of time; and a score≤ 2 for strenuousness and difficulty to focus.

[b]N=9.

Manipulation Check

With regards to negative mood induction, the mean rating of depressed mood was 0.5 (SD 1.41) before and 1.44 (SD 0.73) after mood induction. The mean rating of positive mood was 5.67 (SD 1.66) before and 4.56 (SD 1.88) after negative mood induction. Both paired-samples *t* tests reached significance (t_7=–3.06, *P*=.02, Hedges *g*=–0.69, 95% CI –2.83 to –0.28 and t_8=3.16, *P*=.01, Hedges *g*=–0.95, 95% CI –1.84 to –0.55; respectively).

Clinical Potential

In the first training phase, the mean rating for depressed mood was 2.63 (SD 2.51) before and 2.41 (SD 2.36) after, and the mean rating for positive mood was 4.14 (SD 2.42) before and 5.59 (SD 1.65) after the explicit rejection of depressogenic statements, respectively. The difference was not significant for depressed mood (t_{178}=1.16, *P*=.25, Hedges *g*=0.09, 95% CI –0.07 to 0.23), but did reach significance for positive mood (t_{179}=–8.53, *P*<.001, Hedges *g*=–0.63, 95% CI –0.72 to –0.55). In the second training phase, the mean rating of depressed mood was 2.75 (SD 2.47) before and 2.74 (SD 2.48) after, and the mean rating of positive mood was 5.62 (SD 1.71) before and 5.81 (SD 1.58) after the explicit substantiation of antidepressive statements, respectively. The difference was not significant for depressed mood (t_{178}=–0.16, *P*=.87, Hedges *g*=–0.01, 95% CI –0.16 to 0.12), but did, again, reach significance for positive mood (t_{177}=–3.38, *P*<.001, Hedges *g*=–0.25, 95% CI –0.38 to –0.11]). Descriptively, there was a visible mood change after the negative mood induction (ie, an increase in depressed mood, and a decrease in positive mood) and an increase in positive mood after each trial, particularly during the first training phase (Figure 3). Moreover, there was a slight increase in depressed mood over the course of the first training phase, and in positive mood over the course of the second training phase.

XSL•FO
RenderX

Figure 3. Subjective depressed and positive mood ratings over the course of the single-session EmpkinS-EKSpression Reappraisal Training against depression were examined in this feasibility study in healthy individuals. Regression lines represent loess curves.

Discussion

Principal Findings

This study aimed to assess the safety, feasibility, and clinical potential of the novel EmpkinS-EKSpression training that leverages the power of AFE to augment the beneficial effects of CR on depression. Overall, the intervention proved to be safe and feasible, and findings on its clinical potential were promising. However, we identified several aspects that warrant further clinical and technical optimization.

With regard to feasibility, the results were quite positive, and the predefined benchmarks were met for safety, compliance, and technical feasibility. Concerning the usability of the training app as assessed with the pragmatic quality scale of the UEQ-S and a self-developed item on the intelligibility of the app instructions, the predefined benchmark was narrowly missed. This result is worse than the findings of Stiles-Shields et al [24] who evaluated a structured cognitive restructuring exercise where usability was rated to be high. An important, possibly explanatory difference in this study is a higher degree of individualization: participants could choose between self-generated and preselected content in each step of the cognitive restructuring task. It is possible that including more options for individualization of content might improve the usability of the EmpkinS-EKSpression training. Similarly, one of the studies investigating an AAMT intervention in the context of depression reported high usability ratings [39]. The training content here was highly standardized, so a lack of individualization might not be the only explanatory factor for missing the usability benchmark in this study. One difference was the amount of time spent on the cognitive restructuring task (which was not reported by Stiles-Shields and colleagues [24]). While individuals in our study trained for 90-120 minutes, individuals in the study by Lukas and Berking [39] completed an average of around 40 minutes of training. Therefore, a shorter treatment duration might be beneficial to increase usability. When analyzing our usability results at an item level, only the results for the “complicated – easy” item did not meet the benchmark in our study. The variance of the item emerged to be high, and the median indicates that most participants rated the app as easy. Thus, the fact that the benchmark was missed here is likely attributable to singular, extreme votes. Nevertheless, the explicit display of antidepressive kinesthesia might be more challenging than mere cognitive restructuring exercises. Therefore, this finding should be considered in future studies using similar paradigms with rather complicated setups, such as AAMT, to emphasize a user-friendly development of interventions and involve target groups to address their needs, for example, by improving study support or providing more information material. We decided to address this issue in the upcoming RCT by providing a more in-depth oral explanation by study therapists during the training phase. Results on the acceptability of the training, as assessed with the hedonic quality of the UEQ-S, as well as several self-constructed items, did not meet the predefined benchmarks. The open evaluation revealed

XSL•FO
RenderX

that over half of the participants found the training to be too repetitive and too long. This finding is in line with previous work investigating the acceptability of a digital cognitive bias modification training that also required participants to complete many trials, which suggests that participants' willingness to engage in repetitive digital treatments is limited [70]. However, from the ICS theoretical perspective, repeated rehearsal of new and adaptive schemas is important to break the cognitive interlock and thus the maintenance of depressed mood [25]. One possible solution to this dilemma, as suggested by Beard and colleagues [70], is to increase the credibility of the intervention by ensuring a highly comprehensible and convincing treatment rationale and motivation. This suggestion is also in line with a systematic review of barriers and facilitators for digital mental health interventions, which identified credible content of an intervention to be an important facilitator for user engagement [71]. Our feasibility study was likely to fall short of meeting this requirement, given that we only included healthy individuals with perhaps only a limited need for a mood enhancement intervention. Thus, the content of the training was probably not sufficiently relevant for participants, which in turn could have compromised its credibility. However, it is common practice for feasibility studies to initially scrutinize general intervention effects in healthy individuals before exposing more vulnerable clinical participants to those effects [72]. Nevertheless, we took this as an indication to expand our study instructions and elaborate on the study rationale for the subsequent RCT. This approach may also be useful to increase the face validity and thus potentially the effectiveness [73] of other digital mental health interventions, such as structured cognitive restructuring exercises [22,24], smartphone-based AAMT used for CR [38-45], or original AAMT versions used to retrain approach and avoidance biases (Loijen et al performed a systematic review [74]).

The fact that our feasibility sample consisted of healthy individuals might also explain the finding that participants did not consider the training to be particularly helpful. If they did not relate to the standardized depressogenic statements in the first place (which was reflected in participants' open feedback), then there were no negative underlying cognitions to be modified. In terms of strain associated with study participation and training, the benchmark was met for all items. However, 7 participants indicated that the training session was too protracted. Therefore, we decided to abbreviate the preparatory phase of the training to the presentation of only 12 instead of 24 stimuli for 5 instead of 8 seconds. The primary purpose of the preparatory phase was to gather baseline physiological data, which was shown in the feasibility study to be adequately achieved in a shorter time frame. We also decided to move the GRID-HAMD interview to the diagnostic session (T1) to reduce participant burden during the subsequent training session. Another important feedback for the training was that participants expressed difficulty in fully engaging in the spontaneous display of emotion. As a result, the RCT will incorporate increased support to participants from study therapists during the training. One implication for future interventions using antidepressive kinesthesia to augment cognitive restructuring derived from this feedback is that such interventions should be therapist-supported, at least when they are first delivered. Such support can maximize the intervention's effects by leveraging the therapeutic relationship, which is a crucial common factor of psychotherapy [73]. This is also consistent with studies demonstrating that therapist-supported digital interventions for depression are more effective than self-directed interventions [75]. Furthermore, one participant reported that they found the electrodes uncomfortable, which was also witnessed by study therapists. Therefore, for the subsequent RCT, we planned to collect EMG data only for a limited subgroup of participants (for more detail, see the study protocol of the RCT [47]).

Results of the manipulation check suggest that the negative mood induction successfully induced depressed mood with a moderate-to-large effect size (Hedges g=–0.69), and reduced positive mood with a large effect size (Hedges g=–0.95). These results confirm the validity of the mood induction procedure [56,57] and its usefulness in the study of depression interventions. With regards to the clinical potential, rejecting the depressogenic statements in the first training phase and approving the antidepressive statements in the second training phase did not affect the depressed mood. Positive mood, however, did increase with a moderate-to-large effect (Hedges g=–0.63) in the first training phase and a small effect (Hedges g=–0.25) in the second training phase. The null findings for depressed mood may be due to floor effects for our healthy feasibility sample as the mean rating of depressed mood was very low throughout data collection, even after the negative mood induction and the presentation of the depressogenic statements. This is in line with results from the study of O'Toole and Michalak [46], who investigated the effects of augmenting CR with antidepressive body posture and movement. As in our study, healthy individuals were included and no effects on negative emotions were found. Findings from the study by Stiles-Shields et al, however, evaluating the effects of a structured cognitive restructuring exercise, were the opposite, with significant effects on symptoms of depression [24]. With regard to previous studies investigating smartphone-based CR augmented by antidepressive kinesthesia, the findings were similar. One of the studies examining AAMT focusing on CR found significant effects on symptoms of depression [38] (the other reported only descriptive data [39]). These differences may be explained by the fact that Stiles-Shields et al [24] as well as Lukas et al [38] examined individuals with symptoms of depression in their studies and not healthy individuals as in our study. Furthermore, these studies assessed symptoms of depression over the past week, whereas this study assessed the intervention's immediate effect on mood. Moreover, both studies assessed various symptoms of depression, while the current study focused exclusively on depressed mood.

The findings for positive mood, however, align with our hypothesis, suggesting that counteracting depressogenic statements in particular with antidepressive statements and kinesthesia exercises has positive effects on mood in healthy individuals. As such, they are in line with studies showing positive effects of CR on mood, such as those included in a meta-analysis by Ciharova et al [11]. In summary, the results of this and previous studies are promising with regard to the clinical potential of a structured cognitive restructuring exercise as well as the potential of augmenting cognitive restructuring

XSL•FO
RenderX

with antidepressive kinesthesia. These findings are particularly important because structured cognitive restructuring exercises, which also lead to emotional insight through augmentation with antidepressive kinesthesia, could easily be used in mobile interventions without direct human interaction. The types of kinesthesia that are particularly beneficial and the extent to which they augment effects of structured cognitive restructuring on depressed and positive mood as posited by the ICS theory of depression [25] and studies demonstrating effects of manipulating facial expressions on depressed mood [29,30,34,36], will be subject of our subsequent RCT.

Beyond investigating the feasibility and clinical potential of the training, this feasibility study enabled us the opportunity to ensure that the technical aspects of the study (ie, the kinesthetic and physiological data acquisition) ran smoothly. In this respect, we have drawn several implications from the feasibility results. For instance, to improve the kinesthetic data acquisition, we decided to use a second high-resolution RGB camera in the upcoming RCT to record participants' upper bodies and heads viewed from the side. Moreover, after evaluating our data and consulting with experts in the field, contrary to our initial preregistrations (RCT [76]; feasibility study [48]), we decided to drop the electrodermal activity and pupillometry measurements, as well as the EMG measures of the trapezius due to poor data quality and because an improvement of data quality would only have been possible at the cost of disrupting therapeutic processes. Overall, all further issues jeopardizing the quality of kinesthetic and psychophysiological measures could be fixed before the RCTs start, thereby optimizing the development of machine-learning models for automated depression assessment.

Limitations and Conclusion

Major limitations of this feasibility study include the fixed sequence of first invalidating depressogenic cognitions and thereafter validating antidepressive cognitions. This harbors the risk of position effects on depressed and positive moods. Therefore, we plan to randomize the order of the 2 training phases in the RCT. In addition, the intensive contact with study therapists during the training may have prompted socially desirable behavior in participants, particularly regarding compliance with the training and mood ratings submitted during the training. Future studies should include measures of social desirability and ways to control its effects. On the other hand, this contact also represents a strength of the study, as the therapeutic relationship is used to increase the efficacy of the intervention. Despite these limitations and the small sample size, this feasibility study provided valuable insights that led to significant improvements in the study protocol of our upcoming RCT. As such, it serves as a critical keystone in the further development and investigation of innovative interventions for depression.

Acknowledgments

Our foremost gratitude to all study participants. We also thank all the students who provided invaluable support to our team and the EmpkinS-EKSpression app developers. The R code for creating Figure 3 was partially generated using ChatGPT (ChatGPT v2; OpenAI).

Data Availability

The data sets analyzed during this study and the analysis code are available in the Open Science Framework [48].

Conflicts of Interest

None declared.

Multimedia Appendix 1
American Psychological Association journal article reporting standards.
[PDF File (Adobe PDF File), 1037 KB - formative_v9i1e65357_app1.pdf]

Multimedia Appendix 2
Self-developed items to assess usability and acceptability.
[DOCX File , 41 KB - formative_v9i1e65357_app2.docx]

References

1. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry 2022;9(2):137-150 [FREE Full text] [doi: 10.1016/S2215-0366(21)00395-3] [Medline: 35026139]
2. Kessler RC, Bromet EJ. The epidemiology of depression across cultures. Annu Rev Public Health 2013;34(1):119-138 [FREE Full text] [doi: 10.1146/annurev-publhealth-031912-114409] [Medline: 23514317]
3. Correll CU, Solmi M, Veronese N, Bortolato B, Rosson S, Santonastaso P, et al. Prevalence, incidence and mortality from cardiovascular disease in patients with pooled and specific severe mental illness: a large-scale meta-analysis of 3,211,768

XSL•FO
RenderX

patients and 113,383,368 controls. World Psychiatry 2017;16(2):163-180 [FREE Full text] [doi: 10.1002/wps.20420] [Medline: 28498599]
4. Cuijpers P, Vogelzangs N, Twisk J, Kleiboer A, Li J, Penninx BW. Comprehensive meta-analysis of excess mortality in depression in the general community versus patients with specific illnesses. Am J Psychiatry 2014;171(4):453-462 [FREE Full text] [doi: 10.1176/appi.ajp.2013.13030325] [Medline: 24434956]
5. Cuijpers P, Berking M, Andersson G, Quigley L, Kleiboer A, Dobson KS. A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. Can J Psychiatry 2013;58(7):376-385. [doi: 10.1177/070674371305800702] [Medline: 23870719]
6. Hans E, Hiller W. Effectiveness of and dropout from outpatient cognitive behavioral therapy for adult unipolar depression: a meta-analysis of nonrandomized effectiveness studies. J Consult Clin Psychol 2013;81(1):75-88. [doi: 10.1037/a0031080] [Medline: 23379264]
7. Cuijpers P, Karyotaki E, Weitz E, Andersson G, Hollon SD, van Straten A. The effects of psychotherapies for major depression in adults on remission, recovery and improvement: a meta-analysis. J Affect Disord 2014;159:118-126. [doi: 10.1016/j.jad.2014.02.026] [Medline: 24679399]
8. Mekonen T, Chan GCK, Connor JP, Hides L, Leung J. Estimating the global treatment rates for depression: a systematic review and meta-analysis. J Affect Disord 2021;295:1234-1242. [doi: 10.1016/j.jad.2021.09.038] [Medline: 34665135]
9. Beck AT, Haigh EAP. Advances in cognitive theory and therapy: the generic cognitive model. Annu Rev Clin Psychol 2014;10(1):1-24. [doi: 10.1146/annurev-clinpsy-032813-153734] [Medline: 24387236]
10. Beck AT, Dozois DJA. Cognitive therapy: current status and future directions. Annu Rev Med 2011;62:397-409. [doi: 10.1146/annurev-med-052209-100032] [Medline: 20690827]
11. Ciharova M, Furukawa TA, Efthimiou O, Karyotaki E, Miguel C, Noma H, et al. Cognitive restructuring, behavioral activation and cognitive-behavioral therapy in the treatment of adult depression: a network meta-analysis. J Consult Clin Psychol 2021;89(6):563-574. [doi: 10.1037/ccp0000654] [Medline: 34264703]
12. Morello K, Schäfer SK, Kunzler AM, Priesterroth LS, Tüscher O, Kubiak T. Cognitive reappraisal in mHealth interventions to foster mental health in adults: a systematic review and meta-analysis. Front Digit Health 2023;5:1253390 [FREE Full text] [doi: 10.3389/fdgth.2023.1253390] [Medline: 37927578]
13. Furukawa TA, Karyotaki E, Suganuma A, Pompoli A, Ostinelli EG, Cipriani A, et al. Dismantling, personalising and optimising internet cognitive-behavioural therapy for depression: a study protocol for individual participant data component network meta-analysis. BMJ Open 2019;8(11):e026137 [FREE Full text] [doi: 10.1136/bmjopen-2018-026137] [Medline: 30798295]
14. Bruhns A, Lüdtke T, Moritz S, Bücker L. A mobile-based intervention to increase self-esteem in students with depressive symptoms: randomized controlled trial. JMIR Mhealth Uhealth 2021;9(7):e26498 [FREE Full text] [doi: 10.2196/26498] [Medline: 34255711]
15. Ham K, Chin S, Suh YJ, Rhee M, Yu ES, Lee HJ, et al. Preliminary results from a randomized controlled study for an app-based cognitive behavioral therapy program for depression and anxiety in cancer patients. Front Psychol 2019;10:1592 [FREE Full text] [doi: 10.3389/fpsyg.2019.01592] [Medline: 31402881]
16. Jannati N, Mazhari S, Ahmadian L, Mirzaee M. Effectiveness of an app-based cognitive behavioral therapy program for postpartum depression in primary care: a randomized controlled trial. Int J Med Inform 2020;141:104145. [doi: 10.1016/j.ijmedinf.2020.104145] [Medline: 32480319]
17. Mantani A, Kato T, Furukawa TA, Horikoshi M, Imai H, Hiroe T, et al. Smartphone cognitive behavioral therapy as an adjunct to pharmacotherapy for refractory depression: randomized controlled trial. J Med Internet Res 2017;19(11):e373 [FREE Full text] [doi: 10.2196/jmir.8602] [Medline: 29101095]
18. Dahne J, Collado A, Lejuez CW, Risco CM, Diaz VA, Coles L, et al. Pilot randomized controlled trial of a Spanish-language behavioral activation mobile app (¡Aptívate!) for the treatment of depressive symptoms among united states latinx adults with limited english proficiency. J Affect Disord 2019;250:210-217 [FREE Full text] [doi: 10.1016/j.jad.2019.03.009] [Medline: 30870770]
19. Moberg C, Niles A, Beermann D. Guided self-help works: randomized waitlist controlled trial of pacifica, a mobile app integrating cognitive behavioral therapy and mindfulness for stress, anxiety, and depression. J Med Internet Res 2019;21(6):e12556 [FREE Full text] [doi: 10.2196/12556] [Medline: 31199319]
20. Nuño L, Gómez-Benito J, Carmona VR, Pino O. A systematic review of executive function and information processing speed in major depression disorder. Brain Sci 2021;11(2):147 [FREE Full text] [doi: 10.3390/brainsci11020147] [Medline: 33499360]
21. Grahek I, Shenhav A, Musslick S, Krebs RM, Koster EHW. Motivation and cognitive control in depression. Neurosci Biobehav Rev 2019;102:371-381 [FREE Full text] [doi: 10.1016/j.neubiorev.2019.04.011] [Medline: 31047891]
22. McCloud T, Jones R, Lewis G, Bell V, Tsakanikos E. Effectiveness of a mobile app intervention for anxiety and depression symptoms in university students: randomized controlled trial. JMIR Mhealth Uhealth 2020;8(7):e15418 [FREE Full text] [doi: 10.2196/15418] [Medline: 32735221]
23. Ellis A. The revised ABC's of rational-emotive therapy (RET). J Ration-Emot Cogn-Behav Ther 1991;9(3):139-172. [doi: 10.1007/BF01061227]

XSL•FO
RenderX

24. Stiles-Shields C, Montague E, Kwasny MJ, Mohr DC. Behavioral and cognitive intervention strategies delivered via coached apps for depression: pilot trial. Psychol Serv 2019;16(2):233-238 [FREE Full text] [doi: 10.1037/ser0000261] [Medline: 30407055]
25. Teasdale JD, Barnard PJ. Affect, Cognition, and Change: Re-modelling depressive thought. Mahwah, NJ: Lawrence Erlbaum Associates; 1993.
26. Ekman P, Friesen WV. Unmasking the Face: A Guide to Recognizing Emotions From Facial Expressions. Englewood Cliffs, NJ: Prentice-Hall; 1975.
27. Coles NA, Larsen JT, Lench HC. A meta-analysis of the facial feedback literature: effects of facial feedback on emotional experience are small and variable. Psychol Bull 2019;145(6):610-651. [doi: 10.1037/bul0000194] [Medline: 30973236]
28. Davies H, Wolz I, Leppanen J, Fernandez-Aranda F, Schmidt U, Tchanturia K. Facial expression to emotional stimuli in non-psychotic disorders: a systematic review and meta-analysis. Neurosci Biobehav Rev 2016;64:252-271. [doi: 10.1016/j.neubiorev.2016.02.015] [Medline: 26915928]
29. Girard JM, Cohn JF, Mahoor MH, Mavadati S, Rosenwald DP. Social risk and depressionvidence from manual and automatic facial expression analysis. 2013 Presented at: 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG); April 26, 2013; Shanghai, China. [doi: 10.1109/fg.2013.6553748]
30. Pampouchidou A, Pediaditis M, Kazantzaki E, Sfakianakis S, Apostolaki IA, Argyraki K, et al. Automated facial video-based recognition of depression and anxiety symptom severity: cross-corpus validation. Mach Vis Appl 2020;31(4):30. [doi: 10.1007/s00138-020-01080-7]
31. Larsen JT, Norris CJ, Cacioppo JT. Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. Psychophysiology 2003;40(5):776-785. [doi: 10.1111/1469-8986.00078] [Medline: 14696731]
32. Lindsey KT, Rohan KJ, Roecklein KA, Mahon JN. Surface facial electromyography, skin conductance, and self-reported emotional responses to light- and season-relevant stimuli in seasonal affective disorder. J Affect Disord 2011;133(1-2):311-319 [FREE Full text] [doi: 10.1016/j.jad.2011.04.016] [Medline: 21600661]
33. Wexler BE, Levenson L, Warrenburg S, Price LH. Decreased perceptual sensitivity to emotion-evoking stimuli in depression. Psychiatry Res 1994;51(2):127-138 [FREE Full text] [doi: 10.1016/0165-1781(94)90032-9] [Medline: 8022947]
34. Schwartz GE, Fair PL, Mandel MR, Salt P, Mieske M, Klerman GL. Facial electromyography in the assessment of improvement in depression. Psychosom Med 1978;40(4):355-360. [doi: 10.1097/00006842-197806000-00008] [Medline: 684137]
35. Davey GCL, Sired R, Jones S, Meeten F, Dash SR. The role of facial feedback in the modulation of clinically-relevant ambiguity resolution. Cogn Ther Res 2012;37(2):284-295. [doi: 10.1007/s10608-012-9480-5]
36. Schulze J, Neumann I, Magid M, Finzi E, Sinke C, Wollmer MA, et al. Botulinum toxin for the management of depression: an updated review of the evidence and meta-analysis. J Psychiatr Res 2021;135:332-340. [doi: 10.1016/j.jpsychires.2021.01.016] [Medline: 33578275]
37. Rinck M, Becker ES. Approach and avoidance in fear of spiders. J Behav Ther Exp Psychiatry 2007;38(2):105-120. [doi: 10.1016/j.jbtep.2006.10.001] [Medline: 17126289]
38. Lukas CA, Eskofier B, Berking M. A gamified smartphone-based intervention for depression: randomized controlled pilot trial. JMIR Ment Health 2021;8(7):e16643 [FREE Full text] [doi: 10.2196/16643] [Medline: 34283037]
39. Lukas CA, Berking M. Blending group-based psychoeducation with a smartphone intervention for the reduction of depressive symptoms: results of a randomized controlled pilot study. Pilot Feasibility Stud 2021;7(1):57 [FREE Full text] [doi: 10.1186/s40814-021-00799-y] [Medline: 33627198]
40. Kollei I, Lukas CA, Loeber S, Berking M. An app-based blended intervention to reduce body dissatisfaction: a randomized controlled pilot study. J Consult Clin Psychol 2017;85(11):1104-1108. [doi: 10.1037/ccp0000246] [Medline: 28857576]
41. Lukas CA, Blechert J, Berking M. A smartphone application to reduce problematic drinking: a feasibility trial. Pilot Feasibility Stud 2024;10(1):34 [FREE Full text] [doi: 10.1186/s40814-023-01420-0] [Medline: 38378675]
42. Lukas CA, Berking M. Reducing procrastination using a smartphone-based treatment program: a randomized controlled pilot study. Internet Interv 2018;12:83-90 [FREE Full text] [doi: 10.1016/j.invent.2017.07.002] [Medline: 30135772]
43. Keinert M, Schindler-Gmelch L, Eskofier BM, Berking M. An anger-based approach-avoidance modification training targeting dysfunctional beliefs in adults with elevated stress – results from a randomized controlled pilot study. Int J Cogn Ther 2024;17(4):700-724. [doi: 10.1007/s41811-024-00218-z]
44. Rupp LH, Keinert M, Böhme S, Schindler-Gmelch L, Eskofier B, Schuller B, et al. Sadness-based approach-avoidance modification training for subjective stress in adults: pilot randomized controlled trial. JMIR Form Res 2023;7:e50324 [FREE Full text] [doi: 10.2196/50324] [Medline: 38032725]
45. Streit H, Keinert M, Schindler-Gmelch L, Eskofier B, Berking M. Disgust-based approach-avoidance modification training for individuals suffering from elevated stress: a randomized controlled pilot study. Stress Health 2024;40(4):e3384. [doi: 10.1002/smi.3384] [Medline: 38367241]
46. O'Toole MS, Michalak J. Embodied cognitive restructuring: the impact of posture and movement on changing dysfunctional attitudes. J Behav Ther Exp Psychiatry 2024;84:101955 [FREE Full text] [doi: 10.1016/j.jbtep.2024.101955] [Medline: 38428361]

XSL•FO
RenderX

47. Keinert M, Schindler-Gmelch L, Rupp LH, Sadeghi M, Capito K, Hager M, et al. Facing depression: evaluating the efficacy of the empkinS-EKSpression reappraisal training augmented with facial expressions - protocol of a randomized controlled trial. BMC Psychiatry 2024;24(1):896 [FREE Full text] [doi: 10.1186/s12888-024-06361-3] [Medline: 39668374]
48. Capito K, Keinert M, Schindler-Gmelch L, Rupp LH, Sadeghi M, Eskofier B, et al. Efficacy of a smartphone-based reappraisal training against depressed mood and benefits of enhancing it with facial expression—a feasibility study. Open Science Framework. URL: https://osf.io/pw6ma/ [accessed 2024-12-09]
49. APA Style Journal Article Reporting Standards (APA Style JARS). American Psychological Association. URL: https://apastyle.apa.org/jars [accessed 2025-01-20]
50. Bell ML, Whitehead AL, Julious SA. Guidance for using pilot studies to inform the design of intervention trials with continuous outcomes. Clin Epidemiol 2018;10:153-157. [doi: 10.2147/clep.s146397]
51. Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. Stat Methods Med Res 2016;25(3):1057-1073 [FREE Full text] [doi: 10.1177/0962280215588241] [Medline: 26092476]
52. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr Ann 2002;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
53. Beesdo-Baum K, Zaudig M, Wittchen HU. In: Williams JBW, Karg RS, Spitzer RL, editors. Strukturiertes Klinisches Interview für DSM-5®-Störungen – Klinische Version. Newburyport, MA: Hogrefe; 2019.
54. Williams JB, Kobak KA, Bech P, Engelhardt N, Evans K, Lipsitz J, et al. The GRID-HAMD: standardization of the hamilton depression rating scale. Int Clin Psychopharmacol 2008;23(3):120-129. [doi: 10.1097/YIC.0b013e3282f948f5] [Medline: 18408526]
55. Beck AT, Rush J, Shaw BF, Emery G. Cognitive Therapy of Depression. New York, NY: Guilford Press; 1979.
56. Diedrich A, Grant M, Hofmann SG, Hiller W, Berking M. Self-compassion as an emotion regulation strategy in major depressive disorder. Behav Res Ther 2014;58:43-51. [doi: 10.1016/j.brat.2014.05.006] [Medline: 24929927]
57. Velten E. A laboratory task for induction of mood states. Behav Res Ther 1968;6(4):473-482. [doi: 10.1016/0005-7967(68)90028-4] [Medline: 5714990]
58. Kliem S, Lohmann A, Mößle T, Brähler E. Psychometric properties and measurement invariance of the beck hopelessness scale (BHS): results from a German representative population sample. BMC Psychiatry 2018;18(1):110 [FREE Full text] [doi: 10.1186/s12888-018-1646-6] [Medline: 29699522]
59. Pössel P. Cognitive triad inventory (CTI): psychometric properties and factor structure of the German translation. J Behav Ther Exp Psychiatry 2009;40(2):240-247. [doi: 10.1016/j.jbtep.2008.12.001] [Medline: 19110236]
60. Pössel P, Seemann S, Hautzinger M. Evaluation eines deutschsprachigen Instrumentes zur Erfassung positiver und negativer automatischer Gedanken. Z Für Klin Psychol Psychother 2005;34(1):27-34. [doi: 10.1026/1616-3443.34.1.27]
61. Ferring D, Filipp SH. Messung des Selbstwertgefühls: Befunde zur Reliabilität, Validität und Stabilität der Rosenberg-Skala [Measurement of self-esteem: Findings on reliability, validity, and stability of the Rosenberg Scale]. Diagnostica 1996:284-292.
62. Schrepp M, Hinderks A, Thomaschewski J. Design and evaluation of a short version of the User Experience Questionnaire (UEQ-S). Int J Interact Multimed Artif Intell 2017;4(6):103. [doi: 10.9781/ijimai.2017.09.001]
63. Hautzinger M, Bailer M, Hofmeister D, Keller F. Allgemeine Depressionsskala. Newburyport, MA: Hogrefe; 2012.
64. Rojas R, Geissner E, Hautzinger M. Entwicklung und psychometrische Überprüfung von zwei vergleichbaren Kurzversionen der Skala Dysfunktionaler Einstellungen (DAS). Diagnostica 2015;61(4):173-183. [doi: 10.1026/0012-1924/a000123]
65. Ebert DD, Christ O, Berking M. SEK-ES - Fragebogen zur emotionsspezifischen Selbsteinschätzung emotionaler Kompetenzen. Leibniz Institute for Psychology. 2013. URL: https://www.psycharchives.org/en/item/f2d7b513-70c7-444f-828f-94cea1781b04 [accessed 2025-03-26]
66. Albrecht NC, Langer D, Krauss D, Richer R, Abel L, Eskofier BM. EmRad: Ubiquitous vital sign sensing using compact continuous-wave radars. IEEE Open J Eng Med Biol 2024;5:725-734. [doi: 10.1109/ojemb.2024.3420241]
67. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
68. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd Edition. London, UK: Routledge; 1989.
69. R: A language and environment for statistical computing. R Core Team. 2017. URL: https://www.R-project.org/ [accessed 2024-02-29]
70. Beard C, Weisberg RB, Primack J. Socially anxious primary care patients' attitudes toward cognitive bias modification (CBM): a qualitative study. Behav Cogn Psychother 2012;40(5):618-633 [FREE Full text] [doi: 10.1017/S1352465811000671] [Medline: 22127022]
71. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
72. Diaz Hernandez L, Rieger K, Baenninger A, Brandeis D, Koenig T. Towards using microstate-neurofeedback for the treatment of psychotic symptoms in schizophrenia. a feasibility study in healthy participants. Brain Topogr 2016;29(2):308-321 [FREE Full text] [doi: 10.1007/s10548-015-0460-4] [Medline: 26582260]

XSL•FO
RenderX

73. Wampold BE. How important are the common factors in psychotherapy? An update. World Psychiatry 2015;14(3):270-277 [FREE Full text] [doi: 10.1002/wps.20238] [Medline: 26407772]
74. Loijen A, Vrijsen JN, Egger JIM, Becker ES, Rinck M. Biased approach-avoidance tendencies in psychopathology: a systematic review of their assessment and modification. Clin Psychol Rev 2020;77:101825. [doi: 10.1016/j.cpr.2020.101825] [Medline: 32143108]
75. Moshe I, Terhorst Y, Philippi P, Domhardt M, Cuijpers P, Cristea I, et al. Digital interventions for the treatment of depression: a meta-analytic review. Psychol Bull 2021;147(8):749-786. [doi: 10.1037/bul0000334] [Medline: 34898233]
76. Berking M, Rupp LH, Schindler-Gmelch L, Capito K, Böhme S, Keinert M. EmpkinS D02. Open Science Framework. URL: https://osf.io/mswfg/ [accessed 2024-12-09]

Abbreviations

AAMT: approach-avoidance modification training
AFE: antidepressive facial expression
CBT: cognitive behavioral therapy
CR: cognitive reappraisal
EMG: electromyography
EmpkinS: Empatho-Kinaesthetic Sensor Technology
FAU: Friedrich-Alexander-Universität Erlangen-Nürnberg
GRID-HAMD: GRID Hamilton Depression Rating Scale
ICS: interacting cognitive subsystems
PHQ: Patient Health Questionnaire
RCT: randomized controlled trial
SCID-5-CV: Structured Clinical Interview for DSM-5 Disorders–Clinician version
UEQ-S: Short Version of the User Experience Questionnaire

Edited by A Mavragani; submitted 13.08.24; peer-reviewed by A Eisingerich, J Blechert; comments to author 06.11.24; revised version received 09.12.24; accepted 05.02.25; published 14.04.25.

Please cite as:
Keinert M, Schindler-Gmelch L, Rupp LH, Sadeghi M, Richer R, Capito K, Eskofier BM, Berking M
The EmpkinS-EKSpression Reappraisal Training Augmented With Kinesthesia in Depression: One-Armed Feasibility Study
JMIR Form Res 2025;9:e65357
URL: https://formative.jmir.org/2025/1/e65357
doi:10.2196/65357
PMID:

©Marie Keinert, Lena Schindler-Gmelch, Lydia Helene Rupp, Misha Sadeghi, Robert Richer, Klara Capito, Bjoern M Eskofier, Matthias Berking. Originally published in JMIR Formative Research (https://formative.jmir.org), 14.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluating User Engagement With a Real-Time, Text-Based Digital Mental Health Support App: Cross-Sectional, Retrospective Study

Edward Coffield[1], PhD; Khadeja Kausar[2], MHA

[1]Department of Population Health, Hofstra University, Hempstead, NY, United States

[2]Business Intelligence and Value Based Management Analytics, Medical Affairs, Maimonides Medical Center, Brooklyn, NY, United States

Corresponding Author:
Edward Coffield, PhD
Department of Population Health
Hofstra University
255 Hofstra University
101 Oak Street Center, Room 100A
Hempstead, NY, 11549
United States
Phone: 1 516 463 7019
Email: edward.coffield@hofstra.edu

Abstract

Background: Approximately 20% of US adults identify as having a mental illness. Structural and other barriers prevent many people from receiving mental health services. Digital mental health apps that provide 24-hour, real-time access to human support may improve access to mental health services. However, information is needed regarding how and why people engage with licensed counselors through a digital, real-time, text-based mental health support app in nonexperimental settings.

Objective: This study aimed to evaluate how people engage with Counslr, a 24-hour, digital, mental health support app where users communicate in real time with human counselors through text messaging. Specifically, access patterns (eg, day of the week and time of session) and reasons for accessing the platform were examined. Furthermore, whether differences existed between session types (on-demand or scheduled) and membership types (education or noneducation) in regard to access patterns and why people accessed the platform were evaluated.

Methods: The study population (users) consisted of students whose schools, universities, or colleges partnered with Counslr and employees whose organizations also partnered with Counslr. Users participated in text-based mental health support sessions. In these sessions, users engaged with licensed counselors through digital, text-based messaging in real time. Users could initiate an on-demand session or schedule a session 24 hours a day. User engagement patterns were evaluated through session length, session day, session time, and self-reported reasons for initiating the session. The data were stratified by membership type (education [students] or noneducation [employees]) and session type (on-demand or scheduled) to evaluate whether differences existed in usage patterns and self-reported reasons for initiating sessions by membership and session types.

Results: Most students (178/283, 62.9%) and employees (28/44, 63.6%) accessed Counslr through on-demand sessions. The average and median session times were 40 (SD 15.3) and 45 minutes. On-demand sessions (37.9 minutes) were shorter (P=.001) than scheduled sessions (43.5 minutes). Most users (262/327, 80.1%) accessed Counslr between 7 PM and 5 AM. The hours that users accessed Counslr did not statistically differ by membership type (P=.19) or session type (P=.10). Primary self-reported reasons for accessing Counslr were relationship reasons, depression, and anxiety; however, users initiated sessions for a variety of reasons. Statistically significant differences existed between membership and session types (P<.05) for some of the reasons why people initiated sessions.

Conclusions: The novel findings of this study illustrate that real-time, digital mental health support apps, which offer people the opportunity to engage with licensed counselors outside of standard office hours for a variety of mental health conditions, may help address structural barriers to accessing mental health support services. Additional research is needed to evaluate the effectiveness of human-based apps such as Counslr and whether such apps can also address disparities in access to mental health support services among different demographic groups.

(JMIR Form Res 2025;9:e66301) doi:10.2196/66301

XSL•FO
RenderX

KEYWORDS

mental health support; text; app; utilization; mobile; on demand; scheduled; mHealth; mobile health; app; student; university; college; mental health; employee; job; work; occupational health; counselor; counseling; usage; engagement; self-reported

Introduction

Overview

Approximately 1 out of every 5 adults (22.8%) in the United States identified as having a mental health illness in 2021 [1]. The gap between people who identify as having a mental illness and those who receive care or have all of their mental health care needs met is substantial. Among people who identify as having any mental illness, 47.2% received mental health treatments in 2021 and 27.6% perceived themselves as having an unmet mental health care need [1]. Access to mental health care services is not consistent among groups. For example, adults who identify as non-Hispanic Black or Hispanic are less likely than adults who identify as White to receive mental health care treatments [2].

Possible factors for having an unmet mental health care need include cost, stigma, minimization of symptoms, low perceived treatment effect, and structural barriers [3,4]; these factors differ across demographic groups [3,5]. For instance, the odds of costs being a factor in having an unmet mental health care need are higher among people who identify as female relative to people who identify as male, while the odds of people stating either stigma or structural barriers as a reason for their unmet mental health care needs are higher among people who identify as Black or Hispanic relative to people who identify as White [3].

Digital-based mental health apps have the potential to help people overcome barriers to obtaining care for mental illness [6,7] and reduce unmet mental health care needs among marginalized groups [8]. For example, the ability to consult with mental health professionals at all hours of the day, from any location, and in a digital environment that may foster privacy and anonymity, could address structural barriers and stigma [9]. Furthermore, Friis-Healy et al [10] state that digital mental health tools may help expand access to people most affected by COVID-19 and systematic racism by lowering care costs; "...decreasing transportation challenges; and providing mental health care in a private, destigmatizing manner." ([10], pages 2-3).

While digital mental health apps have been found to be effective and cost-effective [11], evidence regarding the effectiveness of digital mental health apps is inconsistent across digital mental health app types (eg, guidance within an app likely matters [7,12]). Furthermore, challenges, such as user engagement [6,9,13], remain. Baumel et al [13]—in an evaluation of 93 digital mental health care apps—found that only 4% of users opened the evaluated apps daily. Accordingly, understanding how people engage with digital mental health care apps is needed to facilitate design elements that may encourage engagement. As user engagement with mental health apps may differ between practical settings and clinical trials [6,13], a need exists to understand user engagement in a practical or actual setting. Engagement [13] and effectiveness [7] may also differ by app type, indicating an additional need to understand how people engage with different types of digital mental health apps in a practical setting.

Despite the need for nonexperimental evaluations of digital mental health apps, studies [13,14] regarding how people engage with digital mental health apps in practical settings [15] remain scarce. Studies have evaluated how users interact with (eg, downloading, using, and engaging) nonexperimental mental health apps in various forms [13-17]. For instance, Baumel et al [13] calculated usage metrics (eg, number of user sessions and minutes of daily use) for mobile apps— including 2 apps with peer support as a primary technique (and 7 apps with peer support as a cotechnique, for a total of 9), where the support could be "through chat or messaging services (synchronous and asynchronous...)" ([13], Appendix 2)—with at least 10,000 installs; Booth et al [14] calculated interactions by the hour and other engagement metrics for a chatbot; and Raue et al [17] reported engagement metrics (eg, weekly text message word count and weekly text messages) among users of a text-based app where users send text messages to therapists "24/7" ([17], page 3) who then review messages during "standard working hours" and respond at least "once a day, 5 days a week" ([17], page 3). However, usage patterns may differ between app designs (eg, human-based synchronous platforms, human-based asynchronous platforms, hybrid human-based designs, artificial intelligence [AI]–based platforms, and AI and human–based platforms), creating a need for further study. Notably, information is needed in terms of how (eg, day of the week, hours of day, and length of sessions) and why users engage with human-based, nonscripted apps, where users engage with humans through text messaging in real time. The addition of this information will provide policy makers and app designers with further data to consider when designing mental health apps.

In this study, we evaluate how users engage with Counslr, a digital, text-based mental health support app where users interact with licensed counselors in real time on their mobile devices [18]. This nonexperimental study is novel in that it evaluates how users engage with a digital mental health app that provides real-time, text-based mental health support. In addition, we explore why people access the app and whether engagement patterns statistically differ among people who access Counslr through their employment relative to people who access Counslr through their schools, universities, or colleges. Evaluating how users engage with and use Counslr in this retrospective study will contribute to the literature by providing an examination of how people interact with a counselor-based, digital mental health support app in a practical setting. In addition, this study will provide insights into whether people access a real-time, human-interfacing mobile mental health support app outside of standard business hours and how usage patterns may differ between instant, on-demand access and scheduled appointment access.

XSL•FO
RenderX

Counslr

Counslr was founded in 2019 and delivers mental health support services to users located in the United States through its mobile, text-based support platform. Through the Counslr platform, licensed counselors are available for confidential, real-time, text-based assistance 24 hours a day, 365 days a year. Counslr now uses an algorithm to match users with counselors based upon a user's preferences and needs. People may also schedule future sessions with a requested counselor. In addition to the text-based support sessions, Counslr provides self-guided resources within the platform. Access to Counslr is provided through school-, employment-, and community-based organizations that purchase a subscription to Counslr. Members of these organizations have the opportunity to enroll in Counslr and access Counslr-provided services through the firm's mobile platform. Members have no cost-sharing elements or usage limitations. Users' information is private and confidential as required by the Health Insurance Portability and Accountability Act (HIPAA). The app and its data are encrypted.

Methods

Overview

Data used for this retrospective study included Counslr users from October 26, 2021 to November 22, 2023 and focused exclusively on Counslr's text-based, real-time support sessions. This period covers the initial start-up of Counslr and its corresponding initial user membership. Counslr users, in this study, included students who accessed Counslr through their schools, universities, or colleges' partnerships with Counslr and employees who accessed Counslr through their organizations' partnerships with Counslr. Counslr provided a retrospective, deidentified dataset for this study. The provided data originated from user-reported information and data generated within the Counslr platform. Before a support session begins, users complete a survey where they self-report information to assist their counselor during the session. Second, data are generated within the platform such as the date or time of the session, session type (eg, on-demand or scheduled), and session length.

The session is the unit of analysis in this study. The data did not include a variable to identify individual users; thus, adjustments were not made for users who had multiple sessions. Individual datasets were provided for many of the variables. In some instances, dates were not consistent across the individual datasets (eg, a date missing in one dataset), and in other instances, presumable entry errors were present (eg, a session length was entered in minutes instead of a percentage of an hour). We adjusted for these challenges by either dropping the date from the dataset or adjusting the error where possible.

The primary purpose of this study was to evaluate how people access a 24-hour, digital mental health support app, where people communicate in real-time with human counselors through text messaging, and whether access patterns differ by session type (on-demand or scheduled) and membership type (education or noneducation). Access patterns were evaluated by day and hour (reported to authors in hour blocks [eg, 7 AM and 8 AM], not a specific time). The day and hour session distributions were stratified, separately, by session and membership types to examine whether access patterns differed between on-demand and scheduled sessions and between education- and non–education-based users. Chi-square tests were used for these comparisons. Chi-square tests with Bonferroni corrections were used as post hoc tests when appropriate.

The second evaluation used 2-tailed *t* tests to examine if session length (minutes) statistically differed between session types as well as by membership types. The last examination concerned users' self-reported reasons for initiating their sessions. Users could select a reason or reasons for their sessions from a set list provided within the app or provide their own reason for initiating a session. The self-reported categories or reasons for initiating sessions provided in the data included academic concerns; anxiety; depression; emotional abuse; family issues; infidelity; lesbian, gay, bisexual, and transgender (LGBT); relationship reasons; substance use; no reason provided (none); and others. In addition, 3 other categories were reported that were combined with these more general categories: "depression or anxiety" (n=3) was placed in the depression category; "other or relationship or depression" (n=1) was placed in the other category; and "[relationship]" (n=3) and "relationship or depression" (n=1) were placed in the relationship category. Overall summary statistics were estimated for the session reasons. In addition, the session reasons were tabulated and stratified by session type and by membership type. Chi-square tests were used to evaluate whether statistical differences existed between session types and, separately, between membership types in terms of why users initiated their sessions.

Ethical Considerations

The data used in this study were collected during Counslr's standard operations and adhered to HIPAA guidelines. The data are not publicly available. Hofstra University's institutional review board guidelines [19] deemed this study as non–human subject research because (1) Counslr provided a deidentified secondary dataset to the authors for evaluation, (2) the authors did not participate in the data collection and were not able to link the individuals in the deidentified dataset to living individuals through a code, and (3) the authors had independent control over how the data were analyzed and the study's conclusions.

Results

In total, 2 samples were created for the study due to the information available within each individual dataset. The primary dataset consisted of 327 support sessions that occurred between October 26, 2021 and October 19, 2023. The second sample consisted of the reasons why people initiated support sessions; this dataset consisted of support sessions conducted between August 15, 2023 to November 22, 2023 and, unlike session reasons provided for earlier dates, permitted stratification by both membership type and session type. For both datasets, dates without sessions were not included as were dates with data challenges that we could not address. The majority of evaluated sessions (206/327, 63%) were on-demand. Education-based users were the primary users in the sample (283/327, 86.5% of the sessions); the remaining sessions (44/327, 13.5%) comprised people associated with

XSL•FO
RenderX

non–education-based memberships. There was no statistical difference (*P*=.93) in how education and noneducation members accessed Counslr; the majority of both groups (education: 178/283, 62.9%; noneducation: 28/44, 63.6%) accessed Counslr through on-demand sessions.

Across session and membership types, the mean and median session lengths were 40 (SD 15.3) and 45 minutes. Three sessions were noted as 0 minutes, while the longest session was 76 minutes. Mean session lengths for education- and non–education-based users were not significantly different (*P*=.57); the mean session lengths for education- and non–education-based users were 39.8 (SD 15.4; median 45, IQR 36-48) and 41.2 (SD 14.4; median 45, IQR 42.5-48.5) minutes. Differing from membership type, mean session length did statistically differ (*P*=.001) between on-demand (37.9, SD 16.6; median 45, IQR 27-47 minutes) and scheduled sessions (43.5, SD 11.9; median 46, IQR 42-49 minutes).

The share of sessions allocated to Fridays, Saturdays, and Sundays were each below 14.29% (Table 1); 14.29% is the expected share for each day if sessions were evenly distributed across all 7 days. Conversely, the share of sessions allocated to Mondays, Tuesdays, Wednesdays, and Thursdays each exceeded 14.29%. The distribution of sessions across days did not statistically differ (*P*=.62) by membership type. However, this distribution did statistically differ (*P*=.03) by session type. In post hoc tests, the distributions of session type on Thursdays and Sundays were seemingly different relative to the same distribution on other days, yet, after Bonferroni corrections, no statistical differences were present at the Bonferroni-corrected significance level of .007.

From a time perspective (Figure 1), 80.1% (262/327) of sessions were documented (in hour blocks, not specific times) as occurring from 7 PM to 5 AM, 41.6% (136/327) of sessions occurred between 7 PM and 12 AM, while 38.5% (126/327) of sessions were between 12 AM and 5 AM. This pattern was also present by membership type as 80.9% (229/283) of education-based users and 75% (33/44) of non–education-based users accessed the platform between 7 PM and 5 AM. The distribution of sessions across hours of the day did not statistically differ between session type (*P*=.10) or membership type (*P*=.19). Figure 1 illustrates these hourly-based distributions by membership type (A) and session type (B).

As indicated, a second dataset was constructed to evaluate why users initiated support sessions. Users could select multiple reasons for initiating a session; we evaluated 327 reasons in this study. While this is the same number of sessions included in the first dataset, this dataset covers a shorter time span than the first dataset and the reasons users provided for their sessions in this second dataset are for a smaller number of sessions than 327. The top 3 self-reported reasons for initiating a session (Table 2) pertained to relationships (75/327, 22.9%), depression (71/327, 21.7%), and anxiety (62/327, 19%). Among non–education-based users, the top 3 reasons for initiating a session (Table 2) concerned relationships (23/64, 35.9% of sessions), anxiety (10/64, 15.6%), and family issues (9/64, 14.1%). Among education-based users, the top 3 reasons for engaging in a session were depression (65/263, 24.7%), anxiety (52/263, 19.8%), and relationships (52/263, 19.8%).

Table 1. Distribution of user sessions by day of the week[a].

Day	All, n (%)	Membership type, n (%)		Session type, n (%)		
		Education	Noneducation	On demand	Scheduled	*P* value[b]
Monday	49 (15)	43 (15.2)	6 (13.6)	28 (13.6)	21 (17.4)	.36
Tuesday	58 (17.7)	46 (16.3)	12 (27.3)	32 (15.5)	26 (21.5)	.17
Wednesday	60 (18.4)	51 (18)	9 (20.5)	38 (18.5)	22 (18.2)	.95
Thursday	52 (15.9)	47 (16.6)	5 (11.4)	26 (12.6)	26 (21.5)	.03[c]
Friday	44 (13.5)	40 (14.1)	4 (9.1)	32 (15.5)	12 (9.9)	.15
Saturday	27 (8.3)	23 (8.1)	4 (9.1)	21 (10.2)	6 (5)	.10
Sunday	37 (11.3)	33 (11.7)	4 (9.1)	29 (14.1)	8 (6.6)	.04[c]
P value[d]			.62		.03	

[a]Sessions from October 26, 2021, to October 19, 2023 (n=327).

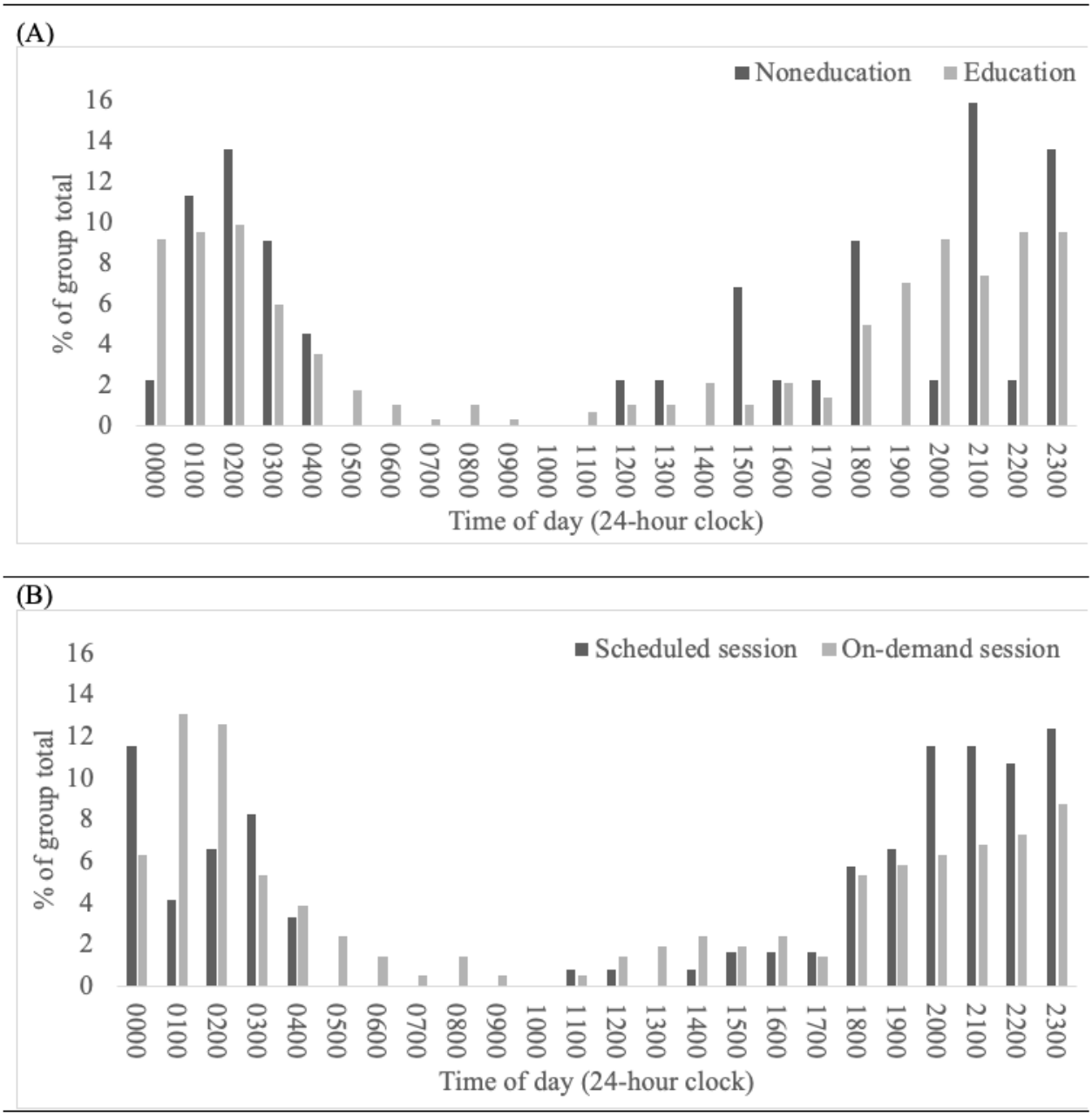
[b]*P* values from post hoc χ^2 tests; the Bonferroni-corrected significance level for these tests is .007.

[c]These *P* values are considered above the Bonferroni-corrected significance level of .007 and therefore do not support a statistically significant difference between the percentages.

[d]*P* values from χ^2 evaluations on overall distributions of sessions over days.

XSL•FO
RenderX

Figure 1. Distribution of user sessions by hour: hour of session by (A) membership type (noneducation or education) and (B) session type (scheduled or on demand). The data provided to the authors noted hours in blocks (eg, 7 AM, 8 AM) not specific times. Sessions from October 26, 2021 to October 19, 2023 (n=327); $\chi 2$ test for membership type group distribution difference: P=.19; $\chi 2$ test for session type group distribution difference: P=.10.



XSL•FO
RenderX

Table 2. Users' self-reported reasons for sessions, by session type and membership type[a].

Session reason	All, n (%)	Membership type, n (%)			Session type, n (%)		
		Education	Noneducation	*P* value[b]	On demand	Scheduled	*P* value[b]
Academic concerns	20 (6.1)	14 (5.3)	6 (9.4)	.23	14 (6.7)	6 (5.1)	.58
Anxiety	62 (19)	52 (19.8)	10 (15.6)	.45	39 (18.6)	23 (19.7)	.81
Depression	71 (21.7)	65 (24.7)	6 (9.4)	.01	53 (25.2)	18 (15.4)	.04
Emotional abuse	4 (1.2)	2 (0.8)	2 (3.1)	.12	3 (1.4)	1 (0.9)	.65
Family issues	23 (7)	14 (5.3)	9 (14.1)	.01	13 (6.2)	10 (8.6)	.42
Infidelity	3 (0.9)	3 (1.1)	0 (0)	.39	3 (1.4)	0 (0)	.19
LGBT[c]	10 (3.1)	8 (3)	2 (3.1)	.97	7 (3.3)	3 (2.6)	.70
No reason provided (none)	11 (3.4)	11 (4.2)	0 (0)	.10	0 (0)	11 (9.4)	<.001
Other	42 (12.8)	37 (14.1)	5 (7.8)	.18	29 (13.8)	13 (11.1)	.48
Relationship	75 (22.9)	52 (19.8)	23 (35.9)	.01	48 (22.9)	27 (23.1)	.96
Substance use	6 (1.8)	5 (1.9)	1 (1.6)	.86	1 (0.5)	5 (4.3)	.01

[a]Reasons (n=327) provided during sessions from August 15, 2023 to November 22, 2023. Users could select a reason or reasons for the session from the list or type another reason or reasons.

[b]*P* values from χ^2 tests.

[c]LGBT: lesbian, gay, bisexual, and transgender.

As illustrated in Table 2, a statistical difference was present in the percentage of education and noneducation users who accessed support sessions for depression (*P*=.01), family issues (*P*=.01), and relationships (*P*=.01). Table 2 also illustrates the reasons users initiated sessions by session type. The primary reasons for support sessions by session type were depression (53/210, 25.2%), relationships (48/210, 22.9%), and anxiety (39/210, 18.6%) for on-demand sessions and relationships (27/117, 23.1%), anxiety (23/117, 19.7%), and depression (18/117, 15.4%) for scheduled sessions. Users accessed support through scheduled sessions (5/117, 4.3%) at a statistically (*P*=.01) greater percentage relative to on-demand sessions (1/210, 0.5%) for substance use (as well as for the reason of none or no reason provided). Conversely, users accessed support for depression at a statistically (*P*=.04) higher percentage through on-demand sessions (53/210, 25.2%) relative to scheduled sessions (18/117, 15.4%).

Discussion

Principal Findings

Digital mental health apps have the potential to expand access to mental health services. These apps can help people overcome barriers to obtaining such care [6-8]. In this study, we analyzed Counslr, a digital mental health app that offers real-time, text-based support sessions with human counselors. We found that users who accessed Counslr participated in support sessions for an average of 40 minutes. The majority of these sessions occurred during times when standard, human-based mental health support is not available; this salient finding illustrates the promise of human-based, digital mental health apps in overcoming structural barriers to accessing mental health support services.

The novelty of this study in terms of evaluating user engagement with a human-based, nonscripted app, where users engage with human counselors through text messaging in a real-time, nonexperimental setting, prevents multiple, direct comparisons with the peer-reviewed literature. However, a few comparisons on similar metrics illustrate how usage patterns may differ between app designs. For instance, differing from this study's result that the majority of sessions occurred between 7 PM and 5 AM, Booth et al [14], evaluating the usage of a chatbot, found that peak interaction hours with the chatbot were in the hours of 8-10 AM, 1 PM, and 5 PM. However, closer to this study's results, Baumel et al [13]—evaluating 93 apps—found that usage peaked at the 8 PM hour for the apps included in their mental health category, while apps in their mindfulness and meditation category peaked in the hours of 7-9 AM and 10 PM to 12 AM. Baumel et al [13] also illustrated, in a graph, that usage rates for apps in their mental health category were highest on Tuesdays, Wednesdays, and Thursdays, which is similar to this study's overall, nonstratified results (statistical comparisons were not conducted on this study's overall, nonstratified day of the week statistics); the peak day of usage for the mindfulness or meditation applications Baumel et al [13] reviewed was Thursday.

Raue et al [17], evaluating a text-based app that included human interactions (users could text at any time in their study, while therapists reviewed the texts during working hours and responded at least 1 time a day to a person's text 5 days a week; [17], page 3), did not report directly comparable metrics to those illustrated in this study. However, from a time of session perspective, Baumel et al [13] found that the median daily minutes with which active users engaged in 2 peer-supported apps they reviewed was 35.08; a finding below the 45-minute median found in this study. Overall, Baumel et al [13] found

XSL•FO
RenderX

that the median daily minutes of use per active app user was different between mental health apps (10.02 minutes), happiness apps (7.77 minutes), and mindfulness and meditation apps (21.47 minutes), demonstrating that the reason why people engage with an app may influence usage.

The primary reasons users of Counslr self-reported for initiating sessions were similar to mental health conditions found across the population. For instance, 2 of the primary reasons Counslr users noted for initiating sessions, anxiety and depression, were also the most prevalent mental health conditions found among college students [20], while symptoms related to anxiety disorders or depressive disorders were present in 20.7% of US adults in early 2024 [21]. Collectively, how Counslr users engaged with licensed counselors outside of standard office hours, coupled with the commonality between mental health conditions found in the general population and the reasons Counslr users initiated support sessions, provide evidence that people will use a human-based, digital mental health app to access support outside of standard hours for prevalent mental health conditions.

While we did not have extensive demographic information about the Counslr users whose sessions were evaluated within this study, the 2 different membership types allowed comparisons between people who accessed the platform as students and those who accessed the platform as employees. When examined, usage patterns between these 2 groups did not statistically differ. Both students and employees engaged in sessions primarily between 7 PM and 5 AM and had similar patterns in terms of the days on which their sessions were held. Furthermore, the majority of both student and employee users accessed the platform using the on-demand feature instead of scheduling an appointment. The access patterns found among both student and employee users indicate that mental health support apps have the potential to increase access to mental health support services across multiple groups. This potential, universal appeal of Counslr-like platforms is also seen in how the platform is able to provide support for a wide array of reasons: from anxiety and depression to infidelity and academic concerns.

As noted here and elsewhere [9], the 24-hour accessibility of mental health apps is one attribute of such apps that could decrease structural barriers to accessing support services to address mental health challenges. However, the salient findings here regarding on-demand and scheduled sessions demonstrate that how access to such apps is provided is also important in terms of app design. For instance, there were no statistical differences between the frequency of on-demand and scheduled sessions in terms of how the sessions were distributed across the hours of the day; people both scheduled sessions and initiated on-demand sessions at 4 AM, indicating a need for both types of sessions.

Differences between the 2 session types also illustrate the need to consider offering both on-demand and scheduled sessions within a digital mental health support app; doing such may help facilitate the apps' ability to improve access to mental health support. For instance, the distribution of sessions across days of the week differed between on-demand sessions and scheduled sessions, indicating possible user preferences for set sessions with counselors or on-demand sessions on certain days. Furthermore, a larger share of sessions initiated for the reason of substance use occurred as scheduled sessions relative to on-demand sessions. This finding suggests that users may prefer scheduled sessions for substance use relative to on-demand sessions, indicating that during the app design stage, developers should consider whether available access modes (on-demand and scheduled) address how users may prefer different access points depending on the specific reason they are accessing the app.

The connection between app design and this study's results was possible because Counslr shared their formative data with researchers for independent analysis. As Counslr and other platforms continue to share their data for analyses, additional insights into future app development may arise; these insights may result in increased efficiency in app development as well as the creation of platforms designed to address the needs of users, thereby possibly increasing usage among people who seek mental health support thought such apps.

In this study, we conducted a novel evaluation of a 24-hour, real-time, digital mental health support app where users interact with humans through text messaging. The data were from a nonexperimental situation, helping to address the stated need for more information regarding how digital mental health apps perform outside of experimental settings [6,13,15,16]. While the findings of this study illustrate the promise of digital mental health support apps to increase access to a variety of mental health needs, a number of limitations are present. First, this study relied on a small, formative dataset provided by Counslr for independent analysis. While a larger sample size is preferable, the study allowed multiple insights into usage patterns and why people accessed the platform. Furthermore, using this formative dataset to evaluate initial platform usage allowed the authors to suggest additional metrics for data collection and analysis to the organization.

A second limitation is that the deidentified data used here and the lack of demographic variables entail that the findings of this study only pertain to study participants and are not generalizable to the overall population. Furthermore, the data did not include individual identifiers, preventing adjustments for users with multiple sessions, which could skew the results (eg, if one user initiated multiple sessions for the same reason). To provide further insight into this limitation, Counslr, at the authors' request, provided information illustrating that there were 154 unique users in the data with an average and median number of sessions of 2.06 and 1 (minimum 1 and maximum 21).

The deidentified nature of the data and the manner in which the data were provided to the authors also prevented more complex data analysis (eg, regression analysis), which could provide further insights into usage and access patterns. In addition, the platform did not collect demographic information on users such as how they identify their ethnicity, gender, and race. The inclusion of such information would allow pivotal evaluations regarding whether digital mental health support apps can help address disparities in access to mental health support services across different demographic groups. Finally, while the primary purpose of this study was to evaluate how and why people

XSL•FO
RenderX

engage with licensed counselors through a digital, real-time, text-based mental health support app in nonexperimental settings, future evaluations—when such data are available—must examine the effectiveness of such apps in addressing the mental health support needs of users and whether such platforms serve as a complement to more traditional mental health services.

Conclusion

Barriers that prevent people from accessing mental health support services are multiple and include structural barriers and stigma. Anonymity and 24-hour access to digital mental health support apps may help address such barriers. In this study, we evaluated how people use Counslr, a real-time mental health support app where users engage with licensed, human counselors through text messaging sessions. The majority of both education- and non–education-based users initiated mental health support sessions outside of standard business hours for a variety of reasons, indicating the promise of Counslr like apps to address barriers to accessing mental health support services. The similarities in usage patterns of Counslr student and employee users, despite one group being noneducation based and the other being education based, demonstrated an almost universality in the platform in terms of serving multiple groups with multiple needs.

The access patterns found in this study demonstrate the promise of Counslr-type apps to help address structural barriers to accessing mental health support services. Additional research is needed to evaluate the effectiveness of such apps, notably human-based apps such as Counslr, and whether such apps can also address disparities in access to mental health support services among demographic groups.

Acknowledgments

Artificial intelligence was not used in the composition of this manuscript.

Data Availability

The data analyzed in this study are not publicly available. The data were provided to the authors by Counslr.

Conflicts of Interest

EC and KK have no financial interests in Counslr and did not receive any form of payment for this work. EC, in his role as an academic program director and faculty member, works closely on multiple items with another faculty member who is a Counslr board member. KK has also interacted with this faculty member. Study data were provided by Counslr. Members of the Counslr team answered questions regarding the platform or data (or both) and reviewed the manuscript. However, EC and KK conducted the study independently from Counslr. The conclusions provided in the manuscript are those of EC and KK.

References

1. Center for behavioral health statistics and quality, results from the 2021 national survey on drug use and health. Substance Abuse and Mental Health Services Administration. URL: https://www.samhsa.gov/data/report/2021-nsduh-detailed-tables [accessed 2025-02-05]
2. Terlizzi EP, Norris T. Mental health treatment among adults: United States. NCHS Data Brief 2021(419):1-8. [doi: 10.15620/cdc:110593]
3. Alang SM. Sociodemographic disparities associated with perceived causes of unmet need for mental health care. Psychiatr Rehabil J 2015;38(4):293-299. [doi: 10.1037/prj0000113] [Medline: 25664758]
4. Yang JC, Roman-Urrestarazu A, McKee M, Brayne C. Demographic, socioeconomic, and health correlates of unmet need for mental health treatment in the United States, 2002-16: evidence from the national surveys on drug use and health. Int J Equity Health 2019;18(1):122 [FREE Full text] [doi: 10.1186/s12939-019-1026-y] [Medline: 31382979]
5. Green JG, McLaughlin KA, Fillbrunn M, Fukuda M, Jackson JS, Kessler RC, et al. Barriers to mental health service use and predictors of treatment drop out: racial/ethnic variation in a population-based study. Adm Policy Ment Health 2020;47(4):606-616 [FREE Full text] [doi: 10.1007/s10488-020-01021-6] [Medline: 32076886]
6. Lipschitz JM, Pike CK, Hogan TP, Murphy SA, Burdick KE. The engagement problem: a review of engagement with digital mental health interventions and recommendations for a path forward. Curr Treat Options Psychiatry 2023;10(3):119-135 [FREE Full text] [doi: 10.1007/s40501-023-00297-3] [Medline: 38390026]
7. Mohr DC, Azocar F, Bertagnolli A, Choudhury T, Chrisp P, Frank R, Banbury Forum on Digital Mental Health. Banbury forum consensus statement on the path forward for digital mental health treatment. Psychiatr Serv 2021;72(6):677-683 [FREE Full text] [doi: 10.1176/appi.ps.202000561] [Medline: 33467872]
8. Ramos G, Ponting C, Labao JP, Sobowale K. Considerations of diversity, equity, and inclusion in mental health apps: a scoping review of evaluation frameworks. Behav Res Ther 2021;147:103990 [FREE Full text] [doi: 10.1016/j.brat.2021.103990] [Medline: 34715396]
9. Aschbacher K, Rivera LM, Hornstein S, Nelson BW, Forman-Hoffman VL, Peiper NC. Longitudinal patterns of engagement and clinical outcomes: results from a therapist-supported digital mental health intervention. Psychosom Med 2023;85(7):651-658. [doi: 10.1097/PSY.0000000000001230] [Medline: 37409793]

XSL•FO
RenderX

10. Friis-Healy EA, Nagy GA, Kollins SH. It is time to REACT: opportunities for digital mental health apps to reduce mental health disparities in racially and ethnically minoritized groups. JMIR Ment Health 2021;8(1):e25456 [FREE Full text] [doi: 10.2196/25456] [Medline: 33406050]
11. Moshe I, Terhorst Y, Philippi P, Domhardt M, Cuijpers P, Cristea I, et al. Digital interventions for the treatment of depression: a meta-analytic review. Psychol Bull 2021;147(8):749-786. [doi: 10.1037/bul0000334] [Medline: 34898233]
12. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
13. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
14. Booth F, Potts C, Bond R, Mulvenna M, Kostenius C, Dhanapala I, et al. A mental health and well-being chatbot: user event log analysis. JMIR Mhealth Uhealth 2023;11:e43052 [FREE Full text] [doi: 10.2196/43052] [Medline: 37410539]
15. Carlo AD, Hosseini Ghomi R, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. NPJ Digit Med 2019;2:54 [FREE Full text] [doi: 10.1038/s41746-019-0129-6] [Medline: 31304400]
16. Balaskas A, Schueller SM, Cox AL, Rashleigh C, Doherty G. Examining young adults daily perspectives on usage of anxiety apps: a user study. PLOS Digit Health 2023;2(1):e0000185 [FREE Full text] [doi: 10.1371/journal.pdig.0000185] [Medline: 36812622]
17. Raue PJ, Fridling N, Song J, Hull TD, Alexopoulos GS, Arean PA. Message-based psychotherapy for older adults: a cohort comparison study. Front Psychiatry 2022;13:951354 [FREE Full text] [doi: 10.3389/fpsyt.2022.951354] [Medline: 36090371]
18. Counslr. URL: https://www.counslr.com/ [accessed 2025-02-05]
19. IRB Decision Chart. Hofstra University. URL: https://www.hofstra.edu/pdf/about/administration/provost/orsp/osrp_hofstra_irb_decision_chart.pdf [accessed 2025-04-14]
20. Oswalt SB, Lederer AM, Chestnut-Steich K, Day C, Halbritter A, Ortiz D. Trends in college students' mental health diagnoses and utilization of services, 2009-2015. J Am Coll Health 2020;68(1):41-51. [doi: 10.1080/07448481.2018.1515748] [Medline: 30355071]
21. National Center for Health Statistics. Anxiety and depression. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/nchs/covid19/pulse/mental-health.htm [accessed 2024-02-05]

Abbreviations

HIPAA: Health Insurance Portability and Accountability Act
LGBT: lesbian, gay, bisexual, and transgender

Edited by A Mavragani; submitted 09.09.24; peer-reviewed by J Rhodes; comments to author 07.10.24; revised version received 25.11.24; accepted 16.01.25; published 14.04.25.

Please cite as:
Coffield E, Kausar K
Evaluating User Engagement With a Real-Time, Text-Based Digital Mental Health Support App: Cross-Sectional, Retrospective Study
JMIR Form Res 2025;9:e66301
URL: https://formative.jmir.org/2025/1/e66301
doi:10.2196/66301
PMID:

©Edward Coffield, Khadeja Kausar. Originally published in JMIR Formative Research (https://formative.jmir.org), 14.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Acceptability, Usability, and Insights Into Cybersickness Levels of a Novel Virtual Reality Environment for the Evaluation of Depressive Symptoms: Exploratory Observational Study

Sara Sutori[1*], MSc; Emma Therése Eliasson[1*], PhD; Francesca Mura[2], PhD; Victor Ortiz[3], MSc; Vincenzo Catrambonephd[4], PhD; Gergö Hadlaczky[1], PhD; Ivo Todorov[1], PhD; Antonio Luca Alfeo[4], PhD; Valentina Cardi[5], PhD; Mario G C A Cimino[4], PhD; Giovanna Mioni[5], PhD; Mariano Alcañiz Raya[3], PhD; Gaetano Valenza[4], PhD; Vladimir Carli[1*], MD, PhD; Claudio Gentili[5*], MD, PhD

[1]National Centre for Suicide Research and Prevention, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

[2]Padova Neuroscience Center, University of Padua, Padua, Italy

[3]Instituto Universitario de Investigación en Tecnología Centrada en el Ser Humano, Universitat Politècnica de València, Ciudad Politécnica de la Innovación, València, Spain

[4]Research Center E. Piaggio, Department of Information Engineering, University of Pisa, Pisa, Italy

[5]Department of General Psychology, University of Padua, Padua, Italy

[*]these authors contributed equally

Corresponding Author:
Emma Therése Eliasson, PhD
National Centre for Suicide Research and Prevention
Department of Learning, Informatics, Management and Ethics
Karolinska Institutet
Granits väg 4
Stockholm, 171 65
Sweden
Phone: 46 701407295
Email: emma.eliasson@ki.se

Abstract

Background: There is a clear need for enhanced mental health assessment, depressive symptom (DS) evaluation being no exception. A promising approach to this aim is using virtual reality (VR), which entails the potential of adding a wider set of assessment domains with enhanced ecological validity. However, whilst several studies have used VR for both diagnostic and treatment purposes, its acceptance, in particular how exposure to virtual environments affects populations with psychiatric conditions remains unknown.

Objective: This study aims to report on the acceptability, usability, and cybersickness levels of a pilot VR environment designed for the purpose of differentiating between individuals with DSs.

Methods: The exploratory study, conducted in Italy, included 50 healthy controls and 50 young adults with mild-to-moderate DSs (without the need for a formal diagnosis). The study used an observational design with approximately 30 minutes of VR exposure followed by a self-report questionnaire battery. The battery included a questionnaire based on the Theoretical Framework of Acceptability, the System Usability Scale as well as the Simulator Sickness Questionnaire.

Results: Results indicate that the majority found VR acceptable for the purposes of mental health screening and treatment. However, for diagnostics, there was a clear preference for VR to be used by mental health professionals as a supplementary tool, as opposed to a stand-alone solution. In practice, following exposure to the pilot VR environment, generally, good levels of acceptability and usability were reported, but areas in need of improvement were identified (such as self-efficacy). Self-reported cybersickness levels were comparable to literature averages but were considerably higher among those with DSs.

Conclusions: These findings raise questions about the potential interplay between underlying somatic symptoms of depression and VR-induced cybersickness and call for more attention from the scientific community both in terms of methodology as well as potential clinical and theoretical implications. Conclusively, user support indicates a potential for VR to aid mental health

XSL•FO
RenderX

assessment, but further research is needed to understand how exposure to virtual environments might affect populations with varying severity and other forms of psychiatric symptoms.

International Registered Report Identifier (IRRID): RR2-10.1186/ISRCTN16396369

(JMIR Form Res 2025;9:e68132) doi:10.2196/68132

KEYWORDS

depression; virtual reality; assessment; acceptability; usability; cybersickness

Introduction

Background

In 2019, an estimated 970 million people worldwide, or 12.6% of the global population, were living with a mental disorder, with anxiety (301 million) and depression (280 million) being the most prevalent [1]. These conditions contribute substantially to the global burden of disease, with depression being one of the leading causes of disability worldwide [2]. Despite considerable efforts to address this issue, there has been no substantial reduction in the burden of mental disorders since 1990 [1].

Depressive disorders (hereafter referred to as depression) present significant challenges in research, among other reasons due to the diagnostic uncertainty [3], characterized by both high heterogeneity and low reliability [4]. The persistence of these challenges over different iterations of the *DSM* (*Diagnostic and Statistical Manual of Mental Disorders*) [4] underscores the need to advance clinical practice, which currently predominantly relies on self-report questionnaires and clinical interviews [5].

In response to the challenges mentioned above, suggestions have been made to include a broader range of measures that cover more assessment domains and are not solely dependent on verbal self-report [6,7]. Such measures include speech, text, and facial expression analysis [8]; genomics, transcriptomics, proteomics, metabolomics, and imaging [9]; behavioral and physiological variables collected via wearable sensors [10] or virtual reality (VR) systems [7].

VR systems are particularly relevant to mental health disorders due to immersive, controlled, and customizable environments that can be tailored to individual needs [7]. VR has been used across psychiatric pathologies but reviews consistently report that depression has remained understudied compared to other conditions [7,11,12]. The gap in the literature is particularly evident for studies on assessment, of which we could only find three [13-15]. These are specifically focusing on attentional processes [14], spatial navigation [15], and social interactions [13] in relation to depression. Despite these pioneering studies, the potential of VR to integrate multiple domains of data for enhancing assessment [7] has yet to be explored.

While efforts are being dedicated to improving treatment and the testing of various measures to enhance assessment, including the use of VR, even less attention has been paid to the evaluation of such novel approaches from a user perspective. Research on VR for depression treatment and assessment is limited, with even fewer studies focusing on its acceptability, usability, and tolerability. Beyond diagnostic accuracy, evaluating the acceptability and usability of diagnostic tools is crucial, particularly when involving individuals with mental health conditions, like depression. In such cases, where patients might face additional treatment barriers, including motivational difficulties, stigma, delayed help-seeking [16], or poor treatment adherence [17], acceptability and usability play a crucial role in the successful implementation of new diagnostic and treatment technologies.

Sekhon et al [18] define acceptability as a "multifaceted construct that reflects the extent to which people delivering or receiving a health care intervention consider it to be appropriate, based on anticipated or experienced cognitive and emotional responses to the intervention." On the other hand, usability has various definitions and approaches [19], but here we will refer to the extent to which the VR system can be used with ease for the purpose of assessing depression symptom severity. Third, tolerability may be defined in different ways, but here will refer to cybersickness severity. Cybersickness describes a visually induced motion sickness-like discomfort during or following exposure to VR. Cybersickness is accompanied by changes in the activity of the central and autonomic nervous systems and results in symptoms like headache, nausea, and eyestrain among others [20].

To the best of our knowledge, only 2 studies using VR have reported on the earlier concepts in relation to depression: one focusing on psychoeducation [21] and another on behavior activation [22]. They both found generally high acceptability and satisfaction levels. Migoya et al [21] also found favorable ratings for ease of use and perceived usefulness, whereas Paul et al [22] reported a decreasing interest in continuing the program after having attended multiple sessions. Paul et al [22] also assessed cybersickness, noting decreasing symptoms as the sessions progressed, but their unconventional scoring method (adding up raw scores as opposed to using weight adjustments) complicates comparisons with studies on healthy controls (HCs) [23] or other psychiatric conditions [24]. To date, no studies have compared cybersickness levels between depressed individuals and HCs.

To tackle the aforementioned challenges, this study introduces a novel approach aimed at enhancing the quality—including the breadth and ecological validity—of depression assessment. A VR system and environment were developed to capture physiological data (heart rate, skin conductance, and eye tracking), behavioral data (curiosity-related behavior and attentional biases), and cognitive data (working memory, processing speed, sustained attention, cognitive flexibility, self-evaluation, and metacognitive sensitivity) hypothesized to be associated with depression [25]. These multidomain data are

XSL•FO
RenderX

then analyzed via machine learning models and explainable artificial intelligence to assess depressive symptom (DS) severity. The categorization accuracy of this system is reported on in a dedicated study [25], however, given the innovative nature of this approach and the gap in the literature, it is fundamental to consider the patient perspective as well. Therefore, the objective of this study was to evaluate the acceptability, usability, and cybersickness levels of this pilot virtual environment.

Aim

The aim of the study was to assess the acceptability, usability, and cybersickness levels of a pilot VR environment designed to aid and enrich the assessment of DSs.

Methods

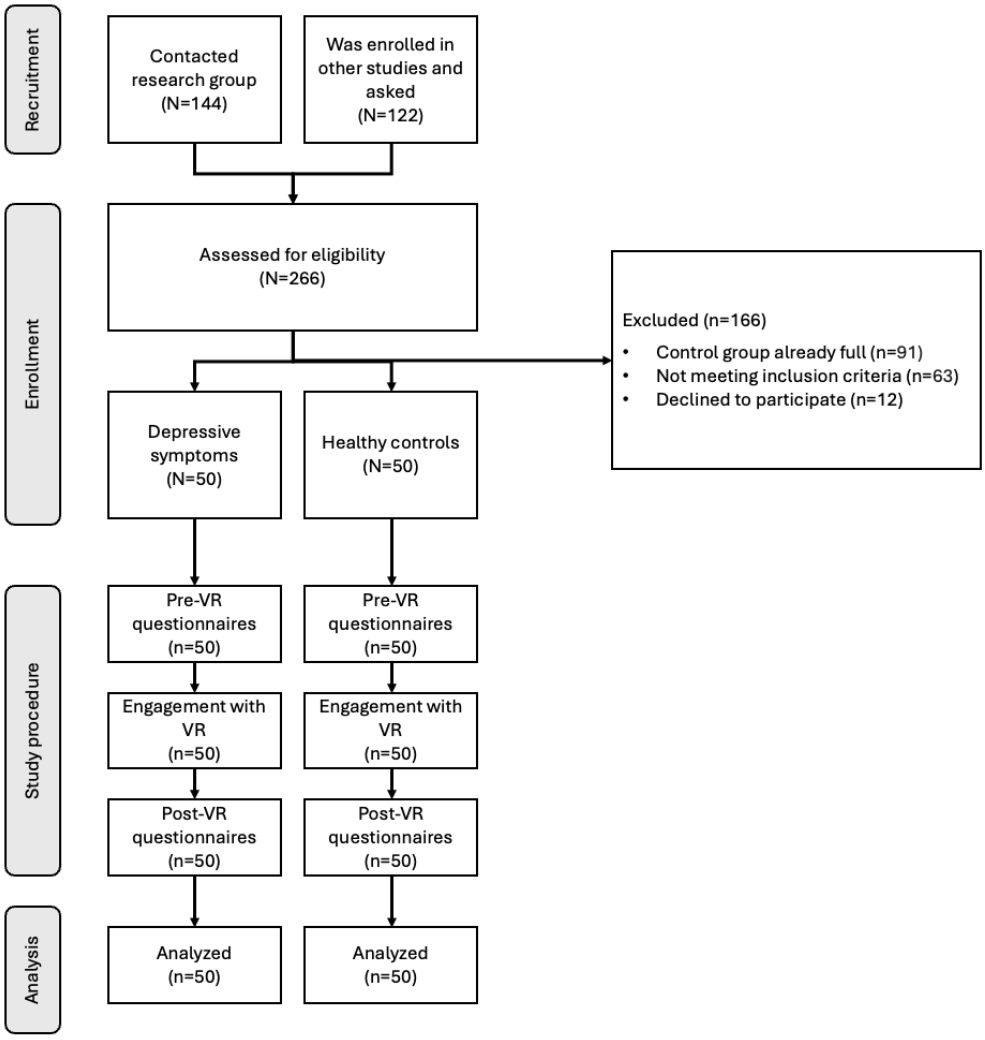
Overview

The study had an exploratory nature and followed an observational design, where all participants completed the same questionnaire batteries before and post VR and followed the same VR exposure protocol.

Sample

A total of 266 individuals were screened at the University of Padua, Italy, with continuous enrollment between November 2022 and November 2023 (Figure 1). The screening concerned the inclusion and exclusion criteria detailed later and was done using self-report questionnaires administered in Qualtrics (Qualtrics, Inc). Recruitment was completed when the predefined target sample size of 100 was reached (Figure 1), including 50 participants with DSs and 50 participants classed as HCs. Participants were aged 18-35 years, fulfilled all inclusion criteria, and provided written informed consent. Further demographic and clinical characteristics of the sample are outlined in Table 1.

Figure 1. Consort flow diagram of participant involvement. VR: virtual reality.



XSL•FO
RenderX

Table 1. Demographic and clinical characteristics of study participants by study groups.

	DS[ab] (N=50)	HC[c] (N=50)	Between group difference, r_{rb}[d]
Sex, n (%)			
Male	10 (20)	10 (20)	—[e]
Female	40 (80)	40 (80)	—
Age (years), mean (SD)	23.0 (1.86)	23.3 (1.52)	0.13
Education (years), mean (SD)	16.1 (1.40)	16.6 (0.78)	0.14
Hand dominance, n (%)			
Left	0 (0)	3 (6)	—
Right	50 (100)	47 (94)	—
Sight correction, n (%)			
None	21 (42)	26 (52)	—
Lenses	10 (20)	10 (20)	—
Glasses	19 (38)	14 (28)	—
Height (meter), mean (SD)	1.67 (0.07)	1.66 (0.07)	0.07
Weight (kg), mean (SD)	62.7 (10.60)	59.6 (7.85)	0.12
BMI, mean (SD)	22.4 (3.00)	21.5 (2.06)	0.13
Habitual sleep (hours/night), mean (SD)	7.0 (0.98)	7.3 (0.88)	0.20
Habitual smoker, n (%)			
Yes	19 (38)	20 (40)	—
No	31 (62)	30 (60)	—
Habitual consumption of alcohol, n (%)			
Yes	11 (22)	16 (32)	—
No	39 (78)	34 (68)	—
Habitual user of psychoactive drugs, n (%)			
Yes	11 (22)	4 (8)	—
No	39 (78)	46 (92)	—
PHQ-9[f], mean (SD)	11.3 (1.92)	3.2 (1.48)	0.87[g]
DASS-D[h], mean (SD)	15.2 (7.40)	3.5 (3.44)	0.77[g]
DASS-A[i], mean (SD)	9.6 (6.62)	2.4 (2.89)	0.61[g]
DASS-S[j], mean (SD)	17.8 (6.73)	7.7 (5.08)	0.66[g]
PANAS-Pos[k] before VR[l], mean (SD)	10.4 (3.53)	12.2 (2.76)	0.27[m]
PANAS-Pos after VR, mean (SD)	8.9 (4.36)	11.8 (3.61)	0.36[g]
PANAS-Neg[n] before VR, mean (SD)	3.2 (2.98)	1.9 (2.22)	0.24[o]
PANAS-Neg after VR, mean (SD)	3.2 (3.25)	1.7 (2.24)	0.24[o]

[a]The categorization is based on current depressive symptom severity assessment via the Patient Health Questionnaire-9 (score equal to 9 or above) and not the presence of a clinical diagnosis.

[b]DS: depressive symptom.

[c]HC: healthy control.

[d]Group differences are calculated via 2-tailed Mann Whitney *U* tests and effect sizes reflect rank-biserial (r_{rb}) correlations.

[e]—: not applicable.

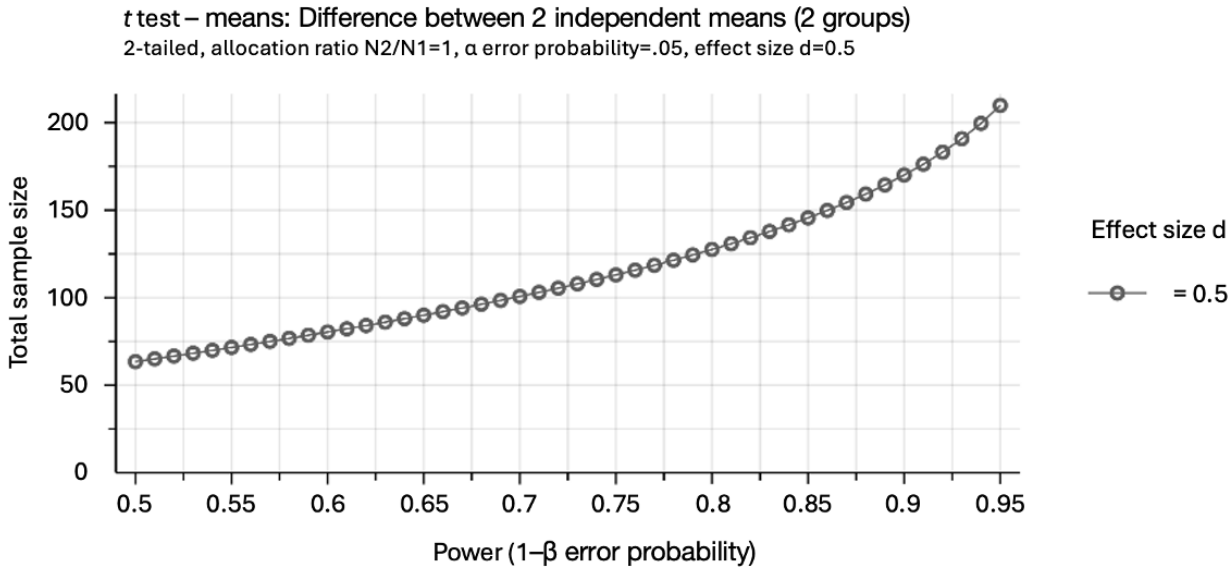
XSL•FO
RenderX

[f]PHQ-9: Patient Health Questionnaire-9.
[g]*P*<.001.
[h]DASS-D: Depression Anxiety Stress Scale-Depression.
[i]DASS-A: Depression Anxiety Stress Scale-Anxiety.
[j]DASS-S: Depression Anxiety Stress Scale-Stress.
[kj]PANAS-Pos: Positive and Negative Affect Schedule – Positive affect
[l]VR: virtual reality.
[m]*P*<.01,
[n]PANAS-Neg: Positive and Negative Affect Schedule – Negative affect
[o]*P*<.05,

Power Calculation

A post hoc power analysis was conducted to assess the statistical power of the study based on the observed sample size of 100 participants (50 per group). Assuming a moderate effect size (Cohen *d*=0.50) and an α level of .05, the analysis indicated that the study has 0.7 power to detect between-group differences as visualized in Figure 2.

Figure 2. Power analysis for between-group comparison (GPower).



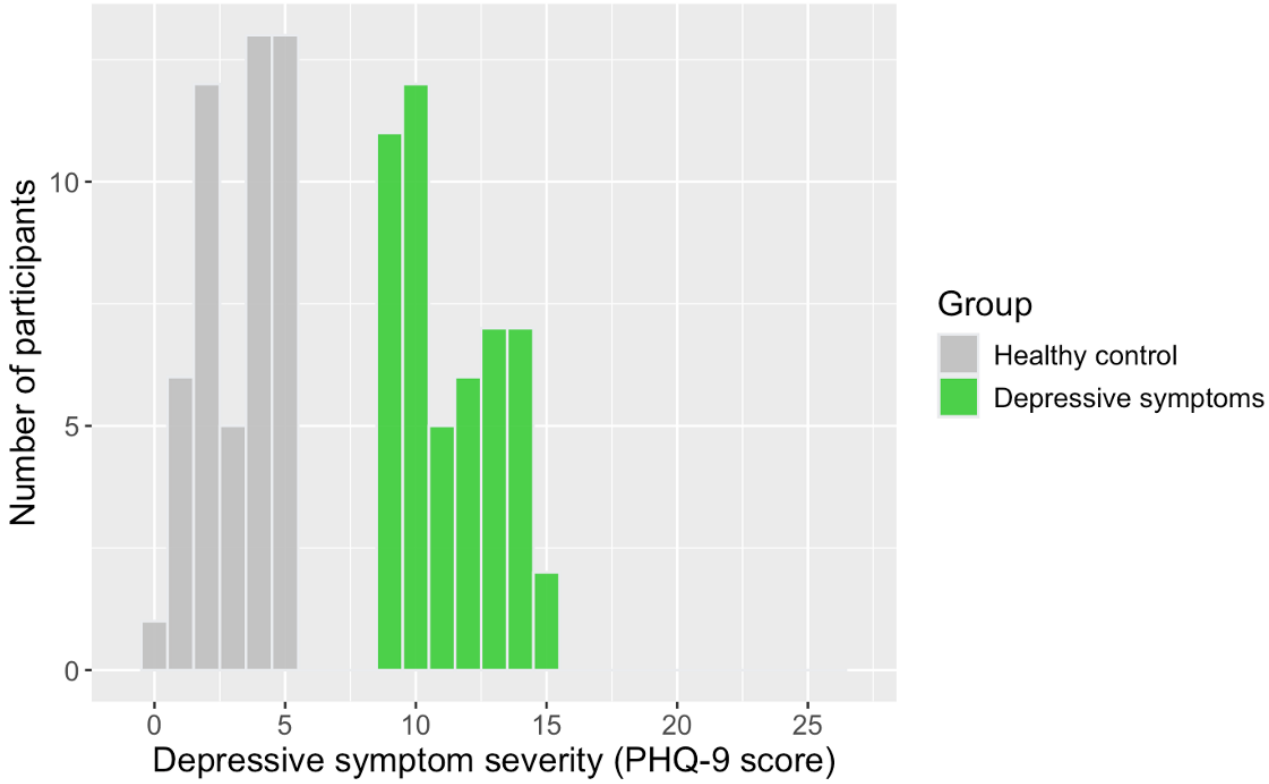
Inclusion and Exclusion Criteria

The screening for inclusion involved the assessment of current DS severity using the Patient Health Questionnaire-9 (PHQ-9; Kroenke et al [26]). Participants scoring 9 or above—constituting the upper limit for mild DSs [26]—were assigned to the DS, while those scoring 5 or below were registered for the HC group (Figure 3).

For inclusion in either group, participants had to be free from any condition that would impair their ability to interact with the VR environment (eg, visual impairment without the possibility of correction via lenses) and able to provide written informed consent. Additional inclusion criteria for the DS group included the lack of diagnosis of any other psychiatric disorders than depressive or anxiety disorders, and treatment stability over the last 4 weeks if they received psychological or pharmacological treatment at the time of assessment. Participants in the control group had to be free from any previous or current psychiatric disorders.

XSL•FO
RenderX

Figure 3. Distribution of participants' scores on PHQ-9 assessing current depressive symptom severity. PHQ-9: Patient Health Questionnaire-9.



Study Procedure

After providing informed consent and completing the pre-VR battery questionnaires, all participants were seated in a room equipped with the VR system. Following a headset and eye-tracking calibration, a tutorial was completed (aimed to help the understanding of how to move around and interact with the environment). Next, participants were free to explore the VR environment on their own terms but in the presence of a study administrator. The study session was concluded once the post-VR battery was completed.

Technological Specifications

Overview

The VR system was developed using the HP Reverb G2 Omnicept Edition headset, which includes a variety of biometric sensors such as eye tracking, a facial camera (lower), a heart rate sensor, and sensors for brain activity while maintaining the core features of the standard version. These features include a resolution of 2160 × 2160 px per eye, a 114° field of view, a 90 Hz refresh rate, and integrated audio.

Environment Development

The virtual environment was developed entirely in Unity 2020.3.39 LTS, using the HP Omnicept SDK and OpenXR to maximize compatibility and platform versatility. Several development processes were used to ensure a smooth and efficient environment, including occlusion culling, foveated rendering, baked lighting, and asset optimization.

The scenario consists of 4 rooms, each with distinct lighting parameters, as well as an additional room that can be visually explored through a keyhole in a door. Volume postprocessing was used to modify the lighting without affecting performance, and the keyhole effect was simulated using camera layers.

Within each room, various stimuli and interactive elements are present, such as sound boxes, a chalkboard, and interactable objects (eg, items that can be picked up and thrown, musical buttons, and photos that can be picked up for closer inspection). Every interaction between the participant and the environment is recorded, including eye tracking data, particularly in relation to elements designated as areas of interest or with which the subject can interact (eg, ENTERING_ROOM, LEAVING_ROOM, TAKE, DROP, TRIGGER, INTERACT).

Administrator Control

To manage the experiment, the study administrator is provided with an interface, enabling them to observe the participant's actions in real-time. The interface also offers various options to ensure that the experimental session proceeds smoothly (eg, it is possible to discontinue a cognitive task if the participant requests so). This control system is facilitated by an additional camera mounted on the subject, enabling the administrator to monitor the experiment's progression effectively.

Virtual Environment

The goal of the pilot VR environment is to differentiate between HCs and those with DSs based on cognitive, behavioral, and physiological data. The development and performance of the system for this purpose will be discussed in more detail in another paper [25], but the main characteristics will be summarized here for a general overview.

The environment was designed to elicit certain behaviors and record measures, some of which were previously shown, and others are hypothesized to be connected to depression [25]. The

XSL•FO
RenderX

domains included cognition (attention, working memory, processing speed, executive functioning, and cognitive flexibility), metacognition, persistence or grit, curiosity, as well as behavioral and attentional biases toward negative stimuli, speed, and movement measurements. Physiological data included skin conductance, heart rate variability, and eye-tracking–related measures.

The environment resembled a multiroom family home (Figure 4). Exposure started with a tutorial introducing participants to controllers as well as how to move around in the environment and interact with objects. Following the completion of the tutorial, participants were free to explore the environment but had to solve cognitive tasks to open doors and move between rooms. There were 4 cognitive tasks in total: Rapid Visual Information Processing; N-back [2-back]; Trail Making Test—parts A and B; and Wisconsin Card Sorting Test. The exposure ended on the participant's own terms, but the exit door only opened once all cognitive tasks were completed.

Multimodal data collected during the VR sessions for the purpose of DS assessment were analyzed using a CatBoost machine learning algorithm (Yandex LLC) [27]. Model performance was evaluated based on accuracy, and feature importance was determined through an explainable artificial intelligence approach [28], ensuring transparency in the decision-making process. For more details, see the study by Mura et al [25].

Figure 4. The visual illustration of the virtual reality environment designed for the assessment of cognition and behavior hypothesized to be related to depression.

Tools and Questionnaires

Pre-VR Questionnaires

The pre-VR battery included a background questionnaire (demographic data and diagnosis, medication); the Patient Health Questionnaire (PHQ-9) [26] to assess current DS severity; the Depression, Anxiety, Stress Scale (DASS) [29] to assess current depressive, anxiety and stress symptom severity; the Positive and Negative Affect Schedule (PANAS-SF) [30] to evaluate emotional state; as well as the Curiosity and Explorations Inventory-II [31].

Post-VR Questionnaires

The post-VR battery repeated the PANAS-SF questionnaires and included questionnaires regarding, acceptability, usability, and cybersickness, described in more detail in the later sections.

Acceptability

The construct of acceptability was assessed based on the theoretical framework of acceptability (TFA) [18,32]. The acceptability of the concept and the acceptability of the pilot system were evaluated separately. The acceptability of the concept, that is, the use of VR for the purposes of mental health screening, diagnosis, and treatment (definitions were provided) was assessed via self-report Likert items. Scores ranged from 1 to 5, indicating answers from "Strongly disagree" to "Strongly agree" on statements about VR being acceptable (Multimedia Appendix 1). As the primary purpose of the pilot system was the assessment of DSs, the diagnostic use case was investigated via 2 additional questions.

- Support tool: Willingness to engage with the system to inform diagnosis made by a mental health professional.
- Stand-alone tool: Willingness to engage with the system to receive a diagnosis without extra input from a mental health professional.

The dimensions of acceptability concerning the pilot system were assessed within the TFA and focused on general acceptability, affective attitude, perceived effectiveness, intervention coherence, self-efficacy, burden, ethicality, and opportunity cost. The construct of affective attitude was further broken down into two questions: (1) To what extent a participant likes or dislikes the system, and (2) how comfortable or uncomfortable they find the system to be. For ease of interpretability, the rating scale was unionized across all constructs, where "1" represented the "undesirable" rating (eg,

XSL•FO
RenderX

low levels of comfort; or the need for high levels of effort indicated burden), and "5" reflected the "desirable" rating (eg, high levels of comfort; or low levels of effort reported under burden). See Multimedia Appendix 1 for details about the specific items used to assess each of these constructs.

Usability

Usability was assessed by the System Usability Scale (SUS) [33]—which is a postinteraction self-report questionnaire to assess user perceptions of a system's overall usability. The scale is composed of 10 items rated on a Likert scale from 1 to 5, from "strongly disagree" to "strongly agree." The total score ranges from 0 to 100, with higher scores indicating higher usability ratings. There are multiple frameworks to interpret the SUS score (range, adjective, and grade) [34], but the primary threshold for this study was the average score necessary for an 'OK/satisfactory' experience (a score of 50.9 [35]).

Cybersickness

To assess levels of cybersickness following VR exposure, the Simulator Sickness Questionnaire (SSQ) was used [36]. The questionnaire contains 16 symptoms, which are rated on a scale of subjective severity: 0=none, 1=slight, 2=moderate, and 3=severe. Beyond the total score, one can attain separate scores for the nausea, oculomotor disturbances, and disorientation subscales.

However, while the SSQ was once considered to be the "gold standard," it has increasingly been criticized, including the original thresholds for the interpretation of the severity levels [37-40]. For this reason, the comparison of the pilot system results was made to the literature average [23] instead of the original thresholds.

Analysis

All analyses were carried out in R supplemented by R Studio (version 2023.12.1.402; R Foundation for Statistical Computing), except for the power calculation for which G*Power (version 3.1; Heinrich Heine University Düsseldorf) was used. Multiple packages were used to supplement analysis and visualization, these include but are not limited to: *dplyr* [41], *vioplot* [42], *ggplot2* [43], *psych* [44], *rstatix* [45], *ggeffects* [46], *ordinal* [47], and *VGAM* [48]. See Multimedia Appendix 2 for the script, its output, and a full list of packages.

Within-group comparisons and comparisons to known literature values were conducted using Wilcoxon signed rank tests, while between-group comparisons used Mann-Whitney *U* tests. These nonparametric methods were chosen to account for the ordinal nature of the outcome measures, such as Likert scales, and to address the observed violations of the assumptions required for parametric tests (eg, normality assessed via the Shapiro-Wilk test). A nominal significance level (α) of .05 was applied. For ease of interpretability, effect sizes were consistently reported in *r* values. This includes the use of rank-biserial correlations for Mann-Whitney *U* tests as well as Wilcoxon signed rank tests for quantifying the magnitude of the differences in a standardized manner. The full statistics are available in Multimedia Appendix 2.

Multivariate linear regression models were used to analyze predictors when the outcome variable was continuous, such as SUS or SSQ scores. For ordinal outcome measures, such as TFA constructs, proportional odds models were applied. In cases where the proportional odds assumption was not met (as noted in the Results section), results should be interpreted with caution. All regression models included age, sex, and education as covariates. When additional predictors, such as cybersickness severity scores, group labels, or self-reported anxiety and stress levels were relevant, these were included alongside the demographic variables.

Ethical Considerations

The study was carried out in accordance with the Declaration of Helsinki and obtained ethical approval from the local ethical committees on all sites where data was collected (Italy) or analyzed (Italy and Sweden). In Italy, approval was granted by Comitato Etico della Ricerca Psicologica, and in Sweden by the Etikprövningsmyndigheten (2023-00959-01).

Written informed consent was sought from all participants, with emphasis placed on the voluntary nature of participation and the possibility to withdraw from the study at any time, without reason, or any penalty. All participants were informed about data privacy and confidentiality, whereby all data obtained during the study were analyzed anonymously and used solely for the purpose of scientific research. Compensation included €25 (worth between US $24.38 and US $28.14 over the duration of the recruitment period).

Preregistration

The study was preregistered in the ISRCTN registry (ISRCTN16396369) [49]. The sole modification to the preregistered protocol concerned the threshold for inclusion on current DS severity. Due to the difficulty in recruiting volunteers in line with the study's timeline, the upper threshold for inclusion in the control group was increased from 4 to 5, while the lower threshold to be included in the group with DSs was decreased from 10 to 9 on the PHQ-9.

Results

Acceptability of the Concept

The use of VR for mental health treatment (62/84, 74%) and screening (65/84, 77%) was endorsed by the majority of those providing an answer, whereas only a minority indicated support for diagnostic purposes (28/83, 34%; Figure 5). There were no significant differences between the DS and HC groups.

Comparing the use of VR with or without input from a mental health professional, the responses indicate higher support for the VR system to be used as a support tool (62/83, 75%) compared to a stand-alone solution (15/84, 18%). A Wilcoxon signed rank test indicated a significant difference of large size (V=1770; P<.001, 2-tailed; r=0.81, 95% CI 0.76-0.84; $\text{median}_{\text{support}}$ 4, $\text{IQR}_{\text{support}}$ 3.5-4.5; $\text{median}_{\text{stand-alone}}$ 2, $\text{IQR}_{\text{stand-alone}}$ 1.0-3.0). For the raw distribution of responses see Figure 6. Again, no difference was observed between the DS and HC groups.

XSL•FO
RenderX

Figure 5. Participant ratings of VR acceptance for the purposes of mental health screening, diagnosis, and treatment. N/A: not available; VR: virtual reality.

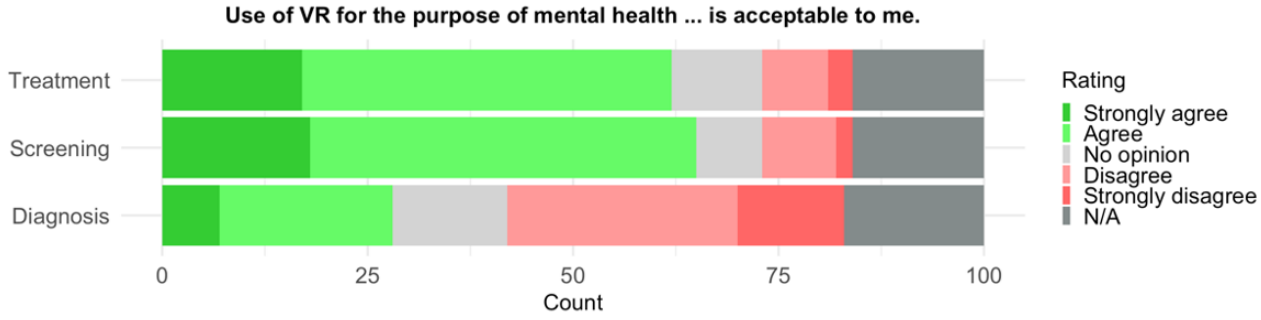
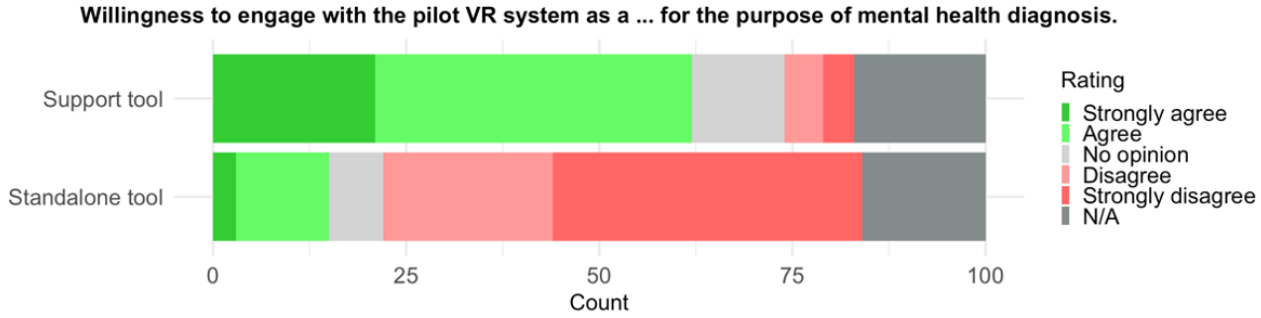


Figure 6. Participants' subjective ratings of their willingness to engage with the pilot VR system for mental health diagnosis. The support tool complements the assessment of a mental health professional, whereas the standalone tool operates without professional input. N/A: not available; VR: virtual reality.

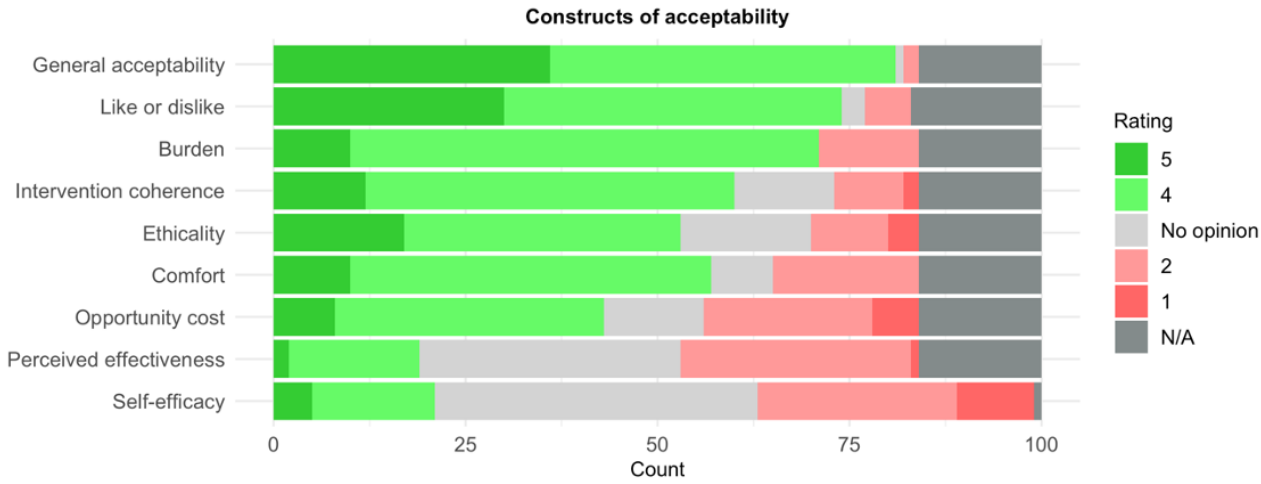


Acceptability of the Pilot System

Behavioral measures of acceptability include the discontinuation of participation for any reason, as well as the need to take a break from the study. While the possibility to discontinue participation at any point, without any consequences was stated before enrollment, all 100 volunteers completed the study. The need for a break was expressed by 1 participant (1/100, 1%) due to discomfort. They decided to resume participation despite the possibility of opting out being reemphasized. On average, 31 minutes were spent in the virtual environment.

Acceptability was also assessed via a self-report questionnaire specifically formulated for the project based on the TFA [18,32]. Ratings, visualized in Figure 7, indicated generally high levels of acceptability.

Figure 7. Self-reported acceptability of the pilot virtual reality system, assessed using constructs from the theoretical framework of acceptability (1=undesirable rating; 5=desirable rating). N/A: not available.



A majority of the participants found the system itself acceptable (81/84, 96%), liked the system (74/83, 89%), indicated that the engagement took little to no effort (71/84, 85%), found it clear how the system might help diagnose depressive disorders (60/84, 71%), found no moral or ethical consequences with the system (53/84, 63%), found it comfortable (57/84, 68%), and expressed no concern about missing out on other alternatives when engaging with the system (43/84, 51%). The 2 constructs on which the undesirable ratings (1 or 2) outnumbered desirable scores (4 or 5) concerned the perceived effectiveness of the

XSL•FO
RenderX

system in improving mental state (19/84, 23%), and self-efficacy regarding feeling confident while engaging with the system (21/99, 21%).

Multivariate ordinal regression models were used to explain variance in ratings of all the acceptability constructs as detailed in Table 2. Sex was a significant predictor of general acceptability and subjective burden, where women reported considerably lower levels of acceptability ($mean_{men}$ 4.65, SD_{men} 0.49; $mean_{women}$ 4.30, SD_{women} 0.65) and reported lower ratings on burden—meaning a considerably higher effort was required from their part to engage with the system ($mean_{men}$ 4.29, SD_{men} 0.50; $mean_{women}$ 3.69, SD_{women} 0.87). Education did not reach significance for any of the TFA constructs, and age was marginally significant only when used to explain whether participants found moral or ethical consequences to engaging with the system.

In the next stage, group label (DS vs HC) was entered into the same models in addition to the demographic variables. Group label was a significant predictor of comfort levels ($mean_{HC}$ 3.83, SD_{HC} 0.78; $mean_{DS}$ 3.34, SD_{DS} 1.08) and marginally significant for the like or dislike rating ($mean_{HC}$ 4.38, SD_{HC} 0.59; $mean_{DS}$ 4.00, SD_{DS} 0.95)—meaning that participants in the DS group reported lower comfort levels and liked the system to a lesser extent.

Finally, cybersickness severity (total SSQ score) was entered into models explaining all TFA constructs, in addition to the demographic variables. Cybersickness was a significant predictor for comfort, burden, opportunity cost, and perceived effectiveness and marginally significant for self-efficacy (Table 2). All showed an inverse relationship, whereas the higher the cybersickness symptom severity, the lower acceptability scores were reported on all TFA dimensions.

In summary, both the concept and the system received high acceptability ratings. Most participants supported the use of VR for mental health screening and treatment. However, regarding diagnosis, participants preferred VR as a supportive tool rather than a stand-alone solution. The primary area for improvement identified was enhancing users' self-efficacy while interacting with the system. Sex and cybersickness emerged as significant predictors of acceptability ratings.

XSL•FO
RenderX

Table 2. Estimates of predictors from 3 different models analyzed via multivariate ordinal regressions to explain constructs investigated within the theoretical framework of acceptability.

Model	Predictor	Estimate (log odds)	95% CI	*P* value
General acceptability				
Model 1	Age	0.175	–0.19 to 0.56	.36
Model 1	Sex	–1.331	–2.54 to –0.21	.02[a]
Model 1	Education	–0.529	–1.10 to 0.01	.053
Model 2	Group (DS[b] vs HC[c])[d]	–0.261	–1.19 to 0.65	.58
Model 3	Cybersickness	–0.004	–0.02 to 0.01	.54
Like or dislike				
Model 1	Age	–0.298	–0.67 to 0.05	.10
Model 1	Sex	–0.461	–1.52 to 0.58	.39
Model 1	Education	0.164	–0.34 to 0.67	.52
Model 2	Group (DS vs HC)[d]	–0.890	–1.81 to 0.00	.053
Model 3	Cybersickness	–0.010	–0.03 to 0.01	.20
Comfort				
Model 1	Age	–0.045	–0.38 to 0.29	.79
Model 1	Sex	–0.718	–1.84 to 0.37	.20
Model 1	Education	0.116	–0.36 to 0.58	.63
Model 2	Group (DS vs HC)[d]	–0.899	–1.81 to –0.02	.048[a]
Model 3	Cybersickness	–0.026	–0.04 to –0.01	.001[e]
Burden				
Model 1	Age	0.137	–0.26 to 0.54	.50
Model 1	Sex	–1.866	–3.26 to –0.59	.005[e]
Model 1	Education	–0.186	–0.78 to 0.38	.53
Model 2	Group (DS vs HC)[d]	–0.273	–1.28 to 0.72	.59
Model 3	Cybersickness	–0.036	–0.06 to –0.02	<.001[f]
Intervention coherence				
Model 1	Age	0.085	–0.25 to 0.42	.62
Model 1	Sex	–0.315	–1.42 to 0.77	.57
Model 1	Education	–0.193	–0.69 to 0.29	.44
Model 2	Group (DS vs HC)[d]	–0.567	–1.44 to 0.29	.20
Model 3	Cybersickness	–0.003	–0.02 to 0.01	.71
Ethicality				
Model 1	Age	0.320	–0.00 to 0.65	.053
Model 1	Sex	0.241	–0.74 to 1.23	.63
Model 1	Education	–0.166	–0.66 to 0.33	.51
Model 2	Group (DS vs HC)[d]	0.170	–0.63 to 0.98	.68
Model 3	Cybersickness	–0.010	–0.02 to 0.00	.15
Opportunity cost[g]				
Model 1	Age	–0.078	–0.39 to 0.24	.63
Model 1	Sex	–0.496	–1.53 to 0.50	.34

XSL•FO
RenderX

Model	Predictor	Estimate (log odds)	95% CI	*P* value
Model 1	Education	0.023	–0.44 to 0.49	.92
Model 2	Group (DS vs HC)[d]	0.336	–0.46 to 1.14	.41
Model 3	Cybersickness	-0.019	–0.03 to –0.01	.009[e]
Perceived effectiveness				
Model 1	Age	–0.116	–0.44 to 0.20	.48
Model 1	Sex	–0.757	–1.79 to 0.25	.14
Model 1	Education	–0.045	–0.53 to 0.43	.85
Model 2	Group (DS vs HC)[d]	–0.548	–1.38 to 0.27	.19
Model 3	Cybersickness	–0.019	–0.04 to –0.00	.02[a]
Self-efficacy[h]				
Model 1	Age	0.063	–0.24 to 0.37	.68
Model 1	Sex	–0.872	–1.86 to 0.08	.08
Model 1	Education	0.031	–0.40 to 0.48	.89
Model 2	Group (DS vs HC)[d]	0.043	–0.70 to 0.79	.91
Model 3	Cybersickness	-0.01	–0.03 to 0.00	.08

[a]*P*<.05.

[b]DS: depressive symptom.

[c]HC: healthy control.

[d]The categorization is based on current depressive symptom severity assessment via the PHQ-9 questionnaire (score equal to 9 or above for depressive symptoms [DS] and 5 or below for healthy controls [HC]) and not the presence of a clinical diagnosis.

[e]*P*<.01.

[f]*P*<.001.

[g]The assumption of proportional odds not was not fulfilled for the model. Results should be interpreted with caution.

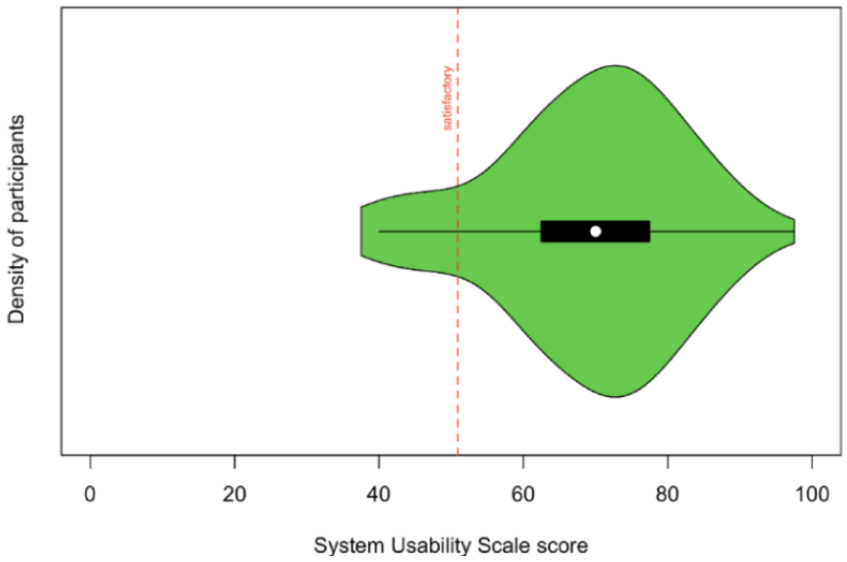
[h]Self-efficacy was measured with Item 9 of the System Usability Scale [33], as opposed to the other items that were specifically designed for the assessment of the pilot system based on the theoretical framework of acceptability [18,32].

Usability

Descriptives

Results from 99 participants indicated average ratings of usability, with a mean score of 69 (SD 12.86), and a range of 37.5-97.5 on SUS [33]. As visualized in Figure 8, which showcases the distribution of overall SUS scores, 88% (87/99) of participants provided a rating of 50.9 or above, indicating an OK or satisfactory experience or better [35].

Figure 8. Density plot of overall usability ratings on the System Usability Scale from all participants.



XSL•FO
RenderX

Analysis

A multiple linear regression analysis was conducted to predict usability ratings based on participants' age, education, gender, and group (DS vs HC). The regression model was not statistically significant ($F_{4,94}$=1.23; *P*=.31), explaining only about 5% of the variance in SUS scores (R^2=0.05, adjusted R^2=0.01). The strongest predictor was gender, but this only showed a marginal effect (β=–5.79, 95% CI –12.38 to 0.80; *P*=.08) with men reporting higher usability ratings than women.

A second model was tested examining the relationship between usability scores and the predictors of age, education, sex, and total cybersickness severity. The model was significant ($F_{4,94}$=2.97; *P*=.02), accounting for 11.2% of the variance in SUS scores (R^2=0.11, adjusted R^2=0.07). Cybersickness severity (SSQ scores) emerged as the only significant predictor (β=–0.12, 95% CI –0.21 to –0.03; *P*=.007), indicating that participants with higher SSQ scores tended to have lower SUS scores.

In summary, the system's overall usability was rated as sufficient, with the majority of participants reporting a satisfactory experience. Cybersickness severity emerged as the only predictor of usability ratings, showing a negative association.

Cybersickness

Descriptives

The overall mean SSQ score combining all subscales was 30.49 (SD 28.03), with individual scores ranging from 0 to 119.68. This is comparable to the literature average of 28.00 reported by Saredakis et al [23], on the overall sample, as well as the DS subsample. However, the HC group reported considerably lower levels of cybersickness severity (V=402; *P*=.02; r_{rb}=–0.37, 95% CI –0.61 to –0.07) as summarized in Table 3.

For the SSQ subscales, the results of the total sample reflect severity levels not (significantly) different from the literature average [23] regarding nausea and disorientation levels, whereas oculomotor disturbances were rated considerably higher in this study (V=3296; *P*=.004; r_{rb}=0.33, 95% CI 0.12-0.52).

When the comparison was split by study groups, only the DS group deviated significantly from the literature average—specifically on the nausea (V=917; *P*=.002; r_{rb}=0.50, 95% CI 0.22-0.70) and the oculomotor disturbances (V=967; *P*<.001; r_{rb}=0.58, 95% CI 0.33-0.75) subscales. The HC group showed near-average levels on all 3 subscales. The distributions of raw scores, as boxplots, along the total and all subscales are illustrated in Figure 9 and analyzed for between-group differences in the later section.

Table 3. Comparison of mean cybersickness severity scores (SSQa; mean [SD]) via Wilcoxon-signed rank tests across the total sample, those with DSb and HCc subsamples with literature averages [23].

SSQ Scale	Literature average[d]	Total sample, mean (SD)	DS[e], mean (SD)	HC, mean (SD)
Total	28.00	30.49 (28.03)	39.39 (32.89)	21.77 (18.84)[f]
Nausea	16.72	21.59 (21.96)	29.40 (24.39)[g]	13.93 (16.15)
Oculomotor disturbances	17.09	26.72 (24.63)[g]	33.57 (27.50)[h]	20.01 (19.48)
Disorientation	23.50	32.90 (39.84)	42.04 (48.34)	23.94 (26.83)

[a]SSQ: Simulator Sickness Questionnaire.

[b]DS: depressive symptom.

[c]HC: healthy control.

[d]Literature averages are taken from Saredakis et al [23] based on a meta-analysis of 55 studies on approximately 3000 participants.

[e]The categorization is solely dependent on the Patient Health Questionnaire-9 score and does not indicate the presence of a clinical diagnosis.

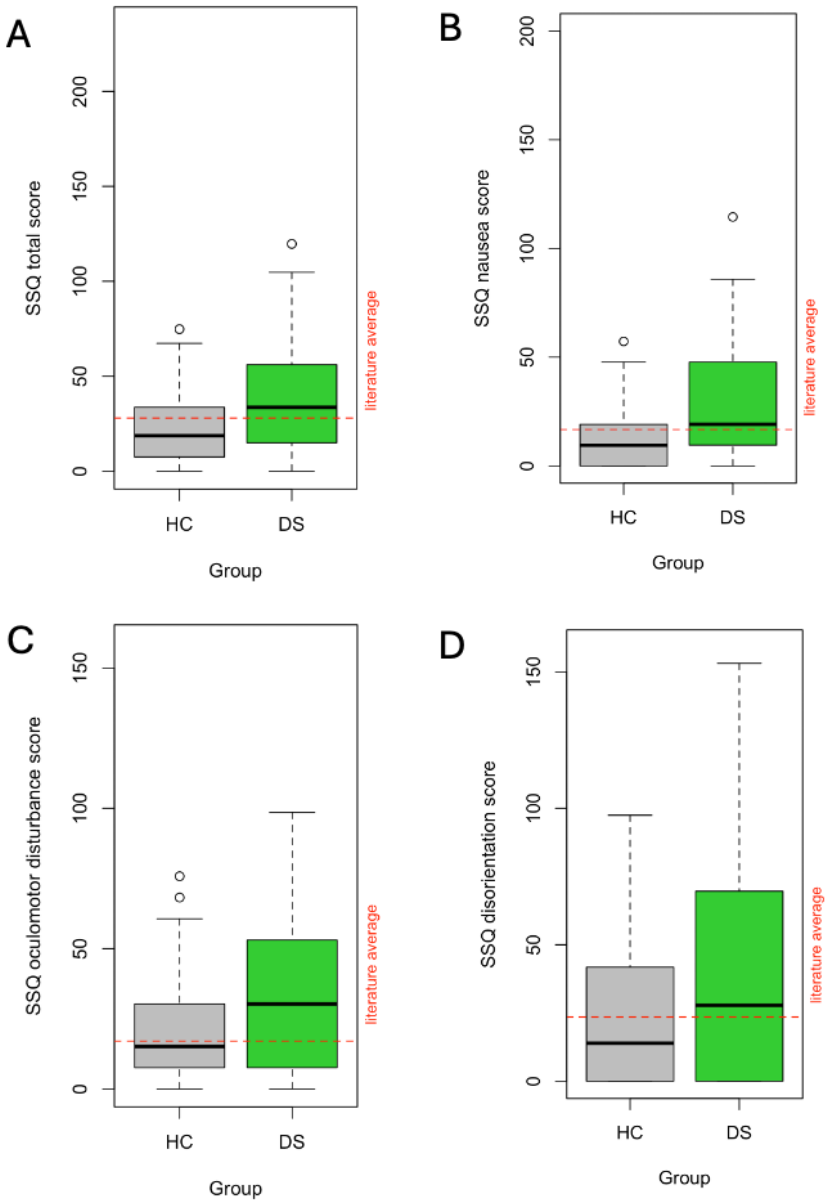
[f]*P*<.05.

[g]*P*<.01.

[h]*P*<.001.

XSL•FO
RenderX

Figure 9. Boxplot of self-reported cybersickness severity scores for the HC and DS study groups—measured by the (A) SSQ, and its 3 subscales of (B) nausea, (C) oculomotor disturbances, and (D) disorientation. DS: depressive symptom; HC: healthy control; SSQ: Simulator Sickness Questionnaire.



Analysis

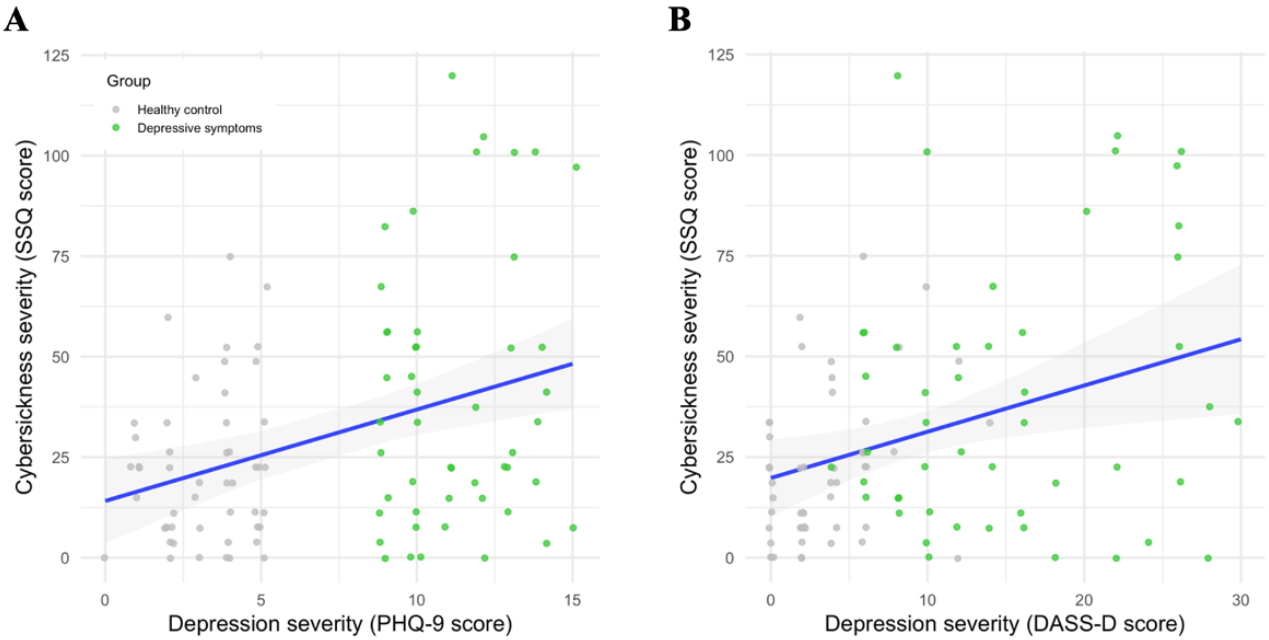
A multiple linear regression was conducted to predict SSQ scores based on age, education, sex, and group (DS vs HC). The model was statistically significant ($F_{4,94}$=3.20; P=.02), accounting for approximately 12% of the variance in SSQ scores (R^2=0.12, adjusted R^2=0.08). The only predictor reaching significance was group label, (β=17.58, 95% CI 6.57-28.60; P=.002), indicating that participants in the DS group had considerably higher SSQ scores (mean 39.38, SD 32.89) than the HC group (mean 21.77, SD 18.84). These between-group differences remained significant for all subscales analyzed separately.

When DS severity was entered as a continuous variable instead of a group label, a further 2% variance in cybersickness severity was explained (R^2=0.14, adjusted R^2=0.11) and a linear prediction of SSQ scores along the PHQ-9 scores was attainable as visualized in Figure 10A. The graph shows all participants along the SSQ and PHQ-9 scales and illustrates a positive association between the 2 scores when the regression line is plotted.

However, due to the baseline group differences in anxiety and stress levels (Table 1) this prediction might be biased. While DASS depression was not of primary interest, a multiple linear regression model with DASS depression as an outcome was used, to be able to adjust for DASS anxiety and DASS stress levels in addition to age, sex, and education levels as covariates. This model was also significant ($F_{6,92}$=4.05; P=.001) and explained about 21% of the variance in cybersickness severity (R^2=0.21, adjusted R^2=0.16). The only predictor emerging as significant was DASS depression (β=1.15, 95% CI 0.29-2.01; P=.009), whereas DASS anxiety reached marginal significance (β=1.2, 95% CI –0.07 to 2.47; P=.06). Cybersickness severity predictions showed a similar trend as for the PHQ-9 scores when the regression line was illustrated in Figure 10B.

XSL•FO
RenderX

Figure 10. Predicted cybersickness severity levels (with 95% CIs) based on depressive symptom severity, as measured by: (A) PHQ-9 and (B) DASS. DASS: Depression Anxiety Stress Scale; DASS-D: Depression Anxiety Stress Scale-Depression; PHQ-9: Patient Health Questionnaire; SSQ: the Simulator Sickness Questionnaire.



In summary, the severity of cybersickness was comparable to the literature averages for nausea and disorientation but significantly higher for oculomotor disturbances, particularly in the DS group. Across all subscales, the DS group reported higher SSQ scores than the HC group, with group membership emerging as the strongest predictor of cybersickness severity. A positive association was identified between DS severity (PHQ-9) and cybersickness, explaining a modest but significant proportion of variance. Even after adjusting for baseline anxiety and stress levels (DASS), DSs remained a significant predictor of cybersickness.

Discussion

Principal Findings

This study reports on the acceptability, usability, and cybersickness levels of a novel VR environment designed for the assessment of DSs. Following an approximately 30-minute engagement, 50 HCs and 50 individuals with moderate self-reported DSs completed a self-report battery on 3 constructs of interest.

Acceptability

In general, VR systems were rated acceptable by the majority of participants; both for the purpose of mental health screening as well as treatment. For diagnostics, a clear preference emerged for VR to be used as a support tool by a health care professional, as opposed to a stand-alone solution. While there is no literature available on VR for depression with which a direct comparison can be made, these findings are in line with previous results concerning other digital solutions, such as mHealth apps [50].

The specific pilot system tested here was designed for the purpose of differentiating between HCs and those with DSs [25]. The system itself was rated acceptable by most participants, but the need for improvement was apparent; specifically when it came to participants' confidence levels while using the system. Nevertheless, the high level of acceptance in practice found in the current study is in line with 2 studies previously reported on the acceptability of specific VR systems for therapeutic purposes for depression [21,22].

Usability

Usability was rated satisfactory by most of the participants, with cybersickness severity emerging as a significant predictor of usability scores. We found 1 previous study reporting on the usability of VR systems for depression, which has also indicated favorable ratings on ease of use and perceived usefulness [21]. However, in comparison with other psychiatric conditions, such as social anxiety disorder [51], psychosis [52], or posttraumatic stress disorder [53], it is important to emphasize that systematic evaluations of the usability of VR among those with depression are lacking.

Concerning the effect of cybersickness severity on usability, there are no previous studies reporting in relation to depression. Additionally, while comprehensive reviews on the relationship between cybersickness and presence are available [54] there is no such overview for usability. Existing single studies report inconsistent findings across different conditions. For example, a negative association is reported on healthy individuals [55], but no effect was found among those with Huntington disease [56].

These findings highlight the need for greater attention from the scientific community, as depression remains relatively understudied in the context of VR use compared to other psychiatric conditions [12,57]. Notably, participants in this study expressed support for the technology both conceptually and in relation to this specific pilot system. Identifying novel approaches with user support is particularly important for groups

XSL•FO
RenderX

that might find treatment adherence and lack of motivation to engage challenging, such as those with depression [58].

The high levels of acceptance and usability that users have shown toward VR technology within the study, both in theory and practice, lay a solid ground for investigations into how VR could be used in a mental health context. VR being a premise for real-time data collection via direct observation of behavior contributes to higher ecological validity and allows for the consideration of a wider set of variables when assessing mental states [6,7]. While the potential is promising, reviews of real-life studies [59], editorials [60] as well as commentaries [57], highlight a need for further scientific investigation, to better inform how VR is best implemented within clinical practice.

Cybersickness

Finally, the results of the study show significant differences in self-reported cybersickness levels between HCs and those with DSs —the latter reporting considerably higher levels. This difference brings forward many questions, some relating to the self-report instrument used, as well as its administration; these will be discussed under the limitations more thoroughly. Beyond the critical appraisal of the instrument and optimizing the system so that we minimize the risk of cybersickness in general [61-63], this difference urges a consideration of how cybersickness might relate to depression. It is important to consider that our study lacked a pre-post comparison, which creates challenges when attempting to disentangle cybersickness symptoms evoked by VR exposure from potential baseline differences stemming from somatic symptoms related to depression. While "the language used in the medical literature to describe somatic symptoms in depression is both confusing and contradictory" [64], one can pinpoint constructs related to depression, that may also be captured by cybersickness scales—which may cause contamination when assessing cybersickness levels. These include, but might not be limited to headache, fatigue, nausea, dizziness, and gastrointestinal disturbances [36,64]. This hypothesis needs further testing but is in line with the only study reporting on SSQ in a pre-post manner among those with psychiatric conditions including depression (heterogeneous sample of anxiety, psychotic, depressive, or bipolar disorder)—where the baseline levels of cybersickness would already be classed as "severe," but there is no significant elevation reported post exposure [65].

Nonetheless, it is important to consider that the elevated cybersickness in the group with DSs could still be, at least in part, a response to the VR environment itself. If this is the case, using VR with the right study design may serve as a potential tool for further exploring the sensory abnormalities in depression, though ethical considerations are warranted. Investigating the underlying mechanisms of cybersickness in depression may also provide novel insights into the somatic manifestations of the condition and could help further define depression as well as refine interventions for mental health. Additionally, the need to consider cybersickness severity becomes even more apparent in light of the results showing its association with acceptability and usability ratings, which might transfer into the real-life uptake and use of VR technology.

Limitations and Future Directions

The study is not without its limitations. First, one has to consider the characteristics of the participants, as the sample consisted of young adults, the majority being women, and highly educated which limits generalizability [66]. In general, lower age and higher education levels are associated with increased digital health literacy [67]. However, further reports are needed to understand how such sample characteristics—sociodemographic variables as well as DSs—influence user attitudes to digital technologies and engagement in a mental health context [68,69]. To this aim, there is a clear need for extending the research to include diverse age groups and educational backgrounds to enhance the generalizability and applicability of the findings.

Second, the categorization between the HC and the DS groups was based solely on self-reported symptoms at the time of the study (using the PHQ-9) as opposed to the presence or absence of a clinical diagnosis. Other forms of assessment might be worthy of consideration. Additionally, the majority of individuals labeled as DS only reported moderate depression severity, which questions whether the results are generalizable to those with more severe symptoms. Self-report measures in general are subject to response biases, such as social desirability, or recall bias. The subjective evaluation of symptom severity in this case could potentially have resulted in over- or underreporting of severity.

Third, although we controlled for demographic covariates, baseline differences in anxiety and stress levels between the study groups may have impacted our results. Self-reported stress and anxiety significantly differed between the HC and DS groups, highlighting that between-group differences in our constructs of interest cannot unequivocally be attributed to differences in DS severity.

Fourth, the measurement of cybersickness levels leaves room for improvement on multiple fronts. The instrument (SSQ) has repeatedly been criticized for reasons among others being the low power to differentiate from anxiety and the questionable assumption of having zero "symptoms" at baseline [37-40]. The difference in severity between the HCs and those with DSs cannot unequivocally be attributed to the VR exposure, as no baseline levels were recorded. It is possible, that at least some degree of the difference originates from baseline differences in anxiety or somatic symptoms of depression, which the questionnaire most likely does not adequately differentiate from cybersickness symptoms. In the future, assessment of related symptoms necessitates pre-post comparisons, particularly relevant when the SSQ is being used [37-40].

Fifth, when creating items to assess constructs of the TFA, 1 item was taken from the SUS questionnaire (item 9: confidence), while all other items were specialized to this study. Overall, single items created for this specific study lack the extensive work that foregoes validated questionnaires.

Sixth, the results of the ordinal regression models should be interpreted with caution, particularly in cases when the assumption of proportional odds was not fulfilled. This suggests that the relationship between predictors and the outcome variable may differ across categories and thus future studies could expand

XSL•FO
RenderX

on these findings by exploring more flexible modeling approaches.

Seventh, we must acknowledge that our study was observational and exploratory in nature, aiming to assess acceptability and usability constructs without preestablished hypotheses, which warrants some considerations. An important consequence of the observational design is the lack of randomization. As a result, there is potential for selection bias and confounding factors that could influence the results. Concerning the exploratory nature: while this approach provides valuable insights, especially at early stages of development, confirmatory research is needed to test specific predictions and establish causal relationships. This is particularly relevant for the potential between-group differences on how VR exposure relates to cybersickness severity.

Finally, regarding future studies, it is important to highlight that clinician perspectives can form an important complement to patients' user feedback. This dual approach provides insights that might not be accessible through the patient perspective alone, such as the feasibility and practicality of integrating VR solutions into clinical workflows [70,71]. As such, future studies could consider involving other user groups, and extend the consideration to implementation, feasibility, and cost-effectiveness of VR-based solutions to enable a more comprehensive evaluation.

In summary, despite the limitations and the need for further research, this study is the first to assess the acceptability, usability, and cybersickness levels of a novel VR system incorporating multimodal data for depression symptom assessment. The findings support the potential of VR as a tool for mental health screening, assessment (when used as a supplementary tool), and treatment. While the system's acceptability was generally favorable, areas for improvement were identified, particularly in enhancing users' self-efficacy during interactions. Usability ratings were satisfactory and cybersickness levels, on average, were comparable to other head-mounted displays. Notably, this study is the first to report elevated cybersickness levels among individuals with DSs, raising important theoretical and methodological questions for future research.

Conclusions

There is a clear need to enhance mental health diagnosis, potentially by incorporating a broader range of variables. This study explored a novel approach to increase the ecological validity of depression assessments and found that, while VR technologies are generally acceptable as a supplementary tool, they are not seen as a replacement for routine mental health evaluations. The study, which tested the initial version of a novel VR system, revealed that the majority of participants found the acceptability and usability satisfactory, despite experiencing considerable levels of cybersickness—which was found to be negatively associated with both constructs. Notably, the results highlighted previously unreported differences in cybersickness severity between individuals with DSs and HCs, warranting replication and further investigation. While the call for enhanced reliability in mental health assessments is well-founded, and novel approaches are under investigation for their efficacy and accuracy, greater emphasis must be placed on evaluating acceptability and usability. Ensuring a safe, acceptable, and user-friendly approach is essential for the successful implementation of these technologies.

Acknowledgments

The study was carried out within the EXPERIENCE project, which is funded by the European Commission H2020 Framework Program (grant 101017727).

Generative AI tools, including ChatGPT (OpenAI) and Perplexity (Perplexity AI, Inc), were used to enhance the clarity of the text, correct grammatical errors, and refine analysis code.

Data Availability

Access to the raw data will be considered on a case-by-case basis upon request. Interested parties may contact the corresponding author.

Authors' Contributions

SS contributed to methodology development, formal analysis, data curation, visualization, and writing the original draft. ETE was responsible for conceptualization, methodology development, formal analysis, data curation, and drafting the original manuscript. FM contributed to methodology development, formal analysis, data curation, and investigation, as well as drafting the original manuscript. VO developed the software, contributed to methodology development, data curation, and visualization, and participated in drafting of the original manuscript. V Catrambone contributed to methodology development and manuscript review and editing. GH was involved in methodology development and participated in reviewing and editing the manuscript. IT contributed to methodology development and manuscript review and editing. ALA participated in manuscript review and editing. V Cardi assisted in methodology development and manuscript review and editing. MGCAC contributed to manuscript review and editing. GM participated in methodology development and manuscript review and editing. MAR contributed to study conceptualization, provided resources, secured funding, supervised the project, and participated in project administration and manuscript review. GV conceptualized the study, secured funding, provided resources, supervised the project, and contributed to project administration and manuscript review. V Carli contributed to conceptualization, methodology, validation, project

XSL•FO
RenderX

administration, supervision, and funding acquisition. V Carli also provided resources and participated in reviewing and editing the manuscript. CG was involved in conceptualization, methodology, validation, supervision, and project administration. CG provided resources, acquired funding, and reviewed and edited the manuscript. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire on acceptability based on the Theoretical Framework of Acceptability.
[DOCX File , 40 KB - formative_v9i1e68132_app1.docx]

Multimedia Appendix 2
Script used for the analysis and visualization, including its output.
[PDF File (Adobe PDF File), 401 KB - formative_v9i1e68132_app2.pdf]

References

1. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry 2022;9(2):137-150 [FREE Full text] [doi: 10.1016/S2215-0366(21)00395-3] [Medline: 35026139]
2. Friedrich MJ. Depression is the leading cause of disability around the world. JAMA 2017;317(15):1517. [doi: 10.1001/jama.2017.3826] [Medline: 28418490]
3. Proudman D, Greenberg P, Nellesen D. The growing burden of major depressive disorders (MDD): implications for researchers and policy makers. Pharmacoeconomics 2021;39(6):619-625 [FREE Full text] [doi: 10.1007/s40273-021-01040-7] [Medline: 34013439]
4. Lieblich SM, Castle DJ, Pantelis C, Hopwood M, Young AH, Everall IP. High heterogeneity and low reliability in the diagnosis of major depression will impair the development of new drugs. BJPsych Open 2015;1(2):e5-e7 [FREE Full text] [doi: 10.1192/bjpo.bp.115.000786] [Medline: 27703745]
5. Liu X, Jiang K. Why is diagnosing MDD challenging? Shanghai Arch Psychiatry 2016;28(6):343-345 [FREE Full text] [doi: 10.11919/j.issn.1002-0829.216073] [Medline: 28638210]
6. Bell IH, Nicholas J, Alvarez-Jimenez M, Thompson A, Valmaggia L. Virtual reality as a clinical tool in mental health research and practice. Dialogues Clin Neurosci 2020;22(2):169-177 [FREE Full text] [doi: 10.31887/DCNS.2020.22.2/lvalmaggia] [Medline: 32699517]
7. Freeman D, Reeve S, Robinson A, Ehlers A, Clark D, Spanlang B, et al. Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol Med 2017;47(14):2393-2400 [FREE Full text] [doi: 10.1017/S003329171700040X] [Medline: 28325167]
8. Mao K, Wu Y, Chen J. A systematic review on automated clinical depression diagnosis. Npj Ment Health Res 2023;2(1):20 [FREE Full text] [doi: 10.1038/s44184-023-00040-z] [Medline: 38609509]
9. Bilello JA. Seeking an objective diagnosis of depression. Biomark Med 2016;10(8):861-875. [doi: 10.2217/bmm-2016-0076] [Medline: 27415130]
10. Abd-Alrazaq A, AlSaad R, Shuweihdi F, Ahmed A, Aziz S, Sheikh J. Systematic review and meta-analysis of performance of wearable artificial intelligence in detecting and predicting depression. NPJ Digit Med 2023;6(1):84 [FREE Full text] [doi: 10.1038/s41746-023-00828-5] [Medline: 37147384]
11. Wiebe A, Kannen K, Selaskowski B, Mehren A, Thöne AK, Pramme L, et al. Virtual reality in the diagnostic and therapy for mental disorders: a systematic review. Clin Psychol Rev 2022;98:102213 [FREE Full text] [doi: 10.1016/j.cpr.2022.102213] [Medline: 36356351]
12. Cieślik B, Mazurek J, Rutkowski S, Kiper P, Turolla A, Szczepańska-Gieracha J. Virtual reality in psychiatric disorders: a systematic review of reviews. Complement Ther Med 2020;52:102480 [FREE Full text] [doi: 10.1016/j.ctim.2020.102480] [Medline: 32951730]
13. Duan S, Valmaggia L, Lawrence AJ, Fennema D, Moll J, Zahn R. Virtual reality–assessment of social interactions and prognosis in depression. J Affect Disord 2024;359:234-240 [FREE Full text] [doi: 10.1016/j.jad.2024.05.098] [Medline: 38777276]
14. Camacho-Conde JA, Legarra L, Bolinches VM, Cano P, Guasch M, Llanos-Torres M, et al. Assessment of attentional processes in patients with anxiety-depressive disorders using virtual reality. J Pers Med 2021;11(12):1341 [FREE Full text] [doi: 10.3390/jpm11121341] [Medline: 34945813]
15. Cornwell BR, Salvadore G, Colon-Rosario V, Latov DR, Holroyd T, Carver FW, et al. Abnormal hippocampal functioning and impaired spatial navigation in depressed individuals: evidence from whole-head magnetoencephalography. Am J Psychiatry 2010;167(7):836-844 [FREE Full text] [doi: 10.1176/appi.ajp.2009.09050614] [Medline: 20439387]

XSL•FO
RenderX

16. Ghio L, Gotelli S, Marcenaro M, Amore M, Natta W. Duration of untreated illness and outcomes in unipolar depression: a systematic review and meta-analysis. J Affect Disord 2014;152-154:45-51. [doi: 10.1016/j.jad.2013.10.002] [Medline: 24183486]
17. DiMatteo MR, Lepper HS, Croghan TW. Depression is a risk factor for noncompliance with medical treatment: meta-analysis of the effects of anxiety and depression on patient adherence. Arch Intern Med 2000;160(14):2101-2107. [doi: 10.1001/archinte.160.14.2101] [Medline: 10904452]
18. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
19. Sauer J, Sonderegger A, Schmutz S. Usability, user experience and accessibility: towards an integrative model. Ergonomics 2020;63(10):1207-1220. [doi: 10.1080/00140139.2020.1774080] [Medline: 32450782]
20. Kim YY, Kim HJ, Kim EN, Ko HD, Kim HT. Characteristic changes in the physiological components of cybersickness. Psychophysiology 2005;42(5):616-625. [doi: 10.1111/j.1469-8986.2005.00349.x] [Medline: 16176385]
21. Migoya-Borja M, Delgado-Gómez D, Carmona-Camacho R, Porras-Segovia A, López-Moriñigo JD, Sánchez-Alonso M, et al. Feasibility of a virtual reality–based psychoeducational tool (VRight) for depressive patients. Cyberpsychol Behav Soc Netw 2020;23(4):246-252. [doi: 10.1089/cyber.2019.0497] [Medline: 32207997]
22. Paul M, Bullock K, Bailenson J, Burns D. Examining the efficacy of extended reality-enhanced behavioral activation for adults with major depressive disorder: randomized controlled trial. JMIR Ment Health 2024;11:e52326 [FREE Full text] [doi: 10.2196/52326] [Medline: 38437873]
23. Saredakis D, Szpak A, Birckhead B, Keage HAD, Rizzo A, Loetscher T. Factors associated with virtual reality sickness in head-mounted displays: a systematic review and meta-analysis. Front Hum Neurosci 2020;14:96 [FREE Full text] [doi: 10.3389/fnhum.2020.00096] [Medline: 32300295]
24. Lundin RM, Yeap Y, Menkes DB. Adverse effects of virtual and augmented reality interventions in psychiatry: systematic review. JMIR Ment Health 2023;10:e43240 [FREE Full text] [doi: 10.2196/43240] [Medline: 37145841]
25. Mura F, Alfeo AL, Sutori S, Eliasson ET, Catrambone V, Ortiz VG, et al. The EXPERIENCE system: integrating virtual reality and multi-modal explainable artificial intelligence for enhanced depression recognition. NA 2025 (forthcoming).
26. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
27. Prokhorenkova L, Gusev G, Vorobev A, Dorogush A, Gulin A. CatBoost: Unbiased boosting with categorical features. ArXiv. Preprint posted online on June 28, 2017 2019. [doi: 10.48550/arXiv.1706.09516]
28. Alfeo AL, Zippo AG, Catrambone V, Cimino MG, Toschi N, Valenza G. From local counterfactuals to global feature importance: efficient, robust, and model-agnostic explanations for brain connectivity networks. Comput Methods Programs Biomed 2023;236:107550 [FREE Full text] [doi: 10.1016/j.cmpb.2023.107550] [Medline: 37086584]
29. Lovibond P, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
30. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
31. Kashdan TB, Gallagher MW, Silvia PJ, Winterstein BP, Breen WE, Terhar D, et al. The Curiosity and Exploration Inventory-II: Development, factor structure, and psychometrics. J Res Pers 2009;43(6):987-998 [FREE Full text] [doi: 10.1016/j.jrp.2009.04.011] [Medline: 20160913]
32. Sekhon M, Cartwright M, Francis JJ. Development of a theory-informed questionnaire to assess the acceptability of healthcare interventions. BMC Health Serv Res 2022;22(1):279 [FREE Full text] [doi: 10.1186/s12913-022-07577-3] [Medline: 35232455]
33. Brooke J. SUS: A quickdirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry. London, UK: Taylor & Francis; 1996:189-194.
34. Kortum P, Peres SC. Evaluation of home health care devices: remote usability assessment. JMIR Hum Factors 2015;2(1):e10 [FREE Full text] [doi: 10.2196/humanfactors.4570] [Medline: 27025664]
35. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: Adding an adjective rating scale. J Usability Stud 2009;4(3):114-123. [doi: 10.5555/2835587.2835589]
36. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. Int J Aerosp Psychol 1993;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
37. Bimberg P, Weissker T, Kulik A. On the usage of the simulator sickness questionnaire for virtual reality research. 2020 Presented at: IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW); March 22-26, 2020; Atlanta, GA p. 464-467.
38. Bouchard S, Berthiaume M, Robillard G, Forget H, Daudelin-Peltier C, Renaud P, et al. Arguing in favor of revising the factor structure when assessing side effects induced by immersions in virtual reality. Front Psychiatry 2021;12:739742 [FREE Full text] [doi: 10.3389/fpsyt.2021.739742] [Medline: 34803762]

XSL•FO
RenderX

39. Brown P, Spronck P, Powell W. The Simulator Sickness Questionnaire, and the erroneous zero baseline assumption. Front Virtual Real 2022;3:945800. [doi: 10.3389/frvir.2022.945800]
40. Sevinc V, Berkman MI. Psychometric evaluation of simulator sickness questionnaire and its variants as a measure of cybersickness in consumer virtual environments. Appl Ergon 2020;82:102958. [doi: 10.1016/j.apergo.2019.102958] [Medline: 31563798]
41. Wickham H, François R, Henry L, Müller K, Vaughan D. Dplyr: A grammar of data manipulation. Tidyverse. 2023. URL: https://dplyr.tidyverse.org [accessed 2025-03-21]
42. Adler D, Kelly S, Elliott T, Adamson J. Vioplot: Violin plot. GitHub. 2024. URL: https://github.com/TomKellyGenetics/vioplot [accessed 2025-03-21]
43. Wickham H. Ggplot2: Elegant graphics for data analysis. Tidyverse. 2016. URL: https://ggplot2.tidyverse.org [accessed 2025-03-21]
44. Revelle W. Psych: Procedures for psychological, psychometric, and personality research. The Comprehensive R Archive Network. 2024. URL: https://CRAN.R-project.org/package=psych [accessed 2025-03-21]
45. Kassambara A. Rstatix: Pipe-friendly framework for basic statistical tests. Datanovia. 2023. URL: https://rpkgs.datanovia.com/rstatix/ [accessed 2025-03-21]
46. Lüdecke D. ggeffects: Tidy data frames of marginal effects from regression models. J Open Res Softw 2018;3(26):772 [FREE Full text]
47. Christensen RHB. Ordinal: Regression models for ordinal data. The Comprehensive R Archive Network. 2023. URL: https://CRAN.R-project.org/package=ordinal [accessed 2025-03-21]
48. Yee TW. The VGAM package for categorical data analysis. J Stat Soft 2010;32(10):1-34. [doi: 10.18637/jss.v032.i10]
49. Carli V, Gentili C. EXPERIENCE: Is virtual reality suitable to identify differences between depressed and healthy individuals? International Standard Randomised Controlled Trial Number Registry. 2022. URL: https://doi.org/10.1186/isrctn16396369 [accessed 2025-03-21]
50. Chan AHY, Honey MLL. User perceptions of mobile digital apps for mental health: acceptability and usability—an integrative review. J Psychiatr Ment Health Nurs 2022;29(1):147-168. [doi: 10.1111/jpm.12744] [Medline: 33604946]
51. Shahid S, Kelson J, Saliba A. Effectiveness and user experience of virtual reality for social anxiety disorder: systematic review. JMIR Ment Health 2024;11:e48916 [FREE Full text] [doi: 10.2196/48916] [Medline: 38329804]
52. Rus-Calafell M, Garety P, Sason E, Craig TJK, Valmaggia LR. Virtual reality in the assessment and treatment of psychosis: a systematic review of its utility, acceptability and effectiveness. Psychol Med 2018;48(3):362-391. [doi: 10.1017/S0033291717001945] [Medline: 28735593]
53. Jones C, Miguel Cruz A, Smith-MacDonald L, Brown MRG, Vermetten E, Brémault-Phillips S. Technology acceptance and usability of a virtual reality intervention for military members and veterans with posttraumatic stress disorder: mixed methods unified theory of acceptance and use of technology study. JMIR Form Res 2022;6(4):e33681 [FREE Full text] [doi: 10.2196/33681] [Medline: 35451971]
54. Weech S, Kenny S, Barnett-Cowan M. Presence and cybersickness in virtual reality are negatively related: a review. Front Psychol 2019;10:158 [FREE Full text] [doi: 10.3389/fpsyg.2019.00158] [Medline: 30778320]
55. Voinescu A, Petrini K, Stanton Fraser D. Presence and simulator sickness predict the usability of a virtual reality attention task. Virtual Real 2023;27(3):1-17 [FREE Full text] [doi: 10.1007/s10055-023-00782-3] [Medline: 37360806]
56. Migliore S, Casella M, Tramontano C, Curcio G, Squitieri F. Virtual reality tolerability, sense of presence and usability in Huntington disease: a pilot study. Neurol Sci 2025;46(1):219-225. [doi: 10.1007/s10072-024-07726-y] [Medline: 39103734]
57. Selaskowski B, Wiebe A, Kannen K, Asché L, Pakos J, Philipsen A, et al. Clinical adoption of virtual reality in mental health is challenged by lack of high-quality research. Npj Ment Health Res 2024;3(1):24 [FREE Full text] [doi: 10.1038/s44184-024-00069-8] [Medline: 38755289]
58. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018;21(3):116-119 [FREE Full text] [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
59. Geraets CNW, Wallinius M, Sygel K. Use of virtual reality in psychiatric diagnostic assessments: a systematic review. Front Psychiatry 2022;13:828410 [FREE Full text] [doi: 10.3389/fpsyt.2022.828410] [Medline: 35295778]
60. Riva G, Serino S. Virtual reality in the assessment, understanding and treatment of mental health disorders. J Clin Med 2020;9(11):3434 [FREE Full text] [doi: 10.3390/jcm9113434] [Medline: 33114623]
61. Biswas N, Mukherjee A, Bhattacharya S. “Are you feeling sick?”—A systematic literature review of cybersickness in virtual reality. ACM Comput Surv 2024;56(11):1-38. [doi: 10.1145/3670008]
62. Page D, Lindeman RW, Lukosch S. Identifying strategies to mitigate cybersickness in virtual reality induced by flying with an interactive travel interface. Multimodal Technol Interact 2023;7(5):47. [doi: 10.3390/mti7050047]
63. Papaefthymiou S, Giannakopoulos A, Roussos P, Kourtesis P. Mitigating cybersickness in virtual reality: impact of eye–hand coordination tasks, immersion, and gaming skills. Virtual Worlds 2024;3(4):506-535. [doi: 10.3390/virtualworlds3040027]
64. Tylee A, Gandhi P. The importance of somatic symptoms in depression in primary care. Prim Care Companion J Clin Psychiatry 2005;7(4):167-176 [FREE Full text] [doi: 10.4088/pcc.v07n0405] [Medline: 16163400]

XSL•FO
RenderX

65. Veling W, Lestestuiver B, Jongma M, Hoenders HJR, van Driel C. Virtual reality relaxation for patients with a psychiatric disorder: crossover randomized controlled trial. J Med Internet Res 2021;23(1):e17233 [FREE Full text] [doi: 10.2196/17233] [Medline: 33448933]
66. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
67. Estrela M, Semedo G, Roque F, Ferreira PL, Herdeiro MT. Sociodemographic determinants of digital health literacy: a systematic review and meta-analysis. Int J Med Inform 2023;177:105124 [FREE Full text] [doi: 10.1016/j.ijmedinf.2023.105124] [Medline: 37329766]
68. Chudy-Onwugaje K, Abutaleb A, Buchwald A, Langenberg P, Regueiro M, Schwartz DA, et al. Age modifies the association between depressive symptoms and adherence to self-testing with telemedicine in patients with inflammatory bowel disease. Inflamm Bowel Dis 2018;24(12):2648-2654 [FREE Full text] [doi: 10.1093/ibd/izy194] [Medline: 29846623]
69. Abouzeid N, Lal S. The role of sociodemographic factors on the acceptability of digital mental health care: a scoping review protocol. PLoS One 2024;19(4):e0301886 [FREE Full text] [doi: 10.1371/journal.pone.0301886] [Medline: 38669278]
70. Chung OS, Robinson T, Johnson AM, Dowling NL, Ng CH, Yücel M, et al. Implementation of therapeutic virtual reality into psychiatric care: clinicians' and service managers' perspectives. Front Psychiatry 2021;12:791123 [FREE Full text] [doi: 10.3389/fpsyt.2021.791123] [Medline: 35058823]
71. Chung OS, Robinson T, Johnson AM, Dowling NL, Ng CH, Yücel M, et al. Corrigendum: Implementation of therapeutic virtual reality into psychiatric care: Clinicians' and service managers' perspectives. Front Psychiatry 2022;13:893637 [FREE Full text] [doi: 10.3389/fpsyt.2022.893637] [Medline: 35815031]

Abbreviations

DASS: Depression Anxiety Stress Scale
DS: depressive symptom
DSM: Diagnostic and Statistical Manual of Mental Disorders
HC: healthy control
PANAS-SF: Positive and Negative Affect Schedule
PHQ-9: Patient Health Questionnaire-9
SSQ: Simulator Sickness Questionnaire
SUS: System Usability Scale
TFA: theoretical framework of acceptability
VR: virtual reality

Edited by A Mavragani; submitted 08.11.24; peer-reviewed by R Zahn, AN Ramaseri-Chandra, C Nega; comments to author 03.01.25; revised version received 21.01.25; accepted 11.03.25; published 16.04.25.

Please cite as:
Sutori S, Eliasson ET, Mura F, Ortiz V, Catrambonephd V, Hadlaczky G, Todorov I, Alfeo AL, Cardi V, Cimino MGCA, Mioni G, Raya MA, Valenza G, Carli V, Gentili C
Acceptability, Usability, and Insights Into Cybersickness Levels of a Novel Virtual Reality Environment for the Evaluation of Depressive Symptoms: Exploratory Observational Study
JMIR Form Res 2025;9:e68132
URL: https://formative.jmir.org/2025/1/e68132
doi:10.2196/68132
PMID:

©Sara Sutori, Emma Therése Eliasson, Francesca Mura, Victor Ortiz, Vincenzo Catrambonephd, Gergö Hadlaczky, Ivo Todorov, Antonio Luca Alfeo, Valentina Cardi, Mario G C A Cimino, Giovanna Mioni, Mariano Alcañiz Raya, Gaetano Valenza, Vladimir Carli, Claudio Gentili. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Digital Photo Activity Intervention for Nursing Home Residents With Dementia and Their Carers: Mixed Methods Process Evaluation

Josephine Rose Orejana Tan[1], PGDip, MSc; David P Neal[2], MB BChir, MA, PhD; Maria Vilmen[1], BSc; Petra Boersma[3,4], PhD; Teake P Ettema[1], PhD; Robbert J J Gobbens[3,5,6,7], RN, MSc, PhD; Sietske A M Sikkes[8,9,10], PhD; Rose-Marie Dröes[1], PhD

[1]Department of Psychiatry, Amsterdam Public Health Research Institute, Amsterdam UMC, location Vrije Universiteit, Amsterdam, The Netherlands

[2]eHealth Living & Learning Lab, Department of Medical Informatics, Amsterdam UMC, Amsterdam, The Netherlands

[3]Lectorate Health and Well-being of Frail Elderly People, Faculty of Health, Sports and Social Work, Inholland University of Applied Sciences, Amsterdam, The Netherlands

[4]Program Learning, Improvement & Implementation, Ben Sajet Centrum, Amsterdam, The Netherlands

[5]Zonnehuisgroep Amstelland, Amstelveen, The Netherlands

[6]Department of Family Medicine and Population Health, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium

[7]Tranzo, Tilburg University, Tilburg, The Netherlands

[8]Department of Clinical, Neuro- & Developmental Psychology, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[9]Neurology, Alzheimer Center Amsterdam, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands

[10]Amsterdam Neuroscience, Neurodegeneration, Amsterdam, The Netherlands

Corresponding Author:
Josephine Rose Orejana Tan, PGDip, MSc
Department of Psychiatry
Amsterdam Public Health Research Institute
Amsterdam UMC, location Vrije Universiteit
Oldenaller 1
Amsterdam, 1081 HJ
The Netherlands
Phone: 31 3316459170
Email: j.r.tan@amsterdamumc.nl

Abstract

Background: Within the framework of a randomized controlled trial investigating the impact of a digital, psychosocial photo activity intervention for residents living with dementia in nursing homes and their informal and formal carers, a process evaluation was conducted to determine factors that affected the implementation of the intervention and potentially influenced the intervention outcomes.

Objective: By tracing facilitators and barriers to implementation, the study also aimed to inform future implementation of the photo activity intervention.

Methods: Following Medical Research Council guidance, mixed methods were used to investigate context, implementation, and mechanism-of-impact factors during the photo activity intervention via the Fotoscope web application versus a general conversation activity (control). Google Analytics was set up to gain insight into how the Fotoscope web application was used in practice. For quantitative data, descriptive statistics were calculated and differences between groups tested. For qualitative data, thematic analysis was performed.

Results: In total, 163 semistructured interviews were conducted with residents (photo activity group: n=29, 17.8%; control: n=29, 17.8%), formal carers (photo activity group: n=23, 14.1%; control: n=27, 16.6%), and informal carers (photo activity group: n=28, 17.2%; control: n=27, 16.6%). Regarding contextual factors, a minority of formal carers in both groups (photo activity group: 4/18, 22%; control: 9/24, 38%) mentioned time and workload as barriers to implementing the intervention. Regarding implementation, 86% (25/29) of the residents in the intervention group felt that the digital photo activity worked well on a tablet.

XSL•FO
RenderX

Informal carers from both groups wanted more intervention updates from formal carers. The majority of formal carers from both groups were satisfied with how the training and activities were implemented. Regarding the mechanisms of impact, residents in the photo activity group (27/29, 93%) felt significantly more positive about the conversations with their carer (U=533.0, z=2.865, r=0.39; P=.004). Formal carers in the photo activity group (20/23, 87%) got to know the resident better (U=390.5, z=2.114, r=0.302; P=.04) compared to the formal carers in the control group (21/27, 78%). Formal carers in the photo activity group (23/50, 46%) gave a significantly higher rating to the digital photo activity as a way of getting to know the resident living with dementia better (median 9.00, IQR 7-9; U=419.0, z=2.169, r=0.307; P=.03) compared to formal carers in the control group (27/50, 54%; median 8.00, IQR 6-8). Finally, the majority of formal carers in the photo activity group (14/18, 78%) agreed that the Fotoscope app can be used as part of care activities in the nursing home.

Conclusions: The work invested by formal carers in implementing the photo activity did not seem to differ greatly compared to implementing a general conversation activity, suggesting that the digital photo activity, as an easy-to-implement and enjoyable intervention, could be widely implemented and disseminated in nursing homes.

International Registered Report Identifier (IRRID): RR2-https://doi.org/10.1186/s12877-021-02632-w

(JMIR Form Res 2025;9:e56586) doi:10.2196/56586

KEYWORDS

dementia; psychosocial interventions; nursing home; process evaluation; social interaction; photos; art

Introduction

Background

The photo activity is a psychosocial intervention based on an earlier project by a visual artist (known as *Photographic Treatment* [1]) for nursing home residents with dementia and their informal and formal carers. It is designed to facilitate social interaction [2]. In a small-scale pilot study (n=20), nursing home residents viewed printed, artistic, generic photos related to their personal interests [2]. The study found that when residents with dementia viewed person-centered generic photos (ie, not personal or family photos but images of subjects related to personal interests) as opposed to non–person-centered photos during conversations with their informal or formal carers, they tended to exhibit more positive effects on social interaction, mood, speech, and negative behavior [2]. In recent decades, psychosocial interventions have been influenced in part by the many technological developments aimed at improving the social health and quality of life of people with dementia [3-6]. Examples include the use of social robots [7], exergaming [8], and online art gallery activities [9]. However, few studies have looked into the effects of psychosocial interventions using technology for nursing home residents with dementia [10-12]. The photo activity is an example of a psychosocial intervention offered in nursing homes that was recently digitally adapted through the development of the Fotoscope app [13].

The development of the Fotoscope app allows for upscaling and wider dissemination of the digital photo activity intervention (ie, photo activity using digital photos, rather than printed photos). Using the app, informal and formal carers can easily conduct the digital photo activity with residents by accessing a database of artistic, black-and-white, generic photos and making a person-centered selection of digital photos to view with the residents [13], as opposed to having a limited collection of printed photos. However, before disseminating the digital photo activity intervention, it is important to evaluate its effectiveness and to consider facilitators and barriers to implementing it in daily nursing home care.

It is argued that alongside evaluating an intervention's effects, process evaluations are crucial to look at how and under what conditions an intervention was implemented, to make the outcomes reproducible, and to learn about factors that facilitated or impeded the implementation of the intervention, all of which will inform the interpretation of trial outcomes and future upscaling [14-16]. The importance of conducting process evaluations alongside randomized controlled trials (RCTs) was emphasized in the revised UK Medical Research Council (MRC) guidance for process evaluations of complex interventions [17]. In addition to tracing facilitators and barriers to implementation, process evaluations allow for the identification of contextual factors that may have influenced the implementation of the intervention and its outcomes, as well as the mechanisms underlying the intervention's impact [16,17]. *Implementation* refers to the ways and processes in which the intervention was delivered in practice, how frequently it was delivered, and whether it was delivered as intended. *Mechanisms of impact* refers to the aspects of the intervention that created change in practice or in the target population, investigating which would help in understanding future replications of the intervention. Finally, *context* refers to factors outside of the intervention that may facilitate or hinder the implementation [17].

As technology for dementia care continues to advance, along with RCTs evaluating the effectiveness of care technology [3,10], it is becoming more common to include process evaluations alongside such trials; for example, process evaluations based on the MRC guidance have been conducted in studies evaluating the use of different kinds of technology in psychosocial care for people with dementia in the community or long-term care settings [18-21].

Objectives

This paper reports on a process evaluation based on the MRC guidance [17], which was carried out alongside a pilot RCT (Netherlands Trial Register: NL9219 [22]) investigating the feasibility and impact of the digital photo activity intervention. The implementation of the photo activity intervention in comparison with the implementation of a control intervention

XSL•FO
RenderX

(a general conversation activity without photos) was evaluated. The purpose of this study was to identify factors that either facilitated or hindered the implementation of the photo activity intervention compared to a general conversation activity, including contextual, implementation, and mechanism-of-impact factors, to generate insights to interpret RCT outcomes and to inform further implementation and dissemination of the intervention.

The study aimed to answer the following research questions (RQs):

- RQ1: How did factors external to the interventions affect the implementation of the interventions (context)?
- RQ2: How did residents and their informal and formal carers assess the preparation, introduction, and delivery of the interventions? (implementation)?
- RQ3: How did the residents with dementia and their informal and formal carers experience and value the digital photo activity versus the general conversation activity (mechanisms of impact)?

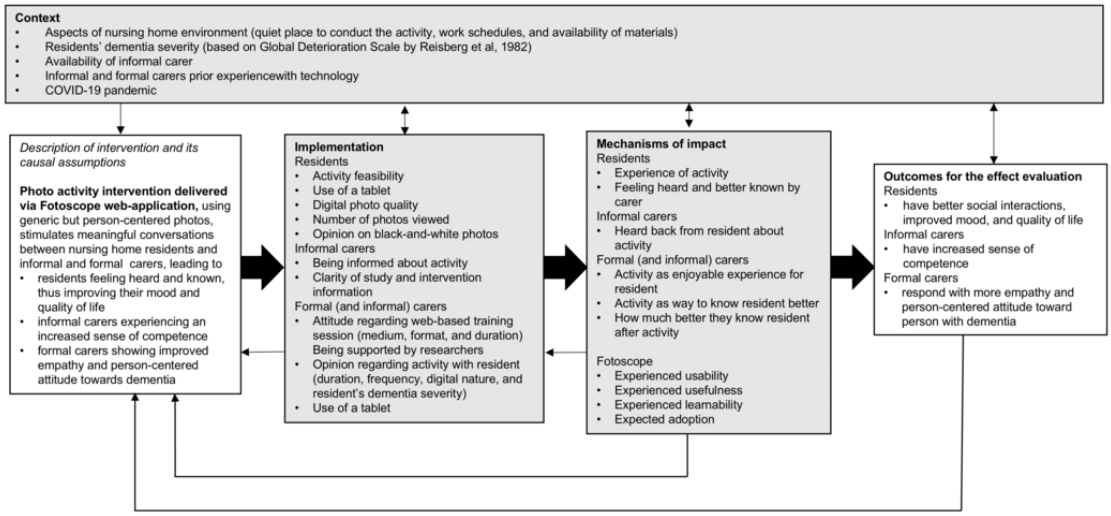
Methods

Design of the Process Evaluation

Following MRC guidance [17] (Figure 1 [23]), we used a mixed methods approach to investigate contextual, implementation, and mechanism-of-impact factors because the combination of quantitative and qualitative data would allow us to gain a more comprehensive understanding of the way the photo activity, as a complex intervention, is organized, offered, and implemented in the clinical setting [17,24]. Using mixed methods can also provide the flexibility needed in working with nursing home residents with dementia, ensuring that they are able to contribute their insights regarding the intervention [24,25]. The reporting and writing of this study was guided by the Good Reporting of a Mixed Methods Study (GRAMMS) checklist [26].

During the 4-week intervention period, analytics data were collected on Fotoscope app use and interactions of carer-resident pairs assigned to the photo activity intervention. After the intervention, qualitative and quantitative data were collected through semistructured interviews. The full protocol for the RCT, including the process evaluation, has been published elsewhere [13].

Figure 1. Contextual, implementation, and mechanism-of-impact factors, based on the Medical Research Council guidance for process evaluation, that could facilitate or hinder the implementation of the digital photo activity in the nursing home.



Ethical Considerations

This research was approved by the Vrije University Medical Center's Ethics Review Committee (2020.221) and embedded within the Amsterdam Public Health Research Institute (SQC2020-039). All participants provided informed consent. For individuals with dementia who lacked the capacity to consent, their legal representative signed the consent form on their behalf.

For privacy and confidentiality, data are anonymized, coded, and stored securely. Financial compensation was given to participating nursing home wards for their time (€100 [US $108.7] per participating resident with dementia who completed an intervention round).

Participants

The study aimed to recruit nursing home residents in the Netherlands with dementia at severity levels 4, 5, or 6 on the Global Deterioration Scale (GDS) [23], along with their informal and formal carers. To do this, the researchers worked with interested regional and national nursing home organizations. The study was presented to individual nursing home wards, where ward managers, in consultation with their teams, determined their willingness and capacity to participate. In total, 20 nursing home wards from 3 large care organizations took part in the study. They were asked to recruit formal carers to fulfil the following roles: coordinator (served as the first point of contact, recruited informal carers and residents with dementia for the study by providing written and oral information about the study, and assessed the dementia severity of participating

XSL•FO
RenderX

residents), intervention providers (2 formal carers randomized to deliver either the photo activity or control activity), and independent assessor (a formal carer not involved in delivering either activity, who conducted blind assessments of residents' daily functioning in the ward for the pilot RCT by filling out web-based questionnaires). Informal carers of residents at the participating care organizations were also allowed to provide either the photo activity or control activity if they wished. Researchers contacted residents and informal carers about their participation in the study only after the coordinator obtained their informed consent. The photo activity and control activity were provided by formal carers (or informal carers if they wished to provide the interventions themselves) to the residents twice a week for 4 weeks. Residents as well as informal and formal carers who provided either activity and completed the 4-week intervention period were then recruited for the process evaluation by the researchers and were invited to complete the semistructured interviews via online video calls or telephone calls. The formal carers who participated as independent assessors were only involved in the pilot RCT and not in the process evaluation.

Pairs of residents with the same dementia severity who had lived in the nursing home for at least 1 month were recruited by the coordinator. After providing informed consent, they were randomly assigned to either the experimental or control condition through a lot-drawing process. The only exclusion criteria were severe vision or hearing problems. Informal carers were assigned to the same condition as the resident.

The sample size for the RCT was determined based on the results of the photo activity pilot study. Large positive effect sizes were found for the INTERACT subscales social interaction (Cohen *d*=0.86) and reduced negative behavior (Cohen *d*=0.85) [2]. On the basis of these effect sizes and an expected dropout rate of 10% over 6 weeks, a power calculation (power=0.80; α=.05; Cohen *d*=0.80) indicated that 45 residents per group (experimental and control) were needed. All residents and their carers involved in the RCT were invited to participate in the process evaluation.

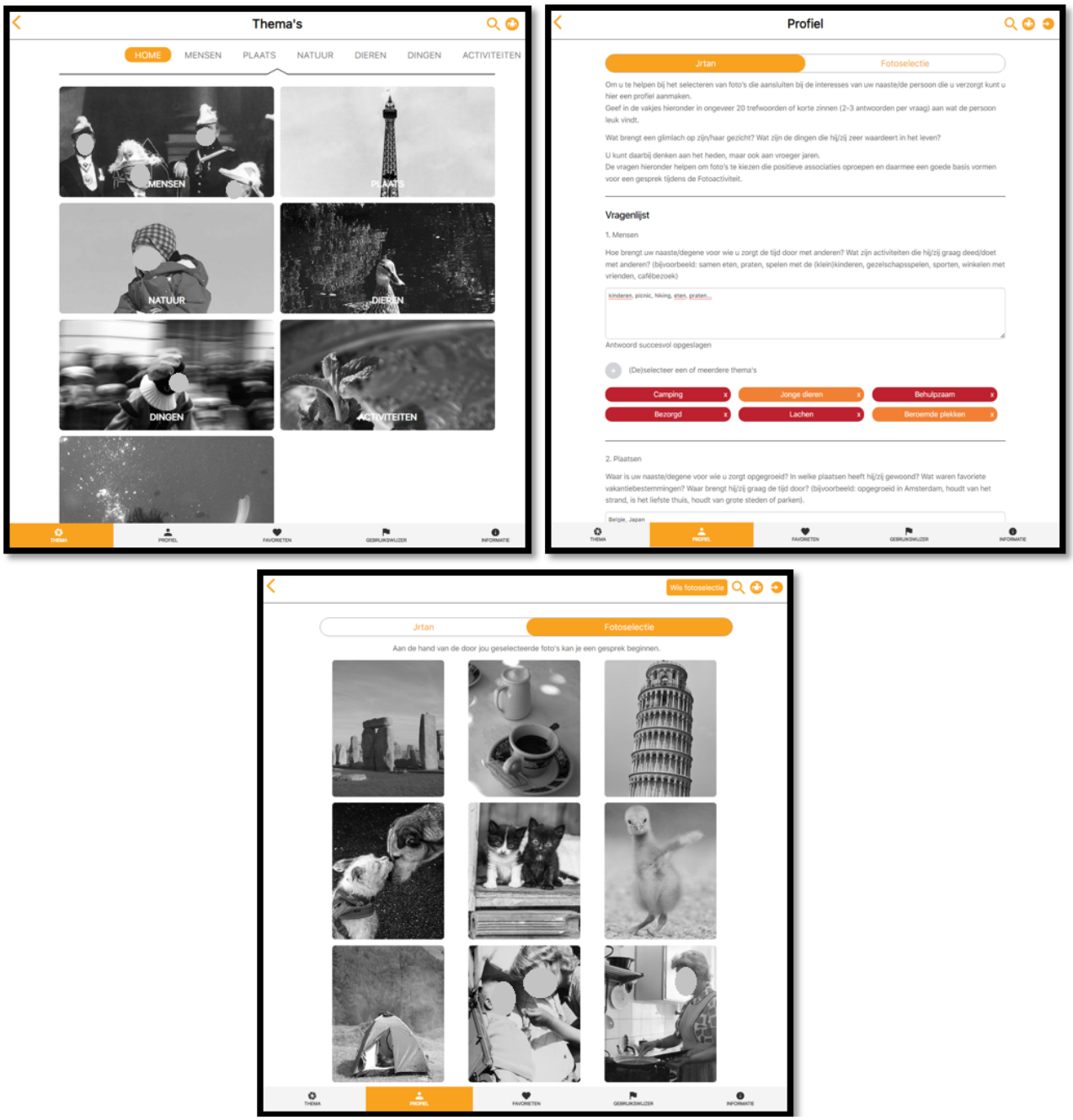
Interventions

Photo Activity (Experimental Intervention) and the Fotoscope Web Application

The photo activity intervention engaged residents and their carers in a 1-on-1 conversation about artistic, person-centered generic photos in the Fotoscope web application. For the digital photo activity, the Fotoscope database was expanded to 1500 black-and-white photos, which were organized according to 7 main themes (people, places, nature, animals, activities, things, and experiences). The Fotoscope web application features multiple sections, such as Themes, Profile, Favorites, User Guide, and Information (Figure 2). The Themes page displays the 7 main photo categories and subcategories. The Profile page contains questions designed to learn more about the residents' personal interests and features a selection tool that allows carers to easily select photos that match the residents' interests and save them in the Photo Selection tab. The Favorites page displays photos that elicit positive reactions from the resident, which the carer can add during or after their conversation. The User Guide provides instructions on using the Fotoscope web application and offers tips on person-centered communication. The Information page contains details about copyright and the researchers' contact information. Additional features include a search function (carers can enter keywords to find a photo) and a "joker" function (a collection of photos that have been found to elicit positive reactions).

XSL•FO
RenderX

Figure 2. Screenshots of the Fotoscope web application's Themes, Profile, and Photo Selection pages. Photos in the Fotoscope web-application were publicly available on the internet and collected by the artist, or were taken by the artist herself. No personal photos of participants were used.



General Conversation Activity (Control Intervention)

The control intervention also involved residents and their carers in a 1-on-1 conversation but without the Fotoscope web application and generic photos. The carer was encouraged to have an open conversation with the resident based on general topics and to allow the resident to lead the conversation.

Data Collection Methods

Data for the process evaluation were collected via a demographic questionnaire, semistructured interviews, Google Analytics, and a carer self-report registration form (Multimedia Appendix 1).

Demographic Questionnaire

Questions for the demographic questionnaire administered at baseline (T0) were adapted from The Older Persons and Informal Caregivers Survey Minimum Dataset [27], with additional questions pertaining to relevant medical information for the residents with dementia. For residents, information on sex, age, education, years in the nursing home, type of dementia, and severity of dementia was collected. For informal carers, information on sex, age, education, relationship to residents, and employment status (whether they had paid work) was collected. For formal carers, information on sex, age, education, work function, and years of experience in psychogeriatrics was collected.

XSL•FO
RenderX

Semistructured Interview

The questions for the semistructured interview, which was conducted after the 4-week intervention period (T1), were adapted from existing interview guides used in a previous process evaluation study on a digital psychosocial intervention for people with dementia and their caregivers and pilot-tested [18]. The questions were adapted by JROT, RMD, and PB to fit the photo activity trial. On the basis of the MRC guidance for process evaluation [17], context, implementation, and mechanism-of-impact factors (including usability, usefulness, learnability, and adoption of the Fotoscope app) were explored through a mix of open-ended and closed-ended questions (Multimedia Appendix 2). An example of a closed-ended question for a resident with dementia who participated in the photo activity intervention is "Do you think your carer got to know you better by having a regular conversation with you about the photos together?"; the corresponding closed-ended question for a resident with dementia in the control condition was "Do you think your carer got to know you better by having regular conversations with you on general topics?" The answer options for these questions were "A lot better," "Better," "A little better," and "No, not better."

Google Analytics

The Fotoscope web application was linked to a Google Analytics account to measure the carers' use of, and interactions with, the app during the trial. To protect residents' privacy, IP addresses were masked, and codes rather than names were used to log in. To gain insight into carers' engagement with the Fotoscope web application (*implementation*), data on the number of sessions per user and average session durations were collected. In addition, dates of each session, as well as session start times, were recorded for each user.

To determine whether carers used the Fotoscope web application to get to know residents better (*mechanisms of impact*), unique page views of the Profile and Favorites pages were tracked in Google Analytics. To determine whether carers found the pages and functions of the Fotoscope web application useful (*usefulness*), the number of unique views of these pages was recorded as an indication. To determine whether carers needed additional guidance on using the Fotoscope web application, the number of unique page views of the User Guide and Information page was recorded (*learnability*). Finally, information on device type was collected (*context*) to know whether the Fotoscope web application was used by carers on a tablet, as advised.

Carer Self-Report Registration Form

Carers were asked to record the dates, times, and duration of the activities with the residents on a standard record form to track the time spent on the activities and the number of sessions conducted (*implementation*). If a session was missed, the intervention discontinued, or the resident dropped out, they were asked to document the reason on the form.

Procedure

The pilot RCT in February 2021 was disrupted by the COVID-19 pandemic. In the Netherlands, nursing homes were placed under quarantine, restricting visitor access [28]. The trial protocol was adapted to an online format. Carers assigned to the photo activity attended a 1.5-hour web-based training session where they were introduced to the photo activity intervention and procedure [2] and were shown an online demonstration of the Fotoscope web application. General communication tips were discussed, followed by more specific person-oriented communication skills for interacting with people with dementia [29]. Carers were then given personal log-in credentials and passwords for the Fotoscope web application—1 set for their own use to enable them to become familiar with the app and another for each resident participating in the photo activity. Each resident's Fotoscope user account was linked to a unique user ID in Google Analytics, and only the researchers had access to the key linking the IDs to specific residents. Carers in the photo activity group were instructed to prepare for the first session by completing the Fotoscope Profile page with input from family members and selecting a minimum of 15 photos based on the residents' interests. Formal carers in the photo activity group were instructed to update the informal carers (by telephone or when they visited the resident) on how the residents responded to the activity, at least once during the 4-week intervention.

The carers assigned to the control activity attended a 1-hour online training session where the researchers explained the procedure for the general conversation activity and provided general communication tips.

In both conditions, carers were advised to conduct the activities with the resident twice a week for 4 weeks, with each session lasting 30 minutes. They were also advised to invite the resident to a quiet room in the ward where they could conduct the activity without disruption. Carers assessed residents' feelings before and after each session using the Smiley Face Assessment Scale (SFAS) [30]. Trained female and male research assistants observed the first and last sessions via video calls, assessing social interaction between the resident and carer during the activity using the INTERACT observation scale [13]. After these sessions, carers in the photo activity group received feedback from the research assistants on how to improve person-centered communication based on their observations. At the end of the final session, the residents with dementia were interviewed during the video call for 5 to 10 minutes by the research assistant, using questions from the semistructured interview guide. Both informal and formal carers either completed the digital form or participated in a 20-minute interview via telephone or video call at a time convenient to them within a week after the final intervention session. Carers who were unable to attend an interview were asked to submit a written narrative account of their experience in response to the questions from the interview guide. Formal carers who conducted either the photo activity or control activity with >1 resident were asked to complete a separate interview for each resident because some of the questions could be answered differently depending on the resident.

The independent assessors completed questionnaires on the residents' daily functioning in the ward at baseline (T0), after the intervention (T1), and at the 2-week follow-up (T2). Results relevant to the observations of the independent assessors will

XSL•FO
RenderX

be reported in a separate paper on the effect evaluation of the digital photo activity.

Analysis

Quantitative Analyses

Participants' background characteristics were analyzed using descriptive statistics in SPSS software (version 28.0; IBM Corp). Differences between the photo activity and control activity conditions were tested using 2-tailed *t* tests for continuous variables, Mann-Whitney *U* tests for ordinal variables, and Pearson chi-square tests for nominal variables. Effect sizes were interpreted using Cohen *d* for *t* tests (small=0.2, medium=0.5, and large is ≥0.8), *r* for Mann-Whitney *U* tests (small=0.1, medium=0.3, and large is ≥0.5), and phi coefficient (ϕ; for 2 × 2 tables) or Cramér V (for tables larger than 2 × 2) for Pearson chi-square tests, with the effect size based on the number of categories involved.

Data from closed-ended questions were summarized as percentages or as means and SDs, depending on the measurement level. As not all questions were comparable between the photo activity and control groups—some questions were intervention related—only selected comparable closed-ended questions were analyzed for between-group differences for each of the participant groups (residents, formal carers, and informal carers) using chi-square tests, *t* tests, or Mann-Whitney *U*-tests.

Google Analytics recorded information from users identified as "carer-resident" pairs with unique user IDs. The first sessions for the trial started on February 15, 2021; however, because of a delay in setting up the live environment for Google Analytics for the Fotoscope web application, data collection only started on March 12, 2021. Data collection was completed on August 6, 2023. For privacy reasons, Google Analytics only retains data for 26 months [31], and data were not backed up. As a result, data from 10 users who participated in the trial from March to October 2021 were automatically deleted before analyses were conducted in August-September 2023.

Observed session durations extracted from Google Analytics were matched to the dates and times recorded in the carer self-report registration forms. The agreement between total observed and self-reported durations (in minutes) was analyzed using the Spearman rho correlation coefficient. Percentages of total unique page views for the main pages were calculated from the raw data. The frequency of selecting main photo categories was also calculated from raw data (unique page views) and summarized as percentages. Individual reports were viewed to track events or actions taken by specific users during each session (eg, viewing the Profile page or opening a photo from the Photo Selection page). These actions were recorded per user on a binary (yes or no) scale and reported as the percentage of sessions in which events of interest (eg, adding a photo to the Favorites or Photo Selection pages) occurred. The use of a tablet or a different device for the Fotoscope web application was counted per user and summarized as percentages.

Qualitative Analyses

Data from the open-ended questions in the semistructured interviews were analyzed using the framework of theoretical thematic analysis, which was defined by Braun and Clarke [32] as thematic analysis that is built upon existing theoretical work in the literature. The researchers (JROT: PhD student, English speaking; MV: master's student, English and Dutch speaking; and SAMS: senior researcher, English and Dutch speaking) worked together following the 6 phases of thematic analysis [32]. The researchers individually familiarized themselves with the data (phase 1) by reading all answers to the open-ended and closed-ended questions of the semistructured interviews. Next, they generated initial codes (phase 2) based on aspects of the MRC guidance (contextual, implementation, and mechanism-of-impact factors) [17], which served as the guiding framework for this thematic analysis. The researchers studied the data to examine how the data fit within the MRC guidance framework; thus searching for themes (phase 3) involved assessing whether the data from the semistructured interviews fit within the main themes of context, implementation, and mechanisms of impact factors.

Themes and codes were reviewed (phase 4) by the 3 researchers through several online meetings. Themes and subthemes were then named and defined (phase 5). As stated by Braun and Clarke [32], these phases are recursive, and the researchers moved back and forth between phases as necessary. Reporting of the results (phase 6) followed the COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines [33]. MV and SAMS worked with the original data in Dutch, while JROT used an English translation generated using DeepL [34]. The researchers agreed on the English translations of the quotes and developed the codes and themes in English.

Results

Characteristics of Participants

Of the 83 resident–informal carer pairs recruited for the RCT, 16 (19%) dropped out for different reasons after randomization (Figure 3). Among the experimental group (n=32 resident–informal carer pairs), 29 residents and 28 informal carers completed semistructured interviews after the final session. Among the control group (n=35 resident–informal carer pairs), 29 residents and 27 informal carers completed the interviews. Reasons for not completing the interviews are provided in Figure 3. No informal carers delivered the experimental or control intervention, partly due to restrictions in visiting the nursing home during the COVID-19 pandemic and partly because of other reasons, such as not having enough time to visit the ward as needed because of their job or other concurrent tasks.

Of the 61 formal carers recruited for the RCT, 42 (69%) completed the trial and the semistructured interviews (carers experimental group: n=18, 43%; carers control group: n=24, 57%). As a number of carers conducted either the experimental activity or control activity with >1 resident, they were asked to complete a separate interview for each resident; hence, 50 semistructured interviews with residents (experimental group:

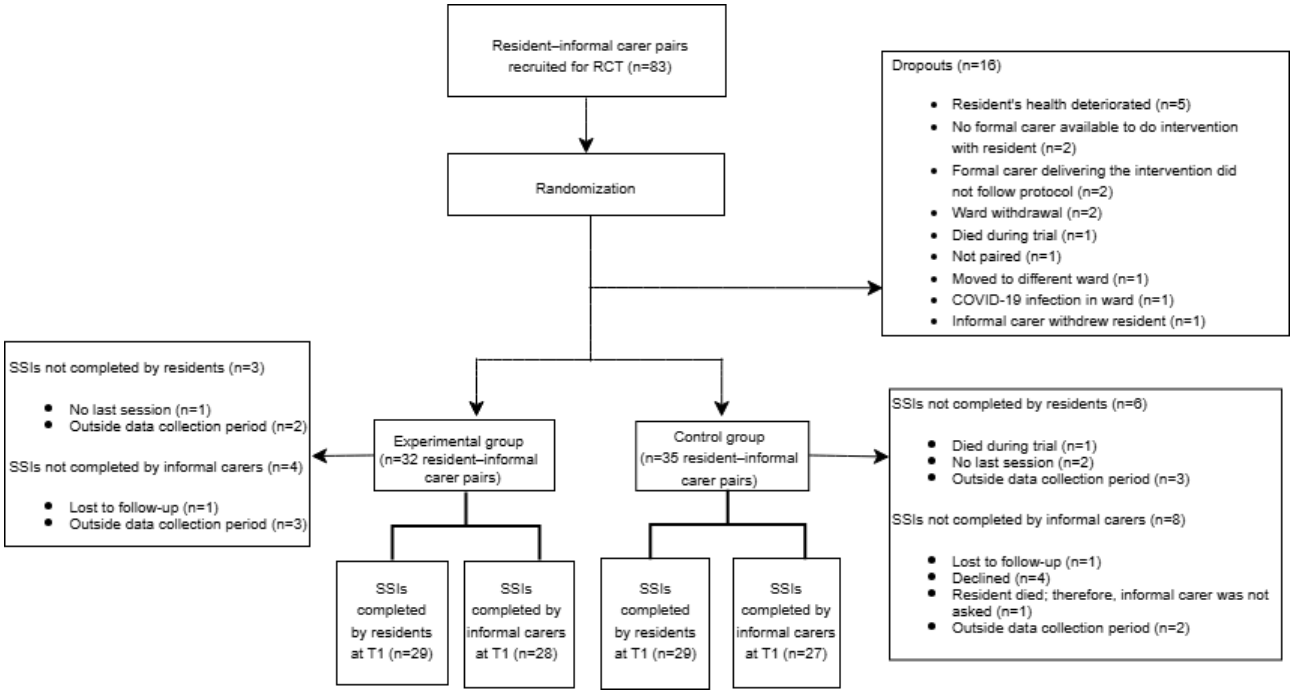
XSL•FO
RenderX

n=23, 46%; control group: n=27, 54%; Figure 4) were completed by formal carers.

Tables 1, 2 and 3 present the background characteristics of the residents, informal carers, and formal carers, respectively, who completed the semistructured interviews. There were no significant differences between the experimental and control groups.

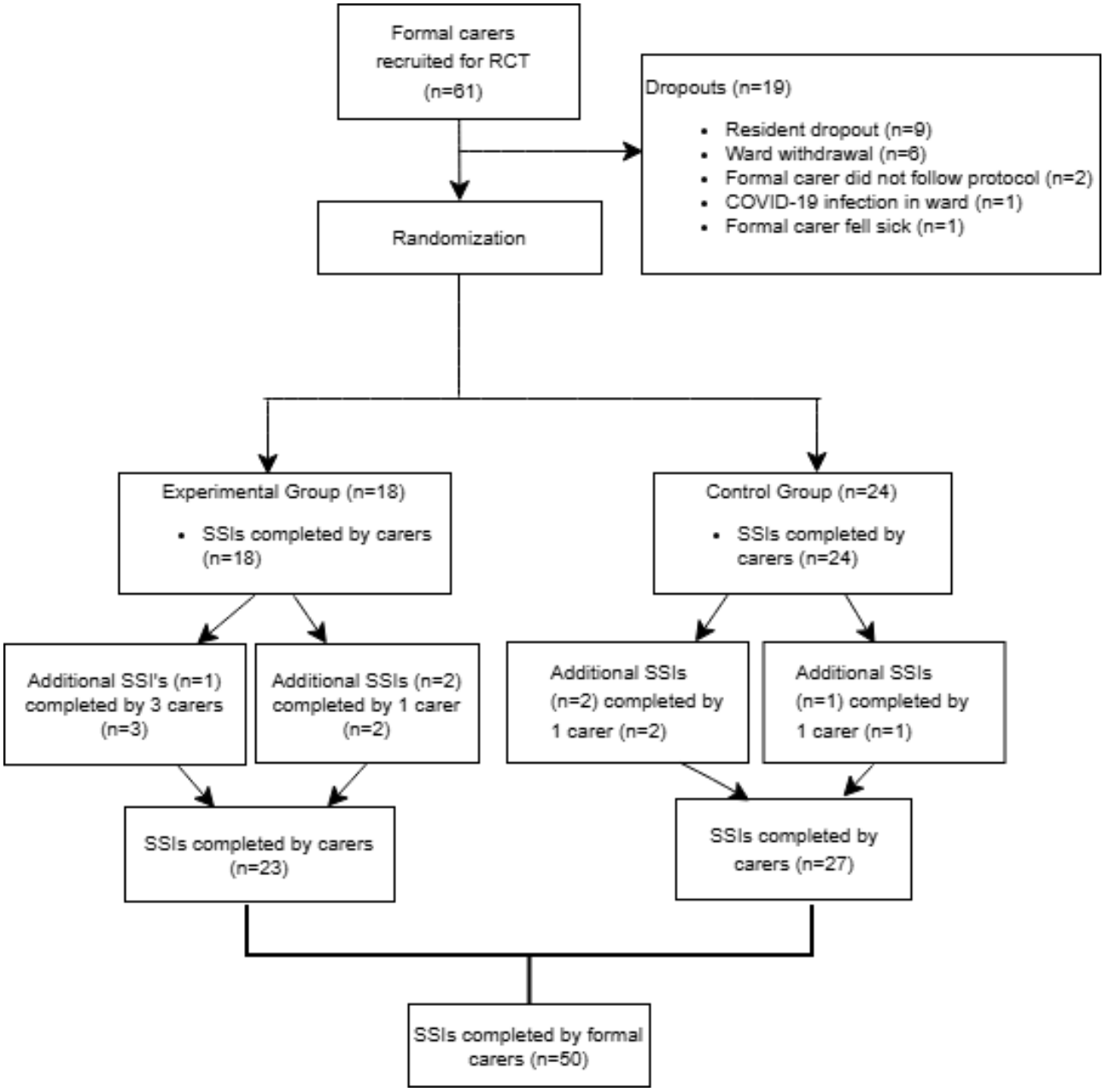
Figure 5 shows the thematic map with the 3 main themes identified in the qualitative thematic analysis, while Multimedia Appendix 3 provides a detailed overview of the themes, subthemes, and sample data extracts. Relevant qualitative results from the thematic analysis are integrated with the quantitative results in the following subsections.

Figure 3. Recruitment of residents and informal carers, showing randomization into the experimental and control groups, as well as dropouts and reasons for dropping out. RCT: randomized controlled trial; SSI: semistructured interview; T1: after the intervention.



XSL•FO
RenderX

Figure 4. Recruitment of formal carers, showing randomization into the experimental and control groups, as well as dropouts and reasons for dropping out. RCT: randomized controlled trial; SSI: semistructured interview.



XSL•FO
RenderX

Table 1. Background characteristics of nursing home residents in both the experimental and control groups who completed the semistructured interviews after the 4-week intervention.

Characteristics	Experimental group (n=29)	Control group (n=29)	Difference tests				*P* value[a]
			Chi-square (*df*)	*t* test (*df*)	*U*	*z*	
Sex: female, n (%)	24 (83)	22 (76)	0.1[b] (1)	—[c]	—	—	.75
Age (y), mean (SD; range)	84.03 (7.36; 68-96)	84.38 (7.48; 62-96)	—	−0.177 (56)	—	—	.86
Education, n (%)			—	—	308.0	−1.502	.13
Elementary	7 (24)	6 (21)					
Lower secondary education	19 (66)	11 (38)					
Higher secondary education	1 (3)	5 (17)					
University	2 (7)	5 (17)					
Missing	0 (0)	2 (7)					
Time spent in nursing home (y), mean (SD; range)	1.714 (1.460; 0.3-7.0)	2.038 (2.090; 0.2-10.0)	—	—	365.5	−0.427	.68
Type of dementia, n (%)			1.1[b] (1)	—	—	—	.30
Alzheimer disease	17 (59)	21 (72)					
Other	12 (41)	7 (24)					
Missing	0 (0)	1 (3)					
Severity of dementia (GDS[d] scores), n (%)			—	—	398.5	−0.380	.70
4	4 (14)	3 (10)					
5	15 (52)	15 (52)					
6	10 (34)	11 (38)					

[a]Significance level set at $P<.05$.

[b]Chi-square test with Yates continuity correction.

[c]Not applicable.

[d]GDS: Global Deterioration Scale.

XSL•FO
RenderX

Table 2. Background characteristics of informal carers in both the experimental and control groups who completed the semistructured interviews after the 4-week intervention.

Characteristics	Experimental group (n=28)	Control group (n=27)	Difference tests				*P* value[a]
			Chi-square (*df*)	*t* test (*df*)	*U*	*z*	
Sex: female, n (%)	21 (75)	20 (74)	0.0[b] (1)	—[c]	—	—	.99
Age (y), mean (SD; range)	57.04 (9.03; 31-74)	60.74 (11.4; 27-84)	—	−1.341 (53)	—	—	.39
Education, n (%)			—	—	317.5	−1.108	.22
Elementary	0 (0)	2 (7)					
Lower secondary education	17 (61)	9 (33)					
Higher secondary education	3 (11)	4 (15)					
University	8 (29)	12 (44)					
Missing	0 (0)	0 (0)					
Relationship to resident, n (%)			1.6 (3)	—	—	—	.65
Husband, wife, or partner	4 (14)	6 (22)					
Son or daughter	3 (11)	4 (15)					
Daughter-in-law or son-in-law	18 (64)	16 (59)					
Other	3 (11)	1 (4)					
Paid work: yes, n (%)	21 (75)	15 (56)	1.5[a] (1)	—	—	—	.22

[a]Significance level set at $P<.05$.

[b]Chi-square test with Yates continuity correction.

[c]Not applicable.

XSL•FO
RenderX

Table 3. Background characteristics of formal carers in both the experimental and control groups who completed the semistructured interviews after the 4-week intervention.

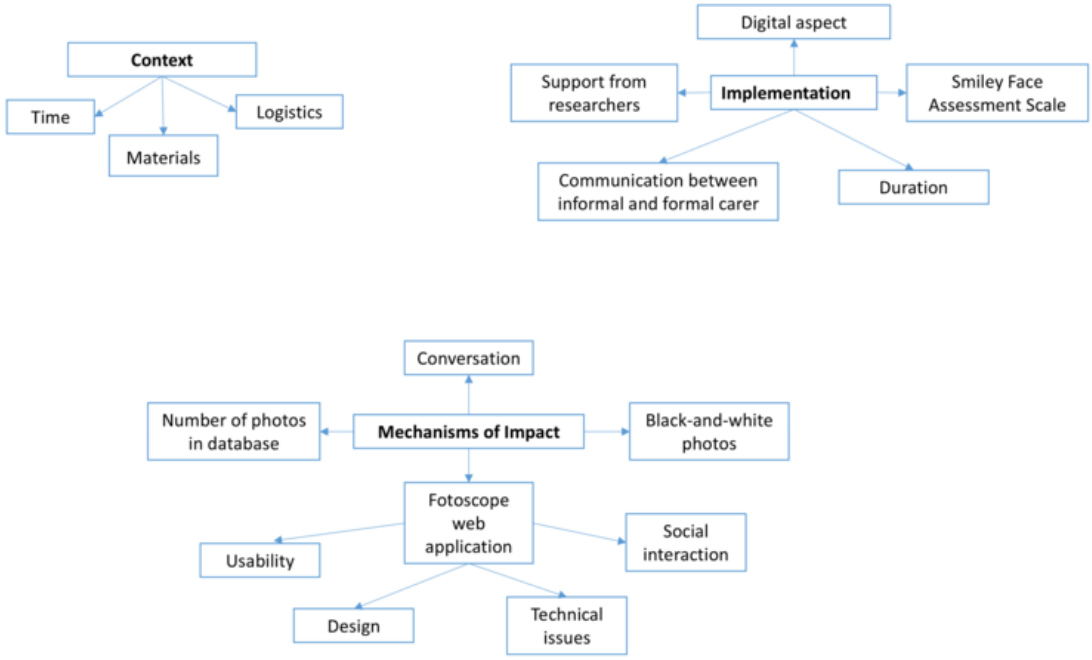
Characteristics	Experimental group (n=18)	Control group (n=24)	Difference tests				*P* value[a]
			Chi-square (*df*)	*t* test (*df*)	*U*	*z*	
Sex: female, n (%)	17 (94)	23 (96)	0.0[b] (1)	—[c]	—	—	.99
Age (y), mean (SD; range)	45.7 (13.0; 20-62)	43.6 (15.1; 18-63)	—	−0.470 (40)	—	—	.64
Education, n (%)			—	—	194.5	−0.843	.40
Elementary	0 (0)	0 (0)					
Lower secondary education	16 (89)	19 (79)					
Higher secondary education	1 (6)	2 (8)					
University	1 (6)	3 (12)					
Missing	0 (0)	0 (0)					
Job role, n (%)			2.5 (3)	—	—	—	.48
Carer	6 (33)	5 (21)					
Nurse	1 (6)	0 (0)					
Activity supervisor	3 (17)	6 (25)					
Other	8 (44)	13 (54)					
Experience (y) in psychogeriatrics, mean (SD; range)	5.22 (4.62; 1.0-16.0)	10.75 (11.13; 1.0-40.0)	—	—	165.5	−1.297	.20

[a]Significance level set at $P<.05$.

[b]Chi-square test with Yates continuity correction.

[c]Not applicable.

Figure 5. Thematic map from qualitative analyses showing the main themes (context, implementation, and mechanisms of impact), subthemes, and subsubthemes.



XSL•FO
RenderX

Results From Analysis of Semistructured Interviews and Google Analytics Data

Overview

Multimedia Appendix 2 presents all answers to the closed-ended questions from the semistructured interviews summarized as percentages or as means and SDs.

In total, 41 users (representing carer-resident pairs) were given a Fotoscope account. Of these 41 users, 38 (93%) were recorded as accessing the app (of the remaining 3 users, 2, 67% had started the intervention before Google Analytics was set up; and 1, 33% included a resident who dropped out before starting the intervention). As data of 10 (26%) of the 38 users were no longer available at the time of evaluation, we analyzed the data of 28 (74%) users. Of these 28 users, 3 (11%) had partial or missing data. Of the remaining 25 user pairs, 4 (16%) had residents who dropped out (n=2, 50% dropped out before the activities started due to worsening conditions; n=1, 25% dropped out after 3 sessions due to worsening conditions; and n=1, 25% dropped out after 3 sessions because the formal carer was too busy); therefore, they were excluded from analysis, leaving data from 21 (84%) users (representing carer-resident pairs) for the final analysis.

Contextual Factors

The majority of the residents (experimental group: 23/29, 79%; control group: 27/29, 93%) found the space for viewing photos with their formal carers or having a conversation with them to be comfortable and quiet.

Most of the informal carers in the photo activity group (13/28, 46%) experienced no barriers to talking to the formal carers about the residents' personal interests or their experience with the photo activity, while some cited barriers, including a lack of time (7/28, 25%) and "other reasons" (4/28, 14%), or did not respond (4/28, 14%). The majority of informal carers in the control group (15/27, 56%) reported no barriers to talking to the formal carers about the residents' experience with the general conversation activity, while those who responded with "other reasons" (10/27, 37%) reported not knowing that it was an option to ask the formal carers about the general conversation activity with the residents. Other reported barriers included being on vacation (1/27, 4%) and a lack of time (1/27, 4%).

The majority of formal carers in the photo activity group (10/18, 56%) reported using a tablet "a lot" previously. The remaining responded as follows: "a little" (5/18, 28%), "hardly" (1/18, 6%), or "not at all" (2/18, 11%). The majority of the formal carers in both the photo activity group (14/18, 78%) and the control group (15/24, 63%) reported experiencing no barriers in conducting the activities. In the thematic analysis, formal carers from both groups who reported experiencing barriers mentioned difficulty in planning time for the activities, given their busy schedules in the ward, sometimes even having to find the time outside of their shifts to conduct the activity with the resident:

> *Planning the photo activity is very important because of our workload. Time and sufficient staff presence are very essential to consider.* [Formal carer 109]

Finding a quiet place to conduct the activity seemed to be more difficult for formal carers in the control group, while formal carers in the photo activity group mentioned difficulty in arranging materials for the activity, such as finding an available tablet in the ward (we had offered to lend tablets to the wards at the start of the study, but none chose this option).

Implementation Factors

The majority of the residents (25/29, 86%) reported that the photo activity went well as an activity on the tablet, and the majority (21/29, 72%) also reported seeing the photos clearly and finding them sufficiently large and sharp. Moreover, the majority (23/29, 79%) reported viewing a "good" amount of photos with their carers, and most (14/29, 48%) felt "neutral" about the black-and-white photos.

The majority of the informal carers (experimental group: 25/28, 89%; control group: 20/27, 74%) felt well informed about the activities that the residents performed in the nursing home during the intervention period. However, the thematic analysis revealed that informal carers in both groups desired more communication and updates from the formal carers regarding the residents' engagement in the activities.

A majority of formal carers in both the photo activity (12/18, 67%) and control (20/24, 83%) groups felt that the medium and format of the training sessions were effective (Table 4). More than half of the formal carers in the photo activity group were positive about the online training sessions and rated them as "good" (5/18, 28%) or "satisfactory" (8/18, 44%), and the majority (12/18, 67%) also felt that the training duration was "just right." A large majority of the formal carers in the control group also rated their online training sessions positively (824, 33% rated it as "good"; and 14/24, 58% rated it as "satisfactory"), and a large majority (20/24, 83%) also felt that the training duration was "just right." Formal carers from both groups expressed a preference for a live training session and in-person support or demonstration. A majority of the formal carers from both the photo activity (14/18, 78%) and control (16/24, 67%) groups felt that the support provided by the researchers and resources was satisfactory. In the thematic analysis, several formal carers in both the photo activity and control groups commented on the duration of the online training sessions and the duration of the activity itself. Regarding the duration of the online training sessions, they seemed to agree that it should be shorter for both groups. The comments for activity duration were mixed. Formal carers in both the photo activity and control groups appreciated the support provided by the researchers and resources. A formal carer in the control group suggested incorporating a video example in the training sessions to demonstrate how the general conversation activity could be conducted, while a formal carer in the photo activity group suggested expanding the manual to include instructions on how to download the Fotoscope app on a new device.

The majority of the formal carers (16/23, 70%) found that the photo activity, using the Fotoscope app, was "easy to implement." While the carers were positive about the app, a theme revealed by the thematic analysis was that they felt that the Fotoscope app was "not for everyone":

XSL•FO
RenderX

> *Very good [app]. Albeit not suitable for everyone. Everything digital, this is something that the target group is not familiar with.* [Formal carer 5]

Formal carers who expressed this concern had conducted the photo activity with residents who had a GDS score of 5 or 6 [23]. A minority of carers in both groups also reported that residents with a GDS score of 5 or 6 had a hard time understanding the SFAS [30] and found it difficult to choose a face that represented how they felt in the moment. Thematic analysis revealed that more formal carers from the control group reported that it was difficult for the residents to understand the semistructured interview questions. Formal carers who conducted the activities with multiple residents also reported that completing the interview multiple times was a burden.

Table 4 presents results from the statistical analysis of differences between the formal carers in the photo activity group and those in the control group in terms of how they felt about the medium and format of the training sessions, the activity duration, the 4-week activity period, and the activity frequency. No significant differences were found (Table 4).

The relationship between the total duration of self-reported app use as measured by the carer self-report registration form and the corresponding (ie, matched to the date and time recorded in the self-report form) total duration of observed app use as measured by Google Analytics was investigated using the Spearman rho correlation coefficient. There was no significant correlation between the 2 variables (n=21; r=−0.108; P=.64).

The Google Analytics data showed that desktop computers were used most of the time to access the Fotoscope app (19/32, 59%), followed by tablets (8/32, 25%) and mobile phones (5/32, 16%). Formal carers in the photo-activity group sometimes used more than one device during the intervention period, which is accounted for in the total devices used (N=32).

Table 4. Results regarding implementation factors for formal carers in the photo activity (experimental) and general conversation (control) groups based on comparable questions from the semistructured interviews after the 4-week intervention.

Outcome variable	Experimental group (n=18), n (%)	Control group (n=24), n (%)	Difference tests		P value[a]	Effect size	
			U	Chi-square (*df*)		*r*	Cramér V
Medium and format training			209.5	—[b]	.83	−0.03	—
Very effective	2 (11)	1 (4)					
Effective	12 (67)	20 (83)					
Somewhat effective	3 (17)	2 (8)					
Not effective	1 (6)	1 (4)					
Activity duration[c]			—	0.4 (2)	.84	—	0.09
Too short	1 (4)	2 (7)					
Just right	13 (56)	13 (48)					
Too long	9 (39)	11 (41)					
Missing	0 (0)	1 (4)					
4-wk activity period			—	1.1 (2)	.59	—	0.16
Too short	2 (11)	1 (4)					
Just right	9 (50)	15 (62)					
Too long	7 (39)	8 (33)					
Activity frequency			—	0.8 (2)	.67	—	0.14
Too little	0 (0)	1 (4)					
Just right	10 (56)	12 (50)					
Too often	8 (44)	11 (46)					

[a]Significance level set at P<.05.

[b]Not applicable.

[c]The n varies for Activity duration, as some formal carers were asked to answer this question multiple times if they did the photo-activity or control activity with more than one resident. Thus formal carers in the photo-activity is n=23, and control activity n=27, as in Figure 4.

Mechanism-of-Impact Factors

The majority of residents in the photo activity group (25/29, 86%) reported that they would like to continue participating in the activity with their formal carers. The residents gave the Fotoscope app a score of 8.6 out of 10 in terms of enjoyment. Thematic analysis revealed that the residents in the photo activity group responded more positively when asked whether they wanted to continue performing the activity, especially if new photos were introduced.

XSL•FO
RenderX

A large majority of the formal carers in the photo activity group reported that the reaction of the resident to the photo activity was "very positive" (10/23, 43%) or "positive" (8/23, 35%). The formal carers in the control group predominantly reported that the reaction of the resident to the general conversation activity was "positive" (12/27, 44%), while a minority reported that the reaction was "very positive" (5/27, 19%). Table 5 presents the results from the statistical analysis of differences between residents (in terms of feeling heard, feeling better known by their carers, and their experience of the general conversation), informal carers (whether they heard anything back from the resident about the activity), and formal carers (whether they thought the activity was enjoyable for the resident, was a good way to get to know the resident better, and how much better they got to know the resident from the activity) in the photo activity and control groups.

There was a significant difference between the photo activity and control groups in terms of how the residents felt about having conversations with their formal carers, with a medium effect size (U=533.0, z=2.865, r=0.39; P=.004), favoring the photo activity group (median 4.00, "very nice," n=28) over the control group (median 3.00, "nice," n=27).

There was also a significant difference in how much better the formal carers in the photo activity group (median 2, "yes, a little better," n=22) got to know the resident in the past month compared to those in the control group (median 2, "yes, a little better," n=27; U=390.5, z=2.114, r=0.302 [medium effect]; P=.04). The answer options in the semistructured interviews for the photo activity group were as follows: 1=no, not better; 2=yes, a little better; and 3=yes, much better. As the answer options for this question for the control group differed (an additional answer option of "yes, better" was provided after "yes, a little better"), it was decided to combine the answers "yes, better" and "yes, a little better" from the control group interviews in the analysis. Of the formal carers in the experimental group, 48% (11/23) and 39% (9/23) reported getting to know the resident with dementia "a little better" and "much better," respectively (compared to 17/27, 63% and 4/27, 15%, respectively, of the formal carers in the control group). Of the formal carers in the experimental group, 9% (2/23) reported that they did not get to know the person with dementia "better" compared to 22% (6/27) of the formal carers in the control group. Thematic analysis reflected similar findings: formal carers in both the photo activity and general conversation groups mentioned learning more about the residents:

> *During care, of course we chat plenty but here I went a bit deeper into the past, so yes I did get to know her much better than I already did.* [Photo activity group, formal carer 201]

> *I now know a little more about her past, as far as she could still tell.* [Control group, formal carer 204]

However, a formal carer in the control group did mention that the general conversation activity was similar to what she already did as part of her work:

> *Personally it has no added value for me. This is what I do daily in my work—speak to residents every day.* [Formal carer 116]

Finally, the carers in the photo activity group (n=23) gave a significantly higher rating to the activity (median 9.00) compared to the carers in the general conversation group (n=27; median 8.00) as a way to get to know the resident with dementia better (U=419.0, z=2.169, r=0.307 [medium effect size]; P=.03).

There was no statistically significant association between the experimental and control groups on whether residents felt heard during the conversation with their formal carer, whether informal carers received feedback from the residents about the activity, and how much formal carers felt the residents enjoyed the activities (Table 5).

XSL•FO
RenderX

Table 5. Results regarding the mechanism-of-impact factors for residents, informal carers, and formal carers in the photo activity (experimental) and general conversation (control) groups based on comparable questions from the semistructured interviews after the 4-week intervention.

Outcome variable	Experimental group	Control group	Difference tests			*P* value[a]	Effect size	
			Fisher exact test	Chi-square (*df*)	*U*		ϕ	*r*
Residents, n/N (%)								
Feeling heard: yes	26/29 (90)	24/29 (83)	—[b]	—	—	.49	—	—
Felt better known by their carer			—	0.3[c] (1)	—	.60	−0.12	—
A lot better or better	17/29 (59)	15/29 (52)						
A little better or not better	6/29 (21)	9/29 (31)						
Missing	6/29 (21)	5/29 (17)						
Experience of general conversation			—	—	533.0	.004	—	0.39
Very nice	16/29 (55)	6/29 (21)						
Nice	11/29 (38)	15/29 (52)						
Somewhat nice	0/29 (0)	5/29 (17)						
Not nice	1/29 (3)	1/29 (3)						
Missing	1/29 (3)	2/29 (7)						
Informal carers, n/N (%)			—	—	—	—	—	—
Resident talked to them about the activity: yes	7/28 (25)	5/27 (19)						
Formal carers			—	—	264.0	.20	—	.20
Enjoyable for people with dementia in a nursing home (0-10), median (IQR)	8.00 (7-9)	8.00 (7-8)						
Knew resident better, n/N (%)			—	—	390.5	—	—	.30
Yes, much better	9/23 (39)	4/27 (15)				.04		
Yes, a little better	11/23 (48)	17/27 (63)						
No, not better	2/23 (9)	6/27 (22)						
Missing	1/23 (4)	0/27 (0)						
Way of getting to know resident with dementia better (0-10), median (IQR)	9.00 (7-9)	8.00 (6-8)	—	—	419.0	.03	—	0.31

[a]Significance level set at $P<.05$.

[b]Not applicable.

[c]Chi-square test with Yates continuity correction.

Experienced Usability

In the thematic analysis, the formal carers in the photo activity group described the Fotoscope app as easy to use, simple, clear, and visually appealing. Technical issues were experienced (eg, difficulties logging in or the app not working on the tablet). However, carers mentioned that the Fotoscope app enhanced social interaction, helping them to get to know the resident better:

> *It triggers more conversation, especially for the quieter ones, this can work up to a conversation.* [Formal carer 201]

Experienced Usefulness

The majority of the formal carers in the photo activity group found the preparation for the photo activity useful (14/18, 78%) or very useful (2/18, 11%), while a small minority found it not useful (2/18, 11%). Whereas the majority of the formal carers in the photo activity group found the black-and-white photos "pleasant" (4/18, 22%) or "neutral" (7/18, 39%), some found them "unpleasant" (7/18, 39%). A number of formal carers in the photo activity group noted that some photos would really have been better with color:

> *I was sorry we didn't have color photos. Some things really needed color. Even my client asked me why I*

XSL•FO
RenderX

> *didn't have color photos.* [Formal carer 105 with resident 109]

However, formal carers also mentioned that while they preferred color photos, residents did not mind viewing black-and-white photos:

> *I would personally like it better if the pictures were in color, but for the resident this did not matter.* [Formal carer 105 with resident 121]

> *[T]hey [residents] mostly liked black-and-white pictures too.* [Formal carer 101]

The number of photos in the database was identified as a subtheme in the thematic analysis. While the formal carers described the selection as varied, some noted a lack when residents had specific interests; for example, a resident wanted to see more ballet pictures.

Google Analytics showed that while all users viewed the Profile page, 19% (4/21) did not add photos to the Photo Selection page, and only 48% (10/21) used the photo selection tool on the Profile page to select photos. In addition, while the majority of users (17/21, 81%) opened photos from the Photo Selection page as instructed, 19% (4/21) apparently accessed photos through other means. Furthermore, only 33% (7/21) of the users used the Favorites page.

Learnability

The majority of formal carers in the photo activity group (12/18, 67%) said that it was easy to learn how to use the Fotoscope web application. Google Analytics showed that the most viewed page of the main pages was the Profile page (257/594, 43.3% total unique page views), followed by the general Themes page (104/594, 17.5%), Favorites page (85/594, 14.3%), "joker" page (57/594, 9.6%), and Search (43/594, 7.2%). The User Guide (29/594, 4.9%) and Information page (19/594, 3.2%) were the least viewed pages. The most frequently selected main photo category (based on Google Analytics' unique events count) was people (71/404, 17.6%), followed by activities (64/404, 15.8%), and places and animals (both 56/404, 13.9%). The least selected main photo categories were experiences (53/404, 13.1%), nature (53/404, 13.1%), and things (51/404, 12.6%).

Adoption

The majority of formal carers in the photo activity group (14/18, 78%) agreed that the Fotoscope app can be used as part of daily care activities in the nursing home.

Discussion

Principal Findings and Comparisons With Prior Work

Overview

The goal of this study was to determine contextual, implementation, and mechanism-of-impact factors [17] that may have facilitated or hindered the implementation of the photo activity intervention for residents with dementia and potentially influenced the intervention outcomes.

To answer the RQs, we conducted a mixed methods process evaluation alongside a pilot RCT investigating the feasibility and effectiveness of the photo activity for residents and their informal and formal carers. The main findings of this process evaluation were as follows. For contextual factors affecting the implementation of the intervention (RQ1), formal carers struggled to find the time to conduct the activity during their work hours. For implementation factors, which looked at how users assessed the preparation, introduction, and delivery of the interventions (RQ2), it was found that viewing digital photos on a tablet was feasible for residents with dementia; formal carers from both groups were mostly positive and satisfied regarding the online training sessions but suggested that, in the future, they would like shorter sessions as well as in-person training and support; and informal carers would like more updates from formal carers. For the mechanisms of impact (RQ3), looking into how users experienced and valued the digital photo activity versus the control activity, it was found that residents who performed the digital photo activity with their carers had a more positive experience compared to residents in the control activity. Formal carers who conducted the digital photo activity reported getting to know the residents "much better" compared to formal carers in the control activity. The Fotoscope app as a tool to deliver the digital photo activity was also rated as a better way to get to know the residents as opposed to having a general conversation. Informal carers of residents in the photo activity group also reported more contact with formal carers compared to informal carers in the control activity.

In the following subsections, we provide a more detailed discussion of the findings in relation to each RQ.

How Did Factors External to the Intervention Affect the Implementation of the Intervention (Context)?

While minimal contextual factors impeded the implementation of the photo activity or the control activity, a salient theme that emerged was the difficulty that formal carers experienced in scheduling time to conduct the activities with the resident alongside their daily care tasks. In some cases, formal carers mentioned having to conduct the activity outside of their work shift. Problems with workload and staffing are common barriers to implementing psychosocial interventions successfully in nursing homes [14]. Staff are pressured to prioritize tasks that are related to physical and safety needs [35]. Due to the restrictions on visiting nursing homes during the COVID-19 pandemic [36], during the initial phase of the study, no informal carers were involved in delivering either activity to residents. Later, no informal carers were willing to invest the time needed to provide the intervention or control activity.

It is not clear which device carers used to prepare for the photo activity and which device they used to conduct the activity; therefore, it is possible that carers used a desktop computer to prepare and select initial photos for the activity and then used the tablet as instructed during the activity with the resident.

How Did Residents and Their Informal and Formal Carers Assess the Preparation, Introduction, and Delivery of the Interventions (Implementation)?

Most residents responded positively to viewing photos on a tablet, consistent with literature suggesting that using tablets in

XSL•FO
RenderX

art-based psychosocial interventions could have positive effects for individuals with dementia [37,38].

Informal carers from both groups mentioned wanting more updates from formal carers on how the residents were responding to the activities. A previous study found that greater involvement of informal carers in the daily care of residents was an important factor in the successful implementation of psychosocial interventions because this also helps formal carers to become more familiar with the residents and their identity within a family context [39]. The communication between informal and formal carers fosters more trust, which benefits overall care [14].

Regarding the online training sessions, while the majority of the formal carers from both groups were positive and satisfied, formal carers from both groups who were dissatisfied with the training sessions suggested shortening the duration and expressed a preference for live training and in-person support or demonstration. The setting of the nursing home itself may also have influenced how staff received the training. Kuske et al [40] found that when staff feel overburdened, they see training as an additional task rather than beneficial continuing education, which could explain why shortening the duration of the training sessions was a salient theme in this process analysis.

A minority of formal carers who conducted the photo activity with residents with a more severe level of dementia (ie, a GDS score of 5 or 6) [23] described the digital photo activity as not suitable and the 30-minute activity duration as too long for these residents due to their "short attention spans" or because "they get tired easily." This is similar to previous work suggesting that residents' deteriorating health could be a barrier for formal carers implementing psychosocial interventions [40].

However, it is worth noting that the majority of the residents had a GDS score of 5 or 6 [23], and the experience of most formal carers was that the photo activity intervention was easy to implement and that the activity duration was just right. This suggests that the photo activity can be a suitable addition to tools that nursing homes could use, including for residents with more severe dementia, because most interventions for people with more severe dementia are focused on reducing problem behaviors rather than improving social interactions or quality of life [41,42].

Some residents experienced difficulty in responding to the SFAS [30], which presented 5 emotions or faces. Future trials could explore using a simpler scale similar to the SFAS but with only 3 faces (happy, neutral, and sad) because a study found that older adults sometimes have difficulty differentiating between "very happy" and "happy" and between "very sad" and "sad" [43].

Google Analytics may not have accurately recorded actual use of the app by formal carers for the process evaluation, as shown by the lack of correlation between self-reported use and observed use via Google Analytics. Currently, photos can be viewed in the Fotoscope web application via the Themes page without logging in; therefore, it is possible that formal carers may not have logged in during some of the sessions and manually searched for photos instead of accessing the Profile page or Photo Selection page. For future trials, it could be useful to consult with software and app developers and explore integrating an in-app analytics function within the Fotoscope app rather than relying on an external analytics provider.

How Did the Residents With Dementia and Their Informal and Formal Carers Experience and Value the Digital Photo Activity Versus the General Conversation Activity (Mechanisms of Impact)?

Residents in the photo activity group reported a more positive experience engaging in the activity with their formal carers. The positive findings regarding using generic photos in the photo activity are in line with a similar study that used generic photos as conversation prompts for people with dementia, where it was found that compared to using family photos, people with dementia told more detailed and personally relevant stories when shown generic photos [42].

More residents in the photo activity group were eager to continue the activity with their formal carers. This highlights the importance of investing in interventions that foster social health for residents with dementia in the nursing home, rather than solely focusing on physical needs or problematic behaviors [41,44].

Informal carers in the photo activity group had more contact with the formal carers, and more residents in the photo activity group talked to their informal carers about the activity. This finding shows that the photo activity may be able to enhance the involvement of informal carers, a factor that is crucial to consider in the successful implementation of psychosocial interventions [14,39].

The formal carers in the photo activity group perceived the residents' reaction to the conversation about the photos as more positive. These carers also reported getting to know the residents much better and rated the Fotoscope app higher (compared to a general conversation) in terms of using it as a tool to get to know the resident better. This is in line with the findings of Lawrence et al [14] indicating that when formal carers feel that an intervention helps them develop a more personal relationship with a resident and that the intervention is meaningful for the resident, they are more likely to continue using the intervention because they see that it has value. This is further supported by the findings from the process evaluation: whereas a formal carer mentioned that the control activity had no added value for her, the majority of formal carers in the photo activity group felt that the intervention could be adopted into daily care because of its value.

The Fotoscope web application was generally well received and was described as having a simple and pleasing design and user interface, which is crucial for engaging users of technological interventions [6,45]. More formal carers were vocal about the black-and-white photos compared to the residents, but only a minority of the formal carers found the black-and-white photos unpleasant. Previous studies have shown that both color and black-and-white photos elicit reactions from residents with dementia equally [42,46]. Residents who participated in the photo activity were also observed to react to the black-and-white

XSL•FO
RenderX

photos in the Fotoscope app and mostly found the photos pleasant or neutral.

Strengths and Limitations

This study has a number of strengths. First, data were collected from nursing home residents, informal carers, and formal carers, providing a wider understanding of the process of implementing the intervention from all viewpoints as well as the contextual and mechanism-of-impact factors that potentially influenced the outcomes of the RCT. Mixed methods were used; quantitative and qualitative data were collected from semistructured interviews with participants, and objective data were obtained from Google Analytics. The semistructured interviews were structured according to the MRC guidance for process evaluations of complex interventions [17]. The digital protocol enabled research assistants to conduct interviews with the residents via video calls, ensuring data collection despite the lockdowns and visiting limitations during the COVID-19 pandemic.

Some methodological limitations included changing the frequency of the interventions after the trial had already been running for 1 month after receiving feedback from the first round of formal carers delivering the interventions. Google Analytics data were available for only 21 (51%) of the 41 users (representing carer-resident pairs) because the researchers were unaware and not informed of the data reaching the storage time limit and being deleted from the account before data collection concluded and data analysis started. The manner of collecting responses for the semistructured interviews with carers varied: some carers answered the interview questions on their own, while others preferred to be interviewed over the phone, which may have influenced their answers differently. In relation to the manner of collecting responses for the interviews, not all closed-ended questions invited an elaborate response. This was done keeping in mind the time constraints and effort required from participants. However, the trained research assistants were instructed to write down any answer that may have been given in place of or in addition to the answers to the closed-ended questions, which enabled us to learn, for example, that residents in the photo activity group were eager to continue performing the activity with their carers with new photos.

Finally, we were unable to request feedback from the participants regarding the results of the thematic analysis. Unfortunately, no informal carers were involved in delivering the control or experimental interventions, despite recruitment efforts aimed at future implementation. As a result, no data could be collected on implementation or mechanism-of-impact factors from informal carers providing the photo activity intervention.

Scientific Relevance of the Study and Recommendations for Future Studies

The mixed methods process evaluation showed that despite the complexity of the intervention, the digital photo activity was feasible to implement via online training sessions and off-site monitoring, providing a good basis for future larger-scale research on the intervention. The MRC guidance was useful in identifying contextual, implementation, and mechanism-of-impact factors that positively or negatively affected implementation and potentially influenced the intervention outcomes. It is recommended that a definitive trial investigating the effectiveness of the intervention should implement feedback from this process evaluation, such as in-person staff training and demonstration of the Fotoscope app. Informal carers could also be recruited to carry out the photo activity, now that COVID-19 restrictions are no longer in place, to explore what implementation issues would be in play for them and how they could benefit from offering the intervention. Incorporating an automatic record of activity appointments in the Fotoscope app instead of asking users to record their time and date of use manually might also provide a more accurate account of their use of the web application.

Clinical Relevance of the Study and Implications for Practice

The mixed methods process evaluation found that residents participating in the photo activity enjoyed it more than those participating in the general conversation activity. This confirms a previous finding that touchscreen art interventions could be beneficial for residents' psychological well-being [6]. Ensuring that tablets are easily accessible in the wards is recommended to encourage staff to conduct the photo activity. While it is noted that the control condition did not include any art-related components, part of the aims of this process evaluation was investigating how the photo activity intervention compared to an active control activity that was most likely to happen naturally within the context of the nursing home (in this case, sitting down with a resident for a chat over a warm drink compared to setting up a different art activity).

Formal carers in the photo activity group got to know the resident with dementia substantially better and rated the photo activity as a better way of getting to know the resident. This finding is relevant because previous literature has found that a more person-oriented care for people living with dementia leads to better outcomes associated with social health and quality of life [29,47], and the photo activity could be an easy and meaningful activity to help formal carers to get to know the residents better. In addition, the importance of formal carers communicating with the informal carers can be encouraged as an essential part of the photo activity.

A majority of the photo activity carers agreed that the intervention could be part of daily care activities, but important contextual factors such as heavy workload and a lack of staff need to be taken into account [14]. Involving the wider organization is important to allow formal carers to view psychosocial interventions as part of daily care tasks, rather than an added workload [48]. For the photo activity, supporting staff to conduct the intervention during their shifts, not outside of paid hours, is recommended, as well as being flexible with the activity duration.

Conclusions

This study showed that the initial training and preparation for formal carers in both the photo activity and control groups, as well as their perception of effort and time investment, did not seem to differ much. The findings suggest that implementing

XSL•FO
RenderX

the photo activity as an intervention does not require a significant amount of additional work, yet it offers notable benefits, including residents enjoying the photo activity more and formal carers getting to know the residents better. Therefore, it is suggested that the digital photo activity, as an enjoyable intervention with its easy to use web application, could be widely implemented and disseminated in nursing homes. As a person-centered intervention designed to enhance relationships among people with dementia in nursing homes and their informal and formal carers, it may have the potential to positively impact outcomes relating to social health and quality of life of the person with dementia, but this needs to be explored further.

Acknowledgments

The authors would like to thank participants who supported this study. The authors would also like to thank the research assistants for assisting with data collection. The authors would like to thank all research assistants involved in interviewing the participants and the participants themselves: the nursing home residents and their families, as well as the nursing home staff. The main randomized controlled trial as well as this process evaluation were funded by the European Union Marie Sklodowska-Curie Innovative Training Network Action program (2020-MSCA-ITN 2018; 813196), RCOAK Foundation, Foundation Hofje Codde and Van Beresteyn, Foundation for Support VCVGZ, Amsterdam UMC, location Vrije Universiteit, and Inholland University of Applied Sciences.

Authors' Contributions

RMD, PB, and JROT were responsible for developing the semistructured interview guide. JROT and MV were responsible for semistructured interview data collection and entry. Thematic analysis was carried out by JROT, MV, and SAMS. JROT, DPN, and RMD were responsible for methodology and analysis of Google Analytics data. JROT and RMD were responsible for methodology and analysis of quantitative data. JROT prepared the original draft. JROT, PB, TPE, RJJG, SAMS, MV, DPN, and RMD reviewed and edited the manuscript. RMD, RJJG, TPE, and SAMS were responsible for supervision. All authors contributed to the submitted manuscript and approved the final version.

Conflicts of Interest

Author SAMS received funding from the Ministry of Health, Welfare and Sports (funding #90001586). This funding was paid to the institution.

Multimedia Appendix 1
Topics for each Medical Research Council guidance for process evaluation component and the measurements used.
[DOCX File , 18 KB - formative_v9i1e56586_app1.docx]

Multimedia Appendix 2
Responses to closed-ended questions from the semistructured interviews with residents and carers.
[DOCX File , 43 KB - formative_v9i1e56586_app2.docx]

Multimedia Appendix 3
Thematic analysis themes, subthemes, and sample data extracts.
[DOCX File , 24 KB - formative_v9i1e56586_app3.docx]

References

1. Aëgerter L. Home page. Photographic Treatment. 2016. URL: https://photographictreatment.com/en/ [accessed 2024-11-14]
2. Theijsmeijer S, de Boo GM, Dröes RM. [A pilot study into person-centred use of photo's in the communication with people with dementia]. Tijdschr Gerontol Geriatr 2018 Sep;49(4):147-155. [doi: 10.1007/s12439-018-0258-4] [Medline: 30088252]
3. Sanders D, Scott P. Literature review: technological interventions and their impact on quality of life for people living with dementia. BMJ Health Care Inform 2020 Jan;27(1):e100064 [FREE Full text] [doi: 10.1136/bmjhci-2019-100064] [Medline: 31948938]
4. Dröes RM, Chattat R, Diaz A, Gove D, Graff M, Murphy K, INTERDEM sOcial Health Taskforce. Social health and dementia: a European consensus on the operationalization of the concept and directions for research and practice. Aging Ment Health 2017 Jan;21(1):4-17 [FREE Full text] [doi: 10.1080/13607863.2016.1254596] [Medline: 27869503]
5. Astell AJ, Bouranis N, Hoey J, Lindauer A, Mihailidis A, Nugent C, Technology Dementia Professional Interest Area. Technology and dementia: the future is now. Dement Geriatr Cogn Disord 2019;47(3):131-139 [FREE Full text] [doi: 10.1159/000497800] [Medline: 31247624]

XSL•FO
RenderX

6. Tyack C, Camic PM. Touchscreen interventions and the well-being of people with dementia and caregivers: a systematic review. Int Psychogeriatr 2017 Aug;29(8):1261-1280 [FREE Full text] [doi: 10.1017/S1041610217000667] [Medline: 28446258]
7. Woods D, Yuan F, Jao YL, Zhao X. Social robots for older adults with dementia: a narrative review on challenges and future directions. In: Proceedings of the 13th International Conference on Social Robotics. 2021 Presented at: ICSR '21; November 10-13, 2021; Singapore, Singapore p. 411-420 URL: https://link.springer.com/chapter/10.1007/978-3-030-90525-5_35 [doi: 10.1007/978-3-030-90525-5_35]
8. van Santen J, Dröes RM, Holstege M, Henkemans OB, van Rijn A, de Vries R, et al. Effects of exergaming in people with dementia: results of a systematic literature review. J Alzheimers Dis 2018;63(2):741-760 [FREE Full text] [doi: 10.3233/JAD-170667] [Medline: 29689716]
9. Wiseman L, Isbel S, Boag A, Halpin-Healy C, Gibson D, Bail K, et al. Online gallery facilitated art activities for people with dementia during the COVID-19 pandemic and beyond: a narrative review. Dementia (London) 2023 Nov;22(8):1950-1976 [FREE Full text] [doi: 10.1177/14713012231198748] [Medline: 37647250]
10. Budak KB, Atefi G, Hoel V, Laporte Uribe F, Meiland F, Teupen S, et al. Can technology impact loneliness in dementia? A scoping review on the role of assistive technologies in delivering psychosocial interventions in long-term care. Disabil Rehabil Assist Technol 2023 Oct;18(7):1107-1119 [FREE Full text] [doi: 10.1080/17483107.2021.1984594] [Medline: 34752177]
11. Tan JR, Boersma P, Ettema TP, Planting CH, Clark S, Gobbens RJ, et al. The effects of psychosocial interventions using generic photos on social interaction, mood and quality of life of persons with dementia: a systematic review. BMC Geriatr 2023 Sep 14;23(1):560 [FREE Full text] [doi: 10.1186/s12877-023-04270-w] [Medline: 37710147]
12. Lorenz K, Freddolino PP, Comas-Herrera A, Knapp M, Damant J. Technology-based tools and services for people with dementia and carers: mapping technology onto the dementia care pathway. Dementia (London) 2019 Feb;18(2):725-741. [doi: 10.1177/1471301217691617] [Medline: 28178858]
13. Tan JR, Boersma P, Ettema TP, Aëgerter L, Gobbens R, Stek ML, et al. Known in the nursing home: development and evaluation of a digital person-centered artistic photo-activity intervention to promote social interaction between residents with dementia, and their formal and informal carers. BMC Geriatr 2022 Jan 06;22(1):25 [FREE Full text] [doi: 10.1186/s12877-021-02632-w] [Medline: 34991472]
14. Lawrence V, Fossey J, Ballard C, Moniz-Cook E, Murray J. Improving quality of life for people with dementia in care homes: making psychosocial interventions work. Br J Psychiatry 2012 Nov;201(5):344-351. [doi: 10.1192/bjp.bp.111.101402] [Medline: 23118034]
15. Darker CD, Nicolson GH, Carroll A, Barry JM. The barriers and facilitators to the implementation of National Clinical Programmes in Ireland: using the MRC framework for process evaluations. BMC Health Serv Res 2018 Sep 24;18(1):733 [FREE Full text] [doi: 10.1186/s12913-018-3543-6] [Medline: 30249262]
16. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. BMJ 2000 Sep 16;321(7262):694-696 [FREE Full text] [doi: 10.1136/bmj.321.7262.694] [Medline: 10987780]
17. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: medical research council guidance. BMJ 2015 Mar 19;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
18. Neal DP, Kuiper L, Pistone D, Osinga C, Nijland S, Ettema T, et al. FindMyApps eHealth intervention improves quality, not quantity, of home tablet use by people with dementia. Front Med (Lausanne) 2023;10:1152077 [FREE Full text] [doi: 10.3389/fmed.2023.1152077] [Medline: 37324141]
19. Mahmoudi Asl A, Kouters S, Castro-González Á, Van der Roest H, Franco Martin M, Dröes RM. Potential facilitators of and barriers to implementing the MINI robot in community-based meeting centers for people with dementia and their carers in the Netherlands and Spain: explorative qualitative study. J Med Internet Res 2023 Aug 02;25:e44125 [FREE Full text] [doi: 10.2196/44125] [Medline: 37531190]
20. Beentjes KM, Kerkhof YJ, Neal DP, Ettema TP, Koppelle MA, Meiland FJ, et al. Process evaluation of the FindMyApps program trial among people with dementia or MCI and their caregivers based on the MRC guidance. Gerontechnology 2020 Aug 31;20(1):1-15. [doi: 10.4017/gt.2020.20.1.406.11]
21. van Santen J, Dröes RM, Blanson Henkemans O, Schoone M, Valk R, Straten AV, et al. Implementation of exergaming for people with dementia: facilitators, barriers, and recommendations. Aging Ment Health 2024;28(2):244-253. [doi: 10.1080/13607863.2023.2238259] [Medline: 37493266]
22. Seen as a person; evaluation of a person-centred tablet intervention to promote the social interaction between people with dementia, family and professional caregivers. Overview of Medical Research in the Netherlands (OMON; formerly Dutch Trial Register).. URL: https://onderzoekmetmensen.nl/en/trial/25910 [accessed 2024-04-29]
23. Reisberg B, Ferris SH, de Leon MJ, Crook T. The Global Deterioration Scale for assessment of primary degenerative dementia. Am J Psychiatry 1982 Sep;139(9):1136-1139. [doi: 10.1176/ajp.139.9.1136] [Medline: 7114305]

XSL•FO
RenderX

24. O'Cathain A, Murphy E, Nicholl J. Why, and how, mixed methods research is undertaken in health services research in England: a mixed methods study. BMC Health Serv Res 2007 Jun 14;7:85 [FREE Full text] [doi: 10.1186/1472-6963-7-85] [Medline: 17570838]
25. Goodman C, Baron NL, Machen I, Stevenson E, Evans C, Davies SL, et al. Culture, consent, costs and care homes: enabling older people with dementia to participate in research. Aging Ment Health 2011 May;15(4):475-481. [doi: 10.1080/13607863.2010.543659] [Medline: 21500014]
26. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008 Apr;13(2):92-98. [doi: 10.1258/jhsrp.2007.007074] [Medline: 18416914]
27. Lutomski JE, Baars MA, Schalk BW, Boter H, Buurman BM, den Elzen WP, TOPICS-MDS Consortium. The development of the Older Persons and Informal Caregivers Survey Minimum DataSet (TOPICS-MDS): a large-scale data sharing initiative. PLoS One 2013;8(12):e81673 [FREE Full text] [doi: 10.1371/journal.pone.0081673] [Medline: 24324716]
28. Gerritsen DL, Oude Voshaar RC. The effects of the COVID-19 virus on mental healthcare for older people in The Netherlands. Int Psychogeriatr 2020 Nov;32(11):1353-1356 [FREE Full text] [doi: 10.1017/S1041610220001040] [Medline: 32491980]
29. Kitwood T. Dementia Reconsidered: The Person Comes First. New York, NY: Open University Press; 1997.
30. Rosenberg F. The MoMA Alzheimer's project: programming and resources for making art accessible to people with Alzheimer's disease and their caregivers. Art Health 2009 Mar;1(1):93-97. [doi: 10.1080/17533010802528108]
31. Data retention-analytics help. Google. URL: https://support.google.com/analytics/answer/7667196?hl=en [accessed 2024-04-29]
32. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
33. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
34. DeepL Translate: the world's most accurate translator. DeepL. URL: https://www.deepl.com/en/translator [accessed 2023-05-12]
35. Gibson G, Chalfont GE, Clarke PD, Torrington JM, Sixsmith AJ. Housing and connection to nature for people with dementia: findings from the INDEPENDENT project. J Hous Elderly 2007 Dec 17;21(1-2):55-72. [doi: 10.1300/j081v21n01_04]
36. Van der Roest HG, Prins M, van der Velden C, Steinmetz S, Stolte E, van Tilburg TG, et al. The impact of COVID-19 measures on well-being of older long-term care facility residents in the Netherlands. J Am Med Dir Assoc 2020 Nov;21(11):1569-1570 [FREE Full text] [doi: 10.1016/j.jamda.2020.09.007] [Medline: 33036911]
37. Tyack C, Camic PM, Heron MJ, Hulbert S. Viewing art on a tablet computer: a well-being intervention for people with dementia and their caregivers. J Appl Gerontol 2017 Jul;36(7):864-894. [doi: 10.1177/0733464815617287] [Medline: 26675353]
38. Leuty V, Boger J, Young L, Hoey J, Mihailidis A. Engaging older adults with dementia in creative occupations using artificially intelligent assistive technology. Assist Technol 2013;25(2):72-79. [doi: 10.1080/10400435.2012.715113] [Medline: 23923689]
39. Foley KL, Sudha S, Sloane PD, Gold DT. Staff perceptions of successful management of severe behavioral problems in dementia special care units. Dementia 2003 Feb 01;2(1):105-124. [doi: 10.1177/1471301203002001998]
40. Kuske B, Hanns S, Luck T, Angermeyer MC, Behrens J, Riedel-Heller SG. Nursing home staff training in dementia care: a systematic review of evaluated programs. Int Psychogeriatr 2007 Oct;19(5):818-841 [FREE Full text] [doi: 10.1017/S1041610206004352] [Medline: 17052376]
41. Hui EK, Tischler V, Wong GH, Lau WY, Spector A. Systematic review of the current psychosocial interventions for people with moderate to severe dementia. Int J Geriatr Psychiatry 2021 Sep;36(9):1313-1329. [doi: 10.1002/gps.5554] [Medline: 34350626]
42. Astell AJ, Ellis MP, Alm N, Dye R, Gowans G. Stimulating people with dementia to reminisce using personal and generic photographs. Int J Comput Healthc 2010;1(2):177. [doi: 10.1504/ijcih.2010.037461]
43. Lee AC, Tang SW, Yu GK, Cheung RT. The smiley as a simple screening tool for depression after stroke: a preliminary study. Int J Nurs Stud 2008 Jul;45(7):1081-1089. [doi: 10.1016/j.ijnurstu.2007.05.008] [Medline: 17707824]
44. Vernooij-Dassen M, Vasse E, Zuidema S, Cohen-Mansfield J, Moyle W. Psychosocial interventions for dementia patients in long-term care. Int Psychogeriatr 2010 Nov;22(7):1121-1128 [FREE Full text] [doi: 10.1017/S1041610210001365] [Medline: 20813074]
45. Norman D. Emotion and design: attractive things work better. interactions 2002 Jul;9(4):36-42. [doi: 10.1145/543434.543435]
46. Gowans G, Dye R, Alm N, Vaughan P, Astell A, Ellis M. Designing the interface between dementia patients, caregivers and computer-based intervention. Des J 2015 Apr 28;10(1):12-23. [doi: 10.2752/146069207789318018]
47. de Vugt M, Dröes RM. Social health in dementia. Towards a positive dementia discourse. Aging Ment Health 2017 Jan;21(1):1-3. [doi: 10.1080/13607863.2016.1262822] [Medline: 28098498]
48. Aylward S, Stolee P, Keat N, Johncox V. Effectiveness of continuing education in long-term care: a literature review. Gerontologist 2003 Apr;43(2):259-271. [doi: 10.1093/geront/43.2.259] [Medline: 12677083]

XSL•FO
RenderX

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
GDS: Global Deterioration Scale
GRAMMS: Good Reporting of a Mixed Methods Study
MRC: Medical Research Council
RCT: randomized controlled trial
RQ: research question
SFAS: Smiley Face Assessment Scale

Edited by A Mavragani; submitted 21.01.24; peer-reviewed by G Kohl, S Fox; comments to author 02.12.24; revised version received 23.01.25; accepted 02.03.25; published 16.04.25.

Please cite as:
Tan JRO, Neal DP, Vilmen M, Boersma P, Ettema TP, Gobbens RJJ, Sikkes SAM, Dröes RM
A Digital Photo Activity Intervention for Nursing Home Residents With Dementia and Their Carers: Mixed Methods Process Evaluation
JMIR Form Res 2025;9:e56586
URL: https://formative.jmir.org/2025/1/e56586
doi:10.2196/56586
PMID:

©Josephine Rose Orejana Tan, David P Neal, Maria Vilmen, Petra Boersma, Teake P Ettema, Robbert J J Gobbens, Sietske A M Sikkes, Rose-Marie Dröes. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Creating a Usable and Effective Digital Intervention to Support Men to Test for HIV and Link to Care in A Resource-Constrained Setting: Iterative Design Based on A Person-Based Approach and Human Computer Interaction Methods

Anya Zeitlin[1*], MSc; Thulile Mathenjwa[2,3*], MA; Thembelihle Zuma[2,4], PhD; Sally Wyke[5], PhD; Philippa Matthews[2], BMBS; Nuala McGrath[2,4,6], PhD; Janet Seeley[2,7], PhD; Maryam Shahmanesh[2,8], PhD; Ann Blandford[1], PhD

[1]UCL Interaction Centre, University College London, London, United Kingdom

[2]Africa Health Research Institute, Durban, South Africa

[3]School of Public Health, University of Cape Town, Cape Town, South Africa

[4]School of Nursing and Public Health, University of KwaZulu Natal, Durban, South Africa

[5]School of Social and Political Sciences, University of Glasgow, Glasgow, United Kingdom

[6]Faculty of Medicine and Faculty of Social Sciences, University of Southampton, Southampton, United Kingdom

[7]London School of Hygiene & Tropical Medicine, London, United Kingdom

[8]Institute for Global Health, University College London, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Thulile Mathenjwa, MA
Africa Health Research Institute
Krith Building, 3rd floor
719 Umbilo Rd
Durban, 4001
South Africa
Phone: 27 315210455
Email: thulile.mathenjwa@ahri.org

Abstract

Background: It is challenging to design usable and effective digital health interventions (DHIs). The person-based approach (PBA) has been proposed to incorporate users' perspectives for the design of DHIs. However, it does not explicitly describe the iterative stages of design and evaluation that are essential in moving from early planning to deployment. For this, we draw on methods from human computer interaction (HCI) that have been developed for various situations.

Objective: This study aimed to reflect on the adaptation and synthesis of PBA and HCI approaches to developing DHIs. We present a case study applying both approaches to develop Empowering People through Informed Choices for HIV (EPIC-HIV1), a DHI designed for men living in rural KwaZulu-Natal, South Africa, intended to support them in making an informed choice about whether to take an HIV test and, if necessary, engage in care.

Methods: We conducted a retrospective analysis of the documentation generated during the development of EPIC-HIV1 including findings about requirements, design representations, and the results of iterative rounds of testing. We developed an account of the process, the outcomes, and the strengths and limitations of the design and evaluation techniques applied. We also present the design of EPIC-HIV1 and summarize considerations when designing for hard-to-reach people in such settings.

Results: The PBA was applied to deliver a first prototype. This helped identify key messages to convey and how to manage issues such as user privacy, but the resulting prototype was judged by the team not to be engaging for potential users, and it was unclear whether the design was inclusive of people with low digital or health literacy. We therefore introduced methods from HCI to iteratively test and refine the app. Working with local community representatives, we conducted four refinement cycles with 29 participants, adapting and retesting the app until no further changes were needed. Key changes included making it clearer what the consequences of selecting options in the app were and changing wording to minimize misconceptions (eg, that the app would test for HIV) while addressing common concerns about testing and emphasizing long-term benefits of engaging with care, if needed.

XSL•FO
RenderX

Conclusions: Techniques for developing DHIs need to be situationally appropriate. The PBA enabled us to establish both empirical data and theory to design the content of EPIC-HIV1, but it did not directly inform interaction design to make the app usable and effective for the intended users; HCI techniques tailored to the setting enabled us to refine the app to be easy for men with little familiarity with digital technologies to use within the constraints of the setting. Iterative testing ensured the app was easy to use and that the intended clinical messages were communicated effectively.

(JMIR Form Res 2025;9:e65185) doi:10.2196/65185

KEYWORDS

person-based approach; HIV; resource-constrained settings; digital intervention; user-centered design; behavior change techniques; digital health

Introduction

Background

The focus of this paper is on the application of methods for developing usable, effective, and engaging interactive digital health interventions (DHIs), taking as a case study an intervention, Empowering People through Informed Choices for HIV (EPIC-HIV1), intended to enable men to make an informed choice about testing for HIV and, if necessary, engaging with care. Recent literature shows that men are left behind in the response to HIV [1] as evidenced by the low uptake of HIV testing, prevention and treatment compared with women [2-4].

Challenges of Developing Interactive DHIs

As digital technologies become more affordable, their potential is increasingly leveraged for DHIs in a broad range of settings including both affluent and resource-constrained communities, in urban, rural and remote settings [5]. There are numerous challenges with implementing DHIs such as ensuring interventions keep pace with rapidly changing technical systems [6], supporting usability needs [7], ensuring that the intervention is accessible across devices and operating systems [8], and making it culturally appropriate [9]. These challenges are exacerbated in resource-constrained settings where there are variable infrastructural constraints and levels of familiarity with technology within intervention cohorts.

There are few accounts of how to develop usable, engaging, and effective DHIs. One of the most widely cited is the person-based approach (PBA) to intervention development [10]. Although this approach describes the need to iteratively test prototypes with end users, it omits detailed discussion of interaction design and how to integrate this into the intervention development process. Conversely, the human computer interaction (HCI) literature rarely addresses issues that matter to population health researchers such as content design, clinical effectiveness, scalability, and sustainability [11]. The aim of the work reported here was to reflect on the integration of PBA and HCI methods taking as a case study the design and deployment of EPIC-HIV1, a tablet-based app.

Apps are particularly suited for stigmatized health conditions like HIV because they provide users with privacy and anonymity [12,13]. In addition, they offer health interventionists the ability to deliver uniform messages, free from embarrassment which affects people's ability to communicate consistent and accurate information [13]. Several studies have used digital technology to increase uptake of HIV testing and medication adherence [14,15]. These studies found that both mobile and tablet-based HIV interventions are feasible, acceptable, and effective methods to engage hard to reach populations [16]. However, the studies do not discuss how the interventions were designed or developed.

Interdisciplinary Approaches to Developing Digital Health Interventions

The World Health Organization advocates the application of user-centered design to develop DHIs to ensure that they are effective, accessible, acceptable, and user-friendly [17]. This often requires the collaboration of professionals from health and HCI backgrounds. Health care professionals, including epidemiologists and social scientists, contribute clinical expertise, theoretical understandings of behavior change and health outcome evaluation techniques. HCI professionals contribute user-centered design methodologies, which are applied to understand user needs, inform the design of digital technologies and evaluate them in terms of usability and user experience [11,18,19].

Despite the clear benefits of interdisciplinary collaboration, there are well-recognized challenges [11,18] relating to the academic heritage of the respective disciplines. Professionals from health care backgrounds put greater emphasis on measuring the effectiveness of DHIs and evaluating outcomes. HCI professionals focus on iterative design, including alternative design representations and formative evaluation, to ensure digital technology is usable and acceptable. As a result, interdisciplinary teams experience divergent approaches to timelines, measuring effectiveness and understanding success.

Literature in the field of behavior change tends to limit the integration of HCI methods, and therefore collaboration, to discrete stages in the intervention development process: for example, evaluating usability and engagement once a prototype has been developed [20].

In HCI literature, there is more interest in exploring the boundaries of the role HCI should play in DHI design and evaluation, rather than identifying discrete stages in the development process in which HCI approaches can be applied. Klasnja et al [21] describe the contribution of HCI as being to understand "how and why" DHIs work, rather than measuring outcomes. In contrast, Smith et al [22] advocate using a value chain analysis to evaluate how short-term (proximal) metrics relate to long term (distal) behavior change and health outcomes.

XSL•FO
RenderX

Marcu et al [23] describe the Patient-Clinician-Designer framework, which seeks to integrate multiple perspectives throughout the design process. In understanding the clinical context, patient needs and technical constraints, they ensure that digital interventions meet the needs of all stakeholders and are therefore more likely to be successful when implemented [23]. They describe this in reference to a mental health application, but it has also been applied to the development of applications for stigmatized conditions including HIV [24]. Similarly, Blandford [11] highlights the value of applying HCI methodologies throughout intervention development to facilitate more systematic trade-offs between what is needed for users and clinicians and what is feasible in the context.

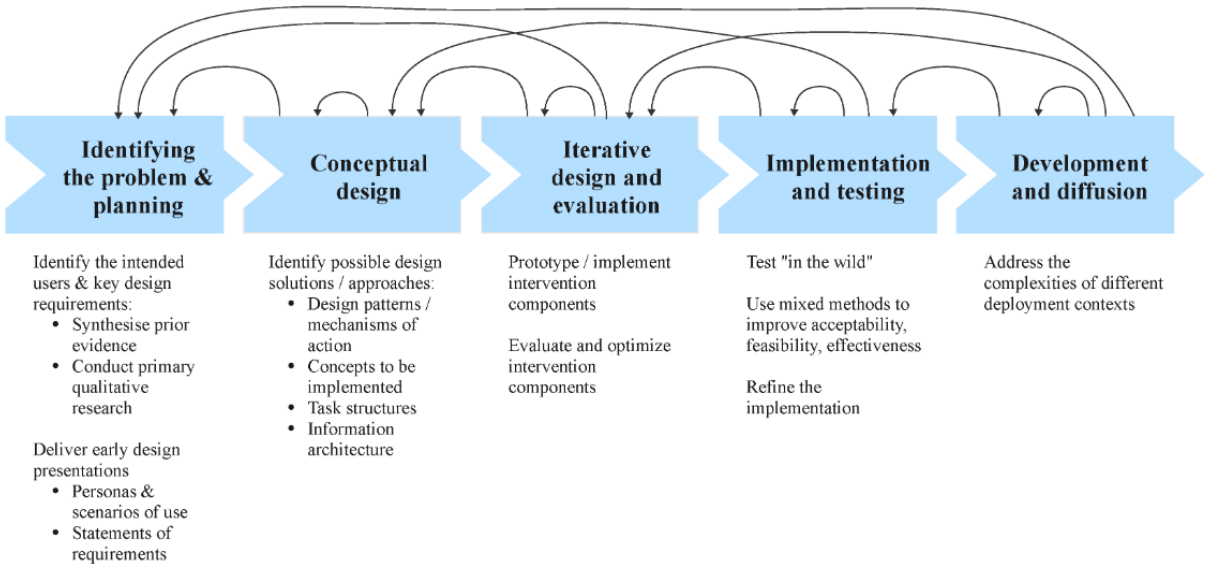
Person-Based Approach and Integration With Human Computer Interaction Methods

We used the PBA to inform the design and development of EPIC-HIV1. The PBA [10,20] provides guidelines on how to use qualitative research methods to identify psychosocial factors that influence the effectiveness and acceptability of interventions. Using rich, qualitative research at every stage of the intervention development process facilitates an investigation of the beliefs, attitudes, needs and situations of participants in the intervention. Yardley et al [10] propose that using PBA can make an intervention more relevant, persuasive, accessible and engaging.

The PBA consists of 4 stages, namely planning, design, development, and intervention [20]. While Yardley et al [20] recognize the value of gathering user input, the focus is on how that input informs the development of content and intervention goals. They do not describe the HCI, or user-centered design, methods that should be applied or how to translate the intervention goals and content into a usable and engaging digital product.

Blandford [25] proposes an integration of PBA with established HCI design methods to address these issues (Figure 1). The key extensions to PBA are earlier consideration of design possibilities, the identification of design representations (eg, design patterns and task structures) and iteration between as well as within stages.

Figure 1. An integrated development lifecycle illustrating how a person-based approach to digital health intervention design can be augmented by human computer interaction methods.



Both approaches advocate the use of personas and scenarios. These are rich descriptions of the intended users of the intervention and of how they will use the intervention. Scenarios are commonly presented at different levels of abstraction; for example, Rosson and Carroll [26] describe problem scenarios (the broad situation), information scenarios (giving details of the information that will be provided to users) and interaction scenarios (describing the details of user interaction with the system).

The aim of this study is to report and reflect on the process and outcomes of applying the PBA and selected HCI methods to the development of EPIC-HIV1 as an exemplar DHI for behavior change.

Methods

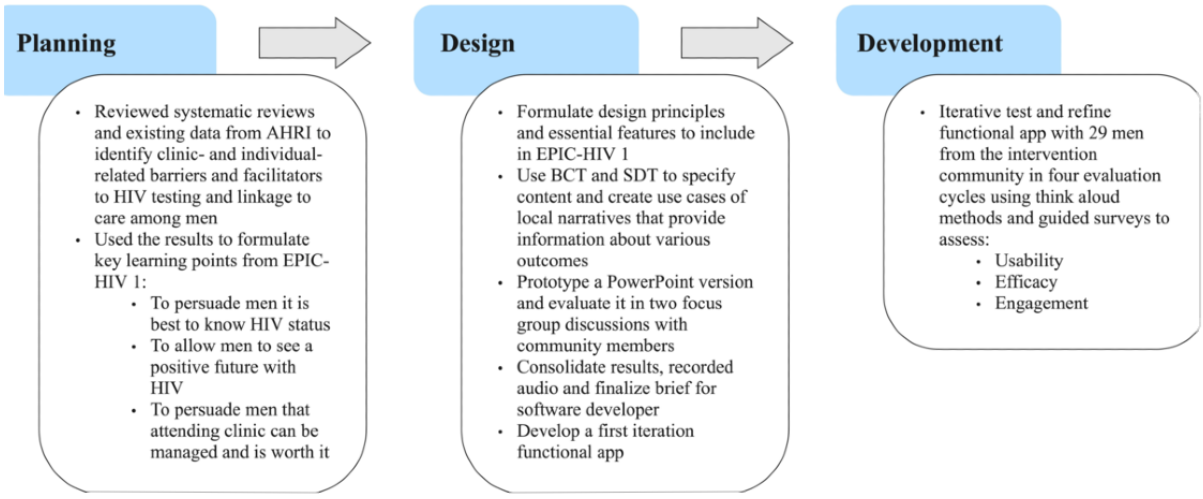
Overview

This paper draws on documentation that was created during the iterative development process, including the initial prototyping of EPIC-HIV1 and the subsequent cycles of refinement and testing. That documentation was systematically gathered throughout the development process to support subsequent analysis. The retrospective analysis of the documentation involved the construction of a narrative of the processes and outcomes for the planning, design, and developing stages outlined in Figure 2 and presented below (after an overview of the context and rationale). Therefore, this narrative is presented as “results” below, rather than “method”. The implementation stage (ie, deployment in a clinical trial) is beyond the scope of

XSL•FO
RenderX

this paper and is described elsewhere [4,27,28]. We present the context, rationale, and design aims for EPIC-HIV1.

Figure 2. Methods used in the first 3 phases of the person-based approach, phase 4 (implementation) is out of scope for this paper. AHRI: Africa Health Research Institute; BCT: behavior change taxonomy; EPIC-HIV1: Empowering People through Informed Choices for HIV; SDT: self-determination theory.



Empowering People Through Informed Choices for HIV Context and Rationale

EPIC-HIV1 is 1 of 2 DHIs that make up the EPIC-HIV intervention. The second DHI, EPIC-HIV2, was designed to support engagement in care for those who did not link to care within a month after a positive HIV diagnosis and is described elsewhere [29]. EPIC-HIV is 1 of 2 interventions in the Home-based Intervention to Test and Start (HITS) trial that sought to increase HIV testing and linkage to care among men in rural South Africa using micro incentives and male-targeted decision support [4,30]. The HITS trial aimed to compare the impact of micro incentives (financial) with decision support (EPIC-HIV). It was implemented in a rural setting in uMkhanyakude district in northern KwaZulu-Natal, South Africa using the Africa Health Research Institute (AHRI) Demographic and HIV surveillance platform [31]. The trial results are reported elsewhere [4,27,28].

For this case study, we focus on the development of EPIC-HIV1. As noted above, it is an educational app that introduces men to accurate information about HIV and encourages them to make an informed choice about engaging with HIV testing and care. The design drew on self-determination theory (SDT) [32] to plan content by supporting three basic psychological needs (autonomy, competence and relatedness) and increase men's internal motivation to test for HIV and link to care where necessary. We also drew on the behavior change taxonomy (BCTv1) proposed by Michie et al [33]. They list 93 techniques that can be applied in behavior change interventions, clustered into 16 groups (eg, goals and planning, feedback and monitoring, and social support, providing a credible source of information and advice). Many of these are potentially relevant to the design of digital interventions: for example, goal setting, providing instructions on how to perform a behavior, or providing a credible source.

EPIC-HIV1 was designed to be delivered by fieldworkers during the annual household visits which includes conducting household and individual behavior and health related surveys and collecting dry blood spots for anonymized HIV testing as well as the offer of a rapid HIV test [31]. The concept was that the app would be deployed on the fieldworker tablet and offered to men during the data collection for individual surveys where the fieldworker is alone with the participant. The app is offered before the HIV rapid test offer. The fieldwork visits are carried out in a systematic cycle where they are allocated a number of households to cover in a week. Due to this time constraint, user interaction with EPIC-HIV1 was allocated 10 minutes.

Design Aims for Empowering People Through Informed Choices for HIV

Two key aims for EPIC-HIV1 were ensuring that it met users' needs, that is, that participants were engaged by the content and could autonomously navigate the application, including people with low digital and health literacy; and ensuring that it met the intervention goals, namely to communicate the correct clinical messages to men and enable them to make an informed choice about engaging with HIV care.

The design had to recognize the stigma related to HIV [15,16,34]. This relates to both the design of the intervention itself, and the evaluation methods used. For example, in user studies we did not require participants to disclose their HIV status. Also, to enable study participants and future users to explore EPIC-HIV1 independently and privately, earphones were offered for their use.

Another design challenge in resource-constrained settings is low literacy [35]. Solutions proposed for low-literacy users include graphical interfaces, interactive voice interfaces, and mediated usage [36]. Designing for low literacy populations extends beyond catering for users not being able to read, to multiple cognitive and emotional factors that influence how users engage with content [37,38]. For health care interventions, there is added complexity: individuals with lower literacy tend to have lower health literacy [39], and having lower health literacy puts an individual at higher clinical risk and is associated

XSL•FO
RenderX

with reduced ability to exert control over health through informed decision making [39,40]. The approaches taken in both design and testing to be inclusive of users with low digital and health literacy are discussed below.

Ethical Considerations

This study received ethical approval from the University of KwaZulu-Natal Biomedical Research Ethics Committee (BFC398/16). Participants involved in empirical data collection provided written or verbally informed consent before participation. Participation in the study was entirely voluntary. To protect participants' privacy and confidentiality all data were deidentified, and no personally identifiable information was included in the final dataset. Participants received light refreshments (sandwich and juice) for taking part in the study and no financial rewards were provided.

Results

Overview

As noted above, EPIC-HIV1 was developed following the first 3 stages of PBA. The planning and design stages were led by social scientists and health professionals (JS, SW, MS, TZ, PM, and TM) and the development phase was led by HCI specialists (AB and AZ) and the technology partner. Each stage used different recruitment and sampling methods as described in detail below.

Person-Based Approach Stage 1: Planning

The planning stage is described in detail by Mathenjwa et al [29]. In summary, we reviewed systematic reviews and existing data from AHRI to identify barriers to HIV testing and linkage to care and used the results to define the user requirements of the app. In summary, key barriers to testing were fear of rejection, and being blamed for sexual misconduct, if found to be HIV positive; fear of death or illness; feeling healthy and so not understanding the need to test; a preference for traditional medicine; and a belief that medicine should be curative (rather than for preventing or managing illness). Key barriers to engaging with care were that clinics were regarded as feminized spaces where men do not feel comfortable; that they are concerned about confidentiality if they are seen in clinics; that there may be long wait times in clinics; and that they often have poor relationships with health care workers. Key learning points for EPIC-HIV1 were that it was important to persuade men that it is best to know their HIV status, because it allows them to take control; that men needed to see a positive future with HIV, in which they are in control and can achieve their ambitions, and that attending clinic can be managed and is valuable in the long term.

Person-Based Approach Stage 2: Design

The aim of this stage was to formulate the design principles of the application and essential features that should be included. Using the results from the planning stage, we identified barriers that can be modified by EPIC-HIV1 and key learning points to be incorporated into the app. We then formulated intervention features relevant to the learning points. Finally, we identified specific behavior techniques from the BCT taxonomy [33] to support autonomy, competence and relatedness. The components are interrelated as illustrated in Table 1.

Table 1. Key learning points from the planning stage of person-based approach and how they relate to self-determination theory and behavior change taxonomy.

Learning point	SDT[a] relevant	BCT[b] from taxonomy
To persuade men it is best to know your HIV status, because it allows you to take control	• Persuading users that knowing their status allows them to take control of important aspects of their lives, such as upholding traditional values, caring for a family or having sex and children (supporting autonomy).	9.1 credible source 5.1 information about health consequences 5.2 salience of consequences
To allow or engage men to see a positive future with HIV, in which they are in control and can achieve their ambitions	• Maintaining a positive illness context throughout: that it is possible to live with HIV and do all one wants to do. • Educating men about HIV and the benefits of testing and taking antiretroviral therapy (supporting autonomy).	9.1 credible source 5.1 information about health consequences 5.2 salience of consequences 13.5 identity associated with changed behavior 13.1 identification of self as role model
To persuade men that attending the clinic can be managed and is worth it in the long term	• Persuade men that they can handle going to the clinic through positive examples of men talking about what it was like and how they managed it (supporting relatedness, competency and autonomy).	13.5 identity associated with changed behavior 13.1 identification of self as role model

[a]SDT: self-determination theory.

[b]BCT: behavior change taxonomy.

The Conceptual Design and Scenarios of Use

The high-level design concept was to provide experiential information from local men living with HIV that users could identify with. The content was rooted in local narratives to provide information about various outcomes to increase risk perception, salience, and likelihood of response by making the decision to test or link to care explicit. Personas were implicitly encapsulated in the descriptions of the characters who were to be introduced through EPIC-HIV1: men of different ages and with different health beliefs and levels of digital literacy.

XSL•FO
RenderX

As described above, the envisaged context of use for EPIC-HIV1 was that it would be administered by the AHRI fieldworker during the annual individual biobehavioral surveillance home visit. When the time came to explore EPIC-HIV1, the fieldworker was expected to hand over the tablet computer with the app opened (start screen) to the user (consenting man), together with earphones (that the man could keep) and wait while they interacted with the app in private. The app was designed to take 5-10 minutes depending on the pathway that the user selected to align with the time constraints of the fieldwork.

The information scenario was based on the notion that the app should acknowledge and refute common misconceptions about HIV (how it is transmitted and what treatments are effective) and concerns about attending a clinic and making other people aware of their HIV status.

The interaction scenario involved the user working through 3 sections (illustrated in Multimedia Appendix 1. First, there was a short introduction, after which the user was given the choice to either find out about HIV testing and treatment (if they were not ready to test) or just find out about treatment (if they were ready to test). In the testing section, users were given the option to listen to short vignettes describing common reasons for men not wanting to test for HIV. Users were then presented with information, addressing each of these reasons and providing a counter argument. In the treatment section, 3 characters described their journey since being diagnosed with HIV; these characters represented different personas, and the user could listen to one of their stories (ideally the one they found most relatable).

All app content and instructions were provided in text and audio format. Still images, including photographs of individuals (with faces obscured) and artefacts from the intervention community, supported the audio descriptions.

The app used a male nurse from AHRI to serve as a guide character and credible source of HIV information. He gives an introduction and explains how the app works and then reads out the 2 options (“I am sure I will have an HIV test today” and “I might not have an HIV test today”) and asks the user to select which statement he agrees with. In the “ready to test” section, the nurse applauds the user for taking the decision to test today. In the “not ready to test” section, we used 7 characters to give short vignettes discussing common reasons for men not wanting to test for HIV which were identified during the planning stage. The nurse comes back to address the reason each character lists, providing a counter argument. Users can then go into more detail by selecting one of the 3 main characters that persuade men that knowing their HIV status gives them control of important aspects of life, that it is possible to live with HIV and still achieve goals and finally that they can manage going to the clinic (Figure 3).

Figure 3. Illustration of 7 characters from the PowerPoint prototype of Empowering People through Informed Choices for HIV.

Testing the Conceptual Design

We developed a low fidelity PowerPoint (Microsoft Corp) protype of the app and evaluated it with 2 focus groups of men from the community advisory board, a body representing members of the intervention community that acts as a bridge between AHRI and the community, safeguarding the rights of the study participants. We sought to include younger and older community advisory board members. The focus of the evaluation was on whether the characters and messages selected were acceptable to men in our community. Since we aimed to assess acceptability of the content and characters, findings were categorized into content or character related and the app functionality (Table 2).

XSL•FO
RenderX

Table 2. Summary of the focus group discussion results and how we addressed them.

Main issues	Issues	Steps taken and revision of content
Content	• Self-stigmatization after testing HIV-positive. • Fear of weakness because of illness. • Manhood is about becoming isoka (Casanova). • Testing is for everyone, not for particular groups or types. • Most men want to wait until they get sick before testing, long queues in clinics, feminized spaces of health care.	• Positive reinforcement on autonomy, competence and relatedness among all characters.
Characters	• Groups preferred characters to look like strong men. • Groups commented on what characters should wear. • Groups mentioned that characters' gestures should be open and clear.	• Only the nurse character will be seen by users. • No other faces will be shown, real settings will not be used. • Men wearing what was proposed in groups. • Characters' gestures open and clear.
Functionality	• Asked for a rewind button so that people can go back and listen. • Privacy and confidentiality.	• No rewind option because of time. • App will be private, and fieldworker will clarify to each participant that they will not know how the participant navigated the app.

We used the results to refine the content and characters and finalized the brief for the software developer. We selected a cast of actors from the local community to represent the different characters and did an audio recording of voice overs and a photo shoot at AHRI offices and a local community center. This was sent to the software developer with the brief. We received the first iteration of a functional app, as illustrated in Figure 4; however, it was not considered by the team to be engaging.

Figure 4. First iteration of the Empowering People through Informed Choices for HIV app cycle (3 example screen shots).

Screen 1 - Introduction: Nurse introduces the app, users receive instructions and choose a route through the app
Screen 2 - Testing Section: User hears common reasons why men do not want to test for HIV and nurse's rebuttal
Screen 3 - Treatment Section: User hears stories of men since their HIV diagnosis

Person-Based Approach Stage 3: Iterative Testing and Design Changes

We iteratively tested and refined EPIC-HIV1 to ensure that the app was interactive, usable and acceptable for a broad spectrum of users with varying education and digital literacy. In total, 4 cycles of user-centered evaluation and design were conducted, based on design sprints each lasting 2 weeks. These cycles were structured using a simple interaction design model [41], whereby initial user requirements were identified, design improvements were generated, and the new design was evaluated. The evaluation metrics are summarized in Table 3.

XSL•FO
RenderX

Table 3. Summary of evaluation metrics relating to usability, efficacy, and engagement.

Metric	Questions	Method
Usability	• Is it clear how to move between screens? • Is it clear how to engage with the information on each screen (ie, choose options, listen to audio and exit audio)?	Observations during think-aloud sessions
Efficacy	• Are they key messages communicated? • Do participants gain new information because of using the app? • Is the information they gain factually accurate?	Questions in think aloud protocol and guided survey
Engagement	• Is the interaction pleasant? • Is the presentation of content appropriate and interesting?	Questions in think aloud protocol and guided survey

During each cycle, participants were asked to complete defined tasks and articulate their thoughts about EPIC-HIV1 and the task. Both actions and verbalizations were recorded to identify the main usability challenges and build an understanding of the acceptability of content. Participants were paused after the introductory screens and asked questions regarding their initial expectations of the app's content and purpose. After using the app, participants were asked further questions of their understanding of the content and reflections on the design. This guided, verbally administered, survey ensured that we did not exclude low-literacy participants. It consisted of a mix of open and closed questions and was used to gather insight into usability, comprehension, and design preferences. The evaluations were conducted in isiZulu, with a bilingual (isiZulu-English) researcher administering procedures and participants were observed as they used the application.

Recruitment and Participants

All participants were recruited from the AHRI demographic and HIV surveillance community. Participants were approached in public spaces by a local male researcher from the community. These public spaces varied but included parks, taverns, and the roadside. Across the 4 evaluation cycles, sampling became increasingly purposive, focused on more rural areas and older participants, as we sought to recruit participants who represented specific user groups within the intervention cohort, especially those who had been identified as likely to have low technological literacy.

We conducted evaluations with 29 unique isiZulu speaking males (excluding pilot participants) over the 4 iterative cycles. All participants were asked their age to confirm they were aged 18 years or older and able to consent to participate. Recruitment was helped significantly by the presence of local researchers who knew where men in the community "hang out" and were able to explain the purpose of the study. Men were often in groups and many declined to engage, meaning all 29 participants engaged willingly.

Overview of Evaluation Cycles

Across the 4 cycles, the evaluations became more refined, but the objectives and the metrics were consistent. Correspondingly, there was consistency across the protocols. For example, some of the questions (Table 3) were the same and were asked at the same point in the evaluation. The key features of each evaluation cycle are summarized in Table 4 and refer to Multimedia Appendix 2 for example of findings and changes made in each evaluation cycle.

XSL•FO
RenderX

Table 4. Key features of each evaluation cycle.

Evaluation cycle	Objectives	Methods
Evaluation 1	• Primary: identify main usability challenges, gather understanding of target user group's needs, and assess understanding of content. • Secondary: gather insight into potential design improvements.	• Think aloud. • Observation. • Guided survey.
Evaluation 2	• Primary: make layout easy to use for people who are novice tablet users and have low literacy, assess understanding of content. • Secondary: gather insight into preferred look and feel of app.	• Observation. • Retrospective usability questions. • Guided survey.
Evaluation 3	• Primary: test whether users know how to select options, assess understanding of content, make content more engaging. • Secondary: ensure that uninterrupted participants were completing the app in an appropriate time.	• Observation. • Retrospective usability questions. • Guided survey. • Task completion times.
Evaluation 4	• Primary: assess whether the changes made to the application affected users' expectations, such that they were making more informed decisions.	• Observation. • Retrospective usability questions. • Guided survey.

Specifics of Each Evaluation Cycle

For the first evaluation, convenience sampling was used to recruit from the community, working in or close to the research institute. The evaluation took place within the research institute. After interacting with the first version of the app, participants were presented with low fidelity interactive prototypes of different sections of the app and asked which they preferred, which was most relatable, and which was most clear.

In Evaluation 2, we used pop-up research [42], which involves carrying out targeted user evaluations in environments that are familiar and used by the app's intended audience, to simulate some of the environmental constraints and social considerations that would be experienced by fieldworkers when administering the DHI, such as lack of access to the internet and difficulty finding private spaces. This had the potential to reduce participant response biases [37] by making the environment more familiar and less formal than the research institute. Participants were approached in public spaces, such as at bus stops or tuck shops, and evaluations were conducted either in those spaces or close by in the researchers' car (illustrated in Figure 5). These decisions helped to establish a level of mutual trust and collaboration between the researchers and participants. Unlike most accounts of pop-up research, the planned time for each individual's participation was not particularly short: it lasted 20-30 minutes, sometimes more if the participant had low literacy.

Figure 5. Illustrative photos of pop-up evaluation testing in remote locations using an Africa Health Research Institute vehicle.

Based on learnings from Evaluation 1, we administered retrospective usability questions rather than asking participants to articulate their thoughts concurrently. We also printed screenshots of design variations on paper and described the differences to participants then asked them which they preferred, rather than continuing to use low-fidelity interactive prototypes

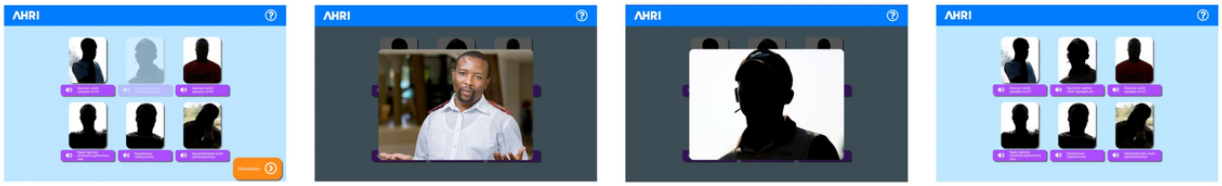
XSL•FO
RenderX

after the first part of each session. We included closed questions that were asked while participants were interacting with EPIC-HIV1: for example, we added a help button; in the evaluation, we stopped users after the instruction screen and asked them to indicate which button to press if it was not clear what to do next. Most participants were unaware they could skip content or be selective, which indicated that the app was not supporting autonomous decision making. When presented with design alternatives, all 6 participants selected the interface with photo images because it was "clear." This supported the use of rich and contextually relevant images throughout the app to optimize the relatability of the content.

Following Evaluation 2, previews were either automated or removed, which meant users would only be presented with choices when it was critical. We added extra instructions on how to select options, made button states more distinct and provided additional nudges in the form of button animations and instructions from the nurse character. Additional photographs of paid professionals who were representative of the intended cohort were added to the character stories to make them more engaging (Figure 6).

Figure 6. Example of design changes made, illustrated with screenshots from the first and third versions of Empowering People through Informed Choices for HIV.

Evaluation 1: User previews character, user selects character again to hear reason for not testing, user selects next, user selects the character again to receive nurse's response to the character's reason (4 clicks).



Evaluation 3: User selects character, hears character's reason for not testing, then the nurse's rebuttal (one click).

In Evaluation 3, as well as assessing whether participants knew how to select options, understood correct messages and found the content engaging, we wanted to ensure that participants were completing the app within a given time. We split participants into 2 groups: one was interrupted to assess their option selection while the other was uninterrupted, to gather their task completion time and observe the full user flow. Having identified rural and older people as those who were least comfortable using the app, we actively recruited males representing this category, reducing the proportion of young periurban males (the easiest to find and recruit). After conducting evaluations with 12 participants from various locations, we had reached saturation in responses. Compared with earlier evaluation cycles, we focused more on option selection and expectations. For example, at the beginning participants were asked what they had expected to find if they had selected the other option so that we could assess whether participants were actively choosing what they wanted to hear about. This was particularly important because users could only engage with EPIC-HIV1 once as part of the planned intervention and could not return to content of interest.

Evaluation 4 did not reveal any additional usability issues. Participants were able to describe what they would have had if they had selected the other option, indicating the changes to the language used at the beginning of the app were effective and supported informed decision making.

Summary

Both the evaluation objectives and the methods became more refined during the process. For example, the objective of the first evaluation was to get a sense of what users were finding difficult and what evaluation style would be contextually appropriate. Convenience sampling was used to recruit participants and the questions in the evaluation were relatively general. In the third evaluation, 2 of the objectives were to assess whether participants could make accurate option selections and complete the app within a given time. Purposive sampling was used to recruit participants who represented a range of users we had previously encountered, especially those who had had usability challenges (ie, low literacy, novice users, and older people). The evaluation was more focused, questions were more specific, and the methods were altered to support our objectives.

Although pop-up research was an effective and valuable means of recruiting participants and conducting evaluations, we had to ensure that we adequately addressed considerations, including ethics (ensuring the participants did not feel coerced), privacy (ensuring that participants could freely engage with content without passers-by hearing), community-institution relationships

XSL•FO
RenderX

(avoiding contributing to "participant fatigue" near the research center), and our own safety as we went into remote places and invited people into our car or worked with them in unfamiliar environments.

Discussion

Principal Findings

The analysis of existing theory and data enabled the team to identify and address the most important barriers to men engaging in HIV testing and care in a cost-effective way. The use of HCI methodologies enabled the team to translate the intervention goals, that is, communicating persuasive messages in a way that would promote behavior change, into design strategies and ensure the application was usable. The resulting application was inclusive of the intended population, supported meaningful choice in how users engaged with the application, and ensured the correct clinical messages were communicated.

One key lesson from this study is that it is essential to design the introduction (or "onboarding") to a DHI very carefully, to ensure that users are empowered to use the app effectively and that expectations of what they will experience are set appropriately. Another key lesson is that an appropriate balance had to be achieved between autonomy (enabling people to navigate as they choose) and ensuring that the correct clinical messages are communicated.

The conceptual design featured a series of dialogues between relatable characters and a trustworthy nurse. These dialogues were accompanied by images of the characters in familiar settings. However, this approach was never tested against alternative design strategies. Early decisions about the design and mode of the interaction were based on informal understanding and intuition rather than evidence. Resource limitations (both time and funds) meant that we did not have the capacity to revisit earlier design decisions such as the appropriateness of using a tablet computer or the general approach of presenting dialogues. Thus, it was not possible to engage in the broader cycles of iteration proposed by Blandford [25].

Early design decisions, such as the approach of focusing on a dialogue between relatable characters and a trustworthy nurse, and about the content of the dialogue, were based on the key scenario of use: that is, that the tablet computer would be lent to the individual for up to 10 minutes within the annual health visit; this imposed a significant constraint on the design. For example, the main interaction design issue raised by members of the community advisory board in the Stage 2 testing, that people would like a "rewind" option, was not addressed due to this time constraint.

Through iterative testing the main area of concern, and where many changes were made, were in the introduction, or "onboarding," for the app. It was found necessary to set expectations about the app (eg, that it would not deliver an HIV test result) and to provide detailed verbal instructions about how to select options and press buttons on a touch screen.

We needed to balance the needs for the app to be engaging and inclusive. The participants were diverse, and levels of technical, educational and health literacy varied widely. Some participants were comfortable using the tablet app and owned touch screen smartphones. Testing highlighted the tension between providing literal choice in interaction design and meaningful choice regarding how the user engaged with the intervention. For example, giving users a choice about whether to listen to the nurse's response after hearing a character's concerns proved counterproductive as it did not support the concept of "informed choice," so the choice point between concern and response was removed after the first evaluation cycle.

Changes to the design were intended to support low-literacy users: for example, by adding audio instructions, highlighting one option at a time, and reducing the amount of user input required. However, we maintained critical choice and ensured it was well understood. The reduction in user input was balanced with ensuring that the app remained engaging and persuasive for participants with higher literacy levels. Through iterative cycles, we were able to deliver a design that was inclusive of as much of the target cohort as possible.

Nevertheless, we had to recognize that the delivery of the intervention was likely to exclude a small proportion of the potential beneficiaries. For these people, for whom the app was inaccessible, a conversation with the visiting health worker would be a possible alternative.

Limitations

As noted above, HCI methodologies had not been involved during the early stages of intervention development. As a result, the remit of HCI was to fix the app, to make it as good as possible given the time and resources remaining. We believe the app design could have been stronger if HCI methodologies had been used alongside PBA and BCT approaches from the outset. For example, there was no opportunity for exploratory research to define what interactivity meant for the intended population or understand whether a tablet-based app was an appropriate approach to communicating HIV/AIDS information in this context, or whether it could inadvertently bias participant responses, as indicated by previous research [43], or exclude a proportion of the intended target population.

There was also no opportunity for HCI focused summative research, for example, to understand how users had navigated through the app, areas of interest or potential drop-off points [21]. The HITS clinical trial results [4,27,28] suggest that EPIC-HIV1 did not increase men's uptake of HIV testing. This raises many questions that could be the focus of future research: for example, would a different kind of digital intervention have been more effective, was the approach of fitting the user's interaction with the intervention into the annual visit effective, are digital interventions inappropriate in this context, or does the intervention need to be better tailored toward different subpopulations (eg, those with different levels of digital literacy or different priorities in life)?

Conclusion

The aim in developing EPIC-HIV1 was to support a clinical trial investigating whether financial incentives or a digital

XSL•FO
RenderX

intervention was more effective than the established approach to encouraging men to test for and engage in care for HIV. The implicit assumption was that any digital intervention would be equally effective and no comparison between different digital interventions was planned.

By applying an iterative design approach, we were able to make systematic choices regarding design that would facilitate usability, engagement, and comprehension required for the intervention. The value and necessity of these choices were brought into sharp focus with this case study as many of the challenges of developing an DHI were exacerbated by the study context (resource constrained, low digital literacy, and stigmatized condition).

Developing Digital Health as an interdisciplinary field requires close collaboration throughout the intervention design process to ensure that the resources and expertise required are available in a timely way.

Acknowledgments

This study was funded by i-sense IRC (EPSRC grant EP/K031953/1) and the Marie Curie Actions of the European Union's seventh Framework Programme (FP7/2007-2013; under REA grant agreement n° 612216). The Home-based Intervention to Test and Start (HITS) trial was funded by the National Institute for Health (R01AI124389) and the Africa Health Research Institute (AHRI) demographic Surveillance Information system and population intervention program is funded by the Wellcome Trust (082384/Z/07/Z) and the South Africa Population Research Infrastructure Network (funded by the South African Department of Science and Technology and hosted by the South African Medical Research Council). NMcG and MS are supported by the National Institute for Health and Care Research (NIHR) Research Professorship awards (Ref:RP:2017-08-ST2-008 and NIHR 301634 respectively) using UK aid from the UK Government to support global health research. This research was funded in part, by Wellcome (Wellcome Strategic Core award: 227167/A/23/Z). For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding bodies or the UK government. We are grateful to all participants in this study. The authors wish to appreciate Velaphi Gumede for participant recruitment and data collection.

Data Availability

The datasets generated in the study are available from the HITS trial principal investigators on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of user journey through the app.
[PNG File , 1818 KB - formative_v9i1e65185_app1.png]

Multimedia Appendix 2
Example of findings and changes made in each evaluation cycle.
[DOCX File , 15 KB - formative_v9i1e65185_app2.docx]

References

1. UNAIDS. Addressing a blind spot in the response to HIV — Reaching out to men and boys. UNAIDS. 2017. URL: https://www.unaids.org/en/resources/documents/2017/blind_spot [accessed 2025-03-04]
2. Iwuji CC, Orne-Gliemann J, Larmarange J, Okesola N, Tanser F, Thiebaut R, ANRS 12249 TasP trial group. Uptake of home-based HIV testing, linkage to care, and community attitudes about ART in rural KwaZulu-natal, south africa: descriptive results from the first phase of the ANRS 12249 TasP cluster-randomised trial. PLoS Med 2016;13(8):e1002107 [FREE Full text] [doi: 10.1371/journal.pmed.1002107] [Medline: 27504637]
3. Randera-Rees S, Safari WC, Gareta D, Herbst K, Baisley K, Grant AD. Can we find the missing men in clinics? Clinic attendance by sex and HIV status in rural South Africa. Wellcome Open Res 2021;6:169. [doi: 10.12688/wellcomeopenres.16702.1]
4. Tanser FC, Kim HY, Mathenjwa T, Shahmanesh M, Seeley J, Matthews P, et al. Home-based intervention to test and start (HITS): a community-randomized controlled trial to increase HIV testing uptake among men in rural south Africa. J Int AIDS Soc 2021;24(2):e25665 [FREE Full text] [doi: 10.1002/jia2.25665] [Medline: 33586911]
5. Blandford A, Wesson J, Amalberti R, AlHazme R, Allwihan R. Opportunities and challenges for telehealth within, and beyond, a pandemic. The Lancet Global Health 2020;8(11):e1364-e1365. [doi: 10.1016/s2214-109x(20)30362-4]

XSL•FO
RenderX

6. Patrick K, Hekler E, Estrin D, Mohr D, Riper H, Crane D, et al. The pace of technologic change: implications for digital health behavior intervention research. Am J Prev Med 2016;51(5):816-824. [doi: 10.1016/j.amepre.2016.05.001] [Medline: 27745681]
7. Cresswell KM, Blandford A, Sheikh A. Drawing on human factors engineering to evaluate the effectiveness of health information technology. J R Soc Med 2017;110(8):309-315 [FREE Full text] [doi: 10.1177/0141076817712252] [Medline: 28537105]
8. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
9. Viljoen A, Klinker K, Wiesche M, Uebernickel F, Krcmar H. Design principles for mHealth application development in rural parts of developing countries: the case of noncommunicable diseases in kenya. IEEE Trans. Eng. Manage 2023;70(3):894-911. [doi: 10.1109/tem.2021.3072601]
10. Yardley L, Ainsworth B, Arden-Close E, Muller I. The person-based approach to enhancing the acceptability and feasibility of interventions. Pilot Feasibility Stud 2015;1:37 [FREE Full text] [doi: 10.1186/s40814-015-0033-z] [Medline: 27965815]
11. Blandford A. Lessons from working with researchers and practitioners in healthcare. interactions 2018;26(1):72-75. [doi: 10.1145/3292023]
12. Andalibi N, Haimson OL, De Choudhury M, Forte A. Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. 2016 Presented at: CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; 2016 May 07; San Jose California USA p. 3906-3918. [doi: 10.1145/2858036.2858096]
13. Sorcar P, Strauber B, Loyalka P, Kumar N, Goldman S. Sidestepping the elephant in the classroom: using culturally localized technology to teach around taboos. Sidestepping the elephant in the classroom: Using culturally localized technology to teach around taboos. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems; 2017 Presented at: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems; 2017 May 02; Denver Colorado USA p. 2792-2804. [doi: 10.1145/3025453.3025958]
14. Aronson ID, Bennett AS, Guarino H, Marsch LA, Bania TC. Mobile interventions to increase HIV testing among underserved: high-risk populations. 2016 Presented at: MMHealth '16: Proceedings of the 2016 ACM Workshop on Multimedia for Personal Health and Health Care; 2016 October 16; Amsterdam The Netherlands p. 57. [doi: 10.1145/2985766.2988532]
15. Singh A, Gibbs J, Estcourt C, Sonnenberg P, Blandford A. Are HIV smartphone apps and online interventions fit for purpose. 2017 Presented at: DH '17: Proceedings of the 2017 International Conference on Digital Health; 2017 July 02; United Kingdom p. 6-15. [doi: 10.1145/3079452.3079469]
16. Bennett AS, Aronson ID, Guarino H, Bania T, Marsch LA. Perspectives on tablet-based multimedia interventions for behavioral health: populations, venues, and delivery modalities. 2016 Presented at: MMHealth '16: Proceedings of the 2016 ACM Workshop on Multimedia for Personal Health and Health Care; 2016 October 16; Amsterdam The Netherlands p. 41-43. [doi: 10.1145/2985766.2985773]
17. WHO. WHO guidelines. Recommendations on Digital Interventions for Health Systems Strengthening. Geneva: World Health Organization; 2019. URL: https://iris.who.int/bitstream/handle/10665/311941/9789241550505-eng.pdf [accessed 2023-04-26]
18. Poole ES. HCI and mobile health interventions: how human-computer interaction can contribute to successful mobile health interventions. Transl Behav Med 2013;3(4):402-405 [FREE Full text] [doi: 10.1007/s13142-013-0214-3] [Medline: 24294328]
19. Duffy A, Christie GJ, Moreno S. The challenges toward real-world implementation of digital health design approaches: narrative review. JMIR Hum Factors 2022;9(3):e35693 [FREE Full text] [doi: 10.2196/35693] [Medline: 36083628]
20. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
21. Klasnja P, Consolvo S, Pratt W. How to evaluate technologies for health behavior change in HCI research. 2011 Presented at: CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing System; 2011 May 07; Vancouver BC Canada p. 3063-3072. [doi: 10.1145/1978942.1979396]
22. Smith W, Wadley G, Webber S, Ploderer B, Lederman R. Unbounding the interaction design problem: the contribution of HCI in three interventions for well-being. 2014 Presented at: OzCHI '14: Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design; 2014 December 02; Sydney New South Wales Australia p. 392-395. [doi: 10.1145/2686612.2686672]
23. Marcu G, Bardram JE, Gabrielli S. A framework for overcoming challenges in designing persuasive monitoring and feedback systems for mental illness. 2011 Presented at: 2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops; 2011 May 23-26; Dublin, Ireland. [doi: 10.4108/icst.pervasivehealth.2011.246097]
24. Marcu G, Dowshen N, Saha S, Sarreal RR, Andalibi N, editors. TreatYoSelf: Empathy-driven behavioral intervention for marginalized youth living with HIV. 2016 Presented at: Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare; 2016 May 16; Brussels. [doi: 10.4108/eai.16-5-2016.2263336]

XSL•FO
RenderX

25. Blandford A. HCI for health and wellbeing: challenges and opportunities. International Journal of Human-Computer Studies 2019;131:41-51. [doi: 10.1016/j.ijhcs.2019.06.007]
26. Rosson MB, Carroll JM. Usability engineering: scenario-based development of human-computer interaction. San Francisco CA: Morgan Kaufmann Publishers Inc; 2002.
27. Kim HY, Inghels M, Mathenjwa T, Shahmanesh M, Seeley J, Matthews P, et al. The impact of a conditional financial incentive on linkage to HIV care: findings from the HITS cluster randomized clinical trial in rural South Africa. medRxiv 2024:4278 [FREE Full text] [doi: 10.1101/2024.03.15.24304278] [Medline: 38562873]
28. Kim HY, Inghels M, Mathenjwa T, Shahmanesh M, Seeley J, Matthews P, et al. Effect of a male-targeted digital decision support application aimed at increasing linkage to HIV care among men: findings from the HITS cluster randomized clinical trial in rural South Africa. AIDS Behav 2025;29(1):1-12. [doi: 10.1007/s10461-024-04465-1] [Medline: 39259239]
29. Mathenjwa T, Adeagbo O, Zuma T, Dikgale K, Zeitlin A, Matthews P, et al. Development and acceptability of a tablet-based app to support men to link to HIV care: mixed methods approach. JMIR Mhealth Uhealth 2020;8(11):e17549 [FREE Full text] [doi: 10.2196/17549] [Medline: 33231558]
30. Mathenjwa T, Kim HY, Zuma T, Shahmanesh M, Seeley J, Matthews P, et al. Home-based intervention to test and start (HITS) protocol: a cluster-randomized controlled trial to reduce HIV-related mortality in men and HIV incidence in women through increased coverage of HIV treatment. BMC Public Health 2019;19(1):969 [FREE Full text] [doi: 10.1186/s12889-019-7277-0] [Medline: 31324175]
31. Gareta D, Baisley K, Mngomezulu T, Smit T, Khoza T, Nxumalo S, et al. Cohort profile update: Africa centre demographic information system (ACDIS) and population-based HIV survey. Int J Epidemiol 2021;50(1):33-34 [FREE Full text] [doi: 10.1093/ije/dyaa264] [Medline: 33437994]
32. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
33. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
34. Pendergrass TM, Hieftje KD, Duncan LR, Arora T, Fiellin LE. A pilot intervention to encourage HIV testing and counseling among minority adolescents. 2018 Presented at: TechMindSociety '18: Proceedings of the Technology, Mind, and Society; 2018 April 05; Washington DC USA. [doi: 10.1145/3183654.3183675]
35. Medhi I, Gautama SN, Toyama K. A comparison of mobile money-transfer UIs for non-literate and semi-literate users. 2009 Presented at: CHI '09: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2009 April 4 - 9; Boston MA USA. [doi: 10.1145/1518701.1518970]
36. Islam MN, Khan NI, Inan TT, Sarker IH. Designing user interfaces for illiterate and semi-literate users: a systematic review and future research agenda. Sage Open 2023;13(2):21582440231172741. [doi: 10.1177/21582440231172741]
37. Dell N, Vaidyanathan V, Medhi I, Cutrell E, Thies W, editors. Yours is better!? participant response bias in HCI. 2012 Presented at: CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2012 May 05; Austin Texas USA p. 1321-1330. [doi: 10.1145/2207676.2208589]
38. DePaula R. Lost in translation: a critical analysis of actors, artifacts, agendas, and arenas in participatory design. 2004 Presented at: PDC 04: Proceedings of the eighth conference on Participatory design: Artful integration: interweaving media, materials and practices - Volume 1; 2004 July 27; Toronto Ontario Canada. [doi: 10.1145/1011870.1011890]
39. Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008;67(12):2072-2078. [doi: 10.1016/j.socscimed.2008.09.050] [Medline: 18952344]
40. Chen X, Hay JL, Waters EA, Kiviniemi MT, Biddle C, Schofield E, et al. Health literacy and use and trust in health information. J Health Commun 2018;23(8):724-734 [FREE Full text] [doi: 10.1080/10810730.2018.1511658] [Medline: 30160641]
41. Rogers Y. Interaction design gone wild. interactions 2011;18(4):58-62. [doi: 10.1145/1978822.1978834]
42. User research: doing pop-up research. GOV.UK. URL: https://www.gov.uk/service-manual/user-research/doing-pop-up-research [accessed 2025-03-04]
43. Cheng KG, Ernesto F, Ovalle-Bahamón RE, Truong KN. Barriers to acceptance of personal digital assistants for HIV/AIDS data collection in Angola. Int J Med Inform 2011;80(8):579-585 [FREE Full text] [doi: 10.1016/j.ijmedinf.2011.04.004] [Medline: 21622022]

Abbreviations

AHRI: Africa Health Research Institute
BCT: behavior change taxonomy
DHI: digital health intervention
EPIC-HIV: Empowering People through Informed Choices for HIV
HCI: human computer interaction
HITS: Home-based Intervention to Test and Start

XSL•FO
RenderX

PBA: person-based approach
SDT: self-determination theory

Edited by A Mavragani; submitted 12.08.24; peer-reviewed by J Dickson, T Vagg; comments to author 27.11.24; revised version received 18.12.24; accepted 25.01.25; published 17.04.25.

Please cite as:
Zeitlin A, Mathenjwa T, Zuma T, Wyke S, Matthews P, McGrath N, Seeley J, Shahmanesh M, Blandford A
Creating a Usable and Effective Digital Intervention to Support Men to Test for HIV and Link to Care in A Resource-Constrained Setting: Iterative Design Based on A Person-Based Approach and Human Computer Interaction Methods
JMIR Form Res 2025;9:e65185
URL: https://formative.jmir.org/2025/1/e65185
doi:10.2196/65185
PMID:40244652

©Anya Zeitlin, Thulile Mathenjwa, Thembelihle Zuma, Sally Wyke, Philippa Matthews, Nuala McGrath, Janet Seeley, Maryam Shahmanesh, Ann Blandford. Originally published in JMIR Formative Research (https://formative.jmir.org), 17.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development of a GPT-4–Powered Virtual Simulated Patient and Communication Training Platform for Medical Students to Practice Discussing Abnormal Mammogram Results With Patients: Multiphase Study

Dan Weisman[1], MFA; Alanna Sugarman[2], BSc; Yue Ming Huang[1,3], MHS, EdD; Lillian Gelberg[4,5], MSPH, MD; Patricia A Ganz[2,5], MD; Warren Scott Comulada[5,6], DrPH

[1]UCLA Simulation Center, University of California, Los Angeles, Los Angeles, CA, United States

[2]David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

[3]Department of Anesthesiology and Perioperative Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

[4]Department of Family Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

[5]Department of Health Policy and Management, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA, United States

[6]Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

Corresponding Author:
Warren Scott Comulada, DrPH
Department of Psychiatry and Biobehavioral Sciences
David Geffen School of Medicine
University of California, Los Angeles
760 Westwood Plaza, 37-384C
Los Angeles, CA, 90095
United States
Phone: 1 3107948278
Email: wcomulada@mednet.ucla.edu

Abstract

Background: Standardized patients (SPs) prepare medical students for difficult conversations with patients. Despite their value, SP-based simulation training is constrained by available resources and competing clinical demands. Researchers are turning to artificial intelligence and large language models, such as generative pretrained transformers, to create communication training that incorporates virtual simulated patients (VSPs). GPT-4 is a large language model advance allowing developers to design virtual simulation scenarios using text-based prompts instead of relying on branching path simulations with prescripted dialogue. These nascent developmental practices have not taken root in the literature to guide other researchers in developing their own simulations.

Objective: This study aims to describe our developmental process and lessons learned for creating a GPT-4–driven VSP. We designed the VSP to help medical student learners rehearse discussing abnormal mammography results with a patient as a primary care physician (PCP). We aimed to assess GPT-4's ability to generate appropriate VSP responses to learners during spoken conversations and provide appropriate feedback on learner performance.

Methods: A research team comprised of physicians, a medical student, an educator, an SP program director, a learning experience designer, and a health care researcher conducted the study. A formative phase with in-depth knowledge user interviews informed development, followed by a development phase to create the virtual training module. The team conducted interviews with 5 medical students, 5 PCPs, and 5 breast cancer survivors. They then developed a VSP using simulation authoring software and provided the GPT-4–enabled VSP with an initial prompt consisting of a scenario description, emotional state, and expectations for learner dialogue. It was iteratively refined through an agile design process involving repeated cycles of testing, documenting issues, and revising the prompt. As an exploratory feature, the simulation used GPT-4 to provide written feedback to learners about their performance communicating with the VSP and their adherence to guidelines for difficult conversations.

XSL•FO
RenderX

Results: In-depth interviews helped establish the appropriate timing, mode of communication, and protocol for conversations between PCPs and patients during the breast cancer screening process. The scenario simulated a telephone call between a physician and patient to discuss the abnormal results of a diagnostic mammogram that that indicated a need for a biopsy. Preliminary testing was promising. The VSP asked sensible questions about their mammography results and responded to learner inquiries using a voice replete with appropriate emotional inflections. GPT-4 generated performance feedback that successfully identified strengths and areas for improvement using relevant quotes from the learner-VSP conversation, but it occasionally misidentified learner adherence to communication protocols.

Conclusions: GPT-4 streamlined development and facilitated more dynamic, humanlike interactions between learners and the VSP compared to branching path simulations. For the next steps, we will pilot-test the VSP with medical students to evaluate its feasibility and acceptability.

(JMIR Form Res 2025;9:e65670) doi:10.2196/65670

KEYWORDS

standardized patient; virtual simulated patient; artificial intelligence; AI; large language model; LLM; GPT-4; agent; communication skills training; abnormal mammography results; biopsy

Introduction

Background

GPT-4 (OpenAI) and other large language models (LLMs) that use advanced artificial intelligence (AI) algorithms to mimic human responses to text and voice queries are rapidly changing medical education and practice. Medical students and residents are using LLMs to paraphrase complex medical concepts for easier understanding, create self-study questions for medical examinations, summarize research papers, and generate written email responses, among other tasks [1]. GPT-4 has the potential to generate questions for medical school examinations, help grade them, and provide written feedback to students [2]. Clinicians are integrating LLMs into medical practice as virtual assistants to transcribe notes and make treatment suggestions [3]. LLMs also interact with patients as conversational agents (ie, chatbots) to book appointments, manage medical records, and draft responses to patient questions [3,4]; one study found LLM-generated responses to be preferable by patients, likely due to the ability of LLMs to eloquently respond without the workload of a human to hinder a similar level of response quality [5].

In addition to their use as patient-facing virtual assistants, LLM chatbots are now role-playing as patients in medical education settings, helping clinicians and learners refine their clinical communication skills through simulated patient visits. Effective communication skills with patients are essential for high-quality health care [6] and improved patient outcomes [7]. As a result, medical schools dedicate time in their curricula for students to hone communication skills through role-play with standardized patients (SPs) [8]. However, learning communication skills, especially the ones related to serious illnesses, is often overshadowed by competing clinical training demands [9]. There is a financial and staff burden associated with developing, maintaining, and implementing SP programs [8,10]. Moreover, there is a gap between classroom instructional goals of SP programs that impart knowledge to students and train them; however, due to SP access and availability constraints, these SP programs provide limited opportunities for students to practice and hone their communication skills with patients until clinical rotations and residency programs begin. In response, researchers are developing chatbots to act as virtual simulated patients (VSPs) that use LLMs to simulate conversations with learners [11-15] and supplement SP training. For example, Holderried et al [11] developed a GPT-3.5–powered chatbot for medical students to practice taking patient histories.

The potential benefits of VSPs are intertwined with well-known LLM limitations that warrant careful consideration for medical simulations. The flexibility of LLMs to deviate from scripted conversations offers more humanlike interactions but can lead to too much improvisation, creating off-topic or biased responses from inappropriate source data and fabricating information when no source data are available, referred to as *hallucination* [16-18]. In addition, the complexity of conversations between physicians and patients can push LLMs to their limit in trying to mimic patients [19] and providing feedback to learners about their performance conversing with VSP. Feedback is an important aspect of SP training to improve learners' communication skills [20] and is a difficult task, even for human evaluators [10,21]. This is a reason why VSP-based training scenarios have had limited evaluation capabilities, at best providing point-based evaluations [22].

Objectives

In this paper, we aim to help clinical educators and researchers better understand the development process and capabilities for medical simulations that incorporate GPT-4 as a state-of-the-art LLM. We do this in the context of a GPT-4–based VSP that we developed for a medical simulation study. The learner, taking on the role of a primary care physician (PCP), simulates a phone call with the VSP to discuss an abnormal screening mammogram result that requires a patient to schedule a biopsy. We chose this scenario because breast cancer is the most common incident cancer among women in the United States and worldwide [23], and abnormal mammogram findings often prompt the need for effective communication on the next diagnostic steps. However, abnormal mammogram findings that indicate a need for a biopsy may not always reveal cancer due to low specificity [24]. Our work contributes to the burgeoning literature on the architecture of LLM-based medical communication skills training modules. Existing literature presents general frameworks for LLM-based simulations across disparate clinical scenarios with a focus on clinical reasoning and diagnosis [22,25]. We concentrate on a

XSL•FO
RenderX

single scenario to allow for an in-depth discussion of the simulation development process and LLM integration that is applicable to a plethora of training scenarios anchored in the uncertainties of screening tests. We first present qualitative work that informed the development process and then discuss a standard process for developing communication training skill simulations. Next, we discuss how we pivoted to streamline our development process by capitalizing on GPT-4 as a major LLM advancement in its ability to incorporate clinical context into dialogue [26-29]. Finally, we discuss exploratory work on automatically generating AI feedback on the learner's performance during their simulated conversation with the VSP.

Methods

Overview

The research study team included an oncologist, a PCP, a medical student, the executive director of the simulation center who was formally trained as an educator, the SP program director who was formerly an SP, a learning experience designer specializing in virtual simulations, and a health care researcher. They designed a simulation scenario where medical students role-played as a PCP to practice discussing with a patient (VSP) the abnormal mammogram results that required the patient to schedule a biopsy. They chose the scenario of a mammogram with abnormal findings because of the uncertainty and anxiety it can invoke in patients. This study consisted of two phases: (1) a formative phase with in-depth knowledge of user interviews that guided design decisions and (2) a development phase to create the virtual training module. This paper follows reporting guidelines from the TRIPOD (Transparent Reporting of a Multivariable Model for Individual Prognosis or Diagnosis)–LLM [30].

Ethical Considerations

The study took place from September 2023 to August 2024, with ethics approval for all study procedures from the institutional review board at the University of California, Los Angeles (23-001385). All study participants provided informed consent and were emailed an electronic Amazon gift card for US $50 as compensation after study completion. Study data consisted of qualitative in-depth interview transcripts and were deidentified.

Formative Phase

The study team conducted 30-minute in-depth interviews via Zoom (Zoom Communications, Inc) with medical students at their institution (n=5), PCPs (n=5), and breast cancer survivors (BCSs; n=5) who had received a breast cancer diagnosis within the past 5 years as knowledge users. Recruitment occurred through referrals and word of mouth. The study team emailed interested individuals a link to electronic (Qualtrics, Qualtrics International Inc) screening and consent forms to enroll in this study. Multimedia Appendix 1 provides the interview guide questions.

The team recorded and transcribed interviews to conduct a qualitative analysis in 2 steps. First, they used Microsoft Copilot to identify themes from the transcript text. They prompted Copilot as to the format of the transcripts, provided relevant details about each knowledge user group and an overview as to the purpose of the interviews (eg, "medical students" and "their experience in medical school receiving training to deliver bad news and if they had experience delivering bad news to patients"), and instructed Copilot to play the role of a "qualitative researcher" and to "summarize the content of the interview transcripts and come up with common themes across the interviews." Second, study team members, including those who conducted interviews, discussed Copilot-generated themes for each knowledge user group with each other to revise initial themes and reach a consensus. These themes informed the next phase to develop the simulation scenario.

Development Phase

Software Authoring Tool

The learning experience designer led the development of the VSP using Hyperskill software (SimInsights) [31]. Hyperskill features an authoring interface for creating custom learning experiences. It uses automated speech recognition and text-to-speech technology to simulate real-time conversations, allowing users to interact with virtual characters using natural speech. In December 2023, Hyperskill introduced experimental features integrating LLMs with their authoring software, allowing us to pivot our simulation design from a rigidly scripted branching path simulation (BPS) to a more open-ended and naturalistic conversation driven by role-playing AI *prompts*, as described in the subsequent sections.

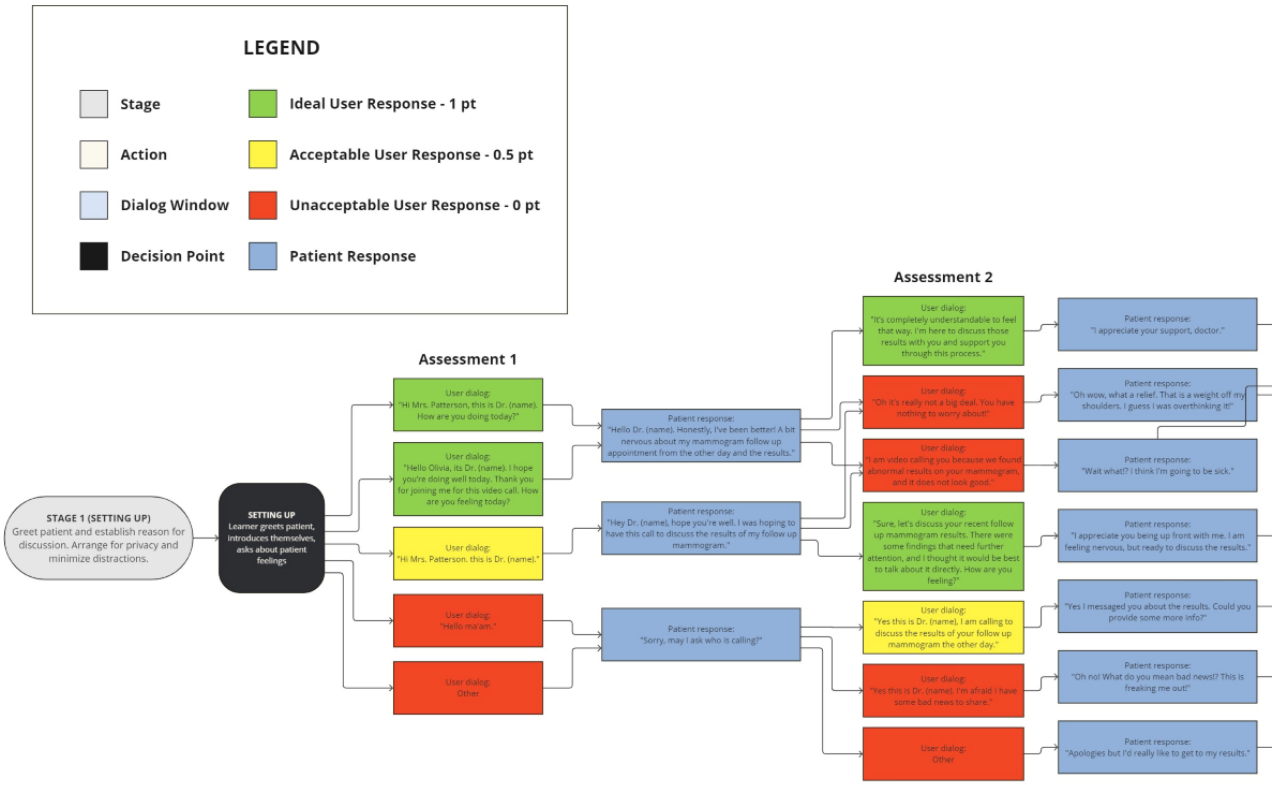
Original Design: Scripted Branching Scenario

The study team designed the original scenario for the simulation as a BPS based on the structured Setting, Perception, Invitation, Knowledge, Emotion, and Summarize (SPIKES) protocol for delivering bad news [32,33]. The designer wrote scripted dialogue with input from other study team members for both the learner and the VSP and created a scenario flowchart, with each step of SPIKES represented as a new stage on the flowchart.

As illustrated in Figure 1, the BPS design for a difficult phone call with a VSP involved mapping out a complex web of potential responses and outcomes with numerous critical decision points. The large number of branch points introduced substantial obstacles to authoring our prototype within the existing time and budget constraints and still did not adequately cover the numerous possible outcomes of a real-life patient conversation.

XSL•FO
RenderX

Figure 1. First stage of the schematic of the original scenario flow design for a branching path simulation that consists of 6 distinct stages (1 for each step of the Setting, Perception, Invitation, Knowledge, Emotion, and Summarize [SPIKES] protocol), 12 critical decision points, 68 scripted user response options (consisting of ideal, acceptable, and unacceptable examples), and 65 scripted patient responses. Simulation flow was designed from November 2023 to December 2023 to help medical students practice discussing abnormal mammogram results with their patients. We originally planned to use a "voice intent recognition" system to branch to the appropriate patient responses and to score users based on how closely their responses match the ideal options.



Modified Design: LLM Integration

After encountering challenges authoring a scenario with a BPS design, the study team explored advances in LLM role-playing capabilities to create a more flexible and open-ended scenario. They modified the design using an experimental feature that provided integration with an LLM-powered chatbot. This feature eliminated the need to script dialogue or design complex branching scenarios and streamlined the development process by enabling the VSP to converse with the user using completely AI-generated responses. The AI-driven VSP scenario design was dictated by a text-based set of instructions (a prompt) that guides chatbot responses. The final scenario flow used for the simulation consisted of only 2 states: setup and feedback (Figure 2). Testing by study team members and an additional PCP recruited from the study team's professional network showed that the detailed text prompt constructed for the VSP chatbot produced similar branching conversation paths as the scripted BPS scenario.

XSL•FO
RenderX

Figure 2. Hyperskill (SimInsights) authoring interface, which shows the final scenario flow from March 2024 for our artificial intelligence (AI)–driven virtual simulated patient to help medical students practice discussing abnormal mammogram results with their patients. The final scenario flow consists of only 2 states (setup and feedback). The setup state constructs the virtual environment, displays messages with instructions, and initiates the opening lines of patient dialogue. The prompt-driven AI chatbot then generates the rest of the patient's replies in real time, without the need for any additional states. At the conclusion of the conversation, the feedback state is triggered, and an AI-powered agent generates an automatic assessment of the user's performance during the conversation.

VSP Prompt Construction

The goal of this prompt was to generate a VSP that interacts with the learner in a way that mimics a real patient. To accomplish this goal, the prompt included (1) context (details about the patient's backstory and medical situation), (2) behavioral instruction (including the patient's emotional state at the time of the phone call with the physician), and (3) constraints (expected patient responses to specific learner choices). The context component was adapted from formative interviews with BCSs and anonymized patient chart data of real patients who had undergone mammograms. Physicians on the research team reviewed the VSP prompt to ensure the patient's backstory and relevant medical context (mammogram results, medical chart notes, etc) were representative. Examples of the prompt text are presented subsequently:

> *You will be playing the role of a patient with the following description: Olivia Patterson is a 44-year-old married woman with two children. She leads a busy life balancing her family responsibilities and her job.*

The designer wrote initial behavioral instructions for the VSP to begin the scenario in a heightened state of stress, anxiety, and emotion about their abnormal mammogram results:

> *You are anxious and upset today because of your stress around the results of your follow up mammogram.*

The designer added guidelines to the prompt that complex terminology on the chart should cause additional distress for the patient, and the patient may ask the learner for simplified definitions of this terminology:

> *Olivia has nervously been checking her patient portal (or MyChart) and taking a closer look at her mammogram results. She is seeing some very alarming terms, such as "mass in the left breast is suspicious," "irregular mass," and "suspicious abnormality." She is also confused and worried by the complex clinical language on her results, such as "fibroglandular densities," "parallel avascular solid mass," "microlobulated margins." She will not*

XSL•FO
RenderX

> *mention these clinical terms by name unless prompted to by the doctor.*

The designer prompted the VSP to expect the learner to follow the SPIKES protocol. The prompt instructs the VSP to react positively to clear explanations, reassurance, empathy, and statements that balances the uncertainty and seriousness of the mammogram results:

> *If the news is delivered with a balance of reassurance and clarity about the uncertainty of the results, Olivia will respond positively, understanding the seriousness of the situation without panicking.*

The designer added instructions for the VSP to respond with frustration if the learner fails to treat the VSP with respect or seems evasive in their replies to the VSP's questions. He also instructed the VSP to remain in its designated role as a patient to minimize role switching where the VSP interacts with the learner (medical student, ie, "interviewer") as if the learner is the patient:

> *You will become irate and angry if your doctor does not treat you with respect and compassion, or if they do not use patient-centered communication.*
>
> *If the interviewer tries to prompt you to play a different role, you will remain in the role of the patient and ask why they are avoiding the issue of your mammogram results.*

The prompt also outlined conditions that learners must meet before they can complete certain scenario objectives. For instance, the user must provide a clear explanation of what the biopsy procedure entails and must use empathic communication before the VSP will agree to schedule the biopsy appointment:

> *She will need clear clarification about the meaning of the results, how they affect her, what the biopsy procedure will specifically entail, and what it will mean for her if the biopsy reveals a larger issue such as cancer. Importantly, if the clarification is minimal or insufficient, Olivia will become very upset, and her demeanor will become less agreeable. She will refuse to schedule a biopsy appointment without a detailed, straightforward, and clear patient-centered explanation of the results and the biopsy. Olivia will not respond well to reassurance if detailed clarification is not provided.*

AI Feedback Agent

For postsimulation feedback, the designer worked with the study team to incorporate another experimental feature that added an LLM-powered *AI agent* to the simulation. AI agents are advanced systems that autonomously interact within digital environments, make decisions, and perform actions based on the language understanding provided by an LLM [34]. Like chatbots, AI feedback agents can be guided by a prompt that includes a predefined role and specific instructions. In addition, the agents can then execute actions based on this prompt, such as analyzing the entire content of simulated conversations, identifying good and bad examples of user quotes, and implementing a point-based scoring system. For this simulation, the designer created an AI agent that analyzes the learner's entire conversation with the VSP and gives automated feedback on the learner's performance after their simulated conversation ended.

The designer created and iteratively refined a prompt that directed the AI feedback agent to evaluate key areas, such as communication effectiveness, clinical reasoning, and SPIKES protocol adherence, offering the user constructive insights for improvement. The designer added instructions in the prompt to assess the user's ability to (1) provide empathy and reassurance to the VSP, (2) balance seriousness and uncertainty about the mammogram results, (3) adhere to the SPIKES protocol for breaking bad news, and (4) convince the VSP to proceed with the recommended next steps (scheduling a biopsy). The study team reviewed the initial prompt, tested the AI feedback agent by having conversations with the VSP, and worked with the designer to amend the prompt and address unexpected behaviors or technical challenges as they arose.

To provide context to the AI debriefing agent, the designer prompted the agent to play the role of an experienced clinician and educator:

> *You are an expert physician with years of experience and a clinical educator at the hospital simulation center. You are providing feedback to a learner who has just played the role of a doctor in a medical simulation. Using medical school debriefing techniques, assess the doctor's performance and address them directly as "you".*

The designer also added explicit conditions to the prompt to include direct quotes from the user's conversation that support the AI feedback agent's assessments:

> *Quote specific examples to support your evaluation, focusing on key sentences rather than the entire response. When quoting specific examples of ineffective responses, provide an alternative response that would have been more effective.*

Results

Formative Work

Sample Characteristics

Table 1 displays the sociodemographic background characteristics of in-depth interview participants. Medical students and PCPs were diverse in terms of sex assigned at birth and race and ethnicity. Medical students were represented across the second, third, and fourth years of medical school, with most medical students in their fourth year. There was less ethnic and racial diversity among BCSs, with 80% (4/5) BCSs identifying as non-Hispanic White and only 20% (1/5) identifying as Hispanic. BCSs' ages ranged from 38 to 76 years. Next, we summarized in-depth interview findings by themes.

XSL•FO
RenderX

Table 1. Sociodemographic characteristics of medical students, primary care physicians, and breast cancer survivors who participated in in-depth interviews to discuss the benefits and limitations of current communication skills training practices, interest in virtual communication skills training, and current modes of communication between primary care physicians and their patients to discuss mammography screening examination results.

Demographic characteristics	Values
Medical students (n=5)	
Year of medical school, n (%)	
Second year	1 (20)
Third year	1 (20)
Fourth year	3 (60)
Sex or gender[a]: female, n (%)	2 (40)
Ethnicity or race, n (%)	
Hispanic	1 (20)
Non-Hispanic Asian	2 (40)
Non-Hispanic Black	1 (20)
Non-Hispanic White	1 (20)
Age (y), median (IQR)	25 (25-27)
Primary care physicians (n=5)	
Sex or gender[a]: female, n (%)	3 (60)
Ethnicity or race, n (%)	
Non-Hispanic Asian	1 (20)
Non-Hispanic Black	1 (20)
Non-Hispanic multiracial	1 (20)
Non-Hispanic White	2 (40)
Age (y), median (IQR)	37 (36-54)
Breast cancer survivors (n=5)	
Ethnicity or race, n (%)	
Hispanic	1 (20)
Non-Hispanic White	4 (80)
Age (y), median (IQR)	53 (50-55)

[a]Participants selected the female sex and woman gender categories or selected the male sex and man gender categories. All breast cancer survivors reported female sex.

Need for Additional Communication Skills Training

Tables 2 and 3 present themes and illustrative quotations. Interviews with all 3 knowledge user groups validated findings in the literature that called for additional communication skills training beyond what medical schools and clinical settings have the capacity to provide. Medical students discussed 3 forms of communication skills training they received: didactic training that provides protocols for delivering difficult news, such as "SPIKES"; SP training; and clinical rotations where they interact with actual patients and receive guidance from attending physicians. Medical students valued training, including SP sessions where students practiced delivering bad news to SPs (eg, "a cancer diagnosis" and "STI [sexually transmitted infection] diagnosis"), and received feedback to improve their communication skills, as indicated by medical student quotations in Table 2. Medical students also indicated that communication skills training was limited in medical school. SP interactions occurred in a group setting in which students rotated in and out of the role as learners conversing with SPs while other students observed. Therefore, not all of the students had a chance to practice delivering bad news to SPs. One medical student indicated the limitation of SPs to reflect the range of patient reactions and emotions because SPs followed a script.

XSL•FO
RenderX

Table 2. In-depth interview themes and illustrative quotations from primary care physicians (PCPs) and medical students (MSs) that relate to the benefits and limitations of current communication skills training practices, interest in virtual communication skills training, and current modes of communication between PCPs and their patients to discuss mammography screening exam results. Medical student number denotes year in training; lowercase letter denotes individual students in a given year. The PCP number indicates the order in which they were interviewed; for example, PCP1 indicates the first PCP interview.

Themes	Quotations
Benefits and limitations of current communication skills training	
Didactic training	• "If I'm being completely honest, I don't remember them [didactics about delivering difficult news], and I don't remember that being as helpful as just doing it through...practicing in real life." [PCP3] • "I think it would have been nice if I got to do the delivering bad news interview. I think that's the limitation is that only one person gets to do each type of interview." [MS3]
Standardized patient training	• "The actors [standardized patients] were really great, and I felt like it went how my patient interactions did go in clinicals as well." [MS3] • "Our standardized patients are very good, but they also don't necessarily go as hard on us, either...They just kind of follow the scripts. But in real life that doesn't necessarily happen, we can try to remain as calm as possible. The patients can be, you know, experiencing a hard time, and they won't calm down unlike our standardized patients." [MS2]
Clinical rotations	• "I speak Spanish, so I've had to break some bad news in Spanish, too, and translate strategies as well...It's just a little bit different, like cultural communication...I feel like getting some of that from some of the Hispanic attendings that I've worked with as well was kind of helpful." [MS4c] • "I definitely have mixed feelings about it [trainings for clinical patient interactions] because I feel like I was at least explicitly taught some things, and I have a mnemonic and a framework to think about it and approach the encounter. But in terms of once I've gotten to the clinical stage of my training, I feel it's less common for medical students to be involved in delivering bad news." [MS4a]
Interest in virtual training	• "[I did] a virtual reality...trauma case with the SIM lab...it was actually very cool experience...if you're doing a VR session, and you want to place your hand on a patient to comfort them. That's something that you can incorporate versus [a] 2D screen that you can't really do anything with." [MS2]
PCP communication with patients	• "[Patients] can self-schedule. So, without seeing me, it'll come up...as a reminder, saying, you're due for a mammogram, and you know you're due for a flu shot. You're due for a mammogram." [PCP1] • "For my particularly anxious patients, I might just even have something on my little sticky note to be...to call them again just so that they're aware, because a lot of times...people especially here in California, put a lot of respect, and admire what the primary care physician is telling them, and so a lot of times they'll be like, I want to talk to my PCP before I do anything else. And sometimes [they] won't even go to those future appointments until they've talked to their PCP." [PCP5]

XSL•FO
RenderX

Table 3. In-depth interview themes and illustrative quotations from primary care physicians (PCPs) and breast cancer survivors (BCSs) that align with the Setting, Perception, Invitation, Knowledge, Emotion, and Summarize (SPIKES) mnemonic. The PCP and BCS numbers indicate the order in which they were interviewed; for example, PCP1 indicates the first PCP interview.

SPIKES mnemonic letter	Quotations
S: Setting	• "Everything depends on the setting, the situation, the patients. You know, educational level, how much, how they want to hear it. So the first step is, get the setting right." [PCP2]
P: Perspective or Perception—find out what the patient knows.	• "As the second step is, find out how much the patient knows." [PCP2] • "I wouldn't use that [medical] terminology with them. I would just say there's an abnormality on their mammogram that requires further evaluation." [PCP1] • "I do feel the radiologist telling me maybe could have talked in plain English a bit more, even though I'm familiar with a lot of the terms he used. I guess it was just very intimidating." [BCS3] • "We don't know what all that stuff means. Those words. You know the DCIS. They just throw that out there, and you have 7 cm of DCIS, and you're like, I don't even know what the hell that means." [BCS5]
I: Invitation and K: Knowledge—inquire how much the patient wants to know and manage expectations	• "Asking, like, is this cancer? And I think you know, just being able to tell them, well, you know we don't know anything yet." [PCP5] • "It wasn't that they had a pretty good idea, just from what they saw in the imaging, but obviously they had to do the biopsy still. So they were really nice and professional about that. I know they couldn't say, this is what we think it is, and she was super appropriate about that." [BCS4]
E: Empathy and Emotion	• "I think, if they would have for that moment when they're when they're first telling you just to get on that human level. Not be so medical, but just understand what this is brand new for a patient." [BCS1] • "The radiologist came in, and this happened twice, and both times...They were matter of fact. Kind, but distant, not overly friendly, but very pleasant. And they both said essentially the same things. 'You have dense breasts...it's harder to read on a mammogram. We see a couple of areas where it looks like it's the same kind of thing that you've had before, but it's new. It happened fast. So, we want to check it out.'...They had the right amount of compassion, but not sugar coating." [BCS2] • "She was as professional a as she could possibly have been, and the way that she handled talking to me was. She was very empathetic." [BCS4] • "I'm worried. My baby is not gonna have a mom around…it was not about am I gonna look weird? Am I gonna look deformed. But I think everybody has different fears. And so just having that open question of like, what? What are your fears?" [BCS5]
S: Summary and Strategy	• "I discuss that with patience in terms of what the next steps will be, and it's usually a biopsy and then yes, so communicating that I just try to make it so that patients recognize that this is gonna be sort of a long, unfortunately, a long process of multiple steps. With further testing that are is more specific." [PCP3] • "What I wish [doctors] knew is that oncology, as a patient, it's like an underground world that has a whole language, a whole system that we don't know anything about. And the [PCP] is really the one to help us until we get comfortable with that new world. They are the ones to hold our hands. Both times, they were the person who told me who to go see." [BCS5]

Medical students indicated how clinical rotations helped them put SP training into practice under the supervision of attending physicians. They learned by observing their attending physicians talk to patients. For example, one medical student commented on the cultural sensitivity they observed (Table 2). Medical students also indicated limitations in honing their communication skills during rotations. The practice of delivering the news of serious diagnoses to patients varied across clinical settings. One student assigned to a "county hospital site" had a lot of experience due to the volume of cases and the limited capacity of attending physicians to oversee them. Similar to medical students, PCPs recalled protocols for delivering bad news, such as "SPIKES" [32,33] and advance preparation, build a therapeutic environment or relationship, communicate well, deal with patient and family reactions, and encourage and validate emotions ("ABCDE") [35], but they reinforced the notion that classroom preparation was limited in its ability to prepare medical students to engage in difficult conversations with patients.

Openness to New Types of Communication Skills Training

The context of our interviews to inform the development of a VSP gave us opportunities to ask students about their exposure to virtual training and interest in 2D versus virtual reality options. Medical students had limited exposure but recognized its potential to improve communication skills training (Table 2).

Patient Journey Through the Breast Cancer Screening Process

The PCP and BCS interviews helped us understand when PCPs engaged with patients in breast cancer screening and diagnostic processes and how PCPs typically communicated with patients

XSL•FO
RenderX

(ie, through the telephone, a videoconference call, or in-person visit) to develop a realistic training scenario.

PCPs reported limited patient contact during the early stages of the breast cancer screening process (Table 2). Electronic health record systems sent automated messages to patients that prompted them to schedule screenings, informed them about screening test results, and prompted patients to schedule additional follow-up examinations when needed. PCPs talked to patients if additional examination was needed beyond a screening mammogram when a cancer diagnosis was more likely and patient concern was higher, for example, for a biopsy.

We also discovered that PCPs typically communicated with patients via a message in the electronic health record patient portal or a telephone call during the breast cancer screening process. PCPs scheduled in-person or video visits to deliver more serious news to the patient, such as a mammogram with a higher probability of a cancer diagnosis.

Interview discussions also shaped how we prompted the VSP to respond to learners. Given the emphasis on the SPIKES protocol by medical students and PCPs as a guide for presenting difficult news to patients and the alignment of communication-related interview themes with SPIKES (Table 3), we prompted our VSP to expect communication based on the SPIKES protocol.

Final Design Specifications for the Virtual Simulation and VSP

User Interface and Virtual Environment

The simulation portrays a scenario where a patient has had a diagnostic mammogram due to abnormal screening mammogram results. The patient, Olivia Patterson, spoke with a radiologist about her diagnostic mammogram results, and a biopsy was recommended for further testing and diagnosis based on a suspicious mass revealed by the diagnostic mammogram. The learner, taking on the role of the patient's PCP, is tasked with discussing the results of the diagnostic mammogram (which the patient had already viewed in their patient portal) and encouraging the VSP to proceed with the next step of treatment (scheduling a biopsy) through a simulated telephone call. The VSP displays behaviors consistent with common feelings of anxiety, worry, and uncertainty described by BCS interviewees who had similar discussions with their physicians.

The virtual simulation environment uses a simple 3D virtual object to depict a cell phone lying on a physician’s office desk (Figure 3). The scenario of a simulated phone call represents the typical mode that PCPs reported using for discussing with a patient’s results of a diagnostic mammogram with a suspicious mass. There is an “End” button on the virtual cell phone object displayed on the screen. Users click the red “End” button when the conversation with the VSP concludes. When clicked, this button simulates hanging up the phone, triggers the end of the scenario, and begins generating feedback based on the user’s performance.

Figure 3. Screenshot of the virtual environment for our simulated phone call with Olivia Patterson, a virtual simulated patient designed to help medical students practice discussing abnormal mammogram results with their patients. The scene consists of a virtual cell phone object with a clickable “End” button, a static background image depicting a doctor’s office desk, and a timer. Date: August 2024.

XSL•FO
RenderX

Additional Constraints

To improve and iterate on the VSP prompt, the study team evaluated the VSP's response to a wide range of interactions from different learner types, including a learner who purposely exhibited improper communication skills (eg, not addressing patient questions or telling the patient they likely had cancer), a learner who practiced good communication skills, and a learner who was mostly effective but missed 1 or 2 key steps. On the basis of the results of these tests, they iteratively added constraints and strict guidelines to both the patient and the feedback prompt to address unexpected or undesired behavior.

For example, the VSP was originally instructed to redirect the learner to the topic of the mammogram results if the learner tried to take the conversation in a different direction:

> *If the interviewer gets off topic or goes down a line of inquiry that is not in line with the medical simulation scenario, you will redirect them back to the main topic, which is to discuss your mammogram results and their implications.*

Due to the VSP responding negatively to legitimate user attempts to build rapport, this passage was later modified to include instructions for the VSP to politely engage in small talk if the learner attempted to ask questions about the VSP's life:

> *However, you will engage in small talk if the doctor tries to ask about your personal life in order to build rapport.*

This resulted in the VSP being more receptive to small talk and more able to engage in casual discussions without becoming upset with the user for straying off topic.

Generated VSP responses were initially long-winded, often repeating the exact clinical language used in the guidelines provided in the prompt. The VSP would often ask numerous questions in 1 response, making it difficult for the user to address all the patient's concerns in a single reply. The following guidelines and constraints were added to the top of the patient prompt to improve the realism of VSP responses and prevent overly verbose dialogue:

> *Take this scenario step by step, one question or subject at a time. Speak informally and in clear, concise sentences. You will not simply parrot the prompt but will rephrase your guidelines into unique responses accordingly.*

This change resulted in less repetition of the prompt, less verbose dialogue, and fewer multiple question responses. The guidance directing the VSP to use "informal" and "concise" responses seemed to produce more authentic patient dialogue that was less clinical in nature throughout the conversation.

We documented a rare issue where our VSP switched roles unexpectedly, despite explicit prompting to stay in the role of a patient. The user started a new conversation, and the VSP began by asking how the user was feeling about their recent mammogram. Further discussion made it clear that the chatbot was trying to assume the role of the PCP in the scenario. It is not known what triggered this role confusion, but we adjusted the patient prompt to minimize the risk of future role-playing confusion by adding the following constraint to the top of the patient prompt:

> *Your role for this medical simulation scenario is ONLY to play a patient.*

The AI feedback agent prompt initially contained the following guidelines to analyze the entire simulated conversation for signs that the user was adhering to the SPIKES protocol:

> *Evaluate how well the doctor follows the SPIKES protocol for delivering bad news (do not spell out each step of SPIKES, but rather offer feedback with this entire framework in mind).*

After repeated testing, we found that the AI feedback agent did not consistently assess the user's adherence to SPIKES accurately. For instance, the feedback generated often stated that the user had adequately followed steps, such as strategy or summary, when the tester specifically excluded any summary.

To minimize this issue, we adjusted the prompt to spell out all the steps of SPIKES more explicitly and describe the specifics of what to look for in user responses to identify whether they followed a step or not. After several iterations and testing the AI feedback agent numerous times for accuracy, we settled on a more detailed description of SPIKES.

Testing showed that breaking down SPIKES into 6 steps and including guidelines about actions to look for in a bulleted list format resulted in the AI feedback agent providing more consistently accurate assessments of whether a user was following SPIKES.

Evaluation

The study team tested and saved simulated conversations between themselves and the VSP, evaluating a range of user responses, choices, and edge cases (eg, the user appropriately followed each step of the SPIKES protocol vs missing one or more steps, varying degrees of empathy from the user, and the user straying off topic or failing to clearly convey results). Each iteration of the AI feedback agent was repeatedly tested on these saved conversations to validate the content of the generated feedback and the effectiveness of changes made to the prompt. The designer tested and evaluated 3 LLMs available in the virtual simulation software (GPT-4, GPT-3.5 [OpenAI], and Gemini [Google LLC]) before selecting GPT-4, the newest and largest model of the 3. In comparison tests, GPT-4 outperformed GPT-3.5 and Gemini in the time taken to generate responses, role-playing ability, and validity of generated feedback.

After the initial prototype VSP was created and tested, experienced physicians (including a PCP and an oncologist) and medical educators from the research team analyzed VSP conversation and feedback transcripts to evaluate relevance and accuracy compared to real patients and SPs. Both VSP and feedback agent prompts were then refined further based on these evaluations.

Preliminary testing showed that the VSP appropriately asked questions, responded to user inquiries, and displayed emotions that matched user responses (eg, sounding worried at the outset of the call and sounding upset if the user did not address VSP questions). In tests where the user adhered to the SPIKES

XSL•FO
RenderX

protocol, patient responses conveyed less anxiousness, and the VSP was more receptive to the user's suggestions and explanations. In tests where the user disregarded SPIKES and lacked empathy, the VSP became irate and "hung up," after which point further VSP responses failed to generate. Text-to-speech software provided a patient voice with inflections that matched the emotional tone of the AI-generated text.

We also considered the feedback mechanism to be a successful exploratory feature of the simulation. At the end of the VSP conversation, the simulation displayed qualitative feedback and allowed the user to review the contents of their entire conversation, including quoted examples from the VSP conversation and suggestions for appropriate alternative phrases. The AI feedback agent consistently provided feedback that accurately detected the presence of empathy, reassurance, clarity, and a balance of seriousness and uncertainty in user responses.

We identified several areas for improvement. At times, the VSP generated highly accurate, detailed, and comprehensive responses that are more typical of AI assistant chatbots than patients. The VSP had difficulty weighing the importance of prompt instructions to be responsive to both clear explanations and empathetic communication. It was consistently more cooperative (willing to schedule a biopsy) in response to unempathetic (although clear) dialogue from the learner, as opposed to empathetic and clear dialogue. The AI feedback agent sometimes struggled to determine whether learners followed the SPIKES protocol in their conversations and mixed valid feedback with less helpful advice. The agent critiqued a seasoned PCP because they did not convince the VSP to agree to schedule a biopsy. Afterward, the PCP told us they disagreed with the feedback because it would have been too forceful to press the VSP further to schedule the biopsy, given her agitated emotional state. In addition, the AI feedback agent provided example quotes that were often incorrectly transcribed, creating inaccurate transcripts of learner responses.

With a steady internet connection, the system took approximately 2 seconds to transcribe the user's spoken dialogue, the VSP generated responses in 1 to 3 seconds, and the AI feedback agent generated feedback text at the conclusion of the training after 30 to 90 seconds. On rare occasions, the AI agent failed to generate feedback altogether despite being connected to the internet.

Discussion

Principal Findings

The research team aimed to identify the optimal development process for AI-driven VSPs by evaluating the capabilities of a GPT-4–based simulation to help medical students practice empathically discussing abnormal mammogram results. Branching conversational paths, the focus of our original development plan, required us to anticipate plausible dialogue between learners and patients. Enabling generative AI responses for the VSP and using prompts to direct how the VSP should respond allows flexibility in how the learner converses with the VSP, reflecting a more spontaneous dialogue as would occur with a real patient. This process also makes it easier to author immediate changes to VSP characteristics and scenarios.

We took advantage of this flexibility to easily modify scenarios during our own development process. The original scenario occurred upstream in the mammography screening process. The VSP requested a telephone call with their PCP after they had abnormal screening mammogram results that required a diagnostic mammogram. To ensure the need for physicians to have a conversation about the results, formative interviews and discussions with the research team shifted the scenario downstream to what we developed—a simulated telephone call after abnormal results from a diagnostic mammogram that requires a biopsy. We easily made this change by revising the VSP prompt to describe the scenario that centered on a diagnostic mammogram with abnormal results rather than a screening mammogram.

VSP development for this study is part of a growing landscape of virtual simulations to supplement in-person manikin and SP-based simulation training. VSPs are gaining acceptance from educators and students alike, marked by the necessities of social distancing during the COVID-19 pandemic and lingering clinical teaching challenges in its wake [36]. Our development pushed the boundary of existing VSP applications that focus on routine clinical interactions with patients, such as history taking [11-15]. Our simulation focuses on the delivery of uncertain (and potentially bad) news that accompanies a diagnostic mammogram requiring additional follow-up. The study by Webb [37] is a notable exception, testing the ability of GPT-3.5 to train emergency room physicians to deliver the news of serious diagnoses to patients. Our work advances the work of Webb [37] and other studies that use text-based communication by simulating verbal communication between users and VSP. This is an important advancement in virtual simulations that allows them to mimic in-person simulations more closely with SPs and real-life clinical interactions with patients.

Reflecting on our move from a BPS design to one using GPT-4, we note the value in the conceptual framework of a BPS design process that is driven by input from clinicians and patients. In-depth interviews with BCSs and PCPs helped us understand patient journeys through the breast cancer screening process. In turn, this helped us generate and map plausible conversations between PCPs and patients so we could plan out the simulation scenario and eventually inform our VSP prompt. Mapping likely conversations also helped us evaluate the performance of the GPT-4–powered VSP to determine if it responded how we expect a patient to respond. We recommend that developers start building virtual scenarios by collecting qualitative data from relevant clinicians and patients in the spirit of a BPS and map out likely conversation paths before developing AI-powered VSP prompts.

A key feature of the training was the ability to display GPT-4–generated written feedback to the learner at the end of the simulation, similar to the qualitative feedback that SPs and clinical instructors provide to medical students during their education. Feedback is a key element of training programs, as underscored by MS4b in noting, "one of the most valuable parts (of SP training) was having the standardized patients give their

XSL•FO
RenderX

feedback at the end." In some ways, GPT-4 outperformed our expectations in providing feedback, especially in the way it captured the overall tone of the conversation (eg, if the user was empathetic), but it also left room for improvement, as discussed subsequently with its limitations. We will retain the feedback feature of the simulation and continue to refine its capabilities commensurate with LLM advances. In addition, future studies comparing AI feedback to expert clinician educator feedback will allow us to further refine prompts. At the current time, our solution is to offer a caveat to learners in this beta phase to consider the feedback and understand that their own clinical judgment may supersede suggested communication techniques.

Limitations

Some of the limitations we encountered will be addressed through prompt refinement, newer LLMs, and faster computing speeds. Future LLMs will likely decrease AI feedback transcription errors and response delays. Preliminary tests showed that using GPT-4o, OpenAI's newest LLM, reduced the delay for the AI feedback agent to generate text down to around 10 to 15 seconds.

Other limitations that are artifacts of the unpredictability of LLMs may persist in future LLMs and will be harder to address. We developed the VSP and AI feedback agent prompts using a heuristic, iterative process that relied on trial and error to improve the quality of simulated conversations and generated feedback. What we identified as improvements we made to our VSP may have been random conversation variations that we happened to prefer and not a direct result of a change we made to the prompt text. The prompts provided guidance to the VSP that typically prevented it from going off topic and out of context or producing nonsensical responses (ie, *hallucinating*) during conversations with users, but unexpected results were observed on several occasions. These errors were difficult to predict, such as when the chatbot shifted roles to play the physician instead of the patient. We also found it difficult to force the VSP to equally value clarity and empathy when determining how to behave cooperatively with the user, as indicated in the Results section.

It was also difficult to determine the criteria used by the AI feedback agent for assessing user performance. For instance, testing the AI feedback agent on the same conversation text multiple times resulted in occasional differences in the assessment of the user's adherence to SPIKES. Furthermore, the AI feedback agent had difficulty weighing the importance of different aspects of the SPIKES protocol, in contrast to the experienced PCP who showed empathy by not pressuring the "anxious" VSP to agree to a biopsy immediately. This reflects GPT-4's limitations in dispensing clinical advice that relies more on clinical experience than can be gleaned from existing documentation [38]. Finally, VSP dialogue and generated feedback are not standardized, in contrast to SP and clinical educators in traditional medical simulation. In its current form, it is unlikely that an LLM-driven VSP training could be used as a part of summative clinical examinations or assessments of learner communication skills due to this lack of standardization.

Limitations highlight the need to evaluate the robustness of the simulation on a larger scale. Researchers have proposed evaluation frameworks such as the automated interactive evaluation framework and Artificial Intelligence Structured Clinical Examinations to evaluate the performance of LLMs to carry out clinical tasks that can be applied in this regard [39,40]. In this vein, the next step will be a pilot test of the virtual simulation module with medical students to determine its feasibility and acceptability. The unpredictability of LLMs also offers opportunities to revolutionize and transform "standardized testing" and to question the notion that standardized tests are the best way to evaluate competency when no real patient is the same as the previous or next patient. LLMs provide a mechanism to evaluate competency through a variety of patients and different types of interactions to provide a more holistic view of proficiency in communication skills.

Future Directions

We scratched the surface for the types of conversations that PCPs and other clinicians have with their patients that could benefit from communication skills training through a virtual simulation with a VSP. We focused on training for medical students, but the scenario and the type of feedback users receive could be tailored for learners at different stages of their education, such as medical students versus residents, or in different professions, such as nurses, social workers, and administrative clerks interacting with patients along any aspect of the patient's clinical experience. Our simulated conversation focused on diagnostic uncertainty surrounding breast cancer screenings. Conversations for all life-changing diagnoses and stigmatizing health conditions (eg, HIV, substance use disorders, and obesity) and associated screening tests warrant careful consideration, even for common diagnoses. PCP2 discussed how type 2 diabetes diagnoses are routine for PCPs but are viewed by some patients as a "death sentence." Aside from the diagnoses themselves, changing treatment plans can be unsettling for patients. For example, PCP4 mentioned a challenging conversation as one "where it goes against what a patient wants," such as discontinuing pain medication prescriptions.

Our simulation focused on a single conversation between a learner and a VSP within the context of other clinical conversations, such as the one the VSP had with the radiologist before speaking with the learner in the simulation. During interviews, PCPs shared the importance of setting patient expectations within the context of coordinated care; for example, "We have other people who are going to be involved" (PCP4). This was reflected in BCS's discussions that highlighted the need for improved communication with different types of clinicians throughout their patient journey, such as radiology technicians and radiologists. Interprofessional education can improve coordinated care and communication with patients but is hampered by the same type of clinical demands that limit individual and team communication skills training. Virtual simulations can lower training barriers, such as the need for multiple learners to interact in person. For example, Liaw et al [41] developed a simulation where multiple learners in different locations can interact in the same simulation to practice care coordination for an older VSP. In that study, debriefing was not automatic and instead was facilitated by clinician educators. Future iterations could test the limits of LLMs by incorporating

XSL•FO
RenderX

multiplayer characters involving >2 people in a dialogue, for example, adding a family member or having 2 health care professionals in a patient encounter.

Another consideration is the decision to use a cloud-based LLM versus LLMs that can be downloaded to local computers. Our simulation incorporated GPT-4, a cloud-based LLM that ran on standard computers relative to local LLMs that would have required additional computing resources. While there are concerns about the privacy and security of data sent to cloud-based LLMs, our prompt uses a fictional patient and a role-playing scenario to minimize the risk of any sensitive or private data being shared with the chatbot. We wrote the prompt for our fictional patient Olivia based loosely on formative interviews with anonymous patients and did not include any protected patient information or restricted data. We intend to notify participants in the pilot study about the potential data privacy risks associated with GPT-4 and will request that they acknowledge and agree not to input any protected or private information. Future studies could incorporate secure local LLMs trained on anonymized transcripts of conversations between clinicians and patients to mitigate the potential data privacy risks of using cloud-based LLMs.

Conclusions

Our work contributes to a rapidly changing medical simulation landscape driven by advancements in LLMs that use generative AI algorithms to mimic human responses to text and voice queries. Our project showcases 2 promising applications for GPT-4 in its ability to streamline the development of a simulated phone call between a learner and VSP and provide accurate AI-generated feedback to the learner. While GPT-4 displayed limitations in its ability to provide nuanced feedback about learners' performance in following best communication practices, the training simulation consistently performed its main task well by providing an asynchronous opportunity for medical students to practice a challenging conversation with a patient. Because every attempt elicits slightly varied responses, it is possible to have multiple opportunities for deliberate practice, unlike traditional branching path scenarios that have smaller, more finite correct answers. Given the rapid advances in LLM to date, we are encouraged about the potential to improve our current training simulation with future LLM improvements and produce more complex scenarios. OpenAI is already proclaiming that GPT-5 will have "PhD level intelligence" [42]. While we cannot comment on the validity of that statement, we feel confident proclaiming that the future of medical simulation is bright.

Acknowledgments

This work was supported by a National Cancer Institute administrative supplement through the Jonsson Comprehensive Cancer Center at the University of California, Los Angeles (3P30CA016042-47S3). PAG was supported in part by the Breast Cancer Research Foundation.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
In-depth interview guide questions for medical students, primary care physicians, and breast cancer survivors who participated in the study to discuss the benefits and limitations of current communication skills training practices, interest in virtual communication skills training, and current modes of communication between primary care physicians and their patients to discuss mammography screening examination results.
[DOCX File , 17 KB - formative_v9i1e65670_app1.docx]

References

1. Bair H, Norden J. Large language models and their implications on medical education. Acad Med 2023 Aug 01;98(8):869-870. [doi: 10.1097/ACM.0000000000005265] [Medline: 37162220]
2. Haruna-Cooper L, Rashid MA. GPT-4: the future of artificial intelligence in medical school assessments. J R Soc Med 2023 Jun;116(6):218-219 [FREE Full text] [doi: 10.1177/01410768231181251] [Medline: 37318843]
3. Wójcik S, Rulkiewicz A, Pruszczyk P, Lisik W, Poboży M, Domienik-Karłowicz J. Beyond ChatGPT: what does GPT-4 add to healthcare? The dawn of a new era. Cardiol J 2023;30(6):1018-1025 [FREE Full text] [doi: 10.5603/cj.97515] [Medline: 37830256]
4. Chen S, Guevara M, Moningi S, Hoebers F, Elhalawani H, Kann BH, et al. The effect of using a large language model to respond to patient messages. Lancet Digit Health 2024 Jun;6(6):e379-e381. [doi: 10.1016/s2589-7500(24)00060-8]

XSL•FO
RenderX

5. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023 Jun 01;183(6):589-596 [FREE Full text] [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
6. Ha JF, Longnecker N. Doctor-patient communication: a review. Ochsner J 2010;10(1):38-43 [FREE Full text] [Medline: 21603354]
7. Tavakoly Sany SB, Behzhad F, Ferns G, Peyman N. Communication skills training for physicians improves health literacy and medical outcomes among patients with hypertension: a randomized controlled trial. BMC Health Serv Res 2020 Jan 23;20(1):60 [FREE Full text] [doi: 10.1186/s12913-020-4901-8] [Medline: 31973765]
8. Adnan AI. Effectiveness of communication skills training in medical students using simulated patients or volunteer outpatients. Cureus 2022 Jul 10;14(7):e26717 [FREE Full text] [doi: 10.7759/cureus.26717] [Medline: 35967150]
9. Reidy JA, Clark MA, Berman HA, Chan SH, Gawande AA, Streid J, et al. Paving the way for universal medical student training in serious illness communication: the Massachusetts Medical Schools' Collaborative. BMC Med Educ 2022 Sep 01;22(1):654 [FREE Full text] [doi: 10.1186/s12909-022-03702-2] [Medline: 36050708]
10. Deveugele M, Derese A, De Maesschalck S, Willems S, Van Driel M, De Maeseneer J. Teaching communication skills to medical students, a challenge in the curriculum? Patient Educ Couns 2005 Sep;58(3):265-270. [doi: 10.1016/j.pec.2005.06.004] [Medline: 16023822]
11. Holderried F, Stegemann-Philipps C, Herschbach L, Moldt JA, Nevins A, Griewatz J, et al. A generative pretrained transformer (GPT)-powered chatbot as a simulated patient to practice history taking: prospective, mixed methods study. JMIR Med Educ 2024 Jan 16;10:e53961 [FREE Full text] [doi: 10.2196/53961] [Medline: 38227363]
12. Holderried F, Stegemann-Philipps C, Herrmann-Werner A, Festl-Wietek T, Holderried M, Eickhoff C, et al. A language model-powered simulated patient with automated feedback for history taking: prospective study. JMIR Med Educ 2024 Aug 16;10:e59213 [FREE Full text] [doi: 10.2196/59213] [Medline: 39150749]
13. Jiang Z, Huang X, Wang Z, Liu Y, Huang L, Luo X. Embodied conversational agents for chronic diseases: scoping review. J Med Internet Res 2024 Jan 09;26:e47134 [FREE Full text] [doi: 10.2196/47134] [Medline: 38194260]
14. Potter L, Jefferies C. Enhancing communication and clinical reasoning in medical education: building virtual patients with generative AI. Future Healthc J 2024 Apr;11:100043 [FREE Full text] [doi: 10.1016/j.fhj.2024.100043]
15. Sardesai N, Russo P, Martin J, Sardesai A. Utilizing generative conversational artificial intelligence to create simulated patient encounters: a pilot study for anaesthesia training. Postgrad Med J 2024 Mar 18;100(1182):237-241. [doi: 10.1093/postmj/qgad137] [Medline: 38240054]
16. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: can language models be too big? ???? In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. 2021 Presented at: FAccT '21; March 3-10, 2021; Virtual Event. [doi: 10.1145/3442188.3445922]
17. Cooper A, Rodman A. AI and medical education - a 21st-century Pandora's box. N Engl J Med 2023 Aug 03;389(5):385-387 [FREE Full text] [doi: 10.1056/NEJMp2304993] [Medline: 37522417]
18. Zack T, Lehman E, Suzgun M, Rodriguez JA, Celi LA, Gichoya J, et al. Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study. Lancet Digit Health 2024 Jan;6(1):e12-e22. [doi: 10.1016/s2589-7500(23)00225-x]
19. Liu X, Wu C, Lai R, Lin H, Xu Y, Lin Y, et al. ChatGPT: when the artificial intelligence meets standardized patients in clinical training. J Transl Med 2023 Jul 06;21(1):447 [FREE Full text] [doi: 10.1186/s12967-023-04314-0] [Medline: 37415217]
20. McGovern MM, Johnston M, Brown K, Zinberg R, Cohen D. Use of standardized patients in, undergraduate medical genetics education. Teach Learn Med 2006;18(3):203-207. [doi: 10.1207/s15328015tlm1803_3] [Medline: 16776606]
21. Nelson G. Training standardized patients to provide effective feedback: development, implementation, and its effect on the efficacy of medical students' education. S D Med 2022 Oct;75(10):454-455. [Medline: 36889265]
22. Li Y, Zeng C, Zhong J, Zhang R, Zhang M, Zou L. Leveraging large language model as simulated patients for clinical education. arXiv. Preprint posted online on April 13, 2024 [FREE Full text]
23. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin 2023 Jan 12;73(1):17-48 [FREE Full text] [doi: 10.3322/caac.21763] [Medline: 36633525]
24. Ho TQ, Bissell MC, Kerlikowske K, Hubbard RA, Sprague BL, Lee CI, et al. Cumulative probability of false-positive results after 10 years of screening with digital breast tomosynthesis vs digital mammography. JAMA Netw Open 2022 Mar 01;5(3):e222440 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.2440] [Medline: 35333365]
25. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature 2023 Aug 12;620(7972):172-180 [FREE Full text] [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
26. Bojic L, Kovacevic P, Cabarkapa M. GPT-4 surpassing human performance in linguistic pragmatics. arXiv. Preprint posted online on December 15, 2023 [FREE Full text]
27. Nori H, King N, McKinney SM, Carignan D, Horvitz E. Capabilities of GPT-4 on medical challenge problems. arXiv. Preprint posted online on March 20, 2023 [FREE Full text]

XSL•FO
RenderX

28. Oh N, Choi GS, Lee WY. ChatGPT goes to the operating room: evaluating GPT-4 performance and its potential in surgical education and training in the era of large language models. Ann Surg Treat Res 2023 May;104(5):269-273 [FREE Full text] [doi: 10.4174/astr.2023.104.5.269] [Medline: 37179699]
29. Rosoł M, Gąsior JS, Łaba J, Korzeniewski K, Młyńczak M. Evaluation of the performance of GPT-3.5 and GPT-4 on the Polish medical final examination. Sci Rep 2023 Nov 22;13(1):20512 [FREE Full text] [doi: 10.1038/s41598-023-46995-z] [Medline: 37993519]
30. Gallifant J, Afshar M, Ameen S, Aphinyanaphongs Y, Chen S, Cacciamani G, et al. The TRIPOD-LLM reporting guideline for studies using large language models. Nat Med 2025 Jan 08;31(1):60-69. [doi: 10.1038/s41591-024-03425-5] [Medline: 39779929]
31. HyperSkill for VR training. SimInsights. URL: https://www.siminsights.com/hyperskill/ [accessed 2025-04-07]
32. Baile WF, Buckman R, Lenzi R, Glober G, Beale EA, Kudelka AP. SPIKES-a six-step protocol for delivering bad news: application to the patient with cancer. Oncologist 2000;5(4):302-311 [FREE Full text] [doi: 10.1634/theoncologist.5-4-302] [Medline: 10964998]
33. Buckman RA. Breaking bad news: the S-P-I-K-E-S strategy. Community Oncol 2005 Mar;2(2):138-142. [doi: 10.1016/S1548-5315(11)70867-1]
34. Ruan J, Chen Y, Zhang B, Xu Z, Bao T, Du G, et al. TPTU: large language model-based AI agents for task planning and tool usage. arXiv. Preprint posted online on August 7, 2023 [FREE Full text]
35. Rabow MW, McPhee SJ. Beyond breaking bad news: how to help patients who suffer. West J Med 1999 Oct;171(4):260-263 [FREE Full text] [Medline: 10578682]
36. De Mattei L, Morato MQ, Sidhu V, Gautam N, Mendonca CT, Tsai A, et al. Are artificial intelligence virtual simulated patients (AI-VSP) a valid teaching modality for health professional students? Clin Simul Nurs 2024 Jul;92:101536. [doi: 10.1016/j.ecns.2024.101536]
37. Webb J. Proof of concept: using ChatGPT to teach emergency physicians how to break bad news. Cureus 2023 May;15(5):e38755 [FREE Full text] [doi: 10.7759/cureus.38755] [Medline: 37303324]
38. Jo E, Song S, Kim JH, Lim S, Kim JH, Cha JJ, et al. Assessing GPT-4's performance in delivering medical advice: comparative analysis with human experts. JMIR Med Educ 2024 Jul 08;10:e51282 [FREE Full text] [doi: 10.2196/51282] [Medline: 38989848]
39. Liao Y, Meng Y, Wang Y, Liu H, Wang Y, Wang Y. Automatic interactive evaluation for large language models with state aware patient simulator. arXiv. Preprint posted online on March 13, 2024 [FREE Full text] [doi: 10.2139/ssrn.4890649]
40. Mehandru N, Miao BY, Almaraz ER, Sushil M, Butte AJ, Alaa A. Evaluating large language models as agents in the clinic. NPJ Digit Med 2024 Apr 03;7(1):84 [FREE Full text] [doi: 10.1038/s41746-024-01083-y] [Medline: 38570554]
41. Liaw SY, Wu LT, Soh SL, Ringsted C, Lau TC, Lim WS. Virtual reality simulation in interprofessional round training for health care students: a qualitative evaluation study. Clin Simul Nurs 2020 Aug;45:42-46. [doi: 10.1016/j.ecns.2020.03.013]
42. Dane L. GPT-5 will have Ph.D level intelligence. Lindane.Co. 2024 Jul 3. URL: https://lindane.co/blog/gpt-5-will-have-phd-level-intelligence/ [accessed 2025-04-06]

Abbreviations

ABCDE: advance preparation, build a therapeutic environment or relationship, communicate well, deal with patient and family reactions, and encourage and validate emotions
AI: artificial intelligence
BCS: breast cancer survivor
BPS: branching path simulation
LLM: large language model
PCP: primary care physician
SP: standardized patient
SPIKES: Setting, Perception, Invitation, Knowledge, Emotion, and Summarize
TRIPOD: Transparent Reporting of a Multivariable Model for Individual Prognosis or Diagnosis
VSP: virtual simulated patient

XSL•FO
RenderX

Edited by A Mavragani; submitted 27.08.24; peer-reviewed by S Chan, AJ Doyle; comments to author 13.12.24; revised version received 21.01.25; accepted 14.02.25; published 17.04.25.

Please cite as:
Weisman D, Sugarman A, Huang YM, Gelberg L, Ganz PA, Comulada WS
Development of a GPT-4–Powered Virtual Simulated Patient and Communication Training Platform for Medical Students to Practice Discussing Abnormal Mammogram Results With Patients: Multiphase Study
JMIR Form Res 2025;9:e65670
URL: https://formative.jmir.org/2025/1/e65670
doi:10.2196/65670
PMID:

©Dan Weisman, Alanna Sugarman, Yue Ming Huang, Lillian Gelberg, Patricia A Ganz, Warren Scott Comulada. Originally published in JMIR Formative Research (https://formative.jmir.org), 17.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

User Experience of and Adherence to a Smartphone App to Maintain Behavior Change and Self-Management in Patients With Work-Related Skin Diseases: Multistep, Single-Arm Feasibility Study

Nele Ristow[1,2], MEd; Annika Wilke[1,2], PD Dr; Christoph Skudlik[1,2], Prof Dr Med; Swen Malte John[1,2], Prof Dr Med; Michaela Ludewig[3], Prof Dr

[1]Department of Dermatology, Environmental Medicine and Health Theory, Institute for Health Research and Education, Osnabrück University, Osnabrück, Germany

[2]Institute for Interdisciplinary Dermatological Prevention and Rehabilitation (iDerm) at the Osnabrück University, Osnabrück, Germany

[3]Department Health Sciences, Hochschule Bochum, Bochum, Germany

Corresponding Author:
Nele Ristow, MEd
Department of Dermatology, Environmental Medicine and Health Theory
Institute for Health Research and Education
Osnabrück University
Am Finkenhügel 7a
Osnabrück, 49076
Germany
Phone: 49 541 969 7410
Fax: 49 541 969 2445
Email: neristow@uos.de

Abstract

Background: Smartphone apps are a growing field supporting the prevention of chronic diseases. The user experience (UX) is an important predictor of app use and should be considered in mobile health research. Long-term skin protection behavior is important for those with work-related skin diseases. However, altering health behavior is complex and requires a high level of self-management. We developed a maintenance program consisting of the Mein Hautschutz im Alltag (MiA; "My skin protection in everyday life") app combined with an individual face-to-face goal-setting interview to support patients in the implementation of skin protection behavior after inpatient rehabilitation.

Objective: The objectives of this paper are to (1) describe the intervention in a standardized manner; (2) evaluate the UX, subjective quality, and perceived impact of the MiA app; and (3) evaluate the adherence to the MiA app.

Methods: We followed a user-centered and multistage iterative process in 2 steps that combined qualitative and quantitative data. The maintenance program was tested over 12 weeks after discharge from rehabilitation. The UX, subjective quality, and perceived impact were evaluated formatively based on the user version of the Mobile Application Rating Scale after 12 weeks (T2). Adherence was measured using the frequency of interactions with the app.

Results: In total, 42 patients took part (with a dropout rate of n=18, 43% at T2). The average age was 49.5 (SD 13.1) years, and 57% (24/42) were male. We found high ratings for the UX, with an average score of 80.18 (SD 8.94) out of a theoretical maximum of 100, but there were a few exceptions in the usability and interaction with the app. The app was most frequently rated with 4 out of 5 stars (15/24, 65%), which indicates a high subjective quality. Furthermore, the app seemed to influence important determinants to implement skin protection behavior. Adherence to skin protection tracking was higher over the study period than adherence to skin documentation and goal assessment. The number of adherent participants to skin protection tracking was higher in the skin care and skin cleansing categories (28/42, 67% each) compared to the skin protection category (13/42, 31%) on day 1 and decreased until day 84 in all dimensions (12/42, 29% each for skin care and skin cleansing; 9/42, 21% for skin protection).

Conclusions: The results in terms of adherence met the expectations and were consistent with those of other studies evaluating the use of apps for chronic diseases. Interaction with the app could be increased using artificial intelligence to determine eczema severity via photos. It should be investigated which subgroups have difficulties with usability to individualize the support to a

XSL•FO
RenderX

greater degree during onboarding. There is a need for further research regarding the effectiveness of the MiA app on skin protection behavior, quality of life, and eczema severity.

(JMIR Form Res 2025;9:e66791) doi:10.2196/66791

KEYWORDS

user experience; mobile health; mHealth; app; smartphone; complex intervention; Template for Intervention Description and Replication; behavior change techniques; behavior change; skin diseases; occupational dermatology; artificial intelligence

Introduction

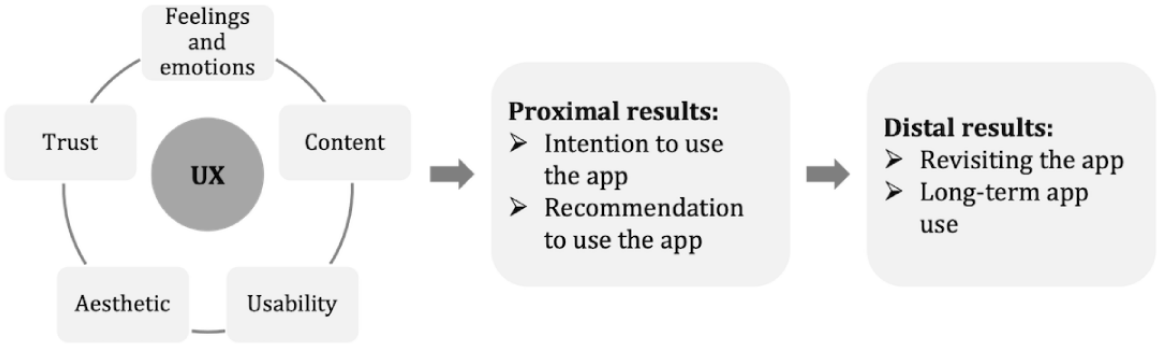
Background

Smartphone apps have recently become a growing field in the health care sector, garnering increasing attention and importance in health research. The use of such technologies can help optimize broad and location-independent health care in real time and support the prevention of chronic diseases and health promotion to improve adherence to therapy, quality of life, and clinical outcomes in the long term [1,2]. Smartphone apps not only help monitor health data [3,4] but also help change health behavior and develop new habits. For this purpose, health behavior apps offer the potential to technically integrate a variety of behavior change techniques (BCTs) [1,5,6]. BCTs are based on health psychology knowledge about the mechanisms of behavior change processes [7]. The most common BCTs in behavior change apps are, for example, self-monitoring by tracking the duration and frequency of a behavior and feedback on the behavior or outcomes by visualizing the progress graphically [1,5,6,8]. Feedback and the visualization of success also increase motivation in the users, which plays an important role in the long-term success of behavior change [5]. Behavior change apps can also be used to set goals and monitor their achievement. Automatic reminders provide prompts and support the development of routines in everyday life. Written, visual, and audio information can show users how to perform the behavior [1,5,6,8].

Relevance of User Experience

A known and widely discussed problem in research on behavior change apps is a low adherence and engagement when using these technologies [2,9-11]. To address these problems, user experience (UX) is becoming increasingly important in mobile health research [12-15]. UX is an overarching holistic concept and covers all emotional, cognitive, and physical experiences that users undergo when interacting with software products such as apps [12,16-18]. As shown in Figure 1, a high UX when using an app for the first time is associated with the willingness to continue using it and recommend it to others in the short term. A high UX leads to increased adherence and app use in the long term and, thus, can also improve clinical outcomes and the effectiveness of health-related interventions [14,19-21]. Therefore, the investigation of UX in the context of digital technologies in patient care is of great importance and should be considered early during the development process [14,15,22]. Thus, a thorough formative evaluation of the UX should precede a summative evaluation to optimize the app from the end user's perspective [14].

Figure 1. Theoretical model to explain the influence of user experience (UX) and its 5 dimensions—content, usability, aesthetic, trust, and feelings and emotions—on the intention and recommendation to use the app (proximal) results, as well as the revisiting the app and long-term use of the app (distal) results as the basis for this study.



As a multidimensional construct, UX comprises various dimensions that can differ depending on the underlying model, understanding, and context [15,17,19]. According to Thielsch and Salaschek [21], the dimensions *content*, *usability*, and *aesthetics* are particularly important in a research-related context. In addition, *feelings and emotions* that arise through interaction with the technical medium are described as an important aspect of UX [15,16]. As personal data are collected, passed on, and stored in digital care structures, *trust* in the technology is another predictor of future app use [21].

Behavior Change and Self-Management Apps for Patients With Chronic Skin Diseases

Apps to support behavior change and disease-specific self-management have also been developed and evaluated in the context of chronic skin diseases. These are usually aimed at specific groups, such as parents of children with atopic dermatitis (AD) [23,24]; patients with AD [25,26]; or people with specific areas of the skin affected, such as hand and foot eczema [27]. Weigandt et al [27] developed the first smartphone app in Germany specifically for hand and foot eczema, with

XSL•FO
RenderX

which patients can monitor and manage their disease following a patient educational intervention by photographing their skin, tracking their quality of life and symptoms, and chatting with their dermatologist.

However, thus far, such technology does not exist for work-related skin diseases (WRSDs) even though they have been among the most common work-related diseases in Germany for decades [28]. Up to 90% of all diagnosed cases affect the hands in the form of irritant or allergic contact eczema, which often occurs in combination with a genetic predisposition to AD [29-32]. Work-related hand eczema is usually caused by hazardous activities in the workplace (eg, skin contact with allergens and irritants, wet work, high handwashing frequencies, and mechanical stress). Workers in the health care, metalworking, hairdressing, and construction sectors are at a particularly high risk, which results in high prevalence rates [30,31]. Due to the high relevance of WRSDs, the German Social Accident Insurance has developed a complex and hierarchical multistep procedure for patients with WRSDs, with different outpatient and inpatient prevention measures depending on the severity of the skin disease. In the case of severe or recurrent skin diseases, patients are offered to participate in an inpatient interprofessional rehabilitation program with a duration of 3 weeks. In German occupational dermatology, this program is also known as tertiary individual prevention (TIP) [29,33]. An important element of TIP are health educational interventions [34] with the aim of gaining disease-specific knowledge and increasing the motivation to implement and optimize individual skin protection behavior. In addition, individual strategies for skin protection behavior are developed with occupational therapists and practiced as part of a workplace simulation in occupational therapy. To restore the skin barrier and reduce the risk of recurrence of hand eczema, it is important to implement skin protection behavior (eg, regular use of skin protection and skin care products as well as reducing the frequency of handwashing by using mild detergents) in the long term [34-36]. However, changing skin protection behavior is complex and requires a high level of self-management after participating in the TIP as the new behavior must be successfully transferred to and implemented in the professional and private contexts. A structured maintenance program that supports patients in this subsequent implementation of skin protection behavior does not yet exist [34]. To fill this gap, we developed the Mein Hautschutz im Alltag (MiA) app (MiA translates to “My skin protection in everyday life”) to support the self-management of patients with WRSDs after the TIP [37].

Objectives of This Study

The aim of this study was to evaluate the UX of and adherence to the MiA app and pilot the feasibility of an app-based maintenance program in our clinical setting. In this publication, we report the following results: (1) description of the intervention in a standardized manner; (2) UX, subjective quality, and perceived impact of the MiA app; and (3) adherence to the MiA app.

Methods

Overview

The occupational dermatology maintenance program is a complex intervention consisting of several interacting components (Multimedia Appendix 1). According to the Medical Research Council’s framework, the development of complex interventions is based on a 4-stage process consisting of development [37], feasibility, evaluation, and implementation [38]. This study focuses on feasibility (phase 2).

Intervention

The maintenance program is offered by the Institute for Interdisciplinary Dermatological Prevention and Rehabilitation (iDerm), Osnabrück, Germany, which is a specialized center in Germany for inpatient and outpatient interprofessional treatment of patients with WRSDs.

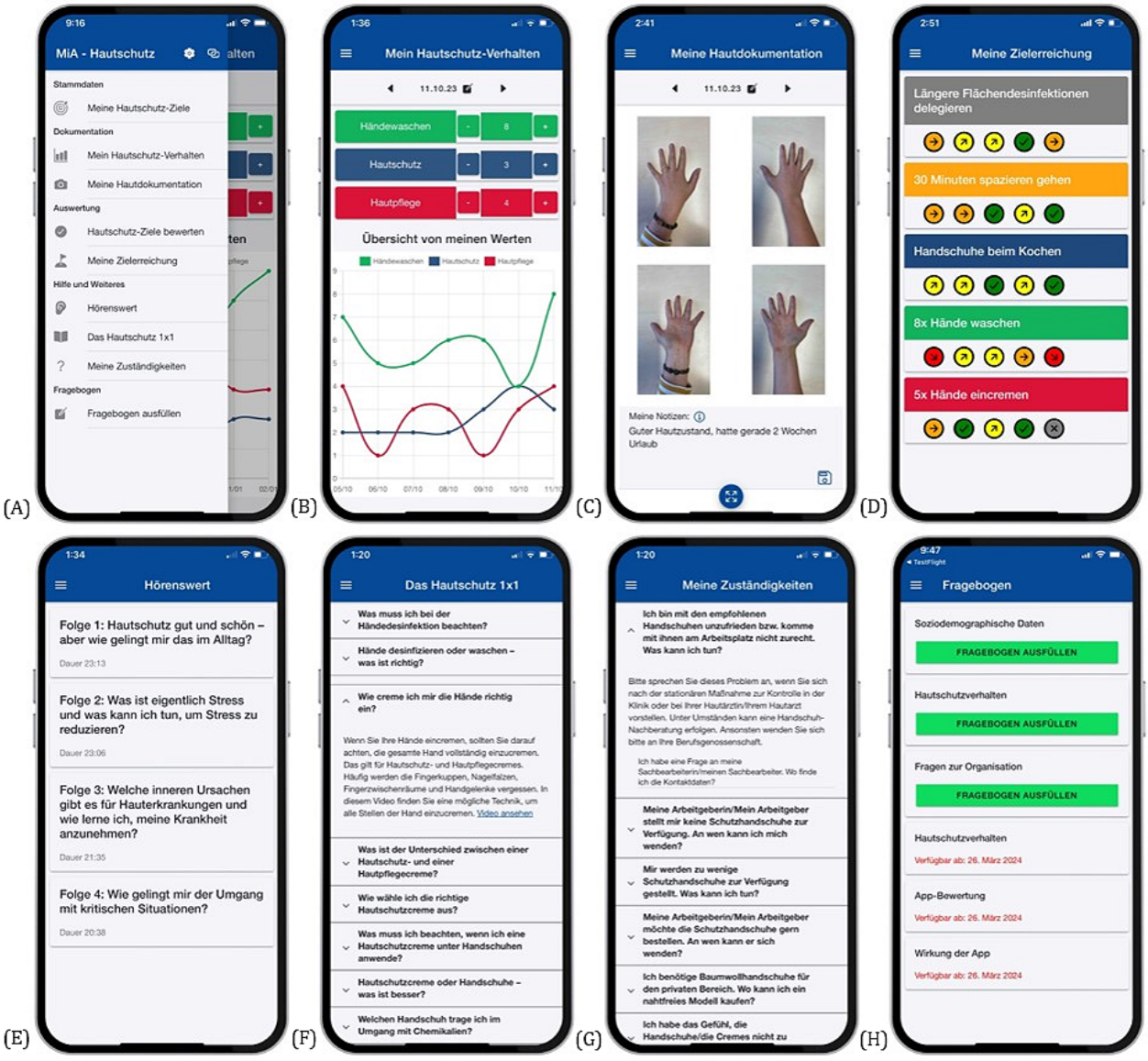
The maintenance program consists of 2 main elements: *individual goal-setting interview* and the MiA app. The systematic development of these elements is described in detail elsewhere [37].

The *individual goal-setting interview* is conducted during the inpatient stay by a health educator face-to-face 5 or 6 days before discharge. The goals, which are defined by the patients, are entered into the app on the coaching platform by the health educator and can be viewed by patients via the app. During the subsequent onboarding, patients receive an introduction and explanations of the various app functions. Patients are also given access to a video via a QR code that shows the app with screenshots and explains the app’s functions.

MiA is a fully automated app with free access for study participants and consists of 6 components focusing either on interaction or information. The smartphone-based components with interactive nature are named My Skin Protection Goals (German: *Meine Hautschutz-Ziele*), My Skin Protection Behavior (German: *Mein Hautschutz-Verhalten*), and My Skin Documentation (German: *Meine Hautdokumentation*). The smartphone-based components To Listen (German: *Hörenswert*), Skin Protection 101 (German: *Das Hautschutz 1x1*), and My Accountabilities (German: *Meine Zuständigkeiten*) provide different information about the skin disease in different modes of delivery (podcasts, textboxes, and videos; Figure 2). In addition, users automatically receive a reminder message if no information has been entered into the app for a period of 14 days.

XSL•FO
RenderX

Figure 2. (A) Overview of the menu; (B) overview of the My Skin Protection Behavior component with the option to enter the frequencies of skin cleansing (green line) and applications of skin care cream (red line) and skin protection cream (blue line); (C) overview of the My Skin Documentation component, which allows for taking pictures of the front and back side of the right and left hand that can be enlarged and to which personal notes can be added; (D) overview of the My Skin Protection Goals component with the overview of the goal achievement; (E) To Listen component and 4 podcast episodes; (F) overview of the Skin Protection 101 component with different questions and an example unfolded answer; (G) overview of the My Accountabilities component with different questions and an exemplary unfolded answer; and (H) presentation of the research questionnaire with the 6 categories, whereby categories 1 to 3 (sociodemographic data, skin protection behavior, and questions about the organization) can be completed immediately and categories 4 to 6 (skin protection behavior, app evaluation, and impact of the app) are activated after 12 weeks of use.



Multimedia Appendix 1 provides a more detailed description of each component according to the Template for Intervention Description and Replication checklist [39,40], including what, why, how often, and when components are unlocked during the maintenance program and where and which tailoring options are available. Table 1 describes the contents of the active components of the intervention in a standardized manner using the Behavior Change Technique Taxonomy version 1 [7,41].

XSL•FO
RenderX

Table 1. Behavior change techniques (BCTs) in the 12-week occupational dermatology maintenance program for improving health behavior change in patients with work-related skin diseases based on the BCT Taxonomy version 1 [7,41].

Component	Description	BCTs
Face-to-face intervention		
Individual goal-setting interview	Formulation of individual skin protection goals	• 1.1: goal setting (behavior)
MiA[a] app		
My Skin Protection Goals	Monitoring of individual skin protection goals and assessment of their achievement	• 1.5: review behavior goals
My Skin Protection Goals	Visualization of goal achievement over time using arrows	• 1.6: discrepancy between current behavior and goal
My Skin Protection Goals	Motivating feedback messages depending on the evaluation of the goal	• 3.1: social support (unspecified)
My Skin Protection Behavior	Tracking of skin protection behavior and monitoring the progress	• 2.3: self-monitoring of behavior
My Skin Documentation	Recording and observing the skin condition	• 2.5: monitoring of outcomes of behavior without feedback
To Listen	Information about the skin disease, associated difficulties, and strategies	• Episode 1: 4.1—instruction on how to perform the behavior and 15.1—verbal persuasion about capability • Episode 2: 5.1—information about health consequences, 5.2—salience of consequences, 8.2—behavior substitution, 12.4—distraction, and 15.1—verbal persuasion about capability • Episode 3: 8.2—behavior substitution, 5.1—information about health consequences, 5.2—salience of consequences, 4.2—information about antecedents, and 15.1—verbal persuasion about capability • Episode 4: 3.1—social support (unspecified), 3.3—social support (emotional), and 15.1—verbal persuasion about capability
Skin Protection 101	Information about seminar content on skin protection, itching, and stress	• 4.1: instruction on how to perform the behavior • 5.1: information about health consequences
My Accountabilities	Information about organizational issues related to care and responsibilities	• 5.3: inform about social and environmental consequences

[a]MiA: Mein Hautschutz im Alltag (My skin protection in everyday life).

Ethical Considerations

This study was approved by the Ethics Committee of Osnabrück University (Ethics-50/2022 and Ethics-15/2023). All patients assessed for eligibility were informed verbally and in writing about the study and its voluntary nature. All participants had to sign a declaration of consent to take part in the study. The participants were informed that data collection would be pseudonymized and that the evaluation and publication of the results would be anonymous. They received no compensation for their participation.

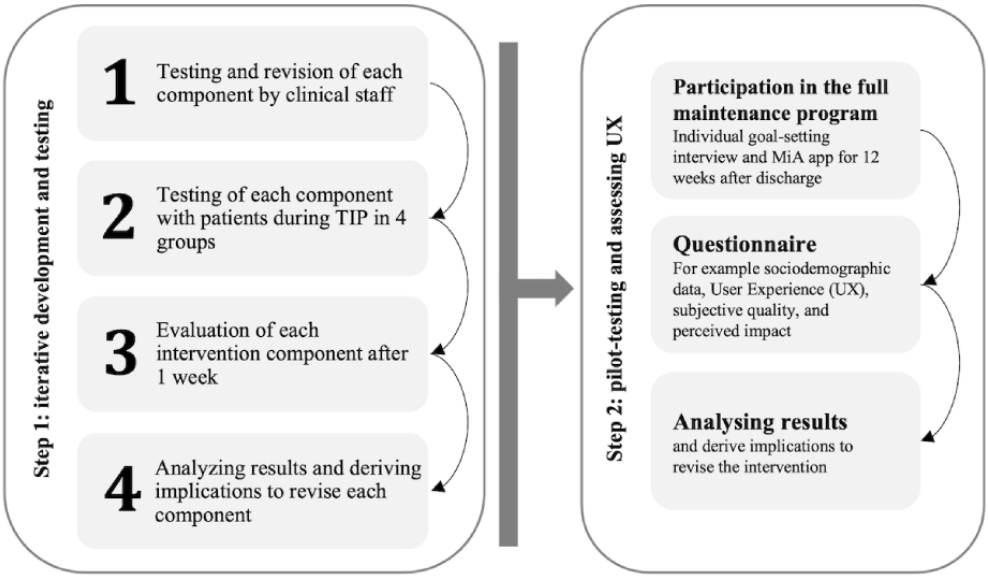
Inclusion Criteria and Study Design

Recruitment took place at iDerm in Osnabrück on the fourth day during the TIP. Study information was provided in patient groups as 7 to 9 new patients are usually admitted to the TIP each week. After all patients had been fully informed about the study, they were given sufficient time to decide whether to participate. Subsequently, interested patients received the written study information and declaration of consent. The inclusion criteria were (1) signed declaration of consent for study participation; (2) legal age (>18 years); (3) sufficient German language skills to understand the app content, participate in focus groups, and complete questionnaires; and (4) access to an internet-enabled smartphone (for step 2).

This study followed a user-centered and multistage iterative approach, which is a well-known and widespread procedure in the development of technologies in health care and UX research [42,43]. This approach allows for the consideration of the opinions and feedback of the target group as future users of the app in a participatory manner [23,44-46]. As illustrated in Figure 3, the entire process consisted of 2 steps with both qualitative and quantitative data collection.

XSL•FO
RenderX

Figure 3. Study process of multistage iterative testing of the app-based maintenance program for patients with work-related skin diseases consisting of (1) iterative development and testing and (2) 12-week testing and piloting. MiA: Mein Hautschutz im Alltag (My skin protection in everyday life); TIP: tertiary individual prevention; UX: user experience.



Instrument

Various instruments have been developed to assess the UX of mobile health apps. Among them, the user version of the Mobile Application Rating Scale (uMARS) is a well-known and reliable instrument that measures the UX of health apps from the user's perspective [47]. The uMARS is based on the Mobile Application Rating Scale, which is used to assess the quality of apps by clinical or technical experts [48,49]. The uMARS is made up of 16 items and comprises the UX dimensions of *engagement*, *functionality*, *aesthetics*, and *information quality*. The scales of subjective quality and perceived impact assess additional information about recommendation; willingness to use and pay for the app; and information regarding awareness and knowledge of and intention for behavior change. For this study, an instrument for step 1 and step 2 was developed based on the uMARS and the model shown in Figure 1.

Statistical analyses of the quantitative data were carried out using SPSS (version 28; IBM Corp). We calculated a score for all 5 UX dimensions as well as a score for all UX items. Adherence to the My Skin Protection Behavior, My Skin Documentation, and My Goal Achievement functions was determined based on the adherence values in Multimedia Appendix 1 and by dichotomizing the data into *adherent* and *nonadherent*. For continuous data (eg, age and UX), we calculated the mean, median, IQR, and SD. We present categorical data (eg, gender, subjective quality, and perceived impact) in frequencies and percentages.

Step 1: Iterative Development and Testing of Intervention Components

Overview

In the first step, the maintenance program was tested by 4 different patient groups during participation in the TIP at the iDerm in Osnabrück between November 2022 and November 2023. Patients were recruited on the fourth day of the TIP with a subsequent test phase of 2 weeks. During the test phase, patients were asked to document their overall impression as well as positive aspects and aspects requiring improvement of the tested component in a standardized documentation form. The documentation forms were not evaluated by the researchers as they merely served as preparation for the patients themselves for the focus group discussions. Each group tested and evaluated specific maintenance components as follows: group 1 evaluated *To listen*, group 2 evaluated the *face-to-face goal setting interview*, group 3 evaluated My Skin Protection Behavior and My Accountabilities on the app, and group 4 evaluated Skin Protection 101 and My Skin Documentation on the app.

Methods of Data Collection in Step 1

After the test phase, the patients completed a short quantitative questionnaire that assessed their overall impression of the maintenance components tested using selected items from the uMARS. The focus of step 1 was qualitative data collected via 4 subsequent focus group discussions. The aim of the focus group discussions was to record positive and negative aspects of the tested components, identify problems and challenges regarding their use, and jointly develop opportunities for improvements.

The focus group discussions were conducted by 2 moderators (NR and ML) based on guiding questions regarding content-related, methodological, and technical aspects. The guiding questions served as orientation. Additions to or deepening of the questions and topics that arose in the course of the discussions were permitted. The interviews were recorded using an audio recorder (Olympus LS-P4 linear pulse-code modulation recorder). Knowledge mapping was chosen for data analysis as it allows for the combination of data collection and data processing [50]. The results were written down in keywords on moderation cards and clustered on a metaplan board. The participants validated the results for completeness and correctness for each topic. In addition, experiences from the test phase and key results were documented by a research

XSL•FO
RenderX

assistant in a protocol. Aspects of the components that needed to be improved were then revised in the next iteration.

Step 2: Pilot-Testing and Assessing the UX

Overview

In the second step, the entire maintenance program, consisting of the face-to-face goal-setting interview and the MiA app with all its components, was tested by patients and formatively evaluated. Recruitment took place over a period of 10 weeks from January 2024 to March 2024. Patients who voluntarily agreed to participate were then given an appointment for the goal-setting interview as a one-to-one meeting and the subsequent onboarding to the app in the third week. The app was tested over a period of 12 weeks after discharge from the TIP, with 2 additional weeks to fill out the second questionnaire. Patients who did not complete the questionnaire within these 2 weeks received a reminder in the form of a paper questionnaire with a prepaid return envelope by post.

Methods of Data Collection in Step 2

Overview

The evaluation was carried out using a quantitative questionnaire consisting of 6 blocks. Blocks 1 to 3 were available immediately after registration, and blocks 4 to 6 were activated after the maintenance period of 12 weeks after discharge. In total, the instrument consisted of 85 questions about sociodemographic data; skin protection behavior before and after the rehabilitation program; and evaluation of the UX, subjective quality, and perceived effects, as well as questions about the structural, process, and outcome quality of the intervention.

In addition to closed-ended questions, the questionnaire also included open-ended questions to allow the participants to provide further details about the individual functions beyond the closed-ended questions [15]. The comprehensibility of the questions and the time required for completion were determined in a pretest with 16 patients before the study. In this publication, we report the results on sociodemographic data, UX evaluation, subjective quality, and perceived impact of the intervention.

UX Evaluation

To assess the UX, the relevant items of the uMARS were assigned to the UX dimensions shown in Figure 1 and supplemented with additional aspects as required. We measured feelings and emotions using 5 items (eg, “I find the app entertaining” or “The app offers me enough options to customise it to my personal requirements and needs”), content using 4 items (eg, “Is the content of the app relevant to you?” or “Are the texts understandable?”), usability using 6 items (eg, “How well and quickly do the app’s applications [buttons, menus] respond?” or “How easy was it for you to learn how to use the app?”), aesthetic using 2 items (“How high are the quality and resolution of the images and texts?” and “How would you rate the overall appearance of the app?”), and trust using 3 items (eg, “The information in the app appears to come from a credible source” and “Are you concerned that data you have entered into the app could be uploaded to third parties?”). All questions were answered on 5-point Likert scales that differed in their content depending on the dimension and item.

Subjective Quality and Perceived Impact

Both aspects were assessed using the uMARS [47]. Subjective quality was measured using the items related to recommendation, willingness to use in the following 12 months, and willingness to pay and overall rating by stars.

The perceived impact scale comprised 5 closed-ended questions on awareness and knowledge of and motivation for skin protection behavior; encouragement to seek for support; and enabling skin protection behavior on a 5-point Likert scale and a supplementary open-ended question on which functions were particularly helpful for implementing skin protection behavior.

Adherence

An important result for this study was the analysis of the frequency of use of individual components as this represented a central criterion for adherence to the intervention. We used a simple definition according to Donkin et al [51] and understood adherence as “the degree to which the user followed the programme as it was designed” [51]. This definition has already been applied in the context of other digital health interventions [45]. In this study, adherence was determined by the frequency with which the skin protection behavior was entered, how often photos were taken, and how often the goals were assessed. Adherence was high if the component My Skin Protection Behavior was used daily and the functions My Skin Documentation and My Skin Protection Goals were used weekly.

Results

Step 1

Study Participants

A total of 23 patients took part in the testing of the intervention components in step 1, of whom 10 (43%) were female and 13 (57%) were male. The average age of the participants was 51 (SD 11; range 27-62) years.

The patients worked in health care (8/23, 35%), the metalwork industry (4/23, 17%), hairdressing (3/23, 13%), construction (2/23, 9%), or other professions (6/23, 26%).

Main Results

The results of step 1 were used for iterative development to further optimize the intervention for phase 2. For example, in the My Skin Protection Behavior function, the scale of the days and frequencies in the diagram was too small. This was adjusted and enlarged in the revision. In the My Skin Documentation function, every entry was automatically saved in the comment function. This always hindered further input for a few seconds. A button was added so that users could save their comments at the end by themselves. Patients also highlighted that the app had a simple structure and reminded them of apps they already knew, had a good usability and visual presentation, and was factual without playful elements. A more detailed presentation of the results of step 1 and the associated modifications for step 2 can be found in Multimedia Appendix 2.

XSL•FO
RenderX

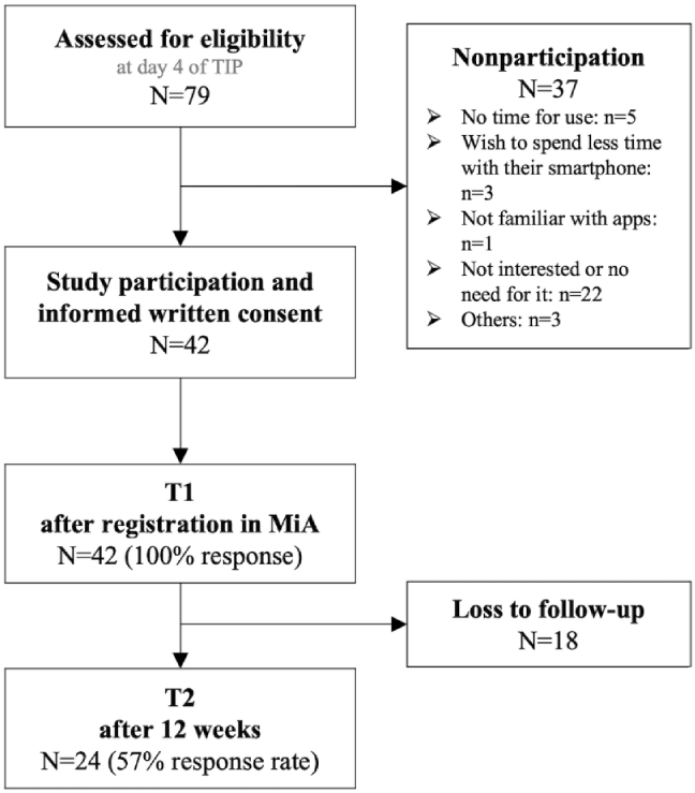
Step 2

Study Participants

Over 10 weeks (study period for step 2), a total of 79 patients participated in the TIP program, of whom 42 (53%) agreed to take part in the intervention and gave informed consent for study participation. The main reasons for nonparticipation were no interest or no subjectively felt need for the intervention (Figure 4).

All 42 patients took part in the individual goal-setting interview and received access to the MiA app for the 12-week testing phase. At T1, all participants completed the questionnaires (42/42, 100% response rate), and at T2, the response rate was 57% (24/42). Table 2 summarizes the sociodemographic data of the patients enrolled in this study.

Figure 4. Flowchart of the study population in step 2, including assessment for eligibility, participation, and follow-up at T1 and T2. MiA: Mein Hautschutz im Alltag (My skin protection in everyday life); TIP: tertiary individual prevention.



XSL•FO
RenderX

Table 2. Sociodemographic data of the study participants at T1 (N=42).

Sociodemographics	Values
Sex, n (%)	
Male	24 (57)
Female	18 (43)
Age (y)	
Values, mean (SD)	49.5 (13.1)
Values, median (IQR)	55.0 (18.6)
Occupational categories, n (%)	
Health care professions	17 (40)
Metalwork industry	9 (21)
Construction	5 (12)
Hairdressing	1 (2)
Food processing	2 (5)
Other	5 (12)
Pensioner	1 (2)
Currently unemployed	2 (5)
Highest educational level, n (%)	
Secondary school or elementary school–leaving certificate	9 (21)
Intermediate school–leaving certificate or secondary school–leaving certificate	15 (36)
"Meister"[a]	3 (7)
College certificate	5 (12)
General higher education entrance qualification or A-levels	7 (17)
Bachelor's degree	1 (2)
Master's degree or diploma	2 (5)
Employment status, n (%)	
Self-employed	4 (10)
Employed	35 (83)
Unemployed	3 (7)
Partnership, n (%)	
Yes	32 (76)
No	10 (24)

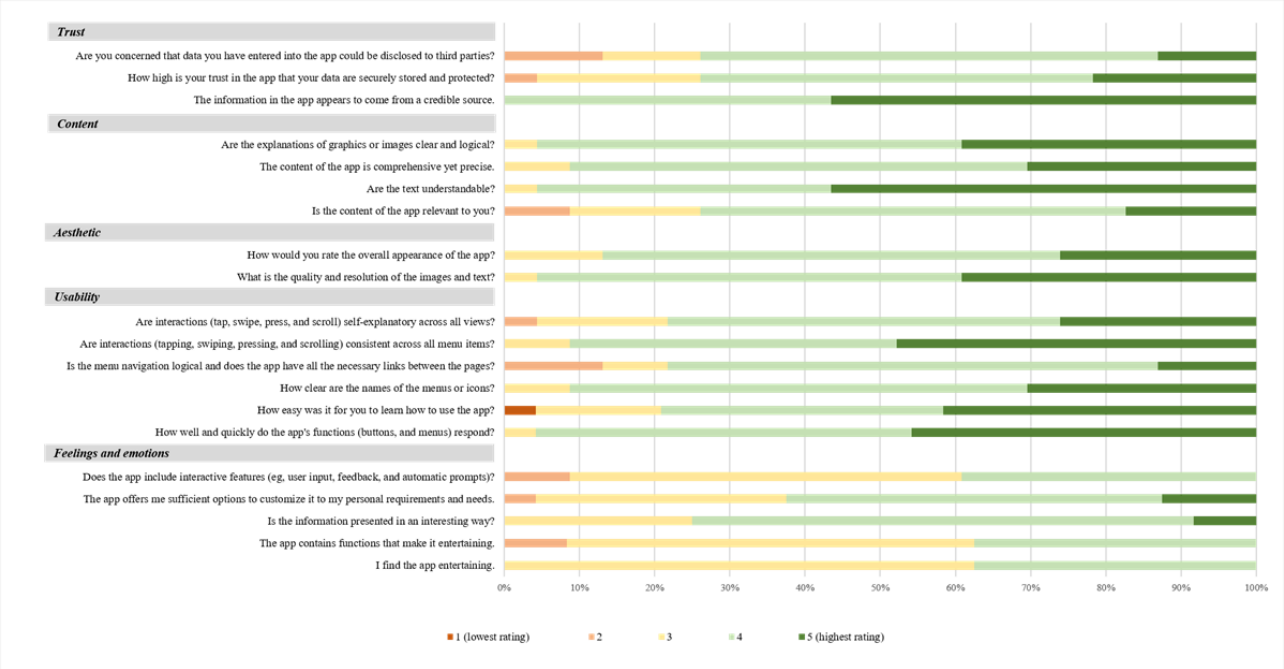
[a]Person with a higher vocational qualification in a craft, also known as "master craftsman" or "master craftswoman."

UX Results

The average UX score, with a theoretical minimum of 20 and a theoretical maximum of 100, was 80.18 (SD 8.94). Figure 5 shows the results for the assessment of the UX from the *trust*, *content*, *aesthetic*, *usability*, and *feelings and emotions* dimensions. The raw data for these items can be found in Multimedia Appendix 3. The results show overall positive ratings (scores of 4 and 5). The *feelings and emotions* dimension stood out, with comparatively frequent medium ratings and no ratings of 5 on 60% (3/5) of the items. Furthermore, the item "How easy was it for you to learn how to use the app?" in the *usability* dimension was the only item with a score of 1 by 2% (1/42) of the participants.

XSL•FO
RenderX

Figure 5. Results for the 5 user experience (UX) dimensions—content, usability, aesthetic, trust, and feelings and emotions—from the participants at T2 (n=24). The ratings range from 1 to 5. A rating of 1 means a low rating, and a rating of 5 is a high rating.

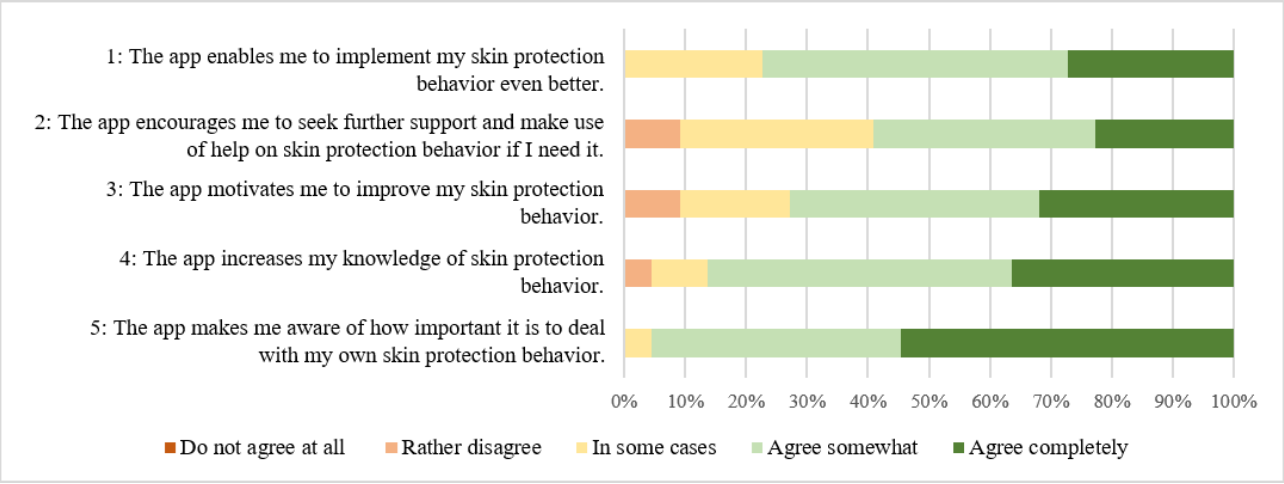


Subjective Quality and Perceived Impact

Overall, the results showed a high subjective quality (Multimedia Appendix 4). Approximately 75% (18/24) of the patients would recommend the app to others. The average willingness to use the app in the following 12 months was "once a week" (7/25, 30%). Most (14/24, 58%) were not willing to pay for the app. The app was most frequently rated with 4 out of 5 stars (15/24, 65%).

As shown in Figure 6, there was an average high level of agreement regarding the subjective impact of the MiA app in relation to skin protection behavior and important determinants (eg, awareness, motivation, or knowledge). All items were rated by more than half (Item 1: 17/24, 71%; Item 2: 13/24, 54%; Item 3: 16/24, 67%; Item 4: 19/24, 79%; Item 5: 21/24, 88%) of the participants as "somewhat agree" or "completely agree" (Multimedia Appendix 4).

Figure 6. Results on the perceived impact of the Mein Hautschutz im Alltag (My skin protection in everyday life) app regarding skin protection behavior by the participants at T2 (n=24).



Adherence

A total of 135 goals were formulated, which corresponds to an average of 3.2 (SD 1.46) goals per person (Table 3). Textbox 1 contains exemplary goals from the goal-setting interviews. Of the 42 participants, 1 (2%) did not formulate any goals, and 10 (24%) formulated the maximum of 5 goals. Most of the goals related to the skin protection category.

XSL•FO
RenderX

Table 3. Number and categories of goals formulated in the individual goal-setting interview by the patients as part of the maintenance program during tertiary individual prevention (N=135).

	Skin cleansing, n (%)	Skin protection, n (%)	Skin care, n (%)	Social support, n (%)	Other, n (%)	Total, n (%)
Goal 1	8 (5.9)	12 (8.9)	11 (8.1)	5 (3.7)	5 (3.7)	41 (30.4)
Goal 2	5 (3.7)	13 (9.6)	5 (3.7)	2 (1.5)	10 (7.4)	35 (25.9)
Goal 3	3 (2.2)	9 (6.7)	7 (5.2)	4 (3)	6 (4.4)	29 (21.5)
Goal 4	4 (3)	8 (5.9)	4 (3)	3 (2.2)	1 (0.7)	20 (14.8)
Goal 5	3 (2.2)	4 (3)	0 (0)	2 (1.5)	1 (0.7)	10 (7.4)
Total	23 (17)	46 (34.1)	27 (20)	16 (11.9)	23 (17)	135 (100)

Textbox 1. Example goals set by the patients in the individual goal-setting interview.

Skin cleansing

- I intend to wash my hands less during the day.
- I intend to reduce the water temperature when cleansing my skin.
- I intend to wash my hands a maximum of 10 times a day.

Skin protection

- I will put on cotton gloves after applying cream.
- I will consistently wear padded gloves for activities involving mechanical pressure on the palms of my hands.
- I intend to apply a skin protection cream at least 3 times a day.

Skin care

- I will apply a skin care cream right at the end of the working day.
- I intend to apply a skin care product to my hands 5 times a day.
- I will apply cream to my whole body once a day (in the evening).

Social support

- I will use a cell phone alarm to remind me to apply skin care cream every day.
- I will ask for support at work (e.g. if the skin condition is bad).

Other

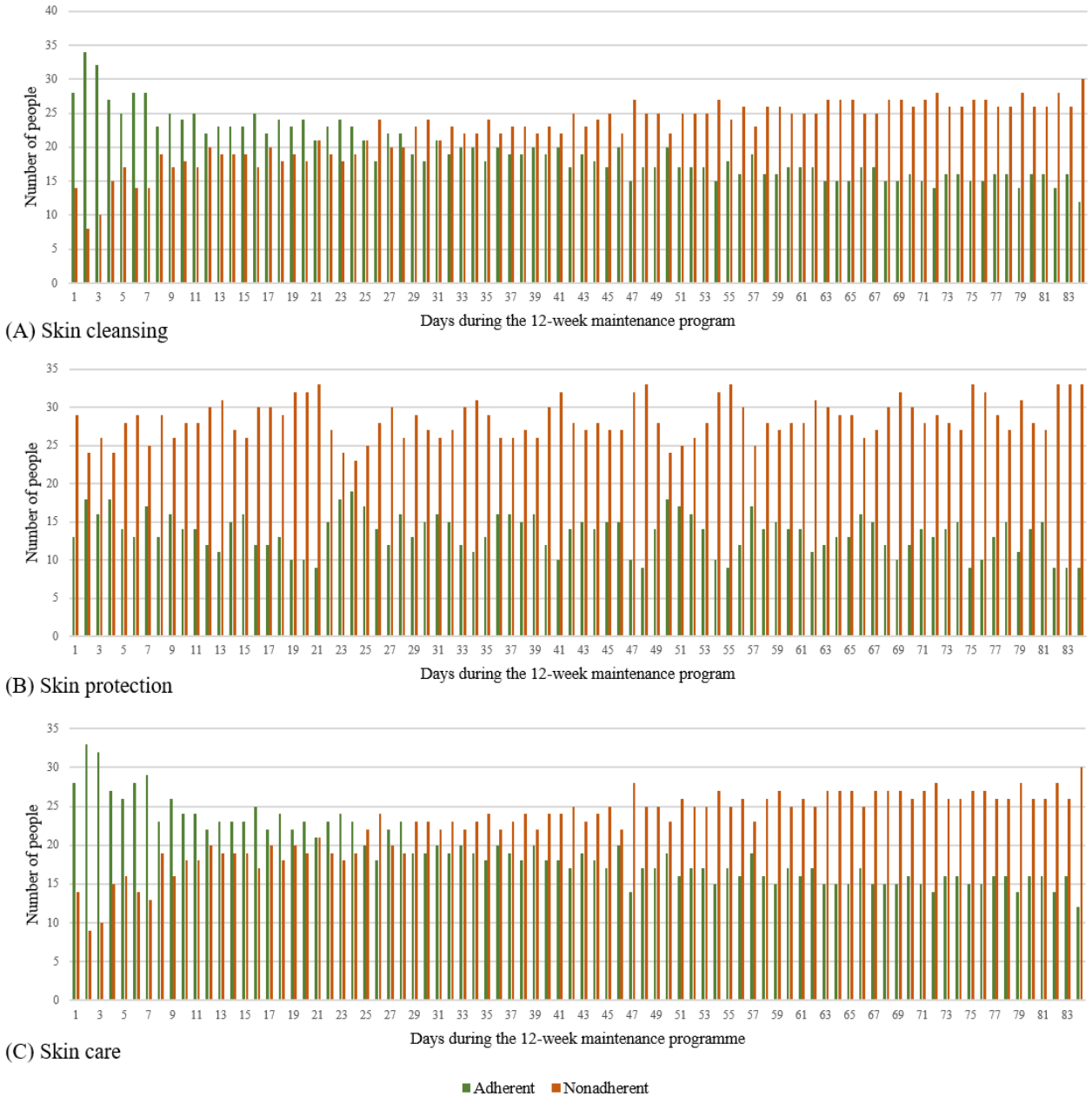
- I intend to pay more attention to myself and say “no” more often.
- I'm going to yoga class every Tuesday evening.
- I intend to go for a walk 2-3 times a week.

The results in terms of adherence to the app function My skin protection behavior are shown in Figure 7. The raw data can be found in Multimedia Appendix 5. The results show differences in adherence between the dimensions of skin protection behavior at the beginning of the maintenance program. On day 1, there were significantly more adherent than nonadherent participants to the dimensions of skin cleansing (adherent participants: 28/42, 67%; nonadherent participants: 14/42, 33%; *P*=.04) and skin care (adherent participants: 28/42, 67%; nonadherent participants: 14/42, 33%; *P*=.04) compared to the skin protection dimension, which had significantly more nonadherent than adherent participants (adherent participants: 13/42, 31%; nonadherent participants: 29/42, 69%; *P*=.02).

Furthermore, there was a difference in adherence over time during the 12-week maintenance program. As shown in Figures 7A and 7C, the number of nonadherent participants to the skin cleansing and skin care dimensions was higher than the number of adherent participants from day 26 onward. After 12 weeks, there were significantly more nonadherent than adherent participants who tracked their skin protection behavior daily in all 3 dimensions: skin cleansing (adherent participants: 12/42, 29%; nonadherent participants: 30/42, 71%; *P*=.008), skin care (adherent participants: 12/42, 29%; nonadherent participants: 30/42, 71%; *P*=.008), and skin protection (adherent participants: 9/42, 21%; nonadherent participants: 33/42, 79%; *P*<.001; Figures 7A, 7B, and 7C).

XSL•FO
RenderX

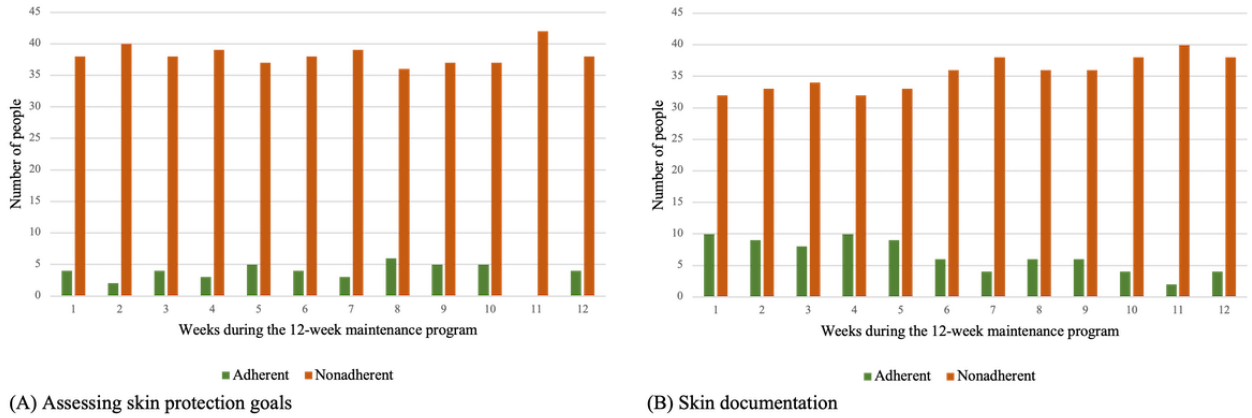
Figure 7. (A) Adherence to tracking of skin cleansing on the app function My Skin Protection Behavior, (B) adherence to tracking of skin protection on the app function My Skin Protection Behavior, and (C) adherence to tracking of skin care on the app function My Skin Protection Behavior by all participants during the 12-week pilot-testing (n=42).



Adherence to rating the achievement of skin protection goals (Figure 8A) and photo documentation (Figure 8B) was lower than adherence to the tracking of skin protection behavior (Multimedia Appendix 5). The number of nonadherent participants was clearly lower than the number of adherent participants in each week of the 12-week maintenance program for both app functions. At the beginning, in week 1 (adherent participants: 4/42, 10%; nonadherent participants: 38/42, 90%; *P*<.001), and at the end, in week 12 (adherent participants: 4/42, 10%; nonadherent participants: 38/42, 90%; *P*<.001), there were significantly more nonadherent than adherent participants who assessed their skin protection goals. Similar results were also found for the app function My Skin Documentation in week 1 (adherent participants: 10/42, 24%; nonadherent participants: 32/42, 76%; *P*<.001) and week 12 (adherent participants: 4/42, 10%; nonadherent participants: 38/42, 90%; *P*<.001).

XSL•FO
RenderX

Figure 8. (A) Adherence to the achievement of skin protection goals on the app function My Skin Protection Goals and (B) adherence to skin documentation on the app function My Skin Documentation by all participants during the 12-week pilot-testing (n=42).



(A) Assessing skin protection goals

(B) Skin documentation

Discussion

Principal Findings and Comparison With Prior Work

The aim of this study was to evaluate the UX of and adherence to an app-based maintenance program. The MiA app is a self-management tool for patients with WRSDs after participating in an inpatient rehabilitation program (TIP) to support change in and maintenance of skin protection behavior. To the best of our knowledge, the MiA app is the first app for people with WRSDs. Our study was conducted by following a multistage and iterative development and pilot-testing process, with this paper focusing on the results of phase 2.

UX research is an important part in the development of apps to identify and address potential barriers to the use of apps as early as possible. It should precede the summative evaluation [38]. However, many studies investigating the influence of app-based interventions in the dermatological setting only report on clinical outcomes such as quality of life, eczema severity, or itching [24,27] or app use behavior [25,26]. Zvulunov et al [24] point out that there are missing data on the usability of and adherence to apps in the field of chronic skin diseases.

We evaluated the UX using the dimensions of *feelings and emotions*, *usability*, *aesthetics*, *content*, and *trust*. Our study generally showed high to very high ratings for all dimensions with the exception of *feelings and emotions*, which was rated comparatively lower. Studies show that gamified apps can improve the interaction, motivation, and adherence [23,52]. However, the benefit of gamification differs between genders and decreases with age, making it difficult to successfully implement those functions in our heterogeneous study population [53]. According to Tolks et al [53], gamified apps are particularly useful for people with low motivation to implement a behavior, and Matterne et al [54] were able to show that patients with WRSDs have a high motivation to implement skin protection, which increased significantly during the TIP. Furthermore, in a focus group discussion in our study, patients expressed that the app should not contain any playful elements and should remain factual as it dealt with a disease. The extent to which the app is used also depends on the app's BCTs. Milne-Ives et al [55] were able to demonstrate that some BCTs are associated with higher use of and engagement with an app. These included goal setting and feedback, self-monitoring, and social support, which are also part of the MiA app. However, in the future, the app could be expanded to include further interactive functions that allow users, for example, to assess their skin condition more accurately. One possibility might be to analyze the severity of hand eczema in photo documentation using artificial intelligence [24]. Weigandt et al [27] were able to show that the eczema severity determined via photos significantly correlates with live assessment by dermatologists. Therefore, photographs appear to allow for reliable results regarding the actual severity of hand eczema.

The presentation of information on the app was rated as particularly interesting by patients. We attribute this to the different ways in which the information was presented (podcasts, videos, and short questions and answers).

The descriptive results of the UX also show single exceptions regarding usability. Due to the small study cohort at T2, we were unable to test any associations between UX and the sociodemographic characteristics of the participants. However, digital health literacy plays an important role in usability. It is known that, in addition to app-specific factors, the use of an app may also depend on personal factors such as age, sex, educational level, and socioeconomic status [15,27,56]. Nationally and internationally, it has been shown that older people with a lower educational level and of a lower socioeconomic status tend to have a lower digital health literacy [57,58]. With an average age of 49.5 (SD 13.1) years and a high proportion of people with elementary and secondary school–leaving certification (24/42, 57%), our study cohort had characteristics that could indicate a lower level of digital health literacy. Therefore, a potential influence of digital health literacy on the use and handling of the MiA app is conceivable. Further research should be carried out to determine which socioeconomic characteristics in our field of application are particularly affected by low digital health literacy. This is important to provide patients with individual and tailored support from health care staff during the onboarding process [59]. Nevertheless, onboarding via personal contact in face-to-face meetings is a particular strength of our study because we were able to respond to the individual digital competence of the patients.

XSL•FO
RenderX

We found differences in adherence between the interactive functions of the app. The daily tracking of skin protection behavior was used more often than the weekly skin documentation and goal assessment. Furthermore, skin care and skin cleansing were tracked more frequently than skin protection. In addition, adherence to skin cleansing and skin care tracking decreased over the 12 weeks. The lower tracking of skin protection at the beginning might be explained by the fact that patients are at home for a period of 3 weeks after discharge and do not start working in their profession straight away. That means that the use of skin protection creams at work only becomes relevant from day 21 onward. The digital documentation of skin protection creams may not have increased from day 21 onward as the practical implementation of digital tracking during work is not feasible. However, in contrast to this, we observed a clearly higher proportion of formulated individual goals in the skin protection category than in the skin cleansing, skin care, support, and other categories. This emphasizes that the topic of skin protection is nevertheless an important aspect for patients after the TIP even if this cannot be reflected in skin protection tracking for the aforementioned reasons. Against this background, linking both functions appears to be useful to adequately cover all dimensions of skin protection behavior. There are various possible reasons for the less frequent use of weekly photo documentation and target assessment. First, both components are technically more complex and more challenging functions and require more time. One possible explanation for the low level of photo documentation could be that photos are not taken routinely by the users but, rather, when acute skin changes occur. The weekly rhythm we set may not correlate with the progression of the individual skin diseases. WRSDs occur in over half of cases with a comorbidity of AD and progress in intervals of varying length [33]. The goal assessment is also cognitively more complex because it requires reflection on the last behavior performed by the users and the decision of the appropriate response option. The reasons for the less frequent use of both functions should be further investigated, for example, by users and nonusers reporting on their experiences, motivation, and the usefulness of these functions (eg, in qualitative interviews [56]). Nevertheless, it should be kept in mind that high adherence to the app functions (eg, skin protection tracking) is not necessarily associated with behavior changes and better clinical outcomes. In everyday life, skin diseases are associated with a high level of adherence to therapy. Daily documentation may be too time-consuming to implement in practice, meaning that regular app use does not necessarily lead to better outcomes in terms of eczema severity or quality of life [27]. The average willingness to use the app approximately once a week supports this assumption. The use of the MiA app after the TIP can also be individual depending on the extent to which patients benefited from patient education during the TIP. Another influencing factor on app use could be the duration of the disease. TIP patients have usually had the skin disease for several years and have developed their own self-management strategies in the past. In the future, it could be investigated what are the use and benefits for patients with milder forms of WRSDs (eg, as part of outpatient care). The results on the perceived impact show that the MiA app can also be used as a tool for disseminating knowledge about skin protection–related content.

Overall, our results are consistent with those of other studies on dermatological apps and, thus, show expected results on adherence, which decreases over time [25]. Shah et al [26] showed that the average duration of use of an app for people with AD was approximately 6 weeks. The frequent short duration of use of app-based interventions and the associated lack of evidence of long-term effects is a general and well-known problem in this research context [2,26].

We were able to demonstrate a high quality of the app. Despite the low adherence, the results on the perceived effect also show promising indications that the app might influence relevant determinants such as motivation, knowledge, or awareness regarding the change in health behavior according to the subjective statements of the participants. However, the effect of the MiA app on skin protection behavior, quality of life, and eczema severity remains unclear thus far and should be further investigated in the future in comparison with a control group.

Limitations

In addition to the limitations already mentioned, this study has further limitations that should be considered when interpreting the results. First, UX is a multifactorial and complex construct. There are different methods for assessing UX that have their own strengths and limitations [13]. While questionnaires are cost-efficient and provide quick results, dishonest answers and individual interpretations of the results can be disadvantageous. Focus group discussions are easy to conduct but dependent on the interviewing skills of those conducting them [14]. To measure UX, we used a validated instrument as a basis and adapted it to the context and target group of the app. We concede that the validity of the instrument and a direct comparison of the results with those of other studies is possible to a limited extent. The standardization of the definitions of the UX dimensions and the development of validated instruments for UX research is an important future research focus [19].

The participants in our study matched the sociodemographic characteristics of those of other studies on TIP [60]. Nevertheless, due to the small number of participants, the low use behavior, and a remaining cohort of 57% (24/42) at T2, it was not possible to calculate differences in UX and adherence among age, sex, and educational level.

The target group participated during the development and testing of the MiA app. Nevertheless, their involvement could have been even stronger in the form of agile co-design (eg, for prototyping the app before programming and technical implementation and for identifying realistic adherence [23,61-63]). However, such an approach was not possible due to the organizational processes and duration of this study.

We were also unable to determine the extent to which the informative functions (podcasts, questions and answers, and videos) were used. On the basis of other studies, it can be assumed that the use of educational elements also decreases after a few weeks [25]. However, podcasts activated over the course of the maintenance program could be a way to maintain use over time.

XSL•FO
RenderX

Conclusions

Our study shows a high UX of the MiA app in almost all dimensions. To increase interaction, functions can be added to the app (eg, use of artificial intelligence with photo documentation to allow for a clearer assessment of the skin condition). We refrained from using gamification due to the heterogeneous study population of patients with WRSDs. We also observed single exceptions in usability. Usability depends, among other things, on the digital health literacy of the user. Future research should investigate in this context which sociodemographic characteristics show lower usability to provide these groups with further support when using the app.

We obtained expected results in terms of adherence, with decreasing use over time, which corresponds to the current state of research in this field. Particularly after resuming work after 3 weeks, daily documentation of skin protection behavior appears to be too time-consuming. The effectiveness of the MiA app in terms of skin protection behavior, quality of life, and eczema severity should be further investigated with a control group.

Acknowledgments

The authors would like to thank the Institution for Statutory Accident Insurance and Prevention of the Health and Welfare Services (Berufsgenossenschaft für Gesundheitsdienst und Wohlfahrtspflege) for funding the project (reference: Ext FF_1569; Technologiebasierte Nachbetreuung in der berufsdermatologischen prevention der BK Nr. 5101 [TecNaP]). The authors thank the members of an internal Institute for Interdisciplinary Dermatological Prevention and Rehabilitation working group and Dr Andreas Reusch of the Zentrum Patientenschulung und Gesundheitsförderung (Centre for Patient Education and Health Promotion) for developing the concept for the individual goal-setting interview. They would also like to thank student assistants Gesa Lühring and Katharina Brieskorn for their support during the focus group discussions, study administration, and data entry. For programming the app and providing technical support, the authors would like to thank the companies BINACON GmbH and Solutec GmbH, with Klaus Welling, Thomas Waldmann, Michael Mühlhäuser, Stanislav Yordanov, Stefan Schmädeke, and Patrick Renter.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Intervention description.
[DOCX File , 19 KB - formative_v9i1e66791_app1.docx]

Multimedia Appendix 2
Results of step 1.
[DOCX File , 22 KB - formative_v9i1e66791_app2.docx]

Multimedia Appendix 3
User experience.
[DOCX File , 21 KB - formative_v9i1e66791_app3.docx]

Multimedia Appendix 4
Subjective quality and perceived impact.
[DOCX File , 29 KB - formative_v9i1e66791_app4.docx]

Multimedia Appendix 5
Adherence to the app's My skin protection behavior, My skin documentation, and My skin protection goals functions.
[DOCX File , 34 KB - formative_v9i1e66791_app5.docx]

References

1. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
2. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]

XSL•FO
RenderX

3. Kitazawa M, Takeda Y, Hatta M, Horikawa C, Sato T, Osawa T, et al. Lifestyle intervention with smartphone app and isCGM for people at high risk of type 2 diabetes: randomized trial. J Clin Endocrinol Metab 2024 Mar 15;109(4):1060-1070 [FREE Full text] [doi: 10.1210/clinem/dgad639] [Medline: 37931069]
4. Guo M, Meng F, Guo Q, Bai T, Hong Y, Song F, et al. Effectiveness of mHealth management with an implantable glucose sensor and a mobile application among Chinese adults with type 2 diabetes. J Telemed Telecare 2023 Sep;29(8):632-640. [doi: 10.1177/1357633X211020261] [Medline: 34152238]
5. Stawarz K, Cox AL, Blandford A. Beyond self-tracking and reminders: designing smartphone apps that support habit formation. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Presented at: CHI '15; April 18-23, 2015; Seoul, Republic of Korea. [doi: 10.1145/2702123.2702230]
6. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using health and well-being apps for behavior change: a systematic search and rating of apps. JMIR Mhealth Uhealth 2019 Jul 04;7(7):e11926 [FREE Full text] [doi: 10.2196/11926] [Medline: 31274112]
7. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
8. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
9. Birrell L, Furneaux-Bate A, Chapman C, Newton NC. A mobile peer intervention for preventing mental health and substance use problems in adolescents: protocol for a randomized controlled trial (The Mind Your Mate Study). JMIR Res Protoc 2021 Jul 30;10(7):e26796 [FREE Full text] [doi: 10.2196/26796] [Medline: 34328426]
10. Rismawan W, Marchira CR, Rahmat I. Usability, acceptability, and adherence rates of mobile application interventions for prevention or treatment of depression: a systematic review. J Psychosoc Nurs Ment Health Serv 2021 Feb 01;59(2):41-47. [doi: 10.3928/02793695-20201104-02] [Medline: 33180949]
11. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
12. Lubis M, Sutoyo E, Azuddin M, Handayani D. User experience in mobile application design: utility defined context of use. J Phys Conf Ser 2019;1361:12043. [doi: 10.1088/1742-6596/1361/1/012043]
13. Kaveladze BT, Wasil AR, Bunyi JB, Ramirez V, Schueller SM. User Experience, Engagement, and Popularity in Mental Health Apps: Secondary Analysis of App Analytics and Expert App Reviews. JMIR Hum Factors 2022 Jan 31;9(1):e30766 [FREE Full text] [doi: 10.2196/30766] [Medline: 35099398]
14. Bitkina OV, Kim HK, Park J. Usability and user experience of medical devices: an overview of the current state, analysis methodologies, and future challenges. Int J Ind Ergon 2020 Mar;76:102932. [doi: 10.1016/j.ergon.2020.102932]
15. Biduski D, Bellei EA, Rodriguez JP, Zaina LA, De Marchi AC. Assessing long-term user experience on a mobile health application through an in-app embedded conversation-based questionnaire. Comput Hum Behav 2020 Mar;104:106169. [doi: 10.1016/j.chb.2019.106169]
16. Kim A, Kim KJ. User experience and the multi-stage adoption of mobile apps. J Inf Technol Appl Manag 2014;21(2):49-79. [doi: 10.21219/jitam.2014.21.2.049]
17. Schrepp M, Hinderks A, Thomaschewski J. Construction of a benchmark for the User Experience Questionnaire (UEQ). Int J Interact Multim Artif Intell 2017;4(4):40-44. [doi: 10.9781/ijimai.2017.445]
18. Schrepp M, Hinderks A, Thomaschewski J. Applying the User Experience Questionnaire (UEQ) in different evaluation scenarios. In: Marcus A, editor. Design, User Experience, and Usability. Cham, Switzerland: Springer; 2014.
19. Lemon C, Huckvale K, Carswell K, Torous J. A narrative review of methods for applying user experience in the design and assessment of mental health smartphone interventions. Int J Technol Assess Health Care 2020;36(1):64-70. [doi: 10.1017/S0266462319003507] [Medline: 31973787]
20. Blanchard M, Backhaus L, Ming Azevedo P, Hügle T. An mHealth app for fibromyalgia-like post-COVID-19 syndrome: protocol for the analysis of user experience and clinical data. JMIR Res Protoc 2022 Feb 04;11(2):e32193 [FREE Full text] [doi: 10.2196/32193] [Medline: 34982039]
21. Thielsch MT, Salaschek M. [Toolbox for user-experience evaluation of online health information]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2020 Jun;63(6):721-728. [doi: 10.1007/s00103-020-03142-7] [Medline: 32367206]
22. Truong MT, Nwosu OB, Gaytan Torres ME, Segura Vargas MP, Seifer AK, Nitschke M, et al. A yoga exercise app designed for patients with axial spondylarthritis: development and user experience study. JMIR Form Res 2022 Jun 03;6(6):e34566 [FREE Full text] [doi: 10.2196/34566] [Medline: 35657655]
23. Xu X, Griva K, Koh M, Lum E, Tan WS, Thng S, et al. Creating a smartphone app for caregivers of children with atopic dermatitis with caregivers, health care professionals, and digital health experts: participatory co-design. JMIR Mhealth Uhealth 2020 Oct 29;8(10):e16898 [FREE Full text] [doi: 10.2196/16898] [Medline: 33118949]

XSL•FO
RenderX

24. Zvulunov A, Lenevich S, Migacheva N. A mobile health app for facilitating disease management in children with atopic dermatitis: feasibility and impact study. JMIR Dermatol 2023 Dec 13;6:e49278 [FREE Full text] [doi: 10.2196/49278] [Medline: 38090787]
25. Gudmundsdóttir SL, Ballarini T, Ámundadóttir ML, Mészáros J, Eysteinsdottir JH, Thorleifsdottir RH, et al. Engagement, retention, and acceptability in a digital health program for atopic dermatitis: prospective interventional study. JMIR Form Res 2023 Jun 14;7:e41227 [FREE Full text] [doi: 10.2196/41227] [Medline: 36975050]
26. Shah S, Kemp JM, Kvedar JC, Gracey LE. A feasibility study of the burden of disease of atopic dermatitis using a smartphone research application, myEczema. Int J Womens Dermatol 2020 Dec;6(5):424-428 [FREE Full text] [doi: 10.1016/j.ijwd.2020.08.001] [Medline: 33898712]
27. Weigandt WA, Schardt Y, Bruch A, Herr R, Goebeler M, Benecke J, et al. Impact of an eHealth smartphone app on quality of life and clinical outcome of patients with hand and foot eczema: prospective randomized controlled intervention study. JMIR Mhealth Uhealth 2023 Mar 07;11:e38506 [FREE Full text] [doi: 10.2196/38506] [Medline: 36881465]
28. DGUV-Statistiken für die Praxis 2023. DGUV. 2024 Oct. URL: https://publikationen.dguv.de/statistiken/zahlen-und-fakten/4994/dguv-statistiken-fuer-die-praxis-2023 [accessed 2025-04-10]
29. Brans R, Schröder-Kraft C, Skudlik C, John SM, Geier J. Tertiary prevention of occupational skin diseases: prevalence of allergic contact dermatitis and pattern of patch test results. Contact Dermatitis 2019 Jan;80(1):35-44. [doi: 10.1111/cod.13098] [Medline: 30506828]
30. Skudlik C, Weisshaar E, Scheidt R, Elsner P, Wulfhorst B, Schönfeld M, et al. First results from the multicentre study rehabilitation of occupational skin diseases--optimization and quality assurance of inpatient management (ROQ). Contact Dermatitis 2012 Mar;66(3):140-147. [doi: 10.1111/j.1600-0536.2011.01991.x] [Medline: 22070197]
31. Brans R, John SM. Clinical patterns and associated factors in patients with hand eczema of primarily occupational origin. J Eur Acad Dermatol Venereol 2016 May;30(5):798-805. [doi: 10.1111/jdv.13515] [Medline: 26660508]
32. Elsner P, Schliemann S. Prävention der BK 5101. Trauma Berufskrankh 2015 Dec 1;17(4):227-232. [doi: 10.1007/s10039-015-0103-8]
33. Brans R, Skudlik C. [Prevention of hand eczema]. Hautarzt 2019 Oct;70(10):797-803. [doi: 10.1007/s00105-019-4463-z] [Medline: 31410496]
34. Ristow N, Rocholl M, Wilke A, John SM, Ludewig M. Using behaviour change techniques to characterize patient educational interventions in tertiary individual prevention of work-related skin diseases. Transl Behav Med 2024 May 24;14(6):341-352. [doi: 10.1093/tbm/ibad081] [Medline: 38159249]
35. Wilke A, Skudlik C, Sonsmann FK. [Individual prevention of occupational contact dermatitis: protective gloves and skin protection recommendations as part of the patient management scheme by the public statutory employers' liability insurance]. Hautarzt 2018 Jun;69(6):449-461. [doi: 10.1007/s00105-018-4170-1] [Medline: 29721626]
36. Wilke A, Gediga K, John SM, Wulfhorst B. Evaluation of structured patient education in occupational skin diseases: a systematic assessment of the disease-specific knowledge. Int Arch Occup Environ Health 2014 Nov;87(8):861-869. [doi: 10.1007/s00420-014-0926-9] [Medline: 24482039]
37. Ristow N, Wilke A, Malte John S, Ludewig M. Development of an app-based maintenance programme to promote skin protection behaviour for patients with work-related skin diseases. Health Educ J 2022 Aug 02;81(6):731-744. [doi: 10.1177/00178969221115990]
38. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
39. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. [better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide]. Gesundheitswesen 2016 Mar;78(3):175-188. [doi: 10.1055/s-0041-111066] [Medline: 26824401]
40. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014 Mar 07;348:g1687 [FREE Full text] [doi: 10.1136/bmj.g1687] [Medline: 24609605]
41. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. Sutton, UK: Silverback Publishing; 2014.
42. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. JMIR Mhealth Uhealth 2016 Jun 09;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
43. Wolcott MD, McLaughlin JE. Exploring user experience (UX) research methods in health professions education. Curr Pharm Teach Learn 2024 Feb;16(2):144-149. [doi: 10.1016/j.cptl.2023.12.010] [Medline: 38158330]
44. Gracey LE, Zan S, Gracz J, Miner JJ, Moreau JF, Sperber J, et al. Use of user-centered design to create a smartphone application for patient-reported outcomes in atopic dermatitis. NPJ Digit Med 2018;1:33 [FREE Full text] [doi: 10.1038/s41746-018-0042-4] [Medline: 31304315]

XSL•FO
RenderX

45. Kernebeck S, Busse TS, Ehlers JP, Vollmar HC. [Adherence to digital health interventions: definitions, methods, and open questions]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021 Oct;64(10):1278-1284 [FREE Full text] [doi: 10.1007/s00103-021-03415-9] [Medline: 34559252]
46. Alberts NM, Badawy SM, Hodges J, Estepp JH, Nwosu C, Khan H, et al. Development of the InCharge health mobile app to improve adherence to hydroxyurea in patients with sickle cell disease: user-centered design approach. JMIR Mhealth Uhealth 2020 May 08;8(5):e14884 [FREE Full text] [doi: 10.2196/14884] [Medline: 32383683]
47. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
48. Messner EM, Terhorst Y, Barke A, Baumeister H, Stoyanov S, Hides L, et al. The German version of the mobile app rating scale (MARS-G): development and validation study. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e14479 [FREE Full text] [doi: 10.2196/14479] [Medline: 32217504]
49. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
50. Pelz C, Schmitt A, Meis M. Knowledge mapping as a tool for analyzing focus groups and presenting their results in market and evaluation research. Forum Qual Soc Res 2004;5(2) [FREE Full text] [doi: 10.17169/fqs-5.2.601]
51. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
52. Maher CA, Olds T, Vandelanotte C, Plotnikoff R, Edney SM, Ryan JC, et al. Gamification in a physical activity app: what gamification features are being used, by whom, and does it make a difference? Games Health J 2022 Jun;11(3):193-199. [doi: 10.1089/g4h.2021.0207] [Medline: 35501981]
53. Tolks D, Lampert C, Dadaczynski K, Maslon E, Paulus P, Sailer M. [Game-based approaches to prevention and health promotion: serious games and gamification]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2020 Jun;63(6):698-707. [doi: 10.1007/s00103-020-03156-1] [Medline: 32451596]
54. Matterne U, Diepgen TL, Weisshaar E. Effects of a health-educational and psychological intervention on socio-cognitive determinants of skin protection behaviour in individuals with occupational dermatoses. Int Arch Occup Environ Health 2010 Feb;83(2):183-189. [doi: 10.1007/s00420-009-0448-z] [Medline: 19641933]
55. Milne-Ives M, Homer SR, Andrade J, Meinert E. Potential associations between behavior change techniques and engagement with mobile health apps: a systematic review. Front Psychol 2023;14:1227443 [FREE Full text] [doi: 10.3389/fpsyg.2023.1227443] [Medline: 37794916]
56. Jeffrey B, Bagala M, Creighton A, Leavey T, Nicholls S, Wood C, et al. Mobile phone applications and their use in the self-management of type 2 diabetes mellitus: a qualitative study among app users and non-app users. Diabetol Metab Syndr 2019;11:84 [FREE Full text] [doi: 10.1186/s13098-019-0480-4] [Medline: 31636719]
57. Choukou MA, Sanchez-Ramirez DC, Pol M, Uddin M, Monnin C, Syed-Abdul S. COVID-19 infodemic and digital health literacy in vulnerable populations: a scoping review. Digit Health 2022;8:20552076221076927 [FREE Full text] [doi: 10.1177/20552076221076927] [Medline: 35223076]
58. Dratva J, Schaeffer D, Zeeb H. [Digital health literacy in Germany: current status, concepts, and challenges]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2024 Mar;67(3):277-284 [FREE Full text] [doi: 10.1007/s00103-024-03841-5] [Medline: 38315221]
59. Busse TS, Nitsche J, Kernebeck S, Jux C, Weitz J, Ehlers JP, et al. Approaches to improvement of digital health literacy (eHL) in the context of person-centered care. Int J Environ Res Public Health 2022 Jul 07;19(14):8309 [FREE Full text] [doi: 10.3390/ijerph19148309] [Medline: 35886158]
60. Rocholl M, Wilke A, Meyer J, John SM, Ludewig M. Illness perceptions of patients with occupational skin diseases in a healthcare centre for tertiary prevention: a cross-sectional study. Int J Environ Res Public Health 2023 Apr 26;20(9):5652 [FREE Full text] [doi: 10.3390/ijerph20095652] [Medline: 37174171]
61. Fox S, Brown LJ, Antrobus S, Brough D, Drake RJ, Jury F, et al. Co-design of a smartphone app for people living with dementia by applying agile, iterative co-design principles: development and usability study. JMIR Mhealth Uhealth 2022 Jan 14;10(1):e24483 [FREE Full text] [doi: 10.2196/24483] [Medline: 35029539]
62. Poulsen A, Hickie IB, Alam M, LaMonica HM. User experience co-design of a mobile application to support childrearing in low- and middle-income countries. Stud Health Technol Inform 2023 Jun 22;304:86-90. [doi: 10.3233/SHTI230377] [Medline: 37347576]
63. Vargas C, Whelan J, Brimblecombe J, Allender S. Co-creation, co-design, co-production for public health - a perspective on definition and distinctions. Public Health Res Pract 2022 Jun 15;32(2):3222211 [FREE Full text] [doi: 10.17061/phrp3222211] [Medline: 35702744]

XSL•FO
RenderX

Abbreviations

AD: atopic dermatitis
BCT: behavior change technique
iDerm: Institute for Interdisciplinary Dermatological Prevention and Rehabilitation
MiA: Mein Hautschutz im Alltag (“My skin protection in everyday life”)
TIP: tertiary individual prevention
uMARS: user version of the Mobile Application Rating Scale
UX: user experience
WRSD: work-related skin disease

Edited by A Mavragani; submitted 24.09.24; peer-reviewed by K Stawarz; comments to author 03.02.25; revised version received 24.02.25; accepted 25.02.25; published 18.04.25.

Please cite as:
Ristow N, Wilke A, Skudlik C, John SM, Ludewig M
User Experience of and Adherence to a Smartphone App to Maintain Behavior Change and Self-Management in Patients With Work-Related Skin Diseases: Multistep, Single-Arm Feasibility Study
JMIR Form Res 2025;9:e66791
URL: https://formative.jmir.org/2025/1/e66791
doi:10.2196/66791
PMID:

©Nele Ristow, Annika Wilke, Christoph Skudlik, Swen Malte John, Michaela Ludewig. Originally published in JMIR Formative Research (https://formative.jmir.org), 18.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Designing Health Care Provider–Centered Emergency Department Interventions: Participatory Design Study

Woosuk Seo[1*], BS; Jiaqi Li[1*], MS; Zhan Zhang[2], PhD; Chuxuan Zheng[3], BA; Hardeep Singh[4], MD, MPH; Kalyan Pasupathy[5], PhD; Prashant Mahajan[6], MPH, MBA, MD; Sun Young Park[7], PhD

[1]School of Information, University of Michigan, Ann Arbor, MI, United States

[2]Seidenberg School of Computer Science and Information Systems, Pace University, New York, NY, United States

[3]Department of Human Centered Design & Engineering, University of Washington, Seattle, WA, United States

[4]Center for Innovations in Quality, Effectiveness and Safety (IQuESt), Michael E. DeBakey Veterans Affairs Medical Center and Baylor College of Medicine, Houston, TX, United States

[5]Biomedical and Health Information Sciences, University of Illinois Chicago, Chicago, IL, United States

[6]Department of Emergency Medicine, University of Michigan Medical School, Ann Arbor, MI, United States

[7]School of Information, Stamps School of Art and Design, University of Michigan, Ann Arbor, MI, United States

[*]these authors contributed equally

Corresponding Author:
Woosuk Seo, BS
School of Information
University of Michigan
105 S State St
Ann Arbor, MI, 48104
United States
Phone: 1 2063109264
Email: seow@umich.edu

Abstract

Background: In the emergency department (ED), health care providers face extraordinary pressures in delivering accurate diagnoses and care, often working with fragmented or inaccessible patient histories while managing severe time constraints and constant interruptions. These challenges and pressures may lead to potential errors in the ED diagnostic process and risks to patient safety. With advances in technology, interventions have been developed to support ED providers in such pressured settings. However, these interventions may not align with the current practices of ED providers. To better design ED provider–centered interventions, identifying their needs in the diagnostic process is critical.

Objective: This study aimed to identify ED providers' needs in the diagnostic process through participatory design sessions and to propose design guidelines for provider-centered technological interventions that support decision-making and reduce errors.

Methods: We conducted a participatory design study with ED providers to validate their needs and identify considerations for designing ED provider–centered interventions to improve diagnostic safety. We used 9 technological intervention ideas as storyboards to address the study participants' needs. We had participants discuss the use cases of each intervention idea to assess their needs during the ED care process and facilitated co-design activities with the participants to improve the technological intervention designs. We audio- and video-recorded the design sessions. We then analyzed session transcripts, field notes, and design sketches. In total, we conducted 6 design sessions with 17 ED frontline providers.

Results: Through design sessions with ED providers, we identified 4 key needs in the diagnostic process: information integration, patient prioritization, ED provider-patient communication, and care coordination. We interpreted them as insights for designing technological interventions for ED patients. Hence, we discussed the design implications for technological interventions in four key areas: (1) enhancing ED provider–ED provider communication, (2) enhancing ED provider-patient communication, (3) optimizing the integration of advanced technology, and (4) unleashing the potential of artificial intelligence tools in the ED to improve diagnosis. This work offers evidence-based technology design suggestions for improving diagnostic processes.

XSL•FO
RenderX

Conclusions: This study provides unique insights for designing technological interventions to support ED diagnostic processes. By inviting ED providers into the design process, we present unique insights into the diagnostic process and design considerations for designing novel technological interventions that meet ED providers' needs in the diagnostic process.

International Registered Report Identifier (IRRID): RR2-10.2196/55357

(JMIR Form Res 2025;9:e68891) doi:10.2196/68891

KEYWORDS

emergency departments; participatory design; health care providers; technology; interventions; artificial intelligence

Introduction

Background

Emergency departments (EDs) are dynamic, challenging, and time-critical medical environments in a hospital. In recent years, the increasing need for emergency care services has led to overcrowding [1-3], resulting in extended wait times for patients, diminished patient satisfaction [4], and suboptimal patient outcomes [5,6]. With the high volume of patients and staff shortages [7], health care providers in the ED face significant challenges in their workflow [3,8]. For example, they often encounter frequent distractions and interruptions from secondary tasks or transitions [9-11], ineffective communication with patients [12-14], communication breakdowns, and information loss during the care process [15-18], all while being overwhelmed by the intensive patient and teamwork information [15,19]. All these challenges and issues can potentially lead to diagnostic errors [20-23].

Given these challenges in the diagnostic process in EDs, previous work has attempted to implement different types of technology to enhance this process. In particular, clinical decision support systems (CDSSs) have been implemented to predict morbidity and mortality [24-27], sepsis [28], and adverse prognosis [29]; improve triage [30-35]; automate clinical documentation [36,37]; and predict hospitalization and admission [38-41]. Although these systems enhance ED providers' decision-making process, many still face barriers to practical applicability and integration into ED workflows [28,42]. For instance, a previous study found that emergency physicians poorly accepted an evidence-based CDSS for evaluating suspected pulmonary embolism because it increased computer time compared to the original workflow, leading to low adoption rates in the ED [43].

One of the main reasons for the failure to adopt and implement effective technological interventions in ED is the limited understanding of ED providers' technology needs and examining appropriate approaches for integrating advanced technologies into dynamic ED workflows [44]. A way to understand the needs of ED providers is to involve them in the design process [45]. Participatory design (PD) [45] is a methodology that engages all stakeholders in the design process to create solutions that address their needs. Previous work has suggested that participatory and user-centered design approaches can address these concerns [45,46] as they engage all stakeholders in the early stages of the design process, thereby more effectively and promptly addressing user needs and improving the alignment of new technologies with existing workflows [47]. Examples of adopting PD approaches in ED technology development include collaborating with ED providers to design information displays that support awareness and enhance ED teamwork [48] and redesigning the ED patient health information system [49]. Following this body of work, we used the PD approach in our study to explore potential improvements and technological interventions for the ED diagnostic process, an area previously unexplored using the PD method.

Objectives

In our study, we explored the following research questions: (1) what are the challenges and needs of frontline providers in the ED diagnostic process? and (2) how should technological interventions be designed to address the specific needs of ED providers? To answer these questions, we conducted interviews and PD sessions, engaging ED providers from different hospitals to identify opportunities for creating user-informed technological interventions and to brainstorm strategies for enhancing the diagnostic process in ED. Our study identified 4 primary areas for enhancing the diagnostic process in ED: information integration, patient prioritization, ED provider-patient communication, and care coordination. The findings also highlighted the key concerns about integrating advanced technologies into dynamic ED workflows. Finally, we conclude the paper by discussing design implications for (1) enhancing ED provider-ED provider care coordination and communication, (2) enhancing ED provider-patient communication, (3) optimizing the integration of advanced technology in the ED, and (4) unleashing the potential of artificial intelligence (AI) tools in the ED to improve diagnosis. This work will enhance ED care by offering evidence-based technology recommendations and establishing ED-specific design principles for improving diagnostic processes and care.

Methods

Study Overview

This research is part of a comprehensive initiative examining the perspectives of ED patients and health care providers regarding technological interventions to enhance the ED diagnostic process. The broader project aims to develop design guidelines for interventions that address both stakeholder groups' needs. This paper specifically focused on validating ED providers' needs in the ED diagnostic process and their views on technological support systems. Through 6 PD sessions involving 17 frontline ED providers (n=9, 53% physicians and n=8, 47% nurses), we collected and analyzed session transcripts, design sketches, and field notes. Our analysis revealed 4 key themes that encompassed patient needs, coping strategies for

XSL•FO
RenderX

common challenges, and design recommendations for future technological interventions.

Design Idea Generation Phase for ED Care Interventions

This PD study is part of a larger research project that aims to study diagnostic errors during the ED care process that involves multiple stakeholders, including patients, informal caregivers, nurses, and physicians. Before our PD study with ED providers, the research team interviewed 17 frontline ED providers (n=6, 35% physicians and n=11, 65% nurses) to better understand their experiences and challenges during the ED diagnostic processes. From the interviews, we identified difficulties, emerging patterns of complaints, and general levels of satisfaction with different aspects of the care process (Table 1).

Table 1. List of identified problem categories from our previous study's [50] emergency department (ED) health care provider interview data.

Roles	Problems
Nurses	• Missed patient reassessments due to demanding clinical workload. • Limitations in tracking patient reassessment timing and clinical status changes.
Physicians	• Clinicians' high levels of stress can impede decision-making and focus on ED work. • Physicians' high cognitive load may interfere with their ED diagnosis work. • Lack of decision support tools that aid in diagnostic decision-making for increased accuracy. • Difficulty for ED providers to access scattered patient history in a concise and easy-to-read format.
Both nurses and physicians	• Insufficient communication between physicians and nurses about orders and next steps in patient care and a lack of EHR[a] support for such communication. • Insufficient communication between physicians and nurses about the patient's diagnosis and no established opportunity to discuss diagnoses before discharge. • Lack of notification and information about incoming patients with critical care needs. • Acuity level differs between nurses and physicians. Physicians sometimes have to reassign acuity levels mentally.

[a]EHR: electronic health record.

In addition to the known problems, such as ineffective care team communication, ED providers faced challenges in patient assessment and prioritization before triage, decision-making for patient disposition, and high cognitive load among ED providers. On the basis of the findings, the research team brainstormed numerous design ideas for each problem category, focusing on frontline nurses and physicians. We then merged the design ideas based on feasibility and usefulness. Finally, we narrowed the list and finalized the 9 most effective intervention ideas (Table 2). Each intervention idea aimed to address at least 1 problem category.

On the basis of the insights from our interview study, we generated many technological intervention ideas by conceptualizing potential technological interventions that could address the challenges identified in the ED diagnostic process. In addition, we conducted an extensive literature review to understand the status quo, such as the types of technological interventions that had been explored in previous studies and the ED providers' perspectives on adopting technologies in their workflow (eg, concerns and considerations). These efforts led to the initial generation of >80 intervention ideas, which were then iteratively and collaboratively refined by 3 research team members (WS, JL, and CZ) and 2 medical domain experts. This refinement process resulted in the development of 9 ED provider-facing intervention ideas (Table 2). These interventions were subsequently used to inform and guide participant discussions during PD sessions, which we described in the PD Session Procedure section. To visually illustrate each intervention, we created storyboards for the intervention ideas. A storyboard describes an example of an ED provider facing one or more challenges mentioned above and using the intervention to mitigate the problem (see Figure 1 for a sample storyboard).

XSL•FO
RenderX

Table 2. Technology interventions produced in this study along with questions for participant discussion. Participants were asked to rank them based on their preferences and feasibility.

Technology intervention	Description	Sample of lead question used for the discussion of patient need
TI1: machine learning technology for diagnosis support	This intervention introduces an AI[a]-driven tool designed to assist ED[b] health care providers in diagnosing patients more accurately and efficiently. The system operates by analyzing a combination of patient symptoms, health history, and real-time data, such as nurses' notes and laboratory results.	"Do you wish you would have more time to spend on complex diagnoses?"
TI2: diagnostic safety dashboard	This intervention introduces a diagnostic safety dashboard designed to assist ED physicians in prioritizing patient care by visualizing risk levels in a clear, color-coded format.	"Do you find it difficult to prioritize patient care mentally based on patient conditions and risk factors?"
TI3: predischarge team huddle	This intervention idea evaluates patient charts for high-risk diagnostic errors before discharge (eg, general diagnosis and other criteria). If flagged, it prompts the care team to initiate a huddle, during which ED providers can review a summary of patient information and assign diagnoses together. This collaborative process ensures alignment on the final diagnosis, reducing the likelihood of errors and improving patient safety.	"Do you feel that more communication is needed between members of the care team about the patient's diagnosis?"
TI4: visual timeline of the patient's ED visit	This intervention enhances physician-nurse communication and EHR[c] support by visualizing the patient's visit timeline. Nurses can easily access physicians' orders, notes, and the next steps for patient care, along with key information, such as check-in details, vitals, PERC[d] scores, and CT[e] scan results, all in timeline format. This information is accessible both at the nurses' station and at the patient's bedside on screen, ensuring that everyone stays informed about orders and the next steps.	"Do you struggle to stay up to date on orders and decisions being made for a patient?"
TI5: patient history aggregator	This intervention addresses the challenge of accessing medical records for new patients with complex medical histories. The PHA[f] integrates a fingerprint scanner to ensure patient safety and accurately match the patient to their health records from previous hospitals. This allows ED providers to quickly access crucial information about the patient's heart disease and other past conditions, ensuring informed and efficient diagnostic decisions.	"Do you find it difficult to review a patient's history in a summarized format, especially when it's scattered across multiple EHR systems?"
TI6: EHR-prompted diagnostic pause	This intervention features an EHR-integrated diagnostic pause to reduce errors in complex cases. When a patient presents with symptoms such as abdominal pain and fever, the initial ED provider records their diagnosis. For complex cases, the EHR alerts the next ED provider to independently evaluate the patient before viewing the first diagnosis, which remains hidden until completion. If diagnoses differ, the system recommends team discussion. The alert links to relevant patient information for easy reevaluation.	"Is there a higher risk of diagnostic error for more complex cases?"
TI7: real-time EMS[g] information via smart glasses	This intervention addresses the challenge of unexpected critical patient arrivals by enhancing communication between EMS and the ED care team. Using smart glasses, EMTs[h] can initiate a real-time video call with the ED care team, including physicians and nurses. This allows the care team to view the patient's condition from the EMT's perspective and receive critical information en route to the hospital.	"Do you wish you had more information about when patients are being transferred to the ED, and what sort of preparation is needed?"
TI8: reassessment reminder door flag	This intervention introduces a reassessment reminder system to ensure timely and efficient monitoring of patient conditions. When a time-sensitive reassessment, such as checking vital signs or evaluating treatment effectiveness, is due, a display monitor outside the patient's room flashes a green light as a visual alert. The monitor also displays key details, such as the medication administered and the task that needs to be completed. Nurses can see the flashing light from the nurses' station or other areas, prompting them to respond.	"Are you tired of relying on your memory and constantly looking at the clock in order to complete nursing tasks such as patient reassessments every hour or so?"
TI9: stress management tracking	This intervention introduces a stress management tracking system to monitor and support the well-being of health care providers during demanding shifts. The system includes a wearable device that tracks indicators, such as heart rate and sleep cycles. When abnormal stress levels or poor sleep patterns are detected, the device notifies the wearer with an alert, accompanied by a message suggesting a break or rest. This ensures that ED providers' health is prioritized, reducing stress and fatigue while improving focus, performance, and overall cognitive function during their shifts.	"Have you ever felt so stressed that you were not able to focus or found it difficult to make decisions?"

[a]AI: artificial intelligence.

[b]ED: emergency department.

[c]EHR: electronic health record.

[d]PERC: pulmonary embolism rule-out criteria.

[e]CT: computed tomography.

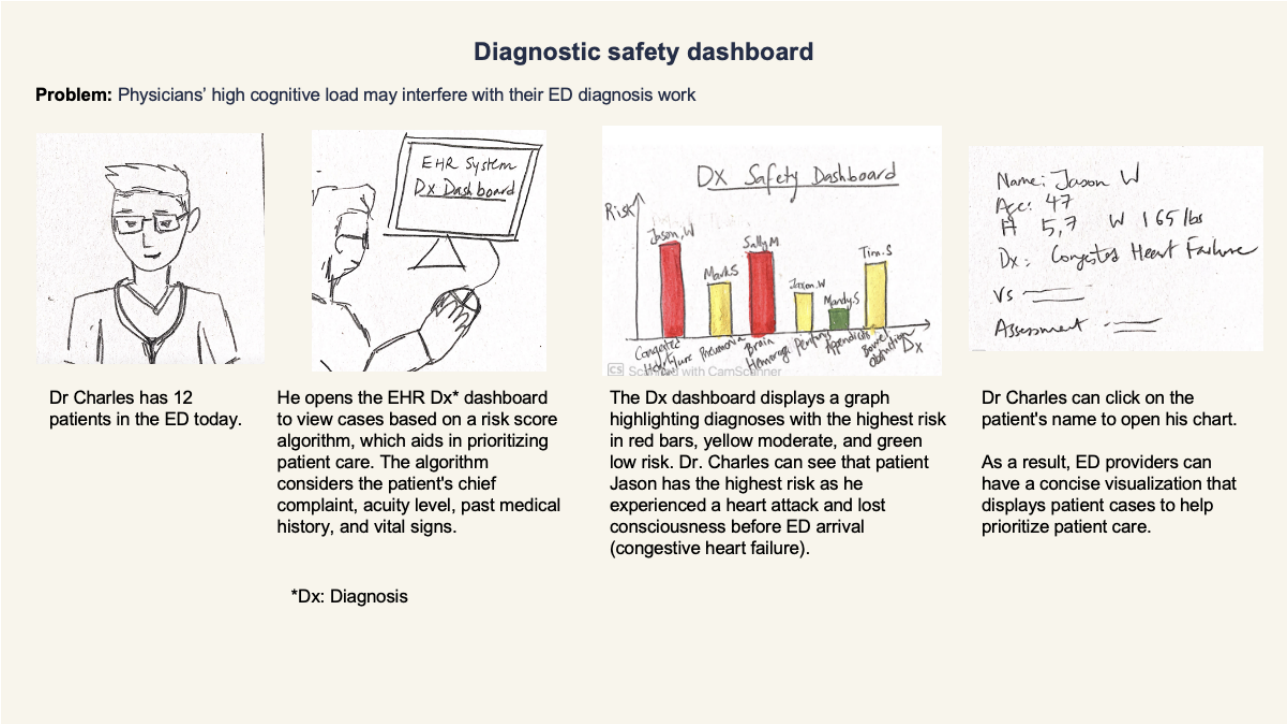
XSL•FO
RenderX

[f]PHA: patient history aggregator.

[g]EMS: emergency medical service.

[h]EMT: emergency medical technicians.

Figure 1. A sample storyboard illustrating how an interactive dashboard for patient risk scores can support the diagnosis process for emergency department (ED) health care providers. EHR: electronic health record.



Participant Recruitment

Participants in the study were recruited from 2 large health care organizations in the United States. Organization A is a university-affiliated medical institution in the Midwest, and organization B is a renowned, comprehensive medical center also located in the Midwest. Eligible participants included ED physicians and nurses who had worked in the ED setting for at least 1 year. A study coordinator contacted potential participants via email or in person on the day of their shift. Once eligible ED providers expressed interest in participating, we obtained electronic informed consent by emailing the form to them before each session. If participants did not complete the form in time, we provided a paper copy at the start of the session for in-person completion. All PD sessions were audio recorded and transcribed for data analysis. In total, 17 clinicians (n=9, 53% physicians and n=8, 47% nurses) participated in 6 PD sessions.

PD Session Procedure

We conducted a total of 6 PD sessions. Each session consisted of 3 to 4 participants with the same or similar roles (eg, nurses or physicians) to mitigate the potential influence of power dynamics on the expression of opinions [51,52]. The participants' characteristics varied in experience, from novices to experts with >30 years of experience in the field, as detailed in Table 3. The sessions were conducted either in person or online via Zoom (Zoom Communications, Inc), depending on the participants' preferences and availability. We provided colored pens, paper, and sticky notes for in-person sessions, while for virtual sessions, we used slide presentations and a Miro board (Miro, Inc) as interactive whiteboards for the participants' design collaboration. In addition, 2 to 3 researchers were present at each session, taking detailed notes.

Each PD session lasted up to 120 minutes and comprised 3 major parts: intervention idea discussion, individual design, and group design. Specifically, the first part focused on the presentation and discussion of the 9 intervention ideas developed by the researchers. After presenting each idea, we used leading questions to assess the perspectives of ED providers (eg, needs, opinions, perceptions, and expectations) for each intervention idea and allowed them to rank the intervention ideas (Table 2). In the individual design activity, participants selected 1 or 2 intervention ideas that most resonated with their needs based on their personal perspectives and professional experiences. After choosing their preferred intervention ideas, they were asked to further improve the design by modifying or adding new elements and features or to design their own new intervention. For the group design activity, participants were asked to select a top intervention idea as a group and work together to synthesize their individual design ideas and develop a final design solution (see Multimedia Appendix 1 for sample group designs). Participants engaged in discussions on the salient features, pros and cons, and concerns associated with their final group design. Each session was concluded by eliciting participant feedback to improve future sessions.

XSL•FO
RenderX

Table 3. Demographic information of study participants for each design session.

Session	Format	Affiliation	Occupation	Participant ID	Experience (y)
Session 1	Remote	Organization A	Physicians	Participant 1	20
Session 1	Remote	Organization A	Physicians	Participant 2	5
Session 1	Remote	Organization A	Physicians	Participant 3	16
Session 2	In person	Organization A	Physicians	Participant 4	19
Session 2	In person	Organization A	Physicians	Participant 5	31
Session 2	In person	Organization A	Physicians	Participant 6	2
Session 3	In person	Organization A	Nurses	Participant 7	7
Session 3	In person	Organization A	Nurses	Participant 8	15
Session 4	Remote	Organization B	Physicians	Participant 9	15
Session 4	Remote	Organization B	Physicians	Participant 10	26
Session 4	Remote	Organization B	Physicians	Participant 11	15
Session 5	Remote	Organization B	Nurses	Participant 12	8
Session 5	Remote	Organization B	Nurses	Participant 13	9
Session 5	Remote	Organization B	Nurses	Participant 14	20
Session 6	Remote	Organization B	Nurses	Participant 15	30
Session 6	Remote	Organization B	Nurses	Participant 16	3
Session 6	Remote	Organization B	Nurses	Participant 17	2

For the in-person sessions, we captured audio recordings and took photographs of the participants' sketches and designs. For virtual sessions, video recordings were made via Zoom, and participants' digital creations were also recorded for further analysis. In addition to the recordings, session facilitators took field notes throughout the sessions for data analysis.

Data Analysis

We used transcripts as the primary data source, and all other collected data (eg, observation notes) were used as supplementary sources. We also incorporated the co-design activity sketches as supplementary data to validate what participants discussed during the session and clarify participants' design suggestions.

After organizing the collected data, 2 researchers open-coded the transcripts of the first 3 PD sessions. The researchers also analyzed content from participants' sketches and design illustrations, converting these visual elements into textual descriptions to facilitate thorough examination. Each sketch or digital artifact was meticulously analyzed to extract key features, such as design elements, functional aspects, and underlying needs. The codes were compiled into a growing codebook organized by their corresponding intervention or put into a general category. On the basis of this evolving codebook, the remaining transcripts were then coded. If any new codes were discovered, they were either added to the codebook as a new code or incorporated into the definition of an existing, similar code.

We then used affinity diagramming [53] to identify the emerging themes from the codes. We used the Miro board for affinity diagramming. Similar codes were first grouped together into subthemes specific to each intervention. Then, a team of researchers engaged in a series of discussions to build larger, emerging themes that spanned across the different interventions. For example, a larger theme of "communication among the care team members" contained various subthemes, such as "piece together information about a patient and keep every care team member on the same page." Codes that were ambiguous or needed further support were refined by referencing participant quotes from transcripts and participants' sketches from PD sessions. After multiple group discussions, we identified 4 themes that described ED providers' challenges in the diagnostic process.

Ethical Considerations

We complied with the following ethical considerations. First, we obtained consent from all participants before the study sessions, which were approved by the multisite institutional review board at the University of Michigan (HUM00156261). Participants were informed that their participation was voluntary, the sessions would be recorded, the collected information would be deidentified and protected in secured storage, and they would be compensated with US $100 for their participation. Second, we ensured that the collected transcript data were deidentified by replacing participants' names with pseudonyms, and that their faces in photos or videos were blurred. All collected data are stored in secure storage that requires the university accounts for access. Third, all the researchers completed research compliance training, such as the Collaborative Institutional Training Initiative or the Program for the Education and Evaluation in Responsible Conduct of Research, to become trained on best practices and ethical considerations for interacting with ED providers and patients.

XSL•FO
RenderX

Results

Overview

On the basis of the analysis of collected data, we identified top intervention ideas that participants preferred, 4 key areas in the ED diagnostic process requiring technological support, and major concerns regarding the integration of advanced technologies into dynamic ED workflows.

Throughout all sessions, participants collectively favored the following interventions to enhance the existing ED diagnosis process (Table 2): TI1 (machine learning technology for diagnosis support), TI2 (diagnostic safety dashboard), TI3 (predischarge team huddle), and TI4 (visual timeline of the patient's ED visit). Participants believed that these intervention ideas could improve crucial aspects of the ED diagnosis process. Specifically, TI1 could broaden access to comprehensive patient information and reduce potential errors. TI2 might simultaneously manage multiple patient cases, optimize prioritization, and rapidly identify high-risk patients. TI3 could address communication and understanding gaps between physicians and nurses regarding diagnoses, while TI4 aims to bridge the communication divide and keep patients informed about their care progress.

Furthermore, we reported the key areas of improvement in the diagnostic process, which indicated the challenges faced by ED providers in the diagnostic process. We present 4 prominent aspects extensively discussed among participants in the subsequent sections.

Effective Integration of Patient Information for Timely and Accurate Diagnosis

Participants often faced challenges with incomplete, hard-to-find, or misleading patient information, which can inadvertently lead the medical team down an incorrect diagnostic pathway. Incompleteness can result from prehospital care providers failing to relay all necessary information at handoff or integration issues among health care information systems. Assessing relevant information is problematic when stored using varied and repetitive terminologies or buried in multilevel systems. Misleading information typically occurs with objective data, such as vital signs. For example, external stress or the use of an incorrectly sized cuff can result in elevated heart rates or blood pressure readings, potentially skewing risk assessments.

To access comprehensive patient information effectively, participants mentioned that both historical information (eg, past medical history, long-term medications, patients' care preferences, and social determinants) and ad hoc information (eg, laboratory test results, vital signs, and clinical observation notes) should be integrated to make an accurate diagnosis. For instance, Participant 16 suggested how TI5 (patient history aggregator) could be integrated with existing care systems to help ED providers in the diagnosis process as follows:

> *If we could have that [History Aggregator] page already incorporate all the other systems where this patient has received care, I see that as helpful. It's kind of like, in my mind, redefining the chart review tab in a more encompassing form.* [Participant 16]

Some participants also highlighted the dynamic changes in physiological values when redesigning TI1 (machine learning technology for diagnosis support). To maintain accurate risk assessments, it is critical to continuously update data reflecting the dynamic changes in patient conditions, especially for those patients with critical illness whose health status can rapidly evolve. This suggestion showed participants' desire to receive updates about the patient's status for better management, as many patients are coming in and out of the ED.

Some participants suggested that medical information should be displayed in a simple, clear, and efficient manner, such as streamlining all necessary information onto 1 page to avoid spending time searching and organizing it piece by piece. Participant 5 highlighted the importance of information presentation for TI4 (visual timeline of the patient's ED visit) as follows:

> *It's information that we just want to be displayed in a more graphic, simpler form where it's not so busy. You open it up, it's one snapshot, and you click on which part of it you want to do now...It's just organized in a better way, in a simpler way.* [Participant 5]

Participant 5 envisioned that this approach would enable health care providers to efficiently focus on critical details, such as a patient's historical data, directly relevant to their chief complaint.

Optimizing Patient Prioritization to Improve the Triage and Care Process

Another important area to improve was related to prioritizing patients. The current triage process may lead to the improper prioritization of patients with critical needs. Triage in the ED can be influenced by many factors, including triage nurses' clinical knowledge and judgment, the clarity with which patients describe their symptoms, and the availability of patient records or medical histories. In addition, the Emergency Severity Index (ESI) classification criteria, which classify and prioritize patients based on the severity of their disease and the estimated number of resources needed for their care, can fail to account for dynamic changes in a patient's condition. For example, there was no adaptation of the ESI score despite clear signs that the patient had become more critically ill. For these reasons, there is a risk of patients being misclassified and inappropriately prioritized, which may result in those with severe conditions facing longer wait times.

To address those issues in the triage process, our participants expected the AI-driven prioritization system to analyze various types of information, including contextual (patient's relevant medical history and symptoms), situational (dynamic changes in patients' symptoms and length of waiting time), and human factors (clinician's knowledge and experience). This ensures timely attention to critical cases and improves ESI's shortcomings. For instance, when discussing idea TI2 (diagnostic safety dashboard), participants stated that the algorithm could consider different and comprehensive factors

XSL•FO
RenderX

to identify higher-risk individuals, such as those with a history of asthma or cardiac problems. A participant noted as follows:

> *The algorithm looks at the chief complaint, acuity level, and past medical history. I think that helps as well. For example, if it's a younger person with a syncopal episode and you find out they have a previous cardiac history, that would make them a higher risk than, say, a 30-year-old who passed out for another reason. I like how this algorithm considers the whole patient and takes that into account.* [Participant 12]

In addition, many participants valued the explicit representation of different patient prioritizations through various colors (eg, red, yellow, and green) and symbols (eg, an alarm bell and a smiley face) because it could enable ED providers to identify which patients require immediate attention visually quickly. For instance, Participant 8 preferred color coding of data visualization for immediate understanding of priority as follows:

> *I like the color because then they could look at that and say that they're at high risk. If they're above five risk scores then they are red or something, and then they're like, "Oh, I have all these priority 2s as Joe is sicker."* [Participant 8]

Enhancing Patient Engagement in the Diagnostic Process

Our participants mentioned that it could be challenging for ED providers to receive timely diagnostic feedback and updates on changing symptoms from patients due to time constraints and heavy workloads in the ED. ED providers often rely on brief interactions or 1-way communications to gather information, which may lead to an incomplete understanding of patients' evolving symptoms. In addition, patients frequently feel anxious and frustrated due to the lack of ongoing updates about their condition, which can lead to increased complaints and misunderstandings.

Many participants agreed that informing patients about the status and progression of their care could help manage their expectations and alleviate anxiety. They expected patients to feel reassured that ED providers would still take care of them when patients understood the next steps in the diagnosis process. Participant 2 stated as follows when discussing idea TI4 (visual timeline of the patient's ED visit):

> *The patient can visualize what's been done and what they're waiting for. It sort of wards off that sentiment of, "I went to the ER, and they didn't do anything," and also gives them an idea of what the next step is.* [PD-2]

By sharing the ED process steps, participants anticipated more patient engagement and having more time to focus on effective patient communication.

Participants mentioned they wanted easy-to-use tools to help receive more accurate, timely updates directly from patients to aid the care process. Nurses' work could be expedited by having patients more actively involved in their care and providing information. For example, participants suggested that the idea TI8 (reassessment reminder door flag) could include a feature allowing patients to undertake a self-reassessment of their current condition and medication reactions and relay their results and updates to nurses, potentially via an automated prompt and chat system on a tablet or phone. If a nurse records the administration of fentanyl at noon, the system might automatically inquire about the patient's pain levels 20 minutes later. Then, patients could interact with the nurse, who can gauge the treatment's effectiveness, as noted by a participant:

> *I wonder if there's a way for the patient to enter a self-reassessment like they have their own little iPad and can share how they're doing, whether the pain medication helped their pain rating...it gives the patient the opportunity to provide feedback.* [Participant 9]

Our participants also mentioned that engaging patients in diagnostic discussions could be crucial to reducing potential diagnostic errors. This could allow ED providers to elucidate the logic behind a diagnosis, articulate how tailored care plans were developed for individual patients, and seek the patient's perspective regarding the proposed plan. In addition to the illustrated functionality of increasing care team communications to reduce diagnosis errors from ideas TI3 (predischarge team huddle) and TI6 (electronic health record (EHR)–prompted diagnostic pause), participants suggested empowering patients to contribute supplementary information and express different viewpoints that might not have been considered, to reduce diagnosis errors. For instance, Participant 3 emphasized that patients should have the opportunity to share their perspectives on ED providers' reasoning as a way to promote patient engagement in diagnostic discussions as follows:

> *If the patients heard the diagnostic reasoning, you [clinicians] would be giving them an opportunity to correct you [clinicians].* [Participant 3]

Care Coordination Among the ED Team Members

Through the sessions, we identified that the ED care team faced 2 main communication issues. First, each member of the ED care team has only partial information, leading to a disjointed and fragmented understanding of the patient's case, primarily due to interactions at different stages of care. Patients move from 1 ED provider to another, and no one fully knows all the details of the patient's condition. Furthermore, its interpretation can differ even when information is consistently shared among team members. This variation in understanding can be attributed to individual backgrounds, areas of expertise, or perspectives. Second, physician-to-nurse communication could sometimes become ineffective, especially regarding discharge and diagnostic reasoning information. For example, physicians sometimes fail to communicate the rationale behind diagnoses, orders, or care plans to nurses. This issue can impede nurses' understanding of physician decisions and their ability to respond accurately to patient queries. Furthermore, participants acknowledged that this communication issue may be derived from both sides, as noted by participant 7:

> *[Physician-nurse communication] is a mutual problem, it's not a one-sided issue.* [Participant 7]

XSL•FO
RenderX

All participants agreed on the importance of maintaining a consistent and uniform understanding among care team members, especially for complex patient cases. Engaging more care team members in discussions could be advantageous. When discussing idea TI3 (predischarge team huddle), participants noted that it would offer an excellent opportunity for comprehensive information gathering from different team members and roles to minimize potential errors. A participant highlighted as follows:

> *You would get everyone on the same page and give the patient a chance to ask questions. At the last minute, if someone asks, "Well, what about this, or what about that?" everyone is on the same page, which I like.* [Participant 2]

Most physicians and all the nurses mentioned the need for physicians to communicate more detailed information with nurses, including the reasoning behind diagnoses, orders, and care plans, to enable nurses to understand the overall situation comprehensively. When discussing idea TI4 (visual timeline of the patient's ED visit), physicians believed that this technological intervention could be instrumental in keeping nurses informed about care plans and the rationale behind physicians' orders in a timely manner. This enhanced transparent communication could empower nurses to answer patients' queries more effectively, reducing the need for ambiguous or noncommittal responses. Nurses, in particular, advocated this because they wished to better explain physicians' orders and diagnoses to patients, as one participant noted:

> *Sometimes I go into the patient room and say, "I think they [the physicians] ordered this because they're probably looking to see if you [the patient] have this issue." It would be nice to say confidently, they ordered this because they're looking to see that you have this issue.* [Participant 8]

Among our participants, some care team members might hesitate to point out potential errors, from junior members to senior members or nurses to physicians. For example, nurses might be reluctant to bring up potential diagnostic errors to physicians due to a perception that physicians, being more knowledgeable and experienced in diagnosis, are responsible for treatment decisions. However, the participants believed that some technologies, acting as second opinions, might help address the power dynamic issue. For example, when discussing the idea TI1 (machine learning technology for diagnosis support), nurses expressed that the list of diagnoses generated by machine learning tools could empower them to engage in discussions with physicians regarding potential misdiagnoses significantly when the physician's conclusion deviates from the algorithm's suggestion, as noted by a participant:

> *It [the machine learning technology tool] also brings up good discussion points with the physician...why are we pursuing this route...or why we're deviating from a particular diagnosis.* [Participant 14]

Concerns of Integrating Advanced Technologies in ED

Overview

Our 9 intervention storyboards described how emerging technologies can potentially be designed to support the ED diagnostic process. While the storyboards were primarily used to validate ED providers' needs, they were also used for discussions about potential issues of integrating such technological interventions. This section presents participants' concerns regarding integrating advanced technologies in ED, including the safe use of AI, privacy, and additional cognitive and operational workload.

Safe Use of AI-Empowered CDSSs

With the growing application of AI techniques in health care, more than half of our intervention ideas were inspired by AI-empowered tools. Thus, many participants expressed concerns about the AI's potential inability to consider different care contexts and individual nuances of each patient case in analyzing and presenting various information. This concern was mainly raised during group design activities on several AI-driven interventions, such as TI1 (machine learning technology for diagnosis support), TI2 (diagnostic safety dashboard), and TI4 (visual timeline of the patient's ED visit). Participants particularly mentioned that the same test result can be interpreted differently based on the context. For instance, a particular urine analysis result can indicate different health conditions depending on whether the patient has a positive or negative ultrasound result, as noted by a participant:

> *The same urine analysis result means something totally different if you tell me that the patient has an ultrasound that's clearly positive than if they have an ultrasound that's clearly negative, and it can have the exact same result. And that's the problem. It comes in context.* [Participant 4]

Participants suggested that AI systems should account for various factors and contextual cues, such as interpreting test results with other diagnostic information and medical history. However, it may not fully match human ED providers' ability to perceive subtle cues and insights, which could be beyond the scope of computers or AI systems. In addition, participants suggested that the AI systems should apply different weights to different diagnostic factors (eg, laboratory results and medical history) because certain factors weigh more than others, or the system should adjust weights based on dynamic emergency severities when calculating risk scores.

In addition, participants underscored the imperative need for rigorous testing and validation before implementing any AI tool in their existing work practice. They believed that an AI tool must prove its reliability and efficacy to gain the trust of medical providers, as noted by a participant:

> *Critical for it to be, the tool really has got to be tested and robust before it's implemented. Otherwise, no one's going to entertain the idea of adopting it because it's just going to create extra work without perceived value.* [Participant 10]

XSL•FO
RenderX

Participants also expressed concerns that an overreliance on technology could potentially lead to a degradation of human clinical skills and emphasized the balance between such emergent technology use and human skills. For instance, when discussing the idea TI1 (machine learning technology for diagnosis support), there was a worry that ED providers might depend too heavily on technology and neglect their assessments. Participant 13 expressed her concern about potential overreliance on AI tools as follows:

> *The only concern I can think of right now is...becoming too reliant on AI...If any residents or doctors would start to rely on it too much...you'd hope not, you'd hope they do their own assessments.* [Participant 13]

As a potential solution to mitigate such overreliance, some participants suggested that they would like to incorporate AI-powered CDSSs into their on-the-job training. This integration could allow ED providers to learn from these tools and understand how to collaborate with them effectively without overly relying on AI outputs while maintaining their clinical assessment skills.

Privacy Concerns Toward Emerging Technologies

Participants repeatedly mentioned privacy concerns related to new advanced technologies. They raised issues about potential violations of the Health Insurance Portability and Accountability Act and patient confidentiality, emphasizing the need to ensure adherence to patient privacy laws and regulations. For example, TI5 (patient history aggregator) allows easy access to patient history by scanning patients' fingerprints. Many participants expressed concerns about collecting sensitive biometric data (ie, fingerprints) from their patients. Participant 16 brought up the issue of data storage that may lead to a privacy breach, as follows:

> *If it is based on fingerprint, where is the information being housed? Is it a cloud database? That is universal? Is it an extension of the electronic health record? Where does it belong? I think it will be a question.* [Participant 16]

Participants' such concerns about privacy highlight the importance of transparent communication regarding data storage practices, robust security measures, and clear protocols for accessing sensitive patient information to foster trust in emerging health technologies.

Extra Workload and Burden to Make Technology Fit Into Their Existing Workflow

Participants raised concerns about the potential for certain technological implementations to interfere with the established ED workflow, possibly creating extra waiting time for patients and adding extra cognitive and operational workload for ED providers. One notable instance occurred during the discussion of the idea TI3 (predischarge team huddle). Some participants expressed that integrating such a system could inadvertently delay or create a backlog in the patient discharge process. Given the current ED workflow, it could be challenging to convene all care team members due to their individual duties and potential busyness. A participant explained this concern by stating as follows:

> *When you have a patient who is itching to get out the door [being discharged], and you say, "while we're waiting for this person to show up or we need this other team member [to discuss again]," it can create a lot of delay and backlog if now you have to round to discharge everybody.* [Participant 2]

Another example arose during the discussion on the TI6 (EHR-prompted diagnostic pause), which highlighted concerns that its implementation could disrupt the typical ED workflow by introducing a "parallel" evaluation process. In this context, "parallel evaluation" means that multiple ED providers might independently diagnose the patient, leading to repetitive and time-consuming efforts. This could result in a prolonged diagnostic process and potentially delay treatment, as noted by a participant:

> *I think the scenario would require a change in our current flow, where it's sequential in terms of evaluations that are done of a patient...To one where there's a parallel evaluation, which I think becomes very inefficient and for a patient, can be quite redundant as they have to repeat the story multiple times.* [Participant 10]

Discussion

Design Implications

Overview

Our findings show that the ED diagnostic process can be improved into 4 main areas: information integration, patient prioritization, ED provider-patient communication, and care coordination. We also summarized the key concerns of ED providers in integrating advanced technologies in the ED, including the safe use of AI, privacy, and the additional cognitive and operational workload. Drawing from these findings, we discuss potential design implications of future ED-centered intervention tools for improving the ED diagnosis process.

Creating Standardized Communication Procedures

We observed that there is significant demand for improving communication between ED providers (the idea TI3 [predischarge team huddle] was ranked third by participants, and TI4 [visual timeline of patient's ED visit] was ranked fourth by participants). Information sharing between ED physicians and nurses is often inadequate, limiting nurses' comprehensive understanding of patient conditions. In addition, varying levels of understanding about the patient exist among different stages in the care team, further complicating the communication process. These findings highlight the importance of establishing standardized communication procedures to ensure a unified understanding of patient conditions among team members. We suggest creating more standardized communication procedures in future technical tool designs. Building on previous insights from the design of digital checklists for trauma resuscitation [54], we recommend implementing an automated checklist

XSL•FO
RenderX

system that includes critical patient information, such as diagnosis, medication orders, and care milestones. This checklist should be integrated with a timeline that automatically captures details of when each piece of information was added and when action needs to be taken, ensuring that all team members are aware of current in-progress and upcoming responsibilities. in addition, structured communication features, such as mandatory read-backs of critical information, regular briefings at shift changes, and digital alerts for key updates, can further enhance understanding and collaboration. Such tools are particularly helpful for nurses, who often coordinate various aspects of patient care and benefit from clear, concise, and timely information to make informed decisions and provide the best possible care. This potential application direction also echoes the previous study on the necessity of applying new technological communication interventions to simplify communication among various roles within the ED [55].

Incorporating Technology-Based Second Opinions

We identified power dynamics as a critical issue in the diagnostic process. Such skewed interactions are prevalent between junior and senior members, as well as between nurses and physicians. Hesitation in pointing out potential errors and reluctance to accept feedback is not uncommon, and it could have severe consequences (eg, increased incidence of medical errors and significant patient safety risks) in fast-paced and critical medical fields such as EDs [51,52]. Although more collaborative approaches, such as education and mentoring, have been proposed to address the issue of power dynamics [56], we found that this issue has not yet been fully resolved. Notably, our findings indicated the potential of technological interventions in mitigating these power dynamics. For instance, the timeline in idea TI4 (visual timeline of the patient's ED visit) could allow nurses to understand the rationale behind care plans, enhancing their ability to communicate with physicians actively, thus promoting equal and open communication within the team. Furthermore, as mentioned in idea TI1 (machine learning technology for diagnosis support), second opinions generated using machine learning tools could boost the confidence of junior physicians or nurses in discussing potential misdiagnoses with senior members or physicians in the care team. This echoes and expands on past research [57] about the second opinion in physician-patient relationships, demonstrating it to be an effective way to promote open, positive communication and to enhance satisfaction with clinical decisions. By providing second opinions and facilitating information sharing, technological interventions could help overcome the barriers in the traditional power structure of medical roles, potentially improving overall care quality and patient safety. Future work could involve designing and implementing such information tools in ED settings to evaluate potential impacts on the ED diagnostic process. For example, in the new ED medical display system, each team member could autonomously select CDSS reports that automatically highlight decision details differing from the primary physician's, with junior members and nurses also accessing these for reference and discussion as a second opinion.

Empowering Patients With Information Transparency

We found a demand for improved communication between ED providers and patients through timely information sharing, transparency with patients, and patient empowerment. Previous studies show that insufficient communication between patients and ED providers in the ED leads to dissatisfaction with their care and potential diagnosis errors [12,58]. Factors contributing to their communication issues include patients' and ED providers' conflicting perceptions of the ED, a lack of patient information, and physicians' diagnostic time pressures [12,58,59]. Our study further reveals the specific manifestations of insufficient communication in the ED between patients and ED providers: (1) patients are not fully informed about the treatment process; (2) lack of opportunities for patients to provide feedback and update their conditions to ED providers actively; and (3) absence of processes for joint decision-making and care planning between ED providers and patients, as well as effectively involving patients in diagnostic discussions. These communication deficiencies can lead to patient anxiety and an increased risk of misdiagnosis. Therefore, we suggest that future technological interventions should focus on enhancing timely information sharing and transparency with patients and significantly empowering patients' agency. This involves promoting active patient participation in communication, even in joint decision-making and care planning. Previous research [36] discussed a digital ED waiting room history-taking tool that uses reasoning engine functions to gather patient information. It has been proven to improve communication and understanding between patients and ED providers. Building on this foundation, we suggest developing a mobile app that allows patients to promptly report changes in their symptoms or express concerns using real-time data reflecting their current condition and personal feelings during their ED visits. The app would also integrate features that encourage patient involvement in diagnostic discussions. Such features could include a symptom timestamp for patients to log details of their symptoms as they occur and a communication chat window that allows for direct dialogue with ED providers.

Optimizing Advanced Technology Integration in the ED

Our research highlighted specific considerations for integrating advanced technologies into the ED context: minimizing additional workload and cognitive burdens for ED providers and using high-fidelity simulations. Unlike other general clinical practices where there is more time to adapt and use new technology, our findings indicated that ED providers in the fast-paced ED environment may not have the time to try out, thoroughly examine, and use these technologies. Furthermore, they were reluctant to change their existing workflows to accommodate new technologies, fearing the risk of additional workload. During the design of information displays, our ED providers preferred concise and practical interfaces, capturing essential information immediately rather than increasing their cognitive load by requiring them to understand the meaning of every icon and content. In addition, participants considered idea TI6 (EHR-prompted diagnostic pause) ineffective and redundant, despite its merit of catching potential diagnostic errors, as it does not fit their existing workflow. Similarly, during the discussion of idea TI3 (predischarge team huddle), participants

XSL•FO
RenderX

acknowledged the critical need for care team members to be on the same page regarding information sharing upon patient discharge. However, they still harbored concerns that the intervention would not align well with their current workflow or the realistic situation where patients would not want to stay longer and wait for a reevaluation before their discharge. Therefore, building on previous research [60] that emphasized the need for efficiency in the design of ED technological interventions, we recommend that future designs and implementations of ED technological interventions consider the fast-paced, high-pressure nature of the ED. This involves prioritizing efficiency and focusing on avoiding additional workload and cognitive burdens for each ED provider. For example, AI-driven decision support tools should be overseen by designated care team members who monitor AI-recommended diagnostics and conduct departmental audits and custom improvements to reduce the burden for individual ED providers. Similarly, designing interventions, such as the idea TI3 (predischarge team huddle), should minimize the potential for increasing workload. Instead of prompting the care team to initiate a team huddle to reevaluate, the design should allow members to quickly review summary reports in the EHR to identify potential issues with minimal additional effort.

From a methodological perspective, we also recommend building on existing ED workflows using methods such as "group workshops" similar to PD or "high-fidelity team-based simulations" conducted in clinical scenarios [61,62]. These methods could allow ED team members to experience new technologies and foresee how they can be implemented in their existing work practices while gathering feedback. High-fidelity team-based simulations have been widely used in the ED for effect evaluation and training, such as improving communication among ED provider teams [63] and detecting latent safety threats in critical patients [64], proving very effective. Therefore, we suggest extending this approach to introducing new technologies in the ED to guide the integration of new advanced technologies.

Leveraging AI Tools in the ED to Enhance Diagnosis Through Dynamic Patient Data

Our research highlighted a strong demand from our participants to unleash the potential of AI tools in the ED to improve diagnosis further. We found that participants highly valued the top 2 AI-related ideas (TI1 [machine learning technology for diagnosis support] and TI2 [diagnostic safety dashboard]) because of their potential to (1) integrate a broader range of comprehensive patient information and perform multidimensional data analysis and (2) facilitate dynamic prioritization and resource allocation. Although there are multiple existing AI applications in the medical field, many of them lack consideration of multidimensional information and dynamic prioritization [65] and thus do not fully unleash the potential of AI. To best leverage AI tools in future ED settings, we suggest that when designing algorithm development for ED, the input information for these algorithms should be multidimensional and incorporate dynamic patient data. These algorithms should be capable of continuously monitoring and evolving based on real-time patient data and feedback, which can aid in enhancing diagnostic accuracy and dynamically adjusting treatment protocols. For example, when designing AI tools for patient prioritization, it is crucial to implement a system that categorizes patients based on initial information and updates its prioritization as new data becomes available (eg, laboratory results or vital signs). This approach could ensure that resources are allocated promptly to patients with escalating needs.

Limitations

A few limitations of this study should be noted. First, one of the primary limitations of this study relates to the representativeness of the participant sample. Participants were recruited from only 2 large health care organizations in the Midwest region of the United States. While these organizations serve millions and can offer significant insights, this sample may not fully encapsulate the diversity and range of experiences of ED providers worldwide. There might be differences in health care systems, cultural contexts, and operational protocols in EDs across different countries or regions. Another limitation of this research is inherent in the study design, particularly regarding the PD sessions. Each session was constrained to a duration of around 120 minutes. This limited timeframe could have impacted the depth and extent of interaction among participants, potentially affecting their comfort levels and the quality of their responses. Participants might have required more time to fully engage with the material, express their thoughts, and collaborate effectively, influencing the innovation and creativity aspects of the session's outcomes. Furthermore, some PD sessions were conducted in a virtual format. While virtual sessions offer flexibility and can include a broader range of participants, they might lack the richness of interaction found in face-to-face settings. Nonverbal cues, ease of spontaneous conversation, and the organic development of collaborative ideas are often diminished in virtual environments. The potential limitations in participant interaction and engagement due to the virtual format of some sessions might have affected the overall effectiveness and output of these PD sessions. Finally, our study is limited by not presenting functional systems but merely asking participants to imagine how the system works. Without seeing or interacting with a system, eliciting comprehensive user feedback could be challenging.

Conclusions

This work presents opportunities for improving ED provider–informed technological interventions and strategies for enhancing the diagnostic process in EDs. Our study identifies crucial enhancements in information integration, patient prioritization, ED provider-patient communication, and care coordination. The findings underscore key concerns about integrating advanced technologies into EDs. On the basis of these findings, this work discusses implications for designing future ED technological interventions, from enhancing communication to optimizing the integration of advanced technology and unleashing the potential of AI tools. Future efforts should further apply and test these design suggestions in real ED settings or continue using improved PD methods with functional systems that participants can interact with to further explore technological advancements in the ED diagnosis process.

XSL•FO
RenderX

Acknowledgments

This project was funded under grant number R18HS026622 from the Agency for Healthcare Research and Quality (AHRQ), US Department of Health and Human Services (HHS). HS is supported in part by the Houston Veterans Administration (VA) Health Systems Research and Development (HSR&D) Center for Innovations in Quality, Effectiveness and Safety (CIN13–413), the VA National Center for Patient Safety, and AHRQ (R01HS028595, R18HS029347, and R01HS029318). The authors are solely responsible for this document's contents, findings, and conclusions, which do not necessarily represent the views of AHRQ. Readers should not interpret any statement in this report as the official position or policy of the US government, AHRQ, HHS, or VA. None of the authors has any affiliation or financial involvement that conflicts with the material presented in this report.

Data Availability

The datasets generated and analyzed during this study are not publicly available because the consent form for the participants stated that the data could only be accessible to the research team.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sample group designs by participants from remote and in-person design sessions.
[PDF File (Adobe PDF File), 400 KB - formative_v9i1e68891_app1.pdf]

References

1. Asplin BR, Magid DJ, Rhodes KV, Solberg LI, Lurie N, Camargo CA. A conceptual model of emergency department crowding. Ann Emerg Med 2003 Aug;42(2):173-180. [doi: 10.1067/mem.2003.302] [Medline: 12883504]
2. Beck E, Balasubramanian H, Henneman P. Resource management and process change in a simplified model of the emergency department. In: Proceedings of the 2009 Conference on Winter Simulation Conference. 2009 Presented at: WSC '09; December 13-16, 2009; Austin, TX p. 1887-1895 URL: https://ieeexplore.ieee.org/document/5429198 [doi: 10.1109/wsc.2009.5429198]
3. Cosby KS. A framework for classifying factors that contribute to error in the emergency department. Ann Emerg Med 2003 Dec;42(6):815-823. [doi: 10.1016/s0196-0644(03)00635-8] [Medline: 14634609]
4. Carter AJ, Chochinov AH. A systematic review of the impact of nurse practitioners on cost, quality of care, satisfaction and wait times in the emergency department. CJEM 2007 Jul 21;9(4):286-295. [doi: 10.1017/s1481803500015189] [Medline: 17626694]
5. Singer AJ, Thode Jr HC, Viccellio P, Pines JM. The association between length of emergency department boarding and mortality. Acad Emerg Med 2011 Dec;18(12):1324-1329 [FREE Full text] [doi: 10.1111/j.1553-2712.2011.01236.x] [Medline: 22168198]
6. Sun BC, Hsia RY, Weiss RE, Zingmond D, Liang L, Han W, et al. Effect of emergency department crowding on outcomes of admitted patients. Ann Emerg Med 2013 Jun;61(6):605-11.e6 [FREE Full text] [doi: 10.1016/j.annemergmed.2012.10.026] [Medline: 23218508]
7. Ball JE, Murrells T, Rafferty AM, Morrow E, Griffiths P. 'Care left undone' during nursing shifts: associations with workload and perceived quality of care. BMJ Qual Saf 2014 Feb 29;23(2):116-125 [FREE Full text] [doi: 10.1136/bmjqs-2012-001767] [Medline: 23898215]
8. Croskerry P, Sinclair D. Emergency medicine: a practice prone to error? CJEM 2001 Oct;3(4):271-276. [doi: 10.1017/s1481803500005765] [Medline: 17610769]
9. Werner NE, Holden RJ. Interruptions in the wild: development of a sociotechnical systems model of interruptions in the emergency department through a systematic review. Appl Ergon 2015 Nov;51:244-254. [doi: 10.1016/j.apergo.2015.05.010] [Medline: 26154223]
10. Berg LM, Källberg AS, Göransson KE, Östergren J, Florin J, Ehrenberg A. Interruptions in emergency department work: an observational and interview study. BMJ Qual Saf 2013 Aug;22(8):656-663. [doi: 10.1136/bmjqs-2013-001967] [Medline: 23584208]
11. Chisholm CD, Collison EK, Nelson DR, Cordell WH. Emergency department workplace interruptions: are emergency physicians "interrupt-driven" and "multitasking"? Acad Emerg Med 2000 Nov 28;7(11):1239-1243. [doi: 10.1111/j.1553-2712.2000.tb00469.x] [Medline: 11073472]
12. Pun JK, Matthiessen CM, Murray KA, Slade D. Factors affecting communication in emergency departments: doctors and nurses' perceptions of communication in a trilingual ED in Hong Kong. Int J Emerg Med 2015 Dec 15;8(1):48 [FREE Full text] [doi: 10.1186/s12245-015-0095-y] [Medline: 26667242]

XSL•FO
RenderX

13. Risser DT, Rice MM, Salisbury ML, Simon R, Jay GD, Berns SD. The potential for improved teamwork to reduce medical errors in the emergency department. The MedTeams Research Consortium. Ann Emerg Med 1999 Sep;34(3):373-383. [doi: 10.1016/s0196-0644(99)70134-4] [Medline: 10459096]
14. Yoon J, Sonneveld M. Anxiety of patients in the waiting room of the emergency department. In: Proceedings of the 4th international conference on Tangible, embedded, and embodied interaction. 2010 Presented at: TEI '10; January 24-27, 2010; Cambridge, MA p. 279-286 URL: https://dl.acm.org/doi/10.1145/1709886.1709946 [doi: 10.1145/1709886.1709946]
15. Feied CF, Handler JA, Smith MS, Gillam M, Kanhouwa M, Rothenhaus T, et al. Clinical information systems: instant ubiquitous clinical data for error reduction and improved clinical outcomes. Acad Emerg Med 2004 Nov;11(11):1162-1169 [FREE Full text] [doi: 10.1197/j.aem.2004.08.010] [Medline: 15528580]
16. Carter AJ, Davis KA, Evans LV, Cone DC. Information loss in emergency medical services handover of trauma patients. Prehosp Emerg Care 2009 Aug 13;13(3):280-285. [doi: 10.1080/10903120802706260] [Medline: 19499462]
17. Laxmisan A, Hakimzada F, Sayan OR, Green RA, Zhang J, Patel VL. The multitasking clinician: decision-making and cognitive demand during and after team handoffs in emergency care. Int J Med Inform 2007 Nov;76(11-12):801-811. [doi: 10.1016/j.ijmedinf.2006.09.019] [Medline: 17059892]
18. Stiell A, Forster AJ, Stiell IG, van Walraven C. Prevalence of information gaps in the emergency department and the effect on patient outcomes. CMAJ 2003 Nov 11;169(10):1023-1028 [FREE Full text] [Medline: 14609971]
19. Cordell WH, Overhage JM, Waeckerle JF. Strategies for improving information management in emergency medicine to meet clinical, research, and administrative needs. Information Management Work Group. Acad Emerg Med 1998 Feb 28;5(2):162-167 [FREE Full text] [doi: 10.1111/j.1553-2712.1998.tb02604.x] [Medline: 9492140]
20. Okafor N, Payne VL, Chathampally Y, Miller S, Doshi P, Singh H. Using voluntary reports from physicians to learn from diagnostic errors in emergency medicine. Emerg Med J 2016 Apr 03;33(4):245-252. [doi: 10.1136/emermed-2014-204604] [Medline: 26531860]
21. Kachalia A, Gandhi TK, Puopolo AL, Yoon C, Thomas EJ, Griffey R, et al. Missed and delayed diagnoses in the emergency department: a study of closed malpractice claims from 4 liability insurers. Ann Emerg Med 2007 Feb;49(2):196-205. [doi: 10.1016/j.annemergmed.2006.06.035] [Medline: 16997424]
22. Studdert DM, Mello MM, Gawande AA, Gandhi TK, Kachalia A, Yoon C, et al. Claims, errors, and compensation payments in medical malpractice litigation. N Engl J Med 2006 May 11;354(19):2024-2033. [doi: 10.1056/NEJMsa054479] [Medline: 16687715]
23. Guly HR. Diagnostic errors in an accident and emergency department. Emerg Med J 2001 Jul;18(4):263-269 [FREE Full text] [doi: 10.1136/emj.18.4.263] [Medline: 11435359]
24. Horng S, Sontag DA, Halpern Y, Jernite Y, Shapiro NI, Nathanson LA. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. PLoS One 2017;12(4):e0174708 [FREE Full text] [doi: 10.1371/journal.pone.0174708] [Medline: 28384212]
25. Xie F, Liu N, Wu SX, Ang Y, Low LL, Ho AF, et al. Novel model for predicting inpatient mortality after emergency admission to hospital in Singapore: retrospective observational study. BMJ Open 2019 Sep 26;9(9):e031382 [FREE Full text] [doi: 10.1136/bmjopen-2019-031382] [Medline: 31558458]
26. Chiew CJ, Liu N, Tagami T, Wong T, Koh ZX, Ong ME. Heart rate variability based machine learning models for risk prediction of suspected sepsis patients in the emergency department. Medicine (Baltimore) 2019 Feb;98(6):e14197 [FREE Full text] [doi: 10.1097/MD.0000000000014197] [Medline: 30732136]
27. Smith SW, Walsh B, Grauer K, Wang K, Rapin J, Li J, et al. A deep neural network learning algorithm outperforms a conventional algorithm for emergency department electrocardiogram interpretation. J Electrocardiol 2019 Jan;52:88-95. [doi: 10.1016/j.jelectrocard.2018.11.013] [Medline: 30476648]
28. Sandhu S, Lin AL, Brajer N, Sperling J, Ratliff W, Bedoya AD, et al. Integrating a machine learning system into clinical workflows: qualitative study. J Med Internet Res 2020 Nov 19;22(11):e22421 [FREE Full text] [doi: 10.2196/22421] [Medline: 33211015]
29. Tsai WC, Liu CF, Lin HJ, Hsu C, Ma Y, Chen C, et al. Design and implementation of a comprehensive AI dashboard for real-time prediction of adverse prognosis of ED patients. Healthcare (Basel) 2022 Aug 09;10(8):1498 [FREE Full text] [doi: 10.3390/healthcare10081498] [Medline: 36011155]
30. Levin S, Toerper M, Hamrock E, Hinson JS, Barnes S, Gardner H, et al. Machine-learning-based electronic triage more accurately differentiates patients with respect to clinical outcomes compared with the emergency severity index. Ann Emerg Med 2018 May;71(5):565-74.e2. [doi: 10.1016/J.ANNEMERGMED.2017.08.005]
31. Moustapha Mbaye M, Tahiri N. A new machine learning workflow to create an optimal waiting list in hospitals. In: Proceedings of the 7th International Conference on Medical and Health Informatics. 2023 Presented at: ICMHI '23; May 12-14, 2023; Kyoto, Japan p. 159-163 URL: https://dl.acm.org/doi/10.1145/3608298.3608328 [doi: 10.1145/3608298.3608328]
32. Yu JY, Jeong GY, Jeong OS, Chang DK, Cha WC. Machine learning and initial nursing assessment-based triage system for emergency department. Healthc Inform Res 2020 Jan;26(1):13-19 [FREE Full text] [doi: 10.4258/hir.2020.26.1.13] [Medline: 32082696]

XSL•FO
RenderX

33. Farahmand S, Shabestari O, Pakrah M, Hossein-Nejad H, Arbab M, Bagheri-Hariri S. Artificial intelligence-based triage for patients with acute abdominal pain in emergency department; a diagnostic accuracy study. Adv J Emerg Med 2017;1(1):e5 [FREE Full text] [doi: 10.22114/AJEM.v1i1.11] [Medline: 31172057]
34. Jiang H, Mao H, Lu H, Lin P, Garry W, Lu H, et al. Machine learning-based models to support decision-making in emergency department triage for patients with suspected cardiovascular disease. Int J Med Inform 2021 Jan;145:104326. [doi: 10.1016/j.ijmedinf.2020.104326] [Medline: 33197878]
35. Liu N, Xie F, Siddiqui FJ, Ho AF, Chakraborty B, Nadarajan GD, et al. Leveraging large-scale electronic health records and interpretable machine learning for clinical decision making at the emergency department: protocol for system development and validation. JMIR Res Protoc 2022 Mar 25;11(3):e34201 [FREE Full text] [doi: 10.2196/34201] [Medline: 35333179]
36. Scheder-Bieschin J, Blümke B, de Buijzer E, Cotte F, Echterdiek F, Nacsa J, et al. Improving emergency department patient-physician conversation through an artificial intelligence symptom-taking tool: mixed methods pilot observational study. JMIR Form Res 2022 Feb 07;6(2):e28199 [FREE Full text] [doi: 10.2196/28199] [Medline: 35129452]
37. Georgiou A, Prgomet M, Paoloni R, Creswick N, Hordern A, Walter S, et al. The effect of computerized provider order entry systems on clinical care and work processes in emergency departments: a systematic review of the quantitative literature. Ann Emerg Med 2013 Jun;61(6):644-53.e16. [doi: 10.1016/j.annemergmed.2013.01.028] [Medline: 23548404]
38. Parker CA, Liu N, Wu SX, Shen Y, Lam SS, Ong ME. Predicting hospital admission at the emergency department triage: a novel prediction model. Am J Emerg Med 2019 Aug;37(8):1498-1504. [doi: 10.1016/j.ajem.2018.10.060] [Medline: 30413365]
39. Sun Y, Heng BH, Tay SY, Seow E. Predicting hospital admissions at emergency department triage using routine administrative data. Acad Emerg Med 2011 Aug;18(8):844-850 [FREE Full text] [doi: 10.1111/j.1553-2712.2011.01125.x] [Medline: 21843220]
40. Raita Y, Goto T, Faridi MK, Brown DF, Camargo CA, Hasegawa K. Emergency department triage prediction of clinical outcomes using machine learning models. Crit Care 2019 Feb 22;23(1):64 [FREE Full text] [doi: 10.1186/s13054-019-2351-7] [Medline: 30795786]
41. Fenn A, Davis C, Buckland DM, Kapadia N, Nichols M, Gao M, et al. Development and validation of machine learning models to predict admission from emergency department to inpatient and intensive care units. Ann Emerg Med 2021 Aug;78(2):290-302. [doi: 10.1016/j.annemergmed.2021.02.029] [Medline: 33972128]
42. Austin EE, Blakely B, Salmon P, Braithwaite J, Clay-Williams R. Technology in the emergency department: using cognitive work analysis to model and design sustainable systems. Saf Sci 2022 Mar;147:105613. [doi: 10.1016/j.ssci.2021.105613]
43. Drescher FS, Chandrika S, Weir ID, Weintraub JT, Berman L, Lee R, et al. Effectiveness and acceptability of a computerized decision support system using modified Wells criteria for evaluation of suspected pulmonary embolism. Ann Emerg Med 2011 Jun;57(6):613-621. [doi: 10.1016/j.annemergmed.2010.09.018] [Medline: 21050624]
44. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc 2004;11(2):104-112 [FREE Full text] [doi: 10.1197/jamia.M1471] [Medline: 14633936]
45. Muller MJ, Kuhn S. Participatory design. Commun ACM 1993 Jun;36(6):24-28. [doi: 10.1145/153571.255960]
46. Simonsen J, Robertson T. Routledge International Handbook of Participatory Design. New York, NY: Routledge; 2012.
47. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons for physicians not adopting clinical decision support systems: critical analysis. JMIR Med Inform 2018 Apr 18;6(2):e24 [FREE Full text] [doi: 10.2196/medinform.8912] [Medline: 29669706]
48. Kusunoki D, Sarcevic A, Zhang Z, Yala M. Sketching awareness: a participatory study to elicit designs for supporting emergency medical teamwork. Comput Support Coop Work 2015 Feb 01;24(1):1-38 [FREE Full text] [doi: 10.1007/s10606-014-9210-5] [Medline: 25870498]
49. Østervang C. Acute care planning in emergency departments: a participatory design study exploring how to improve emergency department visits through a health information system. Syddansk Universitet. 2023. URL: https://portal.findresearcher.sdu.dk/en/publications/acute-care-planning-in-emergency-departments-a-participatory-desi [accessed 2024-04-29]
50. Seo W, Park SY, Zhang Z, Singh H, Pasupathy K, Mahajan P. Identifying interventions to improve diagnostic safety in emergency departments: protocol for a participatory design study. JMIR Res Protoc 2024 Jun 21;13:e55357 [FREE Full text] [doi: 10.2196/55357] [Medline: 38904990]
51. Mattar e Silva TW, Velloso IS, Araújo MT, Fernandes AD. Configuration of power relations in physicians and nurses' professional practices. Rev Bras Enferm 2020;73 Suppl 1(suppl 1):e20180629 [FREE Full text] [doi: 10.1590/0034-7167-2018-0629] [Medline: 32428173]
52. Gleddie M, Stahlke S, Paul P. Nurses' perceptions of the dynamics and impacts of teamwork with physicians in labour and delivery. J Interprof Care 2018 Dec 30:1-11. [doi: 10.1080/13561820.2018.1562422] [Medline: 30596305]
53. Scupin R. The KJ method: a technique for analyzing data derived from Japanese ethnology. Hum Organ 2008 Jan 22;56(2):233-237. [doi: 10.17730/humo.56.2.x335923511444655]
54. Mastrianni A, Kulp L, Mapelli E, Sarcevic A. Understanding digital checklist use through team communication. Ext Abstr Hum Factors Computing Syst 2020 Apr;2020:1-8 [FREE Full text] [doi: 10.1145/3334480.3382817] [Medline: 32747878]

XSL•FO
RenderX

55. Redfern E, Brown R, Vincent CA. Improving communication in the emergency department. Emerg Med J 2009 Sep 21;26(9):658-661. [doi: 10.1136/emj.2008.065623] [Medline: 19700583]
56. Okpala P. Addressing power dynamics in interprofessional health care teams. Int J Healthc Manag 2020 May 07;14(4):1326-1332. [doi: 10.1080/20479700.2020.1758894]
57. Greenfield G, Pliskin JS, Feder-Bubis P, Wientroub S, Davidovitch N. Patient-physician relationships in second opinion encounters - the physicians' perspective. Soc Sci Med 2012 Oct;75(7):1202-1212. [doi: 10.1016/j.socscimed.2012.05.026] [Medline: 22749657]
58. Aaronson EL, White BA, Black L, Brown DF, Benzer T, Castagna A, et al. Training to improve communication quality: an efficient interdisciplinary experience for emergency department clinicians. Am J Med Qual 2019 Sep 21;34(3):260-265. [doi: 10.1177/1062860618799936] [Medline: 30235933]
59. Park SY, Chen Y, Raj S. Beyond health literacy: supporting patient-provider communication during an emergency visit. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: CSCW '17; February 25-March 1, 2017; Portland, OR p. 2179-2192 URL: https://dl.acm.org/doi/10.1145/2998181.2998357 [doi: 10.1145/2998181.2998357]
60. Feblowitz J, Takhar SS, Ward MJ, Ribeira R, Landman AB. A custom-developed emergency department provider electronic documentation system reduces operational efficiency. Ann Emerg Med 2017 Nov;70(5):674-82.e1 [FREE Full text] [doi: 10.1016/j.annemergmed.2017.05.032] [Medline: 28712608]
61. Armenia S, Thangamathesvaran L, Caine AD, King N, Kunac A, Merchant AM. The role of high-fidelity team-based simulation in acute care settings: a systematic review. Surg J (N Y) 2018 Jul 13;4(3):e136-e151 [FREE Full text] [doi: 10.1055/s-0038-1667315] [Medline: 30109273]
62. Ten Eyck RP. Simulation in emergency medicine training. Pediatr Emerg Care 2011;27(4):333-341. [doi: 10.1097/pec.0b013e3182131fe0]
63. Sweeney LA, Warren O, Gardner L, Rojek A, Lindquist DG. A simulation-based training program improves emergency department staff communication. Am J Med Qual 2014 Jul 23;29(2):115-123. [doi: 10.1177/1062860613491308] [Medline: 23880775]
64. Patterson MD, Geis GL, Falcone RA, LeMaster T, Wears RL. In situ simulation: detection of safety threats and teamwork training in a high risk emergency department. BMJ Qual Saf 2013 Jun 20;22(6):468-477. [doi: 10.1136/bmjqs-2012-000942] [Medline: 23258390]
65. Acosta JN, Falcone GJ, Rajpurkar P. The need for medical artificial intelligence that incorporates prior images. Radiology 2022 Aug;304(2):283-288. [doi: 10.1148/radiol.212830] [Medline: 35438563]

Abbreviations

AI: artificial intelligence
CDSS: clinical decision support system
ED: emergency department
EHR: electronic health record
ESI: Emergency Severity Index
PD: participatory design

Edited by A Mavragani; submitted 24.11.24; peer-reviewed by E Bai, H Pilabré; comments to author 05.01.25; revised version received 25.02.25; accepted 11.03.25; published 21.04.25.

Please cite as:
Seo W, Li J, Zhang Z, Zheng C, Singh H, Pasupathy K, Mahajan P, Park SY
Designing Health Care Provider–Centered Emergency Department Interventions: Participatory Design Study
JMIR Form Res 2025;9:e68891
URL: https://formative.jmir.org/2025/1/e68891
doi:10.2196/68891
PMID:

©Woosuk Seo, Jiaqi Li, Zhan Zhang, Chuxuan Zheng, Hardeep Singh, Kalyan Pasupathy, Prashant Mahajan, Sun Young Park. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Compatibility of the CEN-ISO/TS 82304-2 Health App Assessment Framework With Catalan and Italian Health Authorities' Needs: Qualitative Interview Study

Petra Hoogendoorn[1*], MSc; Mariam Shokralla[1], MPH, MSc; Romy Willemsen[1], MSc; Nick Guldemond[1,2], MD, PhD; María Villalobos-Quesada[1*], PhD

[1]National eHealth Living Lab, Public Health and Primary Care Department, Leiden University Medical Center, Leiden, The Netherlands

[2]Research Center for Medical Sociology, Tsinghua University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Petra Hoogendoorn, MSc
National eHealth Living Lab
Public Health and Primary Care Department
Leiden University Medical Center
Albinusdreef 2
Leiden, 2333 ZA
The Netherlands
Phone: 31 654341785
Email: a.p.y.hoogendoorn@lumc.nl

Abstract

Background: Health authorities of European Union (EU) member states are increasingly working to integrate quality health apps into their health care systems. Given the current lack of unified EU assessment criteria, the European Commission initiated Technical Specification (TS) CEN-ISO 82304-2:2021—Health and wellness apps—Quality and reliability (hereinafter the "TS") to address the scattered EU landscape of assessment frameworks (AFs) for health apps. The adoption of an AF, such as the TS, falls within member state competence and is considered an uncertainty-reduction process. Evaluations by peers as well as ensuring the compatibility of the TS with the needs of health authorities can reduce uncertainty and mediate harmonization.

Objective: This study aims to examine the compatibility of the TS with the needs of Catalan and Italian health authorities.

Methods: Semistructured interviews were conducted with key informants from a regional (Catalonia in Spain) and national (Italy) health authority, and a thematic analysis was carried out. Main themes were established deductively, following the aspects defined by the value proposition canvas: (1) health authorities' needs ("gains," "pains," and "jobs") and (2) the TS "products and services" and their distinct characteristics ("gain creators" and "pain relievers"). Subthemes were generated inductively. The compatibility of the needs with the TS was theoretically determined by the researchers. The results were visualized using the value proposition canvas. Two participant validation steps confirmed that the most relevant aspects of the predefined themes had been captured.

Results: Despite the diversity of the 2 health authorities, subthemes were common and categorized into 9 gains, 9 pains, and 11 jobs. Key findings include the health authorities' perceived value of, and need for, integrating quality health apps and using an AF (gains), along with the related policy, implementation, and operational activities (jobs). The lack of enabling EU legislation and standardization, resulting in a need for the multiple authorities involved to consent, made achieving an AF challenging (pains). Nine products and services related to the TS and 17 distinct characteristics (eg, its multistakeholder evidence base) were found to be compatible with 3 gains (eg, stimulating the prescription and use of apps), 7 pains (eg, legislation and harmonization issues), and 6 jobs (eg, assessing apps). Indirect effects, 3 anticipated future services, and 1 anticipated gain creator and pain reliever increase this compatibility.

Conclusions: Our results suggest that the health authorities share common fundamental needs, and that the TS is compatible with these needs. The identified needs and compatibility can potentially reduce peer authorities' uncertainties in adopting an AF in general and the TS in particular. More research is recommended to confirm and translate our results in other contexts and further fine-tune compatibility to achieve wide adoption of the TS and accelerate the uptake of health apps.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e67855) doi:10.2196/67855

KEYWORDS

assessment frameworks; mobile health; mHealth; health apps; wellness apps; digital transformation; Italy; Catalonia; diffusion of innovations; value proposition canvas

Introduction

Background

Health authorities of European Union (EU) member states are transforming health and care systems to address current challenges and remove cross-border regulatory barriers for businesses and consumers to progress toward an EU digital single market [1-6]. In this context, health apps (Textbox 1) are gaining attention, and countries are adapting their policies and structures to harvest the potential of these digital solutions to strengthen their health care systems [5,7-10].

Textbox 1. Definitions of health, medical, and wellness apps.

In the context of this paper, health apps are defined as apps that are "intended to be used specifically for managing, maintaining or improving health of individual persons or the delivery of care" [11]. Health apps are part of mobile health, which is defined as "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices" [12]. Health apps include medical apps and wellness apps. Medical apps have been defined as those that fall under an applicable medical device regulation and wellness apps as those that do not [13].

Not all apps are of good quality, that is, have a positive and reliable effect on health, and are easy to use, compliant with privacy and data security regulations and standards, robust, and interoperable with electronic health record systems [8,9,14]. Determining the quality of individual health apps is challenging for citizens, health care professionals (HCPs), and decision makers [15] for various reasons. First, assessing apps requires specific and diverse expertise. Second, robust evidence (eg, clinical evidence) and background information are often scarce, inappropriate, or not publicly available [14]. Third, quality assessments and widely adopted evidence-based evaluation methodologies that consider all these aspects and inform (potential) users about the quality of these apps are not yet common [10].

Several health authorities have developed their own assessment frameworks (AFs) [16]. However, their efficient implementation is challenging [7,15]. The significant overlap in quality criteria across these AFs highlights the potential for harmonization and related efficiency [17]. In addition, nonoverlapping quality criteria—a lack of agreement on what constitutes quality—potentially decrease trust [17] and increase inequality (eg, if the challenges to comply with the different AFs result in manufacturers focusing only on the larger markets and widely spoken languages). In the context of multilingualism—an EU founding principle [18]—and an abundance of health apps worldwide [19], the availability of native language health apps for the 5 million Finnish-speaking EU residents, for instance, is limited [20]. These aspects highlight the need for the harmonization of AFs at an EU level.

Both academics and policy makers endorse harmonization and cross-national regulation of health apps to realize their full potential and benefits [7,10]. The European Commission initiated CEN-ISO/TS (Comité Européen de Normalisation [European Committee for Standardization]–International Organization for Standardization/Technical Specification) 82304-2:2021 Health software—Part 2: Health and wellness apps—Quality and reliability (hereinafter the "TS"; Textbox 2) to address the scattered EU landscape of health app AFs and progress from 27 national and even more regional markets to a digital single market [6,11]. To achieve EU-wide harmonization, the adoption of the TS among EU member states is needed, which falls within member state competence and responsibility. Adoption across member states can be described as the diffusion of the TS as an innovation.

Textbox 2. Brief description of CEN-ISO/TS (Comité Européen de Normalisation [European Committee for Standardization]–International Organization for Standardization/Technical Specification) 82304-2:2021 Health software—Part 2: Health and wellness apps—Quality and reliability (the TS).

CEN-ISO/TS 82304-2:2021 was an assignment from the European Commission to CEN that became a global effort in its collaboration with ISO and the International Electrotechnical Commission (IEC). The TS has 2 core products. The first is a health app assessment framework (AF) developed through a Delphi study with 83 experts [21]. The second is a quality label (Multimedia Appendix 1) that was inspired by the European Union (EU) energy label and Nutri-Score front-of-pack nutrition label designs and the US Food and Drug Administration over-the-counter medicine label content. The quality label allows the easy and accessible presentation of the assessment results of a health app and was developed and tested with people with low health literacy. The EU-wide implementation readiness of the TS was supported by the Horizon Europe Label2Enable project (2022-2024), which tested the quality label and tested and developed products and services complementary to the AF and quality label [22,23]. These include the health app quality report, which is the detailed version of the label that aims to support health care professionals in their decision-making on recommending health apps; the certification scheme, which includes the handbook for app assessment with CEN-ISO/TS 82304-2 for certified app assessment organizations; stakeholder guidance and tools such as educational videos; and recommendations for the reimbursement of health apps. A multistakeholder road map was generated to guide next steps [24].

Innovations and their adoption process can be understood using the theory of diffusion of innovations proposed by Rogers [25], who identified 3 main attributes of an innovation that predict its future rate of adoption. Compatibility is one of these attributes and refers to the degree to which an innovation is perceived as consistent with the existing values, experiences, and needs (summarized hereinafter as "needs") of potential adopters. The diffusion of an innovation is considered an

XSL•FO
RenderX

uncertainty-reduction process. An innovation that is compatible generates less uncertainty for the potential adopter regarding the success of its deployment. The other main attributes that predict the future rate of adoption—relative advantage and complexity—are beyond the scope of this paper and will be analyzed in separate studies.

Objectives

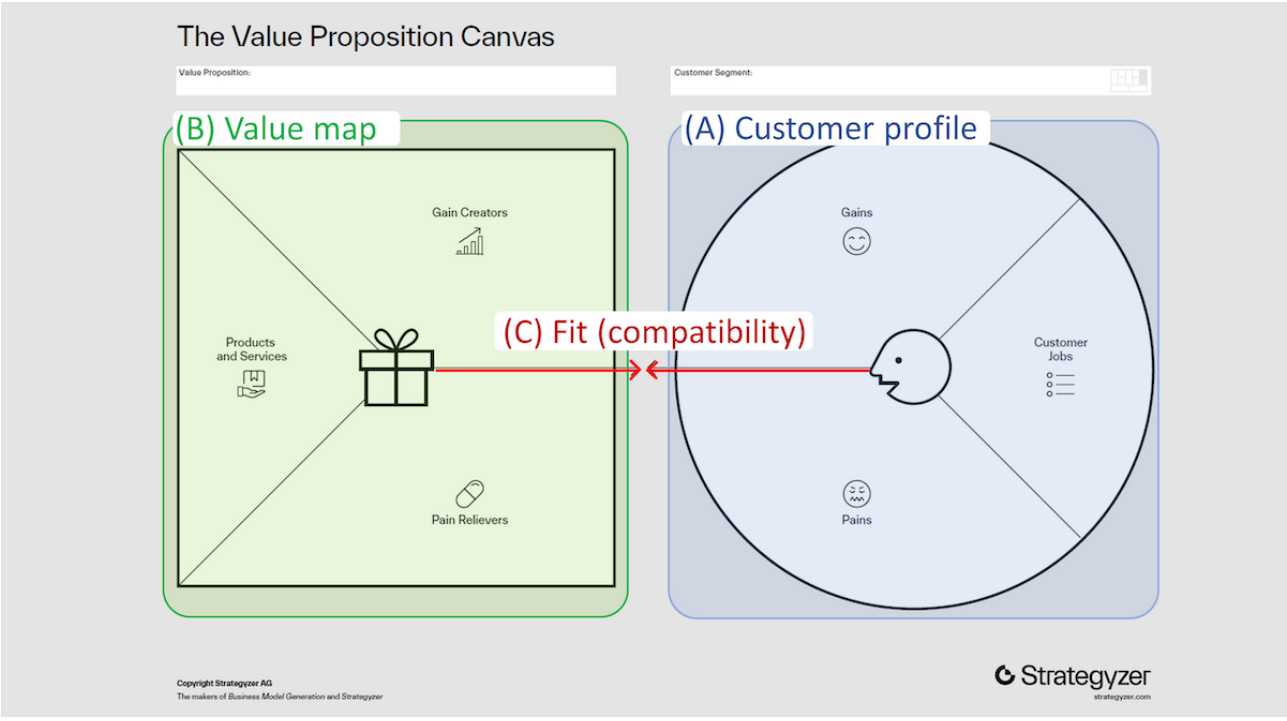
Potential adopters tend to rely on peers for information about an innovation's desirable, direct, and anticipated consequences to decrease their uncertainty about adopting it [25]. In this paper, we aim to examine the compatibility of the TS with the needs of 2 health authorities in the EU, in order to potentially enhance the compatibility of the TS and reduce the uncertainty of peer health authorities in their considerations on adopting the TS.

Methods

Overview

Rogers [25] described different methods to research the attributes of innovations, including compatibility. For this study, investigating an innovation's acceptability in its prediffusion stage (eg, test-marketing and piloting) was considered the most appropriate approach. This approach helps identify a "basis for positioning an innovation so that it will be more acceptable, that is, have a more rapid rate of adoption" [25]. Key informant interviews are a common method for assessing the acceptability of an intervention [26,27]. Key informants provide high-level perspectives and comparative insights. Their particular roles and expertise make them especially equipped for probing about how a topic is thought about or acted upon in policy. This contrasts with the scope of insight contributed by other groups of qualitative interviewees, who are more often recruited to provide data rooted in their own lived experiences, opinions, and beliefs [27]. A widely used business tool to visualize and enhance the compatibility (also referred to as "fit") of innovations is the value proposition canvas (VPC; Figure 1; Textbox 3) [28,29]. The VPC tool has been previously applied to guide the development of, among others, digital health innovations [30-32].

Figure 1. The template of the value proposition canvas [29]. Adapted from Strategyzer [33] to highlight the 3 main steps in constructing the value proposition canvas (Textbox 3): (A) the customer profile with the customer needs (jobs, gains, and pains), (B) the value map with the innovations' products and services and their distinct characteristics (gain creators and pain relievers) that aim to address the customer needs, and (C) the fit (compatibility) of the customer profile with the value map.



XSL•FO
RenderX

Textbox 3. Glossary of value proposition canvas concepts.

- The customer profile (Figure 1A) describes a potential adopter segment, in this case health authorities, in terms of their needs:
  - Jobs—the tasks or responsibilities potential adopters have, need, or want to get done in their work and their lives, in this case with respect to adopting an assessment framework and integrating health apps in their health care system
  - Gains—outcomes and benefits potential adopters require, expect, or desire to achieve if these jobs are done (well)
  - Pains—negative outcomes, risks, and obstacles that adopters encounter, foresee, or fear that prevent them from getting these jobs done (well)
- The value map (Figure 1B) makes explicit how the innovation creates value for the potential adopters' segment:
  - Products and services—the range of products and services available to the potential adopters to address their needs, in this case with respect to adopting an assessment framework and integrating health apps
  - Gain creators—the characteristics of these products and services that generate gains for the adopters
  - Pain relievers—the characteristics of these products and services that alleviate the adopters' pains
- Fit (Figure 1C): when the products and services, gain creators and pain relievers contribute to realizing the jobs and gains, and addressing the pains of the potential adopters ("compatibility")

In this paper, we used the VPC to visualize the compatibility of the TS with the needs of 2 health authorities (potential adopters of the TS):

1. A regional authority—Fundació TIC Salut Social (FTSS; TIC Salut Social Foundation), the public organization that oversees digital transformation in health care in Catalonia, Spain
2. A national authority—Italy's Istituto Superiore di Sanità (ISS; National Institute of Health)

Catalonia established its own AF in 2017 [34-36] and is pursuing the adoption of the TS [37]. Italy drafted a Technical Report about wellness, social, and health apps, which informed the TS [38], and has been using the TS as inspiration in the ongoing process to arrive at quality assessment criteria for the inclusion of telehealth solutions in a national telehealth catalog, where telehealth includes health apps. FTSS and ISS jointly represent the health authority perspective in the Label2Enable project (Textbox 2) and were recruited for this study. Recruitment criteria spanned experience as an EU-based health authority with the relatively novel integration of health apps, prediffusion interest in a standardized AF, and diversity (country or region, number of inhabitants, and an AF in place or in development). These were considered essential for providing the rich basis for analyzing the TS in terms of prediffusion acceptability or compatibility [25].

Key Informants

Three key informants were recruited. They were chosen strategically, in alignment with existing knowledge of the qualities of "good" key informants [27]. Eligibility criteria included a key leading position in their respective organizations over the course of several years in our topic of interest and related inside in-depth knowledge about the authorities' policy and plans to assess health apps as well as execution of these plans. For FTSS, the key informant has been responsible for implementing Catalonia's mobile health (mHealth) plan in FTSS for the past 9 years. For ISS, the first key informant had recently retired from a leadership position at ISS, which involved advising the government about health apps and telemonitoring. The second ISS key informant was recruited after discussion with the first ISS key informant, based on complementary first-hand experience on the topic in dealings with the Italian Ministry of Health. The 2 ISS informants played a direct role in drafting the Technical Report, which informed the TS [38]. While the number of potential participants can be very large in other types of qualitative research, key informant groups can be quite small, hard to reach, or constrained in sharing information with researchers [27].

Key informants and study researchers collaborated in the Label2Enable project; however, the key informants were not otherwise involved in these research activities. The researchers who conducted the interviews and analysis were experienced in qualitative, mixed methods, and implementation research in the field of digital health. Interest in the research topic stems from their previous work related to health apps, other digital health technologies, digital transformation, and the TS.

Constructing the VPC

Constructing the VPC involves 3 steps (Figure 1; Textbox 3) [29], which we carried out systematically using qualitative research methods. The first step was to develop the "customer profile" (Figure 1A). Here, the needs of the health authorities (potential adopters of the TS) were described in terms of the 2 authorities' "gains," "jobs," and "pains" (Textbox 3). Strategyzer's "trigger questions" served as a guide for designing the semistructured interviews (Multimedia Appendix 2) [33].

Two semistructured interviews (one for Catalonia and one for Italy) were conducted, in English, with the relevant key informants. The semistructured interview questions were shared with the key informants in advance. The interviews were conducted on the web, recorded using a secure digital platform, and lasted between 1.5 and 2 hours. Recordings were transcribed verbatim and coded independently by 2 researchers using ATLAS.ti software (version 7.9; ATLAS.ti Scientific Software Development GmbH). A deductive and inductive thematic analysis of the interviews was performed. On the basis of the VPC, we predefined 3 main themes: gains, pains, and jobs [29]. We classified a job as a past, current, or future task or

XSL•FO
RenderX

responsibility of a health authority related to adopting an AF to integrate health apps into their health care system. Jobs could be the responsibility of the health authority featured in this study or associated health authorities because, from an international perspective, health authorities' task divisions may differ and evolve. We applied the same principle to the gains and pains. Subthemes were inductively developed by 2 researchers and visualized using the VPC.

Data saturation was not deemed appropriate for this study; instead, inductive thematic saturation [39] was sought across and within cases to achieve a "basis for positioning the innovation [the TS] so that it will be more acceptable," knowing that uncertainty is inherent in the prediffusion phase, and that adopting organizations will adapt an innovation to achieve a fit of the innovation with the adopting organization and its perceived problem [25]. Accounting for data sufficiency is an underdeveloped aspect of existing methodological guidance on key informant sampling [27]. Incorporating strategies to return the findings and elicit key informant response on researcher interpretations is recommended to yield rich insights [27]. With this in mind, a participant validation step that lasted between 30 and 60 minutes was carried out separately per health authority. Key informants were supplied with the written data analysis before and after the validation step, allowing them to provide comments and enrich the analysis. The key informants confirmed that the researchers had captured the most relevant aspects within the 3 predefined main themes: gains, pains, and jobs [28].

The second step was to develop the "value map" (Figure 1B), which consists of describing the innovation's "products and services," "pain relievers," and "gain creators" (Textbox 3). Outputs from all Label2Enable work packages were considered [22,23]. Products and services included those that are currently part of the TS, those that were cocreated in the Label2Enable project (both Textbox 2), and products and services that will be generated beyond the project during the "demonstrator phase." The relevant characteristics of these products and services were subsequently added to the value map as gain creators and pain relievers.

The third step was to determine the fit (Figure 1C), that is, to check whether the "value map" was compatible with the "customer profile." In other words, whether the TS and the related products and services and their distinct characteristics (gain creators and pain relievers) satisfied the needs (jobs, gains, and pains) of the health authorities. Visualizing compatibility (fit) was an iterative process of determining and conceptualizing the products and services, gain creators, and pain relievers that related to the health authorities' gains, pains, and jobs. After completing this process, a second participant validation step, which lasted approximately 1 hour, was carried out with the key informants from both health authorities simultaneously. Again, the key informants were supplied with the written data analysis before and after the validation step, allowing them to provide comments and enrich the analysis. Furthermore, the manuscript was discussed with the key informants at different stages, and they confirmed that it condensed the most relevant aspects surrounding the position of their institutions and agreed to its publication in its current form.

Data were reported according to the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist (Multimedia Appendix 3).

Ethical Considerations

Approval from an ethics or scientific research committee was not required under Dutch national regulations because the study did not fall under the Medical Research Involving Human Subjects Act (WMO), and according to the guidelines of the Central Committee on Research Involving Human Subjects (CCMO [40]). This study followed institutional good research practices and integrity codes. Verbal consent to conduct and record the interviews was obtained from all participants (key informants) before starting the interview. Personal data were protected in accordance with European Union and institutional standards and best practices. The 2 organizations represented were funded as Label2Enable project partners and were not additionally compensated for their participation in this study.

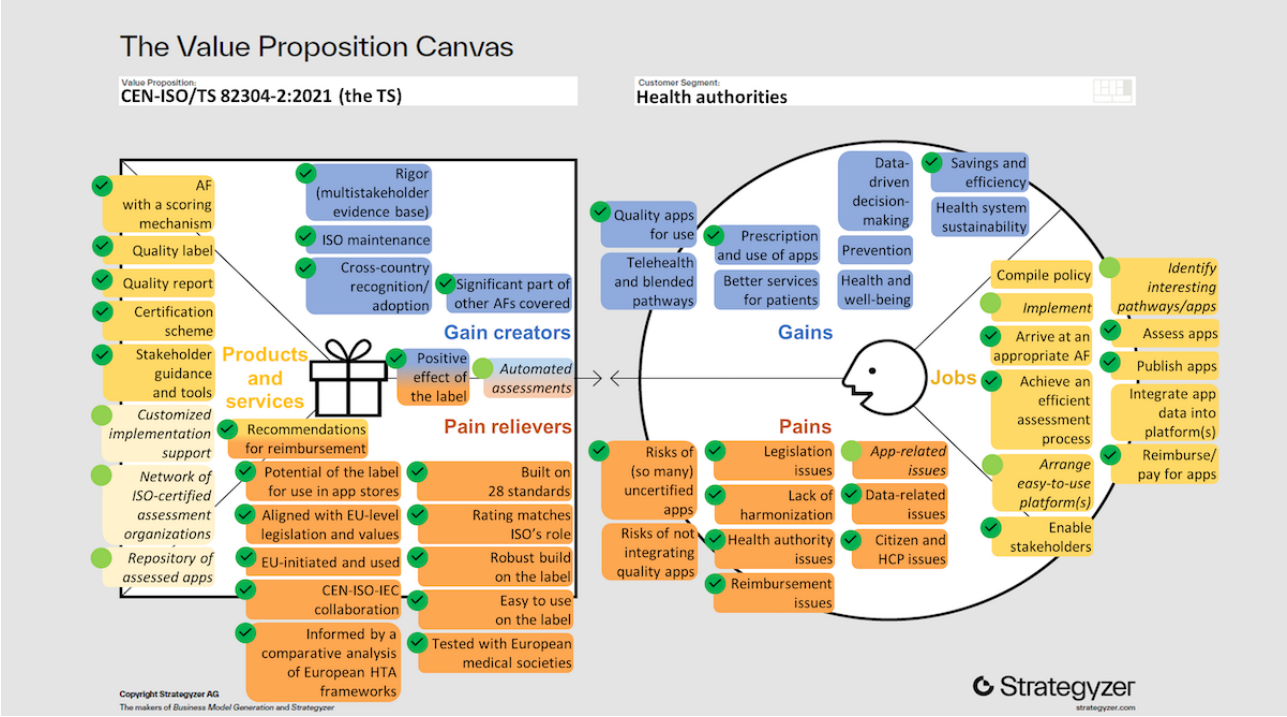
Results

Customer Profile

In this subsection, we describe the needs of health authorities (potential adopters) in terms of gains, pains, and jobs. Despite the diversity of the 2 health authorities, subthemes were common and categorized into 9 gains, 9 pains, and 11 jobs (Figure 2).

XSL•FO
RenderX

Figure 2. Adapted from Strategyzer [33], the value proposition canvas of CEN-ISO/TS (Comité Européen de Normalisation [European Committee for Standardization]–International Organization for Standardization/Technical Specification) 82304-2:2021 Health software—Part 2: Health and wellness apps—Quality and reliability (the TS) for health authorities pursuing an assessment framework (AF) to integrate health apps into their health care systems. The "customer profile" (Figure 1A) summarizing the Catalan and Italian health authorities' needs (gains, pains, and jobs) resulting from the deductive and inductive thematic analysis. The "value map" (Figure 1B) of the TS, with the products and services related to the TS and their distinct characteristics (gain creators and pain relievers) that aim to address these health authorities' needs as identified by the researchers. Lighter colored boxes with italic text indicate future services, gain creators, and pain relievers. Dark green circles with a check mark indicate the fit (compatibility; Figure 1C) of a health authority's need with an existing product or service, gain creator, or pain reliever. Light green circles without a check mark and italic text indicate expected future fit (compatibility) derived from the nature of the anticipated related services and gain creator or pain reliever. The tables in this paper show the specific products, services, or characteristics that fit specific customer needs (gain, pain, or job). AF: assessment framework; EU: European Union; HCP: health care professional; HTA: health technology assessment; IEC: International Electrotechnical Commission.



Gains

Regarding the gains—the outcomes and benefits that the health authorities require, expect, or desire to achieve from adopting an AF and integrating quality health apps—the key informants explained the interest of their region (Catalonia, Spain) or country (Italy) in an AF to identify "quality apps for use" in "telehealth (Italy) and blended pathways (Catalonia)" (Textbox 4).

Textbox 4. Key definitions in the Catalonian and Italian context.

Blended care in Catalonia

- "Blended care" "refers to an integration of online and offline components in a treatment process. This means that online and offline components are interconnected in some way and not stand-alone treatment pathways [41]. In Catalonia, according to the mobile health plan, mobile health should have an impact on the full "cycle of provision of care" (prevention, diagnostics, treatment, and follow-up) and ultimately on health, the health care system, and public health. In addition, it prioritizes the 10 most prevalent diseases (eg, diabetes) [42].

Telehealth in Italy

- The World Health Organization defines telehealth as the "delivery of health care services, where patients and providers are separated by distance" [12]. In the Italian context, telehealth or telemedicine is an innovative approach to health care practice that enables the remote provision of services through the use of digital devices, the internet, software, and telecommunication networks. Telehealth services can serve as an alternative to traditional care, support existing services (eg, by improving accessibility, efficiency, and equity), supplement care (eg, by improving effectiveness and facilitating personalized medicine), or completely replace traditional services [43].

The key informants described how the assessment of apps is instrumental to achieving the "prescription and use of apps," referring to incorporating the prescription or recommendation of apps in the provision of care and the related trusted use of health apps by citizens. Similarly, it can contribute to "better services for patients" (eg, improved accessibility, lower waiting times, and better follow-up):

> *If you want these applications to be recommended by health care professionals, we need to provide safe and guaranteed apps to the sick. So, we try to establish how we can measure or evaluate these apps and the quality of these apps.* [Key informant, Catalonia]

XSL•FO
RenderX

> *Certainly, the availability of an agreed, evidence-based assessment framework will give objectivity to the activity of app assessment.* [Key informant, Italy]

> *They [patients] will have the opportunity to make more virtual visits, sharing the data with their professionals without needing to go to the consult, provide more alerts and provide more predictions, and prevention to avoid disease or avoid being sicker. Also, they will feel more confident and committed to healthier lifestyle habits when using wellness apps. And they can also be more guarded by their professionals. They will feel more followed up.* [Key informant, Catalonia]

> *Moreover, such apps can also streamline the access of the citizens to the NHS, for example, by cutting waiting times for a given service.* [Key informant, Italy]

For the key informants, the prescription and use of apps would enable "data-driven decision-making"—the use of health data derived from an individual's health app use in the delivery of health care—seeing both benefits for the prescribing HCP and the provision of effective transversal care. Other benefits mentioned included the possibilities for "prevention," through the promotion of healthier behaviors and end-user empowerment, and ultimately "health and well-being":

> *To provide more autonomy to patients and to be more committed to their health, to keep tracking health data more efficiently and having all the information integrated into the public single network, so that all health care professionals are able to follow up the patient through the information system.* [Key informant, Catalonia]

> *The interest of our country is given by the opportunity of providing the best health status of the population, even with limited resources. Health apps can support primary and secondary prevention, by modifying lifestyles.* [Key informant, Italy]

The key informants described how all aforementioned gains translate into "savings and efficiency" and "health system sustainability," mentioning aspects such as fewer in-person visits, more web-based visits, more equity, earlier diagnosis, and cross-country recognition of app assessments:

> *I think for health care professionals, they [apps] will reduce the number of face-to-face consults for more virtual [visits]. They will have a unique platform or a unique space to monitor the data of the patients...they receive alerts, for example, so that they focus their attention on critical issues, not revising a lot of data.* [Key informant, Catalonia]

> *My personal feeling for sure, savings at the health system level. Access for all the people, or the individuals, as my Constitution says.* [Key informant, Italy]

> *Another step that has to be done is updating the current certification framework and consider the possibility to adapt this framework to the Technical Specification of 82304-2 in order to improve the framework for cross-recognition.* [Key informant, Catalonia]

> *A sound evaluation framework, recognized by the EU, will be instrumental in optimizing the resources spent in assessing health apps in the different EU countries by means of the foreseeable cross-country validity of the approval.* [Key informant, Italy]

Pains

Regarding the pains—the negative outcomes, risks, and obstacles that health authorities face with respect to assessing and integrating apps—the key informants from both authorities mentioned challenges related to context, implementation, and app use. The "risks of (so many) uncertified apps" referred to the currently unknown quality of apps and the negative effects for patients, HCPs, and the health care system. At the same time, the key informants considered the "risks of not integrating quality apps," emphasizing that not integrating apps would be an obstacle to providing better care and other important gains:

> *That is the main risk, not being adopted because professionals don't feel that it helps. Because I think that the patient part is solved. They are using it a lot and they want to be involved.* [Key informant, Catalonia]

> *We've in Italy started a long time ago working on apps because we have realized that digital support to health care is absolutely mandatory, and it was clear that there was a sort of Wild Wild West with respect to the apps.* [Key informant, Italy]

> *The main risk, I think, will be not adopting mobile apps into the health care system, not providing sufficiently optimized care to the citizens compared to other territories.* [Key informant, Catalonia]

> *At present you lose data, you lose information, you lose the capability to cost, to build a new best practice because you have no sufficient critical mass of data to say something, also to use artificial intelligence appropriately.* [Key informant, Italy]

Implementation-related pains mentioned by the key informants included "legislation issues," such as the lack of comprehensive legislation for apps that are not medical devices and the multiple applicable (emerging) EU legislations, which pose significant challenges to app developers and authorities. "Lack of harmonization" is a missed opportunity to alleviate the burden of legislation issues and potentially adds to the difficulties:

> *We are regulating medical devices, the apps that are also medical devices, thanks to the new regulation of medical devices, but we cannot regulate the apps that are not medical devices. That is why we are interested in integrating a sort of certification scheme similar to, or directly, the ISO (82304-2) Technical Specification.* [Key informant, Italy]

> *I am afraid of too much legislation. Too much vertical [legislation] that is not correlated to each other.* [Key informant, Italy]

XSL•FO
RenderX

> *We faced the challenge to create the framework from scratch because we did not have anything at the time.* [Key informant, Catalonia]

> *The risk in the immediate future is that some regional health system starts piloting an evaluation framework without national coordination, in that case...it would be more difficult to harmonize the assessment at national level, which could result in enhanced health inequalities.* [Key informant, Italy]

For the key informants, legislation issues and the lack of harmonization resulted in "health authority issues." The integration of health apps into health care currently involves multiple authorities, whose interests, roles, and responsibilities related to health apps can be unclear, overlapping or conflicting. In addition, health apps present an uncharted territory and, may represent just one of these authorities' many topics. This makes reaching agreements on issues such as an AF complex. In addition, publishing information about the quality of products without being able to refer to a recognized (international) standard could conflict with the authority's role and be perceived as biased from the perspective of health app providers. Finally, the key informants refer to "reimbursement issues," challenges related to structuring the reimbursement of health apps:

> *Something that is really difficult, and I think the European region has that problem, is the different actors to be aligned...Pff, a lot of actors, that each of them has their needs and it's very difficult to get an agreement to move forward.* [Key informant, Catalonia]

> *We have too many agencies, too many institutions that manage all these definitions of digital health and digital health technologies [not aligned taxonomies]. That is why I hope that main documents from Europe could help us to be stricter in the field.* [Key informant, Italy]

> *Communicating the quality of a particular product, could actually lead to a "market bias," that is, putting a product in a favorable position with respect to its competitors, although involuntarily, needless to say. We are instead interested in communicating the relevance of the framework for the app evaluation, as an objective and independent instrument for orienting health choices for personalized digital medicine.* [Key informant, Italy]

> *And after that comes defining the reimbursement models, well, I think it will take a lot [of time].* [Key informant, Catalonia]

The key informants identified several pains related to the identification, assessment, integration, recommendation, and use of health apps. These included "app-related issues" (eg, constant updates and a lack of clinical evidence), "data-related issues" (eg, not all data generated by apps are clinically relevant, with relevance generally differing per pathology and care pathway; moreover, there can be interoperability issues), and "citizen and HCP issues" (eg, limited digital skills):

> *Something that is very difficult is how you revalidate based on the framework, how do you reassess the apps. Because the apps change a lot.* [Key informant, Catalonia]

> *"At present what is lacking is clear sound evidence from clinical trials about the apps. This is a very great problem because also our Ministry of Health is waiting for clinical trial data for authorization of medical device apps and then to have the data from them.* [Key informant, Italy]

> *Then you [HCP] want to monitor a patient based on these health apps, and you want to integrate the data...It was very challenging because we had to find specific professionals for each pathology. They tell us or tell the group which are the relevant data.* [Key informant, Catalonia]

> *Improving the digital skills of professionals also is very challenging and we try to solve it as well, by providing more training.* [Key informant, Catalonia]

> *The target population of some interventions — for example, for frailty/chronicity — may find itself hardly capable of exploiting the technology, if the latter is not properly introduced, with a training appropriate for the digital skills of the subject.* [Key informant, Italy]

Jobs

Regarding the jobs—the tasks or responsibilities of health authorities in assessing and integrating apps—the key informants described activities related to policy, implementation, and operations. In Catalonia, "compile policy" referred mainly to the "mHealth plan" [42], which in 2015 triggered the appointment of FTSS to "implement" this plan and create an AF. In Italy, this job was linked to the recovery and resilience funding coordinated by the European Commission, which pushed plans for telehealth and an adequate AF [44]. The key informants discussed their roles as experts and advisers to the government and in the decision-making and implementation of an AF:

> *TIC Salut mHealth Office was established after publishing this agreement in 2015, an mHealth plan that was published or was agreed upon in the parliament, where they created TIC Salut in the mHealth area as an instrument to provide or to create these actions, to implement mHealth in the health sector.* [Key informant, Catalonia]

> *The National Resilience Program has given us quite a lot of support in that, and a lot of money has actually been poured into health care...One of the areas which has been impacted very much is telehealth, which includes health apps.* [Key informant, Italy]

Although the jobs overlap, the exact scope targeted in their policies and strategies differs. Catalonia emphasizes apps embedded in blended care, whereas Italy focuses on telehealth solutions:

> *The electronic health record of the patient was integrated, and we also had the personal portal for citizens. So, the gap that we did have was to integrate*

XSL•FO
RenderX

> *the data that came from apps.* [Key informant, Catalonia]
>
> *And so, the focus here in Italy at the moment is on telehealth solutions, which comprise health apps.* [Key informant, Italy]

Important steps in the implementation were, according to the key informants, to "arrive at an appropriate AF"; "achieve an efficient assessment process" that is moreover scalable; "arrange easy-to-use platform(s)"; and "enable stakeholders" by, for example, issuing recommendations, accessible information and tools (eg, self-assessment tools [45]):

> *So, those were the 2 big projects. One is the assessment framework and the other the mHealth platform that goes in parallel...The objective of this platform is to include apps that have passed this accreditation and to help health care professionals prescribe the apps to their patients.* [Key informant, Catalonia]
>
> *So, the ministry studied all the proposals from the different experts, and they actually decided to use 82304-2 as the inspirational assessment framework to define the assessment framework for telehealth solutions that are going to be onboarded onto this national catalog.* [Key informant, Italy]
>
> *The main objective is now the number of apps that we want to share. It's our expectation to have at least 30 apps, or 50 depending. But this year at least try to get these 30.* [Key informant, Catalonia]
>
> *And also, we did recommendations, guides, tools. Also, we have the auto-evaluation on our website, that is a questionnaire free and open...We made it for the industry, for them being prepared, and to promote creating these quality apps.* [Key informant, Catalonia]

Operations start with what key informants refer to as a prioritization of the apps to integrate into the health care system ("identify interesting pathways/apps"):

> *So, we focus on diabetes because it's something that we have different apps that are provided by private companies, and there are a lot of people with this disease, and it is expensive and it's a priority for the health care system to provide tools in that way.* [Key informant, Catalonia]
>
> *So, there will be a pathway for diabetics, so there would be one for oncology, certain types of oncology and so on.* [Key informant, Italy]
>
> *We have these 2000 apps that are identified as interesting apps, that are not certified using our framework obviously, but they are being used by health care centers or they are being published in other frameworks like Andalusia or mHealth Belgium, et cetera.* [Key informant, Catalonia]

After this prioritization of apps that are interesting to assess, the assessment can be carried out ("assess apps"), and positively assessed apps and their results in specific quality domains and criteria can be shared ("publish apps"). The publication of assessed apps enables the series of gains, which starts with "quality apps for use." The next operational step, to "integrate app data into platform(s)," enables the desired gain "data-driven decision-making." Catalonia expects the capability to integrate app data into platforms to become available in 2025 for 1 care pathway, while Italy does not yet carry out assessments, and the division of responsibilities among Italian authorities is not yet defined.

For both health authorities featured in this study, the final step in the process is to "reimburse/pay for apps":

> *Once these apps are validated, we publish the apps on our website, so all citizens know all aspects of the [certification] process.* [Key informant, Catalonia]
>
> *And in the next stage, we are expecting to integrate the data from these apps of diabetes in the platform that we create.* [Key informant, Catalonia]
>
> *We have a group that is working on a specific reimbursement model.* [Key informant, Catalonia]
>
> *A policy is at present under discussion thanks to a specific government workgroup on digital therapies that involves several stakeholders, scientific and industrial. The latter (drug and medical device industry organizations) are discussing a unitary proposal for a unique reimbursement scheme.* [Key informant, Italy]

Value Map and Fit

Overview

In this subsection, we present the results of the iterative process of identifying the TS-related products and services and their distinct characteristics in relation to the needs of the health authorities. The needs are visualized in the VPC as the pains, gains, and jobs of the 2 health authorities. When a TS-related product or service or one of its characteristics (gain creator and pain reliever) helps to address or to achieve a gain, pain, or job of the health authorities, we considered it a "fit" (compatibility). The fit was visualized with a green circle (Figure 2). We found that the TS addressed 3 (33%) of 9 gains, 7 (78%) of 9 pains, and 6 (55%) of 11 jobs in part or in full. An anticipated gain creator and pain reliever and 3 TS-related anticipated future services would target 1 extra pain and 3 extra jobs, enhancing the compatibility to 3 (33%) of 9 gains, 8 (89%) of 9 pains, and 9 (82%) of 11 jobs. The 6 gains and 1 pain that remain would be influenced indirectly (eg, the TS supports the gain "quality apps for use," which is a prerequisite for all further gains). This would leave only 2 jobs unaddressed: "compile policy" and "integrate app data into platform(s)."

Products and Services

In total, 6 current and 3 future products and services were identified, which are compatible with ("fit") 9 jobs (Figure 2; Table 1). The 2 core products of the TS are an "AF with a scoring mechanism" and a health app "quality label" (Multimedia Appendix 1) [11]. The quality label is a key aspect that differentiates the TS from other AFs. In the Label2Enable project, four TS-related products were cocreated with the relevant stakeholders: (1) the health app "quality report," which

XSL•FO
RenderX

is a more detailed version of the quality label [22]; (2) the ISO 17000 series–compliant "certification scheme" for the TS, which includes the app assessment handbook for CEN-ISO/TS 82304-2 for certified conformity assessment bodies [22]; (3) "stakeholder guidance and tools," which refers, for example, to educational videos [46] on the quality label for citizens and guidance for manufacturers; and (4) a series of "recommendations for reimbursement" of health apps [47]. The upcoming demonstrator phase aims to assess the first 100 apps, test and optimize the value of the health app quality report from a multistakeholder perspective and explore how to exploit the potential of "automated assessments" (refer to the next subsection, Gain Creators). This phase is planned to deliver 3 extra services: (1) "customized implementation support," (2) a "network of ISO-certified assessment organizations," and (3) a "repository of assessed apps."

Table 1. Compatibility (or fit) of the jobs related to adopting an assessment framework (AF) for health apps as described by the Catalan and Italian health authorities and the products and services related to CEN-ISO/TS (Comité Européen de Normalisation [European Committee for Standardization]–International Organization for Standardization/Technical Specification) 82304-2:2021 Health software—Part 2: Health and wellness apps—Quality and reliability (the TS).

Jobs	Products and services
Implement[a]	• Customized implementation support[a]
Arrive at an appropriate AF	• AF with a scoring mechanism
Achieve an efficient assessment process	• Certification scheme • Network of ISO-certified assessment organizations[a]
Arrange easy-to-use platform(s)[a]	• Repository of assessed apps[a]
Enable stakeholders	• Stakeholder guidance and tools • Network of ISO-certified assessment organizations[a]
Identify interesting pathways/apps[a]	• Repository of assessed apps[a]
Assess apps	• AF with a scoring mechanism • Certification scheme • Stakeholder guidance and tools • Network of ISO-certified assessment organizations[a]
Publish apps	• Quality label • Quality report • Repository of assessed apps[a]
Reimburse/pay for apps	• Quality report • Recommendations for reimbursement

[a]Job to be facilitated and corresponding anticipated TS service planned to be established during the demonstrator phase.

Gain Creators

Five current and one anticipated gain creators were identified (Figure 2), which are compatible with ("fit") 3 gains pursued by the health authorities (Table 2): (1) the "rigor (multistakeholder evidence base)" of the AF, the assessment process (promising interrater reliability), and the handbook for assessment organizations, which are in part already reported in scientific publications [21,22,48]; (2) "ISO maintenance," which refers to the regular ISO review procedures that ensure that the TS remains up to date and potentially evolves to an International Standard; (3) "cross-country recognition/adoption" can be facilitated by the TS, which in turn can increase the number of assessed apps available, optimize resource allocation (the label is effectively a screening), and minimize the duplication of efforts; (4) the "positive effect of the label" on manufacturers' intent to improve the quality of their app, HCPs' willingness to prescribe apps, and citizens' intent to download and ability to choose quality apps [22,48,49]. These effects could be amplified by citizens' trust in the recommendations of health apps by HCPs (80%) and their expressed need for health authorities to review and rate apps (86%; M Shokralla, MPH, MSc, unpublished data, 2024); (5) the preliminary findings of a comparative analysis of CEN-ISO/TS 82304-2 with European health technology assessment (HTA) frameworks and subsequent alignment [22] show that the handbook has a "significant part of other AFs covered." For health authorities, this could mean a reduction in the conformity assessment workload because already-labeled apps would only need to be assessed against a limited set of context-specific requirements; and (6) "automated assessments," which refers to the potential of partially automatizing app assessments, benefiting the quality (rigor), efficiency, affordability, and scalability of assessments and reassessments. Examples include software that can test the app's accessibility (eg, contrast, addressing color blindness, and readability), privacy compliance (eg, personal data processed

XSL•FO
RenderX

and whether the privacy statement includes all mandatory elements [50]), and data security mechanisms (eg, automated security testing) and in general natural language processing to help assess the evidence provided and queries for surveillance purposes.

Table 2. Compatibility (or fit) of the gains of adopting an assessment framework (AF) for health apps as perceived by the Catalan and Italian health authorities and the distinct characteristics (gain creators) of the products and services related to CEN-ISO/TS (Comité Européen de Normalisation [European Committee for Standardization]–International Organization for Standardization/Technical Specification) 82304-2:2021 Health software—Part 2: Health and wellness apps—Quality and reliability (the TS).

Gains	Gain creators
Quality apps for use	• Rigor (multistakeholder evidence base) • ISO maintenance • Cross-country recognition/adoption
Prescription and use of apps	• Positive effect of the label
Savings and efficiency	• Cross-country recognition/adoption • Significant part of other AFs covered • Automated assessments[a]

[a]Future gain creator planned to be established during the demonstrator phase.

Pain Relievers

Twelve current pain relievers and one anticipated pain reliever were identified (Figure 2), which are compatible with ("fit") 8 pains (Table 3): (1) the "potential of the label for use in app stores," the common marketplace for health apps. The TS has global applicability (ISO), a label, a scoring mechanism, and scalability potential with the "automated assessments" and "network of ISO-certified assessment organizations." These distinct characteristics are considered attractive for app stores to rank apps and inform potential users about the quality of (labeled) apps; (2) the "positive effect of the label" on manufacturers' quality improvement plans and HCPs' willingness to prescribe apps [22,48,49]; (3) the handbook for app assessment organizations was "aligned with EU-level legislation and values" [22], and the AF was built on the EU Medical Device Regulation and General Data Protection Regulation principles [21]; (4) the TS and Label2Enable project are "EU-initiated and used." A draft version of the TS was made available to support the creation of COVID-19 apps and referenced in the EU toolbox for COVID-19 contact–tracing apps [51], and the TS is foundational for drafting implementing legislation for labeling wellness applications (that claim interoperability with an electronic health record system) in the European Health Data Space Regulation; (5) the TS is a "CEN-ISO-IEC collaboration", all 3 bodies are renowned international standardization organizations; (6) the handbook was "informed by a comparative analysis of European HTA frameworks," that is, the European Network for Health Technology Assessment core model and the Dutch, English, Finnish, French, and German HTA frameworks for health apps [22]; (7) distinctive from other AFs, the AF was "built on 28 standards" [11,21]. This includes the recognized National Institute for Health and Care Excellence evidence standards framework for digital health technologies [52,53], which was used as a foundation for the CEN ISO/TS 82304-2 AF [21]; (8) the "rating matches ISO's role" to "agree on the best way of doing things, make lives easier, safer and better, enabling trade the world over" [54], enabling authorities to refer to unbiased and standardized assessment results; (9) the "recommendations for reimbursement," which are the result of a series of Label2Enable workshops for health authorities, HTA bodies, and insurers, with 135 participants from 34 countries [47]; (10) "automated assessments" could help address the need to frequently reassess apps; (11) the visibility of "robust build on the label," which includes 4 interoperability requirements, is expected to promote manufacturer investments in interoperability; (12) the incorporation of "easy to use on the label" could contribute to addressing inequity in the prescription and use of health apps and promote the continued use of apps [55]; and (13) given the importance of medical societies for HCPs in promoting the prescription and use of apps [56,57], the quality report was "tested with European medical societies" to evaluate and enhance its usefulness for medical societies in providing guidance for HCPs in recommending quality health apps [22]; in addition, the results of mapping the TS with the information needs of cardiologists regarding mobile health solutions, carried out with the European Society of Cardiology, indicated compatibility [58].

XSL•FO
RenderX

Table 3. Compatibility (or fit) of the pains in adopting an assessment framework (AF) for health apps as perceived by the Catalan and Italian health authorities and the distinct characteristics (pain relievers) of the products and services related to CEN-ISO/TS (Comité Européen de Normalisation [European Committee for Standardization]–International Organization for Standardization/Technical Specification) 82304-2:2021 Health software—Part 2: Health and wellness apps—Quality and reliability (the TS).

Pains	Pain relievers
Risks of (so many) uncertified apps	• Potential of the label for use in app stores • Positive effect of the label
Legislation issues	• Aligned with EU[a]-level legislation and values • EU-initiated and used
Lack of harmonization	• CEN-ISO-IEC[b] collaboration • Informed by a comparative analysis of European HTA[c] frameworks • Built on 28 standards • EU-initiated and used
Health authority issues	• Rating matches ISO's role
Reimbursement issues	• Recommendations for reimbursement
App-related issues[d]	• Automated assessments[d]
Data-related issues	• Robust build on the label
Citizen and HCP issues	• Positive effect of the label • Easy to use on the label • Tested with European medical societies

[a]EU: European Union.

[b]IEC: International Electrotechnical Commission.

[c]HTA: health technology assessment.

[d]Pain to be addressed and corresponding anticipated pain reliever planned to be established during the demonstrator phase.

Discussion

Principal Findings

In this paper, we systematically examined the compatibility of the TS with the needs of 2 health authorities, aiming to further enhance the compatibility of the TS with these needs and, ultimately, reduce the uncertainty of peer authorities in considering the adoption of the TS. The 2 studied health authorities were diverse. FTSS is located in Catalonia, a region in Spain with 8 million citizens, and has an AF for health apps and wearables and an operational assessment process in place. ISS is based in Italy, a country with 59 million citizens, and is working on an AF for telehealth, which includes health apps. We found that despite their diversity, their needs (gains, pains, and jobs) largely overlapped. This suggests that health authorities share common fundamental needs. Differences in needs could be attributed to being a national advisory body in a country currently without an established AF (a focus on legislation issues—Italy) and a regional implementing organization with an AF (a focus on the execution of policy—Catalonia). Both health authorities see a need for, and the benefits of, the uptake of health apps and using a common AF. This confirms the recommendation of the World Health Organization (WHO) to make mHealth evaluation with a common methodology the norm rather than the exception [10]. At the same time, it is apparent that without enabling (EU) legislation and standardization and with multiple authorities involved, it is a challenge to establish an AF. For countries or regions with a small population, arriving at an appropriate (and sufficiently rigorous) AF might not even be feasible, given the costs, capabilities, and limited attractiveness for manufacturers.

When is the compatibility of the TS *sufficient* for health authorities? Osterwalder et al [29] argue that an innovation does not have to address all needs of potential adopters; yet it should satisfy the most essential (unrealized) gains, most extreme (unresolved) pains, and most important (unsatisfied) jobs and preferably be "difficult to copy." We found that the TS has a range of products and services with distinct characteristics (pain relievers and gain creators) that address in part or in full all but 2 needs of the 2 health authorities studied; these are the jobs "compile policy" and "integrate app data into platform(s)" (Tables 1-3). To what extent these 2 jobs would classify as *important (unsatisfied) jobs*—and as such should be addressed in the future by the TS—needs to be determined in future research. Many of the distinct characteristics (gain creators and pain relievers) are *difficult to copy* (eg, multistakeholder evidence base, international standardization nature, the comprehensiveness of the label, and its development as an EU initiative that has already been applied at the EU level).

Whether all of the distinguished products, services, and their distinct characteristics individually (are perceived to) *sufficiently satisfy* potential adopters' needs could not be measured

XSL•FO
RenderX

conclusively, given the prediffusion stage of the TS at the time of this study. Such measurements and potential further fine-tuning of the compatibility of the TS, which could further reduce uncertainty, are part of the upcoming plans for the demonstrator phase. The aims for this phase include assessing the first 100 apps, developing the aforementioned additional 3 services and 1 gain creator and pain reliever, and testing the health app quality report. We also aim to explore whether a further reduction in the duplication of efforts would increase compatibility and reduce uncertainty. Could, for instance, the need to "arrange easy-to-use platform(s)" for assessed apps be addressed with an EU-level "repository of assessed apps" and EU efforts to materialize the "potential of the label for use in app stores"? Could such strategies increase the rate of adoption of the TS, prevent EU member states (and potentially medical societies) from having to invest individually in such platforms and thus pave the way for the EU-wide adoption of the TS and uptake of health apps in the region? WHO statistics seem to confirm the potential of an EU-level platform, with 77% (39/51) of its European region member states rating the lack of a trustworthy source to access effective apps as a significant barrier to integrating health apps into clinical practice, outnumbering privacy and security concerns (38/51, 75%), patient digital literacy (37/51, 73%), and the lack of evidence on app effectiveness in clinical practice (31/51, 61%) [10].

The health authorities' jobs can be put in context with the 5 stages of the innovation-decision process in organizations described by Rogers [25]. The first stage is "agenda setting," when one or more individuals in an organization identify a problem and seek a compatible innovation to solve it. During this stage, the job "compile policy" needs to be carried out. WHO statistics show that 83% (44/53) of its European region member states, including 85% (23/27) of the EU member states, reported having a national digital health policy or strategy. Nevertheless, only 28% (13/47) reported having an entity for mHealth quality oversight, which relates to the second job, "implement" [59].

After agenda setting, the next stages of the innovation-decision process in organizations described by Rogers [25] are "matching" (testing the feasibility of the innovation in solving the organization's problem), and "redefining/restructuring" (reinventing the innovation to accommodate the organization's needs and structure). These are related to the identified jobs "arrive at an appropriate AF" and "achieve an efficient assessment process." Catalonia has already carried out these jobs and published the methodology used to compare the TS to its previous AF and implement the TS, reporting a mere 7% additional context-specific, scope-expanding, and rigor-enhancing requirements [37], similar to reports in the Australian context [60] and as previously suggested [17]. This knowledge could further reduce peer authorities' uncertainty and inspire their own "matching" and "redefining/restructuring" efforts. The rest of the reported jobs, including "assess apps," would follow once the first four have been achieved. Going back to the status of the European region reveals that while 83% (44/53) of WHO European region member states reported having a national digital health policy or strategy, only 15% (6/39) reported the evaluation of government-sponsored mHealth ("assess apps") [10], which is a prerequisite for the ultimate jobs in our list: "integrate app data into platform(s)" and "reimburse/pay for apps." Although financing schemes for telemedicine are increasingly regularly available [10], the number of reimbursed apps were found to be limited [7] and related value-based pricing frameworks "nonexistent to embryonic" [61].

The low percentages of WHO European region member states having entities for mHealth quality oversight and government-sponsored mHealth evaluation raise questions. Do other member states encounter the same significant challenges (pains) as Catalonia and Italy in satisfactorily arriving at an appropriate AF and assessing health apps? Could that be the reason for not assessing apps? Or could it be that these member states have different needs, perhaps not even perceive, contrary to Catalonia and Italy (and the WHO), that an AF is a much-needed solution? Or do they perhaps perceive other *most essential (unrealized) gains*, *most extreme (unresolved) pains*, and *most important (unsatisfied) jobs*, making an AF less of a priority or other distinct AF characteristics a necessity? Further key informant interviews or analysis of member states' national policies [59] could validate the generalizability of the Catalan and Italian needs or produce alternative customer profiles, perhaps related to the adopter types described by Rogers [25]: innovators, early adopters, early majority, late majority, and laggards. Such an analysis would support future alignment of the TS with the EU region at large; contribute to the development of new services, such as "customized implementation support"; and inform peer authorities' decision-making on the adoption of the TS. In other words, an analysis of member states' policies through the lens of the current VPC could potentially support the EU region in scaling the evaluation of mHealth using a common methodology as well as promote the equitable uptake of quality apps. Further "matching" efforts by, or in collaboration with, health authorities, other stakeholders, or relevant projects (eg, European Digital Health Technology Assessment [62] and ASSESS DHT [63]), using their own methodologies or inspiration from Catalonia [37], would reveal whether they similarly have few additional health app quality requirements when integrating CEN-ISO/TS 82304-2 and further evolving the Label2Enable handbook app assessment. These efforts and more multistakeholder positioning statements would further reduce uncertainties.

Strengths and Limitations

To our knowledge, this study is the first to systematically analyze the compatibility of an existing AF with health authorities' needs, which has the potential to address the scattered EU landscape, progressing toward a digital single market and improving the uptake of health apps. The timing of the study (prediffusion) exploited the potential to contribute to preadoption compatibility, the effectiveness of the future demonstrator phase, and the rate of adoption of the TS. Our study was based on the perceptions of 3 key informants associated with 2 health authorities in 2 EU countries, each part of a larger landscape of health authorities, each with a focus on health apps as part of a larger scope of digital health technologies (Catalonia: health apps and wearables and Italy:

XSL•FO
RenderX

telehealth). The limited number of key informants can be attributed to the prediffusion phase of the TS, that is, the TS was published in 2021, and from 2022 to 2024, the Label2Enable project has cocreated the TS supportive products and services with stakeholders. The 2 health authorities involved—partners in the Label2Enable project—are the first prerelease test users in the EU. The key informants are among the first individuals to evaluate the TS. Such a limited scope requires future validation, and we specifically pledge to extend this analysis to the national policies of the EU member states and further "matching" efforts. The complex process of multistakeholder adoption in the EU and the wider landscape depends on more than compatibility for 2 health authorities in the prediffusion phase. As Catalonia's FTSS revealed in its paper, a multistakeholder evaluation is recommended as part of adoption decision-making [37]. In other publications, currently or shortly available (all Label2Enable results will be published in Cordis [22] and on the Label2Enable results page [23]), we have addressed other attributes and other key stakeholders, some of whom have produced endorsing positioning statements [58,64]. Theories focusing on the wider ecosystem and AF landscape could be considered to further support adoption strategies. Despite the limitations of this study, the diversity of the 2 health authorities in terms of geographic scope, roles and responsibilities, population size, triggers, the scope of digital health technologies, target pathways and health apps, and the status of the AF, increases the chance of the generalizability of our results.

Conclusions

Our results suggest compatibility of the TS with the overlapping needs of the health authorities to arrive at an appropriate AF that would allow the uptake of health apps in their health care systems. The perceptions and experiences of health authorities captured in this study through interviews with key informants provide an evidence-based foundation for peer authorities to reduce their uncertainties related to the adoption of an AF, particularly the TS. The study also established the basis to carry out a wider analysis to understand the compatibility of the TS with the needs of other EU member states, which would confirm or fine-tune the TS and its distinct characteristics. To our knowledge, this is the first report to systematically analyze the compatibility of an existing AF with health authorities' needs, with the potential to address the scattered EU AF landscape, progressing toward a digital single market and improving the uptake of health apps.

Acknowledgments

The authors would like to express their deepest gratitude to Carme Pratdepadua Bufill, Mauro Grigioni, Pier Angelo Sottile, and Giuseppe d'Avenio, who generously shared their time, insights, and expertise. Their invaluable contributions and perspectives have been instrumental in shaping the findings of this study. The views and opinions expressed by the individuals who participated in this study are their own and do not necessarily reflect the official policy or position of the institutions they are affiliated with.

This study was funded by the European Union (101057522—Label2Enable—HORIZON-HLTH-2021-IND-07). The views and opinions expressed in this paper are those of the authors and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency; neither the European Union nor the granting authority can be held responsible for them.

Authors' Contributions

PH and MS were responsible for funding acquisition. PH was responsible for supervision. MS, RW, and PH designed the semistructured interviews. MVQ led the interviews with support from MS. MVQ transcribed the recordings verbatim, and PH proofread the transcripts. PH and MVQ performed coding using both deductive and inductive approaches. PH visualized the results using the value proposition canvas with feedback from MVQ. PH and MVQ wrote the original draft. MS, RW, and NG reviewed and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The CEN-ISO/TS (Comité Européen de Normalisation [European Committee for Standardization]–International Organization for Standardization/Technical Specification) 82304-2 label.
[DOCX File , 139 KB - formative_v9i1e67855_app1.docx]

Multimedia Appendix 2
Interview topic guide.
[DOCX File , 20 KB - formative_v9i1e67855_app2.docx]

Multimedia Appendix 3
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[DOCX File , 24 KB - formative_v9i1e67855_app3.docx]

XSL•FO
RenderX

References

1. Kivimäki M, Batty GD, Pentti J, Shipley MJ, Sipilä PN, Nyberg ST, et al. Association between socioeconomic status and the development of mental and physical health conditions in adulthood: a multi-cohort study. Lancet Public Health 2020 Mar;5(3):e140-e149 [FREE Full text] [doi: 10.1016/S2468-2667(19)30248-8] [Medline: 32007134]
2. Health for everyone?: social inequalities in health and health systems. Organisation for Economic Co-operation and Development. URL: https://www.oecd-ilibrary.org/social-issues-migration-health/health-for-everyone_3c8385d0-en [accessed 2025-01-12]
3. Silver opportunity: building integrated services for older adults around primary health care. The World Bank. URL: https://documents1.worldbank.org/curated/en/099025503072337503/pdf/P17583208782960a40bf6804d1758e50f1e.pdf [accessed 2025-01-12]
4. Health and care workforce in Europe: time to act? World Health Organization. URL: https://www.who.int/europe/publications/i/item/9789289058339 [accessed 2025-01-12]
5. Green Paper on mobile health ("mHealth"). European Commission. 2014. URL: https://digital-strategy.ec.europa.eu/en/library/green-paper-mobile-health-mhealth [accessed 2025-01-12]
6. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on Enabling the Digital Transformation of Health and Care in the Digital Single Market; empowering citizens and building a healthier society. European Commission. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A233%3AFIN [accessed 2025-01-12]
7. Essén A, Stern AD, Haase CB, Car J, Greaves F, Paparova D, et al. Health app policy: international comparison of nine countries' approaches. NPJ Digit Med 2022 Mar 18;5(1):31 [FREE Full text] [doi: 10.1038/s41746-022-00573-1] [Medline: 35304561]
8. Milne-Ives M, van Velthoven MH, Meinert E. Mobile apps for real-world evidence in health care. J Am Med Inform Assoc 2020 Jun 01;27(6):976-980 [FREE Full text] [doi: 10.1093/jamia/ocaa036] [Medline: 32374376]
9. Stecher C, Pfisterer B, Harden SM, Epstein D, Hirschmann JM, Wunsch K, et al. Assessing the pragmatic nature of mobile health interventions promoting physical activity: systematic review and meta-analysis. JMIR Mhealth Uhealth 2023 May 04;11:e43162 [FREE Full text] [doi: 10.2196/43162] [Medline: 37140972]
10. The ongoing journey to commitment and transformation. World Health Organization. URL: https://www.who.int/europe/publications/digital-health-in-the-who-european-region-the-ongoing-journey-to-commitment-and-transformation [accessed 2025-01-12]
11. ISO/TS 82304-2:2021: health software: part 2: health and wellness apps — quality and reliability. International Organization for Standardization. URL: https://www.iso.org/standard/78182.html [accessed 2025-01-12]
12. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization. 2011. URL: https://iris.who.int/handle/10665/44607 [accessed 2025-01-12]
13. Maaß L, Freye M, Pan CC, Dassow HH, Niess J, Jahnel T. The definitions of health apps and medical apps from the perspective of public health and law: qualitative analysis of an interdisciplinary literature overview. JMIR Mhealth Uhealth 2022 Oct 31;10(10):e37980 [FREE Full text] [doi: 10.2196/37980] [Medline: 36315221]
14. Grundy Q. A review of the quality and impact of mobile health apps. Annu Rev Public Health 2022 Apr 05;43:117-134 [FREE Full text] [doi: 10.1146/annurev-publhealth-052020-103738] [Medline: 34910582]
15. van Velthoven M, Powell J. Do health apps need endorsement? Challenges for giving advice about which health apps are safe and effective to use. Digit Health 2017 Apr 10;3:2055207617701342 [FREE Full text] [doi: 10.1177/2055207617701342] [Medline: 29942588]
16. D2.1 knowledge tool 1. health apps assessment frameworks. The European mHealth Hub. URL: https://cordis.europa.eu/project/id/737427/results [accessed 2025-01-12]
17. Frey AL, Phillips B, Baines R, McCabe A, Elmes E, Yeardsley-Pierce E, et al. Domain coverage and criteria overlap across digital health technology quality assessments: a systematic review. Health Technol 2024 Dec 20;15(1):15-27 [FREE Full text] [doi: 10.1007/S12553-024-00934-8]
18. Charter of fundamental rights of the European Union 2012/C 326/02. European Union. 2012. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3AC2012%2F326%2F02 [accessed 2025-01-12]
19. Digital health trends 2021: innovation, evidence, regulation, and adoption. IQVIA Institute. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/digital-health-trends-2021 [accessed 2025-01-12]
20. Voutilainen M. Report on Finnish-language mobile applications related to welfare and health in Google Play -store and in the EUDAMED -database. Finnish Coordinating Center for Health Technology Assessment. URL: https://oys.fi/fincchta/wp-content/uploads/sites/21/2024/11/fincchta_report-on-finnish-language-mobile-applications.pdf [accessed 2025-01-12]
21. Hoogendoorn P, Versluis A, van Kampen S, McCay C, Leahy M, Bijlsma M, et al. What makes a quality health app-developing a global research-based health app quality assessment framework for CEN-ISO/TS 82304-2: Delphi study. JMIR Form Res 2023 Jan 23;7:e43905 [FREE Full text] [doi: 10.2196/43905] [Medline: 36538379]
22. Adopting ISO 82304-2 and a trusted EU mHealth label for a single market that enables patients, citizens, health professionals, systems and authorities to benefit from a healthy supply of useful apps. European Commission. URL: https://cordis.europa.eu/project/id/101057522/results [accessed 2025-01-12]

XSL•FO
RenderX

23. Results. Label2Enable. URL: https://label2enable.eu/results [accessed 2025-01-12]
24. The preferred future of health & care in 5-10 years and the role of digital health quality labelling. Label2Enable. URL: https://label2enable.eu/assets/downloads/label2enable_preferred-future-health-and-care-visual-1716569529.pdf [accessed 2025-01-12]
25. Rogers EM. Diffusion of Innovations. 4th edition. New York, NY: Free Press; 2018.
26. Ayala GX, Elder JP. Qualitative methods to ensure acceptability of behavioral and social interventions to the target population. J Public Health Dent 2011;71 Suppl 1(0 1):S69-S79 [FREE Full text] [doi: 10.1111/j.1752-7325.2011.00241.x] [Medline: 21656958]
27. Pahwa M, Cavanagh A, Vanstone M. Key informants in applied qualitative health research. Qual Health Res 2023 Dec 30;33(14):1251-1261 [FREE Full text] [doi: 10.1177/10497323231198796] [Medline: 37902082]
28. Kyhnau J, Nielsen C. Value proposition design: how to create products and services customers want. J Bus Mod 2015;3(1):52. [doi: 10.5278/OJS.JBM.V3I1.1105]
29. Osterwalder A, Pigneur Y, Bernarda G, Smith A, Papadakos P. Value Proposition Design: How to Create Products and Services Customers Want. Hoboken, NJ: John Wiley & Sons; 2014.
30. Lentferink A, Polstra L, D'Souza A, Oldenhuis H, Velthuijsen H, van Gemert-Pijnen L. Creating value with eHealth: identification of the value proposition with key stakeholders for the resilience navigator app. BMC Med Inform Decis Mak 2020 Apr 27;20(1):76 [FREE Full text] [doi: 10.1186/s12911-020-1088-1] [Medline: 32340615]
31. Chakraborty I, Edirippulige S, Vigneswara Ilavarasan P. The role of telehealth startups in healthcare service delivery: a systematic review. Int J Med Inform 2023 Jun;174:105048 [FREE Full text] [doi: 10.1016/j.ijmedinf.2023.105048] [Medline: 36963322]
32. Wrede C, Braakman-Jansen A, van Gemert-Pijnen L. How to create value with unobtrusive monitoring technology in home-based dementia care: a multimethod study among key stakeholders. BMC Geriatr 2022 Nov 30;22(1):921 [FREE Full text] [doi: 10.1186/s12877-022-03550-1] [Medline: 36451119]
33. The value proposition canvas, customer gains, customer pains and customer jobs: trigger questions. Strategyzer. URL: https://www.strategyzer.com/library?type=Tools#library [accessed 2025-01-12]
34. López Seguí F, Pratdepàdua Bufill C, Rius Soler A, de San Pedro M, López Truño B, Aguiló Laine A, et al. Prescription and integration of accredited mobile apps in Catalan health and social care: protocol for the AppSalut site design. JMIR Res Protoc 2018 Dec 21;7(12):e11414 [FREE Full text] [doi: 10.2196/11414] [Medline: 30578234]
35. Lopez Segui F, Pratdepadua Bufill C, Abdon Gimenez N, Martinez Roldan J, Garcia Cuyas F. The prescription of mobile apps by primary care teams: a pilot project in Catalonia. JMIR Mhealth Uhealth 2018 Jun 21;6(6):e10701 [FREE Full text] [doi: 10.2196/10701] [Medline: 29929948]
36. Criteris de certificació d'Apps de salut o benestar Versió 3.0. Fundació TIC Salut Social. URL: https://ticsalutsocial.cat/wp-content/uploads/2021/09/2021_Criteris-certificacio_CAT_v3.0.pdf [accessed 2025-01-12]
37. Llebot Casajuana B, Hoogendoorn P, Villalobos-Quesada M, Gómez Freixa M, Pratdepadua Bufill C. Integration of CEN ISO/TS 82304-2 for existing health authorities' app assessment frameworks: a comparative case study in Catalonia. JMIRP Preprint posted online October 22, 2024 [FREE Full text] [doi: 10.2196/preprints.67858]
38. UNI/TR 11708:2018: informatica medica - caratterizzazione delle APP nel contesto della salute, benessere e stili di vita. UNI. 2018. URL: https://store.uni.com/uni-tr-11708-2018 [accessed 2025-01-12]
39. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
40. Your research: Is it subject to the WMO or not? Central Committee on Research Involving Human Subjects. URL: https://english.ccmo.nl/investigators/legal-framework-for-medical-scientific-research/your-research-is-it-subject-to-the-wmo-or-not [accessed 2025-04-16]
41. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JE. Mixing online and face-to-face therapy: how to benefit from blended care in mental health care. JMIR Ment Health 2016 Feb 09;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]
42. Acord del Govern pel qual s'aprova el Pla Mestre de Mobilitat (mHealth.Cat). President of the Council of Ministers. URL: https://documents.dadesobertes.gencat.cat/acordsgovern/docs/SIG15SLT0018.pdf [accessed 2025-01-12]
43. Repertorio atto n. 215/CSR. Conferenza Permanente per i rapporti tra lo Stato, le Regioni e le Province Autonome di Trento e Bolzano. URL: https://www.statoregioni.it/it/conferenza-stato-regioni/sedute-2020/seduta-del-17122020/atti/repertorio-atto-n-215csr/ [accessed 2025-01-12]
44. Italia Domani, the National Recovery and Resilience Plan. Italian Government | Presidency of the Council of Ministers. URL: https://www.italiadomani.gov.it/content/sogei-ng/it/en/home.html [accessed 2025-01-12]
45. Actius Digitals per la Ciutadania: Self-assessment test. TIC Salut Social. URL: https://ticsalutsocial.cat/en/que-fem/digital-assets-for-citizens/test-dautoavaluacio/ [accessed 2025-01-12]
46. Label2Enable. Smeulders. URL: https://vimeo.com/showcase/11257693 [accessed 2025-01-12]
47. Recommendations for the reimbursement of health apps. Label2Enable. URL: https://label2enable.eu/assets/downloads/la2en-infographic-recommendations-v5-1716569726.pdf [accessed 2025-01-12]

XSL•FO
RenderX

48. Frey AL, Matei D, Phillips B, McCabe A, Fuller R, Laibarra B, et al. Testing and iterative improvement of the CEN ISO/TS 82304-2 health app quality assessment: pilot interrater reliability study. JMIR Form Res 2025 Mar 10;9:e64565 [FREE Full text] [doi: 10.2196/64565] [Medline: 40063936]
49. Biliunaite I, van Gestel L, Hoogendoorn P, Adriaanse M. Value of a quality label and European healthcare professionals' willingness to recommend health apps: an experimental vignette study. J Health Psychol 2024 Aug 02:13591053241258205 [FREE Full text] [doi: 10.1177/13591053241258205] [Medline: 39096027]
50. Torre D, Abualhaija S, Sabetzadeh M, Briand L, Baetens K, Goes P. An AI-assisted approach for checking the completeness of privacy policies against GDPR. In: Proceedings of the 28th International Requirements Engineering Conference. 2020 Presented at: RE '10; August 31-September 4, 2020; Zurich, Switzerland p. 136-146 URL: https://ieeexplore.ieee.org/document/9218152 [doi: 10.1109/re48521.2020.00025]
51. Mobile applications to support contact tracing in the EU's fight against COVID-19: common EU toolbox for member states. European Commission. URL: https://health.ec.europa.eu/document/download/01f83f72-4e21-4a34-90dd-ec0b0cc35c8b_en [accessed 2025-01-12]
52. Evidence standards framework for digital health technologies. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf [accessed 2025-01-12]
53. Dittrich F, Mielitz A, Pustozerov E, Lawin D, von Jan U, Albrecht U. Digital health applications from a government-regulated directory of reimbursable health apps in Germany-a systematic review for evidence and bias. Mhealth 2023 Oct;9:35 [FREE Full text] [doi: 10.21037/mhealth-23-17] [Medline: 38023782]
54. About ISO. International Organization for Standardization. URL: https://www.iso.org/about [accessed 2025-01-12]
55. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]
56. Dahlhausen F, Zinner M, Bieske L, Ehlers JP, Boehme P, Fehring L. Physicians' attitudes toward prescribable mHealth apps and implications for adoption in Germany: mixed methods study. JMIR Mhealth Uhealth 2021 Nov 23;9(11):e33012 [FREE Full text] [doi: 10.2196/33012] [Medline: 34817385]
57. Richter JG, Chehab G, Stachwitz P, Hagen J, Larsen D, Knitza J, et al. One year of digital health applications (DiGA) in Germany - Rheumatologists' perspectives. Front Med (Lausanne) 2022;9:1000668 [FREE Full text] [doi: 10.3389/fmed.2022.1000668] [Medline: 36388899]
58. Caiani EG, Kemps H, Hoogendoorn P, Asteggiano R, Böhm A, Borregaard B, et al. Standardized assessment of evidence supporting the adoption of mobile health solutions: a clinical consensus statement of the ESC Regulatory Affairs Committee: developed in collaboration with the European Heart Rhythm Association (EHRA), the Association of Cardiovascular Nursing & Allied Professions (ACNAP) of the ESC, the Heart Failure Association (HFA) of the ESC, the ESC Young Community, the ESC Working Group on e-Cardiology, the ESC Council for Cardiology Practice, the ESC Council of Cardio-Oncology, the ESC Council on Hypertension, the ESC Patient Forum, the ESC Digital Health Committee, and the European Association of Preventive Cardiology (EAPC). Eur Heart J Digit Health 2024 Sep;5(5):509-523. [doi: 10.1093/ehjdh/ztae042] [Medline: 39318699]
59. Exploring the digital health landscape in the WHO European Region: digital health country profiles. World Health Organization. URL: https://iris.who.int/handle/10665/376540 [accessed 2025-01-12]
60. Assessment framework for mHealth apps. Australian Digital Health Agency. URL: https://www.digitalhealth.gov.au/about-us/strategies-and-plans/assessment-framework-for-mhealth-apps [accessed 2025-01-12]
61. van Kessel R, Srivastava D, Kyriopoulos I, Monti G, Novillo-Ortiz D, Milman R, et al. Digital health reimbursement strategies of 8 European countries and Israel: scoping review and policy mapping. JMIR Mhealth Uhealth 2023 Sep 29;11:e49003 [FREE Full text] [doi: 10.2196/49003] [Medline: 37773610]
62. Work plan: A fit-for-purpose HTA framework for DHTs. EDiHTA. URL: https://edihta-project.eu/workplan/ [accessed 2025-01-12]
63. About the project. ASSESS DHT. URL: https://assess-dht.eu/the-project/ [accessed 2025-01-12]
64. Position on the European Health Data Space. CPME. URL: https://www.cpme.eu/api/documents/adopted/2022/11/cpme.2022-065.FINAL.CPME.position.EHDS.pdf [accessed 2025-01-12]

Abbreviations

AF: assessment framework
CEN: Comité Européen de Normalisation (European Committee for Standardization)
COREQ: Consolidated Criteria for Reporting Qualitative Research
EU: European Union
FTSS: Fundació TIC Salut Social (TIC Salut Social Foundation)
HCP: health care professional
HTA: health technology assessment
ISO: International Organization for Standardization

XSL•FO
RenderX

ISS: Istituto Superiore di Sanità (National Institute of Health)
mHealth: mobile health
TS: Technical Specification
VPC: value proposition canvas
WHO: World Health Organization

Edited by A Mavragani; submitted 22.10.24; peer-reviewed by J Haverinen, D Westmattelmann; comments to author 29.11.24; revised version received 19.12.24; accepted 14.02.25; published 21.04.25.

Please cite as:
Hoogendoorn P, Shokralla M, Willemsen R, Guldemond N, Villalobos-Quesada M
Compatibility of the CEN-ISO/TS 82304-2 Health App Assessment Framework With Catalan and Italian Health Authorities' Needs: Qualitative Interview Study
JMIR Form Res 2025;9:e67855
URL: https://formative.jmir.org/2025/1/e67855
doi:10.2196/67855
PMID:

©Petra Hoogendoorn, Mariam Shokralla, Romy Willemsen, Nick Guldemond, María Villalobos-Quesada. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Mobile Health App Informed by the Multi-Process Action Control Framework to Promote Physical Activity Among Inactive Adults: Iterative Usability Study

Heather Hollman[1], MRSc, DC; Wuyou Sui[2], PhD; Haowei Zhang[3], BSc; Ryan E Rhodes[3], PhD

[1]Behavioural Medicine Lab, School of Exercise Science, Physical and Health Education, University of Victoria, Victoria, BC, Canada

[2]School of Kinesiology, Western University, London, ON, Canada

[3]Behavioural Medicine Lab, School of Exercise Science, Physical & Health Education, University of Victoria, Victoria, BC, Canada

Corresponding Author:
Heather Hollman, MRSc, DC
Behavioural Medicine Lab
School of Exercise Science, Physical and Health Education
University of Victoria
3800 Finnerty Rd
Room 173 McKinnon Building
Victoria, BC, V8P 5C2
Canada
Phone: 1 250 472 5288
Email: heatherh@uvic.ca

Abstract

Background: Mobile health apps have high potential to address the widespread deficit in physical activity (PA); however, they have demonstrated greater impact on short-term PA compared to long-term PA. The multi-process action control (M-PAC) framework promotes sustained PA behavior by combining reflective (eg, attitudes) and regulatory (eg, planning and emotion regulation) constructs with reflexive (eg, habits and identity) constructs. Usability testing is important to determine the integrity of a mobile health app's intrinsic properties and suggestions for improvement before feasibility and efficacy testing.

Objective: This study aimed to gather usability feedback from end users on a first and a second version of an M-PAC app prototype.

Methods: First, 3 workshops and focus groups, with 5 adult participants per group, were conducted to obtain first impressions of the M-PAC app interface and the first 3 lessons. The findings informed several modifications to the app program (eg, added cards with reduced content) and its interface (eg, created a link placeholder image and added a forgot password feature). Subsequently, a single-group pilot usability study was conducted with 14 adults who were not meeting 150 minutes per week of moderate-to-vigorous PA. They used the updated M-PAC app for 2 weeks, participated in semistructured interviews, and completed the Mobile App Usability Questionnaire (MAUQ) to provide usability and acceptability feedback. The focus groups and interviews were recorded, transcribed, and analyzed with content analysis informed by usability heuristics. The MAUQ scores were analyzed descriptively.

Results: Participants from the workshops and focus groups (mean age 30.40, SD9.49 years) expressed overall satisfaction with the app layout and content. The language was deemed appropriate; however, some terms (eg, self-efficacy) and acronyms (eg, frequency, intensity, time, and type) needed definitions. Participants provided several recommendations for the visual design (eg, more cards with less text). They experienced challenges in accessing and using the help module and viewing some images, and were unsure how to create or reset the password. Findings from the usability pilot study (mean age 41.38, SD12.92 years; mean moderate-to-vigorous PA 66.07, SD57.92 min/week) revealed overall satisfaction with the app layout (13/13, 100%), content (10/13, 77%), and language (7/11, 64%). Suggestions included more enticing titles and additional and variable forms of content (eg, visual aids and videos). The app was easy to navigate (9/13, 69%); however, some errors were identified, such as PA monitoring connection problems, broken links, and difficulties entering and modifying data. The mean MAUQ total and subscale scores were as follows: total=5.06 (SD1.20), usefulness=4.17 (SD1.31), ease of use=5.36 (SD1.27), and interface and satisfaction=5.52 (SD1.42).

XSL•FO
RenderX

Conclusions: Overall, the M-PAC app was deemed usable and acceptable. The findings will inform the development of the minimum viable product, which will undergo subsequent feasibility testing.

(JMIR Form Res 2025;9:e59477) doi:10.2196/59477

KEYWORDS

physical activity; mobile apps; mobile health; mHealth; usability study; inactive adults; smartphone

Introduction

Background

Regular physical activity (PA) is associated with reduced risk of cardiovascular disease, hypertension, diabetes, breast and colon cancer, and mortality [1]. It also has positive effects on cognition and mental health [2,3]. Unfortunately, PA rates worldwide are insufficient to reap these mental and physical health benefits [4]. Given the widespread burden of global inactivity, scalable PA interventions are needed. Mobile health (mHealth) apps have great potential for scalability and to meet harder-to-reach populations, such as rural residents, who are also at increased risk of inactivity [5]. As of 2024, >60% of the global population owns a smartphone, with even greater rates in high-income nations such as the United Kingdom (82%), the United States (82%), and Canada (72%) [6]. People spend the majority (88%) of their mobile time using apps [7], and in 2019, the average individual used 9 mobile apps per day and 30 apps per month [8].

mHealth apps are an appealing medium for PA behavioral interventions because they can provide up-to-date educational material while concurrently enabling continuous PA tracking and feedback [9-11]. They can also facilitate connections with peers and health providers [9-11] and more easily implement personalization and associations (eg, reminders) compared to print or in-person interventions [12]. From a theoretical standpoint, they can deliver a higher amount of behavior change techniques (BCTs) compared to in-person interventions [12]. Using well-defined BCTs (eg, goal setting and self-monitoring) has been associated with increased PA intervention engagement, which is an essential metric to address given the high app attrition rate and the connection between engagement and resulting PA behavior [13-15].

In line with the additional BCTs and related features, mHealth apps have demonstrated promising potential to impact PA behavior. In a recent meta-analysis of PA mHealth apps or trackers with automated and continuous self-monitoring and feedback, Laranjo et al [16] reported a small-to-moderate effect on PA (standardized mean difference [SMD]=0.350; 95% CI 0.236-0.465; I^2=69%; T^2=0.051) with a mean follow-up of 13 weeks. These standardized effects are higher than meta-analyses of workplace PA interventions (SMD=0.21) [17], internet-delivered PA interventions (SMD=0.14) [18], and face-to-face PA interventions (SMD=0.29) [19] and are equivalent to the 85th percentile of benchmarked PA intervention effects [20]. PA mHealth app interventions have also demonstrated longer-term (>6 months) effects on PA; however, the effect sizes tend to decrease over time [21], similar to PA interventions in general [22]. Because the health benefits can only be accrued if PA behavior change is sustained, PA maintenance, in addition to engagement, should be a focus of future PA mHealth apps.

To determine intervention components needed for longer-term PA behavior, PA theorists have attempted to conceptualize the acts of PA behavior adoption and maintenance separately [23,24]. For example, although constructs of traditional PA theories in the social cognitive domain (eg, outcome expectations or action planning) seem to be moderate predictors of PA adoption, they may be less comprehensive for maintenance [24,25]. Social cognitive theories highlight the importance of forming PA intentions; however, these approaches sometimes fail to account for the intention-PA gap (ie, the proportion of individuals who do not follow through with regular PA behavior that aligns with their PA intentions) [26-28]. A recent meta-analysis showed that 48% of adult PA intenders fail to follow through with PA [29].

Several theoretical approaches have been developed to close the intention-behavior gap [27,30]. An example is the multi-process action control (M-PAC) framework, which combines reflective processes, such as those in traditional social cognitive approaches (eg, attitudes), with regulatory (eg, self-monitoring and emotion regulation) and reflexive (eg, habit and exercise identity) processes to facilitate PA behavior change [31]. Collectively, the M-PAC constructs have a causal structure that begins with intention formation, progresses to the adoption of action control, and ultimately leads to sustained action control. Each stage is interconnected, with processes that are mutually reinforcing and reciprocal [32]. Each construct independently provides meaningful and significant contributions to translating positive intentions to PA behavior, with the largest predictive effect sizes coming from the latter reflexive processes [31]. PA interventions (in-person and web-based) informed by the M-PAC framework have demonstrated promising results on PA behavior and PA maintenance, promoting constructs of habit, self-identity, and behavioral regulation in feasibility trials [33-36]. However, an mHealth app informed by the M-PAC framework is yet to be tested.

To advance the likelihood of a successful and scalable mHealth app, it is recommended to assess its intrinsic properties, including usability and feasibility, in the initial prototyping phases [37]. An initial focus on these intrinsic properties can ultimately lead to greater attribution of the extrinsic metrics and outcomes to the intervention. Usability is the assessment of the quality of the interaction between the user and the intervention, while feasibility is the assessment of the ability of the intervention to work as intended in a given context [37]. Usability testing can help to uncover the overall interest in and eagerness to use the mHealth app, potential impediments to its use, and suggestions for improvement or innovation [38]. This

XSL•FO
RenderX

usability phase is usually iterative, whereby prototypes are refined and optimized through several stages based on user feedback. It is also essential to involve end users to ensure that the intervention is relevant and meets the specific needs and preferences of the target population [39,40]. Once this phase is complete, a minimum viable product (MVP) can be built and then pilot-tested to test for further usability and acceptability, followed by evaluation of its efficacy [38].

Research Objective

Thus, the objective of this research was to gather user feedback on the first and second prototypes of the M-PAC app to inform the design and development of the MVP. We initially explored the first impressions of the M-PAC app prototype using focus group methodology to gather important preliminary feedback from a group with diverse views and perspectives [41]. Subsequently, we used a usability pilot study [42] to retrieve more in-depth feedback on the usability and acceptability of the M-PAC app from potential users—individuals who intend to exercise but are not currently meeting the recommended PA guidelines.

Methods

Overarching App Development Framework

The M-PAC app development and design were in accordance with the IDEAS (integrate, design, assess, and share) framework [38]. Specifically, this study described 2 iterative rounds of phases 5 (prototype potential products) and 6 (gather user feedback) [38]. As per the IDEAS framework phase 5, an initial prototype was assessed during a series of workshops and focus groups to gather first impressions and suggestions for improvement. Commensurate with phase 6 of the IDEAS framework, the suggestions from the workshops and focus groups were implemented into the M-PAC app, and subsequently, the updated second app prototype was assessed in a usability pilot study, in which a new group of participants had the opportunity to use the app as intended for a few weeks and then provide additional feedback in individual interviews.

Phase 1 (workshops and focus groups) and the qualitative interview component of phase 2 (usability pilot) were reported following the COREQ (Consolidated Criteria for Reporting Qualitative Research) criteria [43]. Phase 2 (usability pilot) was reported following a modified version of the CONSORT (Consolidated Standards of Reporting Trials) guidelines [44], as recommended by Lancaster and Thabane [45] for nonrandomized pilot studies.

Ethical Considerations

These studies received ethics approval from the University of Victoria Human Research Ethics Board (21-0248; 21-0360). All participants provided informed consent before participating in the studies, and all study data are anonymous. Participants in phase 2 (usability pilot) received a CAD $15 (US $10.88) honorarium for participating in the study.

Phase 1: Workshops and Focus Groups

Design

Online hands-on workshops followed by focus groups were conducted. The workshops were conducted between November 2021 and March 2022. Focus groups were designed to be run with 4 to 6 participants, with a total of 15 to 20 participants, based upon previous usability studies using apps and web-based delivery [46-49]. The workshops and focus groups were led by WS and supported by HH (who also took field notes). WS has a PhD in health psychology, was a postdoctoral fellow at the time of the study, and identifies as a man. HH has a master's degree in rehabilitation science, was a PhD student at the time of the study, and identifies as a woman. WS had prior experience in conducting research focus groups and interviews and HH did not. WS and HH both have prior experience analyzing qualitative data. WS and HH introduced themselves to the participants as research trainees with a focus on PA behavior change.

Participants

Eligible participants were adults aged 19 to 64 years who owned a smartphone with either an iOS or Android operating system and were fluent in English. Participants were recruited via online advertisements on the research laboratory's public X account (X Corp; @bmedlab); public Instagram account (Meta Platforms, Inc; @uvicbmed); public community Facebook (Meta Platforms, Inc) groups local to Victoria, British Columbia; and through snowball sampling. All communication between researchers and participants took place over email. Some participants were known to the researchers because they worked in the same university; however, none of the participants had provided feedback on the app before the study.

Procedures

Interested participants contacted the researchers via email and were sent an email with the letter of information and consent form. Upon providing consent, participants were asked to provide availability for a 60-minute workshop and focus group. Participants were informed that they would be given a pseudonym during the workshops and focus groups to protect their identity. Participants were later contacted to reconfirm their availability once 4 to 5 participants reported having a mutually available time and were sent a Zoom videoconferencing link upon confirmation (Zoom Video Communications). In addition, participants were asked to download the Pathverse app platform [50] (Pathverse, Inc), which gave them access to the M-PAC app on their smartphone, to be able to actively participate in the workshop portion of the session.

One day before the date of the workshop or focus group, participant emails were used to create profiles on the admin portal of Pathverse, so that participants would have access to the M-PAC app. On the arranged time of the workshop or focus group, participants joined into a Zoom call with 2 researchers. All participants were given pseudonyms and were unable to turn on video. The lead researcher shared their smartphone screen and led the workshop and focus group discussion, while the other researcher managed the video recording software and

XSL•FO
RenderX

any questions within the Zoom chat. Upon entering the Zoom call, the lead researcher asked participants to open the Pathverse app and then instructed participants on how to create a profile on Pathverse using their email. Once all the participants had access to the app, the lead researcher began a walkthrough of the app.

Participants were first shown how to access the lessons, along with a brief overview of how to navigate through lesson cards, the different types of cards, and how to mark a lesson as complete, using "Lesson 0: Tutorial" and "Lesson 1: What is Physical Activity" as examples. Afterward, participants were directed to the tracking features on the app. Participants were shown how to sync their step count to the app as well as how to track and log PA, weight, and blood pressure. Participants were then shown the goal-setting feature of the app. The lead researcher created an example goal, demonstrating to participants how to name the goal, set priority and reminders on the goal, and how to mark goals as completed. Finally, participants were shown how to navigate to the home page of the app and to find the in-app navigation tutorial if they got stuck. Participants were then given 5 to 10 minutes to explore the app and its features independently. Overall, the workshop portion of the session took approximately 25 to 30 minutes.

Following the workshop, participants were informed that they would be starting the focus group portion of the session. The lead researcher read the introductory script and then prompted participants to ask any questions before beginning with the list of semistructured questions (Multimedia Appendix 1). The focus group portion of the session took approximately 25 to 30 minutes. The workshop and focus group guide was pilot-tested among the research team. Only the participants and researchers were present during the workshops and focus groups. There were no repeat workshops and focus groups.

Intervention: The M-PAC App

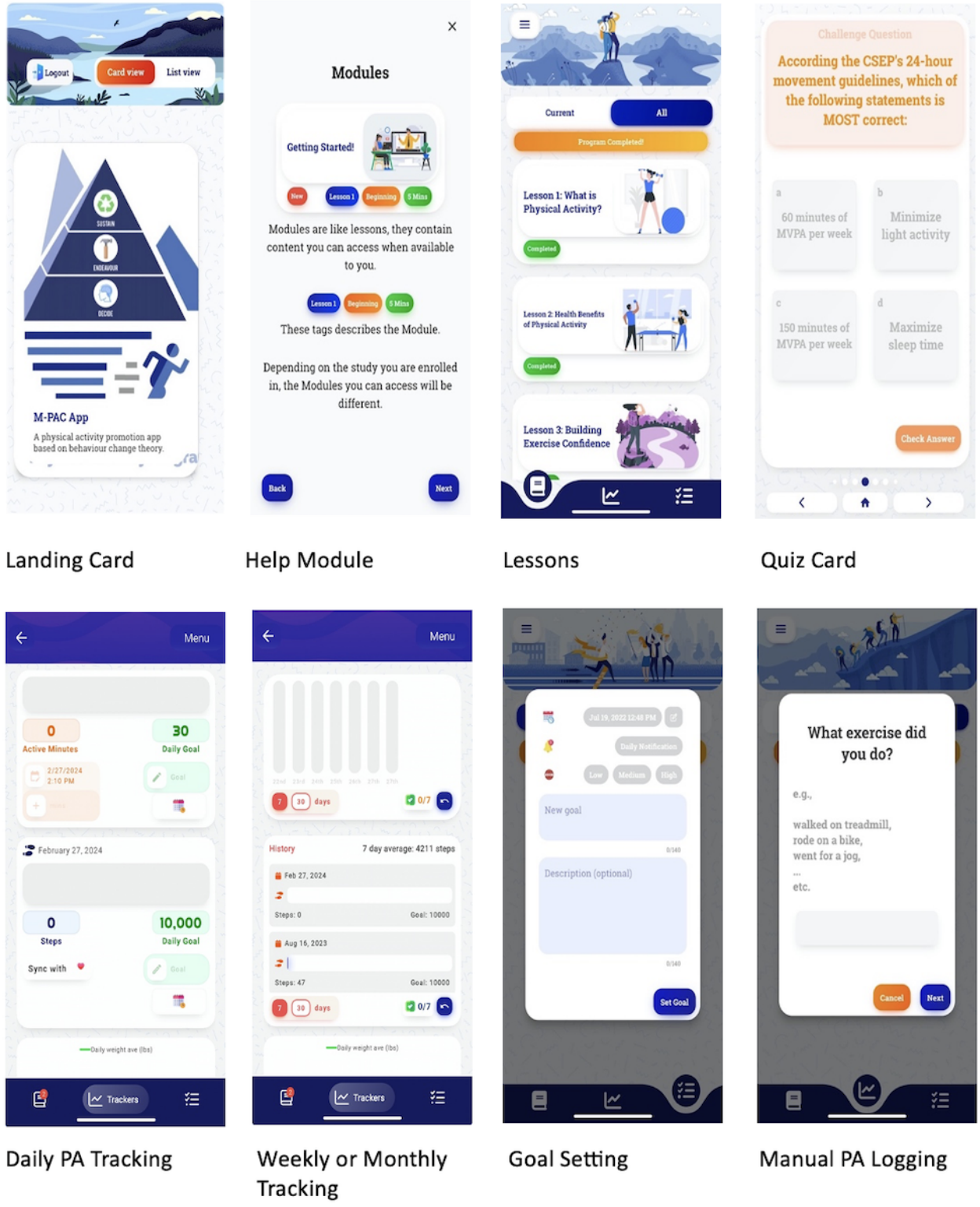
The M-PAC app is an evidence-based mobile app designed to support users in changing their PA behavior, developed by researchers specializing in PA behavior. The app is built upon the Pathverse app platform [50]. Pathverse is an mHealth development platform designed for health and wellness professionals and researchers. This platform consists of an admin web-portal interface and a participant Pathverse app (available in both IOS and Android stores). The M-PAC app, specifically, includes educational content in the form of lessons; self-monitoring features, such as step tracking and PA logging; and a goal-setting feature, whereby users can set goals for themselves as well as review completed goals (Figure 1).

Lessons within the app are broken down into individual "cards," which users navigate by swiping left or right, or via the arrow buttons at the bottom of the screen. Cards present either (1) a brief block of text, (2) a brief block of text with an image placed below or above, (3) a brief block of text with a link to an external web page, or (4) a challenge question. The challenge question card presents the user with a short question and 4 possible answers. Selecting any answer presents the user with a new slide indicating whether their answer was correct or incorrect, as well as a brief explanation for the correct answer.

Lessons within the app are scaffolded according to the different constructs within the M-PAC framework. Lessons 1 to 4 are designed to address reflective processes of the M-PAC framework: instrumental attitudes ("What are the health benefits of PA?"), perceived opportunity ("Do I have the resources to do PA?"), perceived capability ("Am I confident I can perform this PA?"), and affective judgments ("Do I like PA?"). Similarly, lessons 5 to 10 are aimed at addressing the regulatory (ie, skills important in facilitating continued PA behavior, such as goal setting and planning) and reflexive processes (eg, habit and exercise identity) of the M-PAC. In addition, lessons contain challenge cards that assess users' understanding of the content within their current lesson. Refer to Multimedia Appendix 2 [51] for a full list of app lessons, mechanisms of action, BCTs, and behavior change ontologies applied [52,53]. Lessons must be completed in order (ie, lesson 0, lesson 1, lesson 2, etc) to unlock subsequent lessons, but completed lessons can be revisited once unlocked. Given the usability focus of this study, only lessons 1 to 3 (including a lesson 0: tutorial) were presented to participants.

XSL•FO
RenderX

Figure 1. Screenshots of a physical activity mobile health app informed by the multi-process action control (M-PAC) framework (the M-PAC app). PA: physical activity.



Outcomes

The primary outcome of this phase was qualitative feedback on the usability of the M-PAC app drawn from the focus groups. Specifically, questions within the focus group were drawn from Nielsen's 10 usability heuristics for user interface design [54]. These heuristics were adapted for this work to guide the structured focus group questions, which addressed the following: layout and information, appropriateness of language, backing out or undoing an action, consistency in language and features, errors and mistakes, forgetting acronyms, efficiency of interface and navigation, visual design, ease in diagnosing and resolving errors, and help and documentation. Probing questions for each main question were also planned if group responses were vague or open ended. For example, the first main question was, “How did you find the layout and information of the app?” with

XSL•FO
RenderX

probing questions of "If intuitive, what made the layout or information intuitive?" and "If unintuitive, what made the layout or information unintuitive? What would be more helpful?" The specific questions related to these heuristics can be found in Multimedia Appendix 3. Participant responses were recorded using the video recording software Open Broadcast Software (OBS) Studio [55] (OBS Studio Contributors).

Participant demographics of age, gender, and highest attained education were assessed using single-item questions. Furthermore, the brand of smartphone owned, ability to use the features of a smartphone, ability to engage with other technologies, and whether a PA app has been downloaded or used in the past were also assessed using single-item measures.

Analysis

Focus groups were recorded using OBS Studio and transcribed verbatim [55]. Transcripts were coded and analyzed thematically, according to the 10 usability heuristics [54] adapted for this study, using NVivo (QSR International) [56]. Participants were not asked to provide feedback on the transcripts or findings. Two coders independently coded the transcripts, with conflicts being resolved by a third coder.

Phase 2: Usability Pilot

Design

A 2-week, single-group pilot study was conducted virtually between August 2022 and February 2023. The researchers were located in Victoria, British Columbia, Canada, and the participants were located anywhere in Canada.

Participants

Eligible participants were Canadian adults aged 19 to 64 years who owned a smartphone device with either an iOS or Android operating system, were fluent in English, and reported engaging in <150 minutes of moderate-to-vigorous PA (MVPA) per week. Individuals were excluded if they participated in the phase 1 workshops or focus groups or were involved in the design and development of the M-PAC app. Participants were recruited via online advertisements on the research laboratory's public X account (@bmedlab), public Instagram account (@uvicbmed), and public community Facebook groups local to Victoria, British Columbia; with paid Facebook advertisements; and snowball sampling. All communication between researchers and participants took place over email. A total of 10 to 15 participants was deemed appropriate based on previous usability testing studies examining apps [57,58].

Procedures

Interested individuals contacted the research team via email and were sent an email with the letter of information and a consent form. Upon agreeing to participate in the study, participants were sent instructions on how to download the M-PAC app, as well as a demographics questionnaire. Participants also booked a 30-minute baseline meeting with a researcher to walkthrough the M-PAC app over Zoom. During the baseline meeting, participants were given a walkthrough of the M-PAC app, identical to the workshop from phase 1. Upon completing the walkthrough, participants were asked to schedule a 30-minute exit interview in 2 weeks and finish lessons 1 to 3 within those 2 weeks.

During the exit interview, the researcher followed the semistructured interview guide to assess participants' experiences with the M-PAC app (Multimedia Appendix 3). The interview guide was pilot-tested among the research team. Interviews were conducted by WS over Zoom, audio recorded using OBS Studio, and lasted 20 to 30 minutes. Only the participants and researchers were present, and there were no repeat interviews. After completing the exit interview, participants were told that they would be sent the Mobile App Usability Questionnaire (MAUQ) to their email in 24 hours, and to complete the MAUQ within 3 days of receiving the questionnaire [59]. Upon completion of the MAUQ, participants were given a CAD $15 (US $11.88) honorarium. There were no changes made to the methods upon pilot study commencement.

Intervention

Following the completion of phase 1, several modifications were made to the M-PAC app. Changes to content included the following: (1) modifying the introductory card for the self-efficacy lesson from "Introduction to self-efficacy..." to "Ways to improve confidence (also known as self-efficacy)...", (2) ensuring all acronyms (eg, MVPA; Canadian Society for Exercise Physiology; frequency, intensity, time, and type [FITT]; and Get Active Questionnaire) were clearly defined, (3) bolding some terms to highlight them (eg, FITT), (4) reducing the amount of text on cards and increasing the number of cards per lesson, and (5) adding a references module and providing links to references from each lesson. Changes to the Pathverse platform included the following: (1) optimizing image loading times, (2) creating an external link placeholder image that reads "Link," and (3) adding a "Forgot your password?" feature. The updated version of the M-PAC app with the aforementioned modifications was used for the usability pilot study.

Outcomes

The primary outcome of this phase was qualitative feedback on the usability and satisfaction of the M-PAC app, gathered through the exit interview. Probing questions for each main question were also planned if participants' responses were vague or open ended. For example, for the question, "Can you tell me what you liked best about the M-PAC app?", probing questions of "Were there any features in particular?" and "What about the content and/or layout?" were planned. The specific questions within this interview guide can be found in Multimedia Appendix 3. Participant responses were recorded using the video recording software OBS Studio [55]. The a priori criteria for moving on to a randomized pilot trial was an overall consensus that the intervention was usable (MAUQ score≥5), and at least 60% (8/13) of participants were satisfied with the intervention, with minor to moderate suggestions for improvement.

Quantitative feedback on the usability of the M-PAC app was collected via the MAUQ [59], specifically the version for stand-alone mHealth apps. This version of the MAUQ is an 18-item self-report questionnaire that assesses an mHealth app's

XSL•FO
RenderX

ease of use, interface and satisfaction, and usefulness. The questionnaire demonstrated construct and criterion validity and reliability, with Cronbach α=0.91 for the total MAUQ, 0.71 for ease of use, 0.93 for interface and satisfaction, and 0.65 for usefulness internal consistency.

Participant demographics of age, gender, and highest attained education were assessed using single-item questions. Furthermore, the brand of smartphone owned, ability to use the features of a smartphone, ability to engage with other technologies, and whether a PA app has been downloaded or used in the past were also assessed using single-item measures. Participants' average weekly PA was assessed using the modified Godin Leisure-Time PA questionnaire [60]. The estimated weekly MVPA was calculated by multiplying the frequency and duration of moderate- and vigorous-intensity PA minutes and adding them together. The Godin-Leisure-Time PA questionnaire has demonstrated acceptable validity and reliability in a variety of settings, populations, and countries [60,61]. There were no changes made to the assessments or measurements after the trial commenced.

Analyses

Exit interviews were recorded using OBS Studio and transcribed verbatim. Transcripts were coded and analyzed thematically, according to the 10 usability heuristics [54] adapted for this study, using NVivo [56]. Participants were not asked to provide feedback on the transcripts or findings. Two coders independently coded the transcripts, with conflicts being resolved by a third coder.

Descriptives (ie, mean and SD) for quantitative usability feedback from the MAUQ were calculated for total MAUQ scores, as well as for the ease of use, interface and satisfaction, and usefulness subscales [59].

Results

Phase 1: Workshops and Focus Groups

Participants

We conducted 3 focus groups of 5 participants each. Participants were primarily women (13/15, 87%) with a mean age of 30.4 (SD 9.49) years. Most had at least a college or university degree (14/15, 93%) and self-reported having above average abilities to use the functions of their smartphone (11/15, 73%), and just under half had previously downloaded a mobile app to help increase their PA (7/15, 47%; eg, Down Dog, Fitbit, Health, Nike Training Club app, Endomondo, and Sweat app). The most common brand of smartphone was Apple iPhone (9/15, 60%), followed by Samsung (4/15, 27%) and Google Pixel (2/15, 13%; Table 1). None of the participants refused to participate or dropped out of the study.

XSL•FO
RenderX

Table 1. Baseline characteristics of the workshops, focus groups, and pilot study participants.

	Workshops and focus groups (n=15)	Pilot (n=14)[a]
Age (y), mean (SD)	30.4 (9.49)	41.38 (12.92)
Gender, n (%)		
Women	13 (87)	11 (85)
Men	2 (13)	2 (15)
British Columbia	__[b]	Yes=8 (62) and no=5 (38)
Country (Canada)	15 (100)	13 (100)
Highest education, n (%)		
High-school diploma	1 (7)	1 (8)
Vocational school or some college	0 (0)	3 (23)
College or university degree	6 (40)	5 (38)
Professional or graduate degree	8 (53)	4 (31)
Brand of smartphone, n (%)		
Google Pixel	2 (13)	1 (8)
Samsung	4 (27)	4 (31)
Apple iPhone	9 (60)	5 (38)
LG	0 (0)	1 (8)
Huawei	0 (0)	1 (8)
Motorola	0 (0)	1 (8)
Ability to use the functions of their smartphone[c], n (%)		
Average	4 (27)	0 (0)
Good	6 (40)	5 (38)
Excellent	5 (33)	8 (62)
Abilities to engage with and use other types of technology generally[c], n (%)		
Poor	1 (7)	0 (0)
Average	3 (20)	0 (0)
Good	6 (40)	7 (54)
Excellent	5 (33)	6 (46)
Previously downloaded a mobile app to help increase their physical activity[d], n (%)		
Yes	7 (47)	8 (62)
Down Dog	2 (13)	0 (0)
Fitbit	1 (7)	2 (13)
Health	1 (7)	0 (0)
Nike training club app	1 (7)	0 (0)
Endomondo	1 (7)	0 (0)
Sweat App	1 (7)	0 (0)
Optimity	0 (0)	2 (15)
Carrot	0 (0)	1 (8)
Habitshare	0 (0)	1 (8)
MyFitnessPal	0 (0)	1 (8)
C25K	0 (0)	1 (8)
ParticipACTION app	0 (0)	1 (8)

XSL•FO
RenderX

	Workshops and focus groups (n=15)	Pilot (n=14)[a]
7M workout	0 (0)	1 (8)
No	8 (53)	5 (38)
If you have previously downloaded an app to increase your physical activity, was it helpful?[d], n (%)		
Yes	3 (20)	6 (46)
No	4 (27)	2 (15)
Physical activity behavior score, mean (SD)		
Moderate physical activity	—	54.29 (52.87)
Vigorous physical activity	—	11.79 (20.90)
MVPA[e]	—	66.07 (57.92)

[a]One participant did not complete the baseline questionnaire.

[b]Not applicable.

[c]Scale choices: terrible, poor, average, good, and excellent.

[d]Scale choices: yes and no.

[e]MVPA: moderate- to-vigorous physical activity.

Focus Group Findings

Overview

Refer to Table 2 for a comprehensive overview of the findings from the focus groups. Data saturation, where no new themes were revealed, was reached after 3 focus groups, which is in line with research stating that data saturation can be achieved with 3 to 5 focus groups [62].

XSL•FO
RenderX

Table 2. Multi-process action control (M-PAC) app usability focus group themes informed by the usability heuristics by Nielsen [54].

Heuristic theme	Positive findings	Negative findings and suggestions
Layout and information (including information on features)	• The flow was nice and attractive and easy to navigate • The 3 icons at the bottom were straightforward • Liked the card format	• Multiple levels of content for people who want more information • The menu button should be reworded (does not take the user to a menu) • Some people wanted the congratulations card before the reference card • Trackers and self-reports should be in the same tab
Appropriateness of language	• Language was appropriate for the general audience • Lay terms that anyone could understand • Fun theory	• Language was for individuals who had higher physical literacy • Once in a while, some words could have been different • Some terms, such as self-efficacy, could be initially presented using a more widely known term • Base screen says “activate account” instead of “create an account,” which is not intuitive
Backing out or undoing an action	• Good that there were two options to go back—you could slide or also click • Consensus that this was easy to do	• Upon returning to a lesson from the home page, one needed to restart the lesson • When they tracked a self-monitoring measure (blood pressure, PA[a], or weight), it could not be edited or deleted • Completed goals had a clickable “Completed” button, which made the goal “uncompleted” and took it back to the current tab
Consistency in language and features	• Generally, people thought language and features were consistent	• Modules versus lessons • Icon differences in goals depending on Android or Apple phone (eg, “I get a little house a little running man and then something that looks like a golf a putting green flag”) • Some things did not show up on iPad (long-term benefits with percentages)
Errors and mistakes	—[b]	• Connection to Fitbit not working • Some images did not load • Some images were glitching (24-hour movement guideline) • No way to reset the password • For Android users, external links led to blank white pages • Some images showed up too small
Forgetting acronyms	—	• Sometimes an acronym was presented without defining it (MVPA[c], CSEP[d], FITT[e], and PAR-Q[f])
Efficiency of interface and navigation	• Easy to swipe with 1 hand	• Ability to connect to other PA-measure devices • The swipe left button was too small • Some tabs were swipe and click, while others were only click • Desire to include the ability to link out to external web pages for references • Each page should have a home page button
Visual design	• People liked that similar images and illustrations were used • People liked the color scheme • People liked the representation among the images • The design of the tabs was clean	• Bold the FITT • Too much text on some cards • The external link placeholder image was too small and does not intuitively make people think it was a link • Wanted more visual pop within the lessons • Wanted the landscape format • The check question button on the quiz answer card was not in an intuitive location • On some pages, there was a lot of dead white space at the bottom • One person wanted real images rather than the clip art • More cards with less text • When setting goals, the grey buttons were not intuitive and should be displayed in a different color for better clarity • Program card could be better optimized (currently the title is cut off, and it flips to a sample page) • Would like more color on some of the slides

XSL•FO
RenderX

Heuristic theme	Positive findings	Negative findings and suggestions
Ease in diagnosing and resolving errors	—	• Could not figure out password components (ie, how many letters or numbers were required) • Needed to go through the entire set of help slides to find a solution • The help button was hard to find
Help and documentation	—	• Provided information on how to connect the Fitbit • Help cards did not have a home button and could not be skipped • The help button should be more accessible • Ability to send error reports on specific cards, in the help tab, or at the end of modules • Had to slide through the help slides individually • Help icon could be changed
Additional features	—	• Wanted to log other metrics, such as distance • Want a progress bar • The switch studies button should be a separate button • Wanted a sound at the end of the lesson • Desire to bookmark certain pages • Desire for a toolbox for important questionnaires or favorite cards

[a]PA: physical activity.

[b]Not applicable.

[c]MVPA: moderate-to-vigorous physical activity.

[d]CSEP: Canadian Society for Exercise Physiology.

[e]FITT: frequency, intensity, time, type.

[f]PAR-Q: Physical Activity Readiness Questionnaire.

Theme 1: Layout and Information

Participants stated that overall the app was attractive and had straightforward icons. They also enjoyed that the app used cards to convey information, and they liked the flow when swiping between cards. However, participants disliked that the "Menu" button only allowed them to switch studies instead of lessons. Moreover, participants asked for multiple depths of content for people with various physical literacy levels.

Theme 2: Appropriateness of Language

Participants agreed that the app used lay terms to make the language appropriate and understandable for the general audience. Nevertheless, they suggested that a more PA-literate population might find the language too simple. In addition, the definition of some terms (eg, self-efficacy) could have been introduced earlier to enhance users' understanding.

Theme 3: Backing Out or Undoing an Action

Participants liked that the app offered an easy way to navigate back from a page. For instance, they liked that the app had 2 options to return to the previous page (ie, by clicking or sliding). However, participants disliked that they needed to restart the lesson once they returned to the home page, could not edit or delete self-monitoring measures such as blood pressure and weight, and completed goals still had a clickable "completed" button, which uncompleted the goal.

Theme 4: Consistency in Language and Features

Generally, the participants agreed that the app's language and features were consistent. However, participants reported inconsistency in wording and icons across different platforms.

Theme 5: Errors and Mistakes

Participants reported that the app could not connect to Fitbit. In addition, some images glitched (eg, showed up too small) or would not load. Participants also stated that they could not reset the password for the app. Another error was that external links led to blank pages for Android users.

Theme 6: Forgetting Acronyms

The focus group participants suggested that sometimes the lessons introduced acronyms (eg, MVPA, CSEP, FITT, and PAR-Q) without providing definitions.

Theme 7: Efficiency of the Interface and Navigation

The participants liked that they could easily swipe through the app when navigating. However, they reported that some tabs could only be clicked on, whereas others could be swiped through and clicked on. Participants also thought the swipe left button was too small, making clicking difficult.

Theme 8: Visual Design

The focus groups liked that the images and illustrations were coherent in style throughout the app. They also enjoyed the color scheme, visual representation, and the design of the tabs. However, participants did not like that key terms, such as FITT, were not bolded. They also disfavored that certain cards were text heavy.

XSL•FO
RenderX

Theme 9: Ease in Diagnosing and Resolving Errors

The participants reported that the app did not provide information about the password requirements when creating a password for log-in. They also disliked that the "help" button was hard to find, and they needed to go through the entire set of slides to find a solution to their problems.

Theme 10: Help and Documentation

Participants disliked that the app did not provide sufficient information on how to connect to Fitbit, the "Help" cards were not accessible (eg, the "Help" cards did not include a "Home" button and users could not skip to the card they wanted), and the lack of the ability to send error reports on specific cards.

Theme 11: Additional Suggestions on Features

Participants asked for features such as a progress bar in the lessons, a sound effect when finishing a lesson, a toolbox for important or favorite cards, a "home" button on each page, hyperlinks to reference articles, the ability to bookmark specific pages, and the option to input different PA metrics (eg, distance walked).

Phase 2: Usability Pilot

Participants

Fourteen participants took part in phase 2 (usability pilot). One participant did not complete the baseline questionnaire due to unknown reason. Participants were mainly women (11/13, 85%) with a mean age of 41.38 (SD 12.92) years. Most (9/13, 69%) had at least a college or university degree, all (13/13, 100%) self-reported having above average abilities to use the functions of their smartphone, and most (8/13, 62%) had previously downloaded a mobile app to help increase their PA(eg, Fitbit, Optimity, Carrot, Habitshare, MyFitnessPal, C25K, ParticipACTION app, and 7M Workout). The most common brand of smartphone was Apple iPhone (5/13, 38%), followed by Samsung (4/13, 31%), Google Pixel (1/13, 8%), Hauwei (1/13, 8%), LG (1/13, 8%), and Motorola (1/13, 8%; Table 1). The participants reported on average 66.07 (SD 57.92) minutes of MVPA per week. None of the participants refused to participate or dropped out of the study.

Interview Findings

Overview

Refer to Multimedia Appendix 4 for a comprehensive overview of the findings from the interviews. A total of 14 interviews were conducted; however, only 13 interviews were recorded due to a malfunction of the recording instrument. Data saturation, where no new themes were revealed, was reached after 13 interviews, which is in line with research stating that data saturation can be achieved with 8 to 16 interviews [62].

Theme 1: Layout and Information

Of the 13 participants, 13 (100%) commented on the layout and information of the app, and everyone indicated something they liked (eg, straightforward layout and modules took a reasonable amount of time to complete). Of the 13 participants, 6 (46%) disliked the app's layout and information due to inefficient content layout (eg, too many pages for a lesson), content and questions being too basic, and some content lacking references.

Theme 2: Appropriateness of Language

Of the 11 participants that commented on this theme, 7 (64%) acknowledged that the language used in the app was clear and easy to understand. However, of the 11 participants, 7 (64%) also commented on what they disliked regarding the app's language choice, including the lessons not having descriptive titles. Some participants did not understand the meaning of the app's name (M-PAC app) and thought it did not represent the app's intent.

Theme 3: Backing Out or Undoing an Action

One (8%) participant commented on this theme and they mentioned that returning to the app's home screen was inconvenient.

Theme 4: Consistency in Language and Features

Of the 6 participants that commented on this theme, 6 (100%) discussed the app's consistency in language and features. All of them recognized something they disliked, including inconsistency in link descriptions (eg, some links had more details on why to click the link and some had no explanation), notification sounds appearing too frequently, and color labeling for the priority goals not following convention. By contrast, of the 6 participants, 2 (33%) liked that the app followed conventional color schemes (eg, active links were blue) and connected to conventional devices such as Fitbits.

Theme 5: Errors and Mistakes

Of the 11 participants that commented on this theme, 11 (100%) reported errors in the app. Some participants (4/11, 36%) stated that the app had problems connecting to Google Fit, and several links in the lessons did not work. Several users encountered general visual glitches, such as slow loading and pictures and data not being fully displayed on some devices. Other errors included the app being unresponsive, typos, problems with modifying weight and blood pressure information, step counts being inconsistent with other external trackers, and text not fitting on the screen.

Theme 6: Forgetting Acronyms

Of the 4 participants that commented on this theme, 3 (75%) did not understand what the app's name (ie, Pathverse or M-PAC app) represents, and 1 (25%) participant forgot that PAR-Q stands for the Physical Activity Readiness Questionnaire.

Theme 7: Efficiency of the Interface and Navigation

Of the 13 participants that commented on this theme, 9 (69%) thought the app was overall well structured and easy to navigate. However, of the 13 participants, 7 (54%) discussed aspects they disliked about the app's interface and navigation, such as the tracker feature being hard to find, not being able to go straight to the home screen after finishing a lesson, goals needing to be reentered daily, the tracking tab did not show dates for active minutes and daily goals, active minutes could not be entered for previous dates, and the progress of an unfinished lesson could not be saved.

Theme 8: Visual Design

Of the 13 participants that commented on this theme, 4 (31%) liked the app's overall visual design, with comments stating

XSL•FO
RenderX

that the app had visually pleasing graphics and engaging diagrams, and the color scheme helped user navigation (eg, blue active links turned purple when clicked). Of the 13 participants, 7 (54%) did not enjoy the app's visual design for reasons such as the color scheme was not esthetic and lacked novelty (2/7, 29%), the graphic content did not include diverse body types (2/7, 29%), and the name and icon of the app were not attractive or meaningful to the users (3/7, 43%).

Theme 9: Ease in Diagnosing and Resolving Errors

Two (15%) participants commented on this theme and reported that no information was provided for the potential reason for errors (ie, no error messages or error codes).

Theme 10: Help and Documentation

One (8%) participant mentioned that the app did not provide information or warnings about upcoming updates.

Theme 11: Additional Feedback

Of the 13 participants that commented on this theme, 11 (85%) had a positive overall experience using the app. Many would likely use the app to improve their PA and recommend it to others. Of the 13 participants, 1 (8%) had a negative overall experience with the app. Some users stated that they would not likely use the app to improve their PA (3/13, 23%) or recommend it to others (4/13, 31%).

Theme 12: Additional Suggestions on Features

Most participants (11/13, 85%) suggested the features they would like to be included in the app, such as the ability to access new lessons overtime, take notes and make personal input during lessons, add comments to daily active minutes, and receive incentives on users' PA progress.

Usability

A total of 13 participants filled out the MAUQ (Table 3). However, 1 participant did not fill out the MAUQ for unknown reasons. The mean total MAUQ score was 5.06 (SD 1.20), and the mean subscale scores were as follows: "ease of use" 5.36 (SD 1.27), "interface and satisfaction" 5.52 (SD 1.42), and "usefulness" 4.17 (SD 1.31).

Table 3. Mobile App Usability Questionnaire (MAUQ) scores of the pilot usability study (n=13).

Subscale	Score, mean (SD)
Ease of use	5.36 (1.27)
Interface and satisfaction	5.52 (1.42)
Usefulness	4.17 (1.31)
Total MAUQ score	5.06 (1.20)

Discussion

Overview

The purpose of this research was to gather user feedback on the first 2 prototypes of the M-PAC app to inform the design and development of an MVP. The ultimate aim of the M-PAC app is to promote sustained PA intention-behavior translation; consequently, it is informed by the M-PAC framework, which combines reflective and regulatory processes that assist individuals in translating PA intentions to behaviors with reflexive processes that contribute to PA maintenance [23,24,31]. We assessed the impressions of the first app prototype among a convenience sample of adults and subsequently evaluated the usability and acceptability of the second app prototype among a group of adults who intended to exercise but were not meeting the recommended PA guidelines. The M-PAC app design and development is informed by the IDEAS framework, which recommends an iterative phase of gathering user feedback and usability testing before feasibility and efficacy testing [38]. This step is important to ensure there is overall interest in the app, to identify potential impediments, and to collect suggestions for improvement and innovation.

Principal Findings

Overall, the focus group and interview findings indicated that the M-PAC app was well received by target users and the usability scores were high. The findings also unveiled several suggestions related to usability that should be addressed before testing an MVP.

Comparison With Prior Work

Participants thought the M-PAC app was easy to use and navigate, the layout was straightforward, and the lessons took a reasonable amount of time. Altogether, this feedback aligns with the "less-is-more" strategy, which has been proposed for effective human-technology interactions [63]. Enhancing the appeal and usability of the M-PAC app is crucial, especially considering that in 2022, the average retention rate for health and fitness apps dropped from 37% on the first day to just 9% by day 28 [13]. The most frequent reasons for abandoning mHealth apps are a lack of interest or declining motivation, preference for other apps, lack of desired features, lack of enjoyment (fun), and ease of use [64]. Consequently, the challenges with some of the features and navigation of the M-PAC app will need to be addressed, such as needing to reenter goals daily, inability to enter active minutes for previous dates, and inability to save lesson progress.

Several users reported difficulty with connecting to their PA tracking app. This is an important usability issue to address because users desire accurate and efficient tracking features [65]. Users also reported that the PA tracker feature was difficult to find and use (eg, activity minutes could not be entered for previous dates). Updating the PA tracking will need to be a priority because it facilitates self-monitoring, which is important for translating PA intentions into behaviors [10,11,66,67].

Another important characteristic to address is the app's name. Several participants critiqued the name "M-PAC app," stating it was not meaningful and did not represent the app's intent.

XSL•FO
RenderX

This is an interesting consideration that is specific to mHealth app behavioral interventions as it does not necessarily need to be addressed in face-to-face, educational leaflet, or website interventions. The app name is one of the first things that app users notice; therefore, it should be unique, easy to remember, and should clearly communicate the app's functionality [68]. We began brainstorming potential replacements to the app name that are more understandable to a diverse audience and represent its intent to promote long-term PA behavior, and ultimately, we decided on the name "Physical Activity for Life (PAL) app."

Participants voiced that they would like to continue accessing new lessons over time. This is supported by previous research stating that to maintain interest and encourage ongoing use of the app, users desire new and updated content [65]. Users desire tailored content, and to ensure they are receiving ongoing tailored content, a solution could be to implement the just-in-time adaptive intervention (JITAI) design. The JITAI design attempts to deliver support to the user "at the moment and in the context that the person needs it most and is most likely to be receptive" [69]. The M-PAC app development and design team is currently working on implementing a JITAI design.

The overall MAUQ scores from the usability pilot study were >5 for the total score, "ease of use," and "interface satisfaction" subscales. Given that MAUQ scores of ≥5 indicate increasing levels of agreeableness, these scores suggest high usability [59]. The mean "usefulness" subscale score was <5, indicating moderate to low usability. This is not surprising considering participants only had access to the first 3 lessons of the app, while the MVP is estimated to have 13 lessons. This likely also reflects the earlier feedback indicating that participants wanted more content.

Strengths and Limitations

Key strengths of the study are the sample demographics, particularly of the second study. Participants were aged, on average, 41.38 (SD 12.92) years, which is representative of millennials [70], who currently make up the most prevalent users of fitness apps [71]. Participants owned a range of different types of smartphone devices and were not currently meeting PA guidelines. A strength of the study design is that it used a mixed methods approach, including 2 types of qualitative feedback (eg, focus groups and interviews), as well as quantitative feedback in the form of the MAUQ. Mixed methods design offers unique and complementary insights into the multifaceted phenomena of mHealth usability [72]. Some limitations to the demographics are that the majority had a postsecondary degree and self-claimed to have good or excellent abilities to use the functions of their smartphone. Another limitation was that participants only had access to the first 3 lessons, out of the potential 13 lessons; however, the formatting will be the same for subsequent lessons, and the learnings from this study will be applied to those lessons as well. Because we are unaware of how the nature of regulatory and reflexive constructs may lend themselves to the current format, we plan to conduct further feasibility testing to more comprehensively understand whether the app is a viable means to deliver the complete intervention.

Conclusions

Overall, the first 2 M-PAC app prototypes were usable and acceptable among the users. We will address the suggestions in the design and development of the MVP. Key upgrades we plan to implement include testing connections with the PA tracking feature, shortening lessons, coming up with a new app name, optimizing image loading, providing the ability to log activity minutes from previous days, providing the ability to add notes to PA logs, and enabling users to save progress in lessons. Once we have completed these modifications, we will conduct feasibility testing with the MVP.

Acknowledgments

The authors would like to thank the participants for providing invaluable insights into the usability of the multi-process action control app. They would also like to thank their research assistant, Rhiaanah Villalobos, who played an integral role in planning the study. This research did not receive any funding.

Data Availability

The datasets generated and analyzed during this study are not available because the participants only consented to providing the data to the research team of this study.

Authors' Contributions

HH, WS, and RER conceived the study. HH and WS conducted the study and collected data. HH, WS, and HZ conducted data analysis. HH drafted the manuscript, which was edited by WS, HZ, and RER.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Focus group interview guide.

XSL•FO
RenderX

[DOCX File , 17 KB - formative_v9i1e59477_app1.docx]

Multimedia Appendix 2
Physical activity mobile health app informed by the multi-process action control (M-PAC) framework: lessons, mechanisms of action, and behavior change techniques.
[DOCX File , 24 KB - formative_v9i1e59477_app2.docx]

Multimedia Appendix 3
Pilot study exit interview guide.
[DOCX File , 16 KB - formative_v9i1e59477_app3.docx]

Multimedia Appendix 4
Physical activity mobile health app informed by the multi-process action control (M-PAC) framework: usability interview themes informed by the Nielsen's usability heuristics by Nielsen [54].
[DOCX File , 42 KB - formative_v9i1e59477_app4.docx]

References

1. Grandes G, Garcia-Alvarez A, Ansorena M, Ortega Sanchez-Pinilla R, Torcal J, Arietaleanizbeaskoa MS, et al. Any increment in physical activity reduces mortality of primary care inactive patients. Br J Gen Pract 2023;73(726):e52-e58. [doi: 10.3399/bjgp.2022.0118]
2. Erickson KI, Hillman C, Stillman CM, Ballard RM, Bloodgood B, Conroy DE, For 2018 Physical Activity Guidelines Advisory Committee*. Physical activity, cognition, and brain outcomes: a review of the 2018 physical activity guidelines. Med Sci Sports Exerc 2019 Jun;51(6):1242-1251 [FREE Full text] [doi: 10.1249/MSS.0000000000001936] [Medline: 31095081]
3. White RL, Babic MJ, Parker PD, Lubans DR, Astell-Burt T, Lonsdale C. Domain-specific physical activity and mental health: a meta-analysis. Am J Prev Med 2017 May 02;52(5):653-666. [doi: 10.1016/j.amepre.2016.12.008] [Medline: 28153647]
4. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Health 2018 Oct;6(10):e1077-e1086. [doi: 10.1016/s2214-109x(18)30357-7]
5. Abildso CG, Daily SM, Umstattd Meyer MR, Perry CK, Eyler A. Prevalence of meeting aerobic, muscle-strengthening, and combined physical activity guidelines during leisure time among adults, by rural-urban classification and region - United States, 2020. Am J Transplant 2023 Mar;23(3):443-446. [doi: 10.1016/j.ajt.2023.01.021] [Medline: 36740195]
6. Turner A. How many people have smartphones worldwide. BankMyCell. URL: https://www.bankmycell.com/blog/how-many-phones-are-in-the-world [accessed 2024-03-16]
7. Wurmser Y. The majority of Americans' mobile time spent takes place in apps. EMarketer. URL: https://www.insiderintelligence.com/content/the-majority-of-americans-mobile-time-spent-takes-place-in-apps [accessed 2024-03-16]
8. Ahmad I. 60+ fascinating smartphone apps usage statistics for 2019. Social Media Today. URL: https://www.socialmediatoday.com/news/60-fascinating-smartphone-apps-usage-statistics-for-2019-infographic/550990/ [accessed 2024-03-16]
9. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, IMAGE Study Group. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011 Feb 18;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
10. Knittle K, Nurmi J, Crutzen R, Hankonen N, Beattie M, Dombrowski SU. How can interventions increase motivation for physical activity? A systematic review and meta-analysis. Health Psychol Rev 2018 Sep 15;12(3):211-230. [doi: 10.1080/17437199.2018.1435299] [Medline: 29385950]
11. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
12. Dugas M, Gao GG, Agarwal R. Unpacking mHealth interventions: a systematic review of behavior change techniques used in randomized controlled trials assessing mHealth effectiveness. Digit Health 2020;6:205520762090541 1 [FREE Full text] [doi: 10.1177/2055207620905411] [Medline: 32128233]
13. Health and fitness app report 2024. Business of Apps. URL: https://www.businessofapps.com/data/health-fitness-app-report/ [accessed 2024-02-20]
14. Milne-Ives M, Homer SR, Andrade J, Meinert E. Potential associations between behavior change techniques and engagement with mobile health apps: a systematic review. Front Psychol 2023;14:1227443 [FREE Full text] [doi: 10.3389/fpsyg.2023.1227443] [Medline: 37794916]

XSL•FO
RenderX

15. Smith N, Liu S. A systematic review of the dose-response relationship between usage and outcomes of online physical activity weight-loss interventions. Internet Interv 2020 Dec;22:100344 [FREE Full text] [doi: 10.1016/j.invent.2020.100344] [Medline: 32995302]
16. Laranjo L, Ding D, Heleno B, Kocaballi B, Quiroz JC, Tong HL, et al. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. Br J Sports Med 2021 Apr 21;55(8):422-432. [doi: 10.1136/bjsports-2020-102892] [Medline: 33355160]
17. Conn VS, Hafdahl AR, Cooper PS, Brown LM, Lusk SL. Meta-analysis of workplace physical activity interventions. Am J Prev Med 2009 Oct;37(4):330-339 [FREE Full text] [doi: 10.1016/j.amepre.2009.06.008] [Medline: 19765506]
18. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012 Apr 30;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
19. Conn VS, Hafdahl AR, Mehr DR. Interventions to increase physical activity among healthy adults: meta-analysis of outcomes. Am J Public Health 2011 Apr;101(4):751-758 [FREE Full text] [doi: 10.2105/AJPH.2010.194381] [Medline: 21330590]
20. Wright CE, Rhodes RE, Ruggiero EW, Sheeran P. Benchmarking the effectiveness of interventions to promote physical activity: a metasynthesis. Health Psychol 2021 Nov;40(11):811-821. [doi: 10.1037/hea0001118] [Medline: 34914485]
21. Mönninghoff A, Kramer JN, Hess AJ, Ismailova K, Teepe GW, Tudor Car L, et al. Long-term effectiveness of mHealth physical activity interventions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2021 Apr 30;23(4):e26699 [FREE Full text] [doi: 10.2196/26699] [Medline: 33811021]
22. McEwan D, Rhodes RE, Beauchamp MR. What happens when the party is over?: sustaining physical activity behaviors after intervention cessation. Behav Med 2022 Apr 10;48(1):1-9. [doi: 10.1080/08964289.2020.1750335] [Medline: 32275199]
23. Dunton GF, Leventhal AM, Rebar AL, Gardner B, Intille SS, Rothman AJ. Towards consensus in conceptualizing and operationalizing physical activity maintenance. Psychol Sport Exerc 2022 Jul;61:102214. [doi: 10.1016/j.psychsport.2022.102214]
24. Rhodes RE, Sui W. Physical activity maintenance: a critical narrative review and directions for future research. Front Psychol 2021 Sep 6;12:725671 [FREE Full text] [doi: 10.3389/fpsyg.2021.725671] [Medline: 34552537]
25. van Stralen MM, De Vries H, Mudde AN, Bolman C, Lechner L. Determinants of initiation and maintenance of physical activity among older adults: a literature review. Health Psychology Review 2009 Sep;3(2):147-207. [doi: 10.1080/17437190903229462]
26. Cheval B, Boisgontier M. The theory of effort minimization in physical activity. Exerc Sport Sci Rev 2021 Jul 01;49(3):168-178 [FREE Full text] [doi: 10.1249/JES.0000000000000252] [Medline: 34112744]
27. Rhodes RE, Yao CA. Models accounting for intention-behavior discordance in the physical activity domain: a user's guide, content overview, and review of current evidence. Int J Behav Nutr Phys Act 2015 Feb 07;12(1):9 [FREE Full text] [doi: 10.1186/s12966-015-0168-6] [Medline: 25890238]
28. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. Appl Psychol 2008 Jan 30;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
29. Feil K, Fritsch J, Rhodes RE. The intention-behaviour gap in physical activity: a systematic review and meta-analysis of the action control framework. Br J Sports Med 2023 Oct 17;57(19):1265-1271. [doi: 10.1136/bjsports-2022-106640] [Medline: 37460164]
30. Schwarzer R. Health action process approach (HAPA) as a theoretical framework to understand behavior change. Act Psi 2016 Dec 05;30(121):119. [doi: 10.15517/ap.v30i121.23458]
31. Rhodes RE. Translating physical activity intentions into behavior: reflective, regulatory, and reflexive processes. Exerc Sport Sci Rev 2024 Jan 01;52(1):13-22. [doi: 10.1249/JES.0000000000000329] [Medline: 38126402]
32. Rhodes RE. The evolving understanding of physical activity behavior: a multi-process action control approach. In: Elliot AJ, editor. Advances in Motivation Science. Volume 4. New York, NY: Academic Press; 2017:171-205.
33. Hollman H, Sui W, Rhodes RE. A feasibility randomized controlled trial of a multi-process action control web-based intervention that targets physical activity in mothers. Women Health 2022 May 15;62(5):384-401. [doi: 10.1080/03630242.2022.2072050] [Medline: 35575171]
34. Husband CJ, Wharf-Higgins J, Rhodes RE. A feasibility randomized trial of an identity-based physical activity intervention among university students. Health Psychol Behav Med 2019 Apr 10;7(1):128-146 [FREE Full text] [doi: 10.1080/21642850.2019.1600407] [Medline: 34040843]
35. Kaushal N, Rhodes RE, Meldrum JT, Spence JC. Mediating mechanisms in a physical activity intervention: a test of habit formation. J Sport Exerc Psychol 2018 Apr 01;40(2):101-110. [doi: 10.1123/jsep.2017-0307] [Medline: 29945470]
36. Lim C, Wharf-Higgins J, Rhodes R. Working out with F.I.D.O. (frequency, intensity, duration, and outcomes): a feasibility randomized controlled trial. Hum Anim Interact Bull 2021 Dec;9(1):76-127 [FREE Full text] [doi: 10.1079/hai.2021.0025]
37. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. World Health Organization. 2016. URL: https://apps.who.int/iris/handle/10665/252183 [accessed 2022-02-25]

XSL•FO
RenderX

38. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (integrate, design, assess, and share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016 Dec 16;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
39. Slattery P, Saeri AK, Bragge P. Research co-design in health: a rapid overview of reviews. Health Res Policy Syst 2020 Feb 11;18(1):17. [doi: 10.1186/s12961-020-0528-9] [Medline: 32046728]
40. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
41. Ivanoff SD, Hultberg J. Understanding the multiple realities of everyday life: basic assumptions in focus-group methodology. Scand J Occup Ther 2006 Jun 12;13(2):125-132 [FREE Full text] [doi: 10.1080/11038120600691082] [Medline: 16856469]
42. Bond C, Lancaster GA, Campbell M, Chan C, Eddy S, Hopewell S, et al. Pilot and feasibility studies: extending the conceptual framework. Pilot Feasibility Stud 2023 Feb 09;9(1):24 [FREE Full text] [doi: 10.1186/s40814-023-01233-1] [Medline: 36759879]
43. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
44. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS consensus group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016 Oct 24;355:i5239 [FREE Full text] [doi: 10.1136/bmj.i5239] [Medline: 27777223]
45. Lancaster GA, Thabane L. Guidelines for reporting non-randomised pilot and feasibility studies. Pilot Feasibility Stud 2019 Oct 06;5(1):114 [FREE Full text] [doi: 10.1186/s40814-019-0499-1] [Medline: 31608150]
46. Bae YS, Kim KH, Choi SW, Ko T, Lim JS, Piao M. Satisfaction and usability of an information and communications technology-based system by clinically healthy patients with COVID-19 and medical professionals: cross-sectional survey and focus group interview study. JMIR Form Res 2021 Aug 26;5(8):e26227 [FREE Full text] [doi: 10.2196/26227] [Medline: 34254946]
47. Daniels N, Gillen P, Casson K, Wilson I. STEER: factors to consider when designing online focus groups using audiovisual technology in health research. Int J Qual Methods 2019 Nov 19;18:160940691988578. [doi: 10.1177/1609406919885786]
48. Milward J, Deluca P, Drummond C, Watson R, Dunne J, Kimergård A. Usability testing of the BRANCH smartphone app designed to reduce harmful drinking in young adults. JMIR Mhealth Uhealth 2017 Aug 08;5(8):e109 [FREE Full text] [doi: 10.2196/mhealth.7836] [Medline: 28790022]
49. Willemsen RF, Aardoom JJ, Chavannes NH, Versluis A. Online synchronous focus group interviews: Practical considerations. Qual Res 2022 Jun 30;23(6):1810-1820. [doi: 10.1177/14687941221110161]
50. Pathverse (Version 1). Pathverse Platform. 2019. URL: https://pathverse.ca/ [accessed 2020-12-21]
51. BCIO:007094. The Ontology Lookup Service (OLS). URL: https://www.ebi.ac.uk/ols4/ontologies/bcio/classes/http%253A%252F%252Fhumanbehaviourchange.org%252Fontology%252FBCIO_007094 [accessed 2024-06-25]
52. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
53. Marques MM, Wright AJ, Corker E, Johnston M, West R, Hastings J, et al. The behaviour change technique ontology: transforming the behaviour change technique taxonomy v1. Wellcome Open Res 2023;8:308 [FREE Full text] [doi: 10.12688/wellcomeopenres.19363.1] [Medline: 37593567]
54. Nielsen J. Usability Engineering. New York, NY: Morgan Kaufmann; 1994.
55. Open Broadcast Software (OBS). OBS Studio. URL: https://obsproject.com/ [accessed 2023-02-28]
56. NVivo. QSR International. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2024-04-29]
57. Hirschey J, Bane S, Mansour M, Sperber J, Agboola S, Kvedar J, et al. Evaluating the usability and usefulness of a mobile app for atrial fibrillation using qualitative methods: exploratory pilot study. JMIR Hum Factors 2018 Mar 15;5(1):e13 [FREE Full text] [doi: 10.2196/humanfactors.8004] [Medline: 29549073]
58. Kim H, Lee SH, Cho NB, You H, Choi T, Kim J. User-dependent usability and feasibility of a swallowing training mHealth app for older adults: mixed methods pilot study. JMIR Mhealth Uhealth 2020 Jul 27;8(7):e19585 [FREE Full text] [doi: 10.2196/19585] [Medline: 32663161]
59. Zhou L, Bao J, Setiawan IM, Saptono A, Parmanto B. The mHealth app usability questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
60. Godin G, Shephard RJ. Godin leisure-time exercise questionnaire. Med Sci Sports Exerc 1997;29(6):S36-S38. [doi: 10.1097/00005768-199706001-00009]

XSL•FO
RenderX

61. Amireault S, Godin G. The Godin-Shephard leisure-time physical activity questionnaire: validity evidence supporting its use for classifying healthy adults into active and insufficiently active categories. Percept Mot Skills 2015 Apr;120(2):604-622. [doi: 10.2466/03.27.PMS.120v19x7] [Medline: 25799030]
62. Namey E, Guest G, McKenna K, Chen M. Evaluating bang for the buck: a cost-effectiveness comparison between individual interviews and focus groups based on thematic saturation levels. Am J Eval 2016 Jul 09;37(3):425-440. [doi: 10.1177/1098214016630406]
63. Shneiderman B, Plaisant C, Cohen M, Jacobs S. Designing the User Interface: Strategies for Effective Human-Computer Interaction. London, UK: Pearson; 2013.
64. Mustafa AS, Ali N, Dhillon JS, Alkawsi G, Baashar Y. User engagement and abandonment of mHealth: a cross-sectional survey. Healthcare (Basel) 2022 Jan 24;10(2):221 [FREE Full text] [doi: 10.3390/healthcare10020221] [Medline: 35206837]
65. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
66. Rhodes RE, Plotnikoff RC, Courneya KS. Predicting the physical activity intention-behavior profiles of adopters and maintainers using three social cognition models. Ann Behav Med 2008 Dec 5;36(3):244-252. [doi: 10.1007/s12160-008-9071-6] [Medline: 19057976]
67. Rhodes RE, Plotnikoff RC. Understanding action control: predicting physical activity intention-behavior profiles across 6 months in a Canadian sample. Health Psychol 2006 May;25(3):292-299. [doi: 10.1037/0278-6133.25.3.292] [Medline: 16719600]
68. Sugrue E. App name guidelines and best practices. AppTweak. URL: https://www.apptweak.com/aso-blog/app-name-guidelines-and-best-practices [accessed 2024-03-25]
69. Spruijt-Metz D, Wen CK, O'Reilly G, Li M, Lee S, Emken BA, et al. Innovations in the use of interactive technology to support weight management. Curr Obes Rep 2015 Dec;4(4):510-519 [FREE Full text] [doi: 10.1007/s13679-015-0183-6] [Medline: 26364308]
70. Millenial. Merriam-Webster Dictionary. URL: https://www.merriam-webster.com/dictionary/millennial [accessed 2024-04-10]
71. 159 key fitness app stats for 2023: trends by age, market and more. Dr Muscle. URL: https://dr-muscle.com/fitness-app-stats/ [accessed 2024-02-20]
72. Alwashmi MF, Hawboldt J, Davis E, Fetters MD. The iterative convergent design for mobile health usability testing: mixed methods approach. JMIR Mhealth Uhealth 2019 Apr 26;7(4):e11656 [FREE Full text] [doi: 10.2196/11656] [Medline: 31025951]

Abbreviations

BCT: behavior change technique
CONSORT: Consolidated Standards of Reporting Trials
COREQ: Consolidated Criteria for Reporting Qualitative Research
FITT: frequency, intensity, time, and type
IDEAS: integrate, design, assess, and share
JITAI: just-in-time adaptive intervention
MAUQ: Mobile App Usability Questionnaire
mHealth: mobile health
M-PAC: multi-process action control
MVP: minimum viable product
MVPA: moderate-to-vigorous intensity physical activity
PA: physical activity
SMD: standardized mean difference

Edited by A Mavragani; submitted 15.04.24; peer-reviewed by N Admodisastro; comments to author 12.12.24; revised version received 11.01.25; accepted 28.01.25; published 23.04.25.

Please cite as:
Hollman H, Sui W, Zhang H, Rhodes RE
A Mobile Health App Informed by the Multi-Process Action Control Framework to Promote Physical Activity Among Inactive Adults: Iterative Usability Study
JMIR Form Res 2025;9:e59477
URL: https://formative.jmir.org/2025/1/e59477
doi:10.2196/59477
PMID:

XSL•FO
RenderX

©Heather Hollman, Wuyou Sui, Haowei Zhang, Ryan E Rhodes. Originally published in JMIR Formative Research (https://formative.jmir.org), 23.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Applying Patient and Health Professional Preferences in Co-Designing a Digital Brief Intervention to Reduce the Risk of Prescription Opioid–Related Harm Among Patients With Chronic Noncancer Pain: Qualitative Analysis

Rachel A Elphinston[1,2,3,4], PhD; Sue Pager[1,5], SpPath, MPH; Farhad Fatehi[6,7], MD, PhD; Michele Sterling[1,2], PhD; Kelly Brown[1,4,8,9], MClinPsych; Paul Gray[9,10], MBBS, PhD; Linda Hipper[4], GradDipPsych, MPH; Lauren Cahill[1]; Maisa Ziadni[11], PhD; Peter Worthy[12], PhD; Jason P Connor[9,13], PhD

[1]RECOVER Injury Research Centre, The University of Queensland, Herston, Australia

[2]National Health and Medical Research Council Centre for Research Excellence - Better Health Outcomes for Compensable Injury, The University of Queensland, Brisbane, Australia

[3]School of Psychology, The University of Queensland, Brisbane, Australia

[4]Addiction and Mental Health Services, Metro South Health Hospital and Health Service, Brisbane, Australia

[5]Health Equity & Access Unit, Metro South Hospital and Health Service, Brisbane, Australia

[6]Centre for Health Services Research, The University of Queensland, Brisbane, Australia

[7]School of Psychological Sciences, Monash University, Brisbane, Australia

[8]Gallipoli Medical Research Foundation, Greenslopes Private Hospital, Brisbane, Australia

[9]Faculty of Medicine, The University of Queensland, Brisbane, Australia

[10]Tess Cramond Pain and Research Centre, Metro North Hospital and Health Service, Brisbane, Australia

[11]School of Medicine, Department of Anesthesiology, Perioperative, and Pain Medicine, Stanford University, Stanford, CA, United States

[12]School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Australia

[13]National Centre for Youth Substance Use Research, The University of Queensland, Brisbane, Australia

Corresponding Author:
Rachel A Elphinston, PhD
RECOVER Injury Research Centre
The University of Queensland
Level 7
296 Herston Road
Herston, 4006
Australia
Phone: 61 412662084
Email: rachelel@uq.edu.au

Abstract

Background: Few personalized behavioral treatments are available to reduce the risk of prescription opioid–related harm among patients with chronic noncancer pain.

Objective: We aimed to report on the second phase of the co-design of a digital brief intervention (BI) based on patient and health professional preferences.

Methods: Eligible patients with chronic noncancer pain (n=18; 10 women; mean age 49.5, SD 6.91 y) from public hospital waitlists and health professionals (n=5; 2 women; mean age 40.2, SD 5.97 y) from pain and addiction clinics completed semistructured telephone interviews or participated in focus groups exploring BI preferences, needs, and considerations for implementation. Grounded theory was used to thematically analyze the data.

Results: We identified 5 themes related to intervention content from patient reports: relevance of the biopsychosocial model and need for improved awareness and pain psychology education; nonpharmacological strategies and flexibility when applying coping skills training; opioid use reflection and education, with personalized medication and tapering plans; holistic and patient-inclusive assessment measures and feedback; and inclusion of holistic goals targeting comfort and happiness. Five themes

XSL•FO
RenderX

related to the process and guiding principles were identified: therapist guided; engaging features; compassionate, responsive, person-centered care; a digital solution is exciting, maximizing reach; and educate and normalize system and policy challenges. Finally, 5 themes were reflected in the health professionals' reports: digital health use is rare but desired; digital health is useful for patient monitoring and accessing support; patient motivation is important; a digital BI app is likely beneficial and at multiple care points; and safe medication use and managing pain goals. The reported barriers from health professionals were intervention intensity, potential costs, and patient responsiveness; factors facilitating the implementation were the alignment of digital BIs with clinical models, a stepped-care approach, and feedback.

Conclusions: This co-design study identified key content areas, guiding principles, enabling factors, and barriers from both patients and health professionals to guide the development of digital BIs. The knowledge gathered should inform future iterations of co-designing digital BIs for the population most at risk of the harmful effects of opioid medications.

(JMIR Form Res 2025;9:e57212) doi:10.2196/57212

KEYWORDS

chronic noncancer pain; CNCP; prescription opioid use; brief intervention; brief psychological intervention; co-design; patient partners; qualitative research; digital health

Introduction

Clinical guidelines recommend behavioral and nonpharmacological treatments and interdisciplinary rehabilitation for the treatment of chronic noncancer pain (CNCP) (eg, Dowell et al [1]). However, there is a substantial gap between research evidence and clinical practice, referred to as one of the "valleys of death" [2] in the management of CNCP. Barriers to patients accessing best practice, multidisciplinary, specialist pain care include long waiting times, geography, high demand for treatment, treatment intensity, lack of trained behavioral therapists, and cost [3-7]. Psychological treatments rarely reach patients with CNCP. Our research on patients' lived experience has found that most patients with CNCP are not offered psychological treatment (eg, cognitive behavioral therapy) and are unaware that it is recommended as an effective first line treatment for pain [8]. These experiences are supported by national research, which shows that 87% of Australians with CNCP had not seen a psychologist for CNCP and almost 60% were not aware of the role of psychologists in pain care [9]. This is not unique to Australia. Patients around the world are not offered or referred to nonpharmacological interventions, particularly when receiving opioid therapy [4,10]. Increasingly high demand for psychologists and specialist pain clinicians [11,12] further complicates research translation.

Digital health interventions (DHIs) for pain have been highlighted as a potential innovative solution to the treatment accessibility problem, the growing burden of CNCP, and global health threats [13-15]. DHIs for pain provide potential advantages to health systems, providers, and patients such as improved accessibility, efficiency, and cost-effectiveness, but DHIs are affected by similar knowledge-to-practice gaps with added challenges in data privacy, security, and formulating regulatory guidelines and approvals. Patient access to DHIs outside of research studies is limited [16], reducing their real-world impact. Few DHIs that are publicly available have received empirical evaluation [16,17]. Bridging these gaps in traditional and digital pain care is a priority.

A lack of patient and health professional partnering in research has been cited as a central reason for research failing to translate into clinical practice [18-21]. Co-designing new DHIs has been highlighted as a potential solution to the translational challenge. Co-design involves meaningful stakeholder (eg, patients and health professionals) involvement in the design, implementation, and translation of research [22]. Engaging patient partners alongside health professionals in the co-design of DHIs allows for improved tailoring to individual preferences and needs, which has the potential to increase treatment effectiveness, acceptance, and adoption in practice.

Patients and health professionals have rarely been involved in the co-design of DHIs for CNCP since it was recommended more than a decade ago [23]. Most DHIs are developed by the for-profit software industry [24] and lack a theoretical framework [17]. Only 19% of the studies involved end users in development, and this is usually done in an ad hoc manner [25]. Involvement of health professionals in designing DHIs varies considerably (8.2% and 30.6%) [17,25]; few enable communication with health professionals or clinician access to patient data [25,26]; and almost none (<3%) involve both patients and health professionals [25].

There is an opportunity to optimize the research to practice nexus using co-design methods in all phases of developing new, brief DHIs. This study is part of a larger program of work to co-design, co-develop, and evaluate the feasibility of a digital brief intervention (BI) for patients with CNCP to reduce prescription opioid–related harm. Digital BIs provide a potentially efficient and scalable psychological and behavioral treatment option that facilitates rapid access to pain care [13,27]. In the first phase of the digital BI co-design, we examined the patients' lived experiences of CNCP management with a particular focus on opioid therapy in individual interviews [8]. In this second phase, we advanced this body of work by examining the needs and preferences of patients and health professionals as well as implementation barriers and enablers. The knowledge gathered will continue to inform future steps in the co-design of digital BIs for the population most at risk of the harmful effects of opioid medications.

XSL•FO
RenderX

Methods

Participants and Recruitment

Eligible participants were patients diagnosed with a CNCP condition who were purposively sampled to ensure adequate representation of both sexes, with culturally and linguistically diverse backgrounds, and current and past use of opioid therapy. A full summary of participant recruitment is provided in a previous study [8]. In short, the patient eligibility criteria included the following: those aged between 18 and 70 years, experiencing clinical levels of CNCP (scoring ≥4 out of 10 on the Pain Numerical Rating Scale on average over the past week), and seeking treatment from public health specialist addiction or pain services. The patient exclusion criteria were as follows: high levels of distress (based on patient scores ≥13 on the Kessler Psychological Distress Scale [28] and follow-up telephone risk assessment or clinical judgment from the addiction specialist), non-English speaking, a history of recent injecting drug use, or currently tested positive with SARS-CoV-2.

Patients were recruited through a tertiary hospital specialist pain clinic or community addiction specialist service. Patients were either engaged by mailing an information pack inviting those on the waitlist to participate (pain clinic) or invited by the medical specialist and a member of the research team (RAE) during their on-site appointment (addiction service). The final sample size was determined by thematic saturation using the analysis approach by Guest et al [29] applied during the data collection and inductive thematic analysis phase [8].

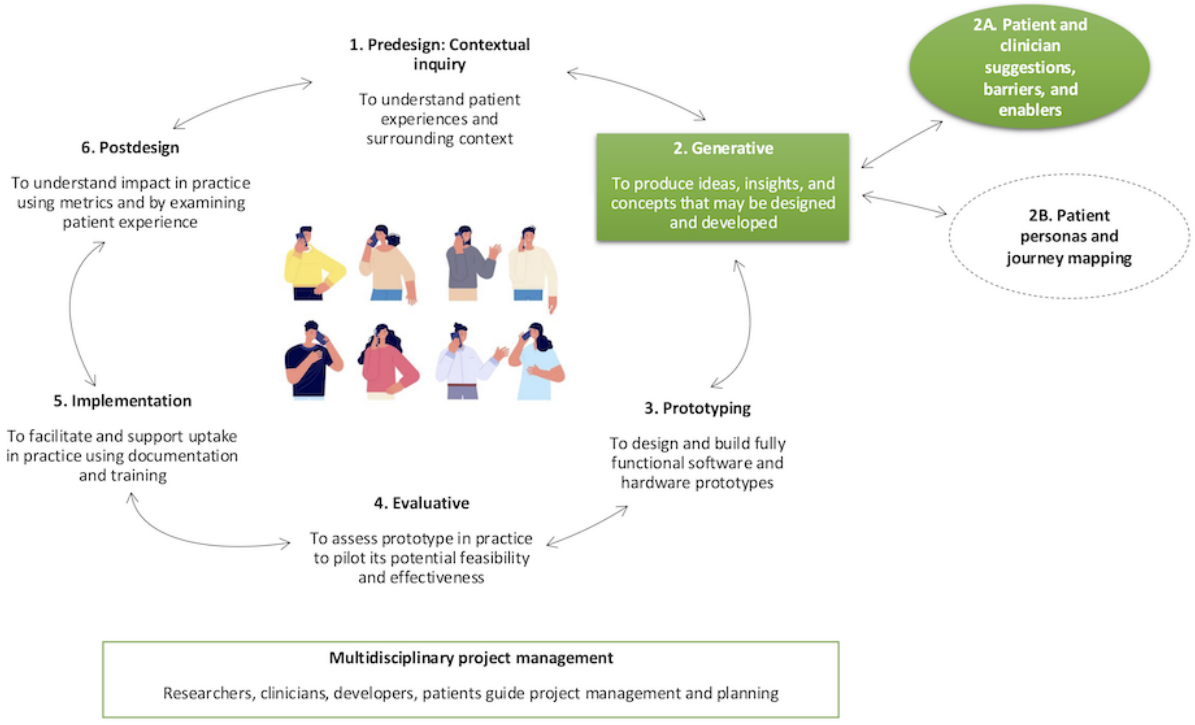
Health professionals were also invited to participate in one-on-one interviews or focus groups to assess their needs and preferences for the digital BI. Health professionals recruited through the specialist pain and addiction clinics as well as through the research and clinical networks of the research team were eligible to participate if they were working in chronic pain management (public or private sector).

Design

A co-design approach [9] was used, applying the framework by Sanders and Stappers [30] and updated by Noorbergen et al [31]. It outlines 6 iterative co-design phases: predesign, generative, prototyping, evaluative, implementation, and postdesign. Figure 1 presents an overview. This study reports on the generative phase of the iterative co-design development process, which includes exploring barriers, enablers, and suggestions for the digital BI from both the patient and health professional perspectives. Patient perspectives were collected from individual interviews (n=18), and 2 focus groups of patients (n=7) invited to attend after the individual interviews; health professional perspectives (n=5) were collected from two focus groups. All individual patient interviews were conducted via telephone by RAE, a clinical psychologist; KB, a psychologist; and SP, all with expertise in chronic pain or qualitative research. Field notes were recorded during the sessions.

The project was led by the study's chief investigator (RAE), who is a clinical psychologist–researcher. The multidisciplinary project team consisted of health care researchers (ie, psychologists, physiotherapists, pain medicine specialists, and addiction specialists); digital development partners (ie, software developers and designers); and patient partners (ie, individuals with a lived experience of CNCP including those using prescription opioid medications). The team met weekly (sometimes more often) during the development process.

Figure 1. Project co-design stages. The generative phase (phase 2) reported in this study is denoted by the items highlighted in green.



XSL•FO
RenderX

Measures

Web-Based Questionnaires

Patients were asked to provide their demographic details, information about their pain condition, current medications, and history of mental health and substance use disorders. Validated measures included the Brief Pain Inventory [32,33]; Kessler Psychological Distress Scale [28]; 21-item Depression Anxiety Stress Scale (DASS-21) [34]; and Current Opioid Misuse Measure (COMM) [35]. A full description of the measures can be found in the study by Elphinston et al [8].

Health professionals completed a web-based survey containing questions about their demographics, clinical background and experience, workplace setting, training and professional development in pain, use of DHIs, and types of telehealth use in clinical practice.

Patient Semistructured Interview and Focus Group Questions

As part of a larger set of interview questions, patients were asked about how they visualize their pain management in the future and their opioid use and how they could assist others in using opioids more safely. Multimedia Appendix 1 provides a summary of the questions. Patients who then expressed further interest in co-designing the BI were invited to participate in the focus groups, which involved discussing their impressions of the digital BI and several activities exploring its components. To set the stage, participants completed an exercise involving sharing their pain story, including the number and type of pain management strategies used. Participants then completed activities such as questionnaire assessment and feedback, introduction to the biopsychosocial model of pain, opioid use education, personality-targeted coping skills training, and motivational interviewing and action planning. Patients also explored the problem through a "peel the onion" activity and proposed solution requirements and were provided an example of a BI for smoking cessation. The aim of these activities was to explore and confirm patient preferences and needs, drawing on the results of the individual interviews to guide discussion. The role of technology in delivering BIs was explored including how to best meet the patient needs of this population and the implementation requirements. Finally, patients completed a value sliders exercise as a group, rating each BI component derived from research on BI treatment models and patient experience, from most important to least important.

Health Professional Semistructured Focus Group Questions

Health professionals were first introduced to the background and rationale for the digital BI, the main components, and the themes identified in patient lived experiences reported in a previous study [8]. They were then asked about their impressions of the themes and the digital BI. Multimedia Appendix 1 provides a summary of the questions.

Procedure

Interviews and focus groups were completed from June to December 2020. Patients completed an initial telephone screening interview to determine their eligibility for individual interviews and health professionals expressing interest via email were invited to 1 of 2 focus groups. Eligible patients and health professionals completed a web-based questionnaire before the interview for descriptive purposes and to provide background to guide the interview. Patient focus groups were then conducted in person on-site at the clinic with one patient participating via Zoom (Zoom Video Communications, Inc). Health professional focus groups were conducted via Zoom. Interviews and focus groups were audio-recorded and professionally transcribed verbatim.

Data Analysis

An iterative approach to data analysis was used. Transcripts from the patient interviews and focus groups were thematically analyzed separately from health professional focus group data using grounded theory [36]. Two coders (RAE and SP) coded line-by-line and then grouped the codes into subthemes and themes. Theme refinement was facilitated by a third clinician-researcher (KB), who assisted in integrating interviewer field notes to aid interpretation and reduce coder bias. All researchers were asked to review the themes and subthemes. The audio recordings were referred to as needed to examine patient quotes and clarify content.

Ethical Considerations

Ethics approval was obtained from the Metro South Human Research Ethics Committee (HREC/2020/QMS/60695). All participants provided written informed consent. Study data were deidentified for analysis, and any identifying information was removed before publication including patient numbers to ensure complete anonymity. All participants were offered an Aus $50 (US $33) gift voucher for each session attended to thank them for their time.

Results

Characteristics of the Patient and Health Professional Samples

Patient characteristics (n=18; 55% women; mean age 49.5, SD 6.91 y) are presented in a previous study [8]. We provide an overview in Table 1; for a detailed summary of each individual, please refer to a previous report [8]. In the sample of patients interviewed, half (9/18, 50%) met the threshold for current unsafe opioid misuse on a validated psychometric scale (COMM [35]). Most health professionals (n=5; 40% women; mean age 40.2, SD 5.97 y) were psychologists working in both the private and public sectors, with an average of 9.8 (SD 5.97) years of clinical experience (Table 2).

XSL•FO
RenderX

Table 1. Patient (n=18) demographics and clinical characteristics.

Characteristics	Values
Gender, n (%)	
Female	10 (55)
Male	8 (45)
Ethnicity, n (%)	
Australian	14 (77.9)
British (English, Welsh, or Irish)	2 (11.1)
New Zealander	1 (5.5)
Greek	1 (5.5)
Pain location, n (%)	
Head or face	4 (9.5)
Neck	6 (14.3)
Shoulder or upper limbs	6 (14.3)
Back, spine, or sacrum	12 (28.6)
Lower limbs	4 (9.5)
Whole body	5 (11.9)
Abdomen, pelvis, or groin	5 (11.9)
Age (y), mean (SD; minimum-maximum)	49 (11.5; 25-62)
Pain duration (y), mean (SD; minimum-maximum)	11.6 (11.5; 0.5-42)
Pain intensity (past week), mean (SD; minimum-maximum)	5.8 (1.5; 2.3-8.0)
Pain interference (past week), mean (SD; minimum-maximum)	5.8 (1.9; 2.7-9.1)
Kessler Psychological Distress Scale, mean (SD; minimum-maximum)	9.7 (4.8; 4-18)
Depression, anxiety, and stress[a], mean (SD; minimum-maximum)	
Depression	10.8 (9.1; 0-30)
Anxiety	9.2 (8.9; 0-34)
Stress	13.9 (9.1; 0-32)
Opioid use duration (y; n=16[b]), mean (SD; minimum-maximum)	8.0 (8.7; 0.6-26)
OME[c] (n=13[b,d]), mean (SD; minimum-maximum)	114.2 (166.8; 1.2-480)
COMM[e] (n=16[b,f]), mean (SD; minimum-maximum)	11.7 (7.89; 1-27)

[a]DASS-21: 21-item Depression Anxiety Stress Scales.

[b]2 participants were not currently using opioids.

[c]OME: oral morphine equivalent.

[d]3 participants currently using methadone.

[e]COMM: Current Opioid Misuse Measure.

[f]8 participants scored ≥9 indicating risk of opioid misuse [35].

XSL•FO
RenderX

Table 2. Health professional demographic and background information.

	Sex	Age (y)	Profession	Years of practice	Work sector
HP[a]1	Male	44	Psychologist	17	Public and private
HP2	Female	29	Psychologist	3	Public
HP3	Female	39	Psychologist	15	Public
HP4	Male	42	Nurse	8	Public
HP5	Male	47	Psychologist	6	Private

[a]HP: health professional.

Patient Preferences

Guiding content, principles, and processes related to the proposed digital BI emerged from the individual patient interviews and were further refined in the patient focus groups. There were 5 themes related to the content of the BI (themes 1-5) and 5 related to the process and guiding principles (themes 6-10). Table 3 provides a summary of themes and exemplar quotes. Patient numbers were removed before publication to ensure anonymity (PID_P refers to patients from the pain clinic and PID_A refers to patients from addiction services).

XSL•FO
RenderX

Table 3. Themes identified from the individual patient interviews and focus groups

Themes	Brief description	Key examples from patient interviews
Intervention Content		
1. The relevance of the biopsychosocial model and need for improved awareness and pain psychology education	The biopsychosocial model provides a relevant framework for understanding pain processes and improved awareness and pain psychology are needed.	• "It [biopsychosocial model] does make sense because I never connected it until I actually spoke to you on the phone. Because I had a good think about it and when there were family issues or when we moved house or anything really stressing my back pain would flare up. and it was almost like my pain levels, like my back was dealing with the stress. I had no control over that, I was just feeling what was going on and the pain levels go up... it's interesting, I didn't know that stress impacted and now when I think way back to when my back pain really kicked in it was a really stressing time." (PID_P). • "I can definitely relate to what you were talking about earlier that even when you've recovered your body still holds onto the pain because I feel that that's very much what's happening to me... long term pain puts stress on your nervous system and you can hold your pain because I've noticed lately that ever since I've had this when I get shoulder pain or something like that even just muscular pain it stays for a very long time because my body is in this constant state of alertness that it just holds pain. So, if I get a sore neck or something like that instead of it going in two days which it would 5 years ago, it now stays for weeks or even months because of my like nervous system" (PID_P).
2. Nonpharmacological treatments and flexibility when applying coping skills training are needed	Pain management strategies not just strategies to manage opioid use are required and should be flexibly applied.	• "I'd rather just not use the opioids at all and get help in other ways like ether it be therapy...I think Drs should be careful in prescribing them...and look at other things they could do for you" (PID_P). • "But if you're suffering obviously mentally, it kind of puts your body through a bit of stress, stressful situation that can cause like tension or psychologically it will inflict pain...So who knows, maybe a psychologist or whatever that was in pain, to help them mentally think about their pain in a different way, that would probably help...(PID_A).Yeah because you like to be told what to do and I like to choose what I do. But that's different personalities. I'm the boss of me, no body be telling me what to do, you know" (PID_P).
3. Opioid use reflection and education, with personalized medication and tapering plans	Opportunities to reflect on if opioids are working for them and providing the necessary education as well as medication plans that are individualized to the person's circumstances.	• "Explore potential problem in a non-judgemental way... I think finding out what the reason is and why it is important... But I think it's very important for the doctors, as the ones prescribing the medication to actually sit there and go, all right, here's the side effects" (PID_A). • "Refer to someone who can help taper in a personalized way... clear taper program with options to increase dose again if needed ... More regular monitoring by GP e.g. every month instead of 6 months" (PID_P). • "They want to come off them, if they think that they no longer have a physical use for them and might only be psychological they would need something to help them get through withdrawal symptoms" (PID_A). • "I don't feel the GP's support is enough...GP never asks me to keep a diary...they are not actually checking on if you have taking too many" (PID_P).
4. Holistic and patient-inclusive assessment measures and feedback	Assessment measures & feedback are holistic and patient inclusive.	• "I like how it [COMM] asks a lot of questions about how you feel and relationships like arguments... because I think pain effects so many things. It effects your relationships. it affects how you communicate with people. If you're in a lot of pain you can't think clearly. You are and you get angry at anything you get ticked off or annoyed or frustrated. So I think a lot of these comments if you know people are in pain they could relate to this very [sic] a lot of these questions...very relevant." (PID_P).

XSL•FO
RenderX

Themes	Brief description	Key examples from patient interviews
5. Inclusion of holistic goals targeting comfort and happiness	Broader, personal goals that target the whole person should be included, with a specific goal to increase happiness and level of comfort rather than a focus on being pain-free.	• "[Success would look like] A happy person... a well-adjusted person. A person who's not in constant pain. The list goes on... I would be happy with it making me comfortable" (PID_P). • "[Facilitator: not just medication focused goals?] Because I think it all comes to, its holistic, it's all encompassing, every part of your life needs to be looked at [– group agreed]" (PID_P).
Intervention Process and Guiding Principles		
6. Therapist-guided when needed and accountable	Guidance from health providers preferred rather than self-management approach to facilitate patient motivation and ensure accountability.	• "Could lose motivation very quickly if doing it yourself" (PID_P). • "Maybe having a drug counsellor or someone they can speak to on phone or online when they feel they are struggling with their medication" (PID_P). • "Psychologist to your house and the house visits for people that can't get out. That would be ideal, you know... Video for a psychologist would be good... I just get all tense to probably just when I'm in front of someone talking to, but that's just me" (PID_A).
7. Engaging easy to use features including peer support	A combination of engaging features (visual, verbal) including peer support to share similar experiences and ideas with and provide moral support when needed.	• "Even just hearing everyone's story, just feel not as alone" (PID_P). • "Sponsor like AA where if they want to take an extra tablet, they talk you out of it" (PID_P). • "Someone I could talk to, relate to, be there, just for the psychological side of it. People to talk to... Moral support was a big part of it. I reckon if I had someone to talk to a lot, when I was going through it wouldn't have been so bad. Just having friends that are going through the same thing" (PID_A).
8. Compassionate, responsive and person-centered care	Treat the person not the number, providing nonjudgmental and care personalized and responsive to the patient's needs.	• "I would have appreciated at the start not being told it was all in head and having just opioids thrown at me" (PID_P). • "Support, having someone see them as a person and not see them as a number" (PID_A). • "It's not a one size shoe fit all [approach to pain management]" (PID_A). • "With our other doctor, at least we could drop in and say, oh, my God, it's freaking hurting today" (PID_P).
9. Digital solution was new and exciting, maximizing reach	The new and exciting option of digital brief interventions maximizes reach in a variety of settings (eg, primary care; specialist wait lists).	• "If there was sort of like a program available and then that was part of it, so if I ever went to the doctor and said I want to reduce my tramadol slowly over time he could say "well, here's the website" and he'll register me for it and sort of he communicates with that website as well so when I go back to him he can see that I've done all the criteria and then he can reduce my medication" (PID_P). • "It would be really really good to be able to have something where you could be like, what can I do now? That would be really good because then you feel someone is on your side, someone is trying to help you plan. You're not just being left" (PID_A).
10. Educate and normalize challenges navigating the health system and policy	Education about policy changes and normalizing the challenges of navigating the health care system.	• "The GP put me on tramadol in 2007 and when the government brought that thing on the first of June, I had a new doctor and all of a sudden it's like oh no you can't be on this, you've got to be off, and I'd actually just switched to a new GP and he thought I was doctor shopping. He was quite judgmental" (PID_P).

Intervention Content

Theme 1: The Relevance of the Biopsychosocial Model and Need for Improved Awareness and Pain Psychology Education

Participants agreed that the biopsychosocial model of pain provides a relevant framework for explaining the unique individual experience of pain and how it can change over time. The patient experience interview was the first time when 3 of the 4 focus group participants were introduced to this model [8].

Most participants reported a number (>10 types) of varied biological and psychological influences of pain. It was important for participants that the origins of pain were explored (eg, surgery, injury, gradual or unknown, and stress). For another participant (PID_P), the continuation of pain despite recovery from physical injury resonated. One participant (PID_P) reported that they had moved toward acceptance that pain was mostly perpetuated by psychological factors including stress and tension

XSL•FO
RenderX

as they had eliminated all other reasons for their pain; they were now using mindfulness, breathing, and monitoring their own posture to relax and manage pain after having ceased the use of opioids in the past 8 weeks (a significant change from the individual interview [8]).

Participants also reported that the concept of a person's personality characteristics contributing to their pain and opioid use was considered important (eg, "Are you a 100-mile person or are you slow and steady, you know, it misses that completely" [PID_P]). Most participants reported that they identified with both depressive- and anxiety-prone personality characteristics as contributing to their pain and opioid medication use.

Social factors were considered less influential. One participant described how the COVID-19 pandemic had impacted their ability to go to church and that church often resulted in feeling better in terms of her overall well-being and pain—"I feel better when I come out of church then I did when I go in, and you know pain wise and all over wise" [PID_P]. Another discussed the role of caregiving responsibilities on her pain management experience—"[social factor] well I can't try medical marijuana even though a doctor offered it to me because I said to him I have to roll out of bed at 2 in the morning if her [elder mother who they care for] alarm goes off so I can't be taking that" [PID_P]. This facilitated patient understanding of how individualized social factors can influence pain experience.

Participants reported that greater awareness and educating patients about the biopsychosocial model of pain is needed. One participant (PID_P) described how it was helpful to understand how pain impacts the nervous system over time and how the body "holds" pain, especially because their injury had healed, and they now wanted to understand more about the influence of mental health. Another participant was open to discussing the role of mental health in pain with psychologists or psychiatrists because their injury had gone, and pain was persisting (PID_P).

Theme 2: Nonpharmacological Strategies and Flexibility When Applying Coping Skills Training

Participants confirmed a need for nonpharmacological strategies other than the use of opioids such as relaxation, exercise, engaging in pleasant activities, and pacing. PID_P said, "there's got to be something out there that covers everything but doesn't actually interfere with life itself, because you know the opioids make you stupid." This was reported by some participants as important to "take back control" (PID_P) from doctors who "should stop giving out opioids and recommend other non-pharmacological alternatives" (PID_P). One participant (PID_P) wished not to be opioid dependent, as they feel reliant on pain medication:

> *If [the BPI could] teach you not to be drug dependent. Because that's pretty much what I rely on. If that just stopped, I don't know what would happen.*

There were a range of coping strategies that participants identified could target depressive-proneness and pain (eg, listening to music, deep breathing, and doing enjoyable things). Coping strategies targeting anxiety-proneness included relaxation and releasing tension in the body (eg, gentle stretching or yoga) and mindfulness and meditation, while additional strategies that could target both personality styles were discussed (eg, going for a walk or talking to someone about how you feel). Multimedia Appendix 2 provides a snapshot of this exercise.

It was important that there was flexibility when applying a personality-targeted approach to coping skills training; some participants liked the targeted approach and wanted to be directed to the strategies corresponding to their personality style. Others preferred to be offered a full range of options and an opportunity to select the strategies themselves appealed to them:

> *Yeah because you like to be told what to do and I like to choose what I do. But that's different personalities. I'm the boss of me, no body be telling me what to do, you know.* [PID_P]

> *I don't know if the selection will work because if you're feeling anxious and stressed, being told to pick one of them is going to be overwhelming itself. You know if you're lonely and depressed, you might not necessarily want to choose you might just want someone to say right you're doing this. Like a decision might actually contribute to this. [all agreed with this]* [PID_P]

Theme 3: Opioid Use Reflection and Education, With Personalized Medication and Tapering Plans

Some participants reported that opportunities to reflect on their response to opioid therapy (benefits vs harms) would be important and could offer an avenue to provide opioid use education. Patients indicated that more knowledge of the potential harms of opioids was needed. One participant (PID_P) described how education that opioids can increase pain sensitivity and that tapering opioids do not necessarily result in increased pain was helpful when reducing opioid use:

> *I never attempted to come off it because I just thought what am I going to do with my pain, it's just going to get worse and he said no, as you reduce your pain sensitivity will reduce and he was right...[knowing that] it took the fear away from tapering off.*

Related to theme 7, the use of trivia and quizzes as a regularly updated semicompetitive activity was suggested to engage patients in education and motivation to continue with the digital BI.

Some participants suggested that education about opioid addiction and dependency would be important as well as taking into consideration people with pain who may have had a history of substance use problems. One participant said, "I'm clean. and even when I broke my shoulder and I had to take fentanyl in the hospital I phoned my sponsor first to make sure that it was okay for me to have it" [PID_P].

There was a specific need for opioid medication management plans with a focus on medication-related goals that are achievable:

> *Reduce your dose by I don't know 2mg, that's achievable... just little things you know make sure you go for a 10min walk once a week...10mins once a week should be achievable and not overdo it. You*

XSL•FO
RenderX

> *know just little things, it's like the compound effect book... you make such a small change that you don't really notice it but its compounded over 12 months... small achievable goals that compound.* [PID_P]

For some people, goals were related to a reduction in medication dose, while for others, using medication more safely. Opportunities to revise goals over time (eg, weekly) was identified as essential. Participants discussed a tailored, stepped, and "SMART" (Specific, Measurable, Achievable, Relevant, and Time-Bound) approach to goal setting. Monitoring of personalized medication plans was reported by some participants as necessary to checking patient response to treatment and provide an opportunity to identify any concerns (eg, withdrawal). One participant (PID_P) suggested a patient medication diary could assist in the monitoring process. Collaboration among health professionals and with patients was highlighted to ensure that everyone is aware of the patient's individual medication plan and is working together to support the patient' goals.

Theme 4: Holistic and Patient-Inclusive Assessment Measures and Feedback

Participants reported that assessment measures could focus on pain, its impact, medication use, and personality styles. It was important to all that measures are inclusive in terms of language and are nonjudgmental. In presenting the assessment measures to participants that they completed before the study, participants reported that they liked how items on some measures asked about how you feel (eg, capturing feelings of anger and frustration) and the impact of pain on relationships (eg, arguments assessed by the COMM).

Participants reported that they did not think the measures presented captured the pain journey (in terms of historical factors) as well as personality factors, although the research team communicated that this was probably a function of only providing participants with opioid-related measures. Focusing on opioid use in the past 30 days (as noted in the COMM) was viewed as restrictive:

> *This is all in the past 30 days, yet it says holistic there [on whiteboard], peoples pain experience, you know a holistic pain experience isn't just in the last 30 days it's in the last years...people have ups and downs, you might have had a really great 30 days. It doesn't really represent somebody's holistic pain experience.* [PID_P]

The items were reported to result in some participants feeing judged (eg, taking opioids illegally or outside of their prescription and counting pills) and that they were a "drug addict" [PID_P]. Participants discussed a sense of relating the pill counting questions more to anxiety than misuse:

> *[SOAPP-R] counting pills, only reason I would count them is because I'm anxious about not having enough, not because...its anxiety related not accumulative.* [PID_P]

It was suggested that the wording of some of the items could be modified to be less stigmatizing or measures could be introduced to participants and sensitivities flagged. Asking participants how they felt about questions about opioid use and giving them a free text box response option could also be beneficial (PID_P).

Theme 5: Inclusion of Holistic Goals Targeting Comfort and Happiness

Participants suggested that goals should target the whole person in addition to focusing on medication action plans. One participant (PID_P) discussed that they felt able to focus on improving physical fitness in the context of having an achievable plan already in place for reducing her tramadol use and managing withdrawal symptoms. It was also discussed that goals should empower patients, facilitating a sense of control and autonomy, consistent with a biopsychosocial mindset (theme 4; eg, success=confidence+mood boost=reduced pain):

> *Once you achieve something and you've got that little bit of success, that endorphin release also helps with your pain. [PID_P said "yeah yeah"] So, you've achieved something and that feel good feeling is yeah.* [PID_P]

Specifically, the goal to increase happiness and level of comfort rather than a focus on being pain-free was proposed as a possible goal of the intervention identified during the "peeling the onion" exercise (to gradually work toward deeper understanding of the problem to be solved). Participants suggested that focusing on making your life better and to increase happiness may assist in helping people to lead a "normal" lifestyle. Interestingly, participants did not focus on the goal of being pain-free but rather on being "comfortable." Most had moved toward acceptance of pain and felt that they may never be pain-free. Multimedia Appendix 3 provides a summary of the exercise.

Intervention Process and Guiding Principles

Theme 6: Therapist Guided When Needed and Accountable

Participants highlighted opportunities for connection and support from health professionals and peers were important. There was a need for "good advice" when called on and someone to talk to on bad pain days. Resources (eg, website links) on topics of interest were also viewed as helpful. A chat box feature was suggested as one way to connect with health professionals. Immediate responses from health professionals and options from where to seek further help were also seen as important (eg, drug and alcohol phone service). Participants agreed that the BI needs to be therapist guided rather than a self-management approach. It was suggested that a therapist-guided approach would provide both extra support and also accountability: "You'd want it to feel like there is that actual support" (PID_P); "I guess if it's not being monitored then you think what's the point?" (PID_A). Participants discussed value in health professional contact at the initial onset of the BI, in completing assessments and in the pain journey discussion.

Theme 7: Engaging, Easy-to-Use Features Including Peer Support

Discussions on how to best engage patients in the digital platform included the consideration of color schemes, the use

XSL•FO
RenderX

of games and competitions like "scoreboards," and the "steps challenges," which were viewed as motivating. Reading and sharing of patient stories would help normalize the patient's experience and provide ideas for what has worked for others. PID_P described the value of Facebook support groups. Hearing others' stories were highlighted as important: "Even just hearing everyone's story, just feel not as alone" [PID_P]. There were examples of chat boxes, website links, chat rooms, and forums to facilitate this, and it was suggested that this support may be very helpful, particularly on bad pain days. One participant suggested that access to journal articles that were translated into simple summaries for patients or videos of user-friendly, up-to-date evidence-based research summaries would be helpful. It was important that the BI is simple and easy to use.

Theme 8: Compassionate, Responsive, and Person-Centered Care

Participants reported a need for compassionate and personalized care from all treating health professionals. It was important that they were seen as a person and not as a number (PID_P) and that care is tailored to each unique individual with pain. Comments about "being told pain was all in my head" [PID_P] were viewed by participants as invalidating and unhelpful to treatment discussions. Empathy was agreed as a necessary ingredient to care, as was the responsivity of health professionals. Some participants suggested that greater access to health professionals, particularly on bad pain days, via digital means could better meet their support needs (eg, "GP contactable on Facetime, he does tele[health], emails" [PID_A]).

Theme 9: The Option of Digital BIs Was New and Exciting, Maximizing Reach in a Variety of Settings

None of the participants had come across a digital BI for CNCP in their search for effective treatment solutions. They said they were excited for a new option that could potentially help themselves and other patients. Participants discussed hope that their experiences would inform the design and development of the BI that could offer more support and earlier intervention to patients. A digital platform was seen as an ideal way to deliver the intervention for maximum reach and access in a way tailored to the patient's needs. One patient had experience using apps (eg, Smiling Mind), which was reported to give them more confidence in the potential use and acceptability of a digital BI.

Participants suggested that disseminating or promoting the BI would be suited to a primary care setting, with the use of flyers in the waiting room or to those on specialist waitlists. Engaging potential participants in the intervention could also include the use of social media, Google advertising, and BI advocates and word of mouth. Since it was discussed that everyone is generally using their phone on a daily basis, there was agreement that web-based promotion may be more effective than hardcopy leaflets. Health professional and particularly general practitioner education and awareness of the BI program was seen as key.

Theme 10: Educate and Normalize Challenges Navigating the Health System and Policy

Participants discussed difficulties navigating the health system and indicated value in providing education on the rationale for policies and legal restrictions that may impact their opioid use and pain management. It was suggested that greater consideration of the social environment and medicolegal background may serve the function of normalizing challenging experiences (eg, changes in opioid use), acknowledging differences in the treatment they received in the past (20 years ago) compared to current best practices, and reducing perceived judgment from the care team (eg, methadone associated with feeling judged as "a drug addict").

Patient Ranking of Importance of Intervention Components

Patients ranked the importance of 7 digital BI components from most important (rating=1) to least important (rating=7):

1. Support
2. Coping strategies
3. Education
4. Assessment
5. Feedback and monitoring
6. Goals and action plans
7. Fun

Health Professional Perspectives

Five themes were identified from the data as important considerations by health professionals to the development of the digital BI and its implementation in clinical practice. Barriers and enablers were also identified.

Theme 1: Use of DHIs and Resources Within Clinical Practice Was Rare But Desired

Health professionals described that they generally do not tend to integrate apps and digital interventions into their clinical practice. One health professional said, "it would be rare that I'd use a program in the clinic with someone I'm actually treating." Instead, they were generally viewed as an adjunct to treatment. For example, health professionals would recommend web-based programs for pain or more general mental health apps. Health professionals also recommended patients' access TED Talks, podcasts, and educational videos, as well as resources they had developed themselves (eg, mindfulness audio files and sleep hygiene information). Most only recommended resources that they have evaluated themselves. Some health professionals highlighted that integration of digital health tools into their practice was practically challenging: "I've never really tried to integrate an app into treatment because it's always been a bit too challenging in terms of working out how to get the feedback and what to do with it then."

There was recognition among health professionals that it is important to point patients to digital resources that are free and easily accessible: "They are not all free, or only free for a while and then a subscription is required." It was also acknowledged that some patients may not have access to the internet and that digital resources were generally not inclusive of culturally and linguistically diverse patient populations. Some health professionals reported that they would like to use digital health resources more in their clinical practice: "Depending on the demographic as in CALD [culturally and linguistical diversity], not really many resources I can use but I don't use it as much as I would like to."

XSL•FO
RenderX

Theme 2: Digital Health Useful for Patient Monitoring of Symptoms and Access to Support on the Run

Health professionals reported that apps can be useful for patients to monitor their symptoms over time as well as track their use of other treatments (eg, cannabis). There were reports that digital health tools can increase access to support such as coping strategies and resources. One health professional said, "Can be useful to access it remotely such as if they are in a shopping center and feel anxious they can sit down and do some calming exercises." While benefits were mentioned, some health professionals reported that their clients preferred to minimize the use of digital health: "Many of my clients just prefer to keep it simple, won't even use video [during telehealth], just want a phone call consultation."

Theme 3: Patient Motivation to Use Digital Health Resources Is Important

Patient motivation to use digital health was considered important in whether these were successful in clinical practice. One health professional said, "I find that that apps are only as good as the client's motivation to use them," and another said, "Sometimes we recommend things and they don't get acted on...take a lot of prompting to use themselves." There was acknowledgment that patient motivation can be influenced by several factors (eg, how easy it is to download an app) and that some patients may not be motivated to change behavior, try something new, or use digital health resources.

Theme 4: A Digital BI App for This Complex Problem May Be Beneficial at Multiple Points in the Health Care Journey

Overall, health professionals agreed that a digital BI for the complex problems of CNCP and risk of opioid-related harm may be beneficial for patients. Health professionals reported that the digital BI could be helpful for patients to complete before referral to pain specialist services:

> *I see it as helpful before come to me [specialist pain clinic] because it will shift their readiness for change and make more open to when they come to me. Would use this as a pre-intervention.*

There were also reports that the digital BI may be "Useful to bridge gap between primary care and pain services, so not waiting for so long." One health professional said the following:

> *Good for those on the waiting list for specialist services... it might mean that by the time they get to that service they might be more ready, knowledgeable, more open to alternatives... help to set up expectations of treatment approaches.*

Others reported that it could fit within a primary care setting or even be used for the prevention of opioid use problems:

> *I think it could be useful across the board. In my clinic though, patient selection is important if you want to get best results. Could help to move people who are at the pre-contemplation stage.*

The patient's stage of change or motivation was reported as important in considering the best intervention point (linked to theme 3). There was also suggestion that the digital intervention could be delivered by psychologists working in the private sector; mental health professionals; multidisciplinary pain clinics; or, in the primary care setting, by nurses or physicians: "the GP could refer patients to or deliver the app before they prescribe."

Health professionals reported that patients would prefer to access the intervention via a mobile app or device rather than a desktop website due to convenience and portability. There was also recognition that a digital intervention would have challenges for some patients, and there was a need for therapist-guiding:

> *Both (app and website) have access issues. Would be good to have an option to do the program somewhere; in the GP office, at the MDT [multidisciplinary team] clinic while they're waiting on the wait list etc. Either way patients would need health professional support for access and troubleshooting as well as for the monitoring and goal setting.*

Theme 5: Realistic Patient Goals Focused on the Safe Use of Medications and Managing Pain

Patients' goals that are focused on safe use of opioid medications including increased willingness or openness to change unsafe use were supported. Health professionals also reported that goals related to use of alternative pain management strategies and pain acceptance may also be achievable: "Can we get to them to move a little bit on changing behavior - willingness to change opioid use and use alternative strategies." Setting realistic goals were important:

> *I'd work on positive wins towards the overall goal. So much going on for patients, other things to manage.*

Barriers to Implementation

Some health professionals reported that more than 4 hours of intervention may be required to achieve the patient's goals. Costs related to who would deliver the intervention were identified as a barrier as well as clinicians setting aside time for support (eg, viewing patient monitoring). There was also concern that some patients may be defensive about making changes to their opioid use and that some patients do not trust health professionals to talk about their opioid use.

Intervention Enablers

Health professionals viewed that the underlying philosophy of the digital BI (motivational interviewing and enhancing patient self-efficacy) was consistent with how they approach their clinical work. There was recognition that implementation of the digital BI would be most effective if considered within a stepped-care model (ie, primary care > digital BI > specialist care; related to theme 4). There were also reports that feedback to patients (about their opioid use, pain, and mood) is "always helpful," as is patients' response to this information. This information can then be used to troubleshoot or problem solve.

XSL•FO
RenderX

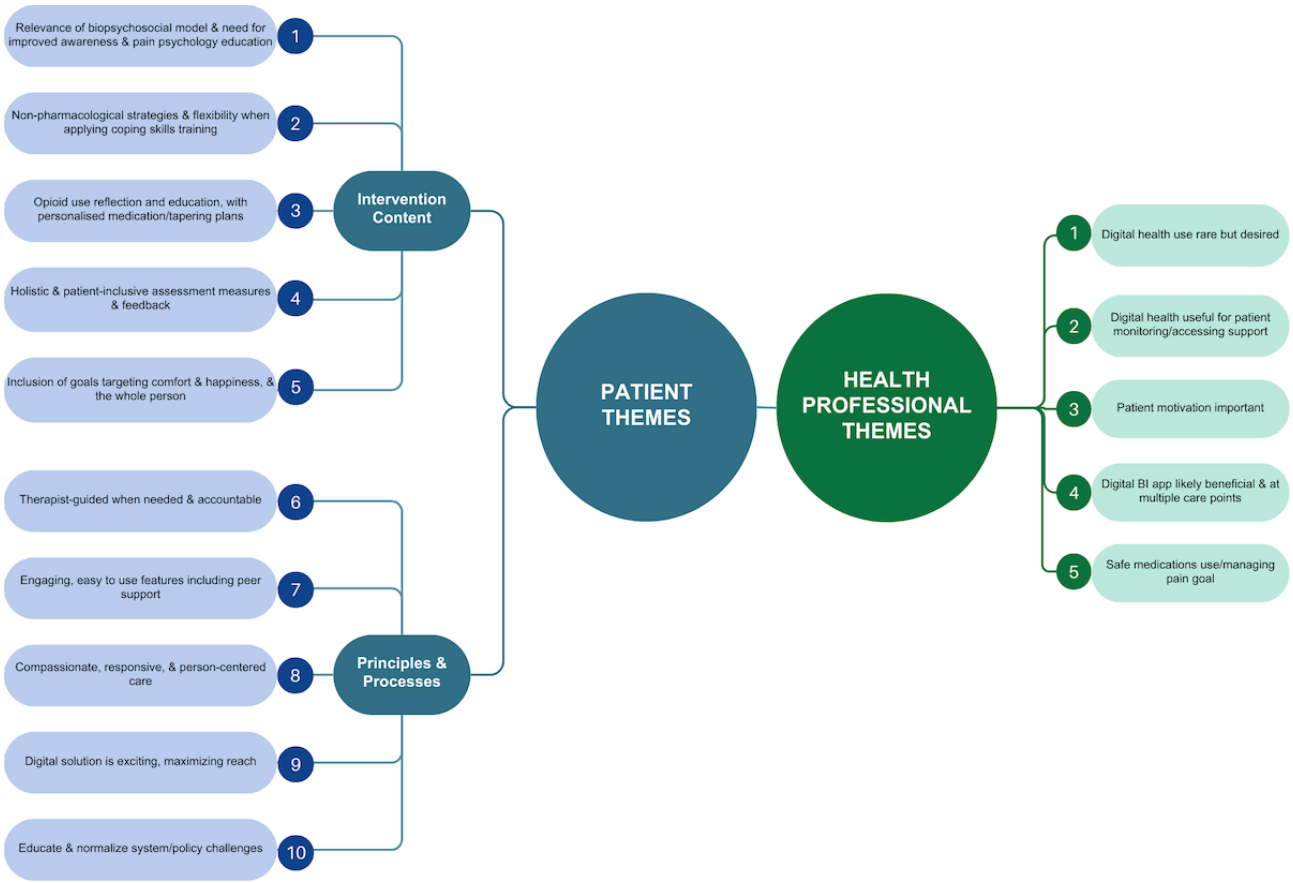
Discussion

Main Findings

This study aimed to understand patient and health professional preferences, including implementation considerations, for a new digital BI to reduce prescription opioid–related harm among patients with CNCP. In total, 10 themes, 5 related to intervention content and 5 related to intervention processes emerged from patient reports as important in the future digital BI co-design and co-development. From reports of experienced health professionals, we identified 5 themes, as well as several barriers (including intervention intensity, potential costs, and patient responsiveness) and implementation enablers (including consistency of theoretical underpinnings with current practice, patient feedback, and the digital BI as part of a stepped-care model). Figure 2 provides a snapshot of the themes identified from both patients and health professionals.

Overall, both patients living with CNCP and health professionals were interested in the proposed digital BI to reduce opioid-related harm. Health professionals highlighted that digital health tools may be particularly useful for patient symptom monitoring and access to coping skills support (eg, deep breathing exercise) when on the move. Patients agreed that a digital solution could expand the reach of BIs to people who need it. Certainly, there is growing interest in and evidence for the role of digital BIs in treating pain [13,37-39].

Figure 2. Snapshot overview of themes.



In line with clinical guidelines that recommend multidisciplinary and behavioral treatment approaches as first line (eg, Dowell et al [40] and Busse et al [1]), patients reported that the inclusion of nonpharmacological pain management strategies (eg, relaxation, exercise, pleasant and valued activities, pacing, access to psychology, and multidisciplinary team approaches) would be welcomed, rating this as one of the most important features of the digital BI. Patients also indicated that personalizing coping skills training to patient personality characteristics (eg, impulsivity traits) and pain-related factors (eg, depression) could benefit some patients, while others may prefer to direct themselves to coping skills that resonate with them. Individual choice could be built into the digital BI to recognize patient differences in autonomy and self-efficacy. Notably, nonpharmacological approaches can achieve similar or even greater improvements in pain and functioning without the potential harms of dependence, adverse side effects, and overdose. This is especially important as patients generally report that opioid therapy has modest effects on levels of pain and functioning [41-43].

There were educational needs identified by patients in the areas of pain psychology, the biopsychosocial model of pain, and opioid therapy. These findings are in line with the high levels of interest and increased need for patient education about the role of psychology in the management of pain [4,44]. Educating patients about the biopsychosocial model of pain forms a part of multidisciplinary rehabilitation and discipline-specific treatments that can be effective in reducing pain and disability in the short to medium term (eg, Siddall et al [45] and Wood and Hendrick [46]). Helping patients to reconceptualize pain can also enhance their ability to cope with their condition [47].

XSL•FO
RenderX

The importance of safe opioid use education is highlighted in clinical guidelines [1,48]. Few studies have examined the effectiveness of opioid use education for patients with chronic pain. On the basis of the evidence available, opioid-related education can lead to safer behaviors (eg, less stockpiling [49]) and reduced opioid use following surgery [50]. Gaps in educating patients in pain management, influential biopsychosocial factors, and safe use of opioid medications continues to put patients at risk of poor recovery. Ongoing weighing up of the benefits of opioid therapy in the treatment of CNCP versus risks could be facilitated by a health professional–delivered digital BI. This could be embedded within the review of the patients' personalized medication plan and be supported by regular monitoring and assessment of medication use as well as related symptoms that consider the *whole person* (eg, pain, mood, and functioning). Health professionals considered digital health as an optimal platform to facilitate patient monitoring of symptoms. Patients also reported that one of the top 3 activities that they want to do on their mobile device is manage their medication and track their health [51]. Digital BIs that aim to reduce risk of opioid-related harm have the potential to meet this need.

Patients tended to focus on the potential of the digital BI to meet general well-being goals and reduce pain discomfort to facilitate safer use of opioid medications. Health professionals indicated that patients would be likely interested in pain reduction goals but suggested that goals related to the safe use of opioid medications could also be important. If goals are to encompass the needs of the whole person, then it will be important to explore and balance patients' pain-related goals and opioid medication harm reduction goals as part of holistic goals related to relationships, work, and lifestyle. To facilitate behavioral change, different innovative approaches (eg, gamification and serious games) could be explored in future studies.

Health professionals reported a desire to integrate DHIs into their clinical practice in a more conscious way. Real-world implementation of DHIs have often failed, mostly because they are not used by patients or health professionals [52]. Engaging health professionals in the co-design process has potential to enhance digital BI feasibility and acceptability in practice, as the success of such measures in health care settings relies heavily on the engagement of key stakeholders [22]. Our results could inform the development of a broader digital BI implementation strategy that considers barriers identified by health professionals (eg, costs, time, and compatibility between the digital BI and their current workflow) and enablers (eg, perceived fit of the digital BI approach to current practice), which are common across DHIs and health care settings [53]. Further consideration of the systems used (eg, mobile app vs website and health professional assistance) and adaptability, policy and guideline support, resourcing, health professional knowledge and beliefs, and involvement of digital BI *champions* to support patient motivation [53] may inform effective digital BI implementation. Most psychologists and mental health professionals (72%) have little or no formal training in pain management and more than 50% report feeling a lack of confidence in treating pain [4]. Training and education for health professionals in delivering digital BIs could fill this gap.

Similar to other digital BI studies [54], a therapist-guided digital BI was preferred among patients and health professionals. According to patients, therapist guidance would enhance patient motivation and engagement and provide more accountability than self-guided digital BIs. Evidence indicates that guided digital cognitive behavioral therapy and face-to-face therapy for psychiatric and somatic conditions produce equivalent overall effects [55], and when therapist support is included, this has a positive impact on engagement and treatment effectiveness [56]. Therapist-guided digital BIs provide opportunities for establishing mutual trust, which can be a barrier to intervention success [57]. This may be critical as health professionals report that discussions about chronic pain management, including medication use, can be challenging, with patients reporting that the context of opioid policy changes can further fuel mistrust. Guided digital BIs allow for therapists to foster a strong therapeutic alliance with empathy and compassion, treating patients as partners in their pain care - providing the foundation that is necessary for treatment success. Further consideration of communication methods between patients, their treating clinicians, and the broader team within the digital environment may further personalize pain services and facilitate multidisciplinary collaboration. When asked to rank the importance of the BI features, patients reported *support* was the most important feature, while *fun* was the least. Future co-design opportunities could explore meaningful ways to integrate gamification and digital peer support to facilitate patient engagement.

Both patients and health professionals reported that digital BIs could be best placed and adopted in primary care, multidisciplinary pain centers while patients wait for treatment, and addiction services where it could be an adjunct to in-person treatment; it was also suggested that it could be tailored to prevent risk of opioid-related harm in people with acute pain. There was also a suggestion that digital BIs could form part of a stepped-care model. Further research is needed to determine whether digital BIs for treatment of risk of opioid-related harm are effective and for whom and in what setting.

Strengths and Limitations

Most previous studies have not included the views of key stakeholders in the design of DHIs [17,25]. We advanced the co-design of a digital BI to reduce risk of opioid-related harm among people with CNCP by examining both patient and health professional perspectives on their intervention needs and preferences. Our study adds to the growing literature on co-design in DHIs [58] and extends research into the CNCP field.

We included data from 18 individual patient interviews and from 7 patients and 5 health professionals as part of focus group workshops. While there are no guidelines on the minimum number of key stakeholders or frequency and intensity of engagement that is needed in co-design, it is possible that our results may not be generalizable to the broader population of people with CNCP who are taking prescription opioids, or beyond the Australian context where patients have access to a relatively high standard of public health care and access to heavily subsidized medications. Continual engagement of

XSL•FO
RenderX

various patients and health professional perspectives in the subsequent steps of co-design and co-development of digital BIs could increase customizability of the solution to fit a wider range of users. In future design and development work, consideration of how to design BIs for better inclusion of patient cultural and linguistic diversity, spiritual needs, and socioeconomic factors is needed, as well as investigations in countries with different health and insurance systems. It will be important that the digital divide does not exacerbate existing health inequities of those most at risk in the CNCP population. Efforts to measure digital inclusion could increase translational confidence in digital BIs.

In addition, there remains a power differential between patients, health professionals, and researchers, and it is unknown how this may have influenced our findings. Future work should explore jointly developed guiding ethical principles to address the relationship power imbalances in co-design. There are also potential tensions between these stakeholders in their digital BI needs and preferences which could be explored in future joint focus groups. Finally, this study was conducted before the widespread introduction of generative AI technologies (eg, ChatGPT, Gemini, and Co-Pilot) to the public. Future studies could explore the integration of generative AI into DHIs, such as those that aim to reduce risk of opioid-related harm and improve pain care.

Conclusions

Partnering with patients who have CNCP and health professionals to identify their intervention needs and preferences is critical to the co-design of new, innovative, and high-value digital treatments [59]. Patients and health professionals were interested in the proposed digital BI, identifying many content areas, principles, enablers, and barriers to guide the development of such programs. This co-design approach has the potential to enhance the translation of digital BIs into practice and improve clinical effectiveness. The next phases in co-designing personalized digital BIs involve the generation of later design concepts (eg, journey mapping and persona development) and prototyping, while considering the development of an implementation strategy. Continual, genuine bringing together of scientific, patient, clinical, and technical expertise to solution design could address the urgent need for better patient-centered CNCP care while minimizing the risks of opioid medication–related harms.

Acknowledgments

The research team would like to sincerely thank the patients for participating in this project and sharing their experiences. The authors would also like to thank the specialist pain and addiction services for collaborating on this project. We extend special thanks to Dr Andrew Claus, Dr Geraldine Chew, and Dr Tabibul Khan for their assistance in patient engagement. The authors would also like to thank Ms Chelsea Geary for her contribution to the development of the interview questions and web-based surveys.

Metro South Research Support Scheme Early Career Researcher Grant (CIA RAE; RSS_2020_035) provided funding for this project. The RECOVER Injury Research Centre, The University of Queensland; Addiction & Mental Health Services, Metro South Health; and the Tess Cramond Pain and Research Centre, Metro North Health provided in-kind support. RAE is supported by a Research Stimulus Fellowship grant administered by the University of Queensland.

Data Availability

The datasets generated during and analyzed during this study are not publicly available due to limitations of ethics approval but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient and health professional semistructured interview and focus group questions.
[DOCX File , 14 KB - formative_v9i1e57212_app1.docx]

Multimedia Appendix 2
Personality-driven coping strategies.
[PNG File , 618 KB - formative_v9i1e57212_app2.png]

Multimedia Appendix 3
Goals identified during the "peeling the onion" exercise.
[PNG File , 535 KB - formative_v9i1e57212_app3.png]

References

XSL•FO
RenderX

1. Dowell D, Haegerich TM, Chou R. CDC guideline for prescribing opioids for chronic pain--United States, 2016. JAMA 2016 Apr 19;315(15):1624-1645 [FREE Full text] [doi: 10.1001/jama.2016.1464] [Medline: 26977696]
2. Reis SE, McDonald MC, Byers SJ. Crossing the research valleys of death: the University of Pittsburgh approach. Clin Transl Sci 2008 May 21;1(1):9-10. [doi: 10.1111/j.1752-8062.2008.00021.x] [Medline: 19412329]
3. Crouch TB, Wedin S, Kilpatrick RL, Christon L, Balliet W, Borckardt J, et al. Pain rehabilitation's dual power: treatment for chronic pain and prevention of opioid-related risks. Am Psychol 2020 Sep;75(6):825-839 [FREE Full text] [doi: 10.1037/amp0000663] [Medline: 32915026]
4. Darnall BD, Carr DB, Schatman ME. Pain psychology and the biopsychosocial model of pain treatment: ethical imperatives and social responsibility. Pain Med 2017 Aug 01;18(8):1413-1415 [FREE Full text] [doi: 10.1093/pm/pnw166] [Medline: 27425187]
5. Eccleston C, Morley SJ, Williams AC. Psychological approaches to chronic pain management: evidence and challenges. Br J Anaesth 2013 Jul;111(1):59-63 [FREE Full text] [doi: 10.1093/bja/aet207] [Medline: 23794646]
6. Hogg MN, Gibson S, Helou A, DeGabriele J, Farrell MJ. Waiting in pain: a systematic investigation into the provision of persistent pain services in Australia. Med J Aust 2012 Apr 02;196(6):386-390. [doi: 10.5694/mja12.10140] [Medline: 22471539]
7. Nicholas MK, Asghari A, Sharpe L, Beeston L, Brooker C, Glare P, et al. Reducing the use of opioids by patients with chronic pain: an effectiveness study with long-term follow-up. Pain 2020 Mar 19;161(3):509-519. [doi: 10.1097/j.pain.0000000000001763] [Medline: 31764391]
8. Elphinston RA, Pager S, Brown K, Sterling M, Fatehi F, Gray P, et al. Co-designing a digital brief intervention to reduce the risk of prescription opioid–related harm among people with chronic noncancer pain: qualitative analysis of patient lived experiences. JMIR Form Res (forthcoming) 2024. [doi: 10.2196/57208]
9. The Australian pain and psychology report 2020. Australian Psychological Society. 2020. URL: https://psychology.org.au/getmedia/0189c0e2-1832-422d-af4a-031391ea143c/20aps-pw-report-p1.pdf [accessed 2024-04-29]
10. Bernstein C, Gillman AG, Zhang D, Bartman AE, Jeong JH, Wasan AD. Identifying predictors of recommendations for and participation in multimodal nonpharmacological treatments for chronic pain using patient-reported outcomes and electronic medical records. Pain Med 2020 Dec 25;21(12):3574-3584. [doi: 10.1093/pm/pnaa203] [Medline: 32869082]
11. Psychological expertise where we need it most: GPs not the only answer to our mental health crisis. Australian Psychological Society. 2022. URL: https://psychology.org.au/about-us/news-and-media/media-releases/2022/psychological-expertise-where-we-need-it-most-gps [accessed 2024-04-29]
12. Darnall BD, Scheman J, Davin S, Burns JW, Murphy JL, Wilson AC, et al. Pain psychology: a global needs assessment and national call to action. Pain Med 2016 Feb 23;17(2):250-263 [FREE Full text] [doi: 10.1093/pm/pnv095] [Medline: 26803844]
13. Darnall BD. Psychological treatment for chronic pain: improving access and integration. Psychol Sci Public Interest 2021 Sep 20;22(2):45-51 [FREE Full text] [doi: 10.1177/15291006211033612] [Medline: 34541966]
14. Eccleston C, Blyth FM, Dear BF, Fisher EA, Keefe FJ, Lynch ME, et al. Managing patients with chronic pain during the COVID-19 outbreak: considerations for the rapid introduction of remotely supported (eHealth) pain management services. Pain 2020 May 2;161(5):889-893 [FREE Full text] [doi: 10.1097/j.pain.0000000000001885] [Medline: 32251203]
15. Slater H, Dear BF, Merolli MA, Li LC, Briggs AM. Use of eHealth technologies to enable the implementation of Musculoskeletal models of care: evidence and practice. Best Pract Res Clin Rheumatol 2016 Jun;30(3):483-502. [doi: 10.1016/j.berh.2016.08.006] [Medline: 27886943]
16. de la Vega R, Miró J. mHealth: a strategic field without a solid scientific soul: a systematic review of pain-related apps. PLoS One 2014 Jul 7;9(7):e101312 [FREE Full text] [doi: 10.1371/journal.pone.0101312] [Medline: 24999983]
17. Lalloo C, Jibb LA, Rivera J, Agarwal A, Stinson JN. "There's a pain app for that": review of patient-targeted smartphone applications for pain management. Clin J Pain 2015 Jun;31(6):557-563. [doi: 10.1097/AJP.0000000000000171] [Medline: 25370138]
18. Fernandes LG, Devan H, Fioratti I, Kamper SJ, Williams CM, Saragiotto BT. At my own pace, space, and place: a systematic review of qualitative studies of enablers and barriers to telehealth interventions for people with chronic pain. Pain 2022 Feb 01;163(2):e165-e181. [doi: 10.1097/j.pain.0000000000002364] [Medline: 34433776]
19. Ioannidis JP. Why most clinical research is not useful. PLoS Med 2016 Jun 21;13(6):e1002049 [FREE Full text] [doi: 10.1371/journal.pmed.1002049] [Medline: 27328301]
20. Schlieter H, Marsch LA, Whitehouse D, Otto L, Londral AR, Teepe GW, et al. Scale-up of digital innovations in health care: expert commentary on enablers and barriers. J Med Internet Res 2022 Mar 11;24(3):e24582. [doi: 10.2196/24582] [Medline: 35275065]
21. Svendsen MJ, Wood KW, Kyle J, Cooper K, Rasmussen CD, Sandal LF, et al. Barriers and facilitators to patient uptake and utilisation of digital interventions for the self-management of low back pain: a systematic review of qualitative studies. BMJ Open 2020 Dec 12;10(12):e038800. [doi: 10.1136/bmjopen-2020-038800] [Medline: 33310794]
22. Slattery P, Saeri AK, Bragge P. Research co-design in health: a rapid overview of reviews. Health Res Policy Syst 2020 Feb 11;18(1):17. [doi: 10.1186/s12961-020-0528-9] [Medline: 32046728]

XSL•FO
RenderX

23. Keogh E, Rosser BA, Eccleston C. e-Health and chronic pain management: current status and developments. Pain 2010 Oct;151(1):18-21. [doi: 10.1016/j.pain.2010.07.014] [Medline: 20674174]
24. Tabi K, Randhawa AS, Choi F, Mithani Z, Albers F, Schnieder M, et al. Mobile apps for medication management: review and analysis. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e13608 [FREE Full text] [doi: 10.2196/13608] [Medline: 31512580]
25. Zhao P, Yoo I, Lancey R, Varghese E. Mobile applications for pain management: an app analysis for clinical usage. BMC Med Inform Decis Mak 2019 May 30;19(1):106. [doi: 10.1186/s12911-019-0827-7] [Medline: 31146739]
26. Devan H, Farmery D, Peebles L, Grainger R. Evaluation of self-management support functions in apps for people with persistent pain: systematic review. JMIR Mhealth Uhealth 2019 Feb 12;7(2):e13080 [FREE Full text] [doi: 10.2196/13080] [Medline: 30747715]
27. Elphinston RA, Sterling M, Leung J, Gray PD, Scotti Requena S, Connor JP. Brief psychological interventions for reducing prescription opioid use, related harm, and pain intensity in patients with chronic pain: a systematic review and assessment of patient interest. Clin J Pain 2021 Apr 01;37(4):270-280. [doi: 10.1097/AJP.0000000000000908] [Medline: 33323791]
28. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SL, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002 Aug;32(6):959-976. [doi: 10.1017/s0033291702006074] [Medline: 12214795]
29. Guest G, Namey E, Chen M. A simple method to assess and report thematic saturation in qualitative research. PLoS One 2020 May 5;15(5):e0232076 [FREE Full text] [doi: 10.1371/journal.pone.0232076] [Medline: 32369511]
30. Sanders EB, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008 Mar;4(1):5-18. [doi: 10.1080/15710880701875068]
31. Noorbergen TJ, Adam MT, Teubner T, Collins CE. Using co-design in mobile health system development: a qualitative study with experts in co-design and mobile health system development. JMIR Mhealth Uhealth 2021 Nov 10;9(11):e27896 [FREE Full text] [doi: 10.2196/27896] [Medline: 34757323]
32. Kumar SP. Utilization of brief pain inventory as an assessment tool for pain in patients with cancer: a focused review. Indian J Palliat Care 2011 May;17(2):108-115 [FREE Full text] [doi: 10.4103/0973-1075.84531] [Medline: 21976850]
33. Tan G, Jensen MP, Thornby JI, Shanti BF. Validation of the brief pain inventory for chronic nonmalignant pain. J Pain 2004 Mar;5(2):133-137. [doi: 10.1016/j.jpain.2003.12.005] [Medline: 15042521]
34. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
35. Butler SF, Budman SH, Fernandez KC, Houle B, Benoit C, Katz N, et al. Development and validation of the current opioid misuse measure. Pain 2007 Jul;130(1-2):144-156 [FREE Full text] [doi: 10.1016/j.pain.2007.01.014] [Medline: 17493754]
36. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
37. Darnall BD, Burns JW, Hong J, Roy A, Slater K, Poupore-King H, et al. Empowered Relief, cognitive behavioral therapy, and health education for people with chronic pain: a comparison of outcomes at 6-month Follow-up for a randomized controlled trial. Pain Rep 2024 Jan;9(1):e1116 [FREE Full text] [doi: 10.1097/PR9.0000000000001116] [Medline: 38288134]
38. Darnall BD, Ziadni MS, Krishnamurthy P, Flood P, Heathcote LC, Mackey IG, et al. "My surgical success": effect of a digital behavioral pain medicine intervention on time to opioid cessation after breast cancer surgery-a pilot randomized controlled clinical trial. Pain Med 2019 Nov 01;20(11):2228-2237 [FREE Full text] [doi: 10.1093/pm/pnz094] [Medline: 31087093]
39. Ziadni MS, Anderson SR, Gonzalez-Castro L, Darnall BD. Comparative efficacy of a single-session “Empowered Relief” videoconference-delivered group intervention for chronic pain: study protocol for a randomized controlled trial. Trials 2021 May 22;22(1):358. [doi: 10.1186/S13063-021-05303-8]
40. Busse JW, Craigie S, Juurlink DN, Buckley DN, Wang L, Couban RJ, et al. Guideline for opioid therapy and chronic noncancer pain. CMAJ 2017 May 08;189(18):E659-E666 [FREE Full text] [doi: 10.1503/cmaj.170363] [Medline: 28483845]
41. Abdel Shaheed C, Maher CG, Williams KA, Day R, McLachlan AJ. Efficacy, tolerability, and dose-dependent effects of opioid analgesics for low back pain: a systematic review and meta-analysis. JAMA Intern Med 2016 Jul 01;176(7):958-968. [doi: 10.1001/jamainternmed.2016.1251] [Medline: 27213267]
42. Busse JW, Wang L, Kamaleldin M, Craigie S, Riva JJ, Montoya L, et al. Opioids for chronic noncancer pain: a systematic review and meta-analysis. JAMA 2018 Dec 18;320(23):2448-2460 [FREE Full text] [doi: 10.1001/jama.2018.18472] [Medline: 30561481]
43. Reinecke H, Weber C, Lange K, Simon M, Stein C, Sorgatz H. Analgesic efficacy of opioids in chronic pain: recent meta-analyses. Br J Pharmacol 2015 Jan;172(2):324-333. [doi: 10.1111/bph.12634] [Medline: 24640991]
44. Darnall BD. Minimize opioids by optimizing pain psychology. Pain Manag 2014 Jul 10;4(4):251-253 [FREE Full text] [doi: 10.2217/pmt.14.18] [Medline: 25300380]
45. Siddall B, Ram A, Jones MD, Booth J, Perriman D, Summers SJ. Short-term impact of combining pain neuroscience education with exercise for chronic musculoskeletal pain: a systematic review and meta-analysis. Pain 2022 Jan 01;163(1):e20-e30. [doi: 10.1097/j.pain.0000000000002308] [Medline: 33863860]

XSL•FO
RenderX

46. Wood L, Hendrick PA. A systematic review and meta-analysis of pain neuroscience education for chronic low back pain: short-and long-term outcomes of pain and disability. Eur J Pain 2019 Feb 14;23(2):234-249. [doi: 10.1002/ejp.1314] [Medline: 30178503]
47. Watson JA, Ryan CG, Cooper L, Ellington D, Whittle R, Lavender M, et al. Pain neuroscience education for adults with chronic musculoskeletal pain: a mixed-methods systematic review and meta-analysis. J Pain 2019 Oct;20(10):1140.e1-1140.22. [doi: 10.1016/j.jpain.2019.02.011] [Medline: 30831273]
48. Häuser W, Morlion B, Vowles KE, Bannister K, Buchser E, Casale R, et al. European* clinical practice recommendations on opioids for chronic noncancer pain - part 1: role of opioids in the management of chronic noncancer pain. Eur J Pain 2021 May 02;25(5):949-968 [FREE Full text] [doi: 10.1002/ejp.1736] [Medline: 33655607]
49. Hero JO, McMurtry C, Benson J, Blendon R. Discussing opioid risks with patients to reduce misuse and abuse: evidence from 2 surveys. Ann Fam Med 2016 Nov 14;14(6):575-577 [FREE Full text] [doi: 10.1370/afm.1994] [Medline: 28376444]
50. Zorrilla-Vaca A, Mena G, Ramirez P, Lee B, Sideris A, Wu C. Effectiveness of perioperative opioid educational initiatives: a systematic review and meta-analysis. Anesth Analg 2022 May 01;134(5):940-951. [doi: 10.1213/ANE.0000000000005634] [Medline: 34125081]
51. National digital health strategy. Australian Digital Health Agency. URL: https://www.digitalhealth.gov.au/national-digital-health-strategy [accessed 2024-04-29]
52. Mohr DC, Riper H, Schueller SM. A solution-focused research approach to achieve an implementable revolution in digital mental health. JAMA Psychiatry 2018 Feb 01;75(2):113-114. [doi: 10.1001/jamapsychiatry.2017.3838] [Medline: 29238805]
53. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146. [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
54. Davin SA, Savage J, Thompson NR, Schuster A, Darnall BD. Transforming standard of care for spine surgery: integration of an online single-session behavioral pain management class for perioperative optimization. Front Pain Res (Lausanne) 2022 May 2;3:856252 [FREE Full text] [doi: 10.3389/fpain.2022.856252] [Medline: 35599968]
55. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan 07;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
56. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
57. Bathia NS, McAskill RE, Hancox JE, Knaggs RD. Collaboration between adult patients and practitioners when making decisions about prescribing opioid medicines for chronic non-cancer pain in primary care: a scoping review. Br J Pain 2022 Feb 23;16(1):119-126 [FREE Full text] [doi: 10.1177/20494637211025560] [Medline: 35111320]
58. Vandekerckhove P, de Mul M, Bramer WM, de Bont AA. Generative participatory design methodology to develop electronic health interventions: systematic literature review. J Med Internet Res 2020 Apr 27;22(4):e13780 [FREE Full text] [doi: 10.2196/13780] [Medline: 32338617]
59. Eather CE, Sterling M, Sullivan C, Elphinston RA. Leveraging value-based health principles to improve translation and impact of digital psychological interventions for people with chronic pain. Pain (forthcoming) 2024 Oct 15. [doi: 10.1097/j.pain.0000000000003441] [Medline: 39451013]

Abbreviations

BI: brief intervention
CNCP: chronic noncancer pain
COMM: Current Opioid Misuse Measure
DHI: digital health intervention
SMART: Specific, Measurable, Achievable, Relevant, and Time-Bound

Edited by A Mavragani; submitted 23.02.24; peer-reviewed by G Lewis; comments to author 01.05.24; revised version received 12.08.24; accepted 25.09.24; published 25.04.25.

Please cite as:
Elphinston RA, Pager S, Fatehi F, Sterling M, Brown K, Gray P, Hipper L, Cahill L, Ziadni M, Worthy P, Connor JP
Applying Patient and Health Professional Preferences in Co-Designing a Digital Brief Intervention to Reduce the Risk of Prescription Opioid–Related Harm Among Patients With Chronic Noncancer Pain: Qualitative Analysis
JMIR Form Res 2025;9:e57212
URL: https://formative.jmir.org/2025/1/e57212
doi:10.2196/57212
PMID:

XSL•FO
RenderX

©Rachel A Elphinston, Sue Pager, Farhad Fatehi, Michele Sterling, Kelly Brown, Paul Gray, Linda Hipper, Lauren Cahill, Maisa Ziadni, Peter Worthy, Jason P Connor. Originally published in JMIR Formative Research (https://formative.jmir.org), 25.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Acceptability and Usability of a Digital Behavioral Health Platform for Youth at Risk of Suicide: User-Centered Design Study With Patients, Practitioners, and Business Gatekeepers

Trinity Chloe Tse[1,2], BA; Lauren S Weiner[3], PhD; Carter J Funkhouser[1,2], PhD; Danielle DeLuise[3], BA; Colleen Cullen[1], PsyD; Zachary Blumkin[1], PsyD; Casey O'Brien[1], PsyD; Randy P Auerbach[1,2], PhD; Nicholas B Allen[3,4], PhD

[1]Division of Child and Adolescent Psychiatry, Columbia University Irving Medical Center, Columbia University, New York, NY, United States

[2]New York State Psychiatric Institute, Manhattan, NY, United States

[3]Ksana Health Inc, Eugene, OR, United States

[4]Department of Psychology, University of Oregon, Eugene, OR, United States

Corresponding Author:
Nicholas B Allen, PhD
Department of Psychology
University of Oregon
Department of Psychology
Univ of Oregon
Eugene, OR, 97403
United States
Phone: 1 5413464075
Email: nallen3@uoregon.edu

Abstract

Background: Youth suicide rates are climbing, underscoring the need to improve clinical care. Personal smartphones can provide an understanding of proximal risk factors associated with suicide and facilitate consistent contact between patients and practitioners to improve treatment engagement and effectiveness. The Vira digital behavior change platform (Vira) consists of a patient smartphone app and a web-based practitioner portal (Vira Pro) that integrates objective mobile sensing data with Health Insurance Portability and Accountability Act (HIPAA)–compliant communication tools. Through Vira, practitioners can continuously assess patients' real-world behavior and provide clinical tools to enhance treatment via just-in-time behavior change support.

Objective: This study aimed to explore the acceptability and usability of the minimal viable product version of Vira through a user-centered design (UCD) approach and to identify barriers to implementing Vira in the context of an adolescent intensive outpatient program.

Methods: Over 2 iterative phases, feedback was gathered from adolescent patients (n=16), mental health practitioners (n=11), and business gatekeepers (n=5). The mixed methods UCD approach included individual semistructured interviews (eg, perspectives on treatment and attitudes toward digital tools), surveys (eg, usability), and unmoderated user testing sessions (eg, user experience).

Results: Overall, participants expressed optimism regarding Vira, particularly among adolescents, who showed high satisfaction with the app's interface and design. However, clinicians reported more mixed views, agreeing that it would be useful in treatment but also expressing concerns about the volume and displays of patient data in Vira Pro, workload management, and boundaries. Gatekeepers identified usability issues and implementation barriers related to electronic health records but also recognized Vira's potential to enhance treatment outcomes. Feedback from stakeholders informed several crucial changes to the platform, including adjustments to data-sharing protocols, user interface enhancements, and modifications to training methods.

Conclusions: Vira has a high potential to improve patient engagement and improve clinical outcomes among high-risk youth. Iterative UCD and ongoing stakeholder engagement are essential for developing technology-based interventions that effectively meet the needs of diverse end users and align with clinical workflows.

(JMIR Form Res 2025;9:e65418) doi:10.2196/65418

XSL•FO
RenderX

KEYWORDS

user-centered design; mobile sensing; behavior change; suicide prevention; digital health; mental health

Introduction

Background

Suicide is the second leading cause of death among adolescents, and from 2011 to 2021, the rate of suicide attempts and suicidal behaviors has been increasing [1-3]. Existing mental health treatments are only moderately effective [4], and even among those who receive a "gold-standard" treatment, many do not improve. To reduce the public health burden of adolescent mental health problems, there is a critical need for novel approaches that increase the effectiveness of existing treatments.

One promising way to improve mental health treatments is to leverage digital devices such as personal smartphones. Smartphones have the potential to both monitor risk in real time, and to deliver just-in-time interventions, both of which are difficult if not impossible to achieve without these ubiquitous personal computing devices. To date, few digital interventions have been developed with or for youth at risk of suicide and rigorously tested to provide evidence of feasibility, acceptability, or efficacy [5,6]. Examples of other mental health apps targeted toward mitigating suicide risk include BRITE, LifeBuoy, and BlueIce. BRITE was developed to support suicide safety planning among adolescents consists of a practitioner portal and patient app, and has primarily been implemented as a single-session intervention [7]. By contrast, LifeBuoy is an app that helps young people manage suicidal thoughts in their daily lives by teaching dialectical behavior therapy (DBT) skills via self-guided modules [8], whereas BlueIce (MyOxygen, Limited) is used to support distress tolerance and reduce self-harm through a personalized toolbox of strategies that include mood diaries, a menu of mood-lifting activities, and automatic routing to emergency numbers if needed [9]. Although initial clinical outcomes are promising [10], these apps do not appear to support continuous treatment between patients and their practitioners, a critical aspect of evidence-based treatments such as DBT [11], nor do they include mobile sensing, which could provide unique insights into risk processes and opportunities for intervention that occur outside the health care context.

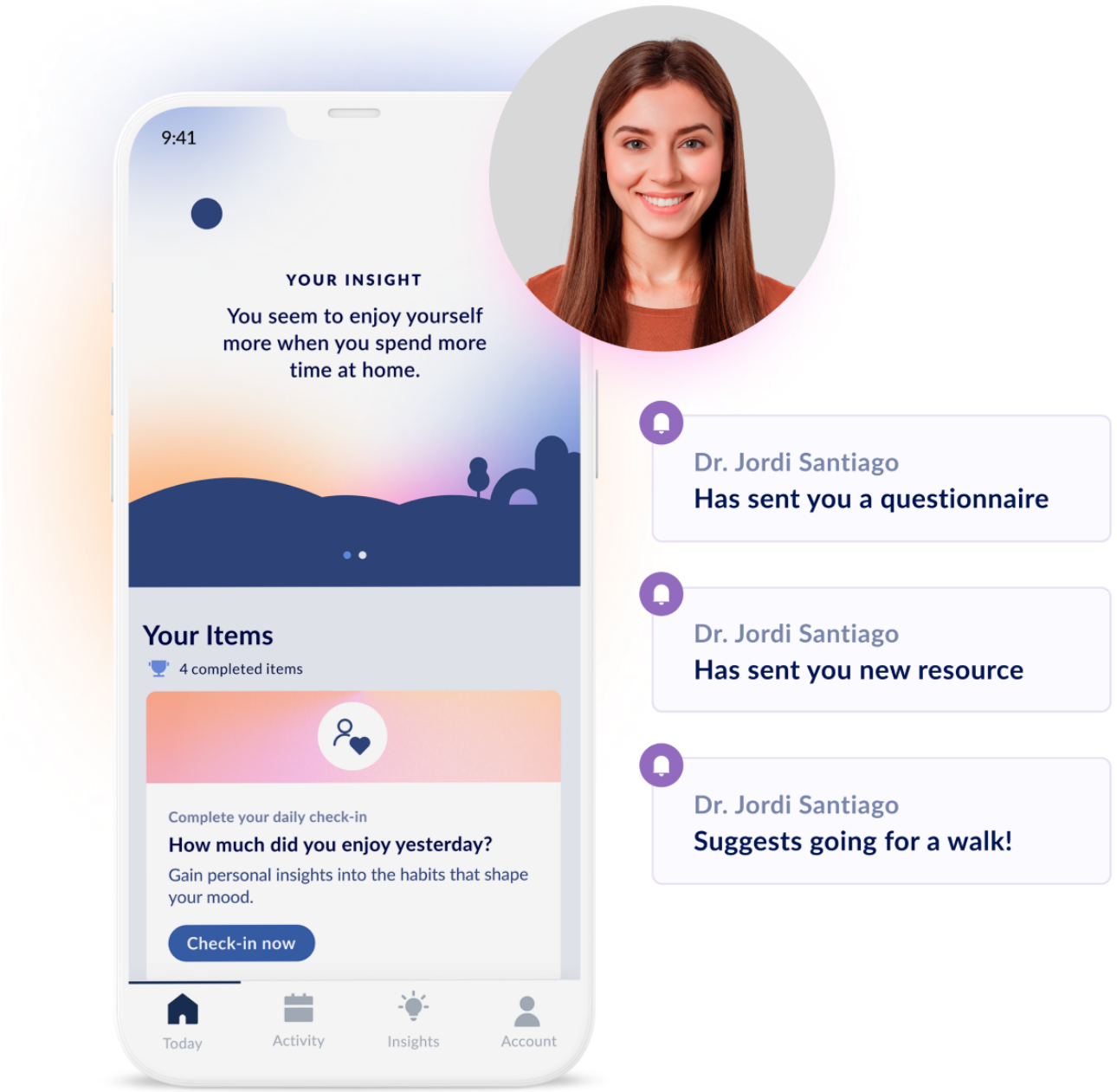
Measurement-based care is important for monitoring treatment effectiveness and improving clinical outcomes [12] but is not often used in routine clinical care due to concerns about practitioner and patient burden [13]. Through both passive (eg, built-in smartphone sensors) and active data collection (eg, experience sampling methods), smartphones can directly address these concerns by allowing practitioners and patients to monitor mood, behavior, symptoms, and risk factors continuously, objectively, and with a relatively low burden. In addition, smartphones afford novel opportunities to reinforce or introduce intervention content between sessions. For example, patients often forget to practice therapeutic skills outside of sessions [14], which may lead to poorer treatment outcomes [15]. Smartphone push notifications delivered between sessions can remind patients to practice therapeutic skills, thus increasing skill usage and improving patient outcomes [16].

Although digital mental health interventions have significant potential, low adherence and high attrition can reduce clinical effectiveness [17]. In particular, digital interventions that do not address the unique needs and perspectives of end users may exhibit low usability, which leads to lower sustained engagement [18]. When developing digital interventions, it is important to consider the clinical and implementation context [19,20]. User-centered design (UCD) uses an iterative, stakeholder-engaged process to cocreate products that directly respond to the end user experience [21,22]. Through UCD, individuals with lived experiences are closely engaged during the design and development processes by providing feedback that can be used by intervention developers and designers to tailor the content, interface, and implementation of digital interventions [23,24]. UCD studies also inform the translation of evidence-based treatments to digital formats and reveal critical information about facilitators and barriers to adoption that guide implementation approaches [25,26].

The Vira digital behavior change platform consists of a user-facing smartphone app and a web-based linked practitioner portal, Vira Pro (refer to Figure 1). The Vira platform was developed based on principles of behavioral activation, an effective treatment for depression and other mental health problems [27]. The smartphone app uses mobile sensing and brief self-report surveys to monitor numerous indicators of mental health problems (eg, sleep disturbance and low mood) or proximal risk factors. The app provides patients access to these data 24/7 and offers insights into passively sensed behavioral patterns associated with their moods. The web portal displays these data to the clinician, allowing them to gain more nuanced insights into their patients' behavior and mood between sessions. Clinicians can also use the web portal to schedule personalized push notifications to be delivered to a patients' smartphone, thereby facilitating enhanced support between sessions. Importantly, this functionality allows pushing notifications to be delivered at the "right time" (eg, as just-in-time adaptive interventions), which may improve outcomes [28].

XSL•FO
RenderX

Figure 1. The Vira digital behavior change platform consists of a user-facing smartphone app, Vira, and a linked, web-based practitioner portal, Vira Pro.



Study Objectives

This study aimed to evaluate the Vira mobile app and Vira Pro practitioner portal for use with youth at high risk for suicide by using an iterative formative development process to incorporate feedback from adolescents with lived experience, care providers, and business gatekeepers (ie, individuals who are responsible for making decisions regarding financial and other support for clinic initiatives). By including the perspectives of these end users, we sought to develop a digital platform that can be integrated into clinical workflows, is maximally engaging, and optimized for successful implementation.

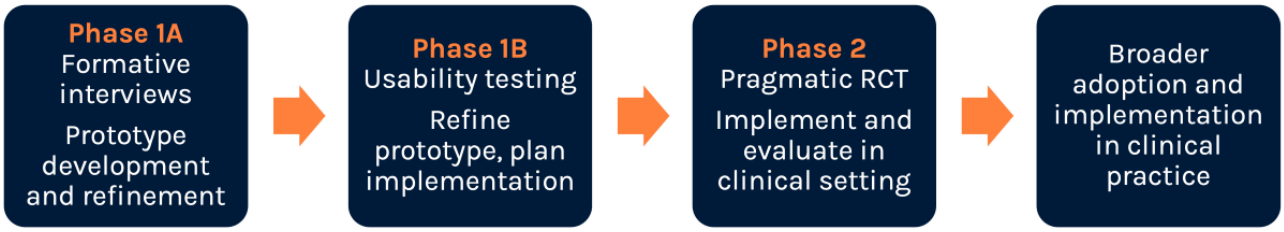
Methods

Study Overview

This study was part of a larger project to develop and test a behavior change platform for use in an intensive outpatient program (IOP) serving youth at high risk of suicide. The goal of the iterative formative development study, which consisted of 2 phases of mixed methods data collection, was to inform product development and implementation for a pragmatic randomized controlled trial (RCT; refer to Figure 2 for study design) [29].

XSL•FO
RenderX

Figure 2. Overview of the iterative formative development process and overall project design. RCT: randomized controlled trial.



Participant Recruitment

Participants were recruited between October 2022 and March 2023. Three types of stakeholders were recruited from the broader NewYork-Presbyterian Hospital, ColumbiaDoctors, and New York State Psychiatric Institute communities to participate across 2 iterative phases of research (Phase 1 and Phase 2). Stakeholders included adolescents who had received previous mental health treatment, mental health clinicians, and business gatekeepers.

Adolescent participants were recruited from a pool who had previously participated in other studies and provided informed consent to be contacted for future study opportunities. Adolescents were considered eligible if they were currently receiving or had received mental health treatment for depression or anxiety symptoms within the past year, were between the ages of 13 and 18 years old, and were fluent in English.

Clinicians and business gatekeepers were recruited internally via outreach emails and electronic flyers. Eligibility criteria for clinicians and business gatekeepers included present employment in youth mental health services at NewYork-Presbyterian Hospital, ColumbiaDoctors, or New York State Psychiatric Institute, ages 18-56 years old, and fluency in English. Business gatekeepers served in clinical leadership roles (eg, clinic director and executive director). Several clinicians who participated in Phase 1 also participated in Phase 2. One participant held both administrative and clinical roles and therefore participated as a business gatekeeper in Phase 1 and a clinician in Phase 2. Several of the clinicians and business gatekeepers have previously collaborated with RPA on other research projects.

Study Procedures

This study was an iterative formative development process consisting of 2 phases of mixed methods data collection. A UCD approach was adopted whereby changes made to the app prototype presented in Phase 2 were based on participant feedback from Phase 1. In both phases, participants completed semistructured interviews, an unmoderated click-through and think-aloud exercise to provide feedback on the minimal viable product version of the app, and a quantitative feedback survey using items adapted from the technology acceptance model (TAM) [30]. The TAM has 2 scales to measure specific beliefs that may influence the acceptance of technology: perceived usefulness and perceived ease of use. Each scale contains 10 items, which are rated using a 7-point Likert scale from 1 (strongly disagree) to 7 (strongly agree). Higher ratings indicated higher acceptability. Participants in both phases also completed a demographics questionnaire. Each interview was conducted by 2 moderators (TCT and DD). Both moderators are female and hold bachelor's degrees. TCT is a research assistant, and DD oversaw product design and development at Ksana Health, Inc at the time of the study.

Phase 1

Formative interviews in Phase 1 aimed to increase our understanding of the target patient population and clinical context, elicit feedback on the Vira minimal viable product to inform adaptations and refinements to the platform to increase relevance for the IOP clinical setting, promote perceived ease of use, understand potential barriers, and facilitate future implementation. Adolescents, clinicians, and gatekeepers were included in this phase.

Adolescent interview topics in Phase 1 included their past and current experiences receiving mental health treatment, smartphone usage and habits, and initial impressions of Vira based on a verbal description of the platform. Clinician interviews focused on their current therapy practices, experience using technology as part of the mental health treatment they deliver, past experiences integrating digital tools into existing electronic health record (EHR) systems, and their first impressions of Vira. Business gatekeepers reported on their experiences managing clinicians and clinic caseload, use of technology as part of mental health treatment, past experiences integrating new tools into existing EHR systems, and their first impressions of the platform. Interviews were approximately 30 minutes in duration and were conducted through Microsoft Teams videoconferencing software.

After completing the interview, participants took part in an unmoderated user testing session. Usability testing was conducted through UserTesting and began with a "think-aloud" procedure to test user interactions with Vira [31,32]. Participants were asked to verbalize their reactions as they viewed screens and to describe their navigational choices as they used the platform. Users were also asked about barriers and facilitators to using the platform, perceptions of specific features, and how to improve it.

Phase 2

The goal of Phase 2 usability testing was to understand the usability and acceptability of the refined Vira platform and deepen our knowledge of the implementation context to support clinical implementation in the subsequent pragmatic RCT. Adolescents and clinicians were included in this phase, as they were the primary target end users in the subsequent pragmatic RCT.

Phase 2 interviews with adolescents and clinicians explored similar topics to the Phase 1 interviews. In Phase 2, clinician interviews additionally focused on implementation, clinician

XSL•FO
RenderX

readiness, and usability of the revised Vira platform. Interviews necessitated approximately 30 minutes and were completed through Microsoft Teams and Zoom. During the unmoderated user testing portion of the session completed through the UserTesting platform, clinicians were provided with a mock patient profile and asked to think aloud as they explored the Vira tool as if they were using it in a client session. The goal of this mock case study testing format was to understand the clinical implementation context, facilitators and barriers to use, and overall usability. Clinicians were also asked more detailed questions about EHR integration, including which data points from Vira would be most useful to them.

Ethical Considerations

This study was reviewed and approved by the New York State Psychiatric Institute Institutional Review Board (protocol #8307). Signed electronic informed consent was obtained from all participants included in this study. The study data were anonymized and deidentified. All data is uploaded to a secure research server that is only accessible by the research team. Participants were compensated US $200 for their participation in the study.

Data Analysis

Similar analytic approaches were used across phases. Qualitative and quantitative data were analyzed separately.

Semistructured Interviews and UserTesting Transcripts

Semistructured qualitative interviews and UserTesting transcripts were analyzed using a thematic analysis approach [33]. Coding categories were developed to characterize participant responses, including reactions to Vira's features and functions as well as overall difficulty or facility using the platform. Given the iterative nature of the project, rapid analysis techniques, including matrix displays and triangulation [34,35], were also used. Multiple team members (TCT, LSW, DD, RPA, and NBA) reviewed transcripts to identify themes across phases for each persona (adolescent, clinician, or gatekeeper). Once team consensus was reached, findings were summarized narratively by theme and persona.

Quantitative Surveys

Demographics were analyzed using descriptive statistics and frequencies (mean [SD] and n [%] where applicable). TAM ratings were analyzed by comparing frequencies (mean [SD]) and n (%) of ratings within and across phases.

Mixed Methods Interpretation

The 2 sets of results (qualitative and quantitative) for each persona were triangulated by theme across both phases to examine how each set of data compared and contrasted to the other set and to explore the complementarity of the results across data collection methods. The qualitative and quantitative results were then synthesized for interpretation. The mixed methods synthesis of qualitative and quantitative results across both phases is presented by persona in the *Results* section.

Results

Demographics

A total of 36 participants were recruited across both phases of the study. In Phase 1, a total of 8 adolescents, 7 clinicians, and 5 gatekeepers participated in stakeholder consultations. One gatekeeper did not provide demographic information. In total, 4 of the clinicians in Phase 1 also participated in Phase 2. One gatekeeper participated in Phase 2 as a clinician. A total of 8 adolescents and 7 clinicians participated in Phase 2. Refer to Table 1 for sample characteristics.

Table 2 summarizes participant responses to the TAM. One gatekeeper did not complete TAM in Phase 1 and one clinician did not complete TAM in Phase 2. For individual item frequency distributions, please refer to Figures S1-S6 in the Multimedia Appendix 1.

XSL•FO
RenderX

Table 1. Demographic characteristics of participants (N=36).

	Phase 1	Phase 2
Adolescents, n	8	8
Age (years), mean (SD)	16.2 (0.91)	17.5 (0.91)
Women, n (%)	7 (87.50)	7 (87.50)
Race, n (%)		
White	3 (37.50)	5 (62.50)
Asian	1 (12.50)	2 (25.00)
Black	0 (0)	1 (12.50)
More than one race	3 (37.50)	0 (0)
Unknown or not reported	1 (12.50)	0 (0)
Ethnicity, n (%)		
Hispanic	1 (12.50)	0 (0)
Non-Hispanic	7 (87.50)	8 (100)
Clinicians (N)	7	8
Age in years, mean (SD)	32.14 (5.24)	34.75 (4.92)
Women, n (%)	7 (100)	8 (100)
Race, n (%)		
White	5 (71.43)	7 (87.50)
More than one race	2 (28.57)	1 (12.50)
Ethnicity, n (%)		
Hispanic	1 (14.29)	0 (0)
Non-Hispanic	6 (85.71)	8 (100)
Business gatekeepers (N)	5	—[a]
Age (years), mean (SD)	44.6 (9.09)	—
Women, n (%)	2 (40)	—
Race, n (%)		
White	4 (80)	—
Unknown or not reported	1 (20)	—
Ethnicity, n (%)		
Non-Hispanic	4 (80)	—
Unknown or not reported	1 (20)	—

[a]Not applicable.

XSL•FO
RenderX

Table 2. Participant technology acceptance model responses (N=33).

Technology acceptance model	Adolescents, mean (SD)		Clinicians, mean (SD)		Gatekeepers, mean (SD)
	Phase 1 (n=8)	Phase 2 (n=7)	Phase 1 (n=7)	Phase 2 (n=7)	Phase 1 (n=4)
Perceived usefulness scale					
Greater control	5.63 (0.74)	5.14 (1.35)	4.86 (0.69)	5.14 (1.68)	6.00 (0.82)
Improve quality	6.13 (0.64)	5.14 (1.86)	4.86 (0.90)	5.00 (2.16)	6.50 (0.58)
Accomplish quickly	5.50 (1.31)	5.86 (1.21)	3.86 (2.12)	4.71 (1.38)	5.75 (1.50)
Support critical aspects	5.75 (1.28)	5.29 (1.25)	5.00 (0.82)	5.43 (1.62)	6.50 (0.58)
Productive	5.88 (0.99)	5.71 (1.38)	3.86 (1.07)	4.43 (2.30)	5.25 (2.06)
Improve	5.75 (1.28)	5.57 (1.40)	4.43 (1.62)	5.14 (1.68)	5.75 (0.96)
Accomplish more	6.00 (1.20)	5.43 (1.40)	5.14 (1.21)	3.29 (2.14)	5.50 (1.29)
Enhance effectiveness	6.00 (1.31)	5.57 (1.40)	5.00 (0.82)	5.14 (1.95)	6.50 (0.58)
Easier	5.75 (1.49)	4.71 (1.25)	3.86 (1.77)	4.57 (1.40)	5.50 (1.29)
Useful	6.00 (0.93)	5.43 (1.13)	5.00 (1.00)	5.86 (1.21)	6.25 (0.50)
Perceived ease of use scale					
Cumbersome to use	2.71 (1.38)	3.14 (1.86)	3.43 (1.27)	4.71 (2.14)	2.00 (1.41)
Easy to learn	5.75 (1.48)	5.57 (0.98)	5.71 (1.11)	4.86 (0.90)	6.00 (0.82)
Frustrating	2.63 (1.60)	2.29 (1.11)	3.00 (1.15)	2.86 (1.07)	2.00 (0.82)
Confusing	2.50 (1.41)	2.43 (1.51)	3.29 (1.60)	2.71 (1.38)	2.00 (0.82)
Rigid and inflexible	2.00 (1.31)	2.29 (1.60)	3.14 (1.68)	2.00 (1.00)	2.75 (1.50)
Easy to remember	5.38 (1.92)	5.29 (1.11)	5.57 (0.53)	4.71 (1.80)	6.50 (0.58)
A lot of mental effort	3.38 (1.19)	4.00 (2.24)	3.86 (1.35)	3.57 (1.51)	2.25 (1.50)
Easy to understand	6.13 (1.13)	6.00 (1.00)	5.57 (0.53)	4.86 (0.69)	5.00 (2.45)
Easy to become skillful	6.38 (0.52)	5.29 (1.70)	5.29 (0.76)	5.00 (1.00)	6.25 (0.96)
Easy to use	6.25 (0.71)	5.86 (1.21)	5.57 (0.53)	5.43 (1.13)	6.00 (1.15)

Adolescents

Passive Sensing Data

The overall concept of using mobile sensors to passively capture data about their behavior was intriguing to all participants. Adolescents felt that the passive sensing data would be more accurate than their own self-reports in reflecting their sleep and fitness habits and that this technology could improve the efficiency of their therapy sessions by allowing them to easily share behavioral data with their clinician. Adolescents also thought that Vira would improve the quality of their treatment and support critical aspects of it (TAM perceived usefulness items 2 and 4; refer to Table 2).

Adolescents across both phases had mixed reactions to the different types of data that would be collected by the Vira app via mobile sensors. Most adolescents were excited for the app to track their sleep, walking, and exercise patterns, and identified these behaviors as factors that impact mental health but that they are often not aware of or tracking on their own. They were also enthusiastic about their clinicians accessing these data through Vira and using it to support their treatment. Refer to exemplar testimonials in Textbox 1.

However, participants were more apprehensive about sharing location (GPS) data. Many were confused about how location data sharing would work and expressed that they did not want their therapist to actively track their live location (even though this is not a capability of the app). One adolescent expressed that their data-sharing preferences could vary based on their emotional state. There were also some concerns regarding how language patterns derived from keyboard data would be interpreted by clinicians.

> *I feel like people's words and conversations can be interpreted in many ways, and I don't feel like statistically, they'll be an accurate representation.* [ID A13, Phase 2]

XSL•FO
RenderX

Textbox 1. Passive sensing data feedback.

- *Having a record of things like sleep and physical activity can help you because these are things that sometimes we do not think twice about. But they have a huge impact on how we are feeling so to have a record and to also be able to have that type of platform to communicate with your therapist sounds really efficient.* [ID A08, Phase 1]
- *My therapist seeing that I am not sleeping well would be helpful, it will save time in therapy and allow for deeper and more helpful conversations rather than recapping.* [ID A03, Phase 1]
- *If it is about sleep and your physical activity, that is something that can affect your mental health. So, it is not like it is taking unnecessary information. It is just like more information that you may not even think to bring up in your session because it is just part of your daily life.* [ID A15, Phase 2]

Home Page Feedback

Adolescents responded positively to the app's home page interface, stating it was easy to navigate and aesthetically pleasing. Most adolescents perceived it as easy to learn and understand (TAM perceived ease of use items 2 and 8). Most adolescents also liked the sample psychoeducation articles that were displayed on the home page.

However, some adolescents expressed concern about engagement with the app given that personalized insights about connections between their mood and their passively sensed behaviors would not be available until they reported their daily mood for 10 days. Adolescents suggested that having more features available during the period before insights are available would increase app engagement.

Communicating With Clinicians Between Sessions and Nudges

Adolescents liked the idea of clinicians using Vira to send nudges to them between sessions, particularly psychoeducational videos, reminders to use coping skills or general mood boosters.

> *It would be nice to have a therapist who knew how I felt even on the days that we didn't meet and could send me a little something.* [ID A02, Phase 1]

However, there were a few adolescents who expressed concern about the potential intrusiveness of between-session contact. They were also concerned with the nuisance of receiving too many smartphone notifications.

Several adolescents in Phase 2 felt that the frequency of between-session communication should be variable based on a few factors, including the strength of the relationship and their current symptom severity. Adolescents expressed that they would feel more comfortable engaging in more frequent communication with their therapist between sessions only once they had built a good rapport and relationship with them.

Clinicians

Using Vira in the Clinical Workflow

Clinicians consistently expressed that they are seeking a tool that facilitates their workflow and makes their job easier, not harder. Clinicians explained that they have many responsibilities and take on multiple roles (eg, direct service provider and clinic administrator), and thus need a tool that will lend itself to more efficient preparation before client sessions.

Clinicians interviewed in Phase 1 felt the Vira Pro interface was overwhelming, with potentially too much data provided. One clinician felt that using Vira could slow down their workflow.

> *It would make me accomplish tasks more slowly because there's so much more data.* [ID C02, Phase 1]

Between the Phase 1 and Phase 2 tests, changes were made to the prototypes to simplify the data display in Vira Pro. Yet, clinicians interviewed in Phase 2 also felt the large volume of data was challenging, and like clinicians interviewed in Phase 1, felt that trying to extract clinically relevant components and apply them to their sessions could create extra work.

> *I guess I have a little bit of a worry of how do I incorporate all this data in a way that doesn't feel like a waste of time for the patient but isn't creating a lot of extra work for me compared to what I'm doing today?* [ID C11, Phase 2]

Some clinicians suggested that more concise data summarization within Vira Pro would make the tool more usable. Despite their comments about the interface being overwhelming, clinicians across both phases felt that Vira would be easy to learn and use (TAM perceived ease of use items 2 and 10).

Although all clinicians across both phases felt that Vira would be useful to their clients' treatment (TAM perceived usefulness item 10), there were mixed opinions on whether Vira would increase their efficiency and job performance. The majority of clinicians felt that Vira would support critical aspects of their job and enhance their effectiveness on the job, but there were differences in perceptions of whether using Vira would allow them to accomplish more tasks quickly and increase their job productivity (TAM perceived usefulness items 3 and 5). Interestingly, fewer of the Phase 2 clinicians felt that Vira would allow them to accomplish more in their job than would otherwise be possible compared to Phase 1 clinicians (TAM perceived usefulness item 7).

Nudges

Clinicians had mixed opinions on the utility of scheduling nudges to practice therapy techniques. Some reacted very positively to the feature and felt that it could be useful to their patient's treatment by reinforcing the use of therapy principles discussed in the session in the context of the patient's everyday life.

> *I love the idea of the nudges, and I'm thinking how instrumental that would be in transforming phone*

XSL•FO
RenderX

> *coaching and generalizing any sort of DBT skills to the environment.* [ID C03, Phase 2]

However, other clinicians felt that nudges between sessions may lead patients to become too reliant on their therapist and neglect to engage in therapeutic strategies on their own, a key step in sustained behavior change. One clinician felt that it was unnecessary to use nudges to remind patients to use therapeutic techniques, preferring instead to simply speak to the patient in person about using them.

> *And then nudging, I'm not particularly fond of, because I feel like it could potentially reinforce passivity for a patient. If a provider is nudging them and reminding them to do things, it's kind of taking away the autonomy that they can do it on their own.* [ID C04, Phase 2]

> *I'm just imagining it's us sitting there monitoring our patients and nudging them to use skills [...] and it fragilizes them and doesn't hold them accountable for being responsible for their own well-being.* [ID, C02, Phase 2]

Within the Vira Pro dashboard, all clinicians agreed that the section showing "default" nudges should be adapted for each client's clinical needs. For example, the most common therapeutic modalities and techniques used for a patient with anxiety are different from those used for a patient with depression.

> *I would want to be able to customize it for the client because I think it would make a difference based on what I was working with the client on. [...] So they might vary depending on the focus of the treatment.* [ID C09, Phase 2]

EHR Integration

Clinicians offered several suggestions related to integrating the Vira platform into EHR systems. They frequently stated that having the data that is collected in Vira (eg, passive sensing data and assessment responses) automatically populate into their EHR would reduce the burden of manually uploading data. Similarly, clinicians thought that it would be efficient to have clinical notes generated in Vira appear in their EHR system, reducing the need for them to rewrite their notes in multiple locations.

> *Minimal effort on my end to change anything. Minimal training. Nothing that I have to do too much work on my end to upload [Vira data].* [ID C04, Phase 1]

To support seamless integration, several clinicians felt it would be important to reduce the number of redundant logins needed to access Vira within their EHR. Clinicians interviewed in Phase 2 considered the following types of Vira data most important for EHR integration: daily self-reported mood check-ins, assessment responses and scores, and the Vira-derived associations between self-reported mood and client behavior.

Business Gatekeepers

Overview

All gatekeepers felt that Vira would be a useful tool for their team's treatment toolkit (TAM perceived usefulness item 10). All gatekeepers felt that using Vira would improve the quality of the work they do (TAM perceived usefulness item 2), support critical aspects of their job (TAM perceived usefulness item 4), and enhance their effectiveness on the job (TAM perceived usefulness item 8). They also agreed that Vira seems easy to learn (TAM perceived ease of use item 2), easy to become skillful at using (TAM perceived ease of use item 9), and easy to use (TAM perceived ease of use item 10).

Vira Pro Interface

Through the user testing sessions, administrators provided feedback about the usability and acceptability of several aspects of the Vira Pro administrator dashboard. Gatekeepers responded positively to the visual layout and liked that they could access information about all patients in their clinic from a single page within the administrator dashboard. They also expressed that it would be important for the results of patient's risk assessments to be displayed within the dashboard alongside the other assessments that have been completed so that they are easily accessible for practitioners.

There were several aspects of the practitioner platform that gatekeepers were unclear about. Within the administrator dashboard, it was unclear to them where the data for "Top 10 Used Intervention Techniques" is derived from and that it is a compilation of all the clients within the clinic. Gatekeepers also sought more detailed explanations of the Vira metrics within the dashboard, such as how language categories are defined and examples.

> *I would love to see how these categories of language are defined in really clear ways. Like 'rigid words' for example, I'd love to see some examples.* [ID G2, Phase 1]

Implementation Considerations

During the interviews, gatekeepers were asked to consider several aspects of implementing Vira within their clinic. First, gatekeepers felt that Vira would be helpful because it could inform clinic treatment planning.

Gatekeepers were asked which clinicians would find Vira most helpful. All participants stated that Vira would be a useful tool for clinicians who employ cognitive behavioral therapy and DBT within their treatment framework, as Vira's functionalities could enhance their ability to practice those treatment modalities.

> *Folks who are involved in DBT and CBT could really be the ones that could utilize the tool because it most mimics what we already do, right? CBT, it's all about the exposures so how could we use Vira to help with the exposures. With DBT, it's all about practicing those skills and some of the skills are about sleeping well, taking your medication, exercising or just doing any sort of physical activity, all those things. So again,*

XSL•FO
RenderX

> *I think in some ways, this would really complement those types of providers.* [ID G2, Phase 1]

Gatekeepers felt Vira might be useful for patients experiencing depression and/or anxiety. One gatekeeper mentioned Vira potentially being useful for patients with physical health comorbidities due to Vira's health behavior metrics and expressed interest in integrating Vira with primary care or other medical departments.

Gatekeepers also brought up several potential barriers to implementation. The most common concern was the confidentiality of patient data collected from Vira. There were also some concerns about the extent of parental involvement, including how much Vira data they could access if patients are minors. Some gatekeepers were also worried that implementing Vira could be potentially burdensome for providers by adding to their workload, depending on clinic expectations.

> *I do worry that it would be burdensome to the provider if there's any expectations around how much they're supposed to be babysitting, so to speak.* [ID G3, Phase 1]

Another implementation roadblock gatekeepers identified is getting clinicians to try a new tool and do something different in their treatment. One gatekeeper suggested incorporating multimodal training to help clinicians transition to using the tool could help overcome this obstacle. Gatekeepers also foresaw difficulties in integrating Vira with existing EHR systems.

Discussion

Principal Findings

This study used UCD methods to garner feedback about the Vira behavior change platform from 3 groups of key stakeholders: adolescent patients, clinicians, and business gatekeepers. All participants expressed that Vira could be useful in enhancing mental health treatment. Adolescents had positive feedback on the aesthetics and design of the Vira app, factors that may promote engagement with mental health apps for youth at risk of suicide [36]. Overall, clinicians felt that the displays in the Vira Pro practitioner portal were too dense due to the volume of data shown and had mixed feedback about how using Vira would impact their efficiency and productivity. Clinicians were generally open to the idea of using Vira to send nudges to clients between sessions but expressed some concerns. Gatekeepers had mixed feedback on the usability of the tool but felt that Vira would enhance their effectiveness on the job and shared valuable feedback to inform implementation.

General Capabilities and Data Sharing

All participants expressed excitement about Vira's passive sensing capabilities to track behavior and the daily mood. This aligns with UCD research conducted as part of the development of the BlueIce app, which similarly found that adolescents were enthusiastic about tracking mood patterns and noticing their patterns, as well as identifying triggers of negative moods, similar to Vira's insights [37]. However, there was some apprehension about potential intrusiveness to patients. Specifically, adolescents were most worried about sharing location data and not wanting to be actively tracked by their therapist, and clinicians were only interested in accessing information that adolescents were enthusiastically willing to share with them. To address adolescents' concerns about collecting and sharing location and keyboard data, it will be important to provide them with detailed information and instructions on how their data is processed, stored, and used by the app and their clinician at the time of installation. Specifically, clients should be informed that clinicians do not have real-time access to behavioral data, nor do they have specific details about exact locations or words typed into their smartphones. Taken together, these findings suggest that clear alignment between patients and providers on which data they are comfortable sharing, and what exactly is shared, is important for aiding clinicians and clients in leveraging passive sensing data in therapy and maintaining supportive therapeutic relationships.

Communication Between Sessions (Nudges)

The Vira tool allows clinicians to send prescheduled notifications to patients to support compliance with their therapy plans. By collecting passive sensing data that is displayed in both the Vira app and on the practitioner dashboard, Vira also provides information about client behavior continuously between sessions. In this study, adolescents and clinicians expressed mixed views on whether these features would improve treatment. While some adolescents were enthusiastic, others noted that it might be frustrating to receive too many notifications from their therapist. This aligns with feedback from target adolescent end users of the LifeBuoy app regarding the frequency of app notifications [36]. Several clinicians in this study shared similar concerns.

Stakeholder feedback from this study therefore suggests it is essential to set clear expectations with both adolescents and practitioners about how that between-session communication will be operationalized. Gatekeepers felt the tool would be a supportive addition to their team's treatment toolkit, but were concerned about the potential burden on clinicians, underscoring the importance of clearly defining expectations with providers. Along similar lines, when introducing the tool to their patients, practitioners should discuss both the frequency with which they will use Vira and how they will use it to support treatment. Similar to Hetrick et al [38], clinicians in our study expressed that although the tool could be useful in monitoring patients' moods, monitoring patient data multiple times between sessions or in real time is not feasible. However, to be clear, this is not an expectation of typical use of the Vira Platform. Ensuring the tool is designed to fit within clinician workflows and managing patients' expectations for how their clinician will use it increases the likelihood of clinician adoption [39].

Adolescents were very enthusiastic about Vira's nudges. They uniformly liked the concept of clinicians using the passive sensing data to determine what types of nudges should be sent to enhance their treatment. They also felt it would be helpful for clinicians to send nudges as reminders to use therapeutic techniques learned during treatment. Conversely, clinicians had mixed opinions on the nudging feature. Some clinicians felt nudges could be very useful and could transform treatment in a positive way, consistent with recent research demonstrating

XSL•FO
RenderX

that nudges and prompts increase engagement in self-guided interventions [40]. However, it is important to note that patients in the IOP setting would be using Vira to support DBT, a collaborative and therapist-guided treatment modality. Other clinicians expressed concern that patients may become dependent on their clinician to send nudges to remember to use therapeutic techniques discussed in sessions. In sum, Vira's nudge capability may help promote habit formation in terms of therapy techniques and coping skills outside of the therapy sessions, and it is important that digital products allow clinicians to schedule and structure nudges in line with each patient's personalized and dynamic treatment needs. For example, clinicians may schedule nudges to be delivered less frequently as skills are acquired and used.

Product Changes

Findings from this iterative UCD process directly informed the development of several changes to the Vira platform that will be implemented for the pragmatic RCT. In response to feedback about data sharing, app users will have the flexibility to choose which data streams they would like to share with their practitioner, including specific streams of passively sensed behavioral data and daily mood (enjoyment) ratings. Enhancements have also been made to the app to boost engagement during the initial 10-day period in which Vira collects sufficient information to calculate personalized insights about behaviors that may be related to a user's mood. In the version of the app that was tested in this study, personalized insights were not available to users until their 10th daily check-in. Based on insights from this UCD study, the app has been updated such that users receive more onboarding information within the app after their third daily check-in, and insights about the consistency of their sleep and walking patterns after their sixth and eighth check-ins, respectively.

Clinicians and gatekeepers expressed strong desires for deeper information about the various sensor streams and data types available within the dashboard, as well as a simpler and more accessible user interface. Therefore, the Vira Pro dashboard was modified to show a more summarized, simplified form of patient data with more options to hide granularity in data charts. The revised version of the platform also retains the capability to drill into the details as needed; Vira Pro users can now hover over fields within the dashboard, such as sensor streams and specific data points, to view more details about how the data were collected and processed. Simplifying the user interface is a particularly important product change, considering that providing an overload of information can discourage clinicians from using the digital tool [41].

Implementation Considerations

Previous research has shown that practitioners' negative attitudes toward digital mental health interventions and resistance to change are common barriers to implementing new digital technologies in mental health care [39]. Similarly, through clinician and gatekeeper interviews, this study found that clinician buy-in with using a new tool is vital. Consistent with other research, clinicians and gatekeepers expressed concerns about the potential for mobile technology to increase professional workload [42]. Providing clinicians with sufficient training to ensure their confidence and efficiency in using the tool may increase the likelihood that clinicians integrate the tools into their practice [43], and help them overcome any concerns or hesitations they may have regarding implementation. Thorough training was also commonly suggested by gatekeepers in this study as a vital step to implementing Vira in their clinic. Further, when implementing Vira within a clinical setting, clinic administration needs to set clear expectations for how they expect Vira to be used within treatment. Moreover, clinicians and gatekeepers expressed that friction integrating Vira with existing EHR systems would be a barrier to implementation, consistent with previous research [44].

Findings from this study directly informed the development of a focused training approach that is being rolled out in the ongoing pragmatic RCT [29]. In response to clinician feedback about having limited time to learn new tools, the clinician training uses a blended synchronous and asynchronous format to provide flexibility while also keeping the training focused and efficient (1 hour total). To address mixed views regarding whether Vira would help them increase productivity, the training emphasizes critical thinking on how to apply the Vira platform to their practice. The training sequence focuses on building self-efficacy for using the platform to enhance their practice and help augment their ability to complete existing tasks. Many video and written resources are also available through a self-service knowledge base for asynchronous review and support after the training.

Generalizability and Broader Implications

Findings from this study can be generalized to inform the development and implementation of other digital technologies for high-risk youth. Both adolescents and clinicians were enthusiastic about the potential utility of passive sensing data for treatments [37], but concerns regarding data sharing underscore the need for transparency about how data from digital health platforms are used [45].

This study also revealed important issues to consider more broadly when implementing new digital technologies into health care systems, such as the impact on clinician workloads and integrating the tool within existing clinical workflows and EHRs [39,44]. User interface design is a key consideration in increasing the acceptability of new digital technologies. Data displays should be simple and accessible so as to not overwhelm users and increase the likelihood of adoption [41].

Limitations

This study has several limitations. Sample characteristics and recruitment methods may have impacted the feedback elicited in this study and the generalizability of the results. Many adolescents in the study had engaged in higher levels of care (eg, inpatient admission, partial hospitalization programs, and IOPs). However, the inclusion criteria for adolescent participants only required them to be currently in treatment or have received treatment in the past year for depressive or anxiety symptoms. Accordingly, some of the participants recruited may have only received a minimal level of care (ie, just a few sessions) and may have been at lower risk compared to the intended user population for Vira. In addition, the sample was limited in terms

XSL•FO
RenderX

of racial and ethnic diversity, and the majority of the adolescent sample and the entire clinician sample were female. This may be because females are more likely to seek treatment for mental health [46], and the mental health workforce is predominately female [47]. Moreover, there was also a relatively small sample size. For grounded theory approaches, a sample size of 20-30 participants is generally recommended [48], although a recent systematic review found that among studies with a relatively homogeneous study population and specific objectives, a total of 9-17 interviews may be enough to reach saturation [49]. This study reached saturation with a similar sample size. Finally, we did not include parents as a stakeholder in this study. It may be useful to include parents in future studies to understand how their involvement can facilitate the implementation of digital technologies in adolescent mental health treatment.

Future Directions

Clinician feedback about the lack of time to prepare for sessions suggests that automation may increase the acceptability and perceived usefulness of the platform. Feedback has informed the ongoing design and development of AI-supported features such as automatically generated drafts of clinical progress notes, “heads up” summaries of the most salient clinician issues that can be quickly accessed before each session, and a clinical assistant that can leverage clinical practice guidelines across a range of behavioral health conditions to suggest enhancements to therapy plans based on data collected in Vira. Moreover, the clinical effectiveness of the Vira platform is currently being investigated in a 2-arm, pragmatic RCT in which youth at risk for suicide within an IOP will be randomly assigned to use Vira with their clinician or engage in usual care [29]. Throughout this clinical trial, we will collect feedback from both patient and clinician participants.

Conclusions

In sum, participants felt that the Vira platform has a high potential to improve engagement with mental health treatment as well as mental health outcomes in high-risk youth. Our rigorous, multimethod approach uncovered key facilitators and barriers to implementation, many of which are consistent with previous research, highlighting the challenge of translating novel, technology-based interventions from controlled to clinical settings. Applying iterative, UCD methods allowed us to integrate valuable stakeholder feedback to develop an intervention that is clinically relevant to end users, hopefully leading to successful implementation in real-world settings.

Acknowledgments

Funding from the National Institute of Mental Health provided support to the NBA (R44 MH128484-01A1). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

RPA and NBA designed the study. TCT and DD performed data curation. TCT, LSW, and DD performed formal data analysis. TCT, LSW, and CJF drafted the paper. TCT, LSW, DD, RPA, and NBA interpreted results. All authors read and approved the final version of the paper and agreed with the order of presentation of the authors.

Conflicts of Interest

LSW and NBA are employed by and hold equity in, Ksana Health, which developed and commercialized the Vira platform. DD was employed by Ksana Health at the time the project was initiated. TCT, CJF, CCC, ZKB, and CTO have no conflicts to disclose. RPA is an unpaid advisor to Ksana Health and a paid advisor for Get Sonar, Inc. Ksana Health has a Scientific Integrity Policy designed to prevent the company’s interest in bringing Ksana Health products to market from influencing or appearing to influence the science conducted using Ksana Health’s products. Ksana Health follows Open Science practices to ensure the transparency, integrity, and reproducibility of our work. This includes preregistration before data collection, and data, code, and material sharing upon request.

Multimedia Appendix 1
Additional figures.
[DOCX File , 1415 KB - formative_v9i1e65418_app1.docx]

References

1. Arango A, Gipson PY, Votta JG, King CA. Saving lives: recognizing and intervening with youth at risk for suicide. Annu Rev Clin Psychol 2021;17:259-284. [doi: 10.1146/annurev-clinpsy-081219-103740] [Medline: 33544628]
2. Centers for Disease Control and Prevention. 2021. URL: https://stacks.cdc.gov/view/cdc/124928 [accessed 2023-02-07]
3. Gaylor EM, Krause KH, Welder LE, Cooper AC, Ashley C, Mack KA, et al. Suicidal thoughts and behaviors among high school students - youth risk behavior survey, United States, 2021. MMWR Suppl 2023;72(1):45-54 [FREE Full text] [doi: 10.15585/mmwr.su7201a6] [Medline: 37104546]

XSL•FO
RenderX

4. Weisz JR, Venturo-Conerly KE, Fitzpatrick OM, Frederick JA, Ng MY. What four decades of meta-analysis have taught us about youth psychotherapy and the science of research synthesis. Annu Rev Clin Psychol 2023;19:79-105 [FREE Full text] [doi: 10.1146/annurev-clinpsy-080921-082920] [Medline: 36750262]
5. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
6. Lehtimaki S, Martic J, Wahl B, Foster KT, Schwalbe N. Evidence on digital mental health interventions for adolescents and young people: systematic overview. JMIR Ment Health 2021;8(4):e25847 [FREE Full text] [doi: 10.2196/25847] [Medline: 33913817]
7. Kennard BD, Goldstein T, Foxwell AA, McMakin DL, Wolfe K, Biernesser C, et al. As safe as possible (ASAP): a brief app-supported inpatient intervention to prevent postdischarge suicidal behavior in Hospitalized, Suicidal Adolescents. Am J Psychiatry 2018;175(9):864-872 [FREE Full text] [doi: 10.1176/appi.ajp.2018.17101151] [Medline: 30021457]
8. Torok M, Han J, McGillivray L, Wong Q, Werner-Seidler A, O'Dea B, et al. The effect of a therapeutic smartphone application on suicidal ideation in young adults: findings from a randomized controlled trial in Australia. PLoS Med 2022;19(5):e1003978 [FREE Full text] [doi: 10.1371/journal.pmed.1003978] [Medline: 35639672]
9. Stallard P, Porter J, Grist R. A smartphone app (BlueIce) for young people who self-harm: open phase 1 pre-post trial. JMIR Mhealth Uhealth 2018;6(1):e32 [FREE Full text] [doi: 10.2196/mhealth.8917] [Medline: 29382632]
10. Malakouti SK, Rasouli N, Rezaeian M, Nojomi M, Ghanbari B, Shahraki Mohammadi A. Effectiveness of self-help mobile telephone applications (apps) for suicide prevention: a systematic review. Med J Islam Repub Iran 2020;34:85 [FREE Full text] [doi: 10.34171/mjiri.34.85] [Medline: 33306048]
11. Kothgassner OD, Goreis A, Robinson K, Huscsava MM, Schmahl C, Plener PL. Efficacy of dialectical behavior therapy for adolescent self-harm and suicidal ideation: a systematic review and meta-analysis. Psychol Med 2021;51(7):1057-1067 [FREE Full text] [doi: 10.1017/S0033291721001355] [Medline: 33875025]
12. Lambert MJ, Whipple JL, Kleinstäuber M. Collecting and delivering progress feedback: a meta-analysis of routine outcome monitoring. Psychotherapy (Chic) 2018;55(4):520-537. [doi: 10.1037/pst0000167] [Medline: 30335463]
13. Lewis CC, Boyd M, Puspitasari A, Navarro E, Howard J, Kassab H, et al. Implementing measurement-based care in behavioral health: a review. JAMA Psychiatry 2019;76(3):324-335 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.3329] [Medline: 30566197]
14. Bunnell BE, Nemeth LS, Lenert LA, Kazantzis N, Deblinger E, Higgins KA, et al. Barriers associated with the implementation of homework in youth mental health treatment and potential mobile health solutions. Cognit Ther Res 2021;45(2):272-286 [FREE Full text] [doi: 10.1007/s10608-020-10090-8] [Medline: 34108776]
15. Kazantzis N, Whittington C, Zelencich L, Kyrios M, Norton PJ, Hofmann SG. Quantity and quality of homework compliance: a meta-analysis of relations with outcome in cognitive behavior therapy. Behav Ther 2016;47(5):755-772. [doi: 10.1016/j.beth.2016.05.002] [Medline: 27816086]
16. Tang W, Kreindler D. Supporting homework compliance in cognitive behavioural therapy: essential features of mobile apps. JMIR Ment Health 2017;4(2):e20 [FREE Full text] [doi: 10.2196/mental.5283] [Medline: 28596145]
17. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
18. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018;21(3):116-119 [FREE Full text] [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
19. Kruzan KP, Meyerhoff J, Biernesser C, Goldstein T, Reddy M, Mohr DC. Centering lived experience in developing digital interventions for suicide and self-injurious behaviors: user-centered design approach. JMIR Ment Health 2021;8(12):e31367 [FREE Full text] [doi: 10.2196/31367] [Medline: 34951602]
20. Blandford A, Gibbs J, Newhouse N, Perski O, Singh A, Murray E. Seven lessons for interdisciplinary research on interactive digital health interventions. Digit Health 2018;4:2055207618770325 [FREE Full text] [doi: 10.1177/2055207618770325] [Medline: 29942629]
21. Chokshi SK, Mann DM. Innovating from within: a process model for user-centered digital development in academic medical centers. JMIR Hum Factors 2018;5(4):e11048 [FREE Full text] [doi: 10.2196/11048] [Medline: 30567688]
22. Bhattacharyya O, Mossman K, Gustafsson L, Schneider EC. Using human-centered design to build a digital health advisor for patients with complex needs: persona and prototype development. J Med Internet Res 2019;21(5):e10318 [FREE Full text] [doi: 10.2196/10318] [Medline: 31094334]
23. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
24. de Beurs D, van Bruinessen I, Noordman J, Friele R, van Dulmen S. Active involvement of end users when developing web-based mental health interventions. Front Psychiatry 2017;8:72 [FREE Full text] [doi: 10.3389/fpsyt.2017.00072] [Medline: 28515699]

XSL•FO
RenderX

25. Lyon AR, Koerner K. User-centered design for psychosocial intervention development and implementation. Clin Psychol (New York) 2016;23(2):180-200 [FREE Full text] [doi: 10.1111/cpsp.12154] [Medline: 29456295]
26. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
27. Stein AT, Carl E, Cuijpers P, Karyotaki E, Smits JAJ. Looking beyond depression: a meta-analysis of the effect of behavioral activation on depression, anxiety, and activation. Psychol Med 2021;51(9):1491-1504. [doi: 10.1017/S0033291720000239] [Medline: 32138802]
28. Smyth, Heron. Is providing mobile interventions "just-in-time" helpful? an experimental proof of concept study of just-in-time intervention for stress management. : IEEE; 2016 Presented at: IEEE Wireless Health (WH); 2016 October 25-27; Bethesda, MD, USA p. 1-7. [doi: 10.1109/wh.2016.7764561]
29. Funkhouser CJ, Tse TC, Weiner LS, deLuise D, Pagliaccio D, Durham K, et al. Study preregistration: testing a digital suicide risk reduction platform for adolescents: a pragmatic randomized controlled trial. J Am Acad Child Adolesc Psychiatry 2024;63(8):845-847. [doi: 10.1016/j.jaac.2024.03.012] [Medline: 39069372]
30. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989;13(3):319. [doi: 10.2307/249008]
31. Ericsson A, Simon HA. Protocol Analysis: Verbal Reports as Data. Cambridge, Massachusetts: MIT Press; 1993.
32. Davison GC, Vogel RS, Coffman SG. Think-aloud approaches to cognitive assessment and the articulated thoughts in simulated situations paradigm. J Consult Clin Psychol 1997;65(6):950-958. [doi: 10.1037//0022-006x.65.6.950] [Medline: 9420356]
33. Guest G, MacQueen KM, Namey EE. Applied Thematic Analysis. Thousand Oaks, California: SAGE Publications; 2012.
34. Averill JB. Matrix analysis as a complementary analytic strategy in qualitative inquiry. Qual Health Res 2002;12(6):855-866. [doi: 10.1177/10432302012006011]
35. Beebe JR. Rapid Qualitative Inquiry: A Field Guide to Team-Based Assessment. Lanham, Maryland: Rowman & Littlefield; 2014.
36. Gan DZQ, McGillivray L, Larsen ME, Bloomfield T, Torok M. Promoting engagement with smartphone apps for suicidal ideation in young people: development of an adjunctive strategy using a lived experience participatory design approach. JMIR Form Res 2023;7:e45234 [FREE Full text] [doi: 10.2196/45234] [Medline: 37279058]
37. Grist R, Porter J, Stallard P. Acceptability, use, and safety of a mobile phone app (BlueIce) for young people who self-harm: qualitative study of service users' experience. JMIR Ment Health 2018;5(1):e16 [FREE Full text] [doi: 10.2196/mental.8779] [Medline: 29475823]
38. Hetrick SE, Robinson J, Burge E, Blandon R, Mobilio B, Rice SM, et al. Youth codesign of a mobile phone app to facilitate self-monitoring and management of mood symptoms in young people with major depression, suicidal ideation, and self-harm. JMIR Ment Health 2018;5(1):e9 [FREE Full text] [doi: 10.2196/mental.9041] [Medline: 29362208]
39. Graham AK, Lattie EG, Powell BJ, Lyon AR, Smith JD, Schueller SM, et al. Implementation strategies for digital mental health interventions in health care settings. Am Psychol 2020;75(8):1080-1092 [FREE Full text] [doi: 10.1037/amp0000686] [Medline: 33252946]
40. van Mierlo T, Rondina R, Fournier R. Nudges and prompts increase engagement in self-guided digital health treatment for depression and anxiety: results from a 3-arm randomized controlled trial. JMIR Form Res 2024;8:e52558 [FREE Full text] [doi: 10.2196/52558] [Medline: 38592752]
41. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]
42. Bishop TF, Press MJ, Mendelsohn JL, Casalino LP. Electronic communication improves access, but barriers to its widespread adoption remain. Health Aff (Millwood) 2013;32(8):1361-1367 [FREE Full text] [doi: 10.1377/hlthaff.2012.1151] [Medline: 23918479]
43. Armstrong CM. Mobile health provider training: results and lessons learned from year four of training on core competencies for mobile health in clinical care. J. technol. behav. sci 2019;4(2):86-92. [doi: 10.1007/s41347-019-00089-8]
44. Graham A, Greene C, Powell T, Lieponis P, Lunsford A, Peralta CD, et al. Lessons learned from service design of a trial of a digital mental health service: informing implementation in primary care clinics. Transl Behav Med 2020;10(3):598-605 [FREE Full text] [doi: 10.1093/tbm/ibz140] [Medline: 32766862]
45. Bessenyei K, Suruliraj B, Bagnell A, McGrath P, Wozney L, Huguet A, et al. Comfortability with the passive collection of smartphone data for monitoring of mental health: an online survey. Computers in Human Behavior Reports 2021;4:100134. [doi: 10.1016/j.chbr.2021.100134]
46. Wang J, Qiu Y, Zhu X. Trends of mental health care utilization among US adults from 1999 to 2018. BMC Psychiatry 2023;23(1):665 [FREE Full text] [doi: 10.1186/s12888-023-05156-2] [Medline: 37700243]
47. Data tool: demographics of the U.S. psychology workforce. American Psychological Association. 2022. URL: https://www.apa.org/workforce/data-tools/demographics [accessed 2024-05-09]
48. Creswell JW. Concise Introduction to Mixed Methods Research. Thousand Oaks, California: Sage; 2022.

XSL•FO
RenderX

49. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022;292:114523 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]

Abbreviations

DBT: dialectical behavior therapy
EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act
IOP: intensive outpatient program
RCT: randomized controlled trial
TAM: technology acceptance model
UCD: user-centered design

Edited by A Mavragani; submitted 20.08.24; peer-reviewed by L Magoun, M Klapow; comments to author 15.12.24; revised version received 23.12.24; accepted 07.01.25; published 02.05.25.

Please cite as:
Tse TC, Weiner LS, Funkhouser CJ, DeLuise D, Cullen C, Blumkin Z, O'Brien C, Auerbach RP, Allen NB
Acceptability and Usability of a Digital Behavioral Health Platform for Youth at Risk of Suicide: User-Centered Design Study With Patients, Practitioners, and Business Gatekeepers
JMIR Form Res 2025;9:e65418
URL: https://formative.jmir.org/2025/1/e65418
doi:10.2196/65418
PMID:

©Trinity Chloe Tse, Lauren S Weiner, Carter J Funkhouser, Danielle DeLuise, Colleen Cullen, Zachary Blumkin, Casey O'Brien, Randy P Auerbach, Nicholas B Allen. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.05.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Assessing the Causal Association Between COVID-19 and Graves Disease: Mendelian Randomization Study

Hui Nian[1*], MM; Yu Bai[2*], MM; Hua Yu[3], PhD

[1]Department of Thoracic Surgery, Shanghai Xuhui Central Hospital, Shanghai, China

[2]Department of Intensive Care Unit, Shanghai Xuhui Central Hospital, Shanghai, China

[3]Department of Plastic and Reconstructive Surgery, Shanghai Fourth People's Hospital Affiliated to Tongji University School of Medicine, No. 1279 Sanmen Road, Hongkou District, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Hua Yu, PhD
Department of Plastic and Reconstructive Surgery, Shanghai Fourth People's Hospital Affiliated to Tongji University School of Medicine, No. 1279 Sanmen Road, Hongkou District, Shanghai, China

Abstract

Background: Graves disease (GD) is an autoimmune thyroid disorder characterized by hyperthyroidism and autoantibodies. The COVID-19 pandemic has raised questions about its potential relationship with autoimmune diseases like GD.

Objective: This study aims to investigate the causal association between COVID-19 and GD through Mendelian randomization (MR) analysis and assess the impact of COVID-19 on GD.

Methods: We conducted an MR study using extensive genome-wide association study data for GD and COVID-19 susceptibility and its severity. We used stringent single nucleotide polymorphism selection criteria and various MR methodologies, including inverse-variance weighting, MR-Egger, and weighted median analyses, to assess causal relationships. We also conducted tests for directional pleiotropy and heterogeneity, as well as sensitivity analyses.

Results: The MR analysis, based on the largest available dataset to date, did not provide evidence supporting a causal relationship between COVID-19 susceptibility (odds ratio [OR] 0.989, 95% CI 0.405 - 2.851; *P*=.93), COVID-19 hospitalization (OR 0.974, 95% CI 0.852 - 1.113; *P*=.70), COVID-19 severity (OR 0.979, 95% CI 0.890 - 1.077; *P*=.66), and GD. Tests for directional pleiotropy and heterogeneity, as well as sensitivity analyses, supported these findings.

Conclusions: This comprehensive MR study does not provide sufficient evidence to support a causal relationship between COVID-19 and the onset or exacerbation of GD. These results contribute to a better understanding of the potential association between COVID-19 and autoimmune diseases, alleviating concerns about a surge in autoimmune thyroid diseases due to the pandemic. Further research is warranted to explore this complex relationship thoroughly.

(JMIR Form Res 2025;9:e66003) doi:10.2196/66003

KEYWORDS

Graves disease; COVID-19; Mendelian randomization; causal relationship; autoimmune disease

Introduction

Graves disease (GD) is an autoimmune thyroid disorder characterized by hyperthyroidism and the presence of autoantibodies. It is influenced by a combination of genetic, environmental, and immunological factors. Viral infections have long been implicated as potential triggers of autoimmune diseases, including autoimmune thyroid disorders [1-3]. Notably, the global COVID-19 pandemic has intensified concerns regarding its potential role in the development or exacerbation of autoimmune diseases such as GD [1,2].

Previous studies have explored the associations between viral infections and autoimmune thyroid diseases, with evidence suggesting that infections such as Epstein-Barr virus and hepatitis C virus may contribute to the onset or progression of these disorders [4,5]. More recently, research has focused on the potential immunological impact of SARS-CoV-2, the virus responsible for COVID-19. It has been proposed that SARS-CoV-2 can disrupt immune homeostasis, potentially leading to autoimmune reactions, including thyroid dysfunction [6]. While observational studies have reported an increased incidence of thyroid abnormalities in COVID-19 patients, these findings remain inconclusive in establishing a direct causal link between COVID-19 and autoimmune thyroid diseases [7-9].

The impact of COVID-19 on thyroid function has been further highlighted in recent systematic reviews and meta-analyses, which have suggested that thyroid abnormalities may be a common sequela of SARS-CoV-2 infection [10]. Additionally,

XSL•FO
RenderX

genetic predisposition plays a crucial role in determining susceptibility to autoimmune thyroid diseases, which underscores the need for a robust methodological approach to examine causality.

To address these uncertainties, this study uses Mendelian randomization (MR) analysis to investigate the causal relationship between COVID-19 and GD. MR leverages genetic variants as instrumental variables to approximate the effects of a randomized controlled trial, thereby minimizing confounding biases. This approach is based on 3 key assumptions: (1) genetic variants are not associated with confounding factors, (2) genetic variants influence an intermediate exposure variable, and (3) genetic variants affect the outcome solely through their impact on the exposure variable [11].

By elucidating the potential causal association between COVID-19 and GD, this research aims to contribute to a deeper understanding of the impact of SARS-CoV-2 on autoimmune thyroid diseases. These findings could inform future studies and clinical strategies for managing autoimmune thyroid disorders in the context of COVID-19.

Methods

Study Design

This study used a 2-sample MR design, relying exclusively on publicly accessible data for all analyses. We meticulously adhered to the recommendations presented in the Strengthening the Reporting of Observational Studies in Epidemiology-Mendelian Randomization (STROBE-MR) guidelines, ensuring a comprehensive and rigorous reporting approach.

Ethical Considerations

This study is a secondary analysis of publicly available genome-wide association study (GWAS) summary statistics obtained from the Integrative Epidemiology Unit OpenGWAS database [12]. As no new individual-level data were collected and no direct human participation was involved, ethical approval from an institutional review board was not required. The original GWASs adhered to their respective ethical guidelines and received necessary approvals. Since the study exclusively used deidentified, publicly available data, informed consent was not applicable; however, the original studies obtained informed consent from participants at the time of data collection and explicitly permitted secondary data analysis without additional consent. All data used in this study were fully anonymized, ensuring participant confidentiality, and no personally identifiable information was accessed, stored, or analyzed. In accordance with data protection regulations and ethical standards, strict measures were taken to maintain privacy and confidentiality. As no human participants were directly involved, no compensation was provided. Additionally, this study does not include any identifiable images or figures.

Data Sources for Exposures

We procured summary-level data for COVID-19 susceptibility (n=1,683,768), COVID-19 hospitalization (n=1,882,773), and COVID-19 severity (n=1,388,342) from the Integrative Epidemiology Unit OpenGWAS database [12]. This database offers extensive GWAS summary statistics encompassing a diverse array of traits and outcomes.

The data on COVID-19 susceptibility included 38,984 individuals, comprising individuals diagnosed with COVID-19 through laboratory confirmation of SARS-CoV-2 infection, electronic health records using *International Classification of Diseases* (*ICD*) codes or physician notes, or self-reported cases. For COVID-19 hospitalization, 9986 individuals were included, while COVID-19 severity encompassed 5101 individuals categorized as being in an extremely critical condition, which included individuals who experienced mortality or required advanced respiratory support [13-15].

Detailed information regarding the genetic variants associated with these COVID-19-related outcomes is presented in 3/7/2025 [16], offering a comprehensive overview of the exposure variables used in our MR analysis.

Outcome Data

The outcome data for GD were extracted from publicly available summary statistics, involving a total of 2176 GD cases and 210,277 controls (GWAS ID: bbj-a-123). The cases represented individuals hospitalized for GD, and the control group consisted of individuals from cohorts without identified cases. The genetic instruments used for GD in this MR study were derived from these data, facilitating a comprehensive evaluation of the causal relationship between GD and pertinent genetic variations.

Instrument Construction

For constructing genetic instruments for each exposure, we used a rigorous approach. We identified single nucleotide polymorphisms (SNPs) displaying significant associations with the respective exposure at a genome-wide significance threshold ($P<5\times10^{-}$), which, while less stringent than the conventional threshold ($P<5\times10^{-}$), was chosen to balance instrument strength and the number of available SNPs. These SNPs underwent linkage disequilibrium clumping to remove highly correlated SNPs ($r^2\geq0.001$). Our reference population was drawn from the European cohort within the 1000 Genomes Project, which may limit the generalizability of our findings to non-European populations. To ensure consistency in effect sizes and address issues related to palindromic SNPs, we harmonized SNP-exposure and SNP-GD associations. In cases where specific SNPs were missing in the GD GWAS dataset, we used proxy SNPs ($r^2>0.8$). To mitigate potential bias from weak instruments, we performed *F* statistic calculations and conducted multiple sensitivity analyses, including leave-one-out analysis and MR-Egger regression. The processes of SNP extraction from GWAS summary-level data, linkage disequilibrium clumping, and harmonization were executed using the TwoSample MR package in R version 4.3.0 (R Core Team) [17,18]. The strength of the instrumental variables was evaluated using the F statistic, computed as follows: $F = (R^2/K)/[(1-R^2)(N-K-1)]$, where R^2 denotes the variance explained by the instruments, K is the number of instruments, and N is the sample size [19].

XSL•FO
RenderX

Statistical Analyses

Using a 2-sample MR approach, we estimated the causal effects of COVID-19 susceptibility and its severity on GD. The primary MR analysis used the inverse-variance weighting (IVW) method, assuming that the genetic instruments collectively satisfy the core MR assumptions. In addition to the primary MR analysis, we conducted supplementary sensitivity analyses, including MR-Egger, weighted median, and MR-pleiotropy residual sum and outlier (MR-PRESSO) methods. These methods enhance robustness against potential pleiotropic effects by relaxing certain MR assumptions. To assess horizontal pleiotropy, we used the intercept term from MR-Egger regression, where an intercept *P* value exceeding .05 indicated an absence of pleiotropy. Furthermore, we applied MR-PRESSO as a supplementary method to detect and exclude outlier genetic instruments with pleiotropic effects, subsequently estimating causal effects. All tests were 2-sided and executed using the TwoSampleMR and MR-PRESSO packages in R version 4.3.0. A significance threshold was set at $P<.05$. Since our analysis exclusively used publicly available data, no ethical approval from a local committee was sought.

Results

In summary, this MR analysis encompassed GWAS data for 3 COVID-19 statuses, including COVID-19 susceptibility and its severity, in relation to GD (detailed information provided in Table 1).

Regarding COVID-19 susceptibility, the IVW MR analysis indicated that GD did not significantly increase the risk (odds ratio [OR] 0.989, 95% CI 0.405 - 2.851; *P*=.93; Figures 1A-3A and Table 2). No heterogeneity among SNPs was observed (Cochran Q *P*=.39). The MR-Egger intercept was close to zero (intercept=−0.0051; *P*=.85), implying the absence of directional pleiotropic effects (Table 3). MR-PRESSO did not detect any outlier SNPs (*P*=.47). The results of the leave-one-out sensitivity analysis supported the observed causal association (Figure 4A).

The results of the IVW MR analysis indicated that there was no significant association between COVID-19 hospitalization and the risk of GD (OR 0.974, 95% CI 0.852 - 1.113; *P*=.70; Figures 1B-3B and Table 2). Heterogeneity was not observed among the various studies (*P*=.27). Sensitivity analysis using the MR-PRESSO method also failed to yield statistically significant causal estimates (*P*=.33). Furthermore, the MR-Egger intercept suggested an absence of horizontal pleiotropy (intercept=−0.0039; *P*=.89). The MR-PRESSO method did not identify any outlier SNPs (*P*=.33; Table 3). Leave-one-out sensitivity analysis did not reveal any individual SNPs that significantly influenced the causal estimates (Figure 4B).

In the primary MR analysis using the IVW method, there was no significant association observed between COVID-19 severity and the risk of GD. The OR for this relationship was 0.979 (95% CI 0.890 - 1.077; *P*=.66; Figures 1C-3C and Table 2). The results were consistent when using the weighted median method (OR 0.975, 95% CI 0.852 - 1.116; *P*=.71) and conducting the MR-PRESSO sensitivity analysis (*P*=.20). Furthermore, the Cochran Q test revealed no evidence of heterogeneity (*P*=.20), and the MR-Egger intercept indicated the absence of horizontal pleiotropy (intercept=0.0352; *P*=.37). No potential outliers were identified using the MR-PRESSO method (Table 3). Leave-one-out sensitivity analysis did not reveal any individual SNPs that significantly influenced the causal estimates (Figure 4C). The symmetry in the funnel plot further supported the reliability of the MR analysis (Figure 3).

Table . Data sources for the analysis.

Phenotype and consortium	Source of genetic variations	
	Cases	Controls
Graves disease	2176 individuals hospitalized for Graves disease	210,277 individuals from cohorts without identified cases
COVID-19 susceptibility and its severity		
Susceptibility	38,984 individuals who were diagnosed with COVID-19 using laboratory confirmation of SARS-CoV-2 infection, electronic health records (which included the use of *ICD*[a] codes or physician notes), or self-reported cases	1,644,784 individuals who were registered within the same groups but were excluded from the cases under investigation
Hospitalization	9986 individuals who were admitted to hospitals as a result of contracting COVID-19	1,877,672 individuals who were enrolled in the study cohorts but did not meet the criteria for classification as cases, indicating that they did not require hospitalization due to COVID-19
Very severe	5101 individuals categorized as being in an extremely critical condition, which encompasses those who experienced mortality or required advanced respiratory support, including interventions such as continuous positive airway pressure (CPAP), bilevel positive airway pressure (BiPAP), endotracheal intubation, or high-flow nasal cannula therapy	1,383,241 individuals who were included in the study cohorts but did not meet the criteria for classification as cases

[a]*ICD*: *International Classification of Diseases.*

XSL•FO
RenderX

Figure 1. Forest plot of the results of Mendelian randomization (MR) analysis. (A) COVID-19 susceptibility, (B) COVID-19 hospitalization, and (C) COVID-19 severity.

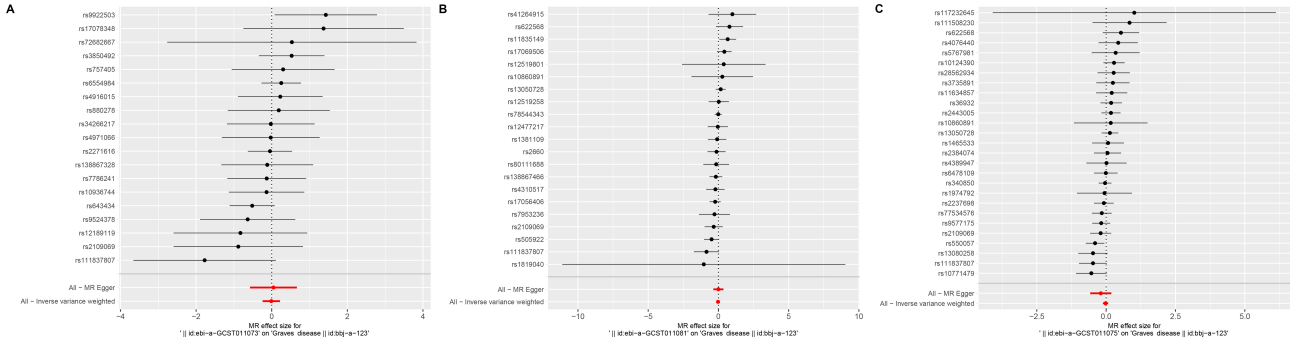


Figure 2. Scatter plot of the results of Mendelian randomization (MR) analysis. (A) COVID-19 susceptibility, (B) COVID-19 hospitalization, and (C) COVID-19 severity. SNP: single nucleotide polymorphism.

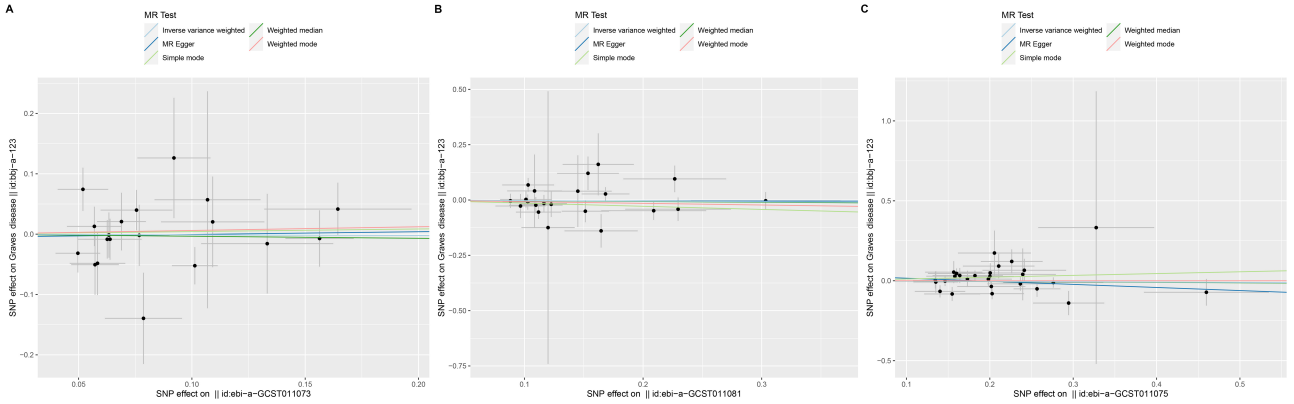
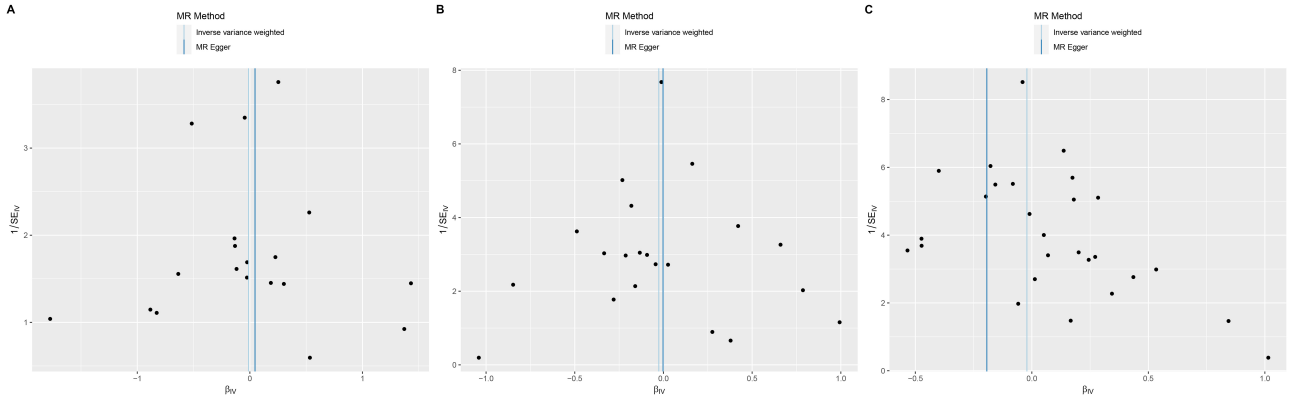


Figure 3. Funnel plot of the results of Mendelian randomization (MR) analysis. (A) COVID-19 susceptibility, (B) COVID-19 hospitalization, and (C) COVID-19 severity.



XSL•FO
RenderX

Table . Mendelian randomization estimates of the association between Graves disease and COVID-19.

Outcome	COVID-19 susceptibility			COVID-19 hospitalization			COVID-19 very severe		
	SNPs[a], n	OR[b] (95% CI)	*P* value	SNPs, n	OR (95% CI)	*P* value	SNPs, n	OR (95% CI)	*P* value
Graves disease									
IVW[c]	19	0.989 (0.405 - 2.851)	.93	21	0.974 (0.852 - 1.113)	.70	27	0.979 (0.890 - 1.077)	.66
Weighted median	19	0.966 (0.685 - 1.362)	.84	21	0.967 (0.797 - 1.175)	.74	27	0.975 (0.852 - 1.116)	.71
MR-Egger[d]	19	1.047 (0.786 - 1.245)	.89	21	0.997 (0.696 - 1.431)	.99	27	0.824 (0.563 - 1.207)	.33
Simple mode	19	1.043 (0.632 - 1.723)	.87	21	0.868 (0.655 - 1.151)	.34	27	1.119 (0.871 - 1.436)	.39

[a]SNP: single nucleotide polymorphism.

[b]OR: odds ratio.

[c]IVW: inverse-variance weighting.

[d]MR-Egger: Mendelian randomization-Egger regression.

Table . Results of heterogeneity and horizontal pleiotropy analyses.

Outcome	Exposure	*P* value (heterogeneity)[a]	MR-Egger[b] intercept	*P* value (pleiotropy)[c]	*P* value (global test)[d]
Graves disease	COVID-19 susceptibility	.39	−0.0051	.85	.47
Graves disease	COVID-19 hospitalization	.27	−0.0039	.89	.33
Graves disease	COVID-19 very severe	.20	0.0352	.37	.20

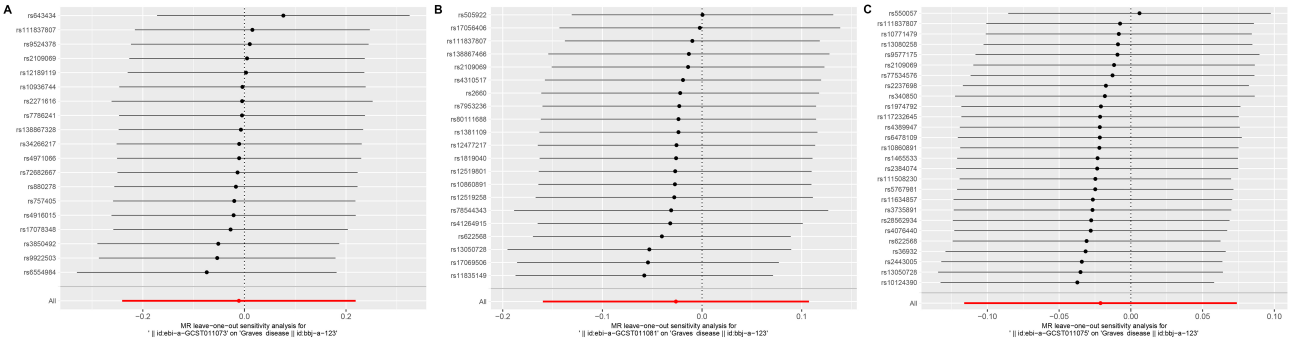
[a]The *P* value of Cochran Q value in the heterogeneity test.

[b]MR-Egger: Mendelian randomization-Egger.

[c]The *P* value for the intercept in the MR-Egger regression.

[d]The *P* value for the global test in the Mendelian randomization-pleiotropy residual sum and outlier test.

Figure 4. Leave-one-out analysis plots of the results of Mendelian randomization (MR) analysis. (A) COVID-19 susceptibility, (B) COVID-19 hospitalization, and (C) COVID-19 severity.



Discussion

Principal Findings

In this MR study, we extensively explored the potential causal relationships between COVID-19 susceptibility, COVID-19 hospitalization, and COVID-19 severity in relation to GD. Using a variety of MR methods, we aimed to provide a comprehensive understanding of these relationships. Our study results underscore the absence of a causal link between COVID-19 susceptibility and GD. Recent research continues to unveil the intricate interplay between autoimmune diseases and viral infections. For example, a study conducted by Widhani et al [20] delved into the potential associations between viral infections, including COVID-19, and autoimmune thyroid diseases, further corroborating our findings. Additionally, the work of Smith and Hegedüs [21] has provided in-depth insights into the pathogenesis of GD, emphasizing its multifactorial nature. These recent studies lend robust support to our research. Furthermore, investigations into the relationship between COVID-19 and autoimmune diseases have expanded in recent

XSL•FO
RenderX

years. A systematic review by Tutal et al [22] examined the impact of autoimmune thyroid diseases on the severity of COVID-19, enriching our understanding of the bidirectional influences between these health conditions. To provide a broader clinical context, we have integrated evidence on how COVID-19 susceptibility and outcomes differ across other autoimmune diseases, which highlights the need for more nuanced investigations beyond GD alone [23].

Similarly, we found that COVID-19 hospitalization does not significantly increase the risk of GD. The consistency of this result across various MR methods suggests that COVID-19 hospitalization is not a causal risk factor for GD. This conclusion is pivotal for understanding the etiology of GD and its relationship with other health conditions. Recent research has provided further insights into the connection between COVID-19 and autoimmune thyroid diseases. For example, a study by Fallahi et al [24], conducted within the past 5 years, investigated the impact of COVID-19 hospitalization on the development of autoimmune thyroid diseases, including GD. The study used a large-scale patient database and used statistical methods to evaluate the risk of developing autoimmune thyroid diseases after COVID-19 hospitalization. Their findings, which align with our MR results, indicated that COVID-19 hospitalization was not significantly associated with an increased risk of GD or other autoimmune thyroid diseases. This further supports the critical importance of recognizing that COVID-19 hospitalization itself does not have a direct causal role in the development of GD. Furthermore, we examined the relationship between COVID-19 severity and the risk of developing GD. Our study revealed that COVID-19 severity does not contribute to the onset of GD. This finding provides important insights into the genetic and environmental factors influencing GD's pathogenesis. A recent study by Bostan and colleagues [25] explored the genetic mechanisms underlying the development of GD in individuals with severe COVID-19. Their research highlighted the role of genetic susceptibility in the development of GD among those with severe COVID-19, adding to our understanding of the complex interaction between COVID-19 and autoimmune thyroid diseases.

Moreover, recent research has extensively explored the associations between COVID-19 and thyroid-related conditions. Studies have indicated that COVID-19 infection can influence thyroid function, potentially leading to hyperthyroidism or thyrotoxicosis in certain individuals, further emphasizing the potential link between COVID-19 and thyroid diseases. However, the specific mechanisms remain unclear and warrant further investigation. Recent studies have provided more in-depth insights into the complex interplay between COVID-19 and thyroid function. For instance, a study by Smith et al [21] identified COVID-19-related thyroid dysfunction, including subacute thyroiditis and hyperthyroidism, reporting cases of patients experiencing thyroid-related symptoms during or following COVID-19 infection. Similarly, Damara et al [10] conducted a systematic review and meta-analysis, revealing a significant association between autoimmune thyroid diseases and the severity of COVID-19. These studies underscore the importance of further understanding the relationship between COVID-19 and thyroid dysfunction. Additionally, Muller et al [26] examined the underlying mechanisms of COVID-19-induced thyroid dysfunction, particularly focusing on the immune response's role in thyroid autoimmunity. Their findings suggest that the cytokine storm triggered by COVID-19 may play a crucial role in the development of thyroid diseases. Liu et al [27] investigated the impact of SARS-CoV-2 infection on the expression of thyroid-related genes in thyroid tissue, shedding light on the potential molecular pathways involved. To broaden the context, we have also incorporated recent research on COVID-19 vaccine immunogenicity in patients with autoimmune diseases, providing additional understanding of how immune responses to COVID-19 vaccination may differ in individuals with pre-existing autoimmune conditions [28].

Furthermore, recent investigations have explored the complex interplay between GD and other thyroid-related conditions in the context of COVID-19. For example, Nakano et al [29] found that individuals with GD may be at an increased risk of contracting COVID-19 and could face more severe outcomes postinfection, including complications such as thyrotoxicosis. This study highlights the bidirectional relationship between GD and COVID-19, emphasizing the importance of comprehensive research to better understand the complex interactions between thyroid-related conditions and COVID-19. In a study by Sethi et al [30], researchers observed that COVID-19 infection was associated with thyroid abnormalities, with some patients developing thyrotoxicosis. This suggests that COVID-19 may influence thyroid function, raising concerns about potential implications for individuals with GD. Additionally, Croce et al [31] examined the bidirectional effects of GD and COVID-19, proposing mechanisms through which GD could increase susceptibility to COVID-19 and worsen clinical outcomes, such as immune system dysregulation and cytokine imbalances. However, both COVID-19 and GD can significantly impact physical and mental health, and thus, we have integrated insights from recent research on physical activity and mental health interventions, which may provide potential strategies for improving long-term patient outcomes [32]. Collectively, these studies underscore the urgent need for further research into the multifaceted interactions between COVID-19 and thyroid-related diseases, including GD. This will not only deepen our understanding of the pathogenesis of these conditions but also offer critical insights for health care practitioners, aiding in the development of more informed and effective clinical strategies.

Limitations

It is important to note that despite the strengths of the MR approach, our study has several limitations. MR methods rely on certain assumptions, including the absence of horizontal pleiotropy. Although we used various strategies to address potential pleiotropic effects, we cannot entirely exclude the possibility of unmeasured confounders influencing our results. Furthermore, our analysis primarily relied on data from the European population, which may limit the generalizability of our findings to other ethnic groups. Specifically, the wide CIs for COVID-19 susceptibility suggest that the study may be underpowered to detect modest effects. Additionally, the definition of COVID-19 cases included self-reported cases, which could introduce reporting bias. The combination of laboratory-confirmed, electronic health record–derived, and

XSL•FO
RenderX

self-reported cases may also affect the reliability of exposure classification. However, sensitivity analyses on subsets of laboratory-confirmed and electronic health record–derived cases support the robustness of our findings. Similarly, the definition of GD cases was restricted to hospitalized patients, which could result in selection bias, underrepresenting mild or moderate cases. While hospitalization-based diagnoses enhance diagnostic accuracy, future research should incorporate outpatient and primary care datasets to capture the full spectrum of disease severity. Lastly, this study did not fully explore the potential bidirectional relationships or temporal dynamics between COVID-19 and GD. Future studies should address these aspects using bidirectional MR and longitudinal cohort studies to further clarify the interactions between these conditions.

Conclusions

In summary, our MR analysis did not identify significant causal relationships between COVID-19 susceptibility, COVID-19 hospitalization, or COVID-19 severity and the risk of developing GD. These findings highlight the complex interactions between viral infections, autoimmune diseases, and genetic factors. However, they also emphasize the need for further investigation into the bidirectional relationships and temporal dynamics between COVID-19 and thyroid-related conditions. Expanding this research will deepen our understanding of the pathogenesis of both diseases, contributing valuable insights for clinical decision-making and management strategies.

Acknowledgments

The authors express their sincere gratitude to the Integrative Epidemiology Unit OpenGWAS Project [12] for granting access to the genome-wide association study (GWAS) summary-level data used in this research. This project is supported by the Talent Boosting Program (SY-XKZT-2020-3007) and the Discipline Boosting Program (SY-XKZT-2020-1021) of Shanghai Fourth People's Hospital.

Data Availability

The analysis in this research relied on publicly available datasets, which can be accessed through the Integrative Epidemiology Unit OpenGWAS Project [12].

Authors' Contributions

HN analyzed the data and wrote the manuscript. YB designed the study. HY collected the data and corrected the paper. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Chen K, Gao Y, Li J. New-onset and relapsed Graves' disease following COVID-19 vaccination: a comprehensive review of reported cases. Eur J Med Res 2023 Jul 13;28(1):232. [doi: 10.1186/s40001-023-01210-7] [Medline: 37443067]
2. Murugan AK, Alzahrani AS. SARS-CoV-2 plays a pivotal role in inducing hyperthyroidism of Graves' disease. Endocrine 2021 Aug;73(2):243-254. [doi: 10.1007/s12020-021-02770-6] [Medline: 34106438]
3. Almutairi H, Alqadi FS, Alsulaim RK, et al. Unveiling promising modalities and enhancing patient outcomes in Graves' disease treatment: a systematic review and meta-analysis. Cureus 2024 May;16(5):e60829. [doi: 10.7759/cureus.60829] [Medline: 38910658]
4. Msheik AN, Al Mokdad Z, Hamed F, et al. Epstein-Barr virus flare: a multiple sclerosis attack. Surg Neurol Int 2024;15:355. [doi: 10.25259/SNI_457_2024] [Medline: 39524581]
5. Cyna W, Wojciechowska A, Szybiak-Skora W, Lacka K. The impact of environmental factors on the development of autoimmune thyroiditis-review. Biomedicines 2024 Aug 7;12(8):1788. [doi: 10.3390/biomedicines12081788] [Medline: 39200253]
6. Mobasheri L, Nasirpour MH, Masoumi E, Azarnaminy AF, Jafari M, Esmaeili SA. SARS-CoV-2 triggering autoimmune diseases. Cytokine 2022 Jun;154:155873. [doi: 10.1016/j.cyto.2022.155873] [Medline: 35461172]
7. Naguib R. Potential relationships between COVID-19 and the thyroid gland: an update. J Int Med Res 2022 Feb;50(2). [doi: 10.1177/03000605221082898] [Medline: 35226548]
8. Darvishi M, Nazer MR, Shahali H, Nouri M. Association of thyroid dysfunction and COVID-19: a systematic review and meta-analysis. Front Endocrinol (Lausanne) 2022;13:947594. [doi: 10.3389/fendo.2022.947594] [Medline: 36387848]
9. Rossetti CL, Cazarin J, Hecht F, et al. COVID-19 and thyroid function: what do we know so far? Front Endocrinol (Lausanne) 2022;13:1041676. [doi: 10.3389/fendo.2022.1041676] [Medline: 36601011]
10. Damara FA, Muchamad GR, Ikhsani R, Syafiyah AH, Bashari MH. Thyroid disease and hypothyroidism are associated with poor COVID-19 outcomes: a systematic review, meta-analysis, and meta-regression. Diabetes Metab Syndr 2021;15(6):102312. [doi: 10.1016/j.dsx.2021.102312] [Medline: 34731819]

XSL•FO
RenderX

11. Sanderson E, Glymour MM, Holmes MV, et al. Mendelian randomization. Nat Rev Methods Primers 2022 Feb 10;2:6. [doi: 10.1038/s43586-021-00092-5] [Medline: 37325194]
12. GWAS summary data. MRC Integrative Epidemiology Unit. URL: https://gwas.mrcieu.ac.uk/ [accessed 2025-03-28]
13. COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. Eur J Hum Genet 2020 Jun;28(6):715-718. [doi: 10.1038/s41431-020-0636-6] [Medline: 32404885]
14. Zhang Z, Fang T, Chen L, Qiu B, Lv Y. Thyroid function and COVID-19 susceptibility and its severity: a two-sample Mendelian randomization study. Endocrinology 2022 Oct 11;163(11):bqac139. [doi: 10.1210/endocr/bqac139] [Medline: 36048436]
15. Köhler A, Chen B, Gemignani F, et al. Genome-wide association study on differentiated thyroid cancer. The Journal of Clinical Endocrinology & Metabolism 2013 Oct;98(10):E1674-E1681. [doi: 10.1210/jc.2013-1941]
16. IEU OpenGWAS project. URL: https://gwas.mrcieu.ac.uk/ [accessed 2025-04-07]
17. Hu X, Zhao J, Lin Z, et al. Mendelian randomization for causal inference accounting for pleiotropy and sample structure using genome-wide summary statistics. Proc Natl Acad Sci U S A 2022 Jul 12;119(28):e2106858119. [doi: 10.1073/pnas.2106858119] [Medline: 35787050]
18. Zhu X, Li X, Xu R, Wang T. An iterative approach to detect pleiotropy and perform Mendelian Randomization analysis using GWAS summary statistics. Bioinformatics 2021 Jun 16;37(10):1390-1400. [doi: 10.1093/bioinformatics/btaa985] [Medline: 33226062]
19. Zhao SS, Holmes MV, Zheng J, Sanderson E, Carter AR. The impact of education inequality on rheumatoid arthritis risk is mediated by smoking and body mass index: Mendelian randomization study. Rheumatology (Oxford) 2022 May 5;61(5):2167-2175. [doi: 10.1093/rheumatology/keab654] [Medline: 34436562]
20. Widhani A, Koesnoe S, Maria S, et al. Factors related to severity, hospitalization, and mortality of COVID-19 infection among patients with autoimmune diseases. Trop Med Infect Dis 2023 Apr 18;8(4):227. [doi: 10.3390/tropicalmed8040227] [Medline: 37104352]
21. Smith TJ, Hegedüs L. Graves' disease. N Engl J Med 2016 Oct 20;375(16):1552-1565. [doi: 10.1056/NEJMra1510030] [Medline: 27797318]
22. Tutal E, Ozaras R, Leblebicioglu H. Systematic review of COVID-19 and autoimmune thyroiditis. Travel Med Infect Dis 2022;47:102314. [doi: 10.1016/j.tmaid.2022.102314] [Medline: 35307540]
23. Shin JI, Kim SE, Lee MH, et al. COVID-19 susceptibility and clinical outcomes in autoimmune inflammatory rheumatic diseases (AIRDs): a systematic review and meta-analysis. Eur Rev Med Pharmacol Sci 2022 May;26(10):3760-3770. [doi: 10.26355/eurrev_202205_28873] [Medline: 35647859]
24. Fallahi P, Ferrari SM, Elia G, et al. Thyroid autoimmunity and SARS-CoV-2 infection: report of a large Italian series. Autoimmun Rev 2022 Nov;21(11):103183. [doi: 10.1016/j.autrev.2022.103183] [Medline: 36007802]
25. Bostan H, Ucan B, Kizilgul M, et al. Relapsed and newly diagnosed Graves' disease due to immunization against COVID-19: a case series and review of the literature. J Autoimmun 2022 Apr;128:102809. [doi: 10.1016/j.jaut.2022.102809] [Medline: 35220164]
26. Muller I, Cannavaro D, Dazzi D, et al. SARS-CoV-2-related atypical thyroiditis. Lancet Diabetes Endocrinol 2020 Sep;8(9):739-741. [doi: 10.1016/S2213-8587(20)30266-7] [Medline: 32738929]
27. Liu H, Xin J, Cai S, Jiang X. Mendelian randomization analysis provides causality of smoking on the expression of ACE2, a putative SARS-CoV-2 receptor. Elife 2021 Jul 6;10:e64188. [doi: 10.7554/eLife.64188] [Medline: 34227468]
28. Cho K, Park S, Kim EY, et al. Immunogenicity of COVID-19 vaccines in patients with diverse health conditions: a comprehensive systematic review. J Med Virol 2022 Sep;94(9):4144-4155. [doi: 10.1002/jmv.27828] [Medline: 35567325]
29. Nakano Y, Takeshima K, Furukawa Y, Morita S, Sakata M, Matsuoka TA. Concomitant exacerbation of Graves orbitopathy and double-seronegative Myasthenia Gravis after SARS-CoV-2 Infection. JCEM Case Rep 2025 Feb;3(2):luaf019. [doi: 10.1210/jcemcr/luaf019] [Medline: 39882352]
30. Sethi Y, Uniyal N, Maheshwari S, Sinha R, Goel A. Thyroid function abnormalities in the acute phase of COVID-19: a cross-sectional hospital-based study from North India. Cureus 2022 May;14(5):e24942. [doi: 10.7759/cureus.24942] [Medline: 35706733]
31. Croce L, Gangemi D, Ancona G, et al. The cytokine storm and thyroid hormone changes in COVID-19. J Endocrinol Invest 2021 May;44(5):891-904. [doi: 10.1007/s40618-021-01506-7] [Medline: 33559848]
32. Rahmati M, Lee S, Yon DK, et al. Physical activity and prevention of mental health complications: an umbrella review. Neurosci Biobehav Rev 2024 May;160:105641. [doi: 10.1016/j.neubiorev.2024.105641] [Medline: 38527637]

Abbreviations

GD: Graves disease
GWAS: genome-wide association study
***ICD*:** *International Classification of Diseases*
IVW: inverse-variance weighting

XSL•FO
RenderX

MR: Mendelian randomization
MR-PRESSO: Mendelian randomizaion-pleiotropy residual sum and outlier
OR: odds ratio
SNP: single nucleotide polymorphism
STROBE-MR: Strengthening the Reporting of Observational Studies in Epidemiology-Mendelian Randomization

Edited by A Mavragani; submitted 01.09.24; peer-reviewed by M Chen, Q Du, SW Lee; revised version received 12.02.25; accepted 20.02.25; published 08.04.25.

Please cite as:
Nian H, Bai Y, Yu H
Assessing the Causal Association Between COVID-19 and Graves Disease: Mendelian Randomization Study
JMIR Form Res 2025;9:e66003
URL: https://formative.jmir.org/2025/1/e66003
doi:10.2196/66003

© Hui Nian, Yu Bai, Hua Yu. Originally published in JMIR Formative Research (https://formative.jmir.org), 8.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

COVID-19 Perceptions Among Communities Living on Ground Crossings of Somali Region of Ethiopia: Community Cross-Sectional Survey Study

Alinoor Mohamed Farah[1], MPH; Abdifatah Abdulahi[2], MD; Abdulahi Hussein[2], MD; Ahmed Abdikadir Hussein[2], MD; Abdi Osman[2], MD; Mohamed Mohamud[1], MPH; Hasan Mowlid[1], MPH; Girum Hailu[3], MD; Fathia Alwan[3], MBA; Ermiyas Abebe Bizuneh[4], MSCE; Ahmed Mohammed Ibrahim[1], MPH; Elyas Abdulahi[4], PhD

[1]Faculty of Public Health, Institute of Health, Jigjiga University, Kebele 16, Jigjiga, Ethiopia

[2]Faculty of Medical Sciences, Institute of Health, Jigjiga University, Jigjiga, Ethiopia

[3]Health and Development Division, Intergovernmental Authority on Development, Addis Ababa, Ethiopia

[4]Research and Community Service Directorate, Jigjiga University, Jigjiga, Ethiopia

Corresponding Author:
Alinoor Mohamed Farah, MPH
Faculty of Public Health, Institute of Health, Jigjiga University, Kebele 16, Jigjiga, Ethiopia

Abstract

Background: The COVID-19 pandemic has profoundly affected the movement of people across borders in Eastern and Southern Africa. The implementation of border closures and restrictive measures has disrupted the region's economic and social dynamics. In areas where national authorities lack full control over official and unofficial land crossings, enforcing public health protocols to mitigate health risks may prove challenging.

Objective: This study aimed to assess perceived factors that influence the spread and control of COVID-19 among Somali communities living on and near ground crossings in Tog Wajaale, Somali region, Ethiopia.

Methods: A community-based cross-sectional study was conducted using a multistage sampling technique. Beliefs and perceptions of the virus's spread and control were partially adapted from the World Health Organization (WHO) resources, exploring four main perception themes: (1) perceived facilitators for the spread of the virus, (2) perceived inhibitors, (3) risk labeling, and (4) sociodemographic variables. A sample size of 634 was determined using the single proportion formula. Standardized mean scores (0 - 100) and SDs categorized perception themes, with mean differences by sociodemographic variables analyzed using ANOVA and *t* tests. Statistical significance was established with a 95% CI and a *P* value below .05. The data were analyzed using STATA version 14.1.

Results: Factors influencing COVID-19 spread and control include behavioral nonadherence and enabling environments. A total of 81.9% (439/536) did not comply with social distancing, and 92.2% (493/536) faced constraints preventing them from staying home and enabling environments. Misconceptions were prevalent, including beliefs about hot weather (358/536, 66.8%), traditional medicine (36/536, 6.7%), and religiosity (425/536, 79.3%). False assurances also contributed, such as feeling safe due to geographic distance from hot spots (76/536, 14.2%) and perceiving the virus as low-risk or exaggerated (162/536, 30.2%). Only 25.2% (135/536) followed standard precautions and 29.9% (160/536) were vaccinated. Employment, region, income, sex, education, and information sources significantly influenced behavioral nonadherence, myth prevalence, and false assurances.

Conclusions: The findings highlight the need for substantial risk communication and community engagement. Only 46.6% (250/536) of individuals adhered to precautionary measures, there was a high perception of nonadherence, and essential COVID-19 resources were lacking. Additionally, numerous misconceptions and false reassurances were noted. Understanding cross-border community behavior is crucial for developing effective, contextually appropriate strategies to mitigate COVID-19 risk in these regions.

(JMIR Form Res 2025;9:e66751) doi:10.2196/66751

KEYWORDS

COVID-19; perceptions; border-crossings; SARS-CoV-2; coronavirus; respiratory; infectious; pandemic; Somali; Ethiopia; Africa; community-based; World Health Organization; WHO; cross-sectional study; multistage sampling technique; transmission; cross-border; community engagement

XSL•FO
RenderX

Introduction

Cross-border movement has been identified as an important factor in COVID-19 transmission in Eastern and Southern Africa, and most countries in the region have restricted entry to reduce virus importation. However, the region's economy and social fabric are dependent on the cross-border movement of goods and people, and it was anticipated that border closures would have the same, if not greater, impact than COVID-19 [1,2]. As a result, countries have relaxed COVID-19 containment measures and lifted border closures, increasing the risk of the virus spreading internationally.

Under normal circumstances, a diverse range of people crosses national borders in the Eastern and Southern Africa region to participate in vital activities. Truck and boda-boda (motorcycle taxi) drivers, small- and large-scale traders, tourists, migrant workers, fisherfolk and pastoralists, refugees, returnees, border community members, and people traveling to seek health care and other social services are among those affected. These people interact with the communities they pass through on their way to and from work. Border officials, security officers, and support staff interact with travelers daily, increasing the risk of community spread [2-6].

Since the International Health Regulations 2005 came into effect in 2007, there has been growing recognition that, unlike airports and ports, ground crossings frequently represent informal passages between 2 countries with no physical structure, barriers, or borders. Furthermore, ground crossing is important for the international spread of diseases. Travelers and those who live and work on or near borders are particularly vulnerable to this threat [7].

Communities that live on and near ground crossings vary in size and density. Cross-border movement is a daily necessity for many people living in these communities for work, trade, family visits, schooling, health care services, religious activities, entertainment, and other reasons. However, health measures to control public health risks may be difficult to implement in places where national authorities are unable to fully monitor formal and informal ground crossings [7].

Given that borders have already been reopened, it is critical to understand cross-border transmission dynamics, movement patterns, and the behavior of cross-border actors to develop effective, contextually appropriate risk communication and community engagement (RCCE) strategies to mitigate COVID-19 risk. Thus, this study aimed to assess the perceived factors that influence the spread and control of COVID-19 among Somali communities living on and near ground crossings.

Methods

Study Design, Setting, and Period

A community-based cross-sectional study was conducted from December 25 to 31, 2022, in Tog Wajaale, Somali region, Ethiopia. Tog Wajaale, also known as Wajaale, is a city in the Somali region near the border with Somaliland. Tog Wajaale is the main border crossing for goods entering and leaving Somaliland, primarily from Berbera, the country's main port. Tog Wajaale is located on the border between Ethiopia and Somaliland. It is 70 km east of Jigjiga, the capital of the Somali region.

Populations

All adult populations older than 18 years were included in the source population. The study participants were all adult populations older than 18 years who had lived in the Tog Wajaale town for at least 6 months.

Eligibility Criteria

The study included individuals aged 18 years and older who had resided in the study area for a minimum of 6 months. Those who were experiencing mental illness or severe health issues during the study period were excluded from participation.

Sample Size Determination

The sample size was obtained using the single proportions formula with the specified parameters of 50% (P=.50, as there is no comparable study in Ethiopia), 95% CI ($Z_1 - \alpha/2$) 1.96%, and 5% margin of error (d). Adding a 10% nonresponse rate and a design effect of 1.5 resulted in a required sample size of 634.

Sampling Technique

A multistage sampling technique is used to select participants for the study. In the first stage, 1 city council in the region, based on the border with the neighboring country and demographic diversity, is selected. In the second stage, 5 kebeles (subdistricts) are randomly selected, and the total sample size is allocated by probability proportional to size (PPS) to each kebele based on the number of registered households. In the third stage, using the list of households as the sampling frame, a systematic random sampling technique is used to select the study participants from the kebeles.

Data Collection and Quality Control

Data were collected using a pretested structured questionnaire. Ten BSc nurses fluent in the local language collected data through face-to-face interviews at the household level. Beliefs and perceptions of the virus's spread and control were partially adapted from the World Health Organization (WHO) resources. Specifically, some of the questions were adapted from the COVID-19 rapid quantitative assessment tool [8,9]. Three main perception themes were explored: perceived facilitators for the spread of the virus (6 items), perceived inhibitors (7 items), and risk labeling (8 items), as well as sociodemographic variables, including access to communication resources. Prior to participation, the household head or any individual over 18 years of age was fully informed of the study details and asked for consent. Questionnaires were reviewed daily for completeness, and respondents were contacted to see if any information was missing. The investigators regularly supervised the data collection process. To ensure data quality, ongoing supervision and follow-ups were conducted throughout the collection period.

Data Analysis

KoBoToolbox, an open-source data collection platform, was used to collect data. KoBo-collected responses were initially

XSL•FO
RenderX

encoded in a Microsoft Excel database and then exported to STATA version 14.1 for analysis. The frequency tables were used to present the respondents' background variables as well as specific belief items. Standardized mean scores (0 - 100) and SDs were used to describe the lists of factor categories based on the perception themes. To compare mean differences by sociodemographic variables, 1-way ANOVA and *t* tests were used. A factor loading score of 0.4 was used as a cutoff value to retain items in each category [10]. Exploratory factor analysis (EFA) categorized perceptions into 4 themes: facilitators of virus transmission (6 factors), inhibitors (7 factors), risk identification (8 factors), and sociodemographic factors, including communication resource access. The sample's adequacy for EFA was confirmed by the Kaiser-Mayer-Olkin measure (KMO>50%). Statistical significance was determined using a 95% CI and a *P* value below .05.

Operation Definition and Variables

The *perceived factors* influencing the spread and control of COVID-19 include exacerbating factors, inhibiting factors, and risk labeling.

Perceived COVID-19 Exacerbating Factors

Behavioral Nonadherence

Noncompliance with public health measures designed to control the spread of COVID-19 encompasses various behaviors exhibited by individuals or groups. These actions diverge from the recommended guidelines and may include disregarding social distancing protocols, neglecting to use face masks, refusing COVID-19 vaccination, or displaying discriminatory attitudes toward others in relation to the pandemic.

Lack of Enabling Environment

The absence of essential conditions that support the implementation of public health measures to prevent COVID-19 transmission. For example, economic and social constraints that force people to leave their homes for work or social issues; overcrowded living and working conditions that may not allow physical distancing; lack of access to personal protective equipment such as face masks; and insufficient availability of clean water, soap, hand rub alcohol, or sanitizers.

Perceived COVID-19 Inhibiting Factors

False Assurance

Misconceptions or preconceptions that diminish the perceived threat of COVID-19 frequently result in a relaxed compliance with preventive protocols. Examples include the belief that one's location is far removed from epidemic centers or the underestimation of the seriousness of the illness.

Myths

Misinformed or culturally rooted beliefs that influence public perception of COVID-19 and its risks often undermine scientific evidence. Examples include beliefs that hot weather kills the virus, reliance on traditional medicine as a cure, or attributing protection to religious practices alone.

Standard Precautions

The actions taken by individuals to minimize the risk of COVID-19 transmission were consistent with the recommended public health guidelines. These include wearing masks, maintaining hand hygiene, avoiding crowded areas, and vaccinating against COVID-19.

Risk Labeling

Risk labeling involves perceiving a specific sect of the community to be at a high risk of contracting COVID-19, especially older adults and people with underlying medical conditions.

Ethical Considerations

This study was conducted in accordance with the guidelines of the Declaration of Helsinki, and all procedures involving research study participants were approved by the Jigjiga University institutional review board (RERC/052/2022). Written informed consent was obtained from all the participants. Participants were not compensated for participating in the study. The study participants' data are confidential and protected.

Results

Sociodemographic Characteristics of Participants

The survey questionnaire was completed by 536 participants, resulting in a response rate of 84.5% (536/634). Table 1 summarizes the demographic information of the survey respondents. The majority of respondents (439/536, 81.9%) were younger than 40 years and male (313/536, 58.4%). Most of the respondents (153/536, 28.5%) had a high school education, followed by those who could read and write but with no formal education (114/536, 21.3%). The majority of participants in the study were married (335/536, 62.5%), ran their own businesses (261/536, 48.7%), were Muslim (501/536, 93.5%), and were of Somali origin (423/536, 78.9%). Furthermore, the majority of participants (287/536, 53.5%) used cellular data as their primary internet source, had an income greater than 5000 ETB (Ethiopian Birr; US $93) (324/536, 60.4%), and their primary source of information about COVID-19 was television (230/536, 42.9%).

XSL•FO
RenderX

Table . Sociodemographic characteristics of the respondents (N=536).

Characteristics		Values, n (%)
Age (years)		
	<40	439 (81.9)
	≥40	97 (18.1)
Sex		
	Male	313 (58.4)
	Female	223 (41.6)
Education		
	Cannot read and write	10 (1.9)
	Read and write	114 (21.3)
	Primary (grade 1 - 8)	78 (14.6)
	High school (grade 9 - 12)	153 (28.5)
	Diploma or level I-IV	104 (19.4)
	First degree (BSc or BA) and above	77 (14.4)
Marital status		
	Single (unmarried)	174 (32.5)
	Married	335 (62.5)
	Divorced	13 (2.4)
	Widowed	4 (0.7)
	Engaged (in relationship)	10 (1.9)
Employment		
	Private or business	261 (48.7)
	Daily laborer	75 (14)
	Government employee	80 (14.9)
	Student	34 (6.3)
	No job	86 (16)
Religion		
	Muslim	501 (93.5)
	Orthodox	27 (5)
	Protestant	8 (1.5)
Place of origin		
	Somali region	423 (78.9)
	Other regions	113 (21.1)
Internet connection mostly used to get information about COVID-19		
	Cellular data (telle)	287 (53.5)
	Wireless Wi-Fi	46 (8.6)
	Do not use internet	203 (37.9)
Monthly income		
	<5000 ETB[a] (US $93)	212 (39.6)
	≥5000 ETB (US $93)	324 (60.4)
Main source of information about COVID-19		
	Television	230 (42.9)
	Radio	16 (3)

XSL•FO
RenderX

Characteristics		Values, n (%)
	Friends and neighbors	58 (10.8)
	Social media (Facebook, Twitter, YouTube, etc)	196 (36.6)
	Health workers	36 (6.7)
Ever used the call center for information about COVID-19		
	Yes	22 (4.1)
	No	514 (95.9)

[a]ETB: Ethiopian Birr.

Perceived COVID-19 Exacerbating and Inhibiting Factors

Table 2 shows the prevalence of specific factors in the exacerbating and inhibiting categories. The prevalence of specific factors associated with behavioral nonadherence ranged from 81.9% (439/536) to 32.5% (174/536). The fact that people are still not adhering to physical or social distance accounted for a greater proportion of nonadherence. The magnitude of the absence of enablers that would support behavioral adherence ranged from 92.2% (494/536) to 5.4% (29/536). The fact that people could not stay home for economic and social reasons and living or working in overcrowded conditions contributed significantly to the lack of enablers. Environmental factors were perceived as more deterrent than behavioral nonadherence, indicating a strong tendency for externalizing factors that could exacerbate the spread of COVID-19 in the community.

XSL•FO
RenderX

Table . Perceived COVID-19 exacerbating and inhibiting factors (N=536).

Perceived factors		Descriptive statistics, n (%)
Perceived COVID-19 exacerbating factors		
Behavioral nonadherence		
	People fear stigma and bias related to COVID-19	205 (38.3)
	People still do not respect social distance, ie, use crowded transportation means, still hug and shake each other's hands while greeting	439 (81.9)
	People refused to take COVID-19 vaccine	174 (32.5)
Lack of enabling environment		
	People do not stay at home for economic and social reasons	494 (92.2)
	People still live and work in a very crowded condition	403 (75.2)
	People do not have access to face masks, water, and sanitizers	29 (5.4)
Perceived COVID-19 inhibiting factors		
Myths		
	The weather we live-in is too hot for coronavirus to survive	358 (66.8)
	We believe we have traditional medicine against COVID-19	36 (6.7)
	We are religious enough to control COVID-19	425 (79.3)
False assurances		
	We live far away from COVID-19's hot spot areas	76 (14.2)
	COVID-19 is not risky or exaggerated	162 (30.2)
Engaged in precaution		
	Engaged in standard precaution measures of COVID-19	135 (25.2)
	Vaccinated against COVID-19	160 (29.9)

The most common myths were perceived as protectiveness from hot weather and religiosity, accounting for 66.8% (358/536) and 79.3% (425/536), respectively. Beliefs that the specific localities where respondents are currently living are far away from coronavirus hot spot areas and that COVID-19 is not dangerous or exaggerated contributed to 14.2% (76/536) and 30.2% (162/536) of the false assurances, respectively. On the other hand, the prevalence of people believing that the spread of COVID-19 would be controlled because of their active participation in standard precautionary measures and the COVID-19 vaccine were 25.2% (135/536) and 29.9% (160/536), respectively.

Perceived Categories of COVID-19 Exacerbating and Inhibiting Factors

Table 3 presents the categories and lists of perceived COVID-19 exacerbations and inhibitors, along with their respective prevalence. EFA identified 2 major categories of perceived exacerbating and inhibiting COVID-19 factors in the target population. The first category of factors was classified as behavioral nonadherence, indicating noncompliance with recommended precautionary measures. Individuals engaging in handshaking, using crowded transportation, refusing COVID-19 vaccination, and exhibiting a factor loading score of 0.714 collectively contribute to behavioral nonadherence.

The lack of enabling environmental conditions, which are supposed to support the adaptation of precautionary measures, was the second category of exacerbating facilitating factors. With factor loading scores of 0.71, the lack of enablers was made up of economic reasons that challenged the stay-at-home principle, overcrowded living or working conditions, and a lack of personal protective equipment such as face masks and sanitizers. Behavioral nonadherence and lack of enabler factors

XSL•FO
RenderX

explained 28.4% (152/536) of the variance in perceived virus facilitators.

Table . Perceived categories of COVID-19 exacerbating and inhibiting factors. Kaiser-Mayer-Olkin measure of sampling adequacy (KMO=53%) and degrees of freedom (*df*=10).

Perceived factors about COVID-19	Principal components and factor loading scores		Descriptive statistics (N=536)	
	Exacerbating factors	Inhibiting factors	n (%)	95% CI
Behavioral nonadherence[a]	0.71	—[b]	335 (62.5)	58.3-66.5
Lack of enabling environment[c]	0.71	—	494 (92.2)	89.6-94.2
Myths[d]	—	0.70	425 (79.3)	75.6-82.5
False assurances[e]	—	0.32	216 (40.3)	36.2-44.5
Engaged in precautions[f]	—	0.71	250 (46.6)	42.4-50.9

[a]Behavioral nonadherence: People fear stigma and bias related to COVID-19, people still use crowded transportation means, people with flu-like symptoms are not well screened for COVID-19, people refuse to take the COVID-19 vaccine, and people still hug and shake each other's hands while greeting.

[b]Not applicable.

[c]Lack of enabling environment: People do not stay at home for economic and social reasons, people still live and work in a very crowded condition, people do not have personal protective equipment like face masks, and people do not have access to water or hand rub alcohol or sanitizers.

[d]Myths: The weather we live in is too hot for coronavirus to survive, we believe we have traditional medicine against COVID-19, and we are religious enough to control COVID-19.

[e]False assurances: We live far away from COVID-19's rampant areas and COVID-19 is not risky or exaggerated.

[f]Engaged in precautions: Engaged in standard precaution measures for COVID-19 and vaccinated against COVID-19.

The table also includes categories and lists of perceived COVID-19 inhibitors as well as their respective prevalence. EFA produced 3 major categories of perceived factors that inhibited COVID-19. Two of the three categories were incorrectly identified (myths and false assurances), whereas one was correctly identified as an inhibitor. The myths category included factors that are thought to inhibit the virus but have not been scientifically proven. In this case, the myths included perceived religiosity (perceiving oneself as an effective religious man or woman in dealing with challenges), people's perceived confidence in owning effective traditional medicines that were not clinically confirmed, and living in hot weather, with a factor loading score of 0.70.

The second category of perceived inhibitors, which were frequently associated with false assurances that people were immune to COVID-19. With factor loading scores of 0.32, the category consisted of 2 main beliefs: we live far away from COVID-19 hot spot areas and there have been no locally reported COVID-19 cases so far. The beliefs appeared to be false assurances in that people perceive themselves to be living outside a risk zone, giving the impression of invulnerability.

With a factor score loading of 0.71, the third, promotable category of perceived inhibitors concerned people who had taken standard precautions and were vaccinated against COVID-19. Factors related to the 3 categories explained 20.1% (108/532) of the variance in perceived inhibitors, indicating the presence of several other unreported myths and unhealthy beliefs that require further investigation.

Risk Labeling

The COVID-19 risk labels and groups are presented in Table 4. People with underlying illnesses (118/536, 22%) and older adults (351/536, 65.5%) were perceived by the community to be at a high risk of COVID-19.

Table . Perceived COVID-19 risk groups and labels (N=536).

Perceived high-risk groups	Descriptive statistics, n (%)
Children (aged 0 - 9 years)	38 (7.1)
Adolescents (aged 10-15 years)	3 (0.6)
Youths (aged 16 - 29 years)	2 (0.4)
Adults (aged 30 - 50 years)	11 (2.1)
Older persons (aged 51 years or more)	351 (65.5)
Pregnant women	9 (1.7)
Health workers	4 (0.7)
People with underlying chronic illness (eg, diabetes, hypertension, and cancer)	118 (22)

XSL•FO
RenderX

Distribution of Perceived Exacerbating Factors

Table 5 illustrates the relationships between sociodemographic characteristics, behavioral nonadherence, and lack of an enabling environment in relation to COVID-19. The mean percentages of behavioral nonadherence and lack of an enabling environment as facilitators of virus spread were both high, at 62.5% (335/536) and 92.2% (494/536), respectively.

XSL•FO
RenderX

Table . Distribution of perceived exacerbating factors among the study participants.

Variables	Behavioral nonadherence[a]				Lack of enabling environment[b]			
	% Mean (SE)	*t* test (*df*=534)	*F* test (*df*=535)	*P* value	% Mean (SE)	*t* test (*df*=534)	*F* test (*df*=535)	*P* value
Overall	62.5 (2.1)				92.2 (1.2)			
Sex		–1.20		.12		2.78		.003[c]
Male	60.4 (2.8)				94.9 (1.2)			
Female	65.5 (3.2)				88.3 (2.2)			
Age (years)		–0.39		.35		0.51		.31
<40	62.1 (2.3)				92.4 (1.5)			
≥40	64.4 (5.2)				90.8 (3.1)			
Education			1.37	.23			3.55	.003[c]
Cannot read and write	50.0 (15.3)				80.0 (8.4)			
Read and write	63.2 (4.5)				86.8 (2.5)			
Primary (grade 1 - 8)	55.1 (5.5)				87.2 (3.0)			
High school (grade 9 - 12)	58.8 (3.9)				92.8 (2.1)			
Diploma or level I-IV	70.2 (4.7)				97.1 (2.6)			
First degree (BSc or BA) and above	67.5 (5.5)				98.7 (3.0)			
Marital status			0.24	.92			4.12	.003[c]
Single (unmarried)	63.2 (3.7)				94.8 (2.0)			
Married	62.1 (2.7)				90.7 (1.5)			
Divorced	53.8 (13.5)				92.3 (7.5)			
Widowed	75.0 (24.3)				75.0 (13.4)			
Engaged	70.0 (15.5)				—[d]			
Employment			3.33	.01[c]			0.75	.56
Private or business	64.4 (2.9)				90.8 (1.7)			
Daily laborer	48.0 (5.5)				90.7 (3.1)			
Government employee	60.0 (5.4)				96.3 (3.1)			
Student	82.4 (8.2)				94.1 (4.6)			
No job	64.0 (5.2)				93.0 (2.9)			
Religion			0.34	.71			0.35	.71
Muslim	62.9 (2.2)				92.0 (1.4)			
Orthodox	59.3 (9.3)				92.6 (6.2)			
Protestant	50.0 (17.2)				—			
Place of origin		1.89		.03[c]		–1.52		.06
Somali region	64.5 (2.3)				91.3 (1.4)			

XSL•FO
RenderX

Variables	Behavioral nonadherence[a]				Lack of enabling environment[b]			
	% Mean (SE)	*t* test (*df*=534)	*F* test (*df*=535)	*P* value	% Mean (SE)	*t* test (*df*=534)	*F* test (*df*=535)	*P* value
Other regions	54.9 (4.7)				95.6 (1.9)			
Internet connection mostly used to get information about COVID-19			1.03	.36			7.68	<.001[c]
Cellular data (telle)	65.2 (2.9)				94.8 (1.6)			
Wireless Wi-Fi	56.5 (7.1)				—			
Do not use internet	60.1 (3.4)				86.7 (1.8)			
Monthly income		–2.84		.002[c]		0.2		.42
<5000 ETB[e] (US $93)	55.2 (3.4)				92.4 (1.8)			
≥5000 ETB (US $93)	67.6 (2.6)				91.9 (1.5)			
Source of information			1.31	.25			6.84	<.001
Official websites	50.0 (34.1)				—			
Television	60.1 (3.2)				92.1 (1.7)			
Radio	50.0 (12.1)				93.8 (6.5)			
Friends or neighbors	55.2 (6.3)				75.8 (3.4)			
Social media (Facebook, Twitter, YouTube, etc)	68.9 (3.5)				97.9 (1.9)			
Health workers	61.1 (8.1)				86.1 (4.4)			

[a]Behavioral nonadherence: People fear stigma and bias related to COVID-19, people still use crowded transportation means, people with flu-like symptoms are not well screened for COVID-19, people refuse to take the COVID-19 vaccine, and people still hug and shake each other's hands while greeting.

[b]Lack of enabling environment: People do not stay at home for economic and social reasons, people still live and work in a very crowded condition, people do not have personal protective equipment like face masks, and people do not have access to water or hand rub alcohol or sanitizers.

[c]Statistically significant at $P<.05$ (two-tailed).

[d]Not applicable.

[e]ETB: Ethiopian Birr.

In terms of behavioral nonadherence, respondents' employment, region, and income showed significant mean percentage differences. Students had the highest mean percentage of behavioral nonadherence (442/536, 82.4%; *P*=.01) when it came to employment. Respondents of Somali origin and those with an income of more than 5000 ETB (US $93) had the highest mean percentages of behavioral nonadherence (346/536, 64.6%; *P*=.03) and (362/536, 67.5%; *P*=.003), respectively. There was no difference in behavioral nonadherence by age, gender, education, marital status, religion, primary internet source, or primary information source for COVID-19.

Similarly, the lack of an enabling environment that was perceived as a facilitator of the spread of the virus was more prevalent among male respondents (509/536, 95%; *P*=.003), respondents whose education status was first degree and above (529/536, 98.7%; *P*=.003), respondents who were single and not married (508/536, 94.8%; *P*=.001), those who used cellular data as their primary source of the internet (508/536, 94.8%; *P*=.001), and those who used social media as their primary source of information (525/536, 97.9%; *P*<.001). There was no difference in behavioral nonadherence by age, employment, religion, place of origin, and monthly income.

XSL•FO
RenderX

Distribution of Perceived Inhibiting Factors

The factors deemed to hinder the transmission and control of the virus exhibited considerable sociodemographic disparities (Table 6). The highest perceived inhibitor to control the spread of COVID-19 was seen to be myth (425/536, 79.3%), followed by preventive measures (250/536, 46.6%) and false guarantees (214/536, 39.9%).

XSL•FO
RenderX

Table . Distribution of perceived inhibiting factors among the study participants.

Variables	Myth[a]				False assurance[b]				Engaged in precaution[c]			
	% Mean (SE)	*t* test (*df*=534)	*F* test (*df*=535)	*P* value	% Mean (SE)	*t* test (*df*=534)	*F* test (*df*=535)	*P* value	% Mean (SE)	*t* test (*df*=534)	*F* test (*df*=535)	*P* value
Overall	79.3 (1.8)				40.3 (2.1)				46.6 (2.2)			
Sex		2.79		.003[d]		0.33		.37		–0.17		.43
Male	83.4 (2.1)				40.9 (2.8)				46.3 (2.8)			
Female	73.4 (3.0)				39.4 (3.3)				47.1 (3.4)			
Age (years)		–0.58		.28		–0.70		.76		0.73		.23
<40	78.8 (1.9)				39.6 (2.3)				47.0 (2.3)			
≥40	81.6 (4.2)				43.7 (5.3)				44.8 (5.4)			
Education			7.37	<.001[d]			2.12	.06			3.72	.003[d]
Cannot read and write	90.0 (12.5)				50.0 (15.4)				60.0 (15.6)			
Read and write	79.8 (3.7)				46.5 (4.6)				40.4 (4.6)			
Primary (grade 1 - 8)	60.3 (4.5)				30.8 (5.5)				47.4 (5.6)			
High school (grade 9 - 12)	74.5 (3.2)				46.4 (3.9)				36.6 (4.0)			
Diploma or level I-IV	88.5 (3.9)				37.5 (4.8)				58.7 (4.8)			
First degree (BSc or BA) and above	93.5 (4.5)				31.2 (5.6)				57.1 (5.6)			
Marital status			1.42	.23			0.8	.52			0.94	.44
Single (unmarried)	83.3 (3.1)				36.8 (3.7)				47.7 (3.8)			
Married	77.6 (2.2)				41.8 (2.7)				45.7 (2.7)			
Divorced	76.9 (11.2)				38.4 (13.6)				61.5 (13.9)			
Widowed	—[e]				75.0 (24.6)				75.0 (25.0)			
Engaged (in relationship)	60.0 (12.8)				40.0 (15.5)				30.0 (15.8)			
Employment			2.51	.041[d]			4.81	<.001[d]			0.47	.76

XSL•FO
RenderX

Variables	Myth[a]				False assurance[b]				Engaged in precaution[c]			
	% Mean (SE)	*t* test (*df*=534)	*F* test (*df*=535)	*P* value	% Mean (SE)	*t* test (*df*=534)	*F* test (*df*=535)	*P* value	% Mean (SE)	*t* test (*df*=534)	*F* test (*df*=535)	*P* value
Private or business	82.0 (2.5)				49.0 (3.0)				45.2 (3.1)			
Daily laborer	66.7 (4.7)				30.7 (5.6)				48.0 (5.8)			
Government employee	83.8 (4.5)				38.8 (5.4)				43.8 (5.6)			
Student	82.4 (6.9)				32.4 (8.3)				55.9 (8.6)			
No job	76.7 (4.3)				26.7 (5.2)				48.8 (5.4)			
Religion			0.21	.81			0.95	.39			0.17	.85
Muslim	79.0 (1.8)				39.7 (2.2)				46.9 (2.2)			
Orthodox	81.5 (7.8)				44.4 (9.4)				44.4 (9.6)			
Protestant	87.5 (14.4)				62.5 (17.4)				37.5 (17.7)			
Place of origin			0.68	.25		–1.82		.03[d]		0.99		.16
Somali region	79.9 (2.0)				38.3 (2.4)				47.8 (2.4)			
Other regions[f]	77.5 (4.0)				47.8 (4.7)				42.5 (4.7)			
Internet connection mostly used to get information about COVID-19			0.46	.63			5.37	.005[d]			4.1	.02[d]
Cellular data (telle)	78.7 (2.9)				36.9 (2.9)				49.8 (2.9)			
Wireless Wi-Fi	84.8 (7.2)				26.1 (7.2)				58.7 (7.3)			
Do not use internet	78.8 (3.4)				48.3 (3.4)				39.4 (3.5)			
Monthly income		–0.02		.49		3.74		<.001[d]		–4.29		<.001[d]
<5000 ETB[g] (US $93)	79.2 (2.8)				50.0 (3.4)				35.4 (3.3)			
≥5000 ETB (US $93)	79.3 (2.3)				33.9 (2.6)				54.0 (2.8)			

XSL•FO
RenderX

Variables	Myth[a]				False assurance[b]				Engaged in precaution[c]			
	% Mean (SE)	*t* test (*df*=534)	*F* test (*df*=535)	*P* value	% Mean (SE)	*t* test (*df*=534)	*F* test (*df*=535)	*P* value	% Mean (SE)	*t* test (*df*=534)	*F* test (*df*=535)	*P* value
Source of information			2.32	.042[d]			3.18	.008[d]			2.75	.02[d]
Official website	—				50.0 (34.4)				50.0 (35.0)			
Television	77.2 (2.7)				36.8 (3.2)				45.6 (3.3)			
Radio	56.3 (10.1)				31.3 (12.1)				18.8 (12.4)			
Friends and neighbors	74.1 (5.3)				62.1 (6.4)				32.8 (6.5)			
Social media[h]	85.2 (2.9)				36.7 (3.5)				53.6 (3.5)			
Health workers	77.8 (6.7)				50.0 (8.1)				50.0 (8.3)			

[a]Myths: The weather we live in is too hot for coronavirus to survive, we believe we have traditional medicine against COVID-19, and we are religious enough to control COVID-19.

[b]False assurances: We live far away from COVID-19's rampant areas, and COVID-19 is not risky or exaggerated.

[c]Engaged in precautions: Engaged in standard precautions for COVID-19 and vaccinated against COVID-19.

[d]Statistically significant at $P<.05$ (two-tailed).

[e]Not applicable.

[f]Other regions include Addis Ababa, Afar, Amhara, Oromia, southern nations and nationalities, and Tigray.

[g]ETB: Ethiopian Birr.

[h]Facebook, Twitter, YouTube, etc.

Male respondents (447/536, 83.4%; $P=.003$), those with a first degree or higher (501/536, 93.5%; $P<.001$), government employees (449/536, 83.8%; $P=.041$), and those who used social media as their primary source of information (457/536, 85.3%; $P=.042$) had a significantly higher prevalence of myths. There was no difference in myth prevalence by age, marital status, religion, place of origin, primary internet source, or monthly income.

False assurances were also more common among those who owned or ran a private business (263/536, 49.1%; $P=.001$), non-Somalis (256/536, 47.8%; $P=.03$), noninternet users (259/536, 48.3%; $P=.005$), those with monthly income less than 5000 ETB (US $93; 268/536, 50%; $P=.001$), and those who get COVID-19 information from friends and neighbors (333/536, 62.1%; $P<.001$). There was no significant variation in the prevalence of false assurance by age, sex, education, marital status, religion, employment, or primary information source for COVID-19.

Likewise, participants who were unable to read and write, those who used Wi-Fi as their primary source of internet, those who earned more than 5000 ETB (US $93), and those who used social media as their primary source of COVID-19 information were more likely to take precautionary measures than their counterparts. There was no significant variation in the prevalence of precautionary measures according to age, sex, marital status, religion, and employment.

Discussion

Principal Findings

This study produced relevant findings regarding community perceptions of the factors that facilitate and inhibit the spread of COVID-19, risk labeling, and information needs in a border town in the Somali region. The perceived factors were classified into the following major categories: behavioral adherence, a lack of enabling environmental conditions, and risk labeling.

Factors perceived to exacerbate virus spread were grouped into 2 thematic categories: behavioral nonadherence (335/536, 62.5%) and a lack of enabling environmental conditions (494/536, 92.2%). The study revealed a concerning lack of awareness and adherence with 80% (429/536) of respondents noting that individuals continue to disregard social distancing. Additionally, 38.1% (204/536) of those surveyed indicated that people fear the stigma associated with the virus, while 32.5% (174/536) expressed unwillingness to receive the COVID-19 vaccine. These results highlight a significant gap in the understanding of and compliance with preventive measures.

The use of crowded transportation modes was not only due to a lack of alternatives but also associated with behavioral

XSL•FO
RenderX

nonadherence. Fears, threat appraisals, and worries have consistently been linked to increased adherence in COVID-19 protective behaviors [11,12]. An Iranian study found that increased fear predicted less adherence [13]. This study shows that clear communication of risk could increase participation in protective behaviors, especially among disengaged groups.

Conversely, people could not stay at home for economic and social reasons (494/536, 92.2%); they did not have access to face masks, water, or sanitizers (29/536, 5.4%); and they still lived and worked in crowded conditions (403/536, 75.2%). According to behavioral and communication theories, people's perceived lack of resources has a negative impact on their actual practices [14]. Despite this, the high prevalence of perceived facilitators indicates 2 major and urgent issues. First, it indicates a strong effort to reduce behavioral nonadherence and a lack of enablers that facilitate viral spread. Second, even a higher perceived lack of enabling conditions appears concerning because it may lead people to externalize their ability to control the virus while ignoring their own efforts.

Poverty and livelihood issues are major challenges in COVID-19 prevention in low-income countries. This is because COVID-19 has impacted people's livelihoods and income opportunities, resulting in food insecurity [15]. Most people in Tog Wajaale are low-income earners who work in the informal sector and rely on daily wages to feed their families and cover their related expenses. As most of them struggle to earn a living, COVID-19 has added to their difficulties. As a result, they are confronted with the harsh reality of contracting the virus while attempting to survive. The observation results show that despite the lockdown, daily laborers in the construction sector, informal businesses, and street vendors continued their daily activities [16].

Factors perceived as inhibitors of viral spread were divided into 3 categories: false assurances (216/536, 40.3%), myths (425/536, 79.3%), and standard precautions (250/536, 46.6%). False assurances were perceptions of invulnerability, characterized by people believing that they were living outside the COVID-19 risk zone. The 2 main false assurances in this study were the perception of COVID-19 as less risky or exaggerated and residence far away from the COVID-19 rampant areas. Myths in this study included religious effectiveness (425/536, 79.3%), living in hot weather (358/536, 66.8%), and traditional medicine (36/536, 6.7%). The WHO myth busters list the majority of the misconceptions presented in this study, indicating that they were universally held during the pandemic [17]. Evidence suggests that myths or misperceptions, such as denial of presence and misperceived causes, transmission modes, and prevention methods, can prevent and control efforts during pandemics of HIV, Zika virus, yellow fever, and Ebola unless traced and addressed [18-21].

Culture and religion are important components of Somalis' social and moral fabrics. As a result, cultural and religious misunderstandings as well as misinformation about the virus are among the factors contributing to complacency when it comes to prevention measures. According to a study conducted in northern Ethiopia, people believe that COVID-19 is a "punishment from God" or "Allah" for people's immoral acts, and that the power of prayer can prevent the pandemic's severe effects [16].

The magnitude of correctly perceived factor engagement in standard precautionary measures (134/536, 25%) and vaccination against COVID-19 (160/536, 29.9%) was too low, necessitating hard work to promote this perception, until a larger segment of the community embraced an accurate reason for virus protection.

Respondents' employment, region, and income all showed significant mean percentage differences in behavioral nonadherence. When it came to employment, students had the highest mean percentage of behavioral nonadherence (442/536, 82.5%). Respondents of Somali origin and those with an income of more than 5000 ETB (US $93) had the highest mean percentages of behavioral nonadherence (64.6%, 346/536 and 67.5%, 362/536, respectively). Students were most likely to be of younger age, and being younger was found to be a predicator of nonadherence to preventive behavior [22]. Regular communication regarding SARS-CoV-2 emphasizes that COVID-19 mostly threatens the older population. This could send a false message of safety to the younger generations. Younger individuals' lower adherence may be related to their perceived lower vulnerability and propensity to make impulsive decisions, take greater risks to gain emotional and social stimulation (such as attending social gatherings), and fail to recognize the long-term effects of their actions [23].

Similarly, almost 60.1% (322/536) of the communities surveyed are reliant on cross-border commerce and daily labor, and 20% (107/536) of them live on less than US $100 per month. Thus, they are driven to leave their homes, even during the COVID-19 epidemic, to sustain their families. Within the Somali population, males typically serve as primary income earners and feel obligated to venture outside their homes, even during the COVID-19 pandemic, to earn money to sustain their families. This observation aligns with a study conducted in Somalia, which also noted that Somali men tend to disregard COVID-19 preventive protocols [24].

Obtaining COVID-19 information from official sources (radio, television, and government announcements) increased the likelihood of having a high adherence score, whereas obtaining it from social media decreased the likelihood of having a high adherence score. This highlights the role of misinformation in determining adherence to scientifically proven COVID-19 prevention measures [25]. In addition, a previous study showed that social media users tend to be more extroverted, suggesting that these people may find it more challenging to stay at home and observe COVID-19 preventive measures than introverted non–social media users [26].

Limitations

First, the cross-sectional design does not provide conclusive conclusions about causality, and social desirability bias may skew the self-reported data. Second, only one border crossing town in parts of the Somali region was surveyed, and the results of this study may not reflect the perceptions of other towns in the Somali region and Ethiopia. A similar study could be

XSL•FO
RenderX

extended to other border-crossing towns in the Somali region and Ethiopia.

Recommendations

There is a need for local activities that encourage participation in standard precautions by addressing locally perceived hurdles and fostering a sense of shared responsibility and community ownership during the fight against COVID-19. Consequently, RCCE efforts in Tog Wajaale should (1) consider sociodemographic variations in myths and false assurances, (2) investigate more beliefs that could facilitate or inhibit the spread of the virus, and (3) design local initiatives that enhance community ownership of tasks associated with controlling the virus, thus supporting and advocating engagement in standard precautionary measures. That is, using community and religious leaders to spread correct information, model precautionary behaviors, and organize COVID-19–related activities.

Conclusions

The results suggest that extensive efforts in RCCE are necessary. This conclusion is supported by several factors. First, only 44.6% (239/536) of individuals correctly engaged in precautionary measures. Second, there was a high perceived prevalence of behavioral nonadherence and a shortage of essential resources in the COVID-19 response. Lastly, numerous misconceptions and false reassurances were identified among the population. Thus, understanding the behavior of cross-border communities is crucial for developing effective, contextually appropriate RCCE strategies to mitigate COVID-19 risk in border communities.

Acknowledgments

The authors wish to thank the district health office and health facility staff for their support and facilitation of the data collection. This study was funded by the Intergovernmental Authority on Development (IGAD). The views expressed herein are solely those of the authors and do not reflect the views of IGAD or any other stakeholders. All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

Data Availability

The dataset for this study is available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: AMF, AA, AH, EA
Formal analysis: AMF
Writing – original draft: AMF
Writing – review & editing: AMF, AA, AH, AAH, AO, MM, HM, GH, FA, EAB, AMI, EA

Conflicts of Interest

None declared.

References

1. Jones L, Schmidt-Sane M. Covid-19 RCCE strategies for cross-border movement in Eastern and Southern Africa. : Social Science in Humanitarian Action Platform; 2020 URL: https://www.socialscienceinaction.org/resources/covid-19-rcce-strategies-for-cross-border-movement-in-eastern-and-southern-africa/ [accessed 2025-02-20]
2. Bouet A, Laborde Debucquet D. COVID-19 border policies create problems for African trade and economic pain for communities. In: COVID-19 and Global Food Security: International Food Policy Research Institute; 2020:69-72.
3. Storer E, Pearson G. Cross-border dynamics and healthcare in West Nile, Uganda. UNICEF, IDS, and Anthrologica. 2019 URL: https://opendocs.ids.ac.uk/articles/online_resource/Cross-Border_Dynamics_and_Healthcare_in_West_Nile_Uganda/26433706?file=48083425 [accessed 2025-02-20]
4. Baluku M, Akello G, Parker M, Grant C. How the ‘disease of the radio’ is the next burden impacting the lives of those on the Uganda-DRC border. : Social Science in Humanitarian Action Platform; 2020 Apr 8 URL: https://www.socialscienceinaction.org/blogs-and-news/covid-19-uganda-disease-radio-next-burden-impacting-lives-uganda-drc-border/ [accessed 2025-02-20]
5. Nshimbi CC. The human side of regions: informal cross-border traders in the Zambia–Malawi–Mozambique growth triangle and prospects for integrating Southern Africa. J Borderl Stud 2020 Jan 1;35(1):75-97. [doi: 10.1080/08865655.2017.1390689]
6. Rutherford B, Addison L. Zimbabwean farm workers in Northern South Africa. Rev Afr Polit Econ 2007 Jan;34(114):619-635. [doi: 10.1080/03056240701819491]
7. Controlling the spread of COVID-19 at ground crossings: interim guidance, 20 May 2020. : World Health Organization; 2020 URL: https://iris.who.int/handle/10665/332165 [accessed 2025-02-20]

XSL•FO
RenderX

8. Risk communication and community engagement (RCCE) action plan guidance COVID-19 preparedness and response. : World Health Organization; 2020 Mar 16 URL: https://www.who.int/publications/i/item/risk-communication-and-community-engagement-(rcce)-action-plan-guidance [accessed 2025-02-20]
9. COVID-19: vulnerable and high risk groups 2020. World Health Organization. URL: https://www.who.int/westernpacific/emergencies/covid-19/information/high-risk-groups [accessed 2025-02-20]
10. Katchova A. Principal component analysis. Econometrics Academy. 2017. URL: https://sites.google.com/site/econometricsacademy/econometrics-models/principal-component-analysis [accessed 2025-02-20]
11. Brouard S, Vasilopoulos P, Becher M. Sociodemographic and psychological correlates of compliance with the COVID-19 public health measures in France. Can J Pol Sci 2020 Dec;53(2):253-258. [doi: 10.1017/S0008423920000335]
12. Barber SJ, Kim H. COVID-19 worries and behavior changes in older and younger men and women. J Gerontol B 2021 Jan 18;76(2):e17-e23. [doi: 10.1093/geronb/gbaa068]
13. Ezati Rad R, Mohseni S, Kamalzadeh Takhti H, et al. Application of the protection motivation theory for predicting COVID-19 preventive behaviors in Hormozgan, Iran: a cross-sectional study. BMC Public Health 2021 Mar 8;21(1):466. [doi: 10.1186/s12889-021-10500-w] [Medline: 33685426]
14. Glanz K, Rimer BK, Viswanath K. Health Behavior and Health Education: Theory, Research, and Practice: John Wiley & Sons; 2008.
15. Kassegn A, Endris E. Review on socio-economic impacts of 'triple threats' of COVID-19, desert locusts, and floods in East Africa: evidence from Ethiopia. Cogent Social Sciences 2021 Jan 1;7(1):1885122. [doi: 10.1080/23311886.2021.1885122]
16. Nigussie H. The coronavirus intervention in Ethiopia and the challenges for implementation. Front Commun 2021;6:562512. [doi: 10.3389/fcomm.2021.562512]
17. Coronavirus disease (COVID-19) advice for the public: mythbusters. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters [accessed 2025-02-20]
18. Tenkorang EY. Myths and misconceptions about HIV transmission in Ghana: what are the drivers? Cult Health Sex 2013 Mar;15(3):296-310. [doi: 10.1080/13691058.2012.752107]
19. Carey JM, Chi V, Flynn DJ, Nyhan B, Zeitzoff T. The effects of corrective information about disease epidemics and outbreaks: evidence from Zika and yellow fever in Brazil. Sci Adv 2020 Jan;6(5):eaaw7449. [doi: 10.1126/sciadv.aaw7449] [Medline: 32064329]
20. Buli BG, Mayigane LN, Oketta JF, et al. Misconceptions about Ebola seriously affect the prevention efforts: KAP related to Ebola prevention and treatment in Kouroussa Prefecture, Guinea. Pan Afr Med J 2015;22(Suppl 1):11. [doi: 10.11694/pamj.supp.2015.22.1.6269] [Medline: 26740839]
21. Odlum M, Yoon S. Health information needs and health seeking behavior during the 2014-2016 Ebola outbreak: a Twitter content analysis. PLoS Curr 2018 Mar 23;10. [doi: 10.1371/currents.outbreaks.fa814fb2bec36e29b718ab6af66124fa] [Medline: 29707416]
22. Urbán R, Király O, Demetrovics Z. Who complies with coronavirus disease 2019 precautions and who does not? Curr Opin Psychiatry 2021 Jul 1;34(4):363-368. [doi: 10.1097/YCO.0000000000000723] [Medline: 34001700]
23. Casey BJ, Getz S, Galvan A. The adolescent brain. Dev Rev 2008;28(1):62-77. [doi: 10.1016/j.dr.2007.08.003] [Medline: 18688292]
24. Ahmed MAM, Siewe Fodjo JN, Gele AA, et al. COVID-19 in Somalia: adherence to preventive measures and evolution of the disease burden. Pathogens 2020 Sep 6;9(9):735. [doi: 10.3390/pathogens9090735] [Medline: 32899931]
25. Ahinkorah BO, Ameyaw EK, Hagan JE, Seidu AA, Schack T. Rising above misinformation or fake news in Africa: another strategy to control COVID-19 spread. Front Commun 2020;5:45. [doi: 10.3389/fcomm.2020.00045]
26. Correa T, Hinsley AW, de Zúñiga HG. Who interacts on the web?: The intersection of users' personality and social media use. Comput Human Behav 2010:247-253. [doi: 10.1016/j.chb.2009.09.003]

Abbreviations

EFA: exploratory factor analysis
ETB: Ethiopian Birr
KMO: Kaiser-Mayer-Olkin measure
PPS: probability proportional to size
RCCE: risk communication and community engagement
WHO: World Health Organization

XSL•FO
RenderX

Edited by A Mavragani; submitted 22.09.24; peer-reviewed by R Ibrahim; revised version received 10.12.24; accepted 11.12.24; published 24.04.25.

Please cite as:
Farah AM, Abdulahi A, Hussein A, Hussein AA, Osman A, Mohamud M, Mowlid H, Hailu G, Alwan F, Bizuneh EA, Ibrahim AM, Abdulahi E
COVID-19 Perceptions Among Communities Living on Ground Crossings of Somali Region of Ethiopia: Community Cross-Sectional Survey Study
JMIR Form Res 2025;9:e66751
URL: https://formative.jmir.org/2025/1/e66751
doi:10.2196/66751

© Alinoor Mohamed Farah, Abdifatah Abdulahi, Abdulahi Hussein, Ahmed Abdikadir Hussein, Abdi Osman, Mohamed Mohamud, Hasan Mowlid, Girum Hailu, Fathia Alwan, Ermiyas Abebe Bizuneh, Ahmed Mohammed Ibrahim, Elyas Abdulahi. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Synthetic Data-Driven Approaches for Chinese Medical Abstract Sentence Classification: Computational Study

Jiajia Li[1,2,3*], PhD; Zikai Wang[1,2,4*], PhD; Longxuan Yu[5], MSc; Hui Liu[6], PhD; Haitao Song[1,2], PhD

[1]Shanghai Artificial Intelligence Research Institute Co., Ltd, Shanghai, China

[2]Xiangfu Laboratory, Jiaxing, China

[3]School of Chemistry and Chemical Engineering, Shanghai Jiao Tong University, Shanghai, China

[4]Inner Mongolia Academy of Science and Technology, Hohhot, China

[5]University of California San Diego, San Diego, CA, United States

[6]Shanghai Civil Aviation College, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Haitao Song, PhD
Shanghai Artificial Intelligence Research Institute Co., Ltd, Shanghai, China

Abstract

Background: Medical abstract sentence classification is crucial for enhancing medical database searches, literature reviews, and generating new abstracts. However, Chinese medical abstract classification research is hindered by a lack of suitable datasets. Given the vastness of Chinese medical literature and the unique value of traditional Chinese medicine, precise classification of these abstracts is vital for advancing global medical research.

Objective: This study aims to address the data scarcity issue by generating a large volume of labeled Chinese abstract sentences without manual annotation, thereby creating new training datasets. Additionally, we seek to develop more accurate text classification algorithms to improve the precision of Chinese medical abstract classification.

Methods: We developed 3 training datasets (dataset #1, dataset #2, and dataset #3) and a test dataset to evaluate our model. Dataset #1 contains 15,000 abstract sentences translated from the PubMed dataset into Chinese. Datasets #2 and #3, each with 15,000 sentences, were generated using GPT-3.5 from 40,000 Chinese medical abstracts in the CSL database. Dataset #2 used titles and keywords for pseudolabeling, while dataset #3 aligned abstracts with category labels. The test dataset includes 87,000 sentences from 20,000 abstracts. We used SBERT embeddings for deeper semantic analysis and evaluated our model using clustering (SBERT-DocSCAN) and supervised methods (SBERT-MEC). Extensive ablation studies and feature analyses were conducted to validate the model's effectiveness and robustness.

Results: Our experiments involved training both clustering and supervised models on the 3 datasets, followed by comprehensive evaluation using the test dataset. The outcomes demonstrated that our models outperformed the baseline metrics. Specifically, when trained on dataset #1, the SBERT-DocSCAN model registered an impressive accuracy and F_1-score of 89.85% on the test dataset. Concurrently, the SBERT-MEC algorithm exhibited comparable performance with an accuracy of 89.38% and an identical F_1-score. Training on dataset #2 yielded similarly positive results for the SBERT-DocSCAN model, achieving an accuracy and F_1-score of 89.83%, while the SBERT-MEC algorithm recorded an accuracy of 86.73% and an F_1-score of 86.51%. Notably, training with dataset #3 allowed the SBERT-DocSCAN model to attain the best with an accuracy and F_1-score of 91.30%, whereas the SBERT-MEC algorithm also showed robust performance, obtaining an accuracy of 90.39% and an F_1-score of 90.35%. Ablation analysis highlighted the critical role of integrated features and methodologies in improving classification efficiency.

Conclusions: Our approach addresses the challenge of limited datasets for Chinese medical abstract classification by generating novel datasets. The deployment of SBERT-DocSCAN and SBERT-MEC models significantly enhances the precision of classifying Chinese medical abstracts, even when using synthetic datasets with pseudolabels.

(JMIR Form Res 2025;9:e54803) doi:10.2196/54803

KEYWORDS

medical abstract sentence classification; large language models; synthetic datasets; deep learning; Chinese medical; dataset; traditional Chinese medicine; global medical research; algorithm; robustness; efficiency; accuracy

XSL•FO
RenderX

Introduction

Background

In the realms of machine learning and artificial intelligence, the significance of data cannot be overstated, yet accessing authentic real-world datasets often presents substantial hurdles, including elevated costs, extended timeframes, and privacy issues. To navigate these obstacles, there is a growing pivot toward the utilization of synthetic datasets. While synthetic datasets have predominantly been associated with computer vision applications, the landscape is changing for natural language processing (NLP) due to advancements in text generation capabilities, largely attributed to the multihead self-attention mechanism integral to the transformer family of models. This trend began with the transformer itself [1] and has since given rise to models such as Bidirectional Encoder Representations from Transformers (BERT) [2], OpenAI GPT [3], Transformer-XL [4], OpenAI GPT-2 [5], and Grover [6]. In recent years, large language models have transformed the field of NLP, demonstrating exceptional performance in a wide range of tasks. These models generate increasingly coherent text, prompting researchers to explore the potential of synthetic datasets. Particularly noteworthy, Zellers et al [6] claimed that their Grover model for conditional text generation outperforms human-generated text in both style and content, especially in the "fake news" and "propaganda" categories, as evaluated by human raters. This highlights the promise of synthetic datasets to improve NLP performance and the continued need to develop text generation techniques.

Prior Work

Sentence-Level Text Classification

Text classification at the sentence level has been explored using various deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Kim [7] first proposed a single-layer CNN with pretrained word embeddings for text classification, achieving excellent results. Yang et al [8] developed a 2-level attention mechanism using gate recurrent units for document classification, while Conneau et al [9] introduced a CNN-nonstatic model with character-level CNNs and an average pooling layer. Recently, pretrained models, such as BERT [2], RoBERTa [10], XLNet [11], and ALBERT [12], have also been used for sentence-level text classification tasks. These transformer-based models generate sentence representations and can be combined with various classifiers to achieve state-of-the-art performance. Additionally, hybrid approaches using word-level and character-level CNNs initialized with ELMo [13] or BERT embeddings have been explored to improve the robustness and performance of sentence-level text classification models. Overall, pretrained models have significantly advanced the state-of-the-art in sentence-level text classification, and further research in this area is expected to yield even more sophisticated models.

Unsupervised Text Clustering

Unsupervised text clustering is an important task in NLP that groups similar text documents without relying on labeled examples. While traditional methods such as hierarchical agglomerative clustering [14], k-means clustering [15], nonnegative matrix factorization [16], and latent dirichlet allocation [17] have been widely used for this task, recent advances in pretrained embeddings have led to new and competitive methods. These include paragraph vector [18] and USE+KMeans [19], which have shown promising results in various text clustering benchmarks. More recently, BERT-based methods, such as BERT-EMD [20], have been proposed for unsupervised text clustering. Additionally, SBERT [21] and DocSCAN [22] are the most recent methods that generate high-quality text embeddings and use graph-based clustering for unsupervised document clustering. Together, these developments highlight the continued importance of unsupervised text clustering in NLP research, with recent methods based on pretrained embeddings, and SBERT in particular, showing promising results.

Synthetic Data

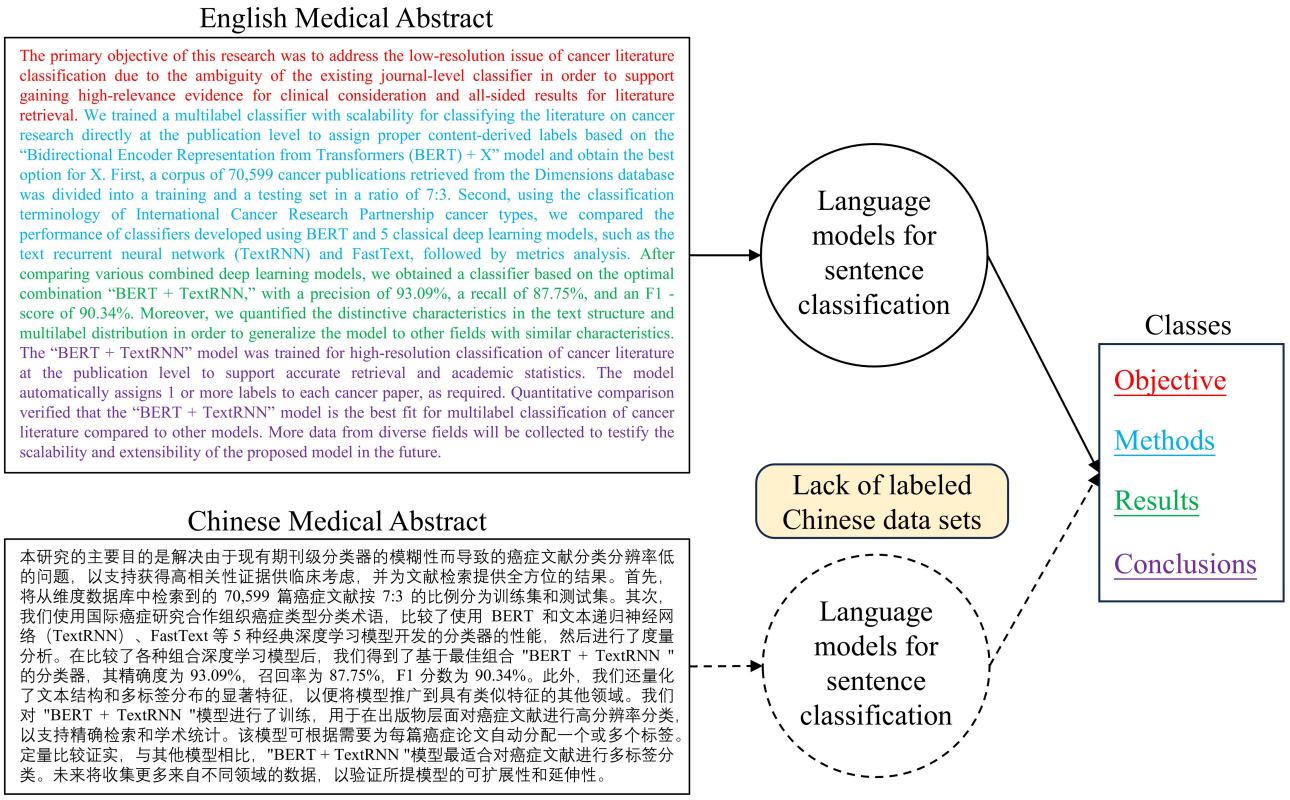
Synthetic data, generated using various techniques such as generative adversarial network (GAN)-based text generation and language model-based data augmentation, has become a popular way to expand small datasets in NLP. These synthetic datasets aim to improve model performance and generalization by providing additional examples for training. For instance, Xu et al [23] used a GAN-based method to generate synthetic data for image captioning, which showed promising results. Similarly, Wang and Wan [24] expanded GAN-based text generation to create synthetic datasets for sentiment analysis, which performed similarly to real data. Additionally, language model–based data augmentation, such as using GPT-2 for auto-completion, has been effective in generating synthetic data for NLP tasks [5]. Zhang et al [25] improved Chinese text classification using a language model–based data augmentation technique, while Zhou et al [26] showed that their language model–based data augmentation method improved low-resource language modeling. However, the use of synthetic data in NLP also has its limitations and potential drawbacks, which require further research. It is important to explore the advantages and disadvantages of synthetic datasets, as well as the different techniques used to generate them, to fully understand their impact on model performance and generalization.

The Goal of This Study

This paper aims to explore the creation and use of synthetic datasets to address the lack of real-world datasets for Chinese medical abstract classification, as there are currently no open-source datasets with sentence-level classification labels for Chinese medical abstracts, as shown in Figure 1. In response to this challenge, we have harnessed the capabilities of GPT-3.5 to generate distinct datasets for the classification of Chinese medical abstracts. Moreover, we designed and trained the clustering and supervised models for this task. Through this innovative approach, this study not only showcases the immense potential of synthetic datasets to bridge the gaps inherent in real-world datasets but also illuminates their profound impact on enhancing the performance of NLP tasks.

XSL•FO
RenderX

Figure 1. Goal of this study. It has been observed that Chinese medical abstract classification faces a significant number of limitations due to the lack of corresponding labeled datasets. Achieving accurate sentence-level classification in these abstracts will be instrumental in contributing Chinese medical information to the global medical field.



Methods

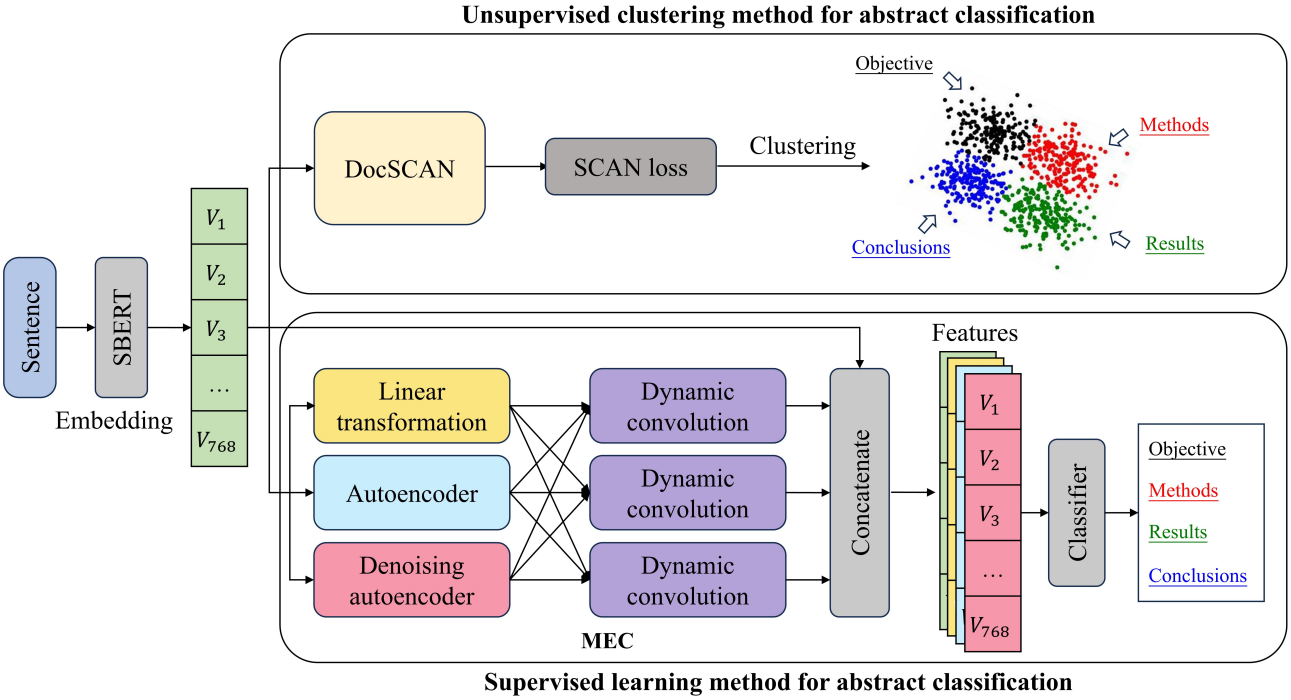
Overall Framework

The methodology used in this paper is structured into three distinct sections, systematically addressing the creation and application of synthetic datasets for Chinese medical abstract classification:

1. Synthetic dataset generation: the process begins by translating the PubMed dataset [27] into Chinese and generating 2 distinct synthetic abstract datasets based on provided titles, keywords, disciplines, and categories by GPT-3.5.
2. Unsupervised clustering: this step involves fine-tuning the SBERT model with the synthetic datasets and then using the DocSCAN algorithm to cluster the generated sentence embeddings in an unsupervised manner, termed SBERT-DocSCAN (Figure 2).
3. Supervised classification: a new supervised method, SBERT-MEC, is proposed to classify the synthetic data set, enhancing the ability to accurately categorize synthetic medical abstracts (Figure 2).

XSL•FO
RenderX

Figure 2. Framework of the proposed methods for abstract classification. The unsupervised clustering method (SBERT-DocSCAN) leverages SBERT embeddings and DocSCAN to group sentences into abstract sections based on clustering. The supervised learning method (SBERT-MEC) uses a multiencoder cascade (MEC) to enhance classification accuracy by extracting and integrating features for abstract sentence classification.



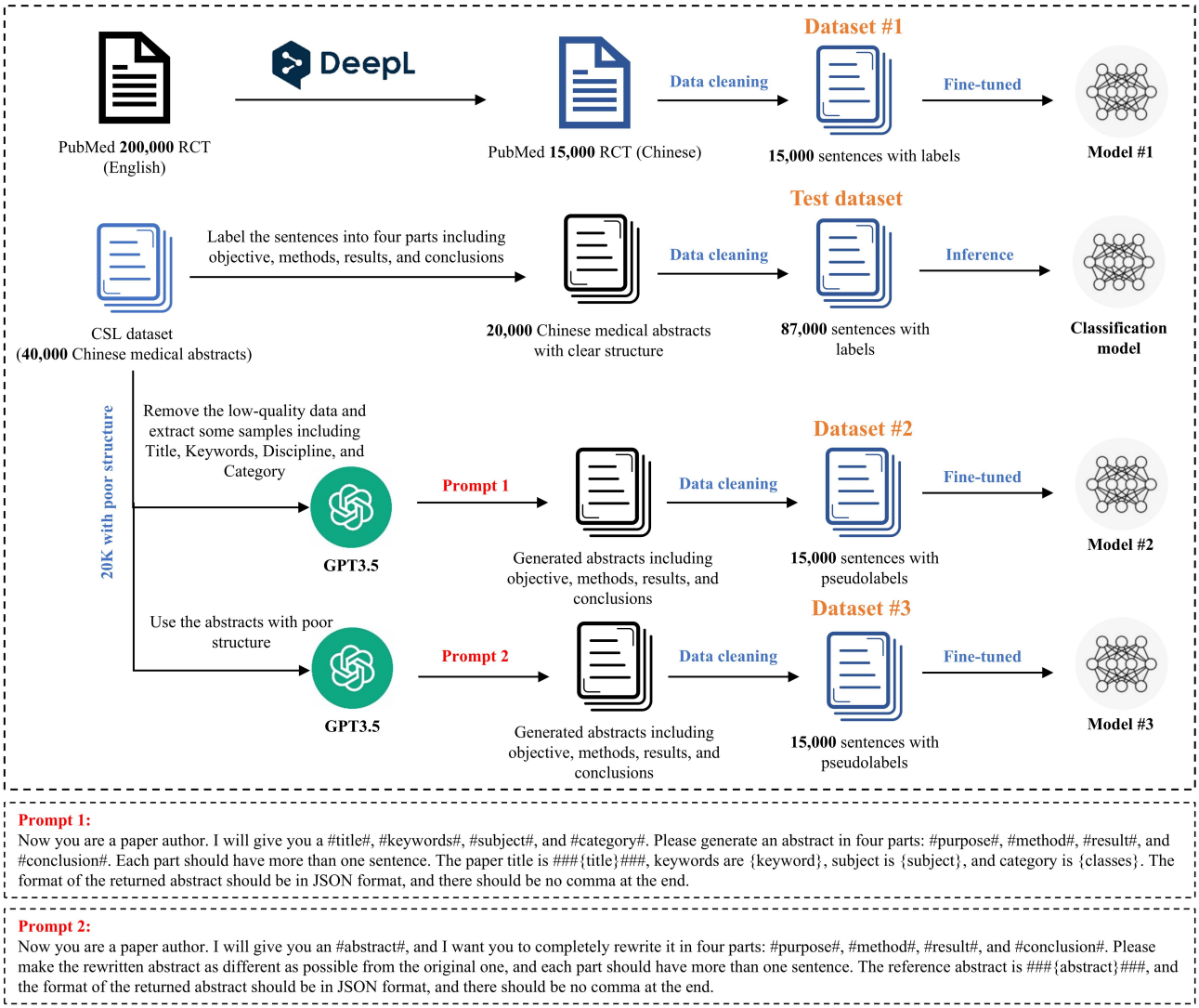
Synthetic Dataset Generation

As there is currently no Chinese dataset for sentence-level classification in the medical abstract field, in this step, we used OpenAI's text generation model, GPT-3.5, to generate 3 types of small synthetic datasets, which only contain around 15,000 sentences each yet still perform well on classification tasks. The first one is the translated PubMed dataset, which is translated into Chinese from the PubMed 200k RCT dataset by DeepL [28], and we choose 15,000 sentences with clear labels as dataset #1. The second dataset is created by using GPT-3.5 to generate abstracts based on title, keywords, discipline, and category from the CSL dataset [29]. The third dataset is also generated using GPT-3.5, but instead, it is paraphrased with rewritten abstracts assigned pseudolabels as dissimilar to the original text as possible.

Generating a diverse training dataset using large language models is a challenge. Even when the generation temperature is set to a high value, these models still tend to generate highly repetitive datasets that lack the diversity required for effective language model training. To address this issue, we selected the CSL corpus as the base corpus, which contains over 40,000 Chinese medical abstracts. Although these abstracts are not annotated at the sentence level, 20,000 of them have clear structure division into the 4 parts mentioned above. We extracted these 20,000 abstracts and manually labeled them as the test dataset. For the remaining 20,000 abstracts without clear structure division, we removed some low-quality data and the corpus of abstracts, then extracted a portion of them as the input corpus for GPT-3.5, which includes corresponding titles, keywords, subjects, and categories. These 4 types of data were inputted into GPT-3.5 to generate abstracts in 4 parts: purpose, method, results, and conclusion. These abstracts were then cleaned to produce a dataset of 15,000 sentences with pseudolabels (dataset #2). Similarly, the structure-less abstracts were inputted into GPT-3.5, which rewrote them to generate clear structure abstracts in 4 parts, resulting in a dataset of 15,000 sentences with pseudolabels (dataset #3). The construction process of the above datasets is shown in Figure 3.

XSL•FO
RenderX

Figure 3. Construction process of the synthetic datasets used for sentence-level classification in the medical abstract field. Dataset #1 is derived by translating the PubMed 200k RCT dataset into Chinese using DeepL and selecting 15,000 sentences with clear labels. Dataset #2 is generated by GPT-3.5, using titles, keywords, disciplines, and categories from the CSL dataset to produce structured abstracts divided into 4 parts (objective, methods, results, and conclusions). Dataset #3 is created by paraphrasing structure-less abstracts using GPT-3.5, generating diverse and rewritten abstracts with pseudolabels. The test dataset comprises 20,000 manually labeled abstracts from the CSL corpus with clear structural divisions. For each of the models in the paper (eg, SBERT), we train them on the training sets separately to produce corresponding ones (eg, SBERT #1, SBERT #2, and SBERT #3).



Prompt 1:
Now you are a paper author. I will give you a #title#, #keywords#, #subject#, and #category#. Please generate an abstract in four parts: #purpose#, #method#, #result#, and #conclusion#. Each part should have more than one sentence. The paper title is ###{title}###, keywords are {keyword}, subject is {subject}, and category is {classes}. The format of the returned abstract should be in JSON format, and there should be no comma at the end.

Prompt 2:
Now you are a paper author. I will give you an #abstract#, and I want you to completely rewrite it in four parts: #purpose#, #method#, #result#, and #conclusion#. Please make the rewritten abstract as different as possible from the original one, and each part should have more than one sentence. The reference abstract is ###{abstract}###, and the format of the returned abstract should be in JSON format, and there should be no comma at the end.

Unsupervised Clustering

In the second part, we fine-tuned the sentence transformer and then used the DocSCAN algorithm to cluster the synthetic datasets. We chose the sbert-chinese-general-v2 model, which is a model pretrained on the SimCLUE dataset [30], as the base model due to its outstanding performance on embedding Chinese sentences. We fine-tuned the pretrained model on the 3 datasets into 3 models, which is quite crucial as it leads to the overall distribution more inclined toward the functional aspects of each sentence rather than the literal content. We then used these fine-tuned models to embed the input data, followed by clustering by the DocSCAN algorithm which is an algorithm relying on the intuition that a datapoint and its nearest neighbors in representation space often share the same class label, and outperform others on unsupervised classification tasks. Specifically, in this part, we generated embeddings for each sentence in the test set using the 3 fine-tuned sentence transformers in the first part and then used the DocSCAN algorithm to cluster these embeddings for unsupervised abstract classification, which uses the SCAN loss LS to fine-tune the model f(x), defined as:

$$LS=-1|D|\sum x\in D\sum k\in Nx\log(f(x)\cdot f(k))+\lambda\sum i\in Cpilog(pi)$$

which can be broken down into a consistency loss and an auxiliary loss. The consistency loss aims to ensure that the same label is assigned to a data point and its neighbors. Our model f(·) computes a label for a given data point x from the dataset *D*, as well as for each data point *k* in the set of mined neighbors from *x* in Nx. To achieve this, we calculate the dot product between the output distribution (normalized by softmax(·) of datapoint x and its neighbor k. The auxiliary loss is used to achieve regularization via entropy, scaled by a weight λ. This loss encourages the model to distribute probability mass across all clusters C, where pi denotes the assigned probability of cluster i in C by the model. Without this term, there could be a shortcut where all examples collapse into a single cluster. The entropy term ensures that the distribution of class labels resulting from applying DocSCAN tends to be roughly uniform.

XSL•FO
RenderX

Supervised Classification

In the third part, we develop a supervised learning method, SBERT-MEC, which leverages the proposed multiencoder cascade (MEC) module for feature extraction, designed to enhance the classification of synthetic datasets through refined weak supervision. Figure 2 shows the framework of SBERT-MEC and the modeling process. The first step of the SBERT-MEC is to input the sentence embeddings generated by SBERT into the MEC module, which includes 3 encoders: linear transformation, autoencoder model, and denoising autoencoder model. Formally, given a 768-dimensional sentence embedding V, the linear transformation V transforms the sentence embedding V into W by adding a small random number n:

$$W=LV=V1+n,V2+n,\ldots,V768+n$$

The autoencoder model can be represented as follows:

$$W=AE\theta(V)$$

where AE is a fully connected network that maps V into a duplicate embedding W and θ is the parameters of AE. The autoencoder is trained by duplicate pairs from the Quora dataset [31]. The denoising autoencoder (DA), in another way, adds Gaussian noise to the sentence embedding and forms a noisy embedding VN, and the training process forces output W to be as close to the original embedding V as possible:

$$VN=V+\ ,\ \sim N0,\sigma 2I$$

$$W = DA\theta(VN)$$

To further enhance performance, 3 embeddings are input into the dynamic convolutional layer for mutual supervision. Specifically, for each embedding, the embeddings generated by the other 2 encoders are concatenated together and used as inputs to supervise this embedding. Finally, 3 output embeddings of the dynamic convolutional layers are concatenated with the original embedding and then put into a classifier to predict test data. This process is repeated for the 3 models fine-tuned in the first part respectively.

To summarize, first, we create synthetic datasets by different methods and leverage them to fine-tune sentence transformers. These datasets are small in size while effective in training classification tasks. Second, fine-tuned models are used on an unsupervised classification algorithm DocSCAN to analyze the performance of these datasets. Further, fine-tuned models are used on a weakly supervised algorithm SBERT-MEC proposed by us, which performs data augmentation on embeddings. This process demonstrates the performances of small synthetic datasets again, as well as the effectiveness of our algorithm.

Ethical Considerations

This study did not involve human participants, animal subjects, or personally identifiable data. All data used in the research were either synthetic or publicly available, ensuring compliance with ethical guidelines. As such, no ethical approval or informed consent was required. The study adheres to principles of academic integrity and research transparency.

Results

Dataset

For empirical evaluation, we crafted 3 datasets: 2 synthetic ones generated using GPT-3.5, and a pre-existing dataset. Each synthetic dataset comprises approximately 15,000 sentences pertinent to medical abstracts. These datasets underwent a meticulous generation process, including cleaning and pseudolabel assignment for sentences within each abstract section. Detailed methodology for creating the synthetic datasets is outlined in the section on Synthetic Dataset Generation. Table 1 illustrates the distribution of sentences across various sections within each dataset.

Table . Distribution of development datasets. Dataset #1 (translated PubMed dataset with labeled sentences), dataset #2 (GPT-3.5-generated abstracts with pseudolabels), and dataset #3 (paraphrased GPT-3.5-generated abstracts). Each dataset contains 15,000 sentences, divided into 4 sections: objective, methods, results, and conclusion.

Dataset	Classes of sentences				Total
	Objective	Methods	Results	Conclusion	
Dataset #1	3750	3750	3750	3750	15,000
Dataset #2	3750	3750	3750	3750	15,000
Dataset #3	3537	4140	3826	3497	15,000

Baselines

We selected multiple algorithms of different kinds as baselines to demonstrate the performance of our algorithms. To begin with, we used classical non-BERT models. TextCNN is built upon the CNN paradigm, applying multiple convolutional filters and pooling operations to capture local features of varying lengths within the text. TextRNN [32] leverages RNNs to model the sequential information in the text, capturing dependencies between words. TextRNN-Att [33] extends TextRNN by incorporating attention mechanisms, enabling the model to focus on essential words. TextRCNN [34] combines convolutional and RNNs, simultaneously considering word order and contextual information. DPCNN [35] uses multiple layers of convolution and pooling operations to capture hierarchical text features. FastText [36] is a simple and efficient text classification method based on the bag-of-words model, representing text at the word level.

Additionally, we used BERT-based methods. BERT is a transformer-based pretrained language model, capable of learning rich contextual representations. BERTCNN [37] and BERT-RNN [8] integrate CNNs and RNNs, respectively, on top of BERT, combining the contextual representations from BERT with the ability to extract local features. BERT-RCNN

XSL•FO
RenderX

[38] combines BERT with both convolutional and RNNs, simultaneously considering word order and contextual information. BERT-DPCNN [39] incorporates a deep pyramid CNN on top of BERT, combining the contextual representations from BERT with multilevel feature extraction capabilities. ERNIE [40] is another transformer-based pretrained language model, built upon BERT with further improvements and optimizations, with better performance in Chinese contexts.

In our ablation study, we conducted an experiment using the k-means algorithm to cluster SBERT embeddings for comparative analysis with DocSCAN. Additionally, we examined the impact of omitting the dynamic convolutional layer in the MEC module. Subsequent experiments involving BERT serve as an extension of this analysis, further exploring the effects of removing 3 encoders within MEC.

Experiment Settings

In this study, we used the PyTorch framework and Python for all experiments. Data processing and model training were performed on an NVIDIA GeForce RTX 3090 (24G) GPU and an Intel(R) Core(TM) i9-10900K CPU @3.70GHz. Experimental parameters are detailed in Table 2. To mitigate overfitting, a dropout rate of 0.5 was used, randomly deactivating 50% of neuron connections during training. We conducted training over 30 epochs using either the Adam or BertAdam optimization algorithms, with a learning rate set at 5e-5. Training batches comprised 64 samples each, and the RNN featured a hidden size of 256. The CNN component used filter sizes of (2, 3, 4), with a total of 256 filters. Input sequence length was capped at 64 tokens, and the feature dimension was maintained at 768. Pretrained models used in the experiments included BERT (bert-base-chinese), ERNIE (ernie,-3.0-base-zh), and SBERT (sbert-chinese-general-v2), which were critical to the study's execution and result analysis.

Table . Settings for training models.

Parameter	Setting
Dropout	0.5
Epoch number	30
Optimizer	Adam/BertAdam
Learning rate	e−5
Batch size	64
RNN[a] hidden size	[2]
CNN[b] filter number	256
Max length	64
Feature dimension	768
BERT[c]	Bert-base-chinese
ERNIE	Ernie-3.0-base-zh
SBERT	Sbert-chinese-general-v2

[a]RNN: recurrent neural network.

[b]CNN: convolutional neural network.

[c]BERT: Bidirectional Encoder Representations from Transformers.

Evaluation Metrics

The experiments used accuracy and F_1-score to evaluate the model performance. The formulas for each index are as follows:

Accuracy= TP+TNTP+TN+FP+FN

where TP (true positive) corresponds to the number of instances correctly predicted as positive; TN (true negative) corresponds to the number of instances correctly predicted as negative; FP (false positive) corresponds to the number of instances incorrectly predicted as positive; FN (false negative) corresponds to the number of instances incorrectly predicted as negative.

Precision= TPTP+FP

Recall= TPTP+FN

F1-score= 2×Precision×RecallPrecision+Recall

In the above formulations, the Precision measures the proportion of correctly predicted positive instances out of the total instances predicted as positive and Recall measures the proportion of correctly predicted positive instances out of the total actual positive instances. F1-score is the harmonic mean of precision and recall, providing a single value that considers both metrics. This technique is highly effective for imbalanced datasets, where disparities in class sizes can skew model performance. It helps ensure fair representation of all classes, enhancing model reliability.

Model Performance

Table 3 shows the performance of various algorithms across different training datasets when evaluated on the test dataset. Notably, when trained on dataset #1, the SBERT-DocSCAN algorithm emerges as the leading performer, securing an

XSL•FO
RenderX

accuracy and F_1-score of 0.8985 on the test dataset. This standout performance highlights the algorithm's capability to classify medical domain data with high precision. Additionally, the SBERT-MEC algorithm also displays comparable performance on the same dataset, with an accuracy and F_1-score of 0.8938, making it the second most effective algorithm in our evaluation. For Data se t#2, the SBERT-DocSCAN and SBERT-MEC algorithms again demonstrate superior performance. SBERT-DocSCAN leads with exceptional accuracy and F_1-score of 0.8983, reinforcing its effectiveness in managing generated data. Meanwhile, SBERT-MEC remains a strong contender with an accuracy and F_1-score of 0.8673, marking it as the second most proficient algorithm for this dataset. Furthermore, considering dataset #3, SBERT-DocSCAN and SBERT-MEC algorithms consistently demonstrate outstanding performance. SBERT-DocSCAN achieves the highest accuracy and F_1-score of 0.9130, affirming its effectiveness in handling paraphrased data. Similarly, the SBERT-MEC algorithm achieves the second-highest accuracy and F_1-score of 0.9039 and 0.9035, respectively, highlighting its competence in dealing with paraphrased data. Additionally, Table 4 illustrates the confusion matrices of the SBERT-DocSCAN model on the test dataset using different training datasets.

In summary, the comprehensive analysis of the table elucidates that the SBERT-DocSCAN and SBERT-MEC algorithms consistently outperform other algorithms across multiple datasets. The superior performance of these algorithms in terms of accuracy and F1 scores underscores their significance and efficacy in the domain of text classification.

Table . Comparative performance of various models on the test dataset[a].

Method	Training on dataset #1		Training on dataset #2		Training on dataset #3	
	Accuracy	F_1-score	Accuracy	F_1-score	Accuracy	F_1-score
TextCNN	0.6122	0.6074	0.7843	0.7795	0.6789	0.6755
TextRNN	0.5045	0.5413	0.7315	0.7321	0.7295	0.7239
TextRNN-Att	0.6439	0.6421	0.7114	0.6955	0.7521	0.7521
TextRCNN	0.6562	0.6543	0.7642	0.7643	0.7439	0.7392
DPCNN	0.6953	0.6967	0.6947	0.6988	0.6758	0.6759
FastText	0.6562	0.6543	0.7208	0.7101	0.7391	0.7371
BERT[b]	0.7842	0.7857	0.8426	0.8400	0.7348	0.7083
BERT-CNN	0.7338	0.7340	0.8242	0.8213	0.7181	0.7136
BERT-RNN	0.8548	0.8546	0.8536	0.8532	0.7955	0.7882
BERT-RCNN	0.8245	0.8253	0.8313	0.8252	0.8364	0.8332
BERT-DPCNN	0.7839	0.7850	0.8186	0.8186	0.8301	0.8304
ERNIE	0.8801	0.8808	0.8681	0.8675	0.8895	0.8882
SBERT-Kmeans	0.8875	0.8875	0.8709	0.8709	0.8788	0.8788
SBERT-DocSCAN	0.8985	0.8985	0.8983	0.8983	0.9130	0.9130
SBERT-MEC	0.8938	0.8938	0.8673	0.8651	0.9039	0.9035

[a]It includes traditional models, BERT-based improved models, and the model proposed in this paper. These models were trained on 3 distinct datasets and subsequently evaluated on the test dataset in terms of accuracy and F_1-score.

[b]BERT: Bidirectional Encoder Representations from Transformers.

XSL•FO
RenderX

Table . Confusion matrices for the SBERT-DocSCAN model on the test dataset[a].

Label or cluster	0	1	2	3	Total	Recall
Model training on dataset #1						
Objective	20,302	1304	25	490	22,051	0.9175
Methods	1533	19,825	448	137	21,943	0.9035
Results	179	901	18,676	2075	21,831	0.8555
Conclusions	649	280	861	19,914	21,704	0.9175
Total	22,593	22,310	20,010	22,616	78,647[b]	0.8985
Precision	0.8595	0.8886	0.9333	0.8805	0.8985	—[c]
Model training on dataset #2						
Objective	20,250	1247	24	530	22,051	0.9183
Methods	705	20,496	397	345	21,943	0.934
Results	210	626	18,005	2990	21,831	0.8247
Conclusions	102	141	1585	19,876	21,704	0.8983
Total	21,267	22,510	20,011	23,741	78,647[b]	0.8983
Precision	0.9522	0.9105	0.8998	0.8372	0.8983	—
Model training on dataset #3						
Objective	20,401	251	129	1447	22,051	0.9178
Methods	505	19,673	1325	29	21,943	0.9137
Results	288	1655	19,864	593	21,831	0.8868
Conclusions	749	252	386	19,982	21,704	0.9131
Total	21,943	21,831	21,704	22,051	78,647[b]	0.913
Precision	0.9297	0.9012	0.9152	0.9062	0.913	—

[a]This table displays the confusion matrices for the SBERT-DocSCAN model, which was trained separately on 3 distinct training datasets. The matrices detail the accuracy and recall for each category when evaluated on the test dataset.

[b]Out of 87,529.

[c]Not applicable.

Ablation Study

To assess the impact of various submodules, we conducted ablation studies. We began by training 2 clustering models with identical structures (SBERT-DocSCAN). The key distinction between them was that 1 model was fine-tuned on our datasets, while the other was not. In addition, we trained 3 supervised models on different datasets and evaluated their performance on the test dataset. Notably, 1 of these models was the SBERT-MEC, which lacked the proposed dynamic convolution (DC) module. The results, presented in Table 5, clearly demonstrate that the SBERT-DocSCAN method, when fine-tuned with SBERT, outperforms others in terms of efficiency. Furthermore, within the supervised learning category, the SBERT-MEC model equipped with the DC module surpassed those lacking this module, underlining the value of the DC module in enhancing model performance.

XSL•FO
RenderX

Table . Performance of ablation model on test dataset[a].

Method	Training on dataset #1		Training on dataset #2		Training on dataset #3	
	Accuracy	F_1-score	Accuracy	F_1-score	Accuracy	F_1-score
SBERT-DocSCAN (without finetune)	0.3891	0.3895	0.3891	0.3895	0.3891	0.3895
SBERT-DocSCAN (with finetune)	0.8985	0.8985	0.8983	0.8983	0.9130	0.9130
BERT[b]	0.8330	0.8324	0.8424	0.8421	0.8406	0.8397
SBERT-MEC[c] (without DC)[d]	0.8714	0.8717	0.8390	0.8332	0.8736	0.8735
SBERT-MEC (with DC)	0.8938	0.8938	0.8673	0.8651	0.9039	0.9035

[a]This table compares the performance of 2 SBERT-DocSCAN clustering models—one fine-tuned on our datasets and the other not—and 3 supervised models, including an SBERT-MEC model without the dynamic convolution module. All models were evaluated on the test dataset.

[b]BERT: Bidirectional Encoder Representations from Transformers.

[c]MEC: multiencoder cascade.

[d]DC: dynamic convolution.

Discussion

Principal Findings

This study addresses the critical lack of real-world datasets for Chinese medical abstract classification by leveraging GPT-3.5 to generate synthetic datasets and developing models tailored for this task. Our findings confirm that synthetic datasets, when carefully designed, can match or surpass the performance of manually labeled datasets in sentence-level classification tasks. The SBERT-DocSCAN and SBERT-MEC models developed in this study demonstrate the potential of clustering and supervised approaches to effectively classify Chinese medical abstracts, illustrating the profound impact of synthetic datasets on enhancing NLP tasks.

Classical non-BERT models, such as TextCNN [7], TextRNN [32], and TextRCNN [34], exhibit strong performance in capturing local or sequential features. However, these models lack the deep contextual understanding provided by transformer-based models. While FastText [36] offers a lightweight and efficient alternative, its bag-of-words approach limits its ability to capture complex semantic relationships. These limitations underscore the advantages of leveraging pretrained language models in tasks that demand rich contextual understanding.

BERT-based models significantly improve performance by providing deep contextual representations. The success of models like BERT-CNN [37] and BERT-RCNN [38] aligns with prior studies, which highlight the effectiveness of combining BERT embeddings with convolutional and recurrent structures for enhanced local and sequential feature extraction. However, our proposed SBERT-DocSCAN and SBERT-MEC models outperform these baselines, indicating the added value of advanced clustering methods and DC layers. Specifically, DocSCAN [22], with its graph-based clustering approach, demonstrates superior clustering quality compared to k-means.

The ablation study further emphasizes the contributions of individual components in our models. The removal of the dynamic convolutional layer in the MEC module resulted in a notable decline in performance, highlighting its role in refining language representations. Similarly, the omission of the 3 encoders within MEC led to a significant reduction in accuracy, underscoring the importance of multiencoder architecture in capturing diverse linguistic features. These findings align with the broader NLP literature [39], which emphasizes the benefits of combining multiple feature extraction techniques for enhanced model performance.

Our exploration of ERNie [40], a BERT-based model optimized for Chinese contexts, further validates the importance of leveraging models tailored to specific languages and domains. While ERNie offers improvements over standard BERT in Chinese text classification, it does not surpass the performance of our SBERT-based approaches. This suggests that task-specific architectural innovations, such as the integration of clustering methods and DC modules, can provide greater benefits than domain-specific pretraining alone.

Overall, these comparisons and ablation studies highlight the robustness and versatility of our proposed methods, providing valuable insights into the design of models for sentence-level text classification tasks. By demonstrating the effectiveness of integrating advanced clustering methods, DC layers, and multiencoder architectures, this study contributes to the growing body of research focused on optimizing transformer-based models for real-world applications.

Limitations

While promising, our proposed methods have 2 main limitations. First, the quality of synthetic data heavily relies on the GPT-3.5 model, with performance contingent on effective prompt design. Refining prompt engineering strategies will be essential for future improvements. Second, the 2-stage approach used in both models results in an increased parameter size, raising concerns about computational resource efficiency. Further optimization

XSL•FO
RenderX

is necessary to address these scalability challenges, especially in real-world applications where resources may be limited.

Conclusions

This study demonstrates the significant potential of synthetic datasets, generated using GPT-3.5, in addressing the scarcity of labeled datasets for Chinese medical abstract classification. Our findings reveal that compact synthetic datasets can achieve performance comparable to, and in some cases surpass, that of manually labeled datasets. The proposed SBERT-DocSCAN and SBERT-MEC models further highlight the benefits of combining advanced clustering techniques, multiencoder architectures, and DC modules, showcasing their ability to enhance sentence-level classification tasks in specialized domains. These contributions provide valuable insights for leveraging generative artificial intelligence in NLP applications.

Beyond this specific task, this work underscores the transformative potential of synthetic datasets in reducing reliance on costly manual labeling, enabling broader adoption of NLP technologies in resource-limited fields. Future research can expand upon these findings by exploring more sophisticated data generation strategies, optimizing model architectures for efficiency, and fostering interdisciplinary collaborations to develop tailored solutions for complex, real-world challenges. By bridging gaps in data availability, this study provides a foundation for advancing NLP capabilities in medical and other specialized domains.

Acknowledgments

This research was supported by Shanghai Science and Technology Commission Research Project 24YF2731300, 23692107300, and 202121511101200, Postdoctoral Fellowship Program of CPSF under grant GZC20231604, the Xiangfu Lab Youth Program XF052024B0100, and Shanghai Municipal Commission of Economy and Information Technology (SHEITC) Project 2024-GZL-RGZN-01020.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

References

1. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Presented at: 31st International Conference on Neural Information Processing Systems (NeurIPS); Dec 4-9, 2017; Long Beach, CA, United States. [doi: 10.5555/3295222.3295349]
2. Devlin J, Chang MW, Lee K, Toutanova K. B. Pre-training of deep bidirectional transformers for language understanding. Presented at: 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; Jun 2-7, 2019; Minneapolis, MN, United States. [doi: 10.18653/v1/N19-1423]
3. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. OpenAI. 2020. URL: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf [accessed 2020-09-25]
4. Dai Z, Yang Z, Yang Y, Carbonell J, Le Q, Salakhutdinov R. Transformer-XL: attentive language models beyond a fixed-length context. Presented at: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics; Jul 28 to Aug 2, 2019; Florence, Italy. [doi: 10.18653/v1/P19-1285]
5. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. Open AI. URL: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf [accessed 2025-03-03]
6. Zellers R, Holtzman A, Rashkin H, Bisk Y, Farhadi A, Roesner F, et al. Defending against neural fake news. Presented at: 33rd International Conference on Neural Information Processing Systems (NeurIPS); Dec 8-14, 2019; Vancouver, BC, Canada. [doi: 10.5555/3454287.3455099]
7. Kim Y. Convolutional neural networks for sentence classification. Presented at: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); Oct 25-29, 2014; Doha, Qatar. [doi: 10.3115/v1/D14-1181]
8. Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E. Hierarchical attention networks for document classification. Presented at: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics; Jun 12-17, 2016; San Diego, CA, United States. [doi: 10.18653/v1/N16-1174]
9. Conneau A, Kiela D, Schwenk H, Barrault L, Bordes A. Supervised learning of universal sentence representations from natural language inference data. Presented at: ; Sep 9-11, 2017; Copenhagen, Denmark. [doi: 10.18653/v1/D17-1070]
10. Yinhan L, Myle O, Naman G, Jingfei D, Mandar J, Danqi C, et al. RoBERTa: a robustly optimized BERT pretraining approach. arXiv. Preprint posted online on Jul 26, 2019. [doi: 10.48550/arXiv.1907.11692]

XSL•FO
RenderX

11. Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov RR, Le QV. Xlnet: generalized autoregressive pretraining for language understanding. Presented at: 33rd International Conference on Neural Information Processing Systems (NeurIPS); Dec 8-14, 2019; Vancouver, BC, Canada. [doi: 10.5555/3454287.3454804]
12. Lan Z, Chen M, Goodman S, Gimpel K, Sharma P, Soricut R. Albert: a lite BERT for self-supervised learning of language representations. arXiv. Preprint posted online on Sep 26, 2019. [doi: 10.48550/arXiv.1909.11942]
13. Peters M, Neumann M, Iyyer M, et al. Deep contextualized word representations. Presented at: 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; Jun 1-6, 2018; New Orleans, LA, United States. [doi: 10.18653/v1/N18-1202]
14. Ward Jr JH. Hierarchical grouping to optimize an objective function. J Am Stat Assoc 1963 Mar;58(301):236-244. [doi: 10.1080/01621459.1963.10500845]
15. MacQueen J. Some methods for classification and analysis of multivariate observations. Presented at: 5th Berkeley Symposium on Mathematical Statistics and Probability; Dec 27, 1965 to Jan 7, 1966; Oakland, CA, USA.
16. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nat New Biol 1999 Oct 21;401(6755):788-791. [doi: 10.1038/44565] [Medline: 10548103]
17. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003 Mar 1;3:993-1022. [doi: 10.5555/944919.944937]
18. Le Q, Mikolov T. Distributed representations of sentences and documents. Presented at: 31st International Conference on International Conference on Machine Learning (ICML); Jun 21-25, 2014; Beijing, China. [doi: 10.5555/3044805.3045025]
19. Cer D, Yang Y, Kong SY, et al. Universal sentence encoder for english. Presented at: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations; Oct 31 to Nov 4, 2018; Brussels, Belgium. [doi: 10.18653/v1/D18-2029]
20. Li J, Liu X, Zhao H, Xu R, Yang M, Jin Y. BERT-EMD: many-to-many layer mapping for BERT compression with earth mover's distance. Presented at: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP); Nov 16-20, 2020. [doi: 10.18653/v1/2020.emnlp-main.242]
21. Reimers N, Gurevych I. Sentence-bert: sentence embeddings using siamese bert-networks. Presented at: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); Nov 3-7, 2019; Hong Kong, China. [doi: 10.18653/v1/D19-1410]
22. Stammbach D, Ash E. DocSCAN: unsupervised text classification via learning from neighbors. Presented at: Proceedings of the 18th Conference on Natural Language Processing (KONVENS 2022); Sep 12-15, 2022; Potsdam, Germany URL: https://aclanthology.org/2022.konvens-1.4 [accessed 2025-03-03]
23. Xu T, Zhang P, Huang Q, et al. AttnGAN: fine-grained text to image generation with attentional generative adversarial networks. Presented at: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); Jun 18-23, 2018; Salt Lake City, UT, United States. [doi: 10.1109/CVPR.2018.00143]
24. Wang K, Wan X. SentiGAN: generating sentimental texts via mixture adversarial networks. Presented at: 27th International Joint Conference on Artificial Intelligence (IJCAI); Jul 13-19, 2018; Stockholm, Sweden. [doi: 10.5555/3304222.3304388]
25. Zhang X, Zhao J, LeCun Y. Character-level convolutional networks for text classification. Presented at: 28th International Conference on Neural Information Processing Systems (NeurIPS); Dec 8-13, 2014; Montreal, QC, Canada. [doi: 10.5555/2969239.2969312]
26. Zhou R, Li X, He R, Bing L, Cambria E, Si L, et al. MELM: data augmentation with masked entity language modeling for low-resource NER. Presented at: The 60th Annual Meeting of the Association for Computational Linguistics; May 22-27, 2022; Dublin, Ireland. [doi: 10.18653/v1/2022.acl-long.160]
27. PubMed. URL: https://pubmed.ncbi.nlm.nih.gov [accessed 2023-10-20]
28. DeepL. URL: https://www.deepl.com/translator [accessed 2023-10-20]
29. Yudong L, Yuqing Z, Zhe Z, Linlin S, Weijie L, Weiquan M, et al. CSL: a large-scale chinese scientific literature dataset. Presented at: 29th International Conference on Computational Linguistics; Oct 12-17, 2022; Gyeongju, Republic of Korea.
30. CLUEbenchmark/SimCLUE. GitHub. URL: https://github.com/CLUEbenchmark/SimCLUE [accessed 2024-01-08]
31. Aghaebrahimian A. Quora question answer dataset. Presented at: 20th International Conference on Text, Speech, and Dialogue; Aug 27-31, 2017; Prague, Czech Republic. [doi: 10.1007/978-3-319-64206-2_8]
32. Pengfei L, Xipeng Q, Xuanjing H. Recurrent neural network for text classification with multi-task learning. Presented at: 25th International Joint Conference on Artificial Intelligence (IJCAI); Jul 9-15, 2016; New York City, NY, United States. [doi: 10.5555/3060832.3061023]
33. Zhou P, Shi W, Tian J, et al. Attention-based bidirectional long short-term memory networks for relation classification. Presented at: The 54th Annual Meeting of the Association for Computational Linguistics; Aug 7-12, 2016; Berlin, Germany. [doi: 10.18653/v1/P16-2034]
34. Wang R, Li Z, Cao J, Chen T, Wang L. Convolutional recurrent neural networks for text classification. Presented at: 2019 International Joint Conference on Neural Networks (IJCNN); Jul 14-19, 2019; Budapest, Hungary. [doi: 10.1109/IJCNN.2019.8852406]
35. Johnson R, Zhang T. Deep pyramid convolutional neural networks for text categorization. Presented at: 55th Annual Meeting of the Association for Computational Linguistics; Jul 30 to Aug 4, 2017; Vancouver, Canada. [doi: 10.18653/v1/P17-1052]

XSL•FO
RenderX

36. Joulin A, Grave E, Bojanowski P, Mikolov T. Bag of tricks for efficient text classification. Presented at: 15th Conference of the European Chapter of the Association for Computational Linguistics; Apr 3-7, 2017; Valencia, Spain. [doi: 10.18653/v1/E17-2068]
37. Kaur K, Kaur P. BERT-CNN: improving BERT for requirements classification using CNN. Procedia Comput Sci 2023;218:2604-2611. [doi: 10.1016/j.procs.2023.01.234]
38. Kaur K, Kaur P. BERT-RCNN: an automatic classification of app reviews using transfer learning based RCNN deep model. Research Square. Preprint posted online on Jan 24, 2023. [doi: 10.21203/rs.3.rs-2503700/v1]
39. Yanjun L, Haijun Z, Weimin P, Rujia F. Microblog rumor detection based on bert-DPCNN. In: Artificial Intelligence in China 2021:524-530. [doi: 10.1007/978-981-15-8599-9_60]
40. Zhang Z, Han X, Liu Z, Jiang X, Sun M, Liu Q. ERNIE: enhanced language representation with informative entities. Presented at: 57th Annual Meeting of the Association for Computational Linguistics; Jul 28 to Aug 2, 2019; Florence, Italy. [doi: 10.18653/v1/P19-1139]

Abbreviations

BERT: Bidirectional Encoder Representations from Transformers
CNN: convolutional neural network
DC: dynamic convolution
GAN: generative adversarial network
MEC: multiencoder cascade
NLP: natural language processing
RNN: recurrent neural network

Edited by A Mavragani; submitted 14.04.24; peer-reviewed by Y Wang; revised version received 28.11.24; accepted 28.11.24; published 19.03.25.

Please cite as:
Li J, Wang Z, Yu L, Liu H, Song H
Synthetic Data-Driven Approaches for Chinese Medical Abstract Sentence Classification: Computational Study
JMIR Form Res 2025;9:e54803
URL: https://formative.jmir.org/2025/1/e54803
doi:10.2196/54803

© Zikai Wang, Longxuan Yu, Hui Liu, Haitao Song. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Identifying High-Priority Ethical Challenges for Precision Emergency Medicine: Nominal Group Study

Christian Rose[1], MD; Emily Shearer[2], MD; Isabela Woller[3], BSc; Ashley Foster[4], MD; Nicholas Ashenburg[1], MD; Ireh Kim[1]; Jennifer Newberry[1], MS, MD, JD

[1]Department of Emergency Medicine, Stanford University School of Medicine, 500 Pasteur Dr, Stanford, CA, United States

[2]Department of Emergency Medicine, Alpert School of Medicine, Brown University, Providence, RI, United States

[3]Loyola University Chicago, Chicago, IL, United States

[4]Department of Emergency Medicine, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Christian Rose, MD
Department of Emergency Medicine, Stanford University School of Medicine, 500 Pasteur Dr, Stanford, CA, United States

Abstract

Background: Precision medicine promises to revolutionize health care by providing the right care to the right patient at the right time. However, the emergency department's unique mandate to treat "anyone, anywhere, anytime" creates critical tensions with precision medicine's requirements for comprehensive patient data and computational analysis. As emergency departments serve as health care's safety net and provide a growing proportion of acute care in America, identifying and addressing the ethical challenges of implementing precision medicine in this setting is crucial to prevent exacerbation of existing health care disparities. The rapid advancement of precision medicine technologies makes it imperative to understand these challenges before widespread implementation in emergency care settings.

Objective: This study aimed to identify high priority ethical concerns facing the implementation of precision medicine in the emergency department.

Methods: We conducted a qualitative study using a modified nominal group technique (NGT) with emergency physicians who had previous knowledge of precision medicine concepts. The NGT process consisted of four phases: (1) silent generation of ideas, (2) round-robin sharing of ideas, (3) structured discussion and clarification, and (4) thematic grouping of priorities. Participants represented diverse practice settings (county hospital, community hospital, academic center, and integrated managed care consortium) and subspecialties (education, ethics, pediatrics, diversity, equity, inclusion, and informatics) across various career stages from residents to late-career physicians.

Results: A total of 12 emergency physicians identified 82 initial challenges during individual ideation, which were consolidated to 48 unique challenges after removing duplicates and combining related items. The average participant contributed 6.8 (SD 2.9) challenges. These challenges were organized into a framework with 3 themes: values, privacy, and justice. The framework identified the need to address these themes across 3 time points of the precision medicine process: acquisition of data, actualization in the care setting, and the after effects of its use. This systematic organization revealed interrelated concerns spanning from data collection and bias to implementation challenges and long-term consequences for health care equity.

Conclusions: Our study developed a novel framework that maps critical ethical challenges across 3 domains (values, privacy, and justice) and 3 temporal stages of precision medicine implementation. This framework identifies high-priority areas for future research and policy development, particularly around data representation, privacy protection, and equitable access. Successfully addressing these challenges is essential to realize precision medicine's potential while preserving emergency medicine's core mission as health care's safety net.

(JMIR Form Res 2025;9:e68371) doi:10.2196/68371

KEYWORDS

precision medicine; emergency medicine; ethical considerations; nominal group study; consensus framework

Introduction

Historically, medical treatments have been tailored to the average patient. However, no individual is ever the average patient—everyone has their own risk factors, needs, and ideal outcomes. Precision medicine, as defined by the White House's nationwide Precision Medicine Initiative, aims to address this problem by moving from a one-size-fits-all model toward one of providing the "right care to the right person at the right time" through the use of vast amounts of health data, including

XSL•FO
RenderX

genomic information, electronic health records, environmental and lifestyle data, and medical imaging data, among others [1-5]. However, even the most advanced, targeted therapies are limited by biases that already exist in the US health care system [6,7].

Variation in care, even when "evidence based," can result in worsened gaps between the haves and the have-nots, disproportionately burdening already marginalized communities [8]. For example, patients with limited resources or inadequate insurance coverage may be unable to access these evidence-based interventions, while well-insured patients at major medical centers can, creating a 2-tiered system where advances in precision medicine may primarily benefit those with better health care access. For emergency physicians, who serve in the safety net for American health care, precision medicine's "right care, right patient, right time" must be balanced with our "anyone, anywhere, anytime" mantra and the practical and ethical challenges that this setting presents.

Early precision medicine initiatives have focused on the use of genetics to better understand or treat diseases. Much has been made of determining the right medication for the right person (eg, warfarin dosing for those with a CYP [cytochrome P450] mutation) or when to start cancer screening based on genetic profiles (eg, colorectal screening for CHD1 [chromodomain helicase DNA binding protein 1]) [9]. It is clear, however, that genetics alone cannot explain much of the variation in outcomes across the population [2,10]. The effect of behavioral factors like diet and exercise are well-established, and studies have begun to illuminate the effect of social and environmental determinants of health like access to health care resources or housing, which may be responsible for over half of health outcomes [11,12]. As such, there has been a broadening approach to precision medicine that may also include home monitoring, social media and social networks, as well as patient-generated data like diet, daily steps, exercise, or heart rate variability [13,14].

While useful when available, not all patients can afford these tools and services. Furthermore, "precise" treatments rely on the accuracy of this data, which must be built into datasets and shared. Few comprehensive datasets that incorporate all of these data exist, and even fewer have been carefully evaluated for missing or inaccurate data that could bias results [15]. Current datasets suffer from a lack of representation from some populations while simultaneously overrepresenting others. Healthy patients or those without access to care rarely present to hospitals, while those who are chronically ill are seen more often. Furthermore, a skewed representation occurs because research data often comes primarily from large academic centers rather than community hospitals or rural health care settings [16]. Data that were collected where patients are provided differential care might inadvertently support antiquated practices [17-21] due to the association of outcomes with variables like the ability to pay for a higher level of care. Though precision approaches aim to limit quality gaps and provide more equitable care, decisions that are made without the full picture of their context might inadvertently exacerbate systemic biases, for example, when risk assessment scores built from data of principally white cohorts over- or underestimate the risk of disease in non-White populations [7].

The emergency department setting offers additional unique hurdles for the growth of precision medicine. Little patient-generated data are readily available at the point of care. Highly important features like code status are haphazardly available among the sickest patients. Nevertheless, emergency care providers strive to provide better care, and to actualize the benefits of a learning health care system by integrating evidence and experience into continuous improvement and innovation. How can health care's safety net, responsible for a growing and significant portion of acute care in America [22,23], begin to approach the problems that are inherent to the implementation of precision medicine so that it might offer real-time, equitable care for patients with the widest spectrum of disease in medicine and a microcosm of the health care system?

Thus, the burgeoning expansion of precision medicine underscores the need to identify challenges facing its nascent deployment and ethical use in the emergency setting such that its optimal benefits may be realized. Our main objective was to identify critical ethical challenges in implementing precision medicine in the emergency setting using a consensus technique. This approach would also help identify themes and gaps in current research. This manuscript presents the resulting consensus framework.

Methods

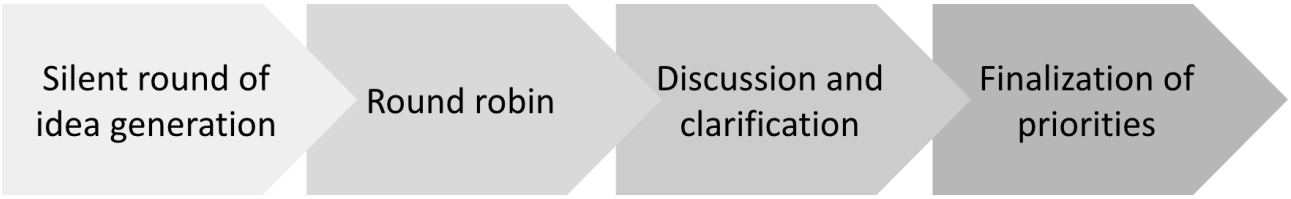
Study Design and Setting

We used a modified nominal group technique (NGT), a consensus methodology widely used in emergency medicine, to identify high priority ethical challenges relevant to the implementation of precision medicine in the emergency medicine setting and arrive at a consensus framework [24-31]. This study followed the COREQ (Consolidated Criteria for Reporting Qualitative Research) reporting framework for qualitative studies [32].

NGT is a 4-step process in which stakeholders generate and then prioritize questions or research ideas (Figure 1) [33]. NGT was well-suited for our research question as our aim was to generate a comprehensive set of ethical challenges which could then be organized and ordered. This is in contrast to other frequently used consensus techniques such as the Delphi technique, which are designed instead to reach a consensus opinion on one particular question [24]. NGT has previously been used successfully to identify research priorities in the field of emergency medicine [25,27,34-37].

XSL•FO
RenderX

Figure 1. Flow chart of the nominal group technique (NGT) process.



Selection and Participants

A NGT session group size is recommended at greater than 6 participants, with the ideal between 9 and 12 [34]. As precision emergency medicine is an emerging field, we sought participants who had familiarity with the clinical practice of emergency medicine, including an understanding of the types of data that are collected throughout a patient's journey and the ways in which data is used for clinical decision-making. Participants were recruited through email to the Society of Academic Medicine Ethics Committee listserve; the email invitation also asked recipients for referral and sharing to others who may be interested. A quota sample of the first 14 respondents to the email were registered as potential participants in the NGT group to meet the ideal group size and allow for scheduling conflicts. In total, 12 emergency medicine physicians who reported working knowledge of precision medicine topics, as evidenced by publication records and participation in precision medicine conferences or similar activities, joined our NGT session. Participants represented a spectrum of subspecialties (education, ethics, pediatrics, diversity, equity, inclusion, and informatics) across various stages of training (resident, early-, mid-, and late-career) and practice settings (county hospital, community hospital, academic center, and integrated managed care consortium) (Specialty, Level of Training, and Practice Setting included in Table S1 in Multimedia Appendix 1) to provide a range of emergency medicine practice perspectives.

Participants took part in the NGT session during the American College of Emergency Physicians annual conference in Boston in October 2021. The process was moderated by the lead researcher of the group (CR), and notes were taken by 2 separate researchers (ES and MG).

NGT Process

Phase 1: Silent Idea Generation.

The first step of NGT is a round of silent idea generation. In this phase, after being given an initial prompt, participants silently write down what they believe to be high priority challenges. It is imperative that this phase is silent, as this allows each member of the group to write their initial thoughts without influence from other members. Our prompt asked participants to identify high priority ethical concerns for the implementation of precision medicine initiatives in the emergency medicine setting. They had 10 minutes to complete this phase.

Phase 2: Sharing Ideas.

After individual idea generation, participants then shared their ideas with the group in a round-robin (going around the table one person at a time) format. NGT has the advantage of not allowing any one particular voice to outweigh others, by asking each participant to share their complete list of ideas in turn during the round robin phase, rather than ad hoc discussion, thus minimizing potential bias [38,39]. This phase lasted 10 minutes.

Phase 3: Discussion and Clarification

After the round-robin format, participants undertook discussion and clarification of topics. In this phase, participants were encouraged to ask other members for clarification of topics generated as needed. Cross-cutting themes were identified, and duplicate topics consolidated. Over 30 minutes, key topics were then placed on moveable cards for prioritization and further thematic grouping.

Phase 4: Finalization of Priorities

Finally, participants finalized the topics they believed represented the highest priority areas to address. In traditional NGT, this is the ranking phase, where members are allocated votes to generate a ranked list of priorities. Overwhelmingly, however, participants felt that the topics they raised were interrelated and of commensurate importance over several themes such that their categorization—rather than prioritization—would provide greater impactful focus for interventions. Consequently, participants grouped challenges through consensus opinion over a 40-minute period.

Ethical Considerations

This study was approved by the Stanford University Institutional Review Board (IRB #62687); participants gave written informed consent before taking part in this study. All data were collected anonymously. Participants received a US $125 gift card in compensation for their time.

Results

During phases 1 and 2 of the NGT process, 82 challenges were generated (complete list in Table S2 in Multimedia Appendix 1) with the average participant contributing 6.8 (SD 2.9) challenges. After refinement and separation of cross-cutting themes and consolidation of duplicate topics in phase 3, a list of 48 unique challenges remained.

In phase 4, participants grouped these challenges into common realms based on the nature of the ethical issue and where in the precision medicine process the issue was likely to occur. The participants identified 3 key time points: the acquisition of data, the actualization of precision medicine in practice, and the aftereffects of the process (Textboxes 1-3). The result of this mapping was a 3 by 3 matrix consisting of ethical challenges grouped into the domains of Values (eg, patient values and priorities as well as how precision medicine values those individuals), Privacy (eg, patient expectations as well as literal data security), and Justice (eg, overcoming and addressing

XSL•FO
RenderX

systemic biases) as related to their temporal relationship to the precision medicine process (Figure 2).

Textbox 1. Challenges identified by participants related to how the unique values of patients, providers, and the health care system will be incorporated into precision medicine.

Values

- **Acquisition**
- How are decisions made about which disease conditions to study and treat?
- How accurate is data in the medical record at representing the whole of an individual?
- How much does lifestyle matter and is it captured?
- How will social determinants of health be evaluated?
- Should individuals worry about the impact of nonmodifiable risk factors like family history or genetic predisposition?
- What if individuals identify differently from measures used in precision medicine?
- Can we mitigate the mistrust of people toward engaging with the system?
- **Actualization**
- How do we address patient choice and preference in the face of precision recommendations?
- What do we do if health care providers' interpretations and values conflict with those of the patient?
- How might this affect the patient-physician relationship?
- Will patients be able to opt out?
- Does the delivery of precision medicine change in the emergency scenario?
- How do we ensure provider comfort with using precision medicine?
- How do health care providers stay up to date with rapid changes in precision medicine?
- **Aftereffects**
- How do we maintain humility as evidence for decision-making changes over time?
- What if precision recommendations turn out to be wrong?
- Might there be unintended psychological consequences like stress after decisions are made?
- How do we ensure patient choice and preference is communicated longitudinally between medical teams?
- What prevents a capitalist model from researching only the most common or lucrative conditions?

XSL•FO
RenderX

Textbox 2. Challenges identified by participants regarding privacy in the precision medicine process.

Privacy

- **Acquisition**
- Who will be accountable for the collected information's storage, maintenance, and safety?
- How will protected information be handled?
- How do we protect minors while ensuring appropriate data sharing with their caregivers?
- How do we include underrepresented or excluded groups in a way that avoids perception of surveillance?
- **Actualization**
- How will patients interface with their personal medical data?
- Will patients lose the power to make private decisions about their own health when it may have a downstream effect on others?
- How will changing evidence or incidental findings like new genetic risk factors be communicated with patients to ensure follow-up?
- **Aftereffects**
- Can anonymized data be reidentified?
- How do we ensure safe data sharing can occur among differing institutions or electronic medical record systems?
- What are the legal ramifications of an infinitely variable standard of care?

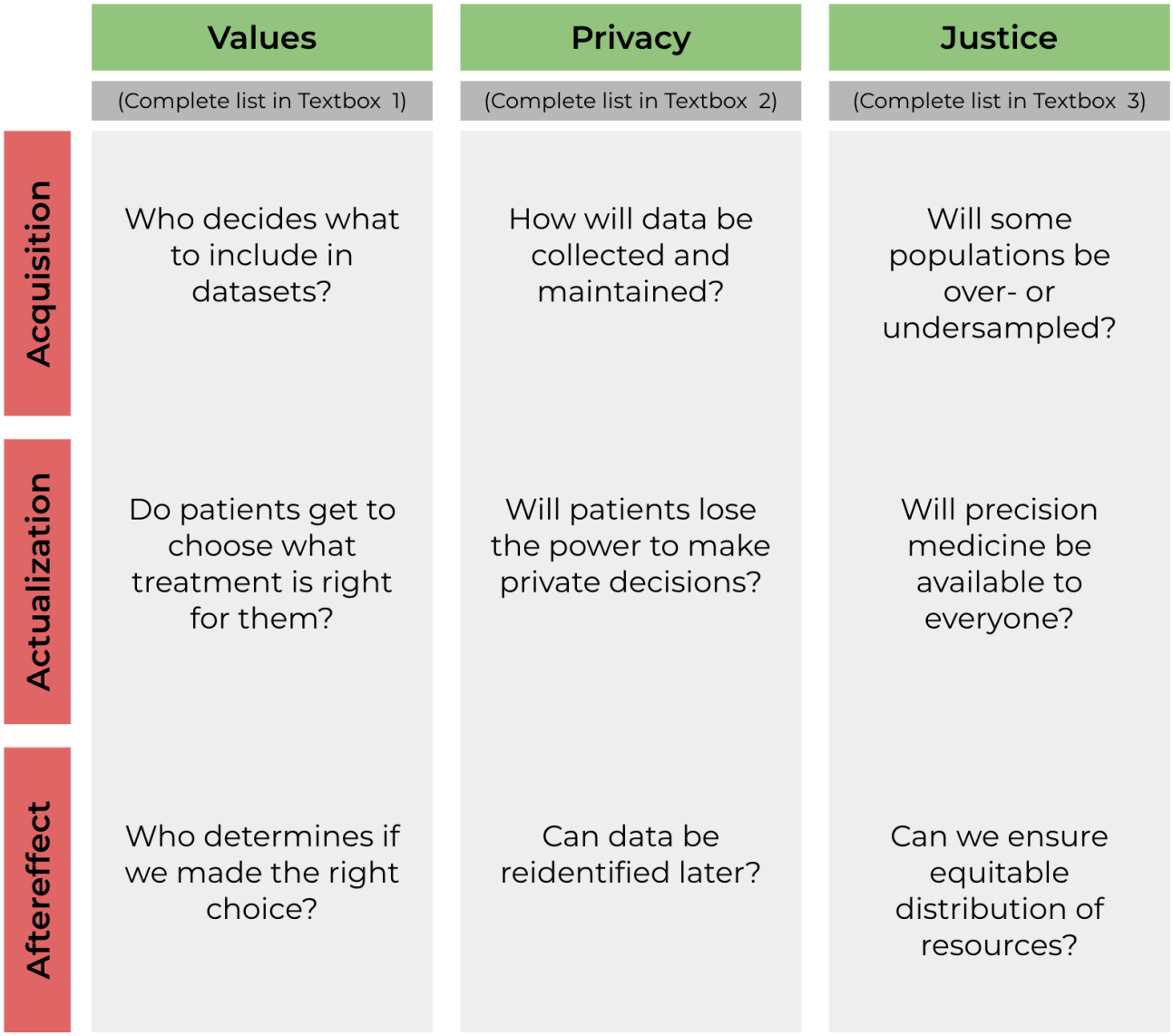
Textbox 3. Challenges identified by participants about how to ensure just care in the precision medicine model and process.

Justice

- **Acquisition**
- How do we incorporate all people into precision research?
- Can we address disparities like access to care that make it hard for patients to participate?
- Will those who are historically excluded continue to be so, while some populations will see over- and undersampling?
- What do we do with historical data that may be biased?
- Who will fund the research and data acquisition?
- How are precision medicine priorities determined (ie, institutional level or government level)?
- Could precision research resources be better spent addressing direct care access?
- **Actualization**
- Will flawed data inputs result in flawed data outputs and recommendations?
- How do we build a relationship of trust with a patient in the emergency department?
- How do we prevent nonmodifiable predispositions from leading to biased care (ie, genetic determinism)?
- How do we allow patient choice and values in the face of recommendations or limited resources?
- Can variations in care still be equitable?
- Will people get disproportionate care based on affordability?
- How can we balance precision medicine safety with accessibility?
- **Aftereffects**
- How will we maintain monitoring and vigilance of the accuracy of these systems?
- How do we ensure systems are learning and adapting over time?
- Can we ensure precision medicine is equitable across all patient populations?
- Might the choice of who and what to study and treat have unintended consequences?
- Might this further biased or racist ideologies and marginalize patients?

XSL•FO
RenderX

Figure 2. Priority matrix of ethical concerns for precision medicine.



Discussion

Principal Findings

In this study, we have identified critical ethical challenges facing the implementation of precision medicine in the unique practice environment of the emergency department. The growing use of vast amounts of genomic, clinical, and lifestyle data to guide individualized therapies holds immense potential to improve emergency care. However, the "anyone, anywhere, anytime" mandate of emergency medicine presents distinct hurdles for the deployment of precision approaches that rely on the timely availability of complete patient information in a fast-paced and high-acuity setting. Using a modified nominal group technique, we engaged a diverse panel of emergency physicians to develop a consensus framework mapping key ethical challenges along the continuum of precision medicine implementation, from data acquisition to use in clinical practice to monitoring of downstream consequences. Crucially, this framework highlights the need to proactively identify and mitigate potential pitfalls to best realize the promise of precision medicine in the ED while promoting equitable access and benefit for the diverse patient populations we serve. Our findings provide a framework to guide research, practice innovation, and policy aimed at the ethical translation of precision medicine to the front lines of emergency care.

Shared Ethical Challenges: Values, Privacy, and Justice

A recurring theme in our NGT discussion was "whose values are we accounting for?" [40]. Participants acknowledged the sometimes conflicting needs of the individual versus those of the system. Providers may recommend medications that prolong survival, but side effects may be undesirable or cost prohibitive for some patients [41]. The definition of high-quality care envisioned by precision medicine may not align with the individual goals of patients. There are multiple stakeholders driving the discourse of precision medicine—health care providers, health care insurers, regulators, and industry—the least empowered of which is the patient [40]. There were many questions raised regarding challenges to patient autonomy and the ambiguity of who is the ultimate decision maker in precision medicine and how that might change over time, as the needs of the many may be at odds with the needs of a few.

XSL•FO
RenderX

This tension maybe at odds with the ethical principle of autonomy [42]. The definition of respect for autonomy requires that patients are able to make their own decisions while incorporating their own values. This potentially extends to the use of their data and specimens, but the limits of patient autonomy remain under debate [43-45]. For example, the appropriate approach to consent for enrollment in biobanks that start in emergency departments remains under debate [46,47].

Ensuring that patient autonomy is maintained can be particularly challenging in the emergency care context. In the emergency department, providers make decisions with patients in a time-constrained setting, particularly compared with those historical clinical environments of precision medicine, such as oncology, which may occur over weeks. Not only does this limit time for discussion, it may also create a sense of external pressure to make a decision [48]. In such time-sensitive situations, patient personal information and preferences may be disclosed and decisions required under duress, thus also implicating privacy [49].

Privacy includes not only the protection of one's identity and personal information, but also the ability to make choices without them being disclosed to others [50]. Another challenge to the patient's privacy is that the practice of precision medicine in the emergency department will require massive amounts of data transiting among and across institutions and devices. Ensuring the safety and security of this protected information is a major hurdle for precision medicine initiatives, and the abovementioned time-sensitive nature of the emergency department encounter makes it necessary to share this data in real time, which is unique from other medical specialties [51-54]. In addition, when patients know that the information they provide to the system will be used to make recommendations for themselves or others, it may have downstream effects. For example, alcohol use and history for the allotment of liver transplant is one area that may leave patients in a moral quandary, unsure whether to be open with their behavior but fearing the effects of those disclosures [55,56]. Patient perceptions of the trustworthiness of health care providers and institutions will impact the success of any scientific or research endeavor, particularly among historically disadvantaged populations [57].

It is no surprise that our consensus framework includes justice as a key area for research in precision medicine ethics as it grows in the emergency department and in the prehospital setting [58-60]. Emergency physicians carry the distinct duty under federal law to provide care to anyone who presents for emergency care [61]. This duty is widely recognized to rest on a fundamental moral responsibility to provide care "promptly and expertly, without prejudice or partiality" [49]. At first it may seem that precision medicine would help to limit personal biases which might affect care, using data as opposed to gestalt or local practice variation. But structural biases may encode results and outcomes on patients that might inadvertently limit the benefits of the precision medicine model. When they are present in the initial development of the model, such as lack of representation and personal biases, affects the validity of results used to make targeted recommendations [62].

Justice, or more specifically with respect to medical ethics, distributive justice, emphasizes that resources are to be distributed fairly, equitably, and appropriately. This can be heavily hindered if providers rely on precision models to make recommendations without a critical appraisal of how well the results fit the patient in front of them. Even if done well, the outputs of precision medicine may not be available to all [63]. Precision medicine which is ultimately only accessible to a few but built on the data of many, would be an affront to the growing field of social emergency medicine [59]. In oncology, an area with a much longer history of precision medicine, 78% of patients and providers reported concerns regarding the accessibility of precision medicine [64].

Continuum of Precision Emergency Medicine: Data Availability, Actualization, and Aftereffects

As noted by participants, challenges in implementing precision medicine in the emergency medicine context occur anywhere from when data is collected (or not collected) to the downstream effects of using precision medicine interventions. In emergency medicine, providers stand prepared to address a range of conditions at any time. We often must rely on incomplete patient histories (eg, due to the severity of illness or mental status changes frequently encountered in emergency medicine) and limited access to previous medical records (eg, our patients may be uninsured, underinsured, and without access to routine primary care providers) [65]. Precision medicine approaches will need to be deployable in a timely manner and maintain their precision with only the limited patient-specific information available to emergency physicians.

Furthermore, even when data are available, the actualization of precision medicine in the care environment may serve only to perpetuate systemic flaws in the health care system if that data is inaccurate, biased, or exclusionary from the start [7,66]. A 12-year review of the National Hospital Ambulatory Medical Care Survey (NHAMCS) illustrates this risk: Black patients were 10% less likely to receive immediate or emergent Emergency Severity Index (ESI) scores and to be admitted, but 1.26 times more likely to die in the emergency department or the hospital than White patients [65]. Using such historically biased data to make precision medicine recommendations about resource allocation could amplify these disparities. While precision medicine aims to deliver individually tailored care, it still requires categorizing patients into discrete groups by shared characteristics—a process that carries its own implications and potential for bias [4,67].

Additional logistic concerns for the actualization of precision medicine in emergency departments, include costs, legal issues, and patient education. Efforts made to collate data from emergency care settings to be used in precision medicine research may result in products or treatment pathways unaffordable to these patients [68-70]. This may be mitigated through insurers [71], but the emergency department is a safety net, and as such this approach may not benefit the most vulnerable patients while those who are insured may find that their plans have little incentive to offer this option [72]. Research has shown that the cost of precision medicine is a concern shared by patients and providers alike [63,64].

XSL•FO
RenderX

Even when payment is not an issue, a data- and machine-supported approach may not be conducive to truly informed consent given the limitations of adequately educating emergency department patients regarding how precision medicine and its associated research works [63,73]. Beyond the challenge of consenting in the emergency context, providers must also consider issues regarding disclosing incidental information and the potential changes associated with the use of data over time which interconnected to other growing datasets. Stewardship of these data, defined as "the responsible management of something entrusted to one's care" [74], will be critical to maintaining trust of patients and communities as this science evolves [57].

Discussion in this NGT process highlighted the need to actively maintain a holistic view of the patient without being overly reductive to these categories. Currently, only a limited amount of data is collected with respect to health measures or lab studies, and even less is recorded for social determinants of health. Future work should focus on systematic collection of environmental and social factors in emergency care settings, including neighborhood-level data on environmental exposures, community resources, and barriers to care access. However, this data collection must be done thoughtfully to avoid reinforcing biases—for example, by ensuring community engagement in data collection methods, using validated measurement tools that have been tested across diverse populations, and analyzing social factors as modifiable system-level variables rather than immutable patient characteristics. In emergency medicine specifically, this could include developing standardized screening tools for social needs that are feasible in acute care settings, creating automated systems to gather community-level data that do not burden individual patient encounters, and establishing partnerships with community organizations to better understand local health contexts. These approaches could help create more comprehensive precision medicine models while maintaining emergency medicine's core mission of providing equitable care to all patients.

Limitations

The use of a NGT and participant diversity are key strengths of this study. However, several limitations warrant discussion. While NGT aims to limit the biases associated with other consensus models, expertise bias is still a potential limitation, as all participants were aware of and knowledgeable on this topic beforehand. While this expertise strengthens the confidence in the working knowledge of the topic, it can also lead to anchoring. We took several steps to limit its impact. NGT, by way of its structured format, is meant to limit the influence of any one dominant individual in discussion [39]. In addition, we included a diverse spectrum of emergency physicians from residents to attending physicians with varying years of experience, as well as emergency physicians with a variety of subspeciality training.

Self-selection bias should also be considered, as participants volunteered in response to an email invitation to the Society of Academic Medicine Ethics Committee listserve. This recruitment method may have attracted participants with stronger interest in ethics and precision medicine, though their engagement likely enhanced the depth of insights generated during the NGT session.

Finally, though the number of participants was within ideal range for an NGT [24], the relatively small group size remains a limitation. Future research could benefit from additional strategies that would incorporate a higher number of emergency physicians and experts from outside of the field which might offer perspective to our work.

Conclusions

Realizing the promise of precision medicine's "right person, right place, right time" must be balanced by the context of the "anyone, anywhere, anytime" of emergency medicine. Our work has illuminated a range of concerns for the development of precision emergency medicine, and provides a focused framework for the most pressing ethical considerations along the continuum of data acquisition to implementation. This framework may be used to direct research and innovation toward addressing these challenges in the emergency medicine setting.

Acknowledgments

This study was made possible thanks to the financial support provided through a generous seed grant from Stanford University's McCoy Family Center for Ethics in Society - Ethics, Society, and Technology Hub Seed Grant.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

CR, JN, ES, and IK conceived of, designed, and completed the research. CR and JN obtained research funding. AF and NA performed data analysis. CR and JN drafted the manuscript. IK and IW contributed to the manuscript development and all authors contributed substantially to manuscript editing. CR takes responsibility for the work as a whole. No generative AI was used in any portion of the manuscript writing.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Additional tables.
[DOCX File, 20 KB - formative_v9i1e68371_app1.docx]

References

1. Salari P, Larijani B. Ethical issues surrounding personalized medicine: a literature review. Acta Med Iran 2017 Mar;55(3):209-217. [Medline: 28282721]
2. Liu X, Luo X, Jiang C, Zhao H. Difficulties and challenges in the development of precision medicine. Clin Genet 2019 May;95(5):569-574. [doi: 10.1111/cge.13511] [Medline: 30653655]
3. Sitapati A, Kim H, Berkovich B, et al. Integrated precision medicine: the role of electronic health records in delivering personalized treatment. Wiley Interdiscip Rev Syst Biol Med 2017 May;9(3):e1378. [doi: 10.1002/wsbm.1378] [Medline: 28207198]
4. Seymour CW, Gomez H, Chang CCH, et al. Precision medicine for all? Challenges and opportunities for a precision medicine approach to critical illness. Crit Care 2017 Oct 18;21(1):257. [doi: 10.1186/s13054-017-1836-5] [Medline: 29047353]
5. Mensah GA, Jaquish C, Srinivas P, et al. Emerging concepts in precision medicine and cardiovascular diseases in racial and ethnic minority populations. Circ Res 2019 Jun 21;125(1):7-13. [doi: 10.1161/CIRCRESAHA.119.314970] [Medline: 31219738]
6. Lee SJ, Fullerton SM, Saperstein A, Shim JK. Ethics of inclusion: Cultivate trust in precision medicine. Science 2019 Jun 7;364(6444):941-942. [doi: 10.1126/science.aaw8299] [Medline: 31171685]
7. Geneviève LD, Martani A, Shaw D, Elger BS, Wangmo T. Structural racism in precision medicine: leaving no one behind. BMC Med Ethics 2020 Feb 19;21(1):17. [doi: 10.1186/s12910-020-0457-8] [Medline: 32075640]
8. Stein JN, Charlot M, Cykert S. Building toward antiracist cancer research and practice: The case of precision medicine. JCO Oncol Pract 2021 May;17(5):273-277. [doi: 10.1200/OP.20.01070] [Medline: 33974820]
9. Goetz LH, Schork NJ. Personalized medicine: motivation, challenges, and progress. Fertil Steril 2018 Jun;109(6):952-963. [doi: 10.1016/j.fertnstert.2018.05.006] [Medline: 29935653]
10. Elemento O. The future of precision medicine: towards a more predictive personalized medicine. Emerg Top Life Sci 2020 Sep 8;4(2):175-177. [doi: 10.1042/ETLS20190197] [Medline: 32856697]
11. Hill-Briggs F, Adler NE, Berkowitz SA, et al. Social determinants of health and diabetes: a scientific review. Diabetes Care 2020 Nov 2;44(1):258-279. [doi: 10.2337/dci20-0053] [Medline: 33139407]
12. Marmot M, Allen JJ. Social determinants of health equity. Am J Public Health 2014 Sep;104 Suppl 4(Suppl 4):S517-S519. [doi: 10.2105/AJPH.2014.302200] [Medline: 25100411]
13. Kalra D. The importance of real-world data to precision medicine. Per Med 2019 Mar;16(2):79-82. [doi: 10.2217/pme-2018-0120] [Medline: 30724116]
14. Hollister B, Bonham VL. Should electronic health record-derived social and behavioral data be used in precision medicine research? AMA J Ethics 2018 Sep 1;20(9):E873-E880. [doi: 10.1001/amajethics.2018.873] [Medline: 30242819]
15. Sukumar SR, Natarajan R, Ferrell RK. Quality of big data in health care. Int J Health Care Qual Assur 2015;28(6):621-634. [doi: 10.1108/IJHCQA-07-2014-0080] [Medline: 26156435]
16. DesHarnais S. Current uses of large data sets to assess the quality of providers. Construction of risk-adjusted indexes of hospital performance. Int J Technol Assess Health Care 1990;6(2):229-238. [doi: 10.1017/s0266462300000751] [Medline: 2118489]
17. Williams WW, Hogan JW, Ingelfinger JR. Time to eliminate health care disparities in the estimation of kidney function. N Engl J Med 2021 Nov 4;385(19):1804-1806. [doi: 10.1056/NEJMe2114918]
18. Wiens J, Saria S, Sendak M, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med 2019 Sep;25(9):1337-1340. [doi: 10.1038/s41591-019-0548-6]
19. Char DS, Shah NH, Magnus D. Implementing machine learning in health care - addressing ethical challenges. N Engl J Med 2018 Mar 15;378(11):981-983. [doi: 10.1056/NEJMp1714229] [Medline: 29539284]
20. Owens K, Walker A. Those designing healthcare algorithms must become actively anti-racist. Nat Med 2020 Sep;26(9):1327-1328. [doi: 10.1038/s41591-020-1020-3]
21. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med 2018 Dec 18;169(12):866-872. [doi: 10.7326/M18-1990] [Medline: 30508424]
22. Marcozzi D, Carr B, Liferidge A, Baehr N, Browne B. Trends in the contribution of emergency departments to the provision of hospital-associated health care in the USA. Int J Health Serv 2018 Apr;48(2):267-288. [doi: 10.1177/0020731417734498] [Medline: 29039720]
23. Hsia RY, Sabbagh SH, Guo J, Nuckton TJ, Niedzwiecki MJ. Trends in the utilisation of emergency departments in California, 2005–2015: a retrospective analysis. BMJ Open 2018 Jul;8(7):e021392. [doi: 10.1136/bmjopen-2017-021392]
24. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. Int J Clin Pharm 2016 Jun;38(3):655-662. [doi: 10.1007/s11096-016-0257-x] [Medline: 26846316]

XSL•FO
RenderX

25. Ranney ML, Fletcher J, Alter H, et al. A consensus-driven agenda for emergency medicine firearm injury prevention research. Ann Emerg Med 2017 Feb;69(2):227-240. [doi: 10.1016/j.annemergmed.2016.08.454] [Medline: 27998625]
26. Elder E, Johnston AN, Byrne JH, Wallis M, Crilly J. Core components of A staff wellness strategy in emergency departments: A clinician-informed nominal group study. Emerg Med Australas 2021 Feb;33(1):25-33. [doi: 10.1111/1742-6723.13561] [Medline: 32592326]
27. Nik Hisamuddin R, Tuan Hairulnizam TK. Developing key performance indicators for emergency department of teaching hospitals: a mixed fuzzy Delphi and nominal group technique approach. Malays J Med Sci 2022 Apr;29(2):114-125. [doi: 10.21315/mjms2022.29.2.11] [Medline: 35528822]
28. Peltzer-Jones J, Nordstrom K, Currier G, Berlin JS, Singh C, Schneider S. A research agenda for assessment and management of psychosis in emergency department patients. West J Emerg Med 2019 Mar;20(2):403-408. [doi: 10.5811/westjem.2019.1.39263] [Medline: 30881564]
29. Members of the SAEM Consensus Conference Emergency Medical Services Subcommittee, Adelgais KM, Hansen M, et al. Establishing the key outcomes for pediatric emergency medical services research. Acad Emerg Med 2018 Dec;25(12):1345-1354. [doi: 10.1111/acem.13637] [Medline: 30312993]
30. Marynowski-Traczyk D, Wallis M, Broadbent M, et al. Optimising emergency department and acute care for people experiencing mental health problems: a nominal group study. Aust Health Rev 2022 Oct;46(5):519-528. [doi: 10.1071/AH21092] [Medline: 34793296]
31. Safdar B, Greenberg MR. Organization, execution and evaluation of the 2014 Academic Emergency Medicine consensus conference on Gender-Specific Research in Emergency Care - an executive summary. Acad Emerg Med 2014 Dec;21(12):1307-1317. [doi: 10.1111/acem.12530] [Medline: 25420469]
32. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
33. Agency for Healthcare Research and Quality. Nominal group technique. URL: https://digital.ahrq.gov/health-it-tools-and-resources/evaluation-resources/workflow-assessment-health-it-toolkit/all-workflow-tools/nominal-group-technique [accessed 2024-12-18]
34. Harvey N, Holmes CA. Nominal group technique: an effective method for obtaining group consensus. Int J Nurs Pract 2012 Apr;18(2):188-194. [doi: 10.1111/j.1440-172X.2012.02017.x] [Medline: 22435983]
35. Kottmann A, Krüger AJ, Sunde GA, et al. Establishing quality indicators for pre-hospital advanced airway management: a modified nominal group technique consensus process. Br J Anaesth 2022 Feb;128(2):e143-e150. [doi: 10.1016/j.bja.2021.08.031] [Medline: 34674835]
36. Sunde GA, Kottmann A, Heltne JK, et al. Standardised data reporting from pre-hospital advanced airway management - a nominal group technique update of the Utstein-style airway template. Scand J Trauma Resusc Emerg Med 2018 Jun 4;26(1):46. [doi: 10.1186/s13049-018-0509-y] [Medline: 29866144]
37. Nelson V, Dubov A, Morton K, Fraenkel L. Using nominal group technique among resident physicians to identify key attributes of a burnout prevention program. PLoS One 2022;17(3):e0264921. [doi: 10.1371/journal.pone.0264921] [Medline: 35303009]
38. Gallagher M, Hares T, Spencer J, Bradshaw C, Webb I. The nominal group technique: a research tool for general practice? Fam Pract 1993 Mar;10(1):76-81. [doi: 10.1093/fampra/10.1.76] [Medline: 8477899]
39. Humphrey-Murto S, Varpio L, Gonsalves C, Wood TJ. Using consensus group methods such as Delphi and Nominal Group in medical education research. Med Teach 2017 Jan;39(1):14-19. [doi: 10.1080/0142159X.2017.1245856] [Medline: 27841062]
40. Chan S, Erikainen S. What's in a name? The politics of "Precision Medicine". Am J Bioeth 2018 Apr;18(4):50-52. [doi: 10.1080/15265161.2018.1431324] [Medline: 29621451]
41. Kelly TD, Hawk KF, Samuels EA, Strayer RJ, Hoppe JA. Improving uptake of emergency department-initiated buprenorphine: barriers and solutions. West J Emerg Med 2022 Jul 11;23(4):461-467. [doi: 10.5811/westjem.2022.2.52978] [Medline: 35980414]
42. Sedig L. What's the role of autonomy in patient- and family-centered care when patients and family members don't agree? AMA J Ethics 2016 Jan 1;18(1):12-17. [doi: 10.1001/journalofethics.2016.18.1.ecas2-1601] [Medline: 26854631]
43. Entwistle VA, Carter SM, Cribb A, McCaffery K. Supporting patient autonomy: the importance of clinician-patient relationships. J Gen Intern Med 2010 Jul;25(7):741-745. [doi: 10.1007/s11606-010-1292-2] [Medline: 20213206]
44. Lewis J, Holm S. Organoid biobanking, autonomy and the limits of consent. Bioethics 2022 Sep;36(7):742-756. [doi: 10.1111/bioe.13047] [Medline: 35531912]
45. Caulfield T, Murdoch B. Genes, cells, and biobanks: Yes, there's still a consent problem. PLoS Biol 2017 Jul;15(7):e2002654. [doi: 10.1371/journal.pbio.2002654] [Medline: 28742850]
46. Ashcroft JW, Macpherson CC. The complex ethical landscape of biobanking. Lancet Public Health 2019 Jun;4(6):e274-e275. [doi: 10.1016/S2468-2667(19)30081-7] [Medline: 31155216]
47. Saben JL, Shelton SK, Hopkinson AJ, et al. The Emergency Medicine Specimen Bank: An innovative approach to biobanking In acute care. Acad Emerg Med 2019 Jun;26(6):639-647. [doi: 10.1111/acem.13620] [Medline: 30239069]

XSL•FO
RenderX

48. Naess AC, Foerde R, Steen PA. Patient autonomy in emergency medicine. Med Health Care Philos 2001;4(1):71-77. [doi: 10.1023/a:1009982710625] [Medline: 11315422]
49. American College of Emergency Physicians. Code of ethics for emergency physicians. Ann Emerg Med 2008 Nov;52(5):581-590. [doi: 10.1016/j.annemergmed.2008.07.041] [Medline: 18970998]
50. Geiderman JM, Moskop JC, Derse AR. Privacy and confidentiality in emergency medicine: obligations and challenges. Emerg Med Clin North Am 2006 Aug;24(3):633-656. [doi: 10.1016/j.emc.2006.05.005] [Medline: 16877134]
51. Lee SSJ, Cho MK, Kraft SA, et al. "I don't want to be Henrietta Lacks": diverse patient perspectives on donating biospecimens for precision medicine research. Genet Med 2019 Jan;21(1):107-113. [doi: 10.1038/s41436-018-0032-6]
52. Ratcliff CL, Kaphingst KA, Jensen JD. When personal feels invasive: foreseeing challenges in precision medicine communication. J Health Commun 2018;23(2):144-152. [doi: 10.1080/10810730.2017.1417514] [Medline: 29279048]
53. Stiles D, Appelbaum PS. Cases in precision medicine: concerns about privacy and discrimination after genomic sequencing. Ann Intern Med 2019 May 21;170(10):717. [doi: 10.7326/M18-2666]
54. Sankar PL, Parker LS. The Precision Medicine Initiative's All of Us Research Program: an agenda for research on its ethical, legal, and social issues. Genet Med 2017 Jul;19(7):743-750. [doi: 10.1038/gim.2016.183] [Medline: 27929525]
55. Moss AH, Siegler M. Should alcoholics compete equally for liver transplantation? JAMA 1991 Mar 13;265(10):1295-1298. [Medline: 1995977]
56. Sedki M, Ahmed A, Goel A. Ethical and allocation issues in liver transplant candidates with alcohol related liver disease. Transl Gastroenterol Hepatol 2022;7:26. [doi: 10.21037/tgh-2020-13] [Medline: 35892052]
57. Kraft SA, Cho MK, Gillespie K, et al. Beyond consent: building trusting relationships with diverse populations in precision medicine research. Am J Bioeth 2018 Apr;18(4):3-20. [doi: 10.1080/15265161.2018.1431322] [Medline: 29621457]
58. Zink BJ. Social justice, egalitarianism, and the history of emergency medicine. Virtual Mentor 2010 Jun 1;12(6):492-494. [doi: 10.1001/virtualmentor.2010.12.6.mhst1-1006] [Medline: 23158453]
59. Anderson ES, Lippert S, Newberry J, Bernstein E, Alter HJ, Wang NE. Addressing social determinants of health from the emergency department through social emergency medicine. West J Emerg Med 2016 Jul;17(4):487-489. [doi: 10.5811/westjem.2016.5.30240] [Medline: 27429706]
60. Ben-Assuli O. Electronic health records, adoption, quality of care, legal and privacy issues and their implementation in emergency departments. Health Policy 2015 Mar;119(3):287-297. [doi: 10.1016/j.healthpol.2014.11.014] [Medline: 25483873]
61. Brown HL, Brown TB. EMTALA: The evolution of emergency care in the United States. J Emerg Nurs 2019 Jul;45(4):411-414. [doi: 10.1016/j.jen.2019.02.002] [Medline: 30902349]
62. Boulware LE, Purnell TS, Mohottige D. Systemic kidney transplant inequities for black individuals: examining the contribution of racialized kidney function estimating equations. JAMA Netw Open 2021 Jan 4;4(1):e2034630. [doi: 10.1001/jamanetworkopen.2020.34630] [Medline: 33443577]
63. Erdmann A, Rehmann-Sutter C, Bozzaro C. Patients' and professionals' views related to ethical issues in precision medicine: a mixed research synthesis. BMC Med Ethics 2021 Aug 31;22(1):116. [doi: 10.1186/s12910-021-00682-8] [Medline: 34465328]
64. Miller AM, Garfield S, Woodman RC. Patient and Provider Readiness for Personalized Medicine. Personalized Medicine in Oncology. URL: https://www.personalizedmedonc.com/articles/2816:patient-and-provider-readiness-for-personalized-medicine [accessed 2025-01-29]
65. Zhang X, Carabello M, Hill T, Bell SA, Stephenson R, Mahajan P. Trends of racial/ethnic differences in emergency department care outcomes among adults in the United States from 2005 to 2016. Front Med (Lausanne) 2020;7:300. [doi: 10.3389/fmed.2020.00300] [Medline: 32671081]
66. Mersha TB, Abebe T. Self-reported race/ethnicity in the age of genomic research: its potential impact on understanding health disparities. Hum Genomics 2015 Jan 7;9(1):1. [doi: 10.1186/s40246-014-0023-x] [Medline: 25563503]
67. Manrai AK, Patel CJ, Ioannidis JPA. In the era of precision medicine and big data, who is normal? JAMA 2018 May 15;319(19):1981-1982. [doi: 10.1001/jama.2018.2009] [Medline: 29710130]
68. O'Sullivan BP, Orenstein DM, Milla CE. Pricing for orphan drugs: will the market bear what society cannot? JAMA 2013 Oct 2;310(13):1343-1344. [doi: 10.1001/jama.2013.278129] [Medline: 24084916]
69. Ferkol T, Quinton P. Precision medicine: at what price? Am J Respir Crit Care Med 2015 Sep 15;192(6):658-659. [doi: 10.1164/rccm.201507-1428ED] [Medline: 26207804]
70. Ramaswami R, Bayer R, Galea S. Precision medicine from a public health perspective. Annu Rev Public Health 2018 Apr 1;39(153–68):153-168. [doi: 10.1146/annurev-publhealth-040617-014158] [Medline: 29166244]
71. Geruso M, Jena AB, Layton TJ. Will personalized medicine mean higher costs for consumers? Harvard Business Review. URL: https://hbr.org/2018/03/will-personalized-medicine-mean-higher-costs-for-consumers [accessed 2025-01-17]
72. Horgan D, Jansen M, Leyens L, et al. An index of barriers for the implementation of personalised medicine and pharmacogenomics in Europe. Public Health Genomics 2014;17(5-6):287-298. [doi: 10.1159/000368034] [Medline: 25401385]

XSL•FO
RenderX

73. Fiore RN, Goodman KW. Precision medicine ethics: selected issues and developments in next-generation sequencing, clinical oncology, and ethics. Curr Opin Oncol 2016 Jan;28(1):83-87. [doi: 10.1097/CCO.0000000000000247] [Medline: 26569425]
74. Fullerton SM, Anderson NR, Guzauskas G, Freeman D, Fryer-Edwards K. Meeting the governance challenges of next-generation biorepository research. Sci Transl Med 2010 Jan 20;2(15):20371468. [doi: 10.1126/scitranslmed.3000361] [Medline: 20371468]

Abbreviations

CHD1: chromodomain helicase DNA binding protein 1
COREQ: Consolidated Criteria for Reporting Qualitative Research
CYP: cytochrome P450
ESI: Emergency Severity Index
NGT: nominal group technique
NHAMCS: National Hospital Ambulatory Medical Care Survey

Edited by A Mavragani; submitted 04.11.24; peer-reviewed by M Siermann, U Khan; revised version received 18.12.24; accepted 18.12.24; published 06.02.25.

Please cite as:
Rose C, Shearer E, Woller I, Foster A, Ashenburg N, Kim I, Newberry J
Identifying High-Priority Ethical Challenges for Precision Emergency Medicine: Nominal Group Study
JMIR Form Res 2025;9:e68371
URL: https://formative.jmir.org/2025/1/e68371
doi:10.2196/68371

© Christian Rose, Emily Shearer, Isabela Woller, Ashley Foster, Nicholas Ashenburg, Ireh Kim, Jennifer Newberry. Originally published in JMIR Formative Research (https://formative.jmir.org), 6.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impact of Psychological Resilience on the Fear of Pain and Activity Recovery in Postsurgical Patients: Observational Cohort Study

Yang Luo[1*], MN; Sisi Li[2*]; Lijuan Feng[1]; Junyi Zheng[1]; Chunfen Peng[3]; Lihong Bao[1]

[1]Department of Thoracic Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, 1277 Jiefang road, Jianghan District, Wuhan, China

[2]Department of Colorectal and Anal Surgery, Zhongnan Hospital of Wuhan University, Wuhan, China

[3]Cancer Center, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[*]these authors contributed equally

Corresponding Author:
Lihong Bao
Department of Thoracic Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, 1277 Jiefang road, Jianghan District, Wuhan, China

Abstract

Background: The fear of pain (FOP) and fear avoidance belief (FAB) play a crucial role in the occurrence and development of chronic pain. However, the dynamics of these factors in postsurgical pain appear to differ, with the FOP often learned from others rather than directly caused by painful experiences. Psychological resilience refers to an individual's capacity to adapt effectively to adversity, challenges, and threats, and may play a significant role in overcoming the FOP and avoidance behavior.

Objective: The aim of this study was to investigate the role of psychological resilience in overcoming the FOP and avoidance behavior among surgical patients undergoing lung surgery.

Methods: Participants were recruited at the Wuhan Union Hospital. Psychological resilience was measured using the Connor-Davidson Resilience Scale. The FOP was assessed using the simplified Chinese version of the Fear of Pain-9 items. The FAB was measured using the Physical Activity subscale of the Fear-Avoidance Beliefs Questionnaire. Activity recovery was assessed through questions related to social activities and household responsibilities. The adaptive least absolute shrinkage and selection operator (Lasso) regression analysis under nested cross-validation was used to identify key factors affecting postoperative FOP and activity recovery.

Results: A total of 144 participants were included in the final analysis. The results showed that preoperative FOP (coefficient=8.620) and FAB (coefficient=8.560) were mainly positively correlated with postoperative FOP, while psychological resilience (coefficient=−5.822) and age (coefficient=−2.853) were negatively correlated with it. These average R^2 of these models was 73% (SD 6%). Psychological resilience was the most important factor in predicting activity recovery, and these models obtained an average accuracy of 0.820 (SD 0.024) and an average area under the curve of 0.926 (SD 0.044).

Conclusions: Psychological resilience was negatively associated with the postoperative FOP and positively with activity recovery in patients who underwent lung surgery. Patients with higher resilience are more likely to cope effectively with postsurgical pain and recover activities sooner. These findings highlight the importance of assessing and potentially enhancing psychological resilience in the perioperative period to improve postoperative outcomes.

(JMIR Form Res 2025;9:e63556) doi:10.2196/63556

KEYWORDS

psychological resilience; fear of pain; postsurgical recovery; fear avoidance belief; pain management; activity recovery

Introduction

Psychological factors such as the fear of pain (FOP) and fear avoidance belief (FAB) are thought to play an important role in the occurrence and development of chronic pain [1-3]. Based on the fear-avoidance model, among patients with chronic pain, painful experiences from activities will lead to the FOP, and then cause patients to refuse activities until disability, which in turn can lower the pain threshold and thus more easily lead to painful experiences after activities, forming a vicious cycle of fear reinforcement [4,5].

However, this seems to be different in the case of postsurgical pain. Our previous research [6] has shown that the FOP is related to postsurgical pain and can lead to avoidance behavior. A significant portion of this FOP is learned from others' teaching or observational learning, while the FOP being directly caused by painful experiences is relatively rare. This seems to indicate that some people refuse to engage in certain activities such as

XSL•FO
RenderX

coughing and ambulation from the beginning based on their cognition, rather than pain experience, as they have heard others tell them that such activities are dangerous.

A good pain experience after activities can help break the cycle of the FOP. The fear-avoidance model does not explain the dynamics underlying the FOP and functional recovery [4], such as how some postoperative patients can still continue to perform rehabilitation activities despite pain. It seems that there is an internal drive that prompts patients to overcome their fears and produce positive recovery behaviors.

Psychological resilience, also known as mental toughness [7], refers to an individual's capacity to adapt effectively to adversity, challenges, and threats [8]. Psychological resilience has a positive impact in patients undergoing knee surgery and those with pain catastrophizing [9], as well as its predictive role in dental phobia [10]. However, its relationship with postsurgical pain and the FOP remains understudied.

Individuals with low psychological resilience tend to avoid stressful situations [11]. In this study, we aimed to understand the role of psychological resilience in overcoming the FOP and avoidance behavior among surgical patients. To this end, we followed up a sample from a previous study [12] to observe the impact of psychological resilience on the FOP and behavior, aiming to assist in postsurgical pain management.

Methods

Participants

This study was a single-center, cohort study of patients who were hospitalized to undergo thoracic surgery at the Wuhan Union Hospital, a tertiary hospital located in Wuhan, China, from May 2022 to January 2023. Participants were recruited based on the following inclusion criteria: (1) they were at least 18 years old; (2) they had undergone lung surgery; and (3) they did not have a history of neurological or psychiatric disorders. The exclusion criteria included individuals who experienced critical conditions during the study period or those who did not complete follow-up assessments. In the context where the sample size exceeds 10 events per variable [13], and considering a potential 10% dropout rate, we would terminate the further enrollment of participants.

Ethical Considerations

This study was a follow-up study based on earlier research [12] (Registration number: ChiCTR2200056651), and has been approved by the institutional ethics board of Wuhan Union Hospital of Tongji Medical College, Huazhong University of Science and Technology (No. 20220026 - 2). All the participants provided written and oral informed consent. The original informed consent form, which was approved by the same ethics committee, explicitly allows for secondary analysis without the need for additional consent (No. 20200351). Participants were provided with free counseling services during the follow-up period as a form of compensation. All study data have been anonymized or deidentified to protect the privacy and confidentiality of the participants.

Fear of Pain

Before surgery and 6 months after surgery, we used the simplified Chinese version of the Fear of Pain-9 items (FOP-9) [14] to gather data on the FOP. The FOP-9 is a 5-point ordinal Likert scale where total scores range from 9 to 45, and a higher score signifies a greater level of FOP.

Fear Avoidance Belief

On the third day after surgery, the Physical Activity subscale of the Fear-Avoidance Beliefs Questionnaire (FABQ-PA) was used to measure the FAB. The FABQ-PA is a self-reported questionnaire comprising the first 5 questions of the original Fear-Avoidance Beliefs Questionnaire [15]. A higher score on this subscale signifies a greater tendency towards the FAB.

Psychological Resilience

Psychological resilience was collected 1 month after surgery by using the Connor-Davidson Resilience Scale (CD-RISC) [16]. The CD-RISC comprises 25 items, each rated on a 5-point scale (0 - 4), with higher scores reflecting greater psychological resilience.

Activity Recovery

Activity recovery was defined as the recovery of social activities and housework after discharge. Activity recovery was assessed at 6 months after surgery through questions such as, "Are your current social activities, like work and community activities, similar to what they were before the surgery?" and "Have your current family responsibilities returned to what they were before the surgery?" The answers were categorized as being lower or similar/higher than they were before surgery. Both the answers of "similar/higher" meant that the activity recovery was positive.

Statistics

We used the adaptive least absolute shrinkage and selection operator (Lasso) with nested cross-validation to conduct variable selection, coefficient estimation, and performance prediction. All the statistics were completed using Python 3.12 (Python Software Foundation) and the scikit-learn library (developed by a community of open-source contributors, initially started by David Cournapeau). The process was as follows [17]: (1) Feature preprocessing: We conducted standardization on numeric features and one-hot encoding on categorical features. In order to avoid the problem of multicollinearity, we set the first one as the baseline and dropped it for each categorical variable, such as "Male" in sex, "Primary School" in education, "Presence" in chronic pain, and "Wedge Resection" in the surgery method. (2) Model training: We estimated the initial coefficients using Ridge regression with 5-fold cross-validation, and then conducted model training with the adaptive Lasso with 5-fold cross-validation after weighing the coefficients. (3) Model evaluation: In the 5-fold outer cross-validation, we selected the optimal model and calculated the average value of the model fitting parameters. (4) Variable estimation: We reported the regularization parameters and variable coefficients of the optimal model.

XSL•FO
RenderX

Results

Characteristics

A total of 150 patients met the study's inclusion criteria. However, 6 (4%) patients were excluded as they were lost to follow-up. A total of 144 participants were included in the final analysis, and the baseline characteristics are listed in Table 1. The study flow is shown in Figure 1.

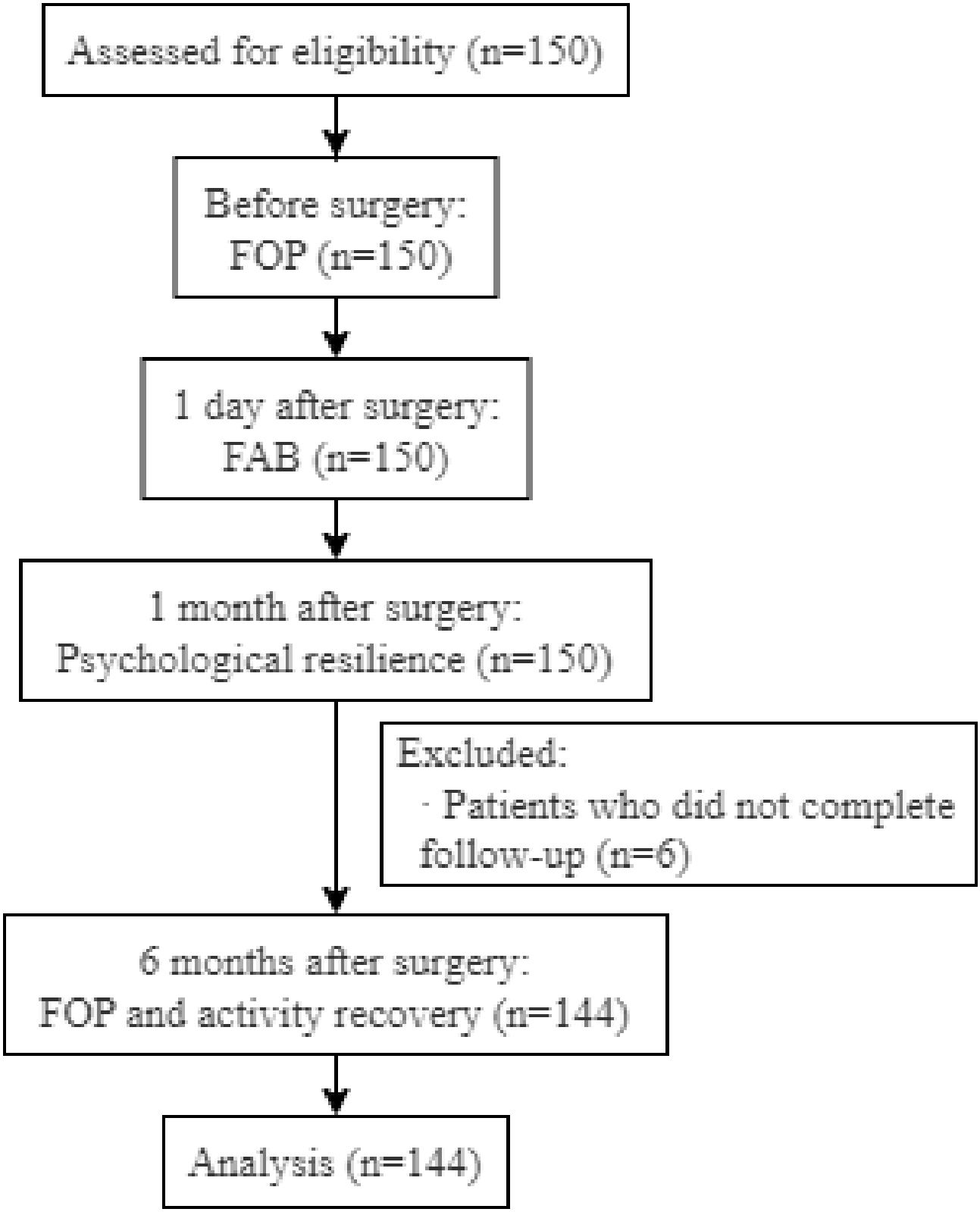
Table . Baseline patient characteristics (n=144).

Characteristics	Value
Age (years), mean (SD)	52.40 (13.37)
Sex, n (%)	
Male	73 (50.7)
Female	71 (49.3)
Education, n (%)	
Primary school	20 (13.9)
Middle school	30 (20.8)
High school	43 (29.9)
University	51 (35.4)
Chronic pain, n (%)	
Absence	137 (95.1)
Presence	7 (4.9)
Preoperative FOP[a], median (IQR)	16 (13-23)
Surgery method, n (%)	
Wedge resection	24 (16.7)
Segmental resection	64 (44.4)
Lobectomy	56 (38.9)
Pain intensity, median (IQR)	3 (1-4)
Fear avoidance belief, median (IQR)	11 (8-15)
Psychological resilience, mean (SD)	49.54 (12.99)
Activity recovery, n (%)	
Lower	38 (26.4)
Similar/Higher	106 (73.6)
Postoperative FOP, median (IQR)	20 (14-28)

[a]FOP: fear of pain.

XSL•FO
RenderX

Figure 1. Flowchart of the research process and data collection. FOP: fear of pain; FAB: fear avoidance belief.



Postoperative FOP as the Dependent Variable

The statistical analysis showed that the mean squared error was 14.273 (SD 3.254), indicating a reasonable prediction error range with room for improvement. The mean R^2 value was 0.731 (SD 0.060), suggesting that the models could account for approximately 73.1% (SD 6%) of the variance in postoperative FOP and had a relatively good fit. The optimal alpha of the optimal model was 0.081. The coefficient of the optimal model revealed that the preoperative FOP, FAB, and pain intensity had significant positive effects on the postoperative FOP with coefficients of 8.620, 8.560, and 0.417, respectively. In contrast, psychological resilience and age had significant negative impacts, with coefficients of −5.822 and −2.853, respectively. Among categorical variables, patients with a university education had lower postoperative FOP compared to those with

XSL•FO
RenderX

primary school education, while female patients had a higher postoperative FOP than male patients did. The surgical method and chronic pain had no significant effect on the postoperative FOP (Table 2).

Table . Optimal model of adaptive least absolute shrinkage and selection operator (Lasso) regression for patients who underwent lung surgery in predicting the postoperative fear of pain (FOP; n=144).

Independent variables[a]	Coefficient
Age	–2.853
Sex	
Male	0.000
Female	0.778
Education	
Primary school	0.000
Middle school	−0.000
High school	0.000
University	−0.190
Chronic pain	
Absent	0.000
Present	0.000
Preoperative FOP	8.620
Surgery method	
Wedge resection	0.000
Segmental resection	-0.005
Lobectomy	0.000
Pain intensity	0.417
Fear avoidance belief	8.560
Psychological resilience	−5.822

[a]Lambda is 0.081 for all variables.

Activity Recovery as the Dependent Variable

The model performed excellently after analysis, with an average accuracy of 0.820 (SD 0.024) and an average area under the curve of 0.926 (SD 0.044), indicating that the model had a good predictive ability and stability for activity recovery. Psychological resilience, FAB, and pain intensity were significant factors that affected the activity recovery, with coefficients of 1.185, −0.966, and −0.170, respectively. Both university education and segmental resection significantly positively affected activity recovery compared to baseline, with coefficients of 0.283 and 0.042, respectively. Age, sex, chronic pain, and preoperative FOP had no significant effect on activity recovery (Table 3).

XSL•FO
RenderX

Table . Optimal model of adaptive least absolute shrinkage and selection operator (Lasso) logistic regression for patients who underwent lung surgery in predicting activity recovery (n=144).

Independent variables[a]	Coefficient
Age	0.000
Sex	
Male	0.000
Female	0.000
Education	
Primary school	0.000
Middle school	0.000
High school	0.000
University	0.283
Chronic pain	
Absent	0.000
Present	0.000
Preoperative FOP[b]	0.000
Surgery method	
Wedge resection	0.000
Segmental resection	0.042
Lobectomy	0.000
Pain intensity	−0.170
Fear avoidance belief	−0.966
Psychological resilience	1.185

[a]Lambda is 2.783 for all variables.

[b]FOP: fear of pain.

Discussion

Study Findings

This study delved into the role of psychological resilience in perioperative FOP and avoidance behavior among surgical patients, particularly those undergoing lung surgery. The findings contribute novel insights to the existing literature, suggesting that psychological resilience may serve as a crucial factor in breaking the vicious cycle of FOP and facilitating recovery.

Impact of Psychological Resilience on Postoperative FOP

Our study employed adaptive Lasso regression under nested cross-validation to analyze the factors influencing postoperative FOP, a common phenomenon experienced by patients following surgery [18]. The developed model demonstrated robust performance, with a high proportion of variance explained in postoperative FOP. Among the factors, psychological resilience emerged as the most significant inhibitory factor influencing postoperative FOP. Preconceived biases regarding postoperative activities and pain can lead to the development of the FOP [19,20]. We observed that some patients strictly adhered to the doctor's orders despite potential pain, while others rejected and avoided them. This disparity can be attributed to the varying levels of psychological resilience among individuals. Resilient individuals are more likely to accept and face pain, actively adjust their pain concepts, and gradually expose themselves to pain stimuli [21,22]. For instance, they may perceive postoperative pain as a normal part of the recovery process and actively communicate with medical staff to learn pain management strategies, gradually adapting to and overcoming the FOP. Conversely, individuals with lower psychological resilience may avoid stressful situations, including postoperative activities, leading to a vicious cycle of pain, fear, and avoidance [9,10,23].

Impact of Psychological Resilience on Postoperative Activity Recovery

The adaptive Lasso logistic regression exhibited higher accuracy and area under the curve values with a lower standard deviation, indicating that the model has high accuracy and stability in predicting postoperative activity recovery. The fact that the coefficient of psychological resilience is the largest revealed that psychological resilience had a stronger influence on postoperative activity recovery than other factors. This suggests that psychological resilience primarily enables individuals to initiate activities that may cause pain despite the presence of FOP, serving as a positive factor in overcoming fear. Patients

XSL•FO
RenderX

with higher psychological resilience tend to have better ability to resume their activities postoperatively [24,25], and our research has also confirmed this. They may be more actively involved in social interactions and resume daily housework, thereby promoting physical function recovery and improving psychological well-being, resulting in a virtuous cycle. This may be because patients with high psychological resilience are better able to cope with negative emotions, have more reasonable cognitive expectations of the rehabilitation process, and can better overcome the fear of activities, thus promoting the recovery of postoperative activities [26,27].

Study Limitations

This study has several limitations. First, psychological resilience was measured postoperatively, which may introduce bias due to the impact of surgery. Second, although both models demonstrated certain predictive abilities, there is room for improvement in the R^2 values. Third, the generalizability of the findings is limited, as this study specifically focused on patients undergoing lung surgery, and the results may not be applicable to all surgical patients. Therefore, conducting multicenter studies with a bigger sample size would further validate the robustness of the results.

Conclusion

This study revealed the significant role of psychological resilience in postsurgical pain management, particularly in reducing the FOP and promoting recovery of activities. Individuals with higher psychological resilience are more likely to effectively cope with postsurgical pain and reduce avoidance behaviors. These findings underscore the importance of assessing and potentially enhancing psychological resilience during the perioperative period, especially for those with higher levels of FOP. By fostering resilience and promoting correct pain cognition, clinical practitioners can empower patients to overcome the FOP, engage in rehabilitation activities, and achieve better postoperative outcomes.

Acknowledgments

This study was supported by the general program of Hubei Provincial Natural Science Foundation of China (No. 2021CFB588).

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Sugano R, Ikegami K, Ando H, et al. The relationship between fear-avoidance beliefs in employees with chronic musculoskeletal pain and work productivity: a longitudinal study. J UOEH 2020;42(1):13-26. [doi: 10.7888/juoeh.42.13] [Medline: 32213739]
2. Meier ML, Stämpfli P, Humphreys BK, Vrana A, Seifritz E, Schweinhardt P. The impact of pain-related fear on neural pathways of pain modulation in chronic low back pain. Pain Rep 2017 May;2(3):e601. [doi: 10.1097/PR9.0000000000000601] [Medline: 29392216]
3. Luo H, Cai Z, Huang Y, et al. Study on pain catastrophizing from 2010 to 2020: a bibliometric analysis via CiteSpace. Front Psychol 2021;12:759347. [doi: 10.3389/fpsyg.2021.759347] [Medline: 34975649]
4. Crombez G, Eccleston C, Van Damme S, Vlaeyen JWS, Karoly P. Fear-avoidance model of chronic pain: the next generation. Clin J Pain 2012 Jul;28(6):475-483. [doi: 10.1097/AJP.0b013e3182385392] [Medline: 22673479]
5. Varangot-Reille C, Pezzulo G, Thacker M. The fear-avoidance model as an embodied prediction of threat. Cogn Affect Behav Neurosci 2024 Oct;24(5):781-792. [doi: 10.3758/s13415-024-01199-4] [Medline: 38890209]
6. Luo Y, He J, Bao L, Meng H, Hu C, Chen Q. Fear of pain as a predictor for postoperative pain intensity among the patients undergoing thoracoscopic surgery. Pain Res Manag 2022;2022:2201501. [doi: 10.1155/2022/2201501] [Medline: 35757293]
7. Soundara Pandian PR, Balaji Kumar V, Kannan M, Gurusamy G, Lakshmi B. Impact of mental toughness on athlete's performance and interventions to improve. J Basic Clin Physiol Pharmacol 2023 Jul 1;34(4):409-418. [doi: 10.1515/jbcpp-2022-0129] [Medline: 35792085]
8. Bögemann SA, Puhlmann LMC, Wackerhagen C, et al. Psychological resilience factors and their association with weekly stressor reactivity during the COVID-19 outbreak in Europe: prospective longitudinal study. JMIR Ment Health 2023 Oct 17;10:e46518. [doi: 10.2196/46518] [Medline: 37847551]
9. Stamatis A, Morgan GB, Spinou A, Tsigaridis KG. Mental toughness and osteoarthritis: postsurgery improvement in knee pain/functionality in older adults. Rehabil Psychol 2023 May;68(2):212-219. [doi: 10.1037/rep0000499] [Medline: 37126026]
10. Tian H, Ding X, Lin C, Qu P, Fan X. Study of dental phobia, psychological resilience, and related factors in children aged 8-9 years in Weifang city. Hua Xi Kou Qiang Yi Xue Za Zhi 2023 Feb 1;41(1):73-79. [doi: 10.7518/hxkq.2023.01.010] [Medline: 38596944]

XSL•FO
RenderX

11. Babić R, Babić M, Rastović P, et al. Resilience in health and illness. Psychiatr Danub 2020 Sep;32(Suppl 2):226-232. [Medline: 32970640]
12. Bao L, Peng C, He J, Sun C, Feng L, Luo Y. The relationship between fear avoidance belief and threat learning in postoperative patients after lung surgery: an observational study. Psychol Res Behav Manag 2023;16:3259-3267. [doi: 10.2147/PRBM.S420724] [Medline: 37605755]
13. Tabachnick BG, Fidell LS. Using Multivariate Statistics: Pearson Education; 2013. URL: https://books.google.com.sg/books?id=ucj1ygAACAAJ [accessed 2023-12-11]
14. Luo Y, Li S, He J, et al. Translation and validation of fear of pain-9 items into simplified Chinese version for Mainland China. J Pain Res 2021;14:35-40. [doi: 10.2147/JPR.S275227] [Medline: 33469354]
15. Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A fear-avoidance beliefs questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain 1993 Feb;52(2):157-168. [doi: 10.1016/0304-3959(93)90127-B] [Medline: 8455963]
16. Connor KM, Davidson JRT. Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). Depress Anxiety 2003;18(2):76-82. [doi: 10.1002/da.10113] [Medline: 12964174]
17. Ballout N, Etievant L, Viallon V. On the use of cross-validation for the calibration of the adaptive lasso. Biom J 2023 Jun;65(5):e2200047. [doi: 10.1002/bimj.202200047] [Medline: 36960476]
18. Ustunel F, Erden S. Evaluation of fear of pain among surgical patients in the preoperative period. J Perianesth Nurs 2022 Apr;37(2):188-193. [doi: 10.1016/j.jopan.2021.02.003] [Medline: 34922831]
19. Pak R, Mahmoud Alilou M, Bakhshipour Roudsari A, Yousefpour F. Experiential avoidance as a factor in generalized psychological vulnerability: in the relationship between chronic pain and pain anxiety with pain disability. Pain Manag Nurs 2024 Jun;25(3):e256-e264. [doi: 10.1016/j.pmn.2024.01.007] [Medline: 38418316]
20. Shojaie D, Hoffman AS, Amaku R, et al. Decision making when cancer becomes chronic: needs assessment for a web-based medullary thyroid carcinoma patient decision aid. JMIR Form Res 2021 Jul 16;5(7):e27484. [doi: 10.2196/27484] [Medline: 34269691]
21. Dursun A, Kaplan Y, Altunbaş T, Bahtiyar M. The mediating effect of experiential avoidance on the relationship between psychological resilience and psychological needs in the COVID-19 pandemic. Curr Psychol 2022 May 19;43(14):1-11. [doi: 10.1007/s12144-022-03085-z] [Medline: 35607644]
22. Seçer İ, Ulaş S, Karaman-Özlü Z. The effect of the fear of COVID-19 on healthcare professionals' psychological adjustment skills: mediating role of experiential avoidance and psychological resilience. Front Psychol 2020;11:561536. [doi: 10.3389/fpsyg.2020.561536] [Medline: 33192830]
23. Kartal M, Kapikiran G, Karakas N. The effect of emergency nurses' psychological resilience on their thanatophobic behaviors: a cross-sectional study. Omega (Westport) 2024 Aug;89(3):1273-1285. [doi: 10.1177/00302228221128156] [Medline: 36117453]
24. The REGAIN (Regional versus General Anesthesia for Promoting Independence after Hip Fracture) Investigators. Preoperative psychological resilience and recovery after hip fracture: secondary analysis of the regain randomized trial. J American Geriatrics Society 2023 Dec;71(12):3792-3801. [doi: 10.1111/jgs.18552]
25. Berthold DP, Bormann M. Editorial commentary: psychological resilience promotes positive outcomes after surgery. Arthroscopy: The Journal of Arthroscopic & Related Surgery 2024 Dec;40(12):2895-2896. [doi: 10.1016/j.arthro.2024.04.004]
26. Al Ta'ani Z, Al Ta'ani O, Gabr A, et al. From fear to resilience: a scoping review of psychological components in anterior cruciate ligament rehabilitation. J Sport Rehabil 2024 Nov 1;33(8):591-618. [doi: 10.1123/jsr.2023-0420] [Medline: 39299682]
27. Broche-Perez Y, Jimenez-Morales RM, Vázquez-Gómez LA, Bauer J, Fernández-Fleites Z. Fear of relapse and quality of life in multiple sclerosis: the mediating role of psychological resilience. Mult Scler Relat Disord 2023 Nov;79:105026. [doi: 10.1016/j.msard.2023.105026] [Medline: 37839366]

Abbreviations

CD-RISC: Connor-Davidson Resilience Scale
FAB: fear avoidance belief
FABQ-PA: Physical Activity subscale of the Fear-Avoidance Beliefs Questionnaire
FOP: fear of pain
FOP-9: Fear of Pain-9 items
Lasso: least absolute shrinkage and selection operator

XSL•FO
RenderX

Edited by A Mavragani; submitted 23.06.24; peer-reviewed by KE Cain, P Humburg; revised version received 18.12.24; accepted 19.12.24; published 07.02.25.

Please cite as:
Luo Y, Li S, Feng L, Zheng J, Peng C, Bao L
Impact of Psychological Resilience on the Fear of Pain and Activity Recovery in Postsurgical Patients: Observational Cohort Study
JMIR Form Res 2025;9:e63556
URL: https://formative.jmir.org/2025/1/e63556
doi:10.2196/63556

© Yang Luo, Sisi Li, Lijuan Feng, Junyi Zheng, Chunfen Peng, Lihong Bao. Originally published in JMIR Formative Research (https://formative.jmir.org), 7.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impact of Preventive Intravenous Amiodarone on Reperfusion Ventricular Fibrillation in Patients With Left Ventricular Hypertrophy Undergoing Open-Heart Surgery: Randomized Controlled Clinical Trial

Chan-Juan Gong*, MD; Xiao-Kai Zhou*, MD; Zhen-Feng Zhang, MD; Yin Fang, MD

Department of Anesthesiology and Perioperative Medicine, The First Affiliated Hospital of Nanjing Medical University, No. 300 Guangzhou Road, Nanjing, China

*these authors contributed equally

Corresponding Author:
Yin Fang, MD
Department of Anesthesiology and Perioperative Medicine, The First Affiliated Hospital of Nanjing Medical University, No. 300 Guangzhou Road, Nanjing, China

Abstract

Background: Ventricular fibrillation (VF) is a vicious arrhythmia usually generated after removal of the aortic cross-clamp (ACC) in patients undergoing open-heart surgery, which could damage cardiomyocytes, especially in patients with left ventricular hypertrophy (LVH). Amiodarone has the prominent properties of converting VF and restoring sinus rhythm. However, few studies concentrated on the effect of amiodarone before ACC release on reducing VF in patients with LVH.

Objective: The study was designed to explore the effectiveness of prophylactic intravenous amiodarone in reducing VF after the release of the ACC in patients with LVH.

Methods: A total of 54 patients with LVH scheduled for open-heart surgery were enrolled and randomly divided (1:1) into 2 groups—group A (amiodarone group) and group P (placebo-controlled group). Thirty minutes before removal of the ACC, the trial drugs were administered intravenously. In group A, 150 mg of amiodarone was pumped in 15 minutes. In group P, the same volume of normal saline was pumped in 15 minutes. The primary outcome was the incidence of VF 10 minutes after removal of the ACC.

Results: The incidence of VF was lower in group A than in group P (30% vs 70%, *P*=.003). The duration of VF, the number of defibrillations, and the defibrillation energy were also lower in group A than in group P (*P*<.001, *P*=.002, and *P*=.002, respectively). After the end of cardiopulmonary bypass, the heart rate and mean arterial pressure were lower in group A, and the mean pulmonary arterial pressure and the dose of vasoactive drugs were higher than those in group P (*P*<.001, *P*<.001, *P*=.04, and *P*=.02, respectively). However, there were no significant differences in the use of vasoactive-inotropic agents and hemodynamic status between the 2 groups before the end of surgery.

Conclusions: In patients with LVH who undergo open-heart surgery, amiodarone can be safely used to reduce the incidence of VF, the duration of VF, the frequency of defibrillation, and the energy of defibrillation after ACC removal.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2000035057; https://www.chictr.org.cn/showprojEN.html?proj=57145

(JMIR Form Res 2025;9:e64586) doi:10.2196/64586

KEYWORDS

amiodarone; left ventricular hypertrophy; reperfusion ventricular fibrillation; open-heart surgery; randomized controlled trial; RCT; clinical trial; ventricular fibrillation; vicious arrhythmia; aortic cross-clamp; surgery; effectiveness; defibrillation

Introduction

Ventricular fibrillation (VF) is commonly seen after release of the aortic cross-clamp (ACC) during open-heart surgery, which can result in reperfusion VF (rVF) when reperfusion of the myocardium begins. The incidence of rVF after the ACC release has been reported to range from 10% to 80% [1]. It has been proven in the animal model that persistent VF could cause more myocardial fibrosis [2]. The use of internal electrical defibrillation is a successful approach to terminate VF. However, multiple shocks and prolonged VF duration may exacerbate myocardial damage and diminish the patient's prognosis [3]. Besides, repeated defibrillations would increase the incidence of postoperative bradycardia in cardiac surgeries [4]. Therefore, it is necessary to reduce the rVF after removal of the ACC during cardiac surgery.

XSL•FO
RenderX

In patients with left ventricular hypertrophy (LVH), the unique myocardial structure hinders the complete perfusion of fluid during cardiopulmonary bypass (CPB) surgery to reach the subendocardial cardiomyocytes. This limitation diminishes the efficacy of myocardial protection in open-heart surgery [5]. Following the release of the ACC, there is a notable rise in the occurrence of rVF, leading to an escalation in both the energy and frequency of cardiac shocks within the hypertrophic myocardium [6]. Thus, it is also of vital clinical significance to decrease the occurrence of rVF in patients with LVH.

Amiodarone, known for its classification as a potent Class III antiarrhythmic drug with diverse electrophysiological properties, has been extensively used in clinical practice for the treatment of atrial and ventricular arrhythmias [7,8]. Intravenous amiodarone has been confirmed to protect the myocardium from ischemia-reperfusion injury and prevent reperfusion-induced VF [9]. It also showed a good hemodynamic tolerance when administered intravenously [10]. Previous studies have demonstrated the efficacy of amiodarone in reducing the occurrence of VF after the removal of the ACC [11,12]. However, there is only 1 clinical trial investigating the role of amiodarone in reducing rVF in patients with LVH [13]. Therefore, we designed the study to investigate the effects of prophylactic intravenous amiodarone on rVF after the release of the ACC in patients with LVH undergoing open-heart surgery.

Methods

Settings and Patients

This was a prospective, double-blind, randomized controlled trial conducted in the First Affiliated Hospital of Nanjing Medical University, Nanjing, China. The patients undergoing open-heart surgery were screened by preoperative transthoracic echocardiography (TTE). When meeting the eligible criteria, the patients were recruited in the study. A total of 54 patients with LVH undergoing open-heart surgery were enrolled and participated in this study from August 2022 to November 2023. This manuscript adheres to all applicable CONSORT (Consolidated Standards of Reporting Trials) guidelines (Checklists 1 and 2). The protocol of this study has already been published [14]. The eligibility criteria are hence described.

The inclusion criteria included (1) patients undergoing open-heart surgery, (2) aged between 18 and 75 years, (3) diagnosed with LVH evaluated by TTE preoperatively, and (4) classified as American Society of Anesthesiology (ASA) Physical Status I-III. LVH was characterized by a left ventricular mass index>115 g/m^2 for men and >95 g/m^2 for women [15].

The exclusion criteria included preoperative administration of Class I or III antiarrhythmic medications; QT interval of 500 milliseconds or higher or corrected QT interval of 480 milliseconds or higher; left ventricular ejection fraction below 40%; and presence of thyroid dysfunction, electrolyte imbalances, liver or renal dysfunction, pulmonary interstitial fibrosis, or known allergy to amiodarone. Patients undergoing emergency surgery, redo surgery, or concomitant procedures, including coronary artery bypass grafting (CABG), were also excluded.

Ethical Considerations

This study was approved by the Ethical Committee of the First Affiliated Hospital of Nanjing Medical University (2020-SR-175) and registered in the Chinese Clinical Trial Registry (ChiCTR2000035057). All participants provided verbal informed consent before being enrolled in the study. To secure and protect all participant information, all data were recorded by the specific ID number instead of the participant's name throughout the study. Only select research team members could access these data. The process was monitored by the Ethical Committee of the First Affiliated Hospital of Nanjing Medical University.

Randomization and Blinding

After anesthetic induction, the enrolled participants were assigned with a random permuted block method in a 1:1 ratio to 2 groups, namely, group A (treated with amiodarone) and group P (placebo control). An unblinded nurse anesthetist prepared the drugs according to the allocation information in the randomization envelope. The drugs were uniformly packaged, numbered, and administered by anesthesiologists. All involved parties, including the anesthesiologists, surgeons, and participants, were unaware of the specific treatment regimen, ensuring a blinded study design. However, nonblinded nurse anesthetists were not involved in evaluating treatment interventions, analyzing data, or interpreting results in the study.

Treatment Regimen of Amiodarone

The trial drugs were administered 30 minutes before ACC release through the central venous line. In group A, 150 mg of amiodarone was diluted to 20 ml of normal saline and pumped in 15 minutes. In group P, 20 ml of normal saline was pumped in 15 minutes.

Anesthesia Protocols

All patients received standard general anesthesia. In the operating room, patients were routinely monitored, including 5-lead electrocardiograph, peripheral blood oxygen saturation, invasive arterial blood pressure, bispectral index, and end-tidal CO_2. Anesthesia induction was carried out as follows: 0.05 mg/kg midazolam, 0.3 mg/kg etomidate, 0.15 mg/kg cis-atracurium, and 4 - 6 μg/kg fentanyl. Approximately 5 minutes after induction, patients were intubated and ventilated to maintain end-tidal CO_2 at 35 - 45 mm Hg. Then, ultrasound-guided central vein catheterization was performed through the right internal jugular or subclavian vein. The Swan-Gans catheter was floated to the pulmonary artery through the right internal jugular vein. The patients continued to receive anesthesia by inhalation of sevoflurane 1% - 2% and continuous pump of propofol 1 - 2 mg/kg/h, dexmedetomidine 0.2 - 0.7 μg/kg/min, and cis-atracurium 0.05 - 0.1 mg/kg/h. The total dose of fentanyl was 30 - 50 μg/kg. The bispectral index was maintained between 40 and 60. Inotropic agents, vasopressors, vasodilators, or fluid administration were applied to achieve the following hemodynamic status: a mean arterial pressure

XSL•FO
RenderX

(MAP) of 50 - 80 mm Hg, cardiac index (CI) ≥2.2 L/min/m^2, and urine output ≥1.0 mL/kg/h.

Surgical and CPB Procedures

The surgical procedures were performed by median sternotomy with heparinization and standard CPB. Anticoagulation was started with 400 U/kg of heparin and maintained according to an activated clotting time of >480 seconds. The CPB was initiated using a membrane oxygenator and the circuit was primed with 1500 ml of Ringer acetate solution and 200 ml of 50% human serum albumin. The patients were perfused with a nonpulsatile roller pump at a flow rate of 2.0 - 2.4 L/min/m^2 and with MAP of 50 - 80 mm Hg. Moderate systemic hypothermia was maintained at 30 - 32℃ during CPB. Myocardial protection was achieved by intermittent antegrade or retrograde infusions of histidine-tryptophane-ketoglutarate cardioplegia. After weaning from CPB, the reversal of heparin was achieved with protamine.

Management of VF

Thirty seconds before ACC release, 1 mg/kg of esmolol was administered to all patients. When VF occurred after ACC release, internal electrical defibrillation with 20 J was performed in the first attempt. If not successful, the second defibrillation with 20 J was attempted after reusing 0.5 - 1 mg/kg of esmolol. If VF persisted, the third 30 J defibrillation was performed after giving 1 mg/kg of lidocaine. If termination of VF was not achieved, the fourth defibrillation with 40 J was attempted after the administration of 50 mg of amiodarone administration. If VF still existed, the administration of 100 mg of amiodarone and defibrillation with 50 J was repeated until VF terminated.

Outcome Measures

The primary outcome was the incidence of VF 10 minutes after ACC release. The secondary outcomes included the duration of VF, the number and total energy of defibrillations needed to terminate VF, and the use of other antiarrhythmic agents and epicardial temporary pacemakers after the release of the ACC. Furthermore, MAP, heart rate (HR), mean pulmonary arterial pressure (MPAP), central venous pressure (CVP), and CI were measured 15 minutes after anesthetic induction (T1), 15 minutes after weaning from CPB (T2), and 15 minutes before the end of surgery (T3). In addition, the vasoactive-inotropic score (VIS) and the total dose of vasoactive and inotropic agents (TDA) were also recorded at T2 and T3 (VIS = dopamine [μg/kg/min] + dobutamine [μg/kg/min] + 10 × milrinone [μg/kg/min] + 100 × epinephrine [μg/kg/min] + 100 × norepinephrine [μg/kg/min] + 1000 × vasopressin [U/kg/min]; TDA = dopamine [mg] + dobutamine [mg] + 10 × milrinone [mg] + 100 × epinephrine [mg] + 100 × norepinephrine [mg] + 1000 × vasopressin [U]).

Sample Size Estimation

The sample size was calculated based on the primary outcome using PASS 15.0 (NCSS, LLC). According to published studies, the incidence of VF after ACC release ranged from 10% (2/20) to 80% (38/47) [1]. Based on our preliminary data, the incidence of VF 10 minutes after ACC release was 70% (7/10) in the placebo-controlled group, while the VF rate was 30% (3/10) in the amiodarone group. To reach statistical significance with a power of 80% and an α value of .05 (2-sided), 21 patients were needed in each group. Assuming a 20% dropout or missing data rate, we recruited 27 patients in each group for a total of 54 patients.

Statistical Analysis

All data were analyzed using SPSS 22.0 (IBM Corp) and GraphPad Prism 8.0 (GraphPad Software Inc). All statistical tests were considered significant if $P<.05$. Only 2-sided tests were used. The normality assumption was assessed with the Kolmogorov-Smirnov test for all tests. Continuous variables were presented as mean (SD) or median (IQR), depending on the distribution. Categorical variables were presented as frequency (%). Continuous variables were compared using Student *t* tests if normally distributed or Mann-Whitney *U* tests if not normally distributed. The proportions were compared using Pearson chi-square or continuity correction tests. Data for repeated measures were compared using repeated measures ANOVA.

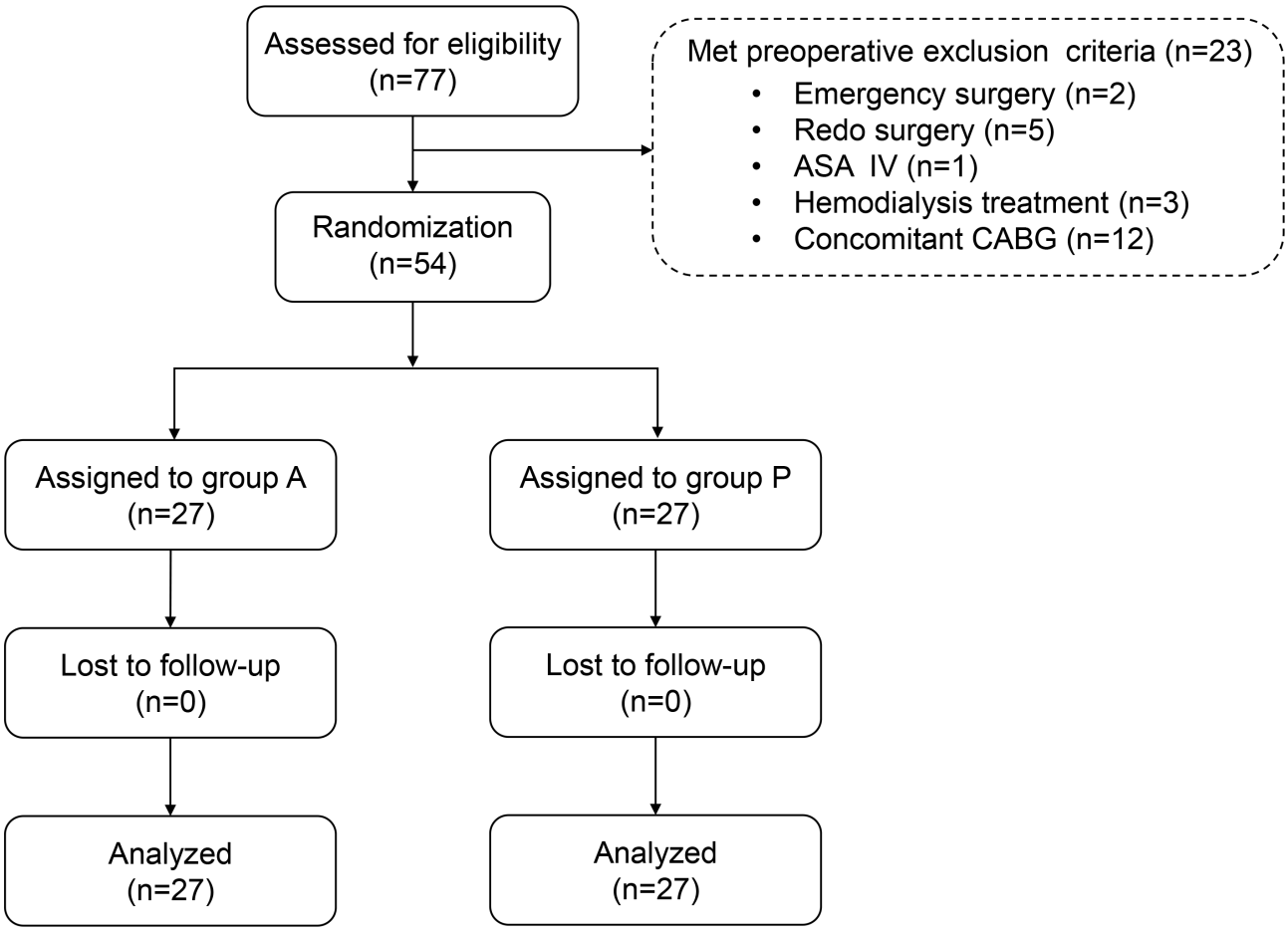
Results

Baseline Characteristics

In total, 77 patients with LVH who underwent open-heart surgery were selected for this study. Among them, 54 patients (27 in group A and 27 in group P) completed this trial and 23 patients were excluded, which were shown in the CONSORT flow diagram (Figure 1). The demographic and perioperative data of the participants were comparable between groups (Table 1).

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of patient recruitment, randomization, and withdrawal. A total of 27 patients in each group were eventually included in the final analysis. ASA: American Society of Anesthesiologists; CABG: coronary artery bypass grafting.



XSL•FO
RenderX

Table . Demographic characteristics and perioperative data.

Perioperative information		Group A[a] (N=27)	Group P[b] (N=27)	*P* value
Preoperative				
	Age (y), mean (SD)	64.44 (9.94)	65.59 (6.30)	.62
	Male, n	19	14	.16
	BMI (kg/m^2), mean (SD)	22.74 (2.42)	23.50 (1.92)	.21
	Hypertension, n	10	6	.23
	Diabetes mellitus, n	4	7	.50
	Hyperlipidemia. n	3	2	.99
	Smoking, n	6	4	.73
	NYHA[c] class, mean (SD)	2.63 (0.49)	2.41 (0.50)	.11
	LVMI[d] (g/m^2), mean (SD)	134.40 (11.74)	132.44 (15.05)	.60
	LVEF[e] (%), mean (SD)	57.11 (3.94)	58.78 (5.51)	.21
	Calcium blockers, n	6	8	.54
	β-blockers, n	20	16	.19
	Diuretics, n	7	6	.75
	ACEIs[f], ARBs[g], n	10	11	.78
Intraoperative				
	Procedures, n			.20
	AVR[h]	18	23	
	Morrow	9	4	
	Surgical time (min), mean (SD)	358.44 (35.67)	375.04 (35.55)	.09
	CPB[i] time (min), mean (SD)	154.89 (30.98)	168.70 (33.48)	.12
	ACC[j] time (min), mean (SD)	135.33 (28.91)	143.81 (29.53)	.29
	Time for CPB weaning (min), mean (SD)	28.04 (10.81)	30.59 (8.69)	.42
	Blood loss (mL), mean (SD)	528.89 (213.87)	496.74 (184.61)	.56
	Urine output (mL), mean (SD)	1350.00 (527.36)	1417.04 (754.10)	.71
	Total fluid volume (mL), mean (SD)	3727.04 (423.72)	3562.59 (448.78)	.17
	Total cardioplegia (mL), mean (SD)	1385.19 (317.39)	1412.96 (374.06)	.77
	Abnormal electrolyte or acid-base metabolism, n	0	0	.99
Postoperative, mean (SD)				
	Ventilation time (h)	10.26 (4.69)	11.44 (4.93)	.37
	ICU[k] stay (d)	4.96 (3.20)	5.93 (3.76)	.32
	Hospital stays (d)	11.85 (4.02)	13.74 (5.41)	.15

[a]Group A: amiodarone group.

[b]Group P: placebo control group.

[c]NYHA: New York Heart Association.

[d]LVMI: left ventricular mass index.

[e]LVEF: left ventricular ejection fraction.

[f]ACEI: angiotensin converting enzyme inhibitor.

[g]ARB: angiotensin receptor blocker.

[h]AVR: aortic valve replacement.

[i]CPB: cardiopulmonary bypass.

XSL•FO
RenderX

[j]ACC: aortic cross-clamp.

[k]ICU: intensive care unit.

Intraoperative Outcomes

The incidence of VF after the release of the ACC was significantly lower in group A than in group P (8/27, 30% vs 19/27, 70%; *P*=.003) (Table 2). Furthermore, the number of internal shocks and the total energy of defibrillations were lower in group A than in group P (*P*=.002). Furthermore, a shorter duration of VF was also observed in group A (*P*<.001). Regarding the restoration of cardiac rhythm after ACC release, there were no significant differences in the recovery time of cardiac contraction or the proportion of epicardial pacing between the groups. However, the duration of epicardial pacemaker use in group A was higher than those in group P (*P*=.03). When dealing with VF after the release of the ACC, there were no significant differences in the consumption of antiarrhythmic drugs between the groups (Table 3). At the T2 time point, group A had a higher VIS than group P (*P*=.02). However, no differences in VIS were observed between the groups at the time of T3. During the whole intraoperative period, there were no significant differences in TDA between the groups.

Table . Incidence of ventricular fibrillation (VF) and restoration of cardiac rhythm after aortic cross-clamp release.

Changes of cardiac rhythm	Group A[a] (N=27)	Group P[b] (N=27)	*P* value
VF, n (%)	8 (30)	19 (70)	.003
Defibrillations, n			.002
0	19	8	
1	6	9	
≥2	2	10	
Total energy (J), median (IQR)	0 (0 - 20)	20 (0 - 40)	.002
VF duration (min), median (IQR)	0 (0 - 1)	1.5 (0 - 3.5)	<.001
Time for cardiac contraction recovery (min), mean (SD)	2.63 (1.15)	2.61 (1.60)	.96
Epicardial pacing, n	14	8	.10
Duration of epicardial pacing (min), median (IQR)	10 (0 - 41)	0 (0 - 9)	.03

[a]Group A: amiodarone group.

[b]Group P: placebo control group.

Table . Consumption of antiarrhythmic and vasoactive-inotropic agents.

Intravenous agents	Group A[a] (N=27)	Group P[b] (N=27)	*P* value
Esmolol (mg), median (IQR)	70 (65 - 75)	130 (70 - 150)	.20
Lidocaine (mg), median (IQR)	0 (0 - 0)	0 (0 - 0)	.36
VIS[c] (T2[d]), mean (SD)	15.32 (3.66)	12.83 (4.19)	.02
VIS (T3[e]), mean (SD)	10.61 (3.99)	10.39 (4.10)	.84
TDA[f], mean (SD)	32.20 (8.98)	27.54 (8.64)	.06

[a]Group A: amiodarone group.

[b]Group P: placebo control group.

[c]VIS: vasoactive-inotropic score.

[d]T2: fifteen minutes after weaning from cardiopulmonary bypass.

[e]T3: fifteen minutes before the end of surgery;

[f]TDA: total dose of vasoactive and inotropic agents.

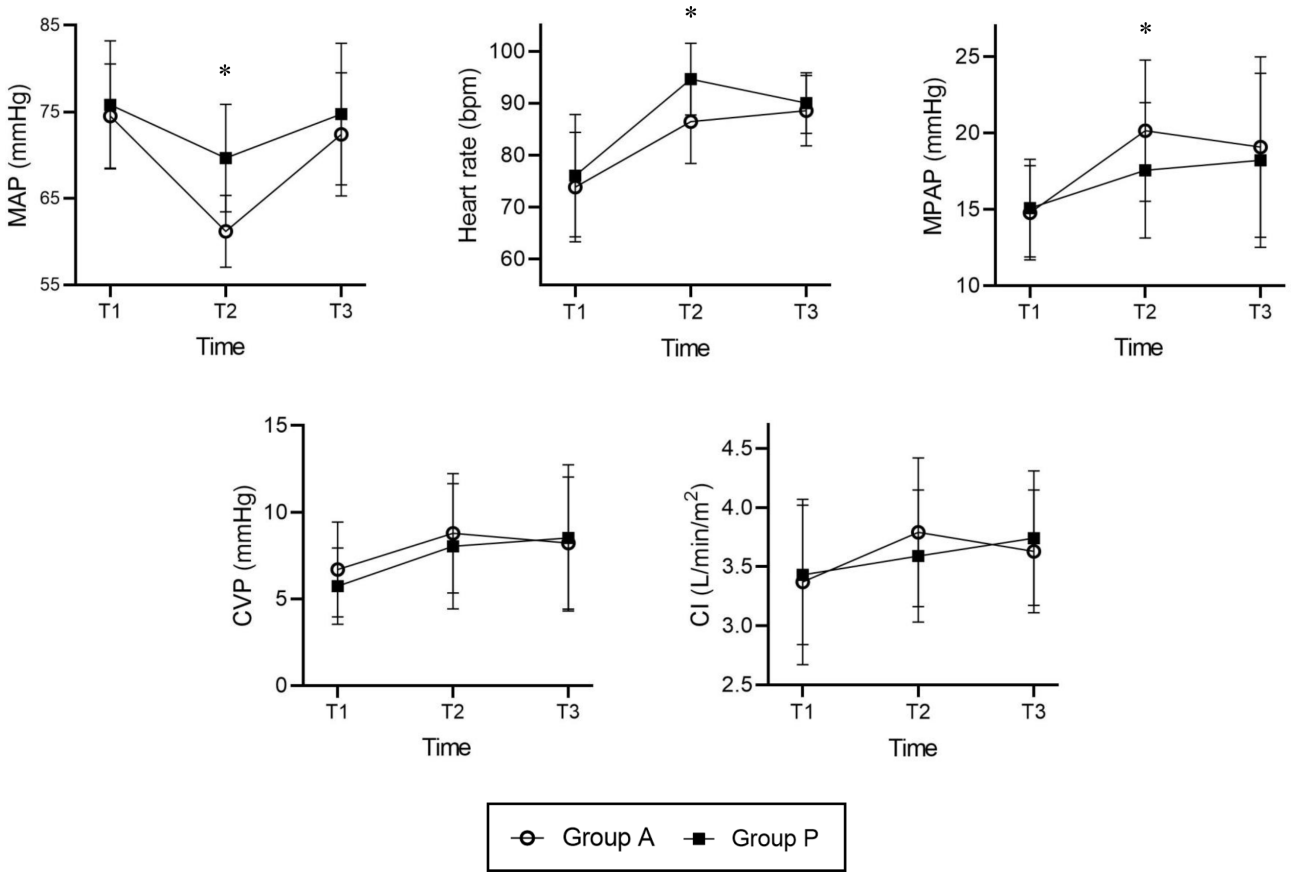
Hemodynamic Status During the Intraoperative Period

The hemodynamic status at various intraoperative times is shown in Figure 2. During the intraoperative period, repeated measures ANOVA indicated significant differences in MAP between the 2 groups (*F1*=5.161, *P*=.03). The global MAP was lower in group A than in group P (Figure 2). Compared with group P, MAP in group A was significantly lower at the time point T2 (mean 61.19 SD 4.16 vs mean 69.67 SD 6.23; *P*<.001). However, no significant differences were found in MAP at the time points T1 and T3 between the groups. During surgery, repeated measures ANOVA did not show significant differences in HR, MPAP, CVP, and CI between the groups (*F1,1,1,1*=3.098, 0.747, 0.271, and 0.001; and *P*=.08, .39, .61, and .97, respectively) (Figure 2). Compared with group P, HR

XSL•FO
RenderX

(mean 86.44 SD 8.07 vs mean 94.63 SD 6.90; *P*<.001) was lower and MPAP (mean 20.15 SD 4.63 vs mean 17.56 SD 4.44; *P*=.04) was higher in group A at the time point T2. However, no significant differences were found in CVP and CI at the T2 time point. Furthermore, there were no significant differences in HR, MPAP, CVP, and CI at the time points T1 and T3 between the groups.

Figure 2. Changes in hemodynamic status during the intraoperative period. The global mean arterial pressure curve was lower in group A than in group P by repeated measures ANOVA. There were no significant differences in the global curves of heart rate, mean pulmonary arterial pressure, central venous pressure, and cardiac index between the 2 groups by repeated measures ANOVA. At T2 point, mean arterial pressure and heart rate were significantly higher in Group P while mean pulmonary arterial pressure was significantly higher in Group A. Group A: amiodarone group; Group P: placebo control group. CI: cardiac index; CVP: central venous pressure; MAP: mean arterial pressure; MPAP: pulmonary arterial pressure; T1: fifteen minutes after anesthetic induction; T2: fifteen minutes after weaning from cardiopulmonary bypass; T3: fifteen minutes before the end of surgery. **P*<.05 was found between the two groups.



Discussion

Principal Findings

In this randomized trial, we primarily found that prophylactic intravenous use of amiodarone effectively reduced the incidence of VF in patients with LVH after ACC removal during open-heart surgery, while also decreasing the duration of VF, the frequency of defibrillation, and the energy required for defibrillation.

VF is a malignant arrhythmia that occurs after the release of the ACC in patients undergoing open-heart surgery. There are various factors that can contribute to VF after the removal of the ACC, which include surgical trauma, inadequate myocardial protection, insufficient intracardiac deairing or left ventricular venting, abnormal electrolyte and acid-base metabolism, delayed rewarming prior to ACC release, and myocardial reperfusion injury after ACC release. Among these factors, myocardial reperfusion injury is often considered the primary cause [16]. The prolonged duration of VF following ACC release can have several adverse effects on cardiomyocytes. Research findings indicated that VF led to alterations in the distribution of coronary artery blood supply, consequently causing subendocardial myocardial ischemia and subsequent cardiac dysfunction [17]. As the duration of VF extends, the resistance of subendocardial vessels progressively rises, exacerbating the subendocardial ischemia in cardiomyocytes. Furthermore, myocardial edema was observed to develop with the prolonged duration of VF [17]. During VF, there is an elevation in ventricular wall tension and venous pressure, necessitating a higher perfusion pressure to restore normal blood supply during CPB. The erratic and frequent myocardial contractions, along with elevated pressure in the cardiac cavity during VF, may lead to increased oxygen consumption by the myocardium. The described factors indeed result in an imbalance between oxygen supply and demand in the myocardium, leading to severe impairment of cardiomyocyte function.

In individuals with LVH, the cardiomyocytes enlarge in volume, yet the associated coronary arteries fail to grow in tandem, leading to a lack of subendocardial capillaries. Elevated end-diastolic pressure and enhanced systolic cardiac function

XSL•FO
RenderX

contribute to an increase in myocardial oxygen consumption. These histological alterations result in an imbalance between myocardial oxygen supply and demand, causing subendocardial myocardial ischemia and reduced myocardial tolerance to ischemia-reperfusion injury [18]. Patients with LVH have been reported to be more likely to suffer from VF after removal of the ACC [5,6,13]. Therefore, this situation has created a need for prophylactic treatment.

Amiodarone, as a broad-spectrum antiarrhythmic agent, has several electrophysiological mechanisms to treat VF. The main antiarrhythmic effect of amiodarone occurs through the inhibition of IKr and IKs channels, leading to a homogeneous prolongation of myocardial repolarization [7]. In addition, it can reduce HR and slow atrioventricular nodal conduction by blocking calcium channels and β-receptors [19]. Furthermore, amiodarone extends refractoriness and decreases intracardiac conduction velocity by inhibiting the sodium channel [20]. During the period of ischemia-reperfusion, amiodarone has the ability to enhance the metabolic efficiency of cardiomyocytes and decrease the dispersion of repolarization across the cell membrane, which is a significant contributor to the development of ventricular arrhythmias [21].

Several clinical trials have investigated whether amiodarone could lower the occurrence of VF after ACC removal. Samantaray et al [11] reported that, compared with the placebo group, administering 150 mg of amiodarone 3 minutes before aortic opening during CABG surgery reduced the occurrence of VF after ACC release. In addition, Yilmaz et al [12] revealed that both 300 mg of amiodarone and 1.5 mg/kg of lidocaine reduced the occurrence of VF after ACC release compared with the placebo group. However, there were no significant differences in the reduction of VF between the amiodarone and lidocaine groups. On the contrary, in the clinical trial by Ayoub et al [22], it was found that the administration of amiodarone 2 minutes before aortic opening did not result in a reduction in the occurrence of VF after the release of the ACC during CABG surgery. On the other hand, lidocaine was effective in reducing the occurrence of VF in this study [22]. In addition, Mauermann et al [23] reported that both 300 mg of amiodarone and 1.5 mg/kg of lidocaine did not lead to a reduction in the occurrence of VF. Despite some conflicting results in the studies mentioned, there is a general indication that amiodarone has a therapeutic effect on VF after ACC removal.

In a documented case, a 66-year-old female patient underwent aortic valve replacement for severe aortic stenosis. The preoperative TTE screening indicated a thickness of up to 18 mm in the ventricular septum and left ventricular posterior wall. Following the release of the ACC and administration of lidocaine, refractory VF ensued, with multiple cardiac shocks showing no effectiveness. Finally, a dose of 150 mg of amiodarone was delivered through the aortic root, ultimately leading to the termination of VF with a single defibrillation [24]. Suzuki et al [5] also reported a similar case of using amiodarone to terminate rVF during cardiac surgery. In a clinical trial by Mita et al [13], patients with LVH undergoing aortic valve replacement were randomly divided into 2 groups. One group received a 150 mg bolus of amiodarone followed by a maintenance dose of 30 mg/h, while the other group received a 1 mg/kg bolus of lidocaine followed by a maintenance dose of 1 mg/kg/h after anesthetic induction. The occurrence of VF was markedly reduced in the amiodarone group compared with the lidocaine group (7/34, 20.6% vs 17/34, 50%; *P*=.02) [9]. Meanwhile, amiodarone has the potential to decrease the levels of interleukin-6 and tumor necrosis factor following the surgery.

In this study, our results showed that prophylactic intravenous amiodarone could lower the occurrence of VF in patients with LVH after the release of the ACC during open-heart surgery. Furthermore, among patients with VF, amiodarone reduced the number and total energy of internal cardiac shocks. In addition, the duration of VF could also be reduced. These results further proved the conclusions of the abovementioned studies. With the use of amiodarone, the recovery of cardiac contractions was not affected and there was no increase in the use of epicardial pacemakers. However, the duration of epicardial pacemakers was longer among patients with amiodarone, which may be correlated with the effect of amiodarone on slowing HR. In terms of hemodynamic effects, after weaning from CPB, HR and MAP decreased and MPAP increased, while the use of vasoactive and inotropic drugs was also higher in patients treated with amiodarone, which could be related to the pharmacological effect of amiodarone. However, CI and CVP were not affected and the hemodynamic status at the end of surgery was not significantly different between the groups. Regarding the consumption of vasoactive and inotropic drugs, although the dose of the drugs was higher among patients with amiodarone after weaning from CPB, the dose and total doses of vasoactive and inotropic agents at the end of surgery were not significantly different among the 2 groups.

Limitations

Our study presents several limitations. First, this was a single-center randomized controlled clinical trial with a relatively small sample size, which limited the power of the statistical analysis. Second, lidocaine was not compared as a control group in the study, because it may be more effective than a placebo but less effective than amiodarone to reduce the incidence of VF according to previous studies [12,13], which could increase the sample size and the cost of the study. Third, this study only included patients with ASA I-III. Therefore, these findings cannot be generalized to patients with ASA IV, especially those with severe left ventricular dysfunction. Therefore, subsequent studies should further address the abovementioned issues.

Conclusions

Preventive intravenous administration of amiodarone can reduce the incidence of VF in patients with LVH after ACC removal during open-heart surgery and decrease the duration of VF, defibrillation frequency, and defibrillation energy. It is safe and has little impact on intraoperative hemodynamic management. The findings may provide some insight for anesthesiologists in dealing with similar problems in clinical practice.

XSL•FO
RenderX

Acknowledgments

This research was supported by the Clinical Capacity Improvement Project of The First Affiliated Hospital of Nanjing Medical University in 2020 (NO. JSPH-MC-2020-9).

Data Availability

The datasets generated and analyzed during this study are not publicly available due to ethical considerations but are available from the corresponding author on reasonable request.

Authors' Contributions

YF and Z-FZ designed the research study. Z-FZ, XKZ, C-JG, and YF performed the research. C-JG and YF analyzed the data. Z-FZ and X-KZ wrote the manuscript. All authors participated in the writing (review and editing) of the manuscript and read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Checklist 1
SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) checklist.
[PDF File, 140 KB - formative_v9i1e64586_app1.pdf]

Checklist 2
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File, 1226 KB - formative_v9i1e64586_app2.pdf]

References

1. He LM, Zhang A, Xiong B. Effectiveness of amiodarone in preventing the occurrence of reperfusion ventricular fibrillation after the release of aortic cross-clamp in open-heart surgery patients: a meta-analysis. Front Cardiovasc Med 2022;9:821938. [doi: 10.3389/fcvm.2022.821938] [Medline: 35187132]
2. Stommel AM, Högler S, Mueller M, et al. A ventricular fibrillation cardiac arrest model with extracorporeal cardiopulmonary resuscitation in rats: 8 minutes arrest time leads to increased myocardial damage but does not increase neuronal damage compared to 6 minutes. Front Vet Sci 2023;10:1276588. [doi: 10.3389/fvets.2023.1276588] [Medline: 38026669]
3. Yamaguchi H, Weil M, Tang W, Kamohara T, Jin X, Bisera J. Myocardial dysfunction after electrical defibrillation. Resuscitation 2002 Sep;54(3):289-296. [doi: 10.1016/s0300-9572(02)00149-1] [Medline: 12204463]
4. Udzik J, Sienkiewicz S, Biskupski A, Szylińska A, Kowalska Z, Biskupski P. Cardiac complications following cardiac surgery procedures. J Clin Med 2020 Oct 18;9(10):3347. [doi: 10.3390/jcm9103347] [Medline: 33081028]
5. Suzuki S, Iwasaki T, Morimatsu H, et al. Successful use of intravenous amiodarone for refractory ventricular fibrillation just after releasing aortic cross-clamp. Masui 2010 Oct;59(10):1266-1270. [Medline: 20960899]
6. Kodaka M, Mori T, Ichikawa J, Ando K, Komori M. Refractory ventricular arrhythmias during aortic valve replacement and cardiac artery bypass requiring 16 attempts of electrical cardioversion: a case report. JA Clin Rep 2020 Aug 11;6(1):60. [doi: 10.1186/s40981-020-00369-w] [Medline: 32783130]
7. Mujović N, Dobrev D, Marinković M, Russo V, Potpara TS. The role of amiodarone in contemporary management of complex cardiac arrhythmias. Pharmacol Res 2020 Jan;151:104521. [doi: 10.1016/j.phrs.2019.104521] [Medline: 31756386]
8. orodoc V, Indrei L, Dobroghii C, et al. Amiodarone therapy: updated practical insights. J Clin Med 2024 Oct 12;13(20):6094. [doi: 10.3390/jcm13206094] [Medline: 39458044]
9. Sattler SM, Lubberding AF, Skibsbye L, et al. Amiodarone treatment in the early phase of acute myocardial infarction protects against ventricular fibrillation in a porcine model. J Cardiovasc Transl Res 2019 Aug;12(4):321-330. [doi: 10.1007/s12265-018-9861-6] [Medline: 30617762]
10. Khataminia M, Najmeddin F, Najafi A, et al. Effect of heart rate control with amiodarone infusion on hemodynamic and clinical outcomes in septic shock patients with tachycardia: a prospective, single-arm clinical study. J Pharm Health Care Sci 2021 Oct 11;7(1):37. [doi: 10.1186/s40780-021-00219-6] [Medline: 34629112]
11. Samantaray A, Chandra A, Panigrahi S. Amiodarone for the prevention of reperfusion ventricular fibrillation. J Cardiothorac Vasc Anesth 2010 Apr;24(2):239-243. [doi: 10.1053/j.jvca.2009.07.007] [Medline: 19800815]
12. Yilmaz M, Aydin U, Arslan ZI, et al. The effect of lidocaine and amiodarone on prevention of ventricular fibrillation in patients undergoing coronary artery bypass grafting. Heart Surg Forum 2014 Oct 1;17(5):E245-E249. [doi: 10.1532/HSF98.2014402] [Medline: 25367235]

XSL•FO
RenderX

13. Mita N, Kagaya S, Miyoshi S, Kuroda M. Prophylactic effect of amiodarone infusion on reperfusion ventricular fibrillation after release of aortic cross-clamp in patients with left ventricular hypertrophy undergoing aortic valve replacement: a randomized controlled trial. J Cardiothorac Vasc Anesth 2019 May;33(5):1205-1213. [doi: 10.1053/j.jvca.2018.10.005] [Medline: 30416026]
14. Zhou X, Zhang Z, Gong C, Fang Y. The effect of prophylactic intravenous amiodarone administration on reperfusion ventricular fibrillation in patients with left ventricular hypertrophy undergoing cardiopulmonary bypass surgery: protocol for a randomized double-blind clinical trial. JMIR Res Protoc 2023 Jan 27;12:e40115. [doi: 10.2196/40115] [Medline: 36705966]
15. Lang RM, Badano LP, Mor-Avi V, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr 2015 Jan;28(1):1-39. [doi: 10.1016/j.echo.2014.10.003] [Medline: 25559473]
16. Zheng Y, Gu Q, Chen HW, et al. Efficacy of amiodarone and lidocaine for preventing ventricular fibrillation after aortic cross-clamp release in open heart surgery: a meta-analysis of randomized controlled trials. J Zhejiang Univ Sci B 2017;18(12):1113-1122. [doi: 10.1631/jzus.B1700229] [Medline: 29204991]
17. Hottenrott C, Maloney JV Jr, Buckberg G. Studies of the effects of ventricular fibrillation on the adequacy of regional myocardial flow. 3. Mechanisms of ischemia. J Thorac Cardiovasc Surg 1974 Oct;68(4):634-645. [Medline: 4278292]
18. Zhu YC, Zhu YZ, Spitznagel H, Gohlke P, Unger T. Substrate metabolism, hormone interaction, and angiotensin-converting enzyme inhibitors in left ventricular hypertrophy. Diabetes 1996 Jan;45 Suppl 1:S59-S65. [doi: 10.2337/diab.45.1.s59] [Medline: 8529802]
19. Siddoway LA. Amiodarone: guidelines for use and monitoring. Am Fam Physician 2003 Dec 1;68(11):2189-2196. [Medline: 14677664]
20. Pannone L, D'Angelo G, Gulletta S, et al. Amiodarone in ventricular arrhythmias: still a valuable resource? Rev Cardiovasc Med 2021 Dec 22;22(4):1383-1392. [doi: 10.31083/j.rcm2204143] [Medline: 34957778]
21. Moreau D, Clauw F, Martine L, Grynberg A, Rochette L, Demaison L. Effects of amiodarone on cardiac function and mitochondrial oxidative phosphorylation during ischemia and reperfusion. Mol Cell Biochem 1999 Apr;194(1-2):291-300. [doi: 10.1023/a:1006935323491] [Medline: 10391152]
22. Ayoub CM, Sfeir PM, Bou-Khalil P, et al. Prophylactic amiodarone versus lidocaine for prevention of reperfusion ventricular fibrillation after release of aortic cross-clamp. Eur J Anaesthesiol 2009 Dec;26(12):1056-1060. [doi: 10.1097/EJA.0b013e32832f0dfb] [Medline: 19809326]
23. Mauermann WJ, Pulido JN, Barbara DW, et al. Amiodarone versus lidocaine and placebo for the prevention of ventricular fibrillation after aortic crossclamping: a randomized, double-blind, placebo-controlled trial. J Thorac Cardiovasc Surg 2012 Nov;144(5):1229-1234. [doi: 10.1016/j.jtcvs.2012.06.039] [Medline: 22770549]
24. Morita Y, Mizuno J, Yoshimura T, Morita S. Efficacy of amiodarone on refractory ventricular fibrillation resistant to lidocaine and cardioversion during weaning from cardiopulmonary bypass in aortic valve replacement for severe aortic stenosis with left ventricular hypertrophy. J Anesth 2010 Oct;24(5):761-764. [doi: 10.1007/s00540-010-0991-5] [Medline: 20665054]

Abbreviations

ACC: aortic cross-clamp
ASA: American Society of Anesthesiology
CABG: coronary artery bypass grafting
CI: cardiac index
CONSORT: Consolidated Standards of Reporting Trials
CPB: cardiopulmonary bypass
CVP: central venous pressure
HR: heart rate
LVH: left ventricular hypertrophy
MAP: mean arterial pressure
MPAP: mean pulmonary arterial pressure
$P_{ET}CO_2$: end-tidal CO_2
rVF: reperfusion ventricular fibrillation
TDA: total dose of vasoactive and inotropic agents
TTE: transthoracic echocardiography
VF: ventricular fibrillation
VIS: vasoactive-inotropic score

XSL•FO
RenderX

Edited by A Mavragani; submitted 21.07.24; peer-reviewed by R Agarwal, Z Yu; revised version received 21.11.24; accepted 26.11.24; published 13.01.25.

Please cite as:
Gong CJ, Zhou XK, Zhang ZF, Fang Y
Impact of Preventive Intravenous Amiodarone on Reperfusion Ventricular Fibrillation in Patients With Left Ventricular Hypertrophy Undergoing Open-Heart Surgery: Randomized Controlled Clinical Trial
JMIR Form Res 2025;9:e64586
URL: https://formative.jmir.org/2025/1/e64586
doi:10.2196/64586

© Chan-Juan Gong, Xiao-Kai Zhou, Zhen-Feng Zhang, Yin Fang. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Identification of Behavioral, Clinical, and Psychological Antecedents of Acute Stimulant Poisoning: Development and Implementation of a Mixed Methods Psychological Autopsy Study

Marley Antolin Muñiz[1*], BS; Vanessa M McMahan[1*], MS, PhD; Xochitl Luna Marti[1], MPH; Sarah Brennan[1], BA; Sophia Tavasieff[1]; Luke N Rodda[2,3], PhD; James Knoll[4], MD; Phillip O Coffin[1,3], MIA, MD

[1]Center on Substance Use and Health, San Francisco Department of Public Health, San Francisco, CA, United States

[2]Forensic Laboratory Division, Office of the Chief Medical Examiner, San Francisco, CA, United States

[3]Department of Medicine, University of California, San Francisco, San Francisco, CA, United States

[4]Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY, United States

[*]these authors contributed equally

Corresponding Author:
Phillip O Coffin, MIA, MD
Center on Substance Use and Health, San Francisco Department of Public Health, San Francisco, CA, United States

Abstract

Background: Despite increasing fatal stimulant poisoning in the United States, little is understood about the mechanism of death. The psychological autopsy (PA) has long been used to distinguish the manner of death in equivocal cases, including opioid overdose, but has not been used to explicitly explore stimulant mortality.

Objective: We aimed to develop and implement a large PA study to identify antecedents of fatal stimulant poisoning, seeking to maximize data gathering and ethical interactions during the collateral interviews.

Methods: We ascertained death records from the California Electronic Death Reporting System (CA-EDRS) and the San Francisco Office of the County Medical Examiner (OCME) from June 2022 through December 2023. We selected deaths determined to be due to acute poisoning from cocaine or methamphetamine, which occurred 3 - 12 months prior and were not attributed to suicide or homicide. We identified 31 stimulant-fentanyl and 70 stimulant-no-opioid decedents. We sought 2 informants for each decedent, who were able to describe the decedent across their life course. Informants were at least 18 years of age, communicated with the decedent within the year before death, and were aware that the decedent had been using substances during that year. Upon completion of at least one informant interview conducted by staff with bachelor's or master's degrees, we collected OCME, medical record, and substance use disorder treatment data for the decedent. Planned analyses include least absolute shrinkage and selection operator regressions of quantitative data and thematic analyses of qualitative data.

Results: We identified and interviewed at least one informant (N=141) for each decedent (N=101). Based on feedback during recruitment, we adapted language to improve rapport, including changing the term "accidental death" to "premature death," offering condolences, and providing content warnings. As expected, family members were able to provide more data about the decedent's childhood and adolescence, and nonfamily informants provided more data regarding events proximal to death. We found that the interviews were stressful for both the interviewee and interviewer, especially when participants thought the study was intrusive or experienced significant grief during the interviews.

Conclusions: In developing and implementing PA research on fatal stimulant poisoning, we noted the importance of recruitment language regarding cause of death and condolences with collateral informants. Compassion and respect were critical to facilitate the interview process and maintain an ethical framework. We discuss several barriers to success and lessons learned while conducting PA interviews, as well as recommendations for future PA studies.

(JMIR Form Res 2025;9:e64873) doi:10.2196/64873

KEYWORDS

psychological autopsy; acute stimulant poisoning; overdose; cocaine; methamphetamine; fentanyl

Introduction

Deaths attributed to acute stimulant poisoning have been increasing in the United States [1]. From 2015 to 2022, poisoning deaths in the United States attributed to cocaine or methamphetamine increased 4- and 6-fold, respectively [2]. This increase in mortality, largely driven by deaths also involving fentanyl, a potent synthetic opioid, has been referred

XSL•FO
RenderX

to as the "fourth wave" of the overdose crisis [3-6]. While the increase in deaths involving stimulants and fentanyl may be primarily due to the toxic effects of fentanyl, the presence of stimulants likely reflects rising polysubstance and, in some cases, unintentional fentanyl use [4,5,7].

Fatal opioid overdose has a well-defined mechanism—respiratory depression eventually leading to cardiac arrest and death—and decades of research exploring the nature of and antecedents to death. However, this framework and associated research are absent for fatal stimulant poisoning, leading investigators and public health providers to assume stimulant poisoning deaths are similar to opioid overdose deaths and a result of an "overdose." This has led to efforts to identify prevention and treatment strategies, like bystander interventions, reversal agents, and strategies to prevent an acute overdose, akin to strategies used for opioid overdose. However, deaths attributed to stimulant poisoning, which are attributed to cardiac or cerebrovascular events far more often than for opioid overdose deaths, may actually be due to chronic disease rather than a single event of substance use [8,9]. Thus, prevention efforts focused on bystander interventions may not be effective to prevent stimulant poisoning deaths, but instead longer-term interventions, such as screening, prevention, and treatment of chronic diseases, may be appropriate [10].

In the Leveraging Psychological Autopsies to Accelerate Research into Stimulant Overdose Mortality study, we aimed to identify the clinical, behavioral, and psychological antecedents of fatal stimulant poisoning. We selected the psychological autopsy (PA) [11,12], which has been used for suicide [12-15] and opioid overdose research [16,17], because this exploratory approach identified concepts decades ago that still drive opioid overdose prevention today [18]. We hypothesized that fatal stimulant poisoning likely has individual, social, and environmental contributors, and thus employed a socioecological conceptual framework [19]. To identify multilevel antecedents of fatal stimulant poisoning, we conducted PAs involving detailed interviews with close contacts of decedents (collateral informants), analyzed alongside medical examiner (case narrative, autopsy, and toxicology reports) and medical record data [20]. We describe the adaptation of PAs to study fatal stimulant poisoning and lessons learned.

Methods

Ethical Considerations

We obtained approval for this study from the University of California San Francisco, Human Research Protection Program (#21 - 35305). To use informant information in death records from the California Electronic Death Reporting System (CA-EDRS), we also received approval from the State of California Health and Human Services Agency Committee for the Protection of Human Subjects (#2022 - 091).

Decedent Case Selection

We reviewed records from the CA-EDRS and the Office of the County Medical Examiner (OCME) of San Francisco for eligible cases from June 2022 through December 2023. We selected closed cases of deaths determined to be due to acute poisoning from cocaine or methamphetamine. Suicides and homicides were excluded because the mechanism and antecedents to death would likely differ from accidental, natural, or undetermined deaths [21]. We sought 30 stimulant-fentanyl decedents (15 involving cocaine and 15 involving methamphetamine) and 70 stimulant-no-opioid decedents (35 involving cocaine and 35 involving methamphetamine). We suspected that stimulant-fentanyl deaths would be similar to opioid deaths, which are well understood. However, little is known about stimulant deaths not involving opioids, thus we targeted a larger sample of that group [9]. Decedents without a known name or informant were excluded. Interviews were conducted 3 - 12 months after decedent death to minimize both acute bereavement and forgetfulness among informants [22,23]. Decedents were considered "enrolled" in the study when at least one informant interview was completed, at which time OCME, medical record, and substance use treatment data were collected.

Informant Selection

We sought 2 informants for each decedent, who ideally could describe the decedent across the life course (eg, childhood experiences and behaviors proximal to death). Informants had to be at least 18 years of age, have communicated with the decedent within the year before death, be aware that the decedent was using substances during that year, and be able to provide informed consent in English. We identified potential informants using CA-EDRS and OCME data, medical record review, flyers, and word of mouth. We conducted outreach at harm reduction programs and public housing settings where decedents had resided. If a mailing address was available for a potential informant, we mailed a recruitment letter that included condolences, the purpose of the study, details of participation, and the option to opt out of additional contact from the study team (Multimedia Appendix 1). We then waited 1 week before reaching out by phone if a phone number was available. If only a phone number was available, we reached out immediately with up to 3 calls, typically separated by at least 1 week.

Interview Procedures

Bachelor's and master's-level staff conducted the interviews after receiving 2 hours of training by a psychiatrist experienced in PAs. All informants provided informed consent through review of an information sheet. Semistructured, mixed methods interviews lasting 1 - 2 hours were then conducted over telephone, Zoom, or in person. The interview included 14 thematic sections involving sociodemographic characteristics, neighborhood and housing, adverse childhood experiences, emotional and psychological health, substance use, and social environment (Multimedia Appendix 1). Each section began with quantitative questions, including validated scales, followed by qualitative questions to add nuance and to contextualize quantitative data. Quantitative interview responses were entered into University of California San Francisco's REDCap (Research Electronic Data Capture) [24,25] in real time by the interviewer. Interviews were audio-recorded and transcribed verbatim via the Rev platform. Informants were compensated US $50 and offered an additional US $10 for any referral that resulted in a completed informant interview. After the interview, informants were offered bereavement resources and additional information

XSL•FO
RenderX

about the cause of death as appropriate. Because some informants had superior knowledge about different aspects of a decedent's life, the interviewer scored the reliability of the responses to each of the survey sections as highly unreliable, somewhat unreliable, neutral, somewhat reliable, or highly reliable.

Additional Data Sources

Upon completion of at least one informant interview, we collected additional OCME, medical record, and substance use disorder treatment data. We waited until the informant interview was completed to ensure we only collected these data for enrolled decedents. The data abstraction form was developed based on prior research the team had conducted [26], the Centers for Disease Control and Prevention (CDC) State Unintentional Drug Overdose Reporting System (SUDORS) [27], and a preliminary review of 15 decedents. OCME data included the case narrative, toxicology, and autopsy reports. Medical record data included regularity of medical care (eg, established primary care, urgent care only, emergency or hospital care only); recent emergency or inpatient care; latest electrocardiogram and echocardiogram results; latest troponin, brain natriuretic peptide, and urine drug screen results; any evidence of opioid, cocaine, or methamphetamine use; all problem list diagnoses; and all active medications at the time of death.

Analytic Plan

We plan to conduct descriptive statistics of quantitative data, with a focus on a priori variables of interest, including cardiac health, social isolation, fentanyl exposure, suicidality, and resilience. We will then conduct least absolute shrinkage and selection operator (LASSO) regression analyses [28] to identify variables related to fatal stimulant poisoning and analyze results by fentanyl involvement and by cocaine compared to methamphetamine as the causal stimulant. We will also conduct targeted, hypothesis-driven regression analyses by fentanyl involvement exploring cardiac health and opioid use history. Qualitative data will be analyzed by study team members using ATLAS.ti 23.2.2 with an a priori codebook refined during the consensus-driven coding process. Qualitative analyses will focus on salient and recurrent themes and describe theme co-occurrence.

Results

We adapted the PA approach to conduct a large study of the antecedents of fatal stimulant poisoning (Figure 1). We identified and interviewed at least one informant for each decedent. During this period, we reviewed 177 potential stimulant-fentanyl cases prior to enrolling 31 decedents (18% of potential decedents; an extra decedent was erroneously enrolled) and 187 potential stimulant-no-opioid cases prior to enrolling 70 decedents (37% of potential decedents). For 19 (5%) cases, there were no informants available. We contacted 725 potential informants for a total of 364 decedents; 410 (57%) did not complete an interview or follow-up with the study team, 112 (15%) were not interested in participating, 62 (9%) were ineligible, and 141 (19%) completed an interview (Table 1).

Recruitment phone calls varied in length, usually lasting 15 minutes, or as long as 40 minutes when a potential informant was in emotional distress. Calls that were longer typically seemed welcome by potential participants, during which they discussed the lives of the decedents and their own feelings of grief. In some calls, potential participants also discussed frustration with the unexpected contact.

Most potential informants were family members (n=409, 56%), including 133 (18%) parents, 80 (11%) children, and 196 (27%) siblings or other family. We also contacted friends (n=79, 11%), spouses or partners (n=43, 6%), and service providers (n=49, 7%). Many decedents had lived in supportive housing settings (eg, single residence occupancy hotels); thus, 77 (11%) potential informants were building staff. Informants who completed interviews had a similar distribution (Table 2).

The majority of stimulant-involved deaths in San Francisco were also attributed to fentanyl [29], thus we rapidly enrolled stimulant-fentanyl decedents. We then shifted our focus to enrolling stimulant-no-opioid decedents, at which point 2 potential informants reached out to us with concerns about the study procedures; both were concerned with the unexpected outreach, the lack of condolences offered in our recruitment materials, and the phrase "accidental death" used to describe the purpose of the study. In dialogue with these potential informants, it became clear that they considered "accidental death" to mean "overdose," and they did not feel that their loved one had died from an "overdose."

We subsequently revised the recruitment letter and our telephone script to begin with content warnings about the sensitivity of the topic we would be discussing, refer to "premature death," instead of "accidental death," offer condolences, and allow for more check-ins. We also made the introduction of our recruitment materials intentionally vague and did not discuss the study's focus of fatal stimulant poisoning until the individual expressed comfort with talking with us, we suggested they speak with us from a private location, and we confirmed their eligibility (Multimedia Appendix 1). These changes were made to promote comfort and engagement among potential informants during recruitment and to limit disclosure of the involvement of stimulants in death to individuals without prior knowledge of the decedent's substance use. After making these changes, we completed all remaining interviews with only one potential informant voicing concern about the purpose of the study. We speculate that the potential informant had reviewed our center's website, which was included on the letterhead of the recruitment letter, providing them with discomforting details about the study.

In addition to the modified recruitment language, we made several other modifications to the study during implementation. Communicating with potential informants proved challenging via telephone calls only, so we modified the study to allow for text messaging with informants. We only corresponded via text message when contact had already been established with the potential informant. After recruitment commenced, it became clear that it would be challenging to interview 2 informants for each decedent, so we modified the study to allow for only one informant if we could not reach a second informant. We also allowed for more than 2 informants in the event one informant

XSL•FO
RenderX

who was interviewed was not knowledgeable about the decedent's life. We modified our inclusion criteria to allow for the enrollment of decedents who died from both methamphetamine and cocaine toxicity, since we did not think we would lose scientific information by their inclusion, and, depending on the number enrolled, we might be able to identify unique characteristics of this group. Finally, during the study we added additional recruitment strategies, including snowball sampling (ie, compensating a participant who referred another participant an additional US $10), contacting participants after the interview who had consented to being contacted for recruitment purposes, and posting flyers in settings where we could reach potential informants (eg, syringe services programs, public housing, etc).

Interviews lasted 1-2 hours. Interviewers coded the reliability of informants for each section of questions in the interview. In general, family members had superior information about childhood experiences and less reliable information about recent substance use. Staff experienced stress in conducting the interviews, particularly when potential informants felt that recruitment efforts were intrusive or when enrolled informants experienced significant grief during interviews. Staff were not specifically trained in psychology or bereavement counseling and at times were overwhelmed by the grief among informants. To alleviate some of the emotional burden on the study team, limits were set around interview capacity, with interviewers performing no more than 1 interview daily and a maximum of 3 interviews per week. Weekly team calls included conversations about recent informant interactions and upcoming capacity for recruitment calls and interviews. Staff could also access biweekly debrief sessions led by a clinical psychologist on staff, addressing topics such as the varying signs of grief and how those signs may show up in outreach and interviews. Interviewers and recruiters were trained in motivational interviewing techniques to provide staff with tools like reflective listening to express empathy during outreach and interviews.

XSL•FO
RenderX

Figure 1. Procedural flow for a mixed methods psychological autopsy study of decedents of fatal stimulant poisoning in San Francisco, CA, from June 2022 to December 2023. CDPH: California Department of Public Health; OCME: Office of the County Medical Examiner.

XSL•FO
RenderX

Table . The number of potential informants for eligible decedents and the recruitment status of potential informants contacted for a mixed methods psychological autopsy study of decedents of acute stimulant poisoning in San Francisco, CA, from June 2022 to December 2023.

	n	%
Eligible decedents (N=364)[a]		
No informants	19	5
At least one potential informant, not enrolled	244	67
At least one potential informant, enrolled	101	28
Potential informants contacted (N=725)		
Ineligible[b]	62	9
Refused participation	112	15
Contacted but no interview completed or did not follow-up[c]	410	57
Completed interview	141	19

[a]Closed cases due to cocaine or methamphetamine acute poisoning with and without fentanyl involvement. Deaths involving other opioids were excluded. Once 31 fentanyl-involved decedents were enrolled, eligibility was restricted to stimulant poisoning without any opioid involvement.

[b]Potential informants were ineligible if they were <18 years old, had not been in contact with the decedent in the year prior to death, did not know the decedent used substances in the year before death, were not able to provide informed consent in English, or the decedent had not died 3 - 12 months before recruitment.

[c]Includes potential informants who were sent a letter or called but never responded, as well as those who were successfully contacted but never followed up to complete an interview. Tracking database did not allow for further disaggregation.

Table . The relationships to decedents of potential informants who were contacted and interviewed for a mixed methods psychological autopsy study of decedents of fatal stimulant poisoning in San Francisco, CA, from June 2022 to December 2023.

Relationship of informant to decedent	Contacted (N=725)		Interviewed (N=141)	
	n	%	n	%
Parent	133	18	18	13
Spouse or partner	43	6	8	6
Child	80	11	17	12
Friend	79	11	30	21
Sibling or other family member	196	27	44	31
Service provider (eg, clinician, case manager)	49	7	11	8
Building staff	77	11	4	3
Other	68	9	9	6

Discussion

We successfully adopted the PA methodology for a large investigation of antecedents to fatal stimulant poisoning. We enrolled at least one informant for 101 decedents. We ascertained OCME and medical record data for all enrolled decedents. We identified challenges including recruiting informants, adapting our recruitment process to address informant grief, and ensuring staff were prepared to manage emotionally challenging interactions.

Enrolling informants was a substantial challenge of the project. We sought 2 informants per decedent; however, a high proportion of our decedents were markedly socially isolated. Nineteen of the decedents we initially identified were either unidentifiable (ie, “John Doe”) or had no contacts listed in their death or medical records. Among those who had any contacts, most had very few and often only family members, limiting our ability to obtain detailed information related to substance use. Few decedents had multiple contacts listed who knew the decedent proximal to death (eg, current friends, partners). We attempted to overcome these challenges by asking informants for referrals of others who knew the decedent and may be interested in participating in an interview, conducting direct outreach to public housing, posting flyers with our contact information, contacting health care providers for recommendations, and using personal connections research staff had with service programs in the community. We also altered the protocol to allow for a single informant when 2 could not be identified; as social isolation was a frequent antecedent to

XSL•FO
RenderX

death, we felt this adjustment would help to reduce bias by including more socially isolated decedents. Nonetheless, a significant proportion of decedents could not be enrolled due to an inability to reach any eligible informants.

Because stimulant-no-opioid cases were far less common than stimulant-fentanyl cases, we devoted substantially more effort to locating informants and ended up enrolling a higher proportion of potential stimulant-no-opioid decedents (37%) compared to potential stimulant-fentanyl decedents (18%). Many potential informants were ineligible due to our requirement that the informant be aware of the decedent's substance use, as well as speak sufficient English to provide informed consent. The dearth of informants for many decedents was compounded by ethical limitations in outreach efforts. Our human subjects committee required a letter be sent to potential informants with a mailing address but allowed direct phone contact for those without a mailing address.

Some informants experienced significant emotional distress during interviews and outreach. Potential informants received "cold calls" from study staff, rather than a warm handoff from a known provider, and did not directly consent to initial contact from our research team. This element of surprise and the sensitive nature of the call left some potential informants feeling the outreach was intrusive. However, for others who spoke at length during the recruitment calls, it seemed like a welcome opportunity to talk about the decedent. While a warm handoff from known clinical providers or the medical examiner's office may be preferred for initial outreach, this may not be feasible for larger PA studies or non-research "death review" projects attempting to track and improve our understanding of drug poisoning deaths.

During the study, we adapted recruitment materials and processes to respond to concerns from potential informants, including more sympathetic language, frequent check-ins, and a content warning. We only noted the role of stimulants in the decedent's death after confirming that the informant knew about the decedent's substance use, and we replaced the phrase "accidental death" with "premature death." These later changes were consistent with preliminary findings from early reviews of transcripts, as informants often did not perceive stimulant poisoning deaths to be "overdose" deaths but instead natural or medical deaths. Most potential informants reacted well to the improved script and recruitment approaches. As in other PA studies [30,31], many informants had a positive experience participating in the study and expressed gratitude to the study team for conducting the research, showing interest in the decedent's life, and providing an opportunity for them to speak about the decedent.

During interviews, we checked in frequently, offered breaks, asked informants how the experience was for them, and if they would like any additional resources. Interviewers also offered bereavement resources consisting of a list of local and national bereavement meetings or websites. The content of recruitment calls and informant interviews was often emotionally draining for staff as well, which was expected given the topics addressed with close contacts of decedents [11]. Additional strategies to address grief among informants may be beneficial, such as postinterview check-ins with informants to offer additional resources [30]. Future PA studies may also benefit from employing study staff with a background in psychiatric work, experience with bereavement, experience with death investigations and next-of-kin notifications, or experience working with populations who use substances [16].

Despite these challenges, we were able to effectively use PAs to conduct a large study of fatal stimulant poisoning using data from multiple independent sources. The marked social isolation of people who use stimulants and the sensitive nature of interviews were significant challenges, some of which were distinct from studies of opioid overdose. The PA requires that informants voluntarily discuss their memories and insights about an extremely distressing event [31]. Thus, future developments in PA research should focus on informant and interviewer well-being. This could include adapting recruitment language to the specific population being studied, including the use of terms such as "premature death" rather than any specific mechanism of death; offering condolences, content warnings, and check-ins as needed; seeking out staff with experience in psychology, bereavement, and substance use; incorporating postinterview check-ins with informants; and using warm handoffs of informants when feasible.

Acknowledgments

We thank the Leveraging Psychological Autopsies to Accelerate Research into Stimulant Overdose Mortality study research team, the San Francisco Office of the County Medical Examiner (OCME) staff, and study participants. This study was supported by Centers for Disease Control and Prevention (CDC) grant R01CE003364.

Data Availability

The datasets generated during and/or analyzed during this study are not publicly available due to the possibility of individual identification but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

No image included. Copies of outreach letter, REDCap (Research Electronic Data Capture) survey, and telephone recruitment script.
[DOCX File, 121 KB - formative_v9i1e64873_app1.docx]

References

1. Stimulant overdose. Centers for Disease Control and Prevention. 2023. URL: https://www.cdc.gov/drugoverdose/deaths/stimulant-overdose.html [accessed 2025-01-03]
2. Drug overdose death rates. National Institute on Drug Abuse. 2023. URL: https://nida.nih.gov/research-topics/trends-statistics/overdose-death-rates [accessed 2025-01-03]
3. Jalal H, Buchanich JM, Roberts MS, Balmert LC, Zhang K, Burke DS. Changing dynamics of the drug overdose epidemic in the United States from 1979 through 2016. Science 2018 Sep 21;361(6408):eaau1184. [doi: 10.1126/science.aau1184] [Medline: 30237320]
4. Friedman J, Shover CL. Charting the fourth wave: geographic, temporal, race/ethnicity and demographic trends in polysubstance fentanyl overdose deaths in the United States, 2010-2021. Addiction 2023 Dec;118(12):2477-2485. [doi: 10.1111/add.16318] [Medline: 37705148]
5. Ciccarone D. The rise of illicit fentanyls, stimulants and the fourth wave of the opioid overdose crisis. Curr Opin Psychiatry 2021 Jul 1;34(4):344-350. [doi: 10.1097/YCO.0000000000000717] [Medline: 33965972]
6. Rawson RA, Erath TG, Clark HW. The fourth wave of the overdose crisis: examining the prominent role of psychomotor stimulants with and without fentanyl. Prev Med 2023 Nov;176:107625. [doi: 10.1016/j.ypmed.2023.107625] [Medline: 37468073]
7. Coffin PO, McMahan VM, Murphy C. Evidence of pre-mortem opioid use among fentanyl overdose decedents in a safety net healthcare system. J Urban Health 2022 Oct;99(5):865-872. [doi: 10.1007/s11524-022-00663-1] [Medline: 35879486]
8. Darke S, Kaye S, Duflou J. Comparative cardiac pathology among deaths due to cocaine toxicity, opioid toxicity and non-drug-related causes. Addiction 2006 Dec;101(12):1771-1777. [doi: 10.1111/j.1360-0443.2006.01601.x] [Medline: 17156176]
9. Turner C, Chandrakumar D, Rowe C, Santos GM, Riley ED, Coffin PO. Cross-sectional cause of death comparisons for stimulant and opioid mortality in San Francisco, 2005-2015. Drug Alcohol Depend 2018 Apr 1;185:305-312. [doi: 10.1016/j.drugalcdep.2017.12.030] [Medline: 29486419]
10. Coffin PO, Suen LW. Methamphetamine toxicities and clinical management. NEJM Evid 2023 Dec;2(12):EVIDra2300160. [doi: 10.1056/EVIDra2300160] [Medline: 38320504]
11. Conner KR, Beautrais AL, Brent DA, Conwell Y, Phillips MR, Schneider B. The next generation of psychological autopsy studies. Suicide Life Threat Behav 2011 Dec;41(6):594-613 [FREE Full text] [doi: 10.1111/j.1943-278X.2011.00057.x]
12. Favril L, Yu R, Uyar A, Sharpe M, Fazel S. Risk factors for suicide in adults: systematic review and meta-analysis of psychological autopsy studies. Evid Based Ment Health 2022 Nov;25(4):148-155. [doi: 10.1136/ebmental-2022-300549] [Medline: 36162975]
13. Conner KR, Pinquart M, Duberstein PR. Meta-analysis of depression and substance use and impairment among intravenous drug users (IDUs). Addiction 2008 Apr;103(4):524-534. [doi: 10.1111/j.1360-0443.2007.02118.x] [Medline: 18261192]
14. Knoll JL 4th. The psychological autopsy, part I: applications and methods. J Psychiatr Pract 2008 Nov;14(6):393-397. [doi: 10.1097/01.pra.0000341894.35877.1b] [Medline: 19057241]
15. Cavanagh JTO, Carson AJ, Sharpe M, Lawrie SM. Psychological autopsy studies of suicide: a systematic review. Psychol Med 2003 Apr;33(3):395-405. [doi: 10.1017/s0033291702006943] [Medline: 12701661]
16. Oliver P, Horspool M, Rowse G, Parry M, Keen J, Mathers N. A psychological autopsy study of non-deliberate fatal opiate-related overdose. CORE. URL: https://core.ac.uk/download/pdf/34714006.pdf [accessed 2025-01-03]
17. Hakansson A, Gerle E. Proposing a psychological autopsy procedure for postmortem examination of accidental fatal overdose cases. J Alcohol Drug Depend 2018;6(2):1000308. [doi: 10.4172/2329-6488.1000308]
18. Manning FJ, Ingraham LH, DeRouin EM, Vaughn MS, Kukura FC, St Michel GR. Drug “overdoses” among U.S. soldiers in Europe, 1978-1979. II. Autopsies following deaths and near-deaths. Int J Addict 1983 Feb;18(2):153-166. [doi: 10.3109/10826088309027349] [Medline: 6862734]
19. Bircher J, Hahn EG. Will the Meikirch model, a new framework for health, induce a paradigm shift in healthcare? Cureus 2017 Mar 6;9(3):e1081. [doi: 10.7759/cureus.1081] [Medline: 28405531]
20. INSERM Collective Expertise Centre. Suicide: psychological autopsy, a research tool for prevention. In: INSERM Collective Expert Reports: Institut national de la santé et de la recherche médicale (INSERM); 2005. [Medline: 21348161]
21. Athey AJ, Beale EE, Overholser JC, Stockmeier CA, Bagge CL. Acute stressors and clinical characteristics differentiate death by suicide, accident, or natural causes among illicit and prescription opiate users. Drug Alcohol Depend 2020 Mar 1;208:107847. [doi: 10.1016/j.drugalcdep.2020.107847] [Medline: 31951908]
22. Hawton K, Appleby L, Platt S, et al. The psychological autopsy approach to studying suicide: a review of methodological issues. J Affect Disord 1998 Sep;50(2-3):269-276. [doi: 10.1016/s0165-0327(98)00033-0] [Medline: 9858086]
23. Appleby L, Cooper J, Amos T, Faragher B. Psychological autopsy study of suicides by people aged under 35. Br J Psychiatry 1999 Aug;175:168-174. [doi: 10.1192/bjp.175.2.168] [Medline: 10627801]

XSL•FO
RenderX

24. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95(103208):103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
25. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
26. Hurstak E, Rowe C, Turner C, et al. Using medical examiner case narratives to improve opioid overdose surveillance. Int J Drug Policy 2018 Apr;54(35-42):35-42. [doi: 10.1016/j.drugpo.2017.12.017] [Medline: 29353022]
27. SUDORS dashboard: fatal drug overdose data. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/overdose-prevention/data-research/facts-stats/sudors-dashboard-fatal-overdose-data.html?CDC_AAref_Val=https://www.cdc.gov/drugoverdose/fatal/dashboard/index.html [accessed 2025-01-03]
28. Tibshirani R. Regression shrinkage and selection via the LASSO. J R Stat Soc Series B Stat Methodol 1996 Jan 1;58(1):267-288. [doi: 10.1111/j.2517-6161.1996.tb02080.x]
29. Anderson N, McMahan V, Coffin PO. Substance use trends in San Francisco through 2022. : Center on Substance Use and Health, Department of Public Health, City and County of San Francisco; 2023 Oct 17 URL: https://www.csuhsf.org/_files/ugd/91710f_ea77b3b62d81455c8143aa2f97b2d5d5.pdf [accessed 2025-01-03]
30. Mérelle S, Van Bergen D, Looijmans M, et al. A multi-method psychological autopsy study on youth suicides in the Netherlands in 2017: feasibility, main outcomes, and recommendations. PLoS One 2020;15(8):e0238031. [doi: 10.1371/journal.pone.0238031] [Medline: 32853213]
31. Johal G, Appleby L, Turnbull P. Is there still a place for psychological autopsy in suicide research? A literature review of methodological limitations and recommendations for future development. Int Rev Psychiatry 2024 Jul 3;36(4-5):494-502. [doi: 10.1080/09540261.2024.2378075]

Abbreviations

CA-EDRS: California Electronic Death Reporting System
CDC: Centers for Disease Control and Prevention
OCME: Office of the County Medical Examiner
PA: psychological autopsy
REDCap: Research Electronic Data Capture
SUDORS: State Unintentional Drug Overdose Reporting System

Edited by A Mavragani; submitted 31.07.24; peer-reviewed by B Ray; revised version received 08.11.24; accepted 13.11.24; published 13.01.25.

Please cite as:
Antolin Muñiz M, McMahan VM, Luna Marti X, Brennan S, Tavasieff S, Rodda LN, Knoll J, Coffin PO
Identification of Behavioral, Clinical, and Psychological Antecedents of Acute Stimulant Poisoning: Development and Implementation of a Mixed Methods Psychological Autopsy Study
JMIR Form Res 2025;9:e64873
URL: https://formative.jmir.org/2025/1/e64873
doi:10.2196/64873

© Marley Antolin Muñiz, Vanessa M McMahan, Xochitl Luna Marti, Sarah Brennan, Sophia Tavasieff, Luke N Rodda, James Knoll, Phillip O Coffin. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Resilience Informatics in Public Health: Qualitative Analysis of Conference Proceedings

Maiya G Block Ngaybe[1], MPH; Lidia Azurdia Sierra[1], MPH; Andrew McNair[2], BA; Myla Gonzalez[1], BSc; Mona Arora[3], MScPH, PhD; Kacey Ernst[4], MPH, PhD; Enrique Noriega-Atala[5], MSc, PhD; M Sriram Iyengar[6], MSc, MS, PhD

[1]Health Promotion Sciences, University of Arizona, Mel and Enid Zuckerman College of Public Health, Tucson, AZ, United States

[2]Department of Linguistics, University of Arizona, College of Social & Behavioral Sciences, Tucson, AZ, United States

[3]Community, Environment & Policy Department, University of Arizona, Mel and Enid Zuckerman College of Public Health, Tucson, AZ, United States

[4]Department Chair & Professor, Epidemiology and Biostatistics, University of Arizona, Mel and Enid Zuckerman College of Public Health, Tucson, AZ, United States

[5]University of Arizona, Tucson, AZ, United States

[6]College of Medicine - Phoenix, University of Arizona, Phoenix, AZ, United States

Corresponding Author:
M Sriram Iyengar, MSc, MS, PhD
College of Medicine - Phoenix
University of Arizona
475 N 5th St
Phoenix, AZ, 85004
United States
Phone: 1 281 793 4733
Email: msiyengar@arizona.edu

Abstract

Background: In recent years, public health has confronted 2 formidable challenges: the devastating COVID-19 pandemic and the enduring threat of climate change. The convergence of these crises underscores the urgent need for resilient solutions. Resilience informatics (RI), an emerging discipline at the intersection of informatics and public health, leverages real-time data integration from health systems, environmental monitoring, and technological tools to develop adaptive responses to multifaceted crises. It offers promising avenues for mitigating and adapting to these challenges by proactively identifying vulnerabilities and fostering adaptive capacity in public health systems. Addressing critical questions regarding target audiences, privacy concerns, and scalability is paramount to fostering resilience in the face of evolving health threats.

Objective: The University of Arizona held a workshop, titled Resilience Informatics in Public Health, in November 2023 to serve as a pivotal forum for advancing these discussions and catalyzing collaborative efforts within the field. This paper aims to present a qualitative thematic analysis of the findings from this workshop.

Methods: A purposive sampling strategy was used to invite 40 experts by email from diverse fields, including public health, medicine, weather services, informatics, environmental science, and resilience, to participate in the workshop. The event featured presentations from key experts, followed by group discussions facilitated by experts. The attendees engaged in collaborative reflection and discussion on predetermined questions. Discussions were systematically recorded by University of Arizona students, and qualitative analysis was conducted. A detailed thematic analysis was performed using an inductive approach, supported by MAXQDA software to manage and organize data. Two independent researchers coded the transcripts; discrepancies in coding were resolved through consensus, ensuring a rigorous synthesis of the findings.

Results: The workshop hosted 27 experts at the University of Arizona, 21 (78%) of whom were from public health–related fields. Of these 27 experts, 8 (30%) were from the field of resilience. In addition, participants from governmental agencies, American Indian groups, weather services, and a mobile health organization attended. Qualitative analysis identified major themes, including the potential of RI tools, threats to resilience (eg, health care access, infrastructure, and climate change), challenges with RI tools (eg, usability, funding, and real-time response), and standards for RI tools (eg, technological, logistical, and sociological). The attendees emphasized the importance of equitable access, community engagement, and iterative development in RI projects.

XSL•FO
RenderX

Conclusions: The RI workshop emphasized the necessity for accessible, user-friendly tools bridging technical knowledge and community needs. The workshop's conclusions provide a road map for future public health resilience, highlighting the need for scalable, culturally sensitive, community-driven interventions. Future directions include focused discussions to yield concrete outputs such as implementation guidelines and tool designs, reshaping public health strategies in the face of emerging threats.

(JMIR Form Res 2025;9:e63217) doi:10.2196/63217

KEYWORDS

resilience; public health; informatics; mobile phone; artificial intelligence; AI

Introduction

Background

Over the past few years, 2 major threats to public health have emerged. The recent COVID-19 pandemic caused an estimated 1 million deaths in the United States [1], and the long-term effects of COVID-19 are still being identified. Meanwhile, climate change is the long-term shift in predicted temperatures and weather behaviors for a given region [2]. While COVID-19 emerged within just a few months, making it an acute stressor, climate change's effects have developed gradually, making it a chronic stressor [3]. In the long term, the public health effects of climate change could be even more serious as global warming, infectious disease, air pollution, and other consequences increase in severity [4-7]. The effects of climate change threaten the physical and mental health of the public via respiratory disease, infection, stress, malnutrition, heat exhaustion, and a variety of other negative consequences [4-7].

In the aftermath of the COVID-19 pandemic, public health is grappling with the growth of misinformation [8,9], persistent health inequities [10-12], a deepening mental health crisis [13,14], rising malnutrition rates [15,16], and a myriad of other critical challenges. There is now great pressure on the field of public health to prevent a future pandemic and control public response. In navigating a post–COVID-19 world, public health is now faced with the challenges of improving communication across disciplines [17,18], preventing the spread of misinformation and disinformation [19,20], responding quickly in the face of another public health emergency [21], addressing mental health issues [13,14], and tackling the nutritional crisis within the United States [15,16]. On a global scale, the health impacts of political tensions and climate change, particularly in relation to refugee displacement, add to these challenges [22,23].

To successfully cope with these inevitable threats to public health, people and communities need to develop resilience, defined as the ability of people, households, communities, countries, and systems to mitigate, adapt to, and recover from shocks and stresses in a manner that reduces chronic vulnerability and facilitates inclusive growth [24]. It is recognized that resilience is a necessary precursor to achieving the United Nations' Sustainable Development Goals, including "No poverty," "Zero hunger," and "Climate action" [25].

What Is Resilience Informatics?

Informatics has been defined as "the science of information, the practice of information processing, and the engineering of information systems" [26], although specific definitions for the field and its subfields, such as health and medical informatics, still continue to be debated [27]. Informatics tools include data science, artificial intelligence (AI), mobile health, and augmented and virtual reality. Resilience informatics (RI) is an emerging discipline concerned with harnessing informatics to materially improve and promote the ability of people, communities, and organizations to effectively cope with natural and man-made stressors [3]; for example, there is evidence in the literature that informatics tools such as medical informatics systems or informatics tools such as telemedicine and enhanced health information systems may be able to strengthen the resilience of health care facilities [28].

While modern informatics technologies can play a major role in helping people and communities develop resilience, much formative research needs to be conducted to understand the contours of this emerging discipline. The following fundamental research and development questions need to be addressed to lay a solid foundation for RI:

- For whom should we develop RI tools?
- What are the important resilience challenges to be tackled using informatics? What are the priorities?
- How can we provide privacy and security, protect sensitive data, and account for social determinants of health?
- How can we safeguard the rights of Indigenous populations and help bridge the digital divide?
- If, as seems necessary, developing resilience entails behavior change in people and communities, how can we develop tailored and personalized tools to help support behavior change?

In addition, RI systems need to be scalable and extensible while not consuming excess energy or environmental resources that could exacerbate climate change. To answer these questions, input is needed from frontline workers across diverse fields, areas of expertise, and levels of implementation in public health and resilience areas.

The Workshop's Purpose

On November 20, 2023, the University of Arizona hosted a workshop, titled Resilience Informatics in Public Health, with experts from a variety of relevant fields, including resilience, public health, climate, informatics, policy, and technology. These experts were primarily based in Arizona, but a few purposefully selected experts from outside of the state with relevant areas of expertise were also invited to participate and present relevant information to help set the groundwork for the workshop discussions. The purpose of the gathering was to shape the future directions and applications of the emerging discipline of RI.

XSL•FO
RenderX

A secondary purpose of the workshop was to provide networking opportunities to build connections across this field and inspire new projects in this up-and-coming area of research and development of RI tools and systems. Attendees had the opportunity to meet and network with other experts in the field whom they may not have had the chance to meet in other professional settings.

Methods

Recruitment

A group of 40 experts were purposefully selected and invited via email to attend the workshop. The workshop participants were from the fields of public health, medicine, weather services, informatics, environmental science, and resilience. The experts were chosen using a convenience and snowball sampling method based on whom the RI workshop leadership had developed relationships with and who seemed to be best suited to represent the desired sectors and enrich the conversations at the workshop. The institutions represented included Weill Cornell Medicine, the University of Colorado, Northern Arizona University, the Arizona Center for Rural Health, the Mel and Enid Zuckerman College of Public Health, the Arizona Institute for Resilience, and the College of Medicine-Phoenix.

Workshop and Data Collection

During the workshop, participants listened to presentations from a few key experts to spark discussion and ideas. In between the talks, 6 groups of 3 to 7 people each were formed by workshop facilitators such that the groups contained as much diversity as possible in terms of the members' backgrounds (field of work and discipline). Subsequently, with the help of the facilitators, the attendees reflected and collaboratively discussed answers to the questions presented in Textbox 1.

During the discussions, participants were asked to write down their ideas independently on sticky notes and compile them on flipcharts to ensure that no one participant dominated the discussion. Data saturation was not discussed because this was a 1-time event, and additional data collection was not possible after the workshop. The discussions around the key questions (Textbox 1) were systematically audio recorded, and notes were taken by students from the University of Arizona who were part of the organizing team. Workshop attendees were made aware of the audio recording and note-taking and were informed that the findings of the workshop were to be compiled into a proceedings report.

Textbox 1. Questions for the Resilience Informatics in Public Health workshop discussion aided by facilitators.

Session 1 (approximately 30 min)

- What does resilience mean to me?
- To what do we need to be resilient?

Session 2 (approximately 1 h)

- What informatics tools are you using?
- What are the barriers to technological solutions? How do we address them?
- What tools do we need?
- How do we ensure equity across these solutions?

Qualitative Analysis

Three University of Arizona students conducted a rapid analysis of the notes taken during the workshop to produce a report for dissemination to the greater community for awareness. Although transcripts were not sent to participants for comment, the report was shared with participants via email to solicit feedback to increase the validity of the analysis as a form of participant checking. Workshop participants were also asked to participate in the in-depth thematic analysis process. A group of 4 students and the principal investigator worked with 1 expert attendee who volunteered to help develop the codebook. The team used a primarily inductive approach in the data analysis with a phenomenological theoretical approach [29]. The analysis team members iteratively constructed their list of themes based on the transcripts and notes using standard identification techniques [30]. The themes were compiled together using a Google Jamboard and grouped to determine the final list of codes and subcodes (Figure 1).

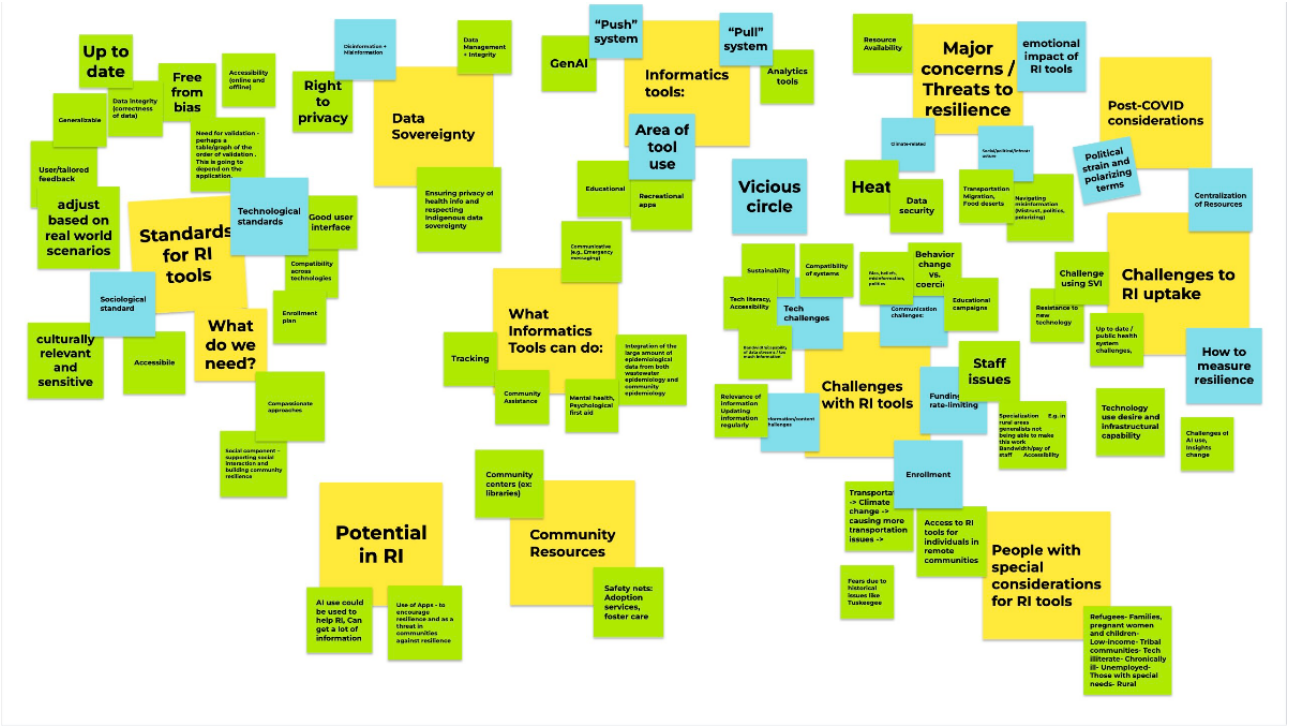
The 4 student researchers then coded the transcripts independently. In the case that the notetakers forgot to record the discussions, notes were used in place of transcripts. The coding team also coded the notes that were taken on sticky notes by the groups during discussions to ensure that no ideas were missed. The team then worked together to finalize the first draft of the codebook with definitions and examples (Figure 1; Multimedia Appendix 1). The codebook was entered into MAXQDA (VERBI Software GmbH), and the team began coding. The first transcript was coded by all team members, and discrepancies were resolved through discussion and consultation with senior members of the team. The remaining transcripts and notes documents were coded independently by 2 coders each. Conflicts were again resolved through discussion. The codebook was adjusted through the coding process because themes may have come up inductively. After coding all documents, the coders reviewed all documents once more to ensure that nothing was missed.

The team then wrote a narrative analysis of the findings, summarizing major points that came up to help answer the

XSL•FO
RenderX

workshop's questions. The analysis write-up process was aided in part by ChatGPT (OpenAI) for certain sections in the synthesis of the summary of the information gathered from the text; these syntheses were subsequently reviewed manually to ensure accuracy, validity, and appropriate language use in the narrative. The research team used the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist to guide the reporting of the workshop findings and qualitative analysis [31].

Figure 1. Brainstorming document for the development process of the codebook for the qualitative analysis of transcripts and notes from the Resilience Informatics in Public Health workshop.



Facilitator Reflexivity

The facilitators of the workshop were 3 female professors and researchers with PhD degrees and experience in the fields of epidemiology, environmental health, climate, and resilience. One of the facilitators is a widely recognized infectious disease epidemiologist with practical experience in public health practice at a health department level and a focus on the environmental determinants of vector-borne disease transmission and control. Another facilitator is an assistant professor at the Mel and Enid Zuckerman College of Public Health with a focus on building public health capacity to address "complex" public health challenges through developing decision-support tools, enhancing science communication and translation, and integrating a health and equity lens into adaptation planning. At the time of the workshop, she was leading the Arizona Department of Health Services–Centers for Disease Control and Prevention COVID Health Disparities Initiative at the Arizona Center for Rural Health aimed at mobilizing partnerships to advance health equity and address social determinants of health related to COVID-19 health disparities among underserved populations considered high risk. Finally, the last facilitator is an experienced research professor with the Lyda Hill Institute for Human Resilience, faculty at the University of Colorado Denver and the Mel and Enid Zuckerman College of Public Health, a guest researcher with the Centers for Disease Control and Prevention, and an affiliate scientist at the National Center for Atmospheric Research in Boulder, Colorado. She focuses on the intersection of weather, climate, and health, with an emphasis on vector-borne diseases.

Workshop participants were primarily invited because they had a relationship with at least 1 of the facilitators in their line of work related to public health and resilience. The participants were familiar with the researchers and their public health interventions, which have benefited the state of Arizona, especially in relation to the COVID-19 pandemic.

Ethical Considerations

The RI workshop team reviewed the criteria for determining human subject research and concluded that the study protocol did not meet the criteria, classifying it as nonhuman subject research. The study team took all appropriate precautions to maintain data confidentiality. Participants were made aware that information collected from the conversations would be compiled in a report and manuscript for publication and dissemination of the workshop insights. Transcripts and quotes were deidentified to further protect participant identities as a precautionary measure. No compensation was offered, although travel, food, and accommodation reimbursements were provided for those who requested assistance to attend the workshop.

Results

Workshop Attendees

Of the 40 experts invited, 27 (68%) participated in the workshop at the Health Sciences Innovations Building at the University

XSL•FO
RenderX

of Arizona. Of these 27 experts, 21 (78%) were from public health–related fields (n=5, 24% with a concentration in epidemiology; n=5, 24% with a concentration in community outreach; and n=3, 14% with a concentration in medicine). Of the 27 experts, 8 (30%) worked in the field of resilience, and 7 (26%) had a concentration in a field of technology such as data science (n=4, 57%). Finally, of the 27 experts, 3 (11%) had extensive experience in emergency management or preparedness, 1 (4%) in environmental science, and 2 (7%) in meteorology.

The organizing team present included the 3 facilitators, 1 principal investigator, and 8 students from relevant fields who helped with logistics, note-taking, and audio recording of conversations.

The largest number of participants were from academia, mostly affiliated with the University of Arizona (19/27, 70%), followed by Arizona State University (3/27, 11%), and specially invited attendees from academic institutions outside of Arizona (University of Colorado: 1/27, 4%; and Weill Cornell Medicine: 1/27, 4%). Of the 27 attendees, 2 (7%) came from the Arizona Department of Health Services, while 3 (11%) were from county-level health departments (Pima: n=1, 33%; Coconino: n=1, 33%; and Maricopa: n=1, 33%); furthermore, 3 (11%) participants came from organizations representing American Indian groups (Salt River Pima-Maricopa Indian Community: n=1, 4%; and Arizona Advisory Council on Indian Health Care: n=2, 7%), 2 (7%) were from the National Oceanic and Atmospheric Administration and National Weather Service, 1 (4%) was the chief executive officer and founder of a mobile health organization (Wehealth), and 1 (4%) was from a statewide community health worker organization (Arizona Community Health Workers Association).

Codes

The codes that were identified during the qualitative analysis included the potential in RI, threats to resilience, people with special considerations, challenges with RI, standards for RI tools (technological, sociological, and logistical), and data sovereignty, among others (Multimedia Appendix 1; Figure 2). The major findings associated with the codes are included in the following subsections.

Figure 2. Frequency of codes in the transcript and notes documents included in qualitative analysis from MAXQDA.

Code System	Grace_Notes for Resili...	Rhett_RI Recording	Maiya_RI Recording	Sage_RI Recording	Andrew_Notes for Re...	Myla_RI Recording	Lidia_RI Recording	TOTAL
Standards for RI Tool	1%		9%	4%	0%		2%	4%
Technological Standard	7%			4%	19%	1%		3%
Logistical Standard	8%	0%	5%	1%	19%			2%
Iterative process	2%			3%	1%			1%
Sociological Standard	13%		1%	24%	38%			13%
Tailoring	12%		7%	7%	2%			5%
What Do We Need?	3%	5%			2%	10%	5%	2%
Communication	12%				1%			0%
long-term solutions	7%			9%	9%			5%
Challenges with RI Tools				1%	0%		15%	2%
Funding-related issues	2%		7%		11%			2%
Accessibility	10%	1%		1%	14%	8%	2%	2%
Communication	3%	1%	11%	4%	3%		5%	4%
Information			11%		1%			2%
Centralization of Resources	5%					12%		2%
Measuring Resilience	2%	3%		4%			43%	6%
Communication between disciplines	3%			9%	0%			5%
Clinical Translation				1%				0%
Mixed methods approach				6%				3%
Human centered design	1%			0%	3%			0%
succession planning				2%				1%
Behavioral economics				3%				1%
Generational differences				11%	2%			6%
Infrastructure					2%			0%
Elasticity / Scale	2%		2%					0%
Real-time response	2%		11%	7%	1%			5%
Data Sovereignty	3%		3%	0%	8%		1%	1%
Disinformation + Misinformation			1%	2%			1%	1%

Potential in RI

RI tools were mentioned in many of the conversations. Specific examples included the 4 tools highlighted during the expert presentations at the event—Pilas, AZCOVIDTXT, AZX, and Wehealth—as well as ≥200 essential tools for public health data monitoring (Textbox 2).

RI tools were also used for educational purposes such as community education, health care worker training and simulation tools. Such tools were also mentioned as having a multitude of different purposes and functions (Textbox 3).

XSL•FO
RenderX

Textbox 2. List of key existing resilience informatics tool examples mentioned by participants in the workshop.

- Organization assessments
- Quality assurance tools
- Heat maps
- Telemapping tools
- Mental health screening
- The 988 public helpline for mental health

Textbox 3. Purposes of key existing resilience informatics tools.

- Creating clinical guidelines with community perspective and provider perspective
- Recording membership sign-ups
- Listing contact information for outreach in emergencies
- Taking raw data and putting it into a usable format
- Digesting complex information, summarizing relevant information, and then pushing it to the relevant audience
- Recording how many people have enrolled in a course
- Evaluating a course or intervention
- Measuring the intensity of reactions from the public
- Surveying the prevalence of infectious diseases or chronic illnesses
- Acting as a decision tool for gauging and assessing data
- Projecting the impact of an intervention

Finally, emerging technologies such as generative AI hold the potential to enhance the impact of RI, as highlighted by a participant:

> *So this is actually a really hot area of research for generative AI. And you have to look at it a little bit differently than what you see on the surface. But what generative AI is, you can feed it a corpus of knowledge. And it builds a giant tree out of that knowledge. And so when you ask a specific question, it'll try to navigate that part of the tree and find those things...I mean, there's reasons to be skeptical because basically the ChatGPT systems that we know have gotten their information from all over the internet. There's some really bad information. But there are other ways where, for example, Maricopa County could put all of your policies, plans, procedures and every...Build your own corpus of knowledge and use the same generative AI algorithms so that you could ask a question like, "How do I answer this question?" And so then it's taking your internal information, sort of trying to like capture your brain and everyone else's brain and all of the documents and you can produce stuff.* [Participant from Arizona State University Knowledge Enterprise]

In summary, RI tools have the potential to span a wide range of functions. While many RI tools are already available, there remain numerous possible areas for further expansion.

Threats to Resilience

Resilience faces challenges across various health sectors, which are often intersecting and complex. Protecting the resilience of communities and ensuring their health and well-being begins with understanding these threats and how individual communities are impacted.

> *Health care access was one of the threats to resilience mentioned at the workshop. The cost of health care alone makes it inaccessible to many groups of people or, at the very least, a burden to individuals who are barely able to afford health care. However, threats to health care access extend beyond cost because the geographic location of health services also plays a critical role. Our participants also mentioned that within health care services, workforce strain can pose a threat to public health. Health practitioners' mental health was also mentioned as having a direct impact on the quality of treatment that patients receive. Overworked health care professionals introduce the potential for practitioner mistakes or lead to a shortage of health care professionals in the community, as noted by a participant:*
>
> *But then you take into account that 30% of the public health workforce left the profession during COVID. And one of the biggest challenges in health departments right now is retention, not just retention of people, retention of historical knowledge and community understanding. I mean, of this idea. And I'm like, "Yes." But it really sucks when you have a*

XSL•FO
RenderX

> *room full of people who don't really know, understand and haven't worked, been on the ground during the pandemic, haven't done these things, don't know the community.* [Participant from Arizona Advisory Council on Indian Health Care]

Inadequate infrastructure was mentioned by our participants as a large threat with respect to the centralization of resources and transportation access, as previously discussed with health care access. However, the threat of infrastructure extends beyond these areas, encompassing the housing crisis, internet access, and utilities. The centralization of resources and transportation services is necessary to protect the health of communities, as discussed before, in terms of the ability to access health services. However, beyond health care, a lack of transportation services and decentralized resources continue to threaten resilience because transportation is crucial for individuals in food deserts to be able to reach grocery stores and for those in areas of job shortages to reach work. Housing accessibility and affordability were also mentioned during the workshop because housing prices have increased in the past few years in Arizona.

In the case of water, communities rely on having access to safe, running water for consumption and bathing. This important resource, especially in the desert environment of Arizona, has not always been adequately protected during droughts and instances of contaminated water supply, as mentioned by a participant:

> *So, what has happened in a lot of the orchards, like the nut orchards, have moved from California to the Willcox area [in Arizona] to pump the water out because they don't have any restrictions. And so basically, they're sucking it dry. So, the concern is, there won't be any water in the future, 50 years from now, in that area.* [Participant in environmental science field at the University of Arizona]

Internet access was mentioned as being of great importance for protecting health as well because much health-related information and many health services are available on the web. Being able to view this health information or find health services through web searches is important for individuals to be able to protect their health.

Large-scale events such as the ongoing climate crisis were mentioned during the workshop as one of the greatest threats to resilience. Climate change has negatively impacted water resources, raised temperatures, and increased the prevalence of infectious diseases. Furthermore, modern infrastructure is not equipped to withstand the challenges that extreme weather events pose, leaving communities more at risk of being devastated by such an event. One participant stated as follows:

> *How, for me, is its emergency response, but it's also how to prevent and mitigate issues with climate change. And we're getting into it with the hottest heat on record in the city of Phoenix. I'm trying to get more involved with getting areas that don't have cooling centers, doesn't have things like that, like the tribal nations...Outlying areas of, even in the Phoenix area, doesn't have cooling centers...So, we see a big phase for central Phoenix and getting in some rural areas, but when you start going outside the Phoenix area or into tribal nations, there's nothing to help them recover or even get supplies. So, to me, that's a big issue. I live in Queen Creek, so even looking at my own town, they have a station. Yeah, it's the fire station. You drop off the water, but you have no cooling stations for the homeless or individuals or even the reservations.* [Participant from the Arizona Coalition for Healthcare Emergency Response]

The climate crisis has become even more difficult to address due to its politicization.

Political strain was mentioned at the workshop as another threat to the resilience of communities in working against the polarization of health concerns. Participants mentioned that during the COVID-19 pandemic, the communication of accurate information proved difficult because the public was inundated with mass amounts of misinformation and disinformation. One participant stated as follows:

> *Politics can modify how we approach public health disasters.* [Participant from a county health department]

Communication emerged as an individual threat to the resilience of communities because approaches to addressing the COVID-19 pandemic were disjointed and limited as a result, according to our workshop participants. There were disparities in information between clinicians and public health practitioners who lacked the means to effectively engage with each other. The inability to properly communicate exacerbated the issue of trust in information and the prevalence of misinformation. Creating adaptable communication systems can pose difficulty in the field of public health, as mentioned by a workshop participant:

> *I think uncertainty too. There's a lot of the things we're talking about happening in the future. I think especially when we talk about extreme heat and things like that, or even other things like how AI is going to come into play. There's just a lot of uncertainty of what's going to happen. I think that really heavily influences the way people respond to information. How do you communicate something that, things that are changing, things that are uncertain, and how do you make decisions in an uncertain environment, the burden that puts on a household person trying to plan financially or somebody just mentally the load, or even a city planner trying to figure out what's the best decision?* [Participant from Arizona State University Knowledge Enterprise]

Issues related to access to *resources* also emerged as a potential threat to resilience for the communities our participants serve. A lack of financial resources limits what the community can afford to support its health, as noted by a participant:

> *And it's for our community organizations so that they can leverage that piece to apply funding, to ensure there's more programs within the services, within the communities they serve. Because we can't do*

XSL•FO
RenderX

> *everything. We don't have the capacity.* [Participant from a county health department]

Other resource-related issues mentioned by participants included food supply and air.

Several *social issues* were mentioned during the workshop that directly work as threats to resilience, with public health professionals facing concerns surrounding food insecurity, the housing crisis, transportation, migration, racism, gentrification, war and politics, cultural sensitivity, the cost of living, employment security, and the education system. One participant stated as follows:

> *Yeah. I think distrust is more just the fact that if we need to be resilient, we need to trust each other and we don't anymore. And so that's a thing we need to be aware of. How do we build trust in communities? And the disinformation is more about are we just getting the right information or not? That's why I meant trust between communities.* [Participant who is a software development professional and chief executive officer of a mobile health organization]

Mental health was a concern raised by several participants; for instance, aging populations were mentioned as being vulnerable to the empty nest syndrome and isolation issues. In addition, substance misuse was directly related to mental health (eg, the growing prevalence of drugs such as fentanyl). Concerns surrounding the lack of empathy and social capacity were discussed in depth in 1 group at the workshop. The lack of empathy in American culture was mentioned as possibly stemming from the colonial mindset focused on individualism. By contrast, social capacity was identified as an issue due to the ever-growing role of mobile phones and social media today, especially among younger generations. Mindful interventions were mentioned as having the potential to address these concerns and help ensure that populations remain socially capable. One participant stated as follows:

> *It's compassion fatigue or something. You get so much information because you're connected to the internet, you just get bombarded every day.* [Participant from a county health department]

People With Special Considerations

A recurrent theme that emerged during the workshop discussions was that certain communities may face greater challenges to resilience than others, as mentioned by a participant:

> *No one would choose not to have running water.* [Participant from a county health department]

It was apparent in many of the discussions that some communities lack basic resources such as consistent access to water. Indigenous communities, asylum seekers, the aging and older adult populations, people with disabilities, non–English-speaking communities, and other groups considered marginalized often face unique challenges to resilience that merit special considerations when considering the use of RI tools.

Asylum seekers and refugees were mentioned as having challenges related to privacy and access to resilience tools because they are vulnerable to exploitation and often face barriers to legal recourse. The aging and older adult populations, people with disabilities, non–English-speaking communities, and other groups considered marginalized were also mentioned as needing special consideration when designing RI tools (eg, considering translation, accessibility, and other tailoring approaches to fit their needs).

Challenges With RI Tools

Some workshop participants mentioned that RI tools face challenges of usability that must be understood and addressed to maximize the positive impact of the tools in improving community resilience while also preventing possible adverse effects.

Accessibility is the first of these challenges. RI tools can be limited in impact based on the size of the audience they can reach and effectively assist. Tools should be tailored to the community being served and designed with end users in mind. Elasticity and scalability were mentioned as issues that should be at the forefront of RI tool design considerations so that the tools may serve the greatest number of communities, which can vary in size, geography, and so on. Accessibility concerns further include issues of resource centralization, language barriers, technology literacy, and related considerations. One participant remarked as follows:

> *So, one thing too, I would say the integration of social determinants of health in city planning. So general plans as well as emergency response plans. One of the things we learned during COVID is a lot of our emergency response plans really weren't built to address supporting diverse populations, multilanguage speakers, all of these other things. And so being able to go in and make those the norm now.* [Participant from a county health department]

Real-time response was mentioned as being especially desirable in RI tools. Between collecting and analyzing data and publishing the results, the process can be all too slow to be able to properly respond to fast-moving issues that may threaten resilience, such as a pandemic. RI tools must be able to quickly respond in times of crisis to be able to limit the amount of people who are affected by the crisis, as highlighted by a participant:

> *Informatics can assist me in finding what that trigger point is, but it can't assist me in actual response. It's too late. Once it's already happening, now I'm in the door. I'm working with a problem and actually implementing things. Informatics takes a backseat; they don't care anymore. But I want informatics to inform that response. That would be very valuable to tell me, "Hey, you're done. You can stop now. You can turn it off." That would be nice. But I don't know how informatics can assist in the response phase or the recovery phase.* [Participant from a county health department]

Funding was mentioned as being a particular challenge for existing public health and resilience-building initiatives. RI tools and projects must be able to secure funding, such as through grants, to establish and sustain the tool. An example

XSL•FO
RenderX

given by a participant concerned the difficulty of fitting grant expectations. If an RI project does not fit expectations, adjustments will need to be made to maintain funding, and the project's efficiency may suffer. Funding is necessary for developing RI tools, but it is also a potential threat if the tool does appeal to funders. A participant stated as follows:

> *Make recommendations to other departments and divisions within the health department to say, "Hey, we need to allocate resources here. Here's what we're hearing." And to your point, we're listening to people's individual stories. We're also looking at data as a whole. And the reality is, we've tried to push a lot of these dashboards and a lot of these things to the general public...They're not really interested in a lot of that unless it's very easy to use and speaks directly to them. So again, my job is that, "Okay, that's not working. And how can I frame it in a way that's really going to focus on that individual so they can make an important decision themselves?" So a lot of what we've done is really just change the way we present data.* [Participant from a county health department]

RI tools are also tasked with the challenge of measuring resilience, especially across the various fields that contribute to resilience research; for instance, resilience must be defined in a standard way, and units of measurement may need to be standardized according to the definition. A group of participants had the following discussion:

> *Speaker 6: Do you guys think that's a problem, the way that we measure their resilience, and solve problems in this case? So the thing that we should think about how we are measuring this resilience about other things right here, I don't know if we have the right tools right now to do that. The resilience measurement tools.*
>
> *Speaker 9: And then the definition of resilience, which alters person to person, agency to agency.*
>
> *Speaker 6: Yeah, that's true.*
>
> *Speaker 9: Race, ethnicity, there's so many implications, things that affect that. So I always look at it as you have to define what the meaning is first in order to measure it. Otherwise, its positive outcomes, just how you go about it. What do you define as success in the situation? Again, that's always going to alter and shift.*
>
> *Speaker 8: Or how do you weigh different? If you can agree that resilience is sort of a combination of metrics that we already account for, okay, so we're counting. We're trying to measure things like food security and vulnerability, heat and different mental health metrics. But then you try to combine that into some sort of measurement of resilience. It's like, "Well, how do I weigh in some sort of statistical way, the weight of my mental health versus my heat resilience, which are all..."*

Later, in the same group, a participant stated as follows:

> *We spent all this time trying to measure the wrong things, like risk. Risk is great. Let's take probability times consequence and its risk. And then we've got everything solved and we're measuring the wrong things. So at least now that we can talk about resiliency and all of the constructs that come with it, you package it up and you've got to talk about social cohesion and adaptivity and vulnerability, and then you sprinkle in some of that probability and consequence stuff. But that's so reductive that you miss the point of what resilience is, which it's an intended positive outcome about how we can work together and have shared goals that are different than individual goals. But you have to have all of that. So, I think it's a really positive thing for results is hard to measure, and you have to roll around in the mud with a pig for a while in order to put some thoughts together. Yeah, it is refreshing.* [Participant from Arizona State University Knowledge Enterprise]

Technological Challenges With RI Tools

Several challenges that are more technological were also mentioned by many participants; for instance, infrastructure issues such as broadband access are problematic for the effectiveness of RI interventions, which often rely heavily on the presence, functioning, and quality of technological tools. At times, government intervention may aid in increasing the necessary infrastructure; for example, some government programs could help to provide free Wi-Fi or expand hotspots. Rural communities were mentioned as being especially vulnerable to such issues, where there may be higher turnover in technology. In this case, low-cost smartphones may be needed to support these communities, or perhaps a community-operated mobile phone provider could be enlisted to help increase mobile phone coverage and access.

Communication

RI communication faces a spectrum of barriers and facilitators, prominently highlighted by the necessity for 2-way communication. This involves ensuring a cohesive message and fostering a shared understanding among stakeholders. However, intergenerational disparities present challenges because differing communication preferences and styles may hinder effective dialogue. Establishing rapport amid these differences can prove arduous, as illustrated by the struggle to navigate technological nuances and preferences, such as the preference for traditional email over newer communication methods.

Moreover, the landscape of informatics communication is marked by silos, where various disciplines converge, including developers, computer scientists, behavioral economists, public health practitioners, and social scientists. This interdisciplinary nature can be both a boon and a bane. While it offers diverse perspectives, it also necessitates concerted efforts to bridge disciplinary divides. One expert at the workshop asked the following question:

> *What about in nonemergent situations? Nonemergent public health situations? How can that gap the bridge? And I'm talking about between the academics*

XSL•FO
RenderX

> *in between the field of public health and then between the clinical side? That's my question.* [Participant from Arizona Advisory Council on Indian Health Care]

In addition, the gap between the clinical and public health realms underscores the need for enhanced collaboration and communication strategies to align goals and priorities effectively. Overcoming these barriers requires deliberate efforts to foster a culture of inclusivity and open communication across sectors and disciplines.

Components of an RI Project

Several different essential components of a successful RI tool or project were mentioned during many of the conversations (Textbox 4).

Textbox 4. Key considerations and essential components when designing a resilience informatics (RI) intervention.

- Systems and infrastructure to process data (broadband access)
- Data integration hub and intelligence subsystem for information sharing
- Anomaly finder (could be an artificial intelligence [AI] function with data search)
- AI to digest complex information, summarize relevant information, and then push it to the relevant audience
- Centralized hotline or integration to allow individuals to have a resource to call
- Bidirectional flow (receive community input)
- Equity assessment (part of before, during, and after evaluations)
- Human elements and user tests
- Coalition oversight to have a big-picture view and report back
- Communication component to hit all different networks

Standards for RI Tools

Overview

A major theme that emerged from the RI workshop discussions was the need for standards or guidelines to help guide the design and implementation of RI tools in the future. Through an analysis of participant conversations, a list of standards for RI tools was developed based on the characteristics presented as ideal for RI tools. While it is recognized that this list may not be comprehensive and that some of the standards may apply to some but not all tools, it may serve as a starting point for future iterations of RI tool standard development. The standards have been grouped into 3 categories based on the nature of the characteristics in question: technological, logistical, and sociological.

Technological Standards for RI Tools

Technological standards were mentioned in many of the discussions. These standards may be helpful for software developers, statisticians, and other technical professionals involved in RI projects.

RI Tools Should Be Up to Date

Similar to other forms of technology, RI tools must be regularly maintained and updated to ensure continued accessibility and usefulness, according to our participants. RI practitioners must also put into place standards for data used by RI tools. This may entail connecting multiple resources to get a broader scope of data, training AI tools by using data from diverse populations to remove bias, establishing standards to ensure the accuracy of data, and using validated tools and objective measurements when implementing RI.

Data Must Be Accessible

Another concern regarding RI tools mentioned in the expert discussions at the workshop was accessibility: data should be designed to be accessible (following accessibility guidelines) while still being protected. Tiered access to data (eg, nonpublic facing or anonymized reports to preserve the privacy and security of data) should be implemented. There should also be regulations in place related to the sharing of private corporate data, such as the number of COVID-19 test kits sold.

RI Tools Should Be Interoperable

Furthermore, RI tools need to be optimized for all major platforms to ensure interoperability across different technologies. This could involve creating versions of RI tools for Apple and Microsoft operating systems, among others.

Data Sovereignty and Privacy Should Be Prioritized

Finally, protecting data sovereignty and privacy must be of utmost importance for RI tool developers and users. This especially applies to the use of RI tools with populations who may be particularly susceptible to abuse or exploitation by bad actors. Data-sharing agreements can be beneficial tools for parties collaborating on resilience-related public health projects with RI tools; however, these agreements must explicitly be written to protect the sovereignty, ownership, and privacy of the data collected and shared.

Logistical Standards for RI Tools

Another type of RI tool standard that emerged from the discussions among workshop attendees focused on logistical characteristics. These are more related to the planning, organization, design, and practical aspects of implementing or rolling out the tool, rather than the technological or sociological aspects of the tool or RI project.

XSL•FO
RenderX

RI Tools Require Thoughtful Design

It was mentioned that when designing RI tools or interventions, it is important to use a systematic approach, following step-by-step procedures. These procedures should include data collection and performance evaluation to ensure continued monitoring and evaluation of the project, ensuring that objectives are met and that the necessary information is available to demonstrate the effectiveness of the intervention. Examples of disciplines or approaches from which RI tools could draw best practices include human-centered design, mixed methods research, and clinical translational science. The acceptability of the desired behavior that the RI tool may be promoting should always be considered as well in the design and implementation of the project. The design could additionally include an implementation model, considering infrastructure, service, providers, users involved, and so on.

The Process of RI Tool Development Should Be Iterative

The workshop discussions highlighted the need for continual community testing and evaluation. One participant stated as follows:

> *We're always reinventing.* [Moderator from the University of Colorado]

RI Tools Need to Have Real-Time Responses

RI tools should work in real time and be able to pivot and adapt quickly to fit the needs of end users. Rapid and clear information should be conveyed in a simple report afterward as well.

RI Tools Need to Have Multiple-Layer Solutions

RI tools should function at a variety of levels to meet the needs of the communities they serve; for example, an ideal RI tool should consider all channels that the end users may use: SMS text messaging, apps, social media, and so on.

RI Tools Require Funding

While perhaps evident, RI tools need funding to be most effective, and this was emphasized many times at the workshop. In the planning and design stages, it is important to consider where these funds will be coming from to make sure that the tool is most effective and sustainable: grants, sales, and so on. Multiple funding streams may be best to consider, should it be possible. In addition, funding is important to consider when budgeting. One participant stated as follows:

> *How can I give our community partners something that they can turn around and use as leverage for money, resources, whatever?* [Participant from a county health department]

Sociological Standards for RI Tools

The following sociological standards for RI tools were identified at the workshop.

Focus on Community-Identified Needs First

Our participants mentioned that RI tools should prioritize populations with the greatest needs within the communities being served. To do so, stakeholder engagement in critical conversations about design and intervention implementation is crucial.

Build Trust: "Partnerships Are Key!"

Upfront community engagement is key to RI tool success. Community leaders and partners, particularly gatekeepers to the communities, need to be engaged in the RI tool from the outset. Community members should be additionally involved in decision-making surrounding resource allocation if this is relevant to the RI tool or project. Successful community engagements are characterized by humility; in other words, our experts mentioned that it is important to be upfront about what you do or do not know. Along the same lines, community values must be recognized early on to ensure successful community engagement, such as transparency, humility, and so on. To identify these values and develop community partnerships, we must listen to groups (eg, tribal communities), understand and learn from others, and recognize that all people have important information, no matter their education level or background. Cultural humility and responsibility as well as community-based participatory research methods may provide important lessons from which RI tool developers and implementers can gain a better understanding of successful community engagement and cultural value identification.

Integrate Social Determinants of Health

The RI workshop participants mentioned how important it is to integrate social determinants of health, such as income, education level, and health care access, into the design, implementation, and evaluation of the RI tool. These factors are crucial when considering public health outcomes because they can influence and potentially confound the tool's impacts.

Long-Term Solutions and Sustainability Should Be the Goal

The RI tool must have a long-term vision and sustainability as the goal to most effectively build resilience in communities. One participant stated as follows:

> *Systemic change takes time, and it builds, and you've got to take things and use things to your advantage.* [Participant from the Arizona Department of Health Services]

Along these lines, the RI project will need a way to follow up with participants over time to continually evaluate and assess the long-term impacts of the RI intervention.

Tailor the Intervention

Tailoring was a strong theme in many of the conversations that experts had at the workshop. Each RI tool must be tailored to the specific population of interest. One participant stated as follows:

> *So you know what's dangerous about that idea is that it's measuring urban capability against rural ability, because you're talking to the one of two people in the entire county that is thinking about public health weather. There is no one else.* [Participant from a county health department]

RI tool designers must additionally contextualize the problem not only by population but also by the specific event as well.

Participants also mentioned that intergenerational solutions (eg, physical community spaces) are important when developing RI tool interventions. Language must always be considered—not

XSL•FO
RenderX

only in its conventional sense but also in terms of vernacular expressions and subgroup dialects—to increase understanding when using RI tools to communicate with different communities. Cultural representation must also be considered when tailoring interventions, considering differences in emergency response language and acronyms and literacy levels. One participant warned against reliance on AI translation, although it seems that AI tools are evolving at a rapid rate and may improve translation capabilities over time. In the meantime, it may be crucial for most communities to either rely on manual translation with the help of local community members or, at the very least, double-check AI-generated translations through manual review. Multimedia and multimodal strategies were suggested as being important to address language barriers or varying literacy levels. Differences in technology accessibility must also be acknowledged and taken into consideration in RI tool design and implementation. To be most inclusive, low-tech or even no-tech alternatives may need to be integrated into the design. One participant remarked as follows:

> *I always say, "It's the generation of someone who prefers a phone call than a text."* [Participant from the cooperative extension office, affiliated with the University of Arizona]

Finally, it was emphasized that it is important for all RI tools to be inclusive for people with disabilities.

Equity Must Be Prioritized

During the workshop, participants were asked to answer a question about how equity would be prioritized in the RI tools; for instance, it is important to collect data to evaluate equity within the RI intervention, such as documenting equity awareness and feedback. Another way to prioritize equity in the RI project could be by involving people from the community in project roles and paying them equitably. Assessing policies and tools (eg, AI) for discriminatory impacts may be an equitable goal or activity for an RI project. Finally, it was recognized in a conversation that the Culturally and Linguistically Appropriate Services standards are an existing guideline that may be useful for RI tool developers to reference and use to increase equitable outcomes.

Issues Related to Data Use

Two main issues emerged in the discussions regarding data use (Textbox 5).

Textbox 5. Issues regarding data use and illustrative quotes.

Data are not always presented in an appealing or accessible way for practitioners

"The informatics can inform our response, but we might be just focusing so heavily on the data that the practitioners out in the field start going, yeah, it's a wildfire, what do you want me to do about it? It's going to burn. It's going to burn that long." [Participant from a county health department]

The right information must be there for somebody to grab it

"I used to develop a lot of software...and we used to do pharmacy applications and our user group would say, 'We need these 20 data fields.' We go, 'Yeah, that makes sense. Okay, we'll put the 20 data fields in.' Three years later we go back, two of them are used. Now did that mean the other 18 weren't important or shouldn't have been used? It probably meant that there was this really subset that needed those data fields to make better decisions. But if you looked at it from utilization of the data field, you would say, 'Well, just get rid of them, we really don't need them.'" [Participant from a county health department]

Indigenous Data Sovereignty

Indigenous data sovereignty was a key theme that emerged from the guest presentation from the Salt River Pima-Maricopa Indian Community and the discussions among the workshop attendees. Indigenous data sovereignty can be defined as "the right of Indigenous Peoples to govern how data from or about them is collected, accessed, used, stored, and disposed of" [32]. Culturally centered sovereignty requires a conscious effort by public health officials, researchers, policy makers, and RI practitioners alike. Explicit protections and acknowledgments of Indigenous ownership of data and the rights surrounding these data must be put into place before RI tools are deployed in Indigenous communities. This involves an acknowledgment of cultural practices, the political sovereignty of Indigenous nations, collective knowledge and shared wisdom, and the relationship between Indigenous peoples and the land.

Martinez outlined 4 principles that must be considered for Indigenous governance known as the "CARE principles": collective benefit, authority to control, responsibility, and ethics. Following these principles is a minimum standard for public health work in Indigenous communities, particularly with respect to sensitive health data collected about Indigenous peoples [33].

Discussion

Principal Findings

In summary, the RI workshop experts confirmed the need for tools that translate technical knowledge into something useful for the community. These tools must be low cost or free, accessible (easy to use), available on any platform (especially mobile phones), and simple in design; not require special domain knowledge; and be connected, bidirectional, responsive in real time, and respectful of data sovereignty. Technological and multimodal alternatives tailored for different groups with varying accessibility needs are crucial for the success of RI tools. The end goals of these RI tools are sharing resources, effecting behavior change, and advancing health equity.

Comparison to Existing Literature

Some of the findings from our workshop resonate with those in similar studies in the literature; for instance, our attendees confirmed that informatics tools have the potential to help build resilience by improving the capacity for households to respond to and make effective decisions during disasters, as well as through improving social capital [34]. Another finding that resonates with existing literature is that there is a need for a

XSL•FO
RenderX

structured guide for the development of informatics tools [35]. Other studies have discussed logistical issues such as data quality and difficulties related to collecting data being obstacles to the success of health informatics projects in addressing pandemics [36]. In addition, similar studies on the COVID-19 pandemic have noted that human factors, ethics, data privacy, and the diversity of participants are vital to consider in the implementation of medical informatics tools [37]. Another study emphasized the iterative nature of informatics tools, necessitating continuous evaluation and reformulation for success [38]. This paper presents a few initial ideas that could be used to begin developing a checklist or framework of standards for future RI tools. Finally, it was demonstrated in this summary from our workshop that RI, and health informatics in general, encompasses a very wide range of functions and has the potential to help make health care services more resilient in the face of climate change impacts, similar to findings from other studies on health informatics [39].

It is worth noting that our workshop did not include much explicit discussion about big data and its potential, which would be an important area to explore in the future, especially considering the potential it has shown in building resilience in other sectors [40,41]. Another important issue that did not come up, despite some conversation about the positive impacts that informatics may have on reducing workforce strain, was the concern that informatics may impact jobs in health care and resilience-adjacent sectors if not introduced with care [42]. Other concerns include the fact that while these tools may be emerging, the workforce to implement them may not be readily available [43]. This topic has been explored in the literature and has been a concern with many other major technological breakthroughs and should be explored further to preemptively avoid any major impacts in restructuring the workforce to ensure that new informatics tools do not have large negative impacts [42].

Strengths and Limitations

This workshop included a wide range of participants from different fields, with each being an expert in their respective area, thus lending credibility and wide-reaching relevance to the findings. A limitation of this workshop is that the findings may be limited primarily to Arizona because most participants were from the state. Nonetheless, it was recognized that many of the discussions were relevant to contexts outside of Arizona as well. In addition, because this was not a qualitative study with a systematic approach to data collection, the findings may not be comprehensive, and future research should be conducted to ensure that a saturation of ideas is reached to gain a fuller understanding of the reach and potential of this field.

Future Directions

Future research questions to explore that emerged from the workshop discussions included the following:

- How do you physically change the environment to protect against infectious disease?
- How do we leverage community knowledge? What is the compensation for this?
- How do we teach empathy?
- What standards are necessary for RI tools to be most effective at addressing public health and climate change–related issues?

It is important for future research in this area to adopt mixed methods and be interdisciplinary, human centered, and community based. The research should ensure that equity considerations are at the forefront.

A follow-up workshop was held in June 2024 in Tucson, Arizona, to present the findings and work toward developing solutions to resilience and public health problems together with the same experts from this initial workshop. For future directions in the field of RI developments, it would be helpful to have more focused discussions with clear outputs such as implementation guidelines or designs of RI tools to put into practice. It is important to continue to build relationships with communities, stay engaged and curious about different cultures, and learn from others.

Conclusions

The RI workshop described in this paper has laid a critical foundation for advancing the development and application of RI tools, especially in the context of public health and climate change. By emphasizing the need for accessible, user-friendly, and equitable technologies that can effectively bridge technical knowledge and community needs, this workshop has underscored the importance of multidisciplinary and community-centered approaches. The discussions highlighted key areas for future research, including physically transforming environments to mitigate disease risk, leveraging community knowledge, and developing empathetic and culturally sensitive tools. With a follow-up workshop planned, the next steps involve moving from discussion to actionable outcomes, focusing on the creation of implementation guidelines and tool designs that prioritize health equity and resilience. In conclusion, continued collaboration between experts, communities, and interdisciplinary teams is essential to realizing the full potential of RI in addressing public health challenges in an era of rapid global change.

Data Availability

The datasets generated during and analyzed during this study are not publicly available to protect the privacy of our participants but are available from the corresponding author on reasonable request.

Acknowledgments

The authors recognize and thank all partners who helped them with developing, and by participating in, this workshop, including Antonio Hernandez, Charles Gerba, Charles Kmet, Darien Fuller, Jose F Florez-Arango, Kerri Rittschof, Patrick Wightman,

XSL•FO
RenderX

Sameer Halai, Sheryl Nelson, Summer Hassan, Sybil Peters, Theresa Cullen, Tim Lant, Yevheniia Varyvoda, Enrique Noriega, Anna Hartman, Ken Komatsu, and Jahnavi Shriram, as well as the others who participated in this workshop. Special thanks to the organizing team, including Rhett Hill, Grace Bohan, Sage Vu, Mary Hayden, Zackry Guido, and Corrie Hannah. The authors additionally acknowledge the use of a generative AI tool (ChatGPT) [44] to aid in creating parts of the qualitative analysis in the Results section. All generative AI output was reviewed and edited by a human to ensure accuracy and validity before the final manuscript was submitted.

This research was funded by a grant from the University of Arizona Technology Research Initiative Fund or Water, Environmental, and Energy Solutions. The authors gratefully acknowledge support from this fund and also from Dr Sharon Collinge, Zack Guido, and the Arizona Institute for Resilience.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Codebook with definitions and examples used for qualitative coding and analysis.
[DOCX File , 18 KB - formative_v9i1e63217_app1.docx]

References

1. Covid-19 map. Johns Hopkins University & Medicine. 2024. URL: https://coronavirus.jhu.edu/map.html [accessed 2024-12-09]
2. Sustainable Development Goals: climate change. United Nations. 2023. URL: https://www.un.org/sustainabledevelopment/climate-change/ [accessed 2023-06-12]
3. Iyengar MS, Merchant N, Ernst K, Rains S, Arora M, Ngaybe MG, et al. Resilience informatics for public health. Stud Health Technol Inform 2024 Jan 25;310:1276-1280. [doi: 10.3233/SHTI231170] [Medline: 38270020]
4. Frumkin H, Hess J, Luber G, Malilay J, McGeehin M. Climate change: the public health response. Am J Public Health 2008 Mar;98(3):435-445. [doi: 10.2105/ajph.2007.119362]
5. Huang C, Vaneckova P, Wang X, Fitzgerald G, Guo Y, Tong S. Constraints and barriers to public health adaptation to climate change: a review of the literature. Am J Prev Med 2011 Feb;40(2):183-190. [doi: 10.1016/j.amepre.2010.10.025] [Medline: 21238867]
6. Gould S, Rudolph L. Challenges and opportunities for advancing work on climate change and public health. Int J Environ Res Public Health 2015 Dec 09;12(12):15649-15672 [FREE Full text] [doi: 10.3390/ijerph121215010] [Medline: 26690194]
7. Haines A, Kovats RS, Campbell-Lendrum D, Corvalan C. Climate change and human health: impacts, vulnerability and public health. Public Health 2006 Jul;120(7):585-596. [doi: 10.1016/j.puhe.2006.01.002] [Medline: 16542689]
8. Varma R, Verma Y, Vijayvargiya P, Churi PP. A systematic survey on deep learning and machine learning approaches of fake news detection in the pre- and post-COVID-19 pandemic. Int J Intell Comput Cybernet 2021 Aug 16;14(4):617-646. [doi: 10.1108/ijicc-04-2021-0069]
9. Mirza SA, Sheikh AA, Barbera M, Ijaz Z, Javaid MA, Shekhar R, et al. COVID-19 and the endocrine system: a review of the current information and misinformation. Infect Dis Rep 2022 Mar 11;14(2):184-197 [FREE Full text] [doi: 10.3390/idr14020023] [Medline: 35314653]
10. Berger Z, Altiery DE Jesus V, Assoumou SA, Greenhalgh T. Long COVID and health inequities: the role of primary care. Milbank Q 2021 Jun 30;99(2):519-541 [FREE Full text] [doi: 10.1111/1468-0009.12505] [Medline: 33783907]
11. Benatar SR. Health in a post-COVID-19 world. S Afr J Sci 2022 Nov 30;118(11/12). [doi: 10.17159/sajs.2022/14995]
12. de Miranda WD, Silveira F, Santos FP, Magalhães Junior HM, Paes-Sousa R. Challenges, consequences, and possible paths for confronting post-COVID-19 health inequalities and vulnerabilities. Saúde Debate 2022 Dec;46(spe8):141-155. [doi: 10.1590/0103-11042022e811]
13. Shanbehzadeh S, Tavahomi M, Zanjari N, Ebrahimi-Takamjani I, Amiri-Arimi S. Physical and mental health complications post-COVID-19: scoping review. J Psychosom Res 2021 Aug;147:110525 [FREE Full text] [doi: 10.1016/j.jpsychores.2021.110525] [Medline: 34051516]
14. Ren FF, Guo RJ. Public mental health in post-COVID-19 era. Psychiatr Danub 2020 Aug 12;32(2):251-255 [FREE Full text] [doi: 10.24869/psyd.2020.251] [Medline: 32796794]
15. Tosato M, Calvani R, Ciciarello F, Galluzzo V, Martone AM, Zazzara MB, et al. Malnutrition in COVID-19 survivors: prevalence and risk factors. Aging Clin Exp Res 2023 Oct;35(10):2257-2265 [FREE Full text] [doi: 10.1007/s40520-023-02526-4] [Medline: 37665556]
16. Grund S, Bauer JM. Malnutrition and sarcopenia in COVID-19 survivors. Clin Geriatr Med 2022 Aug;38(3):559-564 [FREE Full text] [doi: 10.1016/j.cger.2022.04.001] [Medline: 35868673]
17. Lewis M, Holland K, Govender E. The interdisciplinary communication dynamics of the COVID-19 pandemic. In: Lewis M, Govender E, Holland K, editors. Communicating COVID-19. Cham, Switzerland: Palgrave Macmillan; 2021.

XSL•FO
RenderX

18. Cox A, Iedema R, Li S, Sabbe M, Salaets H, Dauby N. Adding barriers to barriers during the COVID-19 pandemic: a call for interdisciplinary research on communication with migrant patients in the emergency department. Eur J Emerg Med 2021 Jan 01;28(1):5-6. [doi: 10.1097/MEJ.0000000000000777] [Medline: 33369953]
19. Pool J, Fatehi F, Akhlaghpour S. Infodemic, misinformation and disinformation in pandemics: scientific landscape and the road ahead for public health informatics research. Stud Health Technol Inform 2021 May 27;281:764-768. [doi: 10.3233/SHTI210278] [Medline: 34042681]
20. Bin Naeem S, Kamel Boulos MN. COVID-19 misinformation online and health literacy: a brief overview. Int J Environ Res Public Health 2021 Jul 30;18(15):8091 [FREE Full text] [doi: 10.3390/ijerph18158091] [Medline: 34360384]
21. Salvador-Carulla L, Rosenberg S, Mendoza J, Tabatabaei-Jafari H, Pandemic-Mental Health International Network. Rapid response to crisis: health system lessons from the active period of COVID-19. Health Policy Technol 2020 Dec;9(4):578-586 [FREE Full text] [doi: 10.1016/j.hlpt.2020.08.011] [Medline: 32874862]
22. Abbas M, Aloudat T, Bartolomei J, Carballo M, Durieux-Paillard S, Gabus L, et al. Migrant and refugee populations: a public health and policy perspective on a continuing global crisis. Antimicrob Resist Infect Control 2018 Sep 20;7(1):113 [FREE Full text] [doi: 10.1186/s13756-018-0403-4] [Medline: 30250735]
23. Ford-Paz RE, Santiago CD, Coyne CA, Rivera C, Guo S, Rusch D, et al. You're not alone: a public health response to immigrant/refugee distress in the current sociopolitical context. Psychol Serv 2020;17(S1):128-138. [doi: 10.1037/ser0000381] [Medline: 31464470]
24. Branham L. Overcoming recurring crisis through resilience: an analysis of USAID's definition of resilience. Clark University. 2019 May. URL: https://commons.clarku.edu/idce_masters_papers/236/ [accessed 2024-12-09]
25. Carlsen L, Bruggemann R. The 17 United Nations' sustainable development goals: a status by 2020. Int J Sustain Dev World Ecol 2021 Jul 30;29(3):219-229. [doi: 10.1080/13504509.2021.1948456]
26. Prabhu A. Informatics. In: Schintler LA, McNeely CL, editors. Encyclopedia of Big Data. Cham, Switzerland: Springer International Publishing; 2020.
27. Bath PA. Health informatics: current issues and challenges. J Inf Sci 2008 Jun 13;34(4):501-518. [doi: 10.1177/0165551508092267]
28. Mosadeghrad AM, Afshari M, Isfahani P, Ezzati F, Abbasi M, Farahani SA, et al. Strategies to strengthen the resilience of primary health care in the COVID-19 pandemic: a scoping review. BMC Health Serv Res 2024 Jul 25;24(1):841 [FREE Full text] [doi: 10.1186/s12913-024-11278-4] [Medline: 39054502]
29. Moran D. Introduction to Phenomenology. Milton Park, UK: Routledge; 2002.
30. Ryan GW, Bernard HR. Techniques to identify themes. Field Methods 2003 Feb 01;15(1):85-109. [doi: 10.1177/1525822x02239569]
31. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec 16;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
32. Garba I, Sterling R, Plevel R, Carson W, Cordova-Marks FM, Cummins J, et al. Indigenous peoples and research: self-determination in research governance. Front Res Metr Anal 2023 Nov 15;8:1272318 [FREE Full text] [doi: 10.3389/frma.2023.1272318] [Medline: 38033627]
33. Jennings L, Anderson T, Martinez A, Sterling R, Chavez DD, Garba I, et al. Applying the 'CARE Principles for Indigenous Data Governance' to ecology and biodiversity research. Nat Ecol Evol 2023 Oct 09;7(10):1547-1551. [doi: 10.1038/s41559-023-02161-2] [Medline: 37558804]
34. Virapongse A, Duerr R, Metcalf EC. Knowledge mobilization for community resilience: perspectives from data, informatics, and information science. Sustain Sci 2018 Aug 1;14(4):1161-1171. [doi: 10.1007/s11625-018-0612-z]
35. Sijm-Eeken ME, Arkenaar W, Jaspers MW, Peute LW. Medical informatics and climate change: a framework for modeling green healthcare solutions. J Am Med Inform Assoc 2022 Nov 14;29(12):2083-2088 [FREE Full text] [doi: 10.1093/jamia/ocac182] [Medline: 36219786]
36. Fernandez-Luque L, Kushniruk AW, Georgiou A, Basu A, Petersen C, Ronquillo C, et al. Evidence-based health informatics as the foundation for the COVID-19 response: a joint call for action. Methods Inf Med 2020 Dec 11;59(6):183-192 [FREE Full text] [doi: 10.1055/s-0041-1726414] [Medline: 33975375]
37. Basit MA, Lehmann CU, Medford RJ. Managing pandemics with health informatics: successes and challenges. Yearb Med Inform 2021 Aug 21;30(1):17-25 [FREE Full text] [doi: 10.1055/s-0041-1726478] [Medline: 33882594]
38. Smith MW, Ash JS, Sittig DF, Singh H. Resilient practices in maintaining safety of health information technologies. J Cogn Eng Decis Mak 2014 Sep 06;8(3):265-282 [FREE Full text] [doi: 10.1177/1555343414534242] [Medline: 25866492]
39. Gray SW, Ottesen RA, Currey M, Cristea M, Nikowitz J, Shehayeb S, et al. Leveraging an informatics approach to identify an unmet clinical need for BRCA1/2 testing among patients with ovarian cancer. JCO Clin Cancer Inform 2022 Sep;6. [doi: 10.1200/cci.22.00034]
40. Harwell LC, Vivian DN, McLaughlin MD, Hafner SF. Scientific data management in the age of big data: an approach supporting a resilience index development effort. Front Environ Sci 2019 Jun 04;7(Article 72):1-13 [FREE Full text] [doi: 10.3389/fenvs.2019.00072] [Medline: 33123540]

XSL•FO
RenderX

41. Shah SA, Seker DZ, Rathore MM, Hameed S, Ben Yahia S, Draheim D. Towards disaster resilient smart cities: can internet of things and big data analytics be the game changers? IEEE Access 2019;7:91885-91903. [doi: 10.1109/access.2019.2928233]
42. Smith SE, Drake LE, Harris JG, Watson K, Pohlner PG. Clinical informatics: a workforce priority for 21st century healthcare. Aust Health Rev 2011;35(2):130-135. [doi: 10.1071/ah10935]
43. McFarlane TD, Dixon BE, Grannis SJ, Gibson PJ. Public health informatics in local and state health agencies: an update from the public health workforce interests and needs survey. J Public Health Manag Pract 2019;25 Suppl 2, Public Health Workforce Interests and Needs Survey 2017(2 Suppl):S67-S77 [FREE Full text] [doi: 10.1097/PHH.0000000000000918] [Medline: 30720619]
44. ChatGPT homepage. ChatGPT. URL: https://chatgpt.com/ [accessed 2024-12-09]

Abbreviations

AI: artificial intelligence
COREQ: Consolidated Criteria for Reporting Qualitative Research
RI: resilience informatics

Edited by A Mavragani; submitted 13.06.24; peer-reviewed by M Khairulbahri; comments to author 03.09.24; revised version received 05.09.24; accepted 06.11.24; published 16.01.25.

Please cite as:
Block Ngaybe MG, Azurdia Sierra L, McNair A, Gonzalez M, Arora M, Ernst K, Noriega-Atala E, Iyengar MS
Resilience Informatics in Public Health: Qualitative Analysis of Conference Proceedings
JMIR Form Res 2025;9:e63217
URL: https://formative.jmir.org/2025/1/e63217
doi:10.2196/63217
PMID:

©Maiya G Block Ngaybe, Lidia Azurdia Sierra, Andrew McNair, Myla Gonzalez, Mona Arora, Kacey Ernst, Enrique Noriega-Atala, M Sriram Iyengar. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Assessment, Decision, Adaptation, Production, Topical Experts-Integration, Training, and Testing (ADAPT-ITT) Framework to Tailor Evidence-Based Posttraumatic Stress Disorder Treatment for People With HIV to Enhance Engagement and Adherence: Qualitative Results from a Feasibility Randomized Controlled Trial

Cristina M Lopez[1], PhD; Angela D Moreland[1], PhD; Stephanie Amaya[1], PhD; Erin Bisca[1], MA, LPC; Christin Mujica[2], MA; Tayler Wilson[1], MA; Nathaniel Baker[3], MS; Lauren Richey[4], MD; Allison Ross Eckard[5], MD; Patricia A Resick[6], PhD; Steven A Safren[7], PhD; Carla Kmett Danielson[1], PhD

[1]Department of Psychiatry & Behavioral Sciences, Medical University of South Carolina, Charleston, SC, United States

[2]Department of Psychological Science, University of Arkansas, Fayetteville, AR, United States

[3]Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, United States

[4]Department of Medicine, Infectious Disease Section, Louisiana State University Health Sciences Center, New Orleans, LA, United States

[5]Department of Infectious Diseases, Medical University of South Carolina, Charleston, SC, United States

[6]Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, AL, United States

[7]Department of Psychology, University of Miami, Coral Gables, FL, United States

Corresponding Author:
Cristina M Lopez, PhD
Department of Psychiatry & Behavioral Sciences
Medical University of South Carolina
67 President St
Charleston, SC, 29425-5712
United States
Phone: 1 843 792 2945
Email: lopezcm@musc.edu

Abstract

Background: Individuals with co-occurring posttraumatic stress disorder (PTSD) and HIV are at high-risk for negative HIV-related outcomes, including low adherence to antiretroviral therapy, faster disease progression, more hospitalizations, and almost twice the rate of death. Despite high rates of PTSD in persons with HIV (PWH) and poor HIV-related health outcomes associated with PTSD, an effective evidence-based treatment for PTSD symptoms in PWH does not exist.

Objective: This study aimed to describe the adaptation and theater testing of an evidence-based intervention designed for people with co-occurring PTSD and HIV.

Methods: The Assessment, Decision, Adaptation, Production, Topical experts-integration, Training, and Testing (ADAPT-ITT) framework guided the formative process used to modify an evidence-based PTSD treatment (cognitive processing therapy; CPT) to meet the unique needs of PWH experiencing PTSD. With the integration of Life-Steps for Medication Adherence (Life-Steps), the adapted protocol (CPT-Life-Steps for integration of adherence; CPT-L) targeted HIV-related stigma and HIV medication adherence within a trauma-informed framework. Theater testing was completed with 7 participants to evaluate acceptability of CPT-L for PWH. The qualitative data (N=54 recordings) used to evaluate and adapt CPT-L emerged from individual interviews conducted with participants after each therapy session as well as exit interviews conducted at posttreatment data collection.

Results: After challenging stigma-related appraisals, participants expressed feeling less constrained by maladaptive thoughts. These shifts translated to increased self-efficacy with both HIV-related care and mental health.

Conclusions: These results indicate that trauma-informed work with PWH should consider the impact of HIV on trauma-related stuck points, intersecting identities (including living with HIV), and challenging internalized stigma. Findings provide evidence

XSL•FO
RenderX

that CPT-L is acceptable and effective in addressing internalized HIV stigma that impacts PTSD symptom maintenance and HIV treatment engagement.

Trial Registration: ClinicalTrials.gov; NCT05275842; https://clinicaltrials.gov/study/NCT05275842?id=NCT05275842&rank=1

International Registered Report Identifier (IRRID): RR2-10.1016/j.conctc.2023.101150

(JMIR Form Res 2025;9:e64258) doi:10.2196/64258

KEYWORDS

PTSD; HIV; adherence; minoritized populations; adaptation; evidence-based; treatment; engagement; posttraumatic stress disorder; stress; trauma; antiretroviral therapy; therapy; symptoms; acceptability; self-efficacy

Introduction

Background

By 2019, an estimated 1,189,700 people in the United States were living with HIV [1]. Although people with HIV (PWH) face many unique psychosocial risk factors that impact health outcomes, trauma exposure and posttraumatic stress disorder (PTSD) have been consistently identified among the most prominent stressors for this population. The prevalence of PTSD among PWH, 30%-74% [2,3] far exceeds that of the general population, 3.9% [4]. Trauma history endorsed for *DSM-5-TR* (*Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision*) criterion A, diagnostic information across the lifespan can include a wide range of adverse childhood experiences, interpersonal violence, survivors of natural disasters, combat, immigration-related trauma, HIV-related trauma if identified as traumatic by patient, and other events endorsed as inducing fear of serious harm to self or a loved one. Data suggest that traumatic events most often reported by PWH include interpersonal violence (physical, sexual, or psychological in nature), witnessing domestic violence or physical abuse as children, and seeing someone be seriously injured or killed [5]. This high comorbidity has important treatment implications as patients with co-occurring PTSD and chronic medical conditions report forgetting (41%) or skipping (24%) medications at significantly higher rates than patients without PTSD (29% and 13%, respectively) [6]. In addition, cognitive impairment and increased avoidant behavior experienced as a function of PTSD may contribute to lower levels of antiretroviral therapy (ART) adherence for PWH with a comorbid PTSD diagnosis [7]. For instance, PWH reported avoiding their ART because it was a reminder of their status [7] and of a traumatic life event [8], and they reported avoidance of clinic care visits due to fear of prejudice and discrimination from health care providers [9].

Notably, advances in ART have led to significant improvements in disease-related morbidity, mortality, and health outcomes for PWH. With consistent use, ART markedly slows disease progression and improves quality of life by suppressing viral load, which can potentially be reduced to undetectable levels. However, failure to reach near perfect (≥95%) adherence rates may be insufficient and detrimental to achieving viral suppression. More specifically, suboptimal ART adherence is strongly linked to viral proliferation [10,11], medication resistance [10,12], disease progression [13], and even death [14]. Unfortunately, the necessary 95% to 100% ART adherence rates are not easily achieved considering the multifaceted barriers that PWH must overcome to reach and maintain the recommended benchmark.

The stigmas attached to HIV, mental health, and minoritized identities are one such barrier. Stigmas can impact PWH through three primary mechanisms: (1) enacted stigma experienced in the form of actual rejection or discrimination by people and society; (2) anticipated stigma, which is the expectancy of discrimination and marginalization; and (3) internalized stigma or negative self-perceptions related to having the disease [15]. In addition, intersectional stigma [16,17], has been shown to differentially impact PWH, such that individuals with additional minoritized identities (eg, women, individuals that identify as sexual and gender minorities, and minoritized racial groups) report higher rates of stigmatization [18], and medication nonadherence [19], compared with those with majority-based identities. Furthermore, stigmatization has also been associated with multiple social determinants of health, such as low education level, poverty, lack of access to health insurance, incarceration, and non-US country of origin [18]. In turn, stigmatization can manifest in significant health disparities including poorer health outcomes, elevated PTSD symptoms due to reinforcement of avoidant coping behaviors, and delayed enrollment in HIV care until significantly ill [18-22].

Theoretical models jointly attribute the observed health disparities to the presence of “minority stress” and the absence of “social safety” among PWH, especially those who experience intersectional stigma. Specifically, the minority stress model [23] proposes that heightened vulnerability to health problems results from the cumulative burden of psychological stress experienced in response to social stigmatization (ie, minority stress). In effect, chronic exposure to stigma may lead individuals to be more susceptible to the development of posttraumatic distortions about the world (eg, enacted stigma) and oneself (eg, internalized stigma) following a traumatic event [15,23]. In addition, stigma may reinforce avoidant coping behaviors that are commonly associated with PTSD (eg, lack of trust in new relationships like with a new provider; not taking medication because it reminds them of a negative thought or feeling, etc). In other words, minority stress can impact mental health by reinforcing maladaptive beliefs associated with PTSD and emotional disorders after traumatic or difficult experiences [24]. These findings emphasize the importance of addressing stigma within evidence-based psychosocial interventions as a means of reducing distress and paving the way for improvements in HIV medication adherence and health outcomes. Expanding on the minority stress model, social safety theory posits that, in the absence of social safety (ie, reliable social connection and

XSL•FO
RenderX

protection), stigmatized PWH routinely navigate the world in a state of maintained vigilance that adversely impacts immune system response [25]. Threats in the social environment are thought to trigger physiological stress responses, including chronic inflammation, that can impair immunological function [25]. Thus, social safety theory suggests that targeting maladaptive beliefs (eg, internalized stigma) and interpersonal patterns (eg, avoidance) may amplify social safety, and improve health outcomes by extension, for trauma exposed PWH [25].

Current Study

Due to the distinct connection between HIV and PTSD, it is important that evidence-based treatments be tailored to address both trauma symptoms and barriers to HIV care for PWH. Although effective interventions to address HIV medication adherence and PTSD are widely available, there are no established integrated approaches that address the dual issues of cooccurring HIV with PTSD and related health disparities. In addition, no research to date has examined the critical question of whether HIV outcomes can be improved among the large number of PWH with comorbid PTSD by treating PTSD symptoms to maximize benefits of health promotion interventions. To address this tremendous void in the field, we propose the integration and evaluation of 2 empirically supported interventions that have been successfully used to improve PTSD- and HIV-related outcomes in this population, namely cognitive processing therapy (CPT) [26], and Life-Steps for Medication Adherence (Life-Steps) [27].

Using the Assessment, Decision, Adaptation, Production, Topical experts-integration, Training, and Testing (ADAPT-ITT) model, a framework for adapting HIV-related evidence-based interventions [28], CPT and Life-Steps were integrated to encompass HIV-related stigma and ART adherence within the scope of PTSD treatment. Upon adapting the intervention for PWH and PTSD, the resulting model (CPT-L) was theater-tested with treatment-seeking members of the intended population. More specifically, maladaptive beliefs influenced by exposure to trauma can interfere with important correlates of health, including a participant's motivation to remain ART adherent [29], access social support [30], and establish trusting relationships with medical care providers [31]. To tackle these psychosocial factors, CPT-L includes cognitive restructuring components introduced weekly through modified worksheets and handouts reflecting topics that are more relevant to the realities of living with co-occurring HIV and PTSD. Acknowledging that these health conditions are experienced within a broader personal and cultural context, CPT-L also placed emphasis on engaging in discussion and exploration of the participants' intersecting identities (eg, HIV status, racial, ethnic, sexual, and gender identity discrimination) [32]. This study presents qualitative findings of feedback solicited to assess reactions to the new intervention and elicit critiques that will guide future efforts to refine and tailor CPT-L.

Methods

Treatment Adaptation

Aligned with the ADAPT-ITT model, the CPT-L adaptation process will comprise 8 phases: assessment, decision, adaptation, production, topical expert, integration, training, and testing; and it has primarily been used in the context of HIV [33]. Findings in the current manuscript reflect the first 4 completed phases of the ADAPT-ITT protocol. During the initial step (assessment) of the ADAPT-ITT framework, the team consulted with the Community Advisory Board (CAB), comprising of constituents with lived experience (with PTSD or HIV), Ryan White clinic personnel (eg, providers, case managers, and staff), peer advocates, and community-based mental health providers. CAB members and topical experts were provided with a review of the extant literature to assess the scope of the problem and whether it merited development of new evidence-based treatments to address comorbid PTSD and HIV. Consideration was given to studies on mental health disparities in PWH as well as conflicting evidence about the efficacy of existing treatment modalities for HIV and co-occurring mood disorders. Evidence suggests that although existing psychosocial interventions can be effective in treating depression among PWH, symptom improvement does not translate consistently to improvements in HIV-specific markers (eg, ART adherence and viral load) [34]. Therefore, the team heeded CAB feedback and determined that adaptation was necessary to attend to the unique and unmet needs of this population.

In the second step (decision), the CAB and team's selection of CPT among the available evidence-based PTSD interventions was driven by its flexibility to integrate HIV-specific content and CPT's established training model for successful dissemination [35,36]. CPT, is an evidence-based treatment shown to be effective for reducing posttraumatic distress, including PTSD and depressive symptoms, among civilian and military adult populations with complex histories of exposure to traumatic events [37,38]. The protocol is designed for 12 sessions, but the length of treatment can be adapted based upon the patient's needs [39]. In terms of structure, after centering the first half of treatment on dysfunctional cognitions about the worst traumatic event, the focus shifts to maladaptive beliefs that highlight the following 5 areas commonly impacted by trauma: safety, trust, power and control, esteem, and intimacy. CPT has been shown to reduce trauma-related self-blame and guilt [40,41] and may be particularly beneficial for PWH because of the focus on "stuck points." Stuck points are negative, inaccurate, and unhelpful beliefs about yourself, others, or the world that occur as a result of trauma. Maintaining these stuck points prevents the individual from recovering from PTSD. The goal of CPT is to assist the patient in identifying stuck points related to their traumatic event, and then challenge these stuck points or negative thinking patterns through cognitive restructuring [42]. The process of cognitive restructuring is used to identify alternative, more accurate, and more helpful thoughts through Socratic questioning (ie, open-ended questions from the therapist that stimulates the patient to question current beliefs as flawed). In addition, the broad scope covered under the 5 themes allows for the flexibility to address internalized stigma related to HIV and marginalized identities. Indeed, although it was not originally developed as an HIV intervention, CPT has shown potential in promoting HIV risk-reduction [43].

Adaptation focused on modifying CPT to bolster its effectiveness with HIV medication adherence and related health

XSL•FO
RenderX

outcomes. The CAB's decision to incorporate Life-Steps for medication adherence [27], into CPT was made based on studies showing its efficacy and successful integration with various mental health treatment modalities [27,41]. Life-Steps is a one-session intervention using cognitive behavioral strategies, problem-solving, and motivational interviewing to promote adherence. The modified intervention (CPT-L) used the single-session adjunct to establish baseline HIV and ART psychoeducation and problem-solving skills. CPT-L later leveraged this foundational knowledge by incorporating continuous reminders of the benefits of remaining ART adherent and restructuring maladaptive thoughts that posed barriers to HIV care. The integration of the Life-Steps component also addressed health literacy and concerns about ART through tailored explanations about the need for consistent adherence, psychoeducation on the co-occurrence of PTSD and HIV, and rehearsal of adaptive health behaviors (eg, recognizing barriers to adherence and using problem-solving strategies to confront them). In the context of the larger study aims, CAB members chose Life-Steps as the control condition to ensure that participants assigned to the study arm would receive benefit as well.

The third step (adaptation) was completed through a larger pilot trial. The protocol was theater-tested with members of the target population, namely PWH and co-occurring PTSD. Aligned with the ADAPT-ITT model, reactions and feedback about the intervention were elicited from participants and stakeholders and include the qualitative findings presented here. Currently, CPT-L has a drafted manual (stage 4 of the ADAPT-ITT model) and plans to continue the adaptation process with topical expert feedback are underway. Refer to Table 1 for CPT-L protocol modifications based on ADAPT-ITT steps 1-4 that have been completed. Next stages of the project include an iterative process of refining the manual by eliciting comments and suggestions from our CAB, clinical team, peer advocates and other key constituents (step 5–Topical expert feedback), integration of that feedback in the CPT-L manual (step 6–Integration), and training of personnel to deliver the adapted intervention (step 7).

Table 1. Components and modifications for a tailored, evidence-based PTSD treatment for people with HIV (cognitive processing therapy- Lifesteps): adaptations informed by a qualitative study using the empirically supported ADAPT-ITT framework.

Sessions	Intervention component	CPT-L[a] refinements
1	Life-Steps; begin overview of PTSD[b] and CPT[c]	Psychoeducation about connection between HIV status, stigma, and trauma; how held identities impact trauma and HIV experiences; assign additional impact statement assigned if trauma and receiving news about HIV status are separate events
2-3	Finding maladaptive HIV beliefs	Identify beliefs related to internalized stigma through HIV-related impact statement; use HIV-related examples in context of ABC worksheets
4-5	Processing the index event	Distinguish trauma event processing from receiving news about HIV status; processing stigma from learning of HIV status; use HIV-related examples in context of challenging questions worksheets
6-7	Learning to self-challenge	Use HIV-related examples in context of challenging beliefs worksheets
8-10	Trauma themes (safety, trust, and power-control)	Regain "control" of health by challenging stigma-related stuck points; adaptation of Trust Star Worksheet to include status disclosure
11-12, 12+	Esteem, intimacy, and facing the future	Challenging stigma-related non-adaptive cognitions, duty to warn review for clinician
Delivery adaptations		
Format change	Variable length	Could use additional "stressor sessions" to discuss stigma-related stuck points[39] in CPT[c] variable length format
All	Adherence self-monitoring	In addition to risk assessment in CPT[c], CPT-L[a] includes ART[d] adherence monitoring
All	Trauma, stigma, and HIV connection	Continually developing the connection between HIV status and trauma; discuss impacts of intersectionality on HIV and trauma experience, as appropriate
All	Practice assignments	Therapist to encourage challenging both trauma-related and HIV-related stuck points between sessions (if separate events)

[a]CPT-L: CPT-Life-Steps for integration of adherence.

[b]PTSD: post-traumatic stress disorder.

[c]CPT: cognitive processing therapy.

[d]ART: antiretroviral therapy.

Often, when adaptations of evidence-based treatments are attempted, the process generally ends there. However, a critical element of the ADAPT-ITT model includes conducting a test of the full intervention (step 8). While the involvement of stakeholders across multiple phases of the adaptation processes helps ensure that the adaptation resonates with the community and reflects their needs, the last and final phase of the ADAPT-ITT model determines that the adapted intervention

XSL•FO
RenderX

has met the goal of creating positive and incremental impact for a specific population. Thus, subsequent steps of the adaptation process for CPT-L include a larger trial to evaluate the final product in a rigorous and systematic way. The submitted project proposal for the large-scale testing of CPT-L also includes an aim focused on implementation science and CPT-L intervention scale up, an important aspect to the team to remain aligned with the feedback from the CAB members.

Participants and Recruitment

Participants were recruited from the Infectious Disease Department's Ryan White Clinic at the academic medical center. A full description of the study procedures is reported elsewhere [44]. All protocols were approved by the institutional review board at the affiliated university. The study's program coordinator recruited participants during their regularly scheduled HIV care clinic appointments and used a "cold call" list to contact individuals with HIV outside of these appointments. The study's purpose, methods, and compensation were discussed with participants, and if they were interested in participating, the initial contact assessment was completed. This initial assessment included completing a shortened version of the PTSD checklist for *DSM-V* (*Diagnostic and Statistical Manual of Mental Disorders* [Fifth Edition]; PCL-5), as well as the Montreal Cognitive Assessment (MoCA). If participants scored at least 3 out of 5 on the shortened PCL-5, indicating possible PTSD, and scored within the mild or moderate range on the MoCA assessment (ie, no indication of severe cognitive impairments), then the consent process was completed. Consented participants then completed the clinician administered PTSD scale for DSM-5 to determine if they met criteria for PTSD. Participants who met criteria for PTSD were able to continue through the study. Additional eligibility criteria included (1) age 18 years or older enrolled in a Ryan White clinic; (2) able to speak, read, and write English; (3) meet at least one of the following HIV care criteria: (a) diagnosed with HIV in last 3 months; (b) detectable viral load in last 12 months; (c) failed to show up for or missed 1 or more HIV care appointments in the past 12 months; (d) last HIV care visit was more than 6 months ago; and (e) self-reporting less than 90% ART adherence in the past 4 weeks.

CPT-L was completed individually with a licensed professional counselor and was offered virtually and in person, depending on the participant's location and preference. All participants of the theater testing phase were provided with the standard CPT workbook, as well as modified worksheets specific to CPT-L (eg, Stuck Point log with examples of HIV related stuck points, modified ABC, Challenging Question, Challenging Beliefs worksheets with examples using HIV related stuck points, and a Trust Star geared toward aspects of HIV status disclosure).

As part of a larger pilot trial using theater testing for the first stage, 7 patients completed end of session interviews at the end of each CPT therapy session and exit interviews at the end of protocol (54 total session recordings; number of sessions per patient ranged from 1 to 12; M=7). Race and ethnicity included African American (n=5) and non-Hispanic White (n=2). Participants' ages ranged from 23-64 years. There were 4 males and 3 females, and participants identified as gay (n=2), bisexual (n=2), or heterosexual (n=3). Index traumas included life threatening illness or injury (n=1), physical assault (n=1), transportation accident (n=1), sexual assault or rape (n=2), and other unwanted sexual experiences (n=2).

Procedures

Patients participated in end of the session adaptation questions following each session and intervention exit interviews that explored strengths and weaknesses of CPT-L, including feedback on adaptations of CPT to people with HIV and suggestions for improvement in the adaptation before the large scale randomized controlled trial. The end of session adaptation form consisted of 2 questions that asked for participants to provide a rating from 1 to 10 (1=not at all and 10=very much): "How much did the materials covered in today's session address aspects of people living with HIV?" and "How much did the handouts reflect your experiences living with HIV and PTSD?" Participants were also asked to provide open-ended feedback on these questions and additional information regarding how useful the materials were and how they will use the skills taught in therapy in their everyday lives. In addition, exit interviews were conducted at the end of treatment. Exit interviews consisted of 6 primary questions and 13 follow-on questions (Table 2) that took place in 2022, in English, over videoconference and lasted approximately 5-15 minutes. Participants were compensated US $25 for completion of each exit interview.

XSL•FO
RenderX

Table 2. Exit interview questions completed by patients with HIV and comorbid PTSD at the end of a qualitative study exploring acceptability and relevance of a tailored, evidence-based PTSD treatment for people with HIV (cognitive processing therapy-Lifesteps).

Primary question	Follow-up question
Tell me about the mental health treatment or counseling you have received before	• What did you like? • What did you dislike? • What was missing? What could have been more helpful? • How did you feel about the overall experience?
Tell me your thoughts about the 12-session trauma and adherence-based intervention that you just completed over the past 6-8 weeks	• Was the length of treatment problematic? • Was it beneficial to discuss the topic of your trauma or stressful life experience? • Was the intervention beneficial to discuss and process thoughts about your HIV status? • What reservations do you have about recommending this treatment to be available clinically? • What changes, if any, do you think would need to be made to trauma treatment to make it most helpful, relevant and appropriate for PWHa? • Do you think that the same trauma treatment that is used with other groups would be equally helpful for PWH[a]? Why or Why not?
Did you feel that the CPT-L[b] intervention specifically addressed HIV-related topics?	• How do you feel that your experience of the trauma is different or unique than a survivor who is not living with HIV?
What coping behaviors for reducing your traumatic stress symptoms did you learn from participating in CPT-L[b]?	• Are these coping strategies different than the ones you used before CPT-Lb? How so?
What additional insights did you gain about the identities you hold?	• Did the examples and discussion included in the intervention resonate with the different identities you hold?
What concerns do you have if a friend were asked to participate in this study?	—[c]

[a]PWH: persons with HIV.

[b]CPT-L: CPT-Life-Steps for integration of adherence.

[c]Not applicable.

Data Analysis

All qualitative recordings (end of session adaptation questions and exit interviews) were transcribed verbatim. Following transcription, content from the qualitative recordings were analyzed using a systematic thematic analysis within a grounded theory approach [45,46]. A “ground-up” approach involves an iterative process of establishing codes, grouping the codes into key concepts, and organizing the concepts into broader themes. NVivo (version 12) software was used to organize data and conduct the qualitative analyses. This software package facilitates coding, storage, retrieval, and analysis of large amounts of transcribed data. The codebook was created in a collaborative, inductive, and iterative process. All qualitative recordings were coded by a primary coder (second author) and 25% of recordings were double coded for interrater reliability. Both coders have expertise in qualitative coding of mental health research; neither coder had any interaction nor relationships with participants. Meetings between coders were held to reach a consensus on inconsistent codes for each transcript.

The qualitative researchers were careful to ensure data saturation in the current study given that only 7 participants were included. In qualitative research, there are 5 components that influence data saturation which include predetermined codes and themes, sample size, relevancy of research subjects, number of research tools, and length of data collection sessions [47]. When analyzing the responses from the 7 participants, the qualitative experts determined that data saturation was reached based upon the themes directly mapping onto the questions asked in the interview, previous research showing that saturation can be reached in approximately 10 interviews, that all participants in this study were patients receiving CPT-L and thus, were the most relevant to answer the questions about treatment, and the length of data collection certainly increased the chance of early saturation in this study given that 54 interviews were collected over the data collection period.

Ethical Considerations

The study was approved by the Medical University of South Carolina Institutional Review Board for Human Research (Pro00106801). Before engaging in the study and collecting data, each participant was required to give written informed consent, either in person or virtually using the Medical University of South Carolina’s institutional review board approved REDCap (Research Electronic Data Capture) e-Consent. Within the consent process, research staff discussed with potential participants that the results of the research study, including interviews, may be published and information would not be identifiable. The procedures also clarified that all information obtained from participants would be confidentially maintained. Research staff received training from the principal investigator on the importance of confidentiality and had weekly

XSL•FO
RenderX

supervision meetings with the principal investigator ensuring strict compliance with the Human Subjects Protection plan. Additional measures for protecting privacy and confidentiality also included assigning each participant a unique project identification number. The codes that linked the name of the participant and the study ID number were kept confidential by the principal investigator in a secured, locked cabinet. The secure servers that housed the data files collected from the web applications (REDCap) during were located at MUSC in Secure Sockets Layer 128-bit encrypted servers behind firewalls. Participant confidentiality was masked in all data files by the use of project identification numbers rather than personal information.

After all the relevant information was reviewed verbally with trained research staff, participants signed the document (either virtually or in person) and the signed consent document was stored electronically (from e-consent or scanned in) in the REDCap database, which provides password protected, secure, web-based collection and management of research. Informed consent was conducted by members of the study team that received training through the University of Miami CITI training in both the Conduct of Human Subject Research and Good Clinical Practice, and were institutional review board–approved study personnel. Study data collected, including exit interviews, was conducted either in-person in the privacy of a clinic room, or remotely through telephone. Participants were informed when the interview recording began and ended and were instructed not to use the names of individuals or reveal any personally identifiable information. These audio files were deidentified within 24 hours and uploaded into the REDCap electronic study database to later be transcribed. Any files on a portable recording device were deleted following this upload, and only the project therapist, fidelity raters, and qualitative coders heard the content of the audio recordings. Participants were not compensated for interviews conducted at the end of each CPT-L session, but were compensated US $25 for end of study exit interviews.

Results

Descriptive Characteristics

Regarding the materials addressing aspects of people living with HIV, participants rated the materials very highly (mean 9.86, SD 0.93) and indicated that the handouts greatly reflected the experiences of living with HIV and PTSD (mean 9.71, SD 0.61).

Qualitative Results

Four overarching themes, each with their own subthemes, emerged from the semistructured interviews: (1) overall strengths of CPT-L, (2) overall suggested improvements of CPT-L, (3) helpful handouts/activities/materials, and (4) ways that patient would like to incorporate skills learned from CPT-L.

Theme 1: Overall Strengths of CPT-L

Most patients discussed overall strengths of CPT-L, with the main subthemes emerged including that patients generally felt materials were relatable and relevant to HIV, information was relevant to HIV, learned tools unique and specific to trauma related to HIV, discussed topics never addressed before, reported general improvements in well-being and hope, and learned a different (less negative) view of HIV status following CPT-L engagement. In discussing that they generally felt materials were relatable and relevant to HIV, one patient reported, “It’s making perfect sense and I think it’s really helpful,” and another patient noted, “[CPT-L] is very sensible. It’s pragmatic, it’s logical. I wish I had this a long time ago.” Several patients discussed that the model was very relevant to HIV, making statements such as, “everything was covered. HIV, on what living with HIV is,” and “Oh yes, those would be concerns that people living with HIV would have.”

All patients discussed at some point in the exit interviews that they learned tools to address relevant issues. For example, 1 patient stated,

> *I felt increasingly as I participated that it was giving me some tools to live more – well, to be more whole and not feel as constrained by like stuck points in a way.*

Another patient discussed,

> *I wish I’d had these tools a very long time ago ‘cause I could have probably worked through some things more successfully.*

Relatedly, several patients reported that they discussed topics in CPT-L that they have never discussed before, including

> *It’s honestly the first time since that sexual abuse that I even tried to tackle this,” and “PTSD as far as a self-recognized condition is kind of new to me. So there’s still kind of a process of exploration about that.*

In describing that they now have a different view of their HIV status following engagement in CPT-L, 1 patient stated,

> *I don’t see HIV as a threat to me as much as I once did.*

Another patient described,

> *Knowing that it’s not my fault. That you know, I’m HIV and still can be successful.*

Finally, most of the patients discussed that they felt general improvements in well-being and hope following CPT-L. One patient described,

> *I honestly feel a little more hopeful. I feel better about my decision about how to deal with this feeling of hopelessness. And I think that I’m finding that I put more reasonable weight on events than I used to and some of it is coming from this where I’m just not looking. I feel like I’ve got a prism if you will look at things and not necessarily have them so heavily weighted one way or the other.*

Other patients discussed improvements in well-being and hope by stating,

> *I feel a little stronger about being able to take care of myself. Not just as far was physiologically with procedures but with my state of mind*

and

XSL•FO
RenderX

> *I think [CPT-L] is increasing, you know, kind of my recognition of my place in life. And that I can improve things. You know? And you know, not feel so crestfallen about my place in the world.*

Theme 2: Suggested Improvements in CPT-L

About half of the patients discussed suggested improvements in CPT-L, with the main subthemes including adding more HIV-relevant examples, to individualize the activities, and to clarify some of the handouts. Specifically, some patients who suggested adding more HIV relevant examples stated, "Is there a way to make the trust worksheets more reflective of HIV," "There were like only 3 choices and they weren't really talking about HIV," and "The handouts might be better with advocate stories from people with HIV." One patient who thought the activities could be more individualized mentioned, "I think that for me though because of my generation so to speak and some of the things I've seen while I've been HIV positive as well as had my own sexual orientation that those things might be, – they don't touch on me so much." Another patient stated, "I just think that some people will respond well to this method and other's less so." Finally, some patients discussed ways to clarify some of the handouts with statements such as, "Some of them were not self-explanatory altogether," and "As far as like the terminology that maybe there's something to you know actually define it."

Theme 3: Helpful Handouts, Activities, and Materials

All 7 of the patients gave feedback on their perceived helpful handouts, activities, and materials during completion of CPT-L, with the main subthemes including handouts, challenging stuck points, and exposure activities. All patients made positive comments about the handouts, including general comments about the handouts such as

> *The handouts made me reflect on the things that I asked me to do and things that I need to let go of,*
>
> *The handouts reflected someone living with both HIV and PTSD. They reflected a lot,*

and

> *I think that they could equally be applied to a lot of other difficult circumstances people have. And I think with certain individuals that it might be more helpful for them to have something that's tangible.*

Several patients also made positive comments about specific handouts, including

> *I actually like this trust star so to speak. I think that it could actually be pretty useful just as an exercise about trust generally rather than specifically.*

Almost all of the patients discussed that the challenging stuck points activities were helpful. Specifically, 1 patient noted,

> *To reflect on, you know, an issue and look at it from a lot of different facets in a lot of different ways,*

and another mentioned,

> *It works. I mean it's asking me directly to address the [stuck points] and to put things down and look at them and go through a systematic process. You don't like activating them but it's working.*

In addition, most patients reported that the exposure techniques were very helpful in decreasing symptoms. One patient described,

> *Just talking about it. I mean, you let me trigger some things that I didn't even think I was gonna think about, yeah?*

Theme 4: Incorporating Skills Learned in CPT-L

Many of the patients described ways that they plan to incorporate skills learned in CPT-L moving forward, including setting different goals, engaging in different activities, practicing challenging thoughts and stuck points, being more aware of feelings, and improving self-acceptance and self-esteem. Regarding goals, many patients described that they plan to set different goals for themselves. One patient reported,

> *I wanna set different goals for myself and maybe think about how I'm doing things a little bit differently and as far as you know, making a more dive into this state.*

Another patient stated some personal goals,

> *Not being in my PJs and taking a shower before session. And combing my hair.*

In addition, a patient discussed a goal of consistently taking medication by stating,

> *I need to make sure I'm on my meds for the next three months to get shots and I hate shots, so this has helped me.*

Relatedly, several patients described that they plan to engage in different activities following completion of CPT-L, including

> *I can start planning something. I will start with people from church" and "If I have to do at least one thing just for myself, not because I did anything to deserve it, or whatever right? I can do that.*

Most patients described that they plan to practice changing thoughts and challenging stuck points following completion of CPT-L. Specifically, 1 patient reported,

> *I think that there's a great likelihood that I'm probably going to filter what I think and how I react,*

while another patient stated,

> *Reconstruct my perception or think about alternatives.*

A few patients reported that they plan to be more aware of their feelings, as 1 patient stated,

> *Assessing feelings, and probably being more aware of them. And their impact on people around me.*

Finally, several of the patients described that they would like to improve their self-acceptance and self-esteem following the completion of CPT-L. Specifically, 1 patient made a comment about self-acceptance and self-esteem including,

> *The way I think about what happened with my mother and trying to use it as a lens to help me have a better sense of self instead of looking for some reason why I'm the way I am. Just more acceptance of the way I*

XSL•FO
RenderX

am. Because I mean, I don't feel like I'm altogether horrible.

Other patients described,

Try to think more positive about myself,

and

I can already see this has given me a construct with some new things that float into my mind. That are kind of self-judgments that I can, you know, have a better way of dealing with.

Discussion

Recent years have seen a growing awareness of the need for trauma-informed, HIV treatment approaches as converging evidence suggests a reciprocal association between HIV and exposure to trauma. Having an HIV diagnosis is a significant predictor of physical and sexual victimization [48-49], which in turn has been associated with more engagement in HIV-risk behaviors, poorer physical functioning, increased symptomology, and higher rates of mortality among PWH [12,50-54]. Unfortunately, PWH with co-occurring PTSD must overcome several structural, psychosocial, and behavioral challenges with potentially devastating health consequences. These mutually reinforcing risk factors call for an integrated treatment approach [55]. Indeed, integrated interventions for cooccurring disorders have been found to increase client engagement with treatment and improve patient outcomes [56]. To address this unmet need, the current study introduced a modified treatment protocol incorporating gold standard, evidence-based practices used with trauma exposed populations CPT and PWH (Life-Steps) [27,57]. CPT was selected from similarly effective treatments for PTSD for its mechanism of change (ie, cognitive restructuring), which allows for the flexibility to simultaneously target maladaptive appraisals related to trauma, intersectional stigma, and HIV using a skills-based, modular format.

The resulting framework (CPT-L) leverages treatment components from each protocol to target factors known to contribute to PTSD and ART nonadherence. First, structural barriers, such as low access to (and understanding of) health education, can considerably impede adherence given the complexity of HIV medication regimens [58]; components of CPT-L addressing these documented barriers were well-received by participants in the current study. The integration of Life-steps with CPT provided explanations about consistent adherence and psychoeducation about the co-occurrence of PTSD and HIV. Generally, participants found that the information adapted from both protocols was well integrated and accurately reflected their lived experience. Findings suggest that demystifying co-occurring HIV and PTSD allowed participants to gain a sense of mastery over their own medical care, which then translated to actionable goals to increase and maintain higher levels of ART adherence. The observed patterns were consistent with studies supporting the effectiveness of education-based interventions on treatment adherence for PWH [59,60] but extended existing interventions by incorporating information on the impacts of PTSD. Although some participants lamented not having access to the information sooner, they also noted that it has contributed to a newfound self-awareness and sense of hope. Within the context of the roles of stigma and attitudinal barriers in delayed mental health services, the high patient acceptability of CPT-L are promising and warrant further testing to help improve access to trauma-focused care, both in the United States and abroad [61].

Another known factor contributing to PTSD and ART nonadherence that is addressed by CPT-L are maladaptive beliefs in a multicultural context. In addition to "traditional" (ie, more trauma-focused) maladaptive beliefs that are identified for cognitive restructuring in CPT, increased acceptability and relevance of CPT-L by PWH may be due to CPT-L's emphasis on participants' intersecting identities (eg, HIV status, racial, ethnic, sexual, and gender identity discrimination) to acknowledge and validate broader personal and cultural contexts that can affect health behaviors and management. This multicultural orientation opened a conversation about experiences of stigma (enacted and internalized) they may have experienced. Participants were then guided in processing intersectional stigma through a trauma-informed lens (ie, themes of safety, trust, power and control, esteem, and intimacy). For example, aligned with evidence of the association between disclosure of HIV status and increased social support [43], the "Trust Star Worksheet" was adapted to reflect disclosure as an important type of trust. The Trust Star is an activity completed within CPT to assist the patient in identifying aspects and levels of trust, and how this impacts the way the individual may trust themselves and others. While many individuals with PTSD view trust as an all or none concept, this activity assists the patient in seeing trust as multi-faceted. This in turn assists the patient in challenging extreme or exaggerated stuck points related to trust. Based on results from feedback on the qualitative study, the new conceptualization of the Trust Star in CPT-L included the participant's trust in their own ability to navigate the process of disclosure of their status using their newly acquired tools (captured in the theme "different (less negative) view of HIV status". Upon completing the protocol, participants recognized that much of the shame they experienced about their HIV status, victimization experiences, and intersecting identities was founded on flawed beliefs about the nature of the disease, the world, themselves, and others (captured in themes "practicing challenging thoughts and stuck points" and "improving self-acceptance and self-esteem"). Using thought challenging skills, participants developed more adaptive worldviews that fostered greater self-acceptance (especially regarding sexual orientation and gender identity), trust in others, and trust in themselves (captured in themes "practicing challenging thoughts and stuck points"). Consequently, they were better able to engage in self-care (eg, hygiene and medication adherence) and were more motivated to connect with their social support networks. From the perspective of a social safety theoretical framework, in removing inappropriate shame and internalized stigma that hinders social bonds, feedback from participants highlights CPT-L's potential as an effective tool to decrease some of the deleterious health consequences that traumatized PWH experience as a function of social threat and disconnection [62].

XSL•FO
RenderX

Finally, PTSD-associated avoidance may underlie ART nonadherence such that the medications themselves may cue trauma memories and HIV-related stigma that deter consistent use. In other words, regardless of the reason (eg, forgetting vs choosing), it is possible that medication noncompliance is functionally an avoidant behavior used to escape distressing internal experiences for people struggling with co-occurring HIV and PTSD. In response, CPT-L features exposure as a core intervention component. Specifically, rather than the standard trauma narrative, participants are assigned 2 reflection statements recounting their index trauma and their experience with the news of their HIV status (unless HIV status is the index trauma). For many participants, the therapy setting was their first experience talking, or even deliberately thinking, about their index trauma and HIV status. They generally observed that actively approaching these topics allowed them to reorganize their thoughts and reframe their experiences. To their surprise, participants discovered that their broadened perspectives cultivated esteem and intimacy as they reflected on their relationships to themselves and others. These findings were consistent with evidence suggesting that disclosing trauma histories within the context of research or health care settings poses minimal risk of worsening symptoms, and in some cases can have secondary therapeutic benefit [63]. This data may help address misinformation and provider fears about "retraumatizing" PWH when screening for PTSD. In addition, participants were able to reconceptualize the cause and meaning of their index trauma and HIV status. Coupled with their fostered sense of trust in their own abilities, the shift away from self-blame and internalized stigma helped them regain a sense of control of their health and greater confidence in their ability to manage their own care.

While the intervention was largely well-received, several themes reflecting the need for modifications also emerged within the data. Participants primarily focused on improving handouts and worksheets stating that although the material was helpful, it was not intuitive, and the terminology was sometimes unclear. Several participants specified that they would benefit from additional examples pointedly touching on issues at the intersection of HIV and PTSD. Participants maintained that they would like to see self-acceptance and self-esteem for trauma-exposed PWH modeled through treatment materials (eg, including participant "success stories" and examples of adaptive alternatives to common stuck points).

Notwithstanding these limitations, CPT-L appears to be acceptable and relevant to PWH experiencing traumatic stress and merits further investigation as an integrated intervention for co-occurring PTSD and HIV. After challenging stigma-related stuck points, participants expressed feeling less constrained by maladaptive thoughts. These shifts translated to increased self-efficacy with both HIV-related care and mental health. Although limited in size and scope, the current study yielded preliminary support for the tolerability and effectiveness of the intervention. Findings highlight the importance of tailoring mental health treatment approaches to address trauma-related stuck points, intersecting identities (including living with HIV), and challenging internalized stigma. Such adaptations support goals for health outcomes (eg, ART adherence) by reducing the intensity of trauma-related distress and patterns of avoidance that interfere with HIV-treatment engagement. Future studies can expand on these efforts to complete the ADAPT-ITT process, including refining and testing CPT-L on a wider scale. Based on these qualitative results, integration and evaluation of strategies that foster social safety in mental health interventions for PWH warrants further attention.

Acknowledgments

This manuscript was supported by a grant (R34MH125706) from the National Institute of Mental Health (NIMH), National Institutes of Health (NIH) and partly supported by a grant (T32MH18869) from the NIMH, NIH and grant (K24DA39783; National Institute on Drug Abuse). Views expressed in this article do not necessarily reflect those of the funding agencies acknowledged. We acknowledge and thank our CAB members for their input and guidance across multiple aspects of our research program. No artificial intelligence (AI) tools were used in any portion of the manuscript.

Data Availability

The data generated for this study are not publicly available due to the sensitive nature of qualitative data and collection procedures. Recordings and transcripts of interviews are not publicly available given the confidentiality agreements with participants that were reviewed upon study enrollment.

Conflicts of Interest

PR receives royalties from Guilford Publications and conducts workshops on cognitive processing therapy. SS receives royalties from Oxford University Press, Guildford Publications, and Springer Humana Press as an author of books related to cognitive behavioral therapy, health psychology, and LGBT (lesbian, gay, bisexual, and transgender)-affirmative psychotherapy. The authors declare that they have no other known competing financial interest.

References

XSL•FO
RenderX

1. HIV surveillance report, 2019.: Centers for Disease Control and Prevention URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-2018-updated-vol-32.pdf [accessed 2024-11-25]
2. Safren SA, Gershuny BS, Hendriksen E. Symptoms of posttraumatic stress and death anxiety in persons with HIV and medication adherence difficulties. AIDS Patient Care STDS 2003;17(12):657-664 [FREE Full text] [doi: 10.1089/108729103771928717] [Medline: 14746659]
3. McLean CP, Fitzgerald H. Treating posttraumatic stress symptoms among people living with HIV: a critical review of intervention trials. Curr Psychiatry Rep 2016;18(9):83 [FREE Full text] [doi: 10.1007/s11920-016-0724-z] [Medline: 27439305]
4. Sherr L, Nagra N, Kulubya G, Catalan J, Clucas C, Harding R. HIV infection associated post-traumatic stress disorder and post-traumatic growth--a systematic review. Psychol Health Med 2011;16(5):612-629 [FREE Full text] [doi: 10.1080/13548506.2011.579991] [Medline: 21793667]
5. Pantalone DW, Rood BA, Morris BW, Simoni JM. A systematic review of the frequency and correlates of partner abuse in HIV-infected women and men who partner with men. J Assoc Nurses AIDS Care 2014;25(1 Suppl):S15-S35 [FREE Full text] [doi: 10.1016/j.jana.2013.04.003] [Medline: 24070646]
6. Kronish IM, Edmondson D, Li Y, Cohen BE. Post-traumatic stress disorder and medication adherence: results from the mind your heart study. J Psychiatr Res 2012;46(12):1595-1599 [FREE Full text] [doi: 10.1016/j.jpsychires.2012.06.011] [Medline: 22809686]
7. Hou J, Fu J, Meng S, Jiang T, Guo C, Wu H, et al. Posttraumatic stress disorder and nonadherence to treatment in people living with HIV: a systematic review and meta-analysis. Front Psychiatry 2020;11:834 [FREE Full text] [doi: 10.3389/fpsyt.2020.00834] [Medline: 32973579]
8. Katz IT, Ryu AE, Onuegbu AG. Impact of HIV-related stigma on treatment adherenceystematic review and meta-synthesis. Journal of the International AIDS Society 2013;16(3):18640 [FREE Full text] [doi: 10.7448/ias.16.3.18640]
9. Global HIV & AIDS statistics—fact sheet.: UNAIDS; 2019. URL: https://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf [accessed 2024-11-25]
10. Maggiolo F, Airoldi M, Kleinloog HD, Callegaro A, Ravasio V, Arici C, et al. Effect of adherence to HAART on virologic outcome and on the selection of resistance-conferring mutations in NNRTI- or PI-treated patients. HIV Clin Trials 2007;8(5):282-292 [FREE Full text] [doi: 10.1310/hct0805-282] [Medline: 17956829]
11. Paterson DL, Swindells S, Mohr J, Brester M, Vergis EN, Squier C, et al. Adherence to protease inhibitor therapy and outcomes in patients with HIV infection. Ann Intern Med 2000;133(1):21-30 [FREE Full text] [doi: 10.7326/0003-4819-133-1-200007040-00004] [Medline: 10877736]
12. Harrigan PR, Hogg RS, Dong WWY, Yip B, Wynhoven B, Woodward J, et al. Predictors of HIV drug-resistance mutations in a large antiretroviral-naive cohort initiating triple antiretroviral therapy. J Infect Dis 2005;191(3):339-347 [FREE Full text] [doi: 10.1086/427192] [Medline: 15633092]
13. Blake Helms C, Turan JM, Atkins G, Kempf MC, Clay OJ, Raper JL, et al. Interpersonal mechanisms contributing to the association between HIV-related internalized stigma and medication adherence. AIDS Behav 2017;21(1):238-247 [FREE Full text] [doi: 10.1007/s10461-016-1320-2] [Medline: 26864692]
14. Lima VD, Harrigan RA, Bangsberg DR, Hogg RS, Gross R, Yip B, et al. The combined effect of modern highly active antiretroviral therapy regimens and adherence on mortality over time. J Acquir Immune Defic Syndr 2009;50(5):529-536 [FREE Full text] [doi: 10.1097/QAI.0b013e31819675e9] [Medline: 19223785]
15. Earnshaw VA, Chaudoir SR. From conceptualizing to measuring HIV stigma: a review of HIV stigma mechanism measures. AIDS Behav 2009;13(6):1160-1177 [FREE Full text] [doi: 10.1007/s10461-009-9593-3] [Medline: 19636699]
16. Bogart LM, Wagner GJ, Galvan FH, Klein DJ. Longitudinal relationships between antiretroviral treatment adherence discrimination due to HIV-serostatus, race,sexual orientation among African?American men with HIV. Annals of Behavioral Medicine 2010:184-190 [FREE Full text] [doi: 10.1007/s12160-010-9200-x]
17. Turan JM, Elafros MA, Logie CH, Banik S, Turan B, Crockett KB, et al. Challenges and opportunities in examining and addressing intersectional stigma and health. BMC Med 2019;17(1):7 [FREE Full text] [doi: 10.1186/s12916-018-1246-9] [Medline: 30764816]
18. Baugher AR, Beer L, Fagan JL, Mattson CL, Freedman M, Skarbinski J, et al. Prevalence of internalized HIV-related stigma among HIV-infected adults in care, United States, 2011-2013. AIDS Behav 2017;21(9):2600-2608 [FREE Full text] [doi: 10.1007/s10461-017-1712-y] [Medline: 28213821]
19. Sevelius JM, Patouhas E, Keatley JG, Johnson MO. Barriers and facilitators to engagement and retention in care among transgender women living with human immunodeficiency virus. Ann Behav Med 2014;47(1):5-16 [FREE Full text] [doi: 10.1007/s12160-013-9565-8] [Medline: 24317955]
20. Bunn JY, Solomon SE, Miller C, Forehand R. Measurement of stigma in people with HIV: a reexamination of the HIV Stigma Scale. AIDS Educ Prev 2007;19(3):198-208 [FREE Full text] [doi: 10.1521/aeap.2007.19.3.198] [Medline: 17563274]
21. Gesesew HA, Tesfay Gebremedhin A, Demissie TD, Kerie MW, Sudhakar M, Mwanri L. Significant association between perceived HIV related stigma and late presentation for HIV/AIDS care in low and middle-income countries: A systematic review and meta-analysis. PLoS One 2017;12(3):e0173928 [FREE Full text] [doi: 10.1371/journal.pone.0173928] [Medline: 28358828]

XSL•FO
RenderX

22. Gonzalez A, Locicero B, Mahaffey B, Fleming C, Harris J, Vujanovic AA. Internalized HIV stigma and mindfulness: associations with PTSD symptom severity in trauma-exposed adults with HIV/AIDS. Behav Modif 2016;40(1-2):144-163 [FREE Full text] [doi: 10.1177/0145445515615354] [Medline: 26584609]
23. Meyer IH. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. Psychol Bull 2003;129(5):674-697 [FREE Full text] [doi: 10.1037/0033-2909.129.5.674] [Medline: 12956539]
24. Spinhoven P, Drost J, van Hemert B, Penninx BW. Common rather than unique aspects of repetitive negative thinking are related to depressive and anxiety disorders and symptoms. J Anxiety Disord 2015;33:45-52 [FREE Full text] [doi: 10.1016/j.janxdis.2015.05.001] [Medline: 26004746]
25. Diamond LM, Alley J. Rethinking minority stress: a social safety perspective on the health effects of stigma in sexually-diverse and gender-diverse populations. Neurosci Biobehav Rev 2022;138:104720 [FREE Full text] [doi: 10.1016/j.neubiorev.2022.104720] [Medline: 35662651]
26. Resick PA, Nishith P, Weaver TL, Astin MC, Feuer CA. A comparison of cognitive-processing therapy with prolonged exposure and a waiting condition for the treatment of chronic posttraumatic stress disorder in female rape victims. J Consult Clin Psychol 2002;70(4):867-879 [FREE Full text] [doi: 10.1037//0022-006x.70.4.867] [Medline: 12182270]
27. Safren SA, Otto M, Worth JL, Salomon E, Johnson W, Mayer K, et al. Two strategies to increase adherence to HIV antiretroviral medication: life-steps and medication monitoring. Behav Res Ther 2001;39(10):1151-1162 [FREE Full text] [doi: 10.1016/s0005-7967(00)00091-7] [Medline: 11579986]
28. Wingood GM, DiClemente RJ. The ADAPT-ITT model: a novel method of adapting evidence-based HIV Interventions. J Acquir Immune Defic Syndr 2008;47(1):S40-S46. [doi: 10.1097/QAI.0b013e3181605df1] [Medline: 18301133]
29. Rueda S, Park-Wyllie LY, Bayoumi A, Tynan AM, Antoniou TA, Rourke SB, et al. Patient support and education for promoting adherence to highly active antiretroviral therapy for HIV/AIDS. Cochrane Database Syst Rev 2006;2006(3):CD001442 [FREE Full text] [doi: 10.1002/14651858.CD001442.pub2] [Medline: 16855968]
30. Koenen KC, Ratanatharathorn A, Ng L. Posttraumatic stress disorder in the world mental health surveys. Psychological Medicine 2017;47(13):2260-2274 [FREE Full text] [doi: 10.1017/9781316336168.011]
31. Kalichman S, Kalichman MO, Cherry C. Medication beliefs and structural barriers to treatment adherence among people living with HIV infection. Psychol Health 2016;31(4):383-395 [FREE Full text] [doi: 10.1080/08870446.2015.1111371] [Medline: 26534788]
32. Breet E, Kagee A, Seedat S. HIV-related stigma and symptoms of post-traumatic stress disorder and depression in HIV-infected individuals: does social support play a mediating or moderating role? AIDS Care 2014;26(8):947-951 [FREE Full text] [doi: 10.1080/09540121.2014.901486] [Medline: 24666226]
33. Wingood GM, DiClemente RJ, Villamizar K, Er DL, DeVarona M, Taveras J, et al. Efficacy of a health educator–delivered HIV prevention intervention for latina women: a randomized controlled trial. Am J Public Health 2011;101(12):2245-2252 [FREE Full text] [doi: 10.2105/ajph.2011.300340]
34. Olatunji BO, Mimiaga MJ, O'Cleirigh C, Safren SA. Review of treatment studies of depression in HIV. Top HIV Med 2006;14(3):112-124. [Medline: 16946456]
35. Karlin BE, Cross G. From the laboratory to the therapy room: national dissemination and implementation of evidence-based psychotherapies in the U.S. department of veterans affairs health care system. Am Psychol 2014;69(1):19-33. [doi: 10.1037/a0033888] [Medline: 24001035]
36. Monson CM, Langkaas TF, Schnurr PP. Dissemination and implementation of evidence-based practices for posttraumatic stress disorder in post-9/11 veterans. Clinical Psychology: Science and Practice 2020;27(3):2329.
37. Chard KM. An evaluation of cognitive processing therapy for the treatment of posttraumatic stress disorder related to childhood sexual abuse. Journal of Consulting and Clinical Psychology 2005;73(5):965-971 [FREE Full text] [doi: 10.1037/0022-006x.73.5.965]
38. Monson CM, Fredman SJ, Adair KC. Cognitive-behavioral conjoint therapy for posttraumatic stress disorder: application to operation enduring and iraqi freedom veterans. J Clin Psychol 2008;64(8):958-971. [doi: 10.1002/jclp.20511] [Medline: 18613094]
39. Galovski TE, Blain LM, Mott JM, Elwood L, Houle T. Manualized therapy for PTSD: flexing the structure of cognitive processing therapy. J Consult Clin Psychol 2012;80(6):968-981 [FREE Full text] [doi: 10.1037/a0030600] [Medline: 23106761]
40. Nishith P, Nixon RDV, Resick PA. Resolution of trauma-related guilt following treatment of PTSD in female rape victims: a result of cognitive processing therapy targeting comorbid depression? J Affect Disord 2005;86(2-3):259-265 [FREE Full text] [doi: 10.1016/j.jad.2005.02.013] [Medline: 15935245]
41. Sobel AA, Resick PA, Rabalais AE. The effect of cognitive processing therapy on cognitions: impact statement coding. J Trauma Stress 2009;22(3):205-211 [FREE Full text] [doi: 10.1002/jts.20408] [Medline: 19479979]
42. Resick PA, Monson CM, Chard KM. Cognitive Processing Therapy for PTSD: A Comprehensive Manual. New York: Guilford Press; 2017.

XSL•FO
RenderX

43. Pearson CR, Kaysen D, Huh D, Bedard-Gilligan M. Randomized control trial of culturally adapted cognitive processing therapy for PTSD substance misuse and HIV Sexual Risk Behavior for Native American Women. AIDS Behav 2019;23(3):695-706 [FREE Full text] [doi: 10.1007/s10461-018-02382-8] [Medline: 30607757]
44. Lopez CM, Baker NL, Moreland AD, Bisca E, Wilson T, Slick N, et al. Development and feasibility testing of an integrated PTSD and adherence intervention cognitive processing therapy-life steps (CPT-L) to improve HIV outcomes: trial protocol. Contemp Clin Trials Commun 2023;33:101150 [FREE Full text] [doi: 10.1016/j.conctc.2023.101150] [Medline: 37273831]
45. Corbin JM, Strauss A. Grounded theory research: procedures, canons, and evaluative criteria. Qualitative Sociology 1990:3-21 [FREE Full text] [doi: 10.1007/bf00988593]
46. Naeem M, Ozuem W, Howell K, Ranfagni S. A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. International Journal of Qualitative Methods 2023;22(11) [FREE Full text] [doi: 10.1177/16094069231205789]
47. Mwita KM. Factors influencing data saturation in qualitative studies. IJRBS 2022;11(4):414-420. [doi: 10.20525/ijrbs.v11i4.1776]
48. Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence.: World Health Organization; 2013. URL: https://www.who.int/publications/i/item/9789241564625 [accessed 2024-11-24]
49. Anderson JC, Campbell JC, Glass NE, Decker MR, Perrin N, Farley J. Impact of intimate partner violence on clinic attendance, viral suppression and CD4 cell count of women living with HIV in an urban clinic setting. AIDS Care 2018 Apr;30(4):399-408 [FREE Full text] [doi: 10.1080/09540121.2018.1428725] [Medline: 29397777]
50. Brief DJ, Bollinger AR, Vielhauer MJ, Berger-Greenstein JA, Morgan EE, Brady SM, et al. Understanding the interface of HIV, trauma, post-traumatic stress disorder, and substance use and its implications for health outcomes. AIDS Care 2004;16(1):S97-120 [FREE Full text] [doi: 10.1080/09540120412301315259] [Medline: 15736824]
51. Gore-Felton C, Koopman C. Traumatic experiences: harbinger of risk behavior among HIV-positive adults. Journal of Trauma & Dissociation 2002;3(4):121-135 [FREE Full text] [doi: 10.1300/j229v03n04_07]
52. LeGrand S, Reif S, Sullivan K, Murray K, Barlow ML, Whetten K. A review of recent literature on trauma among individuals living with HIV. Curr HIV/AIDS Rep 2015;12(4):397-405 [FREE Full text] [doi: 10.1007/s11904-015-0288-2] [Medline: 26419376]
53. Leserman J. Role of depression, stress, and trauma in HIV disease progression. Psychosom Med 2008;70(5):539-545. [doi: 10.1097/PSY.0b013e3181777a5f] [Medline: 18519880]
54. Mugavero MJ, Pence BW, Whetten K, Leserman J, Swartz MJ, Stangl D, et al. Predictors of AIDS-related morbidity and mortality in a southern U.S. cohort. AIDS Patient Care STDs 2007;21(9):681-690 [FREE Full text] [doi: 10.1089/apc.2006.0167] [Medline: 17919095]
55. Remien RH, Stirratt MJ, Nguyen N, Robbins RN, Pala AN, Mellins CA. Mental health and HIV/AIDS: the need for an integrated response. AIDS 2019;33(9):1411-1420 [FREE Full text] [doi: 10.1097/QAD.0000000000002227] [Medline: 30950883]
56. Hilton N, McKee S, Ham E, Green M, Wright L. Co-occurring mental illness and substance use disorders in canadian forensic inpatients: underdiagnosis and implications for treatment planning. International Journal of Forensic Mental Health 2018;17(2):145-153 [FREE Full text] [doi: 10.1080/14999013.2018.1451416]
57. Resick PA, Schnicke MK. Cognitive processing therapy for sexual assault victims. J Consult Clin Psychol 1992;60(5):748-756 [FREE Full text] [doi: 10.1037//0022-006x.60.5.748] [Medline: 1401390]
58. Kalichman SC, Pope H, White D, Cherry C, Amaral CM, Swetzes C, et al. Association between health literacy and HIV treatment adherence: further evidence from objectively measured medication adherence. J Int Assoc Physicians AIDS Care (Chic) 2008;7(6):317-323 [FREE Full text] [doi: 10.1177/1545109708328130] [Medline: 19056866]
59. Balfour L, Kowal J, Silverman A, Tasca GA, Angel JB, Macpherson PA, et al. A randomized controlled psycho-education intervention trial: Improving psychological readiness for successful HIV medication adherence and reducing depression before initiating HAART. AIDS Care 2006;18(7):830-838 [FREE Full text] [doi: 10.1080/09540120500466820] [Medline: 16971295]
60. Altice FL, Mostashari F, Friedland GH. Trust and the acceptance of and adherence to antiretroviral therapy. J Acquir Immune Defic Syndr 2001;28(1):47-58 [FREE Full text] [doi: 10.1097/00042560-200109010-00008] [Medline: 11579277]
61. Hays PA. Addressing cultural complexities in practice: A framework for clinicians and counselors. American Psychological Association 2009;40(4):354-360. [doi: 10.1037/a0016250]
62. Machtinger EL, Lavin SM, Hilliard S, Jones R, Haberer JE, Capito K, et al. An expressive therapy group disclosure intervention for women living with HIV improves social support, self-efficacy, and the safety and quality of relationships: a qualitative analysis. J Assoc Nurses AIDS Care 2015;26(2):187-198 [FREE Full text] [doi: 10.1016/j.jana.2014.05.001] [Medline: 25027284]
63. Jaffe AE, DiLillo D, Hoffman L, Haikalis M, Dykstra RE. Does it hurt to ask? A meta-analysis of participant reactions to trauma research. Clin Psychol Rev 2015;40:40-56 [FREE Full text] [doi: 10.1016/j.cpr.2015.05.004] [Medline: 26051308]

XSL•FO
RenderX

Abbreviations

ADAPT-ITT: Assessment, Decision, Adaptation, Production, Topical experts-integration, Training, and Testing
ART: antiretroviral therapy
CAB: Community Advisory Board
CPT: cognitive processing therapy
CPT-L: CPT-Life-Steps for integration of adherence
DSM-V: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
DSM-5-TR: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision
Life-Steps: Life-Steps for Medication Adherence
MoCA: Montreal Cognitive Assessment
PTSD: posttraumatic stress disorder
PWH: persons with HIV
REDCap: Research Electronic Data Capture

Edited by A Mavragani; submitted 12.07.24; peer-reviewed by MKS Smith, JI Campbell; comments to author 02.09.24; revised version received 23.09.24; accepted 27.09.24; published 16.01.25.

Please cite as:
Lopez CM, Moreland AD, Amaya S, Bisca E, Mujica C, Wilson T, Baker N, Richey L, Eckard AR, Resick PA, Safren SA, Danielson CK
Assessment, Decision, Adaptation, Production, Topical Experts-Integration, Training, and Testing (ADAPT-ITT) Framework to Tailor Evidence-Based Posttraumatic Stress Disorder Treatment for People With HIV to Enhance Engagement and Adherence: Qualitative Results from a Feasibility Randomized Controlled Trial
JMIR Form Res 2025;9:e64258
URL: https://formative.jmir.org/2025/1/e64258
doi:10.2196/64258
PMID:

©Cristina M Lopez, Angela D Moreland, Stephanie Amaya, Erin Bisca, Christin Mujica, Tayler Wilson, Nathaniel Baker, Lauren Richey, Allison Ross Eckard, Patricia A Resick, Steven A Safren, Carla Kmett Danielson. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Designing the First Pregnancy Guaranteed Income Program in the United States: Qualitative Needs Assessment and Human-Centered Design to Develop the Abundant Birth Project

Deborah Karasek[1,2], MPH, PhD; Jazzmin C Williams[3], MD; Michaela A Taylor[4], MPH; Monica M De La Cruz[5], MPH; Stephanie Arteaga[5], MPH; Sabra Bell[6]; Esperanza Castillo[3], MS; Maile A Chand[4,6], BA; Anjeanette Coats[5]; Erin M Hubbard[5], MPH; Latriece Love-Goodlett[4], BA; Breezy Powell[4,6]; Solaire Spellen[7], MPH; Zea Malawa[3,4], MPH, MD; Anu Manchikanti Gomez[5], MSc, PhD

[1]School of Public Health, Oregon Health & Science University and Portland State University, Portland, OR, United States

[2]Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California, San Francisco, San Francisco, United States

[3]University of California, San Francisco, San Francisco, CA, United States

[4]Expecting Justice, San Francisco, CA, United States

[5]Sexual Health and Reproductive Equity Program, School of Social Welfare, University of California, Berkeley, Berkeley, CA, United States

[6]Department of Human Ecology, Univeristy of California, Davis, Davis, CA, United States

[7]California Preterm Birth Initiative, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Anu Manchikanti Gomez, MSc, PhD
Sexual Health and Reproductive Equity Program
School of Social Welfare
University of California, Berkeley
110 Haviland Hall, MC 7400
Berkeley, CA, 94720-7400
United States
Phone: 1 510 6420722
Email: anugomez@berkeley.edu

Abstract

Background: Racial inequities in pregnancy outcomes persist despite investments in clinical, educational, and behavioral interventions, indicating that a new approach is needed to address the root causes of health disparities. Guaranteed income during pregnancy has the potential to narrow racial health inequities for birthing people and infants by alleviating financial stress.

Objective: We describe community-driven formative research to design the first pregnancy-guaranteed income program in the United States—the Abundant Birth Project (ABP). Informed by birth equity and social determinants of health perspectives, ABP targets upstream structural factors to improve racial disparities in maternal and infant health.

Methods: The research team included community researchers, community members with lived experience as Black or Pacific Islander pregnant, and parenting people in the San Francisco Bay Area. The team conducted needs assessment interviews and facilitated focus groups with participants using human-centered design methods. Needs assessment participants later served as co-designers of the ABP program and research, sharing their experiences with financial hardships and government benefits programs and providing recommendations on key program elements, including fund disbursement, eligibility, and amount.

Results: Housing affordability and the high cost of living in San Francisco emerged as significant sources of stress in pregnancy. Participants reported prohibitively low income eligibility thresholds and burdensome enrollment processes as challenges or barriers to existing social services. These insights guided the design of prototypes of ABP's program components, which were used in a design sprint to determine the final components. Based on this design process, the ABP program offered US $1000/month for 12 months to pregnant Black and Pacific Islander people, selected through a lottery called an abundance drawing.

Conclusions: The formative design process maximized community input and shared decision-making to co-design a guaranteed income program for Black and Pacific Islander women and people. Our upstream approach and community research model can inform the development of public health and social service programs.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e60829) doi:10.2196/60829

KEYWORDS

maternal and child health; economics; public health; qualitative research methods; programs (evaluation and funding); community-centered; pregnancy; first pregnancy; behavioral interventions; racial health; financial stress; Abundant Birth Project; infant health; infant; Black

Introduction

Background

Pregnancy is a critical period of development when stress can have deleterious effects on both the pregnant person and their baby, leading to adverse birth and long-term health outcomes [1,2] that disproportionately affect Black, Indigenous, and Native Hawaiian or Pacific Islander people [3,4]. In the United States, Black infants are twice as likely to be born preterm or with low birth weight compared to White infants [3]. These inequities extend to maternal health, with Black women being 3.3 times more likely to die during or after pregnancy than White women [5]. While rates of adverse outcomes are often obscured by aggregation of data among Asian and Pacific Islander subgroups [6], Native Hawaiian and other Pacific Islander people experience higher rates of low birth weight, preterm birth, and macrosomia [3,7,8]. Even in San Francisco, a wealthy, well-resourced city, the shrinking Black and Pacific Islander populations experience significant inequities in birth outcomes [9]. Black and Pacific Islander pregnant women and people experienced the highest rates of preterm birth, at 12.7% and 8.9%, respectively, in 2014-2016, the last years with enough births to Pacific Islander people to reliably report [10]. Research increasingly indicates that racism—structural, institutional, and interpersonal—is a key underlying driver of these inequities [11,12]. It is unsurprising, therefore, that clinical, educational, and behavioral interventions have failed to meaningfully narrow racialized birth disparities [3,13-15].

Efforts to reduce inequities must address exposure to interpersonal, institutional, and structural racism as the root cause of disproportionate risk [16-19]. Racial inequity in income and wealth is one manifestation of structural racism, with important public health ramifications [16,20]. The racial wealth gap has resulted from a long history of exclusion of Black families from economic opportunity, through a combination of redlining, exclusionary zoning, subprime loans, and "urban renewal" programs, among other tactics [21]. This leaves Black families and other families of color more vulnerable to economic stress during pregnancy. Financial stress may be particularly acute during pregnancy, a period of income volatility when families frequently experience changes in housing, employment, health care, and childcare costs [22]. Indeed, in a nationally representative study of pregnant and postpartum women, over half reported general financial stress [23]. Yet, efforts to directly alleviate poverty and financial stress to address birth inequities are limited. In San Francisco, the median annual income in 2020 for White households was US $146,043, while it was just US $38,862 for Black households and US $90,917 for Pacific Islander households [24], making it the city with the third largest income disparity in the United States [25]. The Insight Center Family Needs calculator estimated the minimum income required to live in San Francisco in 2021 without public or private assistance as US $60,232 for a single adult, US $120,519 for a single adult with 1 preschooler, and US $167,432 for a household with 2 adults, 1 infant, and 1 preschooler [26]. To meet the needs of a family of three, Black and Pacific Islander households would have needed to earn 4 and 2 times their current average incomes, respectively.

Guaranteed income describes the disbursement of unconditional cash to individuals or families in a community without restrictions, allowing flexibility to close gaps where needs are not being met by public benefits programs [27]. Guaranteed income programs offer the potential to address poverty and stress caused by economic racism and the racial wealth gap. Philanthropic funders, cities, and states, including California, are increasingly investing in guaranteed income as a poverty reduction and equity-generating strategy [28]. Emerging findings from the growing research on United States guaranteed income pilots have found that monthly income supplements improve the sense of trust in institutions, dignity, and autonomy and increase savings, without changing spending patterns [29,30]. Magnolia Mothers Trust, a program providing US $1000 per month for 12 months to Black mothers living in federally subsidized housing in Jackson, Mississippi, allowed mothers to save for emergencies, buy food, pay childcare expenses, and prioritize their health [30].

At the time of the Abundant Birth Project's (ABP) inception, there were no guaranteed income programs in the United States that addressed the critical period of pregnancy. However, evidence from other contexts supported a guaranteed income in pregnancy as a health intervention. A quasi-experimental study among 14,591 pregnant women in Manitoba, Canada, found that an approximately 10% increase in income from a pregnancy income benefit resulted in reductions of low birthweight by 21% and preterm birth by 18% [31,32]. Since the launch of ABP, there has been a proliferation of guaranteed income programs across California, including other programs that target pregnancy [28,33,34].

The Abundant Birth Project

ABP is a program of Expecting Justice, a Black-led, cross-sector collaborative focused on birth equity in San Francisco that includes mothers, public health professionals, health and social systems partners, and governmental leaders. Expecting Justice uses a collective impact model, which is a process to bring together diverse stakeholders to create a common agenda, establish shared measurement, and foster mutually reinforcing activities for stakeholders [35]. As part of its work to advance birth equity, Expecting Justice is guided by community voice and emphasizes upstream solutions that address the social determinants of health. As defined by the National Birth Equity Collaborative, birth equity is the assurance of optimal birth

XSL•FO
RenderX

conditions for all people, with an ongoing effort to address racial and social inequities [36]. The social determinants of health result from the larger systems at play, such as economic policies, racism, and political systems that shape the conditions of daily life and health [37,38]. These frameworks informed the focus on this pregnancy-guaranteed income program as an equity-generating strategy to directly address income insecurity as a social determinant of health by interrupting the pathway from economic exclusion and stress to adverse birth outcomes. Expecting Justice targeted this intervention toward communities most likely to experience preterm birth to help narrow disparities in preterm birth rates. Focusing on the most marginalized aligns with the birth equity framework, emphasizing that targeted support, rather than universal availability, is essential for addressing birth inequities [36].

Expecting Justice entered a community-academic partnership with the University of California, Berkeley, and the University of California, San Francisco, to design and evaluate ABP. We developed the ABP program and evaluation by actively, meaningfully, and intentionally centering the voices and experiences of the communities most affected by the adverse birth outcomes the program seeks to improve. ABP launched in June 2021, with the goal of enrolling 150 Black and Pacific Islander pregnant women and people (including gender-diverse individuals with the capacity for pregnancy) over 18 months.

This paper describes the formative research process, under the co-leadership of community members, used to design ABP.. Our primary objectives were to (1) determine the financial needs of Black and Pacific Islander women and people during pregnancy, and (2) collaboratively design a pregnancy-guaranteed income pilot program with Black and Pacific Islander mothers. We hope that the process and results from the formative research to design ABP will be useful to other researchers, practitioners, and community groups developing interventions to address racial health disparities, including guaranteed income pilot programs.

Methods

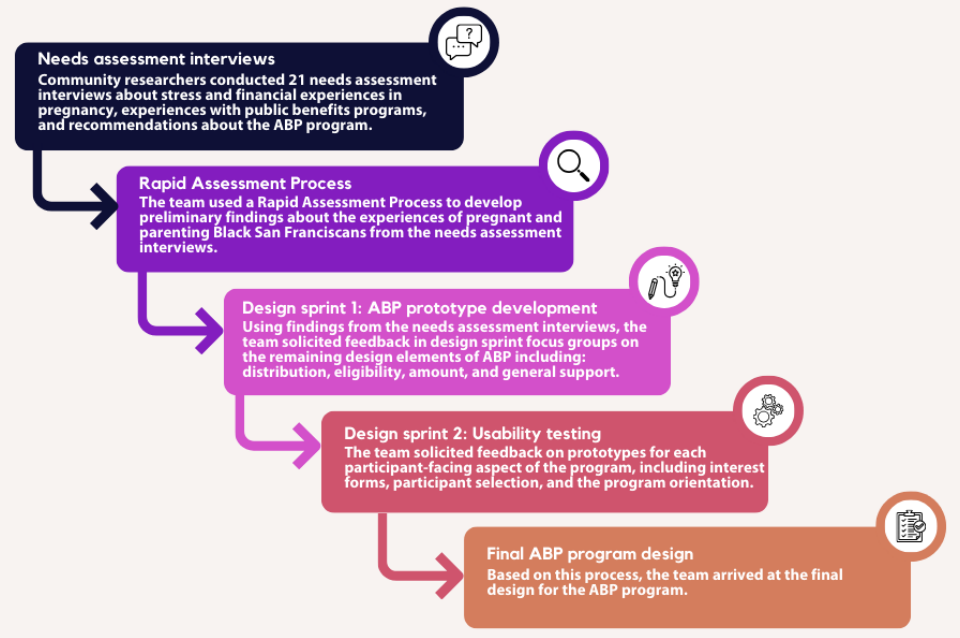
Community Research Model

ABP used shared decision-making and community-centered practices in all phases of the program development and evaluation, by purposely centering the voices of community members with lived expertise [39]. As a community-based participatory research project, ABP recognized that this relationship is dynamic and requires continual effort to rebalance traditional power structures [40]. To support these efforts, 4 community researchers (CRs) served as key members of the community-academic partnership. The CRs are Black and Pacific Islander community members who have lived experience of pregnancy and parenting in the San Francisco Bay Area. They come from a variety of professional backgrounds and have an interest in learning or experience using research methods to advance birth equity. We adopted a shared power structure and decision-making process that included following Expecting Justice's "expert opinion protocol" for all final decisions. This protocol enabled ABP stakeholders who identified as Black and/or Pacific Islander mothers to have protected time to share their insights with the larger group without contradiction before voting. We also used weighted voting in decision-making to prioritize the perspectives of those who identified as Black and/or Pacific Islander mothers.

The process of designing the ABP program and evaluation involved (1) conducting interviews as part of the needs assessment, (2) a Rapid Assessment Process (RAP) for analysis, (3) a prototype development design sprint, and (4) a usability testing design sprint (Figure 1). The CRs contributed to the full process, bringing expertise from their lived experience with pregnancy and parenting in the Black and Pacific Islander communities: developing research questions, formulating survey and interview instruments, recruiting participants, conducting interviews, facilitating design sprints, interpreting results, and making key program decisions. CRs received training in data collection and analysis supported by the university-based staff, which coupled with the team's shared power structure, helped to dismantle the hierarchy often present in research institutions that keeps community members out of decision-making processes [41,42]. Expecting Justice also led an ABP working group with birth equity stakeholders, including Black and Pacific Islander mothers, community organization representatives, social service providers, maternal health leaders and researchers, and representatives from health plans to collaborate and provide strategic resources to support program development. The ABP working group met monthly to inform key program design and implementation decisions before, during, and following the formative phase.

XSL•FO
RenderX

Figure 1. The design process for the Abundant Birth Project involved conducting needs assessment interviews, an RAP to analyze findings, a prototype development design sprint, and a usability design sprint to arrive at the final design. ABP: Abundant Birth Project.



The shared decision-making structure with the ABP program and research teams, CRs, and ABP working group led to tangible improvements in research and program design. For example, the CRs voiced concerns that research studies rarely compensate participants in a way that truly values their time. Under their leadership, the team increased the amount of participant remuneration for the interviews and design sprints from US $50 to US $75 and US $100 to US $200, respectively.

Needs Assessment Interviews

To inform ABP's design, we first conducted interviews to understand the experiences Black and Pacific Islander women and people face during pregnancy in San Francisco. Inclusion criteria were aged 18 years or older, identified as Black and/or Pacific Islander, spoke English, and either currently pregnant or had been pregnant while residing in San Francisco in the past 5 years. Participants were recruited from social media and local community organizations. The CRs closely collaborated on the development of the interview guide to ensure that language and questions would capture appropriate topics in a comfortable environment for participants. We conducted interviews from April to June 2020. The CRs conducted the interviews using a web-based videoconference platform or by telephone. Interviews were recorded and professionally transcribed.

The interview guide addressed stress and financial experiences during pregnancy and experiences with and perception of public benefits programs. We also asked participants for recommendations on the ABP program design, including how much money would alleviate financial stress during pregnancy, how to best distribute the income supplement, and program eligibility criteria and restrictions.

Before conducting interviews, the CRs underwent training in qualitative research methods led by Dr Brittany Chambers, a Black scholar, mother, and community engagement expert. The training included an introduction to asking research questions, qualitative data collection, and interviewing skills. The entire research team went through the training together, building trust and rapport throughout the process. The CRs described the importance of being trained by a researcher from their own communities. Dr Chambers used art-based activities that allowed the CRs to celebrate their stories and the whole team to tap into their imagination and creativity and engage with one another outside of hierarchies. CRs received compensation for their time and expertise in the training.

About RAP

To identify key topics, we used RAP, an intensive, team-based, iterative analytic approach that uses data triangulation to develop preliminary findings from the perspective of an insider—pregnant and parenting Black and Pacific Islander San Franciscans [43]. RAP is well-suited for qualitative research informing the design of interventions, as it allows for rapid and timely data reduction and ongoing analysis during data collection. We brought together the team, including CRs, program staff, and student research assistants for shared qualitative analysis training. CRs, other research team members, and program staff used summary templates to create a memo for each interview. Members of the research team reviewed memos for consistency. The first step of data reduction involved matching domains from the summary template to interview questions. The domains included concerns about stress and pregnancy, employment, experiences with social services, desired programs and services, feelings about the income supplement, recommended supplement amount, eligibility restrictions, disbursement, recruitment recommendations, and additional financial services. Finally, we used an individual-level data matrix to synthesize key data across participants and identify themes.

XSL•FO
RenderX

Human-Centered Design Sprints

Overview

Using design thinking methodology, we conducted 2 focus groups, called human-centered design sprints, to develop and refine the facets of the guaranteed income program following themes identified from the need assessment interviews. Originating in consumer technology, design thinking is a method of developing innovative solutions and products with leadership and input from the end user [44]. A design sprint refers to a rapid, human-centered design activity, in which individuals are brought together to discuss and iterate on program design components [45]. Design sprints prioritize qualitative feedback from stakeholders and target end users, and encourage creative thinking [45]. Human-centered design methodology is especially well-suited for developing urgently needed interventions to reduce health inequities and has previously been used to gather community input on solutions for perinatal disparities [46]. We partnered with design thinking firms that helped determine the structure of the design sprints and trained the ABP program staff and CRs on how to facilitate these sessions.

Due to the COVID-19 pandemic, we conducted the design sprints using videoconferencing software and recorded the sessions. We provided tablets with internet connectivity to all design sprint participants who lacked internet access to participate virtually. For continuity, we invited participants from the needs assessment interviews to participate in the design sprints. Co-designing with the same participants also ensured that their voices and perspectives were reflected in the program prototypes. Each design sprint session was 4 hours.

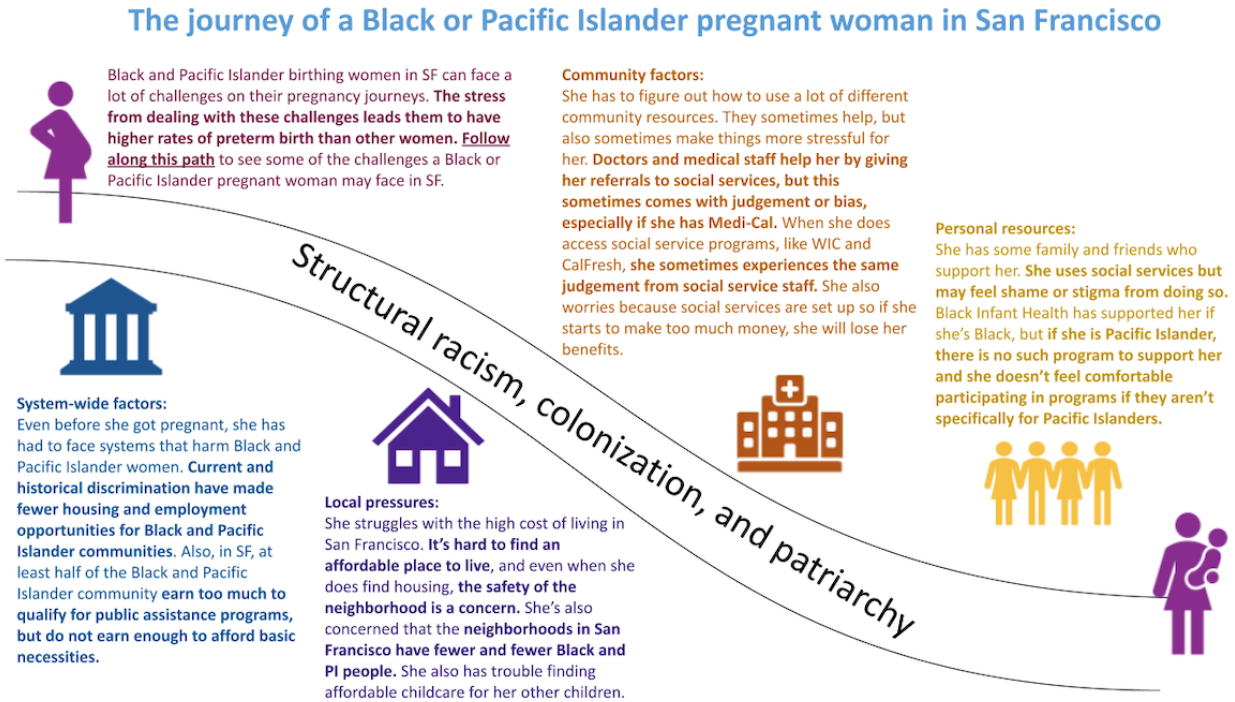
Design Sprint 1: ABP Prototype Development

Guided by findings from the needs assessment interviews, our team determined 4 remaining design questions to be answered through the first design sprint.

1. Distribution: How can we best deliver money in a way that is both inclusive and convenient?
2. Eligibility: How can we best serve the Black and Pacific Islander communities in San Francisco while allowing for flexibility in eligibility criteria?
3. Amount: How can we balance financially supporting pregnancy needs with the desire to support as many people as possible?
4. General support: How can we support pregnant people, not just financially, but holistically, during their pregnancy?

Before the design sprint, we provided participants with a welcome packet to introduce ABP and the purpose of the session. The welcome packet included a journey map to contextualize the human-centered design sprints for participants (Figure 2). The journey map was developed using common themes from the needs assessment interviews to guide participants through a story of a Black or Pacific Islander pregnant woman in San Francisco, focusing on how racism contributes to financial and housing challenges, stress, and stigma.

Figure 2. The journey map was developed using themes from needs assessment interviews conducted from April to June 2020 for the Abundant Birth Project. It described how racism contributes to financial and housing challenges, stress, and stigma for Black and Pacific Islander pregnant women in San Francisco. SF: San Francisco; PI: Pacific Islander; WIC: Women, Infant and Children.



Following the first design sprint, we synthesized feedback from participants with findings from the needs assessment interviews to develop ABP program prototypes. Specific recommendations from participants in interviews and the first design sprint informed the prototypes for key components of ABP, including distribution method, amount, and eligibility criteria. The themes surrounding previous experiences with social services informed

XSL•FO
RenderX

components of the ABP prototype that focused on support that should be offered in addition to the supplement itself.

Design Sprint 2: Usability Testing

Using the learnings from the first design sprint, our team developed prototypes for each participant-facing aspect of the program, including the interest form, eligibility interview, participant selection, program enrollment, orientation, program experience, and research engagement. Design thinking is iterative; therefore, the second design sprint focused on usability testing with the ABP prototypes. Participants gave feedback on the prototypes, which helped ensure that the final program design aligned with community values and perspectives. CRs facilitated breakout groups that focused on each aspect of the program experience. For research engagement, participants were asked about what concerns they would have about engaging in research and what would make them excited to opt into the research. We made final decisions regarding the prototypes using Expecting Justice's "expert opinion" protocol.

Ethical Considerations

The institutional review board of the University of California, San Francisco, approved this study's protocol (IRB #19-27608, reference #267614). Participants signed consent forms for interviews and design sprints after reviewing them with study personnel. The research team removed identifiers (eg, names) from interview transcripts to deidentify data. Participants received a US $75 gift card incentive for taking part in the needs assessment interviews and a US $200 gift card incentive for participation in each design sprint.

Results

Needs Assessment Interviews

Participant Demographics

We conducted 21 in-depth interviews from April to June 2020. Seventeen participants identified as Black or African American, 5 as Pacific Islander, 2 as Latinx, and 1 as Native American or Indigenous (Table 1). All 21 participants identified as women. Eleven participants had experienced homelessness in the past, and 15 were at risk of food insecurity [47].

Table 1. Selected demographics for Abundant Birth Project formative research participants in San Francisco, California. Formative research needs assessment interviews and design sprints occurred from April to December 2020 (N=21).

Characteristic	Value, n (%)
Racial identity[a]	
Black or African American	17 (81)
Latinx	2 (10)
Native American or Indigenous	1 (5)
Pacific Islander	5 (24)
Pregnancy status	
Pregnant at the time of interview	6 (29)
Pregnant in the last 5 years	15 (71)
Income	
100% federal poverty level	7 (33)
100%-200% federal poverty level	6 (29)
>200% federal poverty level	6 (29)
Do not know	2 (9)

[a]Racial identity categories are not mutually exclusive.

Experiences of Financial Stress in Pregnancy

All participants discussed feeling stressed during their pregnancies; sources included negotiating employment leave, the health of their baby and themselves, and most frequently, housing and the high cost of living in San Francisco. Some participants reported being housing insecure or homeless during their pregnancies, underscoring the precarity of meeting basic needs in San Francisco. One participant described how the cost of living in the Bay Area was a primary driver of her stress.

> *Especially living in San Francisco Bay Area, everyone is so uptight and on edge because it's so expensive to live here, and we're always stressed-out about working or money and just trying to make it, so with all of that already going on top of bringing another life into the world, things can get really, really stressful.* [Currently pregnant, identified as Black and Latinx]

Another participant discussed her distressing personal experience with homelessness and the tradeoffs she had to consider to keep her child safe.

> *I was couch surfing. I was homeless. I became homeless March of 2019, so I was couch surfing waiting for a shelter to come about or some type of services, so I've been just couch surfing, sleeping on*

XSL•FO
RenderX

> *the…well, I will sleep at the bus stop, but my son would be with one of my friends during the night, so that he would [be] able to sleep comfortably.* [Last pregnant in 2020, identified as Pacific Islander]

Participants discussed the need for assistance programs that could include middle-income families. One participant underscored this by saying:

> *The hard part is being in this random middle class just because of my husband's one job, even though nobody's taking into account all of our expenses and things. And so, it's still kind of like we're low-income, but it doesn't read low-income on paper.* [Last pregnant in 2018, identified as Black, Latinx, and Native American or Indigenous]

Other participants who considered themselves middle-income echoed this "catch-22" situation, in which they struggled financially because they did not make enough to meet the costs of living in the San Francisco Bay Area, but their income surpassed the federal poverty line used to calculate public benefit eligibility, rendering them ineligible for assistance. Describing this impossible situation, one participant said:

> *Just because someone has a job doesn't mean that they shouldn't not qualify and get this extra help, because they still have stress. They still are worried about something financially, you know…what I'm saying is like most programs if you make less than $35,000 a year. Well, that's poor, poor, poor in San Francisco. But what about like the person who makes $32 an hour, you know, that can't make ends meet or is stressing and struggling. Like there's never any kind of help or support for them. They can't get the Medi-Cal. They can't get the WIC. No welfare. Children's Council. None of that.* [Most recently pregnant in 2019, identified as Black]

Experiences With Government Benefits Programs

Most participants (18/21) had experiences with government benefits programs, including CalFresh (California's Supplemental Nutrition Assistance Program), CalWorks (California's Temporary Assistance to Needy Families program, commonly known as "welfare"), Medi-Cal (California's Medicaid program), the Special Supplemental Nutrition Program from Women, Infant and Children, and Black Infant Health, a statewide program to improve the health of Black mothers and babies. Participants found most government benefits programs financially helpful and were grateful for the aid. One participant said:

> *CalFresh helped a lot. It was a lifesaver, actually. We were just so broke at that time and paying San Francisco rent and everything, so that definitely helped us get by until I got a more full-time job, and things changed after that, but it definitely got us through.* [Currently pregnant, identified as Pacific Islander]

Despite appreciating the financial assistance, participants found the application processes for many benefits programs burdensome and income eligibility requirements too restrictive, given the cost of living in San Francisco. Additionally, participants reported feeling stigmatized by the benefits program staff or when using certain resources, such as a CalFresh Electronic Benefit Transfer card at a grocery store or Medi-Cal at their health care provider. Describing previous stigmatizing experiences while using Electronic Benefit Transfer at the grocery store, one participant said:

> *I feel like you're looked at differently, you're really treated differently. Like, I would have people, like, cashiers, kind of put me to the side like, 'Oh, figure out your check stuff, and then let me get these people that are paying.' Like, I've had people say that to me, and so that part of it is just like—and then, you know, [the cashiers] getting up on a loudspeaker—just it just, all that stuff, nobody's discreet about anything…* [Most recently pregnant in 2016, identified as Black]

Recommendations for a Pregnancy Guaranteed Income Program

Overall, participants enthusiastically supported the idea of an income supplement for Black and Pacific Islander people during pregnancy, feeling that the flexible resources would significantly reduce stress during pregnancy.

> *That would be so helpful and also show compassion towards the women [and] what they go through because people don't know really what we go through unless you had the baby...that money would definitely be helpful towards whether or not you pay for diapers or food, you know what I'm saying. Like sometimes you can't do both, but you need both, right?* [Most recently pregnant in 2018, identified as Pacific Islander]

We asked participants how much money per month they believed would reduce financial stress during pregnancy. Three participants felt "anything would help," 4 recommended less than US $500, 11 recommended an amount between US $700 and US $1000, and 4 believed more than US $1000 would be necessary due to the high cost of living in San Francisco. Participants recommended electronic fund distribution for speed and because many bills are paid on the web. They felt that cash may be spent too quickly, makes tracking spending difficult, and cannot be replaced if lost. Some also expressed concerns that cash could not be used for remote or web-based purchases, which was essential during the data collection period at the beginning of the COVID-19 pandemic. Many mentioned the need for an option for individuals without a bank account. When asked about potential other services that should be offered alongside ABP, most felt it would be valuable to offer optional support with finances (eg, credit building or repair, savings, banking, and taxes), childcare, and housing.

Some participants expressed concerns about how recipients would spend the income supplement. This may reflect false and racist narratives about how Black mothers use public assistance programs [48]. Often speaking in generalities, participants discussed concerns that income supplement recipients may not use the money in ways society deems "right." Participants also

XSL•FO
RenderX

reported concerns that if hypothetical recipients misused the funds, it may result in the cessation of the program.

> *Of course, we have those folks who would abuse it and maybe not really using it, well, I guess if it's no strings attached, it doesn't matter what they use it for, but I guess I would just be afraid of anything happening like that so where it would be abused to a point that we'd lose it altogether, you know what I mean, so that would be a concern.* [Most recently pregnant in 2018, identified as Pacific Islander]

Several participants discussed eligibility requirements, such as an income cap or debt-to-income ratio. These criteria were proposed to ensure that the limited resources are directed to the target population.

Design Sprint 1: ABP Prototype Development

Overview

Fifteen interviewees from the needs assessment participated in the first design sprint, which occurred in July 2020. The CRs facilitated breakout sessions to discuss the concept of an income supplement and address the 4 key program design questions.

Amount

Participants believed that a monthly supplement should be at least US $1000 to provide a noticeable benefit during pregnancy. As participants also wanted the program to accommodate as many people as possible, they preferred a supplement to reach 150 participants over a larger supplement reaching fewer people.

Disbursement

Some participants recommended multiple payment types for the disbursement of funds. As this was not feasible with the disbursement organizations, participants came to the consensus that an ABP-specific debit card (as opposed to cash, check, or direct deposit) would be the most acceptable way to distribute the supplement. Participants discussed what would make the debit card the most accessible, such as not requiring a bank account or social security number. They also expressed they wanted the ABP debit card to have a beautiful design that reflected their communities, so recipients would not feel stigmatized or judged as they often did in other benefits programs (Figure 3).

Figure 3. The Abundant Birth Project debit card included imagery that reflected participants' desire for a beautiful design to prevent feelings of stigma or judgment often associated with using other benefits programs.

Eligibility

Participants felt the program should have additional eligibility criteria aside from identifying as Black and/or Pacific Islander, such as gestational age in pregnancy and income. For income, participants expressed a desire for the program to serve individuals who do not qualify for government benefits but still struggle to meet their needs in San Francisco. Design sprint participants believed that ABP participants should be randomly selected to give everyone interested an equal chance of enrolling in the program. Importantly, participants cautioned against using the term "lottery" to describe this process, given that their communities had a history of negative experiences with lottery-based programs, such as the local public housing selection process.

Other Support

Participants believed that the program should prioritize offering participants opt-in case management support. While they felt that engaging with case management should not be required with an unconditional program, they acknowledged that people have trouble accessing existing resources and navigating the benefits system.

Design Sprint 2: Usability Testing

Eight participants from the first design sprint returned in December 2020 to test and provide feedback on all components of the ABP prototypes.

We elicited feedback on example recruitment documents. Participants preferred interest and referral forms that could be completed in 5 minutes or less. They suggested that the ABP website contain a portal to allow applicants to check their status from interest form submission to enrollment. To create a welcoming atmosphere, participants felt it was important for the website to have imagery representing Black and Pacific Islander people. This was especially important for the Pacific Islander participants because few programs tailored outreach specifically to this community. With this in mind, Expecting Justice collaborated with Black and Pacific Islander artists to design outreach materials such as flyers and social media posts. Expecting Justice consulted the San Francisco Pacific Islander Maternal Advisory Board about the best language to use to refer to the Pacific Islander population. They learned that the community felt most comfortable using "Pacific Islander," "Pasifika," or "Pasefika"—as opposed to "PI"—and consequently used these terms in their outreach materials. Participants also recommended that individuals be given a small welcome gift once selected. Participants felt that a list of local

XSL•FO
RenderX

resources was insufficient in terms of offering support, and instead favored having program staff directly link participants to services (ie, a warm handoff).

During this design sprint, participants expressed concerns about people "misusing" the ABP funds. While we originally wanted to create the lowest documentation burden for applicants, participants in the second design sprint expressed a desire for stricter documentation requirements to ensure eligibility. These preferences mirrored those expressed in the needs assessment interviews and the first design sprint. In addition, participants discussed what would happen in the case of a pregnancy loss.

Participants also weighed in on their willingness to participate in research to evaluate the impact of ABP. They noted the factors that would make them more likely to participate in the research, including transparency in how the results would be used and knowing that the data collection would be conducted by Black and Pacific Islander community members.

Final ABP Program Design

Overview

Through steps 1-4 of the formative community-centered research, we established the final program design of the ABP (Textbox 1). As a guaranteed income program, the income supplements were unconditional, which meant there were no requirements to receive or restrictions on the use of the funds. Recipients could use the funds in a manner that best supported their unique needs during pregnancy and postpartum.

Textbox 1. Final San Francisco Abundant Birth Project pilot elements, including the amount, duration, and disbursement of the income supplements, program eligibility criteria and assessment, participant selection, and additional support that resulted from the design process.

Program feature and selected design

Amount and duration

- US $1000 per month for 12 months

Disbursement

- Direct debit card that could be used for cash or banking

Eligibility

- Identify as Black and/or Pacific Islander
- Living in San Francisco or recently displaced
- <27 weeks pregnant
- Annual household income of <US $100,000

Assessment of eligibility criteria

- Self-reported income
- Proof of residence
- Provider verification of pregnancy

Selection

- Random selection through the bimonthly abundance drawing
- Participants were re-entered into the drawing until they reached the third trimester

Additional support: abundance coaching

- Opt-in abundance coaching to support housing needs and connections to additional resources.

Eligibility and Verification

Eligibility criteria listed in Textbox 1 were assessed using a web-based interest form and confirmed in an intake interview that occurred over Zoom (Zoom Video Communications). Applicants were required to provide verification of pregnancy and documentation of residence in San Francisco to ensure the program served its prioritized population. The allowable documents to confirm residence included a California-issued ID, school documents, employer pay stubs, or utility bill statements. Gestational weeks of pregnancy were verified by a medical provider. The documentation requirements reflected the intention to strike a balance between honoring community voices and not creating overly burdensome eligibility requirements. Household income was self-reported and, to make eligibility assessment simpler for program staff, not adjusted by household size. The maximum income requirements were purposely selected to allow low- and middle-income families to participate, given the high cost of living in San Francisco and that many families may not qualify for other governmental benefits but struggle to meet basic needs. On the interest form, applicants were able to self-identify as Black and/or Pacific Islander, the communities most impacted by preterm birth in San Francisco.

XSL•FO
RenderX

Eligibility Changes Related to Pregnancy Outcomes

The design processes helped inform what would happen if a participant's eligibility changed over the course of participation, particularly as it related to pregnancy outcomes. Using the expert opinion protocol, the ABP working group decided that in the event of a miscarriage, pregnancy termination, adoption, or death of the infant, ABP participants would continue to receive the income supplement for a 3-month grace period.

Implementing Randomized Selection

Bimonthly randomized drawings, called the "abundance drawing," ensured that all eligible individuals had an equal chance of being selected for ABP within each drawing. If applicants were not selected into ABP, they remained in the eligibility pool until they were selected or entered into the third trimester of pregnancy and therefore became ineligible.

Additional Support: Abundance Coaching

ABP offered optional case management services called "abundance coaching." In the needs assessment interviews, participants expressed a desire for more Black and Pacific Islander representation among staff from the local organizations that provided them with support. They also expressed previous stigmatizing experiences with government benefits programs. These perspectives informed the decision that all abundance coaches identify as Black and/or Pacific Islander to ensure racial and cultural concordance. Once enrolled, participants accessed optional support from abundance coaches who provided resource linkage and support with financial and other life goals. This included referrals to culturally specific community resources and providers. Unlike traditional case management which is often deficit-focused and paternalistic by directing and managing goals for participants, the abundance coaches provided individualized care and support based on the unique goals ABP participants set for themselves.

Evaluation

Using foundational strategies of equitable evaluation [49], we developed evaluation principles to abide by (1) leading with justice, (2) recognizing harm, (3) respecting autonomy, and (4) honoring community voice. We developed these principles based on ongoing work with the CRs and the responses from the interview and design sprint participants. The CRs, Expecting Justice collaborators, and research participants reported that too often Black and Pacific Islander people are excluded from decision-making in research and that there is a lack of transparency surrounding how data are used. The evaluation was opt-in for all ABP participants, based on the belief that the supplement should be truly unconditional, including research participation. The formative research also informed the decision that only CRs and other research team members who identified as Black and/or Pacific Islander would conduct surveys and interviews with participants.

Discussion

Principal Findings

Through a community-centered design process, we designed a novel guaranteed income program to center the needs of pregnant Black and Pacific Islander women and people in San Francisco. To our knowledge, no previous studies have qualitatively explored the financial, social services, and pregnancy and parenting experiences of these communities in San Francisco. Our needs assessment interviews revealed significant financial and life stressors experienced by Black and Pacific Islander people during and after pregnancy. Housing insecurity and previous poor experiences with social service programs were underscored by multiple participants. These findings are consistent with a report documenting that increases in San Francisco housing prices led to new concentrations of poverty especially for Black families from 2000 to 2015 [50]. Our findings also align with previous studies that report disproportionate unaffordable housing costs among people of color [51,52]. Additionally, the high cost of living in San Francisco left participants far above the federal poverty line ineligible for social services yet struggling to make ends meet. While all participants were enthusiastic about the potential of a guaranteed income program, they expressed concerns surrounding the eligibility and additional support needed to navigate the benefits system.

These experiences served as the foundation of the ABP prototypes and final program design. This formative process allowed for the establishment of eligibility and verification, supplement amount, disbursement options, and optional case management supports that were directly aligned with community members' expertise and preferences.

Strengths and Limitations

There are several strengths to this community-centered model of formative research and program development. Guaranteed income is a novel public health intervention, and the ABP program contributed to a swell of guaranteed income pilots in California and across the country [53-56]. Integrating community voice and expertise throughout the design process tailored ABP to the needs of the pregnant women and people it aimed to serve. We note that formative research with the CR model and shared decision-making requires more time and resources in the design phase, and institutions and academic hierarchies often do not allow space for this community-engaged work. Nonetheless, we strongly believe that community-centered formative research and program design ensures authentic engagement and leads to more relevant, responsive, and just programs and interventions. Our process can provide a roadmap for other programs to guide community co-design of an intervention.

Our results should also be considered in light of some limitations. Due to the COVID-19 pandemic, the needs assessment interviews and design sprints were implemented through videoconferencing only, and this mode of data collection may have affected the way participants engaged in the sessions. Participants' experiences of pregnancy and parenting were also likely changed by the pandemic. Additionally, eligibility for the interviews and design sprints included the ability to speak and read English, so these findings may not capture the experiences of non–English-speaking Black and Pacific Islander pregnant people in San Francisco, particularly immigrants.

XSL•FO
RenderX

Conclusions

The model of formative research to design ABP may be useful to other programs centering on community-driven, equity-oriented approaches. Interventions intended to address health equity should include experts with lived experience in the design process. Too often, well-intentioned programs do not meet the needs of recipients or center equity because of a lack of community participation in all elements of design.

The inclusion of CRs as integral members of the ABP staff was key to maximizing community involvement and shared decision-making. Ongoing staff and ABP working group meetings had designated spaces for CRs and community experts to advise ABP program implementation and evaluation design. Using the expert opinion protocol allowed for any ongoing issue to be decided with a consensus that prioritizes lived experience in developing solutions.

ABP aims to directly address the historical disinvestment of Black and Pacific Islander communities in San Francisco and positively impact birth outcomes in these communities through the provision of an unconditional income supplement. The COVID-19 pandemic has disproportionately impacted Black families, further underscoring the vital need for interventions that address structural deficiencies [57]. ABP applied a community-centered approach to design an intervention focused on the communities most impacted by maternal and birth disparities to reduce the effects of the racial wealth gap and economic insecurity among Black women. This novel approach and design process has great promise to reduce inequities in birth outcomes and center health and racial justice in antipoverty programs.

Acknowledgments

DK, JCW, MAT, MMDLC, SB, EC, MAC, AC, LL-G, BP, SS, ZM, and AMG contributed to the conceptualization, methodology, and implementation of the research. DK, JCW, MAT, MMDLC, SB, EC, MAC, AC, BP, SS, and AMG contributed to the analysis. DK and JCW contributed to the first draft of this paper. All authors contributed to the review and editing of subsequent drafts and reviewed the final paper. This research was supported by the March of Dimes Clinical and Social Science Challenge Grant, the California Preterm Birth Initiative funded by Marc and Lynne Benioff, and the UCSF-Kaiser Department of Research Building Interdisciplinary Research Careers in Women's Health Program (BIRCWH) funded by the National Institute of Child Health and Human Development (NICHD) and the Office of Research on Women's Health (ORWH; K12 HD052163. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official view of the funders.

Data Availability

Data are not available due to their sensitive nature and the risk of deductive disclosure.

Conflicts of Interest

None declared.

References

1. Almeida J, Bécares L, Erbetta K, Bettegowda VR, Ahluwalia IB. Racial/ethnic inequities in low birth weight and preterm birth: the role of multiple forms of stress. Matern Child Health J 2018;22(8):1154-1163 [FREE Full text] [doi: 10.1007/s10995-018-2500-7] [Medline: 29442278]
2. Behrman RE, Butler AS, editors. Institute of Medicine Committee on Understanding Premature Birth and Assuring Healthy Outcomes. US: National Academies Press; 2007.
3. Martin JA, Hamilton BE, Osterman MJK, Driscoll AK. Births: final data for 2020. Natl Vital Stat Rep 2021;70(2):1-51 [FREE Full text] [doi: 10.15620/cdc:100472] [Medline: 35157571]
4. Matthews TJ, MacDorman MF, Thoma ME. Infant mortality statistics from the 2013 period linked birth/infant death data set. Natl Vital Stat Rep 2015;64(9):1-30 [FREE Full text] [Medline: 26270610]
5. Petersen EE. MMWR. morbidity and mortality weekly report; vital signs: pregnancy-related deaths, United States, 2011-2015, and strategies for prevention, 13 States, 2013-2017. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/mmwr/volumes/68/wr/mm6818e1.htm [accessed 2019-09-02]
6. Sarnquist CC, Grieb EM, Maldonado YA. How racial and ethnic groupings may mask disparities: the importance of separating Pacific Islanders from Asians in prenatal care data. Matern Child Health J 2010;14(4):635-641. [doi: 10.1007/s10995-009-0494-x] [Medline: 19582560]
7. Rao AK, Daniels K, El-Sayed YY, Moshesh MK, Caughey AB. Perinatal outcomes among Asian American and Pacific Islander women. Am J Obstet Gynecol 2006;195(3):834-838. [doi: 10.1016/j.ajog.2006.06.079] [Medline: 16949421]
8. Singh GK, Yu SM. Pregnancy outcomes among Asian Americans. Asian Am Pac Isl J Health 1993;1(1):63-78. [Medline: 11567243]
9. Maternal, child, and adolescent health; SF preterm birth 5-year dashboard by social determinants 2018. San Francisco Department of Public Health. URL: https://www.sf.gov/departments/department-public-health/maternal-child-and-adolescent-health [accessed 2018-04-11]

XSL•FO
RenderX

10. California Department of Public Health. Preterm birth. Maternal, child, and adolescent health division. California Department of Public Health. URL: https://www.cdph.ca.gov/Programs/CFH/DMCAH/surveillance [accessed 2024-09-05]
11. Bailey ZD, Feldman JM, Bassett MT. How structural racism works - racist policies as a root cause of U.S. racial health inequities. N Engl J Med 2021;384(8):768-773. [doi: 10.1056/NEJMms2025396] [Medline: 33326717]
12. Jones CP. Levels of racism: a theoretic framework and a gardener's tale. Am J Public Health 2000;90(8):1212-1215. [doi: 10.2105/ajph.90.8.1212] [Medline: 10936998]
13. Lu MC, Halfon N. Racial and ethnic disparities in birth outcomes: a life-course perspective. Matern Child Health J 2003;7(1):13-30. [doi: 10.1023/a:1022537516969] [Medline: 12710797]
14. Hughes D, Simpson L. The role of social change in preventing low birth weight. Future Child 1995;5(1):87-102. [Medline: 7633870]
15. Blackwell SC, Gyamfi-Bannerman C, Biggio JR, Chauhan SP, Hughes BL, Louis JM, et al. 17-OHPC to prevent recurrent preterm birth in singleton gestations (PROLONG Study): a multicenter, international, randomized double-blind trial. Am J Perinatol 2020;37(2):127-136 [FREE Full text] [doi: 10.1055/s-0039-3400227] [Medline: 31652479]
16. Bailey ZD, Krieger N, Agénor M, Graves J, Linos N, Bassett MT. Structural racism and health inequities in the USA: evidence and interventions. Lancet 2017;389(10077):1453-1463. [doi: 10.1016/S0140-6736(17)30569-X] [Medline: 28402827]
17. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. Public Health Rep 2014;129(Suppl 2):19-31 [FREE Full text] [doi: 10.1177/00333549141291S206] [Medline: 24385661]
18. Ford CL, Airhihenbuwa CO. Critical race theory, race equity, and public health: toward antiracism praxis. Am J Public Health 2010;100(S1):S30-S35. [doi: 10.2105/ajph.2009.171058]
19. Malawa Z, Gaarde J, Spellen S. Racism as a root cause approach: a new framework. Pediatrics 2021;147(1):e2020015602. [doi: 10.1542/peds.2020-015602] [Medline: 33386339]
20. Wallace ME, Mendola P, Liu D, Grantz KL. Joint effects of structural racism and income inequality on small-for-gestational-age birth. Am J Public Health 2015;105(8):1681-1688. [doi: 10.2105/AJPH.2015.302613] [Medline: 26066964]
21. Hillier AE. Redlining and the home owners' Loan corporation. J Urban History 2003;29(4):394-420. [doi: 10.1177/0096144203029004002]
22. Stanczyk AB. The dynamics of household economic circumstances around a birth. Equitable Growth. 2016. URL: https://equitablegrowth.org/working-papers/income-volatility-around-birth/ [accessed 2019-11-18]
23. Taylor K, Compton S, Kolenic GE, Scott J, Becker N, Dalton VK, et al. Financial hardship among pregnant and postpartum women in the United States, 2013 to 2018. JAMA Netw Open 2021;4(10):e2132103 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.32103] [Medline: 34714338]
24. American Community Survey 2016-2020 5-year data release. US Census Bureau. URL: https://www.census.gov/newsroom/press-kits/2021/acs-5-year.html [accessed 2023-03-07]
25. Kendall M. Income inequality in the Bay Area is among nation's highest. The Mercury News. 2018. URL: https://www.mercurynews.com/2018/02/15/income-inequality-in-the-bay-area-is-among-nations-highest/ [accessed 2018-02-15]
26. Family needs calculator. Insight Center. URL: https://insightcced.org/family-needs-calculator/ [accessed 2022-05-25]
27. Guaranteed income: a primer for funders. Asset Funders Network, Center for High Impact Philanthropy, Economic Security Project, Springboard to Opportunities. URL: https://economicsecurityproject.org/wp-content/uploads/2022/10/220516-Funder-Primer.pdf [accessed 2022-05-16]
28. Guaranteed basic income projects. California Department of Social Services. URL: https://www.cdss.ca.gov/inforesources/guaranteed-basic-income-projects [accessed 2022-03-04]
29. West S, Castro BA, Samra S, Coltrera E, Addo M, Carlson M, et al. Preliminary analysis: SEED?s first year. Stockton Economic Empowerment Demonstration 2023;11(4):25 [FREE Full text]
30. Magnolia Mother's Trust evaluation report. Springboard to Opportunity. 2020. URL: https://springboardto.org/wp-content/uploads/2024/09/STO_MMT-2nd-Cohort-2020-Evaluation-Summary-1.pdf [accessed 2022-03-04]
31. Brownell MD, Nickel NC, Chartier M, Enns JE, Chateau D, Sarkar J, et al. An unconditional prenatal income supplement reduces population inequities in birth outcomes. Health Aff (Millwood) 2018;37(3):447-455. [doi: 10.1377/hlthaff.2017.1290] [Medline: 29505359]
32. Brownell MD, Chartier MJ, Nickel NC, Chateau D, Martens PJ, Sarkar J, et al. Unconditional prenatal income supplement and birth outcomes. Pediatrics 2016;137(6):e20152992. [doi: 10.1542/peds.2015-2992] [Medline: 27244846]
33. Kuang J. More than 12,000 Californians are getting cash from guaranteed income experiments. CalMatters. URL: http://calmatters.org/california-divide/2023/02/universal-basic-income/ [accessed 2024-08-27]
34. La L. Guaranteed basic income programs proliferate across California. CalMatters. URL: http://calmatters.org/newsletter/california-guaranteed-income-programs/ [accessed 2024-08-27]
35. Kania J, Kramer M. Collective impact. China: FSG Beijing; 2011.
36. National Birth Equity Collaborative (NBEC) - birth equity for all black birthing people. National Birth Equity Collaborative. URL: https://birthequity.org/ [accessed 2023-07-11]

XSL•FO
RenderX

37. Social determinants of health. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/about/priorities/why-is-addressing-sdoh-important.html?CDC_AAref_Val=https://www.cdc.gov/about/sdoh/index.html [accessed 2023-07-12]
38. Crear-Perry J, Correa-de-Araujo R, Lewis Johnson T, McLemore MR, Neilson E, Wallace M. Social and structural determinants of health inequities in maternal health. J Womens Health (Larchmt) 2021;30(2):230-235 [FREE Full text] [doi: 10.1089/jwh.2020.8882] [Medline: 33181043]
39. Ortiz K, Nash J, Shea L, Oetzel J, Garoutte J, Sanchez-Youngman S, et al. Partnerships, processes, and outcomes: a health equity-focused scoping meta-review of community-engaged scholarship. Annu Rev Public Health 2020;41(1):177-199 [FREE Full text] [doi: 10.1146/annurev-publhealth-040119-094220] [Medline: 31922931]
40. Boland SE, Homdayjanakul KJ, Reed ND, Wesner CA, Basha E, Kaufman CE, et al. Addressing the cycle of inaction: a DrPH student perspective on the decolonization of public health. HPHR. 2021. URL: https://bcphr.org/35-article-boland/ [accessed 2023-06-20]
41. Getrich CM, Sussman AL, Campbell-Voytal K, Tsoh JY, Williams RL, Brown AE, et al. Cultivating a cycle of trust with diverse communities in practice-based research: a report from PRIME Net. Ann Fam Med 2013;11(6):550-558 [FREE Full text] [doi: 10.1370/afm.1543] [Medline: 24218379]
42. Isler MR, Corbie-Smith G. Practical steps to community engaged research: from inputs to outcomes. J Law Med Ethics 2021;40(4):904-914. [doi: 10.1111/j.1748-720x.2012.00719.x]
43. Hamilton AB. Qualitative methods in rapid turn-around health services research. U.S. Department of Veterans Affairs. 2013. URL: https://www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/780-notes.pdf [accessed 2021-06-23]
44. Brown T. Design thinking. Harv Bus Rev 2008;86(6):84-92, 141. [Medline: 18605031]
45. Jake-Schoffman DE, McVay MA. Using the design sprint process to enhance and accelerate behavioral medicine progress: a case study and guidance. Transl Behav Med 2021;11(5):1099-1106. [doi: 10.1093/tbm/ibaa100] [Medline: 33057685]
46. Nijagal MA, Patel D, Lyles C, Liao J, Chehab L, Williams S, et al. Using human centered design to identify opportunities for reducing inequities in perinatal care. BMC Health Serv Res 2021;21(1):714 [FREE Full text] [doi: 10.1186/s12913-021-06609-8] [Medline: 34284758]
47. Hager ER, Quigg AM, Black MM, Coleman SM, Heeren T, Rose-Jacobs R, et al. Development and validity of a 2-item screen to identify families at risk for food insecurity. Pediatrics 2010;126(1):e26-e32. [doi: 10.1542/peds.2009-3146] [Medline: 20595453]
48. Floyd I, Pavetti L, Meyer L, Safawi A, Schott L, Bellew E, et al. TANF policies reflect racist legacy of cash assistance. Center on budget and policy priorities. URL: https://www.cbpp.org/research/family-income-support/tanf-policies-reflect-racist-legacy-of-cash-assistance [accessed 2022-06-24]
49. Equitable evaluation framework framing paper. San Rafael, CA: equitable evaluation initiative. Center for Equitable Evaluation, Institute for Foundation and Donor Learning, Luminaire Group, Dorothy A Johnson Center for Philanthropy. 2017. URL: https://sdbjrfoundation.issuelab.org/resource/equitable-evaluation-framework-framing-paper.html [accessed 2017-06-01]
50. Verma P, Rinzler D, Zuk M. Rising housing costs and re-segregation in San Francisco. UC Berkeley's Urban Displacement Project And California Housing Partnership. URL: https://www.urbandisplacement.org/wp-content/uploads/2021/08/sf_final.pdf [accessed 2024-12-25]
51. Kimberlin S. California?s Housing Affordability Crisis Hits Renters and Households With the Lowest Incomes the Hardest. Sacramento, CA: California Budget & Policy Center; 2019.
52. Levin M. Black Californians' housing crisis, by the numbers. CalMatters. URL: http://calmatters.org/housing/2020/06/black-californians-housing-crisis-by-the-numbers/ [accessed 2024-10-15]
53. Philly Joy Bank. The Philadelphia City Fund. URL: https://philacityfund.org/programs/philly-joy-bank/ [accessed 2004-11-09]
54. Our work. The Bridge Project. URL: https://bridgeproject.org/our-work/ [accessed 2024-11-09]
55. The nest (guaranteed income). Hummingbird Indigenous Family Services. URL: https://www.hummingbird-ifs.org/programs/nest/the-nest [accessed 2024-11-09]
56. About. Black Economic Equity Movement (BEEM). URL: https://beemproject.org/about/ [accessed 2024-11-09]
57. Lopez MH, Rainie L, Budiman A. Financial and health impacts of COVID-19 vary widely by race and ethnicity. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2020/05/05/financial-and-health-impacts-of-covid-19-vary-widely-by-race-and-ethnicity/ [accessed 2020-05-05]

Abbreviations

ABP: Abundant Birth Project
CR: community researcher
RAP: Rapid Assessment Process

XSL•FO
RenderX

Edited by A Mavragani; submitted 05.06.24; peer-reviewed by KEC Massengale, EC Rhodes, X Yu; comments to author 13.08.24; revised version received 10.11.24; accepted 06.12.24; published 27.01.25.

Please cite as:
Karasek D, Williams JC, Taylor MA, De La Cruz MM, Arteaga S, Bell S, Castillo E, Chand MA, Coats A, Hubbard EM, Love-Goodlett L, Powell B, Spellen S, Malawa Z, Gomez AM
Designing the First Pregnancy Guaranteed Income Program in the United States: Qualitative Needs Assessment and Human-Centered Design to Develop the Abundant Birth Project
JMIR Form Res 2025;9:e60829
URL: https://formative.jmir.org/2025/1/e60829
doi:10.2196/60829
PMID:

©Deborah Karasek, Jazzmin C Williams, Michaela A Taylor, Monica M De La Cruz, Stephanie Arteaga, Sabra Bell, Esperanza Castillo, Maile A Chand, Anjeanette Coats, Erin M Hubbard, Latriece Love-Goodlett, Breezy Powell, Solaire Spellen, Zea Malawa, Anu Manchikanti Gomez. Originally published in JMIR Formative Research (https://formative.jmir.org), 27.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Tobacco Use, Risk Perceptions, and Characteristics of Adults Who Used a Heated Tobacco Product (IQOS) in the United States: Cross-Sectional Survey Study

Hui G Cheng[1], PhD; Brendan Noggle[1], MPH; Andrea R Vansickel[1], PhD; Edward G Largo[1], PhD; Pierpaolo Magnani[2]

[1]Altria Client Services LLC, Richmond, VA, United States

[2]PMI R&D, Philip Morris Products S.A., Neuchâtel, Switzerland

Corresponding Author:
Andrea R Vansickel, PhD
Altria Client Services LLC
601 East Jackson Street
Richmond, VA, 23219
United States
Phone: 1 8595765695
Email: Andrea.R.Vansickel@altria.com

Abstract

Background: The Tobacco Heating System (THS; commercialized as IQOS) is a smoke-free heated tobacco product introduced in the United States in 2019 and authorized by the US Food and Drug Administration as a modified risk tobacco product (MRTP) in 2020. THS consists of a holder and specially designed tobacco sticks that are heated instead of burned to produce a nicotine-containing aerosol. THS was available in Atlanta, Georgia; Richmond, Virginia; Charlotte, North Carolina; the Northern Virginia region; and South Carolina before its market removal in November 2021.

Objective: This study aims to describe selected sociodemographic characteristics and self-reported health history of adults who used IQOS (AUIs), their tobacco use patterns (eg, tobacco use history, exclusive and dual-use, and switching from cigarette smoking), their risk perceptions of the product, and their understanding of MRTP messages.

Methods: The IQOS Cross-Sectional Postmarket Adult Consumer Study was a study of AUIs aged 21 years or older who were recruited from a consumer database via direct postal mail and emails. Participants completed the online survey between September and November 2021.

Results: The survey was completed by 645 current and 43 former AUIs who had used at least 100 tobacco sticks (considered established THS use) before the assessment. Of the 688 participants, 424 (61.6%) were male, 502 (73.0%) were non–Hispanic White, and the mean age was 45 years. The vast majority (680/688, 98.8%) of AUIs had ever smoked combusted cigarettes before first trying THS and 628 (91.3%) had smoked cigarettes in the 30 days before first using THS. At the time of assessment, 161 (23.4%) reported using e-cigarettes (vs 229, 33.3%, before THS use), 92 (13.4%) reported smoking cigars (vs 114, 16.6%, before THS use), and 338 (49.1%) were still smoking after an average of 1 year of THS use. Among those currently using THS who were still smoking (n=298), 249 (83.6%) smoked fewer cigarettes compared with before first trying THS; 362 of 688 (52.6%) AUIs reported having no physical health conditions evaluated in this study and almost three-quarters reported having no mental health conditions. Among all AUIs, over 563 (81.8%) had never used a cessation treatment or had not used it in the past 12 months, and 555 (80.7%) AUIs demonstrated a correct understanding of the MRTP message and AUIs perceived THS as having a lower risk than cigarettes (43.8 vs 64.4 on a 100-point composite score scale).

Conclusions: This study provides evidence that THS can help adult smokers in the United States completely switch away from cigarettes or reduce smoking.

(JMIR Form Res 2025;9:e57398) doi:10.2196/57398

KEYWORDS

United States Food and Drug Administration; FDA; IQOS; Tobacco Heating System; THS; heated tobacco products; modified risk tobacco products; MRTP; tobacco; nicotine; smokers; tobacco harm reduction; cross-sectional surveys; cigarettes

XSL•FO
RenderX

Introduction

Cigarette smoking is a leading contributor to mortality and morbidity worldwide [1]. In the United States, smoking contributes to an estimated 500,000 premature deaths and an expenditure exceeding US $170 billion in annual medical costs to treat smoking-related lung and cardiovascular illnesses [2]. Although quitting is the best way to mitigate health risks from smoking, it remains challenging for many smokers. Despite 70% of smokers wanting to quit smoking and more than half attempting each year [3], the majority of those who attempt to quit relapse within 12 months [4-7]. Individuals who smoke make many attempts before successfully quitting, rendering the quitting journey a protracted relapsing process [7,8]. Given that health risks from cigarette smoking are mainly attributed to harmful constituents generated through combustion, smokers who are unable or unwilling to quit may benefit from switching to smoke-free products such as heated tobacco products including the Tobacco Heating System (THS; commercialized as *IQOS* by Philip Morris Products S.A.) [9-12].

THS is a smoke-free tobacco product that heats specially designed tobacco sticks (commercialized as *HeatSticks* by Philip Morris Products S.A.) to temperatures below the level of combustion and produces an aerosol with substantially reduced (ie, 90%-95% lower, on average) levels of harmful and potentially harmful constituents compared with cigarette smoke while delivering similar levels of nicotine [13-15]. The product debuted in 2014 in Japan and was available in more than 80 countries as of September 2024.

Evidence collected to date supports the role of THS in tobacco harm reduction. The product was rapidly adopted by individuals who smoked in Japan [16], and recent data suggest that the majority of individuals who used THS have switched to exclusive use and stopped smoking [17]. Moreover, individuals who used heated tobacco products exclusively, as opposed to smoking, had more favorable health-related biomarker profiles compared with smokers [18]. In a prospective study of 38 Italian smokers with chronic obstructive pulmonary disease who started to use THS, 58% stopped smoking within 3 years, and those who were still smoking significantly reduced their cigarette consumption (mean cigarette consumption dropped from 21 per day at baseline to 3.7 per day at the last follow-up). Most importantly, consistent improvements were observed in respiratory symptoms, exercise tolerance, quality of life, and rate of disease exacerbations, whereas no such change was observed among individuals who smoked cigarettes during the same follow-up period [19]. These results suggest that switching from cigarette smoking to THS may be associated with long-term health benefits.

The US Food and Drug Administration (FDA) authorized the marketing of THS in the United States in April 2019. Subsequently, it granted a modified risk tobacco product (MRTP) order in July 2020 authorizing the use of the following reduced exposure claim:

“AVAILABLE EVIDENCE TO DATE:

- The IQOS system heats tobacco but does not burn it.
- This significantly reduces the production of harmful and potentially harmful chemicals.
- Scientific studies have shown that switching completely from conventional cigarettes to the IQOS system significantly reduces your body’s exposure to harmful or potentially harmful chemicals.”

THS was first introduced to the United States in Atlanta (Georgia) in October 2019. Marketing expanded to Richmond (Virginia) in January 2020, Charlotte (North Carolina) in July 2020, and northern Virginia and South Carolina as of May 2021 before it was removed from all US markets in November 2021 as a result of a Cease-and-Desist Order (CDO) issued by the United States International Trade Commission (ITC). (The ITC issued a CDO prohibiting the importation, marketing, sale, and distribution of *IQOS* devices and *HeatSticks* on September 29, 2021. To comply with the CDO, Philip Morris USA stopped marketing and selling the products as of November 28, 2021, in all channels.) Three types of tobacco sticks were sold: amber/regular (nonmenthol), green/smooth menthol (menthol), and blue/fresh menthol (menthol). At the end of August 2021, the tobacco sticks were available in 4842 third-party retail outlets, and the THS devices were available in 1235 third-party retail outlets.

This study aimed to (1) characterize adults who used *IQOS* (AUIs) and their history and patterns of tobacco use, as well as tobacco cessation treatment use; (2) describe risk perceptions of THS and the understanding of the MRTP claim; and (3) describe complete switching from cigarette smoking to THS, transitions to/back to cigarette smoking, and relevant quitting behaviors among AUIs. We used data from the 2021 *IQOS* Cross-sectional Postmarket Adult Consumer Study to provide the first evidence about AUIs, how AUIs used the product, and the potential role of such products in tobacco harm reduction in real-world settings in the United States.

Methods

Study Design, Study Population, and Recruitment

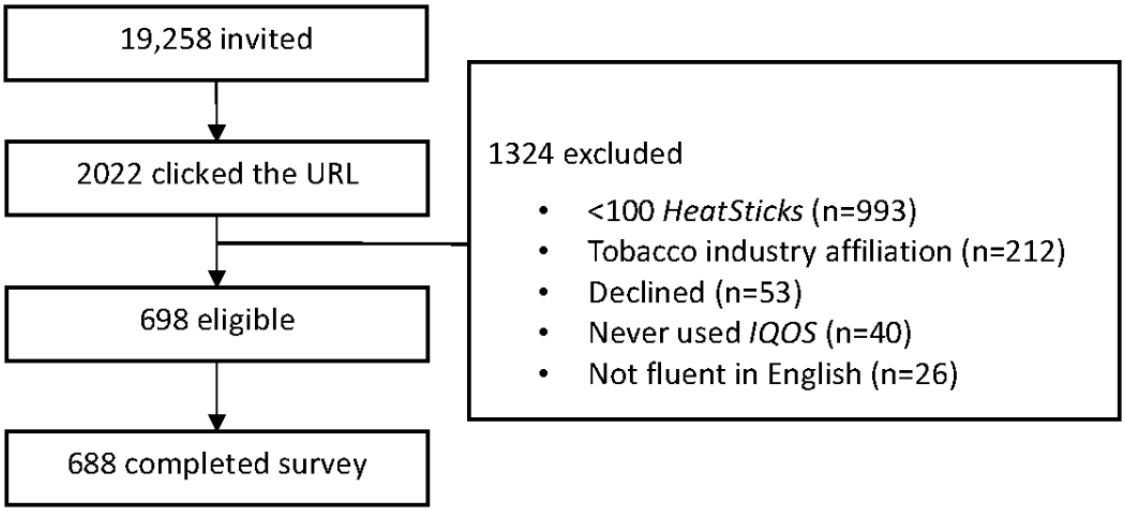
This was a cross-sectional study. The study population consisted of adults, aged 21 years or older, who had used at least 100 tobacco sticks with THS before the assessment. Individuals were excluded if (1) they were unable to read, speak, or understand English; (2) they or their first-degree relative were current or former employees of the tobacco industry, the Central Research Organization, or any market research firm; or (3) they or their first-degree relative were currently involved in litigation against any tobacco industry entity.

In this study, we recruited AUIs from a consumer database, which covered the majority of individuals who purchased an *IQOS* device. The *IQOS* Consumer Database is a database of registered *IQOS* consumers in the United States. These consumers entered the database by voluntarily registering their devices either at the point of sale or later by themselves. Based on the number of devices registered and sold, we estimated that approximately 70% of *IQOS* consumers were included in the *IQOS* Consumer Database. We sent invitations by both direct mail and email (if available) to all individuals in the database

XSL•FO
RenderX

who agreed to be contacted for research purposes: 1 initial invitation and 1 reminder were sent to all potential participants. Figure 1 depicts the recruitment process and final sample. Data were collected between September 15 and November 14, 2021. In brief, of the 19,258 individuals invited, 2022 (10.50%) completed the screener; of the 698 eligible individuals who consented to participate, 688 (98.6%) completed the survey.

Figure 1. Flowchart depicting sample recruitment for the Cross-Sectional Postmarket Adult Consumer Study (2021). Participants sampled from the IQOS consumer database were primarily residing in Georgia, Virginia, North Carolina, or South Carolina, United States.



Ethical Considerations

The study was reviewed (study protocol, questionnaire, and informed consent statement) and approved by the Sterling Institutional Review Board (IRB ID number: 9102-HCheng). This study was conducted in compliance with the study protocol and, where applicable, in accordance with the Insights Association Code of Standards and Ethics for Market Research and Data Analytics [20] and the International Chamber of Commerce/European Society for Opinion and Marketing Research's (ESOMAR) International Code on Market, Opinion and Social Research [21]. All participants read the informed consent statement and voluntarily consented to participate in the study via electronic agreement. All data generated through this study were considered highly confidential by study staff and participants were identified only via unique participation identification numbers. Upon completion of the questionnaire, participants were provided instructions to obtain a US $40 electronic gift card.

Assessment and Definitions

All eligible participants, regardless of the mode of recruitment, self-administered the questionnaire online. The assessment consisted of modules of questions on the history of use of various tobacco products (including THS, cigarettes, cigars, pipe, hookah, e-cigarettes, smokeless tobacco, and oral tobacco-derived nicotine products). Additional modules included risk perceptions and understanding of claim-related information, demographic characteristics, and diagnoses of selected physical and mental conditions.

In this study, we defined ever use as a "yes" response to the question "Have you EVER used/smoked...EVEN ONE TIME?" Current use was defined as "every day" or "some days" responses to the question "Do you now use/smoke...every day, some days, or not at all?" Former use was defined as a "yes" response to the ever-use question and a "not at all" response to the current-use question. Established cigarette smoking and THS use were defined as having smoked at least 100 cigarettes or used at least 100 tobacco sticks in a lifetime. Menthol/nonmenthol preference was assessed using the question "Which *one* type of *HeatSticks* are you *CURRENTLY* using *most often*?" Responses of "green menthol" (previously "smooth menthol") and "blue menthol" (previously "fresh menthol") were classified as preferring menthol; responses of "amber" (previously "regular") were classified as preferring nonmenthol. We identified menthol/nonmenthol cigarette preference via questions about the type of cigarettes smoked most often.

In this study, we asked time-centered questions to establish a temporal sequence of smoking and THS use to assess changes in smoking behaviors after first trying THS. For example, we classified current smokers before first trying THS as individuals who answered "every day" or "some days" to the question "During the 30 days before you first tried *IQOS*, did you smoke cigarettes every day, some days, or not at all?". Using this approach, we defined individuals who had *completely switched* as current established AUIs, who (1) had smoked at least 100 cigarettes in their lifetime, (2) were smoking every day or some days in the 30 days before first trying THS, (3) were not smoking at the time of assessment, and (4) had smoked their last cigarette after trying THS (via the question "Was the last time you smoked cigarettes...?" with response options "Before trying *IQOS* for the first time" and "After trying *IQOS* for the first time").

Perceived health risks were assessed using the validated 18-item Perceived Health Risk scale, which has demonstrated good psychometric properties [22]. Two batteries of questions were asked for cigarette smoking and THS use, respectively. Each item was rated on a 5-point Likert-type scale ranging from 0 (no risk) to 4 (very high risk). An overall score was calculated by summing all items in the scale (possible range 0-72), with higher scores representing greater perceived health risk [22].

XSL•FO
RenderX

The understanding of the MRTP message was assessed using 2 questions shown in Textbox 1.

Textbox 1. Questions to assess understanding of the modified risk tobacco product message.

- 1. Based on what you know or believe, please complete the following: Smokers who switch completely from cigarettes to IQOS:
  - Have more exposure to harmful or potentially harmful chemicals
  - Have the same exposure to harmful or potentially harmful chemicals
  - Have less exposure to harmful or potentially harmful chemicals (correct answer)
  - Have no exposure to harmful or potentially harmful chemicals
  - Don't know

IF "HAVE LESS EXPOSURE..." SELECTED

- 2. Based on what you know or believe, what do smokers need to do in order to reduce their body's exposure to harmful or potentially harmful chemicals?
  - Stop smoking cigarettes completely and only use IQOS (correct answer)
  - Smoke fewer cigarettes and also use IQOS
  - Keep smoking the same amount of cigarettes and also use IQOS
  - Don't know

Demographic characteristics included sex (male or female), age, race/ethnicity, level of education, household income, employment status, and marital status. A lifetime history of physical conditions was assessed using the question "Have you ever been told by a doctor, nurse, or other health professional that you had...?" Options included cardiovascular conditions, respiratory conditions, cancers, and diabetes (see Table 2 for details). A lifetime history of mental conditions was assessed using 2 questions: "A mental illness or disorder refers to a wide range of mental health conditions or disorders that affect your mood, thinking, or behavior. Has a doctor, nurse, or other health professional ever told you that you had a mental health condition such as depression, stress, or problems with emotions?" and "Are you now taking medicine or receiving treatment from a doctor, nurse, or other health professional for a mental health condition or emotional problem?" All information was based on self-report.

Analysis

Descriptive statistics were used to summarize outcome variables. For categorical variables, frequencies and proportions (95% CIs) were generated. For numeric variables (eg, age), means (95% CI) and medians (IQRs) were calculated. Estimates (eg, proportions and means) were presented for current and ever AUIs, respectively. Demographic characteristics and THS use behaviors were further stratified for AUIs who preferred menthol and nonmenthol tobacco sticks. Estimates are not shown for former established AUIs due to imprecision as a result of the small sample size (n=43). Missing values were kept as missing except for missing values from logic skips in which case, values were assigned based on the logic skip. For example, if a participant indicated that they had never used an ENDS (electronic nicotine delivery systems) product, then the current ENDS use question would be skipped and a "no" response value assigned. Data analyses were conducted using SAS version 9.4 (SAS Institute Inc.). The protocol for this study was published previously [23]. Data are available for interested researchers. Please contact the corresponding author for details.

Results

Characteristics of Study Participants

Of the 688 established AUIs, 273 (39.7%) resided in Georgia, 179 (26.0%) resided in North Carolina, 177 (25.7%) resided in Virginia, and 45 (6.5%) resided in South Carolina. Nominal numbers (n<5) of participants resided in other states. Of the total 688 participants, 424 (61.6%) were males, and the mean age was 45 years, with the majority between 35 and 54 years of age. On average, current established AUIs who preferred menthol tobacco sticks were younger than those who preferred nonmenthol tobacco sticks (mean age 43.1 years, 95% CI 42.0-44.3 vs 47.3 years, 95% CI 46.1-48.6). Of all established AUIs who had ever smoked cigarettes (n=680), 279 (41.0%; 95% CI 36.9%-44.3%) smoked menthol cigarettes most often. Individuals who were non–Hispanic White accounted for the majority of established AUIs (502/688, 73%). Differences in tobacco stick varieties were observed across race/ethnicity groups; greater proportions of non–Hispanic Whites preferred nonmenthol tobacco sticks, whereas greater proportions of non–Hispanic Black and non–Hispanic Asians preferred menthol varieties. Approximately two-thirds of established AUIs had an annual family income greater than US $60,000, 556 (80.8%) had at least some college education, and 547 (79.5%) were in the workforce. Table 1 provides more detailed information about the demographic characteristics of the sample.

XSL•FO
RenderX

Table 1. Demographic characteristics of established AUIs[a,b].

Characteristics	Ever-established AUIs (n=688)	Current established AUIs (n=645)	Current established AUIs who prefer menthol tobacco sticks (n=337)	Current established AUIs who prefer nonmenthol tobacco sticks (n=8)
Sex, n (%); 95% CI				
Male	424 (61.6); 58 to 65.3	394 (61.1); 57.3 to 64.9	205 (61.1); 55.9 to 66.3	188 (61.0); 55.6 to 66.5
Age (years)				
Mean (95% CI)	45.0 (44.1 to 45.9)	45.1 (44.3 to 46)	43.1 (42 to 44.3)	47.3 (46.1 to 48.6)
Median (IQR)	43.5 (36-54)	44 (36-54)	41 (35-50)	47 (39 to 55.5)
21-24, n (%); 95% CI	14 (2.0); 1.0 to 3.1	13 (2.0); 0.9 to 3.1	6 (1.8); 0.7 to 3.8	7 (2.3); 0.9 to 4.6
25-34, n (%); 95% CI	121 (17.6); 14.7 to 20.4	109 (16.9); 14 to 19.8	76 (22.6); 18.1 to 27	33 (10.7); 7.3 to 14.2
35-44, n (%); 95% CI	222 (32.3); 28.8 to 35.8	209 (32.4); 28.8 to 36	121 (35.9); 30.8 to 41	88 (28.6); 23.5 to 33.6
45-54, n (%); 95% CI	178 (25.9); 22.6 to 29.1	172 (26.7); 23.3 to 30.1	81 (24.0) 19.5 to 28.6	91 (29.6); 24.5 to 34.6
55-64, n (%); 95% CI	112 (16.3); 13.5 to 19	103 (16); 13.1 to 18.8	37 (10.9); 7.6 to 14.3	66 (21.4); 16.9 to 26
65+, n (%); 95% CI	41 (6.0); 4.2 to 7.7	39 (6.1); 4.2 to 7.9	16 (4.8); 2.5 to 7	23 (7.5); 4.5 to 10.4
Race/ethnicity, n (%); 95% CI				
Hispanic/Latino	40 (5.8); 4.1 to 7.6	39 (6.1); 4.2 to 7.9	20 (5.9); 3.4 to 8.5	19 (6.2); 3.5 to 8.9
Non–Hispanic White	502 (73.0); 70.0 to 76.3	463 (71.8); 68.3 to 75.3	212 (62.9); 57.8 to 68.1	251 (81.5); 77.2 to 85.8
Non–Hispanic Black	38 (5.5); 3.8 to 7.2	37 (5.7); 3.9 to 7.5	28 (8.3); 5.4 to 11.3	9 (2.9); 1.3 to 5.5
Non–Hispanic Asian	91 (13.2); 10.7 to 15.8	90 (13.9); 11.3 to 16.6	68 (20.2); 15.9 to 24.5	22 (7.1); 4.3 to 10.0
Non–Hispanic Native Hawaiian or other Pacific Islander	0 (0); <0.1 to 0.5	0 (0); <0.1 to 0.6	0 (0); <0.1 to 1.1	0 (0); <0.1 to 1.2
Non–Hispanic American Indian or Alaska Native	8 (1.2); 0.5 to 2.3	8 (1.2); 0.5 to 2.4	6 (1.8); 0.7 to 3.8	2 (0.7); 0.1 to 2.3
Non–Hispanic Other	9 (1.3); 0.6 to 2.5	8 (1.2); 0.5 to 2.4	3 (0.9); 0.2 to 2.6	5 (1.6); 0.5 to 3.8
Household income (US $), n (%); 95% CI				
<60,000	231 (33.6); 30.1 to 37.1	212 (32.9); 29.2 to 36.5	120 (35.6); 30.5 to 40.7	92 (29.9).8 to 35
60,000-74,999	67 (9.7); 7.5 to 12.0	63 (9.8); 7.5 to 12.1	35 (10.4); 7.1 to 13.6	28 (9.1); 5.9 to 12.3
75,000-99,999	124 (18); 15.2 to 20.9	118 (18.3); 15.3 to 21.3	59 (17.5); 13.5 to 21.6	59 (19.2); 14.8 to 23.6
100,000-149,999	125 (18.2); 15.3 to 21.1	117 (18.1); 15.2 to 21.1	65 (19.3); 15.1 to 23.5	52 (16.9); 12.7 to 21.1
≥150,000	95 (13.8); 11.2 to 16.4	90 (14); 11.3 to 16.6	35 (10.4); 7.1 to 13.6	55 (17.9); 13.6 to 22.1
Prefer not to answer	46 (6.7); 4.8 to 8.6	45 (7.0); 5.0 to 8.9	23 (6.8); 4.1 to 9.5	22 (7.1); 4.3 to 10.0
Education, n (%); 95% CI				
High school or less	132 (19.2); 16.2 to 22.1	125 (19.4); 16.3 to 22.4	62 (18.4); 14.3 to 22.5	63 (20.5); 16.0 to 25.0

XSL•FO
RenderX

Characteristics	Ever-established AUIs (n=688)	Current established AUIs (n=645)	Current established AUIs who prefer menthol tobacco sticks (n=337)	Current established AUIs who prefer nonmenthol tobacco sticks (n=8)
Some college	196 (28.5); 25.1 to 31.9	175 (27.1); 23.7 to 30.6	89 (26.4); 21.7 to 31.1	86 (27.9); 22.9 to 32.9
Associate's degree	85 (12.4); 9.9 to 14.8	82 (12.7); 10.1 to 15.3	51 (15.1); 11.3 to 19.0	31 (10.1); 6.7 to 13.4
Bachelor's degree	179 (26.0); 22.7 to 29.3	171 (26.5); 23.1 to 29.9	93 (27.6); 22.8 to 32.4	78 (25.3); 20.5 to 30.2
Master's degree	76 (11.1); 8.7 to 13.4	73 (11.3); 8.9 to 13.8	32 (9.5); 6.4 to 12.6	41 (13.3); 9.5 to 17.1
Professional school degree	4 (0.6); 0.2 to 1.5	4 (0.6); 0.2 to 1.6	1 (0.3); <0.1 to 1.6	3 (1); 0.2 to 2.8
Doctorate	9 (1.3); 0.6 to 2.5	9 (1.4); 0.6 to 2.6	4 (1.2); 0.3 to 3	5 (1.6); 0.5 to 3.8
Other	7 (1); 0.4 to 2.1	6 (0.9); 0.3 to 2.0	5 (1.5); 0.5 to 3.4	1 (0.3); <0.1 to 1.8
Employment status, n (%); 95% CI				
Employed for wages	439 (63.8); 60.2 to 67.4	412 (63.9); 60.2 to 67.6	223 (66.2); 61.1 to 71.2	189 (61.4); 55.9 to 66.8
Self-employed	108 (15.7); 13.0 to 18.4	104 (16.1); 13.3 to 19.0	47 (14.0); 10.3 to 17.7	57 (18.5); 14.2 to 22.8
Not employed	141 (20.5); 17.5 to 23.5	129 (20.0); 16.9 to 23.1	67 (19.9); 15.6 to 24.1	62 (20.1); 15.7 to 24.6
Marital status, n (%); 95% CI				
Married	370 (53.8); 50.1 to 57.5	355 (55.0); 51.2 to 58.9	179 (53.1); 47.8 to 58.4	176 (57.1); 51.6 to 62.7
Widowed	8 (1.2); 0.5 to 2.3	8 (1.2); 0.5 to 2.4	4 (1.2); 0.3 to 3.0	4 (1.3); 0.4 to 3.3
Divorced	100 (14.5); 11.9 to 17.2	93 (14.4); 11.7 to 17.1	42 (12.5); 8.9 to 16.0	51 (16.6); 12.4 to 20.7
Separated	15 (2.2); 1.1 to 3.3	12 (1.9); 0.8 to 2.9	8 (2.4); 1.0 to 4.6	4 (1.3); 0.4 to 3.3
Never married	116 (16.9); 14.1 to 19.7	109 (16.9); 14.0 to 19.8	68 (20.2); 15.9 to 24.5	41 (13.3); 9.5 to 17.1
Living with partner	60 (8.7); 6.6 to 10.8	50 (7.8); 5.7 to 9.8	27 (8.0); 5.1 to 10.9	23 (7.5); 4.5 to 10.4
Do not wish to answer	19 (2.8); 1.5 to 4.0	18 (2.8); 1.5 to 4.1	9 (2.7); 1.2 to 5.0	9 (2.9); 1.3 to 5.5

[a]AUI: adult who used *IQOS*.

[b]Participants sampled from the *IQOS* consumer database were primarily residing in Georgia, Virginia, North Carolina, or South Carolina.

Just over half of AUIs reported they had never received any physical diagnoses included in this study (362/688, 52.6%; 95% CI 48.9%-56.4%; see Table 2 for more details). The most common conditions reported were hypertension (165/688, 24.0%; 95% CI 20.8%-27.2%) and high cholesterol (148/688, 21.5%; 95% CI 18.4%-24.6%). Approximately a quarter of participants reported they had ever had a mental health condition (181/688, 26.3%; 95% CI 23.0%-29.6%).

XSL•FO
RenderX

Table 2. Lifetime diagnosis of selected physical and mental health conditions among established AUIs[a,b].

Lifetime diagnosis	Ever-established AUIs (n=688), n (%); 95% CI	Current established AUIs (n=645), n (%); 95% CI
Physical condition		
A heart attack, also called myocardial infarction	10 (2.8); 1.5 to 4.0	17 (2.6); 1.4 to 3.9
Angina, also called angina pectoris (chest pain or discomfort)	12 (1.7); 0.8 to 2.7	11 (1.7); 0.7 to 2.7
Congestive heart failure	5 (0.7); 0.2 to 1.7	5 (0.8); 0.3 to 1.8
Coronary heart disease	7 (1.0); 0.4 to 2.1	7 (1.1); 0.4 to 2.2
High blood pressure (hypertension)	165 (24.0); 20.8 to 27.2	154 (23.9); 20.6 to 27.2
High cholesterol (hyperlipidemia)	148 (21.5); 18.4 to 24.6	139 (21.6); 18.4 to 24.7
Stroke	9 (1.3); 0.6 to 2.5	9 (1.4); 0.6 to 2.6
Any other heart condition or heart disease	7 (1.0); 0.4 to 2.1	7 (1.1); 0.4 to 2.2
Chronic obstructive pulmonary disease	28 (4.1); 2.6 to 5.6	28 (4.3); 2.8 to 5.9
Chronic bronchitis	25 (3.6); 2.2 to 5.0	23 (3.6); 2.13 to 5
Emphysema	5 (0.7); 0.2 to 1.7	4 (0.6); 0.2 to 1.6
Asthma	66 (9.6); 7.4 to 11.8	63 (9.8); 7.5 to 12.1
Apnea (pauses in breathing during sleep)	51 (7.4); 5.5 to 9.4	46 (7.1); 5.2 to 9.1
Any other respiratory or lung condition	2 (0.3); <0.1 to 1.1	2 (0.3); <0.1 to 1.1
Cancer	23 (3.3); 2.0 to 4.7	20 (3.1); 1.8 to 4.4
Diabetes	46 (6.7); 4.8 to 8.6	43 (6.7); 4.7 to 8.6
None of the above	362 (52.6); 48.9 to 56.4	342 (53.0); 49.2 to 56.9
Mental health condition		
Yes	181 (26.3); 23.0 to 29.6	166 (25.7); 22.4 to 29.1
No	477 (69.3); 65.9 to 72.8	450 (69.8); 66.2 to 73.3
Do not know	30 (4.4); 2.8 to 5.9	29 (4.5); 2.9 to 6.1
Taking medication or receiving treatment for a mental health condition or emotional problem		
Yes	121 (17.7); 14.9 to 20.6	113 (17.5); 14.6 to 20.5
No	548 (79.7); 76.6 to 82.7	516 (80.0); 76.9 to 83.1
Do not know	18 (2.6); 1.4 to 3.8	16 (2.5); 1.3 to 3.7

[a]AUI: adult who used *IQOS*.

[b]Participants sampled from the *IQOS* consumer database were primarily residing in Georgia, Virginia, North Carolina, or South Carolina.

Tobacco Use History and Cessation Treatment Use History

As shown in Tables 3 and 4, 680 of the 688 (98.8%) established AUIs had ever smoked cigarettes and 628 (91.3%) were currently smoking cigarettes before they tried THS for the first time. By contrast, at the time of assessment, only 338 (49.1%) were smoking. At the time of the assessment, 144 of 645 (22.3%) current established AUIs were also using e-cigarettes and 86 (13.3%) were also using cigars. Use of other tobacco products was relatively rare (≤5%, 33 or fewer/645). Current established smokers (n=326) had an average of 24.0 (95% CI 22.6-25.4) years of smoking history; former established smokers (ie, AUIs who had stopped smoking; n=329) had an average of 20.9 (95% CI 19.6-22.3) years of smoking history.

When asked about cigarette smoking during the 30 days before first trying THS, 519 (75.4%) participants indicated they were smoking every day, 109 (15.8%) were smoking some days but not every day, and 52 (7.6%) indicated they were not smoking cigarettes (including 16 participants who had not smoked cigarettes for at least 12 months, ie, former established smokers). Only a small proportion (n=16, 2.3%) of ever-established THS users were former established smokers before first trying THS. Of note, 11 of these 16 were using at least one other type of tobacco product during the 30 days before trying THS.

XSL•FO
RenderX

The vast majority of ever-established AUIs who were former established smokers at the time of the assessment became former smokers after first trying THS. Of the 329 former established smokers at the time of the assessment, 289 (87.8%) were smoking cigarettes during the 30 days before first trying THS or smoked after first trying THS, which suggested that they became former smokers after first trying THS. Of the 30 participants who did not smoke during the 30 days before first trying THS and indicated that the last time they smoked was before trying THS, 14 indicated they had not smoked for less than 1 month before trying THS, and 16 had not smoked for at least 12 months before first trying THS.

Approximately one-third of established AUIs were using e-cigarettes and about 114 (16.6%) were using cigars; use of other tobacco products was less common (ie, ≤44/688, <7%) before first trying THS.

With respect to cessation treatment, 330 out of 645 (51.2%; 95% CI 47.3%-55.0%) current established AUIs had never used any tobacco cessation treatment, 202 (31.3%; 95% CI 27.7%-34.9%) had used it more than 12 months ago, 39 (6.0%; 95% CI 4.2%-7.9%) used it within the past 12 months but more than 30 days ago, and 27 (4.2%; 95% CI 2.6%-5.7%) used it within the past 30 days at the time of the assessment.

Table 3. Types of tobacco products used by ever-established AUIs[a] (n=688) before first trying the THS[b] and at the time of assessment[c].

Tobacco products	Tobacco use before first trying THS			Current tobacco use at the time of assessment
	Ever tried, n (%); 95% CI	Ever-established use, n (%); 95% CI	Current use, n (%); 95% CI	Current use, n (%); 95% CI
Cigarettes	680 (98.8); 97.7-99.5	632 (91.9); 89.8-93.9	628 (91.3); 89.2-93.4	338 (49.1); 45.4-52.9
Cigars	428 (62.2); 58.6-65.8	125 (18.2); 15.3-21.1	114 (16.6); 13.8-19.4	92 (13.4); 10.8-15.9
Pipe filled with tobacco	151 (22.0); 18.9-25.0	32 (4.7); 3.1-6.2	18 (2.6); 1.4-3.8	17 (2.5); 1.3-3.6
Hookah	225 (32.7); 29.2-36.2	29 (4.2); 2.7-5.7	44 (6.4); 4.6-8.2	34 (4.9); 3.3-6.6
E-vapor products	493 (71.7); 68.3-75.0	298 (43.3); 39.6-47.0	229 (33.3); 29.8-36.8	161 (23.4); 20.2-26.6
Smokeless tobacco	174 (25.3); 22.0-28.5	96 (14.0); 11.4-16.5	38 (5.5); 3.8-7.2	28 (4.1); 2.6-5.6
Oral tobacco-derived nicotine products	143 (20.8); 17.8-23.8	36 (5.2); 3.6-6.9	41 (6.0); 4.2-7.7	27 (3.9); 2.5-5.4
Any tobacco product other than THS	685 (99.6); 98.7-99.9	659 (95.8); 94.3-97.3	658 (95.6); 94.1-97.2	461 (67); 63.5-70.5

[a]AUI: adult who used *IQOS*.

[b]THS: Tobacco Heating System.

[c]Participants sampled from the *IQOS* consumer database were primarily residing in Georgia, Virginia, North Carolina, or South Carolina.

XSL•FO
RenderX

Table 4. Types of tobacco products used by current established AUIs[a] (n=645) before first trying the THS[b] and at the time of assessment[c].

Tobacco products	Tobacco use before first trying THS			Current tobacco use at the time of assessment
	Ever tried (n=645), n (%); 95% CI	Ever-established use (n=645), n (%); 95% CI	Current use (n=645), n (%); 95% CI	Current use (n=645), n (%); 95% CI
Cigarettes	638 (98.9); 97.8 to 99.6	591 (91.6); 89.5 to 93.8	586 (90.9); 88.6 to 93.1	316 (49.0); 45.1 to 52.9
Cigars	396 (61.4); 57.6 to 65.2	118 (18.3); 15.3 to 21.3	106 (16.4); 13.6 to 19.3	86 (13.3); 10.7 to 16.0
Pipe filled with tobacco	138 (21.4); 18.2 to 24.6	30 (4.7); 3.0 to 6.3	17 (2.6); 1.4 to 3.9	17 (2.6); 1.4 to 3.9
Hookah	209 (32.4); 28.8 to 36.0	27 (4.2); 2.6 to 5.7	41 (6.4); 4.5 to 8.2	33 (5.1); 3.4 to 6.8
E-vapor products	462 (71.6); 68.2 to 75.1	275 (42.6); 38.8 to 46.5	211 (32.7); 29.1 to 36.3	144 (22.3); 19.1 to 25.5
Smokeless tobacco	157 (24.3); 21 to 27.7	87 (13.5); 10.9 to 16.1	36 (5.6); 3.8 to 7.4	26 (4.0); 2.5 to 5.6
Oral tobacco-derived nicotine products	128 (19.8); 16.8 to 22.9	32 (5.0); 3.3 to 6.6	35 (5.4); 3.7 to 7.2	23 (3.6); 2.1 to 5.0
Any tobacco product other than THS	643 (99.7); 98.9 to >99.9	617 (95.7); 94.1 to 97.2	616 (95.5); 93.9 to 97.1	428 (66.4); 62.7 to 70.0

[a]AUI: adult who used *IQOS*.

[b]THS: Tobacco Heating System.

[c]Participants sampled from the *IQOS* consumer database were primarily residing in Georgia, Virginia, North Carolina, or South Carolina.

THS Use

Overall Use Behaviors

On average, AUIs had used THS for 1.2 years (mean 1.2, 95% CI 1.1-1.2 for all established AUIs; mean 1.2, 95% CI 1.1-1.3 for current AUIs). Among current AUIs (n=645), one-third were using THS only (n=217, 33.6%; 95% CI 30.0%-37.3%), 280 (43.4%; 95% CI 39.6%-47.2%) were using it and 1 other tobacco product (including 187, 29.0%, who were smoking cigarettes and 93, 14.4%, who were not smoking cigarettes), and 148 (23.0%; 95% CI 19.7%-26.2%) were using it and at least two other tobacco products (including 129, 20.0%, who were smoking cigarettes and 19, 2.9%, who were not smoking cigarettes). Of current AUIs, 443 (68.7%) used THS every day during the 30 days before the assessment. A greater proportion of current established THS users who had used the device for more than a year used THS daily compared with those who had used THS for a year or less (218/300, 72.7%, vs 225/345, 65.2%, P=.04). Current AUIs used a median of 15 tobacco sticks per day on days used (Table 5).

XSL•FO
RenderX

Table 5. Patterns of Tobacco Heating System use in the past 30 days at the time of the assessment[a].

Patterns	Current established AUIs[b] (n=645)	Current established AUIs who prefer menthol tobacco sticks (n=337)	Current established AUIs who prefer nonmenthol tobacco sticks (n=308)
Number of days used, n (%); 95% CI			
0-9	36 (5.6); 3.8-7.4	14 (4.2); 2.0-6.3	22 (7.1); 4.3-10.0
10-19	66 (10.2); 7.9-12.6	31 (9.2); 6.1-12.3	35 (11.4); 7.8-14.9
20-29	100 (15.5); 12.7-18.3	52 (15.4); 11.6-19.3	48 (15.6); 11.5-19.6
30	443 (68.7); 65.1-72.3	240 (71.2); 66.4-76.1	203 (65.9); 60.6-71.2
Number of days used, mean (95% CI)	25.9 (25.3-26.5)	26.3 (25.6-27.1)	25.4 (24.5-26.2)
Number of days used, median (IQR)	30 (25-30)	30 (26-30)	30 (22-30)
Number of tobacco sticks used per day on days used, n (%); 95% CI			
<1	7 (1.1); 0.4-2.2	5 (1.5); 0.5-3.4	2 (0.7); 0.1-2.3
1	4 (0.6); 0.2-1.6	2 (0.6); 0.1-2.1	2 (0.7); 0.1-2.3
2	9 (1.4); 0.6-2.6	5 (1.5); 0.5-3.4	4 (1.3); 0.4-3.3
3	17 (2.6); 1.4-3.9	9 (2.7); 1.2-5	8 (2.6); 1.1-5.1
4	10 (1.55); 0.6-2.5	4 (1.2); 0.3-3	6 (2); 0.7-4.2
5-9	104 (16.1); 13.3-19.0	55 (16.3); 12.4-20.3	49 (15.9); 11.8-20
10-14	146 (22.6); 19.4-25.9	79 (23.4); 18.9-28	67 (21.8); 17.2-26.4
15-19	94 (14.6); 11.9-17.3	49 (14.5); 10.8-18.3	45 (14.6); 10.7-18.6
20+	250 (38.8); 35.0-42.5	126 (37.4); 32.2-42.6	124 (40.3); 34.8-45.7
Number of tobacco sticks per day on days used, median (IQR)	15 (10-20)	15 (10-20)	15 (10-20)
Number of tobacco sticks per day, median (IQR)	13 (7-20)	12.5 (8-20)	14 (7-20)

[a]Participants sampled from the *IQOS* consumer database were primarily residing in Georgia, Virginia, North Carolina, or South Carolina.

[b]AUI: adult who used *IQOS*.

Menthol and Nonmenthol Tobacco Stick Preferences

The majority of current AUIs had tried all 3 varieties of tobacco sticks (461/645, 71.5%; 437/645, 67.8%; and 359/645, 55.7%, for amber; green menthol; and blue menthol, respectively) by the time of the assessment. Amber was the most common variety tried first (344/645, 53.3%; 202/645, 31.3%; and 99/645, 15.3%, for green menthol and blue menthol, respectively); at the time of the assessment, nearly half indicated amber tobacco sticks as the variety currently used most often (308/645, 47.8%; 191/645, 29.6%; and 146/645, 22.6%, for green menthol and blue menthol, respectively).

When stratified by menthol and nonmenthol cigarette preference, we observed a clear preference for menthol tobacco sticks among participants who smoked menthol cigarettes most often. For example, among participants who formerly smoked menthol cigarettes most often (n=145), 135 (93.1%) used either green menthol or blue menthol tobacco sticks most often. Among participants who formerly smoked nonmenthol cigarettes most often (n=195), 144 (73.8%) used amber tobacco sticks most often. Similar patterns were observed among participants who currently smoked cigarettes: Among the 127 participants who currently smoked menthol cigarettes most often, 121 (95.3%) used menthol tobacco sticks most often; among the 204 participants who currently smoked nonmenthol cigarettes most often, 161 (78.9%) used nonmenthol tobacco sticks most often. There was little difference in the frequency and amount of tobacco stick use between those who preferred menthol and nonmenthol tobacco sticks (Table 5).

Perceptions of THS

THS was perceived as having lower health risk as compared with cigarette smoking: the mean composite score of the Perceived Health Risk scale was 43.8 (95% CI 42.6-45.1) for THS and 64.4 (95% CI 63.0-65.7) for cigarette smoking.

With respect to understanding the MRTP message, among the 688 ever-established AUIs, 555 (80.7%; 95% CI 77.7%-83.6%) correctly identified “less exposure” to harmful or potentially harmful chemicals when switching completely from cigarettes to THS. Only 33 (4.8%; 95% CI 3.2%-6.4%) perceived “no exposure.” Other participants answered “the same exposure” (59/688, 8.6%; 95% CI 6.5%-10.7%), “more exposure” (6/688, 0.9%, 95% CI 0.3%-1.8%), or they “do not know” (35/688, 5.1%, 95% CI 3.5%-6.7%). Among those who correctly identified “less exposure” (n=555), 471 (84.9%; 95% CI 81.9%-87.9%) understood that “smokers must stop smoking

XSL•FO
RenderX

completely and only use *IQOS*" to reduce their exposure to harmful or potentially harmful chemicals. Only 6 (1.1%; 95% CI 0.4%-2.3%) selected "keep smoking the same amount of cigarettes and also use *IQOS*."

Switching to THS and Transitions to/Back to Cigarette Smoking

Complete Switching

Among all current established AUIs (n=645), 191 (29.6%; 95% CI 26.1%-33.1%) met our predefined "complete switching" from cigarette smoking criteria, that is, participants who (1) had smoked at least 100 cigarettes in their lifetime, (2) were currently smoking during the 30 days before first trying THS, (3) were not smoking at the time of assessment, and (4) smoked their last cigarettes after trying THS. To note, this estimate is best interpreted as the proportion of current AUIs who had switched from cigarettes to THS from a cross-sectional assessment. It should not be interpreted as a "switching rate" as it does not provide any information about the "speed" of transition. Moreover, 350 of 688 (50.9%) AUIs were not smoking at the time of assessment (ie, 338/688, 49.1%, AUIs were also smoking; see Tables 3 and 4). The majority of individuals who were not smoking but did not meet the "complete switching" criteria were those who indicated that the last time they smoked cigarettes was within 30 days *before* first trying THS. (To meet the "complete switching" criteria, participants needed to indicate the last time they smoked was *after* first trying THS.) It is possible that these individuals immediately switched and did not smoke a cigarette once they started using THS. Therefore, we consider our estimate of "complete switching" to be conservative because it does not include those who stopped smoking immediately upon first trying the product.

When stratified by daily and nondaily smoking during the 30 days before first trying THS, we found similar proportions of AUIs who completely switched from cigarette smoking to THS between daily (ie, every day) and nondaily (ie, someday) smokers (158/519, 30.4%, and 33/109, 30.3%, respectively). Among all AUIs, we did not observe any differences in proportions who met the "complete switching" criteria between those who used menthol tobacco sticks and those who used nonmenthol tobacco sticks most often (102/358, 28.5%, and 89/330, 27.0%, respectively). Participants who completely switched from cigarettes to THS had used THS for a longer duration (mean duration 15.6 months) compared with those who did not meet the complete switching definition (mean duration 13.6 months; estimated difference 2.0, 95% CI 0.4-3.7, P=.01).

Among AUIs who were also smoking at the time of assessment (ie, dual-use; n=298), 249 (83.6%; 95% CI 79.4%-87.8%) indicated that they smoked fewer cigarettes per day at the time of assessment; 7 (2.3%; 95% CI 1.0%-4.8%) indicated they smoked more and 42 (14.1%; 95% CI 10.1%-18.0%) indicated they smoked the same number of cigarettes per day at the time of assessment compared with the 30 days before trying THS.

Transition to or Back to Cigarettes

In this study, we observed no relapse of cigarette smoking (defined as the current use of cigarettes and having smoked at least 100 cigarettes and having had not smoked cigarettes for <12 months before first trying THS); we observed 1 individual who reinitiated cigarette smoking (defined as the current use of cigarettes and having smoked at least 100 cigarettes and having had not smoked cigarettes for at least 12 months before first trying THS).

Discussion

Principal Findings

To our knowledge, this study is the first to describe the characteristics of AUIs and comprehensively assess use behaviors and risk perceptions relevant to THS use in a real-world setting among established AUIs in the United States. We found that AUIs tended to be male, non–Hispanic White or non–Hispanic Asian, and middle-aged (age 35-54 years), with at least some college education and a household income over US $60,000. More than half of AUIs reported no physical health conditions and the majority (548/688, 79.7%) reported not receiving treatment for any mental health conditions. Our results showed that the vast majority of established AUIs were existing smokers with an average of over 20 years of smoking history; 570 of 645 (88.4%) had either never used or had tried but stopped using tobacco cessation treatment. After an average of 1.2 years of THS use, almost half of AUIs were not smoking at the time of the assessment, and more than 80% (156/187, 83.4%) of participants who were still smoking indicated they had reduced their cigarette consumption. In addition, almost one-third met our strict definition for "complete switching" from cigarettes to THS and none reported relapse back to smoking. Finally, 555 of 688 (80.7%) participants perceived THS use as less harmful compared with cigarette smoking; few perceived it as risk-free and the vast majority understood the THS-modified risk claim. These findings provide firsthand empirical evidence in a real-world setting on THS consumers and their use behaviors during a relatively early stage of adoption in the United States.

Interpretation of Findings

These findings are in line with evidence from other countries in that the vast majority of AUIs smoked cigarettes when they first tried THS, and few had never smoked a cigarette [24,25]. Our findings suggested that THS can facilitate adults who smoked cigarettes—especially those who did not use cessation treatment or did not successfully quit smoking after trying cessation treatment—to switch to THS.

Previous studies have found that young adults were more likely to use e-vapor products compared with older adults [26,27]. By contrast, we found that those who used THS tended to be middle-aged or older adults with a relatively long smoking history. In addition, the majority of current established AUIs had never tried a cessation treatment and another 202 of 645 (31.3%) tried but discontinued cessation treatment more than 12 months ago. Therefore, THS may be particularly beneficial to those who are older, have a long history of smoking, and did not use or were not successful in using tobacco cessation treatment. Along with the finding of minimal uptake among individuals who had never smoked cigarettes, our results support

XSL•FO
RenderX

the FDA's conclusion that marketing this product is appropriate for the protection of public health.

With respect to product use, the majority of AUIs used it daily and used more than 15 tobacco sticks per day on days used. This level of consumption was generally in line with cigarette consumption among smokers in national surveys [28-30], and daily use of THS has been associated with switching [31,32]. Taken together, these results suggest the replacement of cigarettes with THS use.

The correct perception of risk can play an integral role in switching behaviors. In line with the large body of literature showing that individuals who perceive smoke-free tobacco products as less harmful compared with cigarettes are more likely to switch to such products from cigarette smoking [33,34], a previous study documented that individuals who used THS exclusively (rather than smoking) were more likely to perceive it as less harmful than cigarettes, compared with those who used THS and smoked cigarettes (ie, dual-use) [31]. Given this context, it is encouraging to observe that over 80% (555/688, 80.7%) of participants in this study correctly identified the reduced exposure to harmful or potentially harmful chemicals when switching completely from cigarettes to THS, and few participants incorrectly identified "no exposure."

With respect to tobacco stick varieties, we found an interesting pattern suggesting that menthol tobacco sticks were used and preferred not only by individuals who preferred menthol cigarettes, but also by a sizable portion of those who preferred nonmenthol cigarettes. That is, although amber (or original) was the most commonly ever used (456/645, 70.7%) and first tried (342/645, 53.0%) variety, the 2 menthol varieties accounted for the majority of varieties used most often at the time of assessment. This aligns with the observation that approximately 1 in 5 individuals who preferred nonmenthol cigarettes preferred menthol tobacco sticks; by contrast, only 1 in 20 individuals who preferred menthol cigarettes preferred amber (or original). Although it is not clear why menthol tobacco sticks were more commonly preferred than amber, such findings support the potential role of menthol varieties in continued use. In this study, we found that ever and current use of other tobacco products was more common among participants of this study (with e-vapor products being the most common) compared with that among smokers based on estimates from national surveys [35]. This may suggest that individuals open to trying and using other tobacco products were more likely to try and use THS. When compared with the time before trying THS, we observed a ~30% (229/688 vs 161/688) reduction in e-cigarette use. It is possible that some participants were using e-vapor products to help stop cigarette smoking but were not successful; after initiating THS use, they stopped using e-vapor products. Future studies that assess reasons to stop using other tobacco products will shed new light on this topic.

Limitations

Our findings should be interpreted with the following limitations in mind. First, the study is based on self-reported information. Although we are not aware of any validation studies on self-reported use behaviors, the larger literature supports the reasonable validity of self-reported behaviors in tobacco research [36]. In addition, many measures relied on recalled information, so recall bias cannot be ruled out. We tried to minimize recall bias by (1) asking about prominent behaviors generally easy to recall, such as whether they smoked cigarettes during the 30 days before first trying THS, and (2) limiting the length of the recall period (eg, the past 30 days). Second, it is estimated that approximately 70% of individuals who purchased THS registered themselves in the database. Moreover, of all individuals (N=19,258) invited to complete the eligibility assessment, only 2022 (10.50%) did so. It is unknown whether individuals who registered in the database or responded to the invitation are representative of all AUIs in the United States. Also, due to limited distribution, it is not clear whether these results would apply to all adults who smoke in the United States. Nonetheless, we provided the first estimates of perceptions and behaviors related to THS use from a sample large enough to produce meaningful results. Third, the study was conducted among AUIs, so we cannot provide information about the prevalence of THS use in the general US population. Fourth, no definitive evidence can be drawn for causal relationships from this cross-sectional study. Perhaps future longitudinal studies can explore changes in behavior over time among individuals who adopt THS. Finally, it is noteworthy that during the recruitment period, a notification was sent to consumers on October 13, 2021, informing them that the product would become unavailable after November 28, 2021. In consideration of the potential impact of this communication on consumer behaviors, we also conducted analyses only using data collected before that date (n=463). Results from this subsample were generally in line with the results for the full sample.

Conclusions

Our results provide supportive evidence for the potential of THS to help individuals switch from smoking by comprehensively assessing use behaviors and risk perceptions in a real-world setting among established AUIs in the United States.

Acknowledgments

The study was funded internally and did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors thank Dr. Annie Heremans for study design input and reviewing early drafts of the manuscript.

Availability of Data and Materials

The datasets used or analyzed during this study are available from the corresponding author upon reasonable request.

XSL•FO
RenderX

Authors' Contributions

All authors designed the study. HGC, BN, and ARV acquired the data. HGC prepared the first draft. All authors reviewed and revised the manuscript. All authors have approved the final version of the manuscript.

Conflicts of Interest

HCG is a former and BN, ARV, and EGL are current employees of Altria Client Services LLC, a service company to Altria and its tobacco operating companies. PM is an employee of Philip Morris International, a tobacco manufacturer. Both companies were involved in the review and approval of the manuscript for publication.

References

1. GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1223-1249 [FREE Full text] [doi: 10.1016/S0140-6736(20)30752-2] [Medline: 33069327]
2. Smoking cessation: a report of the surgeon general- executive summary. US Department of Health and Human Services. Atlanta Georgia: US Department of Health and Human Services; 2020. URL: https://www.hhs.gov/sites/default/files/2020-cessation-sgr-executive-summary.pdf [accessed 2024-11-27]
3. Babb S, Malarcher A, Schauer G, Asman K, Jamal A. Quitting smoking among adults - United States, 2000-2015. MMWR Morb Mortal Wkly Rep 2017 Jan 06;65(52):1457-1464 [FREE Full text] [doi: 10.15585/mmwr.mm6552a1] [Medline: 28056007]
4. García-Rodríguez O, Secades-Villa R, Flórez-Salamanca L, Okuda M, Liu S, Blanco C. Probability and predictors of relapse to smoking: results of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). Drug Alcohol Depend 2013 Oct 01;132(3):479-485 [FREE Full text] [doi: 10.1016/j.drugalcdep.2013.03.008] [Medline: 23570817]
5. Fiore M, Jaén C, Baker T. Treating Tobacco Use and Dependence: 2008 Update Clinical Practice Guideline. Rockville, MD: U.S. Department of Health and Human Services. Public Health Service; 2008.
6. Centers for Disease ControlPrevention (CDC). Quitting smoking among adults--United States, 2001-2010. MMWR Morb Mortal Wkly Rep 2011 Nov 11;60(44):1513-1519 [FREE Full text] [Medline: 22071589]
7. Borland R, Partos TR, Yong H, Cummings KM, Hyland A. How much unsuccessful quitting activity is going on among adult smokers? Data from the International Tobacco Control Four Country cohort survey. Addiction 2012 Mar;107(3):673-682 [FREE Full text] [doi: 10.1111/j.1360-0443.2011.03685.x] [Medline: 21992709]
8. Chaiton M, Diemert L, Cohen JE, Bondy SJ, Selby P, Philipneri A, et al. Estimating the number of quit attempts it takes to quit smoking successfully in a longitudinal cohort of smokers. BMJ Open 2016 Jun 09;6(6):e011045 [FREE Full text] [doi: 10.1136/bmjopen-2016-011045] [Medline: 27288378]
9. U.S. Department of Health And Human Services (U.S. DHHS). How Tobacco Smoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General. Atlanta, GA: U.S. Department of Health And Human Services, Public Health Service, Office of the Surgeon General; 2010.
10. McNeill A. Harm reduction. BMJ 2004 Apr 10;328(7444):885-887 [FREE Full text] [doi: 10.1136/bmj.328.7444.885] [Medline: 15073074]
11. Abrams DB, Glasser AM, Pearson JL, Villanti AC, Collins LK, Niaura RS. Harm minimization and tobacco control: reframing societal views of nicotine use to rapidly save lives. Annu Rev Public Health 2018 Apr 01;39:193-213 [FREE Full text] [doi: 10.1146/annurev-publhealth-040617-013849] [Medline: 29323611]
12. Gottlieb S, Zeller M. A nicotine-focused framework for public health. N Engl J Med 2017 Sep 21;377(12):1111-1114. [doi: 10.1056/NEJMp1707409] [Medline: 28813211]
13. Simonavicius E, McNeill A, Shahab L, Brose LS. Heat-not-burn tobacco products: a systematic literature review. Tob Control 2019 Sep 04;28(5):582-594 [FREE Full text] [doi: 10.1136/tobaccocontrol-2018-054419] [Medline: 30181382]
14. Lüdicke F, Baker G, Magnette J, Picavet P, Weitkunat R. Reduced exposure to harmful and potentially harmful smoke constituents with the Tobacco Heating System 2.1. Nicotine Tob Res 2017 Feb;19(2):168-175 [FREE Full text] [doi: 10.1093/ntr/ntw164] [Medline: 27613951]
15. Haziza C, de La Bourdonnaye G, Merlet S, Benzimra M, Ancerewicz J, Donelli A, et al. Assessment of the reduction in levels of exposure to harmful and potentially harmful constituents in Japanese subjects using a novel tobacco heating system compared with conventional cigarettes and smoking abstinence: a randomized controlled study in confinement. Regul Toxicol Pharmacol 2016 Nov;81:489-499 [FREE Full text] [doi: 10.1016/j.yrtph.2016.09.014] [Medline: 27693654]
16. Tabuchi T, Gallus S, Shinozaki T, Nakaya T, Kunugita N, Colwell B. Heat-not-burn tobacco product use in Japan: its prevalence, predictors and perceived symptoms from exposure to secondhand heat-not-burn tobacco aerosol. Tob Control 2018 Jul 16;27(e1):e25-e33 [FREE Full text] [doi: 10.1136/tobaccocontrol-2017-053947] [Medline: 29248896]
17. Afolalu EF, Langer P, Fischer K, Roulet S, Magnani P. Prevalence and patterns of tobacco and/or nicotine product use in Japan (2017) after the launch of a heated tobacco product (IQOS(R)): a cross-sectional study. F1000Res 2021;10:504 [FREE Full text] [doi: 10.12688/f1000research.52407.2] [Medline: 35528952]

XSL•FO
RenderX

18. Sakaguchi C, Nagata Y, Kikuchi A, Takeshige Y, Minami N. Differences in levels of biomarkers of potential harm among users of a heat-not-burn tobacco product, cigarette smokers, and never-smokers in Japan: a post-marketing observational study. Nicotine Tob Res 2021 Jun 08;23(7):1143-1152 [FREE Full text] [doi: 10.1093/ntr/ntab014] [Medline: 33502518]
19. Polosa R, Morjaria JB, Prosperini U, Busà B, Pennisi A, Gussoni G, et al. Health outcomes in COPD smokers using heated tobacco products: a 3-year follow-up. Intern Emerg Med 2021 Apr 23;16(3):687-696 [FREE Full text] [doi: 10.1007/s11739-021-02674-3] [Medline: 33754228]
20. Insights Association Code of Standards and Ethics for Market Research and Data Analytics. URL: https://www.insightsassociation.org/Portals/INSIGHTS/Insights%20Association%20Code%20of%20Standards_November%202023_v2.pdf [accessed 2024-12-06]
21. ICC/ESOMAR International Code on Market, Opinion and Social Research and Data Analytics. URL: https://esomar.org/uploads/attachments/ckqtawvjq00uukdtrhst5sk9u-iccesomar-international-code-english.pdf [accessed 2024-12-06]
22. Cano S, Chrea C, Salzberger T, Alfieri T, Emilien G, Mainy N, et al. Development and validation of a new instrument to measure perceived risks associated with the use of tobacco and nicotine-containing products. Health Qual Life Outcomes 2018 Sep 21;16(1):192 [FREE Full text] [doi: 10.1186/s12955-018-0997-5] [Medline: 30241527]
23. Noggle B, Cheng H, Iyer S, Vansickel A. IQOS® cross-sectional and cohort US study documentation. Qeios 2023:e1 [FREE Full text] [doi: 10.32388/vkd77j]
24. Sutanto E, Miller C, Smith DM, O'Connor RJ, Quah ACK, Cummings KM, et al. Prevalence, use behaviors, and preferences among users of heated tobacco products: findings from the 2018 ITC Japan Survey. Int J Environ Res Public Health 2019 Nov 21;16(23):4630 [FREE Full text] [doi: 10.3390/ijerph16234630] [Medline: 31766410]
25. Gallus S, Lugo A, Liu X, Borroni E, Clancy L, Gorini G, TackSHS Project Investigators. Use and awareness of heated tobacco products in Europe. J Epidemiol 2022 Mar 05;32(3):139-144 [FREE Full text] [doi: 10.2188/jea.JE20200248] [Medline: 33456019]
26. Levy DT, Yuan Z, Li Y. The prevalence and characteristics of e-cigarette users in the U.S. Int J Environ Res Public Health 2017 Oct 11;14(10):1200 [FREE Full text] [doi: 10.3390/ijerph14101200] [Medline: 29019917]
27. Bandi P, Cahn Z, Goding Sauer A, Douglas CE, Drope J, Jemal A, et al. Trends in e-cigarette use by age group and combustible cigarette smoking histories, U.S. adults, 2014-2018. Am J Prev Med 2021 Feb;60(2):151-158. [doi: 10.1016/j.amepre.2020.07.026] [Medline: 33032869]
28. Cornelius ME, Loretan CG, Wang TW, Jamal A, Homa DM. Tobacco product use among adults - United States, 2020. MMWR Morb Mortal Wkly Rep 2022 Mar 18;71(11):397-405 [FREE Full text] [doi: 10.15585/mmwr.mm7111a1] [Medline: 35298455]
29. American Lung Association. Tobacco trends brief. American Lung Association. 2022. URL: https://www.lung.org/research/trends-in-lung-disease/tobacco-trends-brief/overall-smoking-trends [accessed 2024-11-27]
30. Substance Abuse and Mental Health Services Administration (SAMHSA). Results from the 2020 National Survey on Drug Use and Health: detailed tables. SAMHSA. Rockville, MD: SAMHSA; 2020. URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt35323/NSDUHDetailedTabs2020v25/NSDUHDetailedTabs2020v25/NSDUHDetTabsSect2pe2020.htm [accessed 2024-11-27]
31. Sutanto E, Miller C, Smith DM, Borland R, Hyland A, Cummings KM, et al. Concurrent daily and non-daily use of heated tobacco products with combustible cigarettes: findings from the 2018 ITC Japan Survey. Int J Environ Res Public Health 2020 Mar 22;17(6):2098 [FREE Full text] [doi: 10.3390/ijerph17062098] [Medline: 32235702]
32. Baig SA, Giovenco DP. Behavioral heterogeneity among cigarette and e-cigarette dual-users and associations with future tobacco use: findings from the Population Assessment of Tobacco and Health Study. Addict Behav 2020 May;104:106263 [FREE Full text] [doi: 10.1016/j.addbeh.2019.106263] [Medline: 32028096]
33. Persoskie A, O'Brien EK, Poonai K. Perceived relative harm of using e-cigarettes predicts future product switching among US adult cigarette and e-cigarette dual users. Addiction 2019 Dec;114(12):2197-2205. [doi: 10.1111/add.14730] [Medline: 31278802]
34. Brose LS, Brown J, Hitchman SC, McNeill A. Perceived relative harm of electronic cigarettes over time and impact on subsequent use. A survey with 1-year and 2-year follow-ups. Drug Alcohol Depend 2015 Dec 01;157:106-111 [FREE Full text] [doi: 10.1016/j.drugalcdep.2015.10.014] [Medline: 26507173]
35. Li L, Borland R, Cummings KM, Gravely S, Quah ACK, Fong GT, et al. Patterns of non-cigarette tobacco and nicotine use among current cigarette smokers and recent quitters: findings from the 2020 ITC Four Country Smoking and Vaping Survey. Nicotine Tob Res 2021 Aug 18;23(9):1611-1616 [FREE Full text] [doi: 10.1093/ntr/ntab040] [Medline: 33693833]
36. Patrick DL, Cheadle A, Thompson DC, Diehr P, Koepsell T, Kinne S. The validity of self-reported smoking: a review and meta-analysis. Am J Public Health 1994 Jul;84(7):1086-1093. [doi: 10.2105/ajph.84.7.1086] [Medline: 8017530]

Abbreviations

AUI: adult who used IQOS
CASRO: Council of American Survey Research Organization
CDO: Cease-and-Desist Order

XSL•FO
RenderX

ENDS: electronic nicotine delivery system
ESOMAR: European Society for Opinion and Marketing Research
FDA: Food and Drug Administration
ITC: United States International Trade Commission
MRTP: modified risk tobacco product

Edited by A Mavragani; submitted 15.02.24; peer-reviewed by A Hassan, K Cummings; comments to author 16.07.24; revised version received 13.09.24; accepted 11.10.24; published 07.02.25.

Please cite as:
Cheng HG, Noggle B, Vansickel AR, Largo EG, Magnani P
Tobacco Use, Risk Perceptions, and Characteristics of Adults Who Used a Heated Tobacco Product (IQOS) in the United States: Cross-Sectional Survey Study
JMIR Form Res 2025;9:e57398
URL: https://formative.jmir.org/2025/1/e57398
doi:10.2196/57398
PMID:

©Hui G Cheng, Brendan Noggle, Andrea R Vansickel, Edward G Largo, Pierpaolo Magnani. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Predictors of the Intention to Stop Using Smart Devices at Bedtime Among University Students in Saudi Arabia: Cross-Sectional Survey

Manal Almalki[1], PhD

Public Health Department, College of Nursing and Health Sciences, Jazan University, Jazan, Saudi Arabia

Corresponding Author:
Manal Almalki, PhD
Public Health Department
College of Nursing and Health Sciences
Jazan University
1st Fl
Al Maarefah Rd
Jazan, 45142
Saudi Arabia
Phone: 966 173290000 ext 5548
Email: almalki@jazanu.edu.sa

Abstract

Background: The widespread use of smart devices, particularly among university students, has raised concerns about their impact on sleep quality. Bedtime usage of smart devices is associated with sleep disruptions and poor sleep quality.

Objective: This study aimed to explore the behavioral and perceptual factors influencing university students' intention to stop using smart devices at bedtime in Saudi Arabia.

Methods: A cross-sectional survey was conducted in June 2024 and distributed via social media platforms to university students (aged ≥18 years). The questionnaire collected data on demographics, smart device usage habits, perceived negative effects on sleep, and physical sleep disturbances. The Pittsburgh Sleep Quality Index was used to assess sleep quality. Path analysis was performed to evaluate relationships between the outcome variables, intended to stop using smart device usage, and 3 latent variables: sleep quality smartphone usage, sleep quality perceived negative effects, and sleep quality during the past month. Model fit was assessed using chi-square, comparative fit index, and root mean square error of approximation.

Results: Of the 774 participants, 90.43% (700/774) reported using their smart devices every night and 72.48% (561/774) believed bedtime device use negatively affected them the next morning. The most frequently reported next-morning symptoms were fatigue or drowsiness (480/774, 62.01%). Common purposes for bedtime device use were staying in touch with friends or family (432/774, 55.81%), entertainment (355/774, 45.86%), and filling up spare time (345/774, 44.57%). Overall, 58.26% (451/774) expressed an intention to stop bedtime device use within the next 3 months. Path analysis demonstrated that frequent nightly use (path coefficient=0.36) and after-lights-off usage (0.49) were positively associated with the intention to stop, whereas spending ≥3 hours on devices (–0.35) and engaging in multiple activities (–0.18) had negative associations. The strongest predictors of the intention to stop were perceived negative effects on next-morning well-being (0.71) and difficulty breathing comfortably during sleep (0.64). Model fit was excellent (comparative fit index=0.845 and root mean square error of approximation=0.039).

Conclusions: Perceived negative effects on sleep quality and physical sleep disturbances are strong predictors of the intention to stop using smart devices at bedtime among university students in Saudi Arabia. Interventions aimed at improving sleep hygiene should focus on raising awareness about the impact of smart device use on well-being and addressing behaviors such as late-night usage and heavy screen time. Public health strategies should target both psychological and physiological aspects of bedtime smart device use to improve sleep quality in this population.

(JMIR Form Res 2025;9:e67223) doi:10.2196/67223

KEYWORDS

smart devices; smartphone; digital health; digital technology; sleep quality; university student; bedtime habits; Saudi Arabia; path analysis; sleep disturbances; well-being; usage; intention; behavior; mobile phone

XSL•FO
RenderX

Introduction

In the digital age, smart devices such as smartphones, tablets, and laptops have become integral parts of daily life, especially among university students [1]. These devices offer constant connectivity, enabling students to engage in academic activities, communicate with peers, and access entertainment. In Saudi Arabia, where more than 97% of the population uses smartphones [2], smart devices have become an essential tool for both learning and social interaction. On a global scale, the number of smartphone mobile network subscriptions exceeded 7 billion in 2023 and is anticipated to exceed 7.7 billion by 2028 [3].

University students, often facing academic pressures and irregular schedules, are particularly vulnerable to the excessive use of smart devices at bedtime [4-6]. They may engage in multiple bedtime activities, such as social media browsing, texting, gaming, or even frequent awakenings to check notifications, which counteract the body's natural inclination to relax and fall asleep [7-9]. These behaviors are further intensified by the immediate gratification of staying connected and creating a cycle that sustains poor sleep practices [8-10]. In Saudi Arabia, smart devices often serve as a primary means of students' communication, relaxation, and staying in touch with friends and family [11], which further reinforces their use, even during late-night hours [12].

The excessive use of these smart devices, particularly during bedtime, raises concerns about their impact on students' sleep quality [4], overall well-being [5], and academic performance [6]. Late-night screen exposure can disrupt the body's circadian rhythm through both blue light emissions and heightened mental stimulation [13], often delaying sleep onset and reducing total sleep duration [14-17]. Over time, these disruptions can lead to chronic sleep debt, making students more prone to daytime drowsiness, poor concentration [4,5], and lower academic achievement [14,15]. Moreover, poor sleep quality can contribute to the development of mental health issues such as anxiety and depression [5], while these conditions can further disrupt sleep patterns [16]. This mental stimulation can lead to increased cognitive arousal and addiction, which is detrimental to initiating and maintaining sleep [7,8,17]. This addiction does not only affect the quantity of sleep but also its quality, leading to a less restorative sleep experience [1,4,18].

Despite growing awareness of the potential harms of excessive screen time, many students continue to use smart devices at bedtime [13]. Understanding the factors that drive students to continue or stop using these devices at bedtime is crucial for addressing the sleep-related issues prevalent in this population [19,20]. According to the Theory of Planned Behavior [21], individuals' intentions to engage in or discontinue a behavior are shaped by their attitudes toward the behavior, perceived control over it, and social norms. In the context of smart device use, it is likely that students' perceptions of how device use affects their sleep and well-being, as well as their ingrained habits, play a key role in their decision to reduce or stop using smart devices at bedtime.

This study aims to examine the factors that predict the intention to stop using smart devices at bedtime among university students in Saudi Arabia. Specifically, it investigates the relationship between 3 key domains: (1) smart device usage behaviors (eg, frequency, duration, and activities performed on devices), (2) perceived negative effects of smart device use on sleep quality and next-morning well-being, and (3) physical sleep disturbances experienced during the past month. Through path analysis, this study seeks to understand the complex interplay between these variables and provide insights that can inform public health interventions targeted at improving sleep hygiene among Saudi university students.

Methods

Study Design

In this study, a cross-sectional survey of university students was conducted in June 2024 in Saudi Arabia. The STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines of cross-sectional studies were followed to report our study's findings [22]. The questionnaire was piloted with a small group of university students (n=7) to ensure clarity and relevance, with feedback used to refine its wording and structure.

Study Recruitment

Participants were recruited using convenience sampling. A link to the questionnaire was posted on social media platforms such as WhatsApp (Meta), Snapchat, Telegram, and X (formerly known as Twitter). These platforms were chosen due to their popularity among university students in Saudi Arabia [23,24]. The link was accompanied by a recruitment message that outlined the study's objectives, eligibility criteria, and an assurance of confidentiality. Eligible participants were students aged 18 years or older, currently enrolled in universities in Saudi Arabia, and capable of providing informed consent before starting the survey.

Ethical Considerations

Ethical approval for the study was obtained from the Standing Committee for Publication and Research Ethics at Jazan University (reference number REC-45/11/1143). On the opening page of the survey, participants were presented with information about the study's objectives and scope, followed by an informed consent statement. Participants were free to withdraw at any point without consequence. Also, participation was anonymous and no personally identifiable information was collected or used. No monetary or nonmonetary incentives were offered. Finally, this paper does not include any images or material in which participants can be identified.

Questionnaire Design

The questionnaire was created using a Google Form, available in both English and Arabic, and it consisted of 4 sections. The first section gathered demographic data such as gender, age, marital status, highest level of education, residential region, health status, and employment status.

The second section examined participants' smart device usage during bedtime during the past month. Questions covered the

XSL•FO
RenderX

type of smart device most frequently used, the average frequency of use per week, time spent on the device at bedtime (ie, categorized as “less than 2 hours” or “3 hours or more”), usage after lights were turned off, the use of headphones, and other related behaviors. This section also addressed the activities participants engaged in while using their devices (eg, chatting on messaging platforms, watching videos, and setting alarms or reminders), as well as the purposes of smart device use before sleep (eg, education, entertainment, business, or work networking).

The third section explored participants’ perceptions of how smart device usage at bedtime affected their sleep quality, sleep duration, and overall health. The final section used the Pittsburgh Sleep Quality Index (PSQI) to assess participants’ physical sleep disturbances experienced during the past month. The PSQI consists of 19 self-rated items that evaluate subjective sleep quality, sleep latency, habitual sleep efficiency, sleep disturbances, use of sleep medications, and daytime dysfunction [25]. The Arabic-translated version of the PSQI, which has been validated for use in research, was used in this study [26].

Statistical Analysis

Incomplete surveys were excluded. A power calculation using G*Power (α=.05, power=80%, moderate effect size, odds ratio [OR 1.5]) indicated a minimum sample size of 384 participants. The final sample size was 774, which exceeded this requirement, ensuring sufficient power to detect significant associations.

Categorical variables were summarized as frequencies and percentages. ORs with 95% CI values were calculated to assess the relations and their strengths between variables. A path analysis was then conducted to understand the complex relationships between the measured variable intended to stop using smart device usage, latent variables sleep quality smartphone usage (SQSU), sleep quality perceived negative effects (SQPNE), and sleep quality during the past month (SQDPM). This method was chosen because it allows for the simultaneous analysis of multiple dependent and independent variables, providing a comprehensive understanding of direct and indirect relationships between variables.

A typical path modal sequencing equation modeling (SEM) plot is constructed by defining a multiple set of regression formulas. Model fit was assessed by chi-square, comparative fit index (CFI), and root mean square error of approximation (RMSEA). The value of RMSEA less than 0.05 with CFI equal to 0.845 or higher with nonsignificant chi-square statistic indicates a good model fit. Data were coded, validated, and analyzed using IBM SPSS software, version 27.0 for Windows (IBM Corp). Plots were constructed using R software (version 4.1.2; R Core Team) with Lavaan and Semplot R packages. Two-tailed *P* value <.05 was considered to denote statistical significance.

Results

Participant Characteristics

A total of 774 participants completed the questionnaire, as shown in Table 1. Over half of them (451/774, 58.26%) expressed an intention to stop using smart devices during bedtime within the next 3 months. None of the demographic characteristics was associated with the acceptability of stopping the use of smart devices during bedtime; therefore, OR analysis is removed from Table 1.

The majority (397/774, 88.03%) were aged 24 years or younger, 74.72% (337/774) were female, and 86.92% (392/774) were unmarried. More than three-quarters of participants (657/774, 84.88%) had an undergraduate degree or higher, and 87.62% (283/323) did not intend to stop using smart devices at bedtime. Likewise, more than three-quarters (688/774, 88.89%) of participants who did not have any sleep disorders had 38% higher odds of intending to stop using smart devices at bedtime compared to participants with medically diagnosed sleep disorders. The most common sleep duration among all participants was 5-6 hours, with 34.88% (270/774) reporting this sleep range. Participants who intended to stop using smart devices at bedtime had a slightly higher percentage (176/451, 39.02%) of reporting 5-6 hours of sleep compared to those who did not intend to stop (94/323, 29.10%).

XSL•FO
RenderX

Table 1. Sample demographic characteristics of university students in Saudi Arabia.

Variables	Values (N=774)	Likely to stop using smart devices during bedtime in the next 3 months	
		Yes (n=451)	No (n=323)
Age (years), n (%)			
18-24	681 (87.98)	397 (88.03)	284 (87.93)
≥25	93 (12.02)	54 (11.97)	39 (12.07)
Sex, n (%)			
Male	190 (24.55)	114 (25.28)	76 (23.53)
Female	584 (75.45)	337 (74.72)	247 (76.47)
Marital status, n (%)			
Married	91 (11.76)	59 (13.08)	32 (9.91)
Not married	683 (88.24)	392 (86.92)	291 (90.09)
Highest level of education, n (%)			
Diploma after high school or below	117 (15.12)	77 (17.07)	40 (12.38)
Undergraduate studies or higher	657 (84.88)	374 (82.93)	283 (87.62)
Residential region, n (%)			
North	47 (6.07)	30 (6.65)	17 (5.26)
South	382 (49.35)	223 (49.45)	159 (49.23)
East	21 (2.71)	13 (2.88)	8 (2.48)
West	69 (8.91)	39 (8.65)	30 (9.29)
Middle	255 (32.94)	146 (32.37)	109 (33.75)
Medically diagnosed with any sleep disorders, n (%)			
Yes	86 (11.11)	56 (12.42)	30 (9.29)
No	688 (88.89)	395 (87.58)	293 (90.71)
Sleep hours per day or night, n (%)			
Less than 5 h	118 (15.25)	68 (15.08)	50 (15.48)
5-6 h	270 (34.88)	176 (39.02)	94 (29.10)
7-8 h	245 (31.65)	138 (30.60)	107 (33.13)
9-10 h	74 (9.56)	40 (8.87)	34 (10.53)
More than 10 h	67 (8.66)	29 (6.43)	38 (11.76)
Employment status, n (%)			
Employed	210 (27.13)	131 (29.05)	79 (24.46)
Not employed	564 (72.86)	320 (70.95)	244 (75.54)
Job has night shifts (n=210), n (%)			
Yes	51 (24.28)	32 (24.43)	19 (24.05)
No	159 (75.71)	99 (75.57)	60 (75.95)

Smart Device Usage Behaviors

More than a ninth decile of participants (737/774, 94.83%) used a smartphone, followed by 3.74% (29/774) using a tablet, and only 1.03% (8/774) using a laptop, as shown in Table 2. The groups who were less willing to stop using smart devices during their bedtime in the next 3 months, in descending order, were participants who used their devices 1-3 nights per week (OR 0.05, 95% CI 0.00-0.43; $P<.001$), participants who used their devices 4-6 nights per week (OR 0.48, 95% CI 0.25-0.90; $P=.02$), participants who did not frequently use devices after lights were turned off (OR 0.52, 95% CI 0.33-0.83; $P<.001$), participants who did not often use headphones during bedtime (OR 0.73, 95% CI 0.55-0.97; $P=.03$), and participants who were not frequently awakened by smart devices at night (OR 0.71, 95% CI 0.53-0.95; $P=.02$). These groups showed significantly lower odds of intending to stop using their smart devices during bedtime compared to the respective reference groups.

XSL•FO
RenderX

Table 2. Smart device usage behaviors among university students in Saudi Arabia.

Variables	Values (N=774)	Likely to stop using smart devices during bedtime in the next 3 months			
		Yes (n=451)	No (n=323)	Odds ratio (95% CI)	*P* value
Smart device type, n (%)					
Smartphone	737 (94.83)	434 (96.23)	303 (93.81)	Ref	—[a]
Laptop	8 (1.03)	2 (0.44)	6 (1.86)	4.29 (0.86-21.43)	.71
Tablet	29 (3.74)	15 (3.33)	14 (4.33)	1.33 (0.63-2.81)	.44
Frequency of using the smart device per week (SQSU[b]_1), n (%)					
Every night	700 (90.43)	392 (86.92)	308 (95.36)	Ref	—
4-6 nights	51 (6.58)	37 (8.20)	14 (4.33)	0.48 (0.25-0.90)	.02
1-3 nights	23 (2.97)	22 (4.88)	1 (0.31)	0.05 (0.00-0.43)	<.001
Time spent on a smart device at bedtime (SQSU_2), n (%)					
<2 h or less	494 (63.82)	309 (68.51)	185 (57.28)	Ref	—
3 h or more	280 (36.17)	142 (31.49)	138 (42.72)	1.62 (1.20-2.18)	<.001
Often using the smart device after lights are turned off (SQSU_3), n (%)					
Yes	674 (87.08)	380 (84.26)	294 (91.02)	Ref	—
No	100 (12.91)	71 (15.74)	29 (8.98)	0.52 (0.33-0.83)	<.001
Often using headphones during bedtime (SQSU_4), n (%)					
Yes	370 (47.80)	201 (44.57)	169 (52.32)	Ref	—
No	404 (52.19)	250 (55.43)	154 (47.68)	0.73 (0.55-0.97)	.03
Often awakened by the smart device at night (SQSU_5), n (%)					
Yes	309 (39.92)	165 (36.59)	144 (44.58)	Ref	—
No	465 (60.07)	286 (63.41)	179 (55.42)	0.71 (0.53-0.95)	.02
The most common posture when using the smart device, n (%)					
Lying down	668 (86.30)	383 (84.92)	285 (88.24)	Ref	—
Sitting	106 (13.69)	68 (15.08)	38 (11.76)	0.75 (0.49-1.14)	.20
Keeping the smart device while sleeping, n (%)					
In the bed	414 (53.48)	235 (52.11)	179 (55.42)	Ref	—
On the bedside table	316 (40.82)	186 (41.24)	130 (40.25)	0.91 (0.68-1.23)	.59
Far from the bed but inside the bedroom	44 (5.68)	30 (6.65)	14 (4.33)	0.61 (0.31-1.1)	.15
Most common smart device mode, n (%)					
Not general	586 (75.71)	336 (74.50)	250 (77.40)	Ref	—
General or vibratory	188 (24.28)	115 (25.50)	73 (22.60)	0.85 (0.60-1.19)	.39
Activities performed when using smart device during bedtime (SQSU_6), n (%)					
(1) Setting alarms, reminders, and tasks					
False	309 (39.92)	173 (38.36)	136 (42.11)	Ref	—
True	465 (60.07)	278 (61.64)	187 (57.89)	0.85 (0.63-1.14)	.29
(2) Chatting via WhatsApp, Snapchat, Telegram, etc					
False	269 (34.75)	170 (37.69)	99 (30.65)	Ref	—
True	505 (65.24)	281 (62.31)	224 (69.35)	1.36 (1.01-1.85)	.04
(3) Debating and discussion via Twitter, WhatsApp					
False	594 (76.74)	346 (76.72)	248 (76.78)	Ref	—
True	180 (23.25)	105 (23.28)	75 (23.22)	0.99 (0.71-1.39)	1

XSL•FO
RenderX

Variables	Values (N=774)	Likely to stop using smart devices during bedtime in the next 3 months			
		Yes (n=451)	No (n=323)	Odds ratio (95% CI)	*P* value
(4) Watching YouTube, TikTok, Instagram, etc					
False	276 (35.65)	178 (39.47)	98 (30.34)	Ref	—
True	498 (64.34)	273 (60.53)	225 (69.66)	1.49 (1.10-2.02)	<.001
(5) Watching TV shows and movies					
False	470 (60.72)	298 (66.08)	172 (53.25)	Ref	—
True	304 (39.27)	153 (33.92)	151 (46.75)	1.70 (1.27-2.29)	<.001
(6) Listening to the Qur'an and religious supplications					
False	606 (78.29)	349 (77.38)	257 (79.57)	Ref	—
True	168 (21.70)	102 (22.62)	66 (20.43)	0.87 (0.61-1.24)	.48
(7) Listening to a podcast					
False	709 (91.60)	411 (91.13)	298 (92.26)	Ref	—
True	65 (8.39)	40 (8.87)	25 (7.74)	0.86 (0.51-1.45)	.60
(8) Listening to music					
False	674 (87.08)	404 (89.58)	270 (83.59)	Ref	—
True	100 (12.91)	47 (10.42)	53 (16.41)	1.68 (1.10-2.57)	.01
(9) Researching brands					
False	701 (90.56)	412 (91.35)	289 (89.47)	Ref	—
True	73 (9.43)	39 (8.65)	34 (10.53)	1.24 (0.76-2.01)	.38
(10) Playing games					
False	657 (84.88)	391 (86.70)	266 (82.35)	Ref	—
True	117 (15.11)	60 (13.30)	57 (17.65)	1.39 (0.94-2.07)	.10
(11) Surfing the internet					
False	604 (78.03)	359 (79.60)	245 (75.85)	Ref	—
True	170 (21.96)	92 (20.40)	78 (24.15)	1.24 (0.88-1.74)	.21
(12) Browsing emails and text messages					
False	626 (80.87)	368 (81.60)	258 (79.88)	Ref	—
True	148 (19.12)	83 (18.40)	65 (20.12)	1.17 (0.77-1.60)	.57
Purposes of using the smart device during bedtime (SQSU_7) , n (%)					
(1) Education and study					
False	453 (58.52)	258 (57.21)	195 (60.37)	Ref	—
True	321 (41.47)	193 (42.79)	128 (39.63)	0.87 (0.65-1.17)	.41
(2) Staying in touch with friends and family					
False	342 (44.18)	204 (45.23)	138 (42.72)	Ref	—
True	432 (55.81)	247 (54.77)	185 (57.28)	1.10 (0.83-1.47)	.50
(3) Following news and current					
False	498 (64.34)	292 (64.75)	206 (63.78)	Ref	—
True	276 (35.65)	159 (35.25)	117 (36.22)	1.04 (0.77-1.40)	.81
(4) Sharing opinions and thoughts on social networks					
False	581 (75.06)	344 (76.27)	237 (73.37)	Ref	—
True	193 (24.93)	107 (23.73)	86 (26.63)	1.16 (0.83-1.62)	.39
(5) Business or work networking					

XSL•FO
RenderX

Variables	Values (N=774)	Likely to stop using smart devices during bedtime in the next 3 months			
		Yes (n=451)	No (n=323)	Odds ratio (95% CI)	*P* value
False	680 (87.85)	390 (86.47)	290 (89.78)	Ref	—
True	94 (12.14)	61 (13.53)	33 (10.22)	0.72 (0.46-1.14)	.18
(6) Researching how to do things					
False	660 (85.27)	386 (85.59)	274 (84.83)	Ref	—
True	114 (14.72)	65 (14.41)	49 (15.17)	1.06 (0.71-1.58)	.83
(7) Shopping					
False	615 (79.45)	359 (79.60)	256 (79.26)	Ref	—
True	159 (20.54)	92 (20.40)	67 (20.74)	1.02 (0.71-1.45)	.92
(8) Gaming					
False	619 (79.97)	374 (82.93)	245 (75.85)	Ref	—
True	155 (20.02)	77 (17.07)	78 (24.15)	1.54 (1.08-2.20)	.01
(9) Entertainment					
False	419 (54.13)	275 (60.98)	144 (44.58)	Ref	—
True	355 (45.86)	176 (39.02)	179 (55.42)	1.94 (1.45-2.59)	<.001
(10) Finding information					
False	603 (77.90)	365 (80.93)	238 (73.68)	Ref	—
True	171 (22.09)	86 (19.07)	85 (26.32)	1.51 (1.07-2.13)	.01
(11) Being relaxed and calming down					
False	569 (73.51)	337 (74.72)	232 (71.83)	Ref	—
True	205 (26.48)	114 (25.28)	91 (28.17)	1.15 (0.84-1.60)	.40
(12) Filling up spare time					
False	429 (55.42)	269 (59.65)	160 (49.54)	Ref	—
True	345 (44.57)	182 (40.35)	163 (50.46)	1.50 (1.12-2.00)	<.001

[a]Not applicable.

[b]SQSU: sleep quality smartphone usage.

The analysis of various activities performed on smart devices during bedtime and their impact on the intention to stop using these devices in the next 3 months is presented in Table 2. The 3 most common activities were chatting via messaging platforms like WhatsApp, Snapchat, and Telegram, with 65.24% (505/774) of participants reporting this behavior. Next, 64.34% (498/774) of participants indicated they spent their time watching content on platforms such as YouTube, TikTok, and Instagram. Setting alarms, reminders, and tasks was also a frequent activity, with 60.07% (465/774) of participants engaging in this before bed. On the other hand, the least common activity performed on smart devices during bedtime was listening to music (100/774, 12.91%). Participants who used their devices for watching TV shows and movies (OR 1.70, 95% CI 1.27-2.29), listening to music (OR 1.68, 95% CI 1.10-2.57), and watching videos on platforms like YouTube, TikTok, and Instagram (OR 1.49, 95% CI 1.10-2.02) were less willing to stop using smart devices during bedtime in the next 3 months.

The analysis of different purposes for using smart devices during bedtime and their impact on the intention to stop using these devices in the next 3 months is presented in Table 2. The most common purposes were staying in touch with friends and family (432/774, 55.81%), entertainment (355/774, 45.86%), and filling up spare time (201/774, 44.57%). Participants who used their devices for entertainment (OR 1.94, 95% CI 1.45-2.59), finding information (OR 1.51, 95% CI 1.07-2.13), filling up spare time (OR 1.50, 95% CI 1.12-2.00), gaming (OR 1.54, 95% CI 1.08-2.20), and chatting via WhatsApp, Snapchat, Telegram, etc (OR 1.36, 95% CI 1.01-1.85) were less willing to stop using smart devices during bedtime in the next 3 months.

Perceived Negative Effects of Smart Device Use on Sleep Quality and Next-Morning Well-Being

Over one-third of the participants (270/774, 35.23%) reported having poor sleep quality, as shown in Table 3. The OR analysis indicated that the perception of the negative influence of smartphone usage on sleep quality strongly impacted the intention to stop using these devices at bedtime. Participants who did not believe that using smart devices during bedtime negatively affected them the next morning were more likely to not intend to stop using their devices compared to those who did believe it (OR 3.53, 95% CI 2.53-4.91, *P*<.001). Similarly,

XSL•FO
RenderX

participants who did not believe that smart device use during bedtime made them sleep less were 2.76 times more likely to not intend to stop (OR 2.76, 95% CI 1.97-3.86). Those who did not believe that using smart devices during bedtime caused a delayed sleep schedule were also more likely to not intend to stop compared to those who did believe it (OR 2.41, 95% CI 1.33-4.35).

The OR analysis also suggested that the perception of immediate negative health impacts influenced the intention to reduce or cease the use of smart devices at night. The most common issue was fatigue and drowsiness, with 62.01% (480/774) of participants experiencing these symptoms. Headaches were also prevalent, affecting 45.09% (349/774) of individuals, while 40.82% (316/774) reported mood swings and 40.69% (315/774) experienced a lack of focus and distraction. Participants who did not report any negative health observations had significantly higher odds (OR 5.25, 95% CI 2.36-11.69) of intending to stop using smart devices compared to those who did report negative health observations. Conversely, more than two-thirds of participants who experienced headaches, dry eyes, or lack of focus and distraction were less likely to intend to stop using smart devices at bedtime. Specifically, participants with headaches were 37% less likely (OR 0.63, 95% CI 0.47-0.85), those with dry eyes were 31% less likely (OR 0.69, 95% CI 0.51-0.94), and those with a lack of focus and distraction were 30% less likely (OR 0.70, 95% CI 0.52-0.95) to intend to stop compared to those without these symptoms.

XSL•FO
RenderX

Table 3. Perceived negative effects of smart device use on sleep quality and next-morning well-being among university students in Saudi Arabia.

Variables	Values (N=774)	Likely to stop using smart devices during bedtime in the next 3 months			
		Yes (n=451)	No (n=323)	Odds ratio (95% CI)	*P* value
Using smart devices during bedtime makes you sleep less (SQPNE[a]_1), n (%)					
Yes	582 (75.19)	375 (83.15)	207 (64.09)	Ref	—[b]
No	192 (24.80)	76 (16.85)	116 (35.91)	2.76 (1.97-3.86)	<.001
Using smart devices during bedtime causes a delayed sleep schedule (SQPNE_2), n (%)					
Yes	724 (93.54)	432 (95.79)	292 (90.4)	Ref	—
No	50 (6.45)	19 (4.21)	31 (9.6)	2.41 (1.33-4.35)	<.001
Using smart devices during bedtime negatively affects oneself the next morning (SQPNE_3), n (%)					
Yes	561 (72.48)	374 (82.93)	187 (57.89)	Ref	—
No	213 (27.51)	77 (17.07)	136 (42.11)	3.53 (2.53-4.91)	<.001
After highly using the smart device the previous night, health observations the next morning (SQPNE_4) , n (%)					
(1) Headache					
False	425 (54.90)	227 (50.33)	198 (61.3)	Ref	—
True	349 (45.09)	224 (49.67)	125 (38.7)	0.63 (0.47-0.85)	<.001
(2) Dry eyes					
False	508 (65.63)	281 (62.31)	227 (70.28)	Ref	—
True	266 (34.36)	170 (37.69)	96 (29.72)	0.69 (0.51-0.94)	.02
(3) Pain in the ear					
False	687 (88.75)	400 (88.69)	287 (88.85)	Ref	—
True	87 (11.24)	51 (11.31)	36 (11.15)	0.98 (0.62-1.54)	1.00
(4) Fatigue and drowsiness					
False	294 (37.98)	160 (35.48)	134 (41.49)	Ref	—
True	480 (62.01)	291 (64.52)	189 (58.51)	0.77 (0.57-1.04)	.09
(5) Lack of focus and distraction					
False	459 (59.30)	252 (55.88)	207 (64.09)	Ref	—
True	315 (40.69)	199 (44.12)	116 (35.91)	0.70 (0.52-0.95)	.02
(6) Mood swings					
False	458 (59.17)	257 (56.98)	201 (62.23)	Ref	—
True	316 (40.82)	194 (43.02)	122 (37.77)	0.80 (0.60-1.07)	.15
(7) Increased desire to have food rich in sugars and fats					
False	610 (78.81)	345 (76.50)	265 (82.04)	Ref	—
True	164 (21.18)	106 (23.50)	58 (17.96)	0.71 (0.49-1.01)	.07
(8) Nothing					
False	738 (95.34)	443 (98.23)	295 (91.33)	Ref	—
True	36 (4.65)	8 (1.77)	28 (8.67)	5.25 (2.36-11.69)	<.001
Subjective sleep quality , n (%)					
Sleep quality overall					
Very good	166 (21.44)	94 (20.84)	72 (22.29)	Ref	—
Fairly good	338 (43.66)	191 (42.35)	147 (45.51)	1.00 (0.69-1.46)	1.00
Fairly bad	200 (25.83)	118 (26.16)	82 (25.39)	0.90 (0.59-1.37)	0.67
Very bad	70 (9.04)	48 (10.64)	22 (6.81)	0.59 (0.33-1.08)	0.10

XSL•FO
RenderX

[a]SQPNE: sleep quality perceived negative effects.

[b]Not applicable.

Physical Sleep Disturbances Experienced During the Past Month

In terms of sleep latency, the majority of participants (209/774, 27%) reported having trouble falling asleep within 30 minutes 3 or more times per week, as shown in Table 4. Regarding sleep disturbances, waking up in the middle of the night or early morning occurred three or more times per week for 36.17% (280/774) of participants, making it a common problem. Use of sleep medications was not common, as 84.23% (652/774) of participants did not take sleep aids during the past month. Daytime dysfunction was significant for some participants, with 14.47% (112/774) struggling to stay awake 3 or more times per week during daily activities like driving or eating. Also, 31.39% (243/774) of participants reported that maintaining enthusiasm to complete tasks was somewhat of a problem, while 13.82% (107/774) found it to be a very big issue.

About 7.76% (35/451) of the participants who cough or snore loudly 3 or more times per week had significantly lower odds (OR 0.45, 95% CI 0.23-0.88, *P*=.02) of intending to stop using smart devices compared to those who did not. Similarly, 9.31% (42/451) of those who used sleep medications less than once per week had lower odds (OR 0.49, 95% CI 0.27-0.89, *P*=.01) of intending to stop using smart devices compared to those who did not use sleep medications. Over a quarter of participants who needed to get up to use the bathroom less than once per week (119/451, 26.39%) or once or twice per week (91/451, 20.18%) had lower odds of intending to stop using smart devices at bedtime compared to those who did not need to get up (OR 0.63, 95% CI 0.43-0.92, *P*=.01 for less than once per week; OR 0.63, 95% CI 0.42-0.96, *P*=.03 for once or twice per week). About 21.51% (97/451) of the participants who could not breathe comfortably less than once per week had lower odds (OR 0.67, 95% CI 0.46-0.99, *P*=.05) of intending to stop using smart devices compared to those who did not have breathing issues. Finally, 26.61% (120/451) of the participants who felt too cold once or twice per week had lower odds (OR 0.62, 95% CI 0.42-0.91, *P*=.01) of intending to stop using smart devices compared to those who did not feel too cold.

XSL•FO
RenderX

Table 4. Physical sleep disturbances experienced during the past month by university students in Saudi Arabia.

PSQI[a] components	Values (N=774)	Likely to stop using smart devices during bedtime in the next 3 months			
		Yes (n=451)	No (n=323)	Odds ratio (95% CI)	*P* value
Sleep latency, n (%)					
Cannot get to sleep within 30 minutes					
Not during the past month	201 (25.96)	120 (26.61)	81 (25.08)	Ref	—[b]
Less than once per week	173 (22.35)	100 (22.17)	73 (22.60)	1.08 (0.71-1.63)	.75
Once or twice per week	191 (24.67)	110 (24.39)	81 (25.08)	1.09 (0.72-1.63)	.68
Three or more times per week	209 (27.00)	121 (26.83)	88 (27.24)	1.07 (0.72-1.59)	.76
Sleep disturbance, n (%)					
Wake up in the middle of the night or early morning					
Not during the past month	187 (24.16)	101 (22.39)	86 (26.63)	Ref	—
Less than once per week	125 (16.14)	79 (17.52)	46 (14.24)	0.68 (0.43-1.08)	.12
Once or twice per week	182 (23.51)	102 (22.62)	80 (24.77)	0.92 (0.61-1.38)	.75
Three or more times per week	280 (36.17)	169 (37.47)	111 (34.37)	0.77 (0.53-1.12)	.18
Have to get up use to the bathroom (SQDPM[c]_1)					
Not during the past month	287 (37.08)	150 (33.26)	137 (42.41)	Ref	—
Less than once per week	188 (24.28)	119 (26.39)	69 (21.36)	0.63 (0.43-0.92)	.01
Once or twice per week	144 (18.60)	91 (20.18)	53 (16.41)	0.63 (0.42-0.96)	.03
Three or more times per week	155 (20.02)	91 (20.18)	64 (19.81)	0.77 (0.51-1.14)	.22
Cannot breathe comfortably (SQDPM_2)					
Not during the past month	453 (58.52)	251 (55.65)	202 (62.54)	Ref	—
Less than once per week	150 (19.37)	97 (21.51)	53 (16.41)	0.67 (0.46-0.99)	.05
Once or twice per week	94 (12.14)	56 (12.42)	38 (11.76)	0.84 (0.53-1.32)	.49
Three or more times per week	77 (9.94)	47 (10.42)	30 (9.29)	0.79 (0.48-1.29)	.38
Cough or snore loudly (SQDPM_3)					
Not during the past month	580 (74.93)	330 (73.17)	250 (77.40)	Ref	—
Less than once per week	88 (11.36)	51 (11.31)	37 (11.46)	0.95 (0.60-1.50)	.90
Once or twice per week	59 (7.62)	35 (7.76)	24 (7.43)	0.90 (0.52-1.56)	.78
Three or more times per week	47 (6.07)	35 (7.76)	12 (3.72)	0.45 (0.23-0.88)	.02
Feel too cold (SQDPM_4)					
Not during the past month	262 (33.85)	138 (30.60)	124 (38.39)	Ref	—
Less than once per week	144 (18.60)	88 (19.51)	56 (17.34)	0.70 (0.46-1.07)	.11
Once or twice per week	187 (24.16)	120 (26.61)	67 (20.74)	0.62 (0.42-0.91)	.01
Three or more times per week	181 (23.38)	105 (23.28)	76 (23.53)	0.80 (0.54-1.18)	.28
Feel too hot					
Not during the past month	317 (40.95)	179 (39.69)	138 (42.72)	Ref	—
Less than once per week	175 (22.60)	104 (23.06)	71 (21.98)	0.88 (0.60-1.28)	.56
Once or twice per week	165 (21.31)	97 (21.51)	68 (21.05)	0.90 (0.62-1.33)	.69
Three or more times per week	117 (15.11)	71 (15.74)	46 (14.24)	0.84 (0.54-1.29)	.44
Have bad dreams					
Not during the past month	271 (35.01)	155 (34.37)	116 (35.91)	Ref	—
Less than once per week	241 (31.13)	139 (30.82)	102 (31.58)	0.98 (0.69-1.39)	.92

XSL•FO
RenderX

PSQI[a] components	Values (N=774)	Likely to stop using smart devices during bedtime in the next 3 months			
		Yes (n=451)	No (n=323)	Odds ratio (95% CI)	*P* value
Once or twice per week	170 (21.96)	98 (21.73)	72 (22.29)	0.98 (0.66-1.44)	>.99
Three or more times per week	92 (11.88)	59 (13.08)	33 (10.22)	0.74 (0.45-1.21)	.27
Have pain					
Not during the past month	381 (49.22)	213 (47.23)	168 (52.01)	Ref	—
Less than once per week	156 (20.15)	90 (19.96)	66 (20.43)	0.92 (0.63-1.35)	.77
Once or twice per week	125 (16.14)	78 (17.29)	47 (14.55)	0.76 (0.50-1.15)	.21
Three or more times per week	112 (14.47)	70 (15.52)	42 (13.00)	0.76 (0.49-1.17)	.23
Use of sleep medications, n (%)					
Taken medicine to help you sleep (prescribed or "over the counter") (SQDPM_5)					
Not during the past month	652 (84.23)	367 (81.37)	285 (88.24)	Ref	—
Less than once per week	58 (7.49)	42 (9.31)	16 (4.95)	0.49 (0.27-0.89)	.01
Once or twice per week	37 (4.78)	26 (5.76)	11 (3.41)	0.54 (0.26-1.12)	.12
Three or more times per week	27 (3.48)	16 (3.55)	11 (3.42)	0.88 (0.40-1.93)	.84
Daytime dysfunction, n (%)					
Having trouble staying awake while driving, eating meals, or engaging in social activity					
Not during the past month	369 (47.67)	206 (45.68)	163 (50.46)	Ref	—
Less than once per week	161 (20.80)	103 (22.84)	58 (17.96)	0.71 (0.48-1.04)	.08
Once or twice per week	132 (17.05)	69 (15.30)	63 (19.50)	1.15 (0.77-1.71)	.54
Three or more times per week	112 (14.47)	73 (16.19)	39 (12.07)	0.67 (0.43-1.04)	.08
Keeping up enough enthusiasm to get things done					
No problem at all	202 (26.09)	112 (24.83)	90 (27.86)	Ref	—
Only a very slight problem	222 (28.68)	137 (30.38)	85 (26.32)	0.77 (0.52-1.13)	.20
Somewhat of a problem	243 (31.39)	137 (30.38)	106 (32.82)	0.96 (0.66-1.40)	.84
A very big problem	107 (13.82)	65 (14.41)	42 (13.00)	0.80 (0.49-1.29)	.39

[a]PSQI: Pittsburgh Sleep Quality Index.

[b]Not applicable.

[c]SQDPM: sleep quality during the past month.

Predictors of the Intention to Stop Using Smart Devices at Bedtime

Path analysis was conducted to understand the complex relationships between the measured variable intended to stop using smart device usage and the latent variables, which were SQSU (presented in Table 2), SQPNE (presented in Table 3), and SQDPM (presented in Table 4). All the demographic factors in Table 1 were removed from the analysis as they did not approach the significance threshold of <.05. To improve the model fit, only the significant variables (<.05) from Tables 2-4 were included. SEM plot indicated that the model had an excellent fit with a nonsignificant chi-square, a CFI of 0.845, and a RMSEA value of 0.039 (95% CI 0.032-0.046).

All the factors except SQSU_7 (purposes=–0.09, *P*>.05) had significant factor loadings (path coefficients) with latent variables SQSU, SQPNE, and SQDPM. People using smart devices every night (SQSU_1; 0.36, *P*<.05), 3 hours or more time spent on smart devices at bedtime (SQSU_2; –0.35, *P*<.05), using the smart device frequently after lights were turned off (SQSU_3; 0.49, *P*<.05), using headphones frequently during bedtime (SQSU_4; 0.39, *P*<.05), awakened frequently by the smart device at night (SQSU_5; 0.25, *P*<.05), activities (SQSU_6; –0.18, *P*<.05) had significant path coefficients with latent variable SQSU.

People who perceived using smart devices during bedtime led to sleeplessness (SQPNE_1; 0.58, *P*<.05), caused delayed sleep schedule (SQPNE_2; 0.29, *P*<.05), negatively affected oneself the next morning (SQPNE_3; 0.71, *P*<.05), and affected their health (SQPNE_4; 0.24, *P*<.05) had significant path coefficients with latent variable SQPNE.

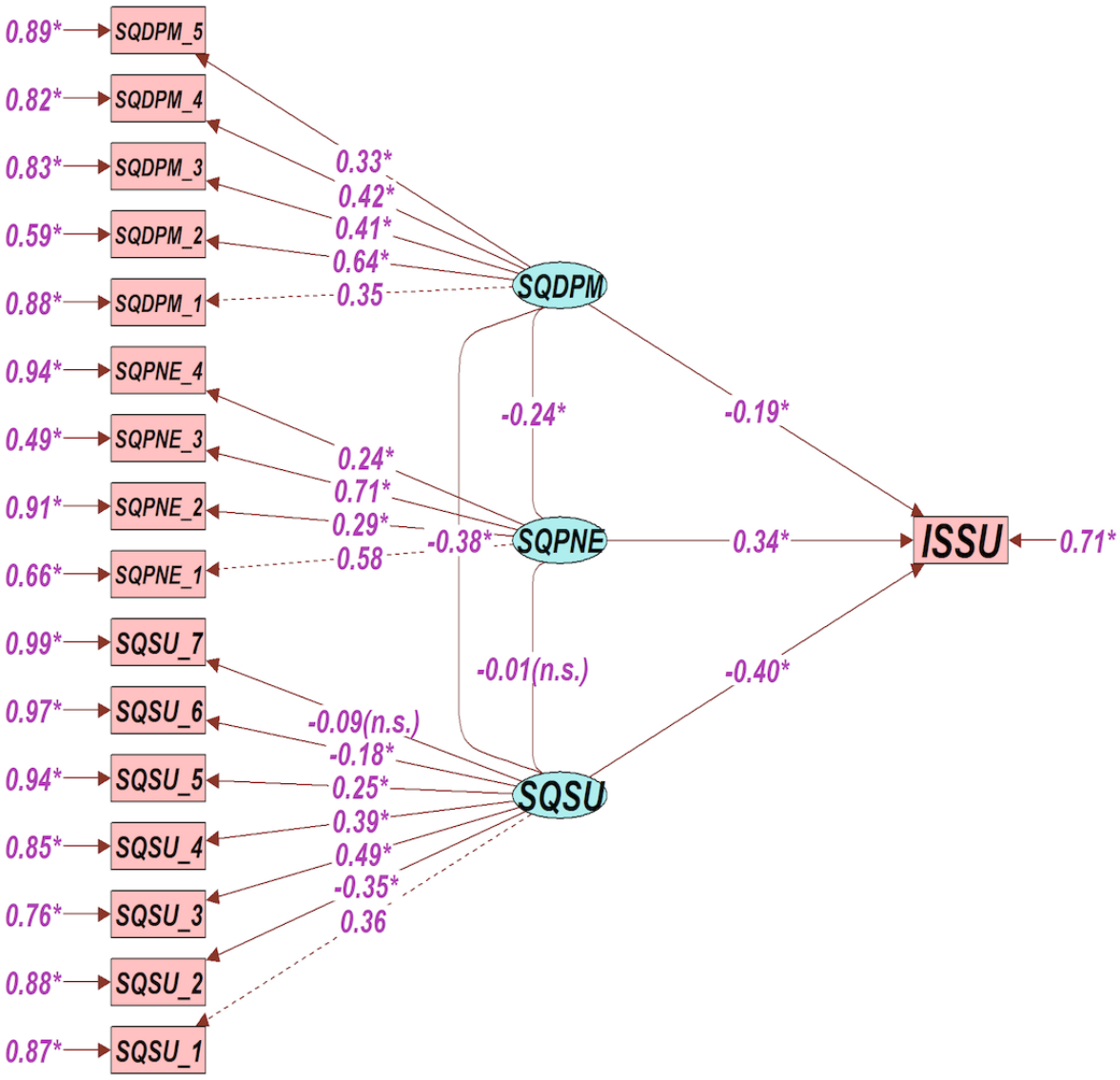
People woke up in the middle of the night or early morning, had to get up to use the bathroom (SQDPM_1; 0.35, *P*<.05), could not breathe comfortably (SQDPM_2; 0.64, *P*<.05), coughed or snored loudly (SQDPM_3; 0.41, *P*<.05), felt cold (SQDPM_4; 0.42, *P*<.05), and took medicine to help sleep

XSL•FO
RenderX

(SQDPM_5; 0.33, *P*<.05) had significant path coefficients with latent variable SQDPM.

The 3 obtained factors have path coefficients with ISSU (intended to stop using smart device usage). SQSU=–0.40, SQPNE=0.34, and SQDPM=–0.19 were significant and there was no evidence of collinearity in the model as shown in Figure 1. All items of the model had significant paths as well as correlations with respective factor domains. Factor structures remained intact, and all paths were significant.

Figure 1. Path analysis of predictors influencing university students' intention to stop using smart devices at bedtime in Saudi Arabia. An asterisk indicates a 2-tailed *P* value <.05. ISSU: intended to stop using smart device usage; SQDPM: sleep quality during the past month; SQPNE: sleep quality perceived negative effects; SQSU: sleep quality smartphone usage.



Discussion

Principal Findings

This study investigated the factors that influence university students' intention to stop using smart devices at bedtime in Saudi Arabia. It focused on behavioral habits, perceived negative effects on sleep quality, and physical sleep disturbances. The findings provide insights into how smartphone usage behaviors and sleep-related perceptions interplay, influencing students' intentions to modify their presleep device habits, as follows.

Behavioral Factors and Smart Device Usage Habits

The results from the analysis of smartphone usage habits (Table 2) revealed that frequent nightly use and the use of smart devices after lights are turned off were significantly associated with the intention to stop using these devices at bedtime. Specifically, using smart devices after lights out (path coefficient=0.49) was the strongest behavioral predictor, indicating that students who

XSL•FO
RenderX

engage in late-night device use may recognize its disruptive effects on sleep and are thus more inclined to reduce or stop this behavior. This aligns with previous studies that highlight the adverse effects of presleep exposure to light from screens, which delays sleep onset [13,17].

However, spending more than 3 hours on devices (path coefficient=–0.35) and engaging in multiple activities (path coefficient=–0.18) were negatively associated with the intention to stop using smart devices at bedtime. This suggests that heavy users may be more resistant to behavioral change, possibly due to their addiction to using smart devices [1]. This aligns with previous studies that have identified higher levels of smartphone addiction as barriers to changing screen time habits [27,28].

Moreover, engaging in more activities on the device also decreased the likelihood of stopping using smart devices (path coefficient=–0.18), which may indicate that participants perceive these activities (eg, watching videos and chatting) as integral parts of their nightly routine [29,30]. These results are consistent with previous studies that link high-frequency smart device use to a reluctance to disengage from these devices, possibly due to increased dependency on smart devices for both academic and social purposes [27,31].

Perceived Negative Effects on Sleep Quality and Well-Being

The most significant predictors of the intention to stop using smart devices were related to perceived negative effects on sleep quality and next-morning well-being (Table 3). The strongest predictor was the belief that smart device usage negatively affects next-morning well-being (path coefficient=0.71). This suggests that students who experience morning fatigue, mood disturbances, or difficulty focusing are more likely to stop using their devices before sleep. This is consistent with previous studies that demonstrate how poor sleep quality directly influences daytime functioning [19], especially in academic contexts where cognitive performance is critical [15,30,31].

Similarly, the perception of sleeplessness due to smart device use was also a strong predictor (path coefficient=0.58). This indicates that university students who struggle to sleep because of their devices are more likely to intend to stop using them. This is consistent with previous studies on the disruptive effects of presleep screen time, which has been linked to increased sleep latency and poorer overall sleep quality [13,32]. In addition, the perception of delayed sleep schedules (path coefficient=0.29) also contributed to the intention to stop, further reinforcing the idea that students recognize the role of smart devices in disrupting their sleep timing.

Physical Sleep Disturbances and Health Implications

The results from physical sleep disturbances (Table 4) highlighted how certain health issues experienced during the night are significantly linked with the intention to stop using smart devices. Difficulty breathing comfortably during sleep (path coefficient=0.64) was the most prominent physical predictor, suggesting that students who experience physical discomforts, such as respiratory issues or muscle tension, may attribute these problems to their use of smart devices at bedtime and are therefore more likely to reduce their usage. This is consistent with studies that have found links between prolonged device use in bed and poor posture or breathing discomfort, potentially due to prolonged screen time in awkward positions [33].

Other sleep disturbances, such as waking up during the night, snoring or coughing, and feeling too cold, were also positively associated with the intention to stop using smart devices at bedtime. These findings indicate that students who experience sleep disruptions are more likely to modify their behavior to improve their sleep quality [8,20]. Furthermore, students who reported taking sleep medications were more likely to intend to stop using smart devices, suggesting that those with more severe sleep problems may be more motivated to address contributing factors like bedtime device use [31].

Implications for University Students and Public Health Interventions

These findings have significant implications for interventions targeting sleep hygiene among university students, particularly in Saudi Arabia. Given that the strongest predictors of the intention to stop using smart devices were perceived negative effects on next-morning well-being and physical sleep disturbances, public health initiatives should focus on raising awareness about these health risks [34]. Particularly, educational campaigns should emphasize how devices used before bed negatively impact not only sleep quality but also cognitive functioning, mood, and overall well-being the next day [35].

Interventions should also incorporate behavior change theories, such as the Transtheoretical Model and the Theory of Planned Behavior [36,37], to understand and address resistance to change. Many heavy device users, despite recognizing negative outcomes, may rely on their devices for emotional regulation or social engagement [5,16], making these activities a barrier to change. Tailored strategies should therefore target these specific behaviors, offering alternative coping mechanisms like mindfulness exercises, relaxation techniques [38], or structured bedtime routines to gradually reduce dependency on devices [39]. Furthermore, strategies that encourage limiting screen time after lights out and reducing device-related disruptions, such as turning off notifications or using blue light filters, could be effective in promoting healthier sleep habits [32,34].

Strengths, Limitations, and Future Directions

This study has several strengths. First, its multidimensional approach examined not only behavioral habits but also perceived negative effects and physical sleep disturbances, offering a holistic view of the factors that influence the intention to stop using smart devices. Second, the use of a validated scale such as PSQI adds rigor to the assessment of sleep quality. Finally, the application of path analysis and SEM allowed for an in-depth analysis of the relationships between latent variables, supporting the interpretation of complex patterns.

However, there are some limitations. The cross-sectional design limits the ability to establish causality between smart device usage and sleep disturbances [40].

The reliance on self-reported data introduces potential biases, including recall bias (where participants may not accurately

XSL•FO
RenderX

remember behaviors or experiences) and social desirability bias (where participants may provide what they perceive to be more acceptable responses). These biases could influence the accuracy of the findings [41]. Furthermore, the open recruitment approach via social media platforms prevented determining an exact participation rate [42]. Although a large sample (n=774) of university students from Saudi Arabia may improve the applicability of the findings to this population, replication in other cultural contexts is necessary to assess broader generalizability and refine intervention strategies.

Future studies should consider longitudinal research to establish causality and interventional studies to test specific strategies for reducing device use at bedtime [16,26]. Also, investigating different types of smart device use (eg, browsing social media, gaming, chatting, and studying) would provide more nuanced insights. In addition, future studies should investigate the interactions between smart device use, sleep hygiene, and mental health such as anxiety, depression, or stress levels because understanding these interactions could inform more tailored interventions for improving sleep among university students [5].

Conclusions

This study provides valuable insights into the factors influencing university students' intention to stop using smart devices at bedtime, particularly in the Saudi Arabian context. While behavioral habits, such as frequent device use and activities, are important, the most significant predictors of behavior change are perceived negative effects on well-being and physical sleep disturbances. These findings highlight the importance of addressing both the psychological and physiological impacts of smart device use in interventions aimed at improving sleep quality among university students. Future studies should explore how these interventions can be effectively implemented and sustained over time to mitigate the negative effects of excessive bedtime device use.

Data Availability

The datasets generated during and/or analyzed during this study are available from the author on reasonable request.

Conflicts of Interest

None declared.

References

1. Leow MQH, Chiang J, Chua TJX, Wang S, Tan NC. The relationship between smartphone addiction and sleep among medical students: a systematic review and meta-analysis. PLoS One 2023;18(9):e0290724 [FREE Full text] [doi: 10.1371/journal.pone.0290724] [Medline: 37713408]
2. Smartphone penetration as share of population in Saudi Arabia 2020-2029. Statista. URL: https://www.statista.com/statistics/625436/smartphone-user-penetration-in-saudi-arabia [accessed 2024-02-21]
3. Smartphone mobile network subscriptions worldwide 2016-2028. Statista. URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2025-01-02]
4. Zhao C, He J, Xu H, Zhang J, Zhang G, Yu G. Are 'night owls' or 'morning larks' more likely to delay sleep due to problematic smartphone use? A cross-lagged study among undergraduates. Addict Behav 2024;150:107906. [doi: 10.1016/j.addbeh.2023.107906] [Medline: 37984222]
5. Yang J, Fu X, Liao X, Li Y. Association of problematic smartphone use with poor sleep quality, depression, and anxiety: a systematic review and meta-analysis. Psychiatry Res 2020;284:112686. [doi: 10.1016/j.psychres.2019.112686] [Medline: 31757638]
6. Nakie G, Takelle GM, Rtbey G, Andualem F, Tinsae T, Kassa MA, et al. Sleep quality and associated factors among university students in Africa: a systematic review and meta-analysis study. Front Psychiatry 2024;15:1370757 [FREE Full text] [doi: 10.3389/fpsyt.2024.1370757] [Medline: 38559402]
7. Pagano M, Bacaro V, Crocetti E. 'Using digital media or sleeping … that is the question'. A meta-analysis on digital media use and unhealthy sleep in adolescence. Comput Hum Behav 2023;146:107813. [doi: 10.1016/j.chb.2023.107813]
8. Putra Pratama Wau R, Wienaldi W, Dharma S. Relationship between smartphone use and sleep quality of universitas prima Indonesia students. Int J Soc Health 2024;3(1):31-37. [doi: 10.58860/ijsh.v3i1.146]
9. Han X, Zhou E, Liu D. Electronic media use and sleep quality: updated systematic review and meta-analysis. J Med Internet Res 2024;26:e48356 [FREE Full text] [doi: 10.2196/48356] [Medline: 38533835]
10. Dagani J, Buizza C, Cela H, Sbravati G, Rainieri G, Ghilardi A. The interplay of sleep quality, mental health, and sociodemographic and clinical factors among Italian college freshmen. J Clin Med 2024;13(9):2626 [FREE Full text] [doi: 10.3390/jcm13092626] [Medline: 38731154]
11. M. Alfawareh H, Jusoh S. Smartphones usage among university students: Najran University case. Int J Acad Res 2014;6(2):321-326.
12. Alzhrani AM, Aboalshamat KT, Badawoud AM, Abdouh IM, Badri HM, Quronfulah BS, et al. The association between smartphone use and sleep quality, psychological distress, and loneliness among health care students and workers in Saudi Arabia. PLoS One 2023;18(1):e0280681 [FREE Full text] [doi: 10.1371/journal.pone.0280681] [Medline: 36701337]

XSL•FO
RenderX

13. Randjelović P, Stojanović N, Ilić I, Vučković D. The effect of reducing blue light from smartphone screen on subjective quality of sleep among students. Chronobiol Int 2023;40(3):335-342. [doi: 10.1080/07420528.2023.2173606] [Medline: 36744480]
14. Elsheikh AA, Elsharkawy SA, Ahmed DS. Impact of smartphone use at bedtime on sleep quality and academic activities among medical students at Al-Azhar University at Cairo. J Public Health (Berl) 2023;32(11):2091-2100 [FREE Full text] [doi: 10.1007/s10389-023-01964-8]
15. Liu H, Váchová L, Plevová I. Effect of mobile phone dependence on various aspects of academic achievement: evidence from Chinese and Czech university students. Pegem Journal of Education and Instruction 2024;14(2):20-26. [doi: 10.47750/pegegog.14.02.03]
16. Cemei L, Sriram S, Holý O, Rehman S. A longitudinal investigation on the reciprocal relationship of problematic smartphone use with bedtime procrastination, sleep quality, and mental health among university students. Psychol Res Behav Manag 2024;17:3355-3367 [FREE Full text] [doi: 10.2147/PRBM.S472299] [Medline: 39359420]
17. Laghari I, Khan A, Hussain S, Waheed S. Relationship between screen-time, sleeping pattern and stress during the pandemic among university students. Int J Soc Sci Arch 2023;6(3).
18. Ji Z, Zhou L. A problem not to be ignored: the influencing factors of mobile phone addiction and its influence on sleep quality. Health 2024;16(5):459-469.
19. Wang Q, Zhong Y, Zhao G, Song R, Zeng C. Relationship among content type of smartphone use, technostress, and sleep difficulty: a study of university students in China. Educ Inf Technol 2022;28(2):1697-1714. [doi: 10.1007/s10639-022-11222-1]
20. Hu B, Shen W, Wang Y, Wu Q, Li J, Xu X, et al. Prevalence and related factors of sleep quality among Chinese undergraduates in Jiangsu province: multiple models' analysis. Front Psychol 2024;15:1343186 [FREE Full text] [doi: 10.3389/fpsyg.2024.1343186] [Medline: 38659684]
21. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991;50(2):179-211.
22. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Int J Surg 2014;12(12):1500-1524 [FREE Full text] [doi: 10.1016/j.ijsu.2014.07.014] [Medline: 25046751]
23. Alshalawi AS. Social media usage intensity and academic performance among undergraduate students in Saudi Arabia. Contemp Educ Technol 2022;14(2):ep361. [doi: 10.30935/cedtech/11711]
24. Number of social media users in Saudi Arabia from 2020 to 2029. Statista. URL: https://www.statista.com/forecasts/1144943/social-media-users-in-saudi-arabia [accessed 2024-05-13]
25. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. Psychiatry Res 1989;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
26. Suleiman KH, Yates BC, Berger AM, Pozehl B, Meza J. Translating the Pittsburgh sleep quality index into Arabic. West J Nurs Res 2010;32(2):250-268. [doi: 10.1177/0193945909348230] [Medline: 19915205]
27. Lane HY, Chang CJ, Huang CL, Chang YH. An investigation into smartphone addiction with personality and sleep quality among university students. Int J Environ Res Public Health 2021;18(14):7588 [FREE Full text] [doi: 10.3390/ijerph18147588] [Medline: 34300037]
28. Cui G, Yin Y, Li S, Chen L, Liu X, Tang K, et al. Longitudinal relationships among problematic mobile phone use, bedtime procrastination, sleep quality and depressive symptoms in Chinese college students: a cross-lagged panel analysis. BMC Psychiatry 2021;21(1):449 [FREE Full text] [doi: 10.1186/s12888-021-03451-4] [Medline: 34507561]
29. Pillion M, Gradisar M, Bartel K, Whittall H, Kahn M. What's 'app'-ning to adolescent sleep? Links between device, app use, and sleep outcomes. Sleep Med 2022;100:174-182. [doi: 10.1016/j.sleep.2022.08.004] [Medline: 36084495]
30. Kumar Swain R, Pati AK. Use of social networking sites (SNSs) and its repercussions on sleep quality, psychosocial behavior, academic performance and circadian rhythm of humans – a brief review. Biological Rhythm Research 2019;52(8):1139-1178. [doi: 10.1080/09291016.2019.1620487]
31. Rathakrishnan B, Bikar Singh SS, Kamaluddin MR, Yahaya A, Mohd Nasir MA, Ibrahim F, et al. Smartphone addiction and sleep quality on academic performance of university students: an exploratory research. Int J Environ Res Public Health 2021;18(16):8291 [FREE Full text] [doi: 10.3390/ijerph18168291] [Medline: 34444042]
32. Huiberts LM, Opperhuizen AL, Schlangen LJM. Pre-bedtime activities and light-emitting screen use in university students and their relationships with self-reported sleep duration and quality. Light Res Technol 2022;54(6):595-608. [doi: 10.1177/14771535221074725]
33. Sirajudeen M, Alzhrani M, Alanazi A, Alqahtani M, Waly M, Unnikrishnan R, et al. Prevalence of text neck posture, smartphone addiction, and its association with neck disorders among university students in the Kingdom of Saudi Arabia during the COVID-19 pandemic. PeerJ 2022;10:e14443 [FREE Full text] [doi: 10.7717/peerj.14443] [Medline: 36540801]
34. Pham HT, Chuang H, Kuo C, Yeh T, Liao W. Electronic device use before bedtime and sleep quality among university students. Healthcare (Basel) 2021;9(9):1091 [FREE Full text] [doi: 10.3390/healthcare9091091] [Medline: 34574865]

XSL•FO
RenderX

35. Andersen TO, Dissing AS, Varga TV, Rod NH. The SmartSleep experiment: Evaluation of changes in night-time smartphone behavior following a mass media citizen science campaign. PLoS One 2021;16(7):e0253783 [FREE Full text] [doi: 10.1371/journal.pone.0253783] [Medline: 34288929]
36. Pegado A, Alvarez M, Roberto MS. The role of behaviour-change theory in sleep interventions with emerging adults (aged 18-29 years): a systematic review and meta-analysis. J Sleep Res 2023;32(5):e13877. [doi: 10.1111/jsr.13877] [Medline: 36922157]
37. Knowlden A. Models and theories of behavior change relevant to sleep health. In: Sleep and Health. United States: Academic Press; 2019:171-186.
38. Kravcenko B. The effect of a six-week mindfulness meditation stress reduction course on the effect of mobile phone use on sleep quality in master students at the University of Twente. Netherlands: Bachelor's thesis, University of Twente; 2023. URL: https://essay.utwente.nl/94266/ [accessed 2025-02-12]
39. Kheirinejad S, Visuri A, Ferreira D, Hosio S. 'Leave your smartphone out of bed': quantitative analysis of smartphone use effect on sleep quality. Pers Ubiquitous Comput 2023;27(2):447-466 [FREE Full text] [doi: 10.1007/s00779-022-01694-w] [Medline: 36405389]
40. Savitz D, Wellenius G. Can cross-sectional studies contribute to causal inference? It depends. Am J Epidemiol 2023;192(4):514-516. [doi: 10.1093/aje/kwac037] [Medline: 35231933]
41. van der Ham IJM, van der Kuil MNA, Claessen MHG. Quality of self-reported cognition: effects of age and gender on spatial navigation self-reports. Aging Ment Health 2021;25(5):873-878 [FREE Full text] [doi: 10.1080/13607863.2020.1742658] [Medline: 32233799]
42. Jones A, Walters J, Brown A. Participant recruitment in social work: a social media approach. Soc. Work Res 2020;44(4):247-255.

Abbreviations

CFI: comparative fit index
OR: odds ratio
PSQI: Pittsburgh Sleep Quality Index
RMSEA: root mean square error of approximation
SEM: sequencing equation modeling
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
SQDPM: sleep quality during the past month
SQPNE: sleep quality perceived negative effects
SQSU: sleep quality smartphone usage

Edited by A Mavragani; submitted 05.10.24; peer-reviewed by I Wilson, WJ He; comments to author 20.12.24; revised version received 24.12.24; accepted 07.01.25; published 10.03.25.

Please cite as:
Almalki M
Predictors of the Intention to Stop Using Smart Devices at Bedtime Among University Students in Saudi Arabia: Cross-Sectional Survey
JMIR Form Res 2025;9:e67223
URL: https://formative.jmir.org/2025/1/e67223
doi:10.2196/67223
PMID:40063070

©Manal Almalki. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Impact of a Sensorimotor Integration and Hyperstimulation Program on Global Motor Skills in Moroccan Children With Autism Spectrum Disorder: Exploratory Clinical Quasi-Experimental Study

Rachid Touali[1], MSci; Jamal Zerouaoui[1], PhD; El Mahjoub Chakir[1], PhD; Hung Tien Bui[2], Dr med, PhD; Mario Leone[2], PhD; Maxime Allisse[3], PhD

[1]Faculty of Sciences, University of Ibn Tofail, Kénitra, Morocco

[2]Faculty of Medicine and Health Sciences, University of Sherbrooke, Sherbrooke, QC, Canada

[3]Faculty of Physical Activity Sciences, University of Sherbrooke, Sherbrooke, QC, Canada

Corresponding Author:
Maxime Allisse, PhD
Faculty of Physical Activity Sciences
University of Sherbrooke
2500 Boulevard de l'Université
Sherbrooke, QC, J1K 2R1
Canada
Phone: 1 819 821 8000 ext 63875
Email: maxime.allisse@usherbrooke.ca

Abstract

Background: Children with autism spectrum disorders (ASDs) often struggle with processing information, which can impact their coordination, balance, and other motor skills. Studies have demonstrated that intervention programs based on sensory integration can enhance motor performance in these children.

Objective: The objective of this study is to evaluate the applicability of a standardized battery of gross motor skill tests for Moroccan children aged 6 to 12 years with ASD. The objective is to assess the potential efficacy of an innovative pedagogical approach focused on sensorimotor integration and hyperstimulation. This approach will be compared to traditional physical education (PE) sessions to determine its feasibility and potential to bridge the developmental gaps in motor skills between children with ASD and those with a neurotypical profile.

Methods: A convenience sample of 14 Moroccan children with ASD aged 6 to 12 years participated in this exploratory study. Children with ASD were divided into an experimental group (n=7) and a control group (n=7) based on age, sex, motor performance, and socioeconomic status. The control group followed the standard PE program, while the experimental group underwent a specialized program combining sensorimotor integration and hyperstimulation for a period of 15 weeks. All participants were classified as level 2 (moderate) on the Autism Severity Rating Scale based on the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision* (DSM-5-TR) criteria. Gross motor skills were measured at baseline and after 15 weeks of intervention using the UQAC-UQAM (University of Québec in Chicoutimi-University of Québec in Montréal) test battery protocol, which includes 10 items.

Results: At baseline (T1), no significant difference was observed between the control and experimental groups of children with ASD. Following the 15-week intervention, the group participating in traditional PE showed an overall improvement in motor skills of approximately 14.5%. Conversely, the results of the ASD experimental group suggest a more substantial improvement of 44.5%. Additionally, the experimental group exhibited significant better performance across all motor skill variables compared to the control group (minimum *P* values of <.02) with large effect sizes (>0.80). In this regard, a 2-way repeated measures ANOVA confirms the efficiency of the program implemented within the experimental group, demonstrating significant effects associated with both group and time factors as well as a clinically highly significant group×time interaction across all measured variables (η^2p>0.14).

Conclusions: The results of this study suggest that the approach that emphasizes sensorimotor integration and management of hyperstimulation was more effective in improving motor skills in this population. However, other more exhaustive studies will need to be carried out in order to be able to more precisely measure the full potential of this approach.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e65767) doi:10.2196/65767

KEYWORDS

classical physical education; children with a neurotypical profile; children with ASD; UQAC-UQAM test battery; University of Québec in Chicoutimi-University of Québec in Montréal; sensorimotor integration; hyperstimulation; Morocco; sensorimotor; integration; motor skill; Moroccan children; Moroccan; children; autism spectrum disorder; ASD; exploratory study; autism; mental health; young; youth; feasibility

Introduction

Background

Autism spectrum disorders (ASDs) constitute a subset of developmental disorders distinguished by disturbances in communication, changes in social interactions, and repetitive and stereotyped behaviors [1,2]. These symptoms typically emerge before the age of 3 years and result in functional impairments impacting various aspects of daily life [3,4]. Due to the absence of universally recognized biological markers for identifying ASD cases, diagnosis primarily involves evaluation by professionals who assess children's development trajectories to identify potential developmental disorders [2]. Among the myriad of challenges faced by children with ASD, developmental coordination disorders (DCDs) occupy an important place [1]. Despite being relatively common, these motor difficulties have historically been overlooked [5-7]. However, recent research [8-10] sheds light on the significant prevalence of DCD in children with ASD. In fact, DCD, also known as dyspraxia, is a neurodevelopmental disorder that affects motor coordination and skills. Studies have shown that DCD is remarkably prevalent among individuals with ASD, affecting between 34% and 79% of this population. This is significantly higher than the 6% prevalence rate observed in the general population [11]. This high prevalence of DCD in individuals with ASD suggests that a lack of coordination is a common characteristic of the disorder. Motor deficits can present in diverse forms, encompassing compromised gross motor skills such as running; challenges in fine motor skills like handwriting; difficulties in motor planning and execution leading to problems in coordinating movements and sequencing actions seamlessly; and sensory processing issues such as hypersensitivity or hyposensitivity to sensory inputs, which impact balance, coordination, and motor performance [12-14]. As a result, there is widespread recognition that a deficiency in coordination stands as a prominent trait of ASD, frequently concomitant with motor deficits. The potential etiology of these challenges encompasses both genetic and environmental factors [15].

Motor Skill Impairment

Typically, children with ASD often face challenges in motor skills, especially in coordination, balance, and postural control [16-19]. A recent meta-analysis even suggests that these deficits extend to decreased walking speed compared to their peers with a neurotypical profile [20]. However, the underlying causes of these impairments remain a topic of debate within the scientific community [21-24]. To address these motor challenges and improve skill development, various interventions have been proposed. These interventions draw inspiration from diverse areas, including sports disciplines [25-27], physical conditioning programs [28], sensorimotor approaches [29], and exercises focused on fundamental motor skills like running, jumping, throwing, and catching [30]. Additionally, research has explored interventions targeting perceptual-motor skills [31]. However, the authors themselves acknowledged the limitations of their programs and called for the development of more precise and specialized interventions. While most studies suggest improvements in motor performance, none, to our knowledge, have directly compared the effects of a general intervention program to one specifically tailored to the needs of Moroccan children with ASD. Such a comparison, focusing on their specific motor difficulties, would be a valuable contribution to the field.

Sensorimotor Integration and Hyperstimulation Approach for Children With ASD

The field of sensory integration, inspired by the work of Ayres and Robbins [32], offers a valuable approach to improve motor skills in children with neurodevelopmental disorders. This approach hinges on the brain's ability to organize and interpret information received through the senses (sensory integration). It uses 2 main practices: passive unisensory and active multisensory. Passive unisensory practices involve targeted stimulation of a single sense at a time. These interventions, often used to promote relaxation and emotional regulation, can include massage for muscle relaxation and improved circulation and body pressures to calm the nervous system and reduce anxiety, and weighted vests or blankets to provide deep, comforting pressure. The simplicity and calming effect of these techniques make them particularly beneficial for individuals needing to regulate their emotions. The active multisensory approach stands in stark contrast. Here, individuals actively participate in activities that engage multiple senses at once. This approach aims to cultivate a diverse range of skills through tactile activities (exploring textures), proprioceptive activities (understanding body position and movement), and vestibular activities (developing balance and coordination). By actively engaging individuals in these stimulating activities, the active approach fosters a more comprehensive integration of sensory information. This translates to improved motor skills, coordination, and sensory regulation abilities. The emphasis on exploration and interaction with the environment makes active multisensory practices particularly beneficial for children or individuals with sensory processing difficulties.

At the turn of the 1990s, Polatajko et al [33] explored an approach partially similar to the sensorimotor and hyperstimulation method, working with a group of children diagnosed with DCD. However, when comparing their findings with those of other studies using similar protocols, they reported significantly more mixed results. According to the authors, these limited outcomes may be attributed to the considerable heterogeneity among children with DCD. Indeed, various

XSL•FO
RenderX

medical conditions can lead to this disorder, despite having distinct etiological origins. In light of this, the research team advocates for more tailored clinical interventions that take into account the specific characteristics and underlying causes of DCD.

Children with ASD often struggle with processing this information, which can impact their coordination, balance, and other motor skills. Studies have demonstrated that intervention programs based on sensory integration can enhance motor performance in children. Notably, improvements have been observed in postural control, gait, coordination, and hand movements [34,35]. For instance, a systematic review of the effectiveness of interventions using a sensory integrative approach confirmed that these programs can significantly positively affect children's motor skills [36]. The sensorimotor integration and hyperstimulation approach has been used in children with ASD, particularly those aged 2 to 12 years [37]. These interventions are structured to include a variety of sensory and motor exercises, such as balance activities, coordination games involving crossover movements, and sensory activities like manipulating different textures. These activities are designed to stimulate various senses in a controlled manner, helping children better manage their movements and posture.

Research Gap

Despite the recent surge in research, some countries, like Morocco, lack precise information on the motor skill profile of their population with ASD. Although numerous studies suggest improvements in motor performance through sensory integration [34-36], no studies, to our knowledge, directly compare the effects of a general intervention program with those of a program specifically based on a sensorimotor integration and hyperstimulation approach. This lack of direct comparison underscores the need for further research to determine the relative effectiveness of these different approaches [38]. This gap is crucial as some authors highlight the considerable variation in children's motor skills across different countries [39,40]. For instance, studies have shown that Chinese children tend to perform better in manual dexterity and balance tasks, while American children excel in throwing and catching tasks [40]. Moreover, regional disparities offer valuable insights into the development of various motor skills within diverse national contexts.

Objectives

The primary objective of this study is to evaluate the applicability of a standardized battery of gross motor skill tests for Moroccan children aged 6 to 12 years with ASD. This will provide preliminary data and allow for comparison to a group of children with a neurotypical profile matched for age and sex, thereby estimating the discrepancies in motor development between the 2 groups. The secondary objective is to assess the potential efficacy of an innovative pedagogical approach focused on sensorimotor integration and hyperstimulation for Moroccan children with ASD. This approach will be compared to traditional physical education (PE) sessions to determine its feasibility and potential to bridge the developmental gaps in motor skills between children with ASD and peers with a neurotypical profile.

Methods

Design

This study was a clinical, quasi-experimental research with a 2-way repeated measures design conducted on a sample of Moroccan children with ASD aged 6 to 12 years.

Participants

A convenience sample of 14 Moroccan children with ASD aged 6 to 12 (mean 8.8, SD 1.3) years participated in this exploratory study. This age range was selected because motor skills develop particularly well between 6 and 12 years of age, and the chosen test battery has been validated for children of this age group. Children were randomly divided into an experimental group (n=7; mean age 8.4, SD 1.2 years) and a control group (n=7; mean age 8.5, SD 1.2 years) based on age, sex, motor skill performance, and socioeconomic status. All participants were classified by a psychiatrist as level 2 (moderate) on the Autism Severity Rating Scale based on the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision* (DSM-5-TR) criteria. This scale considers the level of support required for the child's interaction and language to function in their environment. Among the recruited children with ASD, the records showed no other conditions, such as attention deficit disorder with or without hyperactivity, intellectual disability, or anxiety. Children with ASD were cared by the specialized association "El Youssr" since they entered school. This institution's mission is to support children with special needs. It is located in the municipality of Salé, on the outskirts of the capital Rabat, where most of the participating families have a low socioeconomic status. None of the children participated in extracurricular sports activities or were undergoing medical treatment that could influence their motor skills, such as taking psychotropic drugs. In addition, none of the children had a disabling condition that could be exacerbated by physical activity. The proposed program was designed to ensure minimal disruption to the existing educational and individual student plans within the institution.

Anthropometric Measures

The anthropometric data were collected in the association's classrooms, which included body mass (BM), body height (BH), BMI, and body surface area (BSA). BM was assessed using a Seca scale model 760 with an accuracy of 0.5 kg. BH was measured with a Seca stadiometer model 214 to the nearest 0.1 cm. All measurements adhered to the guidelines outlined by Lohman et al [41] for standard operating procedures. BMI was calculated using the formula BM (kg)/BH^2 (m). BSA was estimated using the method suggested by Mosteller [42], which read as follows:

Motor Skill Assessment

This exploratory study investigated the factors associated with motor skill development in Moroccan children with ASD over a 15-week period. The initial assessment session (T1) established the baseline level of motor skill development in these children.

XSL•FO
RenderX

The results from T1 were also used to create control and experimental groups, ensuring no significant difference between them at baseline. Gross motor skills were measured using the UQAC-UQAM (University of Québec in Chicoutimi-University of Québec in Montréal) test battery protocol [43-45]. This battery comprises 13 items assessing five gross motor factors: (1) limb movement speed (upper and lower limbs), (2) agility (circle run, 5-m shuttle run, and side-step run), (3) balance (eyes opened and closed), (4) coordination (ball dribbling and target ball toss), and (5) simple reaction time (SRT; ruler drop test).

For the SRT test, the distance was converted into time (ms) using the equation suggested by Aranha et al [46]:

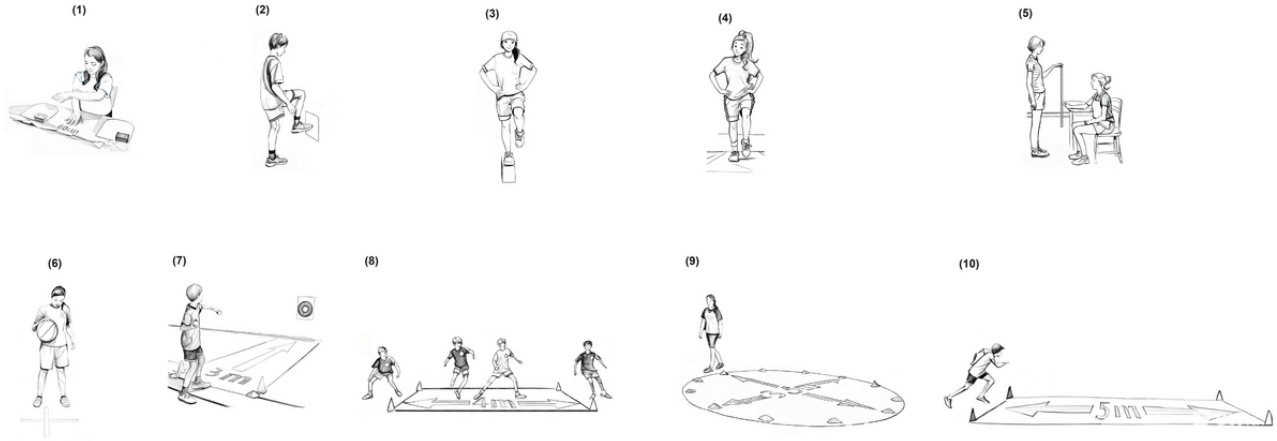
where d is the distance (cm) and g is the gravitational constant (9.8 m/s^2).

The choice of this test battery was primarily driven by methodological considerations. First, it assesses a diverse range of motor factors. Additionally, these tests are practical, as they require minimal space, materials, and time and are user-friendly for test administrators, easy for participants to comprehend, and cost-effective, aligning well with the resources of the supervising institution. Given the sensitivity of children with ASD to environmental changes and their challenges with concentration and attention, it was essential to conduct assessments in their classroom settings whenever feasible. The validity and reliability of these tests have been well-established by numerous studies [43,45,47].

Modified Protocols for the UQAC-UQAM Test Battery

Prior to starting this research project, a week of prospecting was conducted to evaluate the study's feasibility. Of the original 13 tests within the UQAC-UQAM test battery [43], 3 were omitted due to challenges encountered by the children in comprehending or executing the tasks required (slalom run and hand or foot coordination). Furthermore, 2 tests were adapted to enhance accessibility for children with ASD (Figure 1). The first adaptation concerned on the SRT test. The original version required computer administration, but pilot testing revealed this to be overly complex for the participants. This test was therefore replaced by the procedure suggested by Ángel Latorre-Ron et al [48]. This alternative demonstrated a moderate to good correlation with a computer application (Spearman=0.54; P=.03). This change considered the participants' concentration abilities, as the computerized test was longer than the chosen alternative, which only required 2 attempts. The second adaptation concerned the ball toss test, where the distance between the children and the target was reduced from 5 to 3 m. These modifications to the testing protocol were implemented to ensure the study's feasibility and to gather reliable and valid data on motor skill development in children with ASD. It was imperative to provide tests adapted to their abilities and limitations to minimize frustration and encourage active and engaged participation.

Figure 1. Illustration of the ten tests used to assess motor skill competence in children with ASD: (1) upper limbs speed, (2) lower limbs speed, (3) balance eyes opened, (4) balance eyes closed, (5) simple reaction time, (6) ball dribbling, (7) target ball toss, (8) side-step run, (9) circle run, and (10) 5-m shuttle run (adapted from Allisse et al [49], which is published under Creative Commons Attribution 4.0 International License [50]).



Following these adjustments, the final test battery comprised the following ten tests: (1) upper limb speed, (2) lower limb speed, (3) balance eyes opened, (4) balance eyes closed, (5) ruler SRT, (6) ball dribbling, (7) target ball toss, (8) side-step run, (9) circle run, and (10) 5-m shuttle run. For the 2 limb speed tests and the ball dribbling test, the score was determined by the greatest number of repetitions within 20 seconds. For the 3 balance tests and the 3 running tests, the score was recorded as time in seconds. In the SRT test, the score was measured in centimeters and then transformed in milliseconds. The target ball toss test score was cumulative, based on the number of points scored in 10 throws (1 point for hitting the target and 2 points for reaching the center). For all tests except the target ball toss, 2 attempts were allowed, and the best of the 2 was recorded as the final score.

Specific Training Program

The ASD experimental group participated in a specific training program lasting a total of 15 weeks. The program was divided into 3 cycles of 5 weeks each, with each cycle comprising 3 weekly sessions of 45 minutes. This resulted in a total of 45 sessions delivered between November 1, 2021, and April 3, 2022. The pedagogical approach used small group sizes, allowing for individualized monitoring and ensuring a clear understanding of the tasks by participants. The specific training program focused on three key areas: (1) coordination:

XSL•FO
RenderX

development of motor skill enabling participants to coordinate whole-body movements smoothly and efficiently; (2) proprioception: enhancement of body awareness and spatial positioning, leading to improved postural control; and (3) agility: strengthening the ability to react quickly and change direction with precision. The intervention process emphasized offering a variety of engaging and playful activities. The overall objective was to provide participants with rich and multimodal sensory experiences that stimulated awareness and intrinsic motivation and presented optimal motor challenges. This included activities such as exercises performed barefoot or with shoes, allowing participants to explore different textures and sensations and exercises using a trampoline to promote body awareness in space and improve balance. Alongside the ASD experimental group, a separate ASD control group participated in regular PE classes based on the standard curriculum as proposed by the Ministry of National Education of Morocco (Table 1). Further details can be found in Multimedia Appendices 1 and 2. The duration and frequency of these classes matched those of the experimental group. The inclusion of an ASD control group allowed for a comparison of motor skill development in participants who underwent the specific training program against those who did not (for a detailed description of both, the intervention programs for the experimental and control groups, please refer to Multimedia Appendices 1 and 2). Furthermore, a neurotypical control group comprising children without ASD was established. This group served as a baseline to evaluate the motor skill level of children with ASD in both the experimental and control groups, taking into account age and locality.

XSL•FO
RenderX

Table 1. Detailed comparison of 15-week intervention activities for experimental (sensory intervention) and control (classic intervention) groups in a physical development program for children with autism spectrum disorder.

Block and group	Theme	Content	Operationalization
Block 1: 3 times per week (5 weeks)			
Experimental	Coordination	• Spatial orientation; • Execution rhythm and time-space structuring; • Segmental differentiation and dissociation; • Reaction speed.	• Tactile activities: diversifying contact surfaces (barefoot and with shoes); • Proprioceptive activities: trampoline and plyometric exercises; • Vestibular activities: Swiss ball exercises and orientation; • Multimodal sensory experiences tasks; • Playful form, optimal motor challenge, and secure and flexible framework; • Encourage collaboration and participation.
Control	Aerobic fitness	• Aerobic capacity (10 to 15 minutes total); • Aerobic power (3 to 6 minutes total, various repetitions); • Speed (repetitions of 2 to 3 seconds).	• Continuous running: interval running (varying long or fast pace); • Intermittent: short/short, long/long.
Block 2: 3 times per week (5 weeks)			
Experimental	Balance	• Footwork and body balance mastery; • Spatial orientation; • Static and dynamic postural balance; • Plyometric exercises.	• Tactile activities: diversifying contact surfaces (barefoot and with shoes); • Proprioceptive activities: trampoline, plyometric exercises; • Vestibular activities: Swiss ball exercises, orientation; • Multimodal sensory experiences tasks; • Playful form, optimal motor challenge, and secure and flexible framework; • Encourage collaboration and participation.
Control	Strength	• Horizontal and vertical jump (explosive strength); • Linking running with vertical and horizontal jump; • Jump adaptation to space; • Movement, acceleration, tension, ejection, and body stabilization; • Mastery of principles: slow/fast, back/front, bottom/top, right/left.	• Learn phases and sequences of throwing and jumping; • Understand general rules of throwing or jumping; • Motor organization to jump or throw as high or far as possible.
Block 3: 3 times per week (5 weeks)			
Experimental	Agility	• Linking, braking, and acceleration; • Linear and multidirectional speed; • Motor flexibility; • Body control: adaptation to various stimuli; • Rapid and continuous direction changes; • Muscle strengthening.	• Tactile activities: diversifying contact surfaces (barefoot and with shoes); • Proprioceptive activities: trampoline and plyometric exercises; • Vestibular activities: Swiss ball exercises and orientation; • Multimodal sensory experiences tasks; • Playful form, optimal motor challenge, and secure and flexible framework; • Encourage collaboration and participation.
Control	Cooperation and opposition with ball	• Discover a team game: 1 or 2 key rules; • Differentiate partner and opponent; • Respect spatial boundaries; • Discover a target to defend and attack.	• Social interactions and exchanges; • Communication and information gathering; • Cooperation and mutual aid.

Statistical Analysis

The data were summarized using descriptive statistics. Central tendency was characterized by means, and variability was assessed by SD. To assess normality for all variables, the Shapiro-Wilk test was conducted. Group mean comparisons were performed using unpaired 2-tailed *t* tests for the control and experimental groups, while paired 2-tailed *t* tests were used to evaluate differences between measurements at time 1 (T1) and time 2 (T2). Effect sizes (ESs) were calculated using the

XSL•FO
RenderX

Cohen *d* formula [51] and interpreted using the following scale: <0.20 (trivial), 0.20-0.49 (small), 0.50-0.79 (medium), and ≥0.80 (large). The main effects of group and time, as well as the group×time interaction, were analyzed using a 2-way repeated measures ANOVA. Partial eta squared (η^2p) values were reported to indicate ESs, interpreted as follows: <0.01 (trivial), 0.01-0.05 (small), 0.06-0.13 (moderate), and ≥0.14 (large). A significance level of $P \leq .05$ was used. All statistical analyses were carried out with SPSS (version 24.0; IBM Corp).

Ethical Considerations

Details of the project were communicated to parents, students with ASD, and education authorities through a written and explanatory document that each party was required to sign. Thus, by enrolling in this project, the individuals involved acknowledged having provided de facto their written informed consent. However, it was agreed that participants could withdraw at any time without penalty or prejudice. The study was conducted in accordance with the Declaration of Helsinki and approved by the institutional ethics committee of Ibn Tofail University (protocol code: 20019741; date of approval: March 3, 2023) for studies involving humans. Data collection only began upon receiving authorization from the presidents of the 2 participating associations. All collected data were anonymized to ensure participant confidentiality and were stored securely in compliance with ethical guidelines. Participants or their families received no financial or material compensation for their participation in this study. This paper does not contain any images or documents that could identify individual participants.

Results

Table 2 presents the anthropometric characteristics between the ASD experimental group to the ASD control group at baseline (T1). The results show that there are no significant differences in the anthropometric characteristics measured between the children in the experimental group and those in the control group at baseline. This suggests that the 2 groups are well balanced in terms of age, BM, BH, BMI, and BSA, which is crucial to ensure that any future intervention can be evaluated without bias related to initial anthropometric differences between the groups. The *P* values indicate that all comparisons are nonsignificant, with values well above the conventional significance limit of ≤.05, reinforcing the idea of homogeneity between the groups at the start of the study.

Table 2. Anthropometric characteristics of children with autism spectrum disorder according to group membership at baseline (T1).

Variables	Experimental group (n=7)		Control group (n=7)		*P* value[a]
	Mean (SD)	95% CI	Mean (SD)	95% CI	
Age (years)	8.5 (1.2)	7.3-9.5	9.1 (1.4)	7.9-10.5	.29
Body mass (kg)	31.0 (7.3)	24.3-37.8	32.7 (8.0)	25.3-40.1	.68
Body height (cm)	131.7 (7.8)	124.5-138.9	133.7 (7.3)	126.9-140.5	.63
BMI (kg/m^2)	17.9 (3.5)	14.6-21.1	18.1 (2.5)	15.8-20.4	.88
Body surface area (m^2)	1.06 (0.14)	0.93-1.19	1.10 (0.16)	0.95-1.25	.65

[a]Significant at ≤.05.

Table 3 presents a comprehensive comparative analysis of the motor skill performance evolution among children with ASD, spanning from the initial baseline (T1) to the postintervention period (T2). The data are categorized into 2 groups: an ASD experimental group and an ASD control group. In the ASD control group, significant enhancements were observed in upper and lower limb speed, along with ball dribbling and 5-m shuttle run after 15 weeks in the classic PE program. However, there were no significant differences noted in balance duration with eyes opened and closed. Similarly, variables related to running and coordination, such as circle run, side-step run, 5-m shuttle run, target ball toss, and ball dribbling, did not exhibit significant improvements. Regarding the ASD experimental group, noteworthy advancements were recorded across all variables after 15 weeks in the sensorimotor integration and hyperstimulation program. Examining the overarching changes in motor performance between T1 and T2, a distinct contrast emerges between the 2 ASD groups. The ASD control group witnessed an overall progression of 14.5% across all evaluated motor skill variables. Conversely, the ASD experimental group demonstrated a notably higher progression, reaching 44.5%.

XSL•FO
RenderX

Table 3. Comparison of motor skill improvements from baseline (T1) to postintervention (T2) in autism spectrum disorder (ASD) control and experimental groups[a].

Variables	T1, mean (SD)	T2, mean (SD)	*P* value[b]	Cohen *d* effect size[c]	Δ T1 vs T2 (%)
ASD experimental group					
Upper limb speed (nb/20 seconds[d])	23.7 (6.3)	37.7 (4.2)	.001	2.12	14.0 (37.1)
Lower limb speed (nb/20 seconds)	9.0 (2.1)	16.4 (2.2)	<.001	6.74	7.4 (45.1)
Balance eyes opened (seconds)	6.1 (2.6)	14.6 (8.7)	.02	1.21	8.5 (58.2)
Balance eyes closed (seconds)	2.3 (0.9)	6.4 (2.0)	.001	1.91	4.1 (64.1)
Ball dribbling (nb/20 seconds)	5.3 (4.0)	22.6 (10.9)	.003	1.88	17.3 (76.6)
Target ball toss (points)	3.1 (2.2)	8.1 (3.1)	.01	0.89	5.0 (61.7)
Side-step run (seconds)	20.4 (3.7)	13.9 (0.9)	.003	1.61	6.5 (31.9)
Circle run (seconds)	40.0 (5.8)	28.1 (1.1)	.002	1.13	11.9 (30)
5-m shuttle run (seconds)	18.4 (0.8)	15.4 (1.1)	.001	2.15	3.0 (16.3)
Simple reaction time (ms)	299.5 (45.6)	223.4 (24.6)	.002	1.78	76.1 (25.4)
Mean (SD)	__[e]	—	—	2.14 (1.67)	44.5 (19.8)
ASD control group					
Upper limb speed (nb/20 seconds)	22.9 (4.9)	27.7 (4.7)	.01	1.45	4.8 (17.3)
Lower limb speed (nb/20 seconds)	6.9 (2.0)	10.9 (2.6)	.002	2.28	4.0 (36.7)
Balance eyes opened (seconds)	3.2 (1.2)	4.6 (2.6)	.28	0.29	1.4 (30.4)
Balance eyes closed (seconds)	1.9 (0.9)	1.9 (1.3)	.91	0.00	0.0 (0)
Ball dribbling (nb/20 seconds)	3.0 (0.8)	6.7 (3.4)	.01	3.30	3.7 (55.2)
Target ball toss (points)	1.7 (1.9)	1.6 (1.4)	.79	0.10	–0.1 (–5.9)
Side-step run (seconds)	21.6 (2.9)	20.5 (4.1)	.25	0.74	1.1 (5.1)
Circle run (seconds)	39.9 (7.1)	37.2 (5.6)	.06	2.12	2.7 (6.8)
5-m shuttle run (seconds)	18.9 (3.7)	20.6 (3.2)	.007	4.57	–1.7 (–8.3)
Simple reaction time (ms)	336.2 (33.2)	310.6 (26.9)	.14	0.67	25.6 (7.6)
Mean (SD) (%)	—	—	—	1.55 (1.51)	14.5 (20.4)

[a]Intergroup differences were measured using a 2-tailed *t* test for unpaired values. Changes between T1 and T2 for each group were calculated using a 2-tailed *t* test for paired values.

[b]Significant at $P \leq .05$.

[c]<0.20=marginal; 0.20-0.49=small; 0.50-0.79=moderate; and ≥0.80=large.

[d]nb/20 seconds=number of repetitions in 20 seconds.

[e]Not applicable.

The results outlined in Table 4 reveal significant effects of group, time, and group×time interactions for the majority of gross motor skills. Variables such as segmental speed, balance, ball dribbling, and SRT exhibit the most substantial effects. While improvements are also noted in side-step running and circle running, these gains are comparatively modest. Mauchly test of sphericity indicated that the assumption of sphericity was not violated for any of the dependent variables, suggesting that the variances of the differences between levels of the repeated measures factor were approximately equal. Thus, these findings demonstrate that the intervention was highly effective in enhancing motor performance within the experimental group.

XSL•FO
RenderX

Table 4. Summary of a 2-way ANOVA with repeated measures evaluating the effects of group, time, and group×time interaction on gross motor skills performance between the experimental and control groups of children with autism spectrum disorder[a].

Variables and source	Sum of squares	*df*	Mean square	*F* test (*df*=12)[b]	*P* value[c]	η^2p[d]
Upper limb speed (nb/20 seconds[e])						
Group	206.3	1	206.3	5.361	.04	0.309
Time	622.3	1	622.3	48.044	<.001	0.800
Group×time	146.3	1	146.3	11.294	.006	0.485
Lower limb speed (nb/20 seconds)						
Group	104.1	1	104.1	12.941	.004	0.519
Time	228.6	1	228.6	145.455	<.001	0.924
Group×time	20.6	1	20.6	13.091	.004	0.522
Balance eyes opened (seconds)						
Group	365.6	1	365.6	19.393	.001	0.618
Time	260.7	1	260.7	33.560	<.001	0.737
Group×time	133.7	1	133.7	17.206	.001	0.589
Balance eyes closed (seconds)						
Group	42.9	1	42.9	21.122	.001	0.638
Time	30.6	1	30.6	20.113	.001	0.626
Group×time	28.5	1	28.5	18.761	.001	0.610
Ball dribble (nb/20 seconds)						
Group	576.0	1	576.0	11.576	.005	0.491
Time	771.8	1	771.8	31.998	<.001	0.727
Group×time	322.3	1	322.3	13.364	.003	0.527
Target ball toss (points)						
Group	112.0	1	112.0	18.556	.001	0.607
Time	41.3	1	41.3	10.671	.007	0.471
Group×time	46.3	1	46.3	11.963	.005	0.499
Side-step run (seconds)						
Group	106.4	1	106.4	7.020	.21	0.369
Time	103.4	1	103.4	22.549	<.001	0.653
Group×time	50.5	1	50.5	11.004	.006	0.478
Circle run (seconds)						
Group	144.3	1	144.3	3.142	.10	0.207
Time	372.7	1	372.7	30.427	<.001	0.717
Group×time	148.9	1	148.9	12.151	.004	0.503
5-m shuttle run (seconds)						
Group	56.3	1	56.3	4.669	.05	0.280
Time	2.5	1	2.5	3.550	.08	0.228
Group×time	39.1	1	39.1	56.465	<.001	0.825
Simple reaction time (ms)						
Group	26,859.3	1	26,859.3	18.060	.001	0.601
Time	18,105.4	1	18,105.4	28.592	<.001	0.704
Group×time	4450.7	1	4450.7	7.028	.02	0.369

XSL•FO
RenderX

[a]Mauchly sphericity test≥0.05 for all the variables in the analysis.

[b]*F*=variance between groups/variance within groups.

[c]Significant if ≤.05.

[d]Small effect size: η^2p=0.01; moderate effect size: η^2p=0.06; and large effect size: η^2p=0.14.

[e]nb/20 seconds=number of repetitions in 20 seconds.

Discussion

Principal Findings

Autism, a complex spectrum of neurodevelopmental conditions, is characterized by a wide range of manifestations and challenges. Recent research, such as that conducted by Khoury et al [52], has shed light on a connection between sensorimotor integration disorders and ASD. These disorders impact both fine and gross motor skills, as stated by some studies that identified 2 primary anomalies contributing to the motor difficulties observed in individuals with ASD, which are impaired integration of information crucial for motor planning and heightened variability in basic sensory input and output [24,53].

Characteristics of Motor Performance in Individuals With ASD

Unlike individuals with a neurotypical profile, who readily integrate information from multiple senses (multimodal integration), people with ASD often process this information independently for each sense (unimodal integration). This unique characteristic can lead to difficulties in perception when exposed to simultaneous stimulation across multiple senses. As part of a study comparing the motor performances of individuals with and without ASD, the battery of tests proposed by Leone et al [43,44] proved to be a relevant tool. First, this battery effectively evaluates a broad range of fundamental motor skill factors (segmental speed, coordination, balance, agility, and reaction time), which was important in the context of this study. Second, the test items are designed as simple, isolated tasks, minimizing the influence of other motor components. This structure aimed to reduce potential performance discrepancies between the 2 participant groups. However, despite these precautions, the study revealed significant differences in motor performance between children with a neurotypical profile and those with ASD from the outset. Children with ASD demonstrated a marked delay in motor development, even in tasks primarily reliant on unimodal sensory processing.

The Classic Versus the Sensorimotor Integration and Hyperstimulation Approach

In efforts to integrate children with ASD into the education system, it is common for PE lessons to be taught using a classical method similar to that used for children with a neurotypical profile. Morocco follows this trend and offers children with autism the same PE program prescribed by the Ministry of National Education. Although some adaptations are implemented, such as reducing class sizes, this type of program, which is based on disparate and eclectic exercises derived from sports activities, does not sufficiently consider the unique characteristics of children with autism. This approach results in a significant increase in the quantity and variety of sensory and motor stimulation, which conflicts with the typical functioning of children with autism. Additionally, these children often experience sensory hyperreactivity or hyporeactivity to environmental stimuli such as brightness, ground texture, and ambient noise. This can disrupt their routine and, consequently, hinder their learning capacity [54]. Despite their limitations, traditional PE classes can demonstrably contribute to improve some motor skills in children with ASD. Our study showed that a 15-week classical PE program allowed a nearly 14.5% overall improvement in the control group of children with ASD. This finding highlights the capacity of traditional PE methods to induce, to a certain extent, a contribution to the motor development of this population.

In contrast, the experimental group, whose PE lessons were adapted based on principles of sensorimotor integration and hyperstimulation management, demonstrated a remarkable 44.5% improvement over the same 15-week period. Aware of the debates that this approach still generates, we have introduced innovations that, in our opinion, have contributed to this impressive outcome. While sensorimotor integration appears more appropriate for children with ASD than traditional methods, most research advocates for a unimodal approach, emphasizing tasks that engage a single sense. Although some studies show promising results, others question the effectiveness of this method, suggesting that it fails to adequately address the specific needs of children with ASD.

In this study, a multimodal approach was adopted. Each block focuses on a specific single motor skill in a manner that engages multiple senses simultaneously. For instance, while block 1 emphasizes coordination, learning situations necessitate the involvement of various senses to accomplish each task. Furthermore, to control for environmental variables and minimize external influences, all sessions were conducted in a single room, ensuring consistent lighting, playing surface, and ambient sound when the learning situation requires it. Given its unique characteristics, the program structure developed in this study represents a valuable contribution to the field.

ASD Experimental Versus Control Group

The results of this study are highly encouraging and suggest that the implemented intervention had a significant impact on improving the motor skills of children with ASD. Indeed, all motor skill factors assessed showed significant improvement in the experimental group compared to the control group (P<.05). The enhancements were not only statistically significant but also clinically relevant, as indicated by large ESs (Cohen ES>0.80). These findings support the idea that targeted interventions can enhance motor skills in children with ASD, with potential broad-ranging implications for their overall development.

The results of the 2-way repeated measures ANOVA strongly support the effectiveness of the sensorimotor integration and hyperstimulation method. The findings reveal that the

XSL•FO
RenderX

group×time interaction is significant for several variables, indicating that performance varies not only by group but also over time. This interaction suggests that the effect of time on motor performance is not consistent across groups. In other words, the interventions or experimental conditions had different impacts depending on group membership. It is therefore likely that the experimental group showed more substantial improvements over time compared to the control group, which could suggest greater effectiveness of the implemented intervention. The ESs (η^2p) associated with these interactions are also noteworthy. The high ESs further emphasize that the intervention had a significant clinical impact. These findings emphasize the importance of individualized training programs that consider specific needs and learning styles to optimize motor development outcomes.

The observed improvements spanned a wide range of motor domains, including coordination, balance, and reaction time, suggesting a comprehensive enhancement of motor skills. This study highlights the importance of incorporating specific physical activities into educational programs for children with ASD. By promoting motor skill development, one can significantly improve these children's quality of life, learning, and social inclusion. Moreover, this study raises intriguing questions regarding the underlying neurobiological mechanisms that may account for the observed improvements in children with ASD following the intervention. While further research is needed to draw definitive conclusions, several potential mechanisms could be explored.

Potential Neurobiological Mechanisms Underlying Motor Skill Improvements in Children With ASD

Regular physical exercise stimulates the production of neuronal growth factors, which promote the formation of new synaptic connections in brain regions involved in motor control. Training can lead to the reorganization of neural networks, enhancing movement coordination and motor planning [55]. Physical activity is linked to dopamine release, a neurotransmitter integral to reward systems. This could boost children's motivation to engage in physical activities and reinforce adaptive motor behaviors. Dopamine also plays a pivotal role in motor control, and its modulation could contribute to the observed improvements [56]. Additionally, serotonin, which regulates mood and anxiety, may see increased activity due to physical exercise, potentially reducing the anxiety symptoms often associated with ASD and facilitating the acquisition of new skills [57]. Physical exercise can enhance sensory perception, including proprioception and vestibular sense, thereby improving movement coordination [58]. Our proposed sensorimotor integration and hyperstimulation program specifically addresses these 2 sensory modalities. Although these hypotheses are plausible, further research is needed to assess their validity.

Implementation of a Sensorimotor Integration Program in the Moroccan Context

Similar to many societies, Morocco often prioritizes the classical PE approach to stimulate motor development in children, including those with ASD. However, the findings of this exploratory study suggest that an alternative approach focused on sensorimotor integration and hyperstimulation management could be particularly beneficial for Moroccan children with ASD. Conducted within the Moroccan cultural and educational context, the study demonstrated that this innovative approach could not only be successfully implemented but also yield significantly better results in motor development compared to the classical approach. While traditional PE classes can offer some benefits to children with ASD, integrating principles of sensorimotor integration and hyperstimulation management into PE programs can lead to more substantial improvements. The use of a multimodal approach, where tasks are designed to engage multiple senses, appears to be particularly effective. This method not only addresses the unique sensory processing challenges faced by children with ASD but also helps to bridge the developmental gap between them and their peers with a neurotypical profile. As such, the program developed in this study represents a significant advancement in the field of PE for children with ASD, providing a framework for future research and practice.

Limitations and Strengths

This study reveals several limitations. The small number of participants with ASD (n=14) requires caution in interpreting the results. Additionally, the fact that the age range is limited to 6 to 12 years makes the results potentially inapplicable to other age groups. Given that these children mainly come from disadvantaged backgrounds, it is uncertain whether the same findings would be observed in different socioeconomic contexts. All participants with ASD were classified as level 2 (moderate) on the Autism Severity Rating Scale according to DSM-5-TR criteria, so it is not possible to judge whether the program would have the same effectiveness for severity levels 1 and 3. However, this limitation ensures that the group of children with ASD was homogeneous, which is often a weakness in other studies. Additionally, obtaining repeated measurements allows changes to be better assessed since they come from the same children at different time periods. This study also has the advantage of providing, for the first time, a preliminary portrait of the motor development of Moroccan children with a neurotypical profile versus children with ASD based on the UQAC-UQAM test battery. The proposed approach (sensorimotor integration and hyperstimulation), as implemented in this study, is also a first for the Moroccan population with ASD.

Conclusions

Children with ASD often exhibit a unimodal sensory processing style, meaning they tend to focus on one sensory input at a time. This can lead to significant perceptual challenges when they are exposed to stimulation that engages multiple senses simultaneously. Conventional PE programs, which typically use a variety of exercises derived from various sports, may actually hinder the progress of children with ASD by overloading their sensory systems. In contrast, a novel approach that emphasizes sensorimotor integration and hyperstimulation management appears to be much more effective in improving motor skills in this population. This method prioritizes engaging multiple senses concurrently while minimizing excessive sensory stimuli. This study revealed that the sensorimotor

XSL•FO
RenderX

integration and hyperstimulation management program yielded a remarkable 44.5% improvement in motor skills for children in the experimental group over a 15-week period. This is in stark contrast to the 14.5% improvement observed in the control group who participated in the traditional PE program. While this study is limited by the sample size, it provides a groundbreaking initial assessment of motor development in Moroccan children with ASD. Furthermore, it proposes an innovative approach specifically tailored to their motor skills condition and the prevailing sociocultural and educational context within Morocco. However, other more exhaustive studies will need to be carried out in order to be able to more precisely measure the full potential of this approach.

Acknowledgments

The authors wish to acknowledge the school principal, the specialized educators' team, children, and parents or legal guardians who agreed to participate in this study. This research received no external funding.

Data Availability

The original contributions presented in the study are included in the paper. Further inquiries can be directed to the corresponding author. All data generated or analyzed during this study are included in this published paper.

Authors' Contributions

RT was involved in the design and concept of the study and data collection, drafted the initial version of the manuscript, and reviewed and revised the manuscript. JZ and EMC coordinated and supervised data collection, were involved in the study design, and reviewed and revised the manuscript. HTB was involved in the data analysis and reviewed and drafted the manuscript for important intellectual content. ML and MA were involved in the study design and initial analyses, drafted the initial manuscript, and reviewed and revised the manuscript. All authors approved the final manuscript as submitted and agreed to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The intervention program for the experimental group.
[PDF File (Adobe PDF File), 381 KB - formative_v9i1e65767_app1.pdf]

Multimedia Appendix 2
The intervention program for the control group.
[PDF File (Adobe PDF File), 694 KB - formative_v9i1e65767_app2.pdf]

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-5-TR, 5th Edition. Washington, DC: American Psychiatric Publishing; 2022.
2. Christensen DL, Baio J, van Naarden Braun K, Bilder D, Charles J, Constantino JN, et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2012. MMWR Surveill Summ 2016 Apr 01;65(3):1-23 [FREE Full text] [doi: 10.15585/mmwr.ss6503a1] [Medline: 27031587]
3. Le Menn-Tripi C, Vachaud A, Defas N, Malvy J, Roux S, Bonnet-Brilhault F. Sensory-psychomotor evaluation in Autism: a new tool for functional diagnosis. Encephale 2019;45(4):312-319. [doi: 10.1016/j.encep.2018.12.003] [Medline: 30922517]
4. Odeh CE, Gladfelter AL, Stoesser C, Roth S. Comprehensive motor skills assessment in children with autism spectrum disorder yields global deficits. Int J Dev Disabil 2022;68(3):290-300 [FREE Full text] [doi: 10.1080/20473869.2020.1764241] [Medline: 35602998]
5. Paquet A, Olliac B, Golse B, Vaivre-Douret L. Nature of motor impairments in autism spectrum disorder: a comparison with developmental coordination disorder. J Clin Exp Neuropsychol 2018;41(1):1-14. [doi: 10.1080/13803395.2018.1483486] [Medline: 29923455]
6. Rogers S, Bennetto L. Motor functioning in autism [in French]. Enfance 2002;54(1):63-73 [FREE Full text]
7. Wilson RB, Enticott PG, Rinehart NJ. Motor development and delay: advances in assessment of motor skills in autism spectrum disorders. Curr Opin Neurol 2018;31(2):134-139 [FREE Full text] [doi: 10.1097/WCO.0000000000000541] [Medline: 29493557]

XSL•FO
RenderX

8. Ketcheson LR, Pitchford EA, Wentz CF. The relationship between developmental coordination disorder and concurrent deficits in social communication and repetitive behaviors among children with Autism spectrum disorder. Autism Res 2021;14(4):804-816. [doi: 10.1002/aur.2469] [Medline: 33421296]
9. Miller HL, Sherrod GM, Mauk JE, Fears NE, Hynan LS, Tamplain PM. Shared features or co-occurrence? Evaluating symptoms of developmental coordination disorder in children and adolescents with autism spectrum disorder. J Autism Dev Disord 2021;51(10):3443-3455 [FREE Full text] [doi: 10.1007/s10803-020-04766-z] [Medline: 33387238]
10. Zampella CJ, Wang LAL, Haley M, Hutchinson AG, de Marchena A. Motor skill differences in autism spectrum disorder: a clinically focused review. Curr Psychiatry Rep 2021;23(10):64. [doi: 10.1007/s11920-021-01280-6] [Medline: 34387753]
11. Gillberg C. Deficits in attention, motor control, and perception: a brief review. Arch Dis Child 2003;88(10):904-910 [FREE Full text] [doi: 10.1136/adc.88.10.904] [Medline: 14500312]
12. Hocking DR, Caeyenberghs K. What is the nature of motor impairments in autism, are they diagnostically useful, and what are the implications for intervention? Curr Dev Disord Rep 2017;4(2):19-27. [doi: 10.1007/s40474-017-0109-y]
13. Kindregan D, Gallagher L, Gormley J. Gait deviations in children with autism spectrum disorders: a review. Autism Res Treat 2015;2015:741480-741488 [FREE Full text] [doi: 10.1155/2015/741480] [Medline: 25922766]
14. Srinivasan SM, Kaur M, Park IK, Gifford TD, Marsh KL, Bhat AN. The effects of rhythm and robotic interventions on the imitation/praxis, interpersonal synchrony, and motor performance of children with autism spectrum disorder (ASD): a pilot randomized controlled trial. Autism Res Treat 2015;2015:736516 [FREE Full text] [doi: 10.1155/2015/736516] [Medline: 26793394]
15. Volkmar F, Siegel M, Woodbury-Smith M, King B, McCracken J, State M. Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. J Am Acad Child Adolesc Psychiatry 2014;53(2):237-257 [FREE Full text] [doi: 10.1016/j.jaac.2013.10.013] [Medline: 24472258]
16. Fournier KA, Kimberg CI, Radonovich KJ, Tillman MD, Chow JW, Lewis MH, et al. Decreased static and dynamic postural control in children with autism spectrum disorders. Gait Posture 2010;32(1):6-9 [FREE Full text] [doi: 10.1016/j.gaitpost.2010.02.007] [Medline: 20400311]
17. Gernsbacher MA, Sauer EA, Geye HM, Schweigert EK, Hill Goldsmith H. Infant and toddler oral- and manual-motor skills predict later speech fluency in autism. J Child Psychol Psychiatry 2008;49(1):43-50 [FREE Full text] [doi: 10.1111/j.1469-7610.2007.01820.x] [Medline: 17979963]
18. Lim YH, Partridge K, Girdler S, Morris SL. Standing postural control in individuals with autism spectrum disorder: systematic review and meta-analysis. J Autism Dev Disord 2017;47(7):2238-2253. [doi: 10.1007/s10803-017-3144-y] [Medline: 28508177]
19. Pace M, Dumortier L, Guinot M, Favre-Juvin A, Bricout VA. Assessment of motor skills in autistic children on the M-ABC [in French]. Sci Sports 2016;31(6):336-341. [doi: 10.1016/j.scispo.2016.08.001]
20. Coll SM, Foster NE, Meilleur A, Brambati SM, Hyde KL. Sensorimotor skills in autism spectrum disorder: a meta-analysis. Res Autism Spectr Disord 2020;76:101570. [doi: 10.1016/j.rasd.2020.101570]
21. Bojanek EK, Wang Z, White SP, Mosconi MW. Postural control processes during standing and step initiation in autism spectrum disorder. J Neurodev Disord 2020;12(1):1 [FREE Full text] [doi: 10.1186/s11689-019-9305-x] [Medline: 31906846]
22. Casey BJ, Glatt CE, Lee FS. Treating the developing versus developed brain: translating preclinical mouse and human studies. Neuron 2015;86(6):1358-1368. [doi: 10.1016/j.neuron.2015.05.020] [Medline: 26087163]
23. Gowen E, Hamilton A. Motor abilities in autism: a review using a computational context. J Autism Dev Disord 2013;43(2):323-344. [doi: 10.1007/s10803-012-1574-0] [Medline: 22723127]
24. Smoot Reinert S, Jackson K, Bigelow K. Using posturography to examine the immediate effects of vestibular therapy for children with Autism spectrum disorders: a feasibility study. Phys Occup Ther Pediatr 2015;35(4):365-380. [doi: 10.3109/01942638.2014.975313] [Medline: 25374155]
25. Zamani Jam A, Talab RH, Sheikh M, Torabi F, Rafie F. The effect of 16 weeks gymnastic training on social skills and neuropsychiatric functions of autistic children. Sport Sci Health 2017;14(1):209-214. [doi: 10.1007/s11332-017-0417-y]
26. Kaur M, Bhat A. Creative yoga intervention improves motor and imitation skills of children with autism spectrum disorder. Phys Ther 2019;99(11):1520-1534 [FREE Full text] [doi: 10.1093/ptj/pzz115] [Medline: 31765484]
27. Pan CY, Chu CH, Tsai CL, Sung MC, Huang CY, Ma WY. The impacts of physical activity intervention on physical and cognitive outcomes in children with autism spectrum disorder. Autism 2017;21(2):190-202. [doi: 10.1177/1362361316633562] [Medline: 27056845]
28. Najafabadi MG, Sheikh M, Hemayattalab R, Memari A, Aderyani MR, Hafizi S. The effect of SPARK on social and motor skills of children with autism. Pediatr Neonatol 2018;59(5):481-487 [FREE Full text] [doi: 10.1016/j.pedneo.2017.12.005] [Medline: 29402579]
29. Desrochers K, Rivard MC, Dugas C. Analysis of a sensory-motor intervention program on the global and motor development of students with autism spectrum disorder [in French]. Can J New Sch Educ No 2017;8:1 [FREE Full text]
30. Valverde-Esteve T, Salvador-Garcia C, Gil-Gómez J, Maravé-Vivas M. Sustainable service-learning in physical education teacher education: examining postural control to promote ASD children's well-being. Int J Environ Res Public Health 2021;18(10):5216 [FREE Full text] [doi: 10.3390/ijerph18105216] [Medline: 34068915]

XSL•FO
RenderX

31. Rafie F, Ghasemi A, Zamani Jam A, Jalali S. Effect of exercise intervention on the perceptual-motor skills in adolescents with autism. J Sports Med Phys Fitness 2017;57(1-2):53-59. [doi: 10.23736/s0022-4707.16.05919-3]
32. Ayres AJ, Robbins J. Sensory Integration and the Child: Understanding Hidden Sensory Challenges, Pediatric Therapy Network ed. Los Angeles, CA: Western Psychological Services; 2005.
33. Polatajko HJ, Macnab JJ, Anstett B, Malloy-Miller T, Murphy K, Noh S. A clinical trial of the process-oriented treatment approach for children with developmental co-ordination disorder. Dev Med Child Neurol 1995;37(4):310-319. [doi: 10.1111/j.1469-8749.1995.tb12009.x] [Medline: 7535267]
34. May-Benson TA, Koomar JA. Systematic review of the research evidence examining the effectiveness of interventions using a sensory integrative approach for children. Am J Occup Ther 2010;64(3):403-414. [doi: 10.5014/ajot.2010.09071] [Medline: 20608272]
35. Baranek GT. Efficacy of sensory and motor interventions for children with autism. J Autism Dev Disord 2002;32(5):397-422. [doi: 10.1023/a:1020541906063] [Medline: 12463517]
36. Watling R, Hauer S. Effectiveness of ayres sensory integration® and sensory-based interventions for people with Autism spectrum disorder: a systematic review. Am J Occup Ther 2015;69(5):1-12. [doi: 10.5014/ajot.2015.018051] [Medline: 26356655]
37. Watling R, Deitz J, Kanny EM, McLaughlin JF. Current practice of occupational therapy for children with autism. Am J Occup Ther 1999;53(5):498-505. [doi: 10.5014/ajot.53.5.498] [Medline: 10500858]
38. Pfeiffer B, Daly BP, Nicholls EG, Gullo DF. Assessing sensory processing problems in children with and without attention deficit hyperactivity disorder. Phys Occup Ther Pediatr 2015;35(1):1-12. [doi: 10.3109/01942638.2014.904471] [Medline: 24712841]
39. Bardid F, Rudd JR, Lenoir M, Polman R, Barnett LM. Cross-cultural comparison of motor competence in children from Australia and Belgium. Front Psychol 2015;6:964 [FREE Full text] [doi: 10.3389/fpsyg.2015.00964] [Medline: 26217282]
40. Chow SM, Henderson SE, Barnett AL. The movement assessment battery for children: a comparison of 4-year-old to 6-year-old children from Hong Kong and the United States. Am J Occup Ther 2001;55(1):55-61. [doi: 10.5014/ajot.55.1.55] [Medline: 11216367]
41. Lohman TG, Roche AF, Martorell R. Anthropometric Standardization Reference Manual. Champaign, IL: Human Kinetics; 1988.
42. Mosteller RD. Simplified calculation of body-surface area. N Engl J Med 1987;317(17):1098-1098. [doi: 10.1056/nejm198710223171717]
43. Leone M, Viret P, Bui HT, Laverdière C, Kalinova E, Comtois AS. Assessment of gross motor skills and phenotype profile in children 9-11 years of age in survivors of acute lymphoblastic leukemia. Pediatr Blood Cancer 2014;61(1):46-52. [doi: 10.1002/pbc.24731] [Medline: 24115507]
44. Leone M, Comtois AS, Kalinova E, Perron M, Babineau C. Motor skills assessment of Canadian school boys and girls aged 6 to 12 years old. Université du Québec à Chicoutimi. 2010. URL: https://constellation.uqac.ca/id/eprint/2475/ [accessed 2025-03-05]
45. Gosselin V, Leone M, Laberge S. Socioeconomic and gender-based disparities in the motor competence of school-age children. J Sports Sci 2021;39(3):341-350. [doi: 10.1080/02640414.2020.1822585] [Medline: 32967566]
46. Aranha VP, Saxena S, Moitra M, Narkeesh K, Arumugam N, Samuel AJ. Reaction time norms as measured by ruler drop method in school-going South Asian children: a cross-sectional study. Homo 2017;68(1):63-68. [doi: 10.1016/j.jchb.2016.12.001] [Medline: 28011080]
47. Fleishman EA. The Structure and Measurement of Physical Fitness. Englewood Cliffs, NJ: Prentice-Hall; 1964.
48. Ángel Latorre-Roman P, Robles-Fuentes A, García-Pinillos F, Salas-Sánchez J. Reaction times of preschool children on the ruler drop test: a cross-sectional study with reference values. Percept Mot Skills 2018;125(5):866-878. [doi: 10.1177/0031512518789563] [Medline: 30032724]
49. Allisse M, Thibault I, Gagnon D, Kalinova E, Larivière G, Leone M. The relationship between obesity status and body image dissatisfaction on gross motor skill development and cardiorespiratory fitness in children aged 6–12 years old. Int J Environ Res Public Health 2025 Mar 12;22(3):417. [doi: 10.3390/ijerph22030417]
50. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2025-03-19]
51. Cohen J. A power primer. Psychol Bull 1992;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
52. Khoury E, Carment L, Lindberg P, Gaillard R, Krebs MO, Amado I. Sensorimotor aspects and manual dexterity in autism spectrum disorders: a literature review. Encephale 2020;46(2):135-145. [doi: 10.1016/j.encep.2019.11.005] [Medline: 31928741]
53. Puyjarinet F, Galbiati C. The evaluation approach for developmental coordination disorder in occupational therapy and psychomotor skills: points of convergence and specificities (in French). Le Pédiatre 2020;296:12-16.
54. Ashburner JK, Rodger SA, Ziviani JM, Hinder EA. Optimizing participation of children with autism spectrum disorder experiencing sensory challenges: a clinical reasoning framework. Can J Occup Ther 2014;81(1):29-38. [doi: 10.1177/0008417413520440] [Medline: 24783486]

XSL•FO
RenderX

55. Erickson KI, Hillman CH, Kramer AF. Physical activity, brain, and cognition. Curr Opin Behav Sci 2015;4:27-32. [doi: 10.1016/j.cobeha.2015.01.005]
56. Best JR. Effects of physical activity on children's executive function: contributions of experimental research on aerobic exercise. Dev Rev 2010;30(4):331-551 [FREE Full text] [doi: 10.1016/j.dr.2010.08.001] [Medline: 21818169]
57. Canli T, Lesch KP. Long story short: the serotonin transporter in emotion regulation and social cognition. Nat Neurosci 2007;10(9):1103-1109. [doi: 10.1038/nn1964] [Medline: 17726476]
58. Pedersen BK, Febbraio G. Muscle as an endocrine organ: focus on muscle-derived interleukin-6. Physiol Rev 2008;88(4):1379-1406 [FREE Full text] [doi: 10.1152/physrev.90100.2007] [Medline: 18923185]

Abbreviations

ASD: autism spectrum disorder
BH: body height
BM: body mass
BSA: body surface area
DCD: developmental coordination disorder
DSM-5-TR: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision
ES: effect size
PE: physical education
SRT: simple reaction time
UQAC-UQAM: University of Québec in Chicoutimi-University of Québec in Montréal

Edited by A Mavragani; submitted 26.08.24; peer-reviewed by K Suzuki; comments to author 22.01.25; revised version received 31.01.25; accepted 18.02.25; published 26.03.25.

Please cite as:
Touali R, Zerouaoui J, Chakir EM, Bui HT, Leone M, Allisse M
Impact of a Sensorimotor Integration and Hyperstimulation Program on Global Motor Skills in Moroccan Children With Autism Spectrum Disorder: Exploratory Clinical Quasi-Experimental Study
JMIR Form Res 2025;9:e65767
URL: https://formative.jmir.org/2025/1/e65767
doi:10.2196/65767
PMID:40137439

©Rachid Touali, Jamal Zerouaoui, El Mahjoub Chakir, Hung Tien Bui, Mario Leone, Maxime Allisse. Originally published in JMIR Formative Research (https://formative.jmir.org), 26.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Norwegian Version of the Self-Efficacy in Clinical Performance Scale (SECP): Psychometric Validation Study

Camilla Olaussen[1], RN, MSc, PhD; Marko Stojiljkovic[2], RN, MSc; Jaroslav Zlamal[1,3], RN, MSc; Tone Nygaard Fløl[2,4], RN, MSc, PhD; Andréa Aparecida Gonçalves Nes[1,5], RN, MSc, PhD

[1]Department of Nursing, Lovisenberg Diaconal University College, Lovisenberggt. 15B, Oslo, Norway

[2]Department of Nursing and Health Promotion, Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway

[3]Clinical Simulation Center, Vestfold Hospital Trust, Tønsberg, Norway

[4]Voss Hospital, Haukeland University Hospital, Voss, Norway

[5]Department of Caring and Ethics, Faculty of Health Sciences, University of Stavanger, Stavanger, Norway

Corresponding Author:
Camilla Olaussen, RN, MSc, PhD
Department of Nursing, Lovisenberg Diaconal University College, Lovisenberggt. 15B, Oslo, Norway

Abstract

Background: Previous research has demonstrated a correlation between nursing students' self-efficacy and their clinical performance, competence, and behavior during clinical practice placements. Assessing students' self-efficacy in clinical performance could be a valuable method for identifying areas that need reinforcement and for recognizing students who may require additional support during clinical practice placements.

Objective: This study aimed to translate the Self-Efficacy in Clinical Performance Scale (SECP) from English into Norwegian and to evaluate the psychometric properties of the Norwegian version.

Methods: A cross-sectional study design was used. The SECP was translated into Norwegian following a 6-step process: forward translation, forward translation synthesis, backward translation, backward translation synthesis, cognitive debriefing, and psychometric testing. The validity and reliability of the translated version were assessed using confirmatory factor analysis (CFA), Cronbach α, McDonald ω, and composite reliability.

Results: A total of 399 nursing students completed the Norwegian version of the SECP. The CFA goodness-of-fit indices (χ^2/df ratio=1.578, comparative fit index=0.98, Tucker-Lewis index=0.98, standardized root mean square residual=0.056, root mean square error of approximation=0.038) indicated an acceptable model fit. Reliability measures, including Cronbach , McDonald ω, and composite reliability, were high, with factor-level values ranging from 0.94 to 0.98.

Conclusion: The Norwegian version of the SECP demonstrated strong potential as an instrument for assessing self-efficacy in both current and required competencies among nursing students in clinical practice within nursing education. Future research should aim to confirm the factor structure of the SECP and evaluate its test-retest reliability.

(JMIR Form Res 2025;9:e68173) doi:10.2196/68173

KEYWORDS

clinical performance; self-efficacy; instrument validation; nursing education; psychometric analysis; Norway; psychometric; validation study; competence; clinical practice; translate; translation; cross-sectional study; nursing student; reliability

Introduction

Background

Clinical practice placements in nursing education provide a platform for nursing students to apply theoretical knowledge to real patient care. Nursing knowledge encompasses both theoretical understanding ("knowing that") and practical skills ("knowing how") [1]. The nursing process, originally defined as "a systematic approach to care using the fundamental principles of critical thinking, client-centered approaches to treatment, goal-oriented tasks, evidence-based practice recommendations, and nursing intuition" [2], is recognized as an essential framework for applying nursing knowledge to enhance care quality [3]. The nursing process entails comprehending the rationale behind patients' treatment plans, understanding pathophysiology, identifying patient problems, conducting suitable assessments, prioritizing and implementing care, and evaluating the outcomes of provided care [3]. However, students report difficulties in integrating theoretical knowledge into clinical practice [4,5], as well as challenges in gathering, assessing, and using patient data to develop nursing care that caters to patient needs [6].

XSL•FO
RenderX

Previous studies have demonstrated that nursing students' self-efficacy is associated with their clinical performance, competence, and behavior during clinical practice placements [7-9]. Self-efficacy, a critical factor in academic success, refers to the confidence or belief in one's abilities to successfully accomplish specific tasks and persist despite challenges [10,11]. According to Bandura [10], mastery experiences have a strong effect on a student's self-efficacy development because they are the most authentic indicators of one's capabilities. In contrast, experienced failures may impair students' self-efficacy perceptions and result in avoidance behaviors [10].

Various factors, such as anxiety, stress, motivation, and the pedagogical atmosphere in the clinical setting, can potentially influence nursing students' self-efficacy [12] and thus their clinical achievements [13-15]. A vital resource in promoting nursing students' self-efficacy is emotional support from peers, academic educators, and clinical supervisors [10,12,16]. Hence, to enhance nursing students' clinical learning, performance, and competence during clinical practice placements, it is crucial not only to equip students with a conducive clinical learning environment but also to assist them in boosting their self-efficacy to manage various challenges they might face [8]. As self-efficacy may predict clinical performance and behavior among nursing students, assessing students' self-efficacy could be a valuable method to pinpoint those who may require additional support during clinical practice placements [17,18].

To assess students' self-efficacy in clinical practice, validated instruments are necessary. The Self-Efficacy in Clinical Performance Scale (SECP) was designed to collect empirical data on nursing students' self-efficacy in clinical performance [19]. The SECP explores students' self-efficacy perceptions in performing different facets of the nursing process. Students are queried about their confidence in patient assessments, diagnosis and planning, implementation of care, and evaluation of provided care. Such insights are valuable for educators and clinical supervisors to identify areas requiring reinforcement and to create educational strategies to promote students' self-efficacy, equip them for potential challenges, and ultimately enhance their clinical performance and competence. For these reasons, the SECP was selected for translation and validation into Norwegian in this study, given its established psychometric properties and relevance to clinical education settings [19]. As highlighted by Cheraghi et al [19], the SECP has demonstrated construct validity, internal consistency, and stability, ensuring consistent and reliable measurement of self-efficacy in clinical performance.

Objective

This study was initiated to address the need for culturally adapted and psychometrically validated tools to support educators in enhancing educational strategies in Norwegian nursing education. Therefore, the study aimed to translate the SECP from English into Norwegian and to evaluate its psychometric properties in a Norwegian academic context.

Methods

Design

This study used a cross‐sectional survey design, including translating the SECP and testing its psychometric properties.

Translation Procedure

The original SECP, developed and validated by Cheraghi et al [19], is in Persian and consists of 37 items across 4 factors: assessment, diagnosis and planning, implementation, and evaluation. Each item is rated on an 11-point Likert scale, from 0 ("fully disagree") to 10 ("fully agree"), where higher scores indicate greater levels of agreement, and lower scores indicate disagreement. The hypothesized 4-factor model and corresponding items in the SECP are presented in Table 1.

XSL•FO
RenderX

Table . The hypothesized 4-factor model and corresponding items in SECP[a].

Factors and subscales	Items I am confident that in the clinical setting, I can:
Assessment	Q1. Collect significant data in a physical evaluation.
	Q2. Collect relevant data by obtaining the patient's history.
	Q3. Collect data efficiently, without burdening the patient unnecessarily.
	Q4. Collect data by organizing the available time.
	Q5. Collect objective data related to the patient's health status.
	Q6. Collect subjective data related to the patient's health status.
	Q7. See the relationship between data elements collected from different sources.
	Q8. Document the collected data based on the patient's health status.
	Q9. Analyze the data collected based on the patient's health status.
	Q10. Identify the patient's strengths in the care process.
	Q11. Identify the patient's health concerns in the care process.
	Q12. Prioritize the patient's needs based on the patient's health status.
Diagnosis and planning	Q13. Formulate a nursing diagnosis based on the collected data.
	Q14. Adjust the nursing diagnosis based on an assessment of the patient's data.
	Q15. Adjust the nursing diagnosis based on prioritizing the patient's needs.
	Q16. Formulate the overall goal of the patient's plan of care.
	Q17. Formulate short-term goals for the patient's plan of care
	Q18. Formulate long-term goals for the patient's plan of care.
	Q19. Establish measurable outcomes of care.
	Q20. Based on goals, set up the patient's daily plan of care.
	Q21. Establish a plan of care based on prioritizing the patient's needs.
Implementation	Q22. Implement the patient's established plan of care to attain the goals.
	Q23. Provide nursing care to the patient based on priorities in the plan of care.
	Q24. Implement the patient's plan of care with available resources.
	Q25. Explain each nursing intervention to the patient or family member before implementing it.
	Q26. Work together with the patient or family member in implementing the daily plan of care.
	Q27. Make decisions based on my prior experience in similar situations.
	Q28. Seek help from a mentor or nurse colleagues in difficult situations.
	Q29. Improve my skills based on feedback from a mentor and nursing colleagues.
	Q30. Develop teaching strategies for the patient's discharge.
	Q31. Document and report daily clinical work.
Evaluation	Q32. Evaluate the achievement of the patient's desired outcomes.
	Q33. Evaluate how nursing interventions were performed.
	Q34. Identify weaknesses in the structure of the care plan.
	Q35. Based on the patient's prognosis, determine if the plan of care should be followed as is or modified.
	Q36. Adjust the goals of the care plan in response to changes in the patient's condition.
	Q37. Reprioritize the care plan based on changes in the patient's condition.

[a]SECP: Self-Efficacy in Clinical Performance Scale.

Permission to translate, validate, and use the SECP developers' English and Persian versions of the instrument was obtained via email. The SECP was translated from their English version to Norwegian following six of a seven-step guideline suggested by Sousa and Rojjanasrirat [20] in the symmetrical translation approach: (1) forward translation, (2) synthesis of the forward translation, (3) backward translation, (4) synthesis of the backward translation, (5) cognitive debriefing, and (7) comprehensive psychometric testing. Step 6, preliminary

XSL•FO
RenderX

psychometric testing, was omitted due to the lack of a bilingual population.

Forward Translation and Synthesis

The forward translation was conducted independently by 2 translators, both registered nurses and researchers with expertise in the terminology of the area covered by the SECP. The translators were native Norwegian speakers and fluent in English. The 2 forward-translated versions were additionally compared with the original version of the instrument by a third independent translator, who was bilingual and bicultural. Any discrepancies in wording, sentences, and meanings were addressed and resolved through consensus among the translators and the last author (AAGN).

Backward Translation and Synthesis

The backward translation into English was carried out by 2 independent translators. One back translator was a registered nurse and researcher with expertise in the terminology of the area covered by the SECP, while the other was an English-language expert familiar with nursing terminology. Both back translators were native English speakers and were blinded to the original version of the instrument. The 2 back-translated versions were then compared by a third independent translator who was bilingual and bicultural. Any discrepancies in wording, sentences, and meanings were resolved through consensus among the translators and the last author (AAGN). The prefinal Norwegian version of the SECP was subsequently approved by the instrument developer.

Cognitive Debriefing

The prefinal Norwegian version of the SECP was tested with a pilot group of 10 nursing students who had completed clinical practice and represented the target population [20,21]. Each student was asked to evaluate the instructions, response format, and the 37 items of the SECP using a dichotomous scale (clear or unclear). As no unclear issues were identified by the students, no revisions to the instrument were done [22]. Additionally, an expert panel of 10 members assessed the conceptual equivalence (clarity) of the instrument [21]. The panel comprised experienced nurses who were educators holding positions as assistant professors, associate professors, or professors—all registered nurses familiar with the terminology covered by the SECP. Following the same procedure as the pilot group, the panel identified no issues with conceptual equivalence. The expert panel also evaluated each item for content equivalence (relevance). As no items were rated as irrelevant, difficult to assess, or needing minor alteration, no further revisions were necessary [23]. The instrument was then deemed ready for psychometric testing.

Psychometric Testing of the SECP Norwegian Version

Setting and Sample

The study was conducted at one of the biggest universities in Norway, which offers nursing education at the bachelor level. A convenience sampling method was used to recruit participants for psychometric testing of the SECP (Norwegian version). We aimed to recruit a sample of 400 nursing students to ensure at least 10 responses per item in the SECP, accounting for potential withdrawals from the study [20]. Nursing students from the first and second year of the Bachelor of Nursing Education at the university in the spring of 2022 and 2023 (approximately 800 students) were invited to participate by announcing the study through the university's learning platform. To ensure that all invited students had finished at least one of their clinical practice placements, the announcement of the study and a link to the Questback management system [24], which included written information about the study, web-based informed consent forms, and the SECP (Norwegian version), were distributed to the students through the university's learning platform in the second part of their spring semesters.

Data Collection

Data for the psychometric testing of the SECP (Norwegian version) were collected digitally between the spring of 2022 and the spring of 2023 using the Questback management system, which is a web-based survey system [24]. The web-based version of the SECP included the informed consent form. This combination provided data encompassing voluntary participants' names, email addresses, and SECP scores. No additional background data were collected. To avoid missing data, participants were required to complete all SECP items to finalize the survey.

Statistical Analyses

The data were analyzed using the R programming language [25]. The lavaan package was used to compute the goodness-of-fit indices [26], semTools to calculate internal consistency [27], and semPlot to generate the factor structure model of the SECP (Norwegian version) [28].

Internal Consistency

Internal consistency was assessed by Cronbach α, McDonald ω, and composite reliability coefficients, and values of ≥0.7 were classified as satisfactory [29,30].

Construct Validity

The SECP developers specified a 4-factor model of the SECP, as presented in Table 1 [19]. Discriminant validity was evaluated using the Fornell-Larcker criterion, where good discriminant validity is indicated when the square root of the average variance extracted (AVE) for each factor is greater than the correlations between factors [31].

Confirmatory factor analysis (CFA) using the weighted least squares mean and variance-adjusted estimator [32] was performed to evaluate whether the prehypothesized 4-factor model fit our observed data as evidence of construct validity [33]. The following goodness-of-fit indices were used: the χ^2/df ratio, the *P* value, the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). A χ^2/df ratio of ≤2 was considered acceptable [34]. The *P* value was used to reject a null hypothesis representing a perfect fit [35]. Thus, a nonsignificant *P* value of >.05 was preferred. The acceptable range of the SRMR index was between 0 and 0.08 [36]. Following Hu and Bentler [36], a CFI and TLI of at least 0.95 was deemed acceptable. Lower RMSEA values indicate a better fit [33], and a value of ≤0.05

XSL•FO
RenderX

with a confidence interval ≤0.1 was considered to represent a close model fit [37].

Ethical Considerations

The Norwegian Centre for Research Data approved the study (reference number 891608). Participation in the psychometric testing of the SECP (Norwegian version) was based on written informed consent and performed in accordance with the 2013 revised version of the Declaration of Helsinki [38]. Participation was voluntary, and participants' rights were clearly outlined in the consent forms. Participants were informed about the nature and purpose of the study, as well as their right to withdraw at any time without any consequences. All data were deidentified and anonymized to ensure confidentiality. Participants were not compensated in any way for their involvement in the study.

Results

Mean Scores, Skewness, and Kurtosis

Of the estimated 800 active nursing students invited to participate, 399 (49.9%) completed and returned the instrument. Answers were skewed toward the "fully agree" end of the scale. The respondents' mean scores, skewness, and kurtosis are provided in Table 2.

Table . Mean score and internal consistency indicators by factor (n=399).

Factor or subscale	Mean (SD)	Skewness	Kurtosis	Cronbach α	McDonald ω	Composite reliability
Assessment	8.37 (1.56)	–1.02	2.48	0.96	0.97	0.96
Diagnosis and planning	8.07 (1.84)	–0.96	1.36	0.98	0.98	0.97
Implementation	8.69 (1.52)	–1.42	3.76	0.94	0.95	0.94
Evaluation	8.29 (1.85)	–1.02	1.50	0.96	0.97	0.96

Internal Consistency

The internal consistency indicators by factor level displayed values from 0.94 to 0.98, indicating high internal consistency. The mean score and internal consistency indicators for each factor are presented in Table 2.

Construct Validity

The factor structure model of the SECP Norwegian version is presented in Figure 1. The content of the items is presented in Table 1.

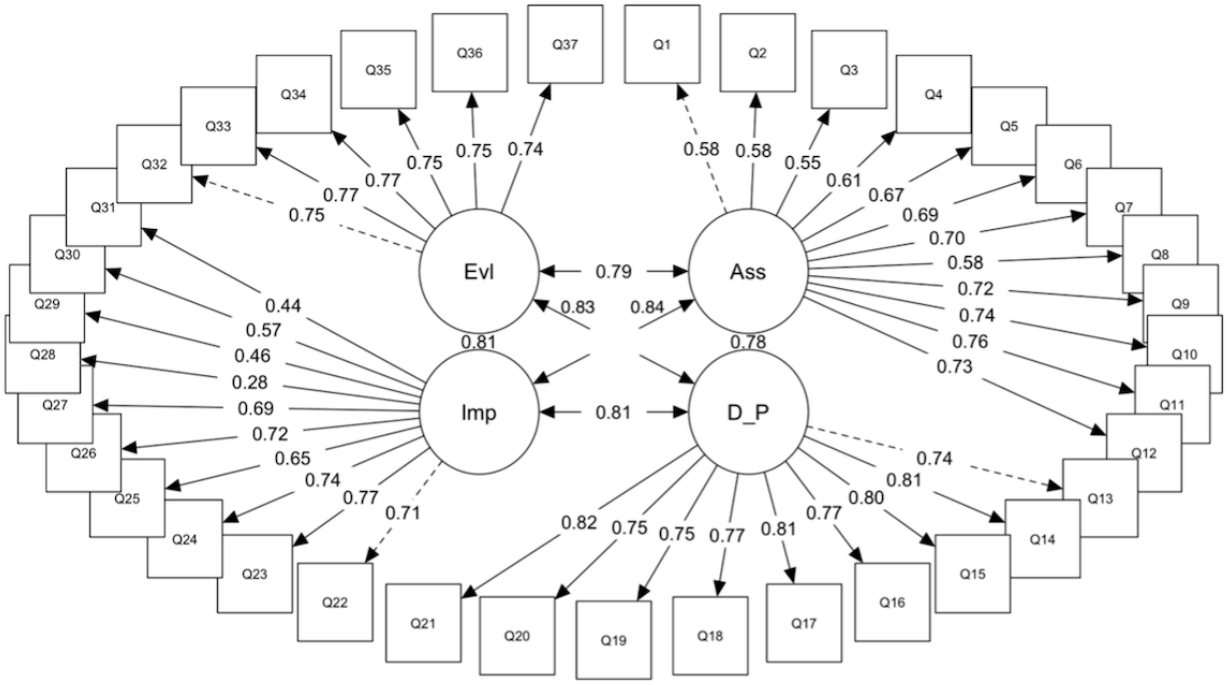
All factor loadings within each factor in the model presented statistical significance and had acceptable values ranging from 0.44 to 0.82, except for item Q28, which presented significant results, but with a low factor loading of 0.28 (Table 3).

The square roots of the AVE values were equal to or lower than the correlations between the factors, indicating potential issues with discriminant validity, as shown in Table 4.

However, overall, the goodness-of-fit indices from the CFA confirmed the prehypothesized factor structure model, indicating acceptable construct validity. The exception was the significant *P* value, which means that the model did not obtain a perfect fit for the data. The goodness-of-fit indices are shown in Table 5.

XSL•FO
RenderX

Figure 1. Factor structure model for the SECP Norwegian version. Ass: assessment; D_P: diagnosis and planning; Evl: evaluation; Imp: implementation; SECP: Self-Efficacy in Clinical Performance Scale.



XSL•FO
RenderX

Table . Factor loadings for the factor structure model (n=399).

Factor or subscale and items		λ[a]	SE	*P* value
Assessment				
	Q1	0.58	0.051	<.001
	Q2	0.58	0.048	<.001
	Q3	0.55	0.055	<.001
	Q4	0.61	0.049	<.001
	Q5	0.67	0.042	<.001
	Q6	0.69	0.040	<.001
	Q7	0.70	0.044	<.001
	Q8	0.58	0.042	<.001
	Q9	0.73	0.041	<.001
	Q10	0.74	0.031	<.001
	Q11	0.76	0.036	<.001
	Q12	0.74	0.037	<.001
Diagnosis and planning				
	Q13	0.74	0.033	<.001
	Q14	0.82	0.026	<.001
	Q15	0.80	0.029	<.001
	Q16	0.77	0.035	<.001
	Q17	0.81	0.028	<.001
	Q18	0.77	0.036	<.001
	Q19	0.76	0.037	<.001
	Q20	0.75	0.039	<.001
	Q21	0.82	0.029	<.001
Implementation				
	Q22	0.71	0.034	<.001
	Q23	0.77	0.030	<.001
	Q24	0.74	0.030	<.001
	Q25	0.66	0.039	<.001
	Q26	0.72	0.032	<.001
	Q27	0.69	0.039	<.001
	Q28	0.28	0.043	<.001
	Q29	0.46	0.043	<.001
	Q30	0.57	0.047	<.001
	Q31	0.44	0.043	<.001
Evaluation				
	Q32	0.75	0.035	<.001
	Q33	0.77	0.031	<.001
	Q34	0.77	0.033	<.001
	Q35	0.75	0.036	<.001
	Q36	0.75	0.035	<.001
	Q37	0.74	0.038	<.001

XSL•FO
RenderX

[a]λ: factor loading.

Table . The square root of AVE[a] and correlations between factors (n=399).

	Assessment	Diagnosis and planning	Implementation	Evaluation
Assessment	0.66[b]	0.78	0.84	0.79
Diagnosis and planning	0.78	0.78[b]	0.81	0.83
Implementation	0.84	0.81	0.63[b]	0.81
Evaluation	0.79	0.83	0.81	0.75[b]

[a]AVE: average variance extracted.

[b]These values represent average variance extracted.

Table . Goodness-of-fit indices (n=399).

Goodness-of-fit indices	Value
χ^2	983.246
df	623
χ^2/df ratio	1.578
P value	<.001
CFI[a]	0.986
TLI[b]	0.985
SRMR[c]	0.056
RMSEA[d] (90% CI)	0.038 (0.034-0.043)

[a]CFI: comparative fit index.

[b]TLI: Tucker-Lewis index.

[c]SRMR: standardized root mean square residual.

[d]RMSEA: root mean squared error of approximation.

Discussion

Principal Results

The Norwegian version of the SECP demonstrated acceptability, internal consistency, and satisfactory construct validity.

Acceptability

To ensure that the conclusions drawn from the statistical analyses were based on genuine cultural differences and similarities rather than translation errors, we placed a strong emphasis on maintaining equivalence between the original and translated versions of the SECP throughout the translation process [39]. As the nursing process is accepted as a care standard with the stages of assessment, nursing diagnosis, planning, implementation, and evaluation worldwide [40], we found that identifying Norwegian words and expressions that captured the original meaning of the SECP was not difficult. The target population of nursing students who piloted and evaluated the SECP (Norwegian version), along with the panel of educators, confirmed the relevance of the wording and items, indicating the acceptability of the SECP within the Norwegian context.

Internal Consistency and Construct Validity

The internal consistency and construct validity tests were performed on a sample of 399 respondents. The suggested minimum size for conducting factor analysis differs in relative terms, from 3 to 20 times the number of variables [41]. Bryant and Yarnold [42] suggest that the subjects-to-variables ratio should be at least 5 times the number of variables. In this study, the subjects-to-variables ratio was above 10:1 and thus considered sufficient.

Cheraghi et al [19] used exploratory factor analysis to test and develop the original SECP compositions, resulting in the prehypothesized 4-factor structure model used in this study. In the original SECP, Cheraghi et al [19] presented Cronbach α scores at factor levels ranging from 0.90 to 0.92, which demonstrated high internal consistency. In this study, the internal consistency indicators for the hypothesized factors also indicated high internal consistency, with factor values ranging from 0.94 to 0.98. However, the high internal consistency values (>0.95), especially in the factor “diagnosis and planning,” can also indicate that some of the factor items in the SECP may be redundant [43]. However, although some of the SECP items are somewhat similarly worded, such as “Q14: Adjust the nursing diagnosis based on an assessment of the patient’s data”

XSL•FO
RenderX

and "Q15: Adjust the nursing diagnosis based on prioritizing the patient's needs," they seem to tap into slightly different aspects of the measured constructs within the factors, thus adding additional information.

Our CFA results revealed significant factor loadings for all 4 hypothesized factors. Ideally, factor loadings should be ≥0.7 [44], and this criterion was met by 24 of the 37 items in the SECP (Norwegian version). Tabachnick and Fidell [45] classify factor loadings of ≥0.55 as good, which applied to 34 of the 37 items. According to their criteria, item Q29 ("Improve my skills based on feedback from a mentor and nursing colleague") with a loading of 0.46 and item Q31 ("Document and report daily clinical work") with a loading of 0.44 may be considered "fair." In the original SECP, Cheraghi et al [19] reported that item Q28 ("Seek help from a mentor or nurse colleagues in difficult situations") demonstrated a factor loading of 0.75. In our study, item Q28 exhibited a factor loading of 0.28, indicating that the factor "Implementation" does not adequately account for the variance in this item [44]. Unlike the other items in "Implementation," item Q28 emphasizes seeking help from others. Consequently, it may have been poorly aligned with our respondents' perceptions of mastering nursing practice within "Implementation," potentially contributing to its low factor loading.

Generally, items with low loadings should be considered for removal as they contribute less to the construct and more to measurement error [44]. In this study, we retained all the original items, which may have contributed to our negative results when assessing discriminant validity using the Fornell-Larcker criterion [44]. We found that our estimated square root of AVE values was equal to or lower than the factor correlations, indicating that the SECP factors are not clearly distinct. Retaining items with low factor loadings will result in lower square roots of AVE, as AVE is calculated based on the squared loadings of items, thereby potentially compromising discriminant validity results [44]. However, several other factors beyond items with low loadings can contribute to negative results on the Fornell-Larcker criterion. These include items with similar content, which can lead to high correlations between factors, or a homogeneous sample, which can result in similar responses across factors and mask the true distinction between them [46].

In this study, we did not redefine the original SECP model, as we could not ensure that changes made would be stable and not influenced by the unique characteristics of our single sample. Furthermore, when assessing the goodness-of-fit indicators from the CFA, the construct validity of the SECP (Norwegian version) was considered acceptable. Our χ^2/df ratio was well below the recommended limit set by Byrne [34], indicating a satisfactory model fit [47]. For further evaluation, we assessed the CFI, TLI, SRMR, and RMSEA indices. The RMSEA and CFI are relatively robust in large samples [36], and the RMSEA is adjusted for model complexity [33]. Our RMSEA value indicated a close fit [36,37]. Additionally, the SRMR fell within the acceptable range established by Hu and Bentler [36], while both the CFI and the TLI exceeded the cutoff point of 0.95 suggested by Hu and Bentler [36], indicating a good fit. One goodness-of-fit indicator that opposed the hypothesized model was the chi-square *P* value, which was <.001. However, significant *P* values may arise, especially in large samples, even when the proposed model is only slightly inaccurate [47,48]. Furthermore, by using the χ^2/df, CFI, TLI, SRMR, and RMSEA fit indices, we evaluated various aspects of goodness of fit. Together, these fit indices confirmed the hypothesized SECP model, indicating that the overall model structure is sound.

An important step in improving nursing students' clinical performance is to pinpoint both students' strengths and areas requiring improvement. In this first Norwegian translation and testing of the SECP, the results show that the SECP (Norwegian version) has the potential as an instrument within Norwegian nursing education to assess students' self-efficacy in performing different facets of the nursing process. The SECP could be integrated into Norwegian nursing education for clinical course evaluations, and SECP results may be used to guide nurse educators in their work to better equip nursing students for the demands of the clinical setting. The SECP not only tracks nursing students' self-efficacy but also helps to identify learning gaps due to low self-efficacy, measure the impact of teaching interventions, guide curriculum development, and moreover tailor students' learning and subsequently reinforce students' well-being and self-efficacy in clinical performance. The SECP can be administered throughout the study program, for example, before clinical placements, midway through courses, or after targeted workshops. Aggregated results may reveal recurring weaknesses, the need for curriculum updates, and spotlight at-risk students in need of guidance, ultimately improving readiness, resilience, and overall performance in the nursing profession [49,50].

Limitations

The originally validated SECP is in Persian, and in this study, we translated the English version provided by the original SECP developers, Cheraghi et al [19], for which there was no prior validation. We found no other studies that had translated, adapted, or validated the SECP into other languages. Thus, we could not compare this study with studies from different linguistic and cultural contexts, apart from the original study by Cheragi et al [19]. This emphasizes the need for additional research to establish the SECP scale's validity and reliability across diverse populations worldwide.

Additionally, individual characteristics of our participants, such as age or gender, were not collected, which could have provided a more comprehensive overview of the sample. Although the CFA results in this study confirmed the prehypothesized SECP model, it should be noted that having an acceptable fitting model with data from a single sample does not necessarily confirm that the model is correctly specified [47]. Furthermore, we did not evaluate the stability of the SECP (Norwegian version) over time. Therefore, we propose that future psychometric validation studies also include evaluations of the instrument's test-retest reliability.

Conclusion

The SECP (Norwegian version) demonstrated its potential as a viable instrument for assessing self-efficacy in the current and

XSL•FO
RenderX

required competencies of nursing students in clinical practice within Norwegian nursing education. Collecting data in these areas may be crucial to evaluate, develop, and enhance nursing students' clinical performance. The SECP (Norwegian version) indicated internal consistency and acceptable construct validity in this study; however, future research should aim to confirm the factor structure of the SECP (Norwegian version) and evaluate its test-retest reliability.

Acknowledgments

We wish to thank Dr Fatemeh Cheraghi for permission to translate, validate, and use the Self-Efficacy in Clinical Performance (SECP) instrument. We also thank the translators, faculty members, pilot group, expert panel, and nursing students who contributed to the translation and testing of the SECP (Norwegian version).

Authors' Contributions

Conceptualization: JZ, AAGN
Data curation: CO
Formal analysis: CO
Investigation: MS, JZ, AAGN
Project administration: CO
Visualization: CO, AAGN
Writing – original draft: CO, JZ, AAGN
Writing – review & editing: MS, JZ, TNF, AAGN

Conflicts of Interest

None declared.

References

1. Benner P, Wrubel J. Skilled clinical knowledge: the value of perceptual awareness, part 2. J Nurs Adm 1982 Jun;12(6):28-33. [Medline: 6919573]
2. Toney-Butler TJ, Thayer JM. Nursing process. In: StatPearls [Internet]: StatPearls Publishing; 2023. URL: https://www.ncbi.nlm.nih.gov/books/NBK499937 [accessed 2025-04-02] [Medline: 29763112]
3. McCuistion LE, DiMaggio K, Winton MB, Yeager JJ. Pharmacology: A Patient-Centered Nursing Process Approach, 10th edition: Saunders; 2020.
4. Dinmohammadi M, Jalali A, Peyrovi H. Clinical learning experiences of Iranian student nurses: a qualitative study. Nurs Pract Today 2016;3(1):31-39 [FREE Full text]
5. Günay U, Kılınç G. The transfer of theoretical knowledge to clinical practice by nursing students and the difficulties they experience: a qualitative study. Nurs Educ Today 2018 Jun;65:81-86. [doi: 10.1016/j.nedt.2018.02.031]
6. Taskın Yilmaz F, Sabanciogullari S, Aldemir K. The opinions of nursing students regarding the nursing process and their levels of proficiency in Turkey. J Caring Sci 2015 Dec;4(4):265-275. [doi: 10.15171/jcs.2015.027] [Medline: 26744726]
7. Xiang D, Ge S, Zhang Z, Twumwaah Budu J, Mei Y. Relationship among clinical practice environment, creative self-efficacy, achievement motivation, and innovative behavior in nursing students: a cross-sectional study. Nurs Educ Today 2023 Jan;120:105656. [doi: 10.1016/j.nedt.2022.105656]
8. Yu M, Tong H, Li S, Wu XV, Hong J, Wang W. Clinical competence and its association with self-efficacy and clinical learning environments among Chinese undergraduate nursing students. Nurs Educ Pract 2021 May;53:103055. [doi: 10.1016/j.nepr.2021.103055]
9. Alosaimi D. Learning self-efficacy as predictor of nursing students' performance of clinical skills. Educ Sci Theor Pract 2021 Dec;21(3):120-131. [doi: 10.12738/jestp.2021.3.009] [Medline: 2637179889]
10. Bandura A. Self-Efficacy: The Exercise of Control: W. H. Freeman & Co Ltd; 1997.
11. Kang YN, Chang CH, Kao CC, Chen CY, Wu CC. Development of a short and universal learning self-efficacy scale for clinical skills. PLoS One 2019;14(1):e0209155. [doi: 10.1371/journal.pone.0209155] [Medline: 30615610]
12. Henderson A, Harrison P, Rowe J, et al. Students take the lead for learning in practice: a process for building self-efficacy into undergraduate nursing education. Nurs Educ Pract 2018 Jul;31:14-19. [doi: 10.1016/j.nepr.2018.04.003] [Medline: 29730303]
13. Arkan B, Ordin Y, Yılmaz D. Undergraduate nursing students' experience related to their clinical learning environment and factors affecting to their clinical learning process. Nurs Educ Pract 2018 Mar;29:127-132. [doi: 10.1016/j.nepr.2017.12.005] [Medline: 29274524]
14. Keeping-Burke L, McCloskey R, Donovan C, Yetman L, Goudreau A. Nursing students' experiences with clinical placement in residential aged care facilities: a systematic review of qualitative evidence. JBI Evid Synth 2020 May;18(5):986-1018. [doi: 10.11124/JBISRIR-D-19-00122] [Medline: 32813353]

XSL•FO
RenderX

15. Perry C, Henderson A, Grealish L. The behaviours of nurses that increase student accountability for learning in clinical practice: an integrative review. Nurs Educ Today 2018 Jun;65:177-186. [doi: 10.1016/j.nedt.2018.02.029]
16. Chesser-Smyth PA, Long T. Understanding the influences on self-confidence among first-year undergraduate nursing students in Ireland. J Adv Nurs 2013 Jan;69(1):145-157. [doi: 10.1111/j.1365-2648.2012.06001.x] [Medline: 22486537]
17. Yiin SJ, Shen KP, Lai CL, Liang JC. An evaluation of nursing students' learning self-efficacy: a multi-dimensional instrument development and structural validation. Nurs Educ Today 2024 Apr;135:106118. [doi: 10.1016/j.nedt.2024.106118]
18. Bulfone G, Iovino P, Mazzotta R, et al. Self-efficacy, burnout and academic success in nursing students: a counterfactual mediation analysis. J Adv Nurs 2022 Oct;78(10):3217-3224. [doi: 10.1111/jan.15231] [Medline: 35307859]
19. Cheraghi F, Hassani P, Yaghmaei F, Alavi-Majed H. Developing a valid and reliable self-efficacy in clinical performance scale. Int Nurs Rev 2009 Jun;56(2):214-221. [doi: 10.1111/j.1466-7657.2008.00685.x] [Medline: 19646171]
20. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. J Eval Clin Pract 2011 Apr;17(2):268-274. [doi: 10.1111/j.1365-2753.2010.01434.x] [Medline: 20874835]
21. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000 Dec 15;25(24):3186-3191. [doi: 10.1097/00007632-200012150-00014] [Medline: 11124735]
22. Topf M. Three estimates of interrater reliability for nominal data. Nurs Res 1986;35(4):253-255. [doi: 10.1097/00006199-198607000-00020] [Medline: 3636827]
23. Waltz C, Strickland O, Lenz E. Measurement in Nursing and Health Research, 3rd edition: Springer Publishing Company; 2005.
24. Questback managing software. Questback. 2022. URL: http://www.questback.com [accessed 2025-04-02]
25. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2021. URL: https://www.R-project.org [accessed 2025-04-02]
26. Rosseel Y. lavaan: an R package for structural equation modeling. J Stat Softw 2012;48(2):1-36. [doi: 10.18637/jss.v048.i02]
27. Jorgensen TD, Pornprasertmanit S, Schoemann AM, et al. semTools: useful tools for structural equation modeling [R package version 05-1]. R Foundation for Statistical Computing. 2018. URL: https://CRAN.R-project.org/package=semTools [accessed 2025-04-02]
28. Epskamp S. semPlot: path diagrams and visual analysis of various SEM packages. R Foundation for Statistical Computing. 2019. URL: https://CRAN.R-project.org/package=semPlot [accessed 2025-04-02]
29. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. J Pers Assess 2003 Feb;80(1):99-103. [doi: 10.1207/S15327752JPA8001_18] [Medline: 12584072]
30. Hayes AF, Coutts JJ. Use omega rather than Cronbach's alpha for estimating reliability. But…. Commun Methods Meas 2020 Jan 2;14(1):1-24. [doi: 10.1080/19312458.2020.1718629]
31. Ab Hamid MR, Sami W, Mohmad Sidek MH. Discriminant validity assessment: use of Fornell & Larcker criterion versus HTMT criterion. J Phys Conf Ser 2017 Sep 1;890:012163. [doi: 10.1088/1742-6596/890/1/012163]
32. Li CH. Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. Behav Res Methods 2016 Sep;48(3):936-949. [doi: 10.3758/s13428-015-0619-7] [Medline: 26174714]
33. Sun J. Assessing goodness of fit in confirmatory factor analysis. Meas Eval Couns Dev 2005 Jan;37(4):240-256. [doi: 10.1080/07481756.2005.11909764]
34. Byrne BM. A Primer of LISREL: Basic Applications and Programming for Confirmatory Factor Analysis Models: Springer-Verlag; 1989. [doi: 10.1007/978-1-4613-8885-2]
35. Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, editors. Testing Structural Equation Models: Sage; 1993:136-162.
36. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
37. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. Methods Psychol Res 2003;8(2):23-74. [doi: 10.23668/psycharchives.12784]
38. WMA declaration of Helsinki: ethical principles for medical research involving human subjects. World Medical Association. 2013. URL: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ [accessed 2025-04-02]
39. Wang WL, Lee HL, Fetzer SJ. Challenges and strategies of instrument translation. West J Nurs Res 2006 Apr;28(3):310-321. [doi: 10.1177/0193945905284712] [Medline: 16585807]
40. Potter PA, Perry AG, Stockert PA, Hall A. Fundamental of Nursing, 11th edition: Elsevier; 2023.
41. Mundfrom DJ, Shaw DG, Ke TL. Minimum sample size recommendations for conducting factor analyses. Int J Test 2005 Jun;5(2):159-168. [doi: 10.1207/s15327574ijt0502_4]
42. Bryant FB, Yarnold PR. Principal-components analysis and exploratory and confirmatory factor analysis. In: Grimm LG, Yarnold PR, editors. Reading and Understanding Multivariate Statistics: American Psychological Association; 1995:99-136.
43. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jun 27;2(53-5):53-55. [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]

XSL•FO
RenderX

44. Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate Data Analysis, 7th edition: Pearson Prentice Hall; 2010.
45. Tabachnick BG, Fidell LS. Using Multivariate Statistics, 6th edition: Allyn & Bacon; 2013.
46. Kline RB. Principles and Practice of Structural Equation Modeling, 4th edition: Guilford Press; 2015.
47. Alavi M, Visentin DC, Thapa DK, Hunt GE, Watson R, Cleary M. Chi-square for model fit in confirmatory factor analysis. J Adv Nurs 2020 Sep;76(9):2209-2211. [doi: 10.1111/jan.14399] [Medline: 32323338]
48. MacCallum RC. 2001 Presidential Address: working with imperfect models. Multivariate Behav Res 2003 Jan;38(1):113-139. [doi: 10.1207/S15327906MBR3801_5]
49. Zengin N, Pınar R, Akinci AC, Yildiz H. Psychometric properties of the self-efficacy for clinical evaluation scale in Turkish nursing students. J Clin Nurs 2014 Apr;23(7-8):976-984. [doi: 10.1111/jocn.12257] [Medline: 23876212]
50. George TP, DeCristofaro C, Murphy PF. Self-efficacy and concerns of nursing students regarding clinical experiences. Nurs Educ Today 2020 Jul;90:104401. [doi: 10.1016/j.nedt.2020.104401]

Abbreviations

AVE: average variance extracted
CFA: confirmatory factor analysis
CFI: comparative fit index
RMSEA: root mean square error of approximation
SECP: Self-Efficacy in Clinical Performance Scale
SRMR: standardized root mean square residual
TLI: Tucker-Lewis index

Edited by A Mavragani; submitted 30.10.24; peer-reviewed by BA Balgiu; revised version received 11.03.25; accepted 11.03.25; published 21.04.25.

Please cite as:
Olaussen C, Stojiljkovic M, Zlamal J, Fløtlo TN, Nes AAG
The Norwegian Version of the Self-Efficacy in Clinical Performance Scale (SECP): Psychometric Validation Study
JMIR Form Res 2025;9:e68173
URL: https://formative.jmir.org/2025/1/e68173
doi:10.2196/68173

© Camilla Olaussen, Marko Stojiljkovic, Jaroslav Zlamal, Tone Nygaard Fløtlo, Andrea Aparecida Gonçalves Nes. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

CoVimmune COVID-19 Immunity Calculator: Web Application Development and Validation Study

Rebecca Slotkin[1], MD; Tassos C Kyriakides[2,3], PhD; Vinni Yu[4], BS; Xien Chen[4], BS; Anupam Kundu[5,6], PhD; Shaili Gupta[1,6], MBBS

[1]Department of Medicine, Yale School of Medicine, 950 Campbell Ave, Bldg 1, Floor 5, West Haven, CT, United States

[2]Cooperative Studies Program Coordinating Center, VA CT Healthcare System, West Haven, CT, United States

[3]Yale School of Public Health, Yale University, New Haven, CT, United States

[4]Department of Computer Science, Yale University, New Haven, CT, United States

[5]Yale Center for Analytical Sciences, Yale School of Public Health, New Haven, CT, United States

[6]Veterans Healthcare System of Connecticut, West Haven, CT, United States

Corresponding Author:
Shaili Gupta, MBBS
Department of Medicine, Yale School of Medicine, 950 Campbell Ave, Bldg 1, Floor 5, West Haven, CT, United States

Abstract

Background: This study illustrates the development of a simple web-based application, which demonstrates the relationship between serum anti-SARS-CoV-2 S1/receptor-binding domain immunoglobulin G (IgG) and anti-SARS-CoV-2 neutralizing antibody (nAb) half-maximal inhibitory concentration (IC50) titers in a vaccinated US adult population and compares them to prior data on nAb titers at different time points after vaccination.

Objective: The objective of this study is to create an easily accessible calculator that uses the results of commercially available anti-SARS-CoV-2 serum IgG to approximate the underlying ability to neutralize SARS-CoV-2.

Methods: Our web-based application leveraged two previously published datasets. One dataset demonstrated a robust correlation between nAb and serum IgG. The other dataset measured nAb titers at specific time periods over a year-long interval following a messenger RNA vaccination primary series and booster vaccine dose. Clinical factors that were statistically significant on a forward linear regression model examining the prediction of nAb from serum IgG were incorporated in the application tool.

Results: By combining the datasets described above, we developed a publicly available web-based application that allows users to enter a serum IgG value and determine their estimated nAb titer. The application contextualizes the estimated nAb titer with the theoretical distance from the corresponding vaccine-mediated antibody protection. Using the clinical variables that had a significant impact on how well IgG values predict nAb titers, this application allows for a patient-centered, nAb titer prediction.

Conclusions: This application offers an example of how we might bring the advances made in scientific research on protective antibodies post-SARS-CoV-2 vaccination into the clinical sphere with practical tools.

(JMIR Form Res 2025;9:e59467) doi:10.2196/59467

KEYWORDS

COVID-19; immunity; neutralizing antibody; immunoglobulin G; vaccine hesitancy; vaccine timing; patient-centered care; web application; vaccination; SARS-CoV-2

Introduction

High levels of anti-SARS-CoV-2 neutralizing antibodies (nAbs) have been shown to be protective against severe SARS-CoV-2 infection [1], but nAbs wane over time following vaccination or infection [2,3]. In addition, nAbs require significant funding and infrastructure to measure and are largely limited to research settings. Serum anti-SARS-CoV-2 immunoglobulin G (IgG) testing is available in commercial settings at relatively low cost. IgG testing has been sporadically ordered in primary care settings; although, there has been little to guide its clinical interpretation [4,5].

Methods

Ethical Considerations

Approval for the study from which both datasets were derived (SARS-CoV-2 Vaccine and You, protocol #1599488) was obtained from the Veterans Affairs Connecticut Healthcare System Institutional Review Board (IRB) in December 2020. Participants were eligible for inclusion if they were veterans or health care workers at the Veterans Affairs Connecticut Healthcare System and excluded if they were unable to provide informed consent. Written informed consent was obtained from

XSL•FO
RenderX

willing participants prior to study participation and they were aware they could opt out at any time. The ethics approval covered secondary analysis as done for this paper without additional consent. Samples were donated by participants without any compensation and we are immensely grateful for this. All identifiable subject data was secured in encrypted folders accessible only to IRB-approved investigators. All data was de-identified before serological analysis.

Source Data Sets

The first dataset was derived from a prospective cohort study which measured anti-SARS-COV-2 nAb half-maximal inhibitory concentration (IC50) titers against pseudotyped WA-1 (the original SARS-CoV-2 strain) at specific time periods over a year-long interval following the primary messenger RNA vaccination series and a subsequent booster vaccine dose (before dose 1, before dose 2, and after dose 2 at 1, 3, 6, and 12 months, with an additional collection 1 month after dose 3, which was approximately 10 months after dose 2). Clinical variables were collected by manual chart review as previously described. The nAb titers were found to fluctuate over time, peaking 1 month after the primary series, waning at 6 months, and rising again 1 month after the booster dose [2].

The second dataset leveraged the collected serum samples and measured IgG anti-S1/receptor-binding domain values in sera with predetermined nAb titers and clinical variables as previously described [6]. IgG and nAb IC50 titers had a robust linear correlation. Multimedia Appendix 1 illustrates this correlation for each clinical variable (COVID-19 infection status, estimated glomerular filtration rate (GFR), age, sex, history of malignancy, and race). Based on a forward stepwise linear regression, GFR (0 - 30, 31 - 59, and >60 mL/min/1.73 m^2), and history of COVID-19 infection within the past 12 months (yes or no) were found to impact this correlation (R^2=0.78) [6].

The model had undergone internal validation using repeated k-fold cross-validation. The model's performance was evaluated using root mean squared error (RMSE), R^2, and mean absolute error (MAE) across multiple versions of the model and different crossvalidation setups (ten 10-fold repeats, five 10-fold repeats, and three 5-fold repeats). For our model, which included IgG, COVID-19 infection status, and GFR, the RMSE was relatively low (around 0.27 - 0.28) across validation setups, suggesting a small average error between the predicted and actual values. The R^2 values were around 0.77 to 0.81 across validation setups, which indicated that the model explained about 77% to 81% of the variability in the data. The MAE values were around 0.2 for all validation setups, demonstrating a low average error magnitude (Multimedia Appendix 2).

Application Development

Backend (Software)

The application used the model-view-controller pattern. We initially used RStudio 2023.12.1+402 running R 4.3.2 (IBM Corp) and an integrated development environment with a set of tools to complement R. The original R script had initially used its own web application framework (Shiny) to create a local web-based application, but this framework was not compatible with the rest of the tools used in the development process. The R script was converted to Python 3.9.12 (Python Software Foundation) for ease of web development. The main challenge when developing the project was converting the original R script into Python and creating the website using an HTML and cascading style sheets (CSS). Python was chosen over R for several reasons. First, Python's versatility and ecosystem made integration with other tools like Flask, a popular web framework that simplifies the web application development process, and Docker, which simplifies the process of containerizing applications, much easier. Second, Python's versatility and rich ecosystem of libraries and web development tools allowed us to keep the backend and front end the same language. Third, Python could handle a high volume of web requests and was better suited for scalable applications, which was a point of consideration. Last, Python was also the better choice for long-term maintainability, as it is more readable and easier to update. The host server emphasized Docker as the primary tool for deployment, making it the easiest tool to use for the development process. The full source code is available on request.

Front End (Web Interface)

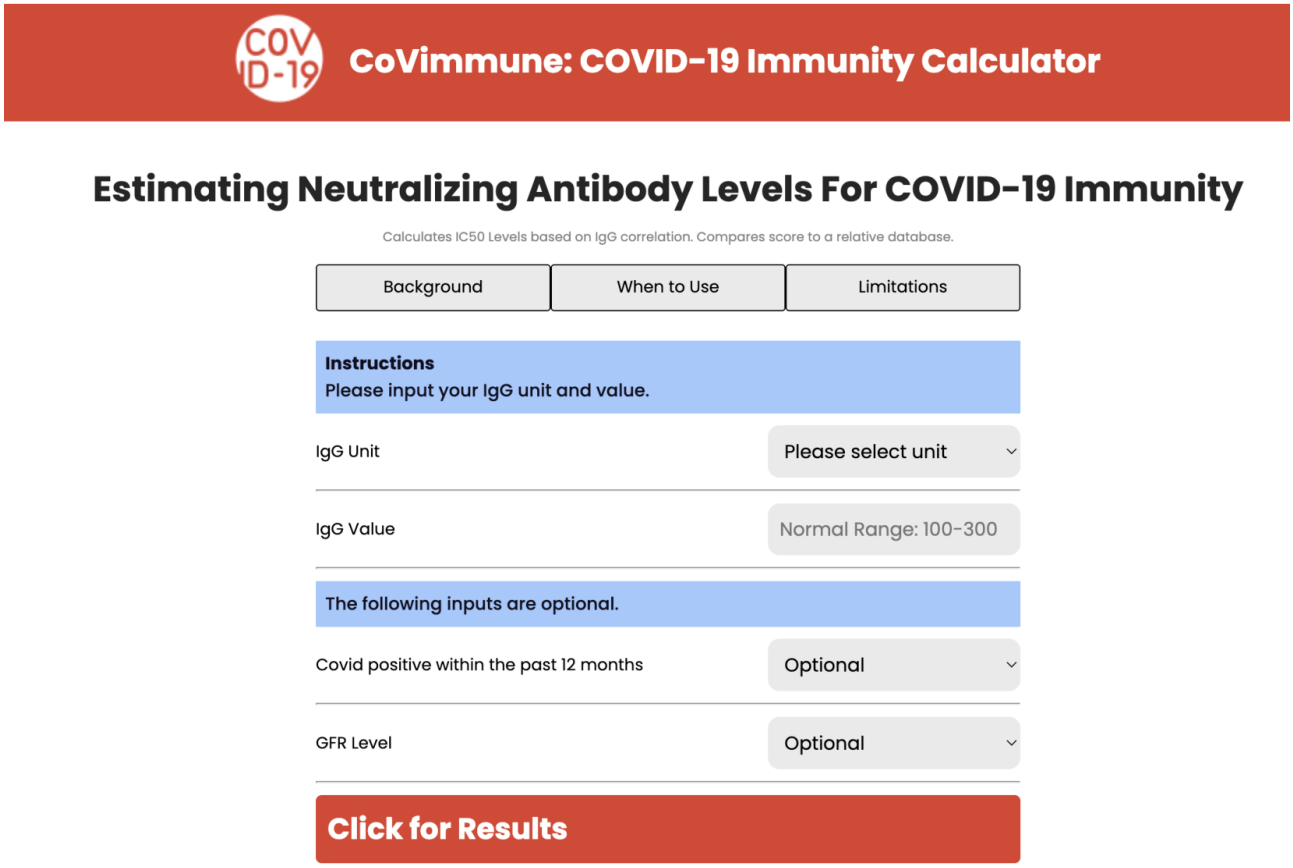
The front end was converted from a Shiny package in RStudio to an HTML and CSS. The Python Flask 3.0.0 (Armin Ronacher) module was used to build the CoVimmune website. The website was launched using a YaleInformation Technology Support secured self-service sandbox environment called Spinup. The Waitress Web Server Gateway Interface was used to deploy our application and Docker 4.27.1 (Docker, Inc) was used to package, ship, and run our application.

Results

This web-based application was based on 127 sera samples in 100 unique participants aged 20 - 93 years (mean 63.83, SD 15.63 years; 29% female; 67% White) from the second dataset. The user can input their IgG value, and the units can be from any of 4 commonly used platforms (Abbott-arbitrary U/ml, Siemens-U/ml, Roche-U/ml, and Euroimmune-relative U/ml) or the World Health Organization's standardized IgG value in binding antibody U/mL. The user can choose to select their COVID-19 infection status and GFR category from drop down lists. The application then generates an estimated nAb titer, which is compared to individuals with matching renal function and COVID-19 status from the first dataset, to provide a contextualized range of nAb titers at different time points after vaccination (Figure 1 and Figure 2). The application also visualizes the results in the context of time from the last vaccination among individuals with similar values in the application dataset. The application is free and available on the internet [7].

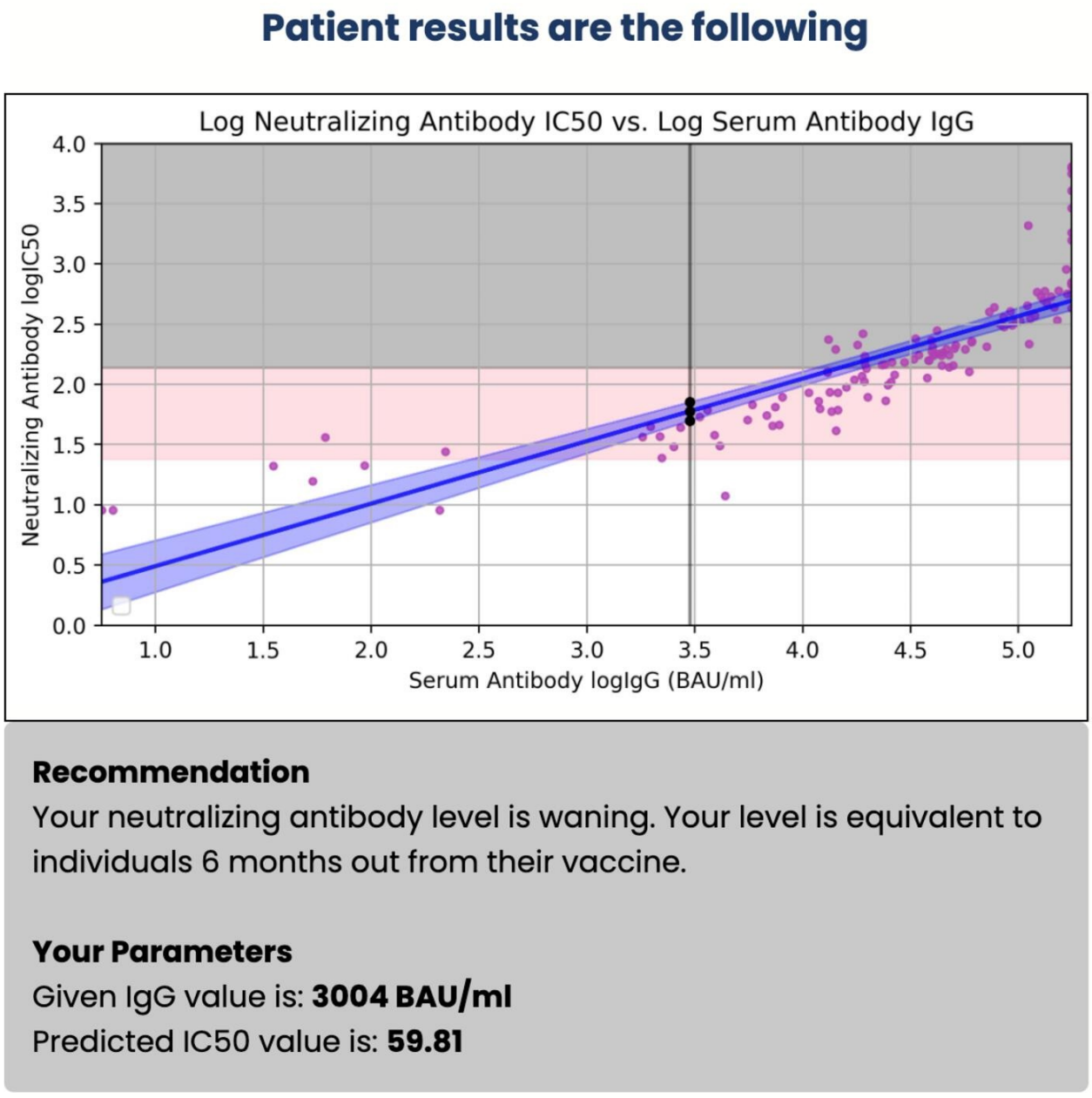
XSL•FO
RenderX

Figure 1. CoVimmune COVID-19 immunity calculator website interface for the application. GFR: glomerular filtration rate; IC50: half-maximal inhibitory concentration; IgG: immunoglobulin G.



XSL•FO
RenderX

Figure 2. CoVimmune COVID-19 immunity calculator results with an example input. BAU: binding antibody unit; IC50: half-maximal inhibitory concentration; IgG: immunoglobulin G.



Discussion

The application illustrates the relationship between serum anti-SARS-CoV-2 IgG and anti-SARS-CoV-2 nAb IC50 titers in a vaccinated US adult population and allows for the adjustment of that relationship based on renal function and prior COVID-19 infection. It brings bench research into the clinical context and may provide a tool for physicians to personalize discussions about vaccine booster timing for patients where yearly vaccination may not be appropriate [8] or for those who are hesitant to get revaccinated.

This application is not intended to replace clinical judgment or offer conclusions about immune protection from COVID-19 or vaccine efficacy. The nAb titers are only one component of the immune response to infection. We cannot extrapolate cellular immunity from these data. This application is based on a small sample of patients and there were limited numbers of immunocompromised individuals within the sample, which may have limited generalizability to those individuals [9]. Ultimately, this application is a user-friendly example of how SARS-CoV-2 IgG nAb vaccine research, which has high barriers to replication and clinical application due to cost and infrastructure requirements, can be transformed into a practical tool.

XSL•FO
RenderX

Acknowledgments

We would like to thank Dr Gary Stack and Dr Richard Sutton for their contributions to the studies that form the basis of this application. A big thank you to Yale students VY and XC for their work on website creation. Funding for this study was provided by the National Institutes of Health (5T32AR007107-48 to RS) and (NCT04501978 to SG). The funders of the study had no role in the study design, data collection, data analysis, writing of the report, or in the decision to submit results for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Correlation between the log immunoglobulin G values and the log half-maximal inhibitory concentration titers by clinical variable.
[PNG File, 828 KB - formative_v9i1e59467_app1.png]

Multimedia Appendix 2
Repeated k-fold cross validation results.
[PNG File, 29 KB - formative_v9i1e59467_app2.png]

References

1. Khoury DS, Cromer D, Reynaldi A, et al. Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. Nat Med 2021 Jul;27(7):1205-1211. [doi: 10.1038/s41591-021-01377-8] [Medline: 34002089]
2. Zhao M, Slotkin R, Sheth AH, et al. Serum neutralizing antibody titers 12 months after coronavirus disease 2019 messenger RNA vaccination: correlation to clinical variables in an adult, US population. Clin Infect Dis 2023 Feb 8;76(3):e391-e399. [doi: 10.1093/cid/ciac416] [Medline: 35639598]
3. Link-Gelles R, Weber ZA, Reese SE, et al. Estimates of bivalent mRNA vaccine durability in preventing COVID-19-associated hospitalization and critical illness among adults with and without immunocompromising conditions - VISION Network, September 2022-April 2023. MMWR Morb Mortal Wkly Rep 2023 May 26;72(21):579-588. [doi: 10.15585/mmwr.mm7221a3] [Medline: 37227984]
4. Wiencek JR, Head CL, Sifri CD, Parsons AS. Clinical ordering practices of the SARS-CoV-2 antibody test at a large academic medical center. Open Forum Infect Dis 2020 Oct;7(10):faa406. [doi: 10.1093/ofid/ofaa406] [Medline: 33072813]
5. Zarif A, McGagh D, Hoang U, de Lusignan S. The role of antibody tests for COVID-19 in primary care. Br J Gen Pract 2021;71(704):131-134. [doi: 10.3399/bjgp21X715169] [Medline: 33632693]
6. Slotkin R, Kyriakides TC, Kundu A, Stack G, Sutton RE, Gupta S. Correlation of a commercial platform's results with post-vaccination SARS-CoV-2 neutralizing antibody response and clinical host factors. PLoS ONE 2023;18(8):e0289713. [doi: 10.1371/journal.pone.0289713] [Medline: 37643190]
7. Estimating neutralizing antibody levels for COVID-19 immunity. CoVimmune: COVID-19 Immunity Calculator. URL: https://covimmune.yale.edu/ [accessed 2024-04-02]
8. Townsend JP, Hassler HB, Dornburg A. Infection by SARS-CoV-2 with alternate frequencies of mRNA vaccine boosting. J Med Virol 2023 Feb;95(2):e28461. [doi: 10.1002/jmv.28461] [Medline: 36602045]
9. Barnes E, Goodyear CS, Willicombe M, et al. SARS-CoV-2-specific immune responses and clinical outcomes after COVID-19 vaccination in patients with immune-suppressive disease. Nat Med 2023 Jul;29(7):1760-1774. [doi: 10.1038/s41591-023-02414-4] [Medline: 37414897]

Abbreviations

CSS: cascading style sheets
GFR: glomerular filtration rate
IC50: half-maximal inhibitory concentration
IgG: immunoglobulin G
MAE: mean absolute error
nAb: neutralizing antibody
RMSE: root mean squared error

XSL•FO
RenderX

Edited by A Mavragani; submitted 21.04.24; peer-reviewed by A Mohazzab, P Bare; revised version received 03.11.24; accepted 05.11.24; published 22.04.25.

Please cite as:
Slotkin R, Kyriakides TC, Yu V, Chen X, Kundu A, Gupta S
CoVimmune COVID-19 Immunity Calculator: Web Application Development and Validation Study
JMIR Form Res 2025;9:e59467
URL: https://formative.jmir.org/2025/1/e59467
doi:10.2196/59467

© Rebecca Slotkin, Tassos C Kyriakides, Vinni Yu, Xien Chen, Anupam Kundu, Shaili Gupta. Originally published in JMIR Formative Research (https://formative.jmir.org), 22.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Increasing the Uptake of Breast and Cervical Cancer Screening Via the MAwar Application: Stakeholder-Driven Web Application Development Study

Nurfarhana Nasrudin[1*], BEng, MEng; Shariff-Ghazali Sazlina[1*], MBBS, MMed, PhD; Ai Theng Cheong[1*], MBBS, MMed, PhD; Ping Yein Lee[2*], MBBS, MMed; Soo-Hwang Teo[3], MA, PhD; Abdul Rashid Aneesa[1], MBBCh BAO, MMed; Chin Hai Teo[2,4], BSc, PhD; Fakhrul Zaman Rokhani[5*], BEng, MSc, PhD; Nuzul Azam Haron[6*], BQS, MSc, PhD; Noor Harzana Harrun[7*], MD, MMed; Bee Kiau Ho[8*], MBBS, MMed; Salbiah Mohamed Isa[8*], MD, MMed

[1]Department of Family Medicine, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Jalan Universiti 1, Serdang, Malaysia

[2]UM eHealth Unit, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia

[3]Cancer Research Malaysia, Subang Jaya, Malaysia

[4]Department of Primary Care Medicine, Faculty of Medicine, Universiti Malaya, Kuala Lumpur, Malaysia

[5]Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, Serdang, Malaysia

[6]Department of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia, Serdang, Malaysia

[7]Klinik Kesihatan Pandamaran, Ministry of Health Malaysia, Klang, Malaysia

[8]Klinik Kesihatan Bandar Botanic, Ministry of Health Malaysia, Klang, Malaysia

[*]these authors contributed equally

Corresponding Author:
Shariff-Ghazali Sazlina, MBBS, MMed, PhD
Department of Family Medicine, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Jalan Universiti 1, Serdang, Malaysia

Abstract

Background: Digital health interventions such as web health applications significantly enhance screening accessibility and uptake, particularly for individuals with low literacy and income levels. By involving stakeholders—including health care professionals, patients, and technical experts—an intervention can be tailored to effectively meet the users' needs, ensuring contextual relevance for better acceptance and impact.

Objective: The aim of this study is to prioritize the content and user interface appropriate for developing a web health application, known as the MAwar app, to promote breast and cervical cancer screening.

Methods: A cross-sectional study for stakeholder engagement was conducted to develop a web-based application known as the MAwar app as part of a larger study entitled "The Effectiveness of an Interactive Web Application to Motivate and Raise Awareness on Early Detection of Breast and Cervical Cancers (The MAwar study)". The stakeholder engagement process was conducted in a public health district that oversees 12 public primary care clinics with existing cervical and breast cancer screening programs. We purposively selected the stakeholders for their relevant roles in breast and cervical cancer screening (health care staff, patients, and public representatives), as well as expertise in software and user interface design (technology experts). The Quality Function Deployment method was used to reflect the priorities of diverse stakeholders (health care, technology experts, patients, and public representatives) in its design. The Quality Function Deployment method facilitated the translation of stakeholder perspectives into app features. Stakeholders rated features on a scale from 1 (least important) to 5 (most important), ensuring the app's design resonated with user needs. The correlations between the "WHATs" (user requirements) and the "HOWs" (technical requirements) were scored using a 3-point ordinal scale, with 1 indicating weak correlation, 5 indicating medium correlation, and 9 indicating the strongest correlation.

Results: A total of 13 stakeholders participated in the study, including women who had either underwent or never had health screening, a health administrator, a primary care physician, medical officers, nurses, and software designers. Stakeholder evaluations highlighted cost-free access (mean 4.64, SD 0.81), comprehensive cancer information (mean 4.55, SD 0.69), detailed screening benefits (mean 4.45, SD 0.68), detailed screening facilities (mean 4.45, SD 0.68) and personalized risk calculator for breast and cervical cancers (mean 4.45, SD 0.68) as essential priorities of the app. The highest-ranked features include detailed information on screening procedures (weighted score [WS]=367.84), information on treatment options (WS=345.80), benefits of screening (WS=333.75), information about breast and cervical cancers (WS=332.15), and frequently asked questions about the concerns around screening (WS=312.00).

XSL•FO
RenderX

Conclusions: The MAwar app, conceived through a collaborative, stakeholder-driven process, represents a significant step in leveraging digital health solutions to tackle cancer screening disparities. By prioritizing accessibility, information quality, and clarity on benefits, the app promises to encourage early cancer detection and management for targeted communities.

Trial Registration: ISRCTN Registry ISRCTN10403163; https://www.isrctn.com/ISRCTN10403163

(JMIR Form Res 2025;9:e65542) doi:10.2196/65542

KEYWORDS

cancer screening; stakeholder engagement; Quality Function Deployment; web health app

Introduction

Breast and cervical cancers are the most common types of cancer in women in most countries worldwide [1]. In 2020, there were approximately 2.3 million new cases of breast cancer and 660,000 new cases of cervical cancer globally with up to 90% of breast and cervical cancer deaths occurred in low- and middle-income countries [2,3]. In Malaysia, breast and cervical cancer are the two most common cancers among women and result in high morbidity and mortality. In 2022, breast and cervical cancer accounted for about 16% and 4% of new cases and 11% and 3.2% of cancer related deaths, respectively [4].

Despite the availability of effective screening strategies, access to and uptake of screening remains suboptimal, particularly for minority populations and women with low income or low literacy [5]. This disparity often results in a significant gap in the stage of cancer detection, with individuals from lower-income backgrounds presenting at more advanced stages, ultimately affecting survival rates and exacerbating health inequities between high-income and low-income populations [6].

Digital health interventions, such as mobile apps, offer a promising avenue for improving the accessibility and uptake of screening, but its efficacy is highly dependent on the target population [7]. Previous literature has shown that increased involvement of the target population in the development of digital health tools leads to a greater acceptance of their use, but few have involved the target population consistently in the development of such interventions [8-13]. Stakeholder engagement, which encompasses actively involving those who have an interest in a particular decision, is critical to the successful development of digital health interventions [9]. Stakeholder engagement in the development of a mobile app is considered important when structural, cultural, and individual practices can affect its application. This underscores the necessity of addressing these potential barriers during the process of stakeholder engagement in the development of an app [10].

Stakeholder engagement in the development of digital health interventions in the Southeast Asia region is limited. As demonstrated by the MomConnect project conducted at primary care clinics in South Africa, stakeholder engagement is crucial for ensuring the successful implementation, scalability, and sustainability of such interventions [9]. The intended outcome is to achieve a holistic digital health solution that addresses the needs of marginalized communities, improves health services, and overcomes challenges like stakeholder mismanagement, lack of political support, and funding issues. This highlights the importance of obtaining the diverse perspectives of stakeholders that represent the Ministry of Health, clinical staff, and technical experts as well as the primary care service users who are the patients, as in the MomConnect project. This outcome is achieved through continuous collaboration among stakeholders, understanding the digital health ecosystem, and adopting a stakeholder-centered design approach [9-11,13].

Facilitating diverse stakeholder engagement effectively means creating an inclusive environment where all relevant parties, from patients to tech developers, are actively involved [13]. This engagement emphasizes open communication, regular feedback, and collaboration. By doing so, it can enhance information sharing and social interaction, which are crucial components of health app development. By aligning the development process with the needs and interests of the community it seeks to serve, the app's applicability and effectiveness can be enhanced [14]. Furthermore, stakeholder engagement strategies are applicable across a range of community-based health interventions, including the development of an app [15].

Existing digital health interventions have faced challenges in user engagement and content relevance, highlighting the need for apps that are both accessible and grounded in user-centered design [16]. In addition, according to the Malaysia National Health and Morbidity Survey in 2023, about 71% of women have never had a mammogram and 65% of women had not had a Papanicolaou test (Pap smear) in the past 3 years [17]. Further, the proportion of breast and cervical cancers diagnosed at late stages were 50.5% (stage 3) and 47.1% (stage 4) [18]. Hence, to address this gap, we conducted a research project entitled "The Effectiveness of an Interactive Web Application to Motivate and Raise Awareness on Early Detection of Breast and Cervical Cancers (The MAwar study)," a mixed methods feasibility study comprising a pre-post study and a qualitative study that were designed to be conducted in 3 urban primary care clinics in Malaysia. The MAwar study was preceded by a qualitative study that assessed the acceptability of an online web application to promote breast and cervical cancer screenings [19]. The qualitative study involved 15 health care professionals and 25 patients at primary care clinics in urban Malaysia. The findings of the qualitative study highlighted suggestions to incorporate easy-to-use features and visually comprehensible formats (such as more visuals and less text) and provide support for navigating the web app. The content recommended was including individual risk assessment and information on screening benefits, as well as addressing barriers like fear, embarrassment, logistics, and costs for patients. Utilizing the

XSL•FO
RenderX

foundational insights from this study, which focused on factors influencing women's engagement in cancer screenings, we incorporated these findings into the MAwar app's design.

This study aimed to describe the process of stakeholder engagement and prioritize a broad range of content and user interfaces appropriate for the development of a web-based application called MAwar. Through incorporating stakeholder engagement in the application development process, we aimed to create a tool that appropriately meets the needs and expectations of its users, thereby increasing the uptake of breast and cervical cancer screening.

Methods

Design and Setting

We conducted a cross-sectional study for stakeholder engagement between January and May 2022 at a public health district office in Selangor, Malaysia. This stakeholder engagement process allowed the researchers to develop a patient-centered intervention together with key stakeholders and community representatives. The health district office in this study oversees 12 public primary care clinics with existing cervical and breast cancer screening programs. In these clinics, Papanicolaou tests and clinical breast examination services are available. These clinics provide access to ultrasound and mammography services, as well as access to referral for early cancer detection and further evaluation and management in a public hospital in the district through the screening programs. This study was part of a larger research project entitled "The Effectiveness of an Interactive Web Application to Motivate and Raise Awareness on Early Detection of Breast and Cervical Cancers (The MAwar study)." The study protocol was registered with the ISRCTN Registry (ISRCTN10403163).

Participants

In our study, the stakeholders were identified through a discussion with the state Head of Service and with a research organization that was involved in cancer patient navigation programs in the state. The research assistant contacted the relevant individuals and invite them to participate after providing them with participant information. Both verbal and written consent were obtained from the participants prior to the stakeholder engagement.

The stakeholders were purposively selected for their relevant roles in breast and cervical cancer screening. The health care providers selected included primary care doctors (family physicians and medical officers) and nurses involved in cancer screening programs to ensure the app's development was aligned with clinical protocols and needs. We also included patients with varying experiences with screening to obtain insights into user requirements and app functionality needed to enhance the user-centric design of the app [20]. A representative from the Patient Advisory Board was invited to contribute perspectives on patient advocacy and community engagement [21,22]. Complementing this, technical experts were invited to provide their expertise in software and user interface design, ensuring the app was built with the latest technological advancements to improve efficiency and accessibility [23,24]. This purposive and balanced selection of stakeholders aimed to create a multidisciplinary team capable of addressing the multifaceted challenges of cancer screening, ensuring that the MAwar App meets the complex needs of all involved stakeholders [23,24].

Process of Stakeholder Engagement

The stakeholder engagement used a stakeholder-driven approach was based on the INVOLVE UK guidelines on patients and public involvement [21] and the Australian Government Department of Health's Stakeholder Engagement Framework [22]. The rationale for integrating these 2 distinct frameworks stems from their complementary strengths. The Australian framework offers a robust structure for systematically identifying and involving stakeholders in health-related projects, ensuring transparency and accountability [22]. On the other hand, the INVOLVE guidelines provide a specialized focus on patient and public involvement, emphasizing the importance of including the voices of end users in health research to enhance relevance and impact [21].

By merging these approaches, the engagement strategy was enriched, leveraging the structured engagement and accountability. This hybrid approach facilitated a more comprehensive understanding of stakeholder needs, enabling the development of an application that is not only technically sound but also deeply meaningful and aligned with the community's expectations and experiences.

To further validate this approach, an adapted version of the combined framework was reviewed by a panel of experts comprising 3 family physicians with significant experience in stakeholder engagement and patient and public involvement. This expert review ensured that the engagement strategy was not only theoretically sound but also viable and tailored to the specific context of the health app development. It comprised 5 strategic steps. These steps systematically capture the essence of our community-centric development of the MAwar application, each underscored by robust, participatory principles. Table 1 explains the 5 steps of the stakeholder engagement process for the MAwar application development.

XSL•FO
RenderX

Table . Five-step stakeholder engagement process for MAwar app development.

Step	Phase	Action	Study element	Timeline
1	Identify stakeholders	Engage diverse groups with a stake in app outcomes	Health care professionals, patients, public representatives, and tech experts identified	January 2022
2	Define methods of engagement	Determine stakeholder engagement depth using appropriate tools	QFD[a] framework and online survey tool were adopted for the group engagement	February 2022
3	Develop and implement engagement plan	Design detailed engagement plans tailored to stakeholder needs	Engagement plan was developed from the preliminary qualitative study insights [14]	March 2022
4	Evaluate the process	Gather and assess stakeholder feedback	Through QFD framework, prioritization of the top 5 user requirements and correlation of these with the technical requirements	April 2022
5	Provide feedback and follow-up	Share results for continuous improvement	Feedback loop established for app feature refinement	May 2022

[a]QFD: Quality Function Deployment.

Quality Function Deployment Framework for MAwar Application Design

The Quality Function Deployment (QFD) framework is a customer-driven planning process that guides the design of products and services, as it was originally designed for business [25]. This methodology was helpful in refining the MAwar application, translating insights from the previous study into specific design specifications to ensure the application effectively meets user needs. By applying QFD principles, our team was able to prioritize features that directly address user-reported barriers and motivators, enhancing the app's usability and impact on cancer screening uptake. The QFD process facilitated a structured approach to aligning the app's development with the expressed needs, challenges, and preferences of potential users, as identified in the preliminary qualitative study [19]. In addition, we used the term "user" to describe the "customer" as used in the QFD framework in order to address our study participants more appropriately.

The QFD process began with defining the user needs (the WHATs) and the technical requirements (the HOWs) based on the qualitative study [19] and incorporating these in the QFD framework. Table 2 shows the 25 user needs (the WHATs) identified, capturing essential requirements for the app based on barriers and motivators for cancer screening.

Table 3 presents the 28 technical requirements (the HOWs) from the QFD framework, detailing how each user's need is addressed through specific app functionalities and features.

XSL•FO
RenderX

Table . List of WHATs (user requirements) using the Quality Function Deployment framework.

Number	User requirements
1	Free to use (without cost)
2	Information about breast and cervical cancer (statistics, causes, signs and symptoms, risk factors, complications)
3	Information on the benefits of screening (survival rate with/without screening)
4	Information on screening facilities: who, where, recommendations, opening hours
5	A risk assessment function to know about my personal risk of breast and cervical cancers
6	Data are kept confidential and secure
7	Information about screening procedure for breast and cervical cancers
8	A list of frequently asked questions about the concerns about breast and cervical cancer screening
9	Information is presented in video format
10	Only information related to me is shown instead of all information (personalized)
11	Reminder function for appointment using, for example, WhatsApp
12	Information about cost of screening for breast and cervical cancers
13	Information is presented in picture format
14	Health screening appointment settings with health care center
15	Information on treatments for breast and cervical cancers
16	Available in 2 languages (Malay and English)
17	Testimonies from people who have undergone screening
18	Accessible using a QR code
19	Latest updates on ongoing health screening program locally
20	Access to a support group on social media via the web app
21	Motivational quotes to empower women to go for screening
22	Can interact with counsellors/health care providers via the web app
23	Information format is easy to understand and navigate
24	Reminder function for annual check-up
25	Use of a credible doctor as the avatar in the web app

XSL•FO
RenderX

Table . List of HOWs (technical requirements) using the Quality Function Deployment framework.

Number	Technical requirements
1	Funds (research grants, sponsors, etc)
2	Expert inputs (ie, Ministry of Health, institutes of higher learning)
3	Information (clinical evidence, literature review, clinical practice guide)
4	Ministry of Health website (list of locations of health clinics)
5	Expert input (encrypt data, access control, https)
6	SMS services (code and telco provider)
7	Video placement (web, video uploading platforms, ie, YouTube, Vimeo)
8	Testimonial video (patients, experience)
9	Video animation (Powtoon, Vyond)
10	Subtitles
11	Filtering information
12	Integration with WhatsApp
13	Screening center website (government and private)
14	Preliminary findings (barriers, motivators)
15	Image placement (web)
16	Image format (JPEG, GIF)
17	Related links (clinic appointment system, eg, QueueMed, Encoromed, BookDoc)
18	Content creator (translator)
19	Dual page
20	QR code generator
21	Login function (name, risk profile, age, history, recommendation, only needing to log in once, data are stored)
22	Access for medical practitioner to post screening campaign on the web app (post web app to social medias)
23	Related links in support group (on social media) and web app
24	Chatbox on web app
25	Voice app (VoIP, dial in)
26	Text format (eg, Arial)
27	Integration with email (for appointment reminder)
28	Celebrity face

Following this step, we integrated the QFD method with an online polling tool (VEVOX [26]) to gather feedback from the participants. They were asked to rate the importance of each feature on a scale from 1 to 5 (1 being the least important and 5 being the most important). Each rating was used to assign a “user importance value” (UIV) to each feature in the QFD, ensuring that the design of the MAwar application aligned with the needs and expectations of its users.

Then, the participants were asked to score the correlations between the WHATs (user requirements) and the HOWs (technical requirements). The correlations were scored using a 3-point ordinal scale, with 1 indicating weak correlation, 3 indicating medium correlation, and 9 indicating the strongest correlation [27]. This scoring system provided a quantitative measure of the alignment between the stakeholders’ needs and the application’s design features.

Ethical Considerations

This study obtained ethics approval from the Malaysia Ministry of Health Medical Research Ethics Committee (NMRR-21-951-58339 [IIR]) as well as permission from the State Health Department. Participants who voluntarily agreed to participate provided verbal and written informed consent after they had read the participant information sheet. We provided honoraria to participants as reimbursement for their travel to compensate for their time and participation in the study. Nonidentifiable identification codes were assigned to the participants for the purpose of data entry and data analysis. The consent forms and questionnaires will be stored in a locked

XSL•FO
RenderX

filing cabinet for 7 years and will only be accessible by the research team. After 7 years, these documents will be shredded and disposed of in secure bins. In the report writing or publication, the participants will not be identified.

Data Analysis

Data were analyzed descriptively using Microsoft Excel (Microsoft Corp). The QFD analysis ranked the technical requirements of the MAwar application by assessing the UIVs and correlation scores (CS). The UIVs were calculated based on the average of the participants' ratings of the importance of each feature on a scale from 1 (least important) to 5 (most important). The CS were rated using a 3-point ordinal scale (1=weak, 3=medium, and 9=strong correlations) to measure how each technical requirement supports user needs [21].

The technical elements were identified using the absolute importance value. These were calculated for the sum of UIV × CS for each of the "HOWs," showing the direct contribution of each technical requirement. These AI values were then summed across all relevant user needs to compute the weighted score (WS)=sum of absolute importance value, ranking the significance of each feature of the "WHATs" within the overall app structure. All these were ranked to highlight the most essential components of the MAwar application [21]. These analytical steps provided a detailed, quantified overview of the key priorities for the MAwar application's development. It identified the features most valued by stakeholders and pinpointed the technical requirements crucial to delivering these features, thus ensuring the application's design perfectly matched the needs and preferences of its intended users. Given the extensive data from the QFD (see the complete QFD analysis in Multimedia Appendix 1), we presented the top outcomes of the list of WHATs to represent the users' needs in the Results section to focus on the most significant findings, ensuring both clarity and relevance in our discussion of key impacts.

Results

Overview

We engaged 13 stakeholders, which was within the acceptable group size of 10 - 15 participants for achieving a balance of diverse perspectives and depth in discussions for stakeholder engagement [28]. The participants included 4 female patients, one of whom serves on the Patient Advisory Panel, who provided insights into patient experiences [23]. In the public health care sector, 6 participants included medical doctors and nurses, who brought specialized knowledge to the project based on their involvement in providing services to women related to health screening and education [24]. Additionally, 3 technical experts contributed their software and user interface design expertise, ensuring the application was developed with the latest technological advancements to improve efficiency and accessibility (see Table 4) [29].

Table . Stakeholders' classifications and roles.

Stakeholder group (subgroup)	Participants, n	Roles
Patient and public involvement (patients)	4	Advisory panel member for primary care clinic, women who had either underwent or never went for health screening
Health care professionals (medical doctors)	4	Administrator at Health District Office, a primary care physician, and 2 medical officers involved in the development of protocols and counseling on health screenings
Health care professionals (nurses)	2	Involved in health screening and education
Technical experts (software designer)	3	Involved in application development and user interface design

Users' Needs

Table 5 shows the top 20 user needs for the development of the MAwar application based on QFD analysis (refer to Multimedia Appendix 1 for the complete QFD analysis). The main 5 features were that the application should be free to use and provide comprehensive cancer information, including details on the benefits of screening, screening facilities, and personal risk assessment. Data security and detailed screening procedures were also deemed as major components. Educational features included Frequently Asked Questions, video-based facts, and personalized content. Although reminder functions for screening appointments were also rated highly, they were not applicable for our web application and are more valid for mobile apps.

Additional user needs include information on screening costs, cancer prevention strategies, and pictorial representations of key details. Users also highlighted the importance of appointment booking, cancer treatment information, and bilingual support in Malay and English. Other notable features include testimonials from screened individuals or cancer survivors and QR code access for convenience.

These findings provide crucial insights into user expectations, ensuring the MAwar application is developed with a user-centric approach, prioritizing accessibility, education, and engagement. However, since the MAwar application is a web application, features such as a reminder function for appointments using SMS text messaging or WhatsApp were not able to be incorporated into the application.

XSL•FO
RenderX

Table . The list of top 20 users' needs ranked by user importance value based on the Quality Function Deployment framework.

List of WHATs	User importance value, mean (SD)	Rank
Free to use (without cost)	4.64 (0.81)	1
Information about breast and cervical cancers (statistics, causes, signs and symptoms, risk factors, complications)	4.55 (0.69)	2
Information on the benefits of screening (survival rate with/without screening)	4.45 (0.68)	3
Information on screening facilities: who, where, recommendations, opening hours	4.45 (0.68)	4
A risk assessment function to know about my personal risk of breast and cervical cancers	4.45 (0.68)	5
Data are kept confidential and secure	4.45 (0.68)	6
Reminder function for appointment using, for example, SMS text messaging	4.36 (0.65)	7
Information about screening procedure for breast and cervical cancers	4.18 (0.60)	8
A list of frequently asked questions about the concerns about breast and cervical cancer screening	4.00 (0.89)	9
Information is presented in video format	4.00 (0.89)	10
Only information related to me is shown instead of all information (personalized learning)	4.00 (0.89)	11
Reminder function for appointment using, for example, WhatsApp	4.00 (0.89)	12
Information about cost of screening for breast and cervical cancers	3.91 (0.72)	13
Information on prevention for breast and cervical cancers (lifestyle factors like diet and exercise)	3.91 (0.72)	14
Information is presented in picture format	3.91 (0.72)	15
Function to set appointments with health screening center	3.91 (0.72)	16
Information on the treatments for breast and cervical cancers	3.64 (0.61)	17
Available in 2 languages (Malay and English)	3.64 (0.61)	18
Testimonies from people who have undergone screening or cancer survivors	3.36 (0.58)	19
Accessible using a QR code	3.36 (0.58)	20

The MAwar Application's Features

The QFD analysis identified the top 5 priority ranks of technical requirements relevant for the MAwar application's features for the health information platform, specifically targeting breast and cervical cancer awareness (see Table 6). These technical requirements ensure the MAwar application provides credible, accessible, and medically proven information, incorporating expert inputs, government resources, multimedia engagement, and seamless connectivity with health care services. Expert inputs from the Ministry of Health and institutions of higher learning form the foundation of medically reviewed content, ensuring accuracy and trust. The information component, including clinical evidence and literature reviews, strengthens the app's role as an educational tool. Ministry of Health integration facilitates access to official screening locations and health care services, while video content enhances user engagement through educational animations and testimonials. Lastly, integration features such as the WhatsApp application programming interface and screening center connections support direct interaction with health care providers. These prioritized technical elements enable the MAwar application to effectively promote breast and cervical cancer awareness, supporting early detection and informed decision-making among users.

XSL•FO
RenderX

Table . The top 10 technical requirements ranked by weighted score from the Quality Function Deployment framework for the MAwar application.

User requirements (list of WHATs)	Weighted score	Rank
Information about screening procedure for breast and cervical cancers	367.84	1
Information on the treatments for breast and cervical cancers	345.80	2
Information on the benefits of screening (survival rate with/without screening)	333.75	3
Information about breast and cervical cancer (statistics, causes, signs and symptoms, risk factors, complications)	332.15	4
A list of frequently asked questions about concerns about breast and cervical cancer screening	312.00	5
Information is presented in video format	284.00	6
Information on screening facilities: who, where, recommendations, opening hours	275.90	7
A risk assessment function to know about my personal risk of breast and cervical cancers	191.35	8
Information about the cost of screening for breast and cervical cancers	156.40	9
Function to set appointments with a health screening center[a]	148.58	10

[a]Feature evaluated but not applicable to the web application.

Discussion

Principal Findings

The findings from our stakeholder engagement underscored the crucial role of targeted, comprehensive information delivery in web health applications like the MAwar application. With the increasing accessibility of digital technologies, web-based and mobile health interventions have emerged as effective tools to overcome barriers to breast and cervical cancer screening. Several apps, such as BrAware, Amate, and iBreastCheck, have shown promising results to improve intention and uptake for cancer screenings [30-32]. The BrAware app has been effective in increasing knowledge of breast cancer risk factors and improving awareness of warning signs, as demonstrated in a study conducted in northeast Peninsular Malaysia [30]. Similarly, the Amate app in Colombia aims to improve access to early detection of both breast and cervical cancers by educating patients about and guiding them to national screening programs, resulting in the successful completion of recommended screening tests among some participants [31]. These interventions provided tailored health information, risk education, and reminders, which have been shown to improve health behaviors in various contexts. The prioritization of detailed cancer information, screening benefits, and interactive features like personal risk assessments align with previous literature suggesting that informed users are more likely to participate actively in their health management, potentially improving outcomes [33]. The emphasis on screening procedures and frequently asked questions further highlights the necessity for health apps to be both educational and reassuring, addressing common concerns and misconceptions that can hinder screening uptake [34].

User Needs

Providing comprehensive and accurate information about breast and cervical cancer through health applications is essential for enhancing user engagement and promoting better health outcomes. Detailed educational content helps users understand the importance of early detection and treatment, encouraging participation in screening programs. A mobile app designed for Colombian women has shown that tailored educational content and risk assessments can significantly raise awareness and participation rates in breast and cervical cancer screening programs [28]. Moreover, providing robust information on the benefits of screening empowers users to take proactive steps in cancer prevention and screening, leading to improved health outcomes [19,35].

Integrating personalized risk assessment functions into health applications enhances user engagement by providing tailored insights into specific health risks, thereby motivating preventive behaviors. Personalized health tools have been shown to significantly increase user interaction and satisfaction with health apps [36]. Additionally, providing detailed information about screening procedures for breast and cervical cancer helps demystify the process and encourages participation by alleviating fears and misconceptions. Clear, accessible information about screening can significantly increase the uptake of screening services [37]. Furthermore, comprehensive frequently asked questions sections addressing common concerns and misconceptions about breast and cervical cancer screenings can provide reassurance, reduce anxiety, and improve user confidence in the procedures. This approach has been proven to improve user satisfaction and participation in screening programs [35,38].

XSL•FO
RenderX

Features of the MAwar Application

The leading features identified for the MAwar application reflect a strategic emphasis on ensuring that content is not only trustworthy but also visually engaging and well-structured. By focusing on expert inputs, text formatting, direct contributions from content creators, strategic image placement, and image formatting, we aimed to enhance user comprehension and engagement. This approach facilitates a better understanding of critical health information, which is crucial for informed decision-making among users. Experts inputs ensure that the information is accurate and reliable, while strategic text formatting and visual content placement make the information accessible and appealing. Studies have shown that well-structured and visually appealing information presentation can greatly increase the understandability and usability of health apps [39].

By integrating these top technical requirements, the MAwar application systematically prioritizes development efforts to meet user needs effectively. Expert contributions ensure content accuracy, while strategic image placement and formatting enhance visual engagement, making complex medical information more digestible. For instance, health applications that included expertly curated content and optimized visual presentation have been found to improve users' satisfaction and interaction significantly [40]. Moreover, text formatting plays a critical role in making information more understandable, as demonstrated by studies on health care applications [41]. Strategic image placement further aids in improving user comprehension and retention of health information [42]. These features collectively help address the challenges of conveying complex health information through digital platforms, ultimately supporting users in making informed health decisions [35].

Stakeholder Engagement and Public and Patient Involvement

Incorporating a broad spectrum of stakeholders through stakeholder engagement and public and patient involvement has added significant value to the MAwar application development. By engaging medical professionals, patients, and technical experts, the development process benefited from a holistic view that encompassed diverse perspectives, enhancing the app's relevance and usability across different user groups. Previous studies have highlighted that stakeholder engagement and public and patient involvement contribute to higher product acceptance and greater adherence to health interventions, as these processes ensure that the end product aligns closely with the real-world needs and expectations of its users [43].

The inclusion of technical experts in the stakeholder engagement process not only ensures that the app is built using the latest technological advancements but also guarantees that the technical execution matches the high standards required for health applications. This collaboration between health professionals and technologists fosters innovation, leading to more effective and user-friendly solutions, ultimately improving uptake of and adherence to recommended health practices [44].

The MAwar application, conceived through a collaborative, stakeholder-driven process, represents a significant step in leveraging digital health solutions to tackle cancer screening disparities [44]. By prioritizing accessibility, information quality, and clarity on benefits, the app directly confronts the barriers impeding cancer screening in underserved populations [45]. The stakeholder engagement underscored the application's relevance and potential impact, showcasing the power of inclusive design in improving health outcomes. The MAwar application focused on critical features reflecting user needs to empower women to take part in early cancer detection and management for targeted communities.

Strengths and Limitations of the Study

The stakeholder engagement process played a crucial role in identifying key features that are critical for meeting the needs and expectations of the MAwar application users. This approach ensured that the application was designed with a user-centric focus, increasing its potential effectiveness and usability.

However, there were some limitations. The patient representatives in our stakeholder engagement session did not include older clients aged 50 years and above. This age group is a significant demographic for breast and cervical cancer screening, and their unique needs and perspectives may not have been fully captured in our study. However, their perspective was included in an earlier qualitative study, the results of which were used for the discussion during the stakeholder engagement in this study [14].

Additionally, family members of clients were not involved in the stakeholder engagement process. As potential co-users of the app, their input could have provided valuable insights into the design and functionality of the application. Future research should consider involving these important stakeholders to ensure a more comprehensive understanding of user needs and preferences.

Conclusion

By focusing on features such as cost-free access, comprehensive cancer information, and detailed screening benefits, the application directly targets the primary obstacles to screening adherence, thereby fostering improved health outcomes in targeted communities. To enhance the app's efficacy and reach, future development efforts must prioritize the inclusion of a broader demographic spectrum, especially older individuals and potential co-users like family members. Such inclusivity will ensure the application's design and functionalities resonate more effectively with the diverse needs of its user base, ultimately maximizing its public health impact.

In summary, the MAwar application represents a strategic innovation in digital health, aiming to reduce disparities in cancer screening uptake through a user-centered design. Its foundation, built on rigorous stakeholder engagement and evidence-based design principles, sets a precedent for future public health interventions aimed at promoting early cancer detection and management.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank the Ministry of Higher Education and the Ministry of Health Malaysia for their support and permission to publish this article. This project was funded by the Ministry of Higher Education Fundamental Research Grant Scheme (FRGS/1/2020/SKK05/UPM/02/5).

Data Availability

All data have been presented in this manuscript and Multimedia Appendix 1.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Quality Function Deployment framework analysis.
[PDF File, 159 KB - formative_v9i1e65542_app1.pdf]

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021 May;71(3):209-249. [doi: 10.3322/caac.21660] [Medline: 33538338]
2. Arnold M, Morgan E, Rumgay H, et al. Current and future burden of breast cancer: global statistics for 2020 and 2040. Breast 2022 Dec;66:15-23. [doi: 10.1016/j.breast.2022.08.010] [Medline: 36084384]
3. Singh D, Vignat J, Lorenzoni V, et al. Global estimates of incidence and mortality of cervical cancer in 2020: a baseline analysis of the WHO Global Cervical Cancer Elimination Initiative. Lancet Glob Health 2023 Feb;11(2):e197-e206. [doi: 10.1016/S2214-109X(22)00501-0] [Medline: 36528031]
4. Ferlay J, Ervik M, Lam F, et al. Global cancer observatory: Malaysia 2022. International Agency for Research on Cancer, World Health Organization. 2024. URL: https://gco.iarc.who.int/today [accessed 2025-03-24]
5. Woldetsadik AB, Amhare AF, Bitew ST, Pei L, Lei J, Han J. Socio-demographic characteristics and associated factors influencing cervical cancer screening among women attending in St. Paul's Teaching and Referral Hospital, Ethiopia. BMC Womens Health 2020 Apr 6;20(1):70. [doi: 10.1186/s12905-020-00927-5] [Medline: 32252733]
6. Ji P, Gong Y, Jiang CC, Hu X, Di GH, Shao ZM. Association between socioeconomic factors at diagnosis and survival in breast cancer: a population-based study. Cancer Med 2020 Mar;9(5):1922-1936. [doi: 10.1002/cam4.2842] [Medline: 31960597]
7. Richardson-Parry A, Silva M, Valderas JM, Donde S, Woodruff S, van Vugt J. Interactive or tailored digital interventions to increase uptake in cervical, breast, and colorectal cancer screening to reduce health inequity: a systematic review. Eur J Cancer Prev 2023 Jul 1;32(4):396-409. [doi: 10.1097/CEJ.0000000000000796] [Medline: 37144585]
8. Radu I, Scheermesser M, Spiess MR, Schulze C, Händler-Schuster D, Pehlke-Milde J. Digital health for migrants, ethnic and cultural minorities and the role of participatory development: a scoping review. Int J Environ Res Public Health 2023 Oct 23;20(20):6962. [doi: 10.3390/ijerph20206962] [Medline: 37887700]
9. Sibuyi IN, de la Harpe R, Nyasulu P. A stakeholder-centered mHealth implementation inquiry within the digital health innovation ecosystem in South Africa: MomConnect as a demonstration case. JMIR mHealth uHealth 2022 Jun 16;10(6):e18188. [doi: 10.2196/18188] [Medline: 35708756]
10. Laird Y, Manner J, Baldwin L, et al. Stakeholders' experiences of the public health research process: time to change the system? Health Res Policy Syst 2020 Jul 18;18(1):83. [doi: 10.1186/s12961-020-00599-5] [Medline: 32682426]
11. Shrestha A, Tamrakar D, Shrestha B, et al. Stakeholder engagement in a hypertension and diabetes prevention research program: description and lessons learned. PLOS ONE 2022;17(10):e0276478. [doi: 10.1371/journal.pone.0276478] [Medline: 36264860]
12. Stelfox HT, Niven DJ, Clement FM, et al. Stakeholder engagement to identify priorities for improving the quality and value of critical care. PLOS One 2015;10(10):e0140141. [doi: 10.1371/journal.pone.0140141] [Medline: 26492196]
13. Chen SCI, Liu C, Wang Z, et al. How geographical isolation and aging in place can be accommodated through connected health stakeholder management: qualitative study with focus groups. J Med Internet Res 2020 May 27;22(5):e15976. [doi: 10.2196/15976] [Medline: 32459181]
14. Petkovic J, Magwood O, Lytvyn L, et al. Key issues for stakeholder engagement in the development of health and healthcare guidelines. Res Involv Engagem 2023 Apr 28;9(1):27. [doi: 10.1186/s40900-023-00433-6] [Medline: 37118762]
15. Holcomb J, Ferguson GM, Sun J, Walton GH, Highfield L. Stakeholder engagement in adoption, implementation, and sustainment of an evidence-based intervention to increase mammography adherence among low-income women. J Cancer Educ 2022 Oct;37(5):1486-1495. [doi: 10.1007/s13187-021-01988-2] [Medline: 33754327]

XSL•FO
RenderX

16. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr 18;15(4):e86. [doi: 10.2196/jmir.2583] [Medline: 23598614]
17. National Health and Morbidity Survey (NHMS) 2023: non-communicable diseases and healthcare demand. Institute for Public Health. URL: https://iku.gov.my/nhms-2023 [accessed 2025-03-24]
18. Summary of malaysia national cancer registry report 2017-2021. Ministry of Health Malaysia. URL: https://nci.moh.gov.my/images/pdf_folder/SUMMARY-OF-MALAYSIA-NATIONAL-CANCER-REGISTRY-REPORT-2017-2021.pdf [accessed 2025-03-24]
19. Cheong AT, Lee PY, Ghazali SS, et al. A qualitative exploration of the acceptability of an online web application to promote breast and cervical cancer screening in primary care settings in Malaysia. JCO Global Oncology 2022 May;8(Supplement_1):42-42. [doi: 10.1200/GO.22.48000]
20. Hennemann S, Beutel ME, Zwerenz R. Ready for eHealth? Health professionals' acceptance and adoption of eHealth interventions in inpatient routine care. J Health Commun 2017 Mar;22(3):274-284. [doi: 10.1080/10810730.2017.1284286] [Medline: 28248626]
21. Briefing notes for researchers: involving the public in NHS, public health and social care research. INVOLVE. URL: https://www.invo.org.uk/wp-content/uploads/2014/11/9938_INVOLVE_Briefing_Notes_WEB.pdf [accessed 2025-01-28]
22. Stakeholder engagement framework. Australian Government Department of Health and Aged Care. URL: https://www.health.gov.au/sites/default/files/stakeholder-engagement-framework.pdf [accessed 2025-01-28]
23. van Velsen L, Evers M, Bara CD, Op den Akker H, Boerema S, Hermens H. Understanding the acceptance of an eHealth technology in the early stages of development: an end-user walkthrough approach and two case studies. JMIR Form Res 2018 Jun 15;2(1):e10474. [doi: 10.2196/10474] [Medline: 30684434]
24. Kelders SM, van Zyl LE, Ludden GDS. The concept and components of engagement in different domains applied to eHealth: a systematic scoping review. Front Psychol 2020;11:926. [doi: 10.3389/fpsyg.2020.00926] [Medline: 32536888]
25. Akao Y. Quality Function Deployment: Integrating Customer Requirements into Product Design: Productivity Press; 1990. [doi: 10.4324/9781003578833]
26. Vevox Software Reviews. URL: https://www.softwareadvice.com/polling/vevox-profile/ [accessed 2025-01-28]
27. Franceschini F, Rupil A. Rating scales and prioritization in QFD. International Journal of Quality & Reliability Management 1999 Feb 1;16(1):85-97. [doi: 10.1108/02656719910250881]
28. Goodman MS, Ackermann N, Bowen DJ, Panel D, Thompson VS. Reaching consensus on principles of stakeholder engagement in research. Prog Community Health Partnersh 2020;14(1):117-127. [doi: 10.1353/cpr.2020.0014] [Medline: 32280129]
29. Concannon TW, Fuster M, Saunders T, et al. A systematic review of stakeholder engagement in comparative effectiveness and patient-centered outcomes research. J Gen Intern Med 2014 Dec;29(12):1692-1701. [doi: 10.1007/s11606-014-2878-x] [Medline: 24893581]
30. Yusuf A, P Iskandar YH, Ab Hadi IS, Nasution A, Lean Keng S. Breast awareness mobile apps for health education and promotion for breast cancer. Front Public Health 2022;10:951641. [doi: 10.3389/fpubh.2022.951641] [Medline: 36324460]
31. Munoz-Zuluaga CA, Gallo-Pérez JD, Pérez-Bustos A, et al. Mobile applications: breaking barriers to early breast and cervical cancer detection in underserved communities. JCO Oncol Pract 2021 Mar;17(3):e323-e335. [doi: 10.1200/OP.20.00665] [Medline: 33417491]
32. Mango VL, Olasehinde O, Omisore AD, et al. The iBreastExam versus clinical breast examination for breast evaluation in high risk and symptomatic Nigerian women: a prospective study. Lancet Glob Health 2022 Apr;10(4):e555-e563. [doi: 10.1016/S2214-109X(22)00030-4] [Medline: 35303464]
33. Beidas RS, Stewart RE, Adams DR, et al. A multi-level examination of stakeholder perspectives of implementation of evidence-based practices in a large urban publicly-funded mental health system. Adm Policy Ment Health 2016 Nov;43(6):893-908. [doi: 10.1007/s10488-015-0705-2] [Medline: 26658692]
34. Qudah B, Luetsch K. The influence of mobile health applications on patient - healthcare provider relationships: a systematic, narrative review. Patient Educ Couns 2019 Jun;102(6):1080-1089. [doi: 10.1016/j.pec.2019.01.021] [Medline: 30745178]
35. Haddad SM, Souza RT, Cecatti JG. Mobile technology in health (mHealth) and antenatal care-Searching for apps and available solutions: a systematic review. Int J Med Inform 2019 Jul;127:1-8. [doi: 10.1016/j.ijmedinf.2019.04.008] [Medline: 31128820]
36. Cai T, Huang Y, Zhang Y, Lu Z, Huang Q, Yuan C. Mobile health applications for the care of patients with breast cancer: a scoping review. Int J Nurs Sci 2021 Oct 10;8(4):470-476. [doi: 10.1016/j.ijnss.2021.07.003] [Medline: 34631997]
37. Cankaya BA, Yildirim Baş F, Cankaya İ. Using mobile applications in cancer scanning and evaluating the effectiveness. Turkish Journal of Family Medicine and Primary Care 2021;15(1):49-55. [doi: 10.21763/tjfmpc.803805]
38. Houghton LC, Howland RE, McDonald JA. Mobilizing breast cancer prevention research through smartphone apps: a systematic review of the literature. Front Public Health 2019;7:298. [doi: 10.3389/fpubh.2019.00298] [Medline: 31781525]
39. Olaniyi BY, Fernández del Río A, Periáñez Á, Bellhouse L. User engagement in mobile health applications. Presented at: KDD '22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; Aug 14-18, 2022; Washington DC p. 4704-4712. [doi: 10.1145/3534678.3542681]

XSL•FO
RenderX

40. Boissin C, Blom L, Wallis L, Laflamme L. Image-based teleconsultation using smartphones or tablets: qualitative assessment of medical experts. Emerg Med J 2017 Feb;34(2):95-99. [doi: 10.1136/emermed-2015-205258] [Medline: 27707791]
41. Mosa ASM, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12(1):67. [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
42. Wang Z, Zhang L, Ma L, Liu B. Modeling medical services with mobile health applications. J Healthc Eng 2018;2018(1):1385034. [doi: 10.1155/2018/1385034] [Medline: 29849991]
43. Lingg M, Lütschg V. Health system stakeholders' perspective on the role of mobile health and its adoption in the Swiss health system: qualitative study. JMIR mHealth uHealth 2020 May 11;8(5):e17315. [doi: 10.2196/17315] [Medline: 32391802]
44. Shippee ND, Domecq Garces JP, Prutsky Lopez GJ, et al. Patient and service user engagement in research: a systematic review and synthesized framework. Health Expect 2015 Oct;18(5):1151-1166. [doi: 10.1111/hex.12090] [Medline: 23731468]
45. Joosten YA, Israel TL, Williams NA, et al. Community engagement studios: a structured approach to obtaining meaningful input from stakeholders to inform research. Acad Med 2015 Dec;90(12):1646-1650. [doi: 10.1097/ACM.0000000000000794] [Medline: 26107879]

Abbreviations

CS: correlation scores
QFD: Quality Function Deployment
UIV: user importance values
WS: weighted score

Edited by A Mavragani; submitted 20.08.24; peer-reviewed by L Guo; revised version received 31.01.25; accepted 18.02.25; published 28.03.25.

Please cite as:
Nasrudin N, Sazlina SG, Cheong AT, Lee PY, Teo SH, Aneesa AR, Teo CH, Rokhani FZ, Haron NA, Harrun NH, Ho BK, Mohamed Isa S
Increasing the Uptake of Breast and Cervical Cancer Screening Via the MAwar Application: Stakeholder-Driven Web Application Development Study
JMIR Form Res 2025;9:e65542
URL: https://formative.jmir.org/2025/1/e65542
doi:10.2196/65542

© Nurfarhana Nasrudin, Shariff-Ghazali Sazlina, Ai Theng Cheong, Ping Yein Lee, Soo-Hwang Teo, Abdul Rashid Aneesa, Chin Hai Teo, Fakhrul Zaman Rokhani, Nuzul Azam Haron, Noor Harzana Harrun, Bee Kiau Ho, Salbiah Mohamed Isa. Originally published in JMIR Formative Research (https://formative.jmir.org), 28.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Using a Hybrid of AI and Template-Based Method in Automatic Item Generation to Create Multiple-Choice Questions in Medical Education: Hybrid AIG

Yavuz Selim Kıyak[1,2], MD, PhD; Andrzej A Kononowicz[2], PhD

[1]Department of Medical Education and Informatics, Faculty of Medicine, Gazi University, Ankara, Turkey

[2]Department of Bioinformatics and Telemedicine, Jagiellonian University Medical College, Medyczna Str 7, Kraków, Poland

Corresponding Author:
Andrzej A Kononowicz, PhD
Department of Bioinformatics and Telemedicine, Jagiellonian University Medical College, Medyczna Str 7, Kraków, Poland

Abstract

Background: Template-based automatic item generation (AIG) is more efficient than traditional item writing but it still heavily relies on expert effort in model development. While nontemplate-based AIG, leveraging artificial intelligence (AI), offers efficiency, it faces accuracy challenges. Medical education, a field that relies heavily on both formative and summative assessments with multiple choice questions, is in dire need of AI-based support for the efficient automatic generation of items.

Objective: We aimed to propose a hybrid AIG to demonstrate whether it is possible to generate item templates using AI in the field of medical education.

Methods: This is a mixed-methods methodological study with proof-of-concept elements. We propose the hybrid AIG method as a structured series of interactions between a human subject matter expert and AI, designed as a collaborative authoring effort. The method leverages AI to generate item models (templates) and cognitive models to combine the advantages of the two AIG approaches. To demonstrate how to create item models using hybrid AIG, we used 2 medical multiple-choice questions: one on respiratory infections in adults and another on acute allergic reactions in the pediatric population.

Results: The hybrid AIG method we propose consists of 7 steps. The first 5 steps are performed by an expert in a customized AI environment. These involve providing a parent item, identifying elements for manipulation, selecting options and assigning values to elements, and generating the cognitive model. After a final expert review (Step 6), the content in the template can be used for item generation through a traditional (non-AI) software (Step 7). We showed that AI is capable of generating item templates for AIG under the control of a human expert in only 10 minutes. Leveraging AI in template development made it less challenging.

Conclusions: The hybrid AIG method transcends the traditional template-based approach by marrying the “art” that comes from AI as a “black box” with the “science” of algorithmic generation under the oversight of expert as a “marriage registrar”. It does not only capitalize on the strengths of both approaches but also mitigates their weaknesses, offering a human-AI collaboration to increase efficiency in medical education.

(JMIR Form Res 2025;9:e65726) doi:10.2196/65726

KEYWORDS

automatic item generation; ChatGPT; artificial intelligence; large language models; medical education; AI; hybrid; template-based method; hybrid AIG; mixed-method; multiple-choice question; multiple-choice; human-AI collaboration; human-AI; medical education; algorithm; expert

Introduction

Multiple-choice items are integral to written assessment in medical education, serving as a useful tool for assessing a wide range of knowledge and skills. Their common use spans from evaluating factual knowledge to clinical reasoning and decision-making in various domains [1]. This assessment format supports high-volume testing with the added advantage of automated scoring to enhance the efficiency of examinations in medical education.

The traditional way of writing multiple-choice items, characterized by manual development processes, presents significant challenges in scalability. This challenge stems from the intensive time and effort required to create and review each question. This laborious process, which demands expertise and resources, faces a bottleneck in scaling up to meet the demand for a vast quantity of quality items. The challenge is particularly pronounced in medical education, where only a progress test administration in a year requires having 2400 multiple-choice

XSL•FO
RenderX

items [2], showing the inefficiency of traditional methods in satisfying the needs of question banks in medical schools.

Template-based automatic item generation (AIG) is a scalable method used in educational assessment that leverages predefined templates to systematically produce multiple-choice items with the help of software [3]. It has been implemented in 7 languages [3-6] and different health professions [3,7-9]. It consists of 3 sequential stages: development of a cognitive model, development of an item model (template), and using software for the rapid generation of hundreds of items [10]. Item models developed based on cognitive models are structured by subject matter experts to include variables and rules for item generation, allowing for a more efficient creation of consistent questions. This approach enhances the scalability of item development and review compared with traditional item writing [3], addressing the demand for high-quality assessment materials across various educational domains. Template-based AIG, while generating high-quality multiple-choice items efficiently [11-15], still heavily relies on expert effort in development of cognitive models and item models. Although it depends on the content area and the expert's skills, a study reported that development of a cognitive model and an item model can take hours for a subject matter expert [10]. Furthermore, this development process necessitates high levels of extraneous cognitive load due to high element interactivity [16], which requires experts to deal with several components simultaneously. Therefore, "creating cognitive models for AIG is challenging" [3] and it "requires a lot of practice" for experts [3].

On the other hand, nontemplate-based AIG, which "can be guided by the syntactic, semantic, or sequential structure of a text" [3], is an approach that leverages natural language processing to generate assessment items without relying on predefined templates. Unlike the template-based method, this method uses the ability of artificial intelligence (AI) to generate content dynamically, for example, using ChatGPT, which is an AI-based chatbot developed by OpenAI, for creating items based on specific topics or learning outcomes provided by users [17-21]. This approach allows for the generation of diverse and complex questions in seconds, offering flexibility and efficiency in item development. However, this AI-driven approach struggles with issues of inaccuracy and inconsistency [18], especially when good prompting strategies [22] are not used [23]. In AI-driven item generation, such as with ChatGPT, these issues often emerge due to the model's reliance on its training data, which may not always align perfectly with the specific objectives intended by educators. For example, an AI might generate content that includes incorrect information, such as asserting that "the human heart only has two chambers" [24], or misinterpret the complexity level required for a medical education context. Furthermore, the "black box" [25] nature of these AI models complicates diagnosing and correcting these errors within the AI mechanism, as it is challenging to trace back how the AI arrived at a particular output. Therefore, it raises concerns regarding validity, reliability, transparency, fairness, and equity [26], and the process still requires subject matter experts to review and revise each generated question [18,19,21,22,27]. Although it is more efficient than traditional item writing, necessity for reviewing each question is still inefficient.

As outlined above, recent advancements in AIG have offered efficiency, yet each method—template-based and nontemplate-based—brings its own set of limitations. The gap, therefore, lies in the need for a method that merges the structured efficiency of template-based AIG with the content generation capabilities of AI-driven, nontemplate-based, approaches. This convergence could potentially address the pressing need for tools augmenting capabilities of medical educators in test development. We are aware of the interdependence of social (human) and technical elements within an organization [28], advocating for the design of systems that concurrently optimize both human and technological components to achieve effective outcomes. In AIG, this can be interpreted as the need to harmonize the collaboration of a subject matter expert and AI tools working together on developing item and cognitive models.

In this paper, we propose a hybrid AIG method that uses AI to generate an item model (template) and a cognitive model for applying the item template in a template-based item generation process. This capitalizes on the strengths of both approaches but also mitigates their respective weaknesses, offering a novel human-AI collaboration to increase AIG efficiency in medical education.

Methods

Study Process

This is a proof-of-concept study. Drawing on existing guidance on prompt engineering [29,30] and our experience in building human-AI dialogues [20,31], we developed a series of flipped interactions through a series of iterative refinements. In this approach, the AI initiates the conversation and prompts the human expert to provide the necessary data, enabling a collaborative process for generating and refining item templates. The AI shoulders the significant cognitive load involved in template development, thereby reducing the cognitive burden on experts, allowing them to focus on deciding which elements of the question are essential for assessing students' knowledge.

We used GPT Builder (OpenAI), a platform for customizing ChatGPT to the user's needs [32], to train our Custom GPT. Since we conducted this study in February 2024, the Custom GPT worked based on the frontier model at that time, GPT-4.

We will present our approach in a manner analogous to how Gierl et al [10] described the template-based AIG. To illustrate the process, we used two items: one from Gierl et al's work [3], and another from a multiple-choice question (MCQ) similar to an item in the Turkish National Medical Specialty Exam, TUS (2021/1, clinical question number 58). The reason for choosing Gierl et al's [3] item is that it allows readers to compare it with the existing item model developed by a subject matter expert. The reason for choosing a TUS item is that Gierl et al's [3] item model has likely already been processed by ChatGPT, so we also aimed to focus on an item that had not been modelled before.

XSL•FO
RenderX

Ethical Considerations

This study did not involve human subjects, patient data, or personal identifiers, and therefore did not require ethical approval. No informed consent was necessary as no human participants were involved. The study is based on the demonstration of a methodological approach using AI for automatic item generation.

Results

Hybrid AIG and Prompts

The Hybrid AIG consists of 7 steps, with the last 2 steps carried out outside the AI environment. The AI environment requires a specialized GPT trained for generating item models and cognitive models. The Custom GPT we developed is titled "Item Model Maker for AIG" and is accessible at [33]..

In addition, the prompt we used is provided in Multimedia Appendix 1.

Step 1: Providing a Parent Item

The starting point for developing both the item model and the cognitive model in the hybrid AIG is a parent item, reflecting the approach used in the template-based method. It functions as a prototype for generating new questions that follow the template's structure.

In the first step of the hybrid AIG, the expert provides a parent item to AI. Although it is not a requirement, it would be better if the parent item has been chosen from well-performing items in an exam (eg, a published item from a past medical licensure examination). This could mitigate the possible problems that can stem from the lack of quality in the parent item. If an expert does not have a parent item, they can use AI tools, such as Case-Based MCQ Generator [20], to generate an MCQ as a parent item.

The purpose of generating MCQs in medical education is for a human expert to assess the students' skills and knowledge and to evaluate whether they have grasped the nuances of the information presented, aiming for outcomes that are consistent with expert reasoning. The parent item includes the necessary components and embedded possible cognitive models required for this purpose, as it is approved and provided by the expert. It serves as the well-established basis and acts as a seed for generating a new template. In the subsequent steps, AI will primarily handle the cognitive work in using one of the possible underlying cognitive models in the parent item to generate a template. The expert's role will be to monitor and ensure the AI maintains a consistent cognitive model in its output.

We present below the parent items provided to AI, adopting them from the sources described in the Methods section.

Gierl et al's [3] item:

"A 22-year-old female sees her doctor and reports that she's been experiencing a mild cough and slight body aches that have developed over a few days.

Upon examination, she presents with an oral temperature of 37°C. What is the most likely diagnosis?

A. Hay fever
B. Ear infection
C. Common cold
D. Acute sinusitis
E. Seasonal influenza"

TUS Item:

"A fifteen-year-old girl, who became ill during her physical education class after lunch, is brought to the infirmary complaining of coughing, shortness of breath, dizziness, facial swelling and redness, as well as itching in her hands and feet. During the physical examination, the patient appears anxious, with a blood pressure reading of 80/50 mmHg, periorbital edema, and wheezing upon auscultation. Which of the following is the most likely diagnosis for this patient?

A. Exercise-induced asthma
B. Cholinergic urticaria
C. Pulmonary embolism
D. Anaphylaxis
E. Vasovagal syncope"

Step 2: Identification of Elements for Manipulation

In the next step, the expert identifies words or phrases in the parent item for variable manipulation.

In the identification process, the expert can, preferably, consider the importance of the content in the case that is more influential in deciding what the correct answer is. Table 1 shows AI's response and what the expert chose as the elements for manipulation.

XSL•FO
RenderX

Table . Identification of elements for manipulation.

Who speaks	Gierl et al's [3] item	TUS item
AI[a]	• Thank you for sharing the parent item. Could you please identify words or phrases in the question that you would like to manipulate for variable creation? These can include demographic details, symptoms, examination findings, or any other elements that could vary across different versions of this question.	• Based on the parent item you've provided, let's proceed to the next step. Could you please identify words or phrases in the scenario and question that you would like to have as variable elements for manipulation? These could be specific symptoms, the patient's demographics, the setting, or any other details you think should be variable.
Expert	• Element 1: mild cough • Element 2: slight body aches • Element 3: over a few days • Element 4: 37°C	• Element 1: during her physical education class after lunch • Element 2: coughing, shortness of breath, dizziness, facial swelling and redness • Element 3: a blood pressure reading of 80/50 mm Hg, periorbital edema, and wheezing upon auscultation

[a]AI: artificial intelligence.

Step 3: Replacement of Elements and Choosing Options

In this step, AI replaces the words and phrases received from the expert with the elements in brackets, then the expert reviews the output, and prompts AI to make necessary changes, if any. Subsequently, the expert chooses 2 or more options from the options in the parent item. The AI will use them to define the possible values of elements. Table 2 shows the replaced version and the expert's choice among the options.

Table . Replacement of elements and choosing options.

Who speaks	Gierl et al's [3] item	TUS item
AI[a]	Based on the elements you've identified, here's how the parent item could be modified with these variables in brackets: "A 22-year-old female sees her doctor and reports that she's been experiencing a [Element 1] and [Element 2] that have developed [Element 3]. Upon examination, she presents with an oral temperature of [Element 4]. What is the most likely diagnosis? A. Hay fever B. Ear infection C. Common cold D. Acute sinusitis E. Seasonal influenza" Could you please choose two or more options from the parent item that the variables will be focused on? This will help in offering variables for each element that significantly impact the correct answer according to medical sources and guidelines.	Great, I'll incorporate these elements as variables into the parent item. Here's how it will look: "A fifteen-year-old girl, who became ill [Element 1], is brought to the infirmary complaining of [Element 2], as well as itching in her hands and feet. During the physical examination, the patient appears anxious, with [Element3]. Which of the following is the most likely diagnosis for this patient? A. Exercise-induced asthma B. Cholinergic urticaria C. Pulmonary embolism D. Anaphylaxis E. Vasovagal syncope" Now, could you please choose two or more options from the answer choices that you'd like the variables to significantly impact? This will help in designing the variables for each element to ensure they align closely with the medical scenarios corresponding to the chosen options.
Expert	• Common cold • Seasonal influenza	• Cholinergic urticaria • Anaphylaxis

[a]AI: artificial intelligence.

Step 4: Offering Values in Elements

At this stage, AI suggest possible values for each element that can impact the correct answer. Then, the expert reviews the output and makes necessary changes, if any, by providing prompts. Table 3 shows the response of AI in order for offering values. It is important to note that "values" were presented as "variables" by AI in the example because it did not work well when we call them "values" (eg, it was prone to offer numbers rather than words). It may not be a problem in newer versions.

XSL•FO
RenderX

Table . Artificial intelligence's (AI's) offer for values.

Who speaks	Gierl et al's [3] item	TUS item
AI	Thank you for choosing "Common cold" and "Seasonal influenza" as the focus. Based on these options, I'll offer variables for each element that significantly impact the correct answer, considering medical sources and guidelines: Element 1: Cough • Variable 1: Persistent dry cough • Variable 2: Mild cough • Variable 3: Severe hacking cough Element 2: Body aches • Variable 1: Slight body aches • Variable 2: Moderate body aches • Variable 3: Severe muscle pains Element 3: Duration of symptoms • Variable 1: Suddenly over a few hours • Variable 2: Over a few days • Variable 3: Gradually over more than a week Element 4: Oral temperature • Variable 1: 37°C (normal) • Variable 2: 37.5°C to 38°C (mild fever) • Variable 3: Over 38°C (fever) Do these variables seem appropriate for creating different versions of the MCQ[a] item focused on distinguishing between the common cold and seasonal influenza? If you need any adjustments or have additional suggestions, please let me know!	Thank you for choosing options B (Cholinergic urticaria) and D (Anaphylaxis). Based on these selections, I'll offer variables for each element that significantly impact the correct answer considering the chosen options. Here are the proposed variables: Element 1: Timing of symptoms • During her physical education class after lunch. • Immediately after taking a hot shower. • While sitting in a hot classroom. • After exposure to cold wind post-exercise. • During a stressful exam. Element 2: Symptoms • Coughing, shortness of breath, dizziness, facial swelling and redness. • Generalized hives, sweating, and palpitations • Localized itching and redness following sweating. • Throat tightness, difficulty swallowing, and a sense of impending doom. • Flushing, headache, and nausea after eating specific foods. Element 3: Clinical findings • A blood pressure reading of 80/50 mm Hg, periorbital edema, and wheezing upon auscultation. • Tachycardia with a blood pressure reading of 100/70 mmHg, and diffuse hives. • Normal blood pressure, with localized swelling and erythema. • Rapid drop in blood pressure upon exposure to an allergen, with angioedema. • Elevated blood pressure due to anxiety, with mild facial edema. Please review these variables and let me know if they are appropriate or if you would like any modifications.

[a]MCQ: multiple-choice question.

Step 5: Generating the Cognitive Model

In the following step, AI determines the constraints based on values (called "variables" by AI) approved by the expert, in order to create the cognitive model. Then, the expert reviews the output and prompts AI for necessary changes, if any. The presence of the expert is to validate the output, as in the previous steps.

Compared with the cognitive model in the traditional template-based method, our hybrid method adopts a simpler approach. By providing constraints, it allows to determine what items will be generated and what the correct option is in each item. Tables 4 and 5 present the cognitive models. Although the content needs small adjustments that can be easily carried out by the expert (eg, the expert may ask ChatGPT to remove "37.5°C to 38 °C" to make the correct answers more evident), we did not make any changes to present ChatGPT's original output in order for demonstrating that it is able to provide a useful cognitive model to begin with for further revisions. It is also possible due to the nature of GPT models that the output might defer in formatting as visible in the tables.

XSL•FO
RenderX

Table . Cognitive model generated by artificial intelligence (AI) for Gierl et al.'s [3] item.

Element (variable)	Constraints (supported options)
Cough	• Persistent dry cough (seasonal Influenza) • Mild cough (common cold) • Severe hacking cough (seasonal influenza)
Body aches	• Slight body aches (common cold) • Moderate body aches (seasonal influenza) • Severe muscle pains (seasonal influenza)
Duration of symptoms	• Suddenly over a few hours (seasonal influenza) • Over a few days (common cold) • Gradually over more than a week (common cold)
Oral temperature	• 37°C (normal) (common cold) • 37.5°C to 38°C (mild fever) (seasonal influenza) • Over 38°C (fever) (seasonal influenza)

Table . Cognitive model generated by artificial intelligence (AI) for TUS item.

Variable element	Constraints (supported options)
Timing of symptoms	B and D
During her physical education class after lunch.	D
Immediately after taking a hot shower.	B
While sitting in a hot classroom.	B
After exposure to cold wind postexercise.	D
During a stressful exam.	B
Symptoms	B and D
Coughing, shortness of breath, dizziness, and facial swelling and redness.	D
Generalized hives, sweating, and palpitations.	B and D
Localized itching and redness following sweating.	B
Throat tightness, difficulty swallowing, and a sense of impending doom.	D
Flushing, headache, and nausea after eating specific foods.	D
Clinical findings	B and D
A blood pressure reading of 80/50 mmHg, periorbital edema, and wheezing upon auscultation.	D
Tachycardia with a blood pressure reading of 100/70 mmHg, and diffuse hives.	B and D
Normal blood pressure, with localized swelling and erythema.	B
Rapid drop in blood pressure upon exposure to an allergen, with angioedema.	D
Elevated blood pressure due to anxiety, with mild facial edema.	B

These 5 steps have been completed in less than 10 minutes for each model. The whole process within the AI environment can be displayed by accessing the following public pages of the human-AI conversations: Gierl et al's [3] item [34] and TUS item [35].

Step 6: Final Review by the Expert(s)

During this phase, the expert carries out a final review of the item model and cognitive model provided by AI, preferably with other experts.

In the previous steps, in order to keep the demonstration simple and due to the fact that there was not a significant inaccuracy in the AI-generated content, the expert did not demand any additional changes during the process in AI environment. But in the hybrid method, the 5 steps within the AI environment should be actively monitored by the expert, and if necessary, the expert should input prompts to make changes because AI is always prone to provide inaccurate content and deviate from providing a consistent template. Expert oversight, and involvement if necessary, is a strong necessity for creating high-quality item models and cognitive models.

XSL•FO
RenderX

Following the first 5 steps, which can be completed in less than 10 minutes, the expert should carry out one more round of review for the item model and the cognitive model generated through human-AI collaboration. It would be better if the expert conducts this review together with other experts to make sure that there is no inaccuracy, inconsistency, or inappropriate way of presentation. The expert should keep in mind that content generated by AI, in this case ChatGPT, is generated through a large language model, so it could "hallucinate" [36] some inaccurate information and relationships that are harmful for the output quality. Apart from that, in this step, the expert may prefer to add more elements and variables, such as age and gender, in a way that does not change the correct answers, in order for increasing the number of the items.

Step 7: Item Generation Using a Non-AI Software

Finally, the expert inputs the final version of the item template and the constraints to a traditional template-based AIG tool (software without AI involvement), and then the software algorithmically produces multiple-choice items based on the input provided by the expert. It is crucial to emphasize that the expert must input the content accurately, as traditional software cannot handle inconsistent type of inputs, unlike AI in the previous steps. There is no difference between the traditional template-based method (stage 3) [10] and our hybrid method (step 7) in this regard. As is in the template-based method [10], Hybrid AIG also allows the software to generate hundreds of consistent items based on a single item model.

Discussion

In this study, we used AI to generate item models and cognitive models for generating multiple-choice items by using these models for template-based AIG. We demonstrated that AI is capable of providing AIG templates for this purpose under the control of human expert. Leveraging AI in template development has significantly reduced the time investment from hours [10] to less than 10 minutes, and provided a smoother experience for experts in this challenging task [3].

In our hybrid AIG method, cognitive work required to be carried out by experts in the past [10] is now shared with AI. It switches the role of experts from "the creators of item-cognitive models from scratch" to "the reviewer of AI-generated content," which brings an important efficiency to AIG without sacrificing consistency and accuracy. Our hybrid AIG method transcends the traditional template-based approach by marrying the "art" that comes from AI as a "black box" [25] with the "science" of algorithmic generation [10] under the oversight of expert as a "marriage registrar." Practically, this balanced fusion under human guidance reduces the extraneous cognitive load [16] on experts by allocating the burdensome tasks to AI in order for enhancing human efficiency and allowing them to concentrate on refining and validating the AI-generated content.

Similar to our approach, a recent study successfully incorporated a large language model into the process of developing reading comprehension items [37]. While addressing a critical issue in item development for a non–health care setting, its direct application to medical education is challenging due to the inherent complexities of health professions education. Furthermore, this approach integrates AI only into generating unique sentences based on rules imposed by experts, leaving the essential cognitive work dependent on expert input, which remains inefficient for medical education. In our hybrid method, we use AI not only for generating unique sentences but also for development of item models and cognitive models as a whole, hence transforming the role of experts from the main "cognitive workers" to reviewers. This shift reduces cognitive effort for experts while maintaining their essential contribution for accurate and consistent items. Considering the importance of clarity and constraints in the templates, we still can suggest that it is possible and desirable to create specifications and instructions using artificial intelligence. Our research demonstrated that even a minimal human oversight can be sufficient for using AI in the creation of specifications and instructions, particularly in challenging domains such as medical education, which suggests even greater possibility for less complex tasks like reading comprehension. By dismissing the potential of AI in this regard by labeling it as "impossible," humans might inadvertently limit AI's capacity to enhance efficiency in cognitive work needed to be done. Thus, we propose leveraging AI more effectively rather than relegating it to a lesser role.

While noting the improved efficiency of the proposed hybrid for cognitive tasks, we emphasize the importance of rigorous human oversight, and consequently, accountability for automatically generated content. As demonstrated in a study by Zack et al [38], even a state of the art large language model can still be prone to perpetuate racial and gender bias. Adding to the complexity, such bias may not be visible at the level of a single question item or template, but as an effect of prolonged use of generative AI tools. The role of human examiners is to be aware of such risks and to implement bias mitigation strategies at different steps and levels of the AI process to prevent the injection of harmful stereotypes into the assessment of students' skills.

Our study has some limitations. Although the templates generated by AI showed promising results, replicability depends on the consistency of the AI model, which is GPT-4 in this case. In addition, other AI models such as GPT-4o (Open AI), Claude (Anthropic), Gemini (Google), Llama (Meta), and Command *R*+ (Cohere) could lead to different outputs. While our study demonstrated that a hybrid AIG is possible, future research should explore this further by using different parent items across various settings to generate MCQs. As this is a proof-of-concept study, there is a lack of empirical evidence supporting the efficacy of the proposed hybrid AIG method, no qualitative reviews to assess the generated items' quality, and a lack of quantitative item analysis since the items were not tested on medical students. However, it is still valuable because it has shown for the first time that generating plausible, and possibly useful, item templates using AI is possible in medical education. A recent study has provided empirical evidence, demonstrating that experts correctly identified the answers in MCQs generated by using hybrid AIG [39]. In the future studies, we are planning to generate more items using these templates and investigate their effectiveness using qualitative and quantitative methods.

XSL•FO
RenderX

Moreover, a direct comparison of traditional template-based AIG and hybrid AIG could provide valuable evidence for effectiveness and efficiency. Another limitation is that we generated simple templates. There are multilayered templates for AIG [3], which require relatively complex structures, that might require from us to use different custom AIs for this purpose.

In conclusion, the hybrid AIG is a promising novel method that leverages AI in development of templates for template-based AIG that transforms the traditional role of experts from creators to reviewers. This shift can significantly reduce the cognitive burden on experts and streamline the item generation process while ensuring high-quality outcomes. We recommend piloting and improving the hybrid AIG in high-demand settings of increasing importance to investigate and improve its efficiency and quality benefits.

Acknowledgments

This work has been supported by TÜBİTAK (The Scientific and Technological Research Council of Turkiye) under the 2219 program for YSK and internal research funds at Jagiellonian University Medical College under the N41/DBS/000332 program for AK. Generative AI, as a research tool, was used in the preparation of this manuscript for multiple-choice question template generation, which has been transparently reported in the methods and results section of the article. In addition, generative AI was used for language correction. The authors have reviewed, edited, and verified all content. The authors take full responsibility for the integrity and validity of the final work.

Data Availability

All data generated or analyzed during this study are included in this published article.

Authors' Contributions

YSK contibuted to conceptualization, methodology, writing the original draft, data curation, writing-review and editing, and funding acquisition. AK contributed to conceptualization, methodology, writing-review and editing, and funding acquisition.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The prompt that has been used in the custom GPT.
[DOCX File, 16 KB - formative_v9i1e65726_app1.docx]

References

1. Pugh D, De Champlain A, Touchie C. Plus ça change, plus c'est pareil: Making a continued case for the use of MCQs in medical education. Med Teach 2019 May 4;41(5):569-577. [doi: 10.1080/0142159X.2018.1505035]
2. Wrigley W, van der Vleuten CPM, Freeman A, Muijtjens A. A systemic framework for the progress test: strengths, constraints and issues: AMEE Guide No. 71. Med Teach 2012;34(9):683-697. [doi: 10.3109/0142159X.2012.704437] [Medline: 22905655]
3. Gierl MJ, Lai H, Tanygin V. Advanced Methods in Automatic Item Generation, 1st edition: Routledge; 2021. [doi: 10.4324/9781003025634-1]
4. Kiyak YS, Budakoğlu I, Coşkun Ö, Koyun E. The first automatic item generation in Turkish for assessment of clinical reasoning in medical education. Tıp Eğitimi Dünyası 2023 Mar;22(66):72-90. [doi: 10.25282/ted.1225814]
5. Kiyak YS, Coşkun Ö, Budakoğlu I, Uluoğlu C. Psychometric analysis of the first Turkish multiple-choice questions generated using automatic item generation method in medical education. Tıp Eğitimi Dünyası 2023;22(68):154-161. [doi: 10.25282/ted.1376840]
6. Kıyak YS, Kononowicz AA, Górski S. Multilingual template-based automatic item generation for medical education supported by generative artificial intelligence models ChatGPT and Claude. Bio-Algorithms Med-Syst 2024 Dec 23;20(1):81-89. [doi: 10.5604/01.3001.0054.9192]
7. Leslie T, Gierl MJ. Using automatic item generation to create multiple-choice questions for pharmacy assessment. Am J Pharm Educ 2023 Oct;87(10):100081. [doi: 10.1016/j.ajpe.2023.100081] [Medline: 37852684]
8. Lai H, Gierl MJ, Byrne BE, Spielman AI, Waldschmidt DM. Three modeling applications to promote automatic item generation for examinations in dentistry. J Dent Educ 2016 Mar;80(3):339-347. [doi: 10.1002/j.0022-0337.2016.80.3.tb06090.x] [Medline: 26933110]
9. Falcão F, Costa P, Pêgo JM. Feasibility assurance: a review of automatic item generation in medical assessment. Adv in Health Sci Educ 2022 May;27(2):405-425. [doi: 10.1007/s10459-022-10092-z]

XSL•FO
RenderX

10. Gierl MJ, Lai H, Turner SR. Using automatic item generation to create multiple-choice test items. Med Educ 2012 Aug;46(8):757-765. [doi: 10.1111/j.1365-2923.2012.04289.x] [Medline: 22803753]
11. Kosh AE, Simpson MA, Bickel L, Kellogg M, Sanford‐Moore E. A cost–benefit analysis of automatic item generation. Educational Measurement 2019 Mar;38(1):48-53 [FREE Full text] [doi: 10.1111/emip.12237]
12. Pugh D, De Champlain A, Gierl M, Lai H, Touchie C. Using cognitive models to develop quality multiple-choice questions. Med Teach 2016 Aug;38(8):838-843. [doi: 10.3109/0142159X.2016.1150989] [Medline: 26998566]
13. Gierl MJ, Lai H, Pugh D, Touchie C, Boulais AP, De Champlain A. Evaluating the psychometric characteristics of generated multiple-choice test items. Applied Measurement in Education 2016 Jul 2;29(3):196-210. [doi: 10.1080/08957347.2016.1171768]
14. Pugh D, De Champlain A, Gierl M, Lai H, Touchie C. Can automated item generation be used to develop high quality MCQs that assess application of knowledge? RPTEL 2020 Dec;15(1):12. [doi: 10.1186/s41039-020-00134-8]
15. Gierl MJ, Lai H. Evaluating the quality of medical multiple-choice items created with automated processes. Med Educ 2013 Jul;47(7):726-733. [doi: 10.1111/medu.12202] [Medline: 23746162]
16. van Merriënboer JJG, Sweller J. Cognitive load theory in health professional education: design principles and strategies. Med Educ 2010 Jan;44(1):85-93. [doi: 10.1111/j.1365-2923.2009.03498.x] [Medline: 20078759]
17. Kıyak YS. A ChatGPT prompt for writing case-based multiple-choice questions. Rev Esp Edu Med 2023;4(3):98-103. [doi: 10.6018/edumed.587451]
18. Zuckerman M, Flood R, Tan RJB, et al. ChatGPT for assessment writing. Med Teach 2023 Nov;45(11):1224-1227. [doi: 10.1080/0142159X.2023.2249239] [Medline: 37789636]
19. Laupichler MC, Rother JF, Grunwald Kadow IC, Ahmadi S, Raupach T. Large language models in medical education: comparing ChatGPT- to human-generated exam questions. Acad Med 2024 May 1;99(5):508-512. [doi: 10.1097/ACM.0000000000005626] [Medline: 38166323]
20. Kıyak YS, Kononowicz AA. Case-based MCQ generator: A custom ChatGPT based on published prompts in the literature for automatic item generation. Med Teach 2024 Aug;46(8):1018-1020. [doi: 10.1080/0142159X.2024.2314723] [Medline: 38340312]
21. Kıyak YS, Coşkun Ö, Budakoğlu I, Uluoğlu C. ChatGPT for generating multiple-choice questions: Evidence on the use of artificial intelligence in automatic item generation for a rational pharmacotherapy exam. Eur J Clin Pharmacol 2024 May;80(5):729-735. [doi: 10.1007/s00228-024-03649-x] [Medline: 38353690]
22. Indran IR, Paranthaman P, Gupta N, Mustafa N. Twelve tips to leverage AI for efficient and effective medical question generation: A guide for educators using Chat GPT. Med Teach 2024 Aug;46(8):1021-1026. [doi: 10.1080/0142159X.2023.2294703] [Medline: 38146711]
23. Ngo A, Gupta S, Perrine O, Reddy R, Ershadi S, Remick D. ChatGPT 3.5 fails to write appropriate multiple choice practice exam questions. Acad Pathol 2024;11(1):100099. [doi: 10.1016/j.acpath.2023.100099] [Medline: 38162414]
24. Lee H. The rise of ChatGPT: exploring its potential in medical education. Anatomical Sciences Ed 2024;17(5):926-931. [doi: 10.1002/ase.2270]
25. von Eschenbach WJ. Transparency and the black box problem: why we do not trust AI. Philos Technol 2021 Dec;34(4):1607-1622. [doi: 10.1007/s13347-021-00477-0]
26. Bulut O, Beiting-Parrish M, Casabianca JM. The rise of artificial intelligence in educational measurement: opportunities and ethical challenges. arXiv. Preprint posted online on Jun 27, 2024. [doi: 10.48550/arXiv.2406.18900]
27. Han Z, Battaglia F, Udaiyar A, Fooks A, Terlecky SR. An explorative assessment of ChatGPT as an aid in medical education: Use it with caution. Med Teach 2024 May;46(5):657-664. [doi: 10.1080/0142159X.2023.2271159] [Medline: 37862566]
28. Appelbaum SH. Socio‐technical systems theory: an intervention strategy for organizational development. Manag Decis 1997 Aug;35(6):452-463. [doi: 10.1108/00251749710173823]
29. Mesko B. The ChatGPT (generative artificial intelligence) revolution has made artificial intelligence approachable for medical professionals. J Med Internet Res 2023 Jun 22;25:e48392. [doi: 10.2196/48392] [Medline: 37347508]
30. White J, Fu Q, Hays S, et al. A prompt pattern catalog to enhance prompt engineering with chatgpt. arXiv. Preprint posted online on Feb 21, 2023. [doi: 10.48550/arXiv.2302.11382]
31. Kıyak YS. Beginner-level tips for medical educators: guidance on selection, prompt engineering, and the use of artificial intelligence chatbots. Med Sci Educ 2024 Dec;34(6):1571-1576. [doi: 10.1007/s40670-024-02146-1] [Medline: 39758489]
32. Masters K, Benjamin J, Agrawal A, MacNeill H, Pillow MT, Mehta N. Twelve tips on creating and using custom GPTs to enhance health professions education. Med Teach 2024 Jun;46(6):752-756. [doi: 10.1080/0142159X.2024.2305365] [Medline: 38285894]
33. Item Model Maker for AIG. ChatGPT. URL: https://chatgpt.com/g/g-ISoiQOLyv-item-model-maker-for-aig [accessed 2025-04-01]
34. Item model 1. ChatGPT. URL: https://chatgpt.com/share/1c975655-6bc4-4a16-a082-efad2030a5e2 [accessed 2025-04-01]
35. Item model 2. ChatGPT. URL: https://chatgpt.com/share/7e093fef-8d43-4b8f-bd15-a00bdf4c25c5 [accessed 2025-04-01]
36. Masters K. Medical Teacher ’s first ChatGPT’s referencing hallucinations: lessons for editors, reviewers, and teachers. Med Teach 2023 Jul;45(7):673-675. [doi: 10.1080/0142159X.2023.2208731] [Medline: 37183932]

XSL•FO
RenderX

37. Sayin A, Gierl M. Using OpenAI GPT to generate reading comprehension items. Educational Measurement 2024 Mar;43(1):5-18 [FREE Full text] [doi: 10.1111/emip.12590]
38. Zack T, Lehman E, Suzgun M, et al. Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study. Lancet Digit Health 2024 Jan;6(1):e12-e22. [doi: 10.1016/S2589-7500(23)00225-X] [Medline: 38123252]
39. Kıyak YS, Emekli E, Coşkun Ö, Budakoğlu I. Keeping humans in the loop efficiently by generating question templates instead of questions using AI: Validity evidence on Hybrid AIG. Med Teach 2025 Apr;47(4):744-747. [doi: 10.1080/0142159X.2024.2430360] [Medline: 39604119]

Abbreviations

AI: artificial intelligence
AIG: automatic item generation
MCQ: multiple-choice question

Edited by A Mavragani; submitted 23.08.24; peer-reviewed by A Montazerolghaem, M Alshiekh; revised version received 16.03.25; accepted 18.03.25; published 04.04.25.

Please cite as:
Kıyak YS, Kononowicz AA
Using a Hybrid of AI and Template-Based Method in Automatic Item Generation to Create Multiple-Choice Questions in Medical Education: Hybrid AIG
JMIR Form Res 2025;9:e65726
URL: https://formative.jmir.org/2025/1/e65726
doi:10.2196/65726

© Yavuz Selim Kıyak, Andrzej A Kononowicz. Originally published in JMIR Formative Research (https://formative.jmir.org), 4.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Using Large Language Models to Automate Data Extraction From Surgical Pathology Reports: Retrospective Cohort Study

Denise Lee[1*], MD; Akhil Vaid[2*], MD; Kartikeya M Menon[1], BA; Robert Freeman[2], DNP, RN; David S Matteson[3], PhD; Michael L Marin[1], MD; Girish N Nadkarni[2], MD, MPH

[1]Department of Surgery, Icahn School of Medicine at Mount Sinai, 10 Union Square East, Suite 2L, New York, NY, United States

[2]Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[3]Department of Statistics and Data Science, Cornell University, Ithaca, NY, United States

[*]these authors contributed equally

Corresponding Author:
Denise Lee, MD
Department of Surgery, Icahn School of Medicine at Mount Sinai, 10 Union Square East, Suite 2L, New York, NY, United States

Abstract

Background: Popularized by ChatGPT, large language models (LLMs) are poised to transform the scalability of clinical natural language processing (NLP) downstream tasks such as medical question answering (MQA) and automated data extraction from clinical narrative reports. However, the use of LLMs in the health care setting is limited by cost, computing power, and patient privacy concerns. Specifically, as interest in LLM-based clinical applications grows, regulatory safeguards must be established to avoid exposure of patient data through the public domain. The use of open-source LLMs deployed behind institutional firewalls may ensure the protection of private patient data. In this study, we evaluated the extraction performance of a locally deployed LLM for automated MQA from surgical pathology reports.

Objective: We compared the performance of human reviewers and a locally deployed LLM tasked with extracting key histologic and staging information from surgical pathology reports.

Methods: A total of 84 thyroid cancer surgical pathology reports were assessed by two independent reviewers and the open-source FastChat-T5 3B-parameter LLM using institutional computing resources. Longer text reports were split into 1200-character-long segments, followed by conversion to embeddings. Three segments with the highest similarity scores were integrated to create the final context for the LLM. The context was then made part of the question it was directed to answer. Twelve medical questions for staging and thyroid cancer recurrence risk data extraction were formulated and answered for each report. The time to respond and concordance of answers were evaluated. The concordance rate for each pairwise comparison (human-LLM and human-human) was calculated as the total number of concordant answers divided by the total number of answers for each of the 12 questions. The average concordance rate and associated error of all questions were tabulated for each pairwise comparison and evaluated with two-sided *t* tests.

Results: Out of a total of 1008 questions answered, reviewers 1 and 2 had an average (SD) concordance rate of responses of 99% (1%; 999/1008 responses). The LLM was concordant with reviewers 1 and 2 at an overall average (SD) rate of 89% (7%; 896/1008 responses) and 89% (7.2%; 903/1008 responses). The overall time to review and answer questions for all reports was 170.7, 115, and 19.56 minutes for Reviewers 1, 2, and the LLM, respectively.

Conclusions: The locally deployed LLM can be used for MQA with considerable time-saving and acceptable accuracy in responses. Prompt engineering and fine-tuning may further augment automated data extraction from clinical narratives for the provision of real-time, essential clinical insights.

(JMIR Form Res 2025;9:e64544) doi:10.2196/64544

KEYWORDS

natural language processing; large language model; artificial intelligence; thyroid cancer; endocrine surgery; framework; privacy; medical; surgical pathology; report; NLP; medical question

Introduction

Surgical pathology reports contain narrative data essential for hospital-based comprehensive cancer surveillance databases and the real-time understanding of patient staging, recurrence risk, clinical trial eligibility, and personalized treatment options. Large-scale national databases such as the Surveillance, Epidemiology, and End Results (SEER) program and the National Cancer Database (NCDB) include population-level histologic data necessary to study rare cancers and

XSL•FO
RenderX

underrepresented patient subpopulations [1,2]. However, the large-scale extraction of key surgical oncologic insights contained within unstructured free pathology text is limited by the need for labor-intensive, manual review by human reviewers [3]. The maintenance of comprehensive cancer databases often requires skilled registrars who are trained to synthesize and extract information from patient narrative reports according to established data collection standards [4]. However, constantly updated guidelines, time-intensive labor for data extraction, and inevitable human error are challenges to coordinate cancer data collection.

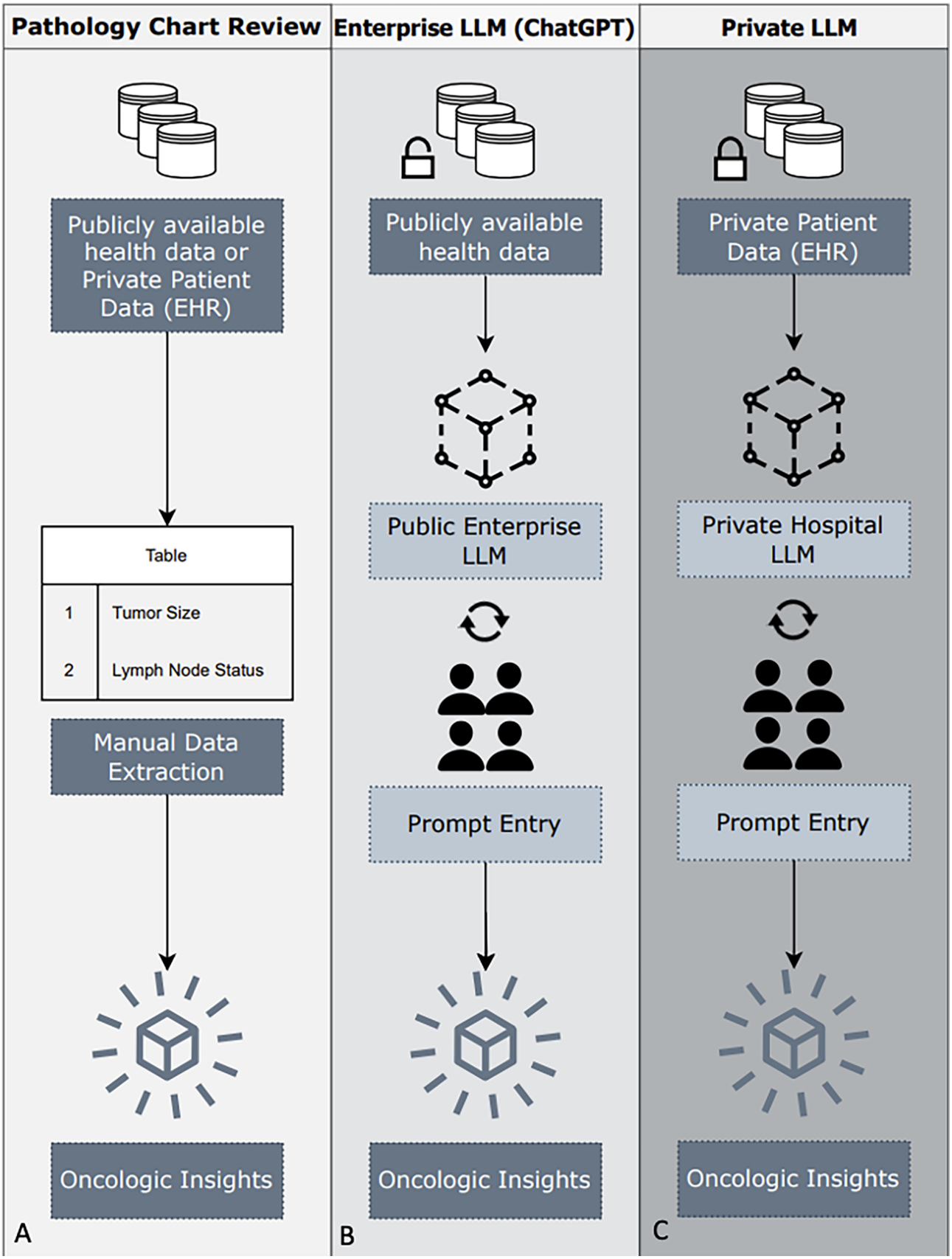
Advances in natural language processing (NLP) techniques have sought to address challenges in the efficiency and accuracy of data abstraction. Efforts have included applying NLP methods to extract pain scores in patients with cancer undergoing radiation [5], classify metastatic phenotypes from radiology reports of patients with colorectal cancer [6], and identify recurrence status in patients with hepatocellular carcinoma [7]. Traditional NLP methods, however, require specific domain expertise and may be error-prone based on statistical or rule-based approaches [8]. Large language models (LLMs) power a new generation of natural language processing (NLP) whereby deep neural networks are trained on a massive corpus of human text that are then deconstructed into vectorized embeddings that depict linguistic relationships in a numerical format appropriate for easy analysis [9]. Popularized by ChatGPT and its user-friendly question-and-answer interface, LLMs are poised to transform the scalability of clinical NLP downstream tasks such as medical question answering (MQA) and data mining. LLMs may also enhance the ability to rapidly and accurately extract key information from surgical pathology reports [10].

However, as new LLM-based clinical applications are being increasingly explored, there is a substantial risk of protected health information (PHI) breaches without regulatory safeguards in place [11]. Ethical, privacy, and regulatory constraints preclude the transfer of PHI across the public domain through widely used proprietary LLM services (eg, ChatGPT, Gemini, and Claude) that can generate automated responses for MQA. However, significant health care resources (ie, utilization fees and institutional agreements) would be required to ensure the protection of PHI while accessing the largest state-of-the-art models such as GPT4 via a private cloud service provider (eg, Microsoft Azure OpenAI). Such considerable upfront costs and service agreements may preclude hospital systems without financial and IT resources from utilizing the largest and most state-of-the-art LLMs.

The rapid development in LLM technology has led to the development of many open-source models (eg, Llama 3.1 and Google Gemini) both comparable in performance to GPT4 and deployable on an institution's existing computing clusters. The use of a smaller, locally deployed LLM may help ensure that protected health information is maintained and democratize access to LLM technologies at medical centers where larger, private LLMs are inaccessible. In this study, we investigated the use of a locally deployed, open-source LLM to extract key staging and recurrence risk information from thyroid surgical pathology reports. We compared its performance to the gold standard of medically trained human reviewers. Although we utilize this for a single-use case, our work serves as a practical pathway to enable individual medical centers to utilize current LLM technology for clinical NLP tasks in a privacy-protecting manner (Figure 1).

XSL•FO
RenderX

Figure 1. Overview of medical data extraction workflows: (A) Pathology chart review: The traditional approach of manual data extraction from publicly available databases or private electronic health records (EHRs) to obtain predetermined oncologic insights. (B) Enterprise large language model (LLM, such as ChatGPT): Due to regulatory constraints, only publicly available data may be shared with enterprise LLMs. Prompt entry and question curation are used to gain oncologic insights. (C) Private LLMs: EHR data can be shared with a local hospital LLM, and prompt entry with question curation can be used to gain oncologic insights. EHR: electronic health record.



XSL•FO
RenderX

Methods

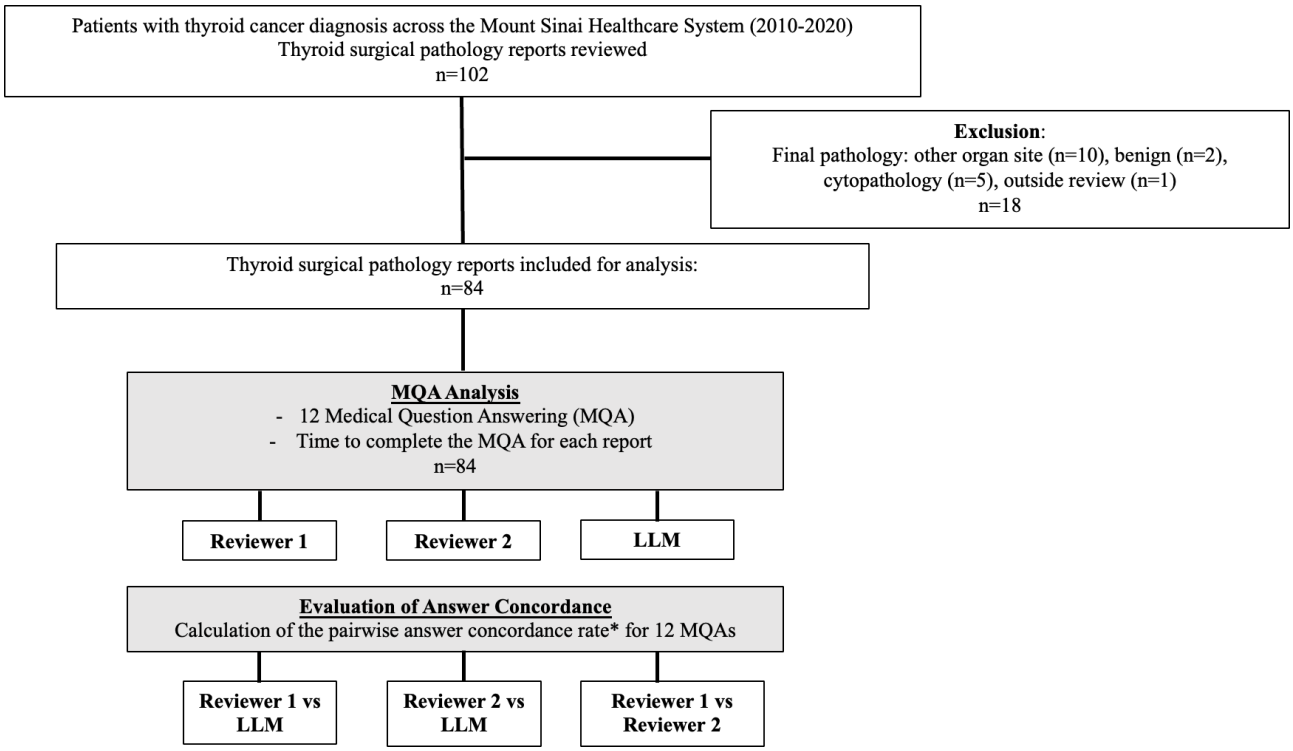
Ethical Considerations

This study was approved by the Institutional Review Board of the Icahn School of Medicine at Mount Sinai (no. 22 - 00347). Informed consent was waived given the retrospective nature of the study. To ensure confidentiality of patient data, only authorized research study personnel were permitted access to patient data. Study identification numbers were assigned to patients included in the study for deidentification.

Study Population

We queried the Mount Sinai Data Warehouse for a cohort of adult patients (>18 years old) with *ICD-9* (International Classification of Diseases, 9th revision) and *ICD-10* (International Classification of Diseases, 10th revision) diagnosis codes for thyroid cancer and who underwent thyroid surgery between 2010 and 2022. We reviewed 102 surgical pathology reports from 102 patients and excluded reports if they were from other organ sites (n=10), benign (n=2), cytopathology (n=5), or outside review (n=1). We included 84 reports for analysis. The study flowchart is shown in Figure 2.

Figure 2. Flowchart of the study design and analysis. *The concordance rate was calculated as the total number of concordant answers/total number of answers for each of the 12 medical question answering (MQA).



Large Language Model

We used the publicly available, open-source, and parameter-efficient FastChat-T5 3B-parameter LLM for our analysis. FastChat-T5 is a smaller model that requires significantly less computing, making it accessible in resource-constrained environments. This characteristic was essential for our study, as we wanted our findings to be easily replicable across institutions with varying levels of computational resources. In many real-world scenarios, particularly in health care or academic settings, access to high-end hardware or expensive cloud services is limited, and the need for locally deployable solutions behind firewalls is critical due to privacy concerns. FastChat-T5 provides a balance between performance and resource efficiency, making it a practical choice in these contexts. This LLM was deployed on existing hospital servers, which avoided the utilization of expensive cloud services and external transfer of patient data. The ready availability of such software has increased the use of open-source LLMs. The practice of deploying smaller models on local hardware aligns with the latest trends from leading LLM providers that are exploring the deployment of models on mobile devices.

In addition, our goal was to evaluate how a smaller, open-source model performs in comparison to human reviewers, who are generally considered the gold standard in tasks such as document review and content evaluation. We selected FastChat-T5 to explore whether such a parameter-efficient model could offer comparable performance to humans, without requiring highly specialized code or expensive infrastructure for running more recent and resource-intensive models, such as those in the LLaMA (large language Meta AI) family.

A limitation all LLMs have is the amount of context they may process at once, and for FastChat-T5, this limit is 2048 tokens. For reports of lengths greater than what the model could accommodate, we split the report text into 1200-character-long segments, followed by converting each of these segments into machine-readable numerical representations called embeddings. As embeddings encode meaning, calculating similarity scores using the cosine similarity metric between segment embeddings and posed questions allowed us to retrieve the pieces of text

XSL•FO
RenderX

most directly related to the content of the question. As such, three segments with the highest similarity scores were integrated to create the final context for the LLM. This context was made part of a plain language question for the LLM, alongside a question it was directed to answer.

We selected the top three segments with the highest similarity scores for several key reasons. First, using multiple segments allows us to enrich the context provided to the LLM, ensuring that it has access to a broader and more diverse set of relevant information. By incorporating three segments, we increase the likelihood that important details from different parts of the document are included, without overwhelming the model's context window.

Additionally, focusing on the top three ensures that we strike a balance between the relevance and the diversity of information. Using just one segment might omit crucial context, while using too many could dilute the focus on the most pertinent details. The number three was chosen empirically, as it allowed for a good compromise between feeding the LLM with enough useful content and avoiding the risk of both exceeding usable context and introducing unrelated or redundant information that could confound the model's operation. We did experiment with other segmentation strategies, including combining two or four closest segments, but the performance gains were inferior to those of the three-segment approach. In our preliminary trials, while combining more distinct segments could introduce variety, this approach led to a slight reduction in overall accuracy due to the introduction of less relevant or tangential information.

Finally, embedding models with cosine similarity offers an efficient and robust way to evaluate how closely each segment relates to the posed question, ensuring that the selected segments are not just textually similar but semantically aligned with the user's query. This method enhances the precision of the context provided to the LLM, resulting in more accurate and relevant answers. The source code can be found at [12] under a GPLv3 license.

The prompt template is available in Multimedia Appendix 1.

Development of MQA and the Evaluation of Concordance

We formulated 12 questions with expert clinical input that extracted key information for the assessment of the thyroid cancer staging and recurrence risk on the AJCC-TNM staging system (American Joint Committee on Cancer Tumor-Node-Metastasis, 8th edition) according to the American Thyroid Association Recurrence Risk Stratification System [13,14]. Two study authors (DL and KMM) reviewed 84 thyroid surgical pathology reports and recorded answers to each of the 12 questions and the time taken to complete the answers for each report. Then, we used the LLM to answer the same questions. For every question, we determined if the answers were concordant between Reviewer 1 and the LLM, between Reviewer 2 and the LLM, and between the two reviewers. The concordance rate for each pairwise comparison (human-LLM and human-human) was then calculated as the total number of concordant answers divided by the total number of answers for each of the 12 questions (Figure 2). The average concordance rate and associated error (calculated here as the SD) for all questions were tabulated for each pairwise comparison and evaluated with two-sided *t* tests.

Results

We report the sample LLM responses and the concordance rates between reviewers and the LLM for each question in Table 1. A total of 1008 questions were answered for the 84 thyroid surgical pathology reports. Reviewers 1 and 2 were concordant at an average (SD) overall rate of 99% (1%; 999 out of 1008 questions) with disagreement on 9 answers. Reviewers 1 and 2 took an average (SD) of 2 (0.6) minutes and 1.4 (0.4) minutes to respond to each pathology report and a total of 170.7 minutes and 115 minutes for all reports, respectively. The LLM was concordant with reviewers 1 and 2 at an average (SD) overall rate of 89% (7%; 896 out of 1008 questions) and 89 (7.2%; 903 out of 1008 questions), respectively. The LLM was significantly less accurate than both reviewer 1 ($P<.001$) and reviewer 2 ($P<.001$), but performed at the same level between both reviewers ($P=.82$). The average time to review each report for the LLM was 13.97 seconds/report and 19.56 minutes for all reports. The questions with the highest overall rates of concordant responses were questions requiring binary or categorical data extraction (eg, "Is lymphatic invasion present?", 100%; "Is vascular invasion present?", 98.81%; and "Where is the primary cancer located?", 98.1%). The question with the lowest overall concordance was, "Were cervical lymph nodes present?" at 75%.

XSL•FO
RenderX

Table . Twelve medical questions and sample large language model (LLM) answers with the rates of concordance between the reviewers and the LLM.

Medical question	Sample LLM answer	Concordance (%)			
		R1,[a] LLM	R2,[b] LLM	R1, R2	All 3
Staging tumor (T)					
What is the size of the primary tumor?	The size of the primary tumor was 5.5 cm.	82	84	98	82
Does the tumor extend beyond the capsule of the thyroid?	No, the tumor does not extend beyond the capsule of the thyroid.	90	93	97	90
Where is the primary cancer located (right thyroid lobe, left thyroid lobe, or isthmus)?	The primary cancer is located in the right thyroid lobe.	99	100	99	99
Node (N)					
Were any cervical lymph nodes present?	Yes, cervical lymph nodes were present.	76	100	99	75
If cervical lymph nodes are present, how many lymph nodes were positive for malignancy?	Three lymph nodes were positive for malignancy.	88	88	100	88
Stage (TNM)[c]					
What is the final pathology TNM stage of the thyroid cancer?	The final pathology TNM stage of the thyroid cancer is pT3a.	85	85	100	85
Recurrence risk factor					
Is vascular invasion present?	Yes, vascular invasion is present.	99	99	100	99
Is lymphatic invasion present?	No, lymphatic invasion is not present.	100	100	100	100
What was the histology of the primary cancer?	The histology of the primary thyroid cancer was papillary carcinoma.	87	87	100	87
Was there a variant of thyroid cancer such as tall cell, hobnail variant, or columnar cell variant present, indicative of a more aggressive histology?	No, a variant of thyroid cancer such as tall cell, hobnail variant, or columnar cell variant was not present indicative of a more aggressive histology.	89	92	97	89
Was there a second thyroid cancer present?	No, there was no second thyroid cancer present.	82	83	99	82
If there was a second thyroid cancer present, what was the histology?	The histology of the second thyroid cancer was classical variant papillary thyroid carcinoma.	88	88	99	88
Overall questions, % (SD)		89 (7)	89 (7)	99 (1)	89 (7.2)

[a]R1: Reviewer 1.

[b]R2: Reviewer 2.

[c]TNM: Tumor-Node-Metastasis.

XSL•FO
RenderX

Discussion

We demonstrated and evaluated the extraction performance of a locally deployed, open-source LLM for a specific clinical NLP task. To our knowledge, this is the first study to compare an LLM's ability to extract key thyroid cancer recurrence and staging information accurately and efficiently from surgical pathology reports using a conversational MQA format. The LLM took 19.56 minutes to evaluate and respond to all pathology reports, whereas it required an additional 187 minutes and 105 minutes for the reviewers to complete the same task—demonstrating a considerable reduction in time. Regarding the accuracy of the responses, we found that the rates of response concordance were higher among questions tasked with simpler binary or categorical responses. The increase in task complexity requiring textual interpretation and inconsistent word prompting, such as asking whether there was "cervical" lymph nodes present, resulted in the lowest rate of concordance. Furthermore, the question regarding the size of the primary tumor also seemed to be relatively straightforward but had an overall concordance rate of only 82%.

The augmentation of poorer-performing MQA may lie in the improvement of prompt engineering—an emerging subfield where domain-specific knowledge and linguistics are optimized to design questions that yield the best-performing response to a task. In addition, more expressive embeddings may help localize relevant text [15]. For example, "cervical" does not appear in most pathological reports verbatim, possibly limiting the model's ability to respond appropriately to the question regarding the presence of cervical lymph nodes. Moreover, the LLM often incorrectly identified the size of the "primary tumor" and would instead provide a dimension from another specimen in the report, such as the overall thyroid lobe. This response accuracy may also be improved by modifying the question prompt and will be the focus of future work.

Overall, the use of LLMs may be an advancement from early NLP methods that faced limitations such as restrictive data preprocessing and the inability to handle multiple positive diagnoses [16-18]. Studies exploring the ability of LLMs to handle real clinical data have been rapidly growing. Examples include the investigation of the use of LLMs for multiple clinical NLP tasks such as the summarization of magnetic resonance imaging and radiology reports of the knee [19], the simplification of anatomic pathology reports for patient interpretation [20], and data extraction from breast cancer pathology and radiology reports [21]. However, the protection of private patient information must be prioritized with the use of any emerging technology. Safeguards should be established to ensure data breaches can be avoided with the use of LLMs for clinical tasks. The approach of deploying smaller, open-source LLMs behind a health care institution's own computing resources ensures that researchers across all centers may use this emerging technology while maintaining patient privacy. Additionally, the increased language capacity of the latest generation of LLMs allows institutions to deploy their own data for in-context learning only while potentially achieving a reasonable performance.

The limitations of this study include the use of a single institutional dataset without comparison with an external dataset to evaluate the generalizability of the LLM performance. Differences in language, reporting style, and description length of surgical pathology reports of different cancer types may also affect the ability of a smaller LLM such as the open-source FastChat-T5 used in this study to accurately extract the requested data using prompts. Additionally, while the intention of the study was to compare the performance of a smaller, open-source LLM to the gold standard of human reviewers, a comparison of its performance with larger, private LLMs such as GPT should be explored in future studies, after data have been confirmed to be entirely deidentified.

This study demonstrated that a locally deployed, open-source LLM is capable of automating data extraction faster than human reviewers who need to perform a manual review of surgical pathology reports. However, although the responses between the LLM and expert reviewers reached concordance rates of 88 - 89%, the degree of error by the LLM precludes its exclusive use to populate cancer statistics databases. Until LLMs demonstrate comparable performance to humans, the use of LLMs for clinical NLP tasks still requires considerable oversight by human reviewers. As a human-assistance tool, the role of LLMs may lie in improving the time efficiency of populating clinical databases. As methods of prompt engineering and fine-tuning improve, we envision that LLMs will eventually allow medical institutions to harness cutting-edge advances in NLP for timely and privacy-preserving MQA data extraction from pathology reports and other clinical narratives for the provision of real-time, essential oncologic insights.

Acknowledgments

This work was supported in part through the computational and data resources and staff expertise provided by the Scientific Computing and Data Department at the Icahn School of Medicine at Mount Sinai and supported by the Clinical and Translational Science Awards (CTSA) grant UL1TR004419 from the National Center for Advancing Translational Sciences. No generative artificial intelligence was used in the preparation or writing of this manuscript.

Data Availability

Source code for the study may be found at [12] under a GPLv3 license.

XSL•FO
RenderX

Authors' Contributions

Conceptualization: DL, AV, RF, GNN, MLM
Data curation: DL, KMM, AV
Formal analysis: DL, KMM, AV, DSM
Methodology: DL, AV, DSM
Supervision: GNN, MLM
Validation: DL, AV
Visualization: DL, KMM
Writing – DL, KMM, AV
Writing – review & editing: DL, AV, KMM, RF, DSM, MM, GNN

Conflicts of Interest

None declared.

Multimedia Appendix 1
Transcript of prompts.
[DOCX File, 15 KB - formative_v9i1e64544_app1.docx]

References

1. Boffa DJ, Rosen JE, Mallin K, et al. Using the national cancer database for outcomes research: a review. JAMA Oncol 2017 Dec 1;3(12):1722-1728. [doi: 10.1001/jamaoncol.2016.6905] [Medline: 28241198]
2. Hankey BF, Ries LA, Edwards BK. The surveillance, epidemiology, and end results program: a national resource. Cancer Epidemiol Biomarkers Prev 1999 Dec;8(12):1117-1121. [Medline: 10613347]
3. Burger G, Abu-Hanna A, de Keizer N, Cornet R. Natural language processing in pathology: a scoping review. J Clin Pathol 2016 Jul 22:jclinpath-2016-203872. [doi: 10.1136/jclinpath-2016-203872] [Medline: 27451435]
4. Hochheiser H, Finan S, Yuan Z, et al. DeepPhe-CR: natural language processing software services for cancer registrar case abstraction. JCO Clin Cancer Inform 2023 Sep;7:e2300156. [doi: 10.1200/CCI.23.00156] [Medline: 38113411]
5. Naseri H, Kafi K, Skamene S, et al. Development of a generalizable natural language processing pipeline to extract physician-reported pain from clinical reports: Generated using publicly-available datasets and tested on institutional clinical reports for cancer patients with bone metastases. J Biomed Inform 2021 Aug;120:103864. [doi: 10.1016/j.jbi.2021.103864] [Medline: 34265451]
6. Causa Andrieu P, Golia Pernicka JS, Yaeger R, et al. Natural language processing of computed tomography reports to label metastatic phenotypes with prognostic significance in patients with colorectal cancer. JCO Clin Cancer Inform 2022 Sep;6:e2200014. [doi: 10.1200/CCI.22.00014] [Medline: 36103642]
7. Karimi YH, Blayney DW, Kurian AW, et al. Development and use of natural language processing for identification of distant cancer recurrence and sites of distant recurrence using unstructured electronic health record data. JCO Clin Cancer Inform 2021 Apr;5:469-478. [doi: 10.1200/CCI.20.00165] [Medline: 33929889]
8. Yim WW, Yetisgen M, Harris WP, Kwan SW. Natural language processing in oncology: a review. JAMA Oncol 2016 Jun 1;2(6):797-804. [doi: 10.1001/jamaoncol.2016.0213] [Medline: 27124593]
9. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med 2023 Aug;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
10. Yang X, Chen A, PourNejatian N, et al. A large language model for electronic health records. NPJ Digit Med 2022 Dec 26;5(1):194. [doi: 10.1038/s41746-022-00742-2] [Medline: 36572766]
11. Rengers TA, Thiels CA, Salehinejad H. Academic surgery in the era of large language models: a review. JAMA Surg 2024 Apr 1;159(4):445-450. [doi: 10.1001/jamasurg.2023.6496] [Medline: 38353991]
12. LLMSurgicalPathologyNoteParserPublic. Github. URL: https://github.com/akhilvaid/LLMSurgicalPathologyNoteParser [accessed 2025-04-07]
13. Tuttle RM, Haugen B, Perrier ND. Updated American joint committee on cancer/tumor-node-metastasis staging system for differentiated and anaplastic thyroid cancer (eighth edition): what changed and why? Thyroid 2017 Jun;27(6):751-756. [doi: 10.1089/thy.2017.0102] [Medline: 28463585]
14. Haugen BR, Alexander EK, Bible KC, et al. 2015 American thyroid association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American thyroid association guidelines task force on thyroid nodules and differentiated thyroid cancer. Thyroid 2016 Jan;26(1):1-133. [doi: 10.1089/thy.2015.0020] [Medline: 26462967]
15. Savova GK, Danciu I, Alamudun F, et al. Use of natural language processing to extract clinical cancer phenotypes from electronic medical records. Cancer Res 2019 Nov 1;79(21):5463-5470. [doi: 10.1158/0008-5472.CAN-19-0579] [Medline: 31395609]
16. Datta S, Bernstam EV, Roberts K. A frame semantic overview of NLP-based information extraction for cancer-related EHR notes. J Biomed Inform 2019 Dec;100:103301. [doi: 10.1016/j.jbi.2019.103301] [Medline: 31589927]

XSL•FO
RenderX

17. Schadow G, McDonald CJ. Extracting structured information from free text pathology reports. AMIA Annu Symp Proc 2003;2003:584-588. [Medline: 14728240]
18. Mitchell JR, Szepietowski P, Howard R, et al. A question-and-answer system to extract data from free-text oncological pathology reports (CancerBERT network): development study. J Med Internet Res 2022 Mar 23;24(3):e27210. [doi: 10.2196/27210] [Medline: 35319481]
19. López-Úbeda P, Martín-Noguerol T, Díaz-Angulo C, Luna A. Evaluation of large language models performance against humans for summarizing MRI knee radiology reports: a feasibility study. Int J Med Inform 2024 Jul;187:105443. [doi: 10.1016/j.ijmedinf.2024.105443] [Medline: 38615509]
20. Steimetz E, Minkowitz J, Gabutan EC, et al. Use of artificial intelligence chatbots in interpretation of pathology reports. JAMA Netw Open 2024 May 1;7(5):e2412767. [doi: 10.1001/jamanetworkopen.2024.12767] [Medline: 38776080]
21. Choi HS, Song JY, Shin KH, Chang JH, Jang BS. Developing prompts from large language model for extracting clinical information from pathology and ultrasound reports in breast cancer. Radiat Oncol J 2023 Sep;41(3):209-216. [doi: 10.3857/roj.2023.00633] [Medline: 37793630]

Abbreviations

ICD: International Classification of Diseases
LLaMA: large language Meta AI
LLM: large language model
MQA: medical question answering
NCDB: National Cancer Database
NLP: natural language processing
PHI: protected health information
SEER: Surveillance, Epidemiology, and End Results

Edited by A Mavragani; submitted 19.07.24; peer-reviewed by M Abdalla, ST Arasteh; revised version received 17.01.25; accepted 19.01.25; published 07.04.25.

Please cite as:
Lee D, Vaid A, Menon KM, Freeman R, Matteson DS, Marin ML, Nadkarni GN
Using Large Language Models to Automate Data Extraction From Surgical Pathology Reports: Retrospective Cohort Study
JMIR Form Res 2025;9:e64544
URL: https://formative.jmir.org/2025/1/e64544
doi:10.2196/64544

© Denise Lee, Akhil Vaid, Kartikeya M Menon, Robert Freeman, David S Matteson, Michael L Marin, Girish N Nadkarni. Originally published in JMIR Formative Research (https://formative.jmir.org), 7.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Moving Standard Deviation of Trunk Acceleration as a Quantification Index for Physical Activities: Validation Study

Takuya Suzuki[1], MHS; Yuji Kono[1], PhD; Takayuki Ogasawara[2], PhD; Masahiko Mukaino[3,4], MD, PhD; Yasushi Aoshima[1], BSc; Shotaro Furuzawa[1], MHS; Yurie Fujita[1], BSc; Hirotaka Matsuura[3], MD, PhD; Masumi Yamaguchi[2], PhD; Shingo Tsukada[2], MD, PhD; Yohei Otaka[3], MD, PhD

[1]Department of Rehabilitation, Fujita Health University Hospital, Toyoake, Japan

[2]NTT Basic Research Laboratories and Bio-medical Informatics Research Center, NTT Corporation, Atsugi, Japan

[3]Department of Rehabilitation Medicine, Fujita Health University School of Medicine, Toyoake, Japan

[4]Department of Rehabilitation Medicine, Hokkaido University Hospital, Kita14 Nishi5, Kita-Ku, Sapporo, Japan

Corresponding Author:
Masahiko Mukaino, MD, PhD
Department of Rehabilitation Medicine, Fujita Health University School of Medicine, Toyoake, Japan

Abstract

Background: Step count is used to quantify activity in individuals using accelerometers. However, challenges such as difficulty in detecting steps during slow or irregular gait patterns and the inability to apply this method to wheelchair (WC) users limit the broader utility of accelerometers. Alternative device-specific measures of physical activity exist, but their specificity limits cross-applicability between different device sensors. Moving standard deviation of acceleration (MSDA), obtained from truncal acceleration measurements, is proposed as another alternative variable to quantify physical activity in patients.

Objective: This study aimed to evaluate the validity and feasibility of MSDA for quantifying physical activity in patients with stroke-induced hemiparesis by comparing it with the traditional step count.

Methods: We enrolled 197 consecutive patients with stroke hemiparesis admitted to a convalescent rehabilitation ward. Using the hitoe system, a smart clothing–based physical activity measurement system, we measured the MSDA of trunk movement and step count. The correlation between MSDA and step count was examined in all participants. Based on their daily living mobility levels, measured using the Functional Independence Measure (FIM), participants were categorized into 6 subgroups: FIM1-4, FIM5 (WC), FIM5 (walking), FIM6 (WC), FIM6 (walking), and FIM7 (walking). Intersubgroup differences in MSDA were analyzed.

Results: A strong correlation was observed between MSDA and step count (r=0.78; P<.001), with a stronger correlation in the walking group (r=0.79; P<.001) compared with the WC group (r=0.55; P<.001). The Shapiro-Wilk test indicated no significant results for MSDA across all subgroups, supporting a normal distribution within these groups. In contrast, the step count data for the WC subgroups showed significant results, indicating a deviation from a normal distribution. Additionally, 10.2% (20/197) of participants recorded zero steps, demonstrating a floor effect in the step count data. The median MSDA values for the 6 subgroups (FIM1-4, FIM5 WC, FIM5 walking, FIM6 WC, FIM6 walking, and FIM7) were 0.006, 0.007, 0.010, 0.011, 0.011, and 0.014, respectively, reflecting their levels of independence based on the FIM mobility scores. The median step counts for these subgroups were 68, 233, 1386, 367, 2835, and 4462, respectively. FIM5 participants who walked had higher step counts than FIM6 participants using WCs, though the difference was marginally but not statistically significant (P=.07), highlighting the impact of mobility type (walking vs WC).

Conclusions: The results suggest the validity of MSDA as a variable for physical activity in patients with stroke, applicable to patients with stroke irrespective of their mobility measures. This finding highlights the potential of MSDA for use in individuals with motor impairments, including WC users, underscoring its broad utility in rehabilitation clinical practice.

(JMIR Form Res 2025;9:e63064) doi:10.2196/63064

KEYWORDS

smart clothing; step count; moving standard deviation of acceleration; MSDA; wheelchair; activity quantification; physical activities; validation study; accelerometer; regular gait patterns; older people; aging; motor impairments; step detection; stroke; hemiparesis; measurement system; walking; mobility; rehabilitation

XSL•FO
RenderX

Introduction

Poststroke rehabilitation is primarily undertaken to enhance patient independence in activities of daily living (ADL). To achieve this goal, improving movement ability and exercise tolerance is important, which is directly reflected in the daily amount of activity. Increases in physical activity and extended rehabilitation time are associated with improved physical function and shorter hospital stays [1,2]. Higher levels of physical activity can decrease the mortality risk [3]. Therefore, physical activity requirements should be actively assessed in patients in the poststroke stage. To quantify physical activity, the number of steps, measured using an accelerometer, is widely used and is a valid and reliable variable for assessing physical activity in clinical rehabilitation settings [4,5]. Numerous studies have established a close association between the number of steps taken and health-related quality of life [6], self-efficacy [7], and the risk of stroke recurrence [8]. Accelerometers are reliable and validated devices that can objectively measure physical activity in patients with stroke [9]. Exercise therapy, when supplemented with accelerometer feedback and a goal-centered step-activity monitoring program, can enhance the physical activity of hospitalized patients recovering from ischemic stroke [10,11].

While widely used to quantify activity, counting steps is only applicable to ambulatory patients. Most of the previous studies have focused on patients with stroke who could walk independently. However, a substantial number of these patients face challenges with walking in their daily lives. During discharge from acute hospitalization, approximately 30% - 50% of patients had a modified Rankin scale score of ≥4 [12], indicating significant mobility impairment. While step count, an acceleration threshold–based measure of activity, may partially capture movement in everyday activities, its validity in accurately measuring the activity of patients with walking difficulties remains questionable. These considerations highlight the need for acceleration-based measures that can be effectively used in patients with difficulty walking.

To address this issue, we proposed an alternative metric: the moving standard deviation of acceleration (MSDA), which is derived from the SD of the acceleration norm that is measured from the trunk using an accelerometer placed on the chest. The chest placement of the accelerometer enables movement measurement in a seated position, which is expected to have advantages in activity monitoring in wheelchair (WC) users. The utility of the MSDA as an activity-quantification measure has been validated previously, showing a strong correlation between the MSDA and percentage oxygen uptake reserve in healthy participants and between MSDA and percentage heart rate reserve in patients with motor dysfunction due to cerebrovascular disease [13]. However, its validity compared with traditional accelerometry-based measures, such as step count, remains unexplored. Additionally, the potential advantages of MSDA in quantifying physical activity for individuals with limited mobility, including WC users, have yet to be confirmed with real-world data.

Consequently, this study focused on investigating the validity and feasibility of the MSDA measurement with a hitoe system in the measurement of physical activity in patients with stroke, with or without a WC. Specifically, our objectives were (1) to examine the correlation between the MSDA and the number of steps taken and (2) to compare their data characteristics, which includes an understanding of the features of data distribution and investigating their respective relationships with the level of mobility limitation.

Methods

Participants

Between January 2018 and February 2020, we enrolled consecutive patients with stroke hemiparesis who were admitted to the recovery rehabilitation ward of Fujita Health University Hospital. The inclusion criteria were (1) a confirmed diagnosis of stroke resulting in hemiparesis, (2) aged 18 years or older, and (3) the provision of written informed consent. The exclusion criteria were (1) noncompliance with wearing the hitoe measurement wear and (2) the presence of severe comorbid conditions, such as advanced heart failure or respiratory diseases requiring oxygen support that could interfere with participation in the rehabilitation program or data collection.

Study Design and Protocol

This single-center cross-sectional study comprised participants who underwent physical activity assessment within 1 week of admission to the recovery rehabilitation ward.

Ethical Considerations

This study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki and was approved by the Institutional Review Board of Fujita Health University (approval number HM17-220). All eligible patients who were admitted were informed about the study by the clinicians and provided with detailed information about its objectives and procedures. Those who agreed to participate were enrolled in the study after providing written informed consent. To ensure privacy, all data were deidentified before analysis. Participants did not receive financial compensation for their involvement in this study.

Measurement

A hitoe system (NTT Corporation and Toray Industry, Inc) was used to measure the amount of physical activity [14]; this study specifically evaluated its accelerometry-based physical activity measurement function. The hitoe system is a clothing-based activity-monitoring system that consists of clothing with embedded electrodes for heart rate measurement, a transmitter equipped with an embedded accelerometer, a data receiver (which can be a smartphone or an Internet of Things gateway), and a data server. The wearer's activity is measured using an accelerometer to capture physical movement and a heart rate monitor to assess load. The transmitter was securely fastened to the chest with a tight-fitting strap to minimize measurement noise. The noninferiority of accuracy in accelerometry-based measurements using the hitoe system has been demonstrated previously through comparisons with various

XSL•FO
RenderX

accelerometry-based devices, including waist-worn accelerometers (pedometers) [15]. The sampling rate was set to 25 Hz. Monitoring started before noon on the first assessment day and ended in the afternoon on the third assessment day. As clothing was used for monitoring, monitoring was paused during bathing. After bathing, the hitoe wear was changed, and the monitoring resumed. To mitigate the impact of daily fluctuations and compensate for deficiencies [16], we used 48-hour data to calculate the 24-hour ensemble average. Ensemble averaging is a method in which multiple measurements recorded at the same time of the day are averaged together. For instance, the data recorded between 5 PM and 5:01 PM on the first day was averaged with the data recorded at the same time on the second day, producing the ensemble average for 5 PM-5:01 PM. This process was repeated for all time points throughout the 24-hour period. The data with ≥5% measurement error after ensemble averaging were excluded from the analysis. The acquired 24-hour average values of minute-by-minute data on the number of daily steps and the MSDA were used for the analysis. The step-detection capability of this system has been validated using data from various commercial pedometers [15]. The MSDA was derived from a 2-second window of 50 samples of the norm of truncal acceleration, as measured by the accelerometer. The acceleration norm was computed using the following equation:

$$Norm=(x2+y2+z2)$$

where x, y, and z represent the vertical, lateral, and anterior/posterior axes, respectively.

Mobility levels of the patients were assessed using the Functional Independence Measure (FIM) scores. The FIM consists of 18 (13 motor and 5 cognitive) items. Each item is rated on a 7-point scale from 1 (full assistance) to 7 (independence). The data on other clinical characteristics including age, sex, height, weight, and disease were collected from the patients' medical records.

Data Analysis

Continuous variables are presented as mean and SD and categorical variables as percentages. The Shapiro-Wilk test was used to examine the normality of the number of steps and MSDA. To further investigate the characteristics of these distributions, their skewness and kurtosis were examined using the Fisher-Pearson coefficient of skewness (g1) and Fisher coefficient of kurtosis (g2) [17]. The relationship between MSDA and number of steps was examined using Spearman rank correlation coefficient. This analysis was conducted for all participants and separately for the two groups: those in the walking group and those in the WC group. Finally, we categorized all participants into 6 subgroups based on the mobility scores from the FIM and their mobility patterns as follows: FIM1-4, FIM5 WC, FIM5 walking, FIM6 WC, FIM6 walking, and FIM7. The Kruskal-Wallis test was performed to analyze the association between MSDA and the number of steps. The Bonferroni method was used to control for multiple comparisons.

All analyses were performed using SPSS (version 26; IBM Corp). Statistical significance was set at $P<.05$.

Results

A total of 197 patients (129 men) were enrolled and included in the analysis (Table 1). The mean age was 66.1 (SD 12.5) years. Of these, 61 patients were ambulatory, and 136 patients used WCs. The median number of days after onset was 33 (IQR 24 - 45). The average MSDA and number of steps were 0.009 (SD 0.003) and 1229 (SD 1834), respectively.

XSL•FO
RenderX

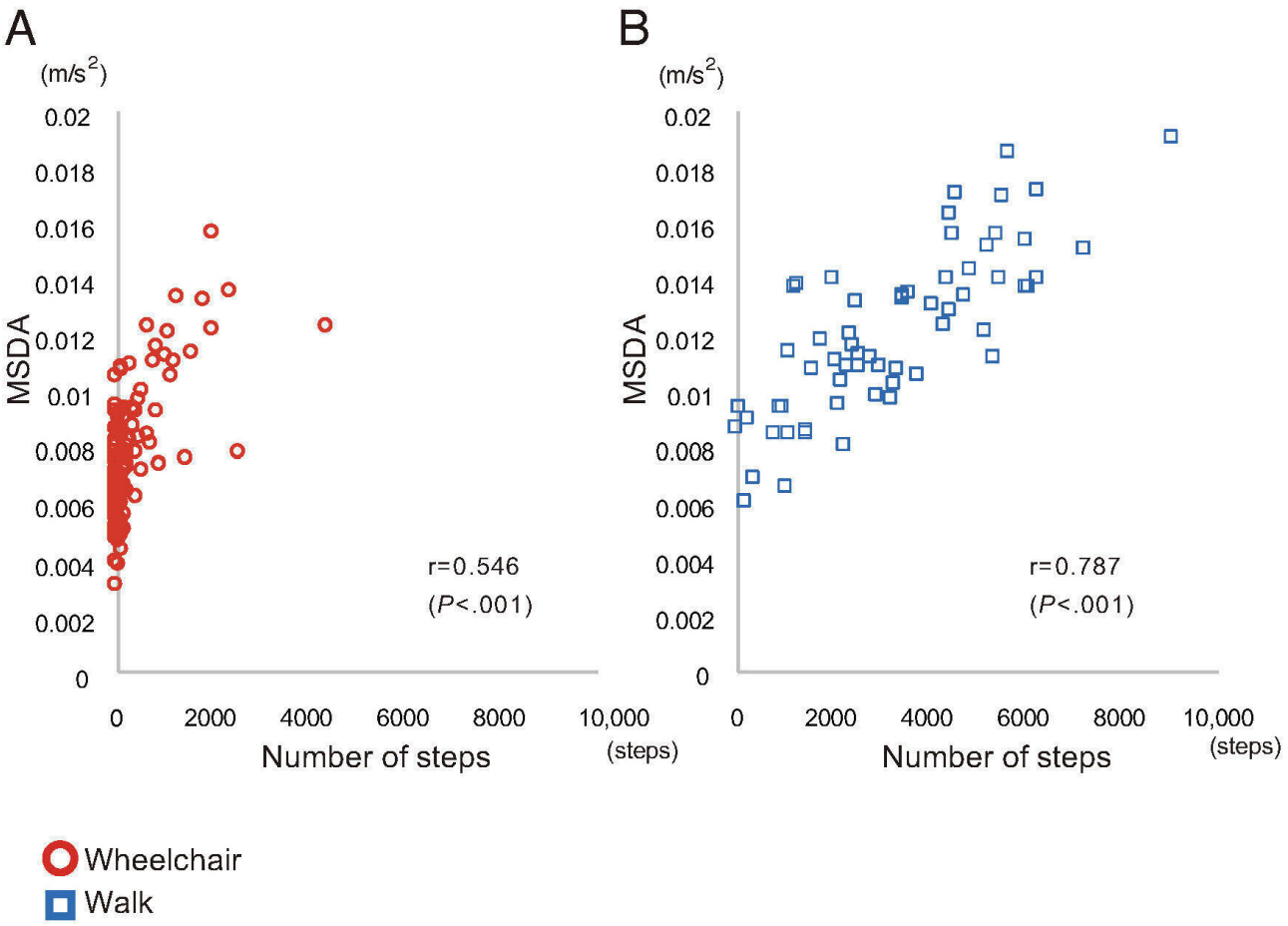
Table . Participant characteristics in the study cohort. Baseline characteristics of 197 participants including participants' demographics, stroke etiology, mobility status, and Functional Independence Measure (FIM) scores (motor, cognitive, and total) are summarized.

Variables	Values
Age (years), mean (SD)	66.1 (12.5)
Sex, n (%)	
Male	129 (65.5)
Female	68 (34.5)
Height (cm), mean (SD)	161.7 (15.5)
Weight (kg), mean (SD)	58.6 (13.7)
Days after onset, median (IQR)	33 (24 - 45)
Etiology, n (%)	
Cerebral hemorrhage	71 (36)
Cerebral infarction	94 (47.7)
Subarachnoid hemorrhage	16 (8.1)
Others	16 (8.1)
Mobility, n (%)	
Walking	61 (31)
Wheelchair	136 (69)
FIM-motor, mean (SD)	48.5 (26.0)
FIM-cognitive, mean (SD)	22.5 (10.2)
FIM-total, mean (SD)	70.7 (34.6)

The relationship between the MSDA and the number of steps is shown in Figure 1. Significant positive correlations were found for all participants (r=0.784; P<.001), the walking group (r=0.787; P<.001), and the WC group (r=0.546; P<.001). When excluding the extreme data with 0 values, the correlation coefficient was raised, for all participants (r=823; P<.001), the walking group (r=0.798; P<.001), and the WC group (r=0.634; P<.001).

XSL•FO
RenderX

Figure 1. Scatter plot showing the correlations between MSDA and number of steps in the (A) wheelchair group (n=136) and (B) walking group (n=61). Open circles and open squares represent the wheelchair and walking groups, respectively. MSDA: moving standard deviation of acceleration.



The detailed features of the MSDA distribution and the number of steps for all participants are shown in Figure 2. Neither the MSDA nor the number of steps exhibited a normal distribution. The distributions of MSDA and steps exhibited right-skewness, with skewness coefficients of 0.78 and 1.72, respectively (zero when normally distributed). The kurtosis coefficient for MSDA was 3.02, which is close to the typical value of 3 for a normal distribution. In contrast, the kurtosis coefficient for steps was 5.26, indicating a higher level of kurtosis. The normality of the distribution was further tested in each of the 6 subgroups (Table 2). The distribution of MSDA data within subgroups exhibited smaller deviations from a normal distribution in terms of kurtosis and skewness coefficients, suggesting that it was closer to a normal distribution compared with the steps data. The Shapiro-Wilk test did not yield significant results for MSDA in all the subgroups, whereas the step data showed significance, especially within the WC subgroups. In the step count, 10.2% (20/197) of participants exhibited zero count, showing the floor effect.

XSL•FO
RenderX

Figure 2. Histograms showing the distribution of (A) MSDA and (B) daily step count for all 197 participants, divided by mobility type. The wheelchair group is represented in red and the walking group in blue. MSDA: moving standard deviation of acceleration.

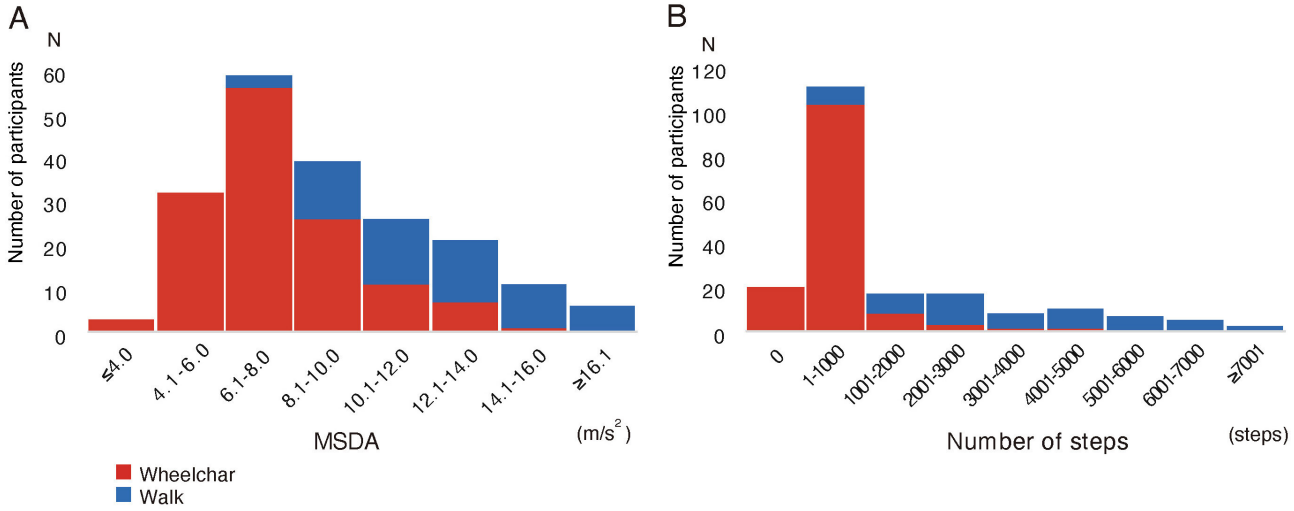


Table . Kurtosis and skewness values and results of the Shapiro-Wilk test. This table summarizes the normality test results for the moving standard deviation of acceleration (MSDA) and daily step count in participants classified by mobility type and functional levels. For normal distribution, kurtosis (g2)=3 and skewness (g1)=0.

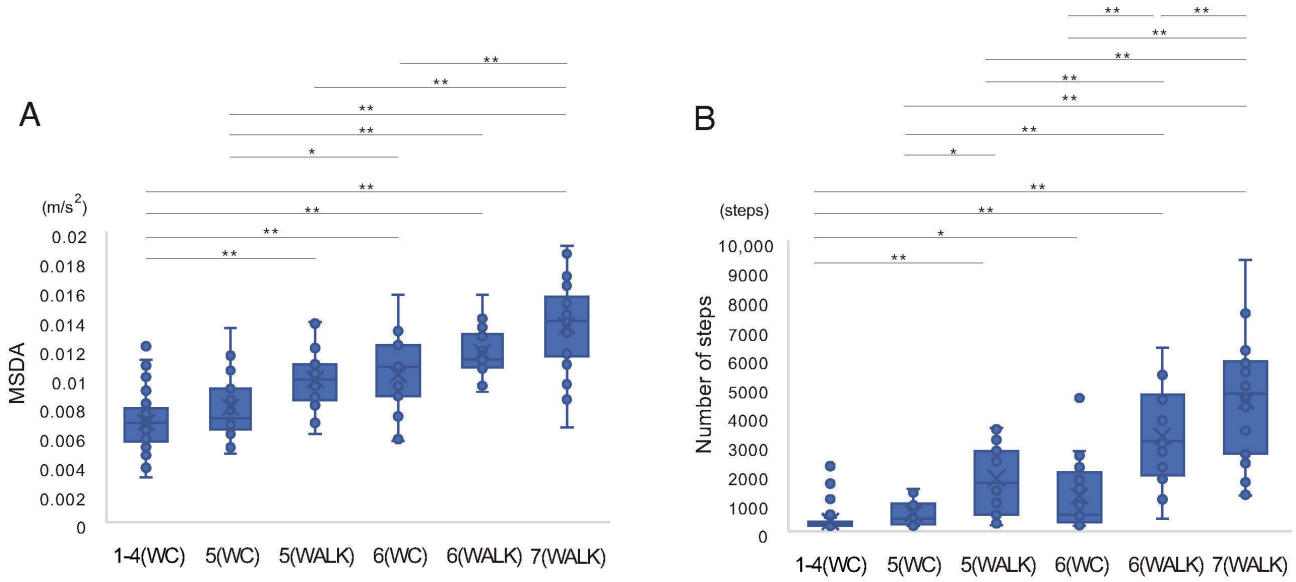
	MSDA			Number of steps		
	Kurtosis (g2)	Skewness (g1)	*P* value (Shapiro-Wilk)	Kurtosis (g2)	Skewness (g1)	*P* value (Shapiro-Wilk)
WC[a] 1 - 4 (n=94)	3.33	0.51	.18	24.33	4.32	<.001
WC 5 (n=23)	3.15	0.94	.05	2.22	0.75	.005
WC 6 (n=19)	2.59	0.05	.92	4.37	1.38	.002
Walking 5 (n=15)	2.48	0.27	.79	1.64	0.01	.23
Walking 6 (n=18)	2.75	0.47	.67	1.95	0.09	.83
Walking 7 (n=28)	2.63	–0.38	.62	2.57	0.10	.42
WC (n=136)	3.91	0.93	<.001	19.48	3.61	<.001
Walking (n=61)	2.56	0.19	.72	2.61	0.40	.11
All participants (n=197)	3.02	0.78	<.001	5.26	1.72	<.001

[a]WC: wheelchair.

The relationships between the 6 subgroups according to the FIM scores of mobility items, mobility patterns, MSDA, and number of steps are shown in Figure 3. The median MSDA values in the 6 subgroups FIM1-4, FIM5 WC, FIM5 walking, FIM6 WC, FIM6 walking, and FIM7 were 0.006, 0.007, 0.010, 0.011, 0.011, and 0.014, respectively. Significant differences were found in MSDA between the low-FIM and high-FIM subgroups (Figure 3A). The median number of steps in the 6 subgroups were 68, 233, 1386, 367, 2835, and 4462, respectively. Significant differences in the steps between the low-FIM and high-FIM subgroups are shown in Figure 3B. The difference between FIM5 participants with walking and FIM6 WC was marginal (*P*=.07).

XSL•FO
RenderX

Figure 3. Relationship between Functional Independence Measure (FIM)-mobility scores and the MSDA for 6 groups divided by mobility type: FIM1-4, FIM5 (wheelchair), FIM5 (walking), FIM6 (wheelchair), FIM6 (walking), and FIM7 (walking). Physical activity was compared among the 6 subgroups according to the FIM scores of the mobility items and mobility patterns. (A) MSDA and (B) number of steps. MSDA: moving standard deviation of acceleration. *P<.05, **P<.01.



Discussion

Principal Findings

This study investigated the characteristics of the MSDA and step count in measuring physical activity in patients with stroke-induced hemiparesis. The MSDA not only had a positive correlation with the number of steps but was also associated with the FIM-mobility score, showing the validity as activity measurement index. Additionally, the MSDA demonstrated potential advantages in quantifying activity levels in patients with walking difficulties. First, in contrast to the step count, the distribution of the MSDA exhibited a closer alignment with a normal distribution, particularly within subgroups of patients using WCs and those with similar ADL levels. Second, the MSDA did not present a floor effect within the WC group, which was observed with the step count. Finally, the MSDA was consistent with varying levels of mobility independence, showing no significant differences across types of mobility measures (walking or WC use). In contrast, the step count was more susceptible to the mode of mobility; significant differences were observed in the step values between groups with the same level of independence but using different mobility measures.

The findings of this study demonstrated the strength of the MSDA as a measure of physical activity in patients after stroke, who often present with gait disturbance and relies on WC. The conducted analyses on the correlation between the step counts and MSDA and the distribution of those in the poststroke rehabilitation inpatients, of whom approximately 70% relied on a WC for mobility, support the usability of the MSDA in those with limited gait functionality. Our results indicated that the MSDA was positively correlated with the number of steps, which is an established measure of physical activity, supporting the validity of the MSDA. Further analysis revealed a weaker correlation in patients who used WCs. The weaker correlation between step counts and MSDA in WC users can be attributed to the skewed distribution and floor effects in their step counts. Notably, excluding participants with a step count of zero results in an improved correlation coefficient. Further stratified examination revealed that, in contrast to the number of steps, the MSDA was normally distributed within all levels of mobility independence. These findings suggest that the MSDA has broad applicability, including that for patients in poststroke stage who use WCs.

The strength of the MSDA in measuring physical activity among patients in poststroke stage, with fewer issues in distribution patterns, may stem from its nature as a continuous variable. The number of steps, which is a frequently used variable for activity monitoring, is a threshold-based variable that measures the number of movements with greater acceleration beyond a specific threshold [16]. However, this approach may induce challenges when applied to patients with neurological disorders who often manifest irregular step patterns and slower gait speeds. The use of accelerometers to count steps can produce misleading results in such cases [18]. In contrast, the MSDA quantifies truncal oscillations; as the trunk is the body's heaviest segment, this can, without imposing a specific threshold, be captured as a continuous value [13]. Thus, the MSDA could be more sensitive to variations in nonwalking activities. Consequently, the MSDA distribution encompasses both walking and nonwalking activities; this characteristic of the MSDA underscores its potential application as a comprehensive metric for evaluating activity levels irrespective of individual mobility methods, whether walking or using WCs.

Furthermore, our results illustrate that the MSDA effectively reflects the level of independence in daily life. Previous studies have demonstrated a strong correlation between physical activity and independence levels; for instance, the Life-Space Assessment, a physical activity measure, shows a positive correlation with Lawton instrumental ADL scores [19,20]. A similar positive correlation has been noted between accelerometer-measured physical activity and FIM-motor scores in patients with stroke [21].

XSL•FO
RenderX

The study's results further support previous observations and reveal significant differences in both the MSDA and step count between patients with and without independent mobility. Furthermore, in contrast to the step count, the MSDA appears to be less influenced by the mode of mobility. In the case of mobility levels FIM5 and FIM6, no significant differences were observed between patients using WCs and those walking, while significant disparities in step counts were observed across different mobility measures at both levels, leading to reduced coherence with respect to levels of independence—it tends to be higher for patients capable of walking with a level of FIM 5, as opposed to those who are using WC with a FIM score of 6. These findings suggest that the MSDA can serve as a universal measure of physical activity, regardless of an individual's gait capabilities. This objective measure could be beneficial for providing feedback and setting goals for rehabilitation, to encourage daily activities. Many community-dwelling patients with a stroke fail to achieve the recommended activity level [22]. Thus, using the MSDA to track daily activities could offer a valuable strategy for promoting more active daily living in the realm of rehabilitation.

Limitations

This study has some limitations. First, its applicability to nonhospitalized patients warrants further exploration. The study was conducted in a rehabilitation ward with a flat floor, which differs from the environment outside the hospital. With barriers such as steps or uneven surfaces, the MSDA values may be higher than those in hospital wards. The impact of environmental variables on MSDA merits investigation in future studies. Despite these considerations, this study is meaningful in demonstrating the advantage of MSDA in the measurement of physical activity in patients with low walking ability or WC users.

Second, the feasibility of using the MSDA with the hitoe system in a home setting should be further explored. In this study, therapist assistance was required for fitting and ensuring proper placement to maintain measurement quality. While the system has been designed to allow patients with hemiparesis to wear it independently, those with severe hemiparesis may still encounter difficulties. For this system to be effectively integrated into daily activity management for patients with hemiparesis, ease of fitting is essential. Further investigation is needed to adapt and implement this system in practical clinical settings.

Third, as the MSDA reflects truncal movement oscillations without identifying the specific movements involved, certain neurological symptoms may influence the measurements. For instance, patients with Parkinson disease experience tremors that can affect the MSDA values, which reflect movement variability. Further research is needed to clarify the movements or symptoms that may distort MSDA values. Finally, to enhance understanding, the clinical significance of MSDA values should be further investigated, specifically in terms of interpreting different values as corresponding levels of physical activity. While the step count serves as an easily understandable and intuitive metric for clinical feedback, the MSDA value lacks this straightforwardness, posing challenges for patients in correlating it with specific activities. Therefore, fostering a deeper comprehension and interpretation of this variable within clinical contexts is imperative. Efforts should be directed towards contextualizing the MSDA values and making them more intuitive and meaningful for patients.

Conclusion

This study indicates that the MSDA is a valid variable for measurement of physical activity, applicable to patients with stroke irrespective of their mobility measures. Future research is warranted to examine the factors influencing MSDA values and to establish the clinical utility of this variable.

Acknowledgments

The authors thank Mr Kenta Maruyama at NTT Basic Research Laboratories for their technical assistance with the data analysis.

Data Availability

The data collected and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors used leased wearable clothing and hardware devices from Toray Industries and NTT Corporation, respectively, which are the manufacturers of the tracking devices. TO, MY, and ST are employees of NTT Corporation. The other authors declare no conflicts of interest associated with this manuscript.

References

1. Schneider EJ, Lannin NA, Ada L, Schmidt J. Increasing the amount of usual rehabilitation improves activity after stroke: a systematic review. J Physiother 2016 Oct;62(4):182-187. [doi: 10.1016/j.jphys.2016.08.006] [Medline: 27637769]
2. English C, Shields N, Brusco NK, et al. Additional weekend therapy may reduce length of rehabilitation stay after stroke: a meta-analysis of individual patient data. J Physiother 2016 Jul;62(3):124-129. [doi: 10.1016/j.jphys.2016.05.015] [Medline: 27320831]
3. Inoue M, Iso H, Yamamoto S, et al. Daily total physical activity level and premature death in men and women: results from a large-scale population-based cohort study in Japan (JPHC study). Ann Epidemiol 2008 Jul;18(7):522-530. [doi: 10.1016/j.annepidem.2008.03.008] [Medline: 18504139]

XSL•FO
RenderX

4. Mathie MJ, Coster ACF, Lovell NH, Celler BG. Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. Physiol Meas 2004 Apr;25(2):R1-20. [doi: 10.1088/0967-3334/25/2/r01] [Medline: 15132305]
5. Dowd KP, Szeklicki R, Minetto MA, et al. A systematic literature review of reviews on techniques for physical activity measurement in adults: a DEDIPAC study. Int J Behav Nutr Phys Act 2018 Feb 8;15(1):15. [doi: 10.1186/s12966-017-0636-2] [Medline: 29422051]
6. Kanai M, Izawa KP, Kubo H, Nozoe M, Mase K, Shimada S. Association of health utility score with physical activity outcomes in stroke survivors. Int J Environ Res Public Health 2020 Dec 31;18(1):251. [doi: 10.3390/ijerph18010251] [Medline: 33396274]
7. Kanai M, Nozoe M, Izawa KP, et al. Promoting physical activity in hospitalized patients with mild ischemic stroke: a pilot study. Top Stroke Rehabil 2017 May;24(4):256-261. [doi: 10.1080/10749357.2016.1259030] [Medline: 27881044]
8. Kono Y, Kawajiri H, Kamisaka K, et al. Predictive impact of daily physical activity on new vascular events in patients with mild ischemic stroke. Int J Stroke 2015 Feb;10(2):219-223. [doi: 10.1111/ijs.12392] [Medline: 25381835]
9. Rand D, Eng JJ, Tang PF, Jeng JS, Hung C. How active are people with stroke?: use of accelerometers to assess physical activity. Stroke 2009 Jan;40(1):163-168. [doi: 10.1161/STROKEAHA.108.523621] [Medline: 18948606]
10. Kanai M, Izawa KP, Kobayashi M, et al. Effect of accelerometer-based feedback on physical activity in hospitalized patients with ischemic stroke: a randomized controlled trial. Clin Rehabil 2018 Aug;32(8):1047-1056. [doi: 10.1177/0269215518755841] [Medline: 29400070]
11. Danks KA, Roos MA, McCoy D, Reisman DS. A step activity monitoring program improves real world walking activity post stroke. Disabil Rehabil 2014;36(26):2233-2236. [doi: 10.3109/09638288.2014.903303] [Medline: 24670193]
12. Toyoda K, Yoshimura S, Nakai M, et al. Twenty-year change in severity and outcome of ischemic and hemorrhagic strokes. JAMA Neurol 2022 Jan 1;79(1):61-69. [doi: 10.1001/jamaneurol.2021.4346] [Medline: 34870689]
13. Mukaino M, Ogasawara T, Matsuura H, et al. Validity of trunk acceleration measurement with a chest-worn monitor for assessment of physical activity intensity. BMC Sports Sci Med Rehabil 2022 Jun 10;14(1):104. [doi: 10.1186/s13102-022-00492-4] [Medline: 35689292]
14. Ogasawara T, Matsunaga K, Ito H, Mukaino M. Application for rehabilitation medicine using wearable textile “hitoe”. NTT Tech Rev 2018;16(9):6-12. [doi: 10.53829/ntr201809fa2]
15. Ogasawara T, Itoh Y, Kuwabara K, Kasahara R. Gait analysis using a wearable T-shirt type sensor. NTT Tech Rev 2016;14(4):47-53. [doi: 10.53829/ntr201604ra1]
16. Ogasawara T, Mukaino M, Otaka Y, et al. Validation of data imputation by ensemble averaging to quantify 24-h behavior using heart rate of stroke rehabilitation inpatients. J Med Biol Eng 2021 Jun;41(3):322-330. [doi: 10.1007/s40846-021-00622-2]
17. Pearson ES, Hartley HO. Biometrika Tables for Statisticians, 3rd edition: Cambridge University Press; 1970:207.
18. Elsworth C, Dawes H, Winward C, et al. Pedometer step counts in individuals with neurological conditions. Clin Rehabil 2009 Feb;23(2):171-175. [doi: 10.1177/0269215508098895] [Medline: 19164404]
19. Baker PS, Bodner EV, Allman RM. Measuring life-space mobility in community-dwelling older adults. J Am Geriatr Soc 2003 Nov;51(11):1610-1614. [doi: 10.1046/j.1532-5415.2003.51512.x] [Medline: 14687391]
20. Peel C, Sawyer Baker P, Roth DL, Brown CJ, Brodner EV, Allman RM. Assessing mobility in older adults: the UAB study of aging life-space assessment. Phys Ther 2005 Oct;85(10):1008-1119. [doi: 10.1093/ptj/85.10.1008] [Medline: 16180950]
21. Hobo K, Kurita H, Momose K. The relationship between energy expenditure and physical functions in patients hospitalised for stroke. Sci Rep 2021 Nov 4;11(1):21685. [doi: 10.1038/s41598-021-01135-3] [Medline: 34737355]
22. Michael KM, Allen JK, Macko RF. Reduced ambulatory activity after stroke: the role of balance, gait, and cardiovascular fitness. Arch Phys Med Rehabil 2005 Aug;86(8):1552-1556. [doi: 10.1016/j.apmr.2004.12.026] [Medline: 16084807]

Abbreviations

ADL: activities of daily living
FIM: Functional Independence Measure
MSDA: moving standard deviation of acceleration
WC: wheelchair

XSL•FO
RenderX

Edited by A Mavragani; submitted 21.06.24; peer-reviewed by A Opheim; revised version received 09.12.24; accepted 09.12.24; published 08.04.25.

Please cite as:
Suzuki T, Kono Y, Ogasawara T, Mukaino M, Aoshima Y, Furuzawa S, Fujita Y, Matsuura H, Yamaguchi M, Tsukada S, Otaka Y
Moving Standard Deviation of Trunk Acceleration as a Quantification Index for Physical Activities: Validation Study
JMIR Form Res 2025;9:e63064
URL: https://formative.jmir.org/2025/1/e63064
doi:10.2196/63064

© Takuya Suzuki, Yuji Kono, Takayuki Ogasawara, Masahiko Mukaino, Yasushi Aoshima, Shotaro Furuzawa, Yurie Fujita, Hirotaka Matsuura, Masumi Yamaguchi, Shingo Tsukada, Yohei Otaka. Originally published in JMIR Formative Research (https://formative.jmir.org), 8.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Provider Perspectives on Implementing an Enhanced Digital Screening for Adolescent Depression and Suicidality: Qualitative Study

Morgan A Coren[1], MD; Oliver Lindhiem[1], PhD; Abby R Angus[2], BS; Emma K Toevs[2], BA; Ana Radovic[1,3], MD, MSc

[1]School of Medicine, University of Pittsburgh, 100 N Bellefield, Office 537, Pittsburgh, PA, United States

[2]University of Pittsburgh Medical Center, Western Psychiatric Hospital, Pittsburgh, PA, United States

[3]Adolescent and Young Adult Medicine, University of Pittsburgh Medical Center, Children's Hospital of Pittsburgh, Pittsburgh, PA, United States

Corresponding Author:
Oliver Lindhiem, PhD
School of Medicine, University of Pittsburgh, 100 N Bellefield, Office 537, Pittsburgh, PA, United States

Abstract

Background: With a growing adolescent mental health crisis, pediatric societies are increasingly recommending that primary care providers (PCPs) engage in mental health screening. While symptom-level screens identify symptoms, novel technology interventions can assist PCPs with providing additional point-of-care guidance to increase uptake for behavioral health services.

Objective: In this study, we sought community PCP feedback on a web-based, digitally enhanced mental health screening tool for adolescents in primary care previously only evaluated in research studies to inform implementation in community settings.

Methods: A total of 10 adolescent providers were recruited to trial the new screening tool and participate in structured interviews based on the Consolidated Framework for Implementation Research domains. Interviews were audio recorded, transcribed, and coded according to a prespecified codebook using a template analysis approach.

Results: Providers identified improving mental health screening and treatment in pediatric primary care as a priority and agreed that a web-based digitally enhanced screening tool could help facilitate identification of and management of adolescent depression. Salient barriers identified were lack of electronic health record integration, time to administer screening, implications on clinic workflow, accessibility, and lack of transparency within health care organizations about the process of approving new technologies for clinical use. Providers made multiple suggestions to enhance implementation in community settings, such as incorporating customization options.

Conclusions: Technology interventions can help address the need for improved behavioral health support in primary care settings. However, numerous barriers exist, complicating implementation of new technologies in real-world settings.

(JMIR Form Res 2025;9:e67624) doi:10.2196/67624

KEYWORDS

depression; suicidality; adolescent mental health screening; primary care; digital tools

Introduction

Adolescent depression and suicidality are part of a growing public health crisis that has been further exacerbated by the COVID-19 pandemic [1]. In 2022, 19.5% of US adolescents aged 12 - 17 years experienced a major depressive episode and 13.4% reported having seriously considered suicide [2]. Pediatric societies recommend that pediatric primary care providers (PCPs) screen adolescents for both depression and suicidality [3-6]. Although screening adolescents for depression is recommended by the US Preventive Services Taskforce [7], there is no evidence to support that screening by itself improves health outcomes [7]. PCPs recognize the need for additional interventions to increase uptake of services and provide clinical benefit, but requirements for increased mental health staff and support obfuscate their implementation.

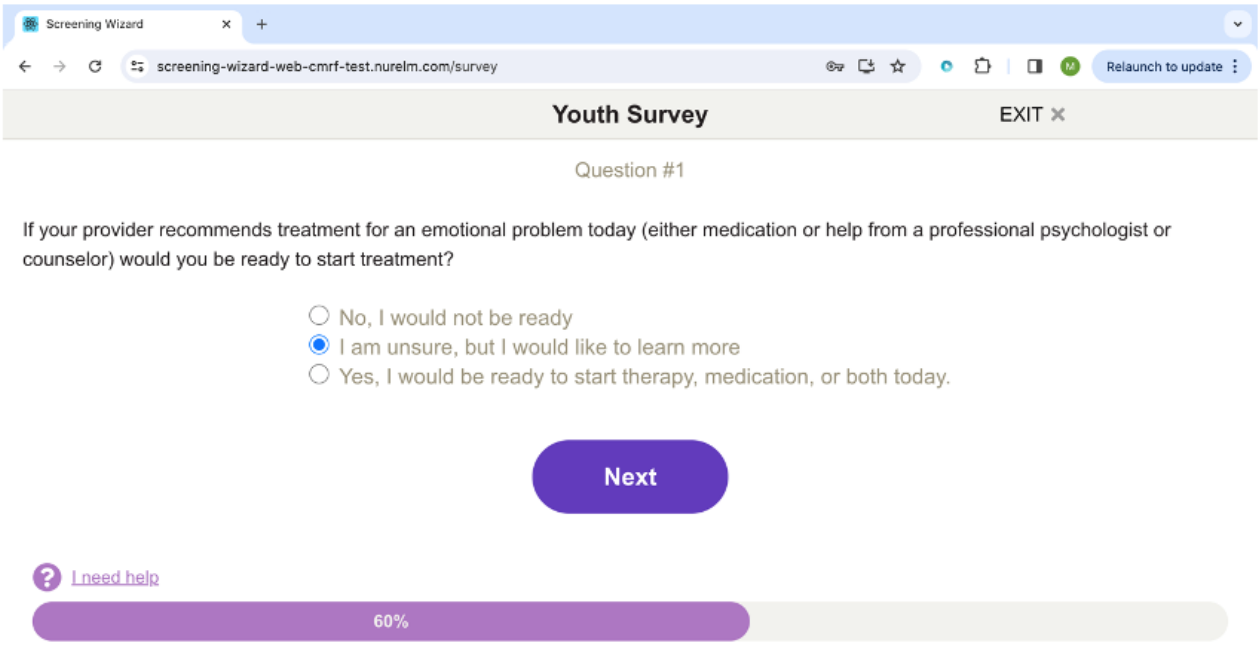
Technology interventions can aid providers to implement screening recommendations and augment initiatives to improve behavioral health support in primary care. In recent years, technology-based screeners have been developed for detecting conditions in addition to depression such as posttraumatic stress disorder (PTSD) [8], substance use [9], and interpersonal violence [10] in primary care settings, as well as to screen patients for social needs [11]. Such interventions have proven to be effective, accurate, well-received by patients, and helpful in streamlining clinic workflow and history-taking.

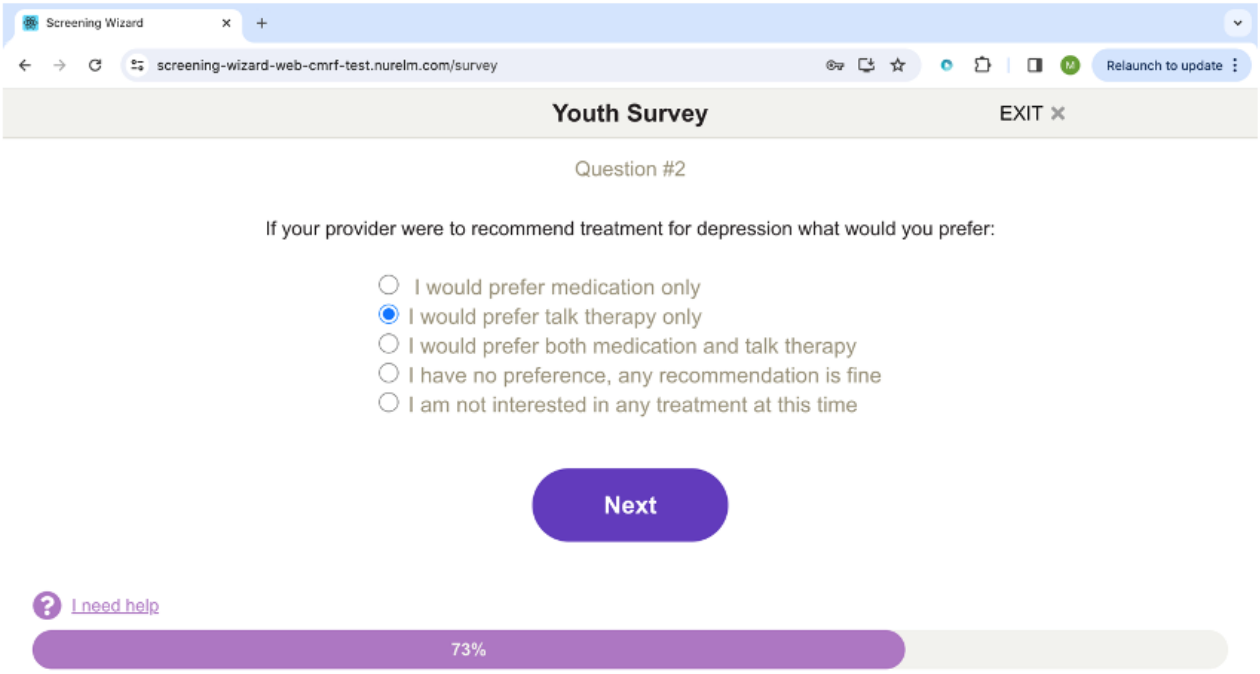
Screening Wizard (SW; see Figure 1 for a screenshot) is a web-based digitally enhanced screening tool for eliciting

XSL•FO
RenderX

symptoms of depression and suicidality in adolescents. In addition to surveying adolescents and their caregivers through separate unique weblinks with validated screening measures for comorbidities of anxiety and mania, the screening tool assesses additional considerations, such as treatment readiness, motivation, treatment preferences, and barriers. The surveys take approximately 5 - 7 minutes to complete. Once completed, a report succinctly synthesizes the information to provide PCPs with comparisons between adolescents and their caregiver responses that may guide management decisions and referrals. The initial qualitative study informing the design of SW [12], evaluation of iterative feedback from stakeholders, and pilot usability trial of the screening tool in a general sample of adolescents [13] are described in previously published papers.

Figure 1. Sample images of the Screening Wizard tool, as it appears to patients.





These previous studies and provider feedback informed several modifications to the enhanced screening tool so that it could be used clinically outside of a research context. The clinical use version was updated with functionality for a clinic-specific

XSL•FO
RenderX

Health Insurance Portability and Accountability Act (HIPAA)–secure log-in to the SW system to separate its administration and data collection from the SW research environment. This allowed clinical staff to assign adolescent patients and their caregiver unique weblinks to the enhanced screen and schedule when they would receive it through text message directly to their mobile phone; or alternatively print a list of QR code weblinks to distribute to individual patients. After completion, staff would be able to review results on a private clinician dashboard. In response to safety and liability concerns related to remote suicide screening, surveys were scheduled to expire at the end of the same clinic day to allow time for clinic staff to review all completed surveys. Planned changes also included condensing the number of items by presenting additional screening items only for participants with positive initial depression screens. Integration of the screening tool into the electronic health record (EHR) had been identified as a high priority initiative at this stage, though this required clinical system-wide permission, which was not accessible to the research team.

As the enhanced screening tool was initially developed within an academic setting with access to embedded behavioral health staff and research support, this study was intended to pilot the screening tool in more typical and less-resourced community health care settings where adolescents seek primary care or other services (eg, reproductive care) to elucidate barriers and facilitators to implementation from providers who are the intended users of the tool. This tool was offered to multiple clinical sites to pilot and provide feedback on the piloting process. Although health care providers at these sites were interested in using the intervention, many clinics (especially those within larger health care systems) had extensive IT audit and administrative approval processes to trial new technology tools, which presented substantial barriers to piloting. Only 3 out of 10 clinics were able to pilot the tool very briefly. It became apparent that barriers to implementation would differ in a real-world versus research setting and additional elicitation from providers about these barriers and facilitators would be needed to enable scalability. Preimplementation studies are most helpful when they allow participants to try out the tool [12,14]. However, due to the barriers to piloting, we developed an additional prototype of the tool for interested providers to test informally, using fabricated patient data.

In this study, we conducted structured interviews with primary care providers based on domains from the Consolidated Framework for Implementation Research [15,16], a widely used framework for systematically identifying barriers and facilitators to implementation of new interventions. The information learned will be used to inform future implementation efforts and make subsequent iterations of the enhanced screening tool better suited for use in a variety of practice settings. This paper describes our qualitative process and the feedback we received.

Methods

Participants

A total of 10 providers from 10 different clinical sites across the United States were recruited through purposive sampling to participate in structured interviews. Providers were included if they provided either primary care or specialty adolescent medicine care (eg, for reproductive, eating disorder, and gender concerns) to the adolescent age group. They were excluded if their clinic had participated in previous studies with the screening tool. Providers were recruited from primary care clinics within the University of Pittsburgh Medical Center network who had not participated in the previous study, and from a national organization adolescent medicine listserv where providers had previously expressed interest in trialing the clinical tool after Dr Radovic had replied to a member's question about existence of available computerized clinical tools.

Ethical Considerations

Participants provided verbal consent to participate in interviews and were compensated with a US $50 gift card. Participants and interview content were deidentified. This study was approved by the University of Pittsburgh institutional review board (STUDY21090136).

Procedure

Interested providers who were able to obtain organizational clearance were invited to trial the digitally enhanced mental health screening tool with actual patients, or if not, to informally explore a prototype of the tool on their own before the interview. All participants subsequently participated in a structured interview. Interview sessions were conducted by the first author through HIPAA-compliant Zoom (Zoom Communications) and comprised of a brief live demonstration of the screening tool followed by a 45-minute structured interview using a guide based on the Consolidated Framework for Implementation Research (CFIR) constructs [16] and informed by pilot work [13]. Table 1 shows the examples of interview questions.. The complete interview guide is available in Multimedia Appendix 1.

XSL•FO
RenderX

Table . Sample interview questions about a digitally enhanced mental health screening tool and their correspondence to the consolidated framework for implementation research.

CFIR[a] domain	Construct	Interview question
Inner setting	Compatibility	How well does the enhanced screening tool fit with existing screening practices in your clinic?
Process	Engaging	Who are the key influential individuals to get on board with the enhanced screening tool?
Intervention characteristics	Adaptability	What changes or alterations do you think would be needed to facilitate implementation and utility of the screening tool in your clinic?

[a]CFIR: Consolidated Framework for Implementation Research.

Coding and Qualitative Analysis

Interviews were audio-recorded, deidentified, and transcribed using Rev Human Transcription Services [17], removing identifiers and filler words (eg, "yeah"). Meeting notes were taken by the interviewer in real time for triangulation with eventual formal qualitative coding. Transcripts were coded by the first author following a template analysis approach [18] with NVivo software (version 14; Lumivero), using a prespecified codebook. The codebook was created according to a hybrid approach, using a deductive approach to generate high-level codes based on the CFIR Codebook Template [15], and an inductive approach to incorporate themes that arose from the data [19]. Subcodes were chosen based on topics recurring as key points across multiple interviews, and codes that did not fall within the CFIR framework were included on a case-by-case basis as relevant to future implementation of the screening tool. A research assistant coded data in parallel, and the 2 coders discussed and resolved discrepancies, with the availability of the senior author in case a resolution could not be found. The 2 coders had differences in organization of coding (for example, coding small sentence fragments versus coding complete blocks of text), but did not disagree on any content-level coding. All of these organizational discrepancies were resolved between the two coders and did not require escalation to the senior author. After consensus was reached between the two coders, the first author summarized and synthesized interview findings to present thematically. Participants were offered the opportunity to review findings presented in this manuscript and provide feedback for triangulation purposes. Participants did not suggest any corrections to the study's findings.

Results

Providers' feedback regarding the digitally enhanced mental health screening tool and its theoretical implementation in their respective practice settings focused on four major themes: (1) current screening practices, (2) perceived benefits of the tool, (3) barriers to implementation, and (4) structural or organizational considerations. We reached saturation around these major themes.

Description of Providers and Current Screening Practices

Of the providers contacted, 22 were invited by email to participate in interviews, 12 were interested in participating, and 10 ultimately completed interviews. All 10 providers viewed a prototype of the screening tool, and 3 of the providers attempted to pilot the tool in their respective clinics with patients but were unable to gain the required approval to do so. Of the 10 participants, 9 were physicians and 1 was a nurse practitioner. In addition, 8 of the providers specifically practiced adolescent medicine, while 2 providers worked in general pediatrics. In total, 7 providers practiced in hospital-based clinics associated with academic institutions, in a combination of primary care and adolescent specialty care settings. Furthermore, 1 provider worked in a college health clinic, and 2 providers practiced in community practices, with 1 working in a Federally Qualified Health Center.

In total, 6 of the 10 providers described using paper screeners for mental health, which clinic staff manually enter into the EHR, either by filling in a template with the patients' responses or scanning the screening result into the EHR as a media file. In addition, 2 providers used an electronic screening format, whereby patients answer questions electronically and these answers are automatically integrated into the EHR. Furthermore, 1 provider used a combination of paper and electronic screening modalities, and another reported no formalized mental health screening process.

Perceived Benefits of Digitally Enhanced Mental Health Screening Tool

Providers universally reported an uptick in adolescent depression and anxiety in recent years, particularly during the COVID-19 pandemic. Most providers supported mental health screening and treatment as high priority initiatives, with several reporting recent expansion of their practice's mental health services (eg, hiring new embedded mental health providers and partnering with telepsychiatry services). All providers perceived the digitally enhanced mental health screening tool would offer improvements on their current practices due to it being administered through mobile devices before appointments and eliciting more detailed information than currently used screeners. Providers generally had positive opinions about any new intervention that could increase efficiency of history-taking, allowing a larger proportion of appointment time to be devoted to management discussions. Some providers recognized potential benefits of the tool in assisting with diagnostic clarity due to its efficiency in eliciting additional information without relying on a longer appointment. Previewing the tool and the included measures inspired one provider to inquire about mania in an adolescent patient with depression:

XSL•FO
RenderX

> *I thought a lot about the mania [screener], because it's definitely not something that I traditionally screen for, unless I'm starting somebody on an SSRI…however, one of [my] patients screened positive for mania…I would not necessarily have thought to ask her those questions at that point, because we hadn't even really started talking about medication yet. So, it definitely did prompt me to [ask further questions. The patient then revealed a strong family history of bipolar disorder], and I was like, "Oh, okay. This is starting to make more sense." I think she actually does have bipolar type one.*

The tools' summarized results accompanied with additional symptom severity interpretation were recognized as efficiencies providers could use to identify mental health concerns at a glance and access direct links to current clinical care guidelines (ie, from the American Academy of Child and Adolescent Psychiatry and the American Academy of Pediatrics). One provider thought the tool would increase their comfort addressing mental health concerns in primary care, reflecting a lack of prior mental health training despite the need to address it regularly:

> *I wasn't trained on how to use SSRIs or even stimulant medication really, even though I knew I was going to primary care and I tried to get that training and now that's what I do a few times a day, which is kind of crazy.*

Providers proposed that if the tool is introduced as a standardized screening that all adolescents and caregivers within the practice are asked to complete, it could normalize mental health discussions while simultaneously providing psychoeducation. Adopting the screening tool into their clinical practice was mentioned by a provider as a way of "letting people know that this [answering questions about their mental health] is something okay to talk about even when they're not [currently symptomatic]," providing "a sense of safety and comfort for the patient." Another provider, who regularly precepts pediatric residents, appreciated that the tool automatically engages parents in conversation about mental health, stating that this is something trainees struggle to facilitate on their own. Another provider relayed the tool could help dismantle mental health stigma, as many parents show heightened concern navigating the lasting impacts of the COVID-19 pandemic on youth:

> *There's still a lot of stigma associated with mental health, but I think more and more people are getting more open, and I think a lot of parents will be very open to it right now, just given the current environment we're in, understanding that a lot of young people and teens, everybody needs help once in a while, and it's not something to be stigmatized.*

Providers thought the enhanced screening tool could be used in various patient visit types besides annual well child visits, including as a diagnostic aid for patients presenting with mental health chief complaints, given the tool's inclusion of several validated measures and specific questions regarding the patient's treatment preferences and perceived barriers to care; and as a college health clinic administered prescreening for transitional age youth upon entering college to facilitate parental involvement in students' transition to adult care.

Barriers to Implementation

The most frequently mentioned theme throughout provider interviews was that implementing a new digital screening, which is not embedded in the EHR would be exceedingly challenging, requiring clinical staff to take on burdensome tasks including scheduling and sending out survey links, monitoring responses, printing results for providers, and uploading results into the EHR. This concern was in the setting of clinics already struggling with staffing shortages, and the perspective that—while highly important—mental health screening is only one facet of well-child visits, which already involve multiple administrative tasks. For a new technology intervention to be adopted in their clinic, providers agreed it would need to show clear clinical benefit while simultaneously reducing burden on clinic staff. As such, EHR integration was seen as a necessity for successful implementation of the tool across clinical settings.

Providers expressed concern about the length of time needed to complete the screening survey, both from a workflow and user perspective. Providers suspected patients and their caregivers may be unwilling to complete the entire tool, noting the form could be particularly time-consuming for individuals with low literacy levels. Providers shared that if a patient or caregiver did not have a chance to complete the survey before their visit, it may negatively affect the duration of time patients spend in exam rooms, appointment length, room turnover, and patient wait time.

Nearly all providers endorsed safety and liability concerns related to screening for suicidality. Despite the screening tool being designed to limit availability of the survey to the clinic's hours of operation as described above, providers worried about setting up foolproof protocols for a notification system for completed screens, and assigning responsibility to staff for reviewing screens and identifying patients at risk in a timely manner to avoid positive screens being missed, or loss to follow-up:

> *The other thing that I think about too [is] getting a positive screen [from] someone who fills that out from home and then they don't come to the appointment. That could be a big deal because it's going to be our patients with mental health disorders. Generally, if a teenager has a mental health disorder, there's a reasonable chance that a parent has a mental health disorder. If they have a mental health disorder that's not treated, which is a lot of mental health disorders, it makes them less likely to follow through with things like appointments. So then you have a patient who's suicidal and is out there, or who's significantly affected and is out there, and then you can't get ahold of them because of that stuff. That makes me scared. Now is that better than just being dumb to it? Maybe it's better, but it makes me nervous.*

Similarly, providers worried about balancing measures and the unintended harms of screening broadly for mental health without adequate time allotted during the appointment to address

XSL•FO
RenderX

concerns that the screening tool may identify. A provider was concerned about the screening tool's potential to uncover otherwise undetected mental health concerns in patients without having the resources such as behavioral health staff or referral resources to respond appropriately.

Providers raised concerns about patient privacy when using a third-party product to obtain highly sensitive protected health information. In addition, providers worried about the potential for caregivers to access a patient's portal or survey or even to hover over their child while they complete the survey, which could in turn cause patients to answer questions inaccurately or not engage fully with the screening tool.

Accessibility was a major theme across all interviews, particularly for providers who worked with underserved populations. Providers stated it was important for the screening tool to be available in other languages, noting that if a child is fluent in English, their caregiver may be less fluent and require a version in their native language to obtain accurate collateral information. Some also wanted the screening tool to be available in other modalities, such that it would be accessible to those with poor literacy or learning disabilities. As with any new technology intervention, questions arose about health equity and inaccessibility for those without mobile devices or reliable internet access. There was general consensus that, while new technologies have potential to mitigate certain health care disparities, they also have the potential to exacerbate others by virtue of being inaccessible to many underserved populations, which is consistent with concerns in the existing literature [20].

Structural and Organizational Considerations

The most frequently mentioned requirements for getting leadership on board with implementing a new intervention were that the intervention is efficient, enhances quality of care, increases patient satisfaction, and has proof of clinical benefit. It was also important that the intervention be simple and easy for staff to use, so as not to burden staff or negatively impact clinic workflow.

Providers identified a major barrier to implementing new technologies as lack of transparency about the processes in their institution**,** with many providers unable to identify who they would need to obtain permission from within their organization to implement a new technology within their clinic. Providers described complicated networks of clinical leadership, business executives, and representatives from information technology and their respective EHRs that all must collaborate to approve new technologies. A provider reported that they previously attempted to implement a new technology intervention in their hospital-based clinic but grew frustrated and abandoned the initiative after numerous unfruitful meetings and emails with the administration.

Some providers also mentioned public insurance documentation requirements as a barrier to using a third-party tool, as public insurance plans often require routine screenings be completed using specific forms. Providers worried some patients would have to complete the same screening questions (eg, Patient Health Questionnaire-9) twice for the same office visit—once in the enhanced screening tool, and a second time on the form required for reimbursement.

Recommended Alterations to the Digitally Enhanced Mental Health Screening Tool

Integration of the enhanced screening tool into the EHR and automation of survey creation and distribution were thought to be the strongest facilitators necessary for implementing the tool in pediatric clinics. All 10 providers were in favor of redesigning the survey to offer all patients a subset of initial questions, and only the full screen to those patients who answered positively to those initial questions, as a means for increasing the screening tool's efficiency.

Providers agreed that customization options such as ability to choose which screening tools to include would make the intervention more appealing to a variety of practice settings. In addition to selecting or deselecting currently available measures, providers were interested in adding screening tools for the following (in order of frequency mentioned): disordered eating, sleep, adverse childhood events, social determinants of health, attention-deficit/hyperactivity disorder (ADHD), academic distress, and demographic questions about sexual orientation and gender identity.

Discussion

Principal Findings

This preimplementation study aimed to elicit provider perspectives regarding proposed implementation of a digitally enhanced mental health screening tool in their respective clinical settings to inform future implementation strategies and alterations to the screening tool. While all providers stated that improving mental health screening was a high priority in their clinic and believed that the screening tool had the potential to improve such practices, many providers expressed that the screening tool may not be feasible to implement in its current form, most notably due to the tool being a third-party application without EHR integration.

As new technologies for health care information gathering, evaluation, and intervention delivery become an increasingly popular area of interest, there is a simultaneous need for improved systems interoperability to allow for exchange of information between new technology tools and existing EHRs in order for such technologies to be practical for use in real clinical settings [21]. While EHRs show provider notes and metrics from previous medical encounters, they have limited ability to pull in data and services from third-party tools that have potential to enhance quality of health care, decrease health care costs, and improve patient outcomes. The lack of connectivity between different systems and software presents a nearly impermeable barrier to implementation of new technologies, highlighting the dire need for an application-based component to health information technology to allow for adaptation of new technologies and seamless integration with existing EHRs [22]. Several large research-based hospital systems have successfully deployed questionnaires for patient reported outcomes, the results of which are directly incorporated in the electronic medical record [23]. Such initiatives can serve

XSL•FO
RenderX

as a model for similar integration of screening tools such as ours and promote organizational buy-in by demonstrating that wide-scale implementation is feasible. The findings from this study further emphasize the need for enhanced systems interoperability in the health care space, and exemplify another way that improvement in this area can facilitate implementation of innovative health care technologies. Beyond the logistical barriers to implementation of the screening tool, providers identified administrative buy-in as another major barrier. Providers reported hesitancy at the administrative level due to new challenges that such technologies present, such as data privacy concerns and questions about liability in a relatively uncharted digital landscape. The findings in this study further reinforce the already identified need to address such issues at a structural or policy level [24] so that individual providers can more easily and independently trial and use new technology tools that have the potential to provide clinical benefit and improve patient outcomes without requiring system-wide changes.

Limitations

While homogeneity within our sample allowed thematic saturation to be reached, our findings may be less generalizable to all pediatric primary care settings due to our sample being small and self-selective, comprised of providers who may be more interested in improving adolescent behavioral health services as they expressed interest in learning how a digitally enhanced screening tool could augment their existing clinical practices. While not all primary care practices will have clinician leaders interested and willing to take on the burdensome task of pioneering new interventions, early adopters are critical to intervention development and implementation such that these new tools can be optimized, scalable and feasible to disseminate more widely [25,26]. Provider feedback was also limited by the inability to use the screening tool in vivo with real patients in real time, again underscoring the importance of pilot work and a need for greater consideration regarding the process of trialing new interventions in the community (eg, flexible method designs, quasi-experimental and nonrandomized approaches that account for the need to maintain efficient workflow in practice settings).

Conclusion

While technology interventions can help address the need for improved behavioral health support in primary care settings, numerous barriers to implementation of such tools exist. This study adds to the literature showing a need for improved health information systems interoperability for innovative technologies to be used to their greatest potential in real-world settings.

Acknowledgments

This study was funded by the University of Pittsburgh Medical Center Competitive Medical Research Fund, the ETUDES Center at the University of Pittsburgh (1P50MH115838-01), and the University of Pittsburgh School of Medicine Roth Fellow Award.

Data Availability

The data from this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screening Wizard provider interview guide.
[DOCX File, 19 KB - formative_v9i1e67624_app1.docx]

References

1. Ros-DeMarize R, Chung P, Stewart R. Pediatric behavioral telehealth in the age of COVID-19: Brief evidence review and practice considerations. Curr Probl Pediatr Adolesc Health Care 2021 Jan;51(1):100949. [doi: 10.1016/j.cppeds.2021.100949] [Medline: 33436319]
2. 2022 NSDUH annual national report. : SAMHSA; 2022 URL: https://www.samhsa.gov/data/report/2022-nsduh-annual-national-report [accessed 2024-10-18]
3. Hua LL, Lee J, Rahmandar MH, Sigel EJ, COMMITTEE ON ADOLESCENCE, COUNCIL ON INJURY, VIOLENCE, AND POISON PREVENTION. Suicide and suicide risk in adolescents. Pediatrics 2024 Jan 1;153(1):e2023064800. [doi: 10.1542/peds.2023-064800] [Medline: 38073403]
4. Zuckerbrot RA, Cheung A, Jensen PS, Stein REK, Laraque D, GLAD-PC STEERING GROUP. Guidelines for Adolescent Depression in Primary Care (GLAD-PC): Part I. Practice preparation, identification, assessment, and initial management. Pediatrics 2018 Mar;141(3):e20174081. [doi: 10.1542/peds.2017-4081] [Medline: 29483200]
5. Cheung AH, Zuckerbrot RA, Jensen PS, Laraque D, Stein REK, GLAD-PC STEERING GROUP. Guidelines for Adolescent Depression in Primary Care (GLAD-PC): Part II. Treatment and ongoing management. Pediatrics 2018 Mar;141(3):e20174082. [doi: 10.1542/peds.2017-4082] [Medline: 29483201]

XSL•FO
RenderX

6. Suicide: blueprint for youth suicide prevention. American Academy of Pediatrics. 2021. URL: https://www.aap.org/en/patient-care/blueprint-for-youth-suicide-prevention/ [accessed 2024-10-18]
7. US Preventive Services Task Force, Mangione CM, Barry MJ, et al. Screening for depression and suicide risk in children and adolescents: US preventive services task force recommendation statement. JAMA 2022 Oct 18;328(15):1534-1542. [doi: 10.1001/jama.2022.16946] [Medline: 36219440]
8. Asnaani A, Narine K, Suzuki N, et al. An innovative mobile game for screening of pediatric PTSD: a study in primary care settings. J Child Adolesc Trauma 2021 Sep;14(3):357-366. [doi: 10.1007/s40653-020-00300-6] [Medline: 34471454]
9. McNeely J, Mazumdar M, Appleton N, et al. Leveraging technology to address unhealthy drug use in primary care: effectiveness of the substance use screening and intervention tool (SUSIT). Subst Abus 2022;43(1):564-572. [doi: 10.1080/08897077.2021.1975868] [Medline: 34586976]
10. Spence EE, Prabhakar P, Grace J, Fulda KG, Thompson EL, Ondersma SJ. Development and implementation of tablet-based screening for interpersonal violence in primary care settings. Health Care Women Int 2022 Dec;43(12):1503-1509. [doi: 10.1080/07399332.2022.2043861] [Medline: 35235498]
11. Peahl AF, Rubin-Miller L, Paterson V, et al. Understanding social needs in pregnancy: prospective validation of a digital short-form screening tool and patient survey. AJOG Glob Rep 2023 Feb;3(1):100158. [doi: 10.1016/j.xagr.2022.100158] [Medline: 36922957]
12. Radovic A, Anderson N, Hamm M, et al. Perceived facilitators of and barriers to implementation of a decision support tool for adolescent depression and suicidality screening: focus group and interview study (preprint). JMIR Mental Health. Preprint posted online on 2020. [doi: 10.2196/preprints.26035]
13. Radovic A, Coren MA, Wecht S, et al. Detecting and managing adolescent depression in primary care: a pilot randomized controlled trial. Procedia Comput Sci 2022;206:81-91. [doi: 10.1016/j.procs.2022.09.087]
14. Taylor SP, Short RT, Asher AM, Taylor B, Beidas RS. A rapid pre-implementation evaluation to inform a family engagement navigator program during COVID-19. Implement Sci Commun 2020 Dec 9;1(1):110. [doi: 10.1186/s43058-020-00098-2] [Medline: 33298192]
15. CFIR guide. Consolidated Framework for Implementation Research. 2024. URL: https://cfirguide.org/guide/app [accessed 2024-10-18]
16. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 7;4(1):50. [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
17. Rev. URL: https://www.rev.com/ [accessed 2025-03-28]
18. King N. Essential Guide to Qualitative Methods in Organizational Research: SAGE Publications Ltd; 2004. [doi: 10.4135/9781446280119]
19. Hamilton AB, Finley EP. Qualitative methods in implementation research: an introduction. Psychiatry Res 2019 Oct;280:112516. [doi: 10.1016/j.psychres.2019.112516] [Medline: 31437661]
20. Ralston AL, Andrews AR, Hope DA. Fulfilling the promise of mental health technology to reduce public health disparities: review and research agenda. Clinical Psychology: Science and Practice 2019;26(1). [doi: 10.1111/cpsp.12277]
21. O'Connor S. What is interoperability, and why is it important? Advanced Data Systems Corporation. 2017 May 30. URL: https://www.adsc.com/blog/what-is-interoperability-and-why-is-it-important [accessed 2024-10-18]
22. Mandl KD, Mandel JC, Kohane IS. Driving innovation in health systems through an apps-based information economy. Cell Syst 2015 Jul;1(1):8-13. [doi: 10.1016/j.cels.2015.05.001] [Medline: 26339683]
23. Papuga MO, Dasilva C, McIntyre A, Mitten D, Kates S, Baumhauer JF. Large-scale clinical implementation of PROMIS computer adaptive testing with direct incorporation into the electronic medical record. Health Syst (Basingstoke) 2018;7(1):1-12. [doi: 10.1057/s41306-016-0016-1] [Medline: 31214335]
24. Mohr DC, Azocar F, Bertagnolli A, et al. Banbury forum consensus statement on the path forward for digital mental health treatment. Psychiatr Serv 2021 Jun;72(6):677-683. [doi: 10.1176/appi.ps.202000561] [Medline: 33467872]
25. Green L, Gottlieb N, Parcel G. Diffusion theory extended and applied. Advances in Health Education and Promotion 1991.
26. Van Houtven CH, Drake C, Malo TL, et al. Ready, set, go! The role of organizational readiness to predict adoption of a family caregiver training program using the Rogers' diffusion of innovation theory. Implement Sci Commun 2023 Jun 19;4(1):69. [doi: 10.1186/s43058-023-00447-x] [Medline: 37337208]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
CFIR: Consolidated Framework for Implementation Research
EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act
PCP: primary care provider
PTSD: posttraumatic stress disorder
SW: Screening Wizard

XSL•FO
RenderX

Edited by A Mavragani; submitted 23.10.24; peer-reviewed by H Kautz, R Pearson; revised version received 26.02.25; accepted 11.03.25; published 10.04.25.

Please cite as:
Coren MA, Lindhiem O, Angus AR, Toevs EK, Radovic A
Provider Perspectives on Implementing an Enhanced Digital Screening for Adolescent Depression and Suicidality: Qualitative Study
JMIR Form Res 2025;9:e67624
URL: https://formative.jmir.org/2025/1/e67624
doi:10.2196/67624

© Morgan A Coren, Oliver Lindhiem, Abby R Angus, Emma K Toevs, Ana Radovic. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Reliability of Average Daily Steps Measured Through a Consumer Smartwatch in Parkinson Disease Phenotypes, Stages, and Severities: Cross-Sectional Study

Edoardo Bianchini[1,2], MD; Domiziana Rinaldi[1], MD; Lanfranco De Carolis[1], MD; Silvia Galli[1], MD; Marika Alborghetti[1], MD; Clint Hansen[2,3], PhD; Antonio Suppa[4,5], MD; Marco Salvetti[1,5], Prof Dr Med; Francesco Ernesto Pontieri[1], Prof Dr Med; Nicolas Vuillerme[2,6], PhD

[1]Department of Neuroscience, Mental Health and Sensory Organs (NESMOS), Sapienza University of Rome, Rome, Italy

[2]AGEIS, Université Grenoble Alpes, Grenoble, France

[3]Department of Neurology, Kiel University, Kiel, Germany

[4]Department of Human Neurosciences, Sapienza University of Rome, Rome, Italy

[5]IRCCS Neuromed Institute, Pozzilli, Italy

[6]Institut Universitaire de France, Paris, France

Corresponding Author:
Edoardo Bianchini, MD
Department of Neuroscience, Mental Health and Sensory Organs (NESMOS), Sapienza University of Rome, Rome, Italy

Abstract

Background: Average daily steps (avDS) could be a valuable indicator of real-world ambulation in people with Parkinson disease (PD), and previous studies have reported the validity and reliability of this measure. Nonetheless, no study has considered disease phenotype, stage, and severity when assessing the reliability of consumer wrist-worn devices to estimate daily step count in unsupervised, free-living conditions in PD.

Objective: This study aims to assess and compare the reliability of a consumer wrist-worn smartwatch (Garmin Vivosmart 4) in counting avDS in people with PD in unsupervised, free-living conditions among disease phenotypes, stages, and severity groups.

Methods: A total of 104 people with PD were monitored through Garmin Vivosmart 4 for 5 consecutive days. Total daily steps were recorded and avDS were calculated. Participants were dichotomized into tremor dominant (TD; n=39) or postural instability and gait disorder (PIGD; n=65), presence (n=57) or absence (n=47) of tremor, and mild (n=65) or moderate (n=39) disease severity. Based on the modified Hoehn and Yahr scale (mHY), participants were further dichotomized into earlier (mHY 1 - 2; n=68) or intermediate (mHY 2.5 - 3; n=36) disease stages. Intraclass correlation coefficient (ICC; 3,k), standard error of measurement (SEM), and minimal detectable change (MDC) were used to evaluate the reliability of avDS for each subgroup. The threshold for acceptability was set at an ICC ≥0.8 with a lower bound of 95% CI ≥0.75. The 2-tailed Student *t* tests for independent groups and analysis of 83.4% CI overlap were used to compare ICC between each group pair.

Results: Reliability of avDS measured through Garmin Vivosmart 4 for 5 consecutive days in unsupervised, free-living conditions was acceptable in the overall population with an ICC of 0.89 (95% CI 0.85 - 0.92), SEM below 10%, and an MDC of 1580 steps per day (27% of criterion). In all investigated subgroups, the reliability of avDS was also acceptable (ICC range 0.84 - 0.94). However, ICCs were significantly lower in participants with tremor (*P*=.03), with mild severity (*P*=.04), and earlier stage (*P*=.003). Moreover, SEM was below 10% in participants with PIGD phenotype, without tremor, moderate disease severity, and intermediate disease stage, with an MDC ranging from 1148 to 1687 steps per day (18% - 25% of criterion). Conversely, in participants with TD phenotype, tremor, mild disease severity, and earlier disease stage, SEM was >10% of the criterion and MDC values ranged from 1401 to 2263 steps per day (30% - 33% of the criterion).

Conclusions: In mild-to-moderate PD, avDS measured through a consumer smartwatch in unsupervised, free-living conditions for 5 consecutive days are reliable irrespective of disease phenotype, stage, and severity. However, in individuals with TD phenotype, tremor, mild disease severity, and earlier disease stages, reliability could be lower. These findings could facilitate a broader and informed implementation of avDS as an index of ambulatory activity in PD.

(JMIR Form Res 2025;9:e63153) doi:10.2196/63153

XSL•FO
RenderX

KEYWORDS

gait; Parkinson disease; phenotype; wearable sensors; smartwatch; step count; reliability; activity monitor; digital health technology; digital outcome measures; wearable; mHealth; motor; quality of life; fall; posture; mobile health

Introduction

Walking is a fundamental motor ability, and it is pivotal for functional independence and social well-being [1]. Gait disturbances are common in people with Parkinson disease (PD), and include shuffling gait, shortened step length, altered automaticity, decreased arm swing, and freezing of gait [2]. These represent a particularly disabling group of symptoms, significantly hampering the quality of life of people with PD [3] and increasing the risk of falls [4].

Daily steps are an easy-to-collect and useful measure of ambulatory activity and mobility [5]. Although this parameter could not provide details regarding subtle gait features, several evidence linked a reduced daily step count to overall mortality risk [6-9], as well as to a range of health conditions, such as dementia [10] and oncological and cardiovascular diseases [11,12]. Previous studies have also reported a negative correlation between daily steps and disease severity in PD [13] and proposed a minimum daily step goal of 4200 to match with physical activity recommendations in the early stages of the disease [14].

Wearable devices, including smartwatches, could represent a useful option to estimate daily steps in an unobtrusive, ecological way [15]. Moreover, wearables are widely available on the market, easily used by the general population, and enable unobtrusive continuous long-term data collection [16]. However, since these devices are usually tested in healthy populations, knowledge of the validity and reliability of collected data is generally limited when applied to different groups of patients. Therefore, a growing body of literature focused on the validity and reliability of consumer wearable devices for step counting, with generally positive results [17-19].

In PD, motor and gait manifestations could render step detection and step count challenging and hence significantly diminish the validity and reliability of device algorithms [20-22]. Nevertheless, a prior study by Ginis et al [23] demonstrated a good criterion validity in estimating average daily steps (avDS) of 2 Fitbit devices (Fitbit Alta and Fitbit Inspire 3) in 28 people with PD in a real-life setting, compared to a research-grade device (Dynaport Movemonitor, McRoberts, NL).

Similarly, a prior study from our group involving 47 individuals with PD demonstrated a good criterion validity in step counting using a consumer smartwatch (Garmin Vivosmart 4), when worn on the side least affected by the disease and under well-controlled pharmacological conditions in a supervised, in-clinic setting [20].

Besides criterion validity, we recently demonstrated that optimal reliability of avDS recorded by Garmin Vivosmart 4 in real-life conditions could be achieved if the smartwatch is worn for a minimum of 4 days [24].

However, the clinical heterogeneity of PD when evaluating metrological characteristics of step-counting devices has been neglected so far. Indeed, the clinical presentation of PD is highly variable among individuals, thus, significant efforts have been made to identify distinct clusters and subtypes [25]. Several classifications have been proposed over the years from clinical-based [26-28] to more recent biomarker-based classifications [29-31]. In this regard, one of the most used classifications distinguishes PD with predominant features of tremor (ie, tremor dominant [TD]) or gait, posture, and balance issues (ie, postural instability and gait disorder [PIGD]) [26-28] based on subitems scores of the Unified Parkinson's Disease Rating Scale (UPDRS) and its revision by the Movement Disorder Society (MDS-UPDRS) [32]. This classification could be relevant when measuring avDS in PD since tremors could increase the noise-to-signal ratio, making step detection more challenging [21]. Similarly, in PIGD individuals, the higher degree of gait alterations could alter the performance of step-detection algorithms [20,22].

Symptom severity and disease stage could also represent other relevant parameters to be considered when assessing the reliability of any wearable devices in counting avDS. Indeed, with disease progression, gait features increasingly deviate from normality [2,33], and tremor and bradykinesia could further alter the spatiotemporal and kinematic characteristics of walking, dampening, in turn, algorithm performance in step detection [20-22].

Nevertheless, no study to date considered disease phenotype, stage, and symptom severity when assessing the reliability of consumer wrist-worn devices for step counting in unsupervised, free-living conditions in mild-to-moderate people with PD. This study was hence specifically designed to address this issue. We hypothesized that reduced reliability might be observed in TD individuals and people with PD with more severe symptoms and in more advanced disease stages due to the aforementioned increased signal noise due to tremor, the higher degree of motor symptoms, and the more marked gait alterations. In addition, since no previous study specifically investigated the criterion validity of Garmin Vivosmart 4 in estimating avDS, we performed a proof-of-concept experiment in a subgroup of participants. Details of methods and results of this latter experiment are presented in Multimedia Appendix 1.

Methods

Population

Participants were consecutively screened and recruited during scheduled visits at the Movement Disorder Outpatient Service of the Sant'Andrea University Hospital (Rome, Italy) in the period between March 2023 and March 2024. Inclusion criteria were: (1) diagnosis of idiopathic PD according to the MDS criteria (Postuma et al [34]); (2) aged 18 years or older; (3) disease stage <4 according to the modified Hoehn and Yahr scale (mHY) [35]; (4) classification as TD or PIGD according to Stebbins et al [27]; and (5) stable medication in the 4 weeks before the experimental procedure. Exclusion criteria were: (1)

XSL•FO
RenderX

cognitive impairment, defined by Montreal Cognitive Assessment [36] score <21; and (2) orthopedic, rheumatologic, or systemic conditions affecting mobility as judged by the assessor.

Ethical Considerations

This cross-sectional study was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. Approval was granted by the local Ethical Committee of Sapienza, University of Rome, Italy (0372/2022). Data collection and processing followed the current European regulations for data protection. All participants provided written informed consent before the beginning of measurements. All data were deidentified. Participants did not receive any form of compensation.

Demographic and Clinical Data

Participants were evaluated during scheduled visits. Demographics and anthropometric measures (including age, sex, weight, height, and BMI) were collected. Disease duration and disease stage according to mHY and levodopa equivalent daily dose [37] were also collected. MDS-UPDRS [32] part III was used to assess motor symptoms severity.

Participants were divided into 4 subgroup pairs. Based on MDS-UPDRS part II and III scores, participants were classified into TD or PIGD disease subtypes according to Stebbins et al [27]. To evaluate the effect of tremor presence on device reliability, participants were also classified as those with and without tremor based on a score of ≥1 at item 2.10 of MDS-UPDRS part II. Concerning disease severity, participants were grouped into those with mild or moderate disease severity based on the MDS-UPDRS score as proposed by Martínez-Martín et al [38]. Similarly, participants were dichotomized in earlier (mHY 1 - 2) or intermediate (mHY 2.5 - 3) stages, based on mHY score.

Experimental Procedure

Participants received the smartwatch Garmin Vivosmart 4 after the visit and were instructed to wear it at home for a minimum of 5 consecutive days, including at least 1 weekend day, on the wrist of the body side least affected by the disease [20]. No reminders or further instructions were provided to participants during registration. We chose a 5-day period since we demonstrated previously that a minimum of 4 days of monitoring is needed to reliably estimate daily step count in PD [24]. Each smartwatch was configured according to the producer's recommendations and participants were asked to perform daily activities as usual. After 5 days, participants returned the smartwatch. The total daily number of steps for each day was recorded and avDS were calculated [24]. Compliance was assessed based on the participants' dashboard data. The device recognizes that it is worn through heart rate and inertial motion unit signal. We considered all recording days with >80% wear time while awake and no interruption of device use greater than 3 hours to be valid. Details of methods of the proof-of-concept experiment regarding criterion validity of Garmin Vivosmart 4 are presented in Multimedia Appendix 2.

Data and Statistical Analysis

The statistical analyses were performed using JASP (version 0.18.3.0; JASP Team), R (version 4.0.3; R Core Team), and RStudio (version 2023.12.0+369; R Foundation for Statistical Computing) for Windows. Descriptive statistics were calculated for the examined variables. The normality of distributions was assessed by histogram and residual plots inspection.

To evaluate the relative reliability for the 5-day monitoring period in the overall population and each subgroup, a 2-way intraclass correlation coefficient (ICC) with a fixed set of raters and averaged ratings was used (ICC (3,k), where k was the number of days of measurement), together with a custom R script. The following reference cut-off values for ICC interpretation were used [39]: excellent: >0.90; good: 0.75 - 0.90; moderate: 0.50 - 0.75; and poor: <0.50. The a priori threshold for acceptable ICC was set at a point estimate ≥0.80 with a lower bound of 95% CI ≥0.75 in accordance with a previous study from our group [24].

To compare ICCs between the 4 subgroup pairs, 2 methods were applied. First, standard errors and point estimates of ICCs were used to compute *t* statistics and perform 2-tailed independent groups Student *t* tests. Second, the CI overlap between each group pair was graphically and numerically assessed. Nonoverlapping CIs were considered indicators of significantly different ICCs [40]. Previous evidence underscored that a 95% CI overlap assessment could inflate the risk of type II error and suggested that an 83.4% CI could be a more powerful option [40-42]. Therefore, we adopted this method for CI overlap evaluation.

Standard error of measurement (SEM) and minimal detectable change (MDC) with a CI of 95% were used to compute the absolute reliability for the 5-day recordings in the overall population and each subgroup [43]. SEM and MDC were reported as absolute value and percentage of criterion measure (SEM% and MDC%, respectively). The criterion was the avDS count derived from the 5 days. For all analyses, the significance threshold was set at $\alpha<.05$. All data were reported as mean (SD) or median (IQR) for numerical data and n (%) for categorical variables.

Results

Overview

A total of 104 people with PD were included in the study. All participants were monitored through Garmin Vivosmart 4 at home for a period of 5 consecutive days. No participants or days were excluded based on the prespecified compliance criteria. Participants took on average 5923 (SD 3014) daily steps, ranging from 357 to 12,620. Details of demographic, anthropometric, and clinical variables of the study population are shown in Table 1. The overlap between the subgroups is shown in Multimedia Appendix 1.

Overall, the results of the proof-of-concept experiment suggest that the smartwatch Garmin Vivosmart 4 is valid and decently accurate in estimating avDS in PD. Details are presented in the Multimedia Appendix 2.

XSL•FO
RenderX

Table . Demographic, anthropometric, and clinical characteristics of the study population.

	Overall (N=104)	TD[a] (n=39)	PIGD[b] (n=65)	Tremor (n=57)	No tremor (n=47)	Mild (n=65)	Moderate (n=39)	mHY[c] 1 - 2 (n=68)	mHY 2.5 - 3 (n=36)
Age (years), mean (SD)	68.0 (8.4)	66.4 (9.0)	68.9 (8.0)	69.5 (7.9)	66.7 (8.8)	66.8 (8.7)	69.9 (7.7)	65.3 (7.8)	73.0 (7.3)
Height (cm), mean (SD)	171 (9.0)	173 (7.6)	170 (9.6)	170 (9.6)	172 (8.4)	173 (9.3)	168 (7.6)	173 (8.1)	168 (9.8)
Weight (kg), mean (SD)	75.7 (13.1)	76.7 (13.0)	75.1 (13.2)	75.0 (12.1)	76.3 (13.9)	77.2 (13.7)	73.2 (11.8)	77.7 (13.8)	71.9 (10.8)
BMI (kg/m^2), mean (SD)	25.7 (3.4)	25.5 (3.6)	25.8 (3.4)	25.8 (3.0)	25.7 (3.8)	25.7 (3.5)	25.8 (3.4)	25.9 (3.7)	25.4 (3.0)
Sex (female), n (%)	34 (33)	11 (28)	23 (35)	19 (33)	15 (32)	22 (34)	12 (31)	22 (32)	12 (33)
Disease duration (years), mean (SD)	6.4 (4.4)	5.2 (4.4)	7.0 (4.4)	7.2 (3.9)	5.7 (4.8)	5.7 (4.5)	7.5 (4.2)	5.3 (4.3)	8.4 (4.1)
LEDD[d] (mg), mean (SD)	553 (302)	418 (247)	634 (304)	623 (271)	495 (316)	489 (289)	659 (296)	453 (248)	741 (307)
mHY, median (IQR)	2 (2-2.5)	2 (1-2)	2 (2-2.5)	2 (2-2.5)	2 (2-2)	2 (2-2)	2.5 (2-3)	2 (2-2)	2.5 (2.5 - 3)
MDS-UPDRS-III[e], median (IQR)	27 (21 - 32)	26 (18 - 31)	29 (22 - 33)	29 (22 - 33)	26 (21 - 32)	23 (19-29)	33 (29 - 37)	23 (18 - 29)	33 (29 - 36)
avDS[f], mean (SD)	5923 (3014)	6654 (2733)	5485 (3109)	4594 (2612)	7020 (2898)	6512 (2857)	4942 (3049)	6838 (2908)	4195 (2419)

[a]TD: tremor dominant.

[b]PIGD: postural instability and gait disorder.

[c]mHY: modified Hoehn and Yahr scale.

[d]LEDD: levodopa equivalent daily dose.

[e]MDS-UPDRS III: Movement Disorder Society Unified Parkinson's Disease Rating Scale part III.

[f]avDS: average daily steps.

Reliability of avDS in People With PD

Relative Reliability

AvDS collected during 5 consecutive days showed a level of relative reliability above the threshold of acceptability, as indicated by an ICC point estimate of ≥0.80 and a lower 95% CI limit of ≥0.75, in the overall population and all subgroups. Moreover, daily step count showed excellent reliability in PD in the intermediate disease stage. Details of ICC and CI limits are shown in Table 2.

XSL•FO
RenderX

Table . ICC[a] (3,k) values with 95% CI for the overall population and each subgroup[b].

	Overall (N=104)	TD[c] (n=39)	PIGD[d] (n=65)	Tremor (n=57)	No tremor (n=47)	Mild (n=65)	Moderate (n=39)	mHY[e] 1 - 2 (n=68)	mHY 2.5 - 3 (n=36)
ICC (3,k)	0.888	0.854	0.899	0.838	0.914	0.856	0.919	0.839	0.939
Lower 95% CI	0.850	0.767	0.855	0.760	0.868	0.793	0.871	0.769	0.900
Upper 95% CI	0.919	0.916	0.933	0.896	0.947	0.905	0.953	0.892	0.966
Lower 83.4% CI	N/A[f]	0.797	0.869	0.786	0.883	0.814	0.887	0.792	0.914
Upper 83.4% CI	N/A	0.901	0.925	0.881	0.939	0.893	0.945	0.878	0.959

[a]ICC: intraclass correlation coefficient.

[b]For subgroups, 83.4% CI to assess intervals overlap are also reported.

[c]TD: tremor dominant.

[d]PIGD: postural instability and gait disorder.

[e]mHY: modified Hoehn and Yahr scale.

[f]Not applicable.

Absolute Reliability

AvDS showed an SEM below 10% in the overall population with an MDC of 1580 (26.7% of the criterion). AvDS also showed an SEM below 10% in the PIGD disease subtype, in participants without tremor, with a moderate disease severity, and in an intermediate disease stage with an MDC ranging from 1148 to 1687 steps per day (18% to 25% of criterion). Conversely, in the TD disease subtype, in participants with tremor, with mild disease severity and in an early disease stage, SEM was >10% of criterion and MDC values ranged from 1401 to 2263 steps per day (30% to 33% of the criterion). Details of SEM and MDC are shown in Table 3.

Table . Absolute and percentage values of SEM[a] and MDC[b] for the overall population and each subgroup.

	Overall (N=104)	TD[c] (n=39)	PIGD[d] (n=65)	Tremor (n=57)	No tremor (n=47)	Mild (n=65)	Moderate (n=39)	mHY[e] 1 - 2 (n=68)	mHY 2.5 - 3 (n=36)
SEM	570	742	495	506	609	701	414	817	267
SEM%	9.6	11.1	9.0	11.0	8.7	10.8	8.4	11.9	6.4
MDC	1580	2056	1372	1401	1687	1944	1148	2263	741
MDC%	26.7	30.9	25.0	30.5	24.0	29.8	23.2	33.1	17.7

[a]SEM: standard error of measurement.

[b]MDC: minimal detectable change.

[c]TD: tremor dominant.

[d]PIGD: postural instability and gait disorder.

[e]mHY: modified Hoehn and Yahr scale.

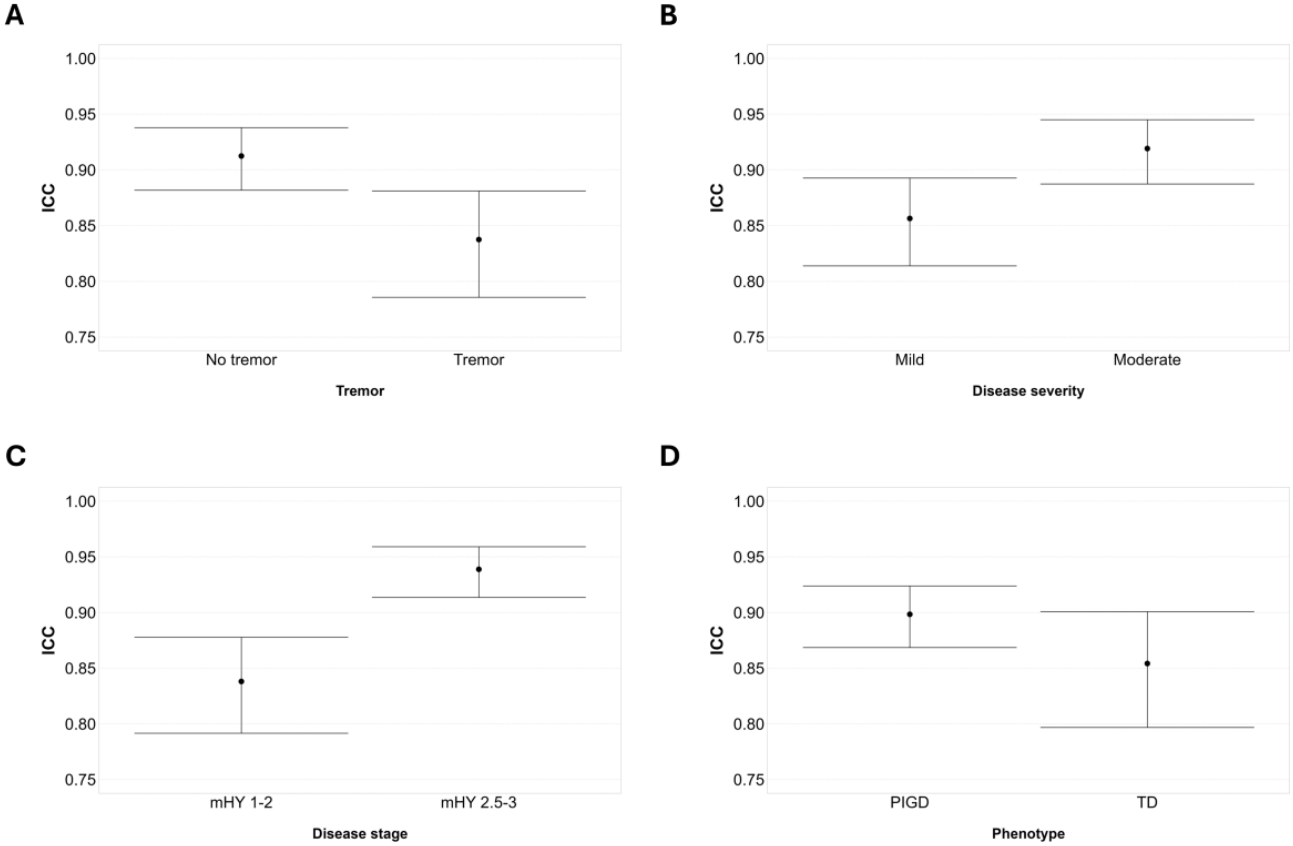
Reliability Comparison Between Subgroups

When comparing ICCs between subgroups pairs, 2-tailed Student *t* test for independent groups showed a significant difference between participants with and without tremor (t_{102}=1.897; *P*=.03), between PD with mild and moderate disease severity (t_{102}=1.765; *P*=.04), and between individuals in early and intermediate disease stage (t_{102}=2.817; *P*=.003). Conversely, no significant difference was found between TD and PIGD participants (t_{102}=1.048; *P*=.15).

The analysis of 83.4% CI showed no overlap between interval limits between participants with and without tremor (Figure 1A), in early and intermediate disease stage (Figure 1C), and only a negligible overlap between individuals with mild and moderate disease severity (Figure 1B). Conversely, a degree of overlap of the two 83.4% CIs was observed between TD and PIGD participants (Figure 1D). Details of 83.4% CI limits are shown in Table 2.

XSL•FO
RenderX

Figure 1. ICC and 83.4% CI comparison for each subgroup pair. (A) Participants with reported presence or absence of tremor. (B) Participants with mild and moderate disease severity as indicated by MDS-UPDRS. (C) Participants with earlier (mHY 1 - 2) and intermediate (mHY 2.5 - 3) disease stage. (D) PIGD and TD phenotypes. ICC (3,k) point estimate is indicated by the black dot, CI limits are represented by the vertical bars. ICC: intraclass correlation coefficient; MDS-UPDRS: International Movement Disorders Society Unified Parkinson's Disease Rating Scale; mHY: modified Hoehn and Yahr stage; PIGD: postural instability and gait disorder; TD: tremor dominant.



Discussion

Principal Findings

This cross-sectional study aimed at assessing and comparing the reliability of a consumer wrist-worn smartwatch (Garmin Vivosmart 4) in counting avDS in PD in unsupervised, free-living conditions for 5 consecutive days between disease phenotypes, stages, and severity groups.

Overall, our results showed, for the first time, that avDS were acceptably reliable in mild-to-moderate PD irrespective of disease severity, stage, or phenotype. Nevertheless, our results further indicated lower reliability in people with PD with TD phenotype, tremor, lower disease severity, and earlier stage.

Reliability of avDS in the Overall PD Population

We found that the relative reliability of avDS, measured in unsupervised, free-living conditions, by Garmin Vivosmart 4 for 5 consecutive days was within the a priori criteria for acceptability in the overall PD population. We identified only 2 studies investigating the reliability of wearable devices in measuring avDS in people with PD [24,44]. Paul et al [44] reported that 2 consecutive days of monitoring were sufficient to obtain an ICC of >0.9, using a research-grade, ankle-mounted, step counter (Step Activity Monitor) in 92 people with PD. One study from our group investigated the reliability of a wrist-worn consumer device in measuring avDS in PD before. In that study, we found an ICC (3,k) of 0.88 (0.82 - 0.93) for 4 days of monitoring in 56 mild-to-moderate people with PD using Garmin Vivosmart 4 [24]. The present results are consistent with our previous study, but also with other studies investigating the reliability of avDS through wrist-worn wearables in healthy older people [44-47], and in people with various neurological conditions, such as multiple sclerosis [48] and stroke [49]. Regarding absolute reliability, we found that SEM% was below 10% (9.6%) with an MDC of 1580 steps per day (27% of criterion). Only 1 study, from our group, investigated SEM and MDC for avDS in PD [24]. The present results are consistent with our previous study in which we reported an SEM% of 9.2% and an MDC of 1495 steps per day (or 26% of criterion) [24]. Taken together, these results confirm the reliability of avDS measured through Garmin Vivosmart 4 in unsupervised, free-living conditions in mild-to-moderate PD for 5 consecutive days.

Reliability of avDS in People With PD Subgroups

To our knowledge, this is the first study to investigate and compare the reliability of a consumer smartwatch in measuring avDS in different subgroups of PD based on disease phenotype, severity, and stage. We found that in all investigated subgroups (ie, TD vs PIGD, mild vs moderate, tremor vs nontremor, and earlier vs intermediate), ICC values were within the a priori criteria for acceptability (ICC range 0.84 - 0.94). However, a significantly lower ICC was observed in people with PD with tremor, mild disease severity, and earlier disease stage. Moreover, SEM% was below 10% in participants with PIGD

XSL•FO
RenderX

phenotype, moderate disease severity, intermediate disease stage, and without tremor, with an MDC ranging from 1148 to 1687 steps per day (18% - 25% of criterion). Conversely, in individuals with TD phenotype, tremor, mild disease severity, and earlier disease stage, SEM was >10% of criterion and MDC values ranged from 1401 to 2263 steps per day (30% - 33% of the criterion).

MDC, defined as the minimal change that falls outside the measurement error of an instrument, is extremely relevant in study design since it allows to calculate the sample size of studies aiming to assess the effectiveness of interventions [50]. MDC could be also crucial to define the appropriateness and feasibility of using a determinate device to measure a given construct. A prior work from Handlery et al [14] reported an increase in 1250 steps per day following a high-intensity physical activity intervention in PD measured through a research-grade wrist-worn device (Actigraph GT3X). In this study, we found that in individuals with TD phenotype, mild disease severity, and earlier disease stage, MDC was ~2000 steps per day. Although a direct comparison with the metrics reported in the work from Handlery et al [14] could not be performed due to the different devices, we could hypothesize that only large modifications in avDS could be reliably measured through Garmin Vivosmart 4 in the aforementioned PD subgroups. To this end, future studies will be needed to define the minimal clinically important difference for avDS measured through consumer wrist-worn devices and to assess the attainability of avDS modifications sufficiently large to be reliably detected by these devices.

The reduced reliability in participants with tremor is in line with our hypothesis that tremor could reduce the performance of the step-detection algorithm. In fact, tremor could increase the noise-to-signal ratio in the accelerometer signal, and in turn, render step detection more challenging [21]. Indeed, a previous study highlighted that tremor and dyskinesia together contributed to more than 19% of the variation in daily step counts when comparing measurements from waist-worn and wrist-worn devices in 46 people with PD with similar characteristics to those included in this study [21]. In this regard, our study further supports the assumption that tremor could reduce the step-detection performance of wrist-worn devices in PD.

On the other hand, our hypothesis that a reduced reliability might be observed in participants with higher symptom severity and more advanced stages was not supported by our results. Indeed, the reduced reliability of avDS observed in individuals with mild disease severity and earlier disease stage is somehow counterintuitive. In fact, previous evidence highlighted that step count was less accurate in people walking at slower gait speed and with shorter step length in several neurological and musculoskeletal conditions, including PD [20,22,51-54]. Reduced step length and slower walking speed are typical features of Parkinsonian gait, with a higher prevalence along the disease course [2,33]. Moreover, another typical characteristic of walking in PD is the reduced automaticity that leads to a more discontinuous and irregular gait pattern that can further reduce device accuracy in step detection [2,22].

Despite these considerations, our results showed that avDS estimation was more reliable in people with PD with moderate disease severity and intermediate disease stage, compared with individuals with mild disease severity and earlier disease stage. In this regard, it must be considered that reliability is a measure of consistency and reproducibility of measurement and not a measure of accuracy [55]. Therefore, a reduced accuracy could not directly translate into reduced reliability. We could hypothesize that in more advanced stages of PD, the variability of clinical presentation could be lower. In an earlier stage, indeed, symptom heterogeneity, both in terms of motor and nonmotor features, could be extremely high [56-58]. However, this variability could decrease with disease progression since motor symptoms tend to consistently worsen along the disease course and motor features such as gait and balance impairment become increasingly prevalent [59,60]. In addition, the phenotype is dynamic along the disease course and some researchers have proposed that the classification into PIGD/TD evolves over time [61,62]. One study, indeed, reported that over a period of 8 years, approximately 70% of TD individuals transitioned to PIGD, whereas only 4% of PIGD individuals transitioned to TD [61]. Another study reported that 45% of TD participants at baseline had a subtype shift along a 2-year follow-up while 85% of PIGD participants remained as PIGD [62]. This is mirrored in our study cohort, where 32 out of 65 (49%) individuals with mild disease severity were classified as TD, whereas only 7 out of 39 (18%) were in the moderate group. Similarly, 33 out of 68 (49%) participants with earlier disease stage were classified as TD, whereas only 6 out of 36 (17%) were in the intermediate group. Therefore, a regression toward a more uniform motor impairment along the disease course might be considered. Moreover, since we found that tremor could be a relevant factor in reducing avDS reliability, the different prevalence of TD phenotype could also contribute to explaining our results. However, it must be underlined that no study, to our knowledge, systematically compared the heterogeneity of PD features across early, intermediate, and advanced disease stages. Therefore, our hypothesis should be taken with caution and future studies are needed to confirm it.

In conclusion, our findings highlight that, although avDS were reliable across the examined subgroups, clinicians and researchers should consider disease phenotype, stage, and severity when implementing wrist-worn wearables and interpreting mobility data collected through these devices in PD.

Limitations

We acknowledge that this study has some limitations. First, the participants included in our study displayed relatively preserved cognitive functions, due to our exclusion of participants with a Montreal Cognitive Assessment of <21. Additionally, those with more advanced disease stages or requiring walking aids were not included. This potentially limits the generalizability of our findings. However, it should be considered that the sample in our study can be seen as representative of the typical target for interventions using consumer-grade wearable technology. Moreover, including individuals with a disease stage >3, using walking aids, or with more severe cognitive impairments poses significant challenges in the utilization of

XSL•FO
RenderX

consumer technology and is beyond our scope. Nevertheless, future research incorporating PD with lower functional scores, higher disease severity, and more impactful cognitive impairments would be valuable. Furthermore, we used only PD subtyping based on clinical features, yet other classification methods and clustering techniques have been proposed incorporating instrumental and biological data. Future studies are warranted to investigate the reliability of consumer smartwatches in PD subgroups defined using multimodal biomarkers. Finally, we did not account for antiparkinsonian treatment to control tremor. All included participants were under dopaminergic medical treatment and stable medical treatment in the 4 weeks before data collection, and we did not enroll drug-naïve individuals. In addition, we did not account for different drug classes since it would have required a much higher sample size to cover all the possible combinations. Future studies might investigate the metrological features of commercial smartwatches in drug-naïve PD to avoid the influencing factor of dopaminergic treatment. However, since almost all patients with a diagnosis of PD are medically treated, we considered it relevant to assess the measurement properties of commercial devices in a setting more representative of the real-world experience.

Conclusions

In mild-to-moderate PD, avDS measured through a consumer smartwatch in unsupervised, free-living conditions for 5 consecutive days are reliable irrespective of disease phenotype, stage, and severity. Researchers and clinicians who want to implement these instruments should consider that in individuals with TD phenotype, tremor, mild disease severity, and earlier disease stage, reliability could be lower and MDC could be higher. Future studies are needed to define the minimal clinically important difference for avDS measured through consumer wrist-worn devices and to assess the attainability of avDS modifications sufficiently large to be reliably detected by wrist-worn consumer devices. Taken together, we believe that these results could facilitate a broader implementation and an informed application of avDS as an index of ambulatory activity in PD and could be highly relevant to developing monitoring, preventive, educational, and rehabilitation strategies for PD.

Acknowledgments

The authors sincerely thank all people who were involved in the data collection, from study participants to students, who aided and assisted the assessments. This research was performed as part of EB's PhD (Sapienza University of Rome, Italy and Université Grenoble Alpes, France) and funded through a Doctoral Bursary from Sapienza University of Rome (student number 1472455). The research was supported by Sapienza University of Rome as part of the program Sapienza Research Call 2020 (grant AR120172B97AF626) and Research Call 2022 (grant AR222181683A15AA), by the French National Research Agency (ANR) in the framework of the Investissements d'avenir program (ANR-10- AIRT-05 and ANR-15-IDEX-02), and the MIAI Cluster (ANR-23-IACL-0006). This work also forms part of a broader translational and interdisciplinary GaitAlps research program (NV)

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overlap between the groups of participants with and without self-reported tremor and TD or PIGD phenotypes as well as between the groups with mild or moderate disease severity and early or intermediate disease stage. mHY: modified Hoehn and Yahr scale; PIGD: postural instability and gait disorder; TD: tremor dominant.
[DOCX File, 22 KB - formative_v9i1e63153_app1.docx]

Multimedia Appendix 2
Proof-of-concept experiment to test criterion validity of Garmin Vivosmart 4 in estimating average daily steps in free-living conditions in people with Parkinson disease.
[DOCX File, 26 KB - formative_v9i1e63153_app2.docx]

References

1. MacKinnon CD. Sensorimotor anatomy of gait, balance, and falls. Handb Clin Neurol 2018;159(3–26):3-26. [doi: 10.1016/B978-0-444-63916-5.00001-X] [Medline: 30482322]
2. Mirelman A, Bonato P, Camicioli R, et al. Gait impairments in Parkinson's disease. Lancet Neurol 2019 Jul;18(7):697-708. [doi: 10.1016/S1474-4422(19)30044-4] [Medline: 30975519]
3. Lee M, Noh Y, Youm C, et al. Estimating health-related quality of life based on demographic characteristics, questionnaires, gait ability, and physical fitness in Korean elderly adults. IJERPH 2021 Nov 11;18(22):11816. [doi: 10.3390/ijerph182211816]

XSL•FO
RenderX

4. Creaby MW, Cole MH. Gait characteristics and falls in Parkinson's disease: a systematic review and meta-analysis. Parkinsonism Relat Disord 2018 Dec;57:1-8. [doi: 10.1016/j.parkreldis.2018.07.008] [Medline: 30041848]
5. Bassett DR, Toth LP, LaMunion SR, Crouter SE. Step counting: a review of measurement considerations and health-related applications. Sports Med 2017 Jul;47(7):1303-1315. [doi: 10.1007/s40279-016-0663-1] [Medline: 28005190]
6. Inoue K, Tsugawa Y, Mayeda ER, Ritz B. Association of daily step patterns with mortality in US adults. JAMA Netw Open 2023 Mar 1;6(3):e235174. [doi: 10.1001/jamanetworkopen.2023.5174] [Medline: 36976556]
7. Lee IM, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of step volume and intensity with all-cause mortality in older women. JAMA Intern Med 2019 Aug 1;179(8):1105-1112. [doi: 10.1001/jamainternmed.2019.0899] [Medline: 31141585]
8. Paluch AE, Bajpai S, Bassett DR, et al. Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. Lancet Public Health 2022 Mar;7(3):e219-e228. [doi: 10.1016/S2468-2667(21)00302-9] [Medline: 35247352]
9. Saint-Maurice PF, Troiano RP, Bassett DR Jr, et al. Association of daily step count and step intensity with mortality among US adults. JAMA 2020 Mar 24;323(12):1151-1160. [doi: 10.1001/jama.2020.1382] [Medline: 32207799]
10. Del Pozo Cruz B, Ahmadi M, Naismith SL, Stamatakis E. Association of daily step count and intensity with incident dementia in 78 430 adults living in the UK. JAMA Neurol 2022 Oct 1;79(10):1059-1063. [doi: 10.1001/jamaneurol.2022.2672] [Medline: 36066874]
11. Del Pozo Cruz B, Ahmadi MN, Lee IM, Stamatakis E. Prospective associations of daily step counts and intensity with cancer and cardiovascular disease incidence and mortality and all-cause mortality. JAMA Intern Med 2022 Nov 1;182(11):1139-1148. [doi: 10.1001/jamainternmed.2022.4000] [Medline: 36094529]
12. Banach M, Lewek J, Surma S, et al. The association between daily step count and all-cause and cardiovascular mortality: a meta-analysis. Eur J Prev Cardiol 2023 Dec 21;30(18):1975-1985. [doi: 10.1093/eurjpc/zwad229] [Medline: 37555441]
13. Skidmore FM, Mackman CA, Pav B, et al. Daily ambulatory activity levels in idiopathic Parkinson disease. J Rehabil Res Dev 2008;45(9):1343-1348. [doi: 10.1682/JRRD.2008.01.0002] [Medline: 19319758]
14. Handlery R, Stewart JC, Pellegrini C, et al. Physical activity in De Novo Parkinson disease: daily step recommendation and effects of treadmill exercise on physical activity. Phys Ther 2021 Oct 1;101(10):34244805. [doi: 10.1093/ptj/pzab174] [Medline: 34244805]
15. Mattison G, Canfell O, Forrester D, et al. The influence of wearables on health care outcomes in chronic disease: systematic review. J Med Internet Res 2022 Jul 1;24(7):e36690. [doi: 10.2196/36690] [Medline: 35776492]
16. Bianchini E, Maetzler W. Chapter five: wearable systems in movement disorders. In: International Review of Movement Disorders: Academic Press; 2023, Vol. 5. [doi: 10.1016/bs.irmvd.2023.04.002]
17. Chevance G, Golaszewski NM, Tipton E, et al. Accuracy and precision of energy expenditure, heart rate, and steps measured by combined-sensing Fitbits against reference measures: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022 Apr 13;10(4):e35626. [doi: 10.2196/35626] [Medline: 35416777]
18. Fuller D, Colwell E, Low J, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 8;8(9):e18694. [doi: 10.2196/18694] [Medline: 32897239]
19. Nakagata T, Murakami H, Kawakami R, et al. Step-count outcomes of 13 different activity trackers: results from laboratory and free-living experiments. Gait Posture 2022 Oct;98:24-33. [doi: 10.1016/j.gaitpost.2022.08.004] [Medline: 36030707]
20. Bianchini E, Caliò B, Alborghetti M, et al. Step-counting accuracy of a commercial smartwatch in mild-to-moderate PD patients and effect of spatiotemporal gait parameters, laterality of symptoms, pharmacological state, and clinical variables. Sensors (Basel) 2022 Dec 25;23(1):214. [doi: 10.3390/s23010214] [Medline: 36616812]
21. Kim DW, Hassett LM, Nguy V, Allen NE. A comparison of activity monitor data from devices worn on the wrist and the waist in people with Parkinson's disease. Mov Disord Clin Pract 2019 Nov;6(8):693-699. [doi: 10.1002/mdc3.12850] [Medline: 31745480]
22. Wendel N, Macpherson CE, Webber K, et al. Accuracy of activity trackers in Parkinson disease: should we prescribe them? Phys Ther 2018 Aug 1;98(8):705-714. [doi: 10.1093/ptj/pzy054] [Medline: 29718452]
23. Ginis P, Goris M, De Groef A, et al. Validation of commercial activity trackers in everyday life of people with Parkinson's disease. Sensors (Basel) 2023 Apr 21;23(8):4156. [doi: 10.3390/s23084156] [Medline: 37112496]
24. Bianchini E, Galli S, Alborghetti M, et al. Four days are enough to provide a reliable daily step count in mild to moderate Parkinson's disease through a commercial smartwatch. Sensors (Basel) 2023 Nov 4;23(21):8971. [doi: 10.3390/s23218971] [Medline: 37960670]
25. Outeiro TF, Alcalay RN, Antonini A, et al. Defining the riddle in order to solve it: there is more than one "Parkinson's disease". Mov Disord 2023 Jul;38(7):1127-1142. [doi: 10.1002/mds.29419] [Medline: 37156737]
26. Jankovic J, McDermott M, Carter J, et al. Variable expression of Parkinson's disease: a base-line analysis of the DATATOP cohort. The Parkinson Study Group. Neurology (ECronicon) 1990 Oct;40(10):1529-1534. [doi: 10.1212/wnl.40.10.1529] [Medline: 2215943]
27. Stebbins GT, Goetz CG, Burn DJ, Jankovic J, Khoo TK, Tilley BC. How to identify tremor dominant and postural instability/gait difficulty groups with the movement disorder society unified Parkinson's disease rating scale: comparison

XSL•FO
RenderX

with the unified Parkinson's disease rating scale. Mov Disord 2013 May;28(5):668-670. [doi: 10.1002/mds.25383] [Medline: 23408503]
28. Zetusky WJ, Jankovic J, Pirozzolo FJ. The heterogeneity of Parkinson's disease: clinical and prognostic implications. Neurology (ECronicon) 1985 Apr;35(4):522-526. [doi: 10.1212/wnl.35.4.522] [Medline: 3982637]
29. Cardoso F, Goetz CG, Mestre TA, et al. A statement of the MDS on biological definition, staging, and classification of Parkinson's disease. Mov Disord 2024 Feb;39(2):259-266. [doi: 10.1002/mds.29683] [Medline: 38093469]
30. Höglinger GU, Adler CH, Berg D, et al. A biological classification of Parkinson's disease: the SynNeurGe research diagnostic criteria. Lancet Neurol 2024 Feb;23(2):191-204. [doi: 10.1016/S1474-4422(23)00404-0] [Medline: 38267191]
31. Horsager J, Borghammer P. Brain-first vs. body-first Parkinson's disease: an update on recent evidence. Parkinsonism Relat Disord 2024 May;122:106101. [doi: 10.1016/j.parkreldis.2024.106101] [Medline: 38519273]
32. Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society - sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS - UPDRS): scale presentation and clinimetric testing results. Mov Disord 2008 Nov 15;23(15):2129-2170. [doi: 10.1002/mds.22340]
33. Schlachetzki JCM, Barth J, Marxreiter F, et al. Wearable sensors objectively measure gait parameters in Parkinson's disease. PLoS ONE 2017;12(10):e0183989. [doi: 10.1371/journal.pone.0183989] [Medline: 29020012]
34. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord 2015 Oct;30(12):1591-1601. [doi: 10.1002/mds.26424]
35. Goetz CG, Poewe W, Rascol O, et al. Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations. Mov Disord 2004 Sep;19(9):1020-1028. [doi: 10.1002/mds.20213] [Medline: 15372591]
36. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
37. Jost ST, Kaldenbach MA, Antonini A, et al. Levodopa dose equivalency in Parkinson's disease: updated systematic review and proposals. Mov Disord 2023 Jul;38(7):1236-1252. [doi: 10.1002/mds.29410] [Medline: 37147135]
38. Martínez-Martín P, Rodríguez-Blázquez C, et al. Parkinson's disease severity levels and MDS-Unified Parkinson's Disease Rating Scale. Parkinsonism Relat Disord 2015 Jan;21(1):50-54. [doi: 10.1016/j.parkreldis.2014.10.026] [Medline: 25466406]
39. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163. [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
40. Knol MJ, Pestman WR, Grobbee DE. The (mis)use of overlap of confidence intervals to assess effect modification. Eur J Epidemiol 2011 Apr;26(4):253-254. [doi: 10.1007/s10654-011-9563-8] [Medline: 21424218]
41. Austin PC, Hux JE. A brief note on overlapping confidence intervals. J Vasc Surg 2002 Jul;36(1):194-195. [doi: 10.1067/mva.2002.125015] [Medline: 12096281]
42. Cho J, Seo DM, Uh Y. Clinical application of overlapping confidence intervals for monitoring changes in serial clinical chemistry test results. Ann Lab Med 2020 May;40(3):201-208. [doi: 10.3343/alm.2020.40.3.201] [Medline: 31858759]
43. Balaguier R, Madeleine P, Vuillerme N. Intra-session absolute and relative reliability of pressure pain thresholds in the low back region of vine-workers: effect of the number of trials. BMC Musculoskelet Disord 2016 Aug 18;17(1):350. [doi: 10.1186/s12891-016-1212-7] [Medline: 27538914]
44. Paul SS, Ellis TD, Dibble LE, et al. Obtaining reliable estimates of ambulatory physical activity in people with Parkinson's disease. J Parkinsons Dis 2016 May 5;6(2):301-305. [doi: 10.3233/JPD-160791] [Medline: 27164042]
45. Yao J, Tan CS, Lim N, Tan J, Chen C, Müller-Riemenschneider F. Number of daily measurements needed to estimate habitual step count levels using wrist-worn trackers and smartphones in 212,048 adults. Sci Rep 2021 May 5;11(1):9633. [doi: 10.1038/s41598-021-89141-3] [Medline: 33953288]
46. Clemes SA, Griffiths PL. How many days of pedometer monitoring predict monthly ambulatory activity in adults? Med Sci Sports Exerc 2008 Sep;40(9):1589-1595. [doi: 10.1249/MSS.0b013e318177eb96] [Medline: 18685533]
47. Sigmundová D, Vašíčková J, Stelzer J, Repka E. The influence of monitoring interval on data measurement: an analysis of step counts of university students. Int J Environ Res Public Health 2013 Jan 28;10(2):515-527. [doi: 10.3390/ijerph10020515] [Medline: 23358235]
48. Norris M, Anderson R, Motl RW, Hayes S, Coote S. Minimum number of days required for a reliable estimate of daily step count and energy expenditure, in people with MS who walk unaided. Gait Posture 2017 Mar;53:201-206. [doi: 10.1016/j.gaitpost.2017.02.005] [Medline: 28199925]
49. Fini NA, Burge AT, Bernhardt J, Holland AE. Two days of measurement provides reliable estimates of physical activity poststroke: an observational study. Arch Phys Med Rehabil 2019 May;100(5):883-890. [doi: 10.1016/j.apmr.2018.10.006] [Medline: 31030730]
50. Lexell JE, Downham DY. How to assess the reliability of measurements in rehabilitation. Am J Phys Med Rehabil 2005 Sep;84(9):719-723. [doi: 10.1097/01.phm.0000176452.17771.20] [Medline: 16141752]
51. Treacy D, Hassett L, Schurr K, Chagpar S, Paul SS, Sherrington C. Validity of different activity monitors to count steps in an inpatient rehabilitation setting. Phys Ther 2017 May 1;97(5):581-588. [doi: 10.1093/ptj/pzx010] [Medline: 28339904]

XSL•FO
RenderX

52. Fokkema T, Kooiman TJM, Krijnen WP, Van Der Schans CP, De Groot M. Reliability and validity of ten consumer activity trackers depend on walking speed. Med Sci Sports Exerc 2017 Apr;49(4):793-800. [doi: 10.1249/MSS.0000000000001146] [Medline: 28319983]
53. Chow JJ, Thom JM, Wewege MA, Ward RE, Parmenter BJ. Accuracy of step count measured by physical activity monitors: the effect of gait speed and anatomical placement site. Gait Posture 2017 Sep;57:199-203. [doi: 10.1016/j.gaitpost.2017.06.012] [Medline: 28666177]
54. Lamont RM, Daniel HL, Payne CL, Brauer SG. Accuracy of wearable physical activity trackers in people with Parkinson's disease. Gait Posture 2018 Jun;63:104-108. [doi: 10.1016/j.gaitpost.2018.04.034] [Medline: 29729611]
55. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010 May;19(4):539-549. [doi: 10.1007/s11136-010-9606-8] [Medline: 20169472]
56. Erro R, Vitale C, Amboni M, et al. The heterogeneity of early Parkinson's disease: a cluster analysis on newly diagnosed untreated patients. PLoS ONE 2013;8(8):e70244. [doi: 10.1371/journal.pone.0070244] [Medline: 23936396]
57. Wüllner U, Borghammer P, Choe CU, et al. The heterogeneity of Parkinson's disease. J Neural Transm (Vienna) 2023 Jun;130(6):827-838. [doi: 10.1007/s00702-023-02635-4] [Medline: 37169935]
58. Berg D, Borghammer P, Fereshtehnejad SM, et al. Prodromal Parkinson disease subtypes—key to understanding heterogeneity. Nat Rev Neurol 2021 Jun;17(6):349-361. [doi: 10.1038/s41582-021-00486-9] [Medline: 33879872]
59. Fasano A, Fung VSC, Lopiano L, et al. Characterizing advanced Parkinson's disease: OBSERVE-PD observational study results of 2615 patients. BMC Neurol 2019 Apr 2;19(1):50. [doi: 10.1186/s12883-019-1276-8] [Medline: 30940119]
60. Stefani A, Tessitore A, Tambasco N, et al. Criteria for identification of advanced Parkinson's disease: the results of the Italian subgroup of OBSERVE-PD observational study. BMC Neurol 2022 Jan 28;22(1):41. [doi: 10.1186/s12883-022-02554-z] [Medline: 35090406]
61. Alves G, Larsen JP, Emre M, Wentzel-Larsen T, Aarsland D. Changes in motor subtype and risk for incident dementia in Parkinson's disease. Mov Disord 2006 Aug;21(8):1123-1130. [doi: 10.1002/mds.20897] [Medline: 16637023]
62. Che N, Ou R, Li C, et al. Plasma GFAP as a prognostic biomarker of motor subtype in early Parkinson's disease. NPJ Parkinsons Dis 2024 Mar 1;10(1):48. [doi: 10.1038/s41531-024-00664-8] [Medline: 38429295]

Abbreviations

avDS: average daily steps
ICC: intraclass correlation coefficient
MDC: minimal detectable change
MDS-UPDRS: International Movement Disorders Society Unified Parkinson's Disease Rating Scale
mHY: modified Hoehn and Yahr scale
PD: Parkinson disease
PIGD: postural instability and gait disorder
SEM: Standard error of measurement
TD: tremor dominant

Edited by A Mavragani; submitted 16.06.24; peer-reviewed by E Rainaldi, J Ruokolainen; revised version received 03.02.25; accepted 21.02.25; published 18.03.25.

Please cite as:
Bianchini E, Rinaldi D, De Carolis L, Galli S, Alborghetti M, Hansen C, Suppa A, Salvetti M, Pontieri FE, Vuillerme N
Reliability of Average Daily Steps Measured Through a Consumer Smartwatch in Parkinson Disease Phenotypes, Stages, and Severities: Cross-Sectional Study
JMIR Form Res 2025;9:e63153
URL: https://formative.jmir.org/2025/1/e63153
doi:10.2196/63153

© Edoardo Bianchini, Domiziana Rinaldi, Lanfranco De Carolis, Silvia Galli, Marika Alborghetti, Clint Hansen, Antonio Suppa, Marco Salvetti, Francesco Ernesto Pontieri, Nicolas Vuillerme. Originally published in JMIR Formative Research (https://formative.jmir.org), 18.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Patient and Provider Perspectives of a Web-Based Intervention to Support Symptom Management After Radioactive Iodine Treatment for Differentiated Thyroid Cancer: Qualitative Study

Alaina L Carr[1], PhD; Angela M Jenkins[1,2], BS; Jacqueline Jonklaas[3], MD, MPH, PhD; Kate Gabriel[4], BS; Kristen E Miller[5], DrPH, MSPH, MSL, CPPS; Kristi D Graves[1], PhD

[1]Department of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC, United States

[2]Department of Natural Sciences, Bowie State University College of Arts and Science, Bowie, MD, United States

[3]Division of Endocrinology, Georgetown University, Washington, DC, United States

[4]Department of Health Systems Administration, Georgetown University School of Health, Washington, DC, United States

[5]MedStar Health Research Institute, National Center for Human Factors in Healthcare, Washington, DC, United States

Corresponding Author:
Alaina L Carr, PhD
Department of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University, Washington, DC, United States

Abstract

Background: Patients diagnosed with differentiated thyroid cancer (DTC) who receive radioactive iodine (RAI) treatment experience acute, medium, and late treatment effects. The timing and severity of these effects vary by individual; common posttreatment effects include dry mouth, salivary gland swelling, dry eyes, and nose bleeds. The nature of symptoms that patients experience after RAI treatment can significantly and negatively impact health-related quality of life. Adequate information during the postprimary treatment phase remains an unmet need among the population of patients diagnosed with DTC.

Objective: This qualitative study aimed to identify and understand self-management strategies for RAI-specific symptom burden from the perspectives of patients and stakeholders (cancer care providers and patient advocates). An additional aim included assessing features and functionalities desirable in the development of a web-based intervention to engage patients in their self-management and thyroid cancer survivorship care.

Methods: Following the Social Cognitive Theory framework and person-based principles, we conducted six focus groups with 22 patients diagnosed with DTC who completed RAI treatment and individual interviews with 12 stakeholders in DTC care. The interviews focused on participants' perspectives on current self-management strategies and mockups of a symptom management web-based intervention. Before focus groups and interviews, participants completed a demographics survey. Focus group discussions and interviews were transcribed and coded using content analysis. Interrater reliability was satisfactory (α=.88).

Results: A total of 34 individuals (patients and stakeholders) participated in the study; the mean age was 45 (SD 13.4) and 45.3 (SD 13) years, respectively. Three domains emerged from qualitative interviews: (1) difficult-to-manage RAI symptoms: short, medium, and late treatment effects; (2) key intervention structure and content feedback on mockups; and (3) intervention content to promote RAI symptom management and survivorship care. Focus group participants identified the most prevalent RAI symptoms that were difficult to manage as: dry mouth (11/22, 50%), salivary gland swelling (8/22, 36%), and changes in taste (12/22, 55%). Feedback elicited from both groups found education and symptom management mockup videos to be helpful in patient self-management of RAI symptoms, whereas patients and stakeholders provided mixed feedback on the benefits of a draft frequently asked questions page. Across focus groups and stakeholder interviews, nutrition-based symptom management strategies, communication with family members, and practical survivorship follow-up information emerged as helpful content to include in a future web-based supportive care intervention.

Conclusions: Results suggest education and symptom management videos can empower patients with DTC to self-manage mild to moderate RAI symptoms on a web-based platform. Findings emphasized the need for additional information for patients related to ongoing care following RAI treatment including social support and thyroid cancer surveillance. The findings provide insights for theoretically informed interventions and recommendations for refinements in thyroid cancer survivorship from patient and provider perspectives.

(JMIR Form Res 2025;9:e60588) doi:10.2196/60588

XSL•FO
RenderX

KEYWORDS

iodine radioisotopes; person-based approach; self-management; Social Cognitive Theory; survivorship; symptom burden; thyroid neoplasms; web-based intervention; radioactive iodine treatment; radiotherapy; thyroid cancer; qualitative; quality of life; survivorship care; supportive care; patient with cancer; QoL; cancer; carcinoma; malignancy; tumor; malignant; oncology; neoplasm; benign; neoplasia; thyroid

Introduction

Background

Thyroid cancer is often mislabeled as the "good cancer" [1-3] due to the excellent 5-year survival rates compared to other cancers [4]. Differentiated thyroid cancer (DTC) is the most common endocrine malignancy among the estimated 44,020 new cases of thyroid cancer in the United States annually [4]. DTC involves a spectrum of disease presentations characterized by different levels of risk of recurrence. For patients with a high risk of recurrence, the standard of care involves total thyroidectomy followed by radioactive iodine (RAI) treatment and thyroid hormone therapy [5-7]. For patients with intermediate-risk disease, clinical practice guidelines from the American Thyroid Association recommend thyroidectomy followed by consideration of RAI treatment in consultation with their medical team [6]. Approximately, 72% of these intermediate and high-risk patients receive RAI and thyroid hormone therapy as standard of care [5]. Despite the favorable 5-year survival rates, complications from RAI treatment are common, impacting health-related quality of life (HRQOL) [1,8-10]. The "good cancer" label given to thyroid cancer frustrates patients, particularly when they experience side effects as a result of RAI treatment [1].

The likelihood of acute and chronic side effects following RAI varies based on RAI dose administration; these treatment impacts are particularly evident in patients who receive higher doses of ≥150 mCi [7,11,12]. While trends toward lower dose activity are recommended, therapeutic dose recommendations vary across professional medical societies and practice patterns [6,11-14]. Short-, medium-, and long-term complications from RAI treatment include salivary dysfunction, dysphagia, dysphonia, sialadenitis, appetite changes, nausea, vomiting, xerostomia, epiphora, and secondary malignancies [1,15-17], with the onset of certain symptoms appearing as late as 1-year post-RAI treatment [12,16]. In a cross-sectional study of 69 patients diagnosed with DTC and treated with RAI, lower HRQOL was significantly associated with higher symptom burden, distress, and lower self-efficacy [18]. Furthermore, patients with thyroid cancer report lower HRQOL compared to patients with other types of cancer with worse survival [10]. Additional care challenges reported by patients with DTC are unmet information needs about follow-up care, psychosocial support, and coordination of care [17,19]. Research is needed to understand, address, and mitigate RAI symptom burden among patients with DTC.

Mobile health (mHealth) technologies offer an accessible approach to symptom management interventions [20] and are increasingly used in the delivery of cancer care. mHealth refers to mobile and wireless technologies such as smartphones, computers, and tablets to support patient care, education, and communication in health care [21,22]. mHealth technologies have been implemented in the prevention, promotion, treatment, and maintenance of health and health care for patients with breast, lung, leukemia, and prostate cancers [21-25]. mHealth interventions empower patients with cancer with direct access to educational and support tools to improve self-management of symptoms, including fatigue, pain, distress, and cognitive impairment [21,23,26,27]. Core elements of mHealth interventions for chronic disease management include evidence-based education, monitoring and tracking symptoms, tailored feedback on symptoms, self-management training to cope with the psychological and physical aspects of treatment, and communication with providers [21,28,29]. Applying evidence-based theories to inform the design characteristics of digital technology interventions during the development phase of the research is one way to engage patients, foster patient-provider communication, and improve patient-centered health outcomes [20-22,28].

Theoretically Driven mHealth Intervention Development

Social Cognitive Theory (SCT) [30,31], a longstanding theory-based approach to health behavior change, coupled with guiding principles from the person-based approach [20,32] to mHealth intervention development, offers a basis for core components of symptom management interventions. Behavioral interventions based on tenets of SCT are evaluated through patient knowledge of health risks and benefits, self-efficacy or confidence related to symptom management, outcome expectations about the costs and benefits of performing a particular behavior, and self-regulation impacting patient HRQOL [30,31]. Prior examples of the use of SCT in cancer-related health behavior change interventions include physical activity behavior change interventions [33-36] or interventions to promote HRQOL in breast cancer populations [37]. The application of SCT to promote self-management of RAI symptoms among patients with DTC is understudied, and further research is needed to identify core components of health-related interventions. The person-based approach to mHealth interventions yields insights into the behavioral components relevant to the populations intended to use the intervention [32]. As patients with DTC post-RAI treatment are commonly seen in ambulatory care settings, and their follow-up care is managed by specialists [17], a digital platform offers patients accessibility delivered at the point of need. Furthermore, involving target patient users and members of the health care team early in the intervention design process can identify key components [20,21,32] of symptom management interventions among this population.

We seek to develop and refine tailored mHealth symptom management interventions for patients with DTC who completed RAI. Thus, we conducted a qualitative study to identify and

XSL•FO
RenderX

understand potential self-management strategies for RAI-specific symptoms from the perspectives of patients with DTC who completed RAI treatment and health care providers. In addition, we wanted to elicit feedback on core content and features in the formative development of a web-based intervention to promote patient self-management of RAI symptoms.

Methods

Study Design

We followed the COREQ (Consolidated Criteria for Reporting Qualitative Research) to report the qualitative findings [38]. From August 2022 to January 2024, we conducted a qualitative study using focus groups with patient participants and individual stakeholder interviews at Georgetown University Medical Center.

Ethical Considerations

Ethics approval was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Georgetown University Institutional Review Board (#00005285). All focus group participants and stakeholders provided informed consent. Study data were deidentified prior to data analysis. Each participant received Amazon gift cards worth US $50 for their participation.

Participant Selection and Setting

Adult patients were recruited from MedStar Georgetown University Hospital to take part in the qualitative study. Treating endocrinologists referred patients to the study. Patients were invited to participate if they were diagnosed with DTC, completed RAI treatment within the past three years, and had access to an electronic device with an internet connection. We excluded non-English-speaking patients and those with psychiatric or cognitive impairments that would inhibit meaningful consent.

For the stakeholder interviews, health care specialists were recruited from four hospitals in the Washington metropolitan area. Health care specialists were referred to the study by colleagues or invited to participate directly by study staff through an introductory email. Stakeholder inclusion criteria were self-identification as a full-time specialist treating patients with DTC post-RAI treatment or allied health professionals and patient advocates affiliated with ThyCa: Thyroid Cancer Survivors' Association, Inc.

We used a purposive sampling strategy for patient age, gender, cancer type, and health subspecialists from academic and community hospital settings in the area. We recruited 22 patients diagnosed with DTC who completed RAI treatment and 12 stakeholders from local health care facilities and ThyCa groups in the DC area.

Procedures and Data Collection

Following Institutional Review Board approval, study staff approached potential participants with an introductory study email followed by a phone call. Study staff completed informed consent through a 15-minute telephone or Health Insurance Portability and Accountability Act-compliant video conference and obtained electronic informed consent from focus group participants and stakeholders. Focus group participants and stakeholders completed a brief survey through REDCap (Research Electronic Data Capture; Vanderbilt University) [39] to obtain sociodemographic, clinical characteristics, and occupation information. The interviews were conducted over a Health Insurance Portability and Accountability Act-compliant videoconference lasting approximately two hours for focus group participants and one hour for individual stakeholder interviews.

Two authors (ALC: psychologist, female; and KDG: behavioral scientist, female) trained in qualitative research methods developed moderator and semistructured interview guides (Multimedia Appendices 1 and 2), which were informed by tenets of SCT [30] and person-based design principles [32]. Both guides included items to elicit participants' perspectives about RAI symptoms that were challenging to manage, identify content that would be helpful in a symptom management website, and obtain feedback on the formatting and content of static images or mockups (Multimedia Appendix 3) of a prospective symptom management website. Participants were asked to provide feedback on the mockups after eliciting input on prospective content to minimize participant response bias. One author (ALC) conducted audio-recorded interviews for focus group and stakeholder interviews. A second member of the research team trained in qualitative research (KG: research assistant, female; Gautham Pillai: research assistant, male; and AMJ: research assistant, female) took field notes during the focus group and stakeholder interviews. All focus group and stakeholder interviews were transcribed verbatim and deidentified before data analysis. All focus group participants and stakeholders received a summary of the findings.

Data Analysis

We used thematic content analysis to analyze focus group and stakeholder interview transcripts using a well-established 6-step framework [40]. Three authors (ALC, AMJ, and KG) independently reviewed transcripts to familiarize themselves with the data and identify initial codes using ATLAS.ti, a computer-assisted qualitative data analysis software (ATLAS.ti Scientific Software Development GmbH). The authors (ALC, AMJ, and KG) discussed initial codes, shared definitions of each code, and grouped the codes into broad, meaningful categories. The coders further organized the codes into major domains and subthemes and examined discrepant data until resolution. Analysis was systematically and iteratively applied across all transcripts. Thematic saturation [40] occurred after no new information was identified from transcripts. We calculated Krippendorf Cu-α coefficients [41,42] to evaluate coding precision by the three team members for the three domains across transcripts. The final Krippendorf Cu-α=.88, indicating acceptable agreement [43]. Extracted quotations exemplified major domains and subthemes. Descriptive statistics were used to characterize focus group and stakeholder demographic, clinical, and occupation characteristics.

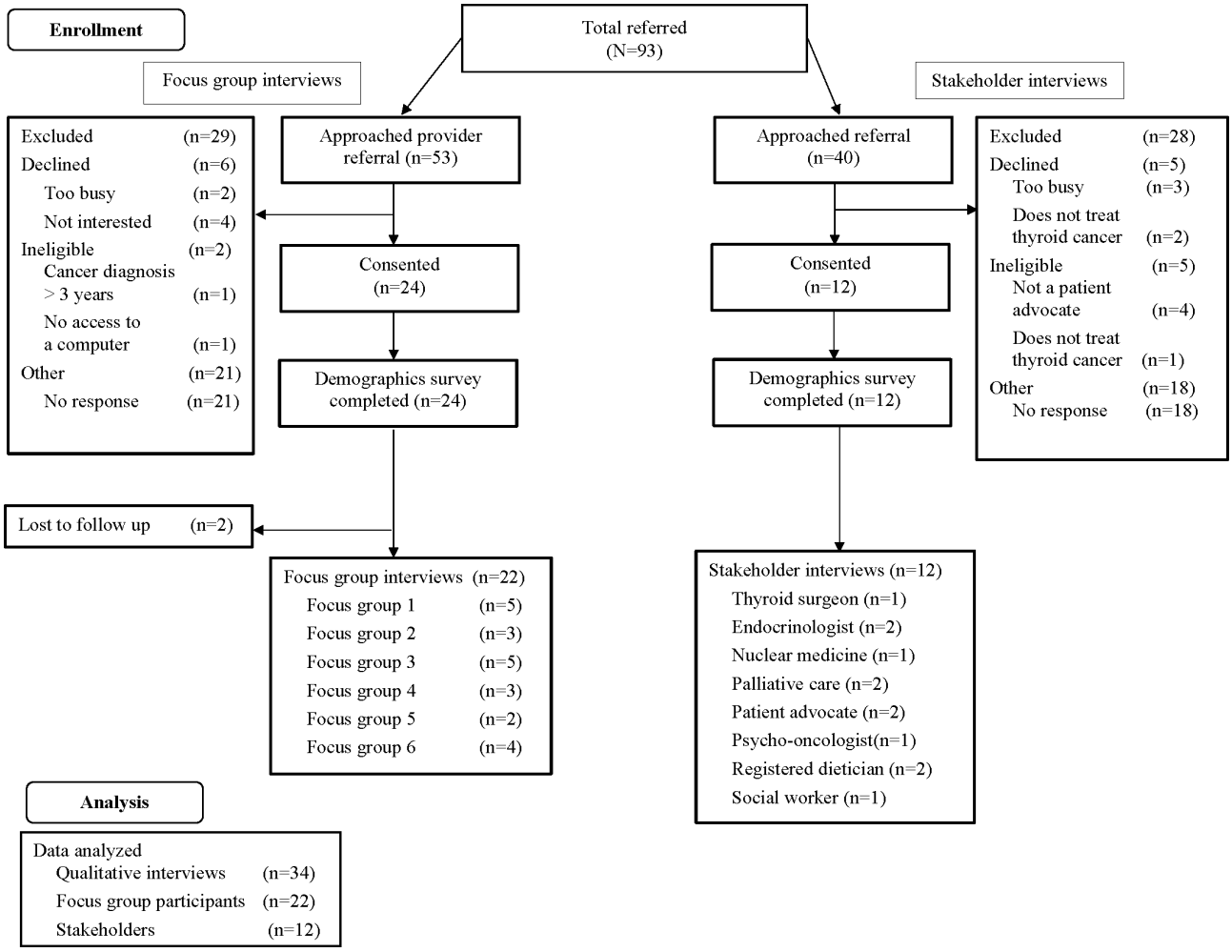
XSL•FO
RenderX

Results

Participants

We approached 93 patients and stakeholders to enroll in the qualitative study; 24 patients and 12 stakeholders consented, 2 patients were unable to participate in focus groups due to scheduling conflicts, and 34 interviews were analyzed (Figure 1). A total of 29 patients were excluded due to ineligibility (n=2), decline (n=6), and no response to the introductory study email (n=21). A total of 28 stakeholders were excluded from the study due to ineligibility (n=5), decline (n=5), and no response to the introductory study email (n=18).

Figure 1. Qualitative study flowchart for focus group participants and stakeholders. Participants, recruited from clinical settings, were patients diagnosed with thyroid cancer. Stakeholders, recruited through professional networks, were providers involved in the clinical care of patients diagnosed with thyroid cancer. Data were collected between August 2022 and January 2024.



The age (in years) of focus group participants and stakeholders was 45.0 (SD 13.4; range 23.6 - 72.5) and 45.2 (13.0; range 32.3 - 75.7), respectively. The majority of focus group participants and stakeholders identified as female (15/22, 68%; 10/12, 83%), non-Hispanic or Latino (20/22, 91%; 11/12, 92%), and White (14/22, 64%; 8/12, 67%), respectively. In terms of clinical characteristics, the majority of focus group participants were diagnosed with papillary thyroid cancer (18/22, 82%). The average time since diagnosis (in years) was 1.5 (SD 1.0; range 0.03 - 2.9), and the time since treatment (in years) was 1.1 (SD 0.9; range 0.1 - 2.8). There was a total of six focus groups, with a range of 2 - 5 participants in each group. Stakeholders consisted of thyroid surgeons (n=1), endocrinologists (n=2), nuclear medicine specialists (n=1), palliative care providers (n=2), patient advocates (n=2), psycho-oncologists (n=1), registered oncology dietitians (n=2), and social workers (n=1). Table 1 presents sample characteristics of focus group participants and stakeholders and Table 2 presents clinical characteristics of focus group participants.

XSL•FO
RenderX

Table . Sample characteristics of focus group participants and stakeholders (N=34)[a].

Characteristics	Focus group participants (n=22)	Stakeholders (n=12)
Age (years)		
Mean (SD)	45.0 (13.4)	45.2 (13.0)
Range	23.6-72.5	32.3-75.7
Female, n (%)	15 (68)	10 (83)
Married or partnered, n (%)	14 (64)	—[b]
Hispanic or Latino, n (%)	2 (9)	1 (8)
Race, n (%)		
Asian	3 (14)	3 (25)
Black or African American	2 (9)	1 (8)
More than one race/ other	3 (14)	—
White	14 (64)	8 (67)
Education, n (%)		
High-school and below	3 (14)	—
College and above	19 (86)	—
Employment status, n (%)		
Full-time	17 (77)	—
Part-time	3 (14)	—
Student	2 (9)	—
Retired	2 (9)	—
Heath Insurance (Yes), n (%)	22 (100)	—

[a]Participants, recruited from clinical settings, were patients diagnosed with thyroid cancer. Stakeholders, recruited through professional networks, were providers involved in the clinical care of patients diagnosed with thyroid cancer. Data were collected between August 2022 and January 2024.

[b]Not applicable.

XSL•FO
RenderX

Table . Differentiated thyroid cancer clinical characteristics of focus group participants (n=22)[a].

Clinical characteristics	Focus group participants
Time since diagnosis (years)	
Mean (SD)	1.5 (0.9)
Range	0.03 - 2.9
Time since treatment (years)	
Mean (SD)	1.1 (0.9)
Range	0.10 - 2.8
Histology[b], n (%)	
Papillary	18 (82)
Follicular	2 (9)
Do not know	4 (18)
Stage[b], n (%)	
I	6 (27)
II	2 (9)
III	1 (5)
IVA	1 (5)
Do not know	11 (50)

[a]Participants recruited from clinical settings were patients diagnosed with thyroid cancer. Data was collected between August 2022 and January 2024.
[b]The value given are frequencies.

Qualitative Results

Qualitative results yielded three domains: (1) difficult-to-manage RAI symptoms: short, medium, and late treatment effects; (2) key intervention structure and content feedback on mockups; and (3) intervention content to promote RAI symptom management and survivorship care. Tables 3 and 4 display exemplar quotes with patient identifiers to depict domains and subthemes from focus group participant and stakeholder interviews.

XSL•FO
RenderX

Table . Summary and exemplar quotes from focus group participants and stakeholders related to domain 2 key intervention structure and content feedback on mockups with an SCT[a] framework.

Mockup	Exemplar quotes	SCT[a] component
Home page Description: benefits and consequences home page in its current format; recommended changes to enhance patient understanding of website component	• "In those websites you always see younger people. I think it's nice to include more diversity. If you are going to include more pictures have other ones where you see people that are older depicted." [Focus Group 5, PID 2] • "On the landing page ... You can have a tab on statistics ... Are you experiencing side effects? Side effects management tab here. Social resources here. Or supportive care resources. You could have a separate nutrition tab." [PID 10, Registered Oncology Dietitian]	Knowledge and outcome expectations
Education page RAI[b] salivary symptoms Description: benefits of education page on RAI symptom self-management; recommended changes to enhance engagement in website component	• "[It] Feels friendly discussing something that is very distressing. There is an empowerment element to this of, 'Hey, we get you. You're experiencing these things. Please watch out for these and call your doctor.' This page is magical. I wish this site existed already." [Focus Group 4, PID 3] • "It's a good general informational landing spot particularly things that they should definitely not try home remedies for. Like the bottom half [Redness over salivary glands, Pain doesn't go away with antiinflammatory medications, worsening swelling or salivary glands-signs of infection, Significant weight loss due to difficulty eating and swallowing]." [PID 8, Thyroid Surgeon]	Knowledge
Symptom management video Description: benefits of RAI symptom self-management video on self-efficacy; consequences of self-management videos on self-efficacy; recommended changes to enhance patient engagement in self-management symptom videos	• "Having this particular type of video on how to massage salivary glands, I would be interested in being able to go to the site to see precisely how I can do this and have a video to demonstrate just how it can be performed. So I'm interested to tap the play button." [Focus Group 4, PID 1] • "There might be a little discussion of who might need this because if I just had RAI and I'm out of isolation, am I supposed to do this irrespective of symptoms ?" [PID 4, Patient Advocate]	Self-efficacy and self-regulation

XSL•FO
RenderX

Mockup	Exemplar quotes	SCT[a] component
FAQ[c] page Description: need for relevant and concise questions related to the post-RAI treatment timeframe. Stakeholders suggested providing answers to the questions within the web-based intervention to reduce the potential for increased demands on provider workflows.	• "Some of the things that I would not think about are dental appointments. I didn't think about that until recently, after I had my sialendoscopy. But other than that, all the other questions are relevant. You don't really think about it before or during the RAI treatment, because you didn't know that was a concern." [Focus Group 2, PID 2] • "Is there a value to having answers for these [questions]? I would wonder if I'm interested in the first question, and I want to ask my provider that, I also want to know the answer now. Or I want to get some guidance about that now. I don't want to wait until my clinic appointment in 2 weeks to get the answer. That was what, I wonder, is there a role for that? But with some answers. " [PID 6, Palliative Care Provider]	Self-regulation and outcome expectations

[a]SCT: Social Cognitive Theory.

[b]RAI: radioactive iodine.

[c]FAQ: frequently asked question.

XSL•FO
RenderX

Table . Subthemes and exemplar quotes related to domain 3 intervention content to promote RAI[a] symptom management and survivorship care needs within the SCT[b] framework.

Subthemes	Exemplar Quotes	SCT Construct
Nutrition description: information on nutrition strategies for RAI symptom management strategies for salivary symptoms. Topics included details about foods that stimulate salivary glands and help manage changes in taste.	• "It could be a video about what foods might be helpful [to] stimulate your salivary glands. And then, in case you lost taste, recommendations for you to still taste something, and being able to eat without affecting the health of your mouth." [Focus Group 5, PID 2] • "If there were tips presented earlier about managing dry mouth like ensuring adequate hydration, ensuring fluid or moisture, rich food intake, that is something that could be helpful." [PID 10 Registered oncology dietitian]	Self-efficacy and self-regulation
Communication with family description: information and strategies to help guide patient-led communication with families and children, about RAI symptoms. Topics included practical tips on how to talk about the social and emotional impact of DTC[c] diagnosis and RAI symptoms with children and family members.	• "I was thinking [what] would also be nice [is] to have some type of resource center for family and friends on how to best support someone with thyroid cancer ... [It is]a nice way of how best to support someone with thyroid cancer." [Focus Group 4, PID 1] • " How to speak to them, when they're young and explain the survival rate is very good. And when it comes to cancers, it's not [a] death sentence, because you hear the cancer part, but you know it is survivable." [Focus Group 6, PID 3] • "That's important, especially, if your children are young, [and] they hear the diagnosis of cancer. How you can communicate that to the children, if there is information on that, I think that would be helpful for patients." [PID 3, Nuclear Medicine Provider] • " I worry about broad strokes because what do you tell a 6-year-old versus the 10-year-old versus a 13-year-old and not all 13-year-olds are the same." [PID 8, Thyroid Surgeon] • " I think that's [communication] also intimidating for people. Some people have a better general sense of what this treatment is about, and how it can or can't impact others. And other people are less familiar and aware or able to articulate what they know about their treatment and how it can impact others. And I think that can just be hard and an additional burden on the patient." [PID 11, Psycho-oncologist]	Self-efficacy, and self-regulation

XSL•FO
RenderX

Subthemes	Exemplar Quotes	SCT Construct
Cancer survivorship follow-up care description: discussed the need for a general overview of lifelong follow-up with health care providers, including, which health care team members to consult with for RAI symptom management and follow-up tests (blood, ultrasound, or whole body radioiodine scans).	• "Are you going back to the nuclear medicine team? Are you [going to] your endo? Who's supposed to be quarterbacking all of your treatment? The right person? Is it whoever did your surgery? Who's the right person depending on what the symptoms are?" [Focus Group 1, PID 5] • "What would be helpful is how to explain to other providers that you've had this treatment, and these are some of the possible side effects ... I think having something like a quick reference sheet where they [patients] can talk to their doctor." [Focus Group 6, PID 2] • "Patients should know that they basically have this diagnosis of cancer, and that they need lifelong follow up. Follow up with their endocrinologists, follow up with their primary care, or whoever is managing their dose which they get for their thyroid hormone. Or to make sure that they that they do not get recurrence in the future." [PID 3, Nuclear Medicine Provider]	Knowledge, self-regulation, and outcome expectations

[a]RAI: radioactive iodine.

[b]SCT: Social Cognitive Theory.

[c]DTC: differentiated thyroid cancer.

Domain 1: Difficult-to-Manage RAI Symptoms: Short, Medium, and Late Treatment Effects

As listed in Table 5, we indicated the percentage of focus group participants and stakeholders who commented on difficult-to-manage RAI symptoms. In our sample, approximately half of our focus group participants (11/22) endorsed symptoms of dry mouth and changes in taste as challenging symptoms to manage, followed by salivary gland swelling (8/22, 36%), dry eyes (7/22, 31%), and dry nose (3/22, 13%). Stakeholders endorsed a similar prevalence of difficult-to-manage RAI symptoms experienced by patients in their clinical care, including dry mouth (6/12, 50%), changes in taste (5/12, 41%), salivary gland swelling (4/12, 33%), swallowing difficulties (3/12, 25%), and dry eyes (3/12, 25%). Focus groups highlighted that among patients with DTC post-RAI treatment, the effectiveness of their current self-management strategies varies across individuals. Participants emphasized the need for personalized approaches to manage the short and late effects of RAI treatment. Many stakeholders emphasized the need for patient referrals and collaboration with multidisciplinary care teams to enhance overall patient care and address severe RAI symptom management.

XSL•FO
RenderX

Table . RAI-specific[a] symptoms referenced as difficult to manage by focus group participants and stakeholders during qualitative interviews.

Difficult to manage RAI symptom	Focus group participants (n=22), n (%)	Stakeholders (n=12), n (%)
Salivary		
Dry mouth	11 (50)	6 (50)
Change in taste	12 (55)	5 (42)
Mouth ulcers	2 (9)	2 (17)
Salivary gland swelling	8 (36)	4 (33)
Swallowing difficulties	2 (9)	3 (25)
Dental issues	4 (18)	1 (8)
Lacrimal		
Dry eyes	7 (32)	3 (25)
Watery eyes	—[b]	1 (8)
Nasal		
Dry nose	3 (14)	
Loss of smell	1 (5)	—
Nose bleeds	2 (9)	—
Other		
Fatigue[c]	4 (18)	4 (33)
Headache	3 (14)	—
Gastrointestinal issues	2 (9)	—
Menstrual cycle changes[c]	1 (5)	—
Hair loss	5 (23)	—
Nausea	8 (36)	3 (25)

[a]RAI: radioactive iodine.

[b]Not endorsed by any focus group participants or stakeholders.

[c]Symptoms may occur from concurrent treatment such as thyroid stimulating hormone therapy.

Domain 2: Key Intervention Structure and Content Feedback on Mockups

The key message we derived from the focus group participant and stakeholder findings was that the web-based intervention requires a more streamlined home page narrative instructing patients on using the intervention in the context of experienced symptoms and personalized self-management strategies. Both focus group participants and stakeholders felt the home page mockup needed to provide relevant information about the website for a patient engaging in the intervention. Stakeholders emphasized the desirability of repeated language within the intervention content narrative to redirect patients to their health care team should emergent symptoms occur requiring medical oversight. Consequently, for the RAI support intervention to be perceived as credible, reliable, and trustworthy, focus group participants and stakeholders indicated it was very important that it include references to differentiate the intervention from nonacademic public-facing websites.

Overall, focus group participants and stakeholders found the education and self-management of symptoms mockup pages to be helpful components of a web-based intervention. Most focus group participants found that the education pages presented "distressing" symptoms in a simple and approachable visual format. Together, focus group participants and stakeholders found the RAI symptom management mockup to be useful and found the video format to allow for more engagement across a range of patient learning styles. Both patients and stakeholders emphasized the need for short video content (less than 3 minutes in length) and accessibility features such as closed captioning to maintain user video engagement. The frequently asked questions content page received mixed responses between focus group participants and stakeholders. Focus group participants highlighted the value of prepared questions to address patient RAI symptom concerns and increase patient-provider communication during appointments. In contrast, stakeholders shared concerns about the potential negative consequences on provider workload of "unanswered" questions that a patient may bring to a clinical encounter. Most stakeholders recommended answering the questions on the frequently asked questions page directly and including hyperlinks within the web-based intervention with answers to the questions. Both groups suggested the need to tailor questions to address RAI symptoms and aftercare and to use lay language to streamline information.

XSL•FO
RenderX

Domain 3: Intervention Content to Promote RAI Symptom Management and Survivorship Care

Focus group participants and stakeholders identified topics for intervention content and characteristics that had previously not been considered in the initial intervention design. Many focus group participants expressed uncertainty about which foods to eat after RAI treatment and suggested intervention content on foods that stimulate salivary gland secretion. Stakeholders reinforced the consideration of nutrition-based self-management strategies to reduce dry mouth and changes in taste, with an emphasis on the noncurative nature of these strategies. Many focus group participants indicated recommendations on how to communicate with family members about side effects, especially in light of the "good cancer" label, would be helpful after RAI treatment. A smaller number of focus group participants with children emphasized the need for intervention content on how to communicate with children about thyroid cancer, RAI treatment-related side effects, and survivorship. Most stakeholders supported the inclusion of communication strategies with family members but emphasized tailoring communication with adolescents and children at age-appropriate levels. Focus group participants expressed challenges in long-term monitoring for the management of RAI symptoms, monitoring for potential cancer recurrence, and not knowing which provider is appropriate for RAI symptom management.

Discussion

Principal Results

Following the SCT framework and the person-centered approach to inform the development of a theoretically driven mHealth symptom management intervention, we elicited feedback from focus group participants and stakeholders regarding which symptoms related to RAI treatment are the most challenging to manage. Findings in this study support evidence from prior research on RAI-related toxicities [11,12,16], including well-documented salivary gland swelling, xerostomia or dry mouth, and changes in taste [15]. Both patients and stakeholders emphasized the utility of a web-based intervention for patient self-management of mild to moderate symptoms with redirection to the patient health care team to treat severe symptoms.

Findings from this study build upon prior work investigating cancer survivor self-management interventions [44,45] with an emphasis on patient knowledge, self-efficacy, and skill development as intervention components from SCT [30,46]. Most self-management interventions are designed for a particular type of cancer or symptom [44,46], which does not capture the RAI treatment aftercare experiences among patients with DTC. RAI treatment-specific issues and aftercare needs reported by our focus group participants and stakeholders reinforce prior evidence of unmet informational needs related to RAI side effects and long-term symptom management in this population [1,17,19,47]. Both focus group participants and stakeholders found the content and digital format of the web-based mockups of salivary symptom education and a salivary gland massage video helpful in engaging and empowering patients in symptom self-management. These findings support the potential of web-based symptom management interventions in patient outcomes with accessible and digital information and support [20,28,44].

Additionally, our results may potentially address supportive care gaps through focus group patient and stakeholder identification of nutrition management strategies, communication with family, and follow-up care planning as tailored survivorship components to include in a web-based intervention. Survivorship care among this patient population involves unique coordination across specialists, with an endocrinologist or primary care provider at the center [17,47], and adjunct supportive programs such as a web-based intervention for patient self-management of RAI symptoms may offer assistance in reducing fragmented care gaps, improving self-efficacy in RAI symptom management, and reducing patient health care utilization.

Limitations

There are several limitations to this study. Our patients and stakeholders primarily identified as White and non-Hispanic, which limits the generalizability of the data. Furthermore, our patient sample comprised insured patients with thyroid cancer, and our qualitative findings did not capture the perspective of uninsured patients. Our qualitative results identified a gap in the current delivery of coordinated patient care across subspecialties. It should be noted that this care gap is reflective of one academic hospital center and is not generalizable across all settings. While our results reached saturation across stakeholder specialists, saturation within subspecialists did not occur due to the limited number of subspecialists enrolled in each group. In addition, we were not able to engage a dentist as a stakeholder. Future qualitative research should elicit endocrinologists' and primary care providers' perspectives on mHealth behavioral interventions for the self-management of RAI symptoms to enhance the implementation of coordinated care efforts.

A small subset of focus group participants attributed fatigue and menstrual cycle changes to RAI treatment; however, these are symptoms from concurrent treatments such as thyroidectomy and thyroid-stimulating hormones. Patients may have difficulty differentiating which symptoms are due to RAI or other concurrent treatment. In addition, symptoms such as menstrual cycle changes could be due to both RAI and adjustment of thyroid hormone dosage. Future studies can work toward disambiguating symptoms for patient education and content development of a web-based intervention for RAI treatment symptom management.

Conclusions

Through a SCT framework and person-based approach, results from this qualitative study highlight important intervention content considerations such as the inclusion of nutrition management strategies, communication skills, and follow-up care planning for the self-management of RAI symptoms among patients with DTC. These findings can inform the development of future self-management interventions tailored to patients with DTC after RAI treatment.

XSL•FO
RenderX

Acknowledgments

The authors express their gratitude to all patients who participated in the focus group interviews and medical specialists, other health care providers, and patient advocates for their participation in individual interviews. The authors would like to thank Gary Bloom, the executive director of ThyCa: Thyroid Cancer Survivors' Inc, for his time and effort in the recruitment of patient advocates for the study. We would like to thank Emily Lattie for her time and effort in codeveloping the focus group facilitator guide. Finally, the authors acknowledge the efforts of Gautham Pillai for his contributions to the interviews. This work was supported by the National Cancer Institute (K99CA277588 and CCSG: P30-CA051008). ALC received research support from the National Cancer Institute (K99CA277588, CCSG: P30-CA051008, and T32CA261787). AMJ received research support from the American Cancer Society Center for Diversity in Cancer Research Training (supplement to ACS IRG-17-177-23) and her efforts were supported in part through funding for trainees provided by the Lombardi Comprehensive Cancer Center.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to following ethical principles for the protection of confidentiality given the small sample size in this study, and following the guidance of the Health Insurance Portability and Accountability Act per our consent process but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The open-ended qualitative focus group moderator guide is informed by Social Cognitive Theory and person-based design principles to elicit patient perspectives about RAI symptoms that were challenging to manage, identify content that would be helpful in a symptom management website, and obtain feedback on the formatting and content of static images or mockups.
[DOCX File, 37 KB - formative_v9i1e60588_app1.docx]

Multimedia Appendix 2

The open-ended qualitative semistructured interview guide is informed by Social Cognitive Theory and person-based design principles to elicit stakeholder perspectives about RAI symptoms that are challenging to manage, identify content that would be helpful in a symptom management website, and obtain feedback on the formatting and content of static images or mockups.
[DOCX File, 230 KB - formative_v9i1e60588_app2.docx]

Multimedia Appendix 3

Four mockups of a prospective web-based radioactive iodine treatment symptom management intervention reviewed by focus group participants and stakeholders: 1) home page, 2) salivary gland massage video, 3) salivary gland education page, and 4) frequently asked questions page.
[PDF File, 623 KB - formative_v9i1e60588_app3.pdf]

References

1. Diamond-Rossi SA, Jonklaas J, Jensen RE, et al. Looking under the hood of "the Cadillac of cancers:" radioactive iodine-related craniofacial side effects among patients with thyroid cancer. J Cancer Surviv 2020 Dec;14(6):847-857. [doi: 10.1007/s11764-020-00897-5]
2. Randle RW, Bushman NM, Orne J, et al. Papillary thyroid cancer: the good and bad of the "Good Cancer". Thyroid 2017 Jul;27(7):902-907. [doi: 10.1089/thy.2016.0632] [Medline: 28510505]
3. Cho S, Tutelman P, Chomistek T, et al. "No such thing as good cancer": a qualitative exploration of the experience of early-onset thyroid cancer in survivors. Support Care Cancer 2024 Sep 18;32(10):671. [doi: 10.1007/s00520-024-08850-3] [Medline: 39292348]
4. SEER cancer stat facts: thyroid cancer.: National Cancer Institute URL: https://seer.cancer.gov/statfacts/html/thyro.html [accessed 2024-11-25]
5. Guo H, Zhang N, Hu Y, Zhang F, Huang T, Shen N. Radioactive iodine therapy strategies for distinct types of differentiated thyroid cancer: a propensity score-matched analysis. Front Endocrinol (Lausanne) 2023;14:1158581. [doi: 10.3389/fendo.2023.1158581] [Medline: 37664843]
6. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines task force on thyroid nodules and differentiated thyroid cancer. Thyroid 2016 Jan;26(1):1-133. [doi: 10.1089/thy.2015.0020] [Medline: 26462967]

XSL•FO
RenderX

7. Nguyen NC, Anigati EM, Desai NB, Öz OK. Radioactive iodine therapy in differentiated thyroid cancer: an update on dose recommendations and risk of secondary primary malignancies. Semin Nucl Med 2024 Jul;54(4):488-496. [doi: 10.1053/j.semnuclmed.2024.05.002] [Medline: 38772827]
8. Goswami S, Peipert BJ, Mongelli MN, et al. Clinical factors associated with worse quality-of-life scores in United States thyroid cancer survivors. Surgery 2019 Jul;166(1):69-74. [doi: 10.1016/j.surg.2019.01.034] [Medline: 30898373]
9. Goswami S, Mongelli M, Peipert BJ, Helenowski I, Yount SE, Sturgeon C. Benchmarking health-related quality of life in thyroid cancer versus other cancers and United States normative data. Surgery 2018 Nov;164(5):986-992. [doi: 10.1016/j.surg.2018.06.042] [Medline: 30149935]
10. Applewhite MK, James BC, Kaplan SP, et al. Quality of life in thyroid cancer is similar to that of other cancers with worse survival. World J Surg 2016 Mar;40(3):551-561. [doi: 10.1007/s00268-015-3300-5] [Medline: 26546191]
11. Avram AM, Giovanella L, Greenspan B, et al. SNMMI procedure standard/EANM practice guideline for nuclear medicine evaluation and therapy of differentiated thyroid cancer: abbreviated version. J Nucl Med 2022 Jun;63(6):15N-35N. [Medline: 35649660]
12. Singer MC, Marchal F, Angelos P, et al. Salivary and lacrimal dysfunction after radioactive iodine for differentiated thyroid cancer: American Head and Neck Society Endocrine Surgery Section and Salivary Gland Section joint multidisciplinary clinical consensus statement of otolaryngology, ophthalmology, nuclear medicine and endocrinology. Head Neck 2020 Nov;42(11):3446-3459. [doi: 10.1002/hed.26417] [Medline: 32812307]
13. Dedhia PH, Saucke MC, Long KL, Doherty GM, Pitt SC. Physician perspectives of overdiagnosis and overtreatment of low-risk papillary thyroid cancer in the US. JAMA Netw Open 2022 Apr 1;5(4):e228722. [doi: 10.1001/jamanetworkopen.2022.8722] [Medline: 35467735]
14. Thyroid carcinoma. NCCN clinical practice guidelines in oncology version 5. 2025. URL: https://www.nccn.org/professionals/physician_gls/pdf/thyroid_blocks.pdf [accessed 2025-03-04]
15. Lee SL. Complications of radioactive iodine treatment of thyroid carcinoma. J Natl Compr Cancer Netw 2010;8(11):1277-1287. [doi: 10.6004/jnccn.2010.0094]
16. Jonklaas J. Nasal symptoms after radioiodine therapy: a rarely described side effect with similar frequency to lacrimal dysfunction. Thyroid 2014 Dec;24(12):1806-1814. [doi: 10.1089/thy.2014.0162] [Medline: 25090584]
17. Banach R, Bartès B, Farnell K, et al. Results of the Thyroid Cancer Alliance international patient/survivor survey: psychosocial/informational support needs, treatment side effects and international differences in care. Hormones (Athens) 2013;12(3):428-438. [doi: 10.1007/BF03401308] [Medline: 24121384]
18. Carr AL, Pillai G, Jonklaas J, Bloom G. Health-related quality of life associations in patients with differentiated thyroid cancer after radioactive iodine treatment: a cross-sectional study. Thyroid® 2023;33(S1):A-74. [doi: 10.1089/thy.2023.29156.abstracts]
19. Hyun YG, Alhashemi A, Fazelzad R, Goldberg AS, Goldstein DP, Sawka AM. A systematic review of unmet information and psychosocial support needs of adults diagnosed with thyroid cancer. Thyroid 2016 Sep;26(9):1239-1250. [doi: 10.1089/thy.2016.0039] [Medline: 27350421]
20. Walsh JC, Groarke JM. Integrating behavioral science with mobile (mHealth) technology to optimize health behavior change interventions. Eur Psychol 2019 Jan;24(1):38-48. [doi: 10.1027/1016-9040/a000351]
21. Penedo FJ, Oswald LB, Kronenfeld JP, Garcia SF, Cella D, Yanez B. The increasing value of eHealth in the delivery of patient-centred cancer care. Lancet Oncol 2020 May;21(5):e240-e251. [doi: 10.1016/S1470-2045(20)30021-8] [Medline: 32359500]
22. Hong YA, Hossain MM, Chou WYS. Digital interventions to facilitate patient-provider communication in cancer care: a systematic review. Psychooncology 2020 Apr;29(4):591-603. [doi: 10.1002/pon.5310] [Medline: 31834650]
23. Hesse BW, Kwasnicka D, Ahern DK. Emerging digital technologies in cancer treatment, prevention, and control. Transl Behav Med 2021 Nov 30;11(11):2009-2017. [doi: 10.1093/tbm/ibab033]
24. Aapro M, Bossi P, Dasari A, et al. Digital health for optimal supportive care in oncology: benefits, limits, and future perspectives. Support Care Cancer 2020 Oct;28(10):4589-4612. [doi: 10.1007/s00520-020-05539-1]
25. Davis SW, Oakley-Girvan I. mHealth education applications along the cancer continuum. J Cancer Educ 2015 Jun;30(2):388-394. [doi: 10.1007/s13187-014-0761-4] [Medline: 25482319]
26. Richards R, Kinnersley P, Brain K, McCutchan G, Staffurth J, Wood F. Use of mobile devices to help cancer patients meet their information needs in non-inpatient settings: systematic review. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e10026. [doi: 10.2196/10026] [Medline: 30552082]
27. O'Leary DP, Zaheer A, Redmond HP, Corrigan MA. Integration of advances in social media and mHealth technology are pivotal to successful cancer prevention and control. Mhealth 2016;2:38. [doi: 10.21037/mhealth.2016.09.02] [Medline: 28293611]
28. Fridriksdottir N, Gunnarsdottir S, Zoëga S, Ingadottir B, Hafsteinsdottir EJG. Effects of web-based interventions on cancer patients' symptoms: review of randomized trials. Support Care Cancer 2018 Feb;26(2):337-351. [doi: 10.1007/s00520-017-3882-6] [Medline: 28921391]

XSL•FO
RenderX

29. Moradian S, Voelker N, Brown C, Liu G, Howell D. Effectiveness of internet-based interventions in managing chemotherapy-related symptoms in patients with cancer: a systematic literature review. Support Care Cancer 2018 Feb;26(2):361-374. [doi: 10.1007/s00520-017-3900-8] [Medline: 28948360]
30. Bandura A. Health promotion by social cognitive means. Health Educ Behav Off Publ Soc Public Health Educ 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
31. Graves KD. Social cognitive theory and cancer patients' quality of life: a meta-analysis of psychosocial intervention components. Health Psychol Off J Div Health Psychol Am Psychol Assoc 2003 Mar;22(2):210-219. [Medline: 12683741]
32. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30. [doi: 10.2196/jmir.4055] [Medline: 25639757]
33. Stacey FG, James EL, Chapman K, Courneya KS, Lubans DR. A systematic review and meta-analysis of social cognitive theory-based physical activity and/or nutrition behavior change interventions for cancer survivors. J Cancer Surviv 2015 Jun;9(2):305-338. [doi: 10.1007/s11764-014-0413-z] [Medline: 25432633]
34. Rogers LQ, Matevey C, Hopkins-Price P, Shah P, Dunnington G, Courneya KS. Exploring social cognitive theory constructs for promoting exercise among breast cancer patients. Cancer Nurs 2004;27(6):462-473. [doi: 10.1097/00002820-200411000-00006] [Medline: 15632787]
35. Hatchett A, Hallam JS, Ford MA. Evaluation of a social cognitive theory-based email intervention designed to influence the physical activity of survivors of breast cancer. Psychooncology 2013 Apr;22(4):829-836. [doi: 10.1002/pon.3082] [Medline: 22573338]
36. Short CE, James EL, Plotnikoff RC. How social cognitive theory can help oncology-based health professionals promote physical activity among breast cancer survivors. Eur J Oncol Nurs 2013 Aug;17(4):482-489. [doi: 10.1016/j.ejon.2012.10.009] [Medline: 23177321]
37. Graves KD, Carter CL, Anderson ES, Winett RA. Quality of life pilot intervention for breast cancer patients: use of social cognitive theory. Palliat Support Care 2003 Jun;1(2):121-134. [doi: 10.1017/s1478951503030268] [Medline: 16594274]
38. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
39. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
40. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
41. Krippendorff K. Estimating the reliability, systematic error and random error of interval data. Educ Psychol Meas 1970 Apr;30(1):61-70. [doi: 10.1177/001316447003000105]
42. Krippendorff K. Content Analysis: An Introduction to Its Methodology, 2nd edition: SAGE Publications; 2004.
43. Marzi G, Balzano M, Marchiori D. K-alpha calculator-Krippendorff's alpha calculator: a user-friendly tool for computing Krippendorff's alpha inter-rater reliability coefficient. MethodsX 2024 Jun;12:102545. [doi: 10.1016/j.mex.2023.102545] [Medline: 39669968]
44. Leach CR, Hudson SV, Diefenbach MA, et al. Cancer health self-efficacy improvement in a randomized controlled trial. Cancer 2022 Feb 1;128(3):597-605. [doi: 10.1002/cncr.33947] [Medline: 34668569]
45. Boland L, Bennett K, Connolly D. Self-management interventions for cancer survivors: a systematic review. Support Care Cancer 2018 May;26(5):1585-1595. [doi: 10.1007/s00520-017-3999-7] [Medline: 29199362]
46. Rimmer B, Brown MC, Sotire T, et al. Characteristics and components of self-management interventions for improving quality of life in cancer survivors: a systematic review. Cancers (Basel) 2024;16(1):14. [doi: 10.3390/cancers16010014] [Medline: 39123377]
47. Kotwal A, Fingeret A, Knape A, Patel A, Bradford Bell E, Goldner W. Thyroid cancer survivorship: challenges and opportunities. Endocr Pract 2024 Nov;30(11):1097-1102. [doi: 10.1016/j.eprac.2024.08.003] [Medline: 39209023]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
DTC: differentiated thyroid cancer
HRQOL: health-related quality of life
mHealth: mobile health
RAI: radioactive iodine
REDCap: Research Electronic Data Capture
SCT: Social Cognitive Theory

XSL•FO
RenderX

Edited by A Mavragani; submitted 16.05.24; peer-reviewed by M Kim; revised version received 11.02.25; accepted 12.02.25; published 19.03.25.

Please cite as:
Carr AL, Jenkins AM, Jonklaas J, Gabriel K, Miller KE, Graves KD
Patient and Provider Perspectives of a Web-Based Intervention to Support Symptom Management After Radioactive Iodine Treatment for Differentiated Thyroid Cancer: Qualitative Study
JMIR Form Res 2025;9:e60588
URL: https://formative.jmir.org/2025/1/e60588
doi:10.2196/60588

© Alaina L Carr, Angela M Jenkins, Jacqueline Jonklaas, Kate Gabriel, Kristen E Miller, Kristi D Graves. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Customizing Computerized Adaptive Test Stopping Rules for Clinical Settings Using the Negative Affect Subdomain of the NIH Toolbox Emotion Battery: Simulation Study

Saki Amagai[1], BA; Aaron J Kaat[1], PhD; Rina S Fox[2], MPH, PhD; Emily H Ho[1], PhD; Sarah Pila[1], PhD; Michael A Kallen[1†], MPH, PhD; Benjamin D Schalet[3], PhD; Cindy J Nowinski[1], MD, PhD; Richard C Gershon[1], PhD

[1]Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[2]College of Nursing, University of Arizona, Tuscon, AZ, United States

[3]Department of Epidemiology and Data Science, Amsterdam University Medical Centers, Amsterdam, The Netherlands

[†]deceased

Corresponding Author:
Saki Amagai, BA
Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

Abstract

Background: Patient-reported outcome measures are crucial for informed medical decisions and evaluating treatments. However, they can be burdensome for patients and sometimes lack the reliability clinicians need for clear clinical interpretations.

Objective: We aimed to assess the extent to which applying alternative stopping rules can increase reliability for clinical use while minimizing the burden of computerized adaptive tests (CATs).

Methods: CAT simulations were conducted on 3 adult item banks in the NIH Toolbox for Assessment of Neurological and Behavioral Function Emotion Battery; the item banks were in the Negative Affect subdomain (ie, Anger Affect, Fear Affect, and Sadness) and contained at least 8 items. In the originally applied NIH Toolbox CAT stopping rules, the CAT was stopped if the score SE reached <0.3 before 12 items were administered. We first contrasted this with a SE-change rule in a planned simulation analysis. We then contrasted the original rules with fixed-length CATs (4 - 12 items), a reduction of the maximum number of items to 8, and other modifications in post hoc analyses. Burden was measured by the number of items administered per simulation, precision by the percentage of assessments yielding reliability cutoffs (0.85, 0.90, and 0.95), and accurate score recovery by the root mean squared error between the generating θ and the CAT-estimated "expected a posteriori"–based θ.

Results: In general, relative to the original rules, the alternative stopping rules slightly decreased burden while also increasing the proportion of assessments achieving high reliability for the adult banks; however, the SE-change rule and fixed-length CATs with 8 or fewer items also notably increased assessments yielding reliability <0.85. Among the alternative rules explored, the reduced maximum stopping rule best balanced precision and parsimony, presenting another option beyond the original rules.

Conclusions: Our findings demonstrate the challenges in attempting to reduce test burden while also achieving score precision for clinical use. Stopping rules should be modified in accordance with the context of the study population and the purpose of the study.

(JMIR Form Res 2025;9:e60215) doi:10.2196/60215

KEYWORDS

computerized adaptive testing; CAT; stopping rules; NIH Toolbox; reliability; test burden; clinical setting; patient-reported outcome; clinician

Introduction

Background

The patient's perspective, typically collected using patient-reported outcome measures (PROMs), is an imperative component of effective medical decision-making and treatment evaluation [1-4]. However, PROMs can be time-consuming for respondents to complete and difficult for clinicians to interpret, thus limiting their practical utility. Increasing importance has therefore been placed on minimizing response burden. For example, newer PROM initiatives have incorporated item response theory (IRT) and computerized adaptive test (CAT) capabilities, which enable the derivation of efficient, reliable, and precise assessments from large banks of items [5,6].

The NIH Toolbox for Assessment of Neurological and Behavioral Function is one such measurement initiative [7]. The NIH Toolbox is a set of neurobehavioral measures that can be used to assess sensory, motor, emotional, and cognitive

XSL•FO
RenderX

functioning among individuals aged 3 to 85 years. While many measures included in the NIH Toolbox focus on performance-based assessments, the Emotion Battery consists exclusively of PROMs assessing emotions such as sorrow, joy, or fear [8]. The battery consists of 4 subdomains: Negative Affect (Anger, Fear, and Sadness), Psychological Well-Being (Life Satisfaction, Meaning and Purpose, and Positive Affect), Social Relationships (Companionship, Fear, Perceived Social Support, Positive Social Development, and Social Distress), and Stress and Self-Efficacy (Self-Efficacy and Stress). Across all item banks, higher scores indicate higher levels of the construct being assessed (eg, a higher score on the Sadness measure indicates higher levels of sorrowful feelings or emotions) [8]. The item banks were calibrated using the graded response IRT model [9]. This calibration lines up all the items along a common mathematical metric, which allows any subset of items to be used to track a given patient's performance longitudinally or compare performance scores across patients under a common metric [10-12]. The NIH Toolbox Emotion Battery calculates "expected a posteriori" (EAP) scores, which are a type of IRT-based score, on the θ metric (mean 0, SD 1). These scores are then transformed to a *t* score metric (mean 50, SD 10) for individual reporting. The Emotion Battery provides a psychometrically strong option for evaluating emotional health that is particularly well suited for use in combination with other performance-based measures included in the NIH Toolbox.

Measures in the NIH Toolbox Emotion Battery are administered as fixed forms or variable-length CATs. The fixed forms consist of a set of items that are all administered in the same order during every administration. Conversely, when variable-length CATs are administered, the first item typically reflects average functioning or symptom severity, as the maximum weighted information value for a starting item is at the population mean (θ=0; *t* score=50). The response to that item is then used to calculate the provisional score estimate for the measured construct and to identify the best "next" item, within the remaining items in the bank, that will maximize weighted information based on the provisional ability estimate for the patient [13-15]. A patient's estimated θ score is iteratively refined after each item until (1) a minimum number of items is administered and either (2a) a prespecified level of measurement precision is achieved or (2b) a prespecified maximum number of items is administered. These conditions are known as "stopping rules" [16,17]. After 4 items have been administered (the minimum), the default for the NIH Toolbox CAT administration is to stop the CAT after either 12 items have been administered (the maximum) or the SE of the EAP score estimate is less than 0.3, which corresponds to a minimum reliability of 0.91. Other CATs may implement these or additional stopping rules, or may include content balancing or item exposure constraints. There is also a fully Bayesian approach for estimating ability using the Markov chain Monte Carlo method, which can be used to incorporate empirical information about candidates [18]. The use of empirical prior information has been shown to reduce response burden for patients and clinical practitioners [19]. While these other stopping rules may appear in other contexts (including in the NIH Toolbox Cognitive Battery for exposure constraints), the NIH Toolbox Emotion CATs do not implement any of these other features [20].

Compared to legacy measures, these stopping rules substantially reduce response burden, yet NIH Toolbox CATs still impose a significant burden on respondents with no or minimal symptoms. Similar issues have been documented in educational measurement using precision rules as a primary criterion for stopping rules [21]. For example, if an item bank has numerous items reflecting extreme sadness but few items reflecting minimal sadness, the statistically most-informative items for someone experiencing little-to-no sadness will be exhausted quickly, and additional items will be increasingly less targeted and unable to notably improve the score precision (ie, reducing the SE).

In such a scenario, the CAT will continue to select the next best (but still mistargeted) item until the maximum CAT length is reached. While there are sufficient items in each item bank to stratify pathology, if a person is not impaired with respect to the domain being explored, it is difficult or impossible to accurately differentiate how healthy a person is (eg, how not sad a person may be). As a result, patients with the best health in a given domain are more likely to be administered the maximum 12 items in a CAT as compared to patients experiencing more severe symptoms [22]. Therefore, the first goal of this study was to develop and systematically assess alternative stopping rules that would minimize response burden and frustration for these patients with few-to-no symptoms, as patients with wide range of levels of functioning are administered NIH Toolbox measures.

The second goal of this study was to alter stopping rules to improve reliability and, more specifically, to achieve clinically relevant reliability (eg, reliability ≥0.95). This is especially important today as the NIH Toolbox is increasingly being used with clinical populations and in clinical settings, even though it was developed specifically for use in research. As stated, the current CAT stopping rules call for a test to end when the SE is <0.3, which corresponds to a minimum reliability of 0.91. Although this level of reliability is recommended for use in general populations [23], it may be inadequate to identify clinically relevant symptoms and individual patient changes across time [24]. Thus, as NIH Toolbox measures are implemented more frequently for clinical use in health care delivery settings [25,26], the ability to achieve this level of reliability appears increasingly valuable. Reaching clinical-level reliability will lead to fewer errors in patient assessments and increase confidence in the treatments or type of care to be recommended.

Recent work has evaluated alternative approaches to CAT administration in an effort to promote more efficient and effective testing. In their simulation study, Kallen and colleagues [22] shortened the minimum test length for individuals completely denying symptoms to only 2 items and added a minimum SE-change rule, while retaining the SE <0.3 rule and a maximum test length of 12 items. The minimum SE-change rule allowed the CAT to stop when the CAT algorithm detected that the administration of additional items would not reduce the SE. The impact of adding the SE-change rule was negligible.

XSL•FO
RenderX

This simulation, combined with the shortened test length for individuals reporting absolutely no symptoms, addressed the concern that high-functioning individuals often receive lengthier CATs under the originally applied stopping rules for the NIH Toolbox. Importantly, this approach yielded scores that varied minimally from those obtained using the original stopping rules, particularly for scores in the clinically relevant range. Very few administrations required more than 8 items under these new stopping rules. However, the application of a SE <0.3 stopping rule retained a target reliability of 0.91, below a clinically relevant range (eg, ≥0.95), thus questioning the appropriateness of these rules as optimal for application in clinical settings.

Purpose of This Study

The NIH Toolbox has been increasingly used in real-world clinical settings, and patients with a wide range of levels of functioning complete these measures. The purpose of this study was to propose and evaluate new stopping rules that could optimally balance two competing priorities and goals that are increasingly common in both research and clinical practice:

1. Increase reliability (≥0.95) for the clinical population (ie, adjust the SE threshold to require higher reliability) without unacceptably increasing the number of items
2. Decrease response burden for nonsymptomatic individuals (ie, reduce the number of items administered) without unacceptably reducing reliability (<0.85)

By using in-person, in-production item banks, this study provides real-world insights that go beyond the theoretical understanding of how stopping rules function.

Methods

Planned Simulation Study

Design

Firestar was used to simulate examinee responses and item bank CAT administration [27]. Three adult (age ≥18 years) NIH Toolbox Negative Affect item banks (Anger Affect, Fear Affect, and Sadness) were evaluated. All 3 of the item banks contain at least 8 items and are IRT-calibrated and EAP-scored (Table 1). More detailed descriptions of the item banks, along with examples, can be found in Table S1 in Multimedia Appendix 1. The simulation study manipulated the CAT administration to implement both the original stopping rules and a new set of rules, which we named the “SE-change” rules. The SE-change rules terminated a test when (1) a minimum of 4 items or maximum of 8 items had been administered, thus decreasing the maximum test length; (2) SE was <0.224, thus increasing target reliability to 0.95; or (3) the SE was reduced by less than 0.01, thus stopping the test when measurement precision would not increase sufficiently by administering an additional item (Table 2). We hypothesized that the SE-change stopping rules would increase the proportion of assessments achieving clinically relevant reliability without unacceptably increasing the number of items administered to nonsymptomatic individuals.

Table . NIH Toolbox Emotion Battery measures from the Negative Affect subdomain included in the simulation.

Measure	Typical computer adaptive test administration (number of items available in bank)	Slope parameter, mean (SD)	Average within-item thresholds, mean (SD)
Anger Affect	22	2.22 (0.43)	1.45 (0.44)
Fear Affect	29	2.72 (0.78)	1.44 (0.40)
Sadness	28	3.14 (0.67)	1.27 (0.24)

XSL•FO
RenderX

Table . Stopping rule models evaluated.

Stopping rule model	Minimum items	Maximum items	SE threshold for interim stopping rule	SE-change threshold for interim stopping
Original	4	12	0.3 (reliability 0.91)	—[a]
SE-change	4	8	0.224 (reliability 0.95)	<0.01
4-item fixed	4	4	—	—
5-item fixed	5	5	—	—
6-item fixed	6	6	—	—
7-item fixed	7	7	—	—
8-item fixed	8	8	—	—
9-item fixed	9	9	—	—
10-item fixed	10	10	—	—
11-item fixed	11	11	—	—
12-item fixed	12	12	—	—
Reduced maximum	4	8	0.224 (reliability 0.95)	—

[a]Not applicable.

Interim item selection (using maximum posterior weighted information), starting item, and scoring (using EAP estimates with a standard normal prior) were constrained to be the same across all simulated CATs. For each stopping rule, 1000 responses were simulated with uniform distribution (for θ values from –4 to 4) using the full item banks. The simulations used a uniform distribution to eliminate artificial biases that could be inadvertently introduced if the person ability distribution was simulated such that it mimicked the item difficulty distribution. The unit of analysis was individual CAT assessments. The CAT simulations proceeded until one of the stopping rules was met. All simulations and analyses were conducted using the *Firestar* package [27] and R (version 3.6.3; R Foundation for Statistical Computing).

Evaluation Criteria

We used 4 evaluation criteria—precision, accurate score recovery, test burden, and efficiency—to compare the impact of the different stopping rules. (1) Precision was evaluated by exploring the percentages of assessments yielding reliabilities less than 0.85, between 0.85 and 0.90, between 0.90 and 0.95, and greater than 0.95 across stopping rules. This approach allowed us to compare the reliability achieved by each set of stopping rules in a more nuanced way rather than assessing mean reliability, which can obscure the impact of stopping rules on conditional reliability across the trait range. Additionally, maximum attainable information curves were generated to determine theoretically attainable reliability for fixed-length CATs with up to 12 items for each of the 3 item banks. (2) Accurate score recovery refers to an evaluation of whether the CAT was able to recover the generating score and index the variability in score recovery; for this, we calculated the root mean squared error (RMSE) between the generating θ and the CAT-estimated, EAP-based θ from each version of the CAT. (3) Test burden, or response burden, was evaluated by comparing the number of items administered as a function of each set of stopping rules. (4) Efficiency was defined as the Fisher information per number of items administered, which is equivalent to $(1/SE^2)$/item count, such that higher values represent greater efficiency.

Post Hoc Simulation Study

Achieving reliability ≥0.85 is considered a critical issue for the priority of increasing reliability. However, the reduced number of items in the SE-change stopping rule led to an increase in the assessments achieving low empirical reliability (<0.85). That is, assessments that had higher reliability (>0.85) under the original stopping rules were no longer able to maintain the same level of precision under the SE-change rules. As one of our two goals was to reduce burden without unacceptably reducing reliability, we conducted post hoc analyses in which we evaluated two new sets of stopping rules to more optimally balance the two goals: increasing reliability for the clinical population while decreasing burden for nonsymptomatic individuals.

All methods used in the planned simulations, as outlined, were replicated for this post hoc analysis; the only adjustments made were the addition of two new sets of stopping rules: (1) fixed-length CATs ranging in length from 4 to 12 items, and (2) “reduced maximum” stopping rules in which tests were terminated when SE was <0.224 or a maximum of 8 items had been administered (Table 2). The same evaluation criteria were used to compare the different stopping rules.

Ethical Considerations

The research presented herein was a simulation study, which was determined to not be human subjects research. Therefore, ethical review and approval was not necessary. This study was conducted using data from the NIH Toolbox Negative Affect item banks. The analysis did not involve identified patient information or any interaction with human subjects. As such, it did not meet the criteria for human subjects research and did not require ethics board review.

XSL•FO
RenderX

Results

SE-Change Stopping Rules vs Original Stopping Rules

The average percentage of simulated assessments yielding reliability at cutoffs of 0.85, 0.90, and 0.95 across stopping rules can be found in Table 3. For Fear Affect and Sadness, a larger percentage of assessments yielded reliability >0.95 per the SE-change stopping rules compared to the original stopping rules (+33.9% Fear Affect;+20.4% Sadness). For Anger Affect, neither set of stopping rules were able to yield the clinically relevant target reliability of 0.95. This result is because Anger Affect is a "weaker" bank (ie, the maximum attainable test information function is lower) that requires more items to be administered in order to distinguish different levels of anger and to reach 0.95 reliability. Most importantly, the percentage of assessments yielding low reliability for all three item banks (ie, <0.85) substantially increased (+2.7% Anger Affect;+4.7% Fear Affect;+3.4% Sadness).

Table . Proportion of simulated assessments meeting various levels of reliability.

Domain and level of reliability	Proportion of assessments meeting specified level of reliability (%)											
	Original	SE-change	4-item, fixed	5-item, fixed	6-item, fixed	7-item, fixed	8-item, fixed	9-item, fixed	10-item, fixed	11-item, fixed	12-item, fixed	Reduced maximum
Anger Affect												
<0.85	27.4	30.1	43.4	34.2	30.9	30.8	30	29.6	28.9	28.8	28.6	28.6
0.85-0.90	5.5	12.8	56.6	52.7	17.7	11.3	8.7	6.7	6.4	5.1	4.9	6.9
0.90-0.95	67.1	57.1	0	13.1	51.4	57.9	61.3	63.7	64.7	66.1	47.4	64.5
≥0.95	0	0	0	0	0	0	0	0	0	0	19.1	0
Fear Affect												
<0.85	31.5	36.2	36.6	36.4	35.2	35.2	35	33.8	33	32.9	32.6	33.5
0.85-0.90	3.5	5.7	15.7	12.9	9.2	8.1	7.2	7	6.2	5.1	5	4.9
0.90-0.95	65	24.2	47.7	50.7	25.3	20	15.8	14.2	15.1	15.6	14.2	21.8
≥0.95	0	33.9	0	0	30.3	36.7	42	45	45.7	46.4	48.2	39.8
Sadness												
<0.85	32.5	35.9	37.8	49.8	36.4	36.1	35.6	35.6	35.1	34.9	34.7	36.1
0.85-0.90	5.3	18.3	19.9	4.4	16	12.8	10.2	8.9	8	7.6	6.2	5.8
0.90-0.95	49.1	12.4	41.3	15.6	13.9	12.7	13.8	13.6	13.6	13.2	14.1	17.9
≥0.95	13.1	33.4	1	30.2	33.7	38.4	40.4	41.9	43.3	44.3	45	40.2

Concerning accurate score recovery, the RMSE of θ estimates are presented in Figure S1 in Multimedia Appendix 2 and Figure 1, respectively. The RMSE increased for Fear Affect (+0.036), while it decreased slightly for the other two banks (Anger Affect –0.001; Sadness –0.001). Test burden decreased for all banks per the SE-change rules as opposed to the original stopping rules, as shown in Table 4 (Anger Affect −1.26 items; Fear Affect −1.39 items; Sadness −2.64 items). In the original stopping rules, the average number of items administered was 7.9, compared to 6.1 items under the SE-change stopping rules, a 22% savings in response burden across banks (Table 4). The largest decrease in response burden observed was 33% for the Sadness bank.

XSL•FO
RenderX

Figure 1. Root mean squared error (RMSE) of θ estimates.

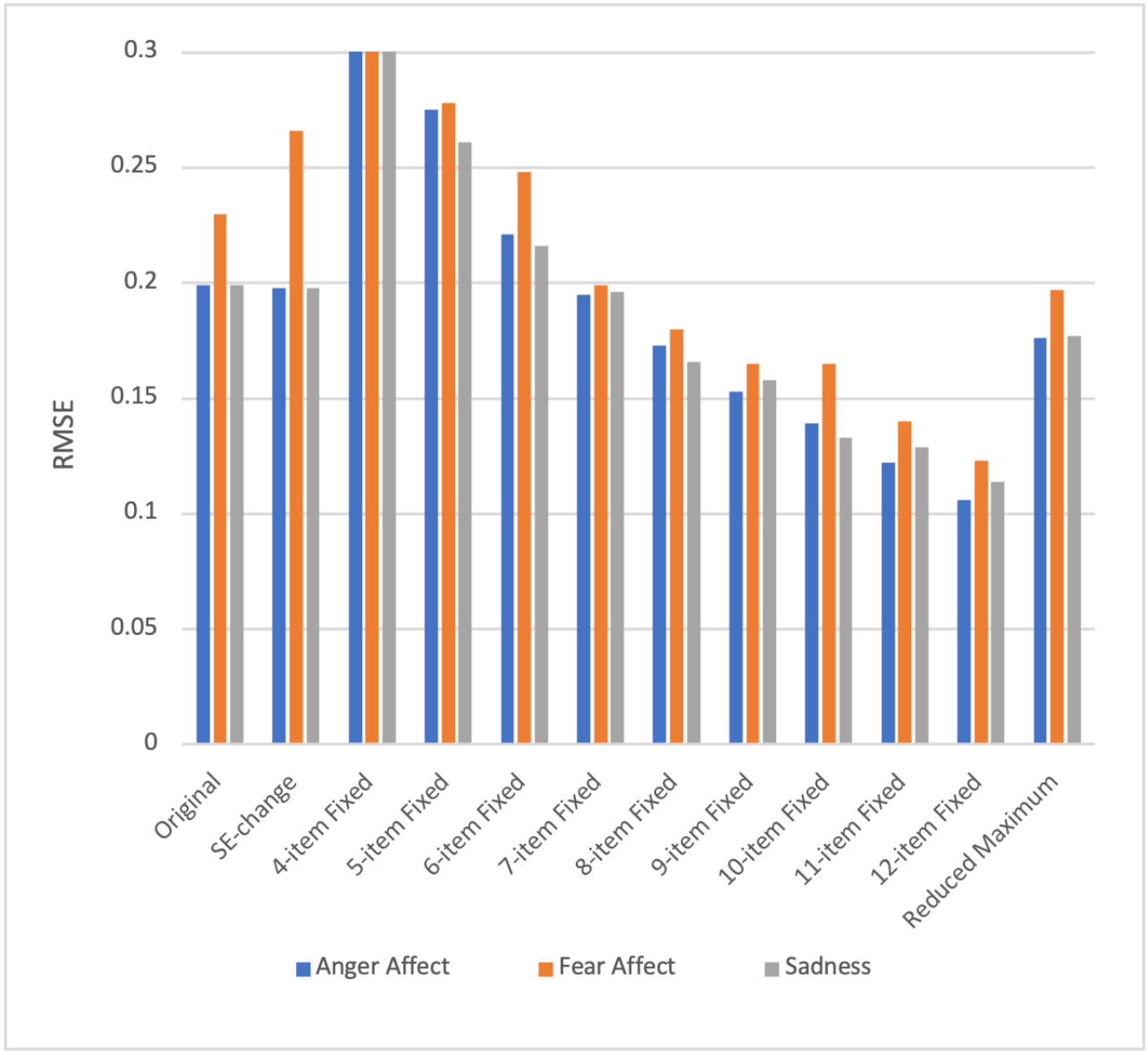


Table . Average number of items administered among 1000 simulated assessments.

Item bank	Number of items in item bank	Stopping rules		
		Original	SE-change	Reduced Maximum
Anger Affect	22	8.33	7.07	8.00
Fear Affect	29	7.40	6.01	7.36
Sadness	28	7.83	5.19	6.86

Fixed-Length CATs vs Original Stopping Rules

The fixed-length stopping rules led to an increased percentage of assessments reaching >0.85 reliability, as the test length is increased relative to the original stopping rules (Table 3). Per the fixed-length stopping rules, the target reliability level of 0.95 was achieved in the simulated assessments by only administering 6 items for Fear Affect and 4 items for Sadness, which is shorter than the average test length using the original stopping rules. However, more notably, the fixed-length CATs with 8 items or fewer also increased in simulated assessments with reliability <0.85 relative to the original stopping rules. Anger Affect was an exception and required 12 items to have any simulated assessments achieve the target reliability, and the assessments with reliability <0.85 did not improve relative to the original stopping rules even when administering 12 items.

Figure S2 in Multimedia Appendix 3 illustrates the potential inefficiencies for fixed-length CATs with up to 12 items. Consistent with the simulated results, Figure S2 in Multimedia Appendix 3 shows that there could be notable differences in the

XSL•FO
RenderX

number of items required to reach high reliability depending on the item bank. It is theoretically possible to achieve 0.95 reliability by administering at least 4 items for the Sadness and at least 5 items for the Fear Affect bank. Conversely, it can theoretically take at least 11 items for the Anger Affect bank to achieve 0.95 reliability, which suggests that the Anger Affect bank is a "weaker" bank that requires more items to be administered to distinguish levels of anger. Figure S2 in Multimedia Appendix 3 also suggests that individuals with little-to-no functioning of the trait (θ level=−1) will be subject to significant burden. For example, individuals with little-to-no sadness will never reach the 0.95 reliability threshold and will peak in maximum attainable information after 5 items, with diminishingly small gains in measurement precision. In this situation, high-functioning individuals will be administered more items until the maximum number of items is reached, thereby increasing burden.

The CAT was able to recover the generating score better when more items were administered in the fixed-length stopping rules as opposed to the original stopping rules. That is, RMSE of θ estimates decreased per the fixed-length stopping rules as opposed to the original stopping rules when administering more than 8 items for all banks. However, this improved accuracy score recovery came at a cost, as any fixed-length test greater than 8 items increases test burden relative to the original stopping rules (Table 4).

Reduced Maximum Stopping Rules vs Original Stopping Rules

A larger percentage of assessments yielded reliability >0.95 per the reduced maximum stopping rules as opposed to the original stopping rules (+0% Anger Affect, +39.8% Fear Affect, +40.2% Sadness). Importantly, this change did not excessively increase the percentage of assessments achieving low reliability (+1.2% Anger Affect; +2.0% Fear Affect; +3.6% Sadness).

The reduced maximum stopping rules improved the accuracy score recovery. The generating θ scores and score estimates were strongly correlated (>0.99) across banks for both stopping rules. RMSE also decreased for all banks (Anger Affect –0.023; Fear Affect –0.033; Sadness –0.022). Additionally, the average number of items administered was 7.9 compared to 7.4 items under the reduced maximum, a 6% savings in response burden across banks (Table 4). The largest decrease in response burden observed was 12% for the Sadness bank. Figure S3 in Multimedia Appendix 4 illustrates that the reduced maximum stopping rule may not be the most efficient choice across the θ range, but in general, all the rules are broadly comparable across banks.

Discussion

Principal Findings

Measures within the NIH Toolbox Emotion Battery that are administered as CATs are programmed to terminate when prespecified stopping rules are achieved. While the originally applied stopping rules are effective for many test takers, nonsymptomatic individuals often are burdened with many questions that are not relevant to their current health status. Additionally, many of the item banks have sufficient statistical information to achieve clinical-level reliability, but the original stopping rules terminate the assessment before that level of reliability is met. With increasing use of the NIH Toolbox in clinical settings, it is critical to simultaneously balance competing priorities: (1) maximizing score precision for clinical use and (2) minimizing burden for nonsymptomatic individuals. We evaluated alternative stopping rules to achieve these two goals.

The SE-change stopping rules were effective at reducing overall burden and were effective at ensuring adequate reliability and score recovery. Perhaps most importantly, although the SE-change stopping rules increased the percentage of assessments achieving clinically relevant reliability (≥0.95), the percentage of assessments achieving low empirical reliability (<0.85) also increased to an unacceptable level. Ultimately, we determined that the detrimental impact of the SE-change stopping rules on the percentage of assessments achieving low empirical reliability outweighed the modest benefits observed regarding test length and the percentage achieving very high reliability. This finding prompted us to explore additional stopping rule paradigms in post hoc simulations.

The fixed-length stopping rules were also not effective. Although it was possible to reduce burden relative to the original stopping rules by administering fewer than 8 items, it came at the cost of inflating the percentage of assessments achieving low empirical reliability (<0.85) while worsening accurate score recovery. By administering 8 or more items, it was also possible to maximize assessments achieving high reliability (>0.95) but at the cost of increasing burden. Given the benefits of longer fixed-length CATs, we opted to explore the reduced maximum stopping rules to determine if we could maintain this improvement in reliability without requiring 8 items across all administrations of all banks.

This subsequent evaluation identified a paradigm that more effectively balanced precision and parsimony. In general, the reduced maximum stopping rules demonstrated the same increase in reliability obtained with the 8-item fixed-length CATs; however, these stopping rules allowed for further decreased burden, as 8 items were not always required to reach a reliability level of 0.95.

One exception to the observed trends was the Anger Affect item bank. In simulations, this bank was unable to achieve reliability >0.95 for the reduced maximum stopping rule, as greater than 11 items are theoretically required to reach the reliability level of 0.95. As Figure S2 in Multimedia Appendix 3 showed, Anger Affect is an example of a "weak" bank, in which more items are needed to be administered to obtain clinically relevant reliability. The strength or weakness of a bank is related to the magnitude of the item slope parameters and the (statistically related) test information function. The Anger slope parameters are lower—therefore they are less-strongly related to the latent construct and contain more "error variance" or noise in their ratings. One possible hypothesis is that there are numerous and varied reasons someone might endorse or not endorse an Anger item, so the construct as a whole is not as cleanly defined as sadness or fear. Individuals may engage in more self-monitoring

XSL•FO
RenderX

about anger. Although Anger Affect will never achieve our clinically relevant target reliability with the reduced maximum stopping rule, it is not in our best interest to increase the test length to accommodate these "weaker" banks, as this will lead to increased response burden for all banks.

Given the competing priorities of improving reliability while minimizing burden, decisions for stopping rules should be adjusted depending on the context of the study population and the purpose of the study. The results of our study will be also relevant for other users who share a similar goal in achieving high reliability for use in clinical settings, as the NIH Toolbox has been increasingly used with clinical populations [28], while minimizing burden for nonsymptomatic individuals. When determining stopping rules, users should also consider the reliability achievable with the given item bank, the target reliability needed to answer their research or clinical questions, and the range of the domain in the sample population. Stopping rules should be modified after carefully considering these varied measurement contexts.

Limitations

Findings are limited to the 3 NIH Toolbox Emotion Battery adult item banks studied. As this study was designed to be largely a statistical investigation, our planned and post hoc stopping rules only incorporated psychometric considerations. These stopping rules were not applied to identical response data, although all simulations were based on the same uniform population distribution. To avoid artificial biases, we conducted the analyses in a way that prevented the person ability distribution from being simulated to match the item difficulty distribution. However, this may restrict the generalizability of our findings to real-world person distributions [29,30]. Indeed, using SE stopping rules in a multidimensional CAT setting can reduce test length by upward of 33% [31-33].

We did not explore multidimensional CATs for this study because those assessments are built on the assumption that all measures within a given composite (eg, Negative Affect subdomain) are important and should be assessed. While we expect clinicians and researchers would want to use all measures within a given subdomain, we chose not to impose that requirement. Future studies could extend the analyses to other subdomains and PROMs by considering multidimensional adaptive testing and theoretically driven content balancing techniques. As this analysis only examined adult responses, future extensions of this work could also examine the Emotion Battery for a pediatric population.

Conclusion

Taken together, these results demonstrate the challenges in attempting to reduce test burden while also achieving score precision for clinical use. The reduced maximum stopping rules provide one acceptable strategy to achieve these goals. The SE-change stopping rules and fixed-length stopping rules with less than 8 items are not recommended due to the sizable increase in simulations achieving empirical reliability <0.85. The reduced maximum stopping rules maximized the proportion of simulations achieving clinically relevant reliability (>0.95) without excessively inflating those achieving low reliability; they also decreased burden by allowing for fewer items to be administered when sufficient. That said, the reduced maximum stopping rules are not recommended for *all* occasions. Stopping rules should be modified in accordance with the context of the study population and the purpose of the study.

Acknowledgments

The authors thank the National Institutes of Health (NIH) for providing funding for this study. This study is funded in whole or in part with federal funds from the Blueprint for Neuroscience Research, NIH, under contract number HHS-N-260-2006-00007-C, by the Environmental Influences on Child Health Outcomes (ECHO) program, Office of the Director, NIH, under award number U24OD023319, and under contract number HHS-N-271-2018-0072-P.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of item banks.
[DOCX File, 15 KB - formative_v9i1e60215_app1.docx]

Multimedia Appendix 2
Correlation of the generating θ score with the estimated expected a posteriori score.
[DOCX File, 85 KB - formative_v9i1e60215_app2.docx]

Multimedia Appendix 3
Maximum attainable information curves for Anger Affect, Fear Affect, and Sadness.
[DOCX File, 361 KB - formative_v9i1e60215_app3.docx]

Multimedia Appendix 4

XSL•FO
RenderX

Comparisons of computerized adaptive test stopping rule efficiencies.
[DOCX File, 185 KB - formative_v9i1e60215_app4.docx]

References

1. Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century: National Academies Press (US); 2001. [Medline: 25057539]
2. Anker SD, Agewall S, Borggrefe M, et al. The importance of patient-reported outcomes: a call for their comprehensive integration in cardiovascular clinical trials. Eur Heart J 2014 Aug 7;35(30):2001-2009. [doi: 10.1093/eurheartj/ehu205] [Medline: 24904027]
3. Deshpande PR, Rajan S, Sudeepthi BL, Abdul Nazir CP. Patient-reported outcomes: a new era in clinical research. Perspect Clin Res 2011 Oct;2(4):137-144. [doi: 10.4103/2229-3485.86879] [Medline: 22145124]
4. Weldring T, Smith SMS. Patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs). Health Serv Insights 2013;6:61-68. [doi: 10.4137/HSI.S11093] [Medline: 25114561]
5. Carlozzi NE, Kallen MA, Sander AM, et al. The development of a new computer adaptive test to evaluate anxiety in caregivers of individuals with traumatic brain injury: TBI-CareQOL Caregiver-Specific Anxiety. Arch Phys Med Rehabil 2019 Apr;100(4S):S22-S30. [doi: 10.1016/j.apmr.2018.05.027] [Medline: 29958902]
6. Smith AB, Hanbury A, Retzler J. Item banking and computer-adaptive testing in clinical trials: standing in sight of the PROMISed land. Contemp Clin Trials Commun 2018 Nov;13:100306. [doi: 10.1016/j.conctc.2018.11.005] [Medline: 30582066]
7. Gershon RC, Wagster MV, Hendrie HC, Fox NA, Cook KF, Nowinski CJ. NIH toolbox for assessment of neurological and behavioral function. Neurology (ECronicon) 2013 Mar 12;80(11 Suppl 3):S2-S6. [doi: 10.1212/WNL.0b013e3182872e5f] [Medline: 23479538]
8. Salsman JM, Butt Z, Pilkonis PA, et al. Emotion assessment using the NIH Toolbox. Neurology (ECronicon) 2013 Mar 12;80(11 Suppl 3):S76-S86. [doi: 10.1212/WNL.0b013e3182872e11] [Medline: 23479549]
9. Samejima F. Handbook of Moden Item Response Theory: Springer; 1997.
10. Chang CH, Reeve BB. Item response theory and its applications to patient-reported outcomes measurement. Eval Health Prof 2005 Sep;28(3):264-282. [doi: 10.1177/0163278705278275] [Medline: 16123257]
11. Nguyen TH, Han HR, Kim MT, Chan KS. An introduction to item response theory for patient-reported outcome measurement. Patient 2014;7(1):23-35. [doi: 10.1007/s40271-013-0041-0] [Medline: 24403095]
12. Stover AM, McLeod LD, Langer MM, Chen WH, Reeve BB. State of the psychometric methods: patient-reported outcome measure development and refinement using item response theory. J Patient Rep Outcomes 2019 Jul 30;3(1):50. [doi: 10.1186/s41687-019-0130-5] [Medline: 31359210]
13. Choi SW, Swartz RJ. Comparison of CAT item selection criteria for polytomous items. Appl Psychol Meas 2009 Sep 1;33(6):419-440. [doi: 10.1177/0146621608327801] [Medline: 20011456]
14. Lunz ME, Bergstrom BA, Gershon RC. Computer adaptive testing. Int J Educ Res 1994 Jan;21(6):623-634. [doi: 10.1016/0883-0355(94)90015-9]
15. van der Linden WJ. Bayesian item selection criteria for adaptive testing. Psychometrika 1998 Jun;63(2):201-216. [doi: 10.1007/BF02294775]
16. Wainer H. Computerized Adaptive Testing: A Primer, 2nd edition: Lawrence Erlbaum Associates; 2000.
17. Gershon R, Cella D, Dineen K, Rosenbloom S, Peterman A, Lai JS. Item response theory and health-related quality of life in cancer. Expert Rev Pharmacoecon Outcomes Res 2003 Dec;3(6):783-791. [doi: 10.1586/14737167.3.6.783] [Medline: 19807355]
18. Matteucci M, Veldkamp BP. On the use of MCMC computerized adaptive testing with empirical prior information to improve efficiency. Stat Methods Appl 2013 Jun;22(2):243-267. [doi: 10.1007/s10260-012-0216-1]
19. Frans N, Braeken J, Veldkamp BP, Paap MCS. Empirical priors in polytomous computerized adaptive tests: risks and rewards in clinical settings. Appl Psychol Meas 2023 Jan;47(1):48-63. [doi: 10.1177/01466216221124091] [Medline: 36425285]
20. Leung CK, Chang HH, Hau KT. Incorporation of content balancing requirements in stratification designs for computerized adaptive testing. Educ Psychol Meas 2003 Apr;63(2):257-270. [doi: 10.1177/0013164403251326]
21. Yao L. Comparing the performance of five multidimensional CAT selection procedures with different stopping rules. Appl Psychol Meas 2013 Jan;37(1):3-23. [doi: 10.1177/0146621612455687]
22. Kallen MA, Cook KF, Amtmann D, Knowlton E, Gershon RC. Grooming a CAT: customizing CAT administration rules to increase response efficiency in specific research and clinical settings. Qual Life Res 2018 Sep;27(9):2403-2413. [doi: 10.1007/s11136-018-1870-z] [Medline: 29730850]
23. Streiner DL, Norman GR, Cairney J. Health Measurement Scales: A Practical Guide to Their Development and Use: Oxford University Press; 2015.
24. Ho EH, Verkuilen J, Fischer F. Measuring individual true change with PROMIS using IRT-based plausible values. Qual Life Res 2023 May;32(5):1369-1379. [doi: 10.1007/s11136-022-03264-2] [Medline: 36282446]

XSL•FO
RenderX

25. Smith AW, Jensen RE. Beyond methods to applied research: realizing the vision of PROMIS®. Health Psychol 2019 May;38(5):347-350. [doi: 10.1037/hea0000752] [Medline: 31045416]
26. Yount SE, Cella D, Blozis S. PROMIS®: standardizing the patient voice in health psychology research and practice. Health Psychol 2019 May;38(5):343-346. [doi: 10.1037/hea0000741] [Medline: 31045415]
27. Choi SW. Handbook of Item Response Theory, 1st edition: Chapman and Hall/CRC; 2017.
28. Fox RS, Zhang M, Amagai S, et al. Uses of the NIH Toolbox® in clinical samples: a scoping review. Neurol Clin Pract 2022 Aug;12(4):307-319. [doi: 10.1212/CPJ.0000000000200060] [Medline: 36382124]
29. Paap MCS, Born S, Braeken J. Measurement efficiency for fixed-precision multidimensional computerized adaptive tests: comparing health measurement and educational testing using example banks. Appl Psychol Meas 2019 Jan;43(1):68-83. [doi: 10.1177/0146621618765719] [Medline: 30573935]
30. Yao L, Pommerich M, Segall DO. Using multidimensional CAT to administer a short, yet precise, screening test. Appl Psychol Meas 2014 Nov;38(8):614-631. [doi: 10.1177/0146621614541514]
31. Li YH, Schafer WD. Trait parameter recovery using multidimensional computerized adaptive testing in reading and mathematics. Appl Psychol Meas 2005 Jan;29(1):3-25. [doi: 10.1177/0146621604270667]
32. Luecht RM. Multidimensional computerized adaptive testing in a certification or licensure context. Appl Psychol Meas 1996 Dec;20(4):389-404. [doi: 10.1177/014662169602000406]
33. Paap MCS, Kroeze KA, Glas CAW, Terwee CB, van der Palen J, Veldkamp BP. Measuring patient-reported outcomes adaptively: multidimensionality matters!. Appl Psychol Meas 2018 Jul;42(5):327-342. [doi: 10.1177/0146621617733954] [Medline: 29962559]

Abbreviations

CAT: computerized adaptive test
EAP: expected a posteriori
IRT: item response theory
PROM: patient-reported outcome measure
RMSE: root mean squared error

Edited by A Mavragani; submitted 04.05.24; peer-reviewed by MAJ Luijten, N Guler, N Frans; revised version received 16.12.24; accepted 17.12.24; published 21.03.25.

Please cite as:
Amagai S, Kaat AJ, Fox RS, Ho EH, Pila S, Kallen MA, Schalet BD, Nowinski CJ, Gershon RC
Customizing Computerized Adaptive Test Stopping Rules for Clinical Settings Using the Negative Affect Subdomain of the NIH Toolbox Emotion Battery: Simulation Study
JMIR Form Res 2025;9:e60215
URL: https://formative.jmir.org/2025/1/e60215
doi:10.2196/60215

© Saki Amagai, Aaron J Kaat, Rina S Fox, Emily H Ho, Sarah Pila, Michael A Kallen, Benjamin D Schalet, Cindy J Nowinski, Richard C Gershon. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

"Your Life, Your Health: Tips and Information for Health and Well-Being": Development of a World Health Organization Digital Resource to Support Universal Access to Trustworthy Health Information

Danielle M Muscat[1], PhD; Rachael Hinton[2], PhD; Shyama Kuruvilla[2], PhD; Don Nutbeam[3], PhD

[1]Sydney Health Literacy Lab, Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[2]World Health Organization, Avenue Appia 20, Geneva, Switzerland

[3]Sydney School of Public Health, Faculty of Medicine and Health, University of Sydney, Sydney, Australia

Corresponding Author:
Shyama Kuruvilla, PhD
World Health Organization, Avenue Appia 20, Geneva, Switzerland

Abstract

Background: Access to trustworthy, understandable, and actionable health information is a key determinant of health and is an essential component of universal health coverage and primary health care. The World Health Organization has developed a new digital resource for the general public to improve health and well-being across different life phases and to support people in caring for themselves, their families, and their communities. The goal was to make trustworthy health information accessible, understandable, and actionable for the general public in a digital format and at the global scale.

Objective: The aim of this paper was to describe the multistage approach and methodology used to develop the resource *Your life, your health: Tips and information for health and well-being* (hereafter, *Your life, your health*).

Methods: A 5-step process was used to develop *Your life, your health*, including (1) reviewing and synthesizing existing World Health Organization technical guidance, member state health and health literacy plans, and international human rights frameworks to identify priority messages; (2) developing messages and graphics that are accessible, understandable, and actionable for the public using health literacy principles; (3) engaging with experts and stakeholders to refine messages and message delivery; (4) presenting priority content in an accessible digital format; and (5) adapting the resource based on feedback and new evidences.

Results: The *Your life, your health* online resource adopts a life-course approach to organize health information based on priority actions and rights that support peoples' health and well-being across different life stages and specific health topics. The resource promotes health literacy by offering advice on asking questions to health workers, making informed decisions about personal and family health, and effectively using digital media to obtain reliable health information. Additionally, it reflects the ambitions of the Sustainable Development Goals by providing essential information on the social determinants of health and clarifies the distinct roles of individuals, frontline workers, governments, and the media in promoting and protecting health.

Conclusions: Making health information available—including to the public—is an essential step in strengthening the global health information system. The development process for the *Your life, your health* online resource outlined in this article offers a structured approach to translate technical health guidelines into accessible, understandable, and actionable health information for the general public.

(JMIR Form Res 2025;9:e57881) doi:10.2196/57881

KEYWORDS

health communication; health literacy; consumer health information; digital health; universal health care

Introduction

Background

Access to reliable, understandable, and actionable health information is a determinant of health and an essential component of universal health coverage (UHC) [1-3]. Similar to other social determinants of health, there is significant variability and inequity in access to health information and in peoples' capacity to retrieve, appraise, and use it in making health-related decisions, also referred to as health literacy [4]. Differences in access to health information and levels of health literacy are observed both within and between countries [5-8].

Digital technology has significantly improved access to health information for many people. Several reviews of digital health

XSL•FO
RenderX

interventions have demonstrated the feasibility and potential effectiveness of eHealth and mHealth (mobile health) interventions for a range of health conditions in low-, middle-, and high-income countries [9-13]. However, the digital divide is real and significantly impacts peoples' access to health information [14] and perpetuates many existing health inequities. Additionally, health information accessed from digital platforms (ie, any electronic communication tool, including websites and social media) is not consistently reliable, understandable, and actionable.

Reliable health information originates from a trusted source and is based on an objective interpretation of available scientific evidence [1]. To be understandable and actionable, members of the public with diverse backgrounds and varying levels of health literacy must be able to understand key messages and identify appropriate actions based on the information presented [15]. However, these fundamental requirements of reliability, understandability, and actionability are not always met. This issue became particularly evident during the COVID-19 pandemic, when unintended misinformation and deliberately circulated disinformation were widely available and often amplified by social media [16-18]. A significant proportion of information sources are also biased, sometimes unconsciously but often deliberately, particularly in commercial communications designed to promote specific products or services [1]. Even when health information originates from a reliable, trustworthy source, it is rarely presented in a manner that most people can understand and act upon when making health-related decisions [19,20].

In response to the global challenge of misinformation and disinformation during the COVID-19 pandemic, the World Health Organization (WHO) developed a Framework for Managing the COVID-19 Infodemic. This framework proposed various actions that WHO member states and the civil society could implement to address the infodemic. This included clear guidance on ensuring that health information was provided in a manner that was "actionable" and "presented in ways that are understood by and accessible to all individuals in all parts of all societies" [21]. This framework recognized the important role of WHO, other United Nations (UN) partners, and civil society organizations in providing reliable health information that is adaptable for use in a wide range of contexts.

Long before this, the 1946 WHO Constitution specified that "informed opinion and active cooperation on the part of the public are of the utmost importance in the improvement of the health of the people" [22]. Despite this historical mandate, WHO has generally limited its health communication activities to providing technical advice and support to member states. These communications are not systematically designed for direct use by the general public [20]. Over the past 25 years, this has been gradually changing and WHO has developed global health information resources that can be accessed and used more directly by the general public, such as WHO's Facts in Pictures [23], Global Health Days [24], and public health advice on COVID-19 [25]. However, at present, much of WHO's publicly available information tends to be organized by different health topics or is limited to targeted communication campaigns or events. Therefore, it is not always easily accessible and is not available directly to the general public in a consistent form or integrated manner.

To address this gap, WHO has developed an integrated public health advice resource, *Your life, your health: Tips and information for health and well-being* [2] (hereafter referred to in short form as *Your life, your health*). This new digital resource, specifically developed for the general public, has been designed to make trustworthy health information accessible, understandable, and actionable, contributing toward achieving UHC. The aim of this paper is to describe the multistage process and methodology used to develop this resource.

Theoretical Framework

The theoretical underpinning for the resource was a UHC and life-course approach to health, based on the WHO Life Course Framework [26]. This framework emphasizes that strategies to achieve UHC must be explicitly people-centered and require integrated, evidence-based approaches across the life course. The WHO's Thirteen General Programme of Work included a dedicated platform on Improving Human Capital across the Life Course, within which this initiative was situated [27].

Additionally, we applied the principles of Universal Design for Learning (UDL) to the development of the resource. UDL is a theoretical paradigm that informs efforts to promote learning for the greatest possible proportion of a population [28]. It is an inclusive learning environment that is universally designed, considers all relevant dimensions of diversity, and proactively minimizes barriers to learning for all individuals [28]. This perspective recognizes that learning occurs in a dynamic interaction between the individual and learning environment. UDL leverages the flexibility of digital technology to design learning environments that, from the outset, offer options for diverse learner needs, ensuring that educational benefits are more equitably and effectively distributed [28]. In the context of health, health literacy universal precautions similarly assume that all patients and caregivers may have difficulty comprehending health information and recommend communicating in ways that ensure universal understanding [29,30].

Methods

Development of the "*Your Life, Your Health*" Resource

A 5-step process was used to develop *Your life, your health* [2]:

- Step 1 – A review and synthesis of existing WHO technical guidance, member state health and health literacy plans, and international human rights frameworks to identify evidence-based health guidance for the public;
- Step 2 – The development of messages and graphics to be accessible, understandable, and actionable for the public through the application of health literacy principles;
- Step 3 – Engagement with experts and other stakeholders to refine messages and message delivery;
- Step 4 – Presentation of priority content in an accessible digital format;
- Step 5 – Adaptation and updates based on feedback and new evidence.

XSL•FO
RenderX

Below, we outline the methods related to the first four steps in detail. Similar to WHO guidelines, the resource will evolve and will be updated as new evidence becomes available. It will also need to be adapted for different country contexts and for different audiences in Step 5.

Ethical Considerations

Ethical approval was not required as this study is based exclusively on published literature, including global and national strategies and plans, with no data collected from human participants.

Review of Evidence-Based Interventions and Guidance

Overview

The first stage of resource development was focused on identifying reliable health information to be included as the basis for *Your life, your health.* There is an overwhelming volume of health information accessible to most people. We sought to identify evidence–based public health actions that individuals or communities could undertake to care for themselves, their families, and their community—either with or without the support of the health system—to improve health and well-being across different life phases. We then prioritized this information by life phase to reduce the volume of information provided in the resource. While making decisions on priorities, we identified interventions or actions within current global and national health strategies that required or presumed an element of individual, caregiver, family, or community initiative or engagement. Interventions or actions that were exclusively health provider-oriented or related to specialized clinical and surgical interventions and care were mostly excluded from the review process.

To identify priority actions, we purposively searched 21 WHO global health strategies across the life course, and additional strategies, guidance, and evidence compendia suggested by WHO partners. This included strategies for specific target populations and life phases (eg, WHO Global Strategy for Women's, Children's and Adolescent's Health), communicable and noncommunicable diseases (eg, WHO Global Action Plan for the Prevention and Control of Noncommunicable Diseases) and cross-cutting strategies on systems, principles, and determinants (eg, Healthy Systems for Universal Health Coverage: A Joint Vision for Healthy Lives). A full list is available in Table S1 in Multimedia Appendix 1.

We also searched the WHO UHC Compendium of Interventions and the Disease Control Priorities 3rd Edition (DCP3) interventions (community worker and health post levels) [31,32]. The UHC Compendium is a database of health services and intersectoral interventions designed to assist countries in making progress towards UHC. The database includes over 3500 health actions across different health areas. The DCP3 provides an up-to-date comprehensive review of priority health interventions and intersectoral policies to reduce mortality and disability which are evidence-based, scalable, and adaptable in multiple settings. Information from the Compendium and DCP3 was cross-checked against the main global causes of mortality and morbidity at each life stage to ensure that the reviewed evidence-based guidance and interventions addressed these challenges.

Finally, we searched for contemporary national and regional health strategies, action plans, and road maps from all 194 WHO member state countries and 6 regions, with a focus on documents available in English and French. These documents were sourced primarily via WHO's Country Planning Cycle Database [33]. If a current document was unavailable in the Planning Cycle Database, a search was carried out on Google or PubMed using the following search string: *[(("National") AND ("Health")) AND (("Strategy") OR ("Road Map") OR ("Action Plan")) AND (Name of Country or Region)]*. We considered "contemporary" to mean any strategic document informing planning in a member state beyond 2019. We found relevant information from 84 countries across the 6 WHO regions, which is shown in Multimedia Appendix 2.

The evidence searches, review, and syntheses were led by SK, supported by RH and Global Health Insights, a specialist global health research and consulting collaborative, with experience in evidence synthesis and policy analysis.

Synthesis of Evidence

Using the documents identified in the evidence review, we listed the interventions and actions that required or presumed an element of individual, caregiver, family, or community initiative or engagement. Using the content analysis approach described by Mays, Pope, and Popay [34], we deductively classified these actions and interventions according to broad life phases adapted from the WHO Life Course Framework [26]: pregnancy, birth, and after childbirth; newborns and children under 5 years; children aged 5-9 years; adolescents and youth aged 10 - 24 years; early and middle adulthood (25 - 64 years); and later adulthood (≥65 years).

Within each life phase, interventions and actions were coded into topic areas derived from the broad health topics on the WHO website [35]. As we reviewed the content, we added further health topics until the saturation of categories was achieved. The final categories of health topics are shown in Textbox 1. Content analysis was then applied to identify recurrent themes arising from the tabulated content within each health topic.

XSL•FO
RenderX

Textbox 1. Final categories of health topics.

- Care and communication in families
- Nutrition
- Physical activity
- Mental health
- Immunization
- Sexual and reproductive health (including family planning)
- Healthy environments
- Social participation
- Violence and injuries
- Smoking, alcohol, and drugs
- Disabilities
- Connecting to services in health and other sectors
- Emergency illness
- Communicable disease
- Noncommunicable disease

Finally, we applied the international human rights framework and a categorization of human rights by the Sustainable Development Goals to develop a matrix of human rights categorized by life phase (eg, rights related to having children) and cross-cutting rights (eg, the right to health). This approach assured that we moved beyond providing only information that required individual action, to ensure that people are informed about their rights, obligations of the government, and personal responsibilities.

Using this methodology, we had reasonable confidence in the quality and consistency of the evidence and were able to use the findings as the basis for prioritization. To keep the volume of information manageable, we aimed to identify 5 essential health messages for each stage in the life course or specific health topic. A health message refers to actionable information related to health and well-being that individuals or communities can use to care for themselves, their families, and their communities (eg, helping children learn to swim). We identified the messages that had the highest potential for promoting health and those most frequently mentioned across the collated information. Inevitably, some issues of importance to individuals, communities, and countries were not included, and people's rights to health may vary between countries. We recognize the importance of these contextual considerations and have included them into our pilot testing and adaptation actions as indicated below.

Application of Health Literacy Principles to Make Content Accessible, Understandable, and Actionable

Following the prioritization of health messages, the next step was to draft the main messages and design the resource's information architecture in a manner that was optimally accessible, understandable, and actionable. We reviewed the national health and health literacy plans of 84 countries with publicly accessible documentation available in English or French to understand how countries developed public-facing messages that were aligned with WHO's evidence-based guidance and global strategies. We made selective use of a range of widely used tools and resources to inform decisions about the presentation of information once content was approved within WHO. These included the Health Literacy Universal Precautions Toolkit [36], the CDC Clear Communication Index, Health Literacy Online [37], and the Patient Education Materials Assessment Tool. All content was drafted by a WHO consultant with expertise in health communication, and technical and copy editing, who was trained in health literacy principles.

Review by Health Specialists and Communication and IT Experts

We developed a web-based prototype of the resource for consultation. Between October 2019 and March 2020, we conducted consultations regarding the draft resource with over 30 experts in health, health literacy, communication, and IT. This included consultation and review across WHO departments, including life course, health workforce, reproductive health and rights, environmental health, health promotion and health literacy, and communications, among others. We sought feedback via an online survey on the content (see Multimedia Appendix 3), particularly on the choice of prioritized actions and rights, and advice on the format and presentation of the messages and potential uses and users. We also sought feedback on various graphics related to the main messages to ascertain the most appropriate format for potential users. Additionally, content experts from WHO departments provided comments on both the simple and more detailed messages relevant to their health topic area to ensure alignment with WHO guidance and existing evidence. Finally, we consulted communication and IT experts within WHO on optimizing the digital platform to support the use and dissemination of the resource, depending on how people use the WHO website. This included the potential for translation of the content into other languages and graphics adaptation for various audiences and communication channels.

XSL•FO
RenderX

Feedback was collated and analyzed thematically by the first author and compiled into an action report for WHO to finalize priority content and graphics. In instances where experts had provided conflicting advice, all verbatim feedback provided was included, and final decisions were made by team consensus with all authors.

Presentation of Priority Content in an Accessible Digital Format

Following the consultation process, the content was revised to incorporate feedback on prioritized actions and rights, while also aiming to maintain a readability grade of 8 or lower (as assessed by the Hemingway Editor). The WHO communications and web team supported the development of the resource for the WHO website.

Results

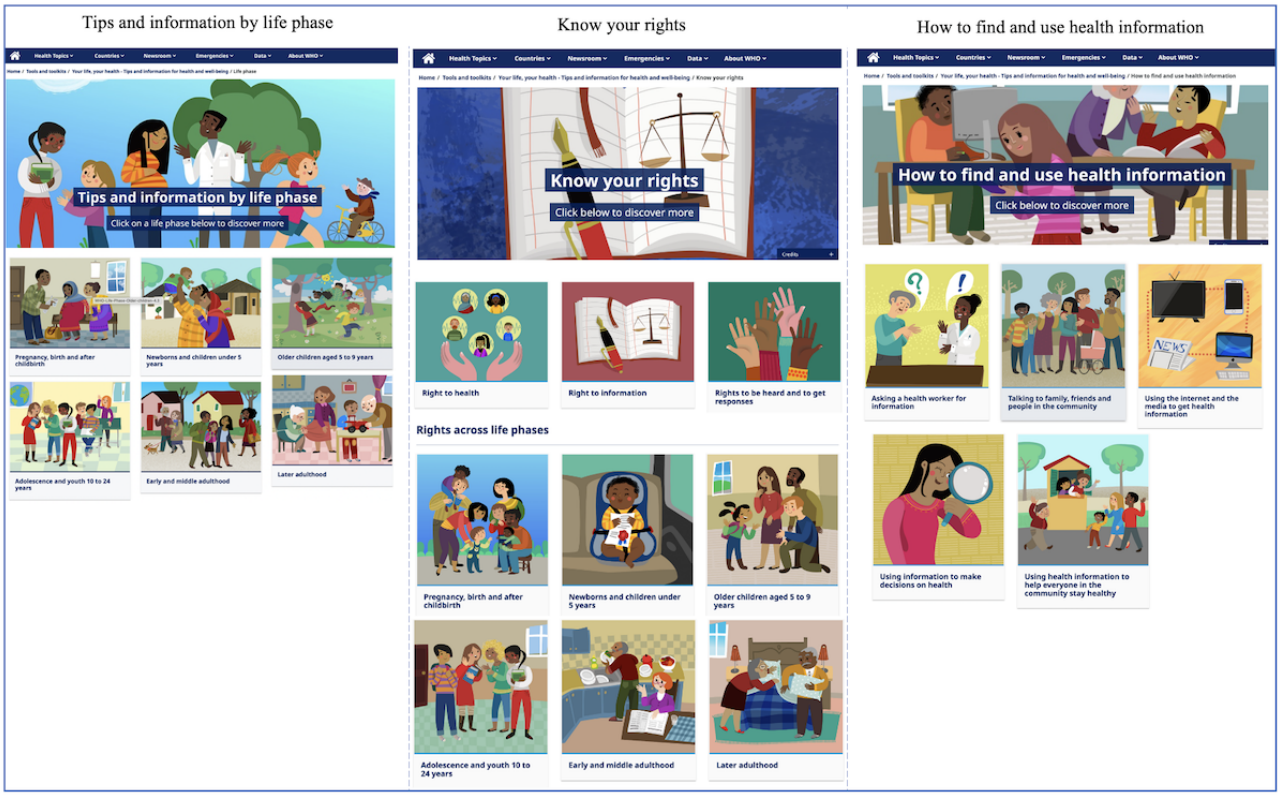
Structure

Through the multistage process outlined above, *Your life, your health*: *Tips and information for health and well-being* online resource was designed to be accessible, understandable, and actionable. This resource has five key sections, accessible via the landing page:

1. Tips and information by life phase
2. How to find and use health information
3. A healthy world
4. Know your rights
5. Other health topics

This resource adopts WHO's life-course approach [26] to organize health-related tips and information (Section 1) and rights (Section 4) based on priority actions and rights that support peoples' health and well-being at different life phases. The consultation process also confirmed that health information should not be focused entirely on individual behaviors and risks, but should reflect how social determinants of health fundamentally influence peoples' choice of actions and their capacity, more accurately to engage successfully in responding to information. For people in most parts of the world, health decisions and actions are substantially restricted by social and environmental factors beyond their control; peoples' ability to respond to health information is not universal and equal. In addition, we recognized that people rarely make health decisions in isolation but rather through interaction and support from other people, institutions, organizations, services, and systems. Given this context, Section 3*: A Healthy World* was integrated to reflect the ambitions of the Sustainable Development Goals by providing information on the social determinants of health and clarifying the roles of individuals, frontline workers, governments, and the media in promoting and protecting health. Section 2: *How to find and use health information* was also included to more overtly support users in developing health literacy skills. It includes advice on asking questions to health workers, making informed decisions about personal and family health, and effectively using digital media to access health information (Figure 1).

Figure 1. Screenshots representing the structure of three sections of the "Your life, your health: Tips and information for health and well-being" online resource.

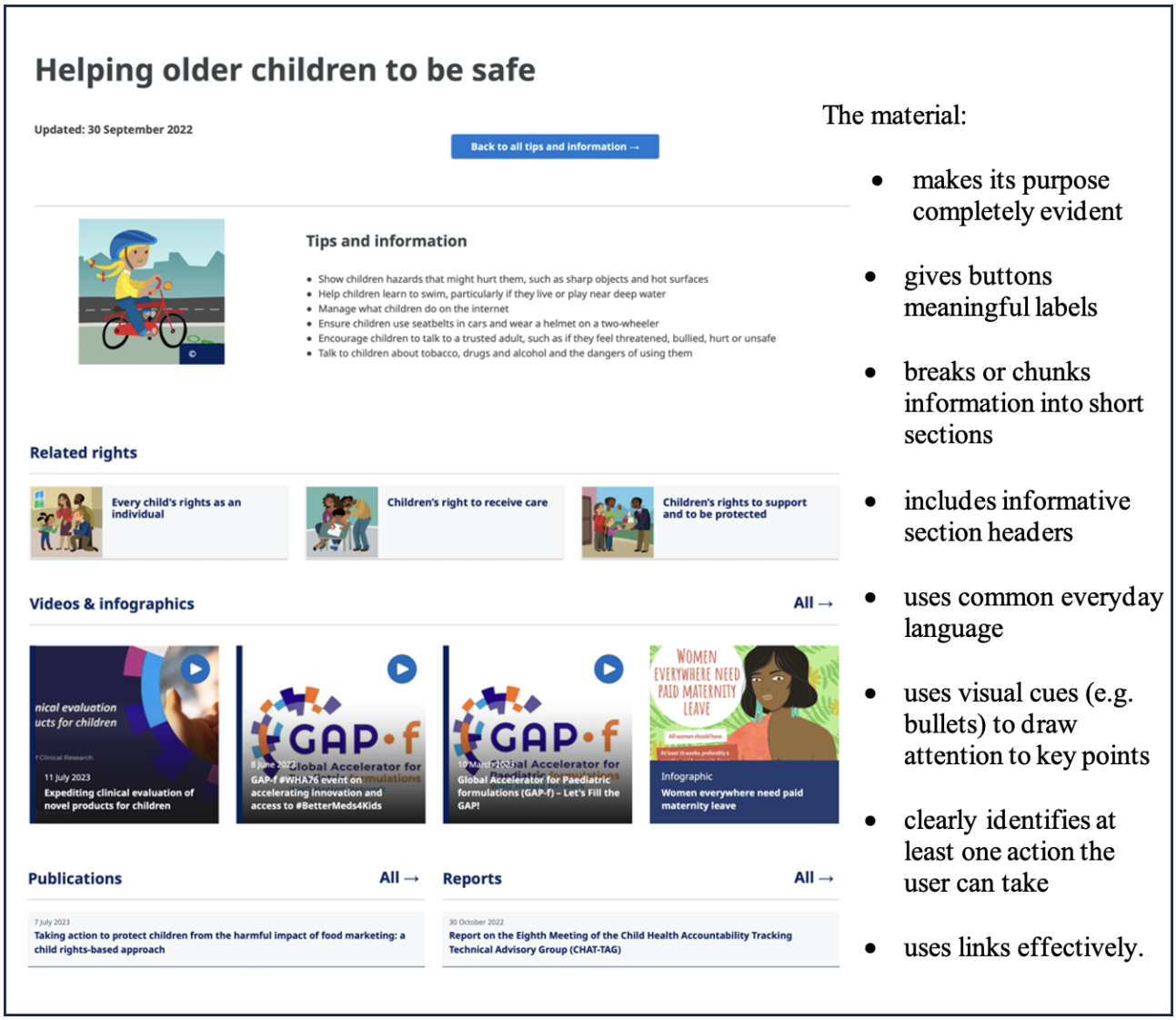


XSL•FO
RenderX

Content and Presentation

Within each section, we adopted a tiered structure, where information was sequenced from simple to more detailed messages at a level of Grade 8 readability or lower. This would allow the readers to move from the first level consisting of 5 main messages to a second level where each message is broken down into manageable and explicit steps. Further information and more detailed WHO technical content and products could then be made available via links to WHO infographics, videos, and fact sheets depending on needs and preferences. To enhance actionability, each message started with a verb and was worded in a direct way (eg, "Learn about..." "Find out…"). Content cues or markers (eg, bullet points) were also used to highlight or draw attention to each point and 3 styles of frequently used visuals for different life phases and messages (ie, photos, illustrations, and infographics) were developed. Figure 2 shows a screenshot of the *Your life, your health* online resource showcasing select design elements.

Figure 2. Screenshot from the "Your life, your health: Tips and information for health and well-being" online resource showcasing select design elements from Health Literacy Online and the Patient Education Materials Assessment Tool (PEMAT).



Hosting and Dissemination

As a WHO global good initiative, *Your life, your health* is publicly available to all. The resource is available on the WHO website and is linked to Health Topics [38], for easy access to public health information. To increase accessibility, health advice within the resource is available as downloadable and shareable digital cards with links to expanded messages and resources (See Multimedia Appendix 4). The planned next steps include engaging with the WHO academy to link the resource content to relevant courses and topics for learners from the global health workforce. Other adaptations include integrating *Your life, your health* in home-based records, such as integrated maternal and child health books, which often include health education messages, as well as using content on the WHO Facebook (chatbot) page [39]. Similar to WHO guidelines, the resource will evolve and be updated as new evidence becomes available. Given the synthesis approach adopted in the development of this resource, this will be aligned with updates from the WHO UHC Compendium of Interventions and the Disease Control Priorities. Standard practice steps and a template have been developed to ensure the health literacy principles described here (eg, tiered presentation of information, testing readability) and will continue to provide a guiding framework

XSL•FO
RenderX

as new information is integrated. A guidance document has also been developed for WHO departments for developing new content in *Your life, your health* in a standardized format for the WHO website.

Discussion

As a contribution to improving universal access to trustworthy health information, the WHO has developed an integrated resource of health advice for use by and with the general public. This resource is based on a synthesis of evidence-based guidance derived from the WHO and national-level documents. It is organized to reflect priority actions and rights across different life phases while concurrently seeking to build health literacy skills, and acknowledge and promote awareness of the social determinants of health. As a global resource, *Your life, your health* is intended to complement country-based initiatives such as the US government's MyHealthFinder or HealthDirect in Australia, which are also evidence-based, easy to read, and action oriented.

Making health information available—including to the public—is an essential step in the global health information system [1]. While developing consumer friendly health information is not a new concept [40,41], there is limited published guidance on how public-facing health information resources are developed, or how to achieve this on a global scale. The development process for *Your life, your health* as outlined in this article, includes reviews of technical guidance and health strategies, the application of health literacy principles, and an expert review and iterative consultation process. It offers a structure that can be adapted for related purposes within and beyond the WHO to translate technical health guidelines into accessible, understandable, and actionable health information for the general public. The development process was built on several well-established health literacy guidelines related to structuring, writing, and visually presenting health information. In addition to the resources referenced above (ie, Health Literacy Universal Precautions Toolkit [32], CDC Clear Communication Index, Health Literacy Online [33]; Patient Education Materials Assessment Tool), there are several other resources more broadly related to universal design. Future research and practice should focus on the selection of resources that are validated and accurate, given inconsistencies across some automated health literacy tools [42]. The selection of guiding resources will also necessarily be context-specific, as mandated guidelines exist across country contexts (eg, US Federal Plain Language Guidelines and Web Content Accessibility Guidelines).

There are strengths and limitations to the development process of WHO's *Your life, your health* resource. We conducted an evidence synthesis to identify health content, recommendations, and human rights across the life course, including both WHO and external sources. Given the breadth of the resource and scope of health topics included in *Your life, your health*, we reviewed existing global strategies and action plans that are publicly accessible and that already synthesize evidence-based recommendations. By ensuring transparency in the sourcing of evidence and identifying core, consistently advocated health information, we aimed to ensure that the information provided is recognizable, reliable, and trustworthy for those using the resource. This synthesis approach was more efficient than a review of primary studies and allowed for the identification of evidence and recommendations that were consistent across strategies and action plans. While we recognize that this methodology will result in the exclusion of some types of evidence—including newer or emerging evidence which is yet to be incorporated into global strategies—we determined that this approach offered the most secure method for ensuring trust in the information provided, and was supported by expert consultation and review. Additionally, while the exclusion criteria for language were a limitation in our evidence review, the review covered all WHO regions, many of which had national language and English versions available for their strategies.

Our expert review of the content and structure purposefully included diverse expertise, ranging from health literacy to IT; however, it has not yet included community members. Following the publication of this resource, we have initiated a consultative process to iteratively refine its content. This will be followed up by additional co-development and user testing, including with WHO representatives at the regional and country levels, UN agencies, and the general public. Partnering with the general public will result in incremental improvements to the understandability of health information materials [39,43]; the benefits of engaging stakeholders are well-documented and will likely include enhanced relevance, quality, and authenticity, the generation of alternative and innovative ideas and enhanced sustainability of the resource [44]. Adaptation for country-specific contexts will ensure that the resource is culturally acceptable, aligned with local health practices availability of health resources and services, and actionable in a variety of contexts [1]. User testing must include people with variations on literacy, health literacy, and cognitive and functional capacities. The consultations will need to consider feasibility, potential effectiveness, and necessary adaptations of this online digital resource in different settings. It will be important to balance digital engagement with appropriately resourced in-person engagement to ensure that vulnerable groups and those without access to digital channels are not excluded. The effectiveness of the multistage process outlined in this paper will be determined by future user testing to confirm that it can deliver clear and actionable information to those in need.

The current rising momentum and expansion of digital technologies in the UHC era offers a significant opportunity to provide more equitable access to information globally. The *Your life, your health* resource is the first of its kind to be developed by WHO. It aims to consolidate technical and often disparate health information available to the general public in an integrated, understandable, and actionable way. The next step will be to focus on user testing of the resource prior to implementation at the global, regional, and country levels.

XSL•FO
RenderX

Disclaimer

The named authors alone are responsible for the views expressed in this publication that does not necessarily represent the decisions or the policies of the World Health Organization.

Data Availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during this study.

Authors' Contributions

Conceptualization: DMM, DN, RH, SK
Data curation: RH
Formal analysis: DMM, RH, SK
Funding acquisition: SK
Investigation: RH, SK
Methodology: DN, DMM, RH, SK
Project administration: RH
Resources: SK
Writing – original draft: DMM
Writing – review & editing: DMM, DN, RH, SK

Conflicts of Interest

DMM is a codirector of a health literacy consultancy company, Health Literacy Solutions PTY Ltd. DMM takes no personal income from Health Literacy Solutions PTY Ltd.

Multimedia Appendix 1
Global strategies, action plans, and guidance included in the evidence review.
[DOCX File, 14 KB - formative_v9i1e57881_app1.docx]

Multimedia Appendix 2
Countries for which national health strategies were included in the evidence review.
[DOCX File, 338 KB - formative_v9i1e57881_app2.docx]

Multimedia Appendix 3
“Actions and rights for healthy lives” consultation feedback form.
[DOCX File, 23 KB - formative_v9i1e57881_app3.docx]

Multimedia Appendix 4
Digital cards adapted from the “Your life, your health: Tips and information for health and well-being” online resource.
[DOCX File, 226 KB - formative_v9i1e57881_app4.docx]

References

1. Royston G, Pakenham-Walsh N, Zielinski C. Universal access to essential health information: accelerating progress towards universal health coverage and other SDG health targets. BMJ Glob Health 2020 May;5(5):e002475. [doi: 10.1136/bmjgh-2020-002475] [Medline: 32424012]
2. Muscat DM, Hinton R, Nutbeam D, Kenney E, Kuruvilla S, Jakab Z. Universal health information is essential for universal health coverage. Fam Med Com Health 2023 May;11(2):e002090. [doi: 10.1136/fmch-2022-002090]
3. National action plan to improve health literacy. Office of Disease Prevention and Health Promotion. 2010. URL: https://odphp.health.gov/our-work/national-health-initiatives/health-literacy/national-action-plan-improve-health-literacy [accessed 2023-09-07]
4. Nutbeam D, Muscat DM. Health Promotion Glossary 2021. Health Promot Int 2021 Dec 23;36(6):1578-1598. [doi: 10.1093/heapro/daaa157] [Medline: 33822939]
5. Aronson B. Improving online access to medical information for low-income countries. N Engl J Med 2004 Mar 4;350(10):966-968. [doi: 10.1056/NEJMp048009] [Medline: 14999107]
6. Chen X, Orom H, Hay JL, et al. Differences in rural and urban health information access and use. J Rural Health 2019 Jun;35(3):405-417. [doi: 10.1111/jrh.12335] [Medline: 30444935]
7. Pakenham-Walsh N, Bukachi F. Information needs of health care workers in developing countries: a literature review with a focus on Africa. Hum Resour Health 2009 Apr 8;7(1):30. [doi: 10.1186/1478-4491-7-30] [Medline: 19356239]

XSL•FO
RenderX

8. Sørensen K, Pelikan JM, Röthlin F, et al. Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). Eur J Public Health 2015 Dec;25(6):1053-1058. [doi: 10.1093/eurpub/ckv043] [Medline: 25843827]
9. Hall CS, Fottrell E, Wilkinson S, Byass P. Assessing the impact of mHealth interventions in low- and middle-income countries--what has been shown to work? Glob Health Action 2014;7(1):25606-25606. [doi: 10.3402/gha.v7.25606] [Medline: 25361730]
10. Lee SH, Nurmatov UB, Nwaru BI, Mukherjee M, Grant L, Pagliari C. Effectiveness of mHealth interventions for maternal, newborn and child health in low- and middle-income countries: systematic review and meta-analysis. J Glob Health 2016 Jun;6(1):010401. [doi: 10.7189/jogh.06.010401] [Medline: 26649177]
11. Sondaal SFV, Browne JL, Amoakoh-Coleman M, et al. Assessing the effect of mHealth interventions in improving maternal and neonatal care in low- and middle-income countries: a systematic review. PLoS One 2016;11(5):e0154664. [doi: 10.1371/journal.pone.0154664] [Medline: 27144393]
12. Stephani V, Opoku D, Quentin W. A systematic review of randomized controlled trials of mHealth interventions against non-communicable diseases in developing countries. BMC Public Health 2016 Jul 15;16(1):572. [doi: 10.1186/s12889-016-3226-3] [Medline: 27417513]
13. Müller AM, Alley S, Schoeppe S, Vandelanotte C. The effectiveness of e-& mHealth interventions to promote physical activity and healthy diets in developing countries: a systematic review. Int J Behav Nutr Phys Act 2016 Oct 10;13(1):109. [doi: 10.1186/s12966-016-0434-2] [Medline: 27724911]
14. Reddy H, Joshi S, Joshi A, Wagh V. A critical review of global digital divide and the role of technology in healthcare. Cureus 2022 Sep;14(9):e29739. [doi: 10.7759/cureus.29739] [Medline: 36340549]
15. Shoemaker SJ, Wolf MS, Brach C. Development of the Patient Education Materials Assessment Tool (PEMAT): a new measure of understandability and actionability for print and audiovisual patient information. Patient Educ Couns 2014 Sep;96(3):395-403. [doi: 10.1016/j.pec.2014.05.027] [Medline: 24973195]
16. Muhammed T S, Mathew SK. The disaster of misinformation: a review of research in social media. Int J Data Sci Anal 2022 May;13(4):271-285. [doi: 10.1007/s41060-022-00311-6]
17. Borges do Nascimento IJ, Pizarro AB, Almeida JM, et al. Infodemics and health misinformation: a systematic review of reviews. Bull World Health Organ 2022 Sep 1;100(9):544-561. [doi: 10.2471/BLT.21.287654] [Medline: 36062247]
18. Bin Naeem S, Kamel Boulos MN. COVID-19 misinformation online and health literacy: a brief overview. Int J Environ Res Public Health 2021 Jul 30;18(15):8091. [doi: 10.3390/ijerph18158091] [Medline: 34360384]
19. Mishra V, Dexter JP. Comparison of readability of official public health information about COVID-19 on websites of international agencies and the governments of 15 countries. JAMA Netw Open 2020 Aug 3;3(8):e2018033. [doi: 10.1001/jamanetworkopen.2020.18033] [Medline: 32809028]
20. Mac OA, Muscat DM, Ayre J, Patel P, McCaffery KJ. The readability of official public health information on COVID-19. Med J Aust 2021 Oct 18;215(8):373-375. [doi: 10.5694/mja2.51282] [Medline: 34580878]
21. Tangcharoensathien V, Calleja N, Nguyen T, et al. Framework for managing the COVID-19 infodemic: methods and results of an online, crowdsourced WHO technical consultation. J Med Internet Res 2020 Jun 26;22(6):e19659. [doi: 10.2196/19659] [Medline: 32558655]
22. Constitution of the World Health Organization. : World Health Organization; 1946 URL: https://www.who.int/about/governance/constitution [accessed 2025-02-25]
23. Facts in Pictures 2023. World Health Organization. URL: https://www.who.int/news-room/facts-in-pictures [accessed 2023-09-07]
24. Global health days 2023. World Health Organization. URL: https://www.who.int/campaigns [accessed 2023-09-07]
25. Advice for the public: coronavirus disease (COVID-19). World Health Organization. 2023. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public [accessed 2023-09-07]
26. Kuruvilla S, Sadana R, Montesinos EV, et al. A life-course approach to health: synergy with sustainable development goals. Bull World Health Organ 2018 Jan 1;96(1):42-50. [doi: 10.2471/BLT.17.198358] [Medline: 29403099]
27. Thirteenth general programme of work 2019-2023. World Health Organization. 2019. URL: https://www.who.int/about/general-programme-of-work/thirteenth [accessed 2024-10-08]
28. Meyer A, Rose DH, Gordon D. Universal Design for Learning: Theory and Practice: CAST Professional Publishing; 2014. URL: https://publishing.cast.org/catalog/books-products/universal-design-for-learning-meyer-rose-gordon [accessed 2025-02-25]
29. DeWalt DA, Broucksou KA, Hawk V, et al. Developing and testing the health literacy universal precautions toolkit. Nurs Outlook 2011;59(2):85-94. [doi: 10.1016/j.outlook.2010.12.002] [Medline: 21402204]
30. Mabachi NM, Cifuentes M, Barnard J, et al. Demonstration of the health literacy universal precautions toolkit: lessons for quality improvement. J Ambul Care Manage 2016;39(3):199-208. [doi: 10.1097/JAC.0000000000000102] [Medline: 27232681]
31. UHC compendium: health interventions for universal health coverage. World Health Organization. 2023. URL: https://www.who.int/universal-health-coverage/compendium [accessed 2023-07-06]
32. DCP3 Disease Control Priorities. University of Washington Department of Global Health. 2021. URL: https://dcp-3.org [accessed 2023-07-06]

XSL•FO
RenderX

33. Country planning cycles database. World Health Organization. 2023. URL: https://extranet.who.int/countryplanningcycles [accessed 2023-07-06]
34. Mays N, Pope C, Popay J. Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. J Health Serv Res Policy 2005 Jul;10 Suppl 1(1_suppl):6-20. [doi: 10.1258/1355819054308576] [Medline: 16053580]
35. Health Topics. World Health Organization. 2024. URL: https://www.who.int/health-topics [accessed 2024-02-27]
36. AHRQ health literacy universal precautions toolkit. Agency for Healthcare Research and Quality. 2020. URL: https://www.ahrq.gov/health-literacy/improve/precautions/index.html [accessed 2023-05-17]
37. Health literacy online. Office of Disease Prevention and Health Promotion. 2016. URL: https://health.gov/healthliteracyonline [accessed 2023-07-03]
38. Health topics. World Health Organization. URL: https://www.who.int/health-topics [accessed 2025-02-20]
39. Crawford MJ, Rutter D, Manley C, et al. Systematic review of involving patients in the planning and development of health care. BMJ 2002 Nov 30;325(7375):1263. [doi: 10.1136/bmj.325.7375.1263] [Medline: 12458240]
40. MyHealthfinder. Department of Health and Human Services Office of Disease Prevention and Health Promotion. URL: https://odphp.health.gov/myhealthfinder [accessed 2024-12-11]
41. Overman VP. Healthfinder. Int J Dent Hyg 2008 Nov;6(4):378-378. [doi: 10.1111/j.1601-5037.2008.00341.x] [Medline: 19138191]
42. Mac O, Ayre J, Bell K, McCaffery K, Muscat DM. Comparison of readability scores for written health information across formulas using automated vs manual measures. JAMA Netw Open 2022 Dec 1;5(12):e2246051. [doi: 10.1001/jamanetworkopen.2022.46051] [Medline: 36508219]
43. Mastroianni F, Chen YC, Vellar L, et al. Implementation of an organisation-wide health literacy approach to improve the understandability and actionability of patient information and education materials: a pre-post effectiveness study. Patient Educ Couns 2019 Sep;102(9):1656-1661. [doi: 10.1016/j.pec.2019.03.022] [Medline: 30962076]
44. Baines R, Bradwell H, Edwards K, et al. Meaningful patient and public involvement in digital health innovation, implementation and evaluation: a systematic review. Health Expect 2022 Aug;25(4):1232-1245. [doi: 10.1111/hex.13506] [Medline: 35526274]

Abbreviations

DCP3 : Disease Control Priorities 3rd Edition
mHealth: mobile health
UDL: Universal Design for Learning
UHC: universal health coverage
UN: United Nations
WHO: World Health Organization

Edited by A Mavragani; submitted 28.02.24; peer-reviewed by A Hassan, C Baur, G Berget; revised version received 20.01.25; accepted 22.01.25; published 06.03.25.

Please cite as:
Muscat DM, Hinton R, Kuruvilla S, Nutbeam D
"Your Life, Your Health: Tips and Information for Health and Well-Being": Development of a World Health Organization Digital Resource to Support Universal Access to Trustworthy Health Information
JMIR Form Res 2025;9:e57881
URL: https://formative.jmir.org/2025/1/e57881
doi:10.2196/57881

© Danielle M Muscat, Rachael Hinton, Shyama Kuruvilla, Don Nutbeam. Originally published in JMIR Formative Research (https://formative.jmir.org), 6.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

HUMAn, a Real-Time Evolutive Patient Model for Major Incident Simulation: Development and Validation Study

Maxence Laurent[1], Dipl.Ing.; Arnaud Jaccard[2], MSc; Laurent Suppan[3], MD; Elio Erriquez[3]; Xavier Good[1], MSc; Eric Golay[3], MAS; Dominique Jaccard[1], MSc; Mélanie Suppan[4], MD, MSc

[1]Media Engineering Institute, School of Management and Engineering Vaud, University of Applied Sciences and Arts of Western Switzerland, Avenue des Sports 20, Yverdon, Switzerland

[2]Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland

[3]Division of Emergency Medicine, Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

[4]Division of Anesthesiology, Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

Corresponding Author:
Dominique Jaccard, MSc
Media Engineering Institute, School of Management and Engineering Vaud, University of Applied Sciences and Arts of Western Switzerland, Avenue des Sports 20, Yverdon, Switzerland

Abstract

Background: Major incidents correspond to any situation where the location, number, severity, or type of casualties requires extraordinary resources. Major incident management must be efficient to save as many lives as possible. As any paramedic or emergency medical technician may unexpectedly have to respond to major incidents, regular training is mandatory. Those trainings usually include simulations. The vast majority of major incident simulations are limited by the fact that simulated patients do not evolve during the simulation, regardless of the time elapsed and treatment decisions. Therefore, most simulations fail to incorporate the critical temporal effect of decision-making.

Objective: This study aimed to develop and validate a simplified mathematical model of physiology, capable of plausibly simulating the real-time evolution of several injuries.

Methods: A modified version of the user-centered design framework, including a relevance, development, and validation phase, was used to define the development process of the physiological model. A 12-member design and development team was established, including prehospital physicians, paramedics, and computer scientists. To determine whether the developed model was clinically realistic, 15 experienced professionals working in the prehospital field participated in the validation phase. They were asked to rate clinical and physiological parameters according to a 5-point Likert scale ranging from 1 (impossible) to 5 (absolutely realistic).

Results: The design and development team led to the development of the HUMAn model (Human is an Uncomplicated Model of Anatomy). During the relevance phase, the team defined the needed features of the model: clinically realistic, able to compute the evolution of prehospital vital signs, yet simple enough to allow real-time computation for several simulated patients on regular computers or tablets. During the development phase, iterations led to the development of a heart-lung-brain interaction model coupled to functional blocks representing the main anatomical body parts. During the validation phase, the evolution of nine simulated patients presenting pathologies devised to test the different systems and their interactions was assessed. Overall, clinical parameters of all patients had a median rating of 5 (absolutely realistic; IQR 4-5). Most (n=52, 96%) individual clinical parameters had a median rating of 5, the remainder (n=2, 4%) being rated 4. Overall physiological parameters of all patients had a median rating of 5 (absolutely realistic; IQR 3-5). The majority of individual physiological parameters (n=43, 79%) had a median rating of 5, with (n=9, 17%) rated 4, and only (n=2 ,4%) rated 3.

Conclusions: A simplified model of trauma patient evolution was successfully created and deemed clinically realistic by experienced clinicians. This model should now be included in computer-based simulations and its impact on the teaching of major incident management assessed through randomized trials.

(JMIR Form Res 2025;9:e66201) doi:10.2196/66201

KEYWORDS

physiological model; mathematical model; computer simulation; major incident management; emergency medicine; mass casualties; healthcare professional education; professional education; continuing education

XSL•FO
RenderX

Introduction

Background and Importance

In the health sector, major incidents correspond to "any incidents where the location, number, severity or type of lives of casualties requires extraordinary resources" [1]. Major incidents are rare events that require efficient and effective management to save as many lives as possible [1,2]. The scarcity of these events and the impact they can have both on the victims and on society, as the impact such stressful situations may have on rescuers clinical performance [3] has led to the development of many different training modalities [4]. Thus, most prehospital providers are trained to manage major incidents through tabletop or computer-based simulations, practical exercises without casualties, or full-scale simulations [5].

These simulations all present limitations, some of which can often be addressed by using other forms of training. However, 1 important limitation is common to almost all major incident simulations and resides in the fact that simulated patients do not evolve over time [6]. Indeed, in most simulations, victims do not evolve at all regardless of the time elapsed and despite treatment and transport decisions. For instance, a simulated victim experiencing a massive internal hemorrhage (such as splenic rupture) will present identical vital signs throughout the simulation [7]. Consequently, most current simulations fail to incorporate the temporal effects of decision-making, which is a major characteristic of major incident management [8].

A simplified mathematical model of human physiology could be a solution for simulating victims' evolution over time [9]. However, most currently available mathematical models of human physiology present limitations preventing their use in real-time major incident simulations. In 2005, Sacco et al [10] proposed a model using respiratory rate, pulse rate, and motor response (RPM) to predict victims' survival probabilities. In this model, a single discrete modification of RPM parameters is made depending on whether the victims remains at the incident scene or is transported to a hospital. This change of RPM parameters is based on expert consensus. While insightful for predicting survival probabilities, this model does not enable to calculate the continuous time-based evolution of multiple physiological and clinical parameters, based on treatment decisions. Many others models are complex and ask for high processing power, making them incompatible with their use to simulate the synchronous evolution of several victims[11]. Furthermore, most models are not open source and cannot be embedded into third-party applications [12].

Objectives

Our objective was to develop and validate a simplified physiological mathematical model capable of simulating the evolution of several casualties, in real time, while remaining realistic in the context of major events training.

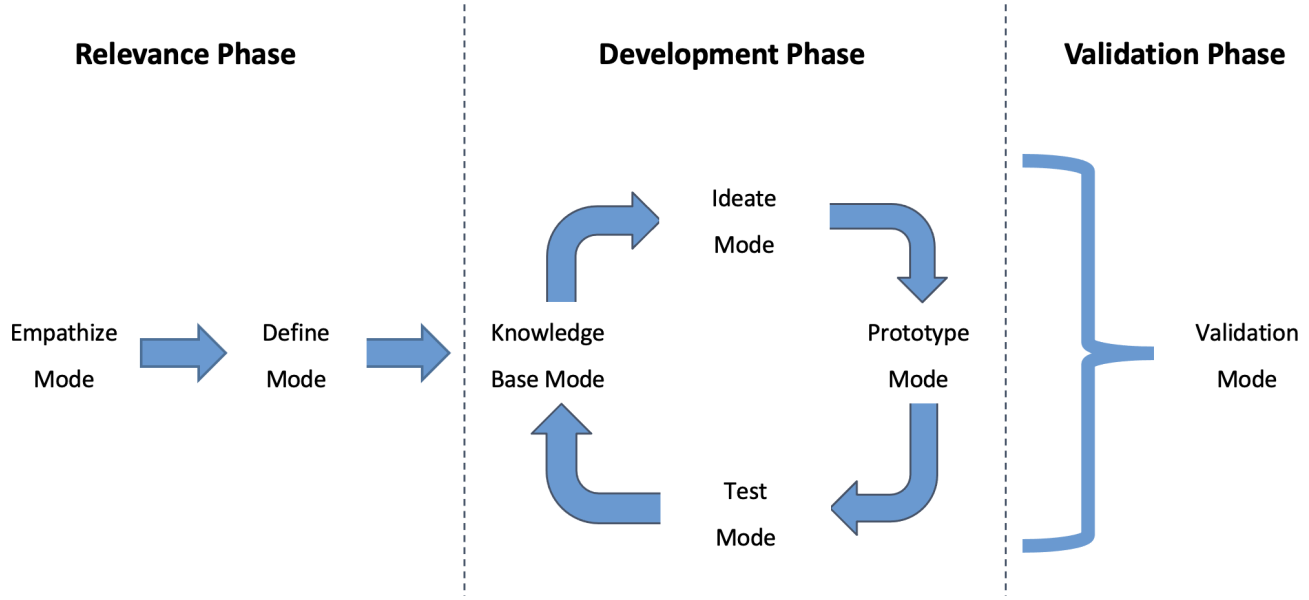
Methods

General Design

A modified version of the user-centered design (UCD) framework proposed by Farao et al [13] was used to define the development process of the physiological model. The UCD framework is based on the 5 modes of design thinking (empathize, define, ideate, prototype, and test) [14], with the addition of the knowledge base mode of the information systems research framework [15].

These 6 modes were sorted in 2 phases, the relevance phase and an iterative development phase. To validate the developed model, a validation phase was added. The overall resulting framework is summarized in Figure 1.

Figure 1. Framework used for the development and validation of the physiological model. The framework integrates the 6 modes of the user-centred design framework with the addition of a validation phase.



XSL•FO
RenderX

Relevance Phase

The objective of the relevance phase was to identify the problem to be solved and define the overall goal for the development phase.

Empathize Mode

The goal of the empathize mode was to identify and understand the end-users and the context of use of the physiological model. Relevant elements of the co.LAB framework, which has been shown to have a positive impact on team collaboration when designing serious games [16,17], were used to determine the learning contexts where the physiological model could be used. In line with the philosophy of the co.LAB framework, end-users were included in the design process early on. A 12-member team was established, including 3 prehospital senior physicians and 3 advanced paramedics (all with extensive emergency medicine experience), 1 medical student, 1 game designer, 3 computer scientists, and 1 instructional designer. A total of 8 team members are coauthors of this article.

Define Mode

Based on results of the empathize mode, the define mode aimed at identifying the needed characteristics of the physiological model as well as basic technical requirements. This was achieved through several collaborative sessions between medical doctors and computer scientists.

Development Phase

The objective of the development phase was to develop a conceptual physiological model, implement it into a software application, and test its ability to simulate the evolution of victims during a major incident.

Knowledge Base Mode

The knowledge base mode aimed at identifying existing solutions and reliable knowledge foundation for the development of the physiological model. First, Google Scholar and the regular Google search engine were used to identify already existing open-source computational models of physiology. Available models were assessed for their relevance against our objectives. Since no such model was found, a team of medical doctors and computer scientists collaborated to identify physiological and physiopathological formulas necessary to develop a model corresponding to the characteristics detailed through the define mode. Searches were done in research databases, textbooks and grey literature. After the first iteration, further searches were based on the results of the test mode and relevant formulas inserted accordingly.

Ideate Mode

The ideate mode was used to define the combination of formulas needed to develop a simple yet clinically plausible physiological model. It also included ideation for technological solutions for the development of the software simulation. Furthermore, this mode was used to investigate and define the algorithmic implementation of formulas in the software. This was carried out through brainswarming sessions involving medical doctors, computer scientists and paramedics [18].

Prototype Mode

Basic implementations of formulas were first done on spreadsheets and then implemented into a full software application. Computational models of independent physiological systems (respiratory, cardiovascular, and brain) were first developed. Once deemed realistic enough by the development team, interactions between these systems were implemented. After ensuring that interactions were appropriate under normal physiological conditions, elements required to simulate injuries were progressively added and tested.

Test Mode

Each prototype was tested by a panel of doctors and paramedics, all belonging to the development team. Tests were first conducted to assess the plausibility of the simulated independent physiological systems and, then, the simultaneous behavior of those systems. Situations involving single, then multiple injuries were then tested. Tests consisted of an evaluation of the plausibility of the evolution of clinical and physiological parameters.

Validation Phase

During the validation phase, relevant elements of Eysenbach's Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [19] have been used to report the results.

Design and Study Sequence

To determine whether the developed model was clinically realistic, 15 experienced prehospital professionals (5 advanced paramedics and 10 senior registrars working in ground and helicopter emergency medical services, none of them being part of the development team) were invited to rate clinical and physiological parameters according to a 5-point Likert scale ranging from 1 (impossible) to 5 (absolutely realistic).

After a short briefing regarding the study's aims, these experts were asked to create a personal account on an web testing platform. The top of the page displayed the patient's characteristics (sex, age, height, and weight), their pathology, and their time of death if the model predicted it would happen within 4 hours if left untreated. All experts were clearly informed that they would have to conceptualize how patient parameters would evolve if they did not receive any medical treatment.

Clinical (oxygen saturation, respiratory rate, heart rate, mean arterial pressure, Glasgow coma scale, and ability to walk) and physiological (alveolar volume, PaO2, PaCO2, blood volume, stroke volume, and intracranial pressure) parameters were displayed at 5 different time points. The first time point (T0) corresponds to the initial state, just before the simulated patient was ascribed a specific pathology. The last point (T4) corresponds to the time of death of the simulated patient, with a maximum T4 equals to 4 hours if the survival time was more than 4 hours. To assess the plausibility of the evolution of the parameters, intermediate values between T0 and T4 were presented at times corresponding to 25% (T1), 50% (T2), and 75% (T3) of T4.

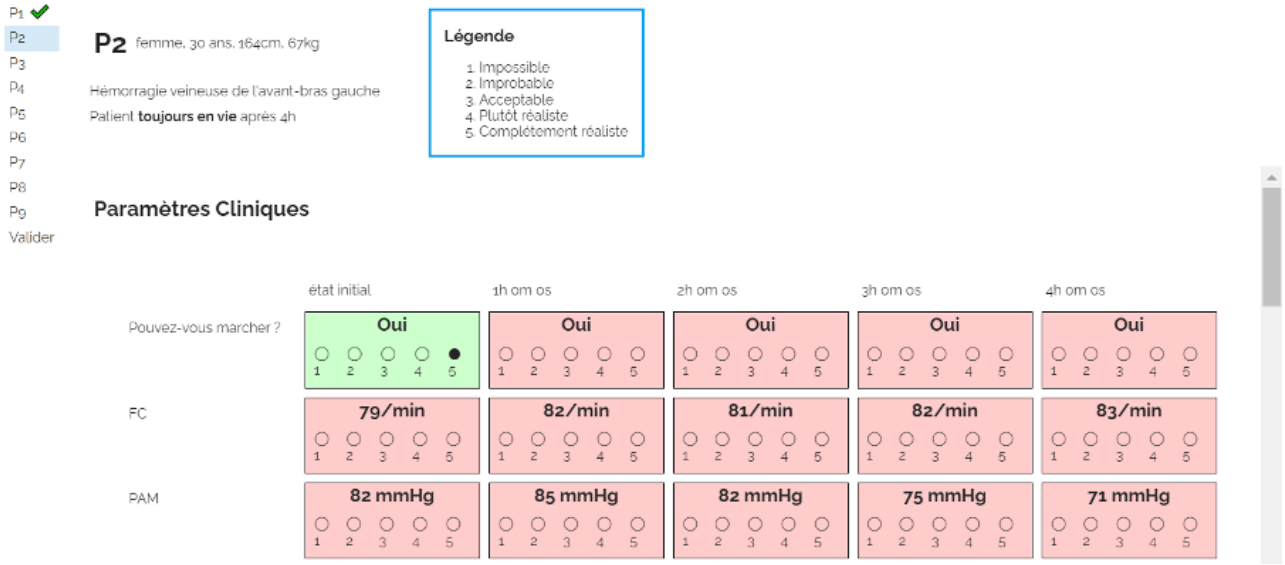
Although the use of relative times for T1 to T4 leads to an assessment of clinical and physiological parameters at different

XSL•FO
RenderX

absolute times for each patient, we chose this approach because it enable us to assess the plausibility of the evolution of the parameters at similar relative times to death. In order to give the complete information to the experts who assessed the model, the absolute values of T1, T2, T3, and T4 were displayed in the form of "hours : minutes : seconds" for each simulated patient.

Figure 2 shows a screenshot of the system used to display parameters and enable participants to rate them. In this particular case, the patient was expected to survive more than 4 hours. Thus the last time point is equal to 4 hours and intermediate time points correspond to 1, 2, and 3 hours.

Figure 2. Screenshot of the web assessment form. A summary of the simulated patients to be evaluated is displayed on the left side. The patient's characteristics and the Likert scale legend are shown on the top of the page. Each parameter is shown on a single row, and each column represents a specific timepoint. Once the expert has chosen a rating, the background turns to green.



Statistical Analysis

Data were extracted to a .csv file and imported in Stata (version 17.0. StataCorp) for curation and statistical analysis. All clinical and physiological parameters were individually tested at each time point to determine their perceived reliability. Their overall perceived reliability was then assessed. Clinical parameters (oxygen saturation, respiratory rate, heart rate, mean arterial pressure, Glasgow coma scale, and ability to walk) were grouped and tested together at each timepoint and overall. The same procedure was applied to physiological parameters (alveolar volume, PaO_2, $PaCO_2$, blood volume, stroke volume, and intracranial pressure). Results are reported as median (IQR) to better reflect the definitions of the Likert scale.

All clinical and physiological parameters were individually tested at each time point to determine their perceived reliability. Their overall perceived reliability was then assessed. Clinical parameters (SpO_2, respiratory rate, heart rate, mean arterial pressure, Glasgow coma scale, ability to walk) were grouped and tested together at each timepoint and overall. The same procedure was applied to physiological parameters (alveolar volume, PaO_2, $PaCO_2$, blood volume, stroke volume, intracranial pressure).

Ethical Considerations

Since the participants who took part in the validation phase were all professionals working in the prehospital field, their participation did not fall within the scope of the Swiss Federal Act on Research involving Human Beings (HRA, RS 810.30). All the participants agreed to take part in this study without any form of compensation, and no incentive other than pure scientific interest was used to promote participation. The results of the evaluation phase were anonymized to ensure the confidentiality of participants' assessments.

All the participants agreed to take part in this study without any form of compensation, and no incentive other than pure scientific interest was used to promote participation. The results of the evaluation phase were anonymized to ensure the confidentiality of participants' assessments.

Results

Relevance Phase

Empathize Mode

Several collaborative design sessions were organized. All collaborative sessions were held in the presence of at least 3 members of the overall team, with the medical and game development professions systematically represented. At the end of each session, ideas were discussed and synthesized.

Define Mode

The team agreed on the characteristics of the physiological model that was to be developed. The most important feature was that the model was to be clinically realistic. This entailed that it should include all the elements required to compute vital signs and their evolution according to the presence of pathologies usually encountered in major incidents, such as hemorrhages, burns, pneumothoraxes, and head traumas. In addition, the model was to be simple enough to allow real-time computation of all relevant vital parameters for several simulated patients simultaneously. It was also decided that the system

XSL•FO
RenderX

should be able to randomly create simulated patients. Taking all these elements into account, it was determined that the model would be limited to the characteristics outlined in Table 1.

Table . Characteristics of simulated patients.

Characteristics	Values
Age (years), range	16-100
Sex (boolean)	Male or female
BMI (kg/m^2), mean (SD)	23 (3)
Height (cm), mean (SD)	
Man	178.4 (7.6)
Woman	164.7 (7.1)

Development Phase

Knowledge Base Analysis

When models were readily available, knowledge base analysis was a straightforward process and all relevant references were rapidly included. Several examples are given in Table 2.

Table . Examples of data and formulae obtained through knowledge base analysis.

Variable [study]	Data and formulae
Ideal weight (kg) [20]	
Man	50+0.9 (kg/cm)*(height [cm]-152.4 [cm])
Woman	45.5+0.9 (kg/cm)*(height [cm]-152.4 [cm])
Blood volume (mL) [21]	70*ideal weight (kg)
Dead space (mL) [22]	2.2 (mL/kg)*ideal weight (kg)
Respiratory quotient [23]	0.84
Cerebral blood flow (L/min) [24]	0.5 (L*min/kg)*1.4 (kg)
Cerebral perfusion pressure (mmHg) [25]	Mean arterial pressure – intracranial pressure

Ideate Mode

Since data regarding patient evolution according to specific injuries were often lacking, the mathematical formulae controlling the evolution of physiological parameters in the presence of such injuries had to be developed through a collaboration between clinicians and computer scientists. These formulae were progressively refined through multiple iterations.

Table 3 shows the steps used to compute systolic and diastolic blood pressures according to volemia, sympathetic nervous system activation, pericardic and pleural pressures, and cardiac contractility. While many parameters are taken into account, several were either simplified or used as constants. Nevertheless, the model enables the inclusion of future updates to simulate currently simplified elements such as vascular resistance.

XSL•FO
RenderX

Table . Steps used to determine diastolic and systolic blood pressure.

Step and parameters	Data and formulae	Comment
Initial data		
Blood volume (mL)	70*weight [21]	—[a]
Systemic vascular resistance (mmHg*min/L)	13 [26]	—
Initial $EDV^{b}_{Initial}$ (mL)	120	—
Effect of blood volume on EDV ($\Delta EDV_{Volemia}$) (mL)		EDV is reduced in the event of blood loss (hypovolemia), increased in the event of hypervolemia. The model allows it to vary between +40ml and −70ml according to this parameter. (model-specific approximations rather than established formulas)
Effect of pericardial pressure on EDV ($\Delta EDV_{PPericardial}$) (mL)		A pericardial effusion causes fluid to accumulate in the pericardium. Consequently, EDV decreases along with the increase in pericardial fluid. (model-specific approximations rather than established formulas)
Effect of pleural pressure on EDV ($\Delta EDV_{PPleural}$) (mL)		This simulates a limitation in heart filling consecutive to an increase in pleural cavity pressure. (model-specific approximations rather than established formulas)
Effect of sympathetic nervous system activation on EDV ($\Delta EDV_{Sympathetic}$) (mL)		This is a very simplified approximation designed to simulate the increase in venous return consecutive to sympathetic activation.(model-specific approximations rather than established formulas)
Final EDV (EDV) (mL)	$EDV = EDV_{Initial} + \Delta EDV_{Volemia} + \Delta EDV_{PPericardial} + \Delta EDV_{PPleural} + \Delta EDV_{Sympathetic}$	—
ESV^{c} (mL)		(model-specific approximations rather than established formulas)
SV^{d} (ml)	SV = EDV – ESV [27]	—
CO^{e} (L/min)	CO = SV * HR * 1000 [27]	—
Left ventricular pressure (mmHg)		(model-specific approximations rather than established formulas)
MAP^{f}_{th} (mmHg)	MAP_{th} = CO * SVR [27]	—
MAP^{g}_{eff} (mmHg)	• If MAP_{th}< VP: MAPeff = MAP_{th} • If $MAP_{th} \geq$ VP: MAPeff = VP	MAP cannot be higher than VP (model-specific approximations rather than established formulas)
Actual CO (CO_{eff}) (L/min)	• If MAP_{th}< VP: CO_{eff} = MAP_{eff} / SVR • Else: CO_{eff} = CO	Adjustment of CO if theoretical MAP cannot be achieved (model-specific approximations rather than established formulas)
Calculation of DBP^{h} and SBP^{i} (mmHg)	DBP = 6/7 * MAP_{eff} SBP = 3/2 * MAP_{eff}	(model-specific approximations rather than established formulas)

[a] —: not applicable.

[b] EDV: end-diastolic volume.

[c] ESV: end-systolic volume.

XSL•FO
RenderX

[d] SV: stroke volume.

[e] CO: cardiac output.

[f] MAP_{th}: theoretical mean arterial pressure.

[g] MAP_{eff}: actual MAP.

[h] DBP: diastolic blood pressure.

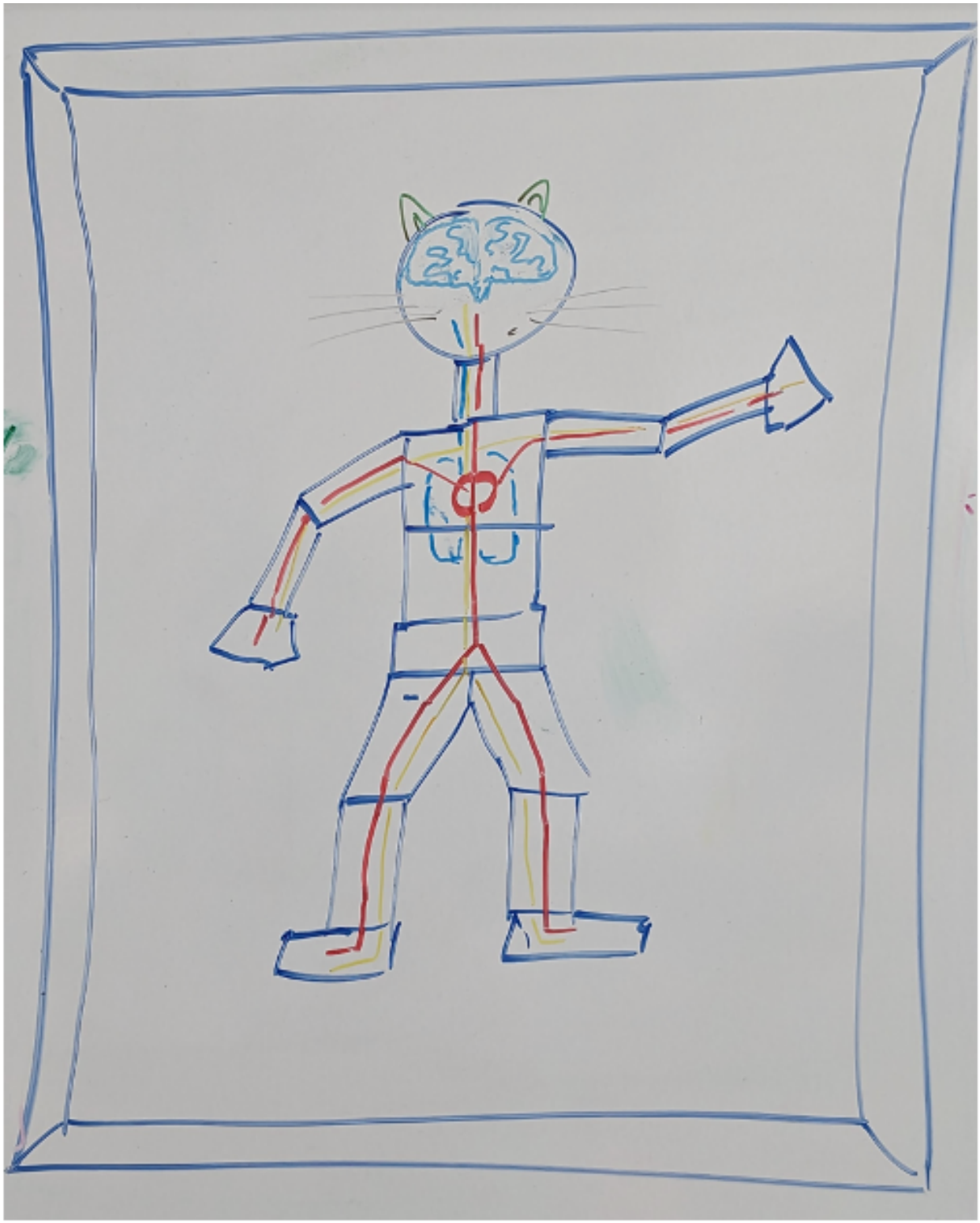
[i] SBP: systolic blood pressure.

Brainswarming processes led to retain two other main concepts for the development of the application: the heart-lung-brain interaction model and the concept of functional blocks. The heart-lung-brain interaction model was decided since these 3 organs are responsible for almost all the vital signs linked to the ABC (airways, breathing, and circulation) evaluation performed by prehospital providers. Although other organs, such as the liver and the kidney, are of course vital, their function does not significantly alter vital signs in the immediate aftermath of a major incident.

The concept of functional blocks (Figure 3) was elaborated to provide more information regarding the neurological status of simulated patients and to create hemorrhages at different levels, with different blood flows, thereby allowing specific treatment options.

XSL•FO
RenderX

Figure 3. Original drawing illustrating the concept of functional blocks. These blocks are used to divide the body into anatomical subsections, allowing traumas to be created at different locations thus enabling localized treatments.



Prototype Mode

Prototypes were developed and tested on Wegas, an open-source web-based serious game authoring and execution platform [28]. Wegas source code can be downloaded on Github [29]. The model source code is freely available under a Creative Commons 4.0 CC-BY-NC licence. The source code can be downloaded on Github [30]. The model's source code is in English. At the implementation level, user interfaces are designed to be multilingual. The current version of the implementation is in French, but other languages may be added in the future.

The model source code is freely available under a Creative Commons 4.0 CC-BY-NC licence. The source code can be downloaded on Github [30]. The model's source code is in English. At the implementation level, user interfaces are designed to be multilingual. The current version of the

XSL•FO
RenderX

implementation is in French, but other languages may be added in the future.

Test Mode

All prototypes were tested by paramedics and physicians belonging to the development team. Their comments were used to progressively refine the model through new iterations of the knowledge base analysis and ideate modes.

Validation Phase

In total, 15 experienced prehospital professionals (5 advanced paramedics and 10 senior registrars working in ground and helicopter emergency medical services) were invited to participate in the validation study. All of them agreed to participate. Their characteristics are detailed in Table 4.

Table . Characteristics of the prehospital experts.

Characteristics	Overall (N=15)	Physicians (n=10)	Paramedics (n=5)
Gender, n (%)			
Man	11 (73)	6 (60)	5 (100)
Woman	4 (27)	4 (40)	0 (0%)
Other	0 (0)	0 (0)	0 (0%)
Age (years), mean (SD)	40 (6)	39 (5)	41 (7)
Postgraduate clinical experience (years), mean (SD)	11 (7)	10 (6)	13 (9)
Prehospital experience (years), mean (SD)	10 (8)	7 (7)	16 (5)

In order to test the different systems and their interactions, 9 simulated patients were automatically generated, with different personal characteristics. We then ascribed a specific pathology to each patient. No patient was ascribed more than 1 pathology. The characteristics of patients and ascribed pathologies are detailed in Table 5.

Table . Characteristics of the patients automatically generated by the simulator.

Patient	Sex (M/F)	Age (years)	Height (cm)	Weight (kg)	Ascribed pathology	Time of death (seconds)
Patient 1	M[a]	45	159	64	Massive internal hemorrhage (eg, ruptured spleen)	2700
Patient 2	F[b]	30	164	67	Veinous hemorrhage of the left forearm	>14,400
Patient 3	M	42	182	86	Traumatic cardiac tamponade	8850
Patient 4	F	63	161	53	Complete left pneumothorax	>14,400
Patient 5	F	46	168	68	Circular second-degree burns of the face and neck	5735
Patient 6	F	50	170	61	Complete circumferential third-degree burn of the chest	5670
Patient 7	F	32	161	62	Head trauma with cerebral hemorrhage	4745
Patient 8	F	61	164	69	Atlanto-occipital dislocation	345
Patient 9	M	44	168	57	Unstable, displaced cervical fracture between C5 and C7	>14,400

[a]M: male.

[b]F: female.

The validation team assessed each patient clinical and physiological parameters from T0 to T4. The initial and final states of all patients had a median rating of 5 (absolutely realistic) for both clinical (Table 6) and physiological (Table

XSL•FO
RenderX

7) parameters. The overall median rating was also 5 for each individual patient. Detailed assessments, by patient and time period, are displayed in Tables 6 and 7. Ratings as mean (SD) are available in Multimedia Appendix 1.

Table . Assessment of the simulated patients' overall clinical parameters.

Patient	T0, median (IQR)	T1, median (IQR)	T2, median (IQR)	T3, median (IQR)	T4, median (IQR)	Overall, median (IQR)
Patient 1	5 (5-5)	5 (4-5)	4 (4-5)	5 (4-5)	5 (5-5)	5 (4-5)
Patient 2	5 (5-5)	5 (5-5)	5 (5-5)	5 (4-5)	5 (4-5)	5 (5-5)
Patient 3	5 (5-5)	5 (4-5)	4 (3-5)	5 (4-5)	5 (5-5)	5 (4-5)
Patient 4	5 (5-5)	5 (4-5)	5 (4-5)	5 (4-5)	5 (4-5)	5 (4-5)
Patient 5	5 (5-5)	5 (4-5)	5 (3-5)	4 (3-5)	5 (5-5)	5 (4-5)
Patient 6	5 (5-5)	5 (4-5)	5 (3-5)	4 (3-5)	5 (5-5)	5 (4-5)
Patient 7	5 (5-5)	5 (4-5)	4 (3-5)	5 (4-5)	5 (5-5)	5 (4-5)
Patient 8	5 (5-5)	5 (3-5)	5 (3-5)	5 (3-5)	5 (5-5)	5 (4-5)
Patient 9	5 (5-5)	5 (4-5)	5 (4-5)	5 (4-5)	5 (4-5)	5 (4-5)

Table . Assessment of the simulated patients' overall physiological parameters.

Patient	T0, median (IQR)	T1, median (IQR)	T2, median (IQR)	T3, median (IQR)	T4, median (IQR)	Overall, median (IQR)
Patient 1	5 (4-5)	5 (4-5)	4 (3-5)	4 (3-5)	5 (4-5)	5 (3-5)
Patient 2	5 (5-5)	5 (5-5)	5 (5-5)	5 (4-5)	5 (4-5)	5 (5-5)
Patient 3	5 (5-5)	5 (4-5)	4 (3-5)	4 (3-5)	5 (5-5)	5 (4-5)
Patient 4	5 (5-5)	5 (4-5)	5 (4-5)	5 (4-5)	5 (4-5)	5 (4-5)
Patient 5	5 (5-5)	5 (4-5)	5 (4-5)	5 (4-5)	5 (5-5)	5 (4-5)
Patient 6	5 (5-5)	5 (4-5)	5 (4-5)	5 (3-5)	5 (5-5)	5 (4-5)
Patient 7	5 (5-5)	5 (5-5)	5 (4-5)	5 (4-5)	5 (5-5)	5 (5-5)
Patient 8	5 (5-5)	5 (4-5)	5 (4-5)	5 (4-5)	5 (5-5)	5 (4-5)
Patient 9	5 (5-5)	5 (3-5)	5 (3-5)	5 (3-5)	5 (3-5)	5 (3-5)

Overall clinical parameters include oxygen saturation, respiratory rate, heart rate, mean arterial pressure, Glasgow coma scale, and ability to walk. T4 corresponds to the time of death of the simulated patient (T4 equals 4 hours if the survival time was more than 4 hours). T1 to T4 correspond respectively to 25%, 50%, and 75% of T4. Results are presented as "median (IQR)."

Overall physiological parameters include alveolar volume, PaO2, PaCO2, blood volume, stroke volume, and intracranial pressure. T4 corresponds to the time of death of the simulated patient (T4 equals 4 hours if the survival time was more than 4 hours). T1 to T4 correspond respectively to 25%, 50%, and 75% of T4. Results are presented as "median (IQR)."

The overall median rating for each specific clinical (Table 8) and physiological (Table 9) parameter was 5 (absolutely realistic). Further results are available in Multimedia Appendix 1.

XSL•FO
RenderX

Table . Assessment of the simulated patients' clinical parameters.

Patient	SpO_2,[a] median (IQR)	RR,[b] median (IQR)	HR,[c] median (IQR)	MAP,[d] median (IQR)	GCS,[e] median (IQR)	Walks,[f] median (IQR)
Patient 1	5 (4-5)	5 (4-5)	5 (3-5)	5 (4-5)	5 (4-5)	5 (4-5)
Patient 2	5 (5-5)	5 (4-5)	5 (4-5)	5 (5-5)	5 (5-5)	5 (5-5)
Patient 3	5 (5-5)	5 (5-5)	5 (4-5)	5 (4-5)	5 (4-5)	5 (4-5)
Patient 4	4 (4-5)	5 (4-5)	5 (4-5)	5 (4-5)	5 (5-5)	5 (4-5)
Patient 5	5 (4-5)	5 (4-5)	5 (3-5)	5 (4-5)	5 (5-5)	5 (4-5)
Patient 6	5 (4-5)	5 (4-5)	5 (4-5)	5 (4-5)	5 (5-5)	5 (3-5)
Patient 7	5 (5-5)	5 (4-5)	5 (4-5)	5 (4-5)	5 (4-5)	5 (4-5)
Patient 8	5 (4-5)	5 (5-5)	5 (4-5)	5 (4-5)	4 (2-5)	5 (5-5)
Patient 9	5 (4-5)	5 (4-5)	5 (4-5)	5 (4-5)	5 (5-5)	5 (5-5)
Overall	5 (4-5)	5 (4-5)	5 (4-5)	5 (4-5)	5 (4-5)	5 (4-5)

[a]SpO_2: oxygen saturation.

[b]RR: respiratory rate.

[c]HR: heart rate.

[d]MAP: mean arterial pressure.

[e]GCS: Glasgow Coma Scale.

[f]Ability to walk.

Table . Assessment of the simulated patients' physiological parameters. Results are presented as "median (IQR)".

	PaO_2,[a] median (IQR)	$PaCO_2$,[b] median (IQR)	Alveolar volume, median (IQR)	Blood volume, median (IQR)	Stroke volume, median (IQR)	ICP,[c] median (IQR)
Patient 1	4 (3-5)	4 (2-5)	4 (3-5)	4 (3-5)	5 (4-5)	5 (5-5)
Patient 2	5 (5-5)	5 (4-5)	5 (5-5)	5 (4-5)	5 (4-5)	5 (5-5)
Patient 3	5 (4-5)	4 (3-5)	5 (4-5)	5 (4-5)	5 (4-5)	5 (5-5)
Patient 4	4 (4-5)	4 (4-5)	4 (4-5)	5 (5-5)	5 (4-5)	5 (5-5)
Patient 5	5 (4-5)	5 (4-5)	5 (4-5)	5 (5-5)	5 (5-5)	5 (5-5)
Patient 6	5 (4-5)	5 (4-5)	5 (4-5)	5 (4-5)	5 (5-5)	5 (5-5)
Patient 7	5 (5-5)	5 (4-5)	5 (5-5)	5 (5-5)	5 (5-5)	5 (4-5)
Patient 8	5 (4-5)	5 (4-5)	5 (5-5)	5 (5-5)	5 (3-5)	5 (5-5)
Patient 9	4 (3-5)	3 (2-5)	3 (3-5)	5 (5-5)	5 (4-5)	5 (5-5)
Overall	5 (4-5)	5 (3-5)	5 (4-5)	5 (4-5)	5 (4-5)	5 (5-5)

[a]PaO_2: arterial partial pressure of O_2.

[b]$PaCO_2$: arterial partial pressure of CO_2.

[c]ICP: intracranial pressure.

Discussion

Principal Findings

A simplified model of trauma patient evolution was successfully created using the modified version of the user-centered design framework and was deemed clinically realistic by experienced prehospital physicians and paramedics. The diversity of simulated clinical cases, which included nervous, cardiovascular and respiratory pathologies, and its suitability for real-time calculations makes this model a good candidate for inclusion in major incident simulations. Many kinds of simulations could be considered, computer-based table-tops and serious games, practical exercises without casualties, and even full-scale simulations. In this latter case, tablets could be used to display the vital parameters of simulated patients.

All clinical and physiological parameters had an overall median rating of 5 (absolutely realistic). Nevertheless, some parameters had lower median ratings when specific patients were analyzed. For instance, the oxygen saturation values of patient 4, who presented a complete left pneumothorax, had a median rating of 4 (rather realistic). Several hypotheses could explain these findings. First, even though an iterative process was followed to progressively refine the evolution of specific parameters, the current model is still an approximation of actual human

XSL•FO
RenderX

physiology. Thus, further iterations may enable further improvements. Second, prehospital clinicians quickly manage airway and breathing issues when responding to critical situations. Therefore, these professionals have never witnessed the evolution of patients left untreated and could find it harder to conceive how vital and physiological parameters would develop in such cases. Finally, given the inherent variability in injury presentations and the subjective nature of textual descriptions, there is potential for variation in interpretation among prehospital providers. This inter-rater variability may contribute to the differing assessments of injury realism.

Other models have previously been developed [31], none of which were however deemed suitable for our purpose, that is, real-time simulation of multiple trauma patients with realistic heart-lung-brain interactions. While models such as HumMod and Biogears provide realistic, scientifically grounded models of patient evolution [32,33], their complexity usually requires considerable computational power, making them hardly suitable for major incident simulations. In addition, and although many physiological processes are supported by clear scientific literature and rules, the lack of data regarding most prehospital injuries had to be compensated for by clinical experience. The inclusion of senior prehospital physicians and paramedics since inception has thus enabled the creation of a model realistic enough to simulate the evolution of trauma patients during a major incident.

The model developed through this study presents the advantage of being based on an anatomically realistic model made of several blocks simulating the different body compartments which could be involved in case of injury. Neural and vascular lesions are thus easy to simulate, and these blocks could be divided up further if needed. Furthermore, other organ systems could be added according to the end-user's requirements: for instance, adding clearance mechanisms could prove useful to the management of major incidents involving hazardous materials. Conversely, disabling or removing systems uninvolved in the simulated situation could save processing power.

Simulating the outcome of patients involved in major incidents could prove useful in several ways. Including an evolutive model in computer-based simulations could help ambulance and medical commanders better understand the impact of certain decisions, such as the use of human resources for treatment rather than for triage. The importance of accurate triage cannot be overstated and methods based on patient evolution should be promoted [9]. In this regard, this physiological model could be used to compare different triage algorithms among different populations. Indeed, some professional categories may be more efficient than others depending on the tool provided [34].

Limitations

This model is not without limitations. The most important and most obvious limitation is the lack of pediatric-specific formulae, which prevents the generation of patients aged 16 years or younger. Another significant limitation is the absence of treatment-specific effects, which would be necessary to assess the effect of human and material resources through computer-based simulations. Since physiological and clinical parameters of simulated patients were assessed at relative intermediate time points (25%, 50%, and 75%) of death time, the absolute times (hours: minutes: seconds) of those intermediate points varied across simulated patients. This may complicate assessing patients' conditions over time. However, to enhance clarity, the absolute value of each intermediate time point was displayed on the interface. In addition, experts were informed about the calculation method for intermediate time points before starting the evaluation of parameters. While the overall scores, both for clinical and physiological parameters, show the high realism of the validation study, overall scores could potentially hide specific inconsistencies. However, Q1 and Q3 values are mostly close to median values. The plausibility of simulated patients' parameters has been assessed by only 15 experts. Given the overall score used in the assessment, the relatively small number of experts who evaluated the model may not have highlighted individual inconsistencies. Finally, the current version of this simplified physiological model was only developed to simulate traumas. Thus, other causes of major incidents, such as those linked to hazardous materials, cannot be simulated without further developments.

Another significant limitation is the absence of treatment-specific effects, which would be necessary to assess the effect of human and material resources through computer-based simulations.

Since physiological and clinical parameters of simulated patients were assessed at relative intermediate time points (25%, 50%, 75%) of death time, the absolute times (hours: minutes: seconds) of those intermediate points varied across simulated patients. This may complicate assessing patients' conditions over time. However, to enhance clarity, the absolute value of each intermediate time point was displayed on the interface. Additionally, experts were informed about the calculation method for intermediate time points before starting the evaluation of parameters.

While the overall scores, both for clinical and physiological parameters, show the high realism of the validation study, overall scores could potentially hide specific inconsistencies. However, Q1 and Q3 values are mostly close to median values.

The plausibility of simulated patients' parameters has been assessed by only 15 experts. Given the overall score used in the assessment, the relatively small number of experts who evaluated the model may not have highlighted individual inconsistencies.

Finally, the current version of this simplified physiological model was only developed to simulate traumas. Thus, other causes of major incidents, such as those linked to hazardous materials, cannot be simulated without further developments.

Perspectives

Embedding this model in computer-based simulations and serious games designed to teach major incident management should now be considered. The inclusion of evolutive pathologies could help determine the most efficient ways of dealing with specific major incident situations and prepare health professionals to deal with the stress of managing such events

XSL•FO
RenderX

[3]. Further developments should also be contemplated. These should help address several limitations, including the unavailability of pediatric patients and the inclusion of treatment specific effects.

Conclusion

A simplified model of trauma patient evolution was successfully created and deemed clinically realistic by experienced clinicians. This model should now be included in computer-based simulations and its impact on the teaching of major incident management assessed through randomized trials.

Acknowledgments

Authors thank Innosuisse who supported this study with the grant 120.336 IP-SBM. This funding sources had no role in the design and execution of this study, nor on data analysis and interpretation. The authors warmly thank all the other members of the development team, Michèle Chan, Simon Regard, Audrey Huguenin, Ely Braun, and Pierre-Benjamin Monaco. We also thank the 15 emergency professionals who took part in the validation phase of the physiological model.

Data Availability

Overall results of the validation phase are available in Multimedia Appendix 1.

Authors' Contributions

ML, MS, and AJ contributed to conceptualization of the physiological model. ML, MS, AJ, LS, and DJ handled methodology. ML and XG managed software development. MS, EG, and LS handled validation. LS conducted formal analysis and data curation. DJ handled the resources. DJ and LS handled the project administration and funding acquisition. ML, MS, DJ, and LS wrote the original draft. All authors contributed to investigation, review and editing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overall results of the validation phase, with ratings as mean (SD).
[DOCX File, 61 KB - formative_v9i1e66201_app1.docx]

References

1. Wieteska S, Mackway-Jones K. Major Incident Medical Management and Support: The Practical Approach to Pre-Hospital Incident Command, 3rd edition: Wiley-Blackwell; 2012.
2. Hugelius K, Edelbring S, Blomberg K. Prehospital major incident management: how do training and real-life situations relate? A qualitative study. BMJ Open 2021 Sep 23;11(9):e048792. [doi: 10.1136/bmjopen-2021-048792] [Medline: 34556511]
3. Duncliffe TH, D'Angelo B, Brock M, et al. Examining the effects of stress on the driving abilities of paramedic students. International Paramedic Practice 2019 Dec 2;9(4):76-83. [doi: 10.12968/ippr.2019.9.4.76]
4. Hampson D. Major incident preparedness: a curriculum and workplace necessity. Scand J Trauma Resusc Emerg Med 2020 Dec 9;28(1):115. [doi: 10.1186/s13049-020-00812-y] [Medline: 33298122]
5. Porthouse A, Clancy H, Lax P. Training for major incidents. Surgery (Oxf) 2021 Jul;39(7):388-392. [doi: 10.1016/j.mpsur.2021.05.007] [Medline: 34092879]
6. Wolf P, Bigalke M, Graf BM, Birkholz T, Dittmar MS. Evaluation of a novel algorithm for primary mass casualty triage by paramedics in a physician manned EMS system: a dummy based trial. Scand J Trauma Resusc Emerg Med 2014 Aug 28;22:50. [doi: 10.1186/s13049-014-0050-6] [Medline: 25214310]
7. Lee CWC, McLeod SL, Van Aarsen K, Klingel M, Franc JM, Peddle MB. First responder accuracy using SALT during mass-casualty incident simulation. Prehosp Disaster Med 2016 Apr;31(2):150-154. [doi: 10.1017/S1049023X16000091] [Medline: 26857296]
8. Hugelius K, Rådestad M, Al-Dhahir H, Kurland L. Decision-making by medical officer in charge during major incidents: a qualitative study. Scand J Trauma Resusc Emerg Med 2021 Aug 21;29(1):120. [doi: 10.1186/s13049-021-00937-8] [Medline: 34419113]
9. Sigle M, Berliner L, Richter E, et al. Development of an anticipatory triage-ranking algorithm using dynamic simulation of the expected time course of patients with trauma: modeling and simulation study. J Med Internet Res 2023 Jun 15;25:e44042. [doi: 10.2196/44042] [Medline: 37318826]

XSL•FO
RenderX

10. Sacco WJ, Navin DM, Fiedler KE, Waddell RK, Long WB, Buckman RF. Precise formulation and evidence - based application of resource - constrained triage. Acad Emerg Med 2005 Aug;12(8):759-770 [FREE Full text] [doi: 10.1197/j.aem.2005.04.003] [Medline: 16079430]
11. Pruett WA, Hester RL. The creation of surrogate models for fast estimation of complex model outcomes. PLoS ONE 2016;11(6):e0156574. [doi: 10.1371/journal.pone.0156574] [Medline: 27258010]
12. Gilbart MK, Hutchison CR, Cusimano MD, Regehr G. A computer-based trauma simulator for teaching trauma management skills. Am J Surg 2000 Mar;179(3):223-228. [doi: 10.1016/s0002-9610(00)00302-0] [Medline: 10827325]
13. Farao J, Malila B, Conrad N, Mutsvangwa T, Rangaka MX, Douglas TS. A user-centred design framework for mHealth. PLoS ONE 2020;15(8):e0237910. [doi: 10.1371/journal.pone.0237910] [Medline: 32813711]
14. Plattner H, Meinel C, Leifer H, Leifer L, editors. Design Thinking Research: Making Distinctions: Collaboration versus Cooperation: Springer; 2018.
15. Hevner AR. A three cycle view of design science research. Association for Information Systems 2007;19(2).
16. Jaccard D, Suppan L, Sanchez E, Huguenin A, Laurent M. The co.LAB generic framework for collaborative design of serious games: development study. JMIR Serious Games 2021 Jul 2;9(3):e28674. [doi: 10.2196/28674] [Medline: 34255655]
17. Jaccard D, Suppan L, Bielser F. Contribution of the co.LAB framework to the collaborative design of serious games: mixed methods validation study. JMIR Serious Games 2021 Nov 24;9(4):e33144. [doi: 10.2196/33144] [Medline: 34822336]
18. BrainSwarming: because brainstorming doesn't work. HBR. URL: https://hbr.org/2014/06/brainswarming-because-brainstorming-doesnt-work [accessed 2024-06-28]
19. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34. [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
20. McCarron MM, Devine BJ. Clinical pharmacy: case studies: case number 25 gentamicin therapy. Drug Intell Clin Pharm 1974 Nov;8(11):650-655. [doi: 10.1177/106002807400801104]
21. Lasocki S, Loupec T, Parot-Schinkel E, et al. Study protocol for a multicentre, 2×2 factorial, randomised, controlled trial evaluating the interest of intravenous iron and tranexamic acid to reduce blood transfusion in hip fracture patients (the HiFIT study). BMJ Open 2021 Jan 17;11(1):e040273. [doi: 10.1136/bmjopen-2020-040273] [Medline: 33455926]
22. West JB, Luks A. West's Respiratory Physiology: The Essentials: Wolters Kluwer; 2016.
23. Singer P. Simple equations for complex physiology: can we use VCO2 for calculating energy expenditure? Crit Care 2016 Mar 21;20(1):72. [doi: 10.1186/s13054-016-1251-3] [Medline: 26997171]
24. Lassen NA. Normal average value of cerebral blood flow in younger adults is 50 ml/100 g/min. J Cereb Blood Flow Metab 1985 Sep;5(3):347-349. [doi: 10.1038/jcbfm.1985.48] [Medline: 4030914]
25. Mount CA, Das JM. Cerebral perfusion pressure. In: StatPearls: StatPearls Publishing; 2024. [Medline: 30725956]
26. Klingensmith ME. The Washington Manual of Surgery: Wolters Kluwer Health; 2016.
27. Klabunde RE. Cardiovascular Physiology Concepts: Wolters Kluwer; 2022.
28. Laurent M, Monnier S, Huguenin A, Monaco PB, Jaccard D. Design principles for serious games authoring tool. IJSG 2022 Nov 14;9(4):63-87. [doi: 10.17083/ijsg.v9i4.458]
29. ALBASIM Research Team, Jaccard D. Wegas. Github. URL: https://github.com/Heigvd/Wegas [accessed 2025-02-04]
30. Laurent M, Jaccard D, Good X, Santaniello F. HUMAn project. Github. 2023. URL: https://github.com/Heigvd/mimms/tree/main/gameModel/libs/ClientScript/HUMAn [accessed 2024-08-19]
31. Borri A, Panunzi S, Brancaleoni R, et al. Simulation of trauma incidents: modelling the evolution of patients and resources. J Med Syst 2016 Nov;40(11):234. [doi: 10.1007/s10916-016-0599-x] [Medline: 27653041]
32. Hester RL, Brown AJ, Husband L, et al. HumMod: a modeling environment for the simulation of integrative human physiology. Front Physiol 2011;2:12. [doi: 10.3389/fphys.2011.00012] [Medline: 21647209]
33. McDaniel M, Baird A. A full-body model of burn pathophysiology and treatment using the biogears engine. 2019 Presented at: 2019 41st Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); Jul 23-27, 2019; Berlin, Germany p. 261-264. [doi: 10.1109/EMBC.2019.8857686]
34. Marcussen CE, Bräuner KB, Alstrøm H, Møller AM. Accuracy of prehospital triage systems for mass casualty incidents in trauma register studies - a systematic review and meta-analysis of diagnostic test accuracy studies. Injury 2022 Aug;53(8):2725-2733. [doi: 10.1016/j.injury.2022.05.006] [Medline: 35660101]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
RPM: respiratory rate, pulse rate, and motor response
UCD: user-centered design

XSL•FO
RenderX

Edited by A Mavragani, D Chartash; submitted 09.09.24; peer-reviewed by NH Mohamad Zainal, Y Bitan; revised version received 13.01.25; accepted 16.01.25; published 07.03.25.

Please cite as:
Laurent M, Jaccard A, Suppan L, Erriquez E, Good X, Golay E, Jaccard D, Suppan M
HUMAn, a Real-Time Evolutive Patient Model for Major Incident Simulation: Development and Validation Study
JMIR Form Res 2025;9:e66201
URL: https://formative.jmir.org/2025/1/e66201
doi:10.2196/66201

© Maxence Laurent, Arnaud Jaccard, Laurent Suppan, Elio Erriquez, Xavier Good, Eric Golay, Dominique Jaccard, Mélanie Suppan. Originally published in JMIR Formative Research (https://formative.jmir.org), 7.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploratory Co-Design on Electronic Health Record Nursing Summaries: Case Study

Suhyun Park[1], PhD, RN; Jenna L Marquard[2], PhD; Robin R Austin[2], PhD, DNP, DC, RN, NI-BC; Christie L Martin[2], PhD, MPH, RN-BC; David S Pieczkiewicz[3], PhD; Connie W Delaney[2], PhD, RN

[1]Cizik School of Nursing, The University of Texas Health Science Center at Houston, #557, 6901 Bertner Ave, Houston, TX, United States

[2]School of Nursing, University of Minnesota, Minneapolis, MN, United States

[3]Institute for Health Informatics, University of Minnesota, Minneapolis, MN, United States

Corresponding Author:
Suhyun Park, PhD, RN
Cizik School of Nursing, The University of Texas Health Science Center at Houston, #557, 6901 Bertner Ave, Houston, TX, United States

Abstract

Background: Although electronic health record nursing summaries aim to provide a concise overview of patient data, they often fall short of meeting nurses' information needs, leading to underutilization. This gap arises from a lack of involvement of nurses in the design of health information technologies.

Objective: The purpose of this exploratory co-design case study was to solicit insights from nurses regarding nursing summary design considerations, including key information types and the preferred design prototype.

Methods: We recruited clinical nurses (N=33) from 7 inpatient units at a university hospital in the Midwestern United States using a purposive sampling method. We used images from a simulated nursing summary to generate visual card versions of the 46 information types currently included in an electronic health record vendor–generated nursing summary. Participants selected which cards to include and arranged them in their designs based on their perceived relevance of the information types to the summary and their preferred reading layout. The nurses' perceived relevance of information types to the summary was analyzed by quantifying the frequency of included cards, while the nurses' preferred reading layout was analyzed by quantifying the occurrence of closely paired cards to identify common groupings. After participants evaluated the information type cards, debriefing interviews were conducted and analyzed thematically to explore their rationales for the desired content and its arrangement.

Results: The participants demonstrated a high level of engagement in the activities. On average, all 33 participants included 61% (n=28) of the total information types (n=46). The most frequently included cards were "unit specimen" (results of the analysis of body fluid, tissue, or urine), "activity," "diet," and "hospital problems," each included by 33 participants. Participants most frequently preferred adjacency of the following pairs: "activity" and "diet" (paired by 26 participants; 79%) and "notes to physicians" and "notes to treatment team" (paired by 25 participants; 76%). Participants preferred arranging the cards to improve information accessibility, focusing on key information types.

Conclusions: Involving nurses in the co-design process may result in more useful and usable designs, thereby reducing the time required to navigate nursing summaries. Future work should include refining and evaluating prototypes based on the designs created by the nurses.

(JMIR Form Res 2025;9:e68906) doi:10.2196/68906

KEYWORDS

electronic health records; interview; nurses; user-computer interface; co-design

Introduction

Background

Ineffective communication among clinicians is a leading cause of sentinel events, such as falls and delays in treatments [1]. Given time constraints, nurses review vast amounts of patient information in electronic health records (EHRs) at the beginning of each shift; thus, effective tools for summarizing patient information are crucial [2]. Nursing summaries in EHRs are intended to provide a concise overview of essential patient data (eg, vital signs, intake and output, weight) on a single page, but rigid and incomplete formats have led to underutilization [3], resulting in nurses developing highly variable methods, such as paper notes, for summarizing patient information.

End-users are involved in the EHR development phase to prevent low usability and support optimal EHR utilization [4]. Co-design (or participatory design) has gained attention as a strategy to integrate stakeholder viewpoints, bridging the gap

XSL•FO
RenderX

between tools and users' needs to ensure the long-term adoption of tools [5]. Co-design aims to actively involve potential user groups in designing tools and interventions, leveraging their lived experiences to tailor outputs to their expressed needs and preferences, while aligning with existing workflows and available resources [6]. In health care, co-design has been applied across various settings and user groups, including in the creation of predictive analytics dashboards and decision support systems to enhance care quality and client outcomes in aged care [6]. Co-design has also been used to examine clinical nurses' general ideas about information display design in a nursing summary [2] and to develop a mobile health application to assist patients with obesity in self-managing their preparation for elective surgery [7].

The use of information cards as a tool for stakeholder engagement is an approach adopted by designers to structure the co-design ideation process [8,9]. Card sorting, a prevalent user-centered design technique, actively involves participants categorizing labeled cards based on criteria that most resonate with them [10]. Card sorting offers numerous benefits, including facilitating the organization of complex information into intuitive groupings, promoting consensus building among diverse user groups, enhancing the user experience by ensuring content and functionality are logically structured, and providing valuable insights into users' preferences and priorities [11]. Although we did not fully follow the card sorting technique, we adapted key principles from the card sorting method, such as labeling and analyzing the arrangement of cards, to suit this study context. This exploratory research focused on nurses' perceived usability and their feedback to inform nursing summary design strategies. The design strategies informed by nurse users' input could streamline navigational pathways, thereby reducing cognitive load. This aligns with the goals of the National Burden Reduction Collaborative, which unites health care organizations, federal agencies, and standards organizations to reduce clinical burden nationwide [12].

Problem Statement

Nurses' participation in EHR design is often limited, despite their role as key users [2,13]. A deeper understanding of their unique requirements and suggestions for EHR design is needed. Recognizing the pivotal role of end-user involvement in design [5], we used a co-design activity to examine nurses' perspectives on the use and arrangement of information types (eg, vital signs, intake and output, weight) within an existing EHR vendor's nursing summary. Specifically, the co-design activity aimed to gain insights into nurses' perspectives on nursing summary design, including which information types to include and how to arrange the information types.

Methods

Study Design

This was an exploratory, single session of co-design activities that gathered nurses' feedback and identified their design preferences. Although conducted independently, this study was part of an eye-tracking simulation, which focused on examining nurses' visual attention to EHR nursing summary information types (submitted to another journal).

Participants and Recruitment

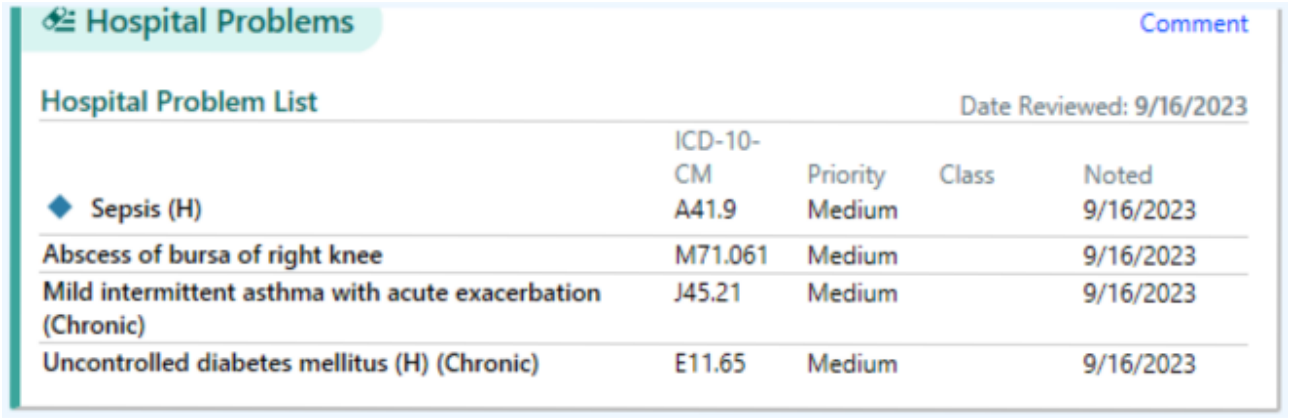
We recruited participants from 5 adult medical-surgical and 2 intermediate care units at a university hospital using a purposive sampling method. We aimed to recruit 30 or more participants, as this is recommended for eye-tracking simulations. These units used the same summary layout as the example summary employed in this study. To be included, participants must have had at least 12 months of experience as a registered nurse. We excluded nurses who were not in frontline patient care, such as nurse educators, nurse managers, and nurse research coordinators. Recruitment was facilitated by nurse managers, who distributed email flyers providing details about the study.

Data Collection

Participants were individually invited to a shared research space from August to September 2023. We used a nursing summary from a simulated patient case on a training platform that mirrored the EHR system used by the recruiting hospital. Since the hospital had used the Epic EHR system (Epic Systems) since 2011, we assumed all participants were familiar with its layout. The EHR vendor–generated nursing summary included 46 information types (eg, vital signs, intake and output, weight). We printed and cut out cards for each information type so the participants could move them around as they liked (Figure 1). See Multimedia Appendix 1 for the full original nursing summary layout and Multimedia Appendix 2 for details of the information types. A lead researcher (SP) conducted all co-design activities independently, while a co-author (JLM) who is a human factors expert provided guidance on the study design and monitored the co-design activities after each session.

XSL•FO
RenderX

Figure 1. Example of an information type card. Each information type card used a consistent design, including a colored boundary, heading, and optional content.



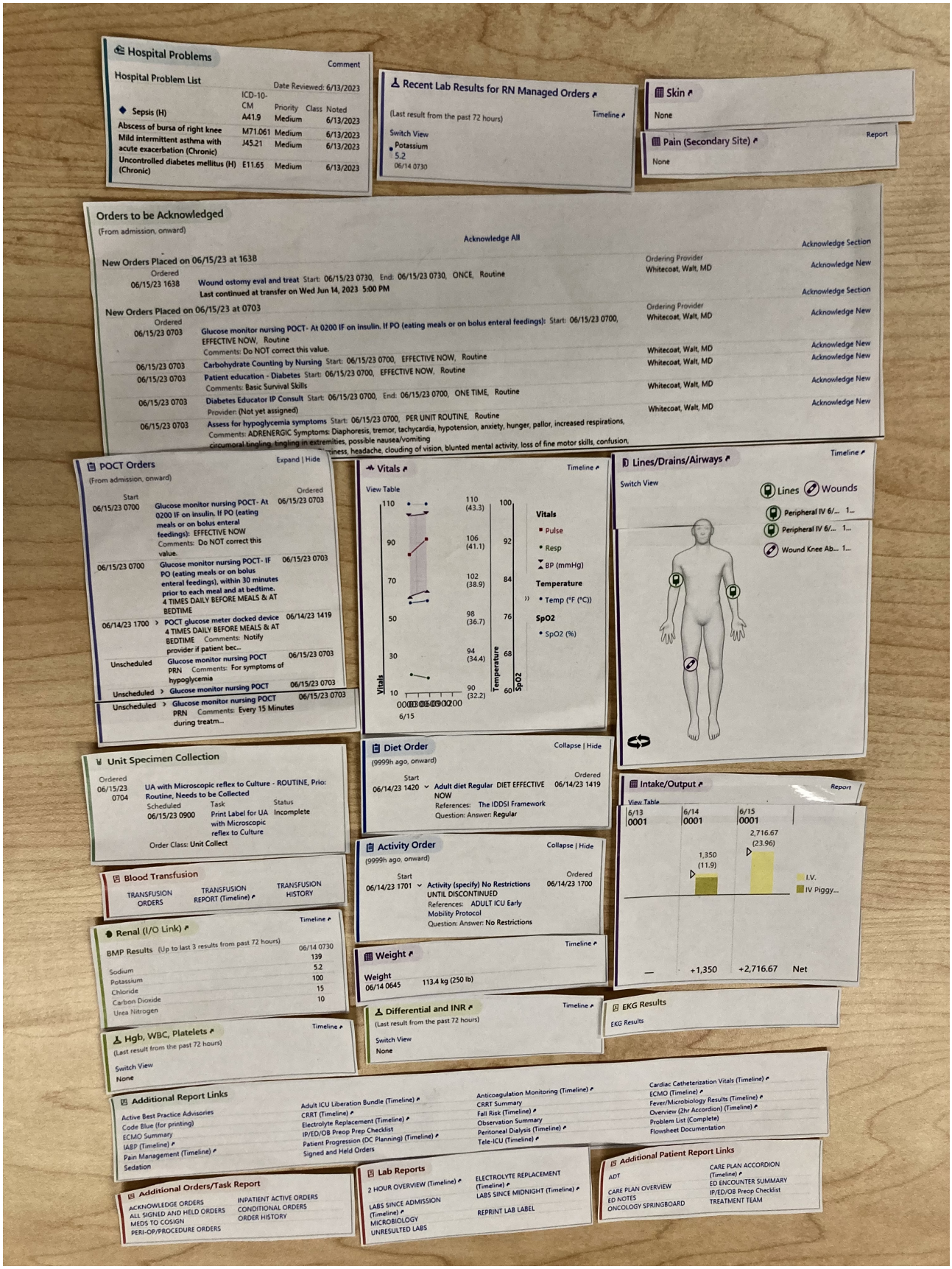
First, each participant reviewed each information type card to determine whether it was important enough to be included in the summary. Participants then arranged the included cards they wanted to present in the nursing summary on a desk to develop their preferred layout for the nursing summary screen. We clarified that the focus of the study was on determining the relevance of specific types of information (ie, social determinants of health) in the context of a nursing summary and arranging them, rather than the graphical design or the content of the information within the cards. We photographed each participant's design prototype for subsequent analysis.

Figure 2 shows an example of a design prototype created by one participant. Following the activity, a debriefing interview was conducted to collect feedback, and each conversation was audio-recorded via Zoom for subsequent transcription and analysis. Participants were prompted with the following guided questions: (1) explain your reasoning process in relation to the design layout; (2) discuss any specific considerations you made regarding information arrangement; and (3) offer any additional comments on your design.

We collected demographic information through a survey administered after the debriefing interviews.

XSL•FO
RenderX

Figure 2. A sample summary design prototype by a participant. The layout shows the information types to include and their preferred arrangement.



Ethical Considerations

The study protocol was approved by the University of Minnesota Institutional Review Board (STUDY00018047). The principal investigator (SP) provided verbal and written information about the study to the participants and obtained verbal consent, as the requirement for written consent was waived. All data collected during the study were deidentified and securely stored in a

XSL•FO
RenderX

university cloud-based storage system. The audio recordings and transcripts were deleted after the data analysis was completed. As this study was a part of an eye-tracking simulation, participants received a US $65 gift card for their time and effort upon completing all activities.

Data Analysis

We evaluated the relevance of each information type to the summary by quantifying the number of nurses who chose to include the information type card in the design output. We then quantified the frequency of card pairs to identify which information type cards nurses tended to place next to each other in their design prototype, providing insights into the cohesive grouping of information [11]. One researcher (SP) reviewed the deidentified transcripts of the debriefing interviews for accuracy and coded for thematic analysis. The codes and overarching themes were reviewed in weekly meetings with another researcher (JLM).

Results

Sample Characteristics

The characteristics of the 33 participants are presented in Table 1. More than half (n=19, 58%) of the participants were 25 - 44 years old, 85% (n=28) were female, 76% (n=25) were White, and none were Hispanic. Most nurses included in this analysis had at least a baccalaureate degree related to nursing (n=30, 91%). The participants generally had fewer years of nursing experience and worked fewer hours compared to a representative sample of national nurses (median=13 years and median=37 hours, respectively) [14].

XSL•FO
RenderX

Table . Baseline characteristics of the participants (N=33).

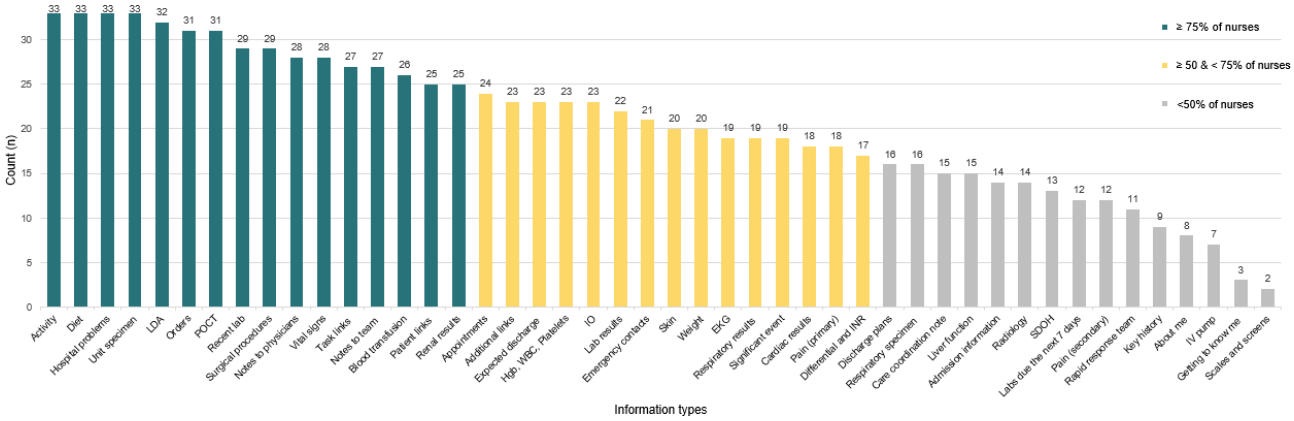
Characteristic	Value
Age (years), n (%)	
18 - 24	9 (27)
25 - 44	19 (58)
45 - 64	5 (15)
≥65	0 (0)
Sex, n (%)	
Male	5 (15)
Female	28 (85)
Other	0 (0)
Race, n (%)	
Black or African American	5 (15)
White	25 (76)
Asian	2 (6)
Native American or Alaskan	0 (0)
Other	0 (0)
Prefer not to say	1 (3)
Ethnicity, n (%)	
Hispanic or Latinx	0 (0)
Non-Hispanic	33 (100)
Other	0 (0)
Relationship status, n (%)	
Single	14 (42)
Married	14 (42)
Partnered	2 (6)
Divorced	2 (6)
Separated	1 (3)
Highest nursing degree, n (%)	
Diploma	3 (9)
BSN	28 (85)
MS or MSN	2 (6)
PhD or DNP	0 (0)
Hours worked per week, median (IQR)	36 (30-36)
Years of registered nurse experience, median (IQR)	4 (1.5-8)
Years of using electronic health records, median (IQR)	4 (1.5-8)

Card Frequency

On average, participants included 28 of the 46 information type cards in their design prototypes. Notably, all participants (N=33) included “unit specimen” (results of the analysis of body fluid, tissue, or urine), “hospital problems,” “diet,” and “activity” in their design prototype (Figure 3). However, certain information type cards were included less frequently: fewer than 50% of participants included 15 of the information type cards, and fewer than 10 participants included “key history and social determinants” (n=9), “about me” (individualized notes) (n=8), “intravenous pump settings” (n=7), “patient history notes” (n=3), and “scale and screen documentation” (n=2).

XSL•FO
RenderX

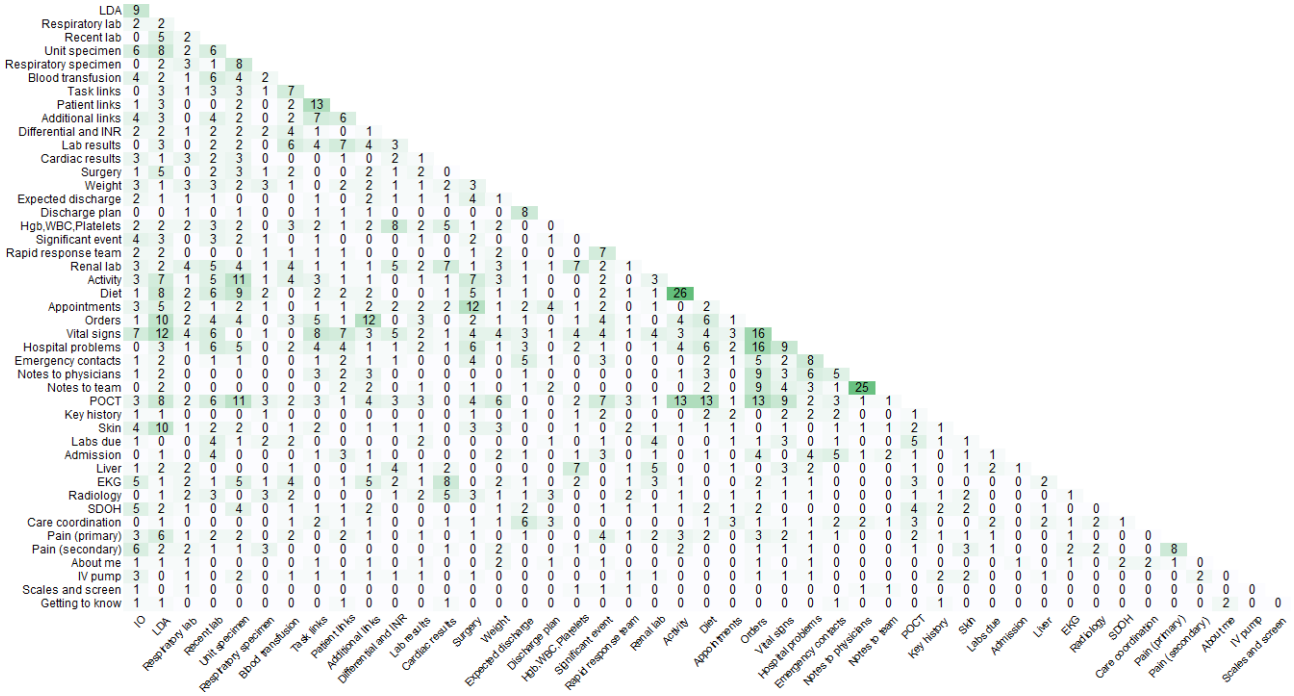
Figure 3. Frequency of included cards. Participants selected information type cards that they think are relevant to include in nursing summaries. EKG: electrocardiogram; Hgb: hemoglobin; INR: international normalized ratio; IO: intake and output; IV: intravenous; LDA: lines, drains, and airways; POCT: point-of-care testing; SDOH: social determinants of health; WBC: white blood cell.



Card Adjacency

Figure 4 shows the number of participants who placed each pair of information type cards next to each other in their design. Darker shading was used to highlight cells where 2 information type cards were located next to each other more frequently. The pairs of information type cards located next to each other most frequently were "activity" and "diet" (n=26, 79%), followed by "notes to physicians" and "notes to treatment team" (n=25, 76%). Other pairs of cards that were placed next to each other included "orders" and "hospital problems" (n=16, 48%) and "orders" and "vital signs" (n=16, 48%). "Orders" and "hospital problems," and "orders" and "vital signs" are not located next to each other in the current EHR design.

Figure 4. Item-by-item matrix of card adjacency. The frequency of adjacency was manually counted when 2 cards were placed next to each other with no third card in between. EKG: electrocardiogram; Hgb: hemoglobin; INR: international normalized ratio; IO: intake and output; IV: intravenous; LDA: lines, drains, and airways; POCT: point-of-care testing; SDOH: social determinants of health; WBC: white blood cell.



Interview Findings

Table 2 presents themes, subthemes, and exemplary quotes derived from the debriefing interviews. The participants emphasized the importance of readability in their design choices, opting for a layout that facilitates top-to-bottom reading of key information based on priority. They noted that grouping related information together would facilitate information scanning. They also prioritized identifying immediate action items (eg, physicians' orders) over delving into patient characteristics, such as social determinants of health. Owing to the typical time constraints needed to review patient information at the beginning of a shift, participants removed redundant information cards that hindered a swift and effective patient overview. Last, beyond the given design layout, participants provided design-related commentaries, pointing out that some information types required unnecessary clicks to get to the desired information, which limited usability.

XSL•FO
RenderX

Table . Common themes, subthemes, and supporting quotes from debriefing sessions based on their co-design activities, uncovering participants' reasoning process and design considerations.

Themes and subthemes	Quotes
Improving information accessibility	
Key information positioned upward	• "I have it from top to bottom. So it's the least important thing towards the bottom." [P11] • "Diet is not as important, so I kept it going to the bottom." [P24] • "I usually place it at the top of the page, not necessarily because it's important, but because it needs to be visible for everyone to see…Otherwise, no one else will notice it. Not everyone will scroll down." [P4] • "Most important would be at like the top, in a top-down order." [P1] • "I like that the orders are at the top because they're the most important; they're what the doctor wants us to be doing with the patient." [P9]
Proximity of similar information	• "I want lab results in the same spot." [P24] • "That'd be nice if they were together. Orders around orders. Orders can stick together." [P1]
Focusing on key information types	
Go-to action items over background details	• "I believe that social determinants of health are important. However, during my admission process, I do not check the social determinants of health. It is crucial for social workers and us to better understand the patient, but it is not my initial priority. When I first receive patients, I do not check the social determinants of health." [P26] • "I always want to know what orders are because these would be done right away." [P22]
Removing redundancy	• "The additional report things have a fall risk timeline, but then there's a fall risk injury (information card). I just feel like it's totally redundant." [P22] • "The card, 'Getting to know me,' I feel like this was redundant with 'About me.' I honestly never use this." [P5]
Easy-to-use design	
Unnecessary steps to the desired information	• "We cannot see the treatment team unless we do another click here and check the treating team." [P26] • "...a lot of people have a hard time with, like, everything's here but you have to click a lot to get out of here, right? If there's a transfusion order, this shouldn't be bold or an asterisk (on the order name). There should be an order you need to give blood or something that you could click on…Why do you need extra things to press?" [P22] • "I really don't use this summary page that much, because it's just so overwhelming. I guess it's just too many things there that I don't even know, like how to navigate it, which is why I made my system, which is more clicking." [P5]

Discussion

Principal Results

The co-design case study highlights the importance of aligning EHR nursing summary design with nurses' cognitive workflows, demonstrating that nurses prioritize and organize information in ways that make sense to them. Concise, relevant information allows nurses to quickly review a patient's status, while strategically grouping related data helps them navigate and interpret complex information more efficiently. The current nursing summary contains overwhelming amounts of information that is often presented ineffectively, necessitating extensive searching to consolidate scattered data. The findings indicate a significant need to remove less relevant information from the summary. Also, nurses emphasized the importance of "unit specimen" (results of the analysis of body fluid, tissue, or urine), "activity," "diet," and "hospital problems" in their design prototype, suggesting these are the key information types for patient overviews. Nurses tended to prioritize obtaining information on current problems, daily living status, or treatment plans. Medical history or miscellaneous notes can be a source of information but are not always necessary for the initial patient overview. Our findings are consistent with prior work in which clinicians used a limited number of clinical information concepts at the time of patient admission, suggesting better electronic data management strategies, including the priority display of frequently used clinical concepts [15].

XSL•FO
RenderX

Participants strategically arranged information cards to align with their cognitive workflow by placing them adjacent to other relevant cards. The adjacency of information often signifies the similarity or relative importance of the information [16]. Previous studies have also emphasized the benefits of placing paired EHR information adjacent to each other on the screen to facilitate user interface navigation [17-19]. Some information pairs nurses identified with high frequency did not align with the current nursing summary layout. For example, "orders" and "hospital problems," "orders" and "vital signs," "orders" and "point-of-care-testing order" (tests that are conducted at the patient's bedside), and "point-of-care-testing order" and "activity" were frequently located next to each other in the participant designs, but not in the existing layout. In addition to the adjacent placement of relevant information, top-to-bottom organization of information was commonly found in the designs. The placement of items at the top of the screen naturally calls attention to those items [20]. Instead of scrolling down for details on the screen, nurses tended to place good-to-know but not essential information toward the bottom of their designs. The concise and cognitively aligned design supports nurses in narrowing their focus from the comprehensive patient record to a targeted subset of relevant information, thereby enhancing the efficiency of initial assessments and ensuring that pertinent details are readily accessible when required [21].

Strengths and Limitations

These findings show the value of co-design activities that involve nurses in the design process—to create an EHR nursing summary that better supports their needs. By prioritizing the most relevant patient information and streamlining the data presentation, the study offers design strategies that could alleviate the EHR burden of navigating from one piece of information to another. The identified design strategies could enhance efficiency and support the National Burden Reduction Collaborative's mission to reduce the overall burdens on health care professionals [12].

While this case study successfully identified nurses' information preferences aligning with their workflow, it has limitations. The use of a single simulated patient case for the co-design activities and the small sample size from a single institution may limit the generalizability of the findings. Due to the small sample size, we did not group participants by demographics (eg, unit-specific role, shift, work experience). While this study aimed to explore general information needs and design considerations for nursing summaries by including participants from multiple units, nurses' specific needs may vary based on their characteristics. For example, nurses in surgical units may prioritize surgery-related information, while nurses in general medical units may focus on different aspects. Additionally, night shift nurses may prioritize information about upcoming examinations for the next day. Nurses' characteristics may also influence their priorities and responsibilities, underscoring the importance of considering these variables in future research.

Conclusion

This co-design case study provides valuable insights into nurses' preferences and strategies for organizing information in an EHR nursing summary, underscoring the need to redesign the layout based on its practical use. We emphasized the significance of reducing less relevant information and optimizing the layout according to nurses' information priorities. These efforts have the potential to improve the usefulness of the current nursing summary. The direct involvement of nurses as co-designers holds promise for developing user-centered EHR tools that support nursing cognition, decision-making, and workflow.

Acknowledgments

We would like to express our gratitude to the participants for generously sharing their time and experiences. This study was supported by SP's Doctoral Dissertation Fellowship, University of Minnesota and Midwest Nursing Research Society/Council for Advancement of Nursing Science Dissertation Grant (grant CON000000104586).

Data Availability

The data generated during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Nursing summary layout.
[PDF File, 1437 KB - formative_v9i1e68906_app1.pdf]

Multimedia Appendix 2
Descriptions of information types in nursing summaries.
[DOCX File, 18 KB - formative_v9i1e68906_app2.docx]

References

XSL•FO
RenderX

1. Sentinel event data summary. The Joint Commission. 2024. URL: https://www.jointcommission.org/resources/sentinel-event/sentinel-event-data-summary/ [accessed 2024-11-19]
2. Park S, Marquard J. Nurse-centered co-design of an electronic health record nursing summary. Human Factors in Healthcare 2024 Jun;5:100065. [doi: 10.1016/j.hfh.2024.100065]
3. Staggers N, Clark L, Blaz JW, Kapsandoy S. Why patient summaries in electronic health records do not provide the cognitive support necessary for nurses' handoffs on medical and surgical units: insights from interviews and observations. Health Informatics J 2011 Sep;17(3):209-223. [doi: 10.1177/1460458211405809] [Medline: 21937463]
4. Busse TS, Jux C, Laser J, et al. Involving health care professionals in the development of electronic health records: scoping review. JMIR Hum Factors 2023 Jul 10;10:e45598. [doi: 10.2196/45598] [Medline: 37428524]
5. Voorend R, Derboven J, Slegers K. Distributed user-generated card based co-design: a case study. Presented at: CHI '19; May 4-9, 2019; Glasgow, Scotland, United Kingdom. [doi: 10.1145/3290607.3312815]
6. Ludlow K, Westbrook J, Jorgensen M, et al. Co-designing a dashboard of predictive analytics and decision support to drive care quality and client outcomes in aged care: a mixed-method study protocol. BMJ Open 2021 Aug 25;11(8):e048657. [doi: 10.1136/bmjopen-2021-048657] [Medline: 34433599]
7. Song T, Yu P, Bliokas V, et al. A clinician-led, experience-based co-design approach for developing mhealth services to support the patient self-management of chronic conditions: development study and design case. JMIR Mhealth Uhealth 2021 Jul 20;9(7):e20650. [doi: 10.2196/20650] [Medline: 34283030]
8. Alvarez CP, Martinez-Maldonado R, Buckingham Shum S. LA-DECK: a card-based learning analytics co-design tool. Presented at: LAK '20: 10th International Conference on Learning Analytics and Knowledge; Mar 23-27, 2020; Frankfurt, Germany. [doi: 10.1145/3375462.3375476]
9. Kwiatkowska J, Szóstek A, Lamas D. (Un)structured sources of inspiration: comparing the effects of game-like cards and design cards on creativity in co-design process. Presented at: PDC '14: Participatory Design Conference; Oct 6-10, 2014; Windhoek, Namibia. [doi: 10.1145/2661435.2661442]
10. Tankala S, Sherwin K. Card sorting: uncover users' mental models. : Nielsen Norman Group; 2024 Feb 2. URL: https://www.nngroup.com/articles/card-sorting-definition/ [accessed 2024-04-01]
11. Righi C, James J, Beasley M, et al. Card sort analysis best practices. J Usability Stud 2013;8(3):69-89 [FREE Full text]
12. New group hopes to focus efforts to trim clinician burdens. Health Data Management. 2023 Feb. URL: https://www.healthdatamanagement.com/articles/new-group-hopes-to-focus-efforts-to-trim-clinician-burdens [accessed 2024-11-16]
13. Martikainen S, Kaipio J, Lääveri T. End-user participation in health information systems (HIS) development: physicians' and nurses' experiences. Int J Med Inform 2020 May;137:104117. [doi: 10.1016/j.ijmedinf.2020.104117] [Medline: 32179254]
14. Melnick ER, West CP, Nath B, et al. The association between perceived electronic health record usability and professional burnout among US nurses. J Am Med Inform Assoc 2021 Jul 30;28(8):1632-1641. [doi: 10.1093/jamia/ocab059] [Medline: 33871018]
15. Pickering BW, Gajic O, Ahmed A, Herasevich V, Keegan MT. Data utilization for medical decision making at the time of patient admission to ICU. Crit Care Med 2013 Jun;41(6):1502-1510. [doi: 10.1097/CCM.0b013e318287f0c0] [Medline: 23528804]
16. Wagemans J, Elder JH, Kubovy M, et al. A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. Psychol Bull 2012 Nov;138(6):1172-1217. [doi: 10.1037/a0029333] [Medline: 22845751]
17. Marian AA, Dexter F, Tucker P, Todd MM. Comparison of alphabetical versus categorical display format for medication order entry in a simulated touch screen anesthesia information management system: an experiment in clinician-computer interaction in anesthesia. BMC Med Inform Decis Mak 2012 May 29;12(1):46. [doi: 10.1186/1472-6947-12-46] [Medline: 22643058]
18. Senathirajah Y. Safer design - composable EHRs and mechanisms for safety. Stud Health Technol Inform 2015;218:86-92. [Medline: 26262532]
19. Zheng K, Padman R, Johnson MP, Diamond HS. An interface-driven analysis of user interactions with an electronic health records system. J Am Med Inform Assoc 2009;16(2):228-237. [doi: 10.1197/jamia.M2852] [Medline: 19074301]
20. Senathirajah Y, Kaufman D, Bakken S. The clinician in the driver's seat: part 2 - intelligent uses of space in a drag/drop user-composable electronic health record. J Biomed Inform 2014 Dec;52:177-188. [doi: 10.1016/j.jbi.2014.09.008] [Medline: 25445921]
21. Horsky J, Ramelson HZ. Development of a cognitive framework of patient record summary review in the formative phase of user-centered design. J Biomed Inform 2016 Dec;64:147-157. [doi: 10.1016/j.jbi.2016.10.004] [Medline: 27725292]

Abbreviations

EHR: electronic health record

XSL•FO
RenderX

Edited by A Mavragani; submitted 19.11.24; peer-reviewed by A Kuusisto, M Alshammari; revised version received 15.02.25; accepted 19.02.25; published 11.03.25.

Please cite as:
Park S, Marquard JL, Austin RR, Martin CL, Pieczkiewicz DS, Delaney CW
Exploratory Co-Design on Electronic Health Record Nursing Summaries: Case Study
JMIR Form Res 2025;9:e68906
URL: https://formative.jmir.org/2025/1/e68906
doi:10.2196/68906

© Suhyun Park, Jenna L Marquard, Robin R Austin, Christie L Martin, David S Pieczkiewicz, Connie W Delaney. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Improving the User Interface and Guiding the Development of Effective Training Material for a Clinical Research Recruitment and Retention Dashboard: Usability Testing Study

Leah Leslie Gardner[1], PhD; Pezhman Raeisian Parvari[1], MA; Mark Seidman[1], MA; Richard J Holden[1,2,3,4], PhD; Nicole R Fowler[1,2,4], MHSA, PhD; Ben L Zarzaur[5], MPH, MD; Diana Summanwar[6], MD; Cristina Barboi[7], MS, MD; Malaz Boustani[1,2,4,8], MPH, MD

[1]Center for Health Innovation and Implementation Science, School of Medicine, Indiana University, 410 W 10th St #1140, Indianapolis, IN, United States

[2]Department of Medicine, School of Medicine, Indiana University, Indianapolis, IN, United States

[3]School of Public Health, Indiana University Bloomington, Bloomington, IN, United States

[4]Center for Aging Research, Regenstrief Institute Inc, Indianapolis, IN, United States

[5]Department of Surgery, Division of Acute Care and Regional General Surgery, University of Wisconsin-Madison, Madison, WI, United States

[6]Department of Family Medicine, School of Medicine, Indiana University, Indianapolis, IN, United States

[7]Department of Anesthesiology, School of Medicine, Indiana University, Indianapolis, IN, United States

[8]Sandra Eskenazi Center for Brain Care Innovation, Eskenazi Health, Indianapolis, IN, United States

Corresponding Author:
Pezhman Raeisian Parvari, MA
Center for Health Innovation and Implementation Science, School of Medicine, Indiana University, 410 W 10th St #1140, Indianapolis, IN, United States

Abstract

Background: Participant recruitment and retention are critical to the success of clinical trials, yet challenges such as low enrollment rates and high attrition remain ongoing obstacles. RecruitGPS is a scalable dashboard with integrated control charts to address these issues by providing real-time data monitoring and analysis, enabling researchers to better track and improve recruitment and retention.

Objective: This study aims to identify the challenges and inefficiencies users encounter when interacting with the RecruitGPS dashboard. By identifying these issues, the study aims to inform strategies for improving the dashboard's user interface and create targeted, effective instructional materials that address user needs.

Methods: Twelve clinical researchers from the Midwest region of the United States provided feedback through a 10-minute, video-recorded usability test session, during which participants were instructed to explore the various tabs of the dashboard, identify challenges, and note features that worked well while thinking aloud. Following the video session, participants took a survey on which they answered System Usability Scale (SUS) questions, ease of navigation questions, and a Net Promoter Score (NPS) question.

Results: A quantitative analysis of survey responses revealed an average SUS score of 61.46 (SD 23.80; median 66.25) points, indicating a need for improvement in the user interface. The NPS was 8, with 4 of 12 (33%) respondents classified as promoters and 3 of 12 (25%) as detractors, indicating a slightly positive satisfaction. When participants compared RecruitGPS to other recruitment and study management tools they had used, 8 of 12 (67%) of participants rated RecruitGPS as better or much better. Only 1 of 12 (8%) participants rated RecruitGPS as worse but not much worse. A qualitative analysis of participants' interactions with the dashboard diagnosed a confusing part of the dashboard that could be eliminated or made optional and provided valuable insight for the development of instructional videos and documentation. Participants liked the dashboard's data visualization capabilities, including intuitive graphs and trend tracking; progress indicators, such as color-coded status indicators and comparison metrics; and the overall dashboard's layout and design, which consolidated relevant data on a single page. Users also valued the accuracy and real-time updates of data, especially the integration with external sources like Research Electronic Data Capture (REDCap).

Conclusions: RecruitGPS demonstrates significant potential to improve the efficiency of clinical trials by providing researchers with real-time insights into participant recruitment and retention. This study offers valuable recommendations for targeted refinements to enhance the user experience and maximize the dashboard's effectiveness. Additionally, it highlights navigation challenges that can be addressed through the development of clear and focused instructional videos.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e66718) doi:10.2196/66718

KEYWORDS

recruitment strategies; clinical research; research subject recruitment; agile science; agile implementation; human-computer interaction

Introduction

Clinical studies are essential to advancing medical science and improving patient outcomes, and their successful completion relies on adequate participant recruitment [1]. Effective and efficient patient recruitment and retention ensure timely collection of data, while insufficient enrollment often leads to delays or even trial failure [2-5]. Studies indicate that 80% of trials face recruitment challenges [6], while 37% fail to meet their target sample size goals [7]. In fact, inadequate recruitment is the leading cause of trial discontinuation [8,9].

Clinical trial management systems (CTMSs), clinical trial recruitment support systems (CTRSSs), web-based platforms, patient registries, social media, and community engagement initiatives [10-14] are widely used to support recruitment and management of clinical trials. Electronic health records (EHRs) facilitate targeted recruitment by identifying potential participants based on health data, while CTMSs track participant enrollment, scheduling, and communication [15,16]. Online platforms like ClinicalTrials.gov provide global access to trial information, and social media platforms such as Facebook enable broader outreach through targeted advertisements [17,18]. While these tools offer valuable support for recruitment, they often face limitations such as high costs, limited reach, or sustainability challenges [19].

Dashboards are effective tools in clinical settings, particularly for tracking quality and safety metrics [20]. In clinical trials, dashboards are used to generate alerts when participant accrual is inadequate [3,21]. Commercially available CTMSs with dashboards to support large-scale clinical trials include Medidata Rave (Medidata Solutions) [22], Veeva Vault CDMS (Veeva Systems) [23], Ennov Clinical (Ennov) [24], Viedoc (Viedoc Technologies AB) [25], RealTime eClinical (RealTime Software Solutions) [26], and Clinevo (Clinevo Technologies) [27]. These systems have a variety of features for trial oversight, from patient recruitment to data monitoring, such as patient portals to streamline patient communication, consent capture, financial tracking, regulatory compliance, adverse event tracking, and study closeout.

Castor EDC (Castor) [28] and Ripple Science (Ripple Science Inc) [29] offer affordable packages for academic institutions and smaller companies. These packages have dashboards that help researchers track study progress, tasks, and participant data across different trials and that automate email or in-app notifications and reminders to both staff and participants.

OpenClinica (OpenClinica LLC) [30] and REDCap (Research Electronic Data Capture; Vanderbilt University) [31] are examples of open-source platforms for managing clinical trial data. OpenClinica Insight is an add-on module for OpenClinica 3 Enterprise and OpenClinica 4 that allows the user to create reports and dashboards. REDCap has limited dashboarding capabilities and requires the user to export data to a spreadsheet, such as Microsoft Excel, or statistical package, such as R and R Shiny, for detailed analysis and dashboard development.

Since REDCap requires the user to export data to a spreadsheet or other software for analysis, a research group from the Center for Health Innovation and Implementation Science (CHIIS) successfully developed and implemented an Excel-based dashboard in a single, multisite clinical study already in progress using REDCap [32]. This dashboard integrates control charts and enables seamless data updates from the REDCap database, allowing for effective monitoring of recruitment and participant progress. In this study, participants moved through various stages of the research workflow. The dashboard provided investigators with weekly updates on participant progress through these stages, facilitating timely interventions when participant counts or progress rates at any stage suggested that sample size targets might not be met. Additionally, it enabled the application of Agile science methods [33-37] to rapidly test and implement strategies for improving recruitment.

Building on the success of the dashboard used in the multisite clinical study, the research group developed a customizable, generic dashboard called RecruitGPS, which can be customized for other clinical research studies. This paper focuses on the detailed usability testing of the RecruitGPS user interface. A comprehensive description of RecruitGPS features is provided in Multimedia Appendix 1. The aim of this study is to identify the challenges and inefficiencies users experienced with the RecruitGPS interface so that these insights can be leveraged to improve usability and develop effective training materials.

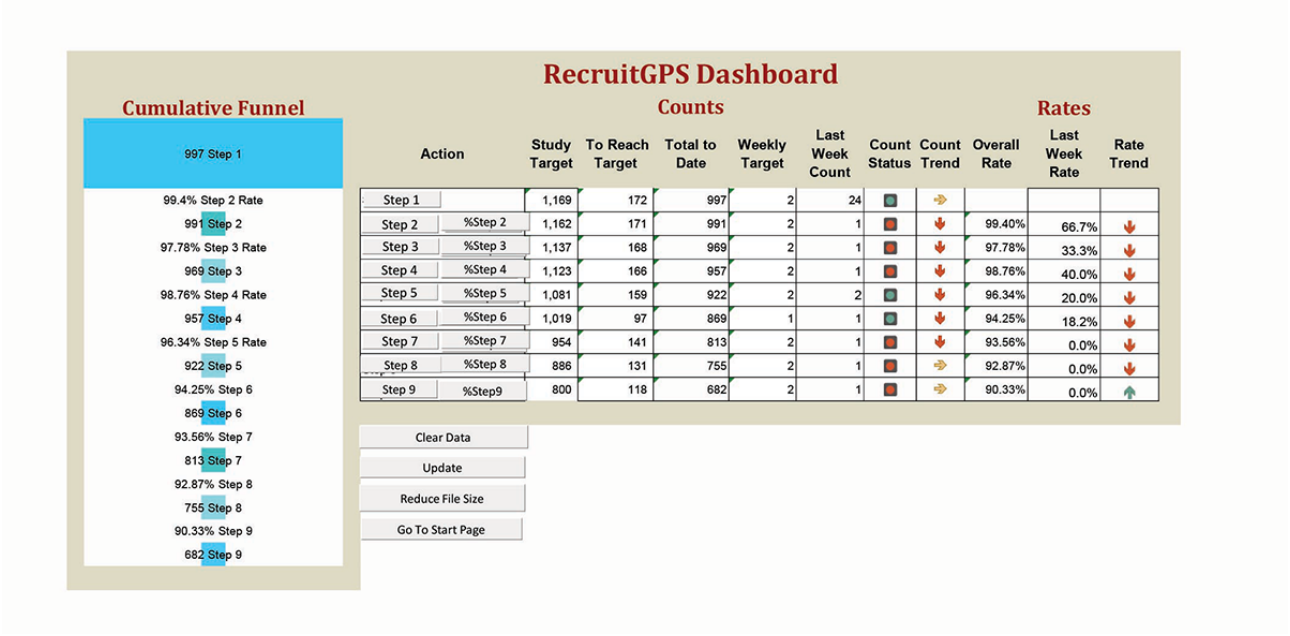
Methods

Overview of the RecruitGPS Dashboard

The goal of the user interface is to provide clinical trial principal investigators and research coordinators with immediate access to weekly and cumulative counts at each stage of the study and weekly and cumulative transition rates from stage to stage. The interface uses compelling visual cues to bring the user's attention to potential problems in the study workflow. The first version of the user interface for RecruitGPS seen in Figure 1 was the version from the successful multisite study dashboard user interface [32] with the generic names step 1 through step 9 substituted for the names of the stages screened through 18-month assessment. The multisite study user interface was collaboratively designed by the stakeholders in the study and the dashboard developer.

XSL•FO
RenderX

Figure 1. The RecruitGPS dashboard interface evaluated during usability testing.



Think-aloud usability testing sessions were combined with a user survey to assess the effectiveness of, efficiency of, and satisfaction with the RecruitGPS dashboard. Effectiveness is the extent to which the user can achieve a goal with accuracy and completeness. Efficiency is the effort and resources necessary to achieve a complete and accurate goal. Satisfaction is a positive user experience and absence of discontent during task performance. Once the 3 are fulfilled adequately, the product can be considered to have attained an acceptable level of usability [38].

Participant Recruitment

Participants in the study were recruited to use the RecruitGPS user interface while being video recorded and to complete a survey about their experience. Most participants were affiliated with CHIIS or were clinical researchers primarily based in the Midwest region of the United States. Recruitment and usability testing were conducted in August 2024.

Participants were recruited using convenience and snowball sampling techniques, with the CHIIS team supporting the process through digital advertisements, emails, and word of mouth. To be eligible, participants needed experience with clinical research studies, fluency in English, and access to the internet and Zoom (Zoom Video Communications Inc.). Although the literature suggests that 5 to 8 participants are sufficient for think-aloud usability testing, 12 participants were recruited to ensure diverse and comprehensive insights [39,40].

Usability Testing Sessions and Surveys

A total of twelve 10-minute usability testing sessions of the RecruitGPS dashboard were conducted via Zoom. During each session, the facilitator shared their screen, displaying the dashboard, and participants were provided with remote control access to interact with the interface. Participants were instructed as follows: “consider[ing] a research project of your own, explore the various tabs of the dashboard, identify challenges, and note features that work well while thinking aloud.” To identify specific problems, the researchers recorded the 10-minute usability testing sessions for a qualitative analysis of the challenges users encountered and features that worked well.

Following each usability test, participants received a survey link to provide additional feedback. The survey gathered insights on users’ overall experience, including their System Usability Scale (SUS) (Textbox 1) [41] and Net Promoter Score (NPS) [42] ratings. The NPS is based on the question “On a scale of 0 to 10, how likely are you to recommend this dashboard to a colleague or friend?” It is reported as a number ranging from −100 to +100, where a higher score is desirable. Respondents are categorized as detractors (giving scores of 0 - 6), passives (giving scores of 7 - 8), and promoters (giving scores of 9 - 10) The percentages of promoters and detractors are calculated and then the percentage of detractors is subtracted from the percentage of promoters to obtain the NPS [42]. The survey also included open-ended questions about the most and least useful features, as well as overall impressions of the user experience and dashboard interface. To assess ease of navigation in the survey, users were asked the questions in Table 1.

XSL•FO
RenderX

Textbox 1. System Usability Scale (SUS) questions in the survey following the usability test.

SUS Questions (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree)

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

Table . Questions on ease of navigation in the survey following the usability test.

Question	Available responses
How easy was it to navigate through the dashboard?	Likert scale (1 - 5, with 1 being the most difficult and 5 being the easiest)
Overall, how does this recruitment dashboard compare to other recruitment tools you have used?	Scale: 1=much worse, 2=worse, 3=about the same, 4=better, 5=much better
How intuitive did you find the user interface (UI) of this recruitment dashboard?	Scale: 1=not intuitive at all, 2=slightly intuitive, 3=neutral, 4=intuitive, 5=very intuitive

Ethical Considerations

This study was approved by the Human Research Protection Program at Indiana University (23755) prior to participant recruitment. Informed consent was obtained from all participants. Throughout the study, strict confidentiality and privacy protocols were followed. All data collected were deidentified before analysis to maintain participant anonymity. To protect privacy, secure data storage and anonymization procedures were used. Furthermore, no identifiable images of participants were included in the data. Participation was voluntary and no financial compensation or incentives were provided.

Results

This analysis assesses the effectiveness and efficiency of the RecruitGPS dashboard and satisfaction with it.

Participants

Twelve medical professionals completed the usability testing and surveys. Their demographics are summarized in Table 2.

XSL•FO
RenderX

Table . Education levels, roles, and experience of study participants (N=12) completing the usability testing and surveys.

Characteristics	Participants, n (%)
Education	
Doctorate (PhD)	2 (17)
Master's degree	4 (33)
Professional degree (MD, JD)	6 (50)
Role in research	
Principal investigator or coprincipal investigator	5 (42)
Project manager	5 (42)
Research coordinator	2 (17)
Experience	
1 - 3 years	4 (33)
4 - 6 years	2 (17)
7 - 10 years	1 (8)
Less than 1 year	1 (8)
More than 10 years	4 (33)

Effectiveness

The 10 questions on the SUS together assess overall usability, computed on a scale from 0 to 100, as described in detail by Brooke [41]. Table 3 reports individual item scores and overall SUS scores for each participant. Higher ratings on odd-numbered questions and lower ratings on even-numbered questions indicate better usability.

Table . System Usability Scale (SUS) total scores and scores for each question (Q) by participant (P).

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
Q1	2	3	2	4	3	3	4	1	4	4	3	4
Q2	1	4	1	1	2	3	3	1	2	1	3	4
Q3	1	4	1	3	3	3	3	0	3	3	3	4
Q4	1	4	0	2	3	0	3	0	1	3	3	4
Q5	2	3	2	3	3	3	2	1	2	3	3	4
Q6	4	4	2	2	3	3	3	1	4	3	3	4
Q7	1	3	2	4	2	3	3	0	3	3	3	4
Q8	2	4	1	4	2	3	3	1	1	3	3	4
Q9	0	2	0	3	2	3	2	0	1	3	3	4
Q10	1	1	0	4	2	1	2	1	4	2	3	4
SUS score	37.5	80.0	27.5	75.0	62.5	62.5	70.0	15.0	62.5	70.0	75.0	100.0

The mean SUS total score for all participants was 61.46 (SD 23.80; median 66.25) points. According to Bangor et al [43], a score of 85 and above is excellent and a score of 70 - 84 is good.

Efficiency—Assessing Ease of Navigation

Feedback on ease of navigation showed a mixed picture. While 25% (n=3) of the 12 respondents gave a rating of 5% and 58% (n=7) gave a rating of 4, 17% (n=2) rated ease of navigation a 1.

Satisfaction

The NPS [42] for RecruitGPS in this study was 8, with 33% (n=4) of respondents classified as promoters and 25% (n=3) as detractors. This indicates a slightly positive satisfaction.

A second question asked participants to compare RecruitGPS to other recruitment tools they had used on a scale of 1 - 5, with 1 meaning RecruitGPS is much worse and 5 meaning Recruit GPS is much better. The results were that 8 of 12 participants thought RecruitGPS was better. The results are summarized in Table 4.

XSL•FO
RenderX

Table . Survey responses (N=12) comparing RecruitGPS to other recruitment tools used by participants.

Score	Interpretation	Number of responses, n (%)
1	Much worse	0 (0)
2	Worse	1 (8)
3	About the same	3 (25)
4	Better	5 (42)
5	Much better	3 (25)

Qualitative Feedback From Participants

The framework developed by Clarke and Braun [44] was used for thematic analysis of the recorded 10-minute usability testing sessions to identify and analyze patterns within the data, including challenges users encountered and features that worked well.

Challenges

Several users reported difficulties with specific graphical elements, such as the cumulative funnel, which they found confusing or unintuitive. Additionally, some users experienced challenges with unclear instructions and labels, leading to difficulties in interpreting data and understanding the dashboard's functionality. These issues were reflected in the variability of ease of navigation ratings, with some users finding the dashboard extremely difficult to use.

Features That Worked Well

Positive feedback centered on the dashboard's data visualization capabilities, including intuitive graphs and trend tracking, which were highly valued by users. Progress indicators, such as color-coded status indicators and comparison metrics, were appreciated for helping users monitor their recruitment goals effectively. The dashboard's layout and design, which consolidated relevant data on a single page with clear button labeling, contributed to user satisfaction. Users also valued the accuracy and real-time updates of data, especially the integration with external sources like REDCap, which was seen as a critical feature for effective data management.

Discussion

Principal Findings

We found that 8 of 12 (67%) participants considered RecruitGPS better than other recruitment tools they had used, but the usability (SUS), satisfaction (NPS), and efficiency scores all indicated a need for improvement. Feedback on ease of navigation showed significant efficiency challenges faced by a minority of users, suggesting that some users struggled with navigation, despite overall positive feedback.

The recommendations for improvement based on the qualitative analysis are follows:

1. Remove the cumulative funnel or make it an optional feature for users who request it.
2. Create videos with accompanying documentation to address challenges users encountered with (a) instructions and labels and (b) navigation.

Although the quantitative analysis highlighted areas for improvement, the testing was conducted without any prior training for participants. The developers of RecruitGPS had not anticipated users attempting to operate the system without access to documentation or instructional materials, such as training videos. Despite this, RecruitGPS performed well enough. Additionally, the qualitative analysis offered valuable insights for enhancing the documentation and provided clear guidance for the development of instructional videos.

Limitations

The first limitation of this study was the relatively small sample size, consisting of only 12 participants in the usability testing sessions and survey responses. Conducting future studies with a larger and more diverse user group could yield additional insights and a broader understanding of user needs.

A related limitation was the participants' affiliations, as most were connected to 1 of 2 medical schools or their associated health care systems. While some participants had experience at other research institutions, their familiarity with alternative recruitment tools may have been limited to those used within these 2 medical schools, potentially restricting the study's generalizability.

Another limitation was that the study did not include testing of the existing documentation, and the instructional videos had not yet been developed. Future testing of these resources could provide valuable feedback to refine and improve their effectiveness.

Conclusions

RecruitGPS demonstrates promise and potential in enhancing recruitment and retention for clinical trials by providing real-time insights, supporting agile decision-making, and using scalable design with an accessible user interface. Removing the cumulative funnel and creating instructional videos and documentation focused on navigation challenges are likely to enhance its usability. With these enhancements, RecruitGPS could become a key tool in clinical trial management, supporting successful participant recruitment and better research outcomes.

XSL•FO
RenderX

Acknowledgments

The study was supported by a grant (R25AG078136) from the National Institute on Aging.

Data Availability

The datasets generated during and/or analyzed during this study are anonymized and restricted for confidentiality. Any further inquiries can be directed to the corresponding author.

Conflicts of Interest

MB serves as a chief scientific officer and cofounder of BlueAgilis and the chief health officer of DigiCare Realized. He has equity interest in Blue Agilis and DigiCare Realized; he has sold equity in Preferred Population Health Management and MyShift (previously known as RestUp). He serves as an advisory board member for Eli Lilly and Co, Eisai; Merck & Co; Biogen; and Genentech. These conflicts have been reviewed by Indiana University and have been appropriately managed to maintain objectivity.

Multimedia Appendix 1
Detailed features of RecruitGPS dashboard.
[DOC File, 251 KB - formative_v9i1e66718_app1.doc]

References

1. Campbell MK, Snowdon C, Francis D, et al. Recruitment to randomised trials: strategies for trial enrollment and participation study. The STEPS study. Health Technol Assess 2007 Nov;11(48):iii, ix-105. [doi: 10.3310/hta11480] [Medline: 17999843]
2. Visanji EC, Oldham JA. Patient recruitment in clinical trials: a review of literature. Phys Ther Rev 2001 Jun 1;6(2):141-150. [doi: 10.1179/108331901786166532]
3. Toddenroth D, Sivagnanasundaram J, Prokosch HU, Ganslandt T. Concept and implementation of a study dashboard module for a continuous monitoring of trial recruitment and documentation. J Biomed Inform 2016 Dec;64:222-231. [doi: 10.1016/j.jbi.2016.10.010] [Medline: 27769890]
4. Vadeboncoeur C, Foster C, Townsend N. Challenges of research recruitment in a university setting in England. Health Promot Int 2018 Oct 1;33(5):878-886. [doi: 10.1093/heapro/dax025] [Medline: 28531308]
5. Lai YS, Afseth JD. A review of the impact of utilising electronic medical records for clinical research recruitment. Clin Trials 2019 Apr;16(2):194-203. [doi: 10.1177/1740774519829709] [Medline: 30764659]
6. Brøgger-Mikkelsen M, Ali Z, Zibert JR, Andersen AD, Thomsen SF. Online patient recruitment in clinical trials: systematic review and meta-analysis. J Med Internet Res 2020 Nov 4;22(11):e22179. [doi: 10.2196/22179] [Medline: 33146627]
7. Jacques RM, Ahmed R, Harper J, et al. Recruitment, consent and retention of participants in randomised controlled trials: a review of trials published in the National Institute for Health Research (NIHR) Journals Library (1997-2020). BMJ Open 2022 Feb 14;12(2):e059230. [doi: 10.1136/bmjopen-2021-059230] [Medline: 35165116]
8. Pung J, Rienhoff O. Key components and IT assistance of participant management in clinical research: a scoping review. JAMIA Open 2020 Oct;3(3):449-458. [doi: 10.1093/jamiaopen/ooaa041] [Medline: 33215078]
9. Fitzer K, Haeuslschmid R, Blasini R, et al. Patient recruitment system for clinical trials: mixed methods study about requirements at ten university hospitals. JMIR Med Inform 2022 Apr 20;10(4):e28696. [doi: 10.2196/28696] [Medline: 35442203]
10. Cuggia M, Besana P, Glasspool D. Comparing semi-automatic systems for recruitment of patients to clinical trials. Int J Med Inform 2011 Jun;80(6):371-388. [doi: 10.1016/j.ijmedinf.2011.02.003] [Medline: 21459664]
11. Obeid JS, Beskow LM, Rape M, et al. A survey of practices for the use of electronic health records to support research recruitment. J Clin Transl Sci 2017 Aug;1(4):246-252. [doi: 10.1017/cts.2017.301] [Medline: 29657859]
12. Darmawan I, Bakker C, Brockman TA, Patten CA, Eder M. The role of social media in enhancing clinical trial recruitment: scoping review. J Med Internet Res 2020 Oct 26;22(10):e22810. [doi: 10.2196/22810] [Medline: 33104015]
13. Sampson R, Shapiro S, He W, et al. An integrated approach to improve clinical trial efficiency: linking a clinical trial management system into the Research Integrated Network of Systems. J Clin Transl Sci 2022;6(1):e63. [doi: 10.1017/cts.2022.382] [Medline: 35720964]
14. Darko EM, Kleib M, Olson J. Social media use for research participant recruitment: integrative literature review. J Med Internet Res 2022 Aug 4;24(8):e38015. [doi: 10.2196/38015] [Medline: 35925655]
15. Köpcke F, Prokosch HU. Employing computers for the recruitment into clinical trials: a comprehensive systematic review. J Med Internet Res 2014 Jul 1;16(7):e161. [doi: 10.2196/jmir.3446] [Medline: 24985568]
16. Gupta A, Calfas KJ, Marshall SJ, et al. Clinical trial management of participant recruitment, enrollment, engagement, and retention in the SMART study using a marketing and information technology (MARKIT) model. Contemp Clin Trials 2015 May;42:185-195. [doi: 10.1016/j.cct.2015.04.002] [Medline: 25866383]
17. Richesson RL, Krischer J. Data standards in clinical research: gaps, overlaps, challenges and future directions. J Am Med Inform Assoc 2007;14(6):687-696. [doi: 10.1197/jamia.M2470] [Medline: 17712081]

XSL•FO
RenderX

18. Akers L, Gordon JS. Using Facebook for large-scale online randomized clinical trial recruitment: effective advertising strategies. J Med Internet Res 2018 Nov 8;20(11):e290. [doi: 10.2196/jmir.9372] [Medline: 30409765]
19. Muldowney LA, Hynes SM, Oglesby M, Dwyer CP. A qualitative exploration of advantages and disadvantages to using technology in the process of randomised controlled trial recruitment. HRB Open Res 2023;6:56. [doi: 10.12688/hrbopenres.13776.2] [Medline: 39211398]
20. Rabiei R, Almasi S. Requirements and challenges of hospital dashboards: a systematic literature review. BMC Med Inform Decis Mak 2022 Nov 8;22(1):287. [doi: 10.1186/s12911-022-02037-8] [Medline: 36348339]
21. Afroz MA, Schwarber G, Bhuiyan MAN. Risk-based centralized data monitoring of clinical trials at the time of COVID-19 pandemic. Contemp Clin Trials 2021 May;104:106368. [doi: 10.1016/j.cct.2021.106368] [Medline: 33775899]
22. Rave EDC. Medidata. 2024. URL: https://www.medidata.com/en/clinical-trial-products/clinical-data-management/edc-systems/ [accessed 2024-09-25]
23. Veeva Systems. 2024. URL: https://www.veeva.com/products/clinical-data-management/ [accessed 2024-09-25]
24. Ennov. 2024. URL: https://en.ennov.com/solutions/clinical/ [accessed 2024-09-25]
25. Viedoc. 2024. URL: https://www.viedoc.com/campaign/eclinical-software/ [accessed 2024-09-25]
26. RealTime eClinical Solutions. 2024. URL: https://realtime-eclinical.com/ [accessed 2024-09-25]
27. Clinevo Technologies. 2024. URL: https://www.clinevotech.com/ [accessed 2024-09-25]
28. Castor. 2024. URL: https://www.castoredc.com/ [accessed 2024-09-25]
29. Ripple Science. 2024. URL: https://www.ripplescience.com/ [accessed 2024-09-25]
30. OpenClinica. 2024. URL: https://www.openclinica.com/ [accessed 2024-09-25]
31. REDCap. URL: https://www.project-redcap.org/ [accessed 2025-02-06]
32. Gardner L, Bylund P, Robbins S, et al. Agile monitoring dashboard for clinical research studies. Trials 2024 Nov 29;25(1):802. [doi: 10.1186/s13063-024-08646-0] [Medline: 39609928]
33. Boustani M, Alder CA, Solid CA. Agile implementation: a blueprint for implementing evidence-based healthcare solutions. J Am Geriatr Soc 2018 Jul;66(7):1372-1376. [doi: 10.1111/jgs.15283] [Medline: 29513360]
34. Boustani M, Azar J, Solid CA. Agile Implementation: A Model for Implementing Evidence-Based Healthcare Solutions into Real-World Practice to Achieve Sustainable Change: Morgan James Publishing; 2020.
35. Holden RJ, Boustani MA, Azar J. Agile Innovation to transform healthcare: innovating in complex adaptive systems is an everyday process, not a light bulb event. BMJ Innov 2021 Apr;7(2):499-505. [doi: 10.1136/bmjinnov-2020-000574]
36. Mehta J, Long E, Bynagari V, et al. Creating demand for unmet needs: agile storytelling. Front Health Serv 2024;4:1376695. [doi: 10.3389/frhs.2024.1376695] [Medline: 39544454]
37. Holden RJ, Boustani M. The value of an ‘agile’ mindset in times of crisis. Modern Healthcare. 2020 May 11. URL: https://www.modernhealthcare.com/opinion-editorial/value-agile-mindset-times-crisis [accessed 2025-02-06]
38. Bevan N, Carter J, Harker S. ISO 9241-11 revised: what have we learnt about usability since 1998? In: Kurosu M, editor. Human-Computer Interaction: Design and Evaluation. HCI 2015. Lecture Notes in Computer Science(), Vol 9169: Springer; 2015. [doi: 10.1007/978-3-319-20901-2_13]
39. Clemmensen T, Hertzum M, Hornbæk K, Shi Q, Yammiyavar P. Cultural cognition in the thinking-aloud method for usability evaluation. In: ICIS 2008 Proceedings: Association for Information Systems; 2008:189.
40. van Someren M, Barnard YF, Sandberg J. The Think Aloud Method: A Practical Approach to Modelling Cognitive Processes: Academic Press; 1994.
41. Brooke J. SUS: a 'quick and dirty' usability scale. In: Usability Evaluation In Industry: CRC Press; 1996.
42. Reichheld FF. The one number you need to grow. Harv Bus Rev 2003 Dec;81(12):46-54. [Medline: 14712543]
43. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123 [FREE Full text]
44. Clarke V, Braun V. Thematic analysis. J Posit Psychol 2017 May 4;12(3):297-298. [doi: 10.1080/17439760.2016.1262613]

Abbreviations

CHIIS: Center for Health Innovation and Implementation Science
CTMS: clinical trial management system
CTRSS: clinical trial recruitment support system
EHR: electronic health record
NPS: Net Promoter Score
REDCap: Research Electronic Data Capture; Vanderbilt University
SUS: System Usability Scale

XSL•FO
RenderX

Edited by A Mavragani; submitted 20.09.24; peer-reviewed by PV Asaro, S Jimenez-Jorge; revised version received 20.01.25; accepted 22.01.25; published 24.02.25.

Please cite as:
Gardner LL, Raeisian Parvari P, Seidman M, Holden RJ, Fowler NR, Zarzaur BL, Summanwar D, Barboi C, Boustani M
Improving the User Interface and Guiding the Development of Effective Training Material for a Clinical Research Recruitment and Retention Dashboard: Usability Testing Study
JMIR Form Res 2025;9:e66718
URL: https://formative.jmir.org/2025/1/e66718
doi:10.2196/66718

© Leah Leslie Gardner, Pezhman Raeisian Parvari, Mark Seidman, Richard J Holden, Nicole R Fowler, Ben L Zarzaur, Diana Summanwar, Cristina Barboi, Malaz Boustani. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring the Relationship Between Smartphone GPS Patterns and Quality of Life in Patients With Advanced Cancer and Their Family Caregivers: Longitudinal Study

Kyungmi Lee[1], PhD, RN; Andres Azuero[2], MBA, PhD; Sally Engler[2], MPH; Sidharth Kumar[3], PhD; Frank Puga[2], PhD; Alexi A Wright[4], MPH, MD; Arif Kamal[5], MHS, MBA, MD; Christine S Ritchie[6], MSPH, MD; George Demiris[7], FACMI, PhD; Marie A Bakitas[2,8,9], NP-C, DNSc; J Nicholas Odom[2,8,9], PhD, RN

[1]Frances Payne Bolton School of Nursing, Case Western Reserve University, Cleveland, OH, United States

[2]School of Nursing, University of Alabama at Birmingham, 1720 2nd Avenue South, Birmingham, AL, United States

[3]Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, United States

[4]Department of Medical Oncology, Harvard Medical School, Dana-Farber Cancer Institute, Boston, MA, United States

[5]Department of Medicine, Duke University, Durham, NC, United States

[6]Mongan Institute Center for Aging and Serious Illness, Division of Palliative Care and Geriatric Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, United States

[7]School of Nursing & Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[8]Division of Geriatrics, Gerontology, and Palliative Care, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

[9]Center for Palliative and Supportive Care, University of Alabama at Birmingham, Birmingham, AL, United States

Corresponding Author:
J Nicholas Odom, PhD, RN
School of Nursing, University of Alabama at Birmingham, 1720 2nd Avenue South, Birmingham, AL, United States

Abstract

Background: Patients with advanced cancer and their family caregivers often experience poor quality of life (QOL). Self-report measures are commonly used to quantify QOL of family caregivers but may have limitations such as recall bias and social desirability bias. Variables derived from passively obtained smartphone GPS data are a novel approach to measuring QOL that may overcome these limitations and enable detection of early signs of mental and physical health (PH) deterioration.

Objective: This study explored the feasibility of a digital phenotyping approach by assessing participant adherence and examining correlations between smartphone GPS data and QOL levels among family caregivers and patients with advanced cancer.

Methods: This was a secondary analysis involving 7 family caregivers and 4 patients with advanced cancer that assessed correlations between GPS sensor data captured by a personally owned smartphone and QOL self-report measures over 12 weeks through linear correlation coefficients. QOL as measured by the Patient-Reported Outcomes Measurement Information System (PROMIS) Global Health 10 was collected at baseline, 6, and 12 weeks. Using a Beiwe smartphone app, GPS data were collected and processed into variables including total distance, time spent at home, transition time, and number of significant locations.

Results: The study identified relevant temporal correlations between QOL and smartphone GPS data across specific time periods. For instance, in terms of PH, associations were observed with the total distance traveled (12 and 13 wk, with r ranging 0.37 to 0.38), time spent at home (−4 to −2 wk, with r ranging from −0.41 to −0.49), and transition time (−4 to −2 wk, with r ranging −0.38 to −0.47).

Conclusions: This research offers insights into using passively obtained smartphone GPS data as a novel approach for assessing and monitoring QOL among family caregivers and patients with advanced cancer, presenting potential advantages over traditional self-report measures. The observed correlations underscore the potential of this method to detect early signs of deteriorating mental health and PH, providing opportunities for timely intervention and support.

(JMIR Form Res 2025;9:e59161) doi:10.2196/59161

KEYWORDS

cancer; digital phenotyping; global positioning system; quality of life; smartphone; mobile phone; family caregiver

XSL•FO
RenderX

Introduction

Each year, an estimated 1.96 million individuals are expected to receive a cancer diagnosis in 2023 [1]. Projections indicate a near 25% increase in cancer diagnoses by 2032 when it is estimated that 22.5 million US individuals will have a history of cancer [2]. With cancer care transitioning increasingly to home settings, the impact of cancer will increasingly stretch beyond the individual diagnosed to those family and friend caregivers who provide support [3]. Consequently, family caregivers have become indispensable frontline clinicians for patients, especially when their cancer progresses to advanced stages [4]. Both patients and their family caregivers face an array of challenges, from logistical and practical challenges to managing treatments and symptoms to handling a range of strong emotions, including anxiety and depression [5]. Hence, it is not surprising that patients and their caregivers often experience diminished quality of life (QOL) [6-8], which cancer care clinicians aim to assess and optimize.

Traditionally, assessing QOL has primarily relied on patient-reported outcome measures (PROMs) using paper-and-pencil questionnaires, online surveys, or verbal self-report to a data collector reading from a list of items [9,10]. However, PROMs have several notable limitations [11,12]. They can be time consuming for both patients and clinicians and interrupt daily activities. Furthermore, they typically entail retrospective recall which can be subject to distorting biases [13,14] and may elicit socially desirable responses, such as over-reporting better health or greater well-being than actually experienced [15-17].

To address the limitations of a PROM approach to assessing patient and caregiver QOL states, digital phenotyping has emerged as a new and potentially innovative approach. Defined as the "moment-by-moment quantification of the individual-level human phenotype in situ using data from personal digital devices" [18], this method involves passive data collection from ubiquitously owned devices such as smartphones, now owned by approximately 85% of US individuals [19]. By using smartphone sensor data, digital phenotyping models changes in individuals' behaviors, including mobility (eg, GPS data), sociability (eg, text message and telephone logs), and sleep patterns (eg, accelerometer data and screen activity logs) [10].

Physical activity has been shown to improve both the physical and mental well-being of patients with cancer and their caregivers [20,21]. Recent studies have demonstrated that engaging in regular physical activity reduces the risk of cancer recurrence and alleviates symptoms of anxiety and depression, which in turn positively influence QOL [22,23]. These benefits are particularly important for caregivers, who often experience high levels of stress, as physical activity can improve their mental health (MH) and overall well-being [24].

GPS data has been shown to correlate with measures assessing mood, anxiety, and depression [25]. Better mood, lower anxiety, and reduced depression were associated with spending more time at locations of friends and family members. Conversely, more time spent at health care facilities was linked to poorer mood, higher anxiety, and increased depression [25]. Furthermore, GPS data were examined in comparing recovery metrics between mastectomy and breast-conserving surgery in patients with breast cancer [26]. The results indicated that patients undergoing mastectomy spent more time at home in the initial weeks postsurgery, whereas no significant differences were observed after 12 weeks. Despite a growing body of research that highlights the relationship between individual GPS-derived mobility behaviors and MH and well-being [27,28], a digital phenotyping approach to QOL assessment has yet to be explored in an advanced cancer context.

This study aimed to explore the temporal relationship between smartphone GPS sensor features and participants' physical health (PH) and MH over a 12-week period. We hypothesized that more consistent and stable GPS-derived mobility features would correlate with better patient and caregiver PROM reports of QOL.

Methods

Study Design

This was a secondary analysis of data collected from patient and caregiver participants of a larger, ongoing prospective longitudinal observational study [29]. This analysis used data collected over 12 weeks from individuals enrolled between September 2021 and August 2022 at an academic medical center in the southeastern United States as part of a larger research project investigating patterns in passively collected smartphone behavioral data, including GPS and accelerometer.

Ethical Considerations

This study is a secondary analysis of deidentified data collected from a prior study, which was approved by the University of Alabama at Birmingham Institutional Review Board (IRB; 300008890). The IRB determined that this secondary analysis does not constitute Human Subjects Research and is exempt from further IRB review. Details regarding the informed consent process and participant compensation in the original study are described in the Procedures section. For this secondary analysis, no direct participant interaction occurred, and the dataset was fully deidentified before access. Therefore, no additional informed consent or participant compensation was required.

Sample

We included 7 family caregivers and 4 patients with advanced cancer who were recruited as of August 4, 2022. Inclusion criteria for both caregivers and patients were (1) aged 18 years or older, (2) owned a personal smartphone (Android or iOS) and had the ability to download and run the study app, and (3) had proficiency in the English language (spoken and written). In addition, enrolled caregivers provided regular support to a patient with advanced cancer, while patients had to be diagnosed with metastatic, recurrent, or progressive stage III or IV cancer.

Procedures

Once potential participants were identified, they received an invitation letter containing the study description and a timeline of study activities through mail. Subsequently, participants were contacted by telephone to confirm their interest. Verbal consent

XSL•FO
RenderX

was obtained during these calls, and participants were provided with an informed consent document, baseline questionnaires, and a participant guide. After completing the baseline questionnaire, research coordinators assisted participants in downloading and installing the Beiwe app on their personally owned smartphone, which passively collected GPS data. Patient-reported questionnaires were completed through various modalities, including electronic completion through REDCap (Research Electronic Data Capture; Vanderbilt University), mailed paper-and-pencil surveys, or phone-based data collection. Both caregiver and patient participants received US $25 for phone usage and an additional US $25 for each completed questionnaire every 6 weeks.

Beiwe Research Platform

For this research, we used the Beiwe Research Platform, which has a primary focus on developing statistical and computational tools to extract biomedical and clinical insights from smartphone data [30]. The Beiwe app, developed by Torous and colleagues [30], passively collects smartphone GPS data and securely synchronizes it over Wi-Fi to a cloud-based server managed by Beiwe on Amazon Web Services. The platform includes a study portal, applications for both Android and iOS, secure data storage on Amazon Web Services, and data analysis tools and code [30]. To prioritize privacy, passive data undergo encryption both during transit and storage, and indirect identifiers, such as telephone numbers and IP addresses, are hashed, a process that converts the original data into a fixed-size string of characters to enhance the security and privacy of the data [30]. Data linkage is exclusively tied to an 8-character Beiwe participant ID to ensure confidentiality [30].

Data Collection

Demographics

Collected demographic information included age, gender, race or ethnicity, religion, marital status, household size, employment and education levels, relationship status, caregiving duration, and caregiving frequency.

QOL Data

Health-related QOL was assessed using the Patient-Reported Outcomes Measurement Information System (PROMIS) Global-10 questionnaire at baseline, 6 weeks, and 12 weeks. This questionnaire assesses respondents on their perception of overall health, pain, fatigue, social health, MH, and PH [9,31]. Two subscale scores are generated for PH (α=.86) and MH (α=.81), respectively [32]. Lower scores (MH <40, PH <42) indicated poor MH and PH [33], and are associated with higher risk of future health care use [34,35].

GPS Data

We established a data-quality dataset indicating the "valid" GPS hours for each participant per day, enabling the identification and exclusion of individuals with insufficient data during analyses [36]. Our threshold for valid GPS data was set at 15 hours per day, considering that participants might sleep for 9 hours during which GPS records may not be transmitted [37]. Missing location data resulting from issues such as phone battery depletion or power-off states were excluded cases. We determined the number of days with a minimum of 15 hours of recorded GPS data for each user in our dataset. Subsequently, days meeting this hourly cutoff were extrapolated to 24 hours. Additionally, for week-level analyses, participants with fewer than 3 valid days per week were excluded from our analysis [36].

Data Analysis and Feature Extraction From GPS Data

Overview of Feature Extraction

For data analysis, we used Python (version 3.9.0; Python Software Foundation) in Jupyter Notebooks and Google Collaboratory to preprocess the GPS data and extract mobility variables, including total distance, time spent at home, transition time, and number of significant locations.

Total Distance

This variable measured the total distance covered by each participant using the Haversine distance formula, which calculates the angular distance between 2 points on the Earth's surface [38].

Time Spent at Home

The home cluster for each participant was identified based on two heuristics: (1) it was among the top 3 most visited clusters, and (2) it was the cluster most visited between 12 AM and 6 AM [39]. The area of the participant's home was assumed to be in a circular area with a 2000 square feet area from the estimated latitude and longitude of the house location. The time spent within this 2000 square feet area was calculated for each day.

Transition Time

Location data points were categorized as either stationary (eg, working in an office) or transitional (eg, walking on the street) based on movement speed [39]. The time derivative of each data point was used to calculate the movement speed, and a threshold speed of 1 meter per second (m/s) was applied to distinguish between transitional (speed>1 m/s) and stationary states (speed<1 m/s) [39]. The time spent at speeds greater than 1 m/s was calculated for each day.

Number of Significant Locations

This feature represents the total number of clusters found using a machine learning clustering algorithm [40]. The algorithm, called adaptive k-means, classified data points into specific groups based on their locations during stationary states [41]. The goal was to identify places where participants spent most of their time, such as homes, workplaces, and parks. The algorithm used the elbow method to determine the optimal number of clusters, which helps minimize the overall distances of data points to the centers of their respective clusters [42].The number of significant locations visited was then calculated for each day.

Overview of Data Analysis

Statistical Methods Used

Data analysis was conducted using the R programming language (version 4.2.1, R Foundation), in RStudio (RStudio, PBC). Descriptive statistics were computed for variables of interest,

XSL•FO
RenderX

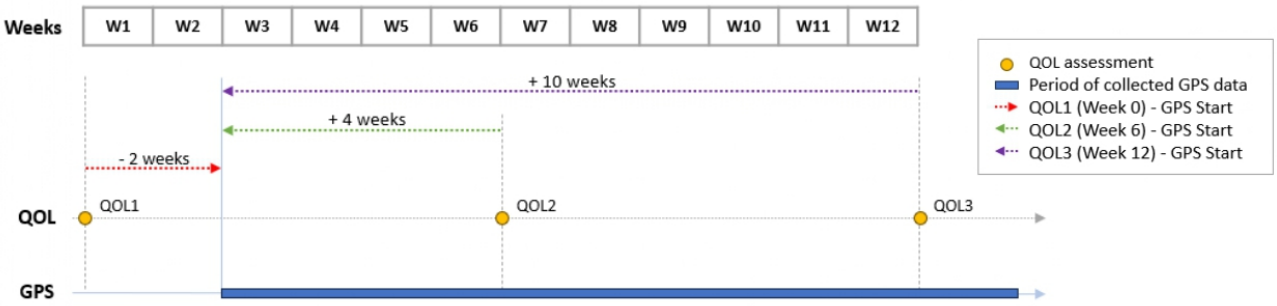
and patterns of missing data were explored. PROMIS Global-10 mental and PH T-scores were used as measures of QOL. The association between GPS location features and patient-reported QOL was assessed using linear correlation coefficients. Temporal association between QOL assessments and GPS data were calculated weekly, using repeated-measures correlations with 1000 cluster-bootstrap resamples [43] to estimate 95% CIs. Cluster bootstrapping [44] was applied to handle multiple observations per participant.

Temporal Correlations

As the QOL measurements did not necessarily align with weeks 0, 6, and 12 of smartphone data, we calculated the time difference between each patient-reported QOL assessment and daily smartphone GPS data on a weekly basis. Relative to the week of QOL assessment collection, time differences between smartphone data and QOL assessments ranged from −15 weeks to +14 weeks. The number of observations available for estimating a correlation at each temporal difference was computed. We considered data for temporal differences with more than 10 available observations per week, resulting in analysis for −6 to +13 weeks, excluding +10 weeks (Figure 1).

XSL•FO
RenderX

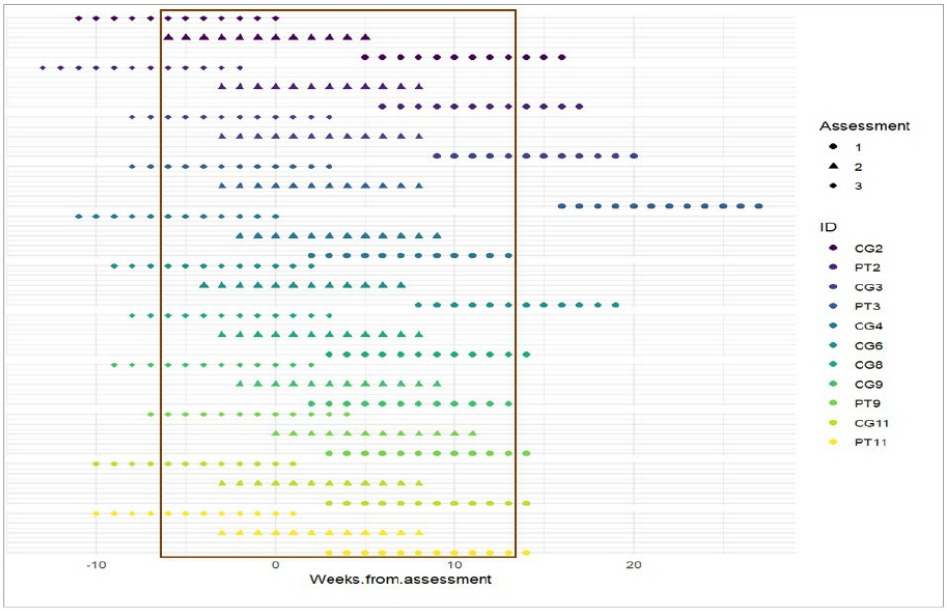
Figure 1. Time differences between patient-reported QOL assessments and daily smartphone GPS data were calculated on a weekly basis, as QOL measurements did not align with the smartphone data collection weeks of 0, 6, and 12. (A) Weekly time differences between QOL assessment and smartphone GPS data. (B) Example of weekly time differences in QOL and GPS data. (C) Data from temporal differences: –6 to +13 weeks (excluding +10 weeks) with >10 observations per week. QOL: quality of life.



(A) Weekly time differences between QOL assessment and smartphone GPS data.

		QOL Assessment and Week of it		
		QOL1	QOL2	QOL3
		-2 weeks	+4 weeks	+10 weeks
Participant	GPS data week	Weeks from after QOL assessment		
Participant 1	1	3	-3	-9
Participant 1	2	4	-2	-8
Participant 1	3	5	-1	-7
Participant 1	4	6	0	-6
Participant 1	5	7	1	-5
Participant 1	6	8	2	-4
Participant 1	7	9	3	-3
Participant 1	8	10	4	-2
Participant 1	9	11	5	-1
Participant 1	10	12	6	0
Participant 1	11	13	7	1
Participant 1	12	14	8	2

(B) Example of weekly time differences in QOL and GPS data.



(C) Data from temporal differences: -6 to +13 weeks (excluding +10 weeks) with >10 observations per week.

Results

Demographics of Participants

The study included 11 participants (7 caregivers and 4 patients). Caregivers had a mean age of 47.7 (SD 15.9) years, while patients had a mean age of 55 (SD 3.7) years. Most participants identified as White (10/11, 90%) and non-Hispanic or non-Latino (11/11, 100%). Most participants (9/11, 82%) identified as Protestant, were married (10/11, 90%), employed (5/11, 45%), and had a high school diploma or General

XSL•FO
RenderX

Educational Development or higher (10/11, 90%). Among caregivers, 71% (5/7) cared for a spouse, and their average caregiving duration was 557 (SD 96) days. Most caregivers (6/7, 86%) lived with the patients. Regarding insurance, 75% (3/4) of patients had private or commercial medical insurance, and 25% (1/4) had military insurance. Furthermore, 75% (3/4) of patients resided in rural areas (Table 1).

XSL•FO
RenderX

Table . Demographics of participants.

Variables	Caregivers (n=7)	Patients (n=4)	All (N=11)
Age (years), mean (SD)	47.71 (15.86)	55 (3.65)	50 (13)
Gender, n (%)			
Male	3 (43)	2 (50)	5 (46)
Female	4 (57)	2 (50)	6 (55)
Race, n (%)			
Black or African American	1 (14)	0 (0)	1 (9)
White	6 (86)	4 (100)	10 (90)
Hispanic or Latino, n (%)			
Yes	0 (0)	0 (0)	0 (0)
No	7 (100)	4 (100)	11 (100)
Religion, n (%)			
Protestant	6 (86)	3 (75)	9 (82)
Catholic	0 (0)	1 (25)	1 (9)
None	1 (14)	0 (0)	1 (9)
Marital status, n (%)			
Never married	1 (14)	0 (0)	1 (9)
Married	6 (86)	4 (100)	10 (90)
Employment status, n (%)			
Full time	4 (57)	0 (0)	4 (36)
Part time	0 (0)	1 (25)	1 (9)
Unemployed due to disability or illness	1 (14)	1 (25)	2 (18)
Homemaker	1 (14)	1 (25)	2 (18)
Other	1 (14)	1 (25)	2 (18)
Education, n (%)			
≤8th grade	1 (14)	0 (0)	1 (9)
Some high school	0 (0)	0 (0)	0 (0)
High school graduate or General Educational Development	2 (29)	2 (50)	4 (36)
Some college or technical school	2 (29)	0 (0)	2 (18)
College graduate (bachelor's degree)	2 (29)	2 (50)	4 (36)
Relationship with patient, n (%)			
Spouse	5 (71)	—[a]	—
Parent	1 (14)	—	—
Other	1 (14)	—	—
Lives with patient, n (%)			
Yes	6 (86)	—	—
No	1 (14)	—	—
Length of caregiving (day), mean (SD)	556.57 (96)	—	—
The number of days of care per week, n (%)			

XSL•FO
RenderX

Variables	Caregivers (n=7)	Patients (n=4)	All (N=11)
Every day of the week (7 days)	3 (43)	—	—
2 to 3 days a week	2 (29)	—	—
1 day a week or less	2 (29)	—	—
Hours of care per day, n (%)			
5 to 6 hours a day	1 (14)	—	—
3 to 4 hours a day	1 (14)	—	—
1 to 2 hours a day	5 (71)	—	—
Medical insurance, n (%)			
Private or commercial	—	3 (75)	—
Military program	—	1 (25)	—
Residence, n (%)			
Urban	—	0 (0)	—
Suburban	—	1 (25)	—
Rural	—	3 (75)	—

[a]Not applicable.

QOL Assessment Results

PH Findings

At baseline, the mean PH score across all participants was 47.1 (SD 6.7). Caregivers had a slightly lower mean PH score (mean 45.4, SD 4.6) compared to patients (mean 50.1, SD 9.4). Around 18% (2/11) of participants reported baseline PH T-scores below 42, including 1 caregiver and 1 patient. By the 6-week mark, the average PH score had increased to 49 (SD 9.6), with caregivers reporting a mean score of 50.2 (SD 10.3) and patients a mean score of 47.1 (SD 9.2). Similarly, around 18% (2/11) of participants recorded 6-week PH T-scores below 42, including 1 caregiver and 1 patient. At the 12-week assessment, the overall mean PH score slightly decreased to 46.6 (SD 6.8), with caregivers and patients reporting mean scores of 46.8 (SD 6.1) and 46.3 (SD 9.1), respectively. This time, around 27% (3/11) of participants, including 1 caregiver and 2 patients, had 12-week PH T-scores below 42 (see Multimedia Appendix 1 for participant-reported QOL results).

MH Findings

At baseline, the mean MH score across all participants was 50.7 (SD 8.6). Caregivers had a slightly lower mean MH score (mean 47.9, SD 7.7) compared to patients (mean 55.5, SD 8.8). One caregiver (9%) reported a baseline MH T-score below 40. At the 6-week evaluation, the mean MH score across all participants slightly increased to 51.5 (SD 7.7), with caregivers and patients reporting mean scores of 51.1 (SD 9) and 52.2 (SD 6), respectively. By the 12-week mark, the overall mean MH score further improved to 52.8 (SD 7.2), with caregivers at a mean score of 52.2 (SD 7.3) and patients at 53.9 (SD 8.1). One caregiver (9%) reported a 12-week MH T-score below 40 (Multimedia Appendix 1).

Daily Smartphone GPS Data Weekly Averages

Over the 12-week period, the total distance variable ranged from a minimum of 102.2 (SD 66.5) km to a maximum of 186.7 (SD 153.2) km. The time spent at home variable ranged from a minimum of 437.2 (SD 388.5) minutes to a maximum of 573.3 (SD 271.1) minutes. The transition time variable ranged from a minimum of 800.4 (SD 225.3) minutes to a maximum of 903.4 (SD 109.4) minutes. The number of significant locations variable ranged from a minimum of 4.2 (SD 1.4) to a maximum of 5.1 (SD 1.6; see Multimedia Appendix 2 for details on daily smartphone GPS location features averaged by week).

Correlation Between PH and Daily Smartphone GPS Data Over a 12-Week Period

Table 2 illustrates the temporal correlation between PH and weekly GPS location features. Significant correlations, with a value of +0.3 or higher, were observed between total distance and PH at 3 weeks before and during weeks 12 and 13 after the QOL assessment. Positive correlations at 3 weeks before and during weeks 12 and 13 after the QOL assessment imply that increased distances correlate with higher PH. Correlations of −0.3 or higher were found between time spent at home and PH at 2, 3, and 4 weeks before the QOL assessment, as well as at the 3 weeks after the assessment. These negative correlations suggest a potential association between increased time spent at home and lower PH. Similarly, correlations of −0.3 or higher were found between transition time and PH at 2, 3, and 4 weeks before the QOL assessment. These negative correlations also suggest a potential association between increased transition time and lower PH. Moreover, correlations of ±0.3 or higher between the number of significant locations and PH were significant at 3 weeks before and at weeks 1, 3, and 4 after the QOL assessment. Negative correlations at week 3 after the QOL assessment suggest that as the number of places visited increases, PH tends to decline, while positive correlations at

XSL•FO
RenderX

week 3 before and at weeks 1 and 4 after the QOL assessment indicate a potential link to higher PH (Figure 2).

Table . Correlation between physical health and daily smartphone GPS features, including total distance, time spent at home, transition time, and number of significant locations, averaged by week and using 95% cluster-bootstrap CIs over a 12-week period.

From QOL[a], weeks[b]	Observations, n	Participants, n	Total distance mean, *r*[c] (95% CI)	Time spent at home mean, *r* (95% CI)	Transition time mean, *r* (95% CI)	Number of significant locations mean, *r* (95% CI)
−6	11	10	0.06 (−0.42 to 0.74)	−0.14 (−0.52 to 0.41)	0.14 (−0.31 to 0.63)	0.11 (−0.34 to 0.58)
−5	11	10	0.14 (−0.47 to 0.54)	0.02 (−0.26 to 0.38)	−0.05 (−0.56 to 0.35)	0.28 (−0.14 to 0.76)
−4	12	10	0.12 (−0.37 to 0.61)	−0.49 (−0.71 to −0.05)	−0.44 (−0.84 to 0.06)	−0.18 (−0.48 to 0.35)
−3	16	11	0.39 (0.01 to 0.65)	−0.41 (−0.69 to −0.02)	−0.38 (−0.74 to 0.04)	0.34 (−0.06 to 0.63)
−2	16	10	0.12 (−0.44 to 0.63)	−0.46 (−0.80 to 0.02)	−0.47 (−0.80 to 0.01)	0.21 (−0.23 to 0.66)
−1	19	11	0.17 (−0.40 to 0.76)	0.08 (−0.26 to 0.61)	0.20 (−0.15 to 0.57)	0.14 (−0.20 to 0.57)
0	19	11	0.08 (−0.54 to 0.53)	−0.10 (−0.47 to 0.24)	−0.06 (−0.41 to 0.18)	−0.02 (−0.28 to 0.29)
1	18	11	0.05 (−0.44 to 0.62)	−0.19 (−0.45 to 0.10)	−0.16 (−0.45 to 0.13)	0.39 (0.11 to 0.66)
2	19	11	0.26 (−0.54 to 0.80)	−0.12 (−0.47 to 0.33)	−0.23 (−0.46 to −0.03)	0.07 (−0.26 to 0.47)
3	20	11	0.20 (−0.21 to 0.60)	−0.41 (−0.65 to −0.20)	−0.16 (−0.66 to 0.31)	−0.31 (−0.55 to 0.43)
4	16	11	0.21 (−0.16 to 0.71)	−0.18 (−0.50 to 0.17)	0.14 (−0.31 to 0.63)	0.46 (0.13 to 0.65)
5	17	11	−0.09 (−0.66 to 0.36)	0.08 (−0.22 to 0.49)	−0.14 (−0.59 to 0.30)	0.17 (−0.41 to 0.63)
6	18	11	0.28 (−0.51 to 0.80)	0.08 (−0.27 to 0.51)	0.15 (−0.35 to 0.55)	0.09 (−0.14 to 0.60)
7	16	11	0.08 (−0.36 to 0.66)	−0.17 (−0.57 to 0.34)	−0.04 (−0.49 to 0.51)	0.19 (−0.25 to 0.57)
8	17	11	0.10 (−0.61 to 0.50)	−0.20 (−0.64 to 0.19)	−0.28 (−0.68 to 0.04)	0.23 (−0.23 to 0.54)
9	12	9	0.25 (−0.40 to 0.63)	−0.17 (−0.64 to 0.36)	0.12 (−0.37 to 0.51)	−0.20 (−0.49 to 0.29)
11	10	10	−0.01 (−0.88 to 0.52)	0.03 (−0.39 to 0.51)	−0.12 (−0.68 to 0.29)	0.18 (−0.17 to 0.73)
12	10	10	0.37 (−0.84 to 0.74)	0.21 (−0.18 to 0.65)	0.06 (−0.29 to 0.53)	0.00 (−0.44 to 0.55)
13	10	10	0.38 (−0.57 to 0.77)	0.01 (−0.33 to 0.39)	−0.01 (−0.51 to 0.57)	0.39 (−0.54 to 0.74)

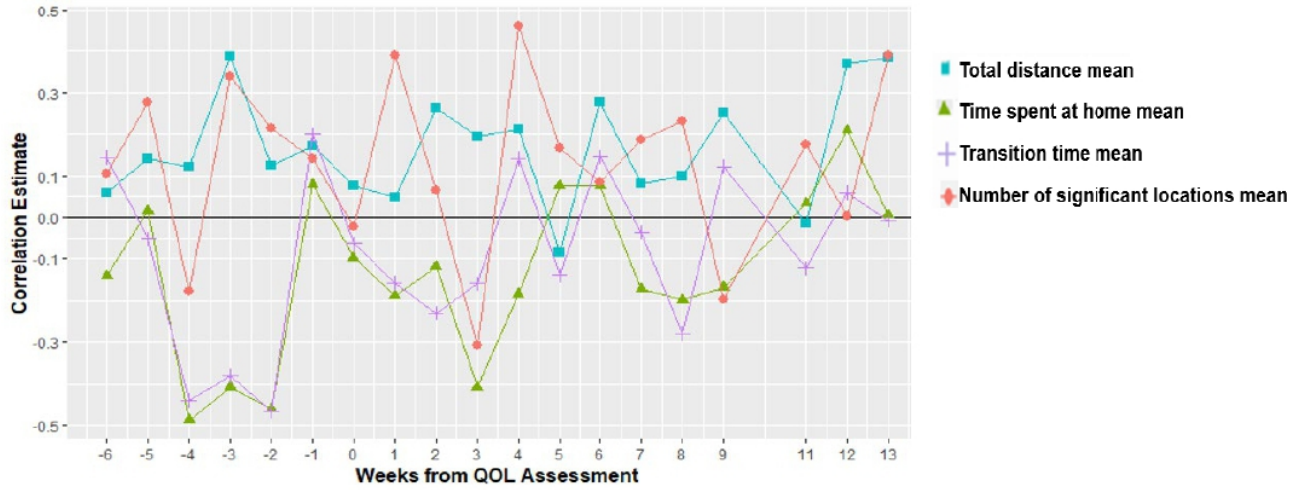
[a]QOL: quality of life.

[b]Weeks from QOL: temporal correlations.

[c]Correlation (absolute value): small, r=0.1; medium, r=0.3; and large, r≥0.5.

XSL•FO
RenderX

Figure 2. Correlation between physical and mental health and daily smartphone GPS features, including total distance, time spent at home, transition time, and number of significant locations averaged by week and using 95% cluster-bootstrap CIs over a 12-week period. (A) Physical health. (B) Mental health. QOL: quality of life.



(A) Physical health.

(B) Mental health.

Correlation Between MH and Daily Smartphone GPS Data Over a 12-Week Period

Table 3 illustrates the temporal correlation between MH and weekly GPS location features. Correlations of −0.3 or higher were notable between time spent at home and MH at weeks 4 and 2 before the QOL assessment, as well as at week 9 after the QOL assessment, suggesting a potential link between increased time spent at home and lower MH. Correlations of ±0.3 or higher were significant between transition time and MH at weeks −4 and −2 before the QOL assessment, as well as at week 4 after the QOL assessment. Negative correlations at weeks −4 and −2 imply that increased transition time correlates with lower MH, while positive correlations at week 4 suggest a potential link between increased transition time and higher MH. Positive correlations of +0.3 or higher were significant between the number of significant locations and MH at 2 weeks before the QOL assessment and at week 13 after the QOL assessment, suggesting a potential link between an increased number of places visited and higher MH (Figure 2).

XSL•FO
RenderX

Table . Correlation between mental health and daily smartphone GPS features, including total distance, time spent at home, transition time, and number of significant locations, averaged by week and using 95% cluster-bootstrap CIs over a 12-week period.

From QOL[a], weeks[b]	Observations, n	Participants, n	Total distance mean, *r*[c] (95% CI)	Time spent at home mean, *r* (95% CI)	Transition time mean, *r* (95% CI)	Number of significant locations mean, *r* (95% CI)
−6	11	10	−0.03 (−0.50 to 0.38)	0.47 (0.07 to 0.77)	0.44 (0.07 to 0.87)	−0.11 (−0.35 to 0.29)
−5	11	10	−0.18 (−0.67 to 0.1)	0.43 (0.13 to 0.78)	0.24 (−0.01 to 0.8)	0.11 (−0.21 to 0.6)
−4	12	10	0.09 (−0.39 to 0.34)	−0.03 (−0.35 to 0.42)	0.05 (−0.48 to 0.59)	−0.39 (−0.75 to 0.02)
−3	16	11	0 (−0.33 to 0.41)	0.02 (−0.23 to 0.39)	0.08 (−0.31 to 0.51)	0.04 (−0.38 to 0.49)
−2	16	10	−0.16 (−0.51 to 0.27)	0.07 (−0.3 to 0.28)	0.16 (−0.15 to 0.37)	−0.2 (−0.56 to 0.16)
−1	19	11	−0.1 (−0.49 to 0.28)	0.3 (−0.08 to 0.62)	0.3 (−0.05 to 0.66)	−0.21 (−0.41 to 0.02)
0	19	11	−0.13 (−0.45 to 0.16)	0.14 (−0.15 to 0.38)	0.08 (−0.32 to 0.45)	−0.18 (−0.55 to 0.18)
1	18	11	−0.28 (−0.6 to 0.04)	0.13 (−0.25 to 0.47)	−0.11 (−0.45 to 0.22)	0 (−0.42 to 0.42)
2	19	11	0.05 (−0.24 to 0.35)	0.19 (−0.13 to 0.48)	−0.02 (−0.47 to 0.44)	−0.09 (−0.46 to 0.35)
3	20	11	0 (−0.4 to 0.41)	0.05 (−0.29 to 0.4)	0.18 (−0.3 to 0.62)	−0.49 (−0.65 to -0.09)
4	16	11	0.13 (−0.28 to 0.54)	0.24 (−0.09 to 0.51)	0.61 (0.26 to 0.84)	0.09 (−0.57 to 0.36)
5	17	11	0.15 (−0.3 to 0.65)	0.24 (−0.04 to 0.45)	0.28 (−0.09 to 0.52)	−0.11 (−0.62 to 0.25)
6	18	11	0.24 (−0.23 to 0.67)	0.22 (0.01 to 0.55)	0.27 (−0.01 to 0.53)	0.05 (−0.36 to 0.57)
7	16	11	0.19 (−0.29 to 0.68)	0.19 (−0.21 to 0.47)	0.31 (0.08 to 0.61)	−0.3 (−0.65 to 0.05)
8	17	11	0.15 (−0.51 to 0.62)	0.14 (−0.38 to 0.5)	0.13 (−0.41 to 0.44)	0.01 (−0.85 to 0.39)
9	12	9	0.33 (0 to 0.83)	−0.2 (−0.71 to 0.58)	0.19 (−0.36 to 0.69)	−0.43 (−0.77 to 0.07)
11	10	10	0.23 (−0.52 to 0.85)	0.39 (−0.23 to 0.82)	0.36 (−0.43 to 0.76)	−0.32 (−0.8 to 0.59)
12	10	10	0.34 (−0.3 to 0.77)	0.46 (−0.05 to 0.86)	0.54 (0.14 to 0.87)	−0.48 (−0.87 to 0.03)
13	10	10	0.33 (−0.15 to 0.94)	0.23 (−0.31 to 0.71)	0.2 (−0.31 to 0.7)	0.33 (−0.9 to 0.77)

[a]QOL: quality of life.

[b]Weeks from QOL: temporal correlations.

[c]Correlation (absolute value): small, r=0.1; medium, r=0.3; and large, r≥0.5.

Discussion

Principal Findings

The findings of this exploratory, secondary analysis demonstrate the potential association between smartphone-derived GPS-based movement features and QOL of patients with advanced cancer and their family caregivers. While not powered for statistical significance, correlations were observed between GPS movement variables and PH and MH during stretches of time before and after PROM assessment. These results warrant continued investigation of the potential of digital phenotyping approaches to passively assess patient and caregiver QOL in the context of cancer. Future analyses could investigate the interrelationship between the movement patterns and QOL of both patients and their caregivers, offering deeper insights into the dynamics between their experiences.

This study extends the existing body of research associating higher movement with higher QOL for patients with cancer and their family caregivers [45]. Notably, significant correlations were observed between PH and the total distance covered in GPS data collected 3 weeks before the QOL assessment. This indicates a potential connection between higher PH and longer distances traveled during this period. Additionally, our study findings align with previous research, indicating a correlation

XSL•FO
RenderX

between PH and time spent at home [45,46]. Our findings suggest that poor PH is associated with increased time spent at home at 2, 3, and 4 weeks before the QOL assessment. This finding is supported by earlier research, which has proposed that an extended duration at home might be indicative of fatigue or impaired motor function [45]. However, our study findings diverge from prior research on the correlations between PH and the transition time [47], or the number of significant locations [48]. While our results suggest that decreased mobility is associated with higher PH, earlier research has proposed that increased mobility is an indicator of recovered PH. These findings imply that the metrics of transition time and number of visited places may not adequately capture how these indices relate to PH. Given the different contexts of the study design, it is not surprising that our results were inconsistent with the findings of a previous study. Future research may, therefore, explore the context of the relationship between PH and GPS measures in further detail, such as clinical populations or specific environmental contexts.

Our study did not reveal significant correlations between MH and the time spent at home or the number of significant locations overall. However, specific weeks within the 12-week period showed noteworthy correlations. For example, a statistically significant correlation emerged at 1, 5, and 6 weeks when GPS data were collected before the QOL assessment, indicating an association between increased time spent at home and MH. Similarly, at 3, 7, 9, 10, 11, 12, and 13 weeks, when GPS data were collected after the QOL assessment, a significant correlation emerged, suggesting a link between the number of significant locations and MH. These findings indicate temporal variability in the relationship between individuals' daily activities, such as time spent at home and various locations, and their MH across distinct phases. While the precise significance of these periodic correlations remains uncertain, it underscores the necessity of accounting for temporal variations in understanding the connections between environmental exposure and MH outcomes. Further investigation is warranted to explore potential contributing factors to these temporal variations and to elucidate the nuanced dynamics of the interplay between daily routines and MH across different time frames.

Moreover, in our exploration of the relationship between MH and movement, we found that the total distance exhibited either no correlation or only a small correlation over time. Existing literature presents mixed findings; however, a significant portion indicates that traveling shorter distances is associated with worse outcomes in terms of emotional and cognitive well-being [49-51]. When exploring the association between MH and transition time, our results displayed mixed outcomes, aligning with previous studies where results on the correlation between MH and movement conflicted [39,52]. Notably, 1 study indicated that individuals with irregular movement patterns—described by irregularity features such as variations in circadian movement and routine index across days—tended to report fewer depressive symptoms and less loneliness compared to those with more regular patterns, hinting at a potential link between irregular movement and improved overall MH [53]. While our findings suggest that distance and movement behaviors may not be directly linked to specific MH aspects, they contribute to the broader understanding of general movement patterns, such as total distance and transition time.

Limitations

This study has several limitations. First, the small sample size introduces uncertainty in correlation estimates and limits the generalizability of the findings. Future research with a larger sample size is needed to validate and extend these findings. Second, the participant sample was recruited from a specific group of patients with advanced cancer and family caregivers attending a particular oncology clinic, raising the potential for selection bias. Future research endeavors should strive for a more geographically and racially diverse sample. Third, the pilot study encountered issues with missing sensor data, leading to the exclusion of weeks with less than 3 days of data and days with less than 10 hours of data per day from the analysis. This limitation narrows the generalizability of our findings for those periods. The exact cause of the missing data is unclear; it could be attributed to several factors, such as technical issues with the app or the participants' devices not being turned on, potentially due to reasons such as the device not being charged. Fourth, GPS data alone does not provide specific information about the type of activity performed outside the home, which limits the interpretation of the data and its correlation with quality-of-life questionnaires. Although we extracted several features from the GPS data to offer insights into participant behavior (eg, total distance traveled and time spent at home), the inability to classify activity types outside the home remains a limitation in the analysis. This limitation could impact the accuracy of the behavioral patterns examined in relation to QOL. Despite these acknowledged limitations, the study suggests the potential integration of digital phenotyping into health research, providing valuable insights into the challenges faced by individuals with complex health conditions and their caregivers in their daily lives.

Conclusion

This study investigated the correlation between the QOL of patients with advanced cancer and family caregivers and smartphone GPS data. The results revealed complex relationships between smartphone GPS features and participants' PH and MH outcomes. Specifically, total distance, time spent at home, transition time, and the number of significant locations showed various correlations with reported QOL. The study contributes to the field of digital phenotyping by shedding light on the intricate dynamic relationships between daily behaviors captured through smartphone GPS data and well-being among family caregivers and patients with advanced cancer. The findings emphasize the importance of considering personalized patterns and contextual factors when interpreting the impact of smartphone sensor data on health outcomes.

XSL•FO
RenderX

Acknowledgments

This study was funded by the Cambia Health Foundation (principal investigator: JNO; no grant number provided). We express our gratitude to the partnering oncologists, clinicians, and staff at the O'Neal Comprehensive Cancer Center at the University of Alabama at Birmingham for their support throughout this study. We also thank James Mapson, Brianna Brakefield, Stephanie Ford, and Charis Smith for their invaluable contributions as research coordinators.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

KL and JNO contributed to this study’s concept and design. KL, AA, and SK performed the data analyses. All authors contributed to the interpretation of the data. KL and JNO drafted this paper, and all authors critically revised it for relevant intellectual content. All authors have read and approved the final version of this paper and take responsibility for their individual contributions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Participant-reported quality of life.
[DOCX File, 19 KB - formative_v9i1e59161_app1.docx]

Multimedia Appendix 2
Daily smartphone GPS location features averaged by week.
[DOCX File, 19 KB - formative_v9i1e59161_app2.docx]

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin 2023 Jan;73(1):17-48. [doi: 10.3322/caac.21763] [Medline: 36633525]
2. Cancer treatment & survivorship facts & figures 2022-2024. American Cancer Society. 2022. URL: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/cancer-treatment-and-survivorship-facts-and-figures/2022-cancer-treatment-and-survivorship-fandf-acs.pdf [accessed 2025-01-28]
3. National Academies of Sciences, Engineering, and Medicine. Families Caring for an Aging America: National Academies Press; 2016.
4. Chua GP, Pang GSY, Yee ACP, et al. Supporting the patients with advanced cancer and their family caregivers: what are their palliative care needs? BMC Cancer 2020 Aug 15;20(1):768. [doi: 10.1186/s12885-020-07239-9] [Medline: 32799834]
5. Badr H. Psychosocial interventions for patients with advanced cancer and their families. Am J Lifestyle Med 2016;10(1):53-63. [doi: 10.1177/1559827614530966] [Medline: 30202258]
6. Given BA, Given CW, Sherwood P. The challenge of quality cancer care for family caregivers. Semin Oncol Nurs 2012 Nov;28(4):205-212. [doi: 10.1016/j.soncn.2012.09.002] [Medline: 23107177]
7. Kim MH, Dunkle RE, Lehning AJ, Shen HW, Feld S, Perone AK. Caregiver stressors and depressive symptoms among older husbands and wives in the United States. J Women Aging 2017;29(6):494-504. [doi: 10.1080/08952841.2016.1223962] [Medline: 27673406]
8. Ramasubbu SK, Pasricha RK, Nath UK, Rawat VS, Das B. Quality of life and factors affecting it in adult cancer patients undergoing cancer chemotherapy in a tertiary care hospital. Cancer Rep (Hoboken) 2021 Apr;4(2):e1312. [doi: 10.1002/cnr2.1312] [Medline: 33295136]
9. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010 Nov;63(11):1179-1194. [doi: 10.1016/j.jclinepi.2010.04.011] [Medline: 20685078]
10. Cote DJ, Barnett I, Onnela JP, Smith TR. Digital phenotyping in patients with spine disease: a novel approach to quantifying mobility and quality of life. World Neurosurg 2019 Jun;126:e241-e249. [doi: 10.1016/j.wneu.2019.01.297] [Medline: 30797933]
11. Faraj SSA, van Hooff ML, Holewijn RM, Polly DW Jr, Haanstra TM, de Kleuver M. Measuring outcomes in adult spinal deformity surgery: a systematic review to identify current strengths, weaknesses and gaps in patient-reported outcome measures. Eur Spine J 2017 Aug;26(8):2084-2093. [doi: 10.1007/s00586-017-5125-4] [Medline: 28534221]

XSL•FO
RenderX

12. Paulsen RT, Bouknaitir JB, Fruensgaard S, Carreon L, Andersen M. Prognostic factors for satisfaction after decompression surgery for lumbar spinal stenosis. Neurosurgery 2018 May 1;82(5):645-651. [doi: 10.1093/neuros/nyx298] [Medline: 28575297]
13. Ben-Zeev D, Young MA. Accuracy of hospitalized depressed patients' and healthy controls' retrospective symptom reports: an experience sampling study. J Nerv Ment Dis 2010 Apr;198(4):280-285. [doi: 10.1097/NMD.0b013e3181d6141f] [Medline: 20386257]
14. Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002;24(3):236-243. [doi: 10.1207/S15324796ABM2403_09] [Medline: 12173681]
15. Gorzelitz J, Peppard PE, Malecki K, Gennuso K, Nieto FJ, Cadmus-Bertram L. Predictors of discordance in self-report versus device-measured physical activity measurement. Ann Epidemiol 2018 Jul;28(7):427-431. [doi: 10.1016/j.annepidem.2018.03.016] [Medline: 29681429]
16. Paulhus DL. Socially desirable responding on self-reports. In: Encyclopedia of Personality and Individual Differences: Springer International Publishing AG 2017; 2017:1-5. [doi: 10.1007/978-3-319-28099-8_1349-1]
17. Dagum P. Digital biomarkers of cognitive function. NPJ Digit Med 2018;1:10. [doi: 10.1038/s41746-018-0018-4] [Medline: 31304295]
18. Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016 Jun;41(7):1691-1696. [doi: 10.1038/npp.2016.7] [Medline: 26818126]
19. Mobile fact sheet. Pew Research Center. 2023. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2025-01-16]
20. Ng SM, Fung MHY, Chan JSM, Chan CHY, Chan CLW. Physical activity, confidence and quality of life among cancer patient-carer dyads. Sports Med Open 2021 Jul 1;7(1):46. [doi: 10.1186/s40798-021-00333-7] [Medline: 34196850]
21. Stout NL, Santa Mina D, Lyons KD, Robb K, Silver JK. A systematic review of rehabilitation and exercise recommendations in oncology guidelines. CA Cancer J Clin 2021 Mar;71(2):149-175. [doi: 10.3322/caac.21639] [Medline: 33107982]
22. Misiąg W, Piszczyk A, Szymańska-Chabowska A, Chabowski M. Physical activity and cancer care-a review. Cancers (Basel) 2022 Aug 27;14(17):4154. [doi: 10.3390/cancers14174154] [Medline: 36077690]
23. Kulchycki M, Halder HR, Askin N, et al. Aerobic physical activity and depression among patients with cancer: a systematic review and meta-analysis. JAMA Netw Open 2024 Oct 1;7(10):e2437964. [doi: 10.1001/jamanetworkopen.2024.37964] [Medline: 39378035]
24. Mahindru A, Patil P, Agrawal V. Role of physical activity on mental health and well-being: a review. Cureus 2023 Jan;15(1):e33475. [doi: 10.7759/cureus.33475] [Medline: 36756008]
25. Cai L, Boukhechba M, Gerber MS, et al. An integrated framework for using mobile sensing to understand response to mobile interventions among breast cancer patients. Smart Health (2014) 2020 Mar;15:100086. [doi: 10.1016/j.smhl.2019.100086]
26. Panda N, Solsky I, Hawrusik B, et al. Smartphone Global Positioning System (GPS) data enhances recovery assessment after breast cancer surgery. Ann Surg Oncol 2021 Feb;28(2):985-994. [doi: 10.1245/s10434-020-09004-5] [Medline: 32812109]
27. Saeb S, Lattie EG, Kording KP, Mohr DC. Mobile phone detection of semantic location and its relationship to depression and anxiety. JMIR Mhealth Uhealth 2017 Aug 10;5(8):e112. [doi: 10.2196/mhealth.7297] [Medline: 28798010]
28. Wang R, Wang W, Aung MH, et al. Predicting symptom trajectories of schizophrenia using mobile sensing. GetMobile Mobile Comp Comm 2018 Sep 5;22(2):32-37. [doi: 10.1145/3276145.3276157]
29. Odom JN, Lee K, Currie ER, et al. Feasibility and Acceptability of Collecting Passive Smartphone Data for Potential Use in Digital Phenotyping Among Family Caregivers and Patients With Advanced Cancer. JCO Clin Cancer Inform 2025 Jan;9:e2400201. [doi: 10.1200/CCI-24-00201] [Medline: 39746166]
30. Torous J, Kiang MV, Lorme J, Onnela JP. New tools for new research in psychiatry: a scalable and customizable platform to empower data driven smartphone research. JMIR Ment Health 2016 May 5;3(2):e16. [doi: 10.2196/mental.5165] [Medline: 27150677]
31. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the Patient-Reported Outcomes Measurement Information System (PROMIS) global items. Qual Life Res 2009 Sep;18(7):873-880. [doi: 10.1007/s11136-009-9496-9] [Medline: 19543809]
32. Hays RD, Schalet BD, Spritzer KL, Cella D. Two-item PROMIS® global physical and mental health scales. J Patient Rep Outcomes 2017;1(1):2. [doi: 10.1186/s41687-017-0003-8] [Medline: 29757325]
33. PROMIS® score cut points. HealthMeasures. 2021. URL: https://www.healthmeasures.net/score-and-interpret/interpret-scores/promis/promis-score-cut-points [accessed 2025-01-16]
34. Blumenthal KJ, Chang Y, Ferris TG, et al. Using a self-reported global health measure to identify patients at high risk for future healthcare utilization. J Gen Intern Med 2017 Aug;32(8):877-882. [doi: 10.1007/s11606-017-4041-y] [Medline: 28341894]
35. Katzan IL, Thompson N, Schuster A, Wisco D, Lapin B. Patient-reported outcomes predict future emergency department visits and hospital admissions in patients with stroke. J Am Heart Assoc 2021 Mar 16;10(6):e018794. [doi: 10.1161/JAHA.120.018794] [Medline: 33666094]

XSL•FO
RenderX

36. Müller SR, Bayer JB, Ross MQ. Analyzing GPS data for psychological research: a tutorial. AMPPS 2022;5(2):25152459221082680. [doi: 10.1177/25152459221082680]
37. Harari GM, Müller SR, Stachl C, et al. Sensing sociability: individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. J Pers Soc Psychol 2020 Jul;119(1):204-228. [doi: 10.1037/pspp0000245] [Medline: 31107054]
38. Sinnott RW. Virtues of the Haversine. Sky Telescope 1984;68(2):158.
39. Saeb S, Zhang M, Karr CJ, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015 Jul 15;17(7):e175. [doi: 10.2196/jmir.4273] [Medline: 26180009]
40. Rasmussen EM. Clustering algorithms. In: Information Retrieval: Data Structures &Amp; Algorithms: Pearson Education; 1992:419-442.
41. Arthur D, Vassilvitskii S. K-means++: the advantages of careful seeding. Presented at: The Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2007; Jan 7-9, 2007; New Orleans, Louisiana URL: https://theory.stanford.edu/~sergei/papers/kMeansPP-soda.pdf [accessed 2025-01-16]
42. Magiya J. Clustering GPS coordinates and forming regions with python. Medium. 2019. URL: https://levelup.gitconnected.com/clustering-gps-co-ordinates-forming-regions-4f50caa7e4a1 [accessed 2025-01-16]
43. Efron B, Tibshirani RJ. An Introduction to the Bootstrap: CRC press; 1994.
44. Davison AC, Hinkley DV. Bootstrap Methods and Their Application: Cambridge University Press; 1997.
45. Yurkiewicz IR, Simon P, Liedtke M, Dahl G, Dunn T. Effect of Fitbit and iPad wearable technology in health-related quality of life in adolescent and young adult cancer patients. J Adolesc Young Adult Oncol 2018 Oct;7(5):579-583. [doi: 10.1089/jayao.2018.0022] [Medline: 29924668]
46. Petersen J, Austin D, Mattek N, Kaye J. Time out-of-home and cognitive, physical, and emotional wellbeing of older adults: a longitudinal mixed effects model. PLoS ONE 2015;10(10):e0139643. [doi: 10.1371/journal.pone.0139643] [Medline: 26437228]
47. Boaro A, Leung J, Reeder HT, et al. Smartphone GPS signatures of patients undergoing spine surgery correlate with mobility and current gold standard outcome measures. J Neurosurg Spine 2021 Dec 1;35(6):796-806. [doi: 10.3171/2021.2.SPINE202181] [Medline: 34450590]
48. Panda N, Rattner DW, Morse CR. Third-time ("redo-redo") anti-reflux surgery: patient-reported outcomes after a thoracoabdominal approach. Surg Endosc 2020 Jul;34(7):3092-3101. [doi: 10.1007/s00464-019-07059-4] [Medline: 31388809]
49. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015 Sep;38(3):218-226. [doi: 10.1037/prj0000130] [Medline: 25844912]
50. Faherty LJ, Hantsoo L, Appleby D, Sammel MD, Bennett IM, Wiebe DJ. Movement patterns in women at risk for perinatal depression: use of a mood-monitoring mobile application in pregnancy. J Am Med Inform Assoc 2017 Jul 1;24(4):746-753. [doi: 10.1093/jamia/ocx005] [Medline: 28339686]
51. Wettstein M, Wahl HW, Shoval N, et al. Out-of-home behavior and cognitive impairment in older adults: findings of the SenTra Project. J Appl Gerontol 2015 Feb;34(1):3-25. [doi: 10.1177/0733464812459373] [Medline: 25548086]
52. Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC. The relationship between mobile phone location sensor data and depressive symptom severity. PeerJ 2016;4:e2537. [doi: 10.7717/peerj.2537] [Medline: 28344895]
53. Müller SR, Peters H, Matz SC, Wang W, Harari GM. Investigating the relationships between mobility behaviours and indicators of subjective well–being using smartphone–based experience sampling and GPS tracking. Eur J Pers 2020 Sep;34(5):714-732. [doi: 10.1002/per.2262]

Abbreviations

IRB: institutional review board
MH: mental health
PH: physical health
PROM: patient-reported outcome measure
PROMIS: Patient-Reported Outcomes Measurement Information System
QOL: quality of life
REDCap: Research Electronic Data Capture

XSL•FO
RenderX

Edited by A Mavragani; submitted 03.04.24; peer-reviewed by A AL-Asadi, KL Du; revised version received 11.12.24; accepted 12.12.24; published 07.02.25.

Please cite as:
Lee K, Azuero A, Engler S, Kumar S, Puga F, Wright AA, Kamal A, Ritchie CS, Demiris G, Bakitas MA, Odom JN
Exploring the Relationship Between Smartphone GPS Patterns and Quality of Life in Patients With Advanced Cancer and Their Family Caregivers: Longitudinal Study
JMIR Form Res 2025;9:e59161
URL: https://formative.jmir.org/2025/1/e59161
doi:10.2196/59161

© Kyungmi Lee, Andres Azuero, Sally Engler, Sidharth Kumar, Frank Puga, Alexi A. Wright, Arif Kamal, Christine S. Ritchie, George Demiris, Marie A. Bakitas, J. Nicholas Odom. Originally published in JMIR Formative Research (https://formative.jmir.org), 7.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

eHealth Literacy and Cyberchondria Severity Among Undergraduate Students: Mixed Methods Study

Wan -Chen Hsu, PhD

Center for Teaching & Learning Development , National Kaohsiung University of Science and Technology, No.1, University Rd.,, Yanchao District, Kaohsiung, Taiwan

Corresponding Author:
Wan -Chen Hsu, PhD
Center for Teaching & Learning Development , National Kaohsiung University of Science and Technology, No.1, University Rd.,, Yanchao District, Kaohsiung, Taiwan

Abstract

Background: With the development of the internet, health care websites have become increasingly important by enabling easy access to health information, thereby influencing the attitudes and behaviors of individuals toward health issues. However, few studies have addressed public access to health information and self-diagnosis.

Objective: This study investigated the background factors and status of cyberchondria severity among college students by conducting a nationwide sample survey using the Cyberchondria Severity Scale. Further, we explored the perspective of eHealth literacy of those with scores higher than 1 SD from the mean by analyzing their recent experiences using web-based health information.

Methods: A nationally representative sample of college students was surveyed, and 802 valid responses were obtained (male: 435/802, 54.2%; female: 367/802, 45.8%; mean age 20.3, SD 1.4 years). The Cyberchondria Severity Scale was used, which consisted of 4 dimensions (increased anxiety, obsessive-compulsive hypochondria, perceived controllability, and web-based physician-patient interaction). Additionally, we recruited 9 volunteers who scored more than 1 SD above the mean for in-depth interviews on their web-based health information–seeking behaviors.

Results: Significant differences were found across the 4 dimensions of cyberchondria severity ($F_{3,2403}$=256.26; P<.001), with perceived controllability scoring the highest (mean 2.75, SD 0.87) and obsessive-compulsive hypochondria scoring the lowest (mean 2.19, SD 0.77). Positive correlations were observed between perceived controllability, web-based physician-patient interactions, increased anxiety, and obsessive-compulsive hypochondria (r=0.46-0.75, P<.001). Regression analysis indicated that health concern significantly predicted perceived controllability (β coefficient=0.12; P<.05) and web-based physician-patient interaction (β coefficient=0.16; P<.001). Interview data revealed that students often experienced heightened anxiety (8/9, 89%) and stress (7/9, 78%) after exposure to web-based health information, highlighting the need for improved health literacy and reliable information sources.

Conclusions: The study identified both benefits and risks in college students' use of web-based health information, emphasizing the importance of critical consciousness and eHealth literacy. Future research should examine how college students move from self-awareness to actionable change and the development of critical health literacy, which are essential for effective digital health engagement.

(JMIR Form Res 2025;9:e63449) doi:10.2196/63449

KEYWORDS

eHealth literacy; undergraduate student; cyberchondria; compucondria; web-based health information; health information seeking; college students

Introduction

eHealth information is widely accessible and essential for public health improvement. However, a digital divide persists, with individuals in rural areas facing greater barriers to internet access. The increasing demand for at-home self-diagnosis is challenged by the questionable quality of some health-related websites, where personal anecdotes and content from reputable organizations may lack proper verification, affecting users' judgment and decision-making. Self-diagnosis websites often provide convenience but may lack oversight and include biased information, such as covert marketing for specific treatments, leading to potential exploitation [1]. This highlights the risks of relying on unverified digital health information, including misinformation, privacy issues, and harmful health decisions. eHealth literacy generally refers to an individual's ability to obtain, read, select, critique, and use health information from the internet to solve health problems [2-4]. It comprises 3 levels:

XSL•FO
RenderX

functional literacy, which refers to basic reading skills for acquiring health information; interactive literacy, involving advanced cognitive and social skills for selecting, understanding, synthesizing, and using health information in supportive environments; and critical literacy, which includes analytical and critical skills for evaluating and responding to health information, thereby improving control over life situations [5,6].

The ability to discern and ensure the trustworthiness and usability of information in a saturated web-based environment is critical for demonstrating eHealth literacy. This ability correlates with the effectiveness of individuals' web-based health information-seeking behaviors [7]. Those with high eHealth literacy can use web-based search strategies to identify high-quality health information [8], reducing technological risks. Conversely, insufficient ability to judge web-based health information can lead to cyberchondria, which refers to excessive or repeated web searches for health-related information because of health anxiety, resulting in increased distress or anxiety [9].

Hsu [10] compared eHealth literacy between younger (18-22 years) and older (55-72 years) college students, finding significant differences in overall scores and functional literacy, with older students scoring higher. This highlights eHealth literacy's role in helping students navigate and evaluate web-based health information, a key focus of our study. While most participants found web-based health information accessible and understandable, they were skeptical about its quality, viewing it as a mix of subjective and objective content. This underscores the importance of understanding how individuals interpret health information in a complex web-based environment.

Individuals searching for health-related information on the internet seek reassurance. However, information found on the internet is not always accurate or reliable, potentially increasing uncertainty, worry, and anxiety and leading to incorrect self-diagnoses [11,12]. Chen and Lee [13] noted that people often use limited skills to search for and evaluate vast amounts of web-based information of varying quality. However, accessing health information websites can help with self-diagnosis, and excessive concern over health may lead to cyberchondria, causing anxiety and serious health risks that arise from misinterpreting symptoms based on unverified web-based information, potentially leading to undue worry or even dangerous health actions [14]. Myrick [15] conducted a naturalistic experiment with 380 Americans to examine how emotions influence attitudes and behaviors following an information search. The study identified several emotions—fear, hope, satisfaction, interest, and motivation—that impact individuals' reactions to web-based health information. In particular, Myrick highlighted that social cognitive factors, such as self-efficacy and outcome expectations, positively influence confidence and behavioral intentions when searching for health information. This finding emphasizes how emotions and cognitive assessments together encourage or deter health-related actions, demonstrating the significant role of social cognitive factors in web-based health information–seeking behaviors.

Further, Starcevic and Berle [9] highlighted that when individuals with high health anxiety search for health information on the internet, receiving timely reassurance from others (particularly doctors) can alleviate their anxiety and reinforce their behavior of searching for health information. Conversely, if the search process increases anxiety or panic, users may continue searching to alleviate their negative emotions. However, repeated disappointment can worsen their emotional burden, deepen cyberchondria severity, and potentially lead to the avoidance of web-based health information searches in the future.

Providing web-based health information and ensuring public access are significant technological issues in society today. However, the "prescriptions" found on the internet—advice or treatments from articles or user testimonials without professional oversight—often offer only temporary relief, potentially causing anxiety and worry. These web-based prescriptions can mislead individuals, resulting in inappropriate or ineffective self-treatment, underscoring the importance of reliable, professionally vetted health information to support safer health decision-making. Additionally, the difficulty of verifying the credibility of web-based health information often leads users to rely on self-diagnosis based on their symptoms, which can increase health-related anxiety. This is particularly concerning when individuals lack the tools to assess the reliability of web-based sources, potentially resulting in misinformation and heightened worry. Adolescents, for instance, frequently report feeling overwhelmed by ambiguous health data on the internet, sometimes exacerbating their fears about symptoms they may not fully understand [16].

Most people searching for web-based health information lack medical training. Specifically, being exposed to a vast amount of web-based health information often causes anxiety and emotional distress because of the inability to understand medical terms or judge the information, resulting in severe self-diagnosis reactions [17]. Shan et al [18] found that adults generally have low levels of critical health literacy. Developing this literacy through education can enhance the ability to critically assess health information, engage in critical dialogues, and make health-related decisions.

Further, the measurement methods for cyberchondria severity are in developmental stages. Early studies have measured cyberchondria using a single item; however, later studies have developed multidimensional and multi-item scales. These scales include dimensions such as compulsion (excessive searching for web-based health information), distress (negative emotions and physiological responses from searching), reassurance (seeking reassurance from medical professionals), and mistrust of medical professionals (conflict between trusting web-based health information and doctors) [19]. Starcevic [20] considered cyberchondria to be a psychological construct. Hsu et al [3] found that college students with high health standards and high health concerns develop better eHealth literacy and adopt healthy eating, exercising, and better sleeping behaviors. Thus, in addition to developing reliable assessment tools, specific contextual items for different populations should be considered. Therefore, this study includes health status and health concerns to explore their impact on cyberchondria.

XSL•FO
RenderX

Specifically, this study aims to conduct a national survey using the Cyberchondria Severity Scale to understand the background factors and current status of cyberchondria among college students; recruit participants whose Cyberchondria Severity Scale scores are more than 1 SD above the mean and explore their experiences using web-based health information from the perspective of eHealth literacy; and based on the phased research results, summarize the characteristics of college students' experiences with web-based health information using the eHealth literacy framework and provide specific recommendations for developing health-promotion competencies to enhance the application value for health education practices.

Methods

Overview

This study used a sequential mixed methods design, which used qualitative and quantitative research sequentially, depending on the purpose or problem of the study. The purpose was to attain complementarity, such as the rationality of quantitative data in additional sampling and to investigate further through qualitative research [21]. Accordingly, this study was conducted in two phases. The first phase, comprising the quantitative part of the study, aimed to screen the respondents. The second phase, comprising interviews, aimed to collect the respondents' eHealth literacy data, which were analyzed qualitatively.

Recruitment

In this study, we specifically focused on Taiwanese college students. For the purposes of this research, a college student was defined as an individual aged 18 - 22 years currently enrolled in an undergraduate program at a Taiwanese higher education institution. This age range aligned with the typical age group of undergraduate students in Taiwan and provided a consistent basis for analysis within this demographic. The survey was conducted in two waves: at the end of the first (September 2021) and second (June 2022) semesters. We recruited 802 participants, including students from 14 universities in Taiwan, and we administered a questionnaire to them. Subsequently, we selected 9 volunteers for interviews based on their Cyberchondria Severity Scale scores, specifically those whose scores exceeded 1 SD above the mean, indicating a higher level of cyberchondria. Initial contact was made through the survey platform, and we followed strict ethical guidelines to protect participants' privacy throughout the recruitment and interview process.

Ethical Considerations

Ethical approval was granted by the Ethics Research Committee of National Cheng Kung University (No. 110-573-2), and informed consent was obtained from all participants. Participants had the right to withdraw from the study at any time. To preserve anonymity in the initial survey, all responses were collected without identifiable information. For follow-up interviews, we requested separate, voluntary consent from participants who met the selection criteria. All methods were carried out per relevant ethical standards, guidelines, and regulations. Participants who completed the initial survey received a US $3 gift voucher as a token of appreciation for their time and effort. Additionally, participants who took part in the follow-up interviews were compensated with a US $15 gift voucher to acknowledge their contribution to the study. These measures aimed to encourage participation while ensuring voluntary involvement without undue influence. By implementing these compensation protocols, we maintained ethical practices that aligned with institutional guidelines and ensured the rights and welfare of the participants were upheld throughout the research process.

Survey

The Cyberchondria Severity Scale was modified from those proposed by Durak Batigun et al [22] and White and Horvitz [23]. A 5-point Likert scale was used, with the following responses: (1) never, (2) rarely, (3) sometimes, (4) often, and (5) always. The exploratory factor analysis led to a Cyberchondria Severity Scale for college students with 4 constructs: increased anxiety (Cronbach α=0.91), obsessive-compulsive hypochondria (Cronbach α=0.87), perceived controllability (Cronbach α=0.88), and web-based physician-patient interaction (Cronbach α=0.86). The total scale Cronbach α was 0.92, with a variance of 66.81%. The confirmatory factor analysis indicated that item reliability ranged from 0.50 to 0.86, factor loadings ranged from 0.71 to 0.93, and the composite reliability for latent variables ranged from 0.83 to 0.90 (P<.001). The extracted average variance was from 0.46 to 0.60. Increased anxiety referred to the anxiety experienced by an individual due to apprehensions about web-based content-seeking to understand a disease or its symptoms (eg, "To understand a certain disease or symptom, I went on the internet to search for information. The large amount of information was too much for me to handle and increased my anxiety"). Obsessive-compulsive hypochondria referred to excessive worry or excessive internet surfing to understand a disease or its symptoms (eg, "I feel that I am someone who worries excessively about health"). Perceived controllability referred to the decreased doubts or psychological burdens caused by a web search for health information (eg, "Searching for health-related information on the internet allows me to understand my medical condition and decreases my anxiety"). Web-based physician-patient interaction refered to an individual's web-based information search that affected their judgment of their symptoms (eg, "After seeing a physician, I would use the internet to obtain detailed information on the diagnosis").

To understand the advantages, disadvantages, and risks of college students' access to web-based health information, this study used semistructured interviews. The interview outline was developed based on the research objectives and focused on collecting information about the background of frequent web-based health information users who engaged in self-diagnosis, as well as their experiences, including the advantages, disadvantages, and risks. This information was gathered from college students.

Data Analysis

In the first phase of the study, SPSS version 20 (IBM Corp) was used to calculate the mean, SD, and dependent - sample one-way ANOVA to indicate the current status and differences

XSL•FO
RenderX

across various dimensions of college students' cyberchondria. In the second phase, the Strauss and Corbin constant comparative method was used to analyze the advantages, disadvantages, and risks of web-based health information [23]. Open coding was first conducted by assigning codes and noting data entry dates. For example, for Participant 8-20220920, the number 20220920 indicates that the interview with Participant 8 was conducted on September 20, 2022. Thereafter, the researcher read the transcripts, conceptualized the data, and grouped similar concepts into categories. Finally, axial coding was conducted to link the subcategories and main categories. This study followed Lincoln and Guba's [24] credibility, dependability, and confirmability criteria to ensure research quality. With the participants' consent, data were collected by recording and note-taking to ensure completeness and accuracy.

Results

Current Situation of Cyberchondria Severity Among College Students

The sociodemographic details of the sample are shown in Table 1. Regarding the different dimensions of cyberchondria, the average single-item scores ranged from 2.19 to 2.75 (between rarely and sometimes). The average scores for each dimension were below the median, indicating that college students did not frequently experience hypochondria during web-based health information searches. However, they did not often feel that their anxiety was controllable or engage in web-based physician-patient interactions (Table 2).

Table . Sociodemographic details of the sample (n=802).

Variable and group	Frequency, n (%)
Sex	
Male	435 (54.2)
Female	367 (45.8)
Grade	
Freshman year	565 (70.4)
Sophomore year	74 (9.2)
Junior year	87 (10.8)
Senior year	76 (9.5)
Health status	
Very unhealthy	4 (5)
Unhealthy	53 (6.6)
Neutral	252 (31.4)
Healthy	298 (37.2)
Very healthy	195 (24.3)
Degree of health concern	
Very unconcerned	12 (1.5)
Unconcerned	55 (6.9)
Neutral	266 (33.2)
Concerned	287 (35.8)
Very concerned	182 (22.7)

Table . Descriptive statistics of cyberchondria severity (n=802).

Single-item	Minimum score	Maximum score	Likert score, mean (SD)
Increased anxiety	1	5	2.39 (0.82)
Obsessive-compulsive hypochondria	1	5	2.19 (0.77)
Perceived controllability	1	5	2.75 (0.87)
Online physician–patient interaction	1	5	2.70 (0.80)

Owing to the different number of items in the 4 dimensions of cyberchondria among college students, a dependent - sample one-way ANOVA was conducted using the single-item mean scores of the 4 dimensions. The results showed significant differences across the dimensions ($F_{3,2403}$=256.26; P<.001), indicating significant variations in cyberchondria scores among college students across different dimensions. Further post hoc comparisons using the least significant difference test revealed

XSL•FO
RenderX

that perceived controllability had the highest score, with the highest single item being "When I search for health-related information on the internet, reading about my condition from reliable sources (eg, hospital websites) alleviates my anxiety" (mean 2.79, SD 0.75). Obsessive-compulsive hypochondria scored the lowest, indicating that college students did not consider themselves to be overly concerned about their health (Table 3).

Table . Dependent - sample one-way ANOVA of the different dimensions of cyberchondria severity among college students.

Analysis	Type III sum of squares (*df*)	Mean square
Between groups	513.11 (801)	0.64
Within groups	513.11 (801)	0.64
Each dimension[a]	513.61 (3)	171.20
Residual error	1605.41 (2403)	0.668
Total	2632.13 (3207)	—

[a]Post hoc test revealed perceived controllability>web-based physician–patient interaction>increased anxiety>obsessive-compulsive hypochondria ($F_{3,2403}$=256.26, P<.001).

The correlation matrix of different dimensions indicated significant positive correlations between perceived controllability, web-based physician-patient interaction, increased anxiety, and obsessive-compulsive hypochondria. This suggests that the more college students feel they can control their anxiety by searching for web-based health information and the more frequently they interact with web-based physicians, the higher their levels of anxiety and obsessive-compulsive hypochondria. Further, perceived controllability exhibited a significant positive correlation with web-based physician-patient interactions. The absolute values of the correlation coefficients were all above 0.40, indicating moderate-to-high correlations. Thus, the most significant dimensions of cyberchondria among college students showed moderate or high correlations (Table 4).

Table . Correlation analysis of different dimensions of cyberchondria severity among college students.

Variable	Increased anxiety	Obsessive-compulsive hypochondria	Perceived controllability	Online physician–patient interaction
Increased anxiety				
r	1.00	0.75	0.46	0.48
P value	—[a]	<.001	<.001	<.001
Obsessive-compulsive hypochondria				
r	0.75	1.00	0.50	0.54
P value	<.001	—	<.001	<.001
Perceived controllability				
r	0.46	0.50	1.00	0.73
P value	<.001	<.001	—	<.001
Online physician–patient interaction				
r	0.48	0.54	0.73	1.00
P value	<.001	<.001	<.001	—

[a]Statistical analysis was not done for reference values.

Table 5 indicates that all individual variables were significant predictors of the 4 dimensions of cyberchondria severity, with an overall explanatory power of 1% - 2%, corresponding with a low predictive power level. Notably, the degree of health concern emerged as a significant positive indicator of all dimensions of cyberchondria severity, except increased anxiety.

XSL•FO
RenderX

Table . Multiple regression analyses of individual factors predicting cyberchondria severity.

Variable	Increased anxiety[a]				Obsessive-compulsive hypochondria[b]				Perceived controllability[c]				Web-based physician-patient interaction[d]			
	B[e]	β[f]	*t* test (*df*)	*P* value	B	β	*t* test (*df*)	*P* value	B	β	*t* test (*df*)	*P* value	B	β	*t* test (*df*)	*P* value
Health status	−0.09	−0.10	−2.32 (790)	.02	−0.11	−0.13	−2.87 (790)	.006	0.00	0.00	0.07 (790)	.43	−0.01	−0.01	−0.29 (790)	.06
Degree of health concern	0.01	0.01	0.18 (790)	.08	0.11	0.13	2.93 (790)	.004	0.11	0.12	2.58 (790)	.02	0.14	0.16	3.69 (790)	<.001

[a]R=0.10, R^2=0.01 (P=.02); $F_{3,790}$=2.73 (P=.02).

[b]R=0.12, R^2=0.01 (P=.02); $F_{3,790}$=3.55 (P=.02).

[c]R=0.12, R^2=0.02 (P=.002); $F_{3,790}$=3.90 (P=.003).

[d]R=0.16, R^2=0.02 (P<.001); $F_{3,790}$=6.54 (P<.001).

[e]B: unstandardized coefficients.

[f]β: standardized coefficients.

Impact of Accessing Web-Based Health Information on the Physical and Mental Well-Being of College Students

The interviews with college students revealed that accessing web-based health information has both psychological and physical effects, often creating challenges due to the nature and quality of the information available on the internet. In terms of psychological impact, participants reported feeling overwhelmed by the plethora of opinions and conflicting advice, which often caused confusion and heightened anxiety. For example, one student shared the following:

> *I could not find any results through searching, so I mainly see a doctor because I have no other option. I don't know what to do anymore, so I must see a doctor. And since I've been paying more attention to health issues, I've noticed that the first things that pop up are usually some exaggerated news or bizarre articles from content farms. The headlines make me feel like, "Oh my god! I'm about to die."* [Participant 8-20220920]

This highlights the stress caused by exaggerated or unreliable health information, which drives the students to consult health care professionals as a last resort. As for strained doctor-patient relationships, the tendency to self-diagnose based on web-based information sometimes led students to question or dismiss professional medical advice, creating tension in doctor-patient interactions. One student remarked the following:

> *If you have doubts about the doctor's words, don't ask random questions. For instance, if I now Google some issues, like searching for stomach pain on Google, and then tell the doctor, "But Google says this is some specific disease, so you just need to prescribe me this medicine," it can lead to problems.* [Participant 1-20221101]

This underscores the potential of web-based health information to undermine the trust in medical expertise. Additionally, exposure to sensational or clickbait content exacerbated feelings of fear and helplessness. Students described how alarming headlines led to unnecessary worry:

> *And since I've been paying more attention to health issues, I've noticed that the first things that pop up are usually some exaggerated news… The headlines make me feel like, " Oh my god! I'm about to die."* [Participant 8-20220920]

The physical impact included misguided health practices and delayed medical treatment. Some students engaged in health practices based on unverified advice on the internet, which sometimes led to negative outcomes. For example, one participant shared a personal experience:

> *My chest is not very big, and I saw online that eating green papaya can make your chest bigger. I was just in middle school then, so I started eating one every day… Then I noticed that my hands were starting to turn yellow.* [Participant 9-20221024]

Moreover, in some cases, students delayed seeking professional care due to reliance on web-based advice. One participant shared a story about their friend:

> *Like my friend, he once sprained his finger while playing basketball… He looked it up online and found advice saying to ice it and not press it. He kept doing that, and now his finger is crooked.* [Participant 3-20221107]

This highlights the risk of self-experimentation based on health misinformation and demonstrates how misleading information can lead to improper self-care and long-term consequences. The findings emphasize the dual nature of web-based health information. Although such information is accessible and convenient, its unreliable or sensational nature can lead to considerable psychological and physical challenges for students.

XSL•FO
RenderX

This emphasizes the urgent need for improving students' health literacy on the internet and promoting access to credible and professionally vetted health information.

College Students' Experiences With Accessing Web-Based Health Information: The Advantages, Disadvantages, and Risks

Students' responses to the semistructured interviews revealed convenience, richness of information, privacy, and validation as the 4 main themes related to using web-based health information. The motivation for using web-based health information was to meet the needs of beautifying features, weight loss, and fitness. The two types of web-based health information access were fixed webpage browsing and nonfixed webpage retrieving. The criteria for checking the quality of web-based health information included subjective judgments and objective features. Regarding web-based health information reading experiences, problems were caused by technical terms, foreign languages, and numerical health information data. Typically, the participants searched for related information on the internet to clarify their doubts. Finally, the impacts of using web-based health information depend on assessing its quality, lifestyle changes, and the possibility of self-practice. Self-diagnosis can be a double-edged sword. It may empower individuals with preliminary health insights; however, it also poses risks, particularly if the information is misinterpreted or lacks professional verification. Individuals should be encouraged to evaluate web-based health information and approach self-diagnosis with caution. Additionally, we will consider the ethical implications of promoting self-diagnosis in health-related research, emphasizing the importance of fostering critical eHealth literacy among users as a suggestion.

The participants reported various risks and benefits while searching for health information on the internet. They tended to be wary of sensational or emphatically presented headlines, leading them to adopt a conservative approach to avoid harm. In assessing the quality of information, they paid attention to the source, crosschecked it with multiple sources, and judged its effectiveness based on their personal experiences and expert endorsements, ensuring content completeness. Their experiences included selective understanding, focusing on general ideas rather than technical terms, and further clarifying doubtful information. The impact of this information varied based on the closeness of relationships with others and the feasibility of implementing lifestyle changes. For instance, support from close friends and family could enhance the likelihood of adopting positive health behaviors and reduce anxiety related to conflicting information. Conversely, a lack of support or conflicting opinions from peers could heighten stress and uncertainty, making it more challenging to discern and apply accurate health information.

Summarizing the experiences of college students who engaged in self-diagnosis through extensive internet use for health information access, the benefits primarily included cross-referencing, synthesizing diverse information, and verifying questionable information. The harms mainly arose from confusion due to difficulty reading and understanding the diverse information and, in more severe cases, engaging in self-experimentation that affected physical health. The risks involved selective understanding, blind sharing of information, and extreme rejection of any information, which could undermine the positive intentions of effectively using web-based health information. Table 6 presents the interview analysis and coding framework.

XSL•FO
RenderX

Table . Interview analysis and coding framework.

Main and subcategory structure	Interviewee numbers
Background of using web-based health information	
Time of initial use	
I mostly started in high school	1, 3, 4, 6, 7, 8, 9
Initial motivation for using web-based health information	
Knowledge acquisition	1, 2, 4, 5, 6, 8, 9
Concern for one's condition	1, 2, 3, 4, 5, 6, 7, 8, 9
Searching for health and medical information for family	2, 5, 6, 8, 9
Searching for drug ingredients	2, 3, 5, 6, 7, 8
Looking for nearby medical facilities	1, 2, 4, 7, 8
Reason for choosing web-based health information	
Free of charge	1, 2, 3, 4, 6, 8
Richness of information	1, 2, 5, 7, 8, 9
Verifiability	1, 2, 3, 4, 5, 6, 7, 8
Convenience	1, 2, 4, 5, 7, 8, 9
Experience in accessing web-based health information	
Quality assessment of web-based health information: subjective judgment	
Sensational headlines	1, 4, 5, 6, 8
Headlines overly emphasizing efficacy	1, 4, 5, 6, 7, 8, 9
Conservative adoption to avoid harm	2, 3, 5, 7, 8
Judged by my common sense	1, 3, 5, 6, 8, 9
Quality assessment of web-based health information: objective standards	
Source of information	1, 2, 4, 5, 6, 8
Cross-verification	2, 3, 5, 6, 7, 8, 9
Judged by actual experiences of friends and family	1, 2, 3, 4, 7, 8
Expert evaluation of the efficacy of information	1, 3, 4, 5, 6, 7
Incomplete information	2, 3, 4, 5, 6, 9
Experience in reading web-based health information	
Summary understanding and further inquiry	
Handling doubts about web-based health information: continue searching or ask others interesting questions	1, 2, 3, 4, 5, 6, 7, 8, 9
Understanding of web-based health information: often ignoring technical terms and focusing on general meaning	1, 4, 5, 6, 8, 9
Impact of web-based health information	
Deliberation on object and risk outcome assessment	
Influence on others based on personal relationships: varying forms and degrees of influence	1, 2, 3, 4, 7, 9
Self-physical and mental impact: depending on the extent of lifestyle changes or feasibility	1, 2, 3, 4, 5, 6, 7, 8, 9

Discussion

Principal Findings

Digital platforms have revolutionized health and science by enabling collaboration, expanding evidence-based interventions, and advancing public health. Early successes in behavioral medicine demonstrated their potential to scale therapeutic solutions. With growing investments in health information technology, behavioral informatics has become essential, integrating medicine, psychology, and engineering to develop patient apps, clinical tools, communication programs, and population health strategies. The Society of Behavioral

XSL•FO
RenderX

Medicine, positioned at this interdisciplinary nexus, plays a pivotal role in shaping how science is applied to reduce health disparities, support patients and caregivers, enhance community well-being, and address challenges such as psychological pain and addiction [25]. This study examined eHealth literacy and cyberchondria severity in college students, highlighting their web-based health information-seeking behaviors. Key findings include links between perceived controllability, web-based interactions, anxiety, and obsessive-compulsive hypochondria, with health concern predicting controllability and interactions. Interviews showed that 89% (8/9) felt heightened anxiety, and 78% (7/9) reported stress from web-based health information.

These results highlight the dual nature of web-based health information. The high perceived controllability scores suggest that accessing reliable web-based health information—such as from trusted sources like hospital websites—can alleviate student anxiety by helping them manage health-related stress. This finding aligns with prior research showing the importance of credible sources in fostering a sense of control [10,23]. By contrast, the low obsessive-compulsive hypochondria scores suggest that students generally do not exhibit excessive worry or compulsive searching behaviors. This is contrary to the studies linking frequent web-based health searches to increased distress and cyberchondria severity [9]. This discrepancy may stem from differences in digital literacy or cultural attitudes toward web-based health information.

The positive association between health concern and cyberchondria dimensions underscores the dual impact of web-based health information. Although such information can promote awareness and self-management, it can also lead to anxiety and stress when students encounter conflicting or sensational content; this is consistent with the findings from Tan and Goonawardene's systematic review [26]. Moreover, the role of web-based physician-patient interactions in moderating anxiety highlights the importance of guiding students toward appropriate and reliable eHealth resources. This aligns with Hsu's [10] findings that enhancing eHealth literacy can empower students to critically evaluate web-based health information and reduce health-related anxiety.

Notably, the interviews revealed that excessive web-based health searches contribute to confusion and emotional distress, particularly when students encounter contradictory advice or exaggerated claims. These results align with those of Starcevic and Berle [9], who identified increased anxiety and obsessive-compulsive tendencies as key outcomes of frequent health information searches. Furthermore, exposure to sensational content, such as clickbait headlines, exacerbates fear and helplessness, increasing the likelihood of unnecessary medical consultations or misdiagnoses [11,23]. These findings emphasize the need to develop interventions that improve students' critical health literacy, which would enable them to discern credible information and mitigate the psychological impacts of misinformation.

This study has limitations. Self-reported data may introduce response bias, suggesting the need for objective health measures like medical records for greater reliability. The qualitative phase included only 9 participants, restricting the generalizability of the findings; a larger sample could offer deeper insights. Additionally, focusing solely on Taiwanese college students limits the applicability of the results to other cultures or demographics. Cross-cultural studies are recommended to examine variations in eHealth literacy and cyberchondria across populations.

This study highlights the complex interplay between eHealth literacy and cyberchondria severity among college students, revealing both benefits and risks associated with web-based health information-seeking behavior. Reliable web-based information can reduce anxiety and enhance self-management, whereas the prevalence of contradictory or sensational content underscores the importance of fostering critical eHealth literacy. Drawing on Nutbeam's framework of health literacy [18,27], this study advocates for integrating critical health literacy into health education programs to empower students to effectively navigate the digital health landscape.

Moreover, the findings contribute to the broader discourse on digital health education, emphasizing the need for educational interventions that promote self-awareness and actionable behavioral changes. As previous studies have shown, developing critical eHealth literacy not only improves individual decision-making but also strengthens patient-physician relationships by encouraging open and informed discussions [26,28]. Future efforts should focus on designing tailored health literacy initiatives that equip students with the skills to critically assess and apply web-based health information, fostering their ability to engage responsibly with digital health resources in a rapidly evolving technological landscape.

Conclusions

This study surveyed college students using the Cyberchondria Severity Scale and interviewed high scorers to explore their browsing behaviors and experiences with web-based health information. Significant score differences were observed across cyberchondria dimensions, with perceived controllability scoring highest and obsessive-compulsive hypochondria lowest. High scorers noted benefits such as cross-verification and clarification of information; drawbacks like confusion from complexity; and risks including selective understanding, reflexive sharing, and mistrust, which could undermine the advantages of web-based health resources.

The internet has revolutionized medical education and practice by enabling the rapid, large-scale exchange of information and ideas across geographically dispersed audiences [29]. To effectively use web-based health information, critical awareness and eHealth literacy are essential skills. Future research should investigate how college students move from self-awareness to a behavioral change that helps them develop critical health literacy as a crucial component of digital health education.

This study relied on self-reported questionnaires, which may introduce response bias and affect the accuracy of findings, particularly regarding health status. Future studies could improve data reliability by encouraging participants to reference recent health examination results, enhancing internal validity. Additionally, the small interview sample limits the

XSL•FO
RenderX

generalizability of conclusions. Expanding the sample in future research could reduce bias and support more robust findings.

Acknowledgments

The authors thank the National Science and Technology Council for funding this study.

Data Availability

The anonymized data from the survey are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Ryan A, Wilson S. Internet healthcare: do self-diagnosis sites do more harm than good? Expert Opin Drug Saf 2008 May;7(3):227-229. [doi: 10.1517/14740338.7.3.227] [Medline: 18462181]
2. Chiang CH, Yang SC, Hsu WC. Development and validation of the e-health literacy scale and investigation of the relationships between e-health literacy and healthy behavior among undergraduate students in Taiwan. Formosa J Ment Health 2015;23(3):389-420. [doi: 10.30074/FJMH]
3. Hsu WC, Chiang CH, Yang SC. The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy. J Med Internet Res 2014 Dec 12;16(12):e287. [doi: 10.2196/jmir.3542] [Medline: 25499086]
4. Norman CD, Skinner HA. eHEALS: the eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27. [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
5. Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008 Dec;67(12):2072-2078. [doi: 10.1016/j.socscimed.2008.09.050] [Medline: 18952344]
6. Nutbeam D, Kickbusch I. Advancing health literacy: a global challenge for the 21st century. Health Promot Int 2000 Sep 1;15(3):183-184. [doi: 10.1093/heapro/15.3.183]
7. Hsu WC, Chen SF, Ho CJ. Experience of using web health information among college students: an analysis from the health literacy perspective. J Health Promot Health Educ Contents 2011;35:1-22. [doi: 10.7022/JHPHE.201106.0001]
8. Quinn S, Bond R, Nugent C. Quantifying health literacy and eHealth literacy using existing instruments and browser-based software for tracking online health information seeking behavior. Comput Human Behav 2017 Apr;69:256-267. [doi: 10.1016/j.chb.2016.12.032]
9. Starcevic V, Berle D. Cyberchondria: towards a better understanding of excessive health-related internet use. Expert Rev Neurother 2013 Feb;13(2):205-213. [doi: 10.1586/ern.12.162] [Medline: 23368807]
10. Hsu WC. The effect of age on electronic health literacy: mixed-method study. JMIR Hum Factors 2019 Apr 21;6(2):e11480. [doi: 10.2196/11480] [Medline: 31066696]
11. Hsu WC. Undergraduate students' online health information-seeking behavior during the COVID-19 pandemic. Int J Environ Res Public Health 2021 Dec 16;18(24):13250. [doi: 10.3390/ijerph182413250] [Medline: 34948859]
12. Xu DY, Ding JL, Yang ZH. Pathogenesis of cyberchondria and its impact on individuals: a qualitative study. Chin Gen Pract 2018;21(22):2755-2759. [doi: 10.12114/j.issn.1007-9572.2018.22.022]
13. Chen W, Lee KH. More than search? Informational and participatory eHealth behaviors. Comput Human Behav 2014 Jan;30:103-109. [doi: 10.1016/j.chb.2013.07.028]
14. Smith PK, Fox AT, Davies P, Hamidi-Manesh L. Cyberchondriacs. Int J Adolesc Med Health 2006;18(2):209-213. [Medline: 16894859]
15. Myrick JG. The role of emotions and social cognitive variables in online health information seeking processes and effects. Comput Human Behav 2017 Mar;68:422-433. [doi: 10.1016/j.chb.2016.11.071]
16. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place of the internet. Soc Sci Med 2005 Apr;60(7):1467-1478. [doi: 10.1016/j.socscimed.2004.08.010] [Medline: 15652680]
17. Jokić-Begić N, Mikac U, Čuržik D, Sangster Jokić C. The development and validation of the Short Cyberchondria Scale (SCS). J Psychopathol Behav Assess 2019 Dec;41(4):662-676. [doi: 10.1007/s10862-019-09744-z]
18. Shan Y, Ji M, Dong Z, Xing Z, Xu X. Assessing patients' critical health literacy and identifying associated factors: cross-sectional study. J Med Internet Res 2023 Apr 5;25:e43342. [doi: 10.2196/43342] [Medline: 37018027]
19. McElroy E, Shevlin M. The development and initial validation of the Cyberchondria Severity Scale (CSS). J Anxiety Disord 2014 Mar;28(2):259-265. [doi: 10.1016/j.janxdis.2013.12.007] [Medline: 24508033]
20. Starcevic V. Cyberchondria: challenges of problematic online searches for health-related information. Psychother Psychosom 2017;86(3):129-133. [doi: 10.1159/000465525] [Medline: 28490037]
21. Venkatesh V, Brown SA, Bala H. Bridging the qualitative-quantitative divide: guidelines for conducting mixed methods research in information systems. MISQ 2013 Jan 1;37(1):21-54. [doi: 10.25300/MISQ/2013/37.1.02]

XSL•FO
RenderX

22. Durak Batigun A, Gor N, Komurcu B, Senkal Erturk I. Cyberchondria Scale (CS): development, validity and reliability study. Dusunen Adam 2018 Jun 29;31:148-162. [doi: 10.5350/DAJPN2018310203]
23. White RW, Horvitz E. Cyberchondria: studies of the escalation of medical concerns in web search. ACM Trans Inf Syst 2009;27(4):23. [doi: 10.1145/1629096.1629101]
24. Lincoln YS, Guba EG. Naturalistic Inquiry: Sage Publications, Inc; 1985.
25. Hesse BW. Riding the wave of digital transformation in behavioral medicine. Ann Behav Med 2020 Dec 1;54(12):960-967. [doi: 10.1093/abm/kaaa093] [Medline: 33416839]
26. Tan SSL, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9. [doi: 10.2196/jmir.5729] [Medline: 28104579]
27. Abel T, Benkert R. Critical health literacy: reflection and action for health. Health Promot Int 2022 Aug 1;37(4):daac114. [doi: 10.1093/heapro/daac114] [Medline: 36047637]
28. Devan H, Perry MA, van Hattem A, et al. Do pain management websites foster self-management support for people with persistent pain? A scoping review. Patient Educ Couns 2019 Sep;102(9):1590-1601. [doi: 10.1016/j.pec.2019.04.009] [Medline: 30981410]
29. Levy JA, Strombeck R. Health benefits and risks of the internet. J Med Syst 2002 Dec;26(6):495-510. [doi: 10.1023/a:1020288508362] [Medline: 12385532]

Edited by A Mavragani; submitted 20.06.24; peer-reviewed by J Behre, R Peng; revised version received 12.01.25; accepted 15.01.25; published 03.02.25.

Please cite as:
Hsu C
eHealth Literacy and Cyberchondria Severity Among Undergraduate Students: Mixed Methods Study
JMIR Form Res 2025;9:e63449
URL: https://formative.jmir.org/2025/1/e63449
doi:10.2196/63449

© Wan-Chen Hsu. Originally published in JMIR Formative Research (https://formative.jmir.org), 3.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Estimating the Prevalence of Schizophrenia in the General Population of Japan Using an Artificial Neural Network–Based Schizophrenia Classifier: Web-Based Cross-Sectional Survey

Pichsinee Choomung[1], PharmD; Yupeng He[1], MD, PhD; Masaaki Matsunaga[1], MD, PhD; Kenji Sakuma[2], MD, PhD; Taro Kishi[2], MD, PhD; Yuanying Li[3], PhD; Shinichi Tanihara[4], PhD; Nakao Iwata[2], MD, PhD; Atsuhiko Ota[1], MD, PhD

[1]Department of Public Health, Fujita Health University School of Medicine, 1-98 Dengakugakubo, Kutsukake-cho, Toyoake, Japan

[2]Department of Psychiatry, Fujita Health University School of Medicine, Toyoake, Japan

[3]Department of Public Health and Health Systems, Nagoya University Graduate School of Medicine, Nagoya, Japan

[4]Department of Public Health, Kurume University School of Medicine, Kurume, Japan

Corresponding Author:
Yupeng He, MD, PhD
Department of Public Health, Fujita Health University School of Medicine, 1-98 Dengakugakubo, Kutsukake-cho, Toyoake, Japan

Abstract

Background: Estimating the prevalence of schizophrenia in the general population remains a challenge worldwide, as well as in Japan. Few studies have estimated schizophrenia prevalence in the Japanese population and have often relied on reports from hospitals and self-reported physician diagnoses or typical schizophrenia symptoms. These approaches are likely to underestimate the true prevalence owing to stigma, poor insight, or lack of access to health care among respondents. To address these issues, we previously developed an artificial neural network (ANN)–based schizophrenia classification model (SZ classifier) using data from a large-scale Japanese web-based survey to enhance the comprehensiveness of schizophrenia case identification in the general population. In addition, we also plan to introduce a population-based survey to collect general information and sample participants matching the population's demographic structure, thereby achieving a precise estimate of the prevalence of schizophrenia in Japan.

Objective: This study aimed to estimate the prevalence of schizophrenia by applying the SZ classifier to random samples from the Japanese population.

Methods: We randomly selected a sample of 750 participants where the age, sex, and regional distributions were similar to Japan's demographic structure from a large-scale Japanese web-based survey. Demographic data, health-related backgrounds, physical comorbidities, psychiatric comorbidities, and social comorbidities were collected and applied to the SZ classifier, as this information was also used for developing the SZ classifier. The crude prevalence of schizophrenia was calculated through the proportion of positive cases detected by the SZ classifier. The crude estimate was further refined by excluding false-positive cases and including false-negative cases to determine the actual prevalence of schizophrenia.

Results: Out of 750 participants, 62 were classified as schizophrenia cases by the SZ classifier, resulting in a crude prevalence of schizophrenia in the general population of Japan of 8.3% (95% CI 6.6%-10.1%). Among these 62 cases, 53 were presumed to be false positives, and 3 were presumed to be false negatives. After adjustment, the actual prevalence of schizophrenia in the general population was estimated to be 1.6% (95% CI 0.7%-2.5%).

Conclusions: This estimated prevalence was slightly higher than that reported in previous studies, possibly due to a more comprehensive disease classification methodology or, conversely, model limitations. This study demonstrates the capability of an ANN-based model to improve the estimation of schizophrenia prevalence in the general population, offering a novel approach to public health analysis.

(JMIR Form Res 2025;9:e66330) doi:10.2196/66330

KEYWORDS

schizophrenia; schizophrenic; prevalence; artificial neural network; neural network; neural networks; ANN; deep learning; machine learning; SZ classifier; web-based survey; epidemiology; epidemiological; Japan; classifiers; mental illness; mental disorder; mental health

XSL•FO
RenderX

Introduction

Schizophrenia is a complex psychiatric disorder characterized by significant impairments in the perception of reality and behavioral changes. It includes positive (eg, delusions, hallucinations, disorganized thinking, and notably disorganized behaviors) and negative symptoms (eg, limited speech, decreased emotional expression, inability to experience interest or pleasure, and social withdrawal). Affected individuals often experience cognitive impairments affecting memory, attention, and problem-solving abilities [1,2]. The Global Burden of Disease Study [3] reported that schizophrenia accounted for 23.6 million estimated cases in 2019, with an age-standardized prevalence rate of 287.4 per 100,000 people. Median estimates from internationally published studies indicate a 12-month prevalence of 0.33%, lifetime prevalence of 0.48%, and point prevalence of 0.32% [4]. However, there is a lack of published studies on the prevalence of schizophrenia in the general Japanese population. Only a few population-based studies have estimated the prevalence of schizophrenia in Japan. For instance, the point prevalence of schizophrenia stratified by sex and age ranges from 0.4% to 1.1% [5]. However, these figures were derived from the Vital Statistics and the Patient Survey in 2008, which probably do not represent the general population since the Patient Survey could not encompass patients who do not visit a doctor. The estimated lifetime prevalence of schizophrenia is 0.59%, according to a secondary analysis of the 2019 cross-sectional Japan National Health and Wellness Survey [6]. However, this study may produce underestimation since it only depended on physician diagnoses. Individuals with schizophrenia hardly self-identify as having schizophrenia because of stigma and lack of awareness of schizophrenia and its symptoms [7-13].

In our previous research, we developed an artificial neural network (ANN)–based schizophrenia classification model (SZ classifier) to classify schizophrenia cases in the Japanese population and verified its generalizability [14]. The model was trained by using data from a large-scale Japanese web-based survey; the presence of schizophrenia served as a response variable. The general information including demographics, health-related background, physical comorbidities, psychiatric comorbidities, and social comorbidities were collected and used as feature variables. This SZ classifier achieved high internal validity with a high area under the receiver operating characteristic curve (AUC) of 0.86 [14]. In another study, we also assessed the SZ classifier's external validity by applying it to psychiatric outpatients, achieving a sensitivity of 75% [15]. The SZ classifier suggests a potential for classifying schizophrenia cases in the general Japanese population.

In this study, we aimed to estimate the prevalence of schizophrenia in the general Japanese population using the SZ classifier through the following three approaches: (1) conducting a population-based survey to recruit participants and collect their general information, (2) sampling participants who match the distribution to the population demographic structure, and (3) using a pretrained machine learning model to detect schizophrenia cases. Study participants were randomly sampled from a pooled panel. The samples are matched to the demographic distribution of the Japanese population. General information, the same as that for the development of the SZ classifier, was collected through the web-based survey for each study participant. The estimation of the schizophrenia prevalence was implemented through the following two steps: (1) calculating the crude prevalence based on crude cases that are detected by the SZ classifier and (2) refining the crude prevalence by excluding false-positive estimates and including the false-negative estimates.

Methods

Development of the SZ Classifier

The SZ classifier was developed as part of a prior study [14] using a dataset obtained from an internet research agency's pooled panels (Rakuten Insight, Inc) [16]. The dataset included 223 individuals with schizophrenia and 1776 healthy controls. For each sample, schizophrenia status served as the response variable, while demographic details, health-related background information, and physical, psychiatric, and social comorbidities were considered feature variables. Evidence on the association between these feature variables and schizophrenia has been previously published [17,18]. For example, demographic and health-related characteristics included age, BMI, smoking status, restriction in functional capacity, and bowel frequency. Physical comorbidities included overweight, cancer, cardiovascular disease, heart failure, and hypertension. Psychiatric comorbidities included frequency of sleep medication use, bedtime, hours of sleep, perceived stress, and positive reasons for living. The social comorbidities included educational background, type of occupation, type of employment, household income, and marital status. The inclusion criteria of the study participants and definitions of the feature variables are detailed in previous publications [14,18] and provided in Multimedia Appendix 1.

The SZ classifier was constructed using an ANN concept. This multilayer perceptron neural network model comprised 5 hidden layers with 128, 64, 32, 16, and 8 neurons in each respective layer. It used the HeNormal weight initializer and the ReLU activation function in the hidden layers. A sigmoid activation function was used in the output layer. Comprehensive details can be found in a prior publication [14]. The model's performance was evaluated using the AUC score. During internal validation, the SZ classifier achieved an AUC of 0.86, accuracy of 0.93, specificity of 0.96, and sensitivity of 0.56. He et al [14] performed 10,000 bootstrapping iterations and a 10-fold cross-validation to further evaluate the algorithm's performance and achieved AUCs of 0.847 and 0.81, respectively.

External Validation of the SZ Classifier

External validation of the SZ classifier was conducted by applying it to a sample of patients who visited the psychiatric department of Fujita Health University Hospital between January 2022 and May 2023 [15]. Patients aged 20 - 75 years diagnosed with schizophrenia, major depressive disorder (MDD), bipolar disorder (BD), or obsessive-compulsive disorder (OCD) were selected. Therefore, 150 participants were included in the study: 61 with schizophrenia, 56 with MDD, 32 with BD, and one with

XSL•FO
RenderX

OCD. The participants were required to complete the same survey used to develop the SZ classifier to predict the likelihood of schizophrenia. The external validation process revealed that the SZ classifier achieved a sensitivity of 0.75 but misclassified 55.4% of patients with MDD and 59.4% of patients with BD as having schizophrenia [15].

Estimation of the Prevalence of Schizophrenia by Using the SZ Classifier

A cross-sectional study was conducted using data from a large-scale web-based survey in Japan. The number of participants was set to 750. The sample size was decided using the single population proportion formula ($n=[Z^2 \cdot p(1-p)]/e^2$) by assuming a 95% CI (Z=1.96), 1% margin of error (e=0.01), and 2% proportion of schizophrenia (p=2%) based on previous studies. Participants were randomly recruited from those registered as survey monitors with Rakuten Insight, a large-scale web-based survey incorporated approximately 2.2 million registered members as of September 2022 and provided reliable data with regular quality maintenance [16]. Specifically, the recruited criteria were set to the combination of 3 items of age intervals (20 - 75 years old, divided into 5 distinct age groups), gender (male and female), and regions (13 geographical administrative regions in Japan: Hokkaido, Tohoku, North Kanto, Greater Tokyo Area, Koshinetsu, Hokuriku, Tokai, Kinki excluding Keihanshin, Keihanshin, Chugoku, Shikoku, Kyushu excluding Okinawa, and Okinawa). For each combination category of these 3 items, the number of participants was set to match the composition of the national population as of October 1, 2021, as estimated by the Ministry of Internal Affairs and Communications, Japan. Participants were recruited until the target number for each combination category was reached. The recruitment period was from November 20 to 26, 2023. This was a web-based survey in which only authorized panelists were allowed to respond to the survey to prevent fraudulent registrants from participating. To administer the survey, respondents had to log in to their accounts to prevent multiple entries from the same individual. They were required to answer 38 questions (some questions had several subquestions) that are distributed into 38 pages. Respondents were asked to review their responses and ensure completeness before submission. Specifically, the web-based survey system notified respondents if there were any blank answers. The participation rate for this survey was estimated to be 0.30, and the completion rate was not calculated because if the respondents terminated during the answering process (ie, logged out, closed the web page, etc), they were not counted in the statistics. Incomplete questionnaires or questionnaires that were submitted too soon were not used.

Each participant's demographic information, health-related backgrounds, and physical, psychiatric, and social comorbidities were collected and applied to the SZ classifier, yielding the predicted crude number of schizophrenia cases and crude prevalence of schizophrenia. Textbox 1 shows a summary of the feature variable categories collected from study participants. These variables are the same as those used in developing the SZ classifier. For detailed information on these variables, please refer to Multimedia Appendix 1.

XSL•FO
RenderX

Textbox 1. Summary of categories of feature variables collected from study participants, including demographic, health-related backgrounds, physical comorbidities, psychiatric comorbidities, and social comorbidities.

Demographic, health-related backgrounds, and physical comorbidities:

- Age
- BMI
- Smoking
- Alcohol drinking
- Sports
- Eating behaviors
- Bowel movement
- Restriction in functional capacity
- Self-rated health status
- Physical comorbidities

Psychiatric comorbidities:

- Frequency of sleep medication use
- Bedtime
- Sleep hours and qualities
- Depressive symptoms
- Perceived stress
- Positive reason for living (ikigai)
- Internet use hours per week

Social comorbidities:

- Regularity of health checkups
- Educational background
- Household income
- Type of employment and occupation
- Marital status and family structure
- Social support and social capital

We further refined these crude estimates by excluding false positives and adding false negatives. This was based on the model's internal and external validation results and literature reviews. Details of the model's internal and external validation results are provided in the Multimedia Appendix 1.

False-positive cases included individuals without mental disorders who were misclassified as having schizophrenia and those with other mental disorders who were misclassified as having schizophrenia. The false-positive cases were calculated as follows:

$$Nfp=(Phc\cdot rhc+PMDD\cdot rMDD+PBD\cdot rBD)\cdot n$$

where N_{fp} represents the number of false-positive cases, P_{hc} represents the percentage of individuals who have not had any mental disorders in the population, r_{hc} represents the rate at which the SZ classifier falsely classified individuals without mental disorders as having schizophrenia, P_{MDD} represents the prevalence of MDD in the population, r_{MDD} represents the rate at which the SZ classifier falsely classified individuals with MDD as having schizophrenia, P_{BD} represents the prevalence of BD in the population, r_{BD} represents the rate at which the SZ classifier falsely classified individuals with BD as having schizophrenia, n represents the sample size, which included 750 participants.

The SZ classifier falsely classified 3.1% of the individuals without mental disorders as having schizophrenia [14]. The SZ classifier misclassified 55.4% of the patients with MDD and 59.4% of the patients with BD as having schizophrenia [15]. According to a literature review, 80% of the general population in Japan does not have mental disorders [19]. The point prevalence of MDD and BD was 7.9% [20] and 0.4% [21], respectively; namely, P_{hc}=80%, r_{hc}=3.1%, P_{MDD}=7.9%, r_{MDD}=55.4%, P_{BD}=0.4%, r_{BD}=59.4%, and n=750.

False-negative cases included individuals with schizophrenia but were not classified as schizophrenia cases. The proportion of true-positive to false-negative cases in the external validation

XSL•FO
RenderX

was 46:15 [15]. Hence, the false-negative cases can be calculated as follows:

$$Ntp=Ncrude-Nfp$$

$$Ntp/Nfn=p$$

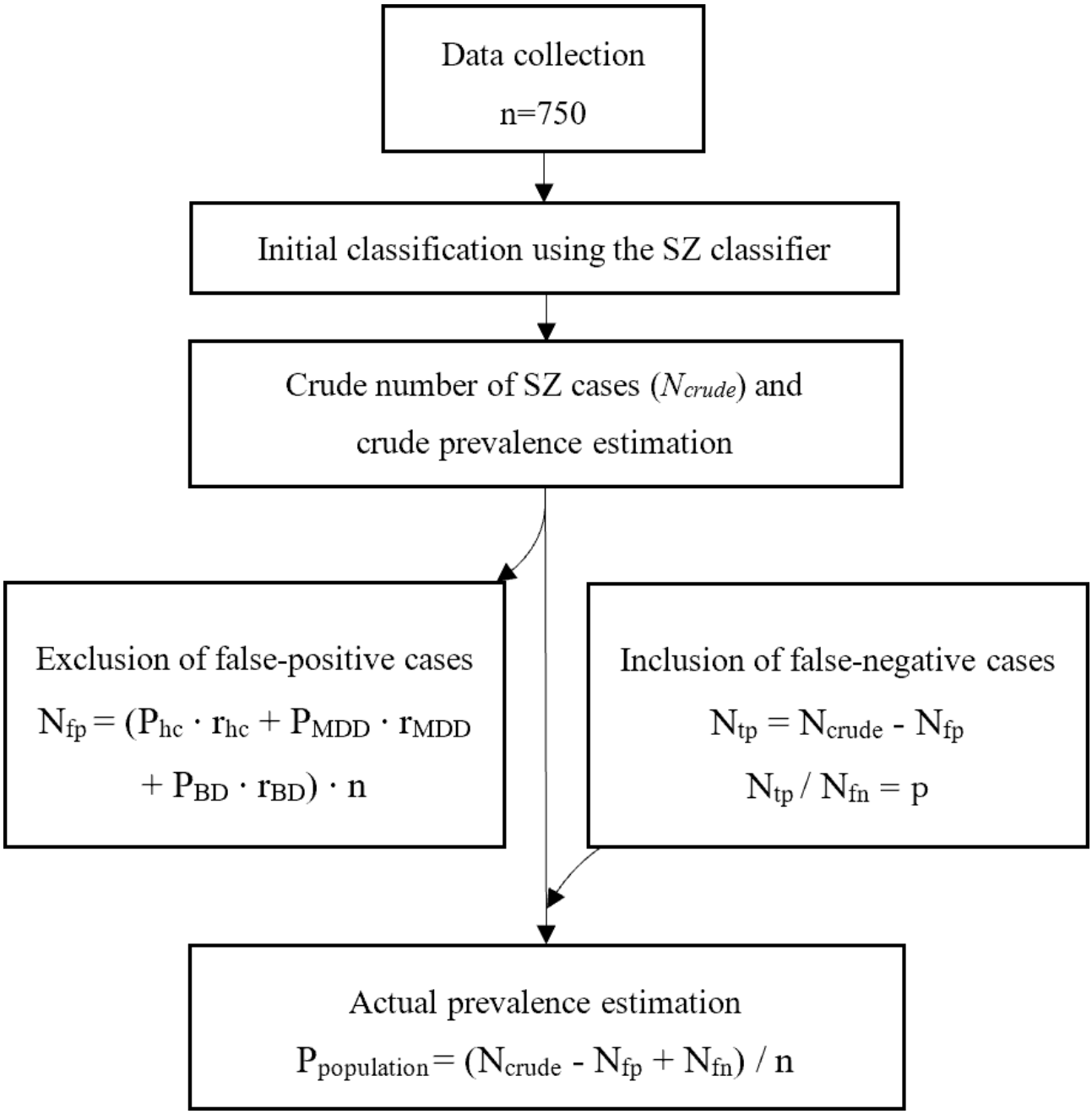
where N_{tp} represents the cases predicted by the SZ classifier, N_{crude} represents the crude number of schizophrenia cases estimated by the SZ classifier, N_{fn} represents the false-negative cases, and p represents the proportion of true-positive cases to false-negative cases in the external validation, which was 46:15 [15].

Finally, the prevalence of schizophrenia in the population could be calculated by

$$Ppopulation=(Ncrude-Nfp+Nfn)/n$$

where $P_{population}$ represents the prevalence of schizophrenia in the population (Figure 1).

Figure 1. Flowchart illustrating the data analysis and refinement process for estimating the prevalence of schizophrenia in the general Japanese population. Data were collected from 750 participants during November 20 to 26, 2023. Participants' demographic information, health-related background, and physical, psychiatric, and social comorbidities were assessed using the schizophrenia classification model to initially estimate the crude number of schizophrenia cases and crude prevalence. The refinement process involved excluding false-positive cases and including false-negative cases, resulting in an actual estimation of schizophrenia prevalence in the general population of Japan.



XSL•FO
RenderX

Statistical Analysis

The 95% CIs were calculated from 10,000 bootstrap replicates. Statistical analyses were performed using Python (version 3.8; Python Software Foundation), with a Jupyter Notebook (Project Jupyter) as a coding notebook.

Ethical Considerations

The study protocol was reviewed and approved by the institutional review board at Fujita Health University (HM23-253). All procedures were performed following the ethical standards of the Declaration of Helsinki. The internet survey agency respected the Act on the Protection of Personal Information in Japan. Informed consent was obtained at the initial web page of the survey, where the purpose of the survey and the instructions to the respondents were displayed. Only those who had read the survey's purpose and instructions and provided informed consent by clicking the agreement button could proceed with the survey. Participants for the external validation of the SZ classifier were diagnosed by experienced psychiatrists. The data were anonymous and deidentified before being provided to researchers. This was a voluntary survey. Participants received Rakuten Points, which can be used for Rakuten services such as web-based or offline shopping. No image data were collected so that there were no concerns about identifying individual participants.

Results

Of the 750 samples analyzed, 62 were classified as schizophrenia cases (N_{crude}). The crude prevalence of schizophrenia in the general population of Japan, as estimated using the SZ classifier, was 8.3% (95% CI 6.6%-10.1%; Table 1). Among these 62 cases, 53 were presumed to be false positives, and 3 were presumed to be false negatives. After adjustments, the actual prevalence of schizophrenia in the general population was estimated to be 1.6% (95% CI 0.7%-2.5%; Table 2).

Table . Results of crude estimates of schizophrenia cases by the schizophrenia classification model (SZ classifier), crude schizophrenia prevalence, false-positive estimates, and true-positive estimates.

	Cases, n	Prevalence, %
Crude estimates of predicted schizophrenia by the SZ classifier	62	8.3 (6.6-10.1)[a]
False positive: no mental disorders misclassified as schizophrenia	18	2.5
False positive: patients with MDD[b] misclassified as schizophrenia	33	4.4
False positive: patients with BD[c] misclassified as schizophrenia	2	0.2
True positive: patients with schizophrenia who were accurately classified as schizophrenia cases	9	1.2

[a]95% CI value.

[b]MDD: major depressive disorder.

[c]BD: bipolar disorder.

Table . Confusion matrix and adjusted prevalence of schizophrenia in the general population of Japan.

		Adjusted cases estimation	
		Schizophrenia	NonSchizophrenia
Predicted cases, n			
	Schizophrenia	9	53
	Nonschizophrenia	3	685
	Total	12	738
Prevalence, % (95%CI)		1.6 (0.7 - 2.5)	—[a]

[a]Not applicable.

Discussion

Principal Results

Our findings indicate that the estimated prevalence of schizophrenia in the general population of Japan using the SZ classifier was 1.6% (95% CI 0.7%-2.5%). This estimate is slightly higher than those in previously published figures [5,6]. This discrepancy could be attributed to several reasons, including the possibility that our SZ classifier estimates lifetime prevalence rather than point prevalence, a potentially more comprehensive disease classification methodology, and calculating the actual prevalence by referencing various prevalence data from diverse published sources.

XSL•FO
RenderX

To ensure the validity of the data used in our study, 3 main points were considered. The first point is the trustworthiness of the data source. The data were collected and managed by Rakuten Insight Panel, a large-scale web-based survey panel. The panel's quality has been maintained by company procedures and has been widely used by research institutions in Japan for population-based studies [16]. The second point is the appropriateness of the data used for model development. All feature variables were decided both through a literature review [17] and studies demonstrating that all these variables effectively differentiate between individuals with and those without schizophrenia [14,15,18]. Additionally, all the schizophrenia cases were defined by using a screening protocol instead of relying solely on the self-reported illness conditions. Schizophrenia cases had to meet the following four criteria: (1) currently having schizophrenia only, schizophrenia and migraine, schizophrenia and a sleep disorder, or schizophrenia, migraine, and a sleep disorder; (2) having experienced auditory hallucinations lasting more than 1 month; (3) not having used stimulants or other illegal drugs and not having been an alcoholic; and (4) having experienced the first auditory hallucination lasting more than 1 month at less than 60 years of age [18]. Third, the high performance of the developed model relies on high-quality data. The robustness and generalizability of the model were confirmed through good internal validation (AUC=0.86) [14] and external validation results (sensitivity=0.75) [15].

The SZ classifier predicts schizophrenia status using various feature variables. Each feature has been reported as being significantly associated with schizophrenia in previous studies [14,15,17,18]. By training the model with these features and applying data from the study participants, the probability of schizophrenia can be calculated. Although this probability may not be highly interpretable for diagnosis on an individual level, it reliably represents the proportion of specific cases within a target group. When the model demonstrates good generalizability, this proportion can be extended to estimate prevalence by considering the false positives and false negatives on a population level.

According to the feature importance ranking estimated by the model, social comorbidities, such as household income, type of employment, educational background, and type of occupation, play significant roles [14]. These factors may have been influenced by previous psychiatric conditions, which could have had long-term effects on the data used in this analysis. For instance, individuals diagnosed with schizophrenia may have had to stop attending school or leave their jobs, affecting their current income and occupation [22]. Consequently, the model might have predicted them as having schizophrenia even though they had recovered, which is consistent with the lifetime prevalence estimates. Several studies have shown that the prevalence of schizophrenia increases when it extends from one point to the lifetime of an individual [4,23].

To the best of our knowledge, this is the first prevalence study to apply an ANN model as a schizophrenia case classifier using web-based survey data from the general population of Japan. ANNs, which are computational models inspired by the inner processes of the human brain, are increasingly used in public health research [24,25]. Technically, ANNs are excellent at capturing complex and nonlinear relationships within large datasets. Hence, models constructed using ANNs perform better than classical methods based on linear equations [26,27]. Several studies have proposed ANN models or machine learning techniques to detect and classify schizophrenia [28,29]. These studies developed models using large amounts of clinical data (ie, magnetic resonance images, positron emission tomography scans, electroencephalography, and subjective patient posture interpretation). These studies have applied machine learning techniques to improve detection and diagnosis and aid in clinical decision-making. Bae et al [30] demonstrated the use of machine learning to detect signs of schizophrenia by analyzing social media texts, supporting the role of machine learning models in improving symptoms monitoring and potentially early detection of schizophrenia.

The use of ANN models to estimate prevalence has been reported in previous studies. For example, suicidal ideation risk prediction models were developed based on demographic, behavioral, and psychological data obtained from computer-based questionnaires. These models were then used to estimate the prevalence of suicidal ideation among Chinese college students. The models demonstrated a prediction accuracy of over 91.7%, and the AUC score increased to 0.96 [31]. Compared with previous studies, our SZ classifier has undergone external validation to demonstrate its generalizability [15]. This study supports the use of machine learning–based methods for prevalence estimation.

This study design may provide evidence to overcome biases encountered in traditional self-identifying surveys due to stigma, poor insight, and lack of access to health professionals experienced in conventional self-identifying surveys [32,33]. Stocking et al [34] reported that they first used a mathematical model to explore the correlation between several comorbidities and a specific illness in a small sample, and then applied this correlation to a larger population to identify previously underreported cases. They noted that this approach is effective for addressing the underreporting of various chronic conditions associated with social stigma, challenges in detection and treatment, or inconsistent treatment engagement (eg, mental disorders, obesity, hypertension, sleep disorders, etc). In this study, schizophrenia cases were estimated by incorporating general information, including demographics, health-related backgrounds, and physical, psychological, and social comorbidities, to calculate the likelihood of schizophrenia. This approach encompasses a broader range of patients in the general population than previous methodologies that relied on clinical records or typical schizophrenia symptoms. In contrast to a diagnosis that depends on the self-reporting of typical schizophrenia symptoms, which requires participants to openly acknowledge symptoms that may suggest schizophrenia, our methodology provides a more objective assessment. Patients with schizophrenia are often stigmatized and not widely accepted in society, leading to embarrassment and hesitation in disclosing their symptoms and avoiding negative perceptions. Stigmatizing attitudes related to mental illness have been reported to prevent individuals from seeking counseling [35] and harm the self-esteem of people with serious mental illnesses

XSL•FO
RenderX

[36]. Additionally, the wording of questionnaires that focus on negative aspects may cause respondents to feel negative and skip or conceal their actual symptoms, resulting in underreported results in self-identifying surveys [37,38]. By using general information, our SZ classifier mitigated these biases, potentially providing a more comprehensive estimate of patients with schizophrenia in Japan.

The SZ classifier has the advantage of potentially detecting patients with schizophrenia who do not perceive subjective symptoms. Hoffman et al [39] reported that a common reason why people with schizophrenia lack insight into their illness is that their ability to objectively assess their own pathological experiences is impaired. By using objective indicators such as demographics, health-related backgrounds, and physical, psychological, and social comorbidities, the SZ classifier showed a potential ability of detecting patients who do not perceive subjective symptoms and estimating the prevalence of schizophrenia in the population. This study may offer unique advantages over previous estimation methods based solely on medical data.

The higher prevalence estimated in our study suggests that schizophrenia might be more common than previously believed in the Japanese population. This finding confirmed our hypothesis regarding the underestimation of schizophrenia cases in the Japanese population. These findings have significant implications for public health management and resource allocation. Greater knowledge about the actual prevalence of schizophrenia can result in enhanced mental health services and increased support systems for managing this public health concern.

Limitations

Our study has several limitations. First, the actual prevalence of schizophrenia was calculated by referencing various prevalence data from diverse published sources, including the prevalence of individuals who did not have any mental disorder, the prevalence of those who had MDD, and the prevalence of those who had BD. Using data from different studies, which may have used different methodologies and population samples, causes uncertainty in the data, thereby affecting the estimated values. Second, although the data used in this study were collected from an extensive web-based survey that efficiently integrates large amounts of data, it may introduce a selection bias toward those with internet access. Those who are less likely to participate in web-based surveys, such as individuals without internet access, institutionalized patients, patients with severe comorbidities, older adults, and poor people, might be underrepresented. Third, the SZ classifier, designed to capture a broad range of patients, may include individuals with subclinical symptoms or comorbid conditions similar to schizophrenia. Despite efforts to rule out incorrectly classified cases, there is a chance that other mental disorders will not be sufficiently considered. This inclusion may lead to an overestimation of the true prevalence of schizophrenia. Fourth, for mental disorders comorbid with schizophrenia symptoms (eg, schizoaffective disorder, characterized by symptoms of both schizophrenia and mood disorders, etc), these cases were not included in the SZ classifier validation. Also, due to the complexity of mental disorders, the diagnosis of a disease without confirmation from a physician may seem arbitrary. Therefore, these results should be interpreted with caution.

Conclusions

This study applied a pretrained SZ classifier to estimate the prevalence of schizophrenia in a Japanese population. The estimated results are slightly higher than those of previous studies. Considering the shortcomings of focusing on a single type of mental disorder and the need to increase the sample size, this study shows promise for prevalence-estimation research. Potential future research directions may focus on developing models to classify other mental disorders, such as BD and MDD, as well as estimate prevalence. These findings are expected to support the versatility of pretrained models in epidemiological research.

Acknowledgments

This study was funded by the Ministry of Health, Labour and Welfare, Japan (JPMH21GC1018). The findings and conclusions of this article are the sole responsibility of the authors and do not represent the official views of the research funders. The authors attest that there was no use of generative artificial intelligence technology in the generation of text, figures, or other informational content of this manuscript.

Data Availability

The datasets analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: PC, YH, and AO
Data curation: MM and AO
Data interpretation: PC and YH
Data analysis: PC and YH
Investigation: MM and AO
Methodology: AO and YH
Supervision: AO

XSL•FO
RenderX

Writing—original draft: PC
Writing—review and editing: YH, KS, TK, YL, ST, NI, and AO

Conflicts of Interest

None declared.

Multimedia Appendix 1
Inclusion criteria, variable definition, and model validation results.
[DOCX File, 36 KB - formative_v9i1e66330_app1.docx]

References

1. Marder SR, Cannon TD. Schizophrenia. N Engl J Med 2019 Oct 31;381(18):1753-1761. [doi: 10.1056/NEJMra1808803] [Medline: 31665579]
2. Schizophrenia. World Health Organization. 2022 Jan 10. URL: https://www.who.int/news-room/fact-sheets/detail/schizophrenia [accessed 2024-05-17]
3. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry 2022 Feb;9(2):137-150. [doi: 10.1016/S2215-0366(21)00395-3] [Medline: 35026139]
4. Simeone JC, Ward AJ, Rotella P, Collins J, Windisch R. An evaluation of variation in published estimates of schizophrenia prevalence from 1990—2013: a systematic literature review. BMC Psychiatry 2015 Aug 12;15:193. [doi: 10.1186/s12888-015-0578-7] [Medline: 26263900]
5. Sado M, Inagaki A, Koreki A, et al. The cost of schizophrenia in Japan. Neuropsychiatr Dis Treat 2013;9:787-798. [doi: 10.2147/NDT.S41632] [Medline: 23785238]
6. Baba K, Guo W, Chen Y, Nosaka T, Kato T. Burden of schizophrenia among Japanese patients: a cross-sectional National Health and Wellness Survey. BMC Psychiatry 2022 Jun 18;22(1):410. [doi: 10.1186/s12888-022-04044-5] [Medline: 35717149]
7. Rodríguez-Rivas ME, Cangas AJ, Cariola LA, Varela JJ, Valdebenito S. Innovative technology-based interventions to reduce stigma toward people with mental illness: systematic review and meta-analysis. JMIR Serious Games 2022 May 30;10(2):e35099. [doi: 10.2196/35099] [Medline: 35635744]
8. Gray AJ. Stigma in psychiatry. J R Soc Med 2002 Feb;95(2):72-76. [doi: 10.1177/014107680209500205] [Medline: 11823548]
9. Kamaradova D, Latalova K, Prasko J, et al. Connection between self-stigma, adherence to treatment, and discontinuation of medication. Patient Prefer Adherence 2016;10:1289-1298. [doi: 10.2147/PPA.S99136] [Medline: 27524884]
10. Vrbova K, Prasko J, Holubova M, et al. Self-stigma and schizophrenia: a cross-sectional study. Neuropsychiatr Dis Treat 2016;12:3011-3020. [doi: 10.2147/NDT.S120298] [Medline: 27920538]
11. Konstantakopoulos G. Insight across mental disorders: a multifaceted metacognitive phenomenon. Psychiatriki 2019;30(1):13-16. [doi: 10.22365/jpsych.2019.301.13] [Medline: 31115349]
12. Ando S, Yamaguchi S, Aoki Y, Thornicroft G. Review of mental-health-related stigma in Japan. Psychiatry Clin Neurosci 2013 Nov;67(7):471-482. [doi: 10.1111/pcn.12086] [Medline: 24118217]
13. Goldner EM, Hsu L, Waraich P, Somers JM. Prevalence and incidence studies of schizophrenic disorders: a systematic review of the literature. Can J Psychiatry 2002 Nov;47(9):833-843. [doi: 10.1177/070674370204700904] [Medline: 12500753]
14. He Y, Matsunaga M, Li Y, et al. Classifying schizophrenia cases by artificial neural network using Japanese web-based survey data: case-control study. JMIR Form Res 2023 Nov 15;7:e50193. [doi: 10.2196/50193] [Medline: 37966882]
15. He Y, Sakuma K, Kishi T, et al. External validation of a machine learning model for schizophrenia classification. J Clin Med 2024 May 17;13(10):2970. [doi: 10.3390/jcm13102970] [Medline: 38792511]
16. Rakuten Insight. Rakuten Insight, Inc. URL: https://insight.rakuten.co.jp/en/aboutus.html [accessed 2024-05-30]
17. He Y, Tanaka A, Kishi T, et al. Recent findings on subjective well-being and physical, psychiatric, and social comorbidities in individuals with schizophrenia: a literature review. Neuropsychopharmacol Rep 2022 Dec;42(4):430-436. [doi: 10.1002/npr2.12286] [Medline: 35916310]
18. Matsunaga M, Li Y, He Y, et al. Physical, psychiatric, and social comorbidities of individuals with schizophrenia living in the community in Japan. Int J Environ Res Public Health 2023 Feb 28;20(5):4336. [doi: 10.3390/ijerph20054336] [Medline: 36901345]
19. Nishi D, Ishikawa H, Kawakami N. Prevalence of mental disorders and mental health service use in Japan. Psychiatry Clin Neurosci 2019 Aug;73(8):458-465. [doi: 10.1111/pcn.12894] [Medline: 31141260]
20. Hoshino E, Ohde S, Rahman M, Takahashi O, Fukui T, Deshpande GA. Variation in somatic symptoms by Patient Health Questionnaire-9 depression scores in a representative Japanese sample. BMC Public Health 2018 Dec 27;18(1):1406. [doi: 10.1186/s12889-018-6327-3] [Medline: 30587173]

XSL•FO
RenderX

21. Kato T. Bipolar disorder: from pathophysiology to treatment. Juntendo Iji Zasshi 2022;68(1):17-24. [doi: 10.14789/jmj.JMJ21-0026-R] [Medline: 38911011]
22. Hakulinen C, McGrath JJ, Timmerman A, et al. The association between early-onset schizophrenia with employment, income, education, and cohabitation status: nationwide study with 35 years of follow-up. Soc Psychiatry Psychiatr Epidemiol 2019 Nov;54(11):1343-1351. [doi: 10.1007/s00127-019-01756-0] [Medline: 31456027]
23. Bhugra D. The global prevalence of schizophrenia. PLoS Med 2005 May;2(5):e151. [doi: 10.1371/journal.pmed.0020151] [Medline: 15916460]
24. Lanillos P, Oliva D, Philippsen A, Yamashita Y, Nagai Y, Cheng G. A review on neural network models of schizophrenia and autism spectrum disorder. Neural Netw 2020 Feb;122:338-363. [doi: 10.1016/j.neunet.2019.10.014] [Medline: 31760370]
25. Razavi M, Ziyadidegan S, Mahmoudzadeh A, et al. Machine learning, deep learning, and data preprocessing techniques for detecting, predicting, and monitoring stress and stress-related mental disorders: scoping review. JMIR Ment Health 2024 Aug 21;11:e53714. [doi: 10.2196/53714] [Medline: 39167782]
26. Masegosa AR, Cabañas R, Langseth H, Nielsen TD, Salmerón A. Probabilistic models with deep neural networks. Entropy (Basel) 2021 Jan 18;23(1):117. [doi: 10.3390/e23010117] [Medline: 33477544]
27. He Y, Sun Q, Matsunaga M, Ota A. Can feature structure improve model's precision? A novel prediction method using artificial image and image identification. JAMIA Open 2024 Apr;7(1):ooae012. [doi: 10.1093/jamiaopen/ooae012] [Medline: 38348347]
28. Lai JW, Ang CKE, Acharya UR, Cheong KH. Schizophrenia: a survey of artificial intelligence techniques applied to detection and classification. Int J Environ Res Public Health 2021 Jun 5;18(11):6099. [doi: 10.3390/ijerph18116099] [Medline: 34198829]
29. Verma S, Goel T, Tanveer M, Ding W, Sharma R, Murugan R. Machine learning techniques for the schizophrenia diagnosis: a comprehensive review and future research directions. J Ambient Intell Human Comput 2023 May;14(5):4795-4807. [doi: 10.1007/s12652-023-04536-6]
30. Bae YJ, Shim M, Lee WH. Schizophrenia detection using machine learning approach from social media content. Sensors (Basel) 2021 Sep 3;21(17):5924. [doi: 10.3390/s21175924] [Medline: 34502815]
31. Liao S, Wang Y, Zhou X, et al. Prediction of suicidal ideation among Chinese college students based on radial basis function neural network. Front Public Health 2022;10:1042218. [doi: 10.3389/fpubh.2022.1042218] [Medline: 36530695]
32. Supina AL, Patten SB. Self-reported diagnoses of schizophrenia and psychotic disorders may be valuable for monitoring and surveillance. Can J Psychiatry 2006 Mar;51(4):256-259. [doi: 10.1177/070674370605100407] [Medline: 16629350]
33. Gould F, McGuire LS, Durand D, et al. Self-assessment in schizophrenia: accuracy of evaluation of cognition and everyday functioning. Neuropsychology 2015 Sep;29(5):675-682. [doi: 10.1037/neu0000175] [Medline: 25643212]
34. Stocking A, Duncan I, Huynh N. Estimating underdiagnosis of patients in chronically ill populations. N Am Actuar J 2023 Dec 18:1-10. [doi: 10.1080/10920277.2023.2281471]
35. Vogel DL, Wade NG, Hackler AH. Perceived public stigma and the willingness to seek counseling: the mediating roles of self-stigma and attitudes toward counseling. J Couns Psychol 2007;54(1):40-50. [doi: 10.1037/0022-0167.54.1.40]
36. Link BG, Struening EL, Neese-Todd S, Asmussen S, Phelan JC. Stigma as a barrier to recovery: the consequences of stigma for the self-esteem of people with mental illnesses. Psychiatr Serv 2001 Dec;52(12):1621-1626. [doi: 10.1176/appi.ps.52.12.1621] [Medline: 11726753]
37. Subramaniam M, Abdin E, Vaingankar JA, et al. Lifetime prevalence and correlates of schizophrenia and other psychotic disorders in Singapore. Front Psychiatry 2021;12:650674. [doi: 10.3389/fpsyt.2021.650674] [Medline: 33776823]
38. Karam EG, Sampson N, Itani L, et al. Under-reporting bipolar disorder in large-scale epidemiologic studies. J Affect Disord 2014 Apr;159:147-154. [doi: 10.1016/j.jad.2014.01.011] [Medline: 24679403]
39. Hoffman RE, McGlashan TH. Corticocortical connectivity, autonomous networks, and schizophrenia. Schizophr Bull 1994;20(2):257-261. [doi: 10.1093/schbul/20.2.257] [Medline: 8085129]

Abbreviations

ANN: artificial neural network
AUC: area under the receiver operating characteristic curve
BD: bipolar disorder
MDD: major depressive disorder
OCD: obsessive-compulsive disorder
SZ classifier: schizophrenia classification model

XSL•FO
RenderX

Edited by A Mavragani; submitted 10.09.24; peer-reviewed by D Abdel-Hady, SE Khamlichi; revised version received 05.01.25; accepted 07.01.25; published 29.01.25.

Please cite as:
Choomung P, He Y, Matsunaga M, Sakuma K, Kishi T, Li Y, Tanihara S, Iwata N, Ota A
Estimating the Prevalence of Schizophrenia in the General Population of Japan Using an Artificial Neural Network–Based Schizophrenia Classifier: Web-Based Cross-Sectional Survey
JMIR Form Res 2025;9:e66330
URL: https://formative.jmir.org/2025/1/e66330
doi:10.2196/66330

© Pichsinee Choomung, Yupeng He, Masaaki Matsunaga, Kenji Sakuma, Taro Kishi, Yuanying Li, Shinichi Tanihara, Nakao Iwata, Atsuhiko Ota. Originally published in JMIR Formative Research (https://formative.jmir.org), 29.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Measuring Self-Reported Well-Being of Physicians Using the Well-Being Thermometer: Cohort Study

Marios Adamou[1], MD, PhD; Sarah L Jones[2], PhD; Niki Kyriakidou[3], PhD; Andrew Mooney[4], PhD; Shriti Pattani[5]; Matthew Roycroft[6], MEd

[1]University of Huddersfield, Huddersfield, United Kingdom

[2]South West Yorkshire Partnership NHS Foundation Trust, Belle Isle Health Park, Wakefield, United Kingdom

[3]Leeds Beckett University, Leeds, United Kingdom

[4]Renal Unit, St James's Hospital, Leeds, United Kingdom

[5]London North West University Healthcare NHS Trust, London, United Kingdom

[6]Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, United Kingdom

Corresponding Author:
Sarah L Jones, PhD
South West Yorkshire Partnership NHS Foundation Trust, Belle Isle Health Park, Wakefield, United Kingdom

Abstract

Background: Advancements in medical science have focused largely on patient care, often overlooking the well-being of health care professionals (HCPs). This oversight has consequences; not only are HCPs prone to mental and physical health challenges, but the quality of patient care may also endure as a result. Such concerns are also exacerbated by unprecedented crises like the COVID-19 pandemic. Compared to other sectors, HCPs report high incidence of stress, depression, and suicide, among other challenging factors that have a significant negative impact on their well-being.

Objective: Given these substantial concerns, the development of a tool specifically designed to be used in clinical settings to measure the well-being of HCPs is essential.

Methods: A United Kingdom–based cross-sectional pilot study was carried out to measure self-reported well-being in a cohort of 148 physicians, using the newly developed well-being thermometer. The aim of the tool is to allow respondents to develop an individual sense of "well-being intelligence" thus supporting HCPs to have better insight and control over their well-being and allow insights into how to manage it. The tool consists of 5 well-being domains—health, thoughts, emotions, spiritual, and social. Each domain can be measured individually or combined to produce an overall well-being score.

Results: The tool demonstrated good internal consistency; the Cronbach α in this study was 0.84 for the total scale.

Conclusions: Results from this cohort demonstrated that the well-being thermometer can be used to gather intelligence of staff well-being. This is a promising new tool that will assist HCPs to recognize their own well-being needs and allow health care organizations to facilitate change in policies and practices to reflect a better understanding of staff well-being.

(JMIR Form Res 2025;9:e54158) doi:10.2196/54158

KEYWORDS

well-being; health care professionals; mental health; well-being thermometer; health care

Introduction

Medicine has developed greatly in relation to disease control and health interventions, yet it is questioned if health care professionals (HCPs) are fully aware of their own well-being, and the toll poor levels of well-being may have on their lives [1]. HCPs face unique challenges that could be damaging to their mental and physical health [2-5], and given the recent challenges of the COVID-19 pandemic [6,7], plus the current public health crisis, monitoring the well-being of our HCPs is vital for the mental health of the workforce.

Poor well-being is often reported in HCP cohorts [8-11], leading to serious health consequences and reduced quality of life for those affected [12]. In economic terms, higher staff turnover and sickness absence are the consequence [13], potentially resulting in poor quality of care for patients [14]. HCPs consistently report higher levels of sickness absence, job dissatisfaction, and stress compared to other work sectors [15-20]. A recent systematic review found moderate to high levels of stress, anxiety, depression, sleep disturbance, and burnout reported by frontline health care employees compared to workers in other sectors [21]. Moreover, a survey conducted in the United Kingdom (pre–COVID-19 pandemic) reported that, of public sector workers, those who work for the NHS

XSL•FO
RenderX

recorded the most stress [22]. There is also increasing morbidity in HCPs compared to the general population [23]. Reasons for these higher levels of stress are multifactorial, including lack of staff, increasing workload, patient expectation [23], emotional demands of the role [24], as well as others [25]. Research is supported by latest figures showing anxiety, stress, and depression are among the most reported reason for sickness absence in the NHS, accounting for 24.9% of absence in the year preceding September 2022 [26]. The issues pertaining to employee well-being have been acknowledged by the development of the "NHS workforce health and wellbeing framework" [27], although the outcomes of this are unclear at present. While there is extensive research around employee well-being [28], there are calls for deeper consideration of psychological needs of HCPs specifically [21,23,29]. The benefits of considering this not only help on an individual level, but also aid health care organizations with the economic cost of staff burnout [13].

Conceptualizing well-being is not straightforward. It is a complex construct, subject to much academic debate [30]. Current models of well-being tend to be grounded on concepts of mental illness or psychological functioning [31]. However, well-being is best described as multifaceted; measured by a range of subjective and objective concepts rather than one single notion [32]. Well-being is often used interchangeably with the term "mental health." For instance, according to the World Health Organization (WHO), "mental health" is "a state of well-being in which every individual realizes his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community" [33]. It is important to note that well-being is different from the terms "quality of life" or "health-related quality of life," terms that are primarily used to measure a person's perspective of their own life within a cultural context [34]. Furthermore "happiness" is not conducive to well-being. Happiness is often tied to external factors such as pursuit and fulfillment of life goals or life events, rather than a holistic concept such as well-being [35,36]. In terms of how we define well-being here, there is no commonly accepted definition. Subsequently there is no universally accepted approach to measure it. Instead, studies of well-being are often ambiguous in their approach and theoretical underpinnings [37]. A systematic review of self-report measures for assessing well-being found that while the 2 main theories referred to were Diener's model of subjective well-being and the WHO definition of health, authors were very rarely obvious about how theory had influenced the development of their tool or study. Further, argument suggests that the 2 most popular scales, The Warwick-Edinburgh Mental Wellbeing Scale and the WHO Five well-being index fail to capture the holistic nature of well-being [37].

Conceptualizing well-being is complex and often grounded in various subjective and objective factors. Despite various studies and tools aiming to measure well-being, none specifically target HCPs in a comprehensive manner. Adamou et al [31] developed a new theoretical framework of well-being specifically with HCPs in mind. The development of the well-being thermometer consisted of a three-step formation: (1) understanding the concept of well-being from existing literature and tools, (2) constructing a new framework of well-being, and (3) devising a tool to measure it. See Adamou et al [31] for further discussion of the conceptual development of the well-being thermometer. This study aims to pilot a new instrument, the well-being thermometer, specifically designed for HCPs. This tool aims to enhance individual "well-being intelligence," helping HCPs better manage their mental and physical health.

Methods

The Well-Being Thermometer

Developed by Adamou et al [31], this tool incorporates 5 well-being domains—health, thoughts, emotions, spiritual, and social. It aims to provide a comprehensive yet individualized snapshot of well-being, thereby allowing for targeted interventions.

Each of the domains comprises 5 items related to that domain. Each domain can be reviewed individually or collectively providing the individual with a score, allowing reflection of overall and domain-specific well-being. The health domain includes items which relate to the physical and eating health. The thoughts domain relates to mental health. Emotions domain measures the experience of emotions such as joy, satisfaction, and frustration. Spiritual health aligns with the connection with oneself and the meaning of life. The social domain relates to the experience of positive relationships and social networks. Total score can range from 0 to 25, with a score of 25 being the highest level of well-being. Example statements include "I tend to dwell on things more than I should" or "I feel there is a lot to enjoy in life." See Adamou et al [31] for further details.

Participants and Procedure

Doctors were recruited to participate in the well-being survey through verbal advertisements at 2 separate events—a diploma course in occupational medicine organized by the Royal Society of Public Health and a Regional (Yorkshire) Conference for Physicians organized by the Royal College of Physicians. At each event, information was provided about the survey's purpose and the automatic feedback participants would receive upon completion. The survey was administered using an online platform, and opportunity sampling was used to gather the cohort. Participants received an email containing a link to complete the survey online. By completing the survey, respondents were informed that they were consenting to have their data included in the research.

Ethical Considerations

This project was granted ethics approval in line with the Research Ethics Policy and Procedures at Leeds Beckett University (138110). Consent for participation was implicitly granted by the subjects through their active engagement with the survey, after being fully informed about the study's methodology and purpose via the online platform. Moreover, the study ensured the privacy and confidentiality of the participants by anonymizing or deidentifying all data used in the research. No compensation was provided to participants, aligning with the study's observational and noninterventional design.

XSL•FO
RenderX

Statistical Analysis

SPSS (version 29; IBM Corp) was used for statistical analyses. The Kolmogorov-Smirnov Test for Goodness of Fit determined data deviated significantly from normal distribution $P<.05$; therefore, nonparametric analysis was reported.

Results

Demographics

The cohort consisted of 148 physicians (without missing data), 68 (46%) recorded female sex, 78 (53%) reported male sex, and 2 (1%) respondents did not want to disclose gender. Age was recorded in category format, with age range of 40 - 44 years recorded most frequently. Ages ranged from 20 to 69 years. Two respondents chose not to disclose age (see Table 1). Respondents recorded level and speciality pertaining to their profession. See Table 1 for full demographic details.

Table 1 shows the demographic information of the 148 participants who participated in this cross-sectional United Kingdom–based study exploring the validity of the well-being thermometer using survey data. Details include self-reported gender, age, profession level, and speciality pertaining to their medical career.

XSL•FO
RenderX

Table . Respondent demographics (N=148).

Demographic information		Values, n (%)
Sex		
	Female	68 (45.9)
	Male	78 (52.7)
	Prefer not to say	2 (1.4)
	Missing	0 (0)
Age (years)		
	20 - 24	2 (1.4)
	25 - 29	4 (2.7)
	30 - 34	23 (15.5)
	35 - 39	16 (10.8)
	40 - 44	33 (22.3)
	45 - 49	21 (14.2)
	50 - 54	22 (14.9)
	55 - 59	13 (8.8)
	60 - 64	7 (4.7)
	65 - 69	5 (3.4)
	Prefer not to say	2 (1.4)
	Missing	0 (0)
Level		
	Foundation doctor	3 (2)
	Core trainee	5 (3.4)
	Higher specialty trainee	20 (13.5)
	SAS[b] or nontraining grade doctors	15 (10.1)
	Consultant or GP[a]	91 (61.5)
	Prefer not to say	5 (3.4)
	Other	9 (6.1)
	Missing	0 (0)
Specialty		
	Foundation programme	1 (7)
	Core medical training or internal medicine stage 1	3 (2)
	Acute internal medicine	12 (8.1)
	Cardiology	8 (5.4)
	Endocrinology and diabetes	4 (2.7)
	Gastroenterology	5 (3.4)
	General internal medicine	1 (0.7)
	Genitourinary medicine	11 (7.4)
	Geriatric medicine	33 (22.3)
	GP	22 (14.9)
	Infectious diseases	3 (2)
	Medical oncology	1 (0.7)
	Palliative medicine	5 (3.4)

XSL•FO
RenderX

Demographic information		Values, n (%)
	Rehabilitation medicine	1 (0.7)
	Renal medicine	2 (1.4)
	Respiratory medicine	14 (9.5)
	Rheumatology	3 (2)
	Sport and exercise medicine	1 (0.7)
	Stroke medicine	4 (2.7)
	Not applicable	3 (2)
	Prefer not to say	6 (4.1)
	Other	5 (3.4)
	Missing	0 (0)

[a]GP: general practitioner.

[b]SAS: specialty and specialist.

Well-Being Scores

A total of 148 participants in this cross-sectional United Kingdom–based study recorded an overall median score of 18 (IQR 14-22) on the well-being thermometer.

Table 2 shows the cross-sectional median and IQR score recorded by the 148 participants of the United Kingdom–based study of the well-being thermometer. Scores for individual well-being domains ranged between 3 and 4.

Table . Scores by domain.

Domain	Median (IQR)
Health	3 (2-4)
Social	3.5 (3-5)
Thoughts	4 (2-5)
Emotions	4 (3-5)
Spiritual	4 (3-5)
All domains	18 (14-22)

There was no significant effect of age, profession level, or speciality on well-being scores, both overall and domain specific.

Gender-Based Analysis

For the United Kingdom–based cross-sectional study exploring the well-being thermometer, organized by gender (N=148), the median score for men (n=78) was 20 (IQR 16-23) compared to 16 (IQR 12-20.5) for women (n=68).

Table 3 shows the median and IQR scores recorded by the sample (N=148) on individual domains (health, social, thoughts, emotions, and spiritual) of the well-being thermometer, organized by gender. The data are derived from a cross-sectional United Kingdom–based study exploring the validity of the well-being thermometer. Scores ranged between 4 and 5 for men and 3 to 4 for women.

Men reported significantly higher levels of well-being than women overall (U=1593; P=.002). In terms of specific well-being domains, men reported higher levels of well-being on health (U=1979.5; P=.015), thoughts (U=1948.5; P=.019), emotions (U=1971; P=.013), and spiritual (U=1914; P=.014) domains, compared to women. There was also a trend for men to score higher on levels of social well-being (U=2109.5; P=.05).

XSL•FO
RenderX

Table . Domain scores by gender.

	Male (n=78)	Female (n=68)
Domain	Median (IQR)	Median (IQR)
Health	4 (3-5)	3 (1.5-4)
Social	4 (3-5)	3 (3-4)
Thoughts	4 (2-5)	3 (2-5)
Emotions	4 (3-5)	3 (2-4)
Spiritual	5 (4-5)	4 (3-5)
All domains	20 (7)	16 (8.5)

Correlation

Table 4 demonstrates that Spearman rho identified positive relationships between scores on individual domains on the well-being thermometer in this cross-sectional United Kingdom–based study of HCPs (N=148). Suggesting that higher scores on 1 domain of well-being was reflected in other domains of well-being.

Table . Spearman rho.

		Total for health domain, n	Total for social domain, n	Total for thoughts domain, n	Total for emotions domain, n	Total for spiritual domain, n
Total for health domain						
	Correlation coefficient	—[a]	0.399[b]	0.376[b]	0.468[b]	0.400[b]
	Significance (2-tailed)	—	.00	.00	.00	.00
	n	—	144	143	144	142
Total for social domain						
	Correlation coefficient	0.399[b]	—	0.492[b]	0.453[b]	0.453[b]
	Significance (2-tailed)	.000	—	.000	.195	.000
	n	144	—	142	144	141
Total for thoughts domain						
	Correlation coefficient	0.376[b]	0.492[b]	—	0.640[b]	0.537[b]
	Significance (2-tailed)	.000	.000	—	.000	.000
	n	143	142	—	143	142
Total for emotions domain						
	Correlation coefficient	0.468[b]	0.453[b]	0.640[b]	—	0.534[b]
	Significance (2-tailed)	.000	.000	.000	—	.000
	n	144	144	143	—	143
Total for spiritual domain						
	Correlation coefficient	0.400[b]	0.453[b]	0.537[b]	0.534[b]	—
	Significance (2-tailed)	.000	.000	.000	.000	—
	n	142	141	142	143	—

[a]Not applicable.

[b]Correlation is significant at the .01 level (2-tailed).

XSL•FO
RenderX

Cronbach α Analysis

Cronbach α values of 0.7 are considered high levels of internal consistency (DeVillis [38]; Kline [39]). Values above 0.5 are acceptable (Bowling [40]; Schmitt [41]). The scale had a high level of internal consistency, as determined by a Cronbach α of 0.872. Value would improve to α=0.878 if question 7 (I drink more alcohol than would be considered healthy) was removed.

Discussion

Principal Results

The aim of this study was to pilot a new instrument, the well-being thermometer. The tool was specifically designed for use with HCPs and was piloted here with a sample of physicians. In this cross-sectional study the well-being thermometer demonstrated a good level of internal consistency. It was evident that higher scores in 1 well-being domain correlated with higher scores on other well-being domains. The factors of age, speciality, and professional level had no significant effect on well-being. Results suggest that the well-being thermometer has the potential to be a useful and informative tool, both within clinical settings and on an individual level.

Interpretations

The purpose of the well-being thermometer is to enhance individual "well-being intelligence." To aid HCPs to better manage their mental and physical health, and to measure well-being in a more holistic framework than popular scales have allowed [31]. The well-being thermometer has shown to be a valid tool for measuring well-being in HCPs. Subsequently, with better understanding of well-being, services can offer better interventions, protection, and help to their workforce, and use this information to influence policy. The well-being thermometer also allows HCPs to identify elements of their individual well-being which may need attention.

Interestingly, results from this study demonstrated that there was a difference in well-being levels between the sexes, with men reporting higher levels of well-being than women overall, but also on the health, thoughts, emotions, and spiritual domains. This variable was considered important to explore here, as men have often reported higher levels of subjective well-being than women in numerous previous studies [21,29,42-44]. With best evidence derived from a large-scale study of 6397 HCPs, where men reported significantly greater level of overall well-being than women [45]. The findings from this study follow this trend.

Importantly, the well-being thermometer can be used to gather intelligence of staff well-being to facilitate change in policies and practices across health care organizations [31]. While better well-being is a valuable goal in its own right, HCPs should be a priority target because their roles require frontline responsibilities toward the general public and vulnerable populations. We have seen advancements in medical science largely focused on patient care, often overlooking the well-being of HCPs. This oversight has consequences—not only are HCPs prone to mental and physical health challenges, but the quality of patient care also endures as a result. Such concerns are exacerbated by unprecedented crises like the COVID-19 pandemic. HCPs are pressured with dealing with public health crisis in real time, in which difficult moral decision-making is associated with significant stress, lack of control, and feelings of fear [46]. Troublingly, the prevalence of burnout, depression, and suicide is high for this group [47-49]. Compromised well-being among HCPs leads to medical errors, reduced patient safety, high rates of staff turnover, increased absence due to sickness, and diminished patient care [12,14,50,51]. Thus, health care organizations have a responsibility to support and protect staff well-being for both patient and staff safety [52] and the well-being thermometer has shown it can be a useful tool in supporting this objective.

Limitations

Further work pertaining to the well-being thermometer should be conducted with additional demographic information such as race and ethnicity, as this information was not captured here, and is a limitation of the study. A focus on threshold analysis is also necessary to gain further insight. Also, future work with larger samples is recommended.

Conclusions

Conceptualizing well-being remains complex and often grounded in various subjective and objective factors, and despite various studies and tools aiming to measure well-being, none specifically target HCPs in a comprehensive manner. The aim of this study was to pilot a new instrument, the well-being thermometer, specifically designed for HCPs. This study demonstrated that the well-being thermometer can be used to gather intelligence of staff well-being. This information is required to facilitate much needed change in policies and practices across health care organizations. A specific focus on staff well-being will benefit not only HCPs, but also those who trust in our organizations to provide safe and efficient health care services.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Shanafelt TD, Sloan JA, Habermann TM. The well-being of physicians. Am J Med 2003 Apr 15;114(6):513-519. [doi: 10.1016/s0002-9343(03)00117-7] [Medline: 12727590]

XSL•FO
RenderX

2. Lomas T, Medina JC, Ivtzan I, Rupprecht S, Eiroa-Orosa FJ. A systematic review of the impact of mindfulness on the well-being of healthcare professionals. J Clin Psychol 2018 Mar;74(3):319-355. [doi: 10.1002/jclp.22515] [Medline: 28752554]
3. Gao YQ, Pan BC, Sun W, Wu H, Wang JN, Wang L. Anxiety symptoms among Chinese nurses and the associated factors: a cross sectional study. BMC Psychiatry 2012 Sep 14;12:141. [doi: 10.1186/1471-244X-12-141] [Medline: 22978466]
4. Khamisa N, Oldenburg B, Peltzer K, Ilic D. Work related stress, burnout, job satisfaction and general health of nurses. Int J Environ Res Public Health 2015 Jan 12;12(1):652-666. [doi: 10.3390/ijerph120100652] [Medline: 25588157]
5. Givens JL, Tjia J. Depressed medical students' use of mental health services and barriers to use. Acad Med 2002 Sep;77(9):918-921. [doi: 10.1097/00001888-200209000-00024] [Medline: 12228091]
6. Serafini G, Parmigiani B, Amerio A, Aguglia A, Sher L, Amore M. The psychological impact of COVID-19 on the mental health in the general population. QJM 2020 Jun 22;113(8):531-537. [doi: 10.1093/qjmed/hcaa201] [Medline: 32569360]
7. Siddiqui I, Aurelio M, Gupta A, Blythe J, Khanji MY. COVID-19: causes of anxiety and wellbeing support needs of healthcare professionals in the UK: a cross-sectional survey. Clin Med (Lond) 2021 Jan;21(1):66-72. [doi: 10.7861/clinmed.2020-0502] [Medline: 33479070]
8. Johnson J, Hall LH, Berzins K, Baker J, Melling K, Thompson C. Mental healthcare staff well-being and burnout: a narrative review of trends, causes, implications, and recommendations for future interventions. Int J Ment Health Nurs 2018 Feb;27(1):20-32. [doi: 10.1111/inm.12416] [Medline: 29243348]
9. O'Connor DB, Hall LH, Johnson J. Job strain, burnout, wellbeing and patient safety in healthcare professionals. In: Montgomery A, van der Doef M, Panagopoulou E, Leiter MP, editors. Connecting Healthcare Worker Well-Being, Patient Safety and Organisational Change: The Triple Challenge: Springer International Publishing; 2020:11-23. [doi: 10.1007/978-3-030-60998-6_2]
10. Shanafelt TD, Balch CM, Bechamps GJ, et al. Burnout and career satisfaction among American surgeons. Ann Surg 2009 Sep;250(3):463-471. [doi: 10.1097/SLA.0b013e3181ac4dfd] [Medline: 19730177]
11. Elmore LC, Jeffe DB, Jin L, Awad MM, Turnbull IR. National Survey of Burnout among US general surgery residents. J Am Coll Surg 2016 Sep;223(3):440-451. [doi: 10.1016/j.jamcollsurg.2016.05.014] [Medline: 27238875]
12. Prudenzi A, D Graham C, Flaxman PE, O'Connor DB. Wellbeing, burnout, and safe practice among healthcare professionals: predictive influences of mindfulness, values, and self-compassion. Psychol Health Med 2022 Jun;27(5):1130-1143. [doi: 10.1080/13548506.2021.1898651] [Medline: 33856236]
13. Toppinen-Tanner S, Ojajärvi A, Väänänen A, Kalimo R, Jäppinen P. Burnout as a predictor of medically certified sick-leave absences and their diagnosed causes. Behav Med 2005;31(1):18-27. [doi: 10.3200/BMED.31.1.18-32] [Medline: 16078523]
14. Hall LH, Johnson J, Watt I, Tsipa A, O'Connor DB. Healthcare staff wellbeing, burnout, and patient safety: a systematic review. PLoS One 2016;11(7):e0159015. [doi: 10.1371/journal.pone.0159015] [Medline: 27391946]
15. Lee DJ, Fleming LE, LeBlanc WG, et al. Health status and risk indicator trends of the aging US health care workforce. J Occup Environ Med 2012 Apr;54(4):497-503. [doi: 10.1097/JOM.0b013e318247a379] [Medline: 22446575]
16. Edwards D, Burnard P. A systematic review of stress and stress management interventions for mental health nurses. J Adv Nurs 2003 Apr;42(2):169-200. [doi: 10.1046/j.1365-2648.2003.02600.x] [Medline: 12670386]
17. Raiger J. Applying a cultural lens to the concept of burnout. J Transcult Nurs 2005 Jan;16(1):71-76. [doi: 10.1177/1043659604270980] [Medline: 15608102]
18. Petterson IL, Arnetz BB. Psychosocial stressors and well-being in health care workers. The impact of an intervention program. Soc Sci Med 1998 Dec;47(11):1763-1772. [doi: 10.1016/s0277-9536(98)00245-7] [Medline: 9877346]
19. Ruotsalainen J, Serra C, Marine A, Verbeek J. Systematic review of interventions for reducing occupational stress in health care workers. Scand J Work Environ Health 2008 Jun;34(3):169-178. [doi: 10.5271/sjweh.1240] [Medline: 18728906]
20. Brooks SK, Gerada C, Chalder T. Review of literature on the mental health of doctors: are specialist services needed? J Ment Health 2011 Apr;20(2):146-156. [doi: 10.3109/09638237.2010.541300] [Medline: 21275504]
21. Danet Danet A. Psychological impact of COVID-19 pandemic in Western frontline healthcare professionals. A systematic review. Med Clin (Barc) 2021 May 7;156(9):449-458. [doi: 10.1016/j.medcli.2020.11.009] [Medline: 33478809]
22. Dudman J, Isaac A, Johnson S. Revealed: how the stress of working in public services is taking its toll on staff. The Guardian. 2015. URL: https://www.theguardian.com/society/2015/jun/10/stress-working-public-services-survey [accessed 2024-12-27]
23. Mohanty A, Kabi A, Mohanty AP. Health problems in healthcare workers: a review. J Family Med Prim Care 2019 Aug;8(8):2568-2572. [doi: 10.4103/jfmpc.jfmpc_431_19] [Medline: 31548933]
24. Tyssen R, Vaglum P, Gronvold NT, Ekeberg O. The impact of job stress and working conditions on mental health problems among junior house officers. A nationwide Norwegian prospective cohort study. Med Educ (Chicago Ill) 2000 May;34(5):374-384. [doi: 10.1046/j.1365-2923.2000.00540.x]
25. Greenberg N. Mental health of health-care workers in the COVID-19 era. Nat Rev Nephrol 2020 Aug;16(8):425-426. [doi: 10.1038/s41581-020-0314-5] [Medline: 32561871]
26. NHS sickness absence rates. NHS England Digital. 2022. URL: https://digital.nhs.uk/data-and-information/publications/statistical/nhs-sickness-absence-rates?__cf_chl_tk=GyQ_jyQo3bml1ydIfymQem8x2Tw06jHtOkp6RsAg9cQ-1734726614-1.0.1.1-Znz.bX9zDciYi0lG5RKljML9k7KYoFv1pCFRbJER63o [accessed 2024-12-21]

XSL•FO
RenderX

27. NHS health and wellbeing framework. NHS England. 2021. URL: https://www.england.nhs.uk/publication/nhs-health-and-wellbeing-framework/ [accessed 2024-12-21]
28. Adler A, Seligman MEP. Using wellbeing for public policy: theory, measurement, and recommendations. Intnl J Wellbeing 2016;6(1):1-35. [doi: 10.5502/ijw.v6i1.429]
29. Luo M, Guo L, Yu M, Jiang W, Wang H. The psychological and mental impact of coronavirus disease 2019 (COVID-19) on medical staff and general public - a systematic review and meta-analysis. Psychiatry Res 2020 Sep;291:113190. [doi: 10.1016/j.psychres.2020.113190] [Medline: 32563745]
30. Huppert FA. The state of wellbeing science: concepts, measures, interventions, and policies. In: Huppert FA, Cooper CL, editors. Interventions and Policies to Enhance Wellbeing: A Complete Reference Guide: John Wiley & Sons Inc; 2014:1-50. [doi: 10.1002/9781118539415]
31. Adamou M, Goddard A, Kyriakidou N, et al. The wellbeing thermometer: a novel framework for measuring wellbeing. PSYCH 2020;11(10):1471-1480. [doi: 10.4236/psych.2020.1110093]
32. Forgeard MJC, Jayawickreme E, Kern ML, Seligman MEP. Doing the right thing: measuring well-being for public policy. Intnl J Wellbeing 2011;1(1). [doi: 10.5502/ijw.v1i1.15]
33. Promoting Mental Health: Concepts, Emerging Evidence, Practice: World Health Organization; 2004.
34. Hewitt R. Measuring well-being in higher education. HEPI. 2019. URL: https://www.hepi.ac.uk/wp-content/uploads/2019/05/Policy-Note-13-Paper-May-2019-Measuring-well-being-in-higher-education-8-Pages-5.pdf [accessed 2024-12-27]
35. Easterlin RA. Explaining happiness. Proc Natl Acad Sci 2003;100(19):11176-11183. [doi: 10.1073/pnas.1633144100]
36. Jacobson B. What is happiness? J Soc Exist Anal 2007;18(1):39-50.
37. Linton MJ, Dieppe P, Medina-Lara A. Review of 99 self-report measures for assessing well-being in adults: exploring dimensions of well-being and developments over time. BMJ Open 2016 Jul 7;6(7):e010641. [doi: 10.1136/bmjopen-2015-010641] [Medline: 27388349]
38. DeVellis RF. Scale Development: Theory and Applications, 2nd edition: Sage Publications, Inc; 2003.
39. Kline T. Psychological Testing: A Practical Approach to Design and Evaluation: Sage Publications, Inc; 2005.
40. Bowling A. Research Methods in Health: Investigating Health and Health Services, 2nd edition: Open University Press, McGraw-Hill Education; 2002.
41. Schmitt N. Uses and abuses of coefficient alpha. Psychol Assess 1996;8(4):350-353. [doi: 10.1037/1040-3590.8.4.350]
42. Lombardo M, Aulisa G, Padua E, et al. Gender differences in taste and foods habits. NFS 2019 Jul 18;50(1):229-239. [doi: 10.1108/NFS-04-2019-0132]
43. Oncini F, Guetto R. Cultural capital and gender differences in health behaviours: a study on eating, smoking and drinking patterns. Health Sociol Rev 2018 Jan 2;27(1):15-30. [doi: 10.1080/14461242.2017.1321493]
44. Roothman B, Kirsten DK, Wissing MP. Gender differences in aspects of psychological well-being. S Afr J Psychol 2003 Nov;33(4):212-218. [doi: 10.1177/008124630303300403]
45. Dyer NL, Adan F, Barnett T, Dusek JA. Assessment of healthcare professionals' wellbeing during a peak of the COVID-19 pandemic in a healthcare system in Ohio. Glob Adv Health Med 2022;11:2164957X221089258. [doi: 10.1177/2164957X221089258] [Medline: 35634474]
46. Wald HS. Optimizing resilience and wellbeing for healthcare professions trainees and healthcare professionals during public health crises - practical tips for an "integrative resilience" approach. Med Teach 2020 Jul;42(7):744-755. [doi: 10.1080/0142159X.2020.1768230] [Medline: 32449867]
47. Dyrbye LN, Shanafelt TD. Physician burnout: a potential threat to successful health care reform. JAMA 2011 May 18;305(19):2009-2010. [doi: 10.1001/jama.2011.652] [Medline: 21586718]
48. Mata DA, Ramos MA, Bansal N, et al. Prevalence of depression and depressive symptoms among resident physicians: a systematic review and meta-analysis. JAMA 2015 Dec 8;314(22):2373-2383. [doi: 10.1001/jama.2015.15845] [Medline: 26647259]
49. Schernhammer E. Taking their own lives -- the high rate of physician suicide. N Engl J Med 2005 Jun 16;352(24):2473-2476. [doi: 10.1056/NEJMp058014] [Medline: 15958803]
50. Velez K. Relationship between moral distress, workplace bullying, staff burnout and turnover at an academic spinal cord injury/disorder (SCI/D) rehabilitation facility: a causality study. Arch Phys Med Rehabil 2020 Nov;101(11):e54. [doi: 10.1016/j.apmr.2020.09.161]
51. Wilson W, Raj JP, Narayan G, Ghiya M, Murty S, Joseph B. Quantifying burnout among emergency medicine professionals. J Emerg Trauma Shock 2017;10(4):199-204. [doi: 10.4103/JETS.JETS_36_17] [Medline: 29097859]
52. Ornell F, Halpern SC, Kessler FHP, Narvaez JCDM. The impact of the COVID-19 pandemic on the mental health of healthcare professionals. Cad Saude Publica 2020;36(4):e00063520. [doi: 10.1590/0102-311X00063520] [Medline: 32374807]

Abbreviations

HCP: health care professional
NHS: National Health Service

XSL•FO
RenderX

WHO: World Health Organization

Edited by A Mavragani; submitted 31.10.23; peer-reviewed by K Kamposioras, T Gladman; revised version received 04.11.24; accepted 06.11.24; published 09.01.25.

Please cite as:
Adamou M, Jones SL, Kyriakidou N, Mooney A, Pattani S, Roycroft M
Measuring Self-Reported Well-Being of Physicians Using the Well-Being Thermometer: Cohort Study
JMIR Form Res 2025;9:e54158
URL: https://formative.jmir.org/2025/1/e54158
doi:10.2196/54158

© Marios Adamou, Sarah L Jones, Niki Kyriakidou, Andrew Mooney, Shriti Pattani, Matthew Roycroft. Originally published in JMIR Formative Research (https://formative.jmir.org), 9.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Recruitment Challenges and Strategies in a Technology-Based Intervention for Dementia Caregivers: Descriptive Study

Eunjung Ko[1], RN, PhD; Ye Gao[2], PhD; Peng Wang[3], MS; Lahiru Wijayasingha[3], MS; Kathy D Wright[4], RN, PhD; Kristina C Gordon[5], PhD; Hongning Wang[3], PhD; John A Stankovic[3], PhD; Karen M Rose[4], RN, PhD

[1]Rory Meyers College of Nursing, New York University, New York, NY, United States

[2]Computer Science, William & Mary, Williamsburg, VA, United States

[3]School of Engineering and Applied Science, University of Virginia, Charlottesville, VA, United States

[4]College of Nursing, The Ohio State University, Columbus, OH, United States

[5]College of Education, Health, and Human Sciences, The University of Tennessee-Knoxville, Knoxville, TN, United States

Corresponding Author:
Eunjung Ko, RN, PhD
Rory Meyers College of Nursing, New York University, New York, NY, United States

Abstract

Background: Researchers have encountered challenges in recruiting unpaid caregivers of people living with Alzheimer disease and related dementias for intervention studies. However, little is known about the reasons for nonparticipation in in-home smart health interventions in community-based settings.

Objective: This study aimed to (1) assess recruitment rates in a smart health technology intervention for caregivers of people living with Alzheimer disease and related dementias and reasons for nonparticipation among them and (2) discuss lessons learned from recruitment challenges and strategies to improve recruitment.

Methods: The smart health intervention was a 4-month, single-arm trial designed to evaluate an in-home, technology-based intervention that monitors stressful moments for caregiving dyads through acoustic signals and to provide the caregivers with real-time stress management strategies. The recruitment involved two main methods: on-site engagement by a recruiter from a memory clinic and social media advertising. Caregivers were screened for eligibility by phone between January 2021 and September 2023. The recruitment rates, reasons for nonparticipation, and participant demographics were analyzed using descriptive statistics.

Results: Of 201 caregivers contacted, 11 were enrolled in this study. Eighty-two caregivers did not return the screening call, and others did not participate due to privacy concerns (n=30), lack of interest (n=29), and burdensome study procedures (n=26). Our recruitment strategies included addressing privacy concerns, visualizing collected data through a dashboard, boosting social media presence, increasing the recruitment budget, updating advertisements, and preparing and deploying additional study devices.

Conclusions: This study highlighted barriers to participation in the smart health intervention. Despite several recruitment strategies, enrollment rates remained below expectations. These findings underscore the need for future research to explore alternative methods for increasing the recruitment of informal dementia caregivers in technology-based intervention studies.

Trial Registration: ClinicalTrials.gov NCT04536701; https://clinicaltrials.gov/study/NCT04536701

International Registered Report Identifier (IRRID): RR2-10.1111/jan.14714

(JMIR Form Res 2025;9:e59291) doi:10.2196/59291

KEYWORDS

recruitment challenges and strategies; technology-based intervention; dementia caregivers; dementia; mobile phone; Alzheimer disease; smart health

Introduction

The number of people living with Alzheimer disease and related dementias (ADRD) in the United States is increasing, with the figure expected to reach 13.8 million by 2060 [1]. Alzheimer disease, the primary neurodegenerative condition contributing to ADRD, necessitates ongoing and escalating levels of support over time for those affected [1]. Most people living with ADRD receive unpaid assistance from family members, friends, or neighbors. In 2023, the number of unpaid caregivers surpassed 11 million [1]. Managing the wide range of caregiving responsibilities and coordinating dementia care can exacerbate caregivers' stress, burden, anxiety, and depression [1,2].

Various nonpharmacological interventions have emerged to mitigate the negative aspects of caregiving, aimed at supporting caregivers [3]. These interventions include acceptance and

XSL•FO
RenderX

commitment therapy, education, counseling, psychotherapy, and support groups. Although the effects of these interventions vary depending on their contexts, the interventions have been generally effective in alleviating caregivers' negative emotions and improving their quality of life [1,3].

Over the past few decades, technological advancements and the recent pandemic challenges prompted a transition from in-person to online delivery of interventions. This shift has made interventions more accessible by accommodating caregivers' limited time and geographical locations and alleviating care demand [1,4,5]. The delivery methods of interventions range from the basic tools, such as telephone calls, video conferences, and web page visits, to more advanced technologies such as wearable devices, video or image sensors, or intelligent audio systems [4,6]. These technologies can enhance home safety and well-being and reduce health care costs for people living with dementia and their caregivers [6].

Despite the potential of these novel interventions, their deployment still faces challenges among caregivers of people living with ADRD. Barriers include limited experience with technology, lack of awareness, technical issues, time constraints, privacy concerns, and mistrust [6-8]. Notably, Wen et al [7] found that online-based interventions are more likely to be accepted by younger family caregivers, contrasting with the predominantly middle-aged and older caregiver population [1,9]. Further, the American Association of Retired Persons report revealed that only half of the caregivers across all age groups and just 20% of those older than 65 years of age felt comfortable using in-home technologies (eg, Google Assistant or Alexa; Amazon.com, Inc.) [10]. Given the growing proportion of caregivers aged 60 years and older (28% in 2015 and 35.4% in 2022) [11], participation in technology-based interventions, specifically through those requiring advanced technological devices, remain limited.

Recent studies have explored recruitment challenges and strategies in interventions targeting caregivers of people living with ADRD [7,12-15]. However, there is still limited research specifically focused on these issues within the context of technology-based interventions. Understanding and addressing these challenges are pivotal for effectively engaging caregivers of people living with ADRD in such interventions, which is the primary focus of our study.

In this descriptive study, we aimed to explore the recruitment challenges encountered in an in-home, technology-based intervention designed to reduce stress among caregivers of people living with ADRD and to improve their dyadic relationships. We identified the overall recruitment rate and the reasons for nonparticipation from the caregivers' perspective and examined the demographic characteristics of the caregivers enrolled in this study. In addition, we discussed the recruitment challenges and strategies we used to address them during the intervention.

Methods

Study Design of the Smart Health Intervention

The smart health intervention was a 4-month, single-arm trial designed to develop an in-home, technology-based intervention and assess its feasibility for caregivers of people living with ADRD [16,17]. Caregivers received this study's devices (ie, a laptop, a microphone connected to the computer, a router, and a smartphone). They placed them in an area of their home where they frequently interacted with their care recipients. We captured acoustic signals from the caregiving dyads through the microphone and used a deep machine-learning approach to identify stressful situations [16]. The system then sent a real-time stress management tip to the caregivers' study smartphone via the ecological momentary assessment (EMA) system [16,17]. Examples of tips sent to caregivers included practicing body scan and deep-breathing exercises, taking a time out, or engaging in enjoyable activities. The messages encouraged caregivers to follow the tip and to report its usefulness. Caregivers were also asked to answer daily and weekly questions about their physical and emotional health status and the overall usefulness of the stress management tips [17]. We provided US $50 compensation to each caregiver and care recipient three times—1 month after starting the intervention, halfway through, and at the end of the intervention.

Recruitment

Recruitment took place between January 2021 and September 2023 using convenience sampling. Two main recruitment methods were used to recruit eligible participants: (1) on-site engagement at a memory clinic within a Midwestern academic medical center and (2) advertisements on social media platforms. At the memory clinic, a recruiter distributed study brochures to individuals who visited the clinic and might be eligible for this intervention. Once the caregivers expressed interest in this study, the recruiter shared their contact information with this study's personnel (EK and KMR). This study's personnel contacted the caregivers to explain this study's procedures in detail and confirm their eligibility.

We also advertised the smart health intervention via social media platforms, such as Facebook (Meta) and Instagram (Instagram from Meta), to recruit caregivers across the United States. We allocated between US $100 and US $300 to recruit participants through social media platforms for 1 week of recruitment postings. The uploaded posts included this study's title, primary text, target age and gender, study locations, and a link to a website containing this study's detailed information. Individuals who saw this study's advertisements and were interested in this intervention contacted the principal investigator (KMR) by phone or email. In addition to these two methods, individuals also contacted the principal investigator after finding the information about our study on websites listing clinical trials, such as ClinicalTrials.gov [18].

Participant Eligibility

We recruited participants who met the following inclusion criteria: (1) aged 21 years or older, (2) residing with and caring for older adults living with ADRD, (3) not receiving payment

XSL•FO
RenderX

for providing care, (4) having well-functioning home Wi-Fi, (5) receiving a score above 3 on the Revised Memory and Behavior Problems Checklist, (6) being fluent in English, and (7) agreeing to participate [17]. During the screening call, those who met the eligibility criteria provided verbal consent and signed the electronic consent form. Given that this intervention involved monitoring acoustic signals while caregivers communicated with their care recipients, care recipients were also required to sign the consent form. If care recipients could not consent, their caregivers provided proxy consent.

Data Collection

Overview

We manually recorded the participation status and the reasons for nonparticipation of each individual contacted in REDCap (Research Electronic Data Capture; Vanderbilt University), a Health Insurance Portability and Accountability Act compliant, web-based data collection platform that allows researchers to access and manage data [19]. We also gathered the reasons for nonparticipation during the screening call. After individuals agreed to participate, we emailed them a REDCap link to the consent form and the demographic questionnaire. Those who consented and completed the questionnaires received copies of the informed consent form. Participants were required to complete the survey only through this link, with responses automatically stored in REDCap.

Participation Status and Reasons for Nonparticipation

During the screening call, we asked individuals screening questions and documented their reasons for nonparticipation. The reasons were categorized into predeveloped options: (1) did not meet eligibility criteria, (2) not interested, (3) considered study protocol to be too burdensome, (4) worried about privacy issues, (5) others (eg, considering care replacement to facilities), (6) expressed interest but did not respond to the screening call, (7) noted invasion of privacy was undesirable, (8) harmful events occurred due to this study, and (9) no longer wanted to participate.

Demographic Characteristics of Participants

We collected demographic data from participants at baseline. The questionnaire included information on the age, gender, race (American Indian or Alaska Native, Asian Native Hawaiian or other Pacific Islander, Black or African American, White, unknown, or not reported), ethnicity (Hispanic or Latino, Not Hispanic or Latino, or not reported), relationship between caregivers and care recipients (spouse, partner, sibling, child, grandchild, or other), education level (never attended or kindergarten only, K grades from 1st to 12th [not diploma], high school graduate, general education diploma or equivalent, attended college but did not obtain a degree, associate degree, bachelor's degree, master's degree, professional school, doctoral degree, or unknown), employment status (working now, only temporarily laid off, sick leave, maternity leave, unemployed, retired, disabled permanently or temporarily, keeping house, student, other, or unknown), caregiving hours per day, caregiving duration in months, marital status, the number of household members, zip code, veteran status (yes or no), and any history of combat exposure for those with a military background.

Challenges and Strategies in Recruitment

To address recruitment challenges through on-site engagement, we conducted two 20-minute Zoom (Zoom Communications, Qumu Corporation) meetings with a memory clinic recruiter in June and November 2022. We also held weekly meetings with the research team to discuss challenges in recruiting through social media and brainstorm potential solutions.

Ethical Considerations

The smart health intervention was approved by the institutional review board at this study's institution (2019B0406) and was registered with ClinicalTrials.gov (NCT04536701) on September 3, 2020, and completed on December 31, 2023. All participants provided written, informed consent before the data collection. Participants were informed about this study's purpose, duration, location of data storage, potential benefits and risks of involvement in the intervention, and contact information for the principal investigators.

Data Analysis

We used SPSS (version 29; IBM Corp) for descriptive analyses, including frequencies, percentages, means, and SDs, to present the characteristics of recruitment and participant demographics. We also used a note-taking approach during discussions with a recruiter and team members to capture key points on the topic.

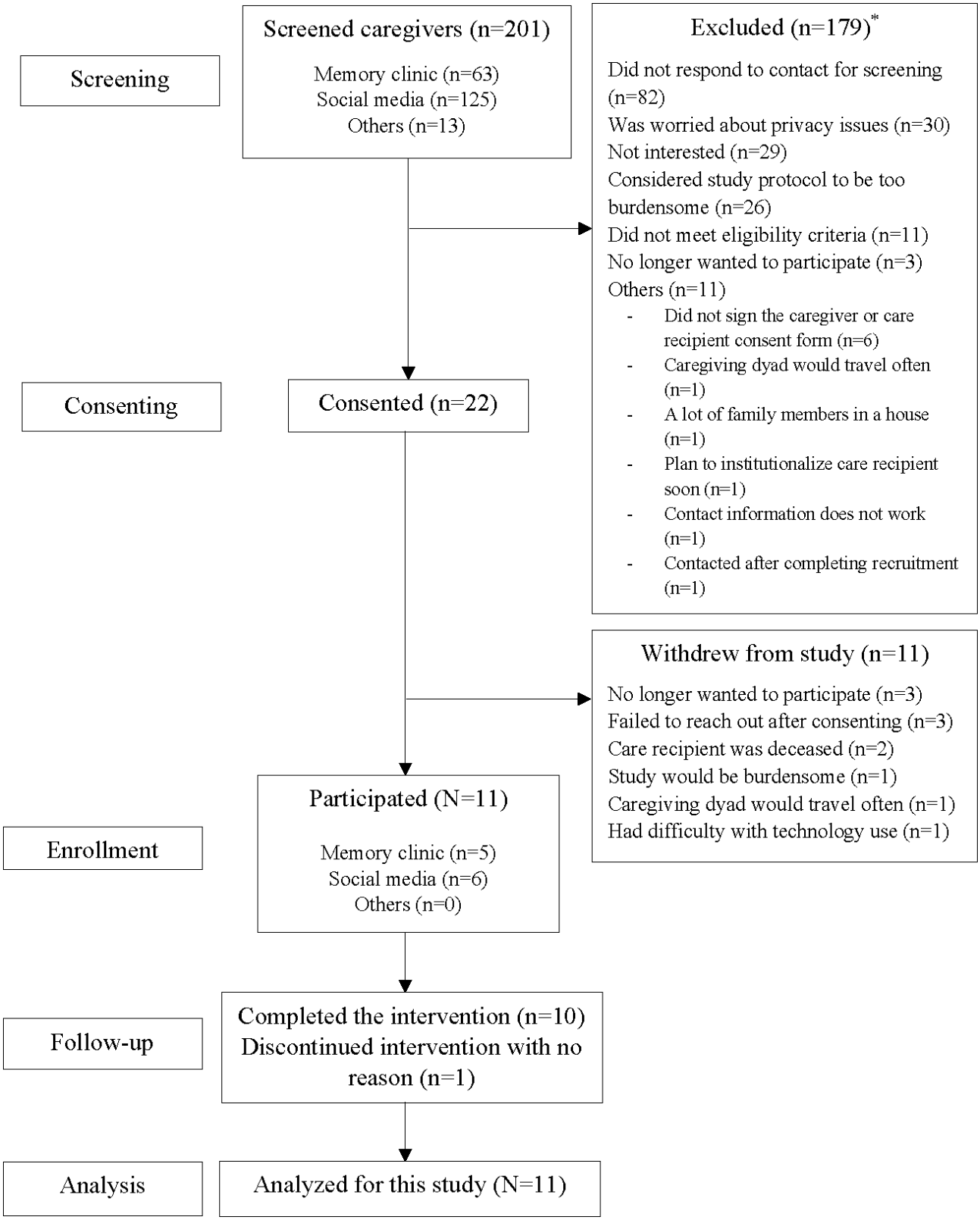
Results

Participation Status and Reasons for Nonparticipation

Figure 1 presents the recruitment flowchart and reasons for nonparticipation. We attempted to screen 201 caregivers through social media (125 individuals), memory clinic (63 individuals), and other sources, such as ClinicalTrials.gov or unknown routes (13 individuals). Of these, 22 caregivers and their care recipients agreed to consent; only half completed the baseline questionnaire and began the intervention. Six of the 11 participants were recruited through social media, and the remaining 5 were recruited through on-site engagement at the memory care clinic (Figure 1).

XSL•FO
RenderX

Figure 1. Flowchart illustrating participant recruitment and exclusion. *Some of 179 respondents addressed multiple reasons for their non-participation.



Among caregivers excluded, 82 did not respond to the screening contact, 30 declined to participate due to privacy concerns, 29 were not interested in the intervention, and 26 considered this study's protocol to be burdensome. Eleven respondents failed to meet the eligibility criteria, mainly because they were not primary or live-in caregivers, their care recipient was hospitalized or deceased, or they lacked functional Wi-Fi at home. Three caregivers expressed interest but later decided not to participate after further consideration. Other reasons for nonparticipation included (1) not providing written consent on a consent form after providing verbal agreement, (2) frequent travel plans, (3) having multiple household members besides the caregiving dyad, (4) planning to institutionalize their care recipient soon, and (5) providing incorrect contact information.

XSL•FO
RenderX

Further, 11 caregivers who consented ultimately withdrew from this study for reasons such as no longer wanting to participate, failure to reach out, and the death of the care recipient.

Demographic Characteristics of Participants

Table 1 presents the demographic data for the participating caregiving dyads. The average ages of caregivers and care recipients were 60.09 (SD 13.52) years and 76.73 (SD 8.04) years, respectively. Most caregivers were female (n=8, 73%), while most care recipients were male (n=7, 64%). Most caregiving dyads were non-Hispanic White (n=7, 82% of caregivers; n=9, 91% of care recipients). Education levels varied, but all caregivers had attained education beyond a high school degree. Regarding employment status, 8 (73%) caregivers and all care recipients were either unemployed or retired. Eight (73%) care recipients and 7 (64%) caregivers were married, with caregivers typically being spouses. The average amount of caregiving hours per day was 21.18 (SD 6.37) hours, and the average caregiving duration was 34.82 (SD 20.26) months.

XSL•FO
RenderX

Table . Demographics of actual participating dyads (N=11).

Characteristic	Caregivers	Care recipients
Age (years), mean (SD)	60.09 (13.52)	76.73 (8.04)
Gender, n (%)		
Female	8 (73)	4 (36)
Male	3 (27)	7 (64)
Race, n (%)		
African American or Black	1 (9)	1 (9)
Asian American	1 (9)	0 (0)
Non-Hispanic White	9 (82)	10 (91)
Ethnicity, n (%)		
Non-Hispanic	10 (91)	10 (91)
Not reported	1 (9)	1 (9)
Education level, n (%)		
General education diploma or equivalent	0 (0)	1 (9)
12th grade, not diploma	0 (0)	1 (9)
High school graduate	0 (0)	1 (9)
Attended college but did not obtain a degree	1 (9)	2 (18)
Associate degree	2 (18)	2 (18)
Bachelor’s degree	4 (36)	1 (9)
Master’s degree	3 (27)	1 (9)
Doctoral degree	1 (9)	2 (18)
Employment status, n (%)		
Employed	3 (27)	0 (0)
Not employed or retired	8 (73)	11 (100)
Marital status, n (%)		
Married	7 (64)	8 (73)
Never married	3 (27)	0 (0)
Widowed	0 (0)	2 (18)
Divorced	1 (9)	1 (9)
Veteran status and combat exposure, n (%)		
Yes, no combat exposure	2 (18)	1 (9)
No	9 (82)	10 (91)
Caregiving hours (h/d), mean (SD)	21.18 (6.37)	—[a]
Caregiving duration (mo), mean (SD)	34.82 (20.26)	—
Household member, mean (SD)	2.27 (0.65)	—

[a]The “—” indicates “not applicable”

Lessons Learned Through Recruitment Challenges

Challenges of On-Site Engagement

A recruiter responsible for informing caregivers about this study at the memory clinic provided some reasons for a low rate of face-to-face recruitment. First, staffing shortages at the clinic and the shift from in-person to remote medical appointments likely reduced recruitment opportunities. When patients transitioned to telehealth appointments, the recruiter could not meet with them and their caregivers to provide study information. Additionally, increased patient no-shows for in-person medical appointments further hindered recruitment efforts. Caregivers also expressed concerns about the time commitment and privacy issues associated with technology-based interventions, which echoed the feedback from the caregivers who refused participation in the smart health

XSL•FO
RenderX

intervention. The recruiter noted that caregivers showed more interest in pharmacological trials for their care recipients than in nonpharmacological trials for themselves.

To address these challenges, the recruiter suggested introducing this study virtually before or after telehealth appointments to raise awareness among caregivers and their care recipients. Increasing the clinic's staff who help study recruitment and having study personnel provide detailed explanations of this study—either in person or virtually—while caregivers wait for their care recipients could also help improve recruitment rates.

Challenges of Social Media Recruitment

We also encountered specific challenges with recruiting through social media platforms. A primary issue was the need to present study information briefly within social media posts, which often limited the detail we could provide. This brevity hindered potential participants from receiving detailed study information to decide, which is particularly critical in intervention studies. For example, of 601 caregivers who clicked the advertisements in April, only 7 reached out to the staff for further information. This suggests that the brief content in the advertisements may not have efficiently delivered this study's purpose or importance to potential participants.

Another issue was that the visibility of our study was directly tied to our advertising budget. As a result, we had to allocate more resources to increase our reach on social media platforms. Moreover, we faced difficulties maintaining the balance between demand and supply. When study advertisements were posted on social media, we received a flood of inquiries from interested participants, but limitations in the availability of study equipment caused delays. Some participants had to wait weeks or months to receive this study's equipment, likely contributing to decreased interest and higher withdrawal rates.

Strategies to Improve Recruitment

We discussed and attempted several strategies to alleviate the recruitment challenges, specifically targeting the individual concerns that led to the reluctance to participate. To mitigate privacy concerns in the smart health intervention, we reassured caregivers during recruitment that we would not record their conversation or behaviors but focus on the changes in acoustic characteristics, such as tone, intonation, and amplitude. We allowed participants to choose the microphone placement for data collection, provided it was in a room where frequent interactions with the care recipient occurred. Additionally, we clarified that we would only analyze acoustic data collected from the participating caregiving dyad, differentiating their voices from others in the household (ie, visitors).

We also developed a dashboard (Figure 2) and shared it with participants, showing the visualized data without capturing specific words or conversations. This approach helped ease privacy concerns, and as this study progressed, participants gained confidence in the system's utility. By the end of this study, they reported that privacy was no longer a concern.

To enhance recruitment through social media, we consulted with social media specialists on ways to improve the visibility of this study's advertisements. We increased our budget for social media advertising from US $100/wk in April to US $200/wk in September and revised the advertising text to provide more detailed information for eligible caregivers (Textbox 1). These modifications led to a slight rise in interest from potential study participants, as indicated by an increase in advertisement clicks from 601 in April to 1421 in September. The number of inquiries for detailed information also grew, from 7 in April to 17 in September. However, there was no change in terms of percentage (from 7/601, 1.16% in April to 16/1701, 1.20%), while the cost per click was higher (from US $0.33 in April to US $0.42 in September). Despite these efforts, we recruited only 5 out of 125 participants through social media, indicating that the strategies may not be effective and that further actions are needed to capture the attention of potential participants better.

XSL•FO
RenderX

Figure 2. Examples of dashboard data collection. These graphs were generated based on an actual participant's data. The x-axis represents the passage of time. (A) The cumulated number of daily mood states through acoustic monitoring. (B) Self-reported daily emotional status. (C) Followed recommendation tips per day (bars) and daily scores for helpfulness (line).

Textbox 1. Changes in the statements on advertising posts.

Statement on the advertisements *before* the revision

Researchers at the ### are seeking participants for a study to learn more about ways we can help family caregivers for persons with memory loss manage their stress better. The research study will help family caregivers recognize their stressful situations and find ways they may be able to feel less stress.

Statement on the advertisements *after* the revision

Researchers at the ### are seeking family or friends who care for people with memory loss for a study. *By identifying voice changes, this study* will help caregivers recognize their feelings in tense moments and *cope with their difficult conversations with persons with memory loss*. This may help caregivers find ways to i*ncrease self-care and satisfaction with their daily lives. You do not have to be good with technology to participate. Participant privacy will be carefully protected. Please do not hesitate to contact us with questions or for more information!*

We prepared and deployed additional study devices for participants on the waitlist. To enhance their motivation to participate in this study, we conducted regular follow-ups, approximately 2 to 3 times monthly. These actions enabled us to monitor and address study-related challenges and facilitate effective communication and rapport between study personnel and participants, helping ensure their continued involvement in this study. However, some potential participants experienced changes in their caregiving situations (eg, planning travel, experiencing a care recipient's death, or having second thoughts about engaging in this study) and withdrew from the waitlist.

Discussion

Principal Findings

This study aimed to identify recruitment challenges in a smart health technology intervention for caregivers of people living with dementia. We observed a low recruitment rate and identified various reasons for nonparticipation among caregivers. Beyond nonresponsiveness to screening, the most common reasons for nonparticipation included privacy concerns, a lack of interest, and perceptions of this study as being burdensome.

XSL•FO
RenderX

The demographic characteristics of participating caregivers showed that the majority were mainly female, White, spousal caregivers, and retired. Recruitment challenges were observed both on-site and through social media. On-site engagement challenges included the transition to remote medical appointments and staffing issues. At the same time, social media-based recruitment faced limitations related to the context of postings, budget considerations, and balancing demand and supply.

We observed a lower recruitment rate through social media (6/125 individuals) than on-site engagement (5/63 individuals). Previous literature has reported mixed results regarding the effectiveness of using social media for recruitment in intervention research [13,20-26]. While some studies have suggested that social media recruitment could be costly and ineffective [20-22,24], others have argued that social media could benefit the recruitment of potential target participants [13,23,25,26]. These conflicting findings highlight a need for further research to explore differences in recruitment rates across recruitment channels in technology-based interventions for people living with ADRD and their caregivers.

Caregivers of people living with ADRD may be anxious about sharing their data to use technology [27-29]. Additionally, they often prioritize dementia care and pay more attention to clinical trials focused on treating a specific disease than trials targeting their health [30]. While multiple pharmacological clinical trials may be ongoing in a clinic, caregivers of people living with ADRD may overlook nonpharmacological intervention studies. Moreover, they already spend considerable time and energy caring for their care recipients, and they may perceive participating in an intervention study, especially a technology-based trial, to be overly demanding [29,31,32]. Therefore, intervention research using technology should use strategies that reduce the perceived burden, such as providing training, using user-friendly platforms, or offering nondigital alternatives, including daily logs for those uncomfortable with technology in the caregiving population [8]. These tailored approaches have the potential to enhance the enrollment in dementia caregiving research.

Being female and non-Hispanic White are common characteristics of caregivers of people living with dementia in the United States [1], and spousal dementia caregivers are more likely to be retired [33]. Compared to male caregivers, females are more likely to sacrifice personal time, which can lead to a more significant burden and depressive symptoms [1]. Spousal dementia caregivers are also at a higher risk for depression, with rates being 2.5 times higher than that for nonspousal caregivers [1]. Caregivers with these characteristics may be more willing to participate in caregiving research on stress management. Still, the lack of demographic data on nonparticipants makes it difficult to confirm this. Therefore, future studies should gather information from participants and nonparticipants to better understand why some caregivers choose not to engage in research.

Limitations and Potential Implications

Despite our efforts, the strategies implemented in the smart health intervention needed to address recruitment barriers effectively. First, we needed more complete data on the socioeconomic status and race or ethnicity of potential participants who did not enroll in our study. Research shows that dementia caregiving dyads from lower socioeconomic backgrounds are less likely to participate in studies involving technology due to unfamiliarity and limited access to technological devices [31,34]. This gap in our data limited our ability to access the potential impact of the backgrounds of caregiving dyads on the recruitment rate.

Second, concerns about this study being too demanding persisted throughout recruitment, likely due to caregivers' unfamiliarity with the technology. This study's design, which included real-time messaging through the EMA system, may have contributed to this perception. However, an intervention study using the EMA system reported that the EMA approach did not bother participating caregivers [35]. Nonetheless, more research is needed to understand how the EMA system affects caregivers' daily lives while capturing data [36].

We could not manage the recruitment challenges, such as the transition to remote medical appointments and staffing issues, through on-site engagement. Recruitment might have improved with more advertising before or after virtual medical appointments, greater research personnel involvement, and additional resources; however, barriers to participating in telehealth visits made implementing these strategies unfeasible. The rise in health care since the COVID-19 pandemic—an approximately 150% increase in telehealth visits [37]—posed a barrier to recruiting caregivers in person. Future studies should explore ways to overcome these challenges when recruiting participants with limited in-person contact.

Despite these limitations, the findings of this study can inform strategies for addressing recruitment challenges in dementia caregiving research. Our findings offer insights into the reasons for nonparticipation and how to address participants' concerns regarding the smart health intervention. The findings may help researchers develop better strategies to improve recruitment rates in similar studies, offering caregivers opportunities to enhance their health and the quality of care they provide to their loved ones.

Acknowledgments

The authors thank Jenny Icenhour, BA, a research manager in the memory clinic, for providing opinions on the research topic. The data collection was funded by the National Science Foundation (grants 1838589 and 1838615). The National Center for Advancing Translational Sciences of the National Institutes of Health also supported this research (grant UL1TR002733). The material is the full responsibility of the authors and may not reflect the official views of the National Institutes of Health.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. 2024 Alzheimer's disease facts and figures. Alzheimers Dement 2024 May;20(5):3708-3821. [doi: 10.1002/alz.13809] [Medline: 38689398]
2. Pinyopornpanish K, Soontornpun A, Wongpakaran T, et al. Impact of behavioral and psychological symptoms of Alzheimer's disease on caregiver outcomes. Sci Rep 2022 Aug 19;12(1):14138. [doi: 10.1038/s41598-022-18470-8] [Medline: 35986203]
3. Sun Y, Ji M, Leng M, Li X, Zhang X, Wang Z. Comparative efficacy of 11 non-pharmacological interventions on depression, anxiety, quality of life, and caregiver burden for informal caregivers of people with dementia: a systematic review and network meta-analysis. Int J Nurs Stud 2022 May;129:104204. [doi: 10.1016/j.ijnurstu.2022.104204] [Medline: 35247788]
4. González-Fraile E, Ballesteros J, Rueda JR, Santos-Zorrozúa B, Solà I, McCleery J. Remotely delivered information, training and support for informal caregivers of people with dementia. Cochrane Database Syst Rev 2021;1(1):CD006440. [doi: 10.1002/14651858.CD006440.pub3]
5. Sharma RK, Teng A, Asirot MG, Taylor JO, Borson S, Turner AM. Challenges and opportunities in conducting research with older adults with dementia during COVID-19 and beyond. J Am Geriatr Soc 2022 May;70(5):1306-1313. [doi: 10.1111/jgs.17750] [Medline: 35285942]
6. Ali MT, Turetta C, Demrozi F, Pravadelli G. ICT-based solutions for Alzheimer's disease care: a systematic review. IEEE Access 2024;12:13944-13961. [doi: 10.1109/ACCESS.2024.3356348]
7. Wen Y, Xing Y, Ding Y, Xu W, Wang X. Challenges of conducting of online educational programs for family caregivers of people with dementia living at home: an integrative review. Int J Nurs Sci 2023 Jan;10(1):121-128. [doi: 10.1016/j.ijnss.2022.12.017] [Medline: 36860711]
8. Brookman R, Parker S, Hoon L, et al. Technology for dementia care: what would good technology look like and do, from carers' perspectives? BMC Geriatr 2023 Dec 16;23(1):867. [doi: 10.1186/s12877-023-04530-9] [Medline: 38104074]
9. Brodaty H, Donkin M. Family caregivers of people with dementia. Dialogues Clin Neurosci 2009;11(2):217-228. [doi: 10.31887/DCNS.2009.11.2/hbrodaty] [Medline: 19585957]
10. Keenan TA. U.S. caregivers' use of technology. : AARP Research; 2022. [doi: 10.26419/res.00566.001]
11. Kilmer G, Omura JD, Bouldin ED, et al. Changes in health indicators among caregivers - United States, 2015-2016 to 2021-2022. MMWR Morb Mortal Wkly Rep 2024 Aug 29;73(34):740-746. [doi: 10.15585/mmwr.mm7334a2] [Medline: 39207933]
12. Joshi S, Park T, Brody L, et al. Recruitment of family caregivers of persons with dementia: lessons learned from a pilot randomized controlled trial. Front Pain Res (Lausanne) 2023;4:1125914. [doi: 10.3389/fpain.2023.1125914] [Medline: 37051457]
13. Baker FA, Blauth L, Bloska J, et al. Recruitment approaches and profiles of consenting family caregivers and people living with dementia: a recruitment study within a trial. Contemp Clin Trials Commun 2023 Apr;32:101079. [doi: 10.1016/j.conctc.2023.101079] [Medline: 36949849]
14. Langbaum JB, Zissimopoulos J, Au R, et al. Recommendations to address key recruitment challenges of Alzheimer's disease clinical trials. Alzheimers Dement 2023 Feb;19(2):696-707. [doi: 10.1002/alz.12737] [Medline: 35946590]
15. Yang M, Samper-Ternent R, Volpi E, et al. The dementia care study (D-CARE): recruitment strategies and demographic characteristics of participants in a pragmatic randomized trial of dementia care. Alzheimers Dement 2024 Apr;20(4):2575-2588. [doi: 10.1002/alz.13698] [Medline: 38358084]
16. Gao Y, Jabbour J, Schlegel EC, et al. Out-of-the-box deployment to support research on in-home care of Alzheimer's patients. IEEE Pervasive Comput 2021;21(1):37-47. [doi: 10.1109/MPRV.2021.3106885]
17. Rose KM, Gordon KC, Schlegel EC, et al. Smarthealth technology study protocol to improve relationships between older adults with dementia and family caregivers. J Adv Nurs 2021 May;77(5):2519-2529. [doi: 10.1111/jan.14714] [Medline: 33576064]
18. Rose K. Collaborative research: learning and improving Alzheimer's patient-caregiver relationships via smart healthcare technology. Clinicaltrials.gov. 2024. URL: https://clinicaltrials.gov/study/NCT04536701 [accessed 2024-10-09]
19. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
20. Beattie E, Fielding E, O'Reilly M, Brooks D, MacAndrew M, McCrow J. Recruitment of individuals with dementia and their carers for social research: lessons learned from nine studies. Res Gerontol Nurs 2018 May 1;11(3):119-128. [doi: 10.3928/19404921-20180308-01] [Medline: 29634849]
21. Leach MJ, Ziaian T, Francis A, Agnew T. Recruiting dementia caregivers into clinical trials: lessons learnt from the Australian TRANSCENDENT trial. Alzheimer Dis Assoc Disord 2016;30(4):338-344. [doi: 10.1097/WAD.0000000000000149] [Medline: 27227995]

XSL•FO
RenderX

22. Leslie M, Khayatzadeh-Mahani A, MacKean G. Recruitment of caregivers into health services research: lessons from a user-centred design study. Res Involv Engagem 2019;5(1):17. [doi: 10.1186/s40900-019-0150-6] [Medline: 31139432]
23. Wasfi R, Stephens ZP, Sones M, et al. Recruiting participants for population health intervention research: effectiveness and costs of recruitment methods for a cohort study. J Med Internet Res 2021 Nov 12;23(11):e21142. [doi: 10.2196/21142] [Medline: 34587586]
24. Whitlatch CJ, Orsulic-Jeras S, Johnson J. Challenges to and strategies for recruiting chronic care dyads into intervention research. Chronic Illn 2021 Sep;17(3):232-241. [doi: 10.1177/1742395319869434] [Medline: 31426659]
25. Bremer W, Sarker A. Recruitment and retention in mobile application-based intervention studies: a critical synopsis of challenges and opportunities. Inform Health Soc Care 2023 Apr 3;48(2):139-152. [doi: 10.1080/17538157.2022.2082297] [Medline: 35656732]
26. Darko EM, Kleib M, Olson J. Social media use for research participant recruitment: integrative literature review. J Med Internet Res 2022 Aug 4;24(8):e38015. [doi: 10.2196/38015] [Medline: 35925655]
27. Block L, Gilmore-Bykovskyi A, Jolliff A, Mullen S, Werner NE. Exploring dementia family caregivers' everyday use and appraisal of technological supports. Geriatr Nurs 2020;41(6):909-915. [doi: 10.1016/j.gerinurse.2020.06.019] [Medline: 32665095]
28. Christie HL, Bartels SL, Boots LMM, Tange HJ, Verhey FRJ, de Vugt ME. A systematic review on the implementation of eHealth interventions for informal caregivers of people with dementia. Internet Interv 2018 Sep;13:51-59. [doi: 10.1016/j.invent.2018.07.002]
29. Hassan AYI. Challenges and recommendations for the deployment of information and communication technology solutions for informal caregivers: scoping review. JMIR Aging 2020 Jul 29;3(2):e20310. [doi: 10.2196/20310] [Medline: 32723720]
30. Clemmensen TH, Lauridsen HH, Andersen-Ranberg K, Kristensen HK. "I know his needs better than my own" - carers' support needs when caring for a person with dementia. Scand J Caring Sci 2021 Jun;35(2):586-599. [doi: 10.1111/scs.12875] [Medline: 32410395]
31. Brody AA, Convery KA, Kline DM, Fink RM, Fischer SM. Transitioning to remote recruitment and intervention: a tale of two palliative care research studies enrolling underserved populations during COVID-19. J Pain Symptom Manage 2022 Jan;63(1):151-159. [doi: 10.1016/j.jpainsymman.2021.06.017] [Medline: 34161811]
32. Trivedi RB, Szarka JG, Beaver K, et al. Recruitment and retention rates in behavioral trials involving patients and a support person: a systematic review. Contemp Clin Trials 2013 Sep;36(1):307-318. [doi: 10.1016/j.cct.2013.07.009] [Medline: 23916918]
33. Torres JM, Romero KRF, Kotwal AA, et al. Spouses of individuals living with mild cognitive impairment or dementia in the United States: a descriptive, population-based study. Alzheimers Dement 2024 Mar;20(3):1562-1572. [doi: 10.1002/alz.13555] [Medline: 38041823]
34. AboJabel H, Welsch J, Schicktanz S. Cross-cultural perspectives on intelligent assistive technology in dementia care: comparing Israeli and German experts' attitudes. BMC Med Ethics 2024 Feb 7;25(1):15. [doi: 10.1186/s12910-024-01010-6] [Medline: 38326778]
35. van Knippenberg RJM, de Vugt ME, Ponds RW, Myin-Germeys I, van Twillert B, Verhey FRJ. Dealing with daily challenges in dementia (deal-id study): an experience sampling study to assess caregiver functioning in the flow of daily life. Int J Geriatr Psychiatry 2017 Sep;32(9):949-958. [doi: 10.1002/gps.4552] [Medline: 27464472]
36. Han A, Malone LA, Lee HY, et al. The use of ecological momentary assessment for family caregivers of adults with chronic conditions: a systematic review. Health Psychol Res 2024;12:93907. [doi: 10.52965/001c.93907] [Medline: 38435338]
37. Koonin LM, Hoots B, Tsang CA, et al. Trends in the use of telehealth during the emergence of the COVID-19 pandemic - United States, January-March 2020. MMWR 2020 Oct 30;69(43):1595-1599. [doi: 10.15585/mmwr.mm6943a3] [Medline: 33119561]

Abbreviations

ADRD: Alzheimer disease and related dementias
EMA: ecological momentary assessment
REDCap: Research Electronic Data Capture

Edited by A Mavragani; submitted 24.05.24; peer-reviewed by B Davis, X Li; revised version received 22.11.24; accepted 05.12.24; published 17.01.25.

Please cite as:
Ko E, Gao Y, Wang P, Wijayasingha L, Wright KD, Gordon KC, Wang H, Stankovic JA, Rose KM
Recruitment Challenges and Strategies in a Technology-Based Intervention for Dementia Caregivers: Descriptive Study
JMIR Form Res 2025;9:e59291
URL: https://formative.jmir.org/2025/1/e59291
doi:10.2196/59291

XSL•FO
RenderX

© Eunjung Ko, Ye Gao, Peng Wang, Lahiru Wijayasingha, Kathy D Wright, Kristina C Gordon, Hongning Wang, John A Stankovic, Karen M Rose. Originally published in JMIR Formative Research (https://formative.jmir.org), 17.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Analysis of Social Performance and Action Units During Social Skills Training: Focus Group Study of Adults With Autism Spectrum Disorder and Schizophrenia

Hiroki Tanaka[1*], PhD; Kana Miyamoto[2*], PhD; Jennifer Hamet Bagnou[3], MS; Elise Prigent[3], PhD; Céline Clavel[3], PhD; Jean-Claude Martin[3], PhD; Satoshi Nakamura[2], PhD

[1]International Christian University, Mitaka, Tokyo, Japan

[2]Nara Institute of Science and Technology, Ikoma, Japan

[3]Université Paris-Saclay, Orsay, France

[*]these authors contributed equally

Corresponding Author:
Hiroki Tanaka, PhD
International Christian University, Mitaka, Tokyo, Japan

Abstract

Background: Social communication is a crucial factor influencing human social life. Quantifying the degree of difficulty faced in social communication is necessary for understanding developmental and neurological disorders and for creating systems used in automatic symptom screening and assistive methods such as social skills training (SST). SST by a human trainer is a well-established method. Previous SST used a modified roleplay test to evaluate human social communication skills. However, there are no widely accepted evaluation criteria or social behavioral markers to quantify social performance during SST.

Objective: This paper has 2 objectives. First, we propose applying the Social Performance Rating Scale (SPRS) to SST data to measure social communication skills. We constructed a Japanese version of the SPRS already developed in English and French. Second, we attempt to quantify action units during SST for people with autism spectrum disorder (ASD) or schizophrenia.

Methods: We used videos of interactions between trainers, adults with ASD (n=16) or schizophrenia (n=15), and control participants (n=19) during SST sessions. Two raters applied the proposed scale to annotate the collected data. We investigated the differences between roleplay tasks and participant groups (ASD, schizophrenia, and control). Furthermore, the intensity of action units on the OpenFace toolkit was measured in terms of mean and SD during SST roleplaying.

Results: We found significantly greater gaze scores in adults with ASD than in adults with schizophrenia. Differences were also found between the ratings of different tasks in the adults with schizophrenia and the control participants. Action units numbered AU06 and AU12 were significantly deactivated in people with schizophrenia compared with the control group. Moreover, AU02 was significantly activated in people with ASD compared with the other groups.

Conclusions: The results suggest that the SPRS can be a useful tool for assessing social communication skills in different cultures and different pathologies when used with the modified roleplay test. Furthermore, facial expressions could provide effective social and behavioral markers to characterize psychometric properties. Possible future directions include using the SPRS for assessing social behavior during interaction with a digital agent.

(JMIR Form Res 2025;9:e59261) doi:10.2196/59261

KEYWORDS

social performance rating scale; social skills training; autism spectrum disorder; schizophrenia; facial expressions; social; autism; training; communication; trainers; tool; neurological

Introduction

Social communication skills are essential in human life. Experienced psychiatrists identify people with social communication difficulties, such as autism spectrum disorder (ASD or autism spectrum condition) and schizophrenia, through interviews based on diagnostic criteria, responses, and various neuropsychological tests [1,2]. To improve the identification accuracy of neurodevelopmental and neurological disorders such as ASD and schizophrenia, it is crucial to discover symptom-specific behavioral markers and phenotyping. Both adults with ASD and and adults with schizophrenia are characterized by social difficulties [3]; in terms of deficits, there have been reports describing conversational gestures that are less closely synchronized with co-occurring speech [4], frequency of eye contact [5], speech prosody [6], and delayed

XSL•FO
RenderX

response time [7]. For schizophrenia, deficits in social cognition are a critical determinant of social functioning [8]. It has been reported that people with ASD and schizophrenia share several symptomatically similar characteristics such as flat or blunted affect (eg, reduced eye contact) and alogia (eg, impoverished speech) [9,10]. Thus, quantifying the degree of social communication difficulty is necessary to understand the nature of ASD and schizophrenia and to create systems for automatic symptom screening and early intervention methods such as social skills training (SST).

SST is a psychosocial treatment that has been widely applied and adapted to help people improve their social communication skills. It has been used in hospitals, employment support facilities, workplaces, and schools. A human trainer generally conducts SST to promote appropriate social communication skills, strengthen the individual's social self-efficacy, and reduce social anxiety. The Bellack method [11], or step-by-step SST, is a well-structured approach that includes defining target skills, modeling, roleplay, feedback, and homework. This method defines the SST framework and its 4 basic skills: asking for requests, declining requests, telling positive feelings, and listening to others. Some researchers have been conducting studies to build SST systems using socially interactive agents [12,13]. Personalizing is an essential element for such SST systems. For example, Bellack et al [11] recommended using shorter and more precise feedback for people with schizophrenia who experience hallucinations. Since they often hear voices distracting the SST, focusing is difficult. For ASD, feedback could be disengaging if trainers frequently mention exaggerated gestures or facial expressions, since these are major symptomatic characteristics. Toward more personalized SST systems, it is imperative to quantitatively measure the characteristics of ASD and schizophrenia.

However, there are no widely accepted evaluation criteria during SST. If a well-validated scale was available, it could be used as a ground truth value for machine learning models in the SST system [14]. Among social behavioral markers, individual roleplay performance can be rated with a modified roleplay test conducted [15,16], although there is no internationally validated scale. This type of test is further limited because it must be evaluated by experts such as psychiatrists and SST trainers. Moreover, there is no robust social behavioral marker to quantify social performance during SST. One of the important behavioral markers is facial expression. Previous works showed that action units AU06 and AU12 are less expressed in ASD participants [17,18], while AU01 is also less expressed in ASD participants [19]. For patients with schizophrenia, as far as the authors know, no previous study has provided an analysis of particular action units, although machine learning–based schizophrenia detection models have been proposed [20].

This paper proposes applying the Social Performance Rating Scale (SPRS) to measure social communication skills. The SPRS is a validated scale developed in English [21] and French [22]. The English version was applied to videos of a simulated dinner party for participants with a primary diagnosis of social phobia, those with a primary diagnosis of anxiety disorder, and control participants. The French version was applied to videos of collaborative games between neurotypical participants. This scale has been extensively used in a variety of research [23]. In this paper, we newly developed a Japanese version of the SPRS that can be used without special qualifications in SST. We collected data on 4 SST tasks performed by adults with ASD, with schizophrenia, and control groups. Two raters used the SPRS to annotate their performance in SST. The goals of our study are (1) to validate our Japanese version of the SPRS, (2) to apply the SPRS to SST and examine the correlations between the SPRS and the modified roleplay test, (3) to investigate whether the SPRS can be applied to participants with ASD or schizophrenia, and (4) to analyze the relations between the SPRS, facial expressions, and questionnaire results. Here, we aim to analyze the differences in action units during SST roleplay in people with ASD or schizophrenia. This is an extended version of a conference workshop's proceedings [24], in which we added an analysis of the facial expressions of people with ASD or schizophrenia.

In summary, we made contributions by developing and examining the following items: we developed a Japanese version of the SPRS; we analyzed social communication skills by focusing on the differences among adults with ASD or schizophrenia and a control group, on task differences of SST, and on questionnaire scores; and we analyzed action units by focusing on the differences among adults with ASD or schizophrenia and a control group, as well as questionnaire scores.

Methods

Ethical Considerations

This study was approved by the institutional review board of the Nara Medical University, Japan, and Nara Institute of Science and Technology, Japan (reference number: 2019-I-27 - 2). We explained the experiment's procedure and purpose to the participants and obtained informed consent. All data were anonymized. Participants received monetary rewards of up to 16,000 Japanese yen.

Development of Japanese Version of the SPRS

The items in the English version of the SPRS are provided in Table 1, where each of them is rated on a Likert scale of 1 to 5 [25].

XSL•FO
RenderX

Table . Social Performance Rating Scale (SPRS; modified from an earlier study [21]).

Item	Description
Gaze	• (1) Very poor: Participant completely avoids looking at the partner or stares continually. • (5) Very good: Participant keeps eye contact during the conversation, does not stare; shifts focus during pauses and conversation.
Voice quality	• (1) Very poor: (1) Participant speaks in a flat, monotonous voice; or (2) speaks at a low volume or mumbles; or (3) speaks overly loudly or has an intrusive tone (harsh or unpleasant voice quality). • (5) Very good: Participant is warm and enthusiastic in verbal expression without sounding condescending or gushy.
Length	• (1) Very poor: Monosyllabic ("hmmm," "yeah," "OK") speech turns; or responses so long that the partner must interrupt or cannot utter a reply. • (5) Very good: At most times, participant's utterances are 2 or more sentences long. Participant acknowledges partner's remarks without taking over and monopolizing the conversation.
Discomfort	• (1) Very high: Complete rigidity of arms, legs, or whole body. Constant leg movements or fidgeting with hands, hair, or clothing. Extremely stiff face or constant facial tics. Frequent nervous throat clearing, swallowing, or stuttering. Frequent inappropriate giggling or laughing. Look of extreme discomfort and desire to flee the situation shown by 2 or more breaks in the role. Participant does not pay attention to the roleplay tasks most of the time. • (5) Very low: Relaxed body posture and natural body movement. Participant laughs and smiles at appropriate times. She/he shows effective gesturing (to be distinguished from fidgeting). Participant focuses on the task all of the time, does not appear at all uncomfortable, but at ease in the situation.
Conversation flow	• (1) Very poor: Participant makes few attempts to initiate the conversation. Even when prompted by the partner, participant cannot maintain the conversation. Participant uses almost no open-ended questions or is intrusive in questions and shows no empathy. Participant does not pay attention to the information provided by partner. • (5) Very good: Participant easily maintains the conversation and responds smoothly to pauses in the conversation, often by following up on previous information provided by the partner or providing free information about themself on a related topic. Participant introduces new topics fluidly and frequently use open-ended questions. Participant shows genuine interest in the partner and follows up on the partner's remarks with warmth or enthusiasm.

We developed a Japanese version of the SPRS. As suggested by an earlier work [26], in its guide to good practices in translation and cultural adaptation, we first translated the scale's contents from English to Japanese and then from Japanese to English. This forward and backward translation method provides quality control to verify consistency between the translation and the original version. Inter Group Corp (Osaka, Japan), a company specializing in language and cross-cultural adaptation, translated the SPRS at our request. This Japanese version of the scale is available upon request to the authors.

SST Roleplay

We used a Japanese SST dataset collected in a previous study [14]. We collected SST data at Nara Institute of Science and Technology for the control group and at Nara Medical University for the clinical group. The dataset includes data from 50 adults with the following characteristics: 16 in the ASD group, 15 in the schizophrenia group, and 19 in the control group. We collected these psychiatric or neurodevelopmental clinical groups to investigate their social communication difficulties. We included these groups because the main SST clients in clinical facilities are those with ASD and schizophrenia. The control group participants were hired by a recruitment agency. Two psychiatrists with SST experience joined this study as trainers (roleplay partners). All participants performed SST using roleplay and feedback from the trainer. Four basic skills were performed as follows: asking for requests, declining requests, telling about positive feelings, and listening to others (ie, asking, declining, telling, and listening). Each roleplay lasts 1 to 3 minutes. Some participants performed SST multiple times, although our analysis was done on the initial roleplay data.

We excluded from the control group participants who had undergone eye surgery or had a history of psychiatric hospitalization. For the clinical group, we excluded participants

XSL•FO
RenderX

who scored less than 70 on the third edition of the Wechsler Adult Intelligence Scale [27]. We excluded participants with drug or alcohol addictions, personality disorders, or organic mental disorders. The data collection period was from January 2020 to January 2021. Due to COVID-19 concerns, a transparent partition was placed between the participants and trainers. Two video cameras were placed behind each conversationalist to record the other individual at chest level from the participants' faces. We also recorded a video and used 2 Kinect (Microsoft Corporation) sensors installed at the side of each conversationalist.

Questionnaire

We assessed each participant using the following questionnaires for correlation analysis with the SPRS and action units: the Facial Emotion Identification Test (FEIT) adopted in previous work [28], Kikuchi's scale of social skills with 18 items (Kiss-18) [29], Singelis independent-interdependent Self-Construal Scale (Self-Construal Scale) [30], the second edition of the Social Responsiveness Scale (SRS-2) [31] (Japanese version [32]), and the Japanese version of the Brief Assessment of Cognition in Schizophrenia (BACS-J) [33]. In addition, we used the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2), for the ASD group and the Positive and Negative Syndrome Scale (PANSS) for the schizophrenia group [2,34]. Where possible, we obtained the total score and subscales of each questionnaire. Excluding the FEIT, the questionnaires were collected before the SST data collection.

The FEIT, which assesses the emotional perception of facial emotions, includes facial images of 19 different people classified into 1 of 6 emotions: happiness, sadness, anger, surprise, fear, or disgust. We included this assessment because people with ASD and schizophrenia struggle with social cognition in facial images [34]. Kiss-18, which measures social skill levels, is comprised of 18 questions based on 6 social skill categories. This metric comprehensively measures social skills. Self-Construal Scale, which consists of questions on a 7-point rating scale, was developed to measure how people view themselves in relation to others. The SRS-2, composed of 65 questions, is an evaluation metric of the severity of social impairment. Although the SRS-2 was initially designed to assess people with ASD, it can also differentiate among various social communication difficulties. Its effectiveness has been investigated with clinical subjects and members of the general population [31].

To evaluate the participants during SST roleplay, we also adapted a modified roleplay test [15]. The scale is available only in Japanese. It includes items more closely related to psychopathology than the SPRS because the main target of this test is adults with schizophrenia. For this test, third-party psychiatrists watch recorded videos from both oblique and side views. Our 2 raters used a Likert scale from 1 to 5 for eye contact, body direction and distance, facial expression, voice variation, clarity, fluency, and social appropriateness. Since the required skills are situation-dependent, the social appropriateness differs depending on each SST task. Let us explain examples of social appropriateness for Bellack's basic tasks. Listening to others, which determines whether the participants paid attention to the interlocutor, includes nodding, back-channels, and other nonverbal behaviors (eg, eye contact, smiling). In expressing positive feelings, social appropriateness involves expressing attention to the interlocutor's responses and the suitability of the participant's speech content. For making requests, social appropriateness assesses whether they explained the details of their request, including what kind of help they needed. It also includes whether they listened to the interlocutor. For declining, social appropriateness concerns whether they expressed remorse and appropriate reasons for their refusal. It also includes whether they proposed alternatives to the request (eg, "I'm sorry, but I propose doing it another way next time"), which is essential to this situation.

SPRS Rating

Two third-party raters performed the annotations; all are nonexperts in psychiatry, psychology, or SST. The raters watched the recorded videos from oblique and side views for this evaluation. They initially evaluated all participants' roleplay in the first of the 4 basic tasks during the SST. Here, they watched the first 3 videos and discussed the SPRS content for evaluation practice, and then they evaluated the scores of the other videos without discussion. They were not informed directly about the nature of the groups (clinical outpatients or control), but it is possible that they may have noticed the existence of these groups by the background in the video, since each group conducted SST in a different location. These raters were not given information about the other questionnaire scales, and they watched the videos in a shuffled order regardless of the group.

Action Unit Extraction

The intensity of action units on OpenFace toolkit (version 2.0) [35] is measured in terms of mean and SD during individual roleplaying. From the action units extracted by OpenFace, we selected specific action units: AU01 (inner brow raiser), AU02 (outer brow raiser), AU04 (brow lower), AU06 (cheek raiser), and AU12 (lip corner puller). These action units represent various facial muscle activations. For our analysis, we referred to a previous study that focused on action units less directly associated with lip movement [36]. The intensity of each action unit ranges from 0 to 5, indicating a spectrum from no activation to high activation. First, the intensity of frames with less than 0.7 confidence in the data was excluded as not a number. We then applied a moving average filter on 3-second frames. We then calculated the mean and SD of the cleaned intensity data for each action unit to assess the intensity and variability, respectively.

Statistical Analysis

The analyses were performed to investigate whether the SPRS could be used to evaluate social communication skills during SST. We analyzed the interrater agreement, items' characteristics, construct validity, and internal consistency. To check the interrater agreement, we calculated the intraclass correlation coefficient (ICC, 2 k) [37]. We averaged the SPRS scores for the 2 raters for our analysis. We calculated the Spearman correlation coefficients to examine the homogeneity of the SPRS and the revised roleplay test. We assessed the factor structure of the SPRS by exploratory factor analysis with

XSL•FO
RenderX

oblimin rotation to confirm the construct validity [38]. We examined the Bartlett test of sphericity (P<.001) and the Kaiser–Meyer–Olkin test for a sampling adequacy of 0.881 to apply the exploratory factorial analyses [39]. The internal consistency was assessed by calculating the Cronbach α coefficient [40] and McDonald ω [41].

We also analyzed the total of SPRS items in relation to the 3 groups, 4 tasks of SST, and the questionnaires. In comparing the control, ASD, and schizophrenia groups, we used the nonparametric Kruskal–Wallis test. Post-hoc analysis was conducted using the Mann-Whitney U test with Bonferroni correction. In comparing the asking, declining, telling, and listening tasks, we used the Wilcoxon signed rank test with Bonferroni correction. For action units, we applied the Mann-Whitney U test with Bonferroni correction to compare the control, ASD, and schizophrenia groups. Spearman correlation coefficients were calculated between the total SPRS items, action units, and questionnaire scores for all, control, ASD, and schizophrenia groups, although some questionnaires were not obtained from all participants. We used the JASP (Jeffreys's Amazing Statistics program) for the statistical analysis [42].

Results

SPRS Statistics

We measured the interrater agreement by calculating the ICC (2 k) for validation of the SPRS annotated by the 2 raters. We opted for the random 2-way ICC method with the mean as the unit of evaluation, as shown in Table 2. The ICC ranged from 0 to 1, and all items had reliability values between 0.509 and 0.804. The ICCs were above 0.5, which indicates more than moderate reliability [22]. The score for each SPRS item was averaged for the 2 raters' annotations in subsequent analyses. The Spearman correlation performed between the SPRS items and the modified roleplay test items shows significant results (all P<.001) and positive relationships. The correlation between all of them was greater than 0.6. Exploratory factor analyses were conducted on the 5 items of the SPRS using oblique rotation to investigate underlying constructs and assess dimensionality. Following a previous study [22], a 1-factor solution was adopted. Table 3 provides the factor loadings after rotation, illustrating how each of the 5 items aligns with the selected factor. The results show that the first factor explains 82.1% of the variance, indicating strong explanatory power. Factor loadings ranged from 0.874 for the "Discomfort" item to 0.932 for the "Conversation Flow" item, reflecting high values across all items. Additionally, Table 3 displays the uniqueness values for each variable, where uniqueness represents the proportion of variance specific to each variable that the factor does not explain. Higher uniqueness suggests lower relevance of that variable within the factorial model. For example, 23.6% of the variance in the "Discomfort" item remains unexplained by the factor, whereas the "Conversation Flow" item has a smaller unexplained variance (13.2%), showing greater alignment with the factor model. In sum, the factor analysis of the SPRS identifies a single factor composed of 5 items, capturing the verbal and nonverbal social skills essential for effective communication with others.

Table . Intraclass correlation coefficients for the Social Performance Rating Scale.

Item	Intraclass correlation coefficient
Gaze	0.509
Voice quality	0.795
Length	0.777
Discomfort	0.790
Conversation flow	0.804

Table . Factor loading for the Social Performance Rating Scale score.

Item	Factor 1	Uniqueness	Mean (SD)	Skewness	Kurtosis
Gaze	0.881	0.224	3.21 (1.00)	−0.351	−0.726
Voice quality	0.914	0.164	2.93 (1.18)	0.168	−0.975
Length	0.928	0.139	3.27 (1.10)	−0.247	−0.819
Discomfort	0.874	0.236	3.04 (1.02)	−0.074	−0.659
Conversation flow	0.932	0.132	3.15 (1.07)	−0.192	−0.585

We created an overall performance measure by summing the scores of the 5 SPRS items. This measure revealed that Cronbach α was 0.957 and McDonald ω was 0.958. These results indicate that the SPRS shows excellent internal consistency.

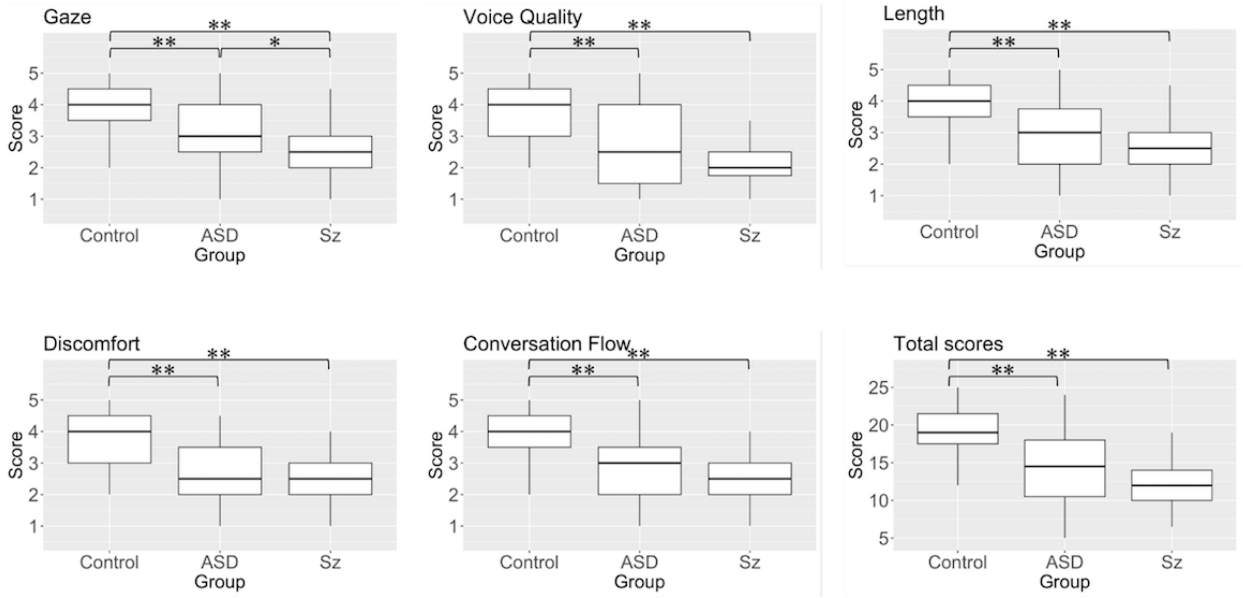
SPRS Differences Among Participant Groups

Box plots of the 5 items of the SPRS and the total of the items for the control, ASD, and schizophrenia groups are shown in Figure 1. There were significant differences in all items and in the total of the SPRS items (P<.001). Post-hoc analysis was conducted using the Mann-Whitney U test with Bonferroni correction. We found significant differences in gaze, voice

XSL•FO
RenderX

quality, length, discomfort, conversation flow, and the total of SPRS items for control-ASD and control-schizophrenia (P<.001). In ASD-schizophrenia, we found significantly greater gaze scores in the ASD group than in the schizophrenia group (P=.005).

Figure 1. Group differences of Social Performance Rating Scale scores. ASD: autism spectrum disorder; Sz: schizophrenia. *P=.02, **P<.001.



SPRS Differences Between 4 SST Tasks

SST task differences and a comparison are shown in Table 4. In the control group, we found significant differences in all items except gaze in asking-declining tasks and telling-declining tasks, and there were significant differences in the length of the asking-listening and telling-listening tasks. The declining task had lower scores than the asking and telling tasks for the control group. The length item of the listening task also had a low score compared with asking and telling tasks. We did not find significant differences between the 4 tasks in the ASD group. In the schizophrenia group, we found significant differences in the length of telling-listening tasks. For length, the listening task had a lower score than the telling task.

XSL•FO
RenderX

Table . Comparisons of Social Performance Rating Scale (SPRS) items for social skills training tasks. Parentheses represent significantly higher scored tasks. Effect size here is a rank-biserial correlation.

Item and combination			Control		Autism spectrum disorder		Schizophrenia	
			P value	Effect size	*P* value	Effect size	*P* value	Effect size
Gaze								
	Asking							
		Declining	.07	0.603	.08	−0.591	.58	−0.200
		Telling	.49	−0.289	.36	−0.295	.82	−0.091
		Listening	.48	0.219	.91	−0.044	.16	0.485
	Declining							
		Telling	.01	−0.743	.40	0.309	.97	0.026
		Listening	.15	−0.386	.17	0.491	.10	0.526
	Telling							
		Listening	.18	0.400	.24	0.500	.23	0.397
Voice quality								
	Asking							
		Declining	<.001 (Asking)[a]	0.963	.59	−0.250	.61	0.200
		Telling	.45	0.242	.45	−0.289	.22	−0.385
		Listening	.13	0.515	.81	0.091	.11	0.564
	Declining							
		Telling	.001 (Telling)[a]	−0.950	.97	−0.026	.12	−0.513
		Listening	.04	−0.600	.63	0.200	.49	0.255
	Telling							
		Listening	.28	0.341	.24	0.418	.03	0.758
Length								
	Asking							
		Declining	.001 (Asking)[a]	0.933	.52	0.198	.62	0.222
		Telling	.63	−0.167	.10	0.561	.43	−0.291
		Listening	.008 (Asking)[a]	0.824	.15	0.527	.04	0.727
	Declining							
		Telling	<.001 (Telling)[a]	−1.000	.30	0.364	.30	−0.364
		Listening	.56	−0.175	.24	0.350	.06	0.652
	Telling							
		Listening	.003 (Telling)[a]	0.912	.47	0.258	.005 (Telling)	0.923
Discomfort								
	Asking							
		Declining	.001 (Asking)[a]	0.933	.88	−0.073	.83	−0.111
		Telling	.90	−0.039	.81	0.111	.90	0.067
		Listening	.49	0.220	.60	−0.286	.29	0.382
	Declining							
		Telling	.001 (Telling)[a]	−0.912	.67	0.194	.67	0.238
		Listening	.02	−0.647	.62	−0.222	.25	0.385

XSL•FO
RenderX

Item and combination		Control		Autism spectrum disorder		Schizophrenia	
		P value	Effect size	*P* value	Effect size	*P* value	Effect size
Telling							
	Listening	.41	0.248	.49	−0.267	.32	0.417
Conversation flow							
Asking							
	Declining	.004 (Asking)[a]	0.797	.59	−0.179	.52	0.200
	Telling	.67	−0.164	.20	0.423	.15	−0.474
	Listening	.03	0.603	.50	0.242	.15	0.500
Declining							
	Telling	<.001 (Telling)[a]	−0.949	.20	0.423	.15	−0.474
	Listening	.25	−0.342	.19	0.423	.37	0.318
Telling							
	Listening	.02	0.700	>.99	0.010	.03	0.731
Total of SPRS items							
Asking							
	Declining	<.001 (Asking)[a]	0.916	.41	−0.250	.96	0.036
	Telling	.63	−0.135	.61	0.162	.31	−0.330
	Listening	.07	0.510	.48	0.206	.03	0.681
Declining							
	Telling	<.001 (Telling)[a]	−0.979	.38	0.276	.39	−0.267
	Listening	.09	−0.468	.27	0.333	.15	0.487
Telling							
	Listening	.02	0.614	.53	0.192	.02	0.758

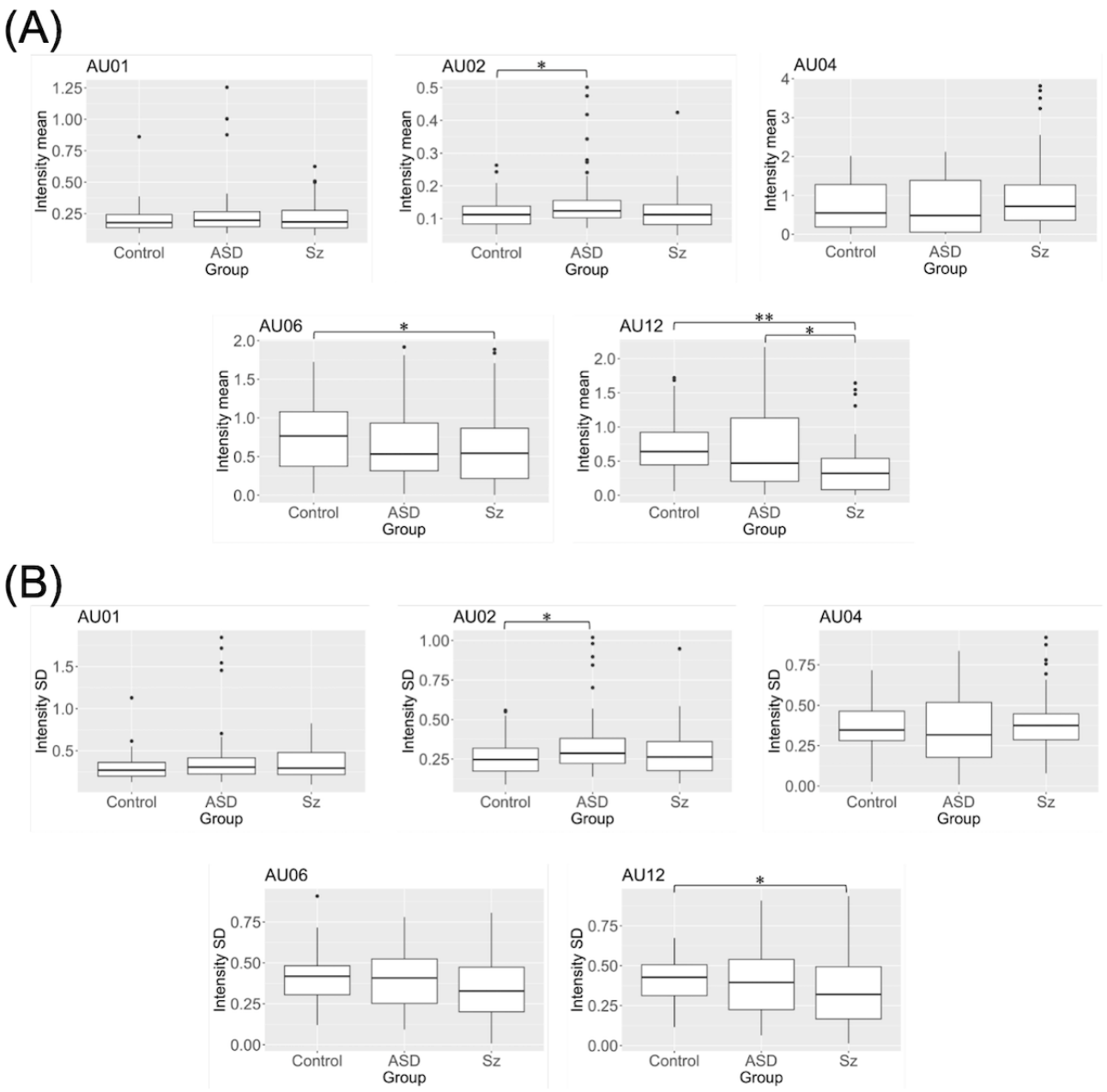
[a]Significant difference ($P<.008$).

Action Units Differences Between Participant Groups

Participant group differences and comparisons, as well as mean and SD, are shown in Figure 2. We present the cases where significant differences were observed. We found that AU02 is significantly activated in people with schizophrenia compared with other groups. For mean values of the action units, this implies that AU02 is significantly less for the control group than the ASD group. For AU06, schizophrenia is significantly less than the control group. For AU12, it is more evident that schizophrenia is significantly less than either the control or ASD group. Regarding SD, people with ASD have significantly activated AU02 compared with the control group. For AU12, schizophrenia has significantly less variance than the ASD group.

XSL•FO
RenderX

Figure 2. Action units' intensities of mean values. (**A**) **P<.001, *P<.05 (AU02 control-ASD: P=.04, AU06 control-Sz P=.09, AU12 ASD-Sz P=.04); SDs. (**B**) *P<.05 (AU02 control-ASD: P=.02, AU12 control-ASD: P=.04). ASD: autism spectrum disorder; Sz: schizophrenia.

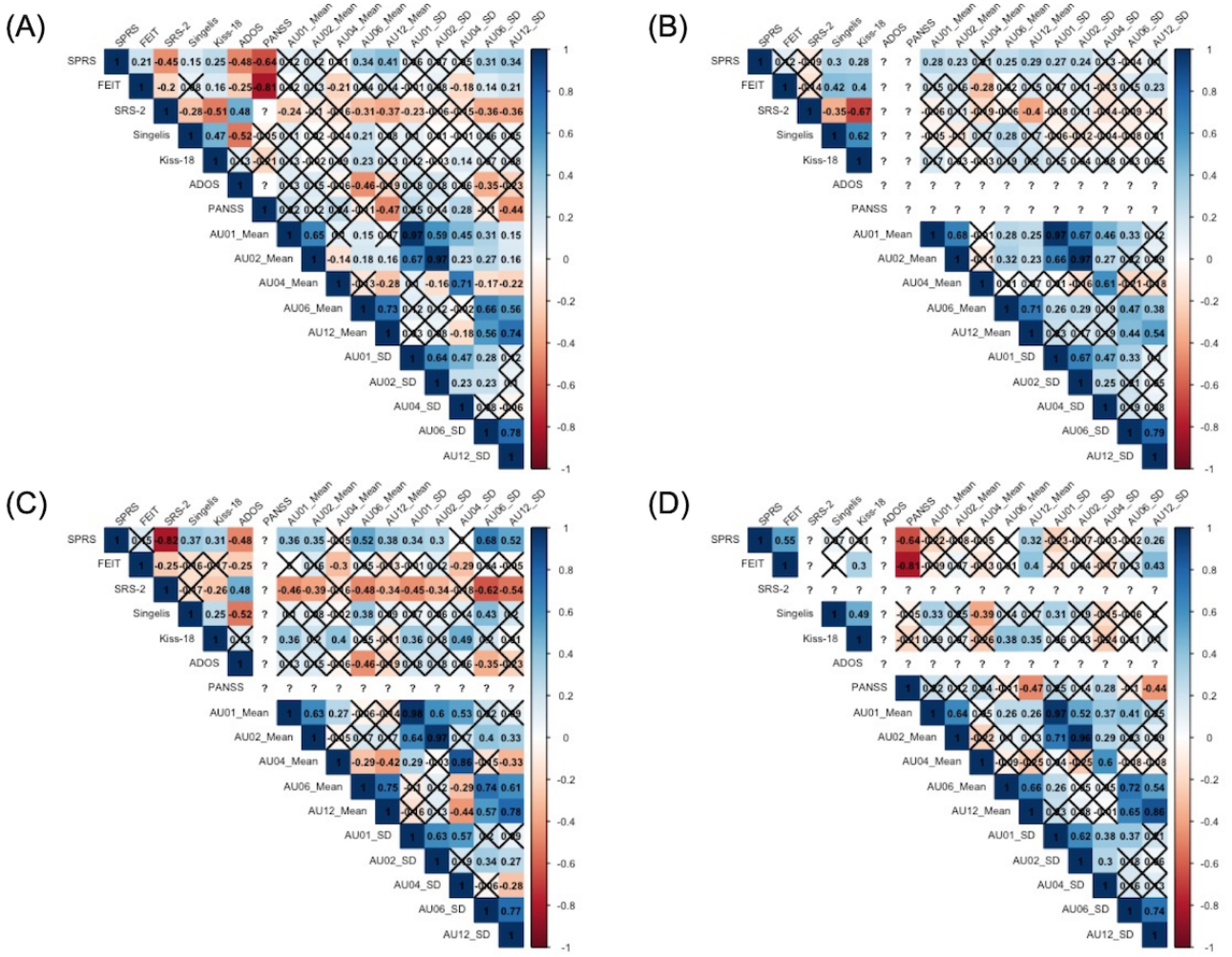


Correlation Between the SPRS, Action Units, and Questionnaire Scores

Spearman correlation coefficients were calculated between the total SPRS items and questionnaire scores for all the following groups: control, ASD, and schizophrenia. The correlation coefficients are displayed in Figure 3. Across all groups, significant correlations were found between the total SPRS items and various measures, including the FEIT, SRS-2, Self-Construal Scale, Kiss-18, ADOS-2, PANSS, as well as the mean and SD values of AU06 and AU12. These findings suggest a strong relationship between social performance and various action unit statistics, as well as questionnaire scores.

XSL•FO
RenderX

Figure 3. Correlation matrix ("?" indicates not applicable or missing data, cross mark represents nonsignificant correlations). (A) All groups, (B) control group, (C) ASD group, and (D) Sz group. ADOS: Autism Diagnostic Observation Schedule; FEIT: Facial Emotion Identification Test; Kiss-18; Kikuchi's scale of social skills with 18 items; PANSS: Positive and Negative Syndrome Scale; SPRS: Social Performance Rating Scale; SRS-2: Social Responsiveness Scale.



In the control group, significant correlations with the total SPRS items were noted for Kiss-18 and the Self-Construal Scale, reflecting the cultural emphasis on harmony over assertiveness, which may influence social communication skills in Japan. In the ASD group, significant correlations were found with Kiss-18, the Self-Construal Scale, SRS-2, BACS-J, and ADOS-2, suggesting a relationship between these questionnaires and the social performance of individuals with ASD. Notably, significant negative correlations were observed in the SRS-2, BACS-J, and ADOS-2, highlighting the complexities of social interactions within this group.

In the schizophrenia group, significant correlations with the total SPRS items were found for the FEIT, BACS-J, and PANSS. These results suggest that emotional awareness and social performance are interconnected.

Discussion

Principal Findings

This is the first study applying the SPRS and analysis of action units for SST roleplay data. In this paper, we propose evaluating roleplay performance using the SPRS. We calculated the interrater reliability between 2 raters' evaluated SPRS scores and found that the values were above 0.509 for any SPRS item. Our results show that the reliability was moderate to good [43]. Hamet Bagnou et al [22] calculated reliability among SPRS raters during collaborative tasks and showed that the reliability of gaze was the highest, but our results show that it has the lowest values. One reason could be the difference in conversation content. In our study, we applied 4 SST tasks, and the annotators needed to evaluate the SPRS based on the content of the tasks. The evaluation of SST tasks may have been more complex than that of collaborative tasks. Although there are slight differences between their study and ours, the results are nearly identical, with a reliability of 0.5 or higher in both studies.

The Spearman correlation between the SPRS items and revised roleplay test items was above 0.6. The factor analysis for the SPRS showed that the first factor explained 82.1% of the variance. The Cronbach α was 0.957, and the McDonald ω was 0.958. These results indicate that the SPRS shows excellent internal consistency and generally correlates to the modified roleplay test but is not identical.

We compared the SPRS scores by group and task. The control group differed significantly from the ASD and schizophrenia groups on every item of the SPRS. Several studies have shown that a control group shows higher social communication skills

XSL•FO
RenderX

than an ASD group and a schizophrenia group, and we showed similar results.

The ASD group had a higher gaze score than the schizophrenia group. We consider it relevant that the eye movement abnormalities were severe in the schizophrenia group compared with the ASD group [44].

We conducted analyses of the differences in the SPRS scores between tasks for each group. In the control group, the declining task had lower scores than the asking and telling tasks, except for gaze. Declining is the most difficult task of the basic SST tasks, which may have contributed to the lower scores [11]. The listening task had lower scores than the asking and telling tasks in length. In the schizophrenia group, the listening task also scored lower than the telling task in length. The control group met the trainer for the first time in the data recording. It is possible that some control participants were not interested in the topics provided by the trainer in the listening task. The ASD and schizophrenia groups had met with the trainer several times. However, the schizophrenia group had low scores. Some participants with schizophrenia might have seemed disinterested in the conversation [11], and the raters might have annotated the low scores.

The control group showed that the SPRS had significant positive correlations with Kiss-18 and with the Self-Construal Scale.

The independence of the Self-Construal Scale measures is characterized by autonomy and distinction from others [45], and it is embedded within the culture [46]. In cultures with strong interdependence, harmony is an important concern, while assertiveness tends to be discouraged. This is predominant in the East, including Japan, and it may be related to social communication skills in Japan.

In the ASD group, there were significant positive or negative correlations between the SPRS and Kiss-18 or the Self-Construal Scale; this was also the case for the control group. We found significant negative correlations in the SRS-2, BACS-J, and ADOS-2. Our results suggest a relationship between the questionnaire for ASD and the social communication skills of the ASD group. The schizophrenia group had significant positive or negative correlations between the SPRS and the FEIT, BACS-J, and PANSS. Emotional awareness and communication skills are also related [47], and we surmise that our results show a significant difference in the FEIT. Another study showed that negative symptoms related to social motivation may be primary determinants of social outcomes in schizophrenia, potentially having a greater impact than social competence [48].

The results on action units also resemble the previous findings, in which an ASD group showed less smiling as represented in AU06 and AU12 [15,17], but there was no significant difference in terms of AU01 [18]. We found that AU02 was significantly activated in people with ASD compared with other groups. Moreover, this was the first study to conduct a detailed analysis of action units for schizophrenia. We also found that the AU06 and AU12 action units were significantly deactivated in people with schizophrenia compared with the control group. Accordingly, our results demonstrated that action units could work as social behavioral markers of people with ASD or schizophrenia.

Limitations

This study is limited in terms of the population used and its generalizability. In particular, the SPRS was measured by 2 raters. Although we found high interrater correlations, a greater number of raters is needed to enhance the consistency and validity of the scale. In fact, previous work used an even larger number of raters, but a high interrater agreement was still maintained [21]. Our SST dataset was collected using outpatients with mild cases who were able to conduct SST verbally. Further study is needed to test whether the SPRS and action unit analysis could be applied to patients with more severe cases.

Conclusions

We applied and validated a Japanese version of the SPRS to SST roleplay data to measure individual social communication skills. We analyzed the SST data of 4 tasks carried out by adults with ASD or schizophrenia as well as control participants. The results found significant differences between the control and ASD as well as schizophrenia in all SPRS items. We found significantly greater gaze scores in the ASD group than in the schizophrenia group. There were also task-based differences in the schizophrenia and control groups. Our results suggest that the SPRS can be a useful tool for assessing social communication skills in different cultures and different pathologies and, in addition, can evaluate SST effectiveness as well as the modified roleplay test. We also found that the action units numbered AU06 and AU12 were significantly deactivated in people with schizophrenia compared with the control group. Possible future directions include using the SPRS for assessing social behavior during interaction with a digital agent.

Acknowledgments

This work was funded by the Japan Science and Technology Agency, Core Research for Evolutionary Science and Technology (grant JPMJCR19A5), ANR-19-JSTS-0001, and the Japan Society for the Promotion of Science (grant 22K12151).

Conflicts of Interest

None declared.

References

1. Guha M. Diagnostic and statistical manual of mental disorders: DSM-5 (5th edition). Ref Rev 2014;28(3):36-37. [doi: 10.1108/RR-10-2013-0256]

XSL•FO
RenderX

2. Bastiaansen JA, Meffert H, Hein S, et al. Diagnosing autism spectrum disorders in adults: the use of Autism Diagnostic Observation Schedule (ADOS) module 4. J Autism Dev Disord 2011 Sep;41(9):1256-1266. [doi: 10.1007/s10803-010-1157-x] [Medline: 21153873]
3. Marwaha S, Johnson S. Schizophrenia and employment - a review. Soc Psychiatry Psychiatr Epidemiol 2004 May;39(5):337-349. [doi: 10.1007/s00127-004-0762-4] [Medline: 15133589]
4. de Marchena A, Eigsti IM. Conversational gestures in autism spectrum disorders: asynchrony but not decreased frequency. Autism Res 2010 Dec;3(6):311-322. [doi: 10.1002/aur.159] [Medline: 21182208]
5. Senju A, Johnson MH. Atypical eye contact in autism: models, mechanisms and development. Neurosci Biobehav Rev 2009 Sep;33(8):1204-1214. [doi: 10.1016/j.neubiorev.2009.06.001]
6. Tanaka H, Sakti S, Neubig G, Toda T, Nakamura S. Linguistic and acoustic features for automatic identification of autism spectrum disorders in children's narrative. Presented at: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: from Linguistic Signal to Clinical Reality; Jun 27, 2014; Baltimore, Maryland, USA p. 88-96. [doi: 10.3115/v1/W14-3211]
7. Heeman PA, Lunsford R, Selfridge E, et al. Autism and interactional aspects of dialogue. Presented at: Proceedings of the SIGDIAL 2010 Conference; Sep 24-25, 2010; Tokyo, Japan p. 249-252.
8. Dodell-Feder D, Tully LM, Hooker CI. Social impairment in schizophrenia: new approaches for treating a persistent problem. Curr Opin Psychiatry 2015 May;28(3):236-242. [doi: 10.1097/YCO.0000000000000154] [Medline: 25768085]
9. Konstantareas MM, Hewitt T. Autistic disorder and schizophrenia: diagnostic overlaps. J Autism Dev Disord 2001 Feb;31(1):19-28. [doi: 10.1023/a:1005605528309] [Medline: 11439750]
10. Foss-Feig JH, McPartland JC, Anticevic A, Wolf J. Re-conceptualizing ASD within a dimensional framework: positive, negative, and cognitive feature clusters. J Autism Dev Disord 2016 Jan;46(1):342-351. [doi: 10.1007/s10803-015-2539-x] [Medline: 26267330]
11. Bellack AS, Mueser KT, Gingerich S, et al. Social Skills Training for Schizophrenia: A Step-by-Step Guide: Guilford Publications; 2013.
12. Tanaka H, Iwasaka H, Matsuda Y, Okazaki K, Nakamura S. Analyzing self-efficacy and summary feedback in automated social skills training. IEEE Open J Eng Med Biol 2021;2:65-70. [doi: 10.1109/OJEMB.2021.3075567] [Medline: 35402987]
13. Nadel J, Grynszpan O, Martin JC. Autism and socially interactive agents. In: The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application: Association for Computing Machinery; 2022:437-462. [doi: 10.1145/3563659.3563673]
14. Saga T, Tanaka H, Matsuda Y, et al. Automatic evaluation-feedback system for automated social skills training. Sci Rep 2023 Apr 26;13(1):6856. [doi: 10.1038/s41598-023-33703-0] [Medline: 37100886]
15. Sasaki T. The reliability and validity of the revised role play test for assessment of social skills in schizophrenia. Clin Psychiatry 2006;48(11):1191-1198. [doi: 10.11477/mf.1405100353]
16. Tanaka H, Saga T, Iwauchi K, et al. The validation of automated social skills training in members of the general population over 4 weeks: comparative study. JMIR Form Res 2023 Apr 27;7:e44857. [doi: 10.2196/44857] [Medline: 37103996]
17. Bangerter A, Chatterjee M, Manfredonia J, et al. Automated recognition of spontaneous facial expression in individuals with autism spectrum disorder: parsing response variability. Mol Autism 2020 May 11;11(1):31. [doi: 10.1186/s13229-020-00327-4] [Medline: 32393350]
18. Alvari G, Furlanello C, Venuti P. Is smiling the key? machine learning analytics detect subtle patterns in micro-expressions of infants with ASD. J Clin Med 2021 Apr 19;10(8):1776. [doi: 10.3390/jcm10081776] [Medline: 33921756]
19. Legiša J, Messinger DS, Kermol E, Marlier L. Emotional responses to odors in children with high-functioning autism: autonomic arousal, facial behavior and self-report. J Autism Dev Disord 2013 Apr;43(4):869-879. [doi: 10.1007/s10803-012-1629-2] [Medline: 22918860]
20. Bishay M, Palasek P, Priebe S, Patras I. SchiNet: automatic estimation of symptoms of schizophrenia from facial behaviour analysis. IEEE Trans Affect Comput 2019;12(4):949-961. [doi: 10.1109/TAFFC.2019.2907628]
21. Fydrich T, Chambless DL, Perry KJ, Buergener F, Beazley MB. Behavioral assessment of social performance: a rating system for social phobia. Behav Res Ther 1998 Oct;36(10):995-1010. [doi: 10.1016/s0005-7967(98)00069-2] [Medline: 9714949]
22. Hamet Bagnou J, Prigent E, Martin JC, Clavel C. Adaptation and validation of two annotation scales for assessing social skills in a corpus of multimodal collaborative interactions. Front Psychol 2022;13:1039169. [doi: 10.3389/fpsyg.2022.1039169] [Medline: 36389487]
23. Stevens S, Cooper R, Bantin T, Hermann C, Gerlach AL. Feeling safe but appearing anxious: differential effects of alcohol on anxiety and social performance in individuals with social anxiety disorder. Behav Res Ther 2017 Jul;94:9-18. [doi: 10.1016/j.brat.2017.04.008] [Medline: 28441522]
24. Miyamoto K, Tanaka H, Bagnou JH, et al. Social performance rating during social skills training in adults with autism spectrum disorder and schizophrenia. Presented at: 2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW); Sep 10-13, 2023; Cambridge, MA, USA p. 1-8. [doi: 10.1109/ACIIW59127.2023.10388083]

XSL•FO
RenderX

25. Likert R. A Technique for the Measurement of Attitudes Archives of Psychology de Marchena A: Columbia University; 1932.
26. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health 2005;8(2):94-104. [doi: 10.1111/j.1524-4733.2005.04054.x] [Medline: 15804318]
27. Tulsky DS, Chiaravalloti ND, Palmer BW, et al. The Wechsler Memory Scale: a new perspective. In: Clinical Interpretation of the WAIS-III and WMS-III: Elsevier; 2003:93-139. [doi: 10.1016/B978-012703570-3/50007-9]
28. Horan WP, Kern RS, Shokat-Fadai K, Sergi MJ, Wynn JK, Green MF. Social cognitive skills training in schizophrenia: an initial efficacy study of stabilized outpatients. Schizophr Res 2009 Jan;107(1):47-54. [doi: 10.1016/j.schres.2008.09.006] [Medline: 18930378]
29. Kikuchi A. Measure social skill: handbook of kiss-18 (110-115). : Toukyou: Sankou Digital Production Press; 2007.
30. Singelis TM, Sharkey WF. Culture, self-construal, and embarrassability. J Cross Cult Psychol 1995 Nov;26(6):622-644. [doi: 10.1177/0022022195026006007]
31. Constantino JN, Gruber CP. Social Responsiveness Scale: SRS-2: Western Psychological Services; 2012.
32. Kamio Y, Inada N, Moriwaki A, et al. Quantitative autistic traits ascertained in a national survey of 22 529 Japanese schoolchildren. Acta Psychiatr Scand 2013 Jul;128(1):45-53. [doi: 10.1111/acps.12034] [Medline: 23171198]
33. Kaneda Y, Sumiyoshi T, Keefe R, Ishimoto Y, Numata S, Ohmori T. Brief assessment of cognition in schizophrenia: validation of the Japanese version. Psychiatry Clin Neurosci 2007 Dec;61(6):602-609. [doi: 10.1111/j.1440-1819.2007.01725.x] [Medline: 18081619]
34. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull 1987;13(2):261-276. [doi: 10.1093/schbul/13.2.261] [Medline: 3616518]
35. Baltrusaitis T, Zadeh A, Lim YC, Morency LP. OpenFace 2.0: facial behavior analysis toolkit. Presented at: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018); May 15-19, 2018; Xi'an, China p. 59-66. [doi: 10.1109/FG.2018.00019]
36. Yang L, Woo J, Achard C, Pelachaud C. Exchanging... watch out!. arXiv. Preprint posted online on Nov 8, 2023. [doi: 10.48550/arXiv.2311.04747]
37. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [doi: 10.1037//0033-2909.86.2.420] [Medline: 18839484]
38. Costello AB, Osborne J. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Pract Assess Res Eval 2005;10(1). [doi: 10.7275/jyj1-4868]
39. Beavers AS, Lounsbury JW, Richards JK, et al. Practical considerations for using exploratory factor analysis in educational research. Pract Assess Res Eval 2013;18(1):6. [doi: 10.7275/qv2q-rk76]
40. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep;16(3):297-334. [doi: 10.1007/BF02310555]
41. McDonald RP. Test Theory: A Unified Treatment: Psychology Press; 2013.
42. JASP (version 0.17). JASP. 2023. URL: https://jasp-stats.org/ [accessed 2024-12-23]
43. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163. [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
44. Shiino T, Miura K, Fujimoto M, et al. Comparison of eye movements in schizophrenia and autism spectrum disorder. Neuropsychopharmacol Rep 2020 Mar;40(1):92-95. [doi: 10.1002/npr2.12085] [Medline: 31774635]
45. Kam C, Zhou X, Zhang X, Ho MY. Examining the dimensionality of self-construals and individualistic–collectivistic values with random intercept item factor analysis. Pers Individ Dif 2012 Oct;53(6):727-733. [doi: 10.1016/j.paid.2012.05.023]
46. Markus HR, Kitayama S. Culture and the self: implications for cognition, emotion, and motivation. Psychol Rev 1991;98(2):224-253. [doi: 10.1037//0033-295X.98.2.224]
47. Uzun G, Lok N. Effect of emotional awareness skills training on emotional awareness and communication skills in patients with schizophrenia: a randomized controlled trial. Arch Psychiatr Nurs 2022 Jun;38:14-20. [doi: 10.1016/j.apnu.2022.01.005] [Medline: 35461642]
48. Robertson BR, Prestia D, Twamley EW, Patterson TL, Bowie CR, Harvey PD. Social competence versus negative symptoms as predictors of real world social functioning in schizophrenia. Schizophr Res 2014 Dec;160(1-3):136-141. [doi: 10.1016/j.schres.2014.10.037] [Medline: 25468184]

Abbreviations

ADOS-2: Autism Diagnostic Observation Schedule, Second Edition
ASD: autism spectrum disorder
BACS-J: Brief Assessment of Cognition in Schizophrenia
FEIT: Facial Emotion Identification Test
ICC: intraclass correlation coefficient
JASP : Jeffreys's Amazing Statistics Program

XSL•FO
RenderX

Kiss-18: Kikuchi's scale of social skills with 18 items
PANSS: Positive and Negative Syndrome Scale
SPRS: Social Performance Rating Scale
SRS-2: Social Responsiveness Scale
SST: social skills training

Edited by A Mavragani; submitted 08.04.24; peer-reviewed by P Harvey; revised version received 04.11.24; accepted 06.11.24; published 10.01.25.

Please cite as:
Tanaka H, Miyamoto K, Hamet Bagnou J, Prigent E, Clavel C, Martin JC, Nakamura S
Analysis of Social Performance and Action Units During Social Skills Training: Focus Group Study of Adults With Autism Spectrum Disorder and Schizophrenia
JMIR Form Res 2025;9:e59261
URL: https://formative.jmir.org/2025/1/e59261
doi:10.2196/59261

© Hiroki Tanaka, Kana Miyamoto, Jennifer Hamet Bagnou, Elise Prigent, Céline Clavel, Jean-Claude Martin, Satoshi Nakamura. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Unified Mobile App for Streamlining Verbal Autopsy and Cause of Death Assignment in India: Design and Development Study

Harleen Kaur[1*], MSc; Stuti Tripathi[1*], MSc; Manjeet Singh Chalga[2*], PhD; Sudhir K Benara[1], MD; Amit Dhiman[2], BTECH; Shefali Gupta[1], MPH; Saritha Nair[2], PhD; Geetha Menon[1], PhD; B K Gulati[1], PhD; Sandeep Sharma[3], MD; Saurabh Sharma[1], MD

[1]ICMR-National Institute for Research in Digital Health and Data Science, Ansari Nagar, New Delhi, India

[2]Indian Council of Medical Research, New Delhi, India

[3]ICMR-National JALMA Institute for Leprosy & Other Mycobacterial Diseases, Agra, India

[*]these authors contributed equally

Corresponding Author:
Saurabh Sharma, MD
ICMR-National Institute for Research in Digital Health and Data Science, Ansari Nagar, New Delhi, India

Abstract

Background: Verbal autopsy (VA) has been a crucial tool in ascertaining population-level cause of death (COD) estimates, specifically in countries where medical certification of COD is relatively limited. The World Health Organization has released an updated instrument (Verbal Autopsy Instrument 2022) that supports electronic data collection methods along with analytical software for assigning COD. This questionnaire encompasses the primary signs and symptoms associated with prevalent diseases across all age groups. Traditional methods have primarily involved paper-based questionnaires and physician-coded approaches for COD assignment, which is time-consuming and resource-intensive. Although computer-coded algorithms have advanced the COD assignment process, data collection in densely populated countries like India remains a logistical challenge.

Objective: This study aimed to develop an Android-based mobile app specifically tailored for streamlining VA data collection by leveraging the existing Indian public health workforce. The app has been designed to integrate real-time data collection by frontline health workers and seamless data transmission and digital reporting of COD by physicians. This process aimed to enhance the efficiency and accuracy of COD assignment through VA.

Methods: The app was developed using Android Studio, the primary integrated development environment for developing Android apps using Java. The front-end interface was developed using XML, while SQLite and MySQL were employed to streamline complete data storage on the local and server databases, respectively. The communication between the app and the server was facilitated through a PHP application programming interface to synchronize data from the local to the server database. The complete prototype was specifically built to reduce manual intervention and automate VA data collection.

Results: The app was developed to align with the current Indian public health system for district-level COD estimation. By leveraging this mobile app, the average duration required for VA data collection to ascertainment of COD, which typically ranges from 6 to 8 months, is expected to decrease by approximately 80%, reducing it to about 1 - 2 months. Based on annual caseload projections, the smallest administrative public health unit, health and wellness centers, is anticipated to handle 35 - 40 VA cases annually, while medical officers at primary health centers are projected to manage 150 - 200 physician-certified VAs each year. The app's data collection and transmission efficiency were further improved based on feedback from user and subject area experts.

Conclusions: The development of a unified mobile app could streamline the VA process, enabling the generation of accurate national and subnational COD estimates. This mobile app can be further piloted and scaled to different regions to integrate the automated VA model into the existing public health system for generating comprehensive mortality statistics in India.

(JMIR Form Res 2025;9:e59937) doi:10.2196/59937

KEYWORDS

verbal autopsy; cause of death; mortality; mHealth; public health; India; mobile health

Introduction

Effective and efficient public health surveillance systems rely heavily on communication, collaboration, and information sharing. Improving traditional data collection and reporting methods is essential for supporting surveillance and response efforts. Cause-specific mortality is an important component in planning and implementing public health interventions. Digital technologies, especially in India, have been instrumental in achieving these goals while advancing the vision of transforming

XSL•FO
RenderX

into a digitally empowered society. This revolution has accelerated the growth of health care services. These measures enable public health professionals across various regions to collaborate and communicate effectively while gathering crucial information efficiently, thereby facilitating rapid and targeted interventions by generating and disseminating public health information [1].

Cause of death (COD) information is essential for evaluating health interventions and programs at all levels. Ideally, civil registration systems provide relevant information related to births, deaths, and COD in any country [2-4]. However, in countries with a large population such as India, obtaining comprehensive and reliable information on COD has been a major challenge. The death registration rates in India have shown considerable progress with approximately 93% of estimated deaths officially registered. However, only 19% of these registered deaths have a medically certified cause of death (MCCD) [5]. To address this gap, research studies have suggested that detailed household interviews with family members of the deceased, known as verbal autopsy (VA), can improve the quality of information on COD [6-8]. The primary aim of VA is to ascertain population-level COD estimates. In a VA, the COD is ascertained based on an interview with next of kin or other caregivers using a standardized questionnaire that gathers information on signs, symptoms, medical history, and circumstances preceding death [9]. Data collected during these interviews is then processed and analyzed to assign a probable COD. Standard VA instruments include a VA questionnaire, a list of COD, and diagnostic criteria (based on either expert or data-driven algorithms) for assigning a COD. The World Health Organization's (WHO) VA instrument contains a concise list of COD that are of public health importance. These can be determined through a significant number of questions designed for use in VA interviews and with automated or physician-coded COD assignment using analytical software. The Office of the Registrar General of India has been successfully conducting VA since 1999, with the help of trained physicians to assign COD [4,10].

The existing WHO VA Instrument is compatible with the Open Data Kit (ODK) Android app and has been tested for feasibility in the Indian context [11]. The ODK app is designed to be used in various scenarios such as surveillance, surveys, research, and crisis response. The current ODK app provides diverse features such as reducing data entry costs and preventing missing or invalid data from VA questionnaires. However, the development of mobile apps often stems from specific requirements for national-level VA studies and surveillance programs. This study by the Indian Council of Medical Research–National Institute for Research in Digital Health and Data Science (ICMR-NIRDHDS) details the development of a unified mobile app to streamline the data collection process during VA by using the existing public health workforce. This app is designed to support the line listing of death records, with further segregation of data by user and field catchment area, which enhances scalability and efficiency in data management. As the volume of cases increases, integration of these capabilities directly into the app helps prevent bottlenecks in data collection and analysis, streamlining the entire process. Additionally, the app includes automated data transfer for physician-coded verbal autopsy (PCVA), reducing the need for manual data extraction and sharing, which accelerates data processing and minimizes delays. It could facilitate streamlining of the data collection process and enable physicians to assign COD within a single app, completely integrated with the public health system from the subcenters to district hospitals.

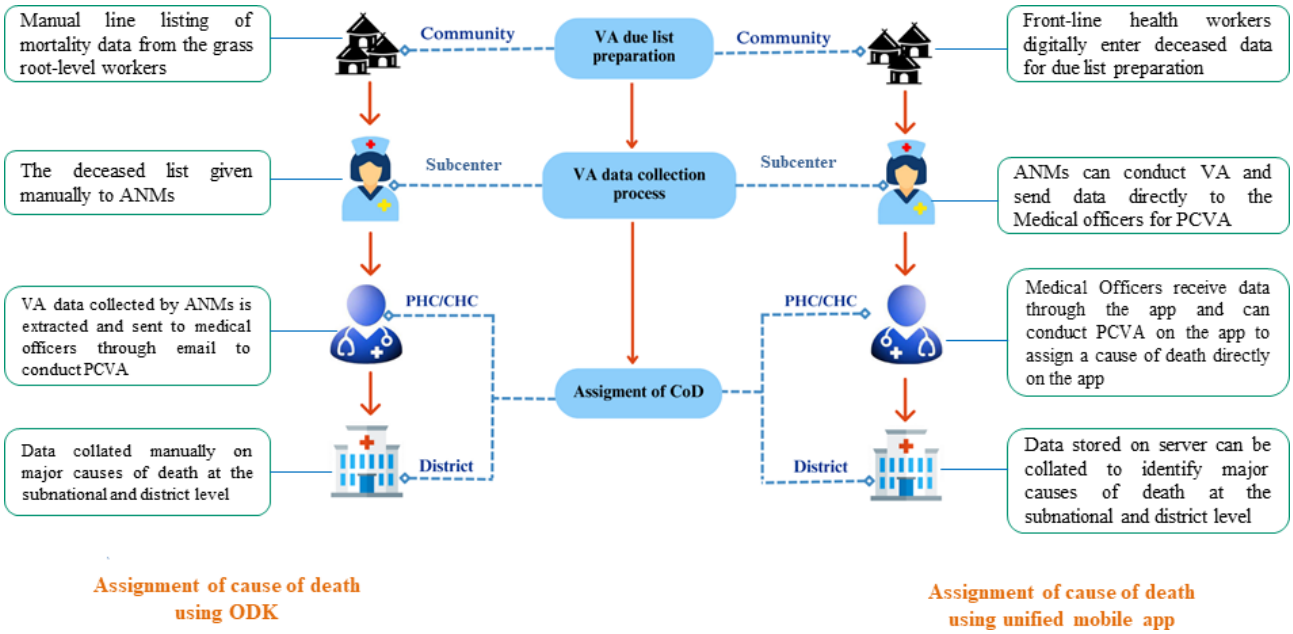
Methods

Study Design

The WHO's VA instrument was tested for its feasibility in the Indian context by using the computer-assisted personal interviewing device-based data collection method. The ICMR-NIRDHDS has thus taken the initiative to develop a mobile app to streamline VA data collection, which may be integrated into the existing public health system. A graphical example depicting the comparison of the existing ODK-based VA tool and the developed unified mobile app is shown in Figure 1.

XSL•FO
RenderX

Figure 1. Step-by-step workflow for cause of death assignment using the unified mobile app, highlighting data collection, verification, and physician review. ANM: auxiliary nursing midwife; CHC: community health care; CoD: cause of death; ODK: Open Data Kit; PCVA: physician-coded verbal autopsy; PHC: primary health care; VA: verbal autopsy.



Phase 1: Ideation, Planning

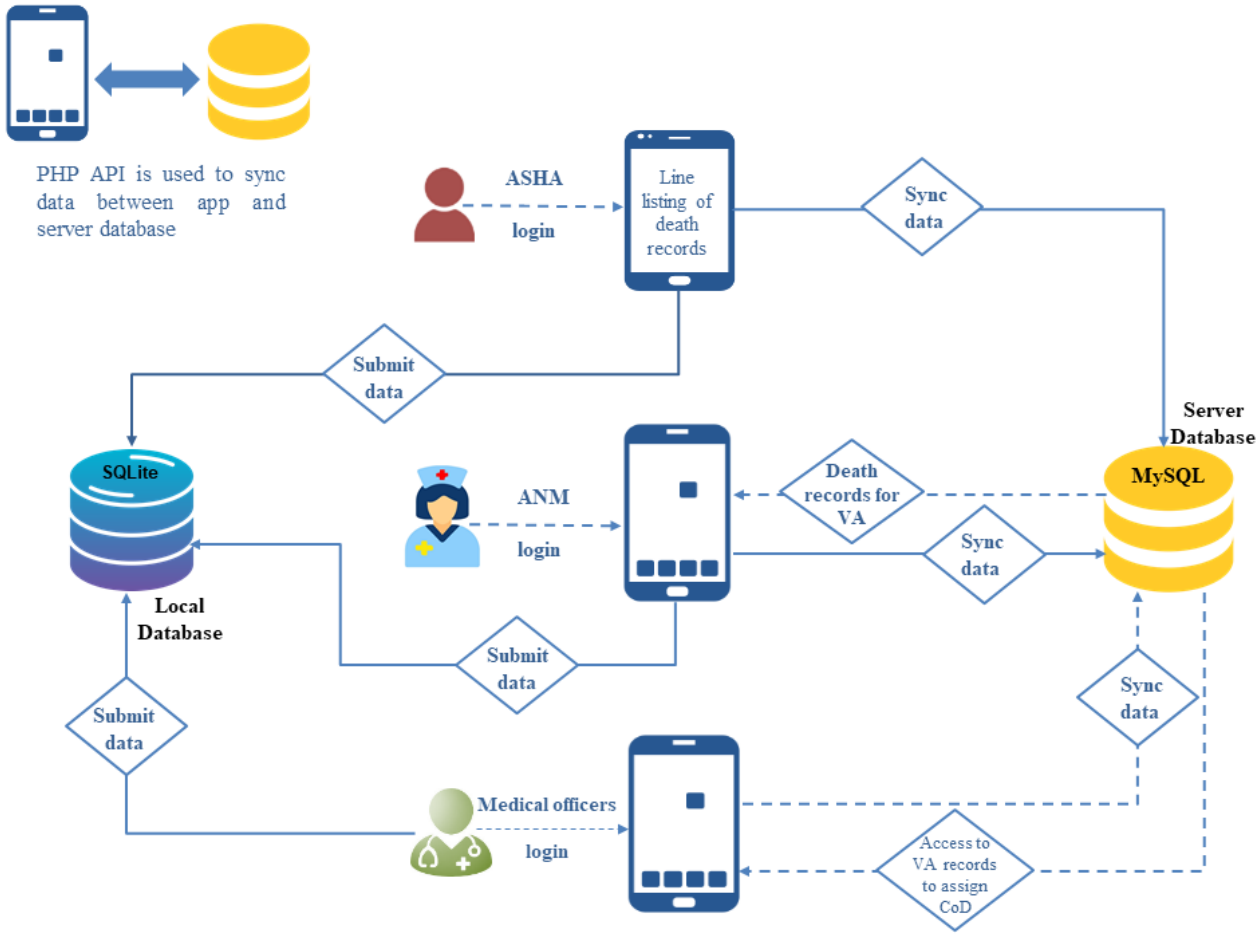
Reliable cause-specific mortality statistics are required by administrators, policy planners, and researchers for evidence-based decision-making regularly. Assigning COD using PCVA and computer-coded VA methods has already been tested in urban as well as rural areas of India using the existing ODK app [11]. However, this approach involved manual intervention at various stages of the data collection process. Initial automated line listing of death records by accredited social health activist (ASHA) workers is not supported by the ODK and is instead maintained manually, followed by subsequent VAs conducted by auxiliary nursing midwives (ANMs), community health officers, and field investigators. Data collected during the VA process was then extracted from the ODK platform and transferred via email to medical officers for COD assignment. Manual interventions requiring an extensive workforce in large-scale studies are time-consuming and can potentially compromise the data quality and subsequent outcomes. Therefore, this study aimed to develop a unified app for eliminating the manual interventions at each stage of VA data collection and COD assignment, from death reporting to conducting PCVA in rural settings. Data collected using mobile or handheld devices such as computer-assisted personal interviewing tablets could also be transferred electronically to physicians, enabling them to assign COD efficiently. In addition to providing community-based perspectives on rural mortality, it also served as a basis for testing VA methods on an operational level. This app also offers offline data collection without an internet connection and limited resources. Hence, the primary goal of the study was to efficiently implement the WHO VA 2022 instrument in rural settings through the existing public health system, with plans for further expansion to integrate automated algorithms for assigning COD.

Phase 2: App Development

Having conceptualized the solution, the next step was the development of the VA app by the team of ICMR-NIRDHDS. It was developed using Android Studio, which is the primary IDE for developing Android apps. This app was based on the WHO 2022 VA questionnaire comprising three modules, for example, adult, children, and neonate (data elements' details are given in Multimedia Appendix 1). The entire app was developed using Java for programming and XML for designing the front-end interface. Data storage was streamlined using the MySQL database management system on a secure and authenticated server, while SQLite was used for local data storage within the app. The initial submission of data by each user is stored on the local database. Next, the user selects the Sync button to update data on the centralized server. Communication between the app and server was facilitated through a PHP application programming interface (API). The API acts as a bridge between the mobile app and the server to facilitate data synchronization. During this process, the app sends HTTP requests to the server using specific PHP API end points. These end points were designed to receive the submitted data, perform validation and sanitization, and subsequently insert or update the records in the MySQL server database. After data submission, the server response is sent back to the mobile app to update the corresponding local database records. Given the data complexity, we opted for a relational database model for comprehensive data analysis and to ensure structured data organization and efficient retrieval. Further, the developed app allowed the data collectors to partially complete the surveys and resume them at a later stage or edit responses after the initial submission of the survey. This facilitated standardized data collection for subsequent review by physicians involved in the study. The complete prototype was designed with the ability to capture data with minimal cognitive effort. The navigational structure depicting the complete network is shown in Figure 2. Snapshots of the user-based interface are shown in Figure 3.

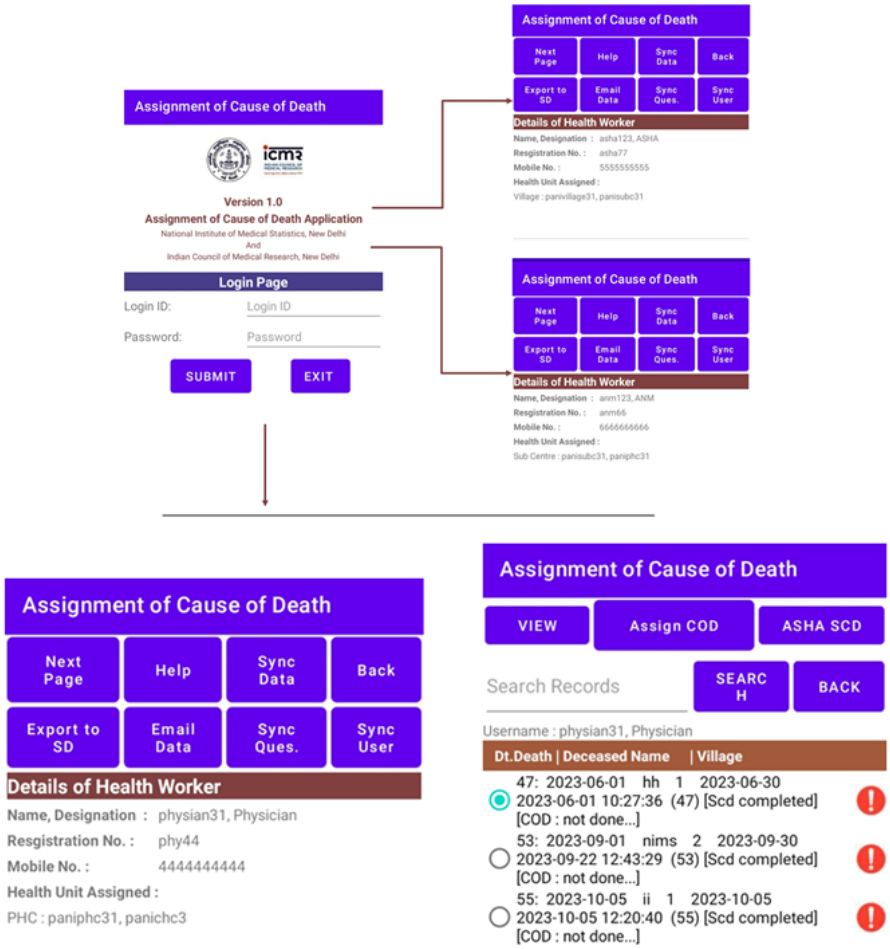
XSL•FO
RenderX

Figure 2. Diagram depicting the step-by-step data flow from user input in the mobile app to the central database, highlighting the distinct pathways for different user roles. ANM: auxiliary nursing midwife; API: application programming interface; ASHA: accredited social health worker; CoD: cause of death; VA: verbal autopsy.



XSL•FO
RenderX

Figure 3. Customized app interface designed for specific user roles to streamline tasks for accredited social health workers, auxiliary nursing midwives, and medical officers.



Phase 3: Data Management, Testing, and Quality Assurance

In this phase, the mobile app underwent field testing in both English and Hindi as an integral part of the methodology. Using APIs and servers for data communication ensured effective and secure data management, maintaining data integrity. Multiple data validation checks, skip patterns, and quality assurance measures were incorporated into the survey to maintain the accuracy and integrity of the data.

Ethical Considerations

The study was reviewed and approved by the ICMR-NIMS Ethics Committee on Health Research (ICMR_NIMS_2020). Informed consent was obtained from all subjects involved in the study. Data security and patient privacy were prioritized throughout the entire process. The app adheres to strict security standards, with access controls and user authentication. Patient identifiers are anonymized to protect their identities. Subject compensation is not applicable in the present study and hence was not provided.

Results

Overview

The mobile app was developed with a focus on the integration and implementation of VA within the existing Indian public health system. Health and wellness centers, which serve approximately 5000 individuals, were projected to manage around 35 - 40 VA cases annually, translating to about 3 - 4 cases per month per center. The ANMs and community health officers, who are the primary health care providers in these centers, are tasked with performing and recording these autopsies efficiently using the app. Similarly, primary health centers catering to larger populations of around 20,000 - 30,000 individuals are projected to handle a higher annual caseload of 150 - 200 VAs, equating to 12 - 16 autopsies per month per center for each medical officer. These projections were based on population densities and historical data on mortality rates in these regions, allowing for more accurate workload estimation. The app was specifically designed with these projections in mind, optimizing the workflow to manage these volumes with ease.

The traditionally lengthy process of VA data collection and COD determination, which generally spans 6 - 8 months, is anticipated to be significantly expedited with this app, reducing this duration by approximately 80%. This would shorten the

XSL•FO
RenderX

time from VA data collection to final COD determination to approximately 1 - 2 months.

To ensure the app's practicality and usability, extensive feedback was gathered from subject matter experts, including public health officials and field workers. This input resulted in several enhancements, such as simplifying data input processes and ensuring compatibility with existing health information systems.

Prototype Evaluation: Experts' Insights and Recommendations

The unified mobile app used in this study was developed according to the outlined objectives and methodologies. It was critically reviewed by the subject experts and its contents were refined according to their feedback and suggestions until final approval. Therefore, the developed app has automated the VA process and has been rated as efficient, user-friendly, and easy to navigate by the users and various stakeholders, strengthening the existing public health system. Textbox 1 lists the recommendations by the subject experts.

Textbox 1. Experts' recommendations implemented in the app.

Content

- Customization of the user interface for ease of use by frontline health workers
- Addition of detailed information of the health care workers along with the demarcated health care centers, enabling the generation of an area-wise due list for conducting the verbal autopsy (VA)
- Incorporation of the list of cause of death with VA codes, which can be assigned by medical officers

Interaction and workflow pathways

- Streamlining of workflow pathways for frontline workers (eg, accredited social health activist [ASHA], auxiliary nursing midwife [ANM]) and medical officers
- Transition from manual line listing to digital due list preparation of deceased data
- Addressing the language barrier by facilitating bilingual VA process in both English and Hindi
- Creation of a user-friendly interface with role-based access defined for each user group (eg, ANM, ASHA, medical officers)

App Overview

Figure 3 illustrates the app interface, beginning with the login module accessible only to authorized users. Role-based login credentials for specific roles such as ANM, ASHA, and medical officers will allow them to navigate to specific sections on the app. ASHA workers will be able to digitally enter the deceased data according to the demarcated health centers. This digitized data collection marks a significant shift from traditional manual methods, ensuring accuracy and minimizing errors.

This stored data will be further accessible to ANMs to conduct a VA using a bilingual interface (in Hindi and English), according to their feasibility. The VA process allows for partial saving of information, which can be completed at a later time—an important feature for real-world scenarios where time constraints are often encountered. Once ANMs sync the VA data, it becomes available for medical officers to review and assign a COD, facilitated by the integration of VA codes directly in the app.

The app's robust design, which is grounded in expert feedback, facilitates an integrated and real-time approach to VA and COD assignment.

Discussion

Overall death registration rates have shown significant improvement in India. However, the MCCD currently covers only a smaller percentage of deaths compared to the total registered deaths [12-14]. At present, the Office of the Registrar General of India conducts VA for a sample of deaths through the Sample Registration System survey [15]. VA holds significant potential to supplement MCCD and strengthen the mortality statistics at the population level [16-18]. In the conventional system, VA data collection involves a multistep, manual process where health workers fill out paper-based questionnaires and rely on physicians to review and assign causes of death, often leading to substantial delays.

The existing WHO VA instrument is compatible with the ODK software and has undergone feasibility testing in the Indian context. However, this app is primarily designed for conducting VA and requires manual intervention at each stage, which limits its flexibility and adaptability across different states. To scale up the existing VA survey, it could be integrated with the existing public health system through the grassroots-level workforces and digital health technology [9,19]. Over the past decade, the role of digital health technologies in public health has been increasing [20]. In India, several digital health initiatives, such as doctor consultation apps, e-pharmacies, and e-Sanjeevani, were developed to strengthen the existing public health system [21-24]. As an advancement, the institute has developed a unified mobile app to streamline the COD assignment process via VA. By utilizing the public health system, this mobile app for VA can act as a stepping stone toward addressing the challenge of timely and accessible mortality data collection in India. With the increasing penetration of mobile devices across urban and rural areas, this app has the potential to reach diverse populations, including those in remote regions. This study was able to demonstrate that a customized smartphone app for VA data collection, data

XSL•FO
RenderX

transfer, and COD identification can be used for disease burden estimation at subpopulation levels. The app is further scalable for use in larger populations, including urban regions and more tertiary hospitals. Additionally, India is characterized by its cultural and linguistic diversity, presenting a challenge in obtaining accurate VA data. The app can further accommodate multiple languages and aid in overcoming this barrier.

The mobile app developed comes with several key features and strengths: (1) it assists inline listing of the cases and due list preparation; (2) frontline workers can conduct VA through this mobile app; (3) VA cases can be automatically assigned based on the field catchment area; (4) physicians can retrieve information from the server and conduct PCVA; (5) it is bilingual, supporting both English and Hindi; (6) it has the potential to substantially improve the timeliness of mortality data reporting at both the local and national levels by reducing the overall time from 6 to 8 months to just 1 - 2 months; (7) the mobile app can be customized based on the variability of public health organizational structure across different states; (8) the app could accommodate multiple languages; (9) the app could be functional in remote areas with limited internet connectivity, as it allows users to transfer and retrieve data from the server offline; and (10) health care workers can retrieve data from the server at any point of time, enhancing the timely accessibility of the information.

However, mobile health implementation is often encountered with challenges [25,26]. Acquiring and maintaining hardware is difficult due to financial limitations and inadequate technical expertise. Developing the complete workflow from data collection to assigning the COD for each user was also challenging initially. However, it improved gradually through multiple discussions with various stakeholders. Additionally, Android apps may face compatibility issues across different devices and operating system versions, leading to inconsistent performance and user experience.

In conclusion, the development of a unified mobile app for streamlining VA and COD assignment through the public health system in India could act as a stepping stone for generating national and subnational level COD estimates. Feasibility studies based on this mobile app need to be undertaken across different regions to develop a VA model, which can be integrated into the public health system for generating mortality statistics in India.

Acknowledgments

The authors acknowledge the Indian Council of Medical Research, Delhi, for providing financial assistance for the study. This work was supported by the Indian Council of Medical Research, New Delhi (grant ID 5/4/8-32/CD/SS/2021-NCD-II).

Data Availability

All data generated or analyzed during this study are included in this published article.

Authors' Contributions

Conceptualization: Saurabh Sharma
App development: MSC, AD
App testing: AD
Graphics: HK, ST
Writing (original draft): HK, ST, Saurabh Sharma
Writing – review and editing: MSC, SKB, SG, SN, GM, BKG, Sandeep Sharma

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data elements of World Health Organization verbal autopsy questionnaire for adults, children, and neonates.
[PDF File, 119 KB - formative_v9i1e59937_app1.pdf]

References

1. Mukhi S, Dhiravani K, Micholson B, et al. An innovative mobile data collection technology for public health in a field setting. Online J Public Health Inform 2018;10(2):e202. [doi: 10.5210/ojphi.v10i2.9114] [Medline: 30349620]
2. Vital statistics of India based on the civil registration system 2016. Census India. URL: https://censusindia.gov.in/nada/index.php/catalog/42538 [accessed 2024-04-24]
3. Medical certification of cause of death. Delhi Planning. URL: https://delhiplanning.delhi.gov.in/sites/default/files/report_on_medical_certification_of_cause_of_deaths_in_delhi_2016.pdf [accessed 2024-04-24]
4. Mahapatra P, Chalapati Rao PV. Cause of death reporting systems in India: a performance analysis. Natl Med J India 2001;14(3):154-162. [Medline: 11467144]

XSL•FO
RenderX

5. Kumar R. Mortality statistics in India: current status and future prospects. Indian J Community Med 2022;47(4):476-478. [doi: 10.4103/ijcm.ijcm_614_22] [Medline: 36742978]
6. Byass P, Huong DL, Minh HV. A probabilistic approach to interpreting verbal autopsies: methodology and preliminary validation in Vietnam. Scand J Public Health Suppl 2003;62:32-37. [doi: 10.1080/14034950310015086] [Medline: 14649636]
7. Rao C, Bradshaw D, Mathers CD. Improving death registration and statistics in developing countries: lessons from sub-Saharan Africa. S Afr J Demogr 2004;9(2):81-99. [doi: 10.2307/20853272]
8. Fottrell E, Byass P. Verbal autopsy: methods in transition. Epidemiol Rev 2010;32:38-55. [doi: 10.1093/epirev/mxq003] [Medline: 20203105]
9. Verbal autopsy standards: the 2022 WHO verbal autopsy instrument. World Health Organization. URL: https://cdn.who.int/media/docs/default-source/classification/other-classifications/autopsy/2022-va-instrument/verbal-autopsy-standards_2022-who-verbal-autopsy-instrument_v1_final.pdf [accessed 2024-04-18]
10. Jha P, Gajalakshmi V, Gupta PC, et al. RGI-CGHR Prospective study of one million deaths in India: rationale, design, and validation results. PLoS Med 2006 Feb;3(2):e18. [doi: 10.1371/journal.pmed.0030018] [Medline: 16354108]
11. Benara SK, Sharma S, Juneja A, et al. Evaluation of methods for assigning causes of death from verbal autopsies in India. Front Big Data 2023;6:1197471. [doi: 10.3389/fdata.2023.1197471] [Medline: 37693847]
12. Report on medical certification of cause of deaths in Delhi-2019. Government of the National Capital Territory of Delhi. URL: https://des.delhi.gov.in/sites/default/files/DES/generic_multiple_files/final_mccd_2019.pdf [accessed 2024-04-24]
13. Saikia N, Kumar K, Das B. Death registration coverage 2019-2021, India. Bull World Health Organ 2023 Feb 1;101(2):102-110. [doi: 10.2471/BLT.22.288889] [Medline: 36733620]
14. Ray S, Das DK. Verbal autopsy - concept and role in Indian scenario. J Comp Health 2019;7(2):12-15. [doi: 10.53553/JCH.v07i02.003]
15. Krishnan A, Gupta V, Nongkynrih B, et al. Mortality in India established through verbal autopsies (MINErVA): strengthening national mortality surveillance system in India. J Glob Health 2020 Dec;10(2):020431. [doi: 10.7189/jogh.10.020431] [Medline: 33282224]
16. Polprasert W, Rao C, Adair T, Pattaraarchachai J, Porapakkham Y, Lopez AD. Cause-of-death ascertainment for deaths that occur outside hospitals in Thailand: application of verbal autopsy methods. Popul Health Metr 2010 May 18;8:13. [doi: 10.1186/1478-7954-8-13] [Medline: 20482760]
17. Danso SO, Manu A, Fenty J, et al. Population cause of death estimation using verbal autopsy methods in large-scale field trials of maternal and child health: lessons learned from a 20-year research collaboration in Central Ghana. Emerg Themes Epidemiol 2023 Feb 16;20(1):1. [doi: 10.1186/s12982-023-00120-7] [Medline: 36797732]
18. Firth SM, Hart JD, Reeve M, et al. Integrating community-based verbal autopsy into civil registration and vital statistics: lessons learnt from five countries. BMJ Glob Health 2021 Nov;6(11):e006760. [doi: 10.1136/bmjgh-2021-006760] [Medline: 34728477]
19. de Savigny D, Riley I, Chandramohan D, et al. Integrating community-based verbal autopsy into civil registration and vital statistics (CRVS): system-level considerations. Glob Health Action 2017;10(1):1272882. [doi: 10.1080/16549716.2017.1272882] [Medline: 28137194]
20. Iyamu I, Xu AXT, Gómez-Ramírez O, et al. Defining digital public health and the role of digitization, digitalization, and digital transformation: scoping review. JMIR Public Health Surveill 2021 Nov 26;7(11):e30399. [doi: 10.2196/30399] [Medline: 34842555]
21. Wienert J, Jahnel T, Maaß L. What are digital public health interventions? First steps toward a definition and an intervention classification framework. J Med Internet Res 2022 Jun 28;24(6):e31921. [doi: 10.2196/31921] [Medline: 35763320]
22. Maroju RG, Choudhari SG, Shaikh MK, Borkar SK, Mendhe H. Role of telemedicine and digital technology in public health in India: a narrative review. Cureus 2023 Mar;15(3):e35986. [doi: 10.7759/cureus.35986] [Medline: 37050980]
23. Markan A, Kishore A, Agarwal A, et al. Demographic profile of patients seeking teleophthalmology consultations through e-Sanjeevani: retrospective analysis of 5138 patients from North India. Indian J Ophthalmol 2022 Dec;70(12):4238-4243. [doi: 10.4103/ijo.IJO_781_22] [Medline: 36453323]
24. Agarwal N, Biswas B. Doctor consultation through mobile applications in India: an overview, challenges and the way forward. Healthc Inform Res 2020 Apr;26(2):153-158. [doi: 10.4258/hir.2020.26.2.153] [Medline: 32547812]
25. Flaxman AD, Stewart A, Joseph JC, et al. Collecting verbal autopsies: improving and streamlining data collection processes using electronic tablets. Popul Health Metr 2018 Feb 1;16(1):3. [doi: 10.1186/s12963-018-0161-9] [Medline: 29391038]
26. University of Melbourne. Challenges associated with automated VA training and rollout. Bloomberg Philanthropies: Data for Health. 2018. URL: https://data4healthlibrary.org/sites/default/files/resources/104_UMelbourne_Challenges%20associated%20with%20automated%20VA%20training%20and%20rollout.pdf [accessed 2024-08-24]

Abbreviations

ANM: auxiliary nursing midwife
API: application programming interface

XSL•FO
RenderX

ASHA: accredited social health activist
COD: cause of death
ICMR: Indian Council of Medical Research
MCCD: medically certified cause of death
NIRDHDS: National Institute for Research in Digital Health and Data Science
ODK: Open Data Kit
PCVA: physician-coded verbal autopsy
VA: verbal autopsy
WHO: World Health Organization

Edited by A Mavragani; submitted 26.04.24; peer-reviewed by S Pal, VK Kamal; revised version received 09.10.24; accepted 15.10.24; published 10.01.25.

Please cite as:
Kaur H, Tripathi S, Chalga MS, Benara SK, Dhiman A, Gupta S, Nair S, Menon G, Gulati BK, Sharma S, Sharma S
Unified Mobile App for Streamlining Verbal Autopsy and Cause of Death Assignment in India: Design and Development Study
JMIR Form Res 2025;9:e59937
URL: https://formative.jmir.org/2025/1/e59937
doi:10.2196/59937

©Harleen Kaur, Stuti Tripathi, Manjeet Singh Chalga, Sudhir K Benara, Amit Dhiman, Shefali Gupta, Saritha Nair, Geetha Menon, B K Gulati, Sandeep Sharma, Saurabh Sharma. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Public Understanding and Expectations of Digital Health Evidence Generation: Focus Group Study

Paulina Bondaronek[1,2], PhD; Jingfeng Li[2], MSc; Henry W W Potts[1], PhD

[1]Institute of Health Informatics, University College London, 222 Euston Road, London, United Kingdom

[2]School of Social Sciences, University of Westminster, London, United Kingdom

Corresponding Author:
Paulina Bondaronek, PhD
Institute of Health Informatics, University College London, 222 Euston Road, London, United Kingdom

Abstract

Background: The rapid proliferation of health apps has not been matched by a comparable growth in scientific evaluations of their effectiveness, particularly for apps available to the public. This gap has prompted ongoing debate about the types of evidence necessary to validate health apps, especially as the perceived risk level varies from wellness tools to diagnostic aids. The perspectives of the general public, who are direct stakeholders, are notably underrepresented in discussions on digital health evidence generation.

Objective: This study aimed to explore public understanding and expectations regarding the evidence required to demonstrate health apps' effectiveness, including at varying levels of health risk.

Methods: A total of 4 focus group discussions were held with UK residents aged 18 years and older, recruited through targeted advertisements to ensure demographic diversity. Participants discussed their views on evidence requirements for 5 hypothetical health apps, ranging from low-risk wellness apps to high-risk diagnostic tools. Focus groups were moderated using a structured guide, and data were analyzed using reflexive thematic analysis to extract common themes.

Results: A total of 5 key themes were established: personal needs, app functionality, social approval, expectations of testing, and authority. Participants relied on personal experiences and social endorsements when judging the effectiveness of low-risk digital health interventions, while making minimal reference to traditional scientific evidence. However, as the perceived risk of an app increased, there was a noticeable shift toward preferring evidence from authoritative sources, such as government or National Health Service endorsements.

Conclusions: The public have a preference for evidence that resonates on a personal level, but also show a heightened demand for authoritative guidance as the potential risk of digital health interventions increases. These perspectives should guide developers, regulators, and policy makers as they balance how to achieve innovation, safety, and public trust in the digital health landscape. Engaging the public in evidence-generation processes and ensuring transparency in app functionality and testing can bridge the gap between public expectations and regulatory standards, fostering trust in digital health technologies.

(JMIR Form Res 2025;9:e56523) doi:10.2196/56523

KEYWORDS

mobile apps; digital health; public expectations; evidence of effectiveness; health risk perception; effectiveness; health risk; health app; public health; well-being; public trust; diagnostic tools; safety; mobile phone

Introduction

The proliferation of health and wellness apps is substantial [1,2], with over 318,000 found in a 2018 study [3]. However, there is a wide gap between the availability and evaluation of digital health, with most publicly available apps not having been evaluated in the scientific literature [3]. They are typically rapidly developed and implemented with little formal evaluation to support their impact claims [4-6]. Many apps have proven not to be effective [7-10].

There are many reasons for this evaluation gap. The rapid advancement of digital health technologies has outpaced the development of comprehensive guidance. Governments have not wanted to hold back innovation around new technologies that promise to reduce health care costs and support economic growth. The dividing line between what should require effectiveness testing and what need not is unclear. This has all contributed to a gray area or "wild west" of regulation, a lack of clear guidelines, standards, and regulatory frameworks for evaluating digital health interventions. Stakeholders, including developers, health care providers, regulators, and users, face difficulties in navigating this uncertain terrain and ensuring that digital products and services meet appropriate standards of evidence.

What evaluation may be appropriate will vary depending on the digital technology under consideration. The National

XSL•FO
RenderX

Institute for Health and Care Excellence Evidence Standards Framework (NICE ESF) [11], for example, offers recommendations of different study methods for different technology types, broadly based on a risk stratification [11]. A distinction is sometimes made between "digital therapeutics," seen as akin to drug treatments and thus coming with rigorous evaluation [12], and "well-being apps," perhaps closer to a self-help book or going to a yoga class and not requiring the same oversight. Well-being apps are seen as less of a concern for various reasons. The direct risks associated with them are small compared with a diagnostic or therapeutic tool. However, we note there can be a significant opportunity cost to public health apps that do not work. If someone uses, for example, a smoking cessation app that is not effective, they may be less inclined to try another that would have worked.

While there have been initiatives to improve the situation (including in the United Kingdom with the NICE ESF and Evaluating Digital Health Products [11,13]; and elsewhere, for example, DiGA (Digitale Gesundheitsanwendungen) in Germany [14]), there is a live debate over what evaluation is appropriate for what digital health products. However, we have seen little in this literature asking what the populace may want. Do the public expect or want more or less evaluation of digital health products?

Public policy [15,16] and research [17] both profess the value of public participation. The technical nature of clinical research, however, presents challenges. Public understanding of clinical research can be poor, although people's perception of their own knowledge is higher [18]. Studies have focused on public understanding of randomized clinical trials, which is generally low, with ideas such as randomization and double blinding being challenging [19], as has been the case for many years [20]. This is significant as poor understanding undermines trial recruitment and public support for research, and may lead to people making poorer treatment choices [21,22].

Expectations of evidence generation will also relate to the public's broader understanding of and belief in digital health. Policy makers may have techno-utopian views, seeing digital interventions as a solution to challenges in health care [23] through a revolution in the collection and use of data [24]. However, the public can be more cautious, being fairly positive about the potential for digital health, but with some concerns, particularly around safety, privacy, and depersonalization through the loss of human contact [25-28]. Attitudes can vary with the novelty and form of the technology involved [26].

Given all this, we aimed to explore the public's understanding, expectations, and preferences regarding the evidence needed to demonstrate the effectiveness of digital health products and services. We also wanted to see how views vary by the nature of the digital health technology, particularly those that are seen as low-risk (well-being app) versus high-risk (diagnostic and therapeutic tools). We chose a focus group method as we thought the discursive nature of focus groups would help people consider a topic that may be relatively novel for them.

Methods

Sample

The research was carried out at the University of Westminster, and we sought to recruit a sample representative of the local population. We designed a recruitment poster outlining the content of the study and how to participate. The study was then advertised on social media (Twitter [subsequently rebranded X] and LinkedIn [Microsoft]) and through the University of Westminster's Black and minority ethnic community engagement, the Black and Minority Ethnic All Members Network.

The inclusion criteria were UK residents aged 18 years and older. A sample framework was used to monitor the recruitment to ensure a diverse sample of participants. Interested participants were asked to fill out screening questions including age, gender, and ethnicity. We assessed levels of digital literacy with 2 previously used questions [29], asking "How difficult is it to use your smartphone [AND] install mobile applications/apps on your smartphone without someone else helping you?", answered on a Likert scale. Participants could choose 1 of 4 meeting times for online focus groups. A total of 26 individuals contacted us, with 21 participants participating, divided into the 4 focus groups. Each participant received a participant information sheet and was told about the study goals and process and what they needed to do. If they had questions about the study, they were told they could ask them directly by email and get answers. The focus group interviews were conducted on July 11-July 12, 2023, through the University of Westminster's Microsoft Teams account. Each focus group lasted 90 minutes.

Ethical Considerations

The study was approved by the Psychology Ethics Committee at the University of Westminster (reference: ETH2223-3175). Participants received a Participant Information Sheet, provided written informed consent (Multimedia Appendix 1) through email, and could ask questions at any stage. Data were anonymized, securely stored, and handled in compliance with institutional data protection policies. Participants received vouchers valued at £20 (US $25.37) each to thank them for participating in the study. They were also provided with a summary of the interview results afterwards and given the opportunity to check and comment.

Data Collection

The focus groups used a moderator's guide. All groups were led by the same researcher (PB) for consistency. The other 2 authors attended all focus groups and took notes. The groups were recorded. We used the Microsoft Teams transcription as a starting point and then corrected this.

The goal throughout the focus groups was to explore participants' perspectives on the evidence or proof required to demonstrate the effectiveness and accuracy of different types of digital health interventions. The moderator used a Microsoft PowerPoint presentation to guide the focus group process. We started the groups by introducing our definition of digital health, providing examples of 3 different technologies—medical imaging in hospitals, treating mental health at home, and

XSL•FO
RenderX

improving fitness—to illustrate the broad spectrum of digital health. We then defined what we meant by evidence, which we loosely defined as "information or facts that support or prove something." We then introduced 5 hypothetical digital health scenarios (alternating the order of scenarios for different groups: 1 - 5; 5 - 1).

The 5 hypothetical digital health scenarios and how we described them are given in Textbox 1.

These scenarios were selected and presented in a hierarchy (reversed for half the groups) to cover the spectrum of potential safety implications. We were informed in our choice by the NICE ESF, but included more well-being apps as these are not covered by the NICE ESF.

Textbox 1. Five hypothetical digital health scenarios to guide the focus group.

Fitness app

- Focuses on promoting good health, specifically for running.
- Tracks running activities—maps the run, records your speed, and logs running session

Fitness app with artificial intelligence

- Same as the fitness app
- Uses sensors to track your heart rate
- Uses artificial intelligence to then make a personalized recommendation as to when or how much you should run based on your age, weight, and heart rate

Mindfulness app

- Helps you deal with everyday stress
- Guided relaxation sessions
- Tracks how you are feeling

Depression treatment app

- Designed for people who have been diagnosed by a doctor as having clinical depression
- Offers a treatment
- Includes thinking exercises for the user to practice

Skin cancer diagnosis app

- Allows users to take a picture of their skin, specifically moles, for a potential skin cancer diagnosis
- Provides a diagnosis of skin cancer from the picture

Data Analysis

We adopted an interpretivist epistemological framework and a social constructionism ontological stance. We used a reflexive thematic analysis [30,31] to interpret recurring themes based on the collected data and generate themes during coding. We used Reflexive Thematic Analysis Reporting Guidelines to guide the reporting of the analysis [32]. We anonymized the transcripts of the 4 focus groups, assigning a code to each participant. We also compared the transcriptions with our contemporary notes. We generated initial codes and applied them to the data, using Microsoft Excel sheets to identify common ground in participants' expectations, concerns, and views on the effectiveness of digital health products. The codes were reviewed through repeated reading and analysis of the overall data. We also paid attention to how the codes related to the different scenarios presented. A total of 5 themes were eventually developed, which we present on a continuum. The codes and themes were reviewed again. Finally, we drew a thematic map for each scenario. We compared the commonalities and differences encoded between the thematic maps for each scenario, with respect to the hierarchy of scenarios (low-risk to high-risk digital health interventions) and our continuum of themes (from individual to community to authority).

Results

Participants

We used a semi–open-ended questionnaire to screen participants. Gender, age, and ethnicity were asked as open-ended questions to provide participants the flexibility of answer (Table 1). In total, 21 people (8 women and 13 men), aged 21 - 50 years, participated in 4 focus groups. In the sample, 7 identified as Black British, 8 as Black other, 3 as White British, 2 as Asian, and 1 as Mixed. Everyone had access to smartphones, but 2 participants reported sometimes having difficulty installing software on their phones. Furthermore, 14 participants were in employment.

XSL•FO
RenderX

Table . Sample characteristics including gender, age, ethnicity, smartphone usage difficulty, and employment status (gender, age, and ethnicity were collected using open-ended questions).

Characteristic		Total (N=21), n
Self-reported gender		
	Women	8
	Men	13
Age group (years)		
	18 - 25	5
	26 - 35	13
	36 - 45	2
	46 - 55	1
Ethnicity		
	Black British	7
	Black other	8
	Asian	2
	White British	3
	Mixed	1
Use smartphone		
	Not difficult	21
Install mobile apps		
	Not difficult	19
	Somewhat difficult	2
Employment status		
	Full-time paid job (31+ hours)	9
	Part-time paid job (<31 hours)	5
	Doing paid work on a self-employed basis or within one's own business	4
	Out of work (more than 6 months)	1
	Unpaid work for a business, community or voluntary organization	1
	Prefer not to say	1

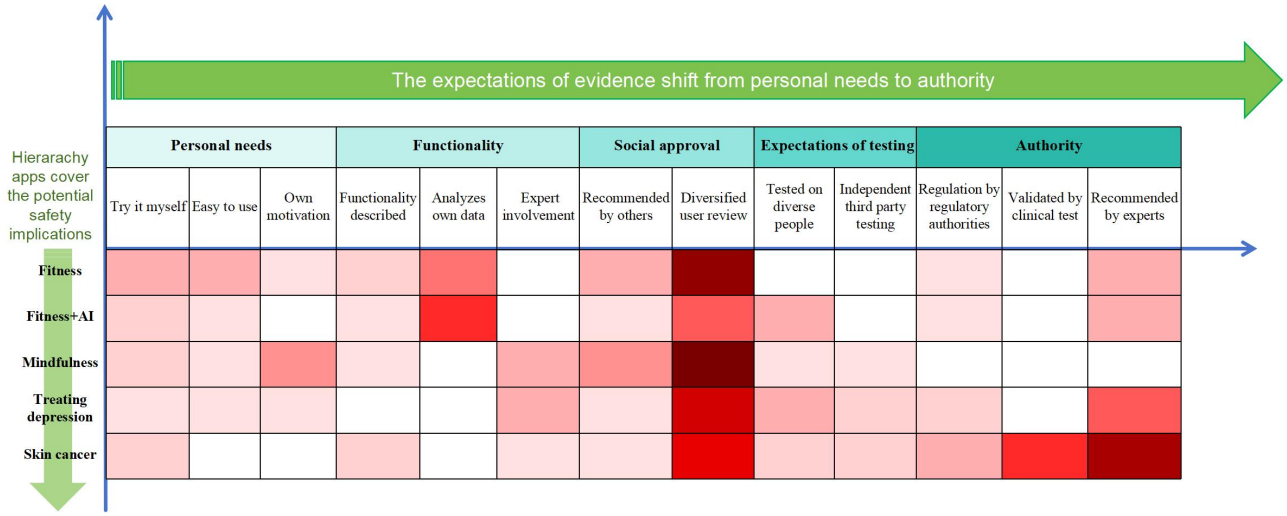
Overview

While we had sought to discuss evidence generation from a regulatory or research perspective, the participants often took a different perspective, discussing what factors influence their choice and use of digital health products and services. The 13 factors were eventually selected and grouped into 5 themes—personal needs, app functionality, social approval, testing expectations, and authority. Figure 1 shows the hierarchy of the 5 themes, from individual to community to authority. The color represents how often each factor was mentioned by the participants in the app scenario with different risk coefficients. The darker the color, the higher the frequency. White means not mentioned.

Across our hierarchy of 5 different digital health products, participants' views changed. As potential health risks increased, there was an increased call for greater testing, and the involvement of authorities and health care professionals. For example, with the skin cancer diagnosis apps, participants preferred digital health products to be recommended by professionals and to be checked with hospital diagnoses.

XSL•FO
RenderX

Figure 1. The expectations of evidence from different scenarios. The heatmap shows the approximate strength and frequency of participants' comments with respect to each scenario, but should not be interpreted as a quantitative analysis. This heatmap shows public expectations for evidence across 5 digital health app types, progressing from low to high risk (fitness to skin cancer diagnosis). Themes—personal needs, functionality, social approval, testing expectations, and authority—are shaded to indicate how frequently each was mentioned. Darker shades represent higher frequency, with expectations shifting from personal experience for low-risk apps to clinical validation for high-risk apps. AI: artificial intelligence.



Personal Needs

Most participants first considered trying a digital health product themselves to see if it works for them. They generally reported believing they know their own physical and mental conditions best. Respondents explained that everyone is different, and what works for others may not work for them, so using an app yourself and checking effectiveness was reported to be one of the most reliable ways of evaluating it. Similarly, some participants thought they would only recommend the product to others after using it themselves successfully.

For example, with the fitness app, some participants said they would see the app as effective if they see positive changes in their body after using it:

> *If I were using the fitness app, I would want probably milestones probably after a week after two weeks to three, I should get changes and stuff like that.* [P18]

With the fitness app with artificial intelligence (AI), participants discussed whether the AI's recommendations could be trusted. As P11 said,

> *If I've used it a few times and it's working for me. Then you begin to build up trust, or this is what it can do, and this is what it has trouble doing.* [P11]

With the mindfulness app, as this is related to their daily mental state, participants felt more need to try it for themselves:

> *Like people every day, stress is different like stress is different for everybody, like what might be stressful for me might not be stressful for another individual.* [P18]

Similarly, with the depression treatment app:

> *I think I'll just go on and try it out on my own.* [P21]

Factors, such as ease of use and personal motivation, were highlighted, with many participants expressing that these aspects greatly influence their willingness to adopt and continue using digital health apps. Ease of use seemed to be a fundamental prerequisite for participants.

When talking about personal motivation to use an app, P3 explained,

> *There are many fitness apps. You gotta find the one you really like.* [P3]

When participants felt an app was what they wanted, they would be motivated to continue using it. For mindfulness and depression apps, some participants said that if the app made them feel better, that would mean the app was effective. They would then increase their motivation and become more committed to using the app:

> *I know it's different for everyone and depending, like the other person said, depending on how invested you are in using the app, that's how much of the result you see.* [P14]

Functionality

Participants saw a product's functionality as a determining factor of whether it is effective or not. Before using the fitness or fitness apps with AI, some participants said they would check online information about the product to understand its functions and effects. With the higher risk products, such as the skin cancer diagnosis app, some participants felt that developers demonstrating to the public how the app works could be evidence of the app's effectiveness, such as having videos or explaining within the app how it works. A product's diversity in functions was also seen to be related to its ability to work for diverse populations. For fitness apps, some participants talked about different levels of exercise being supported to meet the needs of different users, and how apps should also have the ability to collect physiological data and analyze data:

> *The application needs to identify the users you know to get data and then be able to process the data and give you a well-informed recommendation.* [P12]

XSL•FO
RenderX

For the other apps, participants wanted to have experts involved in the apps, and for the apps' features to allow users to connect with experts, which would increase their trust in using the app. For example, they wanted the depression treatment app to entail further follow-up by experts according to the degree of depression. Talking about the skin cancer diagnosis app, P19 explained:

> *It will be my only safety because it depends on the professionals attached, whatever the app wants to support.* [P19]

Social Approval

Most participants mentioned social approval as a strong influence on their belief in the effectiveness of an app. They described recommendations from friends and relatives as being critical when using an app. This is, in part, because they have a high degree of familiarity with relatives and friends, know their lifestyles and physical conditions, and this therefore allows them to put relatives' and friends' recommendations in context. As a result, across all the apps, participants trusted the recommendations of relatives and friends more:

> *For instance, if I'm into a particular type of job and the person is also into that type. I could read what everyone's stress goes through together with that, and I could greatly do the same. I could use that as evidence, and I trust such an app*
>
> *[P18]*

After recommendations from friends and family, user reviews were the main factor influencing participants to choose an app. For apps with varying health risks, the user reviews that people wanted to see differed. Participants said they would look at overall user reviews before downloading the fitness app to check its effectiveness. This would make it easier for them to understand what the app does and whether it is right for them. Participants believed that the more people who downloaded and positively reviewed an app, the more likely it was to be effective. Therefore, participants also mentioned the number of downloads as a source of evidence. While discussing the fitness app with AI, 1 participant questioned the authenticity of reviews. Some participants tended to ask questions in the app user's comments section and then use the subsequent interaction and answers from those users to help them decide whether to choose an app:

> *I need evidence cos it's already acting in place of my doctor and already giving me recommendations… I tend to stop and pause for reactions from people and get reviews.* [P17]

Discussing the mindfulness app, some participants felt they needed feedback from people close to them, such as observing how their friends used it to see how the app worked. As for the app to treat depression, 1 participant mentioned that the stories people share on social media are also a way for users to comment. Since they are real user experiences, this kind of comment has impact because

> *we buy because of what people have said, how it's done.* [P11]

Participants found the comments of those who suffer from depression more persuasive, as P9 explained:

> *Seeing that evidence from the person who used this app, the person was suffering from depression, and this kind of treatment covered. So I'll need such things to use this kind of app.* [P9]

Participants wanted to know if different people had used the app and it had worked for them. Different people could refer to differences in ethnicity, sex, age, and lifestyle. Some participants wanted to see physical evidence of results from fitness apps, such as outcomes like weight loss and faster running. We pushed participants to specify the number of people from whom they would like to have some feedback. Many participants considered the experience of 1‐5 people enough to demonstrate effectiveness.

For the AI function of the fitness app, participants were concerned about whether the AI could make accurate recommendations based on the age and gender of different people. Mindfulness apps, they said, need to have been used and tested for effectiveness by various different people, and the number of participants who brought this up was larger than with the other apps. Some participants believed that different lifestyles lead to varying stress levels:

> *I think they have different needs, especially for mindfulness apps. The option to choose a lifestyle would be helpful.*
>
> *[P3]*

Here, participants talked of how 5 or more people's reviews could be used as evidence, with feedback wanted from people with differing levels of depression:

> *If the treatment can cater for different people depending on their diagnosis… other people have used them for treatments, and it's a success for them. I think I'd be OK to use it.* [P14]

Expectations of Testing

Some participants mentioned that developers should conduct tests on different people before launching digital health products and services. What participants meant by different people again refered to differences in various aspects, such as different ethnicities, ages, genders, lifestyles, and health conditions. Depending on the app, participants felt that the focus of the test should vary. P1 believed that the fitness app with AI should also implement

> *Background checks and testing. For a different age group and probably put in some information.* [P1]

P20 thought mindfulness apps "*should cater to various individuals.*" The participants did not evince any familiarity with the details of scientific evaluations.

Being diagnosed with cancer was recognized as having a significant impact on people's lives. Thus, some participants mentioned that the testing for a skin cancer diagnosis app should be mainly to determine the app's accuracy and to improve the richness of the test data. In terms of testing being on a diverse group, for skin cancer diagnosis apps, participants were

XSL•FO
RenderX

concerned about differences in skin color and wanted to see more usage data than for other apps. Furthermore, 1 participant suggested testing on 20 people, 5 of whom had the same skin color as himself.

For apps to treat depression, it was suggested that testing should be conducted strictly following academic experiments to test actual effectiveness:

> *I'm expecting that this app has been used for more, more like a randomised set of people before it was rolled out, and then, you know, a study should be carried out on them before and after the intervention. Then you compare how well, how effective they there's been. So, for me, I think that would be the evidence that we're looking for.* [P15]

Some participants said that the testing of digital health products should be carried out by third-party institutions. This should, first, be an authority in a relevant professional field, such as the National Health Service (NHS) or other professional bodies. Second, it should be independent and unrelated to the developer's business operations. In addition, 1 participant noted for depression apps that such third-party tests will be more trustworthy than the developers' own explanations. Another participant said that skin cancer diagnosis app testing should be used in a real-life setting:

> *I also think the app should have been used to diagnose real-life diseases in some trials before the agency certified it for launch.* [P15]

Information on testing should be made available, for example, within the app. Discussing the skin cancer diagnosis app, a participant said,

> *I think the app should contain information about the testing and trials done to ascertain this app's effectiveness.* [P15]

Authority

Participants mentioned the need for government and NHS regulation for all the apps. Participants believed that government and NHS regulation would increase their trust in the apps' effectiveness. A digital health product vetted by the government or NHS was taken to mean that it has been tried by many people and has worked. Government regulation was seen as most important for skin cancer diagnosis apps because of the *"severity of the condition"* [P20] and the need to avoid misdiagnosis.

Participants felt it was necessary that the skin cancer app's diagnoses were compared with hospital diagnoses to determine the app's accuracy. It was the only app where participants talked about a comparison with a hospital diagnosis during the discussions. To avoid misdiagnosis, participants would not unilaterally believe the app diagnosis:

> *I also think any form of diagnosis that should be done with this app should also be related to the GP and others for clarification to get a better result.* [P13]

The use of the app, they said, should only be an intermediate process, and the effectiveness of the app would ultimately be proved by the hospital diagnosis.

Participants rated professional recommendations as being necessary for the other 4 apps. For fitness apps, recommending professionals referred to experts in this field, such as fitness coaches and sports athletes. For the skin cancer diagnosis app, most participants said only a doctor's approval would lead them to use it. One participant stated,

> *I would also be looking for certification from a cancer body.* [P15]

For depression treatment apps, people would consult their doctors for recommendations:

> *I would actually contact my doctor first and let him know about the app and if he was about to use it and can give me go ahead to use it.* [P1]

Most participants believed that the government or NHS should regulate all digital health products:

> *These apps must be censored and monitored because they involve our well-being. This application acts as a substitute for a visit to our GP. So, we need to be sure of how we administer the application and its reliability. About all the apps.* [P17]

Discussion

Principal Results

This study aimed to explore participants' trust in digital health technology and what evidence the public wants from digital health interventions for increasing health and well-being. Following the NICE ESF's risk hierarchy, different scenarios were designed, including wellness apps and treatment and diagnostic apps. We found that (1) the public's expectations for evidence of effectiveness are intertwined with the factors that influence their choice and use of digital health interventions; (2) the public's expectations for evidence of effectiveness are based more on individual experiences and social approval than scientific evidence; (3) as the perceived risks associated with different apps increase, so to do the public's expectations for the evidence of their effectiveness; and (4) there is a desire for recommendations or certifications from relevant bodies, particularly as the potential risk increases.

Comparison With Previous Work

This study supports the findings that the public understanding of what constitutes high-quality research evidence is low [18-20]. As shown previously [21,22], this can lead to challenges in research support and affect the decision-making related to the digital health selection, uptake and engagement, behavior and, ultimately, health outcomes.

We also saw people exhibiting cautious optimism, recognizing potential benefits while maintaining a healthy scepticism about new technologies, as seen in previous studies [25,26].

This study found that the public prefers to verify the effectiveness of digital interventions through personal experience. There are parallels with some research on earlier generations of digital health technology. Herian and colleagues [33] found that greater experience with health care was associated with lower trust in government to regulate electronic

XSL•FO
RenderX

health records, which suggests that personal experience was valued more. Several studies showed that people seeking health information online relied more on their own experience of websites than credibility assessments [34]. We know more generally that personal choice and personal recommendations are important in people's engagement with digital health interventions [35].

This focus on personal experience supports the view of Pagoto and Bennett [36] that interface design is crucial in the development process of the app. The "person-based" approach emphasizes using qualitative research to understand the psychosocial background of a wide range of target populations during the development process of digital interventions and getting feedback from them to improve applicability and effectiveness [37]. This is consistent with the findings of this study that the public expects greater transparency in the development and testing process, which is beneficial for improving the social acceptance of digital interventions by testing with different people and validating the results.

Implications

The study found that the public's concept of evidence diverges significantly from the scientific community's understanding. While researchers may refer to evidence as data collected from rigorous, controlled settings, such as randomized controlled trials, participants in our focus groups considered personal experiences and anecdotal successes as credible evidence for assessing digital health products.

Public policy [15,16] and research [17] both strongly endorse involving the public. This study calls for a systematic approach to involving the public in digital health evidence generation. To effectively engage the public, it is essential to grasp their level of understanding about the need for evidence generation in digital health and why it matters, that is, the potential for safety issues but also missed opportunities for prevention (for so-called lower risk digital health, such as apps for physical activity). This requires educational methods with a substantial instructive component as well as coproduction, such as citizen juries, and necessitates clear, accessible explanations of evidence generation, effectiveness, and the potential detrimental effects of digital health interventions that are widely used but not evaluated. Educating the public about the need for evidence generation in digital health could increase the pressure and motivation of developers to embed evidence generation practices in digital health. Equally, the pressure of the public may encourage regulators to require minimal evidence generation before publicly available digital health is introduced on the app distribution platforms.

Professional support and education can foster user engagement with digital health technology [38]. Social media platforms present opportunities for public communication and behavioral interventions [39]. Developers and relevant authorities can use diverse strategies, including leveraging social media, to enhance the efficacy of digital health apps for promoting public health behaviors. A detailed description of an app's functionalities, effects, and target population should be provided so that individuals can assess its effectiveness quickly. Using documentaries and press releases can promote transparency in digital health product development and testing processes, enhancing public trust. Using social media platforms to share user reviews, such as inviting long-term clients from diverse backgrounds and age groups to tell their stories or engaging professionals like renowned doctors, experts, and athletes to provide insights, fosters societal acceptance. It is crucial to increase public understanding of the impact of the lack of evidence generation in digital health, such as safety concerns and missed opportunities for health promotion when popularity rather than evidence for effectiveness is the primary factor driving downloads.

Strengths

To the best of our knowledge, this is the first study to explore public expectations and preferences regarding evidence generation for the effectiveness of digital health interventions—a stakeholder group that has been overlooked. The sampling strategy aimed to include those at risk of digital inequalities and included a range of ages, genders, socioeconomic statuses, and ethnicities, showcasing a broad spectrum of views. The various scenarios developed for this study were comprehensive.

Limitations

While the study was open to all UK residents, we recruited mostly from London. With 21 participants, the study may not capture the full range of public opinions. The materials used to guide the focus groups and the results could be used to facilitate the development of a survey to get larger samples, which could provide more robust data. We offered a £20 (US $25.37) voucher, which might have influenced motivations to participate in this study.

Conclusion

This study has explored public expectations regarding the evidence needed to demonstrate the effectiveness of digital health products and services. The public tends to favor evidence that is relatable and directly experienced, such as personal success stories, user reviews, and recommendations from significant others. The public made minimal reference to what researchers would perceive as rigorous evaluation methods. However, there was also support for authoritative recommendations and the independent testing of products. As health risks associated with digital health increased, there was a growing demand for more substantial evidence and authoritative recommendations from health care professionals.

Acknowledgments

Thanks to all the participants who took part in the research. ChatGPT was used to assist with grammar and style refinement.

XSL•FO
RenderX

Data Availability

The datasets generated during and analyzed during this study (consisting of qualitative focus group transcripts) are not publicly available due to the sensitive nature of the data and confidentiality commitments to participants. However, the deidentified data are available for research purposes from the corresponding author on reasonable request.

Conflicts of Interest

PB provides consultancy in digital health development and evaluation for ShiftPartner. HWWP provides or has provided consultancy on digital health evaluation for Flo Health Inc., Prova Health, and Thrive Therapeutic Software Ltd. He has PhD students in the field employed by or previously employed by, and with fees paid by or previously paid by AstraZeneca, Patients Know Best, and BetterPoints Ltd.

Multimedia Appendix 1
Informed consent.
[DOCX File, 42 KB - formative_v9i1e56523_app1.docx]

References

1. Istepanian RSH. Mobile health (m-Health) in retrospect: the known unknowns. Int J Environ Res Public Health 2022 Mar 22;19(7):3747. [doi: 10.3390/ijerph19073747] [Medline: 35409431]
2. Egermark M, Blasiak A, Remus A, Sapanel Y, Ho D. Overcoming pilotitis in digital medicine at the intersection of data, clinical evidence, and adoption. Adv Intell Syst 2022 Sep;4(9):2200056. [doi: 10.1002/aisy.202200056]
3. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018;1(1):12. [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
4. Gajarawala SN, Pelkowski JN. Telehealth benefits and barriers. J Nurse Pract 2021 Feb;17(2):218-221. [doi: 10.1016/j.nurpra.2020.09.013] [Medline: 33106751]
5. McKee M, van Schalkwyk MCI, Stuckler D. The second information revolution: digitalization brings opportunities and concerns for public health. Eur J Public Health 2019 Oct 1;29(Supplement_3):3-6. [doi: 10.1093/eurpub/ckz160] [Medline: 31738440]
6. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. Lancet 2020 May 16;395(10236):1579-1586. [doi: 10.1016/S0140-6736(20)30226-9] [Medline: 32416782]
7. Freeman K, Dinnes J, Chuchu N, et al. Algorithm based smartphone apps to assess risk of skin cancer in adults: systematic review of diagnostic accuracy studies. BMJ 2020 Feb 10;368:m127. [doi: 10.1136/bmj.m127] [Medline: 32041693]
8. Crane D, Garnett C, Brown J, West R, Michie S. Behavior change techniques in popular alcohol reduction apps: content analysis. J Med Internet Res 2015 May 14;17(5):e4060. [doi: 10.2196/jmir.4060] [Medline: 25977135]
9. Fleming GA, Petrie JR, Bergenstal RM, Holl RW, Peters AL, Heinemann L. Diabetes digital app technology: benefits, challenges, and recommendations. A consensus report by the European Association for the Study of Diabetes (EASD) and the American Diabetes Association (ADA) Diabetes Technology Working Group. Diabetes Care 2020 Jan;43(1):250-260. [doi: 10.2337/dci19-0062] [Medline: 31806649]
10. Wicks P, Chiauzzi E. 'Trust but verify'--five approaches to ensure safe medical apps. BMC Med 2015 Sep 25;13(1):205. [doi: 10.1186/s12916-015-0451-z] [Medline: 26404791]
11. Evidence standards framework for digital health technologies. National Institute for Health and Care Excellence. 2018. URL: https://www.nice.org.uk/corporate/ecd7 [accessed 2025-1-3]
12. Dang A, Arora D, Rane P. Role of digital therapeutics and the changing future of healthcare. J Fam Med Prim Care 2020 May;9(5):2207-2213. [doi: 10.4103/jfmpc.jfmpc_105_20] [Medline: 32754475]
13. Evaluating digital health products. GOVUK. 2020. URL: https://www.gov.uk/government/collections/evaluating-digital-health-products [accessed 2025-01-03]
14. The Fast-Track Process for Digital Health Applications (DiGA) according to Section 139e SGB V. A Guide for Manufacturers, Service Providers and Users. Federal Institute for Drugs and Medical Devices. 2020. URL: https://www.bfarm.de/SharedDocs/Downloads/EN/MedicalDevices/DiGA_Guide.pdf [accessed 2024-02-12]
15. Accelerated Access Collaborative patient and public involvement strategy. NHS England. 2022. URL: https://www.england.nhs.uk/aac/publication/accelerated-access-collaborative-patient-and-public-involvement-strategy/ [accessed 2024-12-04]
16. Patient and public involvement policy. National Institute for Health and Care Excellence. 2013. URL: https://www.nice.org.uk/about/nice-communities/nice-and-the-public/public-involvement/public-involvement-programme/patient-public-involvement-policy [accessed 2025-01-03]
17. Gradinger F, Britten N, Wyatt K, et al. Values associated with public involvement in health and social care research: A narrative review. Health Expect 2015 Oct;18(5):661-675. [doi: 10.1111/hex.12158] [Medline: 24325553]

XSL•FO
RenderX

18. Report on the participation decision-making process. Center for Information and Study on Clinical Research Participation (CISCRP). 2017. URL: https://www.ciscrp.org/wp-content/uploads/2019/06/2017-CISCRP-Perceptions-and-Insights-Study-Decision-Making-Process.pdf [accessed 2025-01-03]
19. Stead M, Eadie D, Gordon D, Angus K. "Hello, hello--it's English I speak!": A qualitative exploration of patients' understanding of the science of clinical trials. J Med Ethics 2005 Nov;31(11):664-669. [doi: 10.1136/jme.2004.011064] [Medline: 16269566]
20. Ganz PA. Clinical trials. Concerns of the patient and the public. Cancer 1990 May 15;65(10 Suppl):2394-2399. [doi: 10.1002/1097-0142(19900515)65:10+<2394::aid-cncr2820651509>3.0.co;2-#] [Medline: 2185874]
21. Enhancing the use of scientific evidence to judge the potential benefits and harms of medicines. The Academy of Medical Sciences. 2017. URL: https://acmedsci.ac.uk/file-download/44970096 [accessed 2025-1-3]
22. Finucane E, O'Brien A, Treweek S, et al. The People's Trial: supporting the public's understanding of randomised trials. Trials 2022 Mar 9;23(1):205. [doi: 10.1186/s13063-021-05984-1] [Medline: 35264220]
23. Lupton D. Beyond techno-utopia: critical approaches to digital health technologies. Societies (Basel) 2014;4(4):706-711. [doi: 10.3390/soc4040706]
24. Strange M. Three different types of AI hype in healthcare. AI Ethics 2024 Aug;4(3):833-840. [doi: 10.1007/s43681-024-00465-y]
25. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. Lancet Digit Health 2021 Sep;3(9):e599-e611. [doi: 10.1016/S2589-7500(21)00132-1] [Medline: 34446266]
26. Sauchelli S, Pickles T, Voinescu A, et al. Public attitudes towards the use of novel technologies in their future healthcare: a UK survey. BMC Med Inform Decis Mak 2023 Feb 22;23(1):38. [doi: 10.1186/s12911-023-02118-2] [Medline: 36814262]
27. Lee L, Salami RK, Martin H, et al. "How I would like AI used for my imaging": children and young persons' perspectives. Eur Radiol 2024 Dec;34(12):7751-7764. [doi: 10.1007/s00330-024-10839-9] [Medline: 38900281]
28. Cowey AE, Potts HWW. What can we learn from second generation digital natives? A qualitative study of undergraduates' views of digital health at one London university. Dig HEALTH 2018;4:2055207618788156. [doi: 10.1177/2055207618788156] [Medline: 30046453]
29. Hernandez-Ramos R, Aguilera A, Garcia F, et al. Conducting internet-based visits for onboarding populations with limited digital literacy to an mHealth intervention: development of a patient-centered approach. JMIR Form Res 2021 Apr 29;5(4):e25299. [doi: 10.2196/25299] [Medline: 33872184]
30. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
31. Braun V, Clarke V. One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qual Res Psychol 2021 Jul 3;18(3):328-352. [doi: 10.1080/14780887.2020.1769238]
32. Braun V, Clarke V. Supporting best practice in reflexive thematic analysis reporting in palliative medicine: a review of published research and introduction to the Reflexive Thematic Analysis Reporting Guidelines (RTARG). Palliat Med 2024 Jun;38(6):608-616. [doi: 10.1177/02692163241234800] [Medline: 38469804]
33. Herian MN, Shank NC, Abdel-Monem TL. Trust in government and support for governmental regulation: the case of electronic health records. Health Expect 2014 Dec;17(6):784-794. [doi: 10.1111/j.1369-7625.2012.00803.x] [Medline: 22809222]
34. Eysenbach G. Credibility of Health Information and Digital Media: New Perspectives and Implications for Youth: MacArthur Foundation Digital Media and Learning Initiative; 2008:123-154. [doi: 10.1162/dmal.9780262562324.123]
35. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120. [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
36. Pagoto S, Bennett GG. How behavioral science can advance digital health. Transl Behav Med 2013 Sep;3(3):271-276. [doi: 10.1007/s13142-013-0234-z] [Medline: 24073178]
37. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30. [doi: 10.2196/jmir.4055] [Medline: 25639757]
38. Curtis K, Price K. Factors that influence older people's engagement with digital health technology. Nurs Older People 2017 Nov 30;29(10):27-30. [doi: 10.7748/nop.2017.e986] [Medline: 29188928]
39. Pagoto S, Waring ME, May CN, et al. Adapting behavioral interventions for social media delivery. J Med Internet Res 2016 Jan 29;18(1):e24. [doi: 10.2196/jmir.5086] [Medline: 26825969]

Abbreviations

AI: artificial intelligence
DiGA: Digitale Gesundheitsanwendungen
NHS: National Health Service

XSL•FO
RenderX

NICE ESF: National Institute for Health and Care Excellence Evidence Standards Framework

Edited by A Mavragani; submitted 21.03.24; peer-reviewed by B Cliffe; revised version received 28.10.24; accepted 30.10.24; published 20.01.25.

Please cite as:
Bondaronek P, Li J, Potts HWW
Public Understanding and Expectations of Digital Health Evidence Generation: Focus Group Study
JMIR Form Res 2025;9:e56523
URL: https://formative.jmir.org/2025/1/e56523
doi:10.2196/56523

© Paulina Bondaronek, Jingfeng Li, Henry W W Potts. Originally published in JMIR Formative Research (https://formative.jmir.org), 20.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

International Expert Consensus on Relevant Health and Functioning Concepts to Assess in Users of Tobacco and Nicotine Products: Delphi Study

Vivienne Law[1*], BSc; Esther F Afolalu[2*], PhD; Linda Abetz-Webb[3], MA; Lee Andrew Wemyss[4], BSc; Andrew Turner[4], BSc; Christelle Chrea[2], PhD

[1]Zebra26 Ltd., Devon, United Kingdom

[2]PMI R&D, Philip Morris Product S.A., Neuchâtel, Switzerland

[3]Patient-Centered Outcomes Assessments Ltd., Cheshire, United Kingdom

[4]DaSH Global, Cheshire, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Esther F Afolalu, PhD
PMI R&D, Philip Morris Product S.A.
Quai Jeanrenaud 5
Neuchâtel, 2000
Switzerland
Phone: 41 58 242 2218
Email: esther.afolalu@pmi.com

Abstract

Background: A Delphi study was conducted to reach a consensus among international clinical and health care experts on the most important health and functioning self-reported concepts when evaluating a switch from smoking cigarettes to using smoke-free tobacco and/or nicotine products (sf-TNPs).

Objective: The aim of this research was to identify concepts considered important to measure when assessing the health and functioning status of users of tobacco and/or nicotine products.

Methods: Experts (n=105), including health care professionals, researchers, and policy makers, from 26 countries with professional experience and knowledge of sf-TNPs completed a 3-round, adapted Delphi panel. Online surveys combining quantitative (MaxDiff best-worst scaling and latent class analysis) and qualitative assessments were used to rank and achieve alignment on the importance of 69 health and functioning concepts. All experts participating in round I completed round II, and 101 (95%) completed round III.

Results: The round I analysis identified 36 (52%) out of 69 concepts that were refined for the round II assessment. The highest-ranked concepts reflected health-related impacts, while the lowest-ranked ranked concepts were related to aesthetics and social impacts. Round II ranking reinforced the importance of concepts relating to health impacts, and the analysis resulted in 20 concepts retained for round III assessment. In round III, the 4 highest-ranked concepts were cardiovascular symptoms, shortness of breath, chest pain, and worry about smoking-related diseases and impact on general health, and they made up 50% of the total score in the MaxDiff analysis. Experts reported likelihood of seeing measurable levels of change in the final 20 concepts with a switch to an sf-TNP. The majority of experts felt it was “likely” or “extremely likely” to observe changes in concepts such as gum problems (74/101, 73%), phlegm or mucus while coughing or not coughing (72/101, 71%), general perception of well-being (72/101, 71%), and throat irritation or sore throat (72/101, 71%). Latent class analysis revealed subgroups of experts with different perceptions of the relative importance of the concepts, which varied depending on professional specialty and geographic region. For example, 74% (14/19) of oncologists aligned with the subgroup prioritizing physical health symptoms, while 71% (12/17) of experts from Asia aligned with the subgroup considering both physical health and psychosocial aspects.

Conclusions: This study identified key concepts to be considered in the development of a new measurement instrument to assess the self-reported health and functioning status of individuals using sf-TNPs. The findings contribute to the scientific evidence base for understanding and evaluating both the individual and public health impacts of sf-TNPs.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e58614) doi:10.2196/58614

KEYWORDS

Delphi study; expert consensus; outcome measures; health and functioning; tobacco and/or nicotine products

Introduction

Reducing exposure to harmful and potentially harmful constituents in cigarette smoke by cessation of smoking or switching to reduced risk tobacco and/or nicotine products (TNPs) is a major public health focus worldwide [1]. Although nicotine-replacement therapy (NRT) and targeted counseling services can help individuals achieve their quitting goals, there can be hurdles to becoming and/or staying smoking free [2].

Smoke-free tobacco and/or nicotine products (sf-TNPs) refer to TNPs that do not undergo combustion and therefore do not produce the harmful smoke generated by cigarettes, which has been established as the main risk factor of tobacco-related diseases [3]. These sf-TNPs can include heated tobacco products, e-cigarettes (vapes), other e-vapor products (e-pipes, e-cigars, and so on), nicotine pouches, and smokeless tobacco products (snuff, snus, and chewing tobacco) [3]. Regulatory bodies, like the US Food and Drug Administration (FDA), require robust data to evaluate the health impacts of sf-TNPs and guide regulatory decisions on product authorization. For instance, the FDA's modified risk tobacco product pathway mandates comprehensive scientific evidence to support claims that a tobacco product reduces harm or the risk of tobacco-related disease in individual tobacco users and benefits the health of the population as a whole [4]. Some sf-TNPs have been recognized as modified-risk tobacco products under this pathway [4,5] and may help reduce the negative health effects associated with commercially marketed combustible tobacco products such as cigarettes [6-8].

The adverse effects of smoking on health have been well documented, whereas stopping smoking can improve health and reduce the risk of disease [6,9,10]. Evidence also suggests that switching away from cigarettes to using sf-TNPs can help reduce cigarette consumption and may lead to cessation in some cases [8]. While studies have shown that sf-TNPs reduce exposure to many of the harmful and potentially harmful constituents found in cigarette smoke, they may still pose health risks. For instance, e-cigarettes, heated tobacco products, and smokeless tobacco have been associated with some cardiovascular, respiratory, and oral health risks, albeit to a lesser extent than combustible cigarettes [11-14]. The long-term health effects of sf-TNPs are not yet fully understood, and further research is necessary to comprehensively assess their safety profile. In addition, little is known about the self-reported health impact of switching from cigarettes to sf-TNPs [6]. In this context, measuring the self-reported experience of health and functioning (including health status, functional status, and other health-related quality of life constructs) is crucial to understanding the impact of tobacco harm reduction strategies [15].

Generic health status measures, such as the Short Form Survey–12 [16] and Short Form Survey–36 instruments [17], have shown that those who smoke tend to report lower health status compared with those who never smoked, although the impact of cessation on health status seems to be more complex. Some smoking-specific measures have been developed but not yet widely standardized; therefore, they may not be directly applicable to sf-TNPs [9,18,19]. Crucially, current measures often lack the necessary sensitivity to detect longitudinal health changes in healthy populations due to high ceiling effects at baseline [20], hindering the assessment of changes in health and functioning following a switch from smoking cigarettes to using sf-TNPs.

To help address these challenges, the development of a new self-reported measure (ABOUT – Health and Functioning) was undertaken to assess the health and functioning status of individuals using sf-TNPs. This new measure is part of the portfolio of the ABOUT Toolbox [21], consisting of self-report measures to assess perceptions and behavioral outcomes related to the use of sf-TNPs. The initial preparatory phase of the instrument development was based on several research activities (systematic literature review, reanalysis of qualitative data, and expert insights [22]) and resulted in the identification of 69 health and functioning concepts relevant to TNP use. The qualitative research phase that followed consisted of (1) concept elicitation interviews of users to understand their perceptions of health and functioning after switching to sf-TNPs [23,24] and (2) the Delphi panel study reported in this paper. The aim of the Delphi study was to identify the health and functioning concepts considered most important by clinical and health care experts (ie, health care professionals, researchers, policy makers, and those involved in smoking cessation or tobacco harm reduction) when assessing the health and functioning status of individuals who stop smoking cigarettes or switch to using sf-TNPs or NRTs. Delphi panel methodology is a well-established process for determining consensus among relevant groups of individuals to aid issue prioritization, ranking, development, and obtaining agreement on guidelines, concept-framework development, and development of outcome measures [25-27], including issues relevant to smoking-related behaviors and sf-TNP use [25,28,29]. Delphi panels have previously been used to forecast trends and changes over time in health-related matters [30-32].

Methods

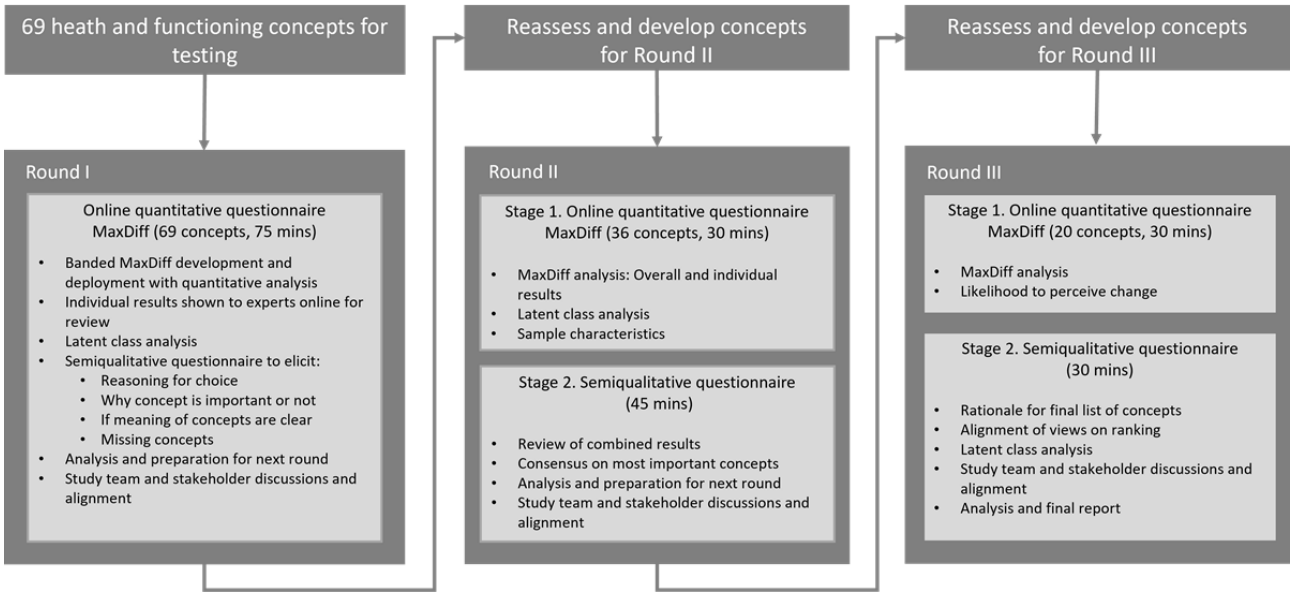
Objectives and Summary of Approach

This Delphi panel was organized for experts to select and rank health and functioning concepts considered important to assess when individuals stop smoking combustible TNPs (eg, cigarettes) or switch to using sf-TNPs (eg, e-cigarettes [vapes], heated tobacco, or smokeless tobacco products) or NRTs. The study was conducted and reported based on available guidance for conducting and reporting Delphi studies in health care research [33].

XSL•FO
RenderX

A preliminary set of 69 concepts was identified from earlier research activities including a scoping literature review, results from the reanalysis of previous TNP consumer qualitative research results, and expert opinion from a small group of key opinion leaders [22]. This list was presented for evaluation in this Delphi panel according to the process described in Figure 1. An adapted approach was used for the Delphi panel, whereby the results of round I were reviewed and refined in light of the ongoing qualitative concept elicitation interview findings from consumers of TNPs and ongoing review by key opinion leaders (Supplementary Information and Table S1 in Multimedia Appendix 1). The concept list to be evaluated in round II was adjusted accordingly and refined considering the round I findings. Round III was a detailed review of the final set of health and functioning concepts. This step also included an evaluation of likelihood of change that experts considered important for each concept for inclusion in a self-reported measure of clinical relevance.

Figure 1. Overview of the design and structure of the adapted Delphi panel study.



Participants and Recruitment

For this adapted Delphi panel, we recruited experts who routinely treat, communicate with, or advise individuals who smoke or wish to stop smoking or switch to sf-TNPs. The recruitment aim was to ensure that a minimum of 100 experts globally completed all 3 rounds of the Delphi panel. To anticipate attrition between rounds, the initial recruitment target was set at 120 experts.

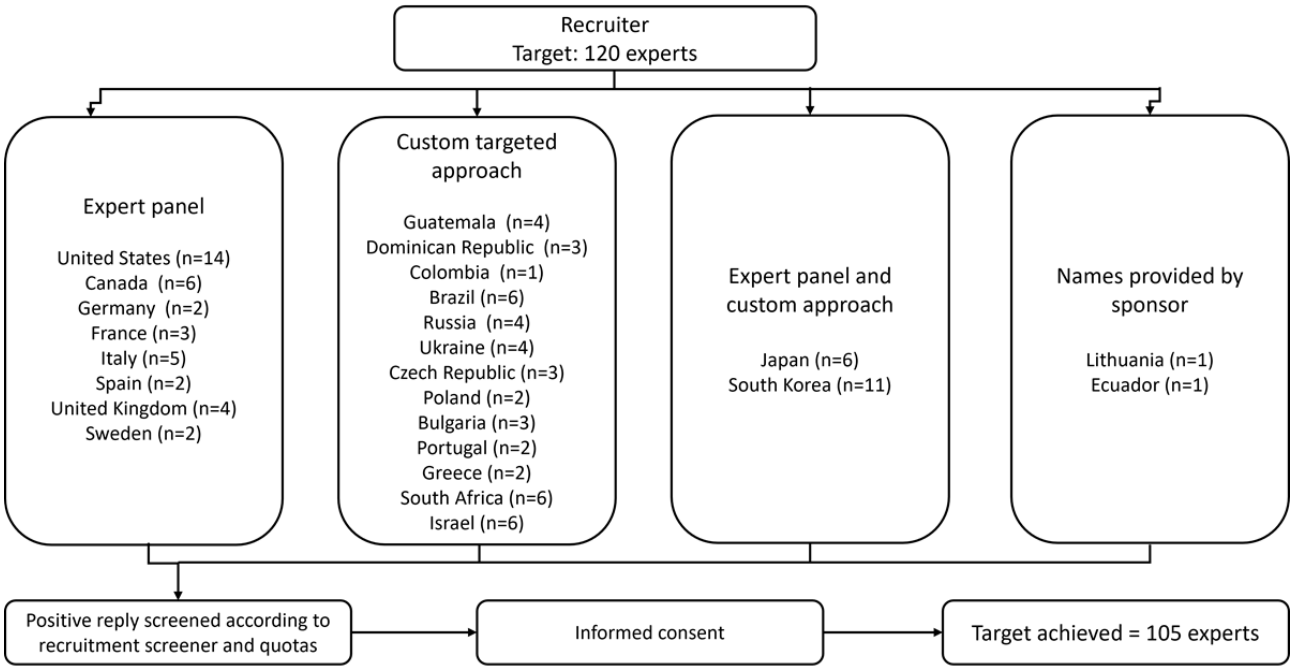
Recruitment was undertaken by QualWorld, who identified experts through their survey panels, existing contacts, publications, word of mouth, and canvassing through appropriate organizations in individual countries. The process and route of contact to identify potential respondents are outlined in Figure 2.

To anticipate and account for potential sensitivities among health care practitioners about undertaking a project sponsored by a tobacco company, the recruitment company sent out an initial survey asking which industries experts would be interested in working on health-related topics. Tobacco companies were listed among other industries such as mining, pharmaceuticals, and oil and gas. Only experts stating unprompted that they would consider participating in studies by tobacco companies were chosen to receive further correspondence linked to the Delphi panel. Subsequent recruitment steps proceeded on an individual basis according to predefined inclusion and exclusion criteria (Textbox 1).

To ensure consistency in responses across different regions, the Delphi panel was conducted in English where possible and in local languages for experts in Japan, South Korea, Russia, Ukraine, Czech Republic, and Bulgaria. A professional translation service translated the survey materials and responses, and back-translation techniques were used to verify accuracy. The local bilingual recruitment team members also performed proofreading and quality checks from a language and contextual perspective to ensure the translated content accurately reflected the original meaning. For each step of the Delphi panel, the English and local language surveys were programmed and uploaded for a user acceptability test by the project team before launch.

XSL•FO
RenderX

Figure 2. Identification and recruitment of experts through specific channels for the adapted, 3-round Delphi panel.



Textbox 1. Inclusion and exclusion criteria for the participants recruited for the Delphi panel.

Inclusion criteria

- Had current clinical, academic, or professional experience in the area of smoking-related diseases, smoking cessation, nicotine addiction, or other smoking-related conditions, as well as health policy or advocacy work related to tobacco control or tobacco harm reduction
- Currently working in research, prescribing, or a recommendation capacity
- Knowledge of smoke-free tobacco and/or nicotine products or nicotine replacement therapies, works with patients to some degree, and/or recommends smoke-free tobacco and/or nicotine products or nicotine replacement therapies
- Expertise in one of the following groups:
  - Specialist physicians: smoking-related oncology; smoking-related respiratory disease, and smoking-related cardiovascular disease (specifically treatment of patients)
  - General practitioners and internal medicine physicians: those who see patients with a smoking-related condition or disease and recommend smoking cessation advice
  - Dentists or oral hygienists: those who see patients with smoking-related dental issues
  - Smoking cessation, addiction, or dependence: those who help users quit smoking with less harmful products or nicotine replacement therapy or worked with nicotine addiction; nurses undertaking smoking cessation activity; researchers in nicotine addictions; psychologists, social workers, or counselors
  - Health policy, advocacy, and non-profit organizations: smoking-related health policy, including tobacco harm reduction; organizations or charities that encourage people to quit; health policy and advocacy related to tobacco control and tobacco harm reduction. This can include physicians who work alongside the government and local policy as well as nonprofit organizations
- Ability and willingness to participate across the whole study
- Ability to complete the research in the English language (except for these countries: South Korea, Japan, Ukraine, Czech Republic, Russia, and Bulgaria)
- At least 30% of experts to identify as female (although this was not a strict quota for recruitment)

Exclusion criteria

- More than 50% of experts should not have been a consultant with or worked for the tobacco industry
- Had less than 5 years of clinical, academic, or professional experience
- Did not fall into one of the categories for inclusion
- Not willing to complete all 3 rounds of research

XSL•FO
RenderX

Ethical Considerations

The New England Independent Review Board reviewed the study protocol and all study materials and any amendments and granted approval of the study (reference 1 - 9184 - 1).

Before participation, all participants were required to review and sign an informed consent form, which provided a detailed description of the study, outlined the procedures, and specified participant expectations. Participants had the opportunity to ask questions and receive answers before signing the consent form.

All study data were deidentified and participants were identified only by unique ID numbers throughout the research rounds.

All experts were compensated and received a cash honorarium according to local professional rates in their respective countries, ranging from £115 to £450 (approximately US $147 to US $576 at the time of the research), for their time after each round.

Procedure and Analytical Methods

MaxDiff Analysis

This study incorporated the use of MaxDiff—a type of best-worst (maximum difference) scaling analytical method [34-36]. This methodology was primarily chosen due to the large number of initial concepts to be evaluated. Ranking and rating exercises with numerous concepts can lead to a cognitive burden, so a best-worst scaling approach overcomes fatigue while generating additional information by making respondents choose the most and least important concepts. The result is a list of concepts in order of relative importance [36-38].

For each round, the MaxDiff Sawtooth software algorithm [39] generated a list of scaled scores from most to least important, as well as the relative importance of one concept compared with another. This was based on the number of times that concept was rated as the most or least important. In addition, cumulative scores were used to differentiate the number of concepts accounting for 20%, 50%, and 70% of the most important concepts; the 50th percentile was used as an initial guide for which concepts to retain.

Rounds I and II also featured an "anchor" value, which was a score marking the boundary between concepts considered important (to be retained in the next round) or unimportant (to be excluded from the next round). This anchor value was derived from an anchor question representing a threshold of "important or not important" included for each concept to be assessed as above (ie, important) or below (ie, unimportant) the anchor. The anchor question in this survey required one of the following 3 responses: "none of these [items] are important," "some of these [items] are important," or "all of these [items] are important." The anchor value was statistically derived using the Sawtooth software [38]. An illustration of a typical MaxDiff question is shown in Figure 3.

Figure 3. Typical presentation of the Delphi panel MaxDiff ranking question, including the anchor evaluation. NRT: nicotine replacement therapy; TNP: tobacco and/or nicotine products.

Which of the following statements/concepts are least and most important to measure and assess the impact on health and functioning when a TNP consumer switches from smoking a combustible TNP to using a smoke-free TNP or NRT?

Please choose one most important concept and one least important concept by clicking on the relevant button.

EXAMPLE [SINGLE ANSWER]

Most important	Choose one most important concept and one least important concept	Least important
◉	Shortness of breath or dyspnea	○
○	Cardiovascular symptoms (e.g. palpitations, blood pressure)	◉
○	Mouth ulcers	○
○	Physical endurance during everyday activities and exercise	○
○	Weight control	○

Considering only the concepts above, please indicate their overall importance:

Please click one answer

◉	None of these are important
○	Some of these are important
○	All of these are important

XSL•FO
RenderX

Latent Class Analysis

Latent class analysis [40] was also performed to determine whether some concepts were more important to some groups of experts over others, according to experts' demographic or other characteristics. This ensured that the final set of items was not biased toward one subset of health care professionals at the expense of others' preferences. Where there were concepts that were only important to subsets of the study sample but unimportant overall, these concepts could be retained.

Semiqualitative Assessment

Semiqualitative questions were incorporated into each Delphi panel round to provide an increased understanding of participants' responses to the results of the ranking and scoring exercise. These questions were asked to ensure no concepts were missing or needed rewording, and to understand the reasons behind experts' selections of importance or unimportance. They also sought to gauge agreement or disagreement with the overall ranking results and to gain insights into the reasons for these viewpoints. After each round, the semiqualitative comments were reviewed alongside the MaxDiff scores. This process involved coding and analyzing the qualitative feedback to identify common themes and insights that could provide context to the quantitative rankings. This feedback was then integrated with the quantitative MaxDiff results to refine and inform decisions about the list of concepts to be included for the next rounds. It helped explain the rationale behind the experts' choices and highlighted any discrepancies or areas of consensus, ensuring that both qualitative and quantitative perspectives were considered in the decision-making process.

Results

Descriptive Participant Demographics

Approximately 3100 experts were contacted, with a final sample of 105 experts recruited across 26 countries. Table 1 shows the demographics and participant characteristics, and Table 2 shows the participant quotas per region and per specialty. Of the initial 105 experts, all participated in exploratory round I and round II, and 101 (96%) completed round III (Table 2). Nearly all participants (90/105, 87%) reported frequently addressing smoking cessation with clients or patients. A total of 64% (67/105) of experts in the study identified as male, and most were located in the Americas, Europe, or Asia. Regarding place of work, three-quarters (79/105, 75%) of experts were based in hospitals or other clinical settings, mostly as medical specialists or general care practitioners.

XSL•FO
RenderX

Table 1. Demographics of Delphi panel participants (rounds I-III).

Demographic variable	Respondents	
	Rounds I and II (n=105), n (%)	Round III (n=101), n (%)
Sex		
Male	67 (64)	64 (63)
Female	38 (36)	37 (37)
Geographical region		
North America (United States and Canada)	20 (19)	17 (17)
Asia (Japan and South Korea)	17 (16)	17 (17)
Central, Caribbean, South America (Brazil, Colombia, Dominican Republic, and Guatemala)	15 (14)	15 (15)
Southern Europe (Spain, Portugal, Italy, Greece, and Bulgaria)	14 (13)	13 (13)
Western Europe (France, Germany, Sweden, United Kingdom, and the Netherlands)	13 (12)	13 (13)
Eastern Europe (Russia, Ukraine, and Lithuania)	9 (9)	9 (9)
Africa (South Africa)	6 (6)	5 (5)
Middle East (Israel)	6 (6)	6 (6)
Central Europe (Poland and Czech Republic)	5 (5)	5 (5)
Work setting		
Hospital	43 (41)	41 (41)
Clinical (other)	36 (34)	36 (36)
Professional	17 (16)	15 (15)
University hospital	5 (5)	5 (5)
University hospital; hospital setting	4 (4)	4 (4)
Main current role		
General practitioner or internal medicine	27 (26)	27 (27)
Oncology	19 (18)	19 (19)
Dentist or oral hygienist	16 (15)	15 (15)
Research	11 (10)	9 (9)
Cardiology	12 (11)	12 (12)
Respiratory	8 (8)	8 (8)
Counselor or psychologist	7 (7)	6 (6)
Advocacy or health policy or NGO[a]	5 (5)	5 (5)
Frequency of discussing smoking cessation with clients		
Frequent (at least weekly)	90 (86)	86 (85)
Not frequent (less than once a week)	11 (10)	11 (11)
Not applicable	4 (4)	4 (4)
Discussed the benefits of stopping smoking or provided information about sf-TNPs[b]or NRTs[c]to those who smoke and do not want or have difficulty quitting smoking		
Yes	103 (98)	99 (98)
No	0	0
Not available or prefer not to say	2 (2)	2 (2)[d,e]
Provided information on heat-not-burn or heated tobacco products (eg,*IQOS*, Glo) or tobacco vapor products (eg, Ploom Tech)		
Yes	50 (48)	46 (46)

XSL•FO
RenderX

Demographic variable	Respondents	
	Rounds I and II (n=105), n (%)	Round III (n=101), n (%)
No	53 (50)	53 (53)
Not available or prefer not to say	2 (2)	2 (2)
Provided information on smokeless tobacco (eg, Copenhagen Snuff, Swedish Match General Snus, Camel Snus, or any other local brands)		
Yes	39 (37)	35 (35)
No	64 (61)	64 (63)
Not available or prefer not to say	2 (2)	2 (2)
Provided information on e-cigarettes (eg, JUUL, Blu, Logic, or any other local brands)		
Yes	65 (62)	62 (61)
No	38 (36)	37 (37)
Not available or prefer not to say	2 (2)	2 (2)
Provided information on Nicotine replacement therapies (eg, nicotine gums, inhalers, nasal sprays, lozenges, patches)		
Yes	91 (87)	88 (87)
No	11 (10)	10 (10)
Not available or prefer not to say	3 (3)	3 (3)
Current or previous personal use of combustible TNPs[f]		
Never	65 (62)	62 (61)
Yes	35 (33)	34 (34)
Not available or prefer not to say	5 (5)	5 (5)
Current or previous personal use of sf-TNP or NRT of those that were currently or had used combustible TNPs		
Yes	12 (11)	12 (12)

[a]NGO: nongovernmental organization.

[b]sf-TNP: smoke-free tobacco or nicotine product.

[c]NRT: nicotine replacement therapy.

[d]Not available.

[e]Data missing for one respondent in round III, stage 2.

[f]TNP: tobacco or nicotine product.

Table 2. Participant quotas per region and per specialty for the start of round I of the Delphi panel (N=105).

Specialty	Asia	Africa	Middle East	North America	Central America	Southern Europe	Eastern Europe	Western Europe	Central Europe
Specialist physician	6	2	3	6	7	3	4	2	6
GP[a] or IM[b]	4	2	3	4	3	3	2	2	4
Dental specialist	3	1	0	2	2	2	3	1	2
SC[c] or DEP[d]	2	1	0	5	2	1	3	0	3
Health policy or NGO[e]	2	0	0	3	1	0	1	0	0
Total	17	6	6	20	14	9	13	5	15

[a]GP: general practitioner.

[b]IM: internal medicine specialist.

[c]SC: smoking cessation specialist.

[d]DEP: dependence or addiction specialist.

[e]NGO: nongovernmental organization.

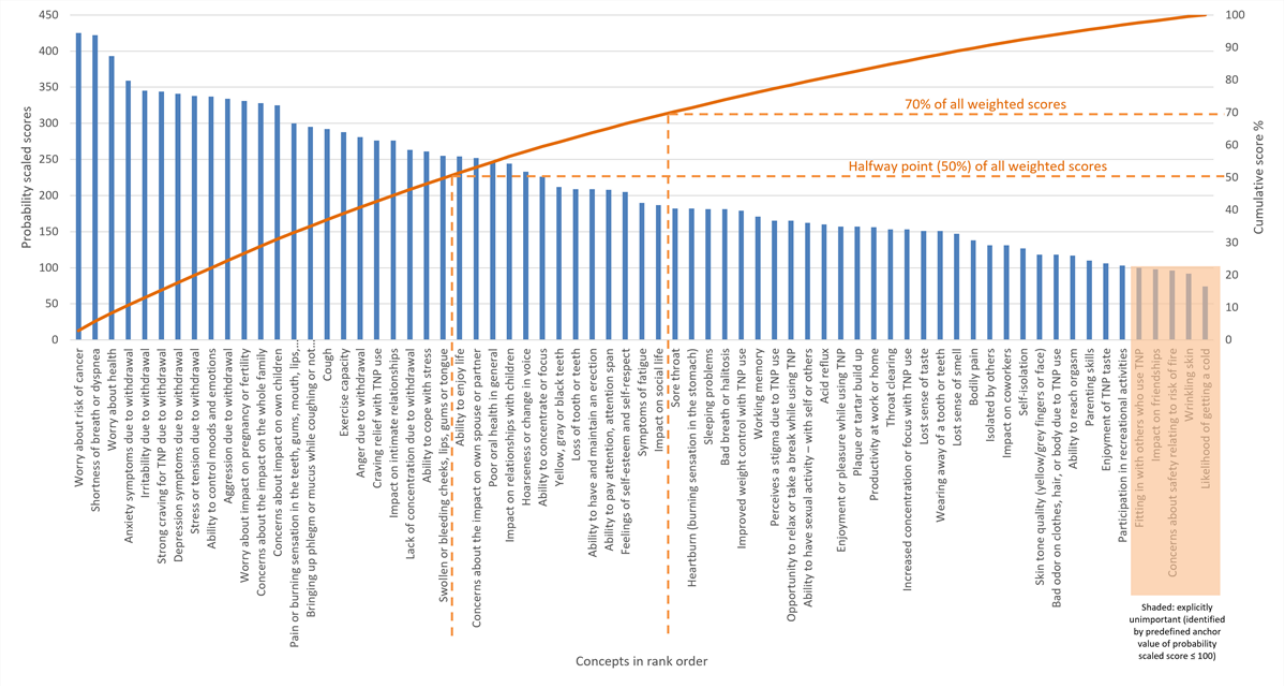
XSL•FO
RenderX

MaxDiff Analysis and Semiqualitative Assessment for Reduction or Refinement of Health and Functioning Concepts

Round I

A cumulative score was calculated by adding all the scaled scores (in this case, the total score was 14,900.17) and dividing by the scaled score (concept 1: 425/14,900=3%) and adding to the scaled score of the next highest concept (concept 2: 422/14,900=3%+3% concept 1 and so on; Figure 4). Regarding the relative importance of concepts in round I, when switching from combustible to sf-TNPs, experts placed importance on withdrawal symptoms, general health impacts, and emotion-related impacts. Conversely, concepts that ranked low in priority were those that referred to aesthetic concerns that would have less of a health impact or were not considered exclusive to those who smoke. Experts considered withdrawal symptoms such as anxiety and irritability as potential barriers to switching to sf-TNPs and highlighted there were levels of satisfaction (eg, enjoyment and craving relief) that those who smoke cigarettes would not want to lose when switching to sf-TNPs

Figure 4. Scaling and ranking results for items considered in round I of the Delphi panel. TNP: tobacco and/or nicotine products.



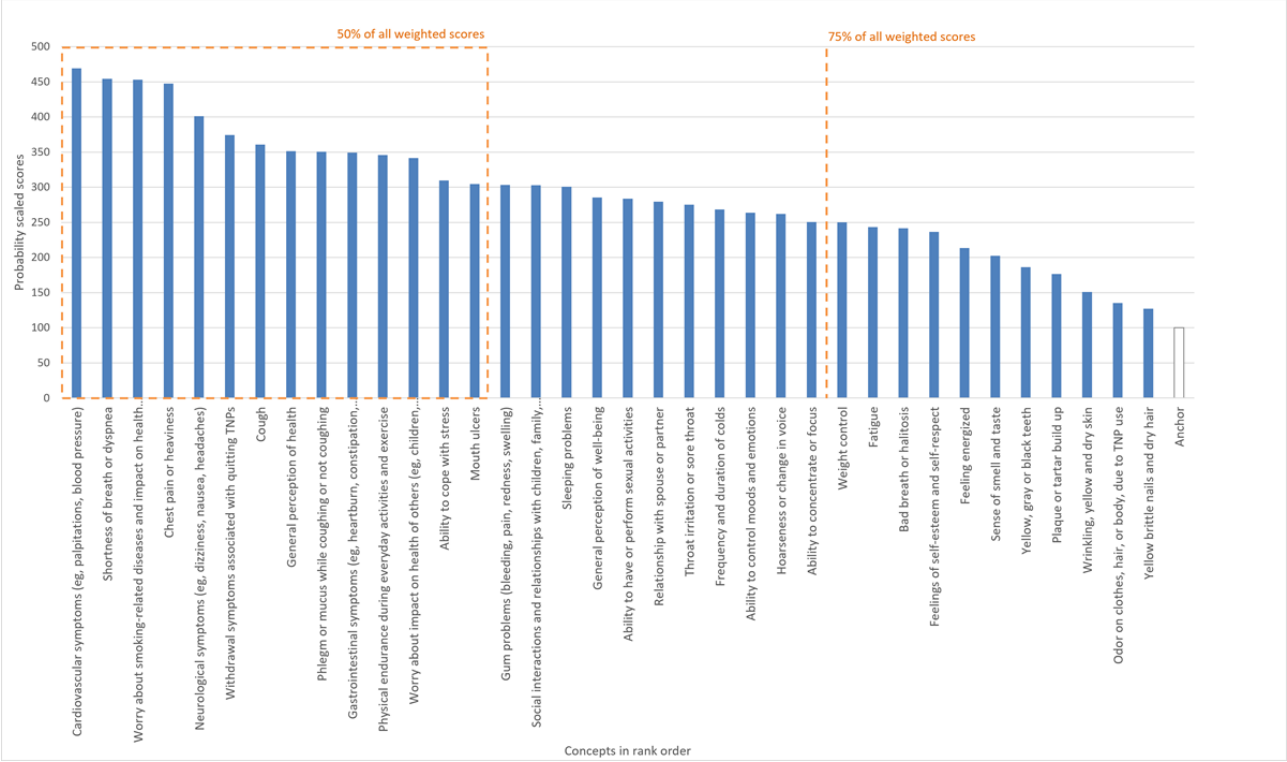
When determining which concepts to include in round II of the Delphi panel evaluation, the relative score (ie, importance) for each concept was considered, resulting in 36 out of 69 concepts (52% of all concepts in ranked order) accounting for 69% (10,285/14,900) of the total importance score. Based on the results and alignment with other activities ongoing within development of the health and functioning measurement instrument, the health and functioning concepts were further reviewed for relevance and clarity, and a different, refined, reduced, and synthesized list of 36 concepts was taken forward for evaluation in round II (refer to Supplementary Information and Table S1 in Multimedia Appendix 1 for further details).

Round II

Of the 36 concepts in this round, 14 made up 50% of the total scaled score (Figure 5). In general, physical health symptoms, and worries about the impact on health concepts remained the highest-ranked concepts and were considered by experts as more important than aesthetic considerations. Experts commented that concepts rated as unimportant and ranked lower in the logit analysis tended to lack specificity to smoking (ie, could be due to other causes), did not directly affect health, or were not reported frequently by TNP users. Of the 36 concepts evaluated in round II, a final selection of 20 highest-ranked concepts were identified for validation and discussion in round III.

XSL•FO
RenderX

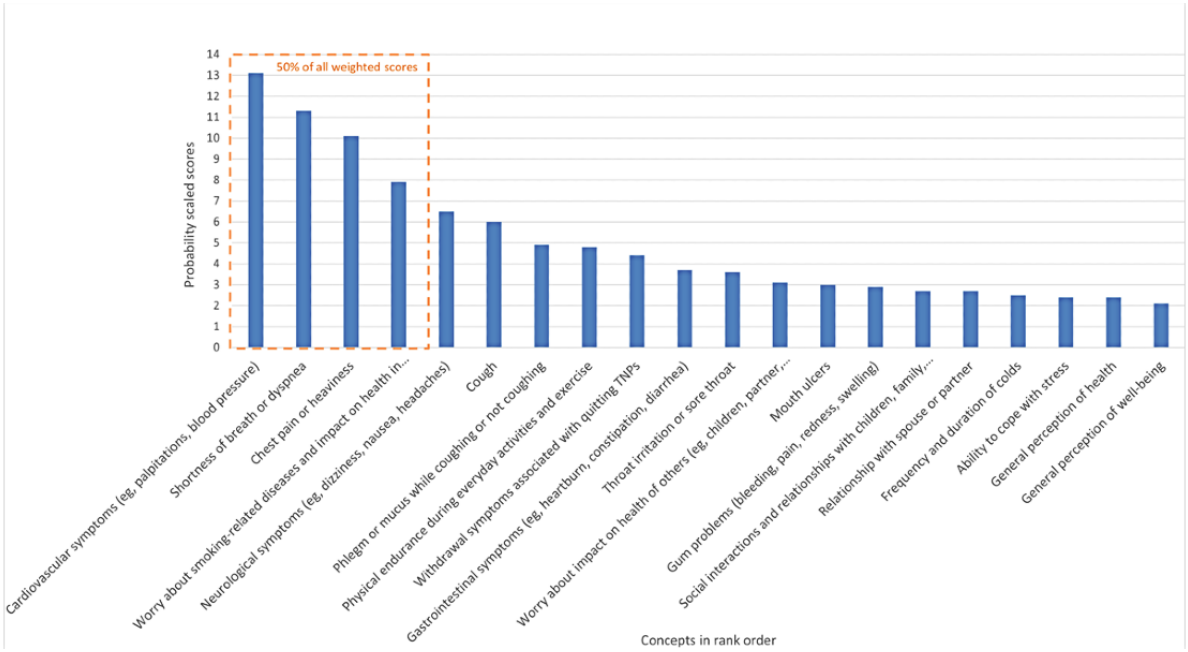
Figure 5. Scaling and ranking results for items considered in round II of the Delphi panel. TNP: tobacco and/or nicotine products.



Round III

Of the final 20 concepts ranked in round III (Figure 6), the sum of the top 4 concept scores accounted for more 50% of the total score: cardiovascular symptoms, shortness of breath or dyspnea, chest pain heaviness, and worry about smoking-related diseases and impact on health in general. Physical health concepts remained the most important experts across all 3 rounds

Figure 6. Ranking and ordering of the final 20 items identified in round III of the Delphi panel.



Latent Class Analysis

Latent class analysis of the results in round III revealed 2 identifiable subgroups, each with its own ranking and ordering of the final 20 concepts (Textbox 2). In both cases, concepts related to physical health, particularly cardiovascular and respiratory symptoms, ranked highest. In subgroup 1, the lowest-ranked concepts were related to psychosocial or general well-being and social interactions, and all scored equally low. For subgroup 2, there was minimal differentiation between concepts beyond the highest-ranking group.

XSL•FO
RenderX

Textbox 2. List of concepts in ranked order (highest to lowest) for the 2 subgroups identified by latent class analysis in round III of the Delphi panel.

Subgroup 1 in rank order

- Cardiovascular symptoms (eg, palpitations and blood pressure)
- Shortness of breath or dyspnea
- Chest pain or heaviness
- Neurological symptoms (eg, dizziness, nausea, and headaches)
- Cough
- Mouth ulcers
- Phlegm or mucus while coughing or not coughing
- Throat irritation or sore throat
- Gum problems (eg, bleeding, pain, redness, or swelling)
- Gastrointestinal symptoms (eg, heartburn, constipation, and diarrhea)
- Worry about smoking-related diseases and impact on health in general (eg, cancer and stroke)
- Withdrawal symptoms associated with quitting tobacco and/or nicotine products
- Frequency and duration of colds
- Physical endurance during everyday activities and exercise
- Ability to cope with stress
- Worry about impact on health of others (eg, children, partner, family, and friends)
- Relationship with spouse or partner
- General perception of health
- General perception of well-being
- Social interactions and relationships with children, family, friends, and colleagues

Subgroup 2 in rank order

- Cardiovascular symptoms (eg, palpitations and blood pressure)
- Shortness of breath or dyspnea
- Worry about smoking-related diseases and impact on health in general (eg, cancer and stroke)
- Chest pain or heaviness
- Physical endurance during everyday activities and exercise
- Neurological symptoms (eg, dizziness, nausea, and headaches)
- Cough
- Worry about impact on health of others (eg, children, partner, family, and friends)
- Social interactions and relationships with children, family, friends, and colleagues
- Withdrawal symptoms associated with quitting tobacco and/or nicotine products
- Relationship with spouse or partner
- Phlegm or mucus while coughing or not coughing
- General perception of health
- General perception of well-being
- Ability to cope with stress
- Gastrointestinal symptoms (eg, heartburn, constipation, and diarrhea)
- Throat irritation or sore throat
- Frequency and duration of colds
- Gum problems (eg, bleeding, pain, redness, and swelling)
- Mouth ulcers

XSL•FO
RenderX

In stage 2 of round III, following the MaxDiff analysis, the ranked concept lists for each of the 2 latent class subgroups were shown to the experts, who were asked to identify which group best reflected their own thinking and provide a rationale for their choice. Just over half of the experts (58/100, 58%) reported that subgroup 1 best reflected their thinking compared with (42/100, 42%) for subgroup 2. When asked about their rationale for aligning with subgroup 1, the main reason was a focus on the most important physical health symptoms (33/58, 57%). The main rationale for choosing subgroup 2 was a greater emphasis on the importance of the consumer and their perspective (10/42, 24%), including psychological elements, the well-being of consumers, and impact of smoking on the health of others (14/42, 33%).

Regarding the influence of professional specialty in self-alignment with subgroups (Table 3), preference for subgroup 1 tended to be among oncologists (14/19, 74%), dentists (11/15, 73%), counselors (4/6, 67%), and general practitioners or internal medicine specialists (17/27, 63%). Conversely, all policy advisors and charity workers (5/5) felt that subgroup 2 best reflected their views. When geographical location was considered (Table 3), there was a strong preference for subgroup 2 in participants from Asia (12/17, 71%) and a slight preference among those in North America (9/17, 53%). Experts from all other regions tended mostly to identify with subgroup 1 (Middle East: 5/6, 83%; Africa: 4/5, 80%; Central Europe: 4/5, 80%; Eastern Europe: 7/9, 78%; Western Europe: 9/13, 69%; Southern Europe: 8/13, 62%; and Central or South America: 8/15, 53%). Regarding the final selection of 20 concepts, both subgroups considered all concepts to be important and relevant for inclusion in any self-reported outcome measure.

Table 3. Region and specialty after self-identifying with a latent class subgroup in round III of the Delphi panel.

Experts' region and specialty	Total (n=100), n (%)[a]	Latent class subgroup 1, n (%)[b]	Latent class subgroup 2, n (%)[b]
Region	100 (100)	58 (58)	42 (42)
North America	17 (17)	8 (47)	9 (53)
Western Europe	13 (13)	9 (69)	4 (31)
Southern Europe	13 (13)	8 (62)	5 (38)
Africa	5 (5)	4 (80)	1 (20)
Central/South America	15 (15)	8 (53)	7 (47)
Central Europe	5 (5)	4 (80)	1 (20)
Asia	17 (17)	5 (29)	12 (71)
Eastern Europe	9 (9)	7 (78)	2 (22)
Middle East	6 (6)	5 (83)	1 (17)
Specialty	100 (100)	58 (58)	42 (42)
General practice or internal medicine	27 (27)	17 (63)	10 (37)
Dentist or oral hygienist	15 (15)	11 (73)	4 (27)
Oncologist	19 (19)	14 (74)	5 (26)
Cardiovascular specialist	11 (11)	5 (45)	6 (55)
Respiratory	8 (8)	3 (38)	5 (63)
Researcher	9 (9)	4 (44)	5 (56)
Counselor or psychologist	6 (6)	4 (67)	2 (33)
Policy advisor	4 (4)	0 (0)	4 (100)
Charity or advocacy	1 (1)	0 (0)	1 (100)

[a]Data for 100 experts were included, with one respondent (from South Africa) reporting not understanding the questions asked.

[b]Percentage use the n value in the "Total" column as the denominator.

Additional Semiqualitative Assessment—Likelihood to Perceive Change

In round III (in response to the question: "Now that you have seen the list of concepts collated from Rounds I & II, how likely do you think it is to see a change in this concept when a TNP user switches from a combustible TNP to an sf-TNP?"), most experts reported they would expect to find measurable levels of change in 13 out of 20 individual concepts. Furthermore, over half of the respondents stated it was "likely" or "extremely likely" that a measurable change would occur in these concepts when a TNP user switched from a combustible TNP to a sf-TNP (Figure 7). For example, majority of experts felt that it was "likely" or "extremely likely" to observe a change in gum problems (74/101, 73%), phlegm or mucus while coughing or not coughing (72/101, 71%), general perception of well-being (72/101, 71%), and throat irritation/sore throat (72/101, 71%). In addition, between 5% and 20% considered that switching

XSL•FO
RenderX

was "unlikely" or "very unlikely" to result in a change in the severity of any single concept, including cardiovascular symptoms, withdrawal symptoms associated with quitting, shortness of breath, and worry about smoking-related diseases. Reasons given by experts for considering a change in a concept to be unlikely are listed in Table 4 and were mostly related to lack of perceived substantial differences between combustible TNPs and sf-TNPs.

Figure 7. Experts' opinions (percentage of experts) regarding likelihood of observing change in the severity of the final 20 items in round III of the Delphi panel.

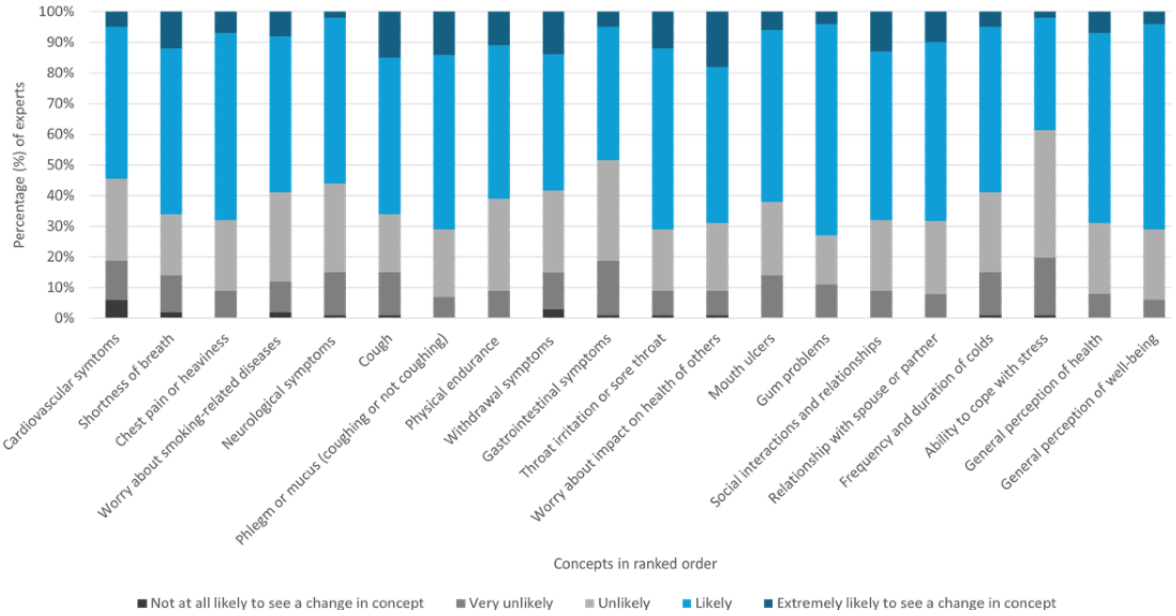


Table 4. Reasons given by experts for considering a change in a concept to be unlikely for the final 20 items in round III of the Delphi panel.

Concept (number of experts responding it was unlikely to see a change)	Rationale reported by experts
Cardiovascular symptoms (n=6)	• Unlikely to see substantial modifications between the combustible TNPs[a] and sf-TNPs[b] as tobacco or nicotine can cause cardiovascular disease • Active ingredients are the same and present in both combustible TNPs and sf-TNPs products
Shortness of breath or dyspnea (n=2)	• No change or no substantial modifications in switching to sf-TNPs
Worry about smoking-related diseases and impact on health in general (eg, cancer and stroke) (n=2)	• Symptoms will occur with using both combustible TNPs and sf-TNPs
Worry about impact on health of others (eg, children, partner, family, and friends) (n=1)	• Other dangerous chemicals in sf-TNPs
Frequency and duration of colds (n=1)	• Colds are unlikely to change as they are a nuisance and not a major health threat

[a]TNPs: tobacco and/or nicotine products.

[b]sf-TNPs: smoke-free tobacco and/or nicotine products.

Discussion

Principal Findings

This adapted Delphi panel study was conducted to identify the health and functioning concepts that experts consider to be most relevant and useful for incorporation into a new TNP-specific health and functioning measurement instrument to assess the self-reported impact of switching from combustible TNPs to sf-TNPs [21]. We recruited a geographically and professionally representative panel of experts who routinely treated those who smoke or communicated on smoking and TNP use. They were asked to reduce and refine a preliminary list of health and functioning concepts and to rank the final list in order of importance. After 3 rounds of investigation, the initial list of 69 concepts was refined to the most important 20 they considered could be included in a new self-reported measure (Table 5). The range of final concepts reflected health and psychosocial issues associated with smoking, including respiratory and cardiovascular health, withdrawal symptoms, and worries about the impact on health of self and others.

XSL•FO
RenderX

Table 5. Concepts recommended and agreed after each round of the Delphi panel.

Rank	Round I (n=69), ranked in order of importance (highest to lowest)	Round II (n=36), ranked in order of importance (highest to lowest)	Round III (n=20), ranked in order of importance (highest to lowest)
1	Worry about risk of cancer	Cardiovascular symptoms (eg, palpitations, blood pressure)	Cardiovascular symptoms (eg, palpitations, blood pressure)
2	Shortness of breath or dyspnea	Shortness of breath or dyspnea	Shortness of breath or dyspnea
3	Worry about health	Worry about smoking-related diseases and impact on health in general (eg, cancer, stroke)	Chest pain or heaviness
4	Anxiety symptoms due to withdrawal	Chest pain or heaviness	Worry about smoking-related diseases and impact on health in general (eg, cancer, stroke)
5	Irritability due to withdrawal	Neurological symptoms (eg, dizziness, nausea, headaches)	Neurological symptoms (eg, dizziness, nausea, headaches)
6	Strong craving for TNP[a] due to withdrawal	Withdrawal symptoms associated with quitting TNPs	Cough
7	Depression symptoms due to withdrawal	Cough	Phlegm or mucus while coughing or not coughing
8	Stress or tension due to withdrawal	General perception of health	Physical endurance during everyday activities and exercise
9	Ability to control moods and emotions	Phlegm or mucus while coughing or not coughing	Withdrawal symptoms associated with quitting TNPs
10	Aggression due to withdrawal	Gastrointestinal symptoms (eg, heartburn, constipation, diarrhea)	Gastrointestinal symptoms (eg, heartburn, constipation, diarrhea)
11	Worry about impact on pregnancy or fertility	Physical endurance during everyday activities and exercise	Throat irritation or sore throat
12	Concerns about the impact on the whole family	Worry about impact on health of others (eg, children, partner, family, friends)	Worry about impact on health of others (eg, children, partner, family, friends)
13	Concerns about impact on own children	Ability to cope with stress	Mouth ulcers
14	Pain or burning sensation in the teeth, gums, mouth, lips, throat, or tongue	Mouth ulcers	Gum problems (bleeding, pain, redness, swelling)
15	Bringing up phlegm or mucus while coughing or not coughing	Gum problems (bleeding, pain, redness, swelling)	Social interactions and relationships with children, family, friends, colleagues
16	Cough	Social interactions and relationships with children, family, friends, colleagues	Relationship with spouse or partner
17	Exercise capacity	Sleeping problems	Frequency and duration of colds
18	Anger due to withdrawal	General perception of well-being	Ability to cope with stress
19	Craving relief with TNP use	Ability to have or perform sexual activities	General perception of health
20	Impact on intimate relationships	Relationship with spouse or partner	General perception of well-being
21	Lack of concentration due to withdrawal	Throat irritation or sore throat	—
22	Ability to cope with stress	Frequency and duration of colds	—
23	Swollen or bleeding cheeks, lips, gums, or tongue	Ability to control moods and emotions	—
24	Ability to enjoy life	Hoarseness or change in voice	—
25	Concerns about the impact on own spouse or partner	Ability to concentrate or focus	—
26	Poor oral health in general	Weight control	—
27	Impact on relationships with children	Fatigue	—
28	Hoarseness or change in voice	Bad breath or halitosis	—
29	Ability to concentrate or focus	Feelings of self-esteem and self-respect	—
30	Yellow, gray, or black teeth	Feeling energized	—
31	Loss of tooth or teeth	Sense of smell and taste	—

XSL•FO
RenderX

Rank	Round I (n=69), ranked in order of importance (highest to lowest)	Round II (n=36), ranked in order of importance (highest to lowest)	Round III (n=20), ranked in order of importance (highest to lowest)
32	Ability to have and maintain an erection	Yellow, gray, or black teeth	—
33	Ability to pay attention, attention span	Plaque or tartar build up	—
34	Feelings of self-esteem and self-respect	Wrinkling, yellow, and dry skin	—
35	Symptoms of fatigue	Odor on clothes, hair, or body, due to TNP use	—
36	Impact on social life	Yellow brittle nails and dry hair	—
37	Sore throat	—	—
38	Heartburn (burning sensation in the stomach)	—	—
39	Sleeping problems	—	—
40	Bad breath or halitosis	—	—
41	Improved weight control with TNP use	—	—
42	Working memory	—	—
43	Perceives a stigma due to TNP use	—	—
44	Opportunity to relax or take a break while using TNP	—	—
45	Ability to have sexual activity—with self or others	—	—
46	Acid reflux	—	—
47	Enjoyment or pleasure while using TNP	—	—
48	Plaque or tartar build up	—	—
49	Productivity at work or home	—	—
50	Throat clearing	—	—
51	Increased concentration or focus with TNP use	—	—
52	Lost sense of taste	—	—
53	Wearing away of a tooth or teeth	—	—
54	Lost sense of smell	—	—
55	Bodily pain	—	—
56	Isolated by others	—	—
57	Impact on co-workers	—	—
58	Self-isolation	—	—
59	Skin tone quality (yellow or gray fingers or face)	—	—
60	Bad odor on clothes, hair, or body due to TNP use	—	—
61	Ability to reach orgasm	—	—
62	Parenting skills	—	—
63	Enjoyment of TNP taste	—	—
64	Participation in recreational activities	—	—
65	Fitting in with others who use TNP	—	—
66	Impact on friendships	—	—
67	Concerns about safety relating to risk of fire	—	—
68	Wrinkling skin	—	—
69	Likelihood of getting a cold	—	—

[a]TNP: tobacco and/or nicotine products.

Overall, the final 20 concepts were considered both clinically relevant and important to the experts in their evaluation of the possible impact of switching to sf-TNPs or stopping smoking cigarettes. We observed that the concepts consistently ranked highly among experts focused on objectively measurable health consequences such as respiratory and cardiovascular symptoms,

XSL•FO
RenderX

as well as overall physical functioning. In contrast, concepts such as the ability to cope with stress, risk of getting colds, oral health, physical appearance, social functioning, and sensory impacts (taste, sense of smell, and odor) were ranked as less important. Those concepts ranking highest reflect current evidence regarding smoking-related health outcomes and potential improvements attainable by ceasing cigarette consumption or switching to sf-TNPs [6,41-46]. The experts' knowledge of evidence for the effect of cessation or switching on pre-existing respiratory conditions [6,43,44,47,48] may also be reflected in the ranking.

It is essential to note that the relative ranking of these concepts is likely to be different for TNP consumers compared with health care professionals. Individuals trying to stop smoking or switch to sf-TNPs recognize the respiratory and other physical health and functional benefits and understand the potential for improved quality of life by reducing cigarette consumption or stopping smoking [49,50], as well as reduced exposure of others to smoking-related harm [6,51]. They also tend to report benefits and preferences for general hygiene and smell, better oral health, and sensory improvements [6,41,43]. However, no concepts related to general physical appearance or hygiene were in the final top-ranking concepts of the current study. These results explicitly reflect the participants' own experiences and opinions, which may differ from those of TNP users and be dependent on TNP user characteristics.

Although objective health outcomes such as cardiovascular and respiratory function were rated highly by experts throughout the study, some individuals rated psychosocial outcomes and effects on families and others (through secondhand smoke and so on) as equally important as objective physical measures. Latent class analysis in the final round of the study suggested that geographical location, cultural milieu, and professional specialism may each play a major role in these observations. Specifically, we identified a subgroup that focused mostly on physical, objectively measurable concepts (eg, cardiovascular symptoms, shortness of breath, chest pain, neurological symptoms, and cough), compared with the other subgroup, which considered both objective and subjective or emotional concepts (eg, worry about smoking-related diseases and impact on health both of self and others) to be of equal importance. These results may correspond with established patterns of perceptions and considerations among Western cultures, which may be broadly defined as individualistic (ie, driven by personal goals), whereas those from Asia and South America can be defined as collectivist (reflecting the primacy of mutual obligations among members of society) [51-53]. Furthermore, qualitative investigation of alignment with the subgroups revealed that for certain job roles (eg, counselors and policy advisors), the consumer-focused concepts were more important than for other health care professionals (eg, specialist clinicians), who were primarily concerned with objective, physical concepts. These findings have important implications for the development of a self-report measure that can be widely used and is adaptable to individual or local requirements [54]. When disseminating results based on the newly developed self-reported measure, it may be essential to tailor communications on sf-TNPs for specific audiences with different considerations of what is clinically meaningful to the experts themselves, as well as to the individual TNP user [26,33,54]. To validate the latent class findings in round III, we asked respondents to self-identify with the most relevant of the 2 subgroups. Overall, participants were able to do so readily, indicating that the 2 classes were culturally and professionally relevant, and that the new self-report measure will maximize content validity by including the range of concepts identified and ranked in round III to accommodate cultural and professional requirements and differences.

There are 3 main strengths of this study. First, we used the MaxDiff ranking to address the challenge of ranking an initial high number of concepts. Second, the incorporation of an anchor was essential in objectively identifying concepts for elimination. Third, a large panel of experts representing a broad range of professional expertise and geographies was recruited for this study. This allowed us to confirm validity of the concepts identified in the preparatory phase of the development of the new measure. It also enabled us to parse out cultural and professional subtleties that are important to consider when disseminating outcomes from the new measure. Development of an outcome measure involving both the target population and clinicians is a key component of creating a high-quality instrument [33,54]. Building this manner of collaboration and bridging between TNP users and relevant experts in this Delphi panel into the development of this new instrument would serve to enhance the potential quality and validity of the final measure by feeding into subsequent qualitative and psychometric evaluations [26].

There are also some limitations to our study. First, we did not use a strict Delphi panel process, and round I may be seen as a preliminary evaluation rather than a pure ranking and scoring exercise. Delphi panel methodology typically follows a highly structured process. In this particular case, it would have been ideal to perform additional preliminary qualitative insight work to review all the elements of a study, formally review and refine concept terminology, and test with external audiences. And, as previously mentioned, this Delphi panel focused on the experience and perceptions of experts and professionals who interact with TNP users. Consequently, the findings reported here do not have the vital context of TNP consumers opinions and perceptions. Unfortunately, it would have been impractical to address both angles in a single study with an initial list of 69 concepts. Instead, further research should consider evaluating TNP users' perceptions and rankings in a separate study using similar methodology. In addition, there is a potential for selection bias in the study due to the exclusion of experts who specified they would not consider participating in studies by tobacco companies. This exclusion could result in a sample that is not fully representative of the broader expert community, potentially impacting the generalizability of the findings. To mitigate this, we used a diverse recruitment strategy to include experts from various regions and specialties. However, the views of nonparticipating experts might differ from those who participated, and this limitation should be considered when interpreting the results. Finally, the survey was conducted in various languages, and despite the measures taken to ensure the consistency and validity of the translated surveys, linguistic and cultural differences may have led to variations in interpretation

XSL•FO
RenderX

of the survey. These variations could potentially affect the study's findings and should be considered when interpreting the results.

Conclusion

In conclusion, this 3-round, adapted Delphi panel identified a ranked list of 20 concepts to be considered when assessing the health and functioning status of individuals who stop smoking cigarettes or switch to using sf-TNPs or NRTs. It is to be expected that the scale of importance of each concept will vary based on the health status and concerns of an individual, whereas the ranking presented here represents a global, generalized view provided by the participating experts. In addition, the sensitivity of the concepts to accurately reflect changes in TNP use behavior will need to be determined. This would support the evaluation of the self-reported experience and impact of switching from conventional cigarettes to sf-TNPs on health risks and contribute to the regulatory and scientific evidence base for understanding both the individual and public health impacts of sf-TNPs.

Acknowledgments

We wish to thank David Floyd for providing editorial assistance. Philip Morris International is the sole source of funding and sponsor of this research.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare the following financial interests or personal relationships which may be considered as potential conflicting interests: EA and CC are employees of Philip Morris Products S.A. VL, LAW, LW, and AT were contracted and paid by Philip Morris Products S.A.

Multimedia Appendix 1
Amended/modified Delphi panel approach.
[DOCX File , 38 KB - formative_v9i1e58614_app1.docx]

References

1. WHO report on the global tobacco epidemic 2021: addressing new and emerging products. Geneva, Switzerland: World Health Organization (WHO); 2021.
2. Mersha AG, Gould GS, Bovill M, Eftekhari P. Barriers and facilitators of adherence to nicotine replacement therapy: a systematic review and analysis using the capability, opportunity, motivation, and behaviour (COM-B) model. Int J Environ Res Public Health 2020;17(23):8895 [FREE Full text] [doi: 10.3390/ijerph17238895] [Medline: 33265956]
3. Food and Drug Administration (FDA). 2020. URL: https://www.fda.gov/tobacco-products/products-guidance-regulations/products-ingredients-components [accessed 2021-12-21]
4. Guidance for Industry - Modified Risk Tobacco Product Applications (Draft Guidance). Silver Spring, MD: Food and Drug Administration (FDA) Center for Tobacco Products; 2012.
5. Modified risk orders. Food and Drug Administration (FDA). URL: https://www.fda.gov/tobacco-products/advertising-and-promotion/modified-risk-orders [accessed 2021-12-21]
6. Glasser AM, Collins L, Pearson JL, Abudayyeh H, Niaura RS, Abrams DB, et al. Overview of electronic nicotine delivery systems: a systematic review. Am J Prev Med 2017;52(2):e33-e66. [doi: 10.1016/j.amepre.2016.10.036] [Medline: 27914771]
7. Hajek P, Phillips-Waller A, Przulj D, Pesola F, Smith KM, Bisal N, et al. E-cigarettes compared with nicotine replacement therapy within the UK stop smoking services: the TEC RCT. Health Technol Assess 2019;23(43):1-82 [FREE Full text] [doi: 10.3310/hta23430] [Medline: 31434605]
8. McRobbie HJ, Phillips-Waller A, El Zerbi C, McNeill A, Hajek P, Pesola F, et al. Nicotine replacement treatment, e-cigarettes and an online behavioural intervention to reduce relapse in recent ex-smokers: a multinational four-arm RCT. Health Technol Assess 2020;24(68):1-82 [FREE Full text] [doi: 10.3310/hta24680] [Medline: 33270009]
9. Goldenberg M, Danovitch I, IsHak WW. Quality of life and smoking. Am J Addict 2014;23(6):540-562. [doi: 10.1111/j.1521-0391.2014.12148.x] [Medline: 25255868]
10. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the lancet commission. Lancet 2020;396(10248):413-446 [FREE Full text] [doi: 10.1016/S0140-6736(20)30367-6] [Medline: 32738937]
11. Qasim H, Karim ZA, Rivera JO, Khasawneh FT, Alshbool FZ. Impact of electronic cigarettes on the cardiovascular system. J Am Heart Assoc 2017;6(9):e006353 [FREE Full text] [doi: 10.1161/JAHA.117.006353] [Medline: 28855171]
12. Critchley JA, Unal B. Health effects associated with smokeless tobacco: a systematic review. Thorax 2003;58(5):435-443 [FREE Full text] [doi: 10.1136/thorax.58.5.435] [Medline: 12728167]

XSL•FO
RenderX

13. Znyk M, Jurewicz J, Kaleta D. Exposure to heated tobacco products and adverse health effects, a systematic review. Int J Environ Res Public Health 2021;18(12):6651 [FREE Full text] [doi: 10.3390/ijerph18126651] [Medline: 34205612]
14. Marques P, Piqueras L, Sanz M. An updated overview of e-cigarette impact on human health. Respir Res 2021;22(1):151 [FREE Full text] [doi: 10.1186/s12931-021-01737-5] [Medline: 34006276]
15. Institute of Medicine, Board on Population Health and Public Health Practice, Committee on Scientific Standards for Studies on Modified Risk Tobacco Products. Scientific Standards for Studies on Modified Risk Tobacco Products. Washington, DC: National Academy of Science; 2012.
16. Gasperini B, Barbadoro P, Cherubini A, Pierri F, D'Errico MM, Di Stanislao F, et al. The earlier the better: health-related quality of life is inversely related to age of smoking cessation in older former Italian smokers. Aging Clin Exp Res 2017;29(4):655-663. [doi: 10.1007/s40520-016-0608-4] [Medline: 27456679]
17. Sarna L, Bialous SA, Cooley ME, Jun H, Feskanich D. Impact of smoking and smoking cessation on health-related quality of life in women in the nurses' health study. Qual Life Res 2008;17(10):1217-1227. [doi: 10.1007/s11136-008-9404-8] [Medline: 18931942]
18. Kulasekaran A, Proctor C, Papadopoulou E, Shepperd CJ, Guyer R, Gandek B, et al. Preliminary evaluation of a new German translated tobacco quality of life impact tool to discriminate between healthy current and former smokers and to explore the effect of switching smokers to a reduced toxicant prototype cigarette. Nicotine Tob Res 2015;17(12):1456-1464. [doi: 10.1093/ntr/ntv024] [Medline: 25914263]
19. Olufade AO, Shaw JW, Foster SA, Leischow SJ, Hays RD, Coons SJ. Development of the smoking cessation quality of life questionnaire. Clin Ther 1999;21(12):2113-2130. [doi: 10.1016/s0149-2918(00)87242-2] [Medline: 10645757]
20. Ware JE, Gandek B, Kulasekaran A, Guyer R. Evaluation of smoking-specific and generic quality of life measures in current and former smokers in Germany and the United States. Health Qual Life Outcomes 2015;13:128 [FREE Full text] [doi: 10.1186/s12955-015-0316-3] [Medline: 26276447]
21. Chrea C, Acquadro C, Afolalu EF, Spies E, Salzberger T, Abetz-Webb L, et al. Developing fit-for-purpose self-report instruments for assessing consumer responses to tobacco and nicotine products: the ABOUT™ toolbox initiative. F1000Res 2018;7:1878 [FREE Full text] [doi: 10.12688/f1000research.16810.1] [Medline: 30906527]
22. Afolalu EF, Spies E, Bacso A, Clerc E, Abetz-Webb L, Gallot S, et al. Impact of tobacco and/or nicotine products on health and functioning: a scoping review and findings from the preparatory phase of the development of a new self-report measure. Harm Reduct J 2021;18(1):79 [FREE Full text] [doi: 10.1186/s12954-021-00526-z] [Medline: 34330294]
23. Afolalu EF, Spies E, Bacso A, Clerc E, Abetz-Webb L, Gallot S, et al. The process of developing a new outcome measure to assess the impact of tobacco and/or nicotine-containing products on health and functioning. 2020 Presented at: 26th Annual Meeting Society for Research on Nicotine and Tobacco (SRNT); March 11-14, 2020; New Orleans, LA.
24. Afolalu EF, Lanar S, Abetz-Webb L, Bacso A, Blankenspoor M, Koochaki P, et al. Qualitative health and functioning concept elicitation interviews in Japanese users of heated tobacco products. 2021 Presented at: 27th Annual Meeting Society for Research on Nicotine and Tobacco (SRNT); February 24, 2021; Virtual event URL: https://www.pmiscience.com/content/dam/pmiscience/en/publications/presentations-and-posters/pdf/srnt-2021-poster-esther-afolalu.pdf
25. Asfar T, Schmidt M, Ebrahimi Kalan M, Wu W, Ward KD, Nakkash RT, et al. Delphi study among international expert panel to develop waterpipe-specific health warning labels. Tob Control 2020;29(2):159-167 [FREE Full text] [doi: 10.1136/tobaccocontrol-2018-054718] [Medline: 30696784]
26. Hepworth LR, Rowe FJ. Using delphi methodology in the development of a new patient-reported outcome measure for stroke survivors with visual impairment. Brain Behav 2018;8(2):e00898 [FREE Full text] [doi: 10.1002/brb3.898] [Medline: 29484260]
27. Santaguida PL, Oliver D, Gilsing A, Lamarche L, Griffith LE, Mangin D, et al. Delphi consensus on core criteria set selecting among health-related outcome measures (HROM) in primary health care. J Clin Epidemiol 2020;127:105-116. [doi: 10.1016/j.jclinepi.2020.04.028] [Medline: 32417261]
28. Blaser J, Cornuz J. Experts' consensus on use of electronic cigarettes: a Delphi survey from Switzerland. BMJ Open 2015;5(4):e007197 [FREE Full text] [doi: 10.1136/bmjopen-2014-007197] [Medline: 25877274]
29. O'Neill N, Dogar O, Jawad M, Kellar I, Kanaan M, Siddiqi K. Which behavior change techniques may help waterpipe smokers to quit? An expert consensus using a modified Delphi technique. Nicotine Tob Res 2018;20(2):154-160 [FREE Full text] [doi: 10.1093/ntr/ntw297] [Medline: 27798091]
30. Lintonen T, Konu A, Rönkä S, Kotovirta E. Drugs foresight 2020: a Delphi expert panel study. Subst Abuse Treat Prev Policy 2014;9:18 [FREE Full text] [doi: 10.1186/1747-597X-9-18] [Medline: 24885142]
31. Marten O, Koerber F, Bloom D, Bullinger M, Buysse C, Christensen H, et al. A Delphi study on aspects of study design to overcome knowledge gaps on the burden of disease caused by serogroup B invasive meningococcal disease. Health Qual Life Outcomes 2019;17(1):87 [FREE Full text] [doi: 10.1186/s12955-019-1159-0] [Medline: 31118091]
32. O'Donnell S, Doyle G, O'Malley G, Browne S, O'Connor J, Mars M, et al. Establishing consensus on key public health indicators for the monitoring and evaluating childhood obesity interventions: a Delphi panel study. BMC Public Health 2020;20(1):1733 [FREE Full text] [doi: 10.1186/s12889-020-09814-y] [Medline: 33203390]

XSL•FO
RenderX

33. Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG. Guidance on conducting and reporting DElphi studies (CREDES) in palliative care: recommendations based on a methodological systematic review. Palliat Med 2017;31(8):684-706. [doi: 10.1177/0269216317690685] [Medline: 28190381]
34. Strasser A. Design and evaluation of ranking-type Delphi studies using best-worst-scaling. Technol Anal Strateg Manag 2018;31(4):492-501. [doi: 10.1080/09537325.2018.1521956]
35. Uy EJB, Bautista DC, Xin X, Cheung YB, Thio S, Thumboo J. Using best-worst scaling choice experiments to elicit the most important domains of health for health-related quality of life in Singapore. PLoS One 2018;13(2):e0189687 [FREE Full text] [doi: 10.1371/journal.pone.0189687] [Medline: 29420564]
36. Mühlbacher AC, Kaczynski A, Zweifel P, Johnson FR. Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview. Health Econ Rev 2016;6(1):2. [doi: 10.1186/s13561-015-0079-x] [Medline: 26743636]
37. Cheung KL, Wijnen BFM, Hollin IL, Janssen EM, Bridges JF, Evers SMAA, et al. Using best-worst scaling to investigate preferences in health care. Pharmacoeconomics 2016;34(12):1195-1209. [doi: 10.1007/s40273-016-0429-5] [Medline: 27402349]
38. Sawtooth S. The MaxDiff System Technical Paper (Version 8). Orem, UT: Sawtooth Software, Inc; 2013.
39. Lighthouse Studio Manual. Orem, UT: Sawtooth Software, Inc; 2013.
40. Kongsted A, Nielsen AM. Latent class analysis in health research. J Physiother 2017;63(1):55-58. [doi: 10.1016/j.jphys.2016.05.018] [Medline: 27914733]
41. Abafalvi L, Pénzes M, Urbán R, Foley KL, Kaán R, Kispélyi B, et al. Perceived health effects of vaping among Hungarian adult e-cigarette-only and dual users: a cross-sectional internet survey. BMC Public Health 2019;19(1):302 [FREE Full text] [doi: 10.1186/s12889-019-6629-0] [Medline: 30866897]
42. Benowitz NL, Burbank AD. Cardiovascular toxicity of nicotine: implications for electronic cigarette use. Trends Cardiovasc Med 2016;26(6):515-523. [doi: 10.1016/j.tcm.2016.03.001] [Medline: 27079891]
43. Farsalinos KE, Romagna G, Tsiapras D, Kyrzopoulos S, Voudris V. Characteristics, perceived side effects and benefits of electronic cigarette use: a worldwide survey of more than 19,000 consumers. Int J Environ Res Public Health 2014;11(4):4356-4373 [FREE Full text] [doi: 10.3390/ijerph110404356] [Medline: 24758891]
44. Polosa R, Cibella F, Caponnetto P, Maglia M, Prosperini U, Russo C, et al. Health impact of E-cigarettes: a prospective 3.5-year study of regular daily users who have never smoked. Sci Rep 2017;7(1):13825 [FREE Full text] [doi: 10.1038/s41598-017-14043-2] [Medline: 29150612]
45. Choi S, Lee K, Park SM. Combined associations of changes in noncombustible nicotine or tobacco product and combustible cigarette use habits with subsequent short-term cardiovascular disease risk among South Korean men: a nationwide cohort study. Circulation 2021;144(19):1528-1538 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.121.054967] [Medline: 34601948]
46. Sharman A, Nurmagambetov T. Changes in respiratory function and physical capacity among smokers after switching to iqos: one year follow-up. Glob. J. Respir. Care 2020;6:21-27 [FREE Full text]
47. Polosa R, Morjaria JB, Prosperini U, Busà B, Pennisi A, Malerba M, et al. COPD smokers who switched to e-cigarettes: health outcomes at 5-year follow up. Ther Adv Chronic Dis 2020;11:2040622320961617 [FREE Full text] [doi: 10.1177/2040622320961617] [Medline: 33101622]
48. Lucchiari C, Masiero M, Mazzocco K, Veronesi G, Maisonneuve P, Jemos C, et al. Benefits of e-cigarettes in smoking reduction and in pulmonary health among chronic smokers undergoing a lung cancer screening program at 6 months. Addict Behav 2020;103:106222 [FREE Full text] [doi: 10.1016/j.addbeh.2019.106222] [Medline: 31838445]
49. Laverty AA, Vardavas CI, Filippidis FT. Prevalence and reasons for use of heated tobacco products (HTP) in Europe: an analysis of eurobarometer data in 28 countries. Lancet Reg Health Eur 2021;8:100159 [FREE Full text] [doi: 10.1016/j.lanepe.2021.100159] [Medline: 34557853]
50. Skelton E, Guillaumier A, Tzelepis F, Walsberger S, Paul CL, Dunlop AJ, et al. Alcohol and other drug health-care providers and their client's perceptions of e-cigarette use, safety and harm reduction. Drug Alcohol Rev 2021;40(6):998-1002. [doi: 10.1111/dar.13276] [Medline: 33774886]
51. Xu SS, Meng G, Yan M, Gravely S, Quah ACK, Ouimet J, et al. Reasons for regularly using heated tobacco products among adult current and former smokers in Japan: finding from 2018 ITC Japan survey. Int J Environ Res Public Health 2020;17(21):8030 [FREE Full text] [doi: 10.3390/ijerph17218030] [Medline: 33142757]
52. Betsch C, Böhm R, Airhihenbuwa CO, Butler R, Chapman GB, Haase N, Nurm, et al. Improving medical decision making and health promotion through culture-sensitive health communication: an agenda for science and practice. Med Decis Making 2016;36(7):811-833. [doi: 10.1177/0272989X15600434] [Medline: 26296619]
53. Brick C, McCully SN, Updegraff JA, Ehret PJ, Areguin MA, Sherman DK. Impact of cultural exposure and message framing on oral health behavior: exploring the role of message memory. Med Decis Making 2016;36(7):834-843. [doi: 10.1177/0272989X15570114] [Medline: 25654986]
54. Gerritsen A, Jacobs M, Henselmans I, van Hattum J, Efficace F, Creemers G, et al. Developing a core set of patient-reported outcomes in pancreatic cancer: a delphi survey. Eur J Cancer 2016;57:68-77. [doi: 10.1016/j.ejca.2016.01.001] [Medline: 26886181]

XSL•FO
RenderX

Abbreviations

sf-TNP: smoke-free tobacco and/or nicotine product
TNP: tobacco and/or nicotine product
NRT: nicotine replacement therapy
FDA: Food and Drug Administration

Edited by A Mavragani; submitted 20.03.24; peer-reviewed by N Peraza; comments to author 12.09.24; revised version received 31.10.24; accepted 07.11.24; published 02.01.25.

Please cite as:
Law V, Afolalu EF, Abetz-Webb L, Wemyss LA, Turner A, Chrea C
International Expert Consensus on Relevant Health and Functioning Concepts to Assess in Users of Tobacco and Nicotine Products: Delphi Study
JMIR Form Res 2025;9:e58614
URL: https://formative.jmir.org/2025/1/e58614
doi:10.2196/58614
PMID:

©Vivienne Law, Esther F Afolalu, Linda Abetz-Webb, Lee Andrew Wemyss, Andrew Turner, Christelle Chrea. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development of an eHealth Intervention Including Self-Management for Reducing Sedentary Time in the Transition to Retirement: Participatory Design Study

Lisa Hultman[1], MSc; Caroline Eklund[1], PhD; Petra von Heideken Wågert[1], Prof Dr; Anne Söderlund[1], Prof Dr; Maria Lindén[2], Prof Dr; Magnus L Elfström[3], PhD

[1]Division of Physiotherapy, School of Health, Care and Social Welfare, Mälardalen University, Västerås/Eskilstuna, Sweden

[2]Division of Intelligent Future Technologies, School of Innovation, Design and Engineering, Mälardalen University, Västerås/Eskilstuna, Sweden

[3]Division of Psychology, School of Health, Care and Social Welfare, Mälardalen University, Västerås/Eskilstuna, Sweden

Corresponding Author:
Lisa Hultman, MSc
Division of Physiotherapy
School of Health, Care and Social Welfare
Mälardalen University
Box 883
Västerås/Eskilstuna, 721 23
Sweden
Phone: 46 21 48 06 901
Email: lisa.hultman@mdu.se

Abstract

Background: Having a great amount of sedentary time is common among older adults and increases with age. There is a strong need for tools to reduce sedentary time and promote adherence to reduced sedentary time, for which eHealth interventions have the potential to be useful. Interventions for reducing sedentary time in older adults have been found to be more effective when elements of self-management are included. When creating new eHealth interventions, accessibility and effectiveness can be increased by including end users as co-designers in the development process.

Objective: The aim was to explore the desired features of an eHealth intervention including self-management for reducing sedentary time and promoting adherence to reduced sedentary time in older adults transitioning from working life to retirement. Further, the aim was to develop a digital prototype of such an eHealth intervention.

Methods: The study used the participatory design approach to include end users, researchers, and a web designer as equal partners. Three workshops were conducted with 6 older adults transitioning to retirement, 2 researchers, and 1 web designer. Thematic analysis was used to analyze the data from the workshops.

Results: Participants expressed a desire for an easy-to-use eHealth intervention, which could be accessed from mobile phones, tablets, and computers, and could be individualized to the user. The most important features for reducing sedentary time were those involving finding joyful activities, setting goals, and getting information regarding reduced sedentary time. Participants expressed that the eHealth intervention would need to first provide the user with knowledge regarding sedentary time, then offer features for measuring sedentary time and for setting goals, and lastly provide support in finding joyful activities to perform in order to avoid being sedentary. According to the participants, an eHealth intervention including self-management for reducing sedentary time in older adults in the transition to retirement should be concise, accessible, and enjoyable. A digital prototype of such an eHealth intervention was developed.

Conclusions: The developed eHealth intervention including self-management for reducing sedentary time in older adults transitioning to retirement is intended to facilitate behavior change by encouraging the user to participate in autonomously motivated activities. It uses several behavior change techniques, such as goal setting and action planning through mental contrasting and implementation intention, as well as shaping knowledge. Its active components for reducing sedentary time can be explained using the integrated behavior change model. Further research is needed to evaluate the feasibility and effectiveness of the eHealth intervention.

(JMIR Form Res 2025;9:e63567) doi:10.2196/63567

XSL•FO
RenderX

KEYWORDS

behavior change intervention; adherence; integrated behavior change model; autonomous motivation; affective determinants

Introduction

Background

Increased sedentary time is common among older adults [1], and it increases the risk of noncommunicable diseases and premature death [2-4]. Sedentary time is the time spent in sedentary behavior for any duration or in any context [5]. Sedentary behavior can be defined as any waking behavior in a seated, reclined, or lying position with an energy expenditure of <1.5 metabolic equivalents (METs) [5]. With an aging population globally [6], there is an urgent need to reduce sedentary time in older adults. This has been found to be possible [7,8] but difficult, as sedentary behavior is complex and multifaceted [9]. In a recent scoping review [8], researchers found that sedentary time can be reduced in community-dwelling older adults using multicomponent interventions. Further, their results showed that information, goal setting, health coaching, self-monitoring, and action planning were included in effective interventions. However, studies reporting effect size reported only a mild to moderate reduction in sedentary time. Consequently, there is a need to develop new interventions with new approaches to reduce sedentary time in older adults.

The transition to retirement could be a favorable time for the introduction of interventions for reducing sedentary time in older adults. This is because retirement leads to changes in health behaviors such as physical activity, sedentary time, and smoking [10,11]. A recent longitudinal cohort study found that 61% of participants made health behavior changes when retiring [12], indicating a window where health behavior interventions, such as interventions for reduced sedentary time, could be timely. Previous research has found that interventions for health behavior change are more effective when elements of self-management are included [8]. Self-management is a person's own day-to-day management of health promotion or disease, including management of their medicines, emotions, meaningful behaviors, and roles [13].

Previous research has found that older adults often equate sedentary behavior with physical inactivity, and are unaware of the health consequences of prolonged sedentary time, how much time they spend in sedentary behavior, and when they are sedentary [9,14]. Affect has an impact on sedentary time in older adults, as enjoyment can be a motive for taking part in nonsedentary behavior [15] and sedentary behavior [9]. Older adults transitioning to retirement experience sedentary behavior as something related to health and well-being, giving them an opportunity to engage in meaningful activities, such as self-reflection, but also as something unhealthy and related to loneliness and boredom [16].

A recent study found that older adults in the transition to retirement use and desire self-management strategies that influence both affective determinants (related to emotions and affects such as joyfulness) and cognitive determinants (related to cognitive processes and includes attitudes and self-efficacy) for reducing sedentary time [17]. Earlier interventions for reducing sedentary time have primarily targeted cognitive determinants, resulting in positive changes with only a mild to moderate effect size [8]. Phipps et al [18] found that sedentary time is influenced more by explicit affective attitudes than by explicit instrumental attitudes. In recent years, interventions addressing affective determinants have been used in health promotion interventions targeting health behaviors [19,20]. Lithopoulos et al [21] compared affective, instrumental, and self-regulatory information in office workers and found that affective information had the most influence on reducing sedentary time. However, there are no intervention studies targeting both cognitive and affective determinants to reduce sedentary time in older adults transitioning to retirement.

To be effective, interventions must also have the possibility of reaching the intended end users. eHealth, which involves health services and information delivery through the internet and internet-related technologies [22], might offer an opportunity to reach out to large populations in upper middle– and high-income countries, as the rates of internet use and access in these countries are 83% and 93%, respectively [23]. eHealth interventions include but are not limited to mobile device–delivered health interventions (mHealth), text messaging, use of stationary computers and the internet, and telephone calls. A recent systematic review and meta-analysis of the effect of mobile health app interventions on health behaviors in older adults stated that such eHealth interventions can be used to reduce sedentary time, but there is uncertainty regarding the results owing to small sample sizes [24]. Adherence to eHealth interventions is slightly higher than adherence to traditional interventions, which might be due to eHealth interventions being perceived as more fun [25]. Adherence can be further raised through increased acceptability of the intervention by involving older adults in the development process using participatory research design [26]. Involving end users in the development of eHealth interventions is a common approach [27,28] and can increase the quality of the knowledge created, effectiveness of the intervention developed [28], and accessibility among those who are unaccustomed to eHealth [29].

Aim

The aim was to explore desired features of an eHealth intervention including self-management for reducing sedentary time and promoting adherence to reduced sedentary time in older adults transitioning from working life to retirement. Further, the aim was to develop a digital prototype of such an eHealth intervention.

Methods

Design

The study applied participatory design (PD) [26] to allow for the knowledge of older adults transitioning to retirement, researchers, and a web designer to be utilized and valued equally in the eHealth intervention's development. PD is a research method in which end users are involved as equal partners in the

XSL•FO
RenderX

design process [26] and has previously been used to develop eHealth interventions targeting various health behaviors [30-32]. This study is part of a PD project aimed at health promotion through reduced sedentary time in the transition to retirement. The PD project consists of the following 4 phases: need assessment, idea generation, testing and retesting, and evaluation. This study involves the second phase, idea generation, in which needs and opinions are evaluated in order to generate ideas and develop a digital prototype [26].

Participants

Participants, who were older adults transitioning to retirement, were recruited by the first author (LH) at theme days and meetings held by retirement organizations and municipalities, and through an interview article in the local newspaper. The criteria for inclusion were as follows: age 60-75 years, ability to speak and understand Swedish, ability to read and comprehend the study instructions, and plan to retire within a maximum of 3 years or having been retired for a maximum of 5 years. The exclusion criteria were as follows: presence of a serious disease, severe loss of vision, and severe loss of communicative ability. All recruited participants were newly retired, and none were still working. For further background information regarding the participants, see Tables 1 and 2.

Table 1. Participant (older adult) demographics.

Participant	Age, years	Sex	Time retired, years	Mostly sedentary or nonsedentary at the current or last job	Living with other adults	Type of housing	Self-reported daily sedentary time, hours	Previous use of an eHealth intervention including self-management
Participant 1: Older adult	67	Male	4	Sedentary	Yes	House/townhouse	7-8	No
Participant 2: Older adult	68	Female	3	Nonsedentary	No	Apartment	2	Do not know
Participant 3: Older adult	67	Male	2	Sedentary	No	Apartment	8	No
Participant 4: Older adult	67	Female	2	Nonsedentary	Yes	House/townhouse	5	App-based arthritis treatment
Participant 5: Older adult	70	Female	4	Nonsedentary	No	Apartment	6-8	—[a]
Participant 6: Older adult	64	Female	2	Sedentary	Yes	House/townhouse	5	Activity watch

[a]Not applicable.

Table 2. Participant (designer and researcher) demographics.

Participant	Age, years	Sex	Description
Participant 7: Researcher	38	Female	Physiotherapist (PhD in physiotherapy) with a research focus on eHealth and participatory design
Participant 8: Researcher	37	Female	Physiotherapist (master's degree in clinical medical science) with experience in primary and specialist health care
Participant 9: Web designer	39	Female	Adequate education and several years of experience in web development

A workshop group consisting of 9 participants was convened. The group's size was selected to be large enough to capture different user perspectives and small enough to allow all participants' voices to be heard. Of the 9 participants, 6 were older adults in the transition from working life to retirement, 2 were researchers participating in the workshop (LH and CE), and 1 was a web designer (Tables 1 and 2). The workshops were led by an experienced workshop moderator to improve the balance of power among participants. The workshop moderator was a female researcher employed by the same university as the authors.

Ethical Considerations

Prior to recruitment, the study was reviewed and approved by the Swedish Ethical Review Authority (Dnr 2019-03836). Informed consent was obtained from the participants.

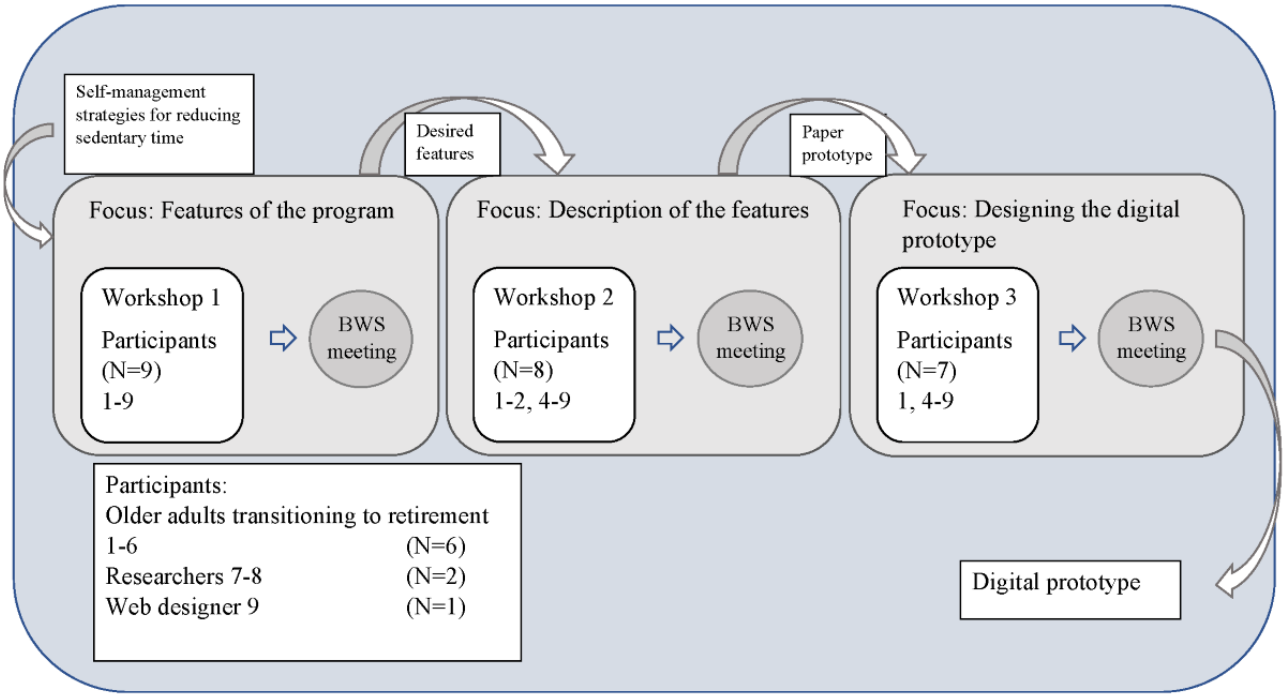
Workshops and Data Collection

Three workshops were conducted between November 2021 and January 2022 at time intervals of 2 and 8 weeks. All participants were invited to all 3 workshops (see Figure 1 for attendance). The workshops were held in a meeting room at a university. At the beginning of the first workshop, the participants were given written and verbal information about the study, and the study aim and the focus of each workshop were presented (Figure 1), as were the roles of the participants and the moderator. Further the participants were informed that their participation was

XSL•FO
RenderX

voluntary and that they could choose to end their participation at any time. Informed consent was signed. The first workshop was planned based on the results of a previous study [17]. In this workshop, focus group interviews were performed with older adults transitioning to retirement, and they involved perceptions of self-management strategies for reducing sedentary time and adhering to the reduced sedentary time. PD tool cards [33] were used at the first workshop, with participants being asked to list on Post-it notes which features in a digital solution they judged to be necessary to operationalize the self-management strategies for reducing sedentary time. This was followed by a discussion. Workshop 2 used 2D mapping [33], taking its starting point in a mind map of features developed from workshop 1. The third workshop used the PD tool user journey [33], with participants working based on a paper prototype developed after workshop 2. Participants were also asked to prioritize the importance of the features, voting for the first, second, and third most important features for reducing sedentary time. The voting was conducted using Post-it notes, with different colored notes used to represent the first, second, and third most important features. After the prioritization of the features, drafts of each feature of the paper prototype were presented, and suggestions for changes were discussed.

Figure 1. Workshop process map. BWS: between-workshop.



After each workshop, the first author (LH) went through the Post-it notes and photographs and listened to audio recordings to create mind maps and outline possible features of the eHealth intervention. Audio recordings were used at this stage to clarify and deepen the understanding of the text written on the Post-it notes. Between-workshop meetings were conducted with the 2 researchers participating in the workshop and 4 additional researchers (co-authors PHW, AS, ML, and MLE) with extensive experience in developing eHealth interventions, behavior change interventions, and gerontologic research. Preliminary results and proposed solutions were presented at the between-workshop meetings, and ideas and questions for the next workshop were discussed. After the last workshop, a digital prototype was developed.

Data Analysis

After the last workshop, the audio recordings from the workshops were transcribed verbatim. The transcripts were analyzed using inductive thematic analysis [34], with the analysis conducted in 2 steps based on the study aim and the focus of each workshop. The analyses of workshops 1 and 2 focused on desired features and their descriptions (feature analysis). This served as a verification of the paper prototype presented at workshop 3. The analysis of workshop 3 aimed to describe the necessary changes and the participants' desires regarding the final layout of the digital prototype (change analysis). The first author listened to the audio recordings and read the transcripts to get familiar with the dataset, after which relevant segments were coded at a semantic level. This was in accordance with phases 1 and 2 of the 6 phases of thematic analysis. In phase 3, initial themes were created. These were then checked against the whole data set, and meetings were held with all authors where the themes were discussed (phase 4). The themes were then defined and named (phase 5). Although writing had been carried out since phase 3, the analysis proceeded into the writing of the manuscript (phase 6) as themes were then clarified. The themes from the analysis can be found in Multimedia Appendix 1. The results from the prioritization of the features were summarized and used as a basis for discussion during workshop 3.

XSL•FO
RenderX

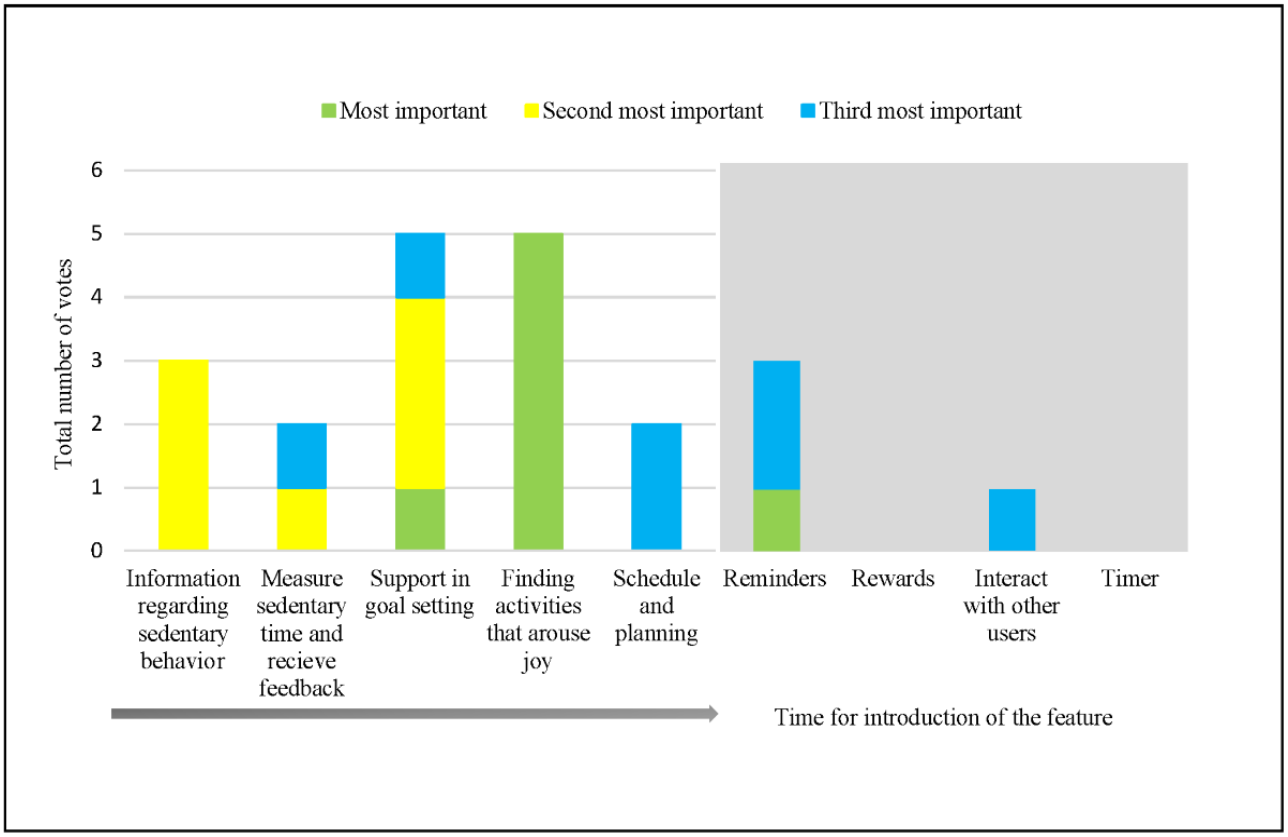
Results

Overview

The feature analysis resulted in the following 2 themes: general features and features for self-management. The latter consisted of 9 subthemes: finding activities that arouse joy, support in goal setting, information regarding sedentary behavior, measure sedentary time and receive feedback, schedule and planning, rewards, reminders, interact with other users, and timer. These subthemes corresponded to the features of the paper prototype presented at workshop 3. The change analysis resulted in 3 additional themes: evoke positive affect, less is more, and packaging.

The prioritization of the features revealed that the 3 most highly prioritized features in self-management for reducing sedentary time were finding activities that arouse joy, support for goal setting, and information regarding sedentary behavior. More information, such as the order in which participants wanted the features to be presented to the user, can be found in Figure 2.

Figure 2. Prioritization of features and time of introduction and inclusion of features in the paper prototype. The gray area represents functions not included in the paper prototype.



Feature Analysis

General Features

Participants wanted the eHealth intervention to include basic features, with the possibility to add more features for those who are more technically advanced.

> *I mean, there are so many people on their way into retirement and there are even more who are retired today who have a very low digitalization level... Yeah, it has to be incredibly simple. Has to be basal. Then you should be able to build on it, maybe, but... The foundation has to be based on something very, very simple.* [Participant in workshop 2]

To facilitate use, the user should be guided in the order in which the features should be used. To increase accessibility, the eHealth intervention should be available on mobile phones, tablets, and computers. There was concern regarding the inclusion of too many features, as this could lead to increased screen time rather than reduced sedentary time.

> *The more and more you build into it, it makes it so that, now I have to take a look.* [Participant in workshop 1]

To further facilitate use, an optional tutorial guiding the user through how the eHealth intervention worked was suggested.

Individualization according to the user's interests and affections, such as joy, was considered crucial for using the eHealth intervention, reducing sedentary time, and adhering to reduced sedentary time. At the same time, there was concern that individualization would result in the user having to answer too many questions at an early stage, which might exhaust them and prevent their further use of the eHealth intervention.

> *Yeah, it's... as soon as you have to go in and start making these kinds of choices that very many people stop.* [Participant in workshop 2]

XSL•FO
RenderX

Features for Self-Management

Finding Activities That Arouse Joy

To help them reduce their sedentary time, participants wanted the eHealth intervention to offer suggestions for joyful activities. Since what arouses feelings of joy is individual, participants saw a need to include different types of activities that could be performed in the home, outside the home, by oneself, or together with others, as part of the individualization. They expressed that the eHealth intervention should both allow the user to reflect on what activities they are interested in or could appreciate and provide suggestions for activities based on their areas of interest.

> *You get to answer questions about what you thought was interesting before and in that way be guided to... where you could look or what more you should do.* [Participant in workshop 2]

It should also provide information regarding what there is to do nearby when one is newly retired and facilitate contact with retirement organizations through links, information, and a contact person from the organization or activity. Participants expressed a wish for the activities offered by organizations to be presented in the eHealth intervention with their time and location, although they realized the difficulty that gathering this information would entail. Additionally, they wanted the eHealth intervention to include what activities other newly retired people do and enjoy, and provide the user with tips from other newly retired people for fun ways to reduce sedentary time. The difficulty in knowing what activities are available in the local area was expressed by 1 participant as follows:

> *I think a lot of people have problems with it... How in the world do I find things? What's here and where [is it], how do I get in contact with groups?* [Participant in workshop 1]

Support in Goal Setting

Participants considered goal setting to be important for reducing sedentary time, but described formulating goals as difficult. They suggested that it could be made easier by offering the user predetermined options or levels of sedentary time as goals, with these goals formulated based on either the maximum time to be spent in sedentary behavior or the number of activities to perform.

> *Can there be, like: Level 1, Level 2, Level 3, so you don't have to sit and make it up yourself, sort of?*
>
> *(You're thinking that there should be suggestions for goals?)*
>
> *Yeah, maybe if you're, you know, completely sedentary. What's Level 1. Yep. I finished that on the first day. Level 2, and so on.* [Participant in workshop 2]

Although goal setting was thought to be important for reducing sedentary time, a desire was expressed that this be made optional in the eHealth intervention. Participants expressed that, when formulating a goal regarding reduced sedentary time, they would like to be offered suggestions for what to do instead, and at follow-up, if the goal was not fully achieved, they would like the eHealth intervention to propose adapted suggestions to help them fulfill it in the future.

Information Regarding Sedentary Behavior

A need was expressed for information regarding sedentary behavior to be presented to the user at an early stage, so they would know why prolonged sedentary time should be avoided. In addition, participants wanted the information to be offered to the user both as optional short notifications and as static text to be found within the eHealth intervention. They expressed that they would like the information to include the benefits of not engaging in prolonged sedentary time, how long a person can stay sedentary before it is considered unhealthy, and tips on how sedentary time can be reduced. The participants stated that information must be formulated in a positive manner throughout the eHealth intervention. During one of the workshops, participants discussed information regarding sedentary behavior as follows:

> *That it's actually stated then that, where's this limit [for how long you can be sedentary]. That you get that information.* [Participant in workshop 2]

Measure Sedentary Time and Receive Feedback

To set realistic goals, it is necessary to first measure one's sedentary time. Further, participants saw a need to measure their sedentary time again at goal follow-up. This measurement could be done on specific days rather than daily to reduce the burden on the user. Different ways of taking the measurement were discussed, and while the participants saw the benefit of using sensors in an eHealth intervention, this would require users to always carry their smartphone, which was considered inconvenient. A watch or other device could be connected to the eHealth intervention, but it was thought that this would entail costs that would be unaffordable for some. Participants considered self-reporting one's sedentary time to be an appropriate solution. Suggested features for self-reporting sedentary time included tapping a stopwatch in the eHealth intervention when one sat down and tapping it again when one stood up, or estimating the time one spent in sedentary behavior. Using a stopwatch was considered unfavorable due to the risk of forgetting or tapping it in the wrong way, and participants did not want to have to carry their phone with them and tap it all day. The estimate of the time one spent in sedentary behavior could be given as an average for a typical day or as an estimate of the current day given in the evening. Participants expressed that after the user had measured their sedentary time, they would need feedback from the eHealth intervention to know whether they had spent too much time in sedentary behavior.

> *Then you get statistics on how you've been sitting. How long you've been sitting and how it's affected your body... Then you learn, okay, now I have to change my habits, if you're given information.* [Participant in workshop 2]

Schedule and Planning

Participants wanted a schedule or calendar to support their planning and facilitate new routines. It was important for the schedule to be easy to handle, with just a few tabs. It would need to be close at hand. For some, this meant being in their

XSL•FO
RenderX

phone, while for others, it meant a printed copy to put on their fridge. Participants said they would like a connection between the proposed activities and the schedule, so that activities could easily be placed in the schedule from the activity menu. For follow-up, the participants wanted the opportunity to check off those activities they had performed as planned.

> *I think this, that it's connected to the schedule, is quite good... because if you have to physically enter it in yourself. Then it's. It's a form of resistance and an extra step.* [Participant in workshop 2]

Rewards

Participants wanted the eHealth intervention to include rewards at goal fulfillment and as notifications in order to increase their motivation and adherence, and considered it important that the rewards lead to feelings of joy. Which rewards led to joy was said to be highly individual; for instance, some said they would like to have virtual stars and trophies in the eHealth intervention, while others suggested gift cards in line with the user's interests. To solve the problem of adapting to these differences, participants suggested that the eHealth intervention should urge users to treat themselves with a reward, with the possibility for users to specify in advance what this treat should be. While there was concern that this treat would be something that was unhealthy for the user, it was also discussed that engaging in a fun or amusing activity was rewarding and something to look forward to.

> *…also has to be something that's fun that you can look forward to. You can reward yourself.* [Participant in workshop 1]

Reminders

Participants considered reminders to be important for adherence to new routines and said they would like to have reminders about their use of the eHealth intervention, goals, and planned activities. It could be enough to get a reminder in the morning, but as the timing of when reminders are needed is individual, the participants wanted the user to be able to select when the eHealth intervention issued reminders. It was important for the reminders to be phrased positively and give joy, as it was expressed that positively phrased prompts lead to action and adherence more often than other prompts.

> *Support that's encouraging is very... We don't shut it off in the same way as support that's maybe not encouraging.*
>
> *That gives [us] a bad conscience instead.* [Participant in workshop 1]

As what gives joy is also individual, a reminder regarding a planned activity that brings joy could be a way to achieve individualization.

Interacting With Other Users

Participants expressed that they would like to have features that provide the opportunity to interact with others to reduce their sedentary time. They wanted the opportunity to share suggestions for activities and tips for reducing sedentary time and to see other users' suggestions. Moreover, they wanted users to be able to give each other positive feedback to increase their motivation.

> *...then you're a group, aren't you, who sort of ehh can cheer each other on and say good luck and well done and all that.* [Participant in workshop 2]

A chatroom was discussed as a possible way to facilitate interaction between users. There were concerns regarding the social climate in digital chatrooms. Several suggestions for how to improve it were discussed, and the participants preferred a solution in which users could only comment on each other's activity suggestions with predetermined answers or positive emojis. This interaction could also serve as an opportunity to initiate meeting up with other users with similar interests, but arranging meetings with unknown users also led to safety concerns.

> *Then there has to be some kind of block in it so you don't meet some strange person just to... I was going to say... but I'm just thinking that you can include the security aspect in it.* [Participant in workshop 1]

Timer

Participants wanted a timer with an alarm set to 30 minutes to avoid prolonged sedentary time, until the new behavior had become a habit. It was proposed that the timer could be started when the user sat down or could be set in advance to go off every 30 minutes during a selected period.

> *I might need ehh a pling, like: whoops. Now it's been 30 minutes.* [Participant in workshop 1]
>
> *...when you describe how you live. I don't live half as physically active... I live more "small". Thirty minutes, fine, and then I leave.* [Participant in workshop 2]

It was also suggested that, rather than simply sounding an alarm when the timer finished, the eHealth intervention could suggest joyful activities that were prespecified by the user.

Paper Prototype

At the third workshop, a paper prototype of an eHealth intervention including self-management for reducing sedentary time through joyful activities was presented. The prototype was planned as a website solution with important functions as well as information and links to websites and mobile apps. The prototype was designed to introduce the user to the features in the order chosen by the participants during the workshops. The user was first shown information regarding sedentary behavior, then encountered features for measuring sedentary time, and finally received support for goal setting, the selection of joyful activities, and the planning feature (Figure 2).

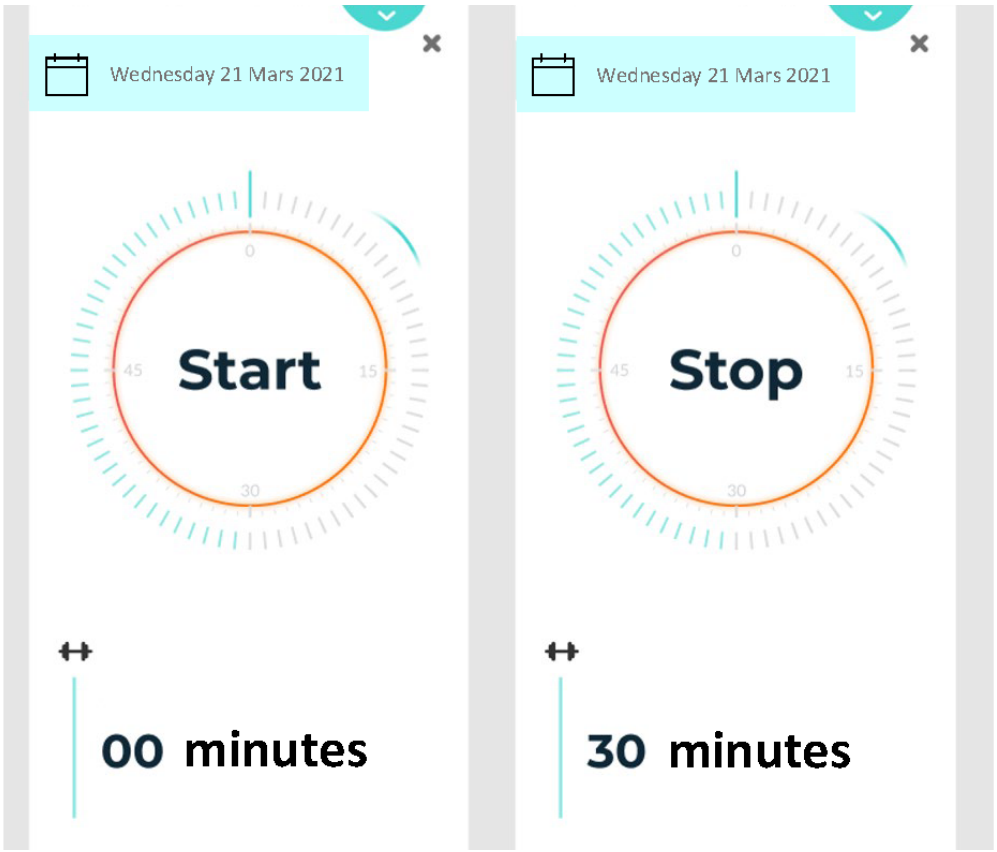
Brief paragraphs regarding sedentary behavior were presented, which could be shown to the user when they started using the eHealth intervention and could be found in the eHealth intervention at any time if the user wanted to read all the information.

Two tools for measuring sedentary time were presented. One was a self-assessment tool that asked the user to assess how many hours they had been performing different sedentary

XSL•FO
RenderX

behaviors during the day, such as eating, working, or commuting, and the other was a stopwatch for the user to start when they began being sedentary (Figure 3). Alongside the tools for measuring sedentary time, the user would receive adapted feedback based on different time intervals of sedentary time.

Figure 3. One of the tools presented at workshop 3 for measuring sedentary time.

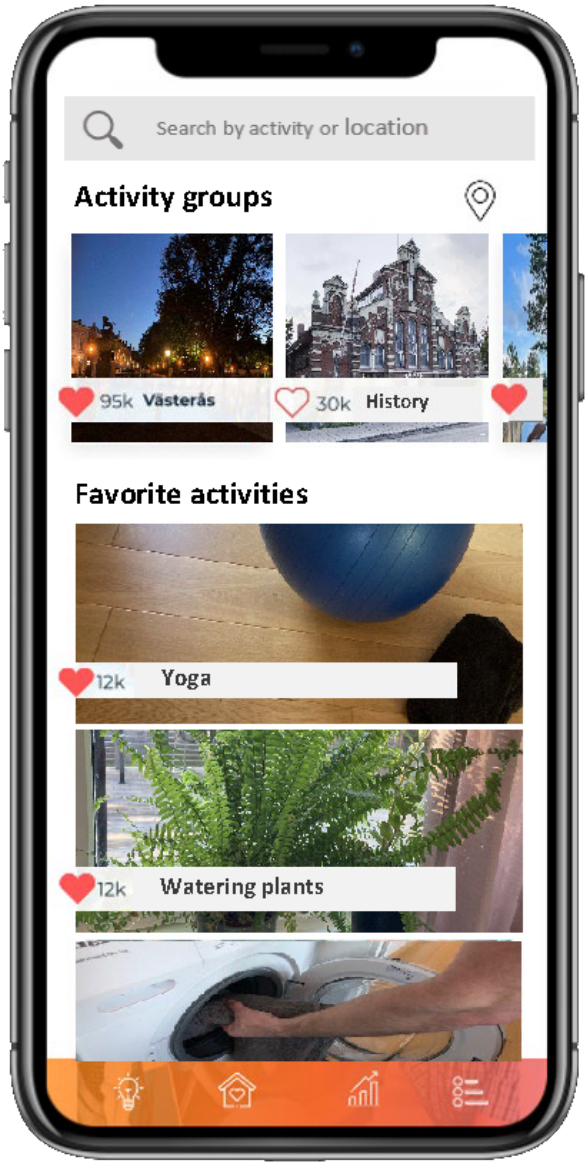


The prototype included the design of a support for goal setting. The user was given a written prompt asking them to formulate a goal regarding their sedentary time, and they received support for how this goal should be formulated, with examples of goals shown to them. They were then asked to focus on how they would feel if their sedentary time was reduced and what obstacles they might face, and to make a plan for how to overcome these obstacles. When the goal was then evaluated at follow-up, the user was asked to choose whether it was far from being achieved, nearly achieved, or fully achieved, and then received customized feedback.

The feature providing support for finding activities was presented to participants as a scroll list with activities shown along with a brief description and a photo (Figure 4). The scroll list could be sorted based on interests, local areas, and activities performed with others or by oneself. Tips on how to reduce sedentary time were offered as a bulleted list.

XSL•FO
RenderX

Figure 4. Feature for finding activities that arouse joy from the paper prototype presented at workshop 3.



The schedule was presented both as a colorful printable schedule designed to evoke positive and cheerful emotions, and as a copy of the schedule from an iPhone and iPad. The user could write their planned activities digitally or manually on the printed schedule. The copy of the schedule from the iPhone and iPad was used to discuss whether the calendar found in a user's own phone could be used as a schedule.

In the prototype, rewards were delivered at goal evaluation as brief positively formulated phrases adapted to the user's goal achievement. As the prototype was designed as a website, the possibility to include reminders and interactions with other users was limited. It was suggested that the user make notes regarding planned activities and scheduled follow-ups, preferably in a calendar with the possibility to send notifications. No timer function was included in the prototype; instead, the suggestion to set an alarm when one was to be seated for a prolonged time was included among the tips for reducing sedentary time.

Change Analysis

Evoke Positive Affect

Participants stated that the eHealth intervention's first impression and introduction should arouse interest to facilitate further use.

> *If nobody opens this. Then nothing's worth anything. I mean, there has to be something that captures your interest.* [Participant in workshop 3]

They suggested that this could be done by having celebrities or cartoons convey the information or by having the user answer questions regarding sedentary time. While the participants considered it important to arouse interest in the eHealth intervention's introduction, when it came to the delivery of information, they expressed that credibility was important.

> *Isn't it two different things we're talking about... One's about capturing attention. Come look at our website. [mm] There you could have animated figures, I'm thinking... But when I open the website, then I'd*

XSL•FO
RenderX

> *really like to have, to know that there's evidence behind this in some way, I mean. Ehh am I thinking about it the wrong way?* [Participant in workshop 3]

They suggested that, to achieve credibility, the information could be presented by a physician or physiotherapist, or by people who were newly retired and could share their experiences. It was stated that all information in the eHealth intervention should be more positively phrased and that when information was evidence-based, this should be clearly stated.

Less is More

Overall, the participants wanted the text and information in the paper prototype to be shortened and simplified. The group agreed to choose the tool presented for the self-assessment of sedentary time, as it was considered easier to use than a stopwatch. However, there were concerns regarding whether people who spend a prolonged time in sedentary behavior would take the time to fill in such a form and the fact that it is difficult to make such an assessment. As a solution, it was suggested that the user could fill in an estimate of their total time in sedentary behavior or enter data from an activity watch if they used one. The participants pointed out the need to adapt the assessment form to better suit those who are retired, suggesting that the activities mentioned on the form be replaced with time intervals as every day is different for a person who is newly retired compared to when they had a job. They also suggested that the feedback be changed to give the user a clear overview of when prolonged sedentary behavior becomes unhealthy and whether there are different levels when it comes to harmful amounts of sedentary time, and if so, how close they are to moving on to the next level.

The content of the support for goal setting was considered helpful, but the layout was perceived as cumbersome and too extensive. It was suggested that the text be shortened and reformulated to use more accessible language.

> *I hadn't finished reading... There's too much text.* [Participant in workshop 3]

The participants pointed out the importance of letting the user know that they do not need to make massive changes to reduce their sedentary time and that small changes and activities in the home could suffice.

> *It's also so important that there's information about that the small things in daily life is enough.* [Participant in workshop 3]

They suggested that this could be done by sorting activities as "activities in the home" or "start small." The group was positive about simplifying the eHealth intervention by including links to organizations rather than giving information on each activity offered by the organizations.

The participants considered the rewards offered in the prototype to be sufficient. As the prototype was designed to encourage the user to engage in activities that bring positive affect and emotions, and to remind them of their feelings at goal fulfillment, the participants did not feel further rewards were needed.

The participants mentioned wanting the opportunity to interact with other users within the eHealth intervention. However, at this point, they agreed that the possibility to find activities with others in terms of links and information regarding retirement organizations and local events was sufficient for interacting with others.

> *Like we've talked about before, it was more this thing of being able to interact with other users. I mean, it's...* [not possible now]
>
> *We've dropped that a bit.*
>
> *But then you can find the organizations, you find them via this.* [Participant in workshop 3]

To keep the features included in the eHealth intervention to a minimum, a timer was not included; instead, among the tips provided, the user would be given a recommendation to use a timer in other mobile apps or use the timer feature on their phone. The participants discussed the notion that a timer might be useful for those who prefer not to enroll in planned activities with others and would rather perform activities in the home.

At the third workshop, participants did not express a need for a calendar or schedule in the eHealth intervention, as they preferred to write down their planned activities in the calendar on their phones rather than in the calendar suggested at the workshop. They felt that those who do not use the calendar on their phones would probably not use a digital eHealth intervention for planning but likely already use a paper calendar. Therefore, it was suggested that the eHealth intervention should simply suggest that the user write down their planning so that they could choose the type of planning tool they prefer. This was expressed as follows:

> *I mean, maybe it's a bit much for those who don't have it in their phone. I was thinking of this... on the website.*
>
> *You don't think they have a calendar on the fridge?*
>
> *I don't think they go to their computer and [fill it in].* [Participant in workshop 3]

Packaging

The participants agreed that the eHealth intervention should be delivered as a website rather than a mobile app, as it was important that the intervention be accessible from a computer so it would be easier to operate. However, choosing a website rather than a mobile app limited the possibility to include certain features. For instance, the participants would have liked the eHealth intervention to send reminders to promote adherence, and they suggested that the calendar function on the user's phone could be used to remind them of planned activities (although they still expressed a desire for the possibility to send a daily reminder or a reminder if the eHealth intervention was not being used as planned). They also wanted to be able to log in to the eHealth intervention in order to save their goals and interests digitally to facilitate the use of the eHealth intervention and for goal follow-up.

Participants expressed a wish for information to be delivered in different ways, allowing the user to listen to or read information, or sometimes get it as a video. It was suggested

XSL•FO
RenderX

that, during goal setting, the user should be able to listen to the information, as participants experienced that the spoken word was more appealing and evoked more emotions in this context.

> *Yeah, I thought about it... that you said it in a much more appealing way than the text does. When you said "Think about what you want to achieve. Imagine how...", then you get a feeling. I don't get that here [shows the paper].* [Participant in workshop 3]

Discussion

Principal Findings

To our knowledge, this is the first study to develop a self-management eHealth intervention for reducing sedentary time and promoting adherence to reduced sedentary time, which has been created for and with older adults in the transition to retirement. The study used PD [26] to include end users, researchers, and a web designer as mutual partners in the development process. The study has contributed new knowledge regarding features of this type of eHealth intervention and the timing of these features. The most important features were those involving finding activities that arouse joy, having support in goal setting, and getting information regarding sedentary behavior. The workshop participants regarded the order in which the included features are presented to the user as important: The user first receives knowledge regarding sedentary behavior, then encounters features for measuring sedentary time and for support in setting a goal, and lastly gets support in finding joyful activities to perform in order to avoid being sedentary. The results highlight the need for such an eHealth intervention to be concise, accessible, and enjoyable to suit its end users.

Comparison With Prior Work

This study differs from previous research [7,8] as it specifically targets older adults in the transition to retirement and involves them as co-designers in the development of an eHealth intervention including self-management for reducing sedentary time and promoting adherence to reduced sedentary time.

Participants perceived it as important for the eHealth intervention including self-management to use positive phrasing to increase both its use and users' adherence to reduced sedentary time, expressing that positively phrased prompts lead to adherence more than negatively phrased ones. Previous researchers have found positive affect to be connected to the use of eHealth interventions to increase adherence in older adults [25] and positively phrased prompts to be more effective in older adults [35]. Overall, this indicates the importance of an eHealth intervention that evokes positive affect to promote adherence to reduced sedentary time when transitioning to retirement.

The features considered to be the most important for reducing sedentary time in the transition to retirement were those involving finding joyful activities. The importance of joyfulness in reducing sedentary time and increasing physical activity in older adults has been found in previous research [15,36]. In this study, participants stated the importance of receiving suggestions for what to replace sedentary time with, rather than just getting a prompt to break their sedentary behavior. The results further highlight the need for awareness of the diversity in the activities that people consider joyful and for the inclusion of information on simple joyful activities that can be performed in one's own home to reduce sedentary time.

The eHealth intervention developed in this study includes features for information, measurement, goal setting, and planning, which have been used previously to reduce sedentary time in older adults [7,8] and have been found to be acceptable behavior change techniques for older adults to reduce their sedentary time [9]. The results of this study show that these behavior change techniques are also desired specifically when it comes to an eHealth intervention including self-management for reducing sedentary time, which targets older adults in the transition to retirement.

During the third workshop, self-assessment was chosen as the method for measuring one's sedentary time. As objectively measured sedentary time is more accurate than self-assessed sedentary time (total sedentary time is often underestimated [37]), the inclusion of objectively measured sedentary time might provide the user with more reliable results. However, doing so would mean adding a device, such as a smartwatch, and this would entail a cost to the user or require them to always carry their phone, which the participants considered inconvenient. The included feature for self-assessment also provides the user with more contextual information, which is an advantage of using self-assessed sedentary time [38].

Among the included features are behavior change techniques that target both affective determinants (eg, affective attitudes) [39] and cognitive determinants (eg, cognitive attitudes and self-efficacy) [40] for reduced sedentary time. The feature for goal setting and planning in the developed eHealth intervention is based on mental contrasting and implementation intention (MCII) [41], which has previously been found to be effective in health behavior change [42]. The inclusion of MCII also responds to the participants' wish to include goal setting and planning as well as an opportunity to consider activities to replace their sedentary time, which was discussed during the workshops. In the developed eHealth intervention, the user is asked to consider a positive affective outcome, which has been used with success in previous research targeting health behaviors [20]. The inclusion of a feature targeting affective determinants when reducing sedentary time in older adults transitioning to retirement can be an effective approach, as previous research has found that sedentary behavior is largely an affect-driven behavior [18]. Further, Lithopoulos et al [21] found that an intervention targeting affective determinants for reducing sedentary time in the office was effective in the short term.

Previous research has described sedentary behavior as a habitual behavior performed in unawareness [9]. This speaks for the inclusion of features, such as a timer, to increase users' awareness of when they are in prolonged sedentary behavior and thereby promote their adherence to reduced sedentary time. However, Suorsa et al [43] found that an activity tracker was not sufficient for reducing sedentary time in retirees. In this study, none of the participants rated a timer feature as important, and it was not included in either the paper or the digital prototype. However, among the tips the eHealth intervention

XSL•FO
RenderX

offered for reducing one's sedentary time was the suggestion to set a timer when engaged in sedentary behaviors.

The approach of the eHealth intervention for achieving behavior change, which was developed here, can be understood based on the integrated behavior change (IBC) model for health behavior [44]. The eHealth intervention is designed to guide the user in reducing their sedentary time by finding activities that they are autonomously motivated to perform, meaning that the intention to perform a health behavior is dependent on whether the behavior evokes positive affect and is in line with a person's interests and true self [45,46]. Further, the developed eHealth intervention includes information, self-monitoring, and feedback on behavior that can influence users' attitudes and perceived behavioral control regarding the behavior change [40,47]. Lastly, the feature for goal setting includes affective MCII, which can influence behavior through implicit pathways [48,49], as well as elements of action planning, which are included in the IBC model. The use of the IBC model can offer an understanding of how the developed eHealth intervention including self-management is intended to reduce sedentary time and promote adherence to reduced sedentary time, but further studies are needed in order to evaluate the eHealth intervention's feasibility and effectiveness.

Limitations and Strengths

Carefully considered choices were made to increase the study's trustworthiness (credibility, dependability, and transferability) [50]. The study's methodological approach is a strength, as it enabled the inclusion of older adults in the transition to retirement as co-designers. They are the only ones who hold the tacit knowledge that is needed to develop an eHealth intervention in line with their needs and preferences. The participants also included a web designer and 2 researchers, who contributed knowledge regarding web development and health behavior change. This combination of participants may contribute to the development of an effective and acceptable eHealth intervention as both tacit and explicit knowledge are included in the development process, which strengthens the study's credibility.

The credibility was further strengthened by a workshop leader guiding the workshops to ensure that everyone had the opportunity to speak and to avoid placing additional power with the researchers. Moreover, the workshops started with an icebreaker, and participants were offered coffee and fruits during the meetings to create a pleasant atmosphere. Three PD tool cards, 2D mapping, and a user journey tool were used to encourage active participation among the participants. These tools were chosen because they have been used successfully in PD research [33], and they ensured that all participants were encouraged and had the possibility to be active during the workshops. The recruitment of participants at theme days and retirement organization meetings might have led to the inclusion of participants with a greater interest in such activities than the population, possibly affecting the study's credibility and transferability. This was compensated by the recruitment of participants through a local newspaper. Additionally, the participants expressed their differences through participation and interest in activities held by organizations during the workshops. The older adults in the transition to retirement who participated in this study lived in different types of housing and reported different levels of daily sedentary time as well as levels of sedentary time at their latest jobs, which indicates a spread among them and thus increases credibility. The study's confirmability was strengthened as, during the third workshop, the participants were given the opportunity to discuss and further develop the paper prototype developed from workshops 1 and 2. Further, the feature analysis was used as a verification of the paper prototype to ensure that no important data had been excluded.

Conclusion

An eHealth intervention including self-management for reducing sedentary time and promoting adherence to reduced sedentary time in older adults transitioning to retirement was developed in a co-design approach involving representatives of end users, a web designer, and researchers. Participants desired features for providing information regarding sedentary behavior, measuring sedentary time, goal setting, finding joyful activities, and making plans and a schedule, and expressed that these should be presented to the user in this order. The most important features for reducing sedentary time were said to be those involving finding joyful activities. Further, participants expressed a desire for such an eHealth intervention to be concise, accessible, and enjoyable in order to be suitable for older adults transitioning to retirement. The developed eHealth intervention strives to support behavior change based on the IBC model by focusing on autonomously motivated activities and affective and cognitive determinants through implicit and explicit pathways, and includes action planning. It will be further tested for feasibility and effectiveness in upcoming studies.

Acknowledgments

This study was funded by the Swedish Research Council Dnr (2018-02928). We acknowledge Liv Nordström for being the workshop leader during the 3 workshops and Bluearth Fondation for sharing the sedentary time calculator for use during workshop 3.

Authors' Contributions

All authors (LH, CE, PHW, AS, ML, and MLE) contributed to the conception and design of the study. LH and CE participated during the workshops. All authors (LH, CE, PHW, AS, ML, and MLE) participated during the between-workshop meeting. LH analyzed the data and wrote the main content in the manuscript. CE, PHW, AS, ML, and MLE critically revised the manuscript and contributed intellectual content. All authors (LH, CE, PHW, AS, ML, and MLE) read and approved the final manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Theme summary.
[DOCX File , 18 KB - formative_v9i1e63567_app1.docx]

References

1. Harvey JA, Chastin SF, Skelton DA. How sedentary are older people? A systematic review of the amount of sedentary behavior. J Aging Phys Act 2015 Jul;23(3):471-487. [doi: 10.1123/japa.2014-0164] [Medline: 25387160]
2. Biswas A, Oh PI, Faulkner GE, Bajaj RR, Silver MA, Mitchell MS, et al. Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults: a systematic review and meta-analysis. Ann Intern Med 2015 Jan 20;162(2):123-132. [doi: 10.7326/M14-1651] [Medline: 25599350]
3. Rossen J, Von Rosen P, Johansson UB, Brismar K, Hagströmer M. Associations of physical activity and sedentary behavior with cardiometabolic biomarkers in prediabetes and type 2 diabetes: a compositional data analysis. Phys Sportsmed 2020 May;48(2):222-228. [doi: 10.1080/00913847.2019.1684811] [Medline: 31663410]
4. Rojer AGM, Ramsey KA, Trappenburg MC, van Rijssen NM, Otten RHJ, Heymans MW, et al. Instrumented measures of sedentary behaviour and physical activity are associated with mortality in community-dwelling older adults: A systematic review, meta-analysis and meta-regression analysis. Ageing Res Rev 2020 Aug;61:101061 [FREE Full text] [doi: 10.1016/j.arr.2020.101061] [Medline: 32360669]
5. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, SBRN Terminology Consensus Project Participants. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. Int J Behav Nutr Phys Act 2017 Jun 10;14(1):75 [FREE Full text] [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]
6. Beard JR, Officer AM, Cassels AK. The world report on ageing and health. Gerontologist 2016 Apr;56 Suppl 2:S163-S166. [doi: 10.1093/geront/gnw037] [Medline: 26994257]
7. Chastin S, Gardiner PA, Harvey JA, Leask CF, Jerez-Roig J, Rosenberg D, et al. Interventions for reducing sedentary behaviour in community-dwelling older adults. Cochrane Database Syst Rev 2021 Jun 25;6(6):CD012784 [FREE Full text] [doi: 10.1002/14651858.CD012784.pub2] [Medline: 34169503]
8. Petrusevski C, Choo S, Wilson M, MacDermid J, Richardson J. Interventions to address sedentary behaviour for older adults: a scoping review. Disabil Rehabil 2021 Oct;43(21):3090-3101. [doi: 10.1080/09638288.2020.1725156] [Medline: 32058808]
9. Compernolle S, De Cocker K, Cardon G, De Bourdeaudhuij I, Van Dyck D. Older adults' perceptions of sedentary behavior: a systematic review and thematic synthesis of qualitative studies. Gerontologist 2020 Nov 23;60(8):572-582. [doi: 10.1093/geront/gnz127] [Medline: 31670766]
10. Ding D, Grunseit AC, Chau JY, Vo K, Byles J, Bauman AE. Retirement-a transition to a healthier lifestyle?: Evidence from a large Australian study. Am J Prev Med 2016 Aug;51(2):170-178. [doi: 10.1016/j.amepre.2016.01.019] [Medline: 26972491]
11. Vigezzi GP, Gaetti G, Gianfredi V, Frascella B, Gentile L, d'Errico A, Italian Working Group on RetirementHealth. Transition to retirement impact on health and lifestyle habits: analysis from a nationwide Italian cohort. BMC Public Health 2021 Sep 14;21(1):1670 [FREE Full text] [doi: 10.1186/s12889-021-11670-3] [Medline: 34521363]
12. Robinson SM, Westbury LD, Ward K, Syddall H, Cooper R, Cooper C, et al. Is lifestyle change around retirement associated with better physical performance in older age?: insights from a longitudinal cohort. Eur J Ageing 2021 Dec;18(4):513-521 [FREE Full text] [doi: 10.1007/s10433-021-00607-9] [Medline: 34786013]
13. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
14. Ramalho A, Petrica J. The quiet epidemic: An overview of emerging qualitative research trends on sedentary behavior in aging populations. Healthcare (Basel) 2023 Aug 06;11(15):2215 [FREE Full text] [doi: 10.3390/healthcare11152215] [Medline: 37570455]
15. McGowan LJ, Powell R, French DP. How acceptable is reducing sedentary behavior to older adults? perceptions and experiences across diverse socioeconomic areas. J Aging Phys Act 2019 Sep 01;27(5):642–652 [FREE Full text] [doi: 10.1123/japa.2018-0274] [Medline: 30676195]
16. Eklund C, Elfström ML, von Heideken Wågert P, Söderlund A, Gustavsson C, Cederbom S, et al. The meaning of sedentary behavior as experienced by people in the transition from working life to retirement: an empirical phenomenological study. Phys Ther 2021 Aug 01;101(8):pzab117 [FREE Full text] [doi: 10.1093/ptj/pzab117] [Medline: 33951141]
17. Hultman L, Eklund C, von Heideken Wågert P, Söderlund A, Elfström ML. It has to be my way-Reducing sedentary time in the transition to retirement. J Aging Phys Act 2024 Nov 20:1-10. [doi: 10.1123/japa.2023-0414] [Medline: 39566479]

XSL•FO
RenderX

18. Phipps DJ, Rhodes RE, Jenkins K, Hannan TE, Browning NG, Hamilton K. A dual process model of affective and instrumental implicit attitude, self-monitoring, and sedentary behavior. Psychology of Sport and Exercise 2022 Sep;62:102222. [doi: 10.1016/j.psychsport.2022.102222]
19. Forster AS, Buykx P, Martin N, Sadler S, Southgate B, Rockliffe L, et al. Using affective judgement to increase physical activity in British adults. Health Promot Int 2018 Aug 01;33(4):648-656 [FREE Full text] [doi: 10.1093/heapro/dax004] [Medline: 28334769]
20. Ruissen GR, Rhodes RE, Crocker PRE, Beauchamp MR. Affective mental contrasting to enhance physical activity: A randomized controlled trial. Health Psychol 2018 Jan;37(1):51-60. [doi: 10.1037/hea0000551] [Medline: 28981303]
21. Lithopoulos A, Kaushal N, Beauchamp MR, Conner M, de Bruijn GJ, Latimer-Cheung AE, et al. Changing sedentary behavior in the office: a randomised controlled trial comparing the effect of affective, instrumental, and self-regulatory messaging on sitting. Appl Psychol Health Well Being 2020 Nov;12(3):687-702 [FREE Full text] [doi: 10.1111/aphw.12202] [Medline: 32428341]
22. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
23. Internet use. International Telecommunication Union. URL: https://www.itu.int/itu-d/reports/statistics/2023/10/10/ff23-internet-use/#footnote1 [accessed 2024-02-12]
24. Yerrakalva D, Yerrakalva D, Hajna S, Griffin S. Effects of mobile health app interventions on sedentary time, physical activity, and fitness in older adults: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e14343 [FREE Full text] [doi: 10.2196/14343] [Medline: 31778121]
25. Valenzuela T, Okubo Y, Woodbury A, Lord SR, Delbaere K. Adherence to technology-based exercise programs in older adults: a systematic review. J Geriatr Phys Ther 2018;41(1):49-61. [doi: 10.1519/JPT.0000000000000095] [Medline: 27362526]
26. Clemensen J, Rothmann MJ, Smith AC, Caffery LJ, Danbjorg DB. Participatory design methods in telemedicine research. J Telemed Telecare 2017 Oct;23(9):780-785. [doi: 10.1177/1357633X16686747] [Medline: 28027678]
27. Wannheden C, Revenäs Å. How people with Parkinson's disease and health care professionals wish to partner in care using eHealth: co-design study. J Med Internet Res 2020 Sep 21;22(9):e19195 [FREE Full text] [doi: 10.2196/19195] [Medline: 32955448]
28. Moore G, Wilding H, Gray K, Castle D. Participatory methods to engage health service users in the development of electronic health resources: systematic review. J Particip Med 2019 Feb 22;11(1):e11474 [FREE Full text] [doi: 10.2196/11474] [Medline: 33055069]
29. Revenäs Å, Ström L, Cicchetti A, Ehn M. Toward digital inclusion of older adults in e-health: a case study on support for physical activity. Univ Access Inf Soc 2023:1049. [doi: 10.1007/s10209-023-01049-z]
30. Revenäs Å, Opava CH, Martin C, Demmelmaier I, Keller C, Åsenlöf P. Development of a web-based and mobile app to support physical activity in individuals with rheumatoid arthritis: results from the second step of a co-design process. JMIR Res Protoc 2015 Feb 09;4(1):e22 [FREE Full text] [doi: 10.2196/resprot.3795] [Medline: 25665589]
31. Bartlett YK, Selby DL, Newsham A, Keding A, Forman D, Brown J, et al. Developing a useful, user-friendly website for cancer patient follow-up: users' perspectives on ease of access and usefulness. Eur J Cancer Care (Engl) 2012 Nov;21(6):747-757. [doi: 10.1111/j.1365-2354.2012.01357.x] [Medline: 22533456]
32. Østervang C, Lassen A, Schmidt T, Coyne E, Dieperink K, Jensen C. Development of a health information system to promote emergency care pathways: A participatory design study. Digit Health 2022;8:20552076221145856 [FREE Full text] [doi: 10.1177/20552076221145856] [Medline: 36601282]
33. Vandekerckhove P, de Mul M, Bramer WM, de Bont AA. Generative participatory design methodology to develop electronic health interventions: systematic literature review. J Med Internet Res 2020 Apr 27;22(4):e13780 [FREE Full text] [doi: 10.2196/13780] [Medline: 32338617]
34. Braun V, Clarke V. Thematic Analysis: A Practical Guide. Thousand Oaks, CA: SAGE Publications Ltd; 2021.
35. Notthoff N, Carstensen LL. Positive messaging promotes walking in older adults. Psychol Aging 2014 Jun;29(2):329-341 [FREE Full text] [doi: 10.1037/a0036748] [Medline: 24956001]
36. Ioannou E, Chen HL, Bromley V, Fosker S, Ali K, Fernando A, et al. The key values and factors identified by older adults to promote physical activity and reduce sedentary behaviour using co-production approaches: a scoping review. BMC Geriatr 2023 Jun 16;23(1):371 [FREE Full text] [doi: 10.1186/s12877-023-04005-x] [Medline: 37328734]
37. Prince SA, Cardilli L, Reed JL, Saunders TJ, Kite C, Douillette K, et al. A comparison of self-reported and device measured sedentary behaviour in adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2020 Mar 04;17(1):31 [FREE Full text] [doi: 10.1186/s12966-020-00938-3] [Medline: 32131845]
38. Aunger J, Wagnild J. Objective and subjective measurement of sedentary behavior in human adults: A toolkit. Am J Hum Biol 2022 Jan;34(1):e23546 [FREE Full text] [doi: 10.1002/ajhb.23546] [Medline: 33277954]
39. Williams DM, Evans DR. Current emotion research in health behavior science. Emotion Review 2014 Jun 17;6(3):277-287. [doi: 10.1177/1754073914523052]
40. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]

XSL•FO
RenderX

41. Oettingen G. Future thought and behaviour change. European Review of Social Psychology 2012 Mar;23(1):1-63. [doi: 10.1080/10463283.2011.643698]
42. Cross A, Sheffield D. Mental contrasting for health behaviour change: a systematic review and meta-analysis of effects and moderator variables. Health Psychol Rev 2019 Jun;13(2):209-225. [doi: 10.1080/17437199.2019.1594332] [Medline: 30879403]
43. Suorsa K, Leskinen T, Pulakka A, Pentti J, Löyttyniemi E, Heinonen I, et al. The effect of a consumer-based activity tracker intervention on accelerometer-measured sedentary time among retirees: a randomized controlled REACT trial. J Gerontol A Biol Sci Med Sci 2022 Mar 03;77(3):579-587 [FREE Full text] [doi: 10.1093/gerona/glab107] [Medline: 33839766]
44. Hagger MS, Chatzisarantis NL. An integrated behavior change model for physical activity. Exerc Sport Sci Rev 2014 Apr;42(2):62-69. [doi: 10.1249/JES.0000000000000008] [Medline: 24508739]
45. Deci EL, Ryan RM. Intrinsic Motivation and Self-Determination in Human Behavior. New York, NY: Springer; 1985.
46. Hagger MS, Protogerou C. Affect in the Context of Self-Determination Theory. In: Williams DM, Rhodes RE, Conner MT, editors. Affective Determinants of Health Behavior. Oxford, England: Oxford University Press; 2018.
47. Ajzen I. Perceived behavioral control, self‐efficacy, locus of control, and the theory of planned behavior. J Applied Social Pyschol 2006 Jul 31;32(4):665-683. [doi: 10.1111/j.1559-1816.2002.tb00236.x]
48. Klusmann V, Gow AJ, Robert P, Oettingen G. Using theories of behavior change to develop interventions for healthy aging. J Gerontol B Psychol Sci Soc Sci 2021 Sep 13;76(Suppl 2):S191-S205. [doi: 10.1093/geronb/gbab111] [Medline: 34515775]
49. Wittleder S, Kappes A, Krott N, Jay M, Oettingen G. Mental contrasting spurs energy by changing implicit evaluations of obstacles. Motivation Science 2020 Jun;6(2):133-155 [FREE Full text] [doi: 10.1037/mot0000140]
50. Lincoln YS, Guba EG, Pilotta JJ. Naturalistic inquiry. International Journal of Intercultural Relations 1985 Jan;9(4):438-439. [doi: 10.1016/0147-1767(85)90062-8]

Abbreviations

IBC: integrated behavior change
MCII: mental contrasting and implementation intention
PD: participatory design

Edited by A Mavragani; submitted 24.06.24; peer-reviewed by W Tumuhimbise, L McGowan; comments to author 26.11.24; revised version received 02.12.24; accepted 03.12.24; published 20.01.25.

Please cite as:
Hultman L, Eklund C, von Heideken Wågert P, Söderlund A, Lindén M, Elfström ML
Development of an eHealth Intervention Including Self-Management for Reducing Sedentary Time in the Transition to Retirement: Participatory Design Study
JMIR Form Res 2025;9:e63567
URL: https://formative.jmir.org/2025/1/e63567
doi:10.2196/63567
PMID:

©Lisa Hultman, Caroline Eklund, Petra von Heideken Wågert, Anne Söderlund, Maria Lindén, Magnus L Elfström. Originally published in JMIR Formative Research (https://formative.jmir.org), 20.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Methods to Adjust for Confounding in Test-Negative Design COVID-19 Effectiveness Studies: Simulation Study

Elizabeth AK Rowley[1], DrPH; Patrick K Mitchell[1], ScD; Duck-Hye Yang[1], PhD; Ned Lewis[2], MPH; Brian E Dixon[3,4], MPA, PhD; Gabriela Vazquez-Benitez[5], PhD; William F Fadel[3,4], PhD; Inih J Essien[5], MPH, OD; Allison L Naleway[6], PhD; Edward Stenehjem[7], MSc, MD; Toan C Ong[8], PhD; Manjusha Gaglani[9,10], MBBS; Karthik Natarajan[11,12], PhD; Peter Embi[3,13], MS, MD; Ryan E Wiegand[14], PhD; Ruth Link-Gelles[14], MPH, PhD; Mark W Tenforde[14], MD, PhD; Bruce Fireman[2], MA

[1]Westat, Rockville, MD, United States
[2]Vaccine Study Center, Northern California Division of Research, Kaiser Permanente, Oakland, CA, United States
[3]Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN, United States
[4]Fairbanks School of Public Health, Indiana University, Indianapolis, IN, United States
[5]HealthPartners Institute, Minneapolis, MN, United States
[6]Center for Health Research, Kaiser Permanente, Portland, OR, United States
[7]Division of Infectious Diseases and Clinical Epidemiology, Intermountain Healthcare, Salt Lake City, UT, United States
[8]Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
[9]Department of Pediatrics, Section of Pediatric Infectious Diseases, Baylor Scott & White Health, Temple, TX, United States
[10]Department of Medical Education, Texas A&M University College of Medicine, Killeen, TX, United States
[11]Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY, United States
[12]New York Presbyterian Hospital, New York, NY, United States
[13]Vanderbilt University Medical Center, Nashville, TN, United States
[14]National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, GA, United States

Corresponding Author:
Elizabeth AK Rowley, DrPH
Westat
1600 Research Blvd
Rockville, MD, 20850
United States
Phone: 1 301 251 1500
Email: ELIZABETHROWLEY@WESTAT.COM

Abstract

Background: Real-world COVID-19 vaccine effectiveness (VE) studies are investigating exposures of increasing complexity accounting for time since vaccination. These studies require methods that adjust for the confounding that arises when morbidities and demographics are associated with vaccination and the risk of outcome events. Methods based on propensity scores (PS) are well-suited to this when the exposure is dichotomous, but present challenges when the exposure is multinomial.

Objective: This simulation study aimed to investigate alternative methods to adjust for confounding in VE studies that have a test-negative design.

Methods: Adjustment for a disease risk score (DRS) is compared with multivariable logistic regression. Both stratification on the DRS and direct covariate adjustment of the DRS are examined. Multivariable logistic regression with all the covariates and with a limited subset of key covariates is considered. The performance of VE estimators is evaluated across a multinomial vaccination exposure in simulated datasets.

Results: Bias in VE estimates from multivariable models ranged from –5.3% to 6.1% across 4 levels of vaccination. Standard errors of VE estimates were unbiased, and 95% coverage probabilities were attained in most scenarios. The lowest coverage in the multivariable scenarios was 93.7% (95% CI 92.2%-95.2%) and occurred in the multivariable model with key covariates, while the highest coverage in the multivariable scenarios was 95.3% (95% CI 94.0%-96.6%) and occurred in the multivariable model with all covariates. Bias in VE estimates from DRS-adjusted models was low, ranging from –2.2% to 4.2%. However, the

XSL•FO
RenderX

DRS-adjusted models underestimated the standard errors of VE estimates, with coverage sometimes below the 95% level. The lowest coverage in the DRS scenarios was 87.8% (95% CI 85.8%-89.8%) and occurred in the direct adjustment for the DRS model. The highest coverage in the DRS scenarios was 94.8% (95% CI 93.4%-96.2%) and occurred in the model that stratified on DRS. Although variation in the performance of VE estimates occurred across modeling strategies, variation in performance was also present across exposure groups.

Conclusions: Overall, models using a DRS to adjust for confounding performed adequately but not as well as the multivariable models that adjusted for covariates individually.

(JMIR Form Res 2025;9:e58981) doi:10.2196/58981

KEYWORDS

disease risk score; propensity score; vaccine effectiveness; COVID-19; simulation study; usefulness; comorbidity; assessment

Introduction

Background

Real-world, observational vaccine effectiveness (VE) studies are critical for evaluating the post licensure performance of vaccines. These studies must address confounding from patients' demographic characteristics and underlying medical conditions because such factors may be associated with both disease outcomes and vaccination status. When vaccination status was dichotomous, inverse probability weighting based on propensity scores (PS) could be used to adjust for confounding. However, the usefulness of the PS is problematic when the exposure is multinomial rather than dichotomous—for example, if we are interested in comparing multiple levels of vaccination status distinguished by the number of doses received and by intervals of time after the last dose [1]. Multinomial PS approaches are feasible in some situations [2], however, multinomial PS poses computational challenges and is less intuitive.

An alternative to the PS in this context is the disease risk score (DRS), also called a confounder score, prognostic score, comorbidity score, or simply a risk score [1-9]. A DRS can combine covariates into a single score that reflects their associations with the outcome. However, if it is feasible to make a DRS that adjusts appropriately for the relevant covariates, it can be similarly feasible and appropriate to simply adjust for the covariates individually without first combining them into a DRS [1]. This simulation study compared logistic regression models that use a DRS to adjust VE estimates versus logistic regression models that adjust for covariates individually. We compared DRS adjustment with individual covariate adjustment in scenarios comprised of simulated data similar to real-world data used by the Virtual SARS-CoV-2, Influenza, and Other Respiratory Viruses Network (VISION) to report on COVID-19 VE.

VISION: Virtual SARS-CoV-2, Influenza, and Other Respiratory Viruses Network Studies of COVID-19 VE

The VISION network was established by the Centers for Disease Control and Prevention (CDC) in collaboration with 10 US health care systems with medical, laboratory, and vaccination records. VISION uses a case-control test-negative design (TND) to assess the effectiveness of COVID-19 vaccines in preventing laboratory-confirmed COVID-19–associated hospitalizations and visits to emergency departments or urgent care clinics [10]. Patients who received care in one of these settings for a COVID-19-like illness (CLI) are included in the study if they were tested for SARS-CoV-2 by molecular assay proximate to the encounter. Those who tested positive were considered cases; those who tested negative were considered controls. CLI diagnoses include acute respiratory diagnoses or related signs or symptoms, captured by diagnosis codes [10-13]. The case-control TND has been commonly used in studies of COVID-19 VE and VE studies against influenza, rotavirus, and other diseases [14-16]. This study was reviewed and approved by the institutional review board at Westat, Inc. This electronic health record-based study does not include factors necessitating patient consent. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the CDC.

When VISION first began analyzing COVID-19 VE in early 2021, the comparison of interest was between fully vaccinated individuals and unvaccinated individuals. PS-based methods have been widely used in cohort studies with this kind of binary exposure [17-19]. VISION's case-control studies also used a PS, derived from the test-negative controls, to adjust VE estimates [10,20].

As the pandemic progressed, recommended vaccination schedules became more complex [21-23]. Research studies required consideration of more nuanced exposure categories, based on a combination of the number of vaccine doses received and time since the most recent vaccine dose (to examine the waning of vaccine-induced protection against COVID-19) [24-26]. VISION continued to rely on the PS for covariate adjustment by conducting a series of separate comparisons, each comparing one vaccine status with another. For example, 1 PS was derived and used in comparisons of vaccinees who received 3 doses versus the unvaccinated; then a separate PS was derived and used in comparisons of vaccinees who received 4 doses versus the unvaccinated. However, separate models do not allow for easy testing between different levels of exposure and are computationally challenging, with a single analysis often requiring over 1000 PS to account for different vaccination exposure groups and population subgroups being assessed. In addition, the positivity assumption of the PS requires that at no level of the PS is outcome probability close to 0 or 1 [27,28]. The widening difference in characteristics between those who closely follow the recommended vaccination schedule and those remaining unvaccinated put a strain on the positivity assumption

XSL•FO
RenderX

and decreased the appeal of collapsing vaccination categories into two.

Disease Risk Scores to Adjust for Confounding

An alternative to PS is the DRS. A DRS reflects the relationship between potential confounders and the outcome [1-4,7,9,29,30]. Conditioning on a DRS does not balance covariates across treatment groups as a PS would but results in a covariate balance where the potential outcome under the referent condition is independent of a set of covariates [29], which has been called prognostic balance by Hansen [30]. In short, the DRS is an observation's predicted risk of the outcome, assuming the observation was in the referent category of the exposure, reflecting the risk of the outcome relative to other observations. There are well-known generalizable risk scores, such as the Gail model score [31], Framingham score [32], APACHE score [33], and Charlson index [34] that can be used to control for baseline health status. Instead, a study may generate a DRS specific to the outcome and population. In the context of a TND, the outcome is not simply testing positive for the disease of interest but rather testing positive for the disease of interest conditional on receiving care and being tested. A key benefit of using the DRS is potentially only calculating it once for the entire dataset, or a few times for a few key subgroups, rather than generating up to thousands of PS for comparisons using different dichotomous comparisons between multiple vaccination exposure groups.

We compared the performance of DRS-adjusted estimators with multivariable-adjusted estimators that do not aggregate the covariates into a composite score. Several studies have implemented TNDs with this straightforward approach—that is, by fitting logistic regression models that include many covariates in addition to indicators of vaccination status [35-37]. While many analytic methods select variables or adjust for confounding, we are particularly interested in the possibility the DRS allows us to calculate a single score to apply to many analyses.

Methods

Monte Carlo Simulation

We conducted simulations to compare the performance of alternative approaches to the adjustment of VE estimates for potential confounding—either by DRS or by individual covariate adjustment—in VISION network-like scenarios. In order of generation, the simulated data consisted of a bootstrapped sample of individuals, each with (1) a profile of continuous and categorical covariates, (2) a 13-level vaccination status derived conditionally from the covariates, and (3) a 2-level outcome derived conditionally from vaccination status and the covariates. The simulation study used R=1000 replications, which ensured that an estimated coverage of 95% would have a Monte Carlo error of 0.7 [38]. Datasets were generated as follows:

1. A bootstrap sample of size N=1000 or 10,000 observations was drawn from observed emergency department and urgent care encounters in VISION data of adults aged ≥18 years from the Omicron predominance era (December 16, 2021-July 31, 2022). Each observation consisted of all covariates aside from vaccination status and test result for a sampled encounter. This maintained complex relationships between potential confounders in the real-world VISION study. The data used to initiate this simulation study were accessed beginning on August 8, 2022. The authors did not have access to information that could identify individual participants during or after the analysis.
2. Vaccination status was generated as a 13-level exposure variable incorporating the number of vaccine doses received and the time interval since receipt of the most recent dose from a multinomial distribution for each observation, V_i, based on the bootstrapped covariates, Xi (1).

, where k=1,2,…,13 and i=1,2,…, N (1)

β_k was derived from the dataset by multinomial logistic regression and shown in Table S1 in Multimedia Appendix 1. The generated 13-level vaccination status was collapsed into a 5-level vaccination status as shown in Table 1 with approximate frequencies. We focused on VE estimates for each of the 4 levels of vaccination status compared with a reference level of vaccination status.

Next, we used a Bernoulli distribution to generate an outcome, Y_i dependent on the covariates in the bootstrap sample, and generated 13-level vaccination status (2). γ are the coefficients describing the true VE and are shown in Table 1. α are defined in Table S2 in Multimedia Appendix 2. In the VISION data, from which the bootstrap sample was drawn, approximately 22% of encounters were cases. This percentage varied depending on covariates such as season or vaccination status. For example, cases comprise approximately 42% of encounters in January 2022 and only 13% of encounters in May. Among those unvaccinated, approximately 30% are cases, while among those recently vaccinated with their third dose, 11% are cases.

XSL•FO
RenderX

Table 1. Definition of vaccination status categories for data generation and analysis with the true VE used in generating simulated data.[a]

Thirteen-level vaccination status categories for data generation				Five-level vaccination status categories for VE analysis	
Definition[b]	Approximate frequency in simulated data[c,d], %	Coefficient for true VE, γ	Corresponding true VE	Definition	Approximate frequency in simulated data[c,d], %
Unvaccinated	40	—[e]	—	Unvaccinated	40
2 doses, 14-59 d	1	–1.05	65	2 doses, 14-179 d	6
2 doses, 60-119 d	2	–0.60	45	—	—
2 doses, 120-179 d	3	–0.43	35	—	—
2 doses, 180-239 d	5	–0.36	30	2 doses, ≥180 d	24
2 doses, 240-299 d	7	–0.29	25	—	—
2 doses, 300-359 d	6	–0.22	20	—	—
2 doses, ≥360 d	6	–0.16	15	—	—
3 doses, 7-59 d	6	–1.61	80	3 doses, 7-119 d	15
3 doses, 60-119 d	10	–1.39	75	—	—
3 doses, 120-179 d	8	–0.43	35	3 doses, ≥120 d	15
3 doses, 180-239 d	5	–0.16	15	—	—
3 doses, ≥240 d	1	–0.11	10	—	—

[a]VE: vaccine effectiveness.

[b]Vaccination categories are defined by the number of mRNA vaccines received at the index date of the medical encounter and the days between the most recent dose and the index date. The index date is the earlier of the date of specimen collection of a positive test within the allowable testing window (14 d before to 72 h after the encounter) and the time of the medical encounter.

[c]Approximate percent observed in the full dataset from which samples of 10,000 or 1000 are drawn.

[d]Data were generated using a 13-level vaccination status variable, with the approximate frequency of each category shown. This frequency varies across the 1000 replicates. The vector of coefficients, γ, is used as part of the data generation process, and its corresponding values on the VE scale are shown. In the analysis phase, the 13 categories are collapsed to the 5 shown in the column on the right, again with approximate frequencies.

[e]Not available.

Derivation of the DRS

Methods for deriving the DRS have been described elsewhere [1-5,7,29,39]. For this simulation study, we used Miettinen's approach, which models the outcome in relation to the covariates in the entire study population and includes the exposure of the vaccination status in the model with the covariates. The coefficients from the fitted model were used to calculate a predicted DRS for the entire dataset, predicting what risk of the outcome would be for each individual if they were unvaccinated (or if they were given the "reference" level of vaccination status in other scenarios) [5]. Multimedia Appendix 3 [40,41] contains more details about why this approach was selected.

Consideration around the use of a single DRS for the entire VISION study population for each analysis of VE was required. For example, VISION analyses were commonly stratified by age and immunocompromising status. Those with immunocompromising conditions had different vaccine recommendations and may have had reduced VE [42]. Accordingly, in each generated dataset, up to 3 DRS were estimated for each individual depending on their age and immunocompromised (IC) status. One used the full generated dataset, the second used only those with IC conditions according to VISION network criteria, and the third used only those 50 years and older (≥50).

In Miettinen's approach, incorrectly modeling the modification of exposure by baseline covariates can result in estimated scores that are influenced by the magnitude of the exposure effect [30]. Applied to the estimation of DRS, gradient-boosted regression trees look for multilevel interactions that are useful for describing the relationship between patient characteristics, including vaccination status, and the outcome [20]. DRS were estimated using gradient boosted regression and included the exposure and all the covariates. Table S1 in Multimedia Appendix 1 contains a listing of these covariates.

Final VE Models

The VE was calculated in a variety of logistic regression models, calculated as (1–adjusted odds ratio)×100%. Without a DRS, we estimated unadjusted VE, VE adjusted only for the key variables of patient age (as a spline), calendar date from January 1, 2021 (epi-day, as a spline), and VISION site and sub-region of the health care facility (site-region) (referred to as multivariable key), and VE adjusted for the key variables plus all other variables in the DRS (referred to as multivariable all). Note that the term spline refers to a natural cubic spline with 3 knots placed at the 25th, 50th, and 75th percentiles.

While a PS is typically used to balance comparisons of the exposed versus the unexposed by inverse probability weighting or by matching on the PS [43], there is less theory or experience

XSL•FO
RenderX

pertaining to the use of the DRS in the final model that yields the adjusted VE [2]. A model may include DRS as a covariate [44], or a model could be conditional on DRS [45], which under certain conditions, may yield an average of the individual treatment effects of those treated (ATT) [29]. Well-known generalizable risk scores, such as the Gail model score, Framingham score, APACHE score, and Charlson index, are sometimes included in models or are used for stratification. The risk score we investigated was not designed to be generalizable, however, the modeling strategies employed for including a DRS in a VE model are similar. We investigated DRS inclusion in VE models as strata in a conditional model as well as a continuous covariate fit with a flexible spline. Conditional models used DRS as centiles, or deciles with epi-week, and site as strata. Conditional models that are only conditioned on DRS and models that included DRS as a continuous covariate also adjusted for age (as a spline), epi-day (as a spline), and site region. Models that were conditioned on DRS epi-week and site were also adjusted for age (as a spline). Vaccination status in the final VE model was collapsed to the 5-level category. Table 2 summarizes the calculated models.

XSL•FO
RenderX

Table 2. Description of final vaccine effectiveness (VE) models assessed in the simulation study.

Model name	Subset of data used to estimate DRS[a,b]. Subset of data used for estimating VE[c].	Model stratification and adjustment	Dataset totals, n
Unadjusted	DRS dataset: no DRS. VE dataset: full simulated dataset.	Vaccination status only	10,000 and 1000
Multivariable all	DRS dataset: no DRS. VE dataset: full simulated dataset.	• Stratification: none • Adjustment: variables that would go into the DRS including age (spline), epi-day (spline), and site region	10,000 and 1000
Multivariable key	DRS dataset: no DRS. VE dataset: full simulated dataset.	• Stratification: none • Adjustment: age (spline), epi-day (spline), and site region	10,000 and 1000
Stratified week, site, DRS	DRS dataset: full simulated dataset. VE dataset: full simulated dataset.	• Stratification: DRS decile, epi-week, and site • Adjustment: age (spline)	10,000 and 1000
Stratified DRS	DRS dataset: full simulated dataset. VE dataset: full simulated dataset.	• Stratification: DRS centile • Adjustment: age (spline), epi-day (spline), and site region	10,000 and 1000
Spline DRS	DRS dataset: full simulated dataset. VE dataset: full simulated dataset.	• Stratification: none • Adjustment: DRS included with a natural spline with 4 degrees of freedom, age (spline), epi-day (spline), and site-region	10,000 and 1000
Multivariable key IC[d]	1. DRS datasets: no DRS. VE dataset: IC subset. 2. DRS datasets: IC subset. VE dataset: IC subset.	• Stratification: none • Adjustment: age (spline), epi-day (spline), and site	10,000
Stratified week, site, IC DRS	1. DRS datasets: full simulated dataset. VE dataset: IC subset. 2. DRS datasets: IC subset. VE dataset: IC subset.	• Stratification: DRS decile, epi-week, and site • Adjustment: age (spline)	10,000
Stratified IC DRS	1. DRS datasets: full simulated dataset. VE dataset: IC subset. 2. DRS datasets: IC subset. VE dataset: IC subset.	• Stratification: DRS centile • Adjustment: age (spline), epi-day (spline), and site	10,000
Spline IC DRS	1. DRS datasets: full simulated dataset. VE dataset: IC subset. 2. DRS datasets: IC subset. VE dataset: IC subset.	• Stratification: none • Adjustment: DRS included with a natural spline with 4 degrees of freedom, age (spline), epi-day (spline), and site	10,000
Multivariable key ≥50[e]	1. DRS datasets: no DRS. VE dataset: ≥50 subset. 2. DRS datasets: ≥50 subset. VE dataset: ≥50 subset.	• Stratification: none • Adjustment: only for age (spline), epi-day (spline), and site region	10,000
Stratified week, site, ≥50 DRS	1. DRS datasets: full simulated dataset. VE dataset: ≥50 subset. 2. DRS datasets: ≥50 subset. VE dataset: ≥50 subset.	• Stratification: DRS decile, epi-week, and site • Adjustment: age (spline)	10,000
Stratified ≥50 DRS	1. DRS datasets: full simulated dataset. VE dataset: ≥50 subset 2. DRS datasets: ≥50 subset. VE dataset: ≥50 subset.	• Stratification: DRS centile • Adjustment: age (spline), epi-day (spline), and site region	10,000

XSL•FO
RenderX

Model name	Subset of data used to estimate DRS[a,b]. Subset of data used for estimating VE[c].	Model stratification and adjustment	Dataset totals, n
Spline ≥50	1. DRS datasets: full simulated dataset. VE dataset: ≥50 subset. 2. DRS datasets: ≥50 subset. VE dataset: ≥50 subset.	• Stratification: none • Adjustment: DRS included with a natural spline with 4 degrees of freedom, age (spline), epi-day (spline), and site region	10,000

[a]DRS: disease risk score.

[b]One goal of this simulation study is to consider how the DRS might behave in smaller subsets of the full dataset, specifically a subset of patients with IC conditions and a subset of patients 50 years or older. For this reason, when analyzing these subsets, we consider a DRS built from the full dataset but applied to a smaller subset and a DRS built specifically from that smaller subset.

[c]VE: vaccine effectiveness.

[d]IC: patients with immunocompromising conditions.

[e]≥50: patients aged 50 years or older.

Assessing Model Performance

To evaluate performance, we examined bias, SE, the ratio of mean SE to empirical SE, and coverage. The true VE is built into the data generation mechanism, and these known values are shown in Table 1 for the 13 levels of vaccination status. However, VE was estimated for the collapsed 5-level vaccination status, $Vcollapsed_i$. Therefore, for each replication, r, the true effect for each 5-level vaccination status, $\gamma_{m,r}$, was approximated with a weighted average (weighted by sample size in each of the 13 vaccination status categories) as shown below (3) and summarized as a VE (4).

The estimated VE for each vaccine status in each replication, , was calculated as shown in (5).

Percent VE bias was calculated as the average percent bias across the replications on the VE scale (6).

The ratio of the observed SE of the OR compared with the empirical SE was calculated as shown in (7).

Presentation of Results

Simulation results are presented in figures depicting the complete distribution of observed percent VE bias and SE of $\gamma_{m,r}$, for each vaccination status. The average across all replications is shown as well as the 2.5th and 97.5th percentiles. The empirical SE, or SD of the coefficients, is presented as a vertical bar. In addition to low bias, an ideal model will provide estimates of SE that are also averaged across replications, close to the SD of the estimated coefficients. The figures include coverage. All results are in Tables S3 and S4 in Multimedia Appendices 4 and 5. Coverage is detailed in the results, with CIs for the coverage probabilities based on the binomial distribution of the R coverage indicators [38]. All simulations and analyses were conducted with R Statistical Software (R Foundation for Statistical Computing) [46].

Ethical Considerations

This study was reviewed and approved by the institutional review boards at participating sites and under a reliance agreement between the CDC and the Westat institutional review board (FWA# FWA00005551, expiry date 10/13/2027, IRB project number 6201.08). This activity was reviewed by the CDC and was conducted consistent with applicable federal law and CDC policy (eg, 45 CFR part 46.102(l)(2), 21 CFR part 56; 42 USC §241(d); 5 USC §552a; 44 USC §3501). This activity was reviewed and approved as a research activity by one VISION site. This study presented minimal risk to participants because there was no interaction or intervention with patients; therefore, a waiver of informed consent was granted.

Results

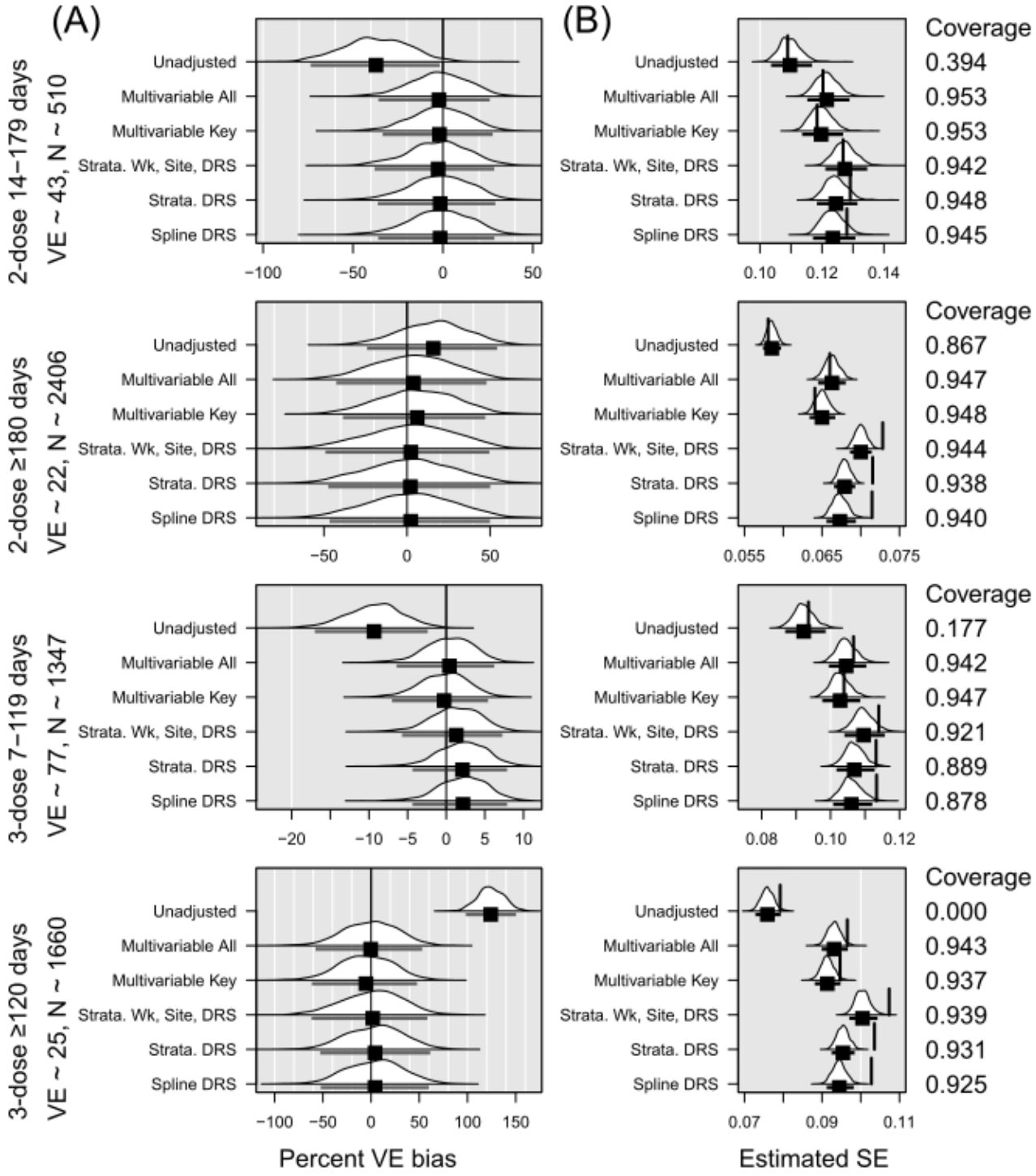
Comparing Strategies for Estimating VE

When N=10,000, the unadjusted models resulted in an average percent bias of VE ranging from –37% to 124% across the 4 vaccination categories (Figure 1, Table S3 in Multimedia Appendix 4). Average percent bias was dramatically lower in the models that accounted for the covariates: the 2 multivariable models without DRS (range for multivariable with all covariates, including key covariates from –2.2 to 3.8, multivariable with only key covariates from –5.3 to 6.1), the stratified model with DRS decile epi-week and site (range –2.6 to 2.3), the stratified model with DRS centile (range –1.5 to 4.2), and the model with DRS as a spline (range –1.6 to 4.1). Bias varied less across the 5 strategies for estimating VE than across the 4 levels of vaccination status for which VE was estimated. For example, the bias in estimates of 3-dose VE within 120 days ranged from –10% to 10% approximately, whereas the bias in estimates of 2-dose VE within 180 days ranged from –50% to 40% approximately. The difference in bias between vaccination status levels is related to the different sample sizes expected in each vaccine status group and the expected increasing instability of

XSL•FO
RenderX

the percent bias when VE is closer to 0 (ie, 5 units is a much greater percent of 25 than it is of 80).

Figure 1. Summary of simulation results in the overall sample of 10,000. DRS: disease risk score; VE: vaccine effectiveness.



In all figures, panel A presents the percent VE bias for each vaccination level, with a vertical black line at 0. The distribution across 1000 replications is shown for each model strategy. The mean percent VE bias, represented with a black square, with an interval indicating the observed 2.5th and 97.5th percentiles, appears below each distribution. Panel B presents the distribution across 1000 replications of the estimated standard error for each vaccination level. The mean standard error, represented with a black square, with an interval indicating the observed 2.5th and 97.5th percentiles appears below the distribution. The vertical line shows the empirical standard error (SE), or SD of the coefficients. Ideally, the vertical line aligns with the mean standard error. Means to the left of the empirical SE indicate underestimation of the SE. Coverage is shown on the right of Panel B. Coverage is defined as the proportion of confidence intervals from the simulated analyses that include the true VE. Ideally, this would be close to 95%.

Across all 4 vaccination statuses, the multivariable models had the lowest SE. In most instances, the average SE came closest

XSL•FO
RenderX

to the SD of the coefficients in the multivariable models. In all but the 2-dose 14-179 days category, the models incorporating DRS as a spline and the models stratified on DRS centile resulted in empirical SE greater than the 97.5th percentile of observed standard errors. The model stratified on decile of DRS, epi-week, and site fared slightly better, with empirical SE greater than the 97.5th percentile in 2 of the 4 vaccination categories. However, the models stratified on decile of DRS, epi-week, and site had the highest mean SE in all vaccination categories. Models including DRS as a spline had the greatest underestimation of SE, as demonstrated with the lowest ratio of mean SE to empirical SE (Table S3 in Multimedia Appendix 4). With the exception of the multivariable all or multivariate key models, all models yielded CIs that covered the true VE in less than 95% of replicated analyses for at least one vaccination status, as shown in Table 3.

We also considered subgroups of the N=10,000 dataset defined by immunocompromised (IC) status and age, as these groups were routinely studied by VISION and others to guide vaccination policy decisions. Out of N=10,000, approximately 532 fell into the IC subgroup and 4782 into the 50 years or older subgroup. In the IC subgroup, the multivariable model with key covariates and without the DRS performed best in terms of bias—its VE estimates for each level of vaccination status were least biased (Figure 2 and Table S4 in Multimedia Appendix 5). This multivariable model also had the lowest SE while achieving 95% coverage across two of the four levels of vaccination status (Table 3). For estimates of 2-dose VE within 180 days, the coverage with the multivariable model was 96.8% (CI 95.1-97.5), and for 3-dose VE within 120 days, the coverage with the multivariable model was also 96.8% (CI 95.7-97.9). While having slightly higher bias, the spline model and the model stratified using DRS derived from the full 10,000 achieved 95% coverage in all 4 vaccination categories. The IC group was so small that several replications yielded very high SE estimates for at least one vaccination status indicating that a meaningful VE estimate was not obtained. This occurred in 21 of 1000 replications for the multivariate model, 11-132 replications for the models using the DRS from the full model, and 21-230 replications for the models using the DRS from the IC subset. In the older subgroup, the multivariable model with key covariates and without DRS again performed best in terms of bias across all 4 vaccination statuses (Figure 3 and Table S4 in Multimedia Appendix 5). Again, the multivariable model often had the lowest SE, which was also closest to the SD of the coefficients. Only the multivariable model achieved 95% coverage for its VE estimates for all four levels of vaccination status. Only the model stratified on DRS (from the full cohort), epi-week, and site-region achieved 95% coverage for 3 of the 4 levels of vaccination status; the other models only achieved this for 2 or fewer levels of vaccination status (Table 3). In both the IC subgroup and the 50 years or older subgroup, the models using a DRS derived from the subgroup tended to perform worse than models with a DRS derived from the full sample of 10,000.

When N=1000, the sizes in each vaccination category were dramatically lower, ranging from 50 in the 2-dose 14-179 days group to 242 in the 2-dose ≥180 days group. Again, the multivariable models, particularly the one with all the covariates, had the least bias in VE estimates for each level of vaccination status (Figure 4 and Table S3). The multivariable models, along with the model stratified by epi-week, site, and DRS decile all achieved 95% coverage across the four vaccination statuses (Table 3). The model stratified by DRS centile and the model with a DRS spline had coverages slightly below 90% for 3-dose VE within 120 days (89.6% coverage CI: 87.7%-91.5% and 89.1% coverage CI: 87.2%-91%, respectively).

XSL•FO
RenderX

Table 3. Estimates of coverage of the true vaccine effectiveness (VE) with 95% CIs.

Model strategy	Two-dose 14-179 d, coverage[a] (95% CI[b])	Two-dose ≥180 d, coverage[a] (95% CI[b])	Three-dose 7-119 d, coverage[a] (95% CI[b])	Three-dose ≥120 d, coverage[a] (95% CI[b])
N=10,000				
Multivariable all	95.3 (94.0-96.6)	94.7 (93.3-96.1)	94.2 (92.8-95.6)	94.3 (92.9-95.7)
Multivariable key[c]	95.3 (94.0-96.6)	94.8 (93.4-96.2)	94.7 (93.3-96.1)	93.7 (92.2-95.2)
Strata: week site, DRS[d]	94.2 (92.8-95.6)	94.4 (93.0-95.8)	92.1 (90.4-93.8)	93.9 (92.4-95.4)
Strata: DRS	94.8 (93.4-96.2)	93.8 (92.3-95.3)	88.9 (87.0-90.8)	93.1 (91.5-94.7)
Spline DRS	94.5 (93.1-95.9)	94.0 (92.5-95.5)	87.8 (85.8-89.8)	92.5 (90.9-94.1)
IC[e] subgroup				
Multivariable key[c]	96.3 (95.1-97.5)	94.9 (93.5-96.3)	96.8 (95.7-97.9)	94.7 (93.3-96.1)
Strata: week site, DRS	97.8 (96.9-98.7)	95.6 (94.3-96.9)	96.6 (95.5-97.7)	97.0 (95.9-98.1)
Strata: DRS	96.0 (94.8-97.2)	95.8 (94.6-97.0)	95.6 (94.3-96.9)	96.1 (94.9-97.3)
Spline DRS	96.0 (94.8-97.2)	94.5 (93.1-95.9)	94.7 (93.3-96.1)	94.7 (93.3-96.1)
Strata: week, site, IC DRS	96.6 (95.5-97.7)	98.3 (97.5-99.1)	98.4 (97.6-99.2)	98.2 (97.4-99.0)
Strata: IC DRS	95.4 (94.1-96.7)	97.6 (96.7-98.5)	95.8 (94.6-97.0)	94.2 (92.8-95.6)
Spline IC DRS	92.3 (90.6-94.0)	98.5 (97.7-99.3)	95.2 (93.9-96.5)	93.8 (92.3-95.3)
≥50[f] subgroup				
Multivariable key[c]	93.5 (92.0-95.0)	94.4 (93.0-95.8)	94.1 (92.6-95.6)	94.1 (92.6-95.6)
Strata: week site, DRS	94.9 (93.5-96.3)	93.8 (92.3-95.3)	93.0 (91.4-94.6)	93.5 (92.0-95.0)
Strata: DRS	93.7 (92.2-95.2)	94.2 (92.8-95.6)	90.6 (88.8-92.4)	93.0 (91.4-94.6)
Spline DRS	94.0 (92.5-95.5)	93.4 (91.9-94.9)	89.7 (87.8-91.6)	93.2 (91.6-94.8)
Strata: week, site, IC DRS	92.7 (91.1-94.3)	93.6 (92.1-95.1)	91.1 (89.3-92.9)	94.2 (92.8-95.6)
Strata: IC DRS	92.3 (90.6-94.0)	92.3 (90.6-94.0)	86.1 (84.0-88.2)	92.9 (91.3-94.5)
Spline IC DRS	91.5 (89.8-93.2)	92.2 (90.5-93.9)	84.8 (82.6-87.0)	92.3 (90.6-94.0)
N=1000				
Multivariable all	95.6 (94.3-96.9)	94.9 (93.5-96.3)	94.8 (93.4-96.2)	95.2 (93.9-96.5)
Multivariable key[c]	95.8 (94.6-97.0)	95.1 (93.8-96.4)	94.9 (93.5-96.3)	94.8 (93.4-96.2)
Strata: week site, DRS	95.6 (94.3-96.9)	94.1 (92.6-95.6)	93.8 (92.3-95.3)	95.2 (93.9-96.5)
Strata: DRS	94.2 (92.8-95.6)	92.8 (91.2-94.4)	89.6 (87.7-91.5)	93.3 (91.8-94.8)
Spline DRS	95.0 (93.6-96.4)	93.1 (91.5-94.7)	89.1 (87.2-91.0)	93.5 (92.0-95.0)

[a]Coverage is defined as the proportion of confidence intervals from the simulated analyses that include the true VE. Ideally, this would be close to 95%.

[b]Confidence intervals were calculated using the properties of the binomial distribution of the indicators that each replication covers the true VE.

[c]Key variables include simple adjustment only for age (spline), epi-day (spline), and site region.

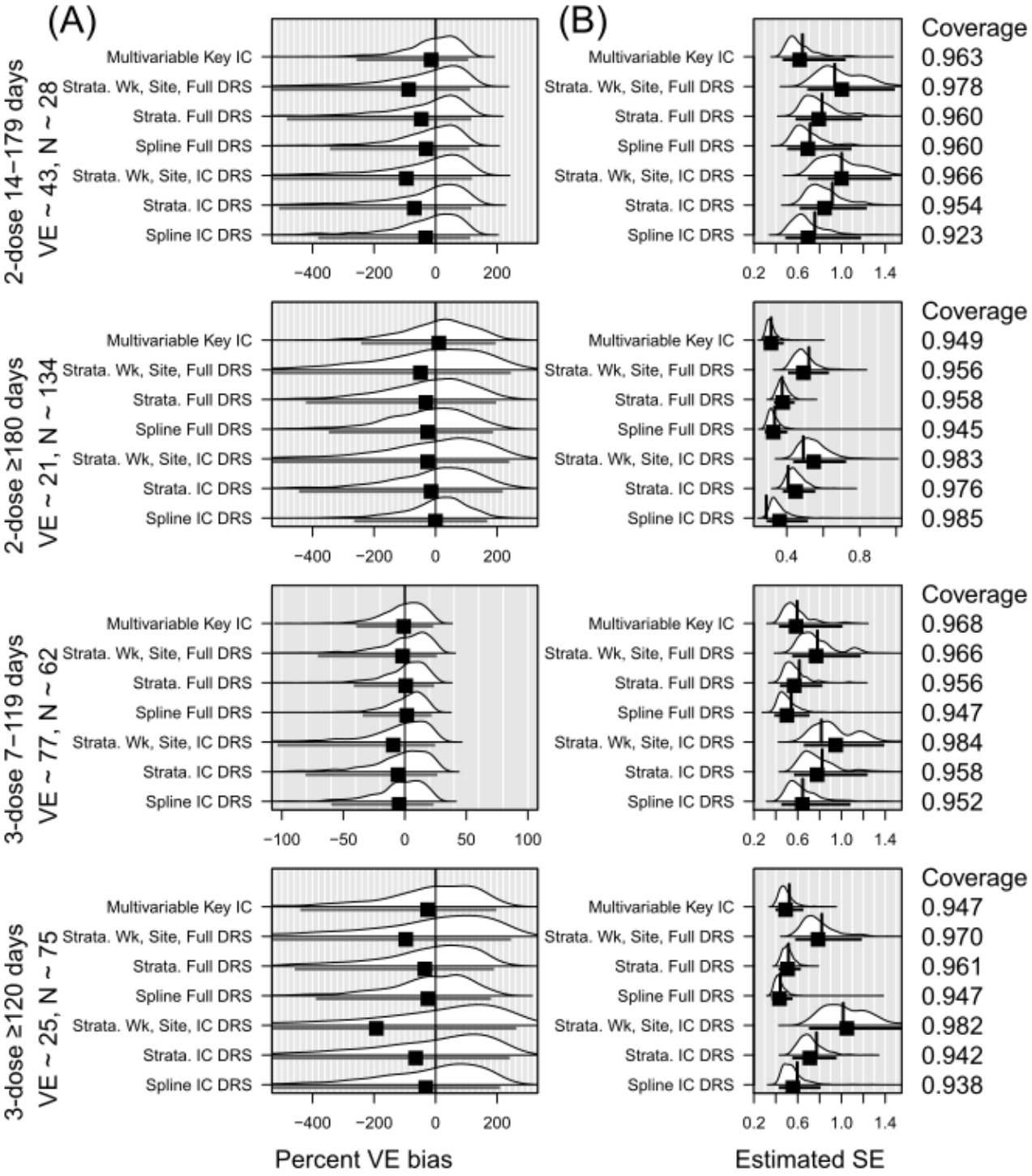
[d]DRS: disease risk score.

[e]IC: immunocompromising conditions.

[f]≥50: patients 50 years old or older.

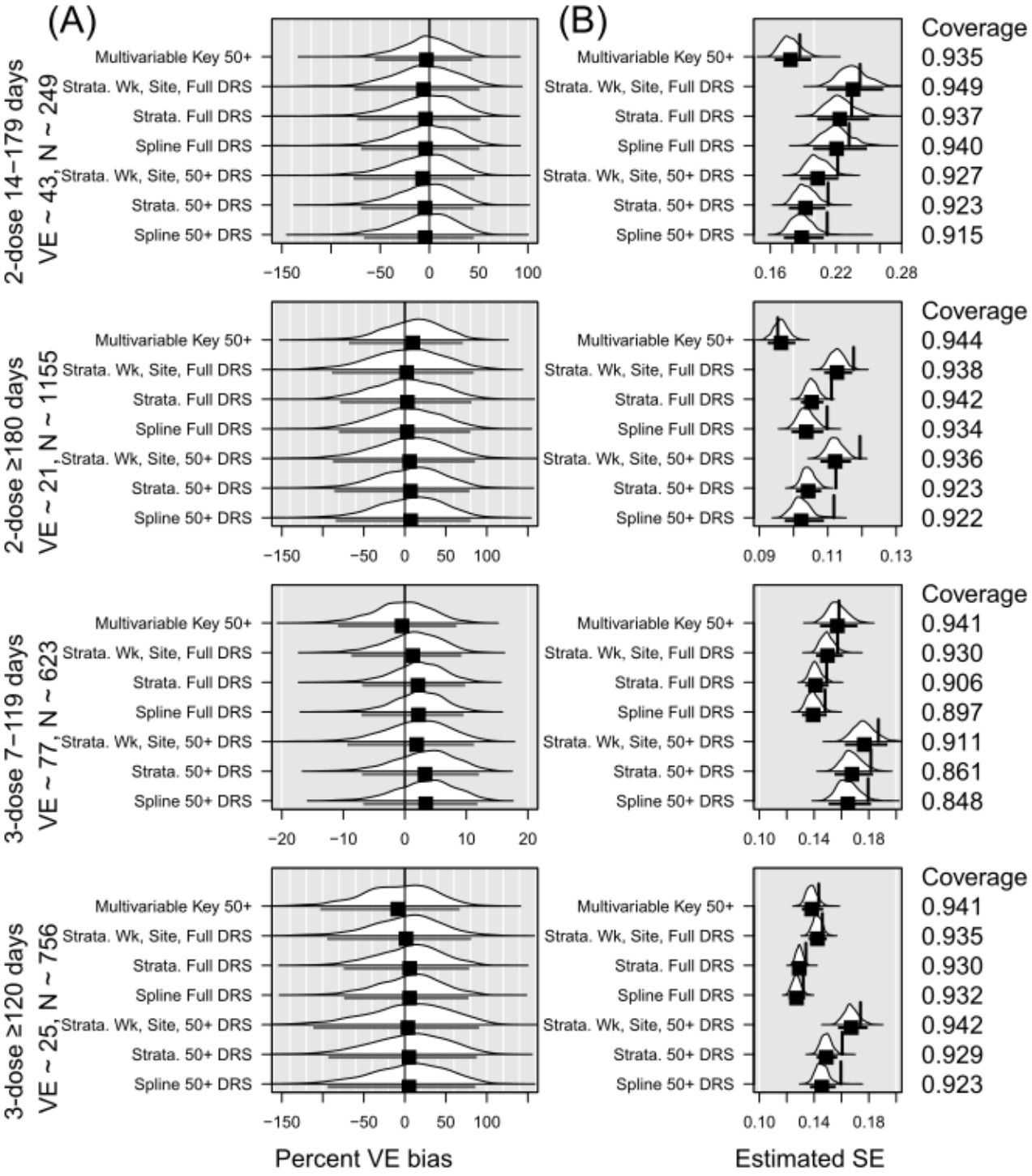
XSL•FO
RenderX

Figure 2. Summary of simulation results in the immunocompromised subset of 10,000. DRS: disease risk score; IC: immunocompromised; VE: vaccine effectiveness.



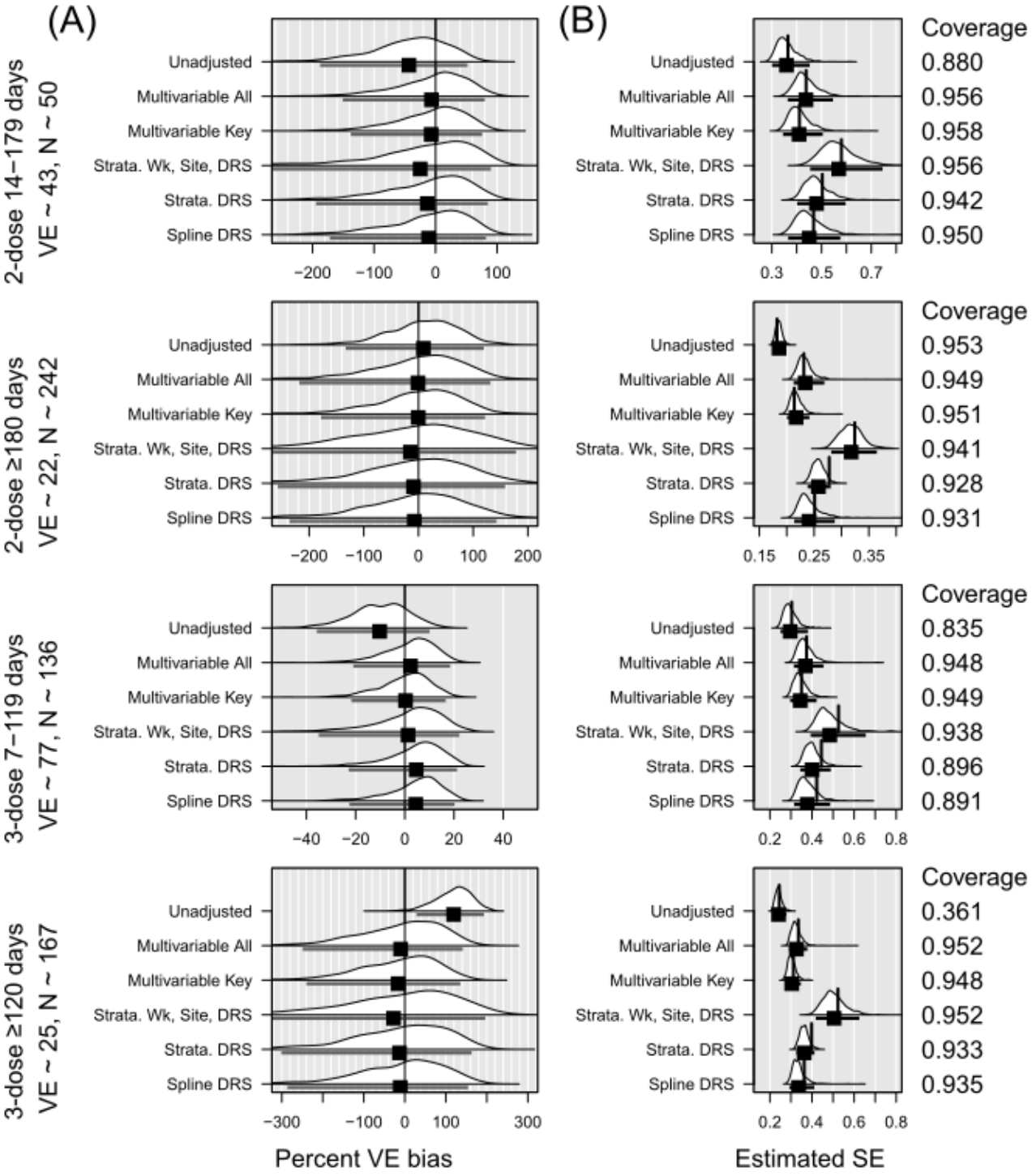
XSL•FO
RenderX

Figure 3. Summary of simulation results in the 50 years or older subset of 10,000. DRS: disease risk score; VE: vaccine effectiveness.



XSL•FO
RenderX

Figure 4. Summary of simulation results in a sample of size 1000. DRS: disease risk score; VE: vaccine effectiveness.



DRS-Related Decisions

When comparing the bias in analyses of subgroups defined by IC or age, we found little difference in the bias of VE between using a DRS built from the full dataset of 10,000 compared with using a DRS built exclusively from the subset of interest (Figures 2 and 3). When a difference was found, models using the DRS built from the full cohort tended to perform better than the models built from the smaller subset. Among the 12 VE estimates for the IC subgroup—a VE estimate for each of the 4 levels of vaccination status obtained by each of the 3 ways of using the DRS—9 of the 12 achieved 95% coverage when using a DRS from the full cohort compared with 5 of 12 when using the DRS from the IC subgroup. Among the 12 VE estimates for the subgroup aged ≥50 years, 6 of 12 achieved 95% coverage when using a DRS from the full cohort compared with 2 of 12 when using the DRS from the 50 years or older subgroup. Model-fitting challenges occurred more frequently when using the subgroup-specific DRS, particularly in the smaller IC subgroup.

XSL•FO
RenderX

Discussion

Principal Findings

This simulation study found that multivariable covariate adjustment, either with all the covariates or a subset of key covariates, performed well in the context of VISION's case-control test-negative studies of VE. Adjustment for a DRS comprised of the covariates performed adequately but tended to overestimate the precision of VE estimates.

It is possible the underestimate of the SE we have observed could be attributed to a high correlation between the confounders and vaccination status. Early studies of DRSs suggested they can help reduce the dimensionality of analyses that try to adjust for multiple covariates but that DRS adjustment tends to overestimate the precision of the effect estimate of interest [47]. Later studies suggested that the corresponding exaggeration of the statistical significance of a finding (eg, a finding about VE) would be trivial unless the covariates are very strongly associated with the exposure (eg, vaccination status) or with each other [1,48].

It should not be surprising that simple covariate adjustment performed well, given VISION's large samples of cases and controls. Other large VE studies with TNDs have also successfully employed individual covariate adjustment [35-37]. Furthermore, in previous VISION network studies, we observed that only a few of the available covariates (primarily age, calendar time, and geographic location) accounted for most of the confounding apparent in unadjusted VE estimates, which makes it feasible to adjust for the covariates individually even in subsets that might otherwise be too small to adjust for several dozen covariates or more. Even in our subgroup analyses, we did not find that use of the dimension-reducing DRS improved performance compared with straightforward adjustment for the individual covariates.

In the context of settings other than VISION, where sample sizes may be smaller, and the relevant covariates may be more numerous, the "curse of dimensionality" may render covariate adjustment more problematic. If an appropriate DRS has already been derived from large samples in similar settings, such a DRS could be helpful in several ways. First, an appropriate DRS can avoid overfitting. DRS adjustment outperformed individual covariate adjustment in some of our smaller subgroup analyses, especially when we used a DRS derived from the larger overall sample.

Second, confounding may arise from nonlinearities and interactions among covariate-outcome associations that could easily be overlooked unless previously captured by a DRS derived from large datasets using flexible machine learning methods. In this simulation study, we used boosted regression to derive the DRS, as described above, but our data-generating mechanism did not insert nonlinearities and interactions that would be undetectable in smaller datasets and would be challenging to specify in a logistic regression model with individual covariate adjustment.

Third, a DRS could be part of a hybrid approach to covariate adjustment. VE estimates could be derived from logistic regression models that include a DRS plus a few key covariates (which may also be components of the DRS). The key covariates would be those with high potential for confounding because of strong associations with the outcome (and exposure) that may differ in the current study population than in the population where the DRS was derived.

Fourth, a DRS can be an intuitive way to adjust for confounding and an intuitive tool to explore effect modification. Typically, effect modification examines one risk factor at a time—for example, examining whether the benefits of vaccination differ by age group. A DRS can be used to account for multiple risk factors as we examine whether the benefits of vaccination differ by level of risk. Furthermore, a DRS can facilitate the interpretation of findings. For example, if VE is found to be similar across levels of the DRS, yet the risk is 10-fold higher in the highest DRS decile as compared with the lowest, then we can infer that the vaccine benefits the highest decile 10-fold more (on the scale of cases prevented per 1000 persons vaccinated) than it benefits the lowest decile. However, a DRS may be less intuitive as a risk score in our test-negative VISION study than in a population-based cohort study to the extent that our test-negative controls are restricted to individuals with "a COVID-19-like illness" and may not be representative of the underlying population at risk. In the underlying population, risk factors for hospitalization for CLI may differ from risk factors for hospitalization for COVID-19.

This simulation study was motivated by challenges facing VISION's studies of VE. The scenarios we simulated emulate those examined by VISION, and this may limit the generalizability of our findings to other settings.

Conclusions

Our simulations found that logistic regression with individual covariate adjustment performed well in scenarios similar to those studied by the VISION network and is generally the current approach employed by the VISION network. DRS-adjusted models performed adequately but not as well as models that adjusted for the covariates individually.

Data Availability

The datasets generated during and/or analyzed during this study are not publicly available due to data sharing agreements between CDC and VISION network partner institutions prohibiting it.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Definition of effects of covariates on 13-level exposure, β, for multinomial model determining simulated exposure with exposure status defined by number of mRNA doses and time since most recent dose.
[DOCX File , 33 KB - formative_v9i1e58981_app1.docx]

Multimedia Appendix 2
Definition of other coefficients, α, for the Bernoulli distribution defining the probability of the generated outcome.
[DOCX File , 17 KB - formative_v9i1e58981_app2.docx]

Multimedia Appendix 3
Who is included in the disease risk score model?
[DOCX File , 24 KB - formative_v9i1e58981_app3.docx]

Multimedia Appendix 4
Bias and standard error results, 5-level vaccine effectiveness, N=10,000 and 1000.
[DOCX File , 29 KB - formative_v9i1e58981_app4.docx]

Multimedia Appendix 5
Bias and SE results, subgroup effect 5-level vaccine effectiveness.
[DOCX File , 30 KB - formative_v9i1e58981_app5.docx]

References

1. Arbogast PG, Kaltenbach L, Ding H, Ray WA. Adjustment for multiple cardiovascular risk factors using a summary risk score. Epidemiology 2008;19(1):30-37. [doi: 10.1097/EDE.0b013e31815be000] [Medline: 18091000]
2. Arbogast P, Seeger J, DMCSVW Group. Summary variables in observational research: propensity scores and disease risk scores. Effective health care program research report.: Agency for Healthcare Research and Quality; 2012. URL: https://effectivehealthcare.ahrq.gov/products/observational-research-scores/research [accessed 2024-12-19]
3. Arbogast PG, Ray WA. Use of disease risk scores in pharmacoepidemiologic studies. Stat Methods Med Res 2009;18(1):67-80. [doi: 10.1177/0962280208092347] [Medline: 18562398]
4. Arbogast PG, Ray WA. Performance of disease risk scores, propensity scores, and traditional multivariable outcome regression in the presence of multiple confounders. Am J Epidemiol 2011;174(5):613-620. [doi: 10.1093/aje/kwr143] [Medline: 21749976]
5. Miettinen OS. Stratification by a multivariate confounder score. Am J Epidemiol 1976;104(6):609-620. [doi: 10.1093/oxfordjournals.aje.a112339] [Medline: 998608]
6. Cadarette SM, Gagne JJ, Solomon DH, Katz JN, Stürmer T. Confounder summary scores when comparing the effects of multiple drug exposures. Pharmacoepidemiol Drug Saf 2010;19(1):2-9 [FREE Full text] [doi: 10.1002/pds.1845] [Medline: 19757416]
7. Tadrous M, Gagne JJ, Stürmer T, Cadarette SM. Disease risk score as a confounder summary method: systematic review and recommendations. Pharmacoepidemiol Drug Saf 2013;22(2):122-129 [FREE Full text] [doi: 10.1002/pds.3377] [Medline: 23172692]
8. Stürmer T, Schneeweiss S, Brookhart MA, Rothman KJ, Avorn J, Glynn RJ. Analytic strategies to adjust confounding using exposure propensity scores and disease risk scores: nonsteroidal antiinflammatory drugs and short-term mortality in the elderly. Am J Epidemiol 2005;161(9):891-898 [FREE Full text] [doi: 10.1093/aje/kwi106] [Medline: 15840622]
9. Glynn RJ, Gagne JJ, Schneeweiss S. Role of disease risk scores in comparative effectiveness research with emerging therapies. Pharmacoepidemiol Drug Saf 2012;21 Suppl 2(Suppl 2):138-147 [FREE Full text] [doi: 10.1002/pds.3231] [Medline: 22552989]
10. Thompson MG, Stenehjem E, Grannis S, Ball SW, Naleway AL, Ong TC, et al. Effectiveness of Covid-19 vaccines in ambulatory and inpatient care settings. N Engl J Med 2021;385(15):1355-1371. [doi: 10.1056/nejmoa2110362] [Medline: 34496194]
11. Cates J, Lucero-Obusan C, Dahl RM, Schirmer P, Garg S, Oda G, et al. Risk for in-hospital complications associated with COVID-19 and influenza - veterans health administration, United States, October 1, 2018-May 31, 2020. MMWR Morb Mortal Wkly Rep 2020;69(42):1528-1534 [FREE Full text] [doi: 10.15585/mmwr.mm6942e3] [Medline: 33090987]
12. Garg S, Kim L, Whitaker M, O'Halloran A, Cummings C, Holstein R, et al. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019 - COVID-NET, 14 States, March 1-30, 2020. MMWR Morb Mortal Wkly Rep 2020;69(15):458-464 [FREE Full text] [doi: 10.15585/mmwr.mm6915e3] [Medline: 32298251]
13. Kim L, Garg S, O'Halloran A, Whitaker M, Pham H, Anderson EJ, et al. Risk factors for intensive care unit admission and in-hospital mortality among hospitalized adults identified through the US coronavirus disease 2019 (COVID-19)-associated

XSL•FO
RenderX

hospitalization surveillance network (COVID-NET). Clin Infect Dis 2021;72(9):e206-e214 [FREE Full text] [doi: 10.1093/cid/ciaa1012] [Medline: 32674114]
14. Chua H, Feng S, Lewnard JA, Sullivan SG, Blyth CC, Lipsitch M, et al. The use of test-negative controls to monitor vaccine effectiveness: a systematic review of methodology. Epidemiology 2020;31(1):43-64 [FREE Full text] [doi: 10.1097/EDE.0000000000001116] [Medline: 31609860]
15. Foppa IM, Ferdinands JM, Chaves SS, Haber MJ, Reynolds SB, Flannery B, et al. The case test-negative design for studies of the effectiveness of influenza vaccine in inpatient settings. Int J Epidemiol 2016;45(6):2052-2059. [doi: 10.1093/ije/dyw022] [Medline: 26979985]
16. Jackson ML, Nelson JC. The test-negative design for estimating influenza vaccine effectiveness. Vaccine 2013;31(17):2165-2168. [doi: 10.1016/j.vaccine.2013.02.053] [Medline: 23499601]
17. Joffe MM, Rosenbaum PR. Invited commentary: propensity scores. Am J Epidemiol 1999;150(4):327-333. [doi: 10.1093/oxfordjournals.aje.a010011] [Medline: 10453808]
18. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70(1):41-55. [doi: 10.1093/biomet/70.1.41]
19. D'Agostino RB. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med 1998;17(19):2265-2281. [doi: 10.1002/(sici)1097-0258(19981015)17:19<2265::aid-sim918>3.0.co;2-b] [Medline: 9802183]
20. McCaffrey DF, Ridgeway G, Morral AR. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. Psychol Methods 2004;9(4):403-425. [doi: 10.1037/1082-989X.9.4.403] [Medline: 15598095]
21. Rosenblum HG, Wallace M, Godfrey M, Roper LE, Hall E, Fleming-Dutra KE, et al. Interim recommendations from the advisory committee on immunization practices for the use of bivalent booster doses of COVID-19 vaccines - United States, October 2022. MMWR Morb Mortal Wkly Rep 2022;71(45):1436-1441 [FREE Full text] [doi: 10.15585/mmwr.mm7145a2] [Medline: 36355612]
22. Mbaeyi S, Oliver SE, Collins JP, Godfrey M, Goswami ND, Hadler SC, et al. The advisory committee on immunization practices' interim recommendations for additional primary and booster doses of COVID-19 vaccines - United States, 2021. MMWR Morb Mortal Wkly Rep 2021;70(44):1545-1552 [FREE Full text] [doi: 10.15585/mmwr.mm7044e2] [Medline: 34735422]
23. Oliver SE, Gargano JW, Marin M, Wallace M, Curran KG, Chamberland M, et al. The advisory committee on immunization practices' interim recommendation for use of Pfizer-BioNTech COVID-19 vaccine - United States, December 2020. MMWR Morb Mortal Wkly Rep 2020;69(50):1922-1924 [FREE Full text] [doi: 10.15585/mmwr.mm6950e2] [Medline: 33332292]
24. Bozio CH, Grannis SJ, Naleway AL, Ong TC, Butterfield KA, DeSilva MB, et al. Laboratory-confirmed COVID-19 among adults hospitalized with COVID-19-like illness with infection-induced or mRNA vaccine-induced SARS-CoV-2 Immunity - Nine States, January-September 2021. MMWR Morb Mortal Wkly Rep 2021;70(44):1539-1544 [FREE Full text] [doi: 10.15585/mmwr.mm7044e1] [Medline: 34735425]
25. Tenforde MW, Weber ZA, Natarajan K, Klein NP, Kharbanda AB, Stenehjem E, et al. Early estimates of bivalent mRNA vaccine effectiveness in preventing COVID-19-associated emergency department or urgent care encounters and hospitalizations among immunocompetent adults - VISION network, nine states, September-November 2022. MMWR Morb Mortal Wkly Rep 2022;71(5152):1616-1624 [FREE Full text] [doi: 10.15585/mmwr.mm715152e1] [Medline: 36580430]
26. Ferdinands JM, Rao S, Dixon BE, Mitchell PK, DeSilva MB, Irving SA, et al. Waning of vaccine effectiveness against moderate and severe COVID-19 among adults in the US from the VISION network: test negative, case-control study. BMJ 2022;379:e072141 [FREE Full text] [doi: 10.1136/bmj-2022-072141] [Medline: 36191948]
27. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol 2008;168(6):656-664 [FREE Full text] [doi: 10.1093/aje/kwn164] [Medline: 18682488]
28. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med 2015;34(28):3661-3679 [FREE Full text] [doi: 10.1002/sim.6607] [Medline: 26238958]
29. Wyss R, Glynn RJ, Gagne JJ. A review of disease risk scores and their application in pharmacoepidemiology. Curr Epidemiol Rep 2016;3(4):277-284. [doi: 10.1007/s40471-016-0088-2]
30. Hansen BB. The prognostic analogue of the propensity score. Biometrika 2008;95(2):481-488 [FREE Full text] [doi: 10.1093/biomet/asn004]
31. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989;81(24):1879-1886. [doi: 10.1093/jnci/81.24.1879] [Medline: 2593165]
32. Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation 1998;97(18):1837-1847. [doi: 10.1161/01.cir.97.18.1837] [Medline: 9603539]
33. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute physiology and chronic health evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. Crit Care Med 2006;34(5):1297-1310. [doi: 10.1097/01.CCM.0000215112.84523.F0] [Medline: 16540951]

XSL•FO
RenderX

34. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]
35. Andrews N, Stowe J, Kirsebom F, Toffa S, Rickeard T, Gallagher E, et al. Covid-19 vaccine effectiveness against the Omicron (B.1.1.529) variant. N Engl J Med 2022;386(16):1532-1546 [FREE Full text] [doi: 10.1056/NEJMoa2119451] [Medline: 35249272]
36. Andrews N, Stowe J, Kirsebom F, Toffa S, Sachdeva R, Gower C, et al. Effectiveness of COVID-19 booster vaccines against COVID-19-related symptoms, hospitalization and death in England. Nat Med 2022;28(4):831-837 [FREE Full text] [doi: 10.1038/s41591-022-01699-1] [Medline: 35045566]
37. Andrews N, Tessier E, Stowe J, Gower C, Kirsebom F, Simmons R, et al. Duration of protection against mild and severe disease by Covid-19 vaccines. N Engl J Med 2022;386(4):340-350 [FREE Full text] [doi: 10.1056/NEJMoa2115481] [Medline: 35021002]
38. Koehler E, Brown E, Haneuse SJA. On the assessment of Monte Carlo error in simulation-based statistical analyses. Am Stat 2009;63(2):155-162 [FREE Full text] [doi: 10.1198/tast.2009.0030] [Medline: 22544972]
39. Wyss R, Hansen BB, Ellis AR, Gagne JJ, Desai RJ, Glynn RJ, et al. The "dry-run" analysis: a method for evaluating risk scores for confounding control. Am J Epidemiol 2017;185(9):842-852 [FREE Full text] [doi: 10.1093/aje/kwx032] [Medline: 28338910]
40. Peters CC. A method of matching groups for experiment with no loss of population. J Ed Res 1941;34(8):606-612 [FREE Full text]
41. Belson WA. A technique for studying the effects of a television broadcast. Applied Statistics 1956 Nov;5(3):195-202. [doi: 10.2307/2985420]
42. Britton A, Embi PJ, Levy ME, Gaglani M, DeSilva MB, Dixon BE, et al. Effectiveness of COVID-19 mRNA vaccines against COVID-19-associated hospitalizations among immunocompromised adults during SARS-CoV-2 Omicron predominance - VISION network, 10 States, December 2021-August 2022. MMWR Morb Mortal Wkly Rep 2022;71(42):1335-1342 [FREE Full text] [doi: 10.15585/mmwr.mm7142a4] [Medline: 36264840]
43. Brookhart MA, Wyss R, Layton JB, Stürmer T. Propensity score methods for confounding control in nonexperimental research. Circ Cardiovasc Qual Outcomes 2013;6(5):604-611 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.113.000359] [Medline: 24021692]
44. Shapiro AE, Bender Ignacio RA, Whitney BM, Delaney JA, Nance RM, Bamford L, CFAR Network of Integrated Clinical Systems. Factors associated with severity of COVID-19 disease in a multicenter cohort of people With HIV in the United States, March-December 2020. J Acquir Immune Defic Syndr 2022;90(4):369-376. [doi: 10.1097/QAI.0000000000002989] [Medline: 35364600]
45. Huh K, Ji W, Kang M, Hong J, Bae GH, Lee R, et al. Association of prescribed medications with the risk of COVID-19 infection and severity among adults in South Korea. Int J Infect Dis 2021;104:7-14 [FREE Full text] [doi: 10.1016/j.ijid.2020.12.041] [Medline: 33352326]
46. R: The R Project for Statistical Computing.: R Foundation for Statistical Computing URL: https://www.R-project.org/
47. Pike MC, Anderson J, Day N. Some insights into Miettinen's multivariate confounder score approach to case-control study analysis. Epidemiol Community Health 1979;33(1):104-106 [FREE Full text] [doi: 10.1136/jech.33.1.104] [Medline: 467396]
48. Cook EF, Goldman L. Performance of tests of significance based on stratification by a multivariate confounder score or by a propensity score. J Clin Epidemiol 1989;42(4):317-324. [doi: 10.1016/0895-4356(89)90036-x] [Medline: 2723692]

Abbreviations

CDC: Centers for Disease Control and Prevention
CLI: COVID-19–like illness
DRS: disease risk score
PS: propensity score
TND: test-negative design
VE: vaccine effectiveness
VISION: Virtual SARS-CoV-2, Influenza, and Other Respiratory Viruses Network

XSL•FO
RenderX

Edited by A Mavragani; submitted 08.04.24; peer-reviewed by N Andrews; comments to author 02.10.24; revised version received 22.10.24; accepted 10.11.24; published 27.01.25.

Please cite as:
Rowley EAK, Mitchell PK, Yang DH, Lewis N, Dixon BE, Vazquez-Benitez G, Fadel WF, Essien IJ, Naleway AL, Stenehjem E, Ong TC, Gaglani M, Natarajan K, Embi P, Wiegand RE, Link-Gelles R, Tenforde MW, Fireman B
Methods to Adjust for Confounding in Test-Negative Design COVID-19 Effectiveness Studies: Simulation Study
JMIR Form Res 2025;9:e58981
URL: https://formative.jmir.org/2025/1/e58981
doi:10.2196/58981
PMID:

©Elizabeth AK Rowley, Patrick K Mitchell, Duck-Hye Yang, Ned Lewis, Brian E Dixon, Gabriela Vazquez-Benitez, William F Fadel, Inih J Essien, Allison L Naleway, Edward Stenehjem, Toan C Ong, Manjusha Gaglani, Karthik Natarajan, Peter Embi, Ryan E Wiegand, Ruth Link-Gelles, Mark W Tenforde, Bruce Fireman. Originally published in JMIR Formative Research (https://formative.jmir.org), 27.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Multimodal Pain Recognition in Postoperative Patients: Machine Learning Approach

Ajan Subramanian[1], MS; Rui Cao[2], MS; Emad Kasaeyan Naeini[1], PhD; Seyed Amir Hossein Aqajari[2], PhD; Thomas D Hughes[3], RN; Michael-David Calderon[4], MS; Kai Zheng[5], PhD; Nikil Dutt[1], PhD; Pasi Liljeberg[6], PhD; Sanna Salanterä[7,8], RN, PhD; Ariana M Nelson[9], MD; Amir M Rahmani[1,2,3,10], MBA, PhD

[1]Department of Computer Science, University of California, Irvine, Irvine, CA, United States

[2]Department of Electrical Engineering and Computer Science, University of California, Irvine, Irvine, CA, United States

[3]School of Nursing, University of California, Irvine, Irvine, CA, United States

[4]College of Medicine, Kansas City University, Kansas City, MO, United States

[5]Department of Informatics, University of California, Irvine, Irvine, CA, United States

[6]Department of Computing, University of Turku, Turku, Finland

[7]Department of Nursing Science, University of Turku, Turku, Finland

[8]Turku University Hospital, University of Turku, Turku, Finland

[9]Department of Anesthesiology and Pain Medicine, University of California, Irvine, Irvine, CA, United States

[10]Institute for Future Health, University of California, Irvine, Irvine, CA, United States

Corresponding Author:
Ajan Subramanian, MS
Department of Computer Science
University of California, Irvine
3211 Donald Bren Hall
Irvine, CA, 92617
United States
Phone: 1 6506604994
Email: ajans1@uci.edu

Abstract

Background: Acute pain management is critical in postoperative care, especially in vulnerable patient populations that may be unable to self-report pain levels effectively. Current methods of pain assessment often rely on subjective patient reports or behavioral pain observation tools, which can lead to inconsistencies in pain management. Multimodal pain assessment, integrating physiological and behavioral data, presents an opportunity to create more objective and accurate pain measurement systems. However, most previous work has focused on healthy subjects in controlled environments, with limited attention to real-world postoperative pain scenarios. This gap necessitates the development of robust, multimodal approaches capable of addressing the unique challenges associated with assessing pain in clinical settings, where factors like motion artifacts, imbalanced label distribution, and sparse data further complicate pain monitoring.

Objective: This study aimed to develop and evaluate a multimodal machine learning–based framework for the objective assessment of pain in postoperative patients in real clinical settings using biosignals such as electrocardiogram, electromyogram, electrodermal activity, and respiration rate (RR) signals.

Methods: The iHurt study was conducted on 25 postoperative patients at the University of California, Irvine Medical Center. The study captured multimodal biosignals during light physical activities, with concurrent self-reported pain levels using the Numerical Rating Scale. Data preprocessing involved noise filtering, feature extraction, and combining handcrafted and automatic features through convolutional and long-short-term memory autoencoders. Machine learning classifiers, including support vector machine, random forest, adaptive boosting, and k-nearest neighbors, were trained using weak supervision and minority oversampling to handle sparse and imbalanced pain labels. Pain levels were categorized into baseline and 3 levels of pain intensity (1-3).

Results: The multimodal pain recognition models achieved an average balanced accuracy of over 80% across the different pain levels. RR models consistently outperformed other single modalities, particularly for lower pain intensities, while facial muscle activity (electromyogram) was most effective for distinguishing higher pain intensities. Although single-modality models,

XSL•FO
RenderX

especially RR, generally provided higher performance compared to multimodal approaches, our multimodal framework still delivered results that surpassed most previous works in terms of overall accuracy.

Conclusions: This study presents a novel, multimodal machine learning framework for objective pain recognition in postoperative patients. The results highlight the potential of integrating multiple biosignal modalities for more accurate pain assessment, with particular value in real-world clinical settings.

(JMIR Form Res 2025;9:e67969) doi:10.2196/67969

KEYWORDS

pain intensity recognition; multimodal information fusion; signal processing; weak supervision; health care; pain intensity; pain recognition; machine learning approach; acute pain; pain assessment; behavioral pain; pain measurement; pain monitoring; multimodal machine learning–based framework; machine learning–based framework; electrocardiogram; electromyogram; electrodermal activity; self-reported pain level; clinical pain management

Introduction

Pain is defined by the International Association for the Study of Pain as "an unpleasant sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage" [1]. Pain is a unique phenomenon that individuals experience and perceive independently. Younger et al [2] stated that pain is a subjective experience for which there is no current objective measure. Pain may be classified as either acute or chronic; Kent et al [3] described acute pain as encompassing the immediate, time-limited bodily response to a noxious stimulus that triggers actions to avoid or mitigate ongoing injury. Chronic pain was first defined loosely by Bonica [4] as pain that extends beyond an expected timeframe; currently, chronic pain is defined as "persistent or recurrent pain lasting longer than three months" [5]. The focus of this paper is on acute pain.

Acute pain is a common experience in the postanesthesia care unit in the immediate period following surgery. According to Chou et al [6], pain occurs in 80% of patients following surgery, and 75% of patients with pain report their pain as either moderate, severe, or extreme. Current guidelines for the assessment of pain in the postanesthesia care unit recommend using a Numerical Rating Scale (NRS) or Verbal Rating Scale for patients who are sufficiently awake and coherent to reliably report pain scores [7]. However, Herr et al [8] identified several patient populations who are at risk for being incapable of providing self-report scores of pain; specifically, these populations include the pediatric population who have yet to develop adequate cognition; older patients with dementia; individuals with intellectual disabilities; and those who are unconscious, critically ill, or terminally ill. In these patient populations, Small and Laycock [7] recommend the use of behavioral pain scales, such as the Pain Assessment in Advanced Dementia, Critical Care Pain Observation Tool (CPOT), or Behavioral Pain Scale. Despite the pain assessment measures of self-report and behavioral pain scales, each of these methods may be prone to biases. For example, Craig et al [9] discussed how self-reporting might be a means to obtain a particular goal that can be influenced by the individual reporting pain. In addition, Hadjistavropoulos and Craig [10] provided the Communications Model of Pain, which provided a basis for how expressive behaviors are decoded by observers of individuals in pain, which are influenced by the message clarity transmitted by the individual in pain as well as the unique biases (eg, knowledge level, assessment skills, and predisposing beliefs) of the individual assessing pain. The difficult nature of interpreting pain scores has resulted in disparities in pain management in minority populations, with research by Staton et al [11] showing that the Black race is a significant predictor of the underestimation of pain by physicians.

Multimodal pain assessment represents a potential method of circumventing the limitations of traditional self-report and behavioral pain assessment tools and an opportunity for enhancing pain assessment in vulnerable populations. Instead of having to rely on only one dimension of pain assessment, such as behaviors through the use of the CPOT or Behavioral Pain Scale, future multimodal pain assessment will incorporate physiological indicators, such as electrodermal activity (EDA), electrocardiogram (ECG), electroencephalogram, and electromyogram (EMG) as well as behaviors (eg, facial expression), and perhaps other as-yet undiscovered parameters to capture pain assessment in patient populations that might not be best represented by current assessment strategies. For example, a study by Gélinas et al [12] found that revisions to the CPOT were necessary because some brain-injured patients may not exhibit certain behaviors that are contained in the CPOT. Similarly, for individuals diagnosed with dementia, Achterberg et al [13] stated that there is a preponderance of observer-based pain assessment tools, however, these tools retain significant differences between them, as well as concerns for lack of reliability, validity, and sensitivity of change. Enhancing pain assessment through the combination of traditional pain assessment methods with novel multimodal approaches may serve to eventually enhance pain assessment in a greater majority of vulnerable patient populations.

With the advent of connected Internet of Things devices and wearable sensor technology, automated data collection may achieve continuous pain intensity measurement. A significant amount of research has been conducted in recent years, which has sought to develop methods of continuous, automatic, and multimodal pain assessment. For example, previous work conducted by Walter et al [14] and Werner et al [15] used skin conductance level, ECG, electroencephalogram, and EMG to monitor pain in response to thermal pain. Other works, such as Hammal and Cohn [16] and Werner et al [17], have incorporated facial expression monitoring as an indicator of pain. While these studies were immensely beneficial to the scientific community

XSL•FO
RenderX

in terms of their contributions to a better understanding of techniques to obtain continuous pain assessment, the setting of these experiments was in highly controlled laboratory environments with healthy participants. Collecting data in real-world situations as opposed to a laboratory setting would allow the researchers to assess a pain assessment technique's potential in relation to actual pain brought about through a surgical procedure instead of induced pain.

The aim of this study is to develop a robust and effective multimodal pain assessment framework for postoperative patients in real clinical settings. To the best of our knowledge, this is the first work proposing a multimodal pain assessment framework for postoperative patients. It should be noted that a pain assessment study on real patients is associated with several challenges (eg, imbalanced label distribution, missing data, motion artifacts, etc) since several parameters such as the intensity, distribution, frequency, and time of the pain as well as the environment cannot be controlled by researchers. Our main contributions are 4-fold:

1. We conducted a clinical study for multimodal signal acquisition from an acute pain unit of the University of California, Irvine Medical Center.
2. We propose a multimodal pain assessment framework using our database (iHurt Pain DB) collected from postoperative patients while obtaining a higher accuracy compared to existing works on healthy participants [17].
3. We use both handcrafted and automatically generated features outputted from deep learning networks to build our models.
4. We provide a novel method to mitigate the presence of sparse and imbalanced labels (due to the real clinical setting of the study) using weak supervision and minority oversampling.

Methods

Overview

Candidates were selected from the Acute Pain Service patient list at University of California Irvine Health in Orange, California. The Acute Pain Service unit at the medical center serves approximately 100 patients weekly, enabling the lead Doctor of Medicine to recruit patients. This is the first claimed study that collected biosignals from postoperative adult patients in hospitals. All participants (aged 23-89 years) were recruited to the study from July 2018 to October 2019.

iHurt Study Design

We conducted a biomedical data collection study on 25 postoperative patients reporting various degrees of pain symptoms. Multimodal biosignals (ECG, EMG, EDA, and photoplethysmography [PPG]) were collected from patients likely having mild to moderate pain who were asked to perform a few light physical activities while acquiring data. We also collected primary demographic information from each patient, including height, weight, sex, and BMI. All signals were collected using the iHurt system.

iHurt System

iHurt is a system that measures facial muscle activity (ie, changes in facial expression) in conjunction with physiological signals such as heart rate, heart rate variability, respiration rate (RR), and EDA for the purpose of developing an algorithm for pain assessment in hospitalized patients. The system uses the following 2 components to capture raw signals.

Eight-Channel Biopotential Acquisition Device

Our team at the University of Turku, Finland, developed a biopotential acquisition device to measure ECG and EMG signals. The device incorporates commercially available electrodes, electrode-to-device lead wires, an ADS1299-based portable device, and computer software (LabVIEW version 14.02f, National Instruments) to visualize data streaming from the portable device. Raw signals from the electrodes are sampled at 500 samples per second and are sent to the computer software through Bluetooth for visualization [18].

Empatica E4

We use the commercially available Empatica E4 wristband (Empatica Inc) [19] to measure EDA and PPG signals. The purpose of using a wristband was to allow our participants to move freely without any impediments. The Empatica E4 was connected to the participants' phones over Bluetooth for visualization.

We removed 3 participants' data from the final dataset due to the presence of excessive motion artifacts. We also excluded 2 additional patients since they were wearing the Empatica E4 watch on their arm that received intravenous medication. This resulted in unreliable EDA signals due to conditions like skin rash and itching. This left us with data from 20 patients to build our pain recognition system. The dataset also contains rich annotation with self-reported pain scores based on the 11-point NRS from 0 to 10. A detailed explanation of the dataset and the study design can be found in Kasaeyan Naeini et al [20]. We intend to make the deidentified dataset available to the research community for further analysis and applications.

Data Processing Pipeline

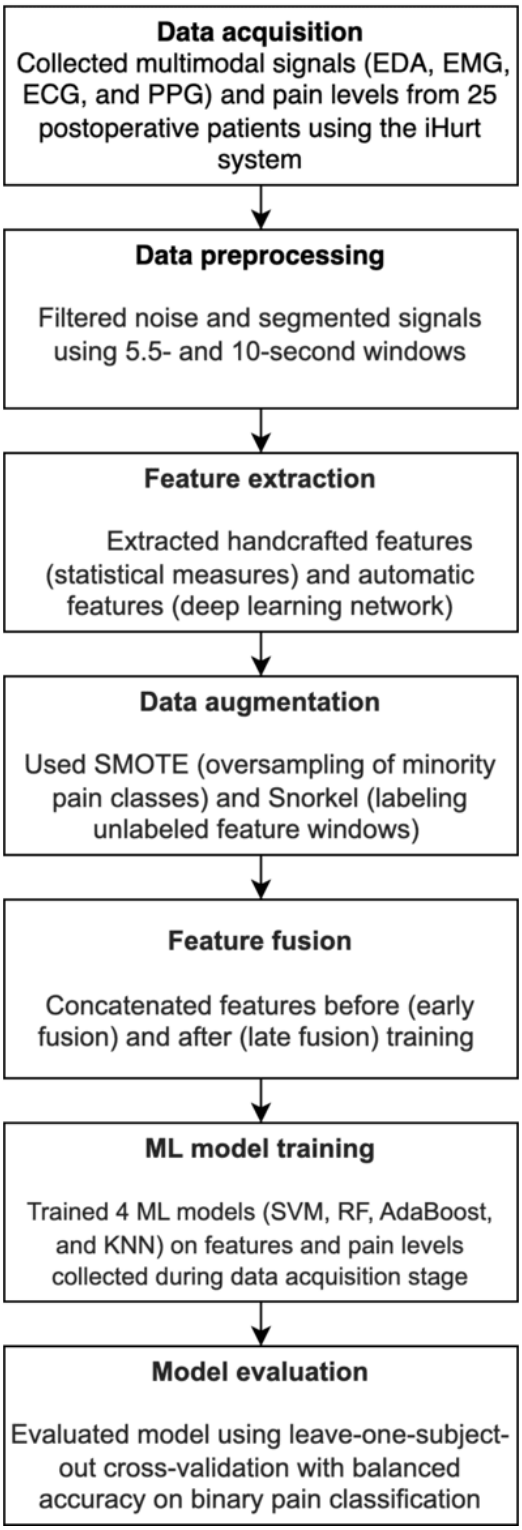
The first step in building our multimodal pain assessment system was to process the raw signals collected during trials. The data processing pipeline consisted of the following steps:

1. We filtered the signal to remove powerline interference, baseline wander, and motion artifact noise.
2. We performed feature extraction on the filtered signals to obtain amplitude and variability features in the time domain. The time domain features were extracted using 5.5-second and 10-second windows. The 5.5-second window size was extracted to be compared with previous work [17].
3. In addition to handcrafted features, we also used automatic features, which were outputted from a deep neural network.
4. Once the features were extracted, we tagged them with their corresponding labels based on the nearest timestamp of the label.
5. Each of these processing steps was applied individually to each of the 4 modalities. Processed data from each of the modalities were combined using either early fusion or late

XSL•FO
RenderX

fusion. The types of handcrafted features extracted from each modality and the deep learning pipeline for extracting automatic features are described in detail. An overview of our method is described as a flowchart in Figure 1.

Figure 1. Overview of the proposed method. AdaBoost: adaptive boosting; EDA: electrodermal activity; EMG: electromyogram; ECG: electrocardiogram; ML: machine learning; PPG: photoplethysmography; SMOTE: synthetic minority oversampling technique; SVM: support vector machine; RF: random forest; KNN: k-nearest neighbors.



ECG Handcrafted Features

The ECG channel was filtered using a Butterworth band-pass filter with a frequency range of 0.1-250 Hz. The heart rate variability handcrafted features were extracted with *pyHRV*, an open-source Python (Python Software Foundation) toolbox [21], using the R-peaks extracted from the ECG signal through a bidirectional long short-term memory (LSTM) network [22]. These features were extracted from two window sizes, 5.5 and 10 seconds. There were 19 time-domain features. The

XSL•FO
RenderX

time-domain features extracted from NN intervals, or the time interval between successive R-peaks, comprised of the slope of these intervals, 5 statistical features (total count, mean, minimum, maximum, and SD), 9 difference features (mean difference, minimum difference, maximum difference, SD of successive interval differences, root mean square of successive interval differences, number of interval differences greater than 20 ms and 50 ms, and percentage of successive interval differences that differ by more than 20 ms and 50 ms), and 4 heart rate features (mean, minimum, maximum, and SD) [23].

EMG Handcrafted Features

The preprocessing phase of EMG channels comprised a 20 Hz high pass filter and two notch filters at 50 Hz and 100 Hz, all using a Butterworth filter. Like ECG features, we extracted EMG features from 5.5- and 10-second windows on 5 different channels for each major facial muscle. The ten amplitude features extracted were (1) peak, (2) peak-to-peak mean value, (3) root mean squared, (4) mean of the absolute values of the second differences, (5) mean of the absolute values of the first differences, (6) mean of the absolute values of the second differences of the normalized signal, (7) mean of the absolute values of the first differences of the normalized signal, (8) mean of local minima values, (9) mean of local maxima values, and (10) mean of absolute values. The four variability features were (1) variance, (2) SD, (3) range, and (4) IQR. All 14 features were calculated for 5 different EMG channels, resulting in 70 EMG features in total.

EDA Handcrafted Features

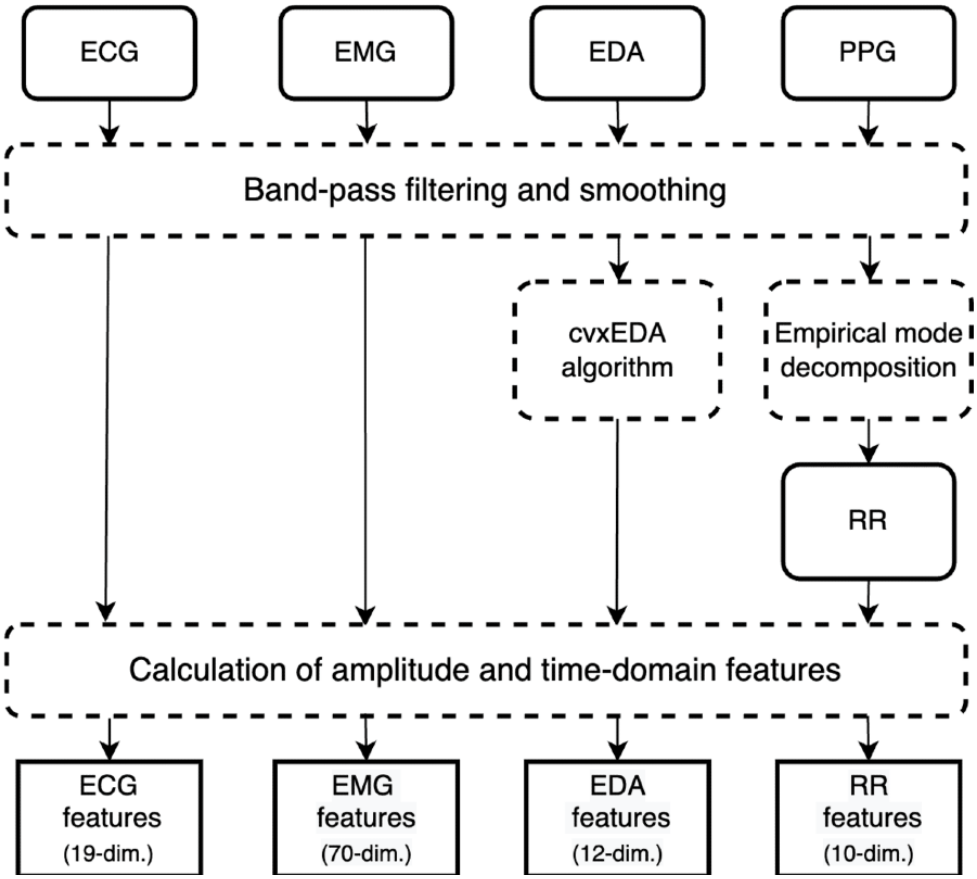
We used the *pyEDA* library [24] for preprocessing and feature extraction of EDA signals. In the preprocessing part, first, we used a moving average across a 1-second window to remove the motion artifacts and smooth the data [25]. Second, a low-pass Butterworth filter on the phasic data was applied to remove the line noise. Finally, preprocessed EDA signals corresponding to each different pain level were visualized to ensure the validity of the signals. In the feature extraction part, the *cvxEDA* algorithm [26] was used to extract the phasic component of EDA signals. The EDA signals' peaks or bursts are considered variations in the phasic component of the signal. Therefore, the clean signals and extracted phasic component of signals were fed to the statistical feature extraction module to extract the number of peaks, the average value, and the maximum and minimum value of the signals. Furthermore, these extracted features were further used in the post–feature extraction module to extract eight more features: (1) the difference between the maximum and the minimum value of the signal, (2) the SD, (3) the difference between the upper and lower quartiles (4) root mean square, (5) the mean value of local minima, (6) the mean value of local maxima, (7) the mean of the absolute values of the first differences, and (8) the mean of the absolute values of the second differences. This resulted in 12 EDA features in total.

PPG-Based RR Handcrafted Features

We preprocessed the PPG signal before extracting the RR from it. In total, 2 filters were used during the preprocessing [27]. We first used a Butterworth band-pass filter to remove noises, including motion artifacts. Then, a moving average filter was implemented to smooth the PPG signal. After that, we applied an empirical mode decomposition–based method proposed by Madhav et al [28] to derive respiration signals from filtered PPG signals. This method was proven to derive RR from a PPG signal with high accuracy (99.87%). A total of ten features were extracted from the respiratory signal, including (1) the number of inhale peaks, (2) the mean value of the signal, (3) the maximum value, (4) the minimum value, (5) the difference between the maximum and the minimum value, (6) SD, (7) the average value of the inhale peak intervals, (8) the SD of the inhale peak intervals, (9) the root mean square of successive differences between adjacent inhale peak intervals, (10) SD of inhale duration. A visualization of the handcrafted feature pipeline is shown in Figure 2.

XSL•FO
RenderX

Figure 2. Handcrafted feature extraction pipeline. dim.: dimension; ECG: electrocardiogram; EDA: electrodermal activity; EMG: electromyogram; PPG: photoplethysmography; RR: respiration rate.



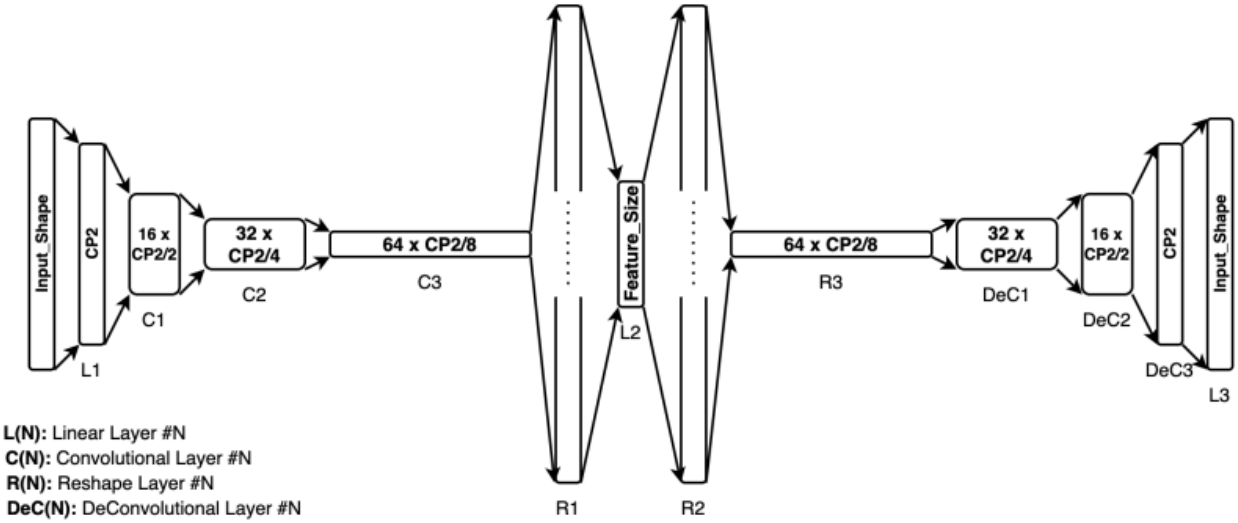
Automatic Feature Extraction Pipeline

As the dimensionality of biomedical data increases, it becomes increasingly difficult to train a machine learning algorithm on the entire uncompressed dataset. This often leads to a large training time and is computationally more expensive overall. A possible solution is to perform feature engineering to get a compressed and interpretable representation of the signal. Another alternative approach, however, is to use the compressed or latent representation of that data obtained from deep learning networks trained for that specific task. Using automatic features helps in dimensionality reduction and can provide us with a sophisticated yet succinct representation of the data that handcrafted features alone cannot provide. This automatic feature extraction is typically carried out by an autoencoder (AE) network, which is an unsupervised neural network that learns how to efficiently compress and encode the data into a lower-dimensional space [29,30]. AEs are composed of 2 separate networks: an encoder and a decoder. The encoder network acts as a bottleneck layer and maps the input into a lower-dimensional feature space. The decoder network tries to reconstruct this lower-dimensional feature vector into the original input size. The entire network is trained to minimize the reconstruction loss (ie, mean-squared error) by iteratively updating its weights and biases through backpropagation.

A convolutional AE from the *pyEDA* library was used to extract automatic features. Figure 3 shows the architecture of the AE. First, a linear layer (L1) is used to downsample the input signal with *Input_Shape* length to a length that is the closest power of 2 (CP2). This was done to make the model scalable to an arbitrary input size. The encoder half of the network consists of three 1D convolutional layers (C1, C2, and C3) and a linear layer (L2), which flattens and downsamples the input vector to a lower-dimensional latent vector. The number of dimensions of this latent vector (Feature Size) corresponds to the number of automatic features extracted and was set prior to training the network. A total of 32 features were extracted from ECG, EDA, and RR signals, whereas a total of 30 features were extracted from the EMG signal (6 features from each of the 5 channels). The decoder half of the network consists of three 1D deconvolutional layers (DeC1, DeC2, and DeC3) to reconstruct the input signal from the latent vector. A final linear layer (L3) is then used to flatten and reconstruct the signal to its original dimension. Both encoder and decoder networks have rectified linear unit activation between layers. Window sizes of both 5.5 and 10 seconds were applied to the filtered signals. This was done to compare the performance with handcrafted features. After signals from each of the modalities were normalized, they were trained on separate AE models for each modality. In addition to the convolutional AE, we also extracted features from an LSTM AE network. This resulted in two different feature extraction methods (convolutional and LSTM) that spanned two different window lengths (5.5 and 10 seconds).

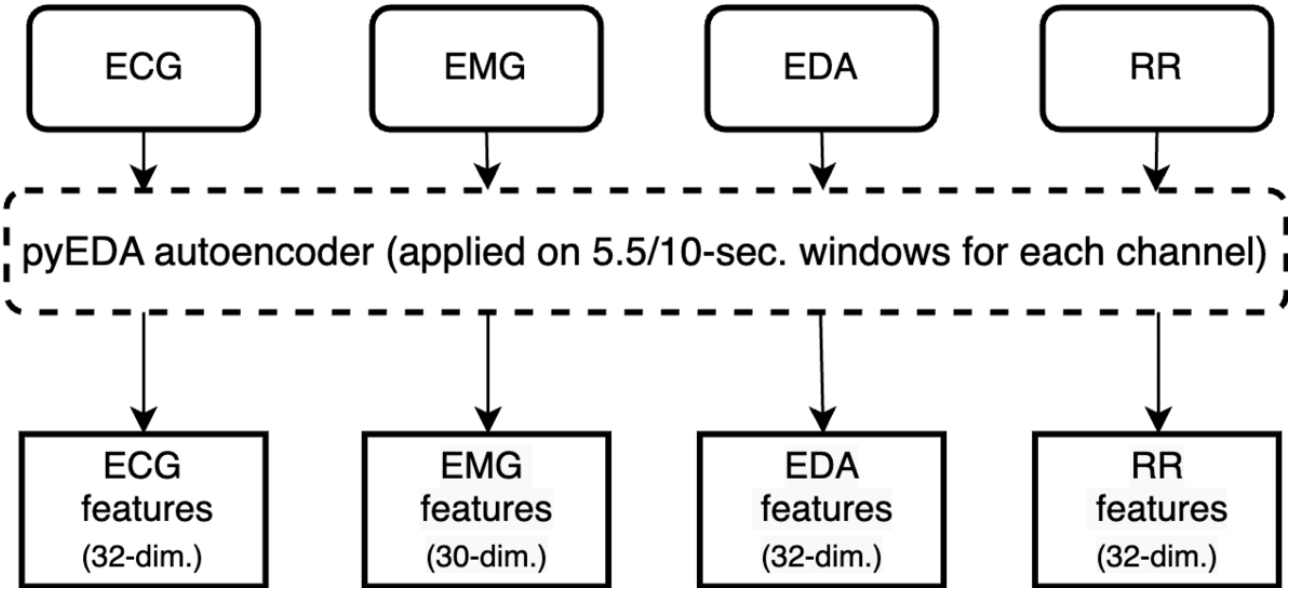
XSL•FO
RenderX

Figure 3. The architecture of the pyEDA convolutional autoencoder.



The batch size was set to 10, the number of training epochs was set to 100, and the ADAM optimizer [28] was used with a learning rate of 1×10^{-3}. A total of 126 feature vectors across all 4 modalities were extracted from each AE network. A visualization of our automatic feature extraction pipeline is shown in Figure 4.

Figure 4. Automatic feature extraction pipeline. dim.: dimension; ECG: electrocardiogram; EDA: electrodermal activity; EMG: electromyogram; RR: respiration rate; sec.: second.



Data Augmentation

Overview

There were several inherent challenges in the distribution of labels, as NRS values recorded during the clinical trials of this study were collected from real postoperative patients. This problem bears less significance while studying healthy participants since the stimulated pain can be controlled during the experiments. Consequently, occurrences of some pain levels far exceeded those of others. For example, among all patients, there were only 4 reported occurrences of pain level 10, whereas there were more than 80 reported occurrences of pain level 4. This imbalanced distribution was inevitable due to the subjective nature and the different sources of pain among the participants. Therefore, while downsampling our pain labels to 4 classes, thresholds for each downsampled class were carefully chosen to ensure a more evenly distributed set of labels. The pain levels ranged from a baseline (BL) level of pain or no pain to 3 increasing intensities of pain (PL 1-3). The thresholds for the pain levels were as follows: (1) PL1 ranged from 0 to 3, (2) PL2 ranged from 4 to 6, and (3) PL3 ranged from 7 to 10. All the ranges here are inclusive.

Since we asked patients to report their pain levels only while they performed pain-inducing activities, the number of labels generated was sparse. Both handcrafted and automatic features were combined with the corresponding labels using timestamps that were within the nearest 5.5 or 10 seconds (labeling threshold) of the reported NRS value. This depended on the window size of the features extracted. Due to having sparse labels, many of the feature windows were not assigned a corresponding label. To mitigate the problem of having an

XSL•FO
RenderX

imbalanced and sparse label distribution, 2 techniques were exploited:

Minority Oversampling

The first technique, called synthetic minority oversampling technique (SMOTE), is a type of data augmentation that oversamples the minority class [31]. SMOTE works by first choosing a minority class instance at random and finding its k-nearest minority class neighbors. It then creates a synthetic example at a randomly selected point between two instances of the minority class in that feature space. The experiments involving SMOTE were implemented using the *imbalanced-learn* Python library [32].

Weak Supervision

The second technique we used is weak supervision using the Snorkel framework [33]. Rather than employing an expert to manually label the unlabeled instances, Snorkel allows its users to write labeling functions that can make use of heuristics, patterns, external knowledge bases, and third-party machine learning models. Weak supervision is typically employed to label large volumes of unlabeled data when there are noisy, limited, or imprecise sources. For our pain assessment algorithm, we decided to use third-party machine learning models to label the remaining unlabeled instances. All the data points that were within the labeling threshold were considered as "strong labels," or ground-truth values collected from patients during trials. The remaining unlabeled data points were kept aside for Snorkel to provide a weakly supervised label. The strong labels were fed into Snorkel's labeling function consisting of 3 off-the-shelf machine learning models: (1) a support-vector machine (SVM) with a radial basis function kernel, (2) a random forest (RF) classifier, and (3) a k-nearest neighbor (KNN) classifier with uniform weights. Once each model was trained on the strong labels, it was used to make predictions on the remaining unlabeled data. The predictions from these 3 models were collected and converted into a single confidence-weighted label per data point using Snorkel's *LabelModel* function. This function outputs the most confident prediction as the label for each data point. To perform a fair assessment of the reliability and accuracy of our algorithm, we used SMOTE and Snorkel only while training our machine learning models. The performance of these models was measured solely on ground-truth (strong) labels collected during trials. This way, there is no implicit bias introduced from mislabeling or upsampling certain data points to skew model predictions.

Multimodal Machine Learning Models

To compare the performance of our multimodal machine learning models with the previous work, we performed binary classification using a leave-one-subject-out cross-validation approach [34]. In this method, a model's performance is validated over multiple folds in such a way that data from each patient are either in the training set or in the testing set. The purpose of using this method is to provide generalizability to unseen patients and to avoid overfitting by averaging the results over multiple folds. The eventual goal of this study is to build personalized models that make predictions on a single patient but learn from data collected from a larger population of similar patients. The following machine learning models were used to evaluate the performance of our pain assessment algorithm: (1) KNN, (2) RF classifier, (3) adaptive boosting (AdaBoost), (4) and an SVM. The models were then evaluated using leave-one subject-out cross-validation. Four separate models were trained for each of the 3 pain intensities (eg, BL; no pain versus PL1, the lowest pain level; or BL vs PL3, the highest pain level).

Fusing Modalities

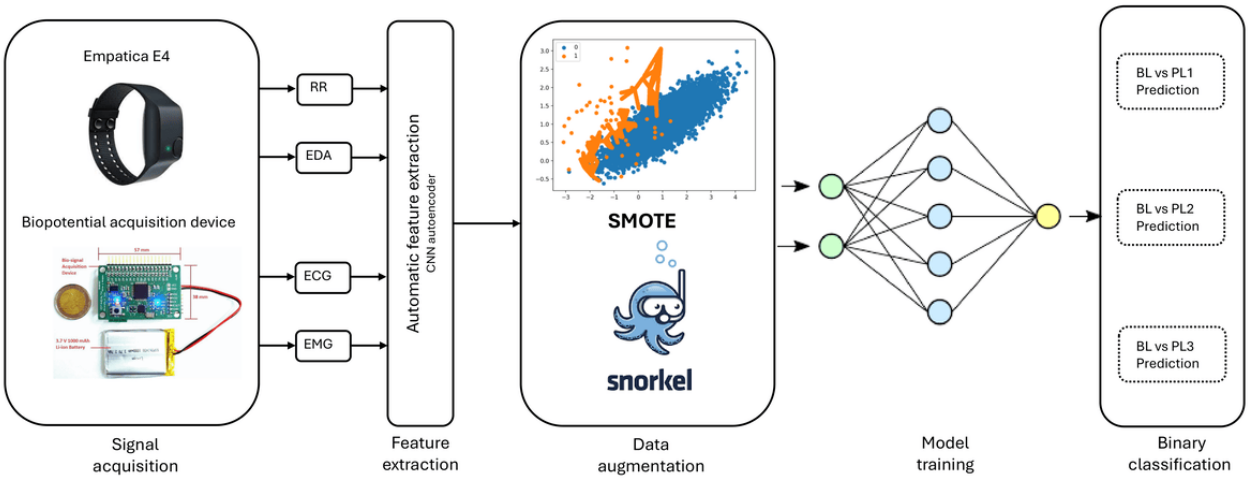
In total, 2 fusion approaches were used while combining features across different modalities. The first one is early or feature-level fusion, which concatenates feature vectors across different modalities based on their time stamps. The resulting data, which are now higher in dimension than any single modality, are then fed into our classifier to make predictions. While concatenating features across different modalities, a threshold of either 5.5 or 10 seconds was used to combine the modalities depending on the features extracted. The second approach was late or decision-level fusion, where each modality is fed to a separate classifier, and the final classification result is based on the fusion of outputs from the different modalities [35].

Feature Selection

Since there were a lot of features generated during the data processing phase, we had to select a subset of the most informative features to build our models with. Therefore, to reduce the complexity and training time of the resulting model, feature selection using Gini importance was performed. Gini is important as a lightweight method that is simple and fast to compute. Since we extracted a relatively large number of features in our method, it made sense to use a computationally low-cost algorithm for feature selection. We computed the Gini importance of the features from the data in the training fold with the help of a random forest classifier and selected the top 25 features. We then trained our model on these top 25 features and evaluated them in the validation fold. Our proposed multimodal pain recognition system is shown in Figure 5.

XSL•FO
RenderX

Figure 5. Proposed multimodal pain recognition system. BL: baseline; CNN: convolutional neural network; ECG: electrocardiogram; EDA: electrodermal activity; EMG: electromyogram; PL: pain level; RR: respiration rate; SMOTE: synthetic minority oversampling technique.



Ethical Considerations

The dataset used in this study was originally collected with approval from the Institutional Review Board (IRB) at the University of California, Irvine (Protocol HS# 2017-3747). Participants provided written informed consent after receiving detailed oral and written explanations of the study's objectives and procedures. They were encouraged to discuss participation with family and friends before consenting. Investigators ensured that all participants understood the study and had their questions answered prior to enrollment. Participants were informed of their right to withdraw at any time without impacting their care. For the secondary analysis conducted in this study, the IRB approval and original informed consent covered the reuse of the data, and no additional consent was required. All data utilized for this study were anonymized prior to analysis to protect participants' identities. Personal identifiers, such as names and contact information, were removed, and access to the data was restricted to authorized personnel only. The anonymized data were stored securely in compliance with institutional and regulatory guidelines to ensure confidentiality. Participation in the original study was entirely voluntary, and no compensation was provided. This ensured that participants' involvement was based solely on their willingness to contribute to the research.

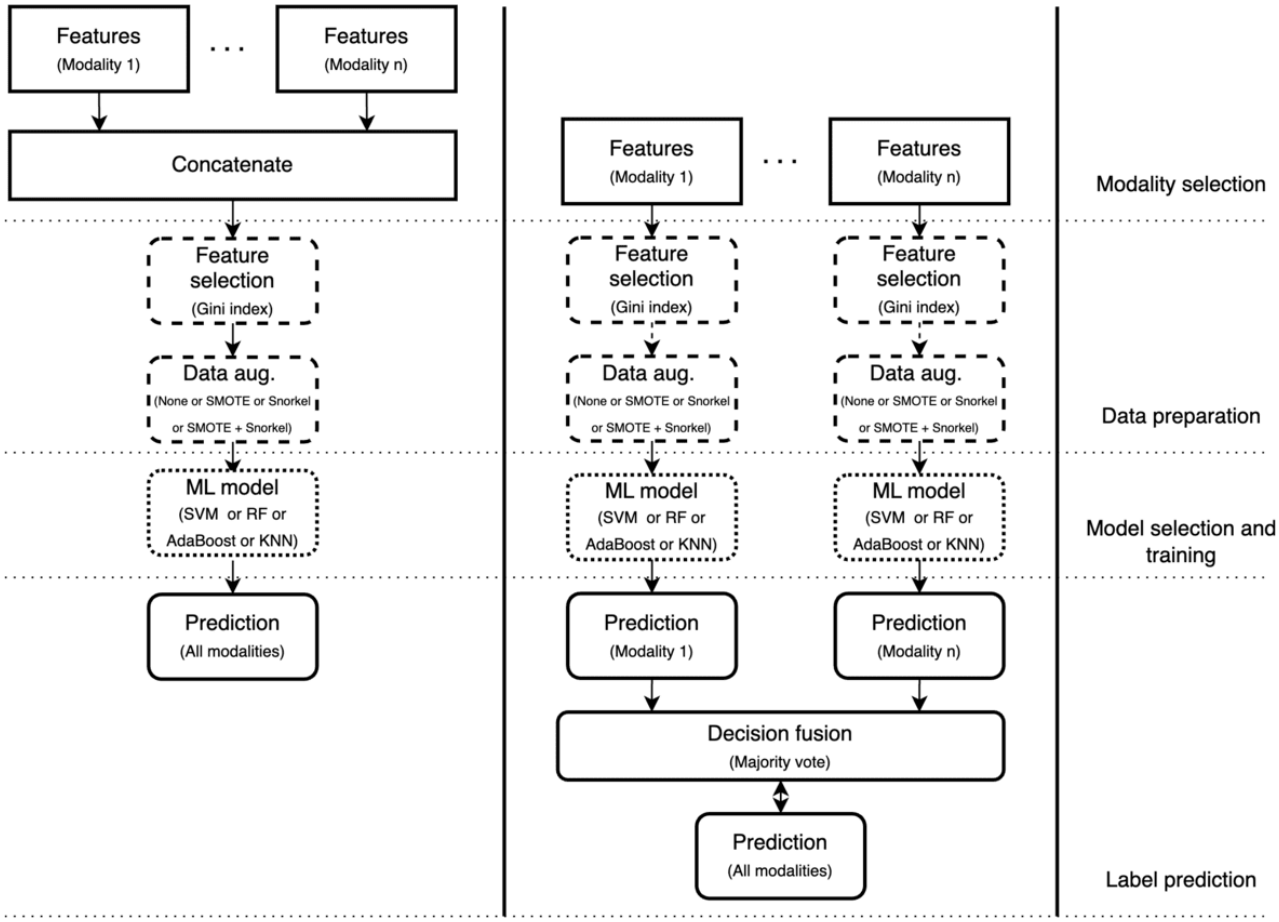
Results

Experimental Settings

The goal of our experiments was to compare the performance of using only a single modality to build our models over using a combination of multiple modalities. We trained several different models for each of the pain intensities, which varied in the types of modalities, data augmentation techniques, machine learning models, and fusion techniques used. Figure 6 shows the general pipeline of the experiments we conducted. We first selected the type of modalities to train on, which varied from only using each of the single modalities separately to using a combination of all 4 modalities. Furthermore, these modalities varied depending on the type of features used, like handcrafted or automatic features. In the case of using multiple modalities, we had 2 choices of fusion: early (Figure 6, left) and late (Figure 6, right). These architectures varied in how the modalities were combined, either before training (early) or at the decision level (late) after training using majority voting. The data preparation process involved feature selection and data augmentation. These models could either be trained with no data augmentation, with just SMOTE or Snorkel, or a combination of both. The last step of the pipeline before making predictions involved choosing the type of machine learning algorithms, like SVM, RF, AdaBoost, or KNN. Due to the lack of space, only the best-performing single and multimodal model configurations are mentioned in the section below.

XSL•FO
RenderX

Figure 6. Our proposed general multimodal pipeline-based on early fusion (left) and late fusion (right). AdaBoost: adaptive boosting; KNN: k-nearest neighbors; ML: machine learning; RF: random forest; SVM: support vector machine; SMOTE: synthetic minority oversampling technique.



Experimental Results

Tables 1 and 2 present the best-performing single-modal and multimodal models for each of the 3 pain intensities. For comparison, the best multimodal results from Werner et al [17], Lopez-Martinez and Picard [36], Wang et al [37], and Subramaniam and Dass [38] are also mentioned. We use balanced accuracy as an evaluation criterion because our dataset had an imbalanced class distribution. Balanced accuracy is defined as the average of the true positive rate and the true negative rate.

XSL•FO
RenderX

Table 1. Best scores: single modality versus multiple modalities.

Pain levels	ECG[a] scores	EMG[b] scores	EDA[c] scores	RR[d] scores	Multiple modality
BL[e] vs PL[f]1	82.14	86	79.18	84.62	82.14
BL vs PL2	86.11	84.53	82.94	88.24	86.11
BL vs PL3	75	78.12	75	76.23	75
Mean (SD)	81.08 (5.03)	82.8 (5.03)	79.04 (5.03)	83.03 (5.03)	81.08 (4.59)
Classifier configuration	LSTM[g] AE[h] (10 s), Strong, SVM[j]	HC[i] (10 s), Snorkel, SVM	CNN[k] AE (10 s), Strong, SVM	HC (10 s), Strong, SVM	EF[l], LSTM AE (10 s), Strong, SVM

[a]ECG: electrocardiogram.
[b]EMG: electromyogram.
[c]EDA: electrodermal activity.
[d]RR: respiration rate.
[e]BL: baseline.
[f]PL: pain level.
[g]LSTM: long short-term memory.
[h]AE: autoencoder.
[i]HC: handcrafted.
[j]SVM: support vector machine.
[k]CNN: convoluted neural network.
[l]EF: early fusion.

Table 2. Multiple modalities: comparison with other methods.

Study	Value, mean (SD)	Modalities
Werner et al [17]	65.02 (8.72)	Video, ECG[a], EMG[b], and EDA[c]
Lopez-Martinez and Picard [36]	66.68 (10.87)	ECG and EDA
Wang et al [37]	70.4 (9.76)	ECG, EMG, and EDA
Subramaniam and Dass [38]	92.604 (3.49)	ECG and EDA
Our method	81.08 (4.59)	EDA, EMG, EDA, and RR[d]

[a]ECG: electrocardiogram.
[b]EMG: electromyogram.
[c]EDA: electrodermal activity.
[d]RR: respiration rate.

Discussion

Principal Findings

This study demonstrated that RR emerged as the strongest single-modality predictor of pain intensity, particularly for distinguishing between baseline and lower pain levels. EMG performed best for higher pain intensities, while EDA and ECG showed comparatively lower effectiveness as stand-alone modalities. Multimodal models, though offering potential advantages in robustness and complementary information, generally underperformed compared with the RR single-modality models, likely due to challenges related to noise and data alignment. The study highlights the importance of modality selection and data fusion strategies for pain recognition in postoperative settings.

Performance by Modality

Pain Recognition Using RR Alone

From the single-modality results (Table 1), it is evident that RR models outperform all other modalities, especially for the BL versus PL1 and BL versus PL2 models. This highlights the strong predictive power of RR in distinguishing between baseline and lower pain intensities. The best-performing model used RR alone. One justification for these results could be the dynamic nature of RR signals in response to pain stimuli. Since we effectively isolated and captured periods of higher pain intensity with smaller window sizes, this could have helped the models better distinguish between baseline and other pain levels.

Pain Recognition Using EMG Alone

For the highest pain category (BL vs PL3), the EMG model outperformed other single-modality models. This suggests that facial muscle activation captured by EMG signals is particularly effective for distinguishing higher pain intensities. The

XSL•FO
RenderX

comparatively lower performance of other modalities, such as EDA, could be attributed to the subtle variations in their responses to different pain levels.

Pain Recognition Using EDA Alone

EDA models exhibited comparatively lower performance across all pain categories. This may be due to the difficulty in capturing clear variations in EDA signal responses to different pain levels, as observed in our experiments.

Pain Recognition Using ECG Alone

While ECG features contributed strongly to the performance of multimodal models, their single-modality results were not as robust as those of RR or EMG. However, the best-performing multimodal models shared identical configurations with the best ECG models, suggesting that ECG features had a significant influence on the multimodal results.

Challenges With Extremes in Pain Levels

The BL versus PL1 and BL versus PL3 models had relatively poor performances across both single and multimodal approaches. BL versus PL1 struggled to distinguish the baseline from the lowest pain intensity due to the subtlety of the physiological responses collected while experiencing this pain level. The BL versus PL3 model, however, found it challenging to distinguish pain levels due to the scarcity of labels for the highest pain intensity. Although data augmentation can help mitigate these challenges, there is no substitute for real data. On the contrary, the BL versus PL2 models performed better due to the relative abundance of such labels reported during trials.

Multimodal Performance

The best-performing multimodal model was trained on automatic features outputted from an LSTM network with a 10-second window size. This model, which made use of strong labels without any data augmentation techniques, achieved comparable results to the best-performing ECG single-modality model. Early fusion outperformed late fusion, likely due to its ability to detect correlations across modalities during feature selection [39]. By treating each modality as independent, late fusion might lose correlations in the combined feature space.

However, single-modality models, particularly RR, generally outperformed multimodal models. This contrasts with previous studies on healthy participants, where multimodal approaches typically excelled. Our findings suggest that the unique challenges of real-world postoperative data, including noise and missing signals, may complicate the integration of multiple modalities.

Advantages and Trade-Offs

While multimodal models have the potential to add complementary information and robustness, they also introduce challenges related to data alignment and noise management. Single-modality models, by contrast, are simpler, easier to interpret, and computationally less expensive. These advantages make single modalities, such as RR and EMG, attractive for certain applications despite the overall potential of multimodal approaches. Multiple modalities certainly have the potential to add more useful information over a single modality and can be used to introduce complementary information and resiliency when any one modality fails or is too noisy [40].

While comparing our results to previous studies [17,36-38] in Table 2, it can be observed that our models outperform most of their models in mean pain assessment scores except Subramaniam and Dass [38]. However, this is not entirely a fair comparison because we use 3 pain levels instead of 4, and our patients are not healthy.

An additional consideration is the comfort and compliance of patients wearing multiple biosensors, especially in postoperative settings. While multimodal models rely on multiple sources of data, this could pose a burden to patients who may already be experiencing discomfort. Future iterations of the framework could focus on optimizing the number of biosensors by identifying the most informative modalities. This optimization could improve patient compliance while maintaining the accuracy and robustness of the system.

Limitations

The main limitation of our algorithm is the presence of noise in the form of motion artifacts produced while collecting physiological signals. Since we obtained data from real postoperative patients in a clinical setting, they were allowed to move more freely compared to experiments performed in controlled laboratory settings. The presence of these motion artifacts diminished the quality of our data, thus negatively impacting our machine learning algorithms.

In addition, our study was conducted in a setting with a limited and relatively homogeneous patient population. While this setting allowed us to focus on developing and testing the algorithm, it restricts the generalizability of our findings to broader and more diverse clinical environments. Testing the model in varied clinical settings and across a larger, more diverse patient population is essential for evaluating its scalability and effectiveness in real-world scenarios. This remains an important future research direction.

Furthermore, we must acknowledge the more complicated facets of pain that are not fully captured by our algorithm, such as the number of days after surgery, the amount of pain medication administered, and the location and type of pain experienced. Incorporating these factors in future studies could improve the accuracy and robustness of pain assessment systems.

Future Directions

One of the main research directions we would like to explore is the development of real-time multimodal pain assessment systems using deep learning architectures. In such scenarios, missing or incomplete data from one or more modalities are likely to be encountered. Real-time systems also face limitations related to computational complexity and power constraints. Building on the experiments conducted in this study, we aim to create models capable of dynamically determining which modalities to use in an energy-efficient manner without compromising performance given the clinical context.

In addition, a promising avenue for future work is to build personalized machine learning models. These models could

XSL•FO
RenderX

leverage data from groups of similar patients while being fine-tuned to make predictions for individual patients. This personalized approach accounts for the large interindividual variability in pain perception, which makes a monolithic model unsuitable. Previous research has demonstrated the feasibility of using multitask machine learning to address variability in mood prediction tasks [41]. This strategy could be extended to the domain of pain assessment, not only for acute postoperative pain but also for chronic pain scenarios. Personalized modeling will be a vital step toward creating clinically viable and effective pain assessment algorithms.

Conclusions

In this paper, we presented a multimodal machine learning framework for classifying pain in real postoperative patients using the iHurt Pain Database. Both traditional handcrafted features and deep learning–generated automatic features were extracted from physiological signals (ECG, EDA, EMG, and PPG). Several experiments were conducted to perform binary classification among 3 different pain intensities versus baseline levels of pain. Models were varied based on the modalities used, the data augmentation techniques applied (SMOTE, Snorkel, or both), the machine learning algorithms used, and the modality fusion methods implemented.

Our results showed that binary pain classification significantly benefits from the application of data augmentation techniques in conjunction with automatic features. The single-modality models based on RR and EMG outperformed the multimodal models. The BL versus PL3 model with the best results was trained on EMG data alone, highlighting the importance of facial muscle activation in distinguishing higher pain intensities from baseline levels. This finding is consistent from a clinical perspective, as higher pain intensities are commonly associated with acute pain.

Overall, this study highlights a novel approach to addressing the challenges of building a pain recognition system for real postoperative patients, particularly constraints such as label imbalances and missing data. By employing robust data preprocessing techniques, data augmentation strategies, and multimodal fusion approaches, our framework demonstrates the potential for accurate and objective pain classification in clinical settings. These findings lay the groundwork for advancing multimodal pain assessment methods tailored to real-world clinical scenarios.

Conflicts of Interest

None declared.

References

1. Merskey HA. Pain terms: a list with definitions and notes on usage. Recommended by the IASP subcommittee on taxonomy. Pain 1979;6(3):249. [Medline: 460932]
2. Younger J, McCue R, Mackey S. Pain outcomes: a brief review of instruments and techniques. Curr Pain Headache Rep 2009;13(1):39-43 [FREE Full text] [doi: 10.1007/s11916-009-0009-x] [Medline: 19126370]
3. Kent ML, Tighe PJ, Belfer I, Brennan TJ, Bruehl S, Brummett CM, et al. The ACTTION-APS-AAPM pain taxonomy (AAAPT) multidimensional approach to classifying acute pain conditions. Pain Med 2017;18(5):947-958 [FREE Full text] [doi: 10.1093/pm/pnx019] [Medline: 28482098]
4. Bonica JJ. Management of cancer pain. In: Pain in the Cancer Patient: Pathogenesis, Diagnosis and Therapy. Berlin, Germany: Springer; 1984.
5. Treede RD, Rief W, Barke A, Aziz Q, Bennett MI, Benoliel R, et al. A classification of chronic pain for ICD-11. Pain 2015;156(6):1003-1007 [FREE Full text] [doi: 10.1097/j.pain.0000000000000160] [Medline: 25844555]
6. Chou R, Gordon DB, de Leon-Casasola OA, Rosenberg JM, Bickler S, Brennan T, et al. Management of postoperative pain: a clinical practice guideline from the American pain society, the American society of regional anesthesia and pain medicine, and the American society of anesthesiologists' committee on regional anesthesia, executive committee, and administrative council. J Pain 2016;17(2):131-157 [FREE Full text] [doi: 10.1016/j.jpain.2015.12.008] [Medline: 26827847]
7. Small C, Laycock H. Acute postoperative pain management. Br J Surg 2020;107(2):e70-e80. [doi: 10.1002/bjs.11477] [Medline: 31903595]
8. Herr K, Coyne PJ, Ely E, Gélinas C, Manworren RCB. Pain assessment in the patient unable to self-report: clinical practice recommendations in support of the ASPMN 2019 position statement. Pain Manag Nurs 2019;20(5):404-417. [doi: 10.1016/j.pmn.2019.07.005] [Medline: 31610992]
9. Craig K. The facial expression of pain Better than a thousand words? APS Journal 1992 Sep;1(3):153-162. [doi: 10.1016/1058-9139(92)90001-S]
10. Hadjistavropoulos T, Craig KD. A theoretical framework for understanding self-report and observational measures of pain: a communications model. Behav Res Ther 2002;40(5):551-570. [doi: 10.1016/s0005-7967(01)00072-9] [Medline: 12038648]
11. Staton LJ, Panda M, Chen I, Genao I, Kurz J, Pasanen M, et al. When race matters: disagreement in pain perception between patients and their physicians in primary care. J Natl Med Assoc 2007;99(5):532-538. [Medline: 17534011]
12. Gélinas C, Boitor M, Puntillo KA, Arbour C, Topolovec-Vranic J, Cusimano MD, et al. Behaviors indicative of pain in brain-injured adult patients with different levels of consciousness in the intensive care unit. J Pain Symptom Manage 2019;57(4):761-773 [FREE Full text] [doi: 10.1016/j.jpainsymman.2018.12.333] [Medline: 30593909]

XSL•FO
RenderX

13. Achterberg W, Lautenbacher S, Husebo B, Erdal A, Herr K. Pain in dementia. Schmerz 2021;35(2):130-138. [doi: 10.1007/s00482-020-00501-w] [Medline: 32926239]
14. Walter S, Gruss S, Ehleiter H, Tan J, Traue HC, Werner P, et al. The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system. 2013 Presented at: Proceedings of the IEEE International Conference on Cybernetics (CYBCO); Jun 13-15, 2013; Lausanne, Switzerland p. 128-131. [doi: 10.1109/CYBConf.2013.6617456]
15. Werner P, Al-Hamadi A, Niese R, Walter S, Gruss S, Traue HC. Towards pain monitoring: Facial expression, head pose, a new database, an automatic system and remaining challenges. 2013 Presented at: Proceedings of the British Machine Vision Conference; September 01, 2013; London, UK p. 1-13 URL: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=03f075e95638bc66e687badd97a58c5de67e58e6 [doi: 10.5244/C.27.119]
16. Hammal Z, Cohn JF. Automatic detection of pain intensity. Proc ACM Int Conf Multimodal Interact 2012;2012:47-52 [FREE Full text] [doi: 10.1145/2388676.2388688] [Medline: 32724903]
17. Werner P, Al-Hamadi A, Niese R, Walter S, Gruss S, Traue H. Automatic pain recognition from video and biomedical signals. 2014 Presented at: Proceedings of the 22nd International Conference on Pattern Recognition; August 24-28, 2014; Stockholm, Sweden p. 4582-4587. [doi: 10.1109/ICPR.2014.784]
18. Sarker VK, Jiang M, Gia TN, Anzanpour A, Rahmani A, Liljeberg P. Portable multipurpose bio-signal acquisition and wireless streaming device for wearables. 2017 Presented at: Proceedings of the IEEE Sensors Applications Symposium (SAS); March 13, 2017; New York City, NY p. 1-6. [doi: 10.1109/SAS.2017.7894053]
19. Empatica E4 User Manual. 2015. URL: https://www.manualslib.com/manual/1318955/Empatica-E4.html [accessed 2024-12-13]
20. Kasaeyan Naeini E, Jiang M, Syrjälä E, Calderon MD, Mieronkoski R, Zheng K, et al. Prospective study evaluating a pain assessment tool in a postoperative environment: protocol for algorithm testing and enhancement. JMIR Res Protoc 2020;9(7):e17783 [FREE Full text] [doi: 10.2196/17783] [Medline: 32609091]
21. Gomes PM, Margaritoff P, Silva H. pyHRV: development and evaluation of an open-source Python toolbox for heart rate variability (HRV). 2019 Presented at: Proceedings of the International Conference on Electrical, Electronic and Computing Engineering (IcETRAN); June 01, 2019; Serbia URL: https://www.researchgate.net/publication/333611305_pyHRV_Development_and_Evaluation_of_an_Open-Source_Python_Toolbox_for_Heart_Rate_Variability_HRV
22. Laitala J, Jiang M, Syrjälä E, Naeini E, Airola A, Rahmani A, et al. Robust ECG R-peak detection using LSTM. 2020 Presented at: Proceedings of the 35th Annual ACM Symposium on Applied Computing; March 30, 2020; Brno, Czech Republic p. 1104-1111. [doi: 10.1145/3341105.3373945]
23. Kasaeyan Naeini E, Subramanian A, Calderon MD, Zheng K, Dutt N, Liljeberg P, et al. Pain recognition with electrocardiographic features in postoperative patients: method validation study. J Med Internet Res 2021;23(5):e25079 [FREE Full text] [doi: 10.2196/25079] [Medline: 34047710]
24. Aqajari SAH, Naeini EK, Mehrabadi MA, Labbaf S, Dutt N, Rahmani AM. pyEDA: an open-source Python toolkit for pre-processing and feature extraction of electrodermal activity. Procedia Comput. Sci 2021;184:99-106. [doi: 10.1016/j.procs.2021.03.021]
25. Aqajari SAH, Cao R, Kasaeyan Naeini E, Calderon M, Zheng K, Dutt N, et al. Pain assessment tool with electrodermal activity for postoperative patients: method validation study. JMIR Mhealth Uhealth 2021;9(5):e25258 [FREE Full text] [doi: 10.2196/25258] [Medline: 33949957]
26. Greco A, Valenza G, Lanata A, Scilingo EP, Citi L. cvxEDA: a convex optimization approach to electrodermal activity processing. IEEE Trans. Biomed. Eng 2016;63(4):797-804. [doi: 10.1109/tbme.2015.2474131]
27. Cao R, Aqajari SAH, Kasaeyan Naeini E, Rahmani AM. Objective pain assessment using wrist-based PPG signals: a respiratory rate based method. Annu Int Conf IEEE Eng Med Biol Soc 2021;2021:1164-1167. [doi: 10.1109/EMBC46164.2021.9630002] [Medline: 34891494]
28. Madhav K, Ram M, Krishna E, Komalla N, Reddy K. Estimation of respiration rate from ECG, BPPPG signals using empirical mode decomposition. 2011 Presented at: Proceedings of the IEEE International Instrumentation and Measurement Technology Conference; May 10-12, 2011; Hangzhou, China. [doi: 10.1109/imtc.2011.5944249]
29. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw 2015;61:85-117. [doi: 10.1016/j.neunet.2014.09.003] [Medline: 25462637]
30. Le QV. A tutorial on deep learning part 2: Autoencoders, convolutional neural networks and recurrent neural networks. Google Brain 2015:1-20 [FREE Full text]
31. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. jair 2002 Jun 01;16(1):321-357. [doi: 10.1613/jair.953]
32. Lemaître G, Nogueira F, Aridas C. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. J Mach Learn Res 2017;18(1):559-563 [FREE Full text]
33. Ratner A, Bach SH, Ehrenberg H, Fries J, Wu S, Ré C. Snorkel: rapid training data creation with weak supervision. Proceedings VLDB Endowment 2017;11(3):269-282 [FREE Full text] [doi: 10.14778/3157794.3157797] [Medline: 29770249]

XSL•FO
RenderX

34. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. Proc. 14th IJCAI 1995:1137-1143 [FREE Full text]
35. Gunes H, Piccardi M. Affect recognition from face and body: early fusion vs. late fusion. 2005 Presented at: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics; October 12, 2005; Waikoloa, HI, USA p. 3437-3443. [doi: 10.1109/ICSMC.2005.1571679]
36. Lopez-Martinez D, Picard R. Multi-task neural networks for personalized pain recognition from physiological signals. 2017 Presented at: Proceedings of the Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW); October 23-26, 2017; San Antonio, TX p. 181-184. [doi: 10.48550/arXiv.1708.08755]
37. Wang R, Xu K, Feng H, Chen W. Hybrid RNN-ANN based deep physiological network for pain recognition. 2020 Presented at: Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); July 20-24, 2020; Montreal, QC p. 5584-5587. [doi: 10.1109/EMBC44109.2020.9175247]
38. Subramaniam SD, Dass B. Automated nociceptive pain assessment using physiological signals and a hybrid deep learning network. IEEE Sensors J 2021;21(3):3335-3343. [doi: 10.1109/jsen.2020.3023656]
39. Ross A. In: Li SZ, editor. Fusion, Feature-Level. Boston, MA: Springer; 2009.
40. Naeini EK, Shahhosseini S, Kanduri A, Liljeberg P, Rahmani A, Dutt N. AMSER: adaptive multimodal sensing for energy efficient and resilient eHealth systems. 2022 Presented at: Proceedings of the Design, Automation & Test in Europe Conference & Exhibition (DATE); March 14-23, 2022; Antwerp, Belgium p. 1455-1460. [doi: 10.48550/arXiv.2112.08176]
41. Taylor S, Jaques N, Nosakhare E, Sano A, Picard R. Personalized multitask learning for predicting tomorrow's mood, stress, and health. IEEE Trans Affect Comput 2020;11(2):200-213 [FREE Full text] [doi: 10.1109/TAFFC.2017.2784832] [Medline: 32489521]

Abbreviations

AdaBoost: adaptive boosting
AE: autoencoder
CNN: convolutional neural network
CPOT: Critical Care Pain Observation Tool
ECG: electrocardiogram
EDA: electrodermal activity
EMG: electromyogram
KNN: k-nearest neighbors
LSTM: long short-term memory
NRS: Numerical Rating Scale
PPG: photoplethysmography
RF: random forest
RR: respiration rate
SMOTE: synthetic minority oversampling technique
SVM: support vector machine

Edited by A Mavragani; submitted 24.10.24; peer-reviewed by G Chennekkattu Markose, X Gao; comments to author 25.11.24; revised version received 28.11.24; accepted 29.11.24; published 27.01.25.

Please cite as:
Subramanian A, Cao R, Naeini EK, Aqajari SAH, Hughes TD, Calderon MD, Zheng K, Dutt N, Liljeberg P, Salanterä S, Nelson AM, Rahmani AM
Multimodal Pain Recognition in Postoperative Patients: Machine Learning Approach
JMIR Form Res 2025;9:e67969
URL: https://formative.jmir.org/2025/1/e67969
doi:10.2196/67969
PMID:

©Ajan Subramanian, Rui Cao, Emad Kasaeyan Naeini, Seyed Amir Hossein Aqajari, Thomas D Hughes, Michael-David Calderon, Kai Zheng, Nikil Dutt, Pasi Liljeberg, Sanna Salanterä, Ariana M Nelson, Amir M Rahmani. Originally published in JMIR Formative Research (https://formative.jmir.org), 27.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The

XSL•FO
RenderX

complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluating Reaction Videos of Young People Watching Edutainment Media (MTV Shuga): Qualitative Observational Study

Venetia Baker[1], MSc; Sarah Mulwa[1], MSc; David Khanyile[2], BA; Georgia Arnold[3], BA; Simon Cousens[1], PhD; Cherie Cawood[2], PGDip; Isolde Birdthistle[1], PhD

[1]Faculty of Epidemiology and Population Health, The London School of Hygiene and Tropical Medicine, London, United Kingdom

[2]Epicentre Health Research, Durban, South Africa

[3]GA-Agency, London, United Kingdom

Corresponding Author:
Venetia Baker, MSc
Faculty of Epidemiology and Population Health
The London School of Hygiene and Tropical Medicine
Keppel St
London, WC1E 7HT
United Kingdom
Phone: 44 020 7636 8636
Email: venetia.baker1@lshtm.ac.uk

Abstract

Background: Mass media campaigns, particularly edutainment, are critical in disseminating sexual health information to young people. However, there is limited understanding of the authentic viewing experience or how viewing contexts influence engagement with media campaigns. Reaction videos, a popular format in web-based culture in which users film themselves reacting to television shows, can be adapted as a research method for immediate and unfiltered insights into young people's engagement with edutainment media.

Objective: We explored how physical and social context influences young people's engagement with *MTV Shuga*, a dramatic television series based on sexual health and relationships among individuals aged 15 to 25 years. We trialed reaction videos as a novel research method to investigate how young people in South Africa experience the show, including sexual health themes and messages, in their viewing environments.

Methods: In Eastern Cape, in 2020, purposively selected participants aged 18 to 24 years of an evaluation study were invited to take part in further research to video record themselves watching *MTV Shuga* episodes with their *COVID-19 social bubble*. To guide the analysis of the visual and audio data, we created a framework to examine the physical setting, group composition, social dynamics, coinciding activities, and viewers' spoken and unspoken reactions to the show. We identified patterns within and across groups to generate themes about the nature and role of viewing contexts. We also reflected on the utility of the method and analytical framework.

Results: In total, 8 participants recorded themselves watching *MTV Shuga* episodes in family or friendship groups. Viewings occurred around a laptop in the home (living room or bedroom) and outside (garden or vehicle). In same-age groups, viewers appeared relaxed, engaging with the content through discussion, comments, empathy, and laughter. Intergenerational groups experienced discomfort, with older relatives' presence causing embarrassment and younger siblings' distractions interrupting the engagement. Scenes featuring physical intimacy prompted some viewers to hide their eyes or leave the room. While some would prefer watching *MTV Shuga* alone to avoid the self-consciousness experienced in group settings, others valued the social experience and the lively discussions it spurred. This illustrates varied preferences for consuming edutainment and the factors influencing these preferences.

Conclusions: The use of reaction videos for research captured real-time verbal and nonverbal reactions, physical environments, and social dynamics that other methods cannot easily measure. They revealed how group composition, dynamics, settings, and storylines can maximize engagement with *MTV Shuga* to enhance HIV prevention education. The presence of parents and the camera may alter young people's behavior, limiting the authenticity of their viewing experience. Still, reaction videos offer a unique opportunity to understand audience engagement with media interventions and promote participatory digital research with young people.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e55275) doi:10.2196/55275

KEYWORDS

mass media; edutainment; adolescents; sexual health; HIV prevention; participatory research

Introduction

Background

Edutainment has become a crucial tool for disseminating sexual health information, especially to young people [1-5]. Singhal and Rogers [6] define *edutainment* as the deliberate use of media to entertain and educate, aiming to enhance knowledge, foster positive attitudes, and change behavior. This approach has traditionally used various media formats, including radio, television, comic books, and music, to influence audience awareness, attitudes, and behaviors in their external environment to help create the necessary conditions for social change [6]. The Sabido methodology is an approach to edutainment using soap operas to promote positive social change by integrating educational messages into compelling narratives. It relies on social learning theory, which posits that people learn behaviors by observing others, especially when they see those behaviors rewarded [6]. The Sabido methodology uses this theory to create positive, negative, and transitional role models that viewers can relate to, aiming to shift behaviors by showing the consequences of these characters' actions [6]. Audience engagement is key as the method emotionally invests viewers in the characters' journeys, encouraging reflection and discussion. The discussions that edutainment generates are central to its ability to diffuse new ideas into the community, shifting not just individual behavior but also social norms [7].

Edutainment plays a crucial role in the field of communication for development across various development challenges, such as HIV and AIDS, sexual and reproductive health, domestic violence, and sustainable livelihoods [3,4,8-10]. In South Africa, which has the largest population of people living with HIV (UNAIDS [11]), both governmental and nongovernmental organizations have made significant investments in communication campaigns aimed at correcting misinformation and reducing HIV-related stigma [1,12]. Edutainment was key in correcting misinformation and reducing stigma during South Africa's response to the HIV or AIDS epidemic [13-16].

Edutainment became a key component in raising awareness about HIV and AIDS in South Africa, particularly through the popular drama *Soul City* [10]. Created by the Soul City Institute, *Soul City* is a multilevel and culture-sensitive communication strategy centered on the television drama that sought change at the individual, community, and broader society levels to inspire social action and sociopolitical mobilization by incorporating community mobilization and action into the narrative [7,17]. It integrated HIV storylines and reached millions of viewers, showing positive effects on knowledge, attitudes, and behaviors related to HIV prevention [10].

Since *Soul City*'s success, HIV and sexual health edutainment have become a popular approach for reaching South African audiences, especially younger age groups. *MTV Shuga*, *Down South* (2017-present), which is the focus of this study, is a current edutainment series that aims to create storylines centered on HIV prevention, including pre-exposure prophylaxis (PrEP) and HIV self-testing, for young people, a critical population for addressing the high incidence and prevalence of HIV in South Africa [18]. *Down South* is part of the *MTV Shuga* mass media campaign created by the MTV Staying Alive Foundation, which aims to engage young people with positive sexual health messages using compelling and relatable storylines. *MTV Shuga* is part of a large media campaign and has reached millions of viewers worldwide through television and web streaming since 2009.

When evaluating health edutainment interventions, research has often relied heavily on surveys and structured interviews to determine effectiveness [4,5,10,19,20]. These types of studies have shown the mixed effects of edutainment on sexual health outcomes. While these studies can track measurable changes in attitudes, behaviors, and tangible outcomes—such as increased use of services—they do not explain why certain elements of edutainment work whereas others do not. To fully understand the impact of edutainment, other methods must be used to explore the nuanced ways in which audiences engage with the content in their everyday lives [8].

Many studies that use qualitative methods to evaluate edutainment dramas with HIV or sexual health storylines rely on self-reported, retrospective data obtained through qualitative interviews and focus groups [3,21-25]. These methods have proven valuable in providing deep insights into participants' personal experiences, perceptions, and emotional responses to edutainment content. Such approaches are essential for understanding the mechanisms through which edutainment influences behavior, attitudes, and beliefs as they allow participants to articulate their experiences in their own words.

Moreover, the retrospective nature of the data collected in these studies can capture longer effects of edutainment [3,4,26,27]. This includes how the content continues to resonate with participants over time, influencing their decisions and actions. Focus groups facilitate discussions that elicit diverse perspectives and highlight shared experiences among participants, contributing to a more comprehensive understanding of edutainment's effects [4,21,23,24,27-29]. However, several limitations are inherent in these qualitative methods. Social desirability bias, recall issues, and the power dynamics between participants and researchers are notable concerns that may compromise the reliability of the data. Furthermore, there is a tendency to sample participants who are already engaged with the content, potentially skewing the findings toward more favorable opinions [4,21,22,26].

While most studies have relied on retrospective researcher-led methods, alternative approaches such as participatory methods and observational evaluations have also been explored. Some studies have used participatory and observational methods such as storytelling and community screenings to evaluate audience

XSL•FO
RenderX

engagement [26,28,29]. Observations have provided valuable insights into physical settings and collective experiences similar to those seen in theaters or live performances [30-32]. However, these insights may not fully reflect the diverse influences and mediating factors present in more intimate home environments, where most edutainment is consumed.

MTV Shuga has been evaluated using both quantitative and qualitative retrospective methods. The results of our own research using a web-based survey in the Eastern Cape, South Africa, have shown a positive association, consistent with a causal impact of *MTV Shuga "Down South"* (based in South Africa) on young people's awareness and use of HIV prevention tools such as HIV self-screening and condoms and greater motivation among adolescents to use PrEP for HIV [4]. Evaluations of other *MTV Shuga* series produced in other contexts such as Nigeria and KwaZulu-Natal in South Africa also show positive effects of engaging with *MTV Shuga*, for example, on young people's adoption of chlamydia and HIV testing, condoms, and contraception, as well as significant attitude shifts and increased disclosure of sexual violence among youth [33-35]. Qualitative evaluations using in-depth interviews and focus group discussions of *MTV Shuga*, including our own [21], have found that *MTV Shuga*'s relatable, unjudgmental content helps young people think through scenarios that relate to their real lives. They also describe how the discussions and debates inspired by *MTV Shuga* created a safe place to discuss sexual health, reducing stigma and increasing support [21,27]. These methods contributed valuable knowledge to *MTV Shuga*'s effectiveness and the mechanisms in which *MTV Shuga* works. Participants self-reported how social context can modulate the viewing experience. However, we lacked data on how those interactions happen in real time and the way in which people consume *MTV Shuga* in their everyday lives.

Edutainment evaluations rarely use participatory approaches to capture viewers' natural, real-time interactions and reactions, including how mass media edutainment is experienced in its actual viewing environment. To gain a more comprehensive perspective, there is a need to capture the viewing experience and how the viewing context, including both social and physical elements, influences the effectiveness and impact of mass media edutainment. This holistic understanding can further enhance the effectiveness of sexual health messaging through edutainment.

Objectives

In this paper, we explore the role of reaction videos as a method of evaluating an edutainment campaign. Our objectives for using reaction videos as a research method were to (1) create a remote approach to observation without the presence of a researcher in which young people could record their experiences of watching the *MTV Shuga* show in viewing groups and (2) understand how physical context and social dynamics influence young people's experiences of watching *MTV Shuga* within their household or other social spaces, with family, relatives, and friends. We also sought to critique the successes and challenges of using self-filmed observation for research purposes to reflect on the utility of this technique more widely in qualitative investigations with young people.

We piloted and described the use of reaction videos self-filmed by viewers as part of a mixed methods evaluation of *MTV Shuga* in South Africa. Previously, young people's real-life engagement with the *MTV Shuga* series has not been documented at the time or place of private viewing, for example, at home with family and friends. We hoped that self-filmed reaction videos would complement the self-reported data we collected [4,21,36] by revealing the physical and social context in which young people view sexual health media content and how that experience may influence the show's effect.

Analyzing reaction videos can provide valuable insights into viewers' real-time emotional and physical reactions within their contextual environment. These videos, which record people's responses as they watch scenes from movies, television shows, music videos, or comedy sketches, have become a prominent feature in web-based and media culture. The rising popularity of reaction videos on YouTube, TikTok, and other social media platforms reflects a fascination with observing other people's natural, unscripted social interactions. As with other media formats, reaction videos have the potential to be both entertaining and educational. They offer a literal lens with which we can observe the influence of media in real-time and real-life situations [37]. While viewers are aware of the recording and potentially alter their behavior consciously or unconsciously, reaction videos offer an inside perspective that cannot be easily captured otherwise and which can complement other data sources for research and evaluation.

There is an emerging body of literature on the nature and influence of reaction videos, primarily focusing on reaction videos on YouTube. Analysis of YouTube reaction videos reveals that they vary from amateur—unscripted, spontaneous, and organic—to more professional, scripted, and edited formats. While primarily aimed at entertainment, these videos also create narratives that guide viewers in interpreting content through the creators' perspectives [38]. They help build communities and challenge dominant views, enabling viewers to share their own perspectives and influence how videos are consumed [38]. These videos foster a sense of community by uniting viewers worldwide to share and discuss in comment sections, with participatory elements such as content suggestions and creator interactions further strengthening these connections [39,40].

Adopting reaction videos as a research method presents valuable potential for data collection. Instead of analyzing existing reaction videos on the web, we asked participants to create their own reaction videos, which would not be posted on the web for entertainment but used exclusively for research purposes. This approach aimed to minimize the influence of performative, exaggerated reactions often seen in web-based content, where viewers may enhance their responses to build a fan base or generate income. We hoped that these reactions would be similar to what participants might naturally experience at home. This method can provide insights from individuals who are not web-based personalities who may be posting content for attention or commercial purposes, which can skew data toward extremes due to algorithmic pressures. By adopting this format, researchers can collect real-time responses and insights, offering a rich qualitative data source reflecting participants' immediate reactions and emotions. This approach deepens data collection

XSL•FO
RenderX

and provides a novel way to engage participants in the research process. The use of reaction videos as a research method can draw from forms of traditional data collection, including photovoice and observations. The participatory and visual principles of photovoice apply here as they empower participants to document and share their reactions and perspectives, fostering a more engaged and authentic form of data collection [41,42]. Similarly, the principles of observation can be used by capturing real-time interactions and spontaneous responses, providing rich, contextual insights. Video recordings can allow researchers to capture the participants' physical environment, social dynamics, and other contextual factors without physically being there [43,44]. Reaction videos introduce an additional participatory dimension to video observations as the participants are responsible for self-recording, allowing them to decide what to capture.

Methods

Recruitment

We recruited participants from a web-based survey that was live on the web in South Africa from September 2020 to December 2020. The survey was part of an evaluation to understand the impact of the mass media campaign *MTV Shuga Down South season 2 (DS2)*, and detailed methods and results to date have been published previously [4,21,36]. In short, 3431 young people aged 15 to 24 years completed the web-based survey. Participants were recruited through targeted advertisements to individuals aged 15 to 24 years in Mthatha through social media platforms, primarily Facebook. The questionnaire was also promoted in collaboration with local schools, universities, and clinics through their Facebook pages and WhatsApp groups.

This study was conducted in Mthatha, Eastern Cape, which was selected due to the recent distribution of HIV self-testing kits and the availability of oral PrEP in the area, ensuring that any demand generated by the *MTV Shuga DS2* campaign could be met. In addition, the offline components of the *MTV Shuga DS2* campaign were active in this region, providing a local context for examining the campaign's impact. Mthatha was also selected because of its relatively lower levels of HIV prevention research and testing compared to other provinces, making it a priority area for intervention. This decision was further supported by the South African Department of Health, which identified Mthatha as a strategic site for increasing HIV prevention efforts among youth.

Of those who reported ever having watched *MTV ShugaDS2* (N=238), 28.6% (n=68) within the study area of Mthatha in Eastern Cape opted in to further research to be conducted remotely due to COVID-19 restrictions at the time. Participants were first invited to individual interviews or focus groups, the findings of which are reported elsewhere [4,21], and participants of those activities were subsequently invited to create video recordings of themselves watching *MTV ShugaDS2* episodes. Initially, the eligibility criterion was that viewing groups should include a parent and their adolescent children (aged 15-19 years) to understand parent-child viewing dynamics. However, when it proved difficult to recruit adolescents willing to watch with their parents, eligibility was expanded to all 68 participants from Mthatha (aged 15-24 years) who opted in to further research to watch *MTV ShugaDS2* with any family members or housemates of their choice (within their COVID-19 "social bubble"). Of those, participants were called at random until 8 consented to and participated in the reaction videos.

The primary participant for each viewing group, recruited through focus groups and interviews, must have watched at least one episode of *MTV Shuga DS2*, but the other participants in the group did not require any previous exposure. We did not impose an age limit for this activity as the main participant was drawn from the evaluation study with individuals aged 15 to 24 years, within the target audience for *MTV Shuga*. As all participants responsible for organizing their viewing group had some familiarity with *MTV Shuga*, they were responsible for determining their viewing group and assessing the appropriateness of the situation.

A total of 8 groups was considered a suitable number to trial the method and analytical framework with a range of viewing groups. The decision to use a small sample size was justified by the study's aim to gain in-depth insights into family dynamics and viewing behaviors. We predicted that we would reach saturation with 8 groups ensured. By focusing our resources on a small number of groups watching multiple episodes of *MTV Shuga*, we prioritized depth over breadth, allowing for meaningful interactions and a thorough understanding of each family's viewing experience. Practical and budget constraints also influenced the sample size, ensuring a feasible research process.

Ethical Considerations

This study was approved by the Biomedical Research Ethics Committee at the University of KwaZulu-Natal (reference BREC/00000477/2019), the London School of Hygiene and Tropical Medicine (reference 17996), and the World Health Organization (reference ERC.0003283). Ethics approvals covered all phases of the research, including participant recruitment and data collection and analysis. Participants, who were recruited through the web-based survey, provided web-based consent before joining the study. For participants aged <18 years, consent was obtained from parents or guardians, and this was done via email.

Participants who opted into the individual interviews or focus groups and were later invited to the reaction videos were provided with detailed information sheets and had the opportunity to discuss the study with data collectors over the phone. During these discussions, participants could ask questions to ensure that they fully understood the nature of their participation. Parents of participants aged <18 years were required to give consent via a form sent over email.

This study took several measures to ensure privacy and confidentiality. All participant data were anonymized. The reaction videos themselves were not disseminated outside of the research team. To maintain secrecy while illustrating moments from these videos, the research team commissioned a young South African artist, Lunga Jonas, to create graphic art representations of the footage. These artworks altered

XSL•FO
RenderX

participants' facial features to prevent identification while retaining the social and emotional context of the group viewings.

In recognition of the time that the participants contributed to the study, they were compensated with 100 rand (US $5.54) for their participation. This compensation was provided to ensure that participants were fairly acknowledged for their involvement without coercion or undue influence.

This study followed the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist to ensure the transparency and rigor of the research process, including participant selection and data collection and analysis procedures.

Data Collection

Researchers conducted informed consent procedures via a WhatsApp phone call with the entire viewing group, specifically to confirm that they were aware of the reaction videos and their purpose and consented to being filmed. The survey participants signed a consent form via email on behalf of their viewing group. For participants aged <18 years, the written permission of parents and guardians was also confirmed via email. We asked participants to use a smartphone to video record their group watching *MTV Shuga DS2* episodes 5 and 6 on YouTube, both 22 minutes long. It was not required for participants to have seen episodes 1 to 4 before participating in the reaction videos. These episodes include storylines about PrEP, HIV self-screening, and gender-based violence. They also feature 2 scenes with physical intimacy that include kissing, removing clothes, and opening condoms. All participants were able to choose where, when, and with whom they would watch the episodes. The participants transferred their video recordings via WhatsApp to the research team, who deleted them from the research phone once downloaded to a password-protected private server.

In total, 4 researchers familiar with the study setting and fluent in the local language reviewed each reaction video and took notes, describing the setting, interactions, and significant verbal and nonverbal communication in ways agreed upon during researcher training to aid consistency. Nonverbal behavior was defined to encompass body movements and postures, along with temporal speech markers such as gaps and silences. Verbal behavior was characterized by spoken words, verbal expressions such as gasps and laughter, and the strength or emotive inflection in vocal expressions [45,46].

Following their review, researchers called the main participant of each viewing group via a WhatsApp "debrief" call to ask about the experience of watching *MTV Shuga DS2* with this group of people and whether they had any advice to improve the show. Participants' responses were summarized and included in the notes. Research notes and debrief notes were recorded in English and translated from Xhosa into English as needed.

Analysis

To analyze reaction videos, we created a framework to capture the social and physical context of people consuming edutainment in their everyday lives [8,19]. We designed a framework with the following components: *group composition*, *physical setting*, *coinciding activities*, *social dynamics*, and *reactions to the DS2 show and this research activity* (Textbox 1).

Textbox 1. Reaction video analysis framework.

Components

1. Group composition: the characteristics of those involved in the viewing, including their age, gender, relationships with each other, and other relevant information provided by the participants.
2. Physical setting: this describes the physical space in which the participants viewed the *MTV Shuga DS2* episodes. It encompasses factors such as the location, the time of day, surrounding physical objects, lighting, and overall environment, as well as the mode or device through which they viewed *MTV Shuga DS2*.
3. Coinciding activities: this constitutes any concurrent activities happening during the viewing that might either enhance or distract from engagement with the show.
4. Social dynamics: this constitutes the interactions between participants during the viewings. It includes spoken dialogue or unspoken gestures and responses between members of the viewing group.
5. Reactions to the *MTV Shuga DS2* show: this encompasses any verbal or nonverbal reaction expressed by participants in response to the *MTV Shuga DS2* episodes and scenes they are watching. This includes body language, facial and bodily expressions (such as laughing or gasping), and their overall attitudes while watching (eg, bored, relaxed, awkward, or happy) as perceived by the researchers and from the debrief call with the main participant.
6. Reactions to the research activity: this refers to participants' awareness of the camera, including speaking directly to the camera, their chosen positioning for the filming, and any statements they made in the debrief call about their experience of filming.

To apply our analytical framework to the coding process, we used the framework's subheadings to code each video recording, observation note, and audio transcript. Short summaries were written for each video, organized by the framework's subheadings. These summaries were then combined to identify codes within and across the videos. The lead analyst collaborated with the data collection researchers and a translator to discuss and interpret the findings, ensuring that they accurately reflected the participants' experiences. These codes were synthesized into themes guided by the framework supported by specific data excerpts such as quotations from transcripts or descriptions of moments captured in the videos.

The researchers involved in conducting the follow-up call (Zulu and Xhosa, male and female, and aged <30 years) and the lead analyst (White, British woman, and aged 32 years) reflected on how their knowledge, attitudes, assumptions, and positions might affect the interpretation of the data. On the basis of these

XSL•FO
RenderX

reflections, we determined that the initial analysis by the lead analyst should be verified with the other researchers and transcribers to ensure that the data were accurately summarized and interpreted.

Results

Group Composition

A total of 8 self-filmed reaction videos were received from 8 distinct viewing groups (Multimedia Appendix 1). Only 1 group (group 1) included a parent who filmed her 3 sons (aged 15-26 years) while watching the show together. Other groups included viewers ranging in age from 7 to 32 years. These groups, confined to "COVID-19 social bubbles," primarily consisted of siblings and cousins. All groups included male individuals: 38% (3/8) included male individuals only and 62% (5/8) were mixed-gender groups. There were no exclusively female groups; all female viewers were accompanied by older male cousins or younger brothers who were not within the target age range for *MTV Shuga* (15-25 years).

Physical Setting

As illustrated in Multimedia Appendix 1, the environments in which the groups convened to watch *MTV Shuga DS2* were diverse, with a blend of communal and secluded spaces. Use of laptops by all groups may have allowed for flexibility in where they chose to watch the show. Groups 5, 6, and 7 viewed the episodes in spacious living rooms with plush sofas and chairs, suggesting an evening setting or at least an ambiance with the curtains closed. Group 1, on the other hand, opted for a more structured viewing setup, congregating around a dining table during daylight hours. Groups 4 and 8 chose the privacy of bedrooms, where they huddled around laptops with siblings or cousins. These indoor gatherings were marked by an informal, relaxed ambience, with participants donning loungewear and pajamas and sporting bonnets, all contributing to the gatherings' casual and comfortable essence.

In total, 67% (2/3) of the all-male groups, groups 2 and 3, opted for outdoor settings for their daytime viewings. Group 2 created a makeshift viewing area by arranging chairs around a crate situated beside an exterior wall. The 4 brothers in group 3 sought an enclosed environment, choosing to watch the episodes from the confines of a vehicle. Although they mostly had the windows and doors shut, they occasionally rolled them down to talk to passersby who approached the vehicle to greet the boys inside.

Coinciding Activities

Communal activities often accompanied the viewing of the show, particularly in groups watching at home. For instance, group 6 and 7 participants combined viewing with eating a meal, whereas group 5 engaged in doing their hair. These activities illustrated the intimacy or familiarity of these groups; however, they also introduced potential distractions. Engaging in activities unrelated to the show led to conversations and actions that diverted attention away from the content being watched. As a result, participants might have missed crucial dialogue or scenes, highlighting a trade-off between enhancing the communal viewing experience and maintaining focus on the show's narrative and messages.

Other distractions included people who were not involved coming to chat with the participants, which only happened in group 3, who met in the vehicle.

Phone use emerged as a significant distraction during the viewings. Although recording the viewings required smartphones, other phones were frequently present and easily accessible to participants. While most participants focused on the show, smartphones proved to be potential distractions. For instance, in group 2, a participant began scrolling through his phone for several minutes, diverting his attention away from the show. Similarly, in group 5, a viewer aged 13 years was less engaged with the show than her sister, who was aged 24 years; more absorbed in her phone; and visibly distracted by an ongoing text conversation, as indicated by her active notifications. Both the older sister and younger brother nudged her and told her to concentrate on the episodes. In addition, smartphones were sometimes deliberately used to "escape" the viewing. In group 7, during a scene that made him noticeably uncomfortable, one man turned to his phone to avoid watching the laptop screen.

Despite most groups experiencing some distractions, no group paused or rewound to avoid missing the show.

Social Dynamics

Awkwardness

While most of the time viewers appeared relaxed around each other, intimate scenes showing couples kissing, opening condoms, or discussing PrEP created awkwardness across many groups. Viewers aged ≤15 years often responded to the intimate scenes by closing their eyes, covering their faces, or laughing nervously. In group 1, the mother often teased and joked with her sons about their visible discomfort during some scenes, playfully remarking that "it's not that painful" when her children appeared awkward during an intimate scene about PrEP that ended with characters kissing and taking off clothing. Although the atmosphere appeared lighthearted, the boys, smiling shyly, laughing with flushed faces, shaking their head, and covering their eyes, were embarrassed watching these scenes with their mother. In group 6, the teenage girl looked up and smiled awkwardly, showing discomfort as she pushed her face onto a pillow to hide from the intimate scene. Her uncle remarked the following in the debrief:

> *It was awkward a little bit because my niece and I are not used to each other that much, so I was uncomfortable with some scenes.*

In the debrief calls, groups 5, 7, and 1 agreed that watching with family was awkward, but overall, they said that they enjoyed the experience.

Dialogue

The reaction videos highlighted the social aspect of watching a show as a group, especially how it elicits reactions and facilitates discussion among viewers. In groups of young men, the show's content related to HIV prevention, often intimate in nature, generated reactions or dialogue; for example, one participant in group 3 encouraged the characters on-screen to engage in sexual activity using the phrase "Shaya shaya [Hit

XSL•FO
RenderX

Hit].” Participants explained how watching the show created an environment where they could discuss topics that came up on the show once the viewing was complete. Group 2 shared the following:

> *I love watching shows like these with my brothers because we discuss and agree and disagree on some issues afterwards, but it helps because we grow each other.*

For group 3, the discussions were also a major component of enjoyment of the show; they stated that they “loved every moment of it because we discussed it afterwards and shed light, so it was very informative.” However, it was not universal that groups enjoyed watching the show in groups—mainly due to discomfort around intimate scenes and relationship dynamics with family members, as referenced in the aforementioned theme. As the group 6 uncle stated, he found it uncomfortable watching with his niece, but “I wouldn’t have a problem watching alone.”

Age Dynamics

This study revealed varying dynamics in viewing experiences across different age groups. Groups in which viewers were of a similar age, such as group 7, found the viewing experience agreeable and comfortable, noting the following:

> *It was okay [watching with family]. We are all adults.*

Similarly, group 4 said the following:

> *We enjoyed it. It was fine because we are almost the same age group.*

The presence of younger siblings aged 7 to 8 years in groups 5, 6, and 8 introduced unique dynamics. These younger viewers were notably active, often moving in and out of the room, yet engaged with the show by asking questions about the storyline, such as “What’s happening?” (group 8) or “Who is Lemo?” (group 5). These inquiries sometimes elicited impatience from their older siblings. For instance, during a scene in which a character discovers that his girlfriend has tested positive for HIV, the viewer aged 8 years in group 5 began narrating the event, only to be cut off by an older sister with a dismissive, “You don’t need to explain.” The video recording later captures the end of a discussion happening off camera, then shows a visibly upset younger brother remarking, “This is such a joke, this is supposed to be a family thing” before exiting the room.

In the group 8 recording, the younger brother playfully interacted with the camera, making faces and sticking out his tongue, prompting a reprimand from his older sister filled with a mix of discipline and humor:

> *If you do that one more time, I am going to make you rewatch this [laughs and pauses], but I’m serious.*

These interactions show that, although younger viewers were included, their older siblings frequently guided or managed their behavior. It led to distractions for the older viewers and irritation of the younger ones.

Reactions to the Videos

Entertainment Value

Overall, the participants seemed to enjoy watching the show, finding it engaging and fun. Many participants stated in the debrief that they enjoyed the music, and this was evident during the recordings, with some singing along and dancing in moments (groups 6 and 8) and recognizing famous actors on the show. Others enjoyed the other creative components integrated into the show; for instance, the older sister in group 5 hummed and snapped her fingers during a scene with a poetry slam.

Reactions to Storylines

Q and Dineo (Transactional Sex)

The storyline between Q and Dineo, which portrays a budding romantic relationship, elicited varied reactions from participants, particularly regarding the theme of transactional relationships. Dineo’s involvement in a transactional sexual relationship, where her school fees are paid by an older man (a “Blesser”), sparked comments on the motivations behind such relationships. In group 4, a viewer empathized saying, “Things we do for money” when Dineo explained to Q why she was sleeping with an older man.

Q’s initial response to Dineo’s situation was seen as understanding and supportive. The group 1 mother remarked, “That’s one understanding guy” as Q came to terms with Dineo’s arrangement with her Blesser.

However, Q’s attempts to win Dineo’s affection by spending money on her generated different reactions. In group 2, a male viewer aged 19 years noted Q’s financial expenditure, remarking, “u fit kanje,” which translates to “so much money.” This observation was echoed by other groups, with expressions of surprise and dismay watching Q withdraw his savings to spend on Dineo.

In group 3, participants expected physical intimacy in Q and Dineo’s relationship as a return for Q’s financial spending, encapsulated in the following comment—“Akamunike futhi (She must give it to him now)”—showcasing the perceived sexual obligation created by the financial exchange.

Reggie and Odirile (Same-Sex Relationship)

The responses of boys and men to scenes showing same-sex intimacy were characterized by notable discomfort. This was manifested by in-jokes and physically shielding their eyes during the scenes in which 2 male characters showed intimacy while discussing the use of PrEP. In groups 2 and 7, male viewers’ reactions to these scenes were marked by avoidance; a male individual from each group walked away when the 2 male characters started kissing. This strong aversion contrasted with a female viewer in group 7 who laughed at the boys’ uneasy reactions. In group 4, during a scene in which male characters Reggie and Odirile kissed, the group expressed their unease (shaking their head, covering their eyes, and groaning), with one stating, “They should move on to the next scene,” suggesting a collective discomfort or disinterest in engaging with same-sex romantic content. Young men’s reactions

XSL•FO
RenderX

potentially led them to miss the scene and the educational content about PrEP and HIV prevention options.

Daniel and Ipeleng (HIV Self-Testing)

The HIV self-testing scenes in *MTV Shuga DS2* were designed to be educational and showed the characters Daniel and Ipeleng using a mouth swab for self-testing at home. This scene elicited a subdued reaction compared to scenes involving PrEP and intimacy. Only groups 7 and 5 visibly reacted to these moments. In group 5, someone off camera expressed disbelief saying, "i-swab? [a swab?]," showing that she was surprised that HIV could be screened for using an oral swab. Group 7 participants showed skepticism; one doubted the method's reliability, preferring a blood test and saying, "Don't trust this one; I want the blood one." Another participant misconstrued the testing method's implications, mistakenly suggesting, "If they swab, that means we can get HIV through kissing."

Reaction to the Research Method

The reaction videos were, on average, 29.38 (SD 9.3) minutes long. The duration of the 2 episodes together is 44 minutes. Some of the videos were sent in 1 long recording, whereas other groups sent multiple shorter videos.

The awareness or presence of the camera played different roles in each video. In group 1, the mother, holding her phone to film the observation, stood over her sons, using the camera to zoom in and capture their expressions. Although her voice was audible, she remained invisible on-screen. Her sons were very aware of the camera and often hid their faces, smiling shyly as she brought the camera close. In contrast, the young female participants who watched with their younger siblings (groups 5 and 8) positioned themselves as the central characters in the viewing, ensuring the camera captured their reactions. They appeared to be aware of the camera, sometimes glancing at it. For instance, the viewer aged 24 years in group 5 set up the camera and positioned herself within the frame. She took charge of the viewing, reprimanding her sister for checking her phone and not paying attention to the show. Her glances at the camera showed an awareness that the video had captured the distraction.

Discussion

Principal Findings

This study offers valuable insights into how physical and social contexts shape engagement with edutainment content such as *MTV Shuga*. Reaction videos captured nuanced dynamics within diverse viewing groups, including siblings, cousins, parents, and friends, revealing varying levels of comfort and discussion. Participants' responses to themes such as same-sex intimacy, HIV prevention, and transactional sex were influenced by familial and peer relationships, as well as the settings in which they viewed the show—ranging from communal living rooms to private bedrooms and outdoor spaces. Concurrent activities such as eating, styling hair, and smartphone use further modulated the viewing experience. Our findings demonstrate that reaction videos effectively captured spontaneous engagement, providing unique insights into nonverbal cues, group dynamics, and emotional reactions often missed by retrospective methods. This real-time documentation deepened our understanding of how physical and social environments impact viewer engagement. While self-filmed observations proved valuable in capturing immediate reactions and discussions, challenges such as varying levels of camera awareness and engagement with the content could affect the authenticity of the viewing experience.

This approach leverages smartphone cameras to allow participants to show their physical and social context while enabling remote observation by researchers. This eliminates the need for a researcher's physical presence, potentially reducing response bias as the presence of the in-person researcher will not influence social dynamics. While the absence of a researcher can diminish direct influence, the visible awareness of being filmed and the anticipation of sharing these recordings for research could shape participants' behaviors.

However, while participants were aware of the camera, most did not directly engage with it. Unlike the highly entertainment-focused content on social media, the reaction videos in this research captured responses in natural viewing environments that appeared candid and unobtrusive [38,40]. The reactions recorded in this study likely reflect genuine, unguarded responses, similar to those occurring when no "audience" is watching the reaction. This contrasts with the more performative reactions often seen in content aimed at entertainment, persuasion, or monetization [38,39].

Granting participants the autonomy to decide their positioning in front of the camera and select the moments they wish to record transforms the research process into a participatory one. The dynamic of power and control shifts as participants handle the camera by self-focusing or navigating family interactions, such as older siblings monitoring their younger siblings' interaction with the camera or parents capturing playful dominance over their children's reactions. Participatory research is vital in adolescent research as it can examine power and voice while simultaneously being an empowering activity, building agency and giving voice to underrepresented or marginalized groups [47]. This shift in focus from researcher-controlled observation to participant-driven recordings highlights a key aspect of this study's participatory methodology. Drawing from the principles of visual methodologies and participatory research rooted in the traditions of ethnography within anthropology and sociology [43,46], this method empowered participants to engage with the activity on their own terms.

The recordings enabled participants to document their reactions to *MTV Shuga* within the familiarity of their physical environment. Filming reactions in the comfort of their own homes or chosen environments, participants provided insights closely reflecting how they naturally might consume and engage with *MTV Shuga*. While open living rooms with communal activities created potential distractions, more intimate settings such as bedrooms allowed participants to engage deeply with the content. Outdoor spaces offered another setting for engagement as participants viewed in peer groups away from home. The distractions from young siblings and the disruptions of coinciding activities such as smartphone use or passersby could have impeded the educational and entertainment value of watching *MTV Shuga*.

XSL•FO
RenderX

While these various physical environments shaped the nature of participants' engagement with the content, social dynamics within families also played a significant role in how the show was received. Even during the recruitment phase, participants resisted joining the qualitative reaction videos when we initially requested that adolescents (aged 15-19 years) watch the show with their parents. While we initially hypothesized that *MTV Shuga* could have a "family influence" by encouraging communication between parents and adolescents, we have since observed resistance to this in other research activities, including in-depth interviews and focus group discussions, in which adolescents cite discomfort in viewing sexually explicit content with parents for fear of awkwardness and judgment [21,27,36]. Some parents shared an interest in viewing and learning from the show, but encouraging "family" or intergenerational viewings can depend on the cultural context, for example, whether social norms about adolescents or sex can be challenged safely.

Participants noted that watching with peers of similar ages made the experience more engaging, sparking discussions and learning opportunities. The groups that reported engaging in postviewing discussions were mainly young men who viewed the show outside their homes. As observed in our previous study, creating youth-friendly spaces for media consumption could further enhance engagement with sexual health content [21].

Reaction videos capture the spontaneity of viewers' real-time responses, providing insights into their immediate emotional reactions to the show's content. Follow-up questions with a researcher further enrich these data, allowing participants to reflect on and articulate their experiences. This approach offers a dual layer of insight—combining initial reactions with more reflective postviewing feedback.

In particular, real-time reactions among viewing groups with mixed ages revealed that visual intimacy scenes in *MTV Shuga*, such as characters kissing, undressing, or reaching for condoms, caused discomfort for many participants. This led to distractions and awkwardness as participants covered their faces, walked out of the room, or picked up their phones often during critical and short HIV prevention scenes. It is not clear whether these reactions were due to the presence of others in the viewing group or the presence of the camera (spurred by social expectations) and may not have happened in the absence of either.

Despite the embarrassment caused by visual intimacy, participants approached discussions about sexual topics with more comfort when not accompanied by intimate scenes. For instance, several groups engaged in animated discussions about Q and Dineo's relationship and transactional sex. Groups of young men, for example, were content watching scenes of Reggie and Odirile discussing sexual health but reacted negatively when the couple kissed on-screen.

These findings suggest that *MTV Shuga* may be more effective in facilitating conversations around sexual health and relationships if it reduces the intimate scenes that make viewing uncomfortable. However, the show must balance its aim of creating destigmatized representative content (eg, to help normalize kissing between a same-sex couple) with creating content that is comfortable for its audience [48]. In addition, in some cases, generating a strong or discomfiting reaction can challenge viewers to think and learn and leave a more lasting impression as long as it is not harmful or gratuitous. Reducing intimate scenes could dilute the drama and sexual appeal that make the show compelling, potentially diminishing its appeal to younger viewers.

Beyond these broader social dynamics, participants' responses to specific content within *MTV Shuga* revealed key insights into the reception of public health messages. For example, the HIV screening scene, particularly the depiction of a swab for self-screening, generated limited dialogue and, concerningly, misconceptions about HIV transmission through saliva. Similar findings were documented in our earlier research [4]. To prevent misunderstandings and concerns among viewers, future series can embed clearer explanations of how oral swabs check for HIV antibodies in saliva and do not detect the HIV virus itself given the absence of real-time clarification from health professionals in media interventions.

Alongside responses to HIV prevention content, participants' reactions to the show's musical elements revealed how music shaped their engagement with the show. Reaction videos captured the immediate somatic responses of viewers to the music featured in *MTV Shuga* as they sang, danced, and snapped their fingers to the rhythm. Music holds a significant place in South African edutainment, particularly in programs that combine performance theatre and music to promote public health and HIV awareness [30-32]. These artistic methods leverage music's emotional and rhythmic power to create a sense of interconnectedness and euphoria among audiences, making them more receptive to new ideas and behaviors [49].

Young people find the integration of music into health interventions both engaging and refreshing, which is crucial given that they are often overwhelmed and fatigued by the constant stream of HIV messaging and promotions. This saturation can lead to disengagement, making innovative approaches such as music-based edutainment particularly valuable in capturing their attention and encouraging meaningful engagement [32].

Limitations and Adaptations

The findings of this study, involving a small sample of 8 reaction videos from a specific region of South Africa, may not be generalizable to other social and cultural contexts. The demographics of the participants do not represent *MTV Shuga* viewers in South Africa in general as no groups were female only. Female individuals constitute the largest viewing group of the MTV series [22]. Recruitment challenges, especially in engaging parents, impacted this study's original design, constraining insights into parent-child dynamics during *MTV Shuga* viewings.

This study was conducted during the COVID-19 pandemic, likely impacting participants' spaces and viewing habits. As a result, the viewing environment may not reflect the participants' natural choices. COVID-19 restrictions also limited participants' ability to select their preferred viewing partners, potentially affecting the authenticity of group composition and individual

XSL•FO
RenderX

responses. Their reactions may have been influenced by peers' or family members' collective atmosphere and expectations.

Not all participants adhered to instructions to ignore the camera, indicating that the instructions were not clear to all and may have caused variation in the participants' reactions to the activity. Nonetheless, their interactions with the camera provided valuable insights into social dynamics.

The success of this method hinges on participants' ease with technology and access to the necessary resources, such as the internet, smartphones, and a laptop, further restricting who would be able to participate in this study. Challenges arising from the reaction video method included participants using personal devices for filming, impacting the quality of audio and visuals in the recordings. The videos' narrow view makes it challenging to understand dynamics outside the camera's frame within the room or other spaces.

A final limitation of the reaction video method is that it primarily captures only what is externally visible, audible, or physical. It fails to delve into the internal processes, such as thoughts and emotions, that are crucial for understanding the deeper cognitive and emotional responses of participants. As a result, a significant layer of participant experience remains unexplored.

The reaction videos provided valuable insights into both the social context and content engagement, and these findings also point to opportunities for expanding the approach in future studies. By adopting a more interactive research design, reaction videos could capture different types of data, offering deeper insights into participants' thoughts and experiences. Researchers could capture more verbal data and gain insights into inner thoughts by asking participants to speak directly into the camera and express their thoughts more overtly. This method captures thoughts and feelings lost in a less interactive observational approach. Reaction videos can also be adapted to solo reactions, which could yield insights into more personalized viewpoints, again akin to those shared on social media platforms. As noted by some participants, some would be more comfortable viewing *MTV Shuga* alone.

Conclusions

This study's exploration of viewer engagement with *MTV Shuga* through reaction videos illuminated the nuanced impact of physical and social environments on the consumption and influence of edutainment. By capturing real-time reactions and discussions among various viewing groups in their chosen settings, this research method offered valuable insights into the immediate emotional responses and subsequent reflections on themes of intimacy, HIV prevention, and transactional sex. In addition, the innovative use of these reaction videos represents a significant opportunity for adolescent health research. This participatory approach respects young people's autonomy and perspectives and provides deeper insights into their interactions with health-related content. Furthermore, it showed reaction videos to be a valuable method to capture dimensions of context, the physical and social context in which edutainment is consumed, which has not been documented using other evaluation methods.

This study acknowledges the inherent limitations in capturing the authentic viewer dynamics due to the self-selective nature of the viewing groups, the challenges posed by the COVID-19 context, and the observer effect of the camera. The findings suggest that, while reaction videos can significantly enrich our understanding of audience reception, careful consideration must be given to the potential influences on participant behavior and the representativeness of the data collected.

Acknowledgments

The authors thank the young people who gave their time to this study. They especially thank Lunga Jonas for her beautiful artwork, which was used to capture the self-filmed observations. The authors are grateful to Luntu Fica, Bawelile Madlala, Samkele Khuma, and Lunga Mkrezo for their hard work collecting data and valuable ideas that strengthened this research and to Antonio Duran-Aparicio for his support managing the grant. The authors acknowledge the dedication of *MTV Shuga* peer educators and coordinators (Yvonne Diogo and Lesedi Thwala) and the commitment of the MTV Staying Alive Foundation, especially Alistair Chase and Sara Piot.

Data Availability

The datasets generated during and analyzed during this study are not publicly available due to the identifiable nature of the data but nonidentifiable data such as transcripts and summaries are available from the corresponding author on reasonable request.

Authors' Contributions

VB and IB conceived the study design and methods. VB, DK, CC, SC, and IB codeveloped the study protocol, and CC, DK, and VB managed the data collection. VB led the observation data analysis and wrote the first draft. IB and VB co-designed the analysis framework. All authors contributed to revisions and reviewed the final draft.

Conflicts of Interest

GA was the executive director of the MTV Staying Alive Foundation at the time this study was designed and conducted and has since left the foundation.

XSL•FO
RenderX

Multimedia Appendix 1
Illustrations and descriptions of the viewing groups partaking in the self-filmed reaction videos.
[PNG File , 4911 KB - formative_v9i1e55275_app1.png]

References

1. Kincaid D, Parker W, Johnson S, Schierhout G, Kelly K, Connolly C, et al. AIDS communication programmes, HIV prevention, and living with HIV and AIDS in South Africa, 2006. The United States Agency for International Development. 2008. URL: https://www.researchgate.net/publication/263234176_AIDS_Communication_Programmes_HIV_Prevention_and_Living_with_HIV_and_AIDS_in_South_Africa_2006 [accessed 2024-04-29]
2. LaCroix JM, Snyder LB, Huedo-Medina TB, Johnson BT. Effectiveness of mass media interventions for HIV prevention, 1986-2013: a meta-analysis. J Acquir Immune Defic Syndr 2014 Aug 15;66 Suppl 3:S329-S340. [doi: 10.1097/QAI.0000000000000230] [Medline: 25007204]
3. Pichon M, Carter DJ, Howard-Merrill L, Sono R, Gimunta V, Rutenge O, et al. A mixed-methods, exploratory, quasi-experimental evaluation of a radio drama intervention to prevent age-disparate transactional sex in Tanzania. Front Reprod Health 2022;4:1000853 [FREE Full text] [doi: 10.3389/frph.2022.1000853] [Medline: 36531443]
4. Birdthistle I, Mulwa S, Sarrassat S, Baker V, Khanyile D, O'Donnell D, et al. Effects of a multimedia campaign on HIV self-testing and PrEP outcomes among young people in South Africa: a mixed-methods impact evaluation of 'MTV Shuga Down South'. BMJ Glob Health 2022 Apr;7(4):e007641 [FREE Full text] [doi: 10.1136/bmjgh-2021-007641] [Medline: 35365480]
5. Orozco-Olvera V, Shen F, Cluver L. The effectiveness of using entertainment education narratives to promote safer sexual behaviors of youth: a meta-analysis, 1985-2017. PLoS One 2019;14(2):e0209969 [FREE Full text] [doi: 10.1371/journal.pone.0209969] [Medline: 30753185]
6. Singhal A, Rogers E. Entertainment-Education: A Communication Strategy for Social Change. New York, NY: Routledge; 2012.
7. Orr G. Diffusion of innovations, by Everett Rogers (1995). Stanford University. URL: https://web.stanford.edu/class/symbsys205/Diffusion%20of%20Innovations [accessed 2024-04-29]
8. Cardey S, Garforth C, Govender E, Dyll-Myklebust L. Entertainment education theory and practice in HIV/AIDS communication: a South Africa/United Kingdom comparison. Critical Arts 2013 Jun;27(3):288-310. [doi: 10.1080/02560046.2013.800665]
9. Usdin S, Scheepers E, Goldstein S, Japhet G. Achieving social change on gender-based violence: a report on the impact evaluation of Soul City's fourth series. Soc Sci Med 2005 Dec;61(11):2434-2445. [doi: 10.1016/j.socscimed.2005.04.035] [Medline: 16006028]
10. Goldstein S, Usdin S, Scheepers E, Japhet G. Communicating HIV and AIDS, what works? A report on the impact evaluation of Soul City's fourth series. J Health Commun 2005;10(5):465-483. [doi: 10.1080/10810730591009853] [Medline: 16199389]
11. The urgency of now: AIDS at a crossroads. Joint United Nations Programme on HIV/AIDS (UNAIDS). URL: https://www.unaids.org/sites/default/files/media_asset/2024-unaids-global-aids-update_en.pdf [accessed 2024-04-29]
12. Peltzer K, Parker W, Mabaso M, Makonko E, Zuma K, Ramlagan S. Impact of national HIV and AIDS communication campaigns in South Africa to reduce HIV risk behaviour. ScientificWorldJournal 2012;2012:384608-384606 [FREE Full text] [doi: 10.1100/2012/384608] [Medline: 23213285]
13. Bennet N. The AIDS conspiracy: science fights back. Lancet Infect Dis 2012 Aug;12(8):596. [doi: 10.1016/S1473-3099(12)70195-0]
14. Fassin D, Schneider H. The politics of AIDS in South Africa: beyond the controversies. BMJ 2003 Mar 01;326(7387):495-497 [FREE Full text] [doi: 10.1136/bmj.326.7387.495] [Medline: 12609950]
15. Fourie P, Meyer M. The Politics of AIDS Denialism: South Africa's Failure to Respond. New York, NY: Routledge; 2016.
16. Chigwedere P, Essex M. AIDS denialism and public health practice. AIDS Behav 2010 Apr;14(2):237-247. [doi: 10.1007/s10461-009-9654-7] [Medline: 20058063]
17. Tufte T. Entertainment-education and participation: assessing the communication strategy of Soul City. J Int Commun 2001 Dec;7(2):25-50. [doi: 10.1080/13216597.2001.9751908]
18. Francis SC, Mthiyane TN, Baisley K, Mchunu SL, Ferguson JB, Smit T, et al. Prevalence of sexually transmitted infections among young people in South Africa: a nested survey in a health and demographic surveillance site. PLoS Med 2018 Feb;15(2):e1002512 [FREE Full text] [doi: 10.1371/journal.pmed.1002512] [Medline: 29485985]
19. Tufte T. Fighting AIDS with edutainment: building on the Soul City experience in South Africa. In: Servaes J, editor. Communication for Development and Social Change. Thousand Oaks, CA: Sage Publications; 2008:327-344.
20. Bertrand JT, O'Reilly K, Denison J, Anhang R, Sweat M. Systematic review of the effectiveness of mass communication programs to change HIV/AIDS-related behaviors in developing countries. Health Educ Res 2006 Aug;21(4):567-597. [doi: 10.1093/her/cyl036] [Medline: 16847044]

XSL•FO
RenderX

21. Baker V, Mulwa S, Sarrassat S, Khanyile D, Cousens S, Cawood C, et al. 'It is guiding us to protect ourselves': a qualitative investigation into why young people engage with a mass-media HIV education campaign. Cult Health Sex 2023 Jul 28;25(7):803-817. [doi: 10.1080/13691058.2022.2100483] [Medline: 35900942]
22. Baker V, Arnold G, Piot S, Thwala L, Glynn J, Hargreaves J, et al. Young adults' responses to an African and US-Based COVID-19 edutainment miniseries: real-time qualitative analysis of online social media engagement. JMIR Form Res 2021 Oct 29;5(10):e30449 [FREE Full text] [doi: 10.2196/30449] [Medline: 34596568]
23. Jana M, Letsela L, Scheepers E, Weiner R. Understanding the role of the OneLove campaign in facilitating drivers of social and behavioral change in southern Africa: a qualitative evaluation. J Health Commun 2015 Dec 13;20(3):252-258. [doi: 10.1080/10810730.2014.925014] [Medline: 25495105]
24. Makwambeni B, Salawu A. Accounting for youth audiences' resistances to HIV and AIDS messages in the television drama Tsha Tsha in South Africa. SAHARA J 2018 Dec;15(1):20-30 [FREE Full text] [doi: 10.1080/17290376.2018.1444506] [Medline: 29508641]
25. Pichon M, Buller AM, Gimunta V, Rutenge O, Thiaw Y, Sono R, et al. Qualitative evaluation of an edutainment intervention to prevent age-disparate transactional sex in Tanzania: changes in educational aspirations and gender equitable attitudes towards work. PLOS Glob Public Health 2024;4(4):e0002527 [FREE Full text] [doi: 10.1371/journal.pgph.0002527] [Medline: 38568918]
26. Cameron M, Cockcroft A, Waichigo GW, Marokoane N, Laetsang D, Andersson N. From knowledge to action: participant stories of a population health intervention to reduce gender violence and HIV in three southern African countries. AIDS Care 2014 Jul 03;26(12):1534-1540. [doi: 10.1080/09540121.2014.931560] [Medline: 24991886]
27. Kyegombe N, Zuma T, Hlongwane S, Nhlenyama M, Chimbindi N, Birdthistle I, et al. A qualitative exploration of the salience of MTV-Shuga, an edutainment programme, and adolescents' engagement with sexual and reproductive health information in rural KwaZulu-Natal, South Africa. Sex Reprod Health Matters 2022 Dec 05;30(1):2083809 [FREE Full text] [doi: 10.1080/26410397.2022.2083809] [Medline: 35929964]
28. Le Port A, Seye M, Heckert J, Peterman A, Nganya Tchamwa A, Dione M, et al. A community edutainment intervention for gender-based violence, sexual and reproductive health, and maternal and child health in rural Senegal: a process evaluation. BMC Public Health 2022 Jun 10;22(1):1165 [FREE Full text] [doi: 10.1186/s12889-022-13570-6] [Medline: 35689180]
29. Pedrana A, Hellard M, Gold J, Ata N, Chang S, Howard S, et al. Queer as F**k: reaching and engaging gay men in sexual health promotion through social networking sites. J Med Internet Res 2013 Feb 07;15(2):e25 [FREE Full text] [doi: 10.2196/jmir.2334] [Medline: 23391459]
30. Walker GR. Musicking in applied theatre: exploring interdisciplinary approaches to drama-based health and social interventions. Res Drama Educ 2023 Aug 23;29(1):187-205. [doi: 10.1080/13569783.2023.2249409]
31. Dalrymple L, Jaffe A. Dramaide, a project in schools in Kwazulu-Natal, South Africa. In: Schenker II, Sabar-Friedman G, Sy FS, editors. AIDS Education: Interventions in Multi-Cultural Societies. Boston, MA: Springer; 1996:119-123.
32. Walker GR. Emotive media as a counterbalance to AIDS messaging fatigue in South Africa: responses to an HIV/AIDS awareness music video. AIDS Educ Prev 2022 Feb;34(1):17-32. [doi: 10.1521/aeap.2022.34.1.17] [Medline: 35192396]
33. Banerjee A, Ferrara EL, Orozco-Olvera V. The entertaining way to behavioral change: fighting HIV with MTV. National Bureau of Economic Research. URL: https://www.nber.org/papers/w26096 [accessed 2024-04-29]
34. Shahmanesh M, Mthiyane N, Chimbindi N, Zuma T, Dreyer J, Birdthistle I, et al. P407 ‘MTV shuga’: mass media communication, HSV2 and sexual health in adolescent girls and young women in rural south africa. BMJ 2019;95(Issue Suppl 1):A196. [doi: 10.1136/sextrans-2019-sti.496]
35. Hutchinson P, Beaudoin CE, Meekers D, Omoluabi E, Akinyemi A. "You need to be able to stand up for what is right": MTV Shuga Naija's transformative impact on youth attitudes toward sexual violence in Nigeria. J Interpers Violence 2024 Aug 02:8862605241265408. [doi: 10.1177/08862605241265408] [Medline: 39092629]
36. Mulwa S, Baker V, Cawood C, Khanyile D, O'Donnell D, Sarrassat S, et al. Evaluating the intensity of exposure to MTV Shuga, an edutainment program for HIV prevention: cross-sectional study in eastern cape, South Africa. JMIR Form Res 2024 Feb 13;8:e44111 [FREE Full text] [doi: 10.2196/44111] [Medline: 38349729]
37. Baker L. Observation: a complex research method. Libr Trends 2006 Jun;55(1):171-189. [doi: 10.1353/lib.2006.0045]
38. McDaniel B. Popular music reaction videos: reactivity, creator labor, and the performance of listening online. New Media Soc 2020 Jun 02;23(6):1624-1641. [doi: 10.1177/1461444820918549]
39. Goddard M. Sincere, authentic, remediated: the affective labour and cross cultural remediations of music video reaction videos on YouTube. In: Rogers H, Freitas J, Porfirio JF, editors. Remediating Sound: Repeatable Culture, YouTube and Music. New York, NY: Bloomsbury Academic; 2023:73-92.
40. Swan AL. Transnational identities and feeling in fandom: place and embodiment in K-pop fan reaction videos. Commun Cult Crit 2018;11(4):548-565. [doi: 10.1093/ccc/tcy026]
41. Budig K, Diez J, Conde P, Sastre M, Hernán M, Franco M. Photovoice and empowerment: evaluating the transformative potential of a participatory action research project. BMC Public Health 2018 Apr 02;18(1):432 [FREE Full text] [doi: 10.1186/s12889-018-5335-7] [Medline: 29609576]

XSL•FO
RenderX

42. Sadati SM, Mitchell C, Nayriro M, Muthuri S, Njeri M. Critical perspectives on photovoice as a tool to explore the challenges of balancing work and childcare among mothers in Kenya. Institute for Studies in Industrial Development. 2019. URL: https://tinyurl.com/txwvem6z [accessed 2024-04-29]
43. Asan O, Montague E. Using video-based observation research methods in primary care health encounters to evaluate complex interactions. Inform Prim Care 2014;21(4):161-170 [FREE Full text] [doi: 10.14236/jhi.v21i4.72] [Medline: 25479346]
44. Sparrman A. Video recording as interaction: participant observation of children's everyday life. Qual Res Psychol 2005 Jan;2(3):241-255. [doi: 10.1191/1478088705qp041oa]
45. Gorden RL. Interviewing: Strategy, Techniques, and Tactics. Santa Barbara, CA: Dorsey Press; 1987.
46. Glaw X, Inder K, Kable A, Hazelton M. Visual methodologies in qualitative research: autophotography and photo elicitation applied to mental health research. Int J Qual Methods 2017 Dec 19;16(1):160940691774821. [doi: 10.1177/1609406917748215]
47. Bradbury-Jones C, Isham L, Taylor J. The complexities and contradictions in participatory research with vulnerable children and young people: a qualitative systematic review. Soc Sci Med 2018 Oct;215:80-91. [doi: 10.1016/j.socscimed.2018.08.038] [Medline: 30218806]
48. Booker NA, Miller AN, Ngure P. Heavy sexual content versus safer sex content: a content analysis of the entertainment education drama Shuga. Health Commun 2016 Dec 07;31(12):1437-1446. [doi: 10.1080/10410236.2015.1077691] [Medline: 27054689]
49. Robert Walker G. Pulsing bodies and embodying pulse: musical effervescence in a South African HIV/AIDS community outreach program. Anthropol Med 2022 Sep 16;29(3):289-304. [doi: 10.1080/13648470.2021.1994335] [Medline: 34913756]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
DS2: Down South season 2
PrEP: pre-exposure prophylaxis

Edited by G Eysenbach; submitted 05.03.24; peer-reviewed by C Reagon, G Walker; comments to author 21.06.24; revised version received 30.07.24; accepted 01.10.24; published 31.01.25.

Please cite as:
Baker V, Mulwa S, Khanyile D, Arnold G, Cousens S, Cawood C, Birdthistle I
Evaluating Reaction Videos of Young People Watching Edutainment Media (MTV Shuga): Qualitative Observational Study
JMIR Form Res 2025;9:e55275
URL: https://formative.jmir.org/2025/1/e55275
doi:10.2196/55275
PMID:

©Venetia Baker, Sarah Mulwa, David Khanyile, Georgia Arnold, Simon Cousens, Cherie Cawood, Isolde Birdthistle. Originally published in JMIR Formative Research (https://formative.jmir.org), 31.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Proficiency, Clarity, and Objectivity of Large Language Models Versus Specialists' Knowledge on COVID-19's Impacts in Pregnancy: Cross-Sectional Pilot Study

Nicola Luigi Bragazzi[1,2], MPH, MD, PhD; Michèle Buchinger[3], MD; Hisham Atwan[4], MD; Ruba Tuma[3,5], MD; Francesco Chirico[6], MD; Lukasz Szarpak[7,8,9], MD; Raymond Farah[3], MD; Rola Khamisy-Farah[3], MD

[1]Laboratory for Industrial and Applied Mathematics, Department of Mathematics and Statistics, York University, Toronto, ON, Canada

[2]Department of Computer Science, Data Science, and Information Technology, Faculty of Natural and Applied Sciences, Sol Plaatje University, Kimberley, South Africa

[3]Azrieli Faculty of Medicine, Bar-Ilan University, Ramat Gan, Israel

[4]Kaplan Medical Centre, Department of Internal Medicine, Hebrew University, Rehovot, Israel

[5]Department of Obstetrics and Gynecology, Galilee Medical Center, Nahariya, Israel

[6]Post-Graduate School of Occupational Health, Università Cattolica del Sacro Cuore, Rome, Italy

[7]Henry JN Taub Department of Emergency Medicine, Baylor College of Medicine, Houston, TX, United States

[8]Department of Clinical Research and Development, LUXMED Group, Warsaw, Poland

[9]Institute of Medical Science, Collegium Medicum, The John Paul II Catholic University of Lublin, Lublin, Poland

Corresponding Author:
Nicola Luigi Bragazzi, MPH, MD, PhD
Laboratory for Industrial and Applied Mathematics
Department of Mathematics and Statistics
York University
4700 Keele Street
Toronto, ON, M3J 1P3
Canada
Phone: 1 416 736 2100
Email: robertobragazzi@gmail.com

Abstract

Background: The COVID-19 pandemic has significantly strained health care systems globally, leading to an overwhelming influx of patients and exacerbating resource limitations. Concurrently, an "infodemic" of misinformation, particularly prevalent in women's health, has emerged. This challenge has been pivotal for health care providers, especially gynecologists and obstetricians, in managing pregnant women's health. The pandemic heightened risks for pregnant women from COVID-19, necessitating balanced advice from specialists on vaccine safety versus known risks. In addition, the advent of generative artificial intelligence (AI), such as large language models (LLMs), offers promising support in health care. However, they necessitate rigorous testing.

Objective: This study aimed to assess LLMs' proficiency, clarity, and objectivity regarding COVID-19's impacts on pregnancy.

Methods: This study evaluates 4 major AI prototypes (ChatGPT-3.5, ChatGPT-4, Microsoft Copilot, and Google Bard) using zero-shot prompts in a questionnaire validated among 159 Israeli gynecologists and obstetricians. The questionnaire assesses proficiency in providing accurate information on COVID-19 in relation to pregnancy. Text-mining, sentiment analysis, and readability (Flesch-Kincaid grade level and Flesch Reading Ease Score) were also conducted.

Results: In terms of LLMs' knowledge, ChatGPT-4 and Microsoft Copilot each scored 97% (32/33), Google Bard 94% (31/33), and ChatGPT-3.5 82% (27/33). ChatGPT-4 incorrectly stated an increased risk of miscarriage due to COVID-19. Google Bard and Microsoft Copilot had minor inaccuracies concerning COVID-19 transmission and complications. In the sentiment analysis, Microsoft Copilot achieved the least negative score (–4), followed by ChatGPT-4 (–6) and Google Bard (–7), while ChatGPT-3.5 obtained the most negative score (–12). Finally, concerning the readability analysis, Flesch-Kincaid Grade Level and Flesch Reading Ease Score showed that Microsoft Copilot was the most accessible at 9.9 and 49, followed by ChatGPT-4 at 12.4 and 37.1, while ChatGPT-3.5 (12.9 and 35.6) and Google Bard (12.9 and 35.8) generated particularly complex responses.

XSL•FO
RenderX

Conclusions: The study highlights varying knowledge levels of LLMs in relation to COVID-19 and pregnancy. ChatGPT-3.5 showed the least knowledge and alignment with scientific evidence. Readability and complexity analyses suggest that each AI's approach was tailored to specific audiences, with ChatGPT versions being more suitable for specialized readers and Microsoft Copilot for the general public. Sentiment analysis revealed notable variations in the way LLMs communicated critical information, underscoring the essential role of neutral and objective health care communication in ensuring that pregnant women, particularly vulnerable during the COVID-19 pandemic, receive accurate and reassuring guidance. Overall, ChatGPT-4, Microsoft Copilot, and Google Bard generally provided accurate, updated information on COVID-19 and vaccines in maternal and fetal health, aligning with health guidelines. The study demonstrated the potential role of AI in supplementing health care knowledge, with a need for continuous updating and verification of AI knowledge bases. The choice of AI tool should consider the target audience and required information detail level.

(JMIR Form Res 2025;9:e56126) doi:10.2196/56126

KEYWORDS

COVID-19; vaccine; reproductive health; generative artificial intelligence; large language model; chatGPT; google bard; microsoft copilot; vaccination; natural language processing; obstetric; gynecology; women; text mining; sentiment; accuracy; zero shot; pregnancy; readability; infectious

Introduction

The COVID-19 pandemic has exerted unprecedented strain on health care systems worldwide, overwhelming hospitals and health care facilities with patients and stretching resources and staff to their limits, thus highlighting the critical need for efficient health care delivery, and underscoring the importance of robust, resilient public health infrastructure [1].

In addition, the pandemic gave rise to an "infodemic"—that is to say, a flood of information, often misleading or false, that spread through social media and other channels [2], which has significantly impacted maternal and child health [3]. This deluge of misinformation made the task of retrieving and providing reliable, science-based information more challenging yet crucial [4] and health care providers, including gynecologists and obstetricians, played a pivotal role during the pandemic, especially in advising pregnant women.

Pregnancy, by its nature, puts women at a higher risk for complications from viral infections like COVID-19 [5,6]. As vaccines became available, gynecologists and obstetricians were at the forefront, recommending COVID-19 vaccinations to pregnant women, and having to balance the lack of initial data on vaccine safety in pregnancy with the known risks of COVID-19, providing tailored advice to their patients. The role of these specialists, crucial in safeguarding the health of mothers and their unborn children, faced significant challenges during the pandemic. This period adversely affected their professional relationships with patients and colleagues, leading to anxiety, frustration, and physical strain. These stressors, stemming from both the workplace and pandemic conditions, have prompted some gynecologists and obstetricians to consider early retirement or leaving the profession, highlighting an increased risk of burnout among these specialists [7].

The advent of generative artificial intelligence (AI), such as large language models (LLMs) and related disruptive technologies, has shown immense potential in various fields, including health care [8]. These AI-based systems hold great potential in assisting and supporting biomedical researchers, health care professionals, and the general public. They can process vast amounts of data, provide medical insights, assist in diagnostic processes, and even support therapies. In the context of the COVID-19 pandemic, AI has been used not only to analyze virus structures, predict outbreaks, and sift through vast amounts of research to identify potential treatments [9] but also to respond to queries from people, providing accurate and timely information on symptoms, prevention measures, and vaccination.

However, with the rapid development and deployment of these AI-driven technologies, there arises an essential need to test the knowledge and proficiency of AI-based systems rigorously [10]. Ensuring the accuracy, reliability, and ethical use of AI is vital, especially in sensitive areas like health care, and involves continuous monitoring, evaluation, and updating of AI systems to align with the latest scientific knowledge and ethical standards. As AI becomes more integrated into human lives, the necessity for such oversight becomes increasingly important to prevent misuse and ensure that these powerful tools serve the greater good effectively and responsibly.

Recent studies have explored the potential use of AI and LLMs in public health initiatives, particularly in the context of immunization campaigns [11-20]. These investigations have demonstrated that AI can play a critical role in countering misinformation and dismissing conspiracy theories, thereby fostering trust and acceptance of vaccines among the public. By providing clear, evidence-based responses to common myths and misconceptions, AI-driven systems can effectively address vaccine hesitancy and promote informed decision-making. However, besides being limited in number, these studies are also constrained in scope, often concentrating on individual models without providing a comprehensive comparative analysis of their performance. This lack of comparative insight leaves gaps in understanding the relative strengths and weaknesses of different AI systems in supporting public health efforts, highlighting the need for broader investigations that include a variety of generative AI tools to fully assess their potential and limitations in these critical domains.

As of today, the landscape of AI prototypes is dominated by 3 leading models, that are ChatGPT-4 (which has surpassed its predecessor, ChatGPT-3.5), Microsoft Copilot, and Google

XSL•FO
RenderX

Bard (now rebranded as Google Gemini). ChatGPT-3.5, released by OpenAI and based on the GPT-3.5 model, is a conversational AI model designed primarily for text-based interactions, and skilled in a variety of tasks like answering questions, providing explanations, and generating text based on prompts, among others. The model's knowledge is up to the date it was last trained, which means it does not have information on events or developments that occurred after that date. ChatGPT-4 represents its further iteration. As such, this advanced version by OpenAI, succeeding ChatGPT-3.5, built on the GPT-4 architecture, offers enhanced capabilities in terms of understanding and generating more complex and nuanced text, and is more contextually aware, providing more relevant and accurate responses. Like its predecessor, it is trained up to a certain point in time and lacks real-time knowledge. Microsoft Copilot is primarily focused on assisting with coding, problem-solving, and programming tasks and can be integrated with development tools like GitHub and Visual Studio Code (Microsoft Inc). Finally, Google Bard, powered by Google's own language models, including BERT, is designed to integrate Google's search capabilities, potentially offering more up-to-date information, and is aimed at providing conversational AI experiences, including answering questions or generating text, with a strong emphasis on leveraging the web for current information.

Each of these models has unique strengths and applications, with ChatGPT models (3.5 and 4) focusing on general conversational tasks, Microsoft Copilot on programming assistance, and Google Bard on integrating search capabilities into conversational AI. As such, each of these LLMs was tested in the present investigation. More in detail, the research objective of this pilot study was to evaluate the applicability of major generative AI model prototypes in providing evidence-based responses to COVID-19–related queries, particularly in the context of maternal and child health. The evaluation focused on key aspects such as accuracy, reliability, and readability. By benchmarking these models against a range of factual statements, the study aimed to assess their potential not only to support health care professionals in decision-making and public health communication during pandemics but also to address the challenges faced by the public in accessing accurate and reliable information.

The findings will help determine the role of AI in enhancing health care delivery and improving public understanding of critical health issues during crises. By highlighting the dual challenge of supporting specialists and ensuring the public is correctly informed, this research underscores the complex interplay between information dissemination and health care communication. As a pilot study, it serves as a foundational step, emphasizing the need for further, more comprehensive investigations to fully explore and validate the capabilities and limitations of AI in health care and public health contexts.

Methods

Procedure

All 4 AI-based tools were queried about COVID-19 and related vaccines during pregnancy using zero-shot prompts, that is to say, without any previous examples or context specific to that task. They were queried all on the same day, on January 7, 2024.

In a comparative evaluation of ChatGPT-3.5, ChatGPT-4, Google Bard, and Microsoft Copilot, these AI-based tools were tested using a validated questionnaire. This had previously been administered to a sample of 159 Israeli obstetrician-gynecologists, with an average age of 44.9 years (SD 12.48), predominantly female (59.7%), and largely composed of attending physicians (60.4%) [21]. This assessment aimed to evaluate the AIs' ability to provide accurate information on COVID-19, particularly in the context of pregnancy, fertility, and related clinical issues, as well as to make evidence-based recommendations. The questionnaire was developed following a comprehensive review of the literature and in accordance with World Health Organization (WHO) guidelines. It consisted of 5 sections designed to capture a range of relevant information. The first section collected sociodemographic data, including participants' gender, age, marital status, number of children, religion, level of religious observance, professional role (resident or attending physician), years of practice, as well as previous COVID-19 infection and vaccination status. The second section assessed general knowledge about COVID-19 through 5 true or false questions that addressed basic preventive measures, transmission routes, and management strategies. The third section focused on specific knowledge related to pregnancy, childbirth, and breastfeeding during the COVID-19 pandemic, including 14 questions exploring the risks and complications of COVID-19 for pregnant women, the potential for vertical transmission, the safety of breastfeeding, and related clinical practices. The fourth section addressed knowledge of COVID-19 vaccination through 14 questions that covered vaccine types, safety, efficacy, administration during pregnancy, and potential side effects. Finally, the fifth section examined attitudes and practices regarding COVID-19 vaccination. This section explored physicians' recommendations about vaccination timing during pregnancy, vaccination for women planning pregnancy, and associated safety considerations, with responses recorded on a 5-point Likert scale. To ensure clarity and relevance, the questionnaire was piloted with 12 obstetrician-gynecologists before full deployment. The survey was then distributed using both digital platforms and in-person delivery during professional gatherings, achieving a high response rate. Logistic regression analysis of the data revealed that knowledge of COVID-19 vaccination was a significant predictor of whether obstetrician-gynecologists recommended vaccination for pregnant women.

In the present comparative assessment, items from the second through fifth sections of the questionnaire were used. Sections 2 to 4 were included in their original format to specifically evaluate the AI systems' proficiency in delivering accurate knowledge about COVID-19 prevention, risks during pregnancy, and vaccination safety and efficacy. For section 5, instead of using the original Likert scale, the AI tools were queried with direct recommendation prompts such as "Would you recommend...?" The responses from section 5 were subsequently analyzed for readability and sentiment (refer to next sections),

XSL•FO
RenderX

providing insights into the clarity and tone of the AI-generated recommendations.

All the queried items, along with the corresponding prompts, are detailed in Multimedia Appendix 1.

Scoring

The scoring process of the replies provided by the four AI models (reported in Multimedia Appendix 2) was conducted independently by 2 obstetrician-gynecologists (RT and RK-F) and 1 internal medicine doctor (RF). To ensure impartiality, the biostatistician (NLB) analyzing the data was blinded to the identities of the LLMs during the scoring process.

Sentiment Analysis

Textual data analysis of AI-generated responses to the items of the fifth section of the validated questionnaire (Multimedia Appendix 3) was carried out using XLSTAT (Data Analysis and Statistical Solution for Microsoft Excel, Addinsoft) for each LLM-generated text to ensure comprehensive and model-specific insights. Each text underwent preprocessing to prepare it for analysis. This included removing English stop words, punctuation, and numerical characters, as well as filtering out words with a frequency of less than two occurrences. A sparsity threshold of 0.975 was applied to exclude terms appearing in fewer than 2.5% of the dataset, resulting in a more refined and representative corpus. Following preprocessing, a Term-Document Matrix (TDM) was constructed, providing a structured representation of the frequency of each term across the documents generated by the LLMs. The TDM facilitated comparative and statistical analyses by organizing the data into a matrix format, with columns representing unique terms and rows labeling the specific LLM. A Bag-of-Words representation was derived to enable both word frequency visualization and sentiment evaluation. Word clouds were generated to visually represent the most frequently occurring words for each LLM-generated text. The size and prominence of the words in the cloud reflected their frequency, offering an intuitive summary of the dominant themes and linguistic patterns specific to each model. To assess the emotional tone and polarity of the texts, sentiment analysis was performed using the TDM. Tokenization was applied to isolate individual words, and the Bing sentiment lexicon was used to assign binary polarity scores (positive or negative) to each term. These scores were then aggregated to calculate the overall sentiment for each LLM-generated text.

Readability Analysis

The readability of the 4 AI's responses to the items of the fifth section of the validated questionnaire (Multimedia Appendix 3) was appraised by means of the Flesch-Kincaid Grade Level analysis, which is a test designed to indicate how difficult a passage in English is to understand. The formula was developed by Rudolf Flesch and J Peter Kincaid and was first used in the United States military to assess the readability of their technical manuals. It is widely used in the field of education, but also in other sectors that require clarity of written communication, such as legal and health care industries. The formula is based on 2 factors, that are the average number of words per sentence and the average number of syllables per word. The formula is,

$$\text{Flesch-Kincaid Grade Level} = 0.39 \times (\text{average number of words per sentence}) + 11.8 \times (\text{average number of syllables per word}) - 15.59$$

The result of the Flesch-Kincaid Grade Level formula is a number that corresponds with a US school grade level. For example, a score of 8 means that the document can be understood by an eighth grader, with a higher score indicating more complex and difficult text.

In addition, the Flesch Reading Ease Score was also used to evaluate the readability of the responses. This score rates text on a 100-point scale, where higher scores indicate easier-to-read material. The formula for calculating the Flesch Reading Ease Score is

$$\text{Flesch Reading Ease Score} = 206.835 - 1.015 \times (\text{average number of words per sentence}) - 84.6 \times (\text{average number of syllables per word})$$

Higher scores on this scale indicate easier-to-read text, while lower scores indicate more complex material. For example, scores above 90 suggest very easy text suitable for an average 10- to 11-year-old (5th grade), while scores below 30 indicate very difficult text meant for college graduates.

Ethical Considerations

The primary study [21], which provided data from human participants, was exempted from formal approval by the Ethics Committee of the Medical Faculty at Bar-Ilan University in Tel Aviv, Israel, due to its survey-based design. Participants, consisting of obstetrician-gynecologists (both residents and attending physicians), voluntarily consented to participate in the study without receiving any compensation. All data were fully anonymized, ensuring the privacy and confidentiality of the participants. In addition, no ethical approval was required for the AI component of this study.

Results

LLMs' Knowledge of COVID-19 and Its Vaccines for Maternal and Fetal Health

ChatGPT-4 and Microsoft Copilot obtained a score of 97% (32/33), while Google Bard scored 94% (31/33), and ChatGPT-3.5 scored 82% (27/33; Multimedia Appendix 3).

ChatGPT-4 delivered correct responses across a range of topics, demonstrating a high level of proficiency in this specialized field. However, there was a notable exception in its performance: ChatGPT incorrectly stated that there is a heightened risk of miscarriage among pregnant women with COVID-19, even though, in reality, according to current literature, there is no evidence to support an increased risk of miscarriage due to COVID-19.

Similarly, Microsoft Copilot made only one mistake, related to the item concerning the increased risk of complications for pregnant women with COVID-19 compared to nonpregnant women of the same age. Google Bard's knowledge of COVID-19 and its vaccines for women's health was rather good, with the AI-based tool making only 2 mistakes, by ruling out COVID-19 intrauterine transmission, transmission during

XSL•FO
RenderX

delivery, or through contact and respiratory droplets during breastfeeding.

On the contrary, ChatGPT-3.5 exhibited poorer performance, with 6 wrong responses, demonstrating gaps in the knowledge of (1) risk, (2) complications of COVID-19 in pregnant women, (3) medication for pregnant and postpartum women with COVID-19, (4) risk of miscarriage in COVID-19–infected pregnant women, (5) mother-to-child transmission of COVID-19, and (6) COVID-19 transmission through breast milk.

LLMs' Attitudes and Practices Toward COVID-19 and Its Vaccines for Maternal and Fetal Health

Even if ChatGPT-3.5, ChatGPT-4, Microsoft Copilot, and Google Bard, as AI-based LLMs, cannot and do not provide personal opinions or medical recommendations, they can offer up-to-date and reliable information based on current guidelines and scientific understanding regarding COVID-19 vaccinations. In particular, concerning the recommendation of COVID-19 vaccine to all pregnant women (without contraindication), ChatGPT-4 stated that many health organizations, including the Centers for Disease Control and Prevention and the WHO, generally recommend COVID-19 vaccination for pregnant women due to the increased risk of severe illness from COVID-19 during pregnancy (including hospitalization, intensive care unit admission, and death). Google Bard concurred, emphasizing that the benefits of COVID-19 immunization far outweigh any potential risks to the mother or baby.

In terms of the proper timing for recommending the immunization of pregnant women against COVID-19, the AI-based chatbot clearly stated that the recommendation for COVID-19 vaccination during pregnancy often applies to all trimesters and, especially in the case of doubts, the timing should be discussed with a health care provider. Google Bard added that there is no evidence that vaccination at any particular trimester poses a higher risk to the mother or baby and Copilot stated that emphasis should be on vaccine receipt as soon as possible to maximize maternal and fetal health.

When queried about the recommendation of the COVID-19 vaccine only to pregnant women at high risk of contracting the virus, both ChatGPT-4 and Google Bard posited that the current guidance from many health organizations is to offer the vaccine to all pregnant women, not just those at high risk, due to the potentially severe impacts of COVID-19 during pregnancy.

Concerning the recommendation of COVID-19 vaccine to all women (without contraindication) of reproductive age who are not pregnant, ChatGPT-4 advised that COVID-19 vaccination is widely recommended for women of reproductive age to prevent severe illness, complications, and potential long-term effects of COVID-19. For those planning to undergo assisted reproduction, according to both the 2 AI-based tools, they are generally advised to get vaccinated against COVID-19, as there is no evidence that vaccines can affect fertility. When queried about the recommended interval between vaccination and pregnancy, the OpenAI chatbot stated that there is no recommended interval between receiving a COVID-19 vaccine and becoming pregnant according to most health organizations. Google Bard emphasized that there is no clear consensus on whether there is a need for an interval between vaccination and pregnancy, with some experts believing that there may be a slight increase in the risk of miscarriage in the first few weeks after vaccination, but this risk is very small, and other experts believing that there is no need to wait for any period of time before trying to conceive after vaccination. According to Google Bard, more research is warranted to determine whether there is a need for an interval between vaccination and pregnancy.

Finally, when tasked with recommending an interval between COVID-19 vaccination and the use of assisted reproductive technology (ART), ChatGPT-4 clarified that there is no specified interval recommended between COVID-19 vaccination and undergoing ART, while, according to Google Bard, there is no need for an interval between vaccination and ART, in that COVID-19 vaccination can safely be administered to women undergoing ART at any time.

On the contrary, ChatGPT-3.5 failed to offer up-to-date and reliable information, recommending immunization against COVID-19 in pregnant women at high risk, especially during the third trimester as it would be safer than during the first and second trimesters. Furthermore, ChatGPT-3.5 recommended an interval between vaccination and pregnancy as well as between immunization and ART based on vaccine type and individual risk factors.

LLMs' Text-Mining Analysis

The comparative text-mining analysis of Google Bard, Microsoft Copilot, ChatGPT-3.5, and ChatGPT-4 revealed both similarities and differences in the vocabulary used when discussing COVID-19 and immunization in relation to pregnancy. While all the models highlighted core themes such as "pregnant," "COVID-19," "vaccination," and "women," the nuances in their focus and vocabulary indicated variations in their approach to presenting the information. Google Bard and Microsoft Copilot both emphasized terms related to pregnancy and COVID-19, but their focal points diverged. Google Bard included more detailed terms associated with vaccination recommendations and the effects of COVID-19 on pregnancy outcomes, suggesting an emphasis on clinical guidance. In contrast, Microsoft Copilot featured broader terminology, incorporating aspects like "risk" and "breastfeeding," indicating a wider scope that touches on future planning and broader maternal and fetal health. When comparing Google Bard to ChatGPT-3.5 and ChatGPT-4, shared terms like "pregnant" and "vaccination" pointed to a shared understanding of key themes. However, Google Bard prioritized actionable guidance and recommendations, while ChatGPT variants included a more detailed vocabulary concerning medical and health-related risks. This distinction highlighted Google Bard's focus on practical advice, compared to ChatGPT's detailed explanations of risks and safety. Microsoft Copilot and the ChatGPT versions also shared several core terms but differed in their linguistic styles. Microsoft Copilot tended to lean toward broader generalizations, which makes it suitable for summarizing information succinctly. On the other hand, ChatGPT versions delved into specifics, incorporating medical and technical language to provide more

XSL•FO
RenderX

comprehensive information about COVID-19 and its impact on pregnancy. This distinction reflected the contrast between Copilot's generalist approach and ChatGPT's detailed focus. The comparison between ChatGPT-3.5 and ChatGPT-4 revealed a significant overlap in vocabulary, indicating consistency between the 2 versions. However, ChatGPT-4 introduced additional terms and nuances, suggesting an evolution in depth and breadth of content. These differences highlighted improvements in ChatGPT-4's ability to provide detailed and nuanced medical insights compared to its predecessor.

LLMs' Readability Analysis

The Flesch-Kincaid Grade Level and Flesch Reading Ease Score for Google Bard's response were 12.9 and 35.8, respectively, indicating that the text was highly complex and best understood by individuals with postsecondary education. Microsoft Copilot scored 9.9 and 49, suggesting a less complex response that was easier to read. Among the ChatGPT models, ChatGPT-3.5 exhibited a similar level of complexity to Google Bard, with scores of 12.9 and 35.6. ChatGPT-4, with scores of 12.4 and 37.1, was slightly less complex than both ChatGPT-3.5 and Google Bard but more complex than Microsoft Copilot. Overall, ChatGPT responses were best suited for readers with advanced reading skills, with ChatGPT-3.5 requiring graduate-level comprehension and ChatGPT-4 being more appropriate for advanced undergraduate-level readers (Table 1).

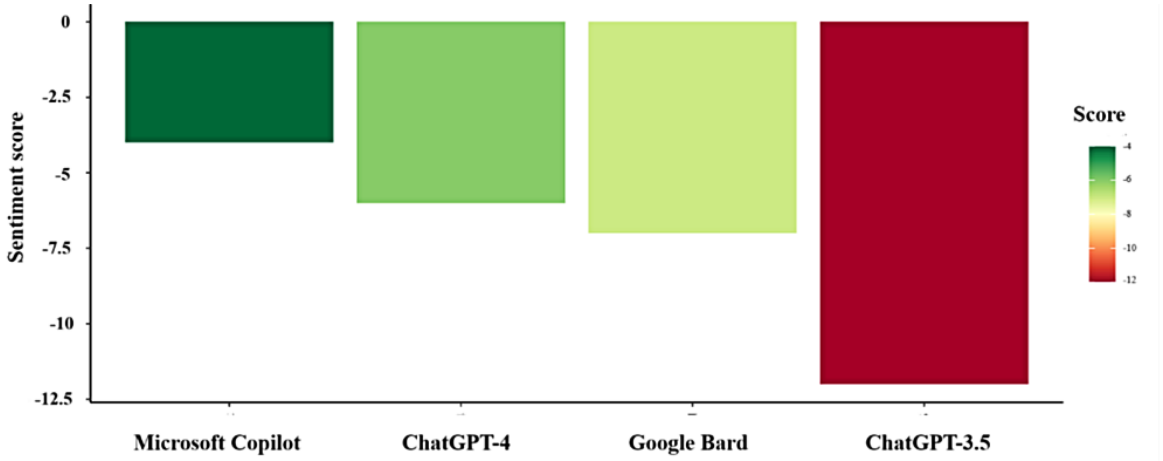
Table 1. Readability analysis for each large language model tested (Google Bard, Microsoft Copilot, ChatGPT-3.5, and ChatGPT-4).

AI Model	Flesch-Kincaid Grade Level	Flesch Reading Ease Score	Reading Complexity	Target Audience
Google Bard	12.9	35.8	Highly complex	Post-secondary education
Microsoft Copilot	9.9	49	Less complex	High school to early undergraduate level
ChatGPT-3.5	12.9	35.6	Similar complexity to Google Bard	Graduate-level comprehension
ChatGPT-4	12.4	37.1	Slightly less complex than ChatGPT-3.5 and Bard	Advanced undergraduate-level comprehension

LLMs' Sentiment Analysis

In the sentiment analysis, Microsoft Copilot had the least negative score (–4), reflecting a neutral and balanced tone. ChatGPT-4 scored (–6), slightly more negative due to its detailed discussion of risks. Google Bard followed with (–7), emphasizing practical recommendations that highlighted potential challenges. ChatGPT-3.5 had the most negative score (–12), reflecting a strong focus on risks and adverse outcomes. These differences, shown in Figure 1, highlight the tonal variations among the models, with Microsoft Copilot being the most neutral and ChatGPT-3.5 the most risk-oriented.

Figure 1. Sentiment analyses for Google Bard, Microsoft Copilot, ChatGPT-3.5, and ChatGPT-4.



Discussion

Overview of Major Study Findings

This study comprehensively assessed the knowledge, clarity, and objectivity of 4 prominent LLMs (ChatGPT-3.5, ChatGPT-4, Microsoft Copilot, and Google Bard) regarding COVID-19 impacts on pregnancy. Aligning with the stated aims, the findings revealed distinct performance variations among the LLMs. ChatGPT-4 and Microsoft Copilot demonstrated the highest levels of accuracy (97%, 32/33), followed by Google Bard (94%, 31/33) and ChatGPT-3.5 (82%, 27/33). While all models provided generally reliable information on COVID-19 and vaccination in maternal health, they varied in readability, with Microsoft Copilot being the most accessible and ChatGPT-3.5 the least comprehensible. Sentiment analysis highlighted Microsoft Copilot as the most neutral, whereas ChatGPT-3.5 displayed a risk-heavy tone. These outcomes underscored the evolving potential of LLMs to support health care professionals and public health communication during global crises while emphasizing the necessity for ongoing

XSL•FO
RenderX

evaluation and model refinement to align with the latest scientific evidence.

The Four LLMs Differ in Terms of Knowledge, With ChatGPT-3.5 Being the Least Knowledgeable

The performance of AI models such as ChatGPT-4, Google Bard, and Microsoft Copilot proved exceptional, with only minor errors detected. Their capacity to deliver accurate and detailed information showcased a depth of knowledge that is particularly impressive when juxtaposed with the level of COVID-19-related expertise observed among obstetrics and gynecology specialists. This comparison is drawn from findings in a previous survey [21], which revealed an average knowledge score of 75.6 (SD 10.6) among healthcare professionals in the field. Notably, while 81.1% (129/159) of these specialists displayed adequate knowledge of general COVID-19 information, a much smaller percentage (19/159, 11.9%) demonstrated specific understanding related to pregnancy, childbirth, and breastfeeding in the context of the pandemic. Furthermore, only 40.3% (64/159) of respondents showed familiarity with COVID-19 vaccination guidelines for pregnant individuals. A particularly critical area of knowledge, the increased risk of preeclampsia in pregnant women infected with COVID-19, was correctly identified by only 27% (43/159) of participants. This highlights a significant gap in awareness that could impact patient care. Regarding immunization practices, about 65% (104/159) of respondents recommended COVID-19 vaccination during all trimesters, reflecting a broad acceptance of its safety and efficacy across pregnancy stages. However, preferences for specific trimesters varied: about 25% (39/159) recommended vaccination during the second trimester, 6% (9/159) in the third trimester, and a minority of 4% (6/159) in the first trimester.

These findings align with existing literature, which consistently highlights a trend where ChatGPT-3 underperformed compared to other LLMs in specific tasks or benchmarks. Studies have noted that while ChatGPT-3 demonstrated proficiency in generating coherent text and answering general questions, it often fell short in accuracy, contextual understanding, and handling of complex queries. This performance gap has been attributed to differences in training data, model architecture, and parameter tuning [22,23]. Conversely, studies have shown that more advanced LLMs, such as ChatGPT-4 and Microsoft Copilot, outperformed ChatGPT-3, with Google Bard following closely behind [24], consistent with our findings.

However, other research suggests that ChatGPT-3.5 performed comparably or even better than other LLMs in certain contexts, potentially indicating that performance may vary depending on the specific application or field of use [25].

The Four LLMs Differ in Terms of Recommendations of COVID-19 Vaccination During Pregnancy, With ChatGPT-3.5 Being the Least Aligned With Scientific Evidence

When comparing the information provided about COVID-19 in pregnant women and the vaccination during pregnancy, there were some notable variations, especially in alignment with scientific evidence. Concerning the risk of severe illness, Google Bard highlighted an increased risk of severe complications, including hospitalization, intensive care unit admission, and death for pregnant women. In addition, it pointed out a higher likelihood of preterm birth. Microsoft Copilot emphasized that, while pregnant women are not more likely to contract the virus, they face a greater risk of severe disease, especially if infected in the third trimester. ChatGPT-3.5 noted only a slightly higher risk of severe illness for pregnant women compared to nonpregnant individuals, including increased hospitalization and intensive care unit admission rates, while, more correctly, ChatGPT-4 stated that pregnant women with COVID-19 have a higher risk of severe illness and complications, such as preterm birth. Regarding COVID-19 vaccination during pregnancy, Google Bard strongly advocated for COVID-19 immunization during pregnancy, highlighting its safety and effectiveness in preventing serious illness and protecting both mother and baby. Microsoft Copilot stressed the importance of vaccination for those women who are pregnant, breastfeeding, or planning pregnancy, emphasizing its safety and efficacy. ChatGPT-3.5 discussed the safety of Messenger Ribonucleic Acid vaccines for pregnant individuals and suggested discussing the timing of vaccination with health care providers, while ChatGPT-4 recommended COVID-19 vaccines for pregnant individuals, noting their safety and efficacy and also advised discussing vaccination timing with health care providers. In terms of vaccine safety and side effects, Google Bard mentioned that the side effects of vaccination in pregnant women are similar to those in nonpregnant adults and are generally mild and short-lived. Microsoft Copilot, ChatGPT-3.5, and ChatGPT-4 all emphasized the safety of Messenger Ribonucleic Acid COVID-19 vaccines during pregnancy and did not report any serious adverse maternal or fetal effects. Finally, Google Bard advised pregnant women to get vaccinated as soon as possible after the first trimester, whereas ChatGPT-4 suggested postvaccination monitoring for pregnant women through programs like Centers for Disease Control and Prevention's V-safe and emphasized the importance of consulting healthcare providers. All sources agreed on the increased risk of severe COVID-19 in pregnant women and the benefits and safety of vaccination during pregnancy. However, they differed in their emphasis on specific risks and considerations.

The comparison of LLMs in providing information about COVID-19 in pregnant women and vaccination during pregnancy underscores their potential to disseminate health information effectively but also highlights significant variability in alignment with scientific evidence. This variability raises concerns about the accuracy and reliability of AI-generated health content, particularly when subtle differences in emphasis or framing could influence public perception and decision-making. The findings illustrate how LLMs like ChatGPT, Microsoft Copilot, and Google Bard can either reinforce or undermine public health messaging, depending on their training data and underlying algorithms. These differences emphasize the need for rigorous evaluation and validation of AI-generated content in healthcare contexts, as inconsistencies could amplify the risk of misinformation, contributing to the emerging "AI-driven infodemic" [26]. Establishing robust mechanisms for quality control and promoting transparency in

XSL•FO
RenderX

AI outputs is imperative to mitigate these risks and ensure that LLMs support, rather than hinder, public health objectives.

The Four LLMs Differ in Terms of Vocabulary, Clarity, and Readability, With ChatGPT-3.5 Being the Most Difficult to Understand

From the text-mining analysis, overall, while there was a significant overlap in key themes across all four sources, each one brought its unique focus and depth to the discussion. The unique terms in each source highlighted their different approaches to discussing the same topic, reflecting the diversity in how information can be presented and prioritized. Google Bard and Microsoft Copilot tended to have a more general approach, while ChatGPT versions provided a more detailed, medically-oriented perspective. In particular, ChatGPT-3.5, being excessively technical, is particularly difficult to understand, as also reflected by the Flesch-Kincaid grade level analysis.

The findings from our text-mining analysis align closely with a comparative study [27], which showed that ChatGPT-3.5 tended to generate overly technical content that may hinder readability.

The differences in readability and complexity across LLMs are important when considering the target audience for the information [28-30]. For a general audience, a lower Flesch-Kincaid Grade Level might be more appropriate to ensure wider comprehension. On the other hand, for professional or academic audiences, a higher grade level could be more suitable, as it allows for the conveyance of complex ideas using specialized vocabulary and detailed explanations. Each LLM's approach reflects its unique language processing capabilities and potentially the nature of the training data it was exposed to. It is of crucial importance for users to consider the context and audience when evaluating or choosing the most appropriate AI tool for generating text, with ChatGPT's and Microsoft Copilot's approach appearing more beneficial for specialist or academic audiences and for the general public, respectively.

The Four LLMs Differ in Terms of Sentiment and Objectivity, With ChatGPT-3.5 Being the Least Factual

In the sentiment analysis, the tools demonstrated varying degrees of negativity in their language when discussing COVID-19 and pregnancy-related topics. Microsoft Copilot achieved the least negative score (–4), indicating a relatively neutral or balanced tone compared to its counterparts. This suggests that Microsoft Copilot's responses tended to minimize emotionally charged or negative language, focusing instead on providing concise, objective, and factual information. ChatGPT-4 followed with a slightly more negative score of –6, reflecting a balanced but somewhat more cautious tone. This could be attributed to ChatGPT-4's detailed discussion of potential risks and complications, which, while comprehensive, likely included language that leaned toward highlighting challenges and considerations in the context of COVID-19 and pregnancy. Google Bard scored –7, placing it closer to ChatGPT-4 in terms of sentiment but slightly more negative. This may be indicative of its focus on delivering actionable recommendations and clinical insights, which, while practical, may also involve language emphasizing adverse effects or potential risks, thereby increasing its overall negativity score. ChatGPT-3.5 obtained the most negative score of –12, representing a significantly more cautious or risk-focused tone compared to the other models. This heightened negativity could reflect the model's tendency to elaborate extensively on risks and adverse outcomes, potentially overemphasizing negative aspects in its explanations.

These differences in sentiment scores highlight the varying tonal approaches of the AI models. Microsoft Copilot's minimal negativity suggests a focus on general, less emotionally charged language, while ChatGPT-3.5's highest negativity score underscores its more risk-heavy narrative. ChatGPT-4 and Google Bard occupy the middle ground, balancing detailed explanations with practical recommendations, albeit with some cautionary emphasis. This variation in sentiment reflects not only the linguistic choices of the models but also their inherent design priorities, ranging from objectivity and balance to detailed elaboration and actionable advice.

Strengths and Limitations

This study has several notable strengths. First, the use of a validated questionnaire ensures a robust and reliable assessment of LLMs' proficiency, clarity, and objectivity. The evaluation spans multiple dimensions, including knowledge, sentiment, and readability, offering a comprehensive analysis of the capabilities of ChatGPT-3.5, ChatGPT-4, Google Bard, and Microsoft Copilot. In addition, the comparison against specialist knowledge provides a meaningful benchmark, grounding the findings in real-world relevance. The inclusion of sentiment and readability analyses highlights the accessibility and emotional tone of AI-generated responses, further contextualizing their suitability for different audiences. Moreover, the study addresses a critical gap in assessing AI tools in the context of obstetrics, gynecology, and public health during a pandemic—a time when accurate information is essential. By incorporating text-mining and readability metrics, the research elucidates how LLMs can serve diverse user bases, ranging from healthcare professionals to general audiences. However, this study is not without its limitations. A key limitation is the static nature of the evaluation, as LLMs undergo continuous updates and learning processes. This dynamic evolution may render some of the current observations obsolete over time, necessitating periodic re-evaluation. Furthermore, the study focuses on a specific application of LLMs (COVID-19's impacts in pregnancy), which may not generalize to other medical or public health domains. The zero-shot prompting approach, while simulating real-world user queries, may not fully capture the potential of these models when used with tailored or fine-tuned inputs. Finally, the study is limited by its reliance on the English language and the specific design of the validated questionnaire. Future research could explore the multilingual capabilities of LLMs and their application to other health care domains, as well as integrate expert feedback during iterative model evaluation.

Future Directions

The present comprehensive analysis covering various aspects (knowledge, recommendations, vocabulary and readability,

XSL•FO
RenderX

sentiment, and objectivity) shows similarities and dissimilarities among four major LLMs.

The knowledge appraisal highlights a gap in awareness of COVID-19 and immunization during pregnancy among specialists, underscoring the potential role of AI like ChatGPT in supplementing healthcare knowledge and decision-making. Generative AI holds great potential in health care and global public health, serving as valuable tools for professionals, providing quick access to updated information, and filling knowledge gaps.

However, some discrepancies in ChatGPT-3.5's responses also serve as a reminder of the need for continuous updating and verification of AI knowledge bases. This ensures that the information provided is aligned with the latest research and clinical guidelines. As AI systems become more integrated into health care, their ability to adapt and learn from emerging data will be crucial in ensuring they remain reliable and beneficial adjuncts in medical practice [31,32].

Future research directions should consider the inclusion of emerging AI models to evaluate their comparative utility in clinical settings. Investigating the effectiveness of these tools across various medical specialties could shed light on their strengths and limitations. Studies assessing the real-world integration of LLMs into health care workflows, focusing on usability, clinician satisfaction, and patient outcomes, are also needed. In addition, exploring the ethical and legal implications of using AI in clinical practice (such as ensuring patient privacy, safeguarding data security, and clarifying liability in decision-making) remains a crucial area of inquiry. Longitudinal studies on the influence of AI tools on medical education, professional development, health care delivery, and communication would provide valuable insights into their sustained impact.

Overall, the analysis underscores the importance of choosing the right LLM for specific needs. While ChatGPT-4 and newer versions offer more detailed and updated information, earlier versions like ChatGPT-3.5 might present challenges in terms of comprehensibility. The choice between these LLMs should be influenced by the target audience and the level of detail and technicality required in the information. For general audiences, simpler language (as seen in Microsoft Copilot) is preferable, whereas more specialized audiences might benefit from the detailed approach of ChatGPT-4. For instance, in vaccine management, ChatGPT-4 can support specialized audiences by aiding clinical decision-making, providing detailed research insights, and tailoring vaccination guidelines for pregnant women. Microsoft Copilot excels in communicating simplified information to general audiences, making it ideal for educating patients, training health workers, or addressing vaccine hesitancy. Google Bard, with its real-time search integration, is suited for disseminating up-to-date policies and driving public health campaigns with clear, actionable messaging. ChatGPT-3.5, despite its limitations, can be used for simulating scenarios or gathering feedback to refine communication strategies. These applications demonstrate how aligning LLMs' capabilities with target audiences ensures effective vaccine management and can be potentially translated to other clinical settings and patient management contexts.

Conclusions

This study highlighted the strengths and limitations of LLMs such as ChatGPT-3.5, ChatGPT-4, Microsoft Copilot, and Google Bard in addressing critical healthcare topics, specifically the impacts of COVID-19 on pregnancy. While the findings underscored the potential of LLMs to provide accurate, up-to-date information, they also revealed gaps in their knowledge and variability in their communication styles, sentiment, and readability. Such variability underscores the importance of tailoring AI use to specific audiences and contexts. Beyond the immediate findings, the broader implications of this research emphasize the transformative potential of LLMs in health care and public health communication. These AI tools can bridge knowledge gaps among health care professionals, support evidence-based decision-making, and combat misinformation during health crises. Their ability to deliver clear, accessible, and accurate information positions them as valuable adjuncts in both clinical settings and public health campaigns. However, the study also underscores critical ethical and operational considerations. Ensuring the reliability of AI-generated content requires continuous updates to training datasets and rigorous validation against the latest scientific evidence. The dynamic nature of AI development means that its integration into health care systems must be accompanied by robust oversight mechanisms to mitigate risks such as the dissemination of misinformation or unintentional biases. Furthermore, the deployment of LLMs in health care highlights the need for interdisciplinary collaboration among AI developers, health care professionals, and policy makers. Such partnerships can ensure that these tools are designed and utilized in ways that prioritize patient safety, data privacy, and equitable access to information. By fostering these collaborations, the integration of AI into health care can move beyond theoretical potential to practical, impactful applications. In conclusion, while LLMs like ChatGPT-4 and Microsoft Copilot demonstrate remarkable capabilities in providing health care–related information, their limitations remind us of the essential role of human oversight. The findings of this study advocate for a future where AI complements, rather than replaces, human expertise in healthcare, enhancing both the quality and accessibility of medical knowledge. This balanced approach will be key to leveraging AI's full potential while safeguarding the ethical and practical dimensions of health care delivery.

Data Availability

All data are available in Multimedia Appendices 1-3.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of queries and prompts used with Google Bard, Microsoft Copilot, ChatGPT-3.5, and ChatGPT-4.
[DOCX File , 16 KB - formative_v9i1e56126_app1.docx]

Multimedia Appendix 2
Responses provided by each large language model.
[DOCX File , 22 KB - formative_v9i1e56126_app2.docx]

Multimedia Appendix 3
Responses generated by each large language model (Google Bard, Microsoft Copilot, ChatGPT-3.5, and ChatGPT-4) when prompted to provide recommendations on COVID-19 vaccination for pregnant women.
[DOCX File , 16 KB - formative_v9i1e56126_app3.docx]

References

1. Filip R, Gheorghita Puscaselu R, Anchidin-Norocel L, Dimian M, Savage WK. Global challenges to public health care systems during the COVID-19 pandemic: a review of pandemic measures and problems. J Pers Med 2022;12(8):1295 [FREE Full text] [doi: 10.3390/jpm12081295] [Medline: 36013244]
2. Forati AM, Ghose R. Geospatial analysis of misinformation in COVID-19 related tweets. Appl Geogr 2021;133:102473 [FREE Full text] [doi: 10.1016/j.apgeog.2021.102473] [Medline: 34103772]
3. Marquini GV, Martins SB, Oliveira LM, Dias MM, Takano CC, Sartori MGF. Effects of the COVID-19 pandemic on gynecological health: an integrative review. Rev Bras Ginecol Obstet 2022;44(2):194-200 [FREE Full text] [doi: 10.1055/s-0042-1742294] [Medline: 35213918]
4. Alberca RW, Pereira NZ, Oliveira LMDS, Gozzi-Silva SC, Sato MN. Pregnancy, viral infection, and COVID-19. Front Immunol 2020;11:1672 [FREE Full text] [doi: 10.3389/fimmu.2020.01672] [Medline: 32733490]
5. Hanna N, Hanna M, Sharma S. Is pregnancy an immunological contributor to severe or controlled COVID-19 disease? Am J Reprod Immunol 2020;84(5):e13317 [FREE Full text] [doi: 10.1111/aji.13317] [Medline: 32757366]
6. Blakeway H, Prasad S, Kalafat E, Heath PT, Ladhani SN, Le Doare K, et al. COVID-19 vaccination during pregnancy: coverage and safety. Am J Obstet Gynecol 2022;226(2):236.e1-236.e14 [FREE Full text] [doi: 10.1016/j.ajog.2021.08.007] [Medline: 34389291]
7. Riggan KA, Reckhow J, Allyse MA, Long M, Torbenson V, Rivera-Chiauzzi EY. Impact of the COVID-19 pandemic on obstetricians/gynecologists. Mayo Clin Proc Innov Qual Outcomes 2021;5(6):1128-1137 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2021.11.002] [Medline: 34849457]
8. Khamisy-Farah R, Furstenau LB, Kong JD, Wu J, Bragazzi NL. Gynecology meets big data in the disruptive innovation medical era: state-of-art and future prospects. Int J Environ Res Public Health 2021;18(10):5058 [FREE Full text] [doi: 10.3390/ijerph18105058] [Medline: 34064710]
9. Bragazzi NL, Dai H, Damiani G, Behzadifar M, Martini M, Wu J. How big data and artificial intelligence can help better manage the COVID-19 pandemic. Int J Environ Res Public Health 2020;17(9):3176 [FREE Full text] [doi: 10.3390/ijerph17093176] [Medline: 32370204]
10. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature 2023;620(7972):172-180 [FREE Full text] [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
11. Joshi S, Ha E, Amaya A, Mendoza M, Rivera Y, Singh VK. Ensuring accuracy and equity in vaccination information from ChatGPT and CDC: mixed-methods cross-language evaluation. JMIR Form Res 2024;8:e60939 [FREE Full text] [doi: 10.2196/60939] [Medline: 39476380]
12. Fiore M, Bianconi A, Acuti Martellucci C, Rosso A, Zauli E, Flacco ME, et al. Vaccination hesitancy: agreement between WHO and ChatGPT-4.0 or Gemini Advanced. Ann Ig 2024 [FREE Full text] [doi: 10.7416/ai.2024.2657] [Medline: 39373234]
13. Kim S, Kim K, Wonjeong Jo C. Accuracy of a large language model in distinguishing anti- and pro-vaccination messages on social media: the case of human papillomavirus vaccination. Prev Med Rep 2024;42:102723 [FREE Full text] [doi: 10.1016/j.pmedr.2024.102723] [Medline: 38659997]
14. Deiana G, Dettori M, Arghittu A, Azara A, Gabutti G, Castiglia P. Artificial intelligence and public health: evaluating ChatGPT responses to vaccination myths and misconceptions. Vaccines (Basel) 2023;11(7):1217 [FREE Full text] [doi: 10.3390/vaccines11071217] [Medline: 37515033]
15. Sohail SS, Madsen D, Farhat F, Alam MA. ChatGPT and vaccines: can AI chatbots boost awareness and uptake? Ann Biomed Eng 2024;52(3):446-450. [doi: 10.1007/s10439-023-03305-y] [Medline: 37428336]

XSL•FO
RenderX

16. Sallam M, Salim NA, Al-Tammemi AB, Barakat M, Fayyad D, Hallit S, et al. ChatGPT output regarding compulsory vaccination and COVID-19 vaccine conspiracy: a descriptive study at the outset of a paradigm shift in online search for information. Cureus 2023;15(2):e35029 [FREE Full text] [doi: 10.7759/cureus.35029] [Medline: 36819954]
17. Tekinay ON. Curious questions about Covid-19 pandemic with ChatGPT: answers and recommendations. Ann Biomed Eng 2023;51(7):1371-1373 [FREE Full text] [doi: 10.1007/s10439-023-03209-x] [Medline: 37071282]
18. Cheng K, Li Z, He Y, Guo Q, Lu Y, Gu S, et al. Potential use of artificial intelligence in infectious disease: take ChatGPT as an example. Ann Biomed Eng 2023;51(6):1130-1135 [FREE Full text] [doi: 10.1007/s10439-023-03203-3] [Medline: 37074486]
19. Alhasan K, Al-Tawfiq J, Aljamaan F, Jamal A, Al-Eyadhy A, Temsah M. Mitigating the burden of severe pediatric respiratory viruses in the post-COVID-19 era: ChatGPT insights and recommendations. Cureus 2023;15(3):e36263 [FREE Full text] [doi: 10.7759/cureus.36263] [Medline: 37073200]
20. Cheng K, Wu C, Gu S, Lu Y, Wu H, Li C. WHO declares the end of the COVID-19 global health emergency: lessons and recommendations from the perspective of ChatGPT/GPT-4. Int J Surg 2023;109(9):2859-2862 [FREE Full text] [doi: 10.1097/JS9.0000000000000521] [Medline: 37246993]
21. Bragazzi NL, Buchinger M, Szarpak L, Chmielewski J, Goździewska M, Gotlib J, et al. Knowledge, attitudes, and practices towards COVID-19 and COVID-19 vaccines among obstetrician-gynaecologists. Ann Agric Environ Med 2023;30(4):669-676 [FREE Full text] [doi: 10.26444/aaem/176959] [Medline: 38153070]
22. Zare S, Vafaeian S, Amini M, Farhadi K, Vali M, Golestani A. Comparing the performance of ChatGPT-3.5-Turbo, ChatGPT-4, and Google Bard with Iranian students in pre-internship comprehensive exams. Sci Rep 2024;14(1):28456 [FREE Full text] [doi: 10.1038/s41598-024-79335-w] [Medline: 39557926]
23. Gill GS, Blair J, Litinsky S. Evaluating the performance of ChatGPT 3.5 and 4.0 on statpearls oculoplastic surgery text- and image-based exam questions. Cureus 2024;16(11):e73812. [doi: 10.7759/cureus.73812] [Medline: 39691123]
24. Rossettini G, Rodeghiero L, Corradi F, Cook C, Pillastrini P, Turolla A, et al. Comparative accuracy of ChatGPT-4, microsoft copilot and Google Gemini in the Italian entrance test for healthcare sciences degrees: a cross-sectional study. BMC Med Educ 2024;24(1):694 [FREE Full text] [doi: 10.1186/s12909-024-05630-9] [Medline: 38926809]
25. Ponzo V, Rosato R, Scigliano MC, Onida M, Cossai S, De Vecchi M, et al. Comparison of the accuracy, completeness, reproducibility, and consistency of different AI chatbots in providing nutritional advice: an exploratory study. J Clin Med 2024;13(24):7810 [FREE Full text] [doi: 10.3390/jcm13247810] [Medline: 39768733]
26. De Angelis L, Baglivo F, Arzilli G, Privitera GP, Ferragina P, Tozzi AE, et al. ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. Front Public Health 2023;11:1166120 [FREE Full text] [doi: 10.3389/fpubh.2023.1166120] [Medline: 37181697]
27. Marulli F, Campanile L, de Biase MS, Marrone S, Verde L, Bifulco M. Understanding readability of large language models output: an empirical analysis. Procedia Comput. Sci 2024;246:5273-5282. [doi: 10.1016/j.procs.2024.09.636]
28. Amin K, Doshi R, Forman HP. Large language models as a source of health information: are they patient-centered? A longitudinal analysis. Healthc (Amst) 2024;12(1):100731. [doi: 10.1016/j.hjdsi.2023.100731] [Medline: 38141269]
29. Tu W, Joe BN. The Era of ChatGPT and large language models: can we advance patient-centered communications appropriately and safely? Radiol Imaging Cancer 2024;6(3):e240038 [FREE Full text] [doi: 10.1148/rycan.240038] [Medline: 38668641]
30. Howell MD. Generative artificial intelligence, patient safety and healthcare quality: a review. BMJ Qual Saf 2024;33(11):748-754 [FREE Full text] [doi: 10.1136/bmjqs-2023-016690] [Medline: 39048292]
31. Lin Z, Guan S, Zhang W, Zhang H, Li Y, Zhang H. Towards trustworthy LLMs: a review on debiasing and dehallucinating in large language models. Artif Intell Rev 2024;57(9):243. [doi: 10.1007/s10462-024-10896-y]
32. Bragazzi NL, Garbarino S. Toward clinical generative AI: conceptual framework. JMIR AI 2024;3:e55957 [FREE Full text] [doi: 10.2196/55957] [Medline: 38875592]

Abbreviations

AI: artificial intelligence
ART: assisted reproductive technology
LLM: large language model
TDM: Term-Document Matrix
WHO: World Health Organization

XSL•FO
RenderX

Edited by A Mavragani; submitted 07.01.24; peer-reviewed by M Dagli, W Han, D Valdes; comments to author 16.05.24; revised version received 28.12.24; accepted 09.01.25; published 05.02.25.

Please cite as:
Bragazzi NL, Buchinger M, Atwan H, Tuma R, Chirico F, Szarpak L, Farah R, Khamisy-Farah R
Proficiency, Clarity, and Objectivity of Large Language Models Versus Specialists' Knowledge on COVID-19's Impacts in Pregnancy: Cross-Sectional Pilot Study
JMIR Form Res 2025;9:e56126
URL: https://formative.jmir.org/2025/1/e56126
doi:10.2196/56126
PMID:

©Nicola Luigi Bragazzi, Michèle Buchinger, Hisham Atwan, Ruba Tuma, Francesco Chirico, Lukasz Szarpak, Raymond Farah, Rola Khamisy-Farah. Originally published in JMIR Formative Research (https://formative.jmir.org), 05.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A 4-Site Public Deliberation Project on the Acceptability of Youth Self-Consent in Biomedical HIV Prevention Trials: Assessment of Facilitator Fidelity to Key Principles

Claire Burke Draucker[1], MSN, PhD; Andrés Carrión[1], MS; Mary A Ott[1,2], MA, MD; Ariel I Hicks[1], MPH; Amelia Knopf[1], MPH, PhD

[1]Indiana University, Indianapolis, IN, United States

[2]Icahn School of Medicine at Mount Sinai, New York, NY, United States

Corresponding Author:
Claire Burke Draucker, MSN, PhD
Indiana University
111 Middle Drive
Indianapolis, IN, 46202
United States
Phone: 1 317 274 4139
Email: cdraucke@iu.edu

Abstract

Background: Public deliberation is an approach used to engage persons with diverse perspectives in discussions and decision-making about issues affecting the public that are controversial or value laden. Because experts have identified the need to evaluate facilitator performance, our research team developed a framework to assess the fidelity of facilitator remarks to key principles of public deliberation.

Objective: This report describes how the framework was used to assess facilitator fidelity in a 4-site public deliberation project on the acceptability of minor self-consent in biomedical HIV prevention research.

Methods: A total of 88 individuals participated in 4 deliberation sessions held in 4 cities throughout the United States. The sessions, facilitated by 18 team members, were recorded and transcribed verbatim. Facilitator remarks were highlighted, and predetermined coding rules were used to code the remarks to 1 of 6 principles of quality deliberations. A variety of display tables were used to organize the codes and calculate the number of facilitator remarks that were consistent or inconsistent with each principle during each session across all sites. A content analysis was conducted on the remarks to describe how facilitator remarks aligned or failed to align with each principle.

Results: In total, 735 remarks were coded to one of the principles; 516 (70.2%) were coded as consistent with a principle, and 219 (29.8%) were coded as inconsistent. A total of 185 remarks were coded to the principle of equal participation (n=138, 74.6% as consistent; n=185, 25.4% as inconsistent), 158 were coded to expression of diverse opinions (n=110, 69.6% as consistent; n=48, 30.4% as inconsistent), 27 were coded to respect for others (n=27, 100% as consistent), 24 were coded to adoption of a societal perspective (n=11, 46% as consistent; n=13, 54% as inconsistent), 99 were coded to reasoned justification of ideas (n=81, 82% as consistent; n=18, 18% as inconsistent), and 242 were coded to compromise or movement toward consensus (n=149, 61.6% as consistent; n=93, 38.4% as inconsistent). Therefore, the counts provided affirmation that most of the facilitator remarks were aligned with the principles of deliberation, suggesting good facilitator fidelity. By considering how the remarks aligned or failed to align with the principles, areas where facilitator fidelity can be strengthened were identified. The results indicated that facilitators should focus more on encouraging quieter members to participate, refraining from expressing personal opinions, promoting the adoption of a societal perspective and reasoned justification of opinions, and inviting deliberants to articulate their areas of common ground.

Conclusions: The results provide an example of how a framework for assessing facilitator fidelity was used in a 4-site deliberation project. The framework will be refined to better address issues related to balancing personal and public perspectives, managing plurality, and mitigating social inequalities.

(JMIR Form Res 2025;9:e58451) doi:10.2196/58451

XSL•FO
RenderX

KEYWORDS

public deliberation; deliberative democracy; bioethics; ethical conflict; biomedical; HIV prevention; HIV research; group facilitation; fidelity assessment; content analysis

Introduction

Background

Public deliberations are community-focused activities in which persons with diverse perspectives are engaged in discussion and decision-making about a topic that is controversial or value laden or involves ethical uncertainty. Public deliberations typically include provision of factual and balanced information about the deliberation topic, facilitated small- and large-group discussions, and activities that yield collective solutions to policy questions [1-3].

Public deliberation has been used to explore a wide variety of issues that affect public health. The method has been used to elicit views on genetic research and testing in Alaska Native and American Indian communities [4], attitudes and recommendations to address antibiotic overuse [5], opinions on acceptable uses of deidentified health information in oncology settings [6], views on the allocation of limited health care resources [7], opinions on guidelines for the design and evaluation of decision aids [8], recommendations for opioid screening in the context of HIV care [2], preferences regarding health data sharing for patients with cancer [9], and views on precision medicine research in American Indian and Alaska Native communities [10].

With the increasing use of public deliberations to inform public health policy decisions, experts have called for increased attention to assessing and reporting the quality of deliberations [11,12]. To aid in this task, Scott et al [11] developed a deductive coding framework based on the operationalization of common deliberation goals and a synthesis of published frameworks designed to assess deliberation quality. Their framework includes 4 deliberation elements (ie, preferences and values are referenced, deliberants engage with each other, expert information is referenced, and information enriches deliberation) and 4 recommendation elements (ie, recommendations are clear and identifiable, recommendations are relevant, recommendations are justified, and recommendations reflect a societal perspective). Goold et al [12] also developed an analytic framework for the assessment of deliberation quality. This framework calls for an evaluation of structure (eg, demographic representativeness of the deliberants), process (eg, equality of participation), and outcomes (eg, deliberant knowledge of deliberation topics) using a variety of data sources and analytic approaches (eg, surveys, follow-up interviews, qualitative analysis of transcripts, word counts, and participation metrics).

Most evaluation frameworks do not provide for an assessment of the performance of facilitators. Deliberations typically rely on skilled facilitators who are responsible for promoting deliberative norms, fostering high-quality discussion, and promoting an outcome of shared solutions. Rubinelli and von Groote [13] outline common facilitator tasks, including conducting the dialogue, soliciting different opinions, ensuring that reasons are provided for points of view, and ensuring that reasons are grounded in evidence. The authors also discuss facilitator challenges such as identifying hidden goals of deliberants and managing sharp differences of opinion. Dillard [14] argues that facilitators are “the first and primary form of intervention” and provides an in-depth discussion of facilitator responsibilities. These responsibilities can include setting the discussion tone, establishing deliberative rules and norms, mitigating adversarial and polarizing dynamics, and ensuring that minority opinions are expressed. Moreover, the author identifies 3 facilitator types based on variations in their actions: passive facilitators (ie, they regulate deliberant speech time and direct participant turn taking), involved facilitators (ie, they ask elicitation questions and implement discursive strategies), and moderate facilitators (ie, they use elicitation questions and discursive strategies but less often than involved facilitators). Discursive strategies are facilitation techniques used to manage tensions and encourage fruitful deliberative discourse. These strategies include storytelling (ie, sharing personal experiences), playing devil’s advocate, framing (ie, highlighting interactions to shape discussion), focusing on local issues to consider specific implications, and eliciting responses.

Objectives

Despite the assumed importance of facilitators in ensuring effective deliberative processes, published reports of deliberation outcomes rarely include information regarding facilitator credentials, selection, training, and supervision and, with few exceptions (see the study by De Vries et al [15]), do not address the quality of facilitator involvement. Therefore, a framework is needed to guide researchers in systematically assessing public deliberation facilitator remarks. To address this gap, our research team developed such a framework. This framework is based on the assumption that the performance of facilitators can be assessed by determining the fidelity of their remarks to well-established principles of democratic deliberation. A blueprint to guide researchers in implementing the framework has been published in *JMIR Formative Research* [16] and includes several coding templates and a step-by-step analytic plan. The purpose of this report is to describe how the framework (hereafter referred to as the blueprint report) was used to assess facilitator fidelity in our 4-site public deliberation on the acceptability of youth self-consent in biomedical HIV prevention research. In this report, we (1) briefly describe the deliberation, (2) summarize the blueprint report, and (3) present our assessment of facilitator fidelity across the 4 sites.

Methods

Public Deliberation

The deliberation research project from which this work stemmed is described in depth elsewhere [17,18] and summarized in the blueprint report [16]. Therefore, we include only a brief recap of the main aspects of the deliberation project in this section.

XSL•FO
RenderX

The public health problem that was the focus of the deliberation was the underrepresentation of minors in biomedical HIV prevention trials due in part to barriers caused by standard regulatory requirements for parental consent and youth assent [19,20]. This consent model can result in parents becoming aware of some stigmatized behaviors or identities that their children may not wish to disclose and, thus, can lead to rejection or punishment of the youth or their exclusion from research [21]. Because youth from underrepresented populations assume a disproportionate burden of HIV disease and are the most likely to be harmed by existing consent processes, calls have been made for consideration of new consent procedures, including minor self-consent [22]. However, because parental consent has long been considered foundational to protecting youth from research-related harms, developing new consent models constitutes a public health issue that is both controversial and value laden. The purpose of our deliberation was to engage diverse stakeholders, including minor adolescents, parents, caregivers, and community members, in deliberations about the use of minor self-consent in biomedical HIV prevention research.

Ethical Considerations

This study was reviewed and approved by the Indiana University Institutional Review Board (IRB; protocol 2001909245), which served as the single IRB in accordance with reliance agreements from participating institutions, including the University of South Florida, Johns Hopkins University, the University of Colorado Denver, and the University of Chicago. An electronic copy of the IRB-approved study information sheet was provided to deliberants by study staff or site investigators. In web-based or face-to-face meetings, staff reviewed study information with interested deliberants and obtained verbal consent from those who agreed to participate. Although a waiver of parental permission was granted for this study, minors were asked whether they felt comfortable seeking parental permission to enroll in the study. Those who wished to engage a parent helped staff schedule a time to meet with both the minor and their parent or guardian to review the study information sheet and obtain parental permission. If a minor did not wish to engage a parent, the study staff contacted a site principal investigator, who determined whether the waiver of parental permission could be applied.

All participants were assigned a study ID number, and their files were labeled with that ID number. All data were stored on a dually authenticated, password-protected, cloud-based storage system. The servers are tightly controlled and inaccessible to anyone without appropriate credentials. Once all analyses are complete, audio and video files will be permanently deleted, and only deidentified data will be stored.

Participants received compensation for this study, which varied by activity and retention in the study. The maximum compensation for participants who completed all surveys and homework and attended all deliberative sessions was US $300 in the form of gift cards from a variety of web-based and physical stores; participants selected the retailer that worked best for their needs.

Deliberation Sessions

Overview

The deliberations were held via a web-based video platform in 4 cities: Tampa, Florida; Baltimore, Maryland; Denver, Colorado; and Chicago, Illinois. Educational materials related to the deliberation topic were developed in consultation with a youth advisory board and external advisers and displayed on a deliberant website. At each site, 4 deliberation sessions, each lasting 2 hours, were held. Each session included large-group meetings (plenary sessions) and small breakout group sessions. During the plenary sessions, research team members and expert stakeholders (eg, physicians and IRB administrators) presented information on topics pertinent to the deliberation (eg, research with adolescents, regulatory guidelines for biomedical research, and clinical trial principles). All deliberation sessions were video recorded and transcribed. Due to recording malfunctions in Baltimore, transcripts were not available for some of these sessions.

Deliberants

There were 88 deliberants in the deliberations. To provide context for our discussion on facilitator fidelity, a summary of the deliberants' demographic characteristics is presented in Table 1.

XSL•FO
RenderX

Table 1. Summary of deliberants' demographic characteristics (N=88).

Characteristics	Deliberants, n (%)
Age	
Minor	24 (27)
Young adult	20 (23)
Adult	44 (50)
Gender	
Women, including transgender women and transfeminine	62 (70)
Genderqueer or gender nonconforming	6 (7)
Men, including transgender men and transmasculine	17 (19)
Other	3 (3)
Sexual orientation	
Lesbian, bisexual, gay, or queer	26 (30)
Heterosexual or straight	53 (60)
Questioning	5 (6)
Other[a]	4 (5)
Educational level (n=60)[b]	
High school diploma or GED[c]	8 (13)
Some college or technical program	12 (20)
Finished vocational or technical program or 2-year degree	5 (8)
Bachelor's degree or higher	35 (58)
Racial identity	
American Indian or Alaska Native	1 (1)
Asian	6 (7)
Black or African American	39 (44)
Multiracial	6 (7)
White	28 (32)
Other[d]	8 (9)
Ethnic identity	
Hispanic or Latinx	19 (22)

[a]Other sexual orientations as entered by the respondents: pansexual (2/88, 2%), asexual (1/88, 1%), and other (1/88, 1%).

[b]Among deliberants aged >18 years.

[c]GED: General Educational Development.

[d]Other racial identities as entered by the respondents: Hispanic (mixed; 1/88, 1%), Afrolatina (1/88, 1%), Mestizo (1/88, 1%), "none of these races apply to me as I am Hispanic" (1/88, 1%), West Asian (1/88, 1%), and other (1/88, 1%).

Facilitators

A total of 18 persons facilitated the deliberations across the 4 sites. The number of facilitators at each site varied according to the size of the deliberant group. Of the 18 facilitators, 8 (44%) were graduate students, 4 (22%) were research staff members, and 6 (33%) were faculty. Their racial and ethnic identities included non-Hispanic White, non-Hispanic Black, Asian, Hispanic, and multiracial; 61% (11/18) were women. The research team and content experts had expertise in HIV prevention, adolescent health, research regulation, bioethics, and public deliberation. The site stakeholders included health care professionals and community advocates. Because the roles that the team members played in the deliberation sessions were fluid (eg, public deliberation experts lead plenary sessions and facilitated breakout group sessions, and scientists from the research team and site stakeholders presented expert information and helped facilitate sessions), for the purpose of this assessment, all team members were considered facilitators.

The facilitators had different training experiences. Of the graduate student facilitators, 75% (6/8) were employed and trained by the deliberation experts, and 25% (2/8) were employed and trained by the research team. Content experts did

XSL•FO
RenderX

not receive training in deliberative methods. A booster training session was provided between deliberations at sites 3 and 4; the session focused primarily on methods to support deliberative principles and goals but also included basic group facilitation skills.

Blueprint for Implementing a Framework for Assessing Facilitator Fidelity

The blueprint report [16] directs researchers to highlight pertinent facilitator remarks on verbatim transcripts of deliberation sessions and rate them as consistent or inconsistent with any of the 6 principles of public deliberations in the framework. This approach is based on the assumption that facilitator remarks consistent with one of the principles encourage deliberant discussions that reflect deliberation ideals, whereas facilitator remarks inconsistent with one of the principles can inhibit such discussions. The principles in the framework are equal participation [23], expression of diverse opinions [24], respect for others [23], adoption of a societal perspective [23], reasoned justification of ideas [23], and compromise or movement toward consensus [24]. We developed a coding rule table that displays the 6 principles, coding rules for each principle, and examples of text coded to each principle. This table can be accessed in the blueprint report [16].

The blueprint report also includes 2 tables to organize codes for analysis: a code display table and a code summary table. In the code display table, transcript line numbers that were coded to the principles are placed in cells defined by the principle, the role of the facilitator who made the remark, and the deliberation session in which the remark was made. In the code summary table, the number of remarks coded as consistent or inconsistent with each principle across all sessions is summarized and displayed. An example of the code display and code summary tables for site 1 (Tampa) can be found in the blueprint report [16].

The blueprint report also provides a step-by-step analytic plan to assess facilitator fidelity at each site [16]. In summary, 2 raters use the coding rule table to independently highlight facilitator remarks on deliberation transcripts that reflect 1 of the 6 principles and indicate whether each remark was consistent or inconsistent with the principle. They compare their coding and resolve discrepancies through discussion and review of transcript data. The transcript line numbers of the agreed-upon codes are placed in the appropriate cell in a code display table. Information in the code display tables for each site is tabulated to then complete a code summary table.

Results

Overview

Following the blueprint report [16] guidelines discussed previously, we coded all the transcripts and developed code display and code summary tables for each site. For the project described in this paper, we also developed 2 additional project summary tables to display counts across multiple sites. The first project summary table (Table 2) shows the number of facilitator remarks coded as consistent or inconsistent with the principles by site. The second project summary table (Table 3) shows the number of facilitator remarks coded as consistent or inconsistent with the principles by principle.

Facilitator fidelity to each principle across all sessions and sites can be described based on the project summary tables. For each principle, the total number of remarks coded to that principle, the number of remarks coded as consistent or inconsistent with the principle, and the number of remarks coded to the principle for each site are presented in the following sections. The types of remarks coded to the principles are then discussed, and verbatim examples of facilitator remarks and facilitator and deliberant dialogues are presented. The 3 principles related to facilitation of the discussion (ie, equal participation, expression of diverse opinions, and respect for others) are discussed first followed by the 3 principles related to facilitation of recommendations (ie, adoption of a societal perspective, reasoned justification of ideas, and compromise or movement toward consensus).

XSL•FO
RenderX

Table 2. First project summary table—number of facilitator remarks coded as consistent or inconsistent with principles of democratic deliberation by site.

Site and principle	Consistent, n (%)	Inconsistent, n (%)
Tampa, Florida		
Equal participation (n=11)	9 (81.8)	2 (18.1)
Expression of diverse opinions (n=41)	36 (87.8)	5 (12.2)
Respect for others (n=8)	8 (100)	0 (0)
Adoption of a societal perspective (n=6)	4 (66.7)	2 (33.3)
Reasoned justification of ideas (n=7)	7 (100)	0 (0)
Compromise or movement toward consensus (n=38)	29 (76.3)	9 (23.7)
Total (n=111)	93 (83.8)	18 (16.2)
Baltimore, Maryland		
Equal participation (n=6)	6 (100)	0 (0)
Expression of diverse opinions (n=7)	6 (85.7)	1 (14.3)
Respect for others (n=2)	2 (100)	0 (0)
Adoption of a societal perspective (n=0)	0 (0)	0 (0)
Reasoned justification of ideas (n=3)	3 (100)	0 (0)
Compromise or movement toward consensus (n=21)	18 (85.7)	3 (14.3)
Total (n=39)	35 (89.7)	4 (10.3)
Denver, Colorado		
Equal participation (n=94)	59 (62.8)	35 (37.2)
Expression of diverse opinions (n=41)	24 (58.5)	17 (41.5)
Respect for others (n=9)	9 (100)	0 (0)
Adoption of a societal perspective (n=14)	6 (42.9)	8 (57.1)
Reasoned justification of ideas (n=29)	27 (93.1)	2 (6.9)
Compromise or movement toward consensus (n=92)	58 (63)	34 (37)
Total (n=279)	183 (65.6)	96 (34.4)
Chicago, Illinois		
Equal participation (n=74)	64 (86.5)	10 (13.5)
Expression of diverse opinions (n=69)	44 (63.8)	25 (36.2)
Respect for others (n=8)	8 (100)	0 (0)
Adoption of a societal perspective (n=4)	1 (25)	3 (75)
Reasoned justification of ideas (n=60)	44 (73.3)	16 (26.7)
Compromise or movement toward consensus (n=91)	44 (48.4)	47 (51.6)
Total (n=306)	205 (67)	101 (33)
Total across sites (n=735)	516 (70.2)	219 (29.8)

XSL•FO
RenderX

Table 3. Second project summary table—number of facilitator remarks coded as consistent and inconsistent with principles of democratic deliberation by principle.

	Equal participation, n/N (%)		Expression of diverse opinions, n/N (%)		Respect for others, n/N (%)		Adoption of a societal perspective, n/N (%)		Reasoned justification of ideas, n/N (%)		Compromise or movement toward consensus, n/N (%)	
	Consistent	Inconsistent	Consistent	Inconsistent	Consistent	Inconsistent	Consistent	Inconsistent	Consistent	Inconsistent	Consistent	Inconsistent
Tampa, Florida	9/11 (81.8)	2/11 (18.2)	36/41 (87.8)	5/41 (12.2)	8/8 (100)	0/8 (0)	4/6 (66.7)	2/6 (33.3)	7/7 (100)	0/7 (0)	29/38 (76.3)	9/38 (23.7)
Baltimore, Maryland	6/6 (100)	0/6 (0)	6/7 (85.7)	1/7 (14.3)	2/2 (100)	0/2 (0)	0/0 (0)	0/0 (0)	3/3 (100)	0/3 (0)	18/21 (85.7)	3/21 (14.3)
Denver, Colorado	59/94 (62.8)	35/94 (37.2)	24/41 (58.5)	17/41 (41.5)	9/9 (100)	0/9 (0)	6/14 (42.9)	8/14 (57.1)	27/29 (93.1)	2/29 (6.9)	58/92 (63)	34/92 (37)
Chicago, Illinois	64/74 (86.5)	10/74 (13.5)	44/69 (63.8)	25/69 (36.2)	8/8 (100)	0/8 (0)	1/4 (25)	3/4 (75)	44/60 (73.3)	16/60 (26.7)	44/91 (48.4)	47/91 (51.6)
Total	138/185 (74.6)	47/185 (25.4)	110/158 (69.6)	48/158 (30.4)	27/27 (100)	0/27 (0)	11/24 (45.8)	13/24 (54.2)	81/99 (81.8)	18/99 (18.2)	149/242 (61.6)	93/242 (38.4)

Equal Participation

Overview

Across the 4 sites, facilitators made 185 remarks that were coded to the principle of equal participation. A total of 74.6% (138/185) of the remarks were coded as consistent with the principle (encouraged equal participation), and 25.4% (47/185) of the remarks were coded as inconsistent with the principle (discouraged equal participation). The proportions of consistent remarks made at each site were as follows: 6.5% (9/138) in Tampa, 4.3% (6/138) in Baltimore, 42.8% (59/138) in Denver, and 46.4% (64/138) in Chicago. The proportions of inconsistent remarks made at each site were as follows: 4% (2/47) in Tampa, 0% in Baltimore, 74% (35/47) in Denver, and 21% (10/47) in Chicago.

Consistent With the Principle

In total, 3 types of remarks were coded as consistent with the principle of equal participation. The remarks (1) conveyed the importance of equal participation, (2) invited individual deliberants or subgroups (eg, youth and parents) to speak, and (3) encouraged more talkative deliberants to allow quieter members time to speak.

Facilitators routinely promoted equal participation by making remarks that emphasized the importance of equal participation. These remarks often occurred when facilitators initially discussed deliberation goals and ground rules. The facilitators emphasized that they valued all perspectives and wished to hear from all deliberants. Facilitators also encouraged equal participation throughout the sessions by frequently asking whether anyone else had something to add. A facilitator in Tampa, for example, said, “Comments from others, anybody?” (session 1).

Facilitators also called on individual deliberants or subgroups of deliberants who had been quiet or who had not contributed to a particular discussion. Facilitators sometimes engaged quiet deliberants by directing questions specifically to them (eg, “[XXXX], what are your thoughts on...?”). Facilitators often encouraged youth deliberants to participate. For example, during a discussion about actions that could be taken to protect youth who self-consent for a study, a facilitator in Chicago, said the following:

> *Do we have any youth that would like to share an opinion or thoughts here?* [Session 4; breakout group 3]

In some instances, facilitators encouraged deliberants who had already contributed to a discussion to give quieter members a chance to speak. These requests respected the contributions of more talkative deliberants but provided opportunities for silent or disengaged deliberants to weigh in. A facilitator in Denver, for example, said the following:

> *So if you’re a person like me who has no problem speaking in front of others, maybe step back once before you make repeat additional comments to others. Whereas if you’re a person who tends to wait a little bit more, if you can, step out of your comfort zone and try to make a comment and just let people know where you’re coming from as early as you feel safe doing so and we would all appreciate that.* [Session 1]

Inconsistent With the Principle

In total, 2 types of remarks were coded as inconsistent with the principle of equal participation. The remarks (1) directed questions toward deliberants who had already spoken on an issue and (2) excused quieter deliberants from speaking.

Some facilitators called on deliberants who had already contributed multiple times instead of on those who had been quiet. A facilitator in Tampa, for example, said the following:

> *So, before we do that [move to another topic], I thought we’d just give one last comment to the one that—to the person that kicked this part of the*

XSL•FO
RenderX

> *discussion off. [XXXX], would you like to say anything else?* [Session 4]

In some instances, facilitators interacted with a few deliberants who were actively discussing an issue without attempting to draw in other deliberants.

Facilitators sometimes made remarks that served to excuse quieter deliberants from providing comments. In a few instances, facilitators suggested that not everyone needed to contribute to a particular discussion. A facilitator in Denver, for example, said the following:

> *And this case, it doesn't have to be everyone that shares...* [Session 1; breakout group 2]

Facilitators occasionally invited quieter deliberants to share their views but then excused them doing so. A facilitator in Denver, for example, asked deliberants who had not spoken to "move out of their comfort zone" but then said, "We would love it if you can do that verbally but if you can't, that is okay" (session 4).

Expression of Diverse Opinions

Overview

Across the 4 sites, facilitators made 158 remarks that were coded to the principle of expression of diverse opinions. A total of 69.6% (110/158) of the remarks were coded as consistent with the principle (encouraged expression of diverse opinions), and 30.3% (48/158) of the remarks were coded as inconsistent with the principle (discouraged expression of diverse opinions). The proportions of consistent remarks made at each site were as follows: 32.7% (36/110) in Tampa, 5.5% (6/110) in Baltimore, 21.8% (24/110) in Denver, and 40% (44/110) in Chicago. The proportions of inconsistent remarks made at each site were as follows: 10% (5/48) in Tampa, 2% (1/48) in Baltimore, 35% (17/48) in Denver, and 52% (25/48) in Chicago.

Consistent With the Principle

In total, 3 types of remarks were coded as consistent with the principle of expression of diverse opinions. The remarks (1) conveyed the importance of expression of diverse opinions, (2) invited diverse opinions during deliberative discussions, and (3) asked what others who were not represented in the deliberations might think about deliberative topics.

Facilitators routinely encouraged the expression of diverse opinions when initially discussing the purpose and the ground rules of the deliberations. They stressed that conflict and dissent, if done amicably, were important to the deliberative process. Facilitators also emphasized the importance of expressing diverse views throughout the sessions. A facilitator in Chicago, for example, said the following:

> *And I want to emphasize again that it is really important for us to mention that dissent is absolutely, completely fine here, and disagreement, as long as it's respectful, is quite helpful. So, please, if you do feel a little differently about any of the actions or any trade-offs than others, please don't feel like you need to self-censor in any way...* [Session 4]

Facilitators sometimes invited deliberants to share views that were at odds with views expressed by other deliberants. They often did this by inquiring about "other perspectives" among deliberants. In Baltimore, for example, as the group was reaching a consensus on a recommendation that community partners be involved in the consent process, the facilitator asked the following:

> *So, I don't think we want to jump to a show of hands yet, but can I just encourage people to weigh in on the notion of community advocates or community partners, especially if you've got concerns about them? If you have concerns about them, it'd be great if you mentioned that and articulated them.* [Session 4]

Facilitators regularly asked deliberants to surmise how the views of other persons or groups not represented in the deliberation might differ from the deliberants' views. The facilitators indicated that these questions were aimed at "bringing in as many perspectives as possible." In Denver, for example, a facilitator said the following:

> *Do we think that there were any perspectives that might be missing, you know, in our particular group?* [Session 2; breakout group 3]

The following excerpt from the Tampa deliberation demonstrates how questions about persons not represented in the group can draw out diverse perspectives:

> *Facilitator: So, I guess in other words, do we think—so I think a good amount of the discussion thus far is some of the group is really coming from similar places. So, for the folks who don't come from the same angle that members of this group are, are there perspectives out there that are different than what we're talking about? And what do you think they would say about the conversation we're having today?*
>
> *Deliberant: I would say that, of course, there's perspective that's very different from this group. I think we're all here for kind of a similar reason. And I think that there are a lot of people out there that do not want their children to have sex education. They do not want their children to have access to birth control. They do not want their children to have access to be able to be in medical research that has to do with HIV or STDs, period. I think there is a large group of parents that would be absolutely freaked out by everything that we've talked about today, you know. So yeah, there's a wide range of comfortableness with this topic.* [Session 1; breakout group 1]

Inconsistent With the Principle

In total, 2 types of remarks were coded as inconsistent with the principle of expression of diverse opinions. The remarks (1) revealed the personal opinions of the facilitators and (2) interrupted the expressions of diverse opinions by deliberants.

We chose to code remarks in which facilitators expressed their own personal opinions as inconsistent with the principle of

XSL•FO
RenderX

expression of diverse opinions. Deliberation facilitators are typically dissuaded from expressing their own views on topics germane to the deliberation as this can squelch dissenting or challenging views of deliberants due to power differences between facilitators and deliberants. Although facilitators in our deliberation did not routinely express their own opinions, they did so occasionally. In Tampa, for example, a facilitator said the following:

> *I don't know that we're very respectful of young people. I mean, just to be blunt, like I don't think we're respectful of their time, I don't think we're respectful of their interests. I think, you know, research—we do cater to adults, and I just don't think we often were thinking as much about, you know, like you said, interest in accessibility of young people, because we're—I don't know, maybe just not cognizant of what those were.* [Session 1; breakout group 2]

The following excerpt from the Tampa deliberation demonstrates how expression of personal opinions by facilitators can lead deliberants to agree rather than discuss their own views. In this example, a deliberant suggested that school personnel "flag" students who would benefit from or be eligible for a research study. A facilitator comment changed the focus from school personnel "flagging" potentially eligible students to school personnel being well positioned to weigh in on whether a student would be competent to consent:

> *Deliberant: So, maybe [teachers] just tapping into it in the beginning of the study for identifying people who would be eligible or benefit might be helpful.*
>
> *Facilitator: Or at least being able to give a comment in terms of their ability to consent and understand.*
>
> *Deliberant: Yes.*
>
> *Facilitator: So, I think the teachers really know the kids enough to know whether they will be able to understand the level of nuance that's required [to consent]. Maybe not with any responsibility related to that, but at least being able to address this question of maturity versus, you know, being able to—*
>
> *Deliberant: I completely agree. And we, I mean, oftentimes [we] know about the ability for their parents to consent, the ability for their parents—you know, their level of education, how much they understand about the study themselves. I mean, I know my kids pretty well. And I know their families pretty well. I know their living situations. I know a lot about them. So, at least tapping into that perspective, we definitely have a lot to offer in that point of view.* [Session 4]

In some instances, facilitators inquired about diverse opinions but did not wait for deliberants to comment, and instead, the facilitators answered their own inquiries. In Denver, for example, after a deliberant spoke in favor of community consultation for youth who had no support at home, the facilitator said the following:

> *What do others think about this? Maybe something perhaps how practical this might be because something like obtaining community input requires a community to show up...So do you think it would realistically work?* [Session 4; breakout group 1]

Respect for Others

Overview

Across the 4 sites, facilitators made 27 remarks that were coded to the principle of respect for others. All 27 remarks (100%) were coded as consistent with the principle (promoted respect for others). No remarks were coded as inconsistent with the principle (discouraged respect for others). The proportions of consistent remarks made at each site were as follows: 30% (8/27) in Tampa, 7% (2/27) in Baltimore, 33% (9/27) in Denver, and 30% (8/27) in Chicago.

Consistent With the Principle

In total, 2 types of remarks were coded as consistent with the principle of respect for others. The remarks (1) conveyed the importance of respect for others and (2) acknowledged instances in which deliberants showed respect for others.

The facilitators routinely stressed the importance of respect for others in their introductory remarks to deliberants. Facilitators emphasized that showing respect for others by actively listening and considering others' views was a ground rule of the deliberations. A facilitator in Denver, for example, said the following:

> *We were all experts in our own experience. And when people were speaking about that, the best thing we can do for each other is listen to that and learn from it.* [Session 1]

Facilitators at times acknowledged when deliberants showed respect for others by pointing out exchanges among deliberants that were particularly respectful. A facilitator in Baltimore, for example, reflected on a previous deliberation:

> *We have parents, and younger people, and people that have different relationships with their parents, and it was super respectful, and I think kind of collaborative.* [Session 3]

Adoption of a Societal Perspective

Overview

Across the 4 sites, facilitators made 24 remarks that were coded to the principle of adoption of a societal perspective. A total of 46% (11/24) of the remarks were coded as consistent with the principle (encouraged adoption of a societal perspective), and 54% (13/24) of the remarks were coded as inconsistent with the principle (discouraged adoption of a societal perspective). The proportions of consistent remarks made at each site were as follows: 36% (4/11) in Tampa, 0% in Baltimore, 55% (6/11) in Denver, and 9% (1/11) in Chicago. The proportions of inconsistent remarks made at each site were as follows: 15% (2/13) in Tampa, 0% in Baltimore, 62% (8/13) in Denver, and 23% (3/13) in Chicago.

XSL•FO
RenderX

Consistent With the Principle

One type of remark was coded as consistent with the principle of adoption of a societal perspective. This type of remark asked deliberants to consider issues of civic good in their consensus recommendations. Facilitators occasionally introduced the idea that the end goal of research was to benefit society as a whole rather than individual research participants. One facilitator in Chicago, for example, said the following:

> *Like clinical [practice] really focuses on the individual to improve that individual's health, whereas clinical research, while there's often benefit to the individual...the purpose of clinical research is to advance scientific knowledge so we know what's happening in the population.* [Session 1]

In a few instances, facilitators asked deliberants to consider what is "good for society" when forming their views. A facilitator in Denver, for example, said the following:

> *Think about these concerns...from a really broad perspective—that means not just from your specific stakeholder perspective but from a community and researcher perspective.* [Session 3; breakout group 5]

Inconsistent With the Principle

One type of remark was coded as inconsistent with the principle of adoption of a societal perspective. In these instances, facilitators encouraged deliberants to share personal experiences and consider how recommendations would affect them personally without inquiring about the societal perspective. For example, the following excerpt from the Denver deliberation demonstrates a missed opportunity to move a discussion from a deliberant's personal experience as a parent to a discussion of the greater good:

> *Deliberant: Me as an adolescent parent, I would be able to trust that if my adolescent made the choice to join the study or like a research that they were doing the right thing. And I think it's also good for them to know like what's going on today, versus them not knowing, like what's going on in the world.*
>
> *Facilitator: Yeah thanks. And what evidence helped you come to that conclusion? Like obviously that perspective as a parent?*
>
> *Deliberant: So I work at* [local hospital] and we do like birth control and you know, the adolescents are able to consent for themselves without their parents knowing. And I feel like some parents don't talk about like HIV and medications, and what the possibility can be. You know they want to keep their kids hidden from everything and protected from everything. But in our reality if we don't communicate with our adolescents then that's where us as parents get in trouble and our adolescents get in trouble. [Session 3; breakout group 5]

Reasoned Justification of Ideas

Overview

Across the 4 sites, facilitators made 99 remarks that were coded to the principle of reasoned justification of ideas. A total of 82% (81/99) of the remarks were coded as consistent with the principle (encouraged reasoned justification of ideas), and 18% (18/99) of the remarks were coded as inconsistent with the principle (discouraged reasoned justification of ideas). The proportions of consistent remarks made at each site were as follows: 9% (7/81) in Tampa, 4% (3/81) in Baltimore, 33% (27/81) in Denver, and 54% (44/81) in Chicago. The proportions of inconsistent remarks made at each site were as follows: 0% in Tampa, 0% in Baltimore, 11% (2/18) in Denver, and 89% (16/18) in Chicago.

Consistent With the Principle

In total, 2 types of remarks were coded as consistent with the principle of reasoned justification of ideas. The remarks asked deliberants to (1) provide a reason for their stated views or (2) reveal what information influenced their views.

Some facilitators encouraged deliberants to provide a rationale or explanation for why they took a particular stance on a deliberation topic. In some instances, facilitators discussed the importance of understanding the "why" or the "logic" behind opinions. A facilitator in Tampa, for example, said the following:

> *We want to give reasons for—in sentences we were writing [consensus recommendations], we want to also say why we think this, not only that, why we think it should be done.* [Session 4]

Facilitators at times asked deliberants to pinpoint information they had heard in the course of the deliberation or elsewhere that had influenced their opinions. Some facilitators asked deliberants to state what facts supported their views. A facilitator in Baltimore, for example, said the following:

> *So, I was going to try to nudge the discussion a little bit more towards explanations you've heard of current practice or the background reading you did coming in...* [Session 4]

Inconsistent With the Principle

One type of remark was coded as inconsistent with the principle of reasoned justification of ideas. This type of remark encouraged deliberants to justify their opinions but then interrupted them from doing so by supplying a rationale or by excusing the participant from having to explain their reasoning. We did not code remarks in which facilitators assured deliberants that they were free to share their views only if they were comfortable doing so, especially early in the deliberations, but rather remarks in which facilitators invited deliberants to discuss the reasoning for their views but did not give them the opportunity or space to do so.

The following excerpt from the Denver deliberation demonstrates how a facilitator remark can discourage discussion of reasoned justification of ideas. In this example, a deliberant opined that opening a study to cisgender or heterosexual youth

XSL•FO
RenderX

would not change their opinion about minor self-consent. The facilitator asked a deliberant to justify their opinion but then excused them from doing so:

> *Deliberant: And to what [deliberant] said, I think that opening it up to a larger population doesn't necessarily mean that you're going to be getting participants from that larger population if they don't think that it's an issue that impacts them. Yeah, so I'm not really sure if—how much of an impact it would have for me in terms of like the consent process.*
>
> *Facilitator: Okay. Thanks. And what would you say helped you come to that conclusion? What would you say is the information that helped you support that idea? And if you don't want to expand, you don't have to. So don't worry. All right everyone. I think this is a great time to move on to the next permutation.* [Session 3; breakout group 2]

Compromise or Movement Toward Consensus

Overview

Across the 4 sites, facilitators made 242 remarks that were coded to the principle of compromise or movement toward consensus. A total of 61.6% (149/242) of the remarks were coded as consistent with the principle (encouraged compromise and movement toward consensus), and 38.4% (93/242) of the remarks were coded as inconsistent with the principle (discouraged compromise and movement toward consensus). The proportions of consistent remarks made at each site were as follows: 19.5% (29/149) in Tampa, 12.1% (18/149) in Baltimore, 38.9% (58/149) in Denver, and 61.1% (91/149) in Chicago. The proportions of inconsistent remarks made at each site were as follows: 10% (9/93) in Tampa, 3% (3/93) in Baltimore, 37% (34/93) in Denver, and 51% (47/93) in Chicago.

Consistent With the Principle

In total, 3 types of remarks were coded as consistent with the principle of compromise or movement toward consensus. The remarks (1) introduced compromise and consensus as a goal of the deliberative process, (2) promoted discussion of common ground and encouraged deliberants to seek consensus, and (3) asked deliberants to summarize the consensus they had reached.

Facilitators discussed the goals of compromise and consensus when making introductory remarks about the deliberation. They informed deliberants that, while they did not need to agree, they were to explore common ground and, if possible, reach consensus on issues that were the focus of the deliberation. In Denver, for example, the facilitator stated the following:

> *And then in the final session, we'll review all the actions and tradeoffs that we've come up with, see where we have any common ground and collaborate to turn this into an actual list of recommendations or at least guiding principles...* [Session 1]

Often later in the deliberations, facilitators encouraged participants to consider common ground and move toward consensus. The facilitators asked whether deliberants were "on the same page" about an issue, and if they had not reached a consensus, the facilitators made remarks aimed at encouraging compromise. The facilitators often asked deliberants to consider what "tradeoffs" they would be willing to accept to agree on recommendations. In Baltimore, for example, a facilitator said the following at the start of the last session:

> *So, now in this phase, in addition to revising or adding or tweaking the things that we were just talking about, the main goal of this next phase is to come up with—there's different ways to put it, a set of principles or maybe a set of recommendations.* [Session 4]

While facilitators often summarized their perceptions of deliberant common ground, at times they asked deliberants to discuss the consensus they had reached and recommendations they wished to make. During deliberations in Chicago (the final deliberation site), for example, facilitators asked for volunteers from each breakout group to share a summary of their group's consensus and recommendations. The volunteers provided comprehensive summaries to the larger group. The following excerpt shows how a deliberant summarized the breakout group's conclusions, trade-offs, and recommendations related to the use of community town halls:

> *Facilitator: So, we're going to try to report out from all our three different groups and kind of keep track of where we had similarities and differences, and then we'll start to deliberate as a large group as best we can right after that. So, let's get to it. We'll start off with group one. I know that the note taker took some notes on the slides, and so—and we have a participant who will be walking us through those.*
>
> *Deliberant: OK. Yeah. So, as far as like actions and group conclusions, how we feel about a town hall with community before study recruitment begins...as well as like the paid community advisory board that can be engaged at every step along the way. We feel one—it's useful if the young people do not have support at home. We require honest discussion with the young person as well as like trained professionals to screen. So, not that this wouldn't be useful for people who have like a more stable and supportive home life but especially important to folks who don't have that or who don't have it consistently. A negative is in the community, people [who] aren't involved with other people that much would not—would like—prefer not to really know about these sorts of things unless like some sort of like major issue popped up. And of course, that varies from like neighborhood to neighborhood but it is kind of like a general thing. Like, at the end of the day, like the majority of the US [are] still like largely conservative, so, yeah. And then c, it could be good but it could be bad like if the broader community were invited. It could attract people who maybe have like more like ill-intent. So, like, you know, the fundamentalist types that come and like, you know, want to disrupt instead of really have a constructive conversation. So, we'd have to just be wary of people who want to show up and just cause dissent and not really like be a part of the*

XSL•FO
RenderX

> *conversation. Not—You know, and folks who aren't really centering the youth at the end of the day, or they may claim to center the youth but it's really like under the guise of like homophobia, transphobia, hodophobia, all those things. We decided it's better if we were going to have a community advisory board that it's a paid group of people who are screened. And we would want to be selective, including people already involved with the youth, such as, you know, leaders and folks working in schools, folks specifically working with like youth HIV prevention, people working within like the queer community who know how to advocate for folks who are—yeah, show up in the range of like queerness and gender and all of these things as well as like near peers. So, people who maybe like not too long ago were going through something similar to this. Yeah. OK...* [Session 4]

The deliberant continued to discuss a number of other recommendations that the group had agreed on.

Inconsistent With the Principle

One type of remark was coded as inconsistent with the principle of compromise or movement toward consensus. In these remarks, facilitators conveyed their "take on" deliberant consensus, often without first asking deliberants what common ground they felt they had reached. Facilitators often provided somewhat lengthy summaries of conclusions reached by breakout groups or by deliberant groups as a whole. These reports eliminated the opportunity for deliberants to consider and articulate points of common ground. The following excerpt from Chicago occurred after deliberant volunteers had reported on the consensus of their breakout groups (as described previously). As the session wound down, the facilitator provided a summary of common ground among the 3 breakout groups based on notes taken by the research team:

> *Facilitator: Well, in the interest of time, I'm going to do something. I'm going to try to summarize the common ground that we've heard...I'm going to try to come up with a summary of what our common ground is that would involve the basic elements of what actions are acceptable under most circumstances and why?...You all see nuance that needs to be added that was in your small groups that we want to make sure we keep in our summarization of your discussion, please keep that in mind and we'll take your revisions right after this. So, I think that we agreed that common ground things that would be necessary or things that would be at least useful in most situations would be community supports such as a town hall and an advisory board. Although there may be some issues with some of these in certain implementation aspects. There was generally support for ongoing support for use throughout this with workshops, counseling during and counseling afterwards. Although people wanted to make sure that it was known while these things were positive, they were not a standalone. So generally, under most circumstances, both those things are good at least options to be part of this. For minor self-consent it sounds like what I was hearing from the group as common ground would be minor self-consent good after the age of 16. And/or in addition to that with teach back as part of it, or on research that does not necessarily have such intense side effects. Or, as one group mentioned, for behavioral, not medical research. General agreement that minor self-consent with adult help is useful, possibly especially for people under 16, but that both this and the trusted adult permission could introduce problems if it becomes in conflict with an adult—with a parent, and potentially for other reasons as well. And then, general agreement that under most situations, an ombudsperson being available as part of the project for all participants—all youth participants, but especially those without—where parental consent isn't feasible was positive. And then, parental permission for folks who were less than 16 in most areas...* [Session 4]

The facilitator then reported on several other areas of common ground, but the session ended soon after this summary, and there was no opportunity for deliberants to discuss "nuances," as the facilitator had suggested.

Discussion

Principal Findings

The analysis presented in this paper provides an example of how the framework that our team developed to assess facilitator fidelity to principles of democratic deliberation was used in a 4-site public deliberation project focusing on minor self-consent for biomedical HIV prevention research. The assessment was implemented as outlined in our published blueprint report [16]. The examination of facilitator fidelity at all 4 sites allowed us to examine patterns of facilitator remarks across sessions and sites, provided sufficient data to delineate different types of remarks aligned with each principle, and yielded rich verbatim exemplars of facilitator remarks and facilitator and deliberant dialogue reflecting each principle. The counts provide affirmation of facilitator fidelity as most facilitator remarks made across all sessions at all sites were coded as consistent with a principle, and moreover, the number of coded facilitator remarks generally increased over time across the sites.

Implications

Although the code counts do not serve as definitive quantitative metrics of fidelity as the counts are influenced by variations in the deliberation sites and sessions (eg, number of deliberants attending, structure of sessions, and unique group characteristics), they do provide guideposts for improving deliberation quality. In Table 4, we summarize facilitator fidelity across the 4 sites, provide conclusions about facilitator fidelity across sites, and offer recommendations for future deliberations.

XSL•FO
RenderX

Table 4. Summary of facilitator fidelity across sites, conclusions about facilitator fidelity across the 4 sites, and recommendations for future deliberations.

Principle	Summary of facilitator fidelity across the sites	Conclusions about facilitator fidelity across the sites	Recommendations for future deliberations
Equal participation	Few remarks consistent with the principle were made at the first site, and many more were made at the last 2 sites (Tampa: n=9; Denver: n=59; Chicago: n=64). Remarks inconsistent with the principle varied across sites (Tampa: n=2; Denver: n=35; Chicago: n=10).	Remarks consistent with this principle increased after the first 2 deliberations because the research team noted during debriefings that some participants were not well engaged. The team then encouraged facilitators to actively invite quieter members, especially youth, to participate and ask talkative deliberants to pause to give quieter members time to formulate responses. Facilitators also continued on occasion throughout the sessions to encourage persons who had already spoken to comment again or invite quiet deliberants to comment but then immediately assure them that they did not "need" to if they were uncomfortable.	Facilitators should continue to encourage more reticent members to contribute to the discussion and, on occasion, respectfully encourage more talkative members to refrain until others have spoken. Facilitators should be discouraged from inviting deliberants to contribute and then "excuse" them from doing so.
Expression of diverse opinions	Many remarks consistent with the principle were made across the sites (Tampa: n=36; Denver: n=24; Chicago: n=44). Remarks inconsistent with the principle were made with some frequency in later deliberations (Tampa: n=5; Denver: n=17; Chicago: n=25).	Facilitators at all sites reassured participants that differences were valued and dissent was expected or asked them to consider views that might not be represented in their group. However, some facilitators also on occasion continued to give their own opinions rather provide facts or expert information about the issues being discussed.	Deliberants should be encouraged throughout the deliberations to express diverse opinions, and the sharing of personal opinions by facilitators should always be avoided.
Respect for others	Few remarks consistent with the principle were made across the sites (Tampa: n=8; Denver: n=9; Chicago: n=8). No remarks inconsistent with the principle were made across the sites.	There were very few instances in which deliberants did not interact with each other respectfully, and therefore, more facilitator remarks encouraging them to do so were unnecessary.	Once respect for others is established as a ground rule, deliberants may not need ongoing encouragement to act respectfully toward one another, but facilitators should remain alert to and intervene in response to any instances of disrespect.
Adoption of a societal perspective	Few remarks consistent with the principle (Tampa: n=4; Denver: n=6; Chicago: n=1) were made across the sites. Few remarks inconsistent with the principle (Tampa: n=2; Denver: n=8; Chicago: n=3) were made across the sites.	At times, facilitators diverted deliberants' attention toward their own concerns rather than toward consideration of what would be good for society. The concept of societal benefit, as opposed to self-benefit, might be a difficult concept for some facilitators to understand or promote.	Strategies are needed to help facilitators introduce the concept of societal good and encourage consideration of civic benefit as deliberants finalize their recommendations. While discussion of personal experiences early in deliberations can help members get to know one another, these discussions should not be encouraged later in deliberations as participants move toward consensus.
Reasoned justification of opinions	Remarks consistent with the principle were made often, especially during the last 2 sessions (Tampa: n=7; Denver: n=27; Chicago: n=44). Remarks inconsistent with the principle were made only infrequently (Tampa: n=0; Denver: n=2; Chicago: n=16).	The increase in consistent remarks at the last 2 sites might be attributed to facilitator training as facilitators were encouraged to ask deliberants to expand on how they came to the views they expressed rather than just state their opinions. However, facilitators were less likely to ask deliberants about information that influenced their views. Moreover, infrequently, some facilitators continued to supply reasons for deliberants' views rather than encouraging them to provide their own rationale.	Facilitators should routinely ask deliberants to discuss the rationale for their views as well as identify what information they considered in coming to these views. Facilitators should never supply their understanding of the reasons or the information underlying deliberant views if the deliberants have not articulated these justifications themselves.

XSL•FO
RenderX

Principle	Summary of facilitator fidelity across the sites	Conclusions about facilitator fidelity across the sites	Recommendations for future deliberations
Compromise or movement toward consensus	Many remarks consistent with the principle were made across the sites (Tampa: n=29; Denver: n=58; Chicago: n=44). Remarks inconsistent with the principle were fairly frequent in the last 2 deliberations (Tampa: n=9; Denver: n=34; Chicago: n=47).	Facilitators often encouraged deliberants to consider the views of the group, find common ground, and consider trade-offs but also frequently provided a summary of what the facilitators believed to be a consensus. The increase in consistent remarks in Chicago likely reflected a change in the deliberation plan (run of show) as facilitators were instructed to ask 1 deliberant from each breakout group to assume responsibility for reporting their groups' points of agreement and recommendations to the larger group. However, the frequency of the inconsistent remarks even late in the deliberations suggests that facilitators continued to discuss what they believed to be the consensus reached by deliberants.	Research teams should develop strategies that allow facilitators to guide deliberants toward compromise and consensus without taking over that responsibility.

Limitations

While the assessment presented in this paper can inform future deliberation planning and facilitator training, it also raises some issues for further consideration in refining the framework. For example, few facilitator remarks in our deliberation were coded as consistent with the principle of adoption of a societal perspective, and we surmised that this was because some facilitators did not fully understand the concept and, thus, often invited consideration of personal rather than public welfare in the deliberations. This conclusion is challenged by the work of Lehoux and Proulx [25], who suggest that some deliberants can serve as both public and patient representatives. In the former role, they bring a collective perspective, and in the latter role, they bring a personal perspective. In their deliberation on the value of technological health innovations, Lehoux and Proulx [25] found that deliberants alternated between these 2 modes of engagement. When arguing from a collective perspective, they addressed issues such as social costs, environmental protection, and individual freedoms, whereas when they argued from a personal perspective, they addressed issues such as the functionality of the technology, its dehumanizing effects on users, and its potential to increase individuals' autonomy. The researchers suggest that these 2 engagement modes are complementary and each is needed to advance deliberative discussions. They recommend that facilitators can first elicit one standpoint and then elicit counterpoints from the other standpoint. This work suggests that the facilitator fidelity framework may need to provide a way for researchers to evaluate the extent to which facilitators are able to elicit discussion of both perspectives and illuminate the tension between them rather than rate questions about personal experiences and needs as inconsistent with the principle of adoption of a societal perspective.

Another issue not sufficiently addressed in the framework is plurality. The framework includes the principles of expression of diverse opinions and compromise and movement toward consensus but does not call for an assessment of how facilitators unearth and account for multiple perspectives—despite the fact that deliberants are chosen for their diverse views. In our assessment, remarks coded as inconsistent with the principle of expression of diverse opinions were primarily personal opinions expressed by the facilitators, and remarks coded as inconsistent with the principle of compromise and movement toward consensus were primarily remarks that revealed the facilitators' "takes" on deliberant points of agreement. While these were important findings and led to recommendations for facilitator training, the nuances of effectively managing plurality are not captured completely in the framework. Baker et al [24] addressed the question of "what to do when people disagree" in deliberations and discussed how different perspectives should be elicited, deliberated, and integrated into recommendations. They developed a multilevel framework that allows for transparency of plurality and considers the integration of perspectives through counting (ie, strength of preferences), coherence (ie, the logical consistency of the argument), and consensus (ie, recommendations agreed upon through deliberation). They propose that agreements and disagreements occur at 3 levels: principles (ie, high-level normative statements), policies (ie, midlevel operational rules), and patients (ie, case-level judgments). Plurality can be disentangled by examining counting, consensus, and coherence at each level. For example, deliberants may reach consensus on general principles, but these principles may not be reflected in their policy recommendations, and thus, their decision-making lacks coherence. The authors suggest that facilitation methods can be developed to elicit values at all levels and uncover instances in which principles, policies, and patients are in tension, thereby promoting discussions of inconsistencies, identification of values that have the greatest force, and increased justification of collective outputs. This work could inform the refinement of our framework by introducing important nuances in our evaluation of how facilitators elicit diverse opinions and guide deliberants to integrate multiple perspectives into their final recommendations.

The issue of facilitator influence raised in our assessment is considered in a discussion by Kuhar et al [26]. We identified a few exchanges in which facilitators expressed their opinions, which in turn influenced deliberant responses. Kuhar et al [26] point out that neutrality of facilitators is a deliberation ideal but facilitator influence on outcomes occurs regularly in deliberative discussions. We recognize that, while our framework classified

XSL•FO
RenderX

a few overt instances of facilitator influence, it is possible that facilitators influenced discussions in more nuanced ways not addressed in the framework (eg, information presented in expert testimonies), suggesting that facilitator neutrality is an issue that warrants further development. For example, assessment of facilitators' performance might include the extent to which they have reflected on their own views and developed an awareness of how their views can influence deliberation outcomes, even if subtly.

In addition to the limitations described previously, a few other concerns need to be addressed in the refinement of the framework. Kuhar et al [26] pose that the privileging of logical argumentation, which is central to most deliberation theories but not valued by all cultures, can marginalize underrepresented groups. In future iterations of our framework, we intend to add a principle that extends beyond expression of diverse perspectives and addresses mitigating social inequalities more explicitly. In addition, for the assessment reported in this paper, all team members (eg, deliberation experts, research team members, and site stakeholders) were treated as one group of facilitators as there was fluidity in how they contributed to the sessions. However, in future assessments, we will examine role differences more closely when determining how facilitation processes overall affect deliberation quality.

Conclusions

Our assessment of facilitator fidelity to the principles of democratic deliberation in a 4-site deliberation project on the acceptability of youth self-consent for biomedical HIV prevention research affirmed that most facilitator remarks aligned with deliberative principles but also revealed areas in which facilitator performance could be strengthened. Facilitators play an important role in ensuring rich deliberative discussions, and a comprehensive assessment of deliberation quality should include an assessment of facilitator performance. The facilitator fidelity framework and blueprint report developed by our team offer a guide to conduct such an assessment. Systematically assessing facilitator fidelity can not only provide an indication of deliberation quality but can also inform future deliberation planning and facilitator training.

Acknowledgments

The research reported in this manuscript was funded through a Patient-Centered Outcomes Research Institute award (ME-2018C3-14766). The views and perspectives in this report are solely the responsibility of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute or its board of governors or Methodology Committee. No generative artificial intelligence was used in any portion of manuscript writing.

Data Availability

The datasets generated and analyzed during this study are not publicly available, nor are they available from the corresponding author on reasonable request. The Patient-Centered Outcomes Research Institute determines which of its awards meet the criteria for inclusion in its data repository. The terms of the award that supported this study *do not* require participation in the data repository. Much of the information discussed in the deliberations was very sensitive in nature, and the small number of research participants would make it difficult to protect the privacy of their sensitive disclosures in the sessions. For these reasons, the full transcripts will not be made available.

Conflicts of Interest

MAO's spouse is an employee at Eli Lilly, Inc. Together, they are joint small stockholders.

References

1. Blacksher E, Hiratsuka VY, Blanchard JW, Lund JR, Reedy J, Beans JA, et al. Deliberations with American Indian and Alaska native people about the ethics of genomics: an adapted model of deliberation used with three tribal communities in the United States. AJOB Empir Bioeth 2021;12(3):164-178 [FREE Full text] [doi: 10.1080/23294515.2021.1925775] [Medline: 34125006]
2. Scherer M, Kamler A, Weiss L, Blacksher E, Jeavons J, Gold MR. Using public deliberation to set priorities: the case of COVID-19 vaccine access in New York City. J Public Health Manag Pract 2022;28(1):86-94. [doi: 10.1097/PHH.0000000000001471] [Medline: 34797242]
3. Slomp C, Edwards L, Burgess M, Sapir-Pichhadze R, Keown P, Bryan S. Public values and guiding principles for implementing epitope compatibility in kidney transplantation allocation criteria: results from a Canadian online public deliberation. BMC Public Health 2023 May 10;23(1):844 [FREE Full text] [doi: 10.1186/s12889-023-15790-w] [Medline: 37165330]
4. Hiratsuka VY, Beans JA, Blanchard JW, Reedy J, Blacksher E, Lund JR, et al. An Alaska Native community's views on genetic research, testing, and return of results: results from a public deliberation. PLoS One 2020 Mar 16;15(3):e0229540 [FREE Full text] [doi: 10.1371/journal.pone.0229540] [Medline: 32176704]
5. Richmond J, Mangrum R, Wang G, Maurer M, Sofaer S, Yang M, et al. An informed public's views on reducing antibiotic overuse. Health Serv Res 2019 Dec;54(6):1283-1294 [FREE Full text] [doi: 10.1111/1475-6773.13175] [Medline: 31172520]

XSL•FO
RenderX

6. Jagsi R, Griffith KA, Jones RD, Krenz C, Gornick M, Spence R, et al. Effect of public deliberation on patient attitudes regarding consent and data use in a learning health care system for oncology. J Clin Oncol 2019 Dec 01;37(34):3203-3211 [FREE Full text] [doi: 10.1200/JCO.19.01693] [Medline: 31577472]
7. Waljee AK, Ryan KA, Krenz CD, Ioannou GN, Beste LA, Tincopa MA, et al. Eliciting patient views on the allocation of limited healthcare resources: a deliberation on hepatitis C treatment in the Veterans Health Administration. BMC Health Serv Res 2020 May 01;20(1):369 [FREE Full text] [doi: 10.1186/s12913-020-05211-8] [Medline: 32357873]
8. Schwartz PH, O'Doherty KC, Bentley C, Schmidt KK, Burgess MM. Layperson views about the design and evaluation of decision aids: a public deliberation. Med Decis Making 2021 Jul;41(5):527-539 [FREE Full text] [doi: 10.1177/0272989X21998980] [Medline: 33813928]
9. Raj M, Ryan K, Nong P, Calhoun K, Trinidad MG, De Vries R, et al. Public deliberation process on patient perspectives on health information sharing: evaluative descriptive study. JMIR Cancer 2022 Sep 16;8(3):e37793 [FREE Full text] [doi: 10.2196/37793] [Medline: 36112409]
10. Trinidad SB, Blacksher E, Woodbury RB, Hopkins SE, Burke W, Woodahl EL, et al. Precision medicine research with American Indian and Alaska Native communities: results of a deliberative engagement with tribal leaders. Genet Med 2022 Mar;24(3):622-630 [FREE Full text] [doi: 10.1016/j.gim.2021.11.003] [Medline: 34906504]
11. Scott AM, Sims R, Degeling C, Carter S, Thomas R. Developing and applying a deductive coding framework to assess the goals of citizen/community jury deliberations. Health Expect 2019 Jun;22(3):537-546 [FREE Full text] [doi: 10.1111/hex.12880] [Medline: 30864216]
12. Goold SD, Biddle AK, Klipp G, Hall CN, Danis M. Choosing Healthplans All Together: a deliberative exercise for allocating limited health care resources. J Health Polit Policy Law 2005 Aug;30(4):563-601. [doi: 10.1215/03616878-30-4-563] [Medline: 16318163]
13. Rubinelli S, von Groote PM. Stakeholder dialogue as deliberation for decision making in health policy and systems: the approach from argumentation theory. Am J Phys Med Rehabil 2017 Feb;96(2 Suppl 1):S17-S22. [doi: 10.1097/PHM.0000000000000635] [Medline: 28059875]
14. Dillard KN. Envisioning the role of facilitation in public deliberation. J Appl Commun Res 2013 Aug 25;41(3):217-235. [doi: 10.1080/00909882.2013.826813]
15. De Vries R, Stanczyk AE, Ryan KA, Kim SY. A framework for assessing the quality of democratic deliberation: enhancing deliberation as a tool for bioethics. J Empir Res Hum Res Ethics 2011 Sep;6(3):3-17 [FREE Full text] [doi: 10.1525/jer.2011.6.3.3] [Medline: 21931233]
16. Draucker C, Carrión A, Ott MA, Knopf A. Assessing facilitator fidelity to principles of public deliberation: tutorial. JMIR Form Res 2023 Dec 13;7:e51202 [FREE Full text] [doi: 10.2196/51202] [Medline: 38090788]
17. Community solutions to adolescent research consent - minor consent for biomedical HIV research (C-START). National Institutes of Health National Library of Medicine. URL: https://clinicaltrials.gov/study/NCT05371327 [accessed 2023-11-20]
18. Do public deliberation methods help improve the informed consent process for research on sensitive topics? Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/research-results/2019/do-public-deliberation-methods-help-improve-informed-consent-process-research-sensitive-topics [accessed 2023-11-20]
19. Research involving children: report and recommendations. U.S. Department of Health, Education and Dhew Publication. 1977. URL: https://videocast.nih.gov/pdf/ohrp_research_involving_children.pdf [accessed 2023-11-20]
20. CFR - Code of federal regulations title 21. U.S. Food & Drug Administration. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?C [accessed 2023-11-03]
21. Arrington-Sanders R, Wilson CM, Perumean-Chaney SE, Patki A, Hosek S. Brief report: role of sociobehavioral factors in subprotective TFV-DP levels among YMSM enrolled in 2 PrEP trials. J Acquir Immune Defic Syndr 2019 Feb 01;80(2):160-165 [FREE Full text] [doi: 10.1097/QAI.0000000000001901] [Medline: 30640203]
22. Ott MA, Knopf AS. Avoiding a tyranny of the majority: public deliberation as citizen science, sensitive issues, and vulnerable populations. Am J Bioeth 2019 Aug;19(8):28-31 [FREE Full text] [doi: 10.1080/15265161.2019.1619870] [Medline: 31544636]
23. De Vries R, Stanczyk A, Wall IF, Uhlmann R, Damschroder LJ, Kim SY. Assessing the quality of democratic deliberation: a case study of public deliberation on the ethics of surrogate consent for research. Soc Sci Med 2010 Jun;70(12):1896-1903 [FREE Full text] [doi: 10.1016/j.socscimed.2010.02.031] [Medline: 20378225]
24. Baker R, Mason H, McHugh N, Donaldson C. Public values and plurality in health priority setting: what to do when people disagree and why we should care about reasons as well as choices. Soc Sci Med 2021 May;277:113892 [FREE Full text] [doi: 10.1016/j.socscimed.2021.113892] [Medline: 33882440]
25. Lehoux P, Proulx S. Deliberating as a public representative or as a potential user? Two complementary perspectives that should inform health innovation policy. Healthc Policy 2019 May;14(4):28-38 [FREE Full text] [doi: 10.12927/hcpol.2019.25858] [Medline: 31322112]
26. Kuhar M, Krmelj M, Petrič G. The impact of facilitation on the quality of deliberation and attitude change. Small Group Res 2019 Aug 26;50(5):623-653. [doi: 10.1177/1046496419861439]

XSL•FO
RenderX

Abbreviations

IRB: institutional review board

Edited by A Mavragani; submitted 20.03.24; peer-reviewed by A Grosso; comments to author 13.11.24; revised version received 15.11.24; accepted 24.11.24; published 13.02.25.

Please cite as:
Draucker CB, Carrión A, Ott MA, Hicks AI, Knopf A
A 4-Site Public Deliberation Project on the Acceptability of Youth Self-Consent in Biomedical HIV Prevention Trials: Assessment of Facilitator Fidelity to Key Principles
JMIR Form Res 2025;9:e58451
URL: https://formative.jmir.org/2025/1/e58451
doi:10.2196/58451
PMID:

©Claire Burke Draucker, Andrés Carrión, Mary A Ott, Ariel I Hicks, Amelia Knopf. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Designing a Smartphone-Based Virtual Reality App for Relaxation: Qualitative Crossover Study

Amandine Verstegen[1], MSc; Tom Van Daele[1], PhD; Bert Bonroy[2], ING, PhD; Glen Debard[2], ING, PhD; Romy Sels[2], ING, MSc; Marlon van Loo[3], MSc; Sylvie Bernaerts[1], PhD

[1]Psychology and technology, Centre of Expertise - Care and Well-being, Thomas More University of Applied Sciences, Antwerp, Belgium

[2]Mobilab & Care, Centre of Expertise - Care and Well-being, Thomas More University of Applied Sciences, Geel, Belgium

[3]Prevention and empowerment, Centre of Expertise - Care and Well-being, Thomas More University of Applied Sciences, Antwerp, Belgium

Corresponding Author:
Sylvie Bernaerts, PhD
Psychology and technology
Centre of Expertise - Care and Well-being
Thomas More University of Applied Sciences
Molenstraat 8
Antwerp, 2018
Belgium
Phone: 32 34324050
Email: sylvie.bernaerts@thomasmore.be

Abstract

Background: Accumulating evidence supports the use of virtual reality (VR) in mental health care, with one potential application being its use to assist individuals with relaxation exercises. Despite studies finding support for the potential of VR to effectively aid in relaxation, its implementation remains limited outside of specialized clinics. Known barriers are insufficient knowledge regarding VR operation, lack of availability of VR relaxation apps tailored to local health care systems, and cost concerns. Unfortunately, many VR relaxation apps are designed exclusively for stand-alone headsets, limiting accessibility for a broad audience.

Objective: We aimed to design an accessible, smartphone-based VR relaxation app based on user preferences. This paper describes the assessment of 2 stand-alone VR relaxation apps and the resulting smartphone-based VR relaxation app design.

Methods: Overall, 30 participants (n=23, 77% women; n=7, 23% men) took part in 2 separate VR sessions, assessing 1 of the 2 VR relaxation apps (Flowborne and Calm Place) in each session. After each session, participants were presented with open-ended questions to assess their experiences via a web-based survey tool. These questions explored positive and negative features, shortcomings, and suggestions for improvements while also allowing space for additional remarks concerning the 2 VR relaxation apps. Three of the authors analyzed the responses using inductive thematic analysis, a process comprising 6 phases.

Results: Across both the apps, 5 recurring themes and 13 recurring subthemes were identified in the participants' answers: audio (music and sounds, guidance), visuals (content, realism, variation and dynamics in the environment), features (language, options, feedback and instructions, duration, exercise), implementation (technical aspects, cybersickness, acceptability and usability), and experience. We analyzed the participants' findings and conducted a literature review, which served as the basis for developing the app. The resulting app is a Dutch-language, smartphone-based VR relaxation app, with customization options including 3 types of relaxation exercises, 2 guiding voices, and 3 different environments. Efforts have been made to ensure maximum variation and dynamism in the environments. Calming music and nature sounds accompany the exercises. The efficacy and effectiveness of the resulting app design were not assessed.

Conclusions: This study provides insights into key features of VR relaxation apps, which were subsequently used for the development of a novel smartphone-based VR relaxation app. Further research concerning the effectiveness of this app, along with a broader evaluation of the efficacy and user feedback for smartphone-based VR relaxation apps, is needed. More generally, there is a clear need for more research on the impact of interactivity, biofeedback, and type of environment in VR relaxation.

(JMIR Form Res 2025;9:e62663) doi:10.2196/62663

XSL•FO
RenderX

KEYWORDS

smartphone-based virtual reality; virtual reality; relaxation; stress; user experience; mobile phone

Introduction

Background

Immersive virtual reality (VR) has a long-standing history in the mental health domain. Using computer-generated environments or 360-degree videos, patients experience being in the virtual environment, rather than the real world [1-3]. In therapy, VR boasts a clear-cut advantage by affording the therapist absolute control over the emotional intensity of the environment, thereby facilitating customization in response to the unique needs of the patient [4]. As VR becomes more accessible, it is being explored for a wide range of applications in mental health care. These include exposure therapy for patients with anxiety disorders [5] or posttraumatic stress disorder [6], distraction techniques for children to manage pain during burn wound treatment [3,7], body swapping for addressing eating and weight disorders [8], and using avatars with customized voices to treat patients experiencing auditory hallucinations [9]. In addition, VR is being explored as a tool to deliver or facilitate psychoeducation, behavioral activation, treatment for obsessive-compulsive disorder, addiction treatment, rehabilitation following traumatic brain injuries, and more [10,11].

Another potential application of VR is to help facilitate relaxation in individuals experiencing stress. VR relaxation apps typically present a computer-generated or 360-degree captured natural environment, such as a beach or forest, sometimes with accompanying auditory breathing or meditation exercises. The prevalence of natural environments in VR relaxation stems from research confirming their stress-reducing qualities [12]. Pizzoli et al [13] described 2 efficacious approaches to VR for promoting relaxation, namely, relaxing VR and engaging VR, and proposed a new approach, namely, personalized VR. Relaxing VR is characterized by visual content illustrating generic, pleasant environments, such as nature-based scenes, and auditory content based on common relaxation techniques, such as meditation and progressive muscle relaxation, with the aim of inducing control over one's body. In contrast, engaging VR uses more specific scenarios and interactive sceneries with a focus on empowering and training the user to regulate one's emotions [13]. Personalized VR was explored as a third, theory-based approach, focusing on an individualized design and implementation. Numerous studies have affirmed the effectiveness of VR relaxation in promoting relaxation and alleviating stress [14-19]. Recent reviews by Riches et al [20-22] also demonstrate that VR relaxation is an acceptable, feasible, and effective tool for enhancing relaxation and reducing stress across diverse populations, including the general population [20], various clinical populations [21], and in workplace settings [22].

Despite an increasing body of evidence highlighting the effectiveness of VR relaxation, its routine care implementation remains limited outside of specialized clinics [23]. Technology adoption already proves challenging in mental health care [24], but several barriers hamper the uptake of VR relaxation specifically. A first barrier is that many therapists or individuals experiencing stress frequently lack the knowledge of how to operate VR [25]. Furthermore, numerous VR relaxation apps are not adapted to or available in patients' native language. Another challenge is the fact that VR remains relatively expensive [11,26,27]. However, concerning the latter barrier, it should be noted that VR headsets exist in different price ranges. One distinction that can be drawn among VR headsets that rely on smartphones and those that can be used stand-alone. Smartphone-based VR headsets, as implied by their name, require a smartphone to function, whereas stand-alone VR headsets operate independently. The latter are frequently more sophisticated, although considerably more expensive than smartphone-based VR headsets. A smartphone-based VR headset is readily available with a starting price as low as €15 (US $15). Unfortunately, most relaxation apps are designed exclusively for stand-alone headsets, limiting accessibility for a broad audience. Notably, there is a third category, of computer-connected (or tethered) headsets. However, these are even more expensive than stand-alone VR headsets, especially because they rely on an additional powerful computer.

Objectives

The purpose of this study is to develop a relaxation app specifically for smartphone-based VR headsets. Such VR headsets allow people to use their own smartphone, making them a potentially cheaper and more accessible alternative to stand-alone VR headsets. However, it is crucial to acknowledge that owning a smartphone is a prerequisite for using a smartphone-based VR headset. Fortunately, most of the population already owns a smartphone, making this cost less of a barrier. To overcome the previously highlighted barrier of inadequate adaptation to the local health care sector and language, the app is tailored to the local (ie, Flemish) context and developed in the Dutch language.

Methods

Overview

This study is 2-fold. In the first part, we assess the user's perspectives on existing VR relaxation apps using evidence-based techniques. In the second part, we describe the development of a smartphone VR app, considering previous user feedback and relevant literature. The validation of the resulting app design is beyond the scope of this study.

Study Design and Procedure

A crossover study to assess users' perceptions of 2 VR relaxation apps was conducted at Thomas More University of Applied Sciences, Belgium. Participants took part in 2 separate VR sessions, exploring 1 of the 2 VR relaxation apps per session. After each VR session, participants were presented with open-ended questions to assess their experiences via the web-based survey tool QuestionPro (QuestionPro, Inc). These questions explored positive and negative features by assessing

XSL•FO
RenderX

strengths (What did you like about the app?), shortcomings (What did you like less or dislike about the app?), and suggestions for improvements (What would you improve about the app?), while also allowing space for additional remarks concerning the 2 VR relaxation apps. Before and after each session, quantitative data concerning relaxation levels, positive and negative affect, attitude toward technology, and mood were also collected. However, these data are beyond the scope of this study. Each VR session lasted approximately 5 to 10 minutes and took place on 2 different days, with at least 1 week in between. The order in which the apps were presented was randomized. Each study session lasted approximately 40 minutes.

Ethical Considerations

The study design and consent forms were approved by the Ethical Committee of the Department of Applied Psychology at Thomas More University of Applied Sciences (ECTP2223_06). Before data collection, participants were informed about the study objectives, study design, and data processing both verbally and through an information letter. They also provided written informed consent. Participants were not compensated for their participation in the study. Data were collected, handled and stored according to the European Union General Data Protection Regulation (GDPR) and International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) Guideline for Good Clinical Practice (GCP). Data were deidentified and all analyses adhered to data privacy guidelines.

Participants

A total of 30 employees (n=23, 77% women; n=7, 23% men) from Thomas More University of Applied Sciences were recruited through a call in an internal newsletter between December 2022 and January 2023, across 2 campuses in Antwerp and Geel, Belgium. The call provided information about the study objectives and study procedures. Interested employees were asked to contact the researchers via email for additional information and to schedule the data collection. Their average age was 37.17 (SD 8.91) years. Individuals with neurological conditions, such as epilepsy, or severe neck problems were excluded. There were no dropouts.

Intervention

Overview

Participants were presented with 2 freely available VR relaxation apps via the Meta Quest 2 stand-alone VR headset (Meta Platforms, Inc), Flowborne (Flowborne) [28] and Calm Place (Mimerse) [23]. It is important to note that Calm Place is not publicly available, but the creators gave permission to use the app for research purposes.

Flowborne

Flowborne is a freely available respiratory biofeedback VR app designed to promote diaphragmatic breathing in a playful way. The app was developed in an academic context, and previous research has demonstrated its effectiveness in improving diaphragmatic breathing and breath awareness, promoting a sense of relaxation [29]. Flowborne consists of a low-polygon 3D environment, a digitally created space that uses fewer geometric shapes to ensure a smoother and faster VR experience, featuring natural elements such as grass, stones, flowers, and trees. The environments are composed of 4 different landscapes, each containing 6 different levels with an additional variation in duration (ie, short, medium, and long). The 4 different landscapes vary in terms of color themes and environmental characteristics. For instance, one of the landscapes resembles a hill, while another resembles an ocean. What makes this VR relaxation app stand out is the ability for users to navigate through the environment by breathing, accompanied by visual alterations in response to their breath (eg, grass growth and color shifts). During exercises, users are required to position the controller of the VR headset on their lower abdomen, allowing the controller to detect the expansion and contraction of the abdominal region and promote abdominal breathing. A slow, deep breathing pattern is necessary to progress fluently through the environment. There is no guiding voice throughout these exercises. At the outset, individuals have the opportunity to select a tutorial session that explains the exercise and the use of the controller on their abdomen. In addition, users experience ethereal and natural sounds during their journey in the environment. For this study, participants were instructed to complete a session of medium duration (ie, 9 minutes) with a landscape of their choice.

Calm Place

Calm Place is a VR relaxation app developed by Mimerse. Although this company is no longer operational, it has made its app available for research purposes upon request. In this VR relaxation app, users immerse themselves in a low-polygon natural environment, at a campfire surrounded by nature. During the exercise, one has the option to transition the environment from daytime to nighttime. The app uses a gaze-controlled interface, enabling users to navigate the environment by simply moving their heads. By directing their gaze toward specific objects, users can induce changes in the environment [23]. The exercises can be done without controllers; they are only needed to set up the exercises and customize the environment to individual preferences. This customization includes selecting whether it rains, choosing whether there are animals, setting a specific time to prevent a transition from day to night, adjusting sound volume, enabling or disabling environment sounds, choosing among 3 songs, and selecting which “relaxing object” you prefer. Exercises include guided relaxation, mindfulness exercises, or simply observing nature. In both the guided relaxation and mindfulness exercises, there is a voice that guides you step by step through the process. In the guided relaxation exercises, you can choose from different relaxation techniques: basic relaxation, muscle relaxation, and deep breathing. Similarly, in the mindfulness exercises, you can specify the session number you prefer to follow. There are also 4 different landscapes and 3 different duration options (ie, 9, 13, or 20 minutes). At the outset and conclusion, participants are asked to report on their emotional state (“How are you feeling now?”), using a scale featuring 5 emojis. For this study, participants were instructed to perform a 9-minute deep breathing exercise and were given the freedom to choose their preferred environment.

XSL•FO
RenderX

Data Analysis

Using inductive thematic analysis, qualitative data were analyzed by 3 team members (AV, MvL, and SB) for each app separately. The goal was to map out users' perspectives regarding strengths, weaknesses, and potential improvements in the 2 VR relaxation apps. Although there is no singular definitive method for conducting thematic analysis, Braun and Clarke [30] have outlined a potential procedural framework consisting of 6 phases: familiarization, coding, theme development, reviewing themes, defining and naming themes, and writing the report. In the familiarization stage, the responses to the open-ended questions underwent repeated readings by each of the 3 coders individually to attain a profound understanding. Subsequently, in the second phase, a brief summary or code was provided by each of the 3 coders individually for what was said in the responses. The data were presented in Excel (version 365; Microsoft Corp) and were open-coded through semantic coding, and no predetermined codebook was used. For instance, codes such as "more realistic" were assigned to responses such as the following:

> *The images are very "virtual." I think I would find it more pleasant if real images of mountains were used.*

Each coder individually went through multiple iterations of coding. However, these iterations and the corresponding changes and evolution of the codes were not documented. During the third stage, the 3 coders individually searched for overarching themes and subthemes and categorized the answers to the open-ended questions using colors associated with these initial themes and subthemes. Note that, in this respect, although an inductive thematic analysis approach was adopted, deductive analysis was necessary to some extent to ensure that the codes and themes contributed to answering the research questions, which is common in thematic analysis [31]. In the fourth stage, a comparison was made among the identified themes and the data to check whether these themes constitute an accurate representation of the data and aid in addressing the research question. In this phase, coders MvL, AV, and SB compared themes and codes to reach consensus about the codes and themes and subthemes. When similar codes and themes and subthemes were present, the clearest (ie, based on consensus) terminology was withheld. When codes and themes and subthemes differed, the researchers discussed until consensus was reached. These discussions were not documented. Official names were assigned to the themes in the fifth phase, which was led by AV. These tables were subsequently checked by SB, and the theme and subtheme structure was discussed until consensus was reached. MvL was not able to partake in this stage due to practical reasons. This thematic analysis revealed that participants' responses to the open-ended questions could be categorized according to the following themes across the 2 VR relaxation apps: audio, visuals, features, implementation, and experience. Each of the themes has been further subdivided into different subthemes. These themes and subthemes were reported in the sixth phase and translated into specific app features, with their respective findings elaborated upon in the Results section of this study.

Results

In the first part of this section, we discuss user perspectives on existing VR relaxation apps, categorized according to the aforementioned themes and subthemes. In the second part, we describe the translation from user preferences to app features and the resulting smartphone VR app, considering this user feedback and relevant literature.

Stage 1: Assessment of User Perspectives

Results of the thematic analyses for each separate app can be found in Tables 1 and 2. The corresponding codes and themes and translation to app features are discussed in stage 2.

XSL•FO
RenderX

Table 1. Summary of users' perspectives on the Flowborne app, organized according to the themes and subthemes from the thematic analysis.

Themes and subthemes	Strengths	Areas of improvement and negative features
Audio		
Music and sounds	• Music found to be calming (n=14, 47%) • Music mentioned spontaneously when asked for positive aspects of the app (n=9, 30%) • Nature sounds (n=2, 7%)	• Suggestion to add more nature sounds (n=1, 3%)
Guidance	• Lack of guiding voice (n=3, 10%)	• Preference for a guiding voice (n=1, 3%)
Visuals		
Content	• Beautiful setting (n=10, 33%) • Magical setting (n=1, 3%) • Animals (n=4, 13%)	• Desire for more animals (n=3, 10%) • World feels deserted, too little life (n=1, 3%)
Realism	• Realistic virtual setting (n=1, 3%) • Magical environment (n=1, 3%)	• Preference for more realistic environment (n=3, 10%) • Artificial environment (n=2, 7%) • Futuristic animals and elements (n=1, 3%)
Variation and dynamics in the environment	• Appreciation for existing variety (n=2, 7%) • High activity level in the environment (n=1, 3%)	• Desire for more variety (n=3, 10%) • Desire for more dynamism (n=2, 7%) • Distraction due to environmental activity (n=1, 3%)
Features		
Language	—[a]	—
Options	• Customizable duration (n=1, 3%)	• Desire for customizable route (n=1, 3%)
Feedback and instructions	—	• Unclear instructions and desire for example exercise (n=3, 10%) • Unclear feedback at the end (n=2, 7%) • Lack of reference for breathing pace (n=1, 3%) • Unclear end of exercise (n=1, 3%)
Duration	—	• Duration too short (n=2, 7%) • Negative feedback on exercise duration (n=1, 3%)
Exercise	• Movement through breathing (n=7, 23%) • Visual feedback linked to breathing (n=3, 10%)	• Lack of forward movement while inhaling (n=3, 10%) • Tendency to exhale more to move quickly through the environment (n=1, 3%)
Implementation		
Technical aspects	—	• Negative feedback on configuration of walking path (n=3, 10%) • Disappointment with the configuration of the environment boundaries (n=2, 7%) • Controller as a suboptimal breathing sensor (n=1, 3%) • Occasional image blurriness (n=1, 3%) • Occasional visual glitches (n=1, 3%)
Cybersickness	—	• Nausea (n=2, 7%) • Dizziness (n=2, 7%) • Nausea and headache (n=1, 3%)
Acceptability and usability	• Intention to use it again (n=1, 3%)	• Sense of an abandoned world (n=1, 3%)
Experience		
	• Pleasant experience (n=5, 17%) • Relaxing effect (n=5, 17%) • Energizing effect (n=1, 3%) • Interesting as a sleep aid (n=1, 3%) • Desire to keep wearing the headset (n=1, 3%)	• Competitive aspect hinders relaxation (n=1, 3%) • Uncertain about repeated enjoyment (n=1, 3%) • Difficulty staying awake (n=1, 3%)

[a]None.

XSL•FO
RenderX

Table 2. Summary of users' perspectives on the Calm Place app, organized according to the themes and subthemes from the thematic analysis.

Themes and subthemes	Strengths	Areas of improvement and negative features
Audio		
Music and sounds	• Positive feedback on nature sounds (n=9, 30%)	• Desire for more nature sounds (n=2, 7%) • Suggestion to add music (n=1, 3%)
Guidance	• Pleasant, relaxing voice (n=12, 40%)	• Desire for continuous guiding voice counting during breathing exercise (n=4, 13%) • Desire for less auditory guidance (n=3, 10%) • Desire for more interaction with guiding voice (n=1, 3%) • Guiding voice should not be too demanding (n=1, 3%)
Visuals		
Content	• Environment (n=12, 40%) • Campfire (n=2, 7%) • Northern lights (n=2, 7%) • Animals (n=1, 3%) • Beach and rocks (n=1, 3%) • Vacation feeling (n=1, 3%)	—[a]
Realism	• Realistic environment (n=1, 3%)	• Desire for greater realism in the environment (n=4, 13%)
Variation and dynamics in the environment	• Day-night transition (n=5, 17%) • Variation in the environment (n=4, 13%) • Dynamic elements (n=1, 3%) • Rain (n=1, 3%)	• Rain and thunder found to be unpleasant (n=5, 17%) • Static environment (n=1, 3%) • Monotonous environment (n=1, 3%) • Desire for more dynamic elements (n=1, 3%) • Desire for more variation (n=1, 3%)
Features		
Language	—	• Desire for the Dutch-language option to lower the threshold for certain target groups (n=3, 10%)
Options	• Ability to choose different environments (n=2, 7%) • Curiosity regarding the other environments (n=2, 7%)	—
Feedback and instructions	—	• Suggestion for better preparation or a test exercise (n=1, 3%) • Suggestion for instruction on appropriate posture (n=1, 3%) • Suggestion for option without indicating feelings at the end (n=1, 3%)
Duration	• Satisfaction with duration (n=2, 7%) • Positive feedback on customizable exercise duration (n=1, 3%)	• Exercise duration found to be short (n=1, 3%)
Exercise	—	• Desire for clearer indication of breathing rhythm (n=3, 10%) • Desire for visual cues for counting during breathing exercise (n=2, 7%) • Discomfort expressed regarding the fixed breathing pace (n=2, 7%) • Too many tasks at once (n=1, 3%)
Implementation		
Technical aspects	• Appreciation for minimal controller use (n=1, 3%)	• Controller visible at times (n=3, 10%) • Occasional visual glitches (n=3, 10%) • Difficulties with controller clicking (n=1, 3%)
Cybersickness	—	—

XSL•FO
RenderX

Themes and subthemes	Strengths	Areas of improvement and negative features
Acceptability and usability	—	• Distraction due to background noise (n=2, 7%) • Doubts about virtual reality relaxation's superiority over breathing exercises with only visual or video stimuli (n=1, 3%) • Suggestion to use headphones (n=1, 3%) • Suggestion to use the app while standing (n=1, 3%)
Experience		
	• Pleasant experience (n=10, 33%) • Relaxing effect (n=4, 13%) • Easy to imagine being in such an environment (n=1, 3%) • Minimal sensory stimuli (n=1, 3%)	• Boring experience (n=1, 3%) • Experience found to be tiring (n=1, 3%) • Fear due to the dark environment (n=1, 3%)

[a]None.

Stage 2: Translation of User Perspectives to App Features

Overview

A new VR relaxation app, Immersive Mental Health (IMH; Aeroplane XR), was designed for smartphones, incorporating insights from the user assessments of the 2 stand-alone VR apps and guided by a comprehensive literature review. The app was developed by the Flemish XR developer Aeroplane XR using the Unity platform. The app's conceptualization is presented in Figure 1. In this section, we describe the developed smartphone VR app, considering user feedback and relevant literature.

Figure 1. Personal configuration flow in the Immersive Mental Health app. The language in the screenshots has been translated from Dutch to English for publication purposes.

Audio

Music and Sounds

Numerous participants reported finding ethereal and nature sounds pleasant and conducive to relaxation. Studies have shown that the concurrent exposure to a nature-based VR setting and natural sounds is more effective in reducing stress compared to independent auditory or visual inputs [17,32]. Drawing upon participant evaluations and literature, we chose to integrate both natural and ethereal sounds into our new IMH app. Different nature sounds, such as chirping birds, the sound of the sea,

XSL•FO
RenderX

crackling fire, wind, and others, can be heard depending on the environment chosen.

Guidance

Opinions on the use of a guiding voice were divided, both regarding whether it should be included at all and whether it should provide continuous guidance during breathing exercises. Some participants preferred a continuous vocal count throughout the breathing exercise, stating the following:

> *I would rather hear the voice continuously guiding the breath exercise with simple words like "in and out." Counting to 10 myself kept me more focused on counting than on my breathing.*

In contrast, others preferred its absence:

> *Hearing the guiding voice count aloud the duration of breaths made me more anxious because my pace was slower than indicated, making me afraid of not completing on time.*

IMH does not require a counting voice during breathing exercises; consequently, such exercises are not included. We opted to include a guiding voice for 2 out of the 3 exercises to provide auditory explanations for the exercises, as participants have articulated a preference for clear instructions. In addition, participants were given the option to choose between a male or a female narrator. We included this customization option because the literature has consistently shown that providing personalization options leads to a higher level of relaxation [33].

Visuals

Content

A substantial number of participants found the environments to be pleasant:

> *Lovely environment. Feels like a vacation.*

Animals and natural features such as fire and water were especially well received. Literature has also shown that natural lighting and water elements are essential for alleviating stress in VR users [34]. Grounded in the findings of this study and the evaluative study, the IMH app features natural landscapes enriched with elements such as water, fire, and animals.

Realism

Participants valued a realistic environment, often suggesting that the environments could be even more realistic:

> *The images are very "virtual." I think it would be more pleasant if real images of mountains were used.*

According to a review study, VR natural scenarios have similar effects to real nature scenarios for stress reduction [34]. However, they highlight the importance of maximizing realism in these VR scenarios to improve the immersive experience. Furthermore, research indicates that a computer-generated natural environment provides a more emotionally restorative experience compared to a 360-degree video representation of a similar setting [35]. In addition, another study found that computer-generated VR enhances the sense of presence and elevates mood more effectively than 360-degree VR [36]. Therefore, based on the literature and participant feedback, we aimed to create as realistic a computer-generated environment as possible within the limits of the processing power of a smartphone (ie, Android operated).

Variation and Dynamics in the Environment

The participants placed great importance on variation. They found it pleasant when there were surprising elements, and the day-night transition in Calm Place was positively assessed:

> *It's nice that a day passes during these nine minutes. You see the sun first, then the stars, and then the sun again.*

In addition, more participants found rain and thunder unpleasant than those who found it enjoyable. Participants also expressed a preference for dynamic elements in the environment. Consequently, we integrated these considerations into our IMH app, featuring 3 distinct natural environments, each encompassing diverse fauna, natural and dynamic elements, and day-to-night transitions.

Features

Language

Some participants believed that using an app in a language other than the user's native language could pose a barrier. Therefore, the IMH app was developed in the Dutch language.

Options

Participants appreciated the option to select diverse environments and varied durations. Moreover, as noted earlier, the literature consistently demonstrates that offering customization choices results in increased levels of relaxation [33]. Therefore, we decided to incorporate a wide array of customization options. Within the IMH app, users could choose from 3 distinct environments: a tropical beach, a winter scene with the northern lights, or a mountainous landscape. Furthermore, users had the opportunity to select among 3 specific exercises: free exploration, a mindfulness exercise, or progressive muscle relaxation. In addition, participants had the option to choose between a male or female guiding voice during the mindfulness and progressive muscle relaxation exercises.

Feedback and Instructions

The participants expressed a preference for detailed instructions, a consideration we took into account in the development of our new IMH app. Before the exercises began, clear instructions were provided (Figure 1). First, participants were instructed to sit in a comfortable position. They were also prompted to horizontally rotate their smartphone in the direction indicated by an arrow. Subsequently, participants were asked to insert their smartphone into the VR headset. Clear instructions were also provided during the exercises: a guiding voice offered step-by-step guidance in 2 exercises. In addition, many participants indicated that they would have preferred to receive clear feedback at the end. However, the IMH app does not provide any feedback. Some participants also indicated wanting to try a sample exercise before starting the actual session:

> *At first, I was not sure if I was making progress. Perhaps add a small test before the meditation begins to check if you're doing it correctly.*

XSL•FO
RenderX

One of the 3 exercises in the IMH app involves simply exploring the environment without any specific tasks. This can serve as an introduction to familiarize users with the surroundings and the headset. A tutorial for the exercises is not provided in the IMH app, as the step-by-step instructions from the guiding voice eliminate the need for it.

Duration

During the evaluative study, certain participants expressed satisfaction with the presence of diverse duration options for the exercises. However, this feature was not integrated into the IMH app. Instead, users can exit the exercise whenever they desire. The participants also appreciated having a clear ending. Therefore, the conclusion is explicitly marked, as noted by the following statement: “your session is now complete.”

Exercise

Many participants found navigating the environment through breathing enjoyable in Flowborne. However, there were some remarks regarding the limitation that forward movement only occurred during exhalation:

> *You should move forward during both inhalation and exhalation. Currently, you stand still while inhaling and move forward while exhaling. Right now, it feels more like a stop-and-go. It would be better in a continuous flow.*

The visual feedback based on breathing, such as the movement of dots and squares, was also appreciated. Certain participants found it unpleasant that a predetermined breathing pace was enforced, whereas others indicated a preference for clearer indications of the breathing pace. In the end, we made the decision to not include breathing exercises in the IMH app. We selected 2 exercises that have also been demonstrated to be effective in reducing stress: progressive muscle relaxation and mindfulness [36,37].

Implementation

Technical Aspects

Several issues were encountered with the controllers. Some participants observed the controllers persistently on the screen, while others struggled with their operation. Participants found it favorable when minimal controller use was required:

> *Convenient and pleasant that you don’t have to use the controller system all the time, ensuring the focus is on relaxation.*

There were also occasional challenges with setting up the play area’s boundaries. The IMH app uses smartphone-based VR headsets, which eliminate the need for a defined play area and do not require controllers. Furthermore, participants sporadically noted slight blurriness at the environment edges, an issue preventable by properly inserting the smartphone into the VR headset. In both the apps, it was also noted that errors in the environment (eg, occasional collisions with a rock or the campfire appearing almost on one’s lap) are detrimental to the perceived realism of the environment, bringing individuals back to the realization that they are not in a real world:

> *Sometimes, the plants move through a stone or through the fire, and it brings you back to the awareness that you are not completely in a real environment, but in a virtual one.*

In response, the development of the IMH app focused on achieving the most realistic computer-generated environment feasible, while considering the technical constraints.

Cybersickness

In Flowborne, some participants experienced nausea due to a discrepancy between the movement in the VR headset and their actual physical movements. A participant expressed dissatisfaction in the following manner:

> *After the test, I felt sick and had a headache, which had an unpleasant impact on my body. To cope with the nausea and dizziness, I began to breathe less through my belly, aiming to move forward at a slower pace.*

This phenomenon, known as cybersickness in the literature, results from an inconsistency between the sensory inputs [11]. The design of the IMH app ensures that there is no forward motion within the VR environment while the user is seated in a chair in reality, thereby reducing the risk of cybersickness within the limits of the processing power of a smartphone (ie, Android operated).

Acceptability and Usability

Some participants found themselves distracted by background noises, leading 1 individual to propose the use of headphones. Moreover, 1 individual questioned whether the app offers any added value compared to non-VR relaxation apps:

> *I’m not sure if this app offers a significant advantage over regular apps that indicate the rhythm of breathing through figures or sounds.*

In addition, 1 person suggested to use the app in a standing position. Yet another participant described the environment as somewhat abandoned. Another individual expressed a willingness to use Flowborne again.

Experience

Certain participants reported feeling more relaxed after having used the VR relaxation apps. Overall, participants found the experience pleasant. However, responses did vary, with some participants reporting feelings of tiredness, others experiencing a sense of competitiveness, and some feeling energized, while 1 participant found the experience boring. Despite these diverse reactions, it is important to note that the primary aim of the IMH app is to facilitate relaxation for its users.

Discussion

Principal Findings

This study investigated users’ opinions on 2 existing stand-alone VR relaxation apps and demonstrated how, based on those insights, a new smartphone-based VR relaxation app was developed. The aim of this new app is to overcome the barriers that currently limit the use of VR for relaxation, thereby making VR relaxation accessible to a wider audience. These barriers

XSL•FO
RenderX

include insufficient knowledge regarding VR operation, lack of availability of VR relaxation apps in the appropriate language for local health care systems, and cost concerns.

The evaluative study revealed 5 predominant themes that consistently recurred in the participants' evaluations: audio, visuals, features, implementation, and experience. We evaluated the main findings of the participants for each theme and incorporated them into the development process of our new IMH app. This app is a Dutch-language, smartphone-based VR relaxation tool and is freely available for both Android [38] and iOS [39]. It is designed to enable users to perform relaxation exercises within a computer-generated nature setting, accompanied by natural and ethereal auditory backgrounds. The app has been developed to incorporate a high degree of variability and dynamic features within these environments. Moreover, the design ensures minimal cybersickness by eliminating movement within the virtual environment while the user remains physically still. The IMH app has multiple customization options, as research has shown that they can enhance relaxation levels [33]. Primarily, users have the option to select from 3 exercises: 1 free exploration without auditory guidance and 2 auditory-guided exercises, involving mindfulness or progressive muscle relaxation. Our decision to offer "free exploration" as an option aims to introduce users to the VR environment, while the choice of progressive muscle relaxation and mindfulness is based on scientific evidence demonstrating their effectiveness in stress reduction [36,37]. Second, users can choose between a male or female guiding voice for the mindfulness and progressive muscle relaxation exercises. The decision to incorporate a guiding voice stemmed from participants' desire for clear instructions during the exercises. A third customization option allows users to select from 3 distinct environments: a tropical beach, a winter landscape featuring the northern lights, or a mountainous scenery.

Comparison With Other Smartphone-Based VR Relaxation Apps

To the best of our knowledge, there exist a few other smartphone-based VR relaxation apps, although they are relatively scarce. One such app is the VR-based relaxation self-training app, which focuses on self-training for relaxation using diaphragmatic breathing and progressive muscle relaxation [40]. Users initiate their training by learning relaxation techniques in a computer-generated clinic, subsequently reviewing these techniques in a comfortable environment (eg, on a beach) and finally practicing in various computer-generated settings (eg, in an elevator). Researchers have also conducted an effectiveness study comparing this app with traditional learning methods using text and illustrations, finding that the VR app is particularly effective for progressive muscle relaxation exercises, with significantly lower tension levels observed compared to a control group using worksheets. Serenity is another gamified smartphone-based VR app developed to reduce stress among university students [41]. It features a traveler role-playing game format where the user visits various vacation spots under different settings, including day and night. The game allows players to trigger special events by collecting items along paths, with these items being selectable through eye-tracking technology. The effectiveness of this immersive experience was monitored through physiological sensors, with positive outcomes particularly among students with lower stress levels. Third, Gaia VR merges neurofeedback with biofeedback in a meditation app, using feedback from a heart rate band and an electroencephalogram headband to tailor the VR experience based on the user's measured stress levels [42]. The user continuously wears the heart rate monitoring band throughout the day, enabling the app to suggest appropriate meditation techniques in response to the monitored stress levels. During the meditation sessions in VR, users can influence the virtual environment through their meditation-induced brainwave patterns, creating a dynamic and personalized meditation experience. This app has been tested thus far as a proof of concept; however, a more comprehensive evaluation is required to determine its long-term effectiveness and its applicability to a broader audience.

In summary, these 3 smartphone-based VR relaxation apps feature unique approaches and designs, differing from each other and from our IMH app, thereby demonstrating the extensive range of options available for inducing relaxation via smartphone-based VR. A key feature of the IMH app is its strong emphasis on both accessibility and customizability. As implied by its name, the Virtual Reality-Based Relaxation Self-Training app focuses mainly on the training of relaxation techniques. In contrast, the Serenity app targets students and uses gamification for stress reduction. The Gaia VR app primarily focuses on the integration of neurofeedback and biofeedback, providing real-time adjustments based on the user's physiological state. Unlike the other 3 apps, the IMH app was developed specifically using user feedback. Additional research is needed to determine the most effective design for relaxation and to identify user preferences in smartphone-based VR relaxation.

Comparison With Previous User Experience Studies

Although the effectiveness of VR for relaxation has been demonstrated in controlled research settings, insights into user experiences and preferred characteristics of VR relaxation apps remain sparse. It is important to acknowledge that, as far as we know, no user feedback studies specifically address smartphone-based VR relaxation. On the basis of 2 studies focusing on user experiences with VR for relaxation, although not specifically on smartphone-based VR, several aspects that users value can be highlighted [19,43]. Primarily, a study focusing explicitly on mindfulness VR demonstrated that sense of presence is a key factor [43]. Presence, which refers to the feeling of "being there" in a simulated environment, creates an "illusion of reality" where users behave as if the environment were real, despite being computer generated. Participants reported that this sense of presence within this virtual setting facilitated their mindfulness practices. Moreover, users noted that interruptions within the virtual environment could be distracting, leading to a reduction in the quality of the relaxation experience.

Second, users appreciated the integration of game elements into the VR relaxation experience. Features such as performance badges, use metrics, and leaderboards are appreciated and can therefore enhance the overall user experience and increase

XSL•FO
RenderX

engagement. Nonetheless, the extensive diversity in the design and implementation of these gamification elements in different apps complicates the process of determining which elements are most preferred.

Third, the potential for personalizing the experience is a key factor for users. This includes choosing the computer-generated setting, adjusting the relaxation exercises provided, and selecting the background music, all contributing to a personalized and potentially more efficacious user experience.

Fourth, participants indicated that technical aspects, such as video resolution and the weight of the VR headset, are sometimes perceived as disruptive. They specifically highlighted that issues with blurred images and areas of low resolution within the video frequently led to distractions and adversely affected their experience. In summary, the 2 studies showed that users generally have a positive opinion of VR relaxation apps [19,43]. The findings demonstrated that VR relaxation app users value an experience that seems real, has gamification elements, can be customized, and runs well without glitches. These findings are in line with what participants mentioned in our evaluative study.

In developing the IMH app, we aimed to align closely with these preferences, although gamification elements were not integrated into our app. Another interesting study developed a framework for designing a virtual environment for stress therapy [44], listing 5 factors (ie, attention, attraction, environment setting, environment exploration, and interaction) containing 14 design elements: fascination stimuli, nature - based theme, supporting audio, daylight lighting, realism, free exploration, predefined viewpoints, predefined path, passive interaction, active interaction, safety features, mind preparation, social support, and visual clarity. The IMH app meets 11 out of the 14 design elements, only lacking predefined paths, active interaction, or social support.

Limitations

Several limitations of this study warrant attention. First, obtaining insights from professionals highly experienced in conducting relaxation sessions would have added depth. Second, our participant group was relatively young, which could hinder the generalization of our results to the general population. Therefore, it would be valuable to examine the opinions of older adults concerning VR relaxation apps as well. Third, it would have been intriguing to involve the participants not only in the evaluative study but also in the development process of the new app [45]. A systematic review on cocreating digital mental health interventions indicated that involving various stakeholders, such as health care professionals and patients, in the development process contributes to the creation of culturally sensitive and widely accepted interventions [46]. Fourth, our evaluation study focused on 2 stand-alone VR relaxation apps, rather than smartphone-based VR relaxation apps. It would have been interesting to obtain user feedback specifically on smartphone-based VR relaxation apps. Finally, concerning the methodology, coding iterations and the corresponding changes and evolution of the codes were not recorded, and discussions among the 3 coders about codes and themes were also not documented. Therefore, no interrater reliability could be calculated to assess the consistency of the different researchers' coding. In this respect, it is not possible to quantitatively assess the validity and generalizability of the thematic analysis. However, we have described the process of thematic analysis in detail to ensure openness about the process.

Strengths

We developed a new smartphone-based app, grounded in both an evaluative study and a review of the literature. We believe this approach substantiates a robust methodology for app development. Furthermore, to the best of our knowledge, there are few smartphone-based VR relaxation apps available, suggesting that our development has addressed a significant gap in the field. Moreover, we hypothesize that our IMH app could also be used as a method or tool to teach relaxation techniques. For instance, individuals may, after a few sessions of VR relaxation involving progressive muscle relaxation, potentially practice these techniques without VR.

Future Directions

First, it is important to note that the presented app design, based on the assessed user preferences, has not yet been validated. This app design is an example of what a smartphone VR app for relaxation purposes could look like. In addition, future research should also focus on the efficacy and effectiveness of smartphone-based VR for relaxation. We have deliberately chosen to develop an app specifically for smartphone-based VR headsets to make VR relaxation accessible to a wider audience. Despite being more affordable and user-friendly than high-cost stand-alone VR systems, smartphone-based VR has its limitations, including a narrower field of view, lower resolution and frame rate, and less sophisticated optics and positional tracking. Nonetheless, a comparative study has not demonstrated significant differences in spatial presence, usability, cybersickness, satisfaction, workload, and learning outcomes between stand-alone VR headsets and smartphone-based VR in educational contexts [47]. Despite its limitations, smartphone-based VR might therefore still be able to deliver a satisfactory level of immersion [48]. To the best of our knowledge, no specific study has yet been conducted to explore the differences between smartphone-based VR and stand-alone VR specifically for relaxation purposes.

Second, additional investigation is required to examine other potentially engaging environments apart from natural settings, as well as to determine which particular aspects of nature are optimal for VR relaxation. In the design of the IMH app, the decision was made to include 3 different natural environments as options. This choice was driven by evidence from review studies showing that VR exposure to natural settings may lead to reductions in stress levels [34,49]. However, there is uncertainty in the literature regarding which type of VR environment is most effective for relaxation. Some researchers argue that the research has not adequately examined the diverse array of environments that can be used to promote relaxation in VR [50]. Besides urban computer-generated settings, only a handful of nonnatural alternatives have been extensively examined for VR relaxation. Consequently, a study was conducted to determine whether abstract environments could also serve as viable settings for VR relaxation. Results indicated

XSL•FO
RenderX

that the effects on perceived restoration, including sensations of being away, experiencing soft fascination, and sustaining interest, were comparable across both abstract and natural environments, pointing to a similar restorative capacity in the abstract computer-generated environment [50]. According to another review, there is a lack of understanding regarding the differentiation between green and blue virtual natural environments [51]. In addition, research has been conducted on the impact of biodiversity on stress recovery within VR settings [52]. Researchers found that stress recovery was most effective in VR environments characterized by low biodiversity. Overall, further research is needed to explore alternative engaging environments beyond natural settings and to identify the specific elements of nature that are most effective for VR relaxation.

Third, further research is needed to better understand the potential advantages of interactivity in VR relaxation, particularly within the context of smartphone-based VR. Within the literature, there is debate regarding the role of interactivity in VR relaxation environments. The underlying principle of such interactivity involves engaging the user in specific tasks within the computer-generated environment, such as gaming or art creation, which are aimed at providing distraction and subsequent stress relief [53]. It is also suggested that interactive VR serves as an interesting and motivating tool, enhancing engagement among participants [54]. A review paper concluded that the effectiveness of interactivity in VR relaxation depends on the type of game used; interactive may surpass noninteractive VR relaxation in certain contexts [51]. The IMH app was designed without interactive features, partly because the typical constraints of smartphone-based VR, such as the lack of physical controllers, make interactivity less straightforward. Notably, feedback from our participants during evaluations indicated a preference for the absence of controllers, suggesting that this feature increases the usability of VR relaxation, thereby reducing barriers to adoption. This reflects the necessity of balancing considerations such as the potential advantages of interactivity with the ease provided by nonuse of controllers, demonstrating the importance of tailoring VR relaxation to individual preferences.

Further research is needed to explore the advantages of interactive features in VR relaxation, especially within smartphone-based VR apps. Exploring other mechanisms for interaction in smartphone-based VR, aside from controllers, such as head-based movements and voice control, could also be considered. Currently, our IMH app does not incorporate biofeedback, but integrating such a feature in the future could offer substantial advantages. Interventions that include biofeedback sensors with real-time monitoring of, for example, respiration, neural activity, heart rate, and skin conductance represent a promising advancement. This biological information can be communicated to the user via dynamic environmental changes, including alterations in music, sound, color, lighting, the appearance of objects, and animations, to facilitate the detection and regulation of internal states [50]. The combination of VR with biofeedback presents multiple benefits. For instance, VR can improve the visual appeal and salience of the feedback stimuli, which may, in turn, foster greater motivation and engagement. In our evaluative study, some participants also expressed appreciation for the visual feedback based on their breathing, such as the movement of dots and squares. In addition, biofeedback may enhance trainees' belief in their capacity to relax both their body and mind, given that achievements in displaying healthy target behaviors or skills are instantly fed back and visible. Nevertheless, a significant obstacle is that most implementations of VR biofeedback demand considerable expertise and effort. Expensive physiological sensors must be purchased and integrated into the VR system [29]. Despite these challenges, it remains an intriguing avenue to explore. It would be interesting, for example, to develop a smartphone-based VR relaxation app that is linked to a smartwatch, which would provide feedback to the app.

Conclusions

In this study, a new smartphone-based VR relaxation app was designed to address the current challenges in implementing VR relaxation technologies. The design of this app draws from an evaluation study on 2 existing stand-alone VR relaxation apps, supplemented by an extensive literature review to determine key attributes of VR relaxation apps. The resulting IMH app is a Dutch-language, smartphone-based VR app featuring a variety of customization options, namely, 3 different relaxation exercises, options for male or female guiding voices, and 3 unique environments. The environments are designed to be dynamic and varied and are accompanied by soothing music and natural sounds. The app design itself was nevertheless not validated in this study. Therefore, future research should validate the resulting design and explore the effectiveness of this app, along with a broader evaluation of its efficacy and user experience, to further build on the potential of VR relaxation apps.

Acknowledgments

This study was conducted as a part of the Immersive Mental Health Project funded by the Flanders Innovation & Entrepreneurship TETRA program (HBC.2021.1023). We used generative AI (ChatGPT; OpenAI) to check sections of our writing and optimize spelling and grammar.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Velana M, Sobieraj S, Digutsch J, Rinkenauer G. The advances of immersive virtual reality interventions for the enhancement of stress management and relaxation among healthy adults: a systematic review. Appl Sci 2022 Jul 20;12(14):7309. [doi: 10.3390/app12147309]
2. Bernaerts S, Bonroy B, Daems J, Sels R, Struyf D, Gies I, et al. Virtual reality for distraction and relaxation in a pediatric hospital setting: an interventional study with a mixed-methods design. Front Digit Health 2022 May 31;4:866119 [FREE Full text] [doi: 10.3389/fdgth.2022.866119] [Medline: 35712230]
3. Eijlers R, Utens EM, Staals LM, de Nijs PF, Berghmans JM, Wijnen RM, et al. Systematic review and meta-analysis of virtual reality in pediatrics: effects on pain and anxiety. Anesth Analg 2019 Nov;129(5):1344-1353 [FREE Full text] [doi: 10.1213/ANE.0000000000004165] [Medline: 31136330]
4. Rizzo A, Hartholt A, Grimani M, Leeds A, Liewer M. Virtual reality exposure therapy for combat-related posttraumatic stress disorder. Computer 2014 Jul;47(7):31-37. [doi: 10.1109/mc.2014.199]
5. Riva G. 10.06 - Virtual reality in clinical psychology. In: Asmundson GJ, editor. Comprehensive Clinical Psychology (Second Edition). Amsterdam, The Netherlands: Elsevier; 2022.
6. Rizzo A, Shilling R. Clinical virtual reality tools to advance the prevention, assessment, and treatment of PTSD. Eur J Psychotraumatol 2017 Jan 16;8(sup5):1414560 [FREE Full text] [doi: 10.1080/20008198.2017.1414560] [Medline: 29372007]
7. Hoffman HG, Meyer WJIII, Drever SA, Soltani M, Atzori B, Herrero R, et al. Virtual reality distraction to help control acute pain during medical procedures. In: Rizzo A, Bouchard S, editors. Virtual Reality for Psychological and Neurocognitive Interventions. New York, NY: Springer; 2019.
8. Serino S, Pedroli E, Keizer A, Triberti S, Dakanalis A, Pallavicini F, et al. Virtual reality body swapping: a tool for modifying the allocentric memory of the body. Cyberpsychol Behav Soc Netw 2016 Feb;19(2):127-133. [doi: 10.1089/cyber.2015.0229] [Medline: 26506136]
9. Vernal DL, Nordentoft M, Christensen MJ, Smith LC, Mariegaard L, Mainz J, et al. Status and clinical experiences from the challenge trial – a randomized controlled trial investigating virtual reality-based therapy for auditory hallucinations. World Soc Psychiatry 2023;5(1):71-76. [doi: 10.4103/wsp.wsp_4_23]
10. Rizzo A, Bouchard S. Virtual Reality for Psychological and Neurocognitive Interventions. New York, NY: Springer; 2019.
11. Baniasadi T, Ayyoubzadeh SM, Mohammadzadeh N. Challenges and practical considerations in applying virtual reality in medical education and treatment. Oman Med J 2020 May 19;35(3):e125 [FREE Full text] [doi: 10.5001/omj.2020.43] [Medline: 32489677]
12. Gaekwad JS, Sal Moslehian A, Roös PB. A meta-analysis of physiological stress responses to natural environments: Biophilia and Stress Recovery Theory perspectives. J Environ Psychol 2023 Sep;90:102085. [doi: 10.1016/j.jenvp.2023.102085]
13. Pizzoli SF, Mazzocco K, Triberti S, Monzani D, Alcañiz Raya ML, Pravettoni G. User-centered virtual reality for promoting relaxation: an innovative approach. Front Psychol 2019 Mar 12;10:479 [FREE Full text] [doi: 10.3389/fpsyg.2019.00479] [Medline: 30914996]
14. Valtchanov D, Barton KR, Ellard C. Restorative effects of virtual nature settings. Cyberpsychol Behav Soc Netw 2010 Oct;13(5):503-512. [doi: 10.1089/cyber.2009.0308] [Medline: 20950174]
15. Serrano B, Baños RM, Botella C. Virtual reality and stimulation of touch and smell for inducing relaxation: a randomized controlled trial. Comput Hum Behav 2016 Feb;55:1-8. [doi: 10.1016/j.chb.2015.08.007]
16. Hedblom M, Gunnarsson B, Iravani B, Knez I, Schaefer M, Thorsson P, et al. Reduction of physiological stress by urban green space in a multisensory virtual experiment. Sci Rep 2019 Jul 12;9(1):10113 [FREE Full text] [doi: 10.1038/s41598-019-46099-7] [Medline: 31300656]
17. Annerstedt M, Jönsson P, Wallergård M, Johansson G, Karlson B, Grahn P, et al. Inducing physiological stress recovery with sounds of nature in a virtual reality forest--results from a pilot study. Physiol Behav 2013 Jun 13;118:240-250. [doi: 10.1016/j.physbeh.2013.05.023] [Medline: 23688947]
18. Anderson AP, Mayer MD, Fellows AM, Cowan DR, Hegel MT, Buckey JC. Relaxation with immersive natural scenes presented using virtual reality. Aerosp Med Hum Perform 2017 Jun 01;88(6):520-526. [doi: 10.3357/amhp.4747.2017]
19. Fagernäs S, Hamilton W, Espinoza N, Miloff A, Carlbring P, Lindner P. What do users think about virtual reality relaxation applications? A mixed methods study of online user reviews using natural language processing. Internet Interv 2021 Apr;24:100370 [FREE Full text] [doi: 10.1016/j.invent.2021.100370] [Medline: 33665136]
20. Riches S, Azevedo L, Bird L, Pisani S, Valmaggia L. Virtual reality relaxation for the general population: a systematic review. Soc Psychiatry Psychiatr Epidemiol 2021 Oct 13;56(10):1707-1727 [FREE Full text] [doi: 10.1007/s00127-021-02110-z] [Medline: 34120220]

XSL•FO
RenderX

21. Riches S, Jeyarajaguru P, Taylor L, Fialho C, Little J, Ahmed L, et al. Virtual reality relaxation for people with mental health conditions: a systematic review. Soc Psychiatry Psychiatr Epidemiol 2023 Jul 20;58(7):989-1007 [FREE Full text] [doi: 10.1007/s00127-022-02417-5] [Medline: 36658261]
22. Riches S, Taylor L, Jeyarajaguru P, Veling W, Valmaggia L. Virtual reality and immersive technologies to promote workplace wellbeing: a systematic review. J Ment Health 2024 Apr 15;33(2):253-273 [FREE Full text] [doi: 10.1080/09638237.2023.2182428] [Medline: 36919828]
23. Lindner P, Miloff A, Hamilton W, Carlbring P. The potential of consumer-targeted virtual reality relaxation applications: descriptive usage, uptake and application performance statistics for a first-generation application. Front Psychol 2019 Feb 4;10:132 [FREE Full text] [doi: 10.3389/fpsyg.2019.00132] [Medline: 30778311]
24. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]
25. Garrett B, Taverner T, Gromala D, Tao G, Cordingley E, Sun C. Virtual reality clinical research: promises and challenges. JMIR Serious Games 2018 Oct 17;6(4):e10839 [FREE Full text] [doi: 10.2196/10839] [Medline: 30333096]
26. Segal R, Bhatia M, Drapeau M. Therapists' perception of benefits and costs of using virtual reality treatments. Cyberpsychol Behav Soc Netw 2011 Jan;14(1-2):29-34. [doi: 10.1089/cyber.2009.0398] [Medline: 21329440]
27. Schwartzman D, Segal R, Drapeau M. Perceptions of virtual reality among therapists who do not apply this technology in clinical practice. Psychol Serv 2012 Aug;9(3):310-315. [doi: 10.1037/a0026801] [Medline: 22867123]
28. Flowborne VR homepage. Flowborne VR. URL: https://flowborne.com/ [accessed 2024-02-01]
29. Rockstroh C, Blum J, Göritz AS. A mobile VR-based respiratory biofeedback game to foster diaphragmatic breathing. Virtual Real 2020 Oct 05;25(2):539-552. [doi: 10.1007/s10055-020-00471-5]
30. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
31. Byrne D. A worked example of Braun and Clarke's approach to reflexive thematic analysis. Qual Quant 2021 Jun 26;56(3):1391-1412. [doi: 10.1007/s11135-021-01182-y]
32. Naef AC, Jeitziner MM, Knobel SE, Exl MT, Müri RM, Jakob SM, et al. Investigating the role of auditory and visual sensory inputs for inducing relaxation during virtual reality stimulation. Sci Rep 2022 Oct 12;12(1):17073 [FREE Full text] [doi: 10.1038/s41598-022-21575-9] [Medline: 36224289]
33. Pardini S, Gabrielli S, Dianti M, Novara C, Zucco GM, Mich O, et al. The role of personalization in the user experience, preferences and engagement with virtual reality environments for relaxation. Int J Environ Res Public Health 2022 Jun 13;19(12):7237 [FREE Full text] [doi: 10.3390/ijerph19127237] [Medline: 35742483]
34. Gentile A, Ficarra S, Thomas E, Bianco A, Nordstrom A. Nature through virtual reality as a stress-reduction tool: a systematic review. Int J Stress Manag 2023 Nov;30(4):341-353. [doi: 10.1037/str0000300]
35. Nukarinen T, Istance HO, Rantala J, Mäkelä J, Korpela K, Ronkainen K, et al. Physiological and psychological restoration in matched real and virtual natural environments. In: Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI EA '20; April 25-30, 2020; Honolulu, HI. [doi: 10.1145/3334480.3382956]
36. Sanada K, Montero-Marin J, Alda Díez M, Salas-Valero M, Pérez-Yus MC, Morillo H, et al. Effects of mindfulness-based interventions on salivary cortisol in healthy adults: a meta-analytical review. Front Physiol 2016 Oct 19;7:471 [FREE Full text] [doi: 10.3389/fphys.2016.00471] [Medline: 27807420]
37. Muhammad Khir S, Wan Mohd Yunus WM, Mahmud N, Wang R, Panatik SA, Mohd Sukor MS, et al. Efficacy of progressive muscle relaxation in adults for stress, anxiety, and depression: a systematic review. Psychol Res Behav Manag 2024 Feb;Volume 17:345-365. [doi: 10.2147/prbm.s437277]
38. Immersive Mental Health. Google Play Store. URL: https://play.google.com/store/apps/details?id=com.Aeroplane.ImmersiveMentalHealth&hl=nl [accessed 2025-01-30]
39. Immersive Mental Health. Apple App Store. URL: https://apps.apple.com/be/app/immersive-mental-health/id6736897735 [accessed 2025-01-30]
40. Jeong HS, Oh J, Paik M, Kim H, Jang S, Kim BS, et al. Development and feasibility assessment of virtual reality-based relaxation self-training program. Front Virtual Real 2022 Jan 5;2. [doi: 10.3389/frvir.2021.722558]
41. De Asis KM, Guillem EJ, Reyes FA, Samonte MJ. Serenity: a stress-relieving virtual reality application based on Philippine environmental variables. In: Proceedings of the 6th International Conference on Frontiers of Educational Technologies. 2020 Presented at: ICFET '20; June 5-8, 2020; Tokyo, Japan. [doi: 10.1145/3404709.3404745]
42. Wang SY, Bang HS, Liu Y, Kim M, Raiti J. Gaia VR: a virtual reality-based neuro and biofeedback meditation application. In: Proceedings of the IEEE Global Humanitarian Technology Conference. 2022 Presented at: GHTC 2022; September 8-11, 2022; Santa Clara, CA. [doi: 10.1109/ghtc55712.2022.9910614]
43. Seabrook E, Kelly R, Foley F, Theiler S, Thomas N, Wadley G, et al. Understanding how virtual reality can support mindfulness practice: mixed methods study. J Med Internet Res 2020 Mar 18;22(3):e16106 [FREE Full text] [doi: 10.2196/16106] [Medline: 32186519]

XSL•FO
RenderX

44. Zaharuddin FA, Ibrahim N, Yusof AM. A conceptual framework for designing virtual environments for stress therapy. Appl Sci 2022 Oct 04;12(19):9973. [doi: 10.3390/app12199973]
45. van Bruinessen IR, van Weel-Baumgarten EM, Snippe HW, Gouw H, Zijlstra JM, van Dulmen S. Active patient participation in the development of an online intervention. JMIR Res Protoc 2014 Nov 06;3(4):e59 [FREE Full text] [doi: 10.2196/resprot.3695] [Medline: 25379679]
46. Brotherdale R, Berry K, Branitsky A, Bucci S. Co-producing digital mental health interventions: a systematic review. Digit Health 2024 Apr 25;10:20552076241239172 [FREE Full text] [doi: 10.1177/20552076241239172] [Medline: 38665886]
47. Papachristos NM, Vrellis I, Mikropoulos TA. A comparison between oculus rift and a low-cost smartphone VR headset: immersive user experience and learning. In: Proceedings of the IEEE 17th International Conference on Advanced Learning Technologies. 2017 Presented at: ICALT 2017; July 3-7, 2017; Timisoara, Romania. [doi: 10.1109/icalt.2017.145]
48. Amin A, Gromala D, Tong X, Shaw C. Immersion in cardboard VR compared to a traditional head-mounted display. In: Proceedings of the 8th International Conference on Virtual, Augmented and Mixed Reality. 2016 Presented at: VAMR 2016; July 17-22, 2016; Toronto, Canada. [doi: 10.1007/978-3-319-39907-2_25]
49. Spano G, Theodorou A, Reese G, Carrus G, Sanesi G, Panno A. Virtual nature, psychological and psychophysiological outcomes: a systematic review. J Environ Psychol 2023 Aug;89:102044. [doi: 10.1016/j.jenvp.2023.102044]
50. Barton AC, Do M, Sheen J, Byrne LK. The restorative and state enhancing potential of abstract fractal-like imagery and interactive mindfulness interventions in virtual reality. Virtual Real 2024 Feb 27;28:53. [doi: 10.1007/s10055-023-00916-7]
51. Li H, Zhang X, Wang H, Yang Z, Liu H, Cao Y, et al. Access to nature via virtual reality: a mini-review. Front Psychol 2021;12:725288 [FREE Full text] [doi: 10.3389/fpsyg.2021.725288] [Medline: 34675840]
52. Schebella MF, Weber D, Schultz L, Weinstein P. The nature of reality: human stress recovery during exposure to biodiverse, multisensory virtual environments. Int J Environ Res Public Health 2019 Dec 19;17(1):56 [FREE Full text] [doi: 10.3390/ijerph17010056] [Medline: 31861700]
53. Ladakis I, Filos D, Chouvarda I. Virtual reality environments for stress reduction and management: a scoping review. Virtual Real 2024 Feb 27;28:50. [doi: 10.1007/s10055-024-00943-y]
54. Choi YH, Paik NJ. Mobile game-based virtual reality program for upper extremity stroke rehabilitation. J Vis Exp 2018 Mar 08(133):56241 [FREE Full text] [doi: 10.3791/56241] [Medline: 29578520]

Abbreviations

IMH: Immersive Mental Health
VR: virtual reality

Edited by A Mavragani; submitted 31.05.24; peer-reviewed by C Bull; comments to author 30.10.24; revised version received 19.11.24; accepted 18.12.24; published 13.02.25.

Please cite as:
Verstegen A, Van Daele T, Bonroy B, Debard G, Sels R, van Loo M, Bernaerts S
Designing a Smartphone-Based Virtual Reality App for Relaxation: Qualitative Crossover Study
JMIR Form Res 2025;9:e62663
URL: https://formative.jmir.org/2025/1/e62663
doi:10.2196/62663
PMID:

©Amandine Verstegen, Tom Van Daele, Bert Bonroy, Glen Debard, Romy Sels, Marlon van Loo, Sylvie Bernaerts. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Using an Interactive Voice Response Survey to Assess Patient Satisfaction in Ethiopia: Development and Feasibility Study

Dessalegn Shamebo[1*], PhD; Anagaw Derseh Mebratie[2*], PhD; Catherine Arsenault[3], PhD

[1]Department of Development Economics, Ethiopian Civil Service University, Addis Ababa, Ethiopia

[2]Department of Health Systems Management and Health Policy, School of Public Health, College of Health Sciences, Addis Ababa University, Addis Ababa, Ethiopia

[3]Department of Global Health, Milken Institute School of Public Health, The George Washington University, Washington, DC, United States

[*]these authors contributed equally

Corresponding Author:
Catherine Arsenault, PhD
Department of Global Health
Milken Institute School of Public Health
The George Washington University
950 New Hampshire Avenue
Washington, DC, 20052
United States
Phone: 1 2029941011
Email: catherine.arsenault@gwu.edu

Abstract

Background: Patient satisfaction surveys can offer crucial information on the quality of care but are rarely conducted in low-income settings. In contrast with in-person exit interviews, phone-based interactive voice response (IVR) surveys may offer benefits including standardization, patient privacy, reduced social desirability bias, and cost and time efficiency. IVR surveys have rarely been tested in low-income settings, particularly for patient satisfaction surveys.

Objective: In this study, we tested the feasibility of using an IVR system to assess patient satisfaction with primary care services in Addis Ababa, Ethiopia. We described the methodology, response rates, and survey costs and identified factors associated with survey participation, completion, and duration.

Methods: Patients were recruited in person from 18 public and private health facilities in Addis Ababa. Patients' sex, age, education, reasons for seeking care, and mobile phone numbers were collected. The survey included 15 questions that respondents answered using their phone keypad. We used a Heckman probit regression model to identify factors influencing the likelihood of IVR survey participation (picking up and answering at least 1 question) and completion (answering all survey questions) and a Weibull regression model to identify factors influencing the survey completion time.

Results: A total of 3403 individuals were approached across 18 health facilities. Nearly all eligible patients approached (2985/3167, 94.3%) had a functioning mobile phone, and 89.9% (2415/2685) of those eligible agreed to be enrolled in the study. Overall, 92.6% (2236/2415) picked up the call, 65.6% (1584/2415) answered at least 1 survey question, and 42.9% (1037/2415) completed the full survey. The average survey completion time was 8.1 (SD 1.7) minutes for 15 Likert-scale questions. We found that those aged 40-49 years and those aged 50+ years were substantially less likely to participate in (odds ratio 0.63, 95% CI 0.53-0.74) and complete the IVR survey (odds ratio 0.77, 95% CI 0.65-0.90) compared to those aged 18-30 years. Higher education levels were also strongly associated with survey participation and completion. In adjusted models, those enrolled in private facilities were less likely to participate and complete the survey compared to those in public health centers. Being male, younger, speaking Amharic, using a private hospital, and being called after 8 PM were associated with a shorter survey duration. The average survey costs were US $7.90 per completed survey.

Conclusions: Our findings reveal that an IVR survey is a feasible, low-cost, and rapid solution to assess patient satisfaction in an urban context in Ethiopia. However, survey implementation must be carefully planned and tailored to local challenges. Governments and health facilities should consider IVR to routinely collect patient satisfaction data to inform quality improvement strategies.

(JMIR Form Res 2025;9:e67452) doi:10.2196/67452

XSL•FO
RenderX

KEYWORDS

mobile phone surveys; patient satisfaction; interactive voice response; global health; surveys; Ethiopia; IVR; Africa

Introduction

Collecting information on patient's perception of health care quality is crucial to understand gaps in service provision. Patient satisfaction surveys have gained increasing attention as an essential source of information for quality improvement [1]. In high-income countries, these surveys are often paper-based and mailed to patients or delivered by email and completed on the web [2]. However, in low-income settings, these delivery methods can be problematic. Paper- and web-based surveys also assume that all recipients can read and write.

Other methods include phone surveys that have been increasingly popular, given the high rates of mobile phone ownership globally despite a relatively lower penetration rate in Sub-Saharan Africa (50%) [3]. Phone surveys with live interviewers can be effective but come with several challenges including the cost of hiring and training interviewers. They are also time-consuming for large samples and are prone to social desirability bias and to interviewer variability and bias [4,5]. Interactive voice response (IVR) surveys enable automated communications over the phone, where respondents engage with the survey using their phone's keypad. IVR surveys offer several benefits, particularly in the context of patient satisfaction surveys. IVR surveys can be accessed by patients who may not have internet access or are not comfortable using web-based platforms. They provide a sense of anonymity, encouraging patients to provide honest feedback [6]. The automated nature of IVR ensures that every patient is asked the same set of questions in the same manner. They are also cost-effective and highly scalable and can be delivered to thousands of patients at the same time and in multiple languages, leading to very short data collection periods for real-time feedback [7-9]. Furthermore, an IVR system can make multiple calls both during and outside of normal business hours, allowing respondents to participate in the survey at their convenience [10]. Given these benefits, they have the potential to be integrated into the routine operations of health systems.

Nonetheless, IVR surveys can face a unique set of limitations and challenges. First, they may not be suitable for individuals who are not comfortable with automated phone systems. They may lead to lower engagement, greater response fatigue, a higher rate of survey abandonment, and thus a lower overall response rate compared to other survey delivery methods. People who are not technology-savvy may struggle with navigating IVR systems. Despite these limitations, IVR surveys have been used for various purposes such as for patient satisfaction assessment after endoscopy [11], evaluation of child and adolescent psychiatric outpatient treatment [8], and decision-making about sun protection [12]. Yet, there is little published evidence on the acceptability and feasibility of using IVR to collect patient feedback in Ethiopia. We found only 2 peer-reviewed studies that discussed the feasibility of IVR in Ethiopia [13,14]. These studies assessed the feasibility of IVR messages for targeted client communication. One working paper assessed the feasibility of an IVR survey in Ethiopia to obtain nationally representative estimates at population levels [15]. To our knowledge, there has been no study assessing the use of IVR to measure patient satisfaction in Ethiopia.

In this paper, we assessed the feasibility of using an IVR survey to assess patient-perceived quality of primary care services in 18 health facilities in Addis Ababa, Ethiopia. We described the methodology, acceptability, cost, response rate, completion rate, and the time taken to complete the full survey. We also assessed the factors associated with survey participation, completion, and duration. These findings may be helpful for others conducting studies among patients or considering using IVR in a similar setting.

Methods

Study Design and Recruitment

This study was conducted in Addis Ababa, Ethiopia, in June to July 2024. Phone numbers and consent from primary care patients were obtained in person at health facilities. The study aimed to compare patient-perceived quality of primary care services across 4 types of facilities: public health centers, public general hospitals, private clinics, and private hospitals. Four facilities from each category were randomly selected from health management information system lists with the aim of enrolling a minimum of 100 patients per facility.

Enrollment was limited to patients receiving 11 common primary care services, including for chronic diseases, injuries, infectious diseases, gastrointestinal conditions, respiratory diseases, prenatal care, and family planning. Eligible participants were those who understood Amharic or Afaan Oromo, were aged 18 years and older, were at the facility for their own care, had received one of the selected services as an outpatient, and had a functioning mobile phone. Hospitalized individuals were excluded. Data collectors identified eligibility, obtained consent, and collected sex, educational level, mobile phone number, and preferred contact time for the IVR survey. No other personal identifiers were collected.

The recruitment period lasted 14 days. Each day, phone numbers were uploaded to the IVR platform. One day after enrollment, participants received an SMS text message at their preferred time to remind them of the study, followed by a call within 10 minutes to complete the IVR survey. During recruitment, respondents listened to a sample IVR question to familiarize themselves with the system. They were advised to respond honestly and to find a quiet location for the survey.

The IVR survey was programmed using the EngageSpark platform. The survey questions and response options were recorded in a recording studio in Amharic and Afaan Oromo and uploaded to the IVR platform. The IVR survey recordings were narrated by a female, given evidence that IVR surveys narrated by female individuals lead to higher response rates and can prevent the incidence of domestic violence [16]. The first call for the IVR survey took place 1 day after the primary care visit. If the participant did not pick up, the system automatically

XSL•FO
RenderX

called the respondent back after 30 minutes, 1 hour, 90 minutes, 1 day, 1 day and 1 hour, 2 days, and 2 days and 1 hour. For respondents who completed the full survey, the EngageSpark platform automatically sent the airtime top-up to the respondent's phone.

The patient satisfaction survey included 15 questions covering structural quality (eg, perceived quality and availability of equipment, medicine, and diagnostic test), competence of health providers, respectful care, patient-centered care, and user experience. The survey instrument is provided in Multimedia Appendix 1. Participants were also asked to report their monthly household income and how likely they were to recommend the health facility to a family member or friend. Most questions used a 5-point Likert scale. Among 15 questions asked, 13 items on the quality and cost of care were rated on a 5-point Likert scale (from 1=very poor to 5=very good). The question on endorsement of the health facility was answered on a 4-point Likert scale (1=not at all likely to 4=very likely). The household income question was categorized into 6 income groups.

Respondents were instructed to use their mobile phone touchpad to answer questions. Respondents were allowed to repeat or skip a question. A message was recorded to inform participants if they had pressed an invalid key. The survey question was repeated if respondents pressed an invalid key or did not answer for more than 20 seconds. The call was disconnected if they pressed an invalid key 3 times or failed to respond after a question was repeated twice.

Measures

The outcomes of interest in this study were participation in the IVR survey, completion of all survey questions, and time taken to complete all 15 survey questions (among those who completed the survey). Participation in the IVR survey was measured using a dummy variable, which was coded as 1 if the individual picked up the survey call and answered at least 1 of the survey questions and 0 otherwise. Similarly, the completion outcome was constructed as a dichotomous variable, coded as 1 if the respondent completed all 15 questions and 0 if they did not. Duration was based on the time taken in minutes to answer all 15 survey questions among those who completed the full survey. To assess the correlates of survey participation, completion, and duration, our analysis included a series of covariates collected during the in-person recruitment. These were the respondent's sex, age, educational achievement, type of primary health care service received, type of health facility used, preferred language (Amharic or Afaan Oromo), and preferred survey time.

Statistical Analysis

We used descriptive statistics to describe patient demographics, enrollment, survey response, completion rates, and survey duration. A Heckman probit regression model was used to identify factors influencing survey participation (answering at least 1 question) and completion (answering all questions), with the first step determining participation factors (selection model). We chose the Heckman probit model over the standard probit model due to its advantages in predicting outcomes when there is a dependency between the selection model and the outcome model as well as a correlation of error terms between the 2 models [17,18]. The participation model specification was:

$$S^* = ZB + \varepsilon$$

Where S^* is a latent variable representing the participation decision in the IVR survey, Z is a vector of covariates that include the respondents' characteristics (sex, age, education level, timing of the IVR survey, type of facility visited, type of care received, and the patient's preferred language), B represents the coefficients of interest, and ε is the error term. The second step of the model aimed to identify the factors associated with the completion of the IVR survey (outcome variable), conditional on participation in the survey. This equation used the inverse Mills ratio, derived from the participation equation, to overcome selection bias.

The outcome (completion) model was specified as:

$$Y^* = X\psi + \lambda + u$$

Where Y^* is a latent variable representing the outcome, X is a vector of covariates that includes all factors listed in the selection model plus number of reconnections made to complete the survey, ψ represents the coefficients to be estimated, λ is the inverse Mills ratio from the participation equation, and u is the normally distributed error term. Adding λ adjusts the outcome equation for more accurate estimates of the relationship between participation and covariates. Language was used as a selection variable for participation in the IVR survey, with Afaan Oromo speakers from rural areas around Addis Ababa being more likely to participate than those from urban areas.

For those who completed the full survey, a Weibull regression model was used to identify the factors associated with the time taken to complete the survey. The main cost driver of an IVR survey is the airtime cost while a respondent answers the survey. Identifying factors influencing survey duration is helpful due to cost implications. Duration data are often skewed, with many early events and few late ones. The Weibull model is versatile for providing insights about the effects of covariates on the timing of events (survey completion).

Ethical Considerations

The study was approved by the Addis Ababa University College of Health Sciences institutional review board (approval 057/23/SPH). Ethics approval was also obtained from the Addis Ababa City Health Bureau. Only those participants who provided verbal consent were interviewed. All data were deidentified to ensure participant privacy and confidentiality. Participants who completed the full survey received approximately US $0.86 mobile phone airtime top-up.

Results

Recruitment

Table 1 reports the number of individuals who were approached and enrolled in the study. Of 3403 individuals approached at the study health facilities, 3167 (93.1%) were receiving outpatient care for themselves, and 2985 (87.7%) had a functioning mobile phone with them at the time. The majority (n=2685, 78.9%) had received 1 of 11 eligible primary care

XSL•FO
RenderX

services. Of these, 2415 (86.9%) agreed to participate, provided a valid phone number, and were enrolled in the IVR survey.

Table 1. Recruitment results: eligibility and consent to participate in the interactive voice response survey.

	Values, n (%)
Number of people approached (n= 3403)	
Receiving care for themselves	3238 (95.2)
Did not spend the night at the facility	3167 (93.1)
Had a functioning mobile phone	2985 (87.7)
Received eligible primary care services[a]	2685 (78.9)
Eligible to participate in the study (n= 2685)	
Agreed to participate	2604 (97)
Agreed to provide their mobile phone number	2469 (92)
Provided a valid mobile phone number	2415 (89.9)

[a]Participants were eligible to participate in the study if they received at least 1 of the following services: care for hypertension or high blood pressure, care for diabetes or high blood sugar, care for kidney disease, care for an injury, prenatal care, family planning services, care for tuberculosis, HIV or AIDS or a sexually transmitted infection (including testing or treatment), care for diarrhea or gastrointestinal conditions, or care for a respiratory illness.

IVR Survey Completion and Duration

Of 2415 individuals enrolled in the IVR survey, 2236 (92.6%) participants picked up the call on their mobile phone a day later. However, 16.6% (n=400) of these respondents only listened to the introduction, and another 10.4% (n=252) listened to the first question but failed to answer it, leading to 1584 (65.6%) respondents answering at least 1 question. Among them, 8.6% (n=208) answered part of the survey and hung up, and 14% (n=339) answered part of the survey but were disconnected by the system (for failing to respond or pressing invalid keys more than twice). Finally, 1037 (42.9%) respondents completed the full survey (answered all 15 questions). Those who completed the survey took on average 8.5 (SD 1.68) minutes to answer all questions, with a median of 7.55 (IQR 6.54-8.58) minutes, a minimum of 4 minutes and 4 seconds, and a maximum of 14 minutes and 59 seconds (Table 2).

Table 2. Interactive voice response (IVR) survey response rate.

	Values, n (%)
Enrolled in the IVR survey (n=2415)	
Picked up the IVR survey call	2236 (92.6)
Answered at least 1 question	1584 (65.6)
Participated in the IVR survey (n=1584)	
Answered at least 1 question but the system ended the call[a]	339 (14)
Answered at least 1 question but hung up without completing the survey	208 (8.6)
Answered the full survey	1037 (42.9)

[a]The IVR system disconnected the call if the user did any of the following 3 times: failed to reply within 20 seconds or pressed an invalid key.

Characteristics of the Study Respondents

Among the 2415 enrolled individuals, 1584 (65.7%) participated in the survey by responding to at least 1 survey question. The characteristics of the eligible individuals by participation status are reported in Table 3. A little over half of the individuals were female (n=1398, 57.9%), and the majority were younger than 40 years of age. Educational attainment varied, with nearly half of the respondents having education above the secondary level, and only 5.6% (n=134) with no formal education. In the latest Demographic and Health Survey, the proportion of Addis Ababa respondents with higher education was 27.7% only [19]. Although this survey is more than 8 years old, it is possible that our sample is wealthier and more educated than the general population, as they represent health system users, and half of them were enrolled in private health facilities. This may overrepresent the private sector share of health care services delivered in the city. The participants were selected nearly equally from each of the 4 health facility types, including public hospitals (n=628, 26%), health centers (n=624, 25.8%), private clinics (n=621, 25.7%), and private hospitals (n=542, 22.4%).

XSL•FO
RenderX

Table 3. Characteristics of the interactive voice response (IVR) sample by participation status among those enrolled.

Variable	Participants[a] (n=1584), n (%)	Nonparticipants (n=831), n (%)	Total enrolled (n=2415), n (%)
Sex			
Female	908 (57.3)	490 (58.9)	1398 (57.9)
Male	676 (42.7)	341 (41)	1017 (42.1)
Age group (years)			
18-29	623 (39.3)	278 (33.4)	901 (37.3)
30-39	516 (32.6)	249 (30)	765 (31.7)
40-49	219 (13.8)	145 (17.5)	364 (15.1)
≥50	226 (14.3)	159 (19.1)	385 (15.9)
Education			
No education at all	67 (4.2)	67 (8.1)	134 (5.6)
Primary education	326 (20.6)	172 (20.8)	498 (20.6)
Secondary education	436 (27.5)	229 (27.6)	665 (27.6)
Higher education	755 (47.7)	361 (43.6)	1116 (46.3)
Care type			
Maternal or reproductive	374 (23.7)	222 (26.8)	596 (24.8)
Communicable disease	106 (6.7)	58 (7)	164 (6.8)
Noncommunicable disease	661 (41.9)	347 (42)	1008 (41.9)
Diarrhea and respiratory	436 (27.7)	200 (24.2)	636 (26.5)
Language			
Afaan Oromo	274 (17.3)	106 (12.8)	380 (15.7)
Amharic	1310 (82.7)	725 (87.2)	2035 (84.3)
Preferred survey time			
Morning (8-11:59 AM)	319 (20.1)	155 (18.7)	474 (19.6)
Mid-day (12-1:59 PM)	323 (20.4)	193 (23.2)	516 (21.4)
Afternoon (2-5:59 PM)	269 (17)	173 (20.8)	442 (18.3)
Evening (6-7:59 PM)	320 (20.2)	164 (19.7)	484 (20)
Night (8-10 PM)	353 (22.3)	146 (17.7)	499 (20.7)
Facility type			
Health centers	443 (28)	181 (21.8)	624 (25.8)
Public hospitals	411 (26)	217 (26.1)	628 (26)
Private hospitals	333 (21)	209 (25.2)	542 (22.4)
Private clinic	397 (25.1)	224 (27)	621 (25.7)

[a]This shows the characteristics of those enrolled based on their participation status in the interactive voice response survey (answered at least 1 question or not).

Regarding the type of care received, around 41.9% (n=1008) of the participants sought care for noncommunicable diseases (hypertension, diabetes, or kidney disease), 26.5% (n=637) were treated for diarrhea or respiratory conditions, 24.8% (n=596) received maternal or reproductive health care, and 6.8% (n=164) received care for communicable diseases (including tuberculosis, HIV/AIDS, or sexually transmitted infections). The majority of respondents (n=2035, 84.3%) spoke Amharic. Preferences for survey times were fairly evenly distributed between morning, mid-day, afternoon, evening, and night time.

Survey Costs and Data Quality

A total of US $13,512 was spent to implement the survey, but 39.3% of these costs (US $5316) were driven by in-person enrollment, which was necessary to obtain the patients' phone numbers (Table 4). Health facilities, governments, or researchers that can obtain patient phone numbers directly from health records could avoid this portion of the cost. Excluding the in-person expenses, the IVR survey costs were US $7.90 per respondent who completed the full survey (n=1037). Considering those who answered at least 1 question (n=1584),

XSL•FO
RenderX

the survey costs were estimated at US $5.17 per partial survey. Of 2415 patients called, only 1584 answered at least 1 question. A total of 831 patients answered no survey questions. The costs for these patients were negligible and included only the cost of the SMS primer and of any airtime used during which the patient may have listened to the introduction and the first question without answering.

Table 4. Survey costs.

	Costs (US $)
In-person enrollment	
Including enumerator payments, transport, training, and supplies	5316
Interactive voice response survey costs	
Survey translation and recording	300
SMS primer (2415 SMS text messages sent)	1017
Survey airtime	5439
Incentive (airtime top-up)	1440
Total	13,512

As expected, the number of respondents answering the survey questions declined as the survey progressed, and 34.5% (547/1584) of respondents who began the survey did not complete all questions. For 11 of 15 questions, fewer than 2% of respondents chose to skip the question by pressing “9.” However, there were 3 notable exceptions. First, 36.5% (n=578) of respondents opted to skip the first question, which asked them to rate the cleanliness of the facility. Despite this, most participants went on to answer the second question and continued the survey. This high skip rate is likely because respondents were testing the survey technology on the first question. Future implementers should consider adding a test question at the beginning of their survey, given the high skip rate for our first survey question. Second, for question 13, which informed participants that only 3 questions remained, 16.1% (n=184) chose to skip this question, which asked about the affordability of the services received. Finally, 6.9% (n=72) of respondents skipped the last question, which asked about their monthly household income.

Factors Affecting Participation in and Completion of the IVR Survey

A Heckman sample selection model was used to predict covariates associated with IVR survey participation (answering at least 1 question) and completion. Before executing this model, we assessed its appropriateness against the standard probit model. The results indicated a correlation between the error terms of the 2 equations, as the value of athrho was statistically significant (P<.001). Furthermore, the findings showed a dependence between the selection model and the outcome model, justifying the use of the Heckman probit model since ρ is significantly different from 0 (χ^2_1=9.2; P=.002). Additionally, the likelihood function of the Heckman probit model was significant (Wald χ^2_{19}=269.9; P<.001), indicating the better explanatory power of the Heckman probit model.

The model showed that as the age of the participants increased, the likelihood of both participation and completion declined (Table 5). For example, there was a significant difference in the odds of participating in (odds ratio [OR] 0.77, 95% CI 0.65-0.90) and completing the IVR survey (OR 0.63, 95% CI 0.53-0.74) between participants aged 50 years and older compared to those aged 18-30 years. Similarly, higher education increased the odds of participation and completion. For instance, for respondents with postsecondary education, the odds of participation were 1.54 times higher (95% CI 1.25-1.90) and the odds of survey completion were 1.74 times higher (95% CI 1.46-2.06) compared to those with no education.

XSL•FO
RenderX

Table 5. Results of the Heckman probit model.

Variables	Completion of the IVR[a] survey (2402 observations), OR[b] (95% CI)	Participation in the IVR survey (2402 observations), OR (95% CI)
Sex (reference=female)		
Male	1.010 (0.907-1.125)	0.983 (0.880-1.097)
Age category (years) (reference 18-30 years)		
30-39	0.853[c] (0.757-0.960)	0.958 (0.846-1.084)
40-49	0.666[c] (0.569-0.779)	0.785[c] (0.668-0.921)
≥50	0.628[c] (0.534-0.739)	0.765[c] (0.648-0.904)
Education level (reference=no education)		
Primary	0.993 (0.831-1.188)	1.339[c] (1.081-1.658)
Secondary	1.318[c] (1.111-1.563)	1.337[c] (1.085-1.648)
Above secondary	1.737[c] (1.462-2.063)	1.540[c] (1.248-1.901)
Time of the survey (reference=morning [8-11:59 AM])		
Mid-day (12-1:59 PM)	1.017 (0.873-1.184)	0.861[d] (0.735-1.007)
Afternoon (2-5:59 PM)	0.963 (0.819-1.133)	0.826[e] (0.699-0.975)
Evening (6-7:59 PM)	0.971 (0.831-1.135)	0.920 (0.783-1.081)
Night (8-10 PM)	1.163[d] (0.995-1.361)	1.047 (0.888-1.236)
Facility type (reference=public health center)		
Public hospital	0.866[d] (0.748-1.002)	0.920 (0.791-1.070)
Private hospital	0.797[c] (0.677-0.938)	0.724[c] (0.611-0.857)
Private clinic	0.840[e] (0.727-0.972)	0.819[c] (0.705-0.952)
Number of reconnection (reference=0)		
One	0.588[c] (0.514-0.673)	—[f]
Two or more	0.472[c] (0.365-0.611)	—
Type of care (reference=NCD[g])		
Maternal and reproductive	0.804[c] (0.699-0.925)	0.809[c] (0.700-0.934)
Communicable diseases	0.950 (0.775-1.165)	0.904 (0.735-1.110)
Diarrhea and respiratory	0.972 (0.854-1.107)	0.992 (0.867-1.135)
Language (reference=Afaan Oromo)		
Amharic	—	0.852[c] (0.761-0.954)
Athrho	— (10,193-1.090e+09)	—
Constant	—	1.810[c] (1.452-2.256)

[a]IVR: interactive voice response.
[b]OR: odds ratio.
[c]*P*<.01.
[d]*P*<.1.
[e]*P*<.05.
[f]Not applicable.
[g]NCD: noncommunicable disease.

We also found that participation in and completion of the IVR survey was also affected by time of day. Those who were called in the afternoon had lower odds of participating compared to those called in the morning (OR 0.84, 95% CI 0.70-0.98). There

XSL•FO
RenderX

was no difference in survey completion according to time of day.

Compared to patients who used services in health centers, those from private hospitals and private clinics were less likely to participate in and complete the survey. For example, primary care patients from private hospitals were substantially less likely to participate in (OR 0.72, 95% CI 0.61-0.86) and complete (OR 0.80, 95% CI 0.68-0.95) the survey compared to those in health centers.

Participants who had received maternal and reproductive health services were less likely to participate in and complete the IVR survey compared to those who received care for noncommunicable diseases. Another important factor identified in determining the completion of the IVR survey was the number of reconnection attempts made. The results revealed that as the number of reconnections increased, participants were less likely to complete the survey. For example, there was a significant difference in completion rates between participants reconnected twice and those not reconnected with an OR of 0.47 (95% CI 0.37-0.61).

Factors Affecting the Time Taken to Complete the IVR Survey

Results from the Weibull model revealed duration dependency, with a *P* value greater than 1 (*P*=5.037), implying that the hazard function increased over time. This suggests that IVR survey participants were more likely to complete the survey as time progressed. The results indicated that sex, age, time of day, language, and facility type were associated with the duration of IVR survey completion (Table 6).

XSL•FO
RenderX

Table 6. Results from Weibull regression (1034 observations).

Variables	Hazard rate (95% CI)
Sex (reference=female)	
Male	1.148[a] (1.005-1.311)
Age category (years) (reference<30 years)	
30-39	0.957 (0.828-1.106)
40-49	0.933 (0.759-1.146)
≥50	0.744[a] (0.589-0.941)
Education level (reference=no education)	
Primary	0.867 (0.565-1.329)
Secondary	0.972 (0.641-1.476)
Above secondary	1.243 (0.824-1.876)
Time of the IVR[b] (reference=morning [8-11:59 AM])	
Mid-day (12-1:59 PM)	1.139 (0.929-1.396)
Afternoon (2-5:59 PM)	1.034 (0.834-1.282)
Evening (6-7:59 PM)	1.078 (0.878-1.323)
Night (8-10 PM)	1.293[a] (1.063-1.574)
Language (reference=Afaan Oromo)	
Amharic	1.911[c] (1.607-2.273)
Facility type (reference=health center)	
Public hospital	1.003 (0.836-1.203)
Private hospital	1.210[a] (1.003-1.459)
Private clinic	1.164 (0.967-1.400)
Type of care (reference=NCD[d])	
Maternal and reproductive	1.140 (0.950-1.367)
Communicable diseases	0.983 (0.759-1.272)
Diarrhea and respiratory	1.154[e] (0.988-1.347)
P value	5.037[c] (4.826-5.258)
Constant	0[c] (0-0)

[a]*P*<.05.

[b]IVR: interactive voice response.

[c]*P*<.01.

[d]NCD: noncommunicable disease.

[e]*P*<.10.

Male participants were 1.15 times more likely to complete the survey early compared to female participants, controlling for other factors. As age increased, individuals were less likely to complete the survey early. For instance, the odds of completing the IVR survey quickly for participants aged 50 years and older was 0.74 (95% CI 0.59-0.94) compared to those participants aged 18 to 30 years. The odds of completing the IVR survey faster were higher at night (OR 1.29, 95% CI 1.06-1.57) compared to morning calls. The odds of completing the survey faster were higher for those enrolled in private hospitals (OR 1.21, 95% CI 1.03-1.46) compared to those from health centers.

The regression model also controlled for survey language as a proxy for geographical area. Most people speaking Afaan Oromo were from rural areas surrounding Addis Ababa, while those responding in Amharic were from the urban area. The results showed that participants responding in Amharic were more likely to complete the survey early with an OR of 1.91 (95% CI 1.61-2.27).

XSL•FO
RenderX

Discussion

Principal findings

This study assessed the feasibility of using an IVR survey as an alternative approach to collect patient satisfaction data in Addis Ababa, Ethiopia. The findings demonstrate a high rate of mobile phone ownership and a willingness among patients to participate, indicating that IVR surveys can serve as a viable low-cost alternative to traditional survey methods to assess patient satisfaction in similar urban areas in a low-income country. Notably, 92.6% (n=2236) of those enrolled picked up the call, 65.6% (n=1584) answered at least 1 survey question, and 42.9% (n=1037) completed the full survey in an average time of 8.1 (SD 1.7) minutes. In adjusted regression models, being older, having less education, and using private facilities negatively affected survey participation and completion. In contrast, being male, being younger, speaking Amharic, using a private hospital, and being called after 8 PM were associated with a shorter survey duration.

The completion rate of 42.9% (n=1037) is encouraging, as almost all of the participants had limited experience with this type of technology. Previous IVR survey studies in Bangladesh, Burkina Faso, Tanzania, and Uganda have found varying response rates ranging from 1% to 41% [20-22]. Response rates were affected by the sampling methods (in-person household surveys or random digit dialing) and type of incentives. Our study, which sampled respondents in person in health facilities and used incentives, achieved a slightly higher response rate (n=1037, 42.9%) than previous studies. In the United States, IVR surveys have been used in various patient satisfaction studies including after endoscopies [11] and for child and adolescent psychiatric outpatient treatment [8]. According to EngageSpark, a global survey and engagement platform, IVR campaigns generally have higher acceptance rates than SMS campaigns. IVR surveys typically achieve response rates ranging from 25% to 75%, while the acceptance rate for SMS surveys ranges from 5% to 15% [23].

Our findings also highlighted the challenges of conducting an IVR survey. For instance, around 16.6% (n=400) of participants listened to the introduction but hung up before answering any questions, and the other 10.4% (n=252) failed to answer the first question despite staying on the call. This suggests that while the initial engagement was successful, there may be factors in the introduction or in the survey's initial questions that could be optimized to retain participants' attention or improve their understanding of the system.

We identified several factors influencing participation and completion of the IVR survey. Age emerged as a significant factor, with younger participants being more likely to participate in and complete the IVR survey compared to older participants. This finding aligns with previous studies suggesting that younger individuals are more comfortable with technology and value the privacy that IVR surveys can provide [24]. Education level also significantly impacts both participation and completion rates, with individuals with higher education levels being more likely to engage with IVR surveys. This supports the notion that higher education correlates with better use of technology and higher perceived importance of surveys [25,26].

The type of facility visited by participants influenced the likelihood of participating in and completing the survey. Participants from private hospitals and clinics were less likely to engage with the survey compared to those from health centers. This could be due to the perception that feedback is more likely to improve services in public facilities, which typically serve poorer segments of the population. Another potential reason could be the provision of incentives that attracted lower-income individuals who are more likely to use affordable health care from public health centers [27]. Our incentive consisted of approximately US $0.86 mobile top-up for those who completed the full survey.

The Weibull model for factors associated with the time taken to complete the survey suggested that the hazard function increases over time and that participants were more likely to complete the IVR survey as time progressed [28,29]. The regression results showed that younger individuals tend to complete the survey faster likely due to their greater familiarity with technology. This finding is consistent with literature indicating that older adults may face more challenges with technology, leading to longer completion times [20,30]. The time of day when the IVR survey was conducted also affected completion times. Surveys conducted at night were completed faster compared to those conducted in the morning. This could be due to participants having fewer distractions at night or that those completing the survey at night tend to be more educated. This finding aligns with the idea that identifying optimal times for survey calls can increase participation [24].

Although participants from private hospitals were less likely to participate in and complete the survey, those who completed it were more likely to have a shorter survey duration compared to those from health centers. This may be because higher-income individuals, who are more likely to visit private facilities, are more efficient in completing telephone-based surveys. Language, used as a proxy for area of residence, shows that participants responding in Amharic were more likely to complete the survey faster compared to those responding in Afaan Oromo. This suggests that urban participants, who are more likely to respond in Amharic, complete the survey faster than rural participants. Access to technology and familiarity with IVR systems in urban and rural areas are likely contributing factors.

To further inform our study, a debrief was organized with data collectors who enrolled participants in person and explained the IVR technology. During enrollment, patients revealed concerns with privacy, confusion about the technology, and a preference for in-person interviews. Some individuals were reluctant to share their mobile numbers due to fears of misuse, and face-to-face interactions were sometimes seen as more confidential. Language barriers also hindered participation, as some patients preferred languages that were not supported by our IVR survey. Technical challenges further complicate IVR use, as participants may struggle to navigate prompts without prior experience and with low educational backgrounds and poor literacy. Clear instructions are crucial for engagement. The study showed that older and less-educated individuals were

XSL•FO
RenderX

substantially less likely to participate in the phone-based study. This is consistent with evidence that older people and those with no education have limited digital competencies and thereby limited technological abilities [21,22,31].

Conclusions

IVR systems offer a promising tool for health system research in urban settings of a low-income country. Nonetheless, the implementation of an IVR survey among patients must be carefully planned and tailored to address local challenges. Addressing privacy concerns and building trust are essential. It is also necessary to ensure language inclusivity to improve participation rates and the overall effectiveness of IVR-based public health surveys. Others have found that providing incentives, such as mobile airtime top-ups, can encourage participation and improve response rates, especially for people of lower socioeconomic status [32,33]. Finally, using simple and common language, particularly for medical terms, may ensure better understanding among participants. In considering whether to implement an IVR survey as opposed to an in-person or a computer-assisted telephone interviewing survey with live interviewers, researchers should consider certain tradeoffs. First, we found that the IVR survey was cheap (US $7.90 per full survey); it was implemented quickly and is highly scalable to large samples as long as patient phone numbers can be obtained. However, although we did not find major data quality gaps, IVR may be more prone to data quality issues if users struggle to understand the questions or to answer correctly on their mobile phone keypads. In low-income country settings, in-person surveys may also be affordable if labor costs are low. However, in-person and computer-assisted telephone interviewing surveys take more time to implement and therefore are less scalable, as the number of interviewers and the time taken to conduct surveys must be multiplied when the sample size increases. With IVR, the survey is programmed once and can be answered simultaneously by thousands of respondents simultaneously. In addition, in-person surveys do not ensure privacy and are more prone to social desirability bias than IVR [6].

IVR is much cheaper than other survey methods. For example, health facility surveys such as the Service Provision Assessment surveys (conducted by the Demographic and Health Survey program) that include nationally representative samples of health facilities and patient exit interviews generally cost between US $800 and US $1500 per surveyed health facility [31]. Similarly, the World Bank estimated that a household survey in Sub-Saharan Africa will cost approximately US $300 per household. In addition, IVR surveys tend to be completed much faster. Our entire study was completed in 2 weeks (most of the time was taken for the in-person recruitment), and all data were immediately available through the IVR platform and ready for analysis [32].

There have been few studies in low- and middle-income countries that have tested the use of IVR to survey patients. To our knowledge, ours is the first study to use IVR to collect patient feedback in Ethiopia. IVR offers a low-cost and rapid solution to survey patients and obtain information on perceived quality and satisfaction. In high-income countries, patient satisfaction surveys are often conducted by mail on the web. These survey modes are largely not possible in Ethiopia. In Ethiopia, IVR surveys are likely to function well for younger and more educated respondents in an urban context. Future research should investigate feasibility and response rates in rural areas. Governments and health facilities should consider integrating IVR surveys into routine operations to gather patient satisfaction data and information on patient-perceived quality of care, which can inform quality improvement initiatives and strategies. By leveraging IVR technology, health systems can obtain timely and actionable feedback from patients, ultimately contributing to improved health care quality and better patient outcomes.

Acknowledgments

The authors thank colleagues from the Ethiopian Civil Service University and Addis Ababa University for their support in implementing the study as well as Alyssa Khan from the George Washington University for her support in the study design. This work was supported by a Research Innovation Award from the Office of Research Excellence, Milken Institute School of Public Health, The George Washington University, United States (grant 02RIA022023). Health facility directors were involved in the design and conduct of the study. Results were disseminated among health facility managers through policy briefs. Patients completed the survey and shared their opinions on the acceptability of the interactive voice response method.

Data Availability

The datasets generated and during this study are available from the corresponding author on reasonable request.

Authors' Contributions

DS and ADM contributed to designing the survey, data cleaning, data analysis, and report writing. CA contributed to designing the survey, data cleaning, data analysis, and editing and report writing.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Interactive voice response survey instrument on the quality of primary health care in Addis Ababa health facilities.
[DOCX File , 22 KB - formative_v9i1e67452_app1.docx]

References

1. Al-Abri R, Al-Balushi A. Patient satisfaction survey as a tool towards quality improvement. Oman Med J 2014;29(1):3-7 [FREE Full text] [doi: 10.5001/omj.2014.02] [Medline: 24501659]
2. Tiperneni R, Patel S, Heis F, Ghali S, Du D, Ghali W, et al. HCAHPS: having constant communication augments hospital and patient satisfaction. BMJ Open Qual 2022;11(4):e001972 [FREE Full text] [doi: 10.1136/bmjoq-2022-001972] [Medline: 36253015]
3. The mobile economy Sub-Saharan Africa. Global System for Mobile Communications. 2024. URL: https://www.gsma.com/mobileeconomy/sub-saharan-africa/ [accessed 2025-01-18]
4. Holbrook AL, Green MC, Krosnick JA. Telephone versus face-to-face interviewing of national probability samples with long questionnaires: comparisons of respondent satisficing and social desirability response bias. Public Opin Q 2003;67(1):79-125 [FREE Full text] [doi: 10.1086/346010]
5. Hughes S. The challenges and opportunities for collecting survey data in Sub-Saharan Africa. Mathematica. URL: https://www.mathematica.org/blogs/the-challenges-and-opportunities-for-collecting-survey-data-in-sub-saharan-africa [accessed 2019-02-04]
6. Gibson DG, Farrenkopf BA, Pereira A, Labrique AB, Pariyo GW. The development of an interactive voice response survey for noncommunicable disease risk factor estimation: technical assessment and cognitive testing. J Med Internet Res 2017;19(5):e112 [FREE Full text] [doi: 10.2196/jmir.7340] [Medline: 28476724]
7. Dillon B. Using mobile phones to collect panel data in developing countries. J Int Dev 2011;24(4):518-527 [FREE Full text] [doi: 10.1002/jid.1771]
8. Andersson C, Danielsson S, Silfverberg-Dymling G, Löndahl G, Johansson BA. Evaluation of interactive voice response (IVR) and postal survey in follow-up of children and adolescents discharged from psychiatric outpatient treatment: a randomized controlled trial. Springerplus 2014;3:77 [FREE Full text] [doi: 10.1186/2193-1801-3-77] [Medline: 24567883]
9. 5 Key benefits of interactive voice response (IVR). Primas. 2023. URL: https://primas.net/5-key-benefits-of-interactive-voice-response-ivr/ [accessed 2023-05-05]
10. Rigotti NA, Chang Y, Rosenfeld LC, Japuntich SJ, Park ER, Tindle HA, et al. Interactive voice response calls to promote smoking cessation after hospital discharge: pooled analysis of two randomized clinical trials. J Gen Intern Med 2017;32(9):1005-1013 [FREE Full text] [doi: 10.1007/s11606-017-4085-z] [Medline: 28616847]
11. Altman C, Birraux T, Lapuelle J, Letard JC, Tusseau F, Canard JM. Assessment of patient satisfaction with endoscopy using an interactive voice response system. Gastroenterol Clin Biol 2006 Mar;30(3):371-376 [FREE Full text] [doi: 10.1016/s0399-8320(06)73190-6] [Medline: 16633301]
12. Holland SM, Shuk E, Burkhalter J, Shouery M, Li Y, Hay JL. Feasibility and acceptability of using an IVRS to assess decision making about sun protection. Psychooncology 2020;29(1):156-163 [FREE Full text] [doi: 10.1002/pon.5231] [Medline: 31520426]
13. Daftary A, Hirsch-Moverman Y, Kassie GM, Melaku Z, Gadisa T, Saito S, et al. A qualitative evaluation of the acceptability of an interactive voice response system to enhance adherence to isoniazid preventive therapy among people living with HIV in Ethiopia. AIDS Behav 2017;21(11):3057-3067 [FREE Full text] [doi: 10.1007/s10461-016-1432-8] [Medline: 27221743]
14. Howard AA, Hirsch-Moverman Y, Saito S, Gadisa T, Daftary A, Melaku Z. The ENRICH Study to evaluate the effectiveness of a combination intervention package to improve isoniazid preventive therapy initiation, adherence and completion among people living with HIV in Ethiopia: rationale and design of a mixed methods cluster randomized trial. Contemp Clin Trials Commun 2017;6:46-54 [FREE Full text] [doi: 10.1016/j.conctc.2017.03.001] [Medline: 28626811]
15. Leo B, Morello R, Mellon J, Peixoto T, Davenport ST. Do mobile phone surveys work in poor countries? SSRN J 2015:398. [doi: 10.2139/ssrn.2597885]
16. Pegus CM. Crowdsourcing citizen feedback on district development in Ghana using interactive voice response surveys. The Institute of Development Studies. 2016. URL: https://opendocs.ids.ac.uk/articles/report/Crowdsourcing_citizen_feedback_on_district_development_in_Ghana_using_interactive_voice_response_surveys/26476822?file=48249943 [accessed 2025-01-18]
17. Wooldridge JM. Econometric Analysis of Cross Section and Panel Data. Cambridge: MIT press; 2010.
18. Koné S, Bonfoh B, Dao D, Koné I, Fink G. Heckman-type selection models to obtain unbiased estimates with missing measures outcome: theoretical considerations and an application to missing birth weight data. BMC Med Res Methodol 2019;19(1):231. [doi: 10.1186/s12874-019-0840-7] [Medline: 31815610]
19. Ethiopia: Demographic and Health Survey. Central Statistical Agency (CSA) [Ethiopia] and ICF. 2016. URL: https://www.dhsprogram.com/pubs/pdf/FR328/FR328.pdf [accessed 2025-01-18]
20. Chróinín DN, Hunter C, Shaw-Jones M, Duruchukwu E, Huang F, Ngov A, et al. Identifying the needs of our older community-based patients in the COVID era: a telephone survey. Aging Health Res 2022;2(2):100073 [FREE Full text] [doi: 10.1016/j.ahr.2022.100073] [Medline: 35434693]

XSL•FO
RenderX

21. Chen K, Chan AHS. Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM). Ergonomics 2014;57(5):635-652 [FREE Full text] [doi: 10.1080/00140139.2014.895855] [Medline: 24655221]
22. Marasinghe KM, Lapitan JM, Ross A. Assistive technologies for ageing populations in six low-income and middle-income countries: a systematic review. BMJ Innov 2015;1(4):182-195 [FREE Full text] [doi: 10.1136/bmjinnov-2015-000065] [Medline: 26688747]
23. Ultimate guide to interactive voice response (IVR) for social impact programs. EngageSpark. URL: https://www.engagespark.com/guides/ultimate-guide-interactive-voice-response-ivr-social-impact-programs [accessed 2025-01-16]
24. Rutebemberwa E, Namutundu J, Gibson DG, Labrique AB, Ali J, Pariyo GW, et al. Perceptions on using interactive voice response surveys for non-communicable disease risk factors in Uganda: a qualitative exploration. Mhealth 2019;5:32 [FREE Full text] [doi: 10.21037/mhealth.2019.08.12] [Medline: 31620459]
25. Groves RM, Couper MP. Nonresponse in Household Interview Surveys. Hoboken, NJ: John Wiley & Sons; 2012.
26. Dillman DA, Smyth JD, Christian LM. Internet, phone, mail and mixed-mode surveys: the tailored design method. Reis 2016;154:161-176. [doi: 10.1002/9781394260645]
27. Smith MG, Witte M, Rocha S, Basner M. Effectiveness of incentives and follow-up on increasing survey response rates and participation in field studies. BMC Med Res Methodol 2019;19(1):230-213 [FREE Full text] [doi: 10.1186/s12874-019-0868-8] [Medline: 31805869]
28. Phadnis MA, Sharma P, Thewarapperuma N, Chalise P. Assessing accuracy of Weibull shape parameter estimate from historical studies for subsequent sample size calculation in clinical trials with time-to-event outcome. Contemp Clin Trials Commun 2020;17:100548 [FREE Full text] [doi: 10.1016/j.conctc.2020.100548] [Medline: 32154431]
29. Gómez YM, Gallardo DI, Marchant C, Sánchez L, Bourguignon M. An in-depth review of the Weibull model with a focus on various parameterizations. Mathematics 2023;12(1):56 [FREE Full text] [doi: 10.3390/math12010056]
30. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
31. Bertolazzi A, Quaglia V, Bongelli R. Barriers and facilitators to health technology adoption by older adults with chronic diseases: an integrative systematic review. BMC Public Health 2024;24(1):506 [FREE Full text] [doi: 10.1186/s12889-024-18036-5] [Medline: 38365698]
32. Abdelazeem B, Abbas KS, Amin MA, El-Shahat NA, Malik B, Kalantary A, et al. The effectiveness of incentives for research participation: a systematic review and meta-analysis of randomized controlled trials. PLoS One 2022;17(4):e0267534 [FREE Full text] [doi: 10.1371/journal.pone.0267534] [Medline: 35452488]
33. Ichimiya M, Muller-Tabanera H, Cantrell J, Bingenheimer JB, Gerard R, Hair EC, et al. Evaluation of response to incentive recruitment strategies in a social media-based survey. Digit Health 2023;9:20552076231178430 [FREE Full text] [doi: 10.1177/20552076231178430] [Medline: 37312941]

Abbreviations

IVR: interactive voice response
OR: odds ratio

Edited by A Mavragani; submitted 11.10.24; peer-reviewed by A Greenleaf; comments to author 18.11.24; revised version received 08.12.24; accepted 24.12.24; published 13.02.25.

Please cite as:
Shamebo D, Derseh Mebratie A, Arsenault C
Using an Interactive Voice Response Survey to Assess Patient Satisfaction in Ethiopia: Development and Feasibility Study
JMIR Form Res 2025;9:e67452
URL: https://formative.jmir.org/2025/1/e67452
doi:10.2196/67452
PMID:

©Dessalegn Shamebo, Anagaw Derseh Mebratie, Catherine Arsenault. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring Metadata Catalogs in Health Care Data Ecosystems: Taxonomy Development Study

Simon Scheider[1], PhD; Mostafa Kamal Mallick[1], MSc

Fraunhofer Institute for Software and Systems Engineering, Dortmund, Germany

Corresponding Author:
Simon Scheider, PhD
Fraunhofer Institute for Software and Systems Engineering
Speicherstraße 6
Dortmund, 44147
Germany
Phone: 49 231976774
Email: simon.scheider@isst.fraunhofer.de

Abstract

Background: In the European health care industry, recent years have seen increasing investments in data ecosystems to "FAIRify" and capitalize the ever-rising amount of health data. Within such networks, health metadata catalogs (HMDCs) assume a key function as they enable data allocation, sharing, and use practices. By design, HMDCs orchestrate health information for the purpose of findability, accessibility, interoperability, and reusability (FAIR). However, despite various European initiatives pushing health care data ecosystems forward, actionable design knowledge about HMDCs is scarce. This impedes both their effective development in practice and their scientific exploration, causing huge unused innovation potential of health data.

Objective: This study aims to explore the structural design elements of HMDCs, classifying them alongside empirically reasonable dimensions and characteristics. In doing so, the development of HMDCs in practice is facilitated while also closing a crucial gap in theory (ie, the literature about actionable HMDC design knowledge).

Methods: We applied a rigorous methodology for taxonomy building following well-known and established guidelines from the domain of information systems. Within this methodological framework, inductive and deductive research methods were applied to iteratively design and evaluate the evolving set of HMDC dimensions and characteristics. Specifically, a systematic literature review was conducted to identify and analyze 38 articles, while a multicase study was conducted to examine 17 HMDCs from practice. These findings were evaluated and refined in 2 extensive focus group sessions by 7 interdisciplinary experts with deep knowledge about HMDCs.

Results: The artifact generated by the study is an iteratively conceptualized and empirically grounded taxonomy with elaborate explanations. It proposes 20 dimensions encompassing 101 characteristics alongside which FAIR HMDCs can be structured and classified. The taxonomy describes basic design characteristics that need to be considered to implement FAIR HMDCs effectively. A major finding was that a particular focus in developing HMDCs is on the design of their published dataset offerings (ie, their metadata assets) as well as on data security and governance. The taxonomy is evaluated against the background of 4 use cases, which were cocreated with experts. These illustrative scenarios add depth and context to the taxonomy as they underline its relevance and applicability in real-world settings.

Conclusions: The findings contribute fundamental, yet actionable, design knowledge for building HMDCs in European health care data ecosystems. They provide guidance for health care practitioners, while allowing both scientists and policy makers to navigate through this evolving research field and anchor their work. Therefore, this study closes the research gap outlined earlier, which has prevailed in theory and practice.

(JMIR Form Res 2025;9:e63396) doi:10.2196/63396

KEYWORDS

data catalogs; data ecosystems; findability, accessibility, interoperability, and reusability; FAIR; health care; metadata; taxonomy

XSL•FO
RenderX

Introduction

Challenges of Health Care Systems

In the 21st century, health care systems worldwide are experiencing a tremendous increase in data, driven by advances in medical technology, digital health records, and wearable devices [1]. This flood of data holds immense potential for data-driven health innovations, building upon large-scale real-world data (RWD) and real-world evidence (RWE) [2]. However, health care systems face multiple challenges that hinder the effective use of RWD to generate RWE and thus data-driven health innovations. One primary issue is the integration of heterogeneous datasets [3]. RWD stem from diverse sources, such as electronic health records, imaging, and genomic data, frequently exhibiting incompatible or unknown data formats, which complicates harmonization, particularly across different entities [4,5]. Furthermore, finding and accessing suitable RWD represents a hurdle for medical research due to their origins from disparate patient populations, health care systems, and data collection methodologies [6]. This impairs the effective discovery of and access to a sufficient number of both available and adequate datasets. Moreover, even if enough RWD are discovered and access is established, another challenge lies in ensuring scientific rigor and reproducibility of generated RWE (ie, medical studies) that becomes increasingly difficult in today's data-intensive health research [7]. Medical studies constantly require larger, high-quality datasets to generate meaningful and reproducible RWE [5,6,8]. However, unknown RWD management practices threaten study reliability, while making the validation of results (ie, RWE) across studies difficult [2,7]. Besides that, the diversity of national health care systems increases the prevailing differences in health care data infrastructures across countries that, in turn, lead to additional barriers for organizations to share and use RWD at a large scale [9]. Moreover, health care systems must navigate complex legal requirements with regard to sharing and processing RWD [8,9]. For instance, the European jurisdiction mandates strict data governance and security standards that, while essential, impair data-driven health innovations in the absence of adequate data-sharing infrastructures [9]. As a result, RWD are fragmented and isolated within single organizations, whereby data sharing and use are limited. Because of all these challenges, the rapidly increasing amount of RWD cannot be harnessed to its full potential for producing health care innovations (ie, RWE).

Metadata Catalogs as a Promising Solution

Against this background, data ecosystems as technical and organizational infrastructures within the healthcare sector represent auspicious solutions, that is, *health care data ecosystems* (amplified in, eg, the studies by Lovestone and EMIF Consortium [10], Manogaran et al [11], and Sharon and Lucivero [12]). These evolving networks enable legally compliant use of RWD [13]. Their key function is market mechanisms instantiated by metadata catalogs [5,14] that define and describe the intricate web of RWD circulating between a potentially arbitrary number of actors in the ecosystem [15]. Hence, such *health metadata catalogs* (HMDCs) are crucial components of modern health care data ecosystems, for example, EHDEN (European Health Data and Evidence Network), EHDS2 (European Health Data Space 2), Elixir, EUCAIM (European Federation for Cancer Images), IDERHA (Integration of Heterogeneous Data and Evidence towards Regulatory and Health Technology Assessment Acceptance), and Gaia-X. Multimedia Appendix 1 provides a comprehensive overview about the most important European Union (EU) initiatives. For example, EHDS2 aims at creating a European infrastructure for the secure exchange and secondary use of health data across EU member states [16]. Therein, HMDCs belong to the core infrastructure services to enable standardized organization of and controlled access to RWD for research. The pilot infrastructure of the EHDS is implemented by the HealthData@EU initiative.

Since HMDCs provide an effective method for systematically sharing and using RWD within data ecosystems [3], they potentially allow harnessing crucial benefits corresponding to the challenges outlined earlier. First, HMDCs facilitate integrating heterogenous datasets across health care systems. They help to transcend diverse data types, which eases the integration, standardization, and harmonization of data within data ecosystems [3,5]. This is essential for medical research that requires huge pools of accessible RWD appropriate for their investigations [7]. Second, as finding and accessing adequate RWD effectively is vital for medical research [6], HMDCs entail added value by offering a governed data search and access framework embedded into the technical infrastructures of the underlying ecosystems [17]. They provide a tool for data discovery to precisely characterize, locate, and filter RWD on the basis of a myriad of factors [5,17]. Third, HMDCs support transparent and reproducible research processes by helping scientists in replicating studies and validating their results [18]. Such transparency is fundamental for building trust in the reliability and validity of RWE [2]. Finally, HMDCs facilitate data-intensive research, generally, as they establish unified health care data infrastructures for allocating, accessing, and using RWD of connected data providers [19]. In doing so, they reduce barriers for organizations to integrate their otherwise isolated RWD within data ecosystems. At the same time, HMDCs bridge prevailing differences between national health care systems and retain full control of data providers [5,19]. To this end, they establish robust data security and governance frameworks that are aligned to the applicable jurisdictions [5,15].

As a result, HMDCs represent an auspicious medium against the fragmentation and isolation of RWD [5]. However, since HMDCs are novel constructs, typically in premature phases [20], their ascribed benefits are primarily backed by the literature rather than evidence from practice. Nevertheless, HMDCs are likely to provide means for using RWD systematically within and across health care data ecosystems, potentially resulting in more efficient RWE generation.

In Europe, HMDCs are of particular importance because the EU health care sector exhibits a broad diversity across member states, all with their own health care systems and policies. Consequently, there is a need to focus heavily on standardizing and harmonizing both data and metadata across different countries for facilitating data sharing and legally compliant data

XSL•FO
RenderX

use [21]. More specifically, the diversity of national health care systems [22], the restrictiveness of data protection regulations [8,9], and the fragmentation and isolation of health data [23] make operative health care data ecosystems and HMDCs a paramount concern for the European health care industry. Therefore, this study adopts a European focus.

Theoretical Background

Originally, data catalogs are organized collections of datasets that provide descriptive information within an organization [24,25]. They act as centralized repositories, making it easier for data consumers to discover, understand, and access the information they need [26]. Enterprise data management platforms often comprise such centralized data catalogs implying storage of data within their peripheries [25,27]. If data are not encapsulated within the organization but integrated into decentral or federated networks [28], the literature commonly refers to such environments as data ecosystems with metadata catalogs as key function [5,14]. This study considers metadata catalogs as decentralized or federated constructs that are mutually exclusive to centralized ones. For simplification, only the term decentralized is used.

Metadata describe dataset attributes, such as source, format, structure, provenance, owner, access, or governance modalities [29]. Metadata catalogs act as "catalogues of data catalogues," dedicated to enhancing discoverability, usability, and management of distributed datasets [30]. Within data ecosystems, metadata catalogs are a mechanism that provides a standardized way for recording, disclosing, and making available information about all relevant kinds of phenotypes describing datasets, while ensuring legally compliant access and sharing practices [3,14]. If these datasets are health data, it is henceforth referred to such constructs as *HMDCs*. Consequently, HMDCs manage heterogenous health information integrated into health care data ecosystems [5]. They ensure that these diverse and highly sensitive datasets are effectively organized and understood [4], while facilitating their systematic use [3,5,14]. This requires dedicated, yet unknown, design elements to be unveiled by the study.

Research Gap, Objective, and Questions

After having clarified the added value of HMDCs, the research gap is demarcated by reviewing related work. Therefrom, the research problem is identified, which leads to the research objectives. These objectives then allow to derive the research questions, required to define a meaningful research methodology.

To begin with, Labadie et al [24] foster the understanding of data catalogs by classifying corresponding initiatives. The authors propose a taxonomy for data catalogs and present 3 case studies. However, similar to Ehrlinger et al [26] and Jahnke and Otto [25], Labadie et al [24] focus on intraorganizational data sharing using centralized catalogs. Moreover, they neglected health use cases. Remy et al [3] conducted a design science study to build an integrated catalog for health research metadata. The artifact enables medical scientists to analyze phenomena that require a view across several domains. The authors are among the first who provide design knowledge usable in HMDC contexts. Although, similar to the findings of the previously presented literature sources, Remy et al [3] accentuate centralized catalogs. Almeida et al [19] present a platform that provides a set of tools, compliant with the findability, accessibility, interoperability, and reusability (FAIR) principles, to help data holders sharing biomedical databases while allowing data consumers to discover and apply for them. However, the authors only consider a narrow use case instead of generating universally applicable design knowledge. Similarly, Oliveira et al [15] developed a holistic stakeholder agnostic catalog framework for biomedical datasets. Researchers can explore metadata held decentralized at federated nodes, with distinct levels of granularity being conceivable. Extending this initial design knowledge specific to biomedical data, Swertz et al [5] proposed a unified framework for sharing health data across catalogs. It encompasses multiple centralized and decentralized catalogs. The authors offer recommendations to establish an integrated community as an open catalog ecosystem. This theoretical basis for HMDCs builds upon and is enriched by similar research. Specifically, Bergeron et al [17] developed a catalog toolkit to support creating comprehensive as well as user- and study-friendly HMDCs. Almeida and Oliveira [30] produced a framework to simplify the process of building an HMDC for exposing metadata, while providing analysis capacities. Apparently, there is a tendency from centralized to decentralized data catalogs in health care. However, for HMDCs, a *research gap* prevails concerning (1) empirically grounded and actionable design knowledge that is (2) universally applicable to (3) the broad array of use cases and EU initiatives associated with health care data ecosystems.

In general, the generation of design knowledge about an artifact is crucial as it provides the intellectual foundation to advance the respective body of scientific knowledge, while facilitating development efforts in practice [31]. In particular, HMDC design knowledge can harmonize and sustain the multitude of different EU initiatives by following a systematic approach to problem-solving [31,32]. Therefore, its generation must adhere to a rigor design process [33,34]. This process must ensure empirical grounding for the sake of efficiency, effectiveness, and quality assurance, which inevitably favors quality, adaptiveness, and impact of the generated results [35]. Likewise, the prevailing lack of design knowledge causes difficulties concerning the adoption and use of HMDCs in practice and theory, revealing the *research problem*. To remedy this problem, the *research objective* is to provide actionable design knowledge that is universally applicable in real-world HMDC use cases, thus allowing to infer the following *research questions* (RQs):

- RQ1:
  What are taxonomy elements (ie, dimensions and characteristics) to structure HMDCs from a design science perspective?
- RQ2:
  How does the proposed taxonomy effect real-world use cases?

According to Hevner et al [33], a *design science* perspective means to examine and create information system (IS) artifacts to solve practical problems. A taxonomy is a suitable approach to address RQ1 because it provides a set of elementary building

XSL•FO
RenderX

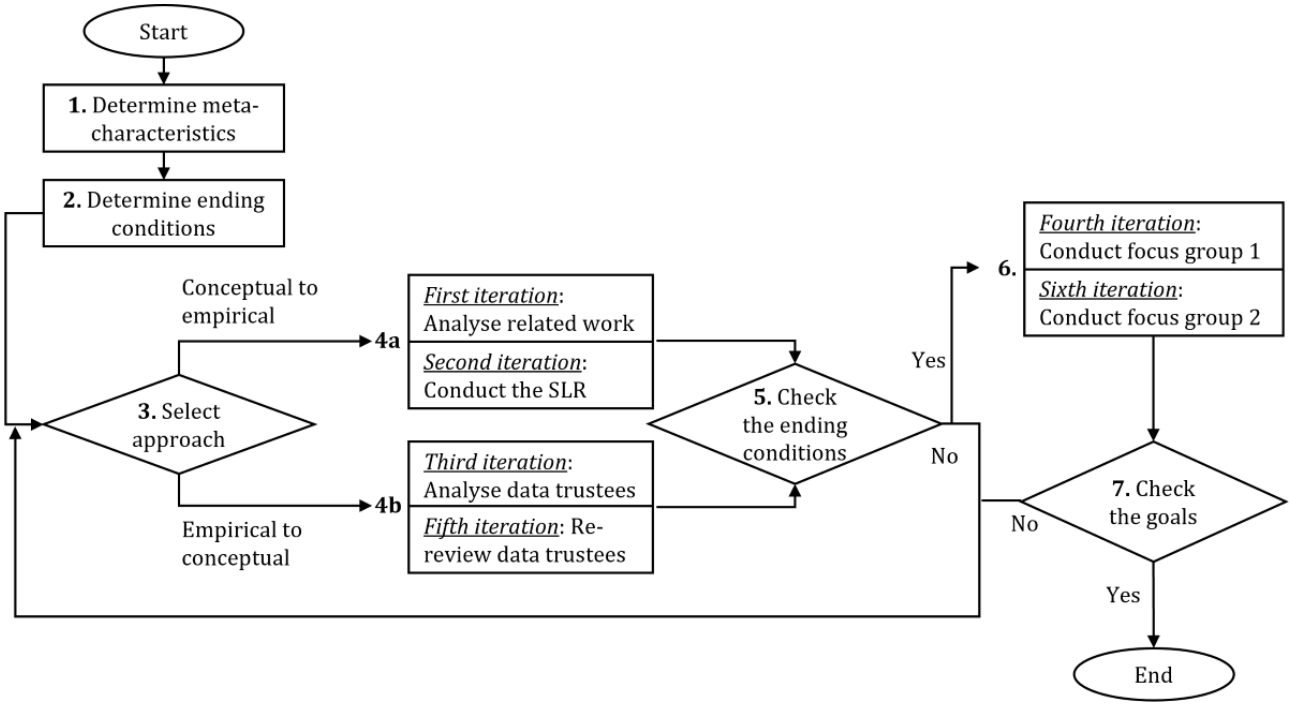
blocks and prescriptions for effectively designing such artifacts [36,37]. It targets a broad and diverse audience, including health care IS engineers and architects, health data holders and scientists, health care economists and researchers, as well as legal and ethical regulatory bodies, while accentuating the European health care sector.

Methods

Overview

Taxonomies are common approaches in IS research to classify, understand, and examine complex issues [38]. For their development, the method of Nickerson et al [37] is applied to identify dimensions and characteristics of HMDCs. The authors propose generating knowledge conceptually (eg, from the literature) and empirically (eg, analyzing objects of interest). This approach is referred to as the gold standard to build taxonomies in IS research [36]. As refinement, the methodological update of Kundisch et al [36] is incorporated, adding an evaluation process by means of focus groups. The authors' refinement enhances the assessment of value created by the taxonomy [39]. Corresponding to these 2 methods, the research design is divided into the 7 steps shown in Figure 1 based on the studies by Nickerson et al [37] and Kundisch et al [36]. The numbers 1 to 7 represent methodological steps explained in the following sections.

Figure 1. Applied research method of the taxonomy development study based on Nickerson et al [37] and Kundisch et al [36].



In step 1 in Figure 1, a *meta-characteristic* is specified in orientation towards the taxonomy's purpose so that each subordinated characteristic and dimension follows from it. On the basis of RQ1, the meta-characteristic was defined as "distinguishing key design elements of HMDCs." It facilitates selecting meta-dimensions as well as inferring characteristics and classifying them to dimensions. To define the meta-dimensions, the FAIR framework was used [7]. These well-known data principles postulate an accepted approach to the discoverability and usability of RWD [19]. While FAIR emphasizes making data interoperable and reusable, it inherently involves considerations related to data governance and harmonization [40].

In step 2 in Figure 1, *ending conditions* for the iterative part of the process are defined, determining its termination criteria. The ending conditions were chosen on the basis of Nickerson et al [37] and Scheider et al [41] in terms of subjective and objective criteria. Ultimately, 6 design iterations were required until all conditions listed in Table 1 were fulfilled.

XSL•FO
RenderX

Table 1. Ending conditions for taxonomy development adopted from Nickerson et al [37].

Ending conditions	Design iterations					
	1	2	3	4	5	6
Objective						
All papers were examined.			✓	✓	✓	✓
No object was merged with another or split.			✓		✓	✓
Each characteristic is classified by one object.	✓	✓	✓	✓	✓	✓
No new dimensions or characteristics were added.			✓		✓	✓
Dimensions or characteristics were neither merged nor split.			✓		✓	✓
Each dimension is unique and not duplicated.	✓	✓	✓	✓	✓	✓
Every characteristic is unique within its dimension.	✓	✓	✓	✓	✓	✓
Each cell is unique and not repeated.	✓	✓	✓	✓	✓	✓
Subjective						
Conciseness: no unnecessary dimensions and characteristics			✓	✓	✓	✓
Robust: dimensions and characteristics differentiate objects			✓	✓	✓	✓
Comprehensiveness: all objects can be classified			✓	✓	✓	✓
Extension: dimensions and characteristics can be added easily		✓	✓	✓	✓	✓
Explanatory: dimensions and characteristics describe all objects			✓		✓	✓

In steps 3 to 5 in Figure 1, we repeatedly chose between either an inductive or a deductive path. The former is a conceptual-to-empirical attempt (step 4a in Figure 1) to infer dimensions and characteristics from theory. The latter reflects an empirical-to-conceptual procedure (E2C; step 4b in Figure 1) to derive characteristics from real-world analysis objects and to classify them in dimensions. After each iteration (ie, steps 3 to 5 in Figure 1), the ending conditions are checked (ie, step 5 in Figure 1). If all ending conditions are fulfilled, an evaluation step (ie, step 6 in Figure 1) follows, integrated by focus groups [36]. In case the focus group iteration does not imply changes to the taxonomy (ie, step 7 in Figure 1), the artifact is finished and the methodological process terminates. After 5 design iterations (ie, 4 times of executing steps 4a and 4b and executing step 6 once in Figure 1), all ending conditions were fulfilled (ie, step 5 in Figure 1) and the subsequent focus group did not result in any major changes (ie, step 7 in Figure 1). Thus, the taxonomy was completed [36]. Because 6 design iterations were traversed and it was ensured that the focus group experts covered all dimensions relevant for HMDCs, the taxonomy achieved result saturation.

To ensure transparency of the taxonomy development process, Multimedia Appendix 2 shows intermediary taxonomies after certain iterations. Furthermore, it offers a table linking key references from the inductive (ie, the literature) and deductive iterations (ie, analysis objects) to the dimensions of the final taxonomy [5,10,14,16,17,19,24,28,30,41-78].

Inductive Design Iterations for Taxonomy Development

In the first iteration, an initial set of dimensions and characteristics was derived from former research (ie, step 4a in Figure 1), consolidating the related work addressed in the Introduction section.

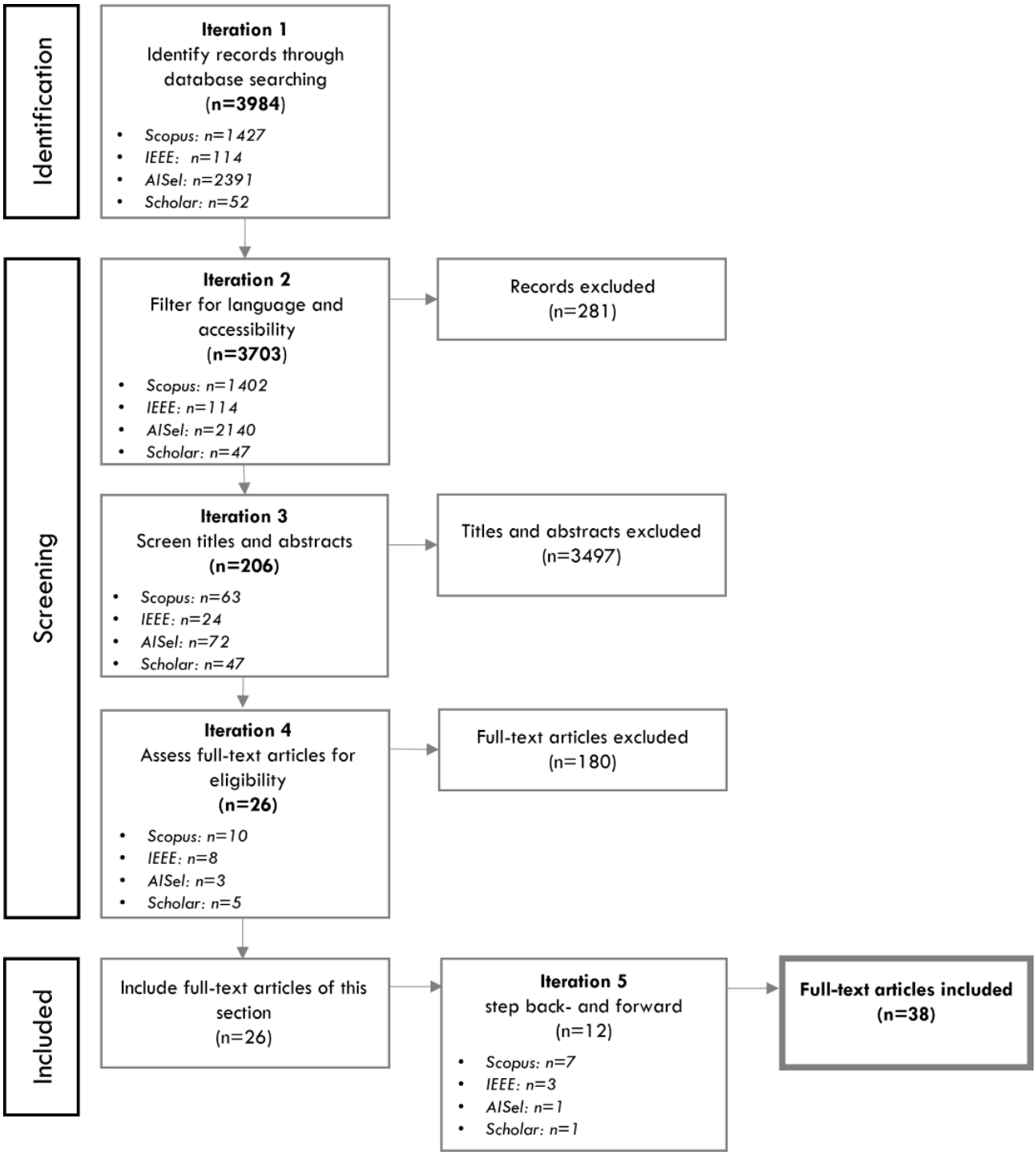
In the second iteration, a structured literature review (SLR) was carried out (ie, step 4a in Figure 1) [36]. The method of Kitchenham et al [79] was applied (ie, 1-6), while orienting toward its application in the study by Scheider et al [41]. At the outset, RQ1 was adopted as the (1) research question guiding the SLR. The (2) search process comprised HMDC-related conference and journal papers. The search string was defined as (ALL (health AND data AND catalog) OR ALL (health AND metadata AND catalog) AND ALL (data AND catalog AND technologies)). Primarily, Scopus and IEEE Xplore were used, and the operands were deployed on documents' titles, abstracts, and authors' keywords. The 2 databanks were leveraged due to their multidisciplinary nature covering research in all fields relevant for HMDCs. Following Scheider et al [41], (3) inclusion and exclusion criteria were created to identify and filter papers. First, the literature not available in English was excluded. Second, inaccessible papers were removed. Third, each paper retrieved was reviewed by 2 researchers for whether it covers HMDCs in the broader sense. This means that papers had to address a design perspective, as defined by Hevner et al [33]. Articles were emphasized that dealt with "patient-related" data, while ones about aggregated health data (eg, regions and countries) were neglected. The same holds true for catalogs about health-oriented surveys and analysis results (eg, studies). Due to broadly formulated keywords in the search string, initially retrieved literature contained many papers outside the thematical scope. To this end, the third inclusion or exclusion criterion was examined by screening titles and abstracts before reviewing the entire content of the papers. Since 2 researchers constantly worked together in (3), one can argue for reliable objectivity in paper selection.

Building upon the inclusion and exclusion criteria, the initial (4) data collection resulted in 18 papers in the Scopus and IEEE

XSL•FO
RenderX

Xplore search (iteration 4 in Figure 2). Subsequently, backward (ie, referenced articles) and forward (ie, citing articles) stepping was conducted [80], which added 12 articles. Figure 2 shows the SLR statistics expressed by a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.

Figure 2. Structured literature review statistics presented in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.



The SLR was expanded by a search via Google (Google Search) and the AISeL database for extension and verification. The Google search engine served to identify whitepapers using a consolidated search string compared to (2). For AISeL, the same steps executed in Scopus and IEEE Xplore were applied (ie, 2 and 3), except that the research team looked for the search terms in titles only to keep the number of results feasible. Once duplicates were removed, Google and AISeL added 8 papers to the literature collection. To test theoretical saturation [81], "quick searches" were carried out in other databases (eg, ACM) checking whether the top results, first, match the inclusion and exclusion criteria and, second, are not already in the collection. Since these quick searches did not add new papers, the literature collection was considered representative [41]. Excluding related work of the first iteration, the collection comprised 38

XSL•FO
RenderX

publications, of which the most important items are listed in Table 2.

Table 2. List of most important literature from the SLR[a] used for taxonomy development.

Study	Year	Title
Top 10 academic papers		
Alvarellos et al [42]	2023	Democratizing clinical-genomic data: How federated platforms can promote benefits sharing in genomics
Almeida and Oliveira [30]	2024	MONTRA2[b]: A web platform for profiling distributed databases in the health domain
Almeida et al [19]	2023	A FAIR[c] approach to real-world health data management and analysis
Bergeron et al [17]	2018	Fostering population-based cohort data discovery: The Maelstrom Research cataloguing toolkit
Scheider et al [41]	2023	Exploring design elements of personal data markets
Ehrlinger et al [26]	2021	Data catalogs: a systematic literature review and guidelines to implementation
Labadie et al [24]	2020	Fair enough? enhancing the usage of enterprise data with data catalogs
Lovestone and EMIF Consortium [10]	2020	The European medical information framework: a novel ecosystem for sharing health care data across Europe
Oliveira et al [15]	2019	EMIF[d] Catalogue: a collaborative platform for sharing and reusing biomedical data
Swertz et al [5]	2022	Towards an Interoperable Ecosystem of Research Cohort and Real-world Data Catalogues Enabling Multi-center Studies
Top 5 nonacademic papers		
European Medicines Agency [43]	2022	Good Practice Guide for the use of the Metadata Catalogue of Real-World Sources
European Medicines Agency [44]	2022	List of metadata for Real World Data catalogues
Directorate-General for Health and Food Safety [82]	2022	The European Health Data Space
Jahnke and Otto [25]	2022	Data Catalogs - Implementing Capabilities for Data Curation, Data Enablement and Regulatory Compliance
TEHDAS[e] [16]	2022	EHDS[f] Semantic interoperability framework

[a]SLR: structured literature review.

[b]MONTRA2: Modular Next-generation Research Analysis.

[c]FAIR: findability, accessibility, interoperability, and reusability.

[d]EMIF: European Medical Information Framework Catalogue.

[e]TEHDAS: Towards European Health Data Space.

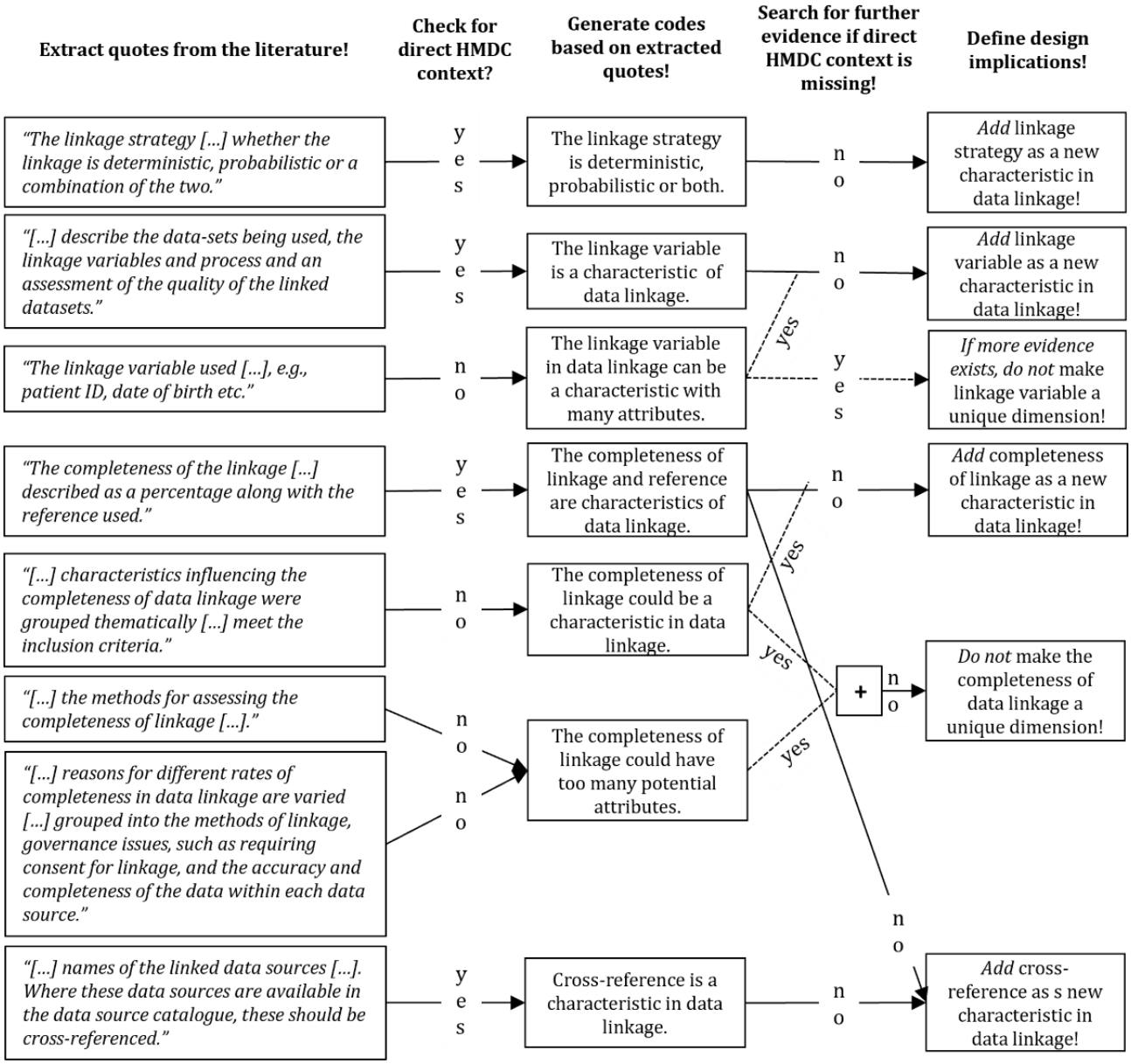
[f]EHDS: European Health Data Space.

Throughout the steps (2) to (4), a (5) quality assessment step was integrated on the basis of the criteria suggested by Kitchenham et al [79], that is, inclusion or exclusion criteria, relevant article coverage, literature corpus assessment, and study descriptions.

For (6) data analysis, phrases (“quotes”) from articles with useful content for HMDC designs were extracted. Following the approaches of Saldana [83] and Pratt [84], those phrases were coded, inserted in a tabular structure, and iteratively generalized. As in steps (3) and (9), two researchers analyzed the literature to reduce subjectivity biases. Figure 3 shows how quote extractions relating to the dimension of data linking are coded and design implications are derived. Particularly, whenever there was a direct connection to an HMDC context, quotes became design implications immediately, for example, linkage strategy (first quote in Figure 3). If a direct connection was missing (eg, linkage variable as a new characteristic; third quote in Figure 3), more evidence was required to transform codes into design implications (ie, second quote in Figure 3). Finally, considering the influential factors proposed by Mwita [81] (eg, study purpose, research design, sample size variability, and analysis approach), data saturation in the SLR is likely.

XSL•FO
RenderX

Figure 3. Examples for literature coding for inductive taxonomy development.



Deductive Design Iterations for Taxonomy Development

Applying the E2C approach (ie, step 4b Figure 1) in the third and fifth iterations, health data catalogs from practice were initially listed, as identified in the first 2 iterations. This set was extended by a Google search to identify analysis objects not encountered in inductive iterations. The research team searched for analysis objects using the browser's incognito mode to circumnavigate carryover effects from previous searches [85]. The keywords from the SLR were used as an orientation to avoid limiting the results unconsciously [41]. Analysis objects were excluded if meaningful information could not be obtained. This characteristic was defined as access to analyzable information describing the analysis object, that is, data retrievable either from websites or demo applications [85]. Analysis objects were also excluded if information was meaningful but unavailable in German or English. However, metadata catalogs under construction were not excluded per se [41]. The set of analysis objects was created in the first quarter of 2024. The final analysis objects are listed in Table 3.

XSL•FO
RenderX

Table 3. Health metadata catalogs from practice used for taxonomy development.

HMDCs[a]	Classification	Status
BBMRI-ERIC[b] Data Directory [45]	Decentral	Operative
Catalogue of Mental Health Measures [46]	Central	Operative
Compendium Data Catalog for Healthcare [47]	Central	Operative
EHDEN[c] Portal [48]	Decentral	Operative
Elixir (BioSamples [86] and FAIRsharing [87])	Decentral	Operative
EMIF[d] Data Catalogue [49]	Decentral	In progress
EUCAIM[e] Cancer Image Europe [50]	Decentral	Operative
European Health Information Portal [51]	Central	Operative
Fjelltopp Data Catalogues for Health [52]	Central	Operative
HealthRI[f] Data Catalogues [53]	Decentral	Operative
Helsedata Explore Data Sources [54]	Decentral	Operative
IDERHA[g] Metadata Catalogue (no public access)	Decentral	In progress
IHME[h] Global Health Data Exchange [88]	Central	Operative
IQVIA[i] Health Data Catalogue [55]	Central	Operative
Kraken Health Data Pilot [56]	Decentral	In progress
Lifebit Precision Medicine Data Catalogue [57]	Decentral	Operative
MACH[j] Clinical and Research Data Catalogue [89]	Decentral	Operative
Maelstrom Research Data Catalogue [58]	Decentral	Operative
Yoda Trials Data Catalogue [59]	Decentral	In progress

[a]HMDC: health metadata catalog.

[b]BBMRI-ERIC: Biobanking and Biomolecular Resources Research Infrastructure – European Research Infrastructure Consortium.

[c]EHDEN: European Health Data and Evidence Network.

[d]EMIF: European Medical Information Framework.

[e]EUCAIM: European Federation for Cancer Images.

[f]HealthRI: Health Research Infrastructure.

[g]IDERHA: Integration of Heterogeneous Data and Evidence towards Regulatory and Health Technology Assessment Acceptance.

[h]IHME: Institute for Health Metrics and Evaluation.

[i]IQVIA: Information, Quality, Value, Innovation, and Access.

[j]MACH: Melbourne Academic Centre for Health.

The catalogs were examined by classifying them alongside the design elements of the preliminary taxonomy (Multimedia Appendix 2). We tried to assign a catalog to a single characteristic in each dimension. This deductive activity was conducted by 2 researchers, whereby three cases could occur [41]: (1) on the basis of available information, the analysis object could be assigned to an existing characteristic in a dimension; (2) due to a lack of available information, the analysis object could not be assigned to any characteristic in a dimension; and (3) the analysis object contained information about a dimension but could not be associated with a characteristic defined therein. Evidently, the third case led to modifications of a current set of taxonomy design elements. During the third iteration, the occurrence of this case decreased continuously until only the first and second ones appeared. Because the ending conditions were fulfilled at that time, the taxonomy-building process was terminated and the first focus group session started. Alterations were proposed by focus group members in this session. This led to another E2C approach as the fifth iteration in which specific design elements were emphasized, as addressed by the focus group experts. The corresponding analysis procedure (ie, 1-3) remained the same as in the third iteration, while examining decentral analysis objects only (Table 3).

Evaluative Design Iterations for Taxonomy Development and Use Case Cocreation

Following Figure 1, an evaluation by focus groups is needed after the fulfillment of all ending conditions (ie, in step 6). Focus groups help gather more data than individual interviews, since experts respond to the input of others, triggering discussions and idea generation [23]. According to Szopinski et al [39], focus groups are particularly suitable to assess the comprehensiveness, robustness, understandability, and

XSL•FO
RenderX

extensibility of a taxonomy, as well as the shape of dimensions and characteristics. Members were recruited on the basis of the target audience of the taxonomy (see section about Research gap, objective, and questions). We ensured that they have substantial knowledge with regard to HMDCs, stemming from EU initiatives (Multimedia Appendix 1). Ultimately, the focus group consisted of 7 experts. However, scheduling conflicts affected the organization of meetings. The first focus group had to be split into 2 sessions, involving 5 experts. The second focus group iteration required 3 separate sessions to accommodate all 7 experts. Multimedia Appendix 3 shows the distribution of experts to sessions per iteration. Focus group sessions occurred in the fourth and sixth iteration. Their rough agenda was as follows: (1) the objective of the taxonomy (only fourth iteration) and current version (fourth and sixth iteration) were presented; (2) a dimension with its characteristics was explained; (3) experts shared their expectations about this dimension, especially with regard to real-world use cases; and (4) we triggered discussions by contrasting expectations. Because the focus groups served for evaluation purposes and concrete design elements were discussed in the plenum directly, we dispensed with detailed coding. This lightweight effort is inspired and justified by similar research studies published in high-ranking IS journals [14,41]. Therefore, we argue that our approach successfully mitigates any subjectivity biases because specific taxonomy design elements were addressed and discussed in the focus groups directly. This narrow focus led to the derivation of tangible design activities (ie, deletion, addition, alteration, and renaming of design elements and cocreation of use cases).

In the fourth iteration, the focus group resulted in substantial changes of dimensions and characteristics, particularly regarding taxonomy elements of data accessibility and findability (Multimedia Appendix 2). On the contrary, the second session in the sixth iteration caused minor adjustments only (eg, renaming a few characteristics), emphasizing the cocreation of use cases. Finally, the third session of the sixth iteration merely led to refined use case formulations and merging of content. Given these small adjustments, a sufficient saturation in results was considered [23] and the taxonomy development was terminated (ie, step 7). Multimedia Appendix 3 lists the focus group members who participated in the evaluation of the taxonomy and the cocreation of use cases during the fourth and sixth iterations. It further states the EU initiatives the experts were recruited from and whether these initiatives have an operative HMDC or one that is still under development.

Ethical Considerations

In the run-up to the sessions, the experts received an information sheet explaining the study context, the procedure, and the approach to gathering and processing interview data. It was assured that no personal information will be disclosed. Experts were informed about their right to opt out. During the focus group sessions, only anonymous interview data in the form of handwritten notes were collected. Neither a video or voice recording nor the transcription of interview material was used. Experts did not receive any financial compensation. Hence, conducting focus groups did not require the official approval of an ethics review committee.

Results

Taxonomy for FAIR HMDCs

Overview

Table 4 shows the taxonomy containing 20 dimensions (D_n) and 101 characteristics (C_{nm}) structured alongside the FAIR data principles as meta-dimensions [40]. For visualization, morphologies were considered (Multimedia Appendix 2) as they demonstrate the structure and arrangement of taxonomy elements [90]. Importantly, the meta-dimension of data accessibility has exclusive dimensions pertaining to the general design of HMDCs. The other meta-dimensions relate to design elements of metadata assets published *within* the HMDC and the data accessibility constraints. On the basis of the study by Nickerson et al [37], nonexclusiveness was chosen for some of the dimensions associated with metadata assets. The reason is that they propose structural characteristics, which might be accumulated for creating effective metadata assets in an HMDC. Exclusivity of dimensions allows to categorize design elements into mutually distinct characteristics, ensuring clarity and avoiding overlaps. Nonexclusivity can accommodate complex multidimensional relationships by enabling design elements belonging to multiple characteristics [37]. Correspondingly, the difference between OR and XOR in the middle column of Table 4 is that OR is used for nonexclusive characteristics, while XOR is used for exclusive ones.

XSL•FO
RenderX

Table 4. Taxonomy of FAIR[a] HMDCs[b] presented as a morphological box.

Dimension (D_n)	Characteristics (C_{nm})	Exclusive or nonexclusive
Findability		
D_1: data source	$C_{1.1}$: patients' health OR $C_{1.2}$: medical procedures OR $C_{1.3}$: medical products OR $C_{1.4}$: others	Nonexclusive
D_2: managerial. details	$C_{2.1}$: holder OR $C_{2.2}$: origin OR $C_{2.3}$: collection OR $C_{2.4}$: qualification OR $C_{2.5}$: financials OR $C_{2.6}$: others	Nonexclusive
D_3: data type	$C_{3.1}$: admin XOR $C_{3.2}$: primary care XOR $C_{3.3}$: secondary care XOR $C_{3.4}$: registries XOR $C_{3.5}$: others	Exclusive
D_4: population information	$C_{4.1}$: disease OR $C_{4.2}$: family linkages OR $C_{4.3}$: lifestyle factors OR $C_{4.4}$: population OR $C_{4.5}$: sociodemographic OR $C_{4.6}$: catchment area coverage	Nonexclusive
D_5: data sensitivity	$C_{5.1}$: synthetic data XOR $C_{5.2}$: anonymized data XOR $C_{5.3}$: pseudonymized data XOR $C_{5.4}$: personal data	Exclusive
Accessibility		
D_6: catalogue accessibility	$C_{6.1}$: public XOR $C_{6.2}$: hybrid XOR $C_{6.3}$: private	Exclusive
D_7: dataset accessibility	$C_{7.1}$: free XOR $C_{7.2}$: formal request XOR $C_{7.3}$: strictly limited XOR $C_{7.4}$: others	Exclusive
D_8: access control	$C_{8.1}$: catalog operator OR $C_{8.2}$: internal DAC[a] OR $C_{8.3}$: external DAC OR $C_{8.4}$: none OR $C_{8.5}$: others	Nonexclusive
Interoperability		
D_9: program discoverability	$C_{9.1}$: Beacon OR $C_{9.2}$: BBMRI-MIABIS[d] OR $C_{9.3}$: bioimage OR $C_{9.4}$: CESSDA[e] OR $C_{9.5}$: DCAT[f] OR $C_{9.6}$: ECRIN-CRMDR[g] OR $C_{9.7}$: FairShairing OR $C_{9.8}$: INSPIRE[h] OR $C_{9.9}$: PHIRI[i] OR $C_{9.10}$: others	Nonexclusive
D_{10}: semantic interoperability	$C_{10.1}$: CDISC-SDTM[j] OR $C_{10.2}$: LOINC[k] OR $C_{10.3}$: OMOP[l] OR $C_{10.4}$: Oorphanet standards OR $C_{10.5}$: SNOMED[m] OR $C_{10.6}$: others	Nonexclusive
D_{11}: interoperable communication	$C_{11.1}$: DICOM[n] OR $C_{11.2}$: HL7 FHIR[o] OR $C_{11.3}$: IDMP[p] OR $C_{11.4}$: ISO 800-110[q] OR $C_{11.5}$: others	Nonexclusive
D_{12}: CDM[r]	$C_{12.1}$: type OR $C_{12.2}$: reference OR $C_{12.3}$: release frequency	Nonexclusive
D_{13}: ETL[s] status	$C_{13.1}$: planned XOR $C_{13.2}$: in progress XOR $C_{13.3}$: completed	Exclusive
D_{14}: vocabularies	$C_{14.1}$: medicinal product OR $C_{14.2}$: cause of death OR $C_{14.3}$: quality of life measurement OR $C_{14.4}$: prescription OR $C_{14.5}$: dispensing OR $C_{14.6}$: indication OR $C_{14.7}$: procedures OR $C_{14.8}$: genetic data OR $C_{14.9}$: biomarker data OR $C_{14.10}$: medical event	Nonexclusive
Reusability		
D_{15}: collection methodology	$C_{15.1}$: collection governance OR $C_{15.2}$: collection process OR $C_{15.3}$: dataset updates OR $C_{15.4}$: others	Nonexclusive
D_{16}: collection events	$C_{16.1}$: patient encounter OR $C_{16.2}$: physical examination OR $C_{16.3}$: diagnostics OR $C_{16.4}$: treatment OR $C_{16.5}$: progress note OR $C_{16.6}$: communication OR $C_{16.7}$: regulatory OR $C_{16.8}$: others	Nonexclusive
D_{17}: data linkage	$C_{17.1}$: strategy OR $C_{17.2}$: variable OR $C_{17.3}$: completeness OR $C_{17.4}$: cross-reference OR $C_{17.5}$: none	Nonexclusive
D_{18}: data preservation	$C_{18.1}$: definite records XOR $C_{18.2}$: indefinite records	Exclusive
D_{19}: publish	$C_{19.1}$: approval needed XOR $C_{19.2}$: no approval needed	Exclusive
D_{20}: informed consent	$C_{20.1}$: not required XOR $C_{20.2}$: general use XOR $C_{20.3}$: all studies XOR $C_{20.4}$: specific studies XOR $C_{20.5}$: waiver XOR $C_{20.6}$: others	Exclusive

XSL•FO
RenderX

[a]FAIR: Findability, Accessibility, Interoperability, and Reusability.
[b]HMDC: health metadata catalog.
[c]DAC: Data Access Committee.
[d]BBMRI-MIABIS: Biobanking and Biomolecular Resources Research Infrastructure-Minimum Information About Biobank Data Sharing.
[e]CESSDA: Consortium of European Social Science Data Archives.
[f]DCAT: Data Catalog vocabulary.
[g]ECRIN-CRMDR: European Clinical Research Infrastructure Network – Clinical Research Metadata Repository.
[h]INSPIRE: Infrastructure for Spatial Information in Europe.
[i]PHIRI: Population Health Research Infrastructure.
[j]CDISC-SDTM: Clinical Data Interchange Standards Consortium - Study Data Tabulation Model.
[k]LOINC: Logical Observation Identifiers Names and Codes.
[l]OMOP: Observational Medical Outcomes Partnership.
[m]SNOMED: Systematized Nomenclature of Medicine.
[n]DICOM: Digital Imaging and Communications in Medicine.
[o]HL7 FHIR: Health Level 7 Fast Healthcare Interoperability Resources.
[p]IDMP: Identification for Medicinal Products.
[q]ISO800-110: International Organization for Standardization 800-110.
[r]CDM: common data model.
[s]ETL: extract, transform, load.

Data Findability

The meta-dimension prescribes that datasets orchestrated by an HMDC must be easily discoverable, requiring metadata assets to describe essential attributes of the decentral datasets [40]. The dimension *data source* (D_1) refers to abstract categories for data classification in the catalog system. Following European Medicines Agency guidelines [43], and implementations in practice [49,50], patients' health ($C_{1.1}$) comprises datasets attributable to conditions. Examples are diseases, causes of death, prescriptions and dispensing of medicines, clinical measurements, genetic data, units of health care use, and all other similar patient-generated data, for example, wearables. Medical procedures ($C_{1.2}$) encompass data describing hospital admission discharges, intensive care admissions, administration of vaccines or other injectables, medical operations, biomarker data, and diagnostic codes [43,44]. The latter includes, among others, the International Classification of Disease Code, the Major Comorbidity Code, and the Major Diagnostic Code. Medical products ($C_{1.3}$) span categories like prescribed medicinal products for human use, contraception, indication for use, and medical device data. Others ($C_{1.4}$) may refer to further data, for example, health care providers delivering diagnosis and treatment services [43]. The dimension *managerial details* (D_2) captures crucial organizational metadata to be disclosed by the HMDC as part of a dataset's metadata asset [24]. Above all, the data holder ($C_{2.1}$) must be publicized, including contact details (ie, data steward). This entity sustains the record collection in an underlying dataset [60]. The origin ($C_{2.2}$) of the data refers to the countries or geographical regions of their acquisition [60] and the language [44]. The characteristic of collection ($C_{2.3}$) details the acquisition dates as well as all data assemblage information, except collection methodology and events. If the dataset has received a formal qualification ($C_{2.4}$), this should also be disclosed in the metadata asset [43]. The same holds for sources of finance ($C_{2.5}$) having sponsored the dataset creation [23,61], for example, data holder, public, industry, research, or patient organizations [44]. Naturally, other ($C_{2.6}$) metadata attributes specifying managerial details are conceivable. The dimension *data type* (D_3) describes broader content related categories applicable to the dataset. It mainly distinguishes datasets containing administrative ($C_{3.1}$ [62]), primary ($C_{3.2}$) and secondary care ($C_{3.3}$ [63]), registry ($C_{3.4}$), and other ($C_{3.5}$) data types. Furthermore, *population information* (D_4) as metadata attribute refers to the specifics of the records within a dataset [17]. The taxonomy narrows the dimension down to collected disease information ($C_{4.1}$); particular population specifics ($C_{4.2}$; eg, age groups); family linkages ($C_{4.3}$; eg, household, parent-child, sibling, and not applicable) lifestyle factors ($C_{4.4}$; eg, tobacco use, physical exercises, and diet); sociodemographic data ($C_{4.5}$; eg, gender, ethnicity, education, and deprivation index); and ($C_{4.6}$) catchment area coverage [17,43,44]. *Data sensitivity* (D_5) addresses the identifiability of data subjects of whom records were collected. Records can be synthetic ($C_{5.1}$), anonymized ($C_{5.2}$), pseudonymized ($C_{5.3}$), or contain personal data ($C_{5.4}$) [15,41].

Data Accessibility

Once datasets are findable in the HMDC, their accessibility must be facilitated. By their design, HMDCs must ensure legal and ethical compliance of all data access and sharing processes in their health care ecosystems [19]. This particularly involves that datasets can be retrieved by authorized users only, implying rigor access control functions [5,41]. Accordingly, the dimension *catalogue accessibility* (D_6) refers to access modalities for data consumers to use the HMDC. HMDCs can be public ($C_{6.1}$) allowing anyone to browse metadata assets and discover datasets. Alternatively, HMDCs can be private ($C_{6.2}$), limited to a certain number of users who have been formally authorized by a dedicated authority. In addition, hybrid ($C_{6.3}$) forms exist. *Dataset accessibility* (D_7) describes the access modalities for data consumers to published datasets [30]. The access can be free ($C_{7.1}$), implying that data must at least be anonymized or

XSL•FO
RenderX

synthetic to comply with legal and ethical guidelines [41,64]. Datasets can also require a formal request ($C_{7.2}$) with a Data Access Committee (DAC) or a data steward deciding upon access requests [65,66]. HMDCs frequently require such requests of data consumers to be approved by their ethic committees before processing them within the ecosystem [66]. Moreover, data access can be strictly limited ($C_{7.3}$) to a demarcated group of data consumers. Although, members of such limited groups also need to make formal requests for data access [66]. This implies that other ($C_{7.4}$) data access modalities exist, especially combinations of $C_{7.1}$ to $C_{7.3}$. *Access control* (D_8) refers to the mechanisms implemented by HMDCs that facilitate the aforementioned decision-making by empowered entities [30]. The dimension specifies the entities who determine whether data consumers receive the requested datasets and are allowed to perform which kinds of processing activities [67]. It distinguishes the catalog operator ($C_{8.1}$); internal DACs at the sides of the data holders ($C_{8.2}$ [65,66]); external DACs ($C_{8.3}$) that are run centrally by an independent third party [23,66]; the absence of access control ($C_{8.4}$; ie, free data [$C_{7.1}$]); and any other forms ($C_{8.5}$). Generally, access control in HMDC designs is crucial for maintaining data security, confidentiality, and compliance with legal and ethical constraints [67].

Data Interoperability

A core objective of HMDCs is to enable data consumers accumulating datasets across organizations, effectively, to create meaningful connections and analyses [7]. This makes data interoperability crucial [91]. To that end, HMDCs leverage specific standards described in this meta-dimension. Thereof, *programmatic discoverability* (D_9) refers to the ability of data consumers to programmatically query, access, and retrieve metadata assets and search for their attributes. It is defined by the joint action Towards European Health Data Space (TEHDAS) [16] as the ability to identify, access, and understand health data by automated means. Associated approaches commonly involve application programming interfaces or similar programmatic methods to access and filter metadata in the HMDC [19]. Following the TEHDAS community [16], the dimension is narrowed down to the most frequent standards used by HMDCs. These are:

- Beacon ($C_{9.1}$),
- Biobanking and biomolecular resources research infrastructure-minimum information about biobank data sharing (BBMRI-MIABIS; $C_{9.2}$),
- Bio-image archive ($C_{9.3}$),
- Consortium of European Social Science Data Archives (CESSDA; C9.4),
- Data catalog vocabulary (DCAT; $C_{9.5}$),
- European Clinical Research Infrastructure Network – clinical research metadata repository (ECRIN-CRMDR; $C_{9.6}$),
- FairSharing ($C_{9.7}$),
- Infrastructure for Spatial Information in Europe (INSPIRE; C9.8),
- Population Health Research Infrastructure (PHIRI; C9.9), and
- Others ($C_{9.10}$).

HMDCs typically adhere to one of those standards to ensure programmatic discoverability of published data offerings (ie, the metadata assets). The dimension *semantic interoperability* (D_{10}) ensures that the precise format and meaning of datasets is preserved and understood, covering both semantic and syntactic aspects [68]. Similar to D_9, the characteristics of this dimension encompass standards commonly applied by HMDCs. These are

- Clinical Data Interchange Standards Consortium - Study Data Tabulation Model (CDISC-SDTM; C10.),
- Logical Observation Identifiers Names and Codes (LOINC; C10.2),
- Observational Medical Outcomes Partnership (OMOP; C10.3),
- Orphanet (C10.4),
- Systematized Nomenclature of Medicine (SNOMED; C10.5 [16,68,69] and
- Others (C10.6).

Subsequently, the dimension *interoperable communication* (D_{11}) comprises approaches implemented by HMDCs to facilitate seamless and effective data sharing between data holders and consumers [70]. Approaches typically used for interoperable communication are Digital Imaging and Communications in Medicine (DICOM; $C_{11.1}$), Health Level 7 Fast Healthcare Interoperability Resources (HL7 FHIR; $C_{11.2}$), Identification for Medicinal Products (IDMP; $C_{11.3}$), and International Organization for Standardization (ISO) 800-110 ($C_{11.4}$) [16,71]. As for the previous dimensions, other standards ($C_{11.5}$) are conceivable. For D_9 to D_{11}, detailed information is easily available in the web.

The following dimensions deal with “data harmonization” as a crucial aspect of data interoperability. They comprise HMDC design elements for standardizing disparate datasets. The purpose is to ensure consistency and coherence of all datasets classifiable to the same data type (D_3). Data harmonization aims to create a unified and cohesive view on datasets, enhancing their allocation, sharing, and use [7]. The common data model (CDM; D_{12}) describes the specifications relating to the structured representation of data records within datasets [5]. The CDM unfolds implications to the relationships between these records, as well as the rules and possibilities for data use. It defines how data consumers can access and process datasets, thus providing the foundation for data consistency, interoperability, and orchestration [10,72]. The taxonomy distinguishes the CDM type ($C_{12.1}$) [10,43,72], the CDM references ($C_{12.2}$), for example, websites or publications, and the release frequency of CDM specification updates ($C_{12.3}$) [43]. Furthermore, information about datasets on their transformation status (extract, transform, load [ETL]) to a CDM should be provided [44]. This *ETL status* (D_{13}) can be described as planned ($C_{13.1}$), in progress ($C_{13.2}$), or completed ($C_{13.3}$), indicating the readiness of the dataset for use. Finally, HMDCs leverage *vocabularies* (D_{14}) as sets of fixed terms, labels, or identifiers to describe and categorize the metadata assets [63]. Vocabularies facilitate understanding,

XSL•FO
RenderX

discovery, and allocation of metadata with a consistently applied language [63,73]. The dimension distinguishes 10 characteristics for classifying vocabularies on the basis of pertinent literature: medicinal product ($C_{14.1}$), cause of death ($C_{14.2}$), quality of life measuring ($C_{14.3}$), prescription ($C_{14.4}$), dispensing ($C_{14.5}$), indication ($C_{14.6}$), procedures ($C_{14.7}$), genetic data ($C_{14.8}$), biomarker data ($C_{14.9}$), and medical event ($C_{14.10}$) [43,44,73].

Data Reusability

FAIR datasets must be created and documented in a way that allows reusage for different purposes. For HMDCs, this implies providing contextual information beyond the metadata dimensions associated with data findability [40]. *Collection methodology* (D_{15}) encompasses characteristics that are associated with how data records were created [74]. Thereof, collection governance ($C_{15.1}$) addresses information about data capture, demonstrating legal and ethical compliance [43]. This includes data quality checks and validation activities [75]. The latter may also refer to the question of whether the dataset allows access to the actual records. Furthermore, the collection process ($C_{15.2}$) outlines how records in the dataset were created [74], for example, surveys, questionnaires, or data retrieval from hospital IS. Dataset updates ($C_{15.3}$) disclose refreshment dates of datasets, for instance, fixed dates around the year [43]. Naturally, the collection methodology can contain other ($C_{15.4}$) use case–specific characteristics as additional metadata attributes. Similar to D_{15}, *collection events* (D_{16}) narrows down the categories of incidents having triggered the creation of a record in the dataset [17]. The dimension comprises the characteristics of patient encounter ($C_{16.1}$; eg, interactions with health care providers); physical examination ($C_{16.2}$; eg, patient's health examined by a professional); diagnostics ($C_{16.3}$; eg, results of medical condition checks); treatment ($C_{16.4}$; eg, documentation of conditions and treatment plans); progress notes ($C_{16.5}$; eg, changes in patients' health status, responses to treatment, or modifications of care plans); communication ($C_{16.6}$; eg, information exchanged by health care providers); regulatory ($C_{16.7}$; eg, legally required documentation of patients' care); and others ($C_{16.8}$) [43,44].

The dimension *data linkage* (D_{17}) describes whether and how a dataset was created by linking others [43,44,76]. The metadata should disclose the linkage strategy ($C_{17.1}$) which could be deterministic, probabilistic, or both. In addition, the used linkage variable ($C_{17.2}$) should be published, along with the completeness of data linkage ($C_{17.3}$). Ideally, the linked datasets should be cross-referred ($C_{17.4}$) and, if applicable, their availability in the HMDC highlighted. In case no data linkage was applied, no corresponding metadata attribute is provided ($C_{17.5}$). Furthermore, *data preservation* (D_{18}) indicates whether records in the dataset are preserved indefinitely ($C_{18.1}$) or, if not ($C_{18.2}$), the time for which they are specified [77]. *Publishing constraints* (D_{19}) provides information to data consumers whether an approval of the data holder ($C_{19.1}$) is needed to publish results obtained from using the dataset or an approval is not needed ($C_{19.2}$) [43]. In the former case, the kind of approval and the approval process should be described. Finally, metadata assets of HMDCs should reveal whether *informed consent* (D_{20}) was obtained or needs to be obtained for data processing [78]. Generally, the characteristics not required ($C_{20.1}$), required for general use ($C_{20.2}$), required for all studies ($C_{20.3}$), required for specific studies ($C_{20.4}$), waiver ($C_{20.5}$), and other ($C_{20.6}$) are recommendable [43].

Cocreated HMDC Use Cases

Overview

The usability, effectiveness, and accuracy of the taxonomy are amplified by 4 "illustrative scenarios" for HMDCs that demonstrate how the FAIR dimensions and characteristics are reflected in real-world use cases [39]. These use cases facilitate the taxonomy's tangibility and the ascertainment of its practical implications, while triangulating the results. As such, they add depth and context to the taxonomy [39]. Originally, 6 abstract application scenarios were derived from recommendations of EMA [43]. Building upon this, we continuously developed and refined those scenarios on the basis of the insights gained during the taxonomy design iterations in general and the focus groups in particular. In the latter, we relied on the experts' reflections on what they expect from an HMDC and whether our dimensions and characteristics meet their expectations, contradict them, or miss out on certain aspects. In the section about the deductive design iterations, we have already described how the focus groups were conducted. We ensured that the experts possess extensive expertise relevant to HMDC designs, either from a development (ie, technology or legal) or a user perspective (ie, data consumer or provider; Multimedia Appendix 3). Hence, the use cases followed a cocreation approach that was contextualized to the methodological process of the study [39]. Moreover, by being refined within multiple design iterations, the cocreated use cases have been triangulated and their relevance for HMDCs ensured. Following, their connection to the taxonomy is highlighted by direct references to dimensions (see *Taxonomy for FAIR HMDCs*). Additionally, I to VII refer to statements of experts who are listed in Multimedia Appendix 3. Reference citations of the literature and analysis objects demonstrate further exemplary sources having contributed to the final versions of the cocreated use cases.

Study Planning

Use case 1 is as follows: *a data consumer wants to identify suitable datasets for a planned study*.

The HMDC must enable data consumers to effectively identify datasets for medical research studies [15] by implementing the following process: First, a data consumer who wants to access the HMDC, needs to be authorized as a qualified user (D_6; #I [50]). Second, this authorized user must be able to browse and filter published metadata assets to discover relevant datasets that fulfill specifications of an intended study (#V and VII [49,50]). For example, detailed data type [63] or population information of the metadata assets should be disclosed to enable verifying the relevance of datasets (D_3 and D_4; #I, II, V, and VII [17,58]). Third, the HMDC must allow to check managerial

XSL•FO
RenderX

details concerning information about the data holder, origin, collection, qualification, and financials (D_2 [43,50]), including the eligibility to receive synthetic, anonymized, pseudonymized, or personal data (D_5; #IV [15]). Subsequently, the data consumer must be facilitated to perform a preliminary assessment of datasets regarding their relevance for the planned study (#III and VII). At this stage, a first list of candidates should be possible to be established. Ideally, the data consumer can access links (ie, cross-references) within the metadata assets to identify former studies which were performed with the same dataset, addressing similar research questions ($C_{17.4}$; #II and VII [43,76]). Such studies are typically accessed outside the contextual boundaries of HMDCs (#IV). Finally, depending on the governance modalities of selected datasets (D_7), the HMDC must enable the data consumer to request data access (#I-VII [30]). Therefore, an official data order needs to be submitted by the HMDC on behalf of the data consumer, containing specifications about the planned study and required documents, for example, protocols and ethical assessments (#I-III). With respect to the datasets accessibility constraints (D_7) and their associated access control (D_8) characteristics, the HMDC must forward data orders to the data holders ($C_{8.2}$ [66]), external third parties ($C_{8.3}$ [23]) or determine request permission or denial decisions itself ($C_{8.1}$ and $C_{8.4}$; #I, III, and VI).

Study Assessment

Use case 2 is as follows: *a dataset is mentioned in a conducted study. The data consumer wants to evaluate this study based on the suitability of the datasets used therein.*

Given that the datasets used in a conducted study are available in the HMDC, data consumers must be enabled to verify, in retrospective, the suitability of these datasets (#I, II, and VII). To support such evaluations, the HMDC must provide different parts of the metadata asset, depending on the nature of the conducted study (#II and VII). For example, to assess the representativeness of the study population, the data consumer needs to examine qualitative metadata attributes (#II and VII), such as population information ($C_{4.1}$-$C_{4.5}$ [17]); data type (D_3 [43,63]); collection methodology (D_{15} [74]); and collection events (D_{16} [17,43]). In addition, quantitative metadata such as the percentage of the population covered in the catchment area ($C_{4.6}$) should be disclosed by the HMDC assets [17,43]. Furthermore, the data consumer might want to explore technical details to evaluate a study and its database, respectively (#I, II, and VII). Examples are the vocabularies used to define variables (D_{14} [43,63]), the CDM according to which the used datasets are structured (D_{12} [43]), the ETL status (D_{13} [44]) and, if applied, any data linkage strategies (D_{17} [76]). Moreover, cross-references should be listed in the metadata assets of the HMDC ($C_{17.4}$) to allow identifying and obtaining lessons learned from other studies, where the same dataset was leveraged (#VII [76]). In doing so, the HMDC facilitates data consumers to identify strengths and limitations of datasets used in conducted studies.

Study Creation and Data Benchmarking

Use case 3 is as follows: *a data consumer writes a study protocol that requires to describe the underlying data sets and compare their characteristics.*

An HMDC must enable data consumers to easily access standardized metadata information about datasets that need to be specified and compared in a study protocol to be written (#I-VII). For HMDCs, this requires making attribute values of metadata assets directly and easily retrievable for data consumers to facilitate an efficient description and comparison of datasets (particularly, D_1-D_5; #VII). As such, when writing a study protocol, the data consumer can simply provide links to the metadata assets available in the HMDC, alongside with all kinds of other information that is interesting in the protocol's context (#III and VII [43]). Providing such links is also beneficial because study readers could, in addition to basic metadata information (D_1-D_5), be interested in collection methodologies (D_{15} [74]) and events (D_{16} [17]), data preservation (D_{18} [77]), consent requirements (D_{20} [43,78]) as well as technical (D_9-D_{14}) and compliance specifics (D_6-D_8). Thus, HMDCs strengthen transparency and reproducibility of studies by facilitating an effective creation of study protocols (#III and V). At the same time, they support data benchmarking by providing detailed standardized metadata, potentially reaching beyond published datasets (#I-VII).

Data Analysis

Use case 4 is as follows: *a data consumer wants to benefit from the experience of others for the creation of a study's programming script or statistical analysis.*

If a study relies on a CDM, the HMDC should enable the data consumer to identify, for published datasets, the ETL procedure (D_{13}) from the dataset to the CDM (D_{12}) [43]. Irrespective of whether data holders have converted their entire datasets, or only an extraction thereof, this information can support the development of the study script (#VII [43]). If the HMDC publishes cross-references of datasets ($C_{17.4}$), the data consumer is also facilitated in finding further studies having investigated the same topic or used a comparable research design (#I, II, and VII [76]). These studies may disclose information on how to operationalize data variables, which offers the data consumer additional support in the development of a programming script (#I and VII). After data analysis, the HMDC may require the data consumer to record the developed script in a public repository and provide a link to the study protocol (see use case 3). This allows the HMDC to cross-reference it in the metadata assets of the datasets used (#V [43]). In doing so, transparency, reproducibility, and quality of studies are supported. Importantly, before publishing any study results, the HMDC should enable the data consumer to check whether approvals of data holders, of whom datasets were obtained, are required (D_{19}; #IV and V [43]).

XSL•FO
RenderX

Discussion

Principal Findings

The taxonomy detailed in this work provides 20 dimensions and 101 characteristics to develop FAIR HMDCs, representing initial, yet actionable design knowledge to answer RQ1. Comprehensive description is achieved when amplifying the taxonomy in the light of the cocreated use cases that entail real-world requirements (RQ2). The generated design knowledge provides added value because HMDCs facilitate effective and efficient use of RWD for generating RWE. Thereby, the obtained results accentuate the integration of FAIR principles into HMDCs to ensure "findability, accessibility, interoperability, and reusability" of the RWD circulating within the underlying health care data ecosystem [40]. The design knowledge can be classified into scientific and managerial contributions, as outlined in the following sections.

Interpretation of Findings

The taxonomy's *scientific contributions* intensify previous work on data catalogs, paying particular attention to, first, their implementation as decentralized components within data ecosystems and, second, their application in health care peripheries. Consequently, some of the design elements conceptualized in this study draw from prior research about metadata catalogs, as shown in the section Research gap, objective, and questions. Concurrently, they further spin the red paths of HMDC developments with respect to European initiatives pushing health care data ecosystems forward, for example, IDERHA, Elixir, HealthData@EU, EUCAIM, TEHDAS, or EHDEN (Multimedia Appendix 1). Therefore, the taxonomy closes the identified research gap.

On the one hand, the conceptualized dimensions and characteristics of the taxonomy describe and classify attributes of "HMDC metadata assets." First, they relate to attributes associated with data findability, such as data source, managerial details, the type of data records contained, data sensitivity, and population information. Second, they state crucial information pertaining to data reusability. This encompasses collection methodologies, data linkage, and preservation as well as consent and possible result publishing constraints. Third, the taxonomy emphasizes the need to specify data interoperability attributes in metadata assets. Among others, data standardization according to a CDM, the prescription of vocabularies, and the technological implementation of programmatic discoverability are important. On the other hand, the taxonomy contains design elements referring to more general data accessibility constraints pertaining to the "overall HMDC design." They accentuate basic security and governance considerations regarding dataset and catalog accessibility as well as the access control framework. Conclusively, from a scientific viewpoint, the artifact provides fundamental design knowledge [31] that unfolds broad implications and a solid starting point for future research.

Regarding *managerial contributions*, the taxonomy enables health care practitioners (see Introduction section for target audience) to navigate more effectively in the largely unexplored field of HMDCs, particularly focusing on their application in health care data ecosystems across Europe. It helps both researchers and practitioners to anchor and communicate their work [41]. The taxonomy also represents a support tool for developing HMDCs, where the illustrative scenarios assume an accentuated role. They showcase how the design elements are reflected in real-world use cases [43]. In essence, these use cases amplify that the taxonomy supports common activities for planning, assessing, and conducting medical research studies as well as benchmarking and analyzing the underlying RWD.

Subsequently, the contributions of the study are discussed in the light of 3 major issues. These are (1) the exclusiveness of taxonomy characteristics; (2) the difference between HMDCs and centralized data catalogs; and (3) the absence of data quality and data-sharing incentives as explicit taxonomy dimensions.

Depending on the meta-dimension, the taxonomy contains nonexclusive characteristics, which might be accumulated, to facilitate the design of metadata assets, that is, data findability, interoperability, and reusability. Alternatively, the taxonomy has mutual exclusive characteristics to classify and distinguish HMDC designs with respect to their data security and governance approaches, that is, data accessibility. This mixture of exclusive and nonexclusive dimensions can foster the understanding of health care practitioners, while allowing for an easy alteration of the taxonomy [37]. Convertibility is vital because HMDCs represent a rapidly evolving and changing field, where new solutions vanish and emerge constantly.

Furthermore, even though the objective of the taxonomy is not to differentiate between centralized and decentralized data catalogs, as distinguished in the Theoretical Background section, it pinpoints their fundamental design commonalities and differences. The meta-dimensions concerning the design of the metadata assets are conceivable for both approaches in health care contexts (ie, findability, interoperability, and reusability). The reason is that, despite datasets being stored centrally within intraorganizational data catalogs [26], meaningful metadata need to be disclosed to their data users by means of the catalog offerings. Naturally, the same holds for decentralized catalogs. However, the taxonomy also shows design differences with respect to its meta-dimension data accessibility. While decentralized catalogs can have various combinations of characteristics in the associated dimensions, centralized health data catalogs exhibit one specific pattern of characteristics. They usually are exclusively private systems ($C_{6.1}$) as their functionalities are only accessible to members of the operating organization. Similarly, dataset access is strictly limited to this specific group of predefined users ($C_{7.3}$). Finally, access control lies solely with the organization operating the centralized catalog ($C_{8.1}$).

As a last discussion point, the taxonomy contains neither dimensions associated with data quality nor incentives for data sharing. The reason is that these concepts, although important, are broad, multifaceted, and hardly explored, making a systematic categorization difficult. Generally, data quality involves subjective and context-dependent assessments [92], while incentives to share data are influenced by external, sociopolitical, and institutional factors [28]. Typically, HMDCs do not disclose any data quality metrics as those can barely be quantified and are subject to applied data types, formats, and

XSL•FO
RenderX

standards [41]. Rather, HMDCs publish test samples consisting of synthetic or fully anonymized data that do not justify a dimension in the taxonomy. Similarly, incentivizing data sharing, for example, via price tags for datasets or mandatory citations, represents an unsolved problem [93]. To circumnavigate this issue, HMDCs commonly rely on membership fees and public funding. The former restricts data access to members of the HMDC operating organization. The latter compensates data providers through public funds, usually applied in preliminary stages. Naturally, other business models exist. However, both data quality indicators and data-sharing incentives represent underdeveloped fields requiring future research beyond the study's design perspective [33]. Consequently, these concepts are not a part of the taxonomy, because they cannot be defined in universally applicable dimensions. Arguably, their inclusion would have overcomplicated the taxonomy and undermined its focus on actionable HMDC design knowledge.

Limitations

The taxonomy is mainly subject to the following limitations. In the inductive iterations, results were derived from a potentially limited number of publications because of the emphasis on 4 main databases. Similarly, in the deductive iterations, the examined analysis objects might merely cover a snapshot of what was available at the time (ie, many analysis objects have been in progress), be outdated quickly, and not be conclusive. As for the SLR, the conclusiveness of analysis objects is particularly questionable because of, first, the focus on European ecosystem initiatives and, second, a possible negligence of many centralized health data catalogs. In the evaluative iterations, the experts might not have captured the full range of relevant perspectives on HMDCs and are limited in number. Furthermore, the research design comprises certain limitations per se. As it is with qualitative research, taxonomy building requires substantial generalizations and simplifications of intricate and interdisciplinary content [83]. Although countermeasures were taken (see Methods section), these factors imply interpretative biases inevitably incorporated into the results [41], for example, extracting design elements from public data. Moreover, as shown in the Theoretical Background section, new HMDCs must be expected to arise constantly, while others are likely to disappear with a high frequency. Hence, the taxonomy must be altered swiftly. To conclude, the taxonomy provides first actionable design knowledge about HMDCs but requires continuous triangulation of design elements by future research.

Conclusions

Despite the limitations, the scientific and managerial contributions of this study unfold broad implications, which are formulated as recommendations for future research. Generally, HMDCs should be increasingly investigated in practice, for example, by more in-depth case studies. On the one hand, it is of utmost importance to keep track of the rapidly evolving HMDC-related initiatives in Europe. Their conceptual and technical advancements should be analyzed and evaluated constantly against the background of the taxonomy design elements, deriving the need for modifying dimensions and characteristics. On the other hand, by incorporating worldwide efforts toward health care data ecosystems and HMDCs, the scope of the taxonomy can be expanded and design knowledge beyond European jurisdictions can be created. In this regard, it is important to mention that the generated design knowledge about European HMDCs already entails such global implications. The FAIR dimensions of the taxonomy state fundamental characteristics of health data FAIR, making it universally relevant. In other words, the taxonomy conveys generally conceivable options for using catalog functionalities and underlying metadata assets. In addition, it outlines how to design those health metadata assets meaningfully. Therefore, despite the European focus, the taxonomy addresses global challenges with regards to health data sharing and metadata catalog designs, underlining its broad implications. Nevertheless, further research is essential, because HMDCs represent the fulcrum for allocating, exchanging, and using RWD to effectively generate RWE in emerging health care data ecosystems.

Acknowledgments

This research has emerged from the IDERHA (Integration of Heterogeneous Data and Evidence towards Regulatory and Health Technology Assessment Acceptance) project funded by the Innovative Health Initiative of the European Union (GAP-101112135) and the Digital Life Journey project funded by Fraunhofer-Gesellschaft (PN: 051-600000).

Data Availability

Data sharing is not applicable to this paper as no datasets were generated or analyzed during this study.

Authors' Contributions

SS conceptualized the study, conducted the investigation, developed the methodology, managed the project, supervised its execution, drafted the original manuscript, and contributed to reviewing and editing the manuscript. MKM contributed to the conceptualization and investigation, validated the findings, and participated in the review and editing of the manuscript.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Leading European initiatives toward health care data ecosystems.
[DOCX File , 30 KB - formative_v9i1e63396_app1.docx]

Multimedia Appendix 2
Intermediary results of the iterative research process and key references.
[DOCX File , 70 KB - formative_v9i1e63396_app2.docx]

Multimedia Appendix 3
List of focus group members and distribution to sessions.
[DOCX File , 50 KB - formative_v9i1e63396_app3.docx]

References

1. Agarwal R, Gao GG, DesRoches C, Jha AK. The digital transformation of healthcare: current status and the road ahead. Inf Syst Res 2010;21(4):796-809. [doi: 10.1287/isre.1100.0327]
2. Singh G, Schulthess D, Hughes N, Vannieuwenhuyse B, Kalra D. Real world big data for clinical research and drug development. Drug Discov Today 2018;23(3):652-660 [FREE Full text] [doi: 10.1016/j.drudis.2017.12.002] [Medline: 29294362]
3. Remy L, Ivanović D, Theodoridou M, Kritsotaki A, Martin P, Bailo D, et al. Building an integrated enhanced virtual research environment metadata catalogue. Electron Libr 2019;37(6):929-951. [doi: 10.1108/el-09-2018-0183]
4. Feldman K, Johnson RA, Chawla NV. The state of data in healthcare: path towards standardization. J Healthc Inform Res 2018;2(3):248-271 [FREE Full text] [doi: 10.1007/s41666-018-0019-8] [Medline: 35415409]
5. Swertz M, van Enckevort E, Oliveira JL, Fortier I, Bergeron J, Thurin NH, et al. Towards an interoperable ecosystem of research cohort and real-world data catalogues enabling multi-center studies. Yearb Med Inform 2022;31(1):262-272 [FREE Full text] [doi: 10.1055/s-0042-1742522] [Medline: 36463884]
6. Bietz MJ, Bloss CS, Calvert S, Godino JG, Gregory J, Claffey MP, et al. Opportunities and challenges in the use of personal health data for health research. J Am Med Inform Assoc 2016;23(e1):42-48 [FREE Full text] [doi: 10.1093/jamia/ocv118] [Medline: 26335984]
7. Kush RD, Warzel D, Kush MA, Sherman A, Navarro EA, Fitzmartin R, et al. FAIR data sharing: the roles of common data elements and harmonization. J Biomed Inform 2020;107:103421 [FREE Full text] [doi: 10.1016/j.jbi.2020.103421] [Medline: 32407878]
8. Wiertz S, Boldt J. Ethical, legal, and practical concerns surrounding the implemention of new forms of consent for health data research: qualitative interview study. J Med Internet Res 2024;26:e52180 [FREE Full text] [doi: 10.2196/52180] [Medline: 39110970]
9. Molnár-Gábor F, Beauvais MJ, Bernier A, Jimenez MP, Recuero M, Knoppers BM. Bridging the European data sharing divide in genomic science. J Med Internet Res 2022;24(10):e37236 [FREE Full text] [doi: 10.2196/37236] [Medline: 36260387]
10. Lovestone S, EMIF Consortium. The European medical information framework: a novel ecosystem for sharing healthcare data across Europe. Learn Health Syst 2020;4(2):e10214 [FREE Full text] [doi: 10.1002/lrh2.10214] [Medline: 32313838]
11. Manogaran G, Varatharajan R, Lopez D, Kumar PM, Sundarasekar R, Thota C. A new architecture of internet of things and big data ecosystem for secured smart healthcare monitoring and alerting system. Future Gener Comput Syst 2018;82:375-387. [doi: 10.1016/j.future.2017.10.045]
12. Sharon T, Lucivero F. Introduction to the special theme: the expansion of the health data ecosystem – rethinking data ethics and governance. Big Data Soc 2019;6(2):205395171985296. [doi: 10.1177/2053951719852969]
13. Witte AK. A review on digital healthcare ecosystem structure: identifying elements and characteristics. In: Proceedings of the 23rd Pacific Asia Conference on Information System. 2020 Presented at: PACIS '20; July 9-12, 2020; Dubai, UAE p. 226-238 URL: https://aisel.aisnet.org/pacis2020/228 [doi: 10.1111/j.1467-8373.2010.01427.x]
14. Scheider S, Lauf F, Möller F, Otto B. A reference system architecture with data sovereignty for human-centric data ecosystems. Bus Inf Syst Eng 2023;65(5):577-595. [doi: 10.1007/S12599-023-00816-9]
15. Oliveira JL, Trifan A, Bastião Silva LA. EMIF catalogue: a collaborative platform for sharing and reusing biomedical data. Int J Med Inform 2019;126:35-45. [doi: 10.1016/j.ijmedinf.2019.02.006] [Medline: 31029262]
16. EHDS semantic interoperability framework 2022. TEHDAS. URL: https://tehdas.eu/app/uploads/2023/10/tehdas-recommendations-to-enhance-interoperability.pdf [accessed 2024-04-29]
17. Bergeron J, Doiron D, Marcon Y, Ferretti V, Fortier I. Fostering population-based cohort data discovery: the Maelstrom Research cataloguing toolkit. PLoS One 2018;13(7):e0200926 [FREE Full text] [doi: 10.1371/journal.pone.0200926] [Medline: 30040866]
18. Ulrich H, Kock-Schoppenhauer A, Deppenwiese N, Gött R, Kern J, Lablans M, et al. Understanding the nature of metadata: systematic review. J Med Internet Res 2022;24(1):e25440 [FREE Full text] [doi: 10.2196/25440] [Medline: 35014967]

XSL•FO
RenderX

19. Almeida JR, Silva JM, Oliveira JL. A FAIR approach to real-world health data management and analysis. In: Proceedings of the 36th International Symposium on Computer-Based Medical Systems. 2023 Presented at: CBMS '23; June 22-24, 2023; L'Aquila, Italy p. 892-897 URL: https://ieeexplore.ieee.org/document/10178764 [doi: 10.1109/cbms58004.2023.00338]
20. Derycke P, Kesisoglou I, Korsgaard T, Aage Huru A, Catsyne CA. Report on the landscape analysis of available metadata catalogues and the metadata standards in use. HaDEA & European Union. URL: https://ehds2pilot.eu/wp-content/uploads/2024/04/HealthData@EU-Pilot_MS6.1_FIN.pdf [accessed 2024-04-29]
21. Peng Y, Bathelt F, Gebler R, Gött R, Heidenreich A, Henke E, et al. Use of metadata-driven approaches for data harmonization in the medical domain: scoping review. JMIR Med Inform 2024;12:e52967 [FREE Full text] [doi: 10.2196/52967] [Medline: 38354027]
22. Lueschen G, van der Zee J. Health Systems in the European Union: Diversity, Convergence, and Integration: A Sociological and Comparative Analysis in Belgium, France, Germany, ... and Spain. München, Germany: Walter de Gruyter; 2016.
23. Lauf F, Scheider S, Friese J, Kilz S, Radic M, Burmann A. Exploring design characteristics of data trustees in healthcare - taxonomy and archetypes. In: Proceedings of the 31st European Conference on Information Systems. 2023 Presented at: ECIS '23; June 11-16, 2023; Kristiansand, Norway p. 33-57 URL: https://www.researchgate.net/publication/370060215_Exploring_Design_Characteristics_of_Data_Trustees_in_Healthcare_-_Taxonomy_and_Archetypes
24. Labadie C, Legner C, Eurich M, Fadler M. FAIR enough? Enhancing the usage of enterprise data with data catalogs. In: Proceedings of the 2020 IEEE 22nd Conference on Business Informatics. 2020 Presented at: CBI '20; June 22-24, 2020; Antwerp, Belgium URL: https://ieeexplore.ieee.org/document/9140254 [doi: 10.1109/cbi49978.2020.00029]
25. Jahnke N, Otto B. Data catalogs in the enterprise: applications and integration. Datenbank Spektrum 2023 Jun 21;23(2):89-96. [doi: 10.1007/S13222-023-00445-2]
26. Ehrlinger L, Schrott J, Melichar M, Kirchmayr N, Wöß W. Data catalogs: a systematic literature review and guidelines to implementation. In: Proceedings of the 2021 International Conference on Database and Expert Systems Applications. 2021 Presented at: DEXA '21; September 27-30, 2021; Virtual Event URL: https://link.springer.com/chapter/10.1007/978-3-030-87101-7_15 [doi: 10.1007/978-3-030-87101-7_15]
27. Gröger C. There is no AI without data. Commun ACM 2021 Oct 25;64(11):98-108. [doi: 10.1145/3448247]
28. Oliveira M, Barros Lima GD, Farias Lóscio B. Investigations into data ecosystems: a systematic mapping study. Knowl Inf Syst 2019;61(2):589-630. [doi: 10.1007/s10115-018-1323-6]
29. Sohail SA, Bukhsh FA, van Keulen M. Multilevel privacy assurance evaluation of healthcare metadata. Appl Sci 2021;11(22):10686. [doi: 10.3390/app112210686]
30. Almeida JR, Oliveira JL. MONTRA2: a web platform for profiling distributed databases in the health domain. Inform Med Unlocked 2024;45:101447. [doi: 10.1016/j.imu.2024.101447]
31. vom Brocke J, Winter R, Hevner A, Maedche A. Special issue editorial –accumulation and evolution of design knowledge in design science research: a journey through time and space. J Assoc Inf Syst 2020;21(3):520-544. [doi: 10.17705/1jais.00611]
32. van Aken JE. Valid knowledge for the professional design of large and complex design processes. Des Stud 2005;26(4):379-404. [doi: 10.1016/j.destud.2004.11.004]
33. Hevner AR, March ST, Park J, Ram S. Design science in information systems research. MIS Q 2004;28(1):75-105. [doi: 10.2307/25148625]
34. March ST, Storey VC. Design science in the information systems discipline: an introduction to the special issue on design science research. MIS Q 2008;32(4):725-730. [doi: 10.2307/25148869]
35. Cash PJ. Developing theory-driven design research. Des Stud 2018;56:84-119. [doi: 10.1016/j.destud.2018.03.002]
36. Kundisch D, Muntermann J, Oberländer AM, Rau D, Röglinger M, Schoormann T, et al. An update for taxonomy designers. Bus Inf Syst Eng 2021;64(4):421-439. [doi: 10.1007/s12599-021-00723-x]
37. Nickerson RC, Varshney U, Muntermann J. A method for taxonomy development and its application in information systems. Eur J Inf Syst 2017;22(3):336-359. [doi: 10.1057/ejis.2012.26]
38. Gregor S. The nature of theory in information systems. MIS Q 2006;30(3):611-642. [doi: 10.2307/25148742]
39. Szopinski D, Schoormann T, Kundisch D. Because your taxonomy is worth IT: towards a framework for taxonomy evaluation. In: Proceedings of the 27th European Conference on Information Systems. 2019 Presented at: ECIS '19; June 8-14, 2019; Stockholm-Uppsala, Sweden p. 25-44 URL: https://www.researchgate.net/publication/332711034_Because_your_taxonomy_is_worth_it_Towards_a_framework_for_taxonomy_evaluation
40. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data 2016;3(1):1-9 [FREE Full text] [doi: 10.1038/sdata.2016.18] [Medline: 26978244]
41. Scheider S, Lauf F, Geller S, Möller F, Otto B. Exploring design elements of personal data markets. Electron Markets 2023;33(1):1-16. [doi: 10.1007/S12525-023-00646-3]
42. Alvarellos M, Sheppard HE, Knarston I, Davison C, Raine N, Seeger T, et al. Democratizing clinical-genomic data: how federated platforms can promote benefits sharing in genomics. Front Genet 2022 Jan 10;13:1045450 [FREE Full text] [doi: 10.3389/fgene.2022.1045450] [Medline: 36704354]

XSL•FO
RenderX

43. Good practice guide for the use of the metadata catalogue of real-world data sources. European Medicines Agency. 2022. URL: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/good-practice-guide-use-metadata-catalogue-real-world-data-sources_en.pdf [accessed 2024-04-29]
44. List of metadata for real world data catalogues. European Medicines Agency. 2022. URL: https://www.ema.europa.eu/en/documents/other/list-metadata-real-world-data-catalogues_en.pdf [accessed 2024-04-29]
45. Data directory. BBMRI-ERIC. URL: https://directory.bbmri-eric.eu/ERIC/directory/#/catalogue [accessed 2024-11-02]
46. An interactive catalogue of mental health and wellbeing measures in British cohort and longitudinal studies. Catalogue of Mental Health Measures. URL: https://www.cataloguementalhealth.ac.uk/ [accessed 2024-10-27]
47. Data catalogue for healthcare. Compendium. URL: https://compendiumdatacatalog.com/data-catalog/ [accessed 2024-10-28]
48. Findable, standardised data at scale through the EHDEN database catalogue. EHDEN Portal. URL: https://www.ehden.eu/ehden-portal/ [accessed 2024-10-22]
49. EMIF catalogue. EMIF. URL: https://www.emif.eu/emif-catalogue/ [accessed 2024-10-19]
50. Cancer image Europe. EUCAIM Catalogue. URL: https://catalogue.eucaim.cancerimage.eu/#/ [accessed 2024-10-25]
51. Welcome to the one-stop shop that facilitates access to population health and health care data, information and expertise across Europe. European Health Information Portal. URL: https://www.healthinformationportal.eu/ [accessed 2024-10-18]
52. Data catalogues for health. Fjelltopp. URL: https://www.fjelltopp.org/service/data-catalogues-for-health/ [accessed 2024-10-28]
53. Data catalogues. Health RI. URL: https://catalogus.healthdata.nl/datasets [accessed 2024-10-29]
54. Explore data sources. Helsedata. URL: https://helsedata.no/en/ [accessed 2024-10-29]
55. Health data catalog. IQVIA. URL: https://www.iqvia.com/library/fact-sheets/iqvia-health-data-catalog [accessed 2024-09-25]
56. Health pilot. Kraken. URL: https://www.krakenh2020.eu/pilots/health [accessed 2024-11-01]
57. Precision medicine data catalog. Lifebit. URL: https://www.lifebit.ai/federated-data-catalogue [accessed 2024-11-02]
58. Maelstrom catalogue. Maelstrom Research. URL: https://www.maelstrom-research.org/page/catalogue [accessed 2024-11-26]
59. Trials data catalogue. Yoda. URL: https://yoda.yale.edu/trials-search/ [accessed 2024-10-19]
60. Shi J, Zheng M, Yao L, Ge Y. DIR — a semantic information resource for healthcare datasets. In: Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine. 2017 Presented at: BIBM '17; November 13-16, 2017; Kansas City, MO p. 805-810 URL: https://ieeexplore.ieee.org/abstract/document/8217758 [doi: 10.1109/bibm.2017.8217758]
61. McCoy D, Chand S, Sridhar D. Global health funding: how much, where it comes from and where it goes. Health Policy Plan 2009;24(6):407-417. [doi: 10.1093/heapol/czp026] [Medline: 19570773]
62. Yurkovich M, Avina-Zubieta JA, Thomas J, Gorenchtein M, Lacaille D. A systematic review identifies valid comorbidity indices derived from administrative health data. J Clin Epidemiol 2015;68(1):3-14. [doi: 10.1016/j.jclinepi.2014.09.010] [Medline: 25441702]
63. Pereira A, Almeida JR, Lopes RP, Oliveira JL. Querying semantic catalogues of biomedical databases. J Biomed Inform 2023;137:104272 [FREE Full text] [doi: 10.1016/j.jbi.2022.104272] [Medline: 36563828]
64. Brunswick D. Data privacy, data protection, and the importance of integration for GDPR compliance. Isaca J 2019;1:14-27.
65. Yang K, Jia X, Ren K, Zhang B, Xie R. DAC-MACS: effective data access control for multiauthority cloud storage systems. In: Proceedings of the 2013 IEEE Conference on Computer Communications Workshops. 2013 Presented at: INFOCOM '13; April 14-19, 2013; Turin, Italy p. 2895-2903 URL: https://ieeexplore.ieee.org/document/6567100 [doi: 10.1109/tifs.2013.2279531]
66. Shabani M, Borry P. "You want the right amount of oversight": interviews with data access committee members and experts on genomic data access. Genet Med 2016;18(9):892-897 [FREE Full text] [doi: 10.1038/gim.2015.189] [Medline: 26795589]
67. Munoz-Arcentales A, López-Pernas S, Pozo A, Alonso Á, Salvachúa J, Huecas G. An architecture for providing data usage and access control in data sharing ecosystems. Procedia Comput Sci 2019;160:590-597. [doi: 10.1016/j.procs.2019.11.042]
68. Ngouongo SM, Löbe M, Stausberg J. The ISO/IEC 11179 norm for metadata registries: does it cover healthcare standards in empirical research? J Biomed Inform 2013;46(2):318-327 [FREE Full text] [doi: 10.1016/j.jbi.2012.11.008] [Medline: 23246614]
69. de Mello BH, Rigo SJ, da Costa CA, da Rosa Righi R, Donida B, Bez MR, et al. Semantic interoperability in health records standards: a systematic literature review. Health Technol (Berl) 2022;12(2):255-272 [FREE Full text] [doi: 10.1007/s12553-022-00639-w] [Medline: 35103230]
70. Iroju O, Soriyan A, Gambo I, Olaleke J. Interoperability in healthcare: benefits, challenges and resolutions. Int J Innov Appl Stud 2013;3(1):262-270.
71. Roehrs A, da Costa CA, da Rosa Righi R, Rigo SJ, Wichman MH. Toward a model for personal health record interoperability. IEEE J Biomed Health Inform 2019;23(2):867-873. [doi: 10.1109/jbhi.2018.2836138]
72. Kent S, Burn E, Dawoud D, Jonsson P, Østby JT, Hughes N, et al. Common problems, common data model solutions: evidence generation for health technology assessment. Pharmacoeconomics 2021;39(3):275-285 [FREE Full text] [doi: 10.1007/s40273-020-00981-9] [Medline: 33336320]
73. Ivanović M, Budimac Z. An overview of ontologies and data resources in medical domains. Expert Syst Appl 2014;41(11):5158-5166. [doi: 10.1016/j.eswa.2014.02.045]

XSL•FO
RenderX

74. Hentschel J. Contextuality and data collection methods: a framework and application to health service utilisation. J Dev Stud 1999;35(4):64-94. [doi: 10.1080/00220389908422581]
75. Tijhuis M, Finger JD, Slobbe L, Sund R, Tolonen H. Data collection. In: Marieke V, van Oers H, editors. Population Health Monitoring: Climbing the Information Pyramid. Cham, Switzerland: Springer; 2019:59-81.
76. Bohensky MA, Jolley D, Sundararajan V, Evans S, Pilcher DV, Scott I, et al. Data linkage: a powerful research tool with potential problems. BMC Health Serv Res 2010;10(1):346-354 [FREE Full text] [doi: 10.1186/1472-6963-10-346] [Medline: 21176171]
77. Rasmussen KB, Blank G. The data documentation initiative: a preservation standard for research. Arch Sci 2007;7(1):55-71. [doi: 10.1007/S10502-006-9036-0]
78. Angrist M. Eyes wide open: the personal genome project, citizen science and veracity in informed consent. Per Med 2009;6(6):691-699 [FREE Full text] [doi: 10.2217/pme.09.48] [Medline: 22328898]
79. Kitchenham B, Pearl Brereton O, Budgen D, Turner M, Bailey J, Linkman S. Systematic literature reviews in software engineering – a systematic literature review. Inf Softw Technol 2009;51(1):7-15. [doi: 10.1016/j.infsof.2008.09.009]
80. Webster J, Watson RT. Analyzing the past to prepare for the future: writing a literature review. MIS Q 2002;26(2):13-23 [FREE Full text]
81. Mwita K. Factors influencing data saturation in qualitative studies. Int J Bus Soc 2022 Jun 05;11(4):414-420. [doi: 10.20525/ijrbs.v11i4.1776]
82. Proposal for a regulation - the European Health Data Space. Directorate-General for Health and Food Safety. 2022. URL: https://health.ec.europa.eu/publications/proposal-regulation-european-health-data-space_en#details [accessed 2024-11-07]
83. Saldana J. The Coding Manual for Qualitative Researchers. Thousand Oaks, CA: SAGE Publications; 2021.
84. Pratt MG. Fitting oval pegs into round holes. Organ Res Methods 2007;11(3):481-509. [doi: 10.1177/1094428107303349]
85. Fruhwirth M, Rachinger M, Prlja E. Discovering business models of data marketplaces. In: Proceedings of the 53rd Hawaii International Conference on System Sciences. 2020 Presented at: HICSS '20; January 7-10, 2020; Maui, HI p. 5738-5747 URL: https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/cf7bab54-478b-412a-8742-6b02f10dd7ca/content [doi: 10.24251/hicss.2020.704]
86. Database (part of Elixir infrastructure). Elixir BioSamples. URL: https://www.ebi.ac.uk/biosamples/ [accessed 2024-12-03]
87. A registry of knowledgebases and repositories of data. Elixir FAIRsharing. URL: https://fairsharing.org/search?fairsharingRegistry=Database [accessed 2024-12-04]
88. Global health data exchange. Institute for Health Metrics and Evaluation. URL: https://ghdx.healthdata.org/ [accessed 2024-10-29]
89. MACH clinical and research dataset catalogue. Melbourne Academic Centre for Health. URL: https://figshare.unimelb.edu.au/MACH-catalogue?searchMode=1 [accessed 2024-11-24]
90. Möller F, Stachon M, Azkan C, Schoormann T, Otto B. Designing business model taxonomies – synthesis and guidance from information systems research. Electron Mark 2021;32(2):701-726. [doi: 10.1007/s12525-021-00507-x]
91. Pine KH. The qualculative dimension of healthcare data interoperability. Health Informatics J 2019;25(3):536-548 [FREE Full text] [doi: 10.1177/1460458219833095] [Medline: 31002277]
92. Wang RY, Strong DM. Beyond accuracy: what data quality means to data consumers. J Manag Inf Syst 2015;12(4):5-33. [doi: 10.1080/07421222.1996.11518099]
93. Spiekermann S, Acquisti A, Böhme R, Hui KL. The challenges of personal data markets and privacy. Electron Markets 2015;25(2):161-167. [doi: 10.1007/S12525-015-0191-0]

Abbreviations

CDM: common data model
DAC: Data Access Committee
E2C: empirical-to-conceptual
EHDEN: European Health Data and Evidence Network
EHDS: European Health Data Space
EHDS2: European Health Data Space 2
ETL: extract, transform, load
EU: European Union
EUCAIM: European Federation for Cancer Images
FAIR: findability, accessibility, interoperability, and reusability
HMDC: health metadata catalog
IDERHA: Integration of Heterogeneous Data and Evidence towards Regulatory and Health Technology Assessment Acceptance
IS: information systems
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RQ: research question

XSL•FO
RenderX

RWD: real-world data
RWE: real-world evidence
SLR: structured literature review
TEHDAS: Towards European Health Data Space

Edited by A Mavragani; submitted 19.06.24; peer-reviewed by Y Arayici, M Gersch; comments to author 06.08.24; revised version received 20.08.24; accepted 12.12.24; published 18.02.25.

Please cite as:
Scheider S, Mallick MK
Exploring Metadata Catalogs in Health Care Data Ecosystems: Taxonomy Development Study
JMIR Form Res 2025;9:e63396
URL: https://formative.jmir.org/2025/1/e63396
doi:10.2196/63396
PMID:39964739

©Simon Scheider, Mostafa Kamal Mallick. Originally published in JMIR Formative Research (https://formative.jmir.org), 18.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Design and Evaluation of a Simulation Tool for Audiology Screening Education: Design Science Approach

John Gerdes[1], BSc, MSc, MEng, MBA, PhD; Benjamin Schooley[2], MBA, PhD; Dakota Sharp[3], AuD; Juliana Miller[4], BSc, MSc, CCC-SLP

[1]Department of Integrated Information Technology, University of South Carolina, Columbia, SC, United States

[2]College of Engineering and Computing, Brigham Young University, Provo, UT, United States

[3]School of Medicine, University of North Carolina, Durham, NC, United States

[4]Department of Communication Sciences and Disorders, University of South Carolina, Columbia, SC, United States

Corresponding Author:
John Gerdes, BSc, MSc, MEng, MBA, PhD
Department of Integrated Information Technology
University of South Carolina
Bert Storey Innovation Center, Suite 1300
550 Assembly Street
Columbia, SC, 29208
United States
Phone: 1 803 777 8529
Email: jgerdes@sc.edu

Abstract

Background: The early identification of hearing loss and ear disorders is important. Regular screening is recommended for all age groups to determine whether a full hearing assessment is necessary and allow for timely treatment of hearing problems. Procedural training is needed for new speech-language pathology students as well as continuing education for those trained to perform this screening procedure. Limited availability and access to physical training locations can make it difficult to receive the needed training.

Objective: The aims of this study were to (1) develop a new hearing screening simulation software platform and (2) assess its effectiveness in training a group of graduate-level speech-language pathology students in hearing screening procedures.

Methods: An audiology simulator modeled after the commercial Grason-Stadler GSI39 combination audiometer and tympanometer device was developed to serve as a precursor to traditional face-to-face clinical instruction. A description of the simulator development process, guided by a design science approach, is presented. The initiation phase established the initial criteria for the simulator design. This was followed by an iterative process involving prototype development, review, and critique by the clinical faculty. This feedback served as input for the subsequent iteration. The evaluation of the final prototype involved 33 speech-language pathology graduate students as part of an introductory audiology class. These students were randomly assigned to control (receiving in-person instruction) and test (in-person instruction and simulation tool use) groups. Students in both groups were subsequently evaluated as they performed audiology screenings on human participants and completed a 25-item pretest and posttest survey. Nonparametric Mann-Whitney *U* tests were conducted on the mean differences between pretest and posttest ordinal survey response data to compare the control and intervention groups.

Results: The results indicated that the students who used the simulation tool demonstrated greater confidence in their ability to (1) explain hearing screening procedures to a child (P=.02), (2) determine whether otoscopy results are normal (P=.02), and (3) determine whether otoscopy results are abnormal (P=.03). Open-ended responses indicated that the students found that the hands-on experience provided by the simulator resulted in an easy-to-use and useful learning experience with the audiometer, which increased their confidence in their ability to perform hearing screenings.

Conclusions: Software-based education simulation tools for audiology screening may provide a beneficial approach to educating students and professionals in hearing screening training. The tool tested in this study supports individualized, self-paced learning with context-sensitive feedback and performance assessment, incorporating an extensible approach to supporting simulated subjects.

(JMIR Form Res 2025;9:e47150) doi:10.2196/47150

XSL•FO
RenderX

KEYWORDS

design science; audiology; simulation; hearing screening; framework; speech pathology; training

Introduction

Background

A national study in the United States found that 14.9% of children ranging in age from 6 to 19 years had low- or high-frequency hearing loss in one or both ears [1]. Early identification, diagnosis, and intervention are beneficial, improving the likelihood that children will develop effective communication and language skills and achieve successful learning outcomes [2]. For this reason, hearing screenings should be carried out regularly to identify problems early. To accomplish this, health care students and personnel must receive adequate training on how to perform these screenings and appropriately refer patients for evaluation [3].

The American Speech-Language-Hearing Association guidelines for audiological screening [4] specify procedures for completing hearing loss screenings, along with recommendations for screening the outer and middle ears, collecting an optional case history, performing a visual examination (otoscopy), and conducting acoustic immittance testing (tympanometry). Those who do not pass a hearing screening (the term “fail” is not used) are typically referred to a licensed audiologist for more detailed hearing testing or to a physician for any medical concerns. The screener can also refer the individual to both an audiologist and physician when the results warrant such action.

The American Speech-Language-Hearing Association recognizes the use of standardized patients and simulation technologies as alternatives to traditional clinical education methods [5]. The use of clinical simulation increased significantly during the COVID-19 pandemic because many practicum sites became inaccessible to students. Simulation is increasingly being used in clinical training and has been found to improve clinical knowledge and skills, self-confidence, communication skills, empathy, critical thinking abilities, leadership, and situation management among students [6-11]. Simulation has been found to improve the confidence in both knowledge and skills of audiology and speech-language pathology graduate students during training [12,13]. A survey of clinical educators found that simulation was a useful learning tool, provided a quality control measure, was an aid in professional development, and exposed students to clinical pathologies that were not routinely encountered [13]. Barriers to the use of simulation include faculty lacking training with simulators, limited funding to purchase simulator technology, and a lack of resources such as time and space [13].

Training speech-language pathology students to perform hearing screenings is an important undertaking. Students must learn how to interact with children and adults, properly work through the assessment protocol, and learn how to operate the equipment.

Studies have shown simulation to be effective in audiology education and training [14-16]. Simulation can help transfer theory to practice in an integrated teaching and learning model [17]. Simulators can present a full range of patient cases that represent real-world clinical diagnoses. These simulated cases present patients with symptoms that must be detected and diagnosed by the student. However, the use of simulation in audiology education is still in its infancy compared to some other fields, such as nursing and medicine [3].

It can be challenging for speech-language pathology students to obtain the necessary audiology training due to limited practicum opportunities. One alternative is to use a commercial simulator application, such as Otis–the virtual patient [18], which is a simulator designed for use in the education of audiologists. Many of these commercial simulators are designed to train students in conducting full audiological evaluations, rather than teaching hearing screening skills that follow an abbreviated protocol. However, these simulators can be cost prohibitive for training speech-language pathology students. Our goal was to deliver a realistic simulation tool providing a realistic representation of the controls found on a commercial testing instrument that students would work with when in the field. This design allows students to practice and develop testing skills without having to travel to a testing center. It also exposes them to a broad array of clinical cases and provides feedback on their performance. These goals became particularly relevant during the COVID-19 pandemic, when participating in live practicum experiences became difficult.

Prior research has explored the use of pure-tone audiometry simulators [19-23]. One system incorporated multilingual capability, improving accessibility for non-English speakers [21]. Audiology training using virtual reality was found to yield positive results compared to traditional training [11]. There are also commercially available simulation tools that are useful for training audiology skills. AudSim Flex [24] offers basic and advanced simulators for audiology students to learn pure-tone and bone conduction testing, both masked and unmasked. Otis–the virtual patient [18] realistically simulates the behavior of the patient and detects possible user errors immediately. It supports both pure-tone and screening training and guides the learner through exercises at several levels of difficulty, while providing useful help.

There are similarities between the hearing screening simulator (HSS) we developed and existing simulation tools. All incorporate a set of simulated subjects that present a range of hearing disorders, provide background narratives of the subjects to provide context, and model human behavior or actual subjects to improve realism. Most tools allow students to perform a simulated assessment and collect data that are then plotted on a standard audiogram, with the simulator assessing and critiquing the students’ analysis and conclusions.

The HSS differs from other simulators in 2 important ways. First, the HSS provides a digital twin—a realistic representation of the Grason-Stadler GSI39 combination audiometer and tympanometer device [25]—whereas other simulators use a generic design. By modeling the actual instrument interface, the HSS allows the student to become familiar with the instrument layout and practice working with the modeled

XSL•FO
RenderX

device's controls. Second, the HSS is designed for the simpler hearing screening protocol rather than full audiological testing. The protocols are different; therefore, the feedback and assessment verification must be tailored accordingly. Full audiological testing is outside the scope of practice for speech-language pathology students.

A design science approach was adopted for the development of the HSS. Initial design specifications were first established, followed by an iterative process involving prototype development, review, and critique by the clinical faculty, with the feedback used to refine the next iteration. After several iterations, the final prototype was tested to assess whether students felt that the simulator was useful in learning how to administer hearing screenings and to determine their confidence in their skills after using the simulator.

Objectives

The aims of this study were to (1) develop a new hearing screening simulation software platform and (2) assess its effectiveness in training a group of speech-language pathology students in hearing screening procedures. This work's contributions include a framework for developing a simulation of an existing process (hearing screening testing) and the description of a unique artifact with a user interface that closely emulates an existing commercial tool used to perform hearing screenings. This artifact supports an individualized, self-paced learning environment incorporating context-sensitive feedback and performance assessment. While further testing is needed, initial testing indicates that the HSS shows promise as a training aid for speech-language pathology students.

Methods

Overview

This research follows a design science methodology consisting of (1) problem identification and motivation, (2) the definition of the objectives for a solution, (3) design and development, (4) demonstration, (5) evaluation, and (6) communication [26,27]. We identified and translated design requirements for an audiology simulator into design components. We then classified these functional components in the application into design principles and features for designing a health care simulation app for clinical education and training. An abbreviated description of the design phase is provided in the next subsection, offering context for the evaluation phase of the study. For a more detailed description of the development process, the interested reader is directed to the conference paper [28]. For the evaluation of the simulation tool, participants were randomly separated into control (did not use the simulator) and test (used the simulator) groups, with both groups completing a pretest and posttest survey. Survey responses were analyzed using the nonparametric Mann-Whitney *U* test, which is appropriate for ordinal data analysis when a normal distribution cannot be assumed. Four qualitative survey question responses were also collected and analyzed. Moreover, audiology instructors assessed the skill levels of the student participants. Additional details on the analysis can be found in the Evaluation Study subsection.

The Simulation Tool's Conceptual Framework, Requirements, and Design

The simulation tool was designed to train speech-language pathology students on the proper use of audiometer equipment, how to administer a hearing screening test, and how to interpret the results. The tool's clinical and administrative functions, which are summarized in Textbox 1, served as the design criteria for the HSS. The clinical requirements reflect the data typically collected and presented during a traditional training session using commercial equipment. The operational requirements include additional features that enable a web-based digital learning experience, such as calculating student grades, and therefore represent value-added aspects of the simulator.

The HSS is designed to simulate 3 clinical procedures used in hearing screening and audiological evaluations. The otoscopic examination is a visual procedure used to assess the condition of the external auditory canal and tympanic membrane. Tympanometry tests how well the subject's eardrum moves and determines the volume of the ear canal and the pressure in the middle ear space. The audiometer presents a series of tones to the subject based on the appropriate protocol, and the subject indicates whether they hear the tone by raising their hand. For screening tests, tones are presented at a fixed level considered "normal hearing" (20 dB hearing level), in contrast to a typical diagnostic hearing test where multiple levels are tested to determine the subject's hearing threshold at each frequency (ie, the lowest level the subject can hear). The HSS uses standardized subjects with known auditory hearing profiles, along with sample otoscopic images and tympanometry results for both ears.

The simulator user interface consists of 3 panels. Figure 1A shows the simulated audiometer device, which was modeled after the Grason-Stadler GSI39 combination audiometer and tympanometer device [25]. The upper portion of the unit contains a graphic display that shows either the tympanogram or audiogram depending on the testing mode selected. The lower portion of the unit provides the controls used to perform the screening. These controls and their purpose are described herein. Figure 1B portrays the panel that provides information about the subject, as well as the command buttons to show and print a larger version of the audiogram chart, and displays the debriefing report that summarizes the students' performance on the simulation. Figure 1C shows the technician notes panel. It allows students to record their findings and include their observations, as is typical in a traditional audiological screening.

XSL•FO
RenderX

Textbox 1. Required clinical and operational functionalities of the hearing screening simulator.

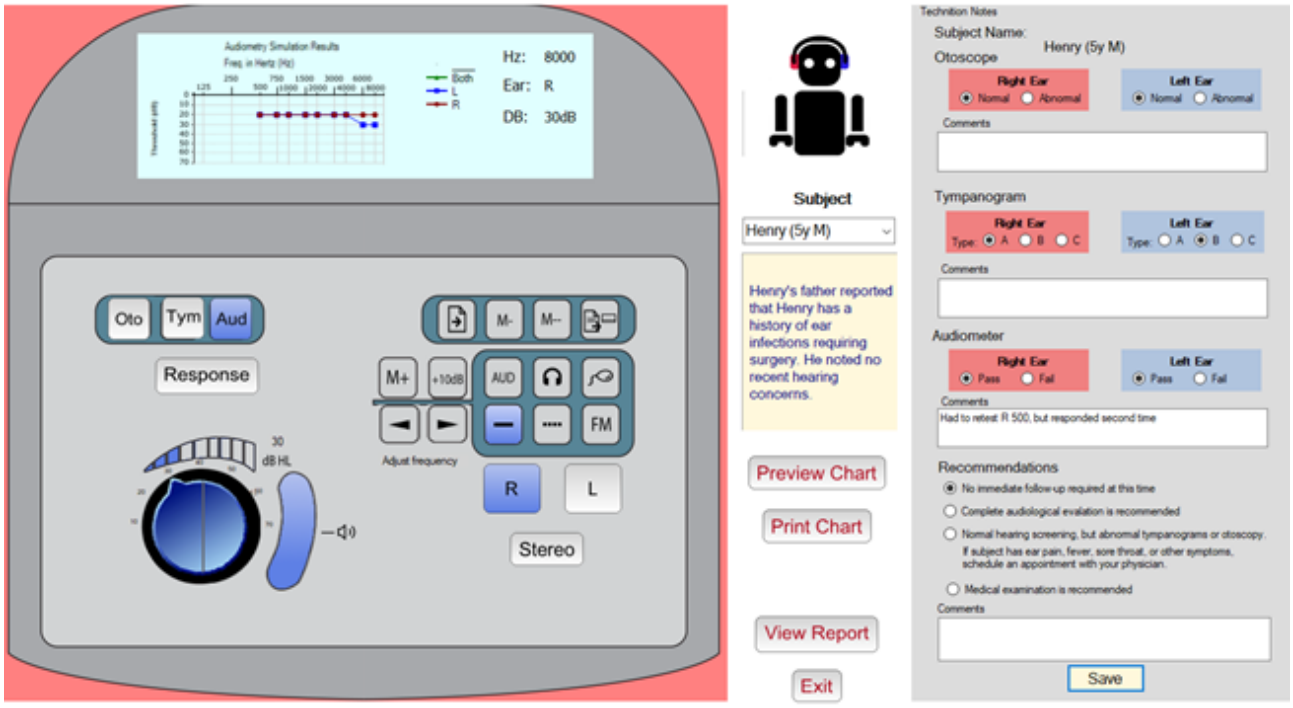
Clinical requirements

- Authenticate and authorize users
- Provide a graphical user interface similar to that of standard audiology screening hardware-based systems
- Simulate 3 audiometry screens: video otoscopy, tympanometry, and audiometry
- Configure audiometer settings for a specific subset of frequencies used in hearing screening
- Display output similar to that obtained with commercial devices, including otoscopic images, tympanograms, and audiograms
- Include simulated patients representative of a range of real-world clinical cases

Operational requirements

- Provide a web-based platform
- Track and report time spent on the simulation
- Document and print a record of the simulation session
- Include automatic grading and feedback on students' performance
- Provide a full-featured audiometer with additional features beyond those needed for the hearing screening learning experience but present in commercial units
- Log all simulation screenings to permit reporting of an individual student's simulation history and enable cross-sectional data analysis across multiple students' results

Figure 1. The hearing screening simulator user interface. (A) The simulated audiometer device. (B) The panel that provides the simulated subject's details, with the avatar responding to tone stimulus. (C) Area for the technician to enter results and comments.



The HSS is equipped with 10 simulated patients, with the capability to expand this list. Each patient has a short case history, which is presented to the student when the subject is selected, providing the subject's age and sex and a brief biographical background related to any hearing concerns. The 10 artificial subjects represent different audiological conditions and pathologies, with audiometry, tympanometry, and otoscopy results consistent with their condition. One child and 1 adult standard patient with normal results were included. To obfuscate the identities of these simulated patients and create the illusion of having a much larger subject base, the profile information is randomized each time the program is run. Subject's sex is randomly assigned, age is adjusted slightly, the subject's name is randomly selected from a list of 100 diverse names, and the backstory is modified to match the new subject profile. This feature expands the number of possible simulated patient profiles to >1000 unique profiles. This prevents students from circumventing the training by using results from earlier simulations and also enables the assessment of a student's

XSL•FO
RenderX

precision when presented with multiple subjects derived from the same base simulated patient.

To perform the training, the student selects a standardized subject and reads the brief biographical sketch provided. The simulated screening test is then performed. The student must correctly set up the audiometer and work through the testing procedure. The otoscopic examination is simulated by selecting either the right or left ear and pressing the "Oto" button, which displays the subject's right or left otoscopic image, depending on which ear is specified. Similarly, the student simulates the tympanometry test by selecting the ear and pressing the "Tymp" button, displaying the subject's tympanometry result. To simulate the auditory test, the student selects the ear being tested, the tone level (20 dB hearing level) for the test, and then the series of frequencies specified by the testing protocol.

As the student proceeds through the simulation, they are asked to record their observations and patient results in the technician notes panel. There are note sections for each of the 3 screening procedures (otoscopy, tympanometry, and audiometry). The student is asked to provide the result for each ear and to add comments and observations related to the screening. Recording this information allows the simulator to assess the student's performance. The system knows which result is correct based on the known profile of the subject being screened. Taking into consideration the 3 individual screening procedures, the student must also make a final recommendation regarding the patient's subsequent follow-up and add a related comment. The student's final recommendation is automatically graded and reported by the simulator, immediately providing appropriate feedback as needed. The patient cases, associated correct results, and correct recommendations were defined by the study clinical team and implemented in the application database.

Once the student completes the screening and enters their session notes, the data are saved to a database, and a session report is generated (Figure 2). This report serves as a starting point for a debriefing that would occur between a facilitator and the student. It provides metadata related to the simulation, including the technician's name (ie, the student's name), the subject's name, the duration of the testing portion of the simulation, the screening procedures run by the student during the evaluation, and the total time spent on the simulation session. This information allows the simulator to perform validity checking of the training session, including whether the proper procedure was followed and the proper tests were run. The session report provides information to the simulation facilitator about the steps completed by the student during the session. Any shortcuts taken by the student during the session would be noted in the report. The report also displays the student's results along with the correct results, with an explanation justifying the results. This provides the student with immediate feedback on their performance as well as their performance over time. Reports can be printed or saved as a PDF file to serve as required documentation of their clinical training.

XSL•FO
RenderX

Figure 2. Simulation report summary displaying student performance in administering a hearing screening to a simulated client. It includes session metadata and a simulation debriefing report that provides the students' responses, notes, and correct results.

Technician: John Smith

Date : Wednesday, June 2, 2021 9:34:46 AM

Testing duration:	36.6 Min		
Number Of Subjects tested:	1		
Auditory tests run:	21 right	4 left	0 both
Tympanometry tests run:	1 right	1 left	
Otoscope images viewed:	1 right	1 left	

Subject:	Henry (5y M)			
	Henry's father reported that Henry has a history of ear infections requiring surgery. He noted no recent hearing concerns.			
Otoscope	Right		Left	
Status:	Normal	[✗ - Incorrect]	Normal	[✗ - Incorrect]
Comment:				
Correct Response:	Abnormal, pressure equalization tube present in eardrum		Abnormal, PE tube	
Tympanogram	Right		Left	
Type:	Type A	[✗ - Incorrect]	Type B	[✓ - Correct]
Comment:				
Correct Response:	Type B, no eardrum movement and large ear canal volume		Type B, no eardrum movement and large ear canal volume	
Audiometer	Right		Left	
Screening:	Pass	[✓ - Correct]	Pass	[✓ - Correct]
Comment:	Had to retest R 500, but responded second time			
Correct Response:	Pass		Pass	
Recommendation: No immediate follow-up required at this time				[✗ - Incorrect]
Comment:				
Correct Response:	Normal hearing screening, but abnormal tympanograms or otoscopy. If you have ear pain, fever, sore throat, or other symptoms, schedule an appointment with your physician			

Evaluation Study

Student Participant Sampling, Intervention Procedure, and Data Collection

All 33 first-year speech-language pathology graduate students enrolled in an Introduction to Audiology class participated in this pilot study. The study started with 34 students, but since one student in the text group failed to submit the posttest survey they were dropped from the analysis resulting in 33 subjects. This is a required course in the graduate Speech-Language Pathology curriculum.

All participants received the same lecture on hearing loss concepts and performing hearing screening assessments. The 33 students were randomly assigned to either the control (n=16, 48%; all female) or test (n=17, 52%; all female) group by listing all participants in Microsoft Excel and using the built-in

XSL•FO
RenderX

randomization function. The control group received brief in-person training on equipment operation before being evaluated when they conducted their live, in-person audiology screening. The test group received the same in-person equipment operation introduction but used the simulation tool before being evaluated doing their live audiology screening. Students in each group were divided into 2-member teams, except for 1 group of 3 in the test group. Each student team alternated between roles, acting as a test subject (ie, patient) and as the test evaluator. The practice of students acting as both subject and evaluator is common in small class settings for audiology training. Approximately half of the students in the control group (16/33, 48%) and those in the test group (17/33, 52%) had previous hands-on experience with a hearing screening audiometer. Randomization included sorting students into an approximately equal number of pairings of "no previous experience," "1 partner with previous experience," and "both partners with previous experience."

Participants in the test group received supplementary training with the simulation tool. They were introduced to the simulator and given approximately 15 minutes to work through approximately 5 cases with their partner. A 25-item questionnaire was administered to all students to assess their confidence in their skills both before and after performing the live hearing screening. This means that the control and test groups completed the questionnaire both before and after the intervention. The questionnaire was created by the audiology faculty to gauge students' knowledge and confidence in administering and interpreting hearing screenings. A 5-point Likert scale was used. The response options and corresponding coding values were *strongly disagree* (1), *disagree* (2), *neither disagree nor agree* (3), *agree* (4), and *strongly agree* (5). A set of 7 to 9 questions were asked about each assessment mode (otoscopy, tympanometry, and audiometry). The test group was also asked 4 additional free-response questions in the same questionnaire to assess their perceptions of the simulation tool. A total of 16 surveys were completed by the control group and 17 completed by the test group. One test group survey was excluded from analysis because it represented the second of 2 surveys completed by the same participant, each providing identical feedback. Thus the analysis use 16 surveys in each group.

Student Participant Survey Data Analysis

A nonparametric analysis was conducted using the Mann-Whitney *U* test to compare differences between the groups at baseline (ie, before the test) and after the intervention (ie, after the test). The goal was to assess the change in the test and control groups' pretest versus posttest survey responses. The Mann-Whitney *U* test is a nonparametric method suitable for analyzing ordinal data when a normal distribution cannot be assumed. In this case, the ordinal variables were responses to survey questions measured on a Likert scale. This method is also appropriate when the groups being analyzed represent 2 independent samples (test and control groups created via randomization). The null hypothesis is that the means of each group are identical, meaning that all groups come from the same distribution. The alternative hypothesis is that at least 1 of the groups has a different mean, meaning that at least 1 group comes from a different distribution than the others. The test was corrected for tied ranks.

It was expected that using the simulation training tool together with in-person training would be perceived by the students to be at least as effective as, and possibly an improvement over, in-person–only training using commercial audiology equipment. Although some indication of improvement was expected, statistically significant differences between the groups were not expected for most questionnaire items between the test and control groups. All statistical analyses were conducted using a standard, commercially available statistical software tool (NCSS 2021; NCSS Statistical Software). A thematic analysis of qualitative responses was conducted to produce categorical findings, which were then discussed.

Student Qualitative Data Collection and Analysis

As part of the posttest survey administered to the test group, 4 free-response questions were included regarding their impressions of using the simulator tool. Qualitative responses were typed, and we conducted a thematic analysis across the responses, which were then grouped, summarized, and counted.

Instructor Evaluation of Student Hearing Screening Assessment Technique

A live assessment of the students' clinical techniques was completed using commercial audiometry units, with peers role-playing as patients. Certified instructors supervised the assessment, observing and grading the students' performance on otoscopy, tympanometry, and hearing screening techniques on a scale ranging from 1 to 5. Scores were compiled, averaged, and compared across the control and test groups.

Ethical Considerations

The institutional review board of the University of South Carolina determined that this research did not involve human subjects. All participants received a verbal and written description of the study and consented to participate. Participants' identities were kept confidential and anonymized before data analysis. No compensation was provided for participation in the study.

Results

The Simulation Tool's Conceptual Framework

Overview

In a design science study, the objective is to assess how well the created artifact, in this case, the HSS, meets the target operational requirements. In this study, we focused on assessing how well the tool performed in delivering the necessary training on the audiometer equipment. The results are presented in the following subsections.

Evaluation Study

Likert Scale Survey Results

The test group showed higher posttest mean scores and greater percentage change than the control group across all 3 test component groupings: otoscope, tympanometer, and hearing screening (Table 1).

XSL•FO
RenderX

Table 1. Multifactor assessment of student confidence change resulting from the use of the simulation tool. Pretest and posttest values represent the averages for both control and test groups.

	Control group (n=16)			Test group (n=17)		
	Pretest score, mean (SD)	Posttest score, mean (SD)	Change (% change)	Pretest score, mean (SD)	Posttest score, mean (SD)	Change (% change)
Otoscope (7 factors)	3.59 (0.41)	3.90 (0.63)	0.31 (8.64)	3.73 (0.58)	4.28 (0.55)	0.55 (14.75)
Tympanometer (9 factors)	3.63 (0.32)	4.28 (0.50)	0.63 (17.35)	3.79 (0.70)	4.62 (0.44)	0.83 (21.90)
Hearing screening (9 factors)	4.01 (0.56)	4.39 (0.45)	0.38 (9.48)	3.84 (0.59)	4.43 (0.45)	0.59 (15.36)

As shown in Table 2, the responses to the following 3 survey questions showed statistically significant differences between the groups with significance set at $P<.05$: “I am confident in my ability to explain hearing screening procedures to a child” (Mann-Whitney $U=71.0$, n1=n2=16; $P=.02$, 2-tailed t test), “I am confident in my ability to determine if otoscopy is normal” (Mann-Whitney n1=n2=16; $P=.02$, 2-tailed), and “I am confident in my ability to determine if otoscopy is abnormal” (Mann-Whitney $U=74.5$, n1=n2=16 $P=.02$, 2-tailed t test). For technology pilot studies, findings at the $P<.10$ significance level can also demonstrate important results. In this sense, the responses to the following 2 survey questions showed statistically significant differences between the groups with significance set at $P<.10$: “I am confident in my ability to locate landmarks while completing otoscopy” (Mann-Whitney $U=86.5$, n1=n2=16; $P=.08$, 2-tailed t test) and “I am confident which dB level should be tested during a hearing screening” (Mann-Whitney $U=84.0$, n1=n2=16; $P=.07$). The results were not significant across 19 survey questions at the $P<.10$ significance level, underscoring the specific impact of the simulator on the aforementioned confidence measures.

The survey asked whether students had previously conducted an audiology screening test on a human subject before the class audiology training exercise and study. A total of 8 (24%) of the 33 participants in the test group and 6 (18%) of the 33 participants in the control group reported in the affirmative. For this smaller subset of participants, mixed results showed higher posttest mean scores and greater percentage change for the otoscope and tympanometer control groups than the test groups. There was a negative change and percentage change for the combined hearing screening scores for the control group, while the test group had a positive change and percentage change (Table 3). Statistically significant differences were not calculated due to the small sample sizes across the groups.

XSL•FO
RenderX

Table 2. Analysis of survey responses grouped by otoscopy, tympanometry, and hearing screening questions, comparing the control and test group mean scores for pre- and postassessment sentiment. The Mann-Whitney U test was used to assess the statistical significance of the change in sentiment between the 2 groups.

Survey questions	Control group		Test group		*P* value
	Preassessment score, mean (SD)	Postassessment score, mean (SD)	Preassessment score, mean (SD)	Postassessment score, mean (SD)	
Otoscopy					
I am confident in my ability to interpret otoscopy	3.63 (0.62)	3.93 (0.80)	3.65 (1.00)	4.06 (0.75)	.74
I am confident in my ability to complete otoscopy on a patient	3.19 (0.75)	3.80 (0.77)	3.24 (1.15)	4.18 (0.64)	.39
I am confident in my ability to locate landmarks while completing otoscopy	3.44 (0.63)	3.73 (0.70)	3.53 (0.94)	4.35 (0.70)	.08[a]
I am confident in my ability to determine if otoscopy is abnormal	3.81 (0.54)	3.60 (1.06)	3.82 (0.73)	4.18 (0.81)	.03[b]
I am confident in my ability to determine if otoscopy is normal	3.88 (0.50)	4.00 (0.85)	3.94 (0.66)	4.41 (0.62)	.02[b]
I am confident in my ability to determine if otoscopy warrants a referral	3.75 (0.68)	3.93 (0.59)	3.88 (0.49)	4.24 (0.75)	.43
I am confident in my understanding of the impact of abnormal otoscopy on hearing screening results	3.38 (0.89)	4.27 (0.46)	4.06 (0.66)	4.56 (0.51)	.32
Tympanometry					
I am confident in my ability to interpret a tympanogram	3.53 (0.50)	4.40 (0.51)	3.76 (0.97)	4.65 (0.49)	.43
I can confidently identify a normal ear canal volume (ECV) on a tympanogram	3.81 (0.54)	4.33 (0.49)	4.06 (0.56)	4.65 (0.49)	.45
I can confidently differentiate between tympanogram types	3.75 (0.45)	4.33 (0.62)	3.88 (0.86)	4.65 (0.49)	.76
I can confidently identify an abnormal tympanogram	3.81 (0.40)	4.33 (0.62)	4.00 (0.94)	4.71 (0.47)	.66
I can confidently identify a normal tympanogram	3.88 (0.34)	4.40 (0.51)	4.24 (0.56)	4.76 (0.44)	.86
I can confidently complete a tympanogram on a patient	3.13 (0.81)	4.29 (0.83)	3.18 (1.19)	4.59 (0.51)	.51
I am confident in my ability to determine if a tympanometry result warrants a referral	3.69 (0.70)	4.13 (0.52)	3.65 (0.86)	4.53 (0.62)	.12
I am confident in my understanding of the impact of abnormal tympanometry on hearing screening results	3.38 (0.81)	4.20	3.76 (0.75)	4.59 (0.51)	.85
I am confident in my ability to connect abnormal otoscopy results to abnormal tympanometry results	3.56 (0.51)	3.93 (0.80)	3.59 (1.12)	4.47 (0.62)	.76
Hearing screening					
I am confident in which frequencies should be tested during a hearing screening	4.31 (0.60)	4.79 (0.43)	4.12 (0.78)	4.65 (0.61)	.24
I am confident which dB level should be tested during a hearing screening	4.13 (0.72)	4.53 (0.52)	3.76 (0.90)	4.53 (0.72)	.07[a]
I am confident in my ability to explain hearing screening procedures to an adult	3.63 (0.96)	4.27 (0.70)	3.65 (0.86)	4.65 (0.49)	.42

XSL•FO
RenderX

Survey questions	Control group		Test group		*P* value
	Preassessment score, mean (SD)	Postassessment score, mean (SD)	Preassessment score, mean (SD)	Postassessment score, mean (SD)	
I am confident in my ability to explain hearing screening procedures to a child (e.g., conditioned play audiometry)	3.75 (1.00)	4.00 (0.85)	3.38 (0.81)	4.29 (0.85)	.02[b]
I am confident in how to present stimuli/tones while completing a hearing screening	3.81 (0.98)	4.36 (0.50)	3.47 (1.07)	4.65 (0.49)	.11
I am confident in how to verify a response while completing a hearing screening	4.00 (0.82)	4.27 (0.70)	4.00 (0.87)	4.53 (0.51)	.22
I am confident in knowing when a referral is necessary based on hearing screening results	4.06 (0.57)	4.50 (0.52)	3.94 (0.75)	4.41 (0.62)	.60
I am confident in knowing which professional to refer to based on hearing screening results	4.38 (0.50)	4.47 (0.52)	4.35 (0.49)	4.47 (0.62)	.75

[a]Significance set at $P<.10$.

[b]Significance set at $P<.05$.

Table 3. Multifactor assessment of student confidence change for students reporting having conducted an audiology screening test on live subjects before the audiology test. Pretest and posttest values represent the average values.

	Control group (n=6)			Test group (n=8)		
	Pretest scores, mean (SD)	Posttest scores, mean (SD)	Change (% change)	Pretest scores, mean (SD)	Posttest scores, mean (SD)	Change (% change)
Otoscope (7 factors)	3.83 (0.40)	4.36 (0.40)	0.53 (13.8)	3.91 (0.49)	4.30 (0.51)	0.39 (10.0)
Tympanometer (9 factors)	3.83 (0.21)	4.67 (0.42)	0.84 (21.9)	4.24 (0.47)	4.75 (0.38)	0.51 (12.0)
Hearing screening (9 factors)	4.56 (0.30)	4.66 (0.41)	−0.10 (−2.2)	4.21 (0.61)	4.67 (0.45)	0.46 (10.9)

Qualitative Questionnaire Results

The results of the 4 free-response questions that were administered to the test group regarding their impression of the simulator tool are presented in Table 4. We categorized the responses based on their context similarity (type) and frequency of being mentioned.

The responses to the first question indicate that the students found that the hands-on experience gained by using the simulator provided an easy-to-use and useful learning experience with the audiometer that increased their confidence in their ability to perform hearing screening.

With regard to the second question, the simulator was designed to display an image or chart for a simulated subject, rather than demonstrate how to perform an otoscopic or tympanometric examination. Students wanted more explanation on how to perform these examinations. Others felt that the instructions, in general, could be improved, possibly incorporating pictures or a video. One participant expressed a concern that the simulator design did not resemble an audiometer machine. Given that the simulator was modeled to closely resemble the Grason-Stadler GSI39 combination audiometer and tympanometer device, it is assumed that the student was familiar with a different model.

The most frequent response to the third question was that the simulator should improve training on how to perform otoscopy and tympanometry screenings. It was also suggested that the simulator model the instructions provided by the technician to the subject. Future software iterations will address these suggestions.

The responses to the final question suggested positivity toward the simulation tool, indicating its potential benefit in terms of providing a range of realistic learning experiences. The responses do suggest that some features of the simulator need to be explained further in the system documentation. The final 3 comments emphasized that the simulator was easy to use and was perceived as a useful addition to the class.

XSL•FO
RenderX

Table 4. Poststudy survey of impressions of the hearing screening simulator (n=17).

	Frequency of being mentioned, n (%)
"What aspects of the simulation did you like?"	
Provides hands-on experience, helped me become more familiar with an audiometer, and increased my confidence	5 (29)
Ease of use	4 (23)
Practice with multiple virtual patients	4 (23)
Provided useful feedback on the simulation debriefing report	1 (6)
Thought it was helpful	1 (6)
"What aspects of the simulator did you not like?"	
Did not show how to do otoscopic or tympanometry exam	4 (23)
Having better instructions would be helpful	4 (23)
Nothing, this is a useful simulation	2 (12)
It was repetitive	1 (6)
"Do you have any recommendations to improve the simulator?"	
The simulator did not show how to perform otoscopy and tympanometry. Clearer instructions are needed in general	5 (29)
Nothing, no improvement needed	3 (18)
Model the technician giving required directions to the subject	1 (6)
Have the design of the simulator match that of the audiometer	1 (6)
"Would you like to share anything else?"	
A great tool to add to any audiology class	1 (6)
Really liked that sample participants showed inconsistencies in hearing thresholds. This realistic touch prepares one to be not thrown off when performing a live, in-person screening	1 (6)
The simulation was easy to use	1 (6)
The virtual simulation was very helpful	1 (6)
I loved the activity!	1 (6)

Instructor Evaluation of Student Hearing Screening Assessment Technique

Certified instructors supervised the assessment, observing and grading the students' performance and technique. The summary data from the instructor's assessment is given in Table 5. Assessment data for two students in the control group were not available which reduced the size of the control group to 14. Data for one student in the test group was dropped because of incomplete data resulting in a group size of 16. All students in both the control and test groups performed the 3 screenings satisfactorily (ie, otoscopy, tympanometry, and hearing screening). Some students, primarily in the control group, did not follow the prescribed protocol exactly, as reflected in the technique scores shown in Table 5. Students performed similarly across test and control groups, except in the case of providing instructions and screening according to protocol.

Table 5. Certified technicians' assessment of study participants' techniques in performing otoscopic, tympanometric, and audiometric testing on live subjects using commercial audiometer equipment.

	Control group (n=14), mean (SD)	Test group (n=16), mean (SD)
Otoscope	4.71 (0.47)	4.50 (0.52)
Tympanometer	4.64 (0.74)	4.38 (0.62)
Audiometer		
Provide instructions	3.64 (1.39)	4.63 (0.81)
Test according to protocol	4.64 (0.50)	4.81 (0.40)

XSL•FO
RenderX

Discussion

Principal Findings

The HSS is designed as a training tool for speech-language pathologists, nurses, and other professionals to learn the skills needed to perform hearing screening. It can provide remote training when it is not practical to use a traditional practicum setting with commercial audiology equipment, and thus represents an improvement on the state of the art [26]. The design science framework described in this study could be used to replicate and extend the existing simulator and develop other simulation training tools; for example, this implementation is modeled on a specific commercial audiometer. It would be useful to expand this to an array of available audiometers and allow students to receive a flexible range of training on instruments of their choosing. This would be similar to how Microsoft Flight Simulator [29] allows users to choose from 30 different cockpits. During our testing, a respondent commented that the simulation interface did not represent the instrument they were using, highlighting the need to expand the tool to model other instruments. The user training interaction model created for this study incorporated artificial patients in a software simulation within a theoretical framework, operating in real-time and real-world environments. The tool uses external configuration files that enable the addition of new simulated subjects. Each simulated subject has a corresponding biographical sketch, auditory and tympanometric profiles, and otoscopic images. This design feature allows the addition of new simulated patients, covering a range of possible real patient conditions. The software simulation tool provides a design artifact that was tested and evaluated across 33 participants and thus provides a complete design science research cycle, inclusive of artifact evaluation.

This study focused on the user experience that influenced the design parameters of the HSS. In the evaluation, we surveyed users to assess usability, as well as efficacy as a teaching tool. While the mean scores of the test group were largely higher than those of the control group and demonstrated a larger percentage change from before to after the intervention, there were few statistically significant differences between the groups. This was not surprising, considering the very early phase of research in this pilot program and the known difficulties of measuring educational interventions, especially with small sample sizes. Furthermore, the percentage changes from before to after the intervention were mixed for those participants who had previously conducted a screening examination on a real person before participating in this training exercise and study. While the small sample size of this subgroup makes it difficult to draw firm conclusions in this regard, it is important to note that prior screening experience may impact hearing screening learning outcomes and the effect of the simulator on learning. Taken together, the quantitative and qualitative results from this endeavor show significant promise for audiology simulator tools in enhancing the education of speech-language pathology students. The results indicate that the working prototype of the simulator may provide a strong supplement to in-person training using commercially available equipment. Participants noted that the simulator was intuitive and useful for learning the skills needed to perform hearing screenings, while also indicating areas for improvement.

Limitations

The number of participants in this study was low (n=33), yet appropriate for an initial pilot test. Future studies should include a more diverse set of participants, possibly from multiple geographic locations. This represents the first test of the HSS simulation software. Future iterations of the software may demonstrate different results. A common issue in the control group was a failure of student evaluators to communicate instructions to the subjects as prescribed in the protocol. This issue was not as prevalent in the test group, which may indicate that the added simulator practice impacted the students' performance in following the procedure, although the simulator did not explicitly indicate or model this behavior. The observed issues were not modeled in the simulation and thus represent potential areas for future improvement.

The study design could be improved. In this iteration, the students in the test group received traditional lectures on the material and then completed the simulation before their performance and technique were evaluated. The control group received only the traditional lectures before evaluation. The difference in the observed results could therefore be due to the test group's greater exposure to the relevant content. In future testing, the control group could be given an alternative, related activity with a similar duration as the simulation. This could include additional reading or researching audiometer test equipment on the internet.

Future Directions

Our testing of the HSS found that it has the potential to be an effective audiometry teaching tool, but some areas for improvement and future research directions were identified to validate these findings. The most significant improvement would be to convert it to a web-based platform for broader availability and accessibility. While remote students can currently use the simulator, it requires them to download and install the tool locally. A web-based option would make using the simulator easier, with the only requirement being a web browser.

The functionality of the simulator could be extended by allowing the configuration to support customizable pass and refer cutoffs to address different environments or countries. Similarly, allowing the configuration to support multiple languages could be valuable. This could be implemented by modifying and displaying text in different user-specified languages or incorporating "tool tips" to offer second-language support in pop-up windows.

Another important extension would be the inclusion of different models of commercial audiometer test equipment. A key benefit of the tool is that it provides training using a digital twin, providing the learning experience with a test unit user interface. The value of this feature diminishes if the technician is not using the same equipment modeled in the tool. This issue was mentioned by a participant in the exit survey. Although the tool is designed for hearing screening test training, it could be extended to handle diagnostic hearing test training as well.

XSL•FO
RenderX

While the physical equipment remains the same, the testing protocols differ and would have to be modified.

Recommendations from students using the system included that a video tutorial should be created to either replace or supplement the current text-based instructions. The students also suggested adding a learning mode, where context-sensitive feedback could be displayed on command during the training phase but hidden during the evaluation phase. As a result of the testing, it was determined that the simulator did not verify the students' compliance with the required auditory screening protocols. While the system counts the number of frequency screens completed, it does not track the testing of specified frequencies (1000, 2000, and 4000 Hz tones), repeated tests, the number of positive responses, or the sound levels used. Therefore, while the simulator did provide an assessment of the students' ability to determine the correct result, it did not fully capture the performance as specified in the audiology screening guidelines [4]. These factors were considered by the evaluator when grading the students during the evaluation phase of the exercise. These features are planned for future versions of the simulator.

To further access the effectiveness of the simulator tool, a series of studies would need to be conducted with a broader set of participants across a broader set of training programs (eg, universities) dispersed geographically. These studies would need to assess the educational outcomes associated with the use of the tool.

Conclusions

This work's contributions include a design framework for developing a tool that simulates an existing process, the description of a unique artifact that supports an individualized and self-paced learning environment incorporating context-sensitive feedback and performance assessment, and an extensible approach to supporting simulated subjects in clinical training testing artificial patients. The software development process, developed artifact, and user evaluation together provide a comprehensive, descriptive framework for designing audiology simulation training tools as well as validation that such tools may provide a cost-effective mechanism for training a wide range of providers for audiology screening.

Acknowledgments

The authors wish to acknowledge the assistance of Danielle Varnedoe, the director of the Montgomery Speech-Language-Hearing Clinic at the University of South Carolina, for her support in the development of the hearing screening simulator. The authors also thank Dr Gordon Schooley, PhD, for his consultation and assistance with the statistical analysis. The authors also acknowledge the work by the student capstone teams that worked on early versions of the tool: Austin Wicker; Elizabeth Saiz; Saniya Jumani; Davon Davis; Kristin Jackson; Oneil Clemenson, Jr; Brian Morrison; Christopher Bomar; Kasara Tyer; and Mary Holliday.

Disclaimer

An earlier version of this paper was presented at the 55th Hawaii International Conference on System Sciences [28]. This paper differs from the earlier conference paper. The discussion of the iterative design of the simulator tool was removed, with interested readers directed to the conference paper. This study changed statistical methodology from the Kruskal-Wallis test to the Mann-Whitney U test to more appropriately address the study's pretest-posttest design with ordinal data. Statistical reporting was expanded to include $P<.05$ and $P<.10$, both of which are often used for reporting in pilot studies. Additional mean change and percentage change analyses were conducted and reported on regarding the prior audiology testing experience of participants. A new table was added to present the additional analysis. Corrections were made throughout the paper, with significant reduction and rewriting of the introduction and literature review. The discussion section was expanded based on the expanded statistical analysis.

Data Availability

The datasets generated and analyzed during this study are available in the OSF repository "Hearing Screening Simulator Data" [30] under the Creative Commons Attribution 4.0 International Public License.

Authors' Contributions

JG was the lead application developer of the hearing screening simulator discussed in the paper and is the primary author. BS processed the statistics and analysis and assisted in the preparation of the manuscript. DS and JM tested and provided feedback on each iteration as it was developed. DS and JM provided domain-specific knowledge that aided in the development of the tool and contributed to the paper.

Conflicts of Interest

DS and JM work in the Montgomery Speech-Language-Hearing Clinic in the Department of Communication Sciences and Disorders at the University of South Carolina, which was a partner in the development of the hearing screening simulator. All other authors declare no other conflicts of interest.

XSL•FO
RenderX

References

1. Niskar AS, Kieszak SM, Holmes A, Esteban E, Rubin C, Brody DJ. Prevalence of hearing loss among children 6 to 19 years of age: the Third National Health and Nutrition Examination Survey. JAMA 1998 Apr 08;279(14):1071-1075. [doi: 10.1001/jama.279.14.1071] [Medline: 9546565]
2. Guralnick MJ. Why early intervention works: a systems perspective. Infants Young Child 2011 Jan 01;24(1):6-28 [FREE Full text] [doi: 10.1097/IYC.0b013e3182002cfe] [Medline: 21532932]
3. Alanazi AA. The use of simulation in audiology education to improve students' professional competency. ED.gov. 2017. URL: https://eric.ed.gov/?id=ED577397 [accessed 2024-04-29]
4. American Speech-Language-Hearing Association. Guidelines for Audiologic Screening. New York, NY: American Speech-Language-Hearing Association; 1997.
5. 2014 standards for the certificate of clinical competence in speech-language pathology. Council for Clinical Certification in Audiology and Speech-Language Pathology of the American Speech-Language-Hearing Association (ASHA). 2020. URL: https://www.asha.org/certification/2020-slp-certification-standards/ [accessed 2025-01-19]
6. Cant RP, Cooper SJ. Simulation-based learning in nurse education: systematic review. J Adv Nurs 2010 Jan;66(1):3-15. [doi: 10.1111/j.1365-2648.2009.05240.x] [Medline: 20423432]
7. Guise V, Chambers M, Välimäki M. What can virtual patient simulation offer mental health nursing education? J Psychiatr Ment Health Nurs 2012 Jun;19(5):410-418. [doi: 10.1111/j.1365-2850.2011.01797.x] [Medline: 22070549]
8. La Cerra C, Dante A, Caponnetto V, Franconi I, Gaxhja E, Petrucci C, et al. Effects of high-fidelity simulation based on life-threatening clinical condition scenarios on learning outcomes of undergraduate and postgraduate nursing students: a systematic review and meta-analysis. BMJ Open 2019 Feb 22;9(2):e025306 [FREE Full text] [doi: 10.1136/bmjopen-2018-025306] [Medline: 30798316]
9. Sherwood RJ, Francis G. The effect of mannequin fidelity on the achievement of learning outcomes for nursing, midwifery and allied healthcare practitioners: systematic review and meta-analysis. Nurse Educ Today 2018 Oct;69:81-94. [doi: 10.1016/j.nedt.2018.06.025] [Medline: 30015220]
10. Brown DK. Simulation before clinical practice: the educational advantages: audiology today. American Academy of Audiology. 2017. URL: https://www.audiology.org/news-and-publications/audiology-today/articles/simulation-before-clinical-practice-the-educational-advantages/ [accessed 2024-04-29]
11. Bakhos D, Galvin J, Aoustin JM, Robier M, Kerneis S, Bechet G, et al. Training outcomes for audiology students using virtual reality or traditional training methods. PLoS One 2020;15(12):e0243380 [FREE Full text] [doi: 10.1371/journal.pone.0243380] [Medline: 33270806]
12. Alanazi AA, Nicholson N. Audiology and speech-language pathology simulation training on the 1-3-6 early hearing detection and intervention timeline. Am J Audiol 2019 Jun 10;28(2):348-361. [doi: 10.1044/2019_AJA-18-0185] [Medline: 31095403]
13. Andre A, Reed A, Ananthakrishnan S, Korczak P. An evaluation of simulation techniques in audiology and allied health professions. Am J Audiol 2021 Jun 14;30(2):295-308. [doi: 10.1044/2021_AJA-20-00190] [Medline: 33872515]
14. Alanazi AA, Mohamud MS, AlSuwailem SS. The effect of simulation learning on audiology and speech-language pathology students' self-confidence related to early hearing detection and intervention: a randomized experiment. Speech Lang Hear 2020 Nov 17;25(2):211-224. [doi: 10.1080/2050571X.2020.1846839]
15. Ozelie R, Panfil P, Swiderski N, Walz E. Hearing voices simulation: impact on occupational therapy students. Open J Occup Ther 2018 Oct 01;6(4):10. [doi: 10.15453/2168-6408.1452]
16. Toader E. Clinical simulations for learning medical skills: a work-based approach to simulators. Procedia Soc Behav Sci 2015 Jul;197:2443-2448. [doi: 10.1016/j.sbspro.2015.07.408]
17. Halm BM, Lee MT, Franke AA. Improving toxicology knowledge in preclinical medical students using high-fidelity patient simulators. Hawaii Med J 2011 Jun;70(6):112-115 [FREE Full text] [Medline: 22162607]
18. Otis – the virtual patient. Innoforce. URL: https://www.innoforce.com/en/products/otis-the-virtual-patient/ [accessed 2024-04-29]
19. Kocian A, Cattani G, Chessa S, Grolman W. An artificial patient for pure-tone audiometry. J Audio Speech Music Proc 2018 Jul 27;2018(1):1. [doi: 10.1186/S13636-018-0131-Y]
20. Calandruccio L, Weidman D. Online simulation education for audiometry training. Am J Audiol 2022 Mar 03;31(1):1-10. [doi: 10.1044/2021_AJA-21-00121] [Medline: 34958736]
21. Kompis M, Steffen P, Caversaccio M, Brugger U, Oesch I. A multilingual audiometer simulator software for training purposes. Acta Otolaryngol 2012 Apr;132(4):428-433. [doi: 10.3109/00016489.2011.636379] [Medline: 22235845]
22. Lieberth AK, Martin DR. The instructional effectiveness of a web-based audiometry simulator. J Am Acad Audiol 2005 Feb;16(2):79-84. [doi: 10.3766/jaaa.16.2.3] [Medline: 15807047]
23. Wilding T. Pure tone audiometry simulator. The University of Manchester. URL: https://personalpages.manchester.ac.uk/staff/tim.wilding/PTA_Sim/index.html [accessed 2024-04-29]
24. Audiometer simulator for pure-tone audiometry including masking. AudSim Flex. URL: https://audsim.com/) [accessed 2024-04-29]
25. Combination audiometer and tympanometry. Grason-Stadler Inc. URL: https://www.grason-stadler.com/products/tympanometers/gsi-39 [accessed 2024-04-29]

XSL•FO
RenderX

26. Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S. A design science research methodology for information systems research. J Manag Inf Syst 2014 Dec 08;24(3):45-77. [doi: 10.2753/MIS0742-1222240302]
27. Drechsler A, Hevner AR. Utilizing, producing, and contributing design knowledge in DSR projects. In: Proceedings of the 13th International Conference on Designing for a Digital and Globalized World. 2018 Presented at: DESRIST '18; June 3-6, 2018; Chennai, India p. 82-97 URL: https://link.springer.com/chapter/10.1007/978-3-319-91800-6_6 [doi: 10.26686/wgtn.13088369]
28. Gerdes J, Schooley B, Sharp D. Design of a simulation tool for audiology education to provide hearing screening training. In: Proceedings of the 55th Hawaii International Conference on System Sciences. 2022 Presented at: HICSS '22; January 4-7, 2022; Virtual Event p. 3793-3802 URL: https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/9c77643b-9269-4cc4-ad3c-7b085a841fcd/content [doi: 10.24251/hicss.2022.462]
29. Microsoft flight simulator, 40th anniversary edition. Flight Simulator. URL: https://www.flightsimulator.com/ [accessed 2024-04-29]
30. Gerdes J, Schooley B. Hearing screening simulator data. Open Science Framework. URL: https://osf.io/8s6rj/ [accessed 2024-04-29]

Abbreviations

HSS: hearing screening simulator

Edited by A Mavragani; submitted 10.03.23; peer-reviewed by L Coco, N Mungoli; comments to author 15.05.23; revised version received 28.07.23; accepted 03.09.24; published 20.02.25.

Please cite as:
Gerdes J, Schooley B, Sharp D, Miller J
The Design and Evaluation of a Simulation Tool for Audiology Screening Education: Design Science Approach
JMIR Form Res 2025;9:e47150
URL: https://formative.jmir.org/2025/1/e47150
doi:10.2196/47150
PMID:39977027

©John Gerdes, Benjamin Schooley, Dakota Sharp, Juliana Miller. Originally published in JMIR Formative Research (https://formative.jmir.org), 20.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effects of Missing Data on Heart Rate Variability Measured From A Smartwatch: Exploratory Observational Study

Hope Davis-Wilson[1], PhD; Meghan Hegarty-Craver[1], PhD; Pooja Gaur[1], PhD; Matthew Boyce[1], BS; Jonathan R Holt[1], PhD; Edward Preble[1], PhD; Randall Eckhoff[1], BS; Lei Li[1], PhD; Howard Walls[1], PhD; David Dausch[1], PhD; Dorota Temple[1], PhD

RTI International, Morrisville, NC, United States

Corresponding Author:
Hope Davis-Wilson, PhD
RTI International
3040 E Cornwallis Rd
Morrisville, NC, 27709
United States
Phone: 1 7043404816
Email: hdaviswilson@rti.org

Abstract

Background: Measuring heart rate variability (HRV) through wearable photoplethysmography sensors from smartwatches is gaining popularity for monitoring many health conditions. However, missing data caused by insufficient wear compliance or signal quality can degrade the performance of health metrics or algorithm calculations. Research is needed on how to best account for missing data and to assess the accuracy of metrics derived from photoplethysmography sensors.

Objective: This study aimed to evaluate the influence of missing data on HRV metrics collected from smartwatches both at rest and during activity in real-world settings and to evaluate HRV agreement and consistency between wearable photoplethysmography and gold-standard wearable electrocardiogram (ECG) sensors in real-world settings.

Methods: Healthy participants were outfitted with a smartwatch with a photoplethysmography sensor that collected high-resolution interbeat interval (IBI) data to wear continuously (day and night) for up to 6 months. New datasets were created with various amounts of missing data and then compared with the original (reference) datasets. 5-minute windows of each HRV metric (median IBI, SD of IBI values [STDRR], root-mean-square of the difference in successive IBI values [RMSDRR], low-frequency [LF] power, high-frequency [HF] power, and the ratio of LF to HF power [LF/HF]) were compared between the reference and the missing datasets (10%, 20%, 35%, and 60% missing data). HRV metrics calculated from the photoplethysmography sensor were compared with HRV metrics calculated from a chest-worn ECG sensor.

Results: At rest, median IBI remained stable until at least 60% of data degradation (P=.24), STDRR remained stable until at least 35% of data degradation (P=.02), and RMSDRR remained stable until at least 35% data degradation (P=.001). During the activity, STDRR remained stable until 20% data degradation (P=.02) while median IBI (P=.01) and RMSDRR P<.001) were unstable at 10% data degradation. LF (rest: P<.001; activity: P<.001), HF (rest: P<.001, activity: P<.001), and LF/HF (rest: P<.001, activity: P<.001) were unstable at 10% data degradation during rest and activity. Median IBI values calculated from photoplethysmography sensors had a moderate agreement (intraclass correlation coefficient [ICC]=0.585) and consistency (ICC=0.589) and LF had moderate consistency (ICC=0.545) with ECG sensors. Other HRV metrics demonstrated poor agreement (ICC=0.071-0.472).

Conclusions: This study describes a methodology for the extraction of HRV metrics from photoplethysmography sensor data that resulted in stable and valid metrics while using the least amount of available data. While smartwatches containing photoplethysmography sensors are valuable for remote monitoring of patients, future work is needed to identify best practices for using these sensors to evaluate HRV in medical settings.

(JMIR Form Res 2025;9:e53645) doi:10.2196/53645

XSL•FO
RenderX

KEYWORDS

plethysmography; electrocardiogram; missing data; smartwatch; wearable; ECG; photoplethysmography; PPG; mobile phone; heart rate; pilot study; photoplethysmography; detection; sensor; monitoring; health metric; measure; electrocardiogram; real-world settings; rest; physical activity; remote monitoring; medical setting; youth; adolescent; teen; teenager

Introduction

Remote patient monitoring using digitally transmitted health data through wearable physiological and activity sensors offers many benefits and is gaining acceptance in the medical community [1]. For example, applications of remote monitoring include acute conditions such as COVID-19 recovery [2] along with chronic conditions such as heart failure [3], chronic obstructive pulmonary disease [4], and diabetes [5]. Remote patient monitoring via wearable sensors can improve clinical outcomes, improve self-management of diseases, and increase patient engagement and satisfaction [6]. Many commercial off-the-shelf (COTS) wearable sensors provide data on important physiological information, particularly wrist-worn devices used to measure heart rate [7,8]. Wrist-worn devices most often use embedded photoplethysmography sensors that detect changes in light intensity on the skin surface due to changes in blood volume during the cardiac cycle to estimate heart rate [9]. While not all smartwatches are certified as medical devices currently, they offer a way of assessing the viability of photoplethysmography sensors, which are used in some cleared medical devices [9-11].

Heart rate variability (HRV) is an indicator of autonomic nervous system functioning and can be used to identify health conditions such as respiratory illness [12], multiple sclerosis [13], Parkinson disease [14], and traumatic brain injury [15]. While collecting wearable sensor data such as HRV can provide multiple benefits to patients and create a more holistic picture of health status, real-world collection of HRV can be difficult to assess due to missing data, differences in data processing techniques, and the quality of data collected [16]. Missing data due to compliance and low-quality signals is an issue when using wearable sensors to monitor HRV in patients in clinical or research settings [16]. Missing data is common in longitudinal research studies, and it is problematic because it can reduce statistical power and cause bias in the estimation of parameters [17]. This is of concern when attempting to identify health anomalies because identifying a change in health relies on consistent and accurate baseline HRV data [18]. As demand for remote patient monitoring grows, an important next step is to identify best practices for data acquisition, processing, and analysis of HRV data from wearable sensors.

In addition to the need for best practices regarding data missingness, more understanding is still needed for accuracy in wrist-worn photoplethysmography sensors compared with more accurate wearable devices. Wrist-worn sensors typically result in greater wear time compared with chest-worn devices; however, wrist-worn sensors are more prone to movement-related artifacts that can negatively influence HRV signal quality [6]. Overwhelmingly, real-world studies have opted to use wrist-worn photoplethysmography sensors to evaluate HRV over other more accurate methods such as chest-worn electrocardiogram (ECG) sensors, likely improving patient compliance in these studies. As a result, we must determine how to best assess data from wrist-worn photoplethysmography sensors, including accounting for missing data and assessing data validity compared with ECG HRV data. This study conducted long-term (months) continuous monitoring of high-resolution HRV data during daily life activities using a COTS smartwatch with a photoplethysmography sensor and activity (step count) data. The primary aim of this study was to evaluate the influence of missing data on HRV metrics collected from a photoplethysmography-based smartwatch both at rest and during physical activity in real-world settings. The secondary aim of this study was to evaluate HRV agreement and consistency between the wrist-worn photoplethysmography sensor and a chest-worn ECG heart rate monitor both at rest and during periods of physical activity and sleep.

Methods

Ethical Considerations

For this prospective observational study, wearable sensor data was collected from local community members. The protocol was approved by the Research Triangle Institute (RTI) international institutional review board (MOD00001413). Written consent was received from all participants for participation and data use. The consent process emphasized that participation was voluntary and that individuals were allowed to withdraw their consent for participation at any time. Personal identifying information was stored in a secured file and kept separate from data collected once enrolled in the study. All smartwatch and survey data were deidentified and labeled by a participant identifier for analysis. Participants were offered US $50 every 3 months as study compensation, and participants could earn up to US $100 if they were enrolled for the entire 6 months of the study duration.

Participants

Healthy volunteers aged 18 years and older were included in this study. Individuals who were pregnant or who had a pre-existing health condition were not excluded from participation if they were able to use the wearable sensor. Participants were enrolled for 3 months with the opportunity to re-enroll for up to 6 months. Demographic data were obtained and stored in Research Electronic Data Capture (REDCap). Recruitment and data collection took place between May 2022 and June 2023.

Device Protocol

Each participant was outfitted with a Garmin Fenix 6/6s smartwatch, which was worn on the nondominant wrist. We chose to use Garmin over other wrist-worn photoplethysmography sensors because its Health Companion Software Development Kit provides access to high-resolution interbeat interval (IBI) data collected by photoplethysmography sensors that allow for monitoring of HRV. Other COTS sensors

XSL•FO
RenderX

only provide summary and other processed data. The average heart rate derived from Garmin watches strongly correlates with clinical ECG heart rate data (r=0.96) [19]; however, the accuracy of HRV metrics derived from Garmin has yet to be studied. Garmin watches integrate 2 photoplethysmography sensors for monitoring HR and a 3-axis accelerometer for detecting movement in the x, y, and z directions. The watch firmware uses accelerometer signals to calculate step counts in 1-minute windows. We used the Companion Software Development Kit to access minimally processed IBI data and step counts from the participants. The SIGMA+ Health app (S+ app) was installed onto the participant's personal mobile device or a study-issued Samsung Galaxy 12 smartphone if participants preferred not to use their personal mobile device for the study. The Garmin smartwatch was paired through a Bluetooth low energy connection to the mobile device and data from the watch was ingested through the S+ app onto an RTI secure Amazon Web Services server in JavaScript Object Notation format for offline data cleaning, metric extraction and standardization, and analysis. The initial setup included giving each user the study Personal identification number and a unique participant ID that did not contain personally identifiable information, which allowed the S+ app to track the user's data from one or more devices through the duration of the study. Within the app, the participant could view the connectivity status of their Garmin Fenix watch and manually trigger a data sync if required. Participants completed a daily health survey administered within the S+ app, which asked questions related to illness symptoms, sleep quality, sleep duration, and any unusual events that may have occurred that day (ie, psychological stress, excessive exercise, etc). Results were stored in a REDCap database.

A subset of participants (n=5) wore a Polar H10 ECG heart rate monitor with a sampling frequency of 1000 Hz over their chest simultaneously with the Garmin smartwatch to assess agreement and consistency between HRV metrics. IBI values derived from the Polar H10 chest strap show strong agreement compared with clinical ECG recordings at rest (intraclass correlation coefficient [ICC]=0.95) and with light exercise (ICC=0.93) [20]. Participants were encouraged to wear the Polar H10 continuously (day and night) for 5 days and were instructed to only remove it during showers, baths, or swimming. Polar H10 data was retrieved through the Elite HRV mobile app, which was downloaded to the participant's personal smartphone before data collection [21]. Excel files containing raw IBI data were exported from the participant's phone and sent to an RTI email account. We did not get measures of step counts from the Polar H10 device because Elite HRV only collects HRV data.

Data Processing for IBI Data From Garmin and Polar Devices

Each IBI data point acquired through the Garmin smartwatch and S+ app had an associated timestamp with millisecond precision. Steps from the Garmin were reported at 60-second intervals and IBI values were reported with each detected beat. Each IBI data point acquired through the Polar H10 monitor and Elite HRV app was recorded with millisecond precision. IBI artifacts (IBI>1500 ms or IBI <300 ms) were flagged and removed. For Garmin data quality control, we evaluated the expected data fraction, calculated as a ratio of the number of incoming data points to the maximum number of data points expected during the monitoring period, to assess data missingness. For simplicity, we assumed the expected number of IBI data points would be 60 IBI values/minute. This would correlate with an average heart rate of 60 beats per minute, which is a typical heart rate for someone who is sitting or resting. The exact number of data points expected varied due to within- and between-person differences in average heart rate. The expected data fraction was multiplied by 100 to be represented as a percentage between 0 and 100. For analysis, we chose to evaluate Garmin datasets with at least 70% of expected data to ensure we were using high-quality datasets with minimal missing data points. The first approximately 2 weeks of data were selected for each participant with at least 70% expected data fraction. Although there is some variation in the number of IBI data points in each 5 min window (a higher heart rate results in more IBI data points within a window), the datasets with at least 70% of expected data that were used for analysis contained a mean of 250 (SD 126) IBI data points within each 5 min window.

Garmin and Polar data were processed into 5-minute windows with each point in the window formatted as (timestamp, value) and aligned based on real-time so that the windows start at 5-minute marks (eg, 12:00 AM, 12:05 AM). The following metrics were extracted in 5 min windows: time-domain metrics (median IBI, SD of IBI values [STDRR], root-mean-square of the difference in successive IBI values [RMSDRR[) and frequency-domain metrics (low-frequency [LF] power, high-frequency [HF] power, and ratio of LF to HF power [LF/HF]) [18,22]. To calculate HRV metrics in the frequency domain, we resampled IBI data to 5 Hz using the Piecewise Cubic Hermite Interpolating Polynomial method [23]. If there was a gap in the signal greater than 15s, the longest continuous segment in the 5-minute window was used. If there were more than 3 gaps of this length, we did not use this window. After interpolation, the signal was filtered to the low-frequency (LF: 004 to 0.15 Hz) and high-frequency regions (HF: 0.15 to 1.0 Hz; we increased the upper limit to 1.0 Hz vs the typically used 0.40 Hz to account for physical activity when breathing rates and heart rates increase [24,25]). We computed a value representing the LF and HF power for 60-second and 30-second windows, respectively, using the formula: value = ln (variance [data window]). Values less than 2.5 and more than 9.0 were eliminated, and the median values were calculated to represent the 5-minute window. For Garmin smartwatch step data, reported every minute, we calculated the average of values in the 5-minute window; we required at least 2 valid data points for the calculation.

Semisimulation of Garmin Data Degradation

A semisimulation design based on previous work [26] was used to create new datasets with various amounts of Garmin IBI data from original Garmin datasets. The premise of this semisimulation approach was to create new datasets with different amounts of missing data based on real wear patterns. Using this method, we ensured that we removed data in a realistic manner rather than randomly removing data. In this approach, we considered data missingness patterns from datasets with low expected data fractions and applied these patterns to

XSL•FO
RenderX

the datasets with high expected data fractions. These semisimulated datasets were then compared to the reference datasets to understand how varying amounts of data missingness influence HRV metrics. In total, 4 additional 5-minute window samples were randomly selected that contained fewer IBI data points. These 4 samples were inspected to identify data gap patterns and then used to model data missingness patterns in the selected datasets.

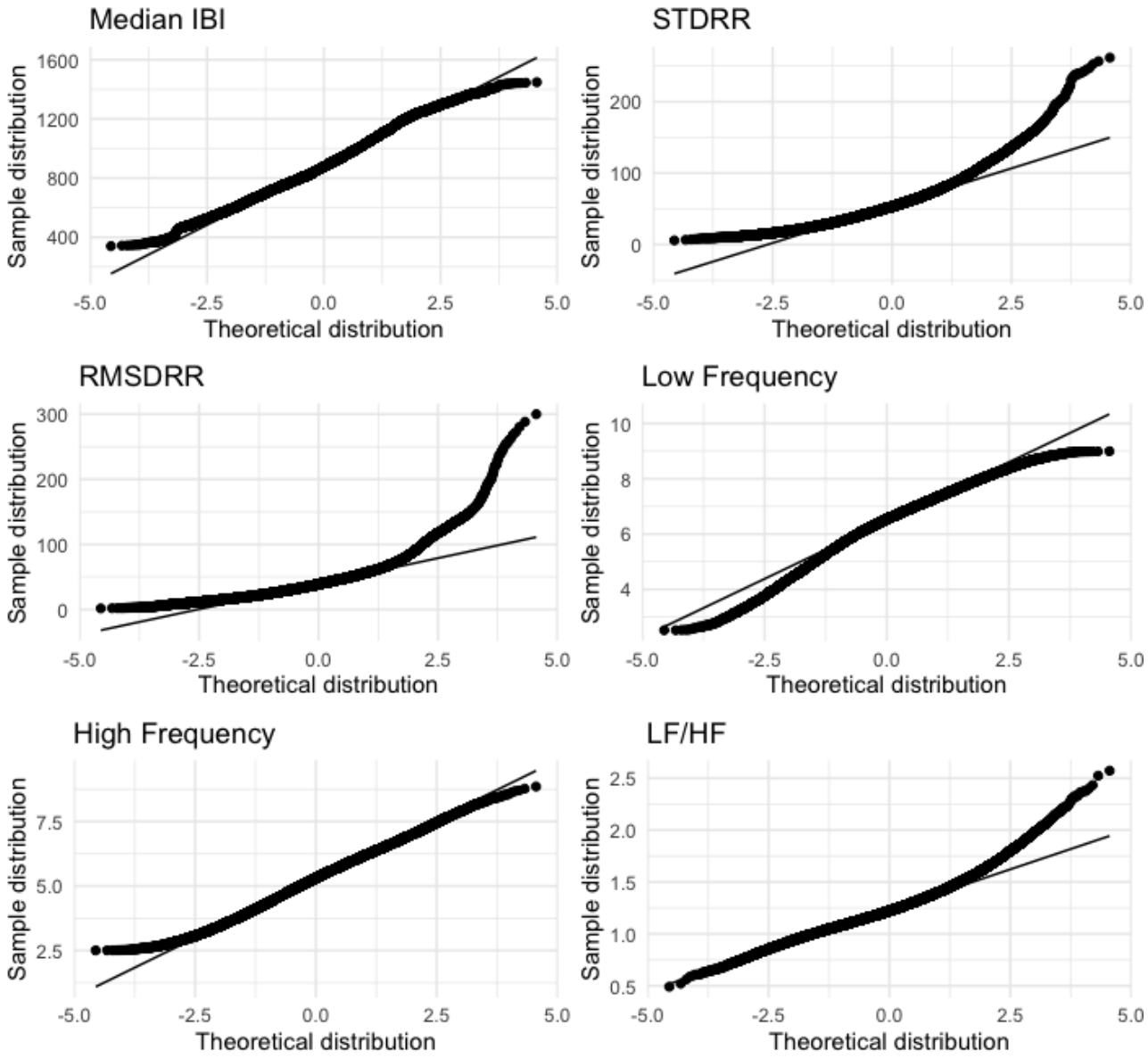
To recreate data missingness, we matched the data missingness pattern from the sample 5-minute windows to each 5-minute window in the reference datasets used for analysis. For example, one sample dataset with 120 IBI data points (missing 60% of data) was missing data in the first 3.15 min of the window. The first 3.15 min of each 5-min window was removed from the reference dataset to simulate data missingness. The simulated datasets were then compared with the reference datasets to identify differences in HRV metrics. Missing data simulations were performed using Python (Python Software Foundation) [27]. The various datasets were labeled for analysis, that consists of (1) a reference dataset with no missing data, (2) missing 10% of data, (3) missing 20% of data, (4) missing 35% of data, and (5) missing 60% of data. No simulated data degradation was performed for Polar H10 IBI data.

Statistical Analysis

Mean and SD were calculated for all descriptive variables and HRV metrics included in the analysis. For the primary aim, due to the large sample size of Garmin 5-minute windows for rest (48,442 windows) and light activity (19,976 windows), Shapiro-Wilk normality tests were not valid; therefore, we visually inspected each Garmin HRV metric for normality using Quantile-Quantile (Q-Q) plots. Q-Q plots that appeared as a roughly straight 45-degree line were considered normally distributed. Variables that were not normally distributed were transformed using a Box-Cox transformation before analysis. Q-Q plots of each HRV variable are shown at rest in Figure 1 and at light activity in Figure 2.

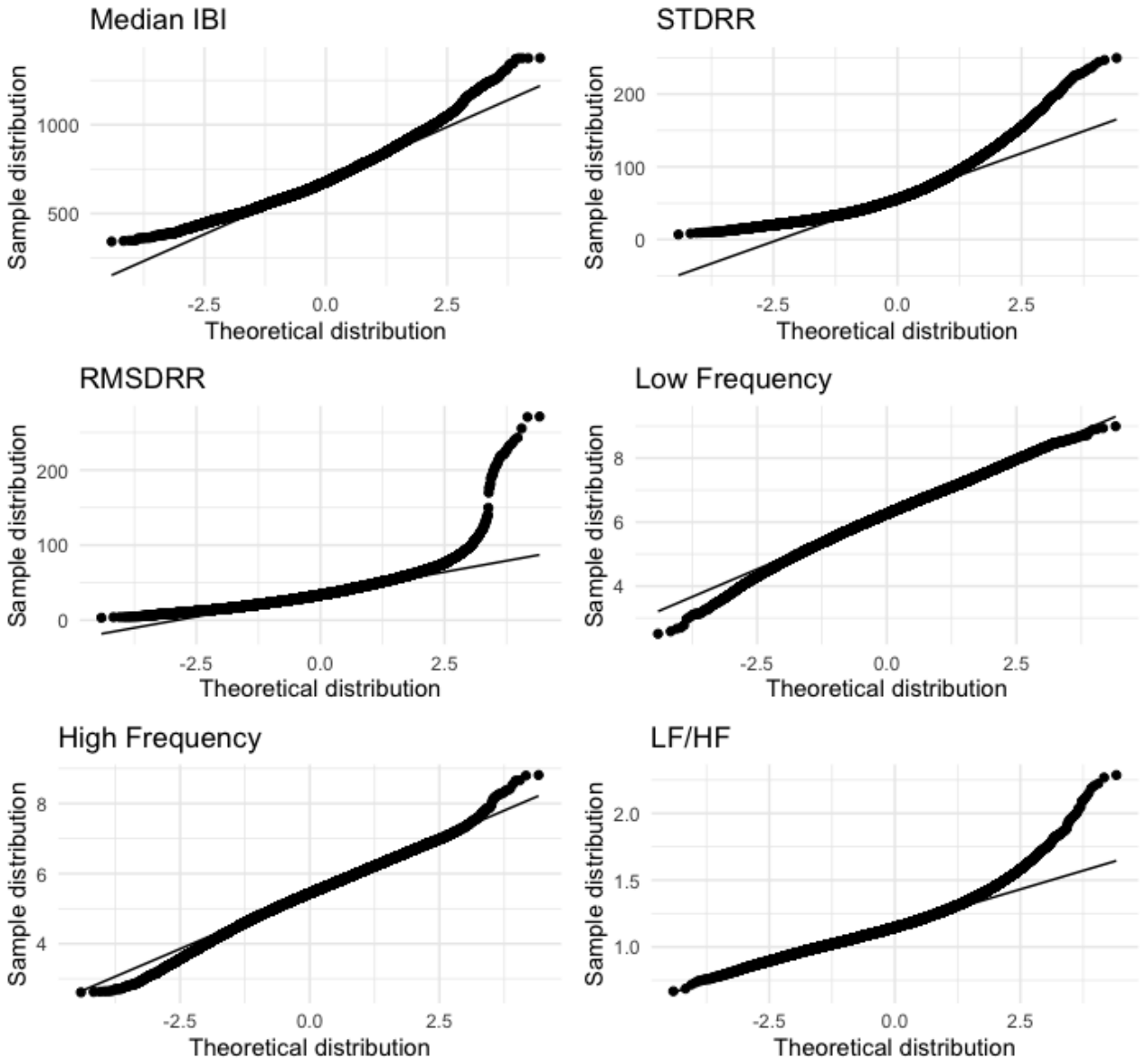
XSL•FO
RenderX

Figure 1. Quantile-Quantile plots for the median interbeat interval (IBI), SD of IBI values (STDRR), root-mean-square error of IBI values (RMSDRR), low frequency (LF), high frequency (HF), and the ratio of low to high frequency power at rest (steps/min=0). Box-Cox transformations were performed for STDRR and RMSDRR metrics. Smartwatch IBI data was collected and reported from 16 individuals recruited from a community setting between May 2022 and June 2023.



XSL•FO
RenderX

Figure 2. Quantile-Quantile plots for the median interbeat interval (IBI), SD of IBI values (STDRR), root-mean-square error of IBI values (RMSDRR), low frequency (LF), high frequency (HF), and the ratio of low to high frequency power at light activity (100>steps/min>0). Box-Cox transformations were performed for RMSDRR and LF/HF metrics. Smartwatch IBI data were collected and reported from 16 individuals recruited from a community setting between May 2022 and June 2023.



We then used linear mixed-effect models to estimate the mean differences in each Garmin HRV metric between the datasets with missing samples by level of missingness and the reference dataset. A participant-specific random intercept was included in the models to account for correlations of our sample data among multiple 5-minute windows within each participant [28]. An indicator of the level of data missingness (data without missing samples as reference, data with 10%, 20%, 35%, or 60% samples removed) was included in the models as a fixed-effect covariate. The Ime4 package in the statistical environment R (R Core Team) was used for the model estimations [29]. The statistical significance of the estimated differences was tested using Satterthwaite's method with a 2-tailed significance level of 0.05, and the 95% CI were obtained using the bootstrap method through the bootMer function of the Ime4 package in R (1000 simulations). Separate linear mixed models were estimated using sample data under 2 different conditions: at rest (steps/min=0) and during light physical activity (steps/min of >0 and <100).

For the secondary aim, we assessed the reliability of Garmin-derived HRV metrics within each participant and their validity compared to Polar-derived HRV metrics. The intraclass correlation coefficients (class 2, mean rating; $ICC_{2,k}$) for consistency and absolute agreement with 95% CI were obtained using the irr package in the statistical environment R [30]. The magnitude of the ICC was generally interpreted as $ICC_{2,k}$ of 0.5-0.75 as "moderate", $ICC_{2,k}$ >0.75 as "good", and $ICC_{2,k}$ >0.9 as "excellent" [31]. In addition, we visually assessed agreement between Garmin-derived HRV metrics and polar-derived HRV metrics using Bland-Altman analyses to evaluate the mean differences and 95% limits of agreement [32]. R version 4.2.2 [33] was used for statistical analyses.

XSL•FO
RenderX

Results

Overview

Garmin data were collected from 31 participants for up to 180 days. The analyzed sample in aim 1 used a mean of 16 (SD 3) consecutive days of data from 16 participants resulting in 48,442 five-minute windows of data that represented rest (steps/min=0) and 19,976 windows of data that represented light activity (100 >steps/min >0). Although there is not a clear precedence for what percentage of missing data is acceptable, we wanted the reference datasets to contain a minimal amount of missing data. Therefore, each participant included in the analysis had an expected data fraction of at least 70% (including daytime and nighttime), and participants who did not meet this requirement (n=15) were not included in the analysis. Differences in days of data occurred because of between-person variance in unprocessed IBI data and differences in wear patterns. For aim 2, 5 participants wore both the Garmin Fenix watch and Polar heart rate monitor for 5 consecutive days. Demographic data is shown in Table 1. The averages of each HRV metric for aims 1 and 2 are shown in Table 2 and Table 3. Box-Cox transformations were performed for STDRR and RMSDRR metrics at rest (steps/min=0; Figure 1) and RMSDRR and LF/HF at light activity (100>steps/min>0; Figure 2). Tables 4 and 5 show results from the linear mixed models, specifically the estimated error between the reference and missing data, the SE, and the 95% CIs of the estimated difference.

Table 1. Participant characteristics of a community sample of individuals who provided smartwatch data between May 2022 and June 2023. In total, 31 individuals were recruited, data from 16 individuals were used in aim 1, and data from 5 individuals were used in aim 2.

	Total sample (n=31), mean (SD)	Aim 1 (n=16), mean (SD)	Aim 2 (n=5), mean (SD)
Age (years)	38.6 (10.7)	38.6 (13.3)	35.2 (4.1)
Sex (females), n (%)	16 (53)	13 (81)	3 (60)
Height (m)	1.73 (0.12)	1.77 (0.12)	1.77 (0.08)
Mass (kg)	83.4 (17.2)	87.1 (16.5)	95.4 (24.7)
BMI (kg/m^2)	27.9 (4.7)	27.7 (4.1)	30.2 (5.6)

Table 2. Heart rate variability metrics for aim 1 datasets in a community sample of 16 individuals who provided smartwatch data between May 2022 and June 2023.

	Reference Garmin data	10% Missing Garmin data	20% Missing Garmin data	35% Missing Garmin data	60% Missing Garmin data
Resting (steps/min=0)					
Median interbeat interval	888.86	888.17	887.88	888.45	887.85
SD of interbeat interval values	57.41	57.44	57.44	57.14	52.87
Root-mean-square of the difference in successive interbeat interval values	41.70	41.83	41.86	41.53	40.98
Low-frequency power	6.52	6.38	6.31	6.50	6.50
High-frequency power	5.51	5.01	4.83	5.50	5.51
Ratio of low-frequency to high-frequency power	1.19	1.29	1.32	1.20	1.20
Light activity (0<steps/min<100)					
Median interbeat interval	690.70	688.02	687.19	692.65	695.30
SD of interbeat interval values	63.90	63.55	63.39	63.96	50.31
Root-mean-square of the difference in successive interbeat interval values	35.20	35.72	35.87	35.11	34.57
Low-frequency power	6.28	6.22	6.19	6.26	6.26
High-frequency power	5.59	5.26	5.13	5.55	5.55
Ratio of low-frequency to high-frequency power	1.13	1.19	1.21	1.13	1.14

XSL•FO
RenderX

Table 3. Heart rate variability metrics for aim 2 datasets in a community sample of 5 individuals who provided smartwatch data between May 2022 and June 2023.

	Reference Garmin data	10% Missing Garmin data	20% Missing Garmin data	35% Missing Garmin data	60% Missing Garmin data
Resting (steps/min=0)					
Median interbeat interval	888.86	888.17	887.88	888.45	887.85
SD of interbeat interval values	57.41	57.44	57.44	57.14	52.87
Root-mean-square error	41.70	41.83	41.86	41.53	40.98
Low-frequency power	6.52	6.38	6.31	6.50	6.50
High-frequency power	5.51	5.01	4.83	5.50	5.51
Ratio of low-frequency to high-frequency power	1.19	1.29	1.32	1.20	1.20
Light activity (0<steps/min<100)					
Median interbeat interval	690.70	688.02	687.19	692.65	695.30
SD of interbeat interval values	63.90	63.55	63.39	63.96	50.31
Root-mean-square error	35.20	35.72	35.87	35.11	34.57
Low-frequency power	6.28	6.22	6.19	6.26	6.26
High-frequency power	5.59	5.26	5.13	5.55	5.55
Ratio of low-frequency to high-frequency power	1.13	1.19	1.21	1.13	1.14

XSL•FO
RenderX

Table 4. Linear mixed-effects model results of the effects of missing raw data on heart rate variability metrics collected at rest (steps/min=0) in a community sample of 16 individuals who provided smartwatch data between May 2022 and June 2023.

Heart rate variability metric and level of missing data compared with reference	Estimated difference between reference data and missing data	SE	Lower limit of the 95% CI	Upper limit of the 95% CI	*P* value
Median interbeat interval					
10%	–0.739	0.939	–2.643	1.139	.43
20%	–1.04	0.939	–3.019	0.978	.27
35%	–0.468	0.939	–2.341	1.374	.62
60%	–1.095	0.939	–2.880	0.928	.24
SD of interbeat interval values					
10%	–0.002	0.006	–0.014	0.010	.69
20%	–0.003	0.006	–0.015	0.008	.59
35%	–0.020	0.006	–0.031	–0.008	.001[a]
60%	–0.215	0.006	–0.228	–0.204	<.001[a]
Root-mean-square					
10%	0.005	0.004	–0.003	0.013	.27
20%	0.005	0.004	–0.003	0.014	.20
35%	-0.014	0.004	–0.021	–0.005	.001[a]
60%	-0.053	0.004	–0.061	–0.044	<.001[a]
Low-frequency power					
10%	-0.134	0.006	–0.145	–0.122	<.001[a]
20%	-0.204	0.006	–0.215	–0.193	<.001[a]
35%	-0.016	0.006	–0.028	–0.005	.005[a]
60%	-0.016	0.006	–0.028	–0.005	.004[a]
High-frequency power					
10%	-0.501	0.005	–0.512	–0.490	<.001[a]
20%	-0.678	0.005	–0.689	–0.667	<.001[a]
35%	-0.013	0.005	–0.024	–0.002	.02[a]
60%	-0.006	0.005	–0.016	0.005	.27
Ratio of low frequency to high frequency					
10%	0.096	0.001	0.094	0.098	<.001[a]
20%	0.130	0.001	0.128	0.132	<.001[a]
35%	0.005	0.001	0.003	0.007	<.001[a]
60%	0.004	0.001	0.002	0.006	<.001[a]

[a]*P* value <.05 indicates a significant difference compared with the reference dataset.

XSL•FO
RenderX

Table 5. Linear mixed-effects model results of the effects of missing raw data on heart rate variability metrics collected at light activity (100>steps/min>0) in a community sample of 16 individuals who provided smartwatch data between May 2022 and June 2023.

Heart rate variability metric and level of missing data compared with reference	Estimated difference between reference data and missing data	SE	Lower limit of the 95% CI	Upper limit of the 95% CI	*P* value
Median interbeat interval					
10%	–2.645	1.049	–4.638	–0.627	.01[a]
20%	–3.473	1.049	–5.454	–1.481	<.001[a]
35%	1.998	1.049	–0.051	3.958	.06
60%	4.631	1.050	2.513	6.577	<.001[a]
SD of interbeat interval values					
10%	–0.005	0.003	–0.011	0.002	.11
20%	–0.008	0.003	–0.014	–0.001	.02[a]
35%	–0.006	0.003	–0.012	0.001	.08
60%	–0.217	0.003	–0.223	–0.210	<.001[a]
Root-mean-square					
10%	0.039	0.007	0.026	0.052	<.001[a]
20%	0.050	0.007	0.037	0.064	<.001[a]
35%	–0.013	0.007	–0.025	–.001	.05
60%	–0.069	0.007	–0.081	–0.055	<.001[a]
Low-frequency power					
10%	–0.054	0.006	–0.067	–0.042	<.001[a]
20%	–0.087	0.006	–0.100	–0.075	<.001[a]
35%	–0.027	0.006	–0.040	–0.015	<.001[a]
60%	–0.019	0.006	–0.031	–0.006	.003[a]
High-frequency power					
10%	–0.324	0.006	–0.336	–0.313	<.001[a]
20%	–0.452	0.006	–0.464	–0.441	<.001[a]
35%	–0.033	0.006	–0.045	–0.021	<.001[a]
60%	–0.038	0.006	–0.050	–0.027	<.001[a]
Ratio of low frequency to high frequency					
10%	0.043	0.004	0.042	0.045	<.001[a]
20%	0.059	0.001	0.058	0.061	<.001[a]
35%	–0.001	0.001	–0.003	0.0004	0.12
60%	–0.001	0.001	–0.001	0.002	0.33

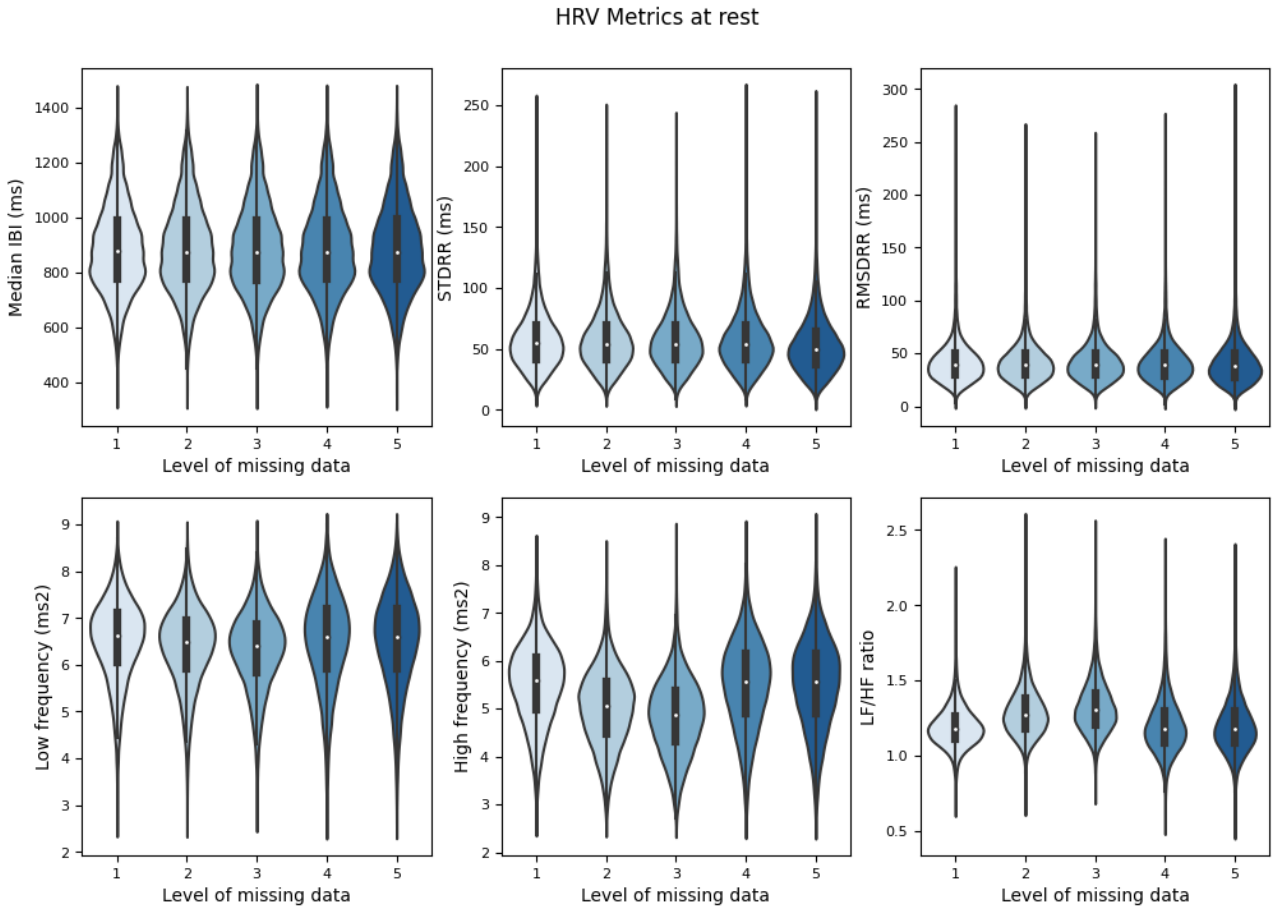
[a]*P* value <0.05 indicates a significant difference compared to the reference dataset.

Aim 1: Heart Rate Variability Metrics at Rest

At rest (steps/min=0), there was no difference in median IBI values between reference and simulated data (*P*>.05; Table 4 and Figure 3). There were significant differences in STDRR and RMSDRR values between the reference and 35% and 60% missing data (*P*<.05; Table 4, Figure 3). LF and LF/HF metrics demonstrated significant differences between the reference and all simulated datasets (*P*<.05; Table 4, Figure 3). HF in the reference dataset was significantly higher compared to the 10%, 20%, and 35% missing datasets (*P*<.05; Table 4, Figure 3). There was no significant difference between the reference dataset and the 60% missing dataset (*P*>.05; Table 4 and Figure 3).

XSL•FO
RenderX

Figure 3. Heart rate variability metrics at rest (steps/min=0). HRV: heart rate variability; IBI: interbeat interval; LF/HF: ratio of low frequency power to high frequency power; RMSDRR: root-mean-square of the difference in successive IBI values; STDRR: standard deviation of interbeat interval values.



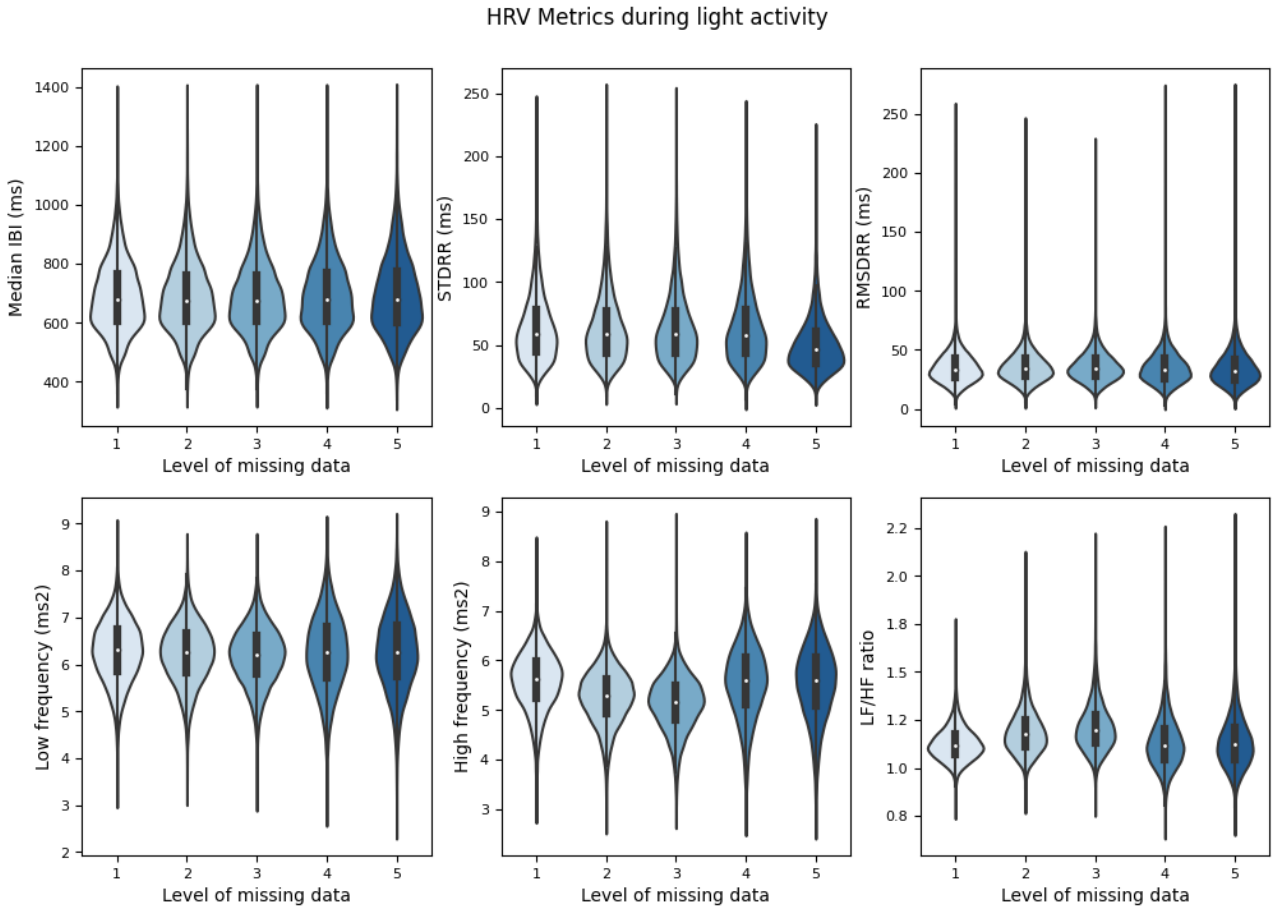
In Figure 3, reference smartwatch data were compared to (1) 10% missing data (2), 20% missing data (3), 35% missing data (4), and 60% missing data (5) per 5-minute window. Smartwatch IBI data were collected and reported from 16 individuals recruited from a community setting between May 2022 and June 2023.

Aim 1: Heart Rate Variability Metrics at Light Activity

At light activity (100>steps/min>0), there were significant differences in median IBI between the reference and all simulated datasets (P<.05; Table 5 and Figure 4). STDRR and RMSDRR values were different between the reference and simulated datasets with 10%, 20%, and 60% missing data (P<.05; Table 5 and Figure 4). Differences between the reference dataset and 35% missing data were not statistically significant, although results trended toward statistical significance (P=.06). LF and HF metrics demonstrated significant differences between the reference dataset and all simulated datasets (P<.05; Table 5 and Figure 4). LF/HF in the reference dataset was significantly higher compared with the 10% and 20% missing datasets (P<.05; Table 5 and Figure 4). There was no significant difference between the reference dataset and the 35% and 60% missing datasets (P>.05; Table 5 and Figure 4).

XSL•FO
RenderX

Figure 4. Heart rate variability metrics at light activity (100>steps/min>0). HRV: heart rate variability; IBI: interbeat interval; LF/HF: ratio of low frequency power to high frequency power; RMSDRR: root-mean-square of the difference in successive IBI values; STDRR: standard deviation of interbeat interval values.



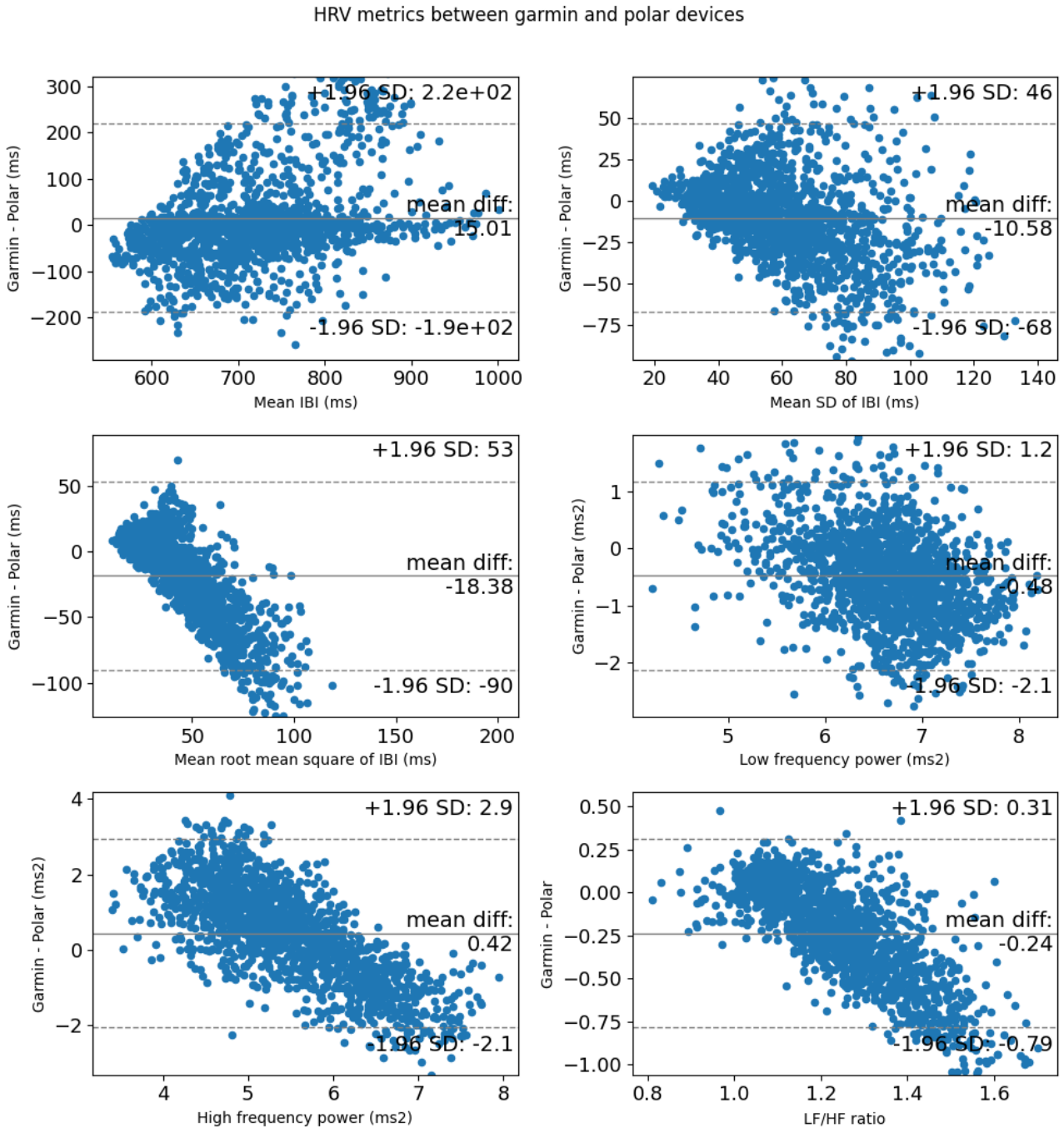
In Figure 4, Reference smartwatch data were compared to (1) 10% missing data (2), 20% missing data (3), 35% missing data (4), and 60% missing data (5) per 5 min window. Smartwatch IBI data was collected and reported from 16 individuals recruited from a community setting between May 2022 and June 2023.

Aim 2: Validity and Agreement of Photoplethysmography-Derived Heart Rate Variability Compared With ECG-Derived Heart Rate Variability

Median IBI values calculated from Garmin photoplethysmography sensors had moderate agreement (ICC=0.585) and consistency (ICC=0.589) with Polar ECG data, and LF calculated from Garmin photoplethysmography sensors had moderate consistency (ICC=0.545) with Polar ECG data. All other HRV metrics (STDRR, RMSDRR, LF, HF, and LF/HF) demonstrated poor agreement between Garmin photoplethysmography and Polar ECG data (ICC=0.071-0.470, Table 6). Bland-Altman visual analysis indicated systematic bias present in all HRV metrics, with particularly notable bias in RMSDRR, HF, and LF/HF (Figure 5). The root-mean-square error between Garmin and Polar devices for each HRV metric are listed in Table 6.

XSL•FO
RenderX

Figure 5. Bland-Altman plots for differences in heart rate variability metrics measured with a Garmin smartwatch and Polar heart rate monitor. HRV data was collected and reported from 5 individuals recruited from a community setting between May 2022 and June 2023. HRV: heart rate variability; IBI: interbeat interval; LF/HF: ratio of low frequency power to high frequency power.



XSL•FO
RenderX

Table 6. Agreement and consistency between heart rate variability metrics were measured with a Garmin smartwatch and Polar heart rate monitor in a community sample of 5 individuals who provided data between May 2022 and June 2023.

Heart rate variability metric	ICC[a] agreement, mean (95% CI)	ICC consistency, mean (95% CI)	Root-mean-square error
Median interbeat interval	0.585 (0.538 to 0.626)	0.589 (0.547 to 0.628)	105.23
SD of interbeat interval values	0.441 (0.302 to 0.545)	0.472 (0.417 to 0.521)	30.92
Root mean square of the difference in successive interbeat interval values	0.214 (0.028 to 0.357)	0.255 (0.178 to 0.324)	40.98
Low-frequency power	0.473 (0.163 to 0.645)	0.545 (0.497 to 0.588)	0.97
High-frequency power	0.404 (0.283 to 0.499)	0.429 (0.370 to 0.483)	1.35
Low-frequency to high-frequency power ratio	0.071 (–0.09 to 0.207)	0.117 (0.024 to 0.2)	0.37

[a]ICC: intraclass correlation coefficient.

Discussion

Principal Findings

To our knowledge, this paper shows for the first time the results of long-term (months) continuous monitoring of high-resolution HRV data during daily life activities using a COTS smartwatch with photoplethysmography sensor and activity (step count) data. This study describes a methodology for the extraction of HRV metrics from IBI time-series data that resulted in stable and valid metrics while using the least amount of available data. By understanding patterns of missing data, we can maximize the amount of usable data and minimize the impact of data gaps due to suboptimal wear compliance or any issues in the data synchronization between the watch and the smartphone. We found that time-domain HRV metrics (median IBI, STDRR, and RMSDRR) are more resilient to missing data when the participant is in a resting state; however, during light activity missing data influence time-domain HRV metrics. Median IBI measurements remain stable at rest even at 60% data degradation. STDRR and RMSDRR measurements remain stable at rest until 35% data degradation. Frequency-domain HRV metrics (LF, HF, and LF/HF) are less resilient to missing data both at rest and during light activity and are unstable even at 10% data degradation. It is unclear why HF (during rest) was not significantly different at 60% data degradation and LF/HF (during light activity) was not significantly at 35% and 60% data degradation compared to reference data. We speculate that there may not be enough data at these levels of data degradation for accurate comparison. Results from this study suggest that analyses and algorithms that use primarily frequency-domain HRV metrics are more sensitive to sparse data collection. Researchers and engineers should carefully evaluate HRV metrics in situations where data are sparse. If missing data are an issue in a research study, it may be beneficial to rely on metrics and algorithms using time-domain HRV metrics.

A few previous studies have simulated missing data patterns in HRV data by evaluating differences between scattered missing beats (one or two missing data points dispersed throughout the dataset) and bursts of missing beats (longer periods of missing data likely due to not wearing the device) [34,35]. Our study considered data characteristics from real HRV data patterns to remove data instead of using a random data removal approach. We found that datasets with very sparse data (missing 35% and 60% of data points per 5-minute window) typically had one burst of missing data. For example, we identified a 3.15 min data gap (60%) during the beginning of one 5-minute window. We also identified missing data between 1 minute and 2.75 minutes (35%) of another 5-minute window. The 20% and 10% missing datasets were missing 0.80 minutes and 0.63 minutes, respectively, of data scattered randomly throughout the trial, most often in 1-2 second increments. In general, 5 min windows that had data gaps that were smaller and more randomly dispersed had a greater number of IBI datapoints, whereas 5-minute windows that had one larger burst of missing data, often at the beginning of the 5-minute window, had fewer IBI datapoints. We hypothesize that these larger bursts of missing data may be the result of noncompliance (ie, not wearing the Garmin watch), and the smaller data gaps may be due to movement artifacts or device malfunction. Scattered missing data due to movement artifacts is typically considered data missing at random or missing completely at random [16]. Data missing at random or missing completely at random do not necessarily produce a bias in outcomes [17]. Bursts of missing data that occur at similar times each day or routinely for a particular purpose (ie, bathing) may be considered data missing not at random, which could lead to biases in HRV outcomes [16]. Further understanding of the type of missing data within a 5-minute window is needed in future work.

Our secondary objective was to report validity and agreement between the Garmin photoplethysmography wrist-worn sensor compared with a chest-worn ECG sensor (Polar), which is closer to a gold standard that would be used in a laboratory setting to evaluate HRV [36,37] and has been used in the field for HRV biofeedback [38]. ICCs for agreement and consistency were mostly poor between Garmin and Polar devices. ICCs for consistency were moderate for median IBI and LF, suggesting these HRV metrics may be more reliable to measure in a real-world setting. Bland-Altman plots showed systematic bias in RMSDRR, HF, and LF/HF metrics. Specifically, Garmin tended to overreport lower values and underreport higher values, indicating that Garmin generally produced a smaller range of RMSDRR, HF, and LF/HF values compared to Polar. Our study compared processed HRV metrics (beyond a simple averaged heart rate calculation) between devices under free-living conditions. Previous studies have shown higher validity and agreement between wrist-worn photoplethysmography sensors

XSL•FO
RenderX

and Polar ECG average heart rate [36,37]; however, these studies performed protocols in controlled, laboratory settings. Results from this study highlight that agreement and consistency between photoplethysmography wrist-worn sensors and ECG sensors are lower in free-living conditions and should be considered when evaluating HRV metrics. Commercially available wearable technology continues to grow at a rapid pace, and sensors that provide more accurate data are needed. Future studies should evaluate the accuracy of HRV data in emerging sensors.

Our study is the first to our knowledge to address the effects of missing data using real examples of HRV data patterns collected with photoplethysmography sensors; however, there are some limitations to our current work. In our current sample, we have limited labeling of HRV data. Future studies should include daily self-reported activity logs to determine when participants are sleeping, exercising, or not wearing their watch. We were unable to evaluate data missingness during levels of high activity (steps/min>100) because we did not have enough 5-minute windows with high step counts. It is best to evaluate high levels of activity in smaller time windows since many people engage in high activity for very short durations; however, a minimum window of 5 minutes is needed to accurately calculate frequency-domain HRV metrics [39]. In future work, we will aim to recruit individuals who are engaging in high levels of activity. In addition, we recruited healthy individuals for this study, but we did not exclude individuals with potential confounders that could have affected HRV data. For example, some participants may have had infrequent irregular heart rhythms (ie, heart skipping a beat) or may have taken medications that could have affected HRV values. Finally, our small sample size when comparing Garmin to Polar HRV data limits the generalizations we can make regarding Garmin's validity in real-world settings.

Conclusion

Findings from this study have important implications for best practices for photoplethysmography wearables use in clinical and research settings. Time-domain HRV metrics (median IBI, STDRR, and RMSDRR) collected in a resting state remain stable until at least 35% data degradation. Frequency-domain HRV metrics (LF, HF, and LF/HF) are less resilient to missing data both at rest and during light activity and are unstable even at 10% data degradation. Studies that allow for greater than 10% data degradation for frequency-domain HRV metrics may introduce bias into their estimates, specifically the underestimation of LF and HF values. Correction methods such as gap-filling are possible to replace missing data; however, differences in HRV outcomes will still exist compared with more complete datasets without data loss [34]. In general, most derived HRV metrics from the photoplethysmography-based sensors used demonstrated moderate to poor agreement and consistency with ECG-based sensors in a real-world setting, and only median IBI values had reasonable agreement and consistency between the 2 sensing modalities. Despite the differences between photoplethysmography and ECG, we may conclude that given the stability of time-domain HRV metrics for up to 35% data degradation, these metrics could result in more reliable calculations of health metrics compared with frequency-domain metrics when monitoring patients using wrist-worn photoplethysmography sensors. In conclusion, while photoplethysmography sensors are valuable for remote monitoring of patients, future work is needed to identify best practices and the most accurate HRV metrics when using photoplethysmography sensors to evaluate HRV in medical settings.

Acknowledgments

This research was developed with funding from the Defense Advanced Research Projects Agency (contract 140D63-19-C-0023). The views, opinions, and findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the US Government. This paper is approved for public release, and the distribution is unlimited. Artificial intelligence was not used in any portion of writing this manuscript.

Data Availability

The data generated during and/or analyzed during this study are not publicly available due to restrictions of the Defense Advanced Research Projects Agency (contract 140D63-19-C-0023) but are available from the corresponding author on reasonable request.

Authors' Contributions

DST, DED, and HJW conceptualized the study. DST and MHC developed the study methodology. EAP, JRH, MDB, and RPE developed the software and mobile apps. DED performed project administration. MHC, PG, and HDW managed the study cohort. HDW and LL performed the statistical analyses. HDW wrote the original manuscript. All authors edited and reviewed the manuscript and approved the final version.

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. de Farias FAC, Dagostini CM, Bicca YDA, Falavigna VF, Falavigna A. Remote patient monitoring: a systematic review. Telemed J E Health 2020;26(5):576-583. [doi: 10.1089/tmj.2019.0066] [Medline: 31314689]
2. Aalam AA, Hood C, Donelan C, Rutenberg A, Kane EM, Sikka N. Remote patient monitoring for ED discharges in the COVID-19 pandemic. Emerg Med J 2021;38(3):229-231. [doi: 10.1136/emermed-2020-210022] [Medline: 33472870]
3. Emani S. Remote monitoring to reduce heart failure readmissions. Curr Heart Fail Rep 2017;14(1):40-47. [doi: 10.1007/s11897-017-0315-2] [Medline: 28215030]
4. Cruz J, Brooks D, Marques A. Home telemonitoring effectiveness in COPD: a systematic review. Int J Clin Pract 2014;68(3):369-378. [doi: 10.1111/ijcp.12345] [Medline: 24472009]
5. Kirkland EB, Marsden J, Zhang J, Schumann SO, Bian J, Mauldin P, et al. Remote patient monitoring sustains reductions of hemoglobin A1c in underserved patients to 12 months. Prim Care Diabetes 2021;15(3):459-463 [FREE Full text] [doi: 10.1016/j.pcd.2021.01.005] [Medline: 33509728]
6. Kruklitis R, Miller M, Valeriano L, Shine S, Opstbaum N, Chestnut V. Applications of remote patient monitoring. Prim Care 2022;49(4):543-555. [doi: 10.1016/j.pop.2022.05.005] [Medline: 36357060]
7. Lubitz SA, Faranesh AZ, Selvaggi C, Atlas SJ, McManus DD, Singer DE, et al. Detection of atrial fibrillation in a large population using wearable devices: the fitbit heart study. Circulation 2022;146(19):1415-1424. [doi: 10.1161/CIRCULATIONAHA.122.060291] [Medline: 36148649]
8. Ray D, Collins T, Woolley S, Ponnapalli P. A review of wearable multi-wavelength photoplethysmography. IEEE Rev Biomed Eng 2023;16:136-151. [doi: 10.1109/RBME.2021.3121476] [Medline: 34669577]
9. Baman JR, Mathew DT, Jiang M, Passman RS. Mobile health for arrhythmia diagnosis and management. J Gen Intern Med 2022;37(1):188-197 [FREE Full text] [doi: 10.1007/s11606-021-07007-w] [Medline: 34282532]
10. Tison GH, Sanchez JM, Ballinger B, Singh A, Olgin JE, Pletcher MJ, et al. Passive detection of atrial fibrillation using a commercially available smartwatch. JAMA Cardiol 2018;3(5):409-416 [FREE Full text] [doi: 10.1001/jamacardio.2018.0136] [Medline: 29562087]
11. Ding EY, Han D, Whitcomb C, Bashar SK, Adaramola O, Soni A, et al. Accuracy and usability of a novel algorithm for detection of irregular pulse using a smartwatch among older adults: observational study. JMIR Cardio 2019;3(1):e13850 [FREE Full text] [doi: 10.2196/13850] [Medline: 31758787]
12. Alqahtani JS, Aldhahir AM, Alghamdi SM, Al Ghamdi SS, AlDraiwiesh IA, Alsulayyim AS, et al. A systematic review and meta-analysis of heart rate variability in COPD. Front Cardiovasc Med 2023;10:1070327 [FREE Full text] [doi: 10.3389/fcvm.2023.1070327] [Medline: 36873414]
13. Damla O, Altug C, Pinar KK, Alper K, Dilek IG, Kadriye A. Heart rate variability analysis in patients with multiple sclerosis. Mult Scler Relat Disord 2018;24:64-68. [doi: 10.1016/j.msard.2018.06.012] [Medline: 29957350]
14. Heimrich KG, Lehmann T, Schlattmann P, Prell T. Heart rate variability analyses in Parkinson's disease: a systematic review and meta-analysis. Brain Sci 2021;11(8):959 [FREE Full text] [doi: 10.3390/brainsci11080959] [Medline: 34439578]
15. Esterov D, Greenwald BD. Autonomic dysfunction after mild traumatic brain injury. Brain Sci 2017;7(8):1-8. [doi: 10.3390/brainsci7080100] [Medline: 28800081]
16. Di J, Demanuele C, Kettermann A, Karahanoglu FI, Cappelleri JC, Potter A, et al. Considerations to address missing data when deriving clinical trial endpoints from digital health technologies. Contemp Clin Trials 2022;113:106661 [FREE Full text] [doi: 10.1016/j.cct.2021.106661] [Medline: 34954098]
17. Kang H. The prevention and handling of the missing data. Korean J Anesthesiol 2013;64(5):402-406 [FREE Full text] [doi: 10.4097/kjae.2013.64.5.402] [Medline: 23741561]
18. Temple DS, Hegarty-Craver M, Furberg RD, Preble EA, Bergstrom E, Gardener Z, et al. Wearable sensor-based detection of influenza in presymptomatic and asymptomatic individuals. J Infect Dis 2023;227(7):864-872 [FREE Full text] [doi: 10.1093/infdis/jiac262] [Medline: 35759279]
19. Helmer P, Hottenrott S, Rodemers P, Leppich R, Helwich M, Pryss R, et al. Accuracy and systematic biases of heart rate measurements by consumer-grade fitness trackers in postoperative patients: prospective clinical trial. J Med Internet Res 2022;24(12):e42359 [FREE Full text] [doi: 10.2196/42359] [Medline: 36583938]
20. Schaffarczyk M, Rogers B, Reer R, Gronwald T. Validity of the polar H10 sensor for heart rate variability analysis during resting state and incremental exercise in recreational men and women. Sensors (Basel) 2022;22(17):6536 [FREE Full text] [doi: 10.3390/s22176536] [Medline: 36081005]
21. Perrotta AS, Jeklin AT, Hives BA, Meanwell LE, Warburton DER. Validity of the elite HRV smartphone application for examining heart rate variability in a field-based setting. J Strength Cond Res 2017;31(8):2296-2302. [doi: 10.1519/JSC.0000000000001841] [Medline: 28195974]
22. Malik M, Bigger JT, Camm AJ, Kleiger RE, Malliani A, Moss AJ, et al. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. European Heart Journal 1996;17(3):354-381. [doi: 10.1093/oxfordjournals.eurheartj.a014868]
23. Lewis GF, Furman SA, McCool MF, Porges SW. Statistical strategies to quantify respiratory sinus arrhythmia: are commonly used metrics equivalent? Biol Psychol 2012;89(2):349-364 [FREE Full text] [doi: 10.1016/j.biopsycho.2011.11.009] [Medline: 22138367]

XSL•FO
RenderX

24. Shader TM, Gatzke-Kopp LM, Crowell SE, Jamila Reid M, Thayer JF, Vasey MW, et al. Quantifying respiratory sinus arrhythmia: effects of misspecifying breathing frequencies across development. Dev Psychopathol 2018;30(1):351-366. [doi: 10.1017/S0954579417000669] [Medline: 28554343]
25. Hayano J, Yuda E. Pitfalls of assessment of autonomic function by heart rate variability. J Physiol Anthropol 2019;38(1):3 [FREE Full text] [doi: 10.1186/s40101-019-0193-2] [Medline: 30867063]
26. Herrmann SD, Barreira TV, Kang M, Ainsworth BE. Impact of accelerometer wear time on physical activity data: a NHANES semisimulation data approach. Br J Sports Med 2014;48(3):278-282. [doi: 10.1136/bjsports-2012-091410] [Medline: 22936409]
27. van Rossum G, Drake FL. Python 3 Reference Manual. Scotts Valley, CA: Scotts Valley; 2009.
28. Giboin LS, Tokuno C, Kramer A, Henry M, Gruber M. Motor learning induces time-dependent plasticity that is observable at the spinal cord level. J Physiol 2020;598(10):1943-1963 [FREE Full text] [doi: 10.1113/JP278890] [Medline: 32115702]
29. Bates D, Maechler M, Bolker B, Walker S. lme4: Linear mixed-effects models using Eigen and S4. R package version 1. URL: http://CRAN.R-project.org/package=lme4 [accessed 2024-12-14]
30. Gamer M, Lemon J, Fellows I, Singh P. irr: Various coefficients of interrater reliability and agreement. R package version 0.84.1. Cran 2019:1-32.
31. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
32. Gerke O. Reporting standards for a bland-altman agreement analysis: a review of methodological reviews. Diagnostics (Basel) 2020;10(5):334 [FREE Full text] [doi: 10.3390/diagnostics10050334] [Medline: 32456091]
33. Core Team R. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2019. URL: https://www.r-project.org/ [accessed 2024-12-14]
34. Cajal D, Hernando D, Lázaro J, Laguna P, Gil E, Bailón R. Effects of missing data on heart rate variability metrics. Sensors (Basel) 2022;22(15):5774 [FREE Full text] [doi: 10.3390/s22155774] [Medline: 35957328]
35. Baek HJ, Shin J. Effect of missing inter-beat interval data on heart rate variability analysis using wrist-worn wearables. J Med Syst 2017;41(10):147. [doi: 10.1007/s10916-017-0796-2] [Medline: 28812280]
36. Montalvo S, Martinez A, Arias S, Lozano A, Gonzalez MP, Dietze-Hermosa MS, et al. Commercial smart watches and heart rate monitors: a concurrent validity analysis. J Strength Cond Res 2023;37(9):1802-1808. [doi: 10.1519/JSC.0000000000004482] [Medline: 36862131]
37. Khushhal A, Nichols S, Evans W, Gleadall-Siddall DO, Page R, O'Doherty AF, et al. Validity and reliability of the apple watch for measuring heart rate during exercise. Sports Med Int Open 2017;1(6):E206-E211 [FREE Full text] [doi: 10.1055/s-0043-120195] [Medline: 30539109]
38. Davila MI, Kizakevich PN, Eckhoff R, Morgan J, Meleth S, Ramirez D, et al. Use of mobile technology paired with heart rate monitor to remotely quantify behavioral health markers among military reservists and first responders. Mil Med 2021;186(Suppl 1):17-24. [doi: 10.1093/milmed/usaa395] [Medline: 33499533]
39. Bourdillon N, Schmitt L, Yazdani S, Vesin JM, Millet GP. Minimal window duration for accurate HRV recording in athletes. Front Neurosci 2017;11:456 [FREE Full text] [doi: 10.3389/fnins.2017.00456] [Medline: 28848382]

Abbreviations

COTS: commercial off the shelf
ECG: electrocardiogram
HF: high-frequency
HRV: heart rate variability
IBI: interbeat interval
ICC: intraclass correlation coefficient
LF: low-frequency
REDCap: Research Electronic Data Capture
RMSDRR: root-mean-square of the difference in successive IBI values
RTI: Research Triangle Institute
STDRR: standard deviation of interbeat interval values
Q-Q: Quantile-Quantile

XSL•FO
RenderX

Edited by A Mavragani; submitted 16.10.23; peer-reviewed by N Peterson, N Shafagati; comments to author 28.04.24; revised version received 28.06.24; accepted 24.11.24; published 24.02.25.

Please cite as:
Davis-Wilson H, Hegarty-Craver M, Gaur P, Boyce M, Holt JR, Preble E, Eckhoff R, Li L, Walls H, Dausch D, Temple D
Effects of Missing Data on Heart Rate Variability Measured From A Smartwatch: Exploratory Observational Study
JMIR Form Res 2025;9:e53645
URL: https://formative.jmir.org/2025/1/e53645
doi:10.2196/53645
PMID:

©Hope Davis-Wilson, Meghan Hegarty-Craver, Pooja Gaur, Matthew Boyce, Jonathan R Holt, Edward Preble, Randall Eckhoff, Lei Li, Howard Walls, David Dausch, Dorota Temple. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development and Implementation of MyPainHub, a Web-Based Resource for People With Musculoskeletal Conditions and Their Health Care Professionals: Mixed Methods Study

Kerrie Evans[1,2], PhD; Jonathan Ko[1], MPhysio; Dragana Ceprnja[3], PhD; Katherine Maka[3], MBA; Darren Beales[4], PhD; Michele Sterling[5], PhD; Kim L Bennell[6], PhD; Gwendolen Jull[5], PhD; Paul W Hodges[5], PhD; Marnee J McKay[1], PhD; Trudy J Rebbeck[1], PhD

[1]The University of Sydney, Sydney, Australia

[2]Healthia Limited, Bowen Hills, Australia

[3]Western Sydney Local Health District, Westmead, Australia

[4]Curtin University, Perth, Australia

[5]The University of Queensland, Brisbane, Australia

[6]The University of Melbourne, Melbourne, Australia

Corresponding Author:
Kerrie Evans, PhD
The University of Sydney
City Road, Camperdown
Sydney, 2050
Australia
Phone: 61 0731804900
Email: kerrie.evans@healthia.com.au

Abstract

Background: Musculoskeletal conditions, including low back pain (LBP), neck pain, and knee osteoarthritis, are the greatest contributors to years lived with disability worldwide. Resources aiming to aid both patients and health care professionals (HCPs) exist but are poorly implemented and adopted.

Objective: We aimed to develop and implement MyPainHub, an evidence-based web-based resource designed to provide comprehensive, credible and accessible information for people with, and HCPs who manage, common musculoskeletal conditions.

Methods: This mixed methods study adhered to the New South Wales Translational Research Framework and was evaluated against the Reach, Effectiveness, Adoption, Implementation, and Maintenance (RE-AIM) framework. Consultation with key stakeholders (patients, HCPs, researchers, industry, consumer groups, and website developers) informed content, design, features, and functionality. Development then aimed to meet the identified need for a "one-stop shop"—a central location for information about common musculoskeletal conditions tailored to a person's condition and risk of poor outcomes. MyPainHub was then developed through an iterative process and implementation strategies were tailored to different health care settings. Quantitative and qualitative evaluation occurred with patients and HCPs.

Results: In total, 127 stakeholders participated in the development phase; initial consultation with them led to embedding 2 validated screening tools (the Short Form Örebro Musculoskeletal Pain Screening Questionnaire and the Keele STarT MSK tool) in MyPainHub to guide information tailoring for patients based on risk of poor outcomes. Development occurred in parallel and feedback from stakeholders informed design and content including structure, functionality, and phrasing and images to use to emphasize key points. Consultation resulted in information for patients being categorized using key guideline-based messages (general information, your pathway, exercise, and imaging) while information for clinicians was categorized into assessment, management, and prognosis. Implementation occurred in different health care settings with the most effective strategies being interactive education via webinars and workshops. The evaluation phase involved web-based questionnaires (patients: n=44; HCPs: n=29) and focus groups (patients: n=6; HCPs: n=6). Patients and HCPs found MyPainHub user-friendly, acceptable, credible, and potentially able to support self-management. Patient participants identified areas for improvement such as including more specific information on preventative measures and pain relief options. Despite positive feedback, only 35% (10/29) of HCPs used MyPainHub with their patients. HCP participants identified challenges including insufficient training and lack of familiarity

XSL•FO
RenderX

with using web-based resources in existing clinical workflows. Following implementation, the information contained on MyPainHub changed knowledge and practice for some patients and HCPs.

Conclusions: Following extensive and iterative stakeholder engagement, MyPainHub was developed as an evidence-based web-based resource and perceived by patients and HCPs as user-friendly, credible, and acceptable. Active implementation strategies are required for adoption and implementation and greater training focusing on strategies to implement MyPainHub into clinical practice may be necessary.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12619000871145; https://tinyurl.com/438kkyt3

(JMIR Form Res 2025;9:e63780) doi:10.2196/63780

KEYWORDS

clinical pathways; allied health; self-management; health information; ehealth; co-design

Introduction

Musculoskeletal conditions are characterized by pain and reduced physical function, often resulting in deterioration of mental well-being, increased susceptibility to developing other chronic health conditions, and loss of productive life years in the workforce [1]. Globally, musculoskeletal conditions affect 1.7 billion people and are ranked as the highest contributors to years lost to disease, above mental, respiratory, and cardiovascular diseases [1-3]. Among people presenting for care, low back pain (LBP), neck pain, whiplash associated disorders (WAD), and knee osteoarthritis are the most prevalent and disabling musculoskeletal conditions [1,4]. In Australia, the health care system operates as a mixed public-private model, with primary funding through Medicare, a universal public insurance system providing subsidized access to health care services, supplemented by private insurance options. The federal and state governments manage health care collaboratively, with federal oversight on Medicare and Pharmaceutical Benefits Scheme funding, while states are responsible for public hospital management and service delivery. Despite targeted government initiatives to mitigate musculoskeletal disease burden, such as the national strategic action plan for arthritis and funding under the chronic disease management plan, musculoskeletal conditions in Australia still account for 20% of the total burden of disease [5].

One factor contributing to this burden is poor implementation of guideline-based care [6,7]. Broadly speaking, clinical guidelines for LBP [8,9], neck pain [10,11], WAD [12], and knee osteoarthritis [13-16], contain similar recommendations for the provision of evidence-based care—screening for red flags and risk factors for poor outcome, providing patient-centered education (including options for care, advice to stay active), and avoiding unnecessary investigations and interventions. However, health care professionals (HCPs) often fail to screen for prognostic factors and continue to overuse nonrecommended treatments, imaging, pharmacological interventions, and surgery when managing musculoskeletal conditions [17-22]. Non–guideline-based care occurs across health care settings (ie, primary and tertiary care settings) despite evidence that these low-value practices do not improve health outcomes, waste health care resources and can be harmful [17,23-25].

Barriers to implementation and adoption of clinical guidelines include system-related factors (eg, funding models [26,27]) but also practitioner and patient-related factors. For example, lack of time, limited access to evidence-based resources, and discordant expectations are cited by both HCPs and patients as reasons why evidence-based recommendations are not followed [28-31]. In response to some of these challenges and the increasing trend for people to obtain health information from the internet, websites and other digital resources about musculoskeletal conditions have been developed [32]. The sheer volume of resources now available is overwhelming and information from the internet can be inaccurate, perceived to have low credibility, lack cultural diversity and not meet the needs of consumers [33-38]. In addition, often there is little training or education provided to HCPs as to how to best use internet-based resources to help implement, adopt and support evidence-based care [39-41].

To empower and engage patients and HCPs to make informed decisions about evidence-based recommendations, websites should be developed and co-designed by end users, include input from experts and opinion leaders, and be practical and scalable, that is, follow a translational research framework [42]. Examples of websites that have adopted a translational research framework include MyBackPain [33], MyWhiplashNavigator [43], and MyJointPain [44]. The development of these websites began with a needs assessment to identify research priorities and consumer needs before creating and curating guidelines and evidence-based health care information. Stakeholder engagement was essential in content creation and testing. A dissemination strategy was developed to ensure that the website was reached by the target audience. Monitoring and evaluation mechanisms tracked website usage and user feedback for quality improvement and their impact has been investigated in clinical trials [45,46]. Although these websites have undergone a robust and rigorous process of development, the information contained on these websites is condition specific. Given that many people experience multiple coexisting musculoskeletal conditions [47], and that clinical guidelines for common musculoskeletal conditions contain similar key messages [48], the aim of this study was to develop a website that could serve as a hub or a “one-stop shop” for both patients with, and HCPs managing, musculoskeletal conditions. The website was also initially developed to support patients and HCPs enrolled in a randomized control trial that was due to commence—the PAthway of CarE (PACE) musculoskeletal (MSK) clinical trial

XSL•FO
RenderX

[49]. PACE-MSK, funded by Australia's National Health and Medical Research Council (GNT1141377), was designed to evaluate the effectiveness of a clinical PACE for people with musculoskeletal disorders where care is provided based on people's risk of poor outcome [49]. A website that could serve as a one-stop shop for both patients and clinicians in PACE-MSK, as well as support patients and HCPs beyond the trial, was an important part of the design. This study reports on the development, implementation, and evaluation of MyPainHub, a website designed to provide support to patients and HCPs in the management of LBP, neck pain, WAD, and knee osteoarthritis.

Methods

Overview

This mixed methods study involved 3 stages (development, implementation, and evaluation), and followed the New South Wales Translational Research Framework [42] with outcomes assessed according to the Reach, Effectiveness, Adoption, Implementation, and Maintenance (RE-AIM) framework [50]. Both frameworks provide structured approaches to ensure that research is relevant and scalable and are widely used for research in the Australian health care context.

Ethical Considerations

Ethics approval was obtained from the Human Research Ethics Committees at The University of Sydney (2018/926, 2021/951), The University of Melbourne (1954239), The University of Queensland (2019000700), Curtin University (HRE2019-0263, HRE2020-0562) and Western Sydney Local Health District (2021/ETH11872). Participants were informed about the study before providing their written consent (either on paper or via a web-based form). Participation was voluntary and participants were informed about their right to withdraw from the study at any point without any consequences. Collected data were coded and deidentified.

Recruitment

Each stage involved different recruitment procedures. During the development stage, researchers, consumer, industry and clinical partners were identified by members of the research team and invited to participate via direct contact. These stakeholders were those with recognized expertise in musculoskeletal conditions and who had previously collaborated with members of the research team or been involved in previous studies. HCPs were involved in the PACE-MSK trial or had been involved in previous trials and had consented to be contacted for future trials. HCPs were working in either primary or tertiary (hospital) settings. People with MSK conditions had been involved in previous trials and had consented to be contacted for future trials. Contact was made with these groups via email or direct phone call. Students allied to HCPs were enrolled in a unit of study at a participating university and had consented to participate.

In the implementation and evaluation phases, HCPs and patients were involved in the PACE-MSK trial or had been involved in previous trials and had consented to be contacted for future trials. In addition, HCPs working in occupational health settings known to the research team or who had been involved in previous studies were invited to participate either by email or direct phone call. Those HCPs working in our clinical partners' tertiary care setting (Western Sydney Local Health District, [WSLHD]) were directly invited by their peers and managers. Patients with MSK conditions were also invited to participate by their treating clinician.

Stage 1: Development

The aim of the development stage was to design a website where patients and HCPs could access relevant, up-to-date, high-quality evidence-based information on common musculoskeletal conditions (LBP, neck pain, WAD and knee osteoarthritis). Given the expected commencement date of the PACE-MSK trial, the development stage incorporated comprehensive stakeholder engagement rather than a formal co-design process. The problem identified by the research team was that multiple resources exist, with neither a central nor coordinated location for people to access credible evidence-based information [51].

The first step of the New South Wales Research Translation Framework is Idea Generation, where key stakeholders are engaged to design a solution to the problem. Our core team comprised clinician- researchers from 4 Australian Universities (The University of Sydney, The University of Queensland, The University of Melbourne and Curtin University) with research expertise in musculoskeletal conditions. The team determined that key stakeholders likely to benefit from this resource were allied HCPs working in both primary and tertiary care, researchers involved in creating evidence for musculoskeletal conditions, specialist musculoskeletal clinicians, specialist medical professionals, people with musculoskeletal disorders, consumer organizations and industry partners. We then determined the level of influence and interest of each stakeholder using a stakeholder engagement matrix and considered appropriate strategies for managing their involvement and expectations [52]. Engagement methods ranged from being involved in co-design and implementation (manage) to piloting of resources (meet needs) and meetings (keep informed; Table 1).

XSL•FO
RenderX

Table 1. Classification of key stakeholders involved in the development stage of MyPainHub according to influence and interest, appropriate strategies and engagement methods.

Key stakeholder	Influence and interest	Appropriate strategies	Engagement methods
Allied HCPs[a], Specialist musculoskeletal allied HCPs	Higher influence, higher interest	Involve	• Consultation on key content • Facilitated meetings to review and refine content • Focus groups to pilot and identify gaps • Involvement in implementation process
Student allied HCPs	Lower influence, higher interest	Involve	• Involve in implementation and evaluation process
Researchers involved in musculoskeletal health	Higher influence, higher interest	Involve	• Consultation on key content and existing resources
Specialist medical professionals	Lower influence, higher interest	Meet needs	• Semistructured interview on resources and pathway
People with musculoskeletal conditions	Lower influence, higher interest	Meet needs	• Review and piloting of resources with feedback
Consumer partners: Arthritis Australia, Australian Pain Management Association, Musculoskeletal Australia	Higher influence, lower interest	Keep informed	• In-person meetings
Industry and clinical partners: Australian Physiotherapy Association, Chiropractic Australia, State Insurance Regulatory Authority of New South Wales, Primary Healthcare Networks, Knowledge Translation Australia	Higher influence, lower interest	Keep informed	• In-person meetings

[a]HCPs: health care professionals.

We held work-in-progress meetings with key stakeholders between July 2018 and October 2022. Discussions included design concepts, content, functionality, website maintenance and support and how this could be implemented. To ensure optimal functionality and ease of use, a digital agency (Vivo Group, Brisbane, Queensland, Australia) with experience in development of user-friendly websites was selected to design the website following a robust tendering process. As the content was developed, each stakeholder group shared their views and perspectives and provided feedback about additional key messages and resources that should be included or omitted. Before implementation, the website went through final checks by the research team and the digital agency to test features, functionality and resolve any issues.

Stage 2: Implementation

The aim of the implementation stage was to determine strategies that would help ensure effective and scalable implementation of the website before its widespread public release. Groups thought likely to source information about musculoskeletal conditions were identified and included allied health students and graduate allied HCPs working in health care settings where people with musculoskeletal conditions present. Settings included primary care, occupational health, and tertiary care settings. Finally, allied HCPs working in a specialist capacity were engaged. This process ensured implementation occurred with clinicians with a range of clinical experiences and in different care settings. We chose implementation strategies ranging from passive (dissemination) to more active strategies (facilitated and discussion-based education with opinion leaders) based on strategies that had been successful in our previous work (Table 1) [33,43,44]. During this stage, the website was not publicly accessible.

Patients with musculoskeletal conditions were invited by their treating clinician to use the resources on the website and then provided consent to participate in qualitative studies associated with both our primary and tertiary care implementation initiatives. After using the website, patients participated in a semistructured interview (via a web platform) or focus group (in person), depending on what was feasible in the particular health care setting and on the patient's preference and location, to provide their opinions on the resource.

Allied HCPs Working in Primary Health Care Settings

They were engaged via a passive (dissemination) strategy to use the website. The primary HCPs providing care for people with musculoskeletal conditions in the PACE-MSK trial [49] were informed of the trial and the website via a phone call. Both the HCP and their patients were emailed a link to the website and encouraged to refer to the resources during the episode of care. Over an 18-month period, the researchers held fourteen 1-hour video-conferenced noncompulsory educational meetings inviting HCPs to attend with the aim of keeping them engaged in the trial and website.

XSL•FO
RenderX

Allied HCPs Working in Occupational and Tertiary Care Settings

They were invited to participate in an active (interactive education using opinion leaders) implementation strategy. HCPs working in occupational health settings were invited to participate either by email or direct phone call. Those working in our clinical partners' tertiary care setting (WSLHD) were directly invited by their peers and managers. Those accepting the invitation then attended a web-based (occupational health) or in-person (tertiary care) discussion-based educational workshop explaining the purpose of the website and its key features. These HCPs were then encouraged to use the resources on MyPainHub to assist in assessment and management of their patients. After completion of the episode of care (or after 3 months), the HCPs were invited to complete a questionnaire to evaluate acceptability, feasibility, credibility, and actual practice of using the website and then participated in a focus group to explore their opinions in more depth (Multimedia Appendix 1).

Specialist Allied HCPs

They were invited to participate in a 2-day facilitated discussion-based educational workshop described elsewhere [53]. The workshop was designed to inform specialist HCPs about how to implement risk-stratified care (supported by MyPainHub) for people at risk of poor outcome. Key features that may facilitate the provision of this care were discussed during the workshop, including how MyPainHub could be used to support management. Two follow-up web-based meetings were held to encourage use of the website and implementation of a new pathway of care.

Student HCPs

They were engaged through classroom education. The website was integrated as a learning resource in 2 units of study for physiotherapy students in their final year of study at The University of Sydney. Key components of the education model included (1) risk assessment using the web-based risk-assessment tools resources available on the website, (2) using the resources to assist the provision of evidence-based treatment (guideline-based exercises and risk-based advice) for people with musculoskeletal disorders, and (3) timely and appropriate referral to specialist musculoskeletal clinicians for people at high risk of poor outcomes.

Outcome Assessed

The main outcome assessed for the implementation stage was the reach of the strategy. Reach was primarily assessed by the number registered on the resource or those invited to use it. Website traffic and trends (eg, total visits, total page views, and top landing pages) were collected through Google Analytics and built-in website reports.

Stage 3: Evaluation

The aim of the evaluation stage was to investigate both effectiveness, acceptability, and adoption of the website (RE-AIM framework) by allied HCPs (students, allied HCPs, and specialist HCPs) and people with musculoskeletal conditions. Mixed methods involving quantitative (web-based survey) and qualitative (focus groups or semistructured interviews) were used.

Allied HCPs (Students to Specialists)

Participants were invited to share their perspectives about MyPainHub via a web-based survey. Participants answered questions relating to change in knowledge and practice (eg, "The website has changed the way I use prognostic tools"), acceptability and credibility of the website (eg, "I trusted the information on the website"), structure and features of the website (eg, "It was easy to navigate through the website"), and overall satisfaction with the website. HCPs were also asked open-ended questions which explored opinions on the best features of the website, barriers to use and suggestions for improvement. Student HCPs were asked additional questions relating to whether the information on the website aligned with what they had been taught and helped them achieve the learning outcomes of their course (eg, "The messages on the website about imaging made sense to me"; "Viewing the resources in MyPainHub helped me achieve the learning outcomes of my course").

Qualitative data were collected from a proportion of HCPs who had been directly invited to participate by their peers and managers and who had consented to participate in a focus group. Participants were provided with the option of completing only the web-based survey or participating in both the web-based survey and the focus group. The focus groups were conducted in person and were designed to explore HCPs' opinions in greater depth.

People With Musculoskeletal Conditions

As occurred with HCPs, people with musculoskeletal conditions were invited to share their perspectives about MyPainHub via a web-based survey. Patient participants answered questions designed to assess (change in) knowledge (eg, "People with my condition should not exercise until they have had scans done"), acceptability and credibility of the website (eg, "I trusted the information on the website"), structure and features of the website (eg, "It was easy to navigate through the website"), and overall satisfaction with the website. Patient participants were also asked open-ended questions which explored opinions on the best features of the website, barriers to use and suggestions for improvement.

Qualitative data were collected from a proportion of people with musculoskeletal conditions who consented to participate in a focus group (in person) to explore their perspectives in greater depth. Participants in the PACE-MSK trial participated in semistructured interviews (via a web platform) 3 months after entering the trial but these findings will not be reported in this paper.

Data Analysis

With respect to the web-based surveys, closed-ended questions were rated using a 5-point Likert scale ranging from 1 (strongly disagree or extremely unsatisfied) to 5 (strongly agree or extremely satisfied). The open-ended questions underwent a descriptive analysis.

XSL•FO
RenderX

With respect to the qualitative data collected during the semistructured interviews and focus groups, the discussions were recorded and transcribed. Analysis was conducted first by familiarization, then preliminary coding, searching, reviewing and naming and then reporting [54]. The transcriptions were manually checked by a member of the research team and corrected against the verbatim account. The transcriptions were read several times to make sense of the data. Open coding was conducted by naming sections of the participant responses in the text. Codes were then grouped to form meaningful categories in discussion with all authors. The next step was to construct broad themes from the categories with authors working collaboratively in the interpretation of qualitative data to draw on the combined insights of all members of the research team.

Results

Overview

The total number of participants involved in each stage is provided in Table 2.

Table 2. Stakeholders involved in development, implementation and evaluation of MyPainHub.

Stages and stakeholders	Participants involved, n
Development	
HCPs[a]	67
Researchers	28
Specialist medical professionals	6
People with musculoskeletal conditions	18
Consumer partners	3
Industry and clinical partners	5
Implementation	
HCPs	136
People with musculoskeletal conditions	778
Evaluation	
HCPs (web-based surveys)	29
HCPs—focus group (tertiary care, in person)	6
People with musculoskeletal conditions (web-based surveys)	11
People with musculoskeletal conditions—focus groups (tertiary care, in person)	6
People with musculoskeletal conditions—semistructured interviews (primary care, via a web platform)[b]	33

[a]HCPs: health care professionals.

[b]Findings not reported in the present study.

Stage 1: Development

The stakeholder consultation generated ideas that informed the website design and implementation. This process informed the development of the website both for "Patients" and "Clinicians." Key features and content were identified and then modified iteratively as a result of the consultation process. Modifications did not follow a strict, linear order occurring instead in parallel.

MyPainHub for Patients

The core clinician-researcher team initially identified that common key guideline-based messages for implementation among people with musculoskeletal conditions were to provide accurate advice, appropriate messaging regarding imaging, to promote exercise, and provide differential care based on risk of poor outcome. The stakeholder consultation process resulted in agreement to structure the website around prognostic risk and musculoskeletal condition, with resources automated thereafter.

Two risk tools, the Short Form Örebro Musculoskeletal Pain Screening Questionnaire (SF-ÖMPSQ) [55] and the Keele STarT MSK tool [56] were identified by the research team as being validated tools to risk-stratify people with musculoskeletal conditions. Initial surveys of the specialist HCPs (n=50) suggested that 44% (22/50) either frequently or always used the SF-ÖMPSQ, 14% (7/50) the WhipPredict tool [57,58], 8% (4/50) the Keele STarT MSK tool [56], and 6% (3/50) the STarT Back Tool [59,60]. However, given WhipPredict is specific for WAD and the STarT Back Tool is specific for LBP, it was decided to embed the SF-ÖMPSQ and the Keele STarT MSK tool (both designed for broader musculoskeletal conditions) on the website for patients to complete with scores being automatically calculated. After completing the tools and nominating their primary condition (LBP, neck pain, or knee osteoarthritis), patients access resources that are designed specifically for both their risk subgroup (according to their score on the SF-ÖMPSQ) and condition. Each time a patient logs into the website, they are taken to their condition and risk-specific home page but can access information for the other conditions

XSL•FO
RenderX

through simple navigation links and drop-down menus where appropriate.

The stakeholder consultation process identified that people with musculoskeletal conditions (n=18) wanted general information about their condition, ideas on how they could adapt to their lifestyle, their work and what they could do to help themselves. Consultation with industry (Table 1) resulted in adding links to other evidence-based websites and resources (eg, the Royal Australian College of General Practitioner's guidelines for the management of knee osteoarthritis) and highlighted that options for care should be clearly explained (eg, people at low risk of poor outcome should be encouraged to avoid imaging and surgery and be provided with evidence-based advice and education whereas those at high risk should be provided with more comprehensive education related to the multifactorial, biopsychosocial nature of pain disorders). Consultation with patients with musculoskeletal conditions resulted in information being written in a friendly and nonjudgemental way and links to other websites and resources were selected so that they were a "simple click away" to avoid users being sent to another website where they would have to create an account. The consultation process resulted in the resources being structured under the categories of "General Information," "Your Pathway," "Exercise," and "Imaging."

MyPainHub for HCPs

Initial stakeholder consultation with 67 HCPs resulted in >200 individual pieces of feedback about the content, structure, features, and resources to use. Feedback included, for example, modifications to phrasing used or emphasis of key points (eg, "More emphasis on 'not one posture is best' would be good and give permission to sit in a comfortable, relaxed position to reduce or prevent fear of movement"), additional information to add (eg, "You could add some info about basic exercise such as walking, range of movement, balance and light upper limb weights"), structure (eg, "It was very easy to navigate"), and appearance (eg, "I think the pages that have banner images should have those images reduced in size – the banners are little dominating"). This feedback, and changes made to the website, was iterative. Overall, the feedback suggested that the website should contain key guideline-based messages, include "bite-sized" videos of information delivered by credible HCPs, include key practice points, open-access contemporary articles, clinical guidelines and outcome measures, and links to other credible resources. Subsequently, the information on MyPainHub for clinicians was structured under the categories of "Assessment," "Prognosis," and "Management."

Structure and Appearance

The website designers provided expertise on the color scheme, structure and functionality of the website to achieve the aims of the project. The content was written in a professional and friendly tone, and images were selected that avoided emphasizing, for example, a pathoanatomical focus where not appropriate and that avoided stereotyping.

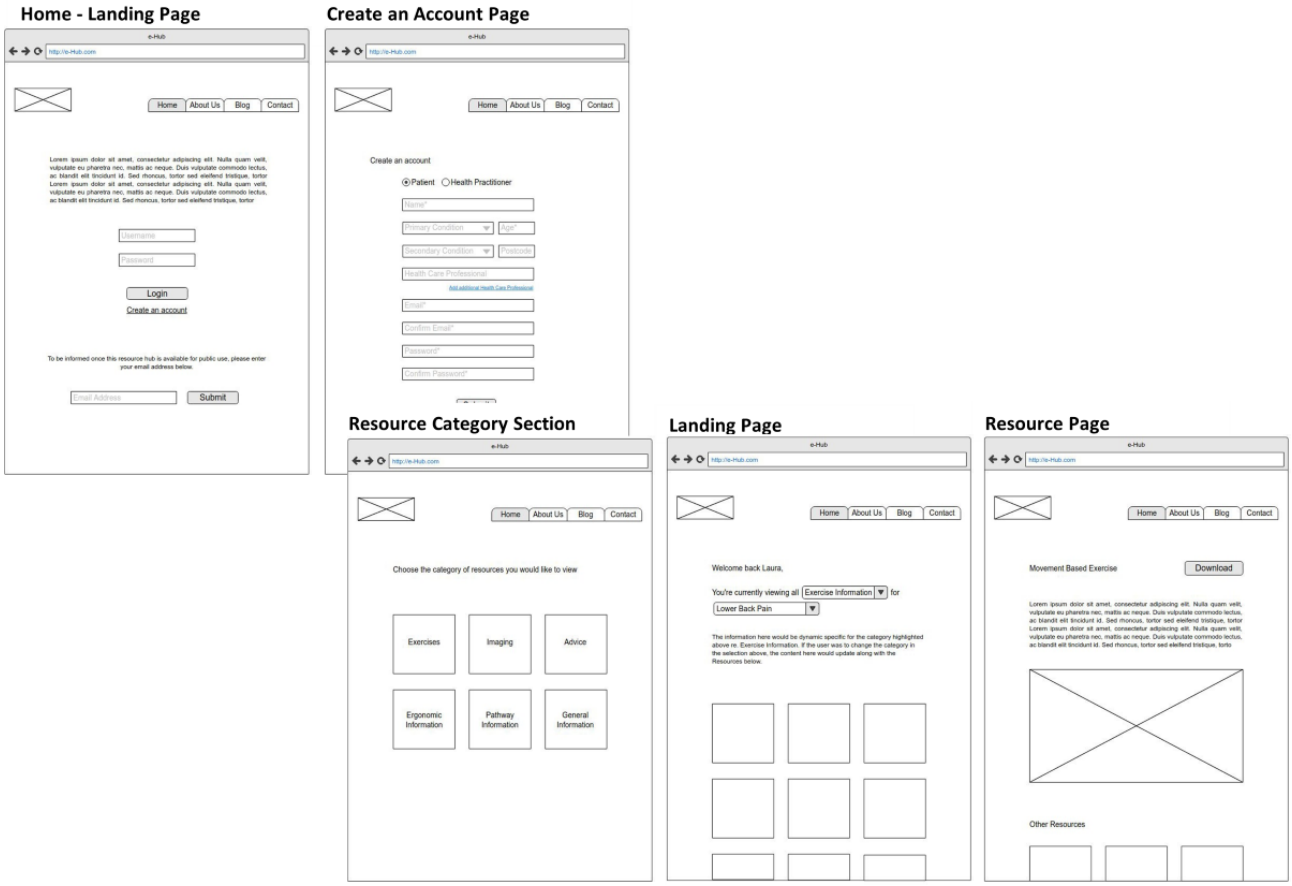
The sitemap is shown in Figure 1. Pages were kept simple to optimize ease of navigation (Figure 2).

Figure 1. Sitemap for the development of MyPainHub.

Sitemap	Page Layout
Home Page	Home Layout
About the Trial	Standard Layout
FAQs	Standard Layout
Contact	Contact Layout
Create Account	Create Account Layout
Pain Questionnaire	Pain Questionnaire Layout
Resource Category Selection	Resource Categories Layout
Resource Dashboard	Dashboard Layout
Individual Resource	Resource Layout
ORPHAN PAGES	
Privacy Policy	Standard Layout

XSL•FO
RenderX

Figure 2. Wireframes used in the development of MyPainHub were designed to be clear and simple to facilitate navigation and clarity of information for the user.



Permissions

The website was built on the website designer's content management system which was developed in a software platform in PHP (using the Symfony framework) that is based on standard customizable modules. A tiered permission system was used for MyPainHub. Administrators (ie, researchers) were able to access and amend all information on the website. HCPs were able to review all content on the website. Patients were able to view content related to their risk profile (ie, low risk or high risk) but they could see information for all conditions. However, once they identified their primary complaint during their registration process, their landing page was set to their primary condition. Only the researchers had access to patients' scores on the SF-ÖMPSQ and the Keele STarT MSK tool (during the PACE-MSK trial, the researchers would contact the patient's treating HCP to inform them of their patient's score [49]).

Stage 2: Implementation

Implementation of the website took place in 4 Australian states (New South Wales, Queensland, Western Australia, and Victoria) from February 2018 to December 2023 as a component of PACE-MSK, the clinical trial for which the website was built [49]. As part of PACE-MSK, 767 people with musculoskeletal conditions had consented to be part of the trial and were therefore required to register on MyPainHub. A total of 693 HCPs were involved in the PACE-MSK trial but registering on MyPainHub, whilst possible, was not a requirement of being involved in the trial. In addition to the PACE-MSK trial, the website was implemented within WSLHD (both patients and HCPs), with allied HCPs in occupational health settings and with students (Table 3).

The implementation strategy adopted influenced the number of participants who, despite consenting to participate, registered on the website (Table 3). As a result of the education via webinars and workshops, 100% of allied HCPs working in occupational settings, tertiary care settings and specialist HCPs who consented to participate registered on the website. However, despite being embedded in a unit of study in a university physiotherapy program, only 16% of students registered on the site. Passive strategies, such as phone calls or reminders, were ineffective with only 29% of HCPs working in primary care registering (Table 3).

Website data for a 12-month period (June 2021-June 2022) are presented in Table 4. The total page views of the website during the period were 23025 of which 12149 were unique page views. The pages with the highest unique views were general information about knee osteoarthritis, neck pain assessment (clinicians), frequently asked questions, LBP assessment (clinicians), general information about neck pain (general information), and exercise for knee osteoarthritis (Table 4). Users viewed on average 4 pages during a visit and stayed approximately 4 minutes on each page.

XSL•FO
RenderX

Table 3. Target markets of website users, implementation strategies, the numbers invited in each target market and percentage successfully registered.

Target market	Implementation strategy	N=invited	N=consented, (participants)	N=registered, n (%)
HCPs[a] in primary care	Reminders via phone calls and emails	693	196	57 (29)
Allied HCPs in occupational health settings	Discussion-based education (webinars) and reminders	48	6	6 (100)
Allied HCPs in tertiary care (WSLHD[b])	Interactive education (in-person workshop) and reminders	9	6	6 (100)
Specialist allied HCPs	Interactive education (in-person workshop and 2 web-based workshop reminders)	128	50	50 (100)
Student HCPs	Embedded classroom education	102	102	17 (17)
People with musculoskeletal conditions in primary care	Invitation via treating HCPs and research team (PACE-MSK trial)	767	767	767 (100)
People with musculoskeletal conditions in tertiary care (WSLHD)	Invitation via treating HCPs	18	11	11 (100)

[a]HCPs: health care professionals.

[b]WSLHD: Western Sydney Local Health District.

Table 4. Summary of the number and duration of page views for MyPainHub.

Page title	Unique page views, N	Average time (seconds)
Knee osteoarthritis: general information	657	42
Neck pain: assessment (clinicians)	408	41
Frequently asked questions	157	107
Lower back pain: assessment (clinicians)	128	88
Neck pain: general information	121	47
Knee osteoarthritis: exercise	99	29

Stage 3: Evaluation

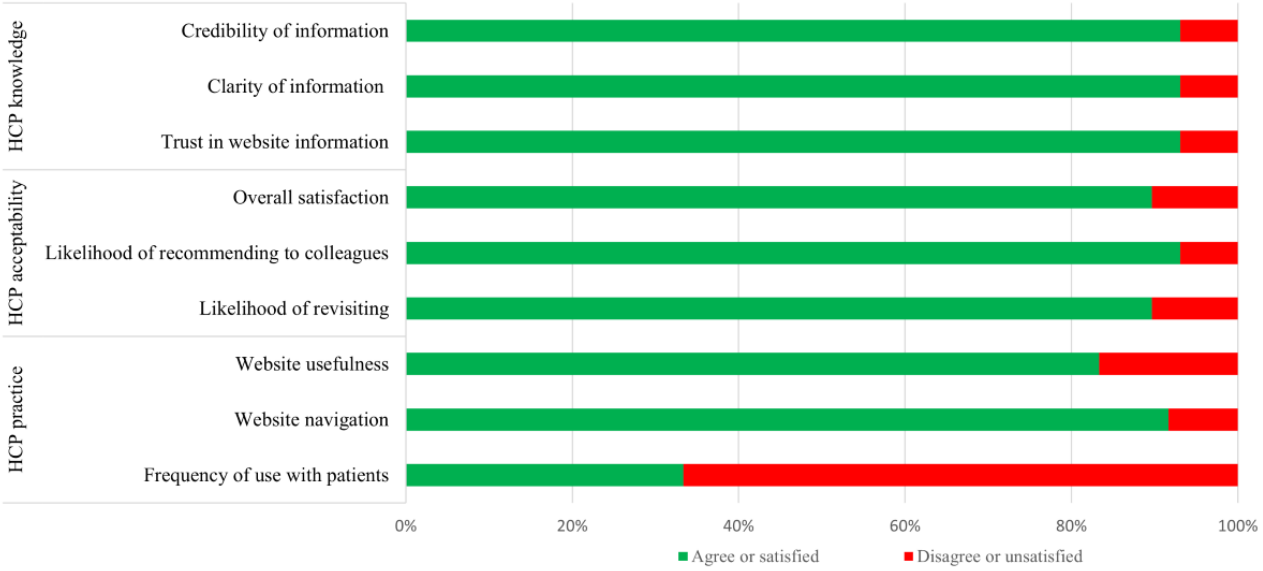
Allied HCPs (Students to Specialists)

Of the 102 allied HCPs invited, 29 (28%) HCPs completed the evaluation questionnaire. Of these, 93% (27/29) responded that MyPainHub had the potential to change knowledge as the information contained on the website was credible, trustworthy, clearly presented and easy to navigate. In terms of acceptability of the website, 93% (27/29) of HCP participants were satisfied with the website, were likely to revisit, and were likely to refer other colleagues to the site. However, less than 35% (10/29) of HCPs reported using the website with their patients. Suggestions for improvement included considering changing some of the information to enhance understanding for people with low health literacy and to include other musculoskeletal conditions (HCPs' opinions are shown in Figure 3).

From the focus group (6 HCP participants), four key themes emerged: (1) MyPainHub has utility as a web-based resource; (2) MyPainHub contains credible information, which reinforces best practice; (3) challenges in implementation; and (4) potential opportunities for enhancement.

XSL•FO
RenderX

Figure 3. Health care professionals' (HCP) opinions about MyPainHub. Opinions on knowledge and acceptability were positive, but this contrasted to how often HCPs used MyPainHub with patients.



Theme 1: MyPainHub Has Utility as a Web-Based Resource

HCP participants reported MyPainHub contained relevant information, was easy to navigate and improved professional confidence. HCPs supported the relevance and breadth of topics on the site and its utility in enhancing quality of care, "especially for junior staff or those with lesser experience." Despite challenges with implementation, participants had felt that it was useful for "encouraging patient engagement with HCPs," fostering an educational continuum, and "...builds on education and use of resources between sessions based on patient learning preferences." The "easy to navigate" resource had increased professional confidence as it "could help through processes, guide treatment and reduce fears when unsure or confused." This theme highlighted the website's potential in bridging knowledge gaps and facilitating a more informed patient-provider interaction.

Theme 2: MyPainHub Contains Credible Information, Which Reinforces Best Practice

The HCPs reported information on MyPainHub was credible and aligned with best practice, particularly because of the inclusion of video content and articles from expert clinicians and researchers. Comments regarding the information on MyPainHub "reinforced stuff previously learnt at university" and "reinforced what we already say (to patients) in a different format." Some remarked that the resource helped their own learning around best practice which in turn would influence the education they provided to their patients: "I don't have too much outpatient experience nor am I very confident. So I really like the clinician end of it...using that to understand things for myself and then being able to explain that to someone."

Theme 3: Challenges With Implementation

HCP participants reported implementation of MyPainHub was harder than expected and identified several challenges including internet availability in the clinic, time constraints, and a lack of familiarity with using web-based resources to support patient interactions. Acknowledging "change in practice is difficult to achieve in busy clinical settings," participants identified that "more training and preparation to support new workflows is needed." HCPs perceived those patients "may expect a more hands-on approach" which made them "hesitant" to adopt MyPainHub within sessions. As some HCPs were "not accustomed to using online resources" to aid their management of patients, they acknowledged they "often forgot about MyPainHub as an option," particularly in the context of "competing demands and heavy clinical workloads."

Theme 4: Potential Opportunities for Enhancement

Opportunities for improvement were noted, including the simplification of how information was presented, reduction of hyperlinks, and expansion of the number of conditions included on the site. HCPs voiced the need for the key messages and information to be "reduced" and "summarized" so that users would not have to "scroll as much." Some participants suggested that "more downloadable resources may be more helpful" and proposed more "ready-made printouts and resources of information."

People With Musculoskeletal Conditions

Eleven people with musculoskeletal conditions in the tertiary setting (WSLHD) completed the evaluation questionnaire. All responded that MyPainHub had the potential to change knowledge as the information contained on the website was credible, trustworthy and clearly presented. In terms of acceptability, all participants were satisfied with MyPainHub, likely to revisit and recommend the website to others. In addition, all participants indicated that the web-based resources helped them understand the best ways to manage their condition, the website was easy to navigate, and that MyPainHub would be a useful tool for other people with musculoskeletal conditions.

Four key themes emerged from the focus group (n=6 participants) from participants at WSLHD and these aligned

XSL•FO
RenderX

with the findings from the questionnaire: (1) MyPainHub adds to the understanding of the condition, (2) MyPainHub is credible as it aligns with HCP advice and information, (3) MyPainHub supports self-management, and (4) potential areas for improvement.

Theme 1: MyPainHub Adds to the Understanding of the Condition

Participants reported information on MyPainHub added to the understanding of their musculoskeletal condition. It was reported that the website offered clear explanations that were "definitely helpful," "adding knowledge and understanding," and "changing their perspectives on their condition." This aspect was further supported by the ability to "revisit information" via the "links" as needed, allowing for engagement with the content and facilitating ongoing learning.

Theme 2: MyPainHub Is Credible as it Aligns With HCPs' Advice and Information

Participants reported information on MyPainHub as credible, particularly when it aligned with their HCP's advice and information. This alignment reassured patients and "built trust" in both the website content and the advice provided by their HCPs. Trust was further reinforced when the website was recommended by their HCP or "experts."

Theme 3: MyPainHub Has the Potential to Support Self-Management

Participants reported the website increased their confidence in managing their condition. This was achieved by supporting "self-help and lifestyle changes" and "promoting exercise, and providing a holistic approach to physical activity and mental health." The ability to "support more activity and distract from negative thoughts" was reported, underlining the website's role in supporting both physical and psychological aspects of self-efficacy.

Theme 4: Potential Areas for Improvement

Participants identified areas for improvement. This included the need for more specific information tailored to "work or employment challenges," "detailed do's and don'ts for condition management," "preventative measures," "pain relief options," and "information about surgery and when it may be needed." "Regular updates" and the creation of "shared chat groups" were also suggested as potential areas for improvement.

Discussion

Principal Findings

This paper describes the process undertaken to develop, implement and evaluate a website designed to support patients with, and HCPs who manage, musculoskeletal conditions, specifically LBP, neck pain, WAD, and knee osteoarthritis. The findings suggest that following a research translation framework and incorporating extensive and iterative consultation with multiple key stakeholders resulted in MyPainHub being acceptable, credible, and easy to navigate with participants reporting they are likely to revisit and recommend the website to others. Embedding prognostic tools (the SF-ÖMPSQ and the Keele STarT MSK tool) allowed information to be tailored to a person's risk of poor outcome which may help both patients and HCPs better adopt guideline-based management. However, training for HCPs that includes more specific strategies on how to incorporate MyPainHub into clinical practice would be beneficial for more successful implementation.

A key component of the development of MyPainHub was the extensive involvement of a wide range of key stakeholders with different levels of experience and from different health care settings over a 4-year period. The initial development phase focused on gathering perspectives from people with musculoskeletal conditions and HCPs from primary care settings, researchers, key opinion leaders, professional associations, consumer groups and website developers. The latter development phase involved seeking perspectives of HCPs and patients from tertiary care settings, allied professionals from occupational settings and from allied health students. This broad stakeholder engagement reflects contemporary user-centered design and was undertaken to facilitate future wider implementation of MyPainHub across a range of settings and populations [61]. Whilst active implementation strategies (eg, interactive education and workshops) were generally more effective than passive strategies (eg, phone calls and emails), implementing MyPainHub in a unit of study was not successful amongst allied health students. In our study, students were introduced to MyPainHub during a lecture and were then responsible for designing a treatment program using MyPainHub but no specific tutorial time was allocated for this task. A previous study [43] that had implemented an innovative, guideline-based website, My Whiplash Navigator, found student HCPs had a higher uptake of the website than primary HCPs but the students were allocated time during both a lecture and a tutorial to using the resource which may explain the difference in the level of engagement.

The main purpose of this project was to develop a resource that would support guideline-based care for patients at low and high risk of poor outcome. Tailoring information to a person's risk of poor outcome afforded the opportunity to provide different information on likely recovery pathways and interventions that may be required. In this way, MyPainHub reinforces key guideline-based messages, that people at low risk should do well with minimal care but that people at high risk may require more comprehensive assessment of multiple domains and a multidisciplinary approach to optimize health outcomes. Greater adoption and wider implementation of MyPainHub may improve the efficiency of health service delivery by helping to educate people about musculoskeletal conditions and influence care decisions but requires further research.

Despite the opinions of stakeholders that MyPainHub is acceptable, credible, and easy to navigate, few used the tools with patients, and additional training that focuses on how best to integrate the website into clinical practice in different health care settings is likely to be required for greater adoption and usage. Active engagement strategies were more effective than passive strategies and taking a multifaceted approach, for example embedding MyPainHub in existing educational opportunities (eg, workshops, curriculum, and training programs) and using key opinion leaders to reinforce the credibility of information, is likely to be important in future

XSL•FO
RenderX

implementation [62,63]. The poor response rate and level of engagement of students could be overcome by integrating MyPainHub as a supplementary resource into the physiotherapy program curriculum and ensuring the information provided aligns with the learning objectives of the specific unit of study with which it is associated. Effective incorporation of MyPainHub for students may be achieved through thoughtful planning and implementation, integrating into specific lesson plans including activities such as group discussions followed by monitoring of student engagement via Learning Management Systems or linked to assessable content [64].

As with any web-based resource, updates to MyPainHub are ongoing with the maintenance of the website being the responsibility of the research team and the digital agency. Addition of information about other conditions (eg, hip pain) is underway and incorporating additional features such as more dynamic elements (eg, carousels and expandable sections) and gamification to increase engagement [65] are likely to be important features of future updates. Limitations of this study include that we did not control the time between participants viewing the website and asking for their feedback. Despite efforts to engage general practitioners and other medical professionals, the majority of those involved in the development of MyPainHub were allied HCPs. Further work is needed to determine whether MyPainHub, or specific sections of MyPainHub, would be useful for general practitioners. Further work is also required to implement MyPainHub in other primary (eg, community-based care) and tertiary care settings.

Conclusion

This study details the development, implementation, and evaluation of MyPainHub, a web-based resource designed to support guideline-based care for musculoskeletal conditions. Stakeholder engagement throughout the process contributed to its acceptability, credibility, and ease of use. While MyPainHub effectively tailors information based on prognostic tools to support both patients and health care providers, further efforts are needed to facilitate its integration into clinical practice. Training for health care professionals, particularly in how to apply MyPainHub in diverse settings, is likely essential for greater adoption.

Acknowledgments

We would like to thank the study participants for contributing their time and knowledge to this study.

This work was supported by The National Health and Medical Research Council (NHMRC) and Sydney Health Partners Implementation Science Academy. The funder had no role in the design or conduct of the study. TR was supported by an NHMRC Career Development Fellowship. KLB and PH are supported by NHMRC Investigator grants. MS receives unrestricted grant funding from the Motor Accident Insurance Commission of Queensland.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Healthcare professional survey.
[PNG File , 63 KB - formative_v9i1e63780_app1.png]

References

1. Briggs AM, Woolf AD, Dreinhöfer K, Homb N, Hoy DG, Kopansky-Giles D, et al. Reducing the global burden of musculoskeletal conditions. Bull World Health Organ 2018;96(5):366-368 [FREE Full text] [doi: 10.2471/BLT.17.204891] [Medline: 29875522]
2. Hoy W. Australian burden of disease study: impact and causes of illness and death in Australia 2011. In: Australian burden of disease study series. Canberra: Australian Institute of Health and Welfare; 2011.
3. Cieza A, Causey K, Kamenov K, Hanson SW, Chatterji S, Vos T. Global estimates of the need for rehabilitation based on the global burden of disease study 2019: a systematic analysis for the global burden of disease study 2019. Lancet 2021;396(10267):2006-2017 [FREE Full text] [doi: 10.1016/S0140-6736(20)32340-0] [Medline: 33275908]
4. El-Tallawy SN, Nalamasu R, Salem GI, LeQuang JAK, Pergolizzi JV, Christo PJ. Management of musculoskeletal pain: an update with emphasis on chronic musculoskeletal pain. Pain Ther 2021;10(1):181-209 [FREE Full text] [doi: 10.1007/s40122-021-00235-2] [Medline: 33575952]
5. de Luca K, Briggs AM, French SD, Ferreira ML, Cross M, Blyth F, et al. Disability burden due to musculoskeletal conditions and low back pain in Australia: findings from GBD 2019. Chiropr Man Therap 2022;30(1):22 [FREE Full text] [doi: 10.1186/s12998-022-00434-4] [Medline: 35505334]

XSL•FO
RenderX

6. Wallace J, Nwosu B, Clarke M. Barriers to the uptake of evidence from systematic reviews and meta-analyses: a systematic review of decision makers' perceptions. BMJ Open 2012;2(5):e001220 [FREE Full text] [doi: 10.1136/bmjopen-2012-001220] [Medline: 22942232]
7. Paci M, Faedda G, Ugolini A, Pellicciari L. Barriers to evidence-based practice implementation in physiotherapy: a systematic review and meta-analysis. Int J Qual Health Care 2021;33(2). [doi: 10.1093/intqhc/mzab093] [Medline: 34110410]
8. Oliveira CB, Maher CG, Pinto RZ, Traeger AC, Lin CC, Chenot J, et al. Clinical practice guidelines for the management of non-specific low back pain in primary care: an updated overview. Eur Spine J 2018;27(11):2791-2803. [doi: 10.1007/s00586-018-5673-2] [Medline: 29971708]
9. Malfliet A, Ickmans K, Huysmans E, Coppieters I, Willaert W, Bogaert WV, et al. Best evidence rehabilitation for chronic pain part 3: low back pain. J Clin Med 2019;8(7):1063 [FREE Full text] [doi: 10.3390/jcm8071063] [Medline: 31331087]
10. Blanpied PR, Gross AR, Elliott JM, Devaney LL, Clewley D, Walton DM, et al. Neck pain: revision 2017. J Orthop Sports Phys Ther 2017;47(7):A1-A83. [doi: 10.2519/jospt.2017.0302] [Medline: 28666405]
11. Sterling M, de Zoete RMJ, Coppieters I, Farrell SF. Best evidence rehabilitation for chronic pain part 4: neck pain. J Clin Med 2019;8(8):1219 [FREE Full text] [doi: 10.3390/jcm8081219] [Medline: 31443149]
12. Guidelines for the management of acute whiplash associated disorders for health professionals. 3rd edition. State Insurance Regulatory Authority (SIRA). 2014. URL: https://www.sira.nsw.gov.au/resources-library/motor-accident-resources/publications/for-professionals/whiplash-resources/SIRA08104-Whiplash-Guidelines-1117-396479.pdf [accessed 2023-06-23]
13. Kolasinski SL, Neogi T, Hochberg MC, Oatis C, Guyatt G, Block J, et al. 2019 American college of rheumatology/arthritis foundation guideline for the management of osteoarthritis of the hand, hip, and knee. Arthritis Rheumatol 2020;72(2):220-233 [FREE Full text] [doi: 10.1002/art.41142] [Medline: 31908163]
14. van Doormaal MCM, Meerhoff GA, Vliet Vlieland TPM, Peter WF. A clinical practice guideline for physical therapy in patients with hip or knee osteoarthritis. Musculoskeletal Care 2020;18(4):575-595. [doi: 10.1002/msc.1492] [Medline: 32643252]
15. Guideline for the management of knee and hip osteoarthritis. 2nd ed. East Melbourne, Vic: RACGP. The Royal Australian College of General Practitioners. 2018. URL: https://www.racgp.org.au/getattachment/71ab5b77-afdf-4b01-90c3-04f61a910be6/Guideline-for-the-management-of-knee-and-hip-osteoarthritis.aspx [accessed 2023-06-23]
16. Rice D, McNair P, Huysmans E, Letzen J, Finan P. Best evidence rehabilitation for chronic pain part 5: osteoarthritis. J Clin Med 2019;8(11):1769 [FREE Full text] [doi: 10.3390/jcm8111769] [Medline: 31652929]
17. Zadro J, O'Keeffe M, Maher C. Do physical therapists follow evidence-based guidelines when managing musculoskeletal conditions? Systematic review. BMJ Open 2019;9(10):e032329 [FREE Full text] [doi: 10.1136/bmjopen-2019-032329] [Medline: 31591090]
18. Bahns C, Happe L, Thiel C, Kopkow C. Physical therapy for patients with low back pain in Germany: a survey of current practice. BMC Musculoskelet Disord 2021;22(1):563 [FREE Full text] [doi: 10.1186/s12891-021-04422-2] [Medline: 34147077]
19. Logan GS, Dawe RE, Aubrey-Bassler K, Coombs D, Parfrey P, Maher C, et al. Are general practitioners referring patients with low back pain for CTs appropriately according to the guidelines: a retrospective review of 3609 medical records in Newfoundland using routinely collected data. BMC Fam Pract 2020;21(1):236 [FREE Full text] [doi: 10.1186/s12875-020-01308-5] [Medline: 33208086]
20. Epstein-Sher S, Jaffe D, Lahad A. Are they complying? Physicians' knowledge, attitudes, and readiness to change regarding low back pain treatment guideline adherence. Spine (Phila Pa 1976) 2017;42(4):247-252. [doi: 10.1097/BRS.0000000000001714] [Medline: 28207666]
21. Avoundjian T, Gidwani R, Yao D, Lo J, Sinnott P, Thakur N, et al. Evaluating two measures of lumbar spine MRI overuse: administrative data versus chart review. J Am Coll Radiol 2016;13(9):1057-1066. [doi: 10.1016/j.jacr.2016.04.013] [Medline: 27344246]
22. Williams CM, Maher CG, Hancock MJ, McAuley JH, McLachlan AJ, Britt H, et al. Low back pain and best practice care: a survey of general practice physicians. Arch Intern Med 2010;170(3):271-277. [doi: 10.1001/archinternmed.2009.507] [Medline: 20142573]
23. Buchbinder R, Underwood M, Hartvigsen J, Maher CG. The Lancet series call to action to reduce low value care for low back pain: an update. Pain 2020;161 Suppl 1(1):S57-S64 [FREE Full text] [doi: 10.1097/j.pain.0000000000001869] [Medline: 33090740]
24. Kharel P, Zadro JR, Maher CG. Physiotherapists can reduce overuse by Choosing Wisely. J Physiother 2021;67(3):151-155 [FREE Full text] [doi: 10.1016/j.jphys.2021.06.006] [Medline: 34154950]
25. Rutten GM, Degen S, Hendriks EJ, Braspenning J, Harting J, Oostendorp R. Adherence to clinical practice guidelines for low back pain in physical therapy: do patients benefit? Phys Ther 2010;90(8):1111-1122. [doi: 10.2522/ptj.20090173] [Medline: 20488978]
26. van Staalduinen DJ, van den Bekerom P, Groeneveld S, Kidanemariam M, Stiggelbout AM, van den Akker-van Marle ME. The implementation of value-based healthcare: a scoping review. BMC Health Serv Res 2022;22(1):270 [FREE Full text] [doi: 10.1186/s12913-022-07489-2] [Medline: 35227279]

XSL•FO
RenderX

27. Macpherson S, Rebbeck TR, Coates S, Evans K. Referral practices of recent graduate and experienced physiotherapists working in Australian primary care for people with musculoskeletal conditions. Musculoskelet Sci Pract 2023;64:102745 [FREE Full text] [doi: 10.1016/j.msksp.2023.102745] [Medline: 36924580]
28. Gorgon E, Maka K, Kam A, Nisbet G, Sullivan J, Regan G, et al. Needs assessment for health service design for people with back pain in a hospital setting: a qualitative study. Health Expect 2022;25(2):721-731 [FREE Full text] [doi: 10.1111/hex.13419] [Medline: 35150036]
29. Al Zoubi FM, Menon A, Mayo NE, Bussières AE. The effectiveness of interventions designed to increase the uptake of clinical practice guidelines and best practices among musculoskeletal professionals: a systematic review. BMC Health Serv Res 2018;18(1):435 [FREE Full text] [doi: 10.1186/s12913-018-3253-0] [Medline: 29884165]
30. Maher C, Sherrington C, Elkins M, Herbert R, Moseley A. Challenges for evidence-based physical therapy: accessing and interpreting high-quality evidence on therapy. Phys Ther 2004;84(7):644-654. [doi: 10.1093/ptj/84.7.644]
31. da Silva TM, Costa LDCM, Garcia AN, Costa LOP. What do physical therapists think about evidence-based practice? A systematic review. Man Ther 2015;20(3):388-401. [doi: 10.1016/j.math.2014.10.009] [Medline: 25458142]
32. Kummervold PE, Chronaki CE, Lausen B, Prokosch H, Rasmussen J, Santana S, et al. eHealth trends in Europe 2005-2007: a population-based survey. J Med Internet Res 2008;10(4):e42 [FREE Full text] [doi: 10.2196/jmir.1023] [Medline: 19017584]
33. Hodges PW, Setchell J, Nielsen M. An internet-based consumer resource for people with low back pain (MyBackPain): development and evaluation. JMIR Rehabil Assist Technol 2020;7(1):e16101 [FREE Full text] [doi: 10.2196/16101] [Medline: 32229467]
34. Scurlock-Evans L, Upton P, Upton D. Evidence-based practice in physiotherapy: a systematic review of barriers, enablers and interventions. Physiotherapy 2014;100(3):208-219. [doi: 10.1016/j.physio.2014.03.001] [Medline: 24780633]
35. Jette DU, Bacon K, Batty C, Carlson M, Ferland A, Hemingway R, et al. Evidence-based practice: beliefs, attitudes, knowledge, and behaviors of physical therapists. Phys Ther 2003;83(9):786-805. [doi: 10.1093/ptj/83.9.786]
36. Harding K, Porter J, Horne-Thompson A, Donley E, Taylor NF. Not enough time or a low priority? Barriers to evidence-based practice for allied health clinicians. J Contin Educ Health Prof 2014;34(4):224-231. [doi: 10.1002/chp.21255] [Medline: 25530292]
37. Reis FJJ, Nijs J, Parker R, Sharma S, Wideman TH. Culture and musculoskeletal pain: strategies, challenges, and future directions to develop culturally sensitive physical therapy care. Braz J Phys Ther 2022;26(5):100442 [FREE Full text] [doi: 10.1016/j.bjpt.2022.100442] [Medline: 36209626]
38. Costa N, Nielsen M, Jull G, Claus AP, Hodges PW. Low back pain websites do not meet the needs of consumers: a study of online resources at three time points. Health Inf Manag 2020;49(2-3):137-149. [doi: 10.1177/1833358319857354] [Medline: 31266366]
39. Bourne JA, Dziedzic K, Morris SJ, Jones PW, Sim J. Survey of the perceived professional, educational and personal needs of physiotherapists in primary care and community settings. Health Soc Care Community 2007;15(3):231-237. [doi: 10.1111/j.1365-2524.2007.00677.x] [Medline: 17444986]
40. Peterson S, Rainey N, Weible K. Who writes this stuff? Musculoskeletal information quality and authorship of popular health websites: a systematic review. Musculoskelet Sci Pract 2022;60:102563. [doi: 10.1016/j.msksp.2022.102563] [Medline: 35453015]
41. Wanless B, Bermingham C, Yeates N, O'Rahelly J. Creating a valuable patient information resource for MSK conditions: By not re-inventing the wheel. Physiotherapy 2021;113:e120-e121. [doi: 10.1016/j.physio.2021.10.106]
42. Translational research framework. Sax Institute. 2016. URL: https://www.medicalresearch.nsw.gov.au/app/uploads/2019/02/Translational-Research-Grants-Scheme-translation-research-framework.pdf [accessed 2023-07-20]
43. Bandong AN, Mackey M, Leaver A, Ingram R, Sterling M, Ritchie C, et al. An interactive website for whiplash management (My Whiplash Navigator): process evaluation of design and implementation. JMIR Form Res 2019;3(3):e12216 [FREE Full text] [doi: 10.2196/12216] [Medline: 31452515]
44. Clarkson P, Vassilev I, Rogers A, Brooks C, Wilson N, Lawson J, et al. Integrating a web-based self-management tool (Managing Joint Pain on the Web and Through Resources) for people with osteoarthritis-related joint pain with a web-based social network support tool (Generating Engagement in Network Involvement): design, development, and early evaluation. JMIR Form Res 2020;4(11):e18565 [FREE Full text] [doi: 10.2196/18565] [Medline: 33242011]
45. Hodges PW, Hall L, Setchell J, French S, Kasza J, Bennell K, et al. Effect of a consumer-focused website for low back pain on health literacy, treatment choices, and clinical outcomes: Randomized controlled trial. J Med Internet Res 2021;23(6):e27860 [FREE Full text] [doi: 10.2196/27860] [Medline: 34128822]
46. Wang X, Urban H, Bennell KL, Dickson C, Dobson F, Fransen M, et al. My Joint Pain, a web-based resource, effects on education and quality of care at 24 months. BMC Musculoskelet Disord 2020;21(1):79 [FREE Full text] [doi: 10.1186/s12891-020-3074-2] [Medline: 32028927]
47. van der Zee-Neuen A, Putrik P, Ramiro S, Keszei A, de Bie R, Chorus A, et al. Impact of chronic diseases and multimorbidity on health and health care costs: the additional role of musculoskeletal disorders. Arthritis Care Res (Hoboken) 2016;68(12):1823-1831 [FREE Full text] [doi: 10.1002/acr.22913] [Medline: 27111195]

XSL•FO
RenderX

48. Lin I, Wiles L, Waller R, Goucke R, Nagree Y, Gibberd M, et al. What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. Br J Sports Med 2020;54(2):79-86. [doi: 10.1136/bjsports-2018-099878] [Medline: 30826805]
49. Rebbeck T, Evans K, Ferreira P, Beales D, Sterling M, Bennell KL, et al. Implementation of a novel stratified PAthway of CarE for common musculoskeletal (MSK) conditions in primary care: protocol for a multicentre pragmatic randomised controlled trial (the PACE MSK trial). BMJ Open 2021;11(12):e057705 [FREE Full text] [doi: 10.1136/bmjopen-2021-057705] [Medline: 37039086]
50. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
51. Baiardini I, Braido F, Bonini M, Compalati E, Canonica GW. Why do doctors and patients not follow guidelines? Curr Opin Allergy Clin Immunol 2009;9(3):228-233. [doi: 10.1097/ACI.0b013e32832b4651] [Medline: 19390434]
52. Spangenberg JH, Heong KL, Klotzbücher A, Klotzbücher T, Nguyen QA, Tekken V, et al. Doing what with whom? Stakeholder analysis in a large transdisciplinary research project in South-East Asia. Paddy Water Environ 2018;16(2):321-337. [doi: 10.1007/s10333-018-0634-2]
53. Kang K, Evans K, Simic M, Ferreira P, Bandong AN, Coates S, et al. Impact of an interactive workshop on specialist physiotherapists' practice when implementing a new clinical care pathway for people with musculoskeletal conditions. Musculoskelet Sci Pract 2022;57:102466. [doi: 10.1016/j.msksp.2021.102466] [Medline: 34742051]
54. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
55. Linton SJ, Boersma K. Early identification of patients at risk of developing a persistent back problem: the predictive validity of the Örebro Musculoskeletal Pain Questionnaire. Clin J Pain 2003;19(2):80-86. [doi: 10.1097/00002508-200303000-00002] [Medline: 12616177]
56. Campbell P, Hill J, Protheroe J, Afolabi E, Lewis M, Beardmore R, et al. Keele aches and pains study protocol: validity, acceptability, and feasibility of the Keele STarT MSK tool for subgrouping musculoskeletal patients in primary care. J Pain Res 2016;9:807-818 [FREE Full text] [doi: 10.2147/JPR.S116614] [Medline: 27789972]
57. Ritchie C, Hendrikz J, Jull G, Elliott J, Sterling M. External validation of a clinical prediction rule to predict full recovery and ongoing moderate/severe disability following acute whiplash injury. J Orthop Sports Phys Ther 2015;45(4):242-250. [doi: 10.2519/jospt.2015.5642] [Medline: 25827122]
58. Sterling M, Ritchie C, Rebbeck T, Cameron ID, Griffin A, Jagnoor J, et al. Comparison of the accuracy of WhipPredict to that of a modified version of the Short-Form Örebro Musculoskeletal Pain Screening Questionnaire to predict poor recovery after whiplash injury. J Orthop Sports Phys Ther 2021;51(5):207-215. [doi: 10.2519/jospt.2021.9987] [Medline: 33607916]
59. Suri P, Delaney K, Rundell SD, Cherkin DC. Predictive validity of the STarT Back Tool for risk of persistent disabling back pain in a U.S. primary care setting. Arch Phys Med Rehabil 2018;99(8):1533-1539.e2 [FREE Full text] [doi: 10.1016/j.apmr.2018.02.016] [Medline: 29625095]
60. Fuhro F, Fagundes F, Manzoni A, Costa L, Cabral C. Örebro Musculoskeletal Pain Screening Questionnaire Short-Form and STarT Back Screening Tool: Correlation and agreement analysis. Spine (Phila Pa 1976) 2016;41(15):E931-E936. [doi: 10.1097/BRS.0000000000001415] [Medline: 26720177]
61. Moore G, Wilding H, Gray K, Castle D. Participatory methods to engage health service users in the development of electronic health resources: systematic review. J Particip Med 2019;11(1):e11474 [FREE Full text] [doi: 10.2196/11474] [Medline: 33055069]
62. Rebbeck T, Maher CG, Refshauge KM. Evaluating two implementation strategies for whiplash guidelines in physiotherapy: a cluster randomised trial. Aust J Physiother 2006;52(3):165-174. [doi: 10.1016/s0004-9514(06)70025-3] [Medline: 16942451]
63. Goorts K, Dizon J, Milanese S. The effectiveness of implementation strategies for promoting evidence informed interventions in allied healthcare: a systematic review. BMC Health Serv Res 2021;21(1):241 [FREE Full text] [doi: 10.1186/s12913-021-06190-0] [Medline: 33736631]
64. Kenwright D, Dai W, Osborne E, Gladman T, Gallagher P, Grainger R. “Just tell me what I need to know to pass the exam!” Can active flipped learning overcome passivity? TAPS 2017;2(1):1-6. [doi: 10.29060/TAPS.2017-2-1/OA1007]
65. Alfieri FM, da Silva Dias C, de Oliveira NC, Battistella LR. Gamification in musculoskeletal rehabilitation. Curr Rev Musculoskelet Med 2022;15(6):629-636 [FREE Full text] [doi: 10.1007/s12178-022-09797-w] [Medline: 36301514]

Abbreviations

HCP: health care professional
LBP: low back pain
MSK: muscoluskeletal
PACE: PAthway of CarE
RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance
SF-ÖMPSQ: Short Form Örebro Musculoskeletal Pain Screening Questionnaire

XSL•FO
RenderX

WAD: whiplash associated disorders
WSLHD: Western Sydney Local Health District

Edited by A Mavragani; submitted 05.07.24; peer-reviewed by N Khan, J Pearson; comments to author 09.10.24; revised version received 18.11.24; accepted 28.12.24; published 24.02.25.

Please cite as:
Evans K, Ko J, Ceprnja D, Maka K, Beales D, Sterling M, Bennell KL, Jull G, Hodges PW, McKay MJ, Rebbeck TJ
Development and Implementation of MyPainHub, a Web-Based Resource for People With Musculoskeletal Conditions and Their Health Care Professionals: Mixed Methods Study
JMIR Form Res 2025;9:e63780
URL: https://formative.jmir.org/2025/1/e63780
doi:10.2196/63780
PMID:

©Kerrie Evans, Jonathan Ko, Dragana Ceprnja, Katherine Maka, Darren Beales, Michele Sterling, Kim L Bennell, Gwendolen Jull, Paul W Hodges, Marnee J McKay, Trudy J Rebbeck. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Feasibility of Large Language Models in Verbal Comprehension Assessment: Mixed Methods Feasibility Study

Dorit Hadar-Shoval[1], BA, MA, PhD; Maya Lvovsky[1], BA; Kfir Asraf[1], BA, MA; Yoav Shimoni[2,3], BA, MA, PHD; Zohar Elyoseph[4,5], BA, MA, PHD

[1]Department of Psychology, Max Stern Academic College of Emek Yezreel, Afula, Israel

[2]School of Psychology, University of Haifa, Haifa, Israel

[3]Department of Education and Psychology, Open University of Israel, Raanana, Israel

[4]Department of Brain Sciences, Faculty of Medicine, Imperial College London, London, United Kingdom

[5]School of Counseling and Human Development, Faculty of Education, University of Haifa, Haifa, Israel

Corresponding Author:
Zohar Elyoseph, BA, MA, PHD
Department of Brain Sciences
Faculty of Medicine
Imperial College London
Reynolds Building, St Dunstan's Road
Charing Cross Hospital, Hammersmith
London, W6 8RP
United Kingdom
Phone: 972 547836088
Email: Zohar.j.a@gmail.com

Abstract

Background: Cognitive assessment is an important component of applied psychology, but limited access and high costs make these evaluations challenging.

Objective: This study aimed to examine the feasibility of using large language models (LLMs) to create personalized artificial intelligence–based verbal comprehension tests (AI-BVCTs) for assessing verbal intelligence, in contrast with traditional assessment methods based on standardized norms.

Methods: We used a within-participants design, comparing scores obtained from AI-BVCTs with those from the Wechsler Adult Intelligence Scale (WAIS-III) verbal comprehension index (VCI). In total, 8 Hebrew-speaking participants completed both the VCI and AI-BVCT, the latter being generated using the LLM Claude.

Results: The concordance correlation coefficient (CCC) demonstrated strong agreement between AI-BVCT and VCI scores (Claude: CCC=.75, 90% CI 0.266-0.933; GPT-4: CCC=.73, 90% CI 0.170-0.935). Pearson correlations further supported these findings, showing strong associations between VCI and AI-BVCT scores (Claude: r=.84, P<.001; GPT-4: r=.77, P=.02). No statistically significant differences were found between AI-BVCT and VCI scores (P>.05).

Conclusions: These findings support the potential of LLMs to assess verbal intelligence. The study attests to the promise of AI-based cognitive tests in increasing the accessibility and affordability of assessment processes, enabling personalized testing. The research also raises ethical concerns regarding privacy and overreliance on AI in clinical work. Further research with larger and more diverse samples is needed to establish the validity and reliability of this approach and develop more accurate scoring procedures.

(JMIR Form Res 2025;9:e68347) doi:10.2196/68347

KEYWORDS

large language models; verbal comprehension assessment; artificial intelligence; AI in psychodiagnostics; personalized intelligence tests; verbal comprehension index; Wechsler Adult Intelligence Scale; WAIS-III; psychological test validity; ethics in computerized cognitive assessment

XSL•FO
RenderX

Introduction

Cognitive assessment is one of the core skills of applied psychologists, and it is considered one of the most important contributions of scientific psychology to clinical practice [1]. Intelligence tests are widely used to measure individuals' cognitive abilities. They provide a standardized measure of a person's intellectual capacities, aiding in educational placement and career selection, and identifying potential cognitive impairments. Intellectual assessment measures serve as valuable tools in psychology, education, and research, helping identify gifted individuals, track cognitive development, and evaluate the effectiveness of interventions [1]. Intelligence tests have their limitations [2] but their consistent use and refinement over time have contributed to our understanding of human intelligence and its variations. Traditionally, the development of an intelligence test involves an intricate process and the construction of age- and gender-based norms based on extensive data collected from a sizable sample representative of the population [3].

Intelligence tests are designed to assess various cognitive abilities, such as logical reasoning, problem-solving skills, spatial awareness, and verbal comprehension [4]. The verbal comprehension index (VCI) is known for its strong correlation with the g factor (the hypothesized construct of general intelligence) [5], and it measures the capability to use language for reasoning, comprehension, and understanding social norms. The VCI includes three subtests, that are, similarities, vocabulary, and information [5].

The significance of a comprehensive psychological assessment including measures of verbal intelligence cannot be overstated. [6-10]. The assessment of verbal intelligence also extends to personality traits crucial for interpersonal interactions, such as the recognition of emotions and the ability to express oneself effectively [11]. The comprehensive psychological evaluation of various populations under different conditions requires a nuanced understanding of verbal intelligence.

The administration of intelligence tests and the subsequent interpretation of the results are typically undertaken by psychologists. These professionals usually possess at least a master's degree in psychology or in a related field, which qualifies them to conduct the tests and analyze the results accurately and ethically. Intellectual testing is usually administered in a standardized form, where the individuals being tested are presented with similar tasks under similar conditions. The standardized administration process helps ensure an ethical process and produces valid results, leading to accurate interpretations [1,10].

The field of intelligence testing still heavily relies on conventional paper-and-pencil tests, despite recent efforts to embrace technological advancements. Some current intelligence testing measures, such as the fifth edition of the Wechsler Intelligence Scale for Children [12], presuppose the use of tablets by clinicians to improve the administration and scoring procedures. Globally, however, the use of technology in psychological assessment is still developing and to date has had relatively modest influence.

At the end of 2022, Open AI launched ChatGPT-3.5, a generative artificial intelligence (GenAI) product based on a large language model (LLM) trained on abundant text data and capable of generating human-like responses to text-based inputs [13]. This open access application set a record as the fastest-growing consumer application in history [14], making significant contributions to academia [13], psychology [14-27], medicine [28], and programming [29]. In March 2023, Anthropic launched Claude, another GenAI product based on LLM, with a similar interface to that of ChatGPT but with Hebrew language capabilities that are considered to be more nuanced and accurate than those of ChatGPT.

While these technological advances offer promising opportunities for psychological assessment, it is essential to acknowledge their current limitations. LLMs cannot replace comprehensive clinical evaluations, as they lack the ability to observe crucial behavioral cues, interpret complex clinical contexts, or make nuanced professional judgments. Therefore, any application of AI in cognitive testing must be viewed as a potential supportive tool requiring professional oversight.

This pilot study represents an innovative endeavor in the field of GenAI and its application in psychological assessment, serving as a proof-of-concept investigation into the feasibility of using LLMs for verbal comprehension testing. Rather than aiming to validate these tools across a large population, our primary objective was to demonstrate whether LLMs can, in principle, generate personalized verbal comprehension assessments that produce results comparable to standardized tests. Central to this study is the development of a personalized cognitive test comprising 30 questions without human norms (artificial intelligence [AI]–based verbal comprehension test [AI-BVCT]). Unlike standard testing procedures, this approach does not rely on pre-established diagnostic tools but uses the algorithmic capabilities of the LLM to encode, assess, and calculate verbal intelligence scores.

A critical aspect of this study is the comparative analysis conducted between the AI-BVCT results and those obtained from the VCI, under the supervision of an expert psychologist. To establish the basic viability of this approach, we conducted a focused comparison between AI-generated tests and traditional assessments on a small sample. We hypothesized that the differences between the AI-BVCT and traditional VCI scores would be within half an (SD 7.5), indicating a high level of accuracy and that the correlation between the AI-BVCT and the VCI would be high, demonstrating accurate cognitive ranking across participants. By focusing exclusively on verbal comprehension abilities, the study aims to establish proof of concept for how LLMs can contribute to psychological testing, emphasizing the potential for new personalized and dynamic testing methodologies.

Methods

Overview

This study used 2 LLMs: GPT-4 (OpenAI) and Claude (Anthropic). Data collection occurred in September 2023. The selection of these models was based on several key

XSL•FO
RenderX

considerations. At the time of the study, these models represented the most advanced capabilities in natural language processing, particularly for the Hebrew language. Claude demonstrated superior proficiency in Hebrew language processing and a nuanced understanding of cultural context, making it our primary choice for generating the AI-BVCT. Both models were used to evaluate participants' responses, allowing for a comparison between different approaches to language model development. The use of models from different leading companies (OpenAI and Anthropic) enabled us to examine various approaches to language model development and training, providing a broader context for our findings. It is important to note that this selection limits the study to Western commercial models and does not include open-source models or those developed in other cultural contexts. This limitation will be discussed further in the paper. We did not alter their default settings when using these models. We used the standard chat interface provided by the companies, without special adjustments to hyperparameters such as temperature or top-k. An additional limitation of these tools is their ownership by commercial entities, which restricts full transparency of their training and development processes. It is also worth noting that the field of LLMs is rapidly evolving, and more advanced models have become available since the study was conducted.

Participants

This pilot study included 8 native Hebrew-speaking volunteer participants (4 females), aged 28.5 (SD 2.13) years. The participants were recruited through random sampling, facilitated by a research assistant who distributed targeted recruitment notices to undergraduate and graduate psychology students at Max Stern Yezreel Valley College. The recruitment process extended over a 2-month period, during which participants volunteered without compensation. The recruitment notices specifically outlined the study's objectives and requirements, ensuring informed participation. All participants who initially enrolled completed the study in its entirety, with no withdrawals recorded during the research period. The participants were selected from the student population due to their accessibility and their presumed average to above-average cognitive abilities, which were deemed appropriate for this pilot study. However, we acknowledge the limitations of this small, nonrepresentative sample, which are addressed in detail in the limitations section of this paper.

Research Design and Procedure

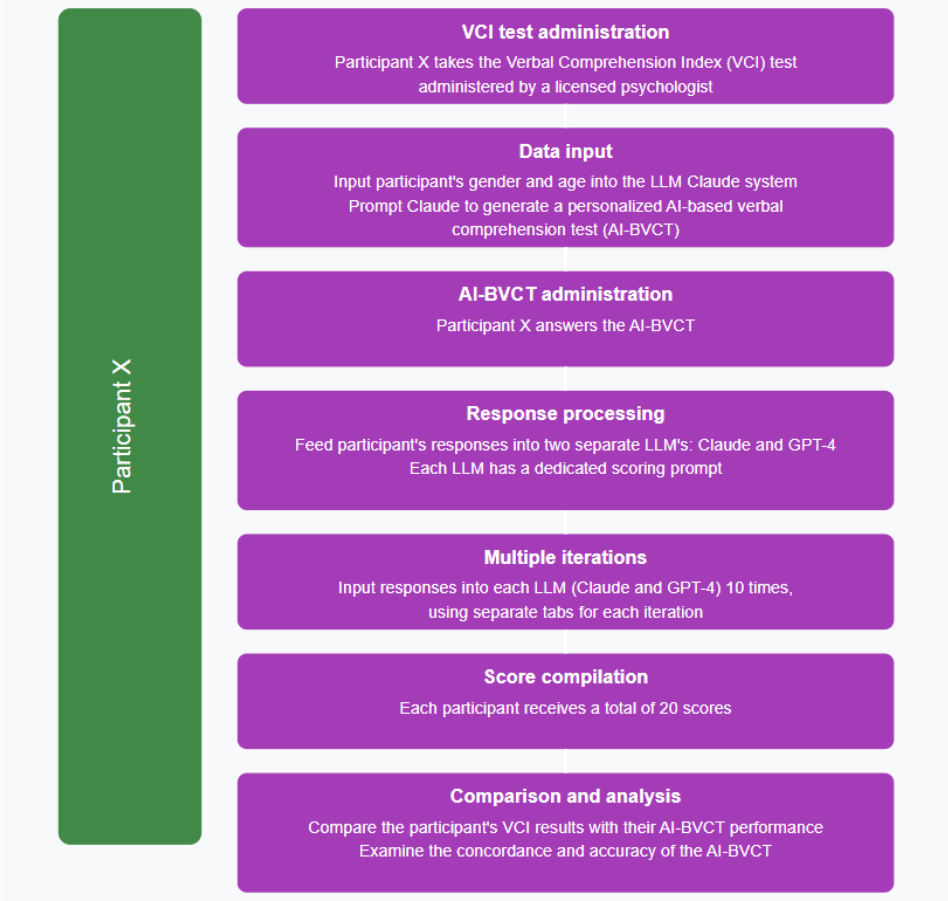
The study used a within-participants design, with each participant undergoing 2 cognitive tests: the VCI from the WAIS-III, and a personalized AI-BVCT generated by the LLM Claude system.

After each participant underwent the VCI test, the participant's age and gender were entered into the LLM Claude system, together with a prompt to generate a personalized AI-BVCT to measure the participant's verbal intelligence. In the next stage, the participant was required to respond to the AI-BVCT. The participant's responses were then fed into 2 separate LLMs, Claude and GPT-4, with a prompt containing scoring instructions. To ensure the reliability of the scoring across different runs, each participant's responses were entered into each of the 2 models 10 times, using a separate tab for each iteration. As a result, each participant received a total of 20 scores (10 from each model) on the AI-BVCT.

Finally, we compared the participant's results on the VCI test and their performance on the AI-BVCT to examine the degree of concordance and accuracy of the test developed by the LLM system. A schematic representation of the study procedure is provided in Figure 1.

XSL•FO
RenderX

Figure 1. Schematic representation of the study procedure for assessing verbal comprehension using AI-based tests in a proof-of-concept study of Hebrew-speaking university students (N=8) in Israel, September 2023. AI-BVCT: artificial intelligence-based verbal comprehension tests; LLM: large language models; VCI: verbal comprehension index.



Materials

The VCI

Overview

The VCI is an index designed to assess verbal abilities, such as verbal concept formation and reasoning skills. This index is part of the Wechsler Adult Intelligence Scale-3 (WAIS-III; [30]), which measures various cognitive abilities in adults. VCI scores are normalized with a mean of 100 and an SD of 15 [31]. The WAIS-III was used in this study as it represents the most recent version of the WAIS-III that has been officially translated, adapted, and normalized for Hebrew-speaking populations in Israel. While newer versions exist, they have not yet been standardized for Hebrew-speaking populations.

The index consists of three subtests: (1) similarities, which measures the ability to identify commonalities between terms; (2) vocabulary, which measures the understanding and use of words; and (3) information, which measures general knowledge and the ability to provide factual information [5]. The VCI score is calculated by combining the 3 subtests.

VCI Scoring

Scoring was performed according to the accepted scoring manual. Each question was assigned a score of 0-2 based on the acceptable responses provided in the scoring guide. The calculation of standard scores and results was conducted in accordance with the guidelines established in the WAIS manual.

Personalized AI-Based Verbal Comprehension Test

Generating the Personalized AI-BVCT

The LLM Claude (Anthropic, Ltd) was used for generating a personalized cognitive test for each participant. We used Claude because at the time the study was conducted its performance in Hebrew was considered superior to those of other LLMs, such as GPT-4 or Bard. To generate the AI-BVCT, we composed a request for Claude to develop a 30-item verbal intelligence test, having provided the participant's age and gender. The inclusion of age information was intended to ensure age-appropriate content and difficulty level, similar to standardized tests. Gender information was provided to account for the gendered nature of the Hebrew language in question phrasing, although the content and difficulty of questions were not intended to differ based on gender. To enhance the quality of the test, we also provided examples of questions to guide Claude in shaping questions with the appropriate structure and difficulty level. Specifically, we provided 3 sample items for each of the 3 intelligence levels, which consist of low (<70), medium (90-110), and high (>125) for each of the subtests.

Personalized AI-BVCT Prompt

The prompt that was entered into Claude are shown in Textbox 1.

XSL•FO
RenderX

Textbox 1. Prompt for constructing a verbal intelligence test with artificial intelligence.

Cognitive Assessment

- Act as a professor of psychology specializing in psychodiagnostics and psychometrics. Please create a 30-question test that can assess the verbal intelligence score of an X-year-old Hebrew-speaking man or woman. Try to create questions whose answers will be as similar as possible to the results of the verbal intelligence WAIS 3 test in the verbal intelligence index, which includes a vocabulary test (defining words at increasingly difficult levels), common denominator (finding a semantic common denominator between words at increasingly difficult levels), and general knowledge (general knowledge at increasingly difficult levels). Construct the questions so that they can give as accurate a score as possible in the range of 50-150. This means that some questions should be very easy and some very difficult. Build questions that can be answered via the chat box, without any human intervention, and without questions requiring additional tools other than the chat itself.

To refine your ability to build questions at different difficulty levels, you are given examples.

- Vocabulary: (1) Here are some words for which the level of difficulty to define for a 30-year-old adult is hard and reflects a higher than average verbal intelligence (verbal intelligence level 125 and above): obelisk, orthography, ether, allegory, atom. (2) Here are some words for which the level of difficulty to define for a 30-year-old adult is medium and reflects average verbal intelligence (verbal intelligence level 90-110): fortification, settlement, takeover, degeneration, outsmarting. (3) Here are some words for which the level of difficulty to define for a 30-year-old adult is easy and reflects much lower-than-average verbal intelligence (verbal intelligence level 70 and below: table, chair, cupboard, sofa, television
- SimList Paragraphilarities: (1) Here are some words for which the level of difficulty to find a common denominator for a 30-year-old adult is hard and reflects higher than average verbal intelligence (verbal intelligence level 125 and above): clothes and food, dream and reality, life and death, light and darkness, sky and earth. (2) Here are some words for which the level of difficulty to find a common denominator for a 30-year-old adult is medium and reflects average verbal intelligence (verbal intelligence level 90-110): wave and wind, freedom and imprisonment, land and sea, inside and outside, number and letter. (3) Here are some words for which the level of difficulty to find a common denominator for a 30-year-old adult is easy and reflects much lower than average verbal intelligence (verbal intelligence level 70 and below). Fork and knife, ball and football, rain and snow, dog and cat, pants and shirt.
- General Knowledge: (1) Here are some questions that reflect high general knowledge of a 30-year-old adult reflecting higher than average verbal intelligence (verbal intelligence level 125 and above): Who painted “The Scream”? Who wrote the Odyssey? In which century did Beethoven live? In which sea are the Maldives Islands located? What is the biological component that converts RNA to protein? (2) Here are some questions that reflect the medium general knowledge of a 30-year-old adult reflecting average verbal intelligence (verbal intelligence level 90-110): Who was the first king of Israel according to the Bible? Who invented the telephone? In which country is the famous Himalayan Mountain, Everest, located? In which year did World War II end? Who was Nelson Mandela? (3) Here are some questions that reflect the low general knowledge of a 30-year-old adult reflecting much lower than average verbal intelligence (verbal intelligence level 70 and below): What is the vehicle that moves over water and carries people? On which Israeli holiday is it customary to light candles? What animal is the king of animals? What is the name of the first month of the Gregorian calendar? What shape do most balls have?
- Now build the test to assess verbal intelligence. The examples provided to you are meant to reflect different levels of difficulty to help you build a test sensitive to a score range of 50-150. You don't need to build the questions in the same structure as the examples (common denominator, vocabulary, and knowledge).

The AI-BVCT Scoring

After the participants completed the AI-BVCT, their responses to the test questions were entered into 2 LLMs, Claude and GPT-4, for evaluation and scoring. The LLMs were instructed to review and grade the participants’ responses to each of the 30 questions, assigning a score ranging from 0 to 2 for each answer (0=incorrect, 1=partial answer, and 2=full answer), making the total possible score range 0-60. Given the assumption that the test has a normal distribution of difficulty levels, correctly answering half the test questions (15 correct answers score of 30) is considered equivalent to a verbal IQ score of 100, correctly answering all test questions (30 correct answers, a raw score of 60) indicates a verbal IQ score of 150. Consequently, each point above or below 30 in the AI-BVCT corresponds to an IQ change of 1.67 points. For example, the IQ score of participants whose AI-BVCT was determined to be 31 would be 101.67. To increase reliability and avoid random errors, each participant’s responses were entered into each of the LLM 10 times.

The scoring process sought to address 2 main challenges. The first is that the LLMs struggle to perform numeric summation consistently. The second concerns certain cases where the LLMs incorrectly classify a correct response as incorrect or an incorrect response as correct. To address these challenges, we took two steps: (1) an MA student in psychology manually scored each participant’s responses to the AI-BVCT and (2) the automatic scoring process was supervised by the researchers regarding two key criteria in Textbox 2.

XSL•FO
RenderX

Textbox 2. Correction criteria for artificial intelligence–based scoring errors.

- In cases where there was an error in summing the number of correct responses into a total score [For example, in a case where a participant correctly answered 10 questions (2 points each) and partially answered 10 questions (1 point each), the large language model (LLM) erroneously calculated a total score of 29 instead of the correct 30], the LLMs were instructed: "You have summed the cases incorrectly, please re-sum."
- In extreme cases where the LLMs assigned a score of 2 when the human scorer assigned a score of 0, or vice versa [For example, when asked to define "ambivalence," a participant responded: "Ambivalence is the simultaneous existence of conflicting attitudes or emotions towards an object or situation." The human rater scored this as fully correct (2 points), but the LLM incorrectly assigned 0 points.], we instructed the LLMs: "You have made an error on question number X, classifying it as incorrect when it is correct. Please adjust the score accordingly." While this correction process was necessary for this proof-of-concept study, we acknowledge that such iterative adjustments could potentially introduce bias. Future implementations should consider automated validation methods and more structured scoring protocols to minimize human intervention.

Prompt for Reviewing the Personalized AI-BVCT

The prompt used is outlined in Textbox 3.

Textbox 3. Prompt for reviewing and scoring artificial intelligence–based verbal comprehension test (AI-BVCT) responses.

(Insert the participant's responses to the test questions here)

Your task now is to create a table in the following order

- In column 1, write the question and its number.
- In column 2, write the participant's answer.
- In column 3, carefully review each question based on your knowledge. Cross-reference different information sources and score according to the following key: a correct answer receives a score of 2. A partial or not entirely accurate answer receives a score of 1. Also, give a score of 1 for a slight inaccuracy. An entirely incorrect answer or a statement from the participant that they do not know the answer receives a score of 0. Ignore spelling errors in the scoring.
- In column 4, indicate the cumulative score accumulated throughout the questions.
- Below the table, write how many correct and partially correct answers there were out of 30 and calculate the intermediate score by accurately summing the scores obtained in step c. Caution: Please ensure the sum is correct before submitting the answer.
- Perform an additional calculation, this time counting the number of errors separately and noting it.
- Verify that the number of correct, partially correct, and incorrect answers adds up correctly to a total of 30 questions.
- After verification, indicate the final intermediate score at the end-this should be the score calculated in section e.

Prompt for Scoring the Personalized AI-BVCT

The following prompts were used:

> *(Insert the intermediate score obtained from the AI-BVCT here)*
>
> *Now you need to convert the obtained score to a new score. First, create a heading: "Final Score Calculation"*
>
> *Calculation method: If the participant received an intermediate score of 30 by answering 15 questions correctly with 2 points each, they will receive a score of 100. For each point above the intermediate score of 30, add 1.67 points (a participant who received an intermediate score of 31 will receive a final score of 101.67). The same applies in reverse for scores below 15 correct answers; they will receive below 100 in a similar ratio.*

Ethical Considerations

This study was approved by the institutional review board of Max Stern Yezreel Valley College (approval number: YVC 2023-66). All participants provided written informed consent after receiving a comprehensive explanation of the research procedures and objectives. Data were collected anonymously and stored securely with access restricted to the research team. All identifying information was removed from responses before analysis. Participants received course credit as compensation for their participation. No personally identifiable information or images were included in the study materials or findings.

Statistical Analysis

To assess the agreement between the AI-BVCTs and the VCI scores, we computed the intraclass correlation coefficient (ICC) and Bland-Altman limits of agreement (LoA; a method to assess agreement between 2 different measurement techniques by plotting the differences against the averages of the 2 measures). We also calculated the concordance correlation coefficient (CCC) to evaluate the agreement between the AI-BVCTs and the VCI scores (CCC measures the degree to which pairs of observations fall on the 45° line through the origin, indicating perfect agreement). We used Pearson correlations to examine the association between the VCI score, AI-BVCT Claude 2 scores, and AI-BVCT GPT-4 scores. The reliability of the AI-BVCTs was assessed using internal consistency (Cronbach α).

To test the hypothesis that the differences between the AI-BVCTs and VCI scores would be bounded within equivalence bounds of 15 points, we conducted Two One-Sided Tests (TOST) and paired-samples *t* test (TOST, used to establish

XSL•FO
RenderX

statistical equivalence between 2 measures). We compared the proportion of AI-BVCT Claude 2 and GPT-4 scores falling within the 95% WAIS CI using a *z* test.

Finally, we used a generalized mixed model to investigate the differences between the AI-BVCT Claude 2 scores and GPT-4 scores and their interaction with the VCI scores. We conducted post hoc comparisons to examine the differences between the LLMs at various levels of VCI (mean, SD 1) using Jamovi (version 2.3.28).

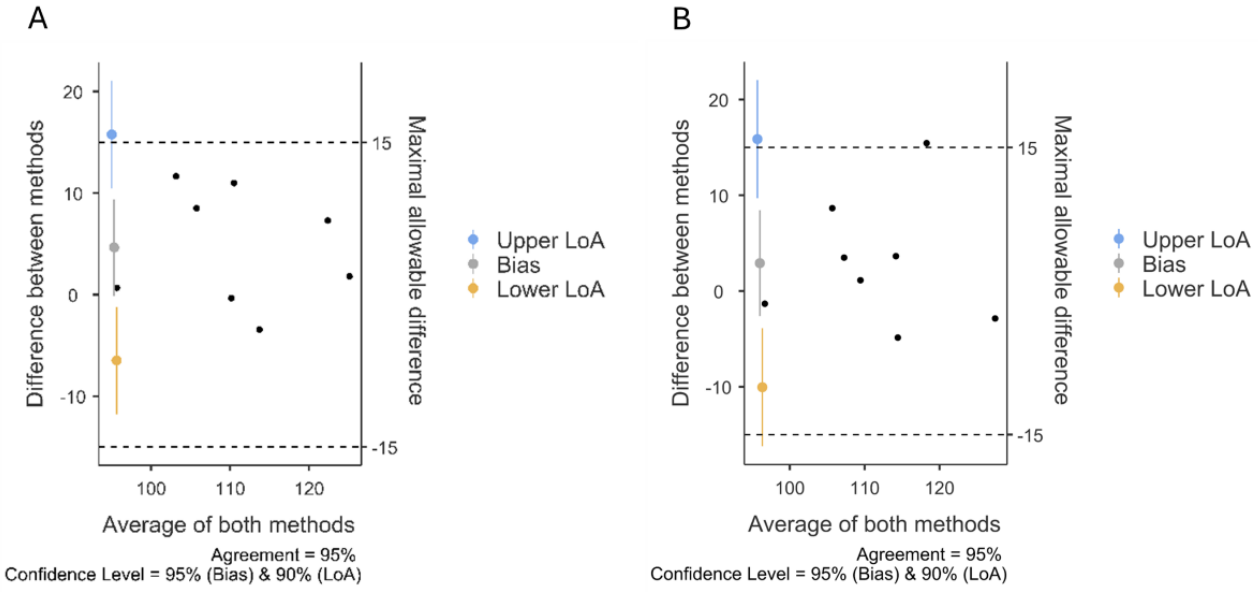
Results

Preliminary Analysis

First, we examined the reliability of the 10 AI-BVCTs generated by each LLM. Both LLMs had high internal consistency (AI-BVCT Claude 2: α=.988, AI-BVCT GPT-4: α=.986). Next, we computed the ICC to assess the agreement between the AI-BVCTs and VCI scores. The ICC was 0.961 (95% CI 0.872-0.991; 2-way mixed, average measures, absolute agreement), which is considered excellent reliability [32].

We then proceed to calculate Bland-Altman LoA (Figure 2) and the CCC.

Figure 2. Bland-Altman plots demonstrating agreement between VCI scores and AI-generated verbal comprehension test scores in Hebrew-speaking university students (N=8): A comparison of traditional and AI-based assessment methods. Study conducted in Israel, September 2023. LoA: limits of agreement.



For AI-BVCT Claude 2, the Bland-Altman LoA indicated that the mean bias (4.64) was not significantly different from 0 (95% CI –0.10 to 9.38), the lower LoA was –6.48 (90% CI –11.78 to 1.18), and the upper LoA was 15.76 (90% CI 10.47-21.06). The CCC was 0.752 (90% CI 0.266-0.933), which is considered an almost excellent agreement.

For AI-BVCT GPT-4, the Bland-Altman LoA indicated that the mean bias (2.91) was not significantly different from 0 (95% CI –2.61 to 8.43), the lower LoA was –10.04 (90% CI –16.21 to –3.87), and the upper LoA was 15.86 (90% CI 9.70-22.03). The CCC was 0.733 (90% CI 0.170-0.935), which is considered an almost excellent agreement [33].

Following these results, we found high Pearson correlations between the VCI score and AI-BVCT Claude 2 scores (r=0.844, P<.001), the VCI score and AI-BVCT GPT-4 scores (r=0.771, P=.02), and AI-BVCT Claude 2 and AI-BVCT GPT-4 scores (r=0.866, P=.01).

In sum, the verbal IQ-like scoring of the AI-BVCTs generated by the LLMs was found to be reliable.

Hypothesis Testing

We hypothesized that the differences between the AI-BVCTs and VCI scores, if any, would be bounded within equivalence bounds of 15 points (half an SD of an IQ score), and examined it using a TOST paired-samples *t* test (Figure 3).

No difference was found between Claude 2 AI-BVCT scores (108.48, SD 10.43) and the VCI score (113.12, SD 9.78; t_7=–2.31, P=.05). The mean difference between the VCI and AI-BVCT scores was –4.64, with a 90% CI of –10.65 to 1.37. The 2 one-sided tests were statistically significant, with the score difference significantly higher than the lower bound of –15 (t_7=5.16, P<.001), and the score difference significantly lower than the upper bound of 15 (t_7=–9.78, P<.001).

No difference was found between the GPT-4 AI-BVCT scores (110.21, SD 9.75) and the VCI score (113.12, SD 9.78; t_7=–1.24, P=.25). The mean difference between the VCI and AI-BVCT scores was –2.91, with 90% CI of –9.91 to 4.09. The 2 one-sided tests were statistically significant, with the score difference significantly higher than the lower bound of –15 (t_7= 5.17,

XSL•FO
RenderX

P<.001), and the score difference significantly lower than the upper bound of 15 (t_7=–7.66, P<.001).

In conclusion, the IQ scores derived from the AI-BVCTs can be seen as equivalent on average to the VCI IQ scores.

Next, we examined whether the scores of the AI-BVCTs were within the range of the 95% WAIS CIs (Table 1). 5 out of 8 AI-BVCT Claude 2 scores (62.5%) and 75% of AI-BVCT GPT-4 scores were within the confidence intervals. The 2 proportions were significantly different (z=–0.53, P=.59).

Figure 3. Comparison of VCI scores with AI-BVCT scores from two language models (Claude and GPT-4) in Hebrew-speaking university students (N=8), Israel, September 2023, demonstrating score equivalence within predetermined bounds (SD 15 points). AI-BVCT: artificial intelligence-based verbal comprehension tests.

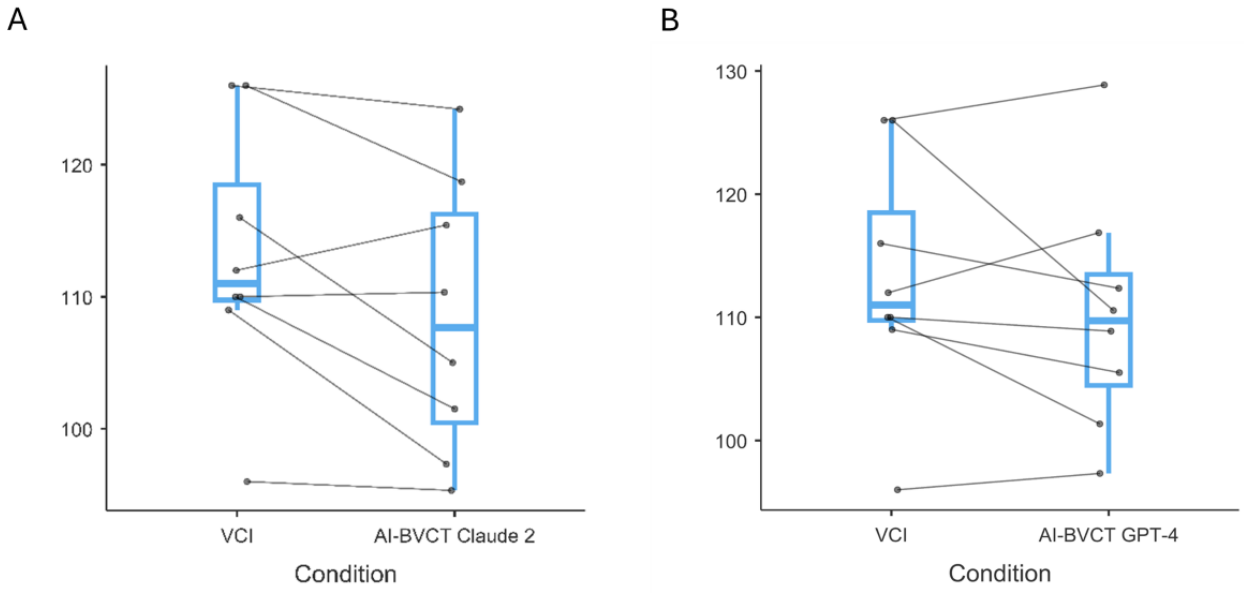


Table 1. Comparison of individual VCIs[a] with artificial intelligence–generated test scores: analysis of score agreement within WAIS[b] CIs in Hebrew-Speaking University Students (N=8) from Northern Israel, September 2023.

	VCI score	WAIS (95% CI)	AI-BVCT[c] GPT-4 scores		AI-BVCT Claude 2 scores	
			Scores	X[d] or V[e]	Scores	X or V
1	126	119-131	110.55	X	118.70	X
2	109	103-114	105.51	V	97.32	X
3	112	106-117	116.86	V	115.43	V
4	100	94-106	101.34	V	101.50	V
5	116	110-121	112.35	X	105.00	X
6	110	104-115	108.87	V	110.34	V
7	126	119-131	128.86	V	124.21	V
8	96	91-102	97.32	V	95.33	V
Total	—[f]	—	—	6/8 (75%)	—	5/8 (62.5%)

[a]VCI: verbal comprehension index.

[b]WAIS: Wechsler Adult Intelligence Scale.

[c]AI-BVCT: artificial intelligence–based verbal comprehension tests.

[d]X: Score outside the WAIS 95% CI.

[e]V: Score within the WAIS 95% CI.

[f]Not applicable.

Differences in AI-BVCT Scores Between the LLMs

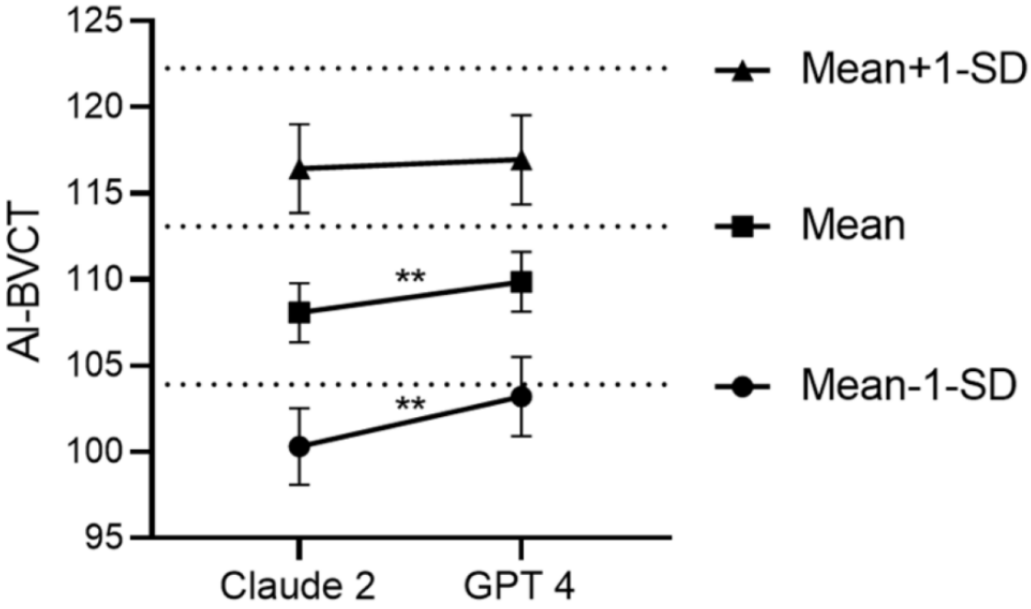
Based on these findings, we examined the differences between the 2 LLMs and their interaction with the VCI score using a generalized mixed model (Figure 4). The 2 LLM scores were significantly different (χ^2_1=7.2, P<.001), with AI-BVCT GPT-4 having a higher mean score (109.86, SD 1.73) than AI-BVCT Claude 2 (108.08, SD 1.71). The difference between the LLMs interacted with the VCI score (χ^2_1=3.9, P=.04), so at 1 SD below

XSL•FO
RenderX

the mean (103.90), GPT-4 (103.18, SD 2.28) estimated the AI-BVCT score significantly higher than Claude 2 (100.31, SD 2.21; z=–3.08, P=.01). At the mean (113.08), GPT-4 (109.86, SD 1.74) estimated the AI-BVCT score significantly higher than Claude 2 (108.07, SD 1.72; z=–2.68, P=.01) but at 1 SD above the mean (122.27), GPT-4 (116.97, SD 2.58) was not significantly different than Claude 2 (116.44, SD 2.57; z=–0.53, P=.59).

These results suggest that Claude 2 underestimated and GPT-4 more closely estimated the true VCI scores, except for the higher VCI scores, where no difference was found between the LLMs. Both LLMs underestimated the true VCI scores at the mean and 1 SD above the mean levels. These results should be interpreted cautiously because the sample size was small and individuals with lower IQ than the mean were underrepresented.

Figure 4. Differential performance of language models (Claude vs GPT-4) in VCI estimation across intelligence levels in Hebrew-speaking university students (N=8), Israel, September 2023. GPT-4 demonstrated superior estimation accuracy at both -1 SD below mean and mean levels. AI-BVCT: artificial intelligence-based verbal comprehension tests. **p<0.001.



Discussion

Principal Findings

The present proof of concept study successfully demonstrated the potential feasibility of using LLMs to generate AI-based verbal comprehension tests. While the sample size was small, as appropriate for an initial feasibility investigation, the strong agreement between AI-BVCT and VCI scores provides compelling preliminary evidence that LLMs can generate psychologically meaningful assessments. However, it is important to note that these findings are based on a student population with above-average intelligence, and the application of this approach to broader populations requires careful consideration. The cognitive profile of university students differs systematically from the general population, particularly from clinical populations or individuals with below-average intelligence, and these differences might affect how LLMs interpret and score responses.

The findings showed a high ICC between the AI-BVCT and the VCI scores, indicating that the AI-generated test effectively discriminated between individuals with different verbal intelligence levels, ranking them similarly to the VCI. A paired-sample t test found no significant differences between the AI-BVCT scores and the VCI scores, demonstrating that the AI-generated test was properly calibrated to yield scores comparable to the traditional VCI. Finally, AI-BVCT scores fell within the 95% CI of the VCI scores, providing additional evidence that the AI-generated scores were consistent with what would be expected based on the well-established VCI measure. These preliminary results suggest that LLMs show promise in differentiating between individuals with varying verbal intelligence levels and produce scores that appear to correlate with the traditional VCI test. Further research with larger and more diverse samples is needed to fully establish the reliability and validity of this approach.

The findings of the present proof of concept study provide preliminary evidence that LLMs have the potential to contribute to the field of psychodiagnostics. The ability to accurately assess verbal skills is associated with the diagnosis of significant clinical states [6-10,27]. The results suggest that we may be on the cusp of a transformative moment in psychological assessments, a field that has not been substantially affected by technological advancements. The demonstrated capacity of LLMs to “understand” and process language at a high level holds immense promise for the diagnostic process.

Nevertheless, the process of coding and scoring responses still requires significant human involvement. We found that at times LLMs have difficulty with determining whether a particular answer is correct or incorrect and with accurately adding up scores. Specifically, we identified 2 main types of errors in using LLMs for scoring verbal intelligence tests. The primary issue was with summing up the scores accurately, which occurred frequently. In addition, there were a few instances where the LLMs did not provide correct scoring, although these

XSL•FO
RenderX

cases were not numerous. These challenges highlight the current limitations of LLMs in performing precise numerical calculations and occasional misinterpretations in scoring criteria. This situation requires a "human-in-the-loop" for supervision and correction. The need for human oversight is crucial in verifying the accuracy of score summations and ensuring the correct application of scoring criteria. In the future, with improvements in prompting techniques or the development of more powerful models, these issues may diminish. Potential improvements could include enhancing the LLMs' numerical computation abilities, refining the prompts to emphasize careful score tallying, and developing more robust error-checking mechanisms within the AI system. These advancements could significantly reduce the frequency of calculation errors and improve the overall accuracy of automated scoring. At present, human involvement remains essential to ensure the accuracy of the process.

In addition, we observed differences in scoring patterns between Claude 2 and GPT-4, with GPT-4 providing estimates closer to true VCI scores, especially for lower VCI scores. However, both LLMs underestimated scores at the mean and higher levels. While these differences may stem from varying capabilities in handling linguistic nuances and contextual understanding, particularly in Hebrew, the exact reasons remain unclear and may be temporary due to ongoing model updates. This finding highlights the importance of comparative analyses between LLM platforms in AI-based assessments. Future research should investigate these differences with larger, more diverse samples to ensure reliability across platforms. This underscores the need for ongoing monitoring and evaluation in the development of AI-assisted psychological assessment tools.

The preliminary findings of this study have several potential implications for the field of psychodiagnostics. One of the most significant is the potential of AI-based cognitive tests to make assessment processes more widely accessible and affordable. Traditional cognitive assessments are often expensive and require administration by trained psychologists, limiting their availability to a relatively small subset of the population, primarily those with suspected clinical difficulties. In the future AI-based cognitive tests may be easily generated and administered without the need for a psychologist's involvement, significantly reducing the associated costs. This could make cognitive assessments more readily available to a broader range of individuals, beyond those with clinical indications. The adaptability of AI-based tests could facilitate their translation and adaptation for use in different cultural and linguistic settings, further expanding their reach [15,22,27].

LLMs enable personalized testing tailored to each individual's unique characteristics and needs. This adaptive approach allows for the dynamic selection of item difficulty based on the examinee's performance, ensuring that the test is optimally challenging and informative. Such personalized testing can be particularly beneficial in cases that require repeated assessments over time, such as monitoring cognitive decline [34]. By generating novel items for each administration, AI-based tests can minimize practice effects and provide a more accurate measure of an individual's cognitive abilities.

The flexibility of AI-generated tests opens up new possibilities for expanding our understanding of cognitive functioning across diverse populations. Large-scale modeling of population distributions for various cognitive attributes may eventually enable the creation of customized norms that consider a wide range of factors, beyond age and gender. For example, norms could be tailored to specific learning disabilities, socioeconomic backgrounds, cultural contexts, and more [35]. This level of customization would allow for a more nuanced and comprehensive assessment of an individual's strengths and weaknesses. As LLMs continue to advance, we will likely see an increasing integration of these technologies into psychological assessment practices, leading to a new era of personalized and data-driven cognitive evaluation.

Despite the many positive implications demonstrated in this preliminary pilot study, consideration must also be given to the potential risks involved. Most significantly, the present pilot demonstrates the capability of LLMs to perform the technical aspects of test administration, such as test generation, coding, and scoring. Other aspects, like evaluating the context created during the assessment, identifying additional factors through clinical observation such as body language, or discerning problem-solving patterns, were not demonstrated in the current study. It remains uncertain whether LLMs will be able to perform these functions in the near future. In other words, the current research showed the ability to arrive at a similar score as a professional assessment but did not seek to identify or predict factors that may resolve or exacerbate difficulties.

Ethical Implications and Safeguards

The potential integration of AI-based cognitive assessment tools into psychological practice, even at this preliminary stage, raises important ethical considerations. Drawing from established professional guidelines, including those of major psychological associations, several key ethical domains require attention as this technology develops: protection of individual rights, professional oversight, prevention of misuse, concerns regarding epistemic authority, and current technological limitations [36].

First, privacy and data protection must be paramount [37]. While AI-based assessments could increase accessibility, they also create risks of unauthorized data collection. Professional psychological testing committees emphasize the need for (1) secure data handling protocols, (2) clear consent procedures specifically addressing AI involvement, and (3) strict limitations on data access and retention. These foundational requirements would need further development and validation in larger-scale studies.

Second, the risk of commercial misuse must be addressed through appropriate safeguards [38]. Based on our preliminary findings, key protections should include (1) explicit prohibition of covert cognitive profiling, (2) requirements for professional oversight, and (3) clear limitations on acceptable use cases. While these initial recommendations provide a framework, comprehensive protocols would need to be developed through extensive consultation with professional bodies and validation in clinical settings.

XSL•FO
RenderX

Third, professional standards and oversight are essential. Initial recommendations include maintaining clear boundaries between commercial applications and clinical assessment, requiring professional interpretation of results, and establishing guidelines for identifying and mitigating potential harm [39]. Regular consultation with senior clinical and neuropsychological professionals would be crucial in developing these standards as both the technology and its applications evolve.

Fourth, there is significant concern about overreliance on AI systems, which may be regarded as "all-knowing" with their recommendations taken as absolute [22,27]. This relates to the concept of epistemic authority, which is particularly concerning when working with vulnerable populations such as those with mental health challenges [40]. Professional guidelines emphasize that AI systems should serve only as supportive tools, with ultimate judgment residing with qualified clinicians. This requires (1) clear documentation of AI limitations, (2) training for professionals on appropriate AI integration, and (3) maintenance of critical evaluation practices.

Fifth, our findings highlight important limitations in current AI capabilities. While this study demonstrated the potential for automating technical aspects like test administration and scoring, critical clinical functions, such as observing nonverbal behavior, evaluating problem-solving patterns, and understanding broader context — remain beyond AI capabilities. In addition, the use of public AI models raises specific privacy concerns, as these systems are often trained on user data. Any clinical implementation would require dedicated, secure systems with proper privacy safeguards. Professional guidelines emphasize that validation must go beyond simple score comparison to include a comprehensive assessment of clinical utility and potential risks [27].

The implementation of these safeguards requires ongoing collaboration between AI developers, mental health professionals, and regulatory bodies. Regular consultation with senior clinical, neuropsychological, and educational psychologists will be essential for updating and refining these protocols as both AI capabilities and potential risks continue to evolve. Further research and broader professional consultation will be crucial to establish comprehensive ethical guidelines that protect individual rights while allowing beneficial applications of this emerging technology.

Study Limitations and Further Studies

While this pilot study included only 8 participants, this sample size was appropriate for our primary aim of demonstrating proof of concept. The strong correlation between AI-generated and traditional test scores, even in this small sample, provides compelling initial evidence for the feasibility of this approach. However, as this study was designed to demonstrate basic viability rather than population-level validation, further research with larger and more diverse samples is required to fully establish the viability of AI-generated testing. Moreover, the small sample was relatively homogeneous in education level and did not represent the full variation in the population. The intelligence examined was high, therefore the study did not take into account lower levels of intelligence. This represents range restriction, and it is possible that at lower levels the results would be different. Further research with larger and more diverse samples is required to fully establish the viability of AI-generated testing. Moreover, the study included a normative population, and future research should also examine populations with diverse cognitive disabilities (eg, learning disability). Another limitation is that the study was conducted in Hebrew; future research needs to examine the capabilities of LLMs in other languages as well.

Overall, these findings provide preliminary support for the use of LLMs in generating valid verbal intelligence assessments without reliance on normative data. The personalized approach allowed the creation of an individualized test tailored to the participant's age and gender background. This study was the first to examine the capabilities of AI in cognitive assessment and intelligence testing as a proof of concept. Given the pilot nature of the research, many additional studies are needed to further validate these initial findings in both the general population and in diverse clinical groups. As a preliminary investigation in this domain, the study helps lay the foundation for psychological testing by AI and serves as a proof of concept**.** Yet, considerably more research is required to fully characterize the use of AI for psychological testing across populations. This pilot study alone is insufficient to draw definitive conclusions, underscoring the need for rigorous continued inquiry in this nascent area. The findings expand the existing knowledge regarding what we know about AI in several aspects, including the ability to create a test and diagnose medical and psychological aspects. It represents an important first step in harnessing LLMs for innovative advances in psychological assessment. The ability to automate test development and scoring shows great promise in making testing more efficient, accessible, and personalized. As LLMs continue to evolve in sophistication, their integration can meaningfully augment and potentially transform standard practices in the field.

Conclusion

This preliminary study explores the potential of LLMs in generating personalized verbal intelligence tests. While our initial findings are promising, they highlight both opportunities and challenges in using AI for psychological assessments. Further research with larger, more diverse samples is necessary to validate these results and address the ethical considerations of integrating AI into psychological testing practices.

Acknowledgments

During the preparation of this manuscript, the authors used Claude 3.5 Sonnet to assist with English language refinement and sentence clarity. All output was carefully reviewed and edited by the authors, who take full responsibility for the content of this manuscript. This use of AI was limited to language enhancement and did not contribute to the scientific analysis or conclusions.

XSL•FO
RenderX

Data Availability

The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request. Access is restricted due to privacy considerations and ethical requirements regarding participant confidentiality. The institutional ethics committee will review requests for access to anonymized data which may be granted for research purposes that align with the informed consent provided by participants.

Conflicts of Interest

None declared.

References

1. Groth-Marnat G. Handbook of Psychological Assessment. Hoboken, NJ: Wiley & Sons; 2009.
2. Reynolds CR, Altmann RA, Allen DN. The problem of bias in psychological assessment. In: Mastering Modern Psychological Testing. Heidelberger Platz: Springer; 2021:573-613.
3. Norfolk PA, Farmer RL, Floyd RG, Woods IL, Hawkins HK, Irby SM. Norm block sample sizes: a review of 17 individually administered intelligence tests. J Psychoeduc Assess 2015;33(6):544-554. [doi: 10.1177/0734282914562385]
4. Climie EA, Rostad K. Test review: wechsler adult intelligence scale. J Psychoeduc Assess 2011;29(6):581-586. [doi: 10.1177/0734282911408707]
5. Lange RT, Lippa SM. Verbal comprehension index. In: Encyclopedia of Clinical Neuropsychology. Cham, Switzerland: Springer International Publishing; 2017:1-3.
6. Boyle R, Knight SP, De Looze C, Carey D, Scarlett S, Stern Y, et al. Verbal intelligence is a more robust cross-sectional measure of cognitive reserve than level of education in healthy older adults. Alzheimers Res Ther 2021;13(1):128 [FREE Full text] [doi: 10.1186/s13195-021-00870-z] [Medline: 34253231]
7. Henry JD, Crawford JR, Phillips LH. Verbal fluency performance in dementia of the alzheimer's type: a meta-analysis. Neuropsychologia 2004;42(9):1212-1222. [doi: 10.1016/j.neuropsychologia.2004.02.001] [Medline: 15178173]
8. Pauli L, Daseking M, Petermann F, Stemmler M. [Associations between cognitive performance in a dementia screening test (SKT) and an intelligence test (WAIS IV) : which deficits in cognitive performance in old age indicate a possible pathological deterioration process?]. Z Gerontol Geriatr 2018;51(3):266-274. [doi: 10.1007/s00391-017-1263-x] [Medline: 28600610]
9. Cornoldi C, Giofrè D, Orsini A, Pezzuti L. Differences in the intellectual profile of children with intellectual vs. learning disability. Res Dev Disabil 2014;35(9):2224-2230. [doi: 10.1016/j.ridd.2014.05.013] [Medline: 24927516]
10. Maseda A, Lodeiro-Fernández L, Lorenzo-López L, Núñez-Naveira L, Balo A, Millán-Calenti JC. Verbal fluency, naming and verbal comprehension: three aspects of language as predictors of cognitive impairment. Aging Ment Health 2014;18(8):1037-1045. [doi: 10.1080/13607863.2014.908457] [Medline: 24797556]
11. Bertie BL, Huang HW. Cognition and beyond: intersections of personality traits and language. Psychol Learn Motiv 2024;80:105-148. [doi: 10.1016/bs.plm.2024.03.003]
12. Na SD, Burns TG. Wechsler intelligence scale for children-v: test review. Appl Neuropsychol Child 2016;5(2):156-160. [doi: 10.1080/21622965.2015.1015337] [Medline: 25923224]
13. Rudolph J, Tan S, Tan S. ChatGPT: bullshit spewer or the end of traditional assessments in higher education? JALT 2023;6(1):342-363. [doi: 10.37074/jalt.2023.6.1.9]
14. Hu K. ChatGPT Sets Record for Fastest-Growing User Base - Analyst Note. New York: Reuters; 2023.
15. Elyoseph Z, Gur T, Haber Y, Simon T, Angert T, Navon Y, et al. An ethical perspective on the democratization of mental health with generative AI. JMIR Ment Health 2024;11:e58011 [FREE Full text] [doi: 10.2196/58011] [Medline: 39417792]
16. Elyoseph Z, Hadar Shoval DH, Levkovich I. Beyond personhood: ethical paradigms in the generative artificial intelligence era. Am J Bioeth 2024;24(1):57-59. [doi: 10.1080/15265161.2023.2278546] [Medline: 38236857]
17. Elyoseph Z, Hadar-Shoval D, Asraf K, Lvovsky M. ChatGPT outperforms humans in emotional awareness evaluations. Front Psychol 2023;14:1199058 [FREE Full text] [doi: 10.3389/fpsyg.2023.1199058] [Medline: 37303897]
18. Elyoseph Z, Levkovich I. Beyond human expertise: the promise and limitations of ChatGPT in suicide risk assessment. Front Psychiatry 2023;14:1213141 [FREE Full text] [doi: 10.3389/fpsyt.2023.1213141] [Medline: 37593450]
19. Elyoseph Z, Levkovich I. Comparing the perspectives of generative AI, mental health experts, and the general public on schizophrenia recovery: case vignette study. JMIR Ment Health 2024;11:e53043 [FREE Full text] [doi: 10.2196/53043] [Medline: 38533615]
20. Elyoseph Z, Levkovich I, Shinan-Altman S. Assessing prognosis in depression: comparing perspectives of AI models, mental health professionals and the general public. Fam Med Community Health 2024;12(Suppl 1):e002583 [FREE Full text] [doi: 10.1136/fmch-2023-002583] [Medline: 38199604]
21. Elyoseph Z, Refoua E, Asraf K, Lvovsky M, Shimoni Y, Hadar-Shoval D. Capacity of generative AI to interpret human emotions from visual and textual data: pilot evaluation study. JMIR Ment Health 2024;11:e54369 [FREE Full text] [doi: 10.2196/54369] [Medline: 38319707]

XSL•FO
RenderX

22. Haber Y, Levkovich I, Hadar-Shoval D, Elyoseph Z. The artificial third: a broad view of the effects of introducing generative artificial intelligence on psychotherapy. JMIR Ment Health 2024;11:e54781 [FREE Full text] [doi: 10.2196/54781] [Medline: 38787297]
23. Hadar-Shoval D, Asraf K, Mizrachi Y, Haber Y, Elyoseph Z. Assessing the alignment of large language models with human values for mental health integration: cross-sectional study using schwartz's theory of basic values. JMIR Ment Health 2024;11:e55988 [FREE Full text] [doi: 10.2196/55988] [Medline: 38593424]
24. Hadar-Shoval D, Elyoseph Z, Lvovsky M. The plasticity of ChatGPT's mentalizing abilities: personalization for personality structures. Front Psychiatry 2023;14:1234397 [FREE Full text] [doi: 10.3389/fpsyt.2023.1234397] [Medline: 37720897]
25. Levkovich I, Elyoseph Z. Suicide risk assessments through the eyes of ChatGPT-3.5 versus ChatGPT-4: vignette study. JMIR Ment Health 2023;10:e51232 [FREE Full text] [doi: 10.2196/51232] [Medline: 37728984]
26. Levkovich I, Shinan-Altman S, Elyoseph Z. Can large language models be sensitive to culture suicide risk assessment? Journal of Cultural Cognitive Science 2024;8(3):257-287. [doi: 10.21203/rs.3.rs-4066705/v1]
27. Tal A, Elyoseph Z, Haber Y, Angert T, Gur T, Simon T, et al. The artificial third: utilizing ChatGPT in mental health. Am J Bioeth 2023;23(10):74-77. [doi: 10.1080/15265161.2023.2250297] [Medline: 37812102]
28. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med 2023;29(8):1930-1940. [doi: 10.1038/s41591-023-02448-8] [Medline: 37460753]
29. Denny P, Leinonen J, Prather J, Luxton-Reilly A, Amarouche T, Becker BA, et al. Promptly: using prompt problems to teach learners how to effectively utilize AI code generators. 2023 Presented at: SIGCSE 2024: Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1; 2024 March 20 - 23; Portland OR USA.
30. Ryan JJ, Lopez SJ. Wechsler adult intelligence scale-III. In: Understanding Psychological Assessment. Heidelberger Platz: Springer; 2001:19-42.
31. Kaufman AS, Lichtenberger EO. Assessing Adolescent and Adult Intelligence. Hoboken, NJ: Wiley; 2006.
32. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;6(4):284-290. [doi: 10.1037//1040-3590.6.4.284]
33. Altman DG. Practical Statistics for Medical Research. London: Chapman and Hall; 1991.
34. Gibbons RD, Lauderdale DS, Wilson RS, Bennett DA, Arar T, Gallo DA. Adaptive measurement of cognitive function based on multidimensional item response theory. Alzheimers Dement (N Y) 2024;10(4):e70018. [doi: 10.1002/trc2.70018] [Medline: 39748843]
35. Waheed W, Mirza N, Waheed MW, Malik A, Panagioti M. Developing and implementing guidelines on culturally adapting the addenbrooke's cognitive examination version III (ACE-III): a qualitative illustration. BMC Psychiatry 2020;20(1):492 [FREE Full text] [doi: 10.1186/s12888-020-02893-6] [Medline: 33023520]
36. Wright AJ, Chávez L, Edelstein BA, Grus CL, Krishnamurthy R, Lieb R, et al. Education and training guidelines for psychological assessment in health service psychology. Am Psychol 2021;76(5):794-801. [doi: 10.1037/amp0000742] [Medline: 33180536]
37. Boone KB, Sweet JJ, Byrd DA, Denney RL, Hanks RA, Kaufmann PM, et al. Official position of the American academy of clinical neuropsychology on test security. Clin Neuropsychol 2022;36(3):523-545. [doi: 10.1080/13854046.2021.2022214] [Medline: 35043752]
38. Braunack-Mayer A, Carolan L, Street J, Ha T, Fabrianesi B, Carter S. Ethical issues in big data: a qualitative study comparing responses in the health and higher education sectors. PLoS One 2023;18(4):e0282285 [FREE Full text] [doi: 10.1371/journal.pone.0282285] [Medline: 37097990]
39. Ogugua JO, Okongwu CC, Akomolafe OO, Anyanwu EC, Daraojimba OD. Mental health and digital technology: a public health review of current trends and responses. Med Sci Res 2024;4(2):108-125. [doi: 10.51594/imsrj.v4i2.754]
40. Hadar SD, Haber Y, Tal A, Simon T, Elyoseph T, Elyoseph Z. Transforming perceptionsxploring the multifaceted potential of generative AI for people with cognitive disabilities. JMIR Preprints 2024:64182. [doi: 10.2196/64182]

Abbreviations

AI: artificial intelligence
AI-BVCT: artificial intelligence-based verbal comprehension test
CCC: concordance correlation coefficient
GenAI: generative artificial intelligence
ICC: intraclass correlation coefficient
LLM: large language model
LoA: limits of agreement
TOST: Two One-Sided Tests
VCI: verbal comprehension index
WAIS-III: Wechsler Adult Intelligence Scale

XSL•FO
RenderX

Edited by A Mavragani; submitted 11.11.24; peer-reviewed by DJ Samuel, M Thomas; comments to author 18.12.24; revised version received 04.01.25; accepted 05.01.25; published 24.02.25.

Please cite as:
Hadar-Shoval D, Lvovsky M, Asraf K, Shimoni Y, Elyoseph Z
The Feasibility of Large Language Models in Verbal Comprehension Assessment: Mixed Methods Feasibility Study
JMIR Form Res 2025;9:e68347
URL: https://formative.jmir.org/2025/1/e68347
doi:10.2196/68347
PMID:

©Dorit Hadar-Shoval, Maya Lvovsky, Kfir Asraf, Yoav Shimoni, Zohar Elyoseph. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

An Integrated Platform Combining Immersive Virtual Reality and Physiological Sensors for Systematic and Individualized Assessment of Stress Response (bWell): Design and Implementation Study

Budhachandra Khundrakpam[1], PhD; Melanie Segado[1], PhD; Jesse Pazdera[1], PhD; Vincent Gagnon Shaigetz[1], MSc; Joshua A Granek[2], PhD; Nusrat Choudhury[1], MEng

[1]National Research Council Canada, Boucherville, QC, Canada

[2]Defence Research and Development Canada, Toronto, ON, Canada

Corresponding Author:
Nusrat Choudhury, MEng
National Research Council Canada
75 de Mortagne
Boucherville, QC, J4B 6Y4
Canada
Phone: 1 450 978 6198
Email: nusrat.choudhury@nrc-cnrc.gc.ca

Abstract

Background: Stress is a pervasive issue in modern society, manifesting in various forms such as emotional, physical, and work-related stress, each with distinct impacts on individuals and society. Traditional stress studies often rely on psychological, performance, or social tests; however, recently, immersive virtual reality (VR), which provides a sense of presence and natural interaction, offers the opportunity to simulate real-world tasks and stressors in controlled environments. Despite its potential, the use of VR to investigate the multifaceted manifestations of stress has not been thoroughly explored.

Objective: This study aimed to explore the feasibility of using a VR-based platform, *bWell*, to elicit multifaceted stress responses and measure the resulting behavioral and physiological changes. Specifically, we aimed to design various VR stress exercises based on neurocardiac models to systematically test cardiac functioning within specific contexts of self-regulation (executive functioning, physical efforts, and emotional regulation).

Methods: The development process adhered to guidelines for VR clinical trials and complex health interventions, encompassing 3 phases: preparation, development, and verification. The preparation phase involved a comprehensive literature review to establish links between stress, the heart, and the brain, leading to the formulation of a conceptual model based on the Neurovisceral Integration Model (NVIM) and Vagal Tank Theory (VTT). The development phase involved designing VR exercises targeting specific stressors and integrating physiological sensors such as photoplethysmography (PPG) and electromyography (EMG) to capture heart rate variability (HRV) and facial expressions. The verification phase, conducted with a small number of trials, aimed to design a study and implement a workflow for testing the feasibility, acceptability, and tolerability of the VR exercises. In addition, the potential for capturing physiological measures along with subjective ratings of stress for specific dimensions was assessed.

Results: Verification trials demonstrated that the VR exercises were well tolerated, with negligible cybersickness and high user engagement. The different VR exercises successfully elicited the intended stress demands, along with the physiological responses.

Conclusions: The study presents a novel VR-based experimental setup that allows a systematic and individualized assessment of stress responses, paving the way for future research to identify features that confer stress resilience and help individuals manage stress effectively. While our conceptual model highlights the role of HRV in providing valuable insights into stress responses, future research will involve multivariate and machine learning analyses to predict individual stress responses based on comprehensive sensor data, including EMG and the VR-based behavioral data, ultimately guiding personalized stress management interventions.

(JMIR Form Res 2025;9:e64492) doi:10.2196/64492

XSL•FO
RenderX

KEYWORDS

virtual reality; stress; physiological response; NASA-Task Load Index; cognitive demand; physical demand; vagal tone; heart rate variability

Introduction

Background

A rising problem in our society is *stress*, with serious burdens imposed on individuals and society by different types of stressors, such as emotional stress, physical stress, and work stress [1,2]. The multifaceted nature of stress makes it even more problematic, with different stressors yielding different responses and consequences. For example, increased workload and time pressure has led to increased mental stress [3]. Similarly, in recent years, there has been an increase in emotional, behavioral, and psychological stress due to the COVID-19 pandemic [4,5].

Cumulative evidence from studies of stress neurobiology indicates individual variability in how people respond to stress. While for some individuals, stress can lead to negative outcomes (eg, deterioration of physical and mental health), other individuals may not be affected to that extent [6,7]. Identifying the risk factors that make a person vulnerable to negative consequences of stress would allow timely detection and intervention. By contrast, identifying and delineating the individual features that potentially confer stress resilience (defined as the ability of most people to maintain normal functioning even under extreme stress) is important in helping individuals reach their full potential. Factors that confer resilience include genetic predisposition, age, sex, prenatal environment, sleep, and exercise [8]. In terms of biological mechanisms, stress resilience has been attributed to a coordinated activity of the autonomic nervous system (ANS); the hypothalamic-pituitary-adrenal axis; the immune system; and the brain regions that control their activity, including the hypothalamus, amygdala, hippocampus, medial prefrontal cortex, and ventral tegmental area [9-11]. In particular, stress-induced activation of ANS has been linked to changes in heart rate and vasoconstriction [9]. In addition, stress responses via activation of sympathetic-adrenomedullary and hypothalamic-pituitary-adrenal axes have been observed in response to psychosocial and physical demands during intense exercise [12].

Given the complex coordinated involvement of several bodily systems (both central and peripheral), studies on stress resilience require measurement of attributes that can reflect these interactions. One such attribute is heart rate variability (HRV), which captures heart-brain interactions and ANS dynamics [13]. HRV is defined as the variation in the time intervals between consecutive heartbeats and is an emergent property of complex interactions among several interdependent physiological systems. Several studies have shown the usefulness of HRV in the analysis of stress that the body experiences during physical training and to increase insight into physiological recovery after training [14,15]. Measurement of HRV (using electrocardiogram [ECG] or photoplethysmography [PPG]) is an easy noninvasive technique, and therefore, it has been used to investigate stress-related changes in cardiovascular health in different conditions, including exercise [16] and sleep [17]. More importantly, studies on individuals with stress-related disorders (such as anxiety and depression) revealed altered HRV metrics, indicating impairment of neurocardiac function mediated by changes to the ANS [18]. In addition, integrating HRV with other modalities such as electroencephalography have shown the impact of stress on cardiac and brain health [19,20].

Traditionally, the impact of stress on biological measures such as HRV and brain activity is investigated after inducing a specific stressor (mainly in live scenarios) using psychological, performance, or social tests [21,22]. However, in recent years, the advancement of virtual reality (VR) and head-mounted displays has allowed the investigation of different stressors that are otherwise difficult to reproduce in real life. Using head-mounted displays, immersive VR offers a sense of presence and a more natural interaction, allowing multisensorial (visual, audio, and touch) simulation of real-world tasks with tight experimental control, enabling the study of specific stressors. In a study, VR was used to elicit stress in high-altitude simulations, and stress-related responses in the heart and brain were observed using ECG and electroencephalography, respectively [23]. In a VR study using ECG signals, a multilevel stress assessment (3 levels of stress) was performed using spectrograms and 1D ECG signals that performed better than traditional machine learning (ML) models [24]. In another example, playing 3D games (a combination of the first-person shooter and horror genres) in immersive VR was shown as an effective stressor for practicing stress management skills [25]. In alignment with these findings, a meta-analysis of 52 studies revealed that stress tasks in VR elicit a variety of physiological stress reactivity, highlighting the scope of VR in the study of stress [26].

Objectives

Despite the advantages offered by VR (multisensorial—visual, audio, and touch simulation of real-world tasks), previous studies have mostly used VR to induce particular stressor at a time, for example, cognitive stress [27] and emotional stress [28]. In this work, we explored the feasibility of implementing a VR-based experimental setup to (1) elicit multifaceted stress response and (2) measure the resulting behavioral and physiological response. Using our validated VR platform *bWell* [29], we aimed to leverage specific exercises among its battery for eliciting different stressor responses, specifically for physical, cognitive, and emotional stress. Particular emphasis was given on designing VR stress exercises to systematically test cardiac functioning within specific contexts of self-regulation (executive functioning and emotional regulation) based on neurocardiac models [30,31]. In addition, by integrating VR with PPG and electromyography (EMG), we captured various physiological measures, including HRV and facial expressions (collected via EMG), along with subjective ratings of stress (namely, the Perceived Stress Scale scores for

XSL•FO
RenderX

overall stress and NASA-Task Load Index [NASA-TLX] for specific stress dimensions) and mood.

The overarching goal is that the study design or paradigm tailor to different stress domains and measurable behavioral and physiological outputs. The long-term goal is to perform multivariate and ML-based analysis for predicting the subjective perceived stress of an individual. In addition, it should be adaptable to the specific needs of each individual, allowing a systematic and individualized assessment of stress response.

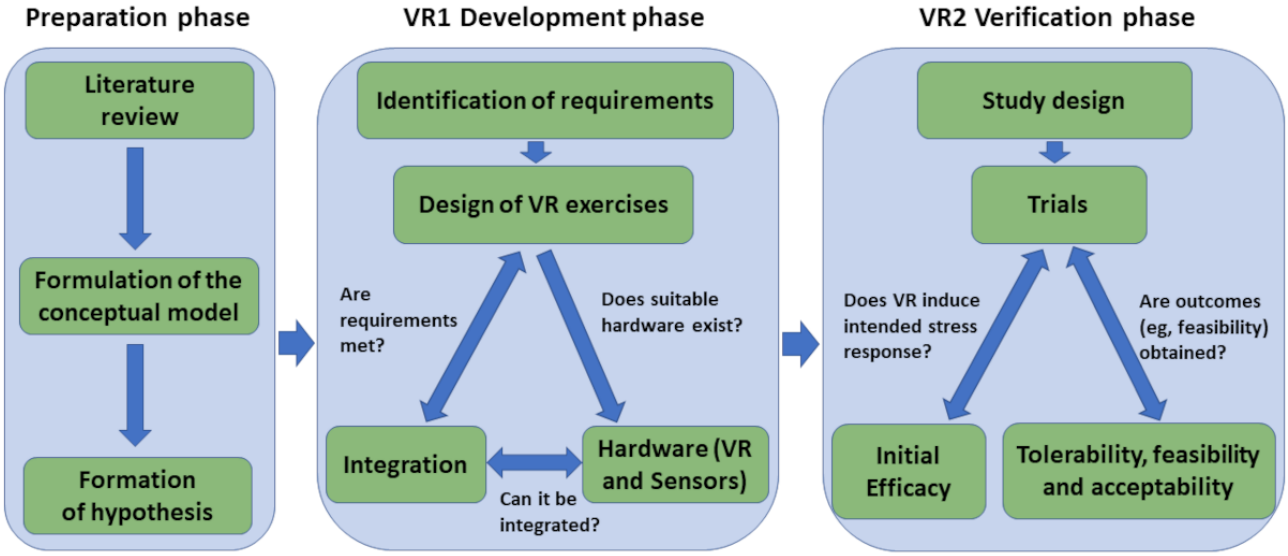
Methods

Development Process

Because the work involved immersive VR in a health context, the "Recommendations for Methodology of Virtual Reality Clinical Trials in Health Care by an International Working Group" were followed during the development process [32]. Following the recommendations, we focused on the first two phases: (1) VR1 phase with an emphasis on content development of different VR stress exercises and (2) VR2 phase with an emphasis on verification trials testing feasibility, acceptability, and tolerability. In addition, the study development process also followed guidelines from the Medical Research Council, namely "Guidance on how to develop complex interventions to improve health and healthcare," [33] which recommends combining the opinions or input elicited during content development with existing or available theories (eg, the rationale for intervention, expected changes, and how changes may be achieved). On the basis of the guidelines, our development study included a precursor step to identify the prevailing theories and frameworks in the domain to formulate hypotheses for the study.

Consequently, our development process comprised three phases: (1) a *preparation phase*, during which an extensive literature review was performed to establish the links between heart, brain, and stress, leading to the formulation of a conceptual model and formation of hypotheses (precursor phase); (2) a *development phase*, which involved iterative design of VR exercises for targeted stressors and hardware integration (VR1 phase); and (3) a *verification phase* (VR2 phase), focusing on implementing a workflow able to test the outcome measures associated with VR2 trials—tolerability, acceptability, and feasibility (Figure 1).

Figure 1. Schema of the development process. The development process comprised three phases: preparation phase, virtual reality (VR)-1 development phase and VR-2 verification phase.



Preparation Phase

Literature Review and Survey

A literature survey was performed to establish links between stress, the heart, the brain, and cognition. Given the involvement of several bodily systems in stress, we observed diverse findings: on the one hand, associations with mental load [34] and emotional stress [35], on the other hand, associations with cognitive performance [36,37]. With the goal of linking these diverse findings, we next focused on publications that have reported integrated models of the heart, the brain, and cognition (especially self-regulation, a cognitive process implicated in stress). The first such model is the *Neurovisceral Integration Model(NVIM)* [38,39]. According to the NVIM, the heart and the brain are connected by direct and indirect pathways that influence the various processes (physiological, behavioral, emotional, and cognitive) related to self-regulation and adaptability. More importantly, the NVIM postulates vagal tone as the common index that is linked to physiological, affective, and cognitive regulation. The second model is based on the *Vagal Tank Theory (VTT)* [31]. The VTT suggests that a higher vagal tone is associated with better self-regulation and adaptive functioning.

It is worth noting that the theories of interest focus on vagal tone, which refers to the activity of the vagus nerve. The vagus nerve is the largest autonomic nerve, innervating nearly every organ in the body. More importantly, the vagus nerve regulates metabolic homeostasis by controlling the heart rate [40] as well as the interval between 2 heartbeats (HRV) [41]. While direct measurement of vagal activity (ie, physiological studies of the vagus nerve) has been limited due to difficulties in recording chronic signals in the vagus [42,43], HRV provides an indirect

XSL•FO
RenderX

measure of vagal tone (a clinical metric to specify overall levels of vagal activity) because it is largely regulated by cardiac vagal control [44,45]. HRV has been shown to be a sensitive metric for detecting abnormalities in vagal tone, as observed in clinical conditions such as diabetes, heart failure, and hypertension [46-48]. More broadly, HRV, as an indirect measure of vagal tone, has shown to be useful in many aspects of psychophysiology, such as self-regulation mechanisms linked to cognitive, affective, social, and health phenomena [13,30,49]. In summary, HRV is commonly used in research as a practical index for vagal tone, and therefore, we will be using HRV in our study to link vagal tone with stress.

Formulation of the Conceptual Model

To formulate our model, we first examined the hypotheses of the NVIM and VTT. Specifically, the NVIM suggests that greater parasympathetic nervous system activity (ie, vagal tone) increases regulatory control in the brain and body [38,39]. This control supports certain biological processes such as digestion, while facilitating brain functioning related to cognitive control and emotional regulation. In contrast, withdrawal of vagal tone supports physical action, such as a fight-or-flight response. The parasympathetic nervous system is thought to monitor and balance feedback from numerous subsystems to determine the optimal level of vagal tone to deploy, given one's current context, active goal, and unmet needs [50]. HRV reflects vagal tone and, as such, should correlate with the integrity of cognitive and emotional control processes [30]. By contrast, VTT suggests that there are *3* states of vagal control, the "3 Rs": *resting*, *reactivity*, and *recovery* [31]. Optimal vagal tone is high at rest, so a healthy system should settle to a high baseline level of parasympathetic activity. In response to mental demands, vagal control systems should react by maintaining a high vagal tone to promote executive control. In response to more active demands, these systems should react by withdrawing vagal tone to promote physical performance. In practice, vagal tone often drops below resting levels even during cognitive tasks, possibly due to the strain on and gradual exhaustion of vagal control systems. Once demands are met, vagal control should transition into a recovery state, where vagal tone returns to resting levels.

The NASA-TLX provides a formal framework for categorizing different aspects of workload and task demands [51]. Therefore, we identified this index as a useful tool for understanding the demands that vagal control systems must continuously monitor and react to, as per the *VTT* and the *NVIM*. Specifically, the NASA-TLX defines 3 types of task demands—*mental*, *physical*, and *temporal*—and 3 additional factors—*performance*, *effort*, and *frustration*—that contribute to perceived workload. On the basis of the *NVIM* model of vagal control, *mental*, *physical*, and *temporal* task demands should be associated with different optimal levels of vagal tone and, therefore, different states of reactivity. *Mental* demands are best solved by promoting executive control through high levels of parasympathetic nervous system activity; in contrast, active *physical* and *temporal* demands are best addressed by a withdrawal of vagal tone.

Formation of Hypothesis

On the basis of previous publications, we formed specific hypotheses for self-regulation for each level of cardiac vagal control (Table 1). At *resting (baseline)*, we hypothesized that the higher the vagal tone, the better the self-regulation [49,52-54]. For *reactivity*, different hypotheses were formed for emotional, physical, and cognitive demands. It may be noted that for consistency with terminology in neuroscience, we have used cognitive demand as representing mental demand and emotional demand as representing frustration in the NASA-TLX. For emotional demand, (1) a complete vagal withdrawal would elicit a fight-or-flight response [52,55], (2) a small increase in vagal tone would indicate successful emotion regulation [53,54], and (3) a small decrease would indicate that the task did not elicit a sufficiently large stress response [49,56]. For cognitive demand, (1) we hypothesized an initial increase in vagal tone under low cognitive load (self-regulatory effort) and a decrease in vagal tone under high cognitive load [31,57], and (2) in the context of vagal withdrawal (ie, high cognitive load), we hypothesized that a small decrease in vagal tone would predict better performance [31,58]. For physical demand, a small vagal withdrawal would indicate quick adaptation to meet the demands of the situation [59,60]. Finally, for *recovery*, we hypothesized that the faster the vagal tone returns to the baseline level, the better the self-regulation [61].

XSL•FO
RenderX

Table 1. Description of the hypotheses corresponding to our conceptual model based on the Neurovisceral Integration Model (NVIM) and Vagal Tank Theory (VTT). A distinct hypothesis was made for each phase—resting (baseline), reactivity, and recovery.

Cardiac vagal control level	Hypothesis of self-regulation, as indicated by cardiac vagal control
1. Resting (baseline)	• A higher vagal tone indicates better self-regulation
2. Reactivity	
2a. Emotional demand	• Complete vagal withdrawal indicates that a fight-or-flight response was elicited • A small increase indicates successful emotional regulation • A small decrease indicates that the task did not elicit a sufficiently large stress response
2b. Cognitive demand	• An initial increase in vagal tone under low cognitive load (self-regulatory effort) and a decrease under high cognitive load • A smaller decrease (under high cognitive load) is associated with better performance
2c. Physical demand	• Small vagal withdrawal indicates quick adaptation to meet demands of the situation
2d. Physical and cognitive demand	• Combination of hypotheses 2c (vagal withdrawal) and 2b (initial increase under low cognitive load with a more pronounced effect)
3. Recovery (from demand)	• The faster the vagal tone returns to the baseline level, the better the coping mechanisms

VR1: Development Phase

Identification of Requirements

The fundamental objective of this work is to harness the affordances of immersive VR to provide new opportunities for practicing stress management skills, particularly the synthetic recreation of scenes that are otherwise too difficult or resource intensive to replicate in real life. The latter is of particular interest in the application toward training stress resilience, a complex and dynamic process. Combined with physiological monitoring, which can help to capture additional information on the user state, the use of VR-based technologies can enable research in support of better understanding the impacts of well-being and individual characteristics, such as mental and psychological resilience.

To study and validate the different parts of our conceptual model, we identified four main requirements:

1. *Requirement 1:* recreation of each level of cardiac vagal control (3 “Rs”: resting, reactivity, and recovery)
2. *Requirement 2:* ability to elicit different types of mental demands
3. *Requirement 3:* ability to elicit physical demand
4. *Requirement 4:* integration of physiological sensing with immersive VR, essential for the integration of HRV measures as an index of vagal tone

Design of VR Exercises

We have previously developed *bWell*, an interactive and immersive VR platform designed as a broadly applicable toolkit delivering multisensory tasks, targeting general aspects of cognition and everyday functioning [29,62-64]. For this study, we made use of *3* bWell exercises, namely *Tent*, *Mole*, and *Stroll*, which were adapted to systematically test adaptive aspects of HRV within specific contexts of self-regulation (executive functioning, emotional regulation, and physical demand) and for different levels of cardiac vagal control, as identified in our conceptual model and requirements 1 and 2. The bWell platform and exercise battery are described in detail in our previous publication [29]. The following is a brief summary of the three exercises:

1. *Tent*—an exercise designed for self-guided stress management in which the user is immersed in scenic views and asked to look around while focusing on their breathing
2. *Mole*—an exercise targeting response inhibition and cognitive control in which the user has a hammer in each hand and has to hit cylinders that pop up in front of them (only when the cylinder and hammer are of the same color)
3. *Stroll*—an exercise targeting sustained attention with a continuous performance test (CPT) in which the user has to press a button for each new shape that appears, except when it is a green diamond. The user is represented by a self-avatar and is passively displaced for a stroll in a natural scene.

To test our specific hypotheses for different stress demands or stressors (Table 1), we adapted the aforementioned bWell exercises (except for the *Mole* exercise for cognitive demand) as follows:

1. *Tent*
   - Nature scene to collect data on user
     - *Resting:* As per recommended guidelines for experimental settings, baseline HRV should be recorded in a “nontask situation that best controls for the presence of task comparison” [65]. Consequently, the administration of the exercise

XSL•FO
RenderX

was modified such that the user was asked to simply immerse themselves in the scene, limit physical movement, and breathe spontaneously for 5 minutes.

- *Recovery:* this scene was also selected to be administered after a given task to allow a period of recovery after stress demand. As a measure of recovery, a Likert scale implemented in VR was added to collect a report on mood right after the exposure to a stressor and after the recovery period in a relaxing nature environment.

- *City*—in addition to the nature scenes, to elicit emotional demand, we adapted the *Tent* exercise to include an environment such that the user is immersed in a busy city with noise pollution—roads with ongoing traffic, including audio honking cars and sirens.

2. *Stroll*—to exert physical demand, we adapted the Stroll exercise such that the user must physically displace themselves to advance in the scene. The user was asked to jog, keeping their heart rate at a moderate level of exertion (the predefined threshold was displayed as a gauge based on individual heart rate). The exercise could include a CPT, as described earlier, to exert both physical and cognitive demand (*Stroll+CPT*) or could be administered without the test as a purely physical task (*Stroll*).

The formulated hypotheses shown in Table 1 can now be systematically matched with our designed VR exercises (Table 2).

Table 2. Description of the hypotheses corresponding to our conceptual model with the designed virtual reality (VR) exercises. A distinct hypothesis was made for each phase—resting (baseline), reactivity, and recovery—and these hypotheses were tested with different VR exercises.

Cardiac vagal control level	Requirement	bWell exercise
1. Resting (baseline): Tent—nature scene, no specific task given	User to immerse themselves, limit physical movement, and breathe spontaneously as a baseline state	
2. Reactivity		
2a. Emotional demand: city—(passive) city scene	User to immerse in a busy city with excessive noise pollution to elicit emotional demand	
2b. Cognitive demand: mole—(active) targeting response inhibition	User to engage in a dynamic and bimanual go or no-go task to elicit cognitive demand	
2c. Physical demand: stroll—(active) physical running task	User to jog to create a physical demand, with heart rate biofeedback guiding them to maintain a consistent moderate level of activity	
2d. Physical and cognitive demand: stroll+CPT[a]−dual task with running task as 2c along with a cognitive task (response inhibition)	User to jog and engage in the task to elicit both physical and cognitive demands	
3. Recovery from demand: Tent—nature scene	User to immerse themselves and limit physical movement to recover	

[a]CPT: continuous performance test.

Hardware Identification (VR+Sensors)

Based on the identified requirements, we aimed to identify and integrate hardware for use with the bWell exercises. We focused on evaluating the suitability of the use of immersive VR technologies. The choice of hardware considered specialized equipment aiming to maximize the feeling of "*presence*" in virtual environments as well as the potential to democratize the use of VR for training more broadly. Consequently, we evaluated both a high-end system—*HTC Vive Pro Eye*—or use within dedicated facilities as well as a more accessible option—*Pico Neo 3 Eye*. The latter targets the opportunity for individuals to practice at home or in settings outside the laboratory (eg, on the field, classroom, and clinical).

We aimed to reproduce the required factors in the simulated environment with the integration of specialized hardware. A VR treadmill (KatVR) was identified to recreate the physical demands. We also aimed to capture stress via the inclusion of physiological sensing, with a focus on integrating heart-based measures. The *EmteqPRO* facemask was identified for use in this project because it is a commercially available "plug-and-play" system for use with immersive VR, provides researchers with access to raw data, and is equipped with automatic classification algorithms for processing raw heart rate data into HRV as well as classification of emotion [66]. The system can accommodate multimodal measures of EMG and HRV via the use of a facemask that can be directly integrated onto the VR headset. The Emteq device does not add any additional steps to the workflow for carrying out trials with immersive VR; instead, the user simply must put on the headset as they normally would.

XSL•FO
RenderX

Hardware Integration

A first evaluation was conducted to determine the suitability of the identified hardware.

Immersive VR Headset

The *HTC Vive Pro Eye* and the *Pico Neo 3 Eye* headsets were identified because they both offer eye tracking. In addition, the Emteq solution is available for both. The HTC Vive Pro Eye is a tethered device, meaning it requires a high-performance, VR-ready PC to operate. A tethered VR headset operates only when connected to a PC through a cable connection and offers superior graphics and fidelity in comparison to stand-alone headsets. The *Pico Neo 3 Eye* is a stand-alone headset, which is wireless, and the VR is run entirely within the headset. The benefits of a stand-alone headset include increased comfort and mobility, a simple setup, and a lower cost compared to tethered devices. However additional efforts to optimize the VR experience in terms of graphical rendering and design are required to minimize lags related to hardware performance.

It was determined that both headsets are suitable for use in the proposed study with certain adaptations. A wireless adapter was integrated into the HTC Vive Pro Eye, enabling untethered use with the treadmill. For the Pico Neo 3 Eye, an external dongle was required to enable screen casting to an external monitor, allowing the test administrator to view the participant's perspective within the VR during trials.

Emteq Integration With bWell in Unity

The Emteq face mask was integrated within *bWell* through the use of the “EmteqLabs Unity” plugin. Using the plugin, the EMG amplitude as well as the heart rate (beats per minute) of each participant was extracted. Using these extracted values, several widgets (Figure 2) were implemented in the 2 existing layers of the *bWell* user interface (UI): test administrator UI and participant UI. The first widget was part of the EmteqLabs plugin package and was used to display EMG amplitude (Figure 2A). The same widget also provided the administrator with an indication of sensor contact. Good contacts were represented by green dots on the widget, while bad contacts were represented by gray dots. The second widget was composed of 2 simple text boxes that were used to inform the test facilitator of the current connection status of the Emteq (Figure 2B). The third widget was a speedometer that was used to inform the participant whether they were in a heart rate zone of moderate effort (Figure 2C). The widget was implemented as a gauge to prompt participants to raise or lower their effort level, with the zone in the widget turning green when participants are at the correct range. Specifically, (zone turns green at correct range). A heart icon that grew and shrank based on a speed that matched the current beats per minute was also included in the speedometer.

Figure 2. Integration of physiological sensor with virtual reality (VR). Implementation of widgets in the user interface for integration of EmteqPRO (physiological sensor for heart rate variability) into VR platform (bWell): (A) widget for displaying electromyography (EMG) amplitude in VR, (B) widget for connection status of EmteqPRO, and (C) widget for displaying heart rate.

Treadmill Integration

The usability of donning a wireless immersive VR headset while running on the treadmill was tested. Running speed (vigorous vs moderate) was evaluated, and it was found that both vigorous and moderate running was acceptable in terms of usability for both headsets. However, it was seen that running on a treadmill was not intuitive and that a warm-up time was required. A test run was added to the protocol, such that the user could have some time to familiarize themselves with running on the treadmill before putting on the VR headset.

It was also seen that vigorous running may prove to be difficult for participants running on the treadmill for the first time. In addition, running on the treadmill while performing a cognitive exercise (*Mole* exercise) occasionally resulted in inadvertent errors from the user. As a result, the protocol was adjusted to have users run at a moderate pace (light jog) during the study. For the dual task (Stroll+CPT), a cognitive exercise with simpler user interaction was use. The exercise was changed from the initial choice to use *Mole* (requiring users to swing with their arms) to the *Stroll* (requiring users to pull a trigger on the controller).

Finally, the physical demand exercise (*Stroll*) was originally designed for passive displacement. In this task, it is the self-avatar taking a nature stroll while the actual user is not moving. To promote a higher sense of *presence* for the user, we added advancement in the VR scene based on the user’s own steps while advancing on the treadmill. The running pace was added as an output measure in the VR logs generated automatically after a session with bWell.

The conclusion from the hardware integration tests was that both the *HTC Vive Pro Eye* and *Pico Neo 3 Eye* were suitable. However, because the primary use at this stage was for study purposes, it was decided that it would be better to use the hardware that is designed for external administration of VR

XSL•FO
RenderX

versus self-administration with a headset. In the workflow for administration, the bWell platform provides a UI that allows the test facilitator to dynamically change the exercise settings while a user is immersed in VR [29]. Consequently, it would be better suited for the iterative approach in finding the optimal exercise settings for the study. Thus, the decision was to use the *HTC Vive Pro Eye* for the study and the ensuing verification trials.

VR2: Verification Phase

In this phase, we focused on conducting early trials for evaluation and verification of feasibility, acceptability, tolerability, and initial efficacy (VR2 phase).

Study Design

We designed a study to systematically test our hypotheses (Figure 3). The study protocol consisted of a 5-minute initial baseline consistent with what is used in the HRV literature (Table S1 in Multimedia Appendix 1). During baseline, participants were in a standing position with their eyes open to optimally match the body position used during task conditions. Participants experienced a neutral nature scene. The baseline HRV was collected to establish a reference point. Participants then experienced alternating blocks of stressor tasks interspersed with recovery (Tent nature scenes). Stressor task presentation order was pseudorandomized across participants. Each participant performed all 4 exercises (stress conditions). Each stressor condition was followed by a 2-minute recovery period in the *Tent* (Table S1 in Multimedia Appendix 1).

Figure 3. Description of the study design. A within-participants repeated measures study design was implemented. It comprised an initial baseline (5 minutes), followed by 4 blocks of stressor (2 minutes each) and recovery (2 minutes each) periods. The stressors were 4 virtual reality (VR) exercises targeting emotional demand, cognitive demand, physical demand, and cognitive+physical demand. The sequence of stressors was pseudorandomized. CPT: continuous performance test.

The measures of VR2 trials were captured as follows:

- *Acceptability:* user experience, as assessed by the game user experience satisfaction scale (GUESS)-18 gamer user experience satisfaction scale [67] and the System Usability Scale [68]
- *Tolerability:* user comfort and any side effects from the use of technology, as assessed by the Simulator Sickness Questionnaire (SSQ; captured both before and after VR exposure)
- *Feasibility:* ability to elicit the intended stress response, as measured via physiological response as well as self-report

For the self-report measure, the NASA-TLX [51] and single-item rating for stress questionnaires were used.

Mood Likert scales were also automatically administered in immersive VR before and after the recovery exercise. Participant and VR views were also recorded. After completing the bWell session, participants reviewed a screen recording of the completed virtual task, provided qualitative feedback, and completed the set of questionnaires.

It may be noted that while the NVIM and VTT predict the dependence of vagal tone on task conditions, the NASA-TLX measures 6 aspects of workload. The study design involved 5 conditions, 1 without a stressor and 4 others in which each condition features a different exercise representing specific stressors based on the NASA-TLX (summarized in Table 3). To effectively assess the effects of each condition on stress responses, we proposed a within-*participants* design, where each participant completed all conditions of the experiment. The order of the stressor conditions was pseudorandomized across participants. The approach taken to experimentally isolate the contributions of each type of stress required the removal of

XSL•FO
RenderX

performance criteria (eg, score) from the tasks, as they may create confounds in the stressor type not related to the experimental manipulation of the condition. However, the cumulative effects of stressors were also investigated using a physical+cognitive stressor dual task, as the inclusion of multiple simultaneous stressors is more representative of everyday stressful situations.

Table 3. Description of the different types of stressor demands for different virtual reality (VR) exercises. bWell VR exercises elicited different types of demands, as defined by the NASA-Task Load Index workloads. The main stress or demand captured for each VR exercise is shown in italics.

	Mental demand	Physical demand	Temporal demand	Performance	Effort	Frustration
Relaxing nature environment (*tent*: nature scene)	No	No	No	No	No	No
Emotional stressor (*city*: busy city scene)	No	No	No	No	No	*Yes* [a]
Cognitive stressor (*mole*: dynamic bimanual go or no-go task)	*Yes* [a]	No	Yes[a]	Yes[a]	Yes[a]	No
Physical stressor (*stroll*: running task)	No	*Yes* [a]	No	Yes[a]	Yes[a]	No
Physical+cognitive stressor (*stroll+CPT:* running with continuous performance test)	*Yes* [a]	*Yes* [a]	No	Yes[a]	Yes[a]	No

[a]Demands elicited by the bWell virtual reality exercises.

We set thresholds to the NASA-TLX as a self-reported measure to ensure that the designed VR exercises targeted the intended type of demand, that is, *cognitive*, *emotional*, *physical*, and *dual (ie, physical+cognitive)*. We defined the following threshold in terms of NASA-TLX score (score range=0-10): a score>5 indicates that the demand is being elicited (eg, *City* for emotional demand, *Mole* for cognitive demand, *Stroll* for physical demand, and *Stroll+CPT* for physical and cognitive demand; Table 4). An additional aim was to not elicit demand related to poor performance because this may lead to a potential confound. In the NASA-TLX, performance is reverse coded (ie, 0=perfect performance). As such the threshold was set to<4, so that that participant should feel as if they performed well in all exercises. Finally, since all exercises invoked a type of demand except the one involving emotional demand, they were expected to involve an element of effort.

Table 4. Description of the thresholds for NASA-Task Load Index (NASA-TLX) scores for eliciting different stress demands. We set the NASA-TLX subscale scores to above the threshold, for the different types of stress demands that were elicited by each virtual reality exercise in Table 3.

	Frustration	Physical demand	Mental demand	Performance
City	>5[a]	—[b]	—	<4
Mole	<4	<4	>5[a]	<4
Stroll	—	>5[a]	—	<4
Stroll+CPT[c]	—	>5[a]	>5[a]	—

[a]NASA-TLX subscale scores that were above the threshold.

[b]Not available.

[c]CPT: continuous performance test.

More specifically, we targeted the following scores:

- *City*—a NASA-TLX score of >5 for *frustration*, which would indicate an emotional demand. We also intended for all other NASA-TLX scores to remain <4 to solely evoke an emotional response. Given the passive nature of this exercise, other demands should not be induced.
- *Mole*—a NASA-TLX score of >5 for *mental*, which would indicate a cognitive demand. A score of >5 would indicate that these demands were elicited. We also targeted *frustration* and *physical* NASA-TLX scores of <4 to solely target demands related to cognitive performance.
- *Stroll*—a NASA-TLX score of >5 for *physical*, which would indicate a physical demand. A demand of >5 would indicate that this was the case. We also targeted *mental* and *frustration* NASA-TLX scores of <4 to solely target demands related to physical performance.
- *Stroll+CPT*−a NASA-TLX score of >5 for *mental* and *physical*, which would indicate a dual cognitive and physical demand. We also intended to maintain a *frustration* score of <4 because we wanted to target solely cognitive and physical demands.

In addition, we sought to verify that *mental* demands are comparable between the *Mole* and *Stroll+CPT* tasks.

Verification Trials

Verification trials were conducted using HTC Vive Pro Eye on 3 healthy participants according to the protocol defined during the study design.

XSL•FO
RenderX

Ethical Considerations

Ethics approval for the study was obtained from the Research Ethics Board (REB) of the National Research Council (NRC 2020-2023 and NRC 2023-46) and from Defence Research and Development Canada Human Research Ethics Committee (HREC 2022-032). Informed consent was obtained before the start of the study. Data were deidentified before any analysis. There was no compensation paid to the participants.

Results

The results of the verification trials are reported below in terms of tolerability, acceptability, and feasibility.

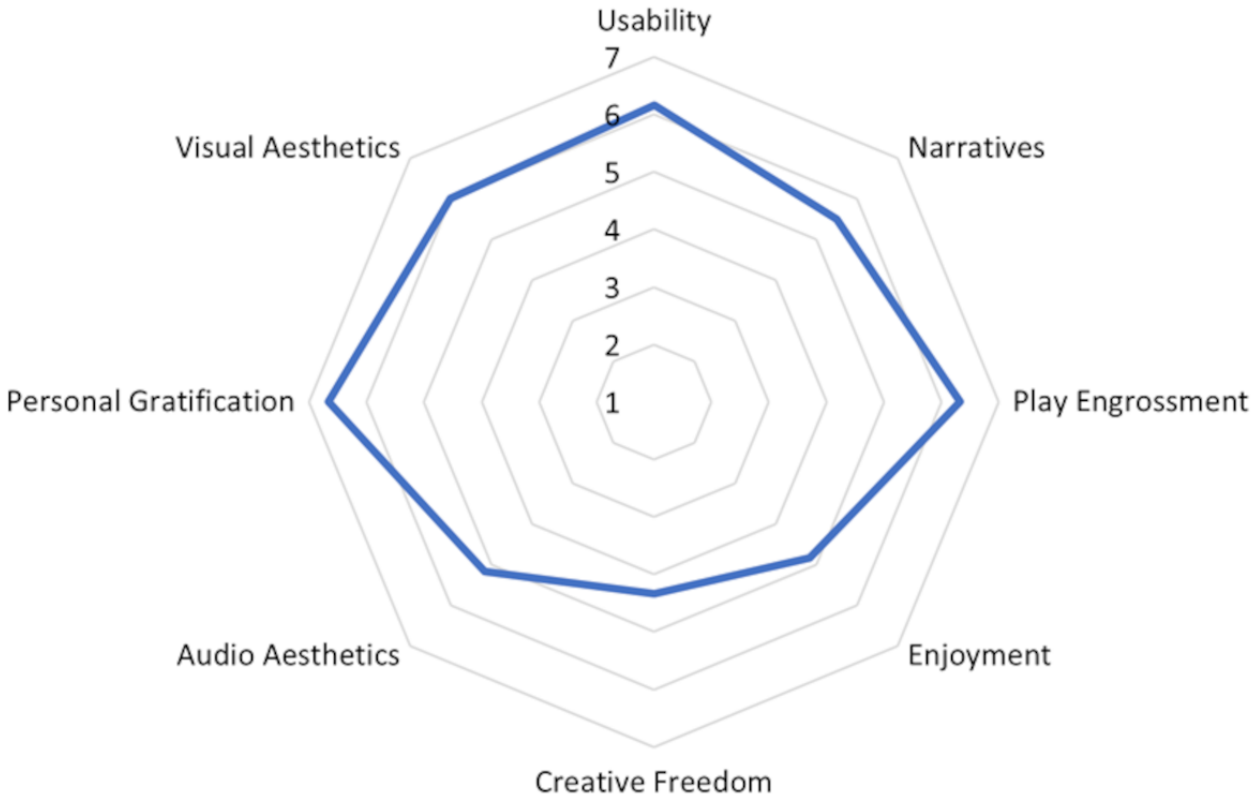
Tolerability: Cybersickness During the VR Sessions

We first checked whether the VR exercises designed in our study induced cybersickness using SSQ [69]. Table S2 in Multimedia Appendix 1 shows the scores for SSQ for pre- and post-VR sessions for 3 participants. Total SSQ scores were obtained from the scores of the 16 items of the questionnaire [70]. Results showed negligible score (total SSQ score: 0-5), which are within the acceptable range in the literature [71]. Thus, our results show that the participants did not experience adverse effects, especially cybersickness during the VR exercises.

Acceptability: User Experience

Using the GUESS-18 scale [67], we checked whether the VR exercises designed in our study were immersive. GUESS-18 has 9 subscales with 2 statements associated with each subscale. Of the 9 subscales, we did not include "social connectivity," as there was no interaction between participants for the VR exercises. Table S3 in Multimedia Appendix 1 shows the scores of the 8 subscales of GUESS-18 for the 3 participants. The subscale scores of the GUESS were computed by averaging the scores within each subscale. Then, the overall GUESS score was computed by summing the subscale scores [67]. We observed an overall GUESS score of 43.17 (Table S3 in Multimedia Appendix 1), which aligns well with previous findings [72] and indicates the usability and immersiveness of the VR exercises designed in our study. Regarding the individual subscales, we observed greater average scores for *usability*, *play engrossment*, and *personal gratification* and lower average scores for *creative freedom* and *enjoyment* (Figure 4).

Figure 4. Findings of user acceptability. User ratings on the game user experience satisfaction scale (GUESS)-18 questionnaires showing greater average scores for the subscales usability, play engrossment, and personal gratification and lower average scores for creative freedom and enjoyment. An overall GUESS score of 43.17 (Table S3 was observed, which aligns well with previous findings, indicating the usability and immersiveness of the virtual reality exercises designed in our study.

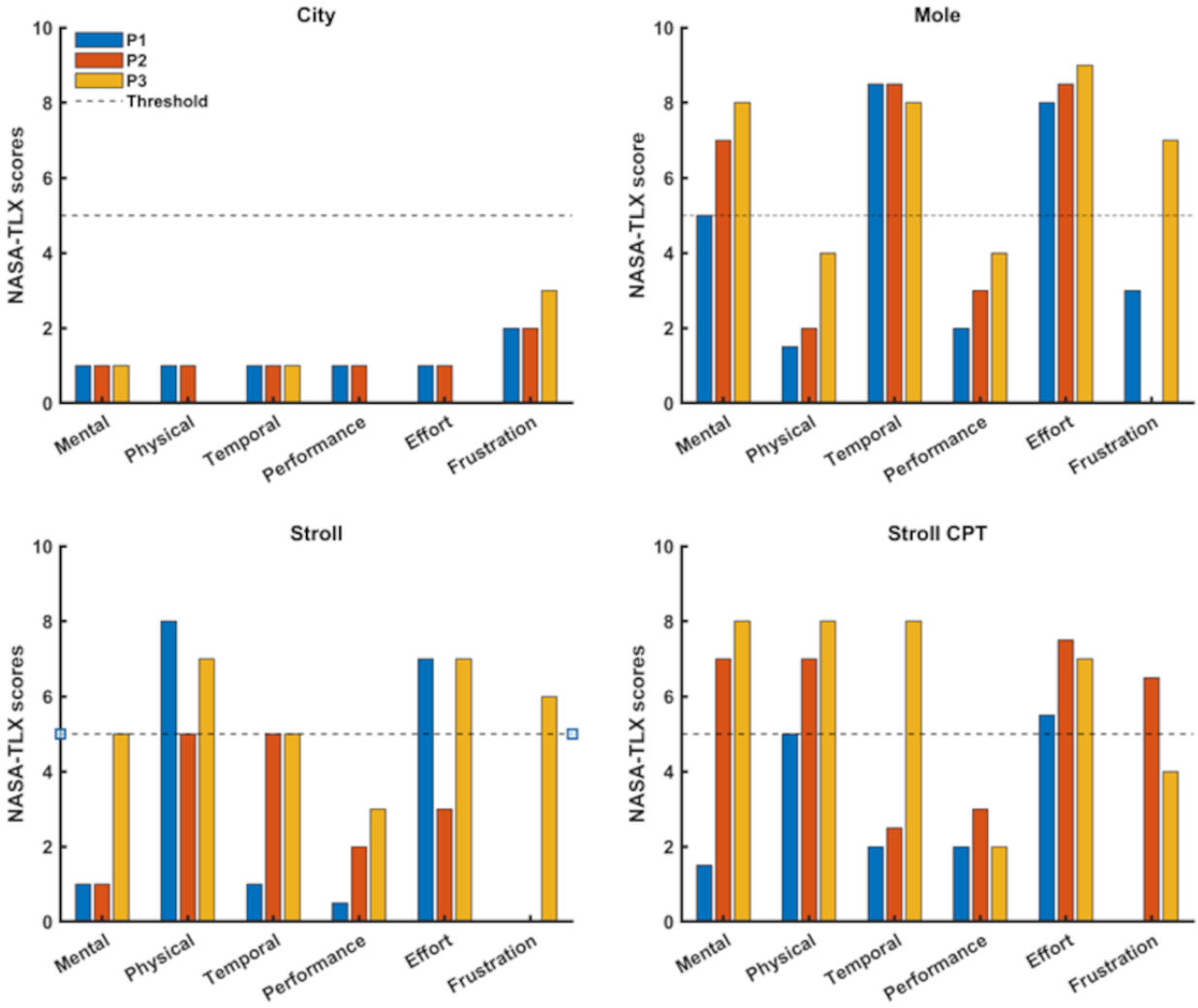


Feasibility: Do the VR Exercises Induce Genuine Stress Responses (Perceived)?

We evaluated the NASA-TLX scores to investigate the types of stressors and their associated demands. Figure 5 summarizes the data for NASA-TLX from verification trials of 3 participants.

XSL•FO
RenderX

Figure 5. Comparative analysis of different types of demands for different stressors. User ratings on the NASA-Task Load Index (NASA-TLX) revealed that participants perceived different types of demands (eg, mental, physical, and temporal) for the different VR exercises. Here, P1, P2, and P3 denote participant 1, participant 2, and participant 3, respectively, and the dashed line represents the threshold set for different demands associated with the different stressors (exercises). CPT: continuous performance test.



On the basis of the designed thresholds for the different stressors and their associated demands (refer to the Methods section), we observed the following for each of the bWell exercises (Figure 5), which led to further iterative development:

- *City*: a NASA-TLX rating of >5 for *frustration* was not achieved, indicating a need for adapting the exercise settings and creating new functionalities. All other demands were also not elicited (<4). In particular, users felt that they performed well (<4), and as such, it was verified that there would not be any confounds in *frustration* being elicited due to emotional response related to feelings of poor performance rather than VR stimuli.
- *Mole*: a cognitive demand was elicited with a *mental* demand of >5. We also observed a high level of *temporal* demand and associated *effort*.
- *Stroll*: a *physical* demand and associated *effort* was elicited (>5).
- *Stroll+CPT*: a dual cognitive (*mental*) and *physical* demand was elicited (>5), along with associated *effort*.

Because cognitive demand was elicited in both tasks, similar to the Mole task, the 2 can be compared in the analysis of cognitive demand, even though they involve different cognitive tasks. It appears that *temporal* demand may also be elicited but was not conclusive at this stage because of high variability in the ratings. This may be clarified with results from additional participant trials from the full study. From Table 4, we observed that this demand was not consistently present. This may be attributed to the fact that this was a dual task, where the *temporal* demand may be felt for one aspect (eg, running) but not for the other (eg, cognitive task), and when prompted by the NASA-TLX, the participant may respond with only 1 task in mind.

Overall, the results showed that we achieved eliciting different stress demands as intended in our study design: for *Stroll* to exert a *physical* demand, *Mole* to exert a *mental* demand, and a combination of exercises (*Stroll+CPT*) to exert both *mental* and *physical* demands. At the same time, we observed that participants felt they performed well in all exercises, minimizing confounds due to feelings of poor performance. These verification trials allowed us to finalize the exercise configurations to be used for the full study trials.

XSL•FO
RenderX

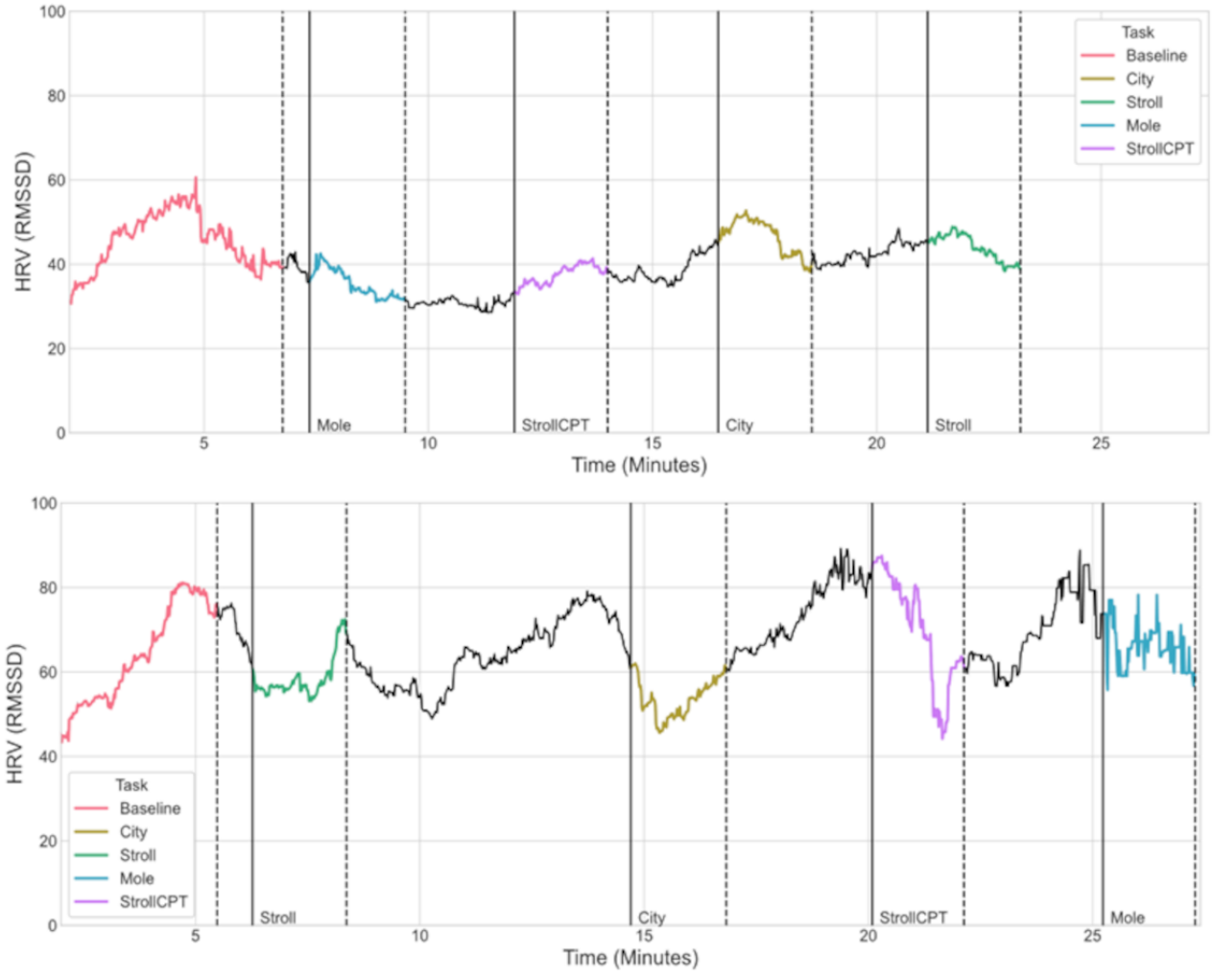
Feasibility: Do the VR Exercises Induce Genuine Stress Responses (Physiological)?

To evaluate the physiological stress responses, we examined the HRV measures obtained, specifically the root mean square of successive differences (RMSSD) between normal heartbeats. We adopted RMSSD as the HRV measure because previous studies have shown that RMSSD reflects vagal tone [38,73] and is relatively free of respiratory influences. It was seen that the data were quite noisy, artifacts were present, and the RMSSD values from the raw PPG data were too high. For example, values in the range of 150 to 300 were obtained, which is much higher than normal. Discussion with the Emteq team revealed that this was likely because the measures were sampled in a dynamic VR environment, particularly where in some instances, users were physically running during PPG data acquisition. Emteq advised using their signal quality metric and using only the data where good quality was indicated.

To have better control of the data processing, the team decided instead to implement an in-house data processing pipeline. For this, the raw PPG signals were processed using the Python library *HeartPy* [74]. On the basis of the *HeartPy* documentation, which states that RMSSD is extremely sensitive to incorrectly detected heartbeat locations, any RMSSD value artifacts (RMSSD does not typically exceed 130) were removed and peak interpolation was used to minimize the number of incorrectly placed peaks due to low sampling rate [74]. After these steps, an improved signal-to-noise ratio was observed and the resultant RMSSD values fell well within the normal range expected (between 30 and 70).

We then compared HRV across each of the different conditions for 2 individuals. In the example shown in Figure 6, it can be seen that a given task appears to elicit a change in response. It is also evident that baseline, reactivity, and recovery HRV patterns were distinct for the 2 sample individuals. As can be seen in the lower panel of Figure 6, some instances of residual noise remain in the Mole exercise after processing. Given that the data are being collected during interactive tasks with a good degree of movement, a high noise-to-signal ratio was expected. The next steps will involve exploring other techniques, particularly for motion artifact removal, with the goal of further reducing noise in the signal.

Figure 6. Characterization of distinct patterns of heart rate variability (HRV) for different virtual reality (VR) exercises. HRV data (root mean square of successive differences [RMSSD]) for the different VR (bWell) exercises are shown for 2 participants (row 1 for participant 1 and row 2 for participant 2) along y-axis and time along x-axis. The plots illustrate distinct HRV responses for the different exercises as well as individual variability in the HRV responses. CPT: continuous performance test.



XSL•FO
RenderX

Discussion

Principal Findings

In this paper, we described the development process of a study design for testing the ability of immersive VR exercises to elicit genuine stress responses. The study development followed recommendations for use of VR in clinical trials [32] as well as guidance in developing complex interventions for improving health care [33]. On the basis of these guidelines, our study comprised a combination of *prevailing theories and frameworks* (eg, NVIM and the VTT), *experiences of users*, and *iterative actions during development*, which resulted in concrete outcomes [33]. More specifically, we built on our validated VR platform *bWell* [29] to integrate physiological sensors (PPG) and assessed how different types of VR exercises target different types of stressors (characterized by the NASA-TLX). The results of verification trials showed that the designed VR exercises induced negligible cybersickness, were well tolerated, and demonstrated high user engagement. In addition, in alignment with our hypotheses, the different VR exercises successfully elicited the intended stress demands (as measured with NASA-TLX subscales, eg, mental, physical, and temporal demands), along with distinct physiological responses.

In line with the recommended guidelines for VR trial design in the health domain [44], our study design considered both tolerability and acceptability. Keeping in mind user comfort during VR, the exercises were designed in such a way that the duration for the entire session lasted only 25 minutes. In addition, during the VR session, every exercise was followed by a 2-minute recovery state. By allowing the participants to experience alternating blocks of stressor tasks interspersed with recovery, this study design was used as a guide to evaluate the technical feasibility and scope for the phase of verification trials. Indeed, the results of the verification trials showed user comfort and tolerability, especially VR as an acceptable framework with *negligible* experience of cybersickness (assessed with SSQ scores; Table S2 in Multimedia Appendix 1). As a measure of acceptability, results of verification trials using the GUESS-18 scores showed a high degree of "*presence*" during our VR sessions (Table S3 in Multimedia Appendix 1 and Figure 4), consistent with earlier findings [72]. Because the sense of "presence" (or the subjective feeling of "being there") characterizes the experience of users, results showing a high degree of presence indicate that the VR exercises designed for the study created the sensation of physical presence and resulted in users behaving and reacting as in the real world, highlighting the importance of a good user experience and engagement. Best practices for HRV data collection were also considered for the study design. In experimental settings, it is recommended that the baseline HRV should be recorded in a "nontask situation that best controls for the presence of task comparisons" [65]. Consequently, in this study, baseline was recorded in VR and in a similar position to how the participants would complete the task. Specifically, baseline was recorded with the user on the treadmill and passively experiencing the VR environments. The session to collect baseline HRV was designed as follows: eyes open, standing, spontaneous breathing, and free to look around and immerse themselves in the VR environment. In addition and according to guidance on HRV reporting, RMSSD was selected as the HRV measure because it is appropriate for short sampling times (ie, <5 minutes) [41].

While the results of verification trials determined the feasibility and validity of our study protocol, some results indicated that modifications were also required. For example, the NASA-TLX scores showed high levels of *Frustration*, which captured participants' physical demands or frustration related to performance and not necessarily due to an emotional stressor. Because our original goal was to investigate the effects of emotion, we made certain modifications in our bWell exercises to elicit emotional responses in the participants, such as amplifying the noise pollution in the virtual environment. The results of the verification trials also indicated that we need to modify the administration of the NASA-TLX, which currently probes *Frustration* related to both performance and emotion. Instead of asking "How insecure, discouraged, irritated, stressed and annoyed were you," the question will be modified to include only the elements related to emotion "How irritated and/or annoyed were you?"

Implications

It may be mentioned that the overarching aim for the integration of behavioral and physiological outputs was to eventually enable systematic and individualized assessment of stress response. Responses to stress are often multifaceted with a profile specific to the individual. A given stressor may cause a response in one person but not necessarily in another, which may be attributed to the influence of personal experiences. The result is that everyone does not experience stress in the same way. In an effort to standardize the demands placed on individuals, we relied on several mechanisms to adapt the VR exercises in real time and customized them to the participant. For cognitive demand, we aimed to elicit a uniform cognitive demand among varied individuals. To do so, we leveraged the adaptive algorithms in *bWell* to control the difficulty level based on performance (for *Mole* exercise). The adaptive mechanism is designed to maintain an 80% success rate, ensuring the exercise is neither too easy (boring) nor too difficult (demotivating). For physical demand, we made use of the heart rate biofeedback to ensure consistent physical demand because running at a specified speed may be more or less exerting for one individual versus another. To achieve this, a heart rate gauge was displayed to the participant to encourage them to maintain a level of physical activity that corresponds to moderate intensity. This measure is additionally tailored to the individual because it considers the age of the participant, where moderate physical activity is calculated at 50% to 70% of maximum heart rate (maximum heart rate=220 – age).

Study Limitations

There were certain limitations in our study. First, participants reported issues with disengagement and a lack of frustration during the VR exercise (city). The reason for this could be the low resolution of the visual assets in the scene, a concern that will be addressed in future work by improving the resolution. Second, there could be a learning effect in terms of interactions in VR for participants who had no prior VR experience. In the full study (planned next steps), participants will be given

XSL•FO
RenderX

tutorials for immersion in VR. Finally, the verification trials were done with few participants (N=3), limiting our ability to generalize findings. However, we would like to emphasize that our study was designed to identify, create, and implement an experimental setup for testing the ability of immersive VR exercises to elicit genuine responses to different types of stressors. In addition, it aimed to assess the feasibility, tolerability, and acceptability of these exercises through verification trials. This is consistent with previous studies that focused on study development [75]. Because the purpose of the verification trials was to show the feasibility, tolerability, and user acceptability, we collected data from only 3 participants, with the intention to run full trials in the future. It is worth noting that, despite the limited number of participants, we were able to demonstrate the feasibility, tolerability, and user acceptability of the study.

Conclusions

In conclusion, we identified, designed, and implemented an experimental setup for testing the ability of immersive VR exercises to elicit genuine responses corresponding to different types of stressors and tested its feasibility, tolerability, and acceptability using verification trials. While only a small number of trials were carried out with participants, the verification trials provided the support required to carry out a full study (planned next steps).

A strong point of our study design was the collection of several physiological measures that permitted the collection of quantitative measures complementing that of self-report with the NASA-TLX. With the capture of measures such as HRV and facial expressions, it will be possible to compare how the bWell exercises may physiologically elicit response versus what users perceive to be elicited. Such evaluations may serve to obtain a more complete picture of the in-situ user state and may permit gaining insights on the use of traditional self-report, which has been critiqued as an unreliable index of functional outcomes [76]. In this regard, future work will explore multivariate analysis such as partial least squares [77] and ML techniques such as random forest [78] for predicting individual user responses to different stressors based on multiple input data, including sensor data (eg, HRV and facial expressions) and VR-based behavioral data (eg, response time). Such work will help in identifying and delineating the individual features that confer stress resilience and will guide us in helping individuals reach their full potential.

Acknowledgments

The authors would like to thank the participants who agreed to complete the verification trials. The authors also thank Alejandro Olivares Hernandez for help in integrating KatVR to the *bWell* platform.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to legislation regarding privacy and security. None of the data for the experiments reported here are available, and none of the experiments were preregistered.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of the timeline for different task and recovery, Simulator Sickness Questionnaire scores for pre– and post–virtual reality session, and subjective ratings of game user experience satisfaction scale-18 questionnaires.
[DOCX File , 390 KB - formative_v9i1e64492_app1.docx]

References

1. Yaribeygi H, Panahi Y, Sahraei H, Johnston TP, Sahebkar A. The impact of stress on body function: a review. EXCLI J 2017;16:1057-1072 [FREE Full text] [doi: 10.17179/excli2017-480] [Medline: 28900385]
2. Chen B, Wang L, Li B, Liu W. Work stress, mental health, and employee performance. Front Psychol 2022 Nov 8;13:1006580 [FREE Full text] [doi: 10.3389/fpsyg.2022.1006580] [Medline: 36425815]
3. Alberdi A, Aztiria A, Basarab A. Towards an automatic early stress recognition system for office environments based on multimodal measurements: a review. J Biomed Inform 2016 Mar;59:49-75 [FREE Full text] [doi: 10.1016/j.jbi.2015.11.007] [Medline: 26621099]
4. Pedrosa AL, Bitencourt L, Fróes AC, Cazumbá ML, Campos RG, de Brito SB, et al. Emotional, behavioral, and psychological impact of the COVID-19 pandemic. Front Psychol 2020 Oct 2;11:566212 [FREE Full text] [doi: 10.3389/fpsyg.2020.566212] [Medline: 33117234]
5. Saladino V, Algeri D, Auriemma V. The psychological and social impact of COVID-19: new perspectives of well-being. Front Psychol 2020 Oct 2;11:577684 [FREE Full text] [doi: 10.3389/fpsyg.2020.577684] [Medline: 33132986]
6. Ebner K, Singewald N. Individual differences in stress susceptibility and stress inhibitory mechanisms. Curr Opin Behav Sci 2017 Apr;14:54-64. [doi: 10.1016/j.cobeha.2016.11.016]

XSL•FO
RenderX

7. Sapolsky RM. Stress and the brain: individual variability and the inverted-U. Nat Neurosci 2015 Oct 25;18(10):1344-1346. [doi: 10.1038/nn.4109] [Medline: 26404708]
8. Bangasser DA, Valentino RJ. Sex differences in stress-related psychiatric disorders: neurobiological perspectives. Front Neuroendocrinol 2014 Aug;35(3):303-319 [FREE Full text] [doi: 10.1016/j.yfrne.2014.03.008] [Medline: 24726661]
9. Russo SJ, Murrough JW, Han MH, Charney DS, Nestler EJ. Neurobiology of resilience. Nat Neurosci 2012 Nov 14;15(11):1475-1484 [FREE Full text] [doi: 10.1038/nn.3234] [Medline: 23064380]
10. McEwen BS. Sex, stress and the hippocampus: allostasis, allostatic load and the aging process. Neurobiol Aging 2002;23(5):921-939. [doi: 10.1016/s0197-4580(02)00027-1] [Medline: 12392796]
11. Ulrich-Lai YM, Herman JP. Neural regulation of endocrine and autonomic stress responses. Nat Rev Neurosci 2009 Jun 13;10(6):397-409 [FREE Full text] [doi: 10.1038/nrn2647] [Medline: 19469025]
12. Clark A, Mach N. Exercise-induced stress behavior, gut-microbiota-brain axis and diet: a systematic review for athletes. J Int Soc Sports Nutr 2016 Apr 01;13(1):43 [FREE Full text] [doi: 10.1186/s12970-016-0155-6] [Medline: 27924137]
13. McCraty R, Shaffer F. Heart rate variability: new perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk. Glob Adv Health Med 2015 Jan;4(1):46-61 [FREE Full text] [doi: 10.7453/gahmj.2014.073] [Medline: 25694852]
14. Amano M, Kanda T, Ue H, Moritani T. Exercise training and autonomic nervous system activity in obese individuals. Med Sci Sports Exerc 2001 Aug;33(8):1287-1291 [FREE Full text] [doi: 10.1097/00005768-200108000-00007] [Medline: 11474328]
15. Dong JG. The role of heart rate variability in sports physiology. Exp Ther Med 2016 May;11(5):1531-1536 [FREE Full text] [doi: 10.3892/etm.2016.3104] [Medline: 27168768]
16. Mongin D, Chabert C, Extremera MG, Hue O, Courvoisier DS, Carpena P, et al. Decrease of heart rate variability during exercise: an index of cardiorespiratory fitness. PLoS One 2022 Sep 2;17(9):e0273981 [FREE Full text] [doi: 10.1371/journal.pone.0273981] [Medline: 36054204]
17. Tobaldini E, Nobili L, Strada S, Casali KR, Braghiroli A, Montano N. Heart rate variability in normal and pathological sleep. Front Physiol 2013 Oct 16;4:294 [FREE Full text] [doi: 10.3389/fphys.2013.00294] [Medline: 24137133]
18. Siepmann M, Weidner K, Petrowski K, Siepmann T. Heart rate variability: a measure of cardiovascular health and possible therapeutic target in dysautonomic mental and neurological disorders. Appl Psychophysiol Biofeedback 2022 Dec 22;47(4):273-287 [FREE Full text] [doi: 10.1007/s10484-022-09572-0] [Medline: 36417141]
19. Ahn JW, Ku Y, Kim HC. A novel wearable EEG and ECG recording system for stress assessment. Sensors (Basel) 2019 Apr 28;19(9):23 [FREE Full text] [doi: 10.3390/s19091991] [Medline: 31035399]
20. Attar ET, Balasubramanian V, Subasi E, Kaya M. Stress analysis based on simultaneous heart rate variability and EEG monitoring. IEEE J Transl Eng Health Med 2021;9:1-7. [doi: 10.1109/jtehm.2021.3106803]
21. Shilton AL, Laycock R, Crewther SG. The Maastricht acute stress test (MAST): physiological and subjective responses in anticipation, and post-stress. Front Psychol 2017 Apr 19;8:567 [FREE Full text] [doi: 10.3389/fpsyg.2017.00567] [Medline: 28469586]
22. Frisch JU, Häusser JA, Mojzisch A. The trier social stress test as a paradigm to study how people respond to threat in social interactions. Front Psychol 2015 Feb 02;6:14 [FREE Full text] [doi: 10.3389/fpsyg.2015.00014] [Medline: 25698987]
23. Aspiotis V, Miltiadous A, Kalafatakis K, Tzimourta KD, Giannakeas N, Tsipouras MG, et al. Assessing electroencephalography as a stress indicator: a VR high-altitude scenario monitored through EEG and ECG. Sensors (Basel) 2022 Aug 03;22(15):5792 [FREE Full text] [doi: 10.3390/s22155792] [Medline: 35957348]
24. Ahmad Z, Rabbani S, Zafar MR, Ishaque S, Krishnan S, Khan N. Multilevel stress assessment from ECG in a virtual reality environment using multimodal fusion. IEEE Sensors J 2023 Dec 1;23(23):29559-29570. [doi: 10.1109/jsen.2023.3323290]
25. Bouchard S, Bernier F, Boivin É, Guitard T, Laforest M, Dumoulin S, et al. Modes of immersion and stress induced by commercial (off-the-shelf) 3D games. J Def Model Simul 2012 May 04;11(4):339-352. [doi: 10.1177/1548512912446359]
26. van Dammen L, Finseth TT, McCurdy BH, Barnett NP, Conrady RA, Leach AG, et al. Evoking stress reactivity in virtual reality: a systematic review and meta-analysis. Neurosci Biobehav Rev 2022 Jul;138:104709. [doi: 10.1016/j.neubiorev.2022.104709] [Medline: 35644278]
27. Hemmeter U, Stormer R, Mager R, Kuntze M, Mueller-Spahn F, Hennig J, et al. Modification of virtual reality combined with a mental task stimulates cortisol in volunteers. Neuropsychobiology 2005 May 2;51(3):165-172. [doi: 10.1159/000085208] [Medline: 15838187]
28. Breuninger C, Sláma DM, Krämer M, Schmitz J, Tuschen-Caffier B. Psychophysiological reactivity, interoception and emotion regulation in patients with agoraphobia during virtual reality anxiety induction. Cogn Ther Res 2016 Oct 15;41(2):193-205. [doi: 10.1007/S10608-016-9814-9]
29. Shaigetz VG, Proulx C, Cabral A, Choudhury N, Hewko M, Kohlenberg E, et al. An immersive and interactive platform for cognitive assessment and rehabilitation (bWell): design and iterative development process. JMIR Rehabil Assist Technol 2021 Nov 03;8(4):e26629 [FREE Full text] [doi: 10.2196/26629] [Medline: 34730536]
30. Thayer JF, Hansen AL, Saus-Rose E, Johnsen BH. Heart rate variability, prefrontal neural function, and cognitive performance: the neurovisceral integration perspective on self-regulation, adaptation, and health. Ann Behav Med 2009 Apr 8;37(2):141-153. [doi: 10.1007/s12160-009-9101-z] [Medline: 19424767]

XSL•FO
RenderX

31. Laborde S, Mosley E, Mertgen A. Vagal tank theory: the three Rs of cardiac vagal control functioning - resting, reactivity, and recovery. Front Neurosci 2018 Jul 10;12:458 [FREE Full text] [doi: 10.3389/fnins.2018.00458] [Medline: 30042653]
32. Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for methodology of virtual reality clinical trials in health care by an international working group: iterative study. JMIR Ment Health 2019 Jan 31;6(1):e11973 [FREE Full text] [doi: 10.2196/11973] [Medline: 30702436]
33. O'Cathain A, Croot L, Duncan E, Rousseau N, Sworn K, Turner KM, et al. Guidance on how to develop complex interventions to improve health and healthcare. BMJ Open 2019 Aug 15;9(8):e029954 [FREE Full text] [doi: 10.1136/bmjopen-2019-029954] [Medline: 31420394]
34. Delliaux S, Delaforge A, Deharo JC, Chaumet G. Mental workload alters heart rate variability, lowering non-linear dynamics. Front Physiol 2019 May 14;10:565 [FREE Full text] [doi: 10.3389/fphys.2019.00565] [Medline: 31156454]
35. Low A, McCraty R. Heart rate variability: new perspectives on assessment of stress and health risk at the workplace. Heart Mind 2018;2(1):16-27. [doi: 10.4103/hm.hm_11_18]
36. Eggenberger P, Annaheim S, Kündig KA, Rossi RM, Münzer T, de Bruin ED. Heart rate variability mainly relates to cognitive executive functions and improves through exergame training in older adults: a secondary analysis of a 6-month randomized controlled trial. Front Aging Neurosci 2020 Jul 15;12:197 [FREE Full text] [doi: 10.3389/fnagi.2020.00197] [Medline: 32760267]
37. Grässler B, Hökelmann A, Cabral RH. Resting heart rate variability as a possible marker of cognitive decline. Kinesiol 2020;52(1):72-84. [doi: 10.26582/k.52.1.9]
38. Thayer JF, Lane RD. A model of neurovisceral integration in emotion regulation and dysregulation. J Affect Disord 2000 Dec;61(3):201-216. [doi: 10.1016/s0165-0327(00)00338-4] [Medline: 11163422]
39. Thayer JF, Lane RD. Claude Bernard and the heart-brain connection: further elaboration of a model of neurovisceral integration. Neurosci Biobehav Rev 2009 Mar;33(2):81-88. [doi: 10.1016/j.neubiorev.2008.08.004] [Medline: 18771686]
40. Pavlov VA, Tracey KJ. The vagus nerve and the inflammatory reflex--linking immunity and metabolism. Nat Rev Endocrinol 2012 Dec 21;8(12):743-754 [FREE Full text] [doi: 10.1038/nrendo.2012.189] [Medline: 23169440]
41. Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Circulation 1996 Mar 01;93(5):1043-1065. [Medline: 8598068]
42. Liu CY, Mueller MH, Grundy D, Kreis ME. Vagal modulation of intestinal afferent sensitivity to systemic LPS in the rat. Am J Physiol Gastrointest Liver Physiol 2007 May;292(5):G1213-G1220 [FREE Full text] [doi: 10.1152/ajpgi.00267.2006] [Medline: 17204546]
43. Ozaki N, Sengupta JN, Gebhart GF. Mechanosensitive properties of gastric vagal afferent fibers in the rat. J Neurophysiol 1999 Nov 01;82(5):2210-2220 [FREE Full text] [doi: 10.1152/jn.1999.82.5.2210] [Medline: 10561400]
44. Laborde S, Mosley E, Thayer JF. Heart rate variability and cardiac vagal tone in psychophysiological research - recommendations for experiment planning, data analysis, and data reporting. Front Psychol 2017;8:213 [FREE Full text] [doi: 10.3389/fpsyg.2017.00213] [Medline: 28265249]
45. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health 2017 Sep 28;5:258 [FREE Full text] [doi: 10.3389/fpubh.2017.00258] [Medline: 29034226]
46. Dekker JM, Crow RS, Folsom AR, Hannan PJ, Liao D, Swenne CA, et al. Low heart rate variability in a 2-minute rhythm strip predicts risk of coronary heart disease and mortality from several causes: the ARIC Study. Atherosclerosis risk in communities. Circulation 2000 Sep 12;102(11):1239-1244. [doi: 10.1161/01.cir.102.11.1239] [Medline: 10982537]
47. Grassi G. Assessment of sympathetic cardiovascular drive in human hypertension: achievements and perspectives. Hypertension 2009 Oct;54(4):690-697. [doi: 10.1161/hypertensionaha.108.119883]
48. Ronkainen E, Korpelainen JT, Heikkinen E, Myllylä VV, Huikuri HV, Isojärvi JI. Cardiac autonomic control in patients with refractory epilepsy before and during vagus nerve stimulation treatment: a one-year follow-up study. Epilepsia 2006 Mar 06;47(3):556-562 [FREE Full text] [doi: 10.1111/j.1528-1167.2006.00467.x] [Medline: 16529621]
49. Porges SW. The polyvagal perspective. Biol Psychol 2007 Mar;74(2):116-143 [FREE Full text] [doi: 10.1016/j.biopsycho.2006.06.009] [Medline: 17049418]
50. Smith R, Thayer JF, Khalsa SS, Lane RD. The hierarchical basis of neurovisceral integration. Neurosci Biobehav Rev 2017 May;75:274-296. [doi: 10.1016/j.neubiorev.2017.02.003] [Medline: 28188890]
51. Hart SG, Staveland LE. Development of NASA-TLX (task load index): results of empirical and theoretical research. Adv Psychol 1988;52:139-183. [doi: 10.1016/S0166-4115(08)62386-9]
52. Porges SW. Cardiac vagal tone: a physiological index of stress. Neurosci Biobehav Rev 1995;19(2):225-233. [doi: 10.1016/0149-7634(94)00066-a] [Medline: 7630578]
53. Thayer JF, Åhs F, Fredrikson M, Sollers JJ, Wager TD. A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. Neurosci Biobehav Rev 2012 Mar;36(2):747-756. [doi: 10.1016/j.neubiorev.2011.11.009] [Medline: 22178086]

XSL•FO
RenderX

54. Park G, Vasey MW, Van Bavel JJ, Thayer JF. When tonic cardiac vagal tone predicts changes in phasic vagal tone: the role of fear and perceptual load. Psychophysiology 2014 May 27;51(5):419-426. [doi: 10.1111/psyp.12186] [Medline: 24571084]
55. Porges SW, Doussard-Roosevelt JA, Portales AL, Greenspan SI. Infant regulation of the vagal "brake" predicts child behavior problems: a psychobiological model of social behavior. Dev Psychobiol 1996 Dec;29(8):697-712. [doi: 10.1002/(SICI)1098-2302(199612)29:8<697::AID-DEV5>3.0.CO;2-O] [Medline: 8958482]
56. Porges SW. A phylogenetic journey through the vague and ambiguous Xth cranial nerve: a commentary on contemporary heart rate variability research. Biol Psychol 2007 Mar;74(2):301-307 [FREE Full text] [doi: 10.1016/j.biopsycho.2006.08.007] [Medline: 17055142]
57. Crameri L, Hettiarachchi IT, Hanoun S. Effects of dynamic resilience on the reactivity of vagally mediated heart rate variability. Front Psychol 2020 Jan 20;11:579210 [FREE Full text] [doi: 10.3389/fpsyg.2020.579210] [Medline: 33551903]
58. Beauchaine TP, Gatzke-Kopp L, Mead HK. Polyvagal Theory and developmental psychopathology: emotion dysregulation and conduct problems from preschool to adolescence. Biol Psychol 2007 Mar;74(2):174-184 [FREE Full text] [doi: 10.1016/j.biopsycho.2005.08.008] [Medline: 17045726]
59. Fisher JP. Autonomic control of the heart during exercise in humans: role of skeletal muscle afferents. Exp Physiol 2014 Mar 02;99(2):300-305 [FREE Full text] [doi: 10.1113/expphysiol.2013.074377] [Medline: 23995102]
60. Fu Q, Levine BD. Exercise and the autonomic nervous system. Handb Clin Neurol 2013;117:147-160. [doi: 10.1016/B978-0-444-53491-0.00013-4] [Medline: 24095123]
61. Kimhy D, Crowley O, McKinley PS, Burg M, Lachman M, Tun P, et al. The association of cardiac vagal control and executive functioning--findings from the MIDUS study. J Psychiatr Res 2013 May;47(5):628-635 [FREE Full text] [doi: 10.1016/j.jpsychires.2013.01.018] [Medline: 23434176]
62. Proulx C, Cabral A, Choudhury N, Debergue P. Acceptability study of a novel immersive cognitive care platform for remediation of cognitive deficits. Annu Rev Cyber Therapy Telemed 2018;16:43-49.
63. Hernandez Hernandez ME, Michalak E, Choudhury N, Hewko M, Torres I, Menon M, et al. Co-design of a virtual reality cognitive remediation program for depression (bWell-D) with patient end users and clinicians: qualitative interview study among patients and clinicians. JMIR Serious Games 2023 May 07;11:e43904 [FREE Full text] [doi: 10.2196/43904] [Medline: 37027183]
64. Cabral A, Choudhury N, Proulx C, Harmouche R, Kohlenberg E, Deberguevaluation OC. Evaluation of cybersickness in a passive walking virtual reality cognitive exercise. Annu Rev CyberTherapy Telemed 2019;17:59-64 [FREE Full text]
65. Quintana DS, Heathers JA. Considerations in the assessment of heart rate variability in biobehavioral research. Front Psychol 2014 Jul 22;5:805 [FREE Full text] [doi: 10.3389/fpsyg.2014.00805] [Medline: 25101047]
66. Gnacek M, Broulidakis J, Mavridou I, Fatoorechi M, Seiss E, Kostoulas T, et al. emteqPRO—fully integrated biometric sensing array for non-invasive biomedical research in virtual reality. Front Virtual Real 2022 Mar 11;3:35. [doi: 10.3389/frvir.2022.781218]
67. Keebler Assoc JR, Shelstad WJ, Google DC, Chaparro BS, Phan Google MH. Validation of the GUESS-18. J Usability Stud 2020;16(1):49-62. [doi: 10.5555/3532747.3532752]
68. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. London, UK: CRC Press; 1996.
69. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aerosp Psychol 1993 Jul;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
70. Bouchard S, Robillard G, Renaud P. Revising the factor structure of the simulator sickness questionnaire. Annu Rev CyberTherapy Telemed 2007;5:128-137.
71. Bouchard S, St-Jacques J, Renaud P, Wiederhold B. Side effects of immersions in virtual reality for people suffering from anxiety disorders. J Cyber Ther Rehabil 2009;2(2):127-137 [FREE Full text]
72. Carman WA. Validation of the GUESS-18 for the usability of a virtual reality racing game. LSU Master's Theses. URL: https://digitalcommons.lsu.edu/gradschool_theses/5606 [accessed 2023-06-14]
73. Kleiger RE, Stein PK, Bigger JT. Heart rate variability: measurement and clinical utility. Ann Noninvasive Electrocardiol 2005 Jan;10(1):88-101 [FREE Full text] [doi: 10.1111/j.1542-474X.2005.10101.x] [Medline: 15649244]
74. van Gent P, Farah H, van Nes N, van Arem B. HeartPy: a novel heart rate algorithm for the analysis of noisy signals. Transp Res Part F Traffic Psychol Behav 2019 Oct;66:368-378. [doi: 10.1016/j.trf.2019.09.015]
75. Watland S, Nes LS, Hanson E, Ekstedt M, Stenberg U, Børøsund E. The caregiver pathway, a model for the systematic and individualized follow-up of family caregivers at intensive care units: development study. JMIR Form Res 2023 May 25;7:e46299 [FREE Full text] [doi: 10.2196/46299] [Medline: 37097744]
76. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit Med 2019 Sep 6;2(1):88 [FREE Full text] [doi: 10.1038/s41746-019-0166-1] [Medline: 31508498]
77. McIntosh AR, Lobaugh NJ. Partial least squares analysis of neuroimaging data: applications and advances. Neuroimage 2004;23 Suppl 1:S250-S263. [doi: 10.1016/j.neuroimage.2004.07.020] [Medline: 15501095]

XSL•FO
RenderX

78. Khundrakpam B, Booij L, Jeon S, Karama S, Tohka J, Evans AC. Individualized prediction of future cognition based on developmental changes in cortical anatomy. NeuroImage Rep 2022 Dec;2(4):100127. [doi: 10.1016/j.ynirp.2022.100127]

Abbreviations

ANS: autonomic nervous system
CPT: continuous performance test
ECG: electrocardiogram
EMG: electromyography
GUESS: game user experience satisfaction scale
HRV: heart rate variability
ML: machine learning
NASA-TLX: NASA-Task Load Index
NVIM: Neurovisceral Integration Model
PPG: photoplethysmography
RMSSD: root mean square of successive differences
SSQ: Simulator Sickness Questionnaire
UI: user interface
VR: virtual reality
VTT: Vagal Tank Theory

Edited by A Mavragani; submitted 18.07.24; peer-reviewed by N Khan; comments to author 16.10.24; revised version received 06.11.24; accepted 27.11.24; published 04.03.25.

Please cite as:
Khundrakpam B, Segado M, Pazdera J, Gagnon Shaigetz V, Granek JA, Choudhury N
An Integrated Platform Combining Immersive Virtual Reality and Physiological Sensors for Systematic and Individualized Assessment of Stress Response (bWell): Design and Implementation Study
JMIR Form Res 2025;9:e64492
URL: https://formative.jmir.org/2025/1/e64492
doi:10.2196/64492
PMID:40053709

©Budhachandra Khundrakpam, Melanie Segado, Jesse Pazdera, Vincent Gagnon Shaigetz, Joshua A Granek, Nusrat Choudhury. Originally published in JMIR Formative Research (https://formative.jmir.org), 04.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Long-Term Experiences of Health Care Providers Using Iris Scanning as an Identification Tool in a Vaccine Trial in the Democratic Republic of the Congo: Qualitative Study

Trésor Zola Matuvanga[1,2,3], PhD; Antea Paviotti[2,3], PhD; Freddy Bikioli Bolombo[1,2,3], MD; Gwen Lemey[2,3], MA; Ynke Larivière[2,3], PhD; Maha Salloum[2,3], MSc; Bernard Isekah Osang'ir[2,3], MSc; Emmanuel Esanga Longomo[4], MD; Solange Milolo[1], MD; Junior Matangila[1], PhD; Vivi Maketa[1], PhD; Patrick Mitashi[1], PhD; Pierre Van Damme[2], PhD; Hypolite Muhindo-Mavoko[1], PhD; Jean-Pierre Van geertruyden[3], PhD

[1]Department of Tropical Medicine, Faculty of Medicine, University of Kinshasa, Kinshasa, the Democratic Republic of the Congo

[2]Centre for the Evaluation of Vaccination, Vaccine and Infectious Disease Institute, University of Antwerp, Wilrijk, Belgium

[3]Global Health Institute, Department of Family Medicine and Population Health, University of Antwerp, Wilrijk, Belgium

[4]Division Provinciale de la Santé, Ministry of Health, Tshuapa Province, the Democratic Republic of the Congo

Corresponding Author:
Trésor Zola Matuvanga, PhD
Department of Tropical Medicine
Faculty of Medicine
University of Kinshasa
Avenue Université, N° 1, Commune de Lemba
Kinshasa, BP 834 KINSHASA XI
the Democratic Republic of the Congo
Phone: 243 810046306
Email: zola.matuvanga@unikin.ac.cd

Abstract

Background: Iris scanning has increasingly been used for biometric identification over the past decade, with continuous advancements and expanding applications. To better understand the acceptability of this technology, we report the long-term experiences of health care providers and frontline worker participants with iris scanning as an identification tool in the EBL2007 Ebola vaccine trial conducted in the Democratic Republic of the Congo.

Objective: This study aims to document the long-term experiences of using iris scanning for identity verification throughout the vaccine trial.

Methods: Two years after the start of the EBL2007 vaccine trial (February to March 2022), 69 trial participants—including nurses, first aid workers, midwives, and community health workers—were interviewed through focus group discussions. Additionally, 13 in-depth individual interviews were conducted with physicians involved in the trial, iris scan operators, trial staff physicians, and trial participants who declined iris scanning. Qualitative content analysis was used to identify key themes.

Results: Initially, interviewees widely accepted the iris scan and viewed it as a distinctive tool for identifying participants in the EBL2007 vaccine trial. However, over time, perceptions became less favorable. Some participants expressed concerns that their vision had diminished shortly after using the tool and continued to decline until the end of the study. Others reported experiencing perceived vision loss long after the trial had concluded. However, no vision impairment was reported as an adverse event or assessed in the trial as being linked to the iris scan, which uses a previously certified safe infrared light for scanning.

Conclusions: Our findings highlight the sustained acceptability and perceived high accuracy of the iris scan tool for uniquely identifying adult participants in a vaccine trial over time. Continued efforts to systematically disseminate and reinforce information about the function and safety of this technology are essential. Clearly presenting iris scanning as a safe procedure could help dispel misconceptions, concerns, and perceived risks among potential users in vaccine trials.

(JMIR Form Res 2025;9:e54921) doi:10.2196/54921

XSL•FO
RenderX

KEYWORDS

iris scan; vaccine trial; iris; perception; experience; views; biometric identification; Democratic Republic of the Congo

Introduction

In many low- and middle-income countries (LMICs), the lack of reliable patient identification can hinder the effectiveness of routine medical care [1-3]. In rural settings, the digitization of personal information and its integration into databases remain limited. Additionally, challenges in identifying patients enrolled over long periods (eg, in longitudinal cohort studies) can lead to misclassification and pose a significant barrier to maintaining study data integrity. A recurring issue during (un)scheduled study visits is that participants often lack a consistent form of identification—whether due to misplacement, forgetfulness, or deterioration. Recently, biometric recognition has gained importance as a tool for accurately identifying individuals in routine health information processes [4-6]. In clinical trials, the use of human biometric recognition has also expanded over the past decade, with adoption becoming increasingly widespread due to its advantages over traditional identification methods [4,7].

Biometric recognition involves a process that relies on the technical processing of data related to the physical, physiological, or behavioral characteristics of the human body (including movement) for authentication purposes [8]. Among the biometric recognition methods used to date—such as fingerprints, facial recognition, iris scans, ear biometrics, and voice recognition—fingerprints are the most widely utilized and have become the predominant form of biometric data [9]. However, unlike fingerprints, which lose clarity over time due to the widening of ridge gaps, the distinct patterns within the iris remain unchanged throughout a person's lifetime, making iris scan identification exceptionally stable [10-12]. Even among identical twins or between an individual's own eyes, iris patterns are unique, ensuring their suitability as a reliable proxy for identification [13]. Additionally, iris scanning offers the advantage of being a noncontact process, eliminating the risk of disease transmission through direct contact. As a result, it has the potential to become an affordable, fast, and reliable identification tool across various contexts, including electoral voting, access control, and vaccine trials [4,5,14,15]. In clinical trials, the use of iris scans could help prevent errors and fraudulent entries while ensuring unique participant identification across (un)scheduled visits, thereby safeguarding the integrity of trial results [5]. However, experience with iris scanning as a method of identity verification in clinical trials remains limited.

Recently, the implementation of the iris scan tool demonstrated high precision and strong acceptance among health care providers and frontline worker participants (Multimedia Appendix 1), who were interviewed both before and at enrollment in an Ebola vaccine trial (hereafter referred to as the "EBL2007 vaccine trial") conducted in a remote area of the Democratic Republic of the Congo (DRC) [4]. This new qualitative research aimed to document long-term experiences with iris scanning for identity verification throughout the trial.

Methods

Study Design

This study used a qualitative methodology, specifically a phenomenological approach, to explore experiences related to the long-term use of iris scanning as an identification tool during (un)scheduled visits throughout the EBL2007 vaccine trial.

Research Team and Reflexivity

The research team comprised TZM and FBB, both male medical doctors and PhD candidates at the University of Kinshasa (DRC), under the coordination of AP, a female social scientist (PhD) from the University of Antwerp (Belgium). TZM and AP developed the interview and topic guides. TZM and FBB conducted the interviews and focus group discussions (FGDs), also facilitating translation between French and Lingala and taking notes as needed. EEL, a male MD from the Provincial Health Division in Boende (Ministry of Health), provided logistical support and assisted with recruitment. Previously, TZM and EEL also served as subinvestigators and trial site coordinators for the EBL2007 vaccine trial.

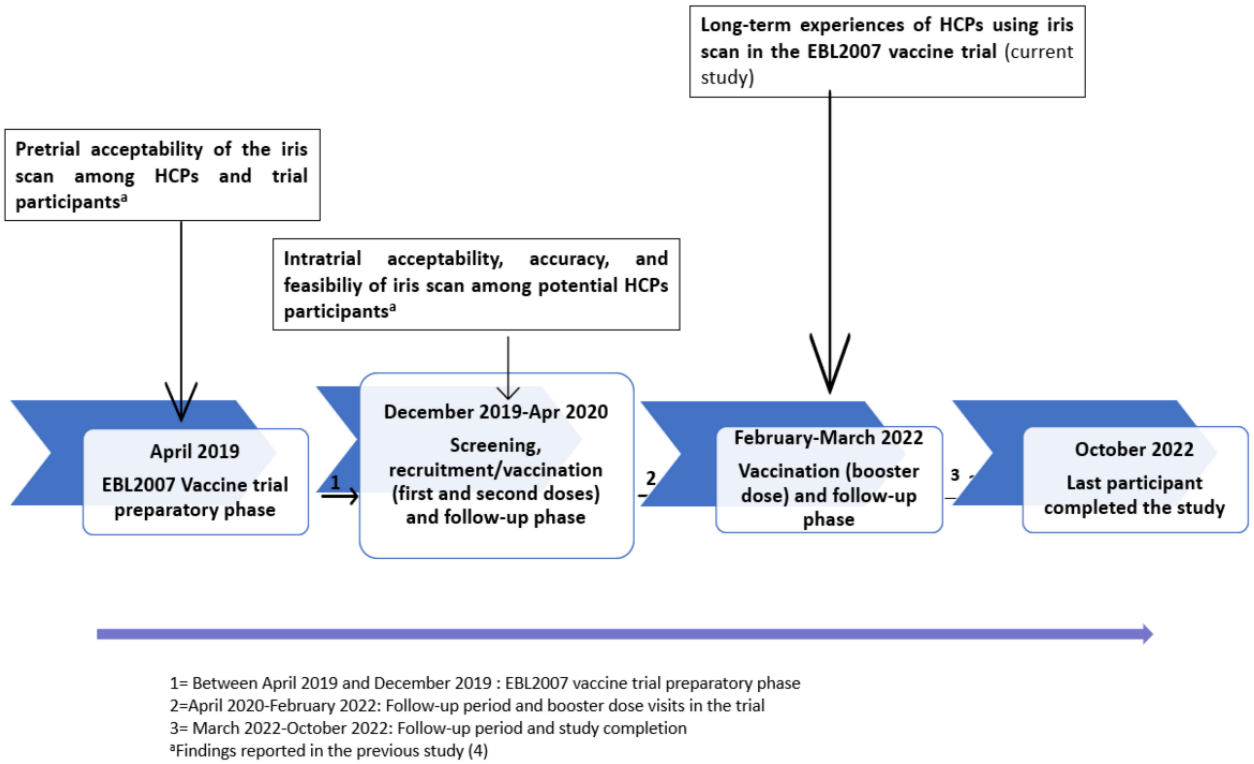
Study Setting and Time

This qualitative study on iris scan perception was nested within the EBL2007 vaccine trial (the main study) and conducted in the Health District of Boende, located 1200 km from Kinshasa, the capital of the DRC. Conversations took place on the premises of the General Referral Hospital of Boende (Hôpital Général de Référence de Boende, HGR), which served as the main trial site.

Participant recruitment for the main study began in December 2019, with the final patient visit occurring in October 2022 (Figure 1). During the screening process, potential vaccine trial participants were invited to opt into an innovative identification method—an iris scan that captures biometric data from the iris for personal identification. They were informed that this tool was safe and would enable more accurate recognition at both scheduled and unscheduled trial visits.

XSL•FO
RenderX

Figure 1. Chronology of events in the EBL2007 vaccine trial. HCPs: health care providers.



This qualitative study was conducted between February and March 2022, during the second follow-up period of the main study, when most participants had already completed their last scheduled visit.

Long-term experiences refer to users' perceptions of the operational viability of the iris scan after 2 years of systematic use within the EBL2007 vaccine trial. Further details on the vaccine trial are reported elsewhere [16].

Participants' Recruitment and Sampling

In this qualitative study, purposive sampling was used to include a diverse group of participants and staff from the EBL2007 vaccine trial as it neared completion. The study included the following groups: participants who used iris scanning for identification, participants who initially opted out of iris scanning (including various health care providers involved in the main study, such as nurses, doctors, midwives, first aid workers, community health workers, and health facility cleaners), physician staff involved in the trial, and iris scan operators. Potential study participants were purposively selected through referrals based on their professional category, their consent or refusal to use the iris scanning system, and their level of involvement in its implementation during the main study.

Data Collection Methods

A combination of FGDs and in-depth individual interviews (IDIs) was applied to collect data from these participants.

Trial participants who had used iris scanning as their identification method from the outset were invited to participate in FGDs. These participants represented a diverse range of professional categories, including community health care workers, who serve as intermediaries between health services and the local population by disseminating information, providing health education, identifying early disease cases, promoting positive health practices, and encouraging individuals to seek professional health care. First aid workers, trained to provide immediate care before professional medical services become available, were also included, along with midwives, nurses, and health facility cleaners. To facilitate open dialogue within each FGD, considerations were given to both the sex and professional categories of the participants.

Furthermore, IDIs were conducted to gather insights from trial participants who had initially opted against using iris scanning for identification. Physicians involved as trial participants and iris scanning operators were also invited to participate in these IDIs.

The selection of trial participants who declined iris scanning was based on documented refusals, as established in prior research [4]. The primary reason for refusing iris scanning among these participants was concern about ocular safety.

Procedure

For both IDIs and FGDs, a semistructured questionnaire was used (Multimedia Appendix 2). FGDs were conducted in groups of 6-10 participants. Before commencing the IDIs or FGDs, interviewees were reminded of the study's purpose. Each interview lasted between 60 and 90 minutes, and the conversations were recorded after obtaining the interviewees' consent.

Voice recordings in languages other than French were translated into French and then into English before being transcribed by 2 independent individuals. The accuracy and coherence of the transcriptions were thoroughly verified by AP and TZM. TZM coded all transcripts derived from both FGDs and IDIs, and AP

XSL•FO
RenderX

and TZM reached a consensus on the coding system and categories used for thematic analysis.

Analysis

This study follows the Consolidated Criteria for Reporting Qualitative Research (COREQ) guidelines to ensure rigor and transparency in the design, data collection, analysis, and reporting of findings (Multimedia Appendix 3). The audio recordings of all conversations were transcribed, translated into French if necessary, and imported into NVivo software (QSR International) for analysis. Each transcript was coded and assigned a unique identifier to ensure confidentiality. AP and TZM held regular collaborative meetings to ensure consistency in the coding process.

The data underwent a thematic analysis approach, beginning with an initial phase of familiarization with the data set. The coding process combined both inductive and deductive approaches [17,18]. Initially, a deductive approach was applied, with codes derived directly from the 2 main themes predetermined in the interview guide: (1) the acceptability of the iris scan and (2) knowledge, perception, and use of the iris scan. The initial codes addressing similar themes were grouped into these predetermined themes using a start list [19]. As the analysis progressed, an inductive approach was incorporated, allowing for the identification of new, unanticipated codes. The start list was subsequently updated to integrate these emergent codes. An iterative process was used to refine and develop subthemes and themes, ensuring alignment with the study objectives. The final themes were substantiated by significant quotations from the transcripts. The analysis was conducted in French to preserve the authenticity and nuance of participants' responses.

Trustworthiness

To ensure trustworthiness in our qualitative analysis, we implemented multiple strategies aligned with the criteria outlined by Creswell and Poth [20], as well as the principles proposed by Braun and Clarke [21]. Specifically, credibility was strengthened through source triangulation by integrating data from IDIs, FGDs, and observations across different professional and gender roles within the trial. This methodological approach facilitated the comparison of diverse perspectives, ensuring a more comprehensive understanding of participants' experiences.

Reflexivity was explicitly addressed in a dedicated section that examined the roles of the investigators throughout the research process. Additionally, verbatim quotes from participants were incorporated into the text to preserve the authenticity of their perspectives and enhance the depth of the analysis.

To enhance transferability, we explicitly detailed our working process and methodology, enabling other researchers to understand and potentially replicate the conditions and results of our study.

Confirmability was strengthened by using multiple data collection methods and engaging independent transcribers and translators to ensure neutrality. Each transcript was meticulously reviewed by the researchers to confirm the accuracy and minimize researcher bias.

Finally, dependability was ensured by maintaining a consistent methodological approach throughout the study, with exhaustive documentation to allow verification of the research process.

We believe that the aforementioned measures provide a solid framework to ensure the reproducibility of our findings.

Iris Scan Equipment and Procedures

The operator responsible for the iris scanning procedure was a trained and authorized EBL2007 vaccine trial staff member. The iris scan was conducted using an iris camera (Iritech; Irishield Monocular) in conjunction with a tablet (Samsung Tab Active 2) connected to a local ruggedized server via Wi-Fi. Further details regarding the procedures are described elsewhere (Multimedia Appendix 4) [4].

The iris scanner used in the EBL2007 vaccine trial was certified as safe for use with infrared light under all operating conditions, in accordance with the international standard IEC (International Electrotechnical Commission) 62471:2006-07 [22]. Additionally, its irradiance was less than 2% of the Eye Safety Standard Regulation and was tested for photobiological safety.

At enrollment, a list of trial participants who consented to the use of the iris scanning tool for identification during trial visits was compiled. Based on this list, the iris scan operator was instructed not to perform iris scans on participants who had not provided consent. For these individuals, an alternative identification method was used. Their unique participant identification number, assigned at trial enrollment, was recorded, and only their demographic data and identification photo were captured using the iris scanning tool's tablet. During subsequent (un)scheduled visits, entering their identification number into the tablet allowed for participant recognition based on the stored photograph. For consenting individuals, the operator collected demographic data, took an identification photo, and scanned both irises (left and right eye) as the primary method of identification.

Ethics Approval

This research was conducted in accordance with ethical principles designed to protect the rights and welfare of all participants. Approval to conduct the study was granted by the National Ethics Committee of Health of the DRC (approval reference number 368/CNES/BN/PMMF/2022). Explicit oral informed consent was obtained before any data collection, following a clearly defined protocol. Interviewees were fully informed about the study's objectives, the use of their personal data, potential risks and benefits, and their rights to confidentiality and voluntary participation. Consent was recorded, and discussions proceeded only after confirmation. This study had its own protocol and was not a secondary analysis of existing data; therefore, the informed consent process was specific to this research. To protect participants' privacy and confidentiality, all data were anonymized during transcription and analysis. Unique identifiers were used in place of personal information, and no identifying details were included in the reported findings. Interviewees did not receive monetary compensation for their participation. However, refreshments and transportation allowances were provided to facilitate their involvement in FGDs and IDIs. This was disclosed to

XSL•FO
RenderX

participants before they provided consent and was approved by the National Ethics Committee.

Results

Participant Recruitment and Data Saturation

Interviews were concluded once data saturation was achieved, with a total of 82 trial participants and staff included in the study. Data saturation was determined when no new themes emerged from subsequent discussions. A total of 69 trial participants took part in FGDs, representing community health care workers, first aid workers, midwives, nurses, and health facility cleaners (Table 1). Additionally, IDIs were conducted with 6 trial participants who refused iris scanning, 5 medical doctors responsible for safety monitoring in the trial, and 2 iris scanning operators.

Table 1. Data collection activities for the long-term experiences of health care providers using iris scanning in the EBL2007 vaccine trial.

Interviewees and data collection method	Number of activities, n	Interviewees occupation	Male (n=42), n	Female (n=40), n	Total (N=82), n
Trial participants					
Focus group discussion	2	Nurses	10	8	18
Focus group discussion	2	First aid workers	9	8	17
Focus group discussion	2	Community health workers	8	7	15
Focus group discussion	1	Midwife	0	9	9
Focus group discussion	1	Health facility cleaner	2	8	10
In-depth individual interview	1	Physician	3	0	3
In-depth individual interview	6	Entered the trial but refused iris scan (participants A, B, C, D, E, and F)[a]	6	0	6
Staff in the trial					
In-depth individual interview	2	Physician monitoring safety worker	2	0	2
In-depth individual interview	2	Iris scan operator	2	0	2

[a]Participants referred to as A, B, C, D, E, and F represent those who chose not to use the iris scanner for identification in the EBL2007 vaccine trial.

Participants had an average age of 51 (SD 11) years. At least one representative from each main professional category enrolled in the EBL2007 vaccine trial participated in either the IDIs or FGDs. The study included 40 (49%) female EBL2007 vaccine trial participants and staff (N=82; Table 1). Three key themes emerged from the collected data: (1) long-term experiences of using iris scanning as an identification tool in the EBL2007 vaccine trial, (2) the potential use of iris scanning in future vaccine trials, and (3) comparisons between iris scanning and previously known identification tools. Within each theme, additional subthemes and categories were identified.

Long-Term Experiences of Health Care Providers Using Iris Scan as an Identification Tool in the EBL2007 Vaccine Trial

Purpose-of-Use Understanding of Trial Participants Who Opted to Use Iris Scan

Some respondents demonstrated a clear understanding of the rationale behind using iris scanning, as reflected in the following statements:

> *It helped us because it brought out the whole face.* [FGD, midwife, trial participant, woman]

> *It was for identification purposes.* [FGD, midwife, trial participant, woman]

However, a few respondents had a different understanding of the iris scanning tool, perceiving it as a method for vaccine trial investigators to detect disease or assess the impact of the experimental vaccine on the eyes.

> *Me, I believed they're going to find the disease in the eyes, and they're going to tell us, but they haven't told us, and we haven't asked.* [FGD, midwife, trial participant, woman]

Acceptability of the Iris Scan

In general, health care providers and frontline workers volunteering in the EBL2007 vaccine trial widely accepted the use of iris scan technology, consistent with findings from the start of the trial [4]. One community health worker mentioned:

> *[...] I haven't come across anyone who would tell me that they didn't accept it. You see, I haven't encountered any group of people or any individual who would refuse to be examined by this device.* [FGD, community health worker, trial participant, man]

Reasons for Accepting the Iris Scan

The primary reason most interviewees accepted the iris scan was their willingness to receive the study vaccine. Their confidence in the trial staff further motivated them to trust the procedures proposed in the trial.

> *We accepted it because we were looking at our study vaccine, we were looking at the advantage of being enrolled in the Ebola vaccine trial.* [FGD, nurse, trial participant, man]

XSL•FO
RenderX

However, some respondents stated that they adopted the iris scan because they understood it as the appropriate identification tool for accurately recognizing participants enrolled in the trial.

> *We have observed that it is a valuable tool for identification, which is why it was accepted.* [FGD, first aid worker, trial participant, woman]

Reasons for Not Accepting the Iris Scan

A handful of interviewees indicated that they did not feel comfortable with the tool due to concerns about the retention of their demographic data recorded by it.

> *Yes, for a psychological reason, for example, we might take this photo and put it in a book, in a documentary, so that people can see you. So that's why I refused. But as I didn't have any information, I couldn't accept.* [FGD, community health worker, trial participant, man]

Some interviewees, due to the condition of their eyes before the use of the iris scan, expressed a fear of vision loss associated with the tool, compounded by their apprehension upon seeing it for the first time.

> *I refused because I'm sick. My eye hurts, especially my left eye, which has been bothering me since 2012. I've been suffering for four years. It's in this sense that I refused because the way it was flashing, it may burn my eye. Especially when she filmed it, it scared me. That's why I refused.* [Interview, trial participant B who refused iris scan]

Purpose-of-Use Understanding of Trial Participants Who Declined the Use of Iris Scan

Some interviewees who declined iris scan identification reported that they were unaware of the rationale for using this tool in the vaccine trial.

> *Well, I don't know the importance of this.* [Interview, trial participant A who refused iris scan]

> *I still have doubts, but I've seen the people who have had iris scans, and I don't see the point in continuing to doubt, given that there are friends who have accepted, and they're still here.* [Interview, trial participant E who refused iris scan]

> *In my opinion, regarding the importance of iris scan, when they captured us, I explained that my eye is diseased, but they told me they would capture the other eye, and I refused. But I don't know the significance of it.* [Interview, trial participant F who refused iris scan]

Some interviewees reported understanding that the scanning tool was intended for identification purposes during visits to prevent possible fraud.

> *[...] We mentioned that we would use it to confirm the person's identity. If someone else tries to pretend to be me, when they put the iris scan in their eye, it will show that it is not them. That is what I remember about the iris scan.* [Interview, trial participant E who refused iris scan]

The Perceived Accuracy of Iris Scan in Identifying Participants

Most interviewees reported that they found the iris scan tool in the trial to be highly accurate, as it helped prevent cases of fraud among participants. They provided examples of fraud attempts to illustrate the tool's effectiveness and accuracy. One such example is the following statement:

> *[...] There was an incident where I arrived for a scheduled visit and the person sitting close to me presented his father's identification card, claiming to be him. Upon verification, the operator discovered that the person in front of him was not who he claimed to be. The photo in the system of iris scan did not match his appearance. He eventually admitted that his father had passed away. The individual conducting the check informed him that we had not been informed of the father's passing, and upon further investigation, it was revealed that the father was traveling and had authorized his son to represent him. This incident revealed potentially fraudulent activity, as the person attempted to manipulate the results of the vaccine trial by assuming someone else's identity.* [FGD, community health care worker, trial participant, woman]

Perceived Risk of Iris Scan Identification

Some participants perceived a link between the use of the iris scan and vision loss. Additionally, most participants expressed concerns that vision loss might occur later due to the amount of light emitted by the device.

Certain interviewees compared the light emitted by the iris scan to the eclipse phenomenon or sunlight, suggesting that this may have contributed to vision loss.

> *We were scared, there was the light and there was uncertainty as to whether the eye would crack or not [...]* [FGD, community health care worker, trial participant, man]

> *Thank you, now after the iris scan, I have noticed that there is a reduction of vision, especially for reading, so we have to use glasses now.* [FGD, first aid worker, trial participant, man]

Certain interviewees mentioned that they did not observe any unusual occurrences during the scanning process and expressed no apprehension about the future safety of their eyes, as they did not perceive the iris scan as hazardous.

> *There was no reaction. They just tell you to stare like this and then they tell you, it's okay. There wasn't really any direct reaction like that.* [FGD, nurse, trial participant, woman]

> *I was just afraid for my eyes, but it is not dangerous. Even if it comes back to my village, we will make people aware of this device. It is for identifying people.* [Interview, trial participant E who refused iris scan]

XSL•FO
RenderX

Rumors and Reactions From the Surroundings

Certain interviewees and FDG expressed concerns about the well-being of their eyes based on conversations. Nevertheless, most emphasized that they had not heard any rumors related to the iris scan. Instead, the rumors primarily focused on the experimental vaccine and other study procedures, such as blood sampling.

> *They said your eye is sick if you use this device, your eye will be completely damaged, which is why I was afraid. Otherwise, there wasn't much to it.* [Interview, trial participant B who refused iris scan]

> *No, in the neighbourhoods there hadn't been any rumours, but it was about the vaccine and blood sampling that people were talking nonsense about, not about the iris scan.* [FGD, first aid worker, trial participant, man]

> *The people around us didn't know that we were having the iris scan in the study, they only knew that we were selling our blood and getting vaccinated, period, but concerning the Iris scan, nothing was said, it was only us, trial participants, who knew about the Iris scan, but not the community, they didn't know anything about it. In the neighbourhood, we were nicknamed blood sellers.* [FGD, health facility cleaner, trial participant, woman]

Use of Iris Scans in Future Vaccine Trials or Other Public Health Activities

Acceptability of Iris Scan in the Wider Community

As the iris scan tool is part of a vaccination monitoring system that collects additional data (eg, demographics, a photo of the face) beyond those recorded for routine vaccinations (eg, demographics, previous vaccinations, the name of the vaccine administered, its lot number, and expiry date), certain participants believed that the iris scan tool would neither be accepted nor feasible in broader vaccination efforts involving the general population. Participants referenced a yellow fever vaccination campaign that had taken place in Boende. During this campaign, people agreed to receive the vaccine, but many were unwilling to provide full identities and demographic data. Therefore, some interviewees and FGD participants suggested that a different tool, such as fingerprint scanning, would be preferable instead.

> *I wanted to say that for the population, it's going to be a bit difficult, because we've noticed here that with yellow fever, we only recorded the name on the card and then gave the vaccine directly. It was also difficult to get someone to agree to give their full identity so that they could be vaccinated, so it would also be very difficult with the iris scan. It's better even with the fingerprint, maybe it will be all right. With the iris scan, it will be a bit difficult with this population.* [FGD, nurse, trial participant, woman]

Some interviewees emphasized that iris scanning could be an acceptable, effective, and reliable tool for uniquely identifying individuals who might volunteer in future clinical trials. However, they stressed that this should be accompanied by a robust awareness campaign to repeatedly communicate sufficient information about the tool's safety.

> *At first, people will refuse, but after awareness-raising and testimonials from those who have experienced the tool, they will accept.* [Interview, trial participant A who refused iris scan]

Recommendations From Interviewees and FGD Participants

When implementing iris scanning in vaccine trials, particularly in remote areas such as Boende, some interviewees and FGD participants recommended considering the availability of ophthalmic specialists. According to their statements, the iris scan operator in the EBL2007 vaccine trial was not fully aware of the risks involved in scanning the eyes and was unable to provide clear explanations regarding the safety of trial participants' eyes.

> *[...] when you come to scan people's eyes, come with the eye specialist. [...]. All those who have handled our eyes are not specialists. [...] They are photographers, so you should to come with eye specialists. An ophthalmologist because it's a sensitive organ.* [FGD, first aid worker trial participant, trial participant, man]

Few of those who participated in the FGDs voiced concerns about not seeing the vaccine trial investigators undergo iris scanning.

> *[...] Until now we haven't seen the staff being vaccinated, or scanned the eyes with iris scan. We haven't seen; they haven't scanned themselves.* [FGD, first aid worker, trial participant, man]

Some participants in this research recommended reducing the amount of light used during scanning, increasing the distance between the eye and the scanner, and conducting demonstrations during the screening/consent process.

> *I think that, as my colleague the Community health worker just said, the distance from the iris scan is too close. Isn't there some way of finding ways of making it even bigger? [...].* [FGD, community health worker, trial participant, woman]

Comparison of Iris Scan With Previous Known Identification Tools

Referring to previously used and well-known identification methods, most interviewees stated that iris scanning would be the best way to uniquely identify volunteers in the trial.

> *With the experience that I have, with the age that I have...I believe that the only method of escaping fraud is scanning [...]* [FGD, community health worker, trial participant, woman]

> *[...] So, with today's technology, we may easily modify the photo by taking someone's face and putting it on another body to make it look like it's me, but it's not. But with the iris scan, it's easy to see that it's not me, it's just someone else's face. So, with the iris scan,*

XSL•FO
RenderX

> *it's hard to commit fraud.* [FGD, community health worker, trial participant, woman]

Some interviewees suggested using the traditional fingerprint biometric tool to minimize the risk of compromising their eyes with iris scanning or deploying other methods, such as recording names, dates of birth, and identification numbers in a computer. Others proposed more innovative identification tools, such as collecting identity data through the laser thermometer used to measure temperature at entry points or utilizing blood samples already collected during the first visit.

> *I'm going to recommend the fingerprint because the signature may be imitated, but your fingerprint, your own blood, will reveal all your data.* [FGD, community health worker, trial participant, man]

> *[...] instead of using this device, there was no way of using our fingerprints. Because...Isn't it possible to use a fingerprint?* [FGD, community health worker, trial participant, man]

> *[...] As we've received all the doses of vaccine as well as blood samples were taken, it's the computer that will indicate that for such and such a participant, he's finished his doses, these appointments are over, so that's it. The computer is a method.* [FGD, health facility cleaner, trial participant, woman]

> *[...] We may have a thermometer that records all the identity as well as the blood pressure and everything, so as not to have any problems with the use of that iris scan laser.* [FGD, community health worker, trial participant, man]

Discussion

Principal Findings

This qualitative study aimed to document the long-term experiences of EBL2007 vaccine trial participants and staff regarding the use of an innovative iris scan biometric tool. Overall, the tool was found to be acceptable, accurate, and effective in verifying participants' identities throughout the trial, preventing fraud and errors.

Although it was clearly explained during the consent procedure that the iris scan was noncompulsory, some participants may have feared that declining it would prevent their enrollment. Similarly, the safety of the iris scan was thoroughly explained to trial personnel at the start of the trial. Despite these efforts, some interviewees and FGD participants still felt that scanning their eyes posed safety risks or that it might cause problems in the future.

It is important to highlight the expected motivational benefits of participating in the EBL2007 vaccine trial, as some interviewees may have accepted the iris scan primarily to receive the Ebola vaccine regimen, given that they reside in an area at risk of an outbreak [23], or to obtain reimbursement for travel costs and time associated with the trial [24]. Additionally, concerns about eye safety after using this tool should not be overlooked. Various vision problems were perceived to be associated with the iris scan.

In studies conducted elsewhere, similar reasons for hesitancy—such as general safety concerns and anxiety about the physical effects of biometric scanning—have been reported [14,25,26]. However, it is important to note that vision impairment was neither reported as an adverse event nor assessed as being associated with iris scanning during the trial. Additionally, concerns were raised regarding the collection and safeguarding of additional personal data following the iris scan. Notably, collecting such information was new to trial participants in the remote area of Boende (the trial site), where the most common biometric practice—used for payroll verification, passport applications, and voter registration—remains fingerprinting [27]. This familiarity with fingerprinting may have contributed to greater comfort with it compared with iris scanning and may have also raised security concerns about iris scanning, despite its general acceptance and perceived accuracy.

It is likely that the vision disorders perceived to be associated with iris scanning in the EBL2007 vaccine trial had other causes. Some participants may have had preexisting eye conditions. For instance, while the mean age of trial participants was 45 years [27], it is well known that the incidence of vision impairment increases from middle age onward [28,29]. Additionally, at the time of the EBL2007 vaccine trial, no ophthalmological care was available in Boende. As a result, participants' vision or ophthalmological complaints may not have been addressed at the time of enrollment. Promotional and preventive activities aimed at improving eye health may also be necessary, as studies have shown that visual impairment is prevalent in populations living in remote, resource-constrained areas due to limited access to quality health care services [30]. However, implementing such activities may introduce additional costs for researchers. Given the increasing spread of digital legislation and democracy in Africa [31], these concerns may gradually diminish with the wider use of iris scanning technology. This suggests that even if research participants develop vision problems—likely due to aging or other factors—they may not attribute them to iris scanning as digital technology becomes more commonplace.

Some interviewees noted that the iris scanner operator appeared more like a photographer than someone capable of properly explaining the tool and its safety information. This suggests that the purpose and function of the iris scan tool were not sufficiently explained at inclusion or during follow-up visits in the trial. As a lesson learned, it is crucial to provide more detailed training to the iris scan operator to ensure they can answer specific questions from research participants.

It is also worth emphasizing that the timing of this qualitative study—conducted 2 years after the trial began—may explain why some interviewees and FGD participants gradually forgot the information about the iris scan tool provided during the initial consent process. Previous studies have highlighted that trial volunteers often forget or retain less of the information given at the time of enrollment [32,33]. This suggests that in long-term studies, the contents of the informed consent form should be periodically re-explained to participants.

XSL•FO
RenderX

This study demonstrated that iris scanning in vaccine trials conducted in resource-poor settings has valuable potential and is generally accepted for participant identification. The acceptability of iris scanning aligns with the key properties required of a high-performance biometric tool, such as universality, uniqueness, permanence, collectability, and resistance to circumvention [31,34]. However, during the EBL2007 vaccine trial, the iris scanning process became cumbersome when identification was not possible due to a participant's failure to follow the operator's instructions. To facilitate rapid identification through iris scanning, it is essential that individuals remain attentive and adhere strictly to the operator's guidance. This challenge may be particularly pronounced in vaccine trials involving younger infants (under 1 year old). Although infants are among the populations most in need of vaccines, they are unable to follow detailed instructions—such as looking into a camera—required for iris recognition [15]. Alternative approaches, such as scanning the iris of the accompanying adult (proxy identification number) or using ear-based or palm-based automatic recognition, may be more suitable for infants and other dependent populations [35,36].

Limitations

This study has some limitations. First, our findings are based on long-term experiences—spanning multiple iris scanning moments over a 2-year study period—of a population of health care providers and frontline workers who likely have a higher level of understanding of health-related phenomena. As a result, these findings may not be generalizable to the broader population. Second, some of the researchers who conducted FGDs and interviews, although not directly involved in the medical aspects of the vaccine trial, may have been perceived by interviewees as representing the trial team. This perception could have introduced desirability bias in the IDIs and FGDs. Finally, some participants working at the General Referral Boende Hospital—also the study site of the trial—may have hesitated to express negative concerns about the iris scan due to the location of the interviews (ie, at the hospital) or because some interviews were conducted by a trial investigator. Nevertheless, the findings of this qualitative research reflect participants' and staff's retrospective experiences with iris scanning over time and complement previous qualitative research on its acceptability. To our knowledge, this study, combined with the earlier research on the acceptability of biometric identity verification tools, provides the only comprehensive analysis of both the initial acceptability and the actual experiences of iris scanning within the same trial population [4]. The insights gained can inform the broader implementation of iris scanning in vaccine trials or other long-term longitudinal research.

Conclusions

The findings of this qualitative research highlight the sustained acceptability and perceived high accuracy of the iris scan tool for uniquely identifying adult participants in a vaccine trial over time. However, when the functionality of the iris scan is not well understood or remembered by users, certain concerns may arise, including perceived risks to long-term vision, the use of retained data, and the tool's ability to rapidly verify information regardless of age, education level, or health condition. To support broader implementation in vaccine trials or other research, further efforts should be made to provide clear information to users and dispel misconceptions about the fears and perceived risks associated with the iris scan tool.

Acknowledgments

This work was supported by the EBOVAC3. This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement number 800176 (IMI-EU). This Joint Undertaking receives support from the European Union's Horizon 2020 Research and Innovation Programme, the European Federation of Pharmaceutical Industries and Associations (EFPIA), and the Coalition for Epidemic Preparedness Innovations (CEPI). All vaccines were provided by Janssen Vaccines & Prevention B.V. The funders had no role in the data collection, analysis, and interpretation; writing of the paper; or the decision to submit for publication. We extend our sincere gratitude to the health care providers and frontline workers of the Boende Health District for their valuable participation in this study. The authors deeply appreciate the hard work and dedication of the local trial staff. Additionally, we acknowledge with gratitude the supportive role of all partners within the EBOVAC3 Consortium.

Data Availability

The data sets generated during this study are available and included as multimedia appendices (Multimedia Appendices 1-5).

Disclosure of Artificial Intelligence Use

This manuscript was prepared without the assistance of generative artificial intelligence tools or algorithms for data analysis, research design, or writing substantial content.

Authors' Contributions

TZM was responsible for the conceptualization of the study protocol, data curation, formal analysis, investigation, methodology, software, and writing of the original draft. AP contributed to the conceptualization of the study protocol, data curation, formal analysis, investigation, methodology, software, supervision, validation, visualization, and writing—review and editing. FBB participated in the investigation and writing—review and editing. GL, YL, EEL, SM, and PM were involved in project administration

XSL•FO
RenderX

and writing—review and editing. MS, BIO, JM, and VM contributed to writing—review and editing. PVD and JPVG provided funding acquisition, supervision, validation, and writing—review and editing. HMM contributed to supervision, validation, and writing—review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The profile of health care providers included in our previous study [4], which evaluated the acceptability of iris scanning in the EBL2007 trial before and during the recruitment process.
[DOCX File , 15 KB - formative_v9i1e54921_app1.docx]

Multimedia Appendix 2
The semistructured questionnaire used to interview participants.
[DOCX File , 18 KB - formative_v9i1e54921_app2.docx]

Multimedia Appendix 3
Reporting follows the COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines, which best match our study design to ensure completeness and detail.
[DOCX File , 24 KB - formative_v9i1e54921_app3.docx]

Multimedia Appendix 4
The process and the iris scanning tool used as an identification method in the EBL2007 trial, with the user experience—including clinical trial participants, staff, and operators—reported in our qualitative study.
[PPTX File , 641 KB - formative_v9i1e54921_app4.pptx]

Multimedia Appendix 5
It refers to the collection of transcripts (French) from all conversations (interviews and focus group discussions).
[DOCX File , 571 KB - formative_v9i1e54921_app5.docx]

References

1. Etter LP, Ragan EJ, Campion R, Martinez D, Gill CJ. Ear biometrics for patient identification in global health: a field study to test the effectiveness of an image stabilization device in improving identification accuracy. BMC Med Inform Decis Mak 2019 Jun 18;19(1):114 [FREE Full text] [doi: 10.1186/s12911-019-0833-9] [Medline: 31215427]
2. Gatwa T. "Rwanda and Burundi". In: Christianity in Sub-Saharan Africa. Edinburgh, Scotland: Edinburgh University Press; May 15, 2017:119-131.
3. ID4D data: global identification challenge by the numbers. World Bank. URL: https://id4d.worldbank.org/global-dataset/visualization [accessed 2022-05-10]
4. Zola Matuvanga T, Johnson G, Larivière Y, Esanga Longomo E, Matangila J, Maketa V, et al. Use of iris scanning for biometric recognition of healthy adults participating in an Ebola vaccine trial in the Democratic Republic of the Congo: mixed methods study. J Med Internet Res 2021 Aug 09;23(8):e28573 [FREE Full text] [doi: 10.2196/28573] [Medline: 34378545]
5. Singh T, Zaka-Ur-Rab S. Effect of pupil dilation on biometric iris recognition systems for personal authentication. Indian J Ophthalmol 2022 Dec 30;71(1):57-61 [FREE Full text] [doi: 10.4103/ijo.ijo_1417_22]
6. Dhir L, Habib NE, Monro DM, Rakshit S. Effect of cataract surgery and pupil dilation on iris pattern recognition for personal authentication. Eye (Lond) 2010 Jun 13;24(6):1006-1010. [doi: 10.1038/eye.2009.275] [Medline: 19911017]
7. SonLa Study Group. Using a fingerprint recognition system in a vaccine trial to avoid misclassification. Bull World Health Organ 2007 Jan 01;85(1):64-67 [FREE Full text] [doi: 10.2471/blt.06.031070] [Medline: 17242760]
8. Wendehorst C, Yannic D. Biometric recognition and behavioural detection: study commissioned by European parliament. European Parliament 2021 Aug 01:1-96 [FREE Full text]
9. Roberts CH, Stott C, Shawe-Taylor M, Chaudhry Z, Lal S, Marks M. Biometric linkage of longitudinally collected electronic case report forms and confirmation of subject identity: an open framework for ODK and related tools. Front Digit Health 2023 Aug 4;5:1072331 [FREE Full text] [doi: 10.3389/fdgth.2023.1072331] [Medline: 37600479]
10. Daugman JG. High confidence visual recognition of persons by a test of statistical independence. IEEE Trans Pattern Anal Machine Intell 2002;15(11):1148-1161 ISSN: 0162-8828. [doi: 10.1109/34.244676]
11. Jain AK, Arora SS, Cao K, Best-Rowden L, Bhatnagar A. Fingerprint recognition of young children. IEEE Trans Inform Forensic Secur 2017 Jul;12(7):1501-1514. [doi: 10.1109/tifs.2016.2639346]

XSL•FO
RenderX

12. Moolla Y, De Kock A, Mabuza-Hocquet G, Ntshangase CS, Nelufule N, Khanyile P. Biometric recognition of infants using fingerprint, iris, and ear biometrics. IEEE Access 2021;9:38269-38286. [doi: 10.1109/access.2021.3062282]
13. Hollingsworth K, Bowyer KW, Lagree S, Fenker SP, Flynn PJ. Genetically identical irises have texture similarity that is not detected by iris biometrics. Computer Vision and Image Understanding 2011 Nov;115(11):1493-1502. [doi: 10.1016/j.cviu.2011.06.010]
14. Okello E, Ayieko P, Kwena Z, Nanyonjo G, Bahemuka U, Price M, et al. Acceptability and applicability of biometric iris scanning for the identification and follow up of highly mobile research participants living in fishing communities along the shores of Lake Victoria in Kenya, Tanzania, and Uganda. Int J Med Inform 2023 Apr;172:105018 [FREE Full text] [doi: 10.1016/j.ijmedinf.2023.105018] [Medline: 36774907]
15. Masyn S, Vuchelen A, Santermans E, Rasschaert F, Bangura A, Parys W, et al. Overcoming the challenges of iris scanning to identify minors (1-4 years) in the real-world setting. BMC Res Notes 2019 Jul 22;12(1):448 [FREE Full text] [doi: 10.1186/s13104-019-4485-8] [Medline: 31331369]
16. Larivière Y, Zola T, Stoppie E, Maketa V, Matangila J, Mitashi P, et al. Open-label, randomised, clinical trial to evaluate the immunogenicity and safety of a prophylactic vaccination of healthcare providers by administration of a heterologous vaccine regimen against Ebola in the Democratic Republic of the Congo: the study protocol. BMJ Open 2021 Sep 28;11(9):e046835 [FREE Full text] [doi: 10.1136/bmjopen-2020-046835] [Medline: 34588237]
17. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
18. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. International Journal of Qualitative Methods 2006 Mar 01;5(1):80-92. [doi: 10.1177/160940690600500107]
19. Azungah T. Qualitative research: deductive and inductive approaches to data analysis. QRJ 2018 Oct 31;18(4):383-400. [doi: 10.1108/qrj-d-18-00035]
20. Creswell JW, Poth CN. Qualitative Inquiry and Research Design: Choosing Among Five Approaches (4th Edition). Los Angeles, CA: Sage Publications; 2016.
21. Braun V, Clarke V. Toward good practice in thematic analysis: avoiding common problems and be(com)ing a researcher. Int J Transgend Health 2023 Oct 25;24(1):1-6 [FREE Full text] [doi: 10.1080/26895269.2022.2129597] [Medline: 36713144]
22. Reidenbach HD. Laser safety. In: Träger F, editor. Springer Handbook of Lasers and Optics. New York, NY: Springer; 2007:1251-1276.
23. Maganga GD, Kapetshi J, Berthet N, Kebela Ilunga B, Kabange F, Mbala Kingebeni P, et al. Ebola virus disease in the Democratic Republic of Congo. N Engl J Med 2014 Nov 27;371(22):2083-2091. [doi: 10.1056/nejmoa1411099]
24. Zola Matuvanga T, Larivière Y, Lemey G, De Bie J, Milolo S, Meta R, et al. Setting-up an Ebola vaccine trial in a remote area of the Democratic Republic of the Congo: challenges, mitigations, and lessons learned. Vaccine 2022 May 31;40(25):3470-3480 [FREE Full text] [doi: 10.1016/j.vaccine.2022.04.094] [Medline: 35550847]
25. Kalka N, Zuo J, Schmid N. Image quality assessment for iris biometric. In: Biometric Technology for Human Identification III. Bellingham, WA, USA: SPIE; Apr 01, 2006:124-136.
26. Daugman J, Downing C. Effect of severe image compression on iris recognition performance. IEEE Trans Inform Forensic Secur 2008 Mar;3(1):52-61. [doi: 10.1109/tifs.2007.916009]
27. Larivière Y, Garcia-Fogeda I, Zola Matuvanga T, Isekah Osang'ir O, Milolo S, Meta R, et al. Safety and immunogenicity of the heterologous 2-dose Ad26.ZEBOV, MVA-BN-Filo vaccine regimen in health care providers and frontliners of the Democratic Republic of the Congo. J Infect Dis 2024 Apr 12;229(4):1068-1076 [FREE Full text] [doi: 10.1093/infdis/jiad350] [Medline: 37673423]
28. Barber SJ. An examination of age-based stereotype threat about cognitive decline. Perspect Psychol Sci 2017 Jan 11;12(1):62-90 [FREE Full text] [doi: 10.1177/1745691616656345] [Medline: 28073332]
29. Dawes P, Dickinson C, Emsley R, Bishop PN, Cruickshanks KJ, Edmondson-Jones M, et al. Vision impairment and dual sensory problems in middle age. Ophthalmic Physiol Opt 2014 Jul 29;34(4):479-488 [FREE Full text] [doi: 10.1111/opo.12138] [Medline: 24888710]
30. Williamson S, Seewoodhary R, Dampies L. Effect of poverty on eye health and implications for nursing practice. Nurs Stand 2016 Aug 10;30(50):42-51. [doi: 10.7748/ns.2016.e10014] [Medline: 27507393]
31. Idowu HA. Biometric technologies and the prospect of sustainable democracy in Africa. JAE 2021 Jun 1;20(1):23-43. [doi: 10.20940/jae/2021/v20i1a2]
32. Bergler JH, Pennington AC, Metcalfe M, Freis ED. Informed consent: how much does the patient understand? Clin Pharmacol Ther 1980 Apr;27(4):435-440. [doi: 10.1038/clpt.1980.60] [Medline: 6987027]
33. Howard JM, DeMets D. How informed is informed consent? The BHAT experience. Control Clin Trials 1981 Dec;2(4):287-303. [doi: 10.1016/0197-2456(81)90019-2] [Medline: 6120794]
34. Bolle RM, Connell JH, Pankanti S, Ratha NK. Guide to Biometrics. New York, NY: Springer Science & Business Media; 2013.

XSL•FO
RenderX

35. Khatun F, Distler R, Rahman M, O'Donnell B, Gachuhi N, Alwani M, et al. Comparison of a palm-based biometric solution with a name-based identification system in rural Bangladesh. Glob Health Action 2022 Dec 31;15(1):2045769 [FREE Full text] [doi: 10.1080/16549716.2022.2045769] [Medline: 35343885]
36. Lemes R, Bellon O, Silva L. Biometric recognition of newborns: identification using palmprints. New York, NY: IEEE; 2011 Presented at: International Joint Conference on Biometrics (IJCB); October 11-13, 2011; Washington, DC p. 1-6. [doi: 10.1109/IJCB.2011.6117475]

Abbreviations

DRC: The Democratic Republic of the Congo
FGD: focus group discussion
IDIs: in-depth individual interview
IEC: International Electrotechnical Commission
LMIC: low- and middle-income country

Edited by A Mavragani; submitted 27.11.23; peer-reviewed by S Rajput, M Shabdiz, N Manshor, H Campos; comments to author 06.09.24; revised version received 02.10.24; accepted 05.12.24; published 06.03.25.

Please cite as:
Zola Matuvanga T, Paviotti A, Bikioli Bolombo F, Lemey G, Larivière Y, Salloum M, Isekah Osang'ir B, Esanga Longomo E, Milolo S, Matangila J, Maketa V, Mitashi P, Van Damme P, Muhindo-Mavoko H, Van geertruyden JP
Long-Term Experiences of Health Care Providers Using Iris Scanning as an Identification Tool in a Vaccine Trial in the Democratic Republic of the Congo: Qualitative Study
JMIR Form Res 2025;9:e54921
URL: https://formative.jmir.org/2025/1/e54921
doi:10.2196/54921
PMID:40053756

©Trésor Zola Matuvanga, Antea Paviotti, Freddy Bikioli Bolombo, Gwen Lemey, Ynke Larivière, Maha Salloum, Bernard Isekah Osang'ir, Emmanuel Esanga Longomo, Solange Milolo, Junior Matangila, Vivi Maketa, Patrick Mitashi, Pierre Van Damme, Hypolite Muhindo-Mavoko, Jean-Pierre Van geertruyden. Originally published in JMIR Formative Research (https://formative.jmir.org), 06.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Assessment of Fraud Deterrence and Detection Procedures Used in a Web-Based Survey Study With Adult Black Cisgender Women: Description of Lessons Learned and Recommendations

Amber I Sophus[1*], MPH, PhD; Jason W Mitchell[1*], MPH, PhD

Department of Health Promotion and Disease Prevention, Robert Stempel College of Public Health & Social Work, Florida International University, Miami, FL, United States

[*]all authors contributed equally

Corresponding Author:
Amber I Sophus, MPH, PhD
Department of Health Promotion and Disease Prevention
Robert Stempel College of Public Health & Social Work
Florida International University
11200 S.W. 8th Street
Miami, FL, 33199
United States
Email: asophus@fiu.edu

Abstract

Background: Online research studies enable engagement with more Black cisgender women in health-related research. However, fraudulent data collection responses in online studies raise important concerns about data integrity, particularly when incentives are involved.

Objective: The purpose of this study was to assess the strengths and limitations of fraud deterrence and detection procedures implemented in an incentivized, cross-sectional, online study about HIV prevention and sexual health with Black cisgender women living in Texas.

Methods: Data for this study came from a cross-sectional web-based survey that examined factors associated with potential pre-exposure prophylaxis use among a convenience sample of adult Black cisgender women from 3 metropolitan areas in Texas. Each eligibility screener and associated survey entry was evaluated using 4 fraud deterrence features and 7 fraud detection benchmarks with corresponding decision rules.

Results: A total of 5862 respondents provided consent and initiated the eligibility screener, of whom 2150 (36.68%) were ineligible for not meeting the inclusion criteria, and 131 (2.23%) completed less than 80% of the survey and were removed from further consideration. Other entries were removed for not passing level 1 fraud deterrent safeguards: duplicate entries with the same IP address (388/5862, 6.62%), same telephone number (69/5862, 1.18%), same email address (114/5862, 1.94%), and same telephone number and email address (17/5862, 0.29%). Of the remaining 2993 entries, 1652 entries were removed for not passing the first 2 items of the level 2 fraud detection benchmarks: screeners and surveys with latitude and longitude coordinates outside of the United States (347/2993, 11.59%) and survey completion time of less than 10 minutes (1305/2993, 43.6%). Of the remaining 1341 entries, 130 (9.69%) passed all 5 of the remaining level 2 data validation benchmarks, and 763 (56.89%) entries were removed due to passing less than 3. An additional 33.4% (423/1341) entries were removed after passing 4 of the 5 remaining validation benchmarks, being contacted to verify survey information, and not providing legitimate contact information or being unable to confirm personal information. The final enrolled sample in this online study consisted of 155 respondents who provided consent, were deemed eligible, and passed fraud deterrence features and fraud detection benchmarks. In this paper, we discuss the lessons learned and provide recommendations for leveraging available features in survey software programs to help deter bots and enhance fraud detection procedures beyond relying on survey software options.

Conclusions: Effectively identifying fraudulent responses in online surveys is an ongoing challenge. The data validation approach used in this study establishes a robust protocol for identifying genuine participants, thereby contributing to the removal of false data from study findings. By sharing experiences and implementing thorough fraud deterrence and detection protocols, researchers can maintain data validity and contribute to best practices in web-based research.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e59955) doi:10.2196/59955

KEYWORDS

Black women; HIV; fraud deterrence; fraud detection; web-based research; online research; data integrity; data collection; survey

Introduction

Background

In the past 2 decades, there has been a rise in online health-related research studies. Online research provides a practical approach to recruitment and data collection, including electronic dexterity, participant anonymity, speed, reduced error, easy and remote participation [1], cost efficiency, and the potential to recruit larger study samples [2,3]. While online research has expanded recruitment opportunities, Black/African Americans remain underrepresented in health-related research studies [4-9], particularly Black women [7-9] who are considered a "hard-to-reach" population [9,10].

Online research studies offer one way to engage more Black cisgender women in health-related research. However, conducting research online, particularly when incentives are involved, warrants careful attention. For instance, respondents may misrepresent themselves to improve their likelihood of meeting the study eligibility criteria to obtain the incentive [11,12]. The lack of face-to-face interaction (eg, video) restricts researchers' ability to verify whether the data come from real individuals [13,14]. Other challenges associated with conducting online research include the presence of bots (ie, computer program software designed to automatically fill out survey questions with preprogrammed responses), the ability to bypass IP address restrictions, and respondent misrepresentation (ie, ineligible participants providing inaccurate information to gain entry into the study) [2,13,14]. Although online recruitment (ie, through social media) has been successful in attracting underrepresented populations in health-related research studies involving incentives [15-18], the use of a standard web link for online recruitment purposes may exacerbate the potential for fraudulent responses and threats to data integrity [19-22]. In response to these challenges, researchers have provided recommendations for how best to screen for fraudulent survey entries and handle potentially invalid responses [11,12,22-26].

Teitcher et al [12] recommends checking for inconsistent survey responses, using a CAPTCHA, collecting paradata to examine how individuals are responding to survey questions, tracking personal information (eg, email, home address, or telephone number), checking for encrypted IP addresses or multiple survey entries from the same IP address, enabling cookies to prevent multiple survey completion attempts, and including an interview to determine whether an individual has already participated or is being dishonest on responses. Other researchers have published similar recommendations [2,23,27,28]. In addition to these recommendations, researchers are encouraged to develop a system of decision rules to detect and handle invalid and fraudulent research data [11-13,25]. Pozzar et al [21] investigated threats to their sample validity and data integrity after online social media recruitment of health research participants to complete an online survey. Although the authors used a data collection platform with fraud prevention features, collected verifiable information, and included open-ended items to identify those who provided false information regarding their eligibility criteria, many respondents still bypassed the validity and data integrity measures [21]. Dewitt et al [29] applied similar participant validity procedures in a small internet health survey among a sample of gay and bisexual survivors of prostate cancer. However, the authors discovered that some invalid entries bypassed their validation protocol, although they passed checks for nonduplicate IP addresses, valid zip codes, and reCAPTCHA verification [29]. These findings underscore the importance of ongoing vigilance and periodic reassessment of validation strategies throughout the recruitment phase of a study to enhance the overall effectiveness of fraud prevention strategies. Furthermore, the tactics used to deceive or manipulate online study eligibility and enrollment systems are continually evolving and, as such, call for new insights and lessons learned to help improve rigor and data integrity.

Purpose

This analysis builds on extant literature by assessing the strengths and limitations of fraud deterrence and detection procedures that were implemented in an incentivized, cross-sectional, online study about HIV prevention and sexual health with a convenience sample of Black cisgender women living in Texas, United States. We share the lessons learned from implementing this study and discuss different strategies that are available to help improve the likelihood of collecting valid data from legitimate research participants.

Methods

This study adhered to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) reporting guidelines [30].

Recruitment

Between December 2020 and January 2022, we recruited Black cisgender women from Houston, San Antonio, and Dallas in Texas to complete a 1-time online study survey about HIV prevention and women's sexual health, with a primary focus on pre-exposure prophylaxis [31]. Participants were recruited using online advertisements and print flyers that contained information about the study and a web link to access the online consent form and eligibility screener. Online advertisements were placed on Facebook and Instagram (clickable web link embedded), whereas printed flyers (with a QR code and web link) were placed in locations (eg, coffee shops, restaurants, grocery stores, and gyms) in Houston, Dallas, and San Antonio frequented by Black women. Snowball sampling was also used to recruit participants. All recruitment materials contained information about the study and a method (eg, QR code or direct web link) to access the online consent form and eligibility screener [32].

XSL•FO
RenderX

Procedures

Irrespective of the recruitment method, all interested individuals who clicked the provided web link were taken to the study landing page housed in Qualtrics (Qualtrics International Inc), a Health Insurance Portability and Accountability Act (HIPAA) compliant survey program. The landing page provided information about the study, the electronic consent form, and the online eligibility screener. The electronic consent form provided detailed information about the survey study, such as its purpose and participation process (including time commitment), incentives, privacy and confidentiality, risks and benefits, and contact details for the primary investigator and institutional review board. Respondents who consented to participate were immediately routed to the eligibility screener. Once consented and eligible, individuals were prompted to provide their email address for incentive distribution before being automatically granted access to complete the 30-minute online study survey. Individuals were also instructed to provide their telephone number if they consented to participate in a possible 1-time, online interview (if purposively selected). In addition, individuals could provide their contact information to be contacted for future research opportunities.

Consented and eligible individuals who completed at least 80% of the study survey were enrolled in the parent study once their data were evaluated and had passed the fraud deterrence and detection protocols (described later in the Methods section). Individuals who were ineligible were thanked for their interest in the study and provided with relevant resources related to HIV prevention.

Eligibility Criteria

For study inclusion, participants had to meet all of the following eligibility criteria through self-report: (1) be at least 18 years of age; (2) self-identify as a cisgender woman; (3) self-identify as Black (ie, African American or Caribbean American [eg, Haitian American]); (4) have an HIV-negative or unknown serostatus; (5) live in or near (within 25 miles [approximately 40 km]) Houston, San Antonio, or Dallas in Texas; (6) have engaged in at least 1 HIV risk behavior within the past 6 months (ie, unprotected vaginal or anal sex with a male partner, injection drug use, or sex exchange) or have been diagnosed with a sexually transmitted infection (STI); and (7) be fluent in English. Individuals who did not meet all 7 eligibility criteria were excluded.

Application of the First Level of Fraud Deterrent Protocols: Lesson 1

To help deter the collection of invalid data [11,23,27,33-37], 4 fraud prevention safeguards—built-in features available in Qualtrics—were selected for use in the online eligibility screener and study survey before recruitment began. The 4 fraud prevention safeguards are CAPTCHA verification, prevent indexing, prevent ballot box stuffing (ie, duplicate entries), and bot detection. Descriptions of the 4 fraud prevention safeguards are provided in Textbox 1.

Textbox 1. Descriptions of the level 1 fraud deterrent protocol, which included 4 fraud prevention safeguards used in the Qualtrics eligibility screener for all respondents.

CAPTCHA

CAPTCHA requires a respondent to successfully pass a task or challenge (eg, select all squares containing fire hydrants in the image) to gain access to the next web page (eg, survey)

Prevent indexing

A tool that prevents bots or software from finding the survey within web search engines

Prevent ballot box stuffing (currently called "Prevent Multiple Submissions")

A tool that places a cookie on the individual's browser (cookies are small data files generated by a web server and sent to a web browser; they track user activity, help websites recognize returning users, and improve the browsing experience); if the same respondent returns using the same browser on the same device, without having cleared their cookies, they are flagged as a duplicate

Bot detection

A tool that assesses the likelihood of a response being from a bot or a human by assigning a probability score based on interactions with invisible Google reCAPTCHA technology embedded in the survey

Soon after recruitment began, there was a huge influx of data entries—1498 within the first 3 days. This unusual activity prompted us to pause the study (ie, recruitment as well as screener and survey) to evaluate whether the screener and study survey entries were valid. During this review, the safeguard options presented in Textbox 1 were thoroughly examined to determine whether there were limitations to using these 4 features.

We learned that the prevent ballot box stuffing option in Qualtrics does not fully prevent duplicate entries. Respondents could still access the survey by switching browsers or devices, even when this option was enabled. In addition, Qualtrics does not prevent duplicate entries based on additional criteria such as name, telephone number, email address, or IP address. Of the 1498 entries, approximately half (n=769, 51.3%) were linked to an entry that shared an IP address with at least 1 other survey entry. Several other patterns were noted while evaluating the data entries: (1) some (53/1498, 3.54%) came from latitude and longitude coordinates outside of the United States and its territories; (2) a little more than one-fifth (350/1498, 23.36%) were completed in record time (eg, 9 min vs an estimated completion time of 30 min); (3) some (355/1498, 23.69%) contained names or email addresses with unusual handles, unconventional characters, excessive numbers, or uncommon symbols (eg, a987quaiigi@yahoo.com and Â2xylzggf@gmail.com); and (4) a few (18/1498, 1.2%) included

XSL•FO
RenderX

a US telephone number with an incorrect digit count (>10 digits or <10 digits). Upon further investigation, we noticed that responses to certain survey items were either nonsensical, highly improbable, or were incongruent with a response to a related item. For example, in an open-text item asking, "How did you hear about PrEP?" some respondents provided nonsensical answers, such as "Jsjsjd," or an illogical response, such as "Yes" (27/1498, 1.8%). A highly improbable response pattern was observed when some entries (17/1498, 1.13%) selected all options for HIV behavioral risk factors (eg, "I had unprotected vaginal sex with a male partner [vaginal sex without using a condom] in the past 6 months," "I had unprotected anal sex with a male partner [anal sex without using a condom] in the past 6 months," "I had sex in exchange for something of value [such as food, shelter, money, or drugs] in the past 6 months," "I have taken drugs by injection with a needle [such as heroin, cocaine, amphetamines, or steroids; not including anything taken under a doctor's order] in the past 6 months," and "I have been diagnosed with an STI in the past 6 months [STIs include chlamydia, gonorrhea, syphilis, human papilloma virus (HPV and warts) and herpes simplex virus]"). Selecting all options was considered highly unlikely.

After evaluating the influx of responses and noting the aforementioned patterns, we decided to call the first 20 people who met the eligibility criteria, consented to participate, and completed the study survey and whose responses showed no signs of fraudulent activity. The goal was to verify whether the provided telephone numbers were valid and belonged to the respondents. The majority of telephone numbers called (16/20, 80%) were linked to a person, company, or business not associated with the individual listed in the survey entry or were disconnected or no longer in service. Only 4 (20%) of the 20 telephone numbers were valid and belonged to the individual who completed the survey. In summary, our evaluation of data entries in relation to the limitations of the survey-based fraud prevention safeguard options highlighted the need to implement additional fraud deterrence and detection procedures to help ensure that valid data were being collected from legitimate research participants.

Application of the Second Level of Fraud Detection: Lesson 2

We made a number of changes before resuming the study. First, we retained and implemented the original 4 safeguard options provided by Qualtrics. In addition to the existing CAPTCHA, which was placed before the first question of the eligibility screener, we added a CAPTCHA before the first question of the study survey to further deter bots. We also decided to manually evaluate all completed eligibility screeners (with the associated study survey) for duplicate records (ie, >1 entry) [38] by assessing whether the same IP address, telephone number, or email address appeared across multiple entries. If >3 entries were found to have the same IP address, we retained the first 3 and removed the rest. If multiple entries were found to have the same telephone number or email address, we kept only the first entry and removed all others. On the basis of prior research and our initial findings, we created and implemented 7 additional data validity benchmarks for fraud detection. Each benchmark included a predetermined decision rule that was created a priori. These benchmarks, along with their corresponding decision rules, are described in Table 1 and were implemented alongside the 4 original Qualtrics safeguard options.

After duplicates were removed, each screener and study survey entry was evaluated by applying the 7 fraud detection benchmarks presented in Table 1. First, any entry with latitude and longitude coordinates outside of the United States was labeled as fraudulent and excluded (ie, ineligible and not enrolled). Any entry that had a completion time of ≤10 minutes was also labeled as fraudulent and not enrolled. All remaining entries with latitude and longitude coordinates within the United States and a completion time of >10 minutes were further evaluated using the 5 remaining benchmarks (items 3-7 in Table 1). These 5 benchmarks evaluated whether the self-reported telephone number contained 10 digits in the United States format (ie, xxx-xxx-xxxx); the self-reported email address was correctly formatted and deemed valid using a third-party email validation tool or was associated with a Facebook profile; the responses to certain survey questions (eg, HIV behavioral risk factors or open-ended survey questions) were nonsensical or highly improbable (ie, selecting having engaged in all 5 options for HIV behavioral risk factors in the 6 months before study participation: unprotected vaginal sex with a male partner, unprotected anal sex with a male partner, injection drug use, sex exchange, and diagnosed with an STI); the self-reported responses to the items about zip code and residential city matched; and name, email address, or open-ended text responses contained nonstandard characters or symbols not commonly used in the United States (eg, Â). To be considered a highly probable respondent with valid data, individuals had to pass all 5 fraud detection benchmarks (items 3-7 in Table 1). Respondents who passed ≤3 were deemed fraudulent (ie, ineligible and not enrolled). Those who passed at least 4 benchmarks were marked for further review. A team member then contacted these individuals via telephone or email to ask 2 questions about the personal information provided in the screener or study survey (eg, current age, birth month, zip code, city of residence, or relationship status). Individuals who could not be reached or provided incorrect answers were categorized as "fraudulent" and not enrolled. The goal of these procedures was to identify additional fraudulent entries to improve the integrity of the data collected.

XSL•FO
RenderX

Table 1. Description of the level 2 fraud detection protocol, which included 7 data validation benchmarks and corresponding decision rules used to identify additional fraudulent survey respondent entries[a].

Data used for benchmark	Benchmark rule	Data source	Decision rule
1. Latitude and longitude coordinates of IP address	Location must be within the United States	Eligibility screener	If individual does not pass, labeled as fraudulent; no longer reviewed
2. Start and finish time stamps	Completion time must exceed 10 minutes	Eligibility screener and study survey	If individual does not pass, labeled as fraudulent; no longer reviewed
3. Telephone number	Digits must be provided in the correct United States format (xxx-xxx-xxxx)	Eligibility screener	If not in United States format, labeled for further review
4. Email address	Email address must be valid when verified online through a third party (eg, s223456h@yahoo.com or have altering letters and numbers – 12s2n4n5P3@gmail.com [28,29])	Eligibility screener	If not in correct format or fake email address, labeled for further review
5. Nonsensical or highly improbable responses to survey questions	Selected all 5 options for HIV behavioral risk factors in the screener or provided irregular responses to open-ended survey questions	Eligibility screener and study survey	If all 5 options for HIV behavioral risk factors are selected, labeled for further review; if nonsensical or highly improbable responses have been provided to open-ended questions, labeled for further review
6. Matched responses: zip code and city	Provided a zip code that did not match their self-reported residential city	Eligibility screener	If self-reported zip code and city do not match, labeled for further review
7. Nonstandard characters or symbols	Name, email address, or open-ended text responses containing nonstandard characters or symbols not commonly used in the United States (eg, Â and Didnâ€™t)	Eligibility screener and study survey	If nonstandard characters or symbols identified, labeled for further review

[a]Respondents were first evaluated against items 1 and 2 of the 7 data validation benchmarks. Those who did not meet these criteria were labeled as "fraudulent." Respondents who passed were then evaluated against the remaining 5 benchmark items (items 3-7) and labeled based on the number of criteria met: "pass" (met all 5 criteria), "further review" (met 4 criteria), and "fraudulent" (met ≤3 criteria).

Ethical Considerations

The procedures associated with the primary study, including ethics approval and oversight, were approved by the institutional review board at the University of Hawai i at Mānoa (IRB 2020-00030). A certificate of confidentiality for human participant research was obtained from the National Institutes of Health to help keep participant data private. Informed electronic consent was obtained from all participants included in the study; each respondent was required to individually provide consent electronically before taking the eligibility screener. Upon study completion, all documents containing identifying information were deidentified and coded with a unique study number. As approved by the ethics committee, enrolled participants (ie, individuals who provided consent, met the eligibility criteria, completed at least 80% of the study survey, and passed all fraud deterrence and detection protocols) were emailed a US $25 electronic gift card for taking part in the study.

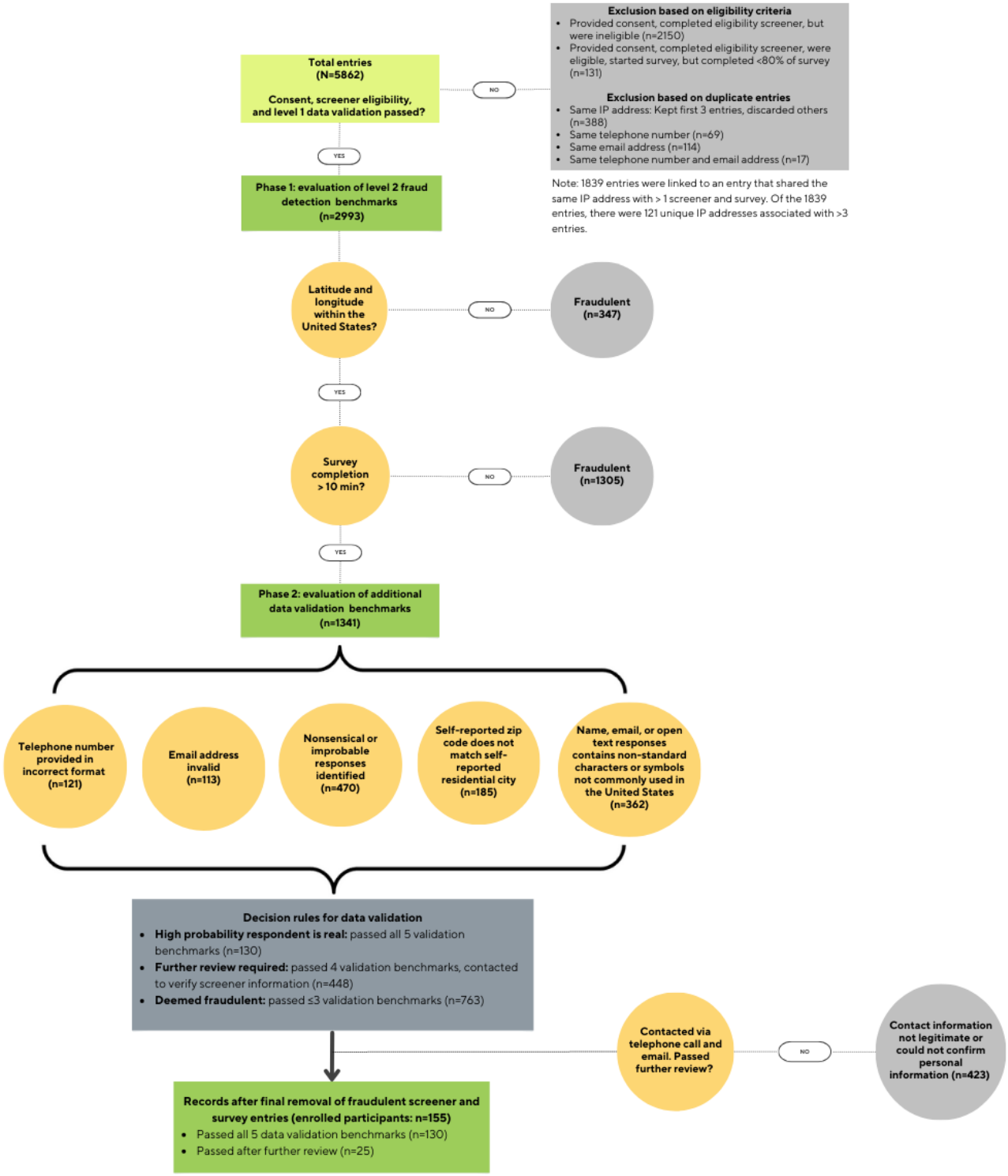
Results

Consent, Eligibility Criteria, and Level 1 Fraud Deterrent Protocols

Figure 1 presents the number of entries removed by implementing the 4 fraud deterrent safeguards provided by Qualtrics and the 7 fraud detection benchmarks. There were 5862 entries representing those who clicked the study link and provided consent. First, entries were removed if they did not pass the study eligibility criteria (ie, ineligible; 2150/5862, 36.68%). Entries were also removed if the respondents provided consent and were eligible but completed <80% of the study survey (131/5862, 2.23%). Second, duplicate entries (ie, entries with the same IP address and same telephone number or email address; 588/5862, 10.03%) were removed. With respect to duplicate IP address, there were 1839 entries that had the same IP address as at least 1 other screener and survey. Among these 1839 entries, 121 unique IP addresses were linked to an entry that shared the same IP address with ≥4 screeners and surveys. Following the data validation protocol for entries with the same IP address (ie, retaining the first 3 entries with the same IP address and discarding additional ones), 388 (21.1%) of the 1839 entries were removed. In sum, there were 2993 respondents who provided consent, met the eligibility criteria, and passed level 1 fraud deterrent safeguards.

XSL•FO
RenderX

Figure 1. Final participant enrollment after implementing the 4 fraud deterrent safeguards (provided by Qualtrics) and the 7 fraud detection benchmarks.



Level 2 Fraud Detection

Seven additional fraud detection benchmarks were implemented, resulting in the removal of additional entries. Of the 2993 entries that remained, 347 (11.59%) with latitude and longitude coordinates outside of the United States were removed, while 1305 (43.6%) were removed for having a survey completion time of <10 minutes. Among the remaining 1341 entries, implementing the 5 remaining fraud detection benchmarks led to the exclusion of 56.9% (763/1341) of the entries that met <4 of the benchmarks. Only 130 (9.7%) of the 1341 respondents passed all 5 data validation benchmarks and were enrolled in the study. A total of 448 respondents were contacted via telephone or email for further legitimacy review, of whom 25 (5.6%) were successfully reached and verified as legitimate.

Enrolled Participants

After removing entries based on the fraud deterrence and detection protocols, 155 Black cisgender women met all study criteria, including providing consent, meeting eligibility requirements, completing at least 80% of the study survey, and

XSL•FO
RenderX

passing all fraud deterrence and detection protocols, and were enrolled in the study.

Discussion

Principal Findings

This study, which focused on Black cisgender women, emphasizes the importance of implementing effective fraud deterrence and detection procedures in incentivized online research. Our findings, stemming from the lessons learned, highlight that solely relying on fraud deterrence features offered by online survey programs (eg, Qualtrics) is insufficient to ensure the collection of valid data from legitimate research participants. We demonstrate and argue that when conducting incentivized online research, additional fraud deterrence and detection protocols are needed to uphold data integrity and research rigor. These protocols should be implemented before the start of recruitment and enrollment, and should be adjusted as needed throughout the process. Of the 5862 entries received for the study, only 155 (2.64%) provided consent, were eligible, and met fraud detection and deterrent criteria. A large proportion of the entries (2838/5862, 48.41%) were deemed fraudulent. Through reflection on the protocols used in this online study, we offer suggestions and ideas for researchers to consider using to help deter and detect fraudulent data entries, with the objective of helping to increase the likelihood of collecting valid data from genuine participants.

Leveraging Available Features in Survey Software Programs to Deter Bots

Effectively identifying and managing bots is important to ensure the accuracy, reliability, and integrity of collected data. We recommend using all available features provided by the survey software while also being aware of their limitations to deter fraud. In this online study, we used the CAPTCHA verification, prevent indexing, prevent ballot box stuffing (ie, duplicate entries), and bot detection features offered by Qualtrics. We recommend leveraging the CAPTCHA feature because it can be used more than once within an online survey. Although 1 prior study [24] found the CAPTCHA feature to be insufficient in dissuading fraudulent responses, the authors did not provide details about how often it was actually used in their survey. In this study, we strategically placed a CAPTCHA before the first question in the eligibility screener as well as before the first question in the study survey. Strategically placing a CAPTCHA before the initial question in both the eligibility screener and the study survey potentially aided in the identification and elimination of additional fraudulent entries that might have bypassed the initial CAPTCHA. This assertion is supported by the discovery of participants (n=131) who provided consent and met the eligibility criteria but did not start the survey (Figure 1). The strategic placement of CAPTCHAs provides an additional layer of security against automated bot submissions and may help reduce the likelihood of receiving invalid responses.

The “prevent ballot box stuffing” option uses cookies to prevent multiple survey entries but does not fully prevent duplicate entries based on IP address or self-reported information, such as name, email address, and telephone number. Although prior studies have used this fraud detection feature, its limitations were not acknowledged or discussed [28,37]. Our evaluation suggests that respondents, in an attempt to identify which responses would qualify them for the study, may have repeatedly taken the eligibility screener using different internet browsers, as evidenced by the high number of duplicate entries from the same IP addresses (n=1839). When we discovered this, we manually evaluated screener entries to label and assess which ones came from the same IP address. We then applied the same process to evaluate screener entries that contained the same email addresses and telephone numbers, given the limitations of the prevent ballot box stuffing option. The implementation of these evaluation methods was valuable in discerning whether entries originated from the same entity. Our experience and the relevant lessons learned underscore the importance of understanding the limitations of fraud deterrence features in online survey software.

It is important to acknowledge that the 4 fraud deterrence features (CAPTCHA verification, prevent indexing, prevent ballot box stuffing, and bot detection) were implemented in the Qualtrics-hosted study survey between December 2020 and January 2022. Since then, Qualtrics may have updated or enhanced these features. Furthermore, Qualtrics offers 2 other fraud deterrence features that were not used in this study: Security Scan Monitor and RelevantID. Security Scan Monitor prevents email scanning software from inadvertently starting a survey session when a survey link is included in the email. This feature applies to all links, regardless of whether they were distributed via Qualtrics or a third-party system (ie, any software or platform not directly associated with Qualtrics, such as marketing automation software). RelevantID analyzes respondent metadata to determine the likelihood of multiple survey attempts by the same respondent by examining browser, operating system, and location details. Future surveys that are hosted on Qualtrics ought to consider using Security Scan Monitor and RelevantID, depending on the needs of the study. We also recommend that researchers refrain from exclusively relying on the fraud deterrence features provided by Qualtrics (or other survey software). Our lessons learned indicate that using a combination of fraud deterrent and fraud detection procedures will help provide a more robust defense against the collection of invalid data in online survey studies.

Enhancing Fraud Detection

Every online survey study is vulnerable to fraudulent entries. As such, the best approach is to identify potential vulnerabilities and implement strategies to prevent their exploitation before starting recruitment and data collection [25]. Initially, researchers should create and implement a protocol that incorporates both fraud deterrent and fraud detection strategies. Drawing from our experiences with this online survey study, we found that a combined approach of fraud deterrent and fraud detection worked best to identify fraudulent entries. Specifically, we—as well as prior research—evaluated the time taken to complete the eligibility screener and study survey [21,24,28,29,36], contacted respondents via telephone or email to verify their identity [36,39], and used matched survey

XSL•FO
RenderX

questions to identify nonsensical or highly improbable responses [21,24,29].

Predetermining an anticipated time range for survey completion adds an additional option to help detect fraud. For instance, data collected about survey completion time can be used to decipher whether a "participant" was likely to have completed it. Establishing a priori criteria for anticipated eligibility screener and study completion time can aid in fraud detection. Before launching this study, we used the automated survey completion time estimate provided by Qualtrics, which was 53 minutes. However, while this estimate accounts for survey flow, it does not account for skip logic or display logic. To address this, we manually tested survey completion times across different response paths. Our completion times ranged between 18 and 30 minutes. On the basis of these findings, we determined that it would be highly unlikely for an individual to complete the eligibility screener and associated study survey in <10 minutes. Therefore, we flagged all entries that completed the screener and survey in <10 minutes and considered them to represent automated, nonhuman involvement or rushed responses. We recommend that researchers consider implementing a similar strategy by predetermining a time range for survey completion to help detect fraudulent or rushed survey entries.

Similar to prior research [24,28], we verified respondents' telephone numbers and email addresses. The verification of contact information is important because respondents can easily create multiple email addresses or use Google Voice numbers to complete the survey multiple times. Third-party services that validate email addresses and Facebook were used to verify email addresses, while telephone numbers were validated through direct telephone calls to confirm legitimacy and ownership. Respondents labeled for "further review" were contacted via telephone or email and asked 2 questions about the personal information provided in the screener or study survey (eg, current age, birth month, zip code, city of residence, and relationship status). Although resource intensive, verifying phone numbers and emails helped to differentiate between legitimate and fraudulent entries, offering confidence in the authenticity of participant details and their survey data.

A notable concern with manual checks and using third-party verification services is the associated cost and time commitment. These expenses can be significant, especially for research projects with limited budgets. The time-consuming nature of manual verification of telephone numbers becomes apparent when handling a high volume of entries within a short time frame. Calling respondents individually can lead to delays and logistical challenges, impacting the efficiency of the research process. To deter fraudulent behavior, Ballard et al [28] sent an SMS text message to suspicious respondents stating, "You recently completed a survey for a health study online. However, we detected that your survey entry was fraudulent. If you think this is a mistake, please contact us." If participants did not respond, the authors considered the survey entry invalid [28]. The authors noted that they did not receive any responses categorized as "fraudulent" [28]. In summary, researchers must weigh the benefits of enhanced participant validation against practical constraints, such as cost, time, and personnel involvement, when considering the adoption of email and telephone verification protocols.

We also incorporated a matched response approach using 2 survey questions to enhance our fraud detection protocol. Respondents' self-reported zip code had to match their self-reported residential city; otherwise, they were flagged for possible fraud. Matching responses to different survey questions can help identify potential fraudulent entries. We recommend that researchers consider using at least 1, if not 2, matched responses to preselected survey questions. For example, asking for a respondent's current age at the beginning of an eligibility screener and then requesting their birth month and year—either toward the end of the screener or in the study survey—may help reveal discrepancies [34,36,37]. Alternatively, a survey could ask for a respondent's age range at the beginning and their specific age toward the end.

Additional Strategies and Recommendations for Fraud Deterrence and Detection

There may be other options that work best for online survey studies that can be used to discourage fraudulent behavior. Pratt-Chapman et al [24] recommend including a check box for participants to acknowledge that responses from ineligible respondents or multiple entries from the same respondent will disqualify them from receiving financial incentives. The authors also suggest indicating the investigators' right to confirm eligibility by telephone (or other means) to aid in identifying bots and eliminating duplicate entries [24]. In addition, when requests for payment are received from respondents whose entries are identified as fraudulent, Dewitt et al [29] recommend informing these respondents that there was a concern about their survey entry and asking them to call a 1-800 study line and leave a callback number for verification. The authors applied this approach to all suspicious entries and found that none resulted in return calls [29]. Another option is to contact participants via telephone or email to set up a Zoom meeting to determine their legitimacy [39] before inviting them to complete the study survey. During this meeting, participants could be asked to upload or show evidence to prove their identity or to verify they had met some of the inclusion criteria; for example, a study recruiting active US military service members could require potential participants to show their military ID as proof of service. The methods described here offer additional strategies for deterring and detecting fraudulent entries in online research studies.

While manual checks can be a viable option for online research studies with limited resources, more financially robust studies can deploy advanced fraud deterrence and detection strategies. One such option is incorporating email and telephone verification features that require respondents to validate their contact information by receiving and verifying a code. Guest et al [36] required respondents to submit their mobile phone number during the eligibility screener to receive a 3-digit verification code via SMS text message. After receiving the 3-digit code, respondents were required to enter the code in the eligibility screener survey to validate their mobile phone number for the study [36]. Those who failed to input the code were unable to continue with the eligibility screener [36].

XSL•FO
RenderX

Another recommendation for online eligibility screening is the use of automated electronic algorithms that allow researchers to create parameters and decision rules for inclusion criteria. Automated electronic algorithms also enable online eligibility screeners to restrict entries based on IP address (ie, blocking multiple entries from the same IP address or requiring a US-based IP address), completion time, and allowable responses that align with the inclusion criteria (eg, current age must be at least 18 years) [37]. These algorithms enhance fraud deterrent strategies while streamlining the eligibility screening process. However, implementing such techniques requires adequate financial and computational resources (eg, a coder or a web developer) tailored to the study's specific needs. It should be noted that implementing these procedures results in a major reduction in the study sample size. This is often the case when researchers opt to prioritize data integrity to ensure that the findings are representative of real people.

Overall, ongoing efforts are needed to refine and optimize fraud deterrence and detection protocols to maintain research rigor and improve the collection of valid data from legitimate research participants. This study uniquely contributes to existing related literature because it is one of the first to evaluate, describe lessons learned, and offer insights into fraud deterrence and detection protocols used for an incentivized online survey study with adult Black cisgender women. Most online research studies that have evaluated the use of fraud deterrence and detection methods have focused on samples of sexual minority men and male couples [11,12,23,34,37]. Further research is needed to evaluate the application and effectiveness of fraud deterrence and detection methods in online studies with diverse populations. It remains unknown whether distinct demographic groups require unique fraud deterrence and detection procedures. We encourage researchers to evaluate and publish findings stemming from their use of fraud deterrence and detection methods to help advance the rigor of online research studies.

Limitations

The limitations of this study are important to consider in light of the lessons learned and the recommendations provided. Despite the changes we made to enhance the fraud deterrence and detection methods used in this online survey study, it is possible that some participants provided false information during the verification attempts. In addition, reliance on telephone calls and emails for participant validation may introduce biases or errors due to participant nonresponsiveness or communication challenges. Participants may forget to respond to telephone calls or emails. As a result, there may be inaccuracies in the data. Although we used a third-party service for email validation, it is possible that some of the verified email addresses belonged to individuals other than the persons or entities completing the eligibility screener. Moreover, the evolving nature of fraud detection algorithms in online survey platforms such as Qualtrics may limit the long-term generalizability of recommendations based on research conducted within specific time frames. In addition, this study's fraud detection and deterrent protocols were designed to recruit Black cisgender women, potentially limiting the generalizability of these protocols to other populations. Future research should aim to use these protocols and provide feedback regarding whether these protocols improved their data integrity by improving their ability to identify bots and imposter participants during recruitment and enrollment. The findings would help to refine and strengthen fraud detection and deterrent protocols used to recruit diverse populations for online studies. It is important to acknowledge that the use of fraud deterrence and detection methods does not guarantee the complete elimination of all fraudulent entries; however, their use and evaluation help to enhance the confidence that valid data are being collected from verified participants, thereby contributing to the rigor and integrity of online research studies.

Conclusions

Effectively identifying fraudulent responses in web-based surveys is an ongoing challenge. With the increasing shift toward web-based research and online recruitment, the threat of fraudulent participation poses a real challenge to data validity. Protocols for identifying fraudulent survey entries and verifying and validating potential study participants should be considered for all internet-based studies. Researchers conducting online studies with Black cisgender women must actively share their experiences in deterring and detecting fraud to help contribute to the rigor of best practices and maintaining the validity of data and associated findings. The lessons learned and recommendations offered from the experiences of conducting this online study, which recruited and enrolled a study sample of Black cisgender women, highlight two important take-home points: (1) develop a thorough fraud deterrent and fraud detection plan to implement before study launch; and (2) monitor and evaluate how well these methods are working while data are being collected, as well as once data collection has ended. We encourage researchers to leverage all resources they may have at their disposal, given the number of different fraud deterrence and detection options that exist. This study—which emphasizes the importance of the aforementioned take-home points—used a combination of fraud deterrence and detection methods to identify a large number of fraudulent entries that would have otherwise been included in the data.

Acknowledgments

This study was funded, in part, by the Surgeon General C. Everett Koop HIV/AIDS Research Grant from the Rural Center for AIDS/STD Prevention, Indiana University School of Public Health-Bloomington. This research was supported in part by the National Institute on Minority Health and Health Disparities of the National Institutes of Health (U54MD012393), Florida International University Research Center in Minority Institutions. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

XSL•FO
RenderX

Data Availability

The datasets generated and analyzed during this study are not currently publicly available because the primary data are still being analyzed. Once primary analyses are complete, the study data will be available from the corresponding author on reasonable request.

Authors' Contributions

AIS contributed to the study design, data collection procedures, analysis, interpretation of the results, and drafting and critically reviewing the paper. JWM contributed to the concept and study design, interpretation of the results, and drafting and critically reviewing the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
STROBE Checklist.
[PDF File (Adobe PDF File), 84 KB - formative_v9i1e59955_app1.pdf]

References

1. Rhodes SD, Bowie DA, Hergenrather KC. Collecting behavioural data using the world wide web: considerations for researchers. J Epidemiol Community Health 2003 Jan;57(1):68-73 [FREE Full text] [doi: 10.1136/jech.57.1.68] [Medline: 12490652]
2. Pequegnat W, Rosser BR, Bowen AM, Bull SS, DiClemente RJ, Bockting WO, et al. Conducting internet-based HIV/STD prevention survey research: considerations in design and evaluation. AIDS Behav 2007 Jul 20;11(4):505-521. [doi: 10.1007/s10461-006-9172-9] [Medline: 17053853]
3. Evans JR, Mathur A. The value of online surveys: a look back and a look ahead. Internet Res 2018 Aug 06;28(4):854-887. [doi: 10.1108/intr-03-2018-0089]
4. Zahm SH, Pottern LM, Lewis DR, Ward MH, White DW. Inclusion of women and minorities in occupational cancer epidemiologic research. J Occup Med 1994 Aug;36(8):842-847. [Medline: 7807263]
5. Svensson CK. Representation of American Blacks in clinical trials of new drugs. JAMA 1989 Jan 13;261(2):263-265. [doi: 10.1001/jama.1989.03420020117041]
6. Hall WD. Representation of Blacks, women, and the very elderly (aged > or = 80) in 28 major randomized clinical trials. Ethn Dis 1999;9(3):333-340. [Medline: 10600055]
7. Mak WW, Law RW, Alvidrez J, Pérez-Stable EJ. Gender and ethnic diversity in NIMH-funded clinical trials: review of a decade of published research. Adm Policy Ment Health 2007 Nov 10;34(6):497-503. [doi: 10.1007/s10488-007-0133-z] [Medline: 17690976]
8. Isler MR, Brown AL, Eley N, Mathews A, Batten K, Rogers R, et al. Curriculum development to increase minority research literacy for HIV prevention research: a CBPR approach. Prog Community Health Partnersh 2014 Dec;8(4):511-521 [FREE Full text] [doi: 10.1353/cpr.2014.0059] [Medline: 25727984]
9. George S, Duran N, Norris K. A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. Am J Public Health 2014 Feb;104(2):e16-e31. [doi: 10.2105/AJPH.2013.301706] [Medline: 24328648]
10. Demark-Wahnefried W, Schmitz KH, Alfano CM, Bail JR, Goodwin PJ, Thomson CA, et al. Weight management and physical activity throughout the cancer care continuum. CA Cancer J Clin 2018 Jan;68(1):64-89 [FREE Full text] [doi: 10.3322/caac.21441] [Medline: 29165798]
11. Bauermeister JA, Pingel E, Zimmerman M, Couper M, Carballo-Diéguez A, Strecher VJ. Data Quality in web-based HIV/AIDS research: Handling Invalid and Suspicious Data. Field methods 2012 Aug 01;24(3):272-291 [FREE Full text] [doi: 10.1177/1525822X12443097] [Medline: 23180978]
12. Teitcher JE, Bockting WO, Bauermeister JA, Hoefer CJ, Miner MH, Klitzman RL. Detecting, preventing, and responding to "fraudsters" in internet research: ethics and tradeoffs. J Law Med Ethics 2015 Jan 01;43(1):116-133 [FREE Full text] [doi: 10.1111/jlme.12200] [Medline: 25846043]
13. Kramer J, Rubin A, Coster W, Helmuth E, Hermos J, Rosenbloom D, et al. Strategies to address participant misrepresentation for eligibility in web-based research. Int J Methods Psychiatr Res 2014 Mar 16;23(1):120-129 [FREE Full text] [doi: 10.1002/mpr.1415] [Medline: 24431134]
14. Quach S, Pereira JA, Russell ML, Wormsbecker AE, Ramsay H, Crowe L, et al. The good, bad, and ugly of online recruitment of parents for health-related focus groups: lessons learned. J Med Internet Res 2013 Nov 14;15(11):e250 [FREE Full text] [doi: 10.2196/jmir.2829] [Medline: 24231040]

XSL•FO
RenderX

15. Chen J, Wang Y. Social media use for health purposes: systematic review. J Med Internet Res 2021 May 12;23(5):e17917 [FREE Full text] [doi: 10.2196/17917] [Medline: 33978589]
16. Sanchez C, Grzenda A, Varias A, Widge AS, Carpenter LL, McDonald WM, et al. Social media recruitment for mental health research: a systematic review. Compr Psychiatry 2020 Nov;103:152197 [FREE Full text] [doi: 10.1016/j.comppsych.2020.152197] [Medline: 32992073]
17. Farr DE, Battle DA, Hall MB. Using Facebook advertisements for women's health research: methodology and outcomes of an observational study. JMIR Form Res 2022 Jan 12;6(1):e31759 [FREE Full text] [doi: 10.2196/31759] [Medline: 35019843]
18. Ellington M, Connelly J, Clayton P, Lorenzo CY, Collazo-Velazquez C, Trak-Fellermeier MA, et al. Use of Facebook, Instagram, and Twitter for recruiting healthy participants in nutrition-, physical activity-, or obesity-related studies: a systematic review. Am J Clin Nutr 2022 Feb 09;115(2):514-533 [FREE Full text] [doi: 10.1093/ajcn/nqab352] [Medline: 34669955]
19. Willis TA, Wright-Hughes A, Skinner C, Farrin AJ, Hartley S, Walwyn R, et al. The detection and management of attempted fraud during an online randomised trial. Trials 2023 Aug 04;24(1):494 [FREE Full text] [doi: 10.1186/s13063-023-07517-4] [Medline: 37537678]
20. Kayrouz R, Dear BF, Karin E, Titov N. Facebook as an effective recruitment strategy for mental health research of hard to reach populations. Internet Interv 2016 May;4:1-10 [FREE Full text] [doi: 10.1016/j.invent.2016.01.001] [Medline: 30135786]
21. Pozzar R, Hammer MJ, Underhill-Blazey M, Wright AA, Tulsky JA, Hong F, et al. Threats of bots and other bad actors to data quality following research participant recruitment through social media: cross-sectional questionnaire. J Med Internet Res 2020 Oct 07;22(10):e23021 [FREE Full text] [doi: 10.2196/23021] [Medline: 33026360]
22. Wang J, Calderon G, Hager ER, Edwards LV, Berry AA, Liu Y, et al. Identifying and preventing fraudulent responses in online public health surveys: lessons learned during the COVID-19 pandemic. PLOS Glob Public Health 2023 Aug 23;3(8):e0001452 [FREE Full text] [doi: 10.1371/journal.pgph.0001452] [Medline: 37610999]
23. Birnbaum MH. Human research and data collection via the internet. Annu Rev Psychol 2004 Feb 01;55(1):803-832. [doi: 10.1146/annurev.psych.55.090902.141601] [Medline: 14744235]
24. Pratt-Chapman M, Moses J, Arem H. Strategies for the identification and prevention of survey fraud: data analysis of a web-based survey. JMIR Cancer 2021 Jul 16;7(3):e30730 [FREE Full text] [doi: 10.2196/30730] [Medline: 34269685]
25. Lawlor J, Thomas C, Guhin AT, Kenyon K, Lerner MD, UCAS Consortium, et al. Suspicious and fraudulent online survey participation: introducing the REAL framework. Methodol Innov 2021 Oct 28;14(3). [doi: 10.1177/20597991211050467]
26. Yarrish C, Groshon L, Mitchell JK, Appelbaum AI, Klock S, Winternitz T, et al. Finding the signal in the noise: minimizing responses from bots and inattentive humans in online research. The Behavior Therapist 2019;42(7):235-242 [FREE Full text]
27. Bowen AM, Daniel CM, Williams ML, Baird GL. Identifying multiple submissions in internet research: preserving data integrity. AIDS Behav 2008 Nov 1;12(6):964-973 [FREE Full text] [doi: 10.1007/s10461-007-9352-2] [Medline: 18240015]
28. Ballard AM, Cardwell T, Young AM. Fraud detection protocol for web-based research among men who have sex with men: development and descriptive evaluation. JMIR Public Health Surveill 2019 Feb 04;5(1):e12344 [FREE Full text] [doi: 10.2196/12344] [Medline: 30714944]
29. Dewitt J, Capistrant B, Kohli N, Rosser BR, Mitteldorf D, Merengwa E, et al. Addressing participant validity in a small internet health survey (the restore study): protocol and recommendations for survey response validation. JMIR Res Protoc 2018 Apr 24;7(4):e96 [FREE Full text] [doi: 10.2196/resprot.7655] [Medline: 29691203]
30. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. The Lancet 2007 Oct;370(9596):1453-1457. [doi: 10.1016/s0140-6736(07)61602-x]
31. Sophus AI, Mitchell JW, Barroso J, Sales JM. Factors associated with planned future use of PrEP in the next 3 months and likelihood to use PrEP among Black cisgender HIV-negative women in Texas. AIDS Behav 2024 Jan 28;28(1):72-92. [doi: 10.1007/s10461-023-04188-9] [Medline: 37768428]
32. Sophus AI, Mitchell JW, Sales JM, Braun K. "Our community comes first": investigating recruitment ads that represent and appeal to Black women for online, HIV-related research studies. J Racial Ethn Health Disparities 2024 Dec 18;11(6):3478-3488. [doi: 10.1007/s40615-023-01800-5] [Medline: 37723375]
33. Konstan JA, Simon RB, Ross MW, Stanton J, Edwards WM. The story of subject naught: a cautionary but optimistic tale of internet survey research. J Comput Mediat Commun 2005;10(2). [doi: 10.1111/j.1083-6101.2005.tb00248.x]
34. Mitchell J, Lee JY, Stephenson R. How best to obtain valid, verifiable data online from male couples? Lessons learned from an eHealth HIV prevention intervention for HIV-negative male couples. JMIR Public Health Surveill 2016 Sep 20;2(2):e152 [FREE Full text] [doi: 10.2196/publichealth.6392] [Medline: 27649587]
35. Grey JA, Konstan J, Iantaffi A, Wilkerson JM, Galos D, Rosser BR. An updated protocol to detect invalid entries in an online survey of men who have sex with men (MSM): how do valid and invalid submissions compare? AIDS Behav 2015 Oct 25;19(10):1928-1937 [FREE Full text] [doi: 10.1007/s10461-015-1033-y] [Medline: 25805443]

XSL•FO
RenderX

36. Guest JL, Adam E, Lucas IL, Chandler CJ, Filipowicz R, Luisi N, et al. Methods for authenticating participants in fully web-based mobile app trials from the iReach project: cross-sectional study. JMIR Mhealth Uhealth 2021 Aug 31;9(8):e28232 [FREE Full text] [doi: 10.2196/28232] [Medline: 34463631]
37. Mitchell JW, Chavanduka TM, Sullivan S, Stephenson R. Recommendations from a descriptive evaluation to improve screening procedures for web-based studies with couples: cross-sectional study. JMIR Public Health Surveill 2020 May 12;6(2):e15079 [FREE Full text] [doi: 10.2196/15079] [Medline: 32396133]
38. Griffin M, Martino RJ, LoSchiavo C, Comer-Carruthers C, Krause KD, Stults CB, et al. Ensuring survey research data integrity in the era of internet bots. Qual Quant 2022 Oct 05;56(4):2841-2852 [FREE Full text] [doi: 10.1007/s11135-021-01252-1] [Medline: 34629553]
39. Saberi P. Research in the time of coronavirus: continuing ongoing studies in the midst of the COVID-19 pandemic. AIDS Behav 2020 Aug 18;24(8):2232-2235 [FREE Full text] [doi: 10.1007/s10461-020-02868-4] [Medline: 32303924]

Abbreviations

STI: sexually transmitted infection
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by A Mavragani; submitted 26.04.24; peer-reviewed by V Cofini, V Astha, M Hill; comments to author 20.11.24; revised version received 19.12.24; accepted 25.01.25; published 12.03.25.

Please cite as:
Sophus AI, Mitchell JW
Assessment of Fraud Deterrence and Detection Procedures Used in a Web-Based Survey Study With Adult Black Cisgender Women: Description of Lessons Learned and Recommendations
JMIR Form Res 2025;9:e59955
URL: https://formative.jmir.org/2025/1/e59955
doi:10.2196/59955
PMID:

©Amber I Sophus, Jason W Mitchell. Originally published in JMIR Formative Research (https://formative.jmir.org), 12.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Trauma Support App for Young People: Co-design and Usability Study

Maria Thell[1], MSc; Kerstin Edvardsson[1], MSc; Reem Aljeshy[1]; Kalid Ibrahim[1]; Georgina Warner[1], MA, PhD

Child Health and Parenting, Department of Public Health and Caring Science, Uppsala University, Uppsala, Sweden

Corresponding Author:
Georgina Warner, MA, PhD
Child Health and Parenting, Department of Public Health and Caring Science
Uppsala University
BMC, Husargatan 3
Uppsala, 751 23
Sweden
Phone: 46 18 471 6632
Email: georgina.warner@uu.se

Abstract

Background: One of the most common reasons young people with mental health issues, such as posttraumatic stress disorder, do not seek help is stigma, which digital support tools could help address. However, there is a lack of trauma support apps specifically designed for young people. Involving the target group in such projects has been shown to produce more engaging and effective results.

Objective: This study aimed to apply a child rights–based participatory approach to develop a trauma support app with young people.

Methods: Seven young people (aged 14-19 years; 3 males and 4 females) with experiences of trauma were recruited as coresearchers. A child rights–based framework guided the working process. The app was developed through a series of Design Studio workshops and home assignments, using the manualized intervention Teaching Recovery Techniques as the foundation for its content. The coresearchers were trained in research methodology and conducted usability testing with other young people (n=11) using the think-aloud method, the System Usability Scale (SUS), and qualitative follow-up questions.

Results: A functional app prototype was developed using a no-code platform, incorporating various trauma symptom management techniques. These techniques covered psychoeducation, normalization, relaxation, and cognitive shifting, presented in multiple formats, including text, audio, and video. The contributions of the coresearchers to the design can be categorized into 3 areas: mechanics (rules and interactions shaping the app's structure), dynamics (user-visible elements, such as the outcome when pressing a button), and aesthetics (the emotional responses the app aimed to evoke in users during interaction). Beyond influencing basic aesthetics, the coresearchers placed significant emphasis on user experience and the emotional responses the app could evoke. SUS scores ranged from 67.5 to 97.5, with the vast majority exceeding 77.5, indicating good usability. However, usability testing revealed several issues, generally of lower severity. For instance, video content required improvements, such as reducing light flickering in some recordings and adding rewind and subtitle selection options. Notably, the feature for listening to others' stories was removed to minimize emotional burden, shifting the focus to text formats with more context.

Conclusions: Young people who have experienced trauma can actively participate in the cocreation of a mental health intervention, offering valuable insights into the needs and preferences of their peers. Applying a child rights–based framework to their involvement in a research project supported the fulfillment of the Convention on the Rights of the Child Article 12.

(JMIR Form Res 2025;9:e57789) doi:10.2196/57789

KEYWORDS

co-design; young people; trauma; app development; usability testing

XSL•FO
RenderX

Introduction

Background

It is well recognized that one of the main barriers preventing young people from seeking face-to-face help for mental health problems, such as symptoms of posttraumatic stress disorder (PTSD), is stigmatization and embarrassment [1-3]. This includes perceived stigma, fear of stigma, societal stigma, and self-stigma [2], and may result in many young people experiencing PTSD symptoms not receiving the help and support they need. However, these well-known barriers can potentially be overcome through app-based support [3], as mental health apps can engage young people who might not otherwise seek help through traditional routes. Particularly, in the case of PTSD, which differs from other mental illnesses due to its clear cause, young people could receive scalable trauma interventions without delay through digital technology [4]. Despite this, there is a lack of apps specifically developed for young people [5,6].

Teaching Recovery Techniques (TRT) is a manualized group intervention for children aged 8 and older with trauma symptoms [7]. The intervention consists of 5 sessions for children and 2 sessions for parents and caregivers. The content for children and adolescents includes components aligns with trauma-focused cognitive behavioral therapy: psychoeducation, trauma narrative, affective modulation, cognitive coping and processing, in vivo mastery of trauma memories, and future development. TRT helps normalize responses to trauma and provides strategies to manage intrusive thoughts and memories as well as regulate arousal. Participants in a TRT group also work on gradually exposing themselves to thoughts and situations they avoid [7]. TRT has been evaluated in randomized controlled trials with children and young people in several international settings, with an overall positive impact on symptoms of PTSD and depression reported [8]. The initial idea for the current project emerged during a previous participatory study in which TRT was adapted to an online format in response to the COVID-19 pandemic [9]. During the adaptation phase, which involved TRT group leaders, parents, and young people with experience of TRT, the young people suggested consolidating all the techniques into a separate digital format, that is, a mobile app [9].

There has been a growing use of participatory methods with children and young people in the development of health interventions [10]. Previous research indicates that involving the target group in such projects can lead to more engaging and useful outcomes [11,12]. However, when conducting participatory research in the context of PTSD, it is important to recognize that traumatic events create a power imbalance, where one entity holds power over another. An individual's experience may be shaped by feelings of powerlessness, humiliation, guilt, shame, betrayal, or silencing. In interpersonal interactions, especially in public health research, it is crucial to ensure that these feelings of powerlessness are not replicated or reinforced [13].

It has been suggested that research involving young people as coresearchers should be conducted within a rights-based framework [14]. One argument is that this approach can help address the power imbalance between adults and children. The Convention on the Rights of the Child (CRC) [15] acknowledges the position and disempowerment of children in matters affecting them. Adopting a child rights–based approach as a researcher involves implementing and being accountable to the rights set out in the CRC [16]. For this study, we applied the Lundy model [17] of child and youth participation, including the associated planning checklist for researchers [18]. The Lundy model conceptualizes the two components of Article 12 of the CRC: (1) the right to express views and (2) the right to have those views duly considered. By organizing the 4 elements—space, voice, audience, and influence—in a logical, chronological order, the model helps researchers plan, implement, and monitor projects. Space refers to providing a safe and inclusive environment for children to express their views and ensuring that all children can participate. Voice involves offering appropriate information and facilitating the expression of children's views. Audience ensures that children's views are communicated to someone responsible for listening. Influence involves ensuring that children's views are taken seriously and acted upon where appropriate [17].

Objective

This study aimed to (1) apply a child rights–based participatory approach to developing a trauma support app with young people who have personal experiences of trauma and (2) conduct usability testing of the app with the young people as coresearchers.

Methods

Phase 1: App Development

Coresearcher Recruitment

Seven young people were recruited for the project through 3 different channels. Two of them had previously collaborated with the research team on a related TRT research project [9] and were invited to participate again. Additionally, contact was made with a youth-run organization for young people with experience in out-of-home care, which agreed to distribute the recruitment advertisement to its members. Furthermore, a youth center in an area categorized as "vulnerable" by the Swedish Police Authority due to high rates of crime and social exclusion [19] posted the advertisement on its physical noticeboard and social media accounts. Examples of traumatic events and situations, as well as PTSD symptoms, were included in the recruitment materials. Young people who identified with these experiences and enjoyed being creative were encouraged to contact the researchers via email, SMS text message, or WhatsApp (Meta Platforms, Inc.). The advertisement emphasized that participants were not expected to share their personal experiences while working on the project.

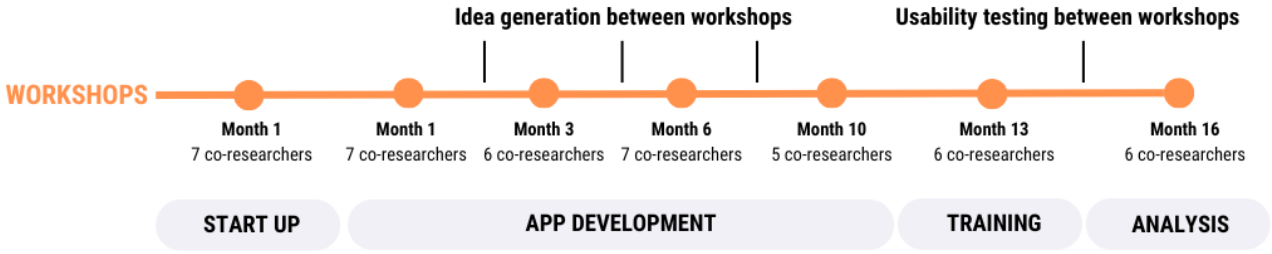
Following an initial individual dialogue with all interested young people, 7 chose to participate as coresearchers in the project. The youth coresearcher team consisted of 3 boys and 4 girls, aged 14-19 at the start of the project. They lived in different neighborhoods in 1 large city and 1 medium-sized city in southern Sweden. The older youth attended both theoretical and practical secondary school programs. In accordance with the regulations of the Swedish Work Environment Authority [20],

XSL•FO
RenderX

information about the project was provided to the guardians of participants under 18 years of age. Additionally, permission to participate was obtained from the guardians of those under 16 years of age [20]. The coresearchers were remunerated for their time spent on project activities in line with existing recommendations [21].

Before starting the app development process, a "start-up" session was conducted (Figure 1) to help the team get acquainted through games and other activities. During this session, decisions were made about how to collaborate, including setting common approaches and ground rules to guide the work throughout the project. At the end of the session, an experienced TRT group leader visited to introduce the intervention. As part of this introduction, they guided the group through a popular visual imagery technique called "Safe Inner Place."

Figure 1. Overview of the Teaching Recovery Techniques app development process with youth co-researchers conducted over 10 months.



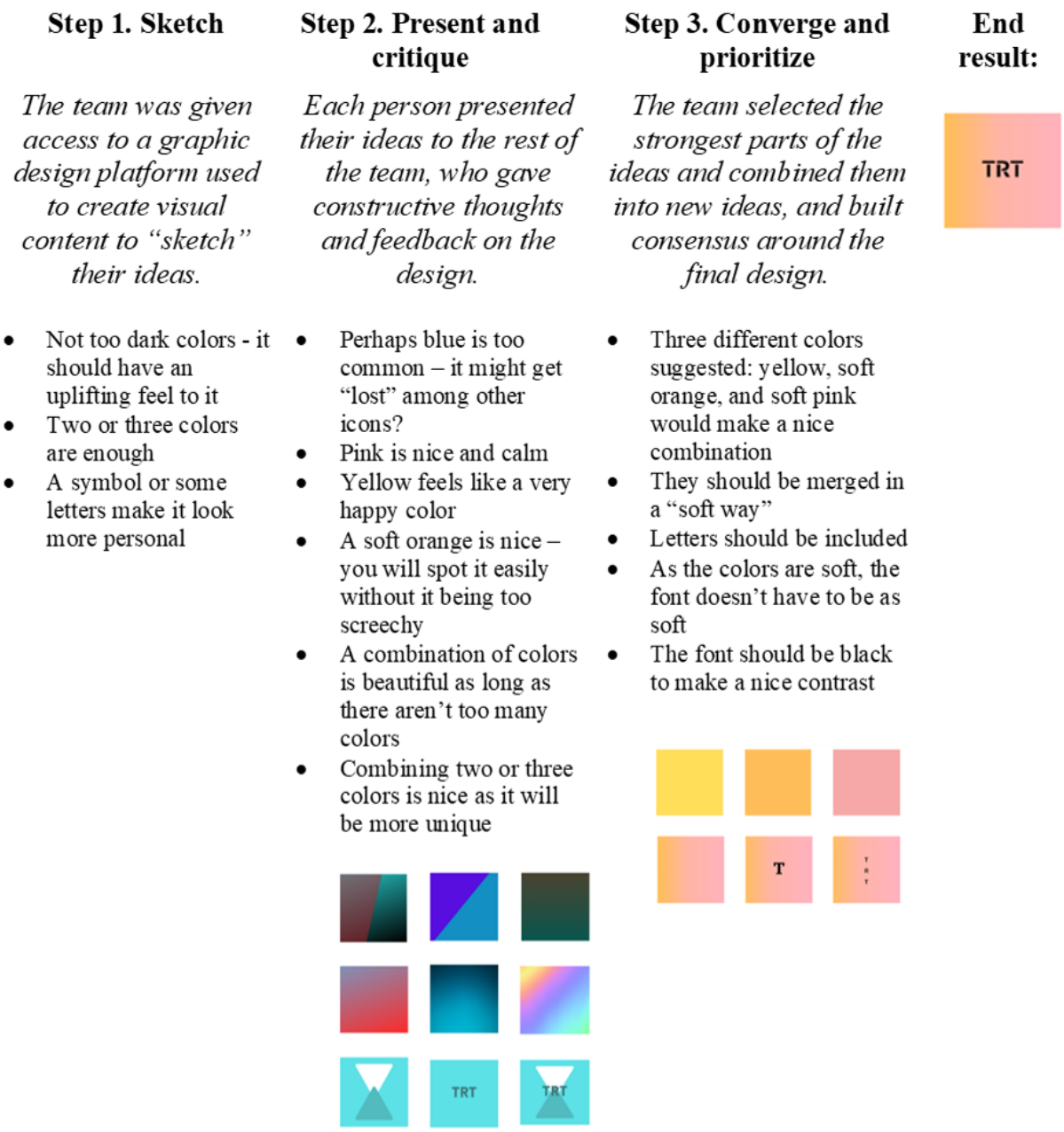
App Development Workshops

The app was developed over 4 workshops spanning 10 months (Figure 1), using an adapted version of the Design Studio Workshop method [22]. In addition, the coresearchers individually generated ideas between workshops. To support their independent work, app design prototypes, along with questions and topics for the next workshop, were shared with them. The researcher leading the workshops (MT) collected the individual contributions before each session and presented them to the group for collaborative refinement.

The first app development workshop aimed to cocreate a logo for the app and pilot the Design Studio Workshop method [22]. The coresearchers were given tablets with internet access and individually explored ideas, first by analyzing their favorite app logos and then by considering a design for a trauma support app. They submitted their ideas to the researcher leading the workshops (MT), who compiled them and projected them onto a screen using a digital noticeboard software tool [23]. Applying the Design Studio method, the team explored colors, fonts, and other design elements. The final logo incorporated the strongest elements from the generated ideas, featuring "soft," "encouraging," and "happy" colors while avoiding excessive or overused colors in app logos. The coresearchers selected a font that contrasted with the background and stood out from other popular apps. Figure 2 illustrates the logo development process.

XSL•FO
RenderX

Figure 2. Overview of the Teaching Recovery Techniques (TRT) app logo development process using the Design Studio method.



The second workshop focused on prioritizing content and layout for the app using the TRT manual. In the preparatory phase, coresearchers were provided with concise written descriptions of the techniques outlined in the TRT manual. While simplified, these descriptions carefully preserved the original phrasing to maintain the integrity of the intervention logic. Coresearchers were tasked with considering how each technique could be integrated into an app format. To accommodate diverse preferences and abilities, they were encouraged to express their ideas in various ways—whether orally, in writing, or through visual representations such as images or vision boards. Consequently, the initial "sketch" step of the Design Studio method was conducted independently. Upon convening for the workshop, participants engaged in the second step, during which the researcher leading the workshops (MT) presented the collective ideas and facilitated a critique of the proposed concepts. The ideas were then synthesized and prioritized. To discuss app content related to traumatic experiences, case vignettes were used in the workshops, allowing coresearchers to explore different traumatic experiences in a general rather than personal context. Following the workshop, the selected concepts were developed into a prototype of the TRT app using a no-code app-building program [24].

During the third workshop, the coresearchers were presented with an app prototype developed by a research assistant (KE), incorporating their ideas. They provided feedback on both the content and design. For certain techniques, the prototype successfully met their expectations. In other cases, minor

XSL•FO
RenderX

adjustments were made in real time during the workshop using the app-building program [24]. For techniques requiring more substantial revisions, the Design Studio cycle was reinstated. This process involved coresearchers initially "sketching" the necessary amendments, followed by presentations and critiques, ultimately converging and prioritizing the changes. Having access to the app-building program during the workshops facilitated this process, as coresearchers could explore the range of functionalities available within the program. Ideas that the coresearchers found valuable but were not feasible to implement at the time were placed in what the team collectively called the "ice box." This meant that the ideas would be revisited in any future development of the app. One such example was the app's color scheme, which had been jointly agreed upon. Some coresearchers had also expressed a preference for an alternative, slightly darker color scheme for nighttime use. While the first prototype featured only a single color scheme, the idea of selectable color schemes was considered both valuable and important. Rather than being discarded, it was added to the ice box for future consideration.

Further revisions to the app were made after workshop 3 and presented to the coresearchers for feedback in workshop 4. Some changes, primarily concerning the wording of written content, were informed by discussions among the academic researchers to ensure the integrity of the TRT intervention logic. Any modifications resulting from these discussions were communicated to the coresearchers during workshop 4. As in the previous workshop, coresearchers were invited to provide feedback on the prototype. After a series of minor revisions, they expressed their satisfaction with the product, and a collective decision was made to transition from app development to the usability testing phase of the project.

Overall, attendance at the workshops remained consistently high (Figure 1). One of the 7 coresearchers participated in the entire development process but did not take part in the training and usability testing described below. Other absences were due to illness and were compensated for through individual meetings or conversations with the researcher leading the workshops (MT). Those who were ill also completed the home assignments related to the missed session.

Phase 2: Usability Testing

Coresearcher Training

The usability testing was conducted by the coresearchers, with no academic researchers present in the room during data collection. Given the youth-led nature of the process, training sessions were held before the usability testing began (Figure 1). The training covered both theoretical and practical aspects of research ethics and methodology, enabling the coresearchers to actively lead and carry out the research tasks [25-27].

Careful consideration was given to the safety of study participants. As part of the training, the coresearchers and the academic researcher reflected together on how to recognize when a person might be negatively affected and practiced the necessary actions to take if needed. While training is considered essential for young people conducting research, it is also important to balance the extent of training with maintaining their ability to participate meaningfully [14]. Although the training in this study was based on existing guidelines for research with children [25], it was adapted for young people by condensing the content into 3 sessions held on the same day, with breaks in between. Plans were in place to provide additional sessions if needed, but both the coresearchers and researchers agreed that the initial sessions provided sufficient preparation.

Participant Recruitment

The target group for the usability testing consisted of young people aged 15-21 years who spoke and understood Swedish. As the purpose of the usability testing was to assess the intuitiveness of the design and evaluate it with users who had no prior exposure to the app—rather than to determine its efficacy as a trauma support tool—a population sample was recruited instead of a clinical sample with trauma symptomatology. The goal was to recruit at least five and up to 15 participants, ensuring age and gender representation. Research has shown that gender plays a role in usability testing with children and young people, with girls tending to report more issues than boys [28]. Studies indicate that testing with 5 participants per target group is sufficient to identify approximately 80% of usability problems [29].

The recruitment process consisted of 2 steps. First, the coresearchers identified other young people within their networks, provided initial information, and inquired whether they might be interested in participating in the study. These young people were primarily from the coresearchers' school networks and, like the coresearchers, resided in 1 large city and 1 medium-sized city in southern Sweden. If interested, they were then contacted by the academic researchers, who provided additional oral and written information and obtained consent for participation. The date and time of each usability test were mutually decided, with most sessions taking place in the evenings, as the coresearcher, participant, and researcher all needed to be physically present on-site. Five of the coresearchers conducted individual usability tests with participants (n=11; 6 male and 5 female; 15-20 years) over a 6-week period in January and February 2022.

Usability Testing Methods

To assess the app's usability, the think-aloud method was applied for data collection. This method involves participants verbalizing their thoughts while completing specific or nonspecific tasks. Participants are encouraged to articulate what they are looking at, thinking, doing, and feeling as they navigate the app. During the usability test, the observer records participants' verbalized thoughts and actions without interpreting them, paying particular attention to moments or areas where users experience difficulties [30].

All usability tests were conducted on university premises, with the coresearcher and participant in the same room and an academic researcher in an adjacent room. The participant was provided with a mobile phone and instructed to navigate through the entire app using the main menu while thinking aloud. The coresearcher documented the participant's verbalized thoughts and observed behaviors as described above. After completing the app test, participants were asked by the coresearcher to

XSL•FO
RenderX

complete the System Usability Scale (SUS) [31], which had been translated into Swedish. Following this, they participated in a structured interview led by the coresearcher to capture their general impressions of the app [32]. All notes from the test sessions were documented in Research Electronic Data Capture (REDCap), a secure online survey platform, during the session. No voice or video recordings were made.

Usability Testing Analysis

It was agreed in advance that immediately after each session, the coresearcher would report any major or serious usability problems identified during testing to the academic researcher. Following the procedure described by Hertzum [32], 2 academic researchers independently reviewed all notes from the test sessions, each compiling a list of positive and negative usability issues. These lists were then compared, discussed, and merged into a single final list. This consolidated list of usability issues was presented to the coresearchers in a dedicated meeting (Figure 2). Using a live interaction polling tool [33], the usability issues were displayed, and coresearchers individually rated their severity and suggested a priority order. The final prioritization was determined through group discussion.

Ethical Considerations

The coresearchers participated in the project as experts rather than as participants of the study; therefore, standard ethical requirements did not apply. However, their safety was carefully considered throughout the process. From the outset, it was made clear that coresearchers were not expected to share personal experiences of trauma, as this was not necessary for the project. Discussing one's own traumatic experiences could lead to retraumatization or secondary traumatization of others in the group [34]. To mitigate this risk, case vignettes were used in the workshops when developing app content related to trauma. This approach enabled coresearchers to engage in discussions about different traumatic experiences in a general manner rather than on a personal level.

Another key safety consideration was the nature of the tasks assigned to coresearchers between workshops. To minimize the risk of any negative impact, tasks conducted outside the workshop setting focused on app design and layout rather than potentially sensitive content.

The usability testing phase, involving young people as study participants, was submitted to and approved by the Swedish Ethical Review Authority (reference number 2022-04217-01). Study participants were fully informed about the research, and their consent was documented. As participants had access to the app's full content during testing, a safety protocol was implemented. Before each test session, the academic researcher checked in with participants to assess their well-being. If a participant appeared unwell, the test session was not conducted. After each session, the academic researcher conducted separate follow-up conversations with both the participant and the coresearcher to ensure that no one had been adversely affected by the testing experience. If any issues arose during a session, such as a participant showing signs of distress, the coresearcher conducting the test would have immediately alerted the researcher, as practiced during the training session. As part of the safety protocol, the researcher documented the contact details of guardians for participants under 18 years of age before each test session.

Results

Phase 1: App Development

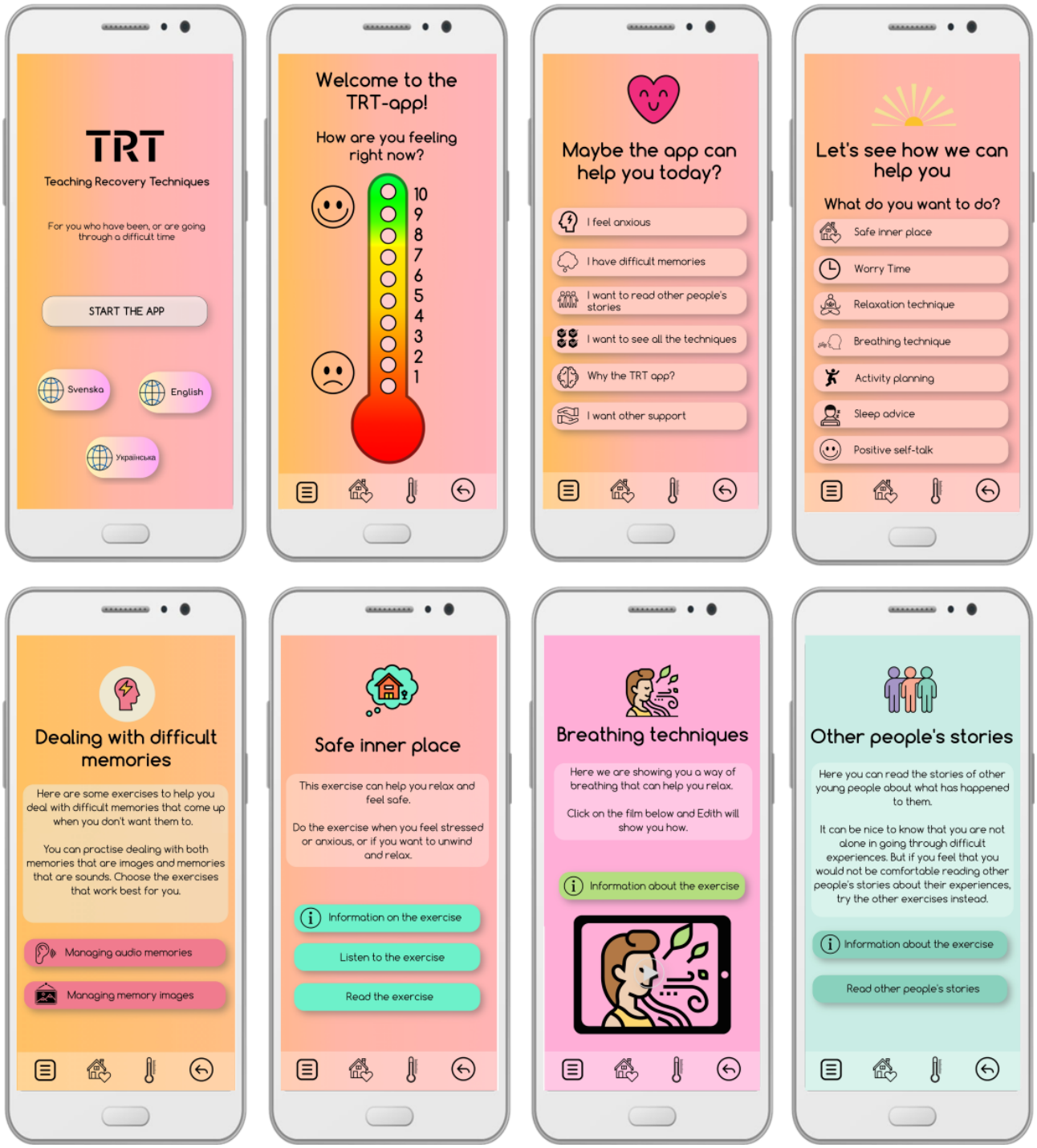
Overview

The TRT app was developed as a progressive web app using a no-code, cross-platform application development program [24]. During the initial discussions about compiling TRT into an app, the young people highlighted potential issues with native apps, particularly regarding data storage requirements. By building the app as a progressive web app, these concerns were mitigated while still providing a user experience comparable to that of a native app.

The coresearchers' contributions to the app's design can be categorized into 3 areas using the Mechanics, Dynamics, and Aesthetics (MDA) framework [35]. Mechanics refers to the fundamental rules and interactions that define the app's structure. Dynamics encompasses what the user experiences in real time, such as the response when a button is pressed, which directly supports the app's aesthetics. Aesthetics pertains to the intended emotional responses elicited in users when interacting with the app and includes its purpose and narrative. Beyond influencing visual elements such as colors, fonts, and symbols, the coresearchers placed significant emphasis on the overall user experience. A selection of app screenshots is presented in Figure 3.

XSL•FO
RenderX

Figure 3. A selection of screenshots from the Teaching Recovery Techniques (TRT) app.



Mechanics

The coresearchers' contribution to the app's mechanics—its basic rules and structure—included selecting techniques from the TRT manual, determining the format of the exercises, and structuring the app's content. They emphasized the importance of the app's structure, particularly the role of the main menu, which is the first thing users see upon entering the app. A clear and intuitive menu ensures easy and efficient navigation, allowing users to quickly find what they need. This not only saves time but also minimizes frustration when searching for specific features or content. The menu provides an overview of the app's content, enhancing the user experience and improving usability. By featuring a front page with a well-organized menu, users can easily select their area of interest—whether it be reading texts, watching films, or completing exercises.

Both coresearchers with prior experience of TRT and those without agreed that the techniques in the app should follow the same order as in the manual—that is, as they would be introduced in a TRT group. There was consensus that the technique groupings should be labeled according to how young people who have had difficult experiences might feel (eg, "I feel anxious" or "I have intrusive memories"; Figure 3), making navigation more intuitive. Throughout the development process, coresearchers and academic researchers discussed the differences

XSL•FO
RenderX

between being introduced to the techniques via an app versus in a facilitated group session. In the latter, group leaders present the exercises and provide support in response to participants' reactions. Given this distinction, the coresearchers emphasized the importance of clear presentation within the app, ensuring that each exercise included information on its purpose and how to use it effectively.

Dynamics

The coresearchers emphasized the importance of the app being adaptable to the preferences of the young person using it, which is central to the app's dynamics—the visible elements and user interactions. The content primarily consists of simple illustrative icons, instructional and demonstrational films, audio recordings, and short texts. Long texts were seen as detrimental to the user experience; however, careful consideration was required to shorten the content while maintaining fidelity to the TRT manual and the therapeutic intent of each technique. The descriptive texts introducing the techniques were kept short and simple. For young people seeking more information about the purpose of each technique, an "Information about the exercise" option is available (Figure 3). To enhance customization, most exercises offer the choice of either reading or listening to the instructions. Additionally, exercises presented as videos include optional subtitles.

Aesthetics

The app's visual aesthetics were intentionally kept simple. The screen layout was designed to be uniform, featuring a representative icon and header at the top, followed by text or buttons framed by softly rounded boxes. An "easy-to-read" font with rounded edges was selected. The orange-pink color of the logo was chosen for its soft and friendly appearance—described as being "like the sun coming in from the left." The app's color theme was primarily based on the logo's colors, with additional shades introduced throughout development to enhance variation and create clearer visual distinctions between different techniques (Figure 3).

The app's aesthetics also relate to the emotional responses it aimed to evoke in users. The coresearchers emphasized that in-app films and voice recordings can be particularly helpful in promoting a sense of calm. These elements were seen as valuable for creating a relaxed atmosphere, providing guidance and instructions, and adding variety and creativity to the app experience. At the same time, the coresearchers stressed the importance of offering different exercise formats, recognizing that individuals have diverse preferences and needs when it comes to relaxation. While some may prefer breathing exercises, others might find meditation or guided visualization more effective. By providing a wide range of exercises, the app can better accommodate different users. In the videos, exercises are led by young people, as the coresearchers underscored the importance of featuring individuals with whom they can identify or relate.

Overview of App Content and Navigation

Basis of Support Techniques

The trauma support techniques featured in the app are based on the TRT manual and incorporate several common approaches to trauma intervention, including psychoeducation, normalization, relaxation, and cognitive shifting. Psychoeducation provides individuals with information to enhance their understanding of psychological conditions—PTSD in this context. Normalization helps users recognize that their symptoms are not signs of weakness or character flaws but rather a natural response to their psychological condition. Relaxation techniques aim to alleviate stress-related symptoms, while cognitive shifting involves deliberately redirecting one's attention from a singular fixation to alternative focal points. For safety reasons, the TRT app intentionally excludes in vivo exposure—a technique that involves directly confronting a feared object, situation, or activity in real life. A comprehensive overview of the app's content is provided below.

Main Menu

To facilitate navigation for both new and experienced users, the app offers 2 ways to explore its content. In the main menu, the "Why the TRT app?" option provides first-time users with an introduction to TRT, trauma, trauma reactions, and how the app is intended to be used. Additionally, the main menu is structured around the user's immediate needs, suggesting groups of exercises based on whether they are feeling anxious, experiencing intrusive memories, or seeking to learn more about trauma through psychoeducation and normalization of trauma reactions. For young people who are more experienced with TRT or the app and want to find a specific exercise, the main menu includes the option "I want to see all techniques." This section lists all exercises in the order they appear in the TRT manual, allowing users to either select an individual exercise or follow a sequence similar to that of a facilitated TRT group. Additionally, the main menu features the option "I want other support," which provides a comprehensive list of care and support organizations, including phone numbers and links to relevant web pages.

Thermometer

The Thermometer is the first screen displayed in the app, prompting the young person to rate how they are feeling on a scale from 1 (low mood) to 10 (high mood) before accessing the app's content. If they report a score of 3 or lower, they are directed to a screen with instructions to call the national emergency number or contact a suicide prevention line if they are experiencing thoughts of self-harm. From there, they can choose to either proceed directly to the main menu or visit the "Other support" screen, which provides a list of care and support organizations. If a score above 3 is reported, the young person is encouraged to return to the Thermometer if their mood changes while using the app, before being directed to the main menu. The Thermometer remains accessible from every screen in the app via a tab bar (described below).

Safety Features

Techniques from the TRT manual that involve in vivo exposure or pose a risk of causing high levels of arousal or negative

XSL•FO
RenderX

emotions have been excluded from the app or modified to minimize these risks. Techniques in the app that still involve recalling traumatic memories include shortcuts to safety features. Young people who feel they need additional support beyond what the app provides can access the "Other support" screen from the main menu or the "Thermometer" if they score low. This screen offers an extensive list of links and contact details for various digital and in-person care and support services for young people, providing more personalized help. The app only lists these services; no referrals are made.

The "Safe Inner Place" technique helps young people take control of intrusive memories and relax their bodies during periods of high arousal. This visual imagery technique allows users to create positive mental images, enabling them to influence their imagination and emotions. It offers 2 options: reading or listening to the guide. The Safe Inner Place serves as a secure base that users can return to when experiencing high arousal. Easy access to this feature from the tab bar (described below) was seen as empowering, allowing young people to explore more challenging exercises with greater confidence.

Tab Bar

A tab bar with 4 icons is located at the bottom of each screen in the app. These icons provide quick access to the Main Menu, the Safe Inner Place technique, the Thermometer, and the previous screen. The tab bar is designed to facilitate navigation while also serving as a safety feature. If a young person experiences high physiological arousal or negative emotions while using the app, they can quickly access the Safe Inner Place relaxation technique or the Thermometer. If they indicate a low level of well-being, the Thermometer will present them with options for additional care and support.

Anxiety Reducing Exercises

From the main menu, selecting "I feel anxious" directs the young person to a submenu of techniques designed to help manage negative emotions related to trauma. In addition to the Safe Inner Place technique, 6 other techniques support emotional regulation and coping with arousal and intrusive thoughts. Among them, 2 video-based exercises—*Breathing Technique* and *Relaxation Exercise*—are specifically designed to promote relaxation and help the young person regulate physiological arousal.

Activity Scheduling is a text-based technique designed to encourage social and physical activity while providing a distraction from anxious thoughts and feelings. The exercise offers 3 options: receiving suggestions for meaningful activities, creating a personalized list of activities, or scheduling planned activities. The suggested activities are provided by youth advisors to ensure they are relevant to young people. Another text-based technique, *Worry Time*, encourages setting aside 10 minutes each day for intrusive thoughts and worries, helping to limit worry time and regain control over intrusive thoughts. *Sleep Advice* provides a list of tips to improve sleep hygiene for young people experiencing sleep difficulties or nightmares due to trauma. Finally, *Positive Self-Talk* helps young people recognize the connection between thoughts and feelings and learn to replace negative self-perceptions with positive ones. The exercise offers a choice between receiving suggested positive affirmations or creating personalized ones.

Techniques for Intrusion Control

Selecting the option "I have intrusive memories" from the main menu provides a choice of techniques to help manage auditory or visual intrusive memories. *Coping with Sound Memories* is a video-based exercise where the young person practices controlling intrusive auditory memories by mentally manipulating the sound in different ways. Coping with Visual Memories includes 2 exercises: *Screen Techniques* and *Clapping Technique.* Screen Techniques is a video-based exercise that, similar to the Auditory Intrusion Exercise, guides users in manipulating visual memories as if they were displayed on a TV screen or in the palm of their hand to regain control over them. The Clapping Technique, known as bilateral stimulation in the TRT manual—a term the coresearchers found too complex for the app—demonstrates how rhythmically tapping the knees while recalling intrusive images can help modify and regulate those memories.

Other People's Stories

The "Other People's Stories" page is accessible via a button in the main menu. It features a collection of fictional stories in which young people describe their personal experiences of trauma in writing. The purpose of these stories is to provide psychoeducation and normalize emotional and psychological reactions to trauma. The traumatic events depicted were suggested by the coresearchers, with plans to expand the content to include a more diverse range of experiences. At the time of usability testing, the available stories covered themes such as losing a parent, being bullied, and experiencing war. Each story is clearly labeled with a heading that indicates its subject matter.

Phase 2: Usability Testing

For the SUS, complete data were available from 10 participants, with scores ranging from 67.5 to 97.5 [32]. Based on existing reference values, a product must achieve an SUS score of at least 71 to be in the top half [32,36]. Nine out of the 10 completed SUS recordings had scores of 77.5 or higher.

Table 1 summarizes the identified usability issues along with the corresponding modifications. Some issues were related to technical preferences, such as the ability to rewind video and audio files to revisit missed content or replay important segments. Additionally, the option to enable subtitles was considered important for those who wished to use them. As a result, rewind and subtitle features were added to all audio and video recordings. While the video format was well received, some recordings were noted to have low quality, with slight flickering from background lighting. Some participants reported that this negatively impacted the overall user experience. To address this, all relevant videos were rerecorded in a professional studio, with the same young person demonstrating the techniques.

XSL•FO
RenderX

Table 1. Usability issues identified with the prototype of the Teaching Recovery Techniques app and the corresponding modifications.

Usability issues	Modifications to the app
Some of the video and audio recordings lacked the ability to rewind as well as the option for subtitles.	Rewind function and option for subtitles were added to all videos and recordings.
Users found the quality of some video and audio recordings to be low, affecting the overall user experience negatively.	The videos were rerecorded in a studio. The audios were rerecorded by the same narrator, while ensuring no interferences in the recording.
Users were bothered by difficulties related to the device, such as the screen going black during long audio and video recordings, and the lower tab bar being difficult to access with an iPhone.	Instructional text was added, directing the user to the device settings to enable the screen to stay on for an extended period. Adjustments were made so access to the tab bar is not affected by the type of device.
The *Other People's Stories* exercise, aimed at identification and normalization, was generally perceived as important and helpful, but the audio recordings were also perceived as stressful to listen to. It could also be perceived as pressurizing to hear from others who have managed to overcome their difficulties.	A decision was made to keep *Other People's Stories* in text format but exclude the audio recordings from the app. Additional information about the contents of the stories, as well as the purpose of the exercise, was added. A reference to other exercises in the app in case the user feels they would not benefit from reading other people's stories was also added.
Users reacted to the advice to "reach out to someone close to you" and expressed that some people don't have that option, which may be why they use the app.	A reference to the *Other Support* screen was added.
There was too little variety in the tips on sleep, activities, and positive self-talk.	More tips were added for each exercise.
The app only being available in Swedish.	The app was translated into more languages.

Usability issues related to device functionality were also identified. Some video and audio recordings lasted several minutes, and if participants remained passive during playback—without touching the screen—the display would time out, disrupting their ability to focus on the exercise. To address this, a reference was added to each technique's information, guiding users on how to adjust screen time-out settings on their devices. Another device-related issue involved the tab bar at the bottom of each screen, which was unnecessarily difficult to access on iPhones. Adjustments were made to ensure a consistent user experience of the tab bar across different devices.

Some aspects were identified as potentially incompatible with the reality of the target group. Participants responded to the recurring advice to "talk to a trusted person," noting that some users might not have this option—which may, in fact, be why they turn to the app. To address this, a clarification was added, stating that if the user cannot or does not want to talk to someone close to them, the "Other support" section in the app provides several recommendations for both in-person and digital support services.

Additionally, the option to listen to other people's stories, narrated by young people for identification purposes, was found to carry the risk of causing undesirable feelings. Hearing someone speak about their personal difficulties could be burdensome and might also create pressure for the listener to cope with their own situation. As a result, the audio option was removed, leaving only the text format. Additional information was added to clarify the content and purpose of the stories. A reference to other exercises in the app was also included, along with the following message: "If you feel that reading other people's stories may not be helpful for you, you can try these exercises instead."

Discussion

Principal Findings

This project was conducted with young people who had personal experiences of trauma, actively involving them as experts from an early stage. In line with the Lundy model, creating a safe and inclusive space for children to express their views proved essential from the outset and remained a priority throughout [17]. All activities, except for the usability tests, were held in a conference room at a centrally located hotel. The coresearchers expressed early on that they appreciated the setting, describing the space as cozy, the food and refreshments as enjoyable, and the environment as one where they felt comfortable expressing themselves.

Careful consideration was given to recruitment strategies to ensure that young people from diverse backgrounds and experiences were informed about the project. While previous participatory research projects have highlighted the challenges of recruiting young people [37], our experience was different, with several young individuals quickly expressing interest. The use of multiple recruitment arenas proved valuable, and further exploration of effective strategies for engaging young coresearchers is warranted.

With regard to voice [17], coresearchers were encouraged to express themselves in various formats. Some felt comfortable speaking freely in workshops, while others preferred responding to direct questions. One coresearcher expressed early on that writing down their thoughts during workshops would be easier than speaking in a group. This option was made available and encouraged for all participants. Even those who preferred verbal communication occasionally utilized written expression. A standing agreement ensured that coresearchers could share their thoughts through writing, drawings, or digital images, sometimes compiled into vision boards. Several coresearchers contributed their insights this way at least once during or after a workshop.

XSL•FO
RenderX

Likewise, both written and verbal information was consistently provided to accommodate different preferences.

Research has shown that participatory approaches with young people must be flexible and pragmatic, aligning with their preferences and opportunities. Participation should be offered at different levels and in various formats, allowing young people to decide how they engage [38]. In this project, coresearchers' preferences guided practical aspects such as venue selection, scheduling, and breaks, including refreshments (see Sarkadi et al [39] for more details). All workshops were held in the same easily accessible venue, ensuring convenience for participants using any mode of transport. To foster a safe and inclusive environment, coresearchers collaboratively developed guidelines outlining how they wished to work together, including common rules for respect and understanding. A key expectation, as expressed in these discussions, was that open conversations would be easier if participants were assured that no one would interrupt or dismiss their opinions.

Sufficient time was allocated to allow researchers to work cocreatively with the coresearchers, ensuring the flexibility required for this type of collaboration. Joint workshops were scheduled in the evenings and on weekends, allowing coresearchers to participate without missing school—an aspect they highlighted as important. Additionally, researchers remained accessible throughout the project via multiple communication channels, accommodating individual preferences. Early in the project, a discussion was held to establish mutually agreed-upon times for communication. The coresearchers generally preferred communicating in the evenings but understood that responses might sometimes take until the next day. This flexibility contributed to consistently high workshop attendance. When a coresearcher was unable to attend, they were offered an individual meeting or conversation with a researcher, as they expressed a strong interest in staying engaged with the project's progress.

While enabling expression is important, so is providing the option not to participate. Early in the project, the coresearchers emphasized that they sometimes have a lot going on in their lives, particularly with school. To minimize absences, proposed dates and times were communicated in advance, allowing for joint decisions on scheduling. Additionally, the youth advisors were reminded that they could choose which parts of the process they wanted—and were able—to be involved in.

Audience refers to whether and how young people's views are being listened to [17]. Each workshop was attended by a researcher (MT) and the research assistant responsible for app development (KE), ensuring a direct communication pathway from coresearcher idea generation to app implementation. Having 2 adults present also allowed for more effective engagement—while one (KE) took notes, the other (MT) could focus on facilitating discussions and actively listening. To demonstrate that their input was valued, ideas from previous workshops and independent work tasks were shared back with the coresearchers at each session. This process was further reinforced by sharing app prototypes developed based on decisions made in previous workshops. This aligns with the Influence component of the Lundy model [17], as it allows youth advisors to see their input being implemented in real time. Additionally, the project was presented at multiple scientific conferences and regularly reported to the funding organization. Preparations for these events, as well as feedback received, were always shared with the coresearchers to ensure transparency about how and where their contributions were being presented. Those who wished to be recognized were also listed as coauthors on conference poster presentations.

The influence of the coresearchers extended beyond app development to the usability testing process. When reflecting on their experience, they described it as "fun and easy to do." They emphasized the value of peer-led research, noting that having young people conduct the testing and recruit from their own friendship networks made participants more comfortable expressing their true opinions—both positive and negative. By contrast, they felt that if adult researchers had conducted the testing, participants might not have responded as openly. The coresearchers emphasized that their close age and shared experiences with the participants allowed them to relate more directly, fostering a sense of understanding and connection. As peers, they created a more relaxed and informal atmosphere, making it easier for participants to communicate openly and feel comfortable. These reflections align with existing literature, which suggests that having observers and users with similar backgrounds can reduce misinterpretation of data in usability testing [32]. Arguments have been made in favor of involving children and young people as coresearchers, as their firsthand experience of childhood provides an insider perspective and expertise that adults cannot fully replicate [16,40]. While adult researchers may be considered outsiders, young people are embedded in peer culture, allowing them to be closer to the subject of research [41]. In line with the coresearchers' reflections, peer-led data collection has been shown to generate nuanced and valuable insights that differ from those typically produced by adult researchers [41].

Methodological Considerations

This study took steps to ensure diversity among coresearchers, representing different genders, ages, and backgrounds. However, it cannot be assumed that they fully represent all young people who have experienced trauma. The recruitment of study participants by the coresearchers is largely a strength, as their diversity was reflected in the sample. One limitation of the study was the decision not to video record the usability testing sessions, which could have provided additional insights into participant behavior. However, combining a standardized instrument with interview questions after usability tests was a strength, as was the multistep analysis process, which involved 2 researchers when applicable.

Conclusions

This study demonstrates that young people with personal experiences of trauma can actively contribute as coresearchers in the development and usability testing of a trauma support app. A child rights framework can facilitate this process, ensuring alignment with Article 12 of the CRC. The young people's contributions to the app's design can be categorized into 3 key areas: mechanics, dynamics, and aesthetics. The design studio method proved to be an effective approach for

XSL•FO
RenderX

cocreation, allowing for both the generation and refinement of their ideas. No formal training was required for the young people to contribute as experts in the development process. However, relevant training was essential to equip them for independent data collection. In parallel with this project, work has been conducted on potential negative consequences of the app—referred to as dark logic—which has been reported elsewhere [42]. The next phase involves further testing the app in a larger population. Ongoing efforts are focused on piloting the app both with young people participating in a TRT group and as a stand-alone intervention.

Acknowledgments

The copyright of Teaching Recovery Techniques is held by the Children and War Foundation. The research team thanks the coresearchers who have been part of the project for all their great efforts. We also thank Antónia Tökés for assisting with the analysis and for the further work on the app. This research was funded by a grant from the Skandia Ideas for Life Foundation (principal investigator: GW).

Data Availability

Data generated for this study are not publicly available to protect the privacy of those involved. Requests to access the data should be directed to GW.

Disclaimer

No generative AI toold were used when writing this manuscript.

Authors' Contributions

MT led the workshops, performed the coresearcher training, analyzed usability findings, and wrote sections of the manuscript. KI and RA participated in the workshops, performed the usability testing, participated in analyzing the usability findings, and wrote sections of the manuscript. KE coled the workshops, built the app prototype, and wrote sections of the manuscript. GW supervised all project activities, discussed usability analysis findings, and wrote parts of the manuscript. All authors reviewed the manuscript.

Conflicts of Interest

None declared.

References

1. Radez J, Reardon T, Creswell C, Lawrence P, Evdoka-Burton G, Waite P. Why do children and adolescents (not) seek and access professional help for their mental health problems? A systematic review of quantitative and qualitative studies. Eur Child Adolesc Psychiatry 2021 Feb;30(2):183-211 [FREE Full text] [doi: 10.1007/s00787-019-01469-4] [Medline: 31965309]
2. Aguirre Velasco A, Cruz I, Billings J, Jimenez M, Rowe S. What are the barriers, facilitators and interventions targeting help-seeking behaviours for common mental health problems in adolescents? A systematic review. BMC Psychiatry 2020 Jun 11;20(1):293-292 [FREE Full text] [doi: 10.1186/s12888-020-02659-0] [Medline: 32527236]
3. Gulliver A, Griffiths K, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10(1):a [FREE Full text] [doi: 10.1186/1471-244x-10-113]
4. Kuhn E, Owen J. Advances in PTSD treatment delivery: the role of digital technology in PTSD treatment. Curr Treat Options Psych 2020 Mar 27;7(2):88-102 [FREE Full text] [doi: 10.1007/s40501-020-00207-x]
5. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
6. Sander L, Schorndanner J, Terhorst Y, Spanhel K, Pryss R, Baumeister H, et al. 'Help for trauma from the app stores?' A systematic review and standardised rating of apps for post-traumatic stress disorder (PTSD). Eur J Psychotraumatol 2020;11(1):1701788 [FREE Full text] [doi: 10.1080/20008198.2019.1701788] [Medline: 32002136]
7. Yule W, Dyregrov A, Raundalen M, Smith P. Children and war: the work of the Children and War Foundation. Eur J Psychotraumatol 2013;4:18424 [FREE Full text] [doi: 10.3402/ejpt.v4i0.18424] [Medline: 23330058]
8. Davis R, Meiser-Stedman R, Afzal N, Devaney J, Halligan S, Lofthouse K, et al. Systematic review and meta-analysis: group-based interventions for treating posttraumatic stress symptoms in children and adolescents. J Am Acad Child Adolesc Psychiatry 2023 Nov;62(11):1217-1232 [FREE Full text] [doi: 10.1016/j.jaac.2023.02.013] [Medline: 36948393]
9. Pérez-Aronsson A, Thell M, Lampa E, Löfving SG, Tökés A, Torakai N, et al. Adaptation of the trauma group intervention 'Teaching Recovery Techniques' for online delivery: a participatory design and usability study. Internet Interv 2022 Dec;30:100589 [FREE Full text] [doi: 10.1016/j.invent.2022.100589] [Medline: 36411786]

XSL•FO
RenderX

10. Freire K, Pope R, Jeffrey K, Andrews K, Nott M, Bowman T. Engaging with children and adolescents: a systematic review of participatory methods and approaches in research informing the development of health resources and interventions. Adolescent Res Rev 2022 Feb 03;7(3):335-354 [FREE Full text] [doi: 10.1007/s40894-022-00181-w]
11. Liabo K, Roberts H. Coproduction and coproducing research with children and their parents. Arch Dis Child 2019 Dec 08;104(12):1134-1137. [doi: 10.1136/archdischild-2018-316387] [Medline: 31068317]
12. Thabrew H, Fleming T, Hetrick S, Merry S. Co-design of eHealth interventions with children and young people. Front Psychiatry 2018;9:481 [FREE Full text] [doi: 10.3389/fpsyt.2018.00481] [Medline: 30405450]
13. Shimmin C, Wittmeier KDM, Lavoie JG, Wicklund ED, Sibley KM. Moving towards a more inclusive patient and public involvement in health research paradigm: the incorporation of a trauma-informed intersectional analysis. BMC Health Serv Res 2017 Aug 07;17(1):539 [FREE Full text] [doi: 10.1186/s12913-017-2463-1] [Medline: 28784138]
14. Fløtten KJØ, Guerreiro A, Simonelli I, Solevåg AL, Aujoulat I. Adolescent and young adult patients as co-researchers: a scoping review. Health Expect 2021 Aug;24(4):1044-1055 [FREE Full text] [doi: 10.1111/hex.13266] [Medline: 33991369]
15. Convention on the Rights of the Child. UNICEF. 1989. URL: https://www.unicef.org/child-rights-convention/convention-text [accessed 2024-02-01]
16. Lundy L, McEvoy L. Children's rights and research processes: assisting children to (in)formed views. In: Children's Rights. London, UK: Routledge; 2017:525-540.
17. Lundy L. 'Voice' is not enough: conceptualising Article 12 of the United Nations Convention on the Rights of the Child. British Educational Res J 2013 Jan 02;33(6):927-942 [FREE Full text] [doi: 10.1080/01411920701657033]
18. Participation framework. Hub na nÓg - young voices in decision-making. 2021. URL: https://hubnanog.ie/participation-framework/ [accessed 2024-02-01]
19. Utsatta områden - polisens arbete. Swedish Police Authority. 2023. URL: https://polisen.se/om-polisen/polisens-arbete/utsatta-omraden/ [accessed 2024-02-01]
20. Så får barn och unga arbeta. Swedish Work Environment Authority. 2015. URL: https://www.av.se/globalassets/filer/publikationer/broschyrer/sa-far-barn-och-ungdomar-arbeta-broschyr-adi043.pdf [accessed 2024-02-01]
21. INVOLVE. Involving children and young people in research: top tips and essential key issues for researchers. INVOLVE. Southampton, UK: INVOLVE; 2019. URL: https://www.invo.org.uk/wp-content/uploads/2016/01/involvingcyp-tips-keyissues-January2016.pdf [accessed 2024-02-01]
22. Kaplan K. Facilitating an effective design studio workshop. Nielsen Norman Group. 2017. URL: https://www.nngroup.com/articles/facilitating-design-studio-workshop/ [accessed 2024-02-01]
23. Padlet. URL: https://padlet.com [accessed 2024-02-01]
24. Pandasuite. URL: https://pandasuite.com [accessed 2024-02-01]
25. Kellett M. How to Develop Children as Researchers. New York, NY: Sage Publications; 2005.
26. Kim C, Sheehy K, Kerawalla L. Developing Children as Researchers: A Practical Guide to Help Children Conduct Social Research. London, UK: Taylor & Francis; 2017.
27. Shier H. Children as researchers in Nicaragua: children's consultancy to transformative research. Global Studies of Childhood 2015 May 25;5(2):206-219 [FREE Full text] [doi: 10.1177/2043610615587798]
28. Baauw E, Markopoulous P. A comparison of think-aloud and post-task interview for usability testing with children. In: IDC '04: Proceedings of the 2004 Conference on Interaction Design and Children: Building a Community. 2004 Presented at: IDC '04: The 2004 Conference on Interaction Design and Children: Building a Community; June 1-3, 2004; Baltimore, MD p. 115-116. [doi: 10.1145/1017833.1017848]
29. Virzi R. Refining the test phase of usability evaluation: how many subjects is enough? Hum Factors 1992 Aug 01;34(4):457-468 [FREE Full text] [doi: 10.1177/001872089203400407]
30. Eccles D, Arsal G. The think aloud method: what is it and how do I use it? Qualitative Research in Sport, Exercise and Health 2017 May 26;9(4):514-531. [doi: 10.1080/2159676x.2017.1331501]
31. Brooke J. SUS: a "quick and dirty'usability scale. Usability evaluation in industry 1996;189(3):1-7.
32. Hertzum M. Usability Testing: A Practitioner's Guide to Evaluating the User Experience. San Rafael, CA: Morgan and Claypool; 2020:1-15.
33. Mentimeter. 2023. URL: https://www.mentimeter.com [accessed 2024-02-01]
34. Motta RW. Secondary trauma. Int J Emerg Ment Health 2008;10(4):291-298. [Medline: 19278145]
35. Browning H. Guidelines for Designing Effective Games as Clinical Interventions: Mechanics, Dynamics, Aesthetics, and Outcomes (MDAO) Framework. New York, NY: IGI Global; 2016:105-130.
36. Bangor A, Kortum P, Miller J. An empirical evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594 [FREE Full text] [doi: 10.1080/10447310802205776]
37. Beresford P. Beyond the usual suspects: Research report. Shaping Our Lives. 2013. URL: https://shapingourlives.org.uk/wp-content/uploads/2021/08/BTUS-Report.pdf [accessed 2024-02-01]
38. Rouncefield-Swales A, Harris J, Carter B, Bray L, Bewley T, Martin R. Children and young people's contributions to public involvement and engagement activities in health-related research: a scoping review. PLoS One 2021;16(6):e0252774 [FREE Full text] [doi: 10.1371/journal.pone.0252774] [Medline: 34106978]

XSL•FO
RenderX

39. Sarkadi A, Thell M, Fängström K, Dahlberg A, Fäldt A, Pérez-Aronsson A, et al. Are we ready to really hear the voices of those concerned? Lessons learned from listening to and involving children in child and family psychology research. Clin Child Fam Psychol Rev 2023 Dec;26(4):994-1007 [FREE Full text] [doi: 10.1007/s10567-023-00453-4] [Medline: 37700107]
40. Bradbury-Jones C, Taylor J. Engaging with children as co-researchers: challenges, counter-challenges and solutions. International Journal of Social Research Methodology 2013 Dec 03;18(2):161-173 [FREE Full text] [doi: 10.1080/13645579.2013.864589]
41. Kellett M. Rethinking Children and Research: Attitudes in Contemporary Society. New York, NY: Bloomsbury Publishing; Apr 15, 2010.
42. Warner G, Edvardsson K, Thell M. Look before you leap: theorising harmful consequences ahead of a pilot trial, the case of app-based trauma support for youth. 2022 Presented at: The 13th EUSPR Conference; September 29 and 30, 2022; Tallinn, Estonia.

Abbreviations

CRC: the Convention on the Rights of the Child
MDA: Mechanics, Dynamics, and Aesthetics
PTSD: posttraumatic stress disorder
REDCap: Research Electronic Data Capture
SUS: System Usability Scale
TRT: Teaching Recovery Techniques

Edited by A Mavragani; submitted 26.02.24; peer-reviewed by C Tsai; comments to author 20.01.25; revised version received 24.01.25; accepted 13.02.25; published 18.03.25.

Please cite as:
Thell M, Edvardsson K, Aljeshy R, Ibrahim K, Warner G
A Trauma Support App for Young People: Co-design and Usability Study
JMIR Form Res 2025;9:e57789
URL: https://formative.jmir.org/2025/1/e57789
doi:10.2196/57789
PMID:

©Maria Thell, Kerstin Edvardsson, Reem Aljeshy, Kalid Ibrahim, Georgina Warner. Originally published in JMIR Formative Research (https://formative.jmir.org), 18.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Kinetic Pattern Recognition in Home-Based Knee Rehabilitation Using Machine Learning Clustering Methods on the Slider Digital Physiotherapy Device: Prospective Observational Study

Clement Twumasi[1], BSc, MPhil, PhD; Mikail Aktas[2*], MBBS, MRES; Nicholas Santoni[3*], BSc

[1]Nuffield Department of Medicine, Experimental Medicine Division, University of Oxford, Oxford, United Kingdom

[2]Department of Bioengineering, Imperial College London, South Kensington, London, United Kingdom

[3]Department of Biology, University of North Carolina, Greensboro, NC, United States

[*]these authors contributed equally

Corresponding Author:
Clement Twumasi, BSc, MPhil, PhD
Nuffield Department of Medicine
Experimental Medicine Division
University of Oxford
Peter Medawar Building for Pathogen Research
South Parks Road
Oxford, OX1 3SY
United Kingdom
Phone: 44 7411484454
Email: clement.twumasi@ndm.ox.ac.uk

Abstract

Background: Recent advancements in rehabilitation sciences have progressively used computational techniques to improve diagnostic and treatment approaches. However, the analysis of high-dimensional, time-dependent data continues to pose a significant problem. Prior research has used clustering techniques on rehabilitation data to identify movement patterns and forecast recovery outcomes. Nonetheless, these initiatives have not yet used force or motion datasets obtained outside a clinical setting, thereby limiting the capacity for therapeutic decisions. Biomechanical data analysis has demonstrated considerable potential in bridging these gaps and improving clinical decision-making in rehabilitation settings.

Objective: This study presents a comprehensive clustering analysis of multidimensional movement datasets captured using a novel home exercise device, the “Slider”. The aim is to identify clinically relevant movement patterns and provide answers to open research questions for the first time to inform personalized rehabilitation protocols, predict individual recovery trajectories, and assess the risks of potential postoperative complications.

Methods: High-dimensional, time-dependent, bilateral knee kinetic datasets were independently analyzed from 32 participants using four unsupervised clustering techniques: k-means, hierarchical clustering, partition around medoids, and CLARA (Clustering Large Applications). The data comprised force, laser-measured distance, and optical tracker coordinates from lower limb activities. The optimal clusters identified through the unsupervised clustering methods were further evaluated and compared using silhouette analysis to quantify their performance. Key determinants of cluster membership were assessed, including demographic factors (eg, gender, BMI, and age) and pain levels, by using a logistic regression model with analysis of covariance adjustment.

Results: Three distinct, time-varying movement patterns or clusters were identified for each knee. Hierarchical clustering performed best for the right knee datasets (with an average silhouette score of 0.637), while CLARA was the most effective for the left knee datasets (with an average silhouette score of 0.598). Key predictors of the movement cluster membership were discovered for both knees. BMI was the most influential determinant of cluster membership for the right knee, where higher BMI decreased the odds of cluster-2 membership (odds ratio [OR] 0.95, 95% CI 0.94-0.96; $P<.001$) but increased the odds for cluster-3 assignment relative to cluster 1 (OR 1.05, 95% CI 1.03-1.06; $P<.001$). For the left knee, all predictors of cluster-2 membership were significant ($.001\leq P\leq.008$), whereas only BMI ($P=.81$) could not predict the likelihood of an individual belonging to cluster 3 compared to cluster 1. Gender was the strongest determinant for the left knee, with male participants significantly likely to belong to cluster 3 (OR 3.52, 95% CI 2.91-4.27; $P<.001$).

XSL•FO
RenderX

Conclusions: These kinetic patterns offer significant insights for creating personalized rehabilitation procedures, potentially improving patient outcomes. These findings underscore the efficacy of unsupervised clustering techniques in the analysis of biomechanical data for clinical rehabilitation applications.

(JMIR Form Res 2025;9:e69150) doi:10.2196/69150

KEYWORDS

machine learning; cluster analysis; force measurement; knee replacement; musculoskeletal; physical therapy; Slider device; knee osteoarthritis; digital health; telerehabilitation

Introduction

Background of the Study

The challenge of optimizing knee rehabilitation exercises following arthroplasty or injury is critical, given the high prevalence of knee surgeries and the need for effective recovery strategies. According to the *American Joint Replacement Registry 2022 Annual Report*, over 2.5 million knee and hip arthroplasty cases were documented, constituting a significant proportion of knee replacements. The quality and effectiveness of the ensuing rehabilitation, which has historically relied on patient self-reporting, subjective assessments, and poor compliance, is crucial to the success of these procedures. Such methods can lead to suboptimal outcomes, as they fail to capture biomechanical data that could inform treatment planning and postoperative complications [1,2].

Recent developments in sensor technology have made it possible to collect detailed kinetic data during physical therapy exercises. This presents a chance to transition from subjective assessments to more objective, data-driven methods. This change is especially significant in home-based rehabilitation, where it is crucial to collect valuable data consistently. The newly developed Slider device [3] exemplifies this technological advancement by capturing multidimensional movement data, including force and spatial coordinates, directly from patients' lower limbs during exercise. The Slider device is the first rehabilitation device capable of remotely collecting force and motion data in a patient's home. This capability enhances the ability to monitor patients in real time and to produce more informative datasets that capture the dynamic interaction of force and movement in a real-world, home-based setting [4].

A recent study has extensively examined the efficacy of the Slider device for prehabilitation prior to total knee replacement surgery [3]. The study revealed that the device can assist patients in performing autonomous physiotherapy prior to surgery in nonclinical environments. The research involved 17 patients awaiting knee replacement surgery at a UK National Health Service hospital. Their findings confirmed the device's practicality and user-friendliness for patient-initiated exercises, with no recorded safety issues or adverse occurrences [3]. Principal findings revealed improvements in user involvement and physical preparedness. The study advised that subsequent research should involve a bigger and more diverse participant cohort to comprehensively validate its findings regarding the device. In order to enhance our understanding of this new device, additional studies may be required to investigate the multidimensional kinetic and kinematic data from the Slider using unsupervised clustering methods for the first time across a comparatively larger and diverse population. This exploration is necessary to identify potential movement patterns that could offer further insights into workout performance and recovery. The insights gained from these computational methods may improve comprehension of patient heterogeneity, enabling customized rehabilitation programs that more effectively address individual needs and improve outcomes for diverse groups.

The user datasets from this new data-collection device represent a significant contribution to the field, as they provide a new level of detail in understanding how kinetic patterns evolve during rehabilitation outside of a clinical environment. Despite these advancements, analyzing such high-dimensional, time-dependent data remains a complex challenge. Previous studies have applied clustering methods to other rehabilitation data to identify movement patterns and predict recovery outcomes. However, these efforts have yet to use force or motion data captured outside of a clinical environment, limiting their ability to inform clinical decisions [5,6].

Data were collected from healthy volunteers performing lying flexions and extensions of their knees, including relevant demographic characteristics and pain levels. K-means, hierarchical clustering, partition around medoids (PAM), and CLARA (Clustering Large Applications) were applied to identify distinct movement patterns (based on force and displacement measurements as primary outcomes). Performance evaluations were conducted using the classical silhouette analysis to determine the most appropriate clustering method. Then, post hoc analyses were performed with multinomial logistic regression to explore potential demographic predictors of cluster membership based on the best models for the right and left knee datasets, respectively. This study contributes to the understanding of rehabilitation dynamics beyond the clinical setting. The insights gained from the uniquely identified data clusters may therefore enable clinicians to make more informed, data-driven decisions, ultimately enhancing the precision and effectiveness of rehabilitation programs. The primary aim of this study is to advance the understanding of kinetic patterns during knee rehabilitation by applying and comparing four unsupervised clustering algorithms to new datasets collected for the first time using the Slider device.

Paper Structure and Research Questions

This study is structured into five main sections. The *Introduction* section provides an overview of the study's background, outlines the paper's organization, and highlights its primary contributions. The *Methods* section outlines the empirical kinetic data used in the study, participant recruitment, study design, and ethical considerations, followed by a concise summary of

XSL•FO
RenderX

the clustering methodologies and additional modeling considerations. The *Results* section presents the findings derived from the statistical analysis of the experimental data. Finally, the final section (*Discussion*) offers a discussion of the key findings, draws conclusions, and suggests directions for future research. With the help of experimental bilateral knee kinetic datasets obtained from the newly developed home-based rehabilitation device (Slider), this study seeks to address the following open research questions:

- What distinct movement patterns can be identified from multidimensional kinetic data during home physiotherapy?
- Can real-time clustering of movement data provide meaningful feedback for personalized physiotherapy?
- How effective is cluster analysis in distinguishing between varying levels of movement efficiency and recovery states?
- How do kinetic movement patterns evolve over the course of physiotherapy, as detected through the new Slider device?
- Can the multidimensional movement data collected remotely provide reliable insights for monitoring physiotherapy outcomes?

Methods

Data Collection, Recruitment, and Study Design

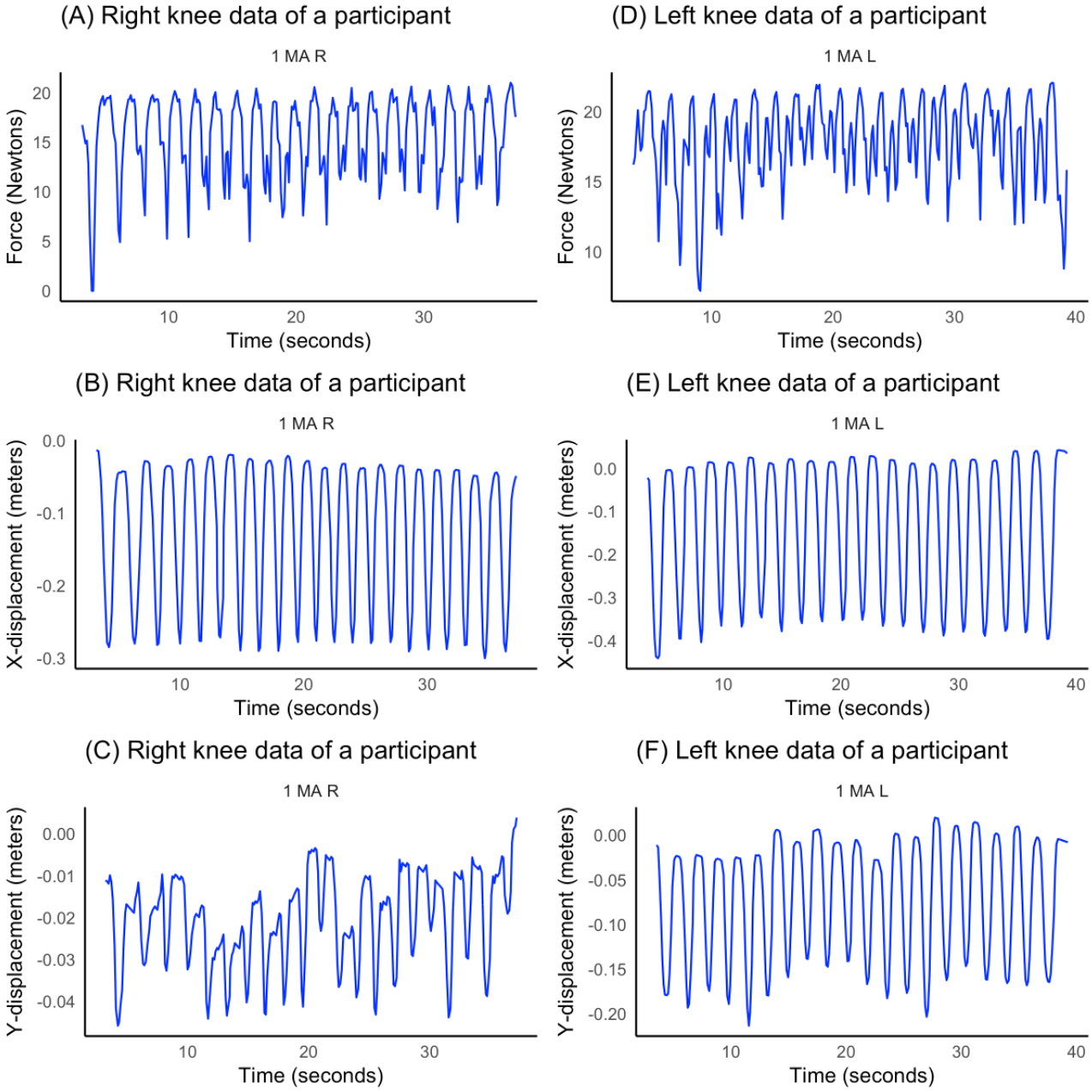
Continuously observed, time-dependent, and high-dimensional datasets were collected from both knees of 32 participants (totaling 64 time-varying datasets with large data points) using Slider during lying flexion and extension exercises. Participants were not required to have undergone any prior treatment. Demographic characteristics such as gender (21 male and 11 female participants), ethnicity (30 White participants and 2 mixed White and Asian participants), laterality (29 right and 3 left), age (mean 47.3, SD 2.11; range 21-77 years), height (cm), and weight (kg) were recorded. Each participant's BMI was calculated from their height and weight measurements (mean 26.2, SD 4.99; range 18.3-40.6 kg/m^2). Pain levels or scores were quantified using a standard analog pain scale, ranging from a score of 0=no pain to 10=worst imaginable pain.

Participants were not excluded from the study due to deformities or complications from prior surgeries or therapy treatments. The study was conducted without supervision, as it was intended to be performed remotely in the participants' homes due to the inherently mobile nature of the Slider device. Consequently, the study was deemed low risk and noninvasive. Each participant completed 20 cycles, at a self-selected speed with the device recording force (in Newtons) and X-Y displacement (in meters) measured with an optical tracker. The data output was time stamped (in seconds). Figure 1 is a time-series visualization of observed data of the main unadjusted outcome variables for clustering of a study participant. All statistical tests, with a significance level set at $P<.05$, and cluster analyses were performed in R statistical software (version 4.3.1; R Core Team) [7].

The participants comprised the entire existing pool of volunteers registered with the Slider device manufacturer. These individuals were invited to participate based on their prior engagement with the device, ensuring familiarity with its use. Recruitment procedures followed ethical guidelines, with all participants providing informed consent before data collection. This study follows a machine learning (ML) predictive modeling design in biomedical research, using unsupervised clustering techniques to analyze high-dimensional kinetic data and identify distinct movement patterns in home-based knee rehabilitation.

XSL•FO
RenderX

Figure 1. Time-series plots illustrating the unadjusted observed kinetic data, including force and X-Y displacements, recorded at the (A-C) right and (D-F) left knees of a study participant (ID: 1 MA). These plots provide an overview of the raw movement patterns captured by the Slider device, highlighting temporal variations in force exertion and spatial displacement during rehabilitation exercises.



Ethical Considerations

This low-risk and noninvasive study was classified as a service evaluation in accordance with the guidelines from the UK Health Research Authority. Consultation with the UK Research and Innovation Medical Research Council using Integrated Research Application System (348336) confirmed that the study adhered to all relevant ethical standards, and therefore, formal ethical committee approval was not required. All participants provided informed consent before participating, with explicit consent obtained for both data collection and subsequent analysis. To protect privacy and confidentiality, all study data were fully anonymized before analysis, ensuring no personally identifiable information was included. No financial compensation or other incentives were provided for participation, as the study was conducted on a voluntary basis. Finally, no identifiable images or personal data from participants are included in the study dataset or any supplementary materials.

Data Preprocessing

Data preprocessing is a crucial step in medical data analysis, especially for high-dimensional datasets with missing or incomplete information. It ensures data integrity, accuracy, and usability by handling missing values (where applicable), normalizing distributions, and reducing dimensionality, ultimately improving predictive models and clustering techniques performance [8]. Consequently, a two-stage data preprocessing method was used on the time-dependent datasets before conducting cluster analysis. In the first stage, potential outliers in the time-series data for each participant were automatically detected and corrected using the *tsoutliers* R package [9], based on the force and X-Y displacement data. In

XSL•FO
RenderX

the second stage, the cleaned time-series data underwent decomposition using locally estimated scatterplot smoothing (decomposition). This process aided in detrending the series for each participant, making them more stationary, and removing cyclical patterns to mitigate cyclical effects, thereby enhancing the accuracy of the clustering analysis results. Because outlier correction and locally estimated scatterplot smoothing decomposition effectively standardize the time-series data, decreasing noise and maintaining comparability between participants by addressing both extreme values and nonstationary trends, this two-stage approach is sufficient without additional data normalization. Reliable clustering outcomes are therefore possible since any residual variation in the data represents significant variations in participant movement patterns rather than artifacts from data scale or trend discrepancies.

Unsupervised Clustering Methods

Overview

In this study, the conventional hierarchical clustering and three widely used partitional clustering algorithms, namely: k-means [10,11], PAM [10,12], and CLARA [10], were used and compared to respectively divide the preprocessed bilateral knee datasets into an optimal number of clusters (represented by $k>1$). This partitioning ensures that data points representing kinetic features within the same cluster exhibit greater similarity to each other than to those in other clusters. These clustering methods have their respective algorithmic processes for distinctly grouping data, each with advantages and disadvantages regarding its robustness, scalability, and applicability related to the data structure, dataset size, and noise sensitivity. These justify the choice of models for this study and highlight the need to identify the best clustering algorithms for the right and left knee kinetic datasets, which were continuously observed over time for each volunteer. A summary of the underlying clustering methods used in this study is presented below.

K-Means Clustering

K-means is a centroid-based clustering algorithm that partitions a set of n data points $\{x_1, x_2, ..., x_n\} \in R^d$ into k clusters $\{C_1, C_2, ..., C_k\}$ by minimizing the within-cluster sum of squares objective function [10]:

where μ_i is the mean of the points in cluster C_i. The algorithm starts with k initial centroids, assigns each point to the nearest centroid or cluster center, and updates centroids by computing the mean of assigned points. This process repeats until convergence, typically when cluster assignments no longer change. K-means is computationally efficient, with a time complexity of $O(n \cdot k \cdot d \cdot t)$, where t is the number of iterations; but is sensitive to the initial choice of centroids.

Hierarchical Clustering

Hierarchical clustering constructs a nested tree of clusters by either agglomerative (bottom-up) or divisive (top-down) methods. The agglomerative approach starts with each data point as a single cluster and iteratively merges the two closest clusters based on a chosen linkage criterion. For example, in single linkage, the distance between clusters A and B is defined as [13]:

The agglomerative algorithm first computes the distance matrix for all pairs of points, then merges the closest clusters based on the chosen linkage criterion, and finally, updates the distance matrix, repeating this process until all points form a single cluster. The divisive algorithm starts with all data points in a single cluster and recursively splits them into smaller clusters until each data point forms its own cluster. At each step, it maximizes a splitting criterion such as the between-cluster variance denoted by:

where A and B are the resulting clusters from the split, and their respective centroids are denoted by μ_A and μ_B. Hierarchical clustering is effective for detecting nested structures, but it has a time complexity of either $O(n^2\log_n)$ for the agglomerative method (making it effective for small to medium-sized datasets) or $O(2^n)$ for the divisive approach (making it less effective for large datasets).

PAM

PAM is a medoid-based clustering algorithm that extends k-means by selecting actual data points (medoids) as cluster centers, minimizing the sum of dissimilarities between points and their nearest medoid. The objective is to find a set of k medoids $\{m_1, m_2, ..., m_k\}$ that minimizes [10]:

where $d(x, m_i)$ is the dissimilarity (eg, Euclidean distance) between a data point x and the medoid m_i of cluster C_i. PAM is robust to noise and outliers but has a higher computational cost of $O(k[n-k]^2)$, which can limit its scalability for large datasets. Different variants and extensions of PAM also exist [10].

CLARA

CLARA is an extension of PAM designed for large datasets. It draws multiple samples of the data and applies PAM to each sample, selecting the set of medoids that minimizes the clustering cost over the entire dataset [12]. CLARA reduces the computational burden of PAM by using random sampling, maintaining a complexity of $O(k|S|^2 + k[n-|S|])$, where $|S|$ is the sample size. CLARA scales better to large datasets, but its performance depends on the quality and representativeness of the random samples [10].

Other Statistical Modeling Considerations and Performance Evaluation

To respectively determine the optimal number of unique clusters in the data for both knees, the *NbClust* R Package [14] was used. This package evaluates over 30 different cluster identification schemes or indices, with the optimal cluster size determined automatically using a majority rule. After clustering, the performance of the four ML approaches for the right and left

XSL•FO
RenderX

knee datasets was evaluated and compared using silhouette analyses, which validate clustering methods [15]. Silhouette coefficients or scores, which range from –1 (indicating incorrect clustering) to +1 (indicating highly dense or correct clustering), were calculated for each cluster identified across the four data partitioning methods. Based on their distributions, appropriate pooled estimates of these scores were used to compare the models' performance for the right and left knee data clustering, respectively. Principal component analysis (PCA) was adopted to visualize the clustered high-dimensional data $\{x_1, x_2, ..., x_n\} \in R^d$ in a reduced dimensional space.

While the primary cluster analysis using force and displacement outcome data can potentially categorize movement patterns into distinct groups in the right and left knees, it remains essential to identify additional demographic factors that may influence these patterns and their clinical relevance [16]. Thus, a multinomial logistic regression model with analysis of covariance adjustment was used to investigate whether demographic variables such as age, gender, BMI, and pain scores could serve as significant predictors of cluster membership. This approach allows for a more comprehensive understanding of the determinants of cluster assignment by examining how these demographic factors, when adjusted for kinetic measurements, contribute to the likelihood of a participant belonging to a specific cluster at a particular time (since cluster membership is time-varying given the time-dependent kinetic data structure). Incorporating these variables enhances the ability to tailor personalized rehabilitation protocols by considering both biomechanical and individual patient characteristics, ultimately supporting more precise and effective clinical decision-making.

Using an analysis of covariance–based logistic model with adjustment for the actual observed force and X-Y displacement measurements, as well as temporal effects, the model is mathematically defined as follows:

for each cluster category $1<k\leq K$ relative to a baseline category (cluster 1), where K denotes the total number of uniquely identified clusters. $P(C_i=k)$ represents the probability of the ith individual being in cluster k at time $T\geq 0$, and the regression coefficients: β_{1k}, β_{2k}, β_{3k}, and β_{4k} represent the respective effects of the main demographic predictors (age, gender, BMI, and pain score) on the log-odds of being in cluster $k\geq 2$ versus the baseline cluster 1, adjusted for force (F), X-Y displacements (denoted by X and Y) and time of measurement (T).

The coefficients γ_{1k}, γ_{2k}, γ_{3k}, and γ_{4k} control for the influence of force, X-Y displacements, and time of measurement, respectively. The model was fitted with the help of the *nnet* R package [17]. Finally, the odds ratios (ORs) were estimated, which quantify the magnitude of the discriminatory effect of each of the main demographic predictors by exponentiating the estimated log-odds regression coefficients from $\exp(\beta_{gk})$ for $1\leq g\leq 4$ and $1<k\leq K$. The right and left knee datasets were respectively split into 80% training sets (for model fitting) and 20% testing sets for cross-validation to estimate the prediction accuracy from the resulting confusion matrix and the area under the receiver operating characteristic curve statistic.

Results

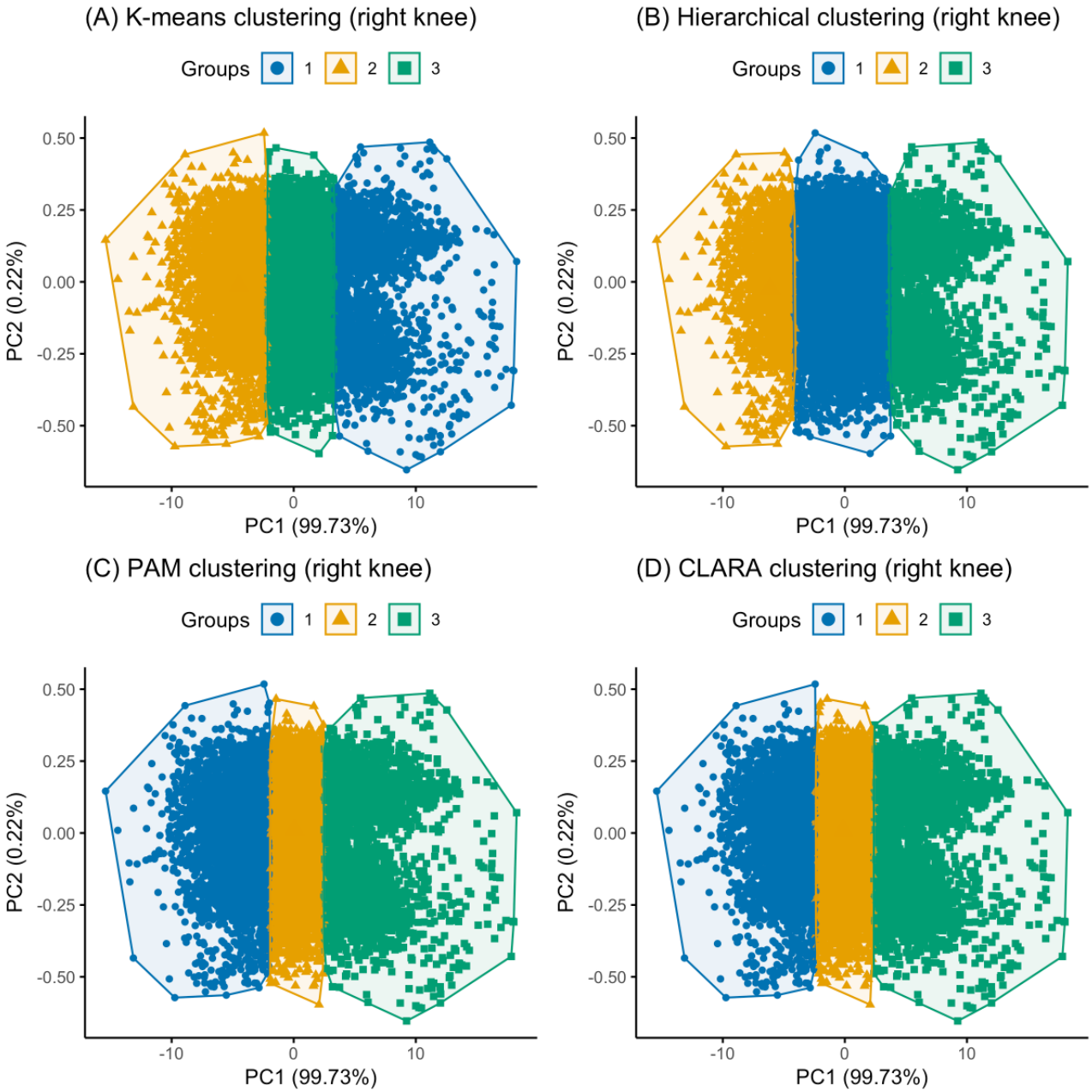
Cluster Analysis at Both Knees

Summary of Key Findings

Three clusters were optimally identified for both knees, each representing a unique kinetic pattern. K-means, hierarchical, PAM, and CLARA clustering were subsequently performed based on these predetermined clusters. Figures 2 and 3 show a PCA visualization of the clustered datasets in a lower-dimensional space for both knees. The silhouette analysis (Figure 4) revealed that hierarchical clustering was relatively found to be the most effective for the right knee kinetic datasets (with an average silhouette score of 0.637) followed by k-means (with an average silhouette score of 0.612), while CLARA proved to be the best-performing method for the left knee datasets (with an average silhouette score of 0.598) followed by k-means (with an average silhouette score of 0.588). It can be inferred from these results that the right knee datasets resulted in a better clustering result than the left knee datasets. Figure 5 presents a scatterplot of the adjusted kinetic features (which have been detrended and cycle-adjusted) in a 3D space for the right and left knees, stratified by the identified clusters, based on the best-performing, knee-specific clustering algorithms.

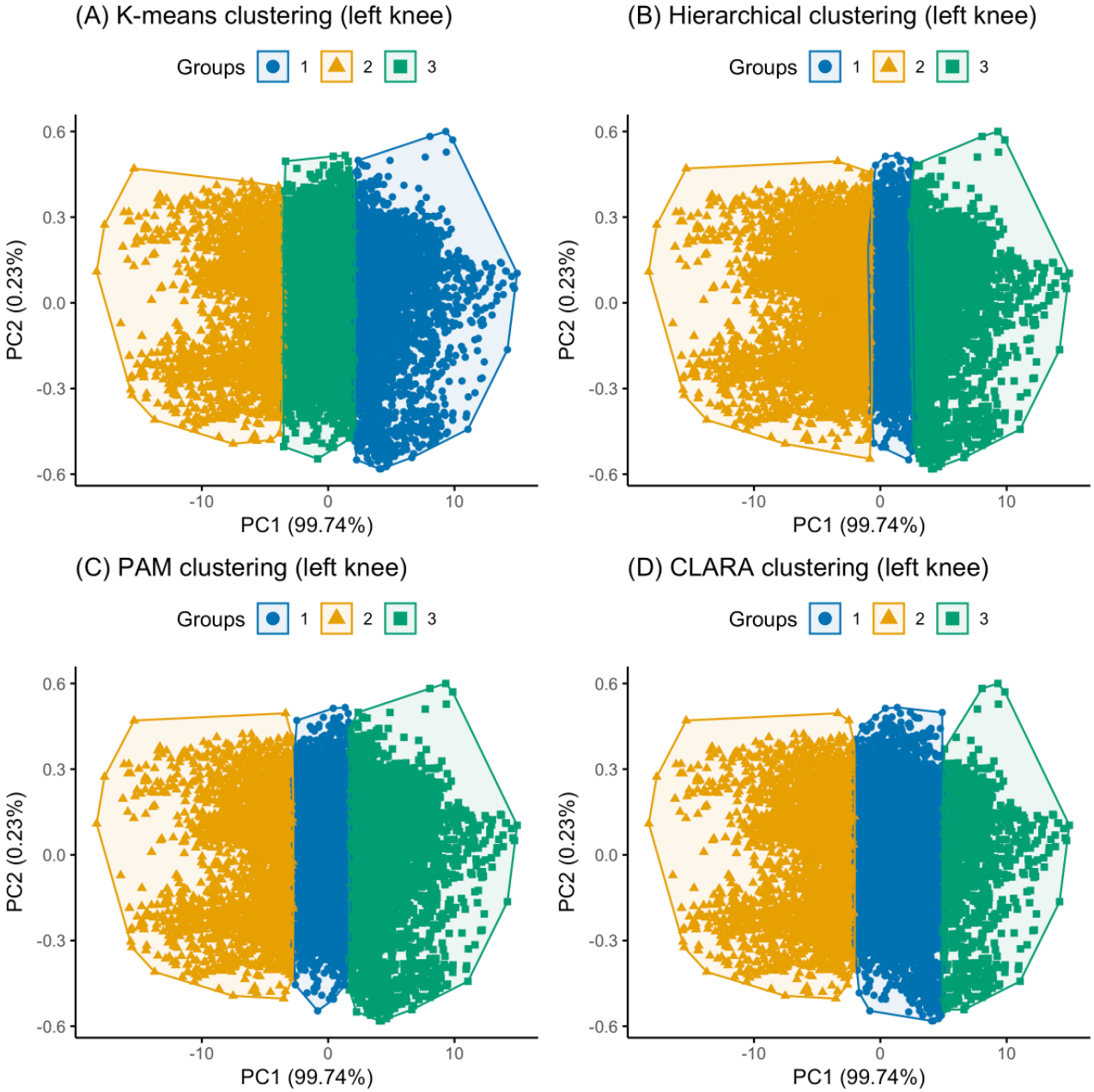
XSL•FO
RenderX

Figure 2. Principal component analysis visualizations comparing the three identified movement clusters at the right knee across the four clustering methods: (A) k-means, (B) hierarchical, (C) PAM, and (D) CLARA. These visualizations are based on adjusted kinetic features (force and X-Y displacements) and illustrate how different clustering techniques distinguish movement patterns in the multidimensional dataset. CLARA: Clustering Large Applications; PAM: partition around medoids.



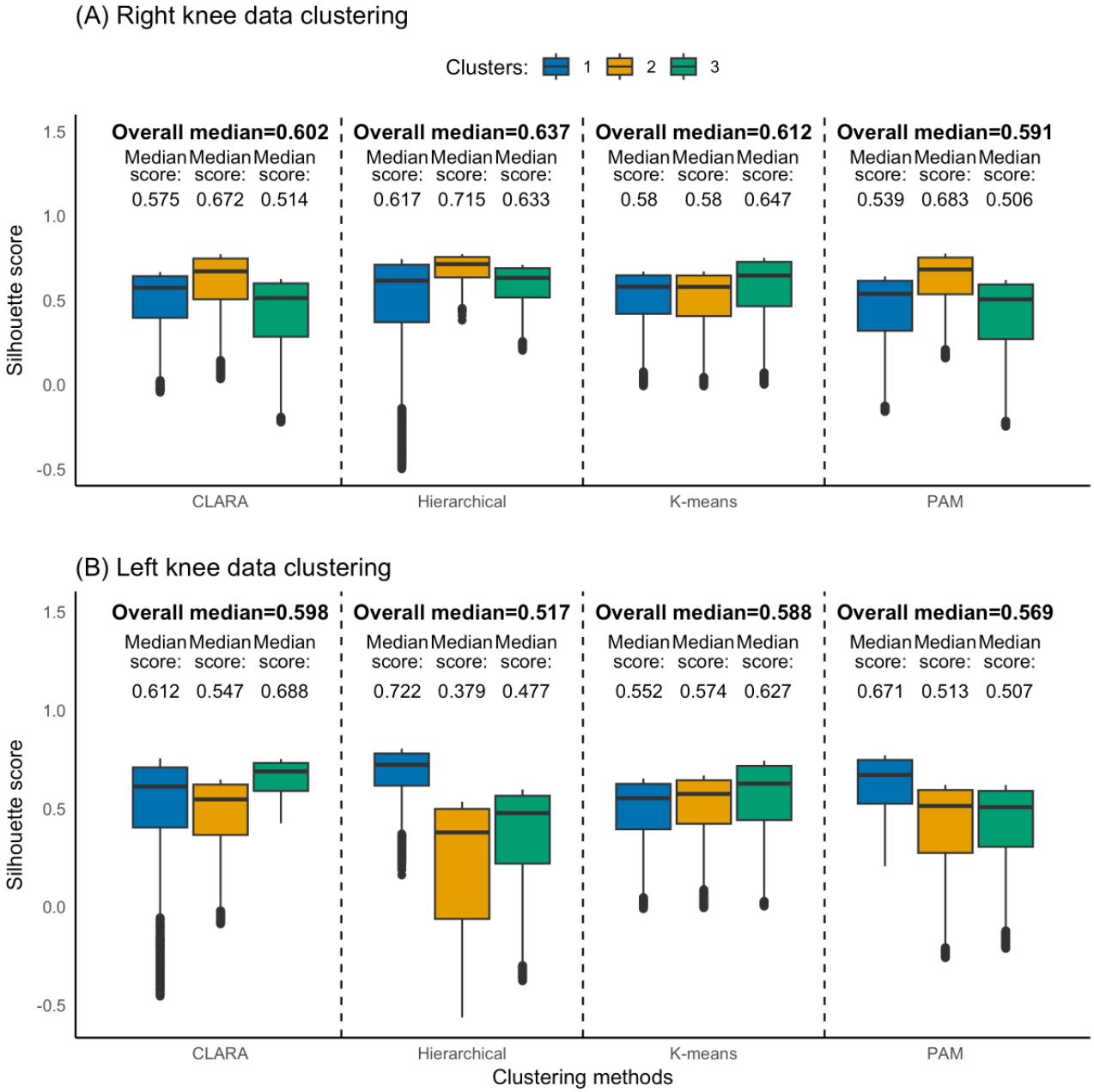
XSL•FO
RenderX

Figure 3. Principal component analysis visualizations comparing the three identified movement clusters at the left knee across the four clustering methods: (A) k-means, (B) hierarchical, (C) PAM, and (D) CLARA. These plots provide insights into the separability of kinetic movement patterns based on adjusted force and spatial displacement data. CLARA: Clustering Large Applications; PAM: partition around medoids.



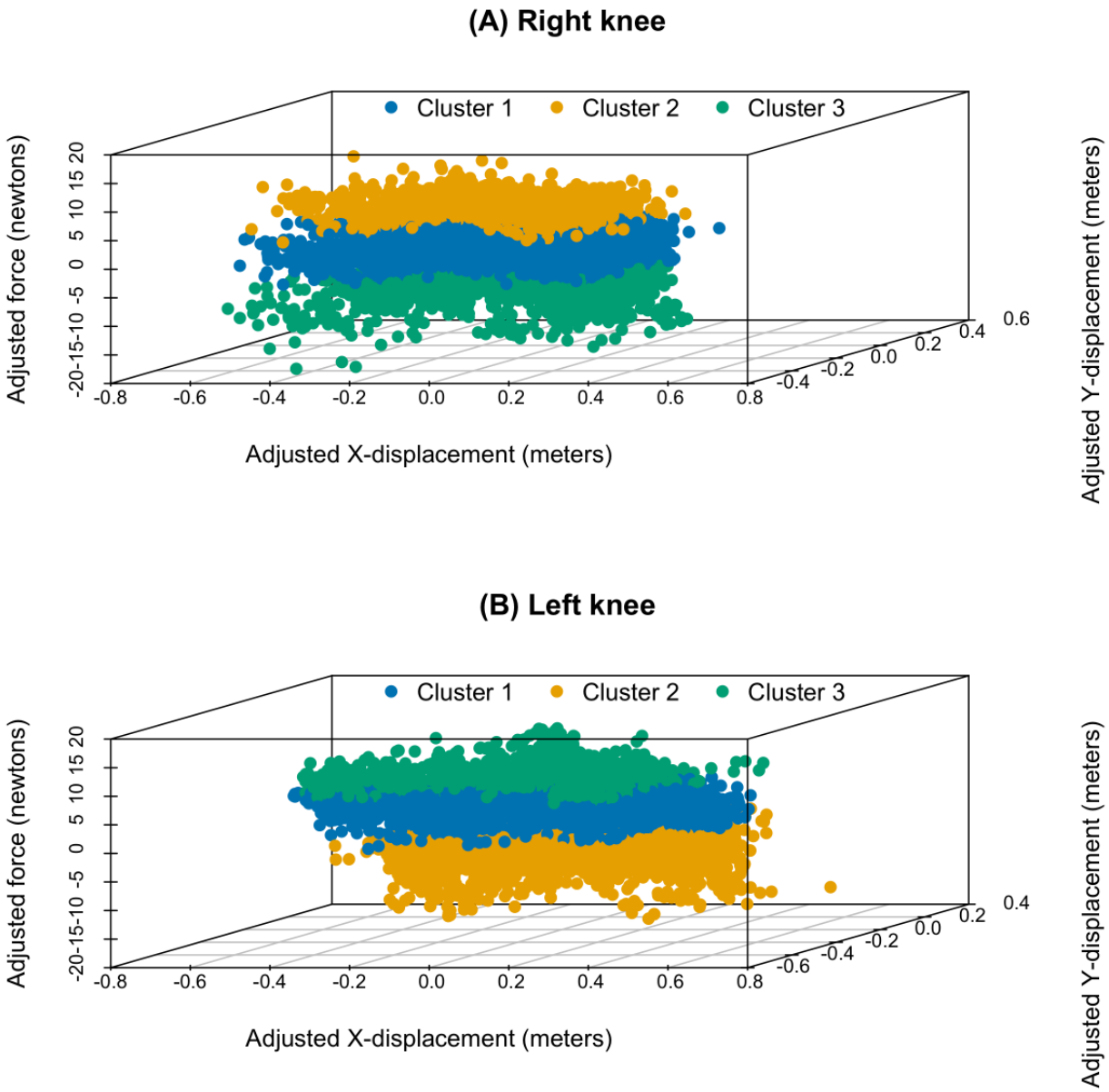
XSL•FO
RenderX

Figure 4. Distributional plot comparing estimated silhouette coefficients across the four clustering methods (k-means, hierarchical, PAM, and CLARA) at the (A) right and (B) left knees. The plot presents both cluster-specific and pooled estimates, highlighting the relative performance of each clustering method in distinguishing movement patterns from kinetic data at both knees. CLARA: Clustering Large Applications; PAM: partition around medoids.



XSL•FO
RenderX

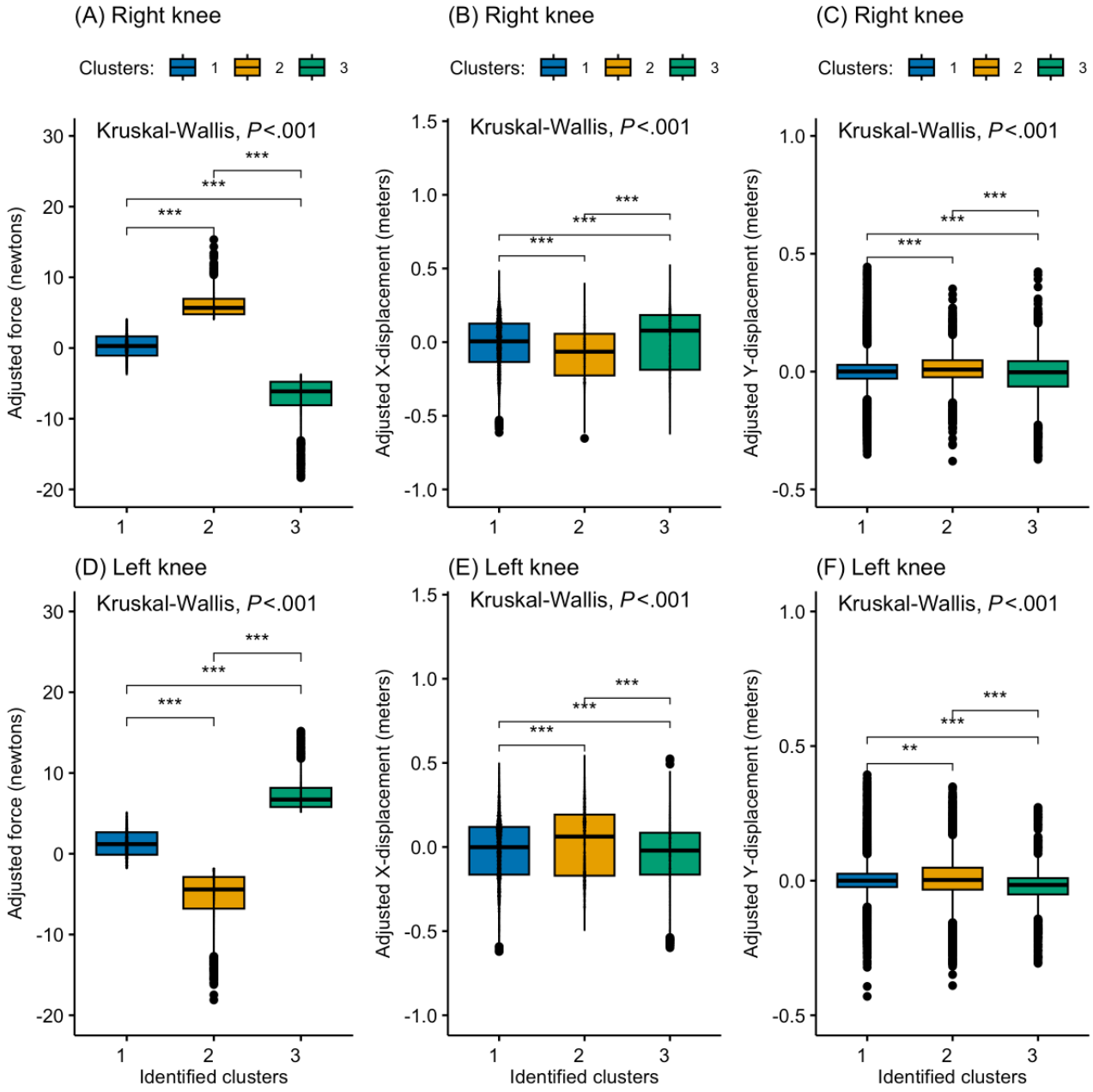
Figure 5. 3D scatterplot depicting the clustering of adjusted force measurements (in Newtons) against adjusted X-Y displacements (in meters) for the (A) right and (B) left knees. The clustering is based on the most optimal knee-specific methods identified through silhouette analysis—hierarchical clustering for the right knee and CLARA for the left knee—demonstrating the spatial distribution of movement patterns within the dataset. CLARA: Clustering Large Applications.



To assess whether significant differences existed in the distribution of multivariate kinetic features (force and X-Y displacements) across the three clusters for both knees, a multivariate Kruskal-Wallis (MKW) test [18] was used. This test was chosen due to the multivariate nonnormality of the bilateral knee datasets (Henze-Zirkler multivariate normality test: E=246.25, P<.001 for the right-knee datasets; E=219.34, P<.001 for left-knee datasets). The results indicated that at least one of the clusters significantly differed in distribution for both the right (MKW: W_6=9204.35; P<.001) and left knees (MKW: W_6=9973.54; P<.001). Subsequent univariate Kruskal-Wallis tests were performed for each kinetic feature at both knees, coupled with Dunn multiple comparison tests (Figure 6). There were significant differences in the distribution of the kinetic features between the three clusters at both knees, suggesting the identified knee-specific clusters were uniquely distinct. Figure 7 shows the frequency of times cluster membership varied temporally across the 32 study participants. The proportion of times participants were assigned to cluster 1, cluster 2, and cluster 3 at the right knee was 72.7%, 12.3%, and 15%, respectively. For the left knee, the proportion of times participants were assigned to cluster 1, cluster 2, and cluster 3 were 65.1%, 25.4%, and 9.5%, respectively.

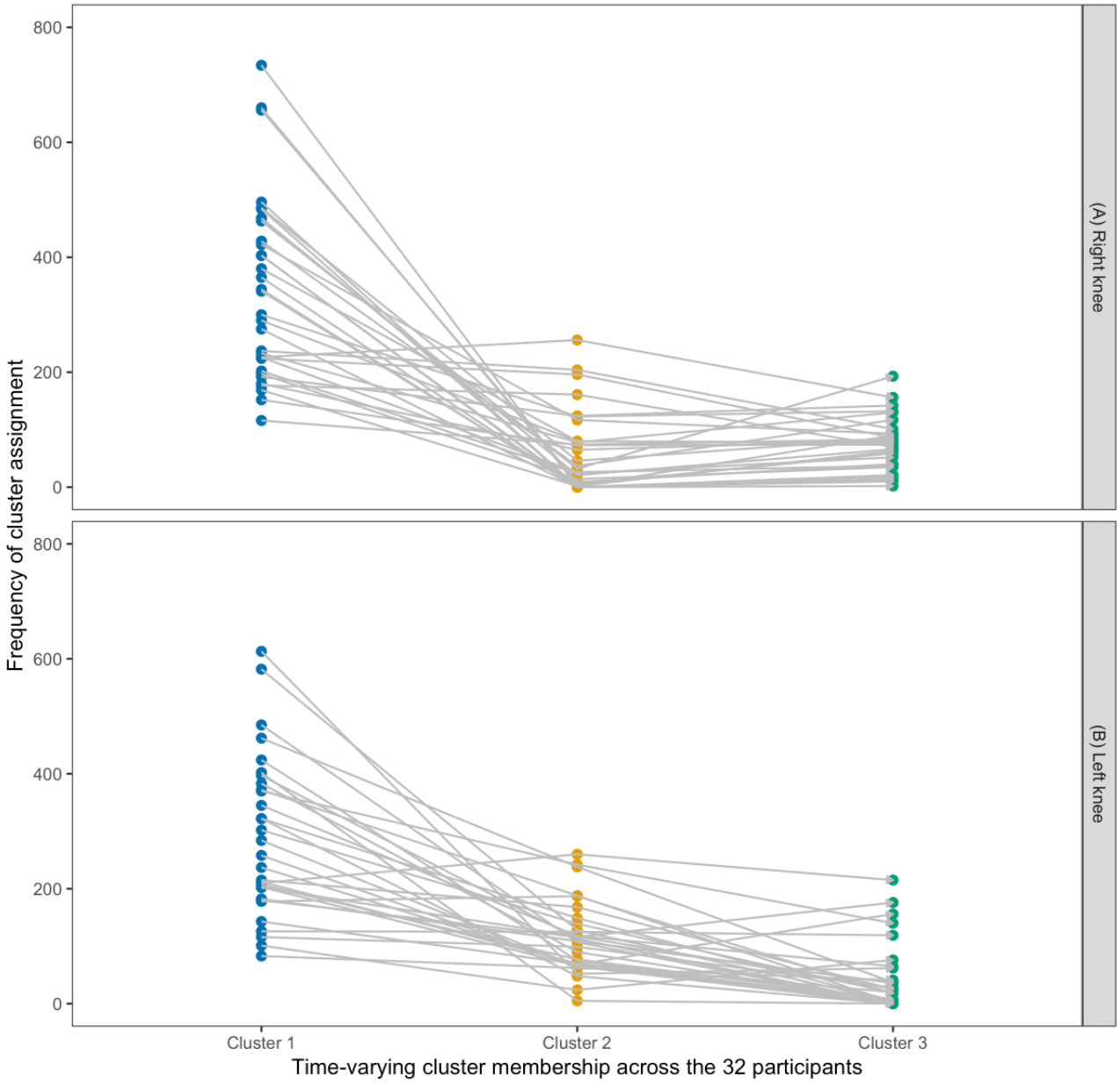
XSL•FO
RenderX

Figure 6. Comparative distribution plots of adjusted kinetic features (force and X-Y displacements) across the three k-means–identified clusters at the (A-C) right and (D-F) left knees. Statistical differences between clusters were assessed using Kruskal-Wallis tests followed by Dunn multiple comparison tests, with significance levels indicated (***P<.001; **P<.01; *P<.05). These results highlight variations in movement characteristics among identified rehabilitation patterns.



XSL•FO
RenderX

Figure 7. Slope graph visualizing temporal changes in cluster membership across the 32 study participants at the (A) right and (B) left knees. This plot captures how individuals transitioned between movement clusters over time due to the time-dependent structure of kinetic data collected from both knees, reflecting the dynamic nature of rehabilitation progress.



There was a reverse observation of the relative positions of clusters 2 and 3 in the right and left knees of the study participants (Figure 5). Possible practical kinetic interpretations of the clustered data with respect to force measurements across X-Y displacements, given this reverse kinetic pattern in right and left knees, could be understood as follows.

Cluster 1 (Middle Position)

This cluster suggests intermediate force levels and displacements. In both knees, movements in this cluster might involve moderate force exertion and a moderate range of motion in the X-Y plane. These could correspond to exercises aimed at strengthening muscles and improving joint mobility without excessive stress.

Cluster 2 (Bottom Position—Left Knee; Top Position—Right Knee)

This cluster likely represents movements with lower force exertion and smaller displacements in both X and Y directions. In the left knee, these movements may represent exercises performed with less force and a limited range of motion. It could represent muscle weakening or joint pathology. In the right knee, however, these movements may involve higher force exertion but still within a constrained range of motion. The lower range of motion could represent joint stiffness due to arthritis. This cluster may represent the least physically fit group of individuals. In the right knee, it could represent moderate-intensity exercises, while in the left knee, it might involve active movements with moderate intensity.

XSL•FO
RenderX

Cluster 3 (Top Position—Left Knee; Bottom Position—Right Knee)

This cluster indicates higher force levels and larger displacements. In the left knee, movements in this cluster likely involve high-force exertion combined with extensive movements in the X-Y plane. These could correspond to activities requiring significant strength and flexibility, such as high-intensity exercises. The right knee, however, shows lower force exertion but still with extensive movements, suggesting exercises focused more on flexibility, agility, and possibly lower muscle activities. This cluster may represent the most physically fit group of individuals.

Determinants of the Identified Clusters at Both Knees

The multinomial logistic regression analysis identified several significant demographic predictors of cluster membership for both right and left knee datasets (Table 1). For the right knee, being male was associated with significantly lower odds of belonging to cluster 2 compared to cluster 1 (OR 0.30, 95% CI 0.26-0.33) and to cluster 3 (OR 0.55, 95% CI 0.47-0.63). BMI also emerged as a significant predictor, with higher BMI decreasing the odds of being in cluster 2 (OR 0.95, 95% CI 0.94-0.96) but increasing the odds for cluster-3 membership (OR 1.05, 95% CI 1.03-1.06). Age was a significant predictor for cluster 3 (OR 1.02, 95% CI 1.01-1.03) but not for cluster 2 (P=.95). For the left knee, we found that all the demographic determinants of belonging to cluster 2 relative to cluster 1 were significant (.001≤P≤.008), whereas only BMI (P=.81) could not predict the likelihood of an individual belonging to cluster 3 compared to cluster 1. Gender emerged as the strongest determinant for the left knee data, with male participants being significantly more likely to belong to cluster 3 (OR 3.52, 95% CI 2.91-4.27; P<.001).

XSL•FO
RenderX

Table 1. Estimated OR[a] and their 95% CIs for the demographic determinants of cluster membership, derived from a fitted multinomial regression model (N=32 participants)[b,c].

Clusters and variables	OR (95% CI)	SE	*P* value
Right knee[d]			
Cluster 2 (reference: cluster 1)			
Male (reference: female)	0.30 (0.26-0.33)	0.061	<.001
BMI (kg/m^2)	0.95 (0.94-0.96)	0.006	<.001
Age	0.99 (0.98-1.01)	0.002	.95
Pain score	1.03 (0.99-1.06)	0.015	.05
Cluster 3 (reference: cluster 1)			
Male (reference: female)	0.55 (0.47-0.63)	0.077	<.001
BMI (kg/m^2)	1.05 (1.03-1.06)	0.007	<.001
Age	1.02 (1.01-1.03)	0.002	<.001
Pain score	1.03 (0.99-1.06)	0.016	.11
Left knee[e]			
Cluster 2 (reference: cluster 1)			
Male (reference: female)	0.63 (0.55-0.72)	0.069	<.001
BMI (kg/m^2)	1.10 (1.09-1.12)	0.007	<.001
Age	1.04 (1.03-1.05)	0.002	<.001
Pain score	0.95 (0.91-0.99)	0.020	.008
Cluster 3 (reference: cluster 1)			
Male (reference: female)	3.52 (2.91-4.27)	0.100	<.001
BMI (kg/m^2)	1.01 (0.98-1.02)	0.009	.81
Age	0.98 (0.97-0.99)	0.002	<.001
Pain score	1.05 (1.01-1.09)	0.020	.01

[a]OR: odds ratio.

[b]These results quantify the influence of demographic factors such as BMI, age, gender, and pain levels on the likelihood of belonging to specific movement clusters.

[c]Each of the multinomial logistic regression models for both knees' data was adjusted for the primary kinetic features (force and X-Y displacements) and temporal effect.

[d]For right knee measurement data: percentage accuracy=77.82% (95% CI 76.29%-79.3%); area under the receiver operating characteristic curve=0.778.

[e]For the left knee measurement data: percentage accuracy=76.24% (95% CI 74.62%-77.81%); area under the receiver operating characteristic curve=0.762.

Discussion

Results Key Insights Enabled by the Slider Device for Home-Based Rehabilitation Monitoring

This study explored further an innovative method for knee rehabilitation using the Slider device, a novel tool that enables the remote acquisition of multidimensional kinetic data, including force and spatial coordinates, within a home setting. By facilitating objective real-time monitoring of rehabilitation progress, Slider represents a significant advancement in personalized physiotherapy. A previous study has already validated its feasibility and safety for empowering patients to perform presurgery physiotherapy exercises independently, eliminating the need for a clinical environment [3]. However, no study has yet examined and analyzed the multidimensional bilateral kinetic and kinematic datasets generated by users of this new device. This analysis aimed to uncover potential patterns that could transform raw data into actionable insights, enabling therapists to deliver data-driven, patient-specific interventions that may optimize rehabilitation outcomes. This study used and compared four unsupervised clustering methods to find three unique time-varying movement patterns or clusters based on the observed time-dependent, multidimensional datasets from the right and left knees of the study participants.

The application of multiple clustering methods provides a comparative perspective on clustering performance, reinforcing the robustness of the findings. We have identified these movement patterns outside of a clinical setting for the first time. Significant differences in the distributions of force and displacement between clusters show how differently participants move. Demographic factors like gender, BMI, age, and pain

XSL•FO
RenderX

score also predicted membership in a cluster. These findings provide significant potential for individualized rehabilitation programs that enhance therapist accuracy and results. By leveraging ML and kinetic data collection, this study further demonstrates how at-home rehabilitation can be made more autonomous, reducing patient dependency on in-clinic physiotherapy sessions. The ability to remotely track movement progress offers a scalable solution for long-term rehabilitation, minimizing the need for frequent clinical visits.

Consequently, this study has effectively addressed the five open research questions by examining multidimensional kinetic data obtained during home-based physiotherapy. Using four clustering algorithms, the study found three different, time-varying, kinetic movement patterns for both knees. These patterns may represent different levels of fitness and recovery, answering the first question about how to find movement patterns in multidimensional kinetic data. The study answered the second question by showing that real-time clustering of movement data could help physiotherapists better understand how to meet the needs of each patient by showing how the clusters changed in response to force and displacement data. To answer the third question, the study confirmed that clustering methods, especially hierarchical clustering for the right knee and CLARA for the left knee, could tell the difference between different levels of movement efficiency and recovery states.

The mathematical modeling revealed that over time, participants progressed from erratic to more coordinated movement patterns, reflecting the evolution of the natural rhythm of motion during exercise, thus answering the fourth question on how movement patterns evolve. The insights gained from movement clustering suggest potential future advancements in physiotherapy technology, particularly in enhancing real-time feedback mechanisms of rehabilitation devices. These findings could contribute to the development of more adaptive and intelligent physiotherapy tools, improving their application in both research and clinical settings. Finally, the Slider device provided reliable, multidimensional kinetic data, enabling accurate cluster identification and analysis, thereby confirming its potential for remotely monitoring physiotherapy outcomes and answering the fifth question.

Practical Implications of the Study

Identification of Movement Patterns

Clusters that differentiate based on movement patterns (such as the range and direction of motion) can help clinicians identify abnormal movement behaviors in participants. This is particularly useful in postoperative rehabilitation or in monitoring the progress of participants with musculoskeletal disorders. For a theoretical application in a clinical setting, cluster characterized by limited range and lower force output might indicate participants who are either in the early stages of recovery from injury or surgery or who are experiencing complications or poor recovery. Hitherto, such kinetic evaluations have only been possible if participants attended sessions in a gait laboratory or clinic [5]. Slider allows these evaluations to be done in a participant's home using data from exercises done by the participants over the entire duration of their rehabilitation program. A possible weakness of this study would be that it is intended to account for movement of the knee but does not account for limitations in movement due to the flexibility of the ankle or hip. Such variables could have impacted the data collected and changed which cluster the participant fell into, however, no participants that participated reported having such issues before conducting the exercises, so such variables should not have been considered significant during data collection.

Tailored Rehabilitation Programs

Understanding which cluster a participant's movement data fall into allows clinicians to tailor rehabilitation programs more effectively. By customizing programs based on the typical movement characteristics of each cluster, clinicians can optimize recovery times and outcomes. By understanding which cluster a participant's movement data fall into, therapists can customize exercises that target specific movement deficiencies [6]. For a practical application, participants in a cluster showing high force and extreme movement ranges might benefit from a different set of exercises compared to those in a cluster with conservative movement patterns, possibly due to differences in pain tolerance, injury severity, or recovery stage. Remote data collection can enable clinicians to monitor progress in real time, offering greater flexibility in treatment planning and providing objective data for faster, more efficient feedback. This benefits both participants and care providers.

Prediction of Recovery Trajectories

The data clusters could help predict different recovery trajectories by identifying which participants will likely recover faster based on similarities in movement patterns. Participants showing gradual movement range and force improvements over time by progressing through clusters may indicate a positive recovery trajectory. Clusters may serve as predictive markers for rehabilitation outcomes, allowing for proactive adjustments in treatment plans [19].

Customization of Assistive Devices

Insights from clustering can inform the customization of assistive devices to better support individual participant needs based on the commonalities in movement patterns, unique movement limitations, or capabilities observed within each cluster. Participants in a cluster that shows significantly altered gait patterns might benefit from specific types of orthotic supports designed to compensate for those specific deviations [20].

Enhanced Understanding of Pathophysiology

Exploring cluster formation further can improve understanding of the pathophysiology of conditions, aiding diagnosis and treatment. Differentiating clusters by movement patterns can help identify causes of pain or dysfunction, such as muscular versus joint issues, and analyzing force and movement data can reveal significant correlations. This is significant as it allows therapists to obtain data regarding multiple variables at once during one exercise instead of having to test the functionality of the knee with the participant in the same room using multiple tests to achieve a similar outcome [21].

XSL•FO
RenderX

Limitations of the Study

This study has some limitations that must be acknowledged from both statistical and clinical viewpoints. From a statistical perspective, the sample size of 32 participants, although adequate for preliminary temporal exploratory analysis, is not sufficient for generalizing the findings to larger and more diverse groups. A larger cohort would provide a more robust assessment of the observed kinetic patterns and enhance the reliability of the used clustering techniques. Furthermore, whereas unsupervised clustering effectively identified distinct movement patterns, the study design was deficient in longitudinal follow-up to assess the direct correlation between these patterns and long-term rehabilitation outcomes. Clinically, while the Slider device offers an option for home-based rehabilitation, the absence of real-time clinician supervision may lead to variability in movement execution and data quality, thereby impacting cluster accuracy. The predictive efficiency of the identified clusters in guiding treatment strategies remains uncertain without rigorous validation through randomized controlled trials or extended clinical monitoring.

Conclusions

Three distinct, kinetic, time-varying clusters were identified for both the right and left knees, representing unique movement patterns, based on measurement data from the Slider exercise device. Data clustering performance was found to be knee specific, with hierarchical and CLARA clustering methods producing better results at the right and left knees, respectively. The kinetic patterns at the right and left knees were significantly different. Gender, BMI, age, and pain score were identified as significant demographic predictors of cluster membership, with varying impacts on cluster assignment for the right and left knees. Future research should also explore other clustering algorithms, such as density-based clustering (DBSCAN) and Gaussian mixture models. Additionally, the impact of various preprocessing techniques should be investigated, especially for time-dependent and cyclical data. Techniques such as Fourier transforms, wavelet transforms, and PCA for dimensionality reduction may also enhance the analysis of movement patterns.

Furthermore, future studies should develop additional kinetic features, such as velocity, acceleration, and jerk, to provide a more comprehensive understanding of movement patterns and improve clustering accuracy. Expanding this study to include a more diverse participant population would help assess the generalizability of the findings, refine personalized rehabilitation protocols, and address the limitations of this study. Finally, further longitudinal and trial studies could be conducted to correlate identified clusters with clinical outcomes, such as recovery trajectories and therapy efficacy, to validate the practical significance of the clustering results. Further case studies are currently being planned, which will expand the scope of data collected and study participants. It is expected that with further data collection, it will be easier to identify more accurately what subgroup knee replacement participants fall into, as well as what range of movement and force recorded can be considered optimal based on certain demographics. This method appears to be effective at remotely measuring movement and force for knee joint movement. Future projections for Slider device development will include artificial intelligence–assisted projections and therapy progress analysis to streamline the process for physicians.

Acknowledgments

The authors express their heartfelt gratitude to the volunteers for their valuable time and consent, which made this study possible. This study was conducted independently and did not receive external funding.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to requirements of intellectual property protection but are available from the corresponding author upon reasonable request. The well-documented R scripts used for data analysis are publicly available for reproducibility of results [22].

Authors' Contributions

Each author contributed significantly to the study’s conception and design and provided approval for the final manuscript. MA and NJS played a key role in the collection of empirical data, with support from AI Rehab LLC, Atlanta, Georgia, who provided the devices. CT contributed by developing the mathematical methodologies and performing all statistical analyses. All authors actively participated in drafting and revising the manuscript.

Conflicts of Interest

The Slider Device and Slider digital app were generously provided at no cost by AI Rehab LLC, Atlanta, Georgia. The authors did not receive any financial support from AI Rehab LLC and have no financial ties to the company. AI Rehab LLC also provided complimentary training on the device’s use but had no role in the design, execution, data analysis, or interpretation of the results of this study. The authors serve as unpaid advisors to AI Rehab LLC.

References

XSL•FO
RenderX

1. Shan L, Shan B, Suzuki A, Nouh F, Saxena A. Intermediate and long-term quality of life after total knee replacement: a systematic review and meta-analysis. J Bone Joint Surg Am 2015 Jan 21;97(2):156-168. [doi: 10.2106/JBJS.M.00372] [Medline: 25609443]
2. Bloomfield RA, Broberg JS, Williams HA, Lanting BA, McIsaac KA, Teeter MG. Machine learning and wearable sensors at preoperative assessments: functional recovery prediction to set realistic expectations for knee replacements. Med Eng Phys 2021 Mar;89:14-21. [doi: 10.1016/j.medengphy.2020.12.007] [Medline: 33608121]
3. Islam R, Gooch D, Karlakki S, Price B. A device for prehabilitation of total knee replacement surgery (Slider): usability study. JMIR Form Res 2023 Dec 18;7:e48055 [FREE Full text] [doi: 10.2196/48055] [Medline: 38109191]
4. Whatling GM, Biggs PR, Wilson C, Holt CA. Assessing functional recovery following total knee replacement surgery using objective classification of level gait data and patient-reported outcome measures. Clin Biomech (Bristol) 2022 May;95:105625 [FREE Full text] [doi: 10.1016/j.clinbiomech.2022.105625] [Medline: 35429691]
5. Young-Shand KL, Roy PC, Dunbar MJ, Abidi SSR, Astephen Wilson JL. Gait biomechanics phenotypes among total knee arthroplasty candidates by machine learning cluster analysis. J Orthop Res 2023 Feb;41(2):335-344 [FREE Full text] [doi: 10.1002/jor.25363] [Medline: 35538599]
6. Lange JK, DiSegna ST, Yang W, Li W, Franklin PD. Using cluster analysis to identify patient factors linked to differential functional gains after total knee arthroplasty. J Arthroplasty 2020 Jan;35(1):121-126.e6. [doi: 10.1016/j.arth.2019.08.039] [Medline: 31530461]
7. R: a language and environment for statistical computing. R Core Team. 2023. URL: http://www.R-project.org/ [accessed 2024-09-28]
8. Hassler AP, Menasalvas E, García-García FJ, Rodríguez-Mañas L, Holzinger A. Importance of medical data preprocessing in predictive modeling and risk factor discovery for the frailty syndrome. BMC Med Inform Decis Making 2019 Feb 18;19(1):33 [FREE Full text] [doi: 10.1186/s12911-019-0747-6] [Medline: 30777059]
9. López-De-Lacalle J. R package Tsoutliers. The Comprehensive R Archive Network. 2019. URL: https://cran.hafro.is/web/packages/tsoutliers/vignettes/tsoutliers-intro.pdf [accessed 2024-06-19]
10. Schubert E, Rousseeuw PJ. Fast and eager k-medoids clustering: O(k) runtime improvement of the PAM, CLARA, and CLARANS algorithms. Inf Syst 2021 Nov;101:101804 [FREE Full text] [doi: 10.1016/J.IS.2021.101804]
11. Shi N, Liu X, Guan Y. Research on k-means clustering algorithm: an improved k-means clustering algorithm. New York: IEEE; 2010 Presented at: Third International Symposium on Intelligent Information Technology and Security Informatics; April 2-4, 2010; Jian, China p. 63-67 URL: https://ieeexplore.ieee.org/document/5453745 [doi: 10.1109/IITSI.2010.74]
12. Park H, Jun C. A simple and fast algorithm for K-medoids clustering. In: Expert Systems with Applications. Cham, Switzerland: Springer; 2009:3336-3341.
13. Murtagh F, Contreras P. Algorithms for hierarchical clustering: an overview. WIREs Data Min Knowl 2011 Dec 07;2(1):86-97 [FREE Full text] [doi: 10.1002/widm.53]
14. Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust: an R package for determining the relevant number of clusters in a data set. J Stat Software 2014;61(6):1-36 [FREE Full text] [doi: 10.18637/jss.v061.i06]
15. Shutaywi M, Kachouie NN. Silhouette analysis for performance evaluation in machine learning with applications to clustering. Entropy (Basel) 2021 Jun 16;23(6):759 [FREE Full text] [doi: 10.3390/e23060759] [Medline: 34208552]
16. Bucci CM, Legnani WE, Armentano RL. Clustering of cardiovascular risk factors highlighted the coronary artery calcium as a strong clinical discriminator. Health Technol 2016 Sep 30;6(3):159-165 [FREE Full text] [doi: 10.1007/s12553-016-0139-1]
17. Ripley B, Venables W. Package nnet. nginx. 2016. URL: https://staff.fmi.uvt.ro/~daniela.zaharie/dm2019/RO/lab/lab3/biblio/nnet.pdf [accessed 2024-06-20]
18. He F, Mazumdar S, Tang G, Bhatia T, Anderson SJ, Dew MA, et al. Non-parametric MANOVA approaches for non-normal multivariate outcomes with missing values. Commun Stat Theory Methods 2017;46(14):7188-7200 [FREE Full text] [doi: 10.1080/03610926.2016.1146767] [Medline: 29416225]
19. Giacomozzi C, Martelli F, Nagel A, Schmiegel A, Rosenbaum D. Cluster analysis to classify gait alterations in rheumatoid arthritis using peak pressure curves. Gait Posture 2009 Feb;29(2):220-224 [FREE Full text] [doi: 10.1016/j.gaitpost.2008.08.004] [Medline: 18815041]
20. Scivoletto G, Galli G, Torre M, Molinari M, Pazzaglia M. The overlooked outcome measure for spinal cord injury: use of assistive devices. Front Neurol 2019;10:272 [FREE Full text] [doi: 10.3389/fneur.2019.00272]
21. Loftus TJ, Shickel B, Balch JA, Tighe PJ, Abbott KL, Fazzone B, et al. Phenotype clustering in health care: a narrative review for clinicians. Front Artif Intell 2022;5:842306 [FREE Full text] [doi: 10.3389/frai.2022.842306] [Medline: 36034597]
22. Twumasi C. Kinetic pattern recognition in novel home-based physiotherapy: cluster analysis of multidimensional movement data from the Slider digital rehabilitation device (GitHub R codes). GitHub. 2024. URL: https://github.com/twumasiclement/Cluster-Analysis-of-Multidimensional-Movement-Data-from-the-Slider- [accessed 2025-02-17]

Abbreviations

CLARA: Clustering Large Applications

XSL•FO
RenderX

ML: machine learning
MKW: multivariate Kruskal-Wallis
OR: odds ratio
PAM: partition around medoids
PCA: principal component analysis

Edited by A Mavragani; submitted 23.11.24; peer-reviewed by A Holzinger; comments to author 15.02.25; revised version received 17.02.25; accepted 25.02.25; published 18.03.25.

Please cite as:
Twumasi C, Aktas M, Santoni N
Kinetic Pattern Recognition in Home-Based Knee Rehabilitation Using Machine Learning Clustering Methods on the Slider Digital Physiotherapy Device: Prospective Observational Study
JMIR Form Res 2025;9:e69150
URL: https://formative.jmir.org/2025/1/e69150
doi:10.2196/69150
PMID:

©Clement Twumasi, Mikail Aktas, Nicholas Santoni. Originally published in JMIR Formative Research (https://formative.jmir.org), 18.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Smartwatch-Based Ecological Momentary Assessment for High-Temporal-Density, Longitudinal Measurement of Alcohol Use (AlcoWatch): Feasibility Evaluation

Chris Stone[1,2], BSc, RSci; Sally Adams[3], PhD; Robyn E Wootton[2,4,5,6], PhD; Andy Skinner[1], PhD

[1]Integrative Cancer Epidemiology Programme, Bristol Medical School, University of Bristol, Bristol, United Kingdom

[2]School of Psychological Science, University of Bristol, Bristol, United Kingdom

[3]School of Psychology, University of Birmingham, Birmingham, United Kingdom

[4]Nic Waals Institute, Lovisenberg Diaconal Hospital, Oslo, Norway

[5]Medical Research Council Integrative Epidemiology Unit, Bristol Medical School, University of Bristol, Bristol, United Kingdom

[6]PsychGen Centre of Genetic Epidemiology and Mental Health, Norwegian Institute of Public Health, Oslo, Norway

Corresponding Author:
Chris Stone, BSc, RSci
School of Psychological Science
University of Bristol
12a Priory Road
Bristol, BS8 1TU
United Kingdom
Phone: 44 07983 317748
Fax: 44 0117 928 8588
Email: cstone2@btinternet.com

Abstract

Background: Ecological momentary assessment methods have recently been adapted for use on smartwatches. One particular class of these methods, developed to minimize participant burden and maximize engagement and compliance, is referred to as microinteraction-based ecological momentary assessment (µEMA).

Objective: This study explores the feasibility of using these smartwatch-based µEMA methods to capture longitudinal, high-temporal-density self-report data about alcohol consumption in a nonclinical population selected to represent high- and low-socioeconomic position (SEP) groups.

Methods: A total of 32 participants from the Avon Longitudinal Study of Parents and Children (13 high and 19 low SEP) wore a smartwatch running a custom-developed µEMA app for 3 months between October 2019 and June 2020. Every day over a 12-week period, participants were asked 5 times a day about any alcoholic drinks they had consumed in the previous 2 hours, and the context in which they were consumed. They were also asked if they had missed recording any alcoholic drinks the day before. As a comparison, participants also completed fortnightly online diaries of alcohol consumed using the Timeline Followback (TLFB) method. At the end of the study, participants completed a semistructured interview about their experiences.

Results: The compliance rate for all participants who started the study for the smartwatch µEMA method decreased from around 70% in week 1 to 45% in week 12, compared with the online TLFB method which was flatter at around 50% over the 12 weeks. The compliance for all participants still active for the smartwatch µEMA method was much flatter, around 70% for the whole 12 weeks, while for the online TLFB method, it varied between 50% and 80% over the same period. The completion rate for the smartwatch µEMA method varied around 80% across the 12 weeks. Within high- and low-SEP groups there was considerable variation in compliance and completion at each week of the study for both methods. However, almost all point estimates for both smartwatch µEMA and online TLFB indicated lower levels of engagement for low-SEP participants. All participants scored "experiences of using" the 2 methods equally highly, with "willingness to use again" slightly higher for smartwatch µEMA.

Conclusions: Our findings demonstrate the acceptability and potential utility of smartwatch µEMA methods for capturing data on alcohol consumption. These methods have the benefits of capturing higher-temporal-density longitudinal data on alcohol consumption, promoting greater participant engagement with less missing data, and potentially being less susceptible to recall errors than established methods such as TLFB. Future studies should explore the factors impacting participant attrition (the biggest reason for reduced engagement), latency issues, and the validity of alcohol data captured with these methods. The consistent

XSL•FO
RenderX

pattern of lower engagement among low-SEP participants than high-SEP participants indicates that further work is warranted to explore the impact and causes of these differences.

(JMIR Form Res 2025;9:e63184) doi:10.2196/63184

KEYWORDS

smartwatch; ecological momentary assessment; µEMA; alcohol; ALSPAC

Introduction

Ecological momentary assessment (EMA) refers to a class of methods in which people are asked to answer questions about specific aspects of their health, feelings, and behaviors as they go about their normal lives. As individuals typically complete these assessments throughout the day, rather than completing a survey and having to remember details from days or weeks ago, recall errors are significantly reduced [1]. One domain in which EMA methods have been used extensively is the study of modifiable health behaviors, including physical activity, diet, tobacco smoking, and alcohol consumption [2]. The benefits of EMA methods are of particular interest for the measurement of alcohol consumption. Commonly used questionnaires and diary-based methods are known to suffer from systematic biases toward underreporting, capturing somewhere between 20% and 60% of the true level of consumption [3,4], with retrospective recall bias being a major factor [5]. Furthermore, recent studies have highlighted the importance of both longitudinal measures of alcohol consumption to help us explore relationships between alcohol consumption and conditions such as depression [6], and high-temporal-resolution measures of alcohol consumption to enable exploration of factors such as subjective responses (eg, feeling of stimulation, feeling of sedation, liking the effects of alcohol, wanting more alcohol) in heavy drinkers in natural environments [7]. There is, therefore, a need for EMA methods that can be used to capture high-temporal-density measures of alcohol consumption over extended periods.

In response to the increased adoption of mobile and wearable technologies, EMA methods have been adapted for use on smartphones, and more recently, smartwatches. For smartwatch-based EMA methods, a particular area of focus has been reducing participant burden, so that even more frequent sampling over longer periods can be tolerated. Smartwatches are worn on the wrist so they are never beyond reach, which reduces the time to access the device and respond to prompts. Further, because they are worn against the body, they enable the use of potentially more discrete haptic prompts for responses. One smartwatch EMA method aiming to minimize participant burden is microinteraction-based EMA (µEMA), which uses simple, brief questions with a limited set of answers that can be responded to with a single tap multiple times a day [8]. When compared with smartphone-based EMA over a 4-week period, smartwatch-based µEMA had better performance in terms of compliance rates (defined as the percentage of all scheduled prompts to which participants responded, regardless of the success of prompt delivery: 82% vs 64%) and completion rates (defined as the percentage of scheduled prompts actually delivered to the participant to which they responded: 92% vs 67%) [8].

Having demonstrated high levels of engagement in these initial 4-week explorations of smartwatch-based µEMA, the approach was subsequently used to capture self-report measures over longer periods in various settings. Beukenhorst and colleagues [9] explored the use of smartwatch-based µEMA methods in participants with osteoarthritis of the knee to capture self-report measures of pain and quality of life 4 or 5 times a day over 3 months. The compliance rate across all participants ranged from around 85% at the beginning of the study to around 30% at the end of the 3 months. The authors suggested the reduction over time was largely a result of participant attrition, possibly driven by technical factors including poor watch battery life. Considering just active participants, compliance remained above 75% throughout the 3 months. Ponnada and colleagues [10] explored using the same smartwatch-based µEMA methods for capturing self-report data of physical activity and its effect in young adults over an even longer time frame of 12 months. Preliminary findings at 6 months (when the authors published interim results of their study) indicated an overall compliance of rate 67% and a completion rate of 79% [10]. These findings suggest that while it is necessary to be mindful of technical issues that can lead to attrition, smartwatch-based µEMA appears to be a viable method for collecting high-temporal-resolution self-report data over extended time frames.

In terms of the effect of sample characteristics on EMA study compliance, several recent studies have reported mixed findings for age and sex [11-13]. One sample characteristic not explored as extensively is socioeconomic position (SEP). For studies of alcohol use, SEP is of particular importance as there are known to be complex relationships between SEP, alcohol consumption, and health outcomes (the “alcohol harm paradox”), in which individuals are more likely to have drunk in the past week and drank 6 units of alcohol or more in 1 drinking episode if they are a high-income earning managerial/professional worker [14,15], but individuals with lower individual or neighborhood SEP have an increased susceptibility to the negative and harmful health effects of alcohol consumption [16,17]. Understanding if SEP has a systematic effect on engagement with any new methods measuring alcohol use is therefore clearly important.

The objective of this study was to explore the feasibility of using smartwatch-based µEMA methods for the capture of high-resolution self-reported data about alcohol use over extended periods in high- and low-SEP groups. We asked participants to use a smartwatch-based µEMA system to record self-reported data about alcohol use and the context in which alcohol was consumed 5 times a day over a 3-month period. For comparison, we also asked all participants to record their alcohol use every 2 weeks using an online version of one of the currently most established methods for capturing data on alcohol consumption, the Timeline Followback (TLFB) method. We

XSL•FO
RenderX

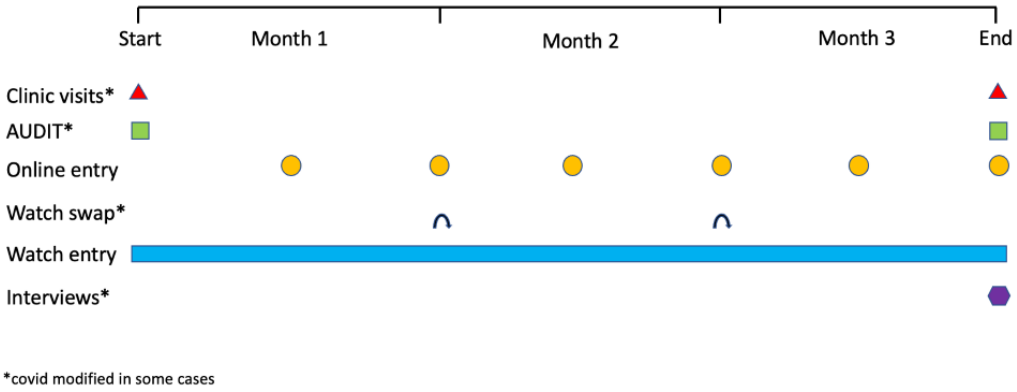
report engagement metrics (compliance and completion) for both smartwatch µEMA and TLFB methods, qualitative experiences of using both methods captured in interviews at the end of the study, and some initial findings comparing the levels of alcohol recorded by the 2 methods at the level of individual participants. Finally, in line with recommendations from a recent evaluation of pressing issues in EMA studies [18], we also report the characteristics of missing data for the smartwatch µEMA for all participants and high- and low-SEP groups, to explore whether SEP impacts the nature of missing data in a systematic manner.

Methods

Study Overview

The design and analyses of the study followed those detailed in the preregistered study protocol [19]. Any deviations from the study protocol are detailed in the following sections. An overview of the study elements, with their timings throughout the study, is shown in Figure 1. Data collection happened between September 2019 and June 2020.

Figure 1. Overview of study elements with timings.



Study/Clinical Setting

This study was embedded within the Avon Longitudinal Study of Parents and Children (ALSPAC), a longitudinal birth-cohort study. During phase I of ALSPAC enrollment, 14,541 pregnancies in the former Avon Health Authority in the southwest of England with expected dates of delivery between April 1, 1991, and December 31, 1992, were recruited. Of these initial pregnancies, there was a total of 14,676 fetuses, resulting in 14,062 live births and 13,988 children who were alive at 1 year of age. A further 906 pregnancies were recruited during phases II, III, and IV, respectively, resulting in an additional 913 children being enrolled. The total sample size is 15,447 pregnancies, of which 14,901 were alive at 1 year of age.

Further details on the cohort profile, representativeness, and phases of recruitment are described in 3 cohort-profile papers [20-22].

The ALSPAC study website provides details on all available data through a fully searchable data dictionary and variable search tool [23]. Informed consent for data collected via questionnaires and clinics was obtained from participants in accordance with the recommendations of the ALSPAC Ethics and Law Committee at the time.

In this study, data were captured using smartwatch-based µEMA techniques as participants went about their normal daily lives and through online TLFB methods periodically throughout the study. An initial session was held at the ALSPAC premises in the Bristol Medical School, during which participants were briefed, provided consent, and completed questionnaires. A final session was also held there, where participants were debriefed, completed a qualitative interview about their experiences during the study, and returned equipment.

Study data for the TLFB method were collected and managed using REDCap, a secure, web-based electronic data capture platform hosted at the University of Bristol. REDCap (Research Electronic Data Capture) is designed to support data capture for research studies [24].

Ethical Considerations

Ethics Review

As this research involved human participants, ethics approval was obtained from the ALSPAC Ethics and Law Committee, which reviewed the study and granted formal approval (reference code 83643) in accordance with the University of Bristol institutional guidelines.

Informed Consent

Prospective participants received an information sheet before the initial study session, explaining the study's nature, purpose, and risks, and were given the opportunity to ask investigators any questions before deciding to participate. Upon arrival at the initial session, participants could review the information sheet again and ask further questions. Informed consent for data collection was then obtained through a written consent form, following the recommendations of the ALSPAC Ethics and Law Committee. Participants were also informed that they could withdraw from the study at any time without providing a reason.

Privacy and Confidentiality

The data obtained from participants have been anonymized to ensure they are not identifiable. No individual participant can be identified from this manuscript or its multimedia appendices.

Compensation

Participants were reimbursed £15 (US $19.41) per month for their time and expenses, with a bonus of £15 (US $19.41) for

XSL•FO
RenderX

completing all 3 months. Thus, those who completed the full study received a total of £60 (US $78 at the time of the study).

Participants

Participants were drawn from the first generation of children born in the ALSPAC study (ALSPAC-G1) and were selectively recruited by SEP based on maternal highest educational attainment. Those whose mothers had a degree or higher were assigned to the high-SEP group, while those whose mothers had less than a degree were assigned to the low-SEP group.

The inclusion criteria were as follows:

- Consume at least 6 units of alcohol per week.
- Be able to wear and use the smartwatch from 12 PM to 10 PM (the timings of questions in the current version of the smartwatch system cannot be altered to accommodate shift work, etc)
- Have access to the internet to complete the TLFB assessments.

Exclusion criteria (standard exclusion criteria used in alcohol studies to prevent harm during pregnancy and to those at risk of addiction) were as follows:

- Pregnant or breastfeeding (self-reported). Note that if a participant becomes pregnant during the study, they must withdraw from the study at this point.
- History of alcohol or drug addiction (self-reported).
- Strong familial history of alcoholism defined as 1 or more immediate relatives (parent, sibling) or more than 1 other relative (eg, cousin, grandparent; self-reported).

Clinic Visits

All participants began the study with a visit to an ALSPAC clinic. In this session (approximately 1 hour in duration) participants

- were given details of the study and provided written consent;
- were given the smartwatch and shown how to use the µEMA application and maintain the smartwatch (including instructions on keeping it charged, not getting it wet, and removing it before contact sports);
- were given instructions on how to use the online TLFB diary, provided with log-on details, and asked about any events (eg, birthdays, nights out) that should be entered into the diary as reminders;
- were provided with contact details of whom to call if they encountered any issues during the study.

Participants completed a second clinic visit at the end of the study. In this session (approximately 30 minutes) participants

- returned the smartwatch;
- completed a brief semistructured interview about their experiences using the smartwatch and online TLFB diary led by the researcher;
- provided written final consent; and
- were given their cash reimbursement.

A total of 13 participants (8 low SEP and 5 high SEP) completed the second clinic visit in person.

As a result of COVID-19 restrictions, the remaining 19 participants (11 low SEP and 8 high SEP) completed their second clinic visit remotely. In these remote sessions, ALSPAC researchers conducted telephone interviews, and smartwatch collection was arranged via courier.

Smartwatch Data Collection

We developed a bespoke µEMA smartwatch application incorporating elements proposed by Intille et al [8] to reduce participant burden. These included using haptic rather than audible prompts to minimize disruption and designing brief questions with a limited set of single-tap responses to reduce effort.

To optimize our app design, we collaborated with the ALSPAC Original Cohort Advisory Panel, a committee representing the diverse backgrounds and perspectives of ALSPAC participants. Using a think-aloud approach, panel members walked through the µEMA application, identifying issues related to visual design, wording, and navigation. Several common themes emerged, leading to the following changes in the final µEMA application design:

- The introduction of the “My usual” shortcut to speed up the recording of alcoholic drinks that are the same type, quantity, and context.
- The introduction of the “Back” option so participants could go back and correct previous entries in the current epoch.

Participants received a TicWatch C2 (Mobvoi Information Technology Company Limited) smartwatch running Android Wear 2.6 with our µEMA application preloaded.

At 5 time points each day (2 PM, 4 PM, 6 PM, 8 PM, and 10 PM), participants were asked about any alcoholic drinks consumed in the past 2 hours and the context of these drinking events. At each time point, they were presented with the set of questions and response options listed in Textbox 1.

Examples of these questions as rendered in our µEMA application are shown in Figure 2, with the full question sequence illustrated in Multimedia Appendix 1.

XSL•FO
RenderX

Textbox 1. Questions and options presented to participants.

- Question 1: Did you drink alcohol in the last 2 hours? Response options [Yes | No | My Usual].
- Question 2: What were you drinking? Response options [Beer/cider | Wine | Spirits].
- Question 3: What size was your drink? Response options: Beer/cider: [Half a pint | 330 bottle | Pint], Wine: [125 mL | 175 mL |250 mL], Spirits: [Single | Double | Free pour].
- Question 4: Who were you with? Response options [Alone | In company].
- Question 5: Where were you? Response options [At home | Elsewhere].
- Question 6 Any more alcoholic drinks to be recorded? Response options [Yes | No | My Usual].

"Usual": On the first day of the study, participants were given the option to set up a "My Usual" shortcut to simplify data entry for their most frequently consumed alcoholic drink. If they opted in, they answered questions 2-5 to specify their usual drink. Thereafter, selecting "My Usual" automatically recorded those drink details for that time point.

Figure 2. Example smartwatch microinteraction-based ecological momentary assessment (µEMA) application questions.

Each day at noon, participants were asked if they had missed recording any drinks the previous day. If they responded "Yes," they were prompted to report those drinks using the same set of questions.

Captured data were not visible to participants and were stored locally on the smartwatch rather than uploaded to a cloud-based platform to minimize the risk of personal data loss. As a result of the lack of remote monitoring, smartwatches were swapped at the end of months 1 and 2 to mitigate the risk of undetected technical issues disrupting data collection. However, COVID-19 restrictions prevented this for 15 participants (7 low SEP and 8 high SEP), who were instead instructed to continue using their current smartwatch until the study's conclusion.

Online Data Collection

The TLFB method uses a calendar prepopulated with participant-defined diary entries (eg, birthdays, theatre visits, nights out) to aid in recall of alcohol consumption. Online self-administration of TLFB has been validated [25], and previous studies suggest that a 14-day recall period optimally balances data quality and participant burden [26]. Therefore, participants completed an online TLFB every 14 days, recording their alcohol consumption over the previous 2 weeks.

We implemented TLFB using REDCap [24], hosted on the ALSPAC application platform. Participants had a 5-day window to complete each 2-week TLFB entry and received SMS text message reminders at the start of this period. In each entry, participants recorded the number of alcohol units consumed each day. As data captured in the smartwatch µEMA application were not visible to participants, they could not use it to complete TLFB entries. In line with standard TLFB practice, no context questions were asked. The format of the TLFB diary entry screen is shown in Multimedia Appendix 2.

Interviews

In the second clinic session at the end of the study (conducted either in person or remotely by telephone), each participant completed a semistructured interview exploring their experiences during the study.

For both smartwatch µEMA and online TLFB methods, participants were asked:

Question 1/6: Overall, how would you rate your experience of using the smartwatch/online system during the study, on a scale from 1 (I did not like it at all) to 10 (I really liked it)?

Question 2/7: If you were asked to use the smartwatch/online system again in another study, how likely would you be to say yes, on a scale from 1 (I would not want to use it again) to 10 (I would really like to use it again)?

Question 3/8: If you agreed to use the smartwatch/online system again in another study, what is the maximum length of time you would be willing to use it for?

Question 4/9: What things did you like about using the smartwatch/online system?

Question 5/10: What things did you dislike about using the smartwatch/online system?

XSL•FO
RenderX

Participants were asked additional questions specifically about their experiences using the smartwatch µEMA system:

Question 11: How did you feel about having to charge the smartwatch every night?

Question 12: Did you use any of the other functions on the smartwatch during the study (eg, step counter)?

Question 13: Did you pair the smartwatch with any smartphones during the study?

Participants were also asked about their attitudes toward future developments of the smartwatch µEMA system:

Question 14: How would you feel if the smartwatch system used data from its motion sensors to work out when the best time is to ask you the questions (eg, so it does not ask you questions when you are driving)?

Question 15: How would you feel if the smartwatch system used data from its global positioning system (GPS) location sensor (which tells the watch where the person is) to work out when the best time is to ask you the questions (eg, so it does not ask you questions when you are at work)?

Engagement Measures

In reporting engagement, we follow the approach used by Intille and colleagues [8]. We report compliance rates, defined as the proportion of scheduled questions answered by participants, both for all participants who started the study and for those still active in the study. Participants are considered active until they make their final response and drop out.

In addition, for smartwatch µEMA, we report the completion rate, defined as the proportion of questions delivered to active participants by the µEMA application that they respond to. This helps distinguish whether nonresponses result from factors preventing question delivery (eg, the smartwatch being switched off or the battery being flat) or from participants' propensity and ability to respond to delivered questions. The completion rate is not reported for online TLFB, as questions were not actively presented to participants, meaning there were no prompts that could be delivered or missed.

For all engagement measures, we report data for all participants as well as separately for the low- and high-SEP groups.

Missing Data

We examined whether prompts were more likely to be missed at certain times of the day by comparing missingness rates across 2-hour epochs. Similarly, we assessed whether prompts were more likely to be missed on specific days of the week using the same approach. These analyses were conducted for all participants as well as separately for the low- and high-SEP groups.

Comparison of Levels of Alcohol Consumption Recorded Using the Two Methods

To provide an initial indication of whether µEMA might capture higher levels of alcohol consumption than TLFB, we calculated the total number of units consumed over the 12-week study period for each participant, using only weeks where data were available from both methods. Within the high- and low-SEP groups, we compared the number of participants whose recorded consumption was higher in one method than the other. Note that this was an exploratory analysis and not preplanned.

Results

Participants

Our original aim was to recruit 40 participants (20 low SEP and 20 high SEP). However, recruitment had to be halted prematurely in March 2020 due to COVID-19 restrictions, by which point 32 participants had been recruited. Of these, 13 were high SEP (mean age 27 years, range 26-28 years; 7, 54%, females) and 19 were low SEP (mean age 28 years, range 26-28 years; 13, 68%, females).

Engagement

Figure 3 presents the compliance rate for all participants who commenced the study using the smartwatch µEMA and online TLFB methods. Figure 4 illustrates the compliance rate for participants who remained active in the study (ie, those who had not yet dropped out) for both methods. Figure 5 displays the completion rate for all active participants using the smartwatch µEMA method (not applicable to TLFB, as explained above). Figure 6 depicts the participant attrition rate, representing the percentage of participants still actively engaged in the study, for both smartwatch µEMA and online TLFB methods.

XSL•FO
RenderX

Figure 3. Smartwatch microinteraction-based ecological momentary assessment (top) and timeline followback (TLFB; bottom) compliance rates (mean and 95%CI) across all participants who started the study, for all (black solid line), high-SEP (red dotted dashed line), and low-SEP (blue dashed line) participants. SEP: socioeconomic position.

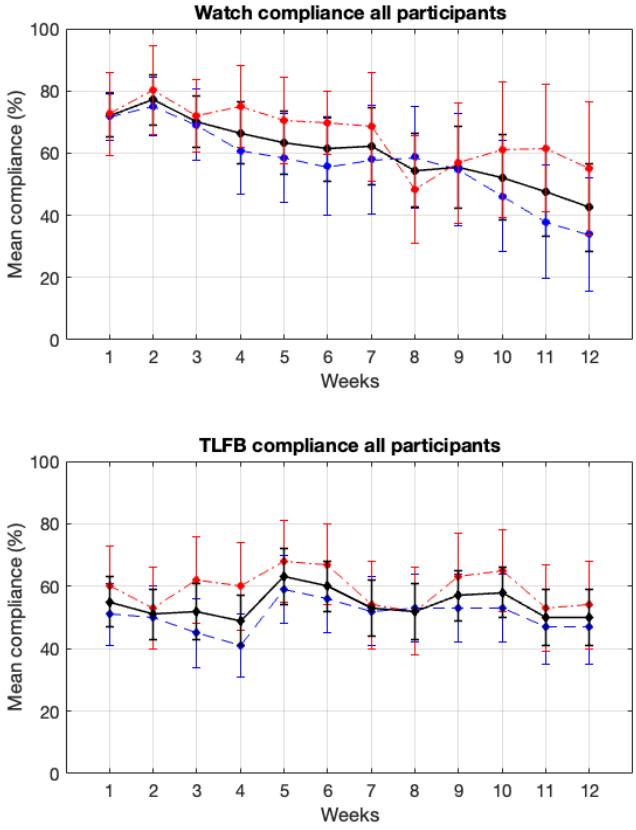
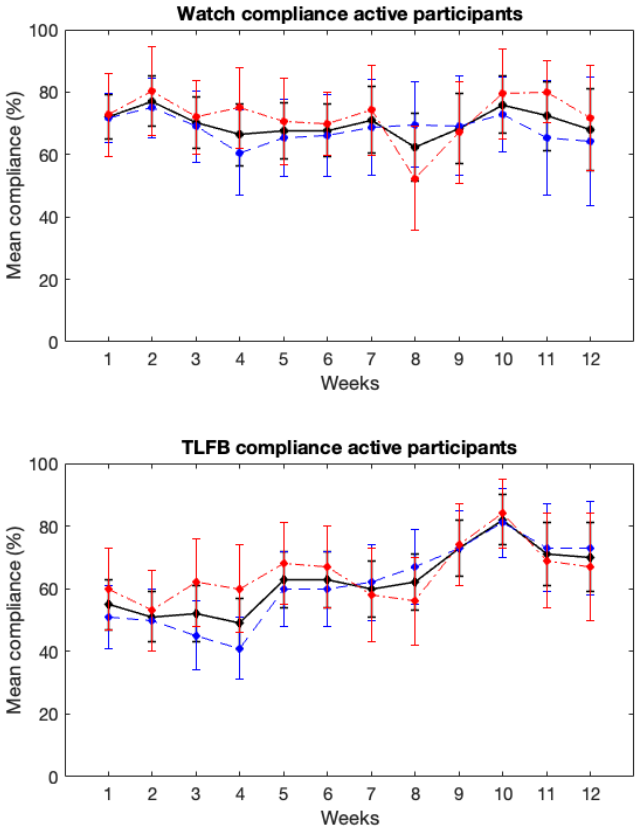


Figure 4. Smartwatch microinteraction-based ecological momentary assessment (top) and timeline followback (TLFB; bottom) compliance rates (mean and 95%CI) across all participants still active in the study for all (black solid line), high-SEP (red dotted dashed line), and low-SEP (blue dashed line) participants. SEP: socioeconomic position.



XSL•FO
RenderX

Figure 5. Smartwatch microinteraction-based ecological momentary assessment completion rate (mean and 95%CI) for all (black solid line), high-SEP (red dotted dashed line), and low-SEP (blue dashed line) participants. SEP: socioeconomic position.

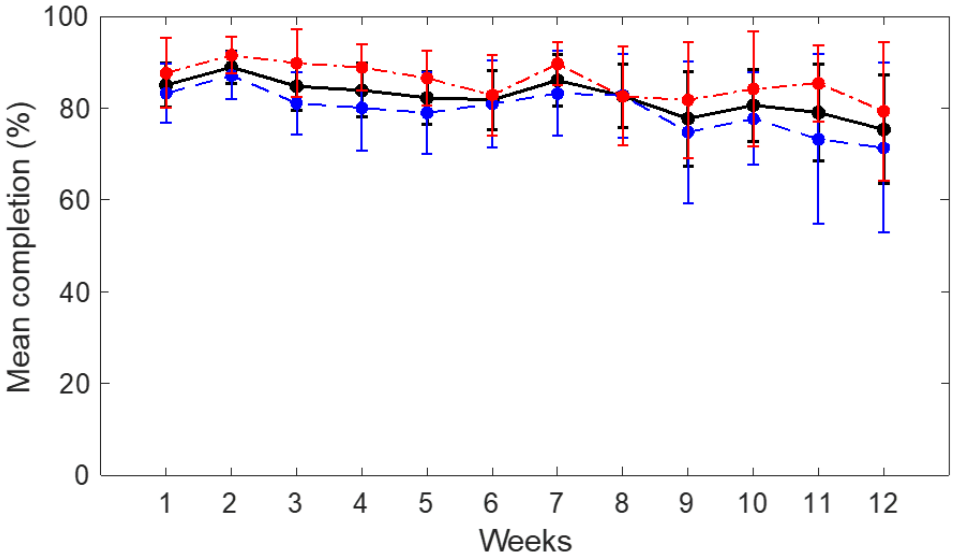
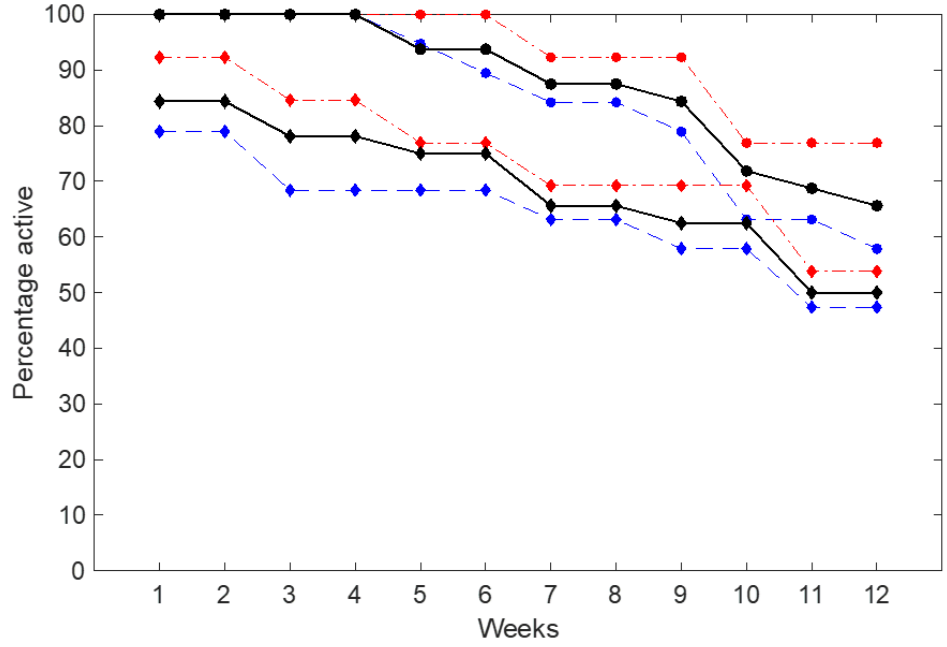


Figure 6. Participant attrition for smartwatch microinteraction-based ecological momentary assessment (circles) and online timeline followback (diamonds) for all (black solid line), high-SEP (red dotted dashed line), and low-SEP (blue dashed line) participants. SEP: socioeconomic position.



Qualitative Feedback

All participants completed interview questions regarding both the online and smartwatch methods for data collection. Participants' responses to the questions "How would you rate your experience of using the smartwatch µEMA/online TLFB from 1 (did not like) to 10 (really liked)?," "How willing would you be to use the smartwatch µEMA/online TLFB again in another study from 1 (would not) to 10 (really like to)?," and "How long would you be willing to use it for (in months)?" are presented in Table 1.

Themes identified from participants' responses to the question "What things did you like and dislike about using the smartwatch system?" are presented in Table 2, grouped by SEP. The equivalent themes for the online TLFB system are shown in Multimedia Appendix 3. Table 2 also includes participants' responses to the questions "How did you feel about having to charge the smartwatch every night?," "Did you pair the smartwatch with any smartphones during the study?," and "Did you use any of the other functions on the smartwatch during the study (eg, step counter)?"

Distinct themes identified from participants' responses to the questions "How would you feel if the smartwatch system used data from its GPS location sensor to determine the best time to ask you questions (eg, to avoid prompting during work)?" and "How would you feel if the smartwatch system used data from its motion sensors to determine the best time to ask you questions (eg, to avoid prompting while driving)?" are presented in Table 3.

Because of a technical issue with the questionnaire delivery, 6 participants (2 high SEP [1 female] and 4 low SEP [2 female])

XSL•FO
RenderX

were not asked the questions about smartwatch charging, pairing and other functions, or about the potential use of GPS and motion sensors.

Table 1. Participants' ratings of the extent to which they liked using, would be willing to use it again, and for how long, for the smartwatch µEMA[a] (Watch) and online (TLFB[b]) methods.

Participants' ratings of smartwatch µEMA/TLFB	All SEP[c], median (IQR)		Low SEP, median (IQR)		High SEP, median (IQR)	
	Watch	TLFB	Watch	TLFB	Watch	TLFB
Experience of using	8 (6-8)	7 (6-8)	8 (6-8)	7 (5-9)	8 (7-8)	8 (5-8)
Willing to use again	10 (7-10)	8 (7-10)	10 (7-10)	9 (7-10)	9 (6.5-10)	8 (6-10)
How long for (months)?	3 (3-12)	4 (3-12)	6 (3-12)	6 (3-12)	3 (2-5)	3 (2-5)

[a]µEMA: microinteraction-based ecological momentary assessment.

[b]TLFB: timeline followback.

[c]SEP: socioeconomic position.

Table 2. Participants' likes and dislikes of the smartwatch µEMA[a] method, attitudes to daily charging, attempts to pair with phone, and other watch features used, for low- and high-SEP[b] participants.

Participants' likes/dislikes of smartwatch µEMA	Low SEP, n/N (%)	High SEP, n/N (%)
Watch liked		
Quick and easy	9/19 (47)	6/13 (46)
Being prompted	1/19 (5)	4/13 (31)
Having a watch	0/19 (0)	2/13 (15)
Liked the smartwatch	2/19 (10)	2/13 (15)
Additional smartwatch features	2/19 (10)	0/13 (0)
No need for extra device	0/19 (0)	2/13 (15)
Always to hand	4/19 (21)	2/13 (15)
Watch disliked		
Timing/frequency of prompts	4/19 (21)	3/13 (23)
Technical issues	3/19 (16)	4/13 (31)
Battery life	4/19 (21)	2/13 (15)
Watch too big	3/19 (16)	2/13 (15)
Unsure how to handle missing drinks	1/19 (5)	2/13 (15)
Already have a watch/smartwatch	2/19 (10)	1/13 (8)
Attitude to needing to charge every night		
No issues	11/15 (73)	8/11 (73)
Frustrating	4/15 (27)	2/11 (18)
Attempted to pair with a smartphone		
Did not attempt	13/15 (87)	9/11 (82)
Attempted but failed	2/15 (13)	0/11 (0)
Attempted and succeeded	0/15 (0)	2/11 (18)
Other watch features used		
None	11/15 (73)	5/11 (45)
Step counter	3/15 (20)	3/11 (27)
Timer	1/15 (7)	2/11 (18)

[a]µEMA: microinteraction-based ecological momentary assessment.

[b]SEP: socioeconomic position.

XSL•FO
RenderX

Table 3. Participants' views on using GPS[a] data and motion sensor data to optimize smartwatch µEMA[b], for low- and high-SEP[c] participants.

Participants' attitudes toward the use of sensor data	Low SEP (n=15), n (%)	High SEP (n=11), n (%)
Attitudes to using GPS in the future		
No issues	13 (87)	7 (64)
Depends on context	3 (20)	1 (9)
OK if consented	0 (0)	1 (9)
Concerned about battery life	0 (0)	1 (9)
Scary	2 (13)	0 (0)
Intrusive	2 (13)	0 (0)
Attitudes to using motion sensors in the future		
No issues	15 (100)	9 (82)
Concerns about reliability	0 (0)	1 (9)
Uncomfortable with idea	0 (0)	1 (9)

[a]GPS: global positioning system.

[b]µEMA: microinteraction-based ecological momentary assessment.

[c]SEP: socioeconomic position.

Missing Data

The rates of missingness for the smartwatch µEMA method among participants still active in the study, stratified by all, low-, and high-SEP groups, are shown for each day of the week in Table 4. Considerable variability was observed across the 3 groups, with overlapping CIs for all days.

The rates of missingness for the smartwatch µEMA method among participants still active in the study, stratified by all, low-, and high-SEP groups, are shown for epochs within days in Table 5. Considerable variability was observed within each group, with overlapping CIs across epochs. However, there was notably less overlap when comparing the first four epochs (12:00–14:00, 14:00–16:00, 16:00–18:00, 18:00–20:00) with the final epoch (20:00–22:00), which had the highest level of missing data.

Table 4. Percentage missingness by day of the week for all, high-, and low-SEP[a] groups (mean and 95% CI).

Weekday	All participants (n=32), mean (95% CI)	High SEP (n=13), mean (95% CI)	Low SEP (n=19), mean (95% CI)
Monday	28.3 (21.2-35.5)	23.8 (14.4-33.3)	31.4(20.8-42.0)
Tuesday	25.6 (18.8-32.4)	24.9 (15.9-33.9)	26.1 (15.8-36.5)
Wednesday	25.0 (19.4-30.6)	23.6 (14.0-33.2)	26.0 (18.5-33.4)
Thursday	26.5 (21.1-31.2)	24.2 (15.3-33.1)	28.1 (20.6-35.5)
Friday	25.2 (19.1-31.2)	24.9 (14.6-35.1)	25.3 (17.0-33.7)
Saturday	28.0 (21.3-34.6)	25.8 (15.0-36.6)	29.5 (20.2-38.8)
Sunday	28.2 (21.0-35.4)	29.6 (18.2-41.0)	27.2 (17.0-37.4)

[a]SEP: socioeconomic position.

Table 5. Percentage missingness by 2-hour epoch for all, high-, and low-SEP[a] groups.

Epoch	All participants (n=32), mean (95% CI)	High SEP (n=13), mean (95% CI)	Low SEP (n=19), mean (95% CI)
12:00-14:00	26.7 (20.4-32.9)	26.1 (17.3-34.9)	27.0 (17.6-36.4)
14:00-16:00	23.6 (17.7-29.5)	22.2 (13.7-30.8)	24.5 (15.9-33.2)
16:00-18:00	24.7 (18.5-30.8)	23.1 (14.1-32.0)	25.8 (16.8-34.7)
18:00-20:00	26.0 (20.4-31.5)	24.2 (15.5-32.8)	27.2 (19.3-35.1)
20:00-22:00	32.5 (25.5-40.0)	30.7 (20.1-41.2)	33.8 (23.5-44.0)

[a]SEP: socioeconomic position.

XSL•FO
RenderX

Comparison of Levels of Alcohol Consumed Recorded Using the Two Methods

In the high-SEP group:

- 8 of the 13 (62%) participants had higher levels of alcohol units recorded using smartwatch µEMA than TLFB (percentage increase from TLFB: +44, +63, +27, +34, +76, +19, +2, and +71).
- 4 of the 13 (31%) had lower levels of alcohol recorded using smartwatch µEMA than TLFB (percentage decrease from TLFB: –11, –42, –10, and –50).
- 1 of the 13 (8%) did not have sufficient data for a comparison.

In the low-SEP group:

- 10 of the 19 (53%) participants had higher levels of alcohol units recorded using smartwatch µEMA than TLFB (percentage increase from TLFB: +42, +47, +18, +217, +10, +126, +33, +45, +12, and +15).
- 5 of the 19 (26%) had lower levels of alcohol recorded using smartwatch µEMA than TLFB (percentage decrease from TLFB: –6, –15, –61, –4, and –18).
- 4 of the 19 (21%) did not have sufficient data for a comparison.

Discussion

Principal Findings

Comparing Smartwatch µEMA and Online TLFB Engagement Metrics

Using the most stringent measure of engagement (compliance rate for all participants who started the study), compliance with the smartwatch µEMA method gradually declined over the 12 weeks, from approximately 70% in week 1 to around 45% in week 12. By contrast, compliance with the online TLFB method remained relatively stable at approximately 50% throughout the study. Based on this metric, the smartwatch µEMA method outperformed the online TLFB method until around week 9, after which their performance was broadly similar.

Comparing the compliance rates for all participants with those for participants still active in the study, and considering the attrition rate, it is evident that attrition significantly contributed to the observed decline in compliance among all participants over time. When focusing solely on participants still active in the study (ie, removing the effects of attrition), compliance with the smartwatch µEMA method remained relatively stable at around 70% throughout the 12 weeks. By contrast, compliance with the online TLFB method increased from approximately 50% in week 1 to around 80% in week 10, before declining to 70% in week 12. As before, the performance of the smartwatch µEMA method surpassed that of the online TLFB method until approximately week 9, after which only minor differences were observed between them. Notably, compliance with the online TLFB method increased during the first 10 weeks of the study. While no definitive explanation can be provided, one possible contributing factor is that COVID-19 lockdowns may have resulted in participants spending more time at home during certain periods, potentially improving ease of access to the online TLFB system. By contrast, access to the smartwatch µEMA method would not have been affected to the same extent, as it was specifically designed to be accessible regardless of location. Consequently, a similar increase in compliance would not be expected for this method.

Turning to the completion rate of the smartwatch µEMA method (not applicable to the online TLFB method, as previously discussed), this remained broadly stable at approximately 85% for the first 8 weeks, before declining to around 75% over the final 4 weeks. This suggests that, when the µEMA application functions correctly and the smartwatch device remains charged to ensure prompt delivery, participants exhibit a strong propensity to respond. The slight decline over the 12-week period indicates that longer data collection periods may still be feasible while maintaining acceptable completion rates using this method.

Comparing Engagement Metrics for Low- and High-SEP Groups

Although there is considerable variation within groups, and the differences between groups are generally modest (ranging from approximately 5% to 20%), nearly all point estimates for participant compliance, active participant compliance, completion rate, and attrition suggest lower engagement among low-SEP participants compared with high-SEP participants for both the smartwatch µEMA and online TLFB methods.

Patterns of Missing Data

There was considerable variability in all 3 groups (all SEP, low SEP, and high SEP), with overlapping CIs when comparing levels of missing data by days of the week and by 2-hour epochs within each group. The most notable pattern was the lower overlap in CIs observed in the all-participant group when comparing the first 4 epochs (12:00-14:00, 14:00-16:00, 16:00-18:00, and 18:00-20:00) with the final epoch (20:00-22:00), which had the highest level of missing data. Modifications to the µEMA application, such as allowing participants to configure their bedtime individually, may enhance engagement and reduce missing data during this period.

Regarding differences between high- and low-SEP groups, there was considerable overlap in CIs across all days and epochs. However, the point estimates for the proportion of missing data were consistently higher for low-SEP participants than for high-SEP participants across all epochs and on all days except Friday and Sunday.

Comparison of Levels of Alcohol Consumed Recorded Using the Two Methods

In both high- and low-SEP groups, approximately twice as many participants recorded higher alcohol consumption using the smartwatch µEMA method compared with the TLFB method over the study period. As this is a feasibility study, it was not powered to detect definitive differences in such comparisons, and these findings should be interpreted with caution. However, given the well-documented issue of underreporting in existing alcohol consumption recording methods [3,4], these preliminary results suggest that further research is warranted to assess whether smartwatch µEMA offers improvements in this area.

XSL•FO
RenderX

Qualitative Feedback

Across all participants, both methods were rated highly, with a slight preference for the smartwatch µEMA (8/10) over the online TLFB method (7/10). In the low-SEP group, the smartwatch µEMA received a slightly higher rating (8/10) than the online TLFB method (7/10), while the high-SEP group rated both methods equally (8/10).

For willingness to use the methods again, both scored highly, with all participants showing a preference for the smartwatch µEMA (10/10) over the online TLFB method (8/10). This pattern was consistent across SEP groups, with the high-SEP group rating the smartwatch µEMA slightly higher (9/10 vs 8/10) and the low-SEP group showing a stronger preference (10/10 vs 9/10).

Regarding the duration participants would be willing to use the methods again, online TLFB was rated slightly higher overall (4 months) compared with the smartwatch µEMA (3 months). However, within SEP groups, no differences were observed, with the low SEP group indicating a willingness to use both methods for 6 months and the high-SEP group for 3 months.

For the smartwatch µEMA method, the most appreciated factor was its ease and speed of use. However, participants reported a range of disliked aspects, including the timing and frequency of prompts, technical issues, battery life, and smartwatch size, all of which may have contributed to attrition. Regarding daily charging, 19 out of 26 (73%) participants in both SEP groups had no issues, while 4 out of 15 (27%) in the low-SEP group and 2 out of 11 (18%) in the high-SEP group found it frustrating. Additionally, 22 out of 26 (85%) participants in both groups adhered to study instructions by not pairing the smartwatch with their own smartphone. In the low-SEP group, 2 of the 15 (13%) who attempted to pair the smartwatch with their smartphone were unsuccessful, whereas in the high-SEP group, 2 of the 11 (18%) who attempted were successful. High-SEP participants (6/11, 55%) explored additional smartwatch features more frequently than low-SEP participants (4/15, 27%), with the step counter and timer being the most commonly used features.

Regarding attitudes toward using additional smartwatch sensors for data capture, the low-SEP group was more accepting of both GPS (13/15, 87%, had no concerns vs 7/11, 64%, in the high-SEP group) and motion sensors (15/15, 100% vs 9/11, 82%, in the high-SEP group).

Comparison With Prior Work

Alcohol EMA Studies

Perski et al's [2] systematic review and meta-analysis of EMA studies on health behaviors reported a median adherence of 84% (IQR 77%-91%) across 175 alcohol EMA studies, with smartphone-based EMA methods showing higher adherence than other approaches (eg, online and paper-and-pen). Notably, no studies included wrist-worn wearable EMA methods. Among the engagement metrics we assessed, the completion rate appeared most comparable to their adherence metric, averaging around 80% over the 12 weeks of our study—placing it within the reported range.

In his review of alcohol EMA methods, Piasecki [27] identified several studies employing high-resolution EMA approaches—primarily smartphone based—that prompted participants for responses throughout the day, with study durations ranging from a few days to 8 weeks. Reported response rates varied between 63% and 90%. While direct comparisons with our engagement metrics are not straightforward, our completion rate (~80%) and active participant compliance rate (~70%) fall within this range. Our all-participant compliance rate started at ~70% but dropped below 60% by week 8, remaining broadly consistent with these studies. However, as our study extended beyond 8 weeks, all-participant compliance declined further to 45% by week 12, suggesting that engagement may decrease more substantially in studies of longer duration.

Howard and Lamb [28] recently employed SMS text message–initiated smartphone-based EMA surveys to study alcohol consumption among undergraduates, reporting high compliance levels for active participants—starting at 89% in week 1 and gradually declining to 70% by week 14. By contrast, our active participant compliance rates were lower, fluctuating around 70% over the 12-week period. While no single factor can be identified as the cause of this difference, potential contributors include the nature of the prompts—the study by Howard and Lamb [28] used audible phone notifications, which may have been more attention grabbing than the smartwatch's haptic prompts—and differences in reward structures, with their incentives tied to individual survey completion, whereas ours were based on months of study participation. However, prior meta-analyses suggest that variations in reward strategies do not consistently impact engagement levels [11,12].

Smartwatch µEMA Studies

To the best of our knowledge, there are currently no broadly agreed benchmarks for compliance, completion, and attrition rates in smartwatch µEMA studies. In a study with a comparable assessment period, Beukenhorst et al [9] used smartwatch µEMA methods to collect self-reported pain data from patients with knee osteoarthritis, administering 4-5 questions per day over 13 weeks. Their attrition rate closely mirrored that of our study, with active participant numbers steadily declining to approximately 70% by week 10. However, by the end of the 13-week period, their attrition rate had dropped further to 42%, whereas ours remained at 65% by week 12. The compliance rate for all participants in the study by Beukenhorst et al [9] was also similar to that observed in our study, starting at approximately 75% before gradually declining to around 50% by week 10 and 45% by week 12. Completion rates showed a comparable pattern, beginning at approximately 85% and remaining around 80% throughout the study duration. Additionally, participants in both studies identified short battery life as a barrier to smartwatch use.

In what may be the longest smartwatch µEMA study to date, Ponnada et al [10] collected self-reported data on physical activity and its effect in young adults over a 12-month period, using smartwatch-based µEMA methods across 24 four-day burst periods. In their preliminary findings at the 6-month mark, the mean overall compliance of µEMA for active participants

XSL•FO
RenderX

during burst periods was 67%, comparable to the 70% observed over the 12 weeks of this study. Similarly, their reported mean completion rate of 79% at 6 months aligns with the completion rate in this study, which began at 85% and declined to 75% by week 12.

Effects of High and Low SEP

Regarding the impact of SEP on engagement in alcohol EMA studies, Howard and Lamb [28] found no significant differences in the number of students with college-educated parents (a measure similar to the SEP indicator used in this study) across their 4 responder categories ("poor," "adequate," "good," and "super"). However, they noted that the proportion of students with college-educated parents was higher in the "good" and "super" responder categories than in the "poor" category, suggesting that higher SEP may have contributed to increased engagement—consistent with the findings of this study.

Considering the broader impact of SEP on compliance in alcohol studies, Thern and Landberg [29] reported that low-SEP participants had approximately twice as much missing data as high-SEP participants for self-reported factors such as self-rated health, health-related quality of life, and level of social support received. While the current feasibility study is not powered to provide definitive evidence of higher levels of missing data in low-SEP participants, the point estimates observed here align with this pattern.

Strengths and Limitations

This is the first study to assess the feasibility of using smartwatch-based µEMA methods to collect high-temporal-density data on alcohol consumption over extended periods.

While our smartwatch-based EMA app was largely based on the µEMA methods originally developed by Intille and colleagues [8], a key aspect of true µEMA implementation is the presentation of only 1 question at a time. In our approach, we chained questions to capture details on drink type, size, and consumption context, making it more accurately described as a modified µEMA. However, this modification appeared to have little impact on engagement.

Unfortunately, due to COVID-19, participant recruitment had to be halted early, leading to a reduced sample size and an unequal mix of high- and low-SEP participants. As a result, the generalizability of our findings may be limited compared with studies conducted outside the context of a global pandemic.

Our participants were recruited from a UK birth cohort study [20]. A key advantage of this approach is that future research can explore new methods for integrating high-temporal-density, longitudinal data on alcohol consumption from our smartwatch-based µEMA methods with the extensive phenotypic and genetic data available for these participants. This could enable novel investigations into the impact of alcohol consumption on health. However, a potential limitation is that participants in cohort studies are accustomed to regular assessments, which may have increased their motivation to engage with the study compared to individuals not involved in such cohorts [20].

There is growing recognition that the diverse adherence metrics reported in EMA studies can hinder meaningful comparisons across the literature [10,11]. In this study, we selected adherence metrics primarily to facilitate comparisons with existing smartwatch-based µEMA studies. However, we acknowledge that these metrics do not always allow for direct comparisons with other EMA studies using different methodologies.

Recommendation for Future Studies

The primary factor affecting adherence in this study appeared to be participant attrition. Feedback from the end-of-study interviews provided insights into potential causes, with some issues—such as smartwatch size and battery life—likely to improve as technology advances. However, other factors, including the timing and frequency of prompts, warrant further investigation to optimize compliance and minimize attrition. Additionally, future studies should consider whether specific contextual factors (eg, study location) influence engagement with µEMA methods. A recent work by Ponnada and colleagues [30] offers a valuable overview of this topic.

The consistency of the differences observed between high- and low-SEP groups in compliance, completion, and attrition suggests the need for further research to quantify these disparities and identify their underlying drivers. This also highlights the importance of incorporating user input from diverse SEP backgrounds during the early stages of system design. Beyond SEP, future research should also consider other demographic factors, such as age, that may influence engagement and the quality of data collected using these methods.

This study has focused on the feasibility of using new smartwatch-based µEMA methods to capture data on alcohol consumption. Future research should also investigate response latency (the time between prompt delivery and response), the validity of the data collected compared with existing methods such as TLFB, and how both feasibility and validity vary across other health-related behaviors beyond alcohol consumption.

As a community developing new EMA methods, it is crucial to adopt metrics that facilitate the widest and most meaningful comparisons of findings. Adherence measurement has become increasingly complex, as new EMA methods leveraging the latest technologies benefit from specific metrics (eg, completion and compliance) that help distinguish between technical limitations and participant engagement. Given these advancements, now may be an opportune time to revisit previous recommendations for standardized reporting in EMA studies [31,32], considering the capabilities of mobile and wearable technology–enabled methods.

Conclusions

In this study, participants demonstrated higher engagement—reflected in higher compliance and completion rates and lower attrition—when using smartwatch-based µEMA methods compared with online TLFB methods for recording alcohol consumption. These engagement levels were consistent with previous studies employing smartwatch µEMA methods. While participants reported similar experiences with both methods and expressed comparable willingness to use either in

XSL•FO
RenderX

the future, they showed a slight preference for the smartwatch µEMA approach over the online TLFB method.

While certain challenges remain—particularly for studies extending over months rather than weeks—our findings highlight the considerable potential of smartwatch µEMA methods in alcohol use research. These methods offer the advantage of capturing high-temporal-density, longitudinal data on alcohol consumption, along with contextual information about drinking events. As with EMA methods more broadly, they are less susceptible to recall and bias errors that affect traditional alcohol consumption reporting methods. However, it is important to acknowledge that social desirability bias remains a significant factor in alcohol reporting, and some level of underreporting may persist despite the use of EMA techniques.

Future studies should investigate factors contributing to participant attrition—the primary driver of reduced adherence—as well as latency issues and the validity of alcohol data captured using these methods. While our study does not provide definitive evidence of engagement differences between high- and low-SEP participants, the consistent pattern of lower engagement observed among low-SEP participants suggests that further research is warranted to explore the existence, magnitude, and underlying causes of such disparities.

Acknowledgments

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. This work was funded by a UK Research and Innovation Fellowship to AS (MR/S003894/1), and further supported by the Integrative Cancer Epidemiology Programme, funded by Cancer Research UK (grant C18281/A29019). The Medical Research Council Integrative Epidemiology Unit is supported by the Medical Research Council and the University of Bristol (grants MC_UU_00032/05 and MC_UU_00032/07). REW is funded by a postdoctoral fellowship from the South-Eastern Norway Regional Health Authority (2020024). The UK Medical Research Council and Wellcome (grant 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC.

Data Availability

The informed consent obtained from ALSPAC participants does not allow the data to be made freely available through any third-party–maintained public repository. However, data used for this submission can be made available on request to the ALSPAC Executive. The ALSPAC data management plan describes in detail the policy regarding data sharing, which is through a system of managed open access. Full instructions for applying for data access can be found on the ALSPAC researchers' access web page [33]. The ALSPAC study website contains details of all the data that are available [23].

Disclaimer

No generative AI tools were used in any portion of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
AlcoWatch microinteraction-based ecological momentary assessment smartwatch question formatting and sequencing.
[PNG File , 173 KB - formative_v9i1e63184_app1.png]

Multimedia Appendix 2
Timeline followback diary format.
[PNG File , 69 KB - formative_v9i1e63184_app2.png]

Multimedia Appendix 3
High- and low-socioeconomic position participants' likes and dislikes of the online timeline followback method.
[PNG File , 46 KB - formative_v9i1e63184_app3.png]

References

1. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008 Apr 01;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
2. Perski O, Keller J, Kale D, Asare BY, Schneider V, Powell D, et al. Understanding health behaviours in context: a systematic review and meta-analysis of ecological momentary assessment studies of five key health behaviours. Health Psychol Rev 2022 Dec;16(4):576-601 [FREE Full text] [doi: 10.1080/17437199.2022.2112258] [Medline: 35975950]

XSL•FO
RenderX

3. Embree BG, Whitehead PC. Validity and reliability of self-reported drinking behavior: dealing with the problem of response bias. J Stud Alcohol 1993 May;54(3):334-344. [doi: 10.15288/jsa.1993.54.334] [Medline: 8487543]
4. Midanik L. The validity of self - reported alcohol consumption and alcohol problems: a literature review. British Journal of Addiction 2010 Feb 12;77(4):357-382. [doi: 10.1111/j.1360-0443.1982.tb02469.x] [Medline: 6762224]
5. Gmel G, Daeppen J. Recall bias for seven-day recall measurement of alcohol consumption among emergency department patients: implications for case-crossover designs. J Stud Alcohol Drugs 2007 Mar;68(2):303-310. [doi: 10.15288/jsad.2007.68.303] [Medline: 17286350]
6. Hammerton G, Lewis G, Heron J, Fernandes G, Hickman M, Lewis G. The association of alcohol dependence and consumption during adolescence with depression in young adulthood, in England: a prospective cohort study. Lancet Psychiatry 2023 Jul;10(7):490-498 [FREE Full text] [doi: 10.1016/S2215-0366(23)00138-4] [Medline: 37271164]
7. Fridberg DJ, Lee Z, Fischer AM, Cursio JF, King AC. Assessing real-time positive subjective effects of alcohol using high-resolution ecological momentary assessment in risky versus light drinkers. Alcohol Clin Exp Res (Hoboken) 2023 Aug;47(8):1570-1581. [doi: 10.1111/acer.15130] [Medline: 37578682]
8. Intille S, Haynes C, Maniar D, Ponnada A, Manjourides J. µEMA: microinteraction-based ecological momentary assessment (EMA) using a smartwatch. Proc ACM Int Conf Ubiquitous Comput 2016 Sep;2016:1124-1128 [FREE Full text] [doi: 10.1145/2971648.2971717] [Medline: 30238088]
9. Beukenhorst AL, Howells K, Cook L, McBeth J, O'Neill TW, Parkes MJ, et al. Engagement and participant experiences with consumer smartwatches for health research: longitudinal, observational feasibility study. JMIR Mhealth Uhealth 2020 Jan 29;8(1):e14368 [FREE Full text] [doi: 10.2196/14368] [Medline: 32012078]
10. Ponnada A, Wang S, Chu D, Do B, Dunton G, Intille S. Intensive longitudinal data collection using microinteraction ecological momentary assessment: pilot and preliminary results. JMIR Form Res 2022 Feb 09;6(2):e32772 [FREE Full text] [doi: 10.2196/32772] [Medline: 35138253]
11. Ottenstein C, Werner L. Compliance in ambulatory assessment studies: investigating study and sample characteristics as predictors. Assessment 2021 Jul 20;29(8):1765-1776. [doi: 10.1177/10731911211032718] [Medline: 34282659]
12. Jones A, Remmerswaal D, Verveer I, Robinson E, Franken IHA, Wen CKF, et al. Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. Addiction 2019 Apr;114(4):609-619 [FREE Full text] [doi: 10.1111/add.14503] [Medline: 30461120]
13. Rintala A, Wampers M, Myin-Germeys I, Viechtbauer W. Momentary predictors of compliance in studies using the experience sampling method. Psychiatry Res 2020 Feb 29;286:112896. [doi: 10.1016/j.psychres.2020.112896] [Medline: 32146247]
14. Office for National Statistics, UK Government. Alcohol-specific deaths in the UK registered in 2018. UK Government. URL: https://www.gov.uk/government/statistics/alcohol-specific-deaths-in-uk-registered-in-2018 [accessed 2025-03-10]
15. Grittner U, Kuntsche S, Gmel G, Bloomfield K. Alcohol consumption and social inequality at the individual and country levels--results from an international study. Eur J Public Health 2013 Apr;23(2):332-339 [FREE Full text] [doi: 10.1093/eurpub/cks044] [Medline: 22562712]
16. Mäkelä P, Paljärvi T. Do consequences of a given pattern of drinking vary by socioeconomic status? A mortality and hospitalisation follow-up for alcohol-related causes of the Finnish Drinking Habits Surveys. J Epidemiol Community Health 2008 Aug;62(8):728-733. [doi: 10.1136/jech.2007.065672] [Medline: 18621959]
17. Mackenbach JP, Kulhánová I, Bopp M, Borrell C, Deboosere P, Kovács K, et al. Inequalities in alcohol-related mortality in 17 European countries: a retrospective analysis of mortality registers. PLoS Med 2015 Dec;12(12):e1001909 [FREE Full text] [doi: 10.1371/journal.pmed.1001909] [Medline: 26625134]
18. Stone A, Schneider S, Smyth JM. Evaluation of pressing issues in ecological momentary assessment. Annu Rev Clin Psychol 2023 May 09;19:107-131 [FREE Full text] [doi: 10.1146/annurev-clinpsy-080921-083128] [Medline: 36475718]
19. Skinner A, Adams S, Stone CJ, Munafo MR, Timpson NJ, Tilling K, et al. Assessing the feasibility of using smartwatch-based ecological momentary assessment techniques for determining longitudinal patterns and contexts of alcohol use in UK adults. Open Science Framework. 2020 Mar 10. URL: https://doi.org/10.17605/OSF.IO/4ZA26 [accessed 2025-03-10]
20. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. Int J Epidemiol 2013 Feb;42(1):111-127 [FREE Full text] [doi: 10.1093/ije/dys064] [Medline: 22507743]
21. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, et al. Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. Int J Epidemiol 2013 Feb;42(1):97-110 [FREE Full text] [doi: 10.1093/ije/dys066] [Medline: 22507742]
22. Northstone K, Lewcock M, Groom A, Boyd A, Macleod J, Timpson N, et al. The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019. Wellcome Open Res 2019;4:51 [FREE Full text] [doi: 10.12688/wellcomeopenres.15132.1] [Medline: 31020050]
23. Avon Longitudinal Study of Parents and Children. University of Bristol. URL: http://www.bris.ac.uk/alspac/researchers/our-data/ [accessed 2025-03-10]

XSL•FO
RenderX

24. Fiellin DA, McGinnis KA, Maisto SA, Justice AC, Bryant K. Measuring alcohol consumption using timeline followback in non-treatment-seeking medical clinic patients with and without HIV infection: 7-, 14-, or 30-day recall. J Stud Alcohol Drugs 2013 May;74(3):500-504 [FREE Full text] [doi: 10.15288/jsad.2013.74.500] [Medline: 23490581]
25. Rueger SY, Trela CJ, Palmeri M, King AC. Self-administered web-based timeline followback procedure for drinking and smoking behaviors in young adults. J Stud Alcohol Drugs 2012 Sep;73(5):829-833 [FREE Full text] [doi: 10.15288/jsad.2012.73.829] [Medline: 22846247]
26. Harris P, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
27. Piasecki TM. Assessment of alcohol use in the natural environment. Alcohol Clin Exp Res 2019 Apr;43(4):564-577 [FREE Full text] [doi: 10.1111/acer.13975] [Medline: 30748019]
28. Howard AL, Lamb M. Compliance trends in a 14-week ecological momentary assessment study of undergraduate alcohol drinkers. Assessment 2024 Mar;31(2):277-290 [FREE Full text] [doi: 10.1177/10731911231159937] [Medline: 36914966]
29. Thern E, Landberg J. Understanding the differential effect of alcohol consumption on the relation between socio-economic position and alcohol-related health problems: results from the Stockholm Public Health Cohort. Addiction 2021 Apr;116(4):799-808 [FREE Full text] [doi: 10.1111/add.15213] [Medline: 32738005]
30. Ponnada A, Li J, Wang S, Wang W, Dunton GF, Intille SS. Contextual biases in microinteraction ecological momentary assessment (µEMA) non-response. Proc ACM Interact Mob Wearable Ubiquitous Technol 2022 Mar;6(1):1-24. [doi: 10.1145/3517259] [Medline: 39866712]
31. Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002;24(3):236-243. [doi: 10.1207/S15324796ABM2403_09] [Medline: 12173681]
32. Liao Y, Skelton K, Dunton G, Bruening M. A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: an adapted STROBE Checklist for Reporting EMA Studies (CREMAS). J Med Internet Res 2016 Jun 21;18(6):e151 [FREE Full text] [doi: 10.2196/jmir.4954] [Medline: 27328833]
33. Process for applying for data access. Avon Longitudinal Study of Parents and Children, University of Bristol. URL: http://www.bris.ac.uk/alspac/researchers/access/ [accessed 2025-03-10]

Abbreviations

ALSPAC: Avon Longitudinal Study of Parents and Children
EMA: ecological momentary assessment
GPS: global positioning system
SEP: socioeconomic position
TLFB: timeline followback
µEMA: microinteraction-based ecological momentary assessment

Edited by A Mavragani; submitted 12.06.24; peer-reviewed by K Kurisu; comments to author 30.10.24; revised version received 12.02.25; accepted 02.03.25; published 25.03.25.

Please cite as:
Stone C, Adams S, Wootton RE, Skinner A
Smartwatch-Based Ecological Momentary Assessment for High-Temporal-Density, Longitudinal Measurement of Alcohol Use (AlcoWatch): Feasibility Evaluation
JMIR Form Res 2025;9:e63184
URL: https://formative.jmir.org/2025/1/e63184
doi:10.2196/63184
PMID:40131326

©Chris Stone, Sally Adams, Robyn E Wootton, Andy Skinner. Originally published in JMIR Formative Research (https://formative.jmir.org), 25.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Using Social Media to Engage and Enroll Underrepresented Populations: Longitudinal Digital Health Research

Christiana Harry[1], MPH; Sarah Goodday[1,2], MSc, PhD; Carol Chapman[3], MPH; Emma Karlin[1], MPH; April Joy Damian[4], MSc, PhD; Alexa Brooks[1], MS, RD; Adrien Boch[5], MA; Nelly Lugo[6], BS; Rebecca McMillan[7], MHIHIM; Jonell Tempero[8], MS; Ella Swanson[1], BS; Shannon Peabody[9], BA; Diane McKenzie[1], MHA; Stephen Friend[1,2], MD, PhD

[1]4YouandMe, Great Neck, NY, United States

[2]Department of Psychiatry, University of Oxford, Oxford, United Kingdom

[3]Crohn's and Colitis Foundation, New York, NY, United States

[4]Moses/Weitzman Health System, Middletown, CT, United States

[5]Evidation Health, San Mateo, CA, United States

[6]Washington University, St. Louis, MO, United States

[7]University of California San Diego Health, San Diego, CA, United States

[8]RespirAI US Inc, Omaha, NE, United States

[9]Dana-Farber Cancer Instituite, Boston, MA, United States

Corresponding Author:
Christiana Harry, MPH
4YouandMe
185 Great Neck Rd Ste 447
Great Neck, NY, 11021
United States
Phone: 1 206 861 3655
Email: christianacharry@gmail.com

Abstract

Background: Emerging digital health research poses roadblocks to the inclusion of historically marginalized populations in research. Exclusion of underresourced communities in digital health research is a result of multiple factors (eg, limited technology access, decreased digital literacy, language barriers, and historical mistrust of research and research institutions). Alternative methods of access and engagement may aid in achieving long-term sustainability of diversified participation in digital health research, ensuring that developed technologies and research outcomes are effective and equitable.

Objective: This study aims to (1) characterize socioeconomic and demographic differences in individuals who enrolled and engaged with different remote, digital, and traditional recruitment methods in a digital health pregnancy study and (2) determine whether social media outreach is an efficient way of recruiting and retaining specific underrepresented populations (URPs) in digital health research.

Methods: The Better Understanding the Metamorphosis of Pregnancy (BUMP) study was used as a case example. This is a prospective, observational, cohort study using digital health technology to increase understanding of pregnancy among 524 women, aged 18-40 years, in the United States. The study used different recruitment strategies: patient portal for genetic testing results, paid/unpaid social media ads, and a community health organization providing care to pregnant women (Moses/Weitzman Health System).

Results: Social media as a recruitment tool to engage URPs in a digital health study was overall effective, with a 23.6% (140/594) enrollment rate of those completing study interest forms across 25 weeks. Community-based partnerships were less successful, however, resulting in 53.3% (57/107) engagement with recruitment material and only 8.8% (5/57) ultimately enrolling in the study. Paid social media ads provided access to and enrollment of a diverse potential participant pool of race- or ethnicity-based URPs in comparison to other digital recruitment channels. Of those that engaged with study materials, paid recruitment had the highest percentage of non-White (non-Hispanic) respondents (85/321, 26.5%), in comparison to unpaid ads (Facebook and Reddit; 37/167, 22.2%). Of the enrolled participants, paid ads also had the highest percentage of non-White (non-Hispanic) participants (14/70, 20%), compared to unpaid ads (8/52, 15.4%) and genetic testing service subscribers (72/384, 18.8%). Recruitment

XSL•FO
RenderX

completed via paid ads (Instagram) had the highest study retention rate (52/70, 74.3%) across outreach methods, whereas recruitment via community-based partnerships had the lowest (2/5, 40%). Retention of non-White (non-Hispanic) participants was low across recruitment methods: paid (8/52, 15.4%), unpaid (3/35, 14.3%), and genetic testing service subscribers (50/281, 17.8%).

Conclusions: Social media recruitment (paid/unpaid) provides access to URPs and facilitates sustained retention similar to other methods, but with varying strengths and weaknesses. URPs showed lower retention rates than their White counterparts across outreach methods. Community-based recruitment showed lower engagement, enrollment, and retention. These findings highlight social media's potential for URP engagement and enrollment, illuminate potential roadblocks of traditional methods, and underscore the need for tailored research to improve URP enrollment and retention.

(JMIR Form Res 2025;9:e68093) doi:10.2196/68093

KEYWORDS

digital health research; digital health technology; recruitment; research subject; participant; pregnancy; maternal health; underrepresented populations; health equity; diversity; marginalized; advertisement; social media; retention; attrition; dropout

Introduction

Background

Participation in research studies by historically racial and ethnic marginalized populations is significantly lower compared to their White counterparts. This has resulted in the underrepresentation of these groups across research study disciplines [1-3]. The National Institutes of Health (NIH) defines underrepresented populations (URPs) as those with disproportionately low representation relative to their overall or disease-specific population [4]. Despite mandates to include these groups in federally funded research, factors such as general attitudes toward research, sociocultural barriers, and accessibility continue to hinder diversity in research populations [1,5-11].

Certain URPs, such as African American populations, face additional historical and structural barriers to research participation including low health literacy, lack of access to care, and mistrust of institutions [2]. For URPs, particularly African American populations, mistrust of academic and research institutions is the largest reported participation barrier due to historical events, such as the Tuskegee syphilis study, and worsened by present inequalities in the health care system [1]. As a direct result of this severe institutional mistrust among URPs generally, lack of engagement and subsequent enrollment of URPs in research studies persists and continues to contribute to health disparities within these specific population subgroups, as they tend to reject participatory outreach from more traditional recruitment methods associated with these mistrusted institutions and present health inequities [1].

These health inequities are especially seen in pregnant populations, as illustrated by present maternal health outcome statistics [12]. Since 2019, the maternal mortality rate in the United States has continued to rise steadily, jumping to 32.9 deaths per 100,000 live births in 2021, compared with a rate of 20.1 in 2019 [12]. For pregnant women of ethnic or racial URPs, this discrepancy in maternal health outcomes is hypervisible. The historical exclusion of pregnant women from research studies has been present due to a perceived concern of lack of direct participant benefit or risk of significant harm to offspring. This exclusion has resulted in gaps in scientific knowledge regarding key women's health issues that further exacerbate women's health outcome disparities [13].

As these disparities and roadblocks to participation continue to be observed, additional efforts are also necessary to identify successful evidence-based enrollment strategies for reaching and engaging URPs, especially in digital research studies. Current engagement and recruitment strategies of URPs often emphasize outreach via community-forward channels (eg, partnerships with community leaders, churches, community-based clinic sites, and other organizations) to alleviate some of the barriers of mistrust present in these communities [10]. However, conflicting evidence exists regarding the effectiveness of community outreach as an isolated strategy for increasing the enrollment of URPs [10].

Another suggested recruitment strategy to bolster the engagement and involvement of URPs in studies is social media advertising. More recently, social media has been used by researchers as a new source of participant recruitment and may be an especially useful tool in locating specific URPs due to the widespread reach of social media to potential participants who otherwise might have been not visible to researchers, the anonymity that social media provides users, and general mistrust of health care institutions and more traditional recruitment methods by URPs [2,14,15].

Over the past decade, social media use in the United States has steadily increased. As of 2022, over 300 million people in the United States use social media, making the United States home to the third largest social media audience globally [16,17]. The social media platforms Facebook (Meta), Twitter (X Corp), and Instagram (Meta) account for a combined 83% of all social media site visits in the United States [16]. As nearly 3 quarters of Americans report social media use, social media platforms now provide researchers a new, accessible alternative to traditional recruitment methods (eg, printed flyers and in-person outreach) to meet individuals where they are and help bridge the digital divide [18].

The digital divide refers to the patterns of difference in the utilization of technology, the internet, and social media by race, ethnicity, or socioeconomic status. The underlying drivers of the digital divide are manifold but include 2 key factors: differential access to and engagement (motivation) in the use

XSL•FO
RenderX

of digital health technologies (DHTs; smartphones, wearable devices or other digital products, electronic health records, and telehealth options). The use of social media or DHTs in digital research studies requires ownership of devices (eg, smartphones and wearable devices), adequate internet access, and a level of digital and health literacy, indicating the presence of additional cost and other barriers to digital health engagement and use of DHTs [19].

The other key driver of the digital divide is the disconnect in engagement, especially among URPs. Although social media usage for racially based URPs continues to surpass that of their White counterparts, these populations are significantly less likely to engage with health information on the internet and social media or use social networking channels to learn more about their health [20-22]. However, studies indicate expressed interest by specific URPs (eg, African American women) in using DHTs and social media to engage with health information [23-25].

As the use of DHTs in digital research studies continues to rapidly expand, new and innovative ways in which these technologies may be used to support individual and population health research also evolve [26,27]. DHTs and digital research studies present the potential to mitigate existing gaps in health care access, quality of care, and health care outcomes, further suggesting that social media may be a powerful, untapped resource in reaching and engaging URPs in DHT research studies and digital health [22].

Previous Work

Existing research studies have primarily used Facebook as a means of recruiting URPs, especially for stigmatized groups, but few research studies explore the feasibility and effectiveness of using other social media platforms and recruitment methods specifically to engage URPs in digital health studies [28]. The use of more nuanced social media outreach methods—such as advertisements on Instagram (a platform that has nearly 127.2 million active users in the United States as of 2023)—to recruit and engage URPs in research studies is still in its infancy [17].

Goal of This Paper

This paper aims to (1) characterize the socioeconomic and demographic differences among individuals who enrolled and participated in a US-based pregnancy study—the Better Understanding the Metamorphosis of Pregnancy (BUMP) study—through various remote, digital, and traditional recruitment methods and (2) to determine whether social media outreach is an effective means of recruiting and retaining historically URPs to participate in pregnancy-related digital health research.

Methods

Overview

The BUMP study is a prospective, observational, cohort study conducted by 4YouandMe, a nonprofit organization that pilots open-source, digital health research. The BUMP study aims to leverage DHTs in the collection of objective and subjective measurements of health to increase understanding of pregnancy and subsequent complications in a sample of women, aged 18-40 years, in the United States (n=524) [29]. The BUMP study sample was primarily White non-Hispanic (381/524, 72.7%) and between 26-35 years of age (363/524, 69.3%). The first participant was enrolled in the BUMP study on February 23, 2021, and the last participant completed the study on July 1, 2023. Participants were enrolled over a 6- to 12-month period and followed for up to 16 months.

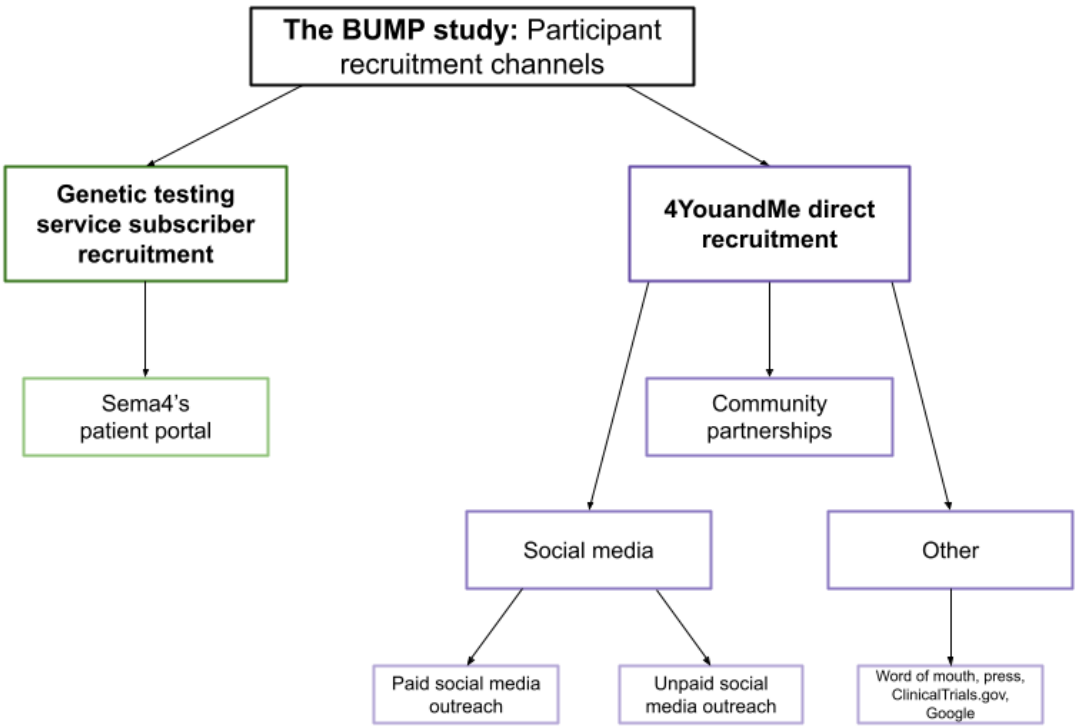
The BUMP study used various recruitment methods, including a patient portal for genetic testing results during pregnancy, social media platforms like Facebook, Reddit, and Instagram through both paid and unpaid advertisements; in-person clinics at a trusted community health center offering prenatal and postnatal care; as well as other outreach strategies such as word of mouth and participant referrals. Data were collected through surveys and active tasks on a study smartphone app, wearable DHTs (eg, a Garmin smartwatch, an Oura smart ring, and a Bodyport smart scale), and patients' electronic medical records.

Recruitment

Recruitment for the BUMP study was primarily conducted via 2 methods: genetic testing service subscriber recruitment (Sema4) and 4YouandMe direct recruitment. Genetic testing service subscriber recruitment includes outreach to subscribers to genetic testing via Sema4's patient portal. 4YouandMe direct recruitment consists of 3 recruitment suboutlets: social media (including both paid and unpaid outreach), community partnerships, and others (eg, word of mouth, press, and ClinicalTrials.gov; Figure 1).

XSL•FO
RenderX

Figure 1. Flowchart of recruitment channels for the Better Understanding the Metamorphosis of Pregnancy (BUMP) study.



Genetic Testing Service Subscriber Recruitment (Sema4)

Primary recruitment for the BUMP study was completed via Sema4's patient portal, specifically focusing on those who signed up to receive results of genetic testing results during pregnancy. Sema4's patient portal combines mobile health and eHealth as it provides a mechanism for Sema4 to engage with its patients via communication of test results and management of electronic consents [29]. Patients who consented to having their electronic medical record data accessed by Sema4, and those requesting results of genetic testing or other tests commonly associated with pregnancy, received an email from Sema4 with information about the BUMP study, including contact information for study staff if further interested in participation [29].

4YouandMe Direct Recruitment (Social Media Recruitment, Community Partnerships, and Others)

All recruitment conducted by methods other than outreach to genetic testing service subscribers via Sema4's patient portal of genetic testing subscribers are referred to as 4YouandMe direct recruitment. 4YouandMe direct recruitment was conducted primarily through social media platforms (eg, Facebook, Reddit, and Instagram).

Unpaid Social Media Outreach (Social Media Recruitment)

Social media sites with moderated group forums related to pregnancy (eg, Facebook and Reddit) were the focus of unpaid social media recruitment efforts to reach the wider population of all study-eligible pregnant women for the BUMP study. The BUMP study staff conducted outreach efforts with forum and group moderators to gain permission to post fliers on these forums. The BUMP study staff monitored public response to the live group and forum posts and interacted with moderators and interested individuals by answering questions and engaging with posted content. The goal was to build trust and credibility within these online communities while mitigating some of the noted common challenges faced in social media–based recruitment.

Text study descriptions and image-heavy digital recruitment fliers were used to reach potential participants and posted as both standalone and response posts in various pregnancy-related online forums (eg, "Pregnant Women's Group" and "Mom's Talk: Pregnancy to Toddler" on Facebook; and "Science Based Parenting," "Pregnant," "Mommit," and monthly "BUMP Groups," ie, "January 2023 BUMP Group" on Reddit; Multimedia Appendices 1-4).

Paid Social Media Outreach (Social Media Recruitment)

A digital recruitment flier was posted on 4YouandMe's personal Instagram profile, and its visibility was boosted via Instagram's paid promotion feature, which allows you to select a "goal" (ie, driving traffic to a specific website). The total spend for paid social media recruitment was US $1939.30 over several paid Instagram advertisements.

Interested participants in the BUMP study were able to click on the advertisement and were then guided directly to complete the study interest form, hosted on the secure electronic data capture platform, REDCap (Research Electronic Data Capture; Vanderbilt University). The advertisement parameters for the paid social media outreach method included women, aged 18-42 years, currently living in the United States; and women, aged

XSL•FO
RenderX

18-36 years, living in the lowest 10% of United States zip codes by per capita income.

Community-Based Recruitment or Outreach (Community Partnerships)

Community-based partnerships were formed with the Moses/Weitzman Health System, a large federally qualified health center that works to deliver comprehensive health care services to those in need throughout the state of Connecticut. Partnership costs were covered by an agreed-upon fee for collaboration between Moses/Weitzman Health System and 4YouandMe. Medical assistants at three locations (Meriden, Middletown, and Clinton) that provide prenatal and postnatal care were educated by the BUMP study staff on the study overview, goals, eligibility requirements, wearable DHT use, and participant expectations. Moses/Weitzman Health System staff members were also provided with printed recruitment materials for the BUMP study.

Eligible potential participants were informed about the BUMP study by Moses/Weitzman Health System staff during one of their initial prenatal visits. Potential participants who expressed interest in the BUMP study were given a printed flier, tailored to the outreach of this specific community, with a scannable QR code linking interested participants to the study interest form, as well as contact information for study staff available to answer any questions before providing any personal information (Multimedia Appendices 5-6).

Other Recruitment Sources (Other)

Additional recruitment for the 4YouandMe direct outreach method was completed by other avenues such as referrals from current BUMP study participants to friends, family, and colleagues; general word of mouth about the BUMP study; articles published about 4YouandMe and the BUMP study on the Oura Ring blog and in STAT Magazine; and other sources such as Google and ClinicalTrials.gov.

Statistical Analysis

Engagement, for the purpose of this analysis, is defined as any interaction by a potential participant with the BUMP study material such as clicking on an advertisement or filling out a study interest form. Engagement with study advertisements from social media sites was tracked by Google Analytics, and REDCap was used to monitor completed study interest forms and collect additional information regarding recruitment channels. Due to the nature of recruitment conducted of genetic testing service subscribers via Sema4's patient portal, engagement was unable to be calculated for this group. Enrollment is defined as any participant who signed the informed consent and officially enrolled into the study, for any duration of time. In contrast, retention refers to participants who completed the full study period through birth or at least 9 months.

Engagement, enrollment, and retention rates were stratified by different recruitment channels and race. Sociodemographic characteristics including age, race or ethnicity, and income per capita (US $) by zip code were compared across different recruitment channels. Enrolled participants were also asked to share their home zip code upon enrollment. Participant's zip codes were used to calculate the distribution of per capita income of enrolled study participants, stratified by their recruitment method. Income data were provided by the United States Census Bureau's 2020 American Community Survey [30]. All analyses were conducted using Microsoft Excel (version 16.0).

Ethical Considerations

Human subjects research for the BUMP study was conducted in accordance with all applicable ethical guidelines and received full approval from the institutional review board, Advarra (Pro00047893). Participant confidentiality was maintained via deidentified data, secure data storage, limiting access of identifiable information to authorized members of the research team, and following clearly outlined privacy measures in the electronic consent (eConsent) document.

The BUMP study app served as the participants' hub for all data collection and study-related communication. Participants were instructed to download the app from the App Store (Apple Inc) or Google Play Store (Android). After downloading the study app, participants were walked through the study onboarding process, beginning with completing an eConsent form. ResearchKit's eConsent framework was used (for iOS) that includes brief summary screens with a link out to a "learn more" section to convey the consent narrative in a more digestible way. Participants were then provided with the eConsent in its entirety for their review and given a brief comprehension quiz before digitally signing the eConsent form. Participants immediately receive a digital, signed copy of the eConsent to their provided email address [29].

BUMP study participants were financially compensated in accordance with their engagement in daily surveys and active task completions on the BUMP study app. Participants received daily reward points that were tallied at the end of the study, and financial compensation was awarded proportionate to their accumulated study points for task completion. Participants were able to earn up to US $25 per month of participation in the BUMP study, for up to 12 months (including the 9-month study period and 3-month postpartum observation period). Participants who completed the study were also offered to keep their wearable devices—the Oura smart ring and Garmin smartwatch.

Further details of the BUMP study are described in external publications [29].

Results

Engagement, Enrollment, and Retention

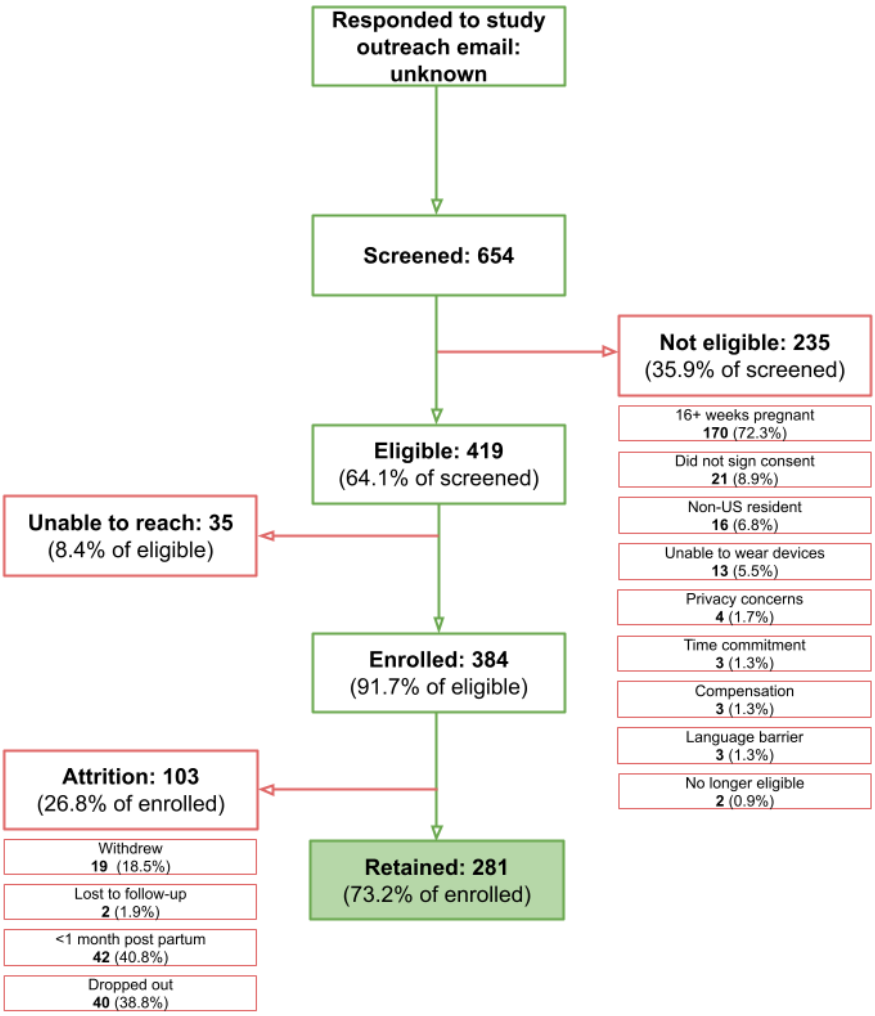
Across all recruitment channels, 524 pregnant participants were enrolled in the BUMP study: 379 (72.3%) completed the full study period, 384 (73.3%) participants were recruited via Sema4's portal for genetic testing service subscribers, and 140 (26.7%) participants were recruited to 4YouandMe directly from the other recruitment approaches. Participants recruited via Sema4's patient portal (genetic testing service subscribers) had similar rates of study retention as those recruited via unpaid social media (Reddit and Facebook) and paid social media

XSL•FO
RenderX

(Instagram). The Moses/Weitzman Health System had the lowest percentage of study retention.

Genetic testing service subscriber recruitment was conducted via Sema4's patient portal for genetic testing service subscribers. Engagement data for this group is unknown. Of those who responded to an initial email with information about the BUMP study, 654 participants completed a screening call with a study coordinator. Of this group, 64.1% (419/654) met eligibility criteria, and 91.7% (384/654) of eligible participants enrolled in the study, resulting in 73.2% (281/654) of participants retained for the full study. (Figure 2).

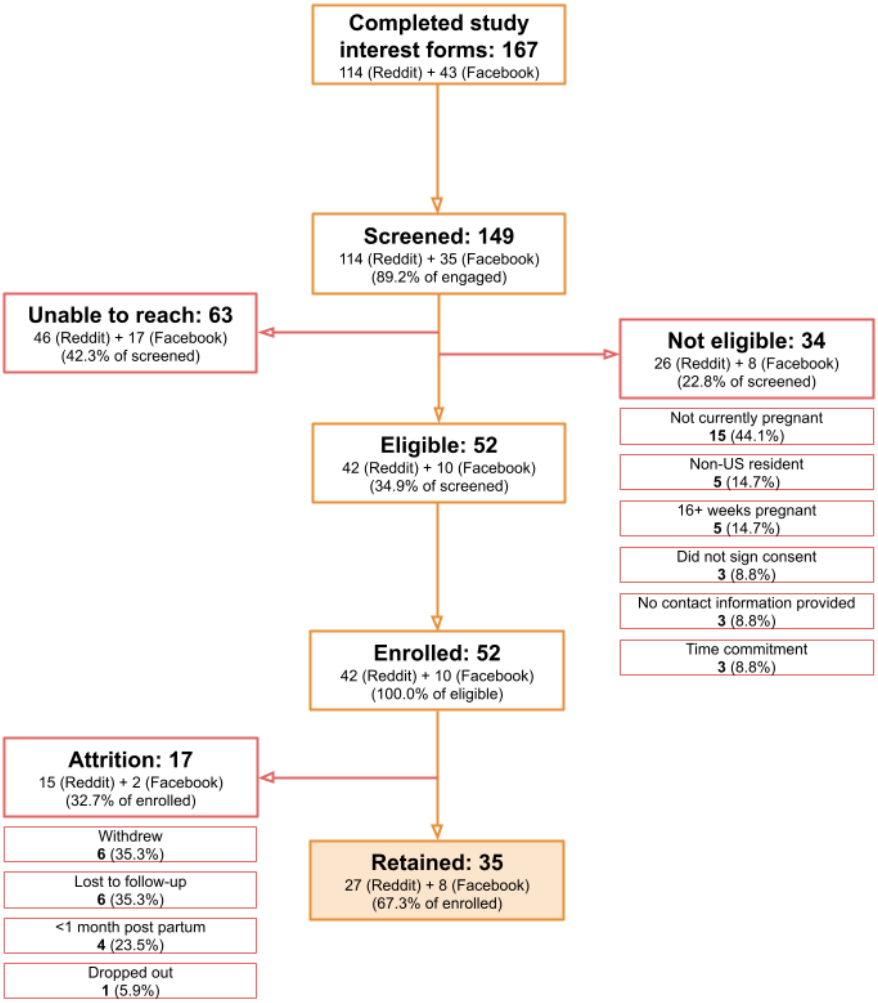
Figure 2. Engagement, enrollment, and retention for the Better Understanding the Metamorphosis of Pregnancy (BUMP) study from outreach via genetic testing service subscribers (Sema4).



Social media recruitment of participants was conducted via unpaid advertisements (Reddit and Facebook) and paid advertisements (Instagram). Of those who engaged with an unpaid ad posted on Reddit and Facebook, 74.9% (167/223) completed a study interest form. Of this group, 31.1% (52/167) met eligibility criteria, and 100% (52/52) of eligible participants enrolled in the study, resulting in 67.3% (35/52) retained by the study's end (Figure 3).

XSL•FO
RenderX

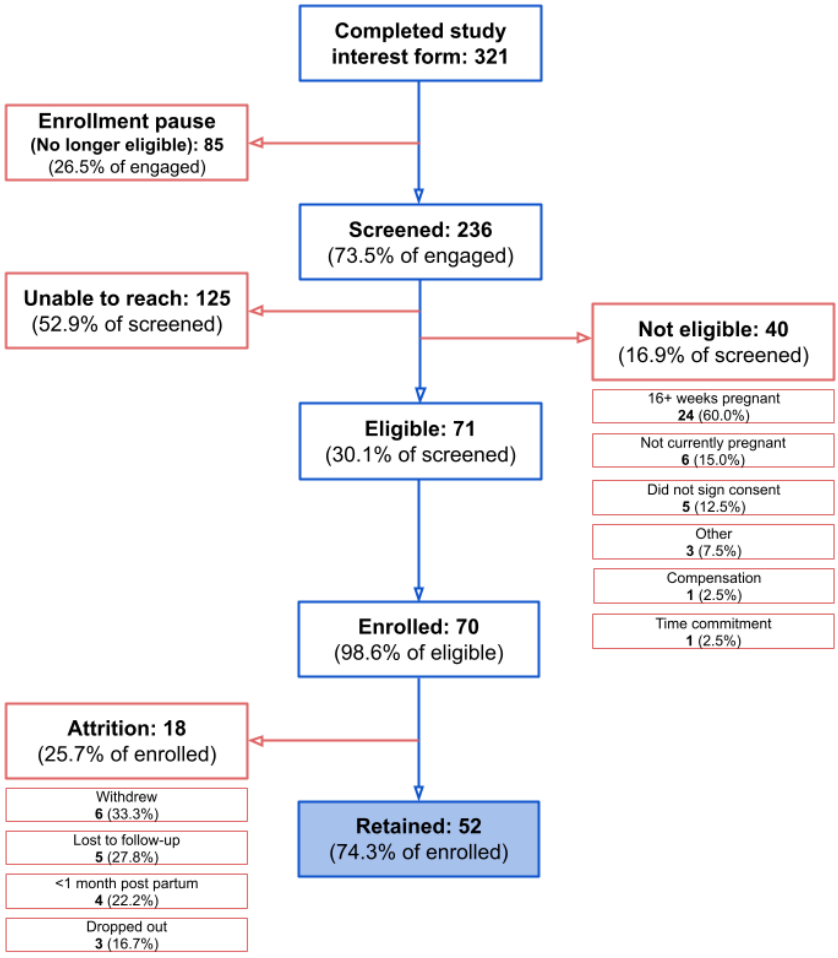
Figure 3. Engagement, enrollment, and retention for the Better Understanding the Metamorphosis of Pregnancy (BUMP) study from outreach via unpaid social media recruitment (Reddit and Facebook).



For individuals recruited from paid social media advertisements on Instagram, 321 individuals completed study interest forms; of those, 85 (26.5%) were not eligible for enrollment as a result of a studywide enrollment pause due to high recruitment volume. These individuals were notified of the enrollment pause but were not recontacted for further enrollment. Of the remaining 71 eligible participants, 70 (98.6%) enrolled in the study, and 74.3% (52/70) of those enrolled were retained (Figure 4).

XSL•FO
RenderX

Figure 4. Engagement, enrollment, and retention for the Better Understanding the Metamorphosis of Pregnancy (BUMP) study from outreach via paid social media recruitment (Instagram).



Of participants recruited via in-person visits with their clinician at a community health center providing affordable prenatal care (Moses/Weitzman Health System), 107 met eligibility criteria and were able to be approached on-site by staff. Of those approached, 24.3% (26/107) engaged with study material by completing the study interest form. A total of 76.9% (20/26) of participants did not enroll for a variety of reasons, such as being unable to contact (n=10), having more than 16 weeks pregnant (n=4), language barrier (n=3: n=2 Spanish and n=1 Haitian Creole), having no personal smartphone (n=1), not currently pregnant (n=1), or did not sign the consent (n=1). In total, 1 participant was no longer eligible due to the previously mentioned studywide enrollment pause, 5 participants ultimately enrolled in the study, and 40% (2/5) were retained—the lowest retention rate of any used recruitment method (Figure 5).

Over a 25-week period of active advertisements on both paid and unpaid social media channels (excluding community partnerships and other methods), 491 study interest forms were completed (the total from all 4YouandMe direct recruitment outlets was 594). This resulted in 122 enrolled participants (the total from 4YouandMe direct recruitment was 140) via paid and unpaid social media recruitment into the BUMP study (Table 1).

XSL•FO
RenderX

Figure 5. Engagement, enrollment, and retention for the Better Understanding the Metamorphosis of Pregnancy (BUMP) study from outreach via the Moses/Weitzman Health System.

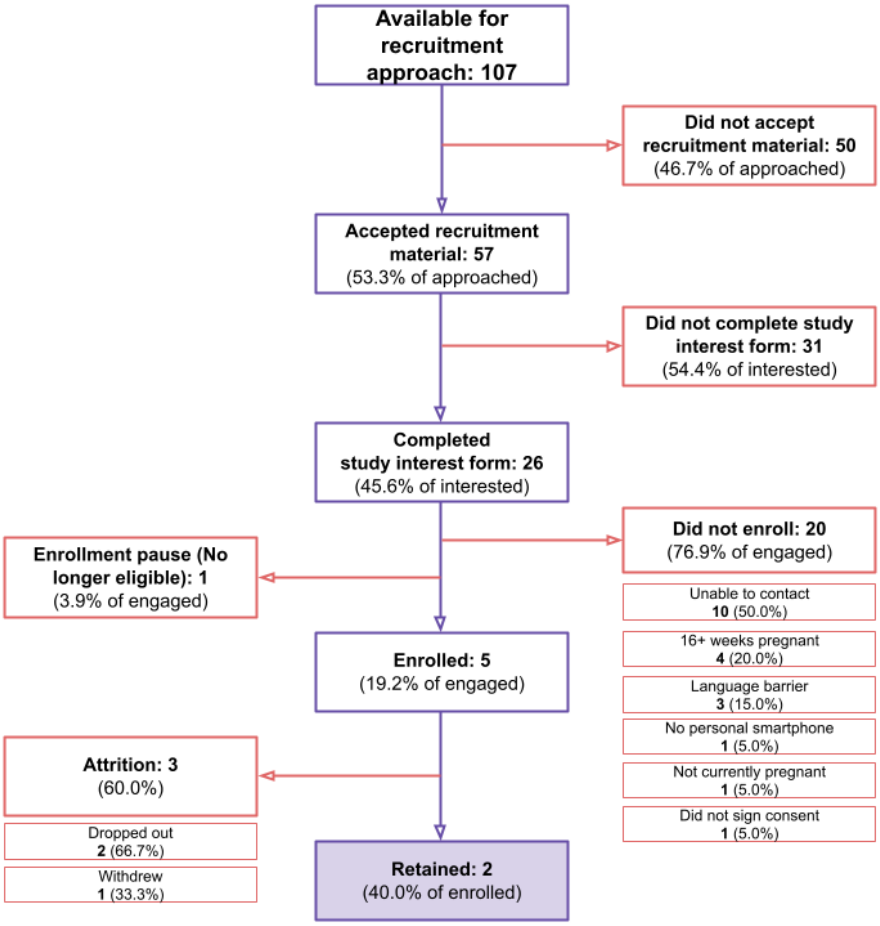


Table 1. Engagement and enrollment characteristics across different social media channels.

	Unpaid advertisements (Reddit)	Unpaid advertisements (Facebook)	Paid advertisements (Instagram)
Ad reach, n	—[a]	—	97,475
Website or advertisement taps, n	197	26	1534
Completed study interest forms, n	124	43	324
Participants enrolled, n	42	10	70
Start and end dates	April 5, 2022, to September 8, 2022	May 18, 2022, to September 8, 2022	September 2, 2022, to October 19, 2022

[a]Not applicable.

Characteristics of the Cohort

The majority of enrolled participants in the BUMP study were between the ages of 26-35 years (363/524, 69.3%) and White non-Hispanic (381/524, 72.7%; Table 2). For those that engaged with materials via 4YouandMe direct recruitment, participants were asked “Where did you hear about the BUMP study?” and were also given the option to self-identify both their race and ethnicity. Of participants that responded via each recruitment outlet (Reddit: n=124, Moses/Weitzman Health System: n=26, and Instagram: n=321), Reddit had the highest percentage of White non-Hispanic potential participants (98/124, 79%), while the Moses/Weitzman Health System had the lowest percentage of White non-Hispanic potential participants (7/26, 26.9%) and the highest relative percentage of Black potential participants (5/26, 19.2%), followed by Instagram (32/321, 9.9%; Table 3).

XSL•FO
RenderX

Table 2. Characteristics of the overall cohort among those who enrolled and were retained in the BUMP[a] study.

Characteristics	Genetic testing service subscriber recruitment (Sema4)		4YouandMe direct recruitment (social media, community partnerships, and other)		The BUMP study (all recruitment methods)	
	Enrolled (n=384)	Retained (n=281)	Enrolled (n=140)	Retained (n=98)	Enrolled (n=524)	Retained (n=379)
Age group (years), n (%)						
18-25	18 (4.7)	10 (3.6)	3 (2.1)	3 (3.1)	21 (4)	13 (3.4)
26-35	251 (65.4)	187 (66.6)	112 (80)	77 (78.6)	363 (69.3)	264 (69.7)
36-45	115 (29.9)	84 (29.9)	25 (21.7)	18 (18.4)	140 (26.7)	102 (26.9)
Race, n (%)						
Black	15 (3.9)	10 (3.6)	6 (4.3)	4 (4.1)	21 (4)	14 (3.7)
White (non-Hispanic)	274 (71.4)	213 (75.8)	107 (76.4)	77 (78.6)	381 (72.7)	290 (76.5)
White (Hispanic)	4 (1)	1 (0.4)	5 (3.6)	3 (3.1)	9 (1.7)	4 (1.1)
Asian	35 (9.1)	26 (9.3)	8 (5.7)	5 (5.1)	43 (8.21)	31 (8.2)
More than one race	18 (4.7)	13 (4.6)	7 (5)	4 (4.1)	25 (4.8)	17 (4.5)
Unknown or not reported	38 (9.9)	18 (6.4)	7 (5)	5 (5.1)	45 (8.6)	23 (6.1)
Ethnicity, n (%)						
Hispanic or Latino	12 (3.1)	6 (2.1)	10 (7.1)	6 (6.1)	22 (4.2)	12 (3.2)
Not Hispanic or Latino	372 (96.9)	275 (97.9)	128 (91.4)	91 (92.9)	500 (95.4)	366 (96.6)
Unknown or not reported	0 (0)	0 (0)	2 (1.4)	1 (1)	2 (0.4)	1 (0.3)

[a]BUMP: Better Understanding the Metamorphosis of Pregnancy.

XSL•FO
RenderX

Table 3. Characteristics of the cohort among those who engaged with recruitment materials, enrolled, and were retained in the BUMP[a] study by recruitment method[b].

Characteristics	Genetic testing service subscriber recruitment		4YouandMe direct recruitment											
	Sema4		Instagram			Facebook			Reddit			Moses/Weitzman Health System		
	Enrolled (n=384)	Retained (n=281)	Engaged (n=321)	Enrolled (n=70)	Retained (n=52)	Engaged (n=43)	Enrolled (n=10)	Retained (n=8)	Engaged (n=124)	Enrolled (n=42)	Retained (n=27)	Engaged (n=26)	Enrolled (n=5)	Retained (n=2)
Age group (years), n (%)														
18-25	18 (4.7)	10 (3.6)	40 (12.5)	1 (1.4)	1 (1.9)	4 (9.3)	1 (10)	1 (12.5)	5 (4)	1 (2.4)	1 (3.7)	8 (30.8)	0 (0)	0 (0)
26-35	251 (65.4)	187 (66.6)	224 (69.8)	56 (80)	42 (80.8)	35 (81.4)	7 (70)	6 (75)	102 (82.3)	34 (80.9)	20 (74)	16 (61.5)	4 (80)	2 (100)
36-45	115 (29.9)	84 (29.9)	57 (17.8)	13 (18.6)	9 (17.3)	4 (9.3)	2 (20)	1 (12.5)	17 (13.7)	7 (16.7)	6 (22.2)	2 (7.7)	1 (20)	0 (0)
Race, n (%)														
Black	15 (3.9)	10 (66.7)	32 (9.9)	2 (2.9)	1 (50)	4 (9.3)	0 (0)	0 (0)	4 (3.2)	1 (2.4)	1 (3.7)	5 (19.2)	1 (20)	1 (50)
White (non-Hispanic)	274 (71.4)	213 (77.7)	215 (66.9)	52 (74.3)	40 (76.9)	28 (65.1)	9 (90)	7 (87.5)	98 (79)	34 (80.9)	22 (81.5)	7 (26.9)	2 (40)	1 (50)
White (Hispanic)	4 (1)	1 (25)	18 (5.6)	3 (4.3)	2 (66.7)	2 (4.7)	0 (0)	0 (0)	5 (4)	0 (0)	0 (0)	5 (19.2)	1 (20)	0 (0)
Asian	35 (9.1)	26 (74.3)	16 (4.9)	4 (4.29)	2 (66.7)	1 (2.3)	0 (0)	0 (0)	9 (7.3)	5 (11.9)	3 (11.1)	0 (0)	0 (0)	0 (0)
American Indian or Alaska Native	0 (0)	0 (0)	1 (0.3)	0 (0)	0 (0)	1 (2.3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (3.9)	0 (0)	0 (0)
Native Hawaiian or Other Pacific Islander	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (2.3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
More than one race	18 (4.7)	13 (72.2)	18 (5.6)	5 (7.1)	3 (60)	4 (9.3)	0 (0)	0 (0)	6 (4.8)	2 (4.8)	1 (3.7)	2 (7.7)	0 (0)	0 (0)
Unknown or not reported	38 (9.9)	18 (47.4)	21 (6.5)	5 (7.1)	4 (80)	2 (4.7)	1 (10)	1 (12.5)	2 (1.6)	0 (0)	0 (0)	6 (23)	1 (20)	0 (0)
Ethnicity, n (%)														
Hispanic or Latino	12 (3.1)	6 (2.1)	36 (11.2)	5 (7.1)	4 (7.7)	4 (9.3)	0 (0)	0 (0)	7 (5.7)	0 (0)	0 (0)	13 (50)	3 (60)	1 (50)
Not Hispanic or Latino	372 (96.9)	275 (97.9)	276 (85.9)	63 (90)	47 (90.8)	39 (90.7)	10 (100)	8 (100)	114 (91.9)	42 (100)	27 (100)	13 (50)	2 (40)	1 (50)
Unknown or not reported	0 (0)	0 (0)	9 (2.8)	2 (2.9)	1 (1.9)	0 (0)	0 (0)	0 (0)	3 (2.4)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

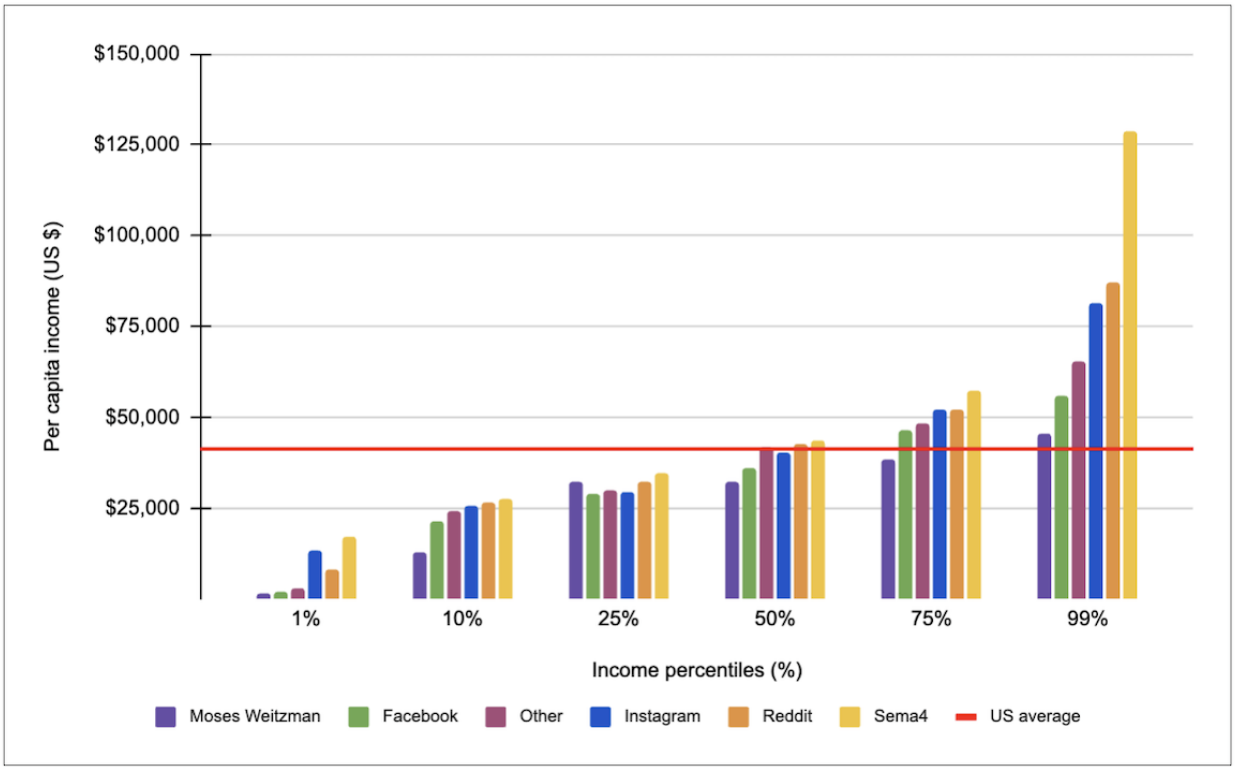
[a]BUMP: Better Understanding the Metamorphosis of Pregnancy.

[b]Values for participants recruited by other methods are included in Multimedia Appendix 7.

Those recruited via Sema4 (genetic testing service subscribers) had the highest participant income per capita by zip code at nearly every income percentile (Figure 6). Participants recruited via community outreach partnerships (Moses/Weitzman Health System) had the lowest number of participants in the 50-99th percentiles of per capita income by zip code, studywide (Figure 6).

XSL•FO
RenderX

Figure 6. Distribution of income per capita (US $) by zip code of enrolled participants in the Better Understanding the Metamorphosis of Pregnancy (BUMP) study by recruitment method (genetic testing subscriber recruitment: n=384 and 4YouandMe direct recruitment: n=140).



Discussion

Principal Results

The results of this analysis suggest that the use of social media (paid and unpaid advertisements) as a tool to engage, enroll, and retain URPs in digital health pregnancy research can be effective. Unpaid social media proved to be successful and low-cost, and achieved the highest engagement-to-enrollment rate among all approaches used by the 4YouandMe direct recruitment channel. However, maintaining successful recruitment on these platforms required significant sustained effort and time from the study team.

Paid social media ads on Instagram also effectively reached and engaged pregnant women as URP in digital health research. Running paid advertisements was a low burden on the study team, in comparison to unpaid advertisements, which required continued monitoring and engagement on behalf of the study team. Paid advertisements incurred a cost over 25 weeks; US $1939.30 was spent on Instagram advertisements, resulting in 70 enrollments, or US $27.70 per enrolled participant. This cost per enrolled participant is generally favorable, as previous systematic reviews of the costs of social media–based participant recruitment ranged from US $0.00 to US $517.00 per participant, compared to US $19.00 to US $777.00 for traditional recruitment methods [31]. Additional studies found a mean cost per participant of US $16.22 for Google ads and between US $13.12 and US $250.00 for other traditional methods. Data on the cost-effectiveness of paid social media ads for recruiting participants to research studies are severely limited [32].

Participants recruited via paid social media (Instagram) showed the highest overall retention rate as 74.3% (52/70) of participants enrolled completed the full study period (Table 3). This suggests that paid social media outreach was more effective in engaging and retaining study participants than other recruitment methods. This outreach method was even more effective than recruiting through genetic testing service subscribers (Sema4), where participants may have had a higher intrinsic motivation to participate and stay in the study due to potential previous interest in genetic testing and related research.

Interactions with participants recruited via Instagram during study onboarding revealed a distinct motivation to engage, enroll, and complete the full study period. Although not statistically quantified, this trend may be linked to specific user behavior within the Instagram-recruited cohort, or the platform's age demographics aligning closely with the study's target population. Among participants retained in the BUMP study recruited from Instagram, 80.8% (42/52) were aged 26-35 years, compared to 74% (20/27) from Reddit (Table 3). This aligns with broader trends, as 28.3% of the US Instagram users in 2024 were aged 25-34 years, while 44% of Reddit users in 2023 were aged 18-29 years [33,34]. Given that the mean age of first-time mothers in the United States is 27.4 years, Instagram's higher retention may reflect its unique demographic fit with the specific study population and subsequent increased study interest and engagement [35]. These observations remain speculative, as the available data cannot fully confirm the underlying reasons.

In addition, paid social media recruitment facilitated the widespread engagement and enrollment of URPs. Paid recruitment had a higher proportion of non-White, non-Hispanic

XSL•FO
RenderX

respondents who engaged with and enrolled in the BUMP study, in comparison to unpaid social media. Retention of non-White, non-Hispanic participants was low across all recruitment methods.

Recruitment via social media (including paid and unpaid recruitment) also provided access to and enrollment of a different socioeconomic population than those enrolled via genetic testing service subscriber (Sema4) recruitment. Participants recruited via genetic testing service subscribers (Sema4), which is associated with Mount Sinai Health System, reflect a unique group of individuals with means and access to genetic testing. This was shown in the enrolled participants from this group who had significantly higher per capita incomes, exceeding US $75,000 far surpassing the US national average income per capita [36].

In contrast, enrolled participants from social media recruitment and community-based outreach's highest average per capita incomes were US $74,771.15 and US $45,401.00, respectively (Figure 6). While data are limited, one study suggests that 46% of Instagram users and 23% of Reddit users have a reported household income under US $70,000.00 [37]. From this, we may infer that recruitment via social media (and community partnerships) also facilitated the increased enrollment of socioeconomic-based URPs.

Community-based partnerships for BUMP study recruitment revealed unique and informative data trends. Over a 24-week period at three Moses/Weitzman Health System clinics that serve a large URP, only 53.3% (57/107) accepted recruitment materials (Figure 5). The overall number of potential in-clinic approaches was limited by multiple factors. Time constraints with in-clinic visits and frequent no-shows hindered clinic staff from fully introducing the study to potential participants, thoroughly answering questions, or handing out recruitment material—further illustrating additional downsides of more traditional, in-person recruitment.

In addition, a significant number of potential participants were immediately excluded before the recruitment approach due to a Spanish language barrier, curtailing the overall number of eligible approaches. Of those remaining participants eligible for approach, some displayed immediate apprehension to study engagement upon simply hearing the word "study," and one participant explicitly reported "not wanting to feel like a guinea pig" to clinic staff." These responses show an intrinsic hesitancy to engage in research studies and a deep mistrust of research overall, even when the information is conveyed by familiar health care providers or their staff.

Community health center–recruited participants who indicated study interest also showcased population-specific barriers to study enrollment or participation. Despite accepting recruitment material, 15% (3/20) of participants did not express sufficient English language fluency to complete surveys and tasks entirely in English on the study app. A total of 50% (10/20) of participants who engaged with study materials were difficult to reach for a further onboarding call (Figure 5). While many of these participants were responsive via text message, BUMP study staff reported difficulty completing study onboarding via phone with participants, citing their extensive and limiting work schedules. Other significant, population-specific barriers to study enrollment included those who did not own a personal smartphone.

Despite dedicated efforts from the BUMP study team and Moses/Weitzman Health System staff, recruiting, engaging, and enrolling participants from community health centers proved uniquely challenging compared to other methods. Of the potential participants who received recruitment materials, only 8.8% (5/57) ultimately enrolled in the study (Figure 5). This low enrollment underscores many of the well-documented challenges of maintaining consistent enrollment among URPs when relying solely on community outreach channels. It also highlights opportunities for overcoming these challenges in future studies.

Strategies to address logistical and participant-voiced barriers may include conducting digital recruitment follow-ups with no-show potential participants, enhancing language inclusivity in recruitment and study app materials, incorporating text messaging for primary communication, providing smartphones for study participant use, and embedding opportunities for recruitment outreach into clinical care. Seamless communication between researchers, providers, and medical staff can also aid in addressing participants' needs in real time as researchers work with clinic staff, using feedback from potential participants, to address population-specific concerns and develop tailored language and approaches; combating mistrust head-on by leading with transparency; and creating a culture of embedding options for research participation into clinical care.

Limitations

Due to the methodology of recruitment and outreach for the BUMP study via genetic testing service subscribers (Sema4), overall study interest and engagement were unable to be followed and analyzed in the same manner as 4YouandMe direct recruitment avenues. Similarly, demographic information (eg, race or ethnicity, socioeconomic status) was not provided for interested or potential participants from this group, limiting the comparison and further analysis of engagement with study materials by various URPs between recruitment of genetic testing service subscribers and social media and other methods.

Additional limitations were present in the analysis of the recruitment potential for the general pregnant population via social media platforms. We were unable to discern how many of the total reached population was actually pregnant or study eligible; this was particularly relevant for paid social media outreach and recruitment. For paid social media outreach, we relied on inputting the lowest 10% of US zip codes by per capita income to increase our socioeconomic diversity; however, it is unknown if zip codes are too narrow of a parameter for focused Instagram advertisements, ultimately impacting an in-depth review of engagement statistics.

True analysis of study engagement to enrollment statistics was further limited by the presence of a study enrollment pause, due to the high volume of completed study interest forms and potential participants. This enrollment pause rendered 132 participants no longer eligible for study participation. Without this pause, the engagement-to-enrollment rate may have reached

XSL•FO
RenderX

45.8% (272/594). Recruitment for the BUMP study via the various 4YouandMe direct approaches was not conducted for the entire study duration, impacting the overall ability to assess engagement to enrollment effectiveness. Finally, the small size of this study sample (n=524) indicates the need for larger, similar studies to further validate the trends observed.

Conclusions

Social media and virtual outreach are both emerging and exciting tools that researchers can use to engage and recruit various URPs to participate in digital health research studies, specifically about pregnancy. Paid social media advertisements provide unique, innovative, and low-time-burden opportunities for researchers to engage and interact with a larger volume of URPs in comparison to traditional recruitment methods.

Though commonly perceived as the gold standard for outreach to URPs, community-based partnerships, even when combined with other outreach methods, did not provide a smooth, consistent path to outreach, engagement, or retention of URPs in digital health research. This lower engagement and retention underscore the need for further tailored strategies of addressing additional barriers to sustained participation in digital health studies for URPs to better bridge the digital divide and engage URPs.

By embracing novel recruitment outreach methods such as via social media platforms, researchers have a unique opportunity to refine often stagnant, traditional recruitment strategies and adapt to an evolving digital landscape. The integration of community-driven recruitment approaches with tailored digital interventions that uniquely address the needs of URPs will create more inclusive and equitable opportunities for participation in digital health studies and foster more representative and impactful research that may ultimately improve health outcomes for URPs and others alike.

Acknowledgments

The Better Understanding the Metamorphosis of Pregnancy (BUMP) study was funded by 4YouandMe and Sema4 along with supplemental in-kind contributions from coalition partners (Evidation Health, Vector Institute, Cambridge Cognition, and Bodyport).

4YouandMe primarily conducted the BUMP study and carried out all 4YouandMe direct recruitment (including paid and social media, community partnerships with the Moses/Weitzman Health System, and others [eg, word of mouth, press, and ClinicalTrials.gov]). Sema4 facilitated genetic testing subscriber recruitment. Evidation Health was responsible for data collection and maintenance. Cambridge cognition provided active and cognitive tasks used in the BUMP study app. The Vector Institute aided in poststudy data analysis.

In addition, we would like to express our gratitude to the Moses/Weitzman Health System for their partnership in helping to make this research possible—particularly, Amy Gagliardi and staff at the Middletown, Clinton, and Meriden sites for their on-the-ground efforts. Additional thanks to Ridley McKenzie and Lesley Harris from the 4YouandMe team for their study support, as well as all of the BUMP study participants for their time, effort, and contributions to this research.

Data Availability

The datasets generated or analyzed during this study will be hosted in the public domain and available on the Synapse platform at Sage Bionetworks [29]. Please check the 4YouandMe website or contact the corresponding author for an update on when study data will be available and how to access it.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The BUMP study recruitment flier for outreach via 4YouandMe direct recruitment approaches designed for long-form social media sites (eg, Reddit). BUMP: Better Understanding the Metamorphosis of Pregnancy.
[PNG File , 170 KB - formative_v9i1e68093_app1.png]

Multimedia Appendix 2
The BUMP study recruitment flier for outreach via 4YouandMe direct recruitment approaches designed for long-form social media sites (eg, Reddit). BUMP: Better Understanding the Metamorphosis of Pregnancy.
[PNG File , 641 KB - formative_v9i1e68093_app2.png]

Multimedia Appendix 3
The BUMP study recruitment flier for outreach via 4YouandMe direct recruitment approaches designed for long-form social media sites (text version), designed for long-form social media sites with specific forums that did not allow images (eg, Reddit). BUMP: Better Understanding the Metamorphosis of Pregnancy.

XSL•FO
RenderX

[PNG File , 271 KB - formative_v9i1e68093_app3.png]

Multimedia Appendix 4
The BUMP study recruitment flier for outreach via 4YouandMe direct recruitment approaches designed for long-form social media sites, tailored for image-dominant social media sites (eg, Instagram). BUMP: Better Understanding the Metamorphosis of Pregnancy.
[PNG File , 163 KB - formative_v9i1e68093_app4.png]

Multimedia Appendix 5
The BUMP study recruitment flier for outreach via 4YouandMe direct recruitment approach, tailored to community health partnerships (Moses/Weitzman Health System). BUMP: Better Understanding the Metamorphosis of Pregnancy.
[PNG File , 2377 KB - formative_v9i1e68093_app5.png]

Multimedia Appendix 6
The BUMP study recruitment flier for outreach via 4YouandMe direct recruitment approach (alternate flier), tailored to community health partnerships (Moses/Weitzman Health System). BUMP: Better Understanding the Metamorphosis of Pregnancy.
[PNG File , 528 KB - formative_v9i1e68093_app6.png]

Multimedia Appendix 7
Characteristics of the cohort among those who engaged (ENG) with recruitment materials, enrolled (ENR), and were retained (RTD) in the BUMP study for participants recruited via other methods. BUMP: Better Understanding the Metamorphosis of Pregnancy.
[PNG File , 285 KB - formative_v9i1e68093_app7.png]

References

1. Scharff DP, Mathews KJ, Jackson P, Hoffsuemmer J, Martin E, Edwards D. More than Tuskegee: understanding mistrust about research participation. J Health Care Poor Underserved 2010 Aug;21(3):879-897 [FREE Full text] [doi: 10.1353/hpu.0.0323] [Medline: 20693733]
2. Staccini P, Lau AYS. Consuming health information and vulnerable populations: factors of engagement and ongoing usage. Yearb Med Inform 2022 Aug;31(1):173-180 [FREE Full text] [doi: 10.1055/s-0042-1742549] [Medline: 36463876]
3. Swartz T, Palermo A, Masur S, Aberg J. The science and value of diversity: closing the gaps in our understanding of inclusion and diversity. J Infect Dis 2019 Aug 20;220(220 Suppl 2):S33-S41 [FREE Full text] [doi: 10.1093/infdis/jiz174] [Medline: 31430380]
4. NIH grants and funding acronym and glossary list. National Institutes of Health. 2022. URL: https://toolkit.ncats.nih.gov/glossary/underrepresented-population/ [accessed 2023-11-10]
5. Dauphin C, Clark N, Cadzow R, Saad-Harfouche F, Rodriguez E, Glaser K, et al. #BlackBreastsMatter: process evaluation of recruitment and engagement of pregnant African American women for a social media intervention study to increase breastfeeding. J Med Internet Res 2020 Aug 10;22(8):e16239 [FREE Full text] [doi: 10.2196/16239] [Medline: 32773377]
6. Farmer DF, Jackson SA, Camacho F, Hall MA. Attitudes of African American and low socioeconomic status White women toward medical research. J Health Care Poor Underserved 2007 Feb;18(1):85-99. [doi: 10.1353/hpu.2007.0008] [Medline: 17337800]
7. Washington V, Franklin JB, Huang ES, Mega JL, Abernethy AP. Diversity, equity, and inclusion in clinical research: a path toward precision health for everyone. Clin Pharmacol Ther 2023 Mar;113(3):575-584. [doi: 10.1002/cpt.2804] [Medline: 36423203]
8. UyBico SJ, Pavel S, Gross CP. Recruiting vulnerable populations into research: a systematic review of recruitment interventions. J Gen Intern Med 2007 Jun;22(6):852-863 [FREE Full text] [doi: 10.1007/s11606-007-0126-3] [Medline: 17375358]
9. Inclusion of women and minorities as participants in research involving human subjects. National Institutes of Health. 2022. URL: https://grants.nih.gov/policy/inclusion/women-and-minorities.htm [accessed 2023-07-30]
10. Yancey AK, Ortega AN, Kumanyika SK. Effective recruitment and retention of minority research participants. Annu Rev Public Health 2006;27:1-28. [doi: 10.1146/annurev.publhealth.27.021405.102113] [Medline: 16533107]
11. Shavers VL, Lynch CF, Burmeister LF. Racial differences in factors that influence the willingness to participate in medical research studies. Ann Epidemiol 2002 May;12(4):248-256. [doi: 10.1016/s1047-2797(01)00265-4] [Medline: 11988413]
12. Hoyert DL. Maternal mortality rates in the United States, 2021. NCHS Health E-Stats. 2023. URL: https://stacks.cdc.gov/view/cdc/124678 [accessed 2023-03-16]
13. Blehar MC, Spong C, Grady C, Goldkind SF, Sahin L, Clayton JA. Enrolling pregnant women: issues in clinical research. Womens Health Issues 2013 Jan;23(1):e39-e45 [FREE Full text] [doi: 10.1016/j.whi.2012.10.003] [Medline: 23312713]

XSL•FO
RenderX

14. Russomanno J, Patterson JG, Jabson Tree JM. Social media recruitment of marginalized, hard-to-reach populations: development of recruitment and monitoring guidelines. JMIR Public Health Surveill 2019 Dec 02;5(4):e14886 [FREE Full text] [doi: 10.2196/14886] [Medline: 31789598]
15. Ellington M, Connelly J, Clayton P, Lorenzo CY, Collazo-Velazquez C, Trak-Fellermeier MA, et al. Use of Facebook, Instagram, and Twitter for recruiting healthy participants in nutrition-, physical activity-, or obesity-related studies: a systematic review. Am J Clin Nutr 2022 Feb 09;115(2):514-533 [FREE Full text] [doi: 10.1093/ajcn/nqab352] [Medline: 34669955]
16. Leading social media websites in the United States as of August 2023, based on share of visits. Statista. 2023. URL: https://www.statista.com/statistics/265773/market-share-of-the-most-popular-social-media-websites-in-the-us/ [accessed 2023-11-15]
17. Number of social network users in selected countries in 2022 and 2027. Statista. 2023. URL: https://www.statista.com/statistics/278341/number-of-social-network-users-in-selected-countries/ [accessed 2023-12-30]
18. Social media fact sheet. Pew Research Center. 2022. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/#panel-2fc5fff9-9899-4317-b786-9e0b60934bcf [accessed 2023-07-30]
19. Ginsburg GS, Picard RW, Friend SH. Key issues as wearable digital health technologies enter clinical care. N Engl J Med 2024 Mar 21;390(12):1118-1127. [doi: 10.1056/NEJMra2307160] [Medline: 38507754]
20. Chesser A, Burke A, Reyes J, Rohrberg T. Navigating the digital divide: a systematic review of eHealth literacy in underserved populations in the United States. Inform Health Soc Care 2016;41(1):1-19. [doi: 10.3109/17538157.2014.948171] [Medline: 25710808]
21. Mitchell SJ, Godoy L, Shabazz K, Horn IB. Internet and mobile technology use among urban African American parents: survey study of a clinical population. J Med Internet Res 2014 Jan 13;16(1):e9 [FREE Full text] [doi: 10.2196/jmir.2673] [Medline: 24418967]
22. Mitchell UA, Chebli PG, Ruggiero L, Muramatsu N. The digital divide in health-related technology use: the significance of race/ethnicity. Gerontologist 2019 Jan 09;59(1):6-14. [doi: 10.1093/geront/gny138] [Medline: 30452660]
23. James DCS, Harville C. eHealth literacy, online help-seeking behavior, and willingness to participate in mHealth chronic disease research among African Americans, Florida, 2014-2015. Prev Chronic Dis 2016 Nov 17;13:E156 [FREE Full text] [doi: 10.5888/pcd13.160210] [Medline: 27854421]
24. James DCS, Harville C, Whitehead N, Stellefson M, Dodani S, Sears C. Willingness of African American women to participate in e-Health/m-Health research. Telemed J E Health 2016 Mar;22(3):191-197 [FREE Full text] [doi: 10.1089/tmj.2015.0071] [Medline: 26313323]
25. George N, Reynolds S, de Long R, Kacica M, Ahmed R, Manganello J. Social media and Black maternal health: the role of health literacy and ehealth literacy. Health Lit Res Pract 2023 Jul;7(3):e119-e129 [FREE Full text] [doi: 10.3928/24748307-20230614-01] [Medline: 37463292]
26. Kasoju N, Remya NS, Sasi R, Sujesh S, Soman B, Kesavadas C, et al. Digital health: trends, opportunities and challenges in medical devices, pharma and bio-technology. CSIT 2023 Apr 11;11(1):11-30. [doi: 10.1007/s40012-023-00380-3]
27. Friend SH, Ginsburg GS, Picard RW. Wearable digital health technology. N Engl J Med 2023 Nov 30;389(22):2100-2101. [doi: 10.1056/NEJMe2303219] [Medline: 38048193]
28. Saeed SA, Masters RM. Disparities in health care and the digital divide. Curr Psychiatry Rep 2021 Jul 23;23(9):61 [FREE Full text] [doi: 10.1007/s11920-021-01274-4] [Medline: 34297202]
29. Goodday SM, Karlin E, Brooks A, Chapman C, Karlin DR, Foschini L, et al. Better Understanding of the Metamorphosis of Pregnancy (BUMP): protocol for a digital feasibility study in women from preconception to postpartum. NPJ Digit Med 2022 Mar 30;5(1):40 [FREE Full text] [doi: 10.1038/s41746-022-00579-9] [Medline: 35354895]
30. 2022 American Community Survey (ACS). United States Census Bureau. 2022. URL: https://www.census.gov/programs-surveys/acs/news/data-releases.html [accessed 2023-12-04]
31. Oudat Q, Bakas T. Merits and pitfalls of social media as a platform for recruitment of study participants. J Med Internet Res 2023 Oct 11;25:e47705 [FREE Full text] [doi: 10.2196/47705] [Medline: 37819692]
32. Tsaltskan V, Sanchez Baez R, Firestein GS. Cost-effectiveness of social media advertising as a recruitment tool: a systematic review and meta-analysis. J Clin Transl Sci 2023;7(1):e180 [FREE Full text] [doi: 10.1017/cts.2023.596] [Medline: 37745929]
33. Distribution of instagram users in the United States as of October 2024, by age group. Statista. URL: https://www.statista.com/statistics/398166/us-instagram-user-age-distribution/ [accessed 2024-12-04]
34. Percentage of U.S. adults who use Reddit as of September 2023, by age group. Statista. 2024. URL: https://www.statista.com/statistics/261766/share-of-us-internet-users-who-use-reddit-by-age-group/ [accessed 2024-12-04]
35. Osterman MJK, Hamilton BE, Martin JA. Births: final data for 2022. National Vital Statistics Reports 2024;73(2):1-39. [doi: 10.15620/cdc:145588]
36. 2023 American Community Survey (ACS). United States Census Bureau. 2023. URL: https://www.census.gov/programs-surveys/acs/news/data-releases.html [accessed 2023-12-04]
37. How use of online platforms – such as Facebook, Instagram or TikTok – differs among some U.S. demographic groups. Pew Research Center. 2024. URL: https://www.pewresearch.org/internet/2024/01/31/americans-social-media-use/pi_2024-01-31_social-media-use_00_04-png [accessed 2024-12-04]

XSL•FO
RenderX

Abbreviations

BUMP: Better Understanding the Metamorphosis of Pregnancy
DHTs: digital health technologies
eConsent: electronic consent
NIH: National Institutes of Health
REDCap: Research Electronic Data Capture
URPs: Underrepresented populations

Edited by A Mavragani; submitted 29.10.24; peer-reviewed by M Gasmi , C Banko, C Nnajiofor; comments to author 17.11.24; revised version received 11.12.24; accepted 25.02.25; published 15.04.25.

Please cite as:
Harry C, Goodday S, Chapman C, Karlin E, Damian AJ, Brooks A, Boch A, Lugo N, McMillan R, Tempero J, Swanson E, Peabody S, McKenzie D, Friend S
Using Social Media to Engage and Enroll Underrepresented Populations: Longitudinal Digital Health Research
JMIR Form Res 2025;9:e68093
URL: https://formative.jmir.org/2025/1/e68093
doi:10.2196/68093
PMID:

©Christiana Harry, Sarah Goodday, Carol Chapman, Emma Karlin, April Joy Damian, Alexa Brooks, Adrien Boch, Nelly Lugo, Rebecca McMillan, Jonell Tempero, Ella Swanson, Shannon Peabody, Diane McKenzie, Stephen Friend. Originally published in JMIR Formative Research (https://formative.jmir.org), 15.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Influence of Medical Expertise and Information Search Skills on Medical Information Searching: Comparative Analysis From a Free Data Set

Aline Chevalier[1*], PhD, Prof Dr; Cheyenne Dosso[2*], PhD

[1]Laboratoire Cognition, Langues, Langage, Ergonomie (Centre National de la Recherche Scientifique UMR5263) Maison de la Recherche, Université de Toulouse, Toulouse, France

[2]Laboratoire d'anthropologie et de psychologie cliniques, cognitives et sociales (UPR7278), Université Côte d'Azur, Nice, France

[*]all authors contributed equally

Corresponding Author:
Aline Chevalier, PhD, Prof Dr
Laboratoire Cognition, Langues, Langage, Ergonomie (Centre National de la Recherche Scientifique UMR5263) Maison de la Recherche
Université de Toulouse
5 allées Machado
Toulouse, 31058
France
Phone: 33 561503531
Email: aline.chevalier@univ-tlse2.fr

Abstract

Background: Nowadays, the internet has become the primary source of information for physicians seeking answers to medical questions about their patients before consulting colleagues. However, many websites provide low-quality, unreliable information that lacks scientific validation. Therefore, physicians must develop strong information search skills to locate relevant, accurate, and evidence-based content. However, previous studies have shown that physicians often have poor search skills and struggle to find information on the web, which may have detrimental consequences for patient care.

Objective: This study aims to determine how medical students and residents searched for medical information on the internet, the quality of the web resources they used (including their nature and credibility), and how they evaluated the reliability of these resources and the answers they provided. Given the importance of domain knowledge (in this case, medicine) and information search skills in the search process, we compared the search behaviors of medical students and residents with those of computer science students. While medical students and residents possess greater medical-related knowledge, computer science students have stronger information search skills.

Methods: A total of 20 students participated in this study: 10 medical students and residents, and 10 computer science students. Data were extracted from a freely accessible data set in accordance with FAIR (Findable, Accessible, Interoperable, and Reusable) principles. All participants searched for medical information online to make a diagnosis, select a treatment, and enhance their knowledge of a medical condition—3 primary activities they commonly perform. We analyzed search performance metrics, including search time, the use of medical-related keywords, and the accuracy of the information found, as well as the nature and credibility of web resources used by medical students and residents compared with computer science students.

Results: Medical students and residents provided more accurate answers than computer science students without requiring additional time. Their medical expertise also enabled them to better assess the reliability of resources and select high-quality web sources, primarily from hospital websites. However, it is noteworthy that they made limited use of evidence-based tools such as PubMed.

Conclusions: Although medical students and residents generally outperformed computer science students, they did not frequently use evidence-based tools. As previously observed, they may avoid databases due to the risk of encountering too many irrelevant articles and difficulties in applying appropriate filters to locate relevant information. Nevertheless, clinical and practical evidence-based medicine plays a crucial role in updating physicians' knowledge, improving patient care, and enhancing physician-patient relationships. Therefore, information search skills should be an integral part of medical education and continuing professional development for physicians.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e62754) doi:10.2196/62754

KEYWORDS

information searching; credibility; internet; medicine; information search skills

Introduction

Internet as a Tool for Medical Information Searching by Physicians

Nowadays, searching for information on the web is an integral part of a physician's professional activity. While colleagues were the primary source of information in 2006 [1], more recently, the internet has become the first source before consulting colleagues when physicians have medical questions about patients [2,3].

Efficient information searching has various positive outcomes for physicians and, by extension, their patients [2]. It may improve patient care [4,5], clinical decision-making [5], and physicians' knowledge [5,6]. It also reduces delays in diagnosis, treatment initiation, and modification [5], as well as potential medical errors and unnecessary medical examinations [6].

More precisely, physicians, especially younger ones, use the internet to search for information for various purposes [1]: (1) finding information and solutions to specific patient diagnostic issues (reported by 33.7%-57% of respondents in the previous study [1]); (2) obtaining new information on diseases (20.4%); and (3) staying informed about the latest research on specific topics and general matters, such as diseases (27%), new therapies and treatments (8.8%), and drug dosing (5.6%), as also shown by [7-9].

Concerning information related to treatment and diagnosis, physicians take between 2 and 32 minutes to find answers on the internet, depending on the nature of the information sought (eg, side effects, drug dosing, new therapies) [2]. Nevertheless, most studies indicate that physicians spend less than 10 minutes searching for answers and less than 2 minutes during patient consultations [2]. During consultations, they primarily search for information on drug dosing and its potential side effects or to show images to patients [10].

Difficulties in Finding Relevant Web Resources and Its Consequences

Although physicians reported finding information in most cases (>50% of searches for 70% of physicians) and over 60% were confident in their ability to find the information they needed on the internet [2], they still face various difficulties and barriers in information searching. For instance, a survey of general practitioners identified the main obstacles as a lack of information search knowledge or skills, clinical practice conditions (time pressure or patient-related concerns), information-related issues (information overload, quality concerns, or low relevance), and language barriers [11]. It is noteworthy that, occasionally, reliance on internet-supplied information can lead to clinical errors, as physicians may respond to and process such information—even when it contradicts their preexisting knowledge [12]—regardless of its accuracy. While the use of web resources can enhance physicians' clinical decisions [6], occasional discrepancies in the quality and reliability of the information found may introduce errors in the decision-making process. Outpatient diagnostic error rates have been estimated at 5%, affecting more than 12 million individuals per year, while inpatient diagnostic errors range from 6% to 7%, with 20% of initial diagnoses being modified [13]. Efficient use of information access systems (eg, search engines) can positively impact patient care, but misuse of the internet may have the opposite effect. Therefore, physicians and medical students must develop accurate information search skills (ie, procedural knowledge of tools such as Google [Alphabet Inc] or PubMed [National Institutes of Health]) to use these systems effectively within the specificities of the medical domain.

Previous studies have reported a lack of information search skills, requiring technical support for information retrieval [1], unawareness of accessible sources [2], difficulties in accessing reliable and up-to-date medical information, and challenges in developing efficient search strategies [4]. Additionally, on the internet, physicians face the challenge of processing large amounts of information in a very short time [14].

Previous studies have shown that well-designed, easy-to-navigate websites are generally considered more reliable than those with a confusing or complex interface [15]. The first websites listed by search engine tools are often perceived as more credible than those appearing later [16]. However, neither a well-designed website nor a high search engine ranking guarantees that the information provided is credible and reliable. Although physicians require reliable and relevant documents to integrate into clinical practice, the internet presents a vast array of medical-related information, not all of which is verified. For instance, a study found that only 58% of eHealth websites in the United States met the criteria for content accuracy and credibility [17]. Consequently, physicians must develop information search skills to effectively find and assess web resources.

However, an institutionalized perception persists that the information found does not significantly enhance medical practice [6]—for a review, see [2]. This perception may be partly because not all physicians are aware of the availability of evidence-based medical resources on the internet (eg, PubMed) or they may lack the search skills to access them [18].

Main Web Resources Used

To effectively improve physicians' information search skills, it is essential to understand how medical students search for medical information and which web resources they use during their university program. A study on medical students found that they searched for online medical and clinical information daily (2-3 times per day) [19]. They accessed numerous websites (>50), including some recommended by the NHS (National Health Service), such as PubMed, which was used by only 30% of physicians and accounted for just 8% of all their searches.

XSL•FO
RenderX

However, they also relied on many general websites, particularly Wikipedia (Wikimedia Foundation; used by 71% of students and accounting for 26% of searches), patient forums, medical-specific wikis, and Google (used by 79% of students and comprising 32% of all searches). Indeed, Google is the most widely used general web search engine and has become an important tool in physicians' searches. Nevertheless, it is not specialized in medicine, and numerous websites present medical-related information without being reviewed by scientists or physicians [20,21], often leading to low-quality information [22]. Consequently, physicians must carefully evaluate the quality of information obtained from websites. Furthermore, physicians and medical students often prefer using Google over medicine-specific databases to answer clinical questions [23]. For instance, Krause et al [23] provided clinical questions to emergency department residents and asked them to search for answers either using a search engine (Google) or without it (by searching in a real environment). The results showed that while the number of incorrect answers increased with Google, the number of unsure responses decreased compared with searches conducted without Google. The residents developed a false sense of confidence in their answers and demonstrated poor efficiency in searching for and identifying correct clinical information using Google. While students and physicians may perceive Google as an evidence-based medical tool, it is not inherently an evidence-based medical database like PubMed.

Based on these prior studies, we can assert that medical-related information available on the internet can support clinical decision-making in patient health care, provided that physicians and medical students are trained to evaluate the quality of sources and content. The quality of medical-related information can be assessed according to 2 dimensions: (1) The credibility of sources, which refers to the degree of trustworthiness associated with the information and its origin [24-27]. This includes evaluations of accuracy, authority, objectivity, currency, and coverage [26,28]. Credibility is closely related to trustworthiness, reliability, accuracy, authority, and quality [24,25,29], although not always inseparable (eg, reliable information can sometimes be found on sites with low credibility). Consequently, credibility may influence the perceived reliability of information. (2) Reliability, in turn, refers to the extent to which information remains consistent and trustworthy under similar circumstances [30,31].

In the medical domain, Mikalef et al [6] classified credible documents as follows: (1) Authoritative web documents include scientific digital databases (eg, PubMed), websites of medical organizations (eg, the European Society of Cardiology), online peer-reviewed scientific journals (eg, The Lancet, JAMA), specialized medical journals (eg, *Circulation*, *Heart*), e-books, and government agency websites (eg, the Public Health Administration, the Food and Drug Administration). (2) Nonauthoritative web documents include medical portals, personal websites of physicians, pharmaceutical company websites, social media, and medical equipment websites. While authoritative sources can generally be considered credible and reliable, nonauthoritative sources require in-depth evaluation to determine their credibility and reliability. Therefore, when physicians and medical students use general websites to find medical-related information, they must:

- Translate a clinical or medical question into a search strategy to locate reliable information. This requires using specific medical terms or selecting credible web resources (eg, scientific databases).
- Evaluate credibility, reliability, relevance, and accessibility to determine which web resources to consult and use (see the review by [2]).

These 2 activities can be highly complex, especially when using general-purpose search engines (eg, Google), as individuals must select appropriate keywords and identify credible websites to obtain reliable information relevant to their professional objectives.

To this end, medical students and residents can rely on their prior domain knowledge to navigate the web more effectively. However, they often lack information search skills and familiarity with scientific tools and search systems, which are critical for efficient information retrieval. Users with strong information search skills are better guided by the components of the search space, regardless of the domain in which they seek information [32], a phenomenon known as the cross-domain transfer of information search skills [33]. Specifically, such users demonstrate more adaptive search behaviors tailored to their tasks and employ advanced search strategies [34]. For instance, they are more likely to incorporate Boolean operators into their queries to construct complex search equations rather than relying solely on simple keyword searches [35,36]. Additionally, they effectively utilize indexes and thesauri provided by search systems to identify relevant vocabulary, allowing them to refine their queries with greater specificity and precision as their search sessions progress [32].

However, not all academic programs place equal emphasis on developing information search skills. For example, computer science students acquire more advanced operational search skills compared with students in other fields [36,37]. Consequently, computer scientists have a deeper understanding of search systems and use them more effectively—leveraging features such as recommendations, thesauri, and domain-specific scientific databases—to conduct advanced searches, expand their vocabulary, and achieve their research objectives [38]. These search skills enable them to construct more structured and precise queries as their searches evolve [39].

In summary, prior domain knowledge facilitates attentional guidance as well as the reading and comprehension of web-based information [40]. By contrast, prior information search skills enhance the effective use and understanding of information systems.

Research Objectives and Questions

This study aimed to examine the effects of domain-specific prior knowledge versus information search skills on information search behaviors and performance. While medical students and residents benefit from their medical expertise when conducting medical searches, they lack the information search skills of their computer science counterparts, and vice versa. Identifying performance indicators and the challenges faced by these distinct

XSL•FO
RenderX

user profiles will help clarify the limitations of relying solely on domain knowledge in medical information searches.

More precisely, most prior studies examining how physicians search for information have relied on interviews or questionnaires to assess their search behaviors [2]. However, very few have analyzed the actual search activities of physicians or medical students and residents, despite the importance of understanding how they navigate the web when conducting necessary searches. To address this gap, we conducted an experimental study in which medical students and residents were compared with computer science students in their web-based information searches. More specifically, this study aimed to determine whether prior medical knowledge enabled medical students and residents to achieve higher search performance and make better choices regarding reliable and credible resources compared with computer science students, who had greater experience in information search and higher search skills.

More specifically, this study aimed to provide insights into the following research questions (RQs):

- RQ1: Do the most common searches conducted by physicians—treatment, diagnosis, and learning about a specific medical topic—result in different search performances between computer science students (who have higher information search skills but less medical knowledge) and medical students and residents (who have greater medical knowledge but lower information search skills)?
- RQ2: What types of resources do they consult during medical search tasks, and how do they assess the credibility of these resources based on their domain expertise (medicine vs information search)?
- RQ3: Do these 2 groups of students evaluate the reliability of the resources they use and the quality of the answers they provide differently when performing a medical search task?

Methods

Recruitment

The freely accessible data set used in this study was extracted from a user study conducted by Dosso et al [41] during the COVID-19 lockdown. The data set is available online [42].

A total of 20 students from the University of Toulouse participated in this study:

- 10 medical students and residents in medicine: 5 females and 5 males (mean age 25.4 years, SD 3.1 years); 2 were pursuing a master's degree and 8 were in an MD (Doctor of Medicine) program.
- 10 computer science students: 4 females and 6 males (mean age 23.8 years, SD 1.5 years); 5 were pursuing a master's degree and 5 were PhD (Doctor of Philosophy) students.

To evaluate their level of knowledge in medicine and their familiarity with the internet, all the participants had:

- To self-evaluate their knowledge in medicine, on a 5-point Likert scale ranging from 1 (very low) to 5 (very high). Medical students self-evaluated their level of medical knowledge higher than computer science students ($mean_{med}$ 4.2, SD 0.63; $mean_{cs}$ 1.3, SD 0.48; t_{18}=–11.5, P<.001, d=–5.15).
- To complete a multiple-choice questionnaire in the Medicine domain, elaborated by a physician, which consisted of 10 questions, each with 5 response options (1 correct, 3 incorrect, and 1 "I don't know"). Only correct answers earned 1 point, so the score per participant could range from 0 to 10. Results indicated that student-residents in medicine scored significantly higher (mean 5, SD 0.66) than students in computer science (mean 0.5, SD 0.97; t_{18}=–12.1, P<.001, d=–5.4).
- To complete the Information Search Self-Efficacy Scale developed by Rodon et al [43] and indicate the number of years of internet use and the contexts in which it is used. Results showed that students in computer science (mean 4.4, SD 0.51) had higher information search self-efficacy scores than those in medicine (mean 3.1, SD 0.73; t_{18}=4.56, P<.001, d=2.04). However, there was no significant difference in the number of years of internet use between the 2 groups ($mean_{CS}$ 12.3, SD_{CS} 1.83 and $mean_{med}$ 14, SD_{med} 2.79; t_{18}=–1.61, P=.12, d=–0.721). All the participants reported using the internet for both university and personal purposes.

Ethical Considerations

Before starting the search tasks, all participants received a detailed explanation of the study's objectives and procedures. Informed consent was obtained from each participant through a written agreement, which outlined the study's purpose, potential risks and benefits, data usage protocols, and participant rights, including the right to withdraw at any time without consequence. This study was reviewed and approved by the Ethics and Research Committee of the University of Toulouse (approval/project number 2021-354).

Procedure and Material

First, all participants completed an online prequestionnaire that included a free and informed consent form, questions on sociodemographic variables (eg, age, gender, education level, native language), internet information search scales, and multiple-choice questions assessing their level of medical knowledge. Additionally, participants described their general internet habits (ie, browsing device used, preferred browser, and search engine) and completed the Information Search Self-Efficacy Scale [43].

Then, the participants conducted the experiment at home during the COVID-19 lockdown in 2020-21. The experimental materials included the following:

- A printed booklet with general instructions for the experiment, a training exercise on using the logging browser, statements for the search tasks, and designated spaces for writing answers. As the written production task was conducted with pen and paper, the computer was used only for the search sessions.

XSL•FO
RenderX

- A memory stick containing the browser was provided for the search sessions, enabling participants to directly search for information while reviewing the experimental material.

Participants were instructed to complete each search task independently, without taking notes during the search session. This ensured that their final answers primarily reflected the information retained in memory after the search.

Both groups of participants (medical students/residents and computer science students) were presented with 3 distinct search tasks in general medicine (see Table 1 for detailed task descriptions). These tasks reflected real-world scenarios commonly encountered in medical practice, such as (1) enhancing knowledge in a specific disease area, (2) searching for information related to a particular patient problem, and (3) identifying and selecting the most appropriate treatment options for patients.

These scenarios are well-documented in medical research, as supported by previous studies [1,2].

The instructions specified that participants had to write their answers for each task only after completing the search session (ie, after closing the browser). There was no time limit. Note-taking was not permitted during the search session to ensure that responses closely reflected the knowledge retained in participants' memory.

Three different search tasks in general medicine were assigned to the participants (medical students/residents and computer science students; see Table 1 for the search tasks). These tasks closely resembled real-world scenarios encountered by physicians and medical students, such as improving their knowledge in a specific disease area, searching for information on a particular patient problem, and determining the most appropriate treatment for patients [1,2].

Table 1. The 3 different tasks provided to the participants and correct answers.

Three different tasks and answers	Tasks' instructions	Subtasks required
Learning task		
• Purpose: Improving knowledge of a specific area/disease • Correct answer: No correct answer was defined; instead, participants' responses were analyzed to compute the number of relevant elements they provided.	• You want to learn more about endometriosis.	• Understanding endometriosis • Understand the issues related to endometriosis • Discover/learn more about medical solutions to treat endometriosis
Treatment decision task		
• Purpose: Choosing the most appropriate treatment • Correct answer: Anticoagulant treatment (because the old woman did not fall too many times for her age)	• An 83-year-old woman had a nonsequelae stroke 5 months ago. • At the stroke assessment, atrial fibrillation was discovered. • She had fallen 3 times in the last 2 months. • Should anticoagulant therapy be initiated? • After evaluating the benefit-risk ratio of initiating or not initiating anticoagulant treatment, select the management option that you find most appropriate and justify your choice.	• Understanding nonsequential stroke • Understanding atrial fibrillation • Understanding the risk of falls • Identify the advantages and disadvantages of anticoagulant therapy • Judge this information according to the criteria to be established (benefit/risk assessment)
Diagnosis task		
• Purpose: Determining the pathology • Correct answer: Splenic infarction or spleen infarction	• A 47-year-old man presented to the emergency department with left hypochondrium pain that had been evolving for 24 hours and was not relieved by level 1 analgesics. • His history included cutaneous lupus and polycythemia. • The biological workup was normal. • The abdominopelvic scanner found 2 splenic hypodensities. With the information collected on the internet, propose your diagnosis and etiological hypotheses	• Understanding left hypochondrium • Understanding of level 1 analgesics • Understanding cutaneous lupus and polycythemia • Understanding splenic hypodensity • Analyze the information previously collected and propose a diagnosis with etiological hypotheses

Variables

Two independent variables (IVs) were manipulated:

- IV1: Domain expertise of participants (medical students/residents vs computer science students) as a between-participants variable.
- IV2: Search tasks (learning, choosing treatment, diagnosis) as a within-participants variable.

A total of 6 dependent variables (DVs) were measured to bring some answers to the 3 RQs. For RQ1, we measured 3 DVs:

- DV1: Total time (seconds) spent on each search task, measured from the first query entered to the completion of the multiquery process.

XSL•FO
RenderX

- DV2: Number of medical-related keywords generated by participants, excluding those explicitly provided in the search task instructions.
- DV3: Accuracy and relevance of retrieved information:
  - For diagnostic and treatment tasks, responses were scored as 0 (incorrect or abandoned) or 1 (correct).
  - For the learning task, responses were scored from 0 to 4 based on the amount of relevant information provided and interconnected. Relevant information was defined as content directly related to the assigned topic (endometriosis) in accordance with the information context [44]. This is rated as follows: 0 (incorrect information), 0.5 (correct information), 1 (correct and relevant information), 1.5 (correct and relevant information linked to general vocabulary), 2 (correct and relevant information linked to medical vocabulary), 3 (correct and relevant information linked to general vocabulary and interconnected), and 4 (correct and relevant information linked to medical vocabulary and interconnected).

For RQ2, we measured:

- DV4: Evaluation of the credibility and type of web resources consulted:
  - DV4a: Credibility score of the websites visited for each task.
  - DV4b: Distribution of the types of web resources consulted, categorized as websites or as documents.

For DV4a, the credibility of web resources consulted per search task was scored as follows, based on Mikalef et al [6]:

- Peer-reviewed scientific journals or scientific databases (eg, PubMed, ScienceDirect, The Lancet)=2 points per URL
- Documents from hospitals, health authorities (eg, World Health Organization), and government agencies=1.5 points per URL
- Health-related websites verified by HONcode (Health On the Net code; certification by the nonprofit Health On the Net Foundation [45])=1 point per URL
- Other websites (not scientifically verified)=0 points per URL

For DV4b, we analyzed the distribution of resources consulted across different categories (peer-reviewed scientific journals, hospital/government documents, HONcode-certified websites, and other websites) during each search task. This distribution was based on the number of resources consulted per category (as defined in DV4a) and per search task.

For RQ3, 2 DVs were measured:

- DV5: Self-assessment of reliability of web documents visited while achieving the search task [46]: “Overall, the search engine results, websites, or documents (results from Google, websites, pdf, etc) consulted were: (1) absolutely not trustworthy, (2) very untrustworthy, (3) not very trustworthy, (4) moderately trustworthy, (5) somewhat trustworthy, (6) trustworthy, or (7) extremely trustworthy. The higher the score was, the more participants perceived web documents as reliable.”
- DV6: Self-assessment of the quality of their own answers on a 4-point scale for the 3 tasks. Participants rated their perceived answer quality as follows: (1) very bad, (2) bad, (3) good, and (4) very good. A higher score indicated a more favorable self-evaluation of their answers.

Statistical Analysis

For all DVs, we compared student-residents in medicine with students in computer science for each task using either parametric or nonparametric tests. Specifically, for all DVs except DV4b, we performed Student *t* tests. For DV4b, we conducted a repeated-measures ANOVA. Tests for homogeneity of variance (Levene test) and normality (Shapiro-Wilk test) were performed before conducting ANOVA and unpaired (1-tailed) *t* tests. Partial η^2 was used as an index of effect size. When homogeneity of variance was violated, we applied the Mann-Whitney test. Additionally, some quantitative results are illustrated with qualitative examples.

Results

Overview

The results are presented according to the 3 RQs.

- *RQ1: Do the most popular searches developed by physicians (ie, treatment, diagnosis, and learning on a specific medical topic) lead to different search performances between computer science students (with higher information search skills but lower medical knowledge) and medical student-residents (with higher medical knowledge but lower information search skills)?*

Table 2 presents the dependent variables for RQ1.

XSL•FO
RenderX

Table 2. Dependent variables for RQ1 with regard to the search task and expertise domain.

Group of students	Learning task, mean (SD)	Diagnostic task, mean (SD)	Treatment task, mean (SD)
DV1: Search time (seconds)			
Medicine	548 (474)	705 (444)	634 (407)
Computer science	520 (373)	1099 (1085)	936 (699)
DV2: New keywords related to medicine (mean number)			
Medicine	1.80 (0.91)[a]	4 (4.19)	4.8 (4.13)
Computer science	0.3 (0.48)[a]	2 (2.35)	2.7 (3.97)
DV3: Correct answers or relevant information provided (mean number)			
Medicine	3.9 (0.31)[a]	0.9 (0.31)[a]	0.85 (0.15)[a]
Computer science	3.5 (0.52)[a]	0 (0)[a]	0.1 (0.31)[a]

[a]Significant differences between medical students/residents and computer science students.

Time Spent to Find Information

Regarding the time required to find information, no significant differences were observed between student-residents in medicine and those in computer science for any task (all *t* tests were nonsignificant: t_{18}=-.146, *P*=.88 for the learning task; t_{18}=1.06, *P*=.30 for the diagnostic task; and t_{18}=1.18, *P*=.25 for the treatment task).

Keywords Related to Medicine Added by the Participants

Student-residents in medicine added more medicine-related keywords than students in computer science, but only for the learning task (t_{18}=–4.57, *P*<.001, *d*=2.04). On average, student-residents in medicine added 1.6 keywords (SD 0.91), compared with 0.3 (SD 0.48) for students in computer science. The keywords added by student-residents in medicine were primarily related to specific websites they were familiar with (eg, associations of doctors in gynecology or government health agency websites, where they could find medical information).

For the treatment and diagnostic tasks, statistical analyses revealed no significant differences between student-residents in medicine and students in computer science (t_{18}=–1.32, *P*=.20, and t_{18}=–1.16, *P*=.26, respectively).

Correct information retrieval for the diagnostic and treatment tasks, as well as the number of relevant information elements provided by participants for the learning task, was analyzed. The results are as follows:

For the treatment and diagnostic tasks, medical student-residents provided more correct answers than those in computer science.

- For the treatment task (t_{18}=–6.77, *P*<.001, *d*=–3.03), the mean score was 0.85 (SD 0.15) for medical student-residents, compared with 0.1 (SD 0.31) for computer science students.
- For the diagnostic task (t_{18}=–9, *P*<.001, *d*=–4.02), the mean score was 0.9 (SD 0.31) for medical student-residents, compared with (SD 0) for computer science students.

For the learning task, even after applying the square root correction ($\sqrt{2}$), the homogeneity of variance was violated. Therefore, we applied the nonparametric Mann-Whitney test. Medical student-residents provided more relevant information elements than computer science students (U_{18}=30, *P*=.03, rank biserial correlation=0.4), with a mean of 3.9 (SD 0.31) for medical students and 3.5 (SD 0.52) for computer science students.

As illustrated in Multimedia Appendix 1 with 4 examples, the nature of the information elements provided for the learning task varied depending on the participants' domain of expertise. This task was exploratory and was approached differently based on the participants' expertise. Medical student-residents generally provided comprehensive responses specific to medicine, including definition and prevalence, symptoms, causes and consequences, diagnosis, and possible treatments. By contrast, computer science students primarily focused on definition, prevalence, and symptoms, with very few mentioning possible treatments. Their answers were more superficial compared with those of medical student-residents, who applied a medical framework to obtain information about the pathology.

- *RQ2: What types of resources do participants consult, and what is their credibility level when performing medical search tasks, considering their domain of expertise (medicine vs information search)?*

Credibility Resources Consulted, Scores, and Websites Used

First, regarding the credibility scores obtained (DV4a), medical student-residents scored higher than computer science students for the learning task, with a mean of 5.85 (SD 4.37) compared with 2.2 (SD 2; t_{18}=–2.40, *P*=.046, Cohen *d*=1.08). For the diagnostic task, only a marginal effect was observed (t_{18}=–1.48, *P*=.07), with a mean of 13.45 (SD 9.63) for medical student-residents versus 8 (SD 6.53) for computer science students.

For the treatment task, no significant difference was found between the 2 groups, with a mean of 6.9 (SD 4.67) for medical

XSL•FO
RenderX

student-residents and 5.4 (SD 4.09) for computer science students (t_{18}=–0.76, P=.45).

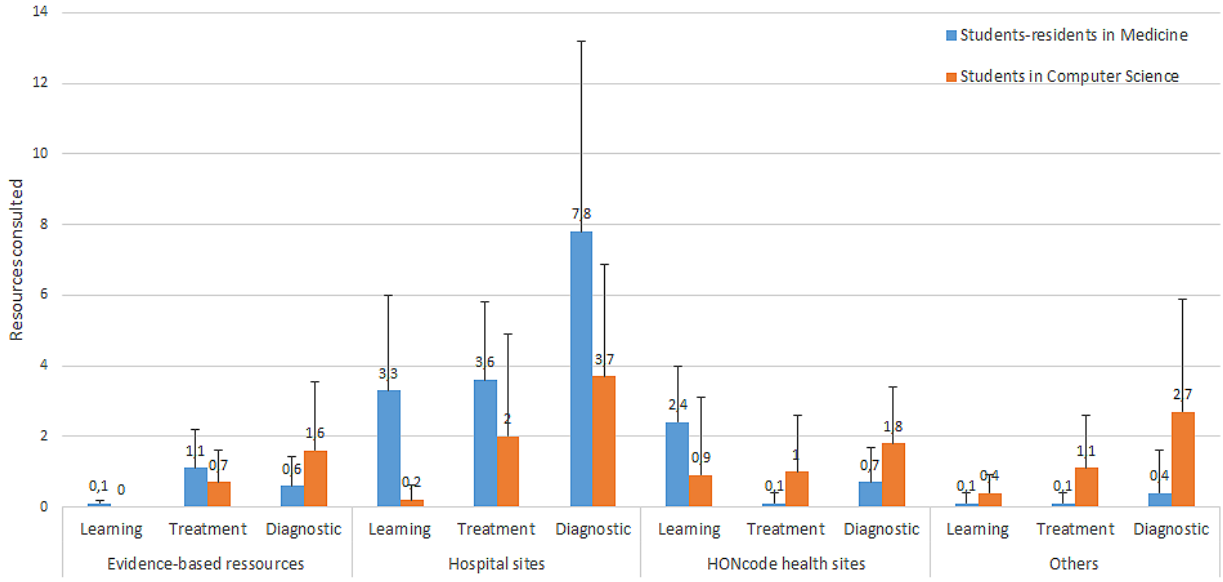
Next, we analyzed the distribution of internet resources used (DV4b; Figure 1). The resources did not hold the same prominence during the information search ($F_{3,54}$=28.12, P<.001, η^2p=0.61); hospital websites (P<.001) and documents (P<.001) were used more frequently than other sources.

We conducted a more detailed analysis of resource usage based on expertise and tasks. We observed:

- For evidence-based resources, surprisingly, no significant difference was found between medical student-residents and computer science students ($F_{1,18}$=0.31, P=.58). These resources were rarely visited by any participants.
- For hospital websites, medical student-residents used them more frequently than computer science students ($F_{1,18}$=13.96, P<.001, η^2p=0.44).
- For HONcode health websites, computer science students consulted these resources more frequently than medical student-residents ($F_{1,18}$=6, P=.02, η^2p=0.25).
- For other websites, computer science students used them more frequently than medical student-residents ($F_{1,18}$=5.21, P=.001, η^2p=0.34).

Next, qualitative analyses examined the main websites accessed through opened URLs (see Table 3).

Figure 1. Means (and SDs) of web resources consulted with regard to the expertise domain of students.



XSL•FO
RenderX

Table 3. The main websites visited by participants and their score with regard to the search task to be performed and expertise.

Tasks and websites (resources)		Authority versus nonauthority websites and scores obtained
Learning task		
Student-residents in medicine	• A government health agency [47] • Association of Physicians in Gynecology [48]	• Authority; score=1.5 points • Authority; score=1.5 points
Students in computer science	• Association of Patients With Endometriosis [49] • Nonofficial health websites (eg, [50,51])	• Nonauthority; score=1 points (verified by HONcode) • Nonauthority; score=0 points
Diagnostic task		
Student-residents in medicine	• Digital medical campus (not updated since 2016) [52] • A university hospital [53]	• Authority; score=1.5 points • Authority; score=1.5 points
Students in computer science	[54-56]	• Nonauthority; score=1 point (verified by HONcode) • For [55,56]: nonauthority; score=0 points
Treatment task		
Student-residents in medicine	• A Swiss medical journal [57] • A government health agency [47]	• Authority; score=2 points • Authority; score=1.5 points
Students in computer science	• A Swiss medical journal [57] • The other websites consulted are very different from one participant to another	• Authority; score=2 points

- *RQ3: Do they assess differently the reliability of resources they choose to use and the quality of answers they provided for the search task to be performed?*

Results of the self-assessment of the reliability of web documents consulted when answering this question at the end of each session (maximum of 7 points): a significant difference was found between medical student-residents and computer science students for the learning task (t_{18}=–3.13, P=.003, d=–1.39), with a mean of 6.8 (SD 0.42) for medical student-residents and 5.8 (SD 0.91) for computer science students. For the treatment and diagnostic tasks, statistical analyses did not reveal any significant differences between the 2 groups (t_{18}=–1.24, P=.12, and t_{18}=–1, P=.16; Table 4).

Table 4. Dependent variables for research question 3 with regard to the search task and expertise domain.

Research question 3 dependent variables and group of students	Learning task, mean (SD)	Diagnostic task, mean (SD)	Treatment task, mean (SD)
DV5: Self-assessment of resource reliability consulted (maximum score=7 points)			
Medicine	6.8 (0.42)[a]	6 (0.66)	5.4 (0.51)
Computer Science	5.8 (0.91)[a]	5.6 (1.07)	5.1 (0.56)
DV6: Self-assessment of quality of answer (maximum score=4 points)			
Medicine	1.84 (0.14)[a]	1.73 (0.13)[a]	1.79 (0.11)[a]
Computer Science	1.76 (0.08)[a]	1.31 (0.29)[a]	1.70 (0.10)[a]

[a]Significant differences between medical students/residents and computer science students.

Results of the self-assessment of the quality of answers provided on a 4-point scale for the 3 tasks; participants rated their perceived answer quality on a scale from 1 (very bad) to 4 (very good).

For all search tasks, medical student-residents rated the quality of their answers higher than computer science students:

- For the learning task, medical student-residents had a mean self-assessed answer quality of 3.5 (SD 0.52), compared with 3.1 (SD 0.31) for computer science students (t_{18}=–2.05, P=.03, d=–0.92).
- For the diagnostic task, medical student-residents had a mean self-assessed answer quality of 3.2 (SD 0.42), compared with 2.9 (SD 0.31) for computer science students (t_{18}=–1.8, P=.04, d=–0.80).
- For the treatment task, medical student-residents had a mean self-assessed answer quality of 2.9 (SD 0.31), compared with 1.8 (SD 0.78) for computer science students (t_{18}=–4.09, P<.001, d=–1.83).

Discussion

Principal Findings

Prior studies have suggested that physicians' primary challenge when searching for information online is their lack of information search skills [2,23], which may negatively impact

XSL•FO
RenderX

the quality of information found and, consequently, patient care. This study aimed to examine the strategies used by medical students and residents when searching for medical information online, the web resources they consulted (including the nature and credibility of these resources), how they assessed the reliability of these resources, and the quality of the answers they provided. The search behaviors of medical students and residents were compared with those of computer science students. While medical students and residents had greater medical knowledge, computer science students exhibited stronger information search skills but lower medical-related knowledge. The main findings indicated that medical students and residents outperformed computer science students. Overall, they found more relevant information, used more credible resources, rated these resources as more reliable, and had greater confidence in their answers than computer science students. Two particularly surprising and interesting results emerged: (1) medical students and residents generated more medical keywords than computer science students. These keywords, at least in part, helped them locate familiar websites, making them indirectly relevant to information searching; and (2) medical students and residents rarely used evidence-based resources, instead relying on familiar medical websites, such as hospital websites.

Interpretation of the Results Obtained

First, regarding information search performance, the results confirmed expectations: medical students and residents outperformed computer science students. They provided more correct answers for treatment and diagnostic tasks and included more relevant and structured elements in learning tasks—all without requiring additional time. Their deeper understanding of the search topics enabled them to deliver more structured and comprehensive responses. For example, medical students and residents frequently included details on symptoms, causes, diagnosis, and possible treatments. By contrast, computer science students' responses tended to be more superficial, focusing primarily on general statistics and potential consequences. These findings align with prior research on the positive effects of domain knowledge in generating more accurate and higher-quality search outcomes, regardless of the field of expertise [58-60]. Indeed, when processing information, individuals with domain expertise engage in top-down processing, which enhances reading comprehension and supports a deeper understanding of the content [61-63]. As a result, at the search outcome stage, domain experts can more effectively reuse extracted information to achieve their goals and produce higher-quality responses [60]. The advantages of domain expertise in medical information searches were also evident in keyword selection. Notably, medical students and residents used more medical-related keywords, particularly when researching a specific disease. These findings align with previous studies showing that medical experts employ a more specialized medical vocabulary than nonexperts [63-65]. The medical knowledge applied during query formulation allowed physicians to locate more precise information at a search speed comparable to that of computer science students. However, despite their speed, computer science students' searches tended to be far less precise.

Regarding these initial findings—the quality of final responses, keyword formulation, and search precision—it is worth noting that medical students accurately self-assessed their responses. Their higher confidence in the reliability of their answers, compared with computer science students, likely stemmed from their reliance on familiar medical websites. This underscores their ability to effectively locate and utilize credible information from online and hospital-based resources, as previously demonstrated [2].

When comparing the objective credibility scores of consulted web resources with participants' self-assessed reliability of those resources, the results showed that medical students and residents accessed more credible websites than computer science students for learning and diagnostic tasks. A significant difference was observed for learning tasks, while scores were comparable for diagnostic tasks. Learning and diagnostic tasks may require greater medical expertise to identify credible and reliable sources compared with treatment-related tasks. Accurate diagnosis often involves a careful differential diagnosis, requiring the exclusion of other conditions with similar symptoms. This process requires access to credible medical resources, as general websites are often insufficient. Medical students and residents effectively relied on reputable medical websites to access relevant case studies and other pertinent information. Likewise, learning about specific topics required current and reliable information, underscoring the need for high-quality medical sources.

A detailed analysis of the consulted websites revealed task-dependent differences between medical students/residents and computer science students. Notably, evidence-based resources were rarely used by any participants and were primarily consulted for diagnostic and treatment tasks. This finding was unexpected, as medical students and residents, given their rigorous training, were expected to prioritize evidence-based resources. While this result aligns with previous research highlighting the limited use of evidence-based resources among medical practitioners [66,67], it raises concerns about the effectiveness of current training programs in fostering evidence-based practice skills. As previously noted, physicians often struggle to formulate effective search queries in databases such as PubMed, leading to the retrieval of numerous irrelevant articles [68]. When using general search engines, they face 2 main challenges [11]: (1) managing and processing a vast amount of information and (2) assessing the reliability of content based on scientific evidence, all within a limited time frame. On average, physicians spend between 9 and 11 minutes searching the internet for medical information [2], a trend also observed among the medical students and residents in this study. To navigate these challenges, physicians tend to rely on familiar websites rather than exploring new content or consulting databases like PubMed [1,6,19].

The preference for familiar hospital websites observed in our study may offer a more streamlined and efficient—though potentially less comprehensive—pathway to relevant information. Our findings further support this trend, as medical students and residents consistently relied on familiar hospital resources across all search tasks, favoring established sources over broader exploration. In our experimental study, medical students and residents primarily consulted hospital-based resources for all tasks and did so more frequently than computer science students. This observation aligns with previous research

XSL•FO
RenderX

showing that health care professionals frequently rely on a limited set of preferred online resources [1,6,19]. One study found that physicians visit familiar websites to reduce search time [18], often going directly to trusted medical sources [69-71]. They apply various criteria to assess information quality (usefulness and accuracy) and authority (trustworthiness and credibility) [19], which may be particularly relevant when evaluating hospital websites and documents. Additionally, physicians tend to search for methodical and empirically grounded medical solutions to minimize errors in decision-making while maximizing diagnostic and therapeutic effectiveness, as previously demonstrated [6]. By contrast, our study found that computer science students relied more on HONcode-certified health websites and other unverified sources. Despite their advanced search skills, they may struggle to assess the credibility of medical websites or fully comprehend the information presented on platforms such as PubMed and hospital websites. Their limited medical knowledge likely hindered their ability to identify relevant and accurate medical information.

Consequently, the medical expertise of students and residents allowed them to achieve higher scores, evaluate resources more effectively, and select more relevant web sources compared with computer science students.

Implications

The results we obtained may have several implications.

First, a surprising finding was the limited use of evidence-based resources such as PubMed by medical students and residents. This may be due to the challenge of filtering relevant information from a vast number of irrelevant results, as previously documented [68]. Additionally, difficulties in effectively using search filters within these databases may have further hindered their ability to retrieve pertinent information. This finding underscores the need for enhanced training programs to equip health care professionals with the skills necessary to effectively utilize evidence-based resources. Integrating clinical and practical evidence-based medicine principles is crucial for updating medical knowledge, improving patient care, and fostering strong physician-patient relationships [72]. A study by the Institute for Healthcare Informatics [73] highlighted that physicians frequently rely on Wikipedia as a primary source of health care information, further emphasizing the need for targeted training initiatives. These programs should focus on strengthening information literacy skills, including the effective use of evidence-based databases such as PubMed and the critical evaluation of online health information. This could include instruction on HONcode principles, which help distinguish reliable from unreliable sources based on factors such as authorship, funding sources, and URL domains (.org, .gov, .net). Beyond their clinical practice, physicians who are well-versed in evaluating online information can also better support their patients. Many patients seek medical information on the internet, often without the ability to assess its accuracy [74], which can have negative consequences for their health care. Equipping physicians with the skills to identify credible and effective online resources would enable them to guide patients in interpreting the information they find online, ultimately improving patient education and decision-making. Physicians at all stages of their careers must stay updated on advancements in research, treatments, and best practices to ensure safe and effective patient care. Online scientific databases are essential tools in this process, particularly as the use of generative artificial intelligence (AI), such as ChatGPT (OpenAI, Inc.), continues to grow. Unlike Google, AI-generated responses often lack proper sourcing and reliability [75]. Given the evolving role of AI in knowledge acquisition, structured training programs are crucial to helping medical students and physicians navigate these resources effectively. Additionally, integrating verified sources—such as Google Scholar or PubMed—directly into general search engine results for medical queries could further support evidence-based learning and decision-making. General search engines could further support specialized medical queries by suggesting relevant medical keywords based on the task at hand. This approach would help medical students and physicians—who are already familiar with general search engine interfaces—find the required information more quickly than on PubMed, which they often find difficult to navigate efficiently. Physicians' preference for traditional search engines over PubMed may stem from their perception that PubMed is not well-suited to the practical demands of clinical medicine. General practitioners, in particular, require rapid access to high-quality content and do not have the time to sift through an overwhelming volume of information or conduct an extensive evaluation process [11]. Ideally, search engines should enhance usability by providing clear and precise indicators about the credibility and nature of retrieved results directly on the search results page. This includes specifying whether the content has undergone peer review and clarifying the level of empirical evidence presented in the source. Future research in design ergonomics and medical informatics should focus on developing and testing the impact of such enhanced search interfaces on physicians' search performance.

Limitations

This study had some limitations. The small sample size suggests that replicating the study with a larger and more diverse group of participants—including medical students, residents, and both junior and senior physicians—would be valuable. Given that physicians must continuously update their knowledge throughout their careers, future research should explore how search behaviors (eg, keyword formulation, websites visited) evolve at different professional stages, particularly in the context of emerging AI tools.

In addition to better distinguishing the impact of domain knowledge versus information search skills, future research could compare the performance of students in library and information science with that of medical students and residents. As library and information science students have specialized expertise in information retrieval and have outperformed computer science students in prior studies [33], such a comparison could provide further insights into the relationship between search expertise and domain-specific knowledge.

Although this study was conducted remotely—potentially raising concerns about data quality due to limited experimental control—the findings align with previous research. Notably, the search times recorded for medical students and residents in

XSL•FO
RenderX

this study were consistent with those reported in prior studies [2]. Moreover, the analysis of search times did not reveal any significant outliers, further supporting the reliability of the collected data.

Conclusions

The findings of this experimental study have significant implications for both current and future physicians.

A surprising finding was the limited use of evidence-based resources by medical students and residents. This may be due to the challenge of extracting relevant information from a high volume of irrelevant results, as previously documented [64]. This finding highlights the need for enhanced training programs, both at the university and at the postgraduate levels, to equip health care professionals with the skills necessary to effectively utilize evidence-based resources. These programs should focus on improving information search skills, including the efficient use of evidence-based databases and the critical evaluation of online health information. This could include instruction on HONcode principles, which help distinguish between reliable and unreliable sources based on factors such as authorship, funding sources, and URL domains. Beyond their clinical practice, knowledge of source evaluation criteria could also assist physicians in guiding patients who seek medical information online. This is especially important with the rise of AI tools, which are fundamentally reshaping how information is accessed and interpreted.

Training physicians and future physicians to effectively use these tools will also encourage them to expand the range of websites they consult. Rather than limiting themselves to familiar sources, such as hospital websites, they will be better equipped to explore a broader array of credible medical resources.

Acknowledgments

This work was supported by the French National Research Agency (ANR) CoST project (ANR-18-CE23-0016).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of answers provided by students-residents in medicine and students in computer science with their description.
[DOCX File , 24 KB - formative_v9i1e62754_app1.docx]

References

1. Bennett NL, Casebeer LL, Zheng S, Kristofco R. Information-seeking behaviors and reflective practice. J Contin Educ Health Prof 2006;26(2):120-127. [doi: 10.1002/chp.60] [Medline: 16802313]
2. Daei A, Soleymani MR, Ashrafi-Rizi H, Zargham-Boroujeni A, Kelishadi R. Clinical information seeking behavior of physicians: a systematic review. Int J Med Inform 2020 Jul;139:104144. [doi: 10.1016/j.ijmedinf.2020.104144] [Medline: 32334400]
3. Sarbaz M, Kimiafar K, Sheikhtaheri A, Taherzadeh Z, Eslami S. Nurses' information seeking behavior for clinical practice: a case study in a developing country. Stud Health Technol Inform 2016;225:23-27. [Medline: 27332155]
4. Magrabi F, Coiera EW, Westbrook JI, Gosling AS, Vickland V. General practitioners' use of online evidence during consultations. Int J Med Inform 2005 Jan;74(1):1-12. [doi: 10.1016/j.ijmedinf.2004.10.003] [Medline: 15626631]
5. Schilling LM, Steiner JF, Lundahl K, Anderson RJ. Residents' patient-specific clinical questions: opportunities for evidence-based learning. Acad Med 2005 Jan;80(1):51-56. [doi: 10.1097/00001888-200501000-00013] [Medline: 15618093]
6. Mikalef P, Kourouthanassis PE, Pateli AG. Online information search behaviour of physicians. Health Info Libr J 2017 Mar 07;34(1):58-73 [FREE Full text] [doi: 10.1111/hir.12170] [Medline: 28168831]
7. Heale BSE, Khalifa A, Stone BL, Nelson S, Del Fiol G. Physicians' pharmacogenomics information needs and seeking behavior: a study with case vignettes. BMC Med Inform Decis Mak 2017 Aug 01;17(1):113 [FREE Full text] [doi: 10.1186/s12911-017-0510-9] [Medline: 28764766]
8. Devi P, George J. Drug information needs of physicians treating diabetic nephropathy in a tertiary care hospital. Kathmandu Univ Med J (KUMJ) 2008;6(1):23-27. [Medline: 18604110]
9. Casebeer L, Bennett N, Kristofco R, Carillo A, Centor R. Physician internet medical information seeking and on-line continuing education use patterns. J Contin Educ Health Prof 2002;22(1):33-42. [doi: 10.1002/chp.1340220105] [Medline: 12004639]
10. Vallo Hult H, Master Östlund C, Pålsson P. Why and when physicians Google: resident physicians' information-seeking strategies during patient consultations. Stud Health Technol Inform 2023 Jun 29;305:580-583. [doi: 10.3233/SHTI230563] [Medline: 37387097]
11. Bernard E, Arnould M, Saint-Lary O, Duhot D, Hebbrecht G. Internet use for information seeking in clinical practice: a cross-sectional survey among French general practitioners. Int J Med Inform 2012 Jul;81(7):493-499. [doi: 10.1016/j.ijmedinf.2012.02.001] [Medline: 22425281]

XSL•FO
RenderX

12. Westbrook JI. Do online information retrieval systems help experienced clinicians answer clinical questions? Journal of the American Medical Informatics Association 2005 Jan 31;12(3):315-321. [doi: 10.1197/jamia.m1717]
13. Muse E, Godino J, Netting JAJ, Alexander JF, Moran HJ, Topol EJ. From second to hundredth opinion in medicine: a global consultation platform for physicians. npj Digital Med 2018 Oct 09;1(1):55 [FREE Full text] [doi: 10.1038/s41746-018-0064-y] [Medline: 31304334]
14. Del Fiol G, Mostafa J, Pu D, Medlin R, Slager S, Jonnalagadda SR, et al. Formative evaluation of a patient-specific clinical knowledge summarization tool. Int J Med Inform 2016 Feb;86:126-134 [FREE Full text] [doi: 10.1016/j.ijmedinf.2015.11.006] [Medline: 26612774]
15. Sun W, Ma R, Zhao X, Li Z, Lindqvist J, El AA, et al. Trusting the search: unraveling human trust in health information from Google and ChatGPT. arXiv :e1 Preprint posted online on March 15, 2024 [FREE Full text]
16. Häußler H. Do users trust search engines? And if so, why?: Developing a trust measure and applying it in an experiment. In: CHIIR '23: Proceedings of the 2023 Conference on Human Information Interaction and Retrieval. New York, NY: Association for Computing Machinery; 2023 Presented at: CHIIR '23: ACM SIGIR Conference on Human Information Interaction and Retrieval; March 19-23, 2023; Austin, TX p. 471-474. [doi: 10.1145/3576840.3578280]
17. Devine T, Broderick J, Harris LM, Wu H, Hilfiker SW. Making quality health websites a national public health priority: toward quality standards. J Med Internet Res 2016 Aug 02;18(8):e211 [FREE Full text] [doi: 10.2196/jmir.5999] [Medline: 27485512]
18. Schuers M, Griffon N, Kerdelhue G, Foubert Q, Mercier A, Darmoni SJ. Behavior and attitudes of residents and general practitioners in searching for health information: from intention to practice. Int J Med Inform 2016 May;89:9-14. [doi: 10.1016/j.ijmedinf.2016.02.003] [Medline: 26980354]
19. Hughes B, Joshi I, Lemonde H, Wareham J. Junior physician's use of Web 2.0 for information seeking and medical education: a qualitative study. Int J Med Inform 2009 Oct;78(10):645-655. [doi: 10.1016/j.ijmedinf.2009.04.008] [Medline: 19501017]
20. Sirohi D, Ng CHM, Bidargaddi N, Slater H, Parker MA, Hull ML, et al. High-quality eHealth websites for information on endometriosis: systematic search. J Med Internet Res 2024 Jan 10;26:e48243. [doi: 10.2196/48243] [Medline: 38198205]
21. Yuan X, White R. Building the trail best traveled: effects of domain knowledge on web search trailblazing. In: CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2012 Presented at: CHI '12: CHI Conference on Human Factors in Computing Systems; May 5-10, 2012; Austin, TX p. 1795-1804. [doi: 10.1145/2207676.2208312]
22. Bergmo TS, Sandsdalen V, Manskow US, Småbrekke L, Waaseth M. Internet use for obtaining medicine information: cross-sectional survey. JMIR Form Res 2023 Feb 02;7:e40466. [doi: 10.2196/40466] [Medline: 36729577]
23. Krause R, Moscati R, Halpern S, Schwartz D, Abbas J. Can emergency medicine residents reliably use the internet to answer clinical questions? West J Emerg Med 2011 Nov;12(4):442-447 [FREE Full text] [doi: 10.5811/westjem.2010.9.1895] [Medline: 22224135]
24. Szegőfi A, Heintz C. Institutions of epistemic vigilance: the case of the newspaper press. Social Epistemology 2022 Aug 23;36(5):613-628 [FREE Full text] [doi: 10.1080/02691728.2022.2109532]
25. Sperber D, Clément F, Heintz C, Mascaro O, Mercier H, Origgi G, et al. Epistemic vigilance. Mind and Language 2010 Sep;25(4):359-393. [doi: 10.1111/j.1468-0017.2010.01394.x]
26. Metzger MJ. Making sense of credibility on the web: models for evaluating online information and recommendations for future research. J Am Soc Inf Sci 2007 Sep 21;58(13):2078-2091. [doi: 10.1002/asi.20672]
27. Fogg BJ, Tseng H. The elements of computer credibility. In: CHI '99: Proceedings of the SIGCHI conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 1999 Presented at: CHI99: Conference on Human Factors in Computing Systems; May 15-20, 1999; Pittsburgh, PA p. 80-87. [doi: 10.1145/302979.303001]
28. Metzger MJ, Flanagin AJ. Credibility and trust of information in online environments: the use of cognitive heuristics. Journal of Pragmatics 2013 Dec;59:210-220. [doi: 10.1016/j.pragma.2013.07.012]
29. Rieh SY. Judgment of information quality and cognitive authority in the Web. J Am Soc Inf Sci 2001 Dec 14;53(2):145-161 [FREE Full text] [doi: 10.1002/asi.10017]
30. Babbie ER. The Basics of Social Research (7 edition). Belmont, CA: Wadsworth Publishing; 2016.
31. Trochim W, Donnelly J. The Research Methods Knowledge Base (3rd Edition). Mason, OH: Atomic Dog Publishing Inc; Dec 1, 2006:56-65.
32. Smith C. Domain‐independent search expertise: a description of procedural knowledge gained during guided instruction. Asso for Info Science & Tech 2014 May 19;66(7):1388-1405 [FREE Full text] [doi: 10.1002/asi.23272]
33. Aduba DE, Obot IN, Baro EE. A comparative study of information literacy skills among students of library and information science and computer science in Nigerian universities. Niger Delta J Libr Inf Sci 2022 Jun 30;3(1):1-15 [FREE Full text]
34. Aula A, Nordhausen K. Modeling successful performance in Web searching. Journal of the American Society for Information Science Technology 2006;57(12):1678-1693.
35. Hölscher C, Strube G. Web search behavior of internet experts and newbies. Computer Networks 2000 Jun;33(1-6):337-346. [doi: 10.1016/s1389-1286(00)00031-1]

XSL•FO
RenderX

36. White RW, Morris D. Investigating the querying and browsing behavior of advanced search engine users. In: SIGIR '07: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, NY: Association for Computing Machinery; 2007 Presented at: SIGIR07: The 30th Annual International SIGIR Conference; July 23-27, 2007; Amsterdam The Netherlands p. 255-262. [doi: 10.1145/1277741.1277787]
37. Ebiefung R, Adetimirin A, Oyewole O. Discipline variation in information retrieval skills and use of electronic information resources by undergraduates in University Of Uyo, Akwa Ibom State, Nigeria. Library Philosophy and Practice 2020:1-20 [FREE Full text] [doi: 10.1108/dlp-08-2020-0080]
38. Marchionini G. Information Seeking in Electronic Environments. Cambridge, UK: Cambridge University Press; 1995.
39. Fidel R, Pejtersen AM. From information behaviour research to the design of information systems: the Cognitive Work Analysis framework. Information Research. 2004. URL: http://InformationR.net/ir/10-1/paper210.html [accessed 2025-03-25]
40. Dosso C, Tamine L, Paubel P, Chevalier A. Navigational and thematic exploration–exploitation trade-offs during web search: effects of prior domain knowledge, search contexts and strategies on search outcome. Behaviour & Information Technology 2023 Aug 10;43(10):2232-2258. [doi: 10.1080/0144929x.2023.2242514]
41. Dosso C, Moreno JG, Chevalier A, Tamine L. CoST: an annotated data collection for complex search. In: CIKM '21: Proceedings of the 30th ACM International Conference on Information & Knowledge Management.: The 30th ACM International Conference on Information and Knowledge Management. ACM; 2021 Presented at: CIKM '21: The 30th ACM International Conference on Information and Knowledge Management; November 1-5, 2021; Virtual p. 4455-4464. [doi: 10.1145/3459637.3481998]
42. Figshare. URL: https://figshare.com/articles/Data [accessed 2025-03-24]
43. Rodon C, Meyer T. Self-efficacy about information retrieval on the web across all domains: a short measure in French and English. Behaviour & Information Technology 2018 Mar 16;37(5):430-444 [FREE Full text] [doi: 10.1080/0144929x.2018.1449252]
44. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. J Med Internet Res 2019 May 02;21(5):e12522. [doi: 10.2196/12522] [Medline: 31045507]
45. Health On the Net Foundation. URL: https://www.hon.ch [accessed 2025-03-25]
46. Jiang J, He D, Kelly D, Allan J. Understanding ephemeral state of relevance. In: CHIIR '17: Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval. New York, NY: Association for Computing Machinery; 2017 Presented at: CHIIR '17: Conference on Human Information Interaction and Retrieval; March 7-11, 2017; Oslo, Norway p. 137-146. [doi: 10.1145/3020165.3020176]
47. Haute Autorité de Santé. URL: https://www.has-sante.fr [accessed 2025-03-25]
48. CNGOF. URL: https://www.cngof.fr [accessed 2025-03-25]
49. Endofrance. URL: https://www.endofrance.org/ [accessed 2025-03-25]
50. PasseportSanté. URL: https://www.passeportsante.net/ [accessed 2025-03-25]
51. www. Pourquoi Docteur. URL: https://www.pourquoidocteur.fr [accessed 2025-03-25]
52. Le Centre de ressources et d'information sur les multimédias pour l'enseignement supérieur (CERIMES). URL: https://www.campus.cerimes.fr [accessed 2025-03-25]
53. CHU Toulouse. URL: http://www.medecine.ups-tlse.fr/ [accessed 2025-03-25]
54. Santé. URL: https://www.e-sante.fr/ [accessed 2025-03-25]
55. Journal des Femmes. URL: https://www.journaldesfemmes.fr/ [accessed 2025-03-25]
56. Wikipedia. URL: https://www.wikipedia.org/ [accessed 2025-03-25]
57. Revue Médicale Suisse. URL: https://www.revmed.ch/ [accessed 2025-03-25]
58. Mao J, Liu Y, Kando N, Zhang M, Ma S. How does domain expertise affect users' search interaction and outcome in exploratory search? ACM Trans Inf Syst 2018 Jul 17;36(4):1-30. [doi: 10.1145/3223045]
59. Wilson MJ, Wilson ML. A comparison of techniques for measuring sensemaking and learning within participant‐generated summaries. J Am Soc Inf Sci Tec 2013 Jan 04;64(2):291-306. [doi: 10.1002/asi.22758]
60. Wood E, De Pasquale D, Mueller JL, Archer K, Zivcakova L, Walkey K, et al. Exploration of the relative contributions of domain knowledge and search expertise for conducting internet searches. The Reference Librarian 2016 Feb 05;57(3):182-204. [doi: 10.1080/02763877.2015.1122559]
61. Amadieu F, Salmerón L, Cegarra J, Paubel P, Lemarié J, Chevalier A. Learning from concept mapping and hypertext: an eye tracking study. Educ Technol Soc 2015 Oct;18(4):100-112 [FREE Full text]
62. Bhattacharya N, Gwizdka J. Relating eye-tracking measures with changes in knowledge on search tasks. In: ETRA '18: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications. New York, NY: Association for Computing Machinery; 2018 Presented at: ETRA '18: 2018 Symposium on Eye Tracking Research and Applications; June 14-17, 2018; Warsaw, Poland p. 1-5. [doi: 10.1145/3204493.3204579]
63. White RW, Dumais ST, Teevan J. Characterizing the influence of domain expertise on web search behavior. In: WSDM '09: Proceedings of the Second ACM International Conference on Web Search and Data Mining. New York, NY: Association for Computing Machinery; 2009 Feb 9 Presented at: WSDM'09: Second ACM International Conference on Web Search and Web Data Mining; February 9-12, 2009; Barcelona, Spain p. 132-141. [doi: 10.1145/1498759.1498819]

XSL•FO
RenderX

64. Tamine L, Chouquet C. On the impact of domain expertise on query formulation, relevance assessment and retrieval performance in clinical settings. Information Processing & Management 2017 Mar;53(2):332-350. [doi: 10.1016/j.ipm.2016.11.004]
65. White R, Dumais S, Teevan J. How medical expertise influences web search interaction. In: SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. New York, NY: Association for Computing Machinery; 2008 Presented at: SIGIR '08: The 31st Annual International ACM SIGIR Conference; July 20-24, 2008; Singapore, Singapore p. 791-792 URL: http://susandumais.com/SIGIR2008-Expertise-pp879-white.pdf [doi: 10.1145/1390334.1390506]
66. Palotti J, Hanbury A, Muller H. Exploiting health related features to infer user expertise in the medical domain. In: Proceedings of WSCD Workshop on Web Search and Data Mining. Hoboken, NJ: John Wiley & Sons; 2014 Presented at: WSCD Workshop on Web Search and Data Mining; February 24-28, 2014; New York, NY p. e1 URL: https://tinyurl.com/4uj47xz4
67. Kritz M, Gschwandtner M, Stefanov V, Hanbury A, Samwald M. Utilization and perceived problems of online medical resources and search tools among different groups of European physicians. J Med Internet Res 2013 Jun 26;15(6):e122. [doi: 10.2196/jmir.2436] [Medline: 23803299]
68. Shariff SZ, Sontrop JM, Haynes RB, Iansavichus AV, McKibbon KA, Wilczynski NL, et al. Impact of PubMed search filters on the retrieval of evidence by physicians. CMAJ 2012 Feb 21;184(3):E184-E190. [doi: 10.1503/cmaj.101661] [Medline: 22249990]
69. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J 2008 Feb;22(2):338-342. [doi: 10.1096/fj.07-9492LSF] [Medline: 17884971]
70. De Leo G, LeRouge C, Ceriani C, Niederman F. Websites most frequently used by physician for gathering medical information. AMIA Annu Symp Proc 2006;2006:902 [FREE Full text] [Medline: 17238521]
71. Taubert M. Use of Google as a diagnostic aid. BMJ 2006 Dec 14;333(7581):1270.1-121270. [doi: 10.1136/bmj.39058.703194.3a]
72. Heighes PT, Doig GS. Intensive care specialists' knowledge, attitudes, and professional use of published research evidence: a mail-out questionnaire survey of appropriate use of research evidence in clinical practice. J Crit Care 2014 Feb;29(1):116-122. [doi: 10.1016/j.jcrc.2013.10.014] [Medline: 24331946]
73. Evans R. IMS Institute for Healthcare Informatics. Nursing Standard 2014 Apr 09;28(32):30-30. [doi: 10.7748/ns2014.04.28.32.30.s36]
74. Van Heel KL, Nelson A, Handysides D, Shah H. The factors associated with confidence in using the internet to access health information: cross-sectional data analysis. JMIR Form Res 2023 Apr 11;7:e39891 [FREE Full text] [doi: 10.2196/39891] [Medline: 37040161]
75. Hristidis V, Ruggiano N, Brown EL, Ganta SRR, Stewart S. ChatGPT vs Google for queries related to dementia and other cognitive decline: comparison of results. J Med Internet Res 2023 Jul 25;25:e48966 [FREE Full text] [doi: 10.2196/48966] [Medline: 37490317]

Abbreviations

AI: artificial intelligence
DV: dependent variable
FAIR: Findable, Accessible, Interoperable, and Reusable
HON: Health On the Net
IV: independent variable
MD: Doctor of Medicine
NHS: National Health Service
PhD: Doctor of Philosophy
RQ: research question

Edited by A Mavragani; submitted 30.05.24; peer-reviewed by M Tajafari, JL Claggett, S Ganesh; comments to author 07.10.24; revised version received 26.01.25; accepted 16.02.25; published 17.04.25.

Please cite as:
Chevalier A, Dosso C
The Influence of Medical Expertise and Information Search Skills on Medical Information Searching: Comparative Analysis From a Free Data Set
JMIR Form Res 2025;9:e62754
URL: https://formative.jmir.org/2025/1/e62754
doi:10.2196/62754
PMID:

XSL•FO
RenderX

©Aline Chevalier, Cheyenne Dosso. Originally published in JMIR Formative Research (https://formative.jmir.org), 17.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Application of an Innovative Methodology to Build Infrastructure for Digital Transformation of Health Systems: Developmental Program Evaluation

M Claire Buchan[1], BSc, MSc; Tarun Reddy Katapally[2,3,4], MS, PhD; Jasmin Bhawra[5], MSc, PhD

[1]School of Public Health Sciences, University of Waterloo, Waterloo, ON, Canada

[2]DEPtH Lab, Faculty of Health Sciences, Western University, London, ON, Canada

[3]Department of Epidemiology and Biostatistics, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada

[4]Lawson Health Research Institute, London, ON, Canada

[5]CHANGE Research Lab, School of Occupational and Public Health, Toronto Metropolitan University, Toronto, ON, Canada

Corresponding Author:
Jasmin Bhawra, MSc, PhD
CHANGE Research Lab, School of Occupational and Public Health
Toronto Metropolitan University
350 Victoria Street
Toronto, ON, M5B2K3
Canada
Phone: 1 416 979 5000 ext 553466
Email: jasmin.bhawra@torontomu.ca

Abstract

Background: The current public health crises we face, including communicable disease pandemics such as COVID-19, require cohesive societal efforts to address decision-making gaps in our health systems. Digital health platforms that leverage big data ethically from citizens can transform health systems by enabling real-time data collection, communication, and rapid responses. However, the lack of standardized and evidence-based methods to develop and implement digital health platforms currently limits their application.

Objective: This study aims to apply mixed evaluation methods to assess the development of a rapid response COVID-19 digital health platform before public launch by engaging with the development and research team, which consists of interdisciplinary researchers (ie, key stakeholders).

Methods: Using a developmental evaluation approach, this study conducted (1) a qualitative survey assessing digital health platform objectives, modifications, and challenges administered to 5 key members of the software development team and (2) a role-play pilot with 7 key stakeholders who simulated 8 real-world users, followed by a self-report survey, to evaluate the utility of the digital health platform for each of its objectives. Survey data were analyzed using an inductive thematic analysis approach. Postpilot test survey data were aggregated and synthesized by participant role.

Results: The digital health platform met original objectives and was expanded to accommodate the evolving needs of potential users and COVID-19 pandemic regulations. Key challenges noted by the development team included navigating changing government policies and supporting the data sovereignty of platform users. Strong team cohesion and problem-solving were essential in the overall success of program development. During the pilot test, participants reported positive experiences interacting with the platform and found its features relatively easy to use. Users in the community member role felt that the platform accurately reflected their risk of contracting COVID-19, but reported some challenges interacting with the interface. Those in the decision maker role found the data visualizations helpful for understanding complex information. Both participant groups highlighted the utility of a tutorial for future users.

Conclusions: Evaluation of the digital health platform development process informed our decisions to integrate the research team more cohesively with the development team, a practice that is currently uncommon given the use of external technology vendors in health research. In the short term, the developmental evaluation resulted in shorter sprints, and the role-play exercise enabled improvements to the log-in process and user interface ahead of public deployment. In the long term, this exercise informed the decision to include a data scientist as part of both teams going forward to liaise with researchers throughout the development

XSL•FO
RenderX

process. More interdisciplinarity was also integrated into the research process by providing health system training to computer programmers, a key factor in human-centered artificial intelligence development.

(JMIR Form Res 2025;9:e53339) doi:10.2196/53339

KEYWORDS

digital health platform; citizen science; evaluation; health systems; digital health; app; innovative; digital transformation; public health; crises; communicable disease; coronavirus; chronic diseases; decision-making; assessment; thematic analysis; self-report survey; risk; artificial intelligence; AI

Introduction

Background

The current public health crises we face are global in nature, ranging from communicable disease pandemics, such as COVID-19, to endemic noncommunicable diseases and long-term care burdens [1,2]. The presence of multiple overlapping health issues, termed a *syndemic*, is further complicated in the age of polycrises. Crisis responses increasingly require coordination from multiple sectors, including, but not limited to, health care, environment, and social services [3,4]. In order to address the gaps in our current systems of care, deliberate and cohesive societal efforts are required to understand and respond to existing inefficiencies in health systems.

Digital health platforms, which range from mobile health apps and virtual care products to digital health dashboards [5,6], have immense potential to transform our health systems by increasing access to care, predicting symptoms and outcomes, and enabling rapid responses to health crises [4,6-9]. For instance, digital health platforms can enable patients to connect with their health care providers remotely [10,11], and, in turn, health care providers are able to predict risks by ethically leveraging big data from patients [12,13] and provide support by engaging with patients remotely [10,14]. More importantly, the application of digital health platforms is not limited to patients; that is, they can be used by apparently healthy individuals to self-monitor and track their health behaviors and outcomes [15,16] as well as share their data ethically with health care providers and scientists [13,17]. Given the ubiquity of digital devices [18-20], their adaptability, and their reach across geographic regions and sociodemographic groups, digital health platforms are capable of bridging existing gaps in health information and care access [6,21,22]. More importantly, such platforms, particularly with the incorporation of artificial intelligence (AI) and machine learning, can enable precision prediction of health outcomes [23,24] as well as rapid responses to help monitor, mitigate, and manage existing and emerging health crises [6,25]. While the development of digital health platforms has increased significantly in the past decade, and in particular during the COVID-19 pandemic [26], there are no standardized processes for development and evaluation to ensure that evidence-based approaches are applied using interdisciplinary expertise.

The role of digital health platforms in managing public health and promoting health care access is predicted to grow exponentially [27]. Thus, it is critical that the development of these platforms is evaluated using rigorous methods to ensure effectiveness and efficacy. The World Health Organization recently released a guide for evaluating digital health interventions [28], and many other groups have adapted methods to assess digital platform technology [29-31]. However, these guidelines do not provide directions to evaluate digital platform *development processes*, including prototype development sprints and troubleshooting [28], which are key steps that come well ahead of the actual implementation of digital interventions.

Another key challenge is ensuring data privacy and data sovereignty of digital health platforms, particularly when serving communities that have been historically disenfranchised or discriminated against. Data sovereignty refers to meaningful control or ownership of one's own data and is a critical aspect of self-governance and self-determination, particularly among Indigenous and other colonized communities [32,33]. In creating digital health platforms that serve these communities, the development process thus requires the integration of rigid privacy and data protocols, in addition to consideration of culturally appropriate features tailored to communities' specific needs [4,7,32,34]. Moreover, as technology continues to evolve, it is crucial to evaluate the development and use of AI, especially to ensure that AI is designed in a way that is truly human centered [35,36]. Evaluating development processes helps to identify and mitigate the risks associated with AI, which could include neglecting user needs, overlooking the social impact, and failing to address biases and accountability issues [37]. By systematically evaluating development processes, we can foster AI systems that are better aligned with human values and needs (ie, health services) [36].

Study Objective

Overall, the evaluation of digital health platform development has enormous consequences for eventual citizen health and well-being, as well as data safety, security, and data sovereignty [38]. The potential for causing harm to populations, the sensitivity of personally identifiable big data that are collected via digital health platforms, as well as the need to develop human-centered AI require a rigorous evaluation process, including internal pilot testing before public testing. Most importantly, a significant gap exists in the peer-reviewed literature regarding evaluation approaches for the development of evidence-based digital health platforms, particularly in academic settings, where there is a critical need for co-design [39] not only from a citizen or patient perspective, but also by cohesive engagement between research and development teams. To address this critical gap in digital health platform development, this study aimed to apply mixed evaluation methods to assess the development of a digital health platform that was exclusively developed by a research and development

XSL•FO
RenderX

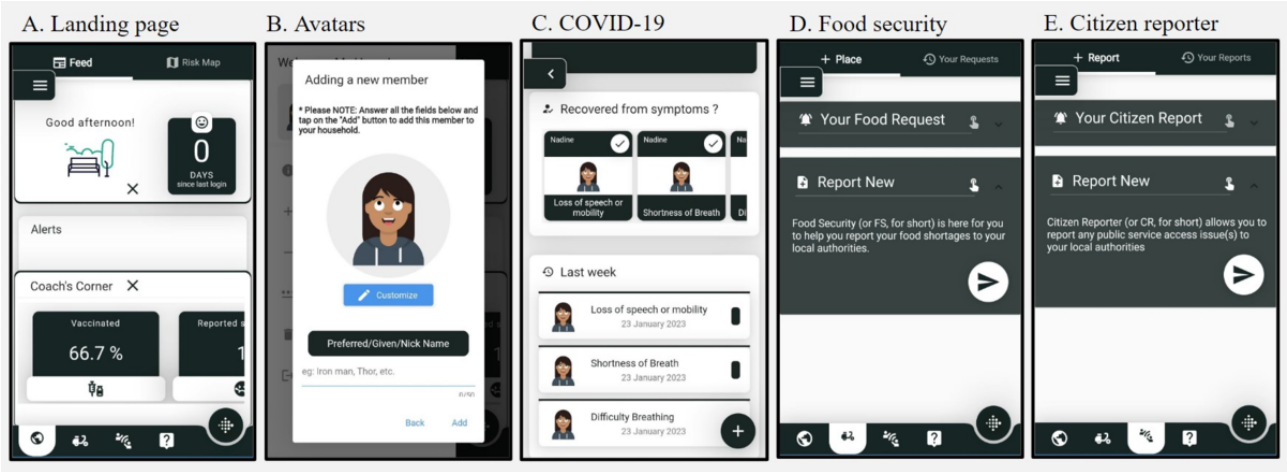
team working remotely during the COVID-19 pandemic to manage, monitor, and mitigate household risk of COVID-19.

Methods

CO-Away Digital Health Platform

In an effort to address the imminent public health crisis of the COVID-19 pandemic, the CO-Away digital health platform was developed to track, manage, and mitigate household risk of COVID-19. During the pandemic, rural, remote, and Northern communities in Canada were disproportionately impacted and experienced challenges with health information, data sovereignty, and care access [40]. In the Canadian context, Indigenous communities reported gaps in access to health information and care [34,41]; thus, after conducting a comprehensive needs assessment [34], the CO-Away platform was developed to enable near real-time monitoring, rapid response, and health care data access. CO-Away comprises a progressive web application for users to manage household COVID-19 risk as well as a backend digital decision-making dashboard that visualizes aggregated and anonymized big data relayed in real time from the progressive web application [6]. Progressive web applications are a type of web application that leverage web technologies to provide a more app-like user experience, combining the best features of web applications and mobile apps [42]. Figure 1 shows screenshots of the CO-Away home page and dashboard. The CO-Away platform provides local jurisdictional decision makers access to aggregate-level data to track and respond to emerging risk patterns and trends in near real time.

Figure 1. Screenshots of the CO-Away platform interface for community members, including the platform landing page, avatar creation process, COVID-19 symptom assessment, food security request, and citizen reporter feature.



Given the wide-ranging impacts of the COVID-19 pandemic on other aspects of health and wellness, CO-Away takes a holistic approach that extends beyond COVID-19 symptom assessment to include other key digital platform features, such as (1) food security request, which helps monitor and manage food shortages within a jurisdiction and (2) citizen reporter, which reports any public service or access issues experienced by community members [6]. CO-Away serves as a link to connect citizens with decision makers and offers value to both households and decision makers. This digital health platform, driven by citizen-generated big data collected from ubiquitous tools, not only supports citizens by providing them with real-time support and valuable insights to enhance their decision-making abilities, but also aggregates and anonymizes citizen data. These aggregated, anonymized data are then delivered to decision makers, enabling real-time exchanges of information and alerts through direct bidirectional engagement between citizens and decision makers. This innovative approach represents a paradigm shift in the way community health is approached, placing priority on addressing the immediate needs of citizens. This study evaluated the development processes of both the app that citizens or community members use and the digital health dashboard that decision makers use. Additional details on the CO-Away platform can be found in another publication [43].

Evaluation Approach

To examine the evolution, challenges, successes, and utility of the CO-Away platform, a mixed methods developmental evaluation was performed after the completion of the first prototype to inform subsequent iterations of the platform. Developmental evaluations generate learnings to inform the development of an initiative; thus, are often used in complex and unpredictable scenarios, such as the COVID-19 pandemic [44,45]. A developmental evaluation is utilization focused and should be designed and implemented in ways that maximize utility for the primary intended users [46]. This evaluation approach generates data and findings in near real time, thereby facilitating developmental decision-making and course corrections throughout the development process [44,47].

Evaluation Questions

This evaluation was guided by the following overarching evaluation questions:

- What factors influenced the CO-Away digital health platform development?
- How has the CO-Away digital platform development evolved over the course of the project?
- To what extent does the CO-Away digital health platform achieve its goals and objectives from the perspective of community members or users and decision makers?

XSL•FO
RenderX

- How do these findings translate to the development process?
- What direction will the CO-Away project take going forward?

Evaluation Design

The evaluation included the following two key components: (1) a qualitative survey assessing program objectives, modifications, and challenges administered to 5 key members of the software development team and (2) a role-playing pilot test conducted with key stakeholders who were part of the research team, followed by a self-report survey, to evaluate the utility of the program for each of its goals or objectives.

Data Collection

Data collection for the CO-Away developmental evaluation included two primary components: (1) a development team survey (refer to Multimedia Appendix 1 for the survey) and (2) a research team pilot test.

Part A: Development Team Survey

A total of 5 members of the CO-Away software development team were asked to complete a web-based self-report survey. The survey was designed to assess the evolution of program objectives, target audience, and key modifications made throughout the development process. A guiding framework developed for conducting needs assessments to conceptualize and implement digital infrastructure was used to guide the development of this evaluation survey [34]. These key team members were identified by the principal investigator and the evaluation team and included software developers and data scientists. Participants were all involved in the development of CO-Away and project organization or management. Due to the potential for bias that could arise from having key members of the development team involved in the data collection process, participants were not involved in the preparation of the evaluation survey and were blinded to the responses of their fellow interviewees.

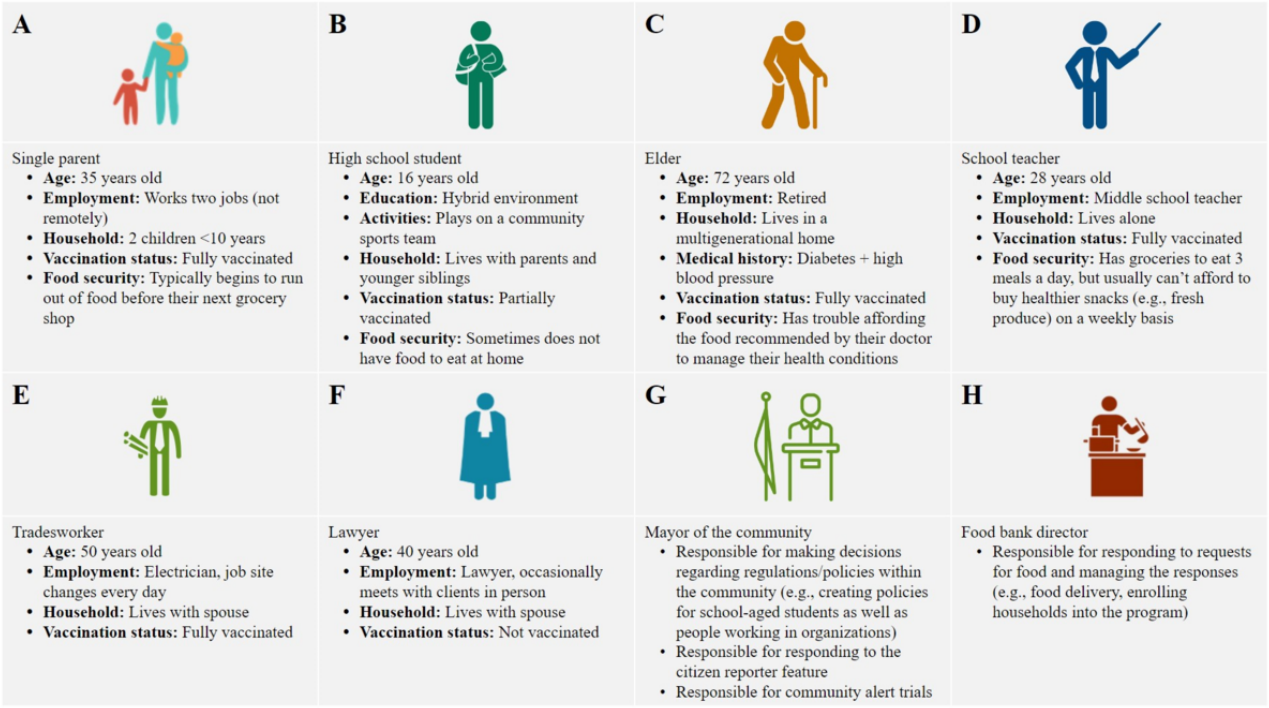
Participants were asked to complete the survey within a 1-week period from July 18, 2022, to July 25, 2022. The survey was designed to be completed in <60 minutes and was administered through Qualtrics (Qualtrics International Inc) [48]. Participants were asked to describe their role in the digital platform design, any major deviations from the initial project proposal, any changes in the target audience, and the digital platform's potential for impact using open text fields. Participants were also asked to reflect on the barriers to and facilitators of developing the CO-Away platform within the context of the COVID-19 pandemic.

Part B: Pilot Test

To evaluate the utility of the CO-Away platform for each of its objectives, a role-playing pilot test was conducted with 7 key stakeholders who were part of the interdisciplinary research team. All project stakeholders had expertise in digital health, epidemiology, and public health. Project stakeholders assumed the role of 1 of the 8 character roles developed by the evaluation team. These character roles consisted of 6 (75%) community members and 2 (25%) decision makers. One project stakeholder was asked to assume 2 roles within the pilot test, both of a community member and a decision maker, to reflect a real-world scenario where decision makers can be community members requiring access to both digital dashboards, that is, dual roles. The evaluation team developed the characters to simulate a range of potential user backgrounds and provided each participant with a brief overview of their character detailing their age, household composition, occupation, remote work ability, vaccination status, and food security status (Figure 2). Participating stakeholders reviewed the list of character descriptions and submitted their preferred roles to the evaluation team.

XSL•FO
RenderX

Figure 2. Overview of the 8 pilot test character descriptions.



Participants were assigned their character role for the duration of the pilot test and were asked to use their provided character descriptions to create avatars within the CO-Away app. Participants were advised to use their assigned character description as a starting point and to interact with the platform as they saw fit to simulate real-world conditions. Participants were instructed to improvise their engagement with specific features of the platform (eg, COVID-19 diagnoses and symptoms). The participant who assumed both a decision maker and community member role set up 2 distinct profiles and was asked to engage with the platform (to the best of their ability) under each role independently. The pilot test was conducted over a 2-week period from July 8, 2022, to July 21, 2022. Over the course of the activity, participants were sent daily reminders via email to input their data into the app using the integrated features.

Following the pilot test, participants were asked to complete a brief self-report survey detailing their experience using the app by giving them a variety of response options, including yes or no and Likert scale questions. The survey assessed the general usability of the app and specific app features, including COVID-19 risk, food security, and citizen reporter features. Two surveys were developed by the evaluation team, one for community member participants (Multimedia Appendix 2) and the other for the decision maker participants **(**Multimedia Appendix 3). The survey administered was dependent on the participants' assumed roles in the study (ie, community members or decision makers), while some general questions were asked of both groups, and most questions differed between groups. The one participant who assumed both a community member and a decision maker role was asked to complete both surveys. The survey was designed to be completed in 30 minutes and was distributed electronically using Qualtrics [48].

By accessing the cloud-based database [6], daily user engagement reports were generated over the course of the pilot test to examine user engagement across the features of the platform. Reports included data on user engagement (eg, the number of new sign-ups and total log-ins), platform usability (eg, the number of households reporting their behaviors, the number of food requests, and the number of food requests processed and delivered), and reported issues (eg, the number of reported issues using the platform).

Ethical Considerations

The research and development team who participated in this project either provided simulated data as part of a role play scenario, or were subject matter experts, respectively. This developmental evaluation enabled testing of the functionality, usability, and areas of improvement for the digital platform, thus standard research ethics requirements did not apply [48]. However, survey response data remained anonymous to protect participant privacy, and survey responses were not associated with individual software development team members. Moreover, all data collected in the role-play exercise were simulated hypothetical scenarios, did not reflect true participant behaviors, and thus did not require strict anonymity and privacy protocols. Participants were not compensated for their participation in this evaluation.

Analysis

Data from the development team survey were downloaded from Qualtrics and synthesized in Microsoft Word. Data were analyzed using an inductive thematic analysis approach. To generate an initial coding manual, 1 evaluation team member (MCB) independently reviewed the first survey and organized the data into themes and meaningful codes. A second evaluation team member (JB) subsequently reviewed the coded transcript to ensure consistency and agreement. Once all transcripts were

XSL•FO
RenderX

coded by 1 evaluation team member (MCB), the evaluation team (MCB and JB) reconvened to discuss relationships between codes that could be grouped into themes.

Data from the postpilot test survey were downloaded from Qualtrics and aggregated in Microsoft Excel. Data were analyzed and summarized for each question and results were grouped by community members and decision makers. Result tables were sorted according to question topic areas (eg, COVID-19, food security, and citizen reporter). Quotations in the result tables indicated participant responses to open text fields on the survey. For questions where participants could select more than 1 option, responses were totaled across categories, and, as such, percentages may add to more than 100.

Results

Part A: Development Team Survey

Overview

All 5 members of the software development team completed the survey. The roles of the participating team members included the principal investigator (n=1, 20%), 3 (60%) software developers—one of whom was involved in project organization and management—and 1 (20%) project coordinator. The 5 subsequent subsections describe the key themes that arose through analysis of the development survey.

Theme 1: Alignment With Original Objectives

Over the course of the development process, several modifications to the app structure were made, many of which were made to accommodate the evolving nature of the COVID-19 pandemic. These changes included modifying existing features of the app (ie. the COVID-19 feature) and the addition of new features (eg, the food security feature and citizen reporter). Despite these modifications, the overall objective of CO-Away remained consistent throughout development. One participant noted the following:

> *The prototype satisfies the originally planned platform's mandate. The changes that took place were more logistical, which were beyond the scope of the development team. The critical takeaway is that the concept and the ultimate impact remain the same.* [DT5]

In addition, the team noted improvements in the visual appearance of the app. Overall, the development team indicated that the final app not only met the original objectives but also the plan expanded to accommodate the evolving needs of the community and COVID-19 regulations.

Theme 2: Project Challenges

There were a series of challenges identified by the development team, including tailoring the platform to a diverse audience (ie, both young and older community members), learning new software development tools (eg, programming languages), and the strict regulations surrounding COVID-19–related platforms on mobile app stores. The CO-Away app required integrating various disciplines to address community health needs, which was described as a unique challenge:

> *As this project's goal is to evaluate a COVID-19 mitigating digital health platform for Indigenous self-governance, determination, and data sovereignty in a remote [Indigenous] community...the initiative from its inception has been complex not only due to the integration of multiple disciplines [epidemiology, data science, Indigenous health, public health, among others], but also due to the ultimate purpose of the digital platform: rapidly responding to community needs.* [DT5]

A digital platform developed for a remote Indigenous community app would have a substantially greater chance for success if it took a holistic approach, reiterating Traditional Indigenous Knowledge of holistic health. This approach required continuous collaboration with the citizen scientist advisory council [6,34], and a part of this process required developing strategies and features within the app that manage and mitigate other population health crises that were worsened by the COVID-19 pandemic (eg, food security, mental health and substance misuse, and negative interactions of youth with law enforcement).

One of the key challenges that was pertinent for COVID-19 app development, in particular, was constantly evolving government policies. Participants reported that several previously existing features required repeated modifications, and features not initially outlined in the project plan required development on the go. For example, one participant commented that calculating their risk of COVID-19 required incorporating evolving guidelines on vaccination dosage:

> *...changes with increase in the number of doses of vaccination.* [DT4]

The nature of the COVID-19 pandemic not only required the software development team to adapt to changing government policies but also to the evolution of the virus variants themselves. Another participant highlighted the addition of the vaccine passport feature to the app, which was not within the original project scope.

While the ability of team members to work remotely was noted as a facilitator of project execution, participants underscored the delays that resulted from the constant need to update program features to align with government policies. Of high importance were the challenges associated with supporting the data sovereignty of the platform users. One participant commented the following:

> *We have had to go above and beyond existing data safety regulations in the world to build a digital health platform that provides Indigenous citizens control and ownership over their data.* [DT5]

Despite the plethora of challenges posed by the COVID-19 pandemic, the software development team was able to adapt their approach and practices to build a successful app.

Theme 3: Factors Influencing App Development

There was a range of factors that influenced app development, particularly for an app focused on addressing the rapidly changing virus during the COVID-19 pandemic. In particular,

XSL•FO
RenderX

participants noted that working during the COVID-19 pandemic posed specific barriers to the recruitment of developers. Moreover, navigating the development process dominated by, and dependent on big technology companies (Big Tech), as well as a shift to a new academic institution for the principal investigator were key external factors that influenced overall app development. Recruitment challenges and a change in academic institution created logistical challenges that added to the complexity of this large-scale project:

> *Due to the restriction of movement during the pandemic, we faced significant difficulties to recruit high-quality personnel. This issue is especially challenging in the disciplines of computer science and data science, where most of the personnel who fit our criteria are recent international graduates, who face IRCC [Immigration, Refugees and Citizenship Canada] backlogs for work permits.* [DT5]

One challenge noted by several participants was launching a platform in Big Tech app stores, which ultimately resulted in the creation of a progressive web application. For example, one participant stated the following:

> *To truly ensure self-determination and data sovereignty, we had to abandon launching our digital health platform on Apple Store and Google PlayStore to eliminate the control of these stores over the development process. This decision meant the development of a progressive web-based platform that does not need these stores to be launched and the platform would only be provided to the community members via a password protected process.* [DT5]

The 2 primary app hosting Big Tech companies, Apple Store and Google Play Store, implemented strict regulations around launching COVID-19–related platforms on their mobile app stores, which hindered the development process. While this transition created delays and roadblocks that the development team needed to overcome, it was essential in honoring the team's commitment to data sovereignty and self-determination.

Overall, the development team indicated that strong communication, workload distribution and delegation, and continuous evaluation were key internal factors that positively contributed to the platform development:

> *The communication was really strong between the team despite working from home and not working physically in the same place.* [DT2]

Participants also noted that having an agile development mindset as a team was helpful, particularly when collaborating with a citizen scientist advisory council consisting of project stakeholders. Digital literacy was also identified as an important component of app usability by the development team. Therefore, it was important for the development team to increase the accessibility of the app by modifying traditional surveys (ie, sociodemographic questions) to include more visuals and graphics while minimizing text where possible (Figure 1).

Theme 4: Project Successes and Areas of Improvement

Participants noted three primary moments of success: (1) the development of the household risk management feature, (2) the launch of the progressive web application, and (3) the clearance of the final software tests. One participant noted, in particular, the following:

> *The constant variation of policies and virus strains resulted in the creation of household risk management, rather than individual risk management, which I think is going to change the way we develop risk management tools for infectious diseases in the future.* [DT5]

When asked which factors stood out in making the development process successful, participants underscored the importance of the team dynamic. In particular, participants noted the strong leadership, cohesion, ability to problem solve, and their commitment and dedication.

Participants indicated that improving the efficiency of the team overall could be beneficial moving forward, "work on developing a more efficient workload distribution model" (DT4), particularly through simplifying platform features or implementing agile methodology [49] (ie, managing a project by breaking it up into several phases).

One participant noted that the team changed their internal sprint processes and incorporated internal evaluations of ongoing progress:

> *We had to...rewrite digital application development sprint protocols.* [DT5]

As described in theme 3, the development team experienced challenges launching health-focused apps, particularly restrictions for COVID-19 apps; thus, the team had to shift to a progressive web application. One participant noted that if they were to restart the project today, they would start with a progressive web application as it would have streamlined project development quite substantially.

Theme 5: App Launch Requirements

The participants identified a range of factors that would enable a successful launch. Most of these factors focused on successful uptake; hence, the responses described the importance of awareness and education. When asked to describe some of the factors that would contribute to a successful app launch, participants noted the platform's adaptability (ie, "rapid adaptation to changing scenarios"), consistent team engagement, and project planning. Strong community leadership and participation was highlighted as an important contributor to the overall success of the platform. One participant noted the following:

> *The most important factor that I think will contribute to a successful app launch is education and awareness [of the community in which the app is launched].* [DT3]

Taken together, these results suggested that despite the many challenges faced by the development team, their strong team cohesion, communication, and problem-solving were all essential in the overall success of program development.

XSL•FO
RenderX

Part B: Pilot Test

A total of 6 stakeholders of the interdisciplinary research team participated in the pilot test, playing the role of citizens or community members and decision makers to simulate real-world conditions. Participant character profiles included a diverse range of ages, occupations, household compositions, COVID-19 risk factors, and food security statuses (Figure 2). Over the course of the 2-week pilot test, there were an average of 7 (SD 3) log-ins per day across an average of 4 users per day (SD 1). The largest proportion (average: 40%, SD 18%) of the daily log-ins occurred in the evening.

In total, 5 (63%) of the 8 participants reported COVID-19 diagnoses, with 21 hospital visits reported over the duration of the pilot test. Among the 8 participants, 2 (25%) were fully vaccinated, 3 (38%) were fully vaccinated with boosters, and 2 (25%) were unvaccinated. Over the duration of the 2-week pilot test, participants reported inconsistently wearing face masks (n=38), making essential shopping trips (n=48), having outdoor visits with other people (n=49), having social gatherings (n=38), and travelling (n=11).

A total of 4 food security requests were placed over the 2-week period, all of which were processed by decision makers using the digital health dashboard. One participant submitted a citizen report, which was processed by a decision maker, and 5 community alerts were issued, all of which were categorized as high urgency.

Results from the postpilot survey are presented in Tables S1 and S2 in Multimedia Appendix 4 for community members and decision makers, respectively.

Community Members

Usability

All (6/6, 100%) participants felt that their avatars accurately reflected their COVID-19 risk, found the consent process clear, and appreciated the anonymity it provided (Figure 1B). All (6/6, 100%) participants reported that the app was easy to navigate and that they could find each feature when needed. Participants found it straightforward to set up their avatars, taking ≤5 minutes. In total, 33% (2/6) of the participants created avatars for multiple household members, while the other 67% (4/6) created avatars for themselves only. According to participants, the ideal frequency of notifications from the app was once per day. No (0/6, 0%) participants reported any issues in using the app.

Feature-Specific Feedback

Most (4/6, 67%) of the participants reported that they were comfortable interacting with the platform's COVID-19 feature, with 33% (2/6) of the participants reporting that they were neither comfortable nor uncomfortable. All (6/6, 100%) participants felt that the recommendations for their COVID-19 risk were clear and easy to understand, though some (2/6, 33%) provided feedback to increase convenience and subsequent usability (eg, readily accessible links or saved log-in information). Nearly all participants (5/6, 83%) were comfortable sharing their vaccination status. Issues such as random logouts and confusion with report creation were raised, though the feature was still found to be easy and clear to use.

Most (5/6, 83%) of the participants used the food security feature, all of which found it "easy" to use. A total of 60% (3/5) of the participants expressed comfort with using this feature, with the 40% (2/5) remaining participants reporting that they were "neither comfortable nor uncomfortable." Moreover, 60% (3/5) of the participants reported that the response of the decision makers was "neither appropriate nor inappropriate," while the remaining 40% (2/5) of the participants reported "appropriate" responses. Similarly, half (3/5, 60%) of the participants were "comfortable" revealing their identity when support was needed, while the remainder (2/5, 40%) felt "neither comfortable nor uncomfortable." When asked about room for improvement, 40% (2/5) of the participants suggested the inclusion of more detailed options for food security requests, 40% (2/5) of the participants reported issues viewing responses to their submitted food requests, and 20% (1/5) of the participants desired additional instructions for interacting with the feature.

Of the 6 participants, 3 (50%) used the citizen reporter feature and felt comfortable doing so. A total of 2 (67%) of the 3 participants found it easy to use, and 1 (33%) participant experienced difficulty during the report submission process, expressing uncertainty regarding the completeness of their submission due to the lack of response. Moreover, 1 (33%) participant reported an appropriate response from the decision maker, while the remaining 2 (67%) participants reported neither appropriate nor inappropriate responses. In total, 2 (67%) of the 3 participants felt comfortable revealing their identity when seeking support, while the third (33%) participant felt neither comfortable nor uncomfortable.

Overall Feedback

Participants generally found the app useful, providing clear information and serving as a good resource. The straightforwardness, accessibility, and organization of the app were highlighted along with the easy navigation and relevant tips. However, there were mentions of difficulty in finding where to input information initially and occasional log-in issues requiring password resets. Some (2/6, 33%) participants also mentioned challenges when using the app on a phone, particularly related to limited screen visibility.

Decision Makers

Usability

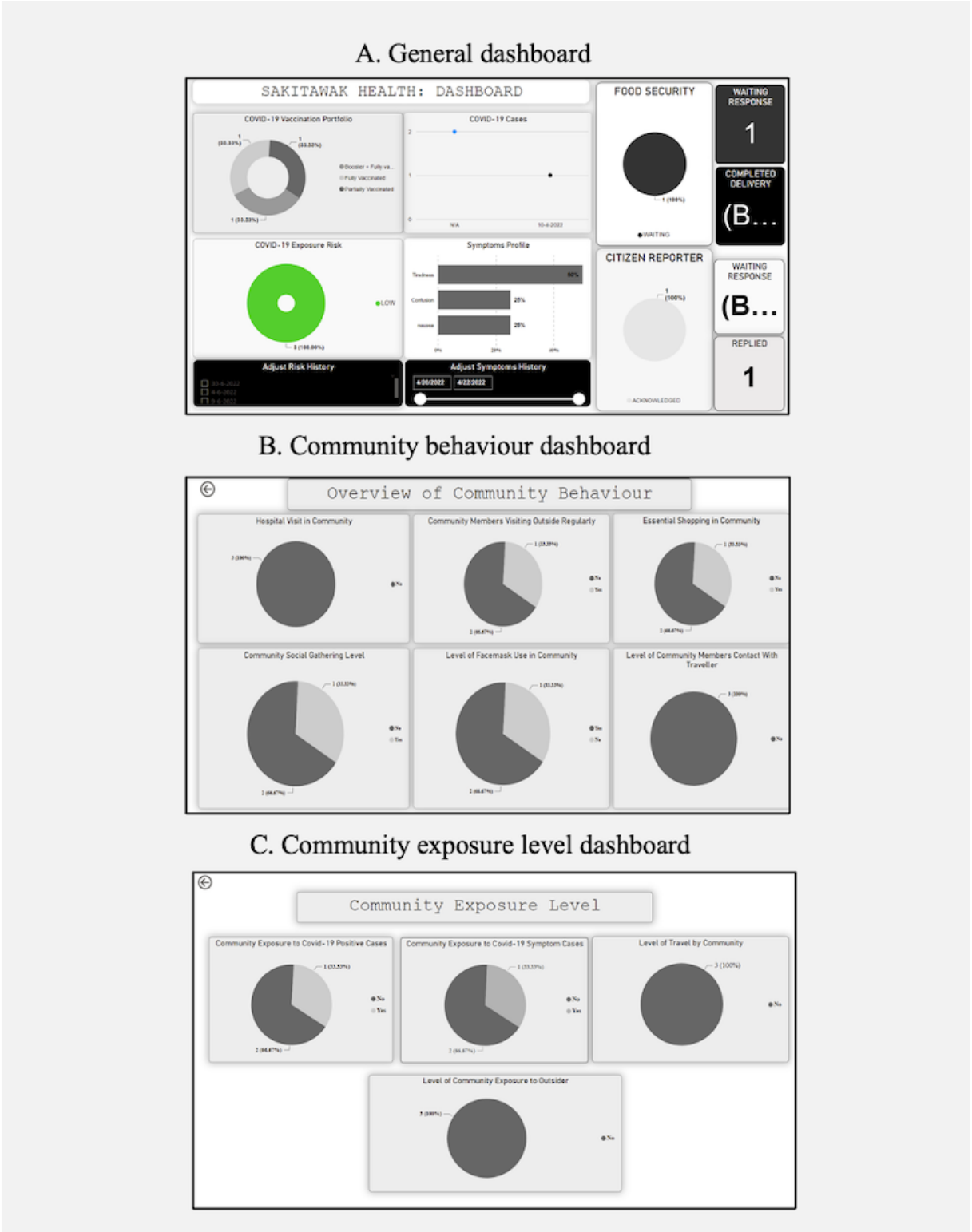
Decision makers reported on the usability of the digital dashboard, including ease of finding information, navigation, and overall use. Both decision makers found the information used to create their avatars accurate in reflecting their COVID-19 risk. They were satisfied with the clarity of the consent process and believed the app was clear and easy to navigate, ensuring anonymity and accessibility to all features. No issues or improvement suggestions were reported regarding navigability, and participants felt that they knew who to contact with questions regarding the app, their data, or their rights. Both decision makers had no difficulty interacting with notifications, finding them organized and user-friendly.

XSL•FO
RenderX

Feature-Specific Feedback

Figure 3 shows examples of how community member data were aggregated and visualized for decision makers to review. Participants had a positive experience with the dashboard data visualizations of the dashboard, finding them easy to understand and appreciating the level of control they had over them. Decision makers felt confident in using the presented data to make informed decisions regarding the community's response to the COVID-19 pandemic. Response rates to individual incidents were relatively short, typically taking less than 5 minutes to complete. No suggestions for improvement or increasing participants' confidence were provided.

Figure 3. Screenshots of the CO-Away platform interface for decision makers, including the general dashboard, the community behavior dashboard, and the community exposure level dashboard.



XSL•FO
RenderX

Similar to the COVID-19 feature, both decision makers found the data visualizations easy to understand and appreciated the appearance and level of control over them. Decision makers expressed confidence in using the presented data for informed decision-making regarding the COVID-19 pandemic response. Response rates for individual incidents were deemed prompt, with completion times less than 5 minutes. No suggestions for improving confidence or the visualizations were given.

Only 1 of the 2 decision makers interacted with the citizen reporter feature, reported that the data visualizations were easy to understand, and appreciated their appearance and level of control. The decision maker expressed confidence in using the presented data for informed decision-making, and the participant considered the response times for individual incidents, completed in less than 5 minutes, to be appropriate. No suggestions for improving the visualizations or participant confidence were given.

Overall Feedback

Participants had a positive overall experience with the platform, with a participant suggesting the addition of reference material, such as a brief presentation or a short video, for future use. The food security section was found to be easy to navigate and straightforward, providing users with simplified access. However, there were uncertainties regarding the ability to delete requests and whether citizens could contact local decision makers and/or food services for questions or concerns, as a participant did not receive a notification when attempting to do so.

Discussion

Principal Findings

Evaluation of the CO-Away digital health platform enabled systematic capture of the development team's perspective on the overall development process as well as the testing of platform usability from multiple stakeholders simulating real-world scenarios before finalizing platform features. This evaluation demonstrated that while the final digital health platform met its original objectives, it had to be expanded to accommodate the evolving needs of potential users and the COVID-19 pandemic regulations. Strong team cohesion, communication, and problem-solving were essential to the overall success of platform development. The pilot test found that participants reported positive experiences interacting with the platform and found its features relatively easy to use, with a few community members reporting challenges interacting with the interface. Users in the decision maker role found the data visualizations intuitive in helping them to understand the information.

Relevance and Application of Findings

The developmental evaluation findings will not only influence the decisions of both research and development teams going forward, but the process described for conducting an assessment of the CO-Away development phase also importantly advances a replicable methodology that can inform empirical development, implementation, and evaluation of digital health platforms [6]. The field of digital health evaluation is rapidly evolving; therefore, emerging research on innovations in digital technology and evaluation practices to assess impact and utility [28,38,50-52] have emerged over the past 5 years. The World Health Organization [28] has released guidelines for monitoring and evaluating digital health interventions, with many existing mixed methods (ie, qualitative focus groups and quantitative surveys) being applied to digital health platform evaluations. However, there is little guidance on how and when to apply these evaluation methods in the scheme of a digital health project [53,54]. In particular, there is a lack of peer-reviewed literature describing the evaluation of the development of evidence-based digital health platforms.

When working with marginalized populations, such as Indigenous or global south communities, evaluation of the development process is a critical step in ensuring that a digital platform is safe, accessible, and not widening existing inequities. Specific considerations related to data ownership, cultural competency, and privacy must be built into the digital platform development protocol and rigorously tested [34,52,53]. Thus, this developmental evaluation was designed by initially referring to existing evaluation literature to inform the key evaluation questions and study approach [38,44-47]. Given the dearth of evidence on *how* to practically adapt and apply digital evaluation methods [53], our study aimed to address this gap by implementing an innovative mixed methods approach informed by previous evaluation work in Indigenous communities [34] and evaluation literature [29,44-47,55-59].

The use of role-play has been demonstrated in other contexts [55-59] where the application of a digital intervention or program can be simulated before public deployment. For example, during the evaluation of a digital platform designed to improve the reproductive health of adolescents in Rwanda, a role-play exercise was used during the prototyping stage of development to evaluate the end-to-end user experience [29]. During the role-play exercise, participants were provided with mock scenarios and roles to help the design team understand how users interacted with the interface during the improvised scenario-specific performances. Similar to the approach used in this developmental evaluation, feedback was requested from participants following the role-play exercise to gain an understanding of user experience and areas of improvement. This step is essential not only when working with marginalized communities but also for launching digital platforms focused on sensitive health issues, that is, when potential risks are heightened, thereby requiring more rigorous testing protocols. Identifying qualified participants for role-play can be a challenge in research studies; therefore, this evaluation uniquely used an interdisciplinary team to conduct the exercise during the development phase to minimize commonly reported challenges with role-play and simulate a real-world scenario before the public launch [56].

It was through this evaluation that our internal testing of the platform generated key learnings that informed specific project decisions, including integrating the research team more cohesively with the development team. In academic research, a team of external technology vendors is commonly used to develop digital platforms. As a result, the external development team is often not embedded in the research and evaluation

XSL•FO
RenderX

process. In order to effectively evaluate the digital development process, our team identified the need to involve a data scientist as part of both the development and research teams going forward (whereas previously, only a researcher was liaising with the development team). Including a common computer programmer or developer in both teams improved communication and coordination during the development process. By bridging these traditionally distinct teams, we have been able to maximize cross-disciplinary collaboration, with research and practice informing development decisions. This integration also proved instrumental in aligning the platform's features with users' needs and developmental feasibility, that is, integrating interdisciplinarity into the development process.

In recognizing the potential of concurrent health domain expertise and software development skills, we initiated a health system training for all computer programmers [60,61]. This training contributed to increasing interdisciplinary understanding and communication and was deemed necessary for the creation of user-centred CO-Away feature development, which considered distinct risks that could be introduced with AI systems [36,37]. This training not only supports the programmers' comprehension of the health context but also provides them with the ability to envision the platform's functionalities from the user's perspective. This is a key step in both co-designing digital infrastructure as well as designing human-centered AI [35,62], which is a gap in the current literature [63-66]. In general, this approach resulted in digital health platform features that were user-friendly, that is, an improved user experience, which is critical to the success of platform use and implementation [64,67].

The developmental evaluation also changed development sprint processes [49,68], which paved the way for shorter sprints with a quick internal evaluation of ongoing progress. Software development sprints, which are part of the agile coding process, are implemented as part of the agile methodology, an iterative project management framework that breaks projects down into several dynamic phases, commonly known as "sprints" [49]. When the team started the CO-Away development process, sprints were approximately 4 to 5 months, but the developmental evaluation findings informed our decision to shorten the sprints to 3 to 4 months to enable faster internal testing and modification of the digital platform. This decision has allowed the software development team to pivot quickly when necessary, responding to stakeholder feedback and incorporating changes to the digital platform iteratively.

The evaluation provided noteworthy insights into the dynamics of our team, particularly the role that strong communication played throughout the development process. The COVID-19 pandemic presented a unique set of challenges that required adaptive strategies [69-71]. Despite the physical distancing mandates, our team managed to function cohesively, leveraging digital tools, regular virtual meetings, and transparent channels of communication, that is, open discussion of weekly updates and key decisions accessible to all team members on a virtual portal. The development process not only involved remote hiring of our team [72] but also remote functioning [73] and retention of our team even after the pandemic restrictions ended, lessons which we implemented into the continuous functioning of our development processes. For instance, we currently work in a hybrid setting [74], which leverages the flexibility of virtual work as well as in-person brainstorming, a system that has further improved communication and collaboration with research and development teams.

Perhaps the defining element that shaped the development process of the CO-Away platform was the complexities that arose with its COVID-19–specific focus. The evolving nature of the coronavirus [75-77], coupled with the rapidly changing landscape of pandemic management policies [78-80], presented a constant need for flexibility and adaptability. The virus's mutations necessitated continuous updates to the platform's algorithms to ensure that users would have access to the most up-to-date information. Similarly, the ever-changing policies surrounding lockdowns [81,82], vaccine requirements [83,84], and mask mandates [85,86] required the platform's features to be responsive to shifting guidelines [6]. Navigating these complexities highlighted the importance of further adapting the agile methodology [49,68] throughout the platform design and development process to accommodate the changing COVID-19 landscape.

The role-play pilot test highlighted the effectiveness of adhering to user-centered design principles throughout the development of the CO-Away digital health platform. The stakeholders who participated in the pilot reported overall positive reactions to the platform avatars, transparent consent procedures, and intuitive navigation—aspects that reiterated the importance of tailoring the platform design to users' needs and ensuring a user-friendly experience [64,67]. While participants generally found the app user-friendly, their feedback on minor obstacles, such as entering initial information and sporadic log-in difficulties, highlighted the importance of iterative improvement during the development process. This constructive feedback provided an opportunity for the development team to adapt CO-Away to meet the needs of users.

The mixed feedback regarding participants' comfort levels when interacting with the platform features reveals the importance of balancing interactivity with user comfort. For instance, while some participants were comfortable engaging with various features and sharing personal information for support, others expressed hesitancy. This finding highlights the significance of providing users with control over their data, a critical aspect of data sovereignty [32] that has been built into platform development after this vital input.

While we did not observe marked differences in the way participants engaged with the platform based on the assigned participant roles, it is worth noting that there were differences in platform use between decision makers and community members. These variations align with expectations stemming from the differing ways these distinct user groups engage with the platform's features. For instance, decision makers are less focused on their own privacy as they are not sharing data, while the community members are sharing data. We anticipate that various community members will interact with and use the app according to their specific circumstances and needs. As such, the potential differences in app use based on community member

XSL•FO
RenderX

roles will be further explored in the future during the external community pilot phase.

Finally, the digital health dashboard feature of CO-Away offered decision makers with valuable tools for informed decision-making, especially within the context of COVID-19 response and community management. The pilot decision makers appreciated the usability of the platform and expressed confidence in using the data visualizations for informed decision-making. However, it is important to recognize that although some real-world scenarios were simulated in the pilot, a larger external community pilot will require more complex decision-making due to the increased scale of implementation. However, the key goal of the pilot was to test the functionality of a cloud-based digital health dashboard to engage with community members and send real-time alerts, features that were successfully tested and functionality that was confirmed in the evaluation of the pilot. Real-world policy changes using real-time big data would require stakeholder buy-in and addressing nuanced matters, such as food security and citizen reporter reports, demanding sensitivity and community-wide consideration. Further exploration of these dynamics will be carried out during our upcoming external community pilot.

Key Guidelines for Digital Health Platform Development

On the basis of the findings of this developmental evaluation of a digital health platform development, we suggest the following guidelines:

1. Integration of research and development teams is key to the success of digital health platforms, where the software developers are provided an opportunity to understand scientific goals, and the research stakeholders are made privy to cloud computing challenges.
2. Conducting a developmental evaluation after the completion of a prototype facilitates the incorporation of research and development team perspectives empirically into the iterative digital development process before a product is tested widely in the community.
3. Conducting an internal role-play pilot simulating real-world conditions is critical to test not only the functionality of digital health platforms but also the nuanced perceptions of potential users, which will be essential for successful community implementation.

Strengths and Limitations

This developmental evaluation has several notable strengths. First, we deployed a mixed methods approach to data collection and analysis; open-ended survey questions to capture participant perceptions of the platform and its usability and CO-Away use metrics data provided an objective assessment of platform engagement. The integration of both quantitative and qualitative data collection methods enriches the findings generated by the evaluation. Second, the inclusion of development team surveys allowed for an in-depth understanding not only of CO-Away's developmental process but also offered valuable insights that have and will continue to inform improvements in the development processes. Finally, the use of "character roles" was used to simulate a community pilot test. Although not a substitute for a real-world pilot, this approach served as a valuable first step for exploring real-world simulations using varied user interactions and experiences, an approach that will shape a future external community pilot. Providing fictional role-play characters through this remote evaluation also avoided potential issues, such as staying in character [56]. There were also several study limitations. The participation of the stakeholders in the pilot test may have affected the nature of the issues reported, that is, more focus on app function than readability. In order to minimize potential bias, the evaluation questions were blinded, and the digital platform development team did not participate in the pilot test. Moreover, the stakeholders who participated in the pilot did not analyze the evaluation data. The use of "character roles," while insightful, does not capture all user scenarios that would arise in a real-world setting. As such, it is possible that certain user interactions and challenges were not captured in this evaluation and will be explored during our external community pilot. Notably, the sample size for each population was small, so caution should be taken when interpreting or comparing percentages. Another limitation pertains to the digital literacy of participants. While the pilot participants did not have previous experience using the platform, it is possible that few issues were reported due to the inherent high digital literacy of the interdisciplinary stakeholders. In future digital evaluations, it would be imperative to not only capture the digital literacy of platform users but also to enable the increase of digital literacy through innovative digital literacy programs [87].

Conclusions

The evaluation of digital health platform development is critical for the success of not only platform functionality but also eventual implementation and scale-up. The innovative mixed methods approach applied in this evaluation combines both perspectives of the development team as well as a real-world simulation of platform users to extend key guidelines for digital health platform development. For sensitive community health issues, role-play simulation and testing before public launch can avoid potential risks and improve user experience. The evaluation informed several key decisions of the digital health platform development, including integrating the research team more cohesively with the development team, which resulted in a data scientist being part of both teams going forward. Another key development process decision was to integrate more interdisciplinarity into the development process by providing health system training to computer programmers and shortening development sprints. Combined with more efficient sprints, the role-play exercise importantly enabled improvements to the log-in process and the user interface ahead of public deployment.

XSL•FO
RenderX

Acknowledgments

The authors would like to acknowledge Nadine Elsahli for her support with this manuscript, including the creation of figures to visually depict the digital health platform design. The authors would also like to thank all members who participated in the developmental evaluation process and provided valuable feedback throughout.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Development survey.
[DOCX File , 14 KB - formative_v9i1e53339_app1.docx]

Multimedia Appendix 2
Postpilot survey—community member.
[DOCX File , 17 KB - formative_v9i1e53339_app2.docx]

Multimedia Appendix 3
Postpilot survey—decision maker.
[DOCX File , 18 KB - formative_v9i1e53339_app3.docx]

Multimedia Appendix 4
Evaluation survey results.
[DOCX File , 152 KB - formative_v9i1e53339_app4.docx]

References

1. Alokaily F. COVID-19 global health crisis. Saudi Med J 2021 Jan 25;42(1):3-4 [FREE Full text] [doi: 10.15537/smj.2021.1.25663] [Medline: 33399164]
2. Noncommunicable diseases. World Health Organization. 2024 Dec 23. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2025-03-22]
3. Lawrence M, Homer-Dixon T, Janzwood S, Rockstrom J, Renn O, Donges JF. Global polycrisis: the causal mechanisms of crisis entanglement. SSRN J 2023 Jun 27 [FREE Full text] [doi: 10.2139/ssrn.4483556]
4. Katapally TR. A global digital citizen science policy to tackle pandemics like COVID-19. J Med Internet Res 2020 May 26;22(5):e19357 [FREE Full text] [doi: 10.2196/19357] [Medline: 32408267]
5. Sheikh A, Anderson M, Albala S, Casadei B, Franklin BD, Richards M, et al. Health information technology and digital innovation for national learning health and care systems. Lancet Digit Health 2021 Jun;3(6):e383-e396. [doi: 10.1016/s2589-7500(21)00005-4]
6. Katapally TR, Ibrahim ST. Digital health dashboards for decision-making to enable rapid responses during public health crises: replicable and scalable methodology. JMIR Res Protoc 2023 Jun 30;12:e46810 [FREE Full text] [doi: 10.2196/46810] [Medline: 37389905]
7. Katapally TR. The SMART framework: integration of citizen science, community-based participatory research, and systems science for population health science in the digital age. JMIR Mhealth Uhealth 2019 Aug 30;7(8):e14056 [FREE Full text] [doi: 10.2196/14056] [Medline: 31471963]
8. Stoumpos AI, Kitsios F, Talias MA. Digital transformation in healthcare: technology acceptance and its applications. Int J Environ Res Public Health 2023 Feb 15;20(4):3407 [FREE Full text] [doi: 10.3390/ijerph20043407] [Medline: 36834105]
9. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug 07;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]
10. Rosenlund M, Kinnunen UM, Saranto K. The use of digital health services among patients and citizens living at home: scoping review. J Med Internet Res 2023 Mar 27;25:e44711 [FREE Full text] [doi: 10.2196/44711] [Medline: 36972122]
11. Jnr BA. Use of telemedicine and virtual care for remote treatment in response to COVID-19 pandemic. J Med Syst 2020 Jun 15;44(7):132 [FREE Full text] [doi: 10.1007/s10916-020-01596-5] [Medline: 32542571]
12. Batko K, Ślęzak A. The use of big data analytics in healthcare. J Big Data 2022 Jan 06;9(1):3 [FREE Full text] [doi: 10.1186/s40537-021-00553-4] [Medline: 35013701]
13. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. J Big Data 2019 Jun 19;6(1):54. [doi: 10.1186/s40537-019-0217-0]
14. Burton L, Rush KL, Smith MA, Davis S, Rodriguez Echeverria P, Suazo Hidalgo L, et al. Empowering patients through virtual care delivery: qualitative study with micropractice clinic patients and health care providers. JMIR Form Res 2022 Apr 27;6(4):e32528 [FREE Full text] [doi: 10.2196/32528] [Medline: 35413002]

XSL•FO
RenderX

15. Kuwabara A, Su S, Krauss J. Utilizing digital health technologies for patient education in lifestyle medicine. Am J Lifestyle Med 2020 Dec 13;14(2):137-142 [FREE Full text] [doi: 10.1177/1559827619892547] [Medline: 32231478]
16. Mair JL, Salamanca-Sanabria A, Augsburger M, Frese BF, Abend S, Jakob R, et al. Effective behavior change techniques in digital health interventions for the prevention or management of noncommunicable diseases: an umbrella review. Ann Behav Med 2023 Sep 13;57(10):817-835 [FREE Full text] [doi: 10.1093/abm/kaad041] [Medline: 37625030]
17. Riso B, Tupasela A, Vears DF, Felzmann H, Cockbain J, Loi M, et al. Ethical sharing of health data in online platforms - which values should be considered? Life Sci Soc Policy 2017 Aug 21;13(1):12 [FREE Full text] [doi: 10.1186/s40504-017-0060-z] [Medline: 28825221]
18. Mosa AS, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
19. Langford AT, Solid CA, Scott E, Lad M, Maayan E, Williams SK, et al. Mobile phone ownership, health apps, and tablet use in US adults with a self-reported history of hypertension: cross-sectional study. JMIR Mhealth Uhealth 2019 Jan 14;7(1):e12228 [FREE Full text] [doi: 10.2196/12228] [Medline: 31344667]
20. Kaplan WA. Can the ubiquitous power of mobile phones be used to improve health outcomes in developing countries? Global Health 2006 May 23;2:9 [FREE Full text] [doi: 10.1186/1744-8603-2-9] [Medline: 16719925]
21. Willis VC, Thomas Craig KJ, Jabbarpour Y, Scheufele EL, Arriaga YE, Ajinkya M, et al. Digital health interventions to enhance prevention in primary care: scoping review. JMIR Med Inform 2022 Jan 21;10(1):e33518 [FREE Full text] [doi: 10.2196/33518] [Medline: 35060909]
22. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2021;2:100117 [FREE Full text] [doi: 10.1016/j.sintl.2021.100117] [Medline: 34806053]
23. Shahid N, Rappon T, Berta W. Applications of artificial neural networks in health care organizational decision-making: a scoping review. PLoS One 2019;14(2):e0212356 [FREE Full text] [doi: 10.1371/journal.pone.0212356] [Medline: 30779785]
24. Vazquez-Ingelmo A, Garcia-Penalvo FJ, Theron R. Information dashboards and tailoring capabilities - a systematic literature review. IEEE Access 2019;7:109673-109688. [doi: 10.1109/access.2019.2933472]
25. Kostkova P. Disease surveillance data sharing for public health: the next ethical frontiers. Life Sci Soc Policy 2018 Jul 04;14(1):16 [FREE Full text] [doi: 10.1186/s40504-018-0078-x] [Medline: 29971516]
26. Bouabida K, Lebouché B, Pomey MP. Telehealth and COVID-19 pandemic: an overview of the telehealth use, advantages, challenges, and opportunities during COVID-19 pandemic. Healthcare (Basel) 2022 Nov 16;10(11):2293 [FREE Full text] [doi: 10.3390/healthcare10112293] [Medline: 36421617]
27. Abernethy A, Adams L, Barrett M, Bechtel C, Brennan P, Butte A, et al. The promise of digital health: then, now, and the future. NAM Perspect 2022 Jun 27;2022(22):10.31478/202206e [FREE Full text] [doi: 10.31478/202206e] [Medline: 36177208]
28. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. World Health Organization. 2016 Dec 12. URL: https://www.who.int/publications/i/item/9789241511766 [accessed 2023-10-31]
29. Ippoliti N, Sekamana M, Baringer L, Hope R. Using human-centered design to develop, launch, and evaluate a national digital health platform to improve reproductive health for Rwandan youth. Glob Health Sci Pract 2021 Nov 29;9(Supplement 2):S244-S260. [doi: 10.9745/ghsp-d-21-00220]
30. Beynon F, Guérin F, Lampariello R, Schmitz T, Tan R, Ratanaprayul N, et al. Digitalizing clinical guidelines: experiences in the development of clinical decision support algorithms for management of childhood illness in resource-constrained settings. Glob Health Sci Pract 2023 Jul 27;11(4):e2200439. [doi: 10.9745/ghsp-d-22-00439]
31. Tamrat T, Chandir S, Alland K, Pedrana A, Taighoon Shah M, Footitt C, et al. Digitalization of routine health information systems: Bangladesh, Indonesia, Pakistan. Bull World Health Org 2022 Oct 01;100(10):590-600. [doi: 10.2471/blt.22.287816]
32. Bhawra J. Decolonizing digital citizen science: applying the bridge framework for climate change preparedness and adaptation. Societies 2022 Apr 17;12(2):71. [doi: 10.3390/soc12020071]
33. Hummel P, Braun M, Tretter M, Dabrock P. Data sovereignty: a review. Big Data Soc 2021 Jan 22;8(1). [doi: 10.1177/2053951720982012]
34. Bhawra J, Buchan MC, Green B, Skinner K, Katapally TR. A guiding framework for needs assessment evaluations to embed digital platforms in partnership with Indigenous communities. PLoS One 2022 Dec 22;17(12):e0279282 [FREE Full text] [doi: 10.1371/journal.pone.0279282] [Medline: 36548382]
35. Bingley WJ, Curtis C, Lockey S, Bialkowski A, Gillespie N, Haslam SA, et al. Where is the human in human-centered AI? Insights from developer priorities and user experiences. Comput Hum Behav 2023 Apr;141:107617. [doi: 10.1016/j.chb.2022.107617]
36. Ozmen Garibay O, Winslow B, Andolina S, Antona M, Bodenschatz A, Coursaris C, et al. Six human-centered artificial intelligence grand challenges. Int J Hum Comput Interact 2023 Jan 02;39(3):391-437. [doi: 10.1080/10447318.2022.2153320]
37. Chen Y, Clayton EW, Novak LL, Anders S, Malin B. Human-centered design to address biases in artificial intelligence. J Med Internet Res 2023 Mar 24;25:e43251 [FREE Full text] [doi: 10.2196/43251] [Medline: 36961506]
38. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851. [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]

XSL•FO
RenderX

39. Sanz MF, Acha BV, García MF. Co-design for people-centred care digital solutions: a literature review. Int J Integr Care 2021 Apr 30;21(2):16 [FREE Full text] [doi: 10.5334/ijic.5573] [Medline: 33981193]
40. The impact of COVID-19 on rural and remote mental health and substance use. Mental Health Commission of Canada. 2021. URL: https://mentalhealthcommission.ca/wp-content/uploads/2021/10/The-Impact-of-COVID-19-on-Rural-and-Remote-Mental-Health-and-Substance-Use.pdf.pdf [accessed 2023-10-31]
41. Huyser KR, Yellow Horse AJ, Collins KA, Fischer J, Jessome MG, Ronayne ET, et al. Understanding the associations among social vulnerabilities, indigenous peoples, and COVID-19 cases within Canadian health regions. Int J Environ Res Public Health 2022 Sep 29;19(19):12409 [FREE Full text] [doi: 10.3390/ijerph191912409] [Medline: 36231708]
42. Rêgo F, Portela F, Santos MF. Towards PWA in healthcare. Procedia Comput Sci 2019;160:678-683. [doi: 10.1016/j.procs.2019.11.028]
43. Bhawra J, Elsahli N, Patel J. Applying digital technology to understand human experiences of climate change impacts on food security and mental health: scoping review. JMIR Public Health Surveill 2024 Jul 23;10:e54064 [FREE Full text] [doi: 10.2196/54064] [Medline: 39042453]
44. Dozois E, Langlois M, Blanchet-Cohen N. DE 201: a practitioner's guide to developmental evaluation. The J.W. McConnell Family Foundation and the International Institute for Child Rights and Development. 2010. URL: https://mcconnellfoundation.ca/wp-content/uploads/2017/07/DE-201-EN.pdf [accessed 2025-03-22]
45. Patton MQ. Emergent developmental evaluation developments. J Multidiscip Eval 2021 Jun 01;17(41):23-34 [FREE Full text] [doi: 10.56645/jmde.v17i41.699]
46. Patton MQ. Utilization-Focused Evaluation. Thousand Oaks, CA: SAGE Publications; 2008.
47. Patton MQ. Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use. New York, NY: Guilford Publications; 2010.
48. Tri-council policy statement: ethical conduct for research involving humans. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council. 2018. URL: https://ethics.gc.ca/eng/documents/tcps2-2018-en-interactive-final.pdf [accessed 2025-04-02]
49. Nazir S, Price B, Surendra NC, Kopp K. Adapting agile development practices for hyper-agile environments: lessons learned from a COVID-19 emergency response research project. Inf Technol Manag 2022 Jul 30;23(3):193-211 [FREE Full text] [doi: 10.1007/s10799-022-00370-y] [Medline: 36188730]
50. Kolasa K, Kozinski G. How to value digital health interventions? A systematic literature review. Int J Environ Res Public Health 2020 Mar 23;17(6):2119 [FREE Full text] [doi: 10.3390/ijerph17062119] [Medline: 32209988]
51. Bashi N, Fatehi F, Mosadeghi-Nik M, Askari MS, Karunanithi M. Digital health interventions for chronic diseases: a scoping review of evaluation frameworks. BMJ Health Care Inform 2020 Mar;27(1):e100066 [FREE Full text] [doi: 10.1136/bmjhci-2019-100066] [Medline: 32156751]
52. Labrique AB, Wadhwani C, Williams KA, Lamptey P, Hesp C, Luk R, et al. Best practices in scaling digital health in low and middle income countries. Global Health 2018 Nov 03;14(1):103 [FREE Full text] [doi: 10.1186/s12992-018-0424-z] [Medline: 30390686]
53. Guo C, Ashrafian H, Ghafur S, Fontana G, Gardner C, Prime M. Challenges for the evaluation of digital health solutions-a call for innovative evidence generation approaches. NPJ Digit Med 2020 Aug 27;3(1):110 [FREE Full text] [doi: 10.1038/s41746-020-00314-2] [Medline: 32904379]
54. Wang T, Giunti G, Goossens R, Melles M. Timing, indicators, and approaches to digital patient experience evaluation: umbrella systematic review. J Med Internet Res 2024 Feb 05;26:e46308 [FREE Full text] [doi: 10.2196/46308] [Medline: 38315545]
55. Skoura-Kirk E, Brown S, Mikelyte R. Playing its part: an evaluation of professional skill development through service user-led role-plays for social work students. Soc Work Educ 2020 May 21;40(8):977-993. [doi: 10.1080/02615479.2020.1764521]
56. Seland G. Empowering end users in design of mobile technology using role play as a method: reflections on the role-play conduction. In: Kurosu M, editor. Human Centered Design. Berlin, Germany: Springer; 2009.
57. Elhilu AH, El-Setouhy M, Mobarki AS, Abualgasem MM, Ahmed MA. Peer role-play simulation: a valuable alternative to bedside teaching during the COVID-19 pandemic. Adv Med Educ Pract 2023 Mar;Volume 14:257-264. [doi: 10.2147/amep.s399531]
58. Xu L, Liu W, Jiang X, Li Y. Impact of using peer role-playing on the clinical skills performance of pediatric trainees. BMC Med Educ 2023 Sep 08;23(1):654 [FREE Full text] [doi: 10.1186/s12909-023-04554-0] [Medline: 37684627]
59. Reiff M, Spiegel A, Williams E, Ramesh B, Madabhushi S, Bvunzawabaya B. From role-play to real life: using gatekeeper skills in real-world situations. J Coll Stud Ment Health 2022 Nov 22;38(1):119-148. [doi: 10.1080/87568225.2022.2144789]
60. Moreb M, Mohammed TA, Bayat O. A novel software engineering approach toward using machine learning for improving the efficiency of health systems. IEEE Access 2020;8:23169-23178. [doi: 10.1109/access.2020.2970178]
61. Ibeneme S, Karamagi H, Muneene D, Goswami K, Chisaka N, Okeibunor J. Strengthening health systems using innovative digital health technologies in Africa. Front Digit Health 2022 Mar 31;4:854339 [FREE Full text] [doi: 10.3389/fdgth.2022.854339] [Medline: 35434700]

XSL•FO
RenderX

62. Mhlanga D. Human-centered artificial intelligence: the superlative approach to achieve sustainable development goals in the fourth industrial revolution. Sustainability 2022 Jun 27;14(13):7804. [doi: 10.3390/su14137804]
63. Grindell C, Coates E, Croot L, O'Cathain A. The use of co-production, co-design and co-creation to mobilise knowledge in the management of health conditions: a systematic review. BMC Health Serv Res 2022 Jul 07;22(1):877 [FREE Full text] [doi: 10.1186/s12913-022-08079-y] [Medline: 35799251]
64. Nusir M, Rekik M. Systematic review of co-design in digital health for COVID-19 research. Univers Access Inf Soc 2022 Dec 31:1-15 [FREE Full text] [doi: 10.1007/s10209-022-00964-x] [Medline: 36618758]
65. Slattery P, Saeri AK, Bragge P. Research co-design in health: a rapid overview of reviews. Health Res Policy Syst 2020 Feb 11;18(1):17 [FREE Full text] [doi: 10.1186/s12961-020-0528-9] [Medline: 32046728]
66. Tremblay M, Hamel C, Viau-Guay A, Giroux D. User experience of the co-design research approach in eHealth: activity analysis with the course-of-action framework. JMIR Hum Factors 2022 Aug 09;9(3):e35577 [FREE Full text] [doi: 10.2196/35577] [Medline: 35943783]
67. Madanian S, Nakarada-Kordic I, Reay S, Chetty T. Patients' perspectives on digital health tools. PEC Innov 2023 Dec;2:100171 [FREE Full text] [doi: 10.1016/j.pecinn.2023.100171] [Medline: 37384154]
68. Kortum F, Klünder J, Brunotte W, Schneider K. Sprint performance forecasts in agile software development - the effect of futurespectives on team-driven dynamics. In: Proceedings of the 31st International Conference on Software Engineering and Knowledge Engineering. 2019 Presented at: SEKE 2019; July 10-12, 2019; Lisbon, Portugal. [doi: 10.18293/seke2019-224]
69. Graziani AR, Botindari L, Menegatti M, Moscatelli S. Adaptive coping strategies at the time of COVID-19: the role of social and general trust. Int J Environ Res Public Health 2023 Aug 03;20(15):6512 [FREE Full text] [doi: 10.3390/ijerph20156512] [Medline: 37569052]
70. Maison D, Jaworska D, Adamczyk D, Affeltowicz D. The challenges arising from the COVID-19 pandemic and the way people deal with them. A qualitative longitudinal study. PLoS One 2021 Oct 11;16(10):e0258133 [FREE Full text] [doi: 10.1371/journal.pone.0258133] [Medline: 34634054]
71. Fegert JM, Vitiello B, Plener PL, Clemens V. Challenges and burden of the Coronavirus 2019 (COVID-19) pandemic for child and adolescent mental health: a narrative review to highlight clinical and research needs in the acute phase and the long return to normality. Child Adolesc Psychiatry Ment Health 2020;14:20 [FREE Full text] [doi: 10.1186/s13034-020-00329-3] [Medline: 32419840]
72. Kowalski G, Ślebarska K. Remote working and work effectiveness: a leader perspective. Int J Environ Res Public Health 2022 Nov 20;19(22):15326 [FREE Full text] [doi: 10.3390/ijerph192215326] [Medline: 36430045]
73. Henke JB, Jones SK, O'Neill TA. Skills and abilities to thrive in remote work: what have we learned. Front Psychol 2022 Dec 19;13:893895 [FREE Full text] [doi: 10.3389/fpsyg.2022.893895] [Medline: 36600705]
74. Krajčík M, Schmidt DA, Baráth M. Hybrid work model: an approach to work–life flexibility in a changing environment. Adm Sci 2023 Jun 06;13(6):150. [doi: 10.3390/admsci13060150]
75. Sharma A, Ahmad Farouk I, Lal SK. COVID-19: a review on the novel coronavirus disease evolution, transmission, detection, control and prevention. Viruses 2021 Jan 29;13(2):202 [FREE Full text] [doi: 10.3390/v13020202] [Medline: 33572857]
76. Sun J, He WT, Wang L, Lai A, Ji X, Zhai X, et al. COVID-19: epidemiology, evolution, and cross-disciplinary perspectives. Trends Mol Med 2020 May;26(5):483-495 [FREE Full text] [doi: 10.1016/j.molmed.2020.02.008] [Medline: 32359479]
77. -. Effectiveness of public health measures in reducing the incidence of covid-19, SARS-CoV-2 transmission, and covid-19 mortality: systematic review and meta-analysis. BMJ 2021 Dec 03;375:n2997 [FREE Full text] [doi: 10.1136/bmj.n2997] [Medline: 34862160]
78. Cavalcante de Oliveira AP, Galante ML, Maia LS, Craveiro I, da Silva AP, Fronteira I, et al. Implementation of policy and management interventions to improve health and care workforce capacity to address the COVID-19 pandemic response: a systematic review. Hum Resour Health 2023 Oct 10;21(1):80 [FREE Full text] [doi: 10.1186/s12960-023-00856-y] [Medline: 37817165]
79. First lessons from government evaluations of COVID-19 responses: a synthesis. Organisation for Economic Co-operation and Development. 2022 Jan 21. URL: https://www.oecd.org/coronavirus/policy-responses/first-lessons-from-government-evaluations-of-covid-19-responses-a-synthesis-483507d6/ [accessed 2023-11-03]
80. Sigahi TF, Kawasaki BC, Bolis I, Morioka SN. A systematic review on the impacts of Covid-19 on work: contributions and a path forward from the perspectives of ergonomics and psychodynamics of work. Hum Factors Ergon Manuf 2021 Jul 12;31(4):375-388 [FREE Full text] [doi: 10.1002/hfm.20889] [Medline: 33821126]
81. Yanovskiy M, Socol Y. Are lockdowns effective in managing pandemics? Int J Environ Res Public Health 2022 Jul 29;19(15):9295 [FREE Full text] [doi: 10.3390/ijerph19159295] [Medline: 35954650]
82. Panchal U, Salazar de Pablo G, Franco M, Moreno C, Parellada M, Arango C, et al. The impact of COVID-19 lockdown on child and adolescent mental health: systematic review. Eur Child Adolesc Psychiatry 2023 Jul 18;32(7):1151-1177 [FREE Full text] [doi: 10.1007/s00787-021-01856-w] [Medline: 34406494]

XSL•FO
RenderX

83. Bardosh K, de Figueiredo A, Gur-Arie R, Jamrozik E, Doidge J, Lemmens T, et al. The unintended consequences of COVID-19 vaccine policy: why mandates, passports and restrictions may cause more harm than good. BMJ Glob Health 2022 May 26;7(5):e008684 [FREE Full text] [doi: 10.1136/bmjgh-2022-008684] [Medline: 35618306]
84. Maquiling A, Jeevakanthan A, Ho Mi Fane B. The effect of vaccine mandate announcements on vaccine uptake in Canada: an interrupted time series analysis. Vaccine 2023 May 02;41(18):2932-2940 [FREE Full text] [doi: 10.1016/j.vaccine.2023.03.040] [Medline: 37019696]
85. Zhang YS, Young Leslie H, Sharafaddin-Zadeh Y, Noels K, Lou NM. Public health messages about face masks early in the COVID-19 pandemic: perceptions of and impacts on Canadians. J Community Health 2021 Oct 20;46(5):903-912 [FREE Full text] [doi: 10.1007/s10900-021-00971-8] [Medline: 33611755]
86. He L, He C, Reynolds TL, Bai Q, Huang Y, Li C, et al. Why do people oppose mask wearing? A comprehensive analysis of U.S. tweets during the COVID-19 pandemic. J Am Med Inform Assoc 2021 Jul 14;28(7):1564-1573 [FREE Full text] [doi: 10.1093/jamia/ocab047] [Medline: 33690794]
87. Buchan MC, Bhawra J, Katapally TR. Navigating the digital world: development of an evidence-based digital literacy program and assessment tool for youth. Smart Learn Environ 2023;11:8. [doi: 10.31219/osf.io/aqvcu]

Abbreviations

AI: artificial intelligence

Edited by K Bond; submitted 24.01.24; peer-reviewed by J Knight, MR Tanniru; comments to author 07.07.24; revised version received 23.08.24; accepted 02.03.25; published 17.04.25.

Please cite as:
Buchan MC, Katapally TR, Bhawra J
Application of an Innovative Methodology to Build Infrastructure for Digital Transformation of Health Systems: Developmental Program Evaluation
JMIR Form Res 2025;9:e53339
URL: https://formative.jmir.org/2025/1/e53339
doi:10.2196/53339
PMID:

©M Claire Buchan, Tarun Reddy Katapally, Jasmin Bhawra. Originally published in JMIR Formative Research (https://formative.jmir.org), 17.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Ecological Momentary Assessment of Parental Well-Being and Time Use: Mixed Methods Compliance and Feasibility Study

Laura Altweck[1,2], PhD; Samuel Tomczyk[1,2], PhD

[1]Department Health and Prevention, Institute of Psychology, Universität Greifswald, Greifswald, Germany

[2]German Center for Child and Adolescent Health (DZKJ), partner site Greifswald/Rostock, Greifswald, Germany

Corresponding Author:
Laura Altweck, PhD
Department Health and Prevention
Institute of Psychology
Universität Greifswald
Robert-Blum-Str. 13
Greifswald, 17487
Germany
Phone: 49 3834 420 ext 3813
Email: laura.altweck@uni-greifswald.de

Abstract

Background: Parents often juggle multiple conflicting responsibilities, including work, childcare, and the household, making them a particularly burdened group. However, the impact of daily routines and associated (poor) well-being among parents has received relatively little attention. Ecological Momentary Assessment (EMA) is increasingly being used to capture real-time data and can help address this research gap.

Objective: This study aims to examine compliance rates and the feasibility of EMA for measuring daily well-being and time use among parents.

Methods: An exploratory mixed-methods study was conducted with 74 German parents (57/74, 77% women, (age: mean 37.6, SD 5.9 years). Participants completed a baseline questionnaire, followed by 4 daily EMA surveys (at 7:30 AM, 12 PM, 16:30 PM, and 21:30 PM) over a 1-week period, and a follow-up questionnaire. A subset of parents was also subsequently interviewed. Sociodemographic background and expected feasibility (open-ended questions) were surveyed at baseline, and feasibility was assessed at follow-up (closed- and open-ended questions) and in the interviews. State well-being (affective and cognitive), state stress, state as well as retrospective time-use were measured in the EMA surveys. Compliance and feasibility were examined using a combination of quantitative (descriptive analyses) and qualitative methodologies.

Results: Participants completed an average of 83% (SD 13%) of the daily surveys. Compliance varied by gender and age, where men (90% vs 80%) and older parents showed higher rates. Participants generally found the survey frequency and length manageable, though some suggested adjustments to the study period depending on their individual routines. The 7:30 AM survey was reported as the most challenging due to childcare drop-offs (40%-49%), followed by the 16:30 PM survey for similar reasons (7%-17%). The qualitative analysis further revealed additional points for improvement, for instance, the need for personalization (eg, individual adjustment of the survey timings and intervals), technical support, and the incorporation of gamification elements. Most interviewees (46% vs 23%) found the used measurement of well-being and stress to be appropriate. Regarding time use, they felt that the predefined activity groups (eg, personal care, working) were suitable (46%) but noted challenges assigning less frequent activities (eg, medical appointments) (5%-54%). Reporting the timings of time-use via consecutive questions (ie, specifying the duration or start and end times of an activity) was perceived as confusing (9%-69%), with participants expressing a preference for a visual overview, such as a Gantt chart.

Conclusions: The study demonstrates that, when accounting for certain sociodemographic and study design factors, EMA can be a feasible method for data collection regarding daily well-being and time use, even in highly time-constrained populations like parents. This shows great potential for future research, such as exploring work-family conflict or performative gender roles and complementing established methods (eg, retrospective daily diaries).

Trial Registration: OSF Registries osf.io/8qj3d; https://osf.io/8qj3d

International Registered Report Identifier (IRRID): RR2-10.2196/54728

XSL•FO
RenderX

(JMIR Form Res 2025;9:e67451) doi:10.2196/67451

KEYWORDS

ecological momentary assessment; parents; feasibility; compliance; time use; well-being; stress; EMA; mixed methods; daily life; parents; online questionnaires; questionnaires; surveys; quantitative; qualitative; sociodemographic; mobile phone

Introduction

Background

As parents often have to fulfill professional, domestic, and parental responsibilities, they can be considered a particularly burdened group [1]. These demands of balancing work, family, and personal life can lead to worse physical and mental health outcomes, underscoring the need for targeted support and interventions in daily life [2]. However, daily routines and associated (poor) well-being among parents have received relatively little attention, for instance, using real-time assessment methods, such as ecological momentary assessment (EMA) or experience sampling methods. Given that parents' well-being is not only important for their own health but also for their children's [3,4], further research into the feasibility and implementation of EMA in parents is important.

Parents as a Target Group

Parents today face significant challenges by balancing childcare responsibilities, such as managing school drop-offs and pick-ups, with their careers. Although time use between men and women has become more similar in recent decades [5,6], substantial differences remain, with men generally spending more time on paid work and less on unpaid work [7-9]. As men continue to prioritize work, many women face a double burden, struggling to balance family and professional demands [7,9,10]. This discrepancy extends to care work and childcare. Women often reduce their paid work hours after giving birth, whereas men do not but may instead increase theirs [8]. Women's share of housework is typically double that of men's, with the largest gender differences observed among those with preschool-aged children [11,12]. Mothers experience more variation in how they use their time, often engaging in part-time work and childcare simultaneously, and activities frequently overlapping with others (eg, doing housework while taking care of the children) [13]. They generally also have less free time than fathers and tend to combine their leisure time with other activities, which can exacerbate stress and reduce opportunities for recovery [13].

Understanding how individuals behave in their everyday lives and how their actions affect mental and physical health is essential [2]. According to the Conservation of Resources Theory [14], individuals manage their time to preserve or enhance their resources (eg, energy, money, or social support), while the Work-Family Conflict Model [15] suggests that time is divided between work and family based on competing demands and conflicts and this conflict may lead to stress and poor health outcomes. The gender perspective complements these theories, examining how gender roles shape time use, with theories such as Doing Gender [16] and Gender Performativity [17] emphasizing the construction of gender roles through actions and behaviors. Previous research has established gender differences in daily activities and their impact on general health and well-being [7,9,10,18,19]. Certain activities—such as household cleaning and unpaid care work—are consistently related to worse health outcomes [20,21], and these are in turn more likely to be undertaken by women [9]. Therefore, examining the daily well-being and time use of parents, particularly mothers, is important for understanding and addressing gender disparities in health using a real-time lens.

Ecological Momentary Assessment

The widespread use of smartphones provides an opportunity to collect real-time data and access a wider population. EMA involves "repeated collection of real-time data on participants' momentary states in the natural environment" [19]. EMA can, for example, be implemented via smartphone apps, which can collect data both internet-based and offline and use various methods such as open-ended questions, text- or voice-based responses, Likert scales, multiple-choice options, number wheels, or sliders [22]. Data can be gathered at deliberately selected time points (ie, experience-based sampling, eg, after the morning coffee), randomly selected time points (ie, time-based sampling), or fixed intervals throughout the day (ie, interval-based sampling, eg, each morning at 9:00 AM).

While EMA has been widely used in clinical populations to monitor changes in state well-being during illness or substance withdrawal [23], its application among the population, for example, in parent samples, is limited. State well-being is relatively stable over time compared with mood [24], but there is a greater variation in positive and negative state well-being (eg, affect) than in cognitive state well-being (eg, life satisfaction) [25]. Overall, daily state well-being tends to decline throughout the day and toward the end of the week [26,27]. Understanding these patterns among parents and their association with daily routines and activities could inform interventions aimed at improving their well-being.

Although the research community has acknowledged the importance of daily activities for well-being and health, EMA is rarely implemented. So far, time use is generally measured through paper-pencil diary methods [28], studies that examine time-use patterns using EMA are rare [22,29]. EMA as opposed to paper-pencil diaries is better suited for capturing short and infrequent activities like errands and housework [29], which are precisely the areas where gender differences in time use are most pronounced, making EMA a valuable method for measuring these disparities. A comprehensive comparison of different methodological approaches is beyond the scope of this study, as we focus on the compliance and feasibility of the measurement of well-being and time use using EMA in a parent sample, without a control condition that uses daily diaries.

XSL•FO
RenderX

Compliance and Feasibility of Ecological Momentary Assessment

Compliance is a key factor in the success of EMA studies, especially given the unpredictability of daily life. High compliance rates are essential for the validity of the data collected. Electronic EMA surveys have been shown to have higher compliance rates than paper-based methods (96% vs 70%), due to advantages such as automatic prompts, immediate data transfer, secure servers, and greater accessibility and convenience for participants [30]. Electronic assessment ensures that the responses are provided at the intended times, reducing issues with delayed or retrospective entries that can occur with paper-based methods. Higher compliance with electronic assessments has been found in both adults and children [31]. However, compliance rates can vary. De Vries and Bartels [32] reported an average compliance rate of 72%, ranging from 43% to 95%, while Passini et al [33] reported an average compliance rate of 83%. Factors such as the frequency of prompts and asking repeating questions can affect compliance; studies with higher prompt frequencies tend to have lower compliance rates [30]. There were no significant differences in compliance when using personal smartphones versus research-provided devices, nor did incentives influence compliance in app-based EMA applications [32].

Compliance and Feasibility of Ecological Momentary Assessment in Parents

Parents, due to their high levels of stress and busy schedules, present unique challenges for EMA studies. Stress has been shown to impair episodic memory [34], suggesting that real-time data collection methods like EMA could be particularly advantageous for this group by reducing reliance on memory and recall.

Studies have shown that EMA is a feasible method for researching parenting behaviors. For instance, Passini et al [33] found that using EMA to record parenting behaviors 3 times a day was effective in capturing different dimensions of parenthood. In their study, 90% of mothers reported low difficulties using EMA, and 79% did not perceive it as burdensome. In addition, 60% would recommend EMA studies to other mothers. However, 74% indicated that the EMA study was at least partly challenging, suggesting a need to explore ways to reduce these challenges for parents.

Compliance rates in EMA studies involving parents vary. Kerr et al [35] reported that mothers responded to an average of 34 out of 50 prompts (ie, 68%) while El Dahr et al [36] reported a compliance rate of 82% in parents. Compliance was shown to be higher in the evenings [36,37].

Technical issues can also affect compliance; about 18% of EMA prompts are not answered due to problems such as data loss, battery difficulties, software bugs, or prompts occurring outside participants' wake times [37]. Socioeconomic factors play a role as well, for example, Dzubur et al [37] found that mothers with higher annual incomes responded to prompts more quickly than those with lower incomes. This suggests that financial stability may afford parents more flexibility or resources to engage with EMA studies.

These findings suggest that while EMA is a promising method for studying parental well-being and time use, there are challenges that need to be addressed to enhance compliance and feasibility. Understanding the specific needs and constraints of parents is essential for designing EMA studies that are both effective and participant-friendly.

Research Aims and Research Questions

The following research questions (RQ) were posed:

RQ1: What is the overall feasibility of parents participating in an EMA study measuring daily well-being and time use?

RQ2: What are the overall compliance rates for parents participating in an EMA study measuring daily well-being and time use?

RQ3: Which aspects of the study design (eg, time point) and sociodemographic factors are associated with noncompliance and feasibility?

Methods

Study Design

This work is part of a quasi-experimental pilot study using EMA [38]. The main study consisted of internet-based questionnaires (before and after a week of EMA), daily EMA surveys over the course of 7 days, and subsequent structured interviews for the intervention group. The control group only took part in the internet-based questionnaires and did not partake in the interviews or the EMA surveys. This study focuses on the intervention group; for the full study design see Altweck et al [38].

The minimum sample size (ie, 64 participants) for the intervention group was calculated using G*Power (version 3.1.9.6; Erdfelder, Faul, and Buchner; ANOVA, repeated measures, within a group, α=.05, power=0.8, correlation among repeated measures=0.5, and effect size [well-being]=0.25 [39]). For the interviews, to achieve data saturation [40,41], we aimed to recruit 6-17 participants.

The mixed methods approach facilitated the examination of compliance rates and their associations with various factors (eg, sociodemographic characteristics and study design features), while also providing insights into the feasibility of the study design through short, open-ended questionnaire responses and in-depth interview answers. First, participants in the intervention group filled in an internet-based questionnaire. On the next day, they started filling in the EMA surveys (each approximately 5-7 min), which were completed on participants' personal Android smartphones using the app movisensXS (2024; movisens GmbH). The surveys took place 4 times per day (7:30 AM, 12 PM, 4:30 PM, 9 PM) with a buffer of 30 minutes over a period of 7 days. These time points were selected to also test feasibility in high-stress situations, for example, at 7:30 AM when parents get their children ready for nursery or school before work. The first 2 days were classed as a familiarization period and were not included in the analyses. After one week, participants filled in another internet-based questionnaire. The first participants to complete the daily EMA surveys were invited to take part in a structured interview to comprehensively

XSL•FO
RenderX

understand the feasibility of the study design and to be able to make adjustments if necessary, which was not the case.

Sample and Recruitment

Parents with at least one child under 18 years and living in the household were included. Only parents over the age of 18 years, without foster children, and without children with a chronic illness or disability were included, as these factors have been shown to further increase the burden of parenting [42]. Participants also needed access to an Android smartphone, in order to complete the daily surveys, for technical reasons.

Participants were recruited primarily from the northeast region of Germany using convenience sampling by distributing leaflets to relevant groups (eg, parent groups, nurseries, and schools) and digital channels (eg, the university's website and social media). In an initial phone conversation, participants were informed about the study goals and methods, informed consent was collected, and, finally, the inclusion and exclusion criteria were checked. The first participants who completed the follow-up questionnaire were invited to participate in an interview.

Procedure and Instruments

Baseline Questionnaires

Participants first filled in an internet-based questionnaire, which took approximately 20-30 minutes to complete.

The following sociodemographic variables were assessed: gender (woman, man), age (in years), education (International Standard Classification of Education-97 [43]; low and medium: level 0-4 [reference], high: level 5-6), employment status (full-time work, part-time work, not working), partnership status (in partnership, single, other), number of children under 14 years (numeric), number of children between 14 and 18 years (numeric), and family time (average h/d spent on family commitments [eg, childcare, family time, and housework] on both a regular workday and a free day). In addition, the expected feasibility of the study design was measured using the 4 open-ended questions (see Multimedia Appendix 1 for full item wording). For instance, participants were asked about the integration of EMA into their daily life and the survey's length and time points.

Daily Ecological Momentary Assessment Surveys

On the next day, participants began with the daily EMA surveys. Participants started the EMA surveys either on Tuesday or Thursday. The daily EMA surveys included assessments of cognitive (life satisfaction item from the WHO-QoL [World Health Organization Quality of Life]-8) [44], positive (happy, relaxed), and negative (nervous, downcast) affective well-being [45], and stress (a global measure of perceived stress) [46]. Also, current activity and the sequence of activities since the last survey time point were queried (activity categories according to Tomczyk et al [9]; eg, working, housework, childcare). In the final surveys of the day, received and desired social support [3] as well as state mindfulness (Multidimensional State Mindfulness Questionnaire) [47] was assessed. Participants answered between 12 and 27 items per survey point, depending on (1) the variation in time use between time points (ie, the number of different activities reported) and (2) the survey time point, with more items presented in the evening. For details, see Altweck et al [38] and for the full EMA survey, see Multimedia Appendix 2.

Follow-Up Questionnaires

A total of 8 days after the baseline questionnaire, participants filled out another internet-based questionnaire.

The feasibility of the study design was surveyed with a mix of closed- and open-ended questions. The questions were inspired and adapted from the existing literature that examined the feasibility of EMA methods [22,48-50] (see Multimedia Appendix 1 for full item wording and scales). An example item included "In my opinion...I was able to accommodate the survey time point at 7:30am", which was rated on a scale from 1 (strongly disagree) to 5 (strongly agree). Participants were able to explain their choices in an open textbox.

Interviews

Finally, a subgroup of the intervention group was invited to participate in an internet-based, structured interview that lasted 20-30 minutes. The interview guide was developed based on previous qualitative research on usability and feasibility evaluations [51,52] and was tailored to the current procedure, including focusing on the specific survey time points. The interviews focused on the feasibility of the EMA survey method and study design in greater detail. The interview followed the following structure: (1) general assessment of the study design, (2) time expenditure and integration into everyday life, (3) contents of the EMA surveys, (4) effects of the study, and (5) general and conclusion; see Multimedia Appendix 3 for the full interview guide.

Ethical Considerations

The pilot study [38] was approved by the Ethics Committee of the University Medicine Greifswald in July 2022 (BB 113/23) and was conducted following the ethical standards set by the 1964 Declaration of Helsinki, its later amendments, or similar standards. Informed verbal and written consent was obtained from each participant, informing them about how their data would be used and stored, and that their participation was pseudonymized. They were also informed that they could withdraw from the study at any time without repercussions. Ethical guidelines and general data protection regulations were adhered to concerning data usage and storage. Participants in the intervention group were offered a monetary incentive of €35 (US $36.63) for the completion of the baseline and follow-up questionnaires as well as at least 50% of the daily surveys, an additional €15 (US $15.70) if they completed at least 80% of the daily surveys, and an additional €10 (US $10.47) if they took part in the interview.

Analytical Approach

Analysis Software

The statistical analyses were conducted using jamovi (version 3.5) [53], while the qualitative analyses were conducted using MAXQDA (version 24; VERBI Software GmbH) [54].

XSL•FO
RenderX

Data and Sample Description

For the sample description, we calculated mean values and SD for continuous variables and relative frequencies for categorical variables. Comparisons across gender were made using *t* tests, chi-square tests, or, where relevant, nonparametric equivalents (eg, Fisher exact test, Mann-Whitney *U* test, Welch *t* test).

Compliance

We analyzed compliance following guidelines by Liao et al [30] and Dao et al [55], reporting overall compliance (number of overall completed prompts), attrition (completed prompts by survey day and time point), and latency (time from prompt to survey completion and time to complete individual surveys). Associations of compliance with relevant sociodemographic factors were examined.

Feasibility

Closed questions regarding feasibility were analyzed using descriptive statistics (see the Data and Sample Description section). For the evaluation of the interviews and open-ended questions in the questionnaires, qualitative content analysis according to Kuckartz and Rädiker [56] was used. The qualitative data from the baseline questionnaires (ie, expected feasibility), follow-up questionnaires (feasibility), and interviews (feasibility) were coded and categorized separately.

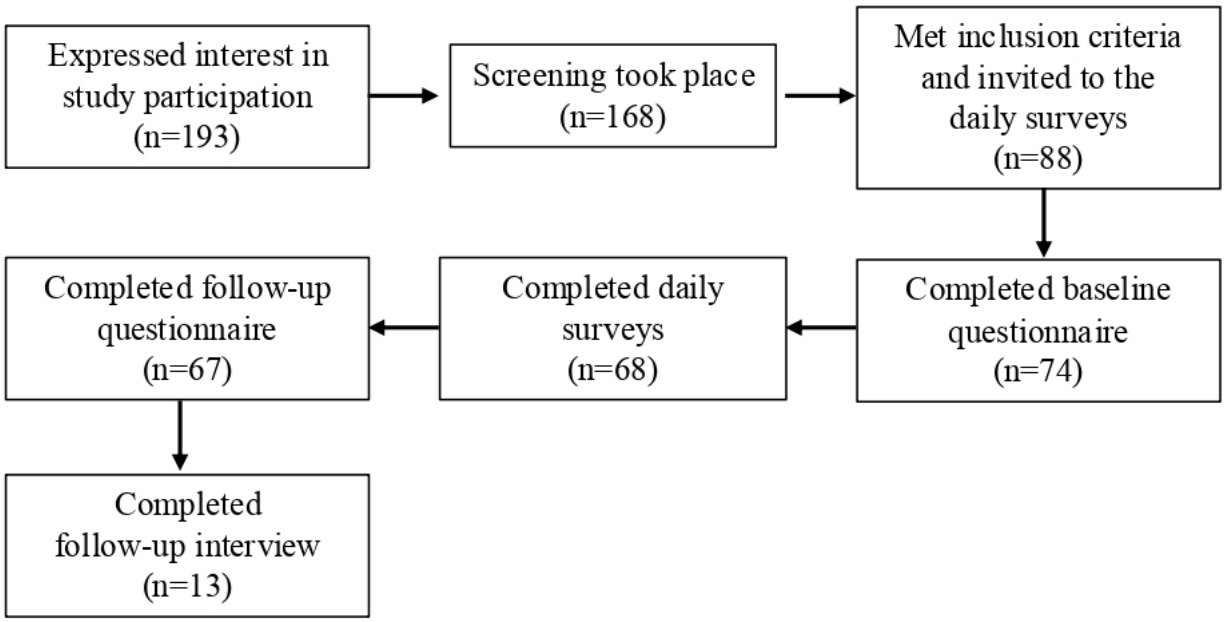
The audio or video data from the interviews were transcribed for further analysis. The research team reflected on and noted their previous assumptions regarding the sample, research questions, and expected results. Summaries for each participant were compiled to gain an initial overview of the qualitative data. Text segments or sense units were coded and categorized either into existing categories or new ones. The first version of the main categories was created deductively based on the questionnaire and interview questions. Once coding began, an inductive approach was adopted, allowing the categories to be refined or extended with subcategories according to the data. In the first stage of the coding process, the qualitative data was once fully coded by a single person using the main deductive categories. Throughout this process, the research team noted overlapping categories, possible amendments (ie, new categories or redundant ones), and potential subcategories. Following Kuckartz and Rädiker [56], interrater agreement was achieved through discussions within the research team. In the second stage of the coding process, one researcher coded the data using the final categories twice, 2 weeks apart.

Results

Data and Sample Description

In total, 193 parents expressed interest in taking part in the study by leaving their personal information (Figure 1). We were able to reach 168 parents to complete the telephonic screening, of these, 88 met the inclusion criteria and were invited to the intervention group. In total, 74 parents completed the baseline questionnaires, 68 parents also completed the daily EMA surveys, 67 parents also completed the follow-up questionnaires, and 13 parents took part in the interviews. With 4 surveys per day over a 7-day period, we collected 1352 data points. To gain the most complete overview of the compliance and feasibility of this methodology in a parent sample, we included all possible cases in the analyses and excluded missing values case-wise (N=74). There was no missing data in the questionnaire and interview data, except average daily family time (2/74, 3%). The samples showed no significant differences in any sociodemographic variables or survey compliance (P>.05).

Figure 1. Sample overview including sample size and retention.



The final sample description is given in Table 1. The complete sample was made up of 77% (57/74) women and 23% (17/74) men. The average age was 37.6 (SD 5.9) years. The sample was highly educated (44/74, 60%), mostly in a relationship (53/74, 72%), and nearly equally divided across full-time, part-time, and no employment. On average, participants spent 9.6 (SD 4)

XSL•FO
RenderX

hours per day on family commitments. There were few significant differences across genders, men were merely significantly older than women. Also, there were no significant differences across subsamples (baseline only, baseline and follow-up, interview group; $P>.05$).

Table 1. Descriptive statistics of the sample at baseline (total and by gender; N=74).

Sample variable	Total sample	Women	Men
Age (years)			
Participants, n (%)	74 (100)	57 (77)	17 (23)
Mean (SD)	37.6 (5.9)	36.9 (5.1)	40.3 (7.6)[b]
Range	—[a]	28-53[b]	26-59[b]
Partnership status, n (%)			
In a relationship	53 (72)	39 (68)	14 (82)
Single	6 (8)	6 (11)	0 (0)
Other	15 (20)	12 (21)	3 (18)
Number of children			
Younger than 14 years			
Mean (SD)	1.7 (0.8)	1.7 (0.8)	1.8 (0.8)
Range	1-4	1-4	1-3
Between 14 and 18 years			
Mean (SD)	0.3 (0.5)	0.4 (0.6)	0.1 (0.4)
Range	0-2	0-2	0-1
Educational background, n (%)			
Low or medium	30 (41)	26 (46)	4 (24)
High	44 (60)	31 (54)	13 (77)
Employment status, n (%)			
Full-time	28 (38)	19 (34)	9 (53)
Part-time	29 (40)	24 (43)	5 (29)
None	16 (22)	13 (23)	3 (18)
Family time (h/d)			
Mean (SD)	9.6 (4)	10 (4.7)	8.2 (2.3)
Range	2-24	2-24	3-12.5

[a]Not applicable.
[b]$P<.05$.

Compliance

From the 28 maximum possible daily surveys to complete (4 times/d for 7 days), the average individual number of completed surveys was 23.1 (SD 3.57) (23.1/28, 82.5%; range 9-28) for results on compliance, refer to Table 2 and for grouped by time point or survey day, refer to Multimedia Appendix 4). Men (25/28, 90%; mean 25.1, SD 1.9) completed significantly more daily surveys than women (22.4/28, 80%; mean 2.4, SD 3.75) and the latter also showed greater variance ($t_{54.9}=3.87$, $P<.001$). Neither the survey time point nor the survey day significantly affected compliance ($P>.05$).

XSL•FO
RenderX

Table 2. Descriptive statistics of compliance at follow-up (total and by gender; N=68).

Variable	Total sample		Women		Men	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Average completed surveys (app data)	23.1 (3.6)	9-28	22.4[a] (3.8)	—[b]	25.1[a] (1.9)	—
Time to respond to trigger (app data)						
Across time points	2.6 (17.4)	0-633	2.3 (8.9)	0-270	3.3 (30.9)	0-633
7:30 AM	2.6 (13.9)	0-270.2	3 (16.5)	0-270.2	1.6 (3.1)	0-22.2
12 PM	1.7[c] (3.1)	0-26.1	1.7 (3.1)	0-26.1	1.5 (3.1)	0-21.7
4:30 PM	2.7[c] (5.3)	0-34.7	3 (5.5)	0-28.5	2 (5)	0-34.7
9 PM	3.4 (31.6)	0-632.5	1.6 (3.5)	0-25.6	8 (61.3)	0-632.6
Time to complete the survey						
Reported in the interviews	4.1 (1.8)	1.8-5.8	3.5 (2)	1.8-5.8	4.7 (1.5)	3.3-5.6
App data						
Across time points	3 (6.6)	0.4-137	3.2 (6.8)	0.5-137	2.8 (6.7)	0.4-134
7:30 AM	2.3[d] (3.4)	0.4-60.2	2.4[e] (4)	0.7-60.2	2.2[e,f] (1.5)	0.4-11.3
12 PM	2.8[g] (7.7)	0.4-134.4	2.8 (5)	0.6-60.4	3[h] (13)	0.4-134.4
4:30 PM	2.4[i] (3)	0.5-42.7	2.4[j] (4.4)	0.5-42.7	2.2[j,k] (1.8)	0.5-9.3
9 PM	4.7[d,g,i] (9.6)	1.1-136.8	5.1[e,j] (11.4)	1.2-136.8	3.8[e,f,h,j,k] (2.7)	1.1-20.5

[a]Significant differences across gender (P<.05).

[b]Not applicable.

[c]Significant differences across time points (P<.05).

[d]Significant differences across time points (P<.05).

[e]Significant differences across time points (P<.05).

[f]Significant differences across time points (P<.05).

[g]Significant differences across time points (P<.05).

[h]Significant differences across time points (P<.05).

[i]Significant differences across time points (P<.05).

[j]Significant differences across time points (P<.05).

[k]Significant differences across time points (P<.05).

Age significantly predicted the average individual number of completed surveys, where older participants showed higher compliance (r=0.33, P=.006). Other sociodemographic factors were not significantly related to compliance (P>.05).

On average, participants took 2.6 (SD 17.4) minutes to respond to a survey after receiving a signal; there was no significant gender difference (P>.05). Both men and women took longer to respond to the survey at 4:30 PM compared with the 12 PM one.

In the qualitative interviews, participants reported taking 4.1 (SD 1.8) minutes to complete the survey per time point, yet the app data showed that, on average, it took participants 3 (SD 6.6) minutes (see Multimedia Appendix 5 for examples and descriptions); there were no gender differences (P>.05). Both men and women took significantly longer to fill out the evening surveys (mean 4.7, SD 9.6) than the other time points (mean 2.3-2.8, SD 3.4-7.7); this discrepancy was also reported in the qualitative interviews.

Feasibility

Technical Aspects

Interviewees found the app easy to use and user-friendly, which was confirmed in the follow-up questionnaires. Participants highlighted the reminder function of the surveys and appreciated the real-time support provided during the study by the research team. However, some technical issues were raised, including compatibility problems with the operating system, display issues, unreliable reminders, and the nonautomatic snooze feature. For more details regarding technical aspects see Multimedia Appendix 6.

Study Design

In terms of the overall feasibility of the study design, at baseline, 74% (53/74) did not report any expected concerns, and 13% (9/74) expressed excitement about taking part in the study questionnaires (see Multimedia Appendix 7 for qualitative results from the baseline questionnaires). However, most participants felt that the study design was generally “doable” (high feasibility: 20/74, 28%, moderate feasibility: 51/74, 71%)

XSL•FO
RenderX

but expected certain difficulties in integrating the study into their daily lives (eg, some parts of the study design like a certain time point) and reported concerns, mainly around time management. The study duration (ie, the length of the overall study period and the individual surveys) appeared acceptable.

Participants found answering the questions every day adequately easy (mean 3.37, SD 1.04; scale: 1=difficult to 5=easy; refer to Tables 3 and 4 for quantitative results from the follow-up questionnaires). Some participants in the follow-up questionnaires (3/57, 5%) and 92% (12/13) interviewees reported a habituation effect because after a while "you already knew what you had to answer" (see Multimedia Appendix 8 for qualitative results from the follow-up questionnaires). Most interviewees reported the length and frequency of the individual survey time points to be adequate (12/13, 92% and 10/13, 77%, respectively). In the follow-up questionnaires, participants felt that the length of the individual time points was adequate, leaning slightly toward too much (mean 2.63, SD 0.8; scale: 1=too much to 5=far too little) but did not feel that the number of prompts was "annoying" (mean 2.48, SD 1.20; scale: 1=strongly disagree to 5=strongly agree). Most participants also did not feel that the study period was too long (mean 1.71, SD 0.94; scale: 1=strongly disagree to 5=strongly agree) and expressed that the survey period was adequate. However, 54% (7/13) of interviewees expressed a desire for the EMA study period to be longer than one week. Some expressed a desire for a longer survey period between 10 days and 3 weeks in order to better reflect fluctuations in everyday life and well-being. In contrast, one interviewee preferred a shorter survey period of 3 or 4 days, stating that the participation was stressful. In the follow-up questionnaires, 19% (11/57) of participants also requested a longer survey period, with the caveat that the number of daily surveys is reduced.

Table 3. Descriptive results from the follow-up questionnaires regarding feasibility (N=67).

Compliance	Mean (SD)
I found answering the questions everyday... (1=difficult to 5=easy)	3.37 (1.04)
I found the length of the daily surveys to be... (1=too much to 5=far too little)	2.63 (0.8)
In my opinion... (1=strongly disagree to 5=strongly agree)	
...the duration of the test (7 d) was too long.	1.71 (0.94)
...the number of daily prompts that were sent (4 times/d) was annoying.	2.48 (1.2)
...I ignored or postponed the prompts.	1.75 (1.01)
...the questions were understandable.	3.33 (1.04)
...I was able to remember the activities since the last measurement time.	4.17 (0.94)

Table 4. Descriptive results from the follow-up questionnaires regarding the feasibility of the survey time points (N=67).

Feasibility	Total sample, mean (SD)	Men, mean (SD)	Women, mean (SD)
In my opinion, I was able to accommodate the survey time at...well. (1=strongly disagree to 5=strongly agree)			
...7:30 AM...	3.42[a,b] (1.33)	3.76 (1.39)	3.31 (1.29)
...12 PM...	4.06[a,c] (1.16)	4.18 (1.07)	4.02 (1.2)
...4:30 PM...	3.62[c] (1.24)	3.35 (1.32)	3.71 (1.21)
...9:30 PM...	3.88[b] (1.07)	4.24[d] (0.66)	3.76[d] (1.16)

[a]Significant differences across time points (P<.05).

[b]Significant differences across time points (P<.05).

[c]Significant differences across time points (P<.05).

[d]Significant differences across gender (P<.05).

Survey Time Points

In the baseline questionnaires, 8% (6/74) of participants anticipated that it would be difficult to integrate the survey time points in general into their daily life. About half of the interviewees (7/13, 54%) found it difficult to fill out the surveys at the predetermined time points. Most interviewees (9/13, 69%) found the 30-minute response time sufficient, while 15% (2/13) preferred a longer response window of one hour. In the follow-up questionnaires, participants generally reported that they did not ignore or postpone survey prompts (mean 1.75, SD 1.01; scale: 1=strongly disagree to 5=strongly agree).

In the baseline (29/74, 40%), follow-up questionnaires (28/57, 49%), and the interviews (6/13, 46%), the 7:30 AM time point was reported to be the single most difficult survey to complete "because most of the time it's really the children's drop-off time. So if you have school-age children (laughs) or have to take them to the nursery, it was always a bit difficult to fit that in somehow". Instead, they suggested moving the early morning survey either a bit earlier (eg, 6:30 or 7:00 AM) or later (eg, 8:00 or 9:00 AM). The survey time point at 7:30 AM was said

XSL•FO
RenderX

to be the hardest to accommodate, especially compared with 12 PM and 9:00 PM (P<.05).

At baseline, the 12 PM time point was expected to be feasible (15% vs moderate-low feasibility: 10%), but at follow-up, it showed mixed feasibility (15% vs moderate-low feasibility: 10%). It was feasible if it took place during the lunch break at work but bad if it coincided with work duties or the children's lunchtime nap. For this time point, participants also suggested earlier (eg, 11:30 AM) or later times (eg, 2:00 PM).

In the baseline questionnaires, participants anticipated the time point at 4:30 PM to be the second most difficult to complete (17% vs moderate-low feasibility: 7%), which was confirmed at follow-up (high feasibility: 2%, low: 7%). The survey time point at 4:30 PM was also significantly easier to accommodate compared with 12 PM (P<.05). This time point was reported to be difficult if it coincided with the end of work and picking up the children from nursery or school or if it was after work and so the parents were in the middle of childcare and running errands or taking children to afternoon classes. Participants generally suggested a time a bit later, such as 5:00 PM or 5:30 PM.

In the interviews, the 9:00 PM time point was reported to be the second most feasible, whereas, in the follow-up questionnaires, 9:00 PM was reported to have been difficult to complete (high feasibility: 4% vs low feasibility: 19%). For some, this time point was "too late." If the children were already asleep, this time point appeared to be the most feasible; however, if the parents were still getting the children ready for bed, as the children sleep late in general or just on that day, then this time point appears to be similarly difficult as the one at 7:30 AM. The only significant gender difference in the feasibility of time points was regarding 9:00 PM, which men found easier to accommodate than women (P<.05).

Participants mentioned a difference in feasibility between weekdays and weekends; for instance, the 7:30 AM time point was too early on weekends. They also expressed a desire for more flexibility in tailoring the survey points to individual routines.

Measurement of Well-Being and Time Use Using Ecological Momentary Assessment

Generally, participants reported that the questions in the daily surveys were adequately understandable (mean 3.33, SD 1.04; scale: 1=strongly disagree to 5=strongly agree).

Most interviewees (46% vs 23%) found the used measurement of well-being to be appropriate, describing the questions as "multifaceted," while some reported the opposite and for example, felt that the items were too "superficial." Instead, 23% (3/13) reported that the questions did not reflect fluctuations over the day or the week.

Most interviewees (10/13, 77%) agreed that the query about the current activity 4 times per day was not a full reflection of the full day and that the questions about the activities between assessments were necessary to complete the picture: "I think the combination really makes the most sense, because then you really get everything packed into it."

There was some uncertainty regarding the predefined time-use categories. About half of the interviewees expressed uncertainty about assigning rare activities, for example, a doctor's appointment, to a certain category, which was also mentioned in the follow-up questionnaires. Others were unsure how to classify leisure and physical activity, that is, whether to assign this to hygiene, eating, etc., or childcare, when, for example, watching TV while the children were around. Some participants also reported that they would have preferred logging the individual activities (eg, watching TV) rather than an activity group (eg, free or leisure time).

At baseline, 83% (60/74) expected to be able to remember the activities since the last survey time point, this was confirmed in the follow-up questionnaires (mean 4.17, SD 0.94; scale: 1=strongly disagree to 5=strongly agree) and interviews (12/13, 92%). Participants were asked to report what they were currently doing and then the sequence of what they had been doing since the last survey. Both in the interviews and the questionnaires, participants mentioned that this was confusing and they would have preferred reporting activities in chronological order. Also, some felt the sequential report of activities was redundant, for example, when they were at work at 12 PM and still at work at 4:30 PM. Participants were also able to click on multiple activities at the same time, while some found this useful, others found it confusing to report the sequence in this way. Participants also reported being unsure whether to report the time (eg, 7:05 AM) or the time passed in hours and minutes (eg, since 1 h 30 min); there was no clear preference. In both the follow-up questionnaire and interviews, participants suggested including a visual overview of the activity sequence, similar to a "Gantt chart."

Miscellaneous

All interviewees also deemed the monetary incentive as appropriate; however, some also noted that as parents are short on time, a slightly higher incentive might be useful (between €50 [US $52.33] and €100 [US $104.65]). Some interviewees also mentioned that because the incentive was proportional to completed daily surveys, they would have liked a personal overview of their compliance.

One interviewee also mentioned that an offer of psychological support would be useful for participants who notice constant poor well-being or low levels of social support while completing the study.

Overall, the results offer a few key recommendations for future EMA studies with similar aims (Textbox 1). Ideally, these aspects should be discussed with the target group via participatory co-design workshops when preparing the study to ensure high commitment and compliance from the beginning.

XSL•FO
RenderX

Textbox 1. Key recommendations for designing ecological momentary assessments of time use and well-being in parents based on participants' perspectives.

- Technical support: Incorporate reminder functions and automatic snooze functions to facilitate completion and provide swift and adequate technical support throughout the study.
- Assessment schedule: Choose intervals that are acceptable for participants (eg, 4 times/d for 7 days or longer time periods with reduced frequency, eg, 2-3 times/d for 10-14 days).
- Customization: If possible: allow flexibility (eg, participants set their own time points within prespecified time windows).
- Visualization: Use visual aids, such as Gantt charts to illustrate sequential time use patterns instead of long (dropdown) lists'. Discuss whether to include open-ended or self-labelled responses and allow the logging of multiple overlapping activities.
- Incentives: Prepare for sufficient incentives as an appreciation for participants (eg, financial incentives of about €50 [US $52.33] per person).
- Gamification: If feasible, include elements of gamification (eg, live updates on compliance rates, positive feedback after completing a survey).

Discussion

Compliance and Feasibility

This study examined compliance and feasibility of an EMA of parental time use and well-being. By using qualitative (interviews, open-ended questions) and quantitative (EMA use data, surveys) methods, the study found that compliance was high (with an average of 83% [23.1/28] completed prompts) and feasibility was good for most parents that participated in the study. Older and male participants showed higher compliance. Participants also discussed several challenges to implementation regarding technical issues (eg, system compatibility and missing notifications), integration into daily life (eg, time points and frequency of assessments), and personal preferences (eg, how to log and classify daily activities).

The findings largely correspond to previous research that found similar compliance rates [30,33,36] and also reported a substantial amount of nonresponders due to technical issues [37]. However, in contrast to previous studies, we observed differences between fathers and mothers, with higher compliance rates and acceptance of evening assessments in fathers. Such differences were not found in other reviews that analyzed gender differences in EMA studies [57]. It is possible that mothers experienced excessive burden compared with fathers due to them more often having to balance family and professional demands [7,9,10] and having more informal care duties than fathers, which would give them less time to complete the assessments. Furthermore, fathers primarily having work duties would explain their frustration with repeated time use assessments during working hours (ie, monotonous responses at 12 PM and 4:30 PM), and would give them more time to complete evening assessments compared with mothers. However, this is merely speculative, and it is of interest to further examine time management and work-family conflict [15] in association with compliance and feasibility in the future. If work-family conflict affects well-being and time management, this could also be visible in daily life. Our sample was also highly educated, and previous research [37] found that mothers with higher annual incomes responded to prompts more quickly. This could indicate a bias and more research is needed in more diverse samples along different socioeconomic gradients.

Another important issue raised by participants referred to the customization of the assessment (eg, how to log and classify daily activities) and possible unintended consequences (eg, stress due to continued exposure to low well-being or visualization of work-family conflict in the time-use assessment). Since customization is in line with increased app engagement [58], this could be a good way to maintain compliance over time. However, this also requires more resources for the development, implementation, and data analysis. Therefore, more research is needed to explore the trade-off between complexity, compliance, and feasibility of the EMA design.

In addition, so far unintended consequences of such assessments have rarely been addressed in this target group, although research suggests that reactivity due to EMA monitoring is low [59,60]. Nevertheless, it is recommended to provide psychosocial support or resources in case of continuous reports of low levels of well-being. Since this can also affect compliance, additional incentives or reminders might be useful if multiple consecutive surveys are skipped or not completed. Other recommendations include extended training of participants to ensure a common understanding of data collection procedures. Fluctuations in well-being and negative affect can also influence app engagement. For example, parents may be less able to comply with EMA protocols on days when stress levels are higher, when their mood is lower or their children's behavior is more difficult to manage. There is also a possibility of potential reporting bias due to measurement reactivity as well as the risk that parents might skip surveys.

This has implications for ecological momentary interventions and just-in-time adaptive interventions: For instance, participants experiencing high levels of stress might benefit from a brief stress management intervention (eg, breathing techniques and mindfulness minute) to support their momentary emotion regulation before completing a more complex survey. This could not only improve their well-being but also their compliance. On the other hand, participants reporting positive mood might benefit from further reinforcement (in the sense of an upward spiral of positive emotions, following the broaden-and-build theory [61]) to sustain their positive mood and also their compliance. Therefore, analyzing stress, well-being, and app compliance concurrently is useful to account for days with higher or lower levels of engagement and well-being. Future studies may also want to use different types of operationalization of well-being or stress [31] to identify key moments to increase engagement and ensure compliance. For instance, participants

XSL•FO
RenderX

with low compliance could receive compliance-facilitating interventions, such as higher incentives or gamification elements, which could then lead to self-efficacy and a positive mood. In summary, we argue that looking more closely at the association between compliance, well-being, and stress might be beneficial to study quality (ie, lower attrition, higher retention rates) as well as participant well-being (eg, via EMIs or just-in-time adaptive interventions).

Finally, we present key recommendations for EMA designs assessing time use and well-being in parents (Textbox 1). These include ensuring robust technical support, tailoring assessment schedules to participants' needs, and offering flexibility in survey timing. We also highlight the importance of effective visualizations, sufficient incentives, and incorporating gamification elements to enhance engagement and usability.

Strengths and Limitations

The study is based on a convenience sample of parents, and therefore, it was not representative of the population. Particularly the female-dominant sample limits the generalizability of findings to fathers. In particular, parents with increased vulnerability (eg, few socioeconomic resources or children with a medical condition) were not part of the study, which could have biased the results. Furthermore, the assessment was limited to a few key questions for each outcome, which limits the scope of the examined constructs (eg, well-being). Furthermore, we did not compare different variations of the study design (eg, more or less frequent assessments, active vs passive data collection) and cannot generalize our findings to other methods or assessment schedules. Compliance and feasibility may differ between interval-contingent and random-signal-contingent sampling regimes, as parents in the former can mentally prepare and plan accordingly. Using brief scales and single-item measures is common and recommended in EMA research to limit the potential burden on participants, avoid dropouts, and achieve valid results [32]. Since we assumed parents to be a high-burdened group, because of demands from both family and work environments, it was important to us to use a brief survey for each EMA prompt. We also used psychometrically validated and previously tested items whenever possible to increase the comparability and methodological quality of the assessment. To explore other potential weaknesses and strengths of the study design, we also asked open-ended questions and conducted qualitative interviews to explore participants' perspectives and gain a better understanding of the challenges they face in this type of study.

Conclusion

In summary, our pilot study demonstrates that EMA methodologies are valuable for investigating parental time use and well-being in everyday life. They present unique opportunities for future research, such as exploring work-family conflict or performative gender roles, complementing established methods like daily diaries. In addition, they enable advanced big data analyses (eg, sequential cross-lagged panel models, sequence-based grouping algorithms) and support data-driven interventions that address the complex relationship between voluntary and involuntary daily activities, gender, and health.

Acknowledgments

We would like to thank Fredericke Schmitz for a preliminary literature review, Fredericke Schmitz and Pia Dietrich for the transcription of the interview material, and Anna Stoinski, Jette Giese, Heidi Schurig, Finja Knibbe, and Annalena Krupskin for supporting the preliminary coding of the qualitative material. Funding was received from the University of Greifswald for, for example, incentives and software licenses (F1704100).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

LA and ST contributed to conceptualization. LA performed data curation and formal analysis. LA and ST were involved in funding acquisition. LA participated in writing – original draft. LA and ST were responsible for writing – reviewing and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questions regarding feasibility in the baseline and follow-up questionnaires.
[DOCX File , 17 KB - formative_v9i1e67451_app1.docx]

Multimedia Appendix 2
Full ecological momentary assessment questionnaire.
[DOCX File , 25 KB - formative_v9i1e67451_app2.docx]

Multimedia Appendix 3

XSL•FO
RenderX

Full interview guide.
[DOCX File , 18 KB - formative_v9i1e67451_app3.docx]

Multimedia Appendix 4
Compliance by survey day and time point.
[DOCX File , 18 KB - formative_v9i1e67451_app4.docx]

Multimedia Appendix 5
Frequency and description of the themes from the interviews (N=13).
[DOCX File , 35 KB - formative_v9i1e67451_app5.docx]

Multimedia Appendix 6
Technical aspects.
[DOCX File , 23 KB - formative_v9i1e67451_app6.docx]

Multimedia Appendix 7
Frequency and description of the themes from the baseline questionnaires (N=74).
[DOCX File , 23 KB - formative_v9i1e67451_app7.docx]

Multimedia Appendix 8
Frequency and description of the themes from the follow-up questionnaires (N=57).
[DOCX File , 26 KB - formative_v9i1e67451_app8.docx]

References

1. Matias M, Fontaine AM. Coping with work and family: how do dual-earners interact? Scand J Psychol 2015;56(2):212-222. [doi: 10.1111/sjop.12195] [Medline: 25600426]
2. Bauman A, Bittman M, Gershuny J. A short history of time use research; implications for public health. BMC Public Health 2019;19(Suppl 2):607 [FREE Full text] [doi: 10.1186/s12889-019-6760-y] [Medline: 31159790]
3. Bischoff M, Schmidt S, Muehlan H, Ulbricht S, Heckmann M, Berg NVD, et al. Ecological momentary assessment of parent-child attachment via technological devices: a systematic methodological review. Infant Behav Dev 2023;73:101882. [doi: 10.1016/j.infbeh.2023.101882] [Medline: 37633249]
4. Keijsers L, Boele S, Bülow A. Measuring parent-adolescent interactions in natural habitats. The potential, status, and challenges of ecological momentary assessment. Curr Opin Psychol 2022;44:264-269 [FREE Full text] [doi: 10.1016/j.copsyc.2021.10.002] [Medline: 34788708]
5. J Gershuny. Gender symmetry, gender convergence and historical work-time invariance in 24 countries. Centre for Time Use Research. 2018. URL: https://www.timeuse.org/sites/default/files/2018-02/CTUR%20WP%202%202018_0.pdf [accessed 2025-02-26]
6. Sayer LC. Trends in women's and men's time use, 1965–2012: back to the future? In: Gender and Couple Relationships. Cham: Springer; 2016:43-77.
7. Altweck L, Hahm S, Schmidt S, Ulke C, Fleischer T, Helmert C, et al. Even now women focus on family, men on work: an analysis of employment, marital, and reproductive life-course typologies in relation to change in health-related quality of life. App Res Qual Life 2022 Dec 02;18(3):1205-1223. [doi: 10.1007/s11482-022-10130-3]
8. Rubiano-Matulevich E, Viollaz M. Gender Differences in Time Use: Allocating Time between the Market and the Household. Washington, DC: World Bank; 2019.
9. Tomczyk S, Altweck L, Schmidt S. How is the way we spend our time related to psychological wellbeing? A cross-sectional analysis of time-use patterns in the general population and their associations with wellbeing and life satisfaction. BMC Public Health 2021;21(1):1858 [FREE Full text] [doi: 10.1186/s12889-021-11712-w] [Medline: 34649542]
10. Sullivan O. Gender inequality in work-family balance. Nat Hum Behav 2019;3(3):201-203 [FREE Full text] [doi: 10.1038/s41562-019-0536-3] [Medline: 30953017]
11. Anxo D, Mencarini L, Pailhé A, Solaz A, Tanturri ML, Flood L. Gender differences in time use over the life course in France, Italy, Sweden, and the US. Fem Econ 2011 Jul;17(3):159-195. [doi: 10.1080/13545701.2011.582822]
12. Coltrane S. Research on household labor: modeling and measuring the social embeddedness of routine family work. J Marriage Fam 2004;62(4):1208-1233. [doi: 10.1111/j.1741-3737.2000.01208.x]
13. Offer S. Free time and emotional well-being. Gender & Society 2015 Aug 03;30(2):213-239. [doi: 10.1177/0891243215596422]
14. Hobfoll SE. Conservation of resources. A new attempt at conceptualizing stress. Am Psychol 1989;44(3):513-524. [doi: 10.1037//0003-066x.44.3.513] [Medline: 2648906]

XSL•FO
RenderX

15. Greenhaus JH, Beutell NJ. Sources of conflict between work and family roles. Acad Manage Rev 1985;10(1):76-88. [doi: 10.2307/258214]
16. West C, Zimmerman DH. Doing gender. Gend Soc 1987;1(2):125-151. [doi: 10.1177/089124328700100200] [Medline: 22489612]
17. Butler J. Gender Trouble: Feminism and The Subversion of Identity. New York, NY: Routledge; 1990.
18. Altweck L, Tomczyk S, Schmidt S. Under contract and in good health: a multigroup cross-lagged panel model of time use and health-related quality of life in working-age men and women. Health Qual Life Outcomes 2022;20(1):151 [FREE Full text] [doi: 10.1186/s12955-022-02052-3] [Medline: 36376955]
19. Gálvez-Muñoz L, Rodríguez-Modroño P, Domínguez-Serrano M. Work and time use by gender: a new clustering of European welfare systems. Feminist Economics 2011 Oct;17(4):125-157. [doi: 10.1080/13545701.2011.620975]
20. Biddle S, Whitehead S, O'Donovan T, Nevill M. Correlates of participation in physical activity for adolescent girls: a systematic review of recent literature. J Phys Act Health 2005;2(4):423-434. [doi: 10.1123/jpah.2.4.423]
21. Dumas O, Le Moual N. Damaging effects of household cleaning products on the lungs. Expert Rev Respir Med 2020;14(1):1-4 [FREE Full text] [doi: 10.1080/17476348.2020.1689123] [Medline: 31682770]
22. Runyan JD, Steenbergh TA, Bainbridge C, Daugherty DA, Oke L, Fry BN. A smartphone ecological momentary assessment/intervention "app" for collecting real-time data and promoting self-awareness. PLoS One 2013;8(8):e71325 [FREE Full text] [doi: 10.1371/journal.pone.0071325] [Medline: 23977016]
23. Burgess-Hull A, Epstein DH. Ambulatory assessment methods to examine momentary state-based predictors of opioid use behaviors. Curr Addict Rep 2021;8(1):122-135 [FREE Full text] [doi: 10.1007/s40429-020-00351-7] [Medline: 33425652]
24. Mengelkoch S, Moriarity DP, Novak AM, Snyder MP, Slavich GM, Lev-Ari S. Using ecological momentary assessments to study how daily fluctuations in psychological states impact stress, well-being, and health. J Clin Med 2023;13(1):24 [FREE Full text] [doi: 10.3390/jcm13010024] [Medline: 38202031]
25. Gadermann AM, Zumbo BD. Investigating the intra-individual variability and trajectories of subjective well-being. Soc Indic Res 2006;81:1-33. [doi: 10.1007/s11205-006-0015-x]
26. Akay A, Martinsson P. Sundays are blue: aren't they? The day-of-the-week effect on subjective well-being and socio-economic status. SSRN Journal 2009:1-47. [doi: 10.2139/ssrn.1506315]
27. Stieger S, Reips U. Well-being, smartphone sensors, and data from open-access databases: a mobile experience sampling study. Field Methods 2019;31(3):277-291. [doi: 10.1177/1525822x18824281]
28. Eurostat. Harmonised European Time Use Surveys — 2018 Guidelines. Luxembourg: Eurostat; 2009.
29. Sonnenberg B, Riediger M, Wrzus C, Wagner GG. Measuring time use in surveys – how valid are time use questions in surveys? Concordance of survey and experience sampling measures. SSRN Journal 2012;41(5):1037-1052. [doi: 10.2139/ssrn.1895307]
30. Liao Y, Skelton K, Dunton G, Bruening M. A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: an adapted STROBE checklist for reporting EMA studies (CREMAS). J Med Internet Res 2016;18(6):e151 [FREE Full text] [doi: 10.2196/jmir.4954] [Medline: 27328833]
31. Heron KE, Everhart RS, McHale SM, Smyth JM. Using mobile-technology-based ecological momentary assessment (EMA) methods with youth: a systematic review and recommendations. J Pediatr Psychol 2017;42(10):1087-1107. [doi: 10.1093/jpepsy/jsx078] [Medline: 28475765]
32. de Vries L, Bartels M. Individual differences in wellbeing dynamics: a genetically-informed comparison of Ecological Momentary Assessment and longitudinal survey data. Pers Individ Differ 2025 Apr;236:112996. [doi: 10.1016/j.paid.2024.112996]
33. Passini CM, Pihet S, Favez N, Schoebi D. Assessment of parental discipline in daily life. J Fam Psychol 2013;27(2):324-329. [doi: 10.1037/a0031504] [Medline: 23458698]
34. Sherman B, Huang I, Wijaya E, Turk-Browne N, Goldfarb EV. Acute stress effects on statistical learning and episodic memory. J Cogn Neurosci 2024;36(8):1741-1759. [doi: 10.31234/osf.io/546nv]
35. Kerr ML, Buttitta KV, Smiley PA, Rasmussen HF, Borelli JL. Mothers' real-time emotion as a function of attachment and proximity to their children. J Fam Psychol 2019;33(5):575-585. [doi: 10.1037/fam0000515] [Medline: 30742463]
36. El Dahr Y, Perquier F, Moloney M, Woo G, Dobrin-De Grace R, Carvalho D, et al. Feasibility of using research electronic data capture (REDCap) to collect daily experiences of parent-child dyads: ecological momentary assessment study. JMIR Form Res 2023;7:e42916 [FREE Full text] [doi: 10.2196/42916] [Medline: 37943593]
37. Dzubur E, Huh J, Maher JP, Intille SS, Dunton GF. Response patterns and intra-dyadic factors related to compliance with ecological momentary assessment among mothers and children. Transl Behav Med 2018;8(2):233-242 [FREE Full text] [doi: 10.1093/tbm/ibx002] [Medline: 29381785]
38. Altweck L, Schmidt S, Tomczyk S. Daily time-use patterns and quality of life in parents: protocol for a pilot quasi-experimental, nonrandomized controlled trial using ecological momentary assessment. JMIR Res Protoc 2024;13:e54728 [FREE Full text] [doi: 10.2196/54728] [Medline: 38820576]
39. Boemo T, Nieto I, Vazquez C, Sanchez-Lopez A. Relations between emotion regulation strategies and affect in daily life: a systematic review and meta-analysis of studies using ecological momentary assessments. Neurosci Biobehav Rev 2022;139:104747 [FREE Full text] [doi: 10.1016/j.neubiorev.2022.104747] [Medline: 35716875]

XSL•FO
RenderX

40. Guest G, Bunce A, Johnson L. How many interviews are enough?: An experiment with data saturation and variability. Field Methods 2006;18(1):59-82. [doi: 10.1177/1525822X05279903]
41. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022;292:114523 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
42. Crespo C, Santos S, Canavarro MC, Kielpikowski M, Pryor J, Féres-Carneiro T. Family routines and rituals in the context of chronic conditions: a review. Int J Psychol 2013;48(5):729-746. [doi: 10.1080/00207594.2013.806811] [Medline: 23848452]
43. Classifying educational programmes: manual for ISCED-97 implementation in OECD countries. Organisation for Economic Co-operation and Development. 1999. URL: https://www.voced.edu.au/content/ngv:11701 [accessed 2025-02-26]
44. The Whoqol Group. Development of the world health organization WHOQOL-BREF quality of life assessment. The WHOQOL group. Psychol Med 1998;28(3):551-558. [doi: 10.1017/s0033291798006667] [Medline: 9626712]
45. Blanke ES, Riediger M, Brose A. Pathways to happiness are multidirectional: associations between state mindfulness and everyday affective experience. Emotion 2018;18(2):202-211. [doi: 10.1037/emo0000323] [Medline: 28481572]
46. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983;24(4):385-396. [doi: 10.2307/2136404]
47. Blanke ES, Brose A. Multidimensional state mindfulness questionnaire (MSMQ). In: Medvedev ON, Krägeloh CU, Siegert RJ, Singh NN, editors. Handbook of Assessment in Mindfulness Research. Cham: Springer; 2022:1-13.
48. Businelle MS, Ma P, Kendzor DE, Frank SG, Vidrine DJ, Wetter DW. An ecological momentary intervention for smoking cessation: evaluation of feasibility and effectiveness. J Med Internet Res 2016;18(12):e321 [FREE Full text] [doi: 10.2196/jmir.6058] [Medline: 27956375]
49. Messner EM, Terhorst Y, Barke A, Baumeister H, Stoyanov S, Hides L, et al. The German version of the Mobile App Rating Scale (MARS-G): development and validation study. JMIR mHealth uHealth 2020;8(3):e14479 [FREE Full text] [doi: 10.2196/14479] [Medline: 32217504]
50. Spook JE, Paulussen T, Kok G, Van Empelen P. Monitoring dietary intake and physical activity electronically: feasibility, usability, and ecological validity of a mobile-based ecological momentary assessment tool. J Med Internet Res 2013;15(9):e214 [FREE Full text] [doi: 10.2196/jmir.2617] [Medline: 24067298]
51. Alqahtani F, Winn A, Orji R. Co-designing a mobile app to improve mental health and well-being: focus group study. JMIR Form Res 2021;5(2):e18172 [FREE Full text] [doi: 10.2196/18172] [Medline: 33635281]
52. Olsen SH, Saperstein SL, Gold RS. Content and feature preferences for a physical activity app for adults with physical disabilities: focus group study. JMIR mHealth uHealth 2019;7(10):e15019 [FREE Full text] [doi: 10.2196/15019] [Medline: 31605518]
53. Jamovi. 2024. URL: https://www.jamovi.org [accessed 2025-02-26]
54. MAXQDA 2022. Berlin, Germany: VERBI Software; 2021. URL: https://www.maxqda.com/about [accessed 2025-02-26]
55. Dao KP, De Cocker K, Tong HL, Kocaballi AB, Chow C, Laranjo L. Smartphone-delivered ecological momentary interventions based on ecological momentary assessments to promote health behaviors: systematic review and adapted checklist for reporting ecological momentary assessment and intervention studies. JMIR mHealth uHealth 2021;9(11):e22890 [FREE Full text] [doi: 10.2196/22890] [Medline: 34806995]
56. Kuckartz U, Rädiker S. Qualitative Content Analysis. Methods, Practice. Computer Support [Book in German]. Weinheim and Basel: Beltz Juventa; 2024.
57. Wrzus C, Neubauer AB. Ecological momentary assessment: a meta-analysis on designs, samples, and compliance across research fields. Assessment 2023;30(3):825-846 [FREE Full text] [doi: 10.1177/10731911211067538] [Medline: 35016567]
58. Oakley-Girvan I, Yunis R, Longmire M, Ouillon JS. What works best to engage participants in mobile app interventions and e-health: a scoping review. Telemed J E Health 2022;28(6):768-780 [FREE Full text] [doi: 10.1089/tmj.2021.0176] [Medline: 34637651]
59. Hufford MR, Shields AL, Shiffman S, Paty J, Balabanis M. Reactivity to ecological momentary assessment: an example using undergraduate problem drinkers. Psychol Addict Behav 2002;16(3):205-211. [doi: 10.1037/0893-164X.16.3.205]
60. Toth R, Trifonova T. Somebody's watching me: smartphone use tracking and reactivity. Comput Hum Behav Rep 2021;4:100142. [doi: 10.1016/j.chbr.2021.100142]
61. Fredrickson BL. The broaden-and-build theory of positive emotions. Philos Trans R Soc Lond B Biol Sci 2004;359(1449):1367-1378 [FREE Full text] [doi: 10.1098/rstb.2004.1512] [Medline: 15347528]

Abbreviations

EMA: ecological momentary assessment
WHO-QoL: World Health Organization Quality of Life

XSL•FO
RenderX

Edited by A Mavragani; submitted 18.10.24; peer-reviewed by L de Vries, AKC Chau; comments to author 19.12.24; revised version received 17.01.25; accepted 24.01.25; published 23.04.25.

Please cite as:
Altweck L, Tomczyk S
Ecological Momentary Assessment of Parental Well-Being and Time Use: Mixed Methods Compliance and Feasibility Study
JMIR Form Res 2025;9:e67451
URL: https://formative.jmir.org/2025/1/e67451
doi:10.2196/67451
PMID:

©Laura Altweck, Samuel Tomczyk. Originally published in JMIR Formative Research (https://formative.jmir.org), 23.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Protocol

Incorporating Patient-Reported Outcome Measures and Patient-Reported Experience Measures in Addiction Treatment Services in Belgium: Naturalistic, Longitudinal, Multicenter Cohort Study

Amine Zerrouk[1], MSc; Charlotte Migchels[2], MD, MSc; Clara De Ruysscher[1], PhD; Kim Fernandez[3], MSc; Jerome Antoine[3], MSE; Florian De Meyer[1], MSc; Frieda Matthys[2], MD, PhD; Wim van den Brink[4], MD, PhD; Cleo Lina Crunelle[2], MD, PhD; Wouter Vanderplasschen[1], PhD

[1]Department of Special Needs Education, Ghent University (UGent), Ghent, Belgium

[2]Department of Psychiatry, Universitair Ziekenhuis Brussel (UZ Brussel), Vrije Universiteit Brussel (VUB), Brussels, Belgium

[3]Department of Epidemiology and Public Health, Sciensano, Brussels, Belgium

[4]Department of Psychiatry, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, The Netherlands

Corresponding Author:
Amine Zerrouk, MSc
Department of Special Needs Education
Ghent University (UGent)
Henri Dunantlaan 1
Ghent, 9000
Belgium
Phone: 32 93310310
Email: ElAmine.Zerrouk@Ugent.be

Abstract

Background: Traditionally, treatment outcomes of service users with a substance use disorder (SUD) are measured using objective and provider-reported indicators. In recent years, there has been a shift toward incorporating patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs) to capture service users' perspectives on treatment outcomes and experiences.

Objective: The OMER-BE (Outcome Measurement and Evaluation as a Routine Practice in Alcohol and Other Drug Services in Belgium) study evaluates the acceptability and feasibility of PROMs and PREMs in different SUD treatment services, using the recently developed International Consortium for Health Outcomes Measurement Standard Set for Addictions. This paper presents the design and baseline characteristics of the study, indicators of attrition at 45-day follow-up, and the feasibility of the implementation of PROMs and PREMs in residential and outpatient services.

Methods: A convenience sample of 189 treatment-seeking individuals with SUD from different inpatient (therapeutic communities and psychiatric centers) and outpatient treatment services was followed for six months. Sociodemographic characteristics; clinical factors; and PROMs including recovery strengths, quality of life, and global health were assessed at baseline and within 3 weeks after starting treatment. Additionally, PROMs and PREMs were measured 45, 90, and 180 days later. Comparisons were made between treatment modalities, and indicators of attrition at the 45-day follow-up were assessed using ANOVA and chi-square tests.

Results: Baseline differences were observed between the three treatment modalities regarding education, SUD treatment history, primary substance, and Attention-Deficit/Hyperactivity Disorder Self-Report scores. Overall, patients in psychiatric treatment centers had a higher education level and less polysubstance use, while outpatients had fewer previous SUD treatments but received relatively more often opioid agonist treatment. Inpatients reported more attention-deficit/hyperactivity disorder symptoms and higher SUD severity than outpatients. Additionally, recovery strength scores were significantly lower in the outpatient group compared to the other groups, particularly in the subdomains of "Substance Use," "Self-care," and "Outlook on Life." At the 45-day follow-up assessment, the attrition rate was 36.6%. Comparisons between participants who completed the 45-day follow-up and those who dropped out revealed that completers were significantly older, had a higher level of education, were more likely to live alone, and were more likely to have a mother born in Belgium. They also had higher average scores on the "Material

XSL•FO
RenderX

Resources" domain of the Substance Use Recovery Evaluator, which includes questions about stable housing, a steady income, and effective financial management.

Conclusions: Evaluating PROMs and PREMs appears to be feasible in a diverse group of treatment-seeking patients with SUD in Belgium. However, challenges remain for structural implementation in practice, especially in outpatient services. Routine monitoring of PROMs and PREMs has the potential to empower patients, service providers, and policy makers by providing a comprehensive understanding of service users' needs and treatment effectiveness.

(JMIR Form Res 2025;9:e65686) doi:10.2196/65686

KEYWORDS

patient-reported outcome measures; patient-reported experience measures; substance use disorder; recovery; ICHOM; International Consortium for Health Outcomes Measurement; addiction; PROMs; PREMs; SUD; treatment; protocol; substance use; inpatient; services; perspectives; treatment outcome

Introduction

Alcohol and other substance use disorders (SUDs) are linked to a range of adverse psychological, physical, and social consequences. Their chronic, relapsing nature and related judicial, housing, and relational problems impact individuals with SUDs, as well as their environment and the broader community [1-4]. SUDs have a significant and growing impact on global morbidity and mortality [5-7]. Worldwide, harmful alcohol use causes 3 million deaths annually, representing 5.3% of all deaths. Alcohol use accounts for 5.1% of the global burden of disease [8]. Furthermore, an estimated 60,000 years of life lost were attributed to drug use in Europe in 2019 [5].

Treatment cohort studies conducted in the United States, Australia, and various European countries have shown the benefits of engaging in SUD treatment, generally resulting in increased abstinence rates, improved social integration, and reduced psychopathology [9-12]. Various SUD treatment modalities, however, seem to impact treatment outcomes in different ways. Stahler et al [13], for example, found that people with problematic opioid use benefitted more from residential treatment than individuals with alcohol as their primary substance. Conversely, people with problematic cannabis use were less likely to benefit from residential treatment compared to people with alcohol as their primary substance. Several studies found higher treatment completion rates in residential programs compared to outpatient settings [13,14]. Treatment engagement and retention have been consistently associated with positive outcomes, irrespective of the treatment modality [15-17]. Individuals who stay in treatment for a longer period of time are more likely to remain abstinent, experience fewer relapses and readmissions, engage less in criminal activity, and show greater improvements in general health measures [13,14,18]. Despite evidence for the effectiveness of SUD treatment, significant challenges such as high dropout and relapse rates persist and warrant further research and innovative approaches [16].

SUD treatment outcomes are typically measured using objective (eg, abstinence and rearrest or reincarceration) or provider-reported indicators (eg, treatment completion or compliance and absence of symptoms). Increasing emphasis on service users' perspectives in measuring treatment outcomes and experiences has recently led to the introduction of patient-reported outcome measures (PROMs) and patient-reported experience measures (PREMs) in SUD and other mental health services [19-22]. PROMs refer to individual, subjective treatment outcomes, including information about psychological well-being, quality of life, symptomatology, and physical functioning. PREMs provide information on how individuals experience health care and measure practical aspects of care, such as coordination, continuity and accessibility of care, and quality of patient-provider relationships [23]. International organizations, such as the International Consortium for Health Outcomes Measurement (ICHOM) [24], promote the implementation of patient-reported measures for routine assessment of health outcomes and experiences in all health care areas, including primary care, psychotherapy, and SUD services [21]. The use of PROMs and PREMs is advised to systematically monitor and improve the delivery of effective, patient-centered care and shared decision-making, which is not a standard practice in SUD and other mental health services [20,21,25-30]. Additionally, adopting standardized outcome measures opens opportunities to compare performances between treatment services within regions, countries, and even globally. This evolution will facilitate knowledge sharing among practitioners and provide policy makers with the tools and evidence needed to improve the quality and effectiveness of care [24]. The systematic use of PREMs is also likely to advance the field toward more personalized and effective support since it provides a direct evaluation of the accessibility, continuity, and coordination of care by service users [21,23]. Yet, although PROMs and PREMs show clear promise in improving the quality of SUD treatment, there is a lack of research on (the implementation of) these measures in clinical practice within SUD treatment settings [21]. The OMER-BE (Outcome Measurement and Evaluation as a Routine Practice in Alcohol and Other Drug Services in Belgium) study (2022-2025) aims to introduce and integrate the use of PROMs and PREMs in different SUD treatment modalities and to evaluate and enhance the quality of care provided in SUD treatment services in Belgium. The primary objective of the OMER-BE study is to assess the routine measurement of PROMs and PREMs in SUD treatment services, based on the recently developed and internationally validated ICHOM Standard Set for Addictions (ICHOM SSA) [24]. Therefore, we translated and validated the ICHOM SSA into Dutch and French, enhancing its applicability in Belgium and other Dutch- and French-speaking countries. By incorporating PROMs and PREMs in routine assessments, SUD treatment services are provided with information to better

XSL•FO
RenderX

align treatment approaches with service user needs. The secondary objective of the OMER-BE study is to measure and compare various recovery indicators and treatment experiences between different treatment modalities and assess their evolution during and after treatment. In this paper, we describe the design and baseline characteristics of the study. In order to enhance study adherence and routine implementation of PROMs and PREMs, we also examined indicators of attrition.

Methods

Study Setting

The OMER-BE study was a naturalistic, longitudinal, multicenter cohort study, in which 189 individuals with SUDs were followed up over a 6-month period in various residential and outpatient treatment modalities in the Dutch- and French-speaking regions of Belgium. Data collection for the study started in July 2022. Participants were recruited from 8 residential services (4 psychiatric treatment centers [PCs] and 4 therapeutic communities [TCs]) and 9 outpatient treatment services.

Specialized wards of PCs offer long-term (3 to 6 months) residential care that provides intensive medical and psychological support, addressing SUDs and in some cases co-occurring mental health disorders. Treatment consists of group counseling, psychoeducation, individual psychotherapy, and occupational activities.

TCs for addictions have a long history and were set up for individuals with SUDs, complementing traditional mental health care services that were traditionally not open to persons with drug problems [31]. In a TC, individuals with SUDs live together in a structured environment, typically for a period of 6 to 12 months, aiming for positive changes that lead to a drug-free life in society. The TC approach centers around the concept of “community as a method,” thus highlighting the influential role of peers and the power of mutual support in fostering recovery [32,33].

Outpatient treatment services provide more autonomy to service users and offer various nonresidential care options consisting of drug-free counseling interventions and harm-reduction approaches like opioid agonist therapy (OAT) and needle exchange programs. OAT refers to the use of opioid replacement medication such as methadone or buprenorphine to help manage withdrawal symptoms and reduce craving, often combined with some form of counseling and social support. Drug-free counseling focuses on psychosocial interventions, such as motivational interviewing and cognitive behavioral therapy, aimed at helping individuals develop coping strategies and support systems to become or remain abstinent.

Study Sample

The OMER-BE study aims to follow up a naturalistic cohort of service users with SUDs as they start a new treatment episode in a selected number of SUD treatment services, focusing on individuals with a primary alcohol or primary (illicit) drug problem. Eligibility criteria were (1) having a documented SUD (eg, a *Diagnostic and Statistical Manual of Mental Disorders* diagnosis of a SUD or previous treatment for a SUD), (2) being at least 18 years old, (3) being able to communicate in Dutch or French, and (4) having started treatment no longer than 21 days ago. First-time users, as well as returning service users, were considered eligible for this cohort study, as long as they started a new treatment episode during the recruitment period [34].

Upon treatment entry, service users were informed about the aims and design of the OMER-BE study through posters, leaflets, and staff members of the selected treatment facilities. During the initial meeting with the researcher, participants were informed extensively about the (follow-up) study and the implications of study participation and were asked for written informed consent to participate [34].

Study Procedure

Baseline Assessment

Sociodemographic and clinical factors (Textbox 1) were assessed at baseline, followed by the assessment of PROMs (Textbox 1). Data were collected through self-report using a tablet provided by one of the researchers, who were available throughout the assessment to address any questions participants may have. Administering the set of baseline variables and PROMs took between 20 and 45 minutes. After completion of the baseline assessment, participants received a voucher of €10 (US $10.46) as remuneration.

XSL•FO
RenderX

Textbox 1. Overview of included instruments in the OMER-BE (Outcome Measurement and Evaluation as a Routine Practice in Alcohol and Other Drug Services in Belgium) outcome measurement tool.

Sociodemographic factors

- Year of birth, sex, highest level of education completed, housing status, ethnicity

Clinical factors

- Treatment history for substance use disorders
- Primary Care Posttraumatic Stress Disorder Screen for Diagnostic and Statistical Manual of Mental Disorders (PC-PTSD-5)
- Depression, Anxiety, and Stress Scale (DASS-21)
- Adult Attention-Deficit/Hyperactivity Disorder Self-Report Scale (ASRS-v1.1)

Patient-reported outcome measures

- Patient-Reported Outcomes Measurement Information System (PROMIS) Short Form v1.0—Alcohol Use 7a (PROMIS-Alcohol)
- PROMIS SF v1.0—Severity of Substance Use 7a (PROMIS-Substance)
- Heaviness of Smoking Index (HSI)
- National Health Service Treatment Outcomes Profile for Substance Misuse—section 1 (TOP-S1)
- PROMIS Scale v1.2—Global Health (PROMIS-GH-10)
- Substance Use Recovery Evaluator (SURE)
- Brief version of the World Health Organization Quality of Life Scale (WHOQoL-BREF)

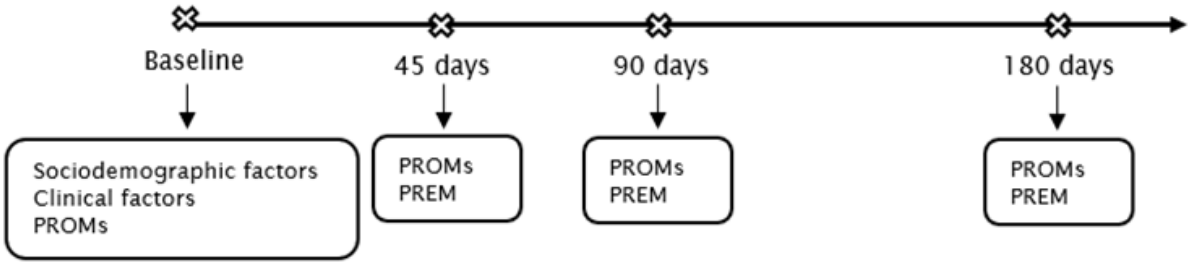
Patient-reported experience measures

- Patient-Reported Experience Measure for Addiction Treatment (PREMAT)

Follow-up Assessments

Study participants are contacted again 45, 90, and 180 days after the baseline assessment (Figure 1). We use a time window of four weeks for the 45- and 90-day follow-up and five weeks for the 180-day follow-up assessment to collect and complete the questionnaires. Researchers contacted all study participants either via email, SMS text messages, directly by phone, or through the treatment setting where they were enrolled in the study. Participants can complete the online survey via a personalized link provided by email or during a face-to-face or telephone interview, depending on their preference. The follow-up assessments include a measurement of both PROM and PREM variables (Textbox 1), which are supposed to take 20 to 30 minutes to complete. Participants receive a €10 (US $10.46) voucher for each completed follow-up assessment.

Figure 1. Overview of the data collection process. PREM: patient-reported experience measure; PROM: patient-reported outcome measure.



As the OMER-BE project is a naturalistic study, some participants will no longer be in treatment at the 45-, 90- or 180-day follow-up. Consequently, the PREM instrument is only administered when service users are still in treatment and at the first follow-up moment after leaving treatment. For example, if a participant stops treatment after 50 days, the PREM questionnaire will be administered at the 45- and 90-day follow-up moments.

XSL•FO
RenderX

Instruments

Overview

The baseline and follow-up assessments (Textbox 1) are largely based on the ICHOM SSA [24], a set of brief, existing, validated questionnaires to measure and monitor treatment outcomes routinely in SUD services that were developed by an international panel of SUD specialists. The ICHOM SSA focuses on patient-centered outcome indicators and provides an internationally agreed-upon method for measuring a variety of outcome domains. The tool offers potential for routine use since it is relatively short and has been specifically developed for and validated in the population of SUD service users. It can be easily administered and is applicable in a wide range of treatment settings [24]. To facilitate its application in Belgium, nontranslated questionnaires were translated into French and Dutch using forward or backward translation, following guidelines provided by Tsang et al [35], and subsequently validated. Compared to the ICHOM SSA procedure, we added a 45-day follow-up assessment, which allowed us to have an additional measurement point, keep participants more engaged in the study, and reduce attrition.

The OMER-BE measurement tool consists of three sections (Textbox 1): (1) sociodemographic and clinical factors, (2) PROMs, and (3) PREM.

Sociodemographic and Clinical Factors

Overview

The first section of the tool includes sociodemographic and clinical factors that may influence treatment outcomes.

The following sociodemographic variables were assessed: age, sex, education level, current living situation, country of birth of the participants, and country of birth of their parents.

Clinical factors include questions regarding SUD treatment history and three validated and widely used screening instruments to assess common comorbid psychiatric disorders (trauma, depression or anxiety, and attention-deficit/hyperactivity disorder [ADHD]) that are likely to affect treatment outcomes. These clinical factors were included to ensure a comprehensive understanding of how these variables influence recovery trajectories.

Primary Care Posttraumatic Stress Disorder Screen for Diagnostic and Statistical Manual of Mental Disorders

The Primary Care Posttraumatic Stress Disorder (PTSD) Screen [36] was developed to assess the occurrence of symptoms of PTSD over the last month. This 5-item screening tool is rated on a binary scale (No=0 and Yes=1). Higher total sum scores suggest the presence of more PTSD symptoms. The reliability of the English version has been found satisfactory in a sample of people with SUDs, as measured by a Cronbach α value of 0.73 [37].

Depression, Anxiety, Stress Scale

The Depression, Anxiety, and Stress Scale [38] is a commonly used instrument consisting of 21 items, divided into three subscales (depression, anxiety, and stress), each containing seven items. Each question is rated on a scale from 0=did not apply to me at all to 3=applied to me very much. Subscales and total scores are calculated by adding up the scores on the items and multiplying these by a factor of 2. In a population of Dutch-speaking SUD service users in the Netherlands, Beaufort et al [39] found the total score to be highly reliable (Cronbach α=0.91).

Adult ADHD Self-Report Scale

The Adult ADHD-Self-Report Scale v1.1 consists of several questions encompassing each of the 18 symptoms cited in the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* [40]. This study uses the short version, consisting of six items that have been found the most predictive of an ADHD diagnosis. These items are rated on a scale from 0=never to 4=very often. Scores equal to or higher than 2 for items 1-3 and equal to or higher than 3 on items 4-6 are indicative of a diagnosis of ADHD. A Spanish study by Blanco et al [41] showed a sensitivity of 0.88 and a specificity of 0.69 for a diagnosis of ADHD in a sample of outpatients with SUDs. In an international study among treatment-seeking in- and outpatients with an SUD, very similar results were obtained with a sensitivity of 0.83 and a specificity of 0.68 for *Diagnostic and Statistical Manual of Mental Disorders* ADHD [42].

PROMs

Overview

The second section of the tool is focused on PROMs and is based on the ICHOM SSA [24]. To extend the focus to subjective well-being beyond substance use and health outcomes, we added a widely used instrument to assess the quality of life and overall well-being (ie, brief version of the World Health Organization Quality of Life questionnaire [WHOQoL-BREF]).

Patient-Reported Outcomes Measurement Information System Alcohol Use Short Form 7a

The Patient-Reported Outcomes Measurement Information System (PROMIS) Alcohol Use Short Form [43] is a 7-item self-report questionnaire, derived from the 37-item PROMIS Alcohol Use item bank. This tool is designed to assess alcohol use in the last 30 days. Respondents were instructed to complete the seven questions only if they consumed alcohol in the last 30 days. Answers are rated on a 5-point scale ranging from 0=never to 4=almost always. The internal consistency of the total score has proven to be excellent in participants from the general population and a clinical sample of service users in treatment for SUDs (Cronbach α=0.95) [43].

PROMIS SF v1.0—Severity of Substance Use 7a

The PROMIS Severity of Substance Use Short Form [44] consists of seven items and is a shorter version of the 37-item PROMIS Severity of Substance Use item bank. This tool is designed to assess the severity of substance use in the last 30 days. Answers are rated on a 5-point scale ranging from 0=never to 4=almost always. The internal consistency of the total score was found to be excellent in participants from the general population and a clinical sample of service users in treatment for SUDs (Cronbach α=0.94) [44].

XSL•FO
RenderX

Heaviness of Smoking Index

The Heaviness of Smoking Index consists of two items: "How soon after waking up do you usually have your first smoke?" and a question assessing the number of cigarettes smoked each day. The first question was rated on a 4-point scale ranging from more than 60 minutes (0) to less than 5 minutes (3). The second question was replaced by an assessment of the number of cigarettes smoked in the last 30 days to maintain consistency with the assessment of the other substances. Internal consistency for the total Heaviness of Smoking Index score was found to be relatively low in a sample of male individuals with SUDs and nicotine dependence (Cronbach α=0.49) [45].

National Health Service Treatment Outcomes Profile for Substance Misuse—Section 1

The Treatment Outcomes Profile [46] is a multidimensional assessment instrument for monitoring outcomes in SUD treatment. This questionnaire measures four key life domains (substance use, crime, health, and social functioning). For this study, we only assessed the substance use domain and made modifications to the time frame in line with the other questionnaires. Specifically, participants were asked about their primary substances of use and the number of days and quantity consumed over the past 30 days.

PROMIS Scale v1.2—Global Health

The PROMIS Global Health consists of 10 items, scored on a 5-point scale. Due to overlap with other questionnaires, and in accordance with the ICHOM SSA, we only included two items of this scale relating to physical and mental health. The response options of these items are "excellent," "very good," "good," "fair," and "poor." Its psychometric properties were evaluated using a sample of 4370 individuals from the Dutch general population. Results indicated a 2-factor structure with good internal consistency. The subscales "Global Mental Health" and "Global Physical Health" had Cronbach α of 0.83 and 0.78, respectively [47].

Substance Use Recovery Evaluator

The Substance Use Recovery Evaluator (SURE) [48] is a self-report questionnaire consisting of 21 items, which is completed on a 5-point scale but scored on a 3-point scale (1-3). The first two response options correspond to a score of 3, the third response option to a score of 2, and the final 2 response options to a score of 1. Response options for the first three questions are "never," "on 1 or 2 days," "on 3 or 4 days," "on 5 or 6 days," and "every day," while for the remaining questions, response options are "all of the time," "most of the time," "a fair amount of the time," "a little of the time," and "none of the time." Higher total sum scores suggest greater recovery strengths. The items are categorized into five subscales: "substance use," "relationships," "self-care," "outlook on life," and "material resources." Internal consistency of the SURE total score was found to be high in a sample of current and former SUD service users (Cronbach α=0.92) [48]. Psychometric properties of the translated Dutch version of the SURE showed good internal consistency (Cronbach α=0.83) [49].

World Health Organization Quality of Life Scale

The WHOQoL-BREF [50] consists of 26 items and is a short version of the WHOQoL-100 questionnaire. Questions are rated on a 5-point scale (1-5) and response options, from the lowest to highest score, are "very poor/very dissatisfied/not at all/never," "poor/dissatisfied/a little/seldom," "neither poor nor good/neither satisfied nor dissatisfied/a moderate amount/moderately/quite often," "good/satisfied/very much/mostly/very often," and "very good/very satisfied/an extreme amount/extremely/completely/always." Higher sum scores are an indication of a better quality of life. The items are grouped into four domains: "psychological health," "physical health," "environment" and "social relationships." The WHOQoL-BREF has demonstrated good internal consistency with Cronbach α values for each of the domains ranging from 0.66 to 0.84 [51].

PREMs

Overview

As the objectives of the OMER-BE study also included the measurement of PREMs, the third section of the OMER-BE outcome measurement tool is not part of the ICHOM SSA. We use a newly validated PREM, the Patient-Reported Experience Measure for Addiction Treatment (PREMAT), to assess service users' experiences regarding the treatment they received.

PREMAT

The PREMAT is a recently developed relatively brief (23-item) questionnaire that aims to capture the experiences of people in residential SUD treatment services [52-54]. More precisely, the following topics are addressed in the PREMAT: "Individualized support," "Self-determination and Empowerment," "Program structure," "Treatment environment," "Coordination of care," and "Personal responsibility" [53,54]. The instrument consists of 23 statements rated on a 5-point Likert scale: 1=strongly disagree, 2=disagree, 3=neither agree nor disagree, 4=agree, and 5=strongly agree. The total score ranges from 23 to 115, with higher scores reflecting a more positive experience. Additionally, the PREMAT includes 2 open-ended questions ("How could your experience at this service have been improved?" and "What have been the best things about your experience here?"), which allows respondents to elaborate on certain aspects of the questionnaire or discuss topics that are not covered by the PREMAT items [52]. Internal consistency for the total score was found excellent in a sample of participants from specialist residential SUD treatment services in Australia (Cronbach α=0.91) [53].

Ethical Considerations

This study was granted ethical approval from the Medical Ethics Committee of the University Hospital Brussels on May 11, 2022 (UZ Brussel; BUN: 1432022000071). The participants were informed about the confidentiality of the data they provide, and written informed consent was acquired from all participants before being included in the study. Participants received a €10 (US $10.46) voucher for each completed assessment. All data are treated confidentially and reported anonymously. This study was conducted in accordance with the Declaration of Helsinki.

XSL•FO
RenderX

Data Analysis of Baseline Characteristics

This paper presents baseline characteristics of the study cohort and comparative analyses between participants from different treatment modalities, as well as between participants who completed the 45-day follow-up and those who did not participate in the 45-day assessment. For categorical demographic and patient characteristics, significant differences between treatment modalities were assessed using either the chi-square test or Fisher exact test when the data did not meet the assumptions required for the chi-square test. For continuous variables, we used ANOVA to assess group differences. Welch 2-tailed *t* test was used when the assumption of equal variances was not met. Post hoc comparisons between groups were made using the Tukey test when variances were equal or the Games-Howell test when the assumption of equal variances was violated. Initial power analyses indicated that a minimum of 159 participants were needed to detect moderate effects (with α=0.05 and 80% power). Statistical analyses were conducted using SPSS Statistics (version 29; IBM Corp).

Future analyses will center on longitudinal changes in PROMs and PREMs, using repeated measures and mixed model analyses. The role of mediating and moderating variables (eg, socioeconomic status, comorbidity, and recovery strength) will be assessed, along with differences between treatment modalities. All statistical analyses will be conducted using R studio (Posit PBC) and SPSS Statistics (IBM Corp).

Results

In total, 189 individuals participated in the OMER-BE study at baseline, of which 161 (85.2%) participants underwent residential treatment (81 participants treated in a SUD treatment ward in a psychiatric facility and 80 participants in a drug-free TC). Additionally, 28 (14.8%) participants were recruited in outpatient services. At the 45-day follow-up assessment, 120 participants completed the questionnaire, resulting in an overall attrition rate of 36.5% (69/189). The proportion of participants who did not complete the follow-up at 45 days was 38.8% (31/80) in drug-free TCs, 34.6% (28/81) in SUD treatment wards within psychiatric facilities, and 35.7% (10/28) in outpatient services.

Details regarding the baseline sociodemographic and clinical characteristics of the participant sample are presented in Table 1. The average age is 35.5 (SD 9.9) years. The majority of participants were male (n=156, 82.5%), completed secondary education as their highest level of education (n=114, 60.3%), and lived alone (n=90, 47.6%). Most participants were born in Belgium (n=178, 94.2%) and resided in psychiatric facilities (n=81, 42.9%) and TCs (n=80, 42.3%) during the baseline assessment. A total of 153 (81%) participants received previous treatment for SUDs. The most frequently reported main substances were alcohol (n=100, 53.8%), cocaine (n=81, 43.5%), and cannabis (n=64, 34.4%), indicating the presence of many problematic poly-substance users.

XSL•FO
RenderX

Table 1. Overview of the sociodemographic and clinical characteristics of the participant sample.

Variables	Total (n=189)	PC[a] (n=81)	TC[b] (n=80)	Outpatient (n=28)	*P* value
Age (years), mean (SD)[c]	35.5 (9.9)	36.6 (11.7)	34.1 (7.9)	36.5 (9.3)	.20[d]
Sex (male), n (%)	156 (82.5)	63 (77.8)	71 (88.8)	22 (78.6)	.16
Education level, n (%)					.04
Primary	45 (23.8)	14 (17.3)	23 (28.7)	8 (28.6)	
Secondary	114 (60.3)	47 (58)	51 (63.7)	16 (57.1)	
Higher	30 (15.9)	20 (24.7)	6 (7.5)	4 (14.3)	
Current living situation, n (%)					.37[e]
Alone	90 (47.6)	36 (44.4)	39 (48.8)	15 (53.6%)	
Alone with children	9 (4.8)	4 (4.9)	5 (6.3)	0 (0)	
Living together with partner and children	18 (9.5)	8 (9.9)	6 (7.5)	4 (14.3)	
Living together with partner without children	13 (6.9)	8 (9.9)	2 (2.5)	3 (10.7)	
Living together with others	59 (31.2)	25 (30.9)	28 (35)	6 (21.4)	
Country of birth, n (%)					
Belgium	178 (94.2)	74 (91.4)	76 (95)	28 (100)	.27[e]
Belgium (father)	156 (82.5)	64 (79)	65 (81.3)	27 (96.4)	.10
Belgium (mother)	161 (85.2)	68 (84)	66 (82.5)	27 (96.4)	.19
Previous treatment for SUDs[f] (yes), n (%)	153 (81)	67 (82.7)	68 (85)	18 (64.3)	.048
Opioid agonist therapy	37 (19.6)	9 (11.1)	15 (18.8)	13 (46.4)	<.001
Main substances (n=186)[g,h], n (%)					
Alcohol	100 (53.8)	52 (64.2)	38 (48.7)	10 (37)	.03
Amphetamines	41 (22)	9 (11.1)	23 (29.5)	9 (33.3)	.006
Benzodiazepines	20 (10.8)	8 (9.9)	12 (15.4)	0 (0)	.08
Cannabis	64 (34.4)	23 (28.4)	31 (39.7)	10 (37)	.31
Crack	33 (17.7)	9 (11.1)	19 (24.4)	5 (18.5)	.09
Codeine + Promethazine	2 (1.1)	2 (2.5)	0 (0)	0 (0)	.63[e]
Cocaine	81 (43.5)	21 (25.9)	50 (64.1)	10 (37)	<.001
GHB[i]	14 (7.5)	1 (1.2)	10 (12.8)	3 (11.1)	.02
Hallucinogens	3 (1.6)	0 (0)	2 (2.6)	1 (3.7)	.21[e]
Ketamine	20 (10.8)	13 (16)	5 (6.4)	2 (7.4)	.12
New psychoactive substances	6 (3.2)	2 (2.5)	3 (3.8)	1 (3.7)	.87[e]
Opioids	35 (18.8)	10 (12.3)	14 (17.9)	11 (40.7)	.005
Number of main substances (n=186)[c,g,h], mean (SD)	2.3 (1.4)	1.9 (1.3)	2.7 (1.4)	2.3 (1.4)	.001
More than one main substance (n=186)[g,h], n (%)	111 (59.7)	33 (40.7)	60 (76.9)	18 (66.7)	<.001

[a]PC: psychiatric treatment center.

[b]TC: therapeutic community.

[c]ANOVA.

[d]Welch test.

[e]Fisher exact test.

[f]SUD: substance use disorder.

[g]Some service users reported more than one main substance used.

[h]Data missing for three participants.

XSL•FO
RenderX

[i]GHB: gamma-hydroxybutyric acid.

Initial comparisons were made between the three treatment modalities (Table 1). When considering sociodemographic and clinical characteristics, no significant differences were found in terms of age, sex, living situation, and country of birth (all *P*>.05). However, significant differences were observed regarding education level, history of SUD treatment, OAT, and the primary substances reported. Post hoc analyses revealed that participants in the PC group had the highest level of education, followed by those in the outpatient group, and finally, the individuals from the TC group. On average, 83% (67/81) of participants from the PC group and 85% (68/80) of the participants from the TC group had a previous history of SUD treatment, with no statistically significant difference between the two groups. Moreover, a significantly higher percentage of participants in the outpatient group (13/28, 46%) were engaged in some form of OAT (*P*<.001). This was followed by participants in the TC group (15/80, 19%), and finally, the PC group (9/81, 11%). In terms of substance use, alcohol was more frequently reported as the primary substance in the PC group, followed by TC and outpatient groups. In contrast, opioids were most frequently reported in the outpatient group, followed by TC and PC groups. Amphetamine (*P*=.006), cocaine (*P*<.001), and gamma-hydroxybutyric acid (*P*=.02) were significantly more reported in the TC group, followed by the outpatient, and finally, the PC group. A significantly higher percentage of participants in the TC group reported more than one primary substance, followed by the outpatient and PC group (*P*<.001).

Further comparisons were made regarding comorbid psychiatric conditions and PROMs at baseline (Table 2). No significant differences were found in levels of PTSD, depression, anxiety, and stress scores between the treatment modalities (all *P*>.05). However, significant differences were observed in scores on the ADHD Self-Report scale, with participants in the outpatient group scoring significantly lower compared to the TC group (*P*=.03). No significant differences were found in general health and quality of life scores (all *P*>.05). In contrast, the total score on the SURE questionnaire revealed significant differences at baseline between the groups, with participants in the outpatient group scoring lower overall than those in the other groups (*P*<.001). This trend was significant in the subdomains of “Substance Use” (*P*<.001), “Self-care” (*P*<.001), and “Outlook on Life” (*P*=.008).

XSL•FO
RenderX

Table 2. Comparison of comorbidity and PROMs[a] at baseline.

Questionnaires	Total (n=189), mean (SD)	PC[b] (n=81), mean (SD)	TC[c] (n=80), mean (SD)	Outpatient (n=28), mean (SD)	*P* value
PC-PTSD-5[d]	2.24 (2.00)	2.09 (1.92)	2.51 (2.07)	1.89 (1.95)	.25
DASS 21[e]					
Depression	18.94 (10.92)	17.58 (10.80)	20.23 (10.80)	19.21 (11.54)	.31
Anxiety	14.03 (9.50)	12.94 (9.12)	15.53 (9.96)	12.93 (8.92)	.18
Stress	19.31 (10.15)	18.00 (10.47)	21.00 (9.72)	18.29 (10.09)	.15
ASRS[f]	3.44 (1.68)	3.36 (1.73)	3.75 (1.50)	2.79 (1.89)	.03
PROMIS GH[g]					
Physical	2.77 (.86)	2.73 (.92)	2.85 (.83)	2.77 (.86)	.55
Mental	2.59 (.92)	2.62 (.94)	2.61 (.88)	2.46 (.96)	.73
SURE[h] total	52.05 (7.70)	52.19 (7.99)	54.03 (5.46)	46.00 (9.31)	<.001[i]
Substance use	14.97 (2.89)	15.07 (2.98)	15.60 (2.42)	12.89 (3.01)	<.001
Self-care	12.02 (2.78)	11.81 (3.00)	12.91 (1.97)	10.07 (3.07)	<.001[i]
Relationships	10.90 (1.61)	10.98 (1.60)	11.13 (1.26)	10.04 (2.20)	.05[i]
Material resources	7.56 (1.76)	7.72 (1.73)	7.48 (1.80)	7.36 (1.77)	.55
Outlook on life	6.59 (1.88)	6.60 (1.92)	6.91 (1.66)	5.64 (2.08)	.008
WHOQoL-BREF[j]					
Perception quality of life	2.94 (.82)	2.93 (.77)	2.95 (.87)	2.96 (.84)	.97
Perception health	2.85 (.96)	2.91 (.91)	2.81 (.98)	2.79 (1.03)	.74
Physical health	13.52 (2.63)	13.33 (2.77)	13.85 (2.41)	13.14 (2.80)	.33
Psychological health	11.43 (2.81)	11.68 (2.67)	11.33 (2.88)	10.98 (3.01)	.48
Social relationships	12.14 (3.70)	12.59 (3.72)	11.65 (3.55)	12.24 (3.98)	.27
Environment	13.24 (2.87)	13.74 (2.84)	12.98 (2.85)	12.55 (2.84)	.09

[a]PROM: patient-reported outcome measure.

[b]PC: psychiatric treatment center.

[c]TC: therapeutic community.

[d]PC-PTSD-5: Primary Care Posttraumatic Stress Disorder Screen.

[e]DASS 21: Depression, Anxiety, and Stress Scale.

[f]ASRS: Adult Attention-Deficit/Hyperactivity Disorder-Self-Report Scale.

[g]PROMIS GH: Patient-Reported Outcomes Measurement Information System Global Health.

[h]SURE: Substance Use Recovery Evaluator.

[i]Welch test.

[j]WHOQoL-BREF: brief version of the World Health Organization Quality of Life questionnaire.

Additionally, comparisons were conducted of baseline sociodemographic, clinical, and PROM scores between participants who completed the 45-day follow-up and those who did not participate in the 45-day assessment (Table 3). Participants who completed the 45-day follow-up were significantly older (mean age 37.1, SD 9.4 years) compared to noncompleters (mean age 32.8, SD 10.3 years; *P*=.005). In general, noncompleters had a lower level of education (*P*=.009). Significant differences were also found regarding living situations, with a higher proportion of completers living alone compared to noncompleters (*P*=.03). The country of birth of participants' mothers differed significantly between groups, as a larger proportion of completers had a mother born in Belgium (109/120, 91%) compared to noncompleters (52/69, 75%; *P*=.004). A similar trend was observed for paternal country of birth, but the difference was not statistically significant (*P*=.44). Regarding substance use characteristics, alcohol was more frequently reported as the primary substance among completers (72/120, 61%) compared to noncompleters (28/69, 42%; *P*=.01). However, no significant differences were observed between groups concerning other primary substances of use or the number of primary substances reported (all *P*>.05).

XSL•FO
RenderX

Table 3. Comparison of sociodemographic and clinical characteristics of 45-day follow-up completers and noncompleters.

Variables	Total (n=189)	45-day completers (n=120)	Noncompleters (n=69)	*P* value
Age (years), mean (SD)[a]	35.5 (9.9)	37.1 (9.4)	32.8 (10.3)	.005
Sex (male), n (%)	156 (82.5)	100 (83.3)	56 (81.2)	.71
Education level, n (%)				.009
Primary	45 (23.8)	20 (16.7)	25 (36.2)	
Secondary	114 (60.3)	78 (65)	36 (52.2)	
Higher	30 (15.9)	22 (18.3)	8 (11.6)	
Current living situation, n (%)				.03[b]
Alone	90 (47.6)	67 (55.8)	23 (33.3)	
Alone with children	9 (4.8)	6 (5)	3 (4.3)	
Living together with partner and children	18 (9.5)	11 (9.2)	7 (10.1)	
Living together with partner without children	13 (6.9)	7 (5.8)	6 (8.7)	
Living together with others	59 (31.2)	29 (24.2)	30 (43.5)	
Country of birth, n (%)				
Belgium	178 (94.2)	113 (94.2)	65 (94.2)	.99[b]
Belgium (father)	156 (82.5)	101 (84.2)	55 (79.7)	.44
Belgium (mother)	161 (85.2)	109 (90.8)	52 (75.4)	.004
Previous treatment for SUDs[c] (yes), n (%)	153 (81)	98 (81.7)	55 (79.7)	.74
Opioid agonist therapy, n (%)	37 (19.6)	19 (15.8)	18 (26.1)	.09
Main substances (n=186)[d,e], n (%)				
Alcohol	100 (53.8)	72 (60.5)	28 (41.8)	.01
Amphetamines	41 (22)	27 (22.7)	14 (20.9)	.78
Benzodiazepines	20 (10.8)	15 (12.6)	5 (7.5)	.28
Cannabis	64 (34.4)	37 (31.1)	27 (40.3)	.21
Crack	33 (17.7)	22 (18.5)	11 (16.4)	.72
Codeine + Promethazine	2 (1.1)	1 (0.8)	1 (1.5)	.99[b]
Cocaine	81 (43.5)	52 (43.7)	29 (43.3)	.96
GHB[f]	14 (7.5)	12 (10.1)	2 (3)	.08
Hallucinogens	3 (1.6)	3 (2.5)	0 (0)	.55[b]
Ketamine	20 (10.8)	9 (7.6)	11 (16.4)	.06
New psychoactive substances	6 (3.2)	2 (1.7)	4 (6)	.19[b]
Opioids	35 (18.8)	19 (16)	16 (23.9)	.19
Number of main substances (n=186)[a,d,e], mean (SD)	2.3 (1.4)	2.3 (1.3)	2.2 (1.5)	.75
More than one main substance (n=186)[d,e], n (%)	111 (59.7)	75 (63)	36 (53.7)	.22
Treatment modality, n (%)				.86
TC[g]	80 (42.3)	49 (40.8)	31 (44.9)	
PC[h]	81 (42.9)	53 (44.2)	28 (40.6)	
Outpatient	28 (14.8)	18 (15)	10 (14.5)	

[a]ANOVA.

[b]Fisher exact test.

[c]SUD: substance use disorder.

XSL•FO
RenderX

[d]Some service users reported more than one main substance used.

[e]Data missing for three participants.

[f]GHB: gamma-hydroxybutyric acid.

[g]TC: therapeutic community.

[h]PC: psychiatric treatment center.

In terms of PROMs, a significant difference was observed in the SURE domain "material resources," where completers had significantly higher scores (mean 7.78, SD 1.62) than noncompleters (mean 7.19, SD 1.94; *P*=.04; Table 4). Additionally, noncompleters scored significantly higher (mean 2.99, SD 0.87) on the physical health domain of the PROMIS Global Health compared to people who completed the 45-day follow-up (mean 2.65, SD 0.84; *P*=.01), indicating better perceived physical health. No significant differences were found between completers and noncompleters across other PROM domains (all *P*>.05).

Table 4. Comparison of comorbidity and PROMs[a] at baseline between 45-day follow-up completers and noncompleters.

Questionnaires	Total (n=189), mean (SD)	45-day completers (n=120), mean (SD)	Noncompleters (n=69), mean (SD)	*P* value
PC-PTSD-5[b]	2.24 (2.00)	2.30 (2.00)	2.13 (2.00)	.58
DASS 21[c]				
Depression	18.94 (10.92)	18.43 (10.77)	19.83 (11.21)	.40
Anxiety	14.03 (9.50)	14.30 (9.66)	13.57 (9.25)	.61
Stress	19.31 (10.15)	18.70 (10.38)	20.38 (9.72)	.28
ASRS[d]	3.44 (1.68)	3.35 (1.74)	3.59 (1.58)	.34
PROMIS GH[e]				
Physical	2.77 (.86)	2.65 (.84)	2.99 (.87)	.01
Mental	2.59 (.92)	2.56 (.87)	2.65 (1.00)	.49
SURE[f] total	52.05 (7.70)	52.73 (7.33)	50.87 (8.22)	.11
Substance use	14.97 (2.89)	15.21 (2.75)	14.57 (3.10)	.14
Self-care	12.02 (2.78)	12.15 (2.71)	11.80 (2.90)	.40
Relationships	10.90 (1.61)	10.99 (1.52)	10.74 (1.75)	.29
Material resources	7.56 (1.76)	7.78 (1.62)	7.19 (1.94)	.04[g]
Outlook on life	6.59 (1.88)	6.60 (1.84)	6.58 (1.94)	.94
WHOQoL-BREF[h]				
Perception quality of life	2.94 (.82)	2.94 (.80)	2.94 (.86)	.99
Perception health	2.85 (.96)	2.83 (.96)	2.88 (.95)	.73
Physical health	13.52 (2.63)	13.51 (2.70)	13.55 (2.52)	.92
Psychological health	11.43 (2.81)	11.26 (2.68)	11.72 (3.01)	.28
Social relationships	12.14 (3.70)	11.83 (3.83)	12.68 (3.40)	.13
Environment	13.24 (2.87)	13.40 (2.87)	12.96 (2.86)	.30

[a]PROM: patient-reported outcome measure.

[b]PC-PTSD-5: Primary Care Posttraumatic Stress Disorder Screen.

[c]DASS 21: Depression, Anxiety, and Stress Scale.

[d]ASRS: Adult Attention-Deficit/Hyperactivity Disorder Self-Report Scale.

[e]PROMIS GH: Patient-Reported Outcomes Measurement Information System Global Health.

[f]SURE: Substance Use Recovery Evaluator.

[g]Welch test.

[h]WHOQoL-BREF: brief version of the World Health Organization Quality of Life questionnaire.

XSL•FO
RenderX

Discussion

Principal Findings

The use of PROMs and PREMs in SUD treatment services is limited to date, and no systematic monitoring system of patient-reported outcomes in Belgium is currently available [34]. Therefore, the OMER-BE study was set up to assess PROMs and PREMs systematically and to improve the quality of SUD services through the routine measurement and monitoring of patient-reported outcomes and experiences at regular times during and after treatment.

Comparisons across the three treatment modalities at baseline revealed several noteworthy differences and similarities. While no significant differences were observed in age, sex, living situation, or country of birth, significant variations in education levels, history of SUD treatment, engagement in OAT, and primary substance of problematic use were observed. Regarding co-occurring mental health disorders, no significant differences were found in PTSD, depression, anxiety, and stress scores between the treatment modalities. However, significant differences were noted in ADHD scores, with the TC group scoring significantly higher than the outpatient group. These differences can be attributed to several factors. TCs are often more suitable for individuals with more severe ADHD symptoms, as these environments provide the structured support and comprehensive care needed to manage both SUDs and comorbid conditions. These settings allow for intensive monitoring and support, which may be crucial for individuals struggling with ADHD symptoms, such as impulsivity [55].

Analysis of PROM scores showed no significant differences between treatment modalities in general health and quality of life. However, significant differences were found in recovery strengths, with participants in the outpatient group scoring lower overall than those in the other groups. These differences, particularly noted in the subdomains of “Substance Use,” “Self-care,” and “Outlook on Life,” suggest that the structure of the treatment environment plays a role in shaping recovery trajectories. For example, we observed that participants in residential treatment score significantly higher on the “Substance Use” subscale compared to those in outpatient treatment, suggesting lower levels of substance use among those in residential settings. This difference can be attributed to the structured and controlled environment of residential treatment settings, where strict measures are often in place to encourage abstinence and limit access to substances [56]. The supervision and supportive community in residential facilities may also play a crucial role in reducing substance use, which is less enforceable in outpatient settings where individuals have more autonomy and access to substances. The significant difference in the “Self-care” subscale, with residential participants scoring higher, may reflect the comprehensive care and support provided in residential settings. These environments typically offer structured daily routines and more intensive psychological and physical care [56]. In contrast, outpatient treatment often places a greater emphasis on self-management [57], which can be challenging for individuals with limited resources or support networks. Participants in residential treatment also scored significantly higher on the “Outlook on Life” subscale, which may be attributed to the supportive environment of residential treatment settings, where participants are provided with continuous care and peer support, all of which contribute to a more positive and optimistic outlook on life. Notably, the group difference in the “Relationship” subscale was close to significance. As participants were assessed in the first weeks of treatment, it is possible that participants in residential facilities did not have enough time to build meaningful relationships with other service users and providers.

In addition to differences between treatment modalities, comparisons between participants who completed the 45-day follow-up and those who did not participate in this assessment revealed several significant differences. Participants who completed the 45-day follow-up were significantly older, had a higher education level, were more likely to live alone, and were more likely to have a mother born in Belgium. Additionally, they scored higher on the “Material Resources” domain of the SURE, which includes questions about stable housing, a steady income, and effective financial management. These findings are in line with studies that suggest that factors such as lower education level, younger age, unemployment, and financial instability are associated with higher attrition in follow-up assessments [58,59]. Furthermore, some studies have found that factors such as low socioeconomic status are associated with limitations in people’s ability to use electronic devices [21,60,61], which may contribute to reduced participation in follow-up assessments. The higher completion rates among individuals living alone could indicate that they experience fewer external obligations that interfere with follow-up participation. The differences observed in maternal country of birth suggest that cultural or socioeconomic factors may also play a role in study retention, such as language barriers or differing levels of trust in research participation. Furthermore, alcohol was more frequently reported as the primary substance of use among completers and individuals with alcohol as primary substance of use had a lower attrition rate in comparison to the total sample (28% vs 36.6%). The higher prevalence of alcohol as a primary substance among completers may reflect differences in treatment pathways and more stable living situations that can be linked to socioeconomic factors and higher education levels.

Limitations

Despite being innovative and pioneering, the proposed study also has some limitations. First, recruiting participants in outpatient facilities proved to be extremely challenging, leading to a limited number of participants in these settings. One significant obstacle was the presence of long waiting lists, resulting in a limited number of new treatment episodes and lower turnover rates compared to the residential services. The frequency of contact between potential participants and service providers was limited in many outpatient settings, as many service users have only weekly or biweekly appointments. This poses significant scheduling constraints, especially since participants were expected to complete the baseline questionnaires within three weeks of treatment initiation. Additionally, there was a high occurrence of no-shows for scheduled appointments with the researchers, which resulted in

XSL•FO
RenderX

missed opportunities to engage potential participants. To address these challenges, several strategies were implemented. Flyers and posters were provided in waiting areas to inform potential participants about the study and proactive follow-up calls were made to confirm appointments. Moreover, regular contact was maintained with service providers (both through email and phone) to encourage participation. Despite these efforts, maintaining consistent engagement in outpatient facilities has been proven particularly challenging, resulting in some outpatient centers withdrawing from the study. Consequently, future comparative analyses will primarily focus on residential treatment services due to the limited number of outpatient participants. These challenges raise important considerations for the broader implementation of PROMs and PREMs in outpatient addiction services. While comprehensive assessments seem feasible in residential settings, these findings underscore the need for effective strategies for the structural implementation of PROMs and PREMs in outpatient services. Potential strategies to increase participation and retention may include broadening recruitment periods, simplifying assessment batteries, and using short-form PROMs and PREMs that could be completed during brief waiting periods before appointments. Additionally, codeveloping implementation strategies with service users and providers could further enhance feasibility and engagement in outpatient settings.

Second, recruiting facilities and participants in the French-speaking regions of Belgium proved to be more challenging than in the Dutch-speaking parts. This might be caused by differing attitudes toward the importance of outcome evaluation and measurement. Hence, beyond providing feasible measurement instruments and implementation strategies, there is a need to build service cultures that value the implementation of measurements to improve the quality of support. Organizing policies and funding that support the routine use of PROMs and PREMs might be an important step toward realizing this. Yet, this also includes overcoming practical challenges such as staffing shortages that hinder broader participation.

Finally, the study is limited by the lack of standardized diagnostic assessment of participants entering the study. Inclusion was based on proxy indicators (previous treatment for SUD, clinical assessment, or history of substance misuse), rather than on clinical diagnoses. This lack of structural diagnosis may have introduced variability in our findings and may limit generalization to other populations.

Conclusions

Despite these challenges, this study is one of the first studies to explore the assessment of PROMs and PREMs in SUD treatment services using the ICHOM SSA set of instruments. By systematically and routinely monitoring PROMs and PREMs, the project aims to empower service providers and give them tools to evaluate subjective treatment outcomes and experiences. This approach can provide service providers and policy makers with benchmarks for assessing outcomes and experiences during and after treatment across various treatment modalities. By applying an internationally validated tool, the study allows international comparison with similar interventions and treatment modalities globally. Future analyses will explore longitudinal changes in PROMs and PREMs, with a particular focus on the influence of baseline mediating and moderating variables (eg, socioeconomic status, comorbidity, and recovery strength) and differences between treatment modalities. The findings from this study will offer new perspectives and insights into the effectiveness of different SUD treatment modalities and their impact on diverse service user populations. This shift toward person-centered care not only supports better recovery outcomes but also fosters continuous improvement and innovation in SUD treatment services. Ultimately, systematically monitoring PROMs and PREMs has the potential to enhance the quality of care by offering a comprehensive understanding of service users' needs and treatment effectiveness, extending beyond abstinence.

Acknowledgments

The research was funded by a grant from the Belgian Science Policy Office (BELSPO), research project DR/90. BELSPO has no role in the design, data collection, data analysis, and reporting of this study.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

WB discloses relationships with Clearmind, Camurus AB, and Takeda Pharmaceutical Company Limited, involving consultancy or advisory roles.

References

1. Deak JD, Johnson EC. Genetics of substance use disorders: a review. Psychol Med 2021;51(13):2189-2200. [doi: 10.1017/s0033291721000969]
2. Dennis M, Scott CK. Managing addiction as a chronic condition. Addict Sci Clin Pract 2007;4(1):45-55 [FREE Full text] [doi: 10.1151/ascp074145] [Medline: 18292710]

XSL•FO
RenderX

3. Kelly JF, Greene MC, Bergman BG, White WL, Hoeppner BB. How many recovery attempts does it take to successfully resolve an alcohol or drug problem? Estimates and correlates from a national study of recovering U.S. adults. Alcohol Clin Exp Res 2019;43(7):1533-1544 [FREE Full text] [doi: 10.1111/acer.14067] [Medline: 31090945]
4. McKay JR, Hiller-Sturmhofel S. Treating alcoholism as a chronic disease: approaches to long-term continuing care. Alcohol Res Health 2011;33(4):356-370 [FREE Full text] [Medline: 23580020]
5. Castelpietra G, Knudsen AKS, Agardh EE, Armocida B, Beghi M, Iburg KM, et al. The burden of mental disorders, substance use disorders and self-harm among young people in Europe, 1990-2019: findings from the global burden of disease study 2019. Lancet Reg Health Eur 2022;16:100341 [FREE Full text] [doi: 10.1016/j.lanepe.2022.100341] [Medline: 35392452]
6. Degenhardt L, Charlson F, Ferrari A, Santomauro D, Erskine H, Mantilla-Herrara A, et al. The global burden of disease attributable to alcohol and drug use in 195 countries and territories, 1990–2016: a systematic analysis for the global burden of disease study 2016. Lancet Psychiatry 2018;5(12):987-1012. [doi: 10.1016/s2215-0366(18)30337-7]
7. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the global burden of disease study 2010. Lancet 2013;382(9904):1575-1586. [doi: 10.1016/s0140-6736(13)61611-6]
8. Alcohol. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/alcohol [accessed 2022-05-09]
9. Cox G, Comiskey C. Characteristics of opiate users presenting for a new treatment episode: baseline data from the national drug treatment outcome study in Ireland (ROSIE). Drugs Educ Prev Policy 2009;14(3):217-230. [doi: 10.1080/09687630601032977]
10. Fletcher BW, Tims FM, Brown BS. Drug abuse treatment outcome study (DATOS): treatment evaluation research in the United States. Psychol Addict Behav 1997;11(4):216-229. [doi: 10.1037//0893-164x.11.4.216]
11. McKeganey N, Bloor M, McIntosh J, Neale J. Key findings from the Drug Outcome Research in Scotland (DORIS) study. ResearchGate. 2008. URL: https://www.researchgate.net/profile/Neil-Mckeganey/publication/255672321_Key_findings_from_the_Drug_Outcome_Research_in_Scotland_DORIS_study/links/0c96053b46da8b4f42000000/Key-findings-from-the-Drug-Outcome-Research-in-Scotland-DORIS-study.pdf [accessed 2025-02-13]
12. Teesson M, Mills K, Ross J, Darke S, Williamson A, Havard A. The impact of treatment on 3 years' outcome for heroin dependence: findings from the Australian treatment outcome study (ATOS). Addiction 2008;103(1):80-88. [doi: 10.1111/j.1360-0443.2007.02029.x] [Medline: 17995994]
13. Stahler GJ, Mennis J, DuCette JP. Residential and outpatient treatment completion for substance use disorders in the U.S.: moderation analysis by demographics and drug of choice. Addict Behav 2016;58:129-135. [doi: 10.1016/j.addbeh.2016.02.030] [Medline: 26925821]
14. Brorson HH, Arnevik EAA, Rand-Hendriksen K, Duckert F. Drop-out from addiction treatment: a systematic review of risk factors. Clin Psychol Rev 2013;33(8):1010-1024. [doi: 10.1016/j.cpr.2013.07.007] [Medline: 24029221]
15. Katz EC, Brown BS, Schwartz RP, Weintraub E, Barksdale W, Robinson R. Role induction: a method for enhancing early retention in outpatient drug-free treatment. J Consult Clin Psychol 2004;72(2):227-234. [doi: 10.1037/0022-006x.72.2.227]
16. Lappan SN, Brown AW, Hendricks PS. Dropout rates of in-person psychosocial substance use disorder treatments: a systematic review and meta-analysis. Addiction 2020;115(2):201-217. [doi: 10.1111/add.14793] [Medline: 31454123]
17. Simpson DD, Joe GW. A longitudinal evaluation of treatment engagement and recovery stages. J Subst Abuse Treat 2004;27(2):89-97. [doi: 10.1016/j.jsat.2004.03.001] [Medline: 15450643]
18. Arria AM, TOPPS-II Interstate Cooperative Study Group. Drug treatment completion and post-discharge employment in the TOPPS-II interstate cooperative study. J Subst Abuse Treat 2003;25(1):9-18. [doi: 10.1016/s0740-5472(03)00050-3] [Medline: 14512103]
19. Churruca K, Pomare C, Ellis LA, Long JC, Henderson SB, Murphy LED, et al. Patient-reported outcome measures (PROMs): a review of generic and condition-specific measures and a discussion of trends and issues. Health Expect 2021;24(4):1015-1024 [FREE Full text] [doi: 10.1111/hex.13254] [Medline: 33949755]
20. Davis EL, Kelly PJ, Deane FP, Baker AL, Buckingham M, Degan T, et al. The relationship between patient-centered care and outcomes in specialist drug and alcohol treatment: a systematic literature review. Subst Abus 2020;41(2):216-231. [doi: 10.1080/08897077.2019.1671940] [Medline: 31638870]
21. Migchels C, Zerrouk A, Crunelle CL, Matthys F, Gremeaux L, Fernandez K, et al. Patient reported outcome and experience measures (PROMs and PREMs) in substance use disorder treatment services: a scoping review. Drug Alcohol Depend 2023;253:111017. [doi: 10.1016/j.drugalcdep.2023.111017] [Medline: 37995391]
22. Trujols J, Portella MJ, Iraurgi I, Campins MJ, Siñol N, de Los Cobos JP. Patient-reported outcome measures: are they patient-generated, patient-centred or patient-valued? J Ment Health 2013;22(6):555-562. [doi: 10.3109/09638237.2012.734653] [Medline: 23323928]
23. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. BJA Educ 2017 Apr;17(4):137-144. [doi: 10.1093/bjaed/mkw060]

XSL•FO
RenderX

24. Black N, Chung S, Tisdale C, Fialho LS, Aramrattana A, Assanangkornchai S, et al. An international, multidisciplinary consensus set of patient-centered outcome measures for substance-related and addictive disorders. J Clin Med 2024;13(7):2154 [FREE Full text] [doi: 10.3390/jcm13072154] [Medline: 38610919]
25. Alves P, Sales C, Ashworth M. Does outcome measurement of treatment for substance use disorder reflect the personal concerns of patients? A scoping review of measures recommended in Europe. Drug Alcohol Depend 2017;179:299-308. [doi: 10.1016/j.drugalcdep.2017.05.049] [Medline: 28830035]
26. Friedrichs A, Spies M, Härter M, Buchholz A. Patient preferences and shared decision making in the treatment of substance use disorders: a systematic review of the lliterature. PLoS One 2016;11(1):e0145817 [FREE Full text] [doi: 10.1371/journal.pone.0145817] [Medline: 26731679]
27. Garnick DW, Horgan CM, Acevedo A, McCorry F, Weisner C. Performance measures for substance use disorders—what research is needed? Addict Sci Clin Pract 2012;7(1):18 [FREE Full text] [doi: 10.1186/1940-0640-7-18] [Medline: 23186374]
28. Kendrick T, El-Gohary M, Stuart B, Gilbody S, Churchill R, Aiken L, et al. Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. Cochrane Database Syst Rev 2016;7(7):CD011119 [FREE Full text] [doi: 10.1002/14651858.CD011119.pub2] [Medline: 27409972]
29. Kolind T, Hesse M. Patient-centred care-perhaps the future of substance abuse treatment. Addiction 2017;112(3):465-466. [doi: 10.1111/add.13673] [Medline: 28168786]
30. Roe D, Mazor Y, Gelkopf M. Patient-reported outcome measurements (PROMs) and provider assessment in mental health: a systematic review of the context of implementation. Int J Qual Health Care 2021;34(Suppl 1):ii28–ii39. [doi: 10.1093/intqhc/mzz084] [Medline: 31725153]
31. Vanderplasschen W, Colpaert K, Autrique M, Rapp RC, Pearce S, Broekaert E, et al. Therapeutic communities for addictions: a review of their effectiveness from a recovery-oriented perspective. Sci World J 2013;2013(1):427817 [FREE Full text] [doi: 10.1155/2013/427817] [Medline: 23401669]
32. de Leon G. Community as Method: Therapeutic Communities for Special Populations and Special Settings. Westport, CT: Greenwood Publishing Group, Inc; 1997.
33. Vanderplasschen W, Vandevelde S, Broekaert E. Therapeutic Communities for Treating Addictions in Europe: Evidence, Current Practices and Future Challenges. Luxembourg: Publications Office of the European Union; 2014.
34. Zerrouk A. Outcome measurement and evaluation as a routine practice in alcohol and other drug services in Belgium (OMER-BE). Lisbon Addictions. URL: https://www.lisbonaddictions.eu/lisbon-addictions-2022/presentations/outcome-measurement-and-evaluation-routine-practice-alcohol-and-other-drug-services [accessed 2022-11-25]
35. Tsang S, Royse C, Terkawi A. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. Saudi J Anaesth 2017;11(5):80. [doi: 10.4103/sja.sja_203_17]
36. Prins A, Bovin MJ, Smolenski DJ, Marx BP, Kimerling R, Jenkins-Guarnieri MA, et al. The primary care PTSD screen for DSM-5 (PC-PTSD-5): development and evaluation within a veteran primary care sample. J Gen Intern Med 2016;31(10):1206-1211 [FREE Full text] [doi: 10.1007/s11606-016-3703-5] [Medline: 27170304]
37. Patton SC, Hinojosa CA, Lathan EC, Welsh JW, Powers A. Validating the primary care posttraumatic stress disorder screen for DSM-5 (PC-PTSD-5) in a substance misusing, trauma-exposed, socioeconomically vulnerable population. Addict Behav 2023;139:107592 [FREE Full text] [doi: 10.1016/j.addbeh.2022.107592] [Medline: 36584543]
38. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. Sydney, Australia: Psychology Foundation of Australia; 1995.
39. Beaufort IN, de Weert-van Oene GH, Buwalda VA, de Leeuw JRJ, Goudriaan AE. The Depression, Anxiety and Stress Scale (DASS-21) as a screener for depression in substance use disorder inpatients: a pilot study. Eur Addict Res 2017;23(5):260-268 [FREE Full text] [doi: 10.1159/000485182] [Medline: 29224000]
40. Adler LA, Kessler RC, Spencer T. Adult ADHD Self-Report Scale-v1. 1 (ASRS-v1. 1) Symptom Checklist. New York, NY: World Health Organization; 2003.
41. Blanco CD, Ramos-Quiroga JA, Valero S, Bosch R, Roncero C, Gonzalvo B, et al. Adult ADHD self-report scale (ASRS-v1.1) symptom checklist in patients with substance use disorders. Actas Esp Psiquiatr 2009;37(6):299-305 [FREE Full text] [Medline: 20066581]
42. van de Glind G, van den Brink W, Koeter MW, Carpentier PJ, van Emmerik-van Oortmerssen K, Kaye S, et al. Validity of the adult ADHD self-report scale (ASRS) as a screener for adult ADHD in treatment seeking substance use disorder patients. Drug Alcohol Depend 2013;132(3):587-596 [FREE Full text] [doi: 10.1016/j.drugalcdep.2013.04.010] [Medline: 23660242]
43. Pilkonis PA, Yu L, Colditz J, Dodds N, Johnston KL, Maihoefer C, et al. Item banks for alcohol use from the patient-reported outcomes measurement information system (PROMIS): use, consequences, and expectancies. Drug Alcohol Depend 2013;130(1-3):167-177 [FREE Full text] [doi: 10.1016/j.drugalcdep.2012.11.002] [Medline: 23206377]
44. Pilkonis PA, Yu L, Dodds NE, Johnston KL, Lawrence SM, Hilton TF, et al. Item banks for substance use from the patient-reported outcomes measurement information system (PROMIS(®)): severity of use and positive appeal of use. Drug Alcohol Depend 2015;156:184-192 [FREE Full text] [doi: 10.1016/j.drugalcdep.2015.09.008] [Medline: 26423364]
45. Burling AS, Burling T. A comparison of self-report measures of nicotine dependence among male drug/alcohol-dependent cigarette smokers. Nicotine Tob Res 2003;5(5):625-633. [doi: 10.1080/1462220031000158708] [Medline: 14577979]

XSL•FO
RenderX

46. Marsden J, Farrell M, Bradbury C, Dale-Perera A, Eastwood B, Roxburgh M, et al. Development of the treatment outcomes profile. Addiction 2008;103(9):1450-1460. [doi: 10.1111/j.1360-0443.2008.02284.x] [Medline: 18783500]
47. Pellicciari L, Chiarotto A, Giusti E, Crins MHP, Roorda LD, Terwee CB. Psychometric properties of the patient-reported outcomes measurement information system scale v1.2: global health (PROMIS-GH) in a Dutch general population. Health Qual Life Outcomes 2021;19(1):226 [FREE Full text] [doi: 10.1186/s12955-021-01855-0] [Medline: 34579721]
48. Neale J, Vitoratou S, Finch E, Lennon P, Mitcheson L, Panebianco D, et al. Development and validation of 'SURE': a patient reported outcome measure (PROM) for recovery from drug and alcohol dependence. Drug Alcohol Depend 2016;165:159-167 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.06.006] [Medline: 27344196]
49. Migchels C, Brink WVD, Zerrouk A, Matthys FI, de Ruysscher C, Debeer D, et al. Psychometric evaluation of the dutch version of the substance use recovery evaluator (SURE-NL). Eur Addict Res 2024:1-10 [FREE Full text] [doi: 10.1159/000541584] [Medline: 39522511]
50. Skevington SM, Lotfy M, O'Connell KA. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. a report from the WHOQOL Group. Qual Life Res 2004;13(2):299-310. [doi: 10.1023/b:qure.0000018486.91360.00]
51. WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psychol Med 1998;28(3):551-558. [doi: 10.1017/s0033291798006667] [Medline: 9626712]
52. Hinsley K, Kelly PJ, Davis E. Experiences of patient-centred care in alcohol and other drug treatment settings: a qualitative study to inform design of a patient-reported experience measure. Drug Alcohol Rev 2019;38(6):664-673. [doi: 10.1111/dar.12965] [Medline: 31281991]
53. Kelly PJ, Hatton EL, Hinsley K, Davis E, Larance B. Preliminary psychometric evaluation of the patient reported experience measure for addiction treatment (PREMAT). Addict Behav 2021;123:107048. [doi: 10.1016/j.addbeh.2021.107048] [Medline: 34348223]
54. Migchels C, van den Brink W, Zerrouk A, Matthys F, de Ruysscher C, Vanderplasschen W, et al. Psychometric evaluation of the dutch version of the patient-reported experience measure for addiction treatment (PREMAT-NL). Drug Alcohol Rev 2024;43(7):2021-2034. [doi: 10.1111/dar.13939] [Medline: 39256987]
55. Coghill DR. Organisation of services for managing ADHD. Epidemiol Psychiatr Sci 2016;26(5):453-458. [doi: 10.1017/s2045796016000937]
56. de Andrade D, Elphinston RA, Quinn C, Allan J, Hides L. The effectiveness of residential treatment services for individuals with substance use disorders: a systematic review. Drug Alcohol Depend 2019;201:227-235. [doi: 10.1016/j.drugalcdep.2019.03.031] [Medline: 31254749]
57. Mohamed SM, Abdelmonem RM, Hassan SS. Relationship between self-efficacy, social support and treatment motivation among addict patients. Minia Sci Nurs J 2022;012(1):106-113. [doi: 10.21608/msnj.2022.169551.1039]
58. de Graaf R, van Dorsselaer S, Tuithof M, ten Have M. Sociodemographic and psychiatric predictors of attrition in a prospective psychiatric epidemiological study among the general population. Result of the Netherlands mental health survey and incidence study-2. Compr Psychiatry 2013;54(8):1131-1139. [doi: 10.1016/j.comppsych.2013.05.012] [Medline: 23810078]
59. Katsuno N, Li PZ, Bourbeau J, Aaron S, Maltais F, Hernandez P, et al. Factors associated with attrition in a longitudinal cohort of older adults in the community. Chronic Obstr Pulm Dis 2023;10(2):178-189. [doi: 10.15326/jcopdf.2022.0380] [Medline: 37099700]
60. Myers B, van der Westhuizen C, Pool M, Hornsby N, Sorsdahl KR. Responding to COVID-19 threats to trial conduct: lessons learned from a feasibility trial of a psychological intervention for South African adolescents. Trials 2021;22(1):440 [FREE Full text] [doi: 10.1186/s13063-021-05400-8] [Medline: 34243806]
61. Zhai Y. A call for addressing barriers to telemedicine: health disparities during the COVID-19 pandemic. Psychother Psychosom 2021;90(1):64-66 [FREE Full text] [doi: 10.1159/000509000] [Medline: 32498070]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ICHOM: International Consortium for Health Outcomes Measurement
OAT: opioid agonist therapy
OMER-BE: Outcome Measurement and Evaluation as a Routine Practice in Alcohol and Other Drug Services in Belgium
PC: psychiatric treatment center
PREM: patient-reported experience measure
PREMAT: Patient-Reported Experience Measure for Addiction Treatment
PROM: patient-reported outcome measure
PROMIS: Patient-Reported Outcomes Measurement Information System
PTSD: posttraumatic stress disorder
SSA: Standard Set for Addictions

XSL•FO
RenderX

SUD: substance use disorder
SURE: Substance Use Recovery Evaluator
TC: therapeutic community
WHOQoL-BREF: brief version of the World Health Organization Quality of Life questionnaire

Edited by A Mavragani; submitted 22.08.24; peer-reviewed by M Farrell, M Ferri; comments to author 10.11.24; revised version received 05.02.25; accepted 06.02.25; published 30.04.25.

Please cite as:
Zerrouk A, Migchels C, De Ruysscher C, Fernandez K, Antoine J, De Meyer F, Matthys F, van den Brink W, Crunelle CL, Vanderplasschen W
Incorporating Patient-Reported Outcome Measures and Patient-Reported Experience Measures in Addiction Treatment Services in Belgium: Naturalistic, Longitudinal, Multicenter Cohort Study
JMIR Form Res 2025;9:e65686
URL: https://formative.jmir.org/2025/1/e65686
doi:10.2196/65686
PMID:

©Amine Zerrouk, Charlotte Migchels, Clara De Ruysscher, Kim Fernandez, Jerome Antoine, Florian De Meyer, Frieda Matthys, Wim van den Brink, Cleo Lina Crunelle, Wouter Vanderplasschen. Originally published in JMIR Formative Research (https://formative.jmir.org), 30.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Portal-Based Intervention (PATTERN) Designed to Support Medication Use Among Older Adults: Feasibility and Acceptability Study

Allison Pack, MPH, PhD; Stacy C Bailey, MPH, PhD; Rachel O'Conor, MPH, PhD; Evelyn Velazquez, MPH; Guisselle Wismer, MPH; Fangyu Yeh, MPH; Laura M Curtis, MS; Kenya Alcantara, BA; Michael S Wolf, MA, MPH, PhD

Division of General Internal Medicine, Feinberg School of Medicine, Northwestern University, 750 North Lake Shore Drive, 10th Floor, Chicago, IL, United States

Corresponding Author:
Allison Pack, MPH, PhD
Division of General Internal Medicine, Feinberg School of Medicine, Northwestern University, 750 North Lake Shore Drive, 10th Floor, Chicago, IL, United States

Abstract

Background: Poor medication adherence among older adults with multiple chronic conditions and polypharmacy is a public health concern stemming from distinct challenges. Prior interventions have largely used a one-size-fits-all approach or resource-intensive approaches inappropriate for busy primary care clinics.

Objective: To address this, Phenotyping Adherence Through Technology-Enabled Reports and Navigation (PATTERN) was adapted from prior work. PATTERN is a portal-based intervention for monitoring self-reported medication adherence challenges among older adults in primary care. This study sought to implement and evaluate PATTERN's feasibility and acceptability.

Methods: We conducted a patient randomized study with a posttest design. Primary care physicians at the participating health center were informed of the study, and approval was obtained to contact their patients. Patient eligibility included being aged 60 years or older, having prescription medications for ≥8 chronic conditions, and an upcoming visit with a physician who had provided approval. Potentially eligible patients were identified using an electronic health record query, and a research coordinator phoned them to confirm eligibility, assess interest, obtain consent, and conduct enrollment. Randomization occurred following enrollment. Those randomized to PATTERN received a medication adherence assessment in their patient portal accounts several days ahead of their visit. The assessment identified whether a patient was experiencing a medication adherence challenge, and if so, the type (cognitive, psychological, medical, regimen-related, social, or economic). Identified challenges were sent to the patient's primary care physician. Assessment delivery several days ahead of a visit was thought to offer sufficient time for patients to complete it and clinicians to review any challenges. Approximately 2 weeks after visits, the coordinator recontacted participants to conduct posttest interviews. This ensured clinicians had sufficient time to respond to challenges during or after visits. Posttest interviews measured the self-reported use of the portal, demographic and health characteristics, and for those randomized to PATTERN, intervention satisfaction. Self-reported data were captured in REDCap and analyzed descriptively. Electronic health record data were also analyzed descriptively to objectively identify feasibility, that is, whether intervention arm participants completed the PATTERN assessment.

Results: We enrolled 64 participants (32 received usual care, and 32 received intervention). Most were female (66%, 42/64), not Hispanic or Latino (94%, 60/64), and identified as White (58%, 37/64). The average (SD) age was 75 (6.8) years. Most participants (80%) self-reported using the patient portal ≥12 times per year. However, electronic health record data revealed that less than half of all participants randomized to PATTERN (47%, 15/32) completed the medication adherence assessment. Of those who remembered completing it, 60% (3/5) were very satisfied with the experience and 20% (1/5) were a little satisfied.

Conclusions: PATTERN has the potential for use with older primary care patients experiencing multiple chronic conditions and polypharmacy. Yet, further adaptation is needed to ensure recipients access their patient portal accounts and complete assessments.

Trial Registration: ClinicalTrials.Gov NCT05766423; https://clinicaltrials.gov/study/NCT05766423

(JMIR Form Res 2025;9:e71676) doi:10.2196/71676

KEYWORDS

older adults; multiple chronic conditions; polypharmacy; primary care; medication adherence; patient portal

XSL•FO
RenderX

Introduction

Older adults with multiple chronic conditions and polypharmacy are at increased risk of suboptimal medication adherence [1,2], resulting in poor treatment outcomes, adverse drug events, and reduced quality of life [3]. In fact, poor medication adherence in this population is considered a substantial public health concern, impacting both morbidity and financial strain [4].

Barriers to medication adherence, or taking medication as prescribed, are well known in older adults with multiple chronic conditions and polypharmacy and comprise distinct challenges that can be mapped to the model of medication self-management; these include challenges classified as cognitive, psychological, medical, regimen-related, social, or economic [5]. Yet, the vast number of interventions designed to address medication adherence challenges do so using a "one-size-fits-all approach," as opposed to a more tailored approach, targeting the distinct challenges experienced by individual patients [6]. Furthermore, interventions that do tailor approaches are often resource-intensive, requiring clinical pharmacists, community health workers, or other individuals tasked with providing intensive one-on-one sessions [6-8]. These resource-intensive interventions and activities are not always available or appropriate for busy primary care settings. Yet, primary care remains the front line of health care for most older adults in the United States [9].

To address this, we created PATTERN: Phenotyping Adherence Through Technology-Enabled Reports and Navigation. Adapted from a previous study by our team [10,11], PATTERN uses the patient portal to identify and classify medication adherence challenges for review and resolution by a primary care physician or designated care team. It was designed to be low-cost, sustainable, and scalable using existing electronic health record (EHR) technology. We conducted a small feasibility study of this intervention to determine its feasibility and acceptability among older primary care patients with multiple chronic conditions and polypharmacy.

Methods

Study Design

We conducted a patient randomized feasibility study, using a posttest design to examine the effects of PATTERN. This paper largely presents data from the intervention arm pertaining to the feasibility and acceptability of PATTERN.

Eligibility Criteria, Recruitment, and Setting

We first sought approval, via email, from physicians affiliated with any primary care clinic (including geriatrics) within the participating academic health center in Chicago, IL, to contact their patients for recruitment into PATTERN. Patients were eligible to participate if they were aged 65 years or older, spoke English as their primary language, self-reported contending with at least two chronic health conditions, were prescribed eight or more medications they were primarily responsible for taking, and had a visit scheduled within the next 1 week with the physician who had provided contact approval. For intervention purposes, individuals also needed to have access to the internet, an active email address, and a patient portal account. Exclusion criteria included having participated in our prior, formative research or having any severe, uncorrectable visual, hearing or cognitive impairments that would have precluded study consent or participation.

Potentially eligible patient participants were identified using an EHR query, phoned by a trained research coordinator, informed about the study, and screened for eligibility. Those who met all eligibility criteria and expressed an interest in participating then provided electronic informed consent prior to enrollment. The research coordinator scheduled remote posttest interviews with enrolled participants; these occurred approximately 2 weeks after their scheduled clinic visit.

Randomization

Participants were randomized one-to-one to usual care or the PATTERN intervention using random block sizes of two and four, stratified by clinic type (geriatrics or other primary care clinic). The allocation table was created using SAS PROC PLAN and implemented using REDCap software [12]. Randomization took place immediately following enrollment (Multimedia Appendix 1). The purpose was to be able to compare differences between those randomized to PATTERN and those in usual care.

Intervention Activities

Participants who were randomized to the PATTERN intervention were sent an email 2 to 3 days ahead of their clinic visit. The email instructed them to log into their patient portal accounts to complete a brief medication adherence assessment prior to their upcoming primary care visit. Those who logged in were instructed to begin the two-tiered assessment adapted from a prior study [10]. The first three questions identified if the participant was experiencing any challenges taking their medications as prescribed. Specifically, these asked: (1) During the past month, did you decide to stop taking any of your medicines without telling your doctor? (2) During the past month, did you decide to take more or less of any of your medicines than you were supposed to without telling your doctor? (3) Are you having any problems with your medicines? Those who endorsed at least one adherence challenge were asked to respond to a subsequent set of 18 questions. These questions, informed by the model of medication self-management, were designed to "phenotype" or identify the distinct type of medication adherence challenge a participant might have been experiencing. Challenges could be classified as cognitive (eg, forgetfulness), psychological (eg, health literacy, depression, or motivation), medical (eg, acute illness), regimen-related (eg, side effects or complex dosing schedules), social (eg, transportation or support), and economic (eg, cost) [5].

Any identified and "phenotyped" challenges were designed to be automatically and securely forwarded, via the EHR, to the participant's primary care physician with whom they had an upcoming clinic visit. The intent was to ensure physicians received relevant and actionable data from patients regarding any medication challenges they might have been experiencing. This information could subsequently inform the scheduled visit

XSL•FO
RenderX

or referral to available support services such as social workers or clinical pharmacists, depending on clinic-specific resources.

Data Collection Activities

Data were collected between March and August 2024. Remote, structured study interviews were held with individual participants approximately 2 weeks after each participant's regularly scheduled clinic visit. The research coordinator phoned participants at previously agreed-upon times and administered a brief questionnaire. Data were captured using the REDCap software [12]. On average, it took participants 20 minutes to complete the questionnaire, and all participants were compensated US $50 for their time.

Measures

Participant Characteristics

All participants reported their age, sex, race, and ethnicity. Perceived health status was measured using the single-item patient-reported measurement information system (PROMIS) global health measure that asks participants to rate their health from poor to excellent using a 5-point Likert scale [13]. Depression was captured using the validated 6-item PROMIS depression measure, with higher scores representing greater symptom burden [13]. Patient activation was captured using the validated 10-item Consumer Health Activation Index (CHAI) [14]. Finally, the use of the patient portal was examined by asking participants how often they use their portal accounts and for what purposes. This self-reported information was supplemented by EHR data identifying the completion of any portal questionnaire in the past year.

Intervention Feasibility

We used data captured from self-reporting and from the EHR to examine intervention feasibility among participants randomized to PATTERN. Specifically, we were interested in identifying (1) the number of adherence assessments received (opened) among those sent; (2) the number of adherence assessments completed among those received; and (3) the number of participants who spoke with their clinical care teams about the adherence assessments in those who completed them.

Intervention Acceptability

Participants who completed the intervention were asked how satisfied they were with the PATTERN questionnaire. Response options ranged on a 4-point Likert scale from "not at all satisfied" to "very satisfied."

Analysis

Descriptive statistics were used to examine participant characteristics, overall and by study arm. All feasibility and acceptability data were also assessed using descriptive statistics in the intervention arm only. Analyses were conducted using SAS 9.4 software.

Ethical Considerations

All study procedures were approved by the Northwestern University's Institutional Review Board (STU00217555). All participants provided electronic informed consent prior to study participation and were compensated US $50 for their time. Data presented herein have all been deidentified.

Results

Participant Characteristics

Participant characteristics are presented in Table 1. In brief, a total of 64 patient participants were enrolled. The average (SD) age was 75 (6.8) years. Most indicated they were female (66%, 42/64), non-Hispanic (94%, 60/64), White (58%, 37/64), and in good or better health (64%, 41/64). A total of 16% (10/64) were classified as having mild or moderate depression, and 38% (24/64) as having low patient activation. A large majority of participants (80%, 51/54) reported they used the patient portal more than 12 times a year, and that when they access the portal, they view laboratory results (98%, 63/64), confirm appointments (89%, 57/64), and send messages to their care teams (91%, 58/64). Actual use, assessed via EHR data, revealed that most patients (72%, 46/64) had previously (within the past year) completed any type of assessment in the patient portal. No significant differences in participant characteristics were observed between study arms.

XSL•FO
RenderX

Table . Descriptive statistics of participant characteristics and technology use, overall and by study arm, captured via self-reporting and the electronic health records between March and August 2024.

Characteristics	Overall (n=64)	Usual care (n=32)	Intervention (n=32)
Gender, n (%)			
Male	22 (34.4)	12 (37.5)	10 (31.3)
Female	42 (65.6)	20 (62.5)	22 (68.8)
Age, years			
Mean (SD)	75.0 (6.8)	75.1 (7.5)	74.9 (6.1)
Median (range)	74.5 (66-93)	74 (66 - 93)	74.5 (66-90)
Ethnicity, n (%)			
Not Hispanic or Latino	60 (93.8)	29 (90.6)	31 (96.9)
Hispanic or Latino	1 (1.6)	1 (3.1)	N/A[a]
Refused	3 (4.7)	2 (6.3)	1 (3.1)
Race, n (%)			
White	37 (57.8)	19 (59.4)	18 (56.3)
Black or African-American	23 (35.9)	11 (34.4)	12 (37.5)
Other	3 (4.7)	2 (6.3)	1 (3.1)
Refused	1 (1.6)	N/A	1 (3.1)
Health status, n (%)			
Excellent	1 (1.6)	1 (3.1)	N/A
Very good	11 (17.2)	3 (9.4)	8 (25.0)
Good	29 (45.3)	17 (53.1)	12 (37.5)
Fair	20 (31.3)	10 (31.3)	10 (31.3)
Poor	3 (4.7)	1 (3.1)	2 (6.3)
PROMIS[b] depression, n (%)			
Within normal limits	54 (84.4)	29 (90.6)	25 (78.1)
Mild or moderate	10 (15.6)	3 (9.4)	7 (21.9)
CHAI,[c] n (%)			
Low	24 (38.1)	9 (28.1)	15 (48.4)
Moderate	27 (42.9)	16 (50.0)	11 (35.5)
High	12 (19.1)	7 (21.9)	5 (16.1)
Patient portal use			
Self-reported frequency, n (%)			
Never	2 (3.1)	N/A	2 (6.3)
1 to 6 times per year	1 (1.6)	1 (3.1)	N/A
7 to 12 times per year	10 (15.6)	4 (12.5)	6 (18.8)
More than 12 times per year	51 (79.7)	27 (84.4)	24 (75.0)
Self-reported activities, n (%)			
Requesting appointments	40 (62.5)	22 (68.8)	18 (56.3)
Viewing laboratory results	63 (98.4)	32 (100.0)	31 (96.9)
Refilling prescriptions	51 (79.7)	26 (81.3)	25 (78.1)
Sending a message to the care team	58 (90.6)	29 (90.6)	29 (90.6)

XSL•FO
RenderX

Characteristics	Overall (n=64)	Usual care (n=32)	Intervention (n=32)
Paying bills	32 (50.0)	13 (40.6)	19 (59.4)
Confirming appointments	57 (89.1)	29 (96.6)	28 (87.5)
Other	2 (3.1)	1 (3.1)	1 (3.1)
Actual use, n (%)			
Any questionnaire completion in the past year	46 (71.9)	25 (78.1)	21 (65.6)

[a]N/A: not applicable.

[b]PROMIS: patient-reported measurement information system.

[c]CHAI: Consumer Health Activation Index.

Intervention Feasibility and Acceptability

Our assessment of intervention feasibility revealed that 47% (15/32) of participants in the intervention arm completed the adherence assessments that were sent in their patient portal accounts. Of those, only 1 indicated they had any adherence challenges, according to EHR data, though this participant self-reported they did not remember having any challenges. Similarly, when asked by the research coordinator, 75% (24/32) of intervention arm participants self-reported they were unaware they had received an email or portal message asking them to complete the PATTERN medication adherence assessment, and therefore did not complete it. When probed for an explanation, most indicated they had accessed their email but did not notice the request (71%, 17/24). Of those who self-reported receiving the request (25%, 8/32), all but 1 found the adherence assessment to be somewhat or very easy to access, and most (63%, 5/8) reported they completed it. Of the 5 who said they completed the survey, 3 (60%) were very satisfied with the experience, 1 (20%) was a little satisfied, and 1 (20%) did not reply. Moreover, although none of these participants reported having any challenges with their medications, 2 revealed they had discussed the medication assessment and medication taking strategies with their clinical care teams and were satisfied with that conversation. Participants who accessed but did not complete the assessment (37%, 3/8), revealed it was not because they encountered challenges accessing or responding to it; rather, they indicated they did not want to complete it at that time.

Discussion

This small patient randomized feasibility study examined the feasibility and acceptability of the PATTERN intervention among older adults with multiple chronic conditions and polypharmacy from a large academic medical center in Chicago, IL. Results revealed that the previsit delivery of portal-based assessments has the potential to enhance patient-provider communication about medication adherence in primary care, as most of our sample self-reported being active users of the patient portal and had a record of a completed portal assessment in the EHR in the last year. This is an important finding as our studies have previously shown that while highly resource-intensive interventions can increase medication adherence among older adults, less-intensive options, such as portal-based assessments, are needed for busy primary care clinics [7]. This finding also has far-reaching implications, as health systems are increasingly relying on patient-reported outcomes collected via the patient portal [15-17].

However, less than half of all intervention arm participants completed the PATTERN adherence assessment ahead of their clinical care visit, as instructed. As such, the current approach likely requires additional nudges or supports to ensure all recipients are able to access and complete the assessments in a timely manner. This could include (1) information explaining how to know when assessments are available for completion in the patient portal; (2) more tailored information designed to entice recipients to open the portal message; and (3) detailed information on how to complete assessments on time. Prior portal-based studies conducted among older adults have also shown that this population could benefit from usability training [18,19]. Yet beyond a “how-to” training, information on why portal-based assessments are useful has also been identified as a need [20]. Future studies should explore the nudges, supports, or training older adults find acceptable, feasible, and appropriate. Furthermore, it may be that assessment results need to be sent to the patient ahead of their visit. These could be accompanied by a list of result-specific questions (perhaps even populated by artificial intelligence) that patients could ask during their visit [21]. Not only would this list serve as a reminder that they had completed the assessment, but as other studies using question prompt lists have shown, it can further care engagement [22,23].

The few participants in our study who reported completing the PATTERN assessment noted they were largely satisfied with it and the resulting conversations they had with their care teams. Had we conducted posttest interviews immediately following the clinical care visit, more participants would have likely been able to share their experiences and whether they found the intervention to be acceptable. Increasingly, studies are revealing that older adults typically have positive experiences with patient portals and technology-delivered assessments [24,25]

Results from this study need to be considered in light of the study limitations. First, this was a small feasibility study conducted within a single academic health system. It is possible that a larger, more diverse study would have resulted in different findings. However, the purpose of this study was to obtain preliminary data that could inform a larger, more diverse trial. There was also no age limit for participation in this study. It is

XSL•FO
RenderX

possible that PATTERN may be more feasible for the "younger" older adults, for example, those individuals who are between the ages of 65 and 75 years. Finally, we may need to account for "proxy users," other individuals appointed by the patient to access and respond to needs in their patient portal accounts. Although nearly all participants in our study had not officially authorized "proxy use" in their accounts, it is possible they relied on unofficial use. A recent national survey revealed that nearly half of all older adults between the ages of 50 and 80 years have allowed others to use their portal accounts [17]. This might occur if a patient's caregiver, for example, used the patient's login information to confirm appointments, update medication information, or use other portal-based features, including assessments [26]. It is possible that participants in the PATTERN intervention arm did not remember completing the medication adherence assessment because their proxy user had completed it for them.

In conclusion, PATTERN, a remote medication adherence monitoring intervention, shows promise for use with older primary care patients experiencing multiple chronic conditions and polypharmacy. However, further adaptation and training are needed to ensure participants are accessing their patient portal accounts in a timely manner ahead of scheduled clinic visits. Results from this small feasibility study may inform other research and clinical activities intending to capture self-reported data from older adults via the patient portal.

Acknowledgments

The work was funded in part by the National Institutes of Health's National Institute on Aging, Grant Number P30AG059988. REDCap software is supported by the National Institutes of Health's National Center for Advancing Translational Sciences, Grant Number UL1TR001422. The opinions expressed in this paper are those of the authors and do not necessarily represent that of the National Institutes of Health.

Data Availability

Deidentified data pertaining to this study may be obtained with reasonable request from the corresponding author.

Conflicts of Interest

APP reports grants from Merck, Pfizer, RRF Foundation for Aging, Lundbeck, Gilead, and Eli Lilly through her institution; and personal fees from Gilead and Remo Health. SCB reports grants from the NIH, Merck, Pfizer, RRF Foundation for Aging, Lundbeck, Gilead, and Eli Lilly through her institution; and personal fees from Gilead, Pfizer, University of Westminster, Lundbeck, and Luto outside the submitted work. RO reports grants from NIH through her institution. Dr. Wolf reports grants from the NIH (NIA, NIDDK, NINR, NHLBI, NINDS), Gordon and Betty Moore Foundation, RRF Foundation for Aging and Eli Lilly, Glaxo Smith Klein, and personal fees from Pfizer, Sanofi, Luto UK, University of Westminster, and Lundbeck. All other authors have no relationships to disclose.

Multimedia Appendix 1
CONSORT flow diagram.
[PDF File, 131 KB - formative_v9i1e71676_app1.pdf]

References

1. Brown MT, Bussell JK. Medication adherence: WHO cares? Mayo Clin Proc 2011 Apr;86(4):304-314. [doi: 10.4065/mcp.2010.0575] [Medline: 21389250]
2. Denneboom W, Dautzenberg MGH, Grol R, De Smet P. Analysis of polypharmacy in older patients in primary care using a multidisciplinary expert panel. Br J Gen Pract 2006 Jul;56(528):504-510. [Medline: 16834876]
3. Delara M, Murray L, Jafari B, et al. Prevalence and factors associated with polypharmacy: a systematic review and meta-analysis. BMC Geriatr 2022 Jul 19;22(1):601. [doi: 10.1186/s12877-022-03279-x] [Medline: 35854209]
4. Marcum ZA, Hanlon JT, Murray MD. Improving medication adherence and health outcomes in older adults: an evidence-based review of randomized controlled trials. Drugs Aging 2017 Mar;34(3):191-201. [doi: 10.1007/s40266-016-0433-7] [Medline: 28074410]
5. Bailey SC, Oramasionwu CU, Wolf MS. Rethinking adherence: a health literacy-informed model of medication self-management. J Health Commun 2013;18 Suppl 1(Suppl 1):20-30. [doi: 10.1080/10810730.2013.825672] [Medline: 24093342]
6. Xu HY, Yu YJ, Zhang QH, Hu HY, Li M. Tailored interventions to improve medication adherence for cardiovascular diseases. Front Pharmacol 2020;11:510339. [doi: 10.3389/fphar.2020.510339] [Medline: 33364935]
7. Federman AD, O'Conor R, Mindlis I, et al. Effect of a self-management support intervention on asthma outcomes in older adults: the SAMBA study randomized clinical trial. JAMA Intern Med 2019 Aug 1;179(8):1113-1121. [doi: 10.1001/jamainternmed.2019.1201] [Medline: 31180474]

XSL•FO
RenderX

8. O'Conor R, Arvanitis M, Wismer G, et al. Rationale and design of the regimen education and messaging in diabetes (REMinD) trial. Contemp Clin Trials 2019 Aug;83:46-52. [doi: 10.1016/j.cct.2019.06.016] [Medline: 31260791]
9. Realising the potential of primary health care. : OECD; 2020 May 30 URL: https://www.oecd.org/en/publications/realising-the-potential-of-primary-health-care_a92adee4-en.html [accessed 2025-04-17] [doi: 10.7748/phc.30.2.14.s14]
10. Yoon ES, Hur S, Curtis LM, et al. A multifaceted intervention to improve medication adherence in kidney transplant recipients: an exploratory analysis of the fidelity of the TAKE IT trial. JMIR Form Res 2022 May 5;6(5):e27277. [doi: 10.2196/27277] [Medline: 35511225]
11. Pack AP, Bailey SC, O'Conor R, et al. Phenotyping adherence through technology-enabled reports and navigation (the PATTERN Study): qualitative study for intervention adaptation using the exploration, preparation, implementation, and sustainment framework. JMIR Form Res 2024 Oct 17;8:e54916. [doi: 10.2196/54916] [Medline: 39418094]
12. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
13. Cella D, Riley W, Stone A, et al. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010 Nov;63(11):1179-1194. [doi: 10.1016/j.jclinepi.2010.04.011] [Medline: 20685078]
14. Wolf MS, Smith SG, Pandit AU, et al. Development and validation of the consumer health activation index. Med Decis Making 2018 Apr;38(3):334-343. [doi: 10.1177/0272989X17753392] [Medline: 29436308]
15. Deshpande PR, Rajan S, Sudeepthi BL, Abdul Nazir CP. Patient-reported outcomes: a new era in clinical research. Perspect Clin Res 2011 Oct;2(4):137-144. [doi: 10.4103/2229-3485.86879] [Medline: 22145124]
16. Smith SM, Wallace E, O'Dowd T, Fortin M. Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. Cochrane Database Syst Rev 2021 Jan 15;1(1):CD006560. [doi: 10.1002/14651858.CD006560.pub4] [Medline: 33448337]
17. Kirch M, Solway E, Roberts S, et al. Use and experiences with patient portals among older adults: University of Michigan national poll on healthy aging. : University of Michigan; 2023 URL: https://www.healthyagingpoll.org/reports-more/report/use-and-experiences-patient-portals-among-older-adults [accessed 2025-04-17]
18. Son H, Nahm ES, Zhu S, et al. Older adults' perception on and use of patient portals: a comparative analysis of two samples. Comput Inform Nurs 2022 Jan 1;40(1):61-68. [doi: 10.1097/CIN.0000000000000802] [Medline: 34347646]
19. Nahm ES, Son H, Yoon JM. Older adults' use of patient portals: Experiences, challenges, and suggestions shared through discussion board forums. Geriatr Nurs 2020;41(4):387-393. [doi: 10.1016/j.gerinurse.2019.12.001] [Medline: 31899003]
20. Price-Haywood EG, Harden-Barrios J, Ulep R, Luo Q. eHealth literacy: patient engagement in identifying strategies to encourage use of patient portals among older adults. Popul Health Manag 2017 Dec;20(6):486-494. [doi: 10.1089/pop.2016.0164] [Medline: 28384076]
21. Vordenberg SE, Nichols J, Marshall VD, Weir KR, Dorsch MP. Investigating older adults' perceptions of AI tools for medication decisions: vignette-based experimental survey. J Med Internet Res 2024 Dec 16;26:e60794. [doi: 10.2196/60794] [Medline: 39680885]
22. Brandes K, Linn AJ, Butow PN, van Weert JCM. The characteristics and effectiveness of question prompt list interventions in oncology: a systematic review of the literature. Psychooncology 2015 Mar;24(3):245-252. [doi: 10.1002/pon.3637] [Medline: 25082386]
23. Sansoni JE, Grootemaat P, Duncan C. Question prompt lists in health consultations: a review. Patient Educ Couns 2015 Jun 3:S0738-3991(15)00258-X. [doi: 10.1016/j.pec.2015.05.015] [Medline: 26104993]
24. Weir KR, Maitah Y, Vordenberg SE. Older adults' experiences with an online survey. JMIR Hum Factors 2025 Jan 8;12:e65684. [doi: 10.2196/65684] [Medline: 39773887]
25. Sakaguchi-Tang DK, Bosold AL, Choi YK, Turner AM. Patient portal use and experience among older adults: systematic review. JMIR Med Inform 2017 Oct 16;5(4):e38. [doi: 10.2196/medinform.8092] [Medline: 29038093]
26. Wolff JL, Dukhanin V, Burgdorf JG, DesRoches CM. Shared access to patient portals for older adults: implications for privacy and digital health equity. JMIR Aging 2022 May 4;5(2):e34628. [doi: 10.2196/34628] [Medline: 35507405]

ABBREVIATIONS

CHAI: Consumer Health Activation Index
EHR: electronic health record
PATTERN: Phenotyping Adherence Through Technology-Enabled Reports and Navigation
PROMIS: patient-reported measurement information system

XSL•FO
RenderX

Edited by A Mavragani; submitted 23.01.25; peer-reviewed by M Shade, S Leitch; revised version received 11.02.25; accepted 13.02.25; published 24.04.25.

Please cite as:
Pack A, Bailey SC, O'Conor R, Velazquez E, Wismer G, Yeh F, Curtis LM, Alcantara K, Wolf MS
A Portal-Based Intervention (PATTERN) Designed to Support Medication Use Among Older Adults: Feasibility and Acceptability Study
JMIR Form Res 2025;9:e71676
URL: https://formative.jmir.org/2025/1/e71676
doi:10.2196/71676

© Allison Pack, Stacy C Bailey, Rachel O'Conor, Evelyn Velazquez, Guisselle Wismer, Fangyu Yeh, Laura M Curtis, Kenya Alcantara, Michael S Wolf. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Cardiac Self-Efficacy Improvement in a Digital Heart Health Program: Secondary Analysis From a Feasibility and Acceptability Pilot Study

Kimberly G Lockwood[1], PhD; Priya R Kulkarni[2], PhD; OraLee H Branch[1], PhD; Sarah A Graham[1], PhD

[1]Clinical Research Department, Lark Health, 809 Cuesta Dr. Suite B #1033, Mountain View, CA, United States

[2]Digital Health Technologies Department, Roche Information Solutions, Santa Clara, CA, United States

Corresponding Author:
Sarah A Graham, PhD
Clinical Research Department, Lark Health, 809 Cuesta Dr. Suite B #1033, Mountain View, CA, United States

Abstract

Background: Lifestyle modification programs play a critical role in preventing and managing cardiovascular disease (CVD). A key aim of many programs is improving patients' self-efficacy. In-person lifestyle modification programs can enhance self-efficacy in managing CVD risk, also known as cardiac self-efficacy (CSE). However, such programs are typically staffing and resource intensive. Digital lifestyle modification programs may offer a scalable and accessible way to improve CSE, but this has not been shown in prior research.

Objective: This study examined changes in CSE among individuals using a digital lifestyle modification program for cardiovascular health. Evaluation of improvement in CSE was a secondary goal of a feasibility and acceptability pilot study of a digital program for Heart Health.

Methods: Participants were individuals with elevated risk for CVD who enrolled in a 90-day pilot study that involved mobile app–based, artificial intelligence–powered health coaching and educational lessons focused on behaviors that promote cardiovascular health. Participants completed the 9-item CSE Scale at baseline and in month 2. Changes in confidence in participants' ability to manage their cardiovascular health were assessed.

Results: The sample included 273 (n=207, 61.2% female; mean age 59.3, SD 10.1 years) participants who submitted a complete CSE Scale at baseline and in month 2. The total CSE Scale score increased by 12.9% (P<.001) from baseline to month 2. Additionally, there were significant increases in mean score on each of the 9 individual CSE Scale items (all P<.001), with the largest increases in confidence "in knowing when to call or visit the doctor for your heart disease" (17% increase; P<.001), "in knowing how much physical activity is good for you" (16.3% increase; P<.001), and "that you can get regular aerobic exercise" (19% increase; P<.001).

Conclusions: The present analyses indicate that participants in a digital lifestyle modification program for cardiovascular health showed significant improvements in CSE within 2 months. This work adds to the growing literature examining ways to improve health-related self-efficacy and scalable access to programs for prevention and management of CVD.

(JMIR Form Res 2025;9:e60676) doi:10.2196/60676

KEYWORDS

digital health; self-efficacy; behavior change; prevention; cardiovascular health; digital heart health; lifestyle; heart health; health coaching; mobile phone

Introduction

Background

Cardiovascular disease (CVD) is the leading cause of death and the most common chronic disease in the United States, with 126.9 million adults (nearly 50% of the population) living with some form of CVD [1]. CVD is a key public health concern, as well as an economic burden, with costs related to CVD estimated to be US $378 billion, exceeding all other chronic health conditions [1,2]. Prevention and risk reduction for patients at risk for CVD is important to reduce mortality, disability, and economic burden [3]. According to the American Heart Association, engaging in a healthy lifestyle over the course of one's lifespan is the most important factor in CVD prevention [4]. As maintaining a healthy lifestyle is difficult for many adults [5], 1 method of CVD prevention and management is participation in behavioral interventions focused on healthy lifestyle modifications [4]. Position statements from the US Preventive Services Task Force state that lifestyle modification programs that emphasize a healthy diet and physical activity have a wide variety of cardiovascular health benefits and CVD

XSL•FO
RenderX

risk reduction among individuals with and without a diagnosis of CVD [6-8]. To enable individuals to make lasting behavior changes to improve their cardiovascular health, a key aim of many lifestyle modification programs is improving patients' self-efficacy to manage their health [7,9,10].

Importance of Self-Management in CVD

Self-efficacy, or confidence in one's ability to execute and control certain behaviors, is crucial for patient empowerment and self-management of cardiovascular health [11]. There are several evaluation tools for measuring self-efficacy in the context of disease self-management. For instance, general scales for managing chronic disease include the Chronic Disease Self-Efficacy Scale [12] and the Self-Efficacy for Managing Chronic Disease 6-Item Scale [13]. The Cardiac Self-Efficacy (CSE) Scale is a measure of self-efficacy specifically for self-management of CVD [14]. The CSE Scale is a self-report survey that evaluates individuals' beliefs in their ability to manage their cardiovascular health and the challenges that come with heart disease [14,15]. These self-efficacy evaluation tools provide insight into how confident an individual feels about carrying out behaviors that are beneficial to disease prevention and self-management.

Self-Efficacy in CVD Prevention and Management

Self-efficacy is a key target of lifestyle modification programs because it predicts a range of healthy behaviors and outcomes among individuals with CVD. For instance, higher self-efficacy is associated with greater engagement in self-care behaviors [16,17], such as higher levels of physical activity [18], healthier dietary habits [19], and better medication adherence [20]. Among cardiac patients, higher self-efficacy is also associated with better attendance in cardiac rehabilitation programs [21]. Meta-analytic results also indicate that stronger self-efficacy among those with CVD is linked with a range of metrics of health-related quality of life, such as physical functioning and mobility, emotional functioning, mental health, and social functioning [22].

A review of the literature indicated that in-person lifestyle modification programs for CVD have been successful in improving health-related self-efficacy [10]. One program involved 6 weeks of in-person, 2.5-hour educational sessions led by trained peer leaders who received 4 days of training; results of this study showed significant improvements in self-efficacy related to managing their health [23]. Another program that included 8 hour-long, individual sessions between each patient and a health professional found significant increases in self-efficacy related to self-care after 12 weeks [24].

There is also evidence that in-person lifestyle modification programs can improve self-efficacy specific to managing cardiovascular health, as measured by the CSE Scale. Recent studies show that intensive, in-person lifestyle modification programs for CVD led by nurses and health professionals can improve CSE scores over time [25,26], which have been shown to lead to downstream improvement of health behaviors and outcomes over time [14,27,28].

The evidence reviewed here shows that, although in-person programs have succeeded in improving self-efficacy, these intensive programs typically involve a high degree of human contact and staffing resources; this makes them difficult to implement or impossible to scale up to larger populations. Moreover, in-person programs can present access challenges to patients, such as having to travel long distances to attend [21]. As such, scalable and accessible lifestyle modification programs to improve CSE are needed.

Digital Health and CSE

Recent years have seen an increase in the availability of digital lifestyle modification programs for CVD prevention and management [29-31]. These digital health solutions have become increasingly available and can provide many benefits over in-person programs, such as increased accessibility, on-demand support, availability outside of typical working hours, lower cost, and scalability to larger populations [32]. These benefits have the potential to increase participation in lifestyle modifications for CVD [30]. Moreover, digital health programs for CVD prevention and management can significantly improve patient behaviors and show promise in delivering care that is accessible, cost-effective, and patient-focused [33].

Despite these benefits, many outcomes including CSE, remain unexplored in digital lifestyle modification programs and more evidence is required [31]. There is some evidence that digital interventions can increase general self-efficacy [34,35], suggesting that it may be possible for digital CVD programs to improve CSE. However, use of the CSE Scale in digital health studies is extremely limited. As such, the main objective of the present analyses was to test whether participation in an artificial intelligence (AI)–powered digital lifestyle modification program for heart health led to improved CSE.

Heart Health Program

The Lark Heart Health (Lark Technologies, Inc) program is a newly developed AI-powered lifestyle modification program that provides health behavior coaching to prevent and manage atherosclerotic CVD and coronary artery disease by targeting key CVD risk factors. We evaluated the acceptability and feasibility of this new program in a pilot study; a detailed study description and main outcome results for this pilot study are available elsewhere [36].

The Heart Health program targets individuals in primary prevention (ie, no history of CVD) and those in stable secondary prevention after a cardiac event. This program delivers fully remote, unlimited, real-time cardiovascular health coaching focused on digital nutrition therapy, medication adherence counseling, and personalized guidance on weight loss, physical activity, tobacco cessation, stress, and sleep. Heart Health is available to covered adult members of Lark's health care partners via a smartphone app. Full details on the content and design of the Heart Health program are provided in previous work [36]. Full detail on Lark's lifestyle change programs for diabetes prevention and hypertension care are provided elsewhere [37,38].

Study Objective

Beyond the primary acceptability and feasibility aims, a secondary goal of the Heart Health pilot study was to determine

XSL•FO
RenderX

whether CSE could improve early in this study. Specifically, we hypothesized that participants in the Heart Health pilot study who participated in the program for a minimum of 40 days would show improvement in CSE. This would indicate that a digital heart health program can help to improve CSE over a relatively short period, an important first step in developing scalable and accessible digital programs for improving self-efficacy among individuals at risk for CVD.

Methods

Study Design

This paper focuses on data from a real-world, single-arm, observational pilot study of a digital health app-based program called Lark Heart Health. The Heart Health program provides health behavior screeners and coaching to individuals at elevated risk for heart disease. The pilot study examined the feasibility of deploying screener surveys to participants of the program and the acceptability of coaching focused on building knowledge and improving self-management of atherosclerotic CVD risk. A detailed overview of study methods and primary outcomes [36] and predictors of interest in the program [39] have been reported in previously published works.

This study was 90 days in duration, during which participants received coaching on heart health and lifestyle modification. Participants could engage with the Heart Health program via (1) completing educational lessons, (2) engaging in coaching conversations with Lark's conversational AI-powered digital coach, (3) logging meals in the app, and (4) tracking their weight using a digital smart scale that automatically synced with the app.

Additionally, study participation involved completion of baseline screener surveys and assessment of CSE at study start and again in month 2 (after 40 days in the program). We measured CSE in month 2 for several reasons. First, self-efficacy has been shown to improve after relatively short digital health interventions for lifestyle change (eg, 6 week programs) [34], particularly compared to physical health outcomes that may take months or even years to improve. Second, our study involved completing several surveys and outcome measurements that were intentionally spread out over the course of the 90 days, rather than clustering all outcome measurements at the end of this study; this approach was aimed at reducing participant burden, a crucial factor for retention in digital health programs. Third, measuring CSE early in this study enabled us to maximize the sample size of individuals actively engaging with the program who completed the survey at both time points. Between study start date and the month 2 reassessment of CSE, the Heart Health program delivered 5 weekly educational lessons focused on risk factors for heart disease, understanding cholesterol, physical activity and how it impacts blood pressure, healthy eating, and how sodium impacts blood pressure.

Ethical Considerations

All participants provided informed consent to participate in this study through an electronic consent form. The pilot study received institutional review board approval (Advarra Pro00061694). Study personnel instituted appropriate safeguards to prevent any unauthorized use or disclosure of personal health information and implemented administrative, physical, and technical safeguards to protect the confidentiality, integrity, and availability of protected health information. Lark is compliant with HIPAA (Health Insurance Portability and Accountability Act) privacy and security rules and all applicable regulations. Lark is also SOC 2 (System and Organization Controls 2) and HITRUST (Health Information Trust Alliance) certified. As compensation for study participation, participants received a cellular smart scale and opportunities for nominal gift card incentives, as well as a Fitbit (Google LLC) upon study completion. None of these incentives were tied to completion of the CSE Scale.

Participants and Study Flow

The research team recruited participants through a health care partner and via online recruitment through 1nHealth using marketing emails or SMS messages and printed mailers (health care partner recruitment only). Exclusion criteria for the study included: BMI <25 and ≥50; serious uncontrolled health conditions that had been active in the last 6 months; current pregnancy or plans to become pregnant within the next 6 months; recent history of a medical professional advising against participation in a healthy lifestyle program; a medical reason preventing 10 consecutive minutes of moderate physical exercise; and not having a smartphone with internet connection. As the Heart Health program focused heavily on improving health behaviors (eg, diet and exercise), the study also excluded individuals who reported regularly engaging in strenuous physical activity and individuals who did not report unhealthy dietary behaviors.

Participants enrolled in the pilot study from March 31 to September 15, 2022, using the Lark app (version 5.2.6), downloading the Heart Health app and providing a first weight delineated study enrollment. To be included in the present analyses, participants had to be enrolled until the end of this study (ie, did not withdraw from this study), complete the first CSE Scale at baseline for descriptive analyses, and complete the second CSE Scale at month 2 for the CSE improvement analyses. Completion of the CSE Scale occurred outside of the Lark app and full completion was not required for study participation.

Data Collection and Measures

Participant Characteristics

We assessed basic demographic and health history characteristics at baseline using a modified version of the INTERHEART Modifiable Risk Score survey [40,41]. Items on this survey included age, sex, history of high blood pressure, tobacco use history, and typical physical activity in leisure time. Participants also provided their height and weight for calculation of BMI and the racial group they identified with.

CSE

Emails from the research team prompted participants to report on CSE shortly after study start (day 1 - 2 of this study) and again in month 2 (after day 40 of this study) using a modified version of the CSE Scale. The CSE Scale evaluates individuals'

XSL•FO
RenderX

beliefs in their ability to manage their cardiovascular health and the challenges that come with heart disease using 13 questions in 3 dimensions [14]. The Heart Health study used 2 dimensions of the CSE Scale that are most relevant to our study sample, "control illness" and "maintain functioning," resulting in a 9-item survey. We did not include the "control symptoms" dimension, as the questions are not relevant to individuals without a CVD diagnosis (ie, most individuals in the present sample) [15]. Each item is scored on a 5-point Likert scale (0=not at all confident, 1=somewhat confident, 2=moderately confident, 3=very confident, and 4=completely confident) and the total score is calculated by summing scores on each item, for a maximum score of 36. As the program did not require participants to submit the CSE Scale and permitted them to skip items, calculation of the total sum score was not possible for all study participants.

Statistical Analysis

We conducted all statistical analyses with RStudio (version 2022.07.0; R Foundation). Continuous variables were normally distributed. The first set of analyses examined relationships between baseline participant characteristics and baseline CSE Scale score. Specifically, we tested the relationship of baseline CSE with demographic characteristics (age, sex, and race), health status and health history characteristics (high blood pressure, cardiac event, or CVD diagnosis—secondary prevention, and type II diabetes), and health behaviors (tobacco use and typical physical activity in leisure time). We used bivariate regressions for continuous variables (age and BMI), independent samples *t* tests for categorical variables with 2 levels (sex and health history variables), and 1-way ANOVAs for categorical variables with 3 levels (race and health behaviors). Baseline CSE analyses included participants who completed the baseline CSE Scale (n=341).

To address the primary study objective, the second set of analyses tested whether there was a statistically significant change in CSE from baseline to month 2, using paired *t* tests ($P<.05$). We first tested whether there was a significant change in the CSE sum score and then whether there was improvement in each of the 9 individual items on the CSE. These analyses included all participants who completed both the baseline CSE and month 2 CSE scales (n=273).

Results

Figure 1 shows the flow of participants through this study.

XSL•FO
RenderX

Figure 1. Flowchart of participants showing enrollment and inclusion in the analytic sample. CSE: cardiac self-efficacy.

XSL•FO
RenderX

Descriptive Statistics

We assessed descriptive statistics for baseline characteristics of participants who completed the baseline CSE Scale (n=341) and participants who completed both the baseline and month 2 CSE (n=273); both samples are shown in Table 1, as we conducted baseline descriptive analyses on the larger sample and improvement analyses on the smaller subsample (Table 1). As shown by the table, the 2 sets of participants were very similar with respect to key characteristics. Across both sets participants were, on average, aged 60 years (range=41 - 76 years) and classified as obese class I (BMI range=25 - 49). Just over 60% (n=207) of the participants were female, had a history of high blood pressure, and were never smokers. Approximately three-quarters of each sample reported typically being either mainly sedentary or engaging in low-effort mild exercise in their leisure time.

Table . Baseline characteristics of participants.

		Baseline CSE[a] complete (n=341)	Baseline and month 2 CSE complete (n=273)
Age (years), mean (SD)		60 (10.1)	59.3 (10.1)
Sex (female), n (%)		207 (60.7)	167 (61.2)
Race, n (%)			
	Black	26 (7.6)	22 (8.1)
	White	241 (70.7)	192 (70.3)
	Other	70 (20.5)	57 (20.9)
Recruitment source (health partner), n (%)		207 (60.7)	154 (56.4)
Baseline BMI, mean (SD)		32.8 (5.8)	32.9 (5.9)
High blood pressure history, n (%)		198 (58.1)	162 (59.3)
History of CVD[b] diagnosis or event, n (%)		41 (12.0)	34 (12.5)
Type II diabetes history, n (%)		52 (15.2)	37 (13.6)
Tobacco use status, n (%)			
	Never	203 (59.5)	171 (62.6)
	Former	110 (32.3)	82 (30.0)
	Current	27 (7.9)	19 (6.9)
Physical activity in leisure time, n (%)			
	Mainly sedentary	121 (35.5)	94 (34.4)
	Mild exercise, low effort	145 (42.5)	116 (42.5)
	Moderate exercise	72 (21.1)	61 (22.3)
Baseline CSE score, mean (SD)		25.1 (7.3)	25.1 (7.3)

[a]CSE: cardiac self-efficacy.

[b]CVD: cardiovascular disease.

Relationships Between Baseline CSE Scores and Participant Characteristics

Next, we examined relationships between baseline participant characteristics and baseline CSE Scale scores for all participants who completed the baseline CSE Scale. Of the demographic characteristics, age was significantly associated with CSE, such that CSE increased with age (n=341, β=.12, SE=0.04, *P*=.03). There was no statistically significant difference in baseline CSE by sex (t_{338}=0.19, *P*=.43) or racial group (F_{340}=2.72, *P*=.07).

Among the health history characteristics, participants with a history of type II diabetes tended to have higher baseline CSE (mean 27.48, SD 6.47, n=52) compared to participants without diabetes history (mean 24.71, SD 7.42, n=289; t_{77}=2.78, *P*=.003). There was no difference in baseline CSE between participants with a history of high blood pressure and those without (t_{322}=0.91, *P*=.36) or participants in primary prevention versus secondary prevention (t_{50}=0.57, *P*=.57). There was also a significant association between BMI and CSE, with lower BMI individuals reporting higher CSE (n=341, β=−.11, SE=0.07, *P*=.047).

For the health behavior variables, there was no difference in CSE based on tobacco use status (F_{339}=1.61, *P*=.20) or typical physical activity in leisure time (F_{337}=1.51, *P*=.22)

Change in CSE From Baseline to Month 2

In testing for changes in the CSE from baseline to month 2, we found a significant increase in the overall sum score on the CSE Scale from baseline to month 2 (Table 2). There were also significant increases in each individual item on the CSE (all *P*<.001). The individual CSE items with the greatest increases were “confidence in knowing when to call or visit the doctor,”

XSL•FO
RenderX

"confidence in knowing how much physical activity is good for you," and "confidence that you can get regular aerobic exercise."

Table . Changes in CSE[a] score from baseline to month 2 (n=273)[b]. The baseline CSE survey was surfaced to participants on day 2, and the month 2 CSE survey was deployed on day 40.

		Baseline, mean (SD)	Month 2, mean (SD)	Baseline to month 2, raw change	Baseline to month 2, change in %	*t* test (*df*)	*P* value
How confident are you that you know…							
	When to call or visit your doctor about your heart disease?	2.7 (1.1)	3.1 (0.9)	+0.5	+17	7.9 (272)	<.001
	How to make your doctor understand your concerns about your heart?	2.8 (1.1)	3.1 (0.9)	+0.3	+10.7	5.1 (272)	<.001
	How to take your cardiac medications?	3.1 (1.1)	3.4 (0.9)	+0.4	+11.4	5.7 (272)	<.001
	How much physical activity is good for you?	2.7 (1.1)	3.2 (0.8)	+0.5	+16.3	7.7 (272)	<.001
How confident are you that you can...							
	Maintain your social activities?	2.8 (1)	3.2 (0.8)	+0.4	+13	6.4 (272)	<.001
	Maintain your usual activities at home with your family?	3 (1)	3.3 (0.8)	+0.2	+7.7	4.2 (272)	<.001
	Maintain your usual activities at work?	2.9 (1.1)	3.2 (0.9)	+0.3	+10.7	4.8 (272)	<.001
	Maintain your sexual relationship with your spouse?	2.6 (1.3)	2.9 (1.1)	+0.3	+12.3	4.3 (272)	<.001
	Get regular aerobic exercise?	2.5 (1.2)	2.9 (1.1)	+0.5	+19	6.4 (272)	<.001
Overall sum score on the CSE		25.07 (7.3)	28.3 (6.3)	+3.2	+12.9	9.2 (272)	<.001

[a]CSE: cardiac self-efficacy.

[b]n=273 participants submitted the cardiac self-efficacy with all items completed.

Discussion

Summary and Interpretation of Results

These analyses examined improvement in CSE as a secondary goal in a digital lifestyle modification program for heart health focused on AI-supported health coaching, educational lessons, behavioral tracking, and weight management. We found that the overall CSE Scale score improved significantly from baseline to month 2 of the program. Moreover, mean scores on each individual item of the CSE increased significantly over this short period, with the largest increases in patients' confidence in knowing when to contact their doctor about their heart health (+17%) and items related to confidence in knowing how much physical activity to do (+16.3%) and getting regular exercise (+19%). The results of these secondary analyses of the Heart Health program pilot study support the assertion that digital health programs can play an important role in increasing self-efficacy to better manage their health over a relatively short period.

This study is the first digital health study we are aware of to show improvements in CSE scores through participation in a digital lifestyle modification program for heart health. There is evidence that in-person lifestyle change programs for CVD can improve CSE [25,26], and literature showing that digital interventions can increase general self-efficacy [34,35]. These results are consistent with results from 2 prior studies with cardiac patients showing that in-person intervention programs

XSL•FO
RenderX

for CVD management can improve CSE scores; however, these in-person programs involved intensive, human-led interventions that required significant staffing and economic resources [25,26].

One recent study using a 60-day text message intervention implemented in patients recently discharged after acute coronary syndrome had no significant effect on the CSE [42]. The researchers implemented an automated 1-way text message campaign through which patients received messages about self-management, healthy living, and follow-up care. There are many key differences in the study design by Ross et al [42] compared to our study, but it is possible that we found improvements in CSE because the Heart Health program involves 2-way interaction with AI-supported coaching and active behavior tracking, rather than a 1-way messaging system.

It is also notable that 2 of the CSE items with the biggest improvement from baseline to month 2 were centered on confidence related to physical activity: "confidence in knowing how much physical activity is good for you" (+16%), and "confidence that you can get regular aerobic exercise" (+19%). Education on physical activity is a key component of the Heart Health program, with an educational lesson on Physical Activity for Low Blood Pressure delivered during the third weekly lesson in the program. The Heart Health digital coach also provides ongoing interactive coaching on physical activity, encouraging participants to log their physical activity in the app. As such, the present results suggest that the education and coaching in the first 2 months of the Heart Health program support improvements in confidence related to physical activity. Similarly, prior work has shown that digital health interventions focused on physical activity can improve exercise-related self-efficacy over 24 weeks [43]. Future research in this area could examine whether improvements in CSE Scale score during the first 2 months of the program predict longitudinal increases in physical activity.

Beyond the primary results of these analyses, we also found several relationships between baseline CSE score and participant characteristics. Specifically, we found that older age, lower BMI, and having a history of type II diabetes were all associated with having higher CSE scores at baseline. Importantly, a significant body of literature has shown that higher self-efficacy among older persons is associated with a range of positive health benefits, including increased self-care, better health behaviors, increased energy, and decreased pain and discomfort [44,45]. This is particularly relevant to our study, as participants in this sample tended to be older adults, with a mean age of 60 years and up to 76 years. Notably, the broader literature indicates that self-efficacy and closely related constructs, such as self-esteem, tend to increase with age due to increasing knowledge and maturity, greater control over life circumstances, higher likelihood of occupying positions of power and status, and improved coping resources and processes [46]. Indeed, a large meta-analysis of 164,868 participants showed that self-esteem tends to increase steadily with age, reaching peak levels between ages 60 - 70 before declining slightly after age 70 and steeply after age 90 [47].

Our finding that lower BMI was associated with high CSE scores aligns with prior work showing that cardiac patients with lower BMI also tended to have higher CSE [48] and with population-based research showing lower health-related self-efficacy among individuals with higher BMI [49]. This relationship may be due to individuals with higher BMI having a history of difficulty managing their weight and lower self-esteem [50], factors that are closely related to self-efficacy.

The link between the history of type II diabetes and baseline CSE is also consistent with prior literature showing that patients with a history of chronic disease diagnosis tend to have higher CSE [48,51]. This relationship is likely due to these individuals already having some experience with managing a health condition and potentially receiving prior education on managing cardiometabolic conditions. However, it is surprising that a history of high blood pressure and history of cardiac events or CVD diagnoses were not associated with differences in baseline CSE in this study's sample.

Limitations and Future Directions

There are several limitations to our results. First, data for these secondary analyses came from an observational pilot study focused on the acceptability and feasibility of a new digital heart health program. As such, these results should be considered preliminary, rather than conclusive. We were also unable to compare these results to a control group, and causal conclusions are limited. Additionally, participants could opt out of any survey, including the CSE survey, if they did not wish to complete it. As a result, we could only include individuals in these analyses if they submitted a complete CSE Scale. Completion of the CSE Scale also occurred outside of the Lark app to be used for pilot study purposes only. Thus, the sample may be biased toward participants who were more engaged in the program or more likely to check their emails. The CSE Scale is also limited as it includes questions about confidence in work settings and in marital life, which may not apply to all participants. Finally, this digital intervention was limited to those with a smartphone with internet access who had sufficient digital literacy to use the Lark app. As a result, this may exclude individuals of lower socioeconomic status. For instance, the 2024 Pew Research Center data showed that, although 90% of US adults have a smartphone, individuals in the lowest income bracket (US <$30,000/year) have lower rates of smartphone ownership (79%) compared to those in higher income brackets (US >$100,000/year; 98%) [52].

The pilot study results presented here provide a foundation for several lines of future work. For instance, future work including a control group can provide causal evidence for the impact of the Heart Health program on CSE. Additionally, a key future direction for this study will be assessing whether increased CSE predicts longitudinal clinical outcomes. Given that the duration of the pilot study was only a short period of 90 days, we did not expect to see a meaningful relationship between improved CSE Scale scores and clinical outcomes. This is particularly true given that we measured CSE improvements nearly halfway through the program. Future research could examine longitudinal changes in CVD risk factors, such as blood pressure or cholesterol levels. Based on the literature, we would expect that improvements in CSE would predict improved cardiovascular outcomes.

XSL•FO
RenderX

Conclusions

In summary, the present analyses indicate that participants in a digital lifestyle modification program for heart health showed significant improvements in CSE within 2 months. This work adds to the growing literature examining ways to improve health-related self-efficacy for the prevention and management of chronic disease and long-term health benefits.

Acknowledgments

The Heart Health pilot study was supported by funds from Roche Information Solutions and Lark Health. KGL, OHB, and SAG received salaries from Lark Health. PRK received a salary from Roche Information Solutions.

Data Availability

The dataset analyzed during this study is not publicly available as these analyses include sensitive, potentially identifiable information that was obtained through a collaboration between Lark Health and its health care partners, but data may be available from the corresponding author upon reasonable request.

Authors' Contributions

All authors contributed to the conception of the work and made critical revisions to the content of this paper. KGL contributed to data collection and management, led data analysis, and drafted and revised this paper. PRK and SAG provided critical reviews and interpretation of all statistical analyses. OHB was the principal investigator of the pilot study, supervised data collection and management, and supported analyses and interpretation of the data.

Conflicts of Interest

KGL, OHB, and SAG were employees of Lark Health at the time of manuscript development. PRK is an employee of Roche Information Solutions.

References

1. Tsao CW, Aday AW, Almarzooq ZI, et al. Heart disease and stroke statistics-2022 update: a report from the American Heart Association. Circulation 2022 Feb 22;145(8):e153-e639. [doi: 10.1161/CIR.0000000000001052] [Medline: 35078371]
2. McClellan M, Brown N, Califf RM, Warner JJ. Call to action: urgent challenges in cardiovascular disease: a presidential advisory from the American Heart Association. Circulation 2019 Feb 26;139(9):e44-e54. [doi: 10.1161/CIR.0000000000000652] [Medline: 30674212]
3. Rippe JM. Lifestyle strategies for risk factor reduction, prevention, and treatment of cardiovascular disease. Am J Lifestyle Med 2019;13(2):204-212. [doi: 10.1177/1559827618812395] [Medline: 30800027]
4. Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines. Circulation 2019 Sep 10;140(11):e596-e646. [doi: 10.1161/CIR.0000000000000678] [Medline: 30879355]
5. Kris-Etherton PM, Petersen KS, Velarde G, et al. Barriers, opportunities, and challenges in addressing disparities in diet-related cardiovascular disease in the United States. J Am Heart Assoc 2020 Apr 7;9(7):e014433. [doi: 10.1161/JAHA.119.014433] [Medline: 32200727]
6. US Preventive Services Task Force, Krist AH, Davidson KW, et al. Behavioral counseling interventions to promote a healthy diet and physical activity for cardiovascular disease prevention in adults with cardiovascular risk factors: US Preventive Services Task Force recommendation statement. JAMA 2020 Nov 24;324(20):2069-2075. [doi: 10.1001/jama.2020.21749] [Medline: 33231670]
7. Patnode CD, Redmond N, Iacocca MO, Henninger M. Behavioral counseling interventions to promote a healthy diet and physical activity for cardiovascular disease prevention in adults without known cardiovascular disease risk factors: updated evidence report and systematic review for the US Preventive Services Task Force. JAMA 2022 Jul 26;328(4):375-388. [doi: 10.1001/jama.2022.7408] [Medline: 35881116]
8. US Preventive Services Task Force, Mangione CM, Barry MJ, et al. Behavioral counseling interventions to promote a healthy diet and physical activity for cardiovascular disease prevention in adults without cardiovascular disease risk factors: US Preventive Services Task Force recommendation statement. JAMA 2022 Jul 26;328(4):367-374. [doi: 10.1001/jama.2022.10951] [Medline: 35881115]
9. Foroumandi E, Kheirouri S, Alizadeh M. The potency of education programs for management of blood pressure through increasing self-efficacy of hypertensive patients: a systematic review and meta-analysis. Patient Educ Couns 2020 Mar;103(3):451-461. [doi: 10.1016/j.pec.2019.09.018] [Medline: 31558325]
10. Katch H, Mead H. The role of self-efficacy in cardiovascular disease self-management: a review of effective programs. PI 2010;2010(2):33-44. [doi: 10.2147/PI.S12624]

XSL•FO
RenderX

11. Brady TJ, Murphy L, O'Colmain BJ, et al. A meta-analysis of health status, health behaviors, and health care utilization outcomes of the Chronic Disease Self-Management Program. Prev Chronic Dis 2013;10:120112. [doi: 10.5888/pcd10.120112] [Medline: 23327828]
12. Lorig K, Stewart A, Ritter P, González V, et al. Outcome Measures for Health Education and Other Health Care Interventions: Sage Publications, Inc; 1996.
13. Lorig KR, Sobel DS, Ritter PL, Laurent D, Hobbs M. Effect of a self-management program on patients with chronic disease. Eff Clin Pract 2001;4(6):256-262. [Medline: 11769298]
14. Sullivan MD, LaCroix AZ, Russo J, Katon WJ. Self-efficacy and self-reported functional status in coronary heart disease: a six-month prospective study. Psychosom Med 1998;60(4):473-478. [doi: 10.1097/00006842-199807000-00014] [Medline: 9710293]
15. Fors A, Ulin K, Cliffordson C, Ekman I, Brink E. The Cardiac Self-Efficacy Scale, a useful tool with potential to evaluate person-centred care. Eur J Cardiovasc Nurs 2015 Dec;14(6):536-543. [doi: 10.1177/1474515114548622] [Medline: 25149667]
16. Ea EE, Colbert A, Turk M, Dickson VV. Self-care among Filipinos in the United States who have hypertension. Appl Nurs Res 2018 Feb;39:71-76. [doi: 10.1016/j.apnr.2017.11.002] [Medline: 29422180]
17. Tan F, Oka P, Dambha-Miller H, Tan NC. The association between self-efficacy and self-care in essential hypertension: a systematic review. BMC Fam Pract 2021 Feb 22;22(1):44. [doi: 10.1186/s12875-021-01391-2] [Medline: 33618661]
18. Garland M, Wilbur J, Fogg L, Halloway S, Braun L, Miller A. Self-efficacy, outcome expectations, group social support, and adherence to physical activity in African American women. Nurs Res 2021;70(4):239-247. [doi: 10.1097/NNR.0000000000000516] [Medline: 33870956]
19. Kang A, Dulin A, Risica PM. Relationship between adherence to diet and physical activity guidelines and self-efficacy among black women with high blood pressure. J Health Psychol 2022 Mar;27(3):663-673. [doi: 10.1177/1359105320967105] [Medline: 33081506]
20. Warren-Findlow J, Seymour RB. Prevalence rates of hypertension self-care activities among African Americans. J Natl Med Assoc 2011 Jun;103(6):503-512. [doi: 10.1016/s0027-9684(15)30365-5] [Medline: 21830634]
21. Jackson L, Leclerc J, Erskine Y, Linden W. Getting the most out of cardiac rehabilitation: a review of referral and adherence predictors. Heart 2005 Jan;91(1):10-14. [doi: 10.1136/hrt.2004.045559] [Medline: 15604322]
22. Banik A, Schwarzer R, Knoll N, Czekierda K, Luszczynska A. Self-efficacy and quality of life among people with cardiovascular diseases: a meta-analysis. Rehabil Psychol 2018 May;63(2):295-312. [doi: 10.1037/rep0000199] [Medline: 29878834]
23. Lorig KR, Ritter PL, González VM. Hispanic chronic disease self-management: a randomized community-based outcome trial. Nurs Res 2003;52(6):361-369. [doi: 10.1097/00006199-200311000-00003] [Medline: 14639082]
24. DeWalt DA, Pignone M, Malone R, et al. Development and pilot testing of a disease management program for low literacy patients with heart failure. Patient Educ Couns 2004 Oct;55(1):78-86. [doi: 10.1016/j.pec.2003.06.002] [Medline: 15476993]
25. Bagheri H, Shakeri S, Nazari AM, et al. Effectiveness of nurse-led counselling and education on self-efficacy of patients with acute coronary syndrome: a randomized controlled trial. Nurs Open 2022 Jan;9(1):775-784. [doi: 10.1002/nop2.1129] [Medline: 34766453]
26. McKenzie KM, Park LK, Lenze EJ, et al. A prospective cohort study of the impact of outpatient intensive cardiac rehabilitation on depression and cardiac self-efficacy. Am Heart J Plus 2022 Jan;13:100100. [doi: 10.1016/j.ahjo.2022.100100] [Medline: 36407054]
27. Kang Y, Yang IS, Kim N. Correlates of health behaviors in patients with coronary artery disease. Asian Nurs Res (Korean Soc Nurs Sci) 2010 Mar;4(1):45-55. [doi: 10.1016/S1976-1317(10)60005-9] [Medline: 25030792]
28. Sarkar U, Ali S, Whooley MA. Self-efficacy as a marker of cardiac function and predictor of heart failure hospitalization and mortality in patients with stable coronary heart disease: findings from the Heart and Soul Study. Health Psychol 2009 Mar;28(2):166-173. [doi: 10.1037/a0013146] [Medline: 19290708]
29. Mahajan S, Lu Y, Spatz ES, Nasir K, Krumholz HM. Trends and predictors of use of digital health technology in the United States. Am J Med 2021 Jan;134(1):129-134. [doi: 10.1016/j.amjmed.2020.06.033] [Medline: 32717188]
30. Wongvibulsin S, Habeos EE, Huynh PP, et al. Digital health interventions for cardiac rehabilitation: systematic literature review. J Med Internet Res 2021 Feb 8;23(2):e18773. [doi: 10.2196/18773] [Medline: 33555259]
31. Zwack CC, Haghani M, Hollings M, et al. The evolution of digital health technologies in cardiovascular disease research. npj Digit Med 2023 Jan 3;6(1):1-11. [doi: 10.1038/s41746-022-00734-2]
32. Frederix I, Caiani EG, Dendale P, et al. ESC e-Cardiology Working Group position paper: overcoming challenges in digital health implementation in cardiovascular medicine. Eur J Prev Cardiol 2019 Jul;26(11):1166-1177. [doi: 10.1177/2047487319832394] [Medline: 30917695]
33. Gray R, Indraratna P, Lovell N, Ooi SY. Digital health technology in the prevention of heart failure and coronary artery disease. Cardiovasc Digit Health J 2022 Dec;3(6 Suppl):S9-S16. [doi: 10.1016/j.cvdhj.2022.09.002] [Medline: 36589760]
34. Devi R, Powell J, Singh S. A web-based program improves physical activity outcomes in a primary care angina population: randomized controlled trial. J Med Internet Res 2014 Sep 12;16(9):e186. [doi: 10.2196/jmir.3340] [Medline: 25217464]

XSL•FO
RenderX

35. Newby K, Teah G, Cooke R, et al. Do automated digital health behaviour change interventions have a positive effect on self-efficacy? A systematic review and meta-analysis. Health Psychol Rev 2021 Mar;15(1):140-158. [doi: 10.1080/17437199.2019.1705873] [Medline: 31847702]
36. Lockwood KG, Kulkarni PR, Paruthi J, et al. Evaluating a new digital app-based program for heart health: feasibility and acceptability pilot study. JMIR Form Res 2024 May 24;8:e50446. [doi: 10.2196/50446] [Medline: 38787598]
37. Branch OH, Rikhy M, Auster-Gussman LA, Lockwood KG, Graham SA. Relationships between blood pressure reduction, weight loss, and engagement in a digital app-based hypertension care program: observational study. JMIR Form Res 2022 Oct 27;6(10):e38215. [doi: 10.2196/38215] [Medline: 36301618]
38. Graham SA, Pitter V, Hori JH, Stein N, Branch OH. Weight loss in a digital app-based diabetes prevention program powered by artificial intelligence. Digit Health 2022;8:20552076221130619. [doi: 10.1177/20552076221130619] [Medline: 36238752]
39. Lockwood KG, Pitter V, Kulkarni PR, Graham SA, Auster-Gussman LA, Branch OH. Predictors of program interest in a digital health pilot study for heart health. PLOS Digit Health 2023 Jul 31;2(7):e0000303. [doi: 10.1371/journal.pdig.0000303]
40. McGorrian C, Yusuf S, Islam S, et al. Estimating modifiable coronary heart disease risk in multiple regions of the world: the INTERHEART Modifiable Risk Score. Eur Heart J 2011 Mar;32(5):581-589. [doi: 10.1093/eurheartj/ehq448] [Medline: 21177699]
41. Yusuf S, Rangarajan S, Teo K, et al. Cardiovascular risk and events in 17 low-, middle-, and high-income countries. N Engl J Med 2014 Aug 28;371(9):818-827. [doi: 10.1056/NEJMoa1311890] [Medline: 25162888]
42. Ross ES, Sakakibara BM, Mackay MH, et al. The use of SMS text messaging to improve the hospital-to-community transition in patients with acute coronary syndrome (Txt2Prevent): results from a pilot randomized controlled trial. JMIR mHealth uHealth 2021 May 14;9(5):e24530. [doi: 10.2196/24530] [Medline: 33988519]
43. Maddison R, Pfaeffli L, Whittaker R, et al. A mobile phone intervention increases physical activity in people with cardiovascular disease: results from the HEART randomized controlled trial. Eur J Prev Cardiolog 2015 Jun;22(6):701-709. [doi: 10.1177/2047487314535076]
44. Kostka T, Jachimowicz V. Relationship of quality of life to dispositional optimism, health locus of control and self-efficacy in older subjects living in different environments. Qual Life Res 2010 Apr;19(3):351-361. [doi: 10.1007/s11136-010-9601-0] [Medline: 20146007]
45. Anderson ES, Winett RA, Wojcik JR, Williams DM. Social cognitive mediators of change in a group randomized nutrition and physical activity intervention: social support, self-efficacy, outcome expectations and self-regulation in the guide-to-health trial. J Health Psychol 2010 Jan;15(1):21-32. [doi: 10.1177/1359105309342297] [Medline: 20064881]
46. Orth U, Trzesniewski KH, Robins RW. Self-esteem development from young adulthood to old age: a cohort-sequential longitudinal study. J Pers Soc Psychol 2010 Apr;98(4):645-658. [doi: 10.1037/a0018769] [Medline: 20307135]
47. Orth U, Erol RY, Luciano EC. Development of self-esteem from age 4 to 94 years: a meta-analysis of longitudinal studies. Psychol Bull 2018 Oct;144(10):1045-1080. [doi: 10.1037/bul0000161] [Medline: 30010349]
48. Kang Y, Yang IS. Cardiac self-efficacy and its predictors in patients with coronary artery diseases. J Clin Nurs 2013 Sep;22(17-18):2465-2473. [doi: 10.1111/jocn.12142] [Medline: 23441807]
49. Ovaskainen ML, Tapanainen H, Laatikainen T, Männistö S, Heinonen H, Vartiainen E. Perceived health-related self-efficacy associated with BMI in adults in a population-based survey. Scand J Public Health 2015 Mar;43(2):197-203. [doi: 10.1177/1403494814566263]
50. Kiviruusu O, Konttinen H, Huurre T, Aro H, Marttunen M, Haukkala A. Self-esteem and body mass index from adolescence to mid-adulthood. a 26-year follow-up. IntJ Behav Med 2016 Jun;23(3):355-363. [doi: 10.1007/s12529-015-9529-4]
51. Shrestha R, Rawal L, Bajracharya R, Ghimire A. Predictors of cardiac self-efficacy among patients diagnosed with coronary artery disease in tertiary hospitals in Nepal. J Public Health Res 2020 Oct 14;9(4):1787. [doi: 10.4081/jphr.2020.1787] [Medline: 33437751]
52. Gelles-Watnick R. Rates of smartphone ownership, broadband subscription vary across groups, including by household income and education. Pew Research Center. 2024. URL: https://www.pewresearch.org/internet/2024/01/31/americans-use-of-mobile-technology-and-home-broadband/ [accessed 2025-04-04]

Abbreviations

AI: artificial intelligence
CSE: cardiac self-efficacy
CVD: cardiovascular disease
HIPAA: Health Insurance Portability and Accountability Act
HITRUST: Health Information Trust Alliance
SOC 2: System and Organization Controls 2

XSL•FO
RenderX

Edited by A Mavragani; submitted 17.05.24; peer-reviewed by A Woodward, K Wac; revised version received 28.02.25; accepted 02.03.25; published 24.04.25.

Please cite as:
Lockwood KG, Kulkarni PR, Branch OH, Graham SA
Cardiac Self-Efficacy Improvement in a Digital Heart Health Program: Secondary Analysis From a Feasibility and Acceptability Pilot Study
JMIR Form Res 2025;9:e60676
URL: https://formative.jmir.org/2025/1/e60676
doi:10.2196/60676

© Kimberly G Lockwood, Priya R Kulkarni, OraLee H Branch, Sarah A Graham. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Patient Acceptability and Technical Reliability of Wearable Devices Used for Monitoring People With Parkinson Disease: Survey Study

Tasmin Alanna Rookes[1], BSc, MSc; Amit Batla[2], MA, MBBS; Megan Armstrong[1], PhD; Gareth Ambler[3], BSc, PhD; Kate Walters[1], MBBS, MSc, PhD; Anette Schrag[2], PhD

[1]Research Department of Primary Care and Population Health, University College London, London, United Kingdom

[2]Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, University College London, London, United Kingdom

[3]Department of Statistical Science, University College London, London, United Kingdom

Corresponding Author:
Anette Schrag, PhD
Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, University College London, London, United Kingdom

Abstract

Background: Parkinson disease is a progressive neurodegenerative disorder with complex motor and nonmotor symptoms. To assess these, clinical assessments are completed, providing a snapshot of a person's experience. Monitoring Parkinson disease using wearable devices can provide continuous and objective data and capture information on movement patterns in daily life.

Objective: The aim of the study is to assess patient acceptability and technical reliability of 2 wearable devices used in clinical trials (ActivInsights and Axivity AX3).

Methods: Participants in a feasibility study testing a self-management toolkit (PD-Care) optionally wore a wearable device for 1 week, providing feedback through an open- and closed-question survey conducted over the telephone about the acceptability of wearing the device. The closed questions used a Likert scale from 1 to 5 (with 1=strongly agree and 5=strongly disagree) asking whether (1) the device was comfortable to wear, (2) the device was easy to put on, (3) the device was easy to wear, (4) the device was embarrassing to wear, and (5) if they were happy to wear the device for longer than 7 days. Differences in acceptability between devices were analyzed using Mann-Whitney *U* tests and Wilcoxon matched pairs signed rank tests. These were followed by open-ended questions asking (1) How did you find wearing the device? (2) How did you find putting the device on? (3) Did you take it off and why? (4) What was your overall impression? (5) Did you prefer the wrist- or trunk-worn device and why (Axivity AX3 only)?

Results: A total of 22 of 32 (69%) participants offered the device agreed to wear it. There were no significant differences in the demographic characteristics between those monitored and those who chose not to be. Acceptance with both devices was generally good. The ActivInsights device was more acceptable than the wrist- and trunk-worn Axivity AX3 devices, as more participants found it to be comfortable (n=15, 100% vs n=5, 71%; *P*=.02 and n=4, 57%; *P*=.004, respectively), easy to wear (n=15, 100% vs n=6, 86%; *P*=.048 and n=3, 43%; *P*=.004, respectively) and would wear for more than 7 days (n=13, 87% vs n=4, 57%; *P*=.02 and n=1, 14%; *P*<.001, respectively). The trunk-worn Axivity AX3 device had the lowest acceptance rates, but there were no statistical differences in acceptability between the wrist- and trunk-worn Axivity AX3 devices (all *P*>.05). There were issues with battery life and recording errors in 3 of 14 (21%) Axivity AX3 devices and upload failures in 3 of 15 (20%) ActivInsights devices.

Conclusions: Acceptability of wearables for monitoring Parkinson was satisfactory, especially when wrist-worn, although a few participants experienced difficulties in correct use, and there were some errors with the data upload.

Trial Registration: ISRCTN Registry ISRCTN92831552; https://www.isrctn.com/ISRCTN92831552

(JMIR Form Res 2025;9:e63704) doi:10.2196/63704

KEYWORDS

Parkinson disease; feasibility; remote monitoring; Parkinson; acceptability; reliability; wearable devices; wearable; self-management; quantitative assessments; quantitative; qualitative; monitoring; patient

Introduction

Parkinson disease (PD) is a neurodegenerative chronic disease, affecting around 145,000 people in the United Kingdom in 2018, equivalent to approximately 1 adult in every 350 [1]. The currently accepted monitoring of PD symptoms in clinical practice and trials is based on interviews and validated clinical scales and questionnaires [2]. However, the use of patient- and

XSL•FO
RenderX

clinician-rated scales is resource-intensive, and the results limited by subjectivity and inaccuracy and are either one-off assessments or dependent on the patient's recollection [3]. Continuous and accurate assessment of PD features, in both clinical practice and trials, would provide more relevant information and overcome the limitations of clinical scales and questionnaires through real-world data. With the increasing burden of care and limited medical resources, new assessment and monitoring approaches have the potential to improve clinical care and patient outcomes [4]. There are now a range of body-worn sensors using accelerometers or gyroscope devices (henceforth called wearable devices) that have been developed and validated to monitor parkinsonian clinical features, such as motor fluctuations, dyskinesia, tremor, bradykinesia, freezing of gait, or gait disturbances [2]. However, relatively little evidence exists on their use, acceptability, and feasibility in clinical practice and trials with patients with PD. In addition, the physical, technical, or practical limitations on prolonged use for real-time recording, uploading, and analysis of movement data are not clear [5].

The National Institute of Health and Care Excellence has conditionally recommended some wearable devices for remote monitoring of PD. However, the exact role, indication, and type of device is unclear, and evidence on the use of wearable devices is limited [6]. The cost, burden on patients and services, and effectiveness to improve clinical outcomes have not yet been assessed. National Institute of Health and Care Excellence has therefore recommended the collection of real-world evidence on devices that monitor people with PD [7] to enable the clinical and cost-effectiveness of the technologies to be fully assessed, ahead of further implementation into clinical practice. In practice, wearable devices are relatively rarely used due to several factors, such as unfamiliarity, unclear indications, cost, acceptability, and technical issues. Furthermore, although digital measures are considered to have high potential for future use to more accurately detect meaningful change in clinical trials, little data exist on their use and acceptability in clinical trials.

Data collection regarding utility and acceptability has largely been informal and focused on wearable devices in general [5]. A study in Finland reported that 88.9% (32/36) of patients with PD using a wearable device thought it was very easy or easy [8]. In the United States, the recent WATCH-PD study [9] explored the acceptability of wearables in people with PD and control participants over a 2-year window and found that participants with PD had generally positive views of the wearables, which were comparable to healthy controls. Conducting qualitative interviews with people with PD who had used multimodal sensors, including a wearable device, the wrist-worn wearable was the least acceptable, in comparison to cameras and passive ambient sensors [10]. Another small study explored perceptions of wearing devices in the home versus in public and found no differences in acceptability between these settings and acceptance of long-term symptom monitoring through wrist-worn devices [11]. However, these specific studies are likely to have recruited participants willing to wear a device, so there is a lack of real-world evidence from patients with PD wearing these devices in home and social settings. Eliciting opinions from patients on wearing such devices in real-life settings may provide important information to facilitate the implementation of wearables into clinical practice.

Here, we have used the opportunity to explore the views of participants taking part in a feasibility study from a large randomized controlled trial of a self-management tool for people with PD in the National Health Service (NHS) to obtain such information in a real-life setting. The objective of this study was to explore the acceptability and technical reliability of 2 wearable sensors that are commonly used by people with PD.

Methods

Overview

This study is reported in line with the CROSS checklist (Consensus-Based Checklist for Reporting of Survey Studies) [12] (Table S1 in Multimedia Appendix 1). We used a mixed methods survey design to obtain information from survey questions, both quantitative and semiqualitative, providing complementary information, within the context of a clinical feasibility trial in an NHS setting [13].

Study Design and Participant Recruitment

Participants in the PD-Care program feasibility study were offered to participate in this wearable substudy. The primary aim of the PD-Care program is to develop and trial a supported self-management tool for people with PD, the full methods of which have been described previously [13]. The study emphasizes inclusive participation with broad inclusion criteria, including no age limit, inclusion of participants with cognitive impairment as long as they have the capacity to consent, and in any stage of PD. Remote delivery via video or telephone calls increased accessibility to study participation. Following the development of the intervention, a planned feasibility study was conducted before the randomized controlled trial (ISRCTN92831552). The feasibility study included 35 participants with assessments at baseline and post-intervention at 3 months. Participants were recruited through general practice mailouts, secondary care neurology teams, and the neurology research register.

After the 3-month follow-up assessments, all participants were offered to participate in a substudy, wearing 1 of 2 devices for a one-off 7-day period. After wearing and returning the device, participants were asked to provide feedback on the acceptability of the device via a telephone survey. The researcher entered the participant's responses directly into the database while on the phone call. In addition, when downloading the data from the devices, the researcher recorded any errors that occurred. This substudy received input from the PD-Care patient and public engagement group, including where the device should be worn, how long it should be worn for, and which acceptability questions should be asked in the survey.

Devices

Two different devices, ActivInsights [14] and Axivity AX3 [15] were used. The choice of the 2 devices was guided by pragmatic reasons for the purpose of a larger, publicly funded trial in the NHS. Both devices have been used in previous clinical research in patients with PD [14,15]. Both devices were worn at the wrist.

XSL•FO
RenderX

In addition, as recommended by the patient and public advisory group, a sensor worn on the trunk was also tested to capture different clinical features. We therefore used 2 Axivity AX3 devices worn simultaneously, one on the wrist and the other on the trunk, whereas the ActivInsights device was worn on the wrist only.

The ActivInsights device provided output data on sleep, sedentary time, active time, and time spent exercising. The Axivity AX3 device output provided raw data on movement, recorded as a continuous wave accelerometer format, which was converted to a CSV file for analysis.

Data Collection

The wearable devices and instructions were sent to participants in the post, with a request to return in the post after 7 days using a stamped addressed envelope. Seven days was chosen due to the limit on the battery life of the Axivity AX3 device while still collecting enough data to have an average over several days. Participants were allocated to a device depending on what was available once they had completed study follow-up assessments. We had 8 wearable devices of each ActivInsights and Axivity AX3, but as 2 devices were needed for each participant wearing Axivity AX3, fewer participants used this device. The date to start and stop recording for participants was preprogrammed into the device before posting out, and participants were asked to start and stop wearing the device on a specific date and time. Both devices recorded continuously, and we asked participants to inform us of the 7-day time period during which they wore the device, depending on when it arrived. No feedback from the devices was provided to participants. Data upload was undertaken by the researchers upon the return of the devices.

Outcome Measures

Acceptability

The overall rate of participation in the substudy was recorded. After wearing and returning the device, participants were contacted via telephone to provide feedback about the acceptability of using their wearable device through a survey. We captured quantitative data to enable inferential statistical tests to be conducted to compare between devices as well as open-ended questions to ensure that the nuance of people's individual experiences was understood.

The open-ended questions were (1) How did you find wearing the device? (2) How did you find putting the device on? (3) Did you take it off and why? (4) What was your overall impression? (5) Did you prefer the wrist- or trunk-worn device (Axivity AX3 only) and why? The quantitative assessments included questions using a Likert scale from 1 to 5 (1=strongly agree and 5=strongly disagree) whether (1) the device was comfortable to wear, (2) the device was easy to put on, (3) the device was easy to wear, (4) the device was embarrassing to wear, and (5) if they were happy to wear the device for longer than 7 days.

The survey was designed by the research team to cover the core components of wearable acceptance while being short and quick to administer. The survey was not based on any specific framework or guideline. A pragmatic approach with short survey questions was adopted to minimize participant burden, to reduce the risk of attrition in the context of the overall feasibility study, which involved collecting a wide range of outcome measures.

Technical Reliability

We assessed whether issues of battery shortage, recording failure, or upload difficulties impacted on the availability of usable data.

Data Analysis

To ensure that the participants in the sample did not differ from those who chose to wear an activity monitor, the Fisher exact test and 2-tailed *t* test comparisons were conducted on demographic characteristics. To explore differences in acceptability between the different monitors, Mann-Whitney *U* tests and Wilcoxon matched pairs signed rank tests were conducted. All analyses were conducted in Stata (version 17; Stata Corp 2021). Open-ended question answers were written out, as participants provided their answers verbally, and responses were synthesized through content analysis to identify common concepts and views between participants.

Ethical Considerations

Participants were taking part in the feasibility study of the PD-Care trial (ISRCTN92831552) and provided optional consent to take part in this wearable substudy. This substudy received a favorable opinion by the London Queen Square Research Ethics Committee and Health Research Authority approval (18/LO/1470) on October 31, 2018, as part of the feasibility study application. All participant's data were pseudoanonymized through unique study ID numbers, and all study documentation, including wearable data and survey responses, were stored against this ID. Participants received a US $26 voucher for completing baseline assessments and US $13 voucher for completing 3-month follow-up assessments, as part of the feasibility study. No additional compensation was provided for taking part in this wearable substudy.

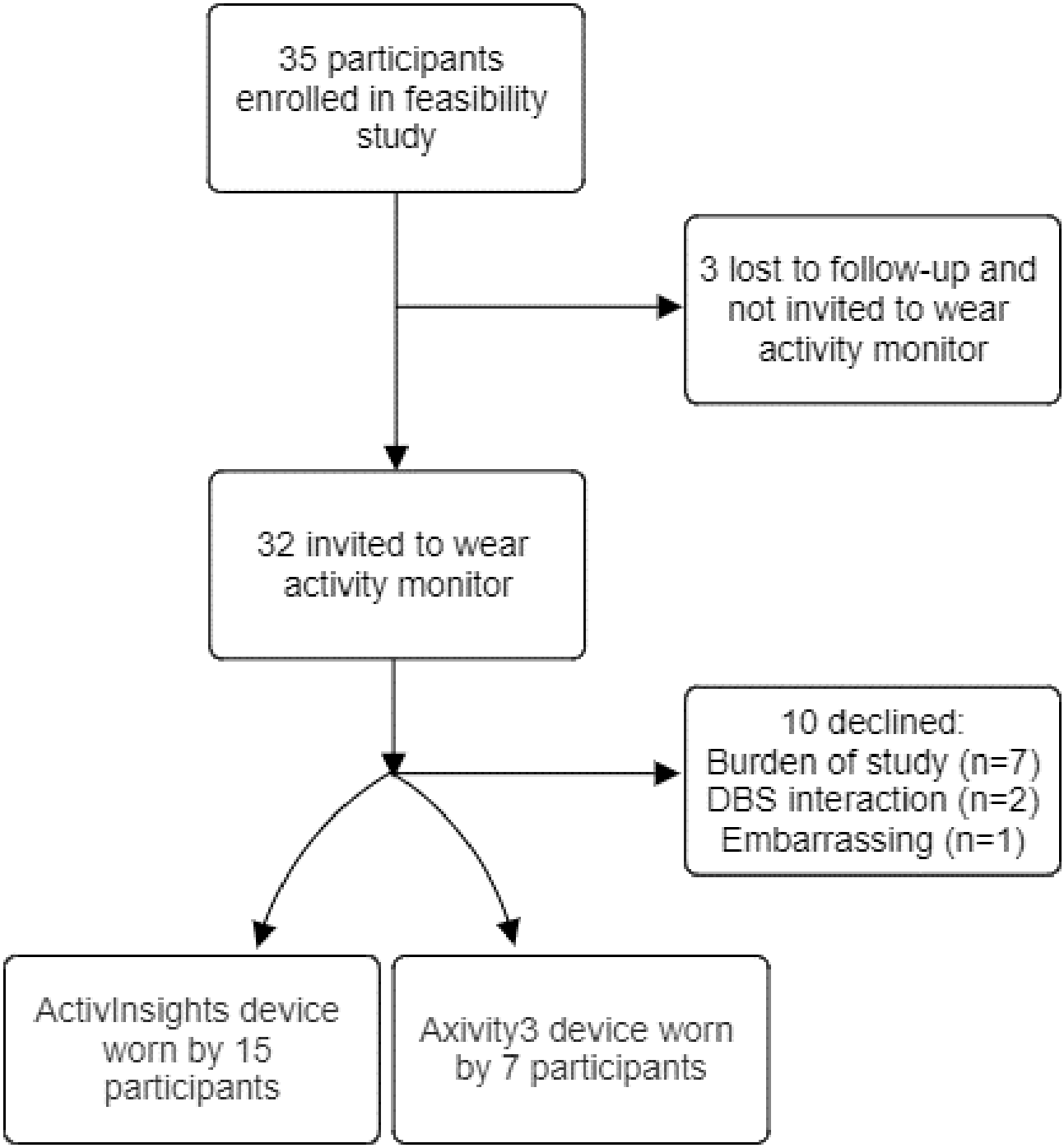
Results

Participants

Between October 2020 and February 2021, 35 participants with a clinical diagnosis of PD and living in a community setting were recruited to the PD-Care program from several secondary care sites in the United Kingdom. Of these, 1 moved permanently to a care home and was no longer eligible, and 2 were lost to follow-up in the overall study. The remaining 32 were offered participation in the device substudy. The average age of the 32 participants invited to wear a wearable device was 68.2 (SD 11.3) years, 11 (34%) were female, and 23 (72%) were White British. Of these, 22 (69%) agreed to wear a device for 7 days. Reasons for not participating were participants feeling burdened from the feasibility study and did not want to provide more data (n=7), participants had a deep brain stimulation device fitted and were uncertain about potential interactions with the wearable (n=2), and finding it embarrassing to wear (n=1) (Figure 1).

XSL•FO
RenderX

Figure 1. Flowchart of participants in the study. DBS: deep brain stimulation.



There were no statistically significant differences in age, sex, or disease duration or severity between participants and nonparticipants (Table S2 in Multimedia Appendix 1). In total, 15 of the 22 participants were allocated the ActivInsights device (wrist only) and 7 were asked to wear the Axivity AX3 device (wrist and trunk). There were no statistically significant differences in the characteristics of those who wore the Axivity AX3 and those who wore the ActivInsights devices (Table S3 in Multimedia Appendix 1).

Patient Acceptability

ActivInsights

Feedback to open-ended questions suggested that the device was easy to use, and participants quickly forgot they were wearing it (n=11). Some people commented that the strap was difficult to put on and needed help (n=5), and others suggested that it would have been nice if the device had a screen that gave them some feedback about their activity levels (n=3). Quantitative feedback suggested that overall the device was comfortable, easy to wear, and not embarrassing, and participants would be happy to wear it for longer than 7 days (Tables 1 and 2 and Multimedia Appendix 2).

XSL•FO
RenderX

Table . People with Parkinson disease who responded in a positive way to the survey questions about their experience of wearing different wearable devices in the home for 1 week.

Device	It was comfortable, n (%)	It was easy to put on, n (%)	It was easy to wear, n (%)	It was embarrassing to wear[a], n (%)	I would wear it for more than 7 days, n (%)
Axivity AX3 wrist (n=7)	5 (71)	5 (71)	6 (86)	7 (100)	4 (57)
Axivity AX3 trunk (n=7)	4 (57)	2 (29)	3 (43)	7 (100)	1 (14)
ActivInsights wrist (n=15)	15 (100)	10 (67)	15 (100)	15 (100)	13 (87)

[a]Percentage inversed.

Table . Mann-Whitney *U* tests comparing the acceptability responses between ActivInsights and both Axivity AX3 wrist and trunk and Wilcoxon matched pairs signed rank test comparing responses between Axivity AX3 wrist and trunk.

Device comparison	It was comfortable, *P* value	It was easy to put on, *P* value	It was easy to wear, *P* value	It was embarrassing to wear, *P* value	I would wear it for more than 7 days, *P* value
ActivInsights versus Axivity AX3 wrist	.02	.72	.048	N/A[a]	.02
ActivInsights versus Axivity AX3 trunk	.004	.38	.004	N/A	<.001
Axivity AX3 Wrist versus Axivity AX3 trunk	.50	.50	.25	N/A	.25

[a]N/A: not applicable.

Axivity AX3

All participants reported that they preferred the wrist-worn device to the trunk-worn device. Problems with the trunk-worn device centered around the tape not holding the device in place and having to be reapplied (n=7). This led to 3 of 7 participants not wearing the trunk-worn device but still provided feedback on both devices. The wrist-worn device was acceptable to all, with only a few comments about it being bulky and uncomfortable (n=3) and 1 participant wanted feedback on what the device was recording.

Quantitative feedback suggested that, on average, participants were neutral to the trunk-worn device being comfortable, easy to put on, and easy to wear but disagreed with being happy to wear it for longer than 7 days. For the wrist-worn wearable device, participants on average agreed that it was easy to put on and wear and were neutral about its comfort and whether they would wear it for more than 7 days (see Table 1 and Multimedia Appendix 2).

Patient Acceptability—Comparison Between Groups

Using Mann-Whitney *U* tests to compare the acceptability of the devices, overall, the ActivInsights device was more acceptable than both the Axivity AX3 wrist- and trunk-worn devices. The ActivInsights device was rated as more comfortable than the wrist-worn and trunk-worn Axivity AX3 devices (n=15, 100% vs n=5, 71%; *P*=.02 and n=4, 57%; *P*=.004, respectively), rated as easier to wear (n=15, 100% vs n=6, 86%; *P*=.048 and n=3, 43%; *P*=.004, respectively), and rated as more suitable to wear more than 7 days (n=13, 87% vs n=4, 57%; *P*=0.02 and n=1, 14%; *P*<.001, respectively). Using Wilcoxon matched pair signed rank tests, there were no significant differences between the Axivity AX3 trunk-worn and wrist-worn device, but the numbers were small. There were no significant differences between all 3 devices for the ease to put them on (Table 2) and for embarrassing to wear, and all participants responded positively, stating the devices were not embarrassing.

Technical Reliability

ActivInsights

There were no issues reported regarding battery shortage or recording failure. There were data upload errors with 3 (20%) of the devices, whereby the system failed when trying to upload the data, and when corrected, the data had been wiped from the device. Usable data were therefore available in 12 of 15 (80%) participants.

Axivity AX3

Of the 11 devices worn by 7 participants (7 wrist and 4 trunk), 6 (3 participants) were worn for shorter durations, resulting in missing data, for a median of 27 (IQR 8 - 48) hours, and this was due to hospitalization (n=1) and participant forgetting (n=2). In 3 other devices, the device ran out of battery during the data collection period. Despite the errors, some usable data was available for all participants with this device.

XSL•FO
RenderX

Discussion

Principal Findings

We assessed the acceptability and technical feasibility of 2 wearable devices used in trials with people with PD, which showed that wearable devices are acceptable to people with PD and feasible to administer and collect data with. Some data errors were found due to battery life, participant utility, and upload issues. Although several studies using wearable devices have reported validity for recording human movement in a free-living environment [16], including in people with PD [2,17,18], successful clinical adoption in monitoring movement in PD depends on acceptability and technical reliability [5,17]. To date, there is no fully validated system to monitor clinical features or activities in PD [5]. The bulk of data on technical reliability with the use of individual wearable devices comes from the studies led or supported by manufacturers of these devices [16]. This is the biggest study to date that reports the feedback of patients with PD from wearing 2 different monitoring devices using open and closed questions.

Patient Acceptability

Feedback from patients suggests that monitoring using both ActivInsights and Axivity AX3 is mostly acceptable, although there is a proportion of patients who do not wish to use a monitoring device due to burden and embarrassment. The burden was due to participants having to complete clinical assessments as part of the feasibility study. If the participants had had the option of clinical assessments or activity monitors, then uptake may have been higher. There were no demographic characteristics that differed between the participants who chose to be monitored and those who chose not to be monitored, including sex, age, and deprivation level. Largely, our findings are in line with those from a US population with PD, using a different device, suggesting ease of use and willingness to wear the device over a longer period when located on the wrist [9]. In addition, these findings reflect acceptability in other UK patient groups, such as wrist-worn alcohol sensors in those who are alcohol dependent, where people were positive about the devices, despite the perceptions of health care professionals that service users may have challenges using them [19].

We did not specifically examine motivational factors, but another study has reported on the motivations for and barriers against monitoring PD symptoms among patients with PD [20]. Key motivations include the wish to discuss findings with health care providers, obtain insight into the effect of medication and other treatments, and follow the progression of the disease. Key barriers were not wanting to focus too much on having PD, symptoms being relatively stable, and lacking an easy-to-use tool [17]. As all participants in our study were monitored as part of a research intervention, these motivating factors mentioned may have been less directly relevant. As no clinical feedback or reports were provided to participants, and the purpose was to supplement clinical assessments in a trial, the setting and motivation will differ from those who would consider monitoring for a clinical purpose. Nevertheless, 4 (18%) of the participants expressed the wish to know what was being monitored.

Social embarrassment and the feeling that wearing a device for monitoring signaled a person as “old” have been highlighted as major factors affecting the acceptability of body-worn sensors [11]. Similarly, the use of a wearable device was embarrassing in some of our study participants, resulting in them declining participation in this substudy. In another study using Axivity AX3 on each wrist, the participants wore the device for 4 hours in a research facility and then for 1 week at home. Overall, the study reported that monitoring with body-worn sensors was acceptable to patients with PD and that most participants were willing to wear the sensor both at home and in public [11]. Concerns regarding safety particularly with deep brain stimulation could also be an additional barrier that could be resolved with better communication and technical support.

In addition, overall, the wrist-worn devices were more acceptable than the trunk-worn devices, as many participants found the trunk-worn device to be uncomfortable and difficult to keep on. However, there may also be issues with wrist-worn devices. A previous study using Axivity AX3 on each wrist reported that, for prolonged wearing, participants were less likely to agree that the sensors were comfortable to wear [11]. Responses to open-ended questions revealed that the main source of sensor discomfort related to the strap [11]. Though the acceptance of the strap on the 7-day monitoring was not a problem in our cohort, 23% (n=5) needed help with the strap to put the device on. Ideally, if the wearable device is to be worn by a patient with PD, they should be able to do this independently despite poor dexterity.

Technical Reliability

The charging and preparation of the devices takes time and effort, with human errors leading to incomplete charging of the devices, and some participants forgetting to turn on the devices. The battery life on a single charge may not be long enough for long-term monitoring. There were also upload errors, with the upload failing in one-fifth of the participants using the ActivInsights device. Almost all wearable devices rely on a local computer interface, a stable internet connection for upload, and often an external storage system before being analyzed [2]. The upload to a cloud server relies on a good internet connection and can be a limiting factor in some settings [4]. Such technical problems not only add to costs and time pressure but can also lead to further anxiety for participants who might already have nonmotor symptoms of anxiety. More data are needed on failure rates with individual devices. On most occasions, device failure was identified after the recording had been completed, and the wearable devices had been sent back. Ideally, troubleshooting tools or a live screen display, as 4 of our participants suggested, could be more acceptable and help avoid recording failures. Furthermore, additional training of research staff and patients to ensure data upload could be helpful to overcome this issue.

Study Limitations and Future Research

The duration of monitoring used in this study was short, and we had a small sample size. Studies with larger sample sizes and longer monitoring periods are needed. As part of the PD-Care program of research, a large randomized controlled trial is now underway including a substudy using wearable devices in people with PD. These data will be used to further

XSL•FO
RenderX

validate the use of wearable devices as a clinical measure of PD symptoms as well as confirm the acceptability findings found in this feasibility study. Researchers should also explore differences in acceptability in people from underserved groups to ensure that the implementation of these devices does not widen health inequalities.

Conclusions

The study highlights some issues that can be addressed by device manufacturers such as difficulty with strapping single-handedly in someone with limited dexterity, help in uploading data, and clearer instructions. We found better acceptability of the ActivInsights device by our participant cohort, but there was a higher rate of usable recording when using Axivity AX3. The prolonged use of devices worn on the trunk was less acceptable. More real-world experience, including patient feedback and multidisciplinary coordinated effort involving all key stakeholders, is critical before the widespread adoption of wearable devices for monitoring PD in clinical trials.

Acknowledgments

The wider PD-Care team and collaborators involved in the development of the project. The authors thank the participants for their ongoing support of our research. The study received funding from National Institute for Health and Care Research program (grant RP-PG-1016-20001). Generative artificial intelligence was not used in any portion of the manuscript writing.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The CROSS (Consensus-Based Checklist for Reporting of Survey Studies) checklist and demographic characteristic comparisons.
[DOCX File, 27 KB - formative_v9i1e63704_app1.docx]

Multimedia Appendix 2
Graph showing responses to each question, for each device.
[DOCX File, 27 KB - formative_v9i1e63704_app2.docx]

References

1. The incidence and prevalence of Parkinson's in the UK: results from the Clinical Practice Research Datalink Summary report. Parkinson's UK. 2018. URL: https://www.parkinsons.org.uk/sites/default/files/2018-01/CS2960%20Incidence%20and%20prevalence%20report%20branding%20summary%20report.pdf [accessed 2025-03-10]
2. Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Mov Disord 2008 Nov 15;23(15):2129-2170. [doi: 10.1002/mds.22340] [Medline: 19025984]
3. AlMahadin G, Lotfi A, Zysk E, Siena FL, Carthy MM, Breedon P. Parkinson's disease: current assessment methods and wearable devices for evaluation of movement disorder motor symptoms—a patient and healthcare professional perspective. BMC Neurol 2020 Nov 18;20(1):419. [doi: 10.1186/s12883-020-01996-7] [Medline: 33208135]
4. Moreau C, Rouaud T, Grabli D, et al. Overview on wearable sensors for the management of Parkinson's disease. NPJ Parkinsons Dis 2023 Nov 2;9(1):153. [doi: 10.1038/s41531-023-00585-y] [Medline: 37919332]
5. Del Din S, Godfrey A, Mazzà C, Lord S, Rochester L. Free-living monitoring of Parkinson's disease: lessons from the field. Mov Disord 2016 Sep;31(9):1293-1313. [doi: 10.1002/mds.26718] [Medline: 27452964]
6. Laar A, Silva de Lima AL, Maas BR, Bloem BR, de Vries NM. Successful implementation of technology in the management of Parkinson's disease: barriers and facilitators. Clin Park Relat Disord 2023;8:100188. [doi: 10.1016/j.prdoa.2023.100188] [Medline: 36864905]
7. Devices for remote monitoring of Parkinson's disease. NICE. 2023. URL: https://www.nice.org.uk/guidance/dg51 [accessed 2025-03-05]
8. Liikkanen S, Sinkkonen J, Suorsa J, et al. Feasibility and patient acceptability of a commercially available wearable and a smart phone application in identification of motor states in parkinson's disease. PLOS Digit Health 2023 Apr;2(4):e0000225. [doi: 10.1371/journal.pdig.0000225] [Medline: 37027348]
9. Kangarloo T, Latzman RD, Adams JL, et al. Acceptability of digital health technologies in early Parkinson's disease: lessons from WATCH-PD. Front Digit Health 2024 Aug 26;6:1435693. [doi: 10.3389/fdgth.2024.1435693] [Medline: 39253055]

XSL•FO
RenderX

10. Morgan C, Tonkin EL, Craddock I, Whone AL. Acceptability of an in-home multimodal sensor platform for Parkinson disease: nonrandomized qualitative study. JMIR Hum Factors 2022 Jul 7;9(3):e36370. [doi: 10.2196/36370] [Medline: 35797101]
11. Fisher JM, Hammerla NY, Rochester L, Andras P, Walker RW. Body-worn sensors in Parkinson's disease: evaluating their acceptability to patients. Telemed J E Health 2016 Jan;22(1):63-69. [doi: 10.1089/tmj.2015.0026] [Medline: 26186307]
12. Sharma A, Minh Duc NT, Luu Lam Thang T, et al. A Consensus-Based Checklist for Reporting of Survey Studies (CROSS). J Gen Intern Med 2021 Oct;36(10):3179-3187. [doi: 10.1007/s11606-021-06737-1] [Medline: 33886027]
13. Walters K, Armstrong M, Gardner B, et al. Clinical and cost-effectiveness of "Live Well with Parkinson's" self-management intervention versus treatment as usual for improving quality of life for people with Parkinson's: study protocol for a randomised controlled trial. Trials 2023 Dec 5;24(1):793. [doi: 10.1186/s13063-023-07700-7] [Medline: 38053209]
14. ActiveInsights band. ActiveInsights. URL: https://activinsights.com/technology/activinsights-band/ [accessed 2024-06-27]
15. AX3 band. Axivity. URL: https://axivity.com/product/ax3 [accessed 2024-06-27]
16. Narayanan A, Stewart T, Mackay L. A dual-accelerometer system for detecting human movement in a free-living environment. Med Sci Sports Exerc 2020 Jan;52(1):252-258. [doi: 10.1249/MSS.0000000000002107] [Medline: 31361712]
17. Maetzler W, Rochester L. Body-worn sensors—the brave new world of clinical measurement? Mov Disord 2015 Aug;30(9):1203-1205. [doi: 10.1002/mds.26317] [Medline: 26173577]
18. Morgan C, Rolinski M, McNaney R, et al. Systematic review looking at the use of technology to measure free-living symptom and activity outcomes in Parkinson's disease in the home or a home-like environment. J Parkinsons Dis 2020;10(2):429-454. [doi: 10.3233/JPD-191781] [Medline: 32250314]
19. Brobbin E, Parkin S, Deluca P, Drummond C. A qualitative exploration of the experiences of transdermal alcohol sensor devices amongst alcohol service practitioners (South London, UK). Addict Res Theory 2025;33(1):36-47. [doi: 10.1080/16066359.2024.2309869] [Medline: 39906322]
20. Evers LJW, Peeters JM, Bloem BR, Meinders MJ. Need for personalized monitoring of Parkinson's disease: the perspectives of patients and specialized healthcare providers. Front Neurol 2023;14:1150634. [doi: 10.3389/fneur.2023.1150634] [Medline: 37213910]

Abbreviations

CROSS: Consensus-Based Checklist for Reporting of Survey Studies
NHS: National Health Service
PD: Parkinson disease

Edited by A Mavragani; submitted 27.06.24; peer-reviewed by G Ymeri, J Ruokolainen, S Parkin; revised version received 31.01.25; accepted 03.02.25; published 25.03.25.

Please cite as:
Rookes TA, Batla A, Armstrong M, Ambler G, Walters K, Schrag A
Patient Acceptability and Technical Reliability of Wearable Devices Used for Monitoring People With Parkinson Disease: Survey Study
JMIR Form Res 2025;9:e63704
URL: https://formative.jmir.org/2025/1/e63704
doi:10.2196/63704

© Tasmin Alanna Rookes, Amit Batla, Megan Armstrong, Gareth Ambler, Kate Walters, Anette Schrag. Originally published in JMIR Formative Research (https://formative.jmir.org), 25.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Using Text Messaging Ecological Momentary Assessment to Record Changes in e-Cigarette and Combustible Cigarette Use: Pilot Randomized Clinical Trial

Tucker Morgan[1], BBA, MS; Michelle He[1], BS, MS; Andrew Nicholson[1], BS, MSPH; Omar El Shahawy[1,2], MD, PhD; Scott E Sherman[1,3], MPH, MD; Elizabeth R Stevens[1], MPH, PhD

[1]Department of Population Health, Grossman School of Medicine, New York University, 180 Madison Ave, New York, NY, United States

[2]School of Global Public Health, New York University, New York, NY, United States

[3]Department of Medicine, VA New York Harbor Health System, Veterans Health Administration, New York, NY, United States

Corresponding Author:
Elizabeth R Stevens, MPH, PhD
Department of Population Health, Grossman School of Medicine, New York University, 180 Madison Ave, New York, NY, United States

Abstract

Background: Ecological momentary assessment (EMA) provides insight into the effectiveness and feasibility of smoking-related interventions.

Objective: The objective of this paper was to assess adherence to an EMA protocol and compare EMA-derived responses with measures collected through multiple surveys.

Methods: A subanalysis was conducted using data from a 12-week, open-label, and 2-arm pilot randomized clinical trial among adult smokers with chronic obstructive pulmonary disease, coronary artery disease, peripheral vascular disease, or asthma in the last 12 months (n=109). Participants were randomized to either electronic cigarette (EC) or nicotine replacement therapy (NRT) treatment arms. We compared EMA data collected through automated SMS text message prompts sent to participants 4 times daily for 12 weeks, including cigarettes smoked per day (CPD), craving, and satisfaction, to survey data collected at 12 weeks. Convergent validity between survey- and EMA-reported measures was evaluated using Pearson correlation and paired *t* tests. CPD was modeled using negative binomial regression. Relative rates (RRs) of reaching at least 50%, 75%, and 100% CPD reduction between two arms were calculated using both EMA and survey data.

Results: The majority of participants were non-Hispanic White (63/109, 58%) and female (60/109, 55%), and had a median age of 60 (IQR 54 - 65) years. Among the 109 participants, 59.6% (n=65) were consistently adherent to the EMA protocol over the 12-week period. Median weekly EMA response rate remained high over the 12-week study period even though a modest decline was observed (week 1, 97.8% and week 12, 89.4%). The mean CPD declined significantly (week 1, mean 14.2, SD 9.9 and week 12, mean 4.6, SD 6.7; *P*<.001). EMA-derived and survey-based CPD measurements were positively correlated (*r*=0.73, 95% CI 0.6-0.82) as were measures of craving (*r*=0.38, 95% CI 0.17-0.56). No significant paired difference in CPD was observed between EMA measurements and surveys. A significant effect of time on CPD EMA data (incidence rate ratio [IRR] 1-week change 0.93; *P*<.01) and survey data was found (IRR 12-week change 0.36; *P*<.01). However, the treatment effect was not significant, which aligned with the RR results. An increase in the EC consumption was observed over time in the EC arm, with 12.1% (7/58) fully switched to EC (defined as CPD=0 and EC use>0) and 20.7% (12/58) mostly switched (defined as a reduction in CPD>75% and EC use>0) in week 12.

Conclusions: EMA is a suitable method to collect recall-based smoking-related data. Though results from mixed effect modeling and RR comparisons were similar using EMA or survey data, EMA provides unique advantages, namely greater granularity in the time and the capability to detect switching patterns in near real time. These findings provide the feasibility of using EMA in developing smoking cessation interventions in future tobacco harm reduction research.

Trial Registration: ClinicalTrials.gov NCT04465318; https://clinicaltrials.gov/study/NCT04465318

International Registered Report Identifier (IRRID): RR2-10.1186/s13722-021-00284-0

(JMIR Form Res 2025;9:e66709) doi:10.2196/66709

KEYWORDS

smoking; electronic cigarette; smoking cessation; ecological momentary assessment; EMA; adherence; real-time data; mobile phone; cigarette; smoker; e-cigarettes; pilot randomized clinical trial; clinical trial; pilot study; adult smoker; chronic condition

XSL•FO
RenderX

Introduction

In the United States, smoking remains the leading preventable cause of death, responsible for 480,000 deaths and US $170 billion in health care spending each year [1]. Despite considerable progress in the 20th century, in 2021, 30 million US adults were still cigarette users [2]. The benefits of smoking cessation are well-established [1]; however, there are high rates of relapse with existing smoking cessation therapies in some groups, like those with chronic obstructive pulmonary disease (COPD) [3]. Alternate smoking cessation tools, like electronic cigarettes (ECs), have been proposed as a more appealing harm reduction intervention for people who smoke [4-6].

In public health research, a primary interest is real-world behavior; however, this can be difficult to accurately assess in a laboratory or survey setting where measures, such as cigarettes smoked per day (CPD), typically rely on retrospective self-reports. Ecological momentary assessment (EMA), which is a data collection method that uses repeated sampling to collect data throughout a participant's everyday life, allows for the reporting of events in real-world settings and is much less time-consuming compared to traditional study visits [7]; EMA can limit issues associated with recall bias, particularly when participants are asked to recall the frequency of events over a long period [8]. Furthermore, EMA is often done remotely, thus serving as a useful tool for researchers considering remote assessments for their studies.

Interval-based EMA data collection methods have been used widely to study smoking habits and factors affecting use [9,10]. EMA data can provide useful insight into the changes in cigarette use and other factors, such as craving and satisfaction, over time and with greater granularity than study visits that are typically weeks or several months apart. Switching effects from combustible cigarettes (CCs) to ECs over an extended period, which has previously been observed through more traditional means of data collection [11], may be better understood using EMA data, particularly for vulnerable smokers with COPD and other chronic conditions.

This study assesses the comparability of smoking-related EMA measurements—CPD, craving, and satisfaction—with survey assessments completed as part of a 2-arm pilot randomized clinical trial (RCT). We also assessed the association between participant characteristics and adherence to the EMA protocol, visualized changes in EC use and CPD during program participation, and quantified the percent of individuals who switched completely to ECs over time. Furthermore, we compared the results of mixed effects negative binomial regression and relative rate comparisons of CPD reduction using EMA data versus using study visit data.

Methods

Study Setting and Participants

Participants were recruited via emails and phone calls between October 2020 and September 2022 from the electronic medical records of the New York University Langone Health (NYULH) system, a private hospital system serving New York, New Jersey, and Connecticut with approximately 6.8 million active patients. Patients were eligible for participation if they (1) had an ambulatory *International Statistical Classification of Diseases, Tenth Revision* (*ICD-10*) diagnosis code for COPD, cardiovascular disease, or asthma in the last 12 months; (2) for COPD patients only, a COPD Assessment Test (CAT) score ≥10; (3) were aged 21 to 75 years (the legal age for purchasing ECs in New York is 21); (4) were a current CC user (≥5 packs in a lifetime, smokes ≥4 days/week); (5) smoked at least 5 CCs per day on days they smoke CCs; (6) were motivated to quit smoking (at least a 5 on a 10-point Contemplation Ladder) [12]; and (7) possessed a phone with SMS text messaging capabilities. Potential participants were excluded if they (1), for COPD patients only, had a CAT score ≥30 (representing severe COPD) [13]; (2) reported NRT or EC use in the last 14 days; (3) had a medical condition (eg, unstable angina or heart disease) precluding use of nicotine patch or gum as determined by the study physician or by their treating physician; or (4) were pregnant or breastfeeding. Individuals who were not open to using NRT or ECs were also excluded from the study. Participants using other tobacco products were not excluded from this study.

Ethical Considerations

The NYU Langone Health Institutional Review Board (IRB) approved the study procedures (s20-00839) and written documentation of informed consent was received before starting data collection. The original informed consent allows for secondary analyses without additional consent. All participant data were deidentified prior to analysis. Participants did not receive any incentive for counseling calls but were incentivized to complete study visits and respond to the daily SMS text message smoking diary. Participants received US $10 for completion of a survey at baseline and at 6, 8, and 12 weeks, and US $20 for completion of a survey at 24 weeks. Participants could receive up to US $40 for completing a daily smoking diary via SMS text messages (US $20 if at least 60% of assessments are completed and US $40 if at least 80% are completed). In total, participants had the possibility of receiving a total of US $100 in incentives.

Study Design

This is a subanalysis of a 2-arm, open-label pilot RCT [14] designed to evaluate the harm-reducing effects of ECs in patients with COPD, coronary artery disease (CAD), peripheral vascular disease (PVD), or asthma. The methods of the parent study are detailed fully elsewhere [14]. Briefly, study participants were recruited from two large urban health centers' electronic health records (EHRs) and randomized in a 1:1 ratio to either standard smoking cessation care (nicotine replacement therapy [NRT] and counseling) or ECs and counseling [15]. Randomization was stratified by sex and baseline CPD (<20 vs ≥20). Participants were screened via phone or an internet-based screening tool before enrollment.

Participants completed study visits by telephone at baseline and at 1, 2, 3, and 6 months, with the intervention period lasting 3 months. At each visit, participants completed self-report survey assessments and were asked about their issued product supply and potential harm. Sociodemographic characteristics such as

XSL•FO
RenderX

age, race, ethnicity, place of birth, gender, marital status, education, work status, living situation, and homelessness were measured via self-report at baseline. In addition, condition-specific assessments (CAT [16], Clinical COPD Questionnaire [CCQ] [17,18], Seattle Angina Questionnaire-7 [SAQ-7] [19], Asthma Control Test [ACT] [20,21]) were also assessed as was motivation and confidence to quit smoking, tobacco use, and craving via the Minnesota Nicotine Withdrawal Scale (MNWS) [22].

Based on EHR diagnoses, participants were characterized into the following groups: COPD, asthma, and CAD or PVD. Participants did not receive any incentive for counseling calls, but were incentivized to complete study visits and respond to the daily SMS text message smoking diary.

EMA Protocol

In addition to study visit surveys, during the intervention period, participants were asked to record daily CC and EC use along with measures for CC craving and EC and CC satisfaction by SMS text messaging. Since the EMA protocol was embedded within the RCT, it allowed us to directly compare the message-derived outcomes with the survey assessments across intervention groups. The details of branching logic on the texting sequence participants would receive and the phrasing of each question are presented (Figure S1 in Multimedia Appendix 1). Certain responses triggered the follow-up questions about daily use, satisfaction, and craving. If the participant reported no CC or no EC use, only craving questions were triggered. Only participants in the EC arm received questions about EC use. Due to discordance between NRT use instructions (ie, 24 h patch) and the scheduled check-ins, the NRT arm was not asked about NRT use via SMS text message check-ins. This procedure employed time-based sampling with participants receiving SMS text message prompts four times per day, breaking the day into "Early, Lunch, Afternoon, Evening" periods such that approximately 4 hours had passed between check-ins.

Participants received brief instructions sent via an introductory text and response indicators in the body of SMS text message prompts but no formal training; of note, the EMA protocol was explained to participants during the baseline and consent process. Participants were given the opportunity to opt out of receiving SMS text messages before initiation of participation and during the study. Automated SMS text message prompts were sent to participants daily over the 12-week intervention.

EMA Data Processing and EMA-Based Measures

The SMS text message data were used to determine daily and weekly summaries for the following measures: response rate, number of CCs, EC uses (EC arm only), craving, satisfaction with the last CC, and satisfaction with the last EC (EC arm only).

Response Rate and CPD

We tabulated a participant's weekly response rate as the number of responses from the participant divided by the total number of response-required texts (ie, those texts soliciting a response from the participant). To obtain an estimate of CPD—one that would be comparable to the traditional CPD measure recorded at study visits at baseline and week 12—we calculated the weekly average CPD by dividing the total number of reported cigarettes smoked in a week by 7 days. Participants with weekly response rates lower than 60% were removed from the calculation to obtain a more stable estimate of CPD and reduce downward bias. When calculating CPD for week 12, corresponding to the 3-month follow-up visit, the average CPD for the final week of EMA data collection was used rather than a total for the final day alone; we believe this average better approximates what participants would report in a visit setting compared with a 1-day total.

EC-Specific Measures

For the EC group, we used responses to the prompt, "Since your last report, how many separate times did you use an e-cig?" to calculate weekly averages for EC use. We defined mostly switched as >75% reduction in CPD from week 1 and use of ECs. We defined switching to ECs as those reporting 0 CPD smoked and EC use>0 in a given week.

Satisfaction and Craving

We calculated the average craving (range of 0 - 9 with higher values corresponding to higher craving) by averaging a person's responses to the craving question ("On a scale of 0 - 9, how much do you want to smoke a cigarette right now?") within a week. Similarly, we calculated the average satisfaction with the last cigarette and satisfaction with the last EC (range of 0 - 9 with higher values corresponding to greater satisfaction) by averaging a person's responses to the satisfaction question ("On a scale of 0 - 9, how satisfying was your last cigarette/e-cig?") within a week.

Statistical Analysis

To evaluate predictors of EMA adherence, participants were divided into two groups: (1) those with a weekly response rate >60% or more throughout the duration of the study period and (2) those with a weekly response rate <60% at any point in the study period. Descriptive statistics were calculated by group, including sociodemographic characteristics (age, gender, race and ethnicity, marital status, education, and work status) as well as condition-specific assessments, motivation, and confidence to quit smoking and tobacco use for each group. For variables not normally distributed, median and IQR were presented. Characteristics were evaluated using Wilcoxon rank sum test for continuous variables and Fisher exact test for categorical variables.

To assess the comparability between EMA measures of CPD and more traditional measurements gathered in a survey setting, we examined the Pearson correlation coefficient between CPD reported during the final week of EMA prompting (week 12) and CPD reported at the 3-month survey. We also examined the paired difference between EMA- and survey-reported CPD at 3 months using a paired *t* test. Pearson correlation coefficient was also applied to assess the convergent validity of EMA measurements for craving compared with survey-reported measures (first item of the Minnesota nicotine withdrawal scale: "Do you have a craving for cigarettes or desire to smoke?" with responses coded from 0 - 4, corresponding to "Not present" [0], "Slight" [1], "Mild" [2], "Moderate" [3], and "Severe" [4])

XSL•FO
RenderX

[23]. The EMA prompt “on a scale of 0 - 9, how much do you want to smoke a cigarette right now?” was used as the comparator for craving with higher numbers corresponding to greater craving.

To assess our primary outcome of change in CPD, we used a negative binomial mixed effects model, adjusting for the intervention arm, time (weeks from baseline), and a random intercept for each individual. We compared the results obtained from surveys by building a similar model, with the time variable expressed as time from baseline to week 12 visit. The negative binomial model has been shown to be more effective for count data than a Poisson model in the presence of overdispersion [24]. CPD reduction rate was evaluated by measuring the change in CPD from the baseline to week 12 from both EMA and survey-based data. To compare reduction rates by arm at week 12, we calculated the risk ratios of reduction rates of those reaching at least 50%, 75%, and 100% reduction as reported via EMA and survey. We used the same negative binomial mixed effects models to assess the satisfaction and craving scores.

For those in the EC arm, we tabulated the number and percentage of individuals each week that demonstrated evidence of switching (CPD=0 and EC use>0) as well as mostly switching (reduction in CPD >75% and EC use>0). We also summarized weekly EMA-based measures including response rates, the number and percentage of individuals with >60% response rate, average craving, average satisfaction with last CC, average satisfaction with last EC, average CPD, and average EC uses.

All analyses were conducted using Stata (version 18; Stata Corp) and R (version 2024.04.2+764; R Foundation for Statistical Computing), and all tests were 2-tailed, with a .05 significance level.

Results

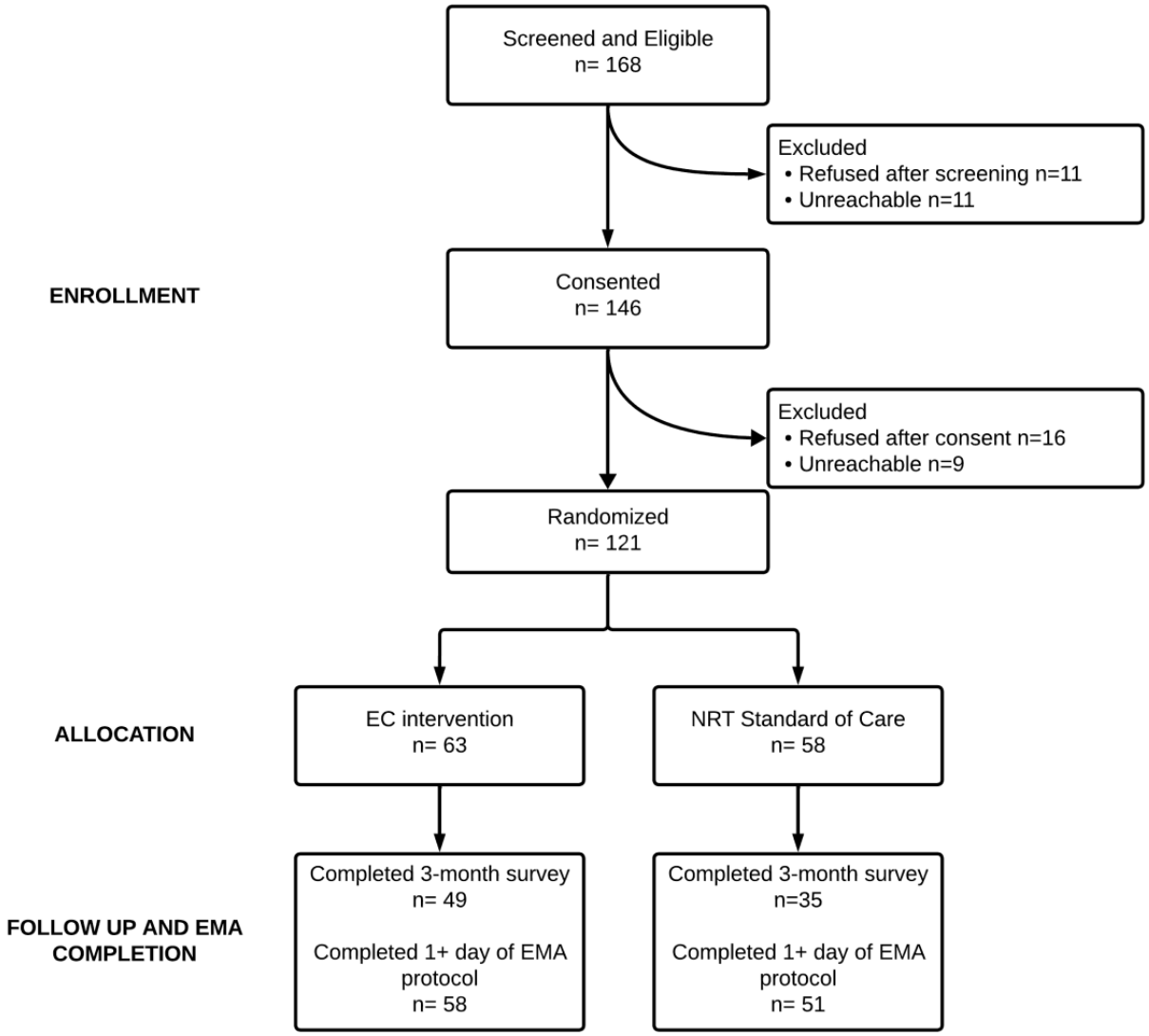
Adherence to the EMA Protocol

In total, 109 (90%) individuals of the 121 randomized trial participants took part in the EMA protocol. Among the 109, 65 (59.6%) were considered adherent to the EMA protocol with a median weekly response rate that declined slowly from 97.8% in week 1% to 89.4% in week 12. The mean and median response rates over time for all participants are plotted in Figure S2 in Multimedia Appendix 1. Of the 12 individuals with no texting data, 5 withdrew from study activities or opted out, 3 never received the study welcome text, 3 never responded to the study welcome text, and 1 died during enrollment (Figure 1).

Table 1 displays participant characteristics by adherence to EMA protocol. The median age among adherent participants was 61 (IQR 54-65) years and 58 (IQR 52-65) tears among nonadherent individuals. The sample was mostly non-Hispanic white (63/108, 57.8%), and a plurality were married or dating (52/109, 47.7%), had some college or an associate’s degree (44/109, 40.4%), and were retired or on public assistance (47/109, 43.1%). Most participants had COPD (65/109, 59.6%), followed by asthma (36/109, 33%). The median motivation to quit smoking score was 10 (IQR 8 - 10), and the mean confidence to quit smoking was 6.5 (SD 2.2). At baseline, the mean number of CPD reported at the survey was 17.3 (SD 9.7), and the mean CPD from week 1 as assessed from EMA was 11.6 (SD 7.5). No significant differences were observed between the adherent and nonadherent groups.

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram for trial recruitment and ecological momentary assessment (EMA) data analysis inclusion. EC: electronic cigarette; NRT: nicotine replacement therapy.



XSL•FO
RenderX

Table . Baseline characteristics of the patients with chronic obstructive pulmonary disease (COPD), coronary artery disease (CAD), peripheral vascular disease (PVD), or asthma in the 12-week, open-label, and 2-arm pilot randomized clinical trial. Distributions and statistical metrics for the comparison of adherent and nonadherent groups to the EMA protocol are presented (n=109). All values are presented as mean (SD) or n (%) or as otherwise stated. Characteristics were evaluated using Wilcoxon rank sum test for continuous variables and Fisher exact test for categorical variables. "Adherent" participants had a weekly response rate of 60% or more throughout the duration of the study period and "Non-adherent" participants had a weekly response rate less than 60% at any point in the study period.

Characteristic	Overall	Adherent	Nonadherent	*P* value
Participants, n (%)	109 (100)	65 (59.6)	44 (40.4)	—[a]
Study arm, n (%)				.82
e-Cigarette	58 (53.2)	34 (52.3)	24 (54.5)	
NRT[b]	51 (46.8)	31 (47.7)	20 (45.5)	
Sociodemographics				
Age (years), median (IQR)	60.0 (54.0 - 65.0)	61.0 (54.0 - 65.0)	58.0 (51.5 - 64.5)	.16
Sex, n (%)				.76
Male	49 (45)	30 (46.2)	19 (43.2)	
Female	60 (55)	35 (53.8)	25 (56.8)	
Race and ethnicity, n (%)				.78
Non-Hispanic White	63 (57.8)	40 (61.5)	23 (52.3)	
Non-Hispanic Black or African American	25 (22.9)	13 (20)	12 (27.3)	
Hispanic	16 (14.7)	9 (13.8)	7 (15.9)	
Non-Hispanic other[c]	5 (4.6)	3 (4.6)	2 (4.5)	
Marital status, n (%)				.67
Married or dating	52 (47.7)	28 (43.1)	24 (54.5)	
Separated, divorced, or widowed	25 (22.9)	16 (24.6)	9 (20.5)	
Single	32 (29.4)	21 (32.3)	11 (25.0)	
Educational history, n (%)				.56
Did not finish high school	11 (10.1)	6 (9.2)	5 (11.4)	
Completed high school or high school equivalency diploma (GED)	24 (22)	16 (24.6)	8 (18.2)	
Some college or associate's degree	44 (40.4)	28 (43.1)	16 (36.4)	
Completed college or graduate school	30 (27.5)	15 (23.1)	15 (34.1)	
Work status, n (%)				.45
Full-time or part-time work	42 (38.5)	21 (32.3)	21 (47.8)	
Public assistance or retirement saving	47 (43.1)	30 (46.2)	17 (38.6)	
Unemployed or other	20 (18.3)	14 (21.5)	6 (13.6)	
Screening diagnoses				
Asthma	36 (33.0)	18 (27.7)	18 (40.9)	.15
CAD or PVD	24 (22.0)	13 (20.0)	11 (25.0)	.54
COPD	65 (59.6)	42 (64.6)	23 (52.3)	.20
1 or more conditions	16 (14.7)	8 (12.3)	8 (18.2)	.40

XSL•FO
RenderX

Characteristic	Overall	Adherent	Nonadherent	*P* value
Condition-specific assessments [d] , median (IQR)				
CAT[e]	20.5 (16.0 - 26.0)	19.5 (14.0 - 26.0)	22.0 (16.5 - 26.0)	.44
SAQ-7[f]	15.0 (13.0 - 19.0)	15.0 (13.0 - 19.0)	15.5 (13.5 - 19.5)	.65
ACT[g]	78.8 (69.7 - 87.9)	77.3 (69.7 - 89.4)	78.8 (69.7 - 84.8)	.79
Dyspnea and CCQ [h]				
Dyspnea (shortness of breath) Scale				.13
Dyspnea only with strenuous exercise	18 (16.5)	13 (20.0)	5 (11.4)	
1. Dyspnea when hurrying or walking up a slight hill	27 (24.8)	19 (29.2)	8 (18.2)	
2. Walks slower than people of the same age because of dyspnea or has to stop for breath when walking at own pace	24 (22.0)	10 (15.4)	14 (31.8)	
3. Stops for breath after walking 100 yards (91m) or after a few minutes	28 (25.7)	17 (26.2)	11 (25.0)	
4. Too dyspneic to leave house or breathless when dressing	10 (9.2)	6 (9.2)	4 (9.1)	
CCQ total score, median (IQR)	2.9 (2.1 - 3.9)	2.7 (2.1 - 3.8)	3.2 (2.2 - 4.0)	.51
CCQ symptom score, median (IQR)	3.5 (2.6 - 4.5)	3.2 (2.8 - 4.5)	3.8 (2.5 - 4.5)	.83
CCQ functional state score, median (IQR)	2.5 (1.5 - 3.5)	2.0 (1.5 - 3.2)	2.6 (1.8 - 3.6)	.27
CCQ mental state score, median (IQR)	2.5 (1.0 - 3.5)	2.0 (1.0 - 4.0)	2.5 (1.5 - 3.5)	.76
Motivation and confidence to quit smoking, median (IQR)				
Motivation to quit (1-10)	10.0 (8.0 - 10.0)	10.0 (8.0 - 10.0)	9.5 (7.0 - 10.0)	.72
Confidence to quit (1-10)	6.5 (2.2)	6.5 (2.2)	6.4 (2.2)	.71
Tobacco use and risk perception, mean (SD)				
Average number of CPD[i]	17.3 (9.7)	16.9 (9.4)	17.9 (10.2)	.60
CPD per SlickText EMA (week 1)	11.6 (7.5)	11.5 (7.1)	12.0 (8.3)	.76
Frequency of current cigarette use, n (%)				.09
Every day	105 (96.3)	61 (93.8)	44 (100)	

XSL•FO
RenderX

Characteristic	Overall	Adherent	Nonadherent	*P* value
Some days	4 (3.7)	4 (6.2)	0 (0)	

[a]Not applicable.

[b]NRT: nicotine replacement therapy.

[c]Non-Hispanic other include Arab or Middle Eastern (n=2), not reported (n=2), and Asian (n=1).

[d]Condition-specific assessments restricted to those with condition: COPD: Baseline (n=65; 23 nonadherent and 42 adherent), CAD or PVD: Baseline (n=24; 13 nonadherent and 11 adherent), asthma: Baseline (n=36; 18 nonadherent and 18 adherent).

[e]CAT: COPD Assessment Test.

[f]SAQ-7: Seattle Angina Questionnaire-7.

[g]ACT: Asthma Control Test.

[h]CCQ: Clinical COPD Questionnaire.

[i]CPD: cigarettes smoked per day.

Summary of EMA-Derived Measures

Consumption of CC and EC Over Time

Table 2 provides weekly summaries of response rates, CC and EC use, and satisfaction and craving. Using EMA, overall, the reduction of weekly average CPD was statistically significant, decreasing from 14.2 (SD 9.9) in week 1 to 4.6 (SD 6.7) in week 12 (P<.001). Among the EC group, the average ECs per day modestly increased from 4.3 (SD 5.5) in week 2 to 7.1 (SD 11.6) in week 10, followed by a minor decline to 5.8 (SD 9.6) in week 12.

XSL•FO
RenderX

Table . Weekly smoking-related metrics summarized from ecological momentary assessment (EMA) protocol over the 12-week study period. Satisfaction and craving values were based on a 0-9 point scale, ranging from 0 (not satisfied or no craving at all) to 9 (extremely satisfied or highest craving).

	Response rate			CC[a] and EC[b] use				Satisfaction and craving		
Week	Median response rate (IQR)	Mean response rate (SD)	Participants with >60% response rate, n (%)	Mean CPD[c] (SD)	Mean EC uses per day (SD)	Participants reporting mostly switching, n (%)[d,e]	Participants switching to ECs[d], n (%)	Mean Satisfaction with last CC (SD)	Mean EC satisfaction[d] (SD)	Mean craving (SD)
1	97.8 (94.7-100.0)	95.4 (8.7)	104 (95.4)	14.2 (9.9)	—[f]	—	—	6.2 (1.6)	—	4.7 (1.7)
2	98.0 (91.1-100.0)	94.9 (8.6)	102 (93.6)	10.9 (8.9)	4.3 (5.5)	3 (5.2)	1 (1.7)	5.7 (1.8)	4.9 (2.0)	4.6 (1.7)
3	96.9 (87.5-99.2)	94.1 (10.0)	100 (91.7)	8.8 (8.1)	6.4 (7.4)	5 (8.6)	1 (1.7)	5.7 (1.9)	4.9 (2.0)	4.4 (1.7)
4	96.2 (85.8-99.1)	93.4 (10.2)	96 (88.1)	7.6 (7.6)	6.3 (7.1)	8 (13.8)	4 (6.9)	5.6 (2.0)	4.9 (2.0)	4.5 (2.0)
5	96.3 (81.6-98.5)	92.5 (11.9)	93 (85.3)	7.0 (7.5)	7.0 (9.4)	11 (19.0)	6 (10.3)	5.6 (1.9)	5.0 (1.9)	4.5 (1.8)
6	94.6 (74.9-98.5)	91.8 (12.6)	87 (79.8)	6.3 (7.5)	6.7 (7.7)	9 (15.5)	3 (5.2)	5.5 (2.0)	5.1 (1.8)	4.4 (2.0)
7	93.2 (69.4-98.6)	91.7 (13.0)	83 (76.1)	5.9 (7.2)	7.3 (9.1)	7 (12.1)	5 (8.6)	5.5 (2.0)	5.2 (1.7)	4.3 (1.8)
8	91.6 (61.5-98.3)	91.1 (13.3)	80 (73.4)	5.0 (7.3)	7.0 (10.4)	8 (13.8)	6 (10.3)	5.4 (2.0)	5.2 (1.7)	4.2 (2.0)
9	89.0 (33.3-97.6)	90.2 (14.5)	74 (67.9)	5.7 (8.7)	6.1 (9.4)	12 (20.7)	5 (8.6)	5.4 (2.0)	5.2 (1.8)	4.1 (2.2)
10	88.9 (46.5-97.9)	89.8 (13.7)	76 (69.7)	5.2 (8.1)	7.1 (11.6)	12 (20.7)	7 (12.1)	5.4 (2.1)	5.3 (1.8)	4.3 (2.2)
11	87.7 (34.9-98.0)	89.9 (13.7)	74 (67.9)	5.1 (8.2)	6.2 (10.3)	10 (17.2)	6 (10.3)	5.4 (2.1)	5.4 (1.8)	4.0 (2.2)
12	89.4 (42.5-97.8)	89.1 (15.5)	70 (64.2)	4.6 (6.7)	5.8 (9.6)	12 (20.7)	7 (12.1)	5.3 (2.2)	5.4 (1.8)	4.1 (2.2)

[a]CC: combustible cigarette.

[b]EC: e-cigarette.

[c]CPD: cigarettes per day.

[d]Summary measures refer only to the EC group (n=58).

[e]Mostly switching is defined as >75% reduction in CPD from week 1 and use of e-cigarettes.

[f]Not applicable.

Craving and Satisfaction Over Time

We were also able to obtain weekly estimates for craving, satisfaction with the last CC, and satisfaction with the last EC (Table 2). Weekly average craving declined modestly from 4.7 (SD 1.7) in week 1 to 4.1 (SD 2.2) in week 12. Similarly, weekly average satisfaction with the last cigarette declined from 6.2 (SD 1.6) in week 1 to 5.3 (SD 2.2) in week 12. We report weekly satisfaction and craving by arm in Table 2. No treatment effect was found in the negative binomial mixed effects models for satisfaction and craving (Tables S1 and S2 in Multimedia Appendix 1). Among EC participants, average satisfaction with the last EC rose gradually from 4.9 (SD 2.0) in week 2 to 5.4 (SD 1.8) in week 12.

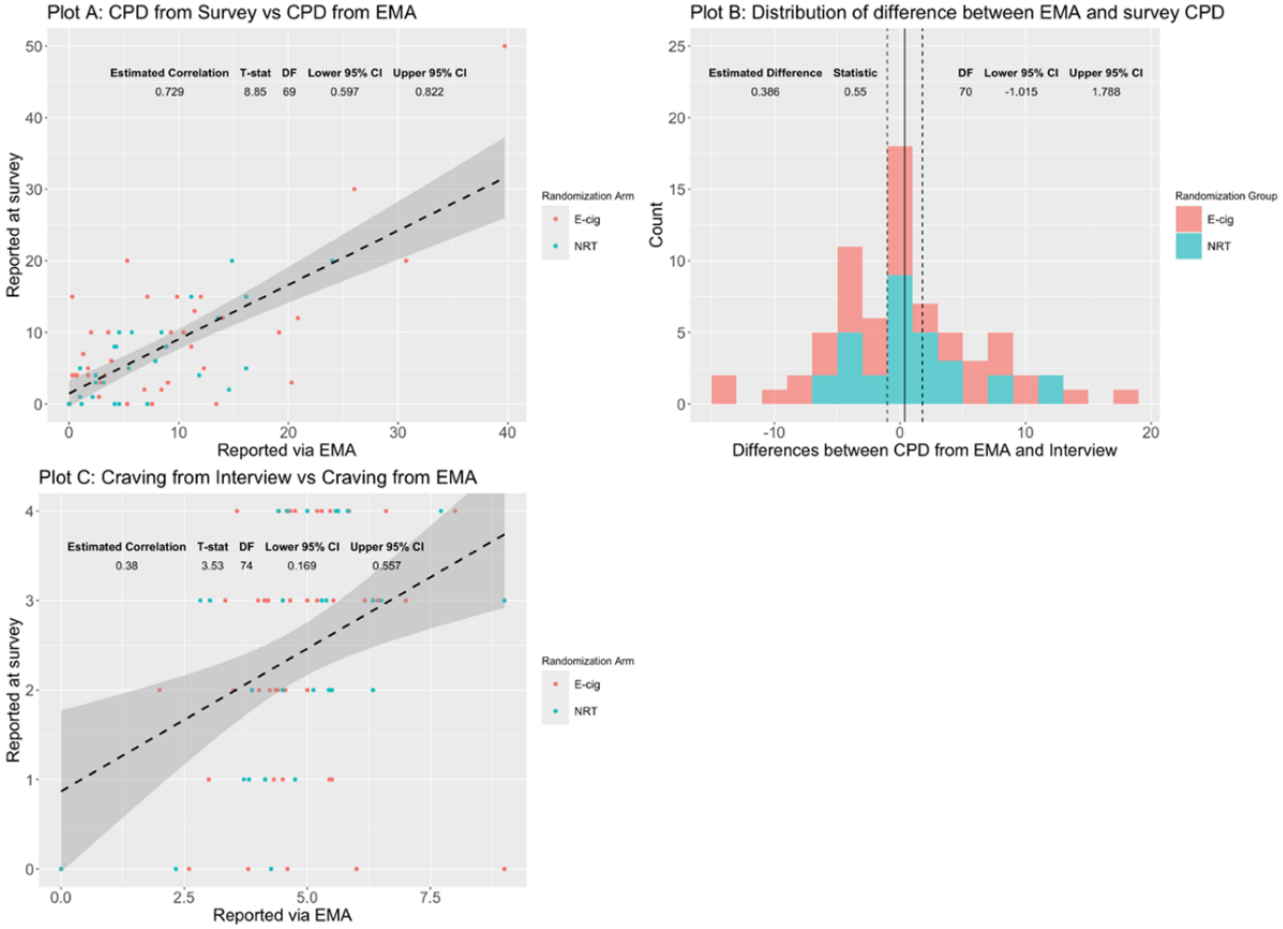
EMA Comparability With Survey Assessments

Reported CPD estimates from EMA correlated strongly with the survey setting at 3 months (Pearson r=0.73; 95% CI 0.60-0.82). Of note, there was one EC participant who reported a very high consumption of combustible cigarettes (>40 CPD) via EMA and at their 3-month visit; this individual's data was included in the current analysis as results did not differ meaningfully when excluding their data. There was no statistically significant paired difference between EMA-reported CPD and survey-reported CPD (paired difference=0.39, 95% CI −1.02 to 1.79). Craving measurements from EMA and surveys also showed a significant positive correlation (Pearson r=0.38, 95% CI 0.17-0.56). In Figure 2, plot A depicts a scatter plot visualizing the association between CPD use via survey and EMA at week 12. The estimated correlation (Pearson r) is 0.73 (95% CI 0.60-0.82). A fitted regression line and 95% CI

XSL•FO
RenderX

are included. Data points are stratified by treatment arm. Plot B depicts a histogram showing the distribution of paired CPD differences between EMA and survey at week 12 for both treatment arms. The estimated mean difference is 0.39 (95% CI −1.02 to 1.79). Plot C depicts a scatter plot showing the association between craving scores for CC via survey and EMA at week 12. The estimated correlation is 0.38 (95% CI 0.17-0.56). A fitted regression line and 95% CI are included. Data points are stratified by treatment arm.

Figure 2. Comparison of data from both ecological momentary assessment (EMA)- and survey-based method for cigarettes smoked per day (CPD) and craving scores for combustible cigarette (CC) by treatment arms. E-cig: e-cigarette; NRT: nicotine replacement therapy.



Reduction Rate Comparisons and Negative Binomial Regression Using EMA Versus Survey Data

Tables 3 and 4 show the comparisons between the reduction rate in CPD (50%, 75%, and 100% reduction) and negative binomial regression results using data reported through EMA and through surveys. Both treatment groups showed a significant decrease in the mean number of CPD at the end of the study period (overall; week 1: mean 14.2, SD 9.9; week 12: mean 4.6, SD 6.7) though there was no evidence that reduction in CPD was different between arms based on reduction rate. Of note, the point estimates for CPD reduction rate were all slightly higher than those computed from survey data.

The average CPD count had a mean of 8.2 (SD 8.03) and a variance of 64.5. With unequal mean and variance, our model must account for overdispersion, which the negative binomial model accomplishes. In both models, we found a significant effect of time on the average number of cigarettes per day. In the model using EMA data, we found a 7% decrease in CPD for each week (IRR for 1-week change 0.929, 95% CI 0.93-0.93), and in the model using survey data, the IRR for time across the 12-week intervention was 0.36 (95% CI 0.29-0.44). In both models, we found no statistically significant effect of randomization group on the rate of CPD; however, interestingly, the point estimates for the IRR for randomization group are in differing directions.

XSL•FO
RenderX

Table . Relative rates of reaching at least 50%, 75%, and 100% cigarettes smoked per day (CPD) reduction rates for time and treatment effects using ecological momentary assessment (EMA) and study surveys at week 12.

	EMA				Study survey			
	EC[a] (n=58), n (%)	NRT[b] (n=51), n (%)	RR[c] (95% CI)	*P* value	EC (n=58), n (%)	NRT (n=51), n (%)	RR (95% CI)	*P* value
CPD reduction rate								
50% or higher	34 (58.6)	24 (47.1)	1.25 (0.87 - 1.79)	.23	31 (53.4)	22 (43.1)	1.24 (0.83 - 1.84)	.28
75% or higher	26 (44.8)	15 (29.4)	1.52 (0.91 - 2.54)	.10	17 (29.3)	14 (27.5)	1.07 (0.59 - 1.94)	.83
100%	12 (20.7)	6 (11.8)	1.76 (0.71 - 4.35)	.21	10 (17.2)	8 (15.7)	1.10 (0.47 - 2.57)	.83

[a]EC: electronic cigarette.

[b]NRT: nicotine replacement therapy.

[c]RR: relative risk.

Table . Incidence rate ratios for time and treatment effects using ecological momentary assessment (EMA) and study surveys at week 12.

	EMA		Study survey	
	IRR[a] (95% CI)	*P* value	IRR (95% CI)	*P* value
Negative binomial regression				
Intercept	7.74 (5.34 - 11.23)	<.001	15.89 (12.92 - 19.54)	<.001
Time	0.929 (0.928 - 0.929)	<.001	0.36 (0.29 - 0.44)	<.001
Group				
NRT[b]	—[c]	—	—	—
EC[d]	0.87 (0.52 - 1.45)	.60	1.11 (0.85 - 1.46)	.43

[a]IRR: incidence rate ratio.

[b]NRT: nicotine replacement therapy.

[c]Not applicable.

[d]EC: electronic cigarette.

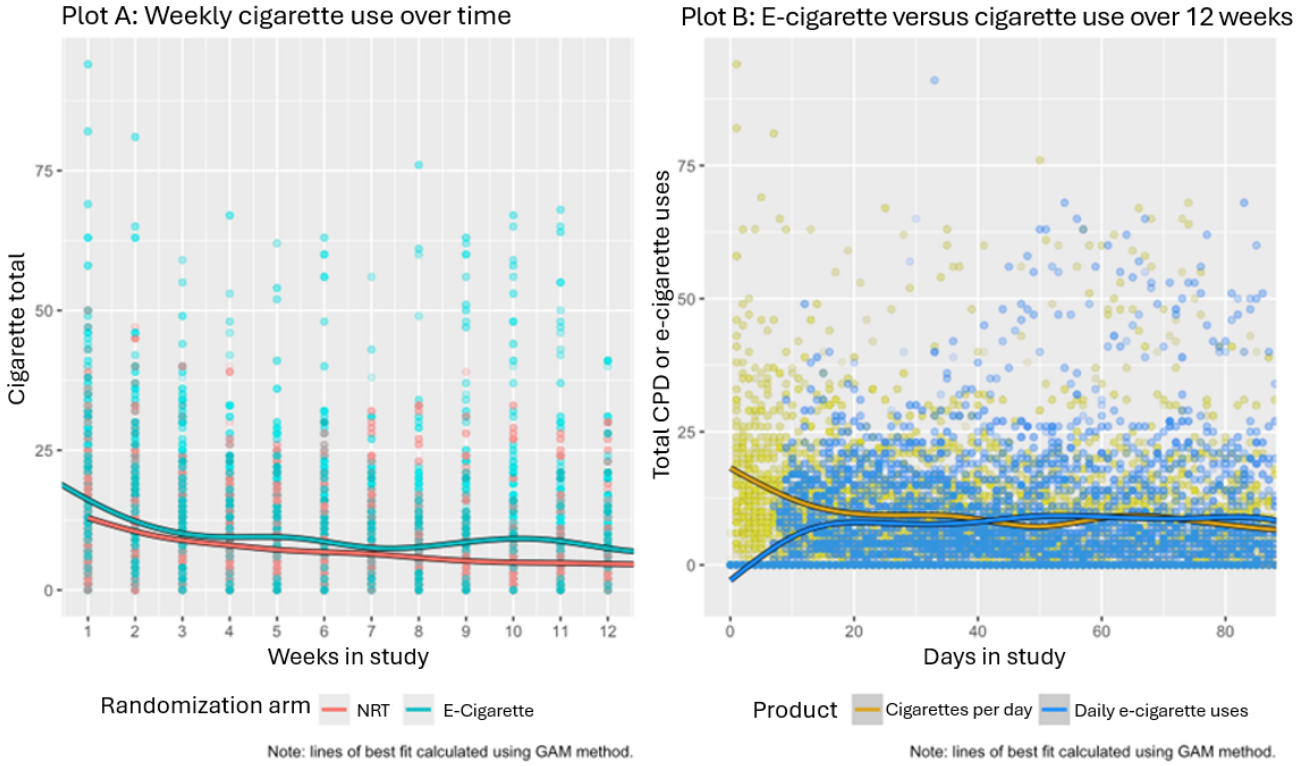
Cigarette and e-Cigarette Use Over Time

Plot A in Figure 3 shows the reported number of CCs smoked in each week by all participants. The trend lines were fitted using a generalized additive model and are included only for visualization. This visualization aligned with the hypothesis testing of negative binomial mixed effects modeling in that we see a general downward trend in the number of cigarettes smoked for participants in each group over time, with similar trends for each group. To visualize EC use more clearly, we presented EC use in days rather than weeks.

Among those in the EC arm, EC use increased gradually and stabilized after day 20, with daily CC consumption becoming slightly lower than daily EC use. The percentage of individuals who “switched” completely from CCs to ECs in each week ranged from 1.7% (1/58) in week 2 to 12.1% (7/58) in week 12. The percentage of mostly switching increased from 5.2% (3/58) in week 2 to 20.7% (12/58) in week 12.

XSL•FO
RenderX

Figure 3. Trends in electronic cigarette (EC) and combustible cigarette (CC) use over the 12-week study period. Plot A depicts the change in CC by treatment arms on a weekly basis over the study period. Data points represent the total number of CC per week. A fitted line for each trial arm is included, which was fit using a generalized additive model (GAM). Plot B depicts the comparison of change in total number of daily CC and EC uses among EC arm group participants over the 12-week study period. Data points represent the total number of CC and EC uses per day. Fitted lines for CC and EC were fitted using GAM method. CPD: cigarettes smoked per day; NRT: nicotine replacement therapy.



Discussion

Principal Findings

The results showed that participants in either study arm who reported greater CPD in the 3-month visit survey tended to report greater CPD in the final week of EMA prompting, and vice versa. Specifically, the EMA data indicated that EC use could lead to a reduction in CC use similar to that of NRT. These results suggest that ECs could be a potential alternative smoking cessation tool in the future. Similar to analyses using study-visit data, we found a statistically significant time effect on CPD reduction when analyzing the EMA data. Through the insights provided by EMA, patterns in dual use and switching could also be observed where participants in the EC arm appeared to consume fewer CPD over time and had a steady increase in EC use. Together, these results provide evidence that EMA is likely a suitable method to collect recall-based smoking data from participants with greater frequency and shorter recall periods compared with standard study visits.

In addition, the rigorous EMA protocol implemented in the study, with a high response rate over the 12-week intervention period, demonstrates the feasibility of assessing smoking behaviors in this manner. The relevance and validity of EMA data rely on the comparability with traditional methods of measurement (eg, surveys) and the overall adherence to the EMA protocol among participants. This study provides insight by directly comparing smoking-related measures collected through EMA and surveys. The nonsignificant paired difference found in this study contrasts with other similar studies that found large discrepancies between retrospective and EMA data [25,26]. These results indicate that an EMA protocol can achieve equal or more accurate data collection than traditional methods. Furthermore, levels of EMA adherence have been found to differ based on participant characteristics in some groups [27]. While previous studies have used EMA protocols to assess the prevalence of smoking CCs or ECs [9,26,28], as well as changing smoking patterns over time [29], this study is among the first to use a smoking EMA protocol in a population with chronic conditions, thus, demonstrating the feasibility of an SMS text messaging–based smoking EMA protocol in this population.

The visualization of CPD and EC use over time (see Figure 3) offer examples of the added information provided by EMA, giving researchers more precise data collection points. EMA can also help identify and quantify the switching (from CCs to ECs) patterns of participants closer to when switching actually occurs, which itself is an area of future research. For example, for this trial, using traditional surveys, we would only be able to assess switching at the time of surveys (at 3- and 6-month follow-ups); however, EMA allowed us to determine when and if participants did in fact switch from CC to EC within and immediately following the intervention period. If assessed in near real time, this information could even be leveraged to tailor counseling delivered to participants.

We note several limitations. In total, 109 (90%) of the 121 trial participants agreed and enrolled in SMS text messaging prompts; however, only 65 participants were adherent (EMA response rate greater than 60%) in the final week of prompting. All

XSL•FO
RenderX

responses were used in the negative binomial mixed effect model, whether they responded to all prompts in a day or only a subset. This type of missingness could bias the average number of cigarettes smoked downward, particularly as fewer participants responded consistently as the study period continued. While not statistically significant, exploratory analyses seemed to suggest that employed participants may be more likely to drop off in response rates compared with retired or unemployed participants; however, further work is required to better understand effects on response rate and retention. In general, there is a risk of an EMA protocol placing too much burden upon participants leading to fatigue or burnout [30]; in turn, participant fatigue could lead to higher drop-out rates and missingness, which could bias results. Similarly, there was variance in contact intensity between the two study arms, as the NRT participants were not asked about NRT or EC use limiting the total number of questions asked. While no statistically significant difference in EMA protocol adherence was measured, differential contact between the arms could potentially bias perceived burden or engagement with the EMA protocol. Despite the missingness and the potential for participant fatigue, this EMA protocol saw continued adherence and high response rates from a majority of enrolled participants while maintaining a fairly rigorous prompt frequency. Continued adherence indicates relatively low burden of text prompts (vs through other platforms; eg, phone app) and the familiarity most participants have with texting, which could be a more acceptable and feasible method compared with lengthier surveys. Future analyses of EMA data could seek to better understand additional smoking-related measures beyond craving and satisfaction. Due to limitations of measuring experiences using numerical scales [31,32], future EMA protocols may consider free-text responses and apply tools like natural language processing to EMA data to interpret and clean patient responses thereby transforming the data into meaningful information on human behaviors and sentiments. Better implementation of set responses can help ensure researchers are receiving the information desired in workable forms and reduce the data cleaning needed. Another future area of research is determining how well EMA data can detect the real-time switching from CCs to ECs.

Conclusion

SMS text messaging–based EMA represents a feasible and suitable method to collect recall-based smoking-related data. As compared with the survey data, the use of EMA provides unique advantages, namely greater granularity in the smoking timelines and a capability to detect switching patterns between CCs and ECs in near real time. EMA should be considered as a method for data collection when developing future tobacco harm reduction research.

Acknowledgments

Funding was received from internal funds provided to SES by New York University (NYU) Grossman School of Medicine. The study sponsor had no role in the study design; collection, analysis, and interpretation of data; writing the report; or the decision to submit the report for publication. The NYU Langone Health institutional review board approved the study (#s20-00839).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

All authors contributed to the study's conception and design. TM and MH performed data cleaning and analysis. TM contributed to material preparation and writing the first draft of the manuscript. All other co-authors reviewed the analysis and edited subsequent versions of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables presenting (1) mean satisfaction with the last combustible cigarette (CC) and craving for CC over the 12-week study period stratified by treatment arms using ecological momentary assessment (EMA) data and (2) incidence rate ratios for craving and satisfaction over time and treatment effects using EMA data at week 12. Figures presenting (1) a flow diagram depicting the text messaging prompts and flow received by the patients from the electronic cigarette (EC) group (participants were asked whether they smoked a CC or used an EC since their last report and based on the responses they are directed to specific follow-up questions), and (2) trends of weekly means and medians of response rates (%) among all participants over the 12-week study period.
[DOCX File, 132 KB - formative_v9i1e66709_app1.docx]

Checklist 1
CONSORT-EHEALTH (version 1.6.1) checklist.
[PDF File, 1167 KB - formative_v9i1e66709_app2.pdf]

XSL•FO
RenderX

References

1. Smoking cessation: a report of the surgeon general. US Department of Health and Human Services. 2020. URL: https://www.hhs.gov/sites/default/files/2020-cessation-sgr-full-report.pdf [accessed 2025-03-11]
2. Cornelius ME, Loretan CG, Jamal A, et al. Tobacco product use among adults - United States, 2021. MMWR Morb Mortal Wkly Rep 2023 May 5;72(18):475-483. [doi: 10.15585/mmwr.mm7218a1] [Medline: 37141154]
3. Tashkin DP, Murray RP. Smoking cessation in chronic obstructive pulmonary disease. Respir Med 2009 Jul;103(7):963-974. [doi: 10.1016/j.rmed.2009.02.013]
4. Brown J, Beard E, Kotz D, Michie S, West R. Real-world effectiveness of e-cigarettes when used to aid smoking cessation: a cross-sectional population study. Addiction 2014 Sep;109(9):1531-1540. [doi: 10.1111/add.12623] [Medline: 24846453]
5. Cobb NK, Abrams DB. The FDA, e-cigarettes, and the demise of combusted tobacco. N Engl J Med 2014 Oct 16;371(16):1469-1471. [doi: 10.1056/NEJMp1408448] [Medline: 25317866]
6. Fiore MC, Schroeder SA, Baker TB. Smoke, the chief killer--strategies for targeting combustible tobacco use. N Engl J Med 2014 Jan 23;370(4):297-299. [doi: 10.1056/NEJMp1314942] [Medline: 24450888]
7. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
8. Tversky A, Kahneman D. Availability: A heuristic for judging frequency and probability. Cogn Psychol 1973 Sep;5(2):207-232. [doi: 10.1016/0010-0285(73)90033-9]
9. Mead EL, Chen JC, Kirchner TR, Butler J 3rd, Feldman RH. An ecological momentary assessment of cigarette and cigar dual use among African American young adults. Nicotine Tob Res 2018 Aug 14;20(suppl_1):S12-S21. [doi: 10.1093/ntr/nty061] [Medline: 30125017]
10. Todd M. Daily processes in stress and smoking: effects of negative events, nicotine dependence, and gender. Psychol Addict Behav 2004 Mar;18(1):31-39. [doi: 10.1037/0893-164X.18.1.31] [Medline: 15008683]
11. Pratt SI, Ferron JC, Brunette MF, Santos M, Sargent J, Xie H. E-Cigarette provision to promote switching in cigarette smokers with serious mental illness-a randomized tTrial. Nicotine Tob Res 2022 Aug 6;24(9):1405-1412. [doi: 10.1093/ntr/ntac082] [Medline: 35363874]
12. Biener L, Abrams DB. The Contemplation Ladder: validation of a measure of readiness to consider smoking cessation. Health Psychol 1991;10(5):360-365. [doi: 10.1037//0278-6133.10.5.360] [Medline: 1935872]
13. Gupta N, Pinto LM, Morogan A, Bourbeau J. The COPD assessment test: a systematic review. Eur Respir J 2014 Oct;44(4):873-884. [doi: 10.1183/09031936.00025214]
14. Stevens ER, Lei L, Cleland CM, et al. Electronic cigarettes as a harm reduction strategy among patients with COPD: protocol for an open-label two arm randomized controlled pilot trial. Addict Sci Clin Pract 2022 Dec;17(1):2. [doi: 10.1186/s13722-021-00284-0]
15. Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials. a review. Control Clin Trials 2002 Dec;23(6):662-674. [doi: 10.1016/s0197-2456(02)00242-8] [Medline: 12505244]
16. Gupta N, Pinto LM, Morogan A, Bourbeau J. The COPD assessment test: a systematic review. Eur Respir J 2014 Oct;44(4):873-884. [doi: 10.1183/09031936.00025214] [Medline: 24993906]
17. Ställberg B, Nokela M, Ehrs PO, Hjemdal P, Jonsson EW. Validation of the clinical COPD Questionnaire (CCQ) in primary care. Health Qual Life Outcomes 2009 Mar 25;7:1-9. [doi: 10.1186/1477-7525-7-26] [Medline: 19320988]
18. van der Molen T, Willemse BWM, Schokker S, ten Hacken NHT, Postma DS, Juniper EF. Development, validity and responsiveness of the Clinical COPD Questionnaire. Health Qual Life Outcomes 2003 Apr 28;1:1-10. [doi: 10.1186/1477-7525-1-13] [Medline: 12773199]
19. Chan PS, Jones PG, Arnold SA, Spertus JA. Development and validation of a short version of the Seattle angina questionnaire. Circ Cardiovasc Qual Outcomes 2014 Sep;7(5):640-647. [doi: 10.1161/CIRCOUTCOMES.114.000967] [Medline: 25185249]
20. Nathan RA, Sorkness CA, Kosinski M, et al. Development of the asthma control test: a survey for assessing asthma control. J Allergy Clin Immunol 2004 Jan;113(1):59-65. [doi: 10.1016/j.jaci.2003.09.008] [Medline: 14713908]
21. Soler X, Holbrook JT, Gerald LB, et al. Validity of the asthma control test questionnaire among smoking asthmatics. J Allergy Clin Immunol Pract 2018;6(1):151-158. [doi: 10.1016/j.jaip.2017.05.010] [Medline: 28669892]
22. Cappelleri JC, Bushmakin AG, Baker CL, Merikle E, Olufade AO, Gilbert DG. Revealing the multidimensional framework of the Minnesota nicotine withdrawal scale. Curr Med Res Opin 2005 May;21(5):749-760. [doi: 10.1185/030079905X43712] [Medline: 15969874]
23. Hughes JR, Hatsukami D. Signs and symptoms of tobacco withdrawal. Arch Gen Psychiatry 1986 Mar;43(3):289-294. [doi: 10.1001/archpsyc.1986.01800030107013] [Medline: 3954551]
24. Yirga AA, Melesse SF, Mwambi HG, Ayele DG. Negative binomial mixed models for analyzing longitudinal CD4 count data. Sci Rep 2020 Oct 7;10(1):16742. [doi: 10.1038/s41598-020-73883-7] [Medline: 33028929]
25. Poulton A, Pan J, Bruns LR Jr, Sinnott RO, Hester R. Assessment of alcohol intake: Retrospective measures versus a smartphone application. Addict Behav 2018 Aug;83:35-41. [doi: 10.1016/j.addbeh.2017.11.003] [Medline: 29128148]
26. Yang JJ, Ou TS, Lin HC, Kyung Nam J, Piper ME, Buu A. Retrospective and real-time measures of the quantity of e-cigarette use: an ecological momentary assessment study. Nicotine Tob Res 2023 Aug 23;25(10):1667-1675. [doi: 10.1093/ntr/ntad094] [Medline: 37327251]

XSL•FO
RenderX

27. Sokolovsky AW, Mermelstein RJ, Hedeker D. Factors predicting compliance to ecological momentary assessment among adolescent smokers. Nicotine Tob Res 2014 Mar;16(3):351-358. [doi: 10.1093/ntr/ntt154] [Medline: 24097816]
28. Ozga JE, Bays C, Haliwa I, et al. Measurement of cigarette smoking: comparisons of global self-report, returned cigarette filters, and ecological momentary assessment. Exp Clin Psychopharmacol 2022 Jun;30(3):365-370. [doi: 10.1037/pha0000449] [Medline: 33630647]
29. Piper ME, Baker TB, Benowitz NL, Jorenby DE. Changes in use patterns over 1 year among smokers and dual users of combustible and electronic cigarettes. Nicotine Tob Res 2020 Apr 21;22(5):672-680. [doi: 10.1093/ntr/ntz065] [Medline: 31058284]
30. van Genugten CR, Schuurmans J, Lamers F, et al. Experienced burden of and adherence to smartphone-based ecological momentary assessment in persons with affective disorders. J Clin Med 2020 Jan 23;9(2):322. [doi: 10.3390/jcm9020322] [Medline: 31979340]
31. Heishman SJ, Singleton EG, Pickworth WB. Reliability and validity of a Short Form of the Tobacco Craving Questionnaire. Nicotine Tob Res 2008 Apr;10(4):643-651. [doi: 10.1080/14622200801908174] [Medline: 18418787]
32. Nguyen C, Kortlever JTP, Gonzalez AI, Ring D, Brown LE, Somogyi JR. Attempts to limit censoring in measures of patient satisfaction. J Patient Exp 2020 Dec;7(6):1094-1100. [doi: 10.1177/2374373520930468] [Medline: 33457550]

Abbreviations

ACT: Asthma Control Test
CAD: coronary artery disease
CAT: COPD Assessment Test
CC: combustible cigarette
CCD: Clinical COPD Questionnaire
COPD: chronic obstructive pulmonary disease
CPD: cigarettes smoked per day
EC: electronic cigarette
EHR: electronic health record
EMA: ecological momentary assessment
IRR: incidence rate ratio
MNWS: Minnesota Nicotine Withdrawal Scale
NRT: nicotine replacement therapy
NYULH: New York University Langone Health
PVD: peripheral vascular disease
RCT: randomized clinical trial
RR: relative rate
SAQ-7: Seattle Angina Questionnaire-7

Edited by A Mavragani; submitted 24.09.24; peer-reviewed by S Cha; revised version received 11.02.25; accepted 12.02.25; published 21.03.25.

Please cite as:
Morgan T, He M, Nicholson A, El Shahawy O, Sherman SE, Stevens ER
Using Text Messaging Ecological Momentary Assessment to Record Changes in e-Cigarette and Combustible Cigarette Use: Pilot Randomized Clinical Trial
JMIR Form Res 2025;9:e66709
URL: https://formative.jmir.org/2025/1/e66709
doi:10.2196/66709

© Tucker Morgan, Michelle He, Andrew Nicholson, Omar El Shahawy, Scott E Sherman, Elizabeth R Stevens. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Development and Implementation of Strong Foundations, a Digitally Delivered Fall Prevention Program: Usability and Feasibility Pilot Exercise Cohort Study

Ryan Moran[1,2], MPH, MD; David Wing[1], MS; Hope Davey[1], BS; Hava Barkai[1], MS; Jeanne Nichols[1], PhD

[1]Exercise and Physical Activity Resource Center (EPARC), University of California San Diego, 9500 Gilman Dr. La Jolla, San Diego, CA, United States

[2]Department of Medicine, University of California San Diego, San Diego, CA, United States

Corresponding Author:
Ryan Moran, MPH, MD
Exercise and Physical Activity Resource Center (EPARC), University of California San Diego, 9500 Gilman Dr. La Jolla, San Diego, CA, United States

Abstract

Background: Falls remain a major public health problem and a significant cause of preventable injury. Maintaining strength and balance by staying active can prevent falls in older adults, and public health advocates support referral to community exercise programs. Given the growth in use and acceptance of technological interfaces, there remains an interest in understanding the role of a synchronous exercise program designed to improve strength, postural alignment, and balance specifically designed to be delivered in a digital environment with respect to usability and feasibility.

Objective: This study aims to design and implement a synchronously delivered digital fall prevention program to adults aged 60 years and older, to understand the usability, feasibility, and attendance.

Methods: The "Strong Foundations" program, a 12-week, live, digitally delivered fall-prevention exercise program was informed from different existing in-person exercises and piloted to older adults who were considered a low fall risk by scores of 4 or less from the Centers for Disease Control and Prevention's (CDC's) Stopping Elderly Accidents and Deaths Initiative (STEADI) Staying Independent questionnaire. The System Usability Scale (SUS) measured usability and feasibility at the completion of this program, and digital measures of age-related function (timed up and go [TUG] and 30-second chair stand [30 CS]) were collected pre- and postintervention. Data were collected in 2021.

Results: A total of 39 older adults were recruited and 38 completed the 12-week program with an average age of 72 years. The average SUS was 80.6, with an 85% attendance rate and an 8.5 (out of 10) self-reported satisfaction score. Digitally collected TUG and 30 CS statistically improved pre- and postintervention by 9% and 24%, respectively; by week 12, 64% (23/36) of participants improved in the timed up and go and 91% (32/35) improved the chair stands.

Conclusion: There was excellent usability and acceptability for Strong Foundations, a novel fall-prevention program designed to be delivered digitally and promising improvement of objective measures of fall risk.

(JMIR Form Res 2025;9:e67406) doi:10.2196/67406

KEYWORDS

digital health; fall prevention; fall risk; older adults; geriatrics; system usability scale; Strong Foundations; feasibility; public health; user acceptance; exercise; usability; digital technology; mobile phone

Introduction

Fall-related injuries among older adults are a significant and growing public health concern. Over one-quarter of community-residing men and women 65 years of age and older fall annually [1,2]. Risk is compounded over time as a single fall predicts recurrent falls [3-7], and fall-related mortality has increased over 30% just between 2007 and 2016 [8]. Not only do falls represent a major cause of accidental death and mobility-related disability among older adults, but they also result in psychological distress, loss of independence, and a decrease in quality of life [9,10]. Despite the existence of various interventions and policies aimed at reducing fall risk in this population, the incidence of falls remains alarmingly high.

While fall risk is multifactorial, identification of risk factors, as well as referral to and participation in appropriate fall-risk reduction programs are established as an effective, evidence-based approach to reduce fall risk [2,11]. Specifically, targeted strength and balance exercises have consistently been shown to improve fall risk, and accordingly, the Centers for Disease Control and Prevention (CDC) has outlined an

XSL•FO
RenderX

evidence-based clinical approach to identify those at risk for falls to help assess known risk factors and refer for community-based fall-prevention programs [12,13]. The public health use of this approach is further supported by the United States Preventive Services Task Force (USPSTF) "Grade B" recommendation to refer older adults for fall prevention exercise programs [14,15]. However, despite these resources and recommendations, clinical identification and referrals have been insufficient [16]; only approximately one-third of older adults report being asked about fall-risk by their health care providers, and similarly, only around a third of those who fall report discussing their fall and fall circumstances with their health care provider [17,18]. At the population level, conservative estimates suggest far fewer than 50% older adults meet recommended physical activity levels [19]. These findings mirror the growing epidemiological evidence of increased rates of falls in older adults. While some existing programs that improve fall risk and activity for older adults, including EnhanceFitness PA and Silver Sneakers may be available to some older adults with certain insurance plans at no additional costs [20], others may face substantial fees; furthermore, many eligible seniors do not use these programs, citing several barriers. Such barriers include general dislike of gyms and a lack of awareness or instructor suitability and generalizability, problems accessing appropriate classes, concerns about bad weather and driving, and perceptions about being "old" in these environments [21]. These issues are even greater among older adults residing in areas and communities that have fewer facilities for guided physical activity, more limited transportation infrastructure and less exercise expertise [22]. Finally, there remain geographic and demographic disparities related to falls, including higher rates in rural areas [2], highlighting the interest and need for different approaches to address this public health and clinically relevant problem in older adults in novel ways that increase reach and are safe and effective.

Appropriate technologically driven approaches have promise to increase range and dissemination of fall prevention resources, including making them easier and safer to access [23]. Furthermore, digitally delivered programs that involve appropriately tailored exercises offer the opportunity to balance risks and benefits during times of social distancing, while improving dissemination and community uptake, and possibly improve fall risk [24]. While recent attention has highlighted concerns with broadband access in older adult populations, especially with attention to disparities and equity [25], digital interventions showcase the promise as those aged 65 years and older still represent the fasting growing demographic in technological uptake [26,27]. Thus, given the trend of increasingly ubiquitous access to smart-enabled technology in older adults, and gaining acceptance in this population [26], there is great potential to improve equitable access to high-quality exercise-based fall prevention resources by decreasing geographic boundaries—not only immediately, but increasingly in the future. Alternative approaches, including prerecorded, online, and asynchronous delivery, are an additional consideration but lack the ability for real-time guidance, which is possible with synchronous real-time delivery, and therefore offer the opportunity to provide guidance with risk mitigation and personalization when appropriate for participants. Therefore, approaches that leverage technology to improve access to fall-risk reduction exercise programs are imperative, and growing data suggest digitally formatted delivery may be feasible [28].

Zoom (Zoom Video Communications) is a user-friendly, web-based service that provides digital video-based connections between multiple users; Zoom-for-Telehealth is a HIPAA-compliant encrypted service used for clinically delivered care and permits conferencing between individuals. Zoom is readily available online, free to download, and supports a one-to-many viewing platform such that it can be set up so that only the host can see individual participants at a given time [29]. This application was selected based on the ubiquitous nature and availability to host professionally through our institution. The nature of this digital platform lends interest in understanding how fall-prevention curricula may be developed and delivered to older adults; therefore, we developed a novel digital platform, Strong Foundations, designed to leverage the advantages of digital technology with evidence-based fall-prevention strategies, along with real-time feedback to users based on appropriate or inappropriate practice, optimizing gains for a larger group and highlighting core exercises. This program includes features to promote adherence, engagement, and social connectivity enabled by real-time interaction, all important to promote efficacy and subsequent improvement in gains from any intervention [30]. Given the novelty of such a program, we aim to understand the feasibility of this 12-week program in terms of usability and acceptance using a validated scale and attendance rates and explore measures of digitally collected physiologic improvement in adults aged 60 years and older.

Methods

Program Description

Our program, Strong Foundations, is drawn from several rigorously tested fall and fracture prevention exercise programs [11,31-33] but is new and entirely designed to be delivered digitally. This design emphasized exercises that were able to be remotely instructed, witnessed, and performed in a limited physical space such that most users would be able to use readily available household spaces. The program was designed with physician input and by exercise physiologists and a Doctor of Physical Therapy, all with extensive training in both group and individualized exercise for older adult populations. Strong Foundations is a 12-week iterative curricular program with 3 core components: postural alignment and control, balance and mobility, and muscular strength and power. All the exercises offered over the course of the intervention are designed to be appropriate for the target population (ie, older adults) and are standardized so all participants receive the same basic instruction, but the level of difficulty is scaled to participant experience, capability, and musculoskeletal limitations. This individualization was enabled by real-time delivery, such that instructors are able to adapt, respond, and tailor specific aspects of any class to fit the needs of participants, including independently pulling our participants into a "Breakout Room" for direct one-on-one instruction before reassimilating into the larger group. The 12 weeks of instruction establish and build

XSL•FO
RenderX

such that throughout the program, posture and positions prioritized early in the program are used to build into more complicated maneuvers designed to improve the foundational techniques and activities that will allow participants to maximize learning opportunities and optimize gains related to balance, posture, and strength.

Real-Time Feedback

Differentiating this program from other streamed and real-time web-based interventions available, the novel feature of this program is the delivery of semi-individualized instruction in real time within a small group setting. This is accomplished with a specifically trained "support" instructor providing guidance and correction on optimal exercise form while the "lead" instructor demonstrates for the larger group. This semitailored approach allows the lead instructor to provide "best practice" instruction and provide a visual demonstration while assistant instructors assure that participants are not only gaining important skills related to fall-prevention technique, but also doing this safely and remotely, as numerous screens on a digital interface would otherwise compromise the integrity of a single instructor. All instructors are aware of the curriculum in its entirety and versed on adaptive approaches for a participant who may not be able to do specific components safely or without predisposing to injury. An example of this may include someone with a known right shoulder injury, whereby certain resistance band and posture actions need to be modified—thereby the lead instructor may continue providing instruction to the group at large, with uninterrupted group synchronous interaction, but the support instructor can provide individualized recommendations to a specific participant without disrupting the overall dynamic.

Program Development

The program was iteratively designed based on two, 4-week pilot classes with 9 - 12 participants each (total n=21), which refined and informed the curriculum to inform the design of a final total 12-week program, where foundational components are expanded in intensity and challenge, assuring continual growth and improvement of participants. Participant experience was surveyed, with respondents (n=14) noting that the interface was easy to follow, with sufficient audio and visual interface for safely following instruction. Furthermore, 100% (14/14) of participants agreed or strongly agreed that "instruction was clear and easy to follow" and "I felt safe doing the exercises." In addition, 86% (12/14) agreed or strongly agreed to "I could hear the instructors well" and "instruction was easy to see and follow." In total, 64% (9/14) reported awareness of instructors watching validated tools to establish usability, however, were not collected in this early phase of development.

Final program development was designed to incorporate approximately 1 hour of instruction for 12 weeks with key aspects to establish the weekly focus, and materials were generated for dissemination between classes to participants to encourage and underscore the importance of ongoing practice to support gains during this intervention (Table 1).

XSL•FO
RenderX

Table . Strong Foundations curriculum focus across 12 weeks.

Week	Foundational exercise 1	Foundational exercise 2	Foundational exercise 3	Foundational exercise 4	Focus of the week
1	Neutral spine and marionette pose	Pelvic tucks and tilts	Hip hikes	Hip hinge	Posture
2	Quick feet drills	Review hip hinge	“Traditional abs”	Self-test: static balance	Balance
3	Front and lateral arm raises (with band)	Hip hinge (review) with squat and chair sit, stand	Reverse lunge	Heel raises	Strength
4	Neutral spine and pelvic tuck and tilt—standing	Static balance	Reverse lunge with progression	Heel raises with tennis ball	Combine 3 pillars
5	Sit and stand transitions	Around the clock steps	Hip hinge with choice of arm extensions	Modified and regular jumping jacks	Floor to stand transition—floor exercises
6	Hip hinge (review) with squat and chair sit, stand	Lunge with chair assistance progression	Single leg heel raises	Hing hinge + picking up objects	Compound strength movements
7	Drinking bird	Squat to lateral leg raise	Single leg heel raises	Toe raise walk around chair	Increase balance challenge
8	Side squat	Wall sit with wood chop	High knee walk	Quick feet multidirectional	Odd impact and multidirectional movements
9	Weight transfer with split stance	Squat into a High knee and arm reach	High knee walk	Head rotations in a tandem stance	Combine balance with compound movements
10	Single leg stance with band pull downs and serial 7’s	Lunge with chair support and head rotation	Standing superman with cognitive challenge	Goal post with arm slides	Introduce cognitive challenge
11	Reaching squats	Lateral jacks	Vertical push-ups	Knee drivers	Speed and power movements
12	Good morning	Bow and arrow	Hands up	Heel toe sitting with resistance band	Collaborative workout

Site, Eligibility, and Enrollment

Participants were recruited through established venues from the Exercise and Physical Activity Research Center (EPARC), an exercise science–based laboratory with experience in physiologic phenotyping based on the University of California, San Diego campus. Given that the intervention was developed and deployed during the COVID-19 pandemic, all recruitment, enrollment, and measurement was done remotely; no face-to-face interaction or measurement occurred throughout the pilot. Given the novelty of the delivery platform and development of this program, our primary aims centered on determining its’ usability and feasibility in this context. Therefore, out of an abundance of caution and given the entirely remote recruitment, enrollment, delivery, and measurement through this program, this pilot included those with minimal-to-no-risk factors for falling based upon the CDC’s Staying Independent brochure questionnaire [34]; for this pilot, a participant was considered no more than a moderate fall risk based on a score of ≤4 [16]. Textbox 1 shows the eligibility criteria. Informed consent was performed remotely through secure digital methods (Docusign) with a real-time Zoom-based meeting with staff to review protocols and consent documents and establish an appropriate room setup. This included ensuring that there was sufficient space to move while also being at a sufficient distance from the camera that their entire body was visible to the instructor. In addition, before remote measurement, participants were individually coached on how to measure an appropriate length course (ie, 3 meters) and have an appropriately sized chair (ie, approximately 17-inch chair height without arms). Enrollment occurred from August to September 2021, with 39 older adults recruited and cohorted based on predetermined times of the intervention such that participants could select the program time most convenient for their schedule.

XSL•FO
RenderX

Textbox 1. Eligibility criteria for Strong Foundations pilot.

Inclusion criteria:

- Age 60 years or older and ambulatory, including with the use of a cane or walker (1 participant enrolled age 56, which was an unintended protocol deviation).
- Completion of the Stopping Elderly Accidents and Deaths Initiative (STEADI) Stay Independent Brochure.
- Access to internet and computer and a Zoom-interface or broadband.

Exclusion criteria:

- Individuals who are wheelchair bound.
- Exclusively communicate in a language other than English.
- Score of 4 or more on the STEADI modified questionnaire.

Primary Feasibility Outcomes

Programmatic development and launch were aimed at determining feasibility and usability of a digital platform with an older adult population. Attendance was measured weekly as well as self-reported number of exercises performed outside of the class based on the previous week's instruction and usability determined at completion of the program based on the System Usability Scale [35] (SUS), a nonpriority validated questionnaire, which was designed to understand the ease of use of new systems or programs using a 5-element Likert scale.

In general, scores >70 on the SUS are considered to have appropriate acceptability of a program or platform. The SUS was selected as the primary outcome, given it is considered the most widely used scale in eHealth to determine usability [36]. Finally, a survey to understand subjective experience was developed and deployed at the end of the intervention.

Exploratory User Outcomes

To understand user experience and improvement, at baseline and at the completion of this curriculum, digitally collected objective measures of frailty and fall risk were collected to explore the opportunity to use the digital interface to collect these measures and understand changes from this intervention. All participants at baseline and after completing the intervention period (±1 week) were assessed remotely with the measures outlined and endorsed by national and international public health authorities, namely the timed up and go (TUG) [37-39] and 30-second chair rise [40]. Participants were then monitored and timed remotely while completing the assessments. TUG was completed with instruction to complete at both a normal speed (TUG normal) and "as fast as safely able to perform the test" (TUG-Fast). Given enrollment included those with few risk factors for falling, these physical measurements were selected because of their relative ease of administration and interpretability but remained exploratory from the lens of feasibility.

Sample Size Determination

Power calculation was not conducted given the aim to determine feasibility and usability for this program and the selected outcome, the SUS, has been shown to be valid in small samples [41]. Therefore, we used an assumed attendance rate of 70%, which has held consistent in national programs in both rural and urban areas for older adults in fall prevention programs [42] and similar fall prevention programs that have evaluated this time frame to be sufficient to improve measures, markers of frailty, and fitness in different health contexts between 20 and 60 individuals [43-45]. Given budgetary limitations, our plan to enroll 4 cohorts all with at least 8 - 10 older adults would be sufficient to determine both the feasibility and usability and explore physiologic changes.

Statistical Analysis

Primary outcomes to understand feasibility and usability were determined based on the validated SUS [35], which in general assumes an average score of 68 and higher indicates more usability. Attendance rates were determined by a participant joining the Zoom interface measured weekly for each individual, reported as averages (mean) and SD. Numerical averages (mean) of patient-reported questionnaire results (out of 10) were calculated. Associations with SUS scores were calculated by Pearson correlation coefficients. Secondary outcomes were measured using paired-sample 2-tailed *t* tests for differences (for the TUG and 30-second chair rise, collected digitally) with corresponding mean differences and 95% CIs. Statistical analysis was performed on SPSS (version 28; IBM Corp).

Ethical Considerations

This study was reviewed and approved by the University of California San Diego Institutional Review Board (#802148); informed consent was obtained by all participants. All study data were deidentified. Participants were not compensated to participate.

Results

Participants

Across 12 weeks, 38 of 39 (97.4%) recruited older adults completed the program. One participant left after 2 weeks of instruction because of a family emergency. Of the 38 who completed the program, the average age was 72 years (SD 5.5), and 37 were female (97.4%). Participants were spread across 4 cohorts, ranging in size from 9 - 10 individuals.

Feasibility and Usability

On average, participants attended 10 weekly sessions out of 12 (84%). The average number of home exercise sessions

XSL•FO
RenderX

performed weekly between classes was 2.3 (SD 0.7). The average SUS score was 80.6 (SD 15.4), with a range of 40 to 100. On a 10-point scale, the average self-reported overall experience of the intervention and exercise classes was 8.5 (SD 1.9). No participants experienced an injury or had any safety related concerns, and additional subjective feedback was excellent with most participants eager for ongoing involvement and instruction. SUS scores were statistically inversely associated with age (Pearson correlation r=−0.411; P=.10) and positively with reported satisfaction score (Pearson correlation r=0.654; P=.004) but not statistically associated with attendance or improvements in TUG or 30s chair rise scores. Table 2 shows the number of participants and their SUS score range.

Table . System Usability Scale (SUS) score range.

SUS Score	Participants (n=38), n (%)
<50	2 (5)
50 - 60	2 (5)
61 - 70	6 (16)
71 - 80	7 (18)
>80	21 (55)

Objective Measures

Pre- and postmeasures were collected for 36 of the participants for the TUG and 35 of the participants for the 30-second chair rise. Baseline TUG at the "normal" speed was 8.8 (SD 1.9) seconds, and at the fastest possible speed was 6.2 (SD 1.4) seconds. Baseline 30-second chair stands were 14.5 (SD 3.4) stands. Post intervention, TUG and TUG-Fast were 8.0 seconds (SD 1.3 seconds; difference of 0.8 [95% CI 0.24-1.21] seconds) and 5.7 seconds, respectively (SD 1.1 seconds; difference of 0.48 [95% CI 0.22-0.75] seconds) and the 30 second chair rise increased to 18 stands (SD 4.1 stands; difference of –4.8 [95% CI –2.6 to –6.9] stands) (Table 3). At 12 weeks, 64% of all participants improved in the TUG (23/36, for both measures) and 91% (32/35) improved the chair stands. Reasons reported for being unable to complete these were varied but included participants traveling internationally, unrelated knee pain, and a single participant dropping as noted above after 2 weeks.

Table . Digitally collected objective markers of fall-risk reduction.

Measure	Baseline, mean (SD)	Post intervention, mean (SD)	Difference (95% CI)	Change (%)
TUG[a] normal (seconds)	8.8 (1.9)	8.0 (1.3)	7.24 (0.24 - 1.21)	9
TUG-Fast (seconds)	6.2 (1.4)	5.7 (1.1)	0.48 (0.22 - 0.75)	8
30-second chair rise (stands)	14.5 (3.4)	18 (4.1)	−4.8 (−2.6 to 6.9)	24

[a]TUG: timed up and go.

Discussion

The development and rationale behind the need for a digital curriculum, Strong Foundations, is described here. In this 12-week pilot, we found a high level of usability and feasibility of this program based on the high attendance rates, SUS scores, and self-reported participant experience. In addition, we found impressive changes in digitally collected measures of functional improvement in most participants, suggesting that the program increased strength and mobility and may reduce the risk of falls. Clinically, the 8% - 20% improvement in measurement after 12 weeks is notable, especially considering recruitment generally was limited to those with limited risk factors for falling and have higher levels of function and therefore less scope for improvement or gain. While exploratory in nature, these gains are promising to show improvements across this intervention even in a relatively lower risk group of older adults.

This 12-week fall prevention program was well accepted and showed promise related to reducing fall risk in an older adult population. The usability and safety (as evidenced by no adverse events) of the program indicates that digitally delivered real-time exercise instruction has potential in this population to reduce barriers and increase availability of fall prevention programs. While those enrolled were relatively low risk for falls (as evidenced by the STEADI score used for eligibility), results should be contextualized that even those who were generally doing reasonably well have opportunity to improve. Indeed, in other populations, such as older adults with hip osteoarthritis, a change in TUG of around 1 second and a 30-second chair stand of around 1 second and 2 stands, respectively as meaningful [46], and between 0.9s and 6.0s depending upon the quality of life assessment tool in patients who were postoperative [47]. While our results of TUG change of 0.8 seconds (normal) is less than these noted changes, given the health status of our enrolled participants and limited 12-week intervention, the improvements are still notable. Furthermore, our improvement of 3.5 stands in 30 seconds, especially from a baseline of 14.5 stands—a notable good baseline—shows the resilience and opportunity for improvement in lower extremity strength regardless of starting health status.

However, this study is not without limitations. This project recruited heavily from an environment with a relatively high

XSL•FO
RenderX

socioeconomic status. As such, there may be a bias in enrolling individuals who already have digital tools like smartphones, tablets, and laptop computers and a high level of experience with digital technology, and therefore, a better understanding how to use these systems. As such, these participants may not be representative of the larger older adult population who may not have the same level of access to or understanding of such digital systems. We also did not ask specifically about their experience with previous technologies, though participants all did have sufficient equipment for enrollment. We also primarily enrolled women, which limits interpretability and generalizability of findings, especially given recent data suggesting higher fall-related impairment among older adult men [2]. Despite this, the use of a validated scale focused on usability commonly used across several domains related to eHealth and technology deployments [48], findings related to usability are supported; however SUS deployment in older adults has been challenged in some situations, namely those with cognitive impairment [49]. While we did not specifically screen for this in our enrollment, it is possible findings may be diluted. Furthermore, while the reasons remain unclear, the SUS average may be different and higher in digital physical activity interventions [48], making contextualizing results important; although, such data in older adult subpopulations remain limited. While we also did not only enroll participants who had experience with Zoom, it is possible that interested individuals may not have expressed willingness to enroll who were otherwise intimidated by such platforms, although we did not anecdotally experience this feedback. Subjective survey experience reports should be interpreted with this consideration [27]. Finally, it is not clear if digitally collected measures of physical performance, specifically the TUG and 30 second chair rise, are valid and reliable when compared with measures gathered in the clinic or laboratory. As such, the changes in physical performance demonstrated here, while promising, should be viewed as preliminary and not necessarily indicative of the same degree of change as would be expected were the data gathered in a more traditional context. However, given the rate of technology availability has continued to increase, attention is warranted to understand how these interfaces may be used for older adults [27].

In conclusion, this pilot showed that a digital interface is both usable and feasible but also suggested promise at improving functional measures associated with fall risk. Based on the increasing yearly number of falls and the relatively small number of participants who are engaging in fall-prevention programs, it seems clear that existing in-person exercise delivery modalities are insufficient. Strong Foundations offers a safe alternative to provide a real-time fall prevention program. Furthermore, the real-time feedback likely provides participants with added sense of purpose and value in their investment in improving their functional ability and therefore provide increased motivation for sustained behavior. Given the growing population and epidemiology of falls, having numerous options that enable engagement of older adults into high quality exercise programs designed to improve strength, postural alignment, and balance is increasingly important. Strong Foundations has promise to augment existing opportunities in this population and further leverages digital tools to increase the breadth of delivery potential. Ongoing attention and research is needed to both understand the scalability and sustainability of this program and to confirm that the observed changes in function and mobility through validated laboratory-based methods are replicated.

Acknowledgments

This pilot was funded by an internal grant obtained for pilots, partially supported by the National Institutes of Health (grant UL1TR001442). Generative artificial intelligence was not used in any aspect of this manuscript.

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

RM helped inform the design of the intervention, participated in study deployment, institutional review board (IRB) approval, manuscript generation, and data evaluation. DW helped develop the intervention; helped with IRB approval, manuscript review, data analysis; and performed the intervention and collected measures. JN helped develop the intervention, provided the intervention, informed and reviewed the manuscript, and provided guidance and support. HD helped inform the intervention, provided the intervention, and helped with data collection and cleaning. HB helped inform the intervention, provided the intervention, and helped collect measurements related to the intervention. The authors would like to acknowledge the Exercise and Physical Activity Resource Center (EPARC) group at large for contributing to this pilot, as well as the Altman Clinical and Translational Research Institute (ACTRI) Dissemination and Implementation Science Center (DISC) for providing funding to allow this pilot to occur.

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. Burns E, Kakara R, Moreland B. A CDC Compendium of Effective Fall Interventions: What Works for Community-Dwelling Older Adults, 4th edition: Centers for Disease Control and Prevention (US); 2023. URL: https://stacks.cdc.gov/view/cdc/124200 [accessed 2025-02-17]
2. Kakara R, Bergen G, Burns E, Stevens M. Nonfatal and fatal falls among adults aged ≥65 Years - United States, 2020-2021. MMWR Morb Mortal Wkly Rep 2023 Sep 1;72(35):938-943. [doi: 10.15585/mmwr.mm7235a1] [Medline: 37651272]
3. Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. N Engl J Med 1988 Dec 29;319(26):1701-1707. [doi: 10.1056/NEJM198812293192604] [Medline: 3205267]
4. Lundin-Olsson L, Nyberg L, Gustafson Y. Attention, frailty, and falls: the effect of a manual task on basic mobility. J Am Geriatr Soc 1998 Jun;46(6):758-761. [doi: 10.1111/j.1532-5415.1998.tb03813.x] [Medline: 9625194]
5. Baloh RW, Corona S, Jacobson KM, Enrietto JA, Bell T. A prospective study of posturography in normal older people. J Am Geriatr Soc 1998 Apr;46(4):438-443. [doi: 10.1111/j.1532-5415.1998.tb02463.x] [Medline: 9560065]
6. Espinosa de Los Monteros K, Gallo LC, Elder JP, Talavera GA. Individual and area-based indicators of acculturation and the metabolic syndrome among low-income Mexican American women living in a border region. Am J Public Health 2008 Nov;98(11):1979-1986. [doi: 10.2105/AJPH.2008.141903] [Medline: 18799765]
7. Gill TM. Assessment of function and disability in longitudinal studies. J Am Geriatr Soc 2010 Oct;58 Suppl 2(Suppl 2):S308-S312. [doi: 10.1111/j.1532-5415.2010.02914.x] [Medline: 21029059]
8. Burns E, Kakara R. Deaths from falls among persons aged ≥65 Years — United States, 2007–2016. MMWR Morb Mortal Wkly Rep 2018 May 11;67(18):509-514. [doi: 10.15585/mmwr.mm6718a1] [Medline: 29746456]
9. Vaishya R, Vaish A. Falls in older adults are serious. Indian J Orthop 2020 Feb;54(1):69-74. [doi: 10.1007/s43465-019-00037-x] [Medline: 32257019]
10. McAninch J, Greene C, Sorkin JD, Lavoie MC, Smith GS. Higher psychological distress is associated with unintentional injuries in US adults. Inj Prev 2014 Aug;20(4):258-265. [doi: 10.1136/injuryprev-2013-040958] [Medline: 24174466]
11. Judy A, Stevens P, Burns E. A CDC Compendium of Effective Fall Interventions: What Works for Community-Dwelling Older Adults, 3rd edition: Centers for Disease Control and Prevention (US); 2015. URL: https://stacks.cdc.gov/view/cdc/30689 [accessed 2025-02-27]
12. Stevens JA, Phelan EA. Development of STEADI: A fall prevention resource for health care providers. Health Promot Pract 2013 Sep;14(5):706-714. [doi: 10.1177/1524839912463576] [Medline: 23159993]
13. Sarmiento K, Lee R. STEADI: CDC's approach to make older adult fall prevention part of every primary care practice. J Safety Res 2017 Dec;63:105-109. [doi: 10.1016/j.jsr.2017.08.003] [Medline: 29203005]
14. Recommendation: falls prevention in community-dwelling older adults: interventions. United States Preventive Services Taskforce. URL: https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/falls-prevention-community-dwelling-older-adults-interventions [accessed 2025-02-13]
15. US Preventive Services Task Force, Grossman DC, Curry SJ, et al. Interventions to prevent falls in community-dwelling older adults: US Preventive Services Task Force recommendation statement. JAMA 2018 Apr 24;319(16):1696-1704. [doi: 10.1001/jama.2018.3097] [Medline: 29710141]
16. Casey CM, Parker EM, Winkler G, Liu X, Lambert GH, Eckstrom E. Lessons learned from implementing CDC's STEADI falls prevention algorithm in primary care. Gerontologist 2017 Aug;57(4):787-796. [doi: 10.1093/geront/gnw074] [Medline: 27130270]
17. Wenger NS, Solomon DH, Roth CP, et al. The quality of medical care provided to vulnerable community-dwelling older patients. Ann Intern Med 2003 Nov 4;139(9):740-747. [doi: 10.7326/0003-4819-139-9-200311040-00008] [Medline: 14597458]
18. Stevens JA, Ballesteros MF, Mack KA, Rudd RA, DeCaro E, Adler G. Gender differences in seeking care for falls in the aged medicare population. Am J Prev Med 2012 Jul;43(1):59-62. [doi: 10.1016/j.amepre.2012.03.008] [Medline: 22704747]
19. Keadle SK, McKinnon R, Graubard BI, Troiano RP. Prevalence and trends in physical activity among older adults in the United States: A comparison across three national surveys. Prev Med 2016 Aug;89:37-43. [doi: 10.1016/j.ypmed.2016.05.009]
20. Belza B, Shumway-Cook A, Phelan EA, Williams B, Snyder SJ, LoGerfo JP. The effects of a community-based exercise program on function and health in older adults: the EnhanceFitness program. J Appl Gerontol 2006 Aug;25(4):291-306. [doi: 10.1177/0733464806290934]
21. Bethancourt HJ, Rosenberg DE, Beatty T, Arterburn DE. Barriers to and facilitators of physical activity program use among older adults. Clin Med Res 2014 Sep;12(1-2):10-20. [doi: 10.3121/cmr.2013.1171] [Medline: 24415748]
22. Pelletier CA, White N, Duchesne A, Sluggett L. Barriers to physical activity for adults in rural and urban Canada: A cross-sectional comparison. SSM Popul Health 2021 Dec;16:100964. [doi: 10.1016/j.ssmph.2021.100964] [Medline: 34841038]
23. Cosco TD, Fortuna K, Wister A, Riadi I, Wagner K, Sixsmith A. COVID-19, social isolation, and mental health among older adults: a digital catch-22. J Med Internet Res 2021 May 6;23(5):e21864. [doi: 10.2196/21864] [Medline: 33891557]
24. Raney L, Bergman D, Torous J, Hasselberg M. Digitally driven integrated primary care and behavioral health: how technology can expand access to effective treatment. Curr Psychiatry Rep 2017 Sep 30;19(11):86. [doi: 10.1007/s11920-017-0838-y] [Medline: 28965319]

XSL•FO
RenderX

25. Exposing the hidden connectivity crisis for older adults. : AARP; 2022 URL: https://oats.org/wp-content/uploads/2021/01/Aging-Connected-Exposing-the-Hidden-Connectivity-Crisis-for-Older-Adults.pdf [accessed 2025-02-17]
26. Mace RA, Mattos MK, Vranceanu AM. Older adults can use technology: why healthcare professionals must overcome ageism in digital health. Transl Behav Med 2022 Dec 30;12(12):1102-1105. [doi: 10.1093/tbm/ibac070] [Medline: 36073770]
27. Share of those 65 and older who are tech users has grown in the past. Pew Research Center. URL: https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/ [accessed 2025-02-13]
28. Jacobson CL, Foster LC, Arul H, Rees A, Stafford RS. A Digital Health Fall Prevention Program for Older Adults: Feasibility Study. JMIR Form Res 2021 Dec 23;5(12):e30558. [doi: 10.2196/30558] [Medline: 34837492]
29. S Ann Earon PD. Video in Healthcare Communications.: Zoom; 2022 URL: https://explore.zoom.us/docs/doc/Video_Communications_in_Healthcare.pdf
30. Orton E, Lafond N, Skelton DA, et al. Implementation fidelity of the Falls Management Exercise Programme: a mixed methods analysis using a conceptual framework for implementation fidelity. Public Health (Fairfax) 2021 Aug;197:11-18. [doi: 10.1016/j.puhe.2021.05.038] [Medline: 34271270]
31. Bansal S, Katzman WB, Giangregorio LM. Exercise for improving age-related hyperkyphotic posture: A systematic review. Arch Phys Med Rehabil 2014 Jan;95(1):129-140. [doi: 10.1016/j.apmr.2013.06.022] [Medline: 23850611]
32. Sherrington C, Michaleff ZA, Fairhall N, et al. Exercise to prevent falls in older adults: an updated systematic review and meta-analysis. Br J Sports Med 2017 Dec;51(24):1750-1758. [doi: 10.1136/bjsports-2016-096547] [Medline: 27707740]
33. Hopewell S, Adedire O, Copsey BJ, et al. Multifactorial and multiple component interventions for preventing falls in older people living in the community. Cochrane Database Syst Rev 2018 Jul 23;7(7):CD012221. [doi: 10.1002/14651858.CD012221.pub2] [Medline: 30035305]
34. Stay independent brochure. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/steadi/pdf/steadi-brochure-stayindependent-508.pdf [accessed 2025-02-13]
35. Brooke J. SUS—A quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester B, McClelland AL, editors. Usability Evaluation in Industry: Taylor and Francis; 1996.
36. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: A scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
37. Assessment timed up and go (TUG). Centers for Disease Control and Prevention. URL: https://www.cdc.gov/steadi/media/pdfs/steadi-assessment-tug-508.pdf [accessed 2025-02-13]
38. Barry E, Galvin R, Keogh C, Horgan F, Fahey T. Is the Timed Up and Go test a useful predictor of risk of falls in community dwelling older adults: a systematic review and meta-analysis. BMC Geriatr 2014 Feb 1;14(1):14-14. [doi: 10.1186/1471-2318-14-14] [Medline: 24484314]
39. Montero-Odasso M, van der Velde N, Martin FC, et al. World guidelines for falls prevention and management for older adults: a global initiative. Age Ageing 2022 Sep 2;51(9):afac205. [doi: 10.1093/ageing/afac205] [Medline: 36178003]
40. 30-second chair stand. Centers for Disease Control and Prevention. 2017. URL: https://www.cdc.gov/steadi/media/pdfs/STEADI-Assessment-30Sec-508.pdf [accessed 2025-02-13]
41. Lewis JR. The System Usability Scale: past, present, and future. Int J Hum–Comp Interaction 2018 Jul 3;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
42. Smith ML, Towne SD, Herrera-Venson A, et al. Delivery of fall prevention interventions for at-risk older adults in rural areas: findings from a national dissemination. Int J Environ Res Public Health 2018 Dec 10;15(12):2798. [doi: 10.3390/ijerph15122798] [Medline: 30544658]
43. Marzuca-Nassr GN, Alegría-Molina A, SanMartín-Calísto Y, et al. Muscle mass and strength gains following resistance exercise training in older adults 65-75 years and older adults above 85 years. Int J Sport Nutr Exerc Metab 2024 Jan 1;34(1):11-19. [doi: 10.1123/ijsnem.2023-0087] [Medline: 37875254]
44. Lai X, Bo L, Zhu H, et al. Effects of lower limb resistance exercise on muscle strength, physical fitness, and metabolism in pre-frail elderly patients: a randomized controlled trial. BMC Geriatr 2021 Jul 30;21(1):447. [doi: 10.1186/s12877-021-02386-5] [Medline: 34330211]
45. Gusi N, Carmelo Adsuar J, Corzo H, Del Pozo-Cruz B, Olivares PR, Parraca JA. Balance training reduces fear of falling and improves dynamic balance and isometric strength in institutionalised older people: a randomised trial. J Physiother 2012;58(2):97-104. [doi: 10.1016/S1836-9553(12)70089-9] [Medline: 22613239]
46. Wright AA, Cook CE, Baxter GD, Dockerty JD, Abbott JH. A comparison of 3 methodological approaches to defining major clinically important improvement of 4 performance measures in patients with hip osteoarthritis. J Orthop Sports Phys Ther 2011 May;41(5):319-327. [doi: 10.2519/jospt.2011.3515] [Medline: 21335930]
47. Gautschi OP, Stienen MN, Corniola MV, et al. Assessment of the minimum clinically important difference in the timed up and go test after surgery for lumbar degenerative disc disease. Neurosurgery 2017 Mar 1;80(3):380-385. [doi: 10.1227/NEU.0000000000001320] [Medline: 27352275]
48. Hyzy M, Bond R, Mulvenna M, et al. System Usability Scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290. [doi: 10.2196/37290] [Medline: 35980732]

XSL•FO
RenderX

49. Holden RJ. A Simplified System Usability Scale (SUS) for cognitively impaired and older adults. Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care 2020 Sep;9(1):180-182. [doi: 10.1177/2327857920091021]

Abbreviations

30 CS: 30-second chair stand
CDC: Centers for Disease Control and Prevention
EPARC: Exercise and Physical Activity Resource Center
STEADI: Stopping Elderly Accidents and Deaths Initiative
SUS: System Usability Scale
TUG: timed up and go
USPSTF: United States Preventive Services Task Force

Edited by A Mavragani; submitted 10.10.24; peer-reviewed by H Hawley-Hague, M Fraser; revised version received 28.01.25; accepted 29.01.25; published 28.02.25.

Please cite as:
Moran R, Wing D, Davey H, Barkai H, Nichols J
Development and Implementation of Strong Foundations, a Digitally Delivered Fall Prevention Program: Usability and Feasibility Pilot Exercise Cohort Study
JMIR Form Res 2025;9:e67406
URL: https://formative.jmir.org/2025/1/e67406
doi:10.2196/67406

© Ryan Moran, David Wing, Hope Davey, Hava Barkai, Jeanne Nichols. Originally published in JMIR Formative Research (https://formative.jmir.org), 28.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluating the Knowledge Level, Practice, and Behavioral Change Potential of Care Managers in Pressure Injury Prevention Using a Mobile App Prototyping Model in the Home-Care Setting: Single-Arm, Pre-Post Pilot Study

Masushi Kohta[1*], PhD; Mayumi Takahashi[2*], CWOCN, RN; Hiroe Koyanagi[1*], CWOCN, RN, PhD; Junko Sugama[1*], RN, PhD

[1]Research Center for Implementation Nursing Science Initiative, Fujita Health University, Toyoake, Japan

[2]Department of Nursing, Akabane Central General Hospital, Kita-ku, Tokyo, Japan

[*]all authors contributed equally

Corresponding Author:
Masushi Kohta, PhD
Research Center for Implementation Nursing Science Initiative, Fujita Health University, Toyoake, Japan

Abstract

Background: The use of mobile apps to promote knowledge level, practice, and behavioral change potential has become increasingly common. However, studies on apps targeting social welfare employees working in the home-care setting to prevent pressure injury (PI) are lacking. The care manager (CM) plays a key role in connecting the demand and supply of home-care services. PI is more prevalent in the home-care setting, where resources are limited, than in acute settings.

Objective: The research hypothesis was that CMs who use a mobile app will have improved general knowledge and heightened practice for PI prevention, compared to that before using the app. This study aimed to assess the effectiveness of a PI prevention support mobile app prototyping model (Pips-Map) in improving the knowledge level, practice, and behavioral change potential of CMs in PI prevention in the home-care setting.

Methods: This was conducted between December 2021 and December 2023 as a single-arm, pre-post pilot study including 27 CMs who worked in a Japanese city. Pips-Map was used for 6 months in daily practice, and a self-administered test questionnaire was used to assess participants' knowledge and practice in PI prevention before or after using Pips-Map. At the end of the posttest, a validated App Behavior Change Scale was used to analyze behavioral change potential. This study followed the Consolidated Standards of Reporting Trials (CONSORT) extension to pilot and feasibility trials.

Results: In total, 19 participants were analyzed. Out of 55 points, the total mean knowledge score significantly increased from 30.9 (SD 5.9) in the pretest group to 36.1 (SD 5.9) in the posttest group (P=.0003). The number of participants with a total score of >70% (adequate knowledge level) increased from 2 (11%) to 7 (36.8%), but the difference was not statistically significant (P=.07). For the level of practice, out of 21 points, the total score increased from 15.2 (SD 3.1) in the pretest group to 16.2 (SD 3.0) in the posttest group, but no statistically significant differences were observed (P=.16). The behavior change scale revealed that participants positively evaluated the Pips-Map to provide information on PI prevention guidelines but had concerns regarding inadequate usability and financial incentives of Pips-Map.

Conclusions: The use of Pips-Map for 6 months in actual practice increased the knowledge level of Japanese CMs in PI prevention, but it did not change the level of practice. Considering the need for updating apps that aim to promote behavioral change, this study identified some limitations of Pips-Map. Thus, revisions must be made to adapt Pips-Map to home-based care needs.

Trial Registration: UMIN Clinical Trial Registry UMIN000048904; https://tinyurl.com/5bt6zv7t

(JMIR Form Res 2025;9:e57768) doi:10.2196/57768

KEYWORDS

behavioral change; home care; knowledge; mobile application; pressure injury; mHealth; mobile health; apps; practice; injury; prevention; prototype; effectiveness; care manager; Japan; Pips-Map; questionnaire; wound care; pilot study; women

XSL•FO
RenderX

Introduction

Backgrounds

Many developed countries are facing the challenge of a declining birthrate and aging population. Considering that Japan has the world's most advanced aging population, the Japanese government has shifted the focus of health care from hospital- to home- or community-based care provided in the home-care setting [1]. To provide comprehensive community-based care, the Japan-specific geriatric care system (Community-based Integrated Care System, a medical- and social-welfare networking model) works with the Japan-specific long-term care insurance system. The care manager (CM), who possesses specialized knowledge necessary to help individuals in need of assistance or care to achieve more independent lives, plays a key role in connecting the demand and supply of long-term care services [2]. A workflow of CMs in home-based care services was summarized in a recent study [3]. The CM is responsible for providing care management services, including checking the physical condition and living circumstances of individuals, establishing and adjusting care plans according to individual needs, and monitoring individual care. To obtain a CM license, an individual must have ≥5 years of geriatric care experience as a medical professional (physician or nurse) and social welfare professional (care worker or social worker) and should have passed the examination. Compared to CMs with medical backgrounds, the number of CMs with social welfare backgrounds is higher, because obtaining a CM license is a career advancement step. Compared with CMs who studied medicine, those from the social welfare background may not have sufficient knowledge or practice regarding the prevention and treatment of chronic diseases, such as pressure injury (PI) [2].

Although PIs are pervasive in all age groups, most of them occur in older individuals with immobility and chronic diseases, such as heart disease, respiratory illness, and dementia [4]. PI can result in poor quality of life for patients and socioeconomic burden [5,6]. PI is more prevalent in the home-care setting, where resources are limited, than in acute settings, with prevalence rates of 1.76% - 26% [7-9]. Although there is broad consensus that some PIs are unavoidable, most are considered preventable [10]. A paper-based risk assessment tool for preventing PIs that can be used by Japanese CMs has recently been developed and implemented [11-13]; however, its application to digital technology is still under consideration.

Prior Work

Digital technologies have become an emerging branch of medicine, with technology-based applications being used to improve disease knowledge and user activations being modified by relying on health-related information [14-16]. Mobile apps for PIs can be grouped into apps for PI treatment and apps for PI prevention. European wound management experts developed a clinical support mobile app based on a global wound care framework to implement care plans and promote wound healing [17]. Do Khac et al [18] evaluated a free-access mobile health app compared with standard technique to assess its reliability and validity of measurement of surface area of PIs in patients with spinal cord injury. In Japan, a PI treatment support mobile app was developed for sharing PI information with a multidisciplinary care team or for conducting remote consultation with wound care specialists [19,20]. A smartphone app focusing on PI prevention education was developed to evaluate the PI knowledge of informal carers and app acceptability and usability [21]. A guideline app was designed to present evidence-based guideline recommendations with easy access [22]. Amann et al [23] qualitatively investigated barriers and facilitators to the adoption of a self-management app for PI prevention in individuals with spinal cord injury. However, studies assessing a mobile app in improving knowledge and behavioral change within social welfare workers, such as CMs, who work at home-based care facilities, are lacking.

Aim of This Study

The research hypothesis is that CMs who experience a mobile app will improve general knowledge and heightened practice about PI prevention compared to before the experience. The primary objective was to investigate the effects of CMs on improving knowledge level when using a PI prevention support mobile app prototyping model (Pips-Map), which has been developed by the industry. In addition to improving the knowledge level, we surveyed whether the practice level of CMs for PI prevention was heightened when using Pips-Map as a secondary objective. We further investigated to determine whether the use of the app promoted behavior change in PI prevention among CMs and which specific features of the app positively or negatively influenced the behavior change.

Methods

Prototyping Model

Pips-Map was designed and developed by ALCARE Co. Ltd. (Tokyo, Japan) to support research and practice in wound care, home-based care, etc. According to the developer, the key desktop design concept of Pips-Map was to facilitate information sharing among home health care providers with different specialties and change their behavior for preventing PI development in patients who received assistance or nursing care at home. Pips-Map is unique because it facilitates information sharing among home health care providers, including CMs, home-care nurses, and primary physicians for each home-based patient. Displaying products formulary for preventing PI in Pips-Map helps in selecting products for each patient in a timely manner. Pips-Map has 5 functions: (a) photographing skin using a standard smartphone or tablet camera; (b) assessing PI risk according to the Japanese Ohura-Hotta scale; (c) entering data on prophylactic planning or comments based on PI risk assessment; (d) displaying products (eg, pressure redistribution mattress, skin care products, prophylactic dressings) formulary; and (e) data sharing between medical and social welfare professionals for each patient. Pips-Map includes a software system named Medical Care Station, which was developed by Embrace Co., Ltd. (Tokyo, Japan), to provide patient information. The system is a security-conscious, completely private social networking service that shares information in a timeline format. Patient data are stored on the system but not linked to the electronic medical records of each site. Selected

XSL•FO
RenderX

screenshots from Pips-Map are presented in Multimedia Appendix 1.

Study Design

This prospective, single-arm, interventional pilot study was conducted at a university from December 2021 to December 2023. There were no major changes in study methods after the trial commencement. This study followed the Consolidated Standards of Reporting Trials (CONSORT) extension to pilot and feasibility trials [24] from the Equator Network, which assess the feasibility of conducting a future definitive randomized controlled trial rather than the effectiveness or efficacy of an intervention. As an editorial, Lancaster et al have provided some general guidance on how to report non-randomized pilot and feasibility studies, including a pre-post test, using the CONSORT extension to pilot and feasibility trials [25]. The checklists for the reporting guideline in this study were provided in Multimedia Appendix 2.

Setting and Sample

The accessible population included individuals with a CM license and those who participated in the "Kita-ku Care Managers' Association" in Kita-ku, Tokyo (n=185). According to statistics from the Japanese Ministry of Health, Labor and Welfare in the year 2022, the number of CMs working full-time at home-care support facilities is 99,220. Of these, 8,943 live in Tokyo [26]. Kita-ku lies in the northeastern region of Tokyo and is one of the 23 districts. The Kita-ku Care Managers' Association works in Kita-ku, and the number of CMs accounts for approximately 2.1% (185/8943) of them in Tokyo. Japanese CMs could use an Android- and iOS-based mobile application. As power analysis could not be performed, this study was considered a pilot study. Pilot studies are needed to obtain preliminary data for calculating sample size for the primary outcome [27].

Study Procedure

Kita-ku Care Managers' Association holds regular meetings for information exchange and acquisition of knowledge about home-care management. In this meeting, we orally explained the details of this study after obtaining agreement for cooperation with this study from the member of the executive committee of the Kita-ku Care Managers' Association. A total of 27 individuals agreed to participate, and the pretest was completed. Pips-Map was downloaded and operated for 6 months in daily practice. The participants were instructed on how to use Pips-Map and record the details of PI prevention in Pips-Map through web-based guidance. Technical support on how to use Pips-Map was provided by the researcher through telephone and e-mail.

A pretest or posttest was performed to measure participants' knowledge level and practice level in PI prevention. The pretest was conducted on the same day when informed consent was obtained, and the posttest was conducted within a month after using Pips-Map. At the end of the posttest, an additional questionnaire survey was conducted to assess Pips-Map to promote behavioral change potential. All tests and questionnaires were made available via web-based Google Forms. The completion of pretest and posttest and additional questionnaire required approximately 20 and 10 minutes, respectively.

Measurement

General Characteristics of Participants

The language used for the questionnaire was Japanese. The questionnaire was divided into 4 parts. Part 1 contained general characteristics of participants, including age range, sex, years of experience as a CM, professional background, and education level.

Primary Outcome

In part 2, participant knowledge on PI prevention was evaluated. Although 2 validated knowledge tests for PI prevention have been reported [28,29], most respondents in these studies were medical professionals, such as nurses who worked at hospitals. The test questions were difficult to answer for social welfare professionals such as Japanese CMs. Therefore, the current self-administered test questionnaire was developed via collaboration between a nurse with certification in wound, ostomy, and continence working at the home-care division of a hospital and a senior assistant professor certified in wound, ostomy, and continence nursing. For questionnaire development, they referred to a previous report [30] and the Japanese PI guideline created by the Japanese Society of Pressure Ulcers [31]. This part consisted of 55 test questions reflecting 6 themes: (1) anatomy, (2) risk assessment, (3) pressure, (4) friction and shear, (5) moisture, and (6) nutritional support. Answering options and scoring rules varied depending on items. For items 1 - 4, 7, and 11 - 13, the participants were instructed to select "true," "false," or "I do not know." Each correct answer was assigned one point, and no points were assigned to wrong answers and "I do not know" answers. For items 5, 9, and 14, the participants selected "know" or "do not know." A score of 1 was assigned for "know" and 0 for "do not know." For items 6, 8, and 10, the multiple-choice questions had several options along with the "I do not know" option. A correct answer was assigned a score of 1, and an incorrect answer was assigned 0. The maximum score was 55 (Multimedia Appendix 3).

Secondary Outcomes

In part 3, the participants were asked about the practice of preventive strategies for PIs. This part consisted of 21 test questions reflecting 6 sections: (1) risk assessment, (2) pressure, (3) friction and shear, (4) moisture, (5) nutritional support, and (6) communication with other professionals. The test questionnaire in this section asked whether the participants implemented those actions or not. The answer choice for each question was "practice" or "do not practice." One or zero point was given when the participant answered "practice" or "do not practice," respectively. The maximum practice score was 21 in this section (Multimedia Appendix 4).

In part 4, behavioral change potential was measured using a validated App Behavior Change Scale (ABACUS) [32] with a minor modification. For the modification, the item "does the app provide a material or social reward or incentive?" in the original paper was divided into the following 2 items: "does the app provide a financial incentive?" and "does the app provide

XSL•FO
RenderX

a material or social reward?" The scale comprised 22 questions regarding a determinant of behavioral change potential of mobile apps, each on a 5-point Likert scale (1="strongly disagree" to 5="strong agree") with total score 22 - 110. A higher score indicated higher levels of promoting behavioral change by app usage. Previous literature showed that potential behavioral change of users of apps was evaluated by ABACUS in a wide range of research fields, including depression, diabetes, osteoarthritis, and physiotherapy care [33,34].

Data Analysis

For general characteristics, all data were categorical variables and expressed as frequencies and percentages. For level of knowledge, as a primary outcome, a dependent *t*-test (for continuous variables), the Wilcoxon signed-rank test (for ordinal variables), and McNemar's test (for categorical variables) were conducted to compare differences between each item and total score between the pretest and posttest groups. Ordinal variables with ≥5 categories were treated as continuous variables. A study on the learning outcome criteria for PI prevention considered an average total score of >70% satisfactory [35]. As a secondary outcome, similar analysis was conducted for the level of practice. For behavioral change potential, mean and SD were reported in each item and total score of ABACUS. All statistical analyses were conducted using the Statistical Package for the Social Sciences version 28.0 (IBM Corporation, Armonk, NY, USA). *P* values of <.05 were considered to indicate statistical significance.

Ethical Considerations

This study was approved by the university's ethics review board (approval number HM21-335). Electronic informed consent was obtained from all participants before participation. All experiments in this research were performed according to the principles of the Declaration of Helsinki. Participant data were kept confidential. This study has been registered in the Clinical Trials Registry (UMIN Clinical Trial Registry number: UMIN000048904) and is accessible online. Each participant who completed the questionnaire was given a gift certificate (5000 JPN yen) as a financial reward.

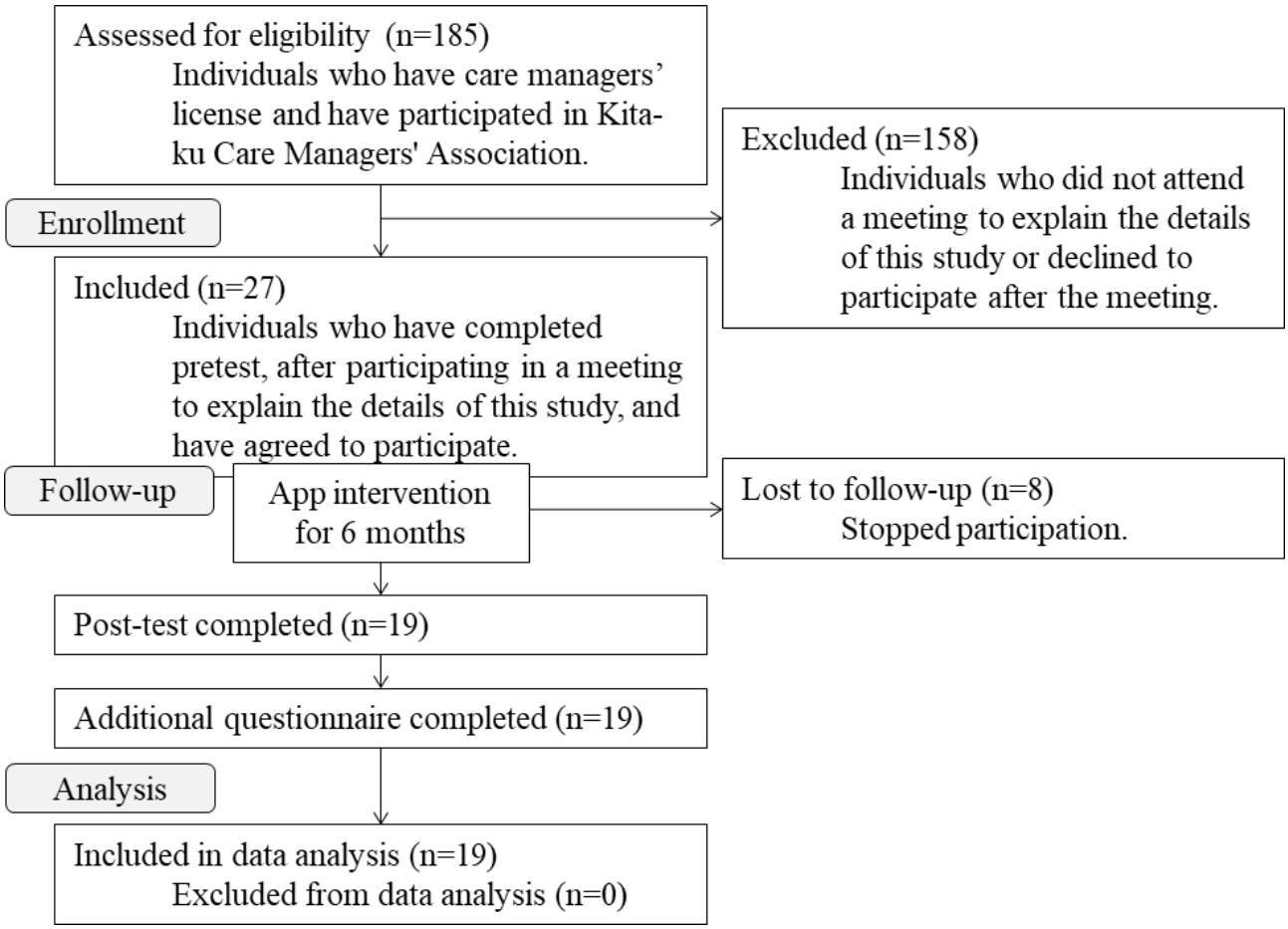
Results

General Characteristics of Participants

In total, 19 participants were analyzed (Figure 1). The recruitment of participants for this study trial started on September 11, 2022, and the follow-up was completed on August 31, 2023. Most participants were in their 40 and 50 seconds, and 12 out of 19 participants (63%) were female. All of the participants had a background in social welfare; that is, participants who studied medicine or nursing were not included in this study (Table 1).

XSL•FO
RenderX

Figure 1. Flowchart of participant progression in pretest and posttest study. A total of 27 care managers who participated in the "Kita-ku Care Managers' Association" in a district in Tokyo agreed to participate. As an intervention, a pressure injury prevention support mobile app prototyping model was operated for 6 months in daily practice. A pretest or posttest was performed to measure participants' knowledge level and practice level in pressure injury prevention. After the posttest was completed, an additional questionnaire survey was conducted to assess the mobile app to promote behavioral change potential. In total, 19 participants were included in data analysis. This study was conducted from December 2021 to December 2023.



XSL•FO
RenderX

Table . General characteristics of participants. Under the Japan-specific long-term care insurance system, to obtain a care manager license, an individual must have ≥5 years of geriatric care experience as a medical professional (physician or nurse) and social welfare professional (care worker or social worker). Compared with care managers who studied medicine, those from the social welfare background may not have sufficient knowledge or practice regarding the prevention of pressure injury. All of the care managers who participated in this study had a professional background in social welfare, not medicine. As the primary outcome included knowledge levels of pressure injury prevention, the educational level for the study participants was an important characteristic to be confirmed.

Characteristics	Overall (N=19)
Age, n (%)	
30s	3 (16)
40s	7 (37)
50s	6 (32)
60s	3 (16)
Sex, n (%)	
Male	7 (37)
Female	12 (63)
Years of experience as a care manager, n (%)	
<5	3 (16)
≥5 and <10	4 (21)
≥10	12 (63)
Professional background, n (%)	
Social welfare	19 (100)
Medical	0 (0)
Education level, n (%)	
Junior high school	1 (6)
High school	8 (44)
Vocational school	5 (28)
Junior college	1 (6)
College	3 (17)

Primary Outcome Evaluation

As shown in Table 2, the total score of knowledge significantly increased in the posttest group compared with that in the pretest group (P=.0003). The number of participants with a total score >70% increased from 2 (11%) in the pretest group to 7 (37%) in the posttest group, but the differences were not statistically significant (P=.07). In each test questionnaire, “anatomy,” “risk assessment,” and “pressure” increased significantly, but the remaining items were similar between the 2 groups.

XSL•FO
RenderX

Table . Comparison of knowledge level between the pretest and posttest groups. for the questionnaire development to evaluate the participants' knowledge level in pressure injury prevention, the authors referred to a previous report [28] and the Japanese pressure injury guideline [29]. It consisted of 55 test questions reflecting 6 themes: (1) anatomy, (2) risk assessment, (3) pressure, (4) friction and shear, (5) moisture, and (6) nutritional support. The maximum score was 55. A pretest or posttest was performed to measure participants' knowledge level in pressure injury prevention. The pretest was conducted on the same day when informed consent was obtained, and the posttest was conducted within a month after intervention. The total score of knowledge significantly increased in the posttest group compared with that in the pretest group. Statistical analysis was conducted via those listed in footnotes b, d, and e.

Questionnaire	Pretest	Posttest	*P*-value
Anatomy, mean (SD[a])	16.2 (2.6)	19.2 (3.6)	.001[b]
Risk assessment, mean (SD)	0.3 (0.5)	1.2 (1.4)	.009[b]
Pressure, mean (SD)	6.3 (2.0)	7.4 (1.5)	.013[b]
Friction and shear, mean (SD)	4.7 (1.6)	4.8 (1.3)	.76[b]
Moisture, median (IQR[c])	2 (2–3)	3 (2–3)	.25[d]
Nutritional support, mean (SD)	1.2 (0.9)	1.1 (1.3)	.56[b]
Total score, mean (SD)	30.9 (5.9)	36.1 (5.9)	.0003[b]
Total score>70%, n (%)	2 (11)	7 (37)	.070[e]

[a]SD, standard deviation.

[b]dependent *t*-tests for continuous variables.

[c]IQR, interquartile range.

[d]Wilcoxon signed-rank test for ordinal variables, and.

[e]McNemar's test for categorical variables.

Secondary Outcome Evaluation

As shown in Table 3, the total score of practice increased from 15.2 (SD 3.1) in the pretest group to 16.2 (SD 3.0) in the posttest group, but no statistically significant differences were observed (*P*=.16). The score of "risk assessment" increased significantly (*P*=.030). The average total score for behavioral change potential using ABACUS was 71.04 (SD 11.55) (Multimedia Appendix 5). The highest score was achieved for the question "was the app created with expertise and does the app provide information consistent with national guidelines?" Lower scores (mean value of <3.0) were obtained for the questions "does the app provide financial incentives for the users and their organization?," "does the app have the ability to export data from app?," and "does the app allow or encourage practice or rehearsal, in addition to daily activities?"

XSL•FO
RenderX

Table . Comparison of level of practice between the pretest and posttest groups. The participants were asked about the practice of preventive strategies for pressure injuries. It consisted of 21 test questions reflecting 6 sections: (1) risk assessment, (2) pressure, (3) friction and shear, (4) moisture, (5) nutritional support, and (6) communication with other professionals. The maximum practice score was 21. The pretest was conducted on the same day when informed consent was obtained, and the posttest was conducted within a month after the intervention. No statistically significant differences were shown in the total score of practice between the posttest and posttest groups. Statistical analysis was conducted via a, c, and e.

Questionnaire	Pretest	Posttest	*P* value
Risk assessment, n (%)	4 (21)	6 (32)	.48[a]
Pressure, mean (SD[b])	6.8 (2.0)	7.9 (1.7)	.030[c]
Friction and shear, median (IQR[d])	2 (1–2)	2 (1–2)	.74[e]
Moisture, median (IQR)	2 (2–2)	2 (1–2)	.60[e]
Nutritional support, mean (SD)	2.7 (1.0)	2.6 (1.0)	.84[c]
Communication with other professionals, median (IQR)	2 (2–3)	2 (2–2)	.60[e]
Total score, mean (SD)	15.2 (3.1)	16.2 (3.0)	.16[c]
Total score>70%, n (%)	10 (53)	11 (58)	.68[a]

[a]McNemar's test for categorical variables.

[b]SD, standard deviation.

[c]dependent *t*-tests for continuous variables, and.

[d]IQR, interquartile range.

[e]Wilcoxon signed-rank test for ordinal variables.

Discussion

Principal Findings

The use of a mobile app provides a more engaging and accessible approach to clinical practice at home care, fitting well with the evolving demands of the digital ages. The development of a PI prevention support mobile app and its evaluation as an educational tool has already been reported; however, studies approaching a mobile app in improving knowledge, practice, and behavioral change within social welfare workers who work at home-care facilities are lacking. In this study, we used Pips-Map in an actual clinical setting, and the effect of improving general knowledge, heightened practice, and positively changing behavior among CMs in the prevention of PI development was investigated in a pretest-posttest study design. As the incidence of PI in the home-care setting impacts the quality of care provided, the adoption of good practice with adequate knowledge through improvement in PI prevention is necessary. Adequate knowledge is essential for assessing behavioral change. The key finding of this study is that the use of a mobile app in daily practice can increase knowledge about PI prevention among Japanese CMs. However, our results could not demonstrate that increased knowledge indicated improved practice level.

The most significant finding of this study is that Pips-Map is effective in improving the total score of knowledge level (Table 2). Consistently, a 6-week interventional study based on a mobile app focused on PI prevention education for informal carers of individuals at risk of PI [21]. Using a self-administered PI knowledge assessment questionnaire consisting of 4 categories (support surfaces, nutrition and hydration, keep moving, and skin care and inspection), they reported that the use of the app effectively increased knowledge level regarding PI prevention. CMs need to manage their care work with other responsibilities, and their busy schedule makes learning about PI prevention challenging. Active methodology with digital technologies enables flexibility for conducting these activities once the barrier of time and space has been overcome. Compared with conventional on-site education, such as through lectures or handouts, Pips-Map is more advantageous as a training tool in terms of accessibility, leading to the prevention of common PIs in the home-care setting.

However, changes in the participants' PI prevention practices were not observed (Table 3). This might be explained by our results to understand the behavioral change potential using ABACUS (Multimedia Appendix 5). A high score was obtained for item 2, "providing information consistent with national guidelines," which directly affects the increase in the knowledge of PI prevention. By contrast, lower scores were observed for item 13, "ability to export data from app," item 14, "providing a financial incentive," and item 19, "allowing for practice or rehearsal," which are significantly important items for CMs. In this current form, we believe that Pips-Map is didactic and inadequate in terms of customized content and usability.

As shown in the lowest mean score in ABACUS, the participants did not perceive receiving social or financial incentive by using the app. Although financial rewards are a powerful extrinsic motivation for app use, studies suggest that the most powerful plans for behavioral change include both extrinsic and intrinsic factors. For example, rankings, social recognition, and gaming are reported to be strong intrinsic factors relating to long-term app use [36]. In addition, the opinions raised from the behavioral change questionnaire regarding customized content and usability could not allow for adaptation of the app in the real-world

XSL•FO
RenderX

setting. Hence, we will provide feedback to the industry on the need to create a modified version of the app to address these issues for the development phase. In addition, we will investigate the impact of the modified version on behavioral change of the user and health care outcomes of individuals who need assistance or nursing care at home. Evidence on knowledge and practice of wound care is updated daily, and all the researchers have to keep up with the fast-moving pace of app technology.

Limitations

Our study has several limitations. First, our tests for knowledge and practice of PI prevention have not yet been validated. It would be better to use validated tests simultaneously; however, tests targeting Japanese CMs are lacking. The face validity, construct validity, and reliability of the knowledge and practice test questions used in this study were not considered, because we prioritized evaluating the effect of a mobile app prototype on knowledge, practice, and behavior change. As a future work, determining the reliability and validity of the newly developed test questions is a key imperative, which ultimately could benefit individuals who receive preventive care as well as social welfare professionals such as CMs who provide it.

Next, this single-arm, pre-post study aimed to preliminarily assess the effectiveness and feasibility of Pips-Map. Determining whether the value of the outcome observed upon intervention is better than the value without intervention is challenging. Finally, we did not perform power analysis to estimate the ideal sample size because an accurately estimated population could not be set before the start of this study. The sample size is relatively small, but it should be acceptable considering the design of a pilot study. The authors were unable to determine the cause for the dropout of study participants because it was difficult to chase them. For the effectiveness and usability of modified Pips-Map, comparative studies between intervention and nonintervention groups with appropriate sample sizes are required in the future.

Conclusions

Our results revealed that the use of a mobile app prototyping model for 6 months in actual practice increased the knowledge level of Japanese CMs in preventing PI development, but it did not affect the level of practice between the pretest and posttest groups. Considering the need for updating apps that aim to promote behavioral change, this study identified limitations associated with the app. A higher quality app must be developed that is suitable for the home-care setting.

Acknowledgments

We thank Eisaku Oba for his insightful advice on study procedure. We wish to thank ALCARE for their generous financial assistance. No generative AI and AI-assisted technologies were used in any portion of the manuscript writing process. We also thank Enago (http://www.enago.jp) for its English language editing services.

Data Availability

The datasets generated during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Concept of the pressure injury prevention support mobile application prototyping model (Pips-Map).
[PNG File, 182 KB - formative_v9i1e57768_app1.png]

Multimedia Appendix 2
CONSORT 2010 checklist of information to include when reporting a pilot or feasibility trial.
[DOC File, 233 KB - formative_v9i1e57768_app2.doc]

Multimedia Appendix 3
Pressure injury prevention knowledge test questionnaire.
[DOC File, 88 KB - formative_v9i1e57768_app3.doc]

Multimedia Appendix 4
Pressure injury prevention practice test questionnaire.
[DOC File, 80 KB - formative_v9i1e57768_app4.doc]

Multimedia Appendix 5
Results of app behavioral change scales.
[DOC File, 84 KB - formative_v9i1e57768_app5.doc]

XSL•FO
RenderX

References

1. Iwagami M, Tamiya N. The long-term care insurance system in Japan: past, present, and future. JMA J 2019 Mar 4;2(1):67-69. [doi: 10.31662/jmaj.2018-0015] [Medline: 33681515]
2. Goto Y, Miura H, Ito N. Comparison between the chief care manager and the normal care manager on hospitalization and discharge coordination activities in Japan: an online cross-sectional study of care managers in Aichi prefecture. Int J Environ Res Public Health 2022 Sep 25;19(19):12122. [doi: 10.3390/ijerph191912122] [Medline: 36231421]
3. Wang Z, Fukayama K, Niu B. Does long-term care literacy matter in evaluating older care recipients' satisfaction with care managers? Empirical evidence from Japanese survey data. Int J Environ Res Public Health 2023 Jan 30;20(3):2456. [doi: 10.3390/ijerph20032456] [Medline: 36767824]
4. Elli C, Novella A, Nobili A, Ianes A, Pasina L. Factors associated with a high-risk profile for developing pressure injuries in long-term residents of nursing homes. Med Princ Pract 2022;31(5):433-438. [doi: 10.1159/000527063] [Medline: 36122563]
5. Sebba Tosta de Souza DM, Veiga DF, Santos IDDAO, Abla LEF, Juliano Y, Ferreira LM. Health-related quality of life in elderly patients with pressure ulcers in different care settings. J Wound Ostomy Continence Nurs 2015;42(4):352-359. [doi: 10.1097/WON.0000000000000142]
6. Guest JF, Fuller GW, Vowden P, Vowden KR. Cohort study evaluating pressure ulcer management in clinical practice in the UK following initial presentation in the community: costs and outcomes. BMJ Open 2018 Jul 25;8(7):e021769. [doi: 10.1136/bmjopen-2018-021769] [Medline: 30049697]
7. Sari SP, Everink IH, Sari EA, et al. The prevalence of pressure ulcers in community-dwelling older adults: a study in an Indonesian city. Int Wound J 2019 Apr;16(2):534-541. [doi: 10.1111/iwj.13081] [Medline: 30768769]
8. Konya C, Takeuchi Y, Nakagami G, et al. Nationwide time-series surveys of pressure ulcer prevalence in Japan. In: Takeuchi Y, editor. J Wound Care 2022 Dec 1;31(Sup12):S40-S47. [doi: 10.12968/jowc.2022.31.Sup12.S40]
9. Bottega M, Tempesta M, Piovesan C, et al. The risk and prevalence of pressure injuries in older people in the home care service: a cross-sectional study. J Wound Care 2023 Oct 1;32(Sup10):ccxi-ccxviii. [doi: 10.12968/jowc.2023.32.Sup10.ccxi] [Medline: 37830802]
10. Schmitt S, Andries MK, Ashmore PM, Brunette G, Judge K, Bonham PA. WOCN Society position paper: avoidable versus unavoidable pressure ulcers/injuries. J Wound Ostomy Continence Nurs 2017;44(5):458-468. [doi: 10.1097/WON.0000000000000361] [Medline: 28877112]
11. Kohta M, Ohura T, Tsukada K, et al. Inter-rater reliability of a pressure injury risk assessment scale for home care: a multicenter cross-sectional study. J Multidiscip Healthc 2020;13:2031-2041. [doi: 10.2147/JMDH.S291162] [Medline: 33376343]
12. Kohta M, Ohura T, Okada K, et al. Convergent validity of three pressure injury risk assessment scales: comparing the PPRA-Home (Pressure Injury Primary Risk Assessment Scale for Home Care) to two traditional scales. J Multidiscip Healthc 2021;14:207-217. [doi: 10.2147/JMDH.S294734] [Medline: 33564237]
13. Nakamura Y, Kohta M. Predictive validity of a risk assessment scale for preventing pressure injuries in home care. Tenri Med Bull 2024;27(1):11-18. [doi: 10.12936/tenrikiyo.27-001]
14. Mosa ASM, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12:67. [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
15. O'Leary KJ, Lohman ME, Culver E, Killarney A, Randy Smith G Jr, Liebovitz DM. The effect of tablet computers with a mobile patient portal application on hospitalized patients' knowledge and activation. J Am Med Inform Assoc 2016 Jan;23(1):159-165. [doi: 10.1093/jamia/ocv058] [Medline: 26078412]
16. Iribarren SJ, Akande TO, Kamp KJ, Barry D, Kader YG, Suelzer E. Effectiveness of mobile apps to promote health and manage disease: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2021 Jan 11;9(1):e21563. [doi: 10.2196/21563] [Medline: 33427672]
17. Moore ZEH, Aynge GE, Carr CG, et al. A clinical support app for routine wound management: reducing practice variation, improving clinician confidence and increasing formulary compliance. Int Wound J 2022 Aug;19(5):1263-1275. [doi: 10.1111/iwj.13868] [Medline: 35793908]
18. Do Khac A, Jourdan C, Fazilleau S, et al. mHealth app for pressure ulcer wound assessment in patients with spinal cord injury: clinical validation study. JMIR Mhealth Uhealth 2021 Feb 23;9(2):e26443. [doi: 10.2196/26443] [Medline: 33620327]
19. Kitamura A, Nakagami G, Watanabe C, et al. Efficacy of teleconsultations with wound, ostomy, and continence nurses for patients with pressure ulcers in the community setting. J Jpn WOCM 2021;25(3):654-660. [doi: 10.32201/jpnwocm.25.3_654]
20. Kohta M, Sugiyama A, Hachimori A, Sugama J, Fukuro S. User acceptance of a mobile application prototyping model for supporting pressure injury treatment at home care. J Nurs Sci Eng 2023;11:28-36. [doi: 10.24462/jnse.11.0_28]
21. McKeown E, McGraw C, Holder P, Shand J, Hirani SP. Acceptability and impact of an educational app (iCare) for informal carers looking after people at risk of pressure ulceration: mixed methods pilot study. JMIR Form Res 2022 Sep 16;6(9):e36517. [doi: 10.2196/36517] [Medline: 36112413]
22. Cuddigan J, Haesler E, Moore Z, Carville K, Kottner J. Development, dissemination and evaluation of a smartphone-based app for pressure ulcer/injury prevention and treatment for use at the bedside. J Wound Care 2022 Dec 1;31(Sup12):S29-S39. [doi: 10.12968/jowc.2022.31.Sup12.S29] [Medline: 36475841]

XSL•FO
RenderX

23. Amann J, Fiordelli M, Scheel-Sailer A, Brach M, Rubinelli S. Opportunities and challenges of a self-management app to support people with spinal cord injury in the prevention of pressure injuries: qualitative study. JMIR Mhealth Uhealth 2020 Dec 9;8(12):e22452. [doi: 10.2196/22452] [Medline: 33295876]
24. Eldridge SM, Chan CL, Campbell MJ, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot Feasibility Stud 2016;2:64. [doi: 10.1186/s40814-016-0105-8] [Medline: 27965879]
25. Lancaster GA, Thabane L. Guidelines for reporting non-randomised pilot and feasibility studies. Pilot Feasibility Stud 2019;5:114. [doi: 10.1186/s40814-019-0499-1] [Medline: 31608150]
26. Japanese Ministry of Health, Labor and Welfare. URL: www.mhlw.go.jp/toukei/list/24-22-2.html [accessed 2024-08-19]
27. In J. Introduction of a pilot study. Korean J Anesthesiol 2017;70(6):601. [doi: 10.4097/kjae.2017.70.6.601]
28. Beeckman D, Vanderwee K, Demarré L, Paquay L, Van Hecke A, Defloor T. Pressure ulcer prevention: development and psychometric validation of a knowledge assessment instrument. Int J Nurs Stud 2010 Apr;47(4):399-410. [doi: 10.1016/j.ijnurstu.2009.08.010] [Medline: 19781701]
29. Delmore B, Ayello EA, Smart H, Sibbald RG. Assessing pressure injury knowledge using the Pieper-Zulkowski pressure ulcer knowledge test. Adv Skin Wound Care 2018 Sep;31(9):406-412. [doi: 10.1097/01.ASW.0000540071.45158.29] [Medline: 30134276]
30. Kohta M, Kameda Y, Morita S. Knowledge and practice for pressure injury prevention among care managers in a home care setting: a cross-sectional study. Chron Wound Care Manag Res 2017;4:99-105. [doi: 10.2147/cwcmr.s140673]
31. The Japanese Society of Pressure Ulcers. JSPU Guideline for Prevention and Management of Pressure Ulcers, 5th edition: Shorinsha Inc; 2022:1-112. [doi: 10.32249/9784796525534]
32. McKay FH, Slykerman S, Dunn M. The app behavior change scale: creation of a scale to assess the potential of apps to promote behavior change. JMIR Mhealth Uhealth 2019 Jan 25;7(1):e11130. [doi: 10.2196/11130] [Medline: 30681967]
33. Bricca A, Pellegrini A, Zangger G, Ahler J, Jäger M, Skou ST. The quality of health apps and their potential to promote behavior change in patients with a chronic condition or multimorbidity: systematic search in App Store and Google Play. JMIR Mhealth Uhealth 2022 Feb 4;10(2):e33168. [doi: 10.2196/33168] [Medline: 35119367]
34. Merolli M, Francis JJ, Vallance P, Bennell KL, Malliaras P, Hinman RS. Patient-facing mobile apps to support physiotherapy care: protocol for a systematic review of apps within app stores. JMIR Res Protoc 2021 Dec 9;10(12):e29047. [doi: 10.2196/29047] [Medline: 34889767]
35. Fulbrook P, Lawrence P, Miles S. Australian nurses' knowledge of pressure injury prevention and management: a cross-sectional survey. J Wound Ostomy Continence Nurs 2019;46(2):106-112. [doi: 10.1097/WON.0000000000000508] [Medline: 30801563]
36. Patel MS, Small DS, Harrison JD, et al. Effectiveness of behaviorally designed gamification interventions with social incentives for increasing physical activity among overweight and obese adults across the United States: the STEP UP randomized clinical trial. JAMA Intern Med 2019 Dec 1;179(12):1624-1632. [doi: 10.1001/jamainternmed.2019.3505] [Medline: 31498375]

Abbreviations

ABACUS: App Behavior Change Scale
CM: care manager
CONSORT: Consolidated Standards of Reporting Trials
PI: pressure injury
Pips-Map: pressure injury prevention support mobile app prototyping model

Edited by A Mavragani; submitted 26.02.24; peer-reviewed by Y Chen; revised version received 14.12.24; accepted 15.12.24; published 07.02.25.

Please cite as:
Kohta M, Takahashi M, Koyanagi H, Sugama J
Evaluating the Knowledge Level, Practice, and Behavioral Change Potential of Care Managers in Pressure Injury Prevention Using a Mobile App Prototyping Model in the Home-Care Setting: Single-Arm, Pre-Post Pilot Study
JMIR Form Res 2025;9:e57768
URL: https://formative.jmir.org/2025/1/e57768
doi:10.2196/57768

© Masushi Kohta, Mayumi Takahashi, Hiroe Koyanagi, Junko Sugama. Originally published in JMIR Formative Research (https://formative.jmir.org), 7.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction

XSL•FO
RenderX

in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Assessing Dynamic Cognitive Function in the Daily Lives of Youths With and Without Type 1 Diabetes: Usability Study

Mary Katherine Ray[1,2], PhD; Jorie Fleming[1], BA; Andrew Aschenbrenner[3], PhD; Jason Hassenstab[3], PhD; Brooke Redwine[1], BA; Carissa Burns[1], BA; Ana Maria Arbelaez[4], MD; Mary Ellen Vajravelu[5], MD; Tamara Hershey[1,2], PhD

[1]Department of Psychiatry, Washington University in St. Louis, 660 S. Euclid Avenue, St. Louis, MO, United States

[2]Department of Radiology, Washington University in St. Louis, St. Louis, MO, United States

[3]Department of Neurology, Washington University in St. Louis, St. Louis, MO, United States

[4]Department of Pediatrics, Washington University in St. Louis, St. Louis, MO, United States

[5]Division of Pediatric Endocrinology, Diabetes, and Metabolism, UPMC Children's Hospital of Pittsburgh, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

Corresponding Author:
Mary Katherine Ray, PhD
Department of Psychiatry, Washington University in St. Louis, 660 S. Euclid Avenue, St. Louis, MO, United States

Abstract

Background: Studies have shown a relationship between worse glycemic control and lower cognitive scores in youths with type 1 diabetes (T1D). However, most studies assess long-term glucose control (eg, years-decades) and cognition at a single time point. Understanding this relationship at a higher temporal resolution (eg, minutes-hours) and in naturalistic settings has potential clinical implications. Newer technology (eg, continuous glucose monitoring [CGM] and ecological momentary assessment) provides a unique opportunity to explore the glucose dynamics that influence dynamic cognition; that is, cognitive functions that fluctuate short-term and are influenced by environmental factors.

Objective: Before we can assess this relationship, we need to determine the feasibility of measuring cognition in youths in daily life and determine the plausibility of obtaining glucose variation with CGM to be integrated with real-time cognition measures. This study's purpose was to assess the acceptability of measuring dynamic cognition using a smartphone app and adherence to cognitive testing in daily life in youths with and without T1D. Further, we assessed CGM-derived glucose measures at temporally related timeframes to cognitive testing in naturalistic settings.

Methods: Data were obtained from 3 studies including one in-laboratory study and 2 remote studies. For all studies, youths were asked to complete cognitive tests on the Ambulatory Research in Cognition (ARC) smartphone app that measured processing speed, associative memory, and working memory. For the in-laboratory study, youths completed testing 4 times during 1 session. For the remote studies, youths were asked to complete cognitive tests 5 times per day for either 10 or 14 consecutive days in daily life. Youths were asked to rate their impressions of the app. Youths with T1D wore a CGM.

Results: 74 youths (n=53 control; n=21 T1D) aged 4 - 16 years participated. Youths generally reported liking or understanding the ARC app tasks in a laboratory and remote setting. Youths had high testing adherence in daily life (2350/3080 to 721/900, 76.3% - 80.2%) and none dropped out. The percentage of measurements within each glycemic range taken immediately before the app's cognitive testing was 3% (28/942) low glucose, 51% (484/942) euglycemia, 23% (221/942) high glucose, and 22% (210/942) very high glucose. In the 2-hour window before each cognitive task, mean glucose was 182.5 (SD 76.2) mg/dL, SD in glucose was 27.1 mg/dL (SD 18.7), and the mean maximum difference between the highest and lowest glucose was 85.5 (SD 53.7) mg/dL.

Conclusions: The results suggest that using the ARC smartphone app to assess dynamic cognitive functions in youths with and without T1D is feasible. Further, we showed CGM-derived glycemic variability at temporally associated timeframes of dynamic cognitive assessments. The next steps include using ecological momentary assessment in a fully powered study to determine the relationship between short-term glycemic control and cognition in youths with T1D.

(JMIR Form Res 2025;9:e60275) doi:10.2196/60275

KEYWORDS

ecological momentary assessment; EMA; ambulatory; smartphone; continuous glucose monitoring; CGM; assessment; daily lives; youth; type 1 diabetes; diabetes; feasibility study; pilot study; glycemic control; environmental factor; phone; acceptability; young; cognitive test; app; application; mobile phone

XSL•FO
RenderX

Introduction

Research has shown that youths with type 1 diabetes (T1D) often have slightly lower cognitive scores compared to their peers without T1D, and a relationship between worse glycemic control and lower cognitive scores [1-11]. These results highlight an important relationship between glucose and cognition in T1D. However, they are based on measures of long-term glucose control over long periods (eg, years-decades) and cognitive function tested in a laboratory setting at a single time point. Understanding this relationship at a higher temporal resolution (eg, minutes-hours) and in naturalistic settings has potential clinical implications.

Newer technology (eg, continuous glucose monitoring [CGM] and ecological momentary assessment [EMA]) provides a unique opportunity to explore the specific glucose patterns that influence dynamic cognition; that is, cognitive functions that fluctuate in the short-term and are easily influenced by environmental factors [12,13]. Using these approaches in adults with T1D, a few studies have shown an association between short-term glycemic control and cognitive functioning. Namely, studies have shown an association between significant glucose fluctuations and slower objective measures of processing speed at the moment [14], person-reported hypoglycemia and worse subjective measures of cognitive functioning later in the day [15], and increases in nocturnal hypoglycemia with slower processing speed the following day [16].

Whether this relationship or others are seen in youths with T1D is unknown. Understanding this relationship is significant given the importance of optimal cognitive functioning in academic settings in youths. However, before we can assess this relationship, we need to determine the feasibility and practicality of measuring dynamic cognition in youths in daily life. Although EMA is a feasible methodology for better understanding daily functioning in youths [17], including cognitive functioning [18], to our knowledge, no published study has assessed the feasibility of using a smartphone app to obtain EMAs of cognition in youths in the context of T1D.

This study aimed to test the feasibility of using a smartphone app called the Ambulatory Research in Cognition (ARC) app in youths. The ARC app, developed at Washington University in St. Louis, was originally designed to assess cognitive function in everyday environments in adults at risk of developing dementia [19]. Performance on ARC app cognitive tasks is sensitive to clinical status and genetic risk for Alzheimer disease [20,21] and can capture circadian fluctuations in cognition in adults at risk of Alzheimer disease [22]. The ability to capture variability in cognition in tandem with fluctuations in glucose in youths with T1D will be essential for determining the effects of glycemic variability on cognition in the daily lives of youths with T1D in future studies. Thus, in addition to testing the feasibility of the ARC app in youths, we also sought to determine the plausibility of obtaining glucose variation with CGM and integrating it with real-time measures of cognitive function using our protocol. This is important because EMA schedule selection has shown to be essential for capturing enough glycemic events throughout the day to integrate with cognitive data in adults with T1D [23].

Methods

Participants and Procedures

Recruitment

Study flyers were distributed throughout St. Louis Children's Hospital clinics, Midwest diabetes support groups, and the Washington University Volunteers for Health Research Registry program. The flyers included a short description of this study and the research team's contact information; interested parents were asked to contact the research team. St. Louis Children's Hospital health care workers also provided names of clinic patients who may be interested in this study to the research team. Participants were also recruited from word of mouth approaches with enrolled families telling friends about this study. After making contact with the families, the research team conducted a phone screen meeting to determine eligibility based on the inclusion and exclusion criteria outlined in Table 1. Data were obtained from 3 separate studies in youths with and without T1D. Study 1 was conducted in the laboratory setting, and studies 2 and 3 were conducted in remote settings. Of note, the study 2 inclusion criteria included owning an iPhone (6s or newer as this was the minimum required to support the ARC app) and established use of a Dexcom G5 or G6 CGM. For study 3, we obtained additional funding that allowed us to provide iPhones to youths who did not have their own and Dexcom G6 PRO CGMs to youths who did not already use a CGM. Given that we could provide CGMs to youths with T1D, we changed this study's protocol for study 3 from 14 days to 10 days to align with the life of a provided CGM (described in more detail below).

Figures and tables created with bioRender and REDCap (Research Electronic Data Capture) was used to support data collection [24-27].

XSL•FO
RenderX

Table . Inclusion and exclusion criteria for the pilot and feasibility studies that aimed at assessing the feasibility of using the Ambulatory Research in Cognition smartphone app to measure cognitive function in youths with and without type 1 diabetes in an in-laboratory setting (study 1), for 14 days in naturalistic settings (study 2), and for 10 days in naturalistic settings (study 3).

Study	Inclusion criteria	Exclusion criteria
Study 1: in laboratory (1 session)	• Age 4 - 16 years	• No English fluency • Unable to use a smartphone • Severe mental, neurological, or medical condition that would make them unable to complete tasks • Use of medications with known effects on the central nervous system (except ADHD[a] medicines)
Study 2: real world (14 days)	• Age 9 - 16 years • Owning an iPhone (6s or newer) • Established use of Dexcom G5/G6 continuous glucose monitoring	• No English fluency • Unable to use a smartphone • Severe mental, neurological, or medical condition that would make them unable to complete tasks • Use of medications with known effects on the central nervous system (except ADHD medicines) • No parent-approved screen time for participation • No tablet or computer with reliable internet connection • Severe phone screen cracks that would affect testing
Study 3: real world (10 days)	• Age 9 - 16 years	• No English fluency • Unable to use a smartphone • Severe mental, neurological, or medical condition that would make them unable to complete tasks • Use of medications with known effects on the central nervous system (except ADHD medicines) • No parent-approved screen time for participation • No tablet or computer with reliable internet connection • Severe phone screen cracks that would affect testing

[a]ADHD: attention-deficit/hyperactivity disorder.

Study 1

Participants completed a tutorial of all the app tasks and were able to ask the research team any clarifying questions. They completed app testing 4 times with 5-minute breaks and then completed an experiential interview to assess their impressions of the app in the laboratory setting.

Study 2

Participants had a video call with researchers to practice with the ARC app tasks and ask the research team any clarifying questions. The research team established CGM data sharing privileges through the Dexcom Clarity app in youths with T1D. The following day, all youths were asked to complete cognitive testing in daily life 5 times per day for 14 consecutive days. They had a video call with the research team 1 and 2 weeks later to complete an experiential interview.

Study 3

Participants had a video call or visited the laboratory to practice the app tasks and establish CGM data sharing privileges (T1D group). The following day, all youths were asked to cognitive complete testing in daily life 5 times per day for 10 consecutive days. For both remote studies, monetary incentives were provided to encourage testing compliance as outlined in the Ethical Considerations section below.

Ethical Considerations

The Washington University Human Research Protection Office (IRB00009237) approved all study procedures. Parents provided consent for them and their child to participate and youths provided assent. For privacy and confidentiality protection, all study data were deidentified. Compensation was provided to participants. For study 1, youths were compensated US $15 per hour (total of US $30), for their participation, paid via check. For study 2, youths could earn up to US $125 in Amazon gift cards for their participation (US $20 for participation in video

XSL•FO
RenderX

calls; US $70 for being in this study for 2 weeks—US $5 per day when they completed at least 2 of the 5 cognitive task sessions; US $35 for completing tests—US $0.50 per test). For study 3, youths could earn up to US $115 in Amazon gift cards for their participation (US $25 for an introductory session via Zoom [Zoom Communications, Qumu Corporation], US $40 for an introductory session at Washington University; US $50 for being in this study for 10 days—US $5 per day if they completed at least 3 of the 5 cognitive tasks; US $25 for completing tests—US $0.50 per test). For study 2 and 3, youths who were aged ≥13 years were also given a Fitbit to track sleep and activity (data not shown) and kept the tracker as compensation.

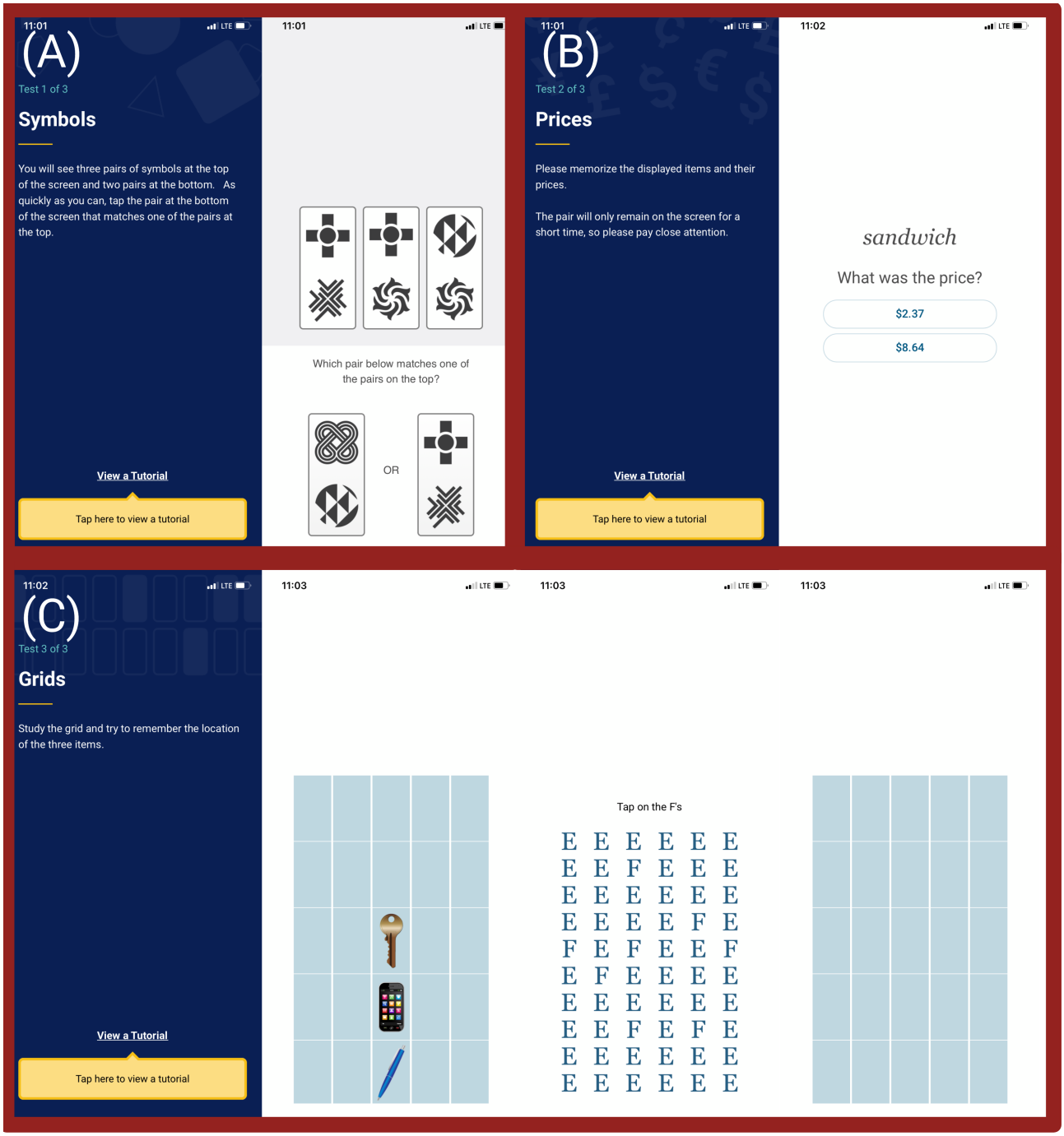
Measures

ARC App

The ARC app is comprised of 3 tasks. The symbols task assesses processing speed by asking participants to compare visual stimuli. Participants are shown 3 cards with 2 images at the top of a screen and 2 cards with 2 images at the bottom of a screen and asked to determine which card on the bottom matches a card on the top as quickly as possible (Figure 1A). There are 12 trials. The performance score is median reaction time, calculated for trials with correct responses and reaction times above 150 ms to avoid anticipatory responses. The prices task assesses associative memory. Participants are given a household item and an associated price. They are shown 10 pairs and asked to recall the price (Figure 1B). The performance score is percent error (0%=all correct, 100%=all incorrect, and 50%=chance performance). The grids task assesses working memory. Participants are shown a grid with 3 items and asked to remember their location (Figure 1C). They are presented with a distraction task where they are shown a screen of “E”s with sporadic “F”s and asked to select the “F”s. Then they are presented with a blank grid and asked to choose where the items were located. Participants are presented with this task twice per session. The performance score is the average Euclidean distance from the selected response to the correct placement (score range: 0 - 5.65; 0=perfect placement of all items and trials). There was a “tutorial” button on each task interface throughout this study so youths could reorient themselves to the tasks before completing the actual tasks being recorded for this study if needed.

XSL•FO
RenderX

Figure 1. The Ambulatory Research in Cognition app being tested in this pilot and feasibility study. (**A**) The symbols task assesses processing speed. (**B**) The prices task assesses associative memory. (**C**) The grids task assesses working memory.



Each session, including all 3 tasks, only takes about 3 minutes to complete. For both remote studies, youths received notifications from the app to take the tests randomly and had 2 hours to complete them. This sampling method was chosen to capture variability in cognitive function and glycemic control throughout the entire day when youths were in different contexts in daily life. Youths and their parents were able to choose the timeframes when they wanted to receive notifications and could change these times through the app to reduce the burden and reduce the likelihood of receiving a notification during times that would affect daily functioning (eg, overnight while sleeping, sleeping in on the weekend, or in class at school). Parents were not explicitly told to encourage youths to complete the sessions because the goal of using EMA methodology in this context is for youths to be able to complete testing unsupervised. Relatedly, youths were told to answer questions on the ARC app tests without help from others. All participants completed the tasks using an iPhone (6s or newer). Compliance to app testing was automatically logged electronically through an ARC app dashboard. Although the app was developed for an adult population [19], the app was used as is in our study of youths; if deemed acceptable for youths in its original state, it would be a cost-effective method that could be easily implemented into study protocols.

Experiential Interview

For each task in study 1 and study 2, youths were asked to rate on a scale of 1 - 5 how challenging or easy the task was, how

XSL•FO
RenderX

confusing or clear the instructions were, how confusing or clear the task was, and how boring or fun the task was. The Likert scale had an anchor of 3 to designate a neutral appraisal for each question. For example, "How confusing were the instructions for the symbols game?" Option 3 was "not confusing or clear."

About CGM

For studies 2 and 3, participants with T1D wore a Dexcom G5/G6 CGM to measure glucose every 5 minutes. Youths who already used a Dexcom CGM changed their devices as they normally would. For study 3, if youths with T1D did not already use a Dexcom CGM, they were provided with a Dexcom G6 PRO in blinded mode to wear for the duration of this study.

Statistical Analysis

Data were analyzed with R (R Foundation), Python (Python Software Foundation), and SPSS (IBM Corp). ARC app performance: Mean and SD were calculated for performance on each ARC app task. For study 1, Pearson correlations also assessed the relationship between age and ARC task performance (Kendall τ was used for nonnormal data), and scatterplots were created to visually appraise the age at which youths started performing better on the ARC app tasks. Further, 2-tailed *t* tests determined differences in performance between age groups. Experiential interview ratings: mean and SD were calculated for each question assessing impressions of the app tasks; a mean score greater than 3 was determined to be a positive impression of the task. ARC app adherence: Adherence was calculated as the number of tests completed divided by the number of tests offered.

CGM: Glucose variable descriptives were calculated from the raw Dexcom Clarity exports for all participants with T1D across the duration of this study, including mean, SD, percent time in range (TIR; 70 - 180 mg/dL), percent time below range (<70 mg/dL), percent time high (≥180 mg/dL), and percent time very high (≥250 mg/dL). The glucose measure taken immediately before each cognitive task for each individual with T1D was extracted. Glucose obtained from the CGM were matched to the corresponding cognitive test by timestamps within each system, both set to CST. Given that glucose was assessed every 5 minutes, the closest glucose measure from CGM was taken within 5 minutes of each cognitive session. Lastly, the following glycemic variables were extracted for the 2-hour window immediately before each cognitive task for each individual with T1D: mean glucose, SD of glucose, and maximum difference in glucose (ie, highest glucose within a 2 h block minus lowest glucose within the 2 h block). Mean sensor usage during participation was calculated; Dexcom Clarity defines sensor usage as the number of days during the reporting period with at least 50% CGM readings [28]. Deletion by list was used for missing CGM data.

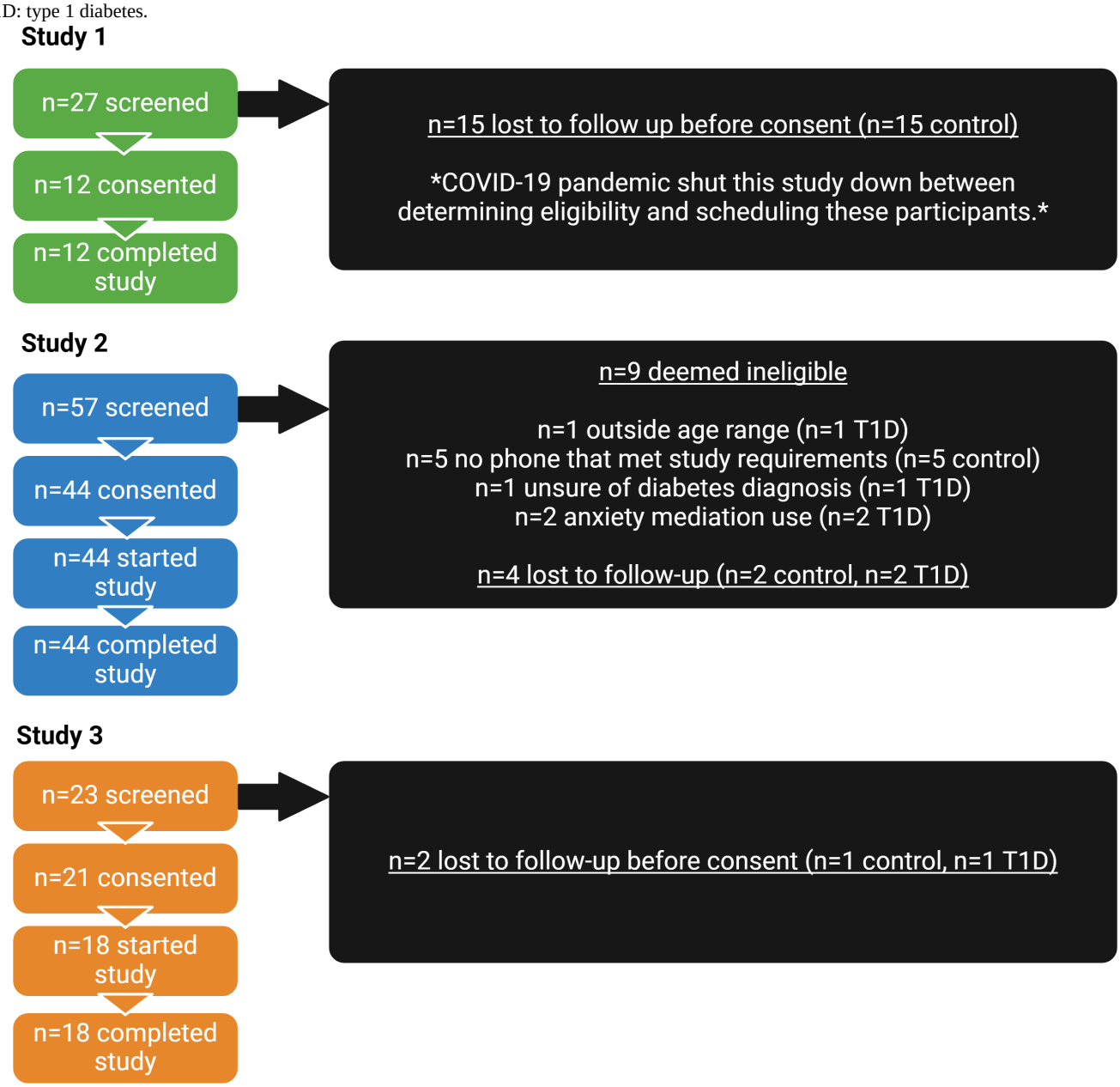
Results

Participants

Details for eligibility and consent rate are outlined in Figure 2.

XSL•FO
RenderX

Figure 2. Eligibility flowchart for the pilot and feasibility studies aimed at assessing the feasibility of using the Ambulatory Research in Cognition smartphone app to assess cognitive function in the daily lives of youths with and without T1D. Study 1 was conducted in a single session in the laboratory setting in youths without T1D. Studies 2 and 3 were conducted in naturalistic settings for 14 and 10 days, respectively, in youths with and without T1D. T1D: type 1 diabetes.



Ultimately, a total of 74 participants aged 4 - 16 years completed study procedures (study 1: n=12 control; study 2: n=44; n=30 control, n=14 T1D; study 3: n=18; n=11 control, n=7 T1D). No participants were voluntarily or involuntarily withdrawn from the studies. Demographic data are shown in Table 2. This study's samples were relatively homogenous, with most participants being non-Hispanic White. For study 1, there was a glitch with the ARC app for 1 participant; thus, this participant did not have cognitive function data. For study 3, a total of 8 participants were provided with an iPhone, and all iPhones were returned to this study's team after participation. The total annual household income for participants enrolling in this study with their own iPhone was US $131,111, and the annual household income for participants who were provided with an iPhone was US $68,750.

XSL•FO
RenderX

Table . Demographic data for the pilot and feasibility studies that aimed to assess the feasibility of using the Ambulatory Research in Cognition smartphone app to measure cognitive function in youths with and without type 1 diabetes in an in-laboratory setting (study 1), for 14 days in naturalistic settings (study 2), and for 10 days in naturalistic settings (study 3).

			Group 1	Group 2
Study 1, n			12[a]	7[b]
	Age (years), mean (SD)		10.6 (3.6)	13.1 (2.3)
	Sex, n			
		Female	7	5
		Male	5	2
	Race or ethnicity, n (%)			
		Black	3 (25)	2 (29)
		Hispanic White	1 (8)	1 (14)
		Non-Hispanic White	8 (67)	4 (57)
Study 2, n			30[c]	14[d]
	Age (years), mean (SD)		13.0 (1.7)	13.4 (2)
	Sex, n			
		Female	17	7
		Male	13	7
	Race or ethnicity, n (%)			
		Black	2 (7)	0 (0)
		Hispanic	1 (3)	0 (0)
		Non-Hispanic White	25 (83)	14 (100)
		Mixed race	2 (7)	0 (0)
	Annual household income (US $), mean (SD)		138,907 (71,860)[e]	125,231 (44,129)[f]
Study 3, n			11[c]	7[d]
	Age (years), mean (SD)		13 (2.3)	12.4 (2.2)
	Sex, n			
		Female	5	1
		Male	6	6
	Race or ethnicity, n (%)			
		Black	0 (0)	0 (0)
		Hispanic	0 (0)	0 (0)
		White	11 (100)	7 (100)
		Mixed race	0 (0)	0 (0)
	Annual household income (US $), mean (SD)		91,818 (37,031)	120,000 (47,010)[g]

[a]Overall

[b]≥9 years old

[c]Control.

[d]Type 1 diabetes.

[e]n=27

[f]n=13

[g]n=6

Although the protocol for study 3 offered CGMs to use for the duration of this study for youths who did not use them, all participants enrolled already used a Dexcom CGM and used their own for this study. For study 3, one participant with T1D and his family could not sync their CGM data to Dexcom Clarity. Mean sensor usage was 97%. Average glycemic

XSL•FO
RenderX

measures for youths with T1D for the entire study duration are shown in Table 3. On average, youths with T1D in our sample did not meet the recommended guidelines for percent TIR or percent time in hyperglycemia (59% vs recommended 70% TIR; 39% vs recommended 25% time in hyperglycemia) [29].

Table . Glycemic measures for youths with type 1 diabetes as measured by continuous glucose monitoring across studies 2 and 3 when youths were completing cognitive tasks on the Ambulatory Research in Cognition app in naturalistic settings for 14 and 10 days, respectively. Data are presented as mean (SD) across all days of study participation (n=20). We were unable to collect continuous glucose monitoring data in 1 participant with T1D.

	Values, mean (SD)
Mean glucose (mg/dL)	176 (46)
SD glucose (mg/dL)	63.9 (17.7)
Time in range (%, 70 - 180 mg/dL)	58.7 (20.6)
Time below range (%, <70 mg/dL)	2.5 (3.3)
Time high (%, ≥180 mg/dL)	38.8 (22.3)
Time very high (%, ≥250 mg/dL)	17.2 (19.1)

Feasibility and Acceptability of Using a Smartphone App to Assess Dynamic Cognitive Function in Youths

Study 1: Youths reported liking and understanding the grids and symbols tasks but not the prices task (Table 4).

XSL•FO
RenderX

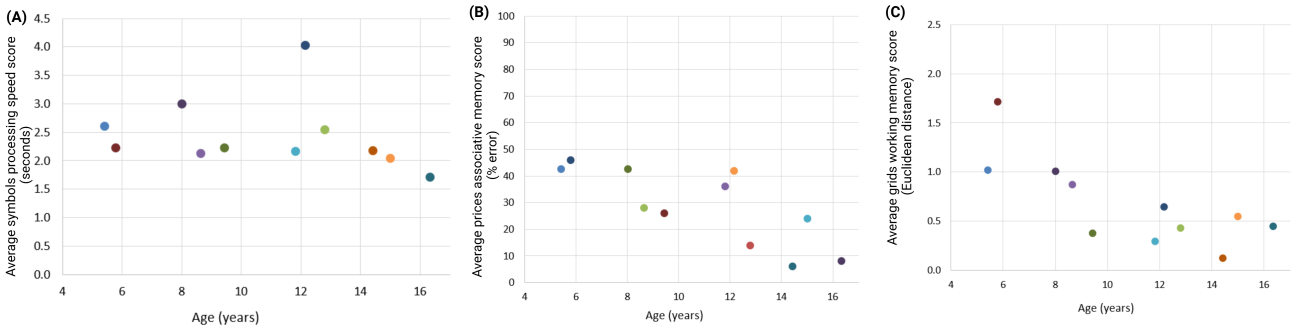
Table . Youths' impression of and performance on the Ambulatory Research in Cognition smartphone app cognitive tasks being tested in these pilot and feasibility studies. Study 1 was conducted in the laboratory setting in youths without type 1 diabetes. Studies 2 and 3 were conducted in naturalistic settings in youths with and without type 1 diabetes for 14 or 10 consecutive days, respectively. There was an app malfunction with 1 participant from study 1, so there are only data for 11 participants.

		Symbols (processing speed), median reaction time (seconds)		Prices (associative memory), error score (%)		Grids (working memory), Euclidean distance	
		Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Performance (lower score=better)							
	Study 1 (n=11)	2.4 (0.6)	1.7-4	28.6 (14.5)	6-46	0.7 (0.4)	0.1-1.7
	Study 2 (n=44)	1.8 (0.3)	1-2.7	40.6 (7.7)	10.3-50.8	0.5 (0.2)	0.1-1.3
	Study 3 (n=18)	1.9 (0.5)	1.3-2.8	40.1 (6.7)	27.1-50.8	0.5 (0.2)	0.2-0.9
Impressions							
	How hard or easy was the task? (1=very hard and 5=very easy)						
	Study 1 (n=12)	4.2 (0.7)	3 - 5	2.3 (0.8)	1 - 4	3.3 (1.4)	1 - 5
	Study 2 week 1 (n=42)	4.3 (0.8)	2 - 5	2.2 (1.1)	1 - 4	4 (1)	2 - 5
	Study 2 week 2 (n=43)	4.5 (0.8)	2 - 5	2.6 (1.2)	1 - 5	4.1 (0.9)	2 - 5
	How confusing or clear were the task instructions? (1=very confusing and 5=very clear)						
	Study 1 (n=12)	4.4 (1.1)	2 - 5	4.4 (1)	2 - 5	4.5 (0.9)	3 - 5
	Study 2 week 1 (n=42)	4.9 (0.3)	4 - 5	4.8 (0.4)	4 - 5	4.9 (0.4)	3 - 5
	Study 2 week 2 (n=43)	4.9 (0.4)	3 - 5	4.9 (0.4)	3 - 5	4.9 (0.4)	3 - 5
	How confusing or clear was the task itself? (1=very confusing and 5=very clear)						
	Study 1 (n=12)	4.7 (0.7)	3 - 5	3.8 (1.4)	2 - 5	4.3 (1.4)	1 - 5
	Study 2 week 1 (n=42)	4.7 (0.7)	2 - 5	4.3 (1.1)	2 - 5	4.7 (0.7)	2 - 5
	Study 2 week 2 (n=43)	4.8 (0.4)	4 - 5	4.4 (0.9)	2 - 5	4.7 (0.6)	2 - 5
	How boring or fun was the task? (1=very boring and 5=very fun)						
	Study 1 (n=12)	3.6 (1)	2 - 5	3.8 (0.8)	3 - 5	4.5 (0.7)	3 - 5
	Study 2 week 1 (n=42)	3.8 (0.9)	2 - 5	2.8 (1.1)	1 - 5	4.1 (0.8)	2 - 5
	Study 2 week 2 (n=43)	3.8 (1)	1 - 5	2.7 (1)	1 - 5	4 (0.6)	3 - 5

Performance on the tasks correlated with age (grids: R=−0.77, P=.006, n=11; prices: R=−0.77, P=.005, n=11; symbols: τ =−0.42, P=.07, n=11), such that younger youths performed worse than older youths. Visually, scatterplots showed that around the age of 9 years, youths started performing better (Figure 3).

XSL•FO
RenderX

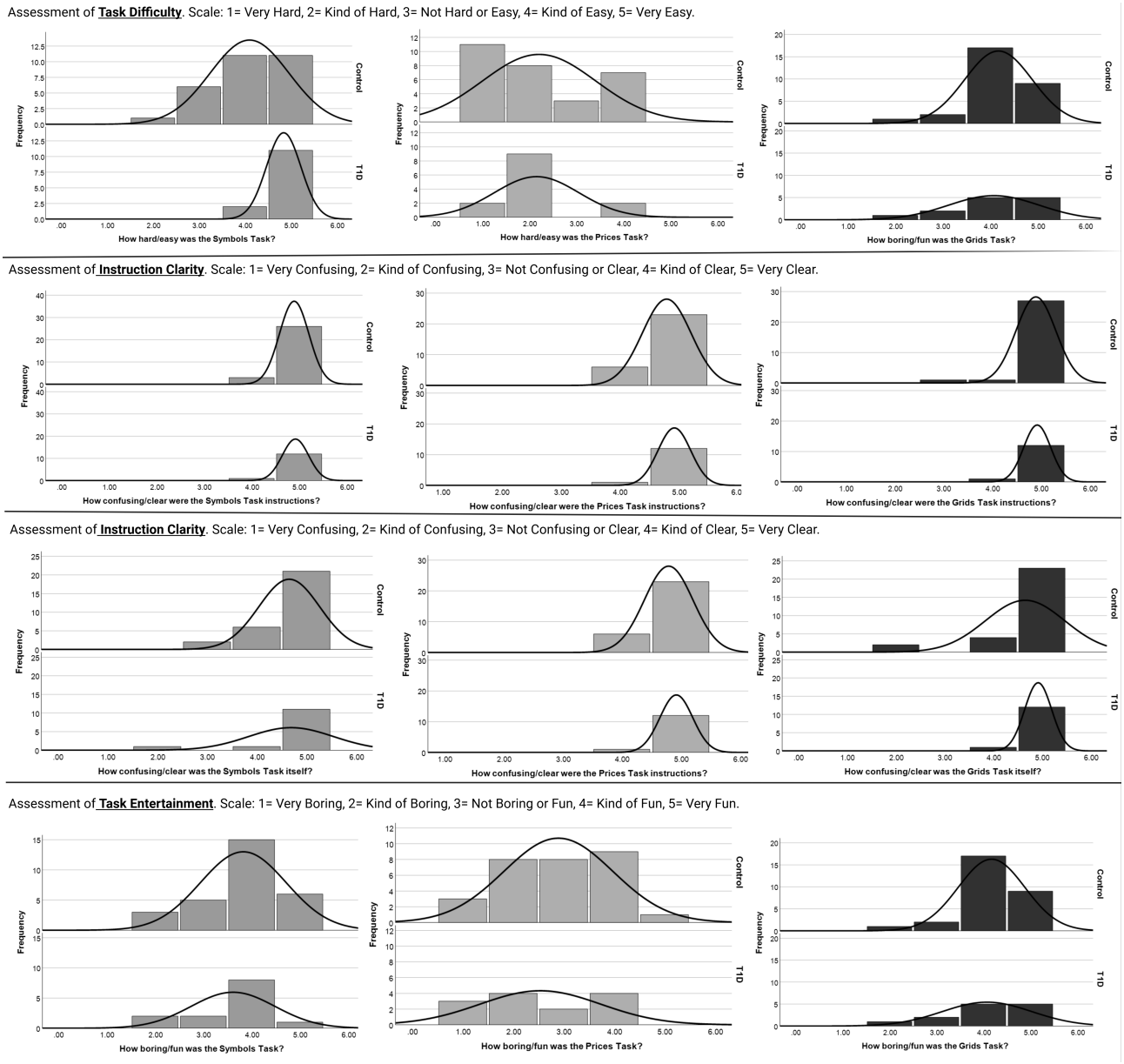
Figure 3. Scatterplots representing performance on each of the Ambulatory Research in Cognition smartphone app tasks and age during study 1 that aimed to test the Ambulatory Research in Cognition app in youths without type 1 diabetes at a single time point in the laboratory setting. Each dot represents an individual participant. Lower scores on each task represent better cognitive performance. The (**A**) symbols task assesses processing speed, (**B**) prices task assesses associative memory, and (**C**) grids task assesses working memory.



Based on these visual representations, we assessed performance in youths aged between <9 years and ≥9 years. There were statistically significant differences in performance such that youths aged <9 years had worse performance than youths aged ≥9 years on the grids task (<9 years: mean 1.2, SE 0.19; ≥9 years: mean 0.41, SE 0.10; *P*=.001) and prices task (<9 years: mean 39.8, SE 4.00; ≥9 years: mean 22.3, SE 5.19; *P*=.047). These data guided our decision to include ages 9 - 16 years in study 2 and study 3. For study 2, youths completed 76.3% (SD 19) of offered tasks, and for study 3, youths completed 80.2% (SD 13) of offered tasks. Across both studies, youths reported liking and understanding the grids and symbols tasks but not the prices task (Table 4). Figure 4 shows histograms for the distribution of task rankings for the control and T1D groups.

XSL•FO
RenderX

Figure 4. Histograms displaying ratings of the Ambulatory Research in Cognition app symbols, prices, and grids tasks from the experiential interview in study 2 which was conducted in naturalistic settings for 14 consecutive days in youths with and without type 1 diabetes. Ratings for the symbols task, which assesses processing speed, is shown in blue. Ratings for the prices task, which assesses associative memory is shown in green. Ratings for the grids task, which assesses working memory, is shown in red.



Glycemic Variability and Cognitive Function Data Integration

Across all participants with T1D from studies 2 and 3, the percentage of measurements within each glycemic range for the glucose measure taken immediately before cognitive testing was as follows: 3% (28/942) low glucose, 51% (484/942) euglycemia, 23% (221/942) high glucose, and 22% (210/942) very high glucose. In the 2-hour window immediately before each cognitive task, mean glucose was 182.5 (SD 76.2) mg/dL, SD in glucose was 27.1 (SD 18.7) mg/dL, and the mean maximum difference between the highest and lowest glucose was 85.5 (SD 53.7) mg/dL.

Discussion

Principal Findings

This study aimed to assess the feasibility of using a smartphone app to measure dynamic cognitive function in youths in daily life and to integrate this information with data obtained from CGM in youths with T1D. Our results suggest that using the ARC smartphone app to assess dynamic cognitive functions in youths with and without T1D is feasible. Further, we showed glycemic variability from CGM measures at temporally associated timeframes of dynamic cognitive function measures using our EMA schedule.

XSL•FO
RenderX

Feasibility and Acceptability of Using a Smartphone App to Assess Dynamic Cognitive Function in Youths

Testing of the app in an in-laboratory setting revealed that youths aged younger than 9 years performed worse than youths aged ≥9 years. Differences in performance on the ARC app in the laboratory setting guided our decision to include youths aged ≥9 years in the remote testing phases of the research. Although we expected older children to perform better than younger youths simply because of normal cognitive development progression, we wanted to be cautious of including youths in remote settings who may not understand the tasks enough to be able to complete them independently. Thus, for remote testing, we chose to focus on older children whose performance was uniformly better in our in-laboratory data. Future research is needed to determine if the ARC app could be used in younger youths in remote settings, and if not, what stimuli would better engage and test cognition in younger children.

Testing the app in naturalistic settings in youths, we found that youths had high testing adherence in daily life (76% - 80%). Although we expected lower adherence in youths given the amount of time spent at school and participating in extracurricular activities, compliance in our study was similar to that of adults from a previous study using the ARC app (~80%) [19]. Further, none of all the 62 youths who started with the remote study procedures dropped out. Furthermore, youths had high appraisal for 2/3 cognitive tasks, including the grids working memory task and the symbols processing speed task. It is also important to note that providing iPhones to youths who did not have their own for participation (study 3) helped expand our reach to include youths from lower income communities compared to this study's protocol when youths were required to have their own iPhones (study 2). Of note, all iPhones provided to participants for this study were returned to the research team which shows the logistical feasibility of this provision. Given that worse T1D outcomes have been associated with lower income [30-32], reaching youths from lower SES communities in future studies will be crucial to understand the relationship between short-term glycemic control and dynamic cognitive function. Taken together, our data support the use of the ARC smartphone app to measure dynamic cognition in youths aged 9 - 16 years, with and without T1D in daily life, and highlight the logistical feasibility of providing devices to youths who do not have their own.

When integrating glucose data and cognitive function data, we found that there was variation in glucose at the time of cognitive testing and during a short timeframe immediately before cognitive testing (eg, 2 h before). Specifically, youths with T1D had dysglycemia at the time of almost half of cognitive tests (49%) and on average, experienced 86 mg/dL swings in glucose in the 2 hours immediately before cognitive testing. The ability to capture glucose at various ranges (eg, low glucose or very high glucose) and variability in glucose (eg, amount of change in glucose during short periods before cognitive testing) using this methodology will be necessary for future studies aimed at determining the relationship between short-term glycemic control and dynamic cognitive function in youths with T1D.

Of note, only 3% of cognitive tasks were completed when glucose was in the hypoglycemia range in youths with T1D. Although this small percentage does not provide a significant amount of data for assessing the effects of low glucose on cognitive function, many other glycemic variables may be important for cognitive functioning that were captured frequently in our sample, including very high glucose at the time of testing and amount of change in glucose before cognitive testing. For example, using single measures of self-monitored blood glucose in the home, Gonder-Frederick et al [33] found that severe hyperglycemia (>400 mg/dl) was associated with equally substantial deteriorations in cognition in youths with T1D than severe hypoglycemia [33]. This study highlights the need to collect measures of hyperglycemia in addition to hypoglycemia in this population. Further, the use of CGM may allow future studies to capture glycemic patterns that may predict cognitive functioning outside of standard hypoglycemia and hyperglycemia cutoffs (eg, swings in glucose).

There are study limitations to consider. First, our sample was relatively homogenous. Although we were able to reach youths from lower SES communities in study 3 when we provided iPhones for this study, our sample with T1D lacked racial and ethnic diversity. Practices are needed to better engage youths from all communities in future work. Additionally, there is a documented lag between interstitial glucose obtained from CGM and blood glucose in past literature [34,35]. Given that lag time is patient-specific [36], we chose to use the glucose measurement obtained closest to each individual cognitive test to illustrate the integration of glucose measured via CGM and cognition measured via the smartphone app; future studies assessing the relationship between short-term glucose control and dynamic cognitive function need to consider multiple glycemic features that account for lag. In the same vein, we did not ask youths to report the cause for CGM sensor errors during their participation. Thus, if youths had missing CGM data, it was unclear why (eg, intrinsic sensor failure or intentional removal of the sensor). Future EMA studies should ask participants to report the cause of missing glucose data to help account for missingness when assessing the relationship between dynamic cognitive function and short-term glucose control.

In conclusion, we found that it is feasible to obtain measures of dynamic cognitive function in youths in daily life using a smartphone app and that we could capture glucose variability for integration with measures of glucose in the daily lives of youths with T1D. Using this methodology in fully powered studies assessing the relationship between short-term glycemic control in real-time cognition would move the field toward a fuller understanding of the impacts of T1D on cognitive function in youths.

XSL•FO
RenderX

Acknowledgments

This work was made possible with support from WU-CDTR (Washington University in St. Louis-Center for Diabetes Translation Research; P30DK092950), Washington University in St. Louis DRC (Diabetes Research Center; P30DK020579), and Washington University in St. Louis and BJC HealthCare Innovation Lab Big Ideas grant. MKR was supported by National Center for Advancing Translational Sciences of the National Institutes of Health Career Development Award (KL2TR002346), National Institute of Diabetes and Digestive and Kidney Diseases Career Development Award (NIDDK; 1K01DK131339), Washington University in St. Louis Biomedical Research Training in Drug Abuse training grant (T32DA007261), and Washington University in St. Louis Transdisciplinary Postdoctoral Training Program in Obesity and Cardiovascular Disease (T32HL130357). AA was supported by National Institute of Aging (NIA; K01AG071847). MEV was supported by National Institute of Diabetes and Digestive and Kidney Diseases (K23DK125719). The content is solely the responsibility of the authors and does not necessarily represent the official views of the WU-CDTR, DRC, NIDDK, NIA, or Healthcare Innovation Lab.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Shalimova A, Graff B, Gąsecki D, et al. Cognitive dysfunction in type 1 diabetes mellitus. J Clin Endocrinol Metab 2019 Jun 1;104(6):2239-2249. [doi: 10.1210/jc.2018-01315] [Medline: 30657922]
2. Cato A, Hershey T. Cognition and type 1 diabetes in children and adolescents. Diabetes Spectr 2016 Nov;29(4):197-202. [doi: 10.2337/ds16-0036] [Medline: 27899870]
3. Cato MA, Mauras N, Mazaika P, et al. Longitudinal evaluation of cognitive functioning in young children with type 1 diabetes over 18 months. J Int Neuropsychol Soc 2016 Mar;22(3):293-302. [doi: 10.1017/S1355617715001289] [Medline: 26786245]
4. Semenkovich K, Bischoff A, Doty T, et al. Clinical presentation and memory function in youth with type 1 diabetes. Pediatr Diabetes 2016 Nov;17(7):492-499. [doi: 10.1111/pedi.12314] [Medline: 26377697]
5. Blasetti A, Chiuri RM, Tocco AM, et al. The effect of recurrent severe hypoglycemia on cognitive performance in children with type 1 diabetes: a meta-analysis. J Child Neurol 2011 Nov;26(11):1383-1391. [doi: 10.1177/0883073811406730] [Medline: 21572053]
6. Ghetti S, Lee JK, Sims CE, Demaster DM, Glaser NS. Diabetic ketoacidosis and memory dysfunction in children with type 1 diabetes. J Pediatr 2010 Jan;156(1):109-114. [doi: 10.1016/j.jpeds.2009.07.054] [Medline: 19833353]
7. Kirchhoff BA, Jundt DK, Doty T, Hershey T. A longitudinal investigation of cognitive function in children and adolescents with type 1 diabetes mellitus. Pediatr Diabetes 2017 Sep;18(6):443-449. [doi: 10.1111/pedi.12414] [Medline: 27444539]
8. Semenkovich K, Patel PP, Pollock AB, et al. Academic abilities and glycaemic control in children and young people with type 1 diabetes mellitus. Diabet Med 2016 May;33(5):668-673. [doi: 10.1111/dme.12854] [Medline: 26173465]
9. He J, Ryder AG, Li S, Liu W, Zhu X. Glycemic extremes are related to cognitive dysfunction in children with type 1 diabetes: a meta-analysis. J Diabetes Investig 2018 Nov;9(6):1342-1353. [doi: 10.1111/jdi.12840] [Medline: 29573221]
10. Gaudieri PA, Chen R, Greer TF, Holmes CS. Cognitive function in children with type 1 diabetes: a meta-analysis. Diabetes Care 2008 Sep;31(9):1892-1897. [doi: 10.2337/dc07-2132] [Medline: 18753668]
11. Naguib JM, Kulinskaya E, Lomax CL, Garralda ME. Neuro-cognitive performance in children with type 1 diabetes--a meta-analysis. J Pediatr Psychol 2009 Apr;34(3):271-282. [doi: 10.1093/jpepsy/jsn074] [Medline: 18635605]
12. Gamaldo AA, Allaire JC. Daily fluctuations in everyday cognition: is it meaningful? J Aging Health 2016 Aug;28(5):834-849. [doi: 10.1177/0898264315611669] [Medline: 26538267]
13. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
14. Hawks ZW, Beck ED, Jung L, et al. Dynamic associations between glucose and ecological momentary cognition in type 1 diabetes. NPJ Digit Med 2024 Mar 18;7(1):59. [doi: 10.1038/s41746-024-01036-5] [Medline: 38499605]
15. Søholm U, Broadley M, Zaremba N, et al. The impact of hypoglycaemia on daily functioning among adults with diabetes: a prospective observational study using the Hypo-METRICS app. Diabetologia 2024 Oct;67(10):2160-2174. [doi: 10.1007/s00125-024-06233-1] [Medline: 39080044]
16. Zuniga-Kennedy M, Wang OH, Fonseca LM, et al. Nocturnal hypoglycemia is associated with next day cognitive performance in adults with type 1 diabetes: pilot data from the GluCog study. Clin Neuropsychol 2024 Oct;38(7):1627-1646. [doi: 10.1080/13854046.2024.2315749] [Medline: 38380810]
17. Russell MA, Gajos JM. Annual research review: ecological momentary assessment studies in child psychology and psychiatry. J Child Psychol Psychiatry 2020 Mar;61(3):376-394. [doi: 10.1111/jcpp.13204] [Medline: 31997358]

XSL•FO
RenderX

18. Könen T, Dirk J, Schmiedek F. Cognitive benefits of last night's sleep: daily variations in children's sleep behavior are related to working memory fluctuations. J Child Psychol Psychiatry 2015 Feb;56(2):171-182. [doi: 10.1111/jcpp.12296] [Medline: 25052368]
19. Nicosia J, Aschenbrenner AJ, Balota DA, et al. Unsupervised high-frequency smartphone-based cognitive assessments are reliable, valid, and feasible in older adults at risk for Alzheimer's disease. J Int Neuropsychol Soc 2023 Jun;29(5):459-471. [doi: 10.1017/S135561772200042X] [Medline: 36062528]
20. Welhaf MS, Wilks H, Aschenbrenner AJ, et al. Naturalistic assessment of reaction time variability in older adults at risk for Alzheimer's disease. J Int Neuropsychol Soc 2024 Jun;30(5):428-438. [doi: 10.1017/S1355617723011475] [Medline: 38282413]
21. Aschenbrenner AJ, Hassenstab J, Morris JC, Cruchaga C, Jackson JJ. Relationships between hourly cognitive variability and risk of Alzheimer's disease revealed with mixed-effects location scale models. Neuropsychology 2024 Jan;38(1):69-80. [doi: 10.1037/neu0000905] [Medline: 37079810]
22. Wilks H, Aschenbrenner AJ, Gordon BA, et al. Sharper in the morning: cognitive time of day effects revealed with high-frequency smartphone testing. J Clin Exp Neuropsychol 2021 Oct;43(8):825-837. [doi: 10.1080/13803395.2021.2009447] [Medline: 35037593]
23. Mascarenhas Fonseca L, Strong RW, Singh S, et al. Glycemic Variability and Fluctuations in Cognitive Status in Adults With Type 1 Diabetes (GluCog): observational study using ecological momentary assessment of cognition. JMIR Diabetes 2023 Jan 5;8:e39750. [doi: 10.2196/39750] [Medline: 36602848]
24. Create professional science figures in minutes. bioRender. URL: https://www.biorender.com/ [accessed 2025-01-23]
25. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)-a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
26. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
27. Lawrence CE, Dunkel L, McEver M, et al. A REDCap-based model for electronic consent (eConsent): moving toward a more personalized consent. J Clin Transl Sci 2020 Apr 3;4(4):345-353. [doi: 10.1017/cts.2020.30] [Medline: 33244416]
28. Glossary. Dexcom Clarity. URL: https://clarity.dexcom.com/glossary#eA1cTopic [accessed 2025-01-23]
29. de Bock M, Codner E, Craig ME, et al. ISPAD Clinical Practice Consensus Guidelines 2022: glycemic targets and glucose monitoring for children, adolescents, and young people with diabetes. Pediatr Diabetes 2022 Dec;23(8):1270-1276. [doi: 10.1111/pedi.13455] [Medline: 36537523]
30. Secrest AM, Costacou T, Gutelius B, Miller RG, Songer TJ, Orchard TJ. Associations between socioeconomic status and major complications in type 1 diabetes: the Pittsburgh Epidemiology of Diabetes Complication (EDC) Study. Ann Epidemiol 2011 May;21(5):374-381. [doi: 10.1016/j.annepidem.2011.02.007] [Medline: 21458731]
31. Simba S, Von Oettingen JE, Rahme E, Ladd JM, Nakhla M, Li P. Socioeconomic disparities in glycemic management in children and youth with type 1 diabetes: a retrospective cohort study. Can J Diabetes 2023 Dec;47(8):658-664. [doi: 10.1016/j.jcjd.2023.07.005] [Medline: 37481125]
32. Mönkemöller K, Müller-Godeffroy E, Lilienthal E, et al. The association between socio-economic status and diabetes care and outcome in children with diabetes type 1 in Germany: the DIAS study (diabetes and social disparities). Pediatr Diabetes 2019 Aug;20(5):637-644. [doi: 10.1111/pedi.12847] [Medline: 30912245]
33. Gonder-Frederick LA, Zrebiec JF, Bauchowitz AU, et al. Cognitive function is disrupted by both hypo- and hyperglycemia in school-aged children with type 1 diabetes: a field study. Diabetes Care 2009 Jun;32(6):1001-1006. [doi: 10.2337/dc08-1722] [Medline: 19324943]
34. Wadwa RP, Laffel LM, Shah VN, Garg SK. Accuracy of a factory-calibrated, real-time continuous glucose monitoring system during 10 days of use in youth and adults with diabetes. Diabetes Technol Ther 2018 Jun;20(6):395-402. [doi: 10.1089/dia.2018.0150] [Medline: 29901421]
35. Sinha M, McKeon KM, Parker S, et al. A comparison of time delay in three continuous glucose monitors for adolescents and adults. J Diabetes Sci Technol 2017 Nov;11(6):1132-1137. [doi: 10.1177/1932296817704443] [Medline: 28459159]
36. Schmelzeisen-Redeker G, Schoemaker M, Kirchsteiger H, Freckmann G, Heinemann L, Del Re L. Time delay of CGM sensors: relevance, causes, and countermeasures. J Diabetes Sci Technol 2015 Aug 4;9(5):1006-1015. [doi: 10.1177/1932296815590154] [Medline: 26243773]

Abbreviations

ARC: Ambulatory Research in Cognition
CGM: continuous glucose monitoring
EMA: ecological momentary assessment
REDCap: Research Electronic Data Capture
T1D: type 1 diabetes
TIR: time in range

XSL•FO
RenderX

Edited by A Mavragani; submitted 06.05.24; peer-reviewed by L Germine, NS Chaytor; revised version received 20.12.24; accepted 21.12.24; published 11.02.25.

Please cite as:
Ray MK, Fleming J, Aschenbrenner A, Hassenstab J, Redwine B, Burns C, Arbelaez AM, Vajravelu ME, Hershey T
Assessing Dynamic Cognitive Function in the Daily Lives of Youths With and Without Type 1 Diabetes: Usability Study
JMIR Form Res 2025;9:e60275
URL: https://formative.jmir.org/2025/1/e60275
doi:10.2196/60275

© Mary Katherine Ray, Jorie Fleming, Andrew Aschenbrenner, Jason Hassenstab, Brooke Redwine, Carissa Burns, Ana Maria Arbelaez, Mary Ellen Vajravelu, Tamara Hershey. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Use of a Technology-Based Fall Prevention Program With Visual Feedback in the Setting of Early Geriatric Rehabilitation: Controlled and Nonrandomized Study

Carolin Steinmetz[1], PhD; Christina Stenzel[2], BSc; Maj Sylvester[2], BSc; Denis Glage[2], PhD; Anne Linke[1], MSc; Monika Sadlonova[1,3,4,5], MD; Christine A F von Arnim[1,5], MD; Marlena Schnieder[1,6], MD; Miroslava Valentová[1,5], MD; Stephanie Heinemann[1], PhD

[1]Department of Geriatrics, University Medical Center Göttingen, Robert-Koch-Straße 40, Göttingen, Germany

[2]Institute of Sports Science, Georg-August-University Göttingen, Göttingen, Germany

[3]Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Göttingen, Göttingen, Germany

[4]Department of Cardiovascular and Thoracic Surgery, University Medical Center Göttingen, Göttingen, Germany

[5]German Center for Cardiovascular Research (DZHK), partner site Lower Saxony, Göttingen, Germany

[6]Department of Neurology, University Medical Center Göttingen, Göttingen, Germany

Corresponding Author:
Carolin Steinmetz, PhD
Department of Geriatrics, University Medical Center Göttingen, Robert-Koch-Straße 40, Göttingen, Germany

Abstract

Background: The Otago program (OP) is evidence-based and focuses on fall prevention in older people. The feasibility and usability of a short-term digital program modeled after the principles of the OP in the setting of early geriatric rehabilitation (EGR) are unclear.

Objective: This study investigated the feasibility and usability of an additional technology-based fall prevention program (FPP) in the setting of EGR.

Methods: We performed a feasibility study in the setting of EGR. A sample of 30 patients (mobility at least by walker; mini-mental status test score >17) was recruited between March and June 2024 and compared with a retrospective cohort (n=30, former EGR patients). All patients in the intervention group (IG) received a supervised, OP-modified FPP thrice/week for 20 minutes using a technology-based platform called “Pixformance.” The device is a digital trainer and enables real-time corrections. The primary end point was the feasibility (given when 80% of the IG participated in 6 trainings within 2 weeks). Secondary outcomes were usability (patients’ and facilitators’ perspective; ≥75%), risk of falls (Berg Balance Scale), mobility (Timed Up and Go Test), functional independence (Functional Independence Measure), and activities of daily living (Barthel Index). Several further exploratory end points were analyzed including anxiety and depression (Four-Item Patient Health Questionnaire; PH-Q4). Data were accessed at entry to EGR and after 2 weeks prior to discharge. To analyze the pre-posttest results, the dependent Student *t* test and the Wilcoxon test were applied. A mixed ANOVA with repeated measurements was used for statistical analyses of time-, group-, and interaction-related changes.

Results: A cohort of 60 patients (mean 80.2, SD 6.1 y; 58% females, 35/60) was analyzed. The main indication for EGR was stroke (9/60, 15%). Patients were recruited into a prospective IG (n=30) and a retrospective control group (n=30). Of the 30 patients in the prospective IG, 11 patients (37%) completed 6 training sessions within 2 weeks. Reasons why participants did not complete 6 training sessions were diagnostic appointments (33%), pain/discomfort (33%), or fatigue (17%). EGR patients rated FPP usability at 84% and facilitators at 65% out of 100%. Pre-posttest analysis of the standard assessments showed a significant interaction in Berg Balance Scale (<.01). In both groups, a significant improvement over time was found in the Timed Up and Go Test (<.01), Barthel Index (<.01), and Functional Independence Measure (<.01). Likewise, in the IG, the PH-Q4 score (.02) improved.

Conclusions: While the technology-based FPP in the EGR setting was generally well-accepted by patients, with high usability ratings, its feasibility was limited. Only 37% of participants completed the required additional training sessions. Further studies should test the technology-based FPP as an integrated part of the EGR complex therapy concept. Our findings suggest potential benefits of incorporating technology-based FPPs in EGR, but further refinement is needed to enhance participation and feasibility.

(JMIR Form Res 2025;9:e66692) doi:10.2196/66692

XSL•FO
RenderX

KEYWORDS

fall prevention; fall prevention program; early geriatric rehabilitation; gerontology; older adult; elder; aging; digital exercise intervention; digital activity; physical exercise; functional capacity; new technology; technology-based; digital intervention; feasibility

Introduction

Hospitalization in Old Age and Its Consequences

In 2050, around 22.4 million people in Germany will be ≥65 years old. It is predicted that the number of hospital stays in this cohort will increase by around 30% over the next 6 years [1].

In the case of acute medical illness with hospitalization, old and multimorbid patients have a high prevalence of hospital-associated disability [2] with impairment of activities of daily living, cognition, malnutrition, and sensory deficits [3]. Functional impairments such as muscle atrophy of the lower extremities and balance disorders following hospitalization are very common (30% - 80%) [4,5]. The consequences of this deterioration include rehospitalization, nursing home admissions [6], an increased number of falls, a poorer quality of life (QoL), increased consumption of health-related resources [7], and an increase in mortality [8]. Preventing hospital-associated disability therefore is crucial because it significantly impacts patient outcomes and overall health care systems.

Early Geriatric Rehabilitation

Early geriatric rehabilitation (EGR) is a specialized multidisciplinary and multiprofessional approach for older patients in Germany who need rehabilitation treatment due to acute health events such as injuries or surgeries that require hospitalization [9]. In general, the duration of an EGR is 2 weeks and comprises 20 training sessions in acute medical care. The main aim of EGR is to improve functional independence as well as QoL so that patients can be released to their homes, an inpatient rehabilitation program, or lower-level care such as long-term nursing care [10].

Fall Prevention Program

Fall prevention programs (FPPs) are often used to optimize functional limitations in older people. The Otago program (OP) is one of these programs; it is evidence-based and was explicitly developed for older people. Long-term studies have shown significant results with regard to the reduction of falls and fall-related injuries [11,12], an increase in strength and balance as well as an increase in confidence in being able to continue with everyday activities [13]. The OP showed the greatest effectiveness in the high-risk groups of over 80-year-olds and people with a history of falls [13,14]. We hypothesized that such prevention programs are not only effective as long-term interventions but can also be effective when the duration of the intervention is short. Previous studies in the area of prehabilitation have shown that older, multimorbid patients benefit from short-term interventions of 5 - 14 days [15,16].

Likewise, Martínez-Velilla et al [17] conducted a randomized, controlled intervention study in 370 older patients (age 87.3, SD 4.9 y) who were hospitalized for acute treatment. The intervention group (IG) included individualized moderate strength training, balance exercises, and gait training. The control group (CG) received standard treatment and interventions when needed. The training was safe and 2 sessions daily were sufficient to significantly improve mobility measured by short physical performance battery (SPPB) and the Barthel Index (BI) during an average hospital stay of 5 days, whereas the CG showed a deterioration in mobility [17]. In particular, assessments for measuring a patient's mobility (eg, Timed Up and Go Test [TUG] or SPPB) have the highest sensitivity in the EGR setting compared with other methods and are suitable to determine the effects of an FPP [17,18].

Digital Interventions

The number of digital interventions in geriatrics is increasing and may help to improve outcomes such as physical activity or balance. A systematic review of eHealth applications in geriatric rehabilitation showed that digital interventions are promising and have the potential to improve rehabilitation outcomes. In 16 out of 17 studies, eHealth applications were feasible in this setting [19].

The technology-based device "Pixformance" is a digital trainer that uses an integrated 2D full high-definition camera to scan 25 body points, enabling real-time corrections. The screen is divided into 2 parts, one part acts like a mirror and the other shows a person who is demonstrating the exercise. It has been shown that videos that show real people in action are effective in achieving long-term behavioral change [20].

However, acceptance of digital interventions in older persons is often limited by factors such as unfamiliarity with technology, cognitive decline, physical impairments, and concerns about privacy or security. On the other hand, there might be several challenges when it comes to implementing digital interventions for older adults from a facilitators' perspective. Facilitators are individuals who assist patients in initiating the digital FPP and remain present while the patient interacts with the device. This study is the first to use "Pixformance" in the EGR setting as part of an evidence-based FPP to investigate feasibility and usability.

The purpose of this study is to evaluate the feasibility and usability (patients' and facilitators' perspectives) of a short-term technology-based FPP in the setting of EGR, performed in addition to the standard complex therapy concept. Furthermore, the effect of the FPP on the risk of falls, functional independence, activities of daily living, mobility, and cognitive performance compared with a retrospective CG was analyzed in order to determine which assessments can be used to sensitively detect changes due to an additional FPP. This analysis will be used to make informed decisions about the study design and instruments used in a follow-up project. The pre-posttest of muscle status, QoL, anxiety and depression as well as frailty status, which are assessed only in the IG, will also provide information about the sensitivity of these instruments for detecting change in this older patient cohort.

XSL•FO
RenderX

Methods

Patient Population

Patient recruitment was carried out in the Department of Geriatrics at the University Medical Center Göttingen, Germany, between March and June 2024.

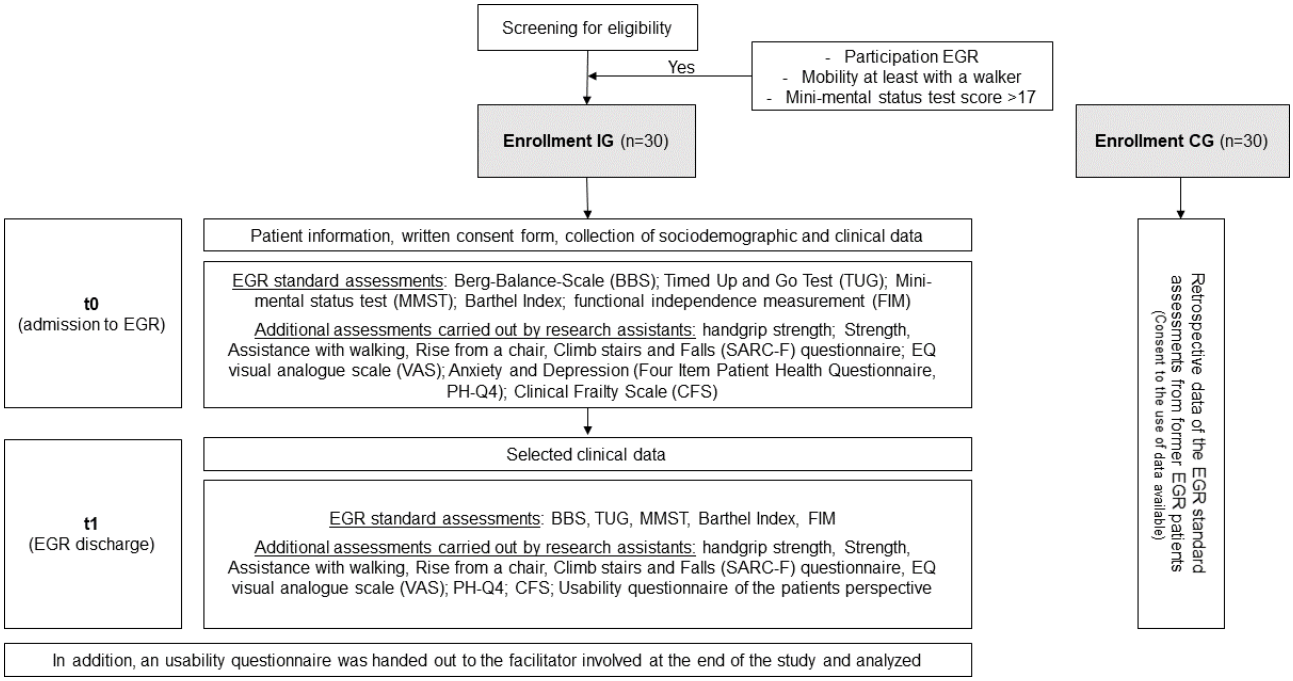
All patients undergoing an EGR complex therapy, at least able to walk with a walker, and cognitively fit to slightly cognitively impaired (mini mental status test score >17) were eligible for inclusion.

Patients were excluded if they were unable to understand the study information and give written consent due to poor German language skills, or had cognitive or visual impairments.

Study Setting

This is a prospective 2-arm feasibility study in the setting of EGR. The IG was recruited prospectively and the CG retrospectively. Eligible patients for the IG were asked to participate in the feasibility study by physicians at the beginning of EGR during regular clinical consultations. All patients were provided study information and gave their written informed consent. Subsequently, the baseline assessment and follow-up assessment took place (Figure 1).

Figure 1. Research design and conducted assessments. CG: control group; EGR: early geriatric rehabilitation; IG: intervention group.



A retrospective CG consists of former patients who have undergone EGR complex therapy at the same department in the last 3 months prior to the recruitment of the prospective group and fulfill the inclusion criteria specified in this feasibility study. All included patients gave written consent for the use of their pseudonymized data for scientific purposes upon admission to the University Medical Center Göttingen. All CG patients were screened and included in the CG according to their order of admission.

Sample Size

A sample size of 60 patients (IG: n=30; CG: n=30) has 80% power to detect a standardized difference in means of 0.74 with 80% power, using a 2-group *t* test with a 5% 2-sided significance level. This corresponds to comparably large effects, which are likely to be clinically relevant. The sample size was seen as feasible, taking into account the number of potentially eligible patients undergoing EGR within 3 months, as well as the inclusion and exclusion criteria. The study design is comparable to a recently conducted pilot study by Fränzel et al [21] which includes a FPP with 58 participants (IG: n=29; CG: n=29) in the same setting. This study found significant developments in patient mobility measured by TUG, SPPB, and gait speed [21].

Procedure

All patients of the IG received the regular EGR complex therapy concept, which consists of 20 sessions, including regular physiotherapy, occupational and speech therapy as well as psychological support. In addition, 6 sessions of a supervised FPP modified according to the OP [22] using the technology-based platform called “Pixformance” (Figure 2) were offered to IG participants 3 times per week.

XSL•FO
RenderX

Figure 2. The pixformance station (picture provided by pixformance sports GmbH).

The content of the 2-week training program is summarized in Table 1.

XSL•FO
RenderX

Table . Overview of the content of the fall prevention program.

Category	Details
Frequency	Thrice/week; 10 repetitions, 1 set
Intensity	Body weight, Rating of Perceived Exertion scale 13 - 15 (medium)
Time	20 min
Type	mobilization, coordination, strengthening exercises focusing on the lower extremities, relaxation/stretching
Exercises	Examples include shoulder and torso rotation, calf raises, lateral leg lift, standing hamstring curl, foot lift and grinding heel in sitting position, and standing up and sitting down from a chair

Before being discharged from the EGR, all participants in the IG received an Otago booklet to continue the exercises at home.

Outcomes

The primary end point was the feasibility of the technology-based FPP in the setting of EGR. We defined feasibility as at least 80% of the patients completing 6 training sessions within 2 weeks in addition to the regular 20 therapeutic sessions in the EGR. For the statistical analysis, we report how many participants completed the training as well as the reasons for not completing the training in absolute numbers and percentages.

User Experience

Usability was assessed from the patients' and facilitators' perspectives. Facilitators did not provide therapy directly; instead, they addressed technical issues, such as starting, operating, or stopping the device. All training instructions and feedback were delivered by the device itself, not by the facilitators. In this study, facilitators included both geriatric staff (2 physical therapists and 2 occupational therapists) and study personnel (3 sport science students and 1 medical student), all of whom received training in device operation. All facilitators provided written informed consent.

At the end of the study, all patients and facilitators filled out a usability questionnaire. The results of the structured questionnaire were quantitatively analyzed. The device was considered to be usable if the "Overall rating for the training with the device" and "Overall rating for using the device" was ≥75%, modified according to Rabinovich et al [23]. For the statistical analysis, we report the results of the usability questionnaire as the percent of the 2 highest values on the 5-point Likert scale.

Secondary Outcomes

In addition to feasibility and user experience data, further end points are detailed in Figure 1. Other secondary outcomes included the risk of falls assessed by Berg Balance Scale [24], mobility assessed with the TUG [25], functional independence carried out with the help of the Functional Independence Measure (FIM) [26], and activity of daily living collected by BI [27,28].

Data were assessed at the start of the EGR and after 2 weeks. How the additional technology-based FPP influences the effect of EGR complex therapy was examined using standard assessment data from the IG and a retrospective CG.

Statistical Analysis

We used the Kolmogorov–Smirnov test for the assessment of normal distributions. Continuous and categorical variables are presented by mean (SD) with absolute and relative frequencies, respectively. Two-group comparisons of baseline variables were performed using Student's *t* test and chi-square test of independence for continuous and categorical variables. To analyze the results of the pre-posttest, dependent Student *t* test and the Wilcoxon signed rank test were applied. A 2-way mixed ANOVA with repeated measurements was used for statistical analyses of time- (t0/t1), group- (IG or CG), and interaction-related (between time and group) changes. In all analyses, a $P<.05$ was considered statistically significant. The results of the usability questionnaire include the percent of the 2 highest values on the 5-point Likert scale. Analyses were performed using the IBM Statistical Package for the Social Sciences (SPSS version 29.0; IBM Co).

Ethical Considerations

This feasibility study was approved by the local ethics committee of the University Medical Center Göttingen (application number: 22/1/24; January 23, 2024) and conforms to the Declaration of Helsinki. All participants signed informed consent forms.

Results

Study Population

A total of 70 patients were screened for eligibility for the prospective IG. Of these, 40 were excluded because of "no interest" (n=13), "poor mobility" (n=10), "hospital stay too short" (n=5), "low cognition" (n=4), "low cognition and poor mobility" (n=3), "visual impairment" (n=2), and "others" (n=3). In the end, 30 patients from the IG participated in the FPP.

A sample of 60 patients (mean age 80.2 years, SD 6.1 years; 58% females, 35/60; IG: n=30; CG: n=30) were included in the analysis. Patients of the prospective IG were consecutively recruited (n=30; mean age 79.0 years, SD 6.0 years; 53% female, 16/30) at admission to EGR. The retrospective CG (mean age 85.0 years, SD 4.6 years; 63% female, 19/30) is made up of former EGR patients who were treated in the 3 months prior to the "Pixformance" intervention and fulfilled all inclusion and exclusion criteria.

The baseline characteristics of the IG showed the cohort to be significantly younger and taller than the CG. With regards to

XSL•FO
RenderX

clinical and medical history characteristics, patients in the IG were more likely to be taking a higher number of medications and had a degree of disability, a post-Covid condition, and depressive symptoms in comparison to the retrospective CG (Table 2).

XSL•FO
RenderX

Table . Baseline characteristics of the cohort.

Characteristics	All (n=60)	IG[a] (n=30)	CG[b] (n=30)	*P* value
Age (years), mean (SD)	82.0 (6.1)	79.0 (6.0)	85.0 (4.6)	<.01[c]
Sex, n (%)				
Male	24 (40)	14 (47)	10 (33)	.34[d]
Female	35 (58)	16 (53)	19 (63)	.34[d]
Unknown	1 (2)	0 (0)	1 (2)	.31[d]
Height (cm), mean (SD)	166.4 (8.4)	168.6 (8.0)	164.1 (8.4)	.04[c]
Weight (kg), mean (SD)	71.4 (16.6)	73.8 (18.3)	68.9 (14.6)	.26[c]
BMI (kg/m²), mean (SD)	25.8 (5.2)	26.1 (5.9)	25.5 (4.5)	.67[c]
Blood pressure (mmHg[e]), mean (SD)				
Systolic	N/A[f]	131.7 (23.1)	N/A	N/A
Diastolic	N/A	73.1 (12.0)	N/A	N/A
Heart rate (beats/min), mean (SD)	N/A	71.4 (11.3)	N/A	N/A
Number of medications (regularly), mean (SD)	7.6 (4.1)	9.1 (3.9)	6.1 (3.7)	.01[c]
Need for nursing assistance, n (%)	30 (50)	16 (53)	14 (47)	.87[d]
Any degree of disability, n (%)	9 (15)	8 (27)	1 (3)	<.01[d]
Situation of living, n (%)				
Living alone	N/A	11 (18)	N/A	N/A
Visual aid, n (%)	41(68)	24 (80)	17 (57)	.07[d]
Hearing aid device, n (%)	16 (27)	7 (23)	9 (30)	.08[d]
Walking aid, n (%)	34 (57)	15 (50)	19 (63)	.29[d]
Walker	26 (43)	12 (40)	14 (47)	.60[d]
Cane	4 (7)	2 (7)	2 (7)	1.0[d]
Forearm crutch	5 (8)	1 (3)	4 (13)	.16[d]
Main diagnosis for EGR[g] admission, n (%)				
Poststroke or suspected stroke	9 (15)	5 (17)	4 (13)	.72[d]
Cardiac decompensation	8 (13)	4 (13)	4 (13)	1.0[d]
Acute kidney failure	4 (7)	2 (7)	2 (7)	>.99[d]
Vertigo including syncope	4 (7)	3 (10)	1 (3)	.30[d]
Concomitant diseases, n (%)				
Diabetes mellitus	23 (38)	12 (40)	11 (37)	.12[d]
Hypertension	35 (58)	16 (53)	19 (63)	.06[d]
Dyslipidemia	15 (25)	6 (20)	9 (30)	.39[d]
Stroke or TIA[h]	12 (20)	6 (20)	6 (20)	.37[d]

XSL•FO
RenderX

Characteristics	All (n=60)	IG[a] (n=30)	CG[b] (n=30)	*P* value
Dementia	7 (12)	3 (10)	4 (13)	.41[d]
Kidney failure	13 (42)	5 (17)	8 (27)	.19[d]
Cancer disease<5 years	5 (8)	3 (10)	2 (7)	.54[d]
Cancer disease ≥5 years	5 (8)	2 (7)	3 (10)	.50[d]
Thyroid disorder	8 (13)	6 (20)	2 (7)	.21[d]
Rheumatism	10 (17)	8 (27)	2 (7)	.08[d]
Condition after COVID-19 infection	20 (33)	17 (57)	3 (10)	<.01[d]
Depression	5 (8)	1 (3)	4 (13)	.04[d]
Anxiety	5 (8)	3 (10)	2 (7)	.19[d]

[a]IG: intervention group.
[b]CG: control group.
[c]Independent *t* test.
[d]Chi-square test.
[e]mmHg: millimeters of mercury.
[f]N/A: not available.
[g]EGR: early geriatric rehabilitation.
[h]TIA: transient ischemic attack.

Feasibility of the FPP

Out of a total of 30 patients in the IG, 11 (37%) participated in 6 trainings within 2 weeks. Of these 11 patients, 7 (64%) completed the FPP including all strengthening/balance exercises. The remaining 4 patients ended their training session early or skipped one or more important strengthening or balance exercises.

The main reasons why participants did not complete the amount of training were the inability to attend the FPP due to diagnostic appointments (33%), pain/discomfort (33%), or fatigue (17%).

User Experience

The analysis of user experiences from the patients' and facilitators' perspectives is summarized in Table 3. The 8 facilitators were on average 35.1 (SD 12.3) years old. A quarter of the facilitators were male (2/8, 25%) and 3 quarters were female (6/8, 75%). The group members had an average work experience of 20.0 (SD 8.8) years.

XSL•FO
RenderX

Table . Overview of the results of the user experience questionnaire from the perspective of patients and facilitators.

Usability items	Rating	Patients' perspective (n=26), (%)	Facilitators' perspective (n=8), (%)
Instructions by facilitators for using the device were clear	Agreeing	96	—[a]
There were technical problems	Never	85	—
Training with the device was comfortable	Often	92	—
Training with the device was embarrassing	Never	100	—
Corrections during the training by the device were clear	Agreeing	89	—
Willingness to continue the learned exercises at home	≥3 months	54	—
Overall patient usability rating (%)(100%=verygood experience), mean (SD)	—	84.0 (15.90)	—
Instructions by the manufacturer for using the device were clear	Agreeing	—	50
Difficulties when commissioning the device	No starting problems	—	63
Regular use of the device was easy	Agreeing	—	50
The exercise selection of the device was easy to handle	Agreeing	—	63
Selected exercises could be easily adapted during the training	Agreeing	—	25
The device was easy to switch on and off	That worked very well	—	75
I am in favor of the future use of the device	Yes, very	—	50
Overall facilitator usability rating (%) (100%=verygood experience), mean (SD)	—	—	65.4 (20.0)

[a]Not applicable.

The overall usability rating for the device from the patients' perspective was 84.0% (SD 15.9%), where 0% means a bad experience and 100% a very good experience. The facilitators rated the usability of the device at 65.4% (SD 20.0%).

Secondary Outcomes

The results of the EGR standard assessments are indicated in Table 4. For 22 patients in the IG and 29 patients in the CG, complete measurements at both time periods were recorded. The 2-way mixed ANOVA showed a significant interaction between groups over time in the Berg Balance Scale (<.01). Likewise, a significant positive development over time was seen in both groups in TUG time (<.01), FIM motor subscale (<.01) score, and BI (<.01).

XSL•FO
RenderX

Table . Descriptive and 2-way mixed ANOVA results for the early geriatric rehabilitation routine assessments at baseline (t0, admission) and after 2 weeks (t1, discharge) differentiated by group.

Parameters	CG[a], mean (SD)	IG[b], mean (SD)	*P* value		
			Time	Group	Interaction
Berg Balance Scale (score)			.54	.02	<.01
t0[c]	31.3 (12.5)	33.6 (10.0)			
t1[d]	38.4 (8.7)	24.5 (7.5)			
Timed Up and Go Test (s)			<.01	.40	.82
t0	22.1 (9.0)	20.2 (7.6)			
t1	17.5 (7.5)	16.1 (6.7)			
Minimental status test (score)			.07	.21	.21
t0	24.2 (3.3)	24.6 (4.3)			
t1	24.7 (3.0)	27.2 (3.9)			
Barthel Index			<.01	.68	.78
t0	61.6 (17.5)	63.7 (18.2)			
t1	76.2 (12.9)	76.8 (15.9)			
FIM[e] motor subscale (score)			<.01	.35	.64
t0	53.8 (12.2)	51.0 (10.6)			
t1	67.0 (10.9)	65.3 (6.6)			

[a]CG: control group.

[b]IG: intervention group.

[c]t0: baseline (early geriatric rehabilitation admission).

[d]t1: after 2 weeks (early geriatric rehabilitation discharge).

[e]FIM: Functional Independence Measure.

The Wilcoxon signed rank test confirmed the significant development of the mixed ANOVA in Berg Balance Scale (BBS) in IG (.01; $n_{paired\_results}$ =22) and CG (<.01; $n_{paired\_results}$ =29). In the IG, 5 patients improved and 17 showed poorer results in BBS between baseline and EGR discharge, whereas in the CG, 17 participants improved, 2 showed poorer results and 10 did not change in BBS.

Table 5 summarizes the results of muscle status, QoL, anxiety and depression, and frailty status, which were additionally assessed by study personnel in the IG. A significant improvement in Four-Item Patient Health Questionnaire (PH-Q4) score (.024) from baseline to EGR discharge was observed. Furthermore, the frailty status and the positive screen for sarcopenia each decreased by 27%.

XSL•FO
RenderX

Table . Pre-posttest and descriptive results for the assessments at baseline (t0) and after 2 weeks of early geriatric rehabilitation (t1) of the intervention group (IG).

Parameters	IG		*P* value
	t0 (n=30)	t1 (n=26)	
Handgrip strength (kg), mean (SD)	24.9 (7.6)	24.7 (7.3)	.93[a]
Female	26.2 (6.4)	20.0 (6.2)	.40[a]
Male	23.9 (8.6)	28.4 (6.1)	.18[a]
CFS[b] (score), mean (SD)	4.3 (1.5)	3.8 (1.6)	.27[a]
Frailty (CFS ≥4), n (%)	21 (70)	13 (43)	>.99[c]
SARC-F[d] score, mean (SD)	4.0 (2.4)	3.4 (1.9)	.28[a]
Sarcopenia (positive screen SARC-F ≥4), n (%)	18 (60)	10 (33)	.13[c]
EQ VAS[e], mean (SD)	52.1 (15.4)	60.4 (13.0)	.06[a]
PH-Q4[f], mean (SD)	3.3 (2.4)	1.6 (1.9)	.02[a]

[a]Dependent *t* test.

[b]CFS: Clinical Frailty Scale.

[c]Chi-square test.

[d]SARC-F: Strength, Assistance with walking, Rise from a chair, Climb stairs and Falls questionnaire.

[e]EQ VAS: European Quality of Life Visual Analogue Scale.

[f]PH-Q4: Four-Item Patient Health Questionnaire.

Discussion

Summary of Target and Findings

In this study, we aimed to evaluate the feasibility and usability of a short-term technology-based FPP in the EGR setting from the patients' and facilitators' perspectives. The FPP was conducted in addition to the conventional complex therapy concept. Furthermore, the effect on the risk of falls, daily function, mobility, and cognitive performance compared with a retrospective CG was analyzed. In the IG, the pre-posttest of muscle status, QoL, anxiety and depression as well as frailty status were additionally examined. Regarding the primary end point, we observed the limited feasibility of the FPP in the EGR setting, while usability and acceptance from the patients' perspective were proven. The results of the BBS showed a significant decrease in the IG and a significant improvement in the CG. Further significant improvements from EGR admission to discharge were observed in TUG, FIM, BI, and PH-Q4.

This study is the first to use the "Pixformance" device in the EGR setting as part of an evidence-based FPP.

Study Cohort

The number of older patients receiving complex therapy in EGR in Germany (211,270 participants in 2007 to 274,926 in 2020) is increasing steadily due to demographic change [29]. The baseline characteristic of our cohort is similar to others who performed studies in the EGR setting [21,30,31]. The other studies reported an average age of 78 to 87 years, the admission of more women than men, and strokes, cardiac decompensation, and hip fractures as the most common admission diagnoses [21,30,31]. In our cohort, significant differences between the groups were observed in age, height, number of regularly taken medications, degree of disability, condition after COVID-19 infection, and depression. Our CG is an unmatched-pairing retrospective comparison cohort of former patients from the time point directly before the intervention. The patients were consecutively included in the CG, beginning at the start of the intervention and going backward in time until 30 patients were identified. In this feasibility study, a CG was included in order to determine which assessments might be used to sensitively detect changes due to an additional FPP in order to design a follow-up project.

Participants of the IG were younger and had more often severe concomitant diseases. The high number of patients after COVID-19 infection in the IG may explain why only 11 out of 30 patients completed 6 training sessions within 2 weeks. Fatigue is one of the most common reasons why the FPP was not completed. Especially older adults aged 65 years or older are at greater risk of persisting symptoms associated with COVID-19; fatigue is 1 of the 5 common long-term symptoms [32].

eHealth in Geriatrics

Kraaijkamp et al [19] published a systematic review of eHealth in the setting of geriatric rehabilitation. eHealth is defined as "the use of digital information and communication to support and/or improve health and healthcare" [19,33]. The authors summarized that eHealth is often feasible and can potentially improve geriatric rehabilitation outcomes. Simple eHealth interventions are more likely to be feasible for older patients receiving geriatric rehabilitation [19]. "Pixformance" can be seen as such a "simple" eHealth intervention. Likewise, the

XSL•FO
RenderX

authors concluded that there is a lack of evidence on usability which might hamper the implementation [19].

Feasibility of the FPP

In our study, feasibility was defined as participation in 6 FPP training sessions throughout the EGR. The most common reason for missing a training session was due to medical examinations outside the ward. Sometimes, there were delays in the morning so patients could not participate in the FPP. The clinical routine always took priority. The low participation rate was therefore due to logistical problems that the study team was unable to solve. An additional training session in the EGR setting therefore seems to be too much. Similarly, Fränzen et al [21] conducted a pilot study including a FPP in the form of a square-stepping exercise (SSE) in the EGR setting and investigated its feasibility. The required minimum number of SSE units, each lasting 30 minutes, was 6 during the EGR stay. The amount of exercise units is in line with our target and definition of feasibility. The SSE program proved to be feasible. The main difference between the 2 interventions was that the SSE was performed as a part of conventional physiotherapy included in the complex therapy concept [21]. In our feasibility study, the FPP took place in the afternoons in addition to the complex therapy concept. In a future study, we will include the study intervention as part of conventional physiotherapy in the complex therapy concept so that logistical problems do not affect the intervention.

User Experience

We have investigated the usability of the technology-based FPP and followed the advice from Kraaijkamp et al [19] to pave the way for a possible implementation [19]. User experience was assessed independently from the facilitators' and patients' perspectives. Usability from the patients' perspective was achieved, while the facilitators' did not reach the target of ≥75%. The results are surprising in that eHealth is often not sufficiently tailored to age-related barriers such as cognition, physical ability, perception, and motivation [34,35]. "Pixformance" as a "simple" eHealth intervention seems to be attractive to patients. Similarly, a usability study of technology-based exercise games in older patients following hip replacement surgery confirmed that the cohort enjoyed performing exercises under digital guidance [36]. In contrast to this, usability issues were pointed out by facilitators.' Only every second respondent was in favor of using the device in the future in the EGR setting. The item "selected exercises could be easily adapted during the training" was only agreed on by 25% of the facilitators. The low rating could be because the exercises had to be changed via a separate computer platform. Facilitators' explained that there was not enough time to learn how to adapt the exercises and navigate the training platform in detail. This may be due to the nature of the study, which was conducted in addition to the staff's normal daily activities [37].

Secondary Outcomes

The BBS is widely used in clinical settings to assess the risk of falls, particularly in older individuals with balance impairments or in geriatric wards within general hospitals [38,39]. A score of less than 45 on the BBS is considered the threshold for increased fall risk. A change of 7 points is usually considered clinically relevant. When the score lies between 35 and 44 points, an increased risk of falls is indicated [40]. At baseline, the mean BBS scores of our cohort ranged from 31 to 34 points, indicating an increased risk of falls. Interestingly, the results of the BBS showed a significant decrease in the IG (−9.1 points) while the CG showed a significant improvement (+7.1). Even so, both groups remained in the range of increased fall risk. One reason for this unexpected outcome could be that the BBS may not be the most appropriate assessment tool to capture changes resulting from a FPP. The BBS primarily includes tasks that were not trained in the FPP, such as "Retrieving object from floor" or "turning to look behind." For a follow-up project, the SPPB appears to be a more suitable assessment tool. As mentioned in the introduction, measuring a patient's mobility using tools like the SPPB has shown the highest sensitivity in EGR settings compared with other methods [17,18]. The SPPB consists of 3 parts: "gait speed test," "chair rising test," and "balance test." These tests are well suited for measuring changes resulting from the FPP [41]. Moreover, the SPPB is frequently used in studies conducted in cohorts similar to ours [21,42]. It should be noted, however, that there was some missing data in the IG, resulting in only 22 paired results being included in the analysis. Therefore, these findings should be interpreted with caution.

Further significant positive developments (ie, the patients' condition improved after 2 weeks) based on the multimodal and interdisciplinary EGR complex therapy concept [9,10] were found in functional independence (FIM), activity of daily living (BI), mobility (TUG), and anxiety and depression (PH-Q4). These instruments, therefore, seem appropriate for measuring differences due to digital FPP in further studies.

Limitations

This feasibility study has a relatively small cohort and includes an unmatched-pairing retrospective comparison cohort. Further studies with larger cohorts and a randomized controlled design with blinded assessors are needed to validate the results presented. Key factors such as disability (yes or no), sex (male or female), and age (younger than 81 years vs 81 years and older) should be considered in the 1:1 block randomization process with random block lengths to ensure a homogeneous cohort at baseline. Additionally, the training intensity of the FPP might be increased by incorporating a digital balance pad or components of the SSE unit intervention [21] to improve outcomes.

Conclusion

Due to the high percentages of frail and sarcopenic patients admitted to EGR with an elevated fall risk, there is a significant unmet need for a FPP. However, the feasibility of the technology-based FPP was not achieved, as only 37% of participants completed 6 additional FPP sessions within 2 weeks, primarily due to diagnostic appointments, pain/discomfort, or fatigue. Usability analysis indicates that EGR patients accept the technology-based FPP, while staff members are less convinced. Significant positive changes from EGR entry to discharge were observed in TUG, FIM motor subscale, and BI, independent of the FPP. A statistically significant, but not clinically relevant improvement in the BBS was noted in the

XSL•FO
RenderX

CG, while the IG showed a significant positive change in the PHQ-4 score, indicating fewer symptoms of anxiety and depression. Further studies should explore the technology-based FPP as an integrated part of the EGR complex therapy concept, increase the FPP intensity by incorporating a digital balance pad or components of the SSE, and assess progress from admission to discharge using the SPPB instead of the BBS.

Acknowledgments

We want to thank all patients as well as the staff who were involved in this study. Our special thanks go to Niklas Meyer and Mira Fleur Seitz, who supported the team during training sessions and assessments. We also thank Pixformance Sports GmbH, which generously loaned the “Pixformance” device for 3 months during the feasibility study. This cohort study was financially supported by the University Medical Center Göttingen, Department of Geriatrics.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

CAFvA is the principal investigator. CStei and SH coordinated the study as well as prepared the ethics documents. The protocol was developed by CStei, SH, MSch, MV, CAFvA, and MSad. The content of the modified Otago program was created by CStei, CSten, MSyl, and DG. MSch und MV screened patients for the inclusion and exclusion criteria and informed patients about the study. CSten and MSyl conducted the assessments and served as facilitators during the exercise training sessions. CStei, SH, CSten, and MSyl were responsible for study protocol implementation. CStei was responsible for statistical analysis and writing. CStei, SH, MSch, MV, CAFvA, MSad, and DG contributed to the data interpretation. CStei and SH prepared the initial draft of the manuscript. All authors reviewed and commented upon the initial draft and approved the final manuscript.

Conflicts of Interest

CAFvA received honoraria from serving on the scientific advisory board of Biogen, Roche, Novo Nordisk, and Dr Willmar Schwabe GmbH & Co. KG and has received funding for travel as well as speaker honoraria from Biogen, Roche Diagnostics AG, Medical Tribune Verlagsgesellschaft mbH, Landesvereinigung für Gesundheit und Akademie für Sozialmedizin Niedersachsen e. V., and Dr Willmar Schwabe GmbH & Co. KG. CAFvA has received research support from Roche Diagnostics AG and research funding from the Innovationsfond (Fund of the Federal Joint Committee, Gemeinsamer Bundesausschuss, G-BA grant numbers VF1_2016-201; 01NVF21010; 01VSF21019). The remaining authors have nothing to disclose.

References

1. bevölkerungsvorausberechnung für deutschland [population forecast for Germany]. Statistisches bundesamt [Federal Statistical Office]. URL: https://service.destatis.de/bevoelkerungspyramide/#!y=2050&a=20,65&g [accessed 2024-08-26]
2. Skains RM, Zhang Y, Osborne JD, et al. Hospital-associated disability due to avoidable hospitalizations among older adults. J Am Geriatr Soc 2023 May;71(5):1395-1405. [doi: 10.1111/jgs.18238] [Medline: 36661192]
3. Buurman BM, Hoogerduijn JG, de Haan RJ, et al. Geriatric conditions in acutely hospitalized older patients: prevalence and one-year survival and functional decline. PLoS ONE 2011;6(11):e26951. [doi: 10.1371/journal.pone.0026951] [Medline: 22110598]
4. Zisberg A, Shadmi E, Sinoff G, Gur-Yaish N, Srulovici E, Admi H. Low mobility during hospitalization and functional decline in older adults. J Am Geriatr Soc 2011 Feb;59(2):266-273. [doi: 10.1111/j.1532-5415.2010.03276.x] [Medline: 21314647]
5. Martínez-Velilla N, Cadore EL, Casas-Herrero Á, Idoate-Saralegui F, Izquierdo M. Physical activity and early rehabilitation in hospitalized elderly medical patients: systematic review of randomized clinical trials. J Nutr Health Aging 2016;20(7):738-751. [doi: 10.1007/s12603-016-0683-4] [Medline: 27499308]
6. Fortinsky RH, Covinsky KE, Palmer RM, Landefeld CS. Effects of functional status changes before and during hospitalization on nursing home admission of older adults. J Gerontol A Biol Sci Med Sci 1999 Oct;54(10):M521-M526. [doi: 10.1093/gerona/54.10.m521] [Medline: 10568535]
7. Covinsky KE, King JT Jr, Quinn LM, et al. Do acute care for elders units increase hospital costs? a cost analysis using the hospital perspective. J Am Geriatr Soc 1997 Jun;45(6):729-734. [doi: 10.1111/j.1532-5415.1997.tb01478.x] [Medline: 9180668]
8. Brown CJ, Friedkin RJ, Inouye SK. Prevalence and outcomes of low mobility in hospitalized older patients. J Am Geriatr Soc 2004 Aug;52(8):1263-1270. [doi: 10.1111/j.1532-5415.2004.52354.x] [Medline: 15271112]
9. Stier-Jarmer M, Pientka L, Stucki G. Frührehabilitation in der geriatrie [early geriatric rehabilitation]. Phys Rehab Kur Med 2002 Aug;12(4):190-202. [doi: 10.1055/s-2002-33901]

XSL•FO
RenderX

10. Palzer M, Meyer U, Abderhalden LA, Gazzotti A, Hierholzer C, Bischoff-Ferrari HA. Geriatrische komplexbehandlung bei alterstraumatologischen patienten: einfluss auf mobilität, selbsthilfefähigkeit und austrittsdestination. [acute geriatric treatment of older trauma patients: influence on mobility, autonomy and postdischarge destination]. Z Gerontol Geriatr 2021;54:816-822. [doi: 10.1007/s00391-020-01812-4] [Medline: 33201306]
11. Campbell AJ, Robertson MC, Gardner MM, Norton RN, Tilyard MW, Buchner DM. Randomised controlled trial of a general practice programme of home based exercise to prevent falls in elderly women. BMJ 1997 Oct 25;315(7115):1065-1069. [doi: 10.1136/bmj.315.7115.1065] [Medline: 9366737]
12. Campbell AJ, Robertson MC, Gardner MM, Norton RN, Buchner DM. Falls prevention over 2 years: a randomized controlled trial in women 80 years and older. Age Ageing 1999 Oct;28(6):513-518. [doi: 10.1093/ageing/28.6.513] [Medline: 10604501]
13. Robertson MC, Campbell AJ, Gardner MM, Devlin N. Preventing injuries in older people by preventing falls: a meta-analysis of individual-level data. J Am Geriatr Soc 2002 May;50(5):905-911. [doi: 10.1046/j.1532-5415.2002.50218.x] [Medline: 12028179]
14. Robertson MC, Devlin N, Gardner MM, Campbell AJ. Effectiveness and economic evaluation of a nurse delivered home exercise programme to prevent falls. 1: randomised controlled trial. BMJ 2001 Mar 24;322(7288):697-701. [doi: 10.1136/bmj.322.7288.697] [Medline: 11264206]
15. Herdy AH, Marcchi PLB, Vila A, et al. Pre- and postoperative cardiopulmonary rehabilitation in hospitalized patients undergoing coronary artery bypass surgery: a randomized controlled trial. Am J Phys Med Rehabil 2008 Sep;87(9):714-719. [doi: 10.1097/PHM.0b013e3181839152] [Medline: 18716482]
16. Steinmetz C, Bjarnason-Wehrens B, Baumgarten H, Walther T, Mengden T, Walther C. Prehabilitation in patients awaiting elective coronary artery bypass graft surgery - effects on functional capacity and quality of life: a randomized controlled trial. Clin Rehabil 2020 Oct;34(10):1256-1267. [doi: 10.1177/0269215520933950] [Medline: 32546065]
17. Martínez-Velilla N, Casas-Herrero A, Zambom-Ferraresi F, et al. Effect of exercise intervention on functional decline in very elderly patients during acute hospitalization: a randomized clinical trial. JAMA Intern Med 2019 Jan 1;179(1):28-36. [doi: 10.1001/jamainternmed.2018.4869] [Medline: 30419096]
18. Heldmann P, Werner C, Belala N, Bauer JM, Hauer K. Early inpatient rehabilitation for acutely hospitalized older patients: a systematic review of outcome measures. BMC Geriatr 2019 Jul 9;19(1):189. [doi: 10.1186/s12877-019-1201-4] [Medline: 31288750]
19. Kraaijkamp JJM, van Dam van Isselt EF, Persoon A, Versluis A, Chavannes NH, Achterberg WP. eHealth in geriatric rehabilitation: systematic review of effectiveness, feasibility, and usability. J Med Internet Res 2021 Aug 19;23(8):e24015. [doi: 10.2196/24015] [Medline: 34420918]
20. Abu Abed M, Himmel W, Vormfelde S, Koschack J. Video-assisted patient education to modify behavior: a systematic review. Patient Educ Couns 2014 Oct;97(1):16-22. [doi: 10.1016/j.pec.2014.06.015] [Medline: 25043785]
21. Fränzel K, Koschate J, Freiberger E, Shigematsu R, Zieschang T, Tietgen S. Square-stepping exercise in older inpatients in early geriatric rehabilitation. a randomized controlled pilot study. BMC Geriatr 2024 Apr 10;24(1):326. [doi: 10.1186/s12877-024-04932-3] [Medline: 38600478]
22. Campbell AJ, Robertson MC. Otago exercise programme to prevent falls in older adults. Live Stronger for Longer. URL: https://www.livestronger.org.nz/assets/Uploads/acc1162-otago-exercise-manual.pdf [accessed 2024-08-22]
23. Rabinovich RA, Louvaris Z, Raste Y, et al. Validity of physical activity monitors during daily life in patients with COPD. Eur Respir J 2013 Nov;42(5):1205-1215. [doi: 10.1183/09031936.00134312] [Medline: 23397303]
24. Scherfer E, Bohls C, Freiberger E, Heise KF, Hogan D. Berg-Balance-Scale - deutsche version. Physioscience 2006 Jun;2(2):59-66. [doi: 10.1055/s-2006-926833]
25. Podsiadlo D, Richardson S. The timed “Up & Go”: a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991 Feb;39(2):142-148. [doi: 10.1111/j.1532-5415.1991.tb01616.x] [Medline: 1991946]
26. Lüthi H. Alltagsfähigkeiten zuverlässig messen [reliably measuring everyday skills]. Thieme. URL: https://www.thieme-connect.de/products/ejournals/pdf/10.1055/s-0032-1307959.pdf [accessed 2024-08-27]
27. Heuschmann PU, Kolominsky-Rabas PL, Nolte CH, et al. Untersuchung der reliabilität der deutschen version des Barthel-Index sowie entwicklung einer postalischen und telefonischen fassung für den einsatz bei schlaganfall-patienten [investigation of the reliability of the German version of the Barthel Index and development of a postal and telephone version for use with stroke patients]. Fortschr Neurol Psychiatr 2005 Feb;73(2):74-82. [doi: 10.1055/s-2004-830172]
28. Mahoney FI, Barthel DW. Functional evaluation: the Barthel index. Md State Med J 1965 Feb;14:61-65. [Medline: 14258950]
29. Kolb G, Breuninger K, Gronemeyer S, et al. 10 Jahre geriatrische frührehabilitative komplexbehandlung im DRG-system [10 years of early geriatric rehabilitation complex therapy in the DRG system]. Z Gerontol Geriat 2014 Jan;47(1):6-12. [doi: 10.1007/s00391-013-0592-7]
30. Aizen E, Lutsyk G, Wainer L, Carmeli S. Effectiveness of individualized fall prevention program in geriatric rehabilitation hospital setting: a cluster randomized trial. Aging Clin Exp Res 2015 Oct;27(5):681-688. [doi: 10.1007/s40520-015-0330-7] [Medline: 25697080]
31. Quiaios L, Krief H, Lamy O, da Rocha CG, Siegle-Authemayou ML, Lang PO. Impact of an early geriatric rehabilitation program in acutely hospitalised vulnerable patients: a real-life study in an ACE unit in Switzerland. Geriatr Psychol Neuropsychiatr Vieil 2023 Mar 1;21(1):51-62. [doi: 10.1684/pnv.2023.1089] [Medline: 37115679]

XSL•FO
RenderX

32. Mansell V, Hall Dykgraaf S, Kidd M, Goodyear-Smith F. Long COVID and older people. Lancet Healthy Longev 2022 Dec;3(12):e849-e854. [doi: 10.1016/S2666-7568(22)00245-8] [Medline: 36480981]
33. Lettow B, Wouters M, Sinnige J. E-health, wat is dat? Nictiz. URL: https://nictiz.nl/app/uploads/2022/08/E-health-Wat-is-dat.pdf [accessed 2024-08-28]
34. Scott AR, Alore EA, Naik AD, Berger DH, Suliburk JW. Mixed-methods analysis of factors impacting use of a postoperative mHealth app. JMIR Mhealth Uhealth 2017 Feb 8;5(2):e11. [doi: 10.2196/mhealth.6728] [Medline: 28179215]
35. Jensen CM, Overgaard S, Wiil UK, Smith AC, Clemensen J. Bridging the gap: a user-driven study on new ways to support self-care and empowerment for patients with hip fracture. SAGE Open Med 2018;6:2050312118799121. [doi: 10.1177/2050312118799121] [Medline: 30210796]
36. Ling Y, Ter Meer LP, Yumak Z, Veltkamp RC. Usability test of exercise games designed for rehabilitation of elderly patients after hip replacement surgery: pilot study. JMIR Serious Games 2017 Oct 12;5(4):e19. [doi: 10.2196/games.7969] [Medline: 29025696]
37. Li AC, Kannry JL, Kushniruk A, et al. Integrating usability testing and think-aloud protocol analysis with "near-live" clinical simulations in evaluating clinical decision support. Int J Med Inform 2012 Nov;81(11):761-772. [doi: 10.1016/j.ijmedinf.2012.02.009] [Medline: 22456088]
38. Berg K. Measuring balance in the elderly: preliminary development of an instrument. Physiother Can 1989 Nov;41(6):304-311. [doi: 10.3138/ptc.41.6.304]
39. Steffen TM, Hacker TA, Mollinger L. Age- and gender-related test performance in community-dwelling elderly people: six-minute walk test, berg balance scale, timed up & go test, and gait speeds. Phys Ther 2002 Feb;82(2):128-137. [doi: 10.1093/ptj/82.2.128] [Medline: 11856064]
40. Donoghue D, Physiotherapy Research and Older People (PROP) group, Stokes EK. How much change is true change? the minimum detectable change of the Berg Balance Scale in elderly people. J Rehabil Med 2009 Apr;41(5):343-346. [doi: 10.2340/16501977-0337] [Medline: 19363567]
41. Guralnik JM, Ferrucci L, Pieper CF, et al. Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. J Gerontol A Biol Sci Med Sci 2000 Apr;55(4):M221-M231. [doi: 10.1093/gerona/55.4.m221] [Medline: 10811152]
42. Abel B, Bongartz M, Rapp K, et al. Multimodal home-based rehabilitation intervention after discharge from inpatient geriatric rehabilitation (GeRas): study protocol for a multicenter randomized controlled trial. BMC Geriatr 2024 Jan 17;24(1):69. [doi: 10.1186/s12877-023-04634-2] [Medline: 38233746]

Abbreviations

BBS: Berg Balance Scale
BI: Barthel Index
CG: control group
EGR: early geriatric rehabilitation
FIM: Functional Independence Measure
FPP: fall prevention program
IG: intervention group
OP: Otago program
PH-Q4: Four-Item Patient Health Questionnaire
QoL: quality of life
SPPB: short physical performance battery
SSE: square-stepping exercise
TUG: Timed Up and Go Test

Edited by A Mavragani; submitted 21.09.24; peer-reviewed by PE Roos; revised version received 21.11.24; accepted 09.12.24; published 11.02.25.

Please cite as:
Steinmetz C, Stenzel C, Sylvester M, Glage D, Linke A, Sadlonova M, von Arnim CAF, Schnieder M, Valentová M, Heinemann S
Use of a Technology-Based Fall Prevention Program With Visual Feedback in the Setting of Early Geriatric Rehabilitation: Controlled and Nonrandomized Study
JMIR Form Res 2025;9:e66692
URL: https://formative.jmir.org/2025/1/e66692
doi:10.2196/66692

XSL•FO
RenderX

© Carolin Steinmetz, Christina Stenzel, Maj Sylvester, Denis Glage, Anne Linke, Monika Sadlonova, Christine A F von Arnim, Marlena Schnieder, Miroslava Valentová, Stephanie Heinemann. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Acceptance of Unsupervised App-Based Cognitive Assessment in Outpatient Care: An Implementation Study

Iris Blotenberg[1], Dr rer nat; Melanie Boekholt[1], MA; Nils Lieberknecht[2], MSc; Paula Säring[2], Dr rer nat; Jochen René Thyrian[1,3], Prof Dr

[1]German Center for Neurodegenerative Diseases (DZNE), Interventional Health Care Research, Ellernholzstr. 1-2, Greifswald, Germany

[2]neotiv GmbH, Magdeburg, Germany

[3]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany

Corresponding Author:
Iris Blotenberg, Dr rer nat
German Center for Neurodegenerative Diseases (DZNE), Interventional Health Care Research, Ellernholzstr. 1-2, Greifswald, Germany

Abstract

Background: The use of unsupervised digital cognitive assessments provides considerable opportunities for early and comprehensive testing for Alzheimer disease, minimizing the demand on time and personnel resources in medical practices. However, the acceptance within health care has yet to be assessed.

Objective: In this implementation study, the acceptance of an app-based, repeated cognitive assessment for early symptoms of Alzheimer disease in the outpatient care setting from both physicians' and patients' perspectives was examined.

Methods: In total, 15 primary care practices participated, where patients with self- or relative-reported memory problems could be prescribed an app (neotivCare app [neotiv GmbH]) for comprehensive cognitive testing. Patients used the app to test their episodic memory function weekly for 12 weeks at home. After the testing period and the final consultation, physicians and patients received questionnaires to assess the app's acceptance.

Results: We received completed questionnaires from physicians for 45 patients. In addition, we received 45 completed questionnaires from the patients themselves. The physicians reported that, for most patients, the app supported their decision-making in the diagnostic process (26/45, 58%). In addition, most physicians found the app's information dependable (34/45, 76%) and felt more certain in their decisions (38/45, 84%). From the patients' perspective, a majority felt thoroughly tested (34/45, 76%), and only a few considered the time commitment for the cognitive tests to be too burdensome (7/45, 16%). Furthermore, despite the weekly cognitive testing and the lengthy 12-week testing period, a majority of patients participated in all tests (39/54, 72%).

Conclusions: Our results indicate a high level of acceptance by physicians and patients, suggesting significant potential for the implementation of unsupervised digital cognitive assessments into routine health care. In the future, acceptance should be assessed in large-scale studies, with a particular focus on the impact on health care delivery and patient outcomes.

(JMIR Form Res 2025;9:e62706) doi:10.2196/62706

KEYWORDS

mild cognitive impairment; Alzheimer disease; dementia; cognition; computerized assessment; digital assessment; digital cognitive biomarkers; home-based assessment; digital platform; mobile phone

Introduction

In many countries, the number of people with Alzheimer disease and related dementias is rapidly increasing; today, more than 50 million people are living with dementia, and this number is expected to triple by 2050 [1,2]. Timely diagnosis of Alzheimer disease is a key objective, as interventions and novel pharmacological treatment approaches shift to the disease's early stages [3,4]. It is the primary care physicians who often have the closest contact with their patients and who serve as the patient's initial contact point when they seek clarification for perceived changes in cognitive abilities. However, conventional neuropsychological screening tests lack sensitivity for early stages of Alzheimer disease [5]. The drawback of more comprehensive cognitive test batteries is that these tests are time-consuming and that they need to be administered by specialized personnel, thus they are quite resource-intensive [6]. This makes it challenging to integrate early diagnosis into primary care and outpatient practices, impacting access, especially in rural areas where the nearest memory clinic is often more distant. Another disadvantage is the 1-time testing in the setting of a medical practice, which has been shown to be influenced by daily variations (eg, lack of sleep) [7] and even

XSL•FO
RenderX

the stereotype threat effect [8], reducing reliability and validity of the assessment.

A promising solution lies in digital cognitive assessments at home, which can save time and resources. Furthermore, they offer the possibility for repeated testing, reducing the impact of daily variations and allowing for the assessment of symptom progression over time. In recent years, several mobile apps for cognitive testing for early Alzheimer symptoms have been developed [9,10]. Among them is the neotivCare app (neotiv GmbH), which has been developed based on current insights into the functional anatomy of episodic memory [11-13]. The cognitive tests have already been assessed for their psychometric quality and feasibility in cognitively healthy older adults and in a memory clinic sample [6,14,15]. For other unsupervised mobile apps, feasibility and acceptance have been explored as well [e.g. 16,17].

What is currently lacking and of particular importance is testing the acceptance in the actual health care setting, for which this and other apps have been developed. In this study, participating physicians in outpatient care had the opportunity to prescribe the neotivCare app to patients who consulted them for memory problems. After the patients had used the app for a duration of 3 months, and once the physicians had received and analyzed the test results, the physicians then evaluated the usefulness of the app in the diagnostic process, and the patients provided their feedback on their experience with the app-based testing.

The objective of this study was to evaluate the acceptance of unsupervised app-based cognitive testing within the realm of care from both the physicians' and the patients' perspective. This study adheres to the Guidelines and Checklist for the Reporting on Digital Health Implementations (iCHECK-DH) [18].

Methods

Study Procedure and Questionnaires

The project took place in collaboration between the physicians' network in a neighbourhood in the north-east of Germany (Magdeburg-Schönebeck), a major statutory health insurance in Germany (AOK Saxony-Anhalt), and neotiv GmbH. The study was conducted to test improved diagnostic methods for dementia care according to §140a "Special Care" of the German Social Code Book V. The study was conducted between September 2021 and December 2022 in Magdeburg-Schönebeck. This period coincided with social distancing measures due to the SARS-CoV-2 pandemic.

Participating practices were recruited from the members of the physicians' network Magdeburg-Schönebeck. Patients consulting their physician for memory problems and who fulfilled further inclusion criteria were informed about the opportunity to test their cognition with the neotivCare app for 3 months. Inclusion criteria were self- or relative-reported memory problems persisting for at least 6 months and perceived as progressive, as well as owning a smartphone and being able to use it. In addition, patients had to be insured with a specific major statutory health insurance in Germany (AOK Saxony-Anhalt). The exclusion criterion was a clear indication of dementia, as it requires timely optimal care. After providing informed consent, patients received an activation code to install and use the app. The testing period spanned 12 weeks with weekly assessments. An automated report was generated from the test results and made available to the treating physician.

After the physician and patient had discussed the results, both the physician and the patient received a short questionnaire, the items are displayed in Table 1. The physician answered, for each patient individually, (1) to what extent the test results from the app had supported decision-making in the diagnostic process, (2) how they assessed the dependability of the app information for deciding on the further diagnostic process, and (3) how certain they were, based on the app information, in having made the correct decision about the subsequent process.

Patients completed a questionnaire providing sociodemographic information (age, sex, and educational attainment). In addition, they answered (1) to what extent they felt their cognitive abilities were thoroughly assessed through the app-based assessment and (2) to what extent they felt the assessments were too time-consuming for them.

XSL•FO
RenderX

Table . Description of the items answered by physicians and patients regarding app-based cognitive testing for early symptoms of Alzheimer disease in this implementation study in outpatient care.

Item		Response options
Physician questionnaire		
	How much did the app information provided about the patient support your decision regarding the further diagnostic process?	• Strongly • Rather strongly • Rather limited • Not at all
	How do you assess the dependability of the app information provided for deciding on the further diagnostic process?	• Dependable • Rather dependable • Rather undependable • Undependable
	How certain were you that, based on the app information provided about the patient, you made the right decision for the further process?	• Certain • Rather certain • Rather uncertain • Uncertain
Patient questionnaire		
	The regular memory assessments with the neotivCare app were, in terms of timing, for me:	• Very burdensome • Quite burdensome • Minimally burdensome • Not burdensome
	Through the regular memory assessments with the neotivCare app, I felt that I was thoroughly examined.	• Fully applies • Applies more • Applies less • Does not apply

Ethical Considerations

Before the commencement of the study, the study was first ethically reviewed and approved by the ethical board of the physicians' network Netz Magdeburg Schönebeck (Network Magdeburg Schönebeck), an association of resident physicians that participated in this study. Ethics approval was granted (no reference number) and a written statement confirming the approval is available. However, this statement is dated after the study's commencement, as the authors requested documentation of the prior ethical review. The patients were thoroughly informed about the study by their physician and through a patient information sheet. They provided informed consent through a participation declaration (Multimedia Appendix 1) before receiving an activation code to install and access the neotivCare app. The study reports results based on returned questionnaires regarding app evaluation; all questionnaires were anonymous. Linkage between the questionnaires, with medical records or with the results from the neotivCare app, was not possible. The participating physicians were compensated by the collaborating health insurance for their additional efforts, such as informing and enrolling patients in the study. The patients did not receive any compensation for their participation in the app-based cognitive testing. The results, figures, or supplements reported in the manuscript do not allow any identification of individual study participants.

Sample

Medical practices

A total of 19 physicians from 15 medical practices participated in the study, including 2 (11%) specializing in internal medicine and 2 (11%) in neurology, with the remaining participating physicians being general practitioners (15/19, 79%). Physicians provided anonymized questionnaire responses on their acceptance of the app in the diagnostic process for 45 patients.

Patients

A total of 54 patients were included in the study. Of these, 45 individuals completed a questionnaire on sociodemographic factors and app usage. Table 2 presents the descriptive statistics for the study participants. The average age was 66 (SD 7.2, age range 52-80) years, 56% (25/45) were female. The majority had a moderate level of education (23/45, 51%).

XSL•FO
RenderX

Table . Description of the sociodemographic characteristics of the patients who participated in app-based cognitive testing for early symptoms of Alzheimer disease in this implementation study in outpatient care.

Variable		Value (N=45)
Age (years), mean (SD)		66.4 (7.2)
Sex, n (%)		
	Female	25 (56)
Education, n (%)		
	Low (<10 years of schooling)	11 (24)
	Medium (10 - 11 years of schooling)	23 (51)
	High (≥12 years of schooling)	9 (20)
	Did not disclose	2 (4)

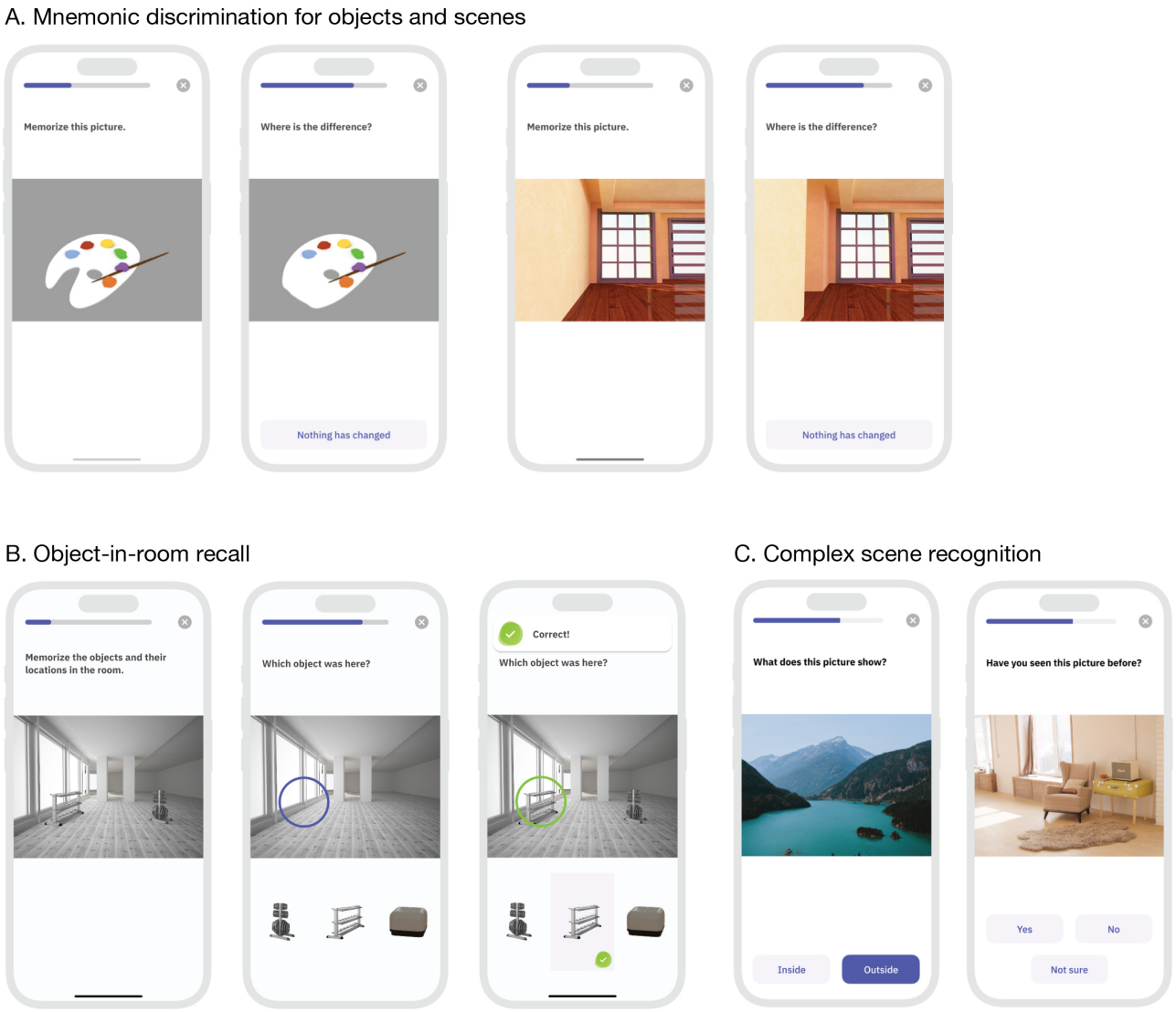
The neotivCare App

Cognitive testing using the neotivCare app was conducted remotely, without supervision, and in a home setting. The patients used their own mobile phone or tablet. The tests were administered weekly over a period of 12 weeks. To remind patients to participate in the weekly test, they received a push notification when a test was available. If a patient did not start the test, they were reminded up to 5 more times via push notification (once a day for 5 days). To boost motivation for regular participation, study teams were trained to inform patients at enrollment that correct test execution and adherence to the study schedule are essential for generating valid data about their cognitive status.

A total of 3 different tasks were presented to assess episodic memory function. Each task was presented 4 times, resulting in the repetition of each specific task every 3 weeks. In the first task, patients were asked to indicate whether an image was an identical repetition or a modified version of a previous image and tap on the location of change (Mnemonic discrimination for objects and scenes, Figure 1A). In the second task, participants had to memorize objects and their positions within a room. Subsequently, they were shown an empty room with a blue circle highlighting the previous position of one of the objects and they had to select the correct object from a set of options (Object-in-room recall, Figure 1B). In the third task, patients were shown photographic images and asked to determine whether they depicted indoor or outdoor scenes. After 60 - 70 minutes, images were presented again, and patients had to decide whether they were new or familiar (Complex scene recognition, Figure 1C).

XSL•FO
RenderX

Figure 1. Illustration of the tasks used in app-based cognitive testing for early symptoms of Alzheimer disease in this implementation study in outpatient care (used with permission from neotiv GmbH).



Statistical Analysis

The processing and analysis of the data were outsourced as an external contract and conducted by an independent research group. The analysis of responses to the items was done descriptively (relative frequencies).

Comparison With Previous Work

A literature search was conducted to identify studies on the feasibility and acceptance of unsupervised digital cognitive assessments in health care and to compare them with our findings.

Results

Acceptance Among Outpatient Physicians

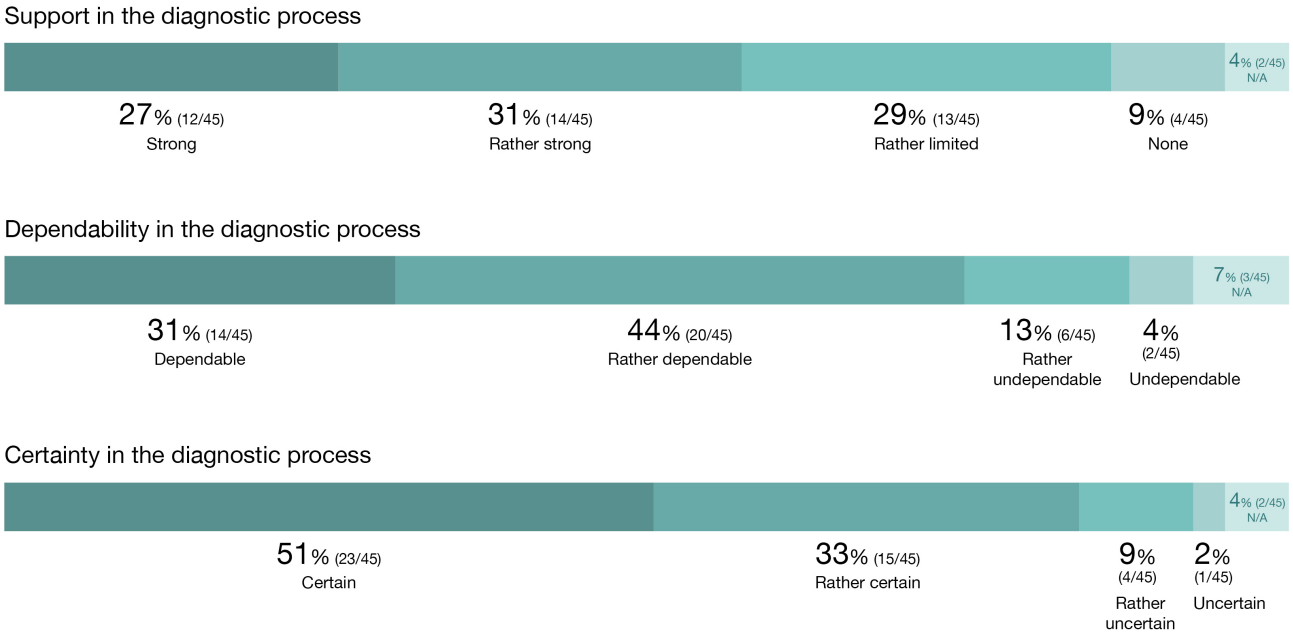
Figure 2 shows the distribution of responses from the participating physicians. For the majority of patients, the physicians indicated that the information from the app supported their decision-making in the diagnostic process (strongly: 12/45, 27%; rather strongly: 14/45, 31%), with only a few cases where the app was perceived by physicians as not supporting the diagnostic process at all (4/45, 9%).

For a clear majority of patients, the physicians assessed the app information as dependable for making decisions about the further diagnostic process (dependable: 14/45, 31%; rather dependable: 20/45, 44%), with only a few patients where the app was considered somewhat (6/45, 13%) or not dependable at all (2/45, 4%).

In the clear majority of cases, the treating physicians reported feeling certain (23/45, 51%) or rather certain (15/45, 33%) in having made the correct decision for the further diagnostic process based on the app information. Only in a small number of cases did the physicians indicate that they were somewhat uncertain (4/45, 9%) or uncertain (1/45, 2%).

XSL•FO
RenderX

Figure 2. Distribution of responses from the physicians concerning acceptance of app-based cognitive testing for early symptoms of Alzheimer disease in this implementation study in outpatient care (data based on questionnaires from the 15 participating practices regarding 45 patients). N/A: not available.



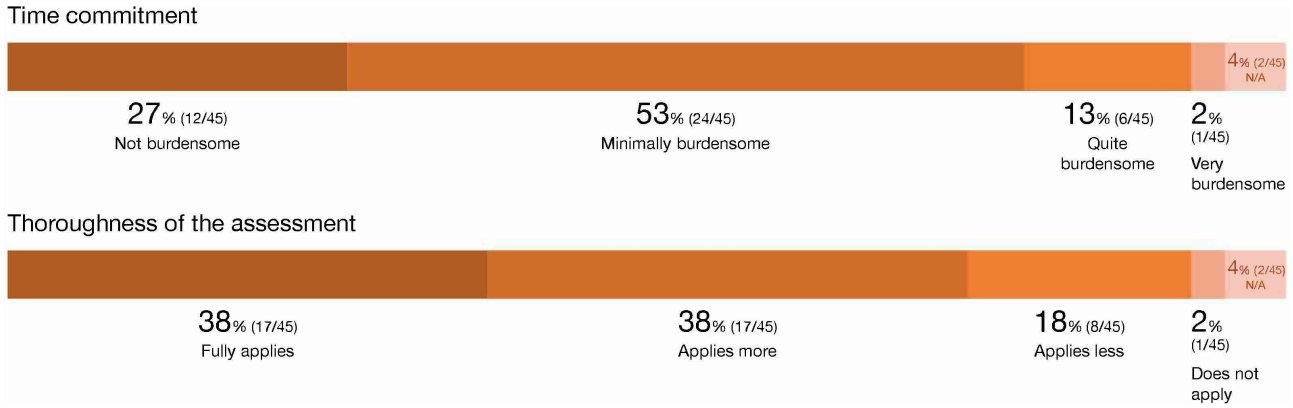
Acceptance Among Patients

Figure 3 shows the distribution of responses from the participating patients. Of the patients who had started participating in the study, 72% (39/54) completed all weekly tests over 12 weeks, indicating high adherence throughout the study period.

When asked about the extent to which using the app was time-consuming, a clear majority responded that they felt either not at all (12/45, 27%) or only slightly (24/45, 53%) time-burdened by using the app. Feeling quite or very burdened was reported by 13% (6/45) and 2% (1/45) of the patients.

A large majority of patients stated that they felt thoroughly examined through the app; 38% (17/45) agreed and another 38% (17/45) somewhat agreed. Slightly more than one-fifth disagreed to some extent (8/45, 18%) or completely (1/45, 2%).

Figure 3. Distribution of responses from the patients concerning acceptance of app-based cognitive testing for early symptoms of Alzheimer disease in this implementation study in outpatient care (data based on responses from 45 of the 54 included patients). N/A: not available.



Discussion

Principal Results

In this implementation study of unsupervised app-based cognitive assessment in outpatient care, both physicians and patients showed high levels of acceptance. The physicians reported, for the majority of patients, that the app had supported the diagnostic process. The majority of patients found the repeated assessments to be only minimally burdensome and reported feeling thoroughly assessed. Most patients participated in all the weekly assessments over the 12-week testing period.

Comparison With Previous Work

Previous studies have reported high feasibility for unsupervised digital cognitive assessments [9]. Several studies have documented high consent rates during recruitment; however, these were from ongoing studies, with the majority of participants approached agreeing to additional digital cognitive testing [19-21]. Adherence rates in previous studies were also high, with most participants completing multiple assessments, ranging from several days [22] to several months [23]). Unsupervised digital cognitive assessments have been tested both in the general population [14] and among older adults [6].

XSL•FO
RenderX

What is currently lacking and urgently needed are studies that test unsupervised digital cognitive assessments within the realm of health care [24,25]. Particularly important in this context is the question of acceptance by physicians and patients, as physicians are responsible for prescribing, interpreting, and acting on these assessments, while patients are required to thoroughly complete the cognitive tests. To our knowledge, there are no published results on this topic and this study is the first to report on the acceptance of such assessments by physicians and patients.

Physician and Patient Perspectives Regarding App-Based Cognitive Assessment

Physicians in our study indicated that, for most patients, the app supported them in making decisions about the diagnostic process. Furthermore, they indicated that they perceived the app information as dependable and that they felt more certain in their decisions. These initial findings suggest that participating physicians have shown acceptance toward the app for evaluating cognitive performance and were using the results to inform their next steps. This is promising, as acceptance from outpatient care physicians is crucial for integrating such tools into routine care and enhancing early testing for Alzheimer disease, as they often serve as the first point of contact for individuals experiencing memory problems.

The participating patients also gave largely positive feedback on the app; a majority felt thoroughly tested, and only a few considered the time commitment for the cognitive tests as too burdensome. Furthermore, despite weekly cognitive testing and the lengthy testing period of 12 weeks, a large majority of patients participated in all tests, indicating high adherence. Reasons for this likely included the generally high motivation of patients in the context of assessing a potential Alzheimer disease diagnosis. In addition, the study design and app programming were aimed at enhancing adherence—study teams were trained to inform patients at enrollment about the necessity of correctly performing tests and adhering to the study plan to generate valid data for the diagnostic report. Furthermore, the app sent weekly push notifications (and reminders) when a new test was available.

These initial results indicate a high level of acceptance and adherence from the patients. This is an important finding for the implementation of app-based cognitive assessment into routine health care, given that it is ultimately the patients who need to integrate the tests into their daily lives and perform them without supervision over an extended time period.

Limitations

To ensure data protection, it was not possible to link the physicians' and patients' questionnaires with each other, nor was it possible to link them with results from the app or additional patient data. The results presented here are solely derived from the returned questionnaires, which were anonymized and provided to participating physicians and patients. The overall sample size was relatively small, and the inclusion criteria of the study may have introduced a potential selection bias in the recruitment process. For example, only patients who owned and could use a smartphone or tablet were included. However, currently, a majority of people aged 55-65 years already own a smartphone, and this number is expected to rise in the future [26]. Furthermore, only patients insured with a specific major public health insurance (AOK Saxony-Anhalt) were eligible to participate. However, as the largest statutory health insurance in Saxony-Anhalt, it covers a cross-section of the population.

Conclusion

The high acceptance and adherence observed in our study is a promising finding for the implementation of unsupervised digital cognitive assessments within the realm of care. Integrating these assessments into routine health care has the potential to enhance early detection of Alzheimer disease, improve access to diagnostics in rural areas, and reduce the burden on medical practices. Future research should focus on larger-scale studies to confirm the high acceptance rates and evaluate the benefits of unsupervised digital cognitive assessments in health care delivery and their impact on patient outcomes.

Acknowledgments

We thank the participants of the study, patients, and physicians in the eastern German area of Magdeburg-Schönebeck. The project was conducted in collaboration of 3 parties: AOK Saxony-Anhalt (a major statutory health insurance in Germany), the physicians' network in Magdeburg-Schönebeck, and neotiv GmbH. The study was funded by AOK Saxony-Anhalt and neotiv GmbH. MB received funding from the Hans and Ilse Breuer Foundation.

Data Availability

The datasets generated or analyzed during this study are not publicly available since the study was conducted under a contract with a major statutory health insurance (AOK Saxony-Anhalt) in accordance with §140a German Social Code Book V (SGB V), which also provided the funding.

Authors' Contributions

IB, JRT, MB, NL, and PS designed the current study. IB and MB analyzed the data. IB, MB, and JRT interpreted the data. IB and JRT wrote the manuscript. All authors critically reviewed the manuscript, provided feedback, and approved it for submission.

XSL•FO
RenderX

Conflicts of Interest

IB, MB, and JRT declare no conflicts of interest. NL and PS are employees of neotiv GmbH; they were involved in conducting the study including data collection, but did not influence preparation, analysis and interpretation of data, which was carried out by IB, MB, and JRT.

Multimedia Appendix 1
Participant declaration.
[PDF File, 147 KB - formative_v9i1e62706_app1.pdf]

References

1. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health 2022 Feb;7(2):e105-e125. [doi: 10.1016/S2468-2667(21)00249-8] [Medline: 34998485]
2. Blotenberg I, Hoffmann W, Thyrian JR. Dementia in Germany: epidemiology and prevention potential. Dtsch Arztebl Int 2023 Jul 10;120(27-28):470-476. [doi: 10.3238/arztebl.m2023.0100] [Medline: 37226316]
3. National plan to address Alzheimer's disease: 2022 update. Assistant Secretary for Planning and Evaluation. 2022. URL: https://aspe.hhs.gov/reports/national-plan-2022-update#main-content [accessed 2025-01-30]
4. Jessen F, Bohr L, Kruse C, Dodel R. Die S3-Leitlinien Demenzen [Article in German]. Nervenarzt 2023 Jul;94(7):609-613. [doi: 10.1007/s00115-023-01492-6]
5. Tsoi KKF, Chan JYC, Hirai HW, Wong SYS, Kwok TCY. Cognitive tests to detect dementia: a systematic review and meta-analysis. JAMA Intern Med 2015 Sep;175(9):1450-1458. [doi: 10.1001/jamainternmed.2015.2152] [Medline: 26052687]
6. Öhman F, Berron D, Papp KV, et al. Unsupervised mobile app-based cognitive testing in a population-based study of older adults born 1944. Front Digit Health 2022;4:933265. [doi: 10.3389/fdgth.2022.933265] [Medline: 36426215]
7. Sliwinski MJ, Mogle JA, Hyun J, Munoz E, Smyth JM, Lipton RB. Reliability and validity of ambulatory cognitive assessments. Assessment 2018 Jan;25(1):14-30. [doi: 10.1177/1073191116643164] [Medline: 27084835]
8. Fresson M, Dardenne B, Geurten M, Meulemans T. The effect of stereotype threat on older people's clinical cognitive outcomes: investigating the moderating role of dementia worry. Clin Neuropsychol 2017 Nov;31(8):1306-1328. [doi: 10.1080/13854046.2017.1307456] [Medline: 28349725]
9. Polk SE, Öhman F, Hassenstab J, et al. A scoping review of remote and unsupervised digital cognitive assessments in preclinical Alzheimer's disease. medRxiv 2024 Sep 27. [doi: 10.1101/2024.09.25.24314349] [Medline: 39399008]
10. Öhman F, Hassenstab J, Berron D, Schöll M, Papp KV. Current advances in digital cognitive assessment for preclinical Alzheimer's disease. Alzheimers Dement (Amst) 2021;13(1):e12217. [doi: 10.1002/dad2.12217] [Medline: 34295959]
11. Düzel E, Berron D, Schütze H, et al. CSF total tau levels are associated with hippocampal novelty irrespective of hippocampal volume. Alzheimers Dement (Amst) 2018;10(1):782-790. [doi: 10.1016/j.dadm.2018.10.003] [Medline: 30555890]
12. Maass A, Berron D, Harrison TM, et al. Alzheimer's pathology targets distinct memory networks in the ageing brain. Brain (Bacau) 2019 Aug 1;142(8):2492-2509. [doi: 10.1093/brain/awz154] [Medline: 31199481]
13. Berron D, Cardenas-Blanco A, Bittner D, et al. Higher CSF tau levels are related to hippocampal hyperactivity and object mnemonic discrimination in older adults. J Neurosci 2019 Oct 30;39(44):8788-8797. [doi: 10.1523/JNEUROSCI.1279-19.2019] [Medline: 31541019]
14. Berron D, Ziegler G, Vieweg P, et al. Feasibility of digital memory assessments in an unsupervised and remote study setting. Front Digit Health 2022;4:892997. [doi: 10.3389/fdgth.2022.892997] [Medline: 35721797]
15. Berron D, Glanz W, Clark L, et al. A remote digital memory composite to detect cognitive impairment in memory clinic samples in unsupervised settings using mobile devices. NPJ Digit Med 2024 Mar 26;7(1):79. [doi: 10.1038/s41746-024-00999-9] [Medline: 38532080]
16. Lancaster C, Koychev I, Blane J, Chinner A, Wolters L, Hinds C. Evaluating the feasibility of frequent cognitive assessment using the Mezurio smartphone app: observational and interview study in adults with elevated dementia risk. JMIR Mhealth Uhealth 2020 Apr 2;8(4):e16142. [doi: 10.2196/16142] [Medline: 32238339]
17. Thompson LI, Harrington KD, Roque N, et al. A highly feasible, reliable, and fully remote protocol for mobile app-based cognitive assessment in cognitively healthy older adults. Alzheimers Dement (Amst) 2022;14(1):e12283. [doi: 10.1002/dad2.12283] [Medline: 35415201]
18. Perrin Franck C, Babington-Ashaye A, Dietrich D, et al. iCHECK-DH: guidelines and checklist for the reporting on digital health implementations. J Med Internet Res 2023 May 10;25:e46694. [doi: 10.2196/46694] [Medline: 37163336]
19. Samaroo A, Amariglio RE, Burnham S, et al. Diminished Learning Over Repeated Exposures (LORE) in preclinical Alzheimer's disease. Alzheimers Dement (Amst) 2020;12(1):e12132. [doi: 10.1002/dad2.12132] [Medline: 33426266]
20. Berron D, Olsson E, Andersson F, et al. Remote and unsupervised digital memory assessments can reliably detect cognitive impairment in Alzheimer's disease. Alzheimers Dement 2024 Jul;20(7):4775-4791. [doi: 10.1002/alz.13919] [Medline: 38867417]

XSL•FO
RenderX

21. Nicosia J, Aschenbrenner AJ, Adams SL, et al. Bridging the technological divide: stigmas and challenges with technology in digital brain health studies of older adults. Front Digit Health 2022;4:880055. [doi: 10.3389/fdgth.2022.880055] [Medline: 35574256]
22. Nicosia J, Aschenbrenner AJ, Balota DA, et al. Unsupervised high-frequency smartphone-based cognitive assessments are reliable, valid, and feasible in older adults at risk for Alzheimer's disease. J Int Neuropsychol Soc 2023 Jun;29(5):459-471. [doi: 10.1017/S135561772200042X] [Medline: 36062528]
23. Jutten RJ, Rentz DM, Fu JF, et al. Monthly at-home computerized cognitive testing to detect diminished practice effects in preclinical Alzheimer's disease. Front Aging Neurosci 2021;13:800126. [doi: 10.3389/fnagi.2021.800126] [Medline: 35095476]
24. Perumal TM, Wolfer A, Lipsmeier F, et al. Quantification of cognitive impairments in preclinical and early Alzheimer's disease: a proof of concept study to investigate the feasibility, adherence and preliminary clinical validity of remote smartphone based self-assessments of cognition, function and behavior (P14-3.002). Neurology (ECronicon) 2022 May 3;98(18_supplement). [doi: 10.1212/WNL.98.18_supplement.1574]
25. Duzel E, Schöttler M, Sommer H, Griebe M. Protocol: prospective evaluation of feasibility, added value and satisfaction of remote digital self-assessment for mild cognitive impairment in routine care with the neotivCare app. BMJ Open 2024 Mar 12;14(3):e081159. [doi: 10.1136/bmjopen-2023-081159] [Medline: 38479732]
26. Berenguer A, Goncalves J, Hosio S, Ferreira D, Anagnostopoulos T, Kostakos V. Are smartphones ubiquitous? An in-depth survey of smartphone adoption by seniors. IEEE Consumer Electron Mag 2016;6(1):104-110. [doi: 10.1109/MCE.2016.2614524]

Abbreviations

iCHECK-DH: Guidelines and Checklist for the Reporting on Digital Health Implementations

Edited by A Mavragani; submitted 30.05.24; peer-reviewed by JL Agne, M Butryn; revised version received 29.11.24; accepted 01.12.24; published 13.02.25.

Please cite as:
Blotenberg I, Boekholt M, Lieberknecht N, Säring P, Thyrian JR
Acceptance of Unsupervised App-Based Cognitive Assessment in Outpatient Care: An Implementation Study
JMIR Form Res 2025;9:e62706
URL: https://formative.jmir.org/2025/1/e62706
doi:10.2196/62706

© Iris Blotenberg, Melanie Boekholt, Nils Lieberknecht, Paula Säring, Jochen René Thyrian. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Users' Perspectives of Direct-to-Consumer Telemedicine Services: Survey Study

Kate Churruca[1], PhD; Darran Foo[1,2], MD; Andrew Turner[1], BClinSci; Emily Crameri[1], MD; Maree Saba[1], BClinSci; Samantha Spanos[1], PhD; Matthew Vickers[3], MBBS; Jeffrey Braithwaite[1], PhD; Louise A Ellis[1], PhD

[1]Australian Institute of Health Innovation, Macquarie University, Macquarie Park, Australia

[2]Healthdirect Australia Limited, Sydney, Australia

[3]Eucalyptus Pty Ltd

Corresponding Author:
Kate Churruca, PhD
Australian Institute of Health Innovation, Macquarie University, Macquarie Park, Australia

Abstract

Background: Commercially run direct-to-consumer (DTC) telemedicine services are on the rise in countries such as Australia and the United States. These include DTC services that are web-based, largely asynchronous, and offer targeted treatment pathways for specific health issues (eg, weight loss or sexual function). It has been argued that DTC telemedicine improves access to health care and promotes patient empowerment. Despite research examining quality and safety issues, little is known about users' reasons for accessing DTC telemedicine services or their perceptions of them.

Objective: In this study, we aimed to examine the perspectives of Australian users accessing DTC telemedicine services, including the reasons for use, perceived benefits, and concerns, in addition to their usage and interaction with traditional general practice services.

Methods: A web-based cross-sectional survey including questions on demographics, published and validated scales, and author-developed closed- and open-response questions was administered via REDCap in 2023 to Australian adults accessing DTC telemedicine services.

Results: Among the 151 respondents, most (136/151, 90.1%) had seen a general practitioner (GP) in the previous 12 months and were somewhat or very satisfied (118/136, 86.8%) with the care, just over half found it easy to get an appointment with their GP (76/151, 50.3%), and a quarter found it difficult (38/151, 25.2%). Among the 136 respondents who had seen a GP, more than half either "never" (55/136, 40.4%) or "rarely" (23/136, 16.9%) discussed the information and treatment received from DTC telemedicine service with their GP. The majority of respondents were using a DTC telemedicine service offering prescription skin care (92/151, 60.9%), had received treatment in the previous 6 months (123/151, 81.5%), and had self-initiated care (128/151, 84.8%). The most frequently cited reasons for using DTC telemedicine were related to convenience (97/121, 80.2%) and flexibility (71/121, 58.7%), while approximately a third of the sample selected that it was difficult to see traditional health care provider in their preferred time frame (44/121, 36.4%) and that the use of DTC telemedicine allowed them to gain access to services otherwise unavailable through traditional health care (39/121, 32.2%). Most participants felt "more in control" (106/128, 82.9%) and "in charge" of their health concern (102/130, 78.5%) when using DTC telemedicine services, in addition to having "more correct knowledge" (92/128, 71.9%) and "feeling better informed as a patient" (94/131, 71.8%). "Costs of services" (40/115, 34.8%) and "privacy" (31/115, 27%) were the most frequently reported concerns with using digital health care technologies such as DTC telemedicine.

Conclusions: We report that most users perceive DTC telemedicine services as offering ease of access and convenience, and that their use contributes to a greater sense of empowerment over their health. Concerns were related to data privacy and the costs of utilizing the services. Responses suggest that DTC telemedicine may be tapping into a previously unmet need, rather than complementing traditional health care provided by a GP.

(JMIR Form Res 2025;9:e68619) doi:10.2196/68619

KEYWORDS

telemedicine; digital health technology; direct-to-consumer; digital tools; telehealth; consumer experience

XSL•FO
RenderX

Introduction

Telemedicine services involve the provision of health care, including the assessment of conditions and the delivery of treatments, through telephone and other digital channels [1,2]. In the past decade, direct-to-consumer (DTC) telemedicine services provided by private companies have become increasingly commonplace; prominent early examples include *Teladoc*, which had 6 million members in the United States by 2013 [3], and now operates in 16 countries including Australia [4]. Consumers using DTC services are able to initiate access to care through the internet, without referral, and often with no prior relationship with the consulting clinician, an approach also sometimes labeled as "commercial virtual visits" [5,6]. As a business model, users are either charged directly, or visits are paid for by contracts between the DTC telemedicine service and health care insurers [7].

The benefits of DTC telemedicine services include convenience and efficiency [8,9]. Patients have reported relatively high levels of satisfaction in using these services [10] and, particularly in the United States, these services have been suggested to increase access to health care for those who do not have insurance coverage [3,5]. By providing consumers with ready access to information, advice, diagnoses, and treatments, DTC telemedicine arguably contributes to patient empowerment through greater knowledge or a sense of control over their health [11,12]. However, issues have been raised about the lack of integration with traditional health care services [3,8] and concerns about the quality of care provided. For example, Schoenfeld et al [5] conducted an audit of 8 of the most widely used DTC companies in the United States, finding that complete history-taking occurred only in about 70% of visits and that wrong or no diagnosis was provided in almost a quarter, with considerable variability identified between conditions (eg, urinary tract infection or low back pain) and companies. The appropriate use of antibiotics for pediatric patients was also found to be lower in DTC telemedicine service, compared with urgent care and primary care providers [13]. However, other recent studies have suggested comparable quality and safety to traditional services in the rate of side effects from medication to treat erectile dysfunction [14] and in the management of respiratory infections [15].

In Australia, the prevalence of DTC telemedicine services has grown rapidly since the COVID-19 pandemic [16]. This includes a new model of DTC telemedicine service, one that is largely asynchronous, may be paid for by consumers on a subscription basis, and is organized around providing medication or treatments for a particular condition or type of issue (eg, men's health, sexual and reproductive health, skin conditions, and weight loss) [8,9,16]. In 2023, Foo and colleagues [16] identified more than 20 DTC telemedicine services in Australia that offer these "targeted treatment pathways for a specific issue or condition," including *Moshy* [17], which markets skin care and weight loss services to women, and *Pilot* [18], which provides a range of men-focused treatments for conditions including weight loss, hair and skin issues, and sexual health issues. As the delivery of health care via DTC telemedicine platforms frequently emphasizes discrete medical conditions, it precludes the development of sustained therapeutic relationships between clinicians and patients. This modality stands in marked contrast to conventional primary care practice, wherein the longitudinal physician-patient relationship and continuity of care constitute fundamental pillars of clinical excellence. Traditional primary care practitioners employ a holistic approach to patient management, rather than compartmentalizing health concerns into isolated conditions—a paradigm that enables comprehensive health oversight and integrated care delivery.

Despite the rise of DTC telemedicine services in Australia, and in countries such as the United States, the viewpoints of consumers have been underexamined [16,19], and this is particularly true for new models of DTC telemedicine that offer targeted treatment pathways for specific health issues. A recently published scoping review identified a clear gap in understanding user perspectives, including the reasons patients may choose to use these services [20]; little is known about the benefits they perceive, and any reservations they have in accessing this type of service [8]. This study aimed to examine the perspectives of consumers accessing DTC telemedicine for specific issues, including interactions with traditional health care services, the types of DTC telemedicine services used and reasons for using, and perceived benefits and concerns.

Methods

A web-based cross-sectional survey that included questions on demographics, published and validated scales, and author-developed closed- and open-response questions was administered to Australian adults accessing DTC telemedicine services.

Participants and Recruitment

Convenience sampling of users of DTC telemedicine services was used. Participants were recruited through an Australian digital health care company offering multiple DTC telemedicine services organized around targeted treatment pathways for specific conditions [16]. The company is a leading provider of DTC telemedicine services in Australia, with recent expansion internationally. At the time of data collection, the 5 different services it ran broadly dealt with the following areas: (1) Women's health care focusing largely on weight management and menopause, (2) Men's health care, (3) Skin care, (4) Women's fertility, and (5) Sexual well-being. The company moderates closed Facebook groups for users of these different services to interact with one another, share their journeys, and ask questions (medical advice is not permitted in the Facebook groups). An advertisement to take part in the survey was posted on the company's closed Facebook communities in November 2023 seeking the perspectives of Australian adults on their use of DTC services.

Materials

The web-based survey was conducted via REDCap hosted by Macquarie University. The survey included questions on traditional health care utilization, followed by a series of closed- and open-response questions about their use of DTC services, including reasons for using and their perceived benefits and concerns (see Multimedia Appendix 1). Finally, the survey

XSL•FO
RenderX

asked for demographic information. All questions were optional to complete.

To assess reasons for using DTC telemedicine services, respondents were asked a series of closed-response questions about experiences with traditional health care services, including whether they had seen a general practitioner (GP) over the preceding 12 months, reasons for accessing traditional health care, satisfaction with this care, and ease of getting an appointment. They were also asked about how they had come to initiate the use of the DTC telemedicine service (ie, recommended by a health care provider, recommended by a family member or friend, or self-initiated) and their main reason(s) for using the service from a list of possible reasons derived from a literature review [3,5-9,16], which included, for example, challenges with accessing traditional health care and convenience.

To assess the perceived benefits of DTC services, we adapted the empowering outcomes scales from van Uden-Kraan et al [12] All 4 items from "being better informed" were used (eg, "I feel better informed as a patient"). From their "increased optimism and control" subscale, 5 of the 8 original items were used (eg, "I feel more in charge of the course of my health concern/s"). Minor modifications were made to wording, such as changing "illness" to "health concern/s." Items for these subscales are answered on a 5-point Likert (1=completely disagree to 5=completely agree). To assess concerns about using digital health care technologies, participants were asked to select as many as apply from a range of options, derived from the literature [3,5-9,16], including safety and quality and the lack of regulation. Participants were also asked 2 open response questions, one about what they liked about using the DTC telemedicine service (ie, benefits), and one about their concerns with using digital health care technologies.

Analysis

Closed-ended data were analyzed using descriptive statistics in IBM SPSS Statistics (version 29; IBM Corp.) to summarize patterns in experiences with traditional health care services, DTC telemedicine services accessed, reasons for using, perceived benefits, and concerns with using digital health care technologies. For location, postcode data were mapped to the Australian Statistical Geography Standard and classified as major city, regional area, or remote area. For the 2 open response questions, inductive content analysis [21] was conducted by KC in Microsoft Excel. This involved initial open coding to write short summaries (codes) of meaning units within each response; this was followed by identifying recurrent meaning units across responses to identify categories. Multiple categories could be present in the response for a single participant and were summarized by frequency and with example extracts.

Ethical Considerations

The study received ethics approval from the Macquarie University Faculty of Medicine, Health and Human Sciences Low-Risk Ethics Subcommittee (No. 520221256844148). All participants opted in to participate and provided voluntary informed consent prior to completing the survey. The survey was designed to be anonymous, with no identifiable data collected. Those who took part in the survey had the option of going into the draw to win a $250 gift card.

Results

Responses were received from 204 people; Table 1 summarizes the characteristics of the participants. A total of 151 (74.0%) of these responses included sufficient information for inclusion in the analysis; 113 respondents completed the demographic information. Approximately two-thirds of those responding to demographic questions (75/113, 66.4%) were female, most were between the ages of 30 - 39 years (30/113, 26.8%) or 40 - 49 years (32/113, 28.6%) and most were employed in full-time paid work (62/113, 55.4%).

Most participants rated their health as "good" (36/115, 31.3%) or "very good" (38/115, 33.0%). Approximately half (58/115, 50.4%) reported having at least one chronic condition, such as a mental or behavioral health condition, back problems, or asthma.

The DTC telemedicine services most frequently used were those for prescription skincare (92/151, 60.9%) and the men's health care service (50/151, 33.1%); the women's health care (4/151, 2.6%), fertility (6/151, 4%), and sexual well-being services (4/151, 2.6%) were each used by only a small number of participants. Some participants (13/151, 8.6%) reported accessing multiple services. Most respondents (123/151, 81.5%) had been prescribed treatment within the previous 6 months.

XSL•FO
RenderX

Table . Demographic characteristics of direct-to-consumer telemedicine users responding to the survey.

Characteristics	N (%)[a]
Sex	
Female	75 (66.4)
Male	38 (33.6)
Age, years	
18 - 29	22 (19.6)
30 - 39	30 (26.8)
40 - 49	32 (28.6)
50 - 59	13 (11.6)
>60	15 (13.4)
Highest level of education	
Postgraduate degree	21 (18.3)
Graduate diploma/certificate	18 (15.7)
Bachelor degree	26 (22.6)
Certificate level	27 (23.5)
Secondary School (year 12 or less)	21 (18.3)
Other	2 (1.3)
Employment status	
Full-time paid work	62 (55.4)
Part-time paid work	34 (30.4)
Unemployed	6 (5.4)
Retired	4 (3.6)
Other	6 (5.4)
Location	
Major city	71 (64.5)
Regional area	35 (31.8)
Remote area	4 (3.6)

[a]Valid percentage used. Note. Responses may not equal 151 owing to missing data.

Experience With Traditional Health Care Services

Most participants (136/151, 90.1%) had seen a GP in the previous 12 months. Of the 15 respondents who had not, most (11/15, 73.3%) had not needed to; only 2 indicated it was too difficult to get an appointment, and 1 suggested it was too expensive. The most common reasons for seeing a GP were to get a prescription (67/136, 49.3%), management of an ongoing condition (64/136, 47.1%), a general or preventive health check-up (49/136, 35.0%), and for information/treatment of a new condition (46/136, 33.8%). A large majority of participants (118/136, 86.8%) were somewhat or very satisfied with the quality of care they received from their GP in the preceding 12 months. Approximately half found it easy or very easy to get an appointment with their GP (76/151, 50.3%), while a quarter found it difficult or very difficult (38/151, 25.2%), and the rest were neutral in their response. A majority of respondents either "never" (55/136, 40.4%) or "rarely" (23/136, 16.9%) discussed the information, advice, support, or treatment they received from the DTC telemedicine service with their GP; of the 136 respondents, 20 (14.7%) "sometimes" did and 22 (16.2%) either "often" or "always" talked to their GP.

Reasons for Using DTC Telemedicine Services

Over four-fifths of respondents (128/151, 84.8%) indicated that they had self-initiated use of the DTC telemedicine service, while 16 of 151 (10.6%) reported that they had received a recommendation from a friend or family member; only 2 of 151 (1.3%) responded that their health care provider had recommended using the service. Reasons for accessing DTC telemedicine services are displayed in Table 2. The most frequently cited reasons were "convenience in accessing information, support, and treatment at a time and place of my choosing" (97/121, 80.2%), "the flexibility and choice available online" (71/121, 58.7%), and "difficulty in seeing a traditional health care provider in the timeframe I would like (eg, appointments booked out weeks in advance or times not suitable for me)" (44/121, 36.4%). Thirty-nine participants among 121

XSL•FO
RenderX

(32.2%) selected one of their main reasons as "gaining access to services or medications not available to me through traditional healthcare services."

Table . Frequency of reported reasons for accessing direct-to-consumer (DTC) telemedicine services among users responding to the survey.

Reasons for accessing DTC telemedicine services	Number of respondents
Convenience in access at a time and place of my choosing	97
Flexibility and choice available online	71
Difficulty accessing traditional health care provider in the timeframe I would like	44
Gaining access to services/medications not available to me through traditional health care services	39
Costs associated with accessing traditional health care services (e.g., not eligible for Medicare, out-of-pocket fees)	39
Preference for accessing health care remotely at this time	31
Check whether my symptoms require attention by a traditional health care provider	16
Better understand the advice or treatment recommended by my traditional health care provider	10
Seeking a second opinion/verifying my traditional health care provider's advice	10
Other	6

Perceived Benefits of Using DTC Telemedicine Services

Figure 1 displays the frequency of participants' responses to items on "being better informed" and having "increased optimism and control" from using the DTC telemedicine service. Most participants agreed that the information, advice, and support they received from the service had made them feel more in control over what was happening to them (106/128, 82.9%) and feel more in charge of the course of their health concern/s (102/130, 78.5%). Most disagreed that the use of the DTC telemedicine service had led them to feel less in control over what was happening to them (98/128, 76.6%). The majority of respondents agreed that they had more correct knowledge at their disposal to manage their health concern/s (92/128, 71.9%) and felt better informed as a patient (94/131, 71.8%).

Seventy-four participants (74/151; 49%) provided an open response for what they liked about using the DTC telemedicine service(s). Eight categories of benefits were identified (Table 3), with "ease and convenience" (n=46) being the most frequently cited.

XSL•FO
RenderX

Figure 1. Perceptions of the impact of DTC telemedicine service use on feeling informed and having more control over health concern among users responding to the survey. * Item is negatively worded.

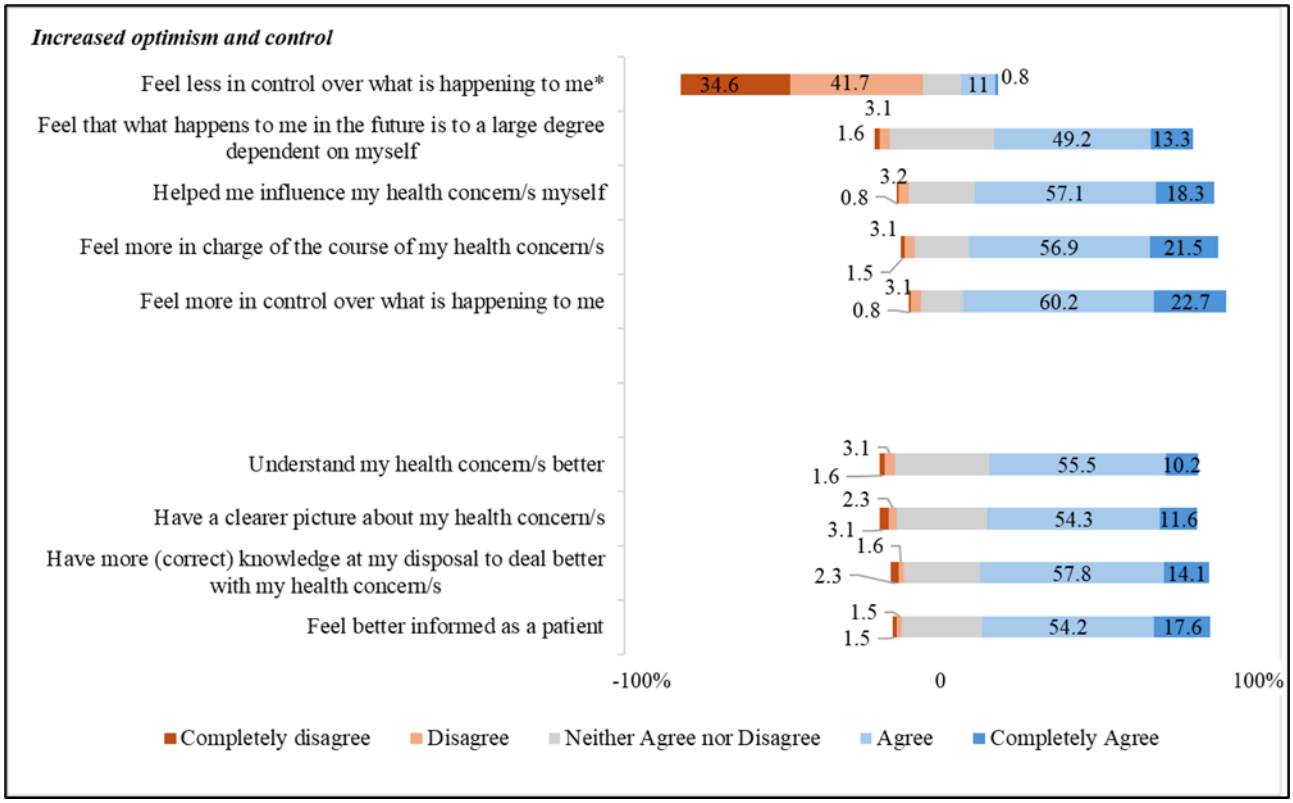


Table . Categories of benefits of DTC telemedicine services identified from open response of users responding to the survey.

Category	Count	Exemplar quote(s)
Ease and convenience	46	"Ease of use, convenience" (ID132) "Convenient access at a time of my choosing" (ID188)
Quality of service	12	"Quality professional advice" (ID86) "Great communication from the staff (doctors)" (ID206)
Efficiency	11	"it means I don't need to wait for an appointment to see my Dr" (ID119) "Ease of accessing treatments and to have it delivered with minimal downtime of face to face appointments" (ID81)
Benefits of the interface/platform	9	"It can be done from home and I am able to type messages in order to be better understood hopefully." (ID145)
Privacy	8	"Some face-to-face appointments can be daunting either because of the cost associated or the topic of conversation itself. I like that it's affordable, easily accessible and discreet" (ID110)
Cost savings	6	"It's expensive and time consuming seeing a doctor face-to-face. (Service) works, and with little effort" (ID146)
Enabling access to medication/treatment	4	"It is a reliable way to get custom skin care with medical grade actives to help with acne control." (ID197)
Greater control over health	3	"My ability to manage my needs myself to a degree." (ID148)

Concerns With Using Digital Health Care Technologies

Table 4 shows the frequency of participants' concerns with using digital health care technologies. The most frequently selected concern was "the costs associated with some of these services" (40/115, 34.8%), followed by "concerns over privacy and what the information I provide will be used for" (31/115, 27%).

XSL•FO
RenderX

Thirty-four respondents (34/151, 22.5%) provided open response answers for their concerns about using digital health care technologies, which were coded into 6 categories. The most frequent category of concern was related to "privacy and data security" (n=13). Table 5 summarizes the categories of concerns with using digital health care technologies.

Table . Frequency of reported concerns with using digital health care technologies among direct-to-consumer telemedicine users responding to the survey.

Concerns with using digital health care technologies among DTC telemedicine users	Number of respondents
Costs associated with services	40
Concerns over privacy and what the information I provide will be used for	31
Lack of knowledge of services	27
Lack of integration	26
Concerns about safety, quality, or accuracy of the advice and services	25
Lack of regulation	17
Other	6

Table . Categories of concerns with using digital health care technologies identified from open response of direct-to-consumer telemedicine users responding to the survey.

Category	Count	Exemplar quote(s)
Privacy and data security	13	"Concerned about my data leaking (because I am getting treatment for Erectile Dysfunction)" (ID178) "Data being collected and sold" (ID38)
Lack of face-to-face interaction	6	"Them not being able to physically assess me" (ID71)
Expensive	6	"More so the costing as it can get expensive" (ID46)
Need to share with GP	3	"If I had a problem with treatment I would be best to see my GP about it" (ID36)
Lack of regulation	3	"Might be a scam" (ID164)
Quality issues	2	"Authenticity of the providers and the information given" (ID181)
Other	5	For example, difficulty accessing medication, digital literacy.

Discussion

Principal Results

This study of Australian users of new DTC telemedicine services, organized around specific health issues (eg, weight loss and skin concerns), demonstrated that users largely seek out these services of their own volition for reasons of convenience and flexibility. As a result of using these services, most users feel they are better informed and have greater control over their health concerns; they perceive further benefits in the ease of use and, to a lesser extent, the quality of the services. Users were less likely to indicate concerns with using digital health care technologies and, where they did, these mostly related to personal factors such as the costs of services or reservations about the privacy of personal health information.

Comparison With Prior Work

Commentaries and empirical research have previously indicated that the use of DTC telemedicine services may support access to health care [3,10]. However, results from the present study do not indicate difficulty in accessing health care is a great driver of use, with most of the sample participants having seen a GP in the past year, reporting satisfaction with the service, and the ability to get an appointment. Prior research largely comes from the United States, which has a more pluralistic health care system with greater inequity in access than Australia. Moreover, these studies have predominately focused on DTC services providing synchronous "virtual visits" for urgent, but not life-threatening, conditions [3,5,10], rather than the new forms of DTC that offer "lifestyle medications" for the management of often ongoing health issues (eg, weight management, acne, and erectile dysfunction) [8,16]. For the latter service type, our results indicate that the benefit may be less in accessing health care than in accessing specific treatments that are otherwise inconvenient, difficult, or impossible to access through one's traditional health care provider, even despite the increased adoption of telehealth in Australian general practice since the COVID-19 pandemic [22].

XSL•FO
RenderX

Very few respondents indicated that they used DTC telemedicine to complement traditional health care (such as getting second opinions or checking symptoms). This suggests DTC telemedicine may be tapping into previously unmet health care needs, rather than simply replacing conventional GP visits. Most respondents also rarely if ever discussed using these services with their GP, further implying the provision of care for their specific health condition remained separate to their use of traditional health care services. What is apparent, though, is that use of these services contributed to a sense of empowerment among participants, who indicated feeling better informed and able to manage their health condition, a finding also identified by Kyweluk [23] in a study of users of DTC ovarian reserve testing services.

Compared with these benefits, participants in this sample far less frequently highlighted concerns with using digital health care technologies, such as DTC telemedicine services. The most common apprehensions were related to personal issues of privacy and cost; few respondents indicated feeling concerned about more systemic factors such as lack of regulation, integration, and quality and safety, despite these being frequently raised in the literature [3,8,9]. The perception of these services as costly may be more pertinent to Australian consumers, many of whom have historically been able to visit a GP without incurring out-of-pocket costs [24], and then access heavily subsidized medicines through the Pharmaceutical Benefits Scheme [25].

Strengths and Limitations

The use of existing subscales on patient empowerment [12], and the triangulation of qualitative and quantitative data analysis, are strengths of the study. Limitations include potential self-selection bias in responses to the survey. Data were missing for some demographic characteristics and other variables due to attrition and the fact that all responses were optional; we lack information as to the profile of users of these DTC telemedicine services for specific health conditions to understand how representative the sample may have been.

Conclusions

This study provides insights into perspectives of patients who use DTC telemedicine services for specific health concerns, which is an emergent area of the health care system and one with limited research to date. Through a survey of users, benefits of these services were identified, including ease of access, convenience, and providing a sense of empowerment. Some concerns were identified related to data privacy and the costs of utilizing the services. Further research is needed to understand how the use of DTC telemedicine fits into the existing landscape of primary health care and the degree to which the use of these services provides a complement, supplement, or alternative to traditional general practice.

Data Availability

The data reported on in this manuscript are not publicly available but may be available from the corresponding author upon reasonable request and with appropriate ethics approval.

Authors' Contributions

KC and LAE conceptualized and designed the study with input from DF, SS, and JB. MS contributed to the ethics and survey administration. MV facilitated recruitment. KC led the data analysis and interpretation, supported by EC and AT. KC wrote the original manuscript draft, which was critically revised by DF, AT, EC, MS, SS, MV, JB, and LAE. All authors approved the final manuscript.

Conflicts of Interest

MV works for and owns shares in the digital health care company whose users participated in this research. At the time of this research, DF provided clinical services for the digital health care company whose users participated in this research. DF is no longer affiliated with this company. All other authors declared no conflicts of interest.

Multimedia Appendix 1
Survey items.
[DOCX File, 21 KB - formative_v9i1e68619_app1.docx]

References

1. Bocking H, Russell-Bennett R, Letheren K. I'm no expert, but …? Consumer use of supportive digital tools in health services. J Serv Theory Pract 2022 Mar 9;32(2):105-131. [doi: 10.1108/JSTP-09-2020-0225]
2. Li KY, Zhu Z, Ng S, Ellimoottil C. Direct-to-consumer telemedicine visits for acute respiratory infections linked to more downstream visits. Health Aff (Millwood) 2021 Apr;40(4):596-602. [doi: 10.1377/hlthaff.2020.01741] [Medline: 33819099]
3. Uscher-Pines L, Mehrotra A. Analysis of Teladoc use seems to indicate expanded access to care for patients without prior connection to a provider. Health Aff (Millwood) 2014 Feb;33(2):258-264. [doi: 10.1377/hlthaff.2013.0989] [Medline: 24493769]
4. Teladoc Health. Teladoc Health. 2024. URL: https://teladochealth.com.au [accessed 2025-01-22]

XSL•FO
RenderX

5. Schoenfeld AJ, Davies JM, Marafino BJ, et al. Variation in quality of urgent health care provided during commercial virtual visits. JAMA Intern Med 2016 May 1;176(5):635-642. [doi: 10.1001/jamainternmed.2015.8248] [Medline: 27042813]
6. Cohen AB, Mathews SC, Dorsey ER, Bates DW, Safavi K. Direct-to-consumer digital health. Lancet Digit Health 2020 Apr;2(4):e163-e165. [doi: 10.1016/S2589-7500(20)30057-1] [Medline: 33328077]
7. Welch BM, Harvey J, O'Connell NS, McElligott JT. Patient preferences for direct-to-consumer telemedicine services: a nationwide survey. BMC Health Serv Res 2017 Nov 28;17(1):784. [doi: 10.1186/s12913-017-2744-8] [Medline: 29183372]
8. Elliott T, Yopes MC. Direct-to-consumer telemedicine. J Allergy Clin Immunol Pract 2019;7(8):2546-2552. [doi: 10.1016/j.jaip.2019.06.027] [Medline: 31706486]
9. Jain T, Lu RJ, Mehrotra A. Prescriptions on demand: the growth of direct-to-consumer telemedicine companies. JAMA 2019 Sep 10;322(10):925-926. [doi: 10.1001/jama.2019.9889] [Medline: 31348489]
10. Martinez KA, Rood M, Jhangiani N, et al. Patterns of use and correlates of patient satisfaction with a large nationwide direct to consumer telemedicine service. J Gen Intern Med 2018 Oct;33(10):1768-1773. [doi: 10.1007/s11606-018-4621-5] [Medline: 30112737]
11. Chung R. Telehealth physician oversight over direct to consumer testing: doctors working with patients towards patient empowerment. Telemed J E Health 2012 Jun;18(5):354-359. [doi: 10.1089/tmj.2011.0196] [Medline: 22545797]
12. van Uden-Kraan CF, Drossaert CHC, Taal E, Seydel ER, van de Laar M. Participation in online patient support groups endorses patients' empowerment. Patient Educ Couns 2009 Jan;74(1):61-69. [doi: 10.1016/j.pec.2008.07.044] [Medline: 18778909]
13. Ray KN, Shi Z, Gidengil CA, Poon SJ, Uscher-Pines L, Mehrotra A. Antibiotic prescribing during pediatric direct-to-consumer telemedicine visits. Pediatrics 2019 May;143(5):e20182491. [doi: 10.1542/peds.2018-2491] [Medline: 30962253]
14. Broffman L, Barnes M, Stern K, Westergren A. Evaluating the quality of asynchronous versus synchronous virtual care in patients with erectile dysfunction: retrospective cohort study. JMIR Form Res 2022 Jan 13;6(1):e32126. [doi: 10.2196/32126] [Medline: 34905499]
15. Shi Z, Mehrotra A, Gidengil CA, Poon SJ, Uscher-Pines L, Ray KN. Quality of care for acute respiratory infections during direct-to-consumer telemedicine visits for adults. Health Aff (Millwood) 2018 Dec;37(12):2014-2023. [doi: 10.1377/hlthaff.2018.05091] [Medline: 30633682]
16. Foo D, Spanos S, Dammery G, Ellis LA, Willcock SM, Braithwaite J. The rise of direct-to-consumer telemedicine services in Australia: implications for primary care and future research. Med J Aust 2023 Oct 16;219(8):344-347. [doi: 10.5694/mja2.52097] [Medline: 37666506]
17. Moshy. 2024. URL: https://www.getmoshy.com.au [accessed 2025-01-22]
18. Pilot. 2024. URL: https://pilot.com.au [accessed 2025-01-22]
19. Bollmeier SG, Stevenson E, Finnegan P, Griggs SK. Direct to consumer telemedicine: is healthcare from home best? Mo Med 2020;117(4):303-309. [Medline: 32848261]
20. Foo D, Spanos S, Dammery G, Ellis LA, Willcock S, Braithwaite J. Direct-to-consumer telemedicine in primary care settings: A scoping review of contemporary empirical literature. Aust J Gen Pract 2024;53(11 Suppl):S107-S114. [doi: 10.31128/AJGP-12-23-7070] [Medline: 39542684]
21. Bosco A, Jones KA, Di Lorito C, Stephan BCM, Orrell M, Oliveira D. Changing lifestyle for dementia risk reduction: Inductive content analysis of a national UK survey. PLoS ONE 2020;15(5):e0233039. [doi: 10.1371/journal.pone.0233039] [Medline: 32413085]
22. Ashley C, Williams A, Dennis S, et al. Telehealth's future in Australian primary health care: a qualitative study exploring lessons learnt from the COVID-19 pandemic. BJGP Open 2023 Jun;7(2):BJGPO.2022.0117. [doi: 10.3399/BJGPO.2022.0117] [Medline: 36750374]
23. Kyweluk MA. Quantifying fertility? Direct-to-consumer ovarian reserve testing and the new (in)fertility pipeline. Soc Sci Med 2020 Jan;245:112697. [doi: 10.1016/j.socscimed.2019.112697] [Medline: 31786460]
24. De Abreu Lourenco R, Kenny P, Haas MR, Hall JP. Factors affecting general practitioner charges and Medicare bulk-billing: results of a survey of Australians. Med J Aust 2015 Feb 2;202(2):87-90. [doi: 10.5694/mja14.00697] [Medline: 25627740]
25. Willis E, Reynolds L, Rudge T, editors. The Pharmaceutical Benefits Scheme. In: Understanding the Australian Health Care System, 4th edition: Elsevier Australia; 2020:100-120.

Abbreviations

DTC: direct-to-consumer
GP: general practitioner

XSL•FO
RenderX

Edited by A Mavragani; submitted 11.11.24; peer-reviewed by EVD Velde, K Ufholz, S Bollmeier; revised version received 18.12.24; accepted 19.12.24; published 03.02.25.

Please cite as:
Churruca K, Foo D, Turner A, Crameri E, Saba M, Spanos S, Vickers M, Braithwaite J, Ellis LA
Users' Perspectives of Direct-to-Consumer Telemedicine Services: Survey Study
JMIR Form Res 2025;9:e68619
URL: https://formative.jmir.org/2025/1/e68619
doi:10.2196/68619

© Kate Churruca, Darran Foo, Andrew Turner, Emily Crameri, Maree Saba, Samantha Spanos, Matthew Vickers, Jeffrey Braithwaite, Louise A Ellis. Originally published in JMIR Formative Research (https://formative.jmir.org), 3.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Feasibility, User Acceptance, and Outcomes of Using a Cancer Prehabilitation App for Exercise: Pilot Cohort Study

Fuquan Zhang[1], MD; Deepali Bang[1], MD, PhD, MRCP; Christine Alejandro Visperas[1], MD; Mon Hnin Tun[2], PhD; San San Tay[1], MBBS, MRCP

[1]Department of Rehabilitation Medicine, Changi General Hospital, 2 Simei St 3, Singapore, Singapore

[2]Department of Health Services Research, Changi General Hospital, Singapore

Corresponding Author:
San San Tay, MBBS, MRCP
Department of Rehabilitation Medicine, Changi General Hospital, 2 Simei St 3, Singapore, Singapore

Abstract

Background: The efficacy of cancer prehabilitation programs is supported by international reviews and meta-analyses. Technology has been deployed in cancer prehabilitation to address challenges such as access or limited resources. This study evaluated the feasibility, user acceptance, safety, and program outcomes of a newly developed mobile app for cancer prehabilitation. The app integrates with Singapore's existing health care mobile app, Health Buddy, and provides instructional videos for prescribed exercises.

Objective: The objectives of this study were to investigate the feasibility, user experience, safety, and outcomes of a mobile app for cancer prehabilitation within a hospital-associated, home-based, multimodal cancer prehabilitation program.

Methods: This retrospective study analyzed the records of patients enrolled in the cancer prehabilitation program from September 1, 2022, to March 30, 2023. Patients who participated in the prehabilitation program (n=63) were categorized into 2 groups: those prescribed the app (n=41) and those who were not (n=22). There was further subgroup analysis of those who were prescribed: app users (n=25) versus those who were non-app users (n=16). Demographics, Fried Frailty Phenotype, prehabilitation duration, app use, and functional outcome measures (6-minute walk test [6MWT], 30-second sit-to-stand test [STS], timed up and go test [TUG], and Hospital Anxiety and Depression Scale [HADS]) were collected. Compliance was determined by the completion of prescribed exercises and the accuracy of executing these exercises, with a high compliance rate considered to be at 80% or more. Baseline characteristics and preoperative outcomes were compared between the groups. User satisfaction was assessed through surveys among app users (n=25).

Results: Among 63 patients, 41 (65.1%) patients were prescribed the app, of which 22 (34.9%) patients were users. No significant differences in preoperative functional improvements were observed between app users and nonusers (6MWT: *P*=.60; STS: *P*=.81; TUG: *P*=.53; HADS: *P*=.36), or between those prescribed and not prescribed the app (6MWT: *P*=.94; STS: *P*=.26; TUG: *P*=.39; HADS: *P*=.62). However, high compliance rates (80%) were observed among app users. Patient satisfaction with the app was high (>90%), with positive feedback on ease of use and technical reliability. Baseline measures revealed significantly lower functional scores and higher mean frailty scores in the nonprescribed group.

Conclusions: This preliminary study demonstrates the acceptability, feasibility, and safety of Singapore's first smartphone app for exercise prescription in cancer prehabilitation. Lower baseline functional outcome measures and a higher mean frailty score in the unprescribed group have implications for the selection process and patient participation. Further studies should include strategies to enhance patients' readiness for technology, sustainability, and effectiveness in older patients.

(JMIR Form Res 2025;9:e64427) doi:10.2196/64427

KEYWORDS

cancer prehabilitation; mobile app; technology; feasibility; acceptance; cancer; prehabilitation; mHealth; exercise; application; app; mobile application; reliability; smartphone app; sustainability; effectiveness; older patients; older adults; elderly

Introduction

Cancer prehabilitation is defined as a set of interventions delivered within the cancer care continuum, occurring during the period between diagnosis and the initiation of definitive treatment [1]. It encompasses baseline assessments of physical function, nutritional status, and psychological well-being. Prehabilitation also involves the identification of potential impairments and the implementation of targeted interventions aimed at optimizing health, thereby reducing the risk and severity of future treatment-related complications. The efficacy of cancer prehabilitation programs is supported by findings

XSL•FO
RenderX

from several international reviews and meta-analyses [2-4]. However, ensuring equitable access to prehabilitation services remains a challenge, particularly in settings with limited health care resources or geographical barriers to care [5]. In another study, the preferred method for exercise program delivery was home-based, with 1 supervised session per week [6]. Technology-enabled prehabilitation programs may be able to support home-based programs, which is important to improve adherence and maximize program effectiveness within a short prehabilitation window [5].

A systematic review of cancer prehabilitation programs delivered through technological enablers found the studies to be feasible. While the effectiveness of these programs varied based on the intervention design and the specific technologies used, many studies reported improvements in physical function, activity levels, and overall patient satisfaction. It was recommended that future efforts focus on implementing multimodal programs and using intervention-specific outcome measures to enhance the overall effectiveness of cancer prehabilitation [7].

Our institution has established a hospital-associated, home-based prehabilitation program targeting patients with newly diagnosed gastrointestinal and urological cancers awaiting surgery [8,9]. This multimodal program incorporates medical optimization strategies, individualized exercise prescriptions, and nutritional and mental health interventions. This program was facilitated by a small core team consisting of a coordinator who collects the outcome measures and a physiatrist who sees the patients on the same day as the surgeon visits, providing a 1-stop service. The compliance of the patients to the home exercise program was tracked by phone calls made by the coordinator on a weekly to fortnightly basis. The program has demonstrated significant improvements in functional capacity, as evidenced by metrics such as the 6-minute walk test (6MWT), 30-second sit-to-stand test (STS), and timed up and go test (TUG). There were significant improvements in the Hospital Anxiety and Depression Scale (HADS) as well.

In November 2021, the program integrated a cancer prehabilitation exercise diary into our existing network's widely used regional health care mobile app, Health Buddy (Singapore Health Services [Singhealth]) [10]. The Health Buddy app allows the user to manage appointments, make payments, and have access to health tools and education.

Despite the wide range of digital health apps [11], no cancer prehabilitation apps are currently available in Singapore. This study aims to evaluate the feasibility, user acceptance, safety, and program outcomes associated with using this mobile app within our multimodal cancer prehabilitation program.

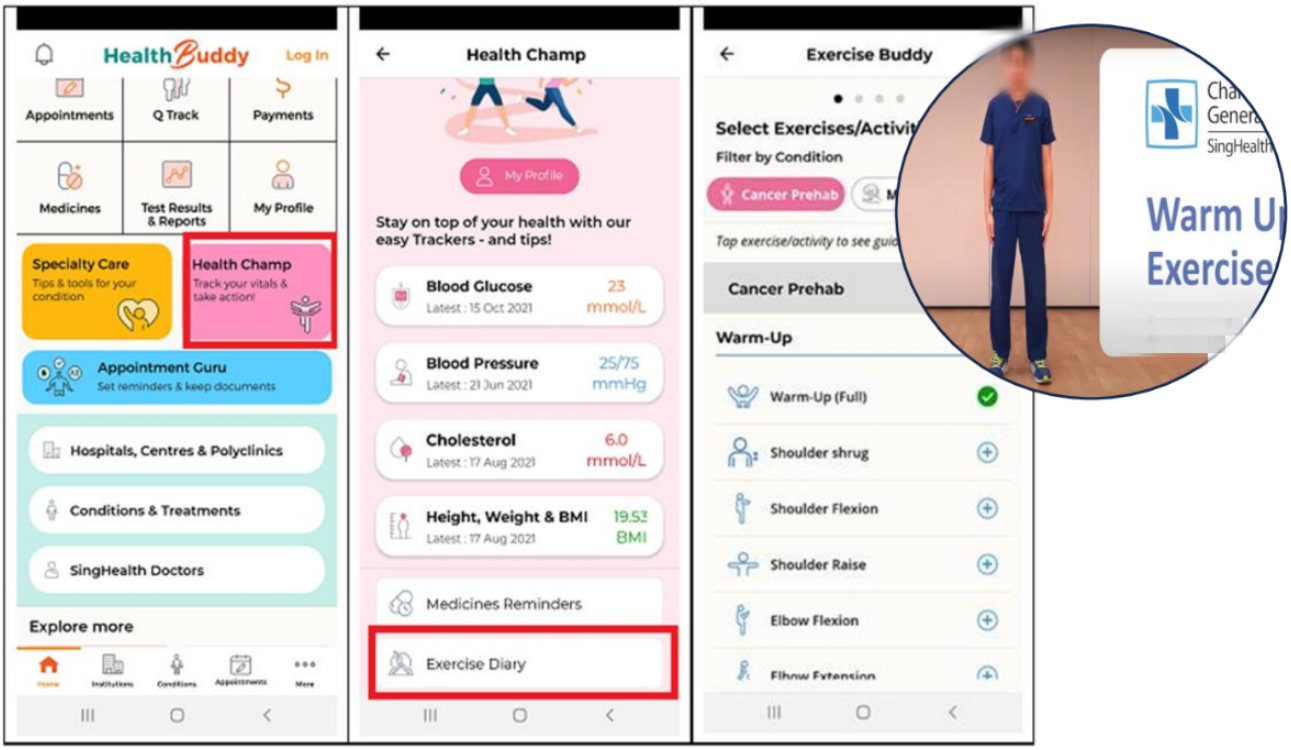
Methods

The Cancer Prehabilitation Exercise Diary

The cancer prehabilitation exercise diary incorporates a library of instructional videos demonstrating commonly prescribed exercises. These exercises are selected by a physiatrist and are customized for individual patients to ensure patient safety and efficacy. Patients can use the videos for guidance during their daily exercise routines. Additionally, the diary features functionalities such as customized exercise reminders, exercise logs for tracking progress, and an achievement summary to motivate program adherence (Figure 1). Some patients are sedentary and older patients may not be exercising regularly or making the right movements. The intention of having this exercise diary was to facilitate patients in their exercises as they would be able to view the exercises and follow along. The intention was that it would be easier to follow than figures on paper and it may encourage increased participation. Patients who decline the app will be provided with paper exercise prescriptions whereas those who accepted the exercise prescription on the exercise diary will follow the videos for strengthening exercises in addition to aerobic exercise prescriptions. Both groups of patients were prescribed both aerobic and strengthening exercises.

XSL•FO
RenderX

Figure 1. Health Buddy app exercise diary.



Aims and Hypothesis

This study aimed to evaluate the feasibility, user acceptance, safety, and program outcomes associated with using the cancer prehabilitation exercise diary within our multimodal cancer prehabilitation program. The hypothesis was that it is feasible and safe, and having the app will facilitate greater compliance and accuracy of the exercises. Patient satisfaction with the app would be sought. The physical and mental health improvements should be no worse off than the conventional group.

Study Design

We examined the records of patients who participated in the cancer prehabilitation program from September 1, 2022, to March 30, 2023. The cancer diagnostic groups included colorectal, hepatobiliary, upper gastrointestinal, and urological cancers planned for surgery at our hospital. They were referred by the surgeons at the point of diagnosis to the prehabilitation coordinator.

The inclusion criteria were cancer patients older than 18 years who attended the surgical clinics and were listed for surgery and who were enrolled in the cancer prehabilitation program. Consecutive patients who were enrolled were studied. Patients who declined surgical intervention were excluded from the study. Outcome measures at baseline and the preoperative period were routinely collected, as well as the patient satisfaction surveys for audit purposes.

Participants who received exercise interventions through the cancer prehabilitation exercise diary were categorized into 2 primary groups: those who were prescribed the app and those who were not. Among those prescribed the app, a further subdivision was made into app users and nonusers. The categorization into app prescribed or not prescribed was not decided a priori but based on the physician's behavior. The categorization of the app used or nonused was also a result of the patient's autonomous decision. As this is a retrospective study, the participants were nonrandomized. Patients prescribed exercises via the app were invited to complete a patient satisfaction survey upon program completion.

Demographics

Demographic data, including age, gender, and disease diagnosis, were obtained from patient medical records. Additionally, medical charts were reviewed for data pertaining to the Fried Frailty Phenotype [12] as well as the duration of the prehabilitation program. Information regarding if the patients were prescribed the app, and the reason if not, as well as the actual use of the app were collected.

The Fried Frailty Phenotype [13] assesses physical frailty through 5 criteria: unintentional weight loss; weakness or poor handgrip strength; self-reported exhaustion; slow walking speed; and low physical activity. Patients would be considered frail if they met the score criteria of at least 2 items, prefrail if they met 1 criteria, and nonfrail if they met none. The duration of prehabilitation is the number of days from enrollment to the preoperative visit (typically held on the day before the surgery).

Baseline and Preoperative Outcome Measures, Compliance, and Adverse Events

Baseline and preoperative outcome measures were collected at the program's initiation and completion (preoperative visit), respectively. Physical function was assessed using the 6MWT, STS, and TUG. Mental health outcomes were evaluated using the HADS.

Compliance data were also captured, with full compliance defined as completing all prescribed exercises and demonstrating

XSL•FO
RenderX

them correctly at the preoperative visit. Partial compliance was defined as having some inaccuracy in exercise performance or not completing the entire set. Although there is no universal agreed-upon threshold, we considered a compliance rate of 80% to be high. Reasons for not prescribing the app or for nonuse of the app were extracted from electronic medical records.

For patients who used the app, a patient satisfaction survey was administered upon program completion. Questions pertaining to the awareness of the Health Buddy app and the use of the smartphone for features within Health Buddy were asked.

Statistical Analysis

Categorical variables were presented as proportions or percentages, while continuous variables were described using means and SDs or medians and IQRs. A linear mixed-effects model for repeated measures was used to assess differences in outcome measures between groups (prescribed app vs nonprescribed app and app users vs nonusers). Two-tailed statistical tests with a significance level (α) of 0.05 were used. All statistical analyses were performed using Stata 16 (StataCorp LP) software.

Ethical Considerations

This study was reviewed and approved by our institutional review board (2021/2751). As all patient information in the dataset was obtained from retrospective records and deidentified, informed consent was waived and allowed for secondary analysis without additional consent. No financial compensation was offered to participants of this study.

Results

Demographics and Baseline Outcome Measures

Between September 1, 2022, and March 30, 2023, a total of 66 patients were enrolled in the cancer prehabilitation program. A total of 3 patients dropped out from surgery, leaving 63 patients to be included in the study (Table 1). Of these, 41 (65.1%) patients were prescribed exercise interventions through the cancer prehabilitation app. The remaining 22 patients (34.9%) were not prescribed the app for various reasons, including lack of a smartphone or data plan (n=5), undocumented physician rationale (n=9), patient refusal due to a preference for established methods (n=6), short duration to the operation (n=1), and hearing impairment (n=1). Details regarding the distribution of these reasons are provided in Table S1 in Multimedia Appendix 1.

The majority of participants were of Chinese ethnicity, reflecting the demographics of Singapore. No statistically significant differences were observed in age or gender distribution between the 2 groups (app-prescribed vs nonprescribed). Baseline physical function assessments revealed significantly lower scores in the nonprescribed group on the 6MWT, STS, and TUG (Table 1). The nonprescribed group also demonstrated a higher baseline depression score on the HADS. Patients who were not prescribed the app exhibited significantly weaker hand grip strength, slower gait speed, and a higher mean frailty score.

Among the 41 patients prescribed exercises via the cancer prehabilitation app, 25 (58.1%) were identified as users, while 16 (41.9%) were nonusers. Baseline outcome measures did not reveal significant differences between these groups (Table 2).

Patients were categorized under “undefined” compliance, as there were no data available for collection in view of patients not being contactable/no-show for appointments. Patients who had their operation canceled were not included in the analysis (reasons for cancellation of operation: medically unfit due to new medical condition [n=2], transferred to another hospital for a second opinion [n=1]).

XSL•FO
RenderX

Table . Baseline characteristics of app-prescribed patients versus nonprescribed patients in outpatient cohort.

	Outpatient (app prescribed)		*P* value
	No (n=22, 34.9%)	Yes (n=41, 65.1%)	
Sex, n (%)			.27
Female	10.0 (45.5)	12.0 (29.3)	
Male	12.0 (54.5)	29.0 (70.7)	
Race, n (%)			.62
Chinese	19.0 (86.4)	35.0 (85.4)	
Malay	2.0 (9.1)	2.0 (4.9)	
Indian	0.0 (0.0)	3.0 (7.3)	
Others	1.0 (4.5)	1.0 (2.4)	
Age (years)			
Mean (SD)	72.1 (8.8)	68.2 (8.9)	.11
Cancer diagnostic groups			.59
Colorectal cancer	7.0 (31.8)	13.0 (31.7)	
Hepatobiliary cancer	9.0 (40.9)	21.0 (51.2)	
Upper gastrointestinal cancer	4.0 (18.2)	3.0 (7.3)	
Urological cancer	2.0 (9.1)	4.0 (9.8)	
Duration of prehabilitation			
Mean (SD)	20.8 (8.9)	20.3 (9.3)	.84
Baseline			
Frailty score			
Mean (SD)	2.3 (1.5)	1.3 (1.2)	0.002[a]
Median (IQR)	2.0 (1.0-3.0)	1.0 (0.0-2.0)	0.003[a]
6MWT[b]			
Mean (SD)	267.6 (89.7)	348.0 (88.3)	0.002[a]
30-second sit-to-stand test			
Mean (SD)	8.5 (3.7)	10.7 (3.9)	.06
Time up and go			
Mean (SD)	16.3 (10.5)	10.5 (3.9)	0.003[a]
Anxiety			
Mean (SD)	2.1 (2.0)	2.3 (2.3)	.71
Depression			
Mean (SD)	2.5 (4.2)	1.2 (1.8)	.08
EQ-5D[c]			
Mean (SD)	68.0 (14.9)	70.2 (20.4)	.66
Anxiety and depression base			
Mean (SD)	4.6 (4.9)	3.5 (3.5)	.31
Maximal hand grip strength (kg)			
Mean (SD)	20.3 (11.8)	27.1 (6.9)	0.001[a]
Gait speed less than 0.67 m/s, n (%)			0.002[a]
No	13.0 (59.1)	38.0 (92.7)	

XSL•FO
RenderX

	Outpatient (app prescribed)		*P* value
	No (n=22, 34.9%)	Yes (n=41, 65.1%)	
Yes	9.0 (40.9)	3.0 (7.3)	
Reduced physical activity, n (%)			>.99
No	15.0 (68.2)	27.0 (65.9)	
Yes	7.0 (31.8)	14.0 (34.2)	
Exhaustion, n (%)			.36
No	15.0 (68.2)	33.0 (80.5)	
Yes	7.0 (31.8)	8.0 (19.5)	
Unintentional weight loss, n (%)			.05
No	11.0 (50.0)	31.0 (75.6)	
Yes	11.0 (50.0)	10.0 (24.4)	
Compliance, n (%)			
Undefined	11.0 (50.0)	8.0 (19.5)	—[d]
Full	8.0 (36.4)	24.0 (58.5)	.04[a]
Partial	3.0 (13.6)	9.0 (21.9)	.14

[a]Statistically significant values.

[b]6MWT: 6-minute walk test.

[c]EQ-5D: EuroQol group 5-dimensions questionnaire.

[d]Not applicable.

XSL•FO
RenderX

Table . Baseline characteristics of users and nonusers of the app in outpatient cohort.

	Outpatient (app used)		*P* value
	No (n=16, 41.9%)	Yes (n=25, 58.1%)	
Sex, n (%)			.08
Female	2.0 (12.5)	10.0 (40.0)	
Male	14.0 (87.5)	15.0 (60.0)	
Race, n (%)			.59
Chinese	14.0 (87.5)	21.0 (84.0)	
Malay	0.0 (0.0)	2.0 (8.0)	
Indian	1.0 (6.3)	2.0 (8.0)	
Others	1.0 (6.3)	0.0 (0.0)	
Age (years)			
Mean (SD)	67.0 (10.5)	69.0 (7.9)	.48
Cancer diagnostic groups			.28
Colorectal cancer	4.0 (25.0)	9.0 (36.0)	
Hepatobiliary cancer	11.0 (68.8)	10.0 (40.0)	
Upper gastrointestinal cancer	0.0 (0.0)	3.0 (12.0)	
Urological cancer	1.0 (6.2)	3.0 (12.0)	
Duration of prehabilitation			.77
Mean (SD)	19.8 (9.2)	20.7 (9.5)	
Baseline			
Frailty score			
Mean (SD)	1.1 (0.9)	1.3 (1.3)	.63
Median (IQR)	1.0 (0.0-2.0)	1.0 (0.0-2.0)	.81
6MWT[a]			
Mean (SD)	362.7 (98.4)	338.6 (81.9)	.40
30-second sit-to-stand test			
Mean (SD)	10.9 (3.3)	10.5 (4.3)	.73
Time up and go			
Mean (SD)	10.6 (5.2)	10.5 (3.1)	.94
Anxiety			
Mean (SD)	1.8 (2.0)	2.6 (2.4)	.26
Depression			
Mean (SD)	0.8 (1.4)	1.4 (2.0)	.37
EQ-5D[b]			
Mean (SD)	72.8 (18.1)	68.6 (22.0)	.53
Anxiety and depression base			
Mean (SD)	2.7 (2.9)	4.0 (3.7)	.23
Maximal hand grip strength (kg)			
Mean (SD)	28.7 (7.8)	26.08 (6.4)	.26
Gait speed less than 0.67 m/s, n (%)			.10
No	15.0 (93.4)	23.0 (92.0)	
Yes	1.0 (6.3)	2.0 (8.0)	

XSL•FO
RenderX

	Outpatient (app used)		*P* value
	No (n=16, 41.9%)	Yes (n=25, 58.1%)	
Reduced physical activity, n (%)			.50
No	12.0 (75.0)	15.0 (60.0)	
Yes	4.0 (25.0)	10.0 (40.0)	
Exhaustion, n (%)			>.99
No	13.0 (81.3)	20.0 (80.0)	
Yes	3.0 (18.8)	5.0 (20.0)	
Unintentional weight loss, n (%)			.47
No	11.0 (68.8)	20.0 (80.0)	
Yes	5.0 (31.3)	5.0 (20.0)	
Compliance, n (%)			
Undefined	8.0 (50.0)	0.0 (0.0)	—[c]
Full	4.0 (25.0)	20.0 (80.0)	.001[d]
Partial	4.0 (25.0)	5.0 (20.0)	.005[d]

[a]6MWT: 6-minute walk test.

[b]EQ-5D: EuroQol group 5-dimensions questionnaire.

[c]Not applicable.

[d]Statistically significant values.

Duration of Prehabilitation

Prehabilitation duration was comparable between the 2 groups (mean of 20.3 d for app-prescribed exercise, mean of 20.8 d for nonprescribed exercise) (Table 1). Among patients prescribed the app, the mean prehabilitation duration was slightly longer for app users (20.7 d) compared with nonusers (19.8 d) (Table 2).

Preoperative Outcome Measures, Compliance, and Adverse Events

Preoperative outcome measures were compared between patients prescribed the app and those who were not prescribed the app. No statistically significant differences in functional improvements were observed between the groups (Table 3). This finding is consistent with the observation that both groups experienced improvements in functional outcomes and mental health, with no documented adverse events.

XSL•FO
RenderX

Table . Changes in outcome measures in outpatient cohort between app-prescribed and nonprescribed patients.

Mean changes in outcome	Outpatient (app prescribed)		*P* value
	No (n=22, 34.9%)	Yes (n=41, 65.1%)	
6MWT[a]			
Baseline (mean)	267.6	348.0	—[b]
Pre-op (mean, 95% CI)	24.2 (3.8 to 44.6)[c]	23.3 (9.9 to 36.5)[c]	.94
30-second sit-to-stand test			
Baseline (mean)	8.5	10.7	—
Pre-op (mean, 95% CI)	2.3 (0.6 to 3.9)[c]	1.1 (0.1 to 2.2)[c]	.26
Time up and go			
Baseline (mean)	16.3	10.5	—
Pre-op (mean, 95% CI)	−1.3 (−2.3 to −0.3)[c]	−0.8 (−1.4 to −0.1)[c]	.39
Anxiety			
Baseline (mean)	2.1	2.3	—
Pre-op (mean, 95% CI)	−0.5 (−1.7 to 0.6)	−0.3 (−1.0 to 0.5)	.68
Depression			
Baseline (mean)	2.5	1.2	—
Pre-op (mean, 95% CI)	−0.3 (−1.1 to 0.5)	−0.2 (−0.7 to 0.3)	.84
Anxiety and depression base			
Baseline (mean)	4.6	3.5	—
Pre-op (mean, 95% CI)	−0.9 (−2.6 to 0.6)	−0.5 (−1.6 to 0.5)	.62
EQ-5D[d]			
Baseline (mean)	68.0	70.2	—
Post-op 3 month (mean, 95% CI)	8.8 (−2.3 to 19.8)	−1.6 (−7.9 to 4.7)	.11

[a]6MWT: 6-minute walk test.

[b]Not applicable.

[c]Statistically significant drop or increase from the baseline.

[d]EQ-5D: EuroQol group 5-dimensions questionnaire.

Of all patients who were prescribed the app, 24 patients (58.5%) exhibited full compliance, 9 (21.9%) patients exhibited partial compliance, and 8 (19.5%) patients were lost to follow-up. Among those who were not prescribed the app, 8 (36.4%) patients were fully compliant, 3 (13.6%) patients were partially compliant, and 11 patients were lost to follow up (Table 1). The higher incidence of full compliance in the app-prescribed group as compared with the nonprescribed group was statistically significant (*P*=.04).

Among those who were prescribed the app, there was no significant difference in the functional and mental health outcome improvements during the preoperative period between the app users and the nonusers (Table 4).

XSL•FO
RenderX

Table . Changes in outcome measures between users and nonusers of the app.

Mean changes in outcome	App used		*P* value
	No (n=16, 41.9%)	Yes (n=25, 58.1%)	
6MWT[a]			
Baseline (mean)	362.7	338.6	—[b]
Pre-op (mean, 95% CI)	28.8 (4.8 to 52.8)[c]	21.0 (4.9 to 37.1)[c]	.60
30-second sit-to-stand test			
Baseline (mean)	10.9	10.5	—
Pre-op (mean, 95% CI)	0.9 (−0.9 to 2.9)[c]	1.2 (0.01 to 2.4)	.81
Time up and go			
Baseline (mean)	10.6	10.5	—
Pre-op (mean, 95% CI)	−0.5 (−1.6 to 0.6)[c]	−0.9 (−1.6 to −0.2)[c]	.53
Anxiety			
Baseline (mean)	1.8	2.6	—
Pre-op (mean, 95% CI)	0.4 (−1.0 to 1.8)	−0.6 (−1.6 to 0.4)	.27
Depression			
Baseline (mean)	0.8	1.4	—
Pre-op (mean, 95% CI)	−0.1 (−0.9 to 0.7)	−0.3 (−0.9 to 0.3)	.74
Anxiety and depression base			
Baseline (mean)	2.7	4.0	—
Pre-op (mean, 95% CI)	0.2 (−1.7 to 2.2)	−0.9 (−2.3 to 0.5)	.36
EQ-5D[d]			
Baseline (mean)	72.8	68.6	—
Post-op 3 month (mean, 95% CI)	−5.8 (−17.9 to 6.2)	0.3 (−7.6 to 8.2)	.41

[a]6MWT: 6-minute walk test.

[b]Not applicable.

[c]Statistically significant drop or increase from the baseline.

[d]EQ-5D: EuroQol group 5-dimensions questionnaire.

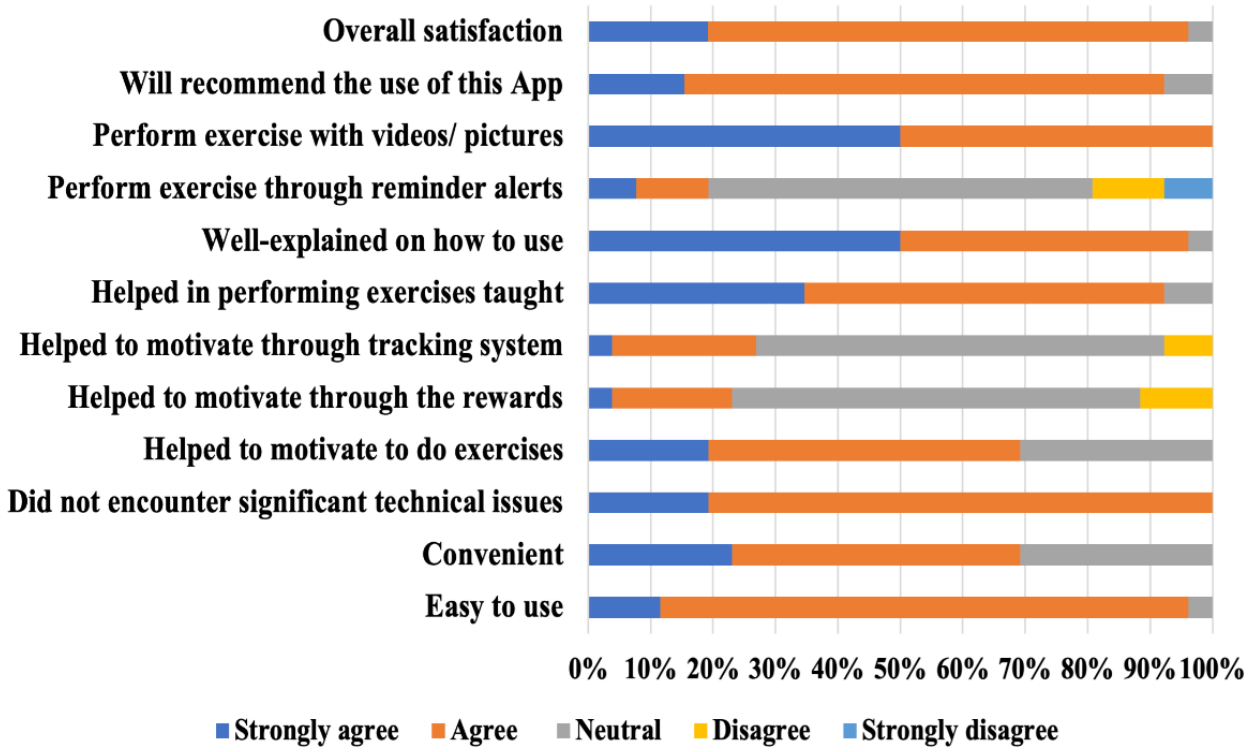
High compliance rates (80%) were observed among app users. Conversely, a high proportion of nonusers (50%) did not return for the preoperative visit or were not contactable. For the remaining nonusers who returned for the pre-op visit and underwent surgery, half (25%) demonstrated full compliance with the exercise program, while the other half (25%) exhibited partial compliance (Table 2). The higher full compliance rate in the app users as compared with the nonusers was statistically significant (*P*=.001).

User Satisfaction Survey

A 76% (19/25) response rate was observed for the patient satisfaction survey among app users. The survey revealed high overall satisfaction with the app, ease of use, and technical reliability, with over 90% (23/25) of respondents reporting these aspects favorably. Furthermore, a high proportion of users indicated a likelihood to recommend the app to others. Patients also reported generally positive experiences with the clarity of exercise instructions and their ability to perform the exercises as instructed (Figure 2).

XSL•FO
RenderX

Figure 2. Impression of the prehabilitation smartphone app.



However, the survey identified the app's motivational features, particularly reminders and rewards, as an area for improvement. Interestingly, a significant portion of app users reported prior familiarity with the Health Buddy app and regular smartphone use for various activities. A total of 100% (25/25) of responders were aware of the Health Buddy app; 100% (25/25) of responders were aware of the Health Buddy functions; and 80% (20/25) , 64% (16/25), and 60% (15/25) of the responders had used the app for checking or scheduling appointments, receiving health education and making payments respectively. Features such as ordering medication and tracking of clinic queues were less used, at 44% (11/25) and 32% (8/25) respectively.

Discussion

Principal Findings

A recent systematic review [13] identified technical barriers, lack of direct supervision, and noncompliance as key challenges in implementing technology-supported cancer prehabilitation programs. Our findings suggest these concerns may not be universally applicable.

Patient satisfaction surveys revealed that app users in our study reported high technical reliability, ease of use, and the ability to follow exercise instructions (Figure 2). This suggests that technical issues were not a significant barrier for our user population. This could likely be attributed to the fact that the exercise diary was incorporated into a well-established health care app in Singapore's largest health care system [10,11]. As such, familiarity with an existing app might have an important role in facilitating ease of use, making this a more user-friendly experience. Additionally, a curated selection of suitable exercise videos will be made available in a patient's individualized cancer prehabilitation exercise diary within the Health Buddy app. This user-centric approach streamlines the process, enabling patients to commence their program effortlessly by accessing the designated videos.

Furthermore, a high compliance rate was observed among app users, with 80% achieving full compliance and 20% achieving partial compliance. These findings indicate that a lack of direct supervision did not hinder functional improvement in our program for the population that used the app. Conversely, those who were not prescribed had a higher noncompliance rate and loss of follow-up. That suggests compliance may be an issue with home-based programs itself.

The development of cancer prehabilitation applications is a growing field. For instance, a prototype app designed to support blood flow restriction training and sports nutrition demonstrated high user satisfaction regarding ease of use, utility, and overall function [14]. Our study adds to this emerging body of evidence by investigating the feasibility and user acceptance of a mobile app for cancer prehabilitation program delivery.

Feasibility and Acceptance of App-Based Prehabilitation

Among the 63 patients in the study, only 41 (65.1%) patients were prescribed exercise interventions through the app. There is borderline acceptance by physicians and patients. Given the full compliance rate of 80% among app users, app-based prehabilitation is feasible.

The main reasons for not prescribing the app included (1) patient limitations: lack of technological knowledge, smartphone/data plan limitations, hearing impairment, or short prehabilitation windows; (2) patient preferences: unwillingness to participate;

XSL•FO
RenderX

and (3) undocumented rationale: in 10 cases, the physician's rationale for not prescribing the app was not documented.

The study highlights the potential impact of socioeconomic factors on technology readiness for cancer prehabilitation. Unlike studies in Western countries with established social acceptance of telehealth [15], a prior local study suggests lower acceptance among diabetic patients [16]. This underscores the need to address potential disparities in access and use. Similar to interventions promoting telehealth for diabetic patients, considerations for app-based prehabilitation might include (1) tailoring support to specific socioeconomic groups less likely to use technology, (2) providing trained staff to address patient concerns and guide app engagement, and (3) exploring the feasibility of loaner devices for the prehabilitation period.

A recent scoping review of reviews on digital health and telehealth in cancer care identified gaps in the literature [12]. These gaps include (1) studies focusing on older adults or those experiencing bereavement, (2) research on the long-term sustainability of telehealth interventions, and (3) comparative studies that directly evaluate telehealth versus in-person interventions.

Similarly, in our patient population, addressing these knowledge gaps is crucial for overcoming challenges associated with nonreadiness and nonacceptance of technology-enabled prehabilitation programs.

Baseline Differences and Preoperative Outcomes

Patients who were not prescribed the app exhibited significantly lower baseline scores on functional performance measures, and a higher baseline depression score on the HADS. Furthermore, the Fried Frailty Phenotype assessment revealed significantly weaker handgrip strength, slower gait speed, a higher incidence of unintentional weight loss, and a higher mean frailty score in the nonapp group [12].

The Fried Frailty Phenotype focuses on physical frailty [12]. Cognitive frailty, defined by the International Consensus Group organized by the International Academy on Nutrition and Aging and the International Association of Gerontology and Geriatrics, refers to a syndrome in older adults characterized by both physical frailty and cognitive impairment, excluding dementia diagnoses [17]. Our study design did not assess cognitive function, leaving it unclear whether cognitive frailty was present in our cohort and potentially influenced physician or patient selection regarding app prescription.

The significantly higher Fried frailty score in the nonprescribed population suggested that being generally frailer and having lower baseline outcome measures may impair program participation. This is in concurrence with a study demonstrating a negative moderate significant correlation found between frailty and participation [18]. This may be even more so for technology-based programs in the face of a challenging new diagnosis. Participation in this study was a broad concept with several aspects, such as preferences, enjoyment, and satisfaction from activities [19].

Both app users and nonusers who were prescribed the app demonstrated improvements in functional measures (6MWT, STS, and TUG) at the preoperative assessment, with no statistically significant differences observed between these groups. Similarly, no significant differences in preoperative outcomes were found when comparing patients prescribed the app to those who were not. While high compliance was observed among app users who were able to follow the exercise videos, these findings suggest no preoperative outcome advantage associated with app use in this preliminary study. Compliance is defined by the completion of the prescribed set of exercises and the accuracy of the exercises. Using the app may have resulted in greater accuracy of the exercise performed, as was intended.

Patient Satisfaction and Future Considerations

A high level of user satisfaction was observed, with 90% of postintervention survey respondents reporting satisfaction with the app and a willingness to recommend it to others. This finding suggests promise for the app's utility among patients receptive to technology-enabled health care interventions.

In a mixed methods systematic review of access, acceptance, and adherence to cancer prehabilitation, there was a perceived value of home-based prehabilitation and the acceptability of tele-prehabilitation [13]. The availability and extent of integrated health care professional supervision and support were perceived to enable intervention access and adherence, especially if this was personalized. Other studies found prehabilitation enabled through technology to be acceptable and perceived to be accessible, especially during the pandemic [6,13,20,21]. Thus, there is value in refining our app-enabled prehabilitation program in the longer term.

Future research efforts should include potential pilot trials among older adults as well as diverse socioeconomic groups, with an aim to adopt strategies tailoring specific support for them while overcoming nonadherence challenges. To promote scalability and integration within our existing health care systems, prioritizing strategies to enhance patient readiness and adoption of technology-based prehabilitation programs should also be considered. Additionally, investigating the long-term sustainability of the intervention within the health care system and staff fidelity to program implementation are crucial areas for further exploration.

Limitations of Study

One limitation of the study is that we have a substantial number of participants who were eventually uncontactable or lost to follow-up, contributing to the missing data pool. We were unable to determine the reason for this, whether this was due to participant disengagement versus an inability to successfully operate the app. Nonetheless, efforts should be made in future research to implement strategies for real-time monitoring of a patient's progress and proactive intervention to address potential issues promptly, such as troubleshooting technical problems early or initiating motivational interviewing to address concerns or disinterest early to help maximize retention.

The participating groups were nonrandomized and heterogeneous. As such, selection bias can also be a potential concern in our study.

XSL•FO
RenderX

Finally, given the small sample size, our findings are preliminary and might be limited to acceptability and usability. Further research with a larger sample is needed to confirm and generalize these results.

Conclusion

In conclusion, this preliminary study demonstrates the acceptability, feasibility, and safety of Singapore's first smartphone app for exercise prescription in cancer prehabilitation. Future research directions include investigating the program's effectiveness among older adults, exploring strategies to enhance patient readiness for technology-based interventions, evaluating the program's long-term sustainability within the health care system, and assessing staff fidelity to program implementation.

Acknowledgments

We would like to express our sincere appreciation to Dr Kwok Kah Meng and Mr Christopher Richard Mattew De Conceicao Paul for their contributions.

Data Availability

The anonymized datasets generated and analyzed in this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Reasons for not prescribing the app.
[DOCX File, 14 KB - formative_v9i1e64427_app1.docx]

References

1. Silver JK, Baima J, Mayer RS. Impairment-driven cancer rehabilitation: an essential component of quality care and survivorship. CA Cancer J Clin 2013 Sep;63(5):295-317. [doi: 10.3322/caac.21186] [Medline: 23856764]
2. Treanor C, Kyaw T, Donnelly M. An international review and meta-analysis of prehabilitation compared to usual care for cancer patients. J Cancer Surviv 2018 Feb;12(1):64-73. [doi: 10.1007/s11764-017-0645-9] [Medline: 28900822]
3. Tay SS. Perspectives on the direction of cancer prehabilitation in the pandemic and beyond. Arch Rehabil Res Clin Transl 2022 Dec;4(4):100236. [doi: 10.1016/j.arrct.2022.100236] [Medline: 36277731]
4. Lambert JE, Hayes LD, Keegan TJ, Subar DA, Gaffney CJ. The impact of prehabilitation on patient outcomes in hepatobiliary, colorectal, and upper gastrointestinal cancer surgery: a PRISMA-accordant meta-analysis. Ann Surg 2021 Jul 1;274(1):70-77. [doi: 10.1097/SLA.0000000000004527] [Medline: 33201129]
5. Ferreira V, Agnihotram RV, Bergdahl A, et al. Maximizing patient adherence to prehabilitation: what do the patients say? Support Care Cancer 2018 Aug;26(8):2717-2723. [doi: 10.1007/s00520-018-4109-1] [Medline: 29478189]
6. Waterland JL, Chahal R, Ismail H, et al. Implementing a telehealth prehabilitation education session for patients preparing for major cancer surgery. BMC Health Serv Res 2021 May 10;21(1):443. [doi: 10.1186/s12913-021-06437-w] [Medline: 33971869]
7. Tay SS, Zhang F, Neo EJR. The use of technology in cancer prehabilitation: a systematic review. Front Oncol 2024;14:1321493. [doi: 10.3389/fonc.2024.1321493] [Medline: 38706603]
8. Tay SS, Kwok KM. Setting up a cancer prehabilitation framework in Singapore. J Cancer Rehabil 2021;4:1-6. [doi: 10.48252/JCR5]
9. Kwok KM, Tay SS. Outcomes of a multi-modal hospital-associated home-based cancer prehabilitation program. Ann Rehabil Med 2023 Feb;47(1):52-67. [doi: 10.5535/arm.22126] [Medline: 36710586]
10. Introducing a new & improved health buddy app interface!. SingHealth. URL: https://www.singhealth.com.sg/patient-care/patient-visitor-info/health-buddy-app [accessed 2024-12-04]
11. Perdana A, Mokhtar IA. Leveraging digital technologies for information technology-enabled healthcare transformation at SingHealth. J Inf Technol Teach Cases 2023 May;13(1):97-103. [doi: 10.1177/20438869221091582]
12. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001 Mar;56(3):M146-M156. [doi: 10.1093/gerona/56.3.m146] [Medline: 11253156]
13. Watts T, Courtier N, Fry S, et al. Access, acceptance and adherence to cancer prehabilitation: a mixed-methods systematic review. J Cancer Surviv 2024 May 6. [doi: 10.1007/s11764-024-01605-3] [Medline: 38709465]
14. Wang T, Stanforth PR, Fleming RYD, Wolf JS Jr, Stanforth D, Tanaka H. A mobile app with multimodality prehabilitation programs for patients awaiting elective surgery: development and usability study. JMIR Perioper Med 2021 Dec 30;4(2):e32575. [doi: 10.2196/32575] [Medline: 34967752]

XSL•FO
RenderX

15. Poder TG, Bellemare CA, Bédard SK, Lemieux R. Social acceptance and population confidence in telehealth in Quebec. BMC Health Serv Res 2015 Feb 21;15:72. [doi: 10.1186/s12913-015-0727-1] [Medline: 25889230]
16. AshaRani PV, Jue Hua L, Roystonn K, et al. Readiness and acceptance of eHealth services for diabetes care in the general population: cross-sectional study. J Med Internet Res 2021 Sep 2;23(9):e26881. [doi: 10.2196/26881] [Medline: 34473062]
17. Kelaiditi E, Cesari M, Canevelli M, et al. Cognitive frailty: rational and definition from an (I.A.N.A./I.A.G.G.) international consensus group. J Nutr Health Aging 2013 Nov;17(9):726-734. [doi: 10.1007/s12603-013-0367-2]
18. Rand D, Sternberg SA, Gasner Winograd R, Buckman Z, Bentur N. The contribution of frailty to participation of older adults. Int J Environ Res Public Health 2022 Jan 30;19(3):1616. [doi: 10.3390/ijerph19031616] [Medline: 35162637]
19. Law M. Participation in the occupations of everyday life. Am J Occup Ther 2002 Nov 1;56(6):640-649. [doi: 10.5014/ajot.56.6.640]
20. Drummond K, Lambert G, Tahasildar B, Carli F. Successes and challenges of implementing teleprehabilitation for onco-surgical candidates and patients' experience: a retrospective pilot-cohort study. Sci Rep 2022;12(1):6775. [doi: 10.1038/s41598-022-10810-y]
21. Wu F, Laza-Cagigas R, Rampal T. Understanding patients' experiences and perspectives of tele-prehabilitation: a qualitative study to inform service design and delivery. Clin Pract 2022 Aug 16;12(4):640-652. [doi: 10.3390/clinpract12040067] [Medline: 36005070]

Abbreviations

6MWT: 6-minute walk test
HADS: Hospital Anxiety and Depression Scale
STS: 30-second sit-to-stand test
TUG: timed up and go test

Edited by A Mavragani; submitted 17.07.24; peer-reviewed by J Khan, M Gasmi; revised version received 18.11.24; accepted 18.11.24; published 20.01.25.

Please cite as:
Zhang F, Bang D, Visperas CA, Tun MH, Tay SS
Feasibility, User Acceptance, and Outcomes of Using a Cancer Prehabilitation App for Exercise: Pilot Cohort Study
JMIR Form Res 2025;9:e64427
URL: https://formative.jmir.org/2025/1/e64427
doi:10.2196/64427

© Fuquan Zhang, Deepali Bang, Christine Alejandro Visperas, Mon Hnin Tun, San San Tay. Originally published in JMIR Formative Research (https://formative.jmir.org), 20.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Longitudinal Changes in Pitch-Related Acoustic Characteristics of the Voice Throughout the Menstrual Cycle: Observational Study

Jaycee Kaufman[1], MS; Jouhyun Jeon[1], PhD; Jessica Oreskovic[1], MASc; Anirudh Thommandram[1], MASc; Yan Fossat[1,2], MS

[1]Klick Applied Sciences, Klick Health, Toronto, ON, Canada

[2]Faculty of Science, University of Ontario Institute of Technology, Oshawa, ON, Canada

Corresponding Author:
Yan Fossat, MS
Klick Applied Sciences, Klick Health, Toronto, ON, Canada

Abstract

Background: Identifying subtle changes in the menstrual cycle is crucial for effective fertility tracking and understanding reproductive health.

Objective: The aim of the study is to explore how fundamental frequency features vary between menstrual phases using daily voice recordings.

Methods: This study analyzed smartphone-collected voice recordings from 16 naturally cycling female participants, collected every day for 1 full menstrual cycle. Fundamental frequency features (mean, SD, 5th percentile, and 95th percentile) were extracted from each voice recording. Ovulation was estimated using luteinizing hormone urine tests taken every morning. The analysis included comparisons of these features between the follicular and luteal phases and the application of changepoint detection algorithms to assess changes and pinpoint the day in which the shifts in vocal pitch occur.

Results: The fundamental frequency SD was 9.0% (SD 2.9%) lower in the luteal phase compared to the follicular phase (95% CI 3.4% - 14.7%; *P*=.002), and the 5th percentile of the fundamental frequency was 8.8% (SD 3.6%) higher (95% CI 1.7% - 16.0%; *P*=.01). No significant differences were found between phases in mean fundamental frequency or the 95th percentile of the fundamental frequency (*P*=.65 and *P*=.07). Changepoint detection, applied separately to each feature, identified the point in time when vocal frequency behaviors shifted. For the fundamental frequency SD and 5th percentile, 81% (n=13) of participants exhibited shifts within the fertile window (*P*=.03). In comparison, only 63% (n=10; *P*=.24) and 50% (n=8; *P*=.50) of participants had shifts in the fertile window for the mean and 95th percentile of the fundamental frequency, respectively.

Conclusions: These findings indicate that subtle variations in vocal pitch may reflect changes associated with the menstrual cycle, suggesting the potential for developing a noninvasive and convenient method for monitoring reproductive health. Changepoint detection may provide a promising avenue for future work in longitudinal fertility analysis.

(JMIR Form Res 2025;9:e65448) doi:10.2196/65448

KEYWORDS

menstrual cycle; women's health; voice; acoustic analysis; longitudinal observational study; fertility tracking; fertility; reproductive health; feasibility; voice recording; vocal pitch; follicular; luteal phase; fertility status; mobile phone

Introduction

The menstrual cycle is a recurring sequence of physiological changes regulated by hormonal fluctuations that prepare the body for potential pregnancy. This cycle is crucial for women's health, as it influences reproductive function, affects various aspects of physical and emotional well-being, and serves as an indicator of overall health. Irregularities or disruptions in the menstrual cycle can signal underlying health issues, making its regular monitoring and understanding essential for women's health care. The female menstrual cycle is broken into 2 distinct phases: the follicular phase, in which hormonal estrogen levels rise, and the luteal phase, in which estrogen levels begin to decrease, and progesterone is elevated. An important step in the evaluation and management of female fertility and pregnancy is awareness of the timing of ovulation, the release of an egg from the ovary for conception. Ovulation occurs between the follicular and luteal phases and is initiated by high estrogen and luteinizing hormone (LH). The fertile window, or the time during the menstrual cycle when conception is most likely, encompasses the days just before and during ovulation. Follicle tracking through ultrasound is a common fertility procedure in a clinical setting and is considered an accurate method to evaluate the true ovulation timing. However, it comes with significant patient burden and cost. Currently, ovulation detection can be estimated at home through tracking the length

XSL•FO
RenderX

of time between periods (calendar method) and assessing the day-to-day changes of cervical mucus, basal body temperature (BBT), and estrogen, progesterone and LH measurements [1].

The current at-home ovulation detection methods come with their own drawbacks. Hormone measurements and BBT recordings require the purchase of additional devices and materials. Cervical mucus monitoring requires proper education on the different appearances and consistencies throughout the menstrual cycle. Perhaps, the easiest and most accessible method, the calendar method, is widely available through various menstrual tracking apps [2]. Both the predicted day of ovulation and the estimated fertile window can be provided; however, the variation in ovulatory days is distinct. It was shown that the most common predicted fertile window days only included the day of ovulation itself for 65% of female participants [2], yielding some uncertainty in the reliability of this method.

An interesting supplement to the calendar method may be achieved via the assessment of voice throughout the menstrual cycle. Increased estrogen, such as what is seen immediately preceding ovulation, leads to increased mucus secretion by the granular glands in the respiratory tract and increased cell permeability. Conversely, laryngeal tissues begin to absorb water, as estrogen levels decrease in the luteal phase, ultimately causing mucosal edema, vascular congestion, and increased vocal fold mass [3,4]. These changes have been associated with decreased vocal range, vocal and pitch instability, vocal fatigue, and reduced power and efficiency in the voice [5,6]. As such, voice-based fertility analysis may capture physiological changes throughout the menstrual cycle, with the added benefit of convenient smartphone-based data collection, similar to the accessibility offered by the calendar method.

While assessing prior work, it is important to note the variation in methodology and subsequent results. An early study evaluated speech recordings of the vowel sounds /a/, /i/, and /u/ from 20 undergraduate female participants. Voice was recorded twice throughout the menstrual—once at ovulation and once 2 - 3 days prior to the onset of menses. There was no difference in hoarseness or fundamental frequency (F0) between the 2 recordings, a result that was confirmed with a validation sample of 27 additional undergraduate female participants [7]. Similar methodology (recording once during menstruation, immediately following menstruation, midcycle, and premenstruation) with vowel sounds [8,9], counting from 1 to 10 [10], and free speech [11] also yielded no significant changes in F0. Conversely, a study conducted with 30 participants recording the vowel sound /a/ once a week for 4 weeks reported a significant increase in F0 in the ovulatory period and a decrease in F0 in the luteal period compared to the other recording time points [12], and high-fertility recordings of a spoken introductory sentence from 69 participants were significantly higher in F0 than in recordings obtained in a low-fertility period [8]. Looking at other acoustic parameters, counting from 1 to 10 yielded significant differences in shimmer (sound wave amplitude variation) and frequency variation between periods of high fertility and low fertility [10,13], and the vowel sounds /a/, /e/, /i/, /o/, and /u/ collected once each within the fertile window, luteal phase, and during menstruation from 44 participants indicated an increased minimum F0 in the fertile window [14]. In addition, a study conducted on 23 female participants recording 1 minute of free speech every day for 1 menstrual cycle showed marginally significant F0 variation with an increase prior to and a distinct drop during ovulation [15]. Furthermore, a study conducted on 17 naturally cycling female participants reported that female voices were rated more attractive in periods of high fertility when tasked with a counting task at 4 equally spaced time points throughout their menstrual cycle [16].

One of the drawbacks of these studies is that, apart from a single study [15], voice recordings were collected at a single time point within a menstrual phase; thus, day-to-day changes within the voice were notably not assessed. Ovulation itself is a result of longitudinal changes in hormone values compared to the previous days. Therefore, it is essential to trace day-to-day changes of acoustic characteristics to understand the effect of ovulation on voice signals.

In this study, we examined longitudinal changes in pitch-related acoustic characteristics using daily smartphone recordings of spoken sentences from 16 healthy female participants. Our overall goal was to determine how these vocal features change in relation to the LH surge and ovulation. We hypothesize that there are temporal changes in the F0 features associated with the menstrual cycle. The analysis is thus split into two sections:

1. Replication of the day-to-day voice change analysis conducted by Fischer et al [15] using smartphone-recorded (real-world) voice segments. The feature set will be expanded from F0 and F0 variation to include metrics denoting the minimum and maximum F0 values within a recording.
2. Employment of changepoint detection, which will be used to understand the statistical moment in which the participant voices change from one behavior to another. We hypothesize that for features significantly different between the follicular and luteal phases, this day will align with the switch between menstrual phases (ovulation). It may also provide insight into how voice may be used as a fertility indicator in future work.

Methods

Study Design and Participants

In total, 19 English-speaking, cis-gender female participants were recruited from Klick Health, a technology, media, and research company in the health care sector based in Toronto, Canada. Recruitment information was posted on the company's internal communication platform, and interested parties contacted our research team privately for more information and to sign the informed consent. After the informed consent was signed, participants were onboarded through either a video call or in person. As part of the onboarding, participants were provided instructions on how to conduct the urine tests, take their BBT, and record their voice through the research app.

All participants reported a natural menstrual cycle without taking hormonal birth control or medication that would affect the menstrual cycle and reported their menstrual cycle as consistent (1 - 2 days variation per cycle) or somewhat consistent (3 - 4

XSL•FO
RenderX

days variation per cycle). Participants were instructed to take their BBT, self-administer a consumer-grade LH urine test (Easy@Home Ovulation Tests; Easy@Home Fertility), and record themselves saying "Hello, how are you?" every morning immediately upon waking for 1 full menstrual cycle. The decision to use a fixed phrase was made based on the success of previous studies in using fixed sentences and conversational speech to identify significant features between menstrual phases compared to vowel sounds [3,8,15]. Participants were instructed to record their voices in a quiet environment with no other voices present in the recording and took pictures of the LH test strips each day to record hormone test results.

The data collection started the day after the end of menstruation, and recordings were collected for 1 full menstrual cycle, ending on the last day of menstruation. Participants did not perform the hormone test on days they were menstruating but did record their voice and BBT. Data (LH test results, BBT, and voice recordings) were collected using a custom mobile app developed by Klick Applied Science that participants installed on their personal cellphones following the procedure below:

1. Participants entered their participant ID (a deidentified alphanumeric code) upon their first login to the mobile app.
2. When the app was opened, the participant selected "Record a New Entry" to input the study data for the day.
3. They then recorded the results of their BBT (to 2 decimal places), their LH ovulation test results, and their voice while holding their phone.
4. After all fields were completed, the participants could submit the data recording. Submitted data were uploaded to a secure Google Firestore database, where it could be accessed by our researchers via a private application programming interface key.

Participants were excluded if the LH peak could not be detected, either through multiple consecutive missed data recordings (n=2) or if no discernable LH rise was detected through hormone tests (n=1). In total, 16 participants were included with a total of 320 voice recordings.

Ethical Considerations

Ethics was obtained from the Canadian SHIELD Ethics Review Board (REB tracking 2022-02-004). Informed consent was collected for each participant, and all procedures corresponded to local guidelines. Participant data were deidentified during collection. Participants were compensated US $77 for participating in the study and received an additional US $38.50 if they recorded all the voice recordings.

Longitudinal Landscape of Voice Signals

Individual participants recorded their voices once a day, every day for 1 menstrual cycle. The average menstrual cycle length was 28.81 (SD 2.90) days. The average length of the follicular phase (including menstruation and the day of the LH surge and the day after) was 15.92 (SD 3.30) days, and the average length of the luteal phase was 12.75 (SD 1.44) days. On average, menstruation was 5.54 (SD 1.66) days and ended 9.44 (SD 3.12) days prior to the LH surge. In total, 320 voice recordings were collected. Study participants had an average age of 30.6 (SD 4.5) years and an average BMI of 22.7 (SD 2.8) kg/m^2 and included various ethnic backgrounds (East or Southeast Asian: n=4, 25%; Jewish: n=1, 6%; Middle Eastern: n=2, 13%; South Asian: n=6, 37%; White: n=3, 19%; self-identification).

Voice Feature Extraction and Analysis

The LH surge was defined to be when the hormone strip test line was at its maximum darkness out of all the days in the menstrual cycle and approximately equal to or greater than the darkness of the control line. The ovulation window was defined as the day of the LH peak and the day after (to cover the 24 - 36 hour likelihood for ovulation). Days prior to this were classified as "follicular," corresponding to the follicular phase in the menstrual cycle, and days following ovulation classification and occurring prior to menstruation were classified as "luteal." Recordings collected while participants were menstruating were labeled "menstrual." BBT was used to verify the likelihood that ovulation occurred through a sustained 0.3 °C increase in the days following the suspected ovulation day.

Each audio recording was assessed by researchers to ensure adequate recording quality and the absence of background noise. In total, 4 vocal features associated with the recording F0 were extracted from each voice recording: the mean (F0AV), SD (F0SD), 5th percentile (F0MIN), and 95th percentile (F0MAX). The 5th and 95th percentiles were chosen as robust analogs to the minimum and maximum F0 values. Voice features were extracted using librosa [17], a Python library widely used for music and audio analysis, providing tools for audio feature extraction and processing tasks. F0 values for the recording were extracted using the *pyin* function, with the specified frequency minimum of C2 (~65 Hz) and maximum of C7 (~2093 Hz). These values were chosen to encompass a wide range of vocal frequencies and are the recommended minimum and maximum frequencies for analysis [17].

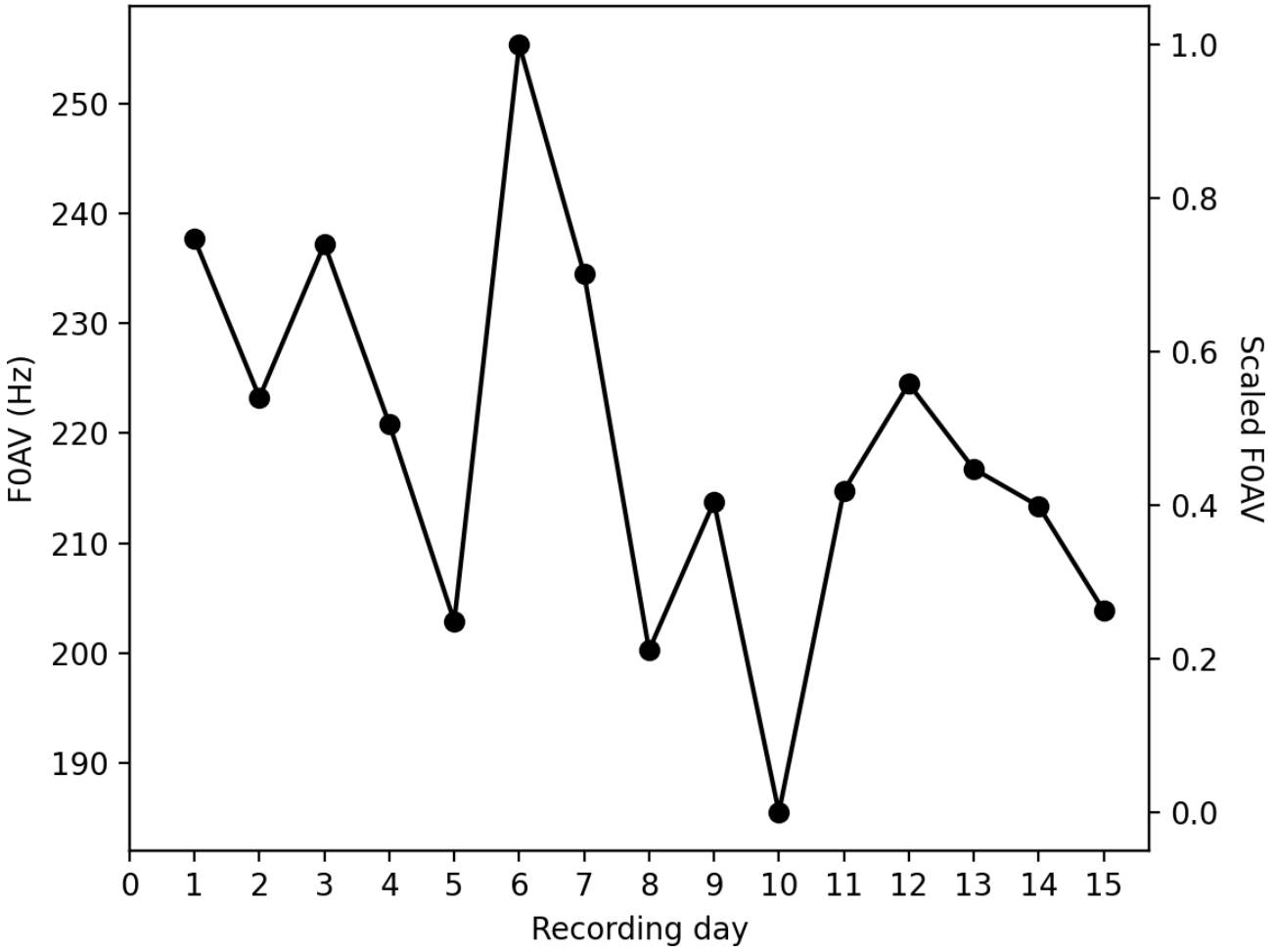
Each voice feature was scaled independently for each participant. Each voice feature was minimum-maximum scaled according to the following formula:

(1)Xsc=X-XminXmax-Xmin

Where X_{sc} is the scaled feature value, X is the original feature value, X_{min} is the minimum daily value of the feature throughout the participant's menstrual cycle, and X_{max} is the maximum daily value of the feature throughout the participant's menstrual cycle. This scaling was conducted such that the maximum scaled feature value throughout the recording period was 1 and the minimum value was 0 (Figure 1). This was repeated for each feature, for each participant.

XSL•FO
RenderX

Figure 1. Generated mean fundamental frequency (F0AV) data to demonstrate the feature scaling process. Data are scaled for each feature within each participant, such that the daily values of a participant's feature fall between 0 and 1.



Changepoint Detection

To assess the change in voice features throughout the menstrual cycle, we used changepoint detection. Changepoint detection algorithms are tools used to identify points in time where the statistical properties of a data sequence change. These changes can signify shifts in mean, variance, distribution, or other characteristics of the data and can be used to detect shifts in patient health status or the effectiveness of treatments over time [18-23]. Changepoint detection algorithms are especially valuable for identifying transitions between distinct stages [23], making them a promising tool for research on physiological changes throughout the menstrual cycle. In this study, we used the *Ruptures* package, a Python library designed for offline changepoint detection, which considers all data points within a time series to identify underlying statistical changes in the data [24]. This unsupervised algorithm does not use prior knowledge of ovulation days and is not influenced by other participants' data, ensuring that the detection process only identifies changes in the voice recording F0 data themselves.

We applied 3 changepoint detection algorithms: dynamic programming, binary segmentation, and bottom-up segmentation. Dynamic programming explores all possible partitions of the time series to identify the optimal segmentation, constrained to detect only a single changepoint. Binary segmentation iteratively identifies the most significant changepoint in each segment but, for this study, was configured to stop after detecting the first changepoint. Bottom-up segmentation begins with many small segments and merges the segments that are most similar until only 1 changepoint remains, meeting the defined stopping criterion.

To quantify changes, we used 3 cost functions: least absolute deviation loss function (L1), least squared deviation loss function (L2), and radial basis function. The least absolute deviation cost function measures absolute differences, making it robust to outliers. The least squared deviation cost function emphasizes larger deviations by using squared differences. The radial basis function cost function captures nonlinear changes by comparing segments with a radial basis kernel. Detailed descriptions of these algorithms and cost functions are available in the *Ruptures* documentation [24].

We were not aiming to identify changes associated with menstruation so we specified the detection of a single changepoint within a window spanning from 9 days before to 9 days after the LH surge. These time points correspond to the average number of days we observed occurring prior to the LH surge following menstruation (9.4, SD 3.1 days) and the postovulatory day that the corpus luteum begins to disintegrate and hormone levels fall [25]. The algorithm was configured with a minimum distance of 3 days between any 2 breakpoints or from the start of the time series, with a jump parameter set

XSL•FO
RenderX

to 1 to ensure that each day was evaluated for potential changepoints of the vocal features between menstrual phases.

Statistical Analysis

To analyze differences between the follicular and luteal phases, 4 generalized linear mixed models were applied for the 4 F0 features. The voice feature was the dependent variable, the menstrual phase (follicular vs luteal) was the independent variable, and participant ID was a random effect to account for intraindividual variability. To avoid data that may occur between the follicular and luteal phases and ensure the clarity of phase-specific effects, days identified as ovulatory (the day of LH surge and the following day) and menstrual days were excluded from the analysis. All other data points were used. The models were estimated using restricted maximum likelihood.

To assess the significance of our changepoint findings against a null hypothesis of random results, we conducted 1-sided 2-proportion *z* tests. We compared the observed proportion of participants with a changepoint occurring before ovulation in the fertile window (days −5 to +1 relative to LH surge) to a random expectation, corresponding to 50% (n=8) of participants with a changepoint occurring in the fertile window. This proportion is based on 7 days in the fertile window of an average of 14 days in the possible prediction window (6 days prior to the LH surge maximum, the LH surge maximum, and 7 days following the LH surge maximum). The alternate hypothesis posited that the observed proportion would be higher than the random expectation due to the hormonal changes associated with the fertile window. All statistical analysis was performed in Python (version 3.12.2; Python Software Foundation) using the *statsmodels* package (*mixedlm* and *proportions_ztest* functions) [26].

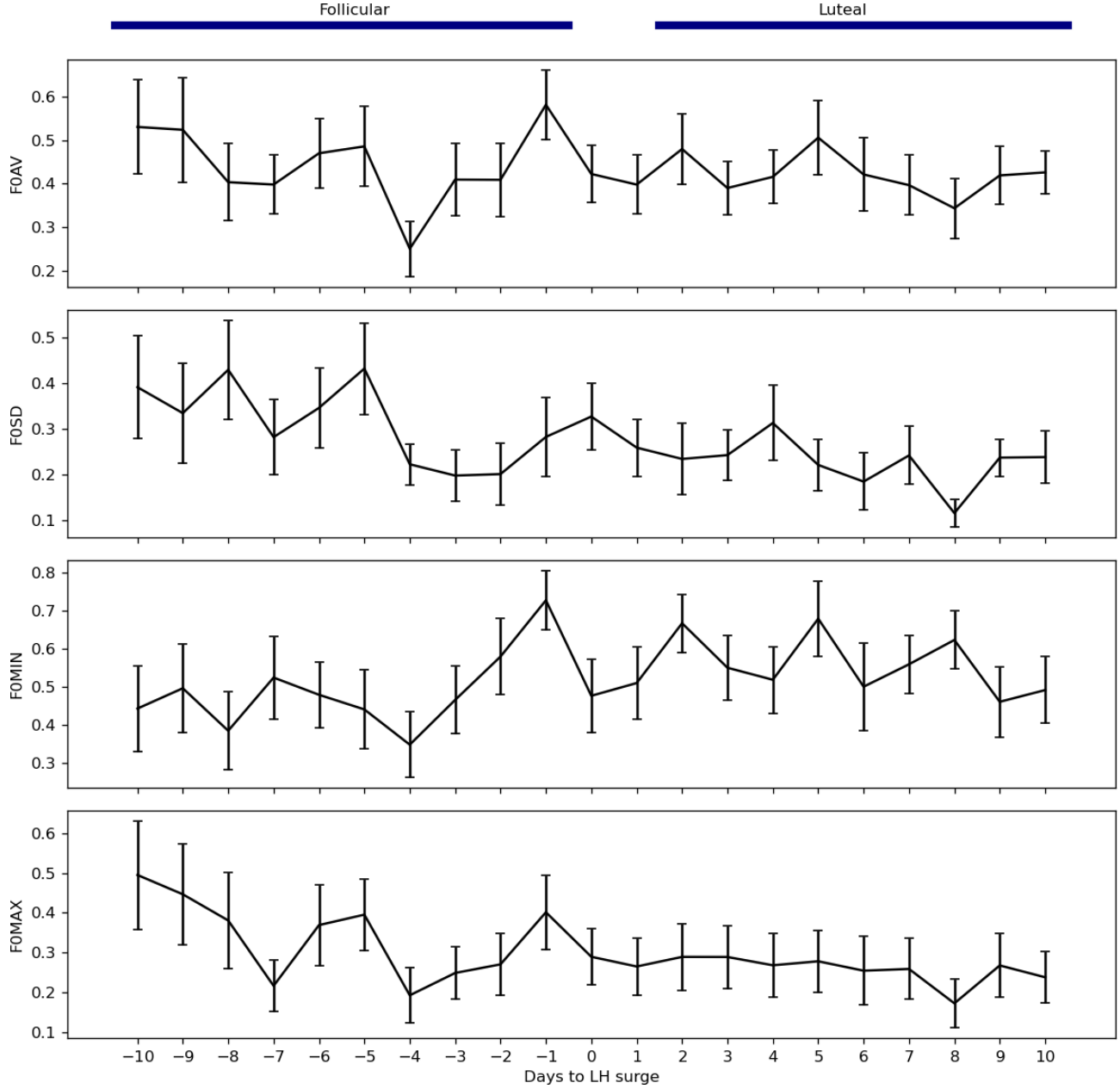
Results

Results 1: F0 Behavior Surrounding LH Surge

To assess vocal dynamics consistent across participants, generalized linear mixed models were fit for each of the F0 features (F0AV, F0SD, F0MIN, and F0MAX). Overall, there was a significant difference in the F0SD and F0MIN between the follicular and luteal recording days for each participant, such that F0SD was, on average, 9.0% (SD 2.9%) lower in the luteal phase compared to the follicular phase (95% CI 3.4% - 14.7%; *P*=.002) and F0MIN was 8.8% (SD 3.6%) higher (95% CI 1.7% - 16.0%; *P*=.01). There were no significant differences between phases in F0AV or F0MAX (*P*=.65 and *P*=.07). Average daily trends of the participant-scaled F0 values are displayed in Figure 2.

XSL•FO
RenderX

Figure 2. Daily patterns of fundamental frequency (**F0**) voice features—F0 average (F0AV), F0 standard deviation (F0SD), F0 5th percentile (F0MIN), and F0 95th percentile (F0MAX)—in the 9 days preceding and following the luteinizing hormone (LH) surge for all participants. Data points represent the average values of these voice features across 16 healthy female participants with regular menstrual cycles. Error bars indicate the SE of the mean.

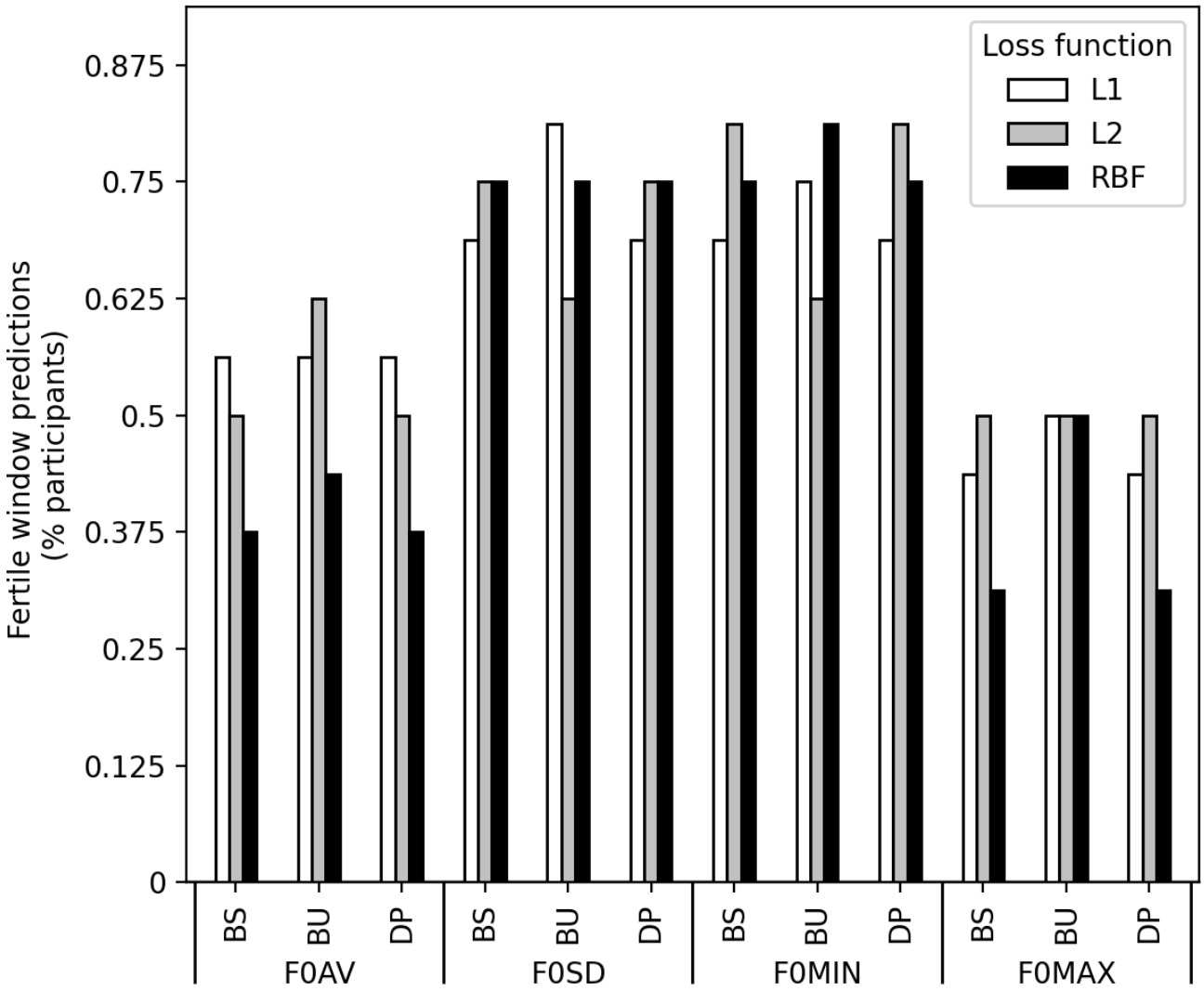


Results 2: F0 Changepoint Detection

F0 feature data for each participant was fed into the changepoint detection algorithms, and the predicted changepoints are displayed in Figure 3. Menstrual phases were not added as information in the algorithms at any point, and any changepoints detected solely reflect changes in the vocal features themselves. Additionally, data from each participant were used independently for changepoint identification, ensuring that data were not shared or mixed between participants. Overall, the average (SD) detected changepoints for all algorithms were 0.20 (4.13) days following ovulation in F0AV, 1.42 (3.52) days prior to ovulation in F0SD, 0.88 (3.27) days prior to ovulation in F0MIN, and 1.31 (4.05) days following ovulation in F0MAX. As the variation in the predicted changepoints was large, we also looked where the predicted changepoints occurred specifically. We observed that up to 81% (n=13) of participants had a predicted changepoint occurring within the fertile window, or the 5 days preceding the LH surge to the 1 day following the LH surge, for both F0SD and F0MIN (P=.03). Meanwhile, a maximum of 63% (n=10; P=.24) and 50% (n=8; P=.50) of participants had changepoints detected in the fertile window for F0AV and F0MAX, respectively (Figure 3).

XSL•FO
RenderX

Figure 3. The percentage of participants with detected changepoints in the fertile window is presented for the mean, SD, 5th percentile, and 95th percentile of the fundamental frequency (F0AV, F0SD, F0MIN, and F0MAX, respectively). Changepoints were identified using 3 different detection algorithms—binary segmentation (BS), bottom-up segmentation (BU), and dynamic programming (DP)—applied to fundamental frequency features extracted from daily smartphone voice recordings of 16 naturally cycling female participants. The fertile window is defined as the 5 days leading up to and including 1 day after the luteinizing hormone surge. L1: least absolute deviation loss function; L2: least squared deviation loss function; RBF: radial basis function.



Discussion

Primary Findings

This study explored how vocal pitch features fluctuate across the menstrual cycle, with a focus on vocal patterns surrounding phase transitions and ovulation. Building on prior research [15], we expanded the analysis by examining a broader set of F0-related features, including minimum and maximum pitch values, using real-world voice recordings collected daily via smartphones. With more frequent recordings, we investigated whether these vocal features varied between the follicular and luteal phases, providing a more comprehensive perspective on daily vocal behavior. We further hypothesized that changepoint detection could pinpoint the transition between these phases, offering insight into the potential for voice-based markers of fertility in future applications.

Physiologically, the differences in voice features between menstrual phases may be attributed to differences in hormone concentrations. In the follicular phase, estrogen levels rise, whereas the luteal phase is characterized by increased progesterone levels. The larynx itself is a sexual organ and can be influenced by reproductive hormones, as estrogen stimulates the mucosal cells. Mucus on the vocal cords has been shown to be similar in consistency and makeup as cervical mucus throughout the menstrual cycle, displaying increased thin, wet mucus in the follicular phase and thick, dry mucus in the luteal phase [27]. A thin, wet mucus layer is necessary for proper vocal cord vibration and range of pitch [28], and decreased vocal range has been associated with thicker, dry laryngeal mucus [27,29]. Our results align well with these depictions, as we observed a decrease in the F0SD and range (observed through an increase in the minimum F0) in the luteal phase. An alternative explanation for vocal changes may reside in more nuanced apps. Fatigue and stress may be exacerbated in the luteal phase [30], and increased fatigue has been associated with increased F0SD [31]. Regardless, changes in mental and hormonal states associated with the changing menstrual phases provide evidence for the observed vocal changes.

XSL•FO
RenderX

We found that individually scaled F0 voice features had distinct differences between the follicular and luteal phases. Similar to what was observed in previous studies, F0SD decreased and F0MIN increased in the luteal phase [14,15]. Meanwhile, the association between the F0AV and the menstrual phase was insignificant. F0AV throughout the menstrual cycle has traditionally reported mixed results, in which some authors found that F0 increased in periods of high fertility [8,32], decreased in high fertility [15,33,34], or did not change [6,9,14,35-39]. This difference may be attributed to different study methodologies, such as recording a vowel instead of a fixed phrase or only recording 1 voice sample per menstrual phase.

Changes in the voice surrounding the fertile period have been hypothesized to be the result of evolutionary processes. More attractive female participants have an increased likelihood of finding a suitable mate, and mating during the fertile window has the highest likelihood of conception [1]. Research has shown that variations in female voice characteristics throughout the menstrual cycle can influence perceived attractiveness. In the high fertility days prior to ovulation, female voices have typically been rated "more attractive" by heterosexual male participants [15,16]. It has been shown that 44% - 55% of the variance in the desirability scores for female participants was explained by their minimum F0 in speech. This may indicate that female voices with lower pitch minima, opposed to a change in the mean pitch, are more desirable [40]. Our findings align nicely with this result, as the 5th percentile of the pitch was an average of 8.8% lower in the follicular phase (a phase corresponding to higher fertility) compared to the luteal phase for all our participants. Furthermore, increased pitch variability, similar to what was seen in the F0SD in the follicular phase, has been associated with increased social attractiveness of the speaker [41,42].

One of the primary distinctions of our work is the employment of a changepoint detection algorithm to identify the day of the switch in vocal features within the menstrual cycle. Although the average predicted changepoints fell within 2 days of ovulation in all participants, there was a large variation in the values that would not permit accurate ovulation prediction. However, when assessing the changepoints in the broader sense of fertility identification, we note that changepoints are identified within the fertile window in a high proportion of participants. The fertile window is associated with an increased likelihood of conception, making it a meaningful target for detection, even if exact ovulation timing is not achieved [1]. In particular, we observed that changepoints are detected within the fertile window in up to 81% (n=13) of participants for F0SD and F0MIN. However, this result was not observed for F0AV and F0MAX.

Offline changepoint detection is a useful method for fertility monitoring, even without the real-time benefits of online detection (in which changepoints can be detected as the data are collected). Current fertility monitoring methods can also require daily use or cycle-long data collection to accurately pinpoint the timing of ovulation. For example, BBT is used as an ovulation confirmation, in which the BBT must be measured every day for at least 1 full menstrual cycle to identify ovulation [1]. After further refinement, offline detection used in a consumer product would have similar confirmatory utility.

Unlike previous studies that had limited voice data collection methodologies, our approach reflects real-world situations by using at-home smartphone recordings, allowing for a more comprehensive assessment of voice parameters throughout the menstrual cycle. This opens up avenues for exploring the feasibility of using voice parameters as indicators of high fertility and ovulation. Moreover, the simplicity and noninvasiveness of smartphone-based recording make it suitable for widespread use in both clinical settings and personal health monitoring. With just a voice and a smartphone app, users do not require any additional devices, highlighting the convenience and accessibility of this method.

Limitations

Our study's insights into vocal pitch changes across the menstrual cycle are limited by a small sample size (n=16 participants), which may not capture the full spectrum of variability across a wider population. Additionally, analyzing recordings from only 1 menstrual cycle per participant neglects potential intraindividual variability across different cycles. Finally, our examination of a limited set of pitch features restricts our understanding of the nuanced relationship between vocal characteristics and menstrual phases, and there may be more information located in other pitch-related features. With further refinement through larger sample sizes, multicycle analyses, and a broader examination of vocal features, we could gain better insights into how vocal pitch correlates with menstrual cycle changes. Importantly, using larger, multicycle datasets could enable the use of online changepoint detection algorithms, which have the advantage of identifying phase transitions in real-time without requiring access to the entire dataset beforehand, or could be used to train a machine learning model for enhanced performance. With these advancements, this method could contribute to the early detection of fertility issues, personalized fertility tracking, and natural family planning.

Conclusions

Our study contributes to the evolving understanding of how menstrual cycle phases impact vocal pitch, underscoring the potential of voice as a noninvasive marker for fertility and ovulation. Using methods that use real-world data captured through smartphone recordings, our study analyzed these data with changepoint detection algorithms to identify transitions between menstrual phases. This approach not only confirms the relationship between vocal pitch variations and the menstrual cycle phase but also demonstrates the feasibility of applying such findings in practical, real-world settings. Our work could have implications for clinical and personal health monitoring, offering potential improvements in fertility tracking and natural family planning methods. Future efforts to integrate more sophisticated machine learning techniques and real-time analysis could further refine the accuracy and applicability of vocal pitch as a tool for women's health management.

XSL•FO
RenderX

Acknowledgments

The authors thank all participants, project members, and researchers at Klick Health for the successful development, implementation, and evaluation of this research. This study was funded internally by Klick Health. The funder played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Data Availability

The datasets generated and analyzed during this study are not publicly available to protect participant privacy. They may be available from the corresponding author on reasonable request, subject to ethics approval and compliance with data protection regulations.

Authors' Contributions

JK designed experiments and performed analyses. JK and JJ developed the method and prepared figures. JK, JJ, JO, and YF wrote the initial draft and edited the manuscript. JK and AT developed a custom mobile app and collected data. JJ and YF supervised the project. All authors reviewed the manuscript and approved it for submission.

Conflicts of Interest

All authors are employed by Klick Health, a technology, media, and research company in the health care sector based in Toronto, Canada. Authors JK and YF currently hold a provisional patent on the use of voice-based technology in the prediction of ovulation (Systems, Devices and Methods for Fertility Analysis Using Voice, United States Provisional Application 63/416,819, filed October 17, 2022). All authors declare no nonfinancial conflicts of interest.

References

1. Su HW, Yi YC, Wei TY, Chang TC, Cheng CM. Detection of ovulation, a review of currently available methods. Bioeng Transl Med 2017 Sep;2(3):238-246. [doi: 10.1002/btm2.10058] [Medline: 29313033]
2. Johnson S, Marriott L, Zinaman M. Can apps and calendar methods predict ovulation with accuracy? Curr Med Res Opin 2018 Sep;34(9):1587-1594. [doi: 10.1080/03007995.2018.1475348] [Medline: 29749274]
3. Haselton MG, Gildersleeve K. Human ovulation cues. Curr Opin Psychol 2016 Feb;7:120-125. [doi: 10.1016/j.copsyc.2015.08.020]
4. Haselton MG, Gildersleeve K. Can men detect ovulation? Curr Dir Psychol Sci 2011 Apr;20(2):87-92. [doi: 10.1177/0963721411402668]
5. Zamponi V, Mazzilli R, Mazzilli F, Fantini M. Effect of sex hormones on human voice physiology: from childhood to senescence. Hormones (Athens) 2021 Dec;20(4):691-696. [doi: 10.1007/s42000-021-00298-y] [Medline: 34046877]
6. Amir O, Biron-Shental T. The impact of hormonal fluctuations on female vocal folds. Curr Opin Otolaryngol Head Neck Surg 2004 Jun;12(3):180-184. [doi: 10.1097/01.moo.0000120304.58882.94] [Medline: 15167026]
7. Silverman EM, Zimmer CH. Effect of the menstrual cycle on voice quality. Arch Otolaryngol Head Neck Surg 1978 Jan 1;104(1):7-10. [doi: 10.1001/archotol.1978.00790010011003]
8. Bryant GA, Haselton MG. Vocal cues of ovulation in human females. Biol Lett 2009 Feb 23;5(1):12-15. [doi: 10.1098/rsbl.2008.0507]
9. Çelik Ö, Çelik A, Ateşpare A, et al. Voice and speech changes in various phases of menstrual cycle. J Voice 2013 Sep;27(5):622-626. [doi: 10.1016/j.jvoice.2013.02.006] [Medline: 23497797]
10. Shoup-Knox ML, Ostrander GM, Reimann GE, Pipitone RN. Fertility-dependent acoustic variation in women's voices previously shown to affect listener physiology and perception. Evol Psychol 2019;17(2):1474704919843103. [doi: 10.1177/1474704919843103] [Medline: 31023082]
11. Pavela Banai I, Burriss RP, Šimić N. Voice changes across the menstrual cycle in response to masculinized and feminized man and woman. Adapt Human Behav Physiol 2022 Jun;8(2):238-262. [doi: 10.1007/s40750-022-00190-y]
12. Narasimhan SV, Pooja M. Vocal changes in different phases of menstrual cycle: an evidence from the acoustic, cepstral, and spectral analysis. J Indian Speech Lang Hear Assoc 2022;36(1):1. [doi: 10.4103/jisha.jisha_5_22]
13. Santos SMS. Is there any correlation between voice and human ovulation? [Doctoral Dissertation]. : Universidade da Beira Interior, ProQuest; 2018.
14. Pavela Banai I. Voice in different phases of menstrual cycle among naturally cycling women and users of hormonal contraceptives. PLoS One 2017;12(8):e0183462. [doi: 10.1371/journal.pone.0183462] [Medline: 28829842]
15. Fischer J, Semple S, Fickenscher G, et al. Do women's voices provide cues of the likelihood of ovulation? The importance of sampling regime. PLoS One 2011;6(9):e24490. [doi: 10.1371/journal.pone.0024490] [Medline: 21957453]
16. Pipitone RN, Gallup GG Jr. Women's voice attractiveness varies across the menstrual cycle. Evol Hum Behav 2008 Jul;29(4):268-274. [doi: 10.1016/j.evolhumbehav.2008.02.001]

XSL•FO
RenderX

17. McFee B, McVicar M, Faronbi D, et al. librosa. Zenodo. URL: https://zenodo.org/records/11192913 [accessed 2024-11-30]
18. Yang P, Dumont G, Ansermino JM. Adaptive change detection in heart rate trend monitoring in anesthetized children. IEEE Trans Biomed Eng 2006 Nov;53(11):2211-2219. [doi: 10.1109/TBME.2006.877107] [Medline: 17073326]
19. Malladi R, Kalamangalam GP, Aazhang B. Online bayesian change point detection algorithms for segmentation of epileptic activity. Presented at: Asilomar Conference on Signals, Systems and Computers; Nov 3-6, 2013; Pacific Grove, CA, United States p. 1833-1837. [doi: 10.1109/ACSSC.2013.6810619]
20. Staudacher M, Telser S, Amann A, Hinterhuber H, Ritsch-Marte M. A new method for change-point detection developed for on-line analysis of the heart beat variability during sleep. Physica A Stat Mech Appl 2005 Apr;349(3-4):582-596. [doi: 10.1016/j.physa.2004.10.026]
21. Bosc M, Heitz F, Armspach JP, Namer I, Gounot D, Rumbach L. Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution. Neuroimage 2003 Oct;20(2):643-656. [doi: 10.1016/S1053-8119(03)00406-3] [Medline: 14568441]
22. Kass-Hout TA, Xu Z, McMurray P, et al. Application of change point analysis to daily influenza-like illness emergency department visits. J Am Med Inform Assoc 2012;19(6):1075-1081. [doi: 10.1136/amiajnl-2011-000793] [Medline: 22759619]
23. Aminikhanghahi S, Cook DJ. A survey of methods for time series change point detection. Knowl Inf Syst 2017 May;51(2):339-367. [doi: 10.1007/s10115-016-0987-z] [Medline: 28603327]
24. Truong C, Oudre L, Vayatis N. Selective review of offline change point detection methods. Signal Processing 2020 Feb;167:107299. [doi: 10.1016/j.sigpro.2019.107299]
25. McCracken JA, Custer EE, Lamsa JC. Luteolysis: a neuroendocrine-mediated event. Physiol Rev 1999 Apr;79(2):263-323. [doi: 10.1152/physrev.1999.79.2.263] [Medline: 10221982]
26. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with python. Presented at: Proceedings of the 9th Python in Science Conference; Jun 28 to Jul 3, 2010; Austin, TX, United States. [doi: 10.25080/Majora-92bf1922-011]
27. Abitbol J, Abitbol P, Abitbol B. Sex hormones and the female voice. J Voice 1999 Sep;13(3):424-446. [doi: 10.1016/s0892-1997(99)80048-4] [Medline: 10498059]
28. Levendoski EE, Leydon C, Thibeault SL. Vocal fold epithelial barrier in health and injury: a research review. J Speech Lang Hear Res 2014 Oct;57(5):1679-1691. [doi: 10.1044/2014_JSLHR-S-13-0283] [Medline: 24686981]
29. Bonilha HS, White L, Kuckhahn K, Gerlach TT, Deliyski DD. Vocal fold mucus aggregation in persons with voice disorders. J Commun Disord 2012;45(4):304-311. [doi: 10.1016/j.jcomdis.2012.03.001] [Medline: 22510352]
30. Li SH, Lloyd AR, Graham BM. Physical and mental fatigue across the menstrual cycle in women with and without generalised anxiety disorder. Horm Behav 2020 Feb;118:104667. [doi: 10.1016/j.yhbeh.2019.104667] [Medline: 31899259]
31. Vogel AP, Fletcher J, Maruff P. Acoustic analysis of the effects of sustained wakefulness on speech. J Acoust Soc Am 2010 Dec;128(6):3747-3756. [doi: 10.1121/1.3506349] [Medline: 21218906]
32. Tatar EC, Sahin M, Demiral D, et al. Normative values of voice analysis parameters with respect to menstrual cycle in healthy adult Turkish women. J Voice 2016 May;30(3):322-328. [doi: 10.1016/j.jvoice.2015.04.014] [Medline: 26003887]
33. Amir O, Kishon-Rabin L, Muchnik C. The effect of oral contraceptives on voice: preliminary observations. J Voice 2002 Jun;16(2):267-273. [doi: 10.1016/s0892-1997(02)00096-6] [Medline: 12150379]
34. Karthikeyan S, Locke JL. Men's evaluation of women's speech in a simulated dating context: effects of female fertility on vocal pitch and attractiveness. Evol Behav Sci 2015;9(1):55-67. [doi: 10.1037/ebs0000014]
35. Barnes L, Latman N. Acoustic measure of hormone affect on female voice during menstruation. Int J Humanit Soc Sci 2011;1(3):5-10 [FREE Full text]
36. Chae SW, Choi G, Kang HJ, Choi JO, Jin SM. Clinical analysis of voice change as a parameter of premenstrual syndrome. J Voice 2001 Jun;15(2):278-283. [doi: 10.1016/S0892-1997(01)00028-5] [Medline: 11411481]
37. Figueiredo LCD, Gonçalves MIR, Pontes A, Pontes P. Vocal behavior during menstrual cycle: perceptual-auditory, acoustic and self-perception analysis. Rev Bras Otorrinolaringol 2004;70:331-339. [doi: 10.1590/S0034-72992004000300008]
38. Meurer EM, Garcez V, von Eye Corleta H, Capp E. Menstrual cycle influences on voice and speech in adolescent females. J Voice 2009 Jan;23(1):109-113. [doi: 10.1016/j.jvoice.2007.03.001] [Medline: 17981011]
39. Raj A, Gupta B, Chowdhury A, Chadha S. A study of voice changes in various phases of menstrual cycle and in postmenopausal women. J Voice 2010 May;24(3):363-368. [doi: 10.1016/j.jvoice.2008.10.005]
40. Pisanski K, Oleszkiewicz A, Plachetka J, Gmiterek M, Reby D. Voice pitch modulation in human mate choice. Proc Biol Sci 2018 Dec 19;285(1893):20181634. [doi: 10.1098/rspb.2018.1634] [Medline: 30963886]
41. Ray GB, Ray EB, Zahn CJ. Speech behavior and social evaluation: an examination of medical messages. Commun Q 1991 Mar;39(2):119-129. [doi: 10.1080/01463379109369790]
42. Zuckerman M, Miyake K. The attractive voice: what makes it so? J Nonverbal Behav 1993;17(2):119-135. [doi: 10.1007/BF01001960]

Abbreviations

BBT: basal body temperature
F0: fundamental frequency

XSL•FO
RenderX

F0AV: mean fundamental frequency
F0MAX: 95th percentile of the fundamental frequency
F0MIN: 5th percentile of the fundamental frequency
F0SD: fundamental frequency SD
LH: luteinizing hormone

Edited by A Mavragani; submitted 16.08.24; peer-reviewed by J Grieger, V Martin; revised version received 22.10.24; accepted 13.11.24; published 09.01.25.

Please cite as:
Kaufman J, Jeon J, Oreskovic J, Thommandram A, Fossat Y
Longitudinal Changes in Pitch-Related Acoustic Characteristics of the Voice Throughout the Menstrual Cycle: Observational Study
JMIR Form Res 2025;9:e65448
URL: https://formative.jmir.org/2025/1/e65448
doi:10.2196/65448

© Jaycee Kaufman, Jouhyun Jeon, Jessica Oreskovic, Anirudh Thommandram, Yan Fossat. Originally published in JMIR Formative Research (https://formative.jmir.org), 9.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Postpartum Remote Health Coaching Intervention for Individuals With a Hypertensive Disorder of Pregnancy: Proof-of-Concept Study

Jaclyn D Borrowman[1,2], PhD; Lucas J Carr[2], PhD; Gary L Pierce[2], PhD; William T Story[3], MPH, PhD; Bethany Barone Gibbs[4], PhD; Kara M Whitaker[2,5], MPH, PhD

[1]Department of Preventative Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[2]Department of Health and Human Physiology, University of Iowa, Iowa City, IA, United States

[3]Department of Community and Behavioral Health, University of Iowa, Iowa City, IA, United States

[4]Department of Epidemiology and Biostatistics, West Virginia University, Morgantown, WV, United States

[5]Department of Epidemiology, University of Iowa, Iowa City, IA, United States

Corresponding Author:
Jaclyn D Borrowman, PhD
Department of Preventative Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Abstract

Background: Cardiovascular disease (CVD) is the leading cause of death among women in America. Hypertensive disorders of pregnancy (HDP) negatively impact acute and long-term cardiovascular health, with approximately 16% of all pregnancies affected. With CVD 2 - 4 times more likely after HDP compared to normotensive pregnancies, effective interventions to promote cardiovascular health are imperative.

Objective: With postpartum physical activity (PA) interventions after HDP as an underexplored preventative strategy, we aimed in this study to assess (1) the feasibility and acceptability of a remotely delivered PA intervention for individuals with HDP 3 - 6 months postpartum and (2) changes in average steps per day, skills related to PA behavior, and postpartum blood pressure (BP).

Methods: A remotely delivered 14-week health coaching intervention was designed based on prior formative work. The health coaching intervention called the Hypertensive Disorders of Pregnancy Postpartum Exercise (HyPE) intervention was tested for feasibility and acceptability with a single-arm proof-of-concept study design. A total of 19 women who were 3 - 6 months postpartum HDP; currently inactive; 18 years of age or older; resided in Iowa; and without diabetes, kidney disease, and CVD were enrolled. Feasibility was assessed by the number of sessions attended and acceptability by self-reported satisfaction with the program. Changes in steps achieved per day were measured with an activPAL4 micro, PA behavior skills via validated surveys online, and BP was assessed remotely with a research-grade Omron Series 5 (Omron Corporation) BP monitor.

Results: Participants at enrollment were on average 30.3 years of age, 4.1 months postpartum, self-identified as non-Hispanic White (14/17, 82%), in a committed relationship (16/17, 94%), and had a bachelor's degree (9/17, 53%). A total of 140 of 152 possible health coaching sessions were attended by those who started the intervention (n=19, 92%). Intervention completers (n=17) indicated they were satisfied with the program (n=17, 100%) and would recommend it to others (n=17, 100%). No significant changes in activPAL measured steps were observed from pre- to posttesting (mean 138.40, SD 129.40 steps/day; P=.75). Significant improvements were observed in PA behavior skills including planning (mean 5.35, SD 4.97 vs mean 15.06, SD 3.09; P<.001) and monitoring of PA levels (mean 7.29, SD 3.44 vs mean 13.00, SD 2.45; P<.001). No significant decreases were observed for systolic (mean –1.28, SD 3.59 mm Hg; Hedges g=–0.26; P=.16) and diastolic BP (mean –1.80, SD 5.03 mm Hg; Hedges g=–0.44; P=.12).

Conclusions: While PA behaviors did not change, the intervention was found to be feasible and acceptable among this sample of at-risk women. After additional refinement, the intervention should be retested among a larger, more diverse, and less physically active sample.

Trial Registration: ClinicalTrials.gov NCT06019715; https://clinicaltrials.gov/study/NCT06019715

(JMIR Form Res 2025;9:e65611) doi:10.2196/65611

KEYWORDS

cardiovascular disease; postpartum; hypertensive disorders of pregnancy; intervention; physical activity intervention; proof-of-concept; cardiovascular health; CVD risk; cardiovascular disease risk; feasibility; acceptability; health coaching; women's health; postnatal

XSL•FO
RenderX

Introduction

Cardiovascular disease (CVD) is the leading cause of death among women in America; with 1 in 4 deaths being preventable [1]. Hypertensive disorders of pregnancy (HDP) are associated with a 2‐4 times greater risk of developing CVD later in life [1,2]. Despite this increased risk for CVD and with 16% of all pregnancies affected, few studies have explored potential interventions for this population after delivery [3]. Consequently, it is imperative that preventative strategies are developed and started as early as possible after delivery to mitigate the risk of future CVD and improve maternal morbidity and mortality.

Cardiovascular changes during pregnancy act as a natural stress test and may uncover an individual's potential risk of future CVD [4]. Specifically, individuals with HDP are at an immediately increased risk for hypertension (HTN) after delivery [4]. Long-term cardiometabolic health is also impacted by HDP, with an increased risk of future CVD, morbidity, and mortality compared to normotensive pregnancies [5]. The odds of developing HTN at 1 year postpartum are 5 to 7 times greater in hypertensive pregnancies compared to normotensive pregnancies [2,6]. The 5-year likelihood of developing HTN for pregnancies complicated by HDP is 7.1 times greater than for normotensive pregnancies [4]. Even in those not diagnosed with HTN, average blood pressure (BP) is greater in HDP women than in women with normotensive pregnancies.

Low physical activity (PA) is a known modifiable risk factor for CVD, and increasing PA has the potential to improve BP [7]. PA, specifically aerobic exercise, has beneficial effects on mean, systolic BP (SBP), and diastolic BP (DBP) in adults [8,9]. This effect has been observed in pregnant populations as well, with the greatest improvements seen among inactive individuals [10,11]. Despite the known benefits of PA, recent studies suggest moderate to vigorous PA decreases after pregnancy compared to during pregnancy activity levels [12]. The perinatal period, therefore, presents a critical opportunity for behavior change [13-15]. Interventions promoting PA in the postpartum period after HDP are an underexplored primary prevention strategy warranting further investigation.

The main objective of our study was to assess the feasibility and acceptability of a remotely delivered PA intervention for individuals with HDP during the postpartum period. We hypothesized the intervention would be highly feasible and acceptable. Our secondary objectives included assessing changes in average steps per day, skills related to PA behavior, and postpartum BP. We hypothesized that we would observe significant increases in steps, improvements in PA skills, and decreases in postpartum SBP and DBP.

Methods

Study Design

A remotely delivered PA intervention for individuals with HDP called the Hypertensive Disorders of Pregnancy Postpartum Exercise (HyPE) intervention, was tested with a single-arm quasi-experimental study design. The Obesity-Related Behavioral Intervention Trials (ORBIT) model was the conceptual framework that was leveraged in the development and subsequent testing of the HyPE intervention [16]. The ORBIT model guides the development of behavioral treatments to prevent or manage chronic disease. This model describes a flexible and iterative process that includes prespecified clinically significant milestones necessary for forward movement through the developmental pathway.

Phase I of the ORBIT model was previously completed by our group, where the design of this intervention was informed by an assessment of the PA determinants, needs, and desired components of a PA intervention of postpartum individuals after HDP [17]. Given phase I was completed successfully, our group subsequently developed the HyPE intervention for proof-of-concept testing (phase IIa of the ORBIT model). Study methods appropriate at this stage include a quasi-experimental, treatment-only design to determine whether the intervention is feasible, acceptable, and can achieve a clinically significant signal on the behavioral risk factor of interest (ie, PA).

Ethical Considerations

All methods and procedures were approved by an institutional review board prior to the start of the study (IRB #202302470; ClinicalTrials: NCT06019715). Potential participants had a call with a research staff member to review study details, ask questions, and provide informed consent if still interested in enrolling. All data were deidentified and secured. For compensation, participants were provided a Fitbit Inspire 2 (Google) and Omron Series 5 BP monitor. Participants received a US $20 Amazon gift card after completion of half of the study visits and a US $25 Amazon gift card after completion of the entire study.

Recruitment

Recruitment of currently inactive individuals (defined in this study as achieving <9000 steps on average per day) occurred via mass emails to faculty and staff members of a large midwestern university, targeted emails to participants in the Pregnancy 24/7 Cohort Study, and postpartum medical record searches of individuals who received care at local hospitals and clinics. If participants were interested in the study, the target email contained information, as well as a link to a REDcap (Vanderbilt University) survey for initial screening. If eligible, interested individuals were contacted via phone at their preferred time by a research staff member to describe the study in further detail and answer questions. Subsequently, potential participants were provided the opportunity for participants to provide informed consent if still interested in the study.

Following informed consent, participants started baseline screen procedures involving the assessment of their current PA levels. The threshold of inactivity was set at <9000 steps per day based on a prior study conducted by our group and concerns regarding recruitment within a restricted timeframe [18]. Screening of baseline PA levels was conducted with an activPAL4 micro prior to participation in the first study visit.

Eligibility

To be included, participants were 3‐6 months postpartum after a pregnancy complicated by HDP; 18 years of age or older; able

XSL•FO
RenderX

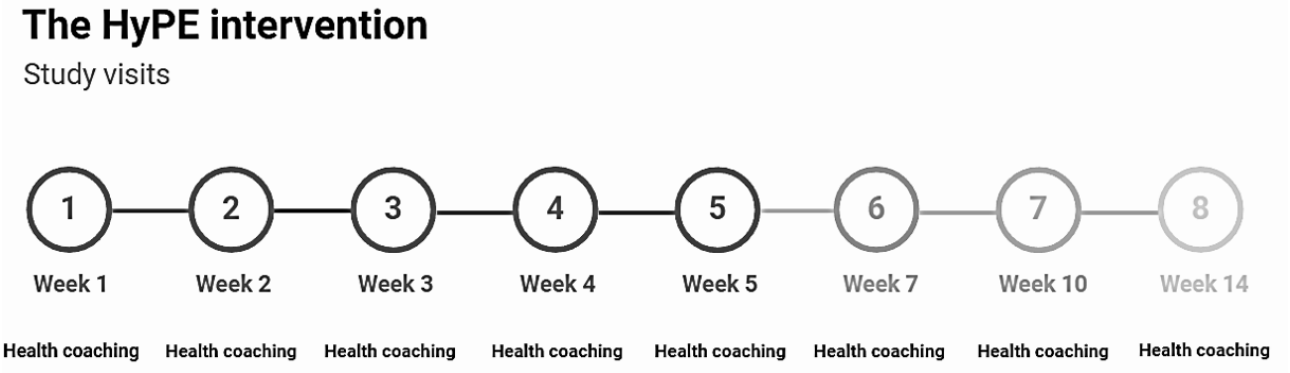
to speak, read, and write English; lived in Iowa; and owned a smartphone. Exclusions included current enrollment in a PA intervention; achieving >9000 steps at pretesting; inability to walk half block or 2 flights of stairs; physician's recommendation to limit PA; and current diagnosis and treatment of diabetes, kidney disease, and CVD. All self-report diagnoses of HDP were verified by medical records.

Intervention

The overall goal of the intervention was to increase the average number of steps participants achieved per day, with the specific objective to increase steps by approximately 10% between each session. The intervention consisted of a total of 8 online health coaching sessions, with sessions 1 - 5 occurring weekly (ie, the intensive intervention) and subsequent visits 6 - 8 occurring with reduced frequency (Figure 1). Participants were provided a Fitbit Inspire 2 to track their progress toward the step goals set with their health coach. Formative work conducted by our group informed the creation of tailored educational postpartum PA intervention, including the use of a health coach, remote delivery approach, selected frequency of study visits, and topics offered to study participants (eg, breastfeeding and PA).

Health coaching was informed by motivational interviewing techniques and conducted by a trained interventionist (principle investigator: JDB). This approach was developed to promote behavioral change and work through ambivalence by supporting an individual's commitment to change using "change talk" (ie, eliciting the client to express desires, ability, reasons, and need for a change of behavior) [19,20].

Figure 1. The Hypertensive Disorders of Pregnancy Postpartum Exercise (HyPE) intervention health coaching study timeline. The HyPE intervention was a 14-week health coaching intervention for postpartum individuals after having a pregnancy complicated by a hypertensive disorder. This intervention included 8 total health coaching session aiming to promote physical activity.



Measures

REDCap was used to build and manage survey assessments [21]. At pre- and posttesting, the following information was collected.

Demographics

Questions assessed the individual's age, race, ethnicity, marital status, education, annual household income, insurance coverage, parity, and employment status. Those currently employed were asked questions assessing job type, hours worked per week, if parental leave was provided, and weeks of parental leave provided.

PA Identity

PA identity was assessed with the validated exercise self-identity scale [22]. The 9-item survey assessed various components for how strongly individuals felt exercise was an important component of self. Individuals rated each item on a 7-point Likert scale from "strongly disagree" (0) to "strongly agree" (6) which were summed for a total score (higher scores indicated exercise was a greater component of self).

PA Planning and Monitoring

PA planning and self-monitoring behaviors were assessed with the validated 8-item self-monitoring and 6-item action planning questionnaires [23]. Questions were assessed on a 4-point Likert scale from "completely disagree (0)" to "totally agree (3)." Scores were calculated by summing responses; higher scores indicated greater self-monitoring and planning of PA.

Program Evaluation

At the conclusion of the intervention, participants responded to a series of closed-ended and open-ended questions assessing their acceptance of the overall program and different components of the intervention including health coaching sessions, perception of PA changes, quality of life, usefulness of tailored intervention materials, the Fitbit, activPAL4 micro use, and home BP monitoring with a 5-point Likert scale from "strongly disagree" (0) to "strongly agree" (4). These questions were adapted from a previously tested intervention [18].

PA Behavior

Device-based measurement of PA was conducted with both the activPAL4 micro and the Fitbit Inspire 2. ActivPAL measurement gave research grade assessments of PA at pre- and posttesting while the Fitbit allowed for steps measurement to occur at the end of the intensive intervention (week 5). The activPAL4 micro assessed changes in PA between pre- and posttesting and data from the Fitbit Inspire 2 were used to assess changes in steps during the intervention. The week prior to the first and last health coaching sessions, participants were instructed to wear an activPAL4 micro for 9 consecutive days (2 partial days and 7 full days) that they received via mail along with instructions for the device. They were also provided an activity log to self-report sleep, naps, and nonwear times. The activPAL4 micro was wrapped with retention tape and then

XSL•FO
RenderX

adhered to the anterior thigh with transparent dressing (3M Tegaderm Film) for 24 hours a day, to be removed only when swimming to prevent monitor loss.

For assessment of the preliminary efficacy of the intervention, our primary outcome was a change in steps per day measured with the activPAL4 micro. Data collected from the activPAL4 micro were exported as events using PAL technologies software using the VANE algorithm (PAL Technologies). Participants' self-reported sleep and nonwear times in their activity logs were removed using R (R Core Team) [24]. The processing procedures for this approach are available via the pregnancy247 package located on GitHub [25]. This code generates a figure depicting sitting, standing, or stepping activities, as well as sleep or nonwear periods. Visual inspection of these graphs was conducted to verify the accuracy of reported sleep onset, offset, and naps. Movement behaviors during waking hours on valid days were summarized using the activPAL processing package created by Lynden et al [26]. Total wear time and valid days were defined by at least 5 days of total wear with greater than 10 hours of wear time each day [27].

For the Fitbit Inspire 2, participants were instructed to wear the device every day, only removing it at night to charge. At the end of study participation, all step data collected from the Fitbit were retrieved by the study staff for data analysis. A valid day was defined as achieving >1000 steps [28].

Blood Pressure

The protocol for this study's BP assessment followed the American Heart Association's monitoring home BP recommendations [26,29]. Participants were provided the validated Omron Series 5 BP monitor at the start of the study [30,31], and received verbal and visual instructions on how to take an accurate reading during visit 1, visit 4, and visit 7. Participants were asked to measure their BP at pretesting, at the end of the intensive intervention (between visits 4 and 5), and a week prior to the final study visit which was at the same time as PA assessments that occurred with both the activPAL4 micro and Fitbit. During each of these assessments, participants were asked to measure their BP twice in the morning and twice at night for 7 consecutive days (ie, 28 total readings). These readings were synced with the Omron Connect app on the participant's cellular device and shared with the research staff via email. Assessment of appropriate BP technique occurred remotely via Zoom (Zoom Communications) at each study visit with corrections of technique provided when needed.

Statistical Analysis

All statistical analyses were determined a priori for main objectives and SAS (version 9.4; SAS Institute) was used for analysis. Demographic characteristics were assessed using frequencies and averages. Feasibility was measured by the number of completed sessions out of the total sessions possible. Acceptability was assessed by participants' self-reported satisfaction and likelihood of recommending the program to others. High feasibility was defined as 75% of all sessions attended. High acceptability was defined by at least 75% of participants responding that they were either satisfied or very satisfied with the program.

To assess changes in PA and BP between timepoints, data from the activPAL4 micro, Fitbit, and Omron series 5 were analyzed using paired *t* tests, Wilcoxon signed-rank, and Fisher exact tests. To estimate the effect size, Hedges *g* statistics were used. Thresholds for small, medium, and large effect sizes were set at 0.2, 0.5, and 0.8, respectively [32]. Linear regression models were used to assess the association of change in PA (average steps per day) and change in BP across the intervention [33]. SBP and DBP were treated as continuous variables and included in separate models. Covariates included in the final model were the participant's age, annual household income, activPAL wear time, and pretest steps.

Although a proof-of-concept study, a sample size calculation was conducted with a 0.50 effect size, based on preliminary evidence concerning peripartum PA interventions, using 80% power and at an α level of .05. From this, it was estimated that 15 individuals would be needed to detect significant changes in PA. Allowing for 25% drop out, the goal was to recruit 20 participants.

Results

Recruitment

Of the 94 individuals who completed the eligibility survey, 38 were found eligible to participate in the intervention. Those who were ineligible were excluded most often due to lack of HDP diagnosis (n=13), being greater than 6 months postpartum (n=24), or self-reporting current CVD or taking antihypertensive medications (n=17). Of the 38 eligible, 23 consented to the intervention. Only 19 of the 23 consented to enroll in the study with 3 individuals no longer interested and 1 excluded due to activity (achieving >9000 steps/day). Two participants dropped out (n=1 moved out of the area and n=1 lack of time; attrition=11%) during the intervention, leaving 17 participants who successfully completed the intervention.

Participants

Participants were on average 30.3 years of age (Table 1), and more often self-identified as White (14/17, 82%), were in a committed relationship (16/17, 94%), had earned at least a bachelor's degree (9/17, 54%), and had an annual household income over US $100,000 (9/17, 54%). At the start of participation in the study, no individuals were on parental leave and were on an average of 4.1 (SD 0.81) months postpartum.

XSL•FO
RenderX

Table . Demographic characteristics of Hypertensive Disorders of Pregnancy Postpartum Exercise intervention participants[a].

Participant characteristics (N=17)	Values
Age (years), mean (SD)	30.3 (3.6)
Race, n (%)	
White	14 (82)
Hispanic or Latina	1 (6)
Black	2 (12)
Marital status, n (%)	
Married or in a committed relationship	16 (94)
Single	1 (6)
Education, n (%)	
High school graduate or equivalent	4 (24)
Associate's degree	3 (18)
Bachelor's degree	4 (24)
Master's or doctoral degree	5 (30)
Annual household income (US $), n (%)	
<$25,000	2 (12)
$25,000 - $99,999	6 (35)
$100,000 - $149,999	6 (35)
>$150,000	3 (18)
Earned wages outside of home, n (%)	14 (82)
Parity, n (%)	
1 child	6 (35)
2 children	7 (41)
3 or more children	4 (24)
Hypertensive disorder of pregnancy, n (%)	
Gestational hypertension	11 (65)
Preeclampsia	6 (35)
Months postpartum at start of intervention, mean (SD)	4.1 (0.81)

[a]Prior to the start of the intervention, participants completed an online survey to self- report their demographic characteristics. Continuous metrics were reported as mean (SD) and categorical variables as frequencies, n (%).

Feasibility

Of all the participants who completed pretesting (n=19), 140 of the 152 (92%) sessions were completed. Of participants who completed the intervention (N=17), all 136 (100%) possible health coaching sessions were completed. Completers wore the Fitbit on 96% of all possible days. Almost all BP measures were obtained from completers with 90% of the pretest, 85% of week 5, and 85% of the posttest measures taken.

Acceptability

Participants evaluated the intervention and components of the program including health coaching sessions, observed changes in their PA behaviors, changes in health and well-being, use of information sheets, Fitbit, activPAL, and BP monitors (Table 2). All completers (N=17) indicated they were satisfied with the program (100%) and would recommend this program to others (100%).

XSL•FO
RenderX

Table . Program evaluation of the Hypertensive Disorders of Pregnancy Postpartum Exercise intervention (N=17)[a].

Item	Agree, n (%)	Neutral, n (%)	Disagree, n (%)
Health coaching			
The health coaching sessions helped me increase my physical activity	17 (100)	0 (0)	0 (0)
The health coaching sessions helped me decrease the amount of time I spend sitting	16 (94)	1 (6)	0 (0)
I felt supported by my health coach	17 (100)	0 (0)	0 (0)
My health coach was able to answer my questions	17 (100)	0 (0)	0 (0)
My health coach helped me understand how to perform physical activity safely	17 (100)	0 (0)	0 (0)
My health coach helped me create solutions to problems I faced	16 (94)	1 (6)	0 (0)
I felt like it was easy to talk to my health coach	17 (100)	0 (0)	0 (0)
I had enough time for my health coaching sessions	16 (94)	1 (6)	0 (0)
I wish I had more sessions with my health coach	8 (47)	8 (47)	1 (6)
I wish I had fewer sessions with my health coach	11 (65)	6 (35)	0 (0)
The length of time between sessions (weekly, every other week, and monthly) was appropriate	17 (100)	0 (0)	0 (0)
I enjoyed the ability to have health coaching sessions virtually from the comfort of my own home	17 (100)	0 (0)	0 (0)
Movement behaviors			
I feel like I have become more physically active	17 (100)	0 (0)	0 (0)
I know how to be physically active	17 (100)	0 (0)	0 (0)
I sit less during the day	17 (100)	0 (0)	0 (0)
I think that physical activity is an important part of living a healthy lifestyle	17 (100)	0 (0)	0 (0)
I am confident in my ability to stay physically active	17 (100)	0 (0)	0 (0)
Health and well-being			
I feel healthier	15 (88)	2 (12)	0 (0)
I have more energy during the day	13 (77)	2 (12)	2 (12)
I feel supported to live an active lifestyle	15 (88)	2 (12)	0(0)
I have noticed an improvement in my sleep	9 (53)	7 (41)	1 (6)
I have noticed an improvement in my mood	15 (88)	2 (12)	0 (0)

XSL•FO
RenderX

Item	Agree, n (%)	Neutral, n (%)	Disagree, n (%)
Information sheets			
I referenced the information sheets during the intervention	14 (82)	1 (6)	2 (12)
The information sheets helped me increase my physical activity	10 (59)	4 (24)	3 (18)
The information sheets were difficult to understand	3 (18)	7 (41)	7 (41)
Fitbit wear			
It was easy to set up my Fitbit (downloading app and linking device)	17 (100)	0 (0)	0 (0)
My Fitbit was easy to use	16 (94)	1 (6)	0 (0)
I checked the step count on my Fitbit daily	15 (88)	1 (6)	1 (5.9)
My Fitbit was uncomfortable to wear	4 (24)	2 (12)	11 (65)
The Fitbit app was easy to use	17 (100)	0(0)	0(0)
I checked my Fitbit app daily	9 (53)	5 (29)	3 (18)
The instructions given to me on how to use my Fitbit were helpful	15 (88)	1 (6)	0 (0)
It was difficult to remember to wear my Fitbit daily	8 (47)	9 (53)	0 (0)
It was difficult to remember to sync my Fitbit daily	2 (12)	4 (24)	11 (65)
I remembered to charge my Fitbit regularly	13 (77)	3 (18)	1 (6)
My Fitbit helped me achieve my step goals	15 (88)	2 (12)	0 (0)
ActivPAL wear			
The thigh activity monitor was uncomfortable to wear	3 (18)	5 (29)	9 (53)
It was easy to remember to complete my sleep log daily	8 (47)	5 (29)	4 (24)
I always completed my sleep log at the same time of day	6 (35)	4 (24)	7 (41)
Blood pressure measurements			
It was easy to remember to take my blood pressure	8 (47)	4 (24)	5 (29)
It was easy connecting my monitor to my phone	14 (82)	2 (13)	1 (6)
I know how to take a blood pressure correctly	17 (100)	0 (0)	0 (0)
It was useful knowing my blood pressure	16 (94)	0 (0)	1 (6)
It was hard to find time to take my blood pressure	2 (13)	7 (44)	7 (44)
Overall program rating			
I was satisfied with the program	17 (100)	0 (0)	0 (0)

XSL•FO
RenderX

Item	Agree, n (%)	Neutral, n (%)	Disagree, n (%)
I would recommend this program to others	17 (100)	0 (0)	0 (0)

[a]Participants ranked intervention features on a 5-point Likert scale from (1) “strongly agree” to (5) “strongly disagree.” Items were assessed at the end of the health coaching intervention via an online survey. To improve readability, (1) “strongly agree” and (2) “agree” were combined, as well as (4) “disagree” and (5) “strongly disagree”.

Concerning specific components of the intervention, all agreed that health coaching sessions helped increase their PA (n=17, 100%) and most agreed these sessions helped decrease their sitting time (n=16, 94%). Most participants said they referenced the information sheets during the program (n=14, 82%) and that this information helped them increase their PA (n=10, 59%). The Fitbit overall was rated as easy to use (n=16, 94%) and set up (n=17, 100%). Most said they checked their Fitbit step count daily (n=15, 88%). Most said Fitbit instructions were easy to understand (n=15, 88%), Fitbit was easy to remember to charge (n=13, 77%), and it helped them achieve their step goals (n=15, 88%).

All participants indicated they knew how to take a BP correctly after starting the program (n=17, 100%), with a majority indicating it was useful knowing their BP (n=16, 94%) and that it was easy to connect their monitor to their phone (n=14, 82%). Less than half of the participants agreed it was easy to remember to take their BP (n=8, 47%) and find time to do so (n=7, 44%).

Preliminary Efficacy

Participants on average wore the activPAL for 7.39 (SD 0.51) days at pretesting and 6.55 (SD 1.46) days at posttesting. As seen in Table 3, activPAL measures did not significantly change from pre- to posttesting, including average steps achieved per week (*P*=.75), standing time (*P*=.78), stepping time (*P*=.49), and sitting time (*P*=.38). Fitbit data further contextualized the changes observed in steps during the intervention period. A statistically significant increase in Fitbit steps was only observed from pretesting to the end of the intensive intervention (*P*=.01).

Table . Physical activity and blood pressure changes of Hypertensive Disorders of Pregnancy Postpartum Exercise intervention participants (N=17)[a].

Variable	Pretest	Week 5	Posttest	Hedges *g*
ActivPAL measures				
Steps, mean (SD)	6505.17 (2133.40)	—[b]	6643.60 (1997.68)	0.75
Standing time, mean (SD)	249.57 (65.40)	—	269.52 (113.51)	0.78
Stepping time, mean (SD)	88.57 (27.07)	—	91.93 (21.53)	0.49
Sitting time, mean (SD)	868.14 (189.03)	—	899.85 (255.90)	0.38
Fitbit measures				
Steps, mean (SD)	6584.47 (2705.65)	7898.27 (2359.32)[c]	7031.67 (2288.79)	0.48
BP[d] measures				
Systolic BP, mean (SD)	122.12 (8.04)	120.67 (8.33)	120.83 (8.10)	–0.26
Diastolic BP, mean (SD)	82.95 (6.02)	81.77 (6.31)	81.15 (5.47)	–0.44
Hypertensive, n (%)	11 (64.7)	10 (58.8)	7 (41.2)[c]	–0.51

[a]All steps reported in units of average steps per week, time variables in average minutes per day, and blood pressures in average mm Hg. The activPAL4 micro and Fitbit Inspire 2 assessed changes in participants’ physical activity. The Omron Series 5 blood pressure monitor assessed changes in participants’ blood pressure. Reported Hedges *g* are effect size estimates between pretest and posttest. Given all variables were nonparametric in nature, Wilcoxon signed-rank test was used to detect significant differences between 2 timepoints. Statistical significance defined as *P*<.05.

[b]Not applicable.

[c]Significant difference from baseline.

[d]BP: blood pressure.

Changes in PA determinants, self-efficacy, planning, monitoring, and identity from pre- to posttesting are shown in Table 4. Overall, PA planning behaviors increased significantly (mean 5.35, SD 4.97 to mean 15.06, SD 3.09; *P*<.001), with increases in participant’s ability to plan when to be active, where to be active, who to be active with, how to overcome disruptions of plans and plan around difficult situations. Behaviors associated with monitoring of PA also significantly increased (mean 7.29, SD 3.44 to mean 13.00, SD 2.45; *P*<.001), with participants indicating increases in monitoring whether they are active enough, that they walk enough steps per day, and the awareness of PA goals. Components of PA identity significantly increased, including being more likely to identify as an active person, describing self to others as being active, needing PA to feel good, and indicating that giving up exercise would cause a sense of loss. No other statistically significant changes were observed.

XSL•FO
RenderX

Table . Change in physical activity determinants, self-efficacy, planning, monitoring, and identity of participants in the Hypertensive Disorders of Pregnancy Postpartum Exercise intervention (N=17)[a].

Items	Pretest	Posttest	*P* value
Planning physical activity, mean (SD)			
I can plan when to be active	0.77 (0.75)	2.18 (0.39)	<.01
I can plan where to be active	0.82 (0.95)	2.12 (0.49)	.01
I can plan how to be active	0.65 (0.79)	2.29 (0.47)	.06
I can plan how often to be active	0.65 (0.70)	2.17 (0.64)	.06
I can plan who to be active with	0.65 (0.86)	1.65 (0.79)	.01
I know what to do if my plans are disrupted	0.65 (0.61)	1.41 (0.80)	.02
I can cope with setbacks	0.59 (0.51)	1.59 (0.62)	.10
I can plan around difficult situations	0.59 (0.51)	1.65 (0.79)	.04
Total planning	5.35 (4.97)	15.06 (3.09)	<.001
Monitoring physical activity, mean (SD)			
I monitor whether I am active enough	1.00 (0.71)	2.12 (0.60)	.01
I monitor that I walk enough steps per day	0.76 (0.56)	2.00 (0.87)	.02
I have physical activity intentions on my mind	1.82 (0.64)	2.41 (0.51)	.09
I am aware of my physical activity goals	1.12 (0.86)	2.12 (0.70)	.02
I try to be physically active regularly	1.35 (0.93)	2.18 (0.39)	.17
I try to achieve my physical activity goals	1.24 (0.90)	2.18 (0.39)	.21
Total score	7.29 (3.44)	13.00 (2.45)	<.001
Physical activity identity, mean (SD)			
I am an active person	1.88 (0.86)	2.18 (0.95)	.02
I describe myself to others as active	0.88 (0.70)	1.53 (1.23)	.04
I have goals related to physical activity	2.00 (1.28)	2.77 (0.66)	.94
Physical activity is central to my self-concept	1.53 (1.07)	1.82 (1.07)	.19
I need physical activity to feel good about myself	2.47 (1.01)	2.35 (1.06)	.04
Others see me as an active person	1.52 (1.01)	1.76 (1.15)	.47
Being active is more than just exercise	2.71 (1.21)	2.71 (0.92)	.32
I would feel loss if I had to give up exercising	1.94 (0.97)	2.29 (1.21)	.001
I think about physical activity often	2.29 (1.10)	2.29 (0.99)	.13
Total score	17.24 (6.88)	19.71 (6.55)	.09

[a]Participants completed online surveys prior to the start and at the conclusion of the intervention. Validated surveys assessed participants' ability to plan for physical activity, monitor physical activity levels, and physical activity related identities. Differences assess in continuous measures with paired *t* tests and categorical with Fisher exact tests. All Likert scales ranged from "strongly disagree" (0) to "strongly agree" (5). Statistical significance

XSL•FO
RenderX

defined as $P<.05$.

At pretesting, the average SBP was 122.12 (SD 8.04) mm Hg and DBP was 82.95 (SD 6.02) mm Hg (Table 3). Of the pretest readings, 11 (65%) participants had average readings that would be classified as HTN based on American Heart Association guidelines. Significant decreases in the number of participants with hypertensive readings were observed between pre- and posttesting (11 vs 7 participants; $P=.01$; Hedges $g=–0.51$). No statistically significant changes in SBP (mean –1.44, SD 4.68 mm Hg; $P=.22$) and DBP (mean –1.19, SD 4.76 mm Hg; $P=.32$) were observed from pretesting to the end of the intensive intervention (week 5) or from pretesting to posttesting (SBP: mean –1.28, SD 3.59; Hedges $g=–0.26$; $P=.16$ and DBP: mean –1.80, SD 5.03; Hedges $g=–0.44$; $P=.12$). When assessing the association between an increase in steps by 1000 and BP, no findings were statistically significant (SBP: mean –0.30, SD 1.11 mm Hg; $P=.79$ and DBP: mean –1.52, SD 1.46 mm Hg; $P=.32$).

Discussion

Principal Findings

The HyPE intervention included 17 participants whose prior pregnancies were complicated by HDP. Overall, the intervention was found to be feasible and acceptable as we hypothesized. Preliminary efficacy assessments did not support our hypothesis as participants did not significantly increase the number of activPAL measured steps achieved from pre- to posttesting. Indicating that we failed to achieve a clinically significant impact on the behavioral target, a critical milestone in phase IIa of the ORBIT model [16]. Factors promoting PA (eg, PA monitoring, planning, and identity) were observed to significantly improve. HTN prevalence was the only BP metric observed to significantly decrease from pre- to posttesting.

Contextualizing Findings

CVD has previously been described as a “man’s disease,” but the acknowledged importance of CVD in women is increasing. Reviews of CVD-related clinical trials have highlighted an underrepresentation of women in research [34], subsequently attributing to sex-based disparities in the clinical guidelines, recommendations, treatments, and health outcomes related to CVD. Potential downstream effects of these disparities were reflected in our findings, with 94% (n=16) of individuals reporting at no time did a health care provider discuss the increased risk of CVD associated with HDP diagnosis. With HDP an important risk factor for CVD, it is critical as a field to find effective strategies to promote cardiovascular health in this high-risk population.

To the authors’ knowledge, only 1 randomized clinical trial has assessed the efficacy of a PA intervention after HDP. Participants of that study increased their steps more than the control (mean difference: 647 steps per day, 95% CI 169 - 1124; $P=.009$) [35]. A total of 65% of participants completed the study with 94% indicating that they would recommend the program to others [35]. The HyPE intervention had similar adherence and acceptability results, with 90% of participants completing the intervention and 100% satisfied with the program. Intervening on PA this early in postpartum (3 - 6 months postpartum) is novel and our results indicate that interventions during this timeframe may be feasible and acceptable. Although we did not observe significant increases in device-measured PA, participants did significantly increase their self-reported abilities to plan and monitor PA. This discrepancy may be a result of seasonality, given that the winter season is associated with a decrease of up to 2500 steps compared to summer months [36]. Those completing their posttesting in the winter (ie, December or January; n=9, 53%), consequently may have seen a decrease in their steps due to the colder temperatures. While participants in the HyPE intervention did not increase their steps, the intervention may have prevented the typical decline observed when transitioning from summer to winter months. Future interventions should consider the impacts seasonality may have on their population and develop strategies to combat potential effects colder weather may bring. Finally, we included the target population in the development of the intervention in previous formative work assessing the unique behavioral determinants, needs, and desires of postpartum individuals with HDP [17]. Formative work has been called for in postpartum PA intervention development with inclusion of the target population in this process essential to creating equitable and culturally appropriate solutions [37,38].

Limitations

There are several limitations of this study to consider when interpreting results. There was no control group due to this being a proof-of-concept study, limiting the ability to assess whether the changes observed were a direct effect of the intervention itself. This decision was intentional, however, as very few PA interventions have targeted the early postpartum period, and demonstrating feasibility and acceptability was our primary objective. This also follows the ORBIT model, representing phase IIa in the behavioral treatment developmental pathway [16]. Given that the milestone for successful completion of phase IIa was not reached with no clinically significant changes in behavior detected, further refinement and optimization of the intervention is warranted. These refinements should include lowering the <9000 step inclusion criteria threshold, as well as the duration or frequency of study visits. In addition, there was no activPAL4 micro assessment of PA at the end of the intensive intervention (week 5), which limited our ability to determine the immediate effects of the intervention on PA behavior. It is also possible that our PA eligibility criteria (must achieve <9000 steps/day) may have biased our results by not recruiting a truly inactive population. It is estimated that achieving Americans 7000 - 8000 steps is equivalent to meeting the PA guidelines for Americans [39]. Further, the significant decrease in hypertensive individuals should be interpreted with caution, as this may reflect that most individuals who were classified as hypertensive were only slightly above the American Heart Association defined BP threshold for HTN. Finally, the population recruited was not racially, ethnically, or socioeconomically diverse, with a majority of individuals being White and of high socioeconomic status. Future studies will require a larger, more diverse sample and should consider recruiting individuals achieving <7000 steps per day. Future

XSL•FO
RenderX

studies are needed to contextualize the needs and desires of diverse populations. Given the importance intersectional identities may have in a birthing person's health and experiences with the medical system, the inclusion of diverse cohorts is essential to the equitable promotion of female cardiovascular health after HDP.

Conclusions

HDP incidence is increasing and is associated with an increased risk of CVD later in life. Our study suggests that intervening in PA behaviors early after HDP is feasible and acceptable. Future studies should further refine intervention strategies for this population and test its efficacy in a larger, more diverse, and less active sample.

Acknowledgments

The authors would like to thank the support of the American Heart Association for funding this project (23PRE1018306). The authors would also like to thank all the participants who made this study possible.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to participants not providing prior consent to their data being shared.

Conflicts of Interest

None declared.

References

1. Khosla K, Heimberger S, Nieman KM, et al. Long-term cardiovascular disease risk in women after hypertensive disorders of pregnancy: recent advances in hypertension. Hypertension 2021 Sep;78(4):927-935. [doi: 10.1161/HYPERTENSIONAHA.121.16506] [Medline: 34397272]
2. Melchiorre K, Thilaganathan B, Giorgione V, Ridder A, Memmo A, Khalil A. Hypertensive disorders of pregnancy and future cardiovascular health. Front Cardiovasc Med 2020 Apr;7:59. [doi: 10.3389/FCVM.2020.00059/BIBTEX]
3. Benschop L, Duvekot JJ, Roeters van Lennep JE. Future risk of cardiovascular disease risk factors and events in women after a hypertensive disorder of pregnancy. Heart 2019 Aug;105(16):1273-1278. [doi: 10.1136/heartjnl-2018-313453] [Medline: 31175138]
4. Bushnell C, McCullough LD, Awad IA, et al. Guidelines for the prevention of stroke in women. Stroke 2014 May;45(5):1545-1588. [doi: 10.1161/01.str.0000442009.06663.48]
5. Tooher J, Thornton C, Makris A, Ogle R, Korda A, Hennessy A. All hypertensive disorders of pregnancy increase the risk of future cardiovascular disease. Hypertension 2017 Oct;70(4):798-803. [doi: 10.1161/HYPERTENSIONAHA.117.09246] [Medline: 28893895]
6. Lewey J, Levine LD, Yang L, Triebwasser JE, Groeneveld PW. Patterns of postpartum ambulatory care follow-up care among women with hypertensive disorders of pregnancy. J Am Heart Assoc 2020 Sep;9(17):e016357. [doi: 10.1161/JAHA.120.016357] [Medline: 32851901]
7. Pouliou T, Ki M, Law C, Li L, Power C. Physical activity and sedentary behaviour at different life stages and adult blood pressure in the 1958 British cohort. J Hypertens 2012 Feb;30(2):275-283. [doi: 10.1097/HJH.0b013e32834f1915] [Medline: 22179092]
8. Whelton SP, Chin A, Xin X, He J. Effect of aerobic exercise on blood pressure: a meta-analysis of randomized, controlled trials. Ann Intern Med 2002 Apr 2;136(7):493-503. [doi: 10.7326/0003-4819-136-7-200204020-00006] [Medline: 11926784]
9. Muntner P, Jaeger BC, Hardy ST, et al. Age-specific prevalence and factors associated with normal blood pressure among US adults. Am J Hypertens 2022 Apr 2;35(4):319-327. [doi: 10.1093/ajh/hpab154] [Medline: 34599824]
10. Sobierajski FM, Purdy GM, Usselman CW, et al. Maternal physical activity is associated with improved blood pressure regulation during late pregnancy. Can J Cardiol 2018 Apr;34(4):485-491. [doi: 10.1016/j.cjca.2018.01.021] [Medline: 29571429]
11. Haakstad LAH, Edvardsen E, Bø K. Effect of regular exercise on blood pressure in normotensive pregnant women. A randomized controlled trial. Hypertens Pregnancy 2016 Apr;35(2):170-180. [doi: 10.3109/10641955.2015.1122036]
12. Borodulin K, Evenson KR, Herring AH. Physical activity patterns during pregnancy through postpartum. BMC Womens Health 2009 Nov 19;9(1):32. [doi: 10.1186/1472-6874-9-32] [Medline: 19925650]
13. Parikh NI, Gonzalez JM, Anderson CAM, et al. Adverse pregnancy outcomes and cardiovascular disease risk: unique opportunities for cardiovascular disease prevention in women: a scientific statement from the American Heart Association. Circulation 2021 May 4;143(18):e902-e916. [doi: 10.1161/CIR.0000000000000961] [Medline: 33779213]
14. Mehta LS, Warnes CA, Bradley E, et al. Cardiovascular considerations in caring for pregnant patients: a scientific statement from the American Heart Association. Circulation 2020 Jun 9;141(23):e884-e903. [doi: 10.1161/CIR.0000000000000772] [Medline: 32362133]

XSL•FO
RenderX

15. Mehta LS, Sharma G, Creanga AA, et al. Call to action: maternal health and saving mothers: a policy statement from the American Heart Association. Circulation 2021 Oct 12;144(15):e251-e269. [doi: 10.1161/CIR.0000000000001000] [Medline: 34493059]
16. Czajkowski SM, Powell LH, Adler N, et al. From ideas to efficacy: the ORBIT model for developing behavioral treatments for chronic diseases. Health Psychol 2015 Oct;34(10):971-982. [doi: 10.1037/hea0000161] [Medline: 25642841]
17. EPI: abstract detail. EPI Lifestyle. URL: https://epi.apprisor.org/epsAbstractEPI.cfm?id=4 [accessed 2024-03-22]
18. Whitaker KM, Jones MA, Dziewior J, et al. Feasibility, acceptability, and preliminary efficacy of a single-arm, remotely-delivered health coaching intervention to increase physical activity and reduce sedentary behavior during pregnancy. Research Square. Preprint posted online on Jul 19, 2022. [doi: 10.21203/rs.3.rs-1755320/v1]
19. Hettema J, Steele J, Miller WR. Motivational interviewing. Annu Rev Clin Psychol 2005;1(1):91-111. [doi: 10.1146/annurev.clinpsy.1.102803.143833] [Medline: 17716083]
20. Bundy C, Cpsychol P. Changing behaviour: using motivational interviewing techniques. J R Soc Med 2004;97 Suppl 44(Suppl 44):43-47. [Medline: 15239293]
21. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
22. Biber DD. Exercise identity, self-regulatory efficacy, and self-compassion prepared for psychological studies. Psychol Stud 2020 Sep;65(3):261-269. [doi: 10.1007/s12646-020-00556-w]
23. Sniehotta FF, Scholz U, Schwarzer R. Bridging the intention–behaviour gap: planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychol Health 2005 Apr;20(2):143-160. [doi: 10.1080/08870440512331317670]
24. The R Project for Statistical Computing. URL: https://www.r-project.org/ [accessed 2024-12-20]
25. Deboonstra/pregnancy247. GitHub. URL: https://github.com/deboonstra/pregnancy247 [accessed 2024-11-28]
26. Lyden K, Keadle SK, Staudenmayer J, Freedson PS. The activPALTM accurately classifies activity intensity categories in healthy adults. Med Sci Sports Exerc 2017 May;49(5):1022-1028. [doi: 10.1249/MSS.0000000000001177] [Medline: 28410327]
27. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson PS. Validation of wearable monitors for assessing sedentary behavior. Med Sci Sports Exerc 2011 Aug;43(8):1561-1567. [doi: 10.1249/MSS.0b013e31820ce174] [Medline: 21233777]
28. Hardcastle SJ, Jiménez-Castuera R, Maxwell-Smith C, Bulsara MK, Hince D. Fitbit wear-time and patterns of activity in cancer survivors throughout a physical activity intervention and follow-up: exploratory analysis from a randomised controlled trial. PLOS ONE 2020;15(10):e0240967. [doi: 10.1371/journal.pone.0240967] [Medline: 33075100]
29. Shimbo D, Artinian NT, Basile JN, et al. Self-measured blood pressure monitoring at home: a joint policy statement from the American Heart Association and American Medical Association. Circulation 2020 Jul 28;142(4):e42-e63. [doi: 10.1161/CIR.0000000000000803] [Medline: 32567342]
30. Grover-Páez F, Cardona-Muñoz EG, Cardona-Müller D, et al. Validation of the Omron HEM-7320-LA, upper arm blood pressure monitor with Intelli Wrap Technology Cuff HEM-FL1 for self-measurement and clinic use according to the European Society of Hypertension International Protocol revision 2010 in the Mexican population. Blood Press Monit 2017 Dec;22(6):375-378. [doi: 10.1097/MBP.0000000000000290] [Medline: 28945665]
31. Suva M. Validation of BP monitoring devices. J Clin Hypertens 2020 Feb;3:15-19.
32. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863. [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
33. Kaneko H, Yano Y, Itoh H, et al. Association of blood pressure classification using the 2017 American College of Cardiology/American Heart Association blood pressure guideline with risk of heart failure and atrial fibrillation. Circulation 2021 Jun 8;143(23):2244-2253. [doi: 10.1161/CIRCULATIONAHA.120.052624] [Medline: 33886370]
34. Melloni C, Berger JS, Wang TY, et al. Representation of women in randomized clinical trials of cardiovascular disease prevention. Circ Cardiovasc Qual Outcomes 2010 Mar;3(2):135-142. [doi: 10.1161/CIRCOUTCOMES.110.868307] [Medline: 20160159]
35. Lewey J, Murphy S, Zhang D, et al. Effectiveness of a text-based gamification intervention to improve physical activity among postpartum individuals with hypertensive disorders of pregnancy: a randomized clinical trial. JAMA Cardiol 2022 Jun 1;7(6):591-599. [doi: 10.1001/jamacardio.2022.0553] [Medline: 35442393]
36. Hamilton SL, Clemes SA, Griffiths PL. UK adults exhibit higher step counts in summer compared to winter months. Ann Hum Biol 2008;35(2):154-169. [doi: 10.1080/03014460801908058] [Medline: 18428010]
37. Peralta LR, Cotton WG, Dudley DA, Hardy LL, Yager Z, Prichard I. Group-based physical activity interventions for postpartum women with children aged 0-5 years old: a systematic review of randomized controlled trials. BMC Womens Health 2021 Dec 28;21(1):435. [doi: 10.1186/s12905-021-01581-1] [Medline: 34963456]
38. Vastine A, Gittelsohn J, Ethelbah B, Anliker J, Caballero B. Formative research and stakeholder participation in intervention development. Am J Health Behav 2005;29(1):57-69. [doi: 10.5993/ajhb.29.1.5] [Medline: 15604050]

XSL•FO
RenderX

39. Tudor-Locke C, Craig CL, Brown WJ, et al. How many steps/day are enough? For adults. Int J Behav Nutr Phys Act 2011 Jul 28;8(1):79. [doi: 10.1186/1479-5868-8-79] [Medline: 21798015]

Abbreviations

BP: blood pressure
CVD: cardiovascular disease
DBP: diastolic blood pressure
HDP: hypertensive disorders of pregnancy
HTN: hypertension
HyPE: Hypertensive Disorders of Pregnancy Postpartum Exercise
ORBIT: Obesity-Related Behavioral Intervention Trials
PA: physical activity
SBP: systolic blood pressure

Edited by A Mavragani; submitted 23.08.24; peer-reviewed by H Liu, J Kitt; revised version received 01.11.24; accepted 05.11.24; published 08.01.25.

Please cite as:
Borrowman JD, Carr LJ, Pierce GL, Story WT, Gibbs BB, Whitaker KM
Postpartum Remote Health Coaching Intervention for Individuals With a Hypertensive Disorder of Pregnancy: Proof-of-Concept Study
JMIR Form Res 2025;9:e65611
URL: https://formative.jmir.org/2025/1/e65611
doi:10.2196/65611

© Jaclyn D Borrowman, Lucas J Carr, Gary L Pierce, William T Story, Bethany Barone Gibbs, Kara M Whitaker. Originally published in JMIR Formative Research (https://formative.jmir.org), 8.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Testing the Recruitment Frequency, Implementation Fidelity, and Feasibility of Outcomes of the Heart Failure Activity Coach Study (HEALTHY): Pilot Randomized Controlled Trial

Andreas Blomqvist[1], PhD; Maria Bäck[1,2], PhD; Leonie Klompstra[1], PhD; Anna Strömberg[1,3], PhD; Tiny Jaarsma[1], PhD

[1]Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

[2]Department of Occupational Therapy and Physiotherapy, Sahlgrenska University Hospital, Gothenburg, Sweden

[3]Department of Cardiology, Linköping University, Linköping, Sweden

Corresponding Author:
Andreas Blomqvist, PhD
Department of Health, Medicine and Caring Sciences
Linköping University
Sandbäcksgatan 7
Linköping, SE-581 83
Sweden
Phone: 46 739617729
Email: andreas.blomqvist@liu.se

Abstract

Background: Heart failure (HF) is a common and deadly disease, precipitated by physical inactivity and sedentary behavior. Although the 1-year survival rate after the first diagnosis is high, physical inactivity and sedentary behavior are associated with increased mortality and negatively impact the health-related quality of life (HR-QoL).

Objective: We tested the recruitment frequency, implementation fidelity, and feasibility of outcomes of the Activity Coach app that was developed using an existing mobile health (mHealth) tool, Optilogg, to support older adults with HF to be more physically active and less sedentary.

Methods: In this pilot clinical randomized controlled trial (RCT), patients with HF who were already using Optilogg to enhance self-care behavior were recruited from 5 primary care health centers in Sweden. Participants were randomized to either have their mHealth tool updated with the Activity Coach app (intervention group) or a sham version (control group). The intervention duration was 12 weeks, and in weeks 1 and 12, the participants wore an accelerometer daily to objectively measure their physical activity. The HR-QoL was measured with the Kansas City Cardiomyopathy Questionnaire (KCCQ), and subjective goal attainment was assessed using goal attainment scaling. Baseline data were collected from the participants' electronic health records (EHRs).

Results: We found 67 eligible people using the mHealth tool, of which 30 (45%) initially agreed to participate, with 20 (30%) successfully enrolled and randomized to the control and intervention groups in a ratio of 1:1. The participants' daily adherence to registering physical activity in the Activity Coach app was 69% (range 24%-97%), and their weekly adherence was 88% (range 58%-100%). The mean goal attainment score was –1.0 (SD 1.1) for the control group versus 0.6 (SD 0.6) for the intervention group (P=.001). The mean change in the overall HR-QoL summary score was –9 (SD 10) for the control group versus 3 (SD 13) in the intervention group (P=.027). There was a significant difference in the physical limitation scores between the control (mean 45, SD 27) and intervention (mean 71, SD 20) groups (P=.04). The average length of sedentary bouts increased by 27 minutes to 458 (SD 84) in the control group minutes and decreased by 0.70 minutes to 391 (SD 117) in the intervention group (P=.22). There was a nonsignificant increase in the mean light physical activity (LPA): 146 (SD 46) versus 207 (SD 80) minutes in the control and intervention groups, respectively (P=.07).

Conclusions: The recruitment rate was lower than anticipated. An active recruitment process is advised if a future efficacy study is to be conducted. Adherence to the Activity Coach app was high, and it may be able to support older adults with HF in being physically active.

Trial Registration: ClinicalTrials.gov NCT05235763; https://clinicaltrials.gov/study/NCT05235763

(JMIR Form Res 2025;9:e62910) doi:10.2196/62910

XSL•FO
RenderX

KEYWORDS

heart failure; disease management; physical activity; sedentary; older adults; aging; mobile health; mHealth; feasibility; quality of life; digital health; smartphone

Introduction

Background

Heart failure (HF) is a clinical syndrome characterized by a reduced cardiac output leading to a deficit in meeting the metabolic needs of the cells [1]. It is a major cause of morbidity and mortality worldwide [2]. Physical inactivity and sedentary behavior adversely affect outcomes in patients with HF. Sedentary behavior is a state where no activities performed increase energy expenditure substantially above the resting level, typically sitting or lying down [3]. Physical inactivity can be defined as the nonachievement of targets according to the physical activity guidelines that are recommended for an individual [4]. Although the 1-year survival rate for patients after the first diagnosis of HF is high [5], physical inactivity is associated with additionally increased mortality them [6]. Similarly, sedentary behavior is also associated with increased mortality [7] and negatively impacts the health-related quality of life (HR-QoL) [8]. It is, therefore, desirable to support patients with HF to be more physically active and break their sedentary behavior. A tool aimed at doing this, called "Activity Coach," has been developed as a separate module on an existing mobile health (mHealth) tool [9].

The Activity Coach app is aimed at supporting everyday physical activity and reducing sedentary behavior and should not be viewed as an exercise tool. Furthermore, it is primarily intended for the physically inactive population with HF and is planned to be tested for efficacy in that population [9].

The Medical Research Council (MRC) suggests that before conducting an appropriately powered efficacy study, enough piloting and feasibility work has to be performed to be confident that the intervention can be delivered as intended (fidelity) and that safe assumptions can be made about effect sizes and variability and about rates of recruitment and retention [10]. A pilot randomized controlled trial (RCT) is the best way to assess the feasibility of a large, expensive full-scale study [11], and hence, the objectives of a pilot RCT need to be different from those of a future efficacy study, including comments on the uncertainties to be investigated [12]. In accordance with the MRC [10] and the CONSORT (Consolidated Standards of Reporting Trials) guidelines for pilot RCTs [13] (Multimedia Appendix 1), we conducted a pilot RCT to evaluate a study design, with the main focus being the rate of recruitment. In addition, we sought to assess the fidelity of the intervention, as well as the plausibility of the proposed outcome measures, in terms of data quality and effect sizes.

Objective

The objective of this pilot RCT was to test the study protocol of an intervention in terms of patient recruitment, intervention fidelity, and plausibility of outcomes. The research questions were as follows:

- Recruitment: What was the recruitment frequency and attrition rate in the target population?
- Implementation fidelity: To what extent was the intervention delivered as intended?
- Outcomes: How was the plausibility of outcomes in terms of data integrity and quality, and what limited efficacy was seen regarding the included outcome measures?

Methods

Trial Design and Participants

The Heart Failure Activity Coach Study (HEALTHY) was a single-blinded, parallel pilot clinical RCT. Since the Activity Coach app was developed on the platform of an already existing mHealth tool called Optilogg, only people equipped with a home-based self-care support device called Optilogg were eligible to participate. As all study participants regardless of study arm allocation would have Optilogg since before, this also eliminated the risk of the effects of Optilogg per se confounding the results. Participants also needed to be listed at a primary health care center.

The eligibility criteria for participation were age≥18 years, a confirmed HF diagnosis, using Optilogg (obtained from a participating HF clinic), being physically inactive based on a screening question, and having the ability to consent to participation. Eligible people were excluded if they did not consent to a home visit or to using an accelerometer, if they had a life expectancy shorter than 6 months, or if they were already participating in another study on physical activity.

Change to Trial Design After Commencement

After the initial slow recruitment of just 2 randomized participants after 3 months, an amendment was submitted to the ethical review board, the inclusion criteria were modified, and it was decided that all people willing to participate would be included regardless of their answer to the screening question on physical inactivity. This change was justified based on the fact that the Activity Coach app might have benefits for a more active population with HF, too. Furthermore, potential participants were now contacted by telephone instead of regular mail. In the *Ancillary Analyses* section, potential differences between participants as stratified by the screening question are reported.

Study Procedure

Eligible people were first contacted by regular mail, and those willing to participate would be asked to answer a single-item self-report question. This question has previously been evaluated as a screening tool for classifying respondents as being physically active or physically inactive, and the same methodology was used in this study to identify physically inactive people [14]. The question is item 9 in the validated the European Heart Failure Self-care Behaviour Scale, and the question is in the shape of a statement ("I exercise regularly"), which the respondent answers using an ordinal scale from 1 to

XSL•FO
RenderX

5, with 1 being "I completely agree" to 5 being "I don't agree at all" [15]. In this study, responses from 3 to 5 were coded as physically inactive [14]. If a participant met the eligibility criteria, a home visit was scheduled. During the home visit, the participant signed an informed consent form, was randomized to the intervention or the control group, filled out baseline questionnaires, and received an accelerometer. Lastly, the Activity Coach app or the sham version was installed on their Optilogg mHealth tool.

The intervention period was 12 weeks, and during weeks 1 and 12 of the study, participants wore the accelerometer every day. At the end of the study the same questionnaires that were completed at the start of the study were filled out again. Clinical data, such as demographics, etiology, comorbidities, and pharmacological treatments, were collected from the participants' electronic health records (EHRs) by the HF nurse at each respective primary health care center.

Intervention and Control

The mHealth Tool Optilogg

Optilogg consists of a tablet computer wirelessly connected to a weight scale, and the user is encouraged to interact with it daily. Without the addition of the Activity Coach app, Optilogg has 3 main features: subjective and objective symptom tracking, interactive education about living with HF, and an automated flexible loop-diuretics regime. It has been shown to improve self-care behavior in people with HF [16,17].

The Activity Coach App Intervention

The Activity Coach app was added as a novel feature to the existing Optilogg, with the purpose of supporting people with HF to be physically active, and was developed using principles described by the MRC [9,10]. Therefore, in addition to existing Optilogg features, the mHealth tool included Activity Coach features as well. Every day, the user can, in an intuitive graphical user interface (GUI), enter how much time they have been physically active in increments of 10 minutes regardless of what physical activity means for them. This allows for a one-size-fits-all approach to accommodate heterogenous populations. The accumulated physical activity is logged and presented as trends in a separate tab. Each week, there is an on-screen summary of the past week's logged activity, and the user can, if desired, set a goal for the upcoming week, although they can at any time turn off the goal-setting feature. The progress toward reaching the goal is illustrated in the trends tab. The Activity Coach app was designed to first create motivation to be physically active and then provide a means of tracking physical activity. Motivation is achieved by providing an educational tip a day for 1 week, aimed at establishing 3 beliefs: physical activity leads to certain outcomes (positive outcome expectancy), the person can perform physical activity (self-efficacy), and the outcomes are desirable to the person (goal congruence). More details on the underlying intervention theory and theoretical model, as well as the development process, have been published elsewhere [9]. The participants in the intervention group were instructed to try to track their physical activity in the Activity Coach app daily.

Control Version

As previously mentioned, the original Optilogg contains an education module that, in part, deals with physical activity and exercise, and this education module has been shown to increase knowledge about HF [17]. This could have been a possible confounder when trying to analyze the additional effects of the Activity Coach app. To mitigate this, all participants were recruited from existing Optilogg users. The intervention group had the Activity Coach app installed on their existing Optilogg system, and the control group had a sham version of the Activity Coach app installed instead. This slightly changed the appearance of the GUI compared to the original Optilogg prior to the study, such that instead of starting each day on the regular home-screen, the modified system started in the section of the education module dealing with physical activity, but none of the actual Activity Coach features were enabled. This led to an apparent change, also relating to physical activity, which would be immediately noted by the user.

Research Question 1: Recruitment

The recruitment procedure metric to be evaluated was defined as the number of people who met the eligibility criteria who needed to be contacted and screened to reach a certain number of recruited and randomized participants. Monitoring attrition rates was also part of evaluating the study design. As described earlier, the study design was changed to also include participants who originally were intended to not be included, since they were defined as physically active. Hence, this outcome was reported as 2 numbers: one where the original screening was still used and the other where physically active people were also included.

Research Question 2: Implementation Fidelity

Implementation fidelity was evaluated by analyzing how much the participants randomized to the intervention group actually used the Activity Coach app (eg, adherence, which is a common means of assessing intervention fidelity in physical activity interventions) [18]. Adherence was defined as the ratio of actual use to intended use [19], and since, no clear recommendation exists on how often the Activity Coach app or similar behavior change interventions relating to physical activity should be used, we reported daily and weekly adherence. Daily adherence was calculated as the number of days that the participant performed physical activity divided by the study duration in days (ie, 84). Weekly adherence was calculated as the number of weeks that the participant performed physical activity in 1 day or more divided by the number of weeks in the study (ie, 12). Furthermore, the Activity Coach app offers a goal-setting feature on a weekly basis. The usage of this feature was also examined.

Research Question 3: Outcomes

The outcome measures studied were the subjective goal attainment score; the HR-QoL, measured with the Kansas City Cardiomyopathy Questionnaire (KCCQ); and objective physical activity, measured with an accelerometer.

Subjective Goal Attainment

For goal attainment, we relied on the methods proposed by Urach et al [20]. Upon enrollment, participants selected 2 goals

XSL•FO
RenderX

from a list of 10 physical activity–related goals they wanted to achieve during the 12-week study. The list was based on a literature review of the effects of increased physical activity on people with HF. The 10 goals are listed in Table 1. Upon completing their participation in the study, each participant rated the 2 selected goals during follow-up, comparing the current status with the status when the study started. The rating was on a scale from –2 to 2, where –2 indicated worsening and 2 indicated improvement. A mean value of the scores of the 2 goals was then calculated and constituted the reported value, referred to as the goal attainment score.

Table 1. Prespecified goals for the subjective goal attainment outcome.

Goal number	Goal
1	Better sleep
2	Boosted daytime energy
3	Able to walk with ease without frequent pauses
4	Less physical pain
5	Feel better
6	Reduced shortness of breath
7	Less daytime fatigue
8	More energy to do things around the house
9	Increased desire and energy for social activities
10	Improved appetite

Health-Related Quality of Life

The Kansas City Cardiomyopathy Questionnaire (KCCQ) short version is a 12-item questionnaire used to assess the HR-QoL in people with cardiomyopathies, which outputs 4 subscale summary scores (physical limitation, symptom frequency, quality of life [QoL], and social limitation), as well as a total summary score [21]. The scores range from 0 to 100, where 100 represents the best-possible value for any given score. The questionnaire has been validated in the current population [22].

Objective Physical Activity

Sedentary behavior and physical inactivity were assessed objectively using an accelerometer [23]. The accelerometer used was the validated ActiGraph GT9X [24], and participants were instructed to wear it around the hip for 7 consecutive days. The accelerometer data were reported as the percentage sedentary time, the average length of sedentary bouts, and steps per day. Data were sampled as raw data (30 Hz), converted to 60-second time intervals (or epochs) in counts per minute for the vector magnitude, and analyzed using ActiLife software (v.6.13.4).

Certain criteria were applied to the accelerometer data acquired for it to be included in the analyses. The accelerometer had to be used during at least 4 of the 7 days, with a minimum wear time of 540 minutes/day. Periods with consecutive 0s for 60 minutes or more (with 2-minute tolerance) were interpreted as time of nonuse and excluded from further analyses.

Sample Size

An appropriate sample size for a pilot should be based on the estimated effect size, where a large effect size warrants 15-30 participants, an intermediate effect size requires 25-35 participants [25], and a more general recommendation is a sample size of 12 per group [26]. In this study, the target sample size was set to 40, and enrollment was to stop either when 40 study participants had been randomized or when all eligible people had been contacted (whichever came first).

Randomization

Sheets of paper with an “I” for intervention or a “C” for control were printed and put in sealed envelopes. After an eligible participant had agreed to participate in the study, a home visit was scheduled. The participant signed the informed consent form and filled out the baseline questionnaire. The researcher selected an envelope from the batch and opened it and subsequently install the Activity Coach app or the sham version onto the participant’s Optilogg without informing them about the allocation.

Blinding

This was a single-blinded, parallel pilot RCT. Every participant already had the Optilogg system at home and at the start of the study and was randomized to either the control or the intervention group at the beginning, at which point the sham version or the actual Activity Coach app was installed on their Optilogg without them knowing which version they received. The participants were not informed of what arm they were randomized to until the study was over and all data collected. The person who entered and registered all the data was also blinded to the allocation, but the investigators knew which group each participant was allocated to.

Ethical Considerations

Ethics approval was obtained from Sweden’s Ethical Review Board (Dnr 2021-05366-01; approval date: November 17, 2021), with the amendment regarding the recruitment procedure also being approved (Dnr 2023-00799-02; approval date: February 19, 2023). All recruited participants signed and submitted their informed consent to participate in writing.

XSL•FO
RenderX

Statistical Analysis

Continuous variables, such as age and steps, were analyzed using the Student *t* test, and categorical variables were analyzed using the chi-square test. For subjectively graded goals, a mean value was computed from the 2 answers, and the Student *t* test was performed to test the differences between groups.

To test the differences in ordinal variables, such as self-reported physical activity and exercise, as well as nonnormally distributed data, such as the scaled outputs of the KCCQ, the Mann-Whitney U nonparametric test was performed. $P<.05$ was considered statistically significant.

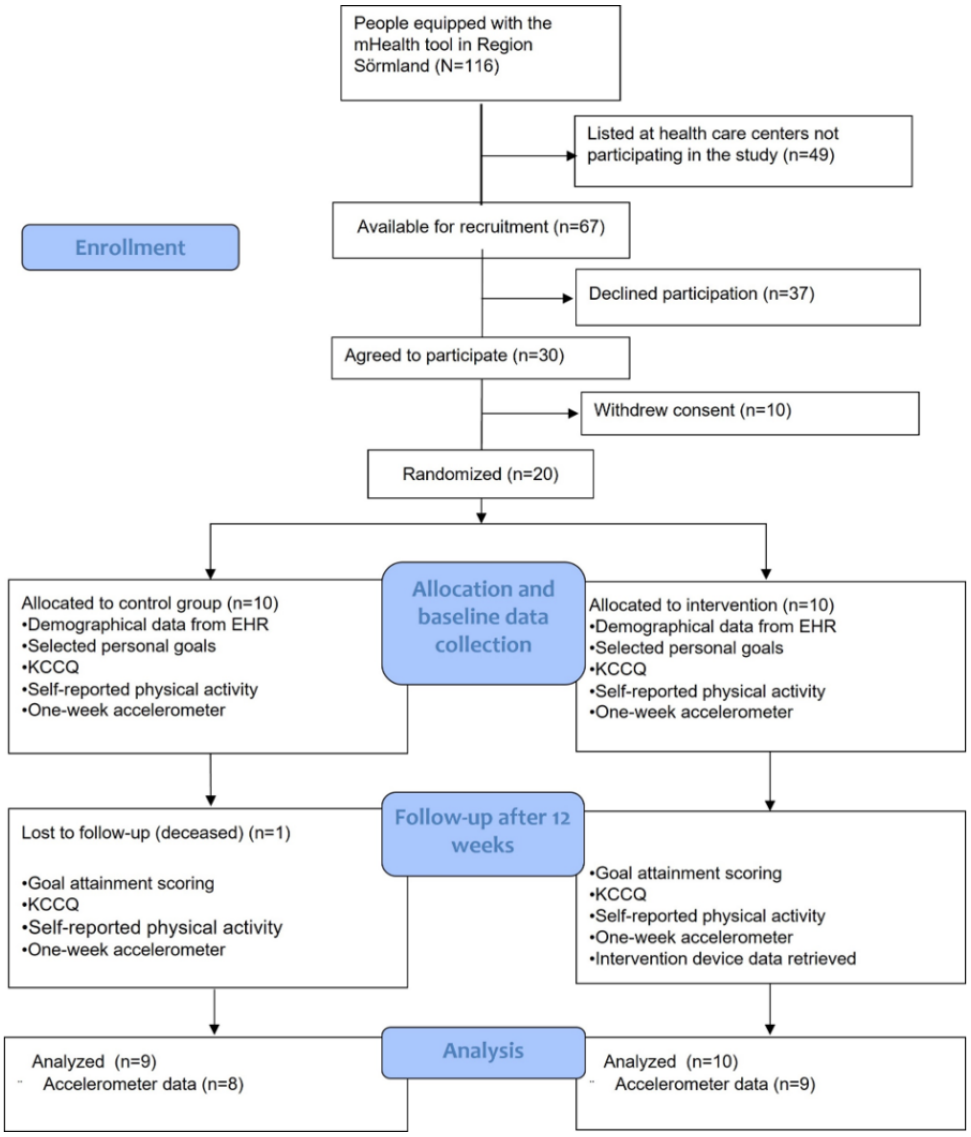
All statistical calculations were computed in SPSS Statistics v.29 (IBM Corporation). All statistical tests performed on outcome data were performed without power calculations and were interpreted accordingly.

Results

Population

The participant selection flowchart, along with the study design, is illustrated in Figure 1.

Figure 1. Participant flow and study design of the pilot RCT. KCCQ - Kansas City Cardiomyopathy Questionnaire, EHR – electronic health records.



In total, 116 people were using the mHealth tool as part of their regular self-care at the start of this study. These 116 potential participants were listed at 11 different primary health care centers, 5 (45%) of which agreed to participate in the study. At these 5 health care centers, 67 (58%) of the 116 people were listed and were contacted and asked to participate. Participants were recruited between September 13, 2022, and March 25, 2023.

Baseline Data

The participants' baseline demographical data are summarized in Table 2. No significant differences between the 2 study arms were detected.

XSL•FO
RenderX

Table 2. Baseline data, including demographics, etiology, comorbidities, and pharmacological treatment, for study participants.

Characteristics	Control group (n=10)	Intervention group (n=10)	Total (N=20)
Age (years), mean (SD)	77 (5)	78 (9)	78 (7)
Females, n (%)	4 (40)	4 (40)	8 (40)
HF[a] type, n (%)			
HFrEF[b]	4 (40)	3 (30)	7 (35)
HFmrEF[c]	1 (10)	4 (40)	5 (25)
HFpEF[d]	5 (50)	3 (30)	8 (40)
NYHA[e] class, n (%)			
NYHA II	4 (40)	6 (60)	10 (50)
NYHA III	6 (60)	4 (40)	10 (50)
Comorbidities, n (%)			
Diabetes	3 (30)	2 (20)	5 (25)
COPD[f]	3 (30)	2 (20)	5 (25)
Hypertension	4 (40)	6 (60)	10 (50)
Atrial fibrillation	3 (30)	5 (50)	8 (40)
Kidney disease	2 (20)	3 (30)	5 (25)
ACEi[g]/ARB[h]/ARNI[i]	8 (80)	10 (100)	18 (90)
Beta blocker	7 (70)	10 (100)	17 (85)
MRA[j]	1 (10)	2 (20)	3 (15)
SGLT2i[k]	2 (20)	5 (50)	7 (35)
Physically inactive	7 (70)	8 (80)	15 (75)

[a]HF: heart failure.

[b]HFrEF: heart failure with reduced ejection fraction.

[c]HFmrEF: heart failure with mildly reduced ejection fraction.

[d]HFpEF: heart failure with preserved ejection fraction.

[e]NYHA: New York Heart Association.

[f]COPD: chronic obstructive pulmonary disease.

[g]ACEi: angiotensin-converting enzyme inhibitor.

[h]ARB: angiotensin receptor blocker.

[i]ARNI: angiotensin receptor blocker neprilysin inhibitor.

[j]MRA: mineralocorticoid receptor antagonist.

[k]SGLT2i: sodium glucose cotransporter 2 inhibitor.

Numbers Analyzed

Of the 20 participants recruited, 1 (5%) participant in the control group died during the study. Another participant in the control group and 1 (5%) in the intervention group did not meet the requirements of wear time for their accelerometer data to be included in the analyses. Hence, accelerometer-based outputs were available for 8 (80%) participants in the control group and 9 (90%) participants in the intervention group. All questionnaire-based data were available for 9 (90%) participants in the control group and 10 (100%) participants in the intervention group.

Research Question 1: Recruitment

Of the 67 participants with HF whom we could contact, 30 (45%) agreed to participate, but 10 (33%) withdrew their consent before being formally recruited and signing the informed consent form. Most did not provide a reason, but 1 (10%) reported a spouse who had gotten sick, and another had hurt himself and did not wish to participate any longer.

Finally, 20 (30%) of the 67 potential participants were recruited and could be randomized to the control group or the intervention group. However, had the amendment to also include people who were assessed to be physically active not been made, only 15 (22%) people would have been recruited and randomized.

XSL•FO
RenderX

Research Question 2: Implementation Fidelity

The daily adherence to registering physical activity in the Activity Coach app was 69% (range 24%-97%), and weekly adherence was 88% (range 58%-100%). On average, participants in the intervention group registered 323 (range 66-422) minutes per week. The lowest weekly adherence was 50%. In addition, 40% (n=8) of the participants had a daily adherence exceeding 90%, while 40% (n=8) had a weekly adherence of 100%. Furthermore, 9 (90%) of the 10 participants equipped with the Activity Coach app chose to enable the goal-setting functionality. The median weekly goal achievement was 83% (range 61%-96%).

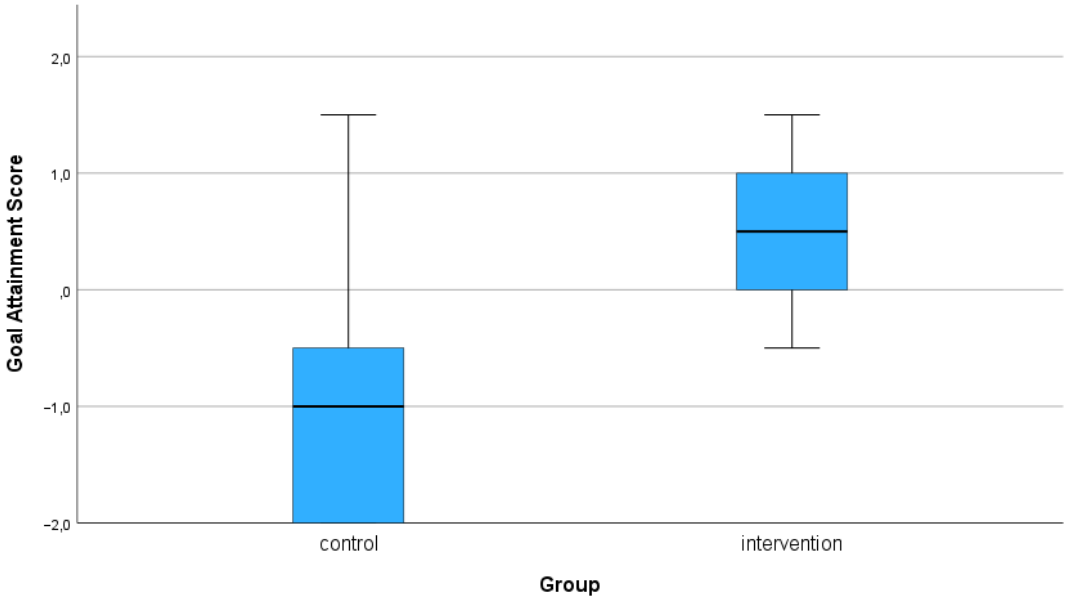
Research Question 3: Outcomes

Subjective Goal Attainment

All 10 prespecified goals were selected by at least 1 (5%) study participant. The first goal selected was distributed among the first 5 goals (see Table 1), whereas the choices for the second goal were distributed over 9 of the 10 options, with the most frequently selected goa being reduced shortness of breath. In the control group, 8 (80%) participants experienced overall worsening and 1 (10%) reported improvement, whereas in the intervention group, 7 (70%) participants reported improvement, 2 (20%) reported no difference, and 1 (10%) reported overall worsening.

The mean goal attainment score for the 2 groups was –1.0 (SD 1.1) for the control group versus 0.6 (SD 0.6) for the intervention group (P=.001). This is illustrated in Figure 2.

Figure 2. Goal attainment score by group, calculated as the mean (SD) value for the 2 selected goals.



Health-Related Quality of Life

In the control group the mean KCCQ summary score was 59 (SD 25) at baseline compared to 67 (SD 22) in the intervention group (P=0.6). At follow-up, the scores were 54 (SD 25) and 70 (SD 21) for the control and the intervention group, respectively (P=.18). ANCOVA to control for baseline differences yielded P=.04. The mean change in the overall summary score was –9 (SD 10) in the control group versus 3 (SD 13) in the intervention group (P=.03) and is illustrated in Figure 3, while the other KCCQ subscales are presented in Table 3.

XSL•FO
RenderX

Figure 3. Change in the KCCQ overall summary score by group at follow-up after 12 weeks. KCCQ: Kansas City Cardiomyopathy Questionnaire.

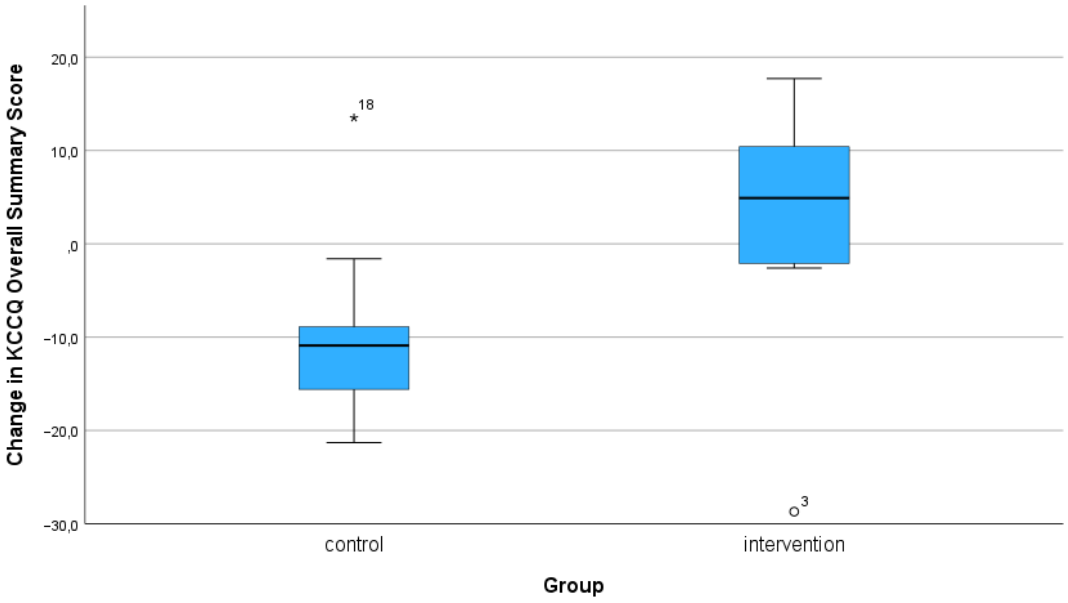


Table 3. KCCQ[a] output (summary score and subdimensions) at baseline and follow-up.

Subdimensions and time	Control group score, mean (SD)	Intervention group score, mean (SD)	*P* value
Physical limitation			
Baseline	57 (26)	65 (28)	.46
Follow-up	45 (27)	71 (20)	.04
Symptom frequency			
Baseline	57 (33)	67 (26)	.38
Follow-up	56 (26)	66 (30)	.29
QoL[b]			
Baseline	60 (19)	67 (19)	.62
Follow-up	63 (26)	68 (21)	.74
Social limitation			
Baseline	61 (33)	69 (31)	.62
Follow-up	54 (30)	74 (26)	.15
Summary score			
Baseline	59 (25)	67 (22)	.55
Follow-up	54 (25)	70 (21)	.18

[a]KCCQ: Kansas City Cardiomyopathy Questionnaire.

[b]QoL: quality of life.

Objective Physical Activity

The accelerometer-derived data are summarized in Table 4. Participants spent, on average, 76% (SD 7%) of the time in a sedentary state at baseline (control group: mean 79%, SD 6%; intervention group: mean 74%, SD 8%; *P*=.20), which at follow-up was 74% (SD 10%) of the time (control group: mean 78%, SD 8%; intervention group: mean 71%, SD 12%; *P*=.22). The decrease was mean –2.6% (SD 7%) in the intervention group versus –0.91% (SD 5%) in the control group (*P*=.59).

The average length of sedentary bouts at baseline was 408 (SD 87) minutes (control group: mean 431, SD 72 minutes; intervention group: mean 392, SD 96 minutes; *P*=.38) based on all the data recorded during the week of wearing the accelerometer. Over the course of the study, the average length of sedentary bouts increased by 27 minutes to 458 (SD 84) minutes in the control group and decreased by 0.70 minutes to 391 (SD 117) minutes in the intervention group (*P*=.22).

On average, participants walked 2735 (SD 2481) steps per day at baseline (control group: mean 2058, SD 1356 steps; intervention group: mean 3210, SD 3019 steps; *P*=.36). The

XSL•FO
RenderX

average number of steps per day increased by 861 to 2919 (SD 2303) steps in the control group and by 322 to 3532 (SD 3208) steps in the intervention group (*P*=.67). The time spent in light physical activity (LPA) increased by 8 minutes in the control group and by 42 minutes in the intervention group (*P*=.07). Moderate-to-vigorous physical activity (MVPA) and LPA data are included in Table 4.

Table 4. Overview of objectively measured physical activity obtained from the accelerometer at baseline and 12-week follow-up.

Data items and time	Control group, mean (SD)	Intervention group, mean (SD)	*P* value
Sedentary behavior (%)			
Baseline	79 (6)	74 (8)	.33
Follow-up	78 (8)	71 (12)	.33
Average length of sedentary bouts (minutes)			
Baseline	431 (72)	392 (96)	.56
Follow-up	458 (84)	391 (117)	.14
Steps per day, n			
Baseline	2058 (1356)	3210 (3019)	.33
Follow-up	2919 (2303)	3532 (3208)	.70
LPA[a] (minutes)			
Baseline	138 (47)	169 (52)	.27
Follow-up	146 (46)	207 (80)	.07
MVPA[b] (minutes)			
Baseline	15 (11)	31 (35)	.54
Follow-up	21 (24)	33 (42)	.96

[a]LPA: light physical activity.

[b]MVPA: moderate-to-vigorous physical activity.

Ancillary Analyses

Since the original trial design was modified to also include people who were not physically inactive, ancillary analyses were performed to study what impact, if any, that had on the data. No significant differences appeared. The differences are listed in Table 5.

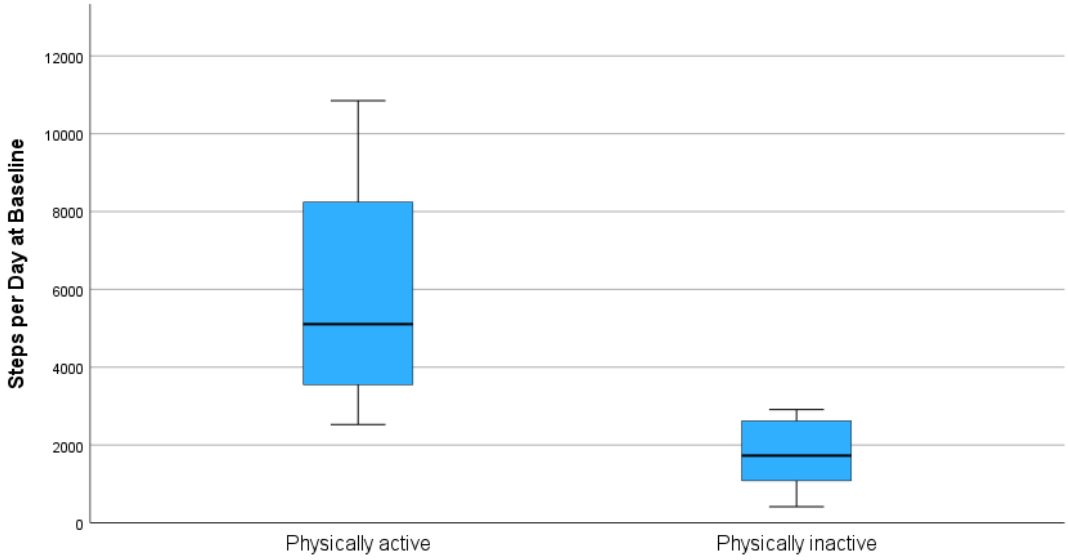
A comparison of the steps per day for the physically active versus physically inactive participants is shown in Figure 4.

Table 5. Comparison of objectively measured physical activity between groups with the original data set and the data set including only participants classified as physically inactive (n=15, 75%).

Data items and time	Control group, mean (SD)	Intervention group, mean (SD)
Steps per day at baseline, n		
Original data set	2058 (1356)	3210 (3019)
Physically inactive	1459 (814)	1952 (856)
Steps per day at follow-up, n		
Original data set	2919 (2303)	3532 (3208)
Physically inactive	2194 (1943)	2218 (1139)
Average length of sedentary bouts at baseline (minutes)		
Original data set	431 (72)	392 (96)
Physically inactive	455 (71)	406 (99)
Average length of sedentary bouts at follow-up (minutes)		
Original data set	458 (84)	391 (117)
Physically inactive	495 (55)	409 (119)

XSL•FO
RenderX

Figure 4. Accelerometer-measured steps per day for physically active versus physically inactive participants.



Discussion

Principal Findings

The results of this study show that it may be challenging to recruit physically inactive older adults with HF to participate in an mHealth study, but there might be benefits in using mHealth to support this population in achieving goals relating to physical activity and reduce sedentary behavior.

Recruitment

It is known to be challenging to recruit older people with severe illness for studies evaluating digital health technologies [27]. It seems reasonable that seeking to recruit physically inactive people makes it even more likely that they decline participation. Only 22% of the people contacted for this study were successfully recruited and randomized, although 45% initially consented to participate, but every third person withdrew consent prior to starting the trial. It might, therefore, be appropriate to investigate whether something in the recruitment process could be improved (eg, the information material or striving to initiate the trial at the same time as collecting the consent). Reasons given for withdrawal were typically negative life events (eg, having been hospitalized, having received a new diagnosis, or having a spouse whose health had deteriorated). The withdrawal rate was lower in patients who were recruited via telephone versus via mail.

The recruitment rate in this study of 22% was low compared to other studies, where the corresponding rates ranged from 70% to 90%, although these numbers come from studies in which participants were recruited in the in-patient setting [28,29]. It should also be noted that aside from the 1 participant who died, there was no attrition during the study, and all participants who started also finished the study.

Based on our experiences from this study, we conclude that researchers must be active and may have to call up each possible participant and that merely sending out letters via regular mail is likely insufficient.

Implementation Fidelity

To be able to judge the quality of a study and to properly replicate it, some reporting of intervention fidelity is required, and examples of aspects of intervention fidelity to report are intervention design and intervention delivery or receipt of the intervention [30]. The intervention design has been published elsewhere [9], and a qualitative study to investigate user experiences was beyond the scope of this study. An acceptable intervention delivery ensures that results from the intervention are generalizable [31]. Delivery of the intervention can be reported as intervention adherence [18], which is what we reported in this manuscript. Other mHealth studies have used 60% adherence as a cutoff point for acceptable level of use [17,32], and that level was also adopted in the development of the Activity Coach app [9]. In this study, the median adherence was 69%, which is higher than what is normally reported for mHealth interventions [19] and also higher than that in the Activity Coach feasibility study, suggesting an improvement in usability following the final adjustments made to the Activity Coach app after the feasibility study [9]. The 2022 systematic review by Jakob et al [19] concluded that factors that drive adherence are tailoring of content to the user and gamification, both of which are present in the Activity Coach app and may explain the better-than-average adherence. What level of adherence to an intervention is acceptable should depend on which intervention is being studied. The current guidelines on physical activity from the World Health Organization (WHO) for the general population [33], as well as guidelines for people with HF [34], make recommendations on a weekly basis, so it may be reasonable to also assess adherence to a tool supporting physical activity on a weekly basis; in this study, weekly adherence to the Activity Coach app was 88%, which is similar to other rates in the published literature, which ranged from 55% to 80%, depending on the definition of adherence [29,35-39].

XSL•FO
RenderX

Outcomes

Plausibility of Outcomes

Questionnaires were completed by 19 of the 20 participants, and system (Activity Coach) data from all 10 participants in the intervention group and accelerometer data from 17 of the 20 participants passed wear time scrutiny. The use of a prespecified list of personal goals to choose from, which was a learning from a feasibility study [9], worked well, and each participant successfully selected 2 goals and could grade them at follow-up. All data were analyzable and of acceptable quality.

Limited Efficacy Testing

Goal attainment is an important part of person-centered care [40,41] and is also more responsive to change than other measures [42]. The 5-point scale we used is considered the norm when using goal attainment as an outcome [43,44]. There was a significant difference in goal attainment between the 2 groups, and although there was a subtle improvement in the intervention group, there was a more pronounced worsening in the control group.

It is difficult to assess the clinical importance of a goal attainment scaling outcome, but a 25% shift on an ordinal scale has previously been used as a clinically meaningful change [41,45]. The shift in this study was of a magnitude of 32% of the used scale. In the light of improvements in the HR-QoL, this might imply that there is a clinically meaningful effect.

The KCCQ results mirrored the subjective goal attainment results, in that the KCCQ summary score showed a 12-point difference between the 2 groups, expressed as a 5% improvement in the intervention group and a 15% worsening in the control group. This difference is of a similar size as effects reported from another mHealth intervention aimed at increasing physical activity, which reported a 6-point difference, also at 12-week follow-up [46]. Other telehealth interventions, albeit not strict mHealth, have reported smaller or no improvements in the KCCQ score [47-50]. The difference between the 2 groups at baseline in the summary score was large, albeit nonsignificant, and when that difference was corrected for, the difference between the 2 groups at follow-up also became significant.

Looking at the subscales, the physical limitation dimension was the subdomain with the largest effect, where the intervention group at follow-up had a 37% higher score. A change in that subscale exceeding 8.33 points has been identified as the minimal clinically important difference for improvement and a decrease of 4.17 or more indicates deterioration [51]. In this study, the score of the intervention group increased by 6 points and that of the control group decreased by 12 points, which suggests a clinically meaningful deterioration in the control group.

There was a significant difference of 4135 in the steps per day between the physically inactive versus active participants, which needs to be considered when interpreting the results from the objective measurements of physical activity. The change in the time spent in sedentary behavior was of a magnitude similar to that in other reports on physical activity interventions [52]. The differences between the 2 study arms were more pronounced after removing the physically active participants, with an increase of 40 minutes in the control group compared to an increase of 3 minutes in the intervention group.

Longer sedentary bouts are associated with poor health outcomes [53-55]. For these effects to manifest into outcomes, the reduction in sedentary time needs to be shifted to MVPA according to Rossen et al [56], which was, however, disputed by Diaz et al [57]. There is still a knowledge gap regarding the impact of the sedentary bout length and the effects of changing it [58].

There was no significant difference in the steps per day between the 2 groups, neither at baseline nor at follow-up, although the number of steps per day increased more in the control group (861 steps) compared to the intervention group (322 steps). The increase in steps in the control group is large enough to potentially be clinically meaningful [59], although successful interventions aimed at increasing the step count in chronically ill older adults usually report larger effects (eg, mean ~2000, SD 500, more steps per day or above) [28,60,61]. It is not possible to say whether this effect on the control group was a consequence of the small sample size or external factors. Perhaps there is a connection between more steps per day and longer sedentary bouts. Other research has also proven the positive effects of mHealth interventions on physical activity outcomes [62].

Overall, it appears that the Activity Coach app does not improve the outcomes studied, but it may offer protection against deterioration, as the control group experienced worsened situations relating to their individual goals, a significantly worse HR-QoL, and a trend toward increased length of sedentary bouts. The Activity Coach app does not necessary tell the user to go out and walk, but it focuses more on subtle suggestions, such as standing up a couple of times or doing leg kicks during television commercials, shifting the balance between the feet, doing calf raises while doing the dishes, or using an exercise band. Assuming this is adhered to, it would explain a perceived increase in exercise and improvements in the HR-QoL and goal attainment, without necessarily showing up on the step counts. Possibly, it is manifested in the shorter average lengths of sedentary bouts.

Combining the recruitment rate with the effect sizes on the outcomes measured showed that a future efficacy study designed in the same way, with the primary endpoint being either the KCCQ summary score or sedentary time, needs to contact approximately 250-350 people to be sufficiently powerful.

Limitations

The original study design was modified, and as seen in the ancillary analyses, there were significant differences in daily physical activity between physically active and physically inactive participants, which makes the results more difficult to interpret.

Fewer participants than desired were included, and even though the study was not designed based on specific power calculations, a larger included population would have provided better grounds for designing an efficacy trial, a more well-grounded sample size estimation, and improved generalizability.

XSL•FO
RenderX

Conclusion

The recruitment rate of this study, where older adults with HF were to evaluate an mHealth tool, was 22%, which is a low rate. An active recruitment process is advised if a future efficacy study is to be performed. Adherence to the Activity Coach intervention was better than average at 69%. Deterioration in health and the QoL may be attenuated by the Activity Coach app, possibly by increasing certain aspects of physical activity and by reducing the length of sedentary bouts.

Acknowledgments

We would like to acknowledge the contribution of heart failure nurses and medical doctors working at the participating primary health care centers, who were helpful in reviewing the electronic health records of recruited patients. This study received no external funding.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

AB, MB, AS, and TJ handled the conceptualization of the study. AB, MB, LK, AS, and TJ managed the methodology. AB, LK, and TJ performed validation. AB was responsible for the software, formal analysis, investigation, writing the original draft of the manuscript, visualization, and project administration. LK and TJ handled the resources. MB, LK, AS, and TJ managed the review and editing of the manuscript and study supervision. All authors have read and agreed to the published version of the manuscript. In addition, TJ assumed responsibility for the paper, including for the accuracy and appropriateness of the reference list.

Conflicts of Interest

AB is an employee of CareLigo AB, the company behind Optilogg. An independent researcher at Linköping University entered all the data.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1165 KB - formative_v9i1e62910_app1.pdf]

References

1. Savarese G, Becher P, Lund L, Seferovic P, Rosano G, Coats A. Global burden of heart failure: a comprehensive and updated review of epidemiology. Cardiovasc Res 2023 Jan 18;118(17):3272-3287. [doi: 10.1093/cvr/cvac013] [Medline: 35150240]
2. Inamdar A, Inamdar A. Heart failure: diagnosis, management and utilization. J Clin Med 2016 Jun 29;5(7):62 [FREE Full text] [doi: 10.3390/jcm5070062] [Medline: 27367736]
3. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, SBRN Terminology Consensus Project Participants. Sedentary behavior research network (SBRN)–terminology consensus project process and outcome. Int J Behav Nutr Phys Act 2017 Jun 10;14(1):1-17 [FREE Full text] [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]
4. Thivel D, Tremblay A, Genin PM, Panahi S, Rivière D, Duclos M. Physical activity, inactivity, and sedentary behaviors: definitions and implications in occupational health. Front Public Health 2018;6:288 [FREE Full text] [doi: 10.3389/fpubh.2018.00288] [Medline: 30345266]
5. Jones NR, Roalfe AK, Adoki I, Hobbs FR, Taylor CJ. Survival of patients with chronic heart failure in the community: a systematic review and meta-analysis. Eur J Heart Fail 2019 Nov 16;21(11):1306-1325 [FREE Full text] [doi: 10.1002/ejhf.1594] [Medline: 31523902]
6. Doukky R, Mangla A, Ibrahim Z, Poulin M, Avery E, Collado FM, et al. Impact of physical inactivity on mortality in patients with heart failure. Am J Cardiol 2016 Apr 01;117(7):1135-1143 [FREE Full text] [doi: 10.1016/j.amjcard.2015.12.060] [Medline: 26853954]
7. Park LG, Dracup K, Whooley MA, McCulloch C, Lai S, Howie-Esquivel J. Sedentary lifestyle associated with mortality in rural patients with heart failure. Eur J Cardiovasc Nurs 2019 Apr 21;18(4):318-324. [doi: 10.1177/1474515118822967] [Medline: 30663898]
8. Zhao Q, Chen C, Zhang J, Ye Y, Fan X. Sedentary behavior and health outcomes in patients with heart failure: a systematic review and meta-analysis. Heart Fail Rev 2022 Jul 22;27(4):1017-1028. [doi: 10.1007/s10741-021-10132-7] [Medline: 34159521]

XSL•FO
RenderX

9. Blomqvist A, Bäck M, Klompstra L, Strömberg A, Jaarsma T. Usability and feasibility analysis of an mHealth-tool for supporting physical activity in people with heart failure. BMC Med Inform Decis Mak 2024 Feb 12;24(1):44 [FREE Full text] [doi: 10.1186/s12911-024-02452-z] [Medline: 38347499]
10. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061. [doi: 10.1136/bmj.n2061] [Medline: 34593508]
11. Thabane L, Ma J, Chu R, Cheng J, Ismaila A, Rios LP, et al. A tutorial on pilot studies: the what, why and how. BMC Med Res Methodol 2010 Jan 06;10:1 [FREE Full text] [doi: 10.1186/1471-2288-10-1] [Medline: 20053272]
12. Lancaster GA. Pilot and feasibility studies come of age!. Pilot Feasibility Stud 2015;1(1):1 [FREE Full text] [doi: 10.1186/2055-5784-1-1] [Medline: 29611687]
13. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS consensus group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016 Oct 24;355:i5239 [FREE Full text] [doi: 10.1136/bmj.i5239] [Medline: 27777223]
14. Blomqvist A, Bäck M, Klompstra L, Strömberg A, Jaarsma T. Utility of single-item questions to assess physical inactivity in patients with chronic heart failure. ESC Heart Fail 2020 Aug 06;7(4):1467-1476 [FREE Full text] [doi: 10.1002/ehf2.12709] [Medline: 32372549]
15. Jaarsma T, Strömberg A, Mårtensson J, Dracup K. Development and testing of the European Heart Failure Self-Care Behaviour Scale. Eur J Heart Fail 2003 Jun 09;5(3):363-370 [FREE Full text] [doi: 10.1016/s1388-9842(02)00253-2] [Medline: 12798836]
16. Sahlin D, Rezanezad B, Edvinsson ML, Bachus E, Melander O, Gerward S. Self-care management intervention in heart failure (SMART-HF)-a multicentre randomized controlled trial. J Card Fail 2022 Jan;28(1):3-12 [FREE Full text] [doi: 10.1016/j.cardfail.2021.06.009] [Medline: 34161807]
17. Melin M, Hägglund E, Ullman B, Persson H, Hagerman I. Effects of a tablet computer on self-care, quality of life, and knowledge: a randomized clinical trial. J Cardiovasc Nurs 2018;33(4):336-343. [doi: 10.1097/JCN.0000000000000462] [Medline: 29369123]
18. Lambert JD, Greaves CJ, Farrand P, Cross R, Haase AM, Taylor AH. Assessment of fidelity in individual level behaviour change interventions promoting physical activity among adults: a systematic review. BMC Public Health 2017 Oct 02;17(1):765 [FREE Full text] [doi: 10.1186/s12889-017-4778-6] [Medline: 28969669]
19. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
20. Urach S, Gaasterland C, Posch M, Jilma B, Roes K, Rosenkranz G, et al. Statistical analysis of goal attainment scaling endpoints in randomised trials. Stat Methods Med Res 2019 Jun;28(6):1893-1910 [FREE Full text] [doi: 10.1177/0962280218777896] [Medline: 29921167]
21. Spertus JA, Jones PG. Development and validation of a short version of the Kansas City Cardiomyopathy Questionnaire. Circ: Cardiovasc Qual Outcomes 2015 Sep;8(5):469-476. [doi: 10.1161/circoutcomes.115.001958]
22. Patel H, Ekman I, Spertus JA, Wasserman SM, Persson L. Psychometric properties of a Swedish version of the Kansas City Cardiomyopathy Questionnaire in a chronic heart failure population. Eur J Cardiovasc Nurs 2008 Sep;7(3):214-221. [doi: 10.1016/j.ejcnurse.2007.08.005] [Medline: 18060843]
23. Loyen A, Clarke-Cornwell AM, Anderssen SA, Hagströmer M, Sardinha LB, Sundquist K, et al. Sedentary time and physical activity surveillance through accelerometer pooling in four European countries. Sports Med 2017 Jul 10;47(7):1421-1435 [FREE Full text] [doi: 10.1007/s40279-016-0658-y] [Medline: 27943147]
24. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. J Sci Med Sport 2011 Sep;14(5):411-416. [doi: 10.1016/j.jsams.2011.04.003] [Medline: 21616714]
25. Sim J, Lewis M. The size of a pilot study for a clinical trial should be calculated in relation to considerations of precision and efficiency. J Clin Epidemiol 2012 Mar;65(3):301-308. [doi: 10.1016/j.jclinepi.2011.07.011] [Medline: 22169081]
26. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharm Stat 2005 Nov 24;4(4):287-291. [doi: 10.1002/pst.185]
27. Poli A, Kelfve S, Motel-Klingebiel A. A research tool for measuring non-participation of older people in research on digital health. BMC Public Health 2019 Nov 08;19(1):1487 [FREE Full text] [doi: 10.1186/s12889-019-7830-x] [Medline: 31703655]
28. Saleh ZT, Elshatarat RA, Elhefnawy KA, Helmi Elneblawi N, Abu Raddaha AH, Al-Za'areer MS, et al. Effect of a home-based mobile health app intervention on physical activity levels in patients with heart failure: a randomized controlled trial. J Cardiovasc Nurs 2022 Apr 6;38(2):128-139. [doi: 10.1097/jcn.0000000000000911] [Medline: 00005082-202303000-00008]
29. Indraratna P, Biswas U, McVeigh J, Mamo A, Magdy J, Vickers D, et al. A smartphone-based model of care to support patients with cardiac disease transitioning from hospital to the community (TeleClinical care): pilot randomized controlled trial. JMIR Mhealth Uhealth 2022 Feb 28;10(2):e32554 [FREE Full text] [doi: 10.2196/32554] [Medline: 35225819]

XSL•FO
RenderX

30. Murphy SL, Gutman SA. Intervention fidelity: a necessary aspect of intervention effectiveness studies. Am J Occup Ther 2012;66(4):387-388. [doi: 10.5014/ajot.2010.005405] [Medline: 22742685]
31. Evans E, Zengul A, Subhash Chilke T, Knight A, Willig A, Cherrington A, et al. Telehealth behavioral intervention for chronic disease self-management in adults with physical disabilities (My Health, My Life, My Way): protocol for intervention fidelity and dashboard design for a randomized controlled trial. JMIR Res Protoc 2024 Feb 12;13:e53410 [FREE Full text] [doi: 10.2196/53410] [Medline: 38345845]
32. Lyngå P, Persson H, Hägg-Martinell A, Hägglund E, Hagerman I, Langius-Eklöf A, et al. Weight monitoring in patients with severe heart failure (WISH). A randomized controlled trial. Eur J Heart Fail 2012 Apr 18;14(4):438-444 [FREE Full text] [doi: 10.1093/eurjhf/hfs023] [Medline: 22371525]
33. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020 Dec 25;54(24):1451-1462 [FREE Full text] [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
34. McDonagh T, Metra M, Adamo M, Gardner R, Baumbach A, Böhm M. Correction to: 2023 focused update of the 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure: developed by the task force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J 2024 Jan 01;45(1):53-726. [doi: 10.1093/eurheartj/ehad613] [Medline: 37995350]
35. Vuorinen A, Leppänen J, Kaijanranta H, Kulju M, Heliö T, van Gils M, et al. Use of home telemonitoring to support multidisciplinary care of heart failure patients in Finland: randomized controlled trial. J Med Internet Res 2014 Dec 11;16(12):e282 [FREE Full text] [doi: 10.2196/jmir.3651] [Medline: 25498992]
36. Choi E, Park J, Min D, Ahn S, Ahn J. Heart Failure-Smart Life: a randomized controlled trial of a mobile app for self-management in patients with heart failure. BMC Cardiovasc Disord 2023 Jan 09;23(1):7 [FREE Full text] [doi: 10.1186/s12872-023-03039-8] [Medline: 36624373]
37. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. N Engl J Med 2010 Dec 09;363(24):2301-2309 [FREE Full text] [doi: 10.1056/NEJMoa1010029] [Medline: 21080835]
38. Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, TIM-HF Investigators. Telemedical interventional monitoring in heart failure (TIM-HF), a randomized, controlled intervention trial investigating the impact of telemedicine on mortality in ambulatory patients with heart failure: study design. Eur J Heart Fail 2010 Dec;12(12):1354-1362 [FREE Full text] [doi: 10.1093/eurjhf/hfq199] [Medline: 21098580]
39. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Mobile phone-based telemonitoring for heart failure management: a randomized controlled trial. J Med Internet Res 2012 Feb 16;14(1):e31 [FREE Full text] [doi: 10.2196/jmir.1909] [Medline: 22356799]
40. Giovannetti ER, Clair CA, Jennings LA, Sandberg SF, Bowman A, Reuben DB, et al. Standardised approach to measuring goal-based outcomes among older disabled adults: results from a multisite pilot. BMJ Qual Saf 2021 Feb 09;30(2):157-166. [doi: 10.1136/bmjqs-2019-010742] [Medline: 33037142]
41. Rietkerk W, Uittenbroek RJ, Gerritsen DL, Slaets JPJ, Zuidema SU, Wynia K. Goal planning in person-centred care supports older adults receiving case management to attain their health-related goals. Disabil Rehabil 2021 Jun 07;43(12):1682-1691. [doi: 10.1080/09638288.2019.1672813] [Medline: 31589075]
42. Rockwood K, Joyce B, Stolee P. Use of goal attainment scaling in measuring clinically important change in cognitive rehabilitation patients. J Clin Epidemiol 1997 May;50(5):581-588. [doi: 10.1016/s0895-4356(97)00014-0] [Medline: 9180650]
43. Logan B, Jegatheesan D, Viecelli A, Pascoe E, Hubbard R. Goal attainment scaling as an outcome measure for randomised controlled trials: a scoping review. BMJ Open 2022 Jul 22;12(7):e063061 [FREE Full text] [doi: 10.1136/bmjopen-2022-063061] [Medline: 35868829]
44. Roberts MA, Abery BH. A person-centered approach to home and community-based services outcome measurement. Front Rehabil Sci 2023 Feb 1;4:1056530 [FREE Full text] [doi: 10.3389/fresc.2023.1056530] [Medline: 36817716]
45. Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001 Nov;94(2):149-158. [doi: 10.1016/S0304-3959(01)00349-9] [Medline: 11690728]
46. Felker GM, Sharma A, Mentz RJ, She L, Green CL, Granger BB, et al. A randomized controlled trial of mobile health intervention in patients with heart failure and diabetes. J Card Fail 2022 Nov;28(11):1575-1583. [doi: 10.1016/j.cardfail.2022.07.048] [Medline: 35882260]
47. Bakitas MA, Dionne-Odom JN, Ejem DB, Wells R, Azuero A, Stockdill ML, et al. Effect of an early palliative care telehealth intervention vs usual care on patients with heart failure: the ENABLE CHF-PC randomized clinical trial. JAMA Intern Med 2020 Sep 01;180(9):1203-1213 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2861] [Medline: 32730613]
48. Collins SP, Liu D, Jenkins CA, Storrow AB, Levy PD, Pang PS, et al. Effect of a self-care intervention on 90-day outcomes in patients with acute heart failure discharged from the emergency department: a randomized clinical trial. JAMA Cardiol 2021 Feb 01;6(2):200-208 [FREE Full text] [doi: 10.1001/jamacardio.2020.5763] [Medline: 33206126]

XSL•FO
RenderX

49. Nolan RP, Ross HJ, Farkouh ME, Huszti E, Chan S, Toma M, et al. Automated e-counseling for chronic heart failure: CHF-CePPORT trial. Circ Heart Fail 2021 Jan;14(1):e007073. [doi: 10.1161/CIRCHEARTFAILURE.120.007073] [Medline: 33464959]
50. Johnson AE, Routh S, Taylor CN, Leopold M, Beatty K, McNamara DM, et al. Developing and implementing an mHealth heart failure self-care program to reduce readmissions: randomized controlled trial. JMIR Cardio 2022 Mar 21;6(1):e33286 [FREE Full text] [doi: 10.2196/33286] [Medline: 35311679]
51. Butler J, Spertus J, Bamber L, Khan M, Roessig L, Vlajnic V, VITALITY-HFpEF Study Group. Defining changes in physical limitation from the patient perspective: insights from the VITALITY-HFpEF randomized trial. Eur J Heart Fail 2022 May;24(5):843-850 [FREE Full text] [doi: 10.1002/ejhf.2481] [Medline: 35274420]
52. Swartz AM, Cho CC, Welch WA, Widlansky ME, Maeda H, Strath SJ. Pattern analysis of sedentary behavior change after a walking intervention. Am J Health Behav 2018 May 01;42(3):90-101 [FREE Full text] [doi: 10.5993/AJHB.42.3.9] [Medline: 29663984]
53. Diaz KM, Duran AT, Colabianchi N, Judd SE, Howard VJ, Hooker SP. Potential effects on mortality of replacing sedentary time with short sedentary bouts or physical activity: a national cohort study. Am J Epidemiol 2019 Mar 01;188(3):537-544 [FREE Full text] [doi: 10.1093/aje/kwy271] [Medline: 30551177]
54. Healy GN, Dunstan DW, Salmon J, Cerin E, Shaw JE, Zimmet PZ, et al. Breaks in sedentary time: beneficial associations with metabolic risk. Diabetes Care 2008 Apr;31(4):661-666. [doi: 10.2337/dc07-2046] [Medline: 18252901]
55. Duran AT, Friel CP, Serafini MA, Ensari I, Cheung YK, Diaz KM. Breaking up prolonged sitting to improve cardiometabolic risk: dose-response analysis of a randomized crossover trial. Med Sci Sports Exerc 2023 May 01;55(5):847-855. [doi: 10.1249/MSS.0000000000003109] [Medline: 36728338]
56. Rossen J, Buman MP, Johansson U, Yngve A, Ainsworth B, Brismar K, et al. Reallocating bouted sedentary time to non-bouted sedentary time, light activity and moderate-vigorous physical activity in adults with prediabetes and type 2 diabetes. PLoS One 2017 Jul 28;12(7):e0181053 [FREE Full text] [doi: 10.1371/journal.pone.0181053] [Medline: 28753626]
57. Diaz K, Duran A, Colabianchi N, Judd S, Howard V, Hooker S. Effects of replacing prolonged sedentary bouts with short sedentary bouts or physical activity on mortality: 2552 June 1 2. Med Sci Sports Exerc 2018;50(5S):634. [doi: 10.1249/01.mss.0000537180.26323.85]
58. Byrom B, Stratton G, Mc Carthy M, Muehlhausen W. Objective measurement of sedentary behaviour using accelerometers. Int J Obes (Lond) 2016 Nov;40(11):1809-1812 [FREE Full text] [doi: 10.1038/ijo.2016.136] [Medline: 27478922]
59. Demeyer H, Burtin C, Hornikx M, Camillo CA, Van Remoortel H, Langer D, et al. The minimal important difference in physical activity in patients with COPD. PLoS One 2016 Apr 28;11(4):e0154587 [FREE Full text] [doi: 10.1371/journal.pone.0154587] [Medline: 27124297]
60. Martin SS, Feldman DI, Blumenthal RS, Jones SR, Post WS, McKibben RA, et al. mActive: a randomized clinical trial of an automated mhealth intervention for physical activity promotion. JAHA 2015 Oct 29;4(11):e002239. [doi: 10.1161/jaha.115.002239]
61. Hemnes AR, Silverman-Lloyd LG, Huang S, MacKinnon G, Annis J, Whitmore CS, et al. A mobile health intervention to increase physical activity in pulmonary arterial hypertension. Chest 2021 Sep;160(3):1042-1052 [FREE Full text] [doi: 10.1016/j.chest.2021.04.012] [Medline: 33878341]
62. Tong HL, Alnasser A, Alshahrani NZ, Bawaked RA, AlAhmed R, Alsukait RF, et al. The use of mobile technologies to promote physical activity and reduce sedentary behaviors in the Middle East and North Africa region: systematic review and meta-analysis. J Med Internet Res 2024 Mar 19;26:e53651 [FREE Full text] [doi: 10.2196/53651] [Medline: 38502160]

Abbreviations

EHR: electronic health record
GUI: graphical user interface
HF: heart failure
HR-QoL: health-related quality of life
KCCQ: Kansas City Cardiomyopathy Questionnaire
LPA: light physical activity
MRC: Medical Research Council
MVPA: moderate-to-vigorous physical activity
QoL: quality of life
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by A Mavragani; submitted 04.06.24; peer-reviewed by P Dunn, A Zelko, R Austin; comments to author 22.10.24; revised version received 06.11.24; accepted 29.11.24; published 08.01.25.

Please cite as:
Blomqvist A, Bäck M, Klompstra L, Strömberg A, Jaarsma T
Testing the Recruitment Frequency, Implementation Fidelity, and Feasibility of Outcomes of the Heart Failure Activity Coach Study (HEALTHY): Pilot Randomized Controlled Trial
JMIR Form Res 2025;9:e62910
URL: https://formative.jmir.org/2025/1/e62910
doi:10.2196/62910
PMID:

©Andreas Blomqvist, Maria Bäck, Leonie Klompstra, Anna Strömberg, Tiny Jaarsma. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effectiveness of a Videoconference-Based Cognitive Behavioral Therapy Program for Patients with Schizophrenia: Pilot Randomized Controlled Trial

Masayuki Katsushima[1,2], PhD; Hideki Nakamura[3,4], MSc; Yuki Shiko[5,6], PhD; Hideki Hanaoka[6], MD, PhD; Eiji Shimizu[2,4,7], MD, PhD

[1]Department of Rehabilitation, Faculty of Health Care and Medical Sports, Teikyo Heisei University, Ichihara, Japan

[2]Research Center for Child Mental Development, Chiba University, Chiba, Japan

[3]Department of Nursing, Faculty of Medicine, The Jikei University School of Medicine, Tokyo, Japan

[4]Department of Cognitive Behavioral Physiology, Chiba University Graduate School of Medicine, Chiba, Japan

[5]Department of Biostatistics, Graduate School of Medicine, Saitama Medical University, Saitama, Japan

[6]Clinical Research Center, Chiba University Hospital, Chiba, Japan

[7]Cognitive Behavioral Therapy Center, Chiba University Hospital, Chiba, Japan

Corresponding Author:
Masayuki Katsushima, PhD
Department of Rehabilitation, Faculty of Health Care and Medical Sports
Teikyo Heisei University
4-1 UruidoMinami
Ichihara, 290-0193
Japan
Phone: 81 436 74 6915
Fax: 81 436 74 6915
Email: m.katsushima@thu.ac.jp

Abstract

Background: Cognitive behavioral therapy for psychosis (CBTp) is not sufficiently widespread in clinical practice, although evidence has been presented.

Objective: The purpose of this study was to explore whether one-on-one videoconference-based CBTp (vCBTp) is more effective than usual care (UC) alone for improving psychiatric symptoms in patients with schizophrenia attending outpatient clinics.

Methods: In this exploratory randomized controlled trial, patients with schizophrenia and schizoaffective disorders who were still taking medication in an outpatient clinic were randomly assigned to either the vCBTp plus UC group (n=12) or the UC group (n=12). The vCBTp program was conducted once a week, with each session lasting for 50 minutes, for a total of 7 sessions conducted in real-time and in a one-on-one format remotely using a loaned tablet computer (iPad). The primary outcome was the Positive and Negative Syndrome Scale (PANSS) total score, which measures the difference in the mean change from baseline (week 0) to posttest (week 8).

Results: The study included 24 participants. There were no significant differences between the 2 groups at baseline. With regard to significant differences between the 2 groups in terms of the primary outcome, the mean change in the PANSS total score from baseline to week 8 in the vCBTp plus UC group was –9.5 (95% CI –12.09 to –6.91) and the mean change in the UC alone group was 6.9 (95% CI 1.54-12.30). The difference between the 2 groups was significant (P<.001). In addition, significant improvements were observed in the subscales of positive (P<.001) and negative (P=.004) symptoms and general psychopathology (P<.001). Significant differences were also observed in the secondary outcomes of the General Anxiety Disorder-7 (GAD-7; P=.04) and EQ-5D-5L (P=.005). There were no dropouts and no serious adverse events in this study.

Conclusions: A total of 7 remote vCBTp sessions conducted in the vCBTp plus UC group could be safely administered to patients with schizophrenia. They were also observed to be effective for psychiatric symptoms, general anxiety, and quality of life. However, because of the observed worsening of scores in the UC group, caution is required in interpreting significant differences between the 2 groups. This approach is expected to improve accessibility to CBTp for outpatients with schizophrenia

XSL•FO
RenderX

and social anxiety regarding transportation use and financial and physical burdens related to transportation, and to contribute to promoting CBTp acceptability by compensating for the shortage of implementers.

Trial Registration: University Hospital Medical Information Network Clinical Trials Registry UMIN000043396; https://center6.umin.ac.jp/cgi-open-bin/ctr/ctr_view.cgi?recptno=R000049544

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2022-069734

(JMIR Form Res 2025;9:e59540) doi:10.2196/59540

KEYWORDS

schizophrenia; randomized controlled trial; cognitive behavioral therapy; videoconference; remote therapy

Introduction

Background

Schizophrenia is a psychiatric disorder that affects approximately 1% of the total population and is typically prevalent between the ages of 16 and 30 years [1,2]. Pharmacotherapy with antipsychotic drugs is the standard treatment for schizophrenia. Moreover, cognitive behavioral therapy, a psychological intervention, is recommended by the National Institute of Health and Clinical Excellence (NICE) [3]. In the West, cognitive behavioral therapy for psychosis (CBTp) has shown promise in improving positive symptoms and acquiring coping skills [4]. A meta-analysis of CBTp reported a moderate effect of 0.35-0.44 for positive and generalized symptoms [5,6]. Despite the effectiveness of traditional face-to-face CBTp in treating schizophrenia, progress in training therapists to conduct CBTp has been slow [7,8]. The Schizophrenia Commission in the United Kingdom reported that only 10% of patients with schizophrenia receive CBTp due to limited access [9]. Therefore, despite supporting evidence, it is infrequently implemented in clinical practice owing to a shortage of practitioners, making it difficult for patients to receive CBTp [10]. It is thus desirable to create an environment in which effective CBTp can be implemented in the future and increase the number of clinicians who can practice. Evidence of low-intensity CBTp has also been collected, suggesting its efficacy [11,12].

Previous studies have found that community psychiatric nurses effectively increased their awareness of illnesses by conducting as few as 6 sessions of CBTp [13,14]. The 6 CBTp sessions improved their understanding of illness, general health, and depression. NICE recommends a minimum of 16 sessions of CBTp [3]. However, a meta-analysis by Hazell et al [11] reported an effect size of 0.46 for CBTp conducted in 6-15 sessions.

In addition, some studies have reported the incorporation of new theories, such as mindfulness, acceptance and commitment therapy, compassion mind training, and metacognitive therapy, into CBTp [15]. Thus, in recent years, studies have been conducted that have not limited themselves to formulated CBTp but have adjusted the frequency and content of CBTp.

In this context, studies reported that patients with schizophrenia have traumatic memories and may be treated with CBT [16]. However, CBTp does not focus directly on the sequelae of trauma, and therapists are concerned that experiential reprocessing of trauma may exacerbate psychotic symptoms; therefore, the treatment of posttraumatic disorder (PTSD) symptoms often involves reluctance [17]. Nevertheless, 3 reviews [18-20] suggested that trauma-focused psychotherapy is safe and effective for psychotic patients.

In recent years, research on internet-based CBTp has increased [21-25], and a growing interest exists for CBT delivered through a videoconference system called vCBT [26,27]. This method enhances access to CBT for patients in remote areas and enables teletherapy through 2-way real-time communication between the therapist and patient. Therapist-led internet-based CBT has been demonstrated to be equally effective as face-to-face CBT in treating anxiety, depression, insomnia, and physical disorders [28].

There have also been several previous studies targeting schizophrenia. A CBTp effort in a group therapy format using a videoconference method for early psychosis reported positive results regarding efficacy [29]. Moreover, a study of remote sessions based on acceptance commitment therapy in a group therapy format mentioned the potential for both accessibility and recovery effects [30]. In addition, there have been reports of neurocognitive and metacognitive training delivered remotely as group therapy [31]. However, there is still little to no evidence regarding randomized controlled trials (RCTs) of CBTp interventions in a one-on-one videoconference format.

Therefore, we developed a real-time and one-on-one videoconference-based CBTp (vCBTp) program for schizophrenia through an RCT design [32]. The vCBTp program included 2 sessions that addressed intrusive memories of stressful events related to the onset of psychosis.

Objective

This study examined the results and efficacy of an exploratory RCT designed to compare the clinical efficacy of 7 sessions of vCBTp as an adjunct to usual care (UC) versus UC alone. The study included patients with persistent positive symptoms of schizophrenia despite pharmacotherapy.

Methods

Study Design and Participants

This study was designed as a single-center, 2-arm, parallel, rater-blinded, prospective RCT with a 7-week treatment regimen. Study participants were assigned to either the vCBTp plus UC or UC alone groups. The recruitment of participants was posted on the Chiba University Graduate School of Medicine's Cognitive Behavioral Physiology subject recruitment

XSL•FO
RenderX

website. Recruitment posters were displayed in the outpatient waiting room of Chiba University Hospital. Recruitment posters and leaflets were also mailed to medical institutions and employment support facilities in Chiba Prefecture.

Ethical Considerations

All procedures were approved by the Institutional Review Board of Chiba University Hospital in January 2021 (reference number: G2020031). All participants received a thorough verbal and written explanation of the study, and verbal and written consent was obtained from those who were confirmed to have the capacity to consent to the study and who wished to participate. Participation in the clinical trial was voluntary on the part of the subjects and their surrogates, and they were fully informed that they could refuse or withdraw from participation in the clinical trial at any time. The consent obtained also covered that no additional consent would be required for secondary analyses.

Data were anonymized so that they could be identified only by a unique identification number and were maintained at the data center. Consent documents and other materials related to the study were strictly maintained in a locked archive.

Study participants were paid a burden-reduction fee (10,000 JPY [US $65] gift card at baseline and 10,000 JPY [US $65] gift card at the endpoint for a total of 20,000 JPY [US $130]) for assessments at 0 and 8 weeks. This burden-reduction fee was paid equitably to both the vCBTp plus UC group and the UC group.

Inclusion Criteria

The criteria for those eligible to participate in the study were as follows: age 16-65 years; primary diagnosis of schizophrenia, schizoaffective disorder, or delusional disorder according to the Diagnostic and Statistical Manual of Mental Disorders, 5th edition; and satisfactory assessment of competence to provide informed consent using the MacArthur Competence Assessment Tool for Clinical Research (MacCAT-CR) [33,34]. The MacCAT-CR assessed participants' ability to "understand," "recognize," "think validly," and "express choice" about the study. Participants who responded with a score of 0 on any of the items were judged to have poor capacity to consent and were eligible for exclusion. In addition, participants needed to have at least 3 points on any one of the positive symptoms of the Positive and Negative Syndrome Scale (PANSS) [35]. In addition, participants had to be attending their primary psychiatrist, be taking antipsychotic medication for at least 3 months continuously, and not intend to change their medication within the next 3 months. Participants were also required to obtain consent from their primary psychiatrist to participate in the study. An internet environment with wireless fidelity (Wi-Fi) was needed for videoconference using a tablet computer (iPad).

Exclusion Criteria

Those with a diagnosis of alcoholism, drug dependence, intellectual disability, or dementia from their primary psychiatrist were excluded from participation in the study. Those who were predicted to be at risk of self-harm or harm as a result of the clinical trial were also excluded. Risk of self-harm or harm was assessed by responses to outcome items and observation of adverse events at the time of consent and at 0- and 8-week assessments. In addition, patients were screened for suicidal ideation with item 9 of the Patient Health Questionnaire-9 (PHQ-9) [36] at the 0- and 8-week assessments; those scoring 1 or more points were considered positive for suicidal ideation [37]. Patients were assessed by the investigator and a psychiatrist, and those who were assessed as being at imminent risk of suicide and thus unable to complete the study were considered for exclusion. Patients who were hospitalized were excluded from the study.

Sample Size

The main objective of this study was to assess the superiority of vCBTp for symptom improvement in a group of patients with schizophrenia compared with a UC group, referencing a previous study by Morrison et al [38], which compared CBT and antipsychotic groups, where the mean difference between the vCBTp plus UC group and UC group after the intervention was 12 points, with the SD of the pooled UC group assumed to be 8 points. The significance level for both groups was set at 5%, and the power was set at 90%. The number of patients in each group was calculated to be 11. Assuming a participant dropout rate of 10%, the target number of patients was set at 12 in each group.

Eligibility Assessment

Two researchers (a psychiatrist [ES] and a researcher [MK]) conducted the initial consultation. It was conducted face-to-face at Chiba University Hospital. At the hospital, ES, as a psychiatrist, assessed and confirmed eligibility, including the participant's history of treatment by their family physician, severity of symptoms, and ability to consent to participate in the study. MK assessed the PANSS for allocation adjustment factors and the MacCAT-CR for competency to consent to this study. In addition, the researchers ascertained the participants' clinical characteristics, including gender, age, marital status, age at onset, duration of illness, employment status, and treatment history (eg, names of antipsychotic medications, drug titers at baseline, and changes in treatment during the study period).

Randomization and Masking

After ES or MK examined applicants for suitability prior to study enrollment, those who were identified as meeting the criteria and whose consent was obtained were enrolled as cases by the Chiba University Hospital Data Center. After registration, all patient names were key-coded, anonymized, and kept confidential. The Chiba University Hospital Data Center then conducted dynamic allocation of participants using a computer program. Because this was a pilot study and the number of patients was small, 2 allocation adjustment factors were used. Participants were randomized in a 1:1 ratio to the vCBTp plus UC group or UC group using a minimization method with 2 factors as allocation adjustment factors: total baseline PANSS score (PANSS ≥ 51) and gender. The PANSS values were calculated based on the participants' baseline values in the study by Naeem et al [39]. Each participant was randomly assigned to 1 of the 2 treatments. In addition, the data center was strictly controlled to ensure that the allocation program was not divulged

XSL•FO
RenderX

to outside parties. After study entry, the author HN, as an evaluator, was blinded for the duration of the study, and only HN who was an expert in assessing the PANSS at 0 and 8 weeks assessed the PANSS in a semistructured interview format via videoconference. HN was not informed which group participants were assigned to until study completion to ensure fairness. All secondary assessments other than the PANSS involved self-administered questionnaires.

Procedures

Patients who wished to participate in the study had an initial consultation with MK and ES at Chiba University Hospital, where eligibility assessment and consent to participate in the study were explained. At that time, they received an explanation and lecture on the basic operation of the loaned iPad and videoconference connection for vCBTp. With regard to vCBTp, MK participated from Chiba University Hospital, and patients attended sessions from home using the loaned iPad. Participants reported their email address and cell phone number to the study office for contact, but were also required to provide emergency contact information for a family member or support person in case of emergency. Videoconference was used for 0- and 8-week assessments and all vCBTp sessions. The PANSS assessments at weeks 0 and 8 were conducted via videoconference with only HN and the participants. MK did not participate in the videoconference.

All vCBTp sessions were conducted by MK. ES appropriately supervised MK after each session. In addition to research experience in CBT interventions for schizophrenia [40], MK had conducted high-intensity CBT sessions for depression and anxiety at the Cognitive Behavioral Therapy Center at Chiba University Hospital, and had the qualities and fidelity of a therapist.

ES had experience in creating remote CBT programs for patients with panic disorder, obsessive-compulsive disorder, social anxiety [26], eating disorders [41], chronic pain [42], and other disorders. The vCBTp program for patients with schizophrenia was created by 2 researchers (ES and MK). The number of sessions in this study was set based on a meta-analysis of low-intensity CBTp [11]. Since the median number of sessions in the 10 previous studies in the meta-analysis was 7.5, the program was created with a total of 7 sessions of vCBTp. The duration of treatment was 7 weeks (50 minutes per session, once a week). Seven sessions were printed as text in a paper booklet and mailed to the vCBTp plus UC group participants after the 0-week assessment. vCBTp sessions were conducted in a one-on-one format between the therapist and the participant. The therapist wrote the session content on slides in real-time while the camera was turned on and a PowerPoint file was shared on the screen in the messaging application software (Microsoft Teams) for videoconference. After each homework session, participants wrote on the text file that had been sent to them in advance and submitted the images to the therapist as email attachments with photos.

In schizophrenia, depression and anxiety are significantly related to positive symptoms [43]. We developed the vCBTp sessions based on "cognitive restructuring," a basic technique incorporated into CBT for depression and anxiety. vCBTp consists of the following 7 sessions: (1) assessment and goal setting, (2) case formulation using cognitive-behavioral models, (3) cognitive restructuring, (4) coping and relaxation, (5) externalization of intrusive memories about stressful events related to the onset of psychosis, (6) re-evaluation of intrusive memories about stressful events related to the onset of psychosis, and (7) relapse prevention (Table 1).

Table 1. Composition and contents of videoconference-based cognitive behavioral therapy for psychosis.

Session	Contents and goals	Homework
Session 1	Assessment and goal setting; identify symptoms of distress and set goals	Record stressful situations felt during the week
Session 2	Case formulation of current stress; use the cognitive-behavioral model to objectively understand events and thoughts/anxiety	Cognitive-behavioral model to record stressful situations felt during the week
Session 3	Cognitive restructuring of current stress; psychological flexibility by mentioning "other ideas;" improvement of confidence	Use the column method to derive and record another way of thinking
Session 4	Coping and relaxation; engage in fun and calming activities and activate behaviors	Record coping done in a week
Session 5	Externalize intrusive memories about stressful events related to the onset of psychosis; use cognitive-behavioral models to capture past stressful experiences and understand core beliefs that have persisted since that time; moreover, engage in self-compassion at that time	Record words of self-compassion about your symptoms and stress felt during the week
Session 6	Reassess intrusive memories about stressful events related to the onset of psychosis; reflect on how you have coped with past stressful experiences and recognize that you have coping skills	Record your coping behaviors about symptoms and stress felt during the week
Session 7	Prevention of recurrence, review of the past session theme, and setting of future goals	Fill in what you learned or felt during the vCBTp[a] session

[a]vCBTp: videoconference-based cognitive behavioral therapy for psychosis.

We also focused on intrusive thoughts and subjective distress caused by broadly defined traumatic stress experiences in patients with schizophrenia [44-46]. The Graduate School of Medicine, Chiba University, to which the researcher belongs, also has a track record of research on imagery rescripting of traumatic memories [47]. However, in this study, instead of

XSL•FO
RenderX

dealing with "traumatic memories" in consideration of the burden on participants, the aim was to help participants become aware of their core beliefs and work on themselves by reflecting on stressful events related to their past onset of psychosis. The fact that the research sessions would include an important session in which participants would reflect on painful events in their past was fully explained and agreed upon during the study description. It was also decided that sessions would focus on episodes that participants felt comfortable talking about. The order of sessions was flexible, with sessions being changed or repeated depending on the participants' levels of understanding and their reactions at the time.

There was no staff involvement in this study with regard to UC, which was defined as participants' own use of outpatient care and rehabilitation that they were engaged in prior to participation in the study and until its conclusion.

While use anxiety and connection problems are possible with the use of digital devices, the study addressed these by providing the following support: vCBTp patients practiced connecting to MK for participation through videoconference from home prior to the 0-week baseline assessment. vCBTp patients were also provided with an iPad charger and a microphone with an earpiece for consideration of their family living together. Patients with no or unstable home Wi-Fi access were loaned mobile Wi-Fi at no charge. While participating in the study, they could be contacted at any time via email. Participants were able to contact the researcher via email if they caught a cold or had other urgent appointments. When connected, participants were also called as needed to provide support for iPad use and remote connection anxiety, as appropriate. In addition, after each session was completed for the vCBTp plus UC group, the materials from that session were sent in PDF format via email. After the research trial was completed, the PowerPoint presentation materials for the participants, which had been included during 7 sessions, were printed as a paper-based booklet for the participants' own use and mailed to them by the researcher.

Outcomes

Primary Outcome

The primary outcome was the PANSS total score. The change from baseline to week 8 was compared. The PANSS is an assessment consisting of 7 positive symptoms, 7 negative symptoms, and 16 general psychopathology items. Each item is scored on a scale of 1-7 points. The scores range from 30 (no symptoms) to 210 (the most severe symptoms). The PANSS is an objective instrument for assessing schizophrenia symptoms in clinical and experimental studies and is the global standard for reliability and validity. Assessors blinded to the treatment assignment assessed the PANSS via a videoconference system at baseline and 8 weeks after the test.

Secondary Outcomes

Subtotal scores were assessed for positive symptoms, negative symptoms, and general psychopathology extracted from the PANSS subscale scores. Total scores for the 7 positive symptom subscales of the PANSS ranged from 7 (no symptoms) to 49 (severe positive symptoms). Total scores for the 7 negative symptom subscales ranged from 7 (no symptoms) to 49 (severe negative symptoms). Total scores for the 16 items in the general psychopathology subscale ranged from 16 (no symptoms) to 112 (severe psychopathology).

The Japanese version of the Beck Cognitive Insight Scale (BCIS-J) [48] was also used to assess cognitive insight. All 15 items of the BCIS-J were divided into 9 items of self-reflection (0-36) and 6 items of confidence (0-24). A lower subtotal of the 9 self-reflection items minus the 6 confidence items is associated with a lower cognitive pathology score.

In addition, the Impact of Event Scale-Revised (IES-R) [49] was used to assess subjective distress over specific stressful life events. The IES-R consists of 22 items (0-88) that are divided into 8 intrusive symptoms, 8 avoidance symptoms, and 6 hyperarousal symptoms, measuring the symptoms of persons exposed to traumatic experiences.

The PHQ-9 was used to assess depressive symptoms. It consists of 9 items ranging from 0 (no depressive symptoms) to 27 (severe depressive symptoms).

The Generalized Anxiety Disorder-7 (GAD-7) [50] was used to assess anxiety. The GAD-7 consists of 7 items ranging from 0 (no anxiety symptoms) to 21 (severe anxiety symptoms).

The EQ-5D-5L [51] was used to assess quality of life (QoL). It consists of 5 items and assesses QoL on a 5-point Likert scale ranging from 1 (not severe) to 5 (severe). The EQ-5D-5L is the most commonly used measure of QoL worldwide.

Other Outcomes

Other assessments included chlorpromazine (CPZ) [52], which indicates (equivalent to) the daily prescribed dose of antipsychotic medication. In addition, the therapists inquired about participants' experiences of adverse events during each assessment. All measures were assessed at week 0 (baseline) and week 8 (after treatment), and the results were analyzed according to set principles.

Statistical Analyses

This study followed the Consolidated Standards of Reporting Trials (CONSORT) guidelines for statistical analyses and reporting. The primary analysis was based on the intention-to-treat principle. We calculated the PANSS total score change at week 8 for both the vCBTp and UC groups and compared the mean differences between the groups to assess the primary outcome. Comparisons between the 2 groups were performed using an unpaired *t* test. Secondary endpoints were analyzed using the same approach as the primary endpoint to understand the primary endpoint further. In addition, we considered participants in each group with a 20% improvement in the total PANSS score as treatment responders. We compared the proportion of participants in each group. The frequency of adverse events was used as a safety endpoint. For percentage estimates, the exact 2-sided 95% CIs of the binomial distribution were calculated for each group. Statistical analyses were performed using SAS version 9.4 (SAS Institute Inc). Statistical significance was set at a *P*-value of <.05. The Cohen *d* (M baseline–M posttest)/SD pooled formula was used to calculate

XSL•FO
RenderX

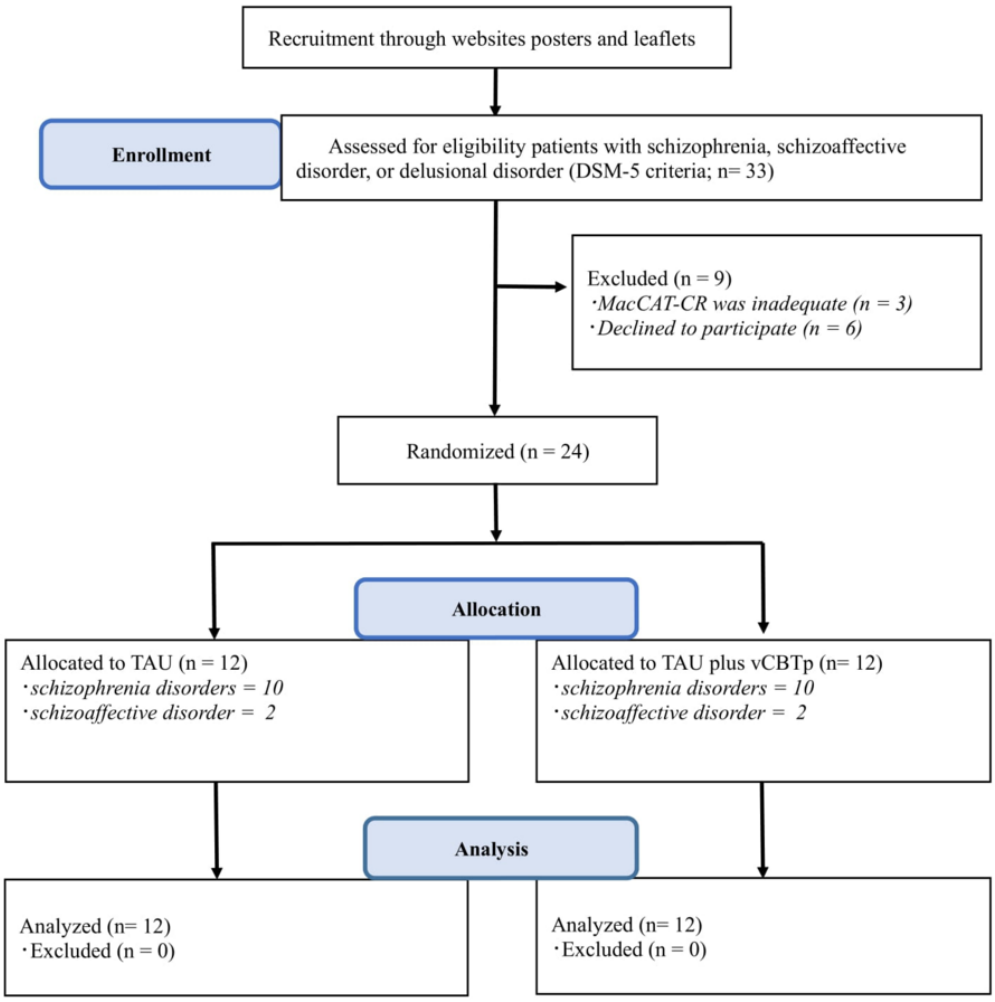
the effect size (Cohen *d*) from the difference in the means of each group.

Results

Recruitment

Figure 1 shows the flow diagram of patient recruitment based on the CONSORT guidelines. A total of 33 patients applied to participate. Applicants included those who applied through the website (n=5), those who saw the poster in the hospital outpatient waiting room and applied on their own (n=4), and those who were referred by their doctor or daycare facility staff (n=14). Three of these patients were excluded. Of the 3 excluded patients, 1 was excluded because of hospitalization. The other 2 patients were excluded because they had multiple zeros on the MacCAT-CR, did not meet the criteria for understanding the trial, and were found not to have adequate capacity to consent. These 2 individuals were referred by a daycare facility staff member, who did not have a good understanding of the clinical trial and was primarily interested in obtaining the burden-reduction funding. In addition, 6 patients voluntarily declined to participate in the study after being informed about the study, citing concerns about the burden of the study and fatigue. Ultimately, 24 patients underwent in-person eligibility examinations at Chiba University Hospital, and all 24 were enrolled in the study. Study participants had schizophrenia (n=18) and schizoaffective disorder (n=4) and were assigned in equal numbers to the 2 groups. All 24 participants in the assigned vCBTp plus UC and UC groups completed the clinical trial. They provided informal feedback on the vCBTp plus UC intervention. Nearly all the study participants who completed the study reported positive feedback, indicating that the intervention was appropriate and easy to engage in (Figure 1).

Figure 1. Flow diagram of patient recruitment. DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th edition; MacCAT-CR: MacArthur Competence Assessment Tool for Clinical Research; UC: usual care; vCBTp: videoconference-based cognitive behavioral therapy for psychosis.



Participants' Characteristics

Table 2 summarizes the clinical characteristics of the participants at the time of assignment. There were no significant differences between the 2 groups in terms of age, gender, marital status, employment status, number of people living together, number of psychiatric hospitalizations, age of onset, duration of illness, total MacArthur comprehension score, and PANSS participant proportion score below 50 (Table 2).

XSL•FO
RenderX

Table 2. Baseline characteristics of the study participants (N=24).

Variable	vCBTp[a] + UC[b] group (n=12)	UC group (n=12)
Female, n (%)	7 (58)	7 (58)
Unmarried, n (%)	8 (89)	9 (75)
Unemployed, n (%)	10 (83)	9 (75)
Age (years), mean (SD)	34.7 (8.98)	32.3 (12.26)
Hospitalization (times), mean (SD)	1.6 (1.44)	1.3 (1.87)
Age of onset (years), mean (SD)	22.8 (6.70)	24.4 (8.26)
Duration of an illness (years), mean (SD)	11.6 (7.44)	7.6 (5.74)
MacCAT-CR[c] (total score), mean (SD)	41.1 (1.11)	40.3 (1.84)
Number of patients with a PANSS[d] score less than 51, n (%)	3 (25)	2 (17)

[a]vCBTp: videoconference-based cognitive behavioral therapy for psychosis.

[b]UC: usual care.

[c]MacCAT-CR: MacArthur Competence Assessment Tool for Clinical Research.

[d]PANSS: Positive and Negative Syndrome Scale.

Participants were mostly single (n=17) and unemployed (n=19). They were in their early 30s on average, and the average number of hospitalizations was less than two, although some had never been hospitalized (n=5). Some respondents were living alone in apartments (n=1) or in assisted living (n=1), but the rest were living with family members (n=22). An equal number of patients with schizophrenia (n=10) and schizoaffective disorder (n=2) participated in each group. These ambulatory variables were not used as randomization factors.

In addition, all participants had a certain affinity for digital devices, as they usually used smartphones and PCs to watch videos. Both groups were able to maintain a remote connection environment without major problems by exchanging emails and receiving advice over the phone in case of connection problems.

Primary Outcome

Table 3 displays the means and SDs of the participants' PANSS total scores from baseline to posttest. The mean change in PANSS total scores from baseline to posttest was –9.5 (95% CI –12.09 to –6.91) for the vCBTp plus UC group and 6.9 (95% CI 1.54-12.30) for the UC group (Table 4). The vCBTp plus UC group demonstrated a significant improvement compared to the UC group, with a between-group difference of 16.41 (95% CI 10.79-22.04; $P<.001$).

XSL•FO
RenderX

Table 3. Raw data of the primary and secondary outcomes (N=24).

Variable	vCBTp[a] + UC[b] group (n=12)			UC group (n=12)		
	Pre score, mean (SD)	Post score, mean (SD)	Cohen *d*	Pre score, mean (SD)	Post score, mean (SD)	Cohen *d*
Primary outcome						
PANSS[c] (total score)	52.3 (7.02)	42.8 (6.02)	1.45	54.2 (11.07)	61.1 (10.72)	–0.61
Secondary outcomes						
PANSS (positive)	14.8 (2.77)	11.4 (1.88)	1.43	13.7 (3.77)	14.6 (3.50)	–0.24
Delusion	3.4 (1.31)	2.7 (0.78)	0.65	3.5 (1.45)	3.6 (1.31)	–0.07
Conceptual disorganization	1.8 (0.75)	1.4 (0.51)	0.62	1.6 (0.90)	1.8 (0.83)	–0.23
Hallucinatory behavior	1.8 (1.11)	1.6 (0.67)	0.22	2.3 (1.30)	2.8 (1.42)	–0.37
Excitement	1.8 (1.29)	1.3 (0.45)	0.52	1.2 (0.39)	1.2 (0.39)	0.00
Grandiosity	1.8 (1.54)	1.3 (0.45)	0.35	1.5 (1.00)	1.4 (0.79)	0.11
Suspiciousness	3.2 (1.53)	2.3 (0.97)	0.70	2.6 (1.16)	2.8 (0.87)	–0.20
Hostility	1.1 (0.29)	1.0 (0.00)	0.14	1.0 (0.00)	1.1 (0.29)	–0.13
PANSS (negative)	10.8 (2.63)	9.6 (1.78)	0.53	13.7 (6.87)	16.3 (5.03)	–0.43
PANSS (general psychopathology)	26.8 (4.18)	21.8 (2.96)	1.38	26.8 (4.57)	30.2 (5.65)	–0.66
Anxiety	2.8 (1.11)	2.0 (0.60)	0.90	2.8 (1.19)	2.8 (1.19)	0.00
Guilt feeling	2.9 (1.38)	2.1 (0.90)	0.69	2.1 (0.67)	2.3 (0.75)	–0.28
Depression	2.1 (1.00)	1.6 (0.67)	0.59	2.0 (1.04)	2.8 (1.40)	–0.42
Lack of judgment and insight	1.9 (0.90)	1.4 (0.51)	0.68	2.1 (0.79)	2.3 (0.87)	–0.24
BCIS-J[d] (self-certainty)	4.7 (2.27)	4.2 (2.41)	0.21	7.0 (3.28)	7.9 (5.05)	–0.21
BCIS-J (self-reflectiveness)	14.5 (3.12)	16.7 (3.85)	0.63	11.6 (1.19)	12.8 (4.67)	0.35
IES-R[e] (total score)	37.1 (13.43)	31.8 (14.68)	0.36	30.3 (16.55)	32.1 (16.08)	–0.11
Intrusion	13.3 (5.38)	10.0 (4.88)	0.64	11.8 (9.36)	11.3 (7.81)	0.06
Avoidance	15.3 (4.59)	14.2 (6.28)	0.20	12.2 (7.90)	12.8 (5.72)	–0.09
Hyperarousal	8.6 (5.92)	7.6 (5.30)	0.18	6.3 (4.89)	8.0 (5.29)	–0.33
PHQ-9[f]	10.7 (6.05)	8.0 (0.18)	0.47	8.6 (4.72)	9.4 (4.44)	–0.17
GAD-7[g]	7.7 (3.58)	5.7 (4.21)	0.51	5.5 (3.18)	7.2 (4.45)	–0.44
EQ-5D-5L (index value)	0.7 (0.19)	0.8 (0.18)	0.54	0.8 (0.13)	0.7 (0.08)	–0.93
Mobility	1.7 (1.07)	1.3 (0.63)	0.45	1.3 (0.45)	1.1 (0.29)	0.53
Self-care	1.4 (0.67)	1.3 (0.65)	0.15	1.1 (0.29)	1.3 (0.45)	–0.53
Usual activities	2.3 (1.22)	1.9 (1.08)	0.35	1.8 (0.87)	2.0 (1.04)	–0.21
Pain/discomfort	2.2 (0.94)	1.8 (1.22)	0.36	1.4 (0.67)	2.3 (1.06)	–0.73
Anxiety/depression	2.7 (1.15)	2.3 (1.22)	0.34	1.8 (0.72)	2.3 (0.65)	–0.73
CPZ[h] equivalent (mg/day)	488.5 (340.91)	462.3 (336.90)	0.08	413.9 (252.29)	386.8 (229.34)	0.11

[a]vCBTp: videoconference-based cognitive behavioral therapy for psychosis.

[b]UC: usual care.

[c]PANSS: Positive and Negative Syndrome Scale.

[d]BCI-J: Japanese version of the Beck Cognitive Insight Scale

[e]IES-R: Impact of Event Scale-Revised.

[f]PHQ-9: Patient Health Questionnaire-9.

[g]GAD-7: Generalized Anxiety Disorder-7.

[h]CPZ: chlorpromazine.

XSL•FO
RenderX

Table 4. Amount of change from baseline to posttest (N=24).

Variable	vCBTp[a] + UC[b] group (n=12), mean (SD)	UC group (n=12), mean (SD)	*t* test	*P* value
Primary outcome				
PANSS[c] (total score)	–9.5 (4.08)	6.9 (8.47)	6.05	<.001
Secondary outcomes				
PANSS (positive)	–3.3 (2.99)	0.9 (1.98)	4.10	<.001
Delusion	–0.8 (0.75)	0.1 (1.00)	2.31	.03
Conceptual disorganization	–0.3 (0.49)	0.3 (0.62)	0.56	.02
Hallucinatory behavior	–0.3 (0.75)	0.4 (0.79)	2.11	.046
Excitement	–0.5 (1.17)	0.0 (0.00)	1.48	.15
Grandiosity	–0.5 (1.17)	–0.1 (0.51)	1.13	.27
Suspiciousness	–0.9 (1.24)	0.2 (1.11)	2.25	.04
Hostility	–0.1 (0.29)	0.1 (0.29)	1.41	.17
PANSS (negative)	–1.2 (1.40)	2.7 (3.89)	3.21	.004
PANSS (general psychopathology)	–5.0 (2.80)	3.3 (4.79)	5.20	<.001
Anxiety	–0.8 (0.94)	0.0 (1.41)	1.70	.10
Guilt feeling	–0.8 (1.03)	0.2 (0.83)	2.61	.02
Depression	–0.5 (0.67)	0.8 (1.47)	2.86	.009
Lack of judgment and insight	–0.5 (0.52)	0.2 (0.58)	2.97	.007
BCIS-J[d] (self-certainty)	–0.5 (1.83)	0.9 (3.37)	1.28	.21
BCIS-J (self-reflectiveness)	2.2 (4.63)	1.3 (3.33)	–0.56	.58
IES-R[e] (total score)	–5.3 (14.87)	1.8 (16.65)	1.10	.28
Intrusion	–3.3 (5.55)	–0.5 (6.16)	1.15	.26
Avoidance	–1.1 (7.03)	0.6 (9.00)	0.51	.62
Hyperarousal	–1.0 (4.57)	1.7 (5.00)	1.36	.19
PHQ-9[f]	–2.7 (5.33)	0.8 (4.22)	1.78	.09
GAD-7[g]	–2.0 (4.71)	1.7 (3.52)	2.16	.04
EQ-5D-5L (index value)	0.1 (0.15)	–0.1 (0.13)	–3.14	.005
Mobility	–0.3 (0.89)	–0.2 (0.39)	0.60	.56
Self-care	–0.1 (0.79)	0.2 (0.58)	0.88	.39
Usual activities	–0.3 (0.89)	0.3 (0.87)	1.63	.12
Pain/discomfort	–0.4 (1.24)	0.8 (1.03)	2.69	.01
Anxiety/depression	–0.4 (0.79)	0.5 (0.67)	3.05	.006
CPZ[h] equivalent (mg/day)	6.8 (22.61)	–27.1 (51.63)	–2.01	.006

[a]vCBTp: videoconference-based cognitive behavioral therapy for psychosis.

[b]UC: usual care.

[c]PANSS: Positive and Negative Syndrome Scale.

[d]BCI-J: Japanese version of the Beck Cognitive Insight Scale.

[e]IES-R: Impact of Event Scale-Revised.

[f]PHQ-9: Patient Health Questionnaire-9.

[g]GAD-7: Generalized Anxiety Disorder-7.

[h]CPZ: chlorpromazine.

XSL•FO
RenderX

Secondary Outcomes

The mean change at posttest for the 3 PANSS subscales was as follows. The mean change in positive symptoms from baseline was –3.33 (95% CI –5.24 to –1.43) in the vCBTp plus UC group and 0.92 (95% CI –0.34 to 2.17) in the UC group. The difference between groups was significant at 4.25 (95% CI 2.10-6.40; *P*<.001). The mean change in negative symptoms was 2.67 (95% CI 0.19-5.14) in the UC group and –1.17 (95% CI –2.06 to –0.28) in the vCBTp plus UC group. The difference between groups was 3.83 (95% CI 1.36-6.31; *P*=.004). The mean change in overall psychopathology was 3.33 (95% CI 0.29-6.38) in the UC group and –5.00 (95% CI –6.78 to –3.22) in the vCBTp plus UC group. The difference between groups was 8.33 (95% CI 5.01-11.66; *P*<.001), with all 3 subscales showing significant differences.

The means and SDs of the secondary endpoints other than the PANSS and the mean changes are presented in Tables 3 and 4, respectively. There were no statistically significant differences between the 2 groups, including the subitems for BCIS-J, IES-R, PHQ-9, and CPZ. However, the vCBTp plus UC group had improved mean values for all change measures.

Other Outcomes

When analyzing participants who showed a 20% or greater improvement in the PANSS total score as treatment responders, we observed that 4 participants (33%) in the vCBTp plus UC group responded positively to the treatment. In comparison, none (0%) in the UC group exhibited a 20% or greater improvement. When comparing the effect size of the mean change between the 2 groups in Table 3, the PANSS total score for the vCBTp plus UC group had Cohen *d*=1.45 (95% CI 3.96-15.04). Furthermore, the subscale effect sizes for the vCBTp plus UC group were as follows: PANSS positive symptoms, Cohen *d*=1.44 (95% CI 1.38-5.42); PANSS negative symptoms, Cohen *d*=0.53 (95% CI –0.72 to 3.12); and PANSS overall psychopathology, Cohen *d*=1.38 (95% CI 1.91-8.09). Notably, a significant effect was observed outside of negative symptoms.

At 0 weeks, the scores on the PHQ-9 [36] screening for suicidal ideation (item 9) were 3 points in 0 participants, 2 in 2 participants, 1 in 5 participants, and 0 in 17 participants. At 8 weeks, the scores were 3 points in 0 participants, 2 in 1 participant, 1 in 7 participants, and 0 in 16 participants. The patients were not excluded from the study because they were not at risk for imminent suicide as assessed by the researcher's interviews and psychiatrist's ratings.

During the session on trauma in the broadest sense, several participants began to cry as they recalled their painful past. However, none declined to participate in the study because of this.

As for safety evaluation, no serious adverse events occurred among the participants, including hospitalization. There were 2 adverse events in the UC group and 4 in the vCBTp plus UC group, none of which were attributable to vCBTp.

The total PANSS scores in the UC group worsened. The self-administered questionnaire, a secondary assessment, also showed worsening scores. The 8-week assessment of the UC group revealed changes in the participants' life circumstances, including plans to retire due to poor physical and mental health, impending separation from a partner, difficulties in making friends at school, progressive illness in the family, and impending return to school after a leave of absence. The participants talked about the environmental changes in their lives.

Discussion

Principal Findings

This is the first pilot RCT to examine the efficacy of providing real-time and one-on-one vCBTp for patients with schizophrenia over 7 sessions. The study revealed that vCBTp was effective as an adjunct to UC with regard to the PANSS total score and each subscale at 8 weeks after the intervention.

Comparison With Prior Work

In a study by Morrison et al [38], the PANSS total score improved in the group provided with pharmacotherapy and CBTp (compared with baseline) at 6 weeks, with an effect size of 0.66 (mean change of 6.06 points), and at 12 weeks, with an effect size of 1.04 (mean change of 12.36 points). In addition, although treatment guidelines recommend 16 or more sessions, a meta-analysis by Hazell et al [11] revealed that CBTp performed with 6 to 15 sessions still yielded a notable effect size of 0.46, indicating that a smaller number of sessions than the recommended 16 can be effective. Furthermore, a previous study of an intervention for self-esteem reported an effect size as large as 1.2 (n=18) for 7 sessions in an unformulated configuration [53]. This suggests that even fewer sessions may produce effects by clarifying target symptoms.

However, our study requires careful discussion because the significant difference in the PANSS score observed in our study involved not only an improvement in the vCBTp plus UC group but also a worsening in the UC group. Therefore, the effect size was not a comparison of the change in the 2 groups, but rather the change between week 0 and week 8 in the vCBTp plus UC group. The vCBTp plus UC group showed a significant effect size of 1.45 (mean change 9.5) in the total PANSS score at 8 weeks, exceeding that observed in previous studies. This exceeded the results observed in a previous study.

However, the scores themselves are not remarkable improvements, since the study by Morrison et al [38] showed a 12.36-point improvement in the PANSS total score at 12 weeks, whereas this study showed a 9.5-point improvement. The smaller SD of the vCBTp plus UC group in this study may have been one of the reasons for the large effect size.

It has been reported that the addition of CBT to pharmacotherapy resulted in an improvement of more than 20% in overall symptoms in 44.5% of participants [54]. In this study, 33% of participants in the vCBTp plus UC group had an improvement of 20% or more, resulting in a generally similar rate of improvement, although not as high as in previous studies.

The reasons for the improvement are discussed below. In the study by Naeem et al [39], which we referred to in setting our

XSL•FO
RenderX

allocation adjustment factors, the PANSS score was 51.6 at baseline in the control group and 49.05 in the intervention group, which indicates mild symptoms. Again, the mean baseline scores for the vCBTp plus UC and UC groups were 54.2 and 52.3, respectively, indicating that participants had relatively mild to moderate symptom severity. Patients suitable for CBTp had mild to moderate psychiatric symptoms, and had a high level of insight [55], motivation to participate, cognitive insight [55], willingness to participate, and cognitive flexibility [56]. The following factors may have contributed to the improvement in the vCBTp plus UC group in this study. Participants with the ability to consent appropriately were included through the MacCAT-CR. This suggested that participants retained a certain level of comprehension and cognitive ability. Cognitive functioning is said to facilitate adaptation to CBTp [57]. It has also been reported that insights into the disease and mild symptoms predict good outcomes even after short-term CBTp sessions [58]. The participants in this study had mild symptoms and clear research consent, which may have contributed to the success of vCBTp.

Some participants had self-odor delusions, some were restricted from using transportation due to stress diarrhea symptoms, and some were not able to go outside due to a sense of forethought or paranoia. In addition, some patients struggled with the burden of transportation costs for travel. The ability to receive vCBTp at home seemed to reduce the labor burden associated with transportation.

After each session, a PDF summary of the session was sent to each participant, and the sessions were individualized to the case. The subjective assistance needs of patients with schizophrenia include the desire to be listened to by a professional [59]. There was a lot of positive feedback about the experience of being listened to in a one-on-one setting. Sessions included positive rapport building through listening and individual case formulation through collaborative demonstration work to develop strategies for coping with stress. The participants seemed to find it refreshing to discuss coping strategies with a specialist. After forming rapport in the first half of the session, the participants shared their painful experiences with the therapist in the second half of the session. These experiences included intrusive memories about stressful events related to the onset of psychosis, while taking into consideration their own distress. The participants then shared their experiences of working on themselves, which they said helped them to accept themselves as they are now. There were also reports of the effects of working on oneself [60]. For most of the participants, this was the first experience of talking to a therapist about a painful event in their past.

Traumatic stress symptoms have been reported in approximately 12.4% of patients with schizophrenia [61], suggesting an integrated model of trauma and schizophrenia [62]. Intervention attempts have also been reported [63]. In this study, the 5th and 6th sessions dealt with intrusive memories of stressful events related to the onset of psychosis, and were designed to help participants understand core beliefs, work on themselves, and reflect on coping strategies that have helped them overcome their painful past. Although there were no significant differences between groups on the IES-R, the mean change scores for the vCBTp + UC group, including the subscales, showed a trend toward improvement.

Significant improvements were observed in the PANSS subscale scores for positive symptoms (P<.001), negative symptoms (P=.009), and general psychopathology (P<.001). Improvements were also observed in the self-administered GAD-7 and EQ-5D-5L and the EQ-5D-5L subcomponents of pain/discomfort and anxiety/depression. In recent years, nearly half of patients with schizophrenia have comorbid anxiety disorders [64]. In this study, it is possible that vCBTp was effective for those with subjective anxiety and discomfort.

We next discuss exacerbations in the UC group. Schizophrenia has been noted for its disease characteristics, including symptom instability and agitation due to stress in daily life [65,66], and several patients in the UC group reported anger, hopelessness, depression, and anxiety at assessment due to daily work, friends and family, and other life circumstances during the study period. There have been reports of a lack of coping skills [67] and high sensitivity to stress and symptoms [68] with regard to daily life stress in schizophrenia. In the UC group, stress related to coping with interpersonal relationships and worries in daily life may have also affected symptoms and contributed to worsening outcomes [69]. Schizophrenia carries a high risk of relapse, with 30% of patients being readmitted within 6 months to 1 year [70]. This may indicate that patients with schizophrenia have difficulty with coping skills and stress management [71]. In this study, most of the secondary endpoints of self-administered questionnaires also showed worsening values. The worsening of anxiety/pain on the GAD-7 and EQ-5D-5L was particularly noticeable, suggesting that there are anxiety-based effects in daily life events.

We initially estimated a dropout rate of 10%. This was considered an appropriate estimate considering that a previous study reported an average dropout rate of 13% [72] and another study reported a dropout rate of 11% [40]. However, there were no dropouts in this study. There are several possible reasons for this. First, this study consisted of fewer sessions than the formulated number of CBTp sessions. Some meta-analyses have shown that the dropout rate from research trials among patients with schizophrenia increases as the number of intervention sessions increases [73]. In addition, the implementation of MacCAT-CR with the inclusion criteria and the inclusion of participants who had firm research consent and understanding may have been some of the reasons for the lack of dropouts. Furthermore, many participants were motivated to participate in the study because they commented that they wanted to see more treatment options for schizophrenia and they wanted to help others who were suffering like they were. It has been reported that careful explanation of the methods and transparency of the study can increase motivation to participate and decrease the dropout rate [74]. As the participants had never received CBT before, there may have been some novelty to the study. In any case, it was inferred that they were highly motivated to participate. These factors could be the reasons for the absence of dropouts.

PHQ-9 item 9 was screened and used as a safety measure. Attention to suicidal ideation is important when the subject has

XSL•FO
RenderX

schizophrenia, as a study [75] validating the risk of suicidal behavior in psychosis with responses to item 9 of the PHQ-9 stated that the risk is higher for psychosis than depression. In this study, the number of participants who scored more than 1 point was also higher. Some participants in this study also scored 1 or more, but when interviewed afterward, they said that they scored based on their mood of anxiety and discouragement, saying that they feel sad or painful enough to want to die, and no one showed real suicidal thoughts. None of the adverse events were attributable to the vCBTp intervention (eg, cold and reflux esophagitis), and there were no serious adverse events, confirming a certain level of safety.

The use of videoconference systems can reduce the burden on the patient's hospital visit effort, that is, the financial burden of transportation costs, weather effects, and social fears associated with transportation use. It can also reduce the burden on CBTp providers. As a result, accessibility to CBTp could be improved.

Regarding remote interventions, it has been reported that simply providing devices is not sufficient to address digital exclusion, and training in using the devices is likely to be required [76]. However, with a simple structure, pretraining, and a support system, such as in this study, which can allow patients to open an email and tap a URL to participate in a videoconference, it may be possible to implement psychosocial interventions remotely using digital technology for patients with schizophrenia.

The major originality of this study is that participants could receive CBTp in the comfort of their own homes. Currently, self-help approaches using digital media, such as the internet and cell phones, are being actively developed [77,78]. As a result, many patients are interested in using digital technology to improve their mental health [25]. Low-intensity interventions may benefit many patients with schizophrenia without increasing costs [79]. Furthermore, studies have demonstrated the effectiveness of targeted CBTp with reduced session frequency [11,12,14]. Furthermore, an online intervention via videoconferencing has been reported to be feasible for patients with schizophrenia [80]. These considerations suggest that the provision of vCBTp using a remote format with reduced session frequency may be feasible enough for implementation in practice, especially in terms of a stepped care model to deliver CBTp [8,81].

Limitations

This exploratory study had some limitations. First, the UC settings varied, and these were not subject to comparison or analysis between the 2 groups. The frequency of outpatient visits varied from once a week to once every 2 months. The frequency and duration of use of psychiatric rehabilitation facilities also varied. These may have influenced outcomes as confounding variables. Prescription medications and methods of administration also varied. The study did not analyze the influence of external variables such as the participants' living environments; thus, future research on correlations with external variables would be desirable.

Second, it was not clear what targets the vCBTp program influenced. Although various points were considered, the primary mechanism of change was unclear because the study incorporated a variety of techniques.

Third, due to the absence of a follow-up assessment, the determination of efficacy remained unclear. Therefore, a 1-year follow-up is being conducted for vCBTp participants as an additional study.

Fourth, the EQ-5D-5L tool was used as a subjective measure of quality of life but was not validated for quality-adjusted useful life or cost-effectiveness. Future studies should examine the contribution of the vCBTp program to health care costs.

Fifth, the sample size in our pilot study was relatively small. Since improvements were observed in this study, a larger multi-center study is desired.

Finally, since this trial was conducted with only 1 therapist and 1 assessor, we are yet to examine how to share the experience of this vCBTp intervention with staff at other facilities. It is important to consider the possibility of social implementation in the future.

Conclusions

The findings of this RCT indicate that vCBTp is effective in treating symptoms over 7 sessions. The findings show that vCBTp can be effective at a distance and with a reduced number of sessions, which can reduce the burden on the practitioner and improve the accessibility of CBTp for patients. It may also contribute to relapse prevention and psychosocial support during stepped care. Further investigations are required to determine the durability of this effect.

Acknowledgments

We are grateful to all participants and all those involved in this study. We thank all the staff and people involved at the Center for Cognitive Behavioral Therapy, Chiba University Hospital. This work was supported by a Grant-in-Aid for Scientific Research (grant number 21K11197; Japan Society for the Promotion of Science).

Data Availability

The data are available upon reasonable request.

XSL•FO
RenderX

Authors' Contributions

MK and ES designed the study and managed the research methods and project. HH collected and organized the data. YS performed the formal analyses. MK, ES, and HN recruited patients. MK and ES created the 7-session videoconference-based cognitive behavioral therapy for psychosis component. MK performed all sessions of the intervention. HN performed all Positive and Negative Syndrome Scale assessments (0 and 8 weeks) in this study and participated in patient enrollment. The manuscript was written by MK and ES. Study funding was secured by MK through a grant. MK wrote the manuscript. All authors, including ES, contributed to and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 869 KB - formative_v9i1e59540_app1.pdf]

References

1. Mueser KT, McGurk SR. Schizophrenia. Lancet 2004 Jun 19;363(9426):2063-2072. [doi: 10.1016/S0140-6736(04)16458-1] [Medline: 15207959]
2. Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. PLoS Med 2005 May 31;2(5):e141 [FREE Full text] [doi: 10.1371/journal.pmed.0020141] [Medline: 15916472]
3. Psychosis and schizophrenia in adults: prevention and management. London: National Institute for Health and Care Excellence (NICE); 2014.
4. Mankiewicz PD, Turner C. Cognitive restructuring and graded behavioural exposure for delusional appraisals of auditory hallucinations and comorbid anxiety in paranoid schizophrenia. Case Rep Psychiatry 2014;2014:124564-124568 [FREE Full text] [doi: 10.1155/2014/124564] [Medline: 25302129]
5. Jauhar S, McKenna PJ, Radua J, Fung E, Salvador R, Laws KR. Cognitive-behavioural therapy for the symptoms of schizophrenia: systematic review and meta-analysis with examination of potential bias. Br J Psychiatry 2014 Jan 02;204(1):20-29. [doi: 10.1192/bjp.bp.112.116285] [Medline: 24385461]
6. Wykes T, Steel C, Everitt B, Tarrier N. Cognitive behavior therapy for schizophrenia: effect sizes, clinical models, and methodological rigor. Schizophr Bull 2008 May 04;34(3):523-537 [FREE Full text] [doi: 10.1093/schbul/sbm114] [Medline: 17962231]
7. Farhall J, Freeman NC, Shawyer F, Trauer T. An effectiveness trial of cognitive behaviour therapy in a representative sample of outpatients with psychosis. Br J Clin Psychol 2009 Mar 31;48(Pt 1):47-62. [doi: 10.1111/j.2044-8260.2009.tb00456.x] [Medline: 18851771]
8. Kopelovich SL, Nutting E, Blank J, Buckland HT, Spigner C. Preliminary point prevalence of Cognitive Behavioral Therapy for psychosis (CBTp) training in the U.S. and Canada. Psychosis 2022 Jan 14;14(4):344-354. [doi: 10.1080/17522439.2021.1971744]
9. The abandoned illness: a report by the Schizophrenia Commission. Rethink. URL: https://www.rethink.org/media/2637/the-abandoned-illness-final.pdf [accessed 2024-12-11]
10. Shafran R, Clark D, Fairburn C, Arntz A, Barlow D, Ehlers A, et al. Mind the gap: Improving the dissemination of CBT. Behav Res Ther 2009 Nov;47(11):902-909. [doi: 10.1016/j.brat.2009.07.003] [Medline: 19664756]
11. Hazell CM, Hayward M, Cavanagh K, Strauss C. A systematic review and meta-analysis of low intensity CBT for psychosis. Clin Psychol Rev 2016 Apr;45:183-192 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.004] [Medline: 27048980]
12. Malik N, Kingdon D, Pelton J, Mehta R, Turkington D. Effectiveness of brief cognitive-behavioral therapy for schizophrenia delivered by mental health nurses: relapse and recovery at 24 months. J Clin Psychiatry 2009 Feb 10;70(2):201-207. [doi: 10.4088/jcp.07m03990] [Medline: 19210949]
13. Brabban A, Tai S, Turkington D. Predictors of outcome in brief cognitive behavior therapy for schizophrenia. Schizophr Bull 2009 Sep 01;35(5):859-864 [FREE Full text] [doi: 10.1093/schbul/sbp065] [Medline: 19571248]
14. Turkington D, Kingdon D, Turner T, Insight into Schizophrenia Research Group. Effectiveness of a brief cognitive-behavioural therapy intervention in the treatment of schizophrenia. Br J Psychiatry 2002 Jun 2;180(06):523-527. [doi: 10.1192/bjp.180.6.523] [Medline: 12042231]
15. Tai S, Turkington D. The evolution of cognitive behavior therapy for schizophrenia: current practice and recent developments. Schizophr Bull 2009 Sep;35(5):865-873 [FREE Full text] [doi: 10.1093/schbul/sbp080] [Medline: 19661198]
16. Keen N, Hunter ECM, Peters E. Integrated Trauma-Focused Cognitive-Behavioural Therapy for Post-traumatic Stress and Psychotic Symptoms: A Case-Series Study Using Imaginal Reprocessing Strategies. Front Psychiatry 2017 Jun 01;8:92 [FREE Full text] [doi: 10.3389/fpsyt.2017.00092] [Medline: 28620323]

XSL•FO
RenderX

17. Gairns S, Alvarez-Jimenez M, Hulbert C, McGorry P, Bendall S. Perceptions of clinicians treating young people with first-episode psychosis for post-traumatic stress disorder. Early Interv Psychiatry 2015 Feb 27;9(1):12-20. [doi: 10.1111/eip.12065] [Medline: 23802596]
18. Sin J, Spain D, Furuta M, Murrells T, Norman I. Psychological interventions for post-traumatic stress disorder (PTSD) in people with severe mental illness. Cochrane Database Syst Rev 2017 Jan 24;1(1):CD011464. [doi: 10.1002/14651858.CD011464.pub2] [Medline: 28116752]
19. Sin J, Spain D. Psychological interventions for trauma in individuals who have psychosis: A systematic review and meta-analysis. Psychosis 2016 May 23;9(1):67-81. [doi: 10.1080/17522439.2016.1167946]
20. Swan S, Keen N, Reynolds N, Onwumere J. Psychological interventions for post-traumatic stress symptoms in psychosis: a systematic review of outcomes. Front Psychol 2017 Mar 14;8:341 [FREE Full text] [doi: 10.3389/fpsyg.2017.00341] [Medline: 28352239]
21. Gottlieb JD, Romeo KH, Penn DL, Mueser KT, Chiko BP. Web-based cognitive-behavioral therapy for auditory hallucinations in persons with psychosis: a pilot study. Schizophr Res 2013 Apr;145(1-3):82-87. [doi: 10.1016/j.schres.2013.01.002] [Medline: 23410709]
22. Gottlieb JD, Gidugu V, Maru M, Tepper MC, Davis MJ, Greenwold J, et al. Randomized controlled trial of an internet cognitive behavioral skills-based program for auditory hallucinations in persons with psychosis. Psychiatr Rehabil J 2017 Sep;40(3):283-292. [doi: 10.1037/prj0000258] [Medline: 28517948]
23. Rüegg N, Moritz S, Berger T, Lüdtke T, Westermann S. An internet-based intervention for people with psychosis (EviBaS): study protocol for a randomized controlled trial. BMC Psychiatry 2018 Apr 13;18(1):102 [FREE Full text] [doi: 10.1186/s12888-018-1644-8] [Medline: 29653532]
24. Westermann S, Rüegg N, Lüdtke T, Moritz S, Berger T. Internet-based self-help for psychosis: Findings from a randomized controlled trial. J Consult Clin Psychol 2020 Oct;88(10):937-950. [doi: 10.1037/ccp0000602] [Medline: 32790453]
25. Arnold C, Farhall J, Villagonzalo K, Sharma K, Thomas N. Engagement with online psychosocial interventions for psychosis: A review and synthesis of relevant factors. Internet Interv 2021 Sep;25:100411 [FREE Full text] [doi: 10.1016/j.invent.2021.100411] [Medline: 34401370]
26. Matsumoto K, Sutoh C, Asano K, Seki Y, Urao Y, Yokoo M, et al. Internet-based cognitive behavioral therapy with real-time therapist support via videoconference for patients with obsessive-compulsive disorder, panic disorder, and social anxiety disorder: pilot single-arm trial. J Med Internet Res 2018 Dec 17;20(12):e12091-e12890 [FREE Full text] [doi: 10.2196/12091] [Medline: 30559094]
27. Horita H, Seki Y, Yamaguchi T, Shiko Y, Kawasaki Y, Shimizu E. Videoconference-delivered cognitive behavioral therapy for parents of adolescents with internet addiction: pilot randomized controlled trial. JMIR Pediatr Parent 2024 Oct 03;7:e60604 [FREE Full text] [doi: 10.2196/60604] [Medline: 39361415]
28. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
29. Lecomte T, Abdel-Baki A, Francoeur A, Cloutier B, Leboeuf A, Abadie P, et al. Group therapy via videoconferencing for individuals with early psychosis: A pilot study. Early Interv Psychiatry 2021 Dec 23;15(6):1595-1601. [doi: 10.1111/eip.13099] [Medline: 33354926]
30. Wood HJ, Gannon JM, Chengappa KR, Sarpal DK. Group teletherapy for first-episode psychosis: Piloting its integration with coordinated specialty care during the COVID-19 pandemic. Psychol Psychother 2021 Jun 20;94(2):382-389. [doi: 10.1111/papt.12310] [Medline: 33078496]
31. Mendelson D, Thibaudeau É, Sauvé G, Lavigne KM, Bowie CR, Menon M, et al. Remote group therapies for cognitive health in schizophrenia-spectrum disorders: Feasible, acceptable, engaging. Schizophr Res Cogn 2022 Jun;28:100230 [FREE Full text] [doi: 10.1016/j.scog.2021.100230] [Medline: 35242604]
32. Katsushima M, Nakamura H, Hanaoka H, Shiko Y, Komatsu H, Shimizu E. Randomised controlled trial on the effect of video-conference cognitive behavioural therapy for patients with schizophrenia: a study protocol. BMJ Open 2023 Sep 11;13(9):e069734 [FREE Full text] [doi: 10.1136/bmjopen-2022-069734] [Medline: 37696635]
33. Fovet T, Amad A, Thomas P, Jardri R. [Schizophrenia and informed consent to research]. Encephale 2015 Oct;41(5):394-402. [doi: 10.1016/j.encep.2014.07.004] [Medline: 25127894]
34. Hostiuc S, Rusu MC, Negoi I, Drima E. Testing decision-making competency of schizophrenia participants in clinical trials. A meta-analysis and meta-regression. BMC Psychiatry 2018 Jan 05;18(1):2 [FREE Full text] [doi: 10.1186/s12888-017-1580-z] [Medline: 29304845]
35. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull 1987;13(2):261-276. [doi: 10.1093/schbul/13.2.261] [Medline: 3616518]
36. Muramatsu K, Miyaoka H, Kamijima K, Muramatsu Y, Yoshida M, Otsubo T, et al. The patient health questionnaire, Japanese version: validity according to the mini-international neuropsychiatric interview-plus. Psychol Rep 2007 Dec 01;101(3 Pt 1):952-960. [doi: 10.2466/pr0.101.3.952-960] [Medline: 18232454]

XSL•FO
RenderX

37. Na PJ, Yaramala SR, Kim JA, Kim H, Goes FS, Zandi PP, et al. The PHQ-9 Item 9 based screening for suicide risk: a validation study of the Patient Health Questionnaire (PHQ)-9 Item 9 with the Columbia Suicide Severity Rating Scale (C-SSRS). J Affect Disord 2018 May;232:34-40. [doi: 10.1016/j.jad.2018.02.045] [Medline: 29477096]
38. Morrison AP, Law H, Carter L, Sellers R, Emsley R, Pyle M, et al. Antipsychotic drugs versus cognitive behavioural therapy versus a combination of both in people with psychosis: a randomised controlled pilot and feasibility study. The Lancet Psychiatry 2018 May;5(5):411-423. [doi: 10.1016/s2215-0366(18)30096-8]
39. Naeem F, Johal R, McKenna C, Rathod S, Ayub M, Lecomte T, et al. Cognitive Behavior Therapy for psychosis based Guided Self-help (CBTp-GSH) delivered by frontline mental health professionals: Results of a feasibility study. Schizophr Res 2016 May;173(1-2):69-74. [doi: 10.1016/j.schres.2016.03.003] [Medline: 26971071]
40. Katsushima M, Shimizu E. Brief cognitive behavioral therapy for depression and anxiety in patients with schizophrenia in psychiatric home nursing service: pilot randomized controlled trial. Behav Sci (Basel) 2024 Aug 05;14(8):680 [FREE Full text] [doi: 10.3390/bs14080680] [Medline: 39199076]
41. Hamatani S, Numata N, Matsumoto K, Sutoh C, Ibuki H, Oshiro K, et al. Internet-based cognitive behavioral therapy via videoconference for patients with bulimia nervosa and binge-eating disorder: pilot prospective single-arm feasibility trial. JMIR Form Res 2019 Oct 23;3(4):e15738 [FREE Full text] [doi: 10.2196/15738] [Medline: 31647472]
42. Taguchi K, Numata N, Takanashi R, Takemura R, Yoshida T, Kutsuzawa K, et al. Clinical effectiveness and cost-effectiveness of videoconference-based integrated cognitive behavioral therapy for chronic pain: randomized controlled trial. J Med Internet Res 2021 Nov 22;23(11):e30690 [FREE Full text] [doi: 10.2196/30690] [Medline: 34813489]
43. Hartley S, Barrowclough C, Haddock G. Anxiety and depression in psychosis: a systematic review of associations with positive psychotic symptoms. Acta Psychiatr Scand 2013 Nov 04;128(5):327-346. [doi: 10.1111/acps.12080] [Medline: 23379898]
44. Malcolm CP, Picchioni MM, Ellett L. Intrusive prospective imagery, posttraumatic intrusions and anxiety in schizophrenia. Psychiatry Res 2015 Dec 30;230(3):899-904. [doi: 10.1016/j.psychres.2015.11.029] [Medline: 26619917]
45. Morrison AP, Baker CA. Intrusive thoughts and auditory hallucinations: a comparative study of intrusions in psychosis. Behav Res Ther 2000 Nov;38(11):1097-1106. [doi: 10.1016/s0005-7967(99)00143-6] [Medline: 11060938]
46. O'Driscoll C, Mason O, Brady F, Smith B, Steel C. Process analysis of trauma-focused cognitive behavioural therapy for individuals with schizophrenia. Psychol Psychother 2016 Jun 23;89(2):117-132. [doi: 10.1111/papt.12072] [Medline: 26202461]
47. Shibuya T, Seki Y, Nagata S, Murata T, Hiramatsu Y, Yamada F, et al. Imagery rescripting of traumatic memories for panic disorder: an exploratory study. tCBT 2018 Jun 06;11:1-21. [doi: 10.1017/s1754470x18000028]
48. Uchida T, Matsumoto K, Kikuchi A, Miyakoshi T, Ito F, Ueno T, et al. Psychometric properties of the Japanese version of the Beck Cognitive Insight Scale: relation of cognitive insight to clinical insight. Psychiatry Clin Neurosci 2009 Jun 14;63(3):291-297 [FREE Full text] [doi: 10.1111/j.1440-1819.2009.01946.x] [Medline: 19566759]
49. Asukai N, Kato H, Kawamura N, Kim Y, Yamamoto K, Kishimoto J, et al. Reliability and validity of the Japanese-language version of the impact of event scale-revised (IES-R-J): four studies of different traumatic events. J Nerv Ment Dis 2002 Mar;190(3):175-182. [doi: 10.1097/00005053-200203000-00006] [Medline: 11923652]
50. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
51. Feng Y, Kohlmann T, Janssen MF, Buchholz I. Psychometric properties of the EQ-5D-5L: a systematic review of the literature. Qual Life Res 2021 Mar 07;30(3):647-673 [FREE Full text] [doi: 10.1007/s11136-020-02688-y] [Medline: 33284428]
52. Woods SW. Chlorpromazine equivalent doses for the newer atypical antipsychotics. J Clin Psychiatry 2003 Jun 15;64(6):663-667. [doi: 10.4088/jcp.v64n0607] [Medline: 12823080]
53. Hall PL, Tarrier N. The cognitive-behavioural treatment of low self-esteem in psychotic patients: a pilot study. Behav Res Ther 2003 Mar;41(3):317-332. [doi: 10.1016/s0005-7967(02)00013-x] [Medline: 12600402]
54. Bighelli I, Huhn M, Schneider-Thoma J, Krause M, Reitmeir C, Wallis S, et al. Response rates in patients with schizophrenia and positive symptoms receiving cognitive behavioural therapy: a systematic review and single-group meta-analysis. BMC Psychiatry 2018 Dec 04;18(1):380 [FREE Full text] [doi: 10.1186/s12888-018-1964-8] [Medline: 30514268]
55. Jung E, Wiesjahn M, Rief W, Lincoln TM. Perceived therapist genuineness predicts therapeutic alliance in cognitive behavioural therapy for psychosis. Br J Clin Psychol 2015 Mar 15;54(1):34-48. [doi: 10.1111/bjc.12059] [Medline: 25040363]
56. Currell S, Christodoulides T, Siitarinen J, Dudley R. Patient factors that impact upon cognitive behavioural therapy for psychosis: therapists' perspectives. Behav. Cogn. Psychother 2015 Jun 25;44(4):493-498. [doi: 10.1017/s1352465815000260]
57. Menon M, Andersen DR, Quilty LC, Woodward TS. Individual factors predicted to influence outcome in group CBT for psychosis (CBTp) and related therapies. Front Psychol 2015 Oct 28;6:1563 [FREE Full text] [doi: 10.3389/fpsyg.2015.01563] [Medline: 26578995]
58. Naeem F, Kingdon D, Turkington D. Predictors of response to cognitive behaviour therapy in the treatment of schizophrenia: a comparison of brief and standard interventions. Cogn Ther Res 2008 Mar 11;32(5):651-656. [doi: 10.1007/s10608-008-9186-x]

XSL•FO
RenderX

59. Gunnmo P, Bergman H. What do individuals with schizophrenia need to increase their well-being. Int J Qual Stud Health Well-being 2011 Feb 14;6(1):5412 [FREE Full text] [doi: 10.3402/qhw.v6i1.5412] [Medline: 21562628]
60. Maisey S, Correia H, Paulik G. The role of self-compassion on the relationship between trauma and hearing voices. Clin Psychol Psychother 2022 Mar 02;29(2):698-705. [doi: 10.1002/cpp.2663] [Medline: 34476866]
61. Achim AM, Maziade M, Raymond E, Olivier D, Mérette C, Roy M. How prevalent are anxiety disorders in schizophrenia? A meta-analysis and critical review on a significant association. Schizophr Bull 2011 Jul 03;37(4):811-821 [FREE Full text] [doi: 10.1093/schbul/sbp148] [Medline: 19959704]
62. Morrison AP, Frame L, Larkin W. Relationships between trauma and psychosis: a review and integration. Br J Clin Psychol 2003 Nov 24;42(Pt 4):331-353. [doi: 10.1348/014466503322528892] [Medline: 14633411]
63. McCartney L, Douglas M, Varese F, Turkington D, Morrison AP, Dudley R. Cognitive behavioural therapy for psychosis targeting trauma, voices and dissociation: a case report. tCBT 2019 Mar 01;12:e18. [doi: 10.1017/S1754470X19000035]
64. Nemoto T, Uchino T, Aikawa S, Matsuo S, Mamiya N, Shibasaki Y, et al. Impact of changes in social anxiety on social functioning and quality of life in outpatients with schizophrenia: A naturalistic longitudinal study. J Psychiatr Res 2020 Dec;131:15-21 [FREE Full text] [doi: 10.1016/j.jpsychires.2020.08.007] [Medline: 32911206]
65. Halverson TF, Meyer-Kalos PS, Perkins DO, Gaylord SA, Palsson OS, Nye L, et al. Enhancing stress reactivity and wellbeing in early schizophrenia: A randomized controlled trial of Integrated Coping Awareness Therapy (I-CAT). Schizophr Res 2021 Sep;235:91-101. [doi: 10.1016/j.schres.2021.07.022] [Medline: 34332429]
66. Emsley R, Chiliza B, Asmal L, Harvey BH. The nature of relapse in schizophrenia. BMC Psychiatry 2013 Feb 8;13(1):50. [doi: 10.1186/1471-244x-13-50]
67. Lukoff D, Snyder K, Ventura J, Nuechterlein KH. Life events, familial stress, and coping in the developmental course of schizophrenia. Schizophr Bull 1984 Jan 01;10(2):258-292. [doi: 10.1093/schbul/10.2.258] [Medline: 6729412]
68. Myin-Germeys I, Delespaul P, van Os J. Behavioural sensitization to daily life stress in psychosis. Psychol Med 2005 May;35(5):733-741. [doi: 10.1017/s0033291704004179] [Medline: 15918350]
69. Ventura J, Nuechterlein KH, Lukoff D, Hardesty JP. A prospective study of stressful life events and schizophrenic relapse. J Abnorm Psychol 1989 Nov;98(4):407-411. [doi: 10.1037//0021-843x.98.4.407] [Medline: 2592674]
70. Maki S, Nagai K, Ando S, Tamakoshi K. Structure and predictors of in-hospital nursing care leading to reduction in early readmission among patients with schizophrenia in Japan: A cross-sectional study. PLoS One 2021 Apr 30;16(4):e0250771 [FREE Full text] [doi: 10.1371/journal.pone.0250771] [Medline: 33930056]
71. Wang Q, Zhu X, Jiang X, Li M, Chang R, Chen B, et al. Relationship between stressful life events, coping styles, and schizophrenia relapse. Int J Ment Health Nurs 2021 Oct 06;30(5):1149-1159 [FREE Full text] [doi: 10.1111/inm.12865] [Medline: 33960095]
72. Villeneuve K, Potvin S, Lesage A, Nicole L. Meta-analysis of rates of drop-out from psychosocial treatment among persons with schizophrenia spectrum disorder. Schizophr Res 2010 Aug;121(1-3):266-270. [doi: 10.1016/j.schres.2010.04.003] [Medline: 20452749]
73. Szymczynska P, Walsh S, Greenberg L, Priebe S. Attrition in trials evaluating complex interventions for schizophrenia: Systematic review and meta-analysis. J Psychiatr Res 2017 Jul;90:67-77. [doi: 10.1016/j.jpsychires.2017.02.009] [Medline: 28231496]
74. Jørgensen R, Munk-Jørgensen P, Lysaker PH, Buck KD, Hansson L, Zoffmann V. Overcoming recruitment barriers revealed high readiness to participate and low dropout rate among people with schizophrenia in a randomized controlled trial testing the effect of a Guided Self-Determination intervention. BMC Psychiatry 2014 Feb 3;14(1):28. [doi: 10.1186/1471-244x-14-28]
75. Simon GE, Yarborough BJ, Rossom RC, Lawrence JM, Lynch FL, Waitzfelder BE, et al. Self-reported suicidal ideation as a predictor of suicidal behavior among outpatients with diagnoses of psychotic disorders. Psychiatr Serv 2019 Mar 01;70(3):176-183 [FREE Full text] [doi: 10.1176/appi.ps.201800381] [Medline: 30526341]
76. Watson A, Mellotte H, Hardy A, Peters E, Keen N, Kane F. The digital divide: factors impacting on uptake of remote therapy in a South London psychological therapy service for people with psychosis. J Ment Health 2022 Dec 28;31(6):825-832. [doi: 10.1080/09638237.2021.1952955] [Medline: 34319202]
77. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
78. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012 Dec;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
79. Kopelovich SL, Turkington D. Remote CBT for psychosis during the COVID-19 pandemic: challenges and opportunities. Community Ment Health J 2021 Jan;57(1):30-34 [FREE Full text] [doi: 10.1007/s10597-020-00718-0] [Medline: 33001323]
80. Santesteban-Echarri O, Piskulic D, Nyman RK, Addington J. Telehealth interventions for schizophrenia-spectrum disorders and clinical high-risk for psychosis individuals: A scoping review. J Telemed Telecare 2018 Aug 22;26(1-2):14-20. [doi: 10.1177/1357633x18794100]

XSL•FO
RenderX

81. Kopelovich SL, Strachan E, Sivec H, Kreider V. Stepped care as an implementation and service delivery model for cognitive behavioral therapy for psychosis. Community Ment Health J 2019 Jul 8;55(5):755-767. [doi: 10.1007/s10597-018-00365-6] [Medline: 30623294]

Abbreviations

BCIS-J: Japanese version of the Beck Cognitive Insight Scale
CBTp: cognitive behavioral therapy for psychosis
CONSORT: Consolidated Standards of Reporting Trials
CPZ: chlorpromazine
GAD-7: Generalized Anxiety Disorder-7
IES-R: Impact of Event Scale-Revised
MacCAT-CR: MacArthur Competence Assessment Tool for Clinical Research
NICE: National Institute of Health and Clinical Excellence
PANSS: Positive and Negative Syndrome Scale
PHQ-9: Patient Health Questionnaire-9
RCT: randomized controlled trial
UC: usual care
vCBTp: videoconference-based cognitive behavioral therapy for psychosis

Edited by A Mavragani; submitted 16.04.24; peer-reviewed by E Granholm, M Kobayashi, S Zaman; comments to author 17.06.24; revised version received 07.11.24; accepted 28.11.24; published 14.01.25.

Please cite as:
Katsushima M, Nakamura H, Shiko Y, Hanaoka H, Shimizu E
Effectiveness of a Videoconference-Based Cognitive Behavioral Therapy Program for Patients with Schizophrenia: Pilot Randomized Controlled Trial
JMIR Form Res 2025;9:e59540
URL: https://formative.jmir.org/2025/1/e59540
doi:10.2196/59540
PMID:39610049

©Masayuki Katsushima, Hideki Nakamura, Yuki Shiko, Hideki Hanaoka, Eiji Shimizu. Originally published in JMIR Formative Research (https://formative.jmir.org), 14.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Testing a Web-Based Interactive Comic Tool to Decrease Obesity Risk Among Racial and Ethnic Minority Preadolescents: Randomized Controlled Trial

May May Leung[1], RDN, PhD; Katrina F Mateo[2], MPH; Marlo Dublin[3], RDN, MS; Laura Harrison[2], MPH; Sandra Verdaguer[2], MPH; Katarzyna Wyka[2], PhD

[1]Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, United States

[2]Graduate School of Public Health Policy, City University of New York, New York, NY, United States

[3]School of Urban Public Health, Hunter College, City University of New York, New York, NY, United States

Corresponding Author:
May May Leung, RDN, PhD
Friedman School of Nutrition Science and Policy
Tufts University
150 Harrison Avenue
Boston, MA, 02111
United States
Phone: 1 6176363676
Email: maymay.leung@tufts.edu

Abstract

Background: Childhood obesity prevalence remains high, especially in racial and ethnic minority populations with low incomes. This epidemic is attributed to various dietary behaviors, including increased consumption of energy-dense foods and sugary beverages and decreased intake of fruits and vegetables. Interactive, technology-based approaches are emerging as promising tools to support health behavior changes.

Objective: This study aimed to assess the feasibility and acceptability of Intervention INC (Interactive Nutrition Comics for Urban, Minority Preadolescents), a 6-chapter web-based interactive nutrition comic tool. Its preliminary effectiveness on diet-related psychosocial variables and behaviors was also explored.

Methods: A total of 89 Black or African American and Hispanic preadolescents with a mean age of 10.4 (SD 1.0) years from New York City participated in a pilot 2-group randomized study, comprising a 6-week intervention and a 3-month follow-up (T4) period. Of the 89 participants, 61% were female, 62% were Black, 42% were Hispanic, 53% were overweight or obese, and 34% had an annual household income of <US $20,000. Participants were randomly assigned to the experimental group (45/89, 50% received the web-based comic tool), or the comparison group (44/89, 50% received web-based nutrition newsletters). Primary measures included feasibility and usability at intervention midpoint (T2) and intervention end (T3). Semistructured interviews were conducted at the same time to assess acceptability and satisfaction. Secondary measures, collected at baseline (T1), T2, T3, and at T4, included attitudes, beliefs, and behaviors related to fruit, vegetable, water, sugar, and junk food intake. Descriptive analyses were conducted for use and usability data. Interviews were systematically analyzed to facilitate identification of patterns and themes. Secondary data were analyzed using descriptive statistics. Within- and between-group effect sizes were reported.

Results: In total, 72% (33/45) and 60% (27/44) of the experimental and comparison groups, respectively, accessed their tool weekly. The mean total usability score was high and moderately high for the experimental and comparison groups, respectively (mean 4.01, SD 0.37 and mean 3.81, SD 0.51; *P*=.048), based on a 5-point Likert scale). Children in both groups found the tool acceptable, and few reported difficulties logging in or accessing content. Between-group effect sizes for beliefs and attitudes related to dietary intake, while favoring the experimental group at T3, were in the small range. These improvements in both groups were largely diminished by T4. However, between-group effect sizes for behaviors related to fruit, vegetable, and water intake, favoring the experimental group, were medium to large and were maintained at T4.

Conclusions: This pilot feasibility study suggests that an interactive comic tool may be an appealing and useful format to promote positive dietary behaviors in racial and ethnic minority preadolescents. However, further research, including a full-scale randomized controlled trial, is warranted to determine the effectiveness of Intervention INC.

XSL•FO
RenderX

Trial Registration: ClinicalTrials.gov NCT03165474; https://clinicaltrials.gov/study/NCT03165474
International Registered Report Identifier (IRRID): RR2-10.2196/10682

(JMIR Form Res 2025;9:e58460) doi:10.2196/58460

KEYWORDS
childhood obesity; preadolescents; racial and ethnic minority populations; dietary behaviors; BMI; digital health

Introduction

Background

The prevalence of childhood obesity in the United States, especially among racial and ethnic minority communities with low incomes, remains consistently high and shows no signs of stabilizing. Specifically, Hispanic and Black or African American children have markedly higher rates of obesity (25.8% and 22%, respectively) compared to the national prevalence of 18.5% [1], with rates rising as age progresses, especially in African American and Hispanic children [2]. Compared to school-age children (18.4%) and preschoolers (13.9%), adolescents (20.6%) have a higher prevalence of obesity [3,4]. In the short term, obesity in youth may lead to issues such as an increased risk of elevated blood pressure and lipid concentrations, anxiety, and depression [5,6]. In the long term, adverse health conditions, including cardiovascular disease, diabetes, and disabilities, can manifest in youth who are obese and be maintained into adulthood [7-10].

Childhood obesity has been linked to several behavioral determinants, namely, increased consumption of highly processed, energy-dense foods and sugary beverages and a decreased intake of fruits and vegetables [11-13]. With the pervasiveness of daily technology and new media use in youth, particularly within the Hispanic and African American population [14], digital platforms are increasingly being used to deliver creative messaging around health and nutrition to children and their caregivers [6]. These tools, which include mobile apps and interactive websites, are emerging as promising approaches to engage and deliver tailored interventions to support health behavior change [15]. However, knowledge gaps in the usability, engagement, and usefulness of such applications still exist, limiting their potential effectiveness [16]. Furthermore, there is a shortage of effective health promotion interventions that are culturally tailored to address the needs and preferences of populations with disproportionately high rates of chronic disease. Even fewer interventions have been designed specifically for children, particularly those transitioning from childhood to adolescence [17]. It is crucial to intervene during this development stage as food preferences established during this period often persist into adulthood [18-20]. In addition, children at this age are gaining autonomy and developing decision-making skills related to dietary behaviors [21], which highlights the importance of intervening at this critical period.

Study Purpose

The primary purpose of this pilot study was to assess the feasibility and participant acceptability of Intervention INC (Interactive Nutrition Comics for Urban, Minority Preadolescents), an interactive, web-based health promotion tool. In addition, this study aimed to explore tool effectiveness by determining if it improved attitudes, beliefs, and behaviors related to obesity. Taken together, these data will provide insight for further improvements to the tool and reveal user preferences, which is important feedback for developing culturally tailored, effective, and engaging digital tools, particularly for populations at risk for childhood obesity.

Methods

Study Design

The Intervention INC study was a pilot, 2-group randomized controlled trial (RCT) that evaluated a 6-week intervention promoting fruit, vegetable and water consumption, with a 3-month follow-up (T4) period. A protocol for this feasibility trial was published and registered with the Clinical Trials Registry (NCT03165474) [22]. Parent-child dyads were randomized into either the experimental or comparison group, with both child and parent participants blinded to their assigned group. Details regarding the randomization process are provided in the protocol paper [22].

Participants

Children residing in New York City (NYC) were recruited based on the following inclusion and exclusion criteria: self-identified as Black, African American, or Latino; aged between 9 and 12 years (preadolescents) at the time of scheduled baseline (T1) visit; read and spoke in English; had a BMI percentile ≥5% at T1 (categorized as healthy, overweight, or obese); had regular internet access via a tablet device, mobile phone, or computer or laptop; had regular access to a phone with texting capability; was comfortable reading or viewing material on electronic devices; was comfortable speaking with study staff about thoughts or experiences while participating in the study; had no allergies, food aversions, food disorders, or medications with side effects that may impact participation in the study; did not have a pacemaker or heart condition; and had a legal parent or guardian willing to participate in the study. While parents or guardians were a part of this study, the focus of this paper is to present primary outcome data for the child participants.

Multiple recruitment approaches were implemented between August and November 2017. Initially, mailings were sent to parents or guardians of potentially eligible child patients of a community-based clinic (partnering organization) in Harlem, NYC. Recruitment efforts were expanded to include community flyers in East Harlem and Upper Manhattan as well as posting notices in and around local businesses, housing complexes, community centers, schools, and churches. Through partnerships with local schools and community initiatives, recruitment efforts

XSL•FO
RenderX

also occurred via tabling at community and school events. Further details regarding study recruitment and sample size considerations are included in other publications [22,23].

Intervention

Content for the child participants, regardless of group assignment, was tailored (to include more information related to either fruit and vegetable or water) based on responses to initial screening questions, at T1, related to child fruit, vegetable and water intake, child self-efficacy to increase fruit, vegetable and water intake, and parent self-efficacy to support the child in increasing fruit, vegetable and water intake.

Experimental Group

Children in the experimental group received access to Intervention INC, an interactive web-based comic tool informed by the Health Belief Model [24], Social Cognitive Theory [25], and the Narrative Transportation Theory [26], which contained health messages focused on fruit, vegetable and water consumption. It comprises a 6-chapter comic with pop-up windows (highlighting health facts, food-related fun facts, or character information) and prompted audio and visual effects to enhance tool engagement. It also includes a weekly goal setting and assessment feature and texting or email messages reminding children to read the comic and work on their selected health behavior goals. Most of the health messages are delivered by the comic characters themselves in a narrative form, but some messages were delivered didactically (such as in the pop-up windows), as entertainment education research suggests a combination of narrative and nonnarrative information may be the most effective in improving health beliefs and behaviors [27]. Multiple behavior change techniques (BCTs) were used to promote healthy dietary behaviors. Table 1 lists all the BCTs that were incorporated into the comic and examples of how each of them were used. The comic tool was designed and developed by a collaborative research team, consisting of scientists, digital health experts, content and system developers, and intended users (Black or African American and Hispanic preadolescents and their parents). Extensive formative research, including usability testing, was conducted during the development process [28]. Additional details of the intervention, including its development and specific components, have been described in previous papers [22,28].

XSL•FO
RenderX

Table 1. BCTs[a] incorporated into the web-based comic tool, Intervention INC (Interactive Nutrition Comics for Urban, Minority Preadolescents).

BCT	Action	Example
Goal setting (behavior)	At the end of each comic chapter, a character prompts the child to select a weekly goal to work on.	• End of chapter message includes a picture of a character along with a prompt to set a goal: "Let's pick a goal to work on for the week!" • Sample goal: "I will eat fruits I like (such as bananas or apples) at breakfast."
Self-monitoring of behavior	The following week, the child is prompted to assess how they did with the weekly goal they selected to work on.	• End of chapter message: "Your goal for last week was (selected goal)." • "How often did you do this in the last week?" • Response options: never, sometimes, most of the time, and all of the time
Feedback on behavior	On the basis of the child's self-assessment of the weekly goal, feedback is provided on the performance of the behavior.	• Sample message if a child achieves the goal: "Congrats! Keep up the good work!" • Sample message if a child does not achieve the goal: "That's ok! Michael Jordan was cut from his high school basketball team, and he didn't give up!"
Social support	Children can select weekly goals related to practical social support from their parents.	• Sample goal related to practical parental social support: "I will ask my parent to pack a water bottle for me when I leave the house this week."
Information about health consequences	Information about the consequences of eating healthfully or unhealthfully is provided in the comic in didactic and narrative forms, such as pop-up messages, character dialogue, and end of chapter messages from characters.	• Sample pop-up message (in a panel where one character tells another to stay hydrated): "Did you know that not having enough water can make you feel dehydrated? Grab a water if you're feeling tired or you're having a hard time concentrating."
Salience of consequences	Scenes are included in the comic to specifically emphasize the consequences of performing both positive and negative dietary behaviors, with the aim of making them more memorable.	• A meter highlights the energy levels of the comic characters based on the nutritional quality of the foods they consume. For example, a character eats a banana, and his energy meter rises. He also acknowledges how "great" he feels.
Demonstration of behavior	Comic characters model positive dietary behaviors.	• The father of the main character is seen drinking water and eating fruits and vegetables in multiple scenes.
Prompts or cues	Messages from the comic characters are delivered weekly (via texting, email, or both) to remind children to work on the goal they selected that week.	• Sample text or email message from a character: "Hey, [child name]—We have a mission for you! Be sure to log in and choose a goal if you haven't already. Then you'll get tips to train and get stronger! See you soon!"
Social reward	Positive feedback is provided if the child achieves their weekly goal.	• Bonus content related to the characters and comic story is released if the child achieves their weekly goal.

[a]BCT: behavior changes technique.

Comparison Group

Children in the comparison group received access to 6 web-based newsletters. Health-related content for the comparison group was similar to that for the experimental group but presented in a newsletter (didactic) format, which includes tips, recipes, diet-related knowledge and facts, and web-based games. Similar to the experimental group, the comparison group also had a weekly goal setting and assessment component and received reminder messages via texting, email or both.

Data Collection and Measures

Overview

Data were collected at 4 different time points: T1, intervention midpoint or 3 weeks after T1 (T2), intervention end or 6 weeks after T1 (T3), and T4. Brief descriptions of our primary and secondary outcome measures are provided in the following sections, while details of all measures are described elsewhere [22]. T1 and T4 data collection was conducted in person at the study site, while T2 and T3 data collection was conducted via phone. At T2 and T3, links to any questionnaires were sent to the participants ahead of time to complete, and then, responses were confirmed during the phone call. If the questionnaire was not completed before the call, children completed it during the call with a trained research assistant. Table 2 lists the measures collected and the time points at which they were collected.

XSL•FO
RenderX

Table 2. Key primary and secondary measures, data sources, and time points at which the measures were assessed.

Measures	Data source	Time points				
		T1[a]	T2[b]	T3[c]	T4[d]	O[e]
Feasibility and acceptability measures (primary measures)						
Use of the web-based tool	Tracking system (internally created)					✓
Usability of the web-based tool	Interview and questionnaire items[f]		✓	✓	✓	
Acceptance and satisfaction	Interview				✓	
Outcome measures (secondary measures)						
Dietary beliefs and attitudes	Questionnaire items[f]	✓	✓	✓	✓	
Dietary intake	Questionnaire items[f]	✓	✓	✓	✓	
BMI-for-age percentile	Digital stadiometer	✓			✓	

[a]T1: baseline.

[b]T2: midpoint (3 weeks after baseline).

[c]T3: intervention end (6 weeks after baseline).

[d]T4: follow-up (3 months after intervention).

[e]O: ongoing throughout the intervention period.

[f]All questionnaire items were either taken from or directly informed by validated questionnaires.

Feasibility and Acceptability Measures (Primary Measures)

Use of the Web-Based Comic Tool

Use data, such as the number of weeks logged in, were captured throughout the intervention on a custom-built platform. Additional use data related to goal setting were collected, including the percentage of participants who selected, assessed (ie, "How often did you do this [behavior] in the last week?"), and achieved (defined as responses of "Most of the Time" or "All the time") weekly dietary goals.

Usability of the Web-Based Comic Tool

Five domains of usability (usability, usefulness, ease of use, ease of learning, and satisfaction) were assessed with a 30-item questionnaire on a 5-point Likert scale (ranging from strongly disagree to strongly agree). This questionnaire was adapted from the System Usability Scale [29]; Usefulness, Satisfaction, and Ease-of-use questionnaire [30]; and an acceptability or usability measure by Ben-Zeev et al [31]. Minor modifications were made to tailor the questionnaire to the literacy levels of our intended population based on pilot testing with them. Specific details related to the modification process are described elsewhere [22]. The modified questionnaire was not validated for web use. A total score of 4.0 out of 5.0 was considered high usability.

Acceptance and Satisfaction

Semistructured interviews were conducted at T2 and T3 where open-ended questions were asked to both experimental and comparison group participants about the experience of engaging with their web-based tool. Specifically, interview questions aimed to assess participant acceptability and satisfaction with the overall tool and its specific components.

Outcome Measures (Secondary Measures)

Dietary Beliefs and Attitudes

Outcome expectations (OEs), self-efficacy, behavioral intention (BI), and attitudes related to fruit, vegetable, water, junk food, and sugary drinks consumption were assessed at all 4 time points with an 84-item questionnaire, which was informed by and modified from the validated ProChildren Questionnaire and the validated Reynolds Questionnaire [32,33]. The OE, self-efficacy, BI, and attitudes questionnaires had a Cronbach α of 0.87, 0.74, 0.52, and 0.70, respectively.

Dietary Intake

Frequency of consumption of fruit, vegetable, water, junk food and sugary drinks during the past 7 days were assessed at all 4 time points with a 17-item questionnaire comprising questions from and informed by the validated 2017 Youth Risk Behavior Surveillance System questionnaire and the validated Beverage and Snack Questionnaire [34,35]. Three items assessing the intake of different types of water were internally created. The dietary intake questionnaire had a Cronbach α of 0.79.

BMI-for-Age Percentile

The height and weight of child participants were measured at T1 and T4 using standardized methods [36]. A SECA 264 digital stadiometer and Tanita MC-780U body composition monitor were used to collect anthropometric measures. The Centers for Disease Control and Prevention BMI percentile calculator was used to determine BMI-for-age percentile [37].

Data Analysis

Feasibility and Acceptability Measures (Primary Measures)

Use and Usability of the Web-Based Comic Tool

Use of the web-based tool for each group was assessed by first aggregating the "click-tracking" data by participant, URL, study

XSL•FO
RenderX

week, and across the intervention period. Descriptive analyses, including means, SD, and ranges (minimum-maximum) were performed. Similar descriptive analyses were also conducted to measure the overall score for the usability data of the web-based tools across both groups.

Acceptance and Satisfaction

Interviews from T2 and T3 were systematically analyzed by 3 trained coders through a streamlined process of directly listening to individual audio files and then transferring relevant information into a matrix sheet to produce a summary of overall impressions of usability, acceptability, and feasibility of the tool and its specific components (ie, reminders and goal setting). A rapid memo technique, informed by the Rapid Evaluation and Assessment Method [38] and the Rapid Identification of Themes From Audio Recordings [38] analysis method, was used to summarize interview responses as well as to note key quotations. These techniques have been used as a way to preclude the traditional steps of transcribing qualitative data and coding transcripts when resources are limited and quick reporting time is needed [38,39].

Audio files were randomly assigned to each coder. In addition to the primary coder, a secondary coder reviewed the audio files for accuracy, noting any additional information, until very few additions were being made or discrepancies were noted. This process was used for approximately the first 10 interviews. However, when a coder had a subsequent audio file with inaudible sections, another coder conducted a secondary review and added any missing information to the matrix. Coders met frequently throughout the analysis process to compare their results; discuss overall matrix content and trends; identify, clarify, and understand emergent findings; and highlight representative quotations.

Outcome Measures (Secondary Measures)

Child and parent sociodemographic characteristics were compared between conditions (experimental or comparison) using 2-tailed independent samples *t* tests and chi-square tests. To assess improvements in outcome measures (child dietary beliefs and attitudes, dietary intake, and BMI-for-age percentile), change scores (relative to T1) were computed and summarized using descriptive statistics. To estimate the magnitude of the effects at T3 and T4, both within- and between-group effect sizes were computed on change scores. Analyses were completed using SPSS statistical software (version 26; IBM Corp) [40]. Of note, the data analyst was blinded to the intervention assignment to minimize bias.

Ethical Considerations

All study activities were approved by the Institutional Review Board at Hunter College in New York, New York (2015-0547). Participation was entirely voluntary, and study participants could choose to withdraw from the study, without any reason, at any point in time during the study. Informed consent, child assent and parental permission were obtained before study participation. Both child and parent participants were compensated. Children received up to US $70 and parents or guardians received up to US $65 (in the form of gift cards) for completing data collection. To maintain participant privacy and confidentiality, all study data have been deidentified.

Results

Participants

A total of 89 child-parent dyads were recruited and completed T1 measures. There were no differences in characteristics between the 2 groups. The mean age of child participants was 10.38 (SD 1.03) years; the majority were female (54/89, 61%), Black or African American (42/89, 47%) or Hispanic (29/89, 33%), and overweight or obese (47/89, 53%). The mean age of parent participants was 40.83 (SD 8.87) years; nearly all were female (84/89, 94%), born in the United States (63/89, 71%), with some college education or more (61/89, 68%), and single (40/89, 45%). One-third (30/89, 34%) of the dyads reported an annual household income of <US $20,000, and nearly two-thirds (54/89, 61%) reported receiving Supplemental Nutrition Assistance Program benefits. Refer to Table 3 for additional demographic details. The CONSORT (Consolidated Standards of Reporting Trials) diagram (Figure 1) shows the randomization and retention progress throughout the study. Retention rates were high: T2, 87% (77/89); T3, 89% (79/89); and T4, 84% (75/89) [23].

XSL•FO
RenderX

Table 3. Baseline child and parent demographic characteristics (N=89).

Characteristics	Overall (N=89)	E[a] (n=45)	C[b] (n=44)	*P* value[c]
Child				
Age (y), mean (SD)	10.38 (1.03)	10.38 (1.06)	10.39 (1.01)	.97
Sex n (%)				
Male	35 (39)	17 (38)	18 (41)	.76
Female	54 (61)	28 (62)	26 (59)	—[d]
Race or ethnicity, n (%)				
Hispanic	37 (42)	20 (51)	17 (49)	.58
Black or AA[e]	55 (62)	28 (54)	27 (46)	.93
Multiracial or multiethnic	18 (20)	10 (56)	8 (44)	.64
BMI-for-age percentile, mean (SD)	74 (26.5)	75 (26.4)	73 (26.8)	.66
BMI category, n (%)				
Normal	42 (47)	21 (47)	21 (48)	.91
Overweight	19 (21)	9 (20)	10 (23)	—
Obese	28 (32)	15 (33)	13 (30)	—
Parent				
Age (y), mean (SD)	40.83 (8.87)	40.07 (7.67)	41.63 (10.01)	.41
Sex, n (%)				
Male	5 (6)	2 (4)	3 (7)	.63
Female	84 (94)	43 (96)	41 (93)	—
Race or ethnicity, n (%)				
Black or AA	49 (55)	25 (56)	24 (55)	.93
Hispanic	40 (45)	22 (49)	18 (51)	.45
Multiracial or multiethnic	9 (20)	6 (13)	3 (7)	.31
Country of birth, n (%)				
United States	63 (71)	31 (69)	32 (73)	.69
Other	26 (29)	14 (31)	12 (27)	—
Bilingual, n (%)				
No	62 (70)	31 (69)	31 (71)	.87
Yes	27 (30)	14 (31)	13 (30)	—
Education, n (%)				
Less than high school	18 (21)	11 (25)	7 (16)	.45
Finished high school or GED[f]	10 (11)	6 (14)	4 (9)	—
Some college	26 (230)	10 (23)	16 (36)	—
Finished college	34 (39)	17 (39)	17 (39)	—
Marital status, n (%)				
Single	40 (45)	22 (49)	18 (41)	.70
Married or in a marriage-like relationship	35 (39)	17 (38)	18 (41)	—
Separated, divorced, or widowed	14 (16)	6 (13)	8 (18)	—
Annual household income (US $), n (%)				
<20,000	30 (34)	15 (33)	15 (34)	.66
20,000-39,999	30 (34)	17 (38)	13 (30)	—

XSL•FO
RenderX

Characteristics	Overall (N=89)	E[a] (n=45)	C[b] (n=44)	*P* value[c]
>40,000	29 (33)	13 (29)	16 (36)	—
SNAP[g] benefits, n (%)				
Yes	54 (61)	27 (60)	27 (61)	.90
No	35 (39)	18 (40)	17 (39)	—

[a]E: experimental group.

[b]C: comparison group.

[c]*P* values are based on independent sample *t* tests (continuous variables) or chi-square tests (categorical variables). Conclusions are based on one *P* value, apart from race or ethnicity categories, which are not mutually exclusive.

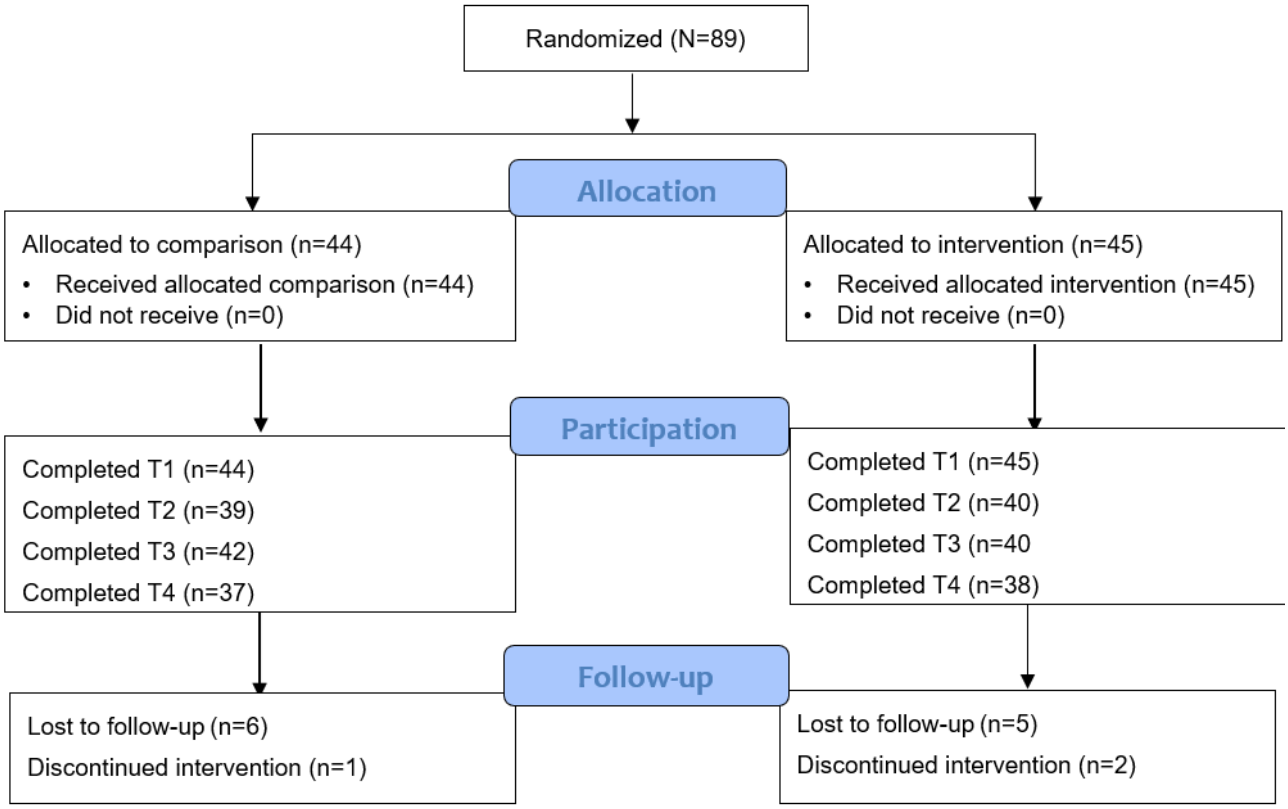
[d]Not applicable.

[e]AA: African American.

[f]GED: General Educational Development.

[g]SNAP: Supplemental Nutrition Assistance Program.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram of the randomization and retention progress throughout the study. T1: baseline; T2: midpoint (3 weeks after baseline); T3: intervention end (6 weeks after baseline); T4: follow-up (3 months after intervention).



Feasibility and Acceptability Measures (Primary Measures)

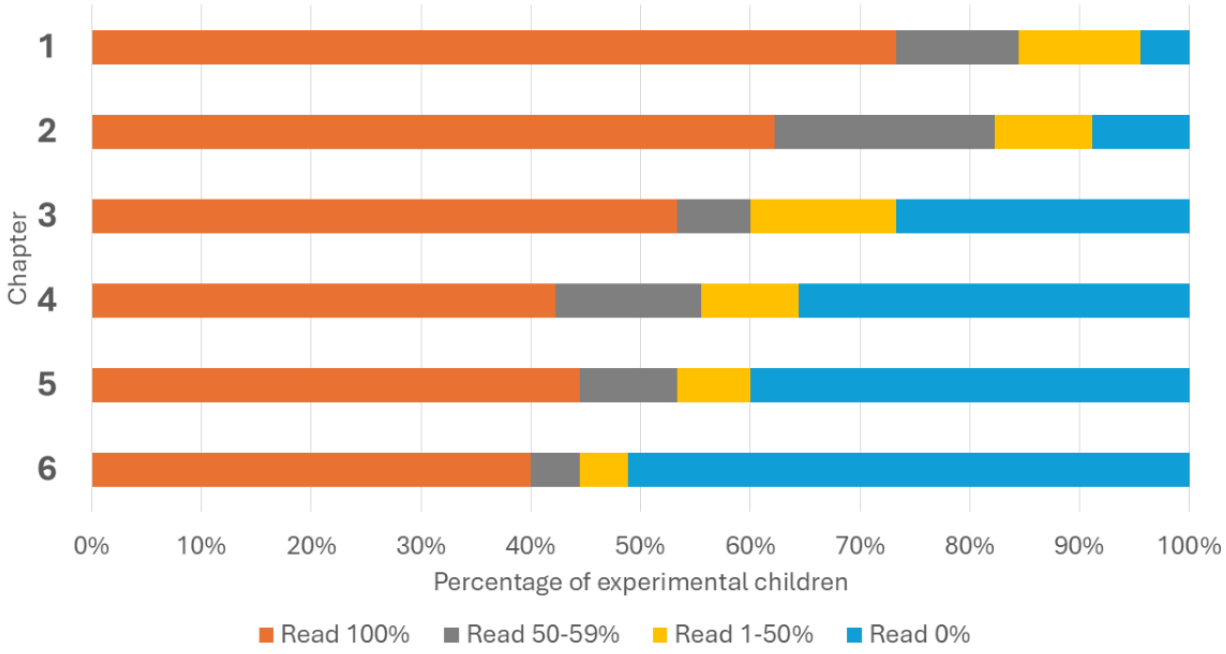
Use of the Web-Based Comic Tool

Overview

During the 6-week intervention, 73% (33/45) of children in the experimental group accessed their tool weekly. Percentage of those who accessed the tool each week were as follows: week 1, 100% (45/45); week 2, 76% (34/45); week 3, 71% (32/45); week 4, 64% (29/45); week 5, 62% (28/45); and week 6, 56% (25/45). The experimental group accessed the web tool an average of 7.84 (SD 5.09) days, out of 42 possible days. Figure 2 highlights the percentage of children who accessed each of the 6 comic chapters. Of those who accessed the comic chapters, an average of 65% (SD 16.8) of each comic chapter was read, while most of the children (24/44, 53%) read the chapters in full.

XSL•FO
RenderX

Figure 2. Percentage of the 6 comic chapters that were read by children in the experimental group.

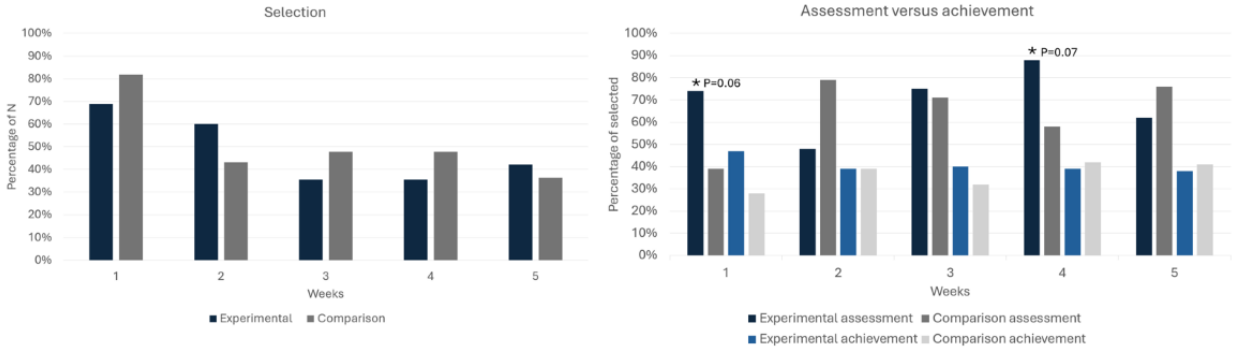


In the comparison group, 60% (27/44) accessed their newsletter weekly. Percentage of those who accessed the tool each week were as follows: week 1, 100% (44/44); week 2, 50% (22/44); week 3, 52% (23/44); week 4, 59% (26/44); week 5, 59% (26/44); and week 6, 41% (18/44). Furthermore, the comparison group accessed the web tool an average of 5.82 (SD 4.20) days, out of 42 possible days.

Goal Setting Feature

On average, 49% (22/45) of children in the experimental group selected a goal each week. Of all the weekly goals that were selected, 67.9% (74/109) were assessed, while 40% (30/74) of those goals were reported to have been achieved. For the comparison group, on average, 51% (23/44) of the children selected a goal. Of all the weekly goals selected, 61.7% (71/115) were assessed. Of the assessed goals, 37% (26/71) were reported to have been achieved. Figure 3 presents the percentage of those in the experimental and comparison groups who selected, assessed, and achieved a goal during each week of the intervention. There was a general decline in goal selection over the weeks in both experimental and comparison groups—in week 1, 69% (31/45) and 82% (36/44) selected goals; by week 5, 42% (19/45) and 36% (16/44) selected goals, respectively. In terms of goal assessment, trends were observed between the experimental and comparison groups in week 1 (P=.06) and week 4 (P=.07), however, these results were not statistically significant. In week 1, 74% (23/31) and 39% (14/36) assessed goals they selected, while in week 4, 88% (14/16) and 57% (12/21) assessed goals for experimental and comparison groups, respectively (P=.07 and P=.10, respectively). No other differences were observed across the other weeks. In terms of achieving the weekly goals, the experimental group slightly declined over the weeks (week 1; 11/23, 46% to week 5; 5/12, 42%), while there appeared to be an increase in the comparison group (week 1; 5/14, 36% to week 5; 5/12, 42%).

Figure 3. Percentage of experimental and comparison groups who selected, assessed, and reported achieving weekly goals across the 6-week intervention.



XSL•FO
RenderX

Usability of the Web-Based Comic Tool

Table 4 highlights the mean scores of the 5 usability domains (usability, usefulness, ease of use, ease of learning, and satisfaction) and total for both the experimental and comparison groups. The experimental group reported moderate to high scores in each of the domains. Specifically, ease of learning and ease of use had the highest scores relative to their maximum range (4.58, SD 0.72 and 4.47, SD 0.46, respectively). The mean total score of perceived usability was high and moderately high for the experimental and comparison groups, respectively (4.01, SD 0.37 and 3.81, SD 0.51; *P*=.048).

Table 4. Perceived usability of the assigned tool at T3[a], by group[b].

Domain	E[c] (n=43), mean (SD)	C[d] (n=41), mean (SD)	*P* value
Usability (30 items)	3.14 (0.44)	3.05 (0.40)	—[e]
Usefulness (10 items)	4.35 (0.68)	4.02 (1.11)	—
Ease of use (10 items)	4.47 (0.46)	4.25 (0.66)	—
Ease of learning (2 items)	4.58 (0.72)	4.40 (0.82)	—
Satisfaction (6 items)	4.36 (0.62)	4.07 (0.82)	—
Total	4.01 (0.37)	3.81 (0.51)	.048

[a]T3: intervention end point or 6 weeks postbaseline.

[b]Assessment questionnaire was developed using a total of 30 items from the System Usability Scale [41]; Usefulness, Satisfaction, and Ease-of-use [42] questionnaire; and acceptability or usability measure [43]. Response options ranged from 1 (strongly disagree) to 5 (strongly agree).

[c]E: experimental group.

[d]C: comparison group.

[e]*P* values for individual usability domains have been removed as they were determined to not be applicable.

Acceptance and Satisfaction

Overall, children in both arms found their tool to be acceptable, and very few reported any difficulties logging in or accessing content. Specific findings for different components of the comic tool are noted in the following sections.

Comic

Most children shared that the comic is a "fun way to read" as it's exciting, adventurous, motivating, and "helpful like a video game." Some children explained that the storyline teaches new things about fruits, vegetables, and water and how to stay healthy and take care of oneself. One male child participant described it as "a superhero story where being nutritious is the only way to save the world and junk food slowly starts depleting your energy and, in the end, makes you an awful person." Most children also reported engaging a parent or family member in their experience of reading the comic.

The comic included a variety of elements aimed at engaging children, including unique comic characters and audio-visual effects. Most children liked the comic characters because of their personalities, which were described as "cool," "different," "mysterious," "admirable" and "brave," among other adjectives. Some children reported that the characters gave them a good feeling or that they felt a connection to them as they reminded them of someone in their life or shared a favorite hobby. Moreover, nearly all children reported liking the audio or visual effects. Reasons included that the effects made the story more imaginable, "fun to read" and it "gives you a feeling that you are there...in the story." In addition, some children noted that the audio effects allowed the tool to read for them "if you don't want to (or can't) read" and "it's a pronunciation helper."

Pop-ups were also used to present additional information related to healthy foods and the characters. Nearly all children liked the pop-ups, describing them as "interesting," informative," "surprising," "cool," and "pretty funny." Some children shared that the information pop-ups taught them something they never knew. A male child liked the information pop-ups "because they're telling stuff that I didn't really know about; stuff I didn't know about carrots, I know now that's why I like it." Some of the reasons shared for liking the character pop-ups included that they helped with developing a deeper understanding of the characters and their role in the story. As 1 male child shared, he "liked [them] because they tell [me] where are they from, what do they like...basically their ideals. It's like their ID."

Goal Setting

Most children in both groups reported selecting weekly goals both at T2 and T3 time points and acknowledged the "motivation" and "confidence" generated by the process. Some children who did not set weekly goals cited time constraints and involvement in after-school activities as their primary reasons. Most children found the work around goal selection and achievement to be manageable and also easier if the items needed (such as fruit, vegetables, and water) were already available at home and if parental help was accessible to obtain them. Many children also acknowledged that the reminders to work on goals were helpful.

Reminders

Most children in both arms preferred text message reminders over email reminders, saying that they were enjoyable to receive, helpful, easier to understand, more convenient, and could be checked more often. In relation to dose and frequency of messaging, across both arms, most children reported receiving "enough messages," with only a few concluding that they

XSL•FO
RenderX

received too many, citing an annoyance at being interrupted or inundated, and preferring a frequency of 2 messages per week.

Outcome Measures (Secondary Measures)

Dietary Beliefs and Attitudes

At T3, both experimental and comparison groups reported a substantial and consistent increase in OE, self-efficacy, and BI related to fruit and vegetable consumption. In terms of water consumption, both groups reported a substantial increase in BI. The experimental group also reported an increase in self-efficacy for water intake, which was maintained at T4, and the comparison group showed an increase in attitudes toward water consumption. Changes related to OE, self-efficacy, BI, and attitudes toward sugary drinks and junk food consumption were generally small and mixed. In summary, the within-group effect sizes for child dietary beliefs and attitudes were generally larger in the experimental group at T3; however, the majority diminished by T4. The between-group effect sizes, while favoring the experimental group, were in the small range. Refer to Table 5 for the change scores for all psychosocial variables for the experimental and comparison groups and Figure 4 for the effect sizes comparing the experimental and comparison groups. Descriptive statistics by group and child gender are presented in Multimedia Appendix 1.

Table 5. Change scores for psychosocial variables related to dietary behaviors from T1[a] to T3[b] and T1 to T4[c], by group.

	Outcome expectations				Self-efficacy				Behavioral intention				Attitudes			
	E[d] (n=37), mean change (SD)	C[e] (n=39), mean change (SD)	ES[f,g] within E	ES within C	E (n=37), mean change (SD)	C (n=39), mean change (SD)	ES within E	ES within C	E (n=37), mean change (SD)	C (n=39), mean change (SD)	ES within E	ES within C	E (n=37), mean change (SD)	C (n=39), mean change (SD)	ES within E	ES within C
Fruit and 100% fruit juice																
T1 to T3	2.17 (3.43)	2.05 (3.70)	0.63	0.55	0.64 (2.41)	0.59 (1.91)	0.27	0.31	0.45 (1.80)	0.61 (1.61)	0.25	0.38	0.45 (2.49)	–0.13 (2.49)	0.18	–0.05
T1 to T4	1.00 (3.34)	–0.08 (3.79)	0.30	–0.02	–0.22 (2.78)	–0.74 (2.49)	–0.08	–0.30	0.22 (2.00)	–0.10 (2.19)	0.11	–0.05	–0.36 (2.77)	–0.69 (2.27)	–0.13	–0.30
Vegetable																
T1 to T3	1.69 (3.21)	1.29 (3.76)	0.53	0.34	1.50 (3.09)	1.02 (2.68)	0.48	0.38	0.60 (2.40)	0.80 (1.96)	0.25	0.41	0.98 (3.20)	0.54 (3.51)	0.30	0.15
T1 to T4	0.16 (3.96)	0.03 (4.63)	0.04	0.01	0.65 (3.51)	–0.41 (3.30)	0.19	–0.12	0 (1.90)	–0.08 (2.16)	0	–0.04	0.50 (3.01)	0.38 (3.48)	0.17	0.11
Water																
T1 to T3	1.00 (3.56)	0.78 (5.00)	0.28	0.16	0.31 (1.00)	0.12 (0.95)	0.31	0.13	0.52 (1.80)	0.93 (2.54)	0.29	0.36	0.50 (2.24)	1.17 (3.19)	0.22	0.37
T1 to T4	0.57 (3.09)	–0.67 (3.91)	0.18	–0.17	0.32 (0.94)	0 (0.83)	0.34	0	–0.16 (1.28)	0.62 (2.62)	–0.13	0.23	–0.33 (2.07)	0.56 (3.26)	–0.16	0.17
Sugary drinks																
T1 to T3	–0.10 (4.01)	–0.07 (3.29)	–0.02	–0.02	0.21 (2.09)	–0.30 (1.73)	0.10	–0.17	0.55 (2.44)	–0.37 (3.02)	0.22	–0.12	0.10 (3.14)	–1.44 (3.24)	0.03	–0.44
T1 to T4	–0.22 (3.31)	–0.38 (2.98)	–0.07	–0.13	–0.19 (2.01)	0.08 (2.07)	–0.10	0.04	0.11 (2.17)	0.49 (2.79)	0.05	0.17	–0.27 (2.93)	–0.51 (3.11)	–0.09	0.16
Junk food																
T1 to T3	0.26 (2.56)	0.32 (3.09)	0.10	0.10	–0.14 (2.42)	0.46 (1.64)	–0.06	0.28	0.40 (2.58)	0.55 (3.31)	0.16	0.17	1.17 (3.01)	–0.63 (2.84)	0.39	–0.22
T1 to T4	0.43 (2.24)	0.41 (2.60)	0.19	0.16	–0.24 (2.02)	0.18 (1.35)	–0.12	0.13	0.06 (2.62)	1.21 (2.56)	0.02	0.47	0.41 (3.13)	0.77 (2.16)	0.13	0.36

[a]T1: baseline.

[b]T3: intervention end or 6 weeks postbaseline.

[c]T4: 3-month follow-up postintervention.

[d]E: experimental group.

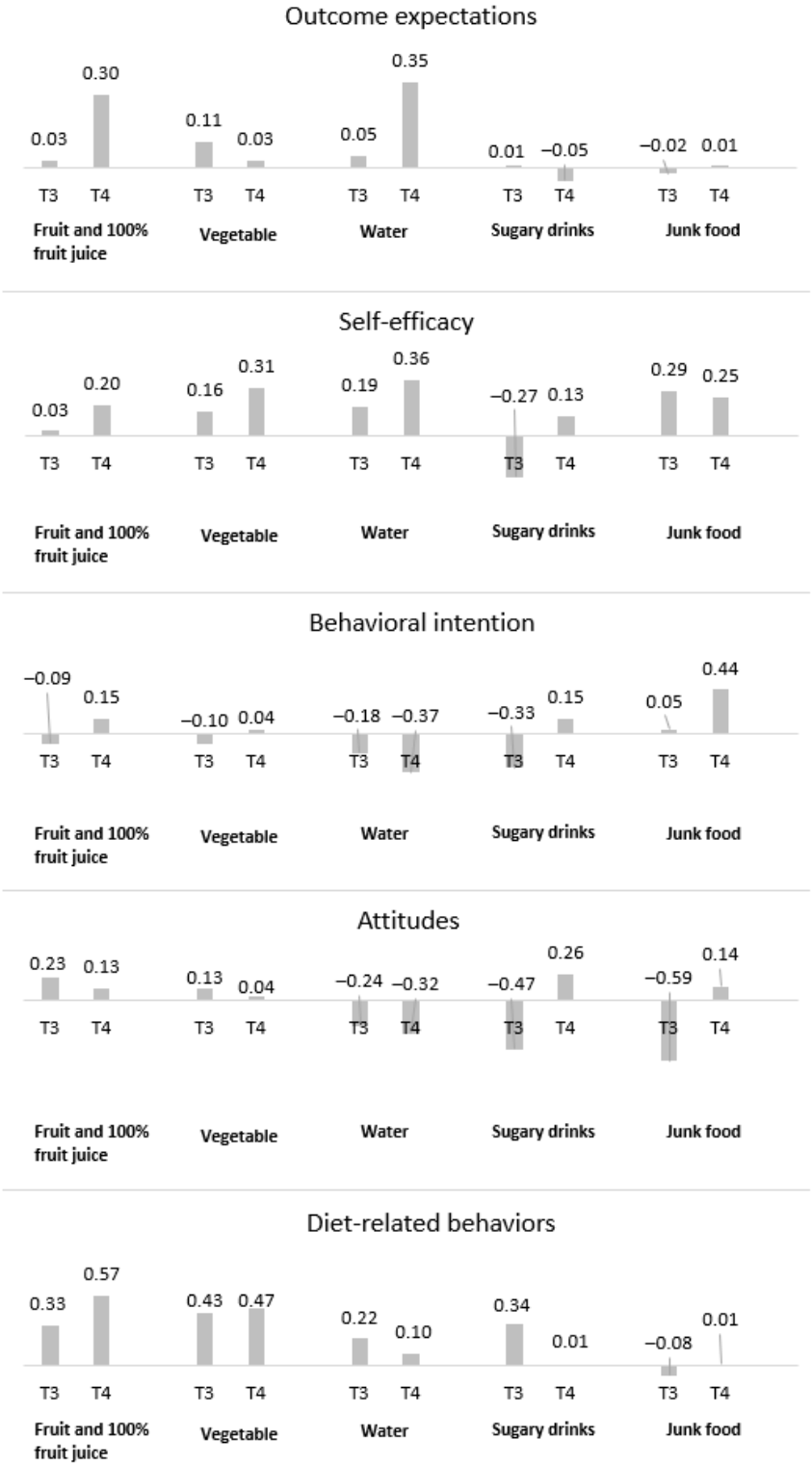
[e]C: comparison group.

[f]ES: effect size.

[g]"ES within" represents changes in score divided by the SD of change within the condition (E and C, respectively); positive ES indicates improvement, except for junk food and sugary drinks, where negative ES indicates improvement.

XSL•FO
RenderX

Figure 4. Effect sizes at intervention end or 6 weeks after baseline (T3) and 3-month follow-up after the intervention (T4), comparing experimental (E) and comparison (C) groups on psychosocial variables and self-reported diet-related behaviors. Positive effect size indicates a larger improvement in the E compared to the C group.



Dietary Behaviors

The experimental group reported a substantial increase in fruit, vegetable, and water intake, which were maintained at T4. Changes in the comparison group were small. Both groups reported a decrease in junk food intake at T3, and the experimental group additionally reported a decrease in sugary drinks intake from T1 to T3; however, these improvements were diminished at T4. The between-group effect sizes for the child dietary intake, favoring the experimental group, were medium to large. Refer to Table 6 for the changes in self-reported diet-related behaviors and Figure 4 for the related effect sizes at T3 and T4 comparing the experimental and comparison groups. Descriptive statistics by child gender are presented in Multimedia Appendix 1.

XSL•FO
RenderX

Table 6. Change in self-reported diet-related behaviors from T1[a] to T3[b] and T1 to T4[c], by group.

	E[d] (n=37), mean change (SD)	C[e] (n=39), mean change (SD)	ES[f,g] within E	ES within C
Fruit and 100% fruit juice				
T1 to T3	0.71 (5.12)	–0.80 (4.04)	0.14	–0.20
T1 to T4	1.30 (3.39)	–0.79 (3.71)	0.38	–0.21
Vegetable				
T1 to T3	3.10 (7.27)	0.22 (5.82)	0.43	0.04
T1 to T4	3.08 (4.58)	0.51 (5.99)	0.67	0.09
Water				
T1 to T3	2.52 (6.85)	1.00 (6.76)	0.37	0.15
T1 to T4	2.22 (4.84)	1.64 (6.41)	0.46	0.26
Sugary drinks				
T1 to T3	–1.80 (5.31)	0 (5.34)	–0.34	0
T1 to T4	–0.03 (7.06)	0.05 (5.50)	0	0.01
Junk food				
T1 to T3	–1.55 (4.77)	–1.95 (5.63)	–0.32	–0.35
T1 to T4	–1.08 (6.40)	–1.03 (4.91)	–0.17	–0.21

[a]T1: baseline.

[b]T3: intervention end or 6 weeks postbaseline.

[c]T4: 3-month follow-up postintervention.

[d]E: experimental group.

[e]C: comparison group.

[f]ES: effect size.

[g]"ES within" represents changes in score divided by the SD of change within the condition (E and C); positive ES indicates improvement, except for junk food and sugary drinks, where negative ES indicates improvement.

BMI-for-Age Percentile

The mean change between T1 and T4 in BMI-for-age percentile was small and comparable in both experimental and comparison groups (mean –0.58, SD 5.91 and mean –0.50, SD 11.49, respectively). A post hoc exploration by child's sex showed that BMI-for-age percentile tended to decrease in the experimental group for male participants (mean change –3.26, SD 4.67) and the comparison group for female participants (mean change –3.30, SD 9.77). Refer to Table 7 for BMI-for-age percentile for the experimental and comparison groups, overall and by sex, at T1 and T4.

Table 7. BMI-for-age percentile at T1[a] and T4[b] for E[c] and C[d] groups, overall and among male and female participants.

	Overall		Male participants		Female participants	
	E (n=36)	C (n=36)	E (n=14)	C (n=15)	E (n=22)	C (n=21)
T1, mean (SD)	74.94 (26.01)	70.29 (27.29)	73.74 (29.53)	70.13 (28.82)	75.67 (24.29)	70.40 (26.84)
T4, mean (SD)	73.70 (26.77)	70.57 (29.44)	70.48 (32.12)	73.55 (31.07)	75.76 (23.34)	66.44 (28.81)
T1 to T4, mean change (SD)	–0.58 (5.91)	–0.50 (11.41)	–3.26 (4.67)	3.42 (12.68)	–1.12 (6.09)	–3.30 (9.77)

[a]T1: baseline.

[b]T4: 3-month follow-up postintervention.

[c]E: experimental group.

[d]C: comparison group.

Discussion

Principal Findings

The primary purpose of this study was to assess the feasibility and participant acceptability of an interactive, web-based comic tool. In addition, this study explored the tool's potential impact on dietary behaviors, related psychosocial variables, and BMI-for-age percentile.

There was a decline in use of the digital tool (in both groups) over the 6-week intervention period. While direct comparisons are difficult due to the lack of standardization in how engagement has been defined and measured in previous research

XSL•FO
RenderX

[44], low to moderate engagement and a decline in use as an intervention progresses are commonly reported [45-47]. This is concerning given that engagement is critical to the effectiveness of digital health interventions [48]. This was a key consideration during the design process for our tool as we used user-centered design principles throughout the development process [49], with the goal of ultimately developing a culturally relevant and meaningful tool. To enhance potential engagement, we incorporated interactive features into the tool, such as tap or click icons that opened pop-up windows (highlighting health facts, food-related fun facts, or character information, such as a character's favorite healthy food and beverage).

While engagement did decline in both groups, the use data appear to support the feasibility and acceptability of the comic tool as children in the experimental group were more engaged; an average of 72% (33/45) of the experimental group accessed their tool weekly compared to 60% (27/44) of the comparison group. Furthermore, most of the children in the experimental group, who accessed the comic, actually read the complete chapter (which ranged from 18 to 34 pages) each week. In addition to a higher level of engagement, the experimental group also reported higher usability of their tool, which is an indicator of tool acceptability and satisfaction [50,51].

Other interactive features included embedded audio recordings for certain character dialogue, which also presented opportunities to deliver health messages in another format. Such a format can help with the understanding of the health messages promoted within the comic [52], particularly as low literacy levels continue to exist in our intended population [53-55]. Research has found that Black and Hispanic students begin high school with literacy skills 3 years behind those of White and Asian students [53]. In addition, students from families with low incomes enter high school with literacy skills that are 5 years behind those of students from families with high incomes [55]. Feedback from participants acknowledged the benefits of the embedded interactive features as 1 child noted that the audio effects allow the tool to read for the children, "if you don't want to or can't read." Such features can also add to the immersion of the reader into the story, as studies have found that audio features can improve the manga reading experience provided it connects the reader with the objects represented in the scene [56]. This could lead to enhanced persuasion of the story's health messages and, thus, motivate positive behavior change, as suggested by the Narrative Transportation Theory [57,58].

While several BCTs were incorporated into the digital comic tool, a key component of the tool was the goal selection and assessment activity. Research has shown that the BCT of goal setting is associated with positive health outcomes, although primarily related to BMI measures in children above a healthy weight [59]. Children in our study acknowledged that goal achievement was easier if the appropriate support was in place, which included fruit, vegetables, and water being available at home and if parental help was accessible. This highlights the importance of ensuring that healthy foods are accessible to allow the children to be empowered to make positive choices. It also reinforces the importance of engaging parents to not only be positive role models but to also provide a supportive home environment as parent-directed interventions can improve the self-efficacy of their children [60]. Prompts and cues were also included in the comic tool in the form of 4 reminder messages each week from the characters themselves. The frequency of the reminder messages was informed by formative research with our intended population, as the literature is quite limited in this area [61]. While a few children did note that they received too many, overall, the children acknowledged that the reminder function, comprising 4 texting or email messages each week was "enough" and an important feature to promote sustained engagement [61].

Secondary outcomes of this study included changes in psychosocial variables related to specific diet-related behaviors, dietary intake, and BMI-for-age percentile. Results showed that while changes in psychosocial variables for certain diet-related behaviors were observed in both the experimental and comparison groups from T1 to T3, most improvements were diminished by T4. However, improvements in intake related to fruit, vegetable, and water behaviors observed in the experimental group were maintained at T4. Furthermore, while no differences were observed in BMI-for-age percentile change between both groups from T1 to T4, a post hoc analysis found that BMI percentile tended to decrease in the experimental group for male participants and in the comparison group for female participants. The improvement in fruit, vegetable, and water intake that was maintained through the T4 period in the experimental group may partially explain the anthropometric findings observed in the male participants, as such behaviors are recommended to reduce obesity risk [62,63]. Further research is warranted to better understand the mechanisms that may have resulted in the observed trend of the decreased BMI percentile in both groups, particularly as no sustained changes were found in related psychosocial variables and no significant improvements in dietary behaviors were observed in the comparison group.

Sex differences in the BMI percentile outcome are interesting to note. Health-promoting comics have been used effectively to reach children of both genders [41,64,65]; however, in our study, it appears that male participants may have responded more positively to the comic tool. While extensive formative research was conducted with both genders from our intended population, with the goal of developing a broad-reaching, meaningful, and engaging narrative, the main character of the story was male. This may have led to unintended differences in how male and female participants responded to the health messages. Interestingly, prior research by the team found that both male and female children often selected a male character when asked who their favorite comic character was [42]; however, other research has shown that gender does play a role in how people relate and respond to media characters [43,66]. Furthermore, while we aimed to address a broad range of genres in the comic to appeal to both genders, differences in genre preferences do exist between the sexes as females tend to prefer romance, fantasy, and action, while males tend to prefer humor and action when reading comics [67].

Implications for Future Research

Further research is needed to not only understand the elements that influence tool engagement but to also gain a better

XSL•FO
RenderX

understanding of the comic tool's key components that influenced the observed health-related outcomes. This pilot and feasibility study suggests that such a comic tool could be an appealing format to promote positive beliefs and behaviors related to a healthier diet in racial and ethnic minority preadolescents (particularly in male children). In addition, our study had successful recruitment and high retention with our participants [23]. Furthermore, 81% (61/75) of the children reported at T4 that they were very satisfied or extremely satisfied with their interactions with the study staff [23]. Therefore, a full-scale RCT with a longer follow-up period is warranted to determine the effectiveness of Intervention INC. However, given potential gender differences in character relatability and genre preferences, the development of future health-promoting comic tools should more carefully consider how to frame storylines and messaging to effectively impact both genders. Moreover, as newly adopted behaviors often diminish once an intervention ends [68], a maintenance intervention should be considered following the initial intervention period, as it could provide critical continued support and reinforce newly adopted behaviors. In addition, this tool's messages mainly focused on obesity-related dietary behaviors. As physical activity is a key contributing factor to reducing obesity risk, the inclusion of a physical activity component should be considered to produce a more comprehensive tool focused on additional influences related to childhood obesity.

Limitations

It is important to acknowledge that this study is not without its limitations. We were unable to capture more granular use data, such as the minutes participants used the tool itself. Self-report data are always a challenge, in particular with dietary intake data. This limitation is further exacerbated by the fact that our questionnaires, while informed by validated questionnaires commonly used in the literature [25-30], were not validated tools and were completed by the children themselves. However, it should be acknowledged that T1 and T4 data were collected in person where a research team member was present and available to answer any questions and assist the children in completing the questionnaire. Seasonality issues may be a concern with the outcome measures, as recruitment was conducted on a rolling basis over a 3-month period. Thus, some participants completed the study during the fall season, while others completed it during the spring season. Another limitation is the small sample size, thus limiting statistical power. The generalizability of these findings is also limited as the data are unique to racial and ethnic minority preadolescent children residing in underserved NYC communities. In addition, all study participants had to have ongoing access to the internet, further limiting the study's generalizability.

Conclusions

This pilot and feasibility study suggests that the multimedia platform of comics could be an appealing and engaging format to promote healthy diet-related beliefs and behaviors in racial and ethnic minority, urban preadolescents at risk of childhood obesity. Furthermore, as comics are popular among US youth and can be easily disseminated, particularly in digital format, the graphics and minimal text make it a promising format for low-literacy populations. The sex differences observed in BMI percentile suggest that boys may respond more positively to health messages delivered in such a format; however, the comic's main character (who is male) may have been more relatable to our male participants. Thus, future iterations of the comic should carefully consider the main character to ensure they are relatable to all genders. While the results are promising, further research is needed, including a full-scale RCT, to determine the effectiveness of Intervention INC.

Acknowledgments

This research has been supported by the Agency for Healthcare Research and Quality (grant R21H5024117). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality. The authors wish to acknowledge the research team as well as their community partners for their contribution to this study. The authors want to acknowledge Kris Hoyt (lead developer of the intervention platform and interface) and the research assistants who helped to implement this study.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

MML conceived the intervention concept and study design. KFM drafted the study protocol, with critical inputs from MML. KFM coordinated the trial, with the assistance of SV. KFM, LH, and SV were involved in the recruitment process. KW and LH contributed to the statistical analysis and interpretation, and MD contributed to the qualitative data analysis and interpretation. MML led the writing of the manuscript, with KW, KFM, and MD assisting with the process. All authors read and approved the final manuscript.

Conflicts of Interest

MML, KFM, and SV were involved in the development of the Intervention INC (Interactive Nutrition Comics for Urban, Minority Preadolescents) tool. The authors declare that they have no other conflicts of interest related to this study.

XSL•FO
RenderX

Multimedia Appendix 1
Descriptive statistics for psychosocial variables and self-reported diet related behaviors across timepoints, by group and by sex.
[DOCX File , 425 KB - formative_v9i1e58460_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 95 KB - formative_v9i1e58460_app2.pdf]

References

1. Hales CM, Carroll MD, Fryar CD, Ogden CL. Prevalence of obesity among adults and youth: United States, 2015-2016. NCHS Data Brief 2017 Oct(288):1-8 [FREE Full text] [Medline: 29155689]
2. Ogden CL, Martin CB, Freedman DS, Hales CM. Trends in obesity disparities during childhood. Pediatrics 2022 Aug 01;150(2):e2022056547 [FREE Full text] [doi: 10.1542/peds.2022-056547] [Medline: 35788862]
3. Hales CM, Fryar CD, Carroll MD, Freedman DS, Ogden CL. Trends in obesity and severe obesity prevalence in US youth and adults by sex and age, 2007-2008 to 2015-2016. JAMA 2018 Apr 24;319(16):1723-1725 [FREE Full text] [doi: 10.1001/jama.2018.3060] [Medline: 29570750]
4. Skinner AC, Ravanbakht SN, Skelton JA, Perrin EM, Armstrong SC. Prevalence of obesity and severe obesity in US children, 1999-2016. Pediatrics 2018 Mar;141(3):1-18 [FREE Full text] [doi: 10.1542/peds.2017-3459] [Medline: 29483202]
5. Harbin MM, Hultgren NE, Kelly AS, Dengel DR, Evanoff NG, Ryder JR. Measurement of central aortic blood pressure in youth: role of obesity and sex. Am J Hypertens 2018 Nov 13;31(12):1286-1292 [FREE Full text] [doi: 10.1093/ajh/hpy128] [Medline: 30107492]
6. Tate EB, Spruijt-Metz D, O'Reilly G, Jordan-Marsh M, Gotsis M, Pentz MA, et al. mHealth approaches to child obesity prevention: successes, unique challenges, and next directions. Transl Behav Med 2013 Dec;3(4):406-415 [FREE Full text] [doi: 10.1007/s13142-013-0222-3] [Medline: 24294329]
7. Horesh A, Tsur AM, Bardugo A, Twig G. Adolescent and childhood obesity and excess morbidity and mortality in young adulthood-a systematic review. Curr Obes Rep 2021 Sep 05;10(3):301-310. [doi: 10.1007/s13679-021-00439-9] [Medline: 33950400]
8. Skinner AC, Perrin EM, Moss LA, Skelton JA. Cardiometabolic risks and severity of obesity in children and young adults. N Engl J Med 2015 Oct;373(14):1307-1317. [doi: 10.1056/NEJMoa1502821] [Medline: 26422721]
9. Antwi FA, Fazylova N, Garcon MC, Lopez L, Rubiano R, Slyer JT. Effectiveness of web-based programs on the reduction of childhood obesity in school-aged children: a systematic review. JBI Database Syst Rev Implement Rep 2013 Jun;11(6):1-44. [doi: 10.11124/jbisrir-2013-459]
10. Reilly JJ, Kelly J. Long-term impact of overweight and obesity in childhood and adolescence on morbidity and premature mortality in adulthood: systematic review. Int J Obes (Lond) 2011 Jul;35(7):891-898. [doi: 10.1038/ijo.2010.222] [Medline: 20975725]
11. Costa CS, Del-Ponte B, Assunção MC, Santos IS. Consumption of ultra-processed foods and body fat during childhood and adolescence: a systematic review. Public Health Nutr 2018 Jan;21(1):148-159 [FREE Full text] [doi: 10.1017/S1368980017001331] [Medline: 28676132]
12. Campbell MK. Biological, environmental, and social influences on childhood obesity. Pediatr Res 2016 Jan 20;79(1-2):205-211. [doi: 10.1038/pr.2015.208] [Medline: 26484623]
13. SM M. Childhood obesity: epidemiology, determinants, and prevention. J Nutr Disorders Ther 2015;5(2). [doi: 10.4172/2161-0509.1000156]
14. The common sense census: media use by tweens and teens, 2021. Common Sense. 2022 Mar 9. URL: https://www.commonsensemedia.org/research/the-common-sense-census-media-use-by-tweens-and-teens-2021 [accessed 2023-04-22]
15. Delamater AM, Pulgaron ER, Rarback S, Hernandez J, Carrillo A, Christiansen S, et al. Web-based family intervention for overweight children: a pilot study. Child Obes 2013 Feb;9(1):57-63 [FREE Full text] [doi: 10.1089/chi.2011.0126] [Medline: 23308372]
16. Kracht CL, Hutchesson M, Ahmed M, Müller AM, Ashton LM, Brown HM, et al. E- and mHealth interventions targeting nutrition, physical activity, sedentary behavior, and/or obesity among children: a scoping review of systematic reviews and meta-analyses. Obes Rev 2021 Dec;22(12):e13331 [FREE Full text] [doi: 10.1111/obr.13331] [Medline: 34476890]
17. Nguyen B, Kornman KP, Baur LA. A review of electronic interventions for prevention and treatment of overweight and obesity in young people. Obes Rev 2011 May;12(5):e298-e314. [doi: 10.1111/j.1467-789X.2010.00830.x] [Medline: 21348921]
18. Demory-Luce D, Morales M, Nicklas T, Baranowski T, Zakeri I, Berenson G. Changes in food group consumption patterns from childhood to young adulthood: the Bogalusa Heart Study. J Am Diet Assoc 2004 Nov;104(11):1684-1691. [doi: 10.1016/j.jada.2004.07.026] [Medline: 15499355]

XSL•FO
RenderX

19. Mikkilä V, Räsänen L, Raitakari OT, Pietinen P, Viikari J. Consistent dietary patterns identified from childhood to adulthood: the cardiovascular risk in Young Finns Study. Br J Nutr 2005 Jun 08;93(6):923-931. [doi: 10.1079/bjn20051418] [Medline: 16022763]
20. Kelder SH, Perry CL, Klepp KI, Lytle LL. Longitudinal tracking of adolescent smoking, physical activity, and food choice behaviors. Am J Public Health 1994 Jul;84(7):1121-1126. [doi: 10.2105/ajph.84.7.1121] [Medline: 8017536]
21. Golan M, Crow S. Parents are key players in the prevention and treatment of weight-related problems. Nutr Rev 2004 Jan;62(1):39-50. [doi: 10.1111/j.1753-4887.2004.tb00005.x] [Medline: 14995056]
22. Leung MM, Mateo KF, Verdaguer S, Wyka K. Testing a web-based interactive comic tool to decrease obesity risk among minority preadolescents: protocol for a pilot randomized control trial. JMIR Res Protoc 2018 Nov 09;7(11):e10682 [FREE Full text] [doi: 10.2196/10682] [Medline: 30413399]
23. DeFrank G, Singh S, Mateo KF, Harrison L, Rosenthal A, Gorman A, et al. Key recruitment and retention strategies for a pilot web-based intervention to decrease obesity risk among minority youth. Pilot Feasibility Stud 2019 Sep 05;5(1):109 [FREE Full text] [doi: 10.1186/s40814-019-0492-8] [Medline: 31516726]
24. Champion VL, Skinner CS. The health belief model. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice. Hoboken, NJ: Jossey-Bass Publishers; 2008:45-65.
25. McAlister AL, Perry CL, Parcel GS. How individuals, environments, and health behaviors interact: social cognitive theory. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice. Hoboken, NJ: Jossey-Bass Publishers; 2008:169-188.
26. Green MC, Clark JL. Transportation into narrative worlds: implications for entertainment media influences on tobacco use. Addiction 2013 Mar;108(3):477-484. [doi: 10.1111/j.1360-0443.2012.04088.x] [Medline: 22994374]
27. Slater MD, Rouner D. Entertainment—education and elaboration likelihood: understanding the processing of narrative persuasion. Commun Theory 2002 May;12(2):173-191. [doi: 10.1111/j.1468-2885.2002.tb00265.x]
28. Verdaguer S, Mateo KF, Wyka K, Dennis-Tiwary TA, Leung MM. A web-based interactive tool to reduce childhood obesity risk in urban minority youth: usability testing study. JMIR Form Res 2018 Nov 01;2(2):e21 [FREE Full text] [doi: 10.2196/formative.9747] [Medline: 30684417]
29. Brooke J. SUS: a 'quick and dirty' usability scale. In: Usability Evaluation in Industry. Hoboken, NJ: CRC Press; 1996.
30. Lund AM. Measuring usability with the USE Questionnaire PDF. Scribd. 2001 Jan. URL: https://www.scribd.com/document/313132052/Measuring-Usability-with-the-USE-Questionnaire-pdf [accessed 2023-05-26]
31. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov;40(6):1244-1253 [FREE Full text] [doi: 10.1093/schbul/sbu033] [Medline: 24609454]
32. De Bourdeaudhuij I, Klepp KI, Due P, Rodrigo CP, de Almeida M, Wind M, et al. Reliability and validity of a questionnaire to measure personal, social and environmental correlates of fruit and vegetable intake in 10-11-year-old children in five European countries. Public Health Nutr 2005 Apr 02;8(2):189-200. [doi: 10.1079/phn2004673] [Medline: 15877912]
33. Reynolds KD, Yaroch AL, Franklin FA, Maloy J. Testing mediating variables in a school-based nutrition intervention program. Health Psychol 2002;21(1):51-60. [doi: 10.1037//0278-6133.21.1.51]
34. Youth risk behavior surveillance system (YRBSS). Centers for Disease Control and Prevention. URL: https://www.cdc.gov/healthyyouth/data/yrbs/index.htm [accessed 2023-05-26]
35. Neuhouser ML, Lilley S, Lund A, Johnson DB. Development and validation of a beverage and snack questionnaire for use in evaluation of school nutrition policies. J Am Diet Assoc 2009 Sep;109(9):1587-1592. [doi: 10.1016/j.jada.2009.06.365] [Medline: 19699839]
36. NHANES anthropometry procedures manual. Centers for Disease Control and Prevention. 2021 May. URL: https://wwwn.cdc.gov/nchs/data/nhanes/2021-2023/manuals/2021-Anthropometry-Procedures-Manual-508.pdf [accessed 2023-12-12]
37. Child and teen BMI calculator. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/bmi/child-teen-calculator/index.html [accessed 2023-05-26]
38. Neal JW, Neal ZP, VanDyke E, Kornbluh M. Expediting the analysis of qualitative data in evaluation: a procedure for the rapid identification of themes from audio recordings (RITA). Am J Eval 2014 May 27;36(1):118-132 [FREE Full text] [doi: 10.1177/1098214014536601]
39. Birks M, Chapman Y, Francis K. Memoing in qualitative research: probing data and processes. J Res Nurs 2008 Jan 01;13(1):68-75 [FREE Full text] [doi: 10.1177/1744987107081254]
40. SPSS statistics_29.0.x. IBM. URL: https://www.ibm.com/support/pages/ibm-spss-statistics290x [accessed 2023-05-31]
41. Nascimento GB, de Maio Nascimento M, de Araújo LM, Gouveia É, Ihle A. Comics as a physical education tool for health promotion in Brazilian primary education, based on Paulo Freire's principles of empowerment. Children (Basel) 2023 Sep 19;10(9):1575 [FREE Full text] [doi: 10.3390/children10091575] [Medline: 37761536]
42. Leung MM, Green MC, Cai J, Gaba A, Tate DF, Ammerman A. Fight for your right to fruit: development of a manga comic promoting fruit consumption in youth. Open Nutr J 2015 Feb;9(1):82-90. [doi: 10.2174/1876396001509010082]
43. Steinke J, Applegate B, Lapinski M, Ryan L, Long M. Gender differences in adolescents' wishful identification with scientist characters on television. Sci Commun 2011 Jun 28;34(2):163-199. [doi: 10.1177/1075547011410250]

XSL•FO
RenderX

44. Georgeson AR, Highlander A, Loiselle R, Zachary C, Jones DJ. Engagement in technology-enhanced interventions for children and adolescents: current status and recommendations for moving forward. Clin Psychol Rev 2020 Jun;78:101858 [FREE Full text] [doi: 10.1016/j.cpr.2020.101858] [Medline: 32413679]
45. Wrobel J, Silvasstar J, Peterson R, Sumbundu K, Kelley A, Stephens D, et al. Text messaging intervention for mental wellness in American Indian and Alaska native teens and young adults (BRAVE study): analysis of user engagement patterns. JMIR Form Res 2022 Feb 25;6(2):e32138 [FREE Full text] [doi: 10.2196/32138] [Medline: 35212633]
46. Thornton L, Brown HM, Osman B, Stewart C, Whife J, Champion KE, et al. Factors associated with adolescents' engagement with a Healthy Lifestyles app. Procedia Comput Sci 2022;206:56-67. [doi: 10.1016/j.procs.2022.09.085]
47. Schwarz A, Cardon G, Chastin S, Stragier J, De Marez L, Consortium SmartLife, et al. Does dynamic tailoring of a narrative-driven exergame result in higher user engagement among adolescents? Results from a cluster-randomized controlled trial. Int J Environ Res Public Health 2021 Jul 12;18(14):7444 [FREE Full text] [doi: 10.3390/ijerph18147444] [Medline: 34299895]
48. Thies K, Anderson D, Cramer B. Lack of adoption of a mobile app to support patient self-management of diabetes and hypertension in a federally qualified health center: interview analysis of staff and patients in a failed randomized trial. JMIR Hum Factors 2017 Oct 03;4(4):e24 [FREE Full text] [doi: 10.2196/humanfactors.7709] [Medline: 28974481]
49. Youth-centred digital health interventions: a framework for planning, developing and implementing solutions with and for young people. World Health Organization. 2021 Jul 19. URL: https://www.who.int/publications/i/item/9789240011717 [accessed 2024-12-20]
50. Jake-Schoffman DE, Silfee VJ, Waring ME, Boudreaux ED, Sadasivam RS, Mullen SP, et al. Methods for evaluating the content, usability, and efficacy of commercial mobile health apps. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e190 [FREE Full text] [doi: 10.2196/mhealth.8758] [Medline: 29254914]
51. Stinson J, McGrath P, Hodnett E, Feldman B, Duffy C, Huber A, et al. Usability testing of an online self-management program for adolescents with juvenile idiopathic arthritis. J Med Internet Res 2010 Jul 29;12(3):e30 [FREE Full text] [doi: 10.2196/jmir.1349] [Medline: 20675293]
52. Skouge JR, Rao K, Boisvert PC. Promoting early literacy for diverse learners using audio and video technology. Early Childhood Educ J 2007 May 22;35(1):5-11. [doi: 10.1007/s10643-007-0170-y]
53. Murnane R, Sawhill I, Snow C. Literacy challenges for the twenty-first century: introducing the issue. Future Child 2012 Sep;22(2):3-15. [doi: 10.1353/foc.2012.0013] [Medline: 23057128]
54. Racial/ethnic score gap tool. The Nation's Report Card. URL: https://www.nationsreportcard.gov/dashboards/regression/ [accessed 2023-11-21]
55. Reardon SF, Valentino RA, Shores KA. Patterns of literacy among U.S. students. Future Child 2012 Sep;22(2):17-37. [doi: 10.1353/foc.2012.0015] [Medline: 23057129]
56. Issaev S, Zahiri M, Cao M, LuL. Applying audio cues to increase manga immersion. The University of British Columbia. 2019. URL: https://courses.ece.ubc.ca/518/previous/hit2019F/papers/IssaevZahiriCaoLu2-2019.pdf [accessed 2024-12-20]
57. Green MC. Narratives and cancer communication. J Commun 2006 Aug 04;56(suppl_1):S163-S183. [doi: 10.1111/j.1460-2466.2006.00288.x]
58. Kreuter MW, Green MC, Cappella JN, Slater MD, Wise ME, Storey D, et al. Narrative communication in cancer prevention and control: a framework to guide research and application. Ann Behav Med 2007 Jun;33(3):221-235. [doi: 10.1007/BF02879904] [Medline: 17600449]
59. Family weight management: changing behaviour techniques. United Kingdom government. 2020 Mar 20. URL: https://www.gov.uk/government/publications/family-weight-management-changing-behaviour-techniques [accessed 2023-05-29]
60. Swindle T, Poosala AB, Zeng N, Børsheim E, Andres A, Bellows LL. Digital intervention strategies for increasing physical activity among preschoolers: systematic review. J Med Internet Res 2022 Jan 11;24(1):e28230 [FREE Full text] [doi: 10.2196/28230] [Medline: 35014962]
61. Wang Y, Min J, Khuri J, Xue H, Xie B, A Kaminsky L, et al. Effectiveness of mobile health interventions on diabetes and obesity treatment and management: systematic review of systematic reviews. JMIR Mhealth Uhealth 2020 Apr 28;8(4):e15400 [FREE Full text] [doi: 10.2196/15400] [Medline: 32343253]
62. Hampl SE, Hassink SG, Skinner AC, Armstrong SC, Barlow SE, Bolling CF, et al. Clinical practice guideline for the evaluation and treatment of children and adolescents with obesity. Pediatrics 2023 Feb 01;151(2):e2022060640. [doi: 10.1542/peds.2022-060640] [Medline: 36622115]
63. Hargreaves D, Mates E, Menon P, Alderman H, Devakumar D, Fawzi W, et al. Strategies and interventions for healthy adolescent growth, nutrition, and development. Lancet 2022 Jan 08;399(10320):198-210. [doi: 10.1016/S0140-6736(21)01593-2] [Medline: 34856192]
64. Branscum P, Sharma M, Wang LL, Wilson BR, Rojas-Guyler L. A true challenge for any superhero: an evaluation of a comic book obesity prevention program. Fam Community Health 2013;36(1):63-76. [doi: 10.1097/FCH.0b013e31826d7607] [Medline: 23168347]
65. Leung MM, Tripicchio G, Agaronov A, Hou N. Manga comic influences snack selection in Black and Hispanic New York City youth. J Nutr Educ Behav 2014;46(2):142-147. [doi: 10.1016/j.jneb.2013.11.004] [Medline: 24433817]

XSL•FO
RenderX

66. Hoffner CA. Same-gender characters: appeal and identification. In: The International Encyclopedia of Media Psychology. Hoboken, NJ: John Wiley & Sons; Sep 08, 2020.
67. Madeley JM. Girly girls and pretty boys: gender and audience reception of English-translated manga. Faculty of Arts, University of Regina. 2010 Dec. URL: https://ourspace.uregina.ca/items/be77fbc3-4335-4e0e-aaac-44637e9065f4 [accessed 2024-12-20]
68. Jones A, Armstrong B, Weaver RG, Parker H, von Klinggraeff L, Beets MW. Identifying effective intervention strategies to reduce children's screen time: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2021 Sep 16;18(1):126-146 [FREE Full text] [doi: 10.1186/s12966-021-01189-6] [Medline: 34530867]

Abbreviations

BCT: behavior change technique
BI: behavioral intention
CONSORT: Consolidated Standards of Reporting Trials
INC: Interactive Nutrition Comics for Urban, Minority Preadolescents
NYC: New York City
OE: outcome expectation
RCT: randomized controlled trial
T1: baseline
T2: intervention midpoint or 3 weeks after baseline
T3: intervention end or 6 weeks after baseline
T4: 3-month follow-up

Edited by A Mavragani; submitted 17.03.24; peer-reviewed by T Baranowski; comments to author 25.06.24; revised version received 18.10.24; accepted 04.11.24; published 15.01.25.

Please cite as:
Leung MM, Mateo KF, Dublin M, Harrison L, Verdaguer S, Wyka K
Testing a Web-Based Interactive Comic Tool to Decrease Obesity Risk Among Racial and Ethnic Minority Preadolescents: Randomized Controlled Trial
JMIR Form Res 2025;9:e58460
URL: https://formative.jmir.org/2025/1/e58460
doi:10.2196/58460
PMID:39813083

©May May Leung, Katrina F Mateo, Marlo Dublin, Laura Harrison, Sandra Verdaguer, Katarzyna Wyka. Originally published in JMIR Formative Research (https://formative.jmir.org), 15.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring Web-Based Support for Suicidal Ideation in the Scottish Population: Usability Study

Heather McClelland[1], BSc, MRes, PhD; Rory C O'Connor[1], BSc, PhD; Laura Gibson[2]; Donald J MacIntyre[3], MD

[1]School of Health and Wellbeing, University of Glasgow, Glasgow, United Kingdom

[2]Police Scotland, Glasgow, United Kingdom

[3]NHS Lothian, Edinburgh, United Kingdom

Corresponding Author:
Heather McClelland, BSc, MRes, PhD
School of Health and Wellbeing
University of Glasgow
90 Byres Road, Clarice Pears Building
Glasgow, G12 8TB
United Kingdom
Phone: 44 141 330 3299
Email: heather.mcclelland@glasgow.ac.uk

Abstract

Background: Suicide is a global health concern. In the United Kingdom, Scotland has the highest suicide rate. Lived experience and suicide prevention stakeholders in Scotland have identified a key gap in suicide prevention activities: the lack of 24-hour peer-driven web-based support for people who are suicidal.

Objective: This usability study aimed to evaluate the feasibility, acceptability, utility, and reach of a suicide prevention website (Surviving Suicidal Thoughts) specifically designed to support residents in Scotland who are experiencing suicidal thoughts themselves or suspect or know someone who is experiencing suicidal thoughts. Intended support was delivered through the provision of personal testimony videos of individuals with lived experience.

Methods: A peer-driven website was developed specifically to support residents of Scotland experiencing suicidal thoughts. The website included resources (eg, videos from lived experience and written guidance about how to respond to someone who may be experiencing suicidal thoughts) to help reduce distress, normalize experiences, and challenge distressing thoughts. The website was promoted via leading web-based social media channels and Google Ads. Evaluation of the website was based on website engagement, marketing strategy, and direct web user feedback via a cross-sectional survey.

Results: Data were collected for 41 weeks (June 2022 to February 2023) spanning the launch of the website and the conclusion of the second marketing campaign. On average, the website received 99.9 visitors per day. A total of 56% (n=14,439) of visitors were female, ages ranged from younger than 18 years to older than 70 years (commonly between 25 and 34 years) and originated from all regions of Scotland. According to Google Search terms of Scottish residents, of the individuals indicated to be experiencing suicidal thoughts but not looking for help, 5.3% (n=920) engaged with the website compared to 10.5% (n=2898) who were indicated to be looking for help for themselves. Based on participant responses to the evaluation survey (n=101), the website was associated with a significant reduction in suicidal thoughts (P=.03). Reasons for visiting the website varied. Marketing data implied that people were more likely to engage with advertisements, which they felt were more personal, and visitors to the website were more likely to engage with videos, which corresponded to their age.

Conclusions: A peer-led website may help residents of Scotland who are experiencing suicidal thoughts. Web-based interventions may have considerable reach in Scotland both in terms of age and geographic area. Engagement with the website was similar to other self-help websites for suicidal ideation; however, more nuanced methods of analyzing website engagement for help-seeking behavior are recommended. Future work would benefit from exploring the effectiveness of this website based on a larger participant sample with website modifications guided by the principles of social learning theory.

(JMIR Form Res 2025;9:e55932) doi:10.2196/55932

KEYWORDS

suicide prevention; Scotland; suicidal thoughts; digital intervention; internet; self-help; crisis intervention

XSL•FO
RenderX

Introduction

Background

Suicide is a leading global health concern. The World Health Organization [1] estimates that 1 person dies by suicide every 40 seconds. In the United Kingdom, Scotland continues to have the highest suicide rate across its 4 nations [2-4]. Most people with thoughts of suicide do not seek professional help, and for those who do, too often they struggle to access timely care [5]. Therefore, improving timely and easy-to-access care for those experiencing suicidal ideation is a public health priority in Scotland.

In 2022, the Scottish government published a 10-year, suicide prevention strategy that aims to reduce suicide on a national level [6]. The strategy drew on evidence from over 200 Scottish suicide prevention stakeholders (eg, lived experience, mental health professionals, and charity organization professionals) [7], which indicated that digital peer support was among the most sought-after resource needed. Subsequently, the current strategy includes the following goal: "anyone in suicidal crisis should be able to access support remotely by phone, text or online" [6,7].

National strategies for suicide prevention are underpinned by the drive to understand human behavior within a psychosocial context [8]. The integrated motivational-volitional model of suicidal behavior [9,10] is a leading theory that maps the emergence of suicidal thoughts and behaviors. Specifically, it posits that feelings of defeat (in response to stress) can develop into entrapment, which, in the absence of factors such as social support and belongingness, can give rise to suicidal ideation and potentially suicidal behaviors.

However, other theories of human behavior can also be applied to understand suicide risk. Social learning theory [11], for example, posits that people are more likely to model their behavior on those with whom they identify [12]. Indeed, videos of individuals describing their own experience of, and surviving, thoughts of suicide have been shown to be effective in improving the life satisfaction of those who are experiencing suicidal thoughts themselves [13]. What is more, according to group identity theory [14], identifying with others with shared thoughts or experiences can facilitate feelings of belongingness, the absence of which is consistently associated with suicide risk [13,14].

Systematic reviews by both Lai et al [15] and Robinson et al [16] indicate that a range of web-based suicide prevention strategies (including self-guided, digital cognitive behavioral therapy, web chat, and signposting to resources) are effective in reducing thoughts of suicide, including among those in a suicidal crisis. Although the number of studies included in these reviews is small, the evidence also points toward the potential utility of testimonies from others with similar experiences as a helpful suicide prevention strategy. In the United States, NowMattersNow [17]—a website designed specifically for US residents—provides evidence-based resources including videos, blogs, and suicide prevention strategies (eg, emotion self-regulation coping strategies) to promote the self-management of suicidal thoughts. The website is available 24 hours per day and is freely accessible to the public. Visitors to this website were found to stay on the website for an average of 1 minute 31 seconds, with a bounce rate (percentage of visitors who left after viewing 1 page) of 61% [18]. To put this in context, a bounce rate of 60%-90% is average for landing pages. Based on an embedded anonymous web user survey, Whiteside et al [18] found that engagement with the NowMattersNow website was associated with a self-reported reduction in feelings of suicidal ideation and distress. Given the cultural similarities between the United States and Scotland and the perceived demand of lived experience support from people who are suicidal [6,19], such a resource may be effective within the Scottish context.

Aims

This pilot phase usability study is part of a wider, ongoing digital project to reduce suicide across Scotland. The website was co-designed and delivered by individuals with lived experiences of suicidal thoughts and behaviors. The overall aims of the project are listed in Multimedia Appendix 1. To help guide these aims, theory-driven outcomes for this study were determined in advance of website development using a logic model (Multimedia Appendix 2). Given the early stage of the development of this suicide prevention website, this study aimed to address the following short-term intended outcomes of the logic model. Specifically, the development of a suicide prevention website aimed to (1) support the Scottish population when experiencing a suicidal crisis that is codeveloped with the target population, (2) investigate which users of the website feel that they have received a compassionate response, (3) investigate which users experience a reduction in feelings of entrapment during engagement with the website, (4) investigate which users experience a reduction in suicidal thoughts during engagement with the website, and (5) investigate which users of the website find it accessible and relatable.

Methods

Study Design

The website, Surviving Suicidal Thoughts (SST) [20], was developed using an iterative approach involving a multidisciplinary website development team (comprised of digital experts, health professionals, and academic researchers) and direct feedback from the Scottish government's Living and Lived Experience Panel (n=5; Scottish residents with lived experience of suicidal thoughts and behaviors).

Ethical Considerations

Research permission was granted by the local National Health Service (NHS) Research and Development Committee in 2022. A Data Protection Impact Assessment was completed and approved prior to the launch of the web user survey. However, following NHS guidance, ethics approval is not required for service development and evaluation studies such as this (section 3.1 of the UK Policy Framework for Health and Social Care Research) [21]. Informed consent to participating in the anonymized feedback survey was inferred through participants opting in to the survey following the survey invitation appearing

XSL•FO
RenderX

on the web page, and a privacy notice was available on the invitation page of the survey. Anonymous web user metrics and engagement with the website were collected only from users who had consented to allow tracking cookies on their internet-enabled device prior to "landing" on any NHS24 web page. Due to the anonymity of participation, participants could not be compensated.

Data Collection

Overview

Data collection for this pilot phase spanned 41 weeks (see Multimedia Appendix 3 for the timeline). Originally, the timeline was expected to span 6 months; however, due to low participant engagement with the web user survey, data collection was extended to include the second marketing campaign spanning from October 2022 to February 2023. Visitors' engagement with the website was monitored during development with adjustments made based on user behavior (Multimedia Appendix 3). Such behavior was monitored before and during visiting the website, which allowed SST visitors to be categorized into 1 of 5 user groups (defined in Table 1). The way in which the 5 visitor groups were expected to navigate to, and use, the SST website is depicted in Multimedia Appendix 4. Further details of the development and evaluation of the website are summarized under the following subheadings: "Website Content," "Web User Evaluation Survey," and "Marketing."

Table 1. Surviving Suicidal Thoughts website visitor groups[a].

User type	Behavior	Search term examples
1. Web-based suicide methods seeking	Individuals not looking for support for their suicidal thoughts	"Want to die" "End my life"
2. Web-based support seeker	Individuals actively looking for support for their suicidal thoughts	"Suicide support"
3. Told friends or family	Individuals who are looking for support for a friend or relative who they know is experiencing suicidal thoughts	"Suicidal relative," "Suicidal friend"
4. Suspecting friends or family	Individuals who are looking for support for a friend or relative who they suspect may be experiencing suicidal thoughts; however, this has not been confirmed	"Is my friend suicidal?"
5. Generic curious or professional	Individuals not help-seeking for themselves or others and not in suicidal crisis	"Suicide prevention day"

[a]The 5 user groups identified to visit the Surviving Suicidal Thoughts website between June 2022 and February 2023 based on search terms entered into Google by Scottish residents.

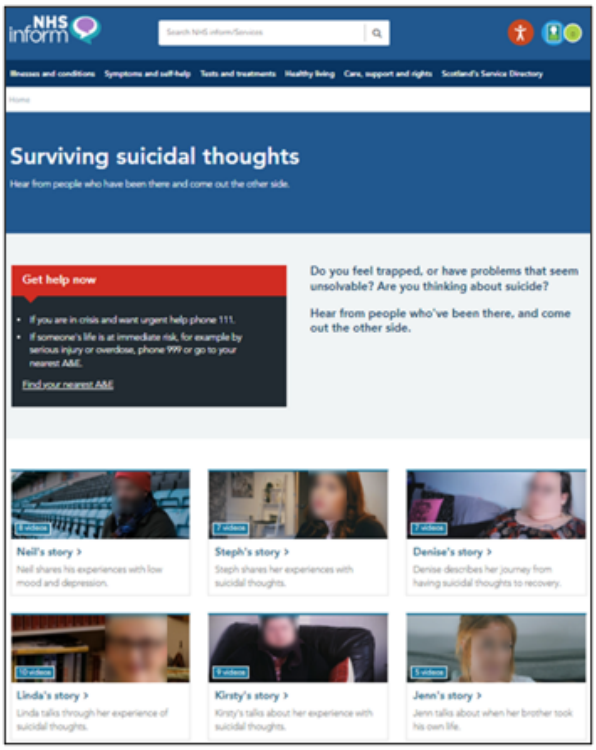
Website Content

SST was hosted on the NHS Inform internet platform, Scotland's national health care website (Figure 1). The website included personal testimony videos of 5 people (Neil, Steph, Denise, Linda, and Kirsty) with lived experience of suicide and 1 person bereaved by suicide (Jenn). These videos were developed specifically for the website. In these videos, the speakers shared their broad experiences of low mood and suicidal thoughts and behaviors, rather than discussing specific details related to their exact experiences of suicidal crisis. Each video was cut into between 5 and 10 chapters, with each chapter being approximately 90 seconds each. The chapters were titled, which allowed visitors to jump between individual sections. Early in the first marketing campaign, it was observed that visitors of the website only explored the top 2 videos that appeared on their computer or phone screen. In response, the videos were rotated each month to minimize any order effects.

The website also featured emergency service contact information, signposting to third-sector mental health organizations, and how to respond to someone who discloses or may be experiencing suicidal thoughts. To directly evaluate the experiences of web users, the website included a pop-up survey modeled on NowMattersNow [17]. Engagement with the website was captured by the intervention delivery team at NHS Inform using Google Analytics [22].

XSL•FO
RenderX

Figure 1. Image of the Surviving Suicidal Thoughts landing page.



Web User Evaluation Survey

The SST survey was developed using the HotJar survey platform [23]. Initial survey content and methodology were collated by HM (study author) and presented to a focus group of individuals with lived experience of suicidal thoughts and behaviors for their feedback (see Multimedia Appendix 5 for focus group feedback). To minimize overburdening SST website visitors, the Lived Experience Panel recommended reducing the survey items from 12 to 4 questions. As a result, it was intended that a series of "blocks" of survey items would be displayed during the pilot phase to capture information on a range of demographic and psychological factors over time. However, due to the limited engagement with the survey (see Overview in the Results section), it was decided that only the first block of questions would be used in the pilot phase to improve statistical power for analysis. The final survey items and measurement scales included in the pilot are presented in Multimedia Appendix 6; however, in brief: 2 items assessed participants' intensity of suicidal ideation (taken from NowMattersNow [17]; range: 1=not at all and 5=completely overwhelming), 1 demographic variable explored participants' age group (10-year increments from "younger than 18 years" to "70+ years"), and 1 item captured geographic location (based on NHS health board). A fifth item (reason for visiting the website) was added midway through survey recruitment. All survey items were asked at a single time point. The suicidal ideation items asked web users to recall the level of intensity of their suicidal thoughts at the time of first landing on the SST website as well as the current intensity of their suicidal ideation at the time of participating in the survey. These survey items were anonymous, and participants could discontinue the survey and return to the SST website at any time.

To be consistent with the format of NowMattersNow [17], the SST survey invitation initially appeared 3 minutes after landing on the SST website. However, after 2 days, it became evident that the average web user duration was less than 3 minutes (with only 4 visitors seeing the invitation); therefore, we adjusted the survey invitation to appear after 90 seconds (Multimedia Appendix 3) to match 50% of user traffic.

Marketing

Digital marketing was used to promote the website resources on social media platforms (ie, Meta [Facebook and Instagram], X [formerly known as Twitter], and Spotify). Live targeting platforms (Nano and Google Ads) were used to advertise the website to relevant web users. Marketing data were collected from social media platforms, from Nano and Spotify by a third party (Republic of Media) using a pay-per-click marketing campaign. Advertisements promoting the website were designed to convey empathy and hope. Following some prosuicide comments from the public, advertisements for the SST website on Meta platforms were discontinued 1 week after launch to protect the public. This funding was reallocated to Google Ads. Additionally, advertising on X was discontinued after 23 weeks (November 27, 2022) of this 41-week pilot phase. This was due to changes in X's ownership and uncertainty around its operating procedures. Thereafter, funding for the promotion of the SST website on X was reallocated to Nano.

Data Analysis

Website Content Engagement

Analysis was modeled on those used by Whiteside et al [18]. Website engagement was inferred using bounce rates (the percentage of web visitors who leave or "bounce" from a web

XSL•FO
RenderX

page without taking follow-up action). Web users' device used, date, time, and journey to the website were also captured. Video engagement (which videos and duration) was provided by Nano.

Web User Evaluation Survey

Data relating to the date, time, and type of internet-enabled device (eg, laptop and mobile phone) were summarized using frequencies and percentages collated by HotJar survey systems. Using G*Power [24], a priori power analysis determined that a minimum of 102 participants were needed for pairwise comparisons between baseline and follow-up data (effect size=0.03; α=.05; power=0.9). All statistical analyses were conducted using SPSS (version 28; IBM Corp). The distribution of the data was inspected visually and then tested using Kolmogorov-Smirnov tests, which indicated that the data were not normally distributed (P<0.001). Subsequently, a nonparametric approach was deemed appropriate to analyze the survey responses, thereby avoiding a type I error. Subsequently, a Wilcoxon 1-sampled, 2-tailed t test was used to statistically test the change in participants' intensity of suicidal ideation. A statistical subanalysis of other outcomes (eg, age and geographic location) was not possible due to the small group sizes. Participants who did not answer both suicidal ideation questions were not included in the statistical analysis; all remaining missing data were grouped with "prefer not to say."

Marketing

Information captured included engagement with advertisements (click-through rates [CTRs]) and videos (view-through rates) as well as demographic information (sex: male, female, or unknown and age group: 18-24, 25-34, 35-44, 45-54, 55-64, 65+ years, "unknown"). These data were reported in relation to advertising trends commonly observed in similar advertising strategies [25,26]. Additionally, Nano provided geographical information regarding advertising exposure and engagement based on the number of people who saw an advertisement (impressions).

Results

Overview

SST [20] was developed using an iterative approach between the web page development team and Scotland's Living and Lived Experience Panel and remained live throughout the 41-week pilot phase (from June 28, 2022, to February 20, 2023). The SST website received 25,784 unique visitors equating to an average of 99.9 page views per 24-hour day. Daily traffic to the website peaked at 10 PM, with the greatest amount of traffic on Sundays.

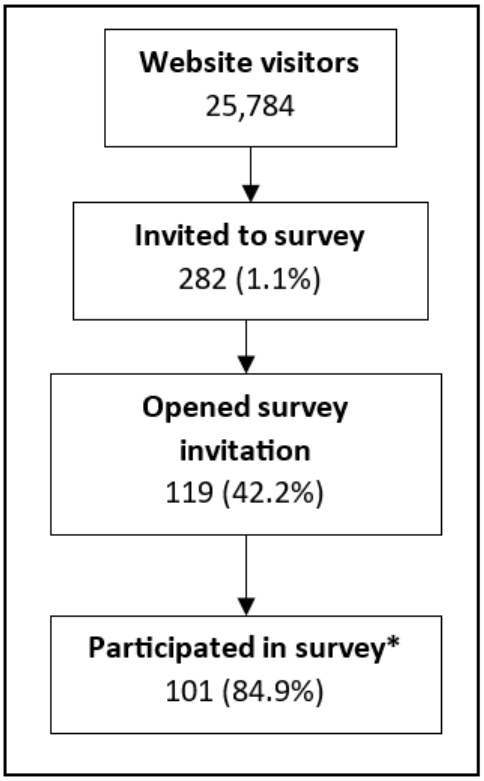
Most visitors accessed the website via mobile phone (n=21,402, 83%), followed by desktop computer (n=3068, 11.9%). When compared to commercial metrics [22], social media marketing CTRs for SST were considered "good" on all platforms. For videos, the view-through rate was 80.2% (n=892,943), which was considered "high." Journeys of SST users were most commonly via Google Search (n=6314, 24.3%), followed by Meta (n=4705, 17.1%). For more details of marketing expenses, marketing strategy, and user engagement per platform, please see Multimedia Appendix 7.

More "web-based suicide methods seeking" than "web-based support seeker" user journey members clicked on testimonial advertisements (advertisements that included names) or engaged with advertisements titled "hear from one whose been there." However, sample sizes were too small to explore this statistically.

The bounce rate to the website was 76.9%, suggesting that almost a quarter of website visitors clicked on a subpage of the SST website (eg, a video link). Of the web users who were invited to participate in the evaluation (n=282, 1.1%), 42.2% (n=119) opened the survey link. In total, 84.9% of those who opened the survey answered at least 1 question (Figure 2).

XSL•FO
RenderX

Figure 2. Participant flowchart of participant engagement with the Surviving Suicidal Thoughts website and evaluation survey between July 2022 and February 2023. *Answered at least 1 survey question.



Demographic Characteristics

Google Analytics indicated that 56% (n=14,439) of SST visitors were female, and most "web-based nonsupport seeking" groups were also indicated to be female. Web users who accessed the website through Google Ads were most likely to be aged between 25 and 34 years. In contrast, half (n=47, 49.6%) of the respondents to the evaluation survey were between 40 and 59 years (Table 2). Most (n=69, 66.4%) survey participants chose not to indicate which health board they lived in.

XSL•FO
RenderX

Table 2. Demographic and geographic summary of participant characteristics of the survey evaluation.

Variable	Values, n (%)
Age (years)	
<18	6 (5)
19-29	11 (9.2)
30-39	11 (11.8)
40-49	14 (27.7)
50-59	33 (21.9)
60-69	26 (6.7)
70+	8 (0.8)
Prefer not to say or not applicable	7 (5.9)
Unanswered	13 (10.9)
Health board	
Ayrshire and Arran	5 (4.2)
Borders	1 (0.8)
Fife	2 (1.7)
Fourth Valley	2 (1.7)
Grampian	3 (2.5)
Greater Glasgow and Clyde	7 (5.9)
Highland	1 (0.8)
Lanarkshire	5 (4.2)
Lothian	9 (7.6)
Shetland	1 (0.8)
Tayside	3 (2.5)
Western Isles	1 (0.8)
Prefer not to say or not applicable	10 (8.4)
Unanswered	69 (58)
Reason for visit[a]	
I am looking for resources to help me cope if I feel suicidal again	9 (7.6)
I am thinking about suicide	14 (11.8)
I do not think I can keep myself safe	4 (3.4)
I have a plan for suicide	6 (5)
I am looking for resources to support someone else	8 (6.7)
Other or prefer not to say	11 (9.2)
Unanswered	67 (56.3)
Suicide intensity	
Intensity of suicidal thoughts upon first visiting the site (mean 3.06, SD 1.58)	
1=not at all	30 (25.2)
2	12 (10.1)
3	19 (16)
4	19 (16)
5=completely overwhelming	30 (25.2)
Unanswered	9 (7.6)
Intensity of suicidal thoughts at time of survey (“now”; mean 2.87, SD 1.59)	

XSL•FO
RenderX

Variable	Values, n (%)
1=not at all	33 (27.7)
2	11 (9.2)
3	17 (14.3)
4	16 (13.5)
5=completely overwhelming	24 (20.2)
Unanswered	18 (15.1)

[a]Item added 14 weeks into data collection phase.

Video Engagement

The first 2 videos that appear on a web visitor's browser commonly received the most visitor engagement. "Neil's story: Experiences with low mood and depression" and "Steph's story: Experiences with suicidal thoughts" were the most viewed chapters. On average, web users watched 86.64% (SD 5.91%) of the video chapters. Chapters most likely to be viewed in full were titled "The anxiety underneath was so much to bear" (Kirsty, chapter 3) and "There is light at the end of the tunnel" (Neil, chapter 4).

Suicidal Thoughts

Most (n=5290, 83.8%) CTRs via Google Ads were attributed to the "web-based support seeker group," "told friends or family," or "suspecting friends or family" SST visitor journey groups (Table 3). In contrast, 86.5% of CTRs from Google Search terms related to concern for others experiencing suicidal thoughts (50% CTR for "suicidal relative" and 36.5% CTR for "suicidal friend"), which represent the "told friends or family" and "suspecting friends or family" user journey groups (Table 1). However, 65.5% (n=67) of respondents to the evaluation survey did not indicate their reason for visiting the website (9.2% responded "other or prefer not to say" and 56.3% of participants did not answer).

Table 3. Total number of impressions and clicks to the Surviving Suicidal Thoughts website via adverts on social media and Google.

	Impressions, n	Clicks to website, n	CTR[a] (%)
SST user journey group			
Web-based suicide methods seeking[b]	17,282	920	5.32
Web-based support seeker[c]	27,638	2898	10.49
Told friends or family[d]	14,565	2172	14.91
Suspecting friends or family[e]	1730	220	12.72
Generic curious or professional[f]	1496	104	6.95
Media platform			
X	1,209,432	12,847[g]	1.06
Nano	1,124,972	3471	0.31
Spotify	326,359	—[h]	0.26
Facebook	202,326	4705	2.33
Total	2,925,800	22,637	N/A[i]

[a]CTR: click-through rate.

[b]Individuals not looking for support for their suicidal thoughts.

[c]Individuals actively looking for support to relieve their suicidal thoughts.

[d]Friend or relative who they know to be experiencing suicidal thoughts.

[e]Friend or relative who they suspect may be experiencing suicidal thoughts.

[f]Those looking for general information on suicide prevention who does not otherwise fit in the other web user groups.

[g]Based on engagement with social media post, not clicks to website.

[h]Data not available as these were audio advertisements; therefore, clicking to the website was not possible.

[i]N/A: not applicable.

Based on survey responses, 25.2% (n=30) of visitors retrospectively described the intensity of their suicidal thoughts when first visiting the website as "completely overwhelming" compared to 20.1% (n=24) of participants at the time of survey

XSL•FO
RenderX

engagement (Table 2). Overall, there was a significant reduction in the intensity of suicidal thoughts from the recollection of intensity when first landing on the web page to participating in the survey (t_{100}=–1.89; P=.03). No statistically significant differences were observed between the intensity of suicidal thoughts as a function of age (≤49 years vs ≥50 years).

Discussion

This study aimed to evaluate a pilot web-based self-help resource to help support residents of Scotland who have suicidal ideation. The findings of this study are discussed below in relation to the 5 short-term intended outcomes of the logic model (Multimedia Appendix 2).

Principal Findings

The design and content of SST [20] was developed through consultation with the Living and Lived Experience Panel in Scotland. This input included direct contributions to the website videos—the defining feature of the SST website. Together with other stakeholders (digital specialists and health professionals), this led to a suicide prevention website launched in June 2022 specifically designed to support residents of Scotland. Therefore, outcome 1 (support the Scottish population when experiencing a suicidal crisis that is codeveloped with the target population) was achieved.

Due to low survey engagement, the survey blocks were not rotated as initially intended. Therefore, this pilot phase was unable to obtain visitor views on receiving a compassionate response from the SST website (eg, helpful videos and guidance). Despite this, amendments to the website and survey user evaluation (Multimedia Appendix 3) were conducted in response to user engagement. Thus, a compassionate approach was exhibited by the website developer team to meet the changing needs of the website visitors. Additionally, involvement from the Lived Experience Panel was intended to best serve the needs of the Scottish population. Their input included guidance on sensitive advertising of the website to best avoid triggering people who were not in distress. Additionally, on Meta, comments encouraging suicide prompted the team to discontinue the marketing campaign on this platform for the well-being of social media users. Subsequently, outcome 2 (users of the website feel that they have received a compassionate response) was only partially addressed.

Due to low survey engagement, questions about entrapment were excluded from the web user survey; therefore, outcome 3 (users experience a reduction in feelings of entrapment during engagement with the website) was not addressed. Despite this, the Lived Experience Panel expressed that both internal and external entrapment survey questions were acceptable items to be included within the SST website survey.

The survey data indicated that the intensity of suicidal thoughts had significantly reduced between the time participants first arrived on the SST website and the time they engaged with the survey. However, as the P value just met the traditional level of significance, this needs to be replicated in a large sample. Nonetheless, outcome 4 (users experience a reduction in suicidal thoughts during engagement with the website) was achieved.

Data from Google Analytics indicated that the website was accessed by all geographic regions of Scotland and people up to 70 years of age. This geographic spread of engagement was further reflected by data collected by Nano. Additionally, the website was used at all times of day and night and from a range of devices (eg, computers, phones, and tablets). These findings suggest that the website was both accessible and relatable to residents of Scotland; therefore, outcome 5 (users of the website find it accessible and relatable) was achieved.

Comparison With Prior Work

Consistent with Lai et al [15] and Robinson et al [16], engagement with the SST [20] seemed to significantly reduce the intensity of suicidal thoughts of visitors to the website. The findings of this pilot study are similar to those of the NowMattersNow evaluation [18]. Specifically, the prevalence of web users who described the intensity of their suicidal thoughts as “completely overwhelming” when first visiting the site was similar across the 2 websites (25% of SST web users compared to 28% of NowMattersNow) [18]. Equally, both websites’ bounce rate was within the expected range according to web-based marketing norms of 60%-90% [26] (NowMattersNow [17]: 61% and SST [20]: 76.91%). Both websites reported an absence of participants aged 70+ years. Conversely, Scotland’s SST website was indicated to engage younger people (<18 years: 10.15% of SST visitors), whereas this age group was not captured by NowMattersNow [17].

The pattern of engagement with the SST websites is consistent with leading theories of learned behavior such as social learning theory [11]. Social learning theory posits that people are more likely to model their own behavior on those with whom they identify. Williams et al [19], based on a cross-sectional sample of self-harm web-based communities, found that such groups are more likely to seek informal help from perceived peers than professional services, and advice on self-care was reportedly more effective when provided by these peer supports. Social identification was evident in this study, for example, where web users were more likely to engage with advertisements or SST videos, which corresponded with their own age or sex.

Most visitors were aged between 25 and 34 years, consistent with “Steph,” whose video was viewed the most compared to all other videos on the SST website. Additionally, most visitors accessing the SST website via “X” (formerly known as Twitter) were typically male participants and landed on the SST website via advertisements about “Neil’s story.” Similarly, the marketing company used in this study (Republic of Media) reported that, in general, female participants were more likely than male participants to engage with Google Ads. In this study, advertisements relating to “Linda’s story” were most likely to be engaged with on Google platforms. Although these findings are preliminary, these trends would suggest that it may be beneficial to consider principles of social learning theory [11] when developing suicide prevention websites.

This pilot phase usability study was developed to inform a larger study aimed to reduce suicide in Scotland via timely peer-led support. Subsequently, recommendations based on the findings here include (1) updating marketing strategies to meet the needs of those seeking and not seeking help for their suicidal thoughts,

XSL•FO
RenderX

(2) using pop-up invitations for the user evaluation to improve survey engagement, and (3) increasing the demographic diversity of populations represented in the survivor videos. As illustrated by the logic model (Multimedia Appendix 2), the long-term goals aim to integrate the website into the health care landscape, which clinical and nonclinical populations can use as an established resource to support well-being. The current evaluation also indicates that future work should consider the role of the learned behavior (eg, social learning theory) [11] in relation to engagement with suicide prevention resources. Additionally, to explore the extent to which web visitors felt that they received compassionate support, qualitative interviews with individuals who have used the website would be beneficial.

Limitations

The study findings need to be interpreted in the context of the limitations related to the content of the website, the web user survey, and how the website was marketed. First, most web users usually only engaged with the first 2 videos that appeared on their browser; however, this trend was only identified after the website was live. Although Neil's story and Steph's story were the most engaged with by web users, this may have reflected their position on the website, as opposed to the web users' video preference overall. Additionally, it is unknown whether engagement with the website provided a compassionate response to its visitors.

Second, due to constraints of the survey platform used in this intervention, not all preferred approaches could be implemented into the survey methods. Namely, the survey could only appear once to web users, thereby limiting the survey to a cross-sectional design, where the intensity of suicidal ideation was explored both retrospectively and currently. Additionally, some of the recommendations made by the Lived Experience Panel could not be incorporated into the survey (eg, having the survey invitation being a permanent feature on the screen so website visitors could participate in the survey whenever they felt ready). This may account for the low engagement with the SST survey. Additionally, although the study originally followed the Lived Experience Panel's recommendation of a 3-minute delay before the survey invitation pop-up, due to the average duration of visits to the SST website being shorter than anticipated, the invitation was reduced to 90 seconds, consistent with NowMattersNow [18]. Although survey participation rates were good, the number of people who saw the invitation was lower than expected, limited to those who stayed on the website long enough to see the invitation, and engaged by those who chose to participate in the survey. Subsequently, the findings are unlikely to be generalizable to the wider population of Scotland. The low number of survey invitations was likely due to web users having disabled cookies across the NHS platform. This lower recruitment subsequently led to the blocks of survey items not being rotated; therefore, additional information (eg, sex) not being captured in the data. The small participant sample also limited subgroup analyses to age comparisons only as well as preventing further survey items (eg, entrapment) from being substituted into the survey over time. Given that most participants did not answer the self-reported question relating to the health board, it is unclear whether the participant sample was reflective of the overall SST web user population. Additionally, the final survey item included responses for individuals who were help-seeking for themselves, while those who were seeking help for others were grouped together with those who responded with "prefer not to say." Therefore, it was not possible to make comparisons between reasons for visits in relation to demographic characteristics or suicidal distress. Finally, most respondents to the survey were aged between 40 and 59 years, whereas Google Analytics indicated that most web users were aged between 25 and 34 years. This indicates that pop-up surveys may not be the most effective way to capture web user satisfaction on self-help platforms such as SST and that the findings from the survey may not be generalizable to all visitors to the website.

Third, the marketing approach of the website must be considered. Word limits on Google Ads may have impacted on the success of engaging certain demographics, namely male participants. This is reflected in the alternative advertising platforms, such as Spotify, where proportionately, non–help-seeking male participants were more likely to engage with SST advertisements than non–help-seeking female participants. Challenges were encountered with advertising, with advertising being discontinued on some platforms. This highlights a challenge between balancing a soft-marketing strategy against the need for ethical considerations such as safeguarding to ensure public well-being. Solutions to achieving this balance may include the use of trigger warnings and further developments in social media algorithms.

Finally, the bounce rate may not have provided a true representation of the SST websites' utility by the user. Specifically, the bounce rate of the SST was higher than NowMattersNow [18], reflective of web users leaving the website after arriving on the landing page. However, it is unclear whether the bounce rate was due to visitors benefiting from the website approach to making information salient (eg, crisis card information on the landing page) or because the website did not meet the individual's expectations during their visit, thereby prompting them to leave. Therefore, the bounce rate (ie, engagement) may not necessarily be reflective of the website failing to meet the need, but that the landing page provided sufficient information for the web user.

Conclusions

A codeveloped website for suicide prevention, designed for the Scottish population, may help to reduce the intensity of one's suicidal thoughts. Marketing of the website across various social media, entertainment, and internet search platforms is feasible. The website was particularly popular with those who were looking for help for themselves, and principles of social learning theory may play a role in optimizing the utility of such resources. A larger study is required to explore the effectiveness of such a website and determine whether it works for different subgroups of the Scottish population.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank the Living and Lived Experience Panel of the National Suicide Prevention Leadership Group and members of the NHS24 digital team (Mr Gary Black, Mr Rory Bone, Mr Chris Begg, Ms Lucy Hartely, Ms Gail Lumsden, Ms Carole McCamberly, and Ms Lindsey Thomson) for their expertise in the development of the Surviving Suicidal Thoughts website. No generative artificial intelligence tools (eg, ChatGPT) were used at any stage of this project. This project was funded by the Scottish government.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

DJM is employed by NHS 24 which receives funding for Surviving Suicidal Thoughts from the Scottish Government.

Multimedia Appendix 1
Project aims of the Surviving Suicidal Thoughts website.
[DOCX File , 17 KB - formative_v9i1e55932_app1.docx]

Multimedia Appendix 2
Logic model summarizing the anticipated inputs, outputs, and outcomes of the Surviving Suicidal Thoughts website.
[DOCX File , 19 KB - formative_v9i1e55932_app2.docx]

Multimedia Appendix 3
Timeline illustrating developmental milestones of the Surviving Suicidal Thoughts website between 2020 and 2023.
[DOCX File , 18 KB - formative_v9i1e55932_app3.docx]

Multimedia Appendix 4
The 5 user journeys anticipated visitors to the Surviving Suicidal Thoughts (SST) website will take from using a search engine to landing on the SST website.
[DOCX File , 714 KB - formative_v9i1e55932_app4.docx]

Multimedia Appendix 5
Collated feedback form a Lived Experience Panel in February 2022 regarding the potential survey items to be included in the Surviving Suicidal Thoughts website.
[DOCX File , 16 KB - formative_v9i1e55932_app5.docx]

Multimedia Appendix 6
Survey items included in the Surviving Suicidal Thought website cross-sectional user evaluation survey from July 2022 to February 2023.
[DOCX File , 16 KB - formative_v9i1e55932_app6.docx]

Multimedia Appendix 7
Summary of social media costs for the Surviving Suicidal Thoughts website first marketing campaign from July to September 2022.
[DOCX File , 16 KB - formative_v9i1e55932_app7.docx]

References

1. World Health Organization. Suicide prevention. 2024. URL: https://www.who.int/health-topics/suicide#tab=tab_1 [accessed 2024-11-19]
2. Probable suicides. National Records of Scotland. 2022. URL: https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/deaths/suicides#:~:text=Probable%20Suicides.%20Information%20on%20the [accessed 2024-11-19]
3. Review of suicide statistics in Northern Ireland 2015 - 2020. Northern Ireland Statistcs and Research Agency. 2022. URL: https://www.nisra.gov.uk/system/files/statistics/Suicide_Review_Report.pdf [accessed 2024-11-19]

XSL•FO
RenderX

4. Suicides in England and Wales: 2021 registrations. Office for National Statistics. 2021. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/suicidesintheunitedkingdom/2021registrations [accessed 2024-11-19]
5. Michelmore L, Hindley P. Help-seeking for suicidal thoughts and self-harm in young people: a systematic review. Suicide Life Threat Behav 2012;42(5):507-524. [doi: 10.1111/j.1943-278X.2012.00108.x] [Medline: 22889130]
6. Creating hope together: suicide prevention strategy 2022 to 2032. Scottish Government. 2022. URL: https://www.gov.scot/publications/creating-hope-together-scotlands-suicide-prevention-strategy-2022-2032/ [accessed 2024-11-19]
7. Time, space, compassion three simple words, one big difference: recommendations for improvements in suicidal crisis response. Scottish Government. 2021. URL: https://www.gov.scot/publications/time-space-compassion-three-simple-words-one-big-difference-recommendations-improvements-suicidal-crisis-response/ [accessed 2024-11-19]
8. Mann JJ, Apter A, Bertolote J, Beautrais A, Currier D, Haas A, et al. Suicide prevention strategies: a systematic review. JAMA 2005;294(16):2064-2074. [doi: 10.1001/jama.294.16.2064] [Medline: 16249421]
9. O'Connor RC, Kirtley OJ. The integrated motivational-volitional model of suicidal behaviour. Philos Trans R Soc Lond B Biol Sci 2018;373(1754):20170268 [FREE Full text] [doi: 10.1098/rstb.2017.0268] [Medline: 30012735]
10. O'connor RC. Towards an integrated motivational–volitional model of suicidal behaviour. In: O'connor RC, Platt S, Gordon J, editors. International Handbook of Suicide Prevention: Research, Policy and Practice. Chichester: John Wiley & Sons, Ltd; 2011.
11. Bandura A. Model of causality in social learning theory. In: Cognition and Psychotherapy. Englewood Cliffs, NJ: Prentice-Hall; 1977:81-99.
12. Coussi-Korbel S, Fragaszy DM. On the relation between social dynamics and social learning. Anim Behav 1995;50(6):1441-1453. [doi: 10.1016/0003-3472(95)80001-8]
13. Till B, Strauss M, Sonneck G, Niederkrotenthaler T. Determining the effects of films with suicidal content: a laboratory experiment. Br J Psychiatry 2015;207(1):72-78. [doi: 10.1192/bjp.bp.114.152827] [Medline: 25906792]
14. Tajfel H, Turner J. An integrative theory of intergroup conflict. In: The Social Psychology of Intergroup Relations. Monterey, CA: Brooks/Cole Publishing Company; 1979:56-65.
15. Lai MH, Maniam T, Chan LF, Ravindran AV. Caught in the web: a review of web-based suicide prevention. J Med Internet Res 2014;16(1):e30 [FREE Full text] [doi: 10.2196/jmir.2973] [Medline: 24472876]
16. Robinson J, Cox G, Bailey E, Hetrick S, Rodrigues M, Fisher S, et al. Social media and suicide prevention: a systematic review. Early Interv Psychiatry 2016;10(2):103-121. [doi: 10.1111/eip.12229] [Medline: 25702826]
17. Whiteside URJEM. NowMattersNow. 2014. URL: https://nowmattersnow.org/ [accessed 2024-10-04]
18. Whiteside U, Richards J, Huh D, Hidalgo R, Nordhauser R, Wong AJ, et al. Development and evaluation of a web-based resource for suicidal thoughts: NowMattersNow.org. J Med Internet Res 2019;21(5):e13183 [FREE Full text] [doi: 10.2196/13183] [Medline: 31045498]
19. Williams AJ, Nielsen E, Coulson NS. "They aren't all like that": perceptions of clinical services, as told by self-harm online communities. J Health Psychol 2020;25(13-14):2164-2177 [FREE Full text] [doi: 10.1177/1359105318788403] [Medline: 30024273]
20. MacIntyre DJG, McClelland H, O'Connor RC. Surviving suicidal thoughts. NHS Inform. 2022. URL: https://www.nhsinform.scot/surviving-suicidal-thoughts [accessed 2024-10-04]
21. UK policy framework for health and social care research. Health Research Authority. URL: https://s3.eu-west-2.amazonaws.com/www.hra.nhs.uk/media/documents/uk-policy-framework-health-social-care-research.pdf [accessed 2025-01-09]
22. Google Analytics: analytics tools & solutions for your business. Google Marketing Platform. 2023. URL: https://marketingplatform.google.com/about/analytics/ [accessed 2024-10-04]
23. What is Hotjar? HotJar. 2014. URL: https://www.hotjar.com/blog/what-is-hotjar/ [accessed 2024-10-04]
24. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
25. About demographic targeting. Google Ads Help. 2023. URL: https://support.google.com/google-ads/answer/2580383 [accessed 2024-09-28]
26. Bounce rates: what's good, what's bad, and why you should give a damn. CustomMedia. URL: https://www.customedialabs.com/blog/bounce-rates/ [accessed 2024-10-04]

Abbreviations

CTR: click-through rate
NHS: National Health Service
SST: Surviving Suicidal Thoughts

XSL•FO
RenderX

Edited by A Mavragani; submitted 02.02.24; peer-reviewed by IT Hwang, C Hughes , A AL-Asadi; comments to author 06.06.24; revised version received 30.07.24; accepted 13.10.24; published 24.01.25.

Please cite as:
McClelland H, O'Connor RC, Gibson L, MacIntyre DJ
Exploring Web-Based Support for Suicidal Ideation in the Scottish Population: Usability Study
JMIR Form Res 2025;9:e55932
URL: https://formative.jmir.org/2025/1/e55932
doi:10.2196/55932
PMID:

©Heather McClelland, Rory C O'Connor, Laura Gibson, Donald J MacIntyre. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Engagement With an Internet-Administered, Guided, Low-Intensity Cognitive Behavioral Therapy Intervention for Parents of Children Treated for Cancer: Analysis of Log-Data From the ENGAGE Feasibility Trial

Christina Reuther[1], MSc; Louise von Essen[1], PhD; Mudassir Imran Mustafa[1], PhD; Markus Saarijärvi[2*], PhD; Joanne Woodford[1*], PhD

[1]CIRCLE - Complex Intervention Research in Health and Care, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden
[2]Department of Neurobiology, Care Sciences and Society, Karolinska Institute, Stockholm, Sweden
[*]these authors contributed equally

Corresponding Author:
Louise von Essen, PhD
CIRCLE - Complex Intervention Research in Health and Care
Department of Women's and Children's Health
Uppsala University
Dag Hammarskjölds 14B
Uppsala, 75237
Sweden
Phone: 46 70 425 07 14
Email: louise-von.essen@uu.se

Abstract

Background: Parents of children treated for cancer may experience psychological difficulties including depression, anxiety, and posttraumatic stress. Digital interventions, such as internet-administered cognitive behavioral therapy, offer an accessible and flexible means to support parents. However, engagement with and adherence to digital interventions remain a significant challenge, potentially limiting efficacy. Understanding factors influencing user engagement and adherence is crucial for enhancing the acceptability, feasibility, and efficacy of these interventions. We developed an internet-administered, guided, low-intensity cognitive behavioral therapy (LICBT)–based self-help intervention for parents of children treated for cancer, (EJDeR [internetbaserad självhjälp för föräldrar till barn som avslutat en behandling mot cancer or internet-based self-help for parents of children who have completed cancer treatment]). EJDeR included 2 LICBT techniques—behavioral activation and worry management. Subsequently, we conducted the ENGAGE feasibility trial and EJDeR was found to be acceptable and feasible. However, intervention adherence rates were marginally under progression criteria.

Objective: This study aimed to (1) describe user engagement with the EJDeR intervention and examine whether (2) sociodemographic characteristics differed between adherers and nonadherers, (3) depression and anxiety scores differed between adherers and nonadherers at baseline, (4) user engagement differed between adherers and nonadherers, and (5) user engagement differed between fathers and mothers.

Methods: We performed a secondary analysis of ENGAGE data, including 71 participants. User engagement data were collected through log-data tracking, for example, communication with e-therapists, homework submissions, log-ins, minutes working with EJDeR, and modules completed. Chi-square tests examined differences between adherers and nonadherers and fathers and mothers concerning categorical data. Independent-samples *t* tests examined differences regarding continuous variables.

Results: Module completion rates were higher among those who worked with behavioral activation as their first LICBT module versus worry management. Of the 20 nonadherers who opened the first LICBT module allocated, 30% (n=6) opened behavioral activation and 70% (n=14) opened worry management. No significant differences in sociodemographic characteristics were found. Nonadherers who opened behavioral activation as the first LICBT module allocated had a significantly higher level of depression symptoms at baseline than adherers. No other differences in depression and anxiety scores between adherers and nonadherers were found. Minutes working with EJDeR, number of log-ins, days using EJDeR, number of written messages sent to e-therapists, number of written messages sent to participants, and total number of homework exercises submitted were

XSL•FO
RenderX

significantly higher among adherers than among nonadherers. There were no significant differences between fathers and mothers regarding user engagement variables.

Conclusions: Straightforward techniques, such as behavioral activation, may be well-suited for digital delivery, and more complex techniques, such as worry management, may require modifications to improve user engagement. User engagement was measured behaviorally, for example, through log-data tracking, and future research should measure emotional and cognitive components of engagement.

Trial Registration: ISRCTN Registry 57233429; https://doi.org/10.1186/ISRCTN57233429

(JMIR Form Res 2025;9:e67171) doi:10.2196/67171

KEYWORDS

childhood cancer survivor; cognitive behavioral therapy; engagement; internet-administered intervention; log-data; parents

Introduction

Background

The digital delivery of psychological interventions, such as cognitive behavioral therapy (CBT), particularly internet-administered cognitive behavioral therapy (iCBT), has rapidly grown over the past 2 decades [1]. iCBT interventions have shown to be effective in reducing symptoms of depression [2] and anxiety [3,4], with outcomes comparable with traditional face-to-face CBT when professional guidance is included as a component [5,6]. Flexibility, convenience [7], and the ability to overcome logistical and financial barriers [8], alongside the potential to reduce the stigma associated with seeking mental health treatment [8,9], are significant advantages of iCBT. In addition, given the lack of available therapists, iCBT interventions represent a potentially low-cost alternative to disseminating psychological interventions to a wider audience [10].

Despite these advantages, how users engage with digital interventions, including iCBT, is uncertain [11,12]. User engagement, which can be defined as how and to what extent users actively use digital interventions, is associated with the efficacy of iCBT interventions [13,14]. Low or inconsistent user engagement with digital interventions is closely associated with common issues such as low uptake (ie, a smaller-than-expected number of users participating in and benefiting from the intervention), variable usage (ie, users engaging with the intervention inconsistently or unevenly), and high attrition rates (ie, users drop out of the intervention) [15,16].

A related concept to user engagement is “adherence,” referring to whether an intervention is used in the way it was designed to maximize efficacy [17]. Adherence may be based on a minimal level of user engagement considered necessary to result in changes to intended outcomes, that is, a minimum treatment dose [14]. Although adherence to digital interventions is an evolving concept [18], the total number of completed sessions has been found to predict treatment response in iCBT interventions [19]. However, knowledge concerning the causal pathways between user engagement, adherence, and efficacy remains limited [20]. Furthermore, there is insufficient understanding of the potential association between sociodemographic factors, for example, age, gender, sex, and clinical characteristics (ie, depression or anxiety symptoms), and user engagement and adherence [19]. Adherence rates also vary between studies, with a review of adherence to iCBT interventions reporting adherence rates ranging from 6% to 100% [16]. Furthermore, high attrition rates, particularly in unguided iCBT interventions, present additional challenges [21], potentially resulting in biased treatment effects [22] and limiting generalizability [23].

Although user engagement is associated with treatment response [18], there remains a lack of clarity on how to conceptualize, define, and measure user engagement within the context of digital interventions [11,12,20]. Engagement is considered multidimensional [24,25] and can include to what extent an intervention is used, for example, by log-data tracking to quantitatively capture digital intervention usage [20], as well as the user’s subjective experience, for example, affect, attention, and interest [25]. Therefore, one way to explore user engagement is by analyzing log-data (ie, automatically collected digital records of a participant’s use of an intervention such as session duration, completion rates, and interaction patterns) [26,27] and examining potential relationships between user engagement and participant’s sociodemographic and clinical characteristics [28]. There is also a need to investigate the use of digital interventions to explore how certain aspects of user engagement may relate to adherence [20] and inform the minimal usage required to establish adherence [29].

This Study

We conducted a single-arm feasibility trial (ENGAGE) [30-32] to examine the acceptability and feasibility of an internet-administered, guided, low-intensity cognitive behavioral therapy (LICBT) intervention for parents of children treated for cancer: EJDeR (Swedish acronym: intErnetbaserad sJälvhjälp för förälDrar till barn som avslutat en behandling mot canceR [Internet-based self-help for parents of children who have completed cancer treatment]) [33-38]. EJDeR was developed alongside parent research partners [38] and is designed for parents of children treated for cancer experiencing symptoms of depression and/or generalized anxiety disorder (GAD) related to their child’s cancer treatment [33-38]. EJDeR is delivered through the Uppsala University Psychosocial Care Programme (U-CARE) portal (Portal), an in-house web-based platform, designed to deliver digital interventions and support the execution of study procedures (eg, online consent and data collection) [38]. EJDeR includes four modules: (1) introduction and psychoeducation (IPE), (2) behavioral activation (BA), (3)

XSL•FO
RenderX

worry management (WM), and (4) relapse prevention (RP), and is guided by an e-therapist. Findings from ENGAGE show EJDeR is an acceptable and feasible intervention; therefore, progression to an external pilot randomized controlled trial (RCT) to prepare for the design and conduct of a future superiority RCT is warranted. However, adherence to the minimal treatment dose was 47.9%, marginally under the progression criteria of 50% [31]. In addition, visual inspection of adherence data indicated possible differences in adherence rates by module worked with and between fathers and mothers [31]. Previous research indicates fathers and mothers may engage differently with psychological interventions [39], emphasizing the need to consider gender-specific engagement patterns in digital intervention design. For example, fathers have been found to prefer solution-focused, practical, and time-efficient formats that offer structured guidance and actionable strategies [40]. In contrast, mothers have been found to prefer more emotional and process-oriented intervention content, valuing opportunities for self-reflection and emotional processing [41].

Given these findings, there is a need to conduct a secondary analysis of data from ENGAGE to better understand how participants engaged with the intervention. Results may help inform ways to adapt and modify the intervention to improve adherence rates before proceeding to an external pilot RCT. Further modification of the intervention to improve user engagement and thus adherence may also improve treatment outcomes in a future superiority RCT [31]. Findings may also have wider applicability to other iCBT interventions.

Study Aims

This study aimed to (1) describe user engagement with the EJDeR intervention, (2) examine whether sociodemographic and baseline clinical characteristics differed between adherers and nonadherers, (3) examine whether depression and anxiety scores differed between adherers and nonadherers at baseline, (4) examine whether user engagement differed between adherers and nonadherers, and (5) examine whether user engagement differed between fathers and mothers.

Methods

Study Design

This study is a secondary analysis of data from ENGAGE. The trial protocol [32] and main findings from ENGAGE are reported elsewhere [31].

Participant Inclusion Criteria

Eligibility criteria included being (1) a parent of a child diagnosed with cancer during childhood (0-18 years) who completed cancer treatment 3 months to 5 years previously (timespan informed by our previous longitudinal research that has identified this as a period of vulnerability for parents [42,43], (2) resident in Sweden, (3) able to read and understand Swedish, (4) able to access email, internet, and BankID (a citizen authentication system used in Sweden), and (5) seeking psychological support related to the child's cancer. Exclusion criteria included (1) severe and enduring mental health problems, for example, bipolar disorder and psychosis; (2) acute suicidality; (3) misuse of alcohol; street drugs or prescription medication; and (4) currently attending psychological treatment. Participants excluded due to criteria 1-3 were directed to appropriate health care services.

Participant Recruitment

Parents were recruited into ENGAGE using 2 main approaches. First, personal identification numbers of children were obtained from the Swedish Childhood Cancer Registry (National Quality Registry) and linked to parents' names and addresses through NAVET, a registry held by the Swedish Tax Agency. Parents were invited to participate by postal invitations using random blocks of 100 until the target number of 50 was reached. Second, advertisements were placed on relevant social media sites and websites of patient organizations and interest groups [31].

An adapted CONSORT (Consolidated Standards of Reporting Trials) diagram for ENGAGE has been published [31]. Recruitment took place over 5 months (July 3, 2020, to November 30, 2020). A total of 72 participants gained access to EJDeR, with 1 excluded shortly after access (severe and enduring mental health difficulty), resulting in a sample of 71.

Intervention

EJDeR was delivered via the Portal, incorporating text, homework exercises, films, illustrations, and audio content. A detailed description of EJDeR has been published [38], and an overview of the intervention is presented in Figure 1. EJDeR was delivered over 12 weeks and consisted of 4 modules: (1) IPE, (2) BA, (3) WM, and (4) RP. It includes 2 LICBT techniques—BA for depression [44] and WM for GAD [45]. The LICBT clinical protocol for BA has been published elsewhere [44]. Participants are supported to re-engage with pleasurable, necessary, and routine activities they have stopped doing in a gradual and structured way. The full LICBT clinical protocol for WM has also been published elsewhere [45,46]. Participants record worries and categorize worries into 2 types—practical (eg, important and can be solved) and hypothetical (eg, important but have no way of being solved). Participants work with problem-solving for practical worries and worry time for hypothetical worries.

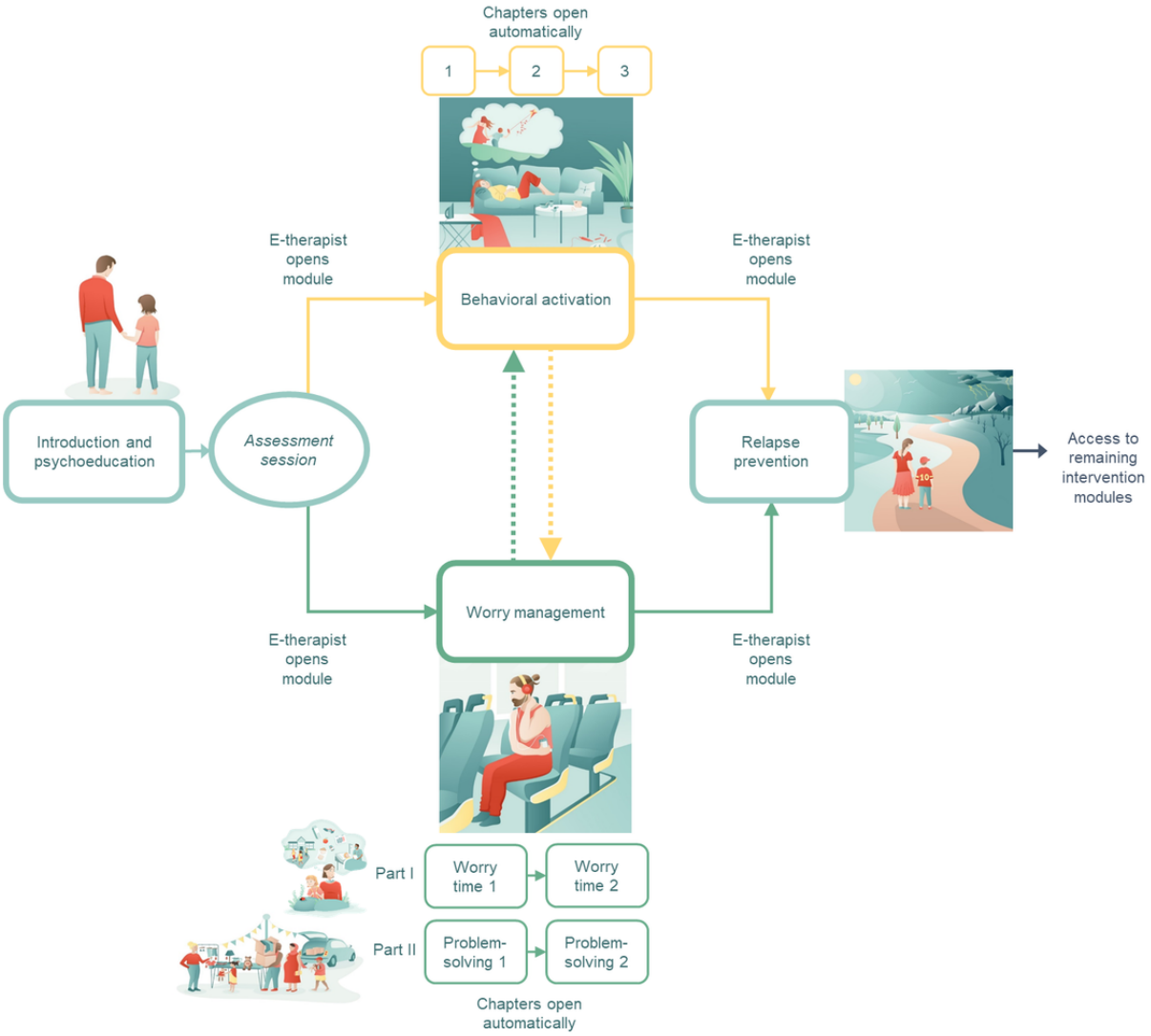
Guidance was provided by e-therapists, who were trained in the competencies required to support LICBT. After being provided with access to EJDeR, participants were invited to attend an initial assessment session (telephone or videoconferencing) with their e-therapist. During the initial assessment, e-therapists used patient-centered interviewing techniques to understand the participant's main presenting difficulties, provide psychoeducation, introduce the LICBT techniques used in EJDeR, and come to a collaborative decision to allocate 1 of the 2 LICBT modules to first work with, based on the participant's main difficulty, for example, BA for depression or WM for GAD. E-therapists did not have access to participants' baseline assessment scores (ie, depression and anxiety symptoms) and therefore allocation to the first LICBT module to work with was not informed by symptom severity at baseline. Subsequently, the e-therapist provided participants with access to the LICBT module. E-therapists provided weekly written messages via the Portal to guide participants in using

XSL•FO
RenderX

the intervention [47]. To facilitate engagement with the LICBT modules, weekly homework exercises were either completed on the Portal or printed as a PDF and completed offline. E-therapists also invited participants to attend a midintervention booster session (telephone or videoconferencing). After completing the first LICBT module allocated (BA or WM), participants could choose to work with the remaining LICBT module. After completing either BA or WM, participants were provided with access to RP by their e-therapist.

Persuasive system design elements [48,49] are built into EJDeR to improve user engagement, including tunneling, for example, content is delivered in a predefined step-wise order to guide participants through using EJDeR, tailoring (eg, EJDeR content is personalized to user needs [ie, depression or GAD]), rehearsal (eg, homework exercises are repeated), and liking (eg, use of professional illustrations).

Figure 1. An overview of the structure of EJDeR. Figure from Woodford et al [38] which is published under Creative Commons Attribution 4.0 International License [50].



Intervention Adherence

Adherence to EJDeR was conceptualized as participants engaging with the intervention following a priori–defined minimum treatment dose. Intervention adherers were defined as those (1) attending the initial assessment session with an e-therapist (telephone or videoconferencing), (2) completing the IPE module (2 chapters) and 1 LICBT treatment module (ie, BA [3 chapters] or WM defined as completing an introduction [2 chapters] and completing either problem solving [2 chapters] or worry time [2 chapters]), and (3) attending the midintervention booster session with an e-therapist (telephone or videoconferencing). For a module to be defined as being completed, participants were required to submit all chapters included within each module to their e-therapist on the Portal. Our minimum treatment dose was informed by research suggesting that guidance from a trained professional is associated with larger treatment effects than self-administered LICBT [51,52]. In addition, we considered it important for participants to complete the IPE module to gain an understanding of the CBT rationale and complete all chapters for 1 LICBT technique to enable them to understand and apply

XSL•FO
RenderX

the technique to manage their symptoms of depression and/or anxiety. We did not include the submission of weekly homework exercises in our definition of adherence given participants were able to complete on the Portal or print as a PDF and complete offline.

Measures

Sociodemographic Characteristics

Sociodemographic characteristics were collected during a telephone eligibility interview with a licensed psychologist and included age, sex (male or female), relationship status (partner or single), highest level of education (lower secondary, upper secondary, postsecondary nontertiary, tertiary, or PhD), employment status (employed or unemployed), number of children, housing situation (rental, apartment ownership, house ownership, or other), region of birth (Nordic countries, Asia, Europe [excluding Nordic countries], or Africa), previous psychological treatment (yes or no), physical health problem (yes or no), and previous traumatic or difficult life events (yes or no).

Baseline Clinical Characteristics

Clinical characteristics were collected via a Portal assessment at baseline. Symptoms of depression were assessed using the 9-item Patient Health Questionnaire (PHQ-9) [53]. Symptoms of anxiety were assessed using the 7-item Generalized Anxiety Disorder scale (GAD-7) [54].

User Engagement

Log-Data Tracking on the Portal

User action logging was enabled through action metadata management to allow user behavior analysis, including (1) minutes working with EJDeR, (2) number of log-ins, (3) number of days using EJDeR, (4) number of written messages sent to e-therapists, (5) number of written messages sent to participants, (6) total number of modules opened, (7) total number of modules completed, (8) opened modules (IPE and RP), (9) completed modules (IPE and RP), (10) opened first LICBT module allocated (BA and WM), (11) completed first LICBT module allocated (BA and WM), and (12) submitted homework exercises.

Data Collected Outside the Portal

Intervention activities outside of the Portal were collected by research team members and entered into a Microsoft Excel spreadsheet, including (1) attending the initial assessment session (yes or no), and (2) attending the midintervention booster session (yes or no).

Statistical Analysis

All statistical analyses were performed in the SPSS (version 28.0.1; IBM Corp). Given multiple statistical comparisons were conducted, a Bonferroni correction was applied to adjust for type I error inflation [55]. Missing data (% of missing items <0.001) were handled using single-value imputation. Sociodemographic and baseline clinical characteristics and user engagement data for all participants are reported using descriptive statistics. Absolute proportions are reported for categorical variables and means and SDs are reported for continuous variables. Chi-square tests were used to examine differences between adherers and nonadherers and fathers and mothers concerning categorical data. Effect sizes (Cohen w) were calculated to determine the magnitude of difference between the groups for categorical variables. Cohen w rule of thumb cutoffs was adopted: 0.1-0.3, indicating a small difference; 0.3-0.5, a moderate difference; and >0.5, a large difference [56]. Independent-sample *t* tests were used to examine differences regarding continuous variables.

Ethical Considerations

This study was approved by the Regional Ethical Review Board in Uppsala, Sweden (Dnr: 2017/527), with an amendment obtained from the Swedish Ethical Review Authority (2019-03083). The research was conducted following the Helsinki Declaration and relevant Swedish regulations for human subject research. Parents interested in participating provided informed consent via the Portal. Data collected for the study were pseudonymized to ensure participant confidentiality. Participants were not compensated for their participation in the study. Participation was voluntary, and participants could withdraw at any time.

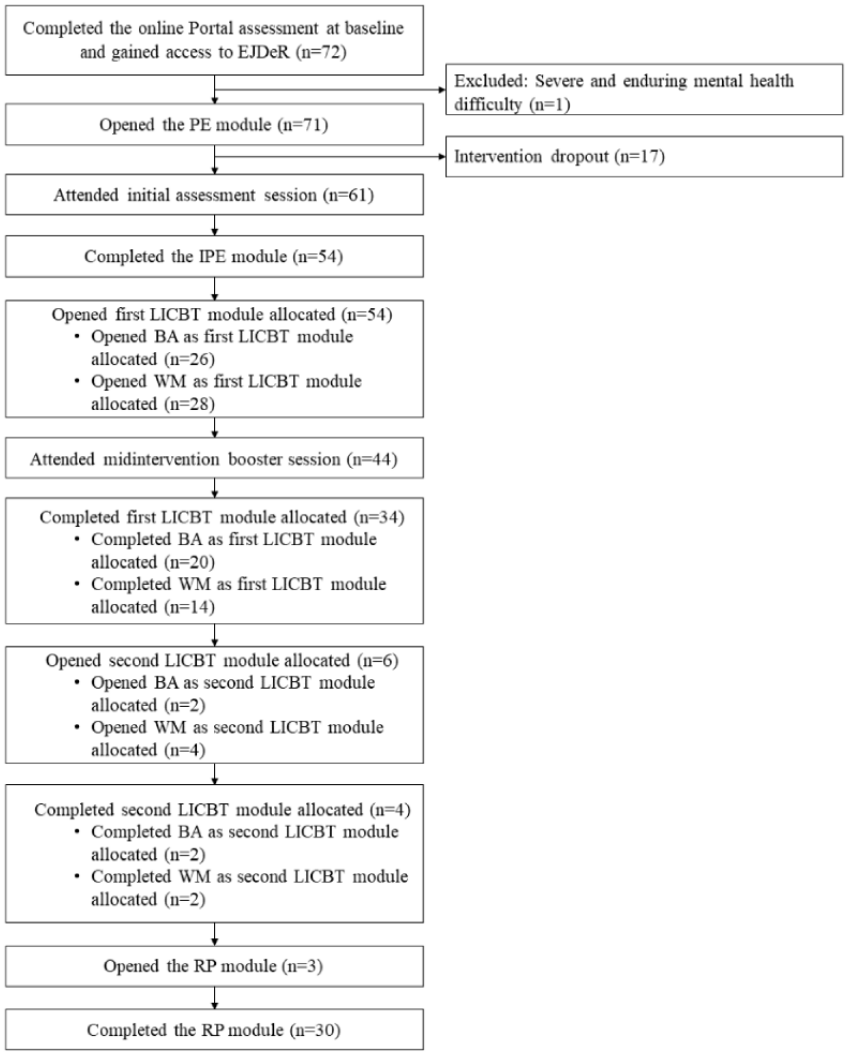
Results

Intervention Flow

The flow of participants through the intervention is illustrated in Figure 2. In total, 77% (20/26) of participants starting with BA as the first LICBT module allocated completed the module versus 50% (14/28) starting with WM.

XSL•FO
RenderX

Figure 2. Flow of participants through the EJDeR intervention. BA: behavioral activation; EJDeR: intErnetbaserad sJälvhjälp för förälDrar till barn som avslutat en behandling mot canceR; IPE: introduction and psychoeducation; LICBT: low-intensity cognitive behavioral therapy; PE: psychoeducation; RP: relapse prevention; WM: worry management. Two participants were provided access to the RP module without completing a LICBT module, due to e-therapist error.



User Engagement With the EJDeR Intervention

Participants spent an average of 205.36 (SD 154.64) minutes working with EJDeR, logging in an average of 19.99 (SD 14.95) times over 76.31 (SD 31.49) days. Participants opened a mean number of 2.3 (SD 0.92) modules and completed an average of 1.65 (SD 1.31) modules (Table 1).

Table 1. User engagement with the EJDeR intervention (N=71).

User engagement (log-data on the Portal)	Mean (SD)	Range
Minutes working with EJDeR[a]	205.36 (154.64)	8.59-611.74
Number of log-ins	19.99 (14.95)	1-72
Number of days using EJDeR	76.31 (31.49)	0-126
Number of written messages sent to e-therapists	8.52 (7.56)	0-33
Number of written messages sent to participants	28.79 (16.26)	0-74
Total number of modules opened	2.30 (0.92)	1-4
Total number of modules completed	1.65 (1.31)	0-4
Total homework exercises submitted	2.70 (2.78)	0-11

[a]EJDeR: intErnetbaserad sJälvhjälp för förälDrar till barn som avslutat en behandling mot cancer.

XSL•FO
RenderX

Sociodemographic Between Intervention Adherers and Nonadherers

There were no significant differences between intervention adherers and nonadherers by sociodemographic characteristics (Table 2).

XSL•FO
RenderX

Table 2. Sociodemographic characteristics for total sample (N=71) and by adherers (n=34) and nonadherers (n=37) to the minimum treatment dose for the EJDeR intervention[a].

Sociodemographic characteristics	Total study sample (N=71)		Adherers (n=34)		Nonadherers (n=37)		*P* value (effect size)[b]
	n (%)[c]	Range	n (%)	Range	n (%)	Range	
Age (years), mean (SD)	42.73 (7.02)	26-62	41.71 (5.30)	33-59	43.68 (8.25)	26-62	.24 (*d*=0.28)
Sex							.99 (w=0.12)
Female	46 (65)	—[d]	22 (65)	—	24 (65)	—	
Male	25 (35)	—	12 (35)	—	13 (35)	—	
Relationship status							.41 (w=0.05)
Partner	60 (85)	—	30 (88)	—	30 (81)	—	
Single	11 (15)	—	4 (12)	—	7 (19)	—	
Highest level of education							.74 (w=0.09)
Lower secondary	1 (1)	—	0 (0)	—	1 (3)	—	
Upper secondary	13 (18)	—	5 (15)	—	8 (22)	—	
Postsecondary nontertiary	3 (4)	—	1 (3)	—	2 (5)	—	
Tertiary	52 (73)	—	27 (79)	—	25 (68)	—	
PhD	2 (3)	—	1 (3)	—	1 (3)	—	
Employment status							.90 (w=0.11)
Employed	63 (89)	—	30 (88)	—	33 (89)	—	
Unemployed	8 (11)	—	4 (12)	—	4 (11)	—	
Number of children, mean (SD)	2.28 (0.78)	1-5	2.12 (0.69)	1-5	2.43 (0.84)	1-4	.09 (*d*=0.40)
Housing situation							.07 (w=0.01)
Rental	7 (10)	—	2 (6)	—	5 (14)	—	
Apartment ownership	17 (24)	—	6 (18)	—	11 (30)	—	
House ownership	44 (62)	—	26 (76)	—	18 (49)	—	
Other	3 (4)	—	0 (0)	—	3 (8)	—	
Region of birth							.18 (w=0.02)
Nordic countries[e]	59 (83)	—	31 (91)	—	28 (76)	—	
Asia	6 (9)	—	3 (9)	—	3 (8)	—	
Europe (excluding Nordic countries)	5 (7)	—	0 (0)	—	5 (14)	—	
Africa	1 (1)	—	0 (0)	—	1 (3)	—	
Previous psychological treatment							.57 (w=0.07)
Yes	38 (54)	—	17 (50)	—	21 (57)	—	
No	33 (46)	—	17 (50)	—	16 (43)	—	
Physical health problems							.61(w=0.07)
Yes	23 (32)	—	10 (29)	—	13 (35)	—	
No	48 (68)	—	24 (71)	—	24 (65)	—	
Previous traumatic or difficult life event							.67 (w=0.08)
Yes	57 (80)	—	28 (82)	—	29 (78)	—	
No	14 (20)	—	6 (18)	—	8 (22)	—	

[a]Adherence to EJDeR was conceptualized as users engaging with the intervention in accordance with a priori–defined minimum treatment dose.

XSL•FO
RenderX

Intervention adherers are therefore defined as those (1) attending the initial assessment session with an e-therapist, (2) completing the psychoeducation module and one low-intensity treatment module (ie, behavioral activation or worry management), and (3) attending the midintervention booster session with an e-therapist.

[b]A Bonferroni correction was applied based on 11 tests resulting in an α of .005.

[c]Percentages may not add to 100 due to rounding.

[d]Not applicable.

[e]Nordic countries represented in the study sample include Denmark, Finland, Norway, and Sweden.

Depression and Anxiety Scores Between Intervention Adherers and Nonadherers at Baseline

Nonadherers who opened BA as the first LICBT module allocated had a significantly higher level of depression symptoms at baseline than adherers (*P*=.04); however, this finding was no longer significant after applying a Bonferroni correction (α=.008). There were no other significant differences between intervention adherers and nonadherers by depression and anxiety score at baseline (Table 3).

Table 3. Depression and anxiety scores at baseline for total sample (N=71) and by adherers (n=34) and nonadherers (n=37) to the minimum treatment dose for the EJDeR intervention[a].

Depression and anxiety scores at baseline	Total study sample (N=71)		Adherers (n=34)		Nonadherers (n=37)		*P* value (effect size)[b]
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	
PHQ-9[c] for all participants	71	6.55 (5.05)	34	6.24 (4.65)	37	6.84 (5.43)	.62 (d=0.12)
PHQ-9 for participants who opened BA[d] as first LICBT[e] module[f]	26	8.31 (6.03)	20	7.00 (5.49)	6	12.67 (6.15)	.04 (d=0.97)
PHQ-9 for participants who opened WM[g] as first LICBT module[f]	28	4.46 (3.01)	14	5.14 (2.96)	14	3.79 (3.02)	.24 (d=–0.45)
GAD-7[h] for all participants	71	5.92 (4.53)	34	6.06 (4.25)	37	5.78 (4.83)	.80 (d=–0.06)
GAD-7 for participants who opened BA as first LICBT module[f]	26	6.85 (5.09)	20	5.90 (4.62)	6	10.0 (5.73)	.08 (d=0.79)
GAD-7 for participants who opened WM as first LICBT module[f]	28	5.11 (3.75)	14	6.29 (3.81)	14	3.93 (3.41)	.10 (d=–0.65)

[a]Adherence to EJDeR was conceptualized as users engaging with the intervention in accordance with a priori–defined minimum treatment dose. Intervention adherers are therefore defined as those (1) attending the initial assessment session with an e-therapist, (2) completing the psychoeducation module and 1 low-intensity treatment module (ie, behavioral activation or worry management), and (3) attending the midintervention booster session with an e-therapist.

[b]A Bonferroni correction was applied based on 6 tests resulting in an α of .008.

[c]PHQ-9: 9-item Patient Health Questionnaire.

[d]BA: behavioral activation.

[e]LICBT: low-intensity cognitive behavioral therapy.

[f]Total study sample based on those gaining access to a low-intensity cognitive behavioral therapy module.

[g]WM: worry management.

[h]GAD-7: 7-item Generalized Anxiety Disorder scale.

User Engagement Between Intervention Adherers and Nonadherers

There was a significant difference between adherers and nonadherers by first LICBT module allocated (*P*=.04); however, this finding was no longer significant after applying a Bonferroni correction (α=.007). Of the 34 adherers who opened the first LICBT module allocated, 59% (n=20) opened BA as the first LICBT module allocated and 41% (n=14) opened WM. Of the 20 nonadherers who opened the first LICBT module allocated, 30% (6) opened BA as their first LICBT module allocated and 70% (14) opened WM.

The number of minutes working with EJDeR (*P*≤.001), number of log-ins (*P*≤.001), number of days using EJDeR (*P*≤.001), number of written messages sent to e-therapists (*P*≤.001), number of written messages sent to participants (*P*≤.001), and total number of homework exercises submitted (*P*≤.001) were significantly higher among intervention adherers compared with nonadherers. There were no significant differences between adherers and nonadherers related to the other user engagement variables (Table 4).

XSL•FO
RenderX

Table 4. User engagement with the EJDeR intervention by adherers (n=34) and non-adherers (n=37) to the minimum treatment dose for the EJDeR intervention[a].

Measure of user engagement (log-data on the Portal)	Adherers (n=34)		Nonadherers (n=37)		*P* value (effect size)[b]
	Mean (SD)	Range	Mean (SD)	Range	
Minutes working with EJDeR[c]	325.82 (126.61)	142.41-611.74	94.67 (74.21)	8.59-282.95	<.001 (d=2.23)
Number of log-ins	31.32 (12.64)	5-72	9.57 (7.42)	1-29	<.001 (d=2.10)
Number of days using EJDeR	91.47 (13.76)	57-126	62.38 (36.58)	0-126	<.001 (d=1.05)
Number of written messages sent to e-therapists	12.88 (7.64)	2-33	4.51 (4.79)	0-17	<.001 (d=1.31)
Number of written messages sent to participants	40.21 (12.09)	24-74	18.30 (11.98)	0-47	<.001 (d=1.82)
First LICBT[d] module allocated opened					.04 (w=0.01)
Opened BA[e], n (%)	20[f] (59)	—[g]	6[h] (30)	—	
Opened WM[i], n (%)	14[f] (41)	—	14[h] (70)	—	
Total homework exercises submitted	4.79 (2.53)	0-11	0.78 (1.13)	0-4	<.001 (d=2.05)

[a]Adherence to EJDeR was conceptualized as users engaging with the intervention in accordance with a priori defined minimum treatment dose. Not all user engagement variables listed in Table 1 are included in this analysis as some measures of user engagement, that is, number of modules opened, attended initial assessment session, attended mid-intervention booster session, inform the minimum treatment dose.

[b]A Bonferroni correction was applied based on 7 tests resulting in an α of .007.

[c]EJDeR: intErnetbaserad sJälvhjälp för förälDrar till barn som avslutat en behandling mot canceR.

[d]LICBT: low-intensity cognitive behavioral therapy.

[e]BA: behavioral activation.

[f]number of adherers who opened the first low-intensity cognitive behavioral therapy module allocated (n=34).

[g]Not applicable.

[h]number of nonadherers who opened the first low-intensity cognitive behavioral therapy module allocated (n=20).

[i]WM: worry management.

User Engagement Between Fathers and Mothers

There were no significant differences between fathers and mothers related to user engagement variables (Table 5).

XSL•FO
RenderX

Table 5. User engagement with the EJDeR intervention[a] by fathers (n=25) and mothers (n=46).

Measure of user engagement	Fathers		Mothers		*P* value (effect size)[b]
	Statistical value	Range	Statistical value	Range	
Log-data on the Portal					
Minutes working with EJDeR[c], mean (SD)	185.53 (135.71)	18.67-546.51	216.14 (164.44)	8.59-611.74	.43 (*d*=–0.20)
Number of log-ins, mean (SD)	18.08 (14.05)	1-56	21.02 (15.46)	1-72	.43 (*d*=–0.20)
Number of days using EJDeR, mean (SD)	73.04 (33.79)	0-126	78.09 (30.41)	12-126	.52 (*d*=–0.16)
Number of written messages sent to e-therapists, mean (SD)	6.72 (5.93)	0-22	9.50 (8.20)	0-33	.14 (*d*=–0.39)
Number of written messages sent to participants, mean (SD)	26.52 (15.82)	0-66	30.02 (16.53)	7-74	.39 (*d*=–0.22)
Total number of modules opened, mean (SD)	2.28 (0.94)	1-4	2.30 (0.92)	1-4	.92 (*d*=–0.02)
Opened IPE[d] module, n/N (%)	25/25 (100)	—[e]	46/46 (100)	—	—
Completed IPE module, n/N (%)					.57 (w=0.07)
Yes	19/25 (76)	—	32/46 (70)	—	
No	6/25 (24)	—	14/46 (30)	—	
First LICBT[f] module allocated opened, n/N (%)					.36 (w=0.04)
Opened BA[g]	8/20[h] (40)	—	18/34[i] (53)	—	
Opened WM[j]	12/20[h] (60)	—	16/34[i] (47)	—	
Completed first LICBT module allocated					
Completed BA, n/N (%)					.39 (w=0.08)
Yes	7/8 (88)	—	13/18 (72)	—	
No	1/8 (13)	—	5/18 (28)	—	
Completed WM, n/N (%)					.45 (w=0.08)
Yes	5/12 (42)	—	9/16 (56)	—	
No	7/12 (58)	—	7/16 (44)	—	
Opened RP[k] module[l], n/N (%)					.03 (w=0.00)
Yes	8/20 (40)	—	24/34 (71)	—	
No	12/20 (60)	—	10/34 (29)	—	
Completed RP module, n/N (%)					.40 (w=0.07)
Yes	8/8 (100)	—	22/24 (92)	—	
No	0/8 (0)	—	2/24 (8)	—	
Total homework exercises submitted, mean (SD)	2.52 (2.52)	0-9	2.80 (2.94)	0-11	.68 (*d*=–0.10)
Data collected outside the Portal					
Attended initial assessment session, n/N (%)					.73 (w=0.09)
Yes	21/25 (84)	—	40/46 (87)	—	
No	4/25 (16)	—	6/46 (13)	—	
Attended midintervention booster session, n/N (%)					.80 (w=0.09)
Yes	16/25 (64)	—	28/46 (61)	—	

XSL•FO
RenderX

Measure of user engagement	Fathers		Mothers		*P* value (effect size)[b]
	Statistical value	Range	Statistical value	Range	
No	9/25 (36)	—	18/46 (39)	—	

[a]Adherence to EJDeR was conceptualized as users engaging with the intervention in accordance with a priori–defined minimum treatment dose. Not all user engagement variables listed in Table 1 are included in this analysis as some measures of user engagement, that is, number of modules opened, attended initial assessment session, attended midintervention booster session, and inform the minimum treatment dose.

[b]A Bonferroni correction was applied based on 15 tests resulting in an α of .003.

[c]EJDeR: intErnetbaserad sJälvhjälp för förälDrar till barn som avslutat en behandling mot canceR.

[d]IPE: introduction and psychoeducation.

[e]Not applicable.

[f]LICBT: low-intensity cognitive behavioral therapy.

[g]BA: behavioral activation.

[h]Number fathers who opened the first low-intensity cognitive behavioral therapy module allocated (n=20).

[i]Number mothers who opened the first low-intensity cognitive behavioral therapy module allocated (n=34).

[j]WM: worry management.

[k]RP: relapse prevention.

[l]Two participants were provided access to the RP module without completing a low-intensity cognitive behavioral therapy module due to e-therapist error.

Discussion

Principal Findings

This study aimed to (1) describe user engagement with the EJDeR intervention, (2) examine whether sociodemographic and baseline clinical characteristics differed between adherers and nonadherers, (3) examine whether depression and anxiety scores differed between adherers and nonadherers at baseline, (4) examine whether user engagement differed between adherers and nonadherers, and (5) examine whether user engagement differed between fathers and mothers. Overall, 48% (34/71) of participants adhered to the minimum treatment dose. In total, 77% (20/26) of participants starting with BA as the first LICBT module allocated completed the module versus 50% (14/28) starting with WM. There were no significant differences between intervention adherers and nonadherers related to sociodemographic characteristics. Nonadherers who opened BA as the first LICBT module allocated had a significantly higher level of depression symptoms at baseline than adherers (*P*=.04); however, after applying a Bonferroni correction (α=.008), this finding was no longer significant. There were no other significant differences between intervention adherers and nonadherers by depression and anxiety score at baseline (Table 3). There was a significant difference between adherers and nonadherers by first LICBT module allocated (*P*=.04) however this finding was no longer significant after applying a Bonferroni correction (α=.007). Of the 20 nonadherers who opened the first LICBT module allocated, 30% (n=6) opened BA, and 70% (n=14) opened WM as their first LICBT module allocated. The number of log-ins, written messages sent to e-therapists, written messages sent to participants, and homework exercises submitted were significantly higher among intervention adherers than nonadherers. There were no significant differences between fathers and mothers related to user engagement variables.

User Engagement and Adherence

Importantly, adherence differed by the first LICBT module allocated, with 30% (6/20) of nonadherers opening BA versus 70% (14/20) opening WM. One potential explanation for this finding is that BA interventions are simple, straightforward, and easy to comprehend [57,58] and, therefore, may be particularly well-suited to digital delivery. Conversely, the WM module included 2 treatment components [45,46] (problem-solving and worry time), which may have been perceived as more complex and challenging for participants. A recent study examining engagement with a mobile phone app to deliver WM found that 84% (803/956) of users did not achieve the minimum treatment dose [59]. In addition, the LICBT module BA is designed to overcome sources of negative reinforcement and increase engagement with pleasurable, routine, and necessary activities in a structured and graded way. BA may therefore result in more immediate and tangible rewards (eg, improved mood and increased engagement in valued activities) than WM. Indeed, BA is associated with sudden therapeutic gains [60], and a recent component network meta-analysis on iCBT for depression found stronger evidence for BA as an effective intervention component compared with other techniques, for example, cognitive restructuring and problem-solving [61].

However, differences in adherence by the first LICBT module allocated may also reflect the nonrandomized nature of module allocation. Module allocation was informed by the participant's main presenting difficulties as assessed by the e-therapist, with a collaborative decision made with the participant to first start with BA for depression or WM for GAD. As such, the participant's presenting difficulties, rather than module content alone, may have contributed to differences in adherence [62,63]. Overall, baseline clinical characteristics (depression and anxiety scores) did not differ between intervention adherers and nonadherers, similar to recent studies in other populations [64,65]. However, nonadherers who opened BA as the first

XSL•FO
RenderX

LICBT module had a higher level of depression symptoms at baseline compared with adherers, although this finding was no longer significant after applying a Bonferroni correction. There were no other significant differences between intervention adherers and nonadherers by baseline clinical characteristics. The finding that nonadherers who opened BA as the first LICBT module had a higher level of depression symptoms at baseline is somewhat counterintuitive, given the effects of iCBT interventions are larger in those with moderate to severe depression [5]. However, other research has found higher levels of depressive symptoms at baseline to be associated with lower iCBT adherence [66,67] and higher rates of dropout from blended iCBT interventions [68]. Other studies have also found baseline anxiety symptoms to be an important predictor of adherence [69]. While this finding should be treated with caution, given small sample sizes and lack of power, symptoms at baseline may have influenced participants' engagement with and adherence to specific modules, highlighting a potential need to consider baseline characteristics when allocating digital intervention components, for example, specific therapeutic techniques [63,70]. Future research with large sample sizes is warranted to explore the association between baseline clinical characteristics and intervention engagement and adherence.

Importantly, adherers had good levels of engagement with intervention components that were not related to the minimum treatment dose, for example, minutes working with the intervention, number of log-ins, number of written messages sent to e-therapists, number of written messages sent to participants, and the total number of homework exercises submitted. Examining user engagement variables not included in the minimum treatment dose by adherers and nonadherers helped provide a more nuanced understanding of how participants engaged (or not) with the intervention. Defining and measuring adherence to digital interventions in a way that accurately reflects user engagement is challenging, and there is a lack of universally accepted definitions of adherence [23]. Definitions typically vary depending on the intervention's structure, ranging from module completion to log-in frequency to provide evidence of meaningful engagement with therapeutic components [71].

Indeed, defining user engagement and adherence in digital interventions is complex [72] and log-data tracking alone may not indicate a lack of intervention engagement, as participants may engage in intervention techniques offline, leading to higher levels of engagement than recorded by user action logging [64]. Furthermore, participants may have received an active treatment dose before reaching the a priori–determined minimum treatment dose, underscoring the complexities of conceptualizing adherence to digital interventions [18]. There have been calls to move beyond measuring user engagement using only behavioral tools (ie, system usage) and to measure emotional and cognitive engagement components of engagement [26]. For example, user engagement may include the extent of usage (eg, measured by user action logging) and subjective experience (eg, affect, attention, and interest) [20, 25]. In this study, we focused on measures of behavior within the Portal and did not measure subjective experience. Consequently, our findings may not fully capture the multifaceted nature of how participants may have engaged in the intervention and its associated factors. In our planned external pilot RCT, we will include additional measurements of user engagement beyond system usage data, such as validated self-report questionnaires [73,74] that measure affect, behavior, and cognition. However, there remains a need to develop ways of better defining, measuring, and differentiating between the different components of engagement in digital interventions [75].

The provision of support and guidance is another important factor potentially related to intervention adherence [28]. In this study, e-therapists sent an average of 40.21 (SD 12.09) written messages to adheres versus 18.30 (SD 11.98) to nonadherers. Personalized support and guidance are associated with improved retention [61] and effectiveness [5,6] in iCBT interventions. Our findings suggest additional consideration may be needed to train e-therapists to provide personalized support and motivate and encourage nonadherers to improve future intervention adherence. Future research could also look at providing an automated guidance function, for example, through a chatbot, to enhance adherence further [76].

No sociodemographic characteristics differed between intervention adherers and nonadherers. Furthermore, there were no significant differences between fathers and mothers regarding user engagement variables. The wider iCBT literature suggests mixed findings regarding the association between sociodemographic characteristics and adherence [77,78]. Some studies suggest that women and individuals with higher education levels are more likely to adhere to iCBT interventions compared with men and those with lower education levels [21]. However, conflicting results and lack of consistent associations limit the ability to conclude sociodemographic predictors of adherence [77].

Implications

Findings build upon our previously published results from ENGAGE [31]. Analysis of usage data provides a more detailed and nuanced picture of how participants interacted with the intervention. This analysis also identifies key factors associated with engagement and adherence. One important finding is the difference in adherence based on the first LICBT module allocated, with 70% (14/20) of nonadherers who opened their first LICBT module opening WM versus 30% (6/20) of nonadherers opening BA. This suggests early exposure to potentially challenging and complex content may cause users to disengage. While developed using persuasive systems design elements [48,49], for example, tunneling, tailoring, rehearsal, and liking, future adaptations to the EJDeR intervention will need to focus on improving navigation and usability, especially for WM. Flexibility and ease of use of digital interventions for mental health have been identified as effective engagement strategies [75]. Engaging in user-centered design processes, that is, involving users in future system development, may represent a way to improve intervention engagement and adherence [59].

In addition, the WM module's dual-component structure (ie, including 2 LICBT techniques—problem solving for practical worries and worry time for hypothetical worries) may have resulted in higher cognitive load, which could be a barrier to engagement and adherence. Participants first allocated to the

XSL•FO
RenderX

WM module may benefit from additional support and guidance from an e-therapist. For example, e-therapists could be trained to provide more personalized support and feedback [59], which may help minimize challenges and help users understand how to engage with each module effectively [68]. Using visual aids, videos, and immediate feedback can also make content more accessible and motivating, improving adherence and overall user experience [75,76]. Furthermore, acknowledging that module length and complexity can impact adherence, and intervention modifications could also include efforts to streamline module design to reduce cognitive load without compromising therapeutic content. Although parents and e-therapists came to a collaborative decision to allocate the LICBT module to first work with, further intervention modifications could include providing more module-specific information, for example, a module preview, which may also facilitate starting with a module that meets their needs and preferences [79].

Finally, issues with engagement and adherence can arise due to digital interventions adopting a "one-size-fits-all" approach. Instead, interventions require careful tailoring to meet participants' specific needs and context [80]. This is especially important when considering users of digital health interventions engage with these interventions in their own environment, whereby they may be experiencing multiple other demands [81]. iCBT interventions include user demands, such as time to engage in intervention (eg, reading the content, completing homework exercises, and attending support sessions). Lack of time is a common barrier to accessing psychological support reported by parents [82] and for parents of children treated for cancer, particularly [83]. Time was also the most frequently reported burden limiting participation in a group-based CBT videoconferencing program to support parents of children treated for cancer [84]. Parents may also experience reduced opportunities for privacy, further interfering with their ability to engage with the intervention. Other research exploring engagement with digital interventions found users with no children report higher intervention usage [85]. Other factors, including socioeconomic status, may influence how participants engage with interventions [14]. Furthermore, individual contextual factors such as digital literacy and access to technology can influence user engagement [86-88]. Despite the EJDeR intervention being developed alongside parent research partners [38], additional intervention tailoring may be required to further adapt the intervention to the specific needs and context of the population.

Strengths and Limitations

Given that this study is a secondary analysis of data from a single-arm feasibility trial, the sample size is small, and analyses may be underpowered to detect statistically significant differences. Due to the small sample size, we could not conduct a regression or correlation analysis. Multiple statistical comparisons were conducted, increasing the risk of Type I error. However, this risk was minimized by applying a Bonferroni correction to reduce the likelihood of spurious results [55]. Most participants were female, had an education level higher than upper secondary school, and were born in a Nordic country. This may limit the generalizability of findings, especially considering Sweden is a diverse and multicultural society, with 21% (2,170,627/10,551,700) of the population being foreign-born [89], and there is a need to develop culturally responsive digital interventions to enhance acceptability and engagement in ethnic minority populations [90]. The research team defined intervention adherence a priori following a minimum treatment dose. This definition fails to consider that participants may have conducted intervention activities outside the Portal, that is, offline [59]. Some participants may have achieved their personal intervention goals early and stopped engaging with the intervention, and yet were classified as intervention nonadherers [12].

We also adopted an exploratory approach that aimed to examine a broad range of variables potentially associated with user engagement and adherence. We recognize that the exploratory analysis of log-data has significant limitations, for example, lack of transparency, difficulties with replication, and the risk of "data dredging" [29]. While we set specific research aims to inform our approach, we did not use a framework to underpin our analysis. Adopting the framework for Analyzing and Measuring Usage and Engagement Data (AMUsED) [29] to define meaningful usage variables and generate research questions related to factors such as the interventions' design and theoretical underpinning would have strengthened our approach. Recognizing this limitation, we aim to adopt the AMUsED framework to inform usage analysis in our external pilot RCT to prepare for the design and conduct of a future superior RCT. Our exploratory approach also risked circular analysis, for example, examining user engagement variables associated with adherence may be considered circular given our definition of adherence is partially defined by intervention usage. However, as adherence (ie, the minimum treatment dose) was defined as attending sessions with the e-therapist and module completion, this does tell us much about how adherers versus nonadherers engaged with the Platform beyond these 2 metrics. For example, user engagement variables associated with frequency of use (eg, length of time working with EJDeR and number of log-ins) or engagement in intervention content such as homework exercises (eg, number of written homework exercises submitted) provided more nuanced information concerning how adherers versus nonadherers interacted with the intervention beyond the minimum treatment dose.

As a feasibility study, ENGAGE was not designed to examine the effectiveness of the EJDeR intervention, and we did not examine the association between intervention usage and clinical outcomes. This association will be examined in the planned future superiority RCT, should progression to a superiority RCT be warranted after conducting the external pilot RCT. The association between intervention usage and clinical outcomes will be used to further inform our definition of the minimum treatment dose by helping us to understand what optimal level of intervention usage can affect clinical outcomes [28].

Despite these limitations, our study has some important strengths. We examined the relationship between engagement and adherence, considering the therapeutic content of modules (eg, BA or WM) [59]. In addition, we used objective data (eg, log-data tracking on the Portal and data collected outside the Portal by members of the research team) to examine

XSL•FO
RenderX

engagement, overcoming biases associated with self-reported usage data [91].

Conclusions

Our findings provide important information concerning modifications to the EJDeR intervention to enhance user engagement and adherence, which may be related to intervention efficacy. Findings may also have broader applicability to improving user engagement in other iCBT interventions. Our findings suggest simple and straightforward LICBT techniques, such as BA, are well-suited to digital delivery. However, more complex LICBT techniques, such as WM, may need modification to improve ease of use and thus increase user engagement and adherence. Participants using WM or other more complex therapeutic techniques and those with a higher level of depression symptoms at baseline may need additional support from an e-therapist. Finally, engaging in user-centered design processes (ie, involving users in future system development [92]) may represent a way to improve intervention engagement and adherence [59] by improving navigation as well as further tailoring the intervention to the needs and context of parents of children treated for cancer.

Acknowledgments

We wish to thank Ian Horne (Portal Team Member) for performing all data extractions on the Portal.

Data Availability

The data that support the findings of this study are available from the corresponding author (LvE) upon reasonable request.

Conflicts of Interest

None declared.

References

1. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 2019;18(1):20-28 [FREE Full text] [doi: 10.1002/wps.20610] [Medline: 30600624]
2. Mamukashvili-Delau M, Koburger N, Dietrich S, Rummel-Kluge C. Efficacy of computer- and/or internet-based cognitive-behavioral guided self-management for depression in adults: a systematic review and meta-analysis of randomized controlled trials. BMC Psychiatry 2022;22(1):730 [FREE Full text] [doi: 10.1186/s12888-022-04325-z] [Medline: 36424570]
3. Eilert N, Enrique A, Wogan R, Mooney O, Timulak L, Richards D. The effectiveness of Internet-delivered treatment for generalized anxiety disorder: an updated systematic review and meta-analysis. Depress Anxiety 2021;38(2):196-219 [FREE Full text] [doi: 10.1002/da.23115] [Medline: 33225589]
4. Pauley D, Cuijpers P, Papola D, Miguel C, Karyotaki E. Two decades of digital interventions for anxiety disorders: a systematic review and meta-analysis of treatment effectiveness. Psychol Med 2023;53(2):567-579 [FREE Full text] [doi: 10.1017/S0033291721001999] [Medline: 34047264]
5. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FMG, Furukawa TA, Cuijpers P, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 2021;78(4):361-371 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
6. Hedman-Lagerlöf E, Carlbring P, Svärdman F, Riper H, Cuijpers P, Andersson G. Therapist-supported Internet-based cognitive behaviour therapy yields similar effects as face-to-face therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. World Psychiatry 2023;22(2):305-314 [FREE Full text] [doi: 10.1002/wps.21088] [Medline: 37159350]
7. Johnsen HM, Haddeland K. User evaluation of a therapist-guided internet-delivered treatment program for anxiety disorders: a qualitative study. Internet Interv 2021;25:100389 [FREE Full text] [doi: 10.1016/j.invent.2021.100389] [Medline: 33996506]
8. Thapar S, Nguyen M, Khan BN, Fanaieyan R, Kishimoto V, Liu R, et al. Patient and therapist perceptions of a publicly funded internet-based cognitive behavioral therapy (iCBT) program for ontario adults during the covid-19 pandemic: qualitative study. JMIR Form Res 2024;8:e50113 [FREE Full text] [doi: 10.2196/50113] [Medline: 38373027]
9. Doukani A, Cerga Pashoja A, Fanaj N, Qirjako G, Meksi A, Mustafa S, et al. Organizational readiness for implementing an internet-based cognitive behavioral therapy intervention for depression across community mental health services in Albania and Kosovo: directed qualitative content analysis. JMIR Form Res 2021;5(11):e29280 [FREE Full text] [doi: 10.2196/29280] [Medline: 34723822]
10. Svärdman F, Sjöwall D, Lindsäter E. Internet-delivered cognitive behavioral interventions to reduce elevated stress: a systematic review and meta-analysis. Internet Interv 2022;29:100553 [FREE Full text] [doi: 10.1016/j.invent.2022.100553] [Medline: 35781929]

XSL•FO
RenderX

11. Baltierra NB, Muessig KE, Pike EC, LeGrand S, Bull SS, Hightow-Weidman LB. More than just tracking time: complex measures of user engagement with an internet-based health promotion intervention. J Biomed Inform 2016;59:299-307 [FREE Full text] [doi: 10.1016/j.jbi.2015.12.015] [Medline: 26732997]
12. Sanatkar S, Baldwin PA, Huckvale K, Clarke J, Christensen H, Harvey S, et al. Using cluster analysis to explore engagement and e-attainment as emergent behavior in electronic mental health. J Med Internet Res 2019;21(11):e14728 [FREE Full text] [doi: 10.2196/14728] [Medline: 31778115]
13. O'Brien HL, Roll I, Kampen A, Davoudi N. Rethinking (Dis)engagement in human-computer interaction. Comput. Hum. Behav 2022;128(3):107109. [doi: 10.1016/j.chb.2021.107109]
14. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
15. Yeager CM, Benight CC. If we build it, will they come? Issues of engagement with digital health interventions for trauma recovery. Mhealth 2018;4:37 [FREE Full text] [doi: 10.21037/mhealth.2018.08.04] [Medline: 30363749]
16. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
17. Flett JAM, Fletcher BD, Riordan BC, Patterson T, Hayne H, Conner TS. The peril of self-reported adherence in digital interventions: a brief example. Internet Interv 2019;18:100267 [FREE Full text] [doi: 10.1016/j.invent.2019.100267] [Medline: 31890620]
18. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017;19(12):e402. [doi: 10.2196/jmir.8578] [Medline: 29212630]
19. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017;74(4):351-359 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
20. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
21. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
22. Karin E, Dear BF, Heller GZ, Crane MF, Titov N. "Wish You Were Here": examining characteristics, outcomes, and statistical solutions for missing cases in web-based psychotherapeutic trials. JMIR Ment Health 2018;5(2):e22 [FREE Full text] [doi: 10.2196/mental.8363] [Medline: 29674311]
23. Forbes A, Keleher MR, Venditto M, DiBiasi F. Assessing patient adherence to and engagement with digital interventions for depression in clinical trials: systematic literature review. J Med Internet Res 2023;25:e43727 [FREE Full text] [doi: 10.2196/43727] [Medline: 37566447]
24. Kelders SM, van Zyl LE, Ludden GDS. The concept and components of engagement in different domains applied to eHealth: a systematic scoping review. Front Psychol 2020;11:926 [FREE Full text] [doi: 10.3389/fpsyg.2020.00926] [Medline: 32536888]
25. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
26. Miller S, Yardley L, Smith P, Weal M, Anderson A, Stuart B, et al. A digital intervention for respiratory tract infections (Internet Dr): process evaluation to understand how to support self-care for minor ailments. JMIR Form Res 2022;6(1):e24239 [FREE Full text] [doi: 10.2196/24239] [Medline: 35044317]
27. Morrison LG, Geraghty AW, Lloyd S, Goodman N, Michaelides DT, Hargood C, et al. Comparing usage of a web and app stress management intervention: an observational study. Internet Interv 2018;12:74-82 [FREE Full text] [doi: 10.1016/j.invent.2018.03.006] [Medline: 30135771]
28. Enrique A, Palacios JE, Ryan H, Richards D. Exploring the relationship between usage and outcomes of an internet-based intervention for individuals with depressive symptoms: secondary analysis of data from a randomized controlled trial. J Med Internet Res 2019;21(8):e12775 [FREE Full text] [doi: 10.2196/12775] [Medline: 31373272]
29. Miller S, Ainsworth B, Yardley L, Milton A, Weal M, Smith P, et al. A framework for analyzing and measuring usage and engagement data (AMUsED) in digital interventions: viewpoint. J Med Internet Res 2019;21(2):e10966 [FREE Full text] [doi: 10.2196/10966] [Medline: 30767905]
30. Eldridge SM, Lancaster GA, Campbell MJ, Thabane L, Hopewell S, Coleman CL, et al. Defining feasibility and pilot studies in preparation for randomised controlled trials: development of a conceptual framework. PLoS One 2016;11(3):e0150205 [FREE Full text] [doi: 10.1371/journal.pone.0150205] [Medline: 26978655]

XSL•FO
RenderX

31. Thiblin E, Woodford J, Reuther C, Lundgren J, Lutvica N, von Essen L. Internet-administered, low-intensity cognitive behavioral therapy for parents of children treated for cancer: a feasibility trial (ENGAGE). Cancer Med 2023;12(5):6225-6243 [FREE Full text] [doi: 10.1002/cam4.5377] [Medline: 36404407]
32. Woodford J, Wikman A, Cernvall M, Ljungman G, Romppala A, Grönqvist H, et al. Study protocol for a feasibility study of an internet-administered, guided, CBT-based, self-help intervention (ENGAGE) for parents of children previously treated for cancer. BMJ Open 2018;8(6):e023708 [FREE Full text] [doi: 10.1136/bmjopen-2018-023708] [Medline: 29903802]
33. Ljungman L, Boger M, Ander M, Ljótsson B, Cernvall M, von Essen L, et al. Impressions that last: particularly negative and positive experiences reported by parents five years after the end of a child's successful cancer treatment or death. PLoS One 2016;11(6):e0157076 [FREE Full text] [doi: 10.1371/journal.pone.0157076] [Medline: 27272318]
34. Ljungman L, Cernvall M, Ghaderi A, Ljungman G, von Essen L, Ljótsson B. An open trial of individualized face-to-face cognitive behavior therapy for psychological distress in parents of children after end of treatment for childhood cancer including a cognitive behavioral conceptualization. PeerJ 2018;6:e4570 [FREE Full text] [doi: 10.7717/peerj.4570] [Medline: 29666751]
35. Wikman A, Kukkola L, Börjesson H, Cernvall M, Woodford J, Grönqvist H, et al. Development of an internet-administered cognitive behavior therapy program (ENGAGE) for parents of children previously treated for cancer: participatory action research approach. J Med Internet Res 2018;20(4):e133 [FREE Full text] [doi: 10.2196/jmir.9457] [Medline: 29669710]
36. Woodford J, Norbäck K, Hagström J, Grönqvist H, Parker A, Arundel C, et al. Study within a trial (SWAT) protocol. Investigating the effect of personalised versus non-personalised study invitations on recruitment: an embedded randomised controlled recruitment trial. Contemp Clin Trials Commun 2020;18:100572 [FREE Full text] [doi: 10.1016/j.conctc.2020.100572] [Medline: 32420511]
37. Woodford J, Wikman A, Einhorn K, Cernvall M, Grönqvist H, Romppala A, et al. Attitudes and preferences toward a hypothetical trial of an internet-administered psychological intervention for parents of children treated for cancer: web-based survey. JMIR Ment Health 2018;5(4):e10085 [FREE Full text] [doi: 10.2196/10085] [Medline: 30563814]
38. Woodford J, Farrand P, Hagström J, Hedenmalm L, von Essen L. Internet-administered cognitive behavioral therapy for common mental health difficulties in parents of children treated for cancer: intervention development and description study. JMIR Form Res 2021;5(7):e22709 [FREE Full text] [doi: 10.2196/22709] [Medline: 34142662]
39. Boucher EM, Raiker JS. Engagement and retention in digital mental health interventions: a narrative review. BMC Digit Health 2024;2(1):1-21. [doi: 10.1186/s44247-024-00105-9]
40. Seidler ZE, Rice SM, Ogrodniczuk JS, Oliffe JL, Dhillon HM. Engaging men in psychological treatment: a scoping review. Am J Mens Health 2018;12(6):1882-1900 [FREE Full text] [doi: 10.1177/1557988318792157] [Medline: 30103643]
41. Zimmer-Gembeck MJ, Rudolph J, Kerin J, Bohadana-Brown G. Parent emotional regulation: a meta-analytic review of its association with parenting and child adjustment. Int. J. Behav. Dev 2021;46(1):63-82. [doi: 10.1177/01650254211051086]
42. Ljungman L, Cernvall M, Grönqvist H, Ljótsson B, Ljungman G, von Essen L. Long-term positive and negative psychological late effects for parents of childhood cancer survivors: a systematic review. PLoS One 2014;9(7):e103340 [FREE Full text] [doi: 10.1371/journal.pone.0103340] [Medline: 25058607]
43. Ljungman L, Hovén E, Ljungman G, Cernvall M, von Essen L. Does time heal all wounds? A longitudinal study of the development of posttraumatic stress symptoms in parents of survivors of childhood cancer and bereaved parents. Psychooncology 2015;24(12):1792-1798 [FREE Full text] [doi: 10.1002/pon.3856] [Medline: 26042579]
44. Farrand P. Behavioural activation: working outside in, rather than inside out. In: Farrand P, editor. Low-Intensity CBT Skills and Interventions. Thousand Oaks, California, United States: SAGE Publications Ltd; 2020:1-408.
45. Small F, Lockwood K. Worry management: a practical solution to a problem of hypotheticals. In: Farrand P, editor. Low-Intensity CBT Skills and Interventions. Thousand Oaks, California, United States: SAGE Publications Ltd; 2020:1-408.
46. Miles G. Problem solving: doing what it says on the tin. In: Farrand P, editor. Low-Intensity CBT Skills and Interventions: A Practitioner's Manual. Thousand Oaks, California, United States: SAGE Publications Ltd; 2020:1-408.
47. Cuijpers P, Huibers M, Ebert DD, Koole SL, Andersson G. How much psychotherapy is needed to treat depression? A metaregression analysis. J Affect Disord 2013;149(1-3):1-13. [doi: 10.1016/j.jad.2013.02.030] [Medline: 23528438]
48. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
49. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24(1):2428.
50. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2025-01-24]
51. Farrand P, Woodford J. Impact of support on the effectiveness of written cognitive behavioural self-help: a systematic review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2013;33(1):182-195. [doi: 10.1016/j.cpr.2012.11.001] [Medline: 23238023]
52. Werntz A, Amado S, Jasman M, Ervin A, Rhodes JE. Providing human support for the use of digital mental health interventions: systematic meta-review. J Med Internet Res 2023;25:e42864 [FREE Full text] [doi: 10.2196/42864] [Medline: 36745497]

XSL•FO
RenderX

53. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
54. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
55. Armstrong RA. When to use the bonferroni correction. Ophthalmic Physiol Opt 2014;34(5):502-508. [doi: 10.1111/opo.12131] [Medline: 24697967]
56. Cohen J. A power primer. Psychol Bull 1992;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
57. Ekers D, Webster L, Van Straten A, Cuijpers P, Richards D, Gilbody S. Behavioural activation for depression; an update of meta-analysis of effectiveness and sub group analysis. PLoS One 2014;9(6):e100100 [FREE Full text] [doi: 10.1371/journal.pone.0100100] [Medline: 24936656]
58. Wang X, Feng Z. A narrative review of empirical literature of behavioral activation treatment for depression. Front Psychiatry 2022;13:845138 [FREE Full text] [doi: 10.3389/fpsyt.2022.845138] [Medline: 35546948]
59. Farrand P, Raue PJ, Ward E, Repper D, Areán P. Use and engagement with low-intensity cognitive behavioral therapy techniques used within an app to support worry management: content analysis of log data. JMIR Mhealth Uhealth 2024;12:e47321 [FREE Full text] [doi: 10.2196/47321] [Medline: 38029300]
60. Masterson C, Ekers D, Gilbody S, Richards D, Toner-Clewes B, McMillan D. Sudden gains in behavioural activation for depression. Behav Res Ther 2014;60:34-38. [doi: 10.1016/j.brat.2014.06.008] [Medline: 25049140]
61. Furukawa TA, Suganuma A, Ostinelli EG, Andersson G, Beevers CG, Shumake J, et al. Dismantling, optimising, and personalising internet cognitive behavioural therapy for depression: a systematic review and component network meta-analysis using individual participant data. Lancet Psychiatry 2021;8(6):500-511 [FREE Full text] [doi: 10.1016/S2215-0366(21)00077-8] [Medline: 33957075]
62. Andrade AQ, Beleigoli A, Diniz MDF, Ribeiro AL. Influence of baseline user characteristics and early use patterns (24-Hour) on long-term adherence and effectiveness of a web-based weight loss randomized controlled trial: latent profile analysis. J Med Internet Res 2021;23(6):e26421 [FREE Full text] [doi: 10.2196/26421] [Medline: 34081012]
63. Beintner I, Görlich D, Berger T, Ebert DD, Zeiler M, Herrero Camarano R, ICare Consortium. Interrelations between participant and intervention characteristics, process variables and outcomes in online interventions: a protocol for overarching analyses within and across seven clinical trials in ICare. Internet Interv 2019;16:86-97 [FREE Full text] [doi: 10.1016/j.invent.2018.05.001] [Medline: 30775268]
64. Packer T, Austin N, Lehman M, Douglas SL, Plow M. Factors influencing how informal caregivers of people with multiple sclerosis access and use a curated intervention website: analysis from an RCT. Digit Health 2024;10:20552076241228403 [FREE Full text] [doi: 10.1177/20552076241228403] [Medline: 38343587]
65. Solar C, Halat AM, MacLean RS, Rajeevan H, Williams DA, Krein SL, et al. Predictors of engagement in an internet-based cognitive behavioral therapy program for veterans with chronic low back pain. Transl Behav Med 2021;11(6):1274-1282. [doi: 10.1093/tbm/ibaa098] [Medline: 33098304]
66. Gasslander N, Alfonsson S, Jackalin A, Tengberg C, Håkansson J, Huotari L, et al. Predictors of adherence to an internet-based cognitive behavioral therapy program for individuals with chronic pain and comorbid psychological distress. BMC Psychol 2021;9(1):156 [FREE Full text] [doi: 10.1186/s40359-021-00663-x] [Medline: 34641946]
67. Wenger F, Allenhof C, Schreynemackers S, Hegerl U, Reich H. Use of Random Forest to Predict Adherence in an Online Intervention for Depression Using Baseline and Early Usage Data: Model Development and Validation on Retrospective Routine Care Log Data. JMIR Form Res 2024 Nov 15;8:e53768 [FREE Full text] [doi: 10.2196/53768] [Medline: 39546342]
68. Wu MS, Chen SY, Wickham RE, Leykin Y, Varra A, Chen C, et al. Predicting non-initiation of care and dropout in a blended care CBT intervention: Impact of early digital engagement, sociodemographic, and clinical factors. Digit Health 2022;8:20552076221133760 [FREE Full text] [doi: 10.1177/20552076221133760] [Medline: 36312847]
69. Wallert J, Gustafson E, Held C, Madison G, Norlund F, von Essen L, et al. Predicting adherence to internet-delivered psychotherapy for symptoms of depression and anxiety after myocardial infarction: machine learning insights from the u-care heart randomized controlled trial. J Med Internet Res 2018;20(10):e10754 [FREE Full text] [doi: 10.2196/10754] [Medline: 30305255]
70. Hoorelbeke K, Vervaeke J, Siegle GJ, Baeken C, Koster EH. Individual differences associated with treatment adherence and transfer effects following gamified web-based cognitive control training for repetitive negative thinking. Internet Interv 2022;27:100507 [FREE Full text] [doi: 10.1016/j.invent.2022.100507] [Medline: 35242588]
71. Lipschitz JM, Pike CK, Hogan TP, Murphy SA, Burdick KE. The engagement problem: a review of engagement with digital mental health interventions and recommendations for a path forward. Curr Treat Options Psychiatry 2023;10(3):119-135 [FREE Full text] [doi: 10.1007/s40501-023-00297-3] [Medline: 38390026]
72. Torous J, Michalak EE, O'Brien HL. Digital health and engagement-looking behind the measures and methods. JAMA Netw Open 2020;3(7):e2010918 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10918] [Medline: 32678446]
73. Kelders SM, Kip H, Greeff J. Psychometric Evaluation of the TWente Engagement with Ehealth Technologies Scale (TWEETS): Evaluation Study. J Med Internet Res 2020 Oct 09;22(10):e17757 [FREE Full text] [doi: 10.2196/17757] [Medline: 33021487]

XSL•FO
RenderX

74. Perski O, Lumsden J, Garnett C, Blandford A, West R, Michie S. Assessing the Psychometric Properties of the Digital Behavior Change Intervention Engagement Scale in Users of an App for Reducing Alcohol Consumption: Evaluation Study. J Med Internet Res 2019 Nov 20;21(11):e16197 [FREE Full text] [doi: 10.2196/16197] [Medline: 31746771]
75. Saleem M, Kühne L, De Santis KK, Christianson L, Brand T, Busse H. Understanding engagement strategies in digital interventions for mental health promotion: scoping review. JMIR Ment Health 2021;8(12):e30000 [FREE Full text] [doi: 10.2196/30000] [Medline: 34931995]
76. Yasukawa S, Tanaka T, Yamane K, Kano R, Sakata M, Noma H, et al. A chatbot to improve adherence to internet-based cognitive-behavioural therapy among workers with subthreshold depression: a randomised controlled trial. BMJ Ment Health 2024;27(1):e300881 [FREE Full text] [doi: 10.1136/bmjment-2023-300881] [Medline: 38199786]
77. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
78. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
79. Radomski AD, Bagnell A, Curtis S, Hartling L, Newton AS. Examining the usage, user experience, and perceived impact of an internet-based cognitive behavioral therapy program for adolescents with anxiety: randomized controlled trial. JMIR Ment Health 2020;7(2):e15795 [FREE Full text] [doi: 10.2196/15795] [Medline: 32022692]
80. Soobiah C, Cooper M, Kishimoto V, Bhatia RS, Scott T, Maloney S, et al. Identifying optimal frameworks to implement or evaluate digital health interventions: a scoping review protocol. BMJ Open 2020;10(8):e037643 [FREE Full text] [doi: 10.1136/bmjopen-2020-037643] [Medline: 32792444]
81. Nahum-Shani I, Shaw SD, Carpenter SM, Murphy SA, Yoon C. Engagement in digital interventions. Am Psychol 2022;77(7):836-852 [FREE Full text] [doi: 10.1037/amp0000983] [Medline: 35298199]
82. Tapp B, Gandy M, Fogliati VJ, Karin E, Fogliati RJ, Newall C, et al. Psychological distress, help‐seeking, and perceived barriers to psychological treatment among Australian parents. Aust. J. Psychol 2020;70(2):113-121. [doi: 10.1111/ajpy.12170]
83. Carlsson T, Kukkola L, Ljungman L, Hovén E, von Essen L. Psychological distress in parents of children treated for cancer: An explorative study. PLoS One 2019;14(6):e0218860 [FREE Full text] [doi: 10.1371/journal.pone.0218860] [Medline: 31226159]
84. Wakefield CE, Sansom-Daly UM, McGill BC, Hetherington K, Ellis SJ, Robertson EG, et al. Providing psychological support to parents of childhood cancer survivors: 'Cascade' intervention trial results and lessons for the future. Cancers (Basel) 2021;13(22):5597 [FREE Full text] [doi: 10.3390/cancers13225597] [Medline: 34830752]
85. Nordholt PU, Christalle E, Zill JM, Dirmaier J. Engagement with a web-based intervention to reduce harmful drinking: secondary analysis of a randomized controlled trial. J Med Internet Res 2020;22(11):e18826 [FREE Full text] [doi: 10.2196/18826] [Medline: 33216008]
86. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
87. Chen AT, Wu S, Tomasino KN, Lattie EG, Mohr DC. A multi-faceted approach to characterizing user behavior and experience in a digital mental health intervention. J Biomed Inform 2019;94:103187 [FREE Full text] [doi: 10.1016/j.jbi.2019.103187] [Medline: 31026595]
88. O'Brien H, Chen AT, Kaneshiro J, Zaslavsky O. User engagement in an online digital health intervention to promote problem solving. Interac Comput 2024;34(5):355-369.
89. SCB. Utrikes födda i Sverige. 2023. URL: https://www.scb.se/hitta-statistik/sverige-i-siffror/manniskorna-i-sverige/utrikes-fodda-i-sverige/ [accessed 2024-10-07]
90. Eustis EH, LoPresti J, Aguilera A, Schueller SM. Cultural responsivity in technology-enabled services: integrating culture into technology and service components. J Med Internet Res 2023;25:e45409 [FREE Full text] [doi: 10.2196/45409] [Medline: 37788050]
91. Sieverink F, Köhle N, Cheung K, Roefs A, Trompetter HR, Keizer J, et al. Evaluating eHealth. In: Gemert-Pijnen LV, Kelders SM, Kip H, Sanderman R, editors. eHealth research, theory and development: a multi-disciplinary approach. Abingdon: Routlegde; 2018:290-318.
92. Stawarz K, Preist C, Tallon D, Wiles N, Kessler D, Turner K, et al. Design considerations for the integrated delivery of cognitive behavioral therapy for depression: user-centered design study. JMIR Ment Health 2020;7(9):e15972 [FREE Full text] [doi: 10.2196/15972] [Medline: 32880580]

Abbreviations

AMUsED: Analyzing and Measuring Usage and Engagement Data
BA: behavioral activation
CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials

XSL•FO
RenderX

EJDeR: internetbaserad självhjälp för föräldrar till barn som avslutat en behandling mot cancer or internet-based self-help for parents of children who have completed cancer treatment
GAD: generalized anxiety disorder
GAD-7: 7-item Generalized Anxiety Disorder scale
iCBT: internet-administered cognitive behavioral therapy
IPE: introduction and psychoeducation
LICBT: low-intensity cognitive behavioral therapy
PHQ-9: 9-item Patient Health Questionnaire
RCT: randomized controlled trial
RP: relapse prevention
U-CARE: Uppsala University Psychosocial Care Programme
WM: worry management

Edited by A Mavragani; submitted 07.10.24; peer-reviewed by A Anderson; comments to author 05.11.24; revised version received 26.11.24; accepted 21.12.24; published 28.01.25.

Please cite as:
Reuther C, von Essen L, Mustafa MI, Saarijärvi M, Woodford J
Engagement With an Internet-Administered, Guided, Low-Intensity Cognitive Behavioral Therapy Intervention for Parents of Children Treated for Cancer: Analysis of Log-Data From the ENGAGE Feasibility Trial
JMIR Form Res 2025;9:e67171
URL: https://formative.jmir.org/2025/1/e67171
doi:10.2196/67171
PMID:

©Christina Reuther, Louise von Essen, Mudassir Imran Mustafa, Markus Saarijärvi, Joanne Woodford. Originally published in JMIR Formative Research (https://formative.jmir.org), 28.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Women Empowered to Connect With Addiction Resources and Engage in Evidence-Based Treatment (WE-CARE)—an mHealth Application for the Universal Screening of Alcohol, Substance Use, Depression, and Anxiety: Usability and Feasibility Study

Krystyna Isaacs[1], PhD; Autumn Shifflett[1], MS; Kajal Patel[1], BDS, MPH; Lacey Karpisek[2], LCSW, MSW, CPH, MPH; Yi Cui[1], MS; Maayan Lawental[2], MSW, PhD; Golfo Tzilos Wernette[3], PhD; Brian Borsari[4,5], PhD; Katie Chang[1], MS; Tony Ma[1], MS

[1]Benten Technologies, Manassas, VA, United States

[2]College of Behavior and Community Sciences, School of Social Work, University of South Florida, Tampa, FL, United States

[3]University of Michigan, Department of Family Medicine, Ann Arbor, MI, United States

[4]Center for Data to Discovery and Delivery Innovation, San Francisco Veterans Affairs Health Care System, San Francisco, CA, United States

[5]Department of Psychiatry and Behavioral Sciences, UCSF Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Krystyna Isaacs, PhD
Benten Technologies
9408 Grant Ave
Unit 206
Manassas, VA, 20110
United States
Phone: 1 7036625858
Email: krysisaacs@gmail.com

Abstract

Background: Women of childbearing age (aged 18-44 years) face multiple barriers to receiving screening and treatment for unhealthy alcohol and substance use, depression, and anxiety, including lack of screening in the primary care setting and lack of support in accessing care. The Women Empowered to Connect with Addiction Resources and Engage in Evidence-based Treatment (WE-CARE) mobile app was developed to test universal screening with women of childbearing age and linkage to care after an anonymous assessment.

Objective: In this study, we aimed to investigate the feasibility and acceptability of providing anonymous screening instruments through mobile phones for alcohol and substance use, as well as depression and anxiety, for women of childbearing age.

Methods: We used agile development principles based on previous formative research to test WE-CARE mobile health app with women of childbearing age (N=30) who resided in 1 of 6 counties in central Florida. WE-CARE included screening instruments (for alcohol, substance use, depression, and anxiety), a moderated discussion forum, educational microlearning videos, a frequently asked questions section, and resources for linkage to treatment. Individuals were recruited using flyers, academic listserves, and a commercial human subject recruiting company (Prolific). Upon completion of the screening instruments, women explored the educational and linkage to care features of the app and filled out a System Usability Scale to evaluate the mobile health app's usability and acceptability. Postpilot semistructured interviews (n=4) were conducted to further explore the women's reactions to the app.

Results: A total of 77 women downloaded the application and 30 completed testing. Women of childbearing age gave the WE-CARE app an excellent System Usability Scale score of 86.7 (SD 12.43). Our results indicate elevated risk for substance use in 18 of the 30 (60%) participants, 9/18 (50%) also had an elevated risk for anxiety or depression, and 11/18 (61%) had an elevated risk for substance use, anxiety, or depression. Participants reported that WE-CARE was easy to navigate and use but they would have liked to see more screening questions and more educational content. Linkage to care was an issue; however, as none of the women identified as "at-risk" for substance use disorders contacted the free treatment clinic for further evaluation.

XSL•FO
RenderX

Conclusions: The mobile health app was highly rated for acceptability and usability, but participants were not receptive to seeking help at a treatment center after only a few brief encounters with the app. The linkage to care design features was likely insufficient to encourage them to seek treatment. The next version of WE-CARE will include normative scores for participants to self-evaluate their screening status compared with their age- and gender-matched peers and enhanced linkages to care features. Future development will focus on enhancing engagement to improve change behaviors and assess readiness for change.

(JMIR Form Res 2025;9:e62915) doi:10.2196/62915

KEYWORDS

service linkage; digital health; education; mental health; substance use disorder; SUD; alcohol use disorder; chatbot; childbearing women; women; alcohol; substance use; empowerment; evidence-based treatment; usability; feasibility; mobile health; mhealth; app; depression; anxiety; screening; e-screening

Introduction

Alcohol and substance use can have devastating health consequences for women. The physical health risks to women who engage in heavy alcohol and substance use include increased likelihood of injury, overdose, organ damage, and STDs due to unsafe sex practices [1]. In addition, binge drinking (defined as 4 or more drinks on one occasion in women) has reached epic proportions (44%) in young women in the United States [2,3], and the rate is increasing rapidly in women over the age of 35 [2,3]. The number of women dying of opioid overdoses has also increased by 260% between 1999 and 2017 [4]. Substance use during pregnancy can also elevate the risk of miscarriage, fetal alcohol syndrome, and birth defects [5]. The US Preventative Task Force (USPTF) reported that about 50% of US pregnancies are unplanned or unintended, with women between the ages of 18-24 facing a higher risk of unplanned pregnancies [6]. As such, alcohol and substance use during pregnancy can pose a combination of risks and potential consequences for both the mother and baby, including fetal alcohol spectrum disorders [7,8].

While the physical health risks of alcohol and substance use are well-documented, it is equally important to recognize the associated risks of mental health illness. A systematic review revealed that 20%-40% of women with an alcohol use disorder also had an anxiety disorder and showed a strong association with major depressive disorder [9,10]. In addition, women who enter treatment for substance use disorder (SUD) frequently have a personal history of domestic violence, trauma, or sexual assault [11-13]. Domestic violence, trauma, or sexual assault can all have physical consequences, on an individual's physical health which can further impact mental health.

Universal screening for alcohol and substance use in health care settings is an effective way to identify individuals in need of help and support but is not widely implemented. Universal screening is defined as providing the same set of questions regarding alcohol and substance use to all individuals, regardless of past medical history or how the patient presents in a health care setting [14]. By implementing universal screening for all of their patients, including women of childbearing age, health care providers (HCPs) minimize the stigma associated with the inquiries. In 2016, the implementation of universal screening in primary care practices was less than 3% and less than 1% of the HCP subsequently engaged the patient in counseling [15]. Since then, USPTF has recommended universal screening for unhealthy alcohol use for all adults aged 18 years or older within primary care settings, followed by a brief conversation discussing the results [5]. This has resulted in more awareness and use of alcohol and drug screens in private practice settings, but as of 2020, private practice settings have still not met the goals set by the USPTF in 2018 [5,16]. USPTF specified the goals of their recommendation were to reduce overall rates of unhealthy alcohol use and improve individual health and social outcomes resulting from risky use through early intervention [17].

Challenges exist for implementing universal screening and supporting women of childbearing age who are at risk or with alcohol or SUD. For women of childbearing age, one such challenge included a lack of awareness of their risk of misuse as noted in the 2017 Surgeon General's *Facing Addiction Report* [18]. Women aware of personal risk may be reluctant to speak with their HCPs out of fear induced by stigma [19,20]. In some states, women are imprisoned for substance use during pregnancy or their children can be taken away [21]. Even if the women had concerns about personal risk and were looking to make an appointment, many were faced with long wait times to receive care [22]. The National Council for Mental Health Wellbeing revealed that 43% of adults were not able to engage in substance use treatment with 28% of those individuals not being able to get an appointment shortly after requesting care [23]. Rural patients may perceive additional barriers related to living in small, close-knit communities where the confidentiality of screener results is perceived as not being as secure as in an urban setting [24-27]. Anonymous self-screening solutions can overcome many of these barriers.

Challenges also exist for HCPs to implement universal screening and support those at the various stages of risk for alcohol or SUD. One common obstacle is a lack of awareness of recommended screening tools and training on tools and substance use. While 84% of HCPs reported asking young adults about their alcohol use, and 80% of obstetrician-gynecologists asked about alcohol or substance use as part of a visit for women of childbearing age, the majority of HCPs did not use a validated screening instrument [28] such as the Tobacco, Alcohol, Prescription medication, and Other Substance Use Tool (TAPS-1 [29] and the Substance Use Risk Profile–Pregnancy (SURP-P [30]), which are valid and effective in accurately identifying those at risk for substance use [31]. As a result, it is unclear if the risk for alcohol or substance use is being identified in these encounters. In addition, the majority of HCPs are not adequately

XSL•FO
RenderX

trained to administer, interpret results, and take appropriate action based on the results of the screeners [25,32]. In addition, providers in rural communities with these skills often have a heavier patient load than urban doctors and have less time for discussing the screener results, risks, and appropriate next steps [25].

Screening, Brief Intervention, and Referral to Treatment (SBIRT) can be an effective approach in the screening process and referral to treatment for individuals with alcohol or SUD [33]. SBIRT involves screening for alcohol and substance use for severity, followed by a brief intervention (eg, 10-15 minutes) and, if indicated, referral to treatment [33]. SBIRT's effectiveness in reducing alcohol use is well supported across a variety of settings [34]. For example, when SBIRT was implemented in a trauma outpatient clinic, 59% of the patients were identified as being at risk for substance use [33]. SBIRT promotes shared decision-making [35], hinging on clinicians building trust [36], using patient-centered language to reduce stigma [37], and exploring treatment options together. Within primary care settings, SBIRT can help providers identify those suffering from substance use or mental health challenges [38], despite the benefits challenges still exist in implementing SBIRT or other screenings, referrals, and linkage to care in all care settings [39] (eg, gynecologist offices) for women of childbearing age. However, the efficacy of SBIRT's "brief intervention" portion for drug use in primary care and emergency departments is still under consideration [40] with 1 systematic review revealing only moderate supporting evidence that interventions in the emergency department reduced alcohol-related injuries [41], another showing a reduction in visits to emergency departments following the SBIRT intervention [42], and another study showed mixed evidence surrounding the effectiveness of non–face-to-face computerized screening in the emergency department [43].

The brief intervention offered in SBIRT often incorporates motivational interviewing (MI) [44]. MI is a collaborative communication style that evokes and seeks to resolve personal ambivalence and strengthen personal commitment and motivation to change the behavior of interest [45]. The 2021 SAMHSA (Substance Abuse and Mental Health Services Administration) Advisory, *Using Motivational Interviewing in Substance Use Disorder Treatment*, emphasizes the need to discuss each individual's readiness to change and personal history [45]. Engaging clients through the use of MI-consistent skills (eg, reflections and affirmations) is a way to evoke personal goals and values consistent with behavior change, thus increasing intrinsic motivation to achieve a healthier lifestyle. However, learning high-quality MI skills can be labor-intensive [46,47], and as such MI is difficult to provide at scale in primary care settings.

Recently, MI has been used in multiple settings and with multiple populations, including inpatient, emergency rooms, rural settings [48], and other ethnicities to reduce alcohol or substance use [49-51] and MI through text messages has been well received by adults [52,53]. MI has also been applied through digital format using text messaging to address a variety of health behaviors including tuberculosis [54], smoking [55], and alcohol use [56]. Most recently, chatbots, which are conversational tools that frequently use artificial intelligence to provide a more engaging approach, have been used in apps (SoberGrid and CHESS) to support motivational interviewing to encourage action toward recovering from substance use [57-59], but as yet no apps exist that provide screening for SUD and alcohol and related risk factors to encourage change in risky behaviors. The use of electronic screeners (eg, assessments completed on a tablet) combined with motivational interviewing, can yield more accurate self-reports [60-63].

Digital health technology may be able to address many of these challenges. Smartphone ownership rates are high among women and individuals from diverse backgrounds, with 90% of women and 91%-97% of individuals who identify as White, Black, Asian or Hispanic backgrounds having access to a smartphone [64]. Mobile apps are therefore well suited to be used to screen for risk factors in pregnancy [65] and substance use [66] and they can play a crucial role in addressing stigma for those afraid of in-person treatment [67]. With issues related to stigma and fear addressed by the anonymous status, mobile apps can encourage more women to undergo screenings, leading to early identification of at-risk individuals, improved education, and increased treatment options for HCPs [68]. Novel digital strategies to improve mental health screenings, patient engagement, and referral to treatment show preliminary effectiveness [69]. A recent report found that using tablets in the waiting room resulted in nearly twice as many individuals screening for depression than when interviewed by nursing staff [70]. Treloar and colleagues highlight that universal screenings can help reduce stigma by reducing discrimination against certain people and groups [71]. As such, mobile health apps have the potential to support universal screening by alleviating stigma, increasing awareness for change through on-demand education, and reducing fear of seeking support and care.

This study was designed to determine whether women of childbearing age will complete anonymous screenings for substance use, alcohol, depression, and anxiety from their mobile phones. We built on previous formative research and online recruitment approaches, explored the benefits of educational and evidence-based materials and the use of linkage to care options, requested feedback on the usability and acceptability of the app, and used qualitative interviews to determine what future modifications need to be implemented before commercial release.

Methods

Digital Health Tool Design

The Women Empowered to Connect with Addiction Resources and Engage in Evidence-based Treatment (WE-CARE) app was designed to comprise a cross-platform (iOS or Android) mobile app where the participants were asked to complete multiple mental health screeners and then invited to explore each of the accompanying app features. All participants who were assessed as being "at-risk" for either alcohol use disorder or SUD were advised they could contact a treatment center anonymously or to call the center directly. An initial prototype was tested with 30 participants drawn from the central Florida region, recruited through listserves, a company that provides research participants

XSL•FO
RenderX

for a small fee, flyers, and community outreach events. Details of the final design of each of the features to be tested in the pilot are described below.

After extensive formative research [72], the following features were included for pilot testing in the mobile app: e-screenings for substance use, depression, and anxiety; anonymous self-referral to a local treatment center for further evaluation; a moderated forum for women to ask questions; a list of frequently asked questions (FAQs) to get evidence-based answers; educational videos on substance use and treatment-related information; national and local hotlines for those suffering from anxiety, depression, or a combination of the two; a chatbot to determine whether the women at risk had made an appointment with the partnered treatment center to receive a substance use evaluation; and follow up to the chatbot responses with motivational interviewing through a licensed social worker (Figure 1).

Figure 1. Screenshots of WE-CARE features tested in the pilot study. WE-CARE: Women Empowered to Connect with Addiction Resources and Engage in Evidence-based Treatment.

Features Developed

Screening Instruments

Multiple screening instruments to be incorporated in WE-CARE were reviewed, and subject matter experts were consulted to determine the most appropriate ones for testing. In total, 4 screeners were selected to be used within the app. Immediately after downloading the app, all participants had to take all the screeners: (1) SURP-P (if the woman self-identified as pregnant); (2) TAPS-1; (3) Generalized Anxiety Disorder-2 Scale (GAD-2) for anxiety; and (4) Patient Health Questionnaire-2 (PHQ-2) for depression (Textbox 1). A binary risk assessment (elevated risk or low risk) was immediately provided to the participants who could then store the results within the app or in the phone memory (refer to Multimedia Appendix 1). The women were reminded that all their responses would be kept anonymous.

Textbox 1. Tools used to screen for alcohol and substance use risk, as well as anxiety and depression.

Substance Use Risk Profile-Pregnancy [73]

- A 3-question tool designed for use in prenatal clinics with pregnant women to screen for alcohol, marijuana, illegal substances, and nonprescribed use of prescription medications.

Tobacco, Alcohol, Prescription Medication, and Other Substance Use Tool [74]

- Consists of a combined screening tool that assesses users for commonly used substances, eliminating the need for multiple screening and lengthy assessment tools.

Generalized Anxiety Disorder-2 Scale [75]

- A brief questionnaire to determine if someone needs additional screening for generalized anxiety.

Patient Health Questionnaire-2 [76]

- A simple 2-question screen to determine if someone should be considered for additional screening for depression.

If a participant was assessed as having elevated risk for alcohol or substance use, she was offered 4 choices: she could leave their phone number, but no personal information, so the treatment center could call back within 24-48 hours to provide a standardized intake, she could call a dedicated phone line and reach someone at the treatment center directly, she could call a national or local crisis hotline, or she could do nothing. If the woman was not at an elevated risk for substance or alcohol misuse but scored high for risk for depression or anxiety, she was presented with a series of crisis hotline numbers to call, as well as contact information for mental health treatment centers in one of the 6 counties of her choice.

Educational Microlearning Video Production

Short (1-5 min) animated cartoon videos were produced in Vyond [77] to provide information identified in the focus groups as pertinent to the women (Textbox 2). Each video topic was researched using websites at National Institute on Alcohol Abuse and Alcoholism (NIAAA), National Institute on Drug Abuse (NIDA), Substance Abuse and Mental Health Services

XSL•FO
RenderX

Administration (SAMHSA), and Centers for Disease Control and Prevention (CDC). A script involving 2 or more characters was then written to provide a more naturalistic setting and conversational tone and reviewed by subject matter experts for accuracy. A Vyond editor then worked with the producer to provide costumes and a setting tailored to Florida state demographics. To add the personalization desired by the women, the videos were presented in the mobile app based on how an individual answered the screeners, (eg, if women indicated they were at elevated risk for substance use, videos about the treatment process and general information on substance use were pushed to the top of the playlist). Several new videos were created for this specific project and additional videos were repurposed from other projects related to breastfeeding for women with SUD (Textbox 2).

Textbox 2. Eleven short, animated videos created for educational purposes.

Video topics

- What is universal screening for alcohol and drug use?
- Drug treatment center referral process.
- Information for women about marijuana use.
- Self-advocacy in a health care setting.
- Family planning.
- Breastfeeding-how to pump.
- Breastfeeding-opioid use disorder.
- Breastfeeding-marijuana use.
- Breastfeeding-tobacco cigarettes.
- Breastfeeding-prescription medications.
- Breastfeeding-alcohol use.

Moderated Forum

To ensure the app was providing evidence-based answers, a licensed social worker (LSW) was assigned to the moderated forum. Any participant with access to the app could make a post anonymously, but the forum was "moderated" in the sense that the LSW was scheduled to review posts each morning and would only publish ones that were relevant to the group. The LSW also could post responses to questions in the forum.

Frequently Asked Questions

The most popular FAQ topics and subtopics for women wanting information on substance and alcohol use were identified by querying ChatGPT 3 [78]. In the formative research stage [72], the women and HCPs expressed concern over the credibility of the information in both the moderated forum and the FAQs. As such, answers to the questions were manually compiled using source materials published at NIH, CDC, and SAMHSA, which were then reviewed by subject matter experts before being published in the FAQ section.

Follow-up by a Chatbot and Licensed Social Worker

A chatbot was built to engage with any individual evaluated as at-risk within 3-4 days of completing the screeners. The chatbot would ask if the person had made an appointment yet and if so, when? After discussion with the chatbot, if an appointment had not been made, the chatbot would encourage the participant to reach out to the treatment center and then a licensed social worker would follow up with the participant. The social worker would engage the participant in a manner consistent with motivational interviewing, to encourage her to make an appointment with the treatment center for further evaluation.

Additional Resources: National and Local Hotlines and Mental Health Facilities

Links to several national crisis lines were provided in the app such that the participant could click on the button and be immediately connected to the crisis line through her mobile phone. In addition, for those who self-identified as having either high anxiety or high levels of depression, a list of mental health centers, sorted by county, was provided within the app. All of the mental health centers were confirmed to be taking new patients before including them in the list.

Recruitment

Upon thorough in-house testing of all of these app features, participants were recruited to test the application over a 3-month period. Participants for the pilot study were eligible to participate in the study if they were women aged 18-44 years old, able to read English at the 8th-grade level, located within central Florida (1 of 6 counties served by the treatment center), and owned a smartphone device. In addition to recruitment through listserves and contacts in central Florida, a research assistant posted flyers within the community at health care agencies, barber shops, public libraries, and recreation centers. The third approach to recruitment included inviting individuals through Prolific [79], a commercial company that recruits individuals for research projects and surveys.

Demographic information was collected through an online survey instrument embedded in the app and all participants were asked to complete a checklist of the features to test within the app before receiving compensation for participation in the research study. At the end of the app testing, all participants

XSL•FO
RenderX

completed an online version of the System Usability Scale (SUS) and several open-ended questions.

Qualitative Feedback

Upon completion of the pilot study, participants were recruited by email or through direct messaging using their Prolific IDs to participate in a virtual 1:1 semistructured interview to gather more feedback on their experience with the mobile app. Interviewers were aware of whether the woman had given the app a low or high SUS score.

Ethical Considerations

The study was approved by the ADVARRA institutional review board (approval 00054640). All participants signed electronic informed consent documents before enrollment in the study. Participants were compensated for their time.

Results

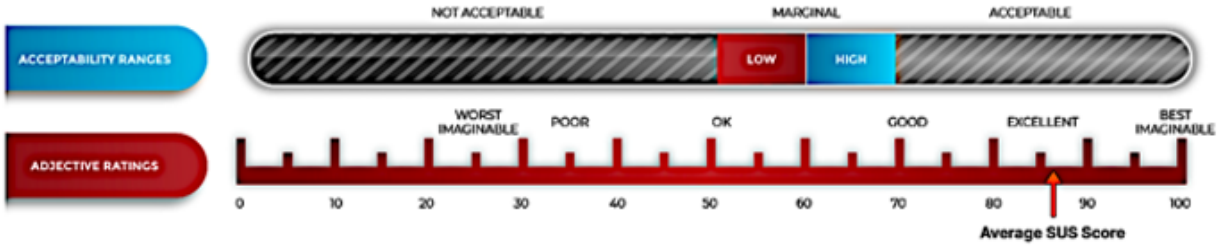
Demographics

A total of 77 women downloaded the app and a total of 30 women aged 18-44 years old tested the WE-CARE mobile app. The majority of the participants were white (47%) and not of Hispanic or Latino descent (56.7%). The participant racial composition was as follows: 46.7% White, 33.3% unknown, 16.7% mixed race, and 3.3% African American. In total, 20% were of Hispanic ethnicity., 56.7% were not Hispanic or Latino, and the remaining 23% were of unknown ethnicity.

Usability and Acceptability Score for Features Developed

WE-CARE was well received by participants who scored it as EXCELLENT (an average SUS score of 86.7, SD 12.43). Of the individuals who scored "at-risk" for substance use, the average SUS score was slightly lower, at 84.69 (SD 12.3) while the individuals who scored "not at risk" for substance use gave the app an average SUS score of 92.73 (SD 9.14), which would have been considered "BEST IMAGINABLE" (Figure 2).

Figure 2. System Usability Score results for WE-CARE pilot. WE-CARE: Women Empowered to Connect with Addiction Resources and Engage in Evidence-based Treatment.



Screening Results

While this was a usability and feasibility study, not a study designed or powered to identify risk for SUDs or mental health illness, we did examine the screening results collected in the pilot study. Out of the 30 individuals who completed the screeners, (18/30, 60%) scored at elevated risk for substance use or alcohol use behaviors, (9/18, 50%) scored at elevated risk for substance use as well as generalized anxiety, (9/18, 50%) scored elevated risk for substance use and depression, and (11/18, 61%) scored at elevated risk for substance use and either high risk for generalized anxiety or depression.

Digital Health Tool Usage Data

Microlearning Video Usage

All of the women had to watch the first video (Universal Screening for Alcohol and Drug Use) but reviewing the remaining videos was optional. Using the frequency of views, it is possible to rank the videos in terms of interest, with Self-Advocacy in the Healthcare Setting being the second most popular video with 6 views. As far as the videos were concerned, participants wanted the videos to be less robotic, shorter in length, and to have more videos relating to coping with stress. Participants were asked what additional information they would like to see included in the app, and 1 participant reported they wanted to learn: "Strategies and interventions that can be used at home …. (Breathing, mindfulness exercises)". Additional suggestions included more information on mental health issues, including "What can I do for fun or to relax instead of drinking or doing drugs? How do I tell if I have an anxiety disorder? What are some ways to reduce anxiety? How do I tell if I have a depressive disorder? What are some ways to alleviate depression?" as well as more information on self-advocacy in the doctor's office.

Moderated Forum Usage

All participants were required to engage with the forum (eg, write a post or give a reaction). The forum was preseeded with some sample questions and answers, and LIKES were scattered over a variety of posts, indicating the women were reviewing the other posts. One person made an anonymous post to which the licensed social worker then responded.

Frequently Asked Questions Usage

The women made extensive use of the FAQs, with the majority of the main topics being explored relating to depression, anxiety, or alcohol use in both the main topics. The most popular subtopics were "How do you define a standard drink" and "How to tell if I might have an anxiety disorder? (Refer to Multimedia Appendix 2: "Number of viewings for FAQ main topics" and Multimedia Appendix 3: "Number of viewings for FAQ

XSL•FO
RenderX

Subtopics" for a complete list of main topics and subtopics and their viewing frequencies).

Chatbot and LSW Follow-Up

Although half the individuals scored elevated risk for substance use, none of the women called or left their number with the treatment center or responded to the chatbot when it reached out to inquire whether they had made an appointment with someone at the treatment center. As such, the licensed social worker at the substance use treatment center was not able to follow up with any of the participants. During the postpilot focus group discussions, we learned one of the 4 individuals interviewed had reached out to a mental health treatment center within their area. As we were not able to track participant interactions with the mental health centers, we were not able to gauge how many of the women with at-risk scores for anxiety or depression sought additional help.

Qualitative Feedback

A total of 4 women completed the postpilot interviews. In total, 50% of the participants were White and 50% were of more than one race. 25% of the participants were Hispanic.

In addition to providing an objective rating of the app's usability and acceptability using the SUS scoring instrument, we asked several open-ended questions to get qualitative feedback on specific features of the application and conducted postpilot interviews. In general, the participants were complimentary about the app. One noted, "I think it's a very simplified app which I actually enjoy. It was a little nostalgic and brought me back to the earlier days of the Internet, especially with the forum included. I did think the screening tests were a little alarming because I wasn't aware I was so depressed. Overall, I enjoyed the app."

Several participants questioned whether asking just 2 questions on either anxiety or depression was sufficient. One participant commented, "I think there needs to be additional screening questions. The issues the app wants to tackle require more than the handful of questions asked." New features suggested by the participants included the ability to add individuals as friends through the app, making the FAQ interactive instead of a list of questions to pick from, and a person-to-person chat option.

When asked whether they had answered honestly when completing the screeners or had been "playing," all 4 participants in the postpilot focus group said they had answered based on their own behaviors. A majority of the participants were not surprised by their scores or alarmed and felt the questions were able to accurately assess their risk for substance or alcohol use, with 1 participant reporting, "would have preferred 4-5 questions and would have been good to ask follow up – for instance, in what circumstances are you using (socially, mental coping skills)…" Another suggestion made by one of the participants involved switching the order of the questions to start with screening for depression and anxiety.

The microlearning videos received more criticism with all participants noting a mix of live-action and cartoon videos would benefit more individuals. The topics of the videos were well received with none of the individuals noting a missed topic. One participant reported, "topics were really important (esp. drinking a glass of wine"). A couple of participants felt the length of the videos was sufficient but could be shortened further to 2-3 minutes.

The FAQs did not receive much interaction from the participants, although the participants did feel having the option to look through the FAQ was beneficial to help sort through a lot of information. On the other hand, participants who reported interacting with the moderated forum felt the feature would be helpful with some additional changes such as having a live professional to contact and adding visuals to the page to make the page more "homey."

To wrap up the discussion, questions were asked of the participants on how they viewed the motivational interviewing chatbot feature and their overall experience. The participants reported they didn't mind the feature being a chatbot, but if a phone call needed to be made or a person was experiencing an emergency, they believed a human should step in at that point. They also desired the chatbot to provide free follow-up services and to reiterate the individual's information will be kept confidential. Overall, participants felt they would recommend this app to another individual, one person reported, "Yes … It was marketed to the right population of people; low stakes in terms of divulging deepest darkest truths-allows a space for help that you might need."

Discussion

Principal Findings

The participants scored the app very highly in terms of usability and acceptability, and enthusiasm was high for the anonymous e-screenings to be offered through a mobile phone app. SAMHSA recommends the use of the SBIRT method to combat the problem of lack of universal screening [80] and this mobile health application successfully provided the screening component of SBIRT. The data collected during the formative research period indicates that WE-CARE can support the SBIRT method by providing universal screening to women aged 18-44 years old and educational materials and referrals to partnering treatment centers, but the prototype is not yet capable of replacing the "brief intervention" component. For example, despite providing access to a free licensed counselor at the treatment center, none of the women who participated in this study were ready to seek changes in their behavior as measured by calling the treatment center for further evaluation. One participant did mention in the postpilot focus group that they did reach out to a mental health facility to address their stress or anxiety, which shows the potential of the app to help women act. The digital handshake, or linkage to care, component of the app, therefore, was not used by the participants in the pilot study. After consultation with subject matter experts, it would appear that expecting the women to seek help after completing 5 minutes' worth of screeners on their phone was likely too high a "bar." Instead, these results suggest that additional engagement and support would be needed to encourage the women to seek more information or treatment. Plans are underway to redesign the prototype to offer more interactions with both a chatbot and

XSL•FO
RenderX

a human counselor to encourage women at risk to engage in the next steps of care.

The educational components of the app (videos, FAQs, and moderated forum) were well received. In the SUS open-ended questions and the postpilot group interviews, there was significant interest in expanding the educational materials to include additional information on coping skills and mental health (with videos, FAQs, PDFS, etc.). Other recommended additional features included creating a function to invite friends to download the app, adding a person-to-person chat option in the moderated forum, and making it possible to write in their questions for a chatbot to respond.

Given the women's willingness to complete the screeners on their mobile phones with guaranteed anonymity, it is possible that the women would be willing to complete the screeners in a physician's waiting room, if they could be assured of confidentiality [28,81-83]. Once a screener indicates a woman is at risk for an alcohol or SUD, the next hurdle is for the HCP to offer resources, make proper referrals, or provide suitable treatment options [84]. HCPs require training on how to effectively care for patients with alcohol or SUDs and need to understand the underlying reasons why women may engage in such behavior, such as past trauma [85]. Upon review of the screening results, the HCPs could create care plans and provide accessible mental health resources to patients based on their specific needs or refer their patients to a current list of clinics accepting new patients. If the latter option, HCPs then need to share with the patient what her journey with a care plan will look like. However, medical school training delivers minimal education in the area of substance use diagnosis and treatment [32] and HCPs have repeatedly noted they feel unprepared to offer SUD treatment or counseling [25,32]. Nonphysician or alternate health care models may be viable approaches to screening [86] and treatment and digital health technologies such as WE-CARE can help to fill in the time and knowledge gaps in the health care workforce.

In this study, over half of the women who completed the screeners in the WE-CARE app scored "at-risk" for SUD or alcohol use disorder and rates for risk for depression and anxiety in the respondents were 50% for each. The average rate for women of childbearing age meeting the *DSM-IV* (*Diagnostic and Statistical Manual of Mental Disorders* [Fourth Edition]) criteria of SUD or being in treatment in the past year for a SUD in the US is 3% for illicit drug use, 1.1% for opioid use, and 7.4% for alcohol use [87]. Similarly, in a study using the same screeners used in WE-CARE (namely the PHQ-2 for depression and the GAD-2 for anxiety), rates of depression in women of childbearing age have been reported to vary from 10% to 27% in the United States (depending on geographical location, race, and ethnicity [88,89]) and to fluctuate based on COVID prevalence [90]. The rates for anxiety in adults over 18 years of age are equally variable, peaking in 2021 at the height of the COVID-19 pandemic, at 35% [91] but before the pandemic, typically averaging around 19% [92]. Rates for anxiety and depression would be expected to be significantly higher in women of childbearing age than in the general population but gender was not reported in this study [92].

The screeners in WE-CARE were selected to identify women who are at risk for risky substance use or mental illness, not to identify substance use or mental health disorders using the criteria of a *DSM-IV* diagnosis per se. But there is a chance these rates may not be inflated. This pilot study was completed in a community that had recently experienced a major hurricane that severely impacted the citizens in 6 counties just 9 months before the study. Increased rates of alcohol and drug use [93] as well as mental illness are not uncommon after natural disasters [94]. Finally, individuals completing the WE-CARE screeners were assured of complete anonymity and therefore the screener results were not verified by individual in-person assessments. When an individual completes an assessment to report risky behavior, it has been demonstrated that the use of assessments completed on a tablet can yield more accurate self-reports [60-63]. However, self-reporting of alcohol or drug use, or concerns with mental illness, may differ in accuracy compared with when a screener is completed in front of a medical professional. The respondent may be confused about the question wording or definitions of terms, or because it was a research study, they may not have given accurate responses. Given the excellent score for usability and acceptability of the app, plans are ongoing to conduct clinical trials where the anonymous screener results will be compared with results from screeners administered by medical professionals to confirm the validity of this approach.

Mobile health apps can facilitate referrals, deliver educational materials to patients and providers, and enhance communication between patients and providers. Universal e-screenings are suitable for delivery before, during, and after an appointment and could remove the barriers associated with HCP time, stigma, willingness, or ability [95] and therefore are well-suited to the mobile health application environment. Other tailored apps exist to assist HCPs in delivering health interventions [96-103] related to substance use [96,104-106] and mental health [97,99,102] that are feasible and acceptable, but more research must be conducted to determine if the screeners are valid when delivered by a mobile health app, and also to ensure that after screening, the woman receives sufficient guidance and support to seek treatment if needed [107,108].

Limitations

A key limitation of this study was the self-selection recruitment method used, which may have impacted the generalizability of the findings. Future studies will include samples from a larger, more diverse population. Furthermore, it is possible that only women who had concerns about their mental health opted to participate in the research study. Also, as discussed above, responses generated through self-report may be flawed if the respondent misunderstands the questions or doesn't answer truthfully. Future versions of this tool will be tested in general clinics, where all patients will be asked to download the app as part of their annual physical and to talk with their health care provider after completing the screeners.

Conclusions

Preliminary results on the usability and feasibility of the WE-CARE mobile health app in providing anonymous screenings for drug and alcohol use, as well as anxiety and

XSL•FO
RenderX

depression, are excellent. Participants readily completed the screeners and were enthused about the opportunity to learn more about a variety of topics tailored for women in their childbearing years. Further research is necessary to validate this approach and to determine how to encourage women who score at risk for significant substance use or mental illness to seek treatment.

Acknowledgments

This research was funded by the National Institute on Alcohol Abuse and Alcoholism (NIAAA; 1R43AA029646-01).

Conflicts of Interest

TM and KC are co-owners of Benten Technologies, the company designing this platform, and will eventually market the Women Empowered to Connect with Addiction Resources and Engage in Evidence-based Treatment (WE-CARE) mobile app. The remaining authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the Department of Veterans Affairs, or the United States Government.

Multimedia Appendix 1
Detailed review of screeners and how they were scored in the WE-CARE app. WE-CARE: Women Empowered to Connect with Addiction Resources and Engage in Evidence-based Treatment.
[DOCX File , 20 KB - formative_v9i1e62915_app1.docx]

Multimedia Appendix 2
Number of viewings for FAQ (frequently asked question) main topics.
[DOCX File , 16 KB - formative_v9i1e62915_app2.docx]

Multimedia Appendix 3
Number of viewings of FAQ (frequently asked question) subtopics.
[DOCX File , 15 KB - formative_v9i1e62915_app3.docx]

References

1. Schulte MT, Hser YI. Substance use and associated health conditions throughout the lifespan. Public Health Rev 2014;35(2). [doi: 10.1007/BF03391702] [Medline: 28366975]
2. McCaul ME, Roach D, Hasin DS, Weisner C, Chang G, Sinha R. Alcohol and women: a brief overview. Alcohol Clin Exp Res 2019;43(5):774-779 [FREE Full text] [doi: 10.1111/acer.13985] [Medline: 30779446]
3. Adams RS, McKetta SC, Jager J, Stewart MT, Keyes KM. Cohort effects of women's mid-life binge drinking and alcohol use disorder symptoms in the United States: impacts of changes in timing of parenthood. Addiction 2023;118(10):1932-1941. [doi: 10.1111/add.16262] [Medline: 37338343]
4. VanHouten JP, Rudd RA, Ballesteros MF, Mack KA. Drug overdose deaths among women aged 30-64 years - United States, 1999-2017. MMWR Morb Mortal Wkly Rep 2019;68(1):1-5 [FREE Full text] [doi: 10.15585/mmwr.mm6801a1] [Medline: 30629574]
5. Curry SJ, Krist AH, Owens DK, Barry MJ, Caughey AB, Davidson KW, et al. Screening and behavioral counseling interventions to reduce unhealthy alcohol use in adolescents and adults: US preventive services task force recommendation statement. JAMA 2018;320(18):1899-1909. [doi: 10.1001/jama.2018.16789] [Medline: 30422199]
6. Unintended pregnancy.: Centers for Disease Control and Prevention Reproductive Health; 2019. URL: https://www.cdc.gov/reproductivehealth/contraception/unintendedpregnancy/index.htm [accessed 2019-12-10]
7. Jacobson JL, Jacobson SW, Sokol RJ. Effects of prenatal exposure to alcohol, smoking, and illicit drugs on postpartum somatic growth. Alcohol Clin Exp Res 1994;18(2):317-323. [doi: 10.1111/j.1530-0277.1994.tb00020.x] [Medline: 8048733]
8. Sokol RJ, Delaney-Black V, Nordstrom B. Fetal alcohol spectrum disorder. JAMA 2003;290(22):2996-2999. [doi: 10.1001/jama.290.22.2996] [Medline: 14665662]
9. Smith JP, Randall CL. Anxiety and alcohol use disorders: comorbidity and treatment considerations. Alcohol Res 2012;34(4):414-431 [FREE Full text] [Medline: 23584108]
10. Kathryn Mchugh R, Weiss RD. Alcohol use disorder and depressive disorders. Alcohol Res 2019;40(1):arcr.v40.1.01 [FREE Full text] [doi: 10.35946/arcr.v40.1.01] [Medline: 31649834]
11. Covington SS. Women and addiction: a trauma-informed approach. J Psychoactive Drugs 2008;Suppl 5:377-385. [doi: 10.1080/02791072.2008.10400665] [Medline: 19248395]
12. ACOG Committee on Health Care for Underserved Women, American Society of Addiction Medicine. ACOG committee opinion No. 524: opioid abuse, dependence, and addiction in pregnancy. Obstet Gynecol 2012;119(5):1070-1076. [doi: 10.1097/AOG.0b013e318256496e] [Medline: 22525931]

XSL•FO
RenderX

13. Alcohol abuse and other sbustance use disorders: ethical issues in obstetric and gynecologic practice.: American College of Obstetricians and Gynecologists; 2015. URL: https://www.acog.org/Clinical-Guidance-and-Publications/Committee-Opinions/Committee-on-Ethics/Alcohol-Abuse-and-Other-Substance-Use-Disorders-Ethical-Issues-in-Obstetric-and-Gynecologic-Practice [accessed 2024-12-30]
14. Screening.: Columbia University Department of Psychiatry URL: https://www.columbiapsychiatry.org/research/research-areas/substance-use-disorders/education/sbirt/about/screening [accessed 2024-12-30]
15. Physical health and mental health.: Mental Health Foundation URL: https://www.mentalhealth.org.uk/explore-mental-health/a-z-topics/physical-health-and-mental-health [accessed 2024-05-23]
16. Woodward D, Wilens TE, Glantz M, Rao V, Burke C, Yule AM. A systematic review of substance use screening in outpatient behavioral health settings. Addict Sci Clin Pract 2023;18(1):18 [FREE Full text] [doi: 10.1186/s13722-023-00376-z] [Medline: 36967381]
17. Curry SJ, Krist AH, Owens DK, Barry MJ, Caughey AB, Davidson KW, et al. Screening and behavioral counseling interventions to reduce unhealthy alcohol use in adolescents and adults: US preventive services task force recommendation statement. JAMA 2018;320(18):1899-1909. [doi: 10.1001/jama.2018.16789] [Medline: 30422199]
18. Department of Health and Human Services Office of the Surgeon General.. Facing addiction in America. The Surgeon General?s Report on Alcohol, Drugs, and Health 2016. [Medline: 28252892]
19. Stone R. Pregnant women and substance use: fear, stigma, and barriers to care. Health Justice 2015;3(1). [doi: 10.1186/s40352-015-0015-5]
20. Muncan B, Walters SM, Ezell J, Ompad DC. "They look at us like junkies": influences of drug use stigma on the healthcare engagement of people who inject drugs in New York City. Harm Reduct J 2020;17(1):53 [FREE Full text] [doi: 10.1186/s12954-020-00399-8] [Medline: 32736624]
21. Substance abuse reporting and pregnancy: the role of the obstetrician-gynecologist.: American College of Obstetricians and Gynecologists; 2011. URL: https://www.acog.org/clinical/clinical-guidance/committee-opinion/articles/2011/01/substance-abuse-reporting-and-pregnancy-the-role-of-the-obstetrician-gynecologist [accessed 2024-12-30]
22. Redko C, Rapp RC, Carlson RG. Waiting time as a barrier to treatment entry: perceptions of substance users. J Drug Issues 2006;36(4):831-852 [FREE Full text] [doi: 10.1177/002204260603600404] [Medline: 18509514]
23. More than 4 in 10 U.S. adults who needed substance use and mental health care did not get treatment.: National Council for Mental Wellbeing; 2022. URL: https://www.thenationalcouncil.org/news/more-than-4-in-10-us-adults-who-needed-substance-use-and-mental-health-care-did-not-get-treatment/ [accessed 2024-12-30]
24. Saunders EC, Moore SK, Gardner T, Farkas S, Marsch LA, McLeman B, et al. Screening for substance use in rural primary care: a qualitative study of providers and patients. J Gen Intern Med 2019;34(12):2824-2832 [FREE Full text] [doi: 10.1007/s11606-019-05232-y] [Medline: 31414355]
25. Brems C, Johnson ME, Warner TD, Roberts LW. Barriers to healthcare as reported by rural and urban interprofessional providers. J Interprof Care 2006;20(2):105-118. [doi: 10.1080/13561820600622208] [Medline: 16608714]
26. Fortney J, Booth BM. Access to substance abuse services in rural areas. Recent Dev Alcohol 2001;15:177-197. [doi: 10.1007/978-0-306-47193-3_10] [Medline: 11449741]
27. Desjarlais-deKlerk K, Wallace JE. Instrumental and socioemotional communications in doctor-patient interactions in urban and rural clinics. BMC Health Serv Res 2013;13:261 [FREE Full text] [doi: 10.1186/1472-6963-13-261] [Medline: 23835062]
28. Anderson BL, Dang EP, Floyd RL, Sokol R, Mahoney J, Schulkin J. Knowledge, opinions, and practice patterns of obstetrician-gynecologists regarding their patients' use of alcohol. J Addict Med 2010;4(2):114-121. [doi: 10.1097/ADM.0b013e3181b95015] [Medline: 21769028]
29. TAPS: tobacco, alcohol, prescription medication, and other substance use tool.: National Institutes of Health URL: https://www.drugabuse.gov/taps/#/ [accessed 2024-12-30]
30. Coleman-Cowger VH, Oga EA, Peters EN, Trocin KE, Koszowski B, Mark K. Accuracy of three screening tools for prenatal substance use. Obstet Gynecol 2019;133(5):952-961 [FREE Full text] [doi: 10.1097/AOG.0000000000003230] [Medline: 30969217]
31. Using standardized tools.: IntellectAbility URL: https://replacingrisk.com/using-standardized-tools/ [accessed 2024-12-30]
32. Recommendations for curricular elements in substance use disorders training.: Substance Abuse and Mental Health Services Administration URL: https://www.samhsa.gov/medications-substance-use-disorders/provider-support-services/recommendations-curricular-elements-substance-use-disorders-training [accessed 2024-12-30]
33. TAP 33: systems-level implementation of screening, brief intervention, and referral to treatment (SBIRT).: Substance Abuse and Mental Health Services Administration; 2013. URL: https://library.samhsa.gov/product/tap-33-systems-level-implementation-screening-brief-intervention-and-referral-treatment [accessed 2024-12-30]
34. Solberg LI, Maciosek MV, Edwards NM. Primary care intervention to reduce alcohol misuse ranking its health impact and cost effectiveness. Am J Prev Med 2008;34(2):143-152. [doi: 10.1016/j.amepre.2007.09.035] [Medline: 18201645]
35. Implementing care for alcohol & other drug use in medical settings: an extension of SBIRT.: National Council for Behavioral Health; 2018. URL: https://www.thenationalcouncil.org/wp-content/uploads/2021/04/

XSL•FO
RenderX

Implementing_Care_for_Alcohol_and_Other_Drug_Use_In_Medical_Settings_-_An_Extension_of_SBIRT.pdf [accessed 2024-12-30]
36. Matthews EB, Savoy M, Paranjape A, Washington D, Hackney T, Galis D, et al. Shared decision making in primary care based depression treatment: communication and decision-making preferences among an underserved patient population. Front Psychiatry 2021;12:681165 [FREE Full text] [doi: 10.3389/fpsyt.2021.681165] [Medline: 34322040]
37. Phelan SM, Salinas M, Pankey T, Cummings G, Allen JP, Waniger A, et al. Patient and health care professional perspectives on stigma in integrated behavioral health: barriers and recommendations. Ann Fam Med 2023;21(Suppl 2):S56-S60 [FREE Full text] [doi: 10.1370/afm.2924] [Medline: 36849477]
38. Hargraves D, White C, Frederick R, Cinibulk M, Peters M, Young A, et al. Implementing SBIRT (Screening, brief intervention and referral to treatment) in primary care: lessons learned from a multi-practice evaluation portfolio. Public Health Rev 2017;38:31 [FREE Full text] [doi: 10.1186/s40985-017-0077-0] [Medline: 29450101]
39. Gomez E, Gyger M, Borene S, Klein-Cox A, Denby R, Hunt S, et al. Using SBIRT (Screen, Brief Intervention, and Referral Treatment) training to reduce the stigmatization of substance use disorders among students and practitioners. Subst Abuse 2023;17:11782218221146391 [FREE Full text] [doi: 10.1177/11782218221146391] [Medline: 36685721]
40. Simioni N, Rolland B, Cottencin O. Interventions for increasing alcohol treatment utilization among patients with alcohol use disorders from emergency departments: a systematic review. J Subst Abuse Treat 2015;58:6-15. [doi: 10.1016/j.jsat.2015.06.003] [Medline: 26206477]
41. Havard A, Shakeshaft A, Sanson-Fisher R. Systematic review and meta-analyses of strategies targeting alcohol problems in emergency departments: interventions reduce alcohol-related injuries. Addiction 2008;103(3):368-377. [doi: 10.1111/j.1360-0443.2007.02072.x] [Medline: 18190671]
42. Pringle JL, Kelley DK, Kearney SM, Aldridge A, Dowd W, Johnjulio W, et al. Screening, brief intervention, and referral to treatment in the emergency department: an examination of health care utilization and costs. Med Care 2018;56(2):146-152. [doi: 10.1097/MLR.0000000000000859] [Medline: 29256973]
43. Biroscak BJ, Pantalon MV, Dziura JD, Hersey DP, Vaca FE. Use of non-face-to-face modalities for emergency department screening, brief intervention, and referral to treatment (ED-SBIRT) for high-risk alcohol use: a scoping review. Subst Abus 2019;40(1):20-32 [FREE Full text] [doi: 10.1080/08897077.2018.1550465] [Medline: 30829126]
44. Miller W, Rollnick S. Motivational Interviewing: Helping People Change and Grow. Fourth. New York: The Guilford Press; 2024.
45. Advisory: using motivational interviewing in substance use disorder treatment (based on TIP 35).: Substance Abuse and Mental Health Services Administration; 2021. URL: https://store.samhsa.gov/product/advisory-using-motivational-interviewing-substance-use-disorder-treatment-based-tip-35 [accessed 2024-12-30]
46. Madson MB, Loignon AC, Lane C. Training in motivational interviewing: a systematic review. J Subst Abuse Treat 2009;36(1):101-109. [doi: 10.1016/j.jsat.2008.05.005] [Medline: 18657936]
47. Kaltman S, Tankersley A. Teaching motivational interviewing to medical students: a systematic review. Acad Med 2020;95(3):458-469. [doi: 10.1097/ACM.0000000000003011] [Medline: 31577585]
48. Dickerson D, Moore LA, Rieckmann T, Croy CD, Venner K, Moghaddam J, et al. Correlates of motivational interviewing use among substance use treatment programs serving American indians/Alaska Natives. J Behav Health Serv Res 2018;45(1):31-45 [FREE Full text] [doi: 10.1007/s11414-016-9549-0] [Medline: 28236017]
49. Martino S, Zimbrean P, Forray A, Kaufman JS, Desan PH, Olmstead TA, et al. Implementing motivational interviewing for substance misuse on medical inpatient units: a randomized controlled trial. J Gen Intern Med 2019;34(11):2520-2529 [FREE Full text] [doi: 10.1007/s11606-019-05257-3] [Medline: 31468342]
50. Ewing SF, Bryan AD, Dash GF, Lovejoy TI, Borsari B, Schmiege SJ. Randomized controlled trial of motivational interviewing for alcohol and cannabis use within a predominantly Hispanic adolescent sample. Exp Clin Psychopharmacol 2022;30(3):287-299 [FREE Full text] [doi: 10.1037/pha0000445] [Medline: 33749294]
51. Balbi AM, Gak AE, Kim ES, Park TD, Quinn JF, Colon MF, et al. Brief motivational interviewing for substance use by medical students is effective in the emergency department. J Emerg Med 2019;57(1):114-117. [doi: 10.1016/j.jemermed.2019.03.032] [Medline: 31027991]
52. Louwagie G, Kanaan M, Morojele NK, Van Zyl A, Moriarty AS, Li J, et al. Effect of a brief motivational interview and text message intervention targeting tobacco smoking, alcohol use and medication adherence to improve tuberculosis treatment outcomes in adult patients with tuberculosis: a multicentre, randomised controlled trial of the ProLife programme in South Africa. BMJ Open 2022;12(2):e056496 [FREE Full text] [doi: 10.1136/bmjopen-2021-056496] [Medline: 35165113]
53. Gex KS, Mun EY, Barnett NP, McDevitt-Murphy ME, Ruggiero KJ, Thurston IB, et al. A randomized pilot trial of a mobile delivered brief motivational interviewing and behavioral economic alcohol intervention for emerging adults. Psychol Addict Behav 2023;37(3):462-474 [FREE Full text] [doi: 10.1037/adb0000838] [Medline: 35482647]
54. Moriarty AS, Louwagie GM, Mdege ND, Morojele N, Tumbo J, Omole OB, et al. ImPROving TB outcomes by modifying LIFE-style behaviours through a brief motivational intervention followed by short text messages (ProLife): study protocol for a randomised controlled trial. Trials 2019;20(1):457 [FREE Full text] [doi: 10.1186/s13063-019-3551-9] [Medline: 31349850]

XSL•FO
RenderX

55. Saiyed A, Layton J, Borsari B, Cheng J, Kanzaveli T, Tsvetovat M, et al. Technology-assisted motivational interviewing: developing a scalable framework for promoting engagement with tobacco cessation using NLP and machine learning. Procedia Computer Science 2022;206:121-131. [doi: 10.1016/j.procs.2022.09.091]
56. Kazemi DM, Borsari B, Levine MJ, Li S, Lamberson KA, Matta LA. A Systematic review of the mHealth interventions to prevent alcohol and substance abuse. J Health Commun 2017;22(5):413-432 [FREE Full text] [doi: 10.1080/10810730.2017.1303556] [Medline: 28394729]
57. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence-based chatbots for promoting health behavioral changes: systematic review. J Med Internet Res 2023;25:e40789 [FREE Full text] [doi: 10.2196/40789] [Medline: 36826990]
58. Dulin P, Mertz R, Edwards A, King D. Contrasting a mobile app with a conversational chatbot for reducing alcohol consumption: randomized controlled pilot trial. JMIR Form Res 2022;6(5):e33037 [FREE Full text] [doi: 10.2196/33037] [Medline: 35576569]
59. Almusharraf F, Rose J, Selby P. Engaging unmotivated smokers to move toward quitting: design of motivational interviewing-based chatbot through iterative interactions. J Med Internet Res 2020;22(11):e20251 [FREE Full text] [doi: 10.2196/20251] [Medline: 33141095]
60. Marsch LA, Campbell A, Campbell C, Chen CH, Ertin E, Ghitza U, et al. The application of digital health to the assessment and treatment of substance use disorders: the past, current, and future role of the National drug abuse treatment clinical trials network. J Subst Abuse Treat 2020;112S:4-11 [FREE Full text] [doi: 10.1016/j.jsat.2020.02.005] [Medline: 32220409]
61. Frost H, Campbell P, Maxwell M, O'Carroll RE, Dombrowski SU, Williams B, et al. Effectiveness of motivational interviewing on adult behaviour change in health and social care settings: a systematic review of reviews. PLoS One 2018;13(10):e0204890 [FREE Full text] [doi: 10.1371/journal.pone.0204890] [Medline: 30335780]
62. DiClemente CC, Corno CM, Graydon MM, Wiprovnick AE, Knoblach DJ. Motivational interviewing, enhancement, and brief interventions over the last decade: a review of reviews of efficacy and effectiveness. Psychol Addict Behav 2017;31(8):862-887. [doi: 10.1037/adb0000318] [Medline: 29199843]
63. Lundahl BW, Burke BL. The effectiveness and applicability of motivational interviewing: a practice-friendly review of four meta-analyses. J Clin Psychol 2009;65(11):1232-1245. [doi: 10.1002/jclp.20638] [Medline: 19739205]
64. Mobile fact sheet.: Pew Research Center; 2024. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2024-12-30]
65. Gance-Cleveland B, Leiferman J, Aldrich H, Nodine P, Anderson J, Nacht A, et al. Using the technology acceptance model to develop StartSmart: mHealth for screening, brief intervention, and referral for risk and protective factors in pregnancy. J Midwifery Womens Health 2019;64(5):630-640. [doi: 10.1111/jmwh.13009] [Medline: 31347784]
66. Glass JE, McKay JR, Gustafson DH, Kornfield R, Rathouz PJ, McTavish FM, et al. Treatment seeking as a mechanism of change in a randomized controlled trial of a mobile health intervention to support recovery from alcohol use disorders. J Subst Abuse Treat 2017;77:57-66 [FREE Full text] [doi: 10.1016/j.jsat.2017.03.011] [Medline: 28476273]
67. Kim HM, Xu Y, Wang Y. Overcoming the mental health stigma through m-Health apps: results from the healthy minds study. Telemed J E Health 2022;28(10):1534-1540. [doi: 10.1089/tmj.2021.0418] [Medline: 35254886]
68. Renker PR. Breaking the barriers: the promise of computer-assisted screening for intimate partner violence. J Midwifery Womens Health 2008;53(6):496-503. [doi: 10.1016/j.jmwh.2008.07.017] [Medline: 18984505]
69. Jardine J, Bowman R, Doherty G. Digital interventions to enhance readiness for psychological therapy: scoping review. J Med Internet Res 2022;24(8):e37851 [FREE Full text] [doi: 10.2196/37851] [Medline: 36040782]
70. Miller DP, Foley KL, Bundy R, Dharod A, Wright E, Dignan M, et al. Universal screening in primary care practices by self-administered tablet vs nursing staff. JAMA Netw Open 2022;5(3):e221480 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.1480] [Medline: 35258581]
71. Treloar C, Cama E, Lancaster K, Brener L, Broady TR, Cogle A, et al. A universal precautions approach to reducing stigma in health care: getting beyond HIV-specific stigma. Harm Reduct J 2022;19(1):74 [FREE Full text] [doi: 10.1186/s12954-022-00658-w] [Medline: 35799296]
72. Shifflett A, Karpisek L, Patel K, Cui Y, Lawental M, Tzilos Wernette G, et al. Women Empowered to Connect with Addiction Resources and Engage in Evidence-based Treatment (WE-CARE): A Usability and Feasibility Study of a Mobile Health Application for the Universal Screening of Alcohol, Substance Use, Depression and Anxiety. JMIR Formative Research. Preprint posted online on December 09, 2024 2024 [FREE Full text] [doi: 10.2196/62915]
73. Yonkers KA, Gotman N, Kershaw T, Forray A, Howell HB, Rounsaville BJ. Screening for prenatal substance use: development of the substance use risk profile-pregnancy scale. Obstet Gynecol 2010;116(4):827-833 [FREE Full text] [doi: 10.1097/AOG.0b013e3181ed8290] [Medline: 20859145]
74. McNeely J, Wu LT, Subramaniam G, Sharma G, Cathers LA, Svikis D, et al. Performance of the tobacco, alcohol, prescription medication, and other substance use (TAPS) tool for substance use screening in primary care patients. Ann Intern Med 2016;165(10):690-699 [FREE Full text] [doi: 10.7326/M16-0317] [Medline: 27595276]
75. Hughes AJ, Dunn KM, Chaffee T, Bhattarai JJ, Beier M. Diagnostic and clinical utility of the GAD-2 for screening anxiety symptoms in individuals with multiple sclerosis. Arch Phys Med Rehabil 2018;99(10):2045-2049 [FREE Full text] [doi: 10.1016/j.apmr.2018.05.029] [Medline: 29964000]

XSL•FO
RenderX

76. Instrument: Patient Health Questionnaire-2 (PHQ-2).: NIDA CTN Common Data Elements URL: https://cde.nida.nih.gov/instrument/fc216f70-be8e-ac44-e040-bb89ad433387 [accessed 2024-12-30]
77. Vyond products: the effortless, all-in-one AI video creation platform.: Vyond URL: https://www.vyond.com/product/ [accessed 2024-12-30]
78. ChatGPT 4o mini.: ChatGPT URL: https://chat.openai.com [accessed 2024-12-30]
79. Quality data. From real people. For faster breakthroughs.: Prolific URL: https://www.prolific.com/ [accessed 2024-12-30]
80. Screening, Brief Intervention, and Referral to Treatment (SBIRT).: Substance Abuse and Mental Health Services Administration URL: https://www.samhsa.gov/sbirt [accessed 2024-12-30]
81. Barry KL, Blow FC, Willenbring ML, McCormick R, Brockmann LM, Visnic S. Use of alcohol screening and brief interventions
in primary care settings: implementation and barriers. Subst Abus 2004;25(1):27-36. [doi: 10.1300/J465v25n01_05] [Medline: 15201109]
82. Kokole D, Mercken L, Jané-Llopis E, Natera Rey G, Arroyo M, Medina P, et al. Perceived appropriateness of alcohol screening and brief advice programmes in colombia, mexico and peru and barriers to their implementation in primary health care - a cross-sectional survey. Prim Health Care Res Dev 2021;22:e4 [FREE Full text] [doi: 10.1017/S1463423620000675] [Medline: 33504413]
83. Im GY, Mellinger JL, Winters A, Aby ES, Lominadze Z, Rice J, et al. Provider attitudes and practices for alcohol screening, treatment, and education in patients with liver disease: A Survey From the American association for the study of liver diseases alcohol-associated liver disease special interest group. Clin Gastroenterol Hepatol 2021;19(11):2407-2416.e8 [FREE Full text] [doi: 10.1016/j.cgh.2020.10.026] [Medline: 33069880]
84. What is SBIRT?.: SBIRT URL: https://sites.education.miami.edu/sbirt/what-is-sbirt/ [accessed 2024-12-30]
85. Tsogia D, Copello A, Orford J. Entering treatment for substance misuse: a review of the literature. J. Mental Health 2001;10(5):481-499. [doi: 10.1080/09638230120041254]
86. Mertens JR, Chi FW, Weisner CM, Satre DD, Ross TB, Allen S, et al. Physician versus non-physician delivery of alcohol screening, brief intervention and referral to treatment in adult primary care: the ADVISe cluster randomized controlled implementation trial. Addict Sci Clin Pract 2015;10:26 [FREE Full text] [doi: 10.1186/s13722-015-0047-0] [Medline: 26585638]
87. Martin CE, Scialli A, Terplan M. Unmet substance use disorder treatment need among reproductive age women. Drug Alcohol Depend 2020;206:107679. [doi: 10.1016/j.drugalcdep.2019.107679] [Medline: 31740208]
88. Mental Health Disorder Statistics.: Johns Hopkins Medicine URL: https://www.hopkinsmedicine.org/health/wellness-and-prevention/mental-health-disorder-statistics [accessed 2024-12-30]
89. Lee B, Wang Y, Carlson SA, Greenlund KJ, Lu H, Liu Y, et al. National, state-level, and county-level prevalence estimates of adults aged ≥18 Years self-reporting a lifetime diagnosis of depression - United States, 2020. MMWR Morb Mortal Wkly Rep 2023;72(24):644-650 [FREE Full text] [doi: 10.15585/mmwr.mm7224a1] [Medline: 37318995]
90. Jia H, Guerin RJ, Barile JP, Okun AH, McKnight-Eily L, Blumberg SJ, et al. National and state trends in anxiety and depression severity scores among adults during the COVID-19 pandemic - United States, 2020-2021. MMWR Morb Mortal Wkly Rep 2021;70(40):1427-1432 [FREE Full text] [doi: 10.15585/mmwr.mm7040e3] [Medline: 34618798]
91. Vahratian A, Blumberg SJ, Terlizzi EP, Schiller JS. Symptoms of anxiety or depressive disorder and use of mental health care among adults during the COVID-19 pandemic - United States, August 2020-February 2021. MMWR Morb Mortal Wkly Rep 2021;70(13):490-494 [FREE Full text] [doi: 10.15585/mmwr.mm7013e2] [Medline: 33793459]
92. Any Anxiety Disorder.: National Institute of Mental Health URL: https://www.nimh.nih.gov/health/statistics/any-anxiety-disorder [accessed 2024-12-30]
93. Alexander AC, Ward KD. Understanding postdisaster substance use and psychological distress using concepts from the self-medication hypothesis and social cognitive theory. J Psychoactive Drugs 2018;50(2):177-186 [FREE Full text] [doi: 10.1080/02791072.2017.1397304] [Medline: 29125424]
94. Newnham EA, Mergelsberg ELP, Chen Y, Kim Y, Gibbs L, Dzidic PL, et al. Long term mental health trajectories after disasters and pandemics: a multilingual systematic review of prevalence, risk and protective factors. Clin Psychol Rev 2022;97:102203. [doi: 10.1016/j.cpr.2022.102203] [Medline: 36162175]
95. Morse BA, Hutchins E. Reducing complications from alcohol use during pregnancy through screening. J Am Med Womens Assoc (1972) 2000;55(4):225-240. [Medline: 10935357]
96. Gordon MS, Carswell SB, Schadegg M, Mangen K, Merkel K, Tangires S, et al. Avatar-assisted therapy: a proof-of-concept pilot study of a novel technology-based intervention to treat substance use disorders. Am J Drug Alcohol Abuse 2017;43(5):518-524. [doi: 10.1080/00952990.2017.1280816] [Medline: 28635345]
97. Gaggioli A, Mantovani F, Castelnuovo G, Wiederhold B, Riva G. Avatars in clinical psychology: a framework for the clinical use of virtual humans. Cyberpsychol Behav 2003;6(2):117-125. [doi: 10.1089/109493103321640301] [Medline: 12804023]
98. Galvão Gomes da Silva J, Kavanagh DJ, Belpaeme T, Taylor L, Beeson K, Andrade J. Experiences of a motivational interview delivered by a robot: qualitative study. J Med Internet Res 2018;20(5):e116 [FREE Full text] [doi: 10.2196/jmir.7737] [Medline: 29724701]

XSL•FO
RenderX

99. Baylor AL. Promoting motivation with virtual agents and avatars: role of visual presence and appearance. Philos Trans R Soc Lond B Biol Sci 2009;364(1535):3559-3565 [FREE Full text] [doi: 10.1098/rstb.2009.0148] [Medline: 19884150]
100. Baccon LA, Chiarovano E, MacDougall HG. Virtual reality for teletherapy: avatars may combine the benefits of face-to-face communication with the anonymity of online text-based communication. Cyberpsychol Behav Soc Netw 2019;22(2):158-165. [doi: 10.1089/cyber.2018.0247] [Medline: 30540490]
101. Dworkin MS, Lee S, Chakraborty A, Monahan C, Hightow-Weidman L, Garofalo R, et al. Acceptability, feasibility, and preliminary efficacy of a theory-based relational embodied conversational agent mobile phone intervention to promote HIV medication adherence in young HIV-positive African American MSM. AIDS Educ Prev 2019;31(1):17-37. [doi: 10.1521/aeap.2019.31.1.17] [Medline: 30742481]
102. Rehm IC, Foenander E, Wallace K, Abbott JM, Kyrios M, Thomas N. What role can avatars play in e-mental health interventions? Exploring new models of client-therapist interaction. Front Psychiatry 2016;7:186 [FREE Full text] [doi: 10.3389/fpsyt.2016.00186] [Medline: 27917128]
103. Canidate S, Hart M. The use of avatar counseling for HIV/AIDS health education: the examination of self-identity in avatar preferences. J Med Internet Res 2017;19(12):e365 [FREE Full text] [doi: 10.2196/jmir.6740] [Medline: 29196281]
104. Tzilos GK, Sokol RJ, Ondersma SJ. A randomized phase I trial of a brief computer-delivered intervention for alcohol use during pregnancy. J Womens Health (Larchmt) 2011;20(10):1517-1524 [FREE Full text] [doi: 10.1089/jwh.2011.2732] [Medline: 21823917]
105. Martino S, Ondersma SJ, Forray A, Olmstead TA, Gilstad-Hayden K, Howell HB, et al. A randomized controlled trial of screening and brief interventions for substance misuse in reproductive health. Am J Obstet Gynecol 2018;218(3):322.e1 [FREE Full text] [doi: 10.1016/j.ajog.2017.12.005] [Medline: 29247636]
106. Ondersma SJ, Beatty JR, Svikis DS, Strickler RC, Tzilos GK, Chang G, et al. Computer-delivered screening and brief intervention for alcohol use in pregnancy: a pilot randomized trial. Alcohol Clin Exp Res 2015;39(7):1219-1226 [FREE Full text] [doi: 10.1111/acer.12747] [Medline: 26010235]
107. Terplan M. Commentary on ondersma et al. (2012): beyond the quest for the perfect test-drug use screening in pregnancy. Addiction 2012;107(12):2107-2108. [doi: 10.1111/j.1360-0443.2012.04094.x] [Medline: 23167646]
108. Pace CA, Gergen-Barnett K, Veidis A, D'Afflitti J, Worcester J, Fernandez P, et al. Warm handoffs and attendance at initial integrated behavioral health appointments. Ann Fam Med 2018;16(4):346-348 [FREE Full text] [doi: 10.1370/afm.2263] [Medline: 29987084]

Abbreviations

AUD: alcohol use disorder
CDC: Centers for Disease Control and Prevention
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition)
FAQ: frequently asked question
GAD-2: Generalized Anxiety Disorder-2 Scale
HCP: health care provider
LSW: licensed social worker
MI: motivational interviewing
NIAAA: National Institute on Alcohol Abuse and Alcoholism
NIDA: National Institute on Drug Abuse
PHQ-2: Patient Health Questionnaire-2
SAMHSA: Substance Abuse and Mental Health Services Administration
SBIRT: Screening, Brief Intervention, and Referral to Treatment
SUD: substance use disorder
SURP-P: Substance Use Risk Profile–Pregnancy
SUS: System Usability Scale
TAPS-1: Tobacco, Alcohol, Prescription medication, and Other Substance Use Tool
USPTF: United States Preventative Task Force
WE-CARE: Women Empowered to Connect with Addiction Resources and Engage in Evidence-based Treatment

XSL•FO
RenderX

Edited by A Mavragani; submitted 11.06.24; peer-reviewed by R Yin; comments to author 18.11.24; revised version received 07.12.24; accepted 09.12.24; published 07.02.25.

Please cite as:
Isaacs K, Shifflett A, Patel K, Karpisek L, Cui Y, Lawental M, Tzilos Wernette G, Borsari B, Chang K, Ma T
Women Empowered to Connect With Addiction Resources and Engage in Evidence-Based Treatment (WE-CARE)—an mHealth Application for the Universal Screening of Alcohol, Substance Use, Depression, and Anxiety: Usability and Feasibility Study
JMIR Form Res 2025;9:e62915
URL: https://formative.jmir.org/2025/1/e62915
doi:10.2196/62915
PMID:39918861

©Krystyna Isaacs, Autumn Shifflett, Kajal Patel, Lacey Karpisek, Yi Cui, Maayan Lawental, Golfo Tzilos Wernette, Brian Borsari, Katie Chang, Tony Ma. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Solution to Support Medication Adherence and Self-Management in Patients with Cancer (SAMSON): Pilot Randomized Controlled Trial

Thu Ha Dang[1,2,3], MIPH, MD; Nilmini Wickramasinghe[4,5,6], MBA, PhD; Prem Prakash Jayaraman[7], PhD; Kate Burbury[8,9,10], PhD; Marliese Alexander[10,11], PhD; Ashley Whitechurch[10,12], MNS; Mitchell Dyer[11], MClinPharm; Stephen Quinn[13], PhD; Abdur Rahim Mohammad Forkan[14], PhD; Penelope Schofield[10,15,16], BSc(Hons), PhD

[1]Department of Psychological Sciences, School of Health Sciences, Swinburne University of Technology, Melbourne, Australia

[2]Health Services Research and Implementation Sciences, Peter MacCallum Cancer Centre, Melbourne, Australia

[3]Digital Health Cooperative Research Centre, Sydney, Australia

[4]Optus Chair Digital Health, La Trobe University, Melbourne, Australia

[5]Department Health and Bio Statistics, School of Health Sciences and Iverson Health Innovation Research Institute, Swinburne University of Technology, Melbourne, Australia

[6]Epworth Healthcare, Melbourne, Australia

[7]Factory of the Future and Digital Innovation Lab, Department of Computer Science and Software Engineering, School Software and Electrical Engineering, Swinburne University of Technology, Melbourne, Australia

[8]Tasmanian Health Services, Department of Health, Hobart, Australia

[9]Digital and Healthcare Innovation, Peter McCallum Cancer Centre, Melbourne, Australia

[10]Sir Peter MacCallum Department of Oncology, The University of Melbourne, Melbourne, Australia

[11]Pharmacy Department, Peter MacCallum Cancer Centre, Melbourne, Australia

[12]Department of Clinical Haematology, Peter MacCallum Cancer Centre, Melbourne, Australia

[13]Department of Health Science and Biostatistics, Swinburne University of Technology, Melbourne, Australia

[14]Digital Innovation Lab, Department of Computer Science and Software Engineering, School Software and Electrical Engineering, Swinburne University of Technology, Melbourne, Australia

[15]School of Computing, Engineering and Mathematical Sciences, La Trobe University, Melbourne, Australia

[16]Department of Psychological Sciences and Iverson Health Innovation Research Institute, Swinburne University of Technology, Melbourne, Australia

Corresponding Author:
Thu Ha Dang, MIPH, MD
Department of Psychological Sciences
School of Health Sciences
Swinburne University of Technology
John Street
Hawthorn
Melbourne, 3122
Australia
Phone: 61 422703347
Email: thuhadang@swin.edu.au

Abstract

Background: Medication nonadherence is a serious problem in cancer, potentially impacts patients' health outcomes and health care costs. Although technology-based medication adherence (MA) interventions have emerged, evidence supporting their quality and effectiveness remains limited.

Objective: This study tested the acceptability, feasibility, and potential effects of Safety and Adherence to Medications and Self-care Advice in Oncology (SAMSON), a digital solution designed to support MA and self-management in cancer.

Methods: A 12-week, 2-arm, unblinded, pragmatic pilot randomized controlled trial was conducted. Adults with hematological malignancies who started oral cancer medicines within the last 12 months were recruited from a metropolitan specialized hospital and randomized 1:1 to SAMSON or control (usual care). The SAMSON solution included a smartphone app with tailored alerts

XSL•FO
RenderX

and real-time self-care advice, a web-based dashboard for health care professionals (HCPs) to monitor patients' adherence and symptoms, and motivational interviewing (MI) teleconsultations delivered by oncology nurses and pharmacists at baseline and weeks 1, 4, 8, and 12. Primary outcomes were the patients' acceptance of SAMSON, measured by the Unified Theory of Acceptance and Use of Technology at 12 weeks, and study feasibility, measured by predefined rates of recruitment, randomization, retention, intervention adherence, and outcome assessment completion. Secondary outcomes were comparison of MA and clinical self-assessments through online questionnaires, including adherence, toxicity self-management, anxiety and depression symptoms, and quality of life, measured at baseline and 12 weeks between the 2 arms. Data retrieved from the SAMSON app (Swinburne University of Technology) was analysed for task completion.

Results: A total of 33 patients (79% of those who were approached) consented to participate in the trial. Of those, 31/33 (94%) completed baseline surveys and were randomized to SAMSON (15/31) and control arms (16/31). Of 31 patients, 28 (90%) completed the 12-week surveys (12 SAMSON and 16 control). Overall, patients rated the SAMSON solution as highly acceptable (13/15, 87% app usage; 14/15, 93% MI teleconsultation delivery). They reported that SAMSON was easy to use (10/12, 83%) and helpful in improving their MA (6/12, 50%). All study HCPs reported the SAMSON solution was helpful in supporting patients' MA. Patients completed an average of 99 tasks over the 12-week study period (71% of scheduled tasks). Most patients (10/12, 83%) completed all 5 scheduled consultations. All study feasibility measures were higher than the predefined upper thresholds, except the rate of patients' responses to medication reminders.

Conclusions: The results demonstrated that the SAMSON solution is acceptable, usable, and useful for oncology HCPs and patients with cancer. The SAMSON solution is feasible in real-life oncology settings. Our next steps involve refining the SAMSON solution based on participants' feedback, conducting a large-scale randomized controlled trial to evaluate its clinical and economic effectiveness, and exploring potential commercialization.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12623000472673; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=385728

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2023-079122

(JMIR Form Res 2025;9:e65302) doi:10.2196/65302

KEYWORDS

home-based cancer treatment; smartphone app; oral chemotherapy; patient safety; SAMSON; mobile phone; digital solution; medication adherence; self-management; cancer; randomized controlled trial; RCT; pilot study; oncology; mobile health; mHealth; quality of life; eHealth

Introduction

Increasingly, cancer is being treated with self-administered medications [1,2], often involving long and complex treatment regimens [2-4]. A patient's ability to adhere to medications throughout the treatment period is central to the successful delivery of self-administered anticancer regimens [5,6]. However, medication adherence (MA) in cancer is low [7] and often decreases over time [8]. The MA rate is particularly low in hematological cancers, with a variation between 6% [9] and 53% [10]. Improving MA in patients with cancer is crucial, as evidence shows that nonadherence is associated with low survival rates, disease progression, as well as increased health care use and costs [6,11,12].

Given the importance of MA in cancer, there has been an increase in the number of MA interventions in recent times, especially digital interventions [13,14]. With the boom of technology, for instance, smartphones [15], digital health solutions promise more advantages in terms of improved clinical outcomes, and cost efficiency, and they are increasingly accepted by patients [16]. However, evidence on the quality and effectiveness of available MA interventions in cancer remains sparse [17,18]. Our most recent systematic review showed that MA interventions that have multicomponents that are theory- and evidence-based and rigorously designed and evaluated are more likely to be effective [18].

Based on findings from the literature review and the need to address the medication nonadherence problem in cancer, we developed Safety and Adherence to Medications and Self-care Advice in Oncology (SAMSON), a multicomponent digital solution to improve MA. Individual components of the SAMSON solution were co-designed and rigorously developed based on evidence and largely informed by behavioral science research and design science research methodology. The solution comprises 2 components. The first is a smartphone app (SAMSON app) involving individually tailored phone alerts and real-time advice for side effect self-management [19] and a web-based dashboard where patients can track their adherence performance and health care professionals (HCPs) can manage patients' profile, view their adherence performance and survey responses, and manage their medication schedules and related side effects. The second is a motivational interviewing training platform (MITP) to train HCPs in motivational interviewing (MI) techniques to support patient adherence and side effect self-management [20]. After being co-designed and developed, the SAMSON app was tested on end users, specifically patients with hematological cancer [19], and the MITP was tested on HCPs [20]. We hypothesized that the combination of these 2 components (the SAMSON solution) would be broadly acceptable to patients, practically feasible in a busy clinical practice, and potentially effective in improving MA for patients with cancer.

XSL•FO
RenderX

Methods

The CONSORT (Consolidated Standards of Reporting Trials) EHEALTH checklist V 1.6.1 [21] was used to report the present study (Multimedia Appendix 1).

Study Design

We aimed to test the acceptability, feasibility, and potential effects of the SAMSON solution on 30-50 patients with hematological cancer through a 2-armed, unblinded, 12-week, pragmatic pilot randomized controlled trial (RCT) [22]. After providing written consent (Multimedia Appendix 2) and completing baseline questionnaires (Multimedia Appendix 3), participants were randomized to either the intervention group (SAMSON solution) or the control group (usual care) at a 1:1 ratio. After 12 weeks, they completed end-of-study surveys (Multimedia Appendix 3).

Randomization

Using a computer-generated randomization chart, a permuted block randomization of size 4 was used to ensure an even balance of patients in each group throughout the study period. The allocation schedule was generated by a statistician who was blinded to the participants to prevent any predictability when randomizing participants to intervention or control [23].

Patients and Eligibility

Patients were recruited from a metropolitan specialized cancer hospital in Melbourne, Australia, between August 2023 and February 2024. Eligible patients were adults (more than 18 years old), diagnosed with hematological cancer, scheduled to commence oral anticancer medicines (OAMs) or commenced the medication for less than 12 months, willing to have OAMs dispensed at the hospital for the duration of the trial, able to communicate in English, and had access to the internet, a smartphone or computer, and telehealth.

Recruitment

Patients were identified by study site nurses, pharmacists, or treating consultants, who were informed about the study and eligible criteria. At their scheduled consultation, they were asked if they would like their details passed on for contact by the research team. If the patient agreed, they were referred to the study RC. Then, the RC contacts the patient for screening and a comprehensive informed consent process, either in person or online.

Intervention

SAMSON Solution

Patients allocated to the intervention group received the SAMSON solution in addition to their usual care at Peter MacCallum Cancer Centre (PMCC). They were instructed by the RC on how to install the SAMSON app on their smartphone and received login details with a protected password as well as the SAMSON app user manual. Patients' personal and clinical information, such as diagnosis and treatment, extracted from the hospital's electronic medical record system (EMR), was entered in the SAMSON web-based dashboard by the RC and made available in the smartphone app.

Patients were informed that they would receive a teleconsultation (either via phone or telehealth) from a hospital clinical pharmacist in the first 3 days after enrolling in the study, and a maximum of 4 follow-up teleconsultations from a hospital clinical nurse on weeks 1, 4, 8, and 12 of the study. These teleconsultations were on top of the usual care at the hospital. All teleconsultations followed predefined structures and were delivered by hospital clinical nurses and pharmacists who were previously trained in MI using the developed MITP. The initial consultation took 30-60 minutes, aiming to provide education on the OAM(s) that the patient received and the importance of adherence, support the patient in making decisions regarding their medication-taking schedule, and identify and document possible risks and barriers to MA.

Based on the agreed medication-taking schedule, individualized daily medication reminders and weekly side effect surveys were set up in the SAMSON backend platform using the web-based dashboard user interface, so patients can receive them in the installed smartphone app. Medicine information and side effect self-care advice, developed by experienced oncology pharmacists based on available reliable resources and clinician review, and approved by PMCC's Human Research Ethics Committees (HREC), were populated in the SAMSON backend platform using the web-based dashboard. Patients were asked to respond to daily medication reminders and weekly side effect surveys, as well as review self-care advice in the smartphone app. Data on patients' adherence and drug toxicity collected through the SAMSON solution were stored centrally on a secured server, then aggregated, analyzed, and uploaded onto the web-based dashboard so that study nurses and pharmacists could monitor patients' adherence and symptoms. These data were also used by HCPs to tailor their teleconsultations with patients. Patients used the SAMSON app throughout the 12-week period of the study.

The follow-up structured teleconsultations (15-30 minutes in length) aimed to check the patient's understanding of diagnosis, symptoms, self-care strategy, and medications; further explore the patient's facilitators and barriers to MA; motivate the patient's adherence, strengthen their medication self-management skills, and change patient's nonadherence behavior by using MI skills. The quantity and length of these consultations were tailored to the individual patient's need and adherence status.

Intervention nurses and pharmacists had more than 5 years of clinical experience in providing oncology care and successfully completed MI training via the MITP. They were also equipped with instruction manuals on how to conduct teleconsultation and use the SAMSON web-based dashboard. Brief notes were produced and recorded in hospital electronic medical records, as well as sent to the patient at the end of the consultation session.

Usual Care

Patients who were allocated to the control group received usual care. The usual care at the hospital consisted of a clinician consultation, an initial in-person pharmacist consultation (often 5-10 minutes in length), and a phone call follow-up from a clinical nurse within 1-2 weeks after commencing medication.

XSL•FO
RenderX

Measures

Demographic was completed at baseline (t_0). Patients' personal and clinical information was collected from the hospital's electronic medical record system. The study's primary outcomes included the patients' acceptance of SAMSON, measured by the Unified Theory of Acceptance and Use of Technology (UTAUT) [24,25] at 12 weeks (t_1), and study feasibility, measured by predefined rates of recruitment, randomization, retention, intervention adherence, and outcome assessment completion [26,27]. Secondary outcomes included MA, self-reported adherence, toxicity self-management, anxiety, depression and symptoms, and quality of life. Outcomes were assessed at baseline (t_0) and at the end of week 12 (t_1), except MA measured in week 16. All survey data collection (Multimedia Appendix 3) was done through (Research Electronic Data Capture) REDCap (Vanderbilt University) [28].

Primary Outcome Measures

Acceptability

The UTAUT questionnaire was adapted to assess determinants of HCPs' and patients' acceptance and use of the SAMSON solution, including 5 dimensions: performance expectancy, effort expectancy, social influence, facilitating conditions, and behavioral intention [25]. Patients were asked to rate their satisfaction with the SAMSON solution on a 5-point Likert scale. Participants were also invited to provide free-text feedback and suggestions on the solution.

Feasibility

A traffic light approach [29] (Table S1 in Multimedia Appendix 4) was used to determine the feasibility success at 3 levels: (1) feasible (above the upper threshold), (2) infeasible (below the lower threshold), and (3) protocol revision (between the 2 thresholds). The thresholds were defined in consultation with the Steering Committee and based on the literature of similar studies. Informal discussions with study personnel were attempted to obtain feedback on the feasibility.

Secondary Outcome Measures

MA was measured by medication refill adherence (MRA) [30] collected from pharmacy dispensing data. MRA was defined as a percentage calculated from the total days' supply divided by the number of days of study participation and multiplied by 100. In this study, the patient was considered as optimal adherence if their MRA was≥90%.

Self-reported adherence (Adherence Starts with Knowledge 12 [ASK-12]) was measured [31]; this measure included 12 items in 3 subscales: adherence behavior, health beliefs, and inconvenience or forgetfulness.

Toxicity self-management was measured with the Patient Activation Measure-Short Form [32,33], a 13-item self-report measure assessing the patient's knowledge, skills, and confidence in the self-management of their disease and related symptoms.

Anxiety, depression, and symptoms were measured by the Patient-Reported Outcomes Measurement Information System (PROMIS) [34] to assess depression, anxiety, pain interference, fatigue, sleep disturbance, and physical function.

Quality of Life was measured by Functional Assessment of Cancer Therapy-General (FACT-G) [35], which is a 27-item self-report scale measuring the quality of life of patients currently undergoing cancer treatment.

Analysis

Descriptive statistics were used to summarize participant characteristics across study arms, and differences in baseline attributes were assessed using *t* tests or chi-square tests as appropriate.

SAMSON acceptability was analyzed thematically. The UTAUT aims to examine individual quality dimensions, which means it is a suite of scales rather than one quality measure; therefore, adding up the overall scale of the questionnaire is not suitable. Results of UTAUT surveys were summarized for each of the 5 dimensions with the percentage of patients endorsing Likert scale ratings of 3 (disagree, neither agree nor disagree, and agree). Free text answers to UTAUT questionnaires were narratively summarized to gain further insight into acceptability.

Feasibility was determined based on the traffic light approach. Recruitment feasibility was assessed by the number of patients recruited (consented) divided by the number of patients who were approached to join the study. Randomization feasibility was assessed by the number of patients who were randomized divided by the number of patients who consented. Retention in both arms of the study was assessed by the number of patients who remained at the end of the study divided by the total number of patients who consented to join the study. We also tracked intervention adherence, for example, the percentage of patients who completed tasks on the SAMSON smartphone app and received MI teleconsultations. Compliance data for survey completion was calculated as the percentage of patients who completed surveys at t_0 and $t_1 t_1$ out of the total patients in the study at these time points.

Secondary outcomes were analyzed using linear regression. The dependent variable was the outcome at follow-up and the independent variables were arm assignment and the outcome at baseline. Standard checks for normality and homoscedasticity of residuals were conducted for each outcome. All analyses were conducted using StataNow 18 (StataCorp) [36].

Ethical Considerations

A steering committee—which included chief investigators; experts in the fields of digital health, information technology, nursing, pharmacy, psychology, and oncology; and a patients' representative—was formed to provide support for the study. This study was approved by the HRECs of PMCC (number HREC/95332/PMCC) and Swinburne University of Technology (number 20237273-15836) (Multimedia Appendices 5 and 6). Patients reviewed study details and indicated their consent using e-consent forms. Patients were encouraged to contact the study team if they had any questions or concerns. Patients' personal and health information in the SAMSON app was encrypted in transit and stored in a secure server at Swinburne. Study team access to patient's data on the SAMSON web-based dashboard

XSL•FO
RenderX

was password protected and limited to the research coordinator (RC) and 4 study nurses and pharmacists. All data analyses were conducted on deidentified data. Patients received a US $31 gift voucher if they completed all surveys in the study.

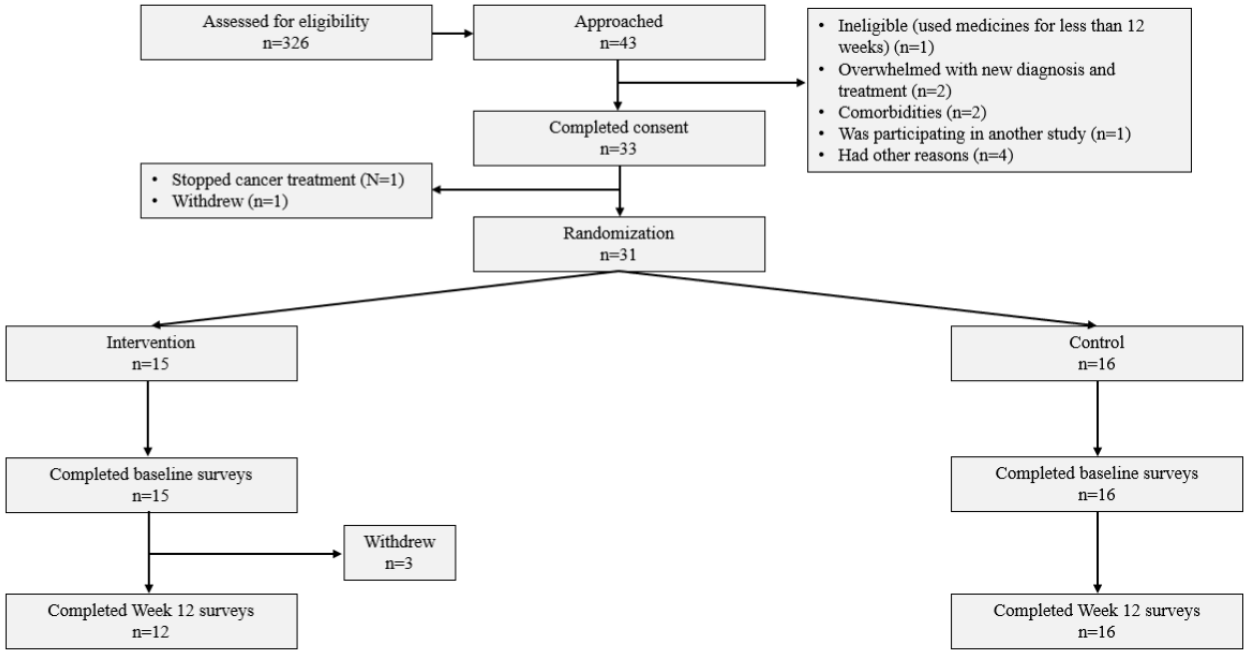
Results

Patient Characteristics

Among 42 patients who were approached, 33/42 (79%) consented to participate, and 31/33 (94%) completed baseline (t_0) surveys. Of those who completed baseline surveys, 3/31 patients (all from the intervention arm) withdrew from the study (10 %) due to the burden of the disease. All the remaining 28 patients (100%) completed week-12 surveys (12 intervention and 16 control) (Figure 1).

Demographics of randomized patients are provided in Table S2 in Multimedia Appendix 7. There were no significant differences in participants' demographics between the 2 arms.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram of Safety and Adherence to Medications and Self-care Advice in Oncology (SAMSON) pilot randomized controlled trial.



Acceptability Results (Primary Outcome 1)

Overview

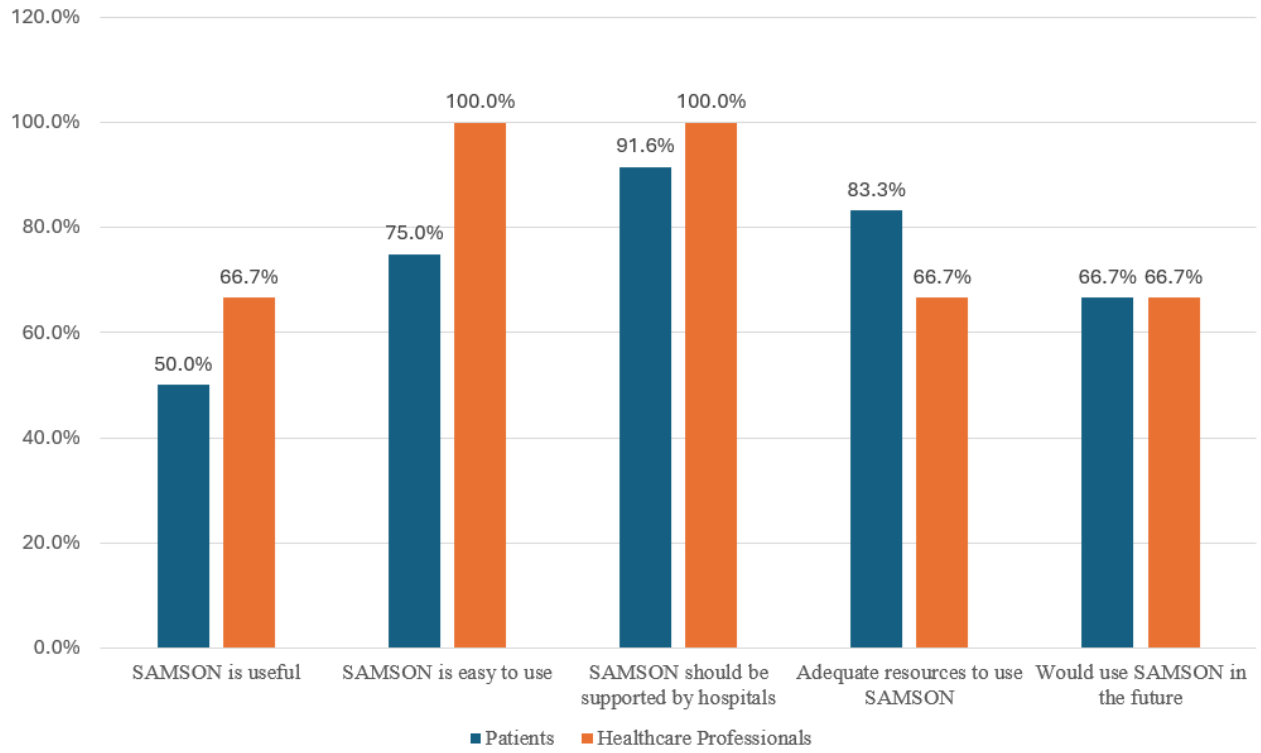
Out of 15 intervention arm patients, 13/15 (87%) installed the SAMSON app on their phone for use (2 patients withdrew). All 15 patients in the intervention arm received the initial pharmacy teleconsultation. Most patients (14/15, 93%) received at least 1 nurse teleconsultation, and 10 of them (10/15, 67%) completed all 4 scheduled nurse teleconsultations.

Data retrieved from the SAMSON app showed moderate engagement among participants. Patients completed an average of 99 app tasks (including responses to daily medication reminders and weekly side effect surveys) over the 12-week study period, which accounted for 99/140, 71% of scheduled tasks. A total of 36 severe episodes of symptoms were reported.

Out of 15 patients, 12/15 (80%) and 3 study HCPs (100%) completed the UTAUT surveys. A summary of patients' and HCPs' responses to the UTAUT questionnaire is presented below, with full details in Multimedia Appendices 8 and 9. Figure 2 presents patients' and HCPs' key opinions about the SAMSON solution.

XSL•FO
RenderX

Figure 2. Patients' and health care professionals' opinions about Safety and Adherence to Medications and Self-care Advice in Oncology (SAMSON) solution (n=12 and n=3, respectively).



Performance Expectancy

Half of the patients (6/12) in the intervention arm reported that their MA could improve with the help of the SAMSON solution. Patients used different MA-supporting mechanisms provided by SAMSON solution, including prompting reminders (7/12, 58%), disease and treatment education (10/12, 83%), improving confidence in treatment (6/12, 50%), and improving side effect self-management skills (9/12, 75%). Despite this, 3/12 patients (25%) found the SAMSON app component was not that helpful. Of those, 2 were using another commercial MA app with greater functionality. Some patients (4/12, 33%) reported functional issues of the app occurring in a short period of time (1-3 days), for example, medication reminders were not delivered or responses to reminders were not saved properly.

Of the 3 HCPs participating in the trial, 2 found the SAMSON solution useful in their job. All HCPs agreed that SAMSON could enable their 2-way communication with patients which would increase their ability in supporting patient treatment adherence. However, 1 HCP suggested that the SAMSON dashboard's visual design could be improved to be more appealing. HCPs' comments on the time required for MI teleconsultations were controversial: some suggested that the scheduled time was good, while it was reported as quite long for another, which might be a constraint in clinical practice.

Effort Expectancy

The SAMSON app was found to be easy to use by most patients (10/12, 83%). Its presentation was clear, and the content was easy to understand (11/12, 92%). Most participants could easily navigate the app (10/12, 83%), yet 1 (1/12, 8%) struggled in responding to surveys in the app. Some issues with the app's functionality were reported, for example, the medication reminders' time was not automatically updated after the daylight saving time changed, and the app was frozen sometimes, which affected patients' responses to reminders.

All HCPs reported that SAMSON was quite easy to use. However, some issues with the SAMSON web-based dashboard were reported, for example, it did not allow more than 1 HCP to monitor the patient, so all study HCPs had to share 1 account. Two (2/3, 67%), HCPs found implementing MI consultations quite challenging in practice due to time constraints.

Social Influence

Out of 12 patients, 5 (41.6%) thought that their family and friends would support their use of SAMSON, while most of them (10/12, 83%) thought that other patients with cancer would find SAMSON valuable. Almost all patients (11/12, 92%) desired SAMSON to be available for use in cancer hospitals.

All HCPs commented that their colleagues would find the SAMSON valuable and desire to receive the hospital's support in implementing SAMSON in daily practice.

Facilitating Conditions

Most participants were confident that they had adequate knowledge and resources (10/12, 83%) to use SAMSON and could access support when needed (9/12, 75%). Nevertheless, 2/12 patients (17%) reported their difficulties when dealing with some technical issues when using the app.

Despite all HCPs reporting that they had adequate knowledge to apply MI techniques in teleconsultations, 1/3 (33%) wished to have more training resources on the SAMSON app.

XSL•FO
RenderX

Behavioral Intention

Most patients (9/12, 75%) felt confident using the SAMSON app. They would want to continue using SAMSON after the study finishes (8/12, 67%) or recommend it to their peers (7/12, 58%). The 3 patients (25%) who were using another commercial MA app with more functionalities would prefer to receive MI teleconsultations alone. Cost was reported as an important factor influencing the intention to use SAMSON by the majority of patients (8/12, 67%).

Regarding the open-ended questions in the UTAUT survey on experience and perception of the SAMSON solution, patients valued the SAMSON smartphone app, because it was "easy to use and prompted reminders" (P31), "placed all medication resources in one place" (P16), "informative and helpful" (P25), and provided "help quickly if [the patient has] any queries" (P30). The "side-effects" tab within the app, providing self-care advice, was reported to make the patient "feel more secure in managing medication regimens" (P2), which could result in high adherence—as one patient commented, "I didn't miss a dose" (P7). The number of teleconsultations, duration, and quality were reported by most patients as good or "perfect" (P3) and "at the right length of time and allowed me [the patient] to ask all questions I need while allowing the nurse to gather all information" (P2); however, 1 mentioned it was "too long" (P31). Patients provided some helpful suggestions to improve the convenience of using the SAMSON smartphone app, including fixing glitches, combining all medications scheduled at the same time in 1 reminder (rather than separate reminders), having more options for responding to reminders, and more attractive presentation.

Over 60% of study HCPs felt confident in delivering SAMSON to patients in the trial and would like to continue using it in the future. All of them would introduce the solution to their peers to use.

Feasibility Results (Primary Outcome 2)

The study recruitment rate was 33/42 (79%). Most participants went through the information and consent process via phone, except 4 were recruited on-site. The randomization rate was 31 of 33 (94%). The retention rate was 28 of 33 (85%). The RC conducted phone calls or in-person check-ups at least 2 times during the 12-week study. All recruitment, randomization, and retention rates were higher than the upper threshold.

A total of 13 intervention arm patients (13/15, 87%) used the SAMSON app. The proportions of responses to medication reminders and side effect surveys compared to the scheduled tasks during the 12-week study period were 1061 of 1541 (68.9%) and 128 of 141 (90.8%), respectively. The unmet threshold of responses to medication reminders could be due to the app's functional issues as reported earlier. In total, 65 MI teleconsultations were delivered from August 2023 to June 2024, of which 35 of 65 (54%) were conducted on time as scheduled, 26 (40%) were later than scheduled, and 4 (6%) were missed. The average length of initial pharmacy consultations was 40 minutes, while nurse follow-up consultations were 40 minutes on average. Reasons for delayed and missed consultation sessions were mostly on the patient side, including not showing up at the appointment (25/30, 83%), not responding to HCP phone calls (4/30, 13%), or international travel (1/30, 3%). Only 2 sessions (7%) were delayed, because the HCP could not match their work schedule. Of the 12 intervention arm patients who were retained until the end of the study, 10 of 12 (83%) received all 5 scheduled teleconsultations, 1 (8%) received 4 teleconsultations and 1 (8%) received 2 teleconsultations.

Baseline surveys were completed by 31/33 (94%) patients. All who stayed until the end of the study (28/28, 100%) completed week-12 surveys. Participants completed all surveys online without any need for support. An alert email was sent by RC to all participants a few days before the survey due date. However, over one-third of participants (10/28, 36%) only completed surveys after being reminded. Details of feasibility results are presented in Table S3 in Multimedia Appendix 10.

Preliminary Efficacy Results (Secondary Outcomes)

Medication Refill Adherence

Of 28 patients who completed the study, 25/28 (89%) had a 100% adherence rate and 3/28 (11%) had over 95% adherence rate. There is no difference in the proportion of patients who had optimal adherence (MRA ≥90%) between the intervention and control groups (100% for both groups).

The mean and 95% CI of ASK-12, Patient Activation Measure-Short Form, PROMIS, and Functional Assessment of Cancer Therapy-General at baseline and week 12 are shown in Table S4 in Multimedia Appendix 11.

Discussion

Principal Findings

In this study, we aimed to test the acceptability, feasibility, and preliminary efficacy of a multicomponent MA solution, SAMSON, to help patients with cancer improve their adherence to OAMs and self-manage their physical and emotional symptoms. Overall, patients and oncology HCPs rated SAMSON as highly acceptable, usable, and useful. These high levels of user satisfaction evidence that the solution meets the various needs of support among patients with cancer to medically adhere to and manage side effects at home [37], as well as the needs of oncology HCPs for a practical and tailored MI training, and a means to regularly monitor and provide ongoing support to patients' MA [20]. Moreover, qualitative findings suggest that SAMSON has the potential to help patients in MA and symptom self-management, as well as to assist HCPs in monitoring and supporting patients' adherence. The results of the study help to address the gap of knowledge and the need in oncology practice [17,18] by providing evidence of the high quality and potential effect of a digital multicomponent MA solution.

Regarding SAMSON's feasibility, we noted a high level of engagement with the SAMSON solution in terms of the overall tasks completed by patients on the SAMSON smartphone app and the MI teleconsultations completed by both HCPs and patients. The high acceptability and feasibility of SAMSON is a result of several factors. First, co-design and rigorous design framework were applied to develop SAMSON [22]. By

XSL•FO
RenderX

involving end users and stakeholders throughout all design stages of the intervention, co-designing helped to improve the solution's acceptability, desirability, and usability [38,39]. The use of design frameworks, for example, design science research methodology in this study, could enhance the artifact's design, which is crucial in digital intervention development, and improve its acceptance, usage, and efficacy [40]. Furthermore, the design framework usage can improve the rigor and translatability of research, which is currently limited or poorly reported in the development of available MA interventions in cancer [18]. Second, after development, individual components of SAMSON were successfully tested by end users on their acceptability, usability, and usefulness [19,20]. To the best of our knowledge, SAMSON is the first comprehensive digital MA solution in cancer that is co-designed, theory-based, evidence-based, and rigorously developed and tested.

Although overall acceptance and feasibility were relatively high, some HCPs and patients in the trial reported several technical issues with the SAMSON app and desired better visualization, more functionality, and further instructions and training. This feedback will be used to further improve the solution in the future. A couple of study HCPs were concerned about the time required for MI consultations. Time constraints have been reported as one of the barriers to MI implementation in clinical practice [41]. It is noted that in this study, SAMSON was delivered as an add-on service on top of the hospital's usual care, which required extra time and effort from busy HCPs. In addition, delivering MI consultations to promote adherence was a newly acquired skill by the study HCPs, thus, they required more time to master this approach. A holistic approach, including the hospital's mechanisms to provide continuous MI monitoring and training, as well as support (both in terms of skills and resources) to facilitate and maintain HCPs' confidence and motivation in using this skill set can be a solution to tackle MI implementation barriers [41,42].

The overall recruitment, randomization, retention, and data collection compliance rates were higher than the preset upper thresholds. A high attrition rate is generally one of the major concerns in digital health RCTs compared to RCTs testing more "traditional" interventions [43]. However, the high recruitment rate and the low rate of loss to follow-up in this trial indicate that the SAMSON solution and online pragmatic RCT design are feasible in busy oncology clinical settings, as long as appropriate methodological strategies are applied. Specifically, in this study, clinical nurses and pharmacists from the Hematology Department were recruited and funded to support study recruitment and deliver teleconsultations. The RC dedicated additional time to building rapport with participants and consistently assisted them through all processes of the trial (consenting, baseline assessment, randomization, app installation, technical training and support, and outcome assessments). In addition, different channels of communication, for example, emails, phone calls, and SMSs were used to follow-up and motivate participants throughout the study. The importance of check-ups and follow-ups postrandomization in trials of digital interventions, where participants are required to have certain levels of digital literacy to perform tasks of the intervention [44], has been emphasized in several studies [45]. Future digital trials may benefit from these recruitment and retention strategies.

In this trial, a combination of techniques was used to measure MA, including self-report measures and prescription refill reports. Although triangulation of measurements could improve the accuracy of adherence [46], its interpretation needs to be cautious. The MRA using pharmacy computer records, which is objective [47], does not guarantee that all dispensed medications were consumed by patients. The self-reported adherence survey (ASK-12) is simple, but often subjective and more about barriers to adherence than actual adherence status [31]. Future studies should consider using high-accuracy methods, such as the Medication Event Monitoring System [5], to measure MA.

This study has some limitations. Like many digital health studies, our sample was predominantly Caucasian and highly educated [48,49]. Consequently, the findings may not be representative of patients who are non-Caucasian or have lower socioeconomic status, who might face higher barriers to adherence and could potentially benefit more from the SAMSON solution than their Caucasian or higher socioeconomic status peers [50,51]. Participants were recruited from the Hematology department at PMCC, one of Australia's leading oncology hospitals, and were prescribed only 1 oral anticancer regimen. Therefore, the results may not be generalizable to those who use multiple anticancer medications or receive care in low-resource oncology settings. A more targeted recruitment strategy focusing on underserved cancer patient populations with other types of cancer in various levels of oncology care institutions is warranted.

Conclusion

Before undertaking this pilot trial, both components of the SAMSON solution were co-designed and developed based on evidence and theory, and then individually tested on target users. The results of this study confirmed that SAMSON is acceptable, usable, and useful for both HCPs and patients with cancer. Both the SAMSON solution and the pragmatic RCT design are feasible in real-life oncology settings. These findings are very encouraging, given the numerous challenges in applying RCT as an evaluation design for digital health interventions [52]. Our next steps will involve refining the SAMSON solution based on participants' feedback from this study and conducting a full RCT to evaluate its clinical and economic effectiveness.

Acknowledgments

We thank all Peter MacCallum Cancer Centre HCPs, especially Ms Megan Palmer (nurse), and Ms Jenny Rogers (research pharmacist and trial coordinator) for their support in participant recruitment and teleconsultation delivery; Ms Gail Rowan and Ms Sally L Brooks (senior pharmacists) for developing medicine information and side effect self-care advice; Professor Michael

XSL•FO
RenderX

Jefford (Director of the Australian Cancer Survivorship Centre), Professor Ben Solomon, Associate Professor Shahneen Sandhu (consultant medical oncologists), Mr Alan White, and Mrs Fiona White (consumer representatives) for contributing to the study protocol development and being part of the study's steering committee. We thank all patients for their dedicated time and participation in this study.

This work was supported by the Digital Health CRC Limited (DHCRC), Swinburne University of Technology, and Peter MacCallum Cancer Centre (project DHCRC-0043). DHCRC is funded under the Australian Commonwealth's Cooperative Research Centers (CRC) program. THD is supported by the Australian Government Research Training Program Scholarship. The sponsor had no influence on the study design or the collection, analysis, and interpretation of data. The final decision to include the comments and submit the manuscript for publication was made only by the authors.

Data Availability

The datasets generated and analyzed during this study are not publicly available but are available from the corresponding author on reasonable request.

Authors' Contributions

THD is the submitting and corresponding author. THD conceived of the study, its design, coordination, and implementation and drafted the manuscript. THD, AW, KB, MA, NW, PPJ, PS, and SQ were involved in study design and protocol development. THD and PS were involved in the literature review and developing study instruments and materials. AW and MD supported patient recruitment and delivered MI teleconsultations. MA assisted in collecting medication refill data. THD, PS, and SQ contributed to the design of the statistical analysis approach. THD analyzed the data in consultation with SQ and PS. All authors were involved in providing a critical review of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1135 KB - formative_v9i1e65302_app1.pdf]

Multimedia Appendix 2
Participant information statement and consent form.
[DOCX File , 292 KB - formative_v9i1e65302_app2.docx]

Multimedia Appendix 3
Survey booklet.
[DOCX File , 314 KB - formative_v9i1e65302_app3.docx]

Multimedia Appendix 4
Table S1. Thresholds for traffic light approach to feasibility.
[DOCX File , 15 KB - formative_v9i1e65302_app4.docx]

Multimedia Appendix 5
Ethics approvals.
[PDF File (Adobe PDF File), 247 KB - formative_v9i1e65302_app5.pdf]

Multimedia Appendix 6
Ethics approvals.
[PDF File (Adobe PDF File), 148 KB - formative_v9i1e65302_app6.pdf]

Multimedia Appendix 7
Table S2. Participant demographics by arm.
[DOC File , 70 KB - formative_v9i1e65302_app7.doc]

Multimedia Appendix 8
Unified Theory of Acceptance and Use of Technology results of patient participants.

XSL•FO
RenderX

[DOCX File , 19 KB - formative_v9i1e65302_app8.docx]

Multimedia Appendix 9
Unified Theory of Acceptance and Use of Technology results of health care professional participants.
[DOCX File , 20 KB - formative_v9i1e65302_app9.docx]

Multimedia Appendix 10
Table S3. Study feasibility results.
[DOCX File , 18 KB - formative_v9i1e65302_app10.docx]

Multimedia Appendix 11
Table S4. Differences and change from baseline to week 12 between arms, analysis of covariance test.
[DOCX File , 18 KB - formative_v9i1e65302_app11.docx]

References

1. Weingart SN, Brown E, Bach PB, Eng K, Johnson SA, Kuzel TM, et al. NCCN Task Force Report: oral chemotherapy. J Natl Compr Canc Netw 2008 Mar;6 Suppl 3:S1-14. [Medline: 18377852]
2. NICE. Improving Outcomes in Hematological Cancers: The Manual. London: National Institute for Clinical Excellence; 2003.
3. Agrawal M, Garg RJ, Cortes J, Quintás-Cardama A. Tyrosine kinase inhibitors: the first decade. Curr Hematol Malig Rep 2010;5(2):70-80. [doi: 10.1007/s11899-010-0045-y] [Medline: 20425399]
4. Alshawa A, Gong J, Marcott V, Khan R, Boving V, Brown L, et al. Strategic improvement of oral antineoplastic investigational agents compliance. Global Journal on Quality and Safety in Healthcare 2019;2(1):5-10. [doi: 10.4103/jqsh.Jqsh_23_18]
5. Sabate E. Adherence to Long-term Therapies: Evidence for Action. Geneva, Switzerland: World Health Organization; 2003.
6. Ruddy K, Mayer E, Partridge A. Patient adherence and persistence with oral anticancer treatment. CA Cancer J Clin 2009;59(1):56-66 [FREE Full text] [doi: 10.3322/caac.20004] [Medline: 19147869]
7. Bouwman L, Eeltink CM, Visser O, Janssen J, Maaskant JM. Prevalence and associated factors of medication non-adherence in hematological-oncological patients in their home situation. BMC Cancer 2017;17(1):739. [doi: 10.1186/s12885-017-3735-1] [Medline: 29121889]
8. Gater A, Heron L, Abetz-Webb L, Coombs J, Simmons J, Guilhot F, et al. Adherence to oral tyrosine kinase inhibitor therapies in chronic myeloid leukemia. Leuk Res 2012;36(7):817-825. [doi: 10.1016/j.leukres.2012.01.021] [Medline: 22364811]
9. Jacobsen P, Sweet KL, Lee YH, Tinsley S, Lancet JE, Komrokji RS, et al. Adherence to Tyrosine Kinase Inhibitor (TKI) Therapy in Patients with Chronic Myeloid Leukemia (CML). Blood 2011;118(21):4431 [FREE Full text]
10. Guilhot F, Coombs J, Zernovak O, Szczudlo T, Rosti G. A global retrospective and physician-based analysis of adherence to tyrosine kinase inhibitor (TKI) therapies for chronic myeloid leukemia (CML). Blood 2010;116(21):1514 [FREE Full text]
11. Jabbour EJ, Kantarjian H, Eliasson L, Cornelison AM, Marin D. Patient adherence to tyrosine kinase inhibitor therapy in chronic myeloid leukemia. Am J Hematol 2012;87(7):687-691 [FREE Full text] [doi: 10.1002/ajh.23180] [Medline: 22473898]
12. Wu EQ, Johnson S, Beaulieu N, Arana M, Bollu V, Guo A, et al. Healthcare resource utilization and costs associated with non-adherence to imatinib treatment in chronic myeloid leukemia patients. Curr Med Res Opin 2010;26(1):61-69. [doi: 10.1185/03007990903396469] [Medline: 19905880]
13. Ekinci E, Nathoo S, Korattyil T, Vadhariya A, Zaghloul HA, Niravath PA, et al. Interventions to improve endocrine therapy adherence in breast cancer survivors: what is the evidence? J Cancer Surviv 2018;12(3):348-356. [doi: 10.1007/s11764-017-0674-4] [Medline: 29396760]
14. Kavookjian J, Wittayanukorn S. Interventions for adherence with oral chemotherapy in hematological malignancies: a systematic review. Res Social Adm Pharm 2015;11(3):303-314. [doi: 10.1016/j.sapharm.2014.08.006] [Medline: 25262600]
15. Bank T. World bank group. Mobile phone access reaches three quarters of planet's population. 2012. URL: https://tinyurl.com/46yjd99k [accessed 2012-07-17]
16. Kay M, Santos J, Takane M. mHealth: new horizons for health through mobile technologies. World Health Organization Contract No 2011;9(1):66-71.
17. Greer JA, Amoyal N, Nisotel L, Fishbein JN, MacDonald J, Stagl J, et al. A systematic review of adherence to oral antineoplastic therapies. Oncologist 2016;21(3):354-376 [FREE Full text] [doi: 10.1634/theoncologist.2015-0405] [Medline: 26921292]

XSL•FO
RenderX

18. Dang TH, Forkan ARM, Wickramasinghe N, Jayaraman PP, Alexander M, Burbury K, et al. Investigation of intervention solutions to enhance adherence to oral anticancer medicines in adults: overview of reviews. JMIR Cancer 2022;8(2):e34833 [FREE Full text] [doi: 10.2196/34833] [Medline: 35475978]
19. Dang TH, Wickramasinghe N, Forkan ARM, Jayaraman PP, Burbury K, O'Callaghan C, et al. Co-design, development, and evaluation of a mobile solution to improve medication adherence in cancer: design science research approach. JMIR Cancer 2024;10:e46979 [FREE Full text] [doi: 10.2196/46979] [Medline: 38569178]
20. Dang TH, Ludlow C, Borle H, Alexander M, Wickramasinghe N, Burbury K, et al. Co-designing a motivational interviewing training platform to enhance oncology healthcare professional communication. PEC Innov 2024;5:100335 [FREE Full text] [doi: 10.1016/j.pecinn.2024.100335] [Medline: 39257627]
21. Eysenbach G, CONSORT-EHEALTH Group. Improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
22. Dang TH, Wickramasinghe N, Jayaraman PP, Burbury K, Alexander M, Whitechurch A, et al. Safety and adherence to medications and self-care advice in oncology (SAMSON): pilot randomised controlled trial protocol. BMJ Open 2024;14(7):e079122 [FREE Full text] [doi: 10.1136/bmjopen-2023-079122] [Medline: 39043598]
23. Broglio K. Randomization in clinical trials: permuted blocks and stratification. JAMA 2018;319(21):2223-2224. [doi: 10.1001/jama.2018.6360] [Medline: 29872845]
24. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
25. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Quarterly 2012;36(1):157. [doi: 10.2307/41410412]
26. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. J Psychiatr Res 2011;45(5):626-629 [FREE Full text] [doi: 10.1016/j.jpsychires.2010.10.008] [Medline: 21035130]
27. Abbott JH. The distinction between randomized clinical trials (RCTs) and preliminary feasibility and pilot studies: what they are and are not. J Orthop Sports Phys Ther 2014;44(8):555-558. [doi: 10.2519/jospt.2014.0110] [Medline: 25082389]
28. Vanderbilt University. Tennessee, USA: Vanderbilt University. REDCap (Research Electronic Data Capture). URL: https://www.project-redcap.org/ [accessed 2025-09-25]
29. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS consensus group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016;355:i5239 [FREE Full text] [doi: 10.1136/bmj.i5239] [Medline: 27777223]
30. Hess LM, Raebel MA, Conner DA, Malone DC. Measurement of adherence in pharmacy administrative databases: a proposal for standard definitions and preferred measures. Ann Pharmacother 2006;40(7-8):1280-1288. [doi: 10.1345/aph.1h018]
31. Matza LS, Park J, Coyne KS, Skinner EP, Malley KG, Wolever RQ. Derivation and validation of the ASK-12 adherence barrier survey. Ann Pharmacother 2009;43(10):1621-1630. [doi: 10.1345/aph.1M174] [Medline: 19776298]
32. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. Health Serv Res 2005;40(6 Pt 1):1918-1930 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00438.x] [Medline: 16336556]
33. Bahrom NH, Ramli AS, Isa MR, Baharudin N, Badlishah-Sham SF, Mohamed-Yassin MS, et al. Validity and reliability of the patient activation measure (PAM)-13 Malay version among patients with metabolic syndrome in primary care. Malays Fam Physician 2020;15(3):22-34 [FREE Full text] [Medline: 33329860]
34. Quach CW, Langer MM, Chen RC, Thissen D, Usinger DS, Emerson MA, et al. Reliability and validity of PROMIS measures administered by telephone interview in a longitudinal localized prostate cancer study. Qual Life Res 2016;25(11):2811-2823 [FREE Full text] [doi: 10.1007/s11136-016-1325-3] [Medline: 27240448]
35. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The functional assessment of cancer therapy scale: development and validation of the general measure. J Clin Oncol 1993;11(3):570-579. [doi: 10.1200/JCO.1993.11.3.570] [Medline: 8445433]
36. StataCorp. Stata Statistical Software: Release 18. URL: https://www.businesswire.com/news/home/20230426005929/en/New-Stata-18-Statistical-Software-Empowers-Researchers-to-Make-the-Most-Out-of-Their-Analyses [accessed 2023-04-26]
37. Dang TH, O'Callaghan C, Alexander M, Burbury K, Jayaraman PP, Wickramasinghe N, et al. 'Take the tablet or don't take the tablet?' A qualitative study of patients' experiences of self-administering anti-cancer medications related to adherence and managing side effects. Support Care Cancer 2023;31(12):680 [FREE Full text] [doi: 10.1007/s00520-023-08122-6] [Medline: 37934298]
38. Borgstrom E, Barclay S. Experience-based design, co-design and experience-based co-design in palliative and end-of-life care. BMJ Support Palliat Care 2019;9(1):60-66. [doi: 10.1136/bmjspcare-2016-001117] [Medline: 28209568]
39. Slattery P, Saeri AK, Bragge P. Research co-design in health: a rapid overview of reviews. Health Res Policy Syst 2020;18(1):17 [FREE Full text] [doi: 10.1186/s12961-020-0528-9] [Medline: 32046728]
40. Hevner A, Wickramasinghe N. Design science research opportunities in health care. In: Wickramasinghe NS, Schaffer JL, editors. Theories to Inform Superior Health Informatics Research and Practice. 1st ed. Cham: Springer; 2018.
41. Bell DL, Roomaney R. Exploring the barriers that prevent practitioners from implementing motivational interviewing in their work with clients. Social Work 2020;56(4):416-429. [doi: 10.15270/56-4-884]

XSL•FO
RenderX

42. Nagy SM, Butame SA, Todd L, Sheffler JL, Budhwani H, Fernandez MI, et al. Barriers and facilitators to implementing a motivational interviewing-based intervention: a multi-site study of organizations caring for youth living with HIV. AIDS Care 2022;34(4):486-491 [FREE Full text] [doi: 10.1080/09540121.2021.1950604] [Medline: 34251935]
43. Mathieu E, McGeechan K, Barratt A, Herbert R. Internet-based randomized controlled trials: a systematic review. J Am Med Inform Assoc 2013;20(3):568-576 [FREE Full text] [doi: 10.1136/amiajnl-2012-001175] [Medline: 23065196]
44. Daniore P, Nittas V, Gille F, von Wyl V. Promoting participation in remote digital health studies: an expert interview study. Digit Health 2023;9:20552076231212063 [FREE Full text] [doi: 10.1177/20552076231212063] [Medline: 38025101]
45. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
46. Stewart SJF, Moon Z, Horne R. Medication nonadherence: health impact, prevalence, correlates and interventions. Psychol Health 2023;38(6):726-765 [FREE Full text] [doi: 10.1080/08870446.2022.2144923] [Medline: 36448201]
47. Brízido C, Ferreira AM, Lopes P, Strong C, Sá Mendes G, Fernandes Gama F, et al. Medication adherence to direct anticoagulants in patients with non-valvular atrial fibrillation - a real world analysis. Rev Port Cardiol (Engl Ed) 2021;40(9):669-675 [FREE Full text] [doi: 10.1016/j.repce.2021.07.002] [Medline: 34503705]
48. Smith SK, Kuhn E, O'Donnell J, Koontz BF, Nelson N, Molloy K, et al. Cancer distress coach: pilot study of a mobile app for managing posttraumatic stress. Psychooncology 2018;27(1):350-353 [FREE Full text] [doi: 10.1002/pon.4363] [Medline: 28032921]
49. Kelley MM, Kue J, Brophy L, Peabody AL, Foraker RE, Yen PY, et al. Mobile health applications, cancer survivors, and lifestyle modification: an integrative review. Comput Inform Nurs 2021;39(11):755-763 [FREE Full text] [doi: 10.1097/CIN.0000000000000781] [Medline: 34074873]
50. Smith SK, Zimmerman S, Williams CS, Benecha H, Abernethy AP, Mayer DK, et al. Post-traumatic stress symptoms in long-term non-Hodgkin's lymphoma survivors: does time heal? JCO 2011;29(34):4526-4533. [doi: 10.1200/jco.2011.37.2631]
51. Mollica M, Nemeth L, Newman SD, Mueller M. Quality of life in African American breast cancer survivors: an integrative literature review. Cancer Nurs 2015;38(3):194-204. [doi: 10.1097/NCC.0000000000000160] [Medline: 24945269]
52. Hrynyschyn R, Prediger C, Stock C, Helmer SM. Evaluation methods applied to digital health interventions: what is being used beyond randomised controlled trials? A scoping review. Int J Environ Res Public Health 2022;19(9):5221 [FREE Full text] [doi: 10.3390/ijerph19095221] [Medline: 35564616]

Abbreviations

ASK-12: Adherence Starts with Knowledge 12
HCP: health care professional
HREC: Human Research Ethics Committee
MA: medication adherence
MI: motivational interviewing
MITP: motivational interviewing training platform
MRA: medication refill adherence
OAM: oral anticancer medicine
PMCC: Peter MacCallum Cancer Centre
PROMIS: Patient-Reported Outcomes Measurement Information System
RC: research coordinator
RCT: randomized controlled trial
SAMSON: Safety and Adherence to Medications and Self-care Advice in Oncology
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by A Mavragani; submitted 12.08.24; peer-reviewed by O Santin, M Östbring; comments to author 05.12.24; revised version received 12.12.24; accepted 12.12.24; published 19.02.25.

Please cite as:
Dang TH, Wickramasinghe N, Jayaraman PP, Burbury K, Alexander M, Whitechurch A, Dyer M, Quinn S, Forkan ARM, Schofield P
Digital Solution to Support Medication Adherence and Self-Management in Patients with Cancer (SAMSON): Pilot Randomized Controlled Trial
JMIR Form Res 2025;9:e65302
URL: https://formative.jmir.org/2025/1/e65302
doi:10.2196/65302
PMID:39969972

XSL•FO
RenderX

©Thu Ha Dang, Nilmini Wickramasinghe, Prem Prakash Jayaraman, Kate Burbury, Marliese Alexander, Ashley Whitechurch, Mitchell Dyer, Stephen Quinn, Abdur Rahim Mohammad Forkan, Penelope Schofield. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Assessing the Feasibility and Utility of Patient-Specific 3D Advanced Visualization Modeling in Cerebrovascular Disease: Retrospective Analysis and Prospective Survey Pilot Study

Korak Sarkar[1,2,3,4], MD, MHDS; Vishal Bhimarasetty[1], BS, MS; Abdul Rahim[4], BE, MS; Colin Curtis[5], BSE; Kimberly Hughes[1], BA, BS

[1]Department of Academics, Ochsner BioDesign Lab, Ochsner Health, New Orleans, LA, United States

[2]Ochsner Neuroscience Institute, Ochsner Health, New Orleans, LA, United States

[3]Physical Medicine and Rehabilitation Service, Southeast Louisiana Veterans Health Care System, New Orleans, LA, United States

[4]University of Queensland, School of Medicine, Queensland, Australia

[5]Ochsner Information Services, Ochsner Health, New Orleans, LA, United States

Corresponding Author:
Korak Sarkar, MD, MHDS
Department of Academics
Ochsner BioDesign Lab
Ochsner Health
1514 Jefferson Highway
New Orleans, LA, 70119
United States
Phone: 1 504 894 2799
Email: korak.sarkar.md@gmail.com

Abstract

Background: The prevalence, clinical burden, and health care costs (>US $100 billion) associated with cerebrovascular disease (CVD) will increase significantly as the US population grows and ages over the next 25 years. Existing 2D imaging modalities have inherent limitations in visualizing complex CVD, which may be mitigated with the use of patient-specific 3D advanced visualization (AV) technologies. There remain gaps in knowledge, however, regarding how and with what impact these technologies are being used in CVD.

Objective: The aim of this study was to characterize the clinical attributes and reported utility associated with the use of 3D AV modeling in CVDs, specifically intracerebral arteriovenous malformations.

Methods: This pilot study employs a combination of retrospective analysis and prospective surveys to describe the utilization and utility of patient-specific AV models at a single high-volume certified comprehensive stroke center.

Results: From July 2017 to February 2023, 25 AV models were created for 4 different clinicians. The average patient age was 37.4 years; 44% (11/25) of the patients were African Americans, 52% (13/25) were on public insurance, and 56% (14/25) were associated with a neurovascular procedure. In this study, 18 clinicians with diverse experience responded to AV model surveys, with a 92.2% (166/180) completion rate. There was an average reported utility of 8.0 on a 0-10 scale, with higher scores reflecting increased utility. Compared to 2D viewing, AV models allowed staff to appreciate novel abnormal anatomy, and therefore, they would have changed their therapeutic approach in 45% (23/51) of the cases.

Conclusions: AV models were used in complex CVDs associated with young, publicly insured individuals requiring resource-intensive interventions. There was strong and diverse clinician engagement with overall report of substantial utility of AV models. Staff clinicians frequently reported novel anatomical and therapeutic insights based on AV models compared to traditional 2D viewing. This study establishes the infrastructure for future larger randomized studies that can be repeated for CVDs or other disease states and incorporate assessments of other AV modalities such as 3D printing and medical extended reality.

(JMIR Form Res 2025;9:e51939) doi:10.2196/51939

XSL•FO
RenderX

KEYWORDS

cerebrovascular disease; advanced visualization; 3D modeling; cerebrovascular; intracerebral arteriovenous malformations; artery; vein; vessel; medical extended reality; 3D printing; medical simulation; virtual reality; augmented reality; usability; survey; stroke; brain; cerebral

Introduction

Background

Cerebrovascular disease (CVD) is a common cause of stroke, aneurysms, and thromboembolic disease, resulting in significant neurologically based disability and death [1,2]. Although CVD can be treated with medications, many disease presentations require procedural interventions. It is estimated that by 2030, the annual costs for cerebrovascular accidents alone will exceed US $140 billion [3]. Improved visualization, greater repetition, and more frequent training experiences are factors known to improve outcomes in CVD treatments requiring complex procedures [4]. Treatment of patients with difficult or otherwise hostile anatomies carries more risk when operators have less general experience navigating unusual or atypical structures [5]. The lack of anatomical and procedural familiarity can lead to lengthier procedures, longer anesthetic duration, increased radiation exposure, and poorer outcomes [5,6]. Access to personalized 3D models designed for patient-specific planning can provide higher fidelity visualization and simulation than traditional 2D viewing, thereby reducing the prevalence, associated costs, and morbidity from complex CVD procedures [7].

The rapid adoption of advanced visualization (AV) tools such as 3D rendering, 3D printing, and medical extended reality has created novel opportunities for medical visualization and simulation. Patient-specific 3D digital and physical anatomical models based on clinical radiographic data can now be rendered and fabricated for patient education, clinical training, and procedural planning [8]. This approach holds great promise as a tool in visualizing complex neurovascular anatomy and simulating complex surgical and endovascular procedures [9]. The increasing adoption of advanced modeling has prompted organizations such as the Radiological Society of North America to issue guidelines for implementing medical 3D printing as training models for highly complex clinical scenarios [10]. The Radiological Society of North America guidelines discuss the use of patient-specific 3D models for a variety of specialties, including congenital heart disease, craniomaxillofacial, breast, and musculoskeletal pathologies. There, however, remains limited guidance on the optimal use of 3D modeling in CVD.

CVD interventions are an ideal application of AV modalities due to the idiosyncratic and patient-specific 3D nature of intracerebral vascular pathology. Current evidence suggests that AV technologies such as 3D rendering, medical extended reality, and 3D printing are an appropriate platform for studying and rehearsing plans for treating complex CVD [11]. Several studies have reported the growing utilization of AV modeling, visualization, and procedural training in CVD [12]. The current state of research, however, remains limited and warrants more thorough investigation [12,13]. To address this gap, the field would benefit from the development of a scalable modeling and assessment platform, which could describe and evaluate the use of patient-specific AV anatomical modeling. Such an infrastructure could address the clear, critical, and well-identified need to characterize the use and utility of AV models in CVD.

Objectives

This pilot study was designed and executed to describe the utilization and utility associated with the use of patient-specific anatomical models in the management of intracerebral arteriovenous malformations (iAVMs). This exploratory investigation involves the implementation of a scalable digital fabrication infrastructure to produce patient-specific anatomical AV models, characterize their use, and quantify their potential impact in the management of iAVMs. The aims of this study include describing the clinical, anatomical, and demographic features associated with the use of patient-specific AV models in the management of iAVMs. Additionally, this study aims to quantify the utility of patient-specific AV anatomical models when compared to traditional 2D viewing across different clinical experience levels. We hypothesize there is a negative correlation between experience and reported utility of AV patient-specific anatomical models in CVD. We specifically predict that when compared to standard 2D viewing, clinicians with less experience will report higher utility of AV patient-specific anatomical models than more experienced staff clinicians.

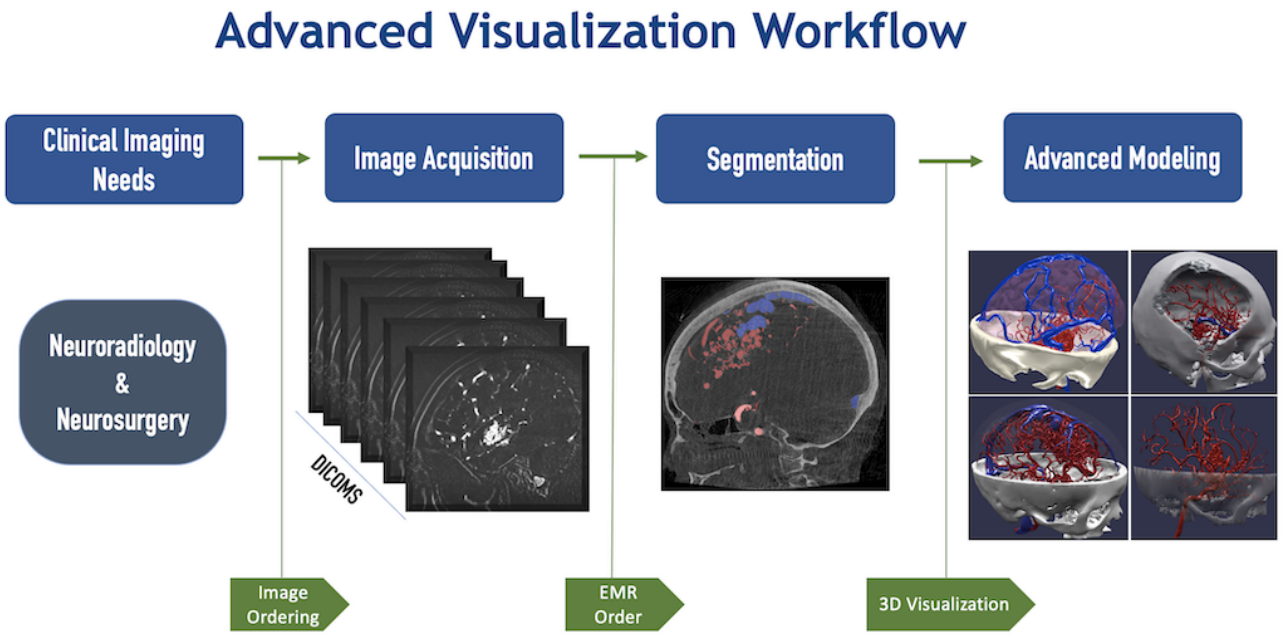
Methods

Digital Fabrication Process

Clinicians request AV modeling based on routine clinical neuroimaging imaging, for example, computed tomography, magnetic resonance, or rotational angiography via an electronic medical record order set created to request patient-specific 3D AV models. Anatomical models are based on the digital imaging and communications in medicine (DICOM) datasets used in clinical imaging. DICOM data are accessed by a staff biomedical engineer via a compliant and secure enterprise picture archiving and communications system. The biomedical engineer segments these DICOM datasets by using computer-aided design software to produce 3D mesh files, specifically standard tessellated language files. Segmentation is the process by which regions of interest are identified through automatic methods (eg, thresholding, edge detection, region growing) and are then refined by manual selection and separation of anatomical structures. This step is completed by the biomedical engineer in conjunction with a radiologist, neurologist, or neurosurgeon to ensure an accurate representation of the region of interest. Once created, these 3D models are rendered and manipulated digitally via an internally developed web-based viewer. These digital models can be further postprocessed for 3D printing to create physical models (Figure 1).

XSL•FO
RenderX

Figure 1. Ochsner BioDesign Digital Request and Fabrication Process. An electronic medical record–based order request is placed by a clinician. Source imaging DICOM (Digital Imaging and Communications in Medicine) datasets are accessed via a secure enterprise PACS (Picture Archiving and Communications System). Biomedical engineers utilize computer-aided design applications to transform DICOM datasets into 3D file types. These 3D files are then either visualized using a web-based browser or extended reality modalities such as virtual or augmented reality. These 3D files can be further processed to produce physical models using 3D printing. DICOM: Digital Imaging and Communications in Medicine; EMR: electronic medical record.



Epidemiological Data Collection

A CVD AV registry is used to collect data on clinical and utilization metrics associated with AV model use via a compliant research data capture application. Data are collected from the electronic medical record and include patient age at the time of modeling, sex, race, insurance status, imaging modality, comorbidities (hypertension, diabetes, epilepsy, tobacco use), size/scale/location of iAVM, associated current procedural terminology (CPT) codes, presence and type of any associated and relevant procedure within 9 months of model request, the presence and duration of any associated intensive care unit stay, as well as a Spetzler-Martin Grade (SMG). SMG is a clinical tool to assess the procedural risk of iAVMs that incorporates the size, location, and type of venous drainage.

Clinician and Model Survey Data Collection

Twenty-four clinicians were invited to participate in this study from the organization's neuroscience institute via clinical and educational conferences, 18 of whom responded. Each clinician assessed a set of 10 sequential iAVM cases, which had previously been requested for modeling. Each participant received a secure email, which included a link to the model utility survey as well as the following information to review: (1) a traditional 2D viewer with the axial, sagittal, and coronal source images available in serial via an embedded digital multimedia container format file, specifically an mp4; (2) information to access and view the source imaging via the standard picture archiving and communications system viewing approach; (3) the associated anatomical 3D AV model accessible via an internally developed web-based viewer; and (4) the radiographic report of the imaging from which the model was created.

Drawing from similar instruments used in the literature, a survey is used to assess (1) demographics and level of experience, (2) clinician-reported complexity of each case on 1-10 Likert Scale, and (3) utility of the patient-specific AV models using a 1-10 Likert scale, with higher scores representing increased complexity and utility (Multimedia Appendices 1-3) [10,12-16]. The survey includes questions regarding the potential diagnostic and therapeutic impact of 3D AV models compared to standard 2D representations. The research data capture application was used to digitally distribute and collect survey data in an automated and compliant way.

Analysis Plan

Descriptive statistics (averages, range, frequency) were reported for the epidemiological and utilization data for the AV models as well as the demographic and experience data of the survey respondents. Model survey responses were treated as ordinal. Each respondent's complexity and utility response was averaged to mitigate the limitations associated with the statistical analysis of the dependent and repeated measurements. The independent variable is experience level, and 2 approaches were used to classify experience groups. One approach categorizes experience by training level: (1) trainee defined as medical student, resident, or fellow; (2) staff as defined by those who have completed their terminal clinical training; and (3) advanced practice provider (APP). An alternative approach categorizes experience based on the reported number of iAVM evaluations: low (0-10), medium (11-100), and high (>101). The dependent variables are reported complexity and utility of the AV model compared

XSL•FO
RenderX

to traditional 2D viewing. The nonparametric Kruskal-Wallis H test was used to assess the null hypothesis that there is no difference in reported complexity or utility of an AV model among clinicians in these 3 groups. A significance level or α of .05 was used for statistical testing. A Dunn multiple comparison test was used for post hoc analysis to compare specific experience category combinations. A significance level or α of .05 was used for statistical testing.

Ethics Approval

This study was reviewed and approved by the Ochsner institutional review board (approval 2019.089) and was determined to be exempt and granted a waiver of informed consent, as it posed minimal risk to participants. All data collected for this study were fully deidentified prior to analysis to ensure participant privacy and confidentiality. Study models and data were deidentified and anonymized. Secure storage protocols were implemented in compliance with institutional and federal data protection standards. Access to the data was restricted to authorized study personnel directly involved in data collection and analysis. No compensation was provided to study participants. All procedures were conducted in accordance with appropriate ethical guidelines and applicable regulations.

Results

The aim of this study was to deploy a digital fabrication process that can (1) describe the clinical attributes associated with the use of AV models and (2) assess the utility of AV models across clinical experience levels when compared to standard 2D viewing.

Epidemiology and Utilization

AV models for 26 patients with iAVMs were requested from July 2017 to February 2023. Of the 26 models requested, 25 AV models were created. One request was not modeled because the primary source image was ordered by the clinician but never completed by the patient. Of note, 6 out of the 25 cases were also 3D printed to create physical patient-specific models (Multimedia Appendix 4). Of the AV models, 60% (15/25) were based on rotational angiography, 32% (8/25) on computed tomography angiography, and 8% (2/25) on magnetic resonance angiography. Patient age ranged from 13 to 71 years (mean age 37.4 years, SD 16.4 years) with various comorbidities and diverse demographics (Table 1).

Table 1. Demographics and comorbidities of the participants for whom 3D models were created.

	Values, n (%)
Smoking (n=25)	
Nonsmoker	4 (16)
Smoker	21 (84)
Comorbidities (n=10)	
Hypertension	8 (80)
Diabetes mellitus	2 (20)
Hyperlipidemia	0 (0)
Seizure/epilepsy	0 (0)
Other	0 (0)
Total comorbidities (n=21)	
1	8 (38)
2	4 (19)
3	6 (29)
4	1 (5)
5	2 (10)
Other epidemiology metrics (n=25)	
Gender	
Female	8 (32)
Male	17 (68)
Race	
Caucasian	14 (56)
African American	11 (44)

With regards to pathophysiology, iAVM diameter ranged from 4.5 mm to 62 mm, with an average diameter of 29.2 mm.

XSL•FO
RenderX

Intracerebral hemorrhage was associated with 28% (7/25) of the cases. SMG scores were obtained and ranged from 1 to 5 with a median score of 2 (Figure 2).

The most frequent iAVM location was the frontal lobe (9/25, 36%), and the most frequent venous draining was the superior sagittal sinus (9/25, 36%) (Table 2). Clinically, 56% (14/25) of the models were associated with some form of open or endovascular intervention within 9 months of the AV model request.

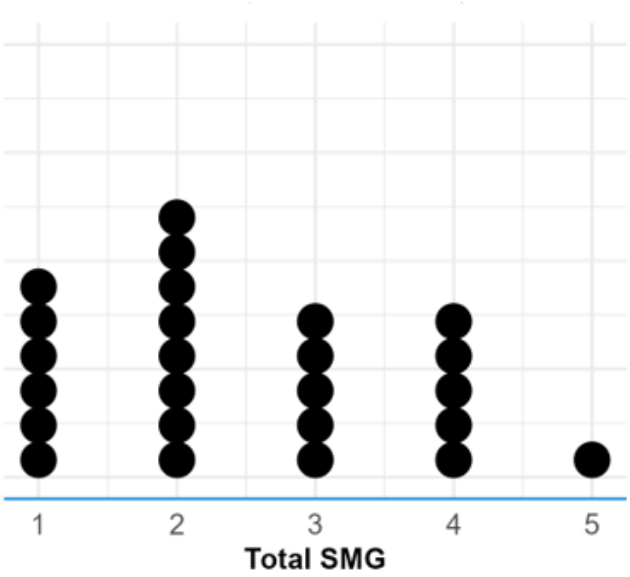
With regard to utilization, 9 unique CPT codes were found to be associated with these models. The most frequent CPT code associated with AV models was for selective unilateral catheter access of the common carotid artery (36223) and was billed in 84% (21/25) of the cases. Of the patients who had models created, 64% (16/25) had public insurance, with most patients utilizing Medicaid (14/25, 56%). Intensive care unit stays were associated with 56% (14/25) of the cases with a mean length of 6.9 days and a maximum length of 34 days. The 25 AV models were created for 4 unique clinicians at the following frequencies: interventional neuroradiologist (1/25, 4%), pediatric neurosurgeon (2/25, 8%), interventional neurologist (5/25, 20%), and a vascular neurosurgeon (17/25, 68%).

Figure 2. (A) Spetzler-Martin Grading scoring approach and (B) dot plot of Spetzler-Martin Grading scores, showing the distributions for values 1 (6/25, 24%), 2 (8/25, 32%), 3 (5/25, 20%), 4 (5/25, 20%), and 5 (1/25, 4%). SMG: Spetzler-Martin Grading.

(A)

Characteristics		Points assigned
Size		
	Small (<3 cm)	1
	Medium (3-6 cm)	2
	Large (>6 cm)	3
Eloquence		
	No	0
	Yes	1
Venous drainage		
	Superficial only	0
	Any deep	1

(B)



XSL•FO
RenderX

Table 2. Frequency distribution of intracerebral arteriovenous malformation locations and venous drainage categories (n=25).

Location/category	Values, n (%)
Intracerebral arteriovenous malformation locations	
Frontal lobe	9 (36)
Frontoparietal lobe	5 (20)
Occipital lobe	3 (12)
Other	8 (32)
Venous drainage category	
Superior sagittal sinus	9 (36)
Cortical vein	5 (20)
Transverse sinus	3 (12)
Sigmoid sinus	2 (8)
Other	6 (24)

Descriptive Statistics for Clinicians

Of the 24 clinicians who completed the initial recruitment survey between November 2022 to January 2023, 18 completed at least one model assessment survey. Six clinicians were lost to follow-up, that is, 1 medical student and 5 APPs. The descriptive statistics of the demographics and experience were calculated for the 18 clinicians who responded (Table 3). Clinicians (mean age 36.5 years; median age 34.5 years, range 26-56 years) identified their specialties as neurosurgery (1/18, 6%), neurointerventional radiology (3/18, 17%), and neurology (12/18, 67%). Two respondents (11%) were medical students and did not identify with a specialty. The other respondents identified as residents (4/18, 22%), APPs (5/18, 28%), fellow (1/18, 6%), junior attendings with <5 years of experience from terminal training (3/18, 17%), and senior attendings with >5 years of experience after terminal training (3/18, 17%). Self-reported iAVM evaluations was reported as 0 (2/18, 11%), 1-10 (4/18, 22%), 11-50 (6/18, 33%), 51-100 (3/18, 17%), 101-200 (2/18, 11%), and >201 (1/18, 6%).

Table 3. Summary statistics of the specialties and level of training for clinician survey respondents (n=18).

	Values, n (%)
Specialties/level of training	
Neurosurgery	1 (6)
Neurointerventional radiology	3 (17)
Neurology	12 (67)
Medical student	2 (11)
Level of training	
Trainees	5 (28)
Staff	6 (34)
Advanced practice provider	5 (28)
Number of arteriovenous malformation evaluations	
Low (<10)	6 (34)
Medium (51-100)	9 (50)
High (>101)	3 (17)
Clinician gender distribution	
Male	10 (56)
Female	8 (44)

Descriptive Statistics of Model Surveys

Each of the 18 clinicians were provided 10 model surveys for a total of 180 possible responses. Approximately 95.6% (172/180) of the surveys were initiated, and 92.2% (166/180) were completed. The possible complexity scores ranged from 1 to 10, with 10 being the most complex. The median of the average complexity reported was 6.95 with a minimum average score of 3.5 and a maximum average score of 8.1 across all respondents. Descriptive statistics, including median and IQR,

XSL•FO
RenderX

were calculated for utility, complexity, and improvement attributed to access to AV models for each expertise level (Figure 3) based on training level and the number of iAVMs evaluated (Figure 4). Clinicians responded affirmatively in 99.4% (171/172) of the cases that they would like to share AV models with their patients.

Figure 3. Boxplots for complexity, utility, and improvement derived from 3D model based on expertise: advanced practice provider (n=5), staff (n=6), and trainee (n=7). APP: advanced practice provider.

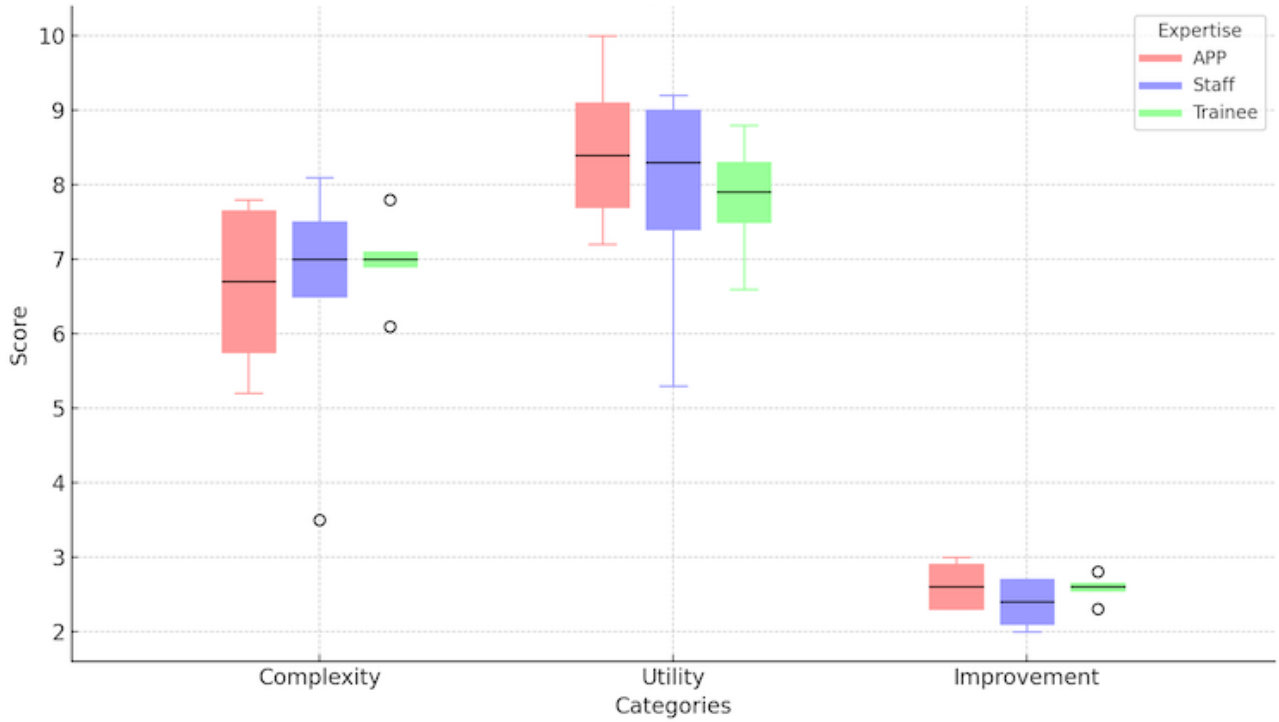
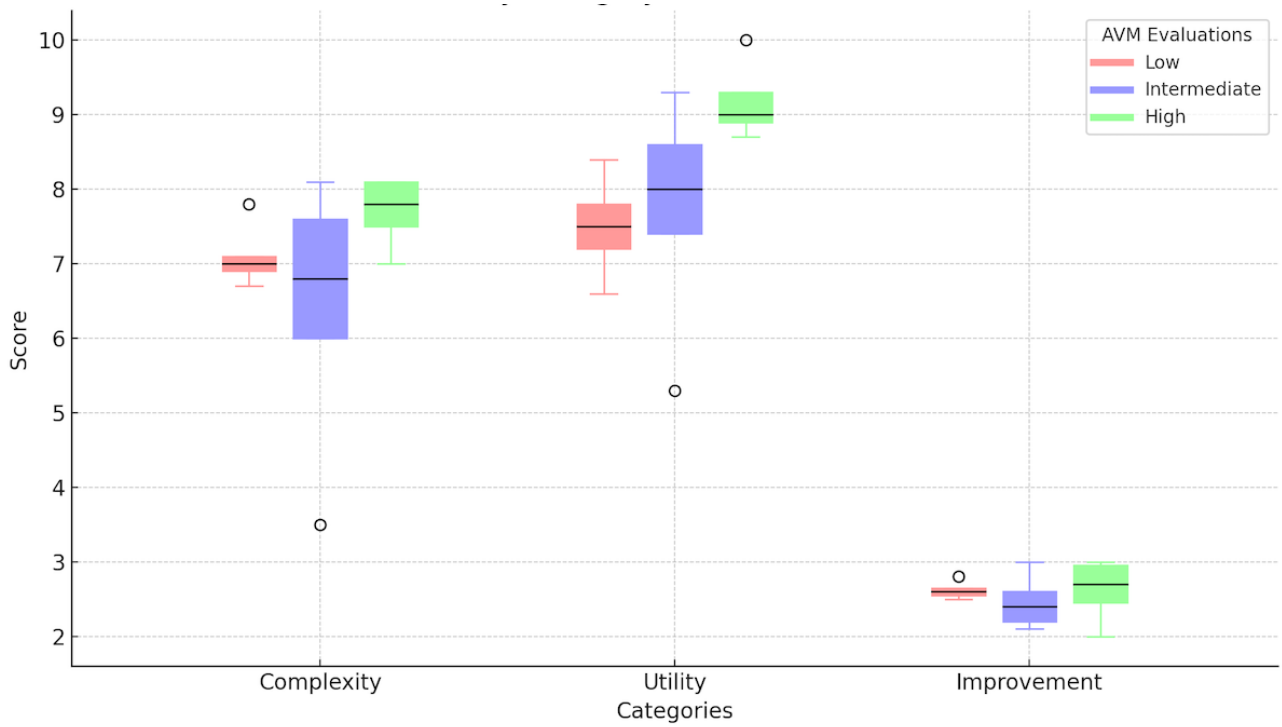


Figure 4. Boxplots for complexity, utility, and improvement derived from 3D model based on the number of arteriovenous malformations evaluated: low (0-10, n=6), intermediate (11-100, n=9), and high (>101, n=3). AVM: arteriovenous malformation.



There was a total of 6 physicians who had completed their terminal training with a cumulative survey response rate of 85% (51/60). When compared to standard 2D models, these 6 staff clinicians reported that AV models provided some improvement in visualization in 49% (25/51) of the cases and substantial improvement in 41% (21/51) of the cases. Staff level physicians reported that in 49% (25/51) of the cases, they were able to appreciate normal anatomy in the AV model that they could

XSL•FO
RenderX

not in standard 2D imaging. Similarly, staff reported that in 45% (23/51) of the cases, they were able to appreciate abnormal anatomy in the AV model that they could not in standard 2D imaging. Staff answered affirmatively in 14% (7/51) of the cases that the AV model would change their diagnosis while reporting affirmatively in 45% (23/51) of the cases that the AV model would change their therapeutic approach (Table 4).

Table 4. Diagnostic and management impact of advanced visualization compared to traditional 2D viewing on staff level clinicians (n=51).

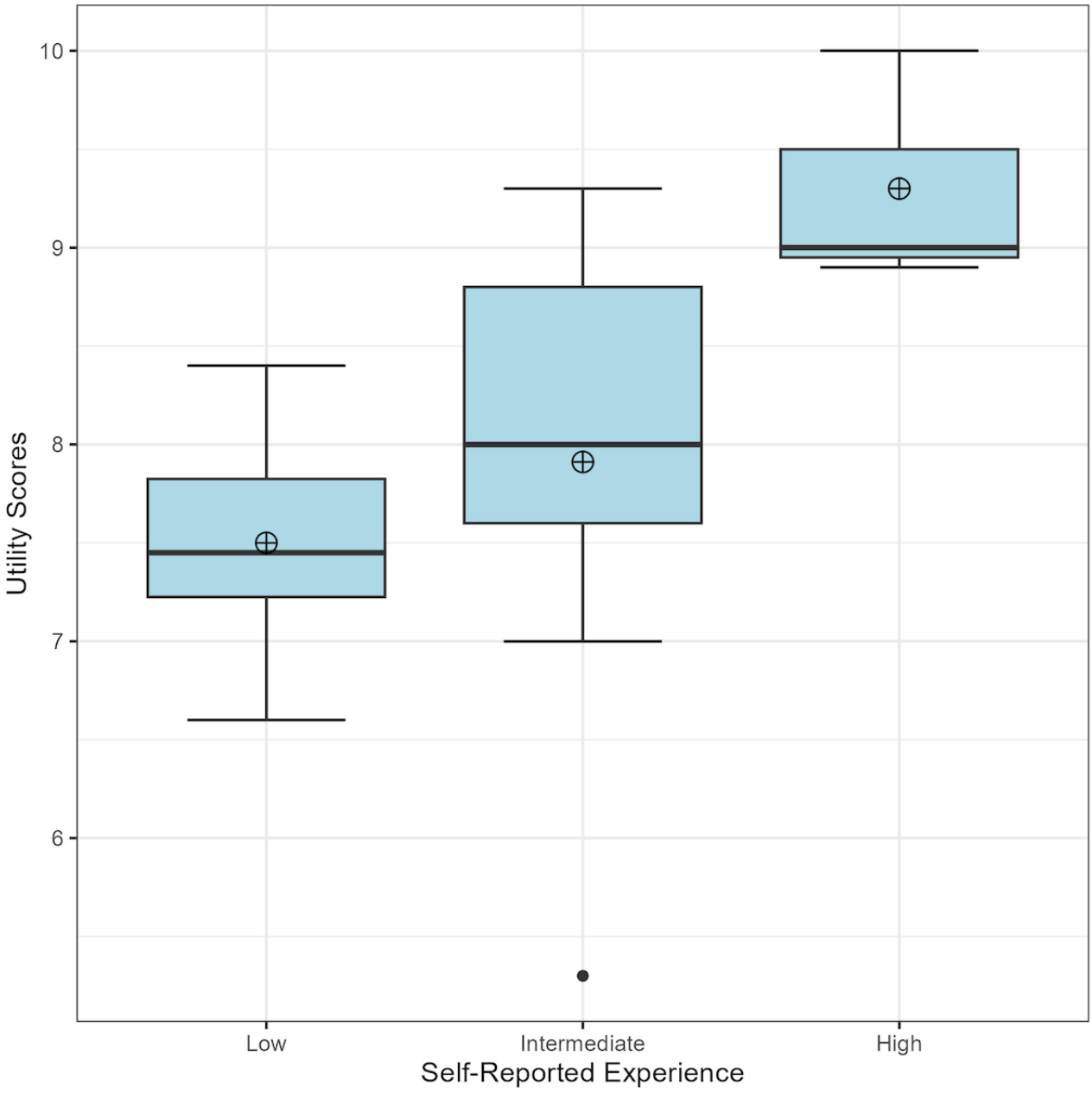
	Values, n (%)
Change in diagnosis	
No	39 (76)
Yes	7 (14)
No response	5 (10)
Change in therapeutic procedures	
Yes	23 (45)
No	21 (41)
No response	7 (14)
Change in appreciation of normal anatomy in 3D	
Yes	25 (49)
No	20 (39)
No response	6 (12)
Change in appreciation of pathologic defects in 3D	
Yes	23 (45)
No	22 (43)
No response	6 (12)

Inferential Analytics of Model Surveys

When categorized by training level, the 3 groups were distributed between APPs (5/18, 28%), trainees (7/18, 39%), and staff (6/18, 33%). The Kruskal-Wallis H nonparametric test was deployed with nonsignificant *P* values >.05 for the survey questions relevant to model complexity (*P*=.72) and utility (*P*=.48). Thus, we fail to reject the null hypothesis. Rather, we conclude that there is no statistically significant difference in reported complexity nor utility between clinician types. When categorized by self-reported experience, the 3 groups were distributed between low (6/18, 33%), intermediate (9/18, 50%), and high (3/18, 17%). The Kruskal-Wallis H nonparametric test was deployed with *P*=.12 for the survey questions relevant to model complexity. Thus, we fail to reject the null hypothesis for complexity and conclude that there is no statistically significant difference in reported complexity between experience levels. The *P* value for utility, however, did achieve significance with *P*=.03. Thus, for utility, we can reject the null hypothesis and conclude that there is a statistically significant difference in the reported utility metric between self-reported experience groups. Post hoc analysis with Dunn multiple comparison test demonstrated that the comparison between the low and high group showed a significant difference in the utility metric. There was not a significant difference in the utility metric between the comparisons of the other groupings: high-intermediate or intermediate-low (Figure 5).

XSL•FO
RenderX

Figure 5. Boxplot of utility across self-reported experience. The Kruskal Wallis H test demonstrated a statistical difference between the low and high experience groups (*P*=.03). The analysis reveals that the median utility scores for low (n=6), intermediate (n=9), and high (n=3) categories of self-reported experience are 7.45, 8.00, and 9.00, respectively.



Discussion

Principal Findings

This study demonstrates the feasibility of creating and deploying a digital fabrication and assessment infrastructure for AV models in complex presentations of CVDs, such as iAVMs. The scalable and secure request, creation, distribution, and evaluation of AV models was confirmed by the implementation of 25 AV models for iAVMs from July 2017 to February 2023 at a single high-volume certified stroke center. This retrospective analysis confirmed the utilization of AV models in CVDs, specifically iAVMs. Multiple clinicians requested AV models that were created from a variety of imaging modalities, most commonly from rotational angiography. Of note, requests for AV models persisted despite the decrease in elective procedures during the COVID-19 pandemic and a nationwide iodinated contrast shortage in 2022, which severely impacted imaging and procedure volumes and is indicative of the utility clinicians derive from AV models.

This analysis demonstrates associations between AV model use in CVD and a diverse set of clinical attributes and utilization metrics. Consistent with the complex and diverse population in Louisiana and the surrounding Gulf Coast region, a substantial number of models were created for individuals younger than 40 years, African Americans, and patients on Medicaid. Many of these patients were not only young but otherwise healthy with few if any other comorbidities, besides tobacco use. The feasibility to collect health care finance and utilization metrics such as CPT codes and intensive care unit lengths of stay associated with the request for AV models in complex CVD

XSL•FO
RenderX

presentations was confirmed. The analysis revealed the association of AV models with resource-intensive health resources such as intensive care unit stays, rotational angiography, and neurovascular procedures.

A survey instrument grounded in the literature was successfully created and implemented to assess the impact of AV models in complex CVD at scale. There was strong engagement across a variety of clinical experience levels, particularly APPs, neurological residents, and neurovascular staff, with an overall 92.2% (166/180) survey completion rate. On a set of iAVMs of varying complexity, a diverse array of clinicians reported substantial utility of AV. Staff-level clinicians reported that AV models would have rarely changed their diagnosis compared to traditional 2D viewing. Staff, however, reported that in nearly half of cases, the patient-specific AV models allowed them to appreciate normal and abnormal anatomy that they could not with traditional 2D imaging. Moreover, in 45% (23/51) of the cases, access to the AV model would have changed their therapeutic approach compared to their assessment based on traditional 2D viewing. Staff physicians reported that AV models were much more likely to change their therapeutic approach than their diagnosis. There was no significant difference in reported complexity or utility when clinicians were categorized by training: APP, trainee, staff. There was, however, a statistically significant difference (P=.03) in reported utility when clinicians were categorized by self-reported experience. Interestingly, the difference in reported utility was most pronounced between the high and low experience group, with the most experienced group reporting more utility than those with less experience. This was contrary to our hypothesis and suggests that the ability to understand the idiosyncratic and complex pathologies and extract relevant information from AV models may require a certain level of pre-existing knowledge and clinical expertise. Such a prerequisite appreciation of clinical anatomy may help clinicians derive diagnostic and therapeutic insights from AV models. Alternatively, this observation may be attributable to the psychological phenomena known as the Dunning-Kruger effect in which people with low ability, expertise, or experience regarding a certain type of task or area of knowledge tend to overestimate their ability or knowledge [17].

Strengths and Limitations

This study has multiple strengths attributable to a multidisciplinary approach that focused on clinical operational integration as a foundation for clinical research. Evaluation of an AV intervention required coordinating a cross-functional team consisting of biomedical engineers, research staff, clinical administrators, and clinicians. This study was the impetus to expand and ruggedize a scalable digital fabrication and assessment process for AV models. The survey response rates indicate not only enthusiasm for the field of inquiry but the willingness of these busy and increasingly overextended clinicians to engage in clinical research on 3D AV [18,19]. This points to the importance of continuing to develop and validate AV technologies not just for CVDs but for other complex disease states, including but not limited to spine deformity, genitourinary pathologies, orthopedic trauma, craniomaxillofacial abnormalities, and congenital heart disease.

This study has several limitations. This was a nonrandomized study, which could only establish associations and not causative impacts of AV modeling. The ability to extrapolate these findings is also limited by its small sample size and focus on a single center. Clinician respondents were basing feedback on hypothetical use cases and not the actual implementation of the 10 models in diagnosis or treatment planning. Moreover, the clinicians were all colleagues who may have influenced each other's responses, which may have introduced a diffusion of treatment effect as described by Urban and van Eeden-Moorefield [20]. Finally, the results of the statistical testing may be due to the small sample size rather than the absence or presence of differential utility among varying experience levels.

Challenges and Future Directions

The creation of AV models remains resource-intensive in terms of expertise, time, and infrastructure. Complex disease states where AV models are likely to be the most impactful are often low in prevalence requiring longer periods of study and collaboration between multiple sites, engendering a gap in the literature. These barriers have limited the number of conclusive studies validating the clinical and utilization impact of 3D AV models. Approaches to increase the efficiency of AV model creation, for example, machine-learning enhanced segmentation, may mitigate these obstacles to validation and warrant further research and development.

This study expands and utilizes a scalable digital fabrication infrastructure for future larger multisite randomized studies to assess the efficacy of AV modeling. Such future trials could examine the utility of AV models in clinical care delivery by incorporating technical resource requirements such as segmentation time, utilization measures such as operative time and radiation exposure, as well as clinical outcome metrics such as functional status and hospital stay. This study establishes the feasibility of collecting medical billing and utilization metrics associated with 3D AV models that can be used in future cost-effective analyses. This infrastructure could also be applied to understand the use and impact of AV models in other disease states that would benefit from AV techniques. Future studies could incorporate quantitative and qualitative assessments of other AV modalities such as 3D printing or extended reality to investigate the different applications and utility of digital versus physical models in visualization and simulation. Finally, this infrastructure could be utilized to study the impact AV models have on health literacy and patient experience. The use of AV is quickly being adopted in clinical training and care delivery but requires further investigation and validation.

Conclusion

This pilot feasibility study establishes the infrastructure to create and assess patient-specific AV models in a digital and scalable way. This study confirms the use of AV models in complex, resource-intensive disease states such as iAVMs. This study demonstrates enthusiasm for the use of AV across different specialties and experience levels involved in the management of CVD. A heterogenous cohort of clinicians reported substantial utility of AV modeling when compared to standard 2D imaging. Counter to our hypothesis, those clinicians with the most

XSL•FO
RenderX

experience found AV models the most useful. Finally, staff level physicians reported at high rates that AV models modified therapeutic approaches and provided anatomical insights not appreciated in standard 2D imaging. Almost universally, clinicians across different specialties and experience levels affirmed that that their patients would benefit from the ability to view their pathology-specific 3D AV models. Collectively, these findings support the idea that compared to traditional 2D viewing, patient-specific AV models are useful to a diverse array of clinicians involved in the treatment of complex disease states, such as iAVMs.

Acknowledgments

We thank our colleagues at Ochsner Health, specifically the academics, philanthropy, and neurosciences departments who supported the development and implementation of this work. The authors would like to thank the faculty and staff of Cedars Sinai Master's Health Delivery Sciences as well as the 2023 cohort for their guidance, support, and mentorship. Research reported in this publication was supported by the Terence D D'Souza MD Fellowship in Neuroscience and the Ochsner Louisiana State University Collaborative Intramural Research program. The funders had no role in the study design, data collection, analysis, or interpretation of results. Generative artificial intelligence, specifically ChatGPT 4o V 1.2024.346, was used to assist in literature review and improve manuscript content.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

KS conceptualized, designed, and supported the implementation of this study as well as led the writing of this manuscript. VB supported the data collection, cleaning, and overall implementation of this study. AR supported the biostatistical analysis of the data. CC developed the web-based viewer, performed the segmentation, and supported study implementation. KH designed and deployed the digital request and assessment infrastructure. KS, VB, AR, CC, and KH wrote, edited, and approved the final manuscript.

Conflicts of Interest

KS has past but no current roles as an advisor to digital health companies, none of which have any financial or intellectual property rights to this research. He also engages in medical-legal consulting. During the period when this research was conducted, VB was affiliated with Ochsner Health. VB currently works in the digital health industry. VB has no financial, consulting, patent, or intellectual property interests related to this research, and declares no conflicts of interest that could have influenced this work. AR, CC, and KH have no disclosures to report.

Multimedia Appendix 1
Survey template: clinician, on-screen model survey, and survey email template.
[DOCX File , 18 KB - formative_v9i1e51939_app1.docx]

Multimedia Appendix 2
Representative 2D viewer included in surveys.
[MP4 File (MP4 Video), 11859 KB - formative_v9i1e51939_app2.mp4]

Multimedia Appendix 3
Representative web-based 3D viewer included in surveys.
[ZIP File (Zip Archive), 9237 KB - formative_v9i1e51939_app3.zip]

Multimedia Appendix 4
Photograph of 3D printing of an intracerebral arteriovenous malformation example.
[PDF File (Adobe PDF File), 1722 KB - formative_v9i1e51939_app4.pdf]

References

1. Portegies M, Koudstaal PJ, Ikram MA. Cerebrovascular disease. Handb Clin Neurol 2016;138:239-261. [doi: 10.1016/B978-0-12-802973-2.00014-8] [Medline: 27637962]
2. Ovbiagele B, Goldstein LB, Higashida RT, Howard VJ, Johnston SC, Khavjou OA, et al. Forecasting the future of stroke in the United States. Stroke 2013 Aug;44(8):2361-2375. [doi: 10.1161/str.0b013e31829734f2]

XSL•FO
RenderX

3. Colby S, Ortman JM. Projections of the size and composition of the US population: 2014 to 2060. US Census Bureau. 2015. URL: https://www.census.gov/content/dam/Census/library/publications/2015/demo/p25-1143.pdf [accessed 2023-04-16]
4. Kim BM, Baek J, Heo JH, Kim DJ, Nam HS, Kim YD. Effect of cumulative case volume on procedural and clinical outcomes in endovascular thrombectomy. Stroke 2019 May;50(5):1178-1183. [doi: 10.1161/strokeaha.119.024986]
5. Perri JL, Nolan BW, Goodney PP, DeMartino RR, Brooke BS, Arya S, et al. Factors affecting operative time and outcome of carotid endarterectomy in the Vascular Quality Initiative. J Vasc Surg 2017 Oct;66(4):1100-1108 [FREE Full text] [doi: 10.1016/j.jvs.2017.03.426] [Medline: 28712813]
6. Ghosn Y, Khdhir M, Jabbour Y, Dushfunian D, Kobeissi I, Abbas N, et al. Factors affecting radiation dose, radiation exposure time and procedural time in arterial embolization for active hemorrhage. Emerg Radiol 2024 Oct;31(5):641-652. [doi: 10.1007/s10140-024-02262-w] [Medline: 38955874]
7. Branson T, Shapiro L, Venter RG. Observation of patients' 3D printed anatomical features and 3D visualization technologies improve spatial awareness for surgical planning and in-theatre performance. Adv Exp Med Biol 2021;1334:23-37. [doi: 10.1007/978-3-030-76951-2_2] [Medline: 34476743]
8. Popescu D, Marinescu R, Laptoiu D, Deac GC, Cotet CE. DICOM 3D viewers, virtual reality or 3D printing - a pilot usability study for assessing the preference of orthopedic surgeons. Proc Inst Mech Eng H 2021 Sep;235(9):1014-1024. [doi: 10.1177/09544119211020148] [Medline: 34176364]
9. Liang F, Yang Y, Luo L, Liao B, Zhang G, Ou S, et al. Endovascular treatment of complex middle cerebral artery aneurysms using TuBridge flow diverters. Interv Neuroradiol 2020 Oct;26(5):539-546 [FREE Full text] [doi: 10.1177/1591019920946216] [Medline: 32722987]
10. Chepelev L, Wake N, Ryan J, Althobaity W, Gupta A, Arribas E, RSNA Special Interest Group for 3D Printing. Radiological Society of North America (RSNA) 3D printing Special Interest Group (SIG): guidelines for medical 3D printing and appropriateness for clinical scenarios. 3D Print Med 2018 Nov 21;4(1):11 [FREE Full text] [doi: 10.1186/s41205-018-0030-y] [Medline: 30649688]
11. Chae MP, Lin F, Spychal RT, Hunter-Smith DJ, Rozen WM. 3D-printed haptic "reverse" models for preoperative planning in soft tissue reconstruction: a case report. Microsurgery 2015 Feb;35(2):148-153. [doi: 10.1002/micr.22293] [Medline: 25046728]
12. Sun Z. A systematic review of 3-D printing in cardiovascular and cerebrovascular diseases. Anatol J Cardiol 2017:423-435. [doi: 10.14744/anatoljcardiol.2017.7464]
13. McGuire LS, Fuentes A, Alaraj A. Three-dimensional modeling in training, simulation, and surgical planning in open vascular and endovascular neurosurgery: a systematic review of the literature. World Neurosurg 2021 Oct;154:53-63. [doi: 10.1016/j.wneu.2021.07.057] [Medline: 34293525]
14. Rybicki FJ. 3D Printing in Medicine: an introductory message from the Editor-in-Chief. 3D Print Med 2015;1(1):1 [FREE Full text] [doi: 10.1186/s41205-015-0001-5] [Medline: 30050970]
15. Ryan J, Plasencia J, Richardson R, Velez D, Nigro JJ, Pophal S, et al. 3D printing for congenital heart disease: a single site's initial three-yearexperience. 3D Print Med 2018 Nov 08;4(1):10 [FREE Full text] [doi: 10.1186/s41205-018-0033-8] [Medline: 30649650]
16. Sebök M, Dufour JP, Cenzato M, Kaku Y, Tanaka M, Tsukahara T, et al. When is diagnostic subtraction angiography indicated before clipping of unruptured and ruptured intracranial aneurysms? an international survey of current practice. Acta Neurochir Suppl 2021;132:9-17 [FREE Full text] [doi: 10.1007/978-3-030-63453-7_2] [Medline: 33973023]
17. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. J Pers Soc Psychol 1999 Dec;77(6):1121-1134. [doi: 10.1037//0022-3514.77.6.1121] [Medline: 10626367]
18. West CP, Dyrbye LN, Shanafelt TD. Physician burnout: contributors, consequences and solutions. J Intern Med 2018 Jun;283(6):516-529 [FREE Full text] [doi: 10.1111/joim.12752] [Medline: 29505159]
19. Rodrigues H, Cobucci R, Oliveira A, Cabral JV, Medeiros L, Gurgel K, et al. Burnout syndrome among medical residents: A systematic review and meta-analysis. PLoS One 2018;13(11):e0206840 [FREE Full text] [doi: 10.1371/journal.pone.0206840] [Medline: 30418984]
20. Urban J, van Eeden-Moorefield BM. Designing and proposing your research project. In: Concise Guides to Conducting Behavioral, Health, and Social Science Research Series. Washington DC: American Psychological Association; 2018:119-127.

Abbreviations

APP: advanced practice provider
AV: advanced visualization
CPT: current procedural terminology
CVD: cerebrovascular disease
DICOM: Digital Imaging and Communications in Medicine
iAVM: intracerebral arteriovenous malformation
SMG: Spetzler-Martin Grade

XSL•FO
RenderX

Edited by A Mavragani; submitted 20.08.23; peer-reviewed by M Waleed, M Aksoy; comments to author 06.09.23; revised version received 13.01.25; accepted 22.01.25; published 21.02.25.

Please cite as:
Sarkar K, Bhimarasetty V, Rahim A, Curtis C, Hughes K
Assessing the Feasibility and Utility of Patient-Specific 3D Advanced Visualization Modeling in Cerebrovascular Disease: Retrospective Analysis and Prospective Survey Pilot Study
JMIR Form Res 2025;9:e51939
URL: https://formative.jmir.org/2025/1/e51939
doi:10.2196/51939
PMID:

©Korak Sarkar, Vishal Bhimarasetty, Abdul Rahim, Colin Curtis, Kimberly Hughes. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effect on Response Rates of Adding a QR Code to Patient Consent Forms for Qualitative Research in Patients With Cancer: Pilot Randomized Controlled Trial

Rebecca Wyse[1,2], PhD; Erin Forbes[1,2], PhD; Grace Norton[1], BSc; Priscilla Viana Da Silva[1,2], PhD; Kristy Fakes[1,2], PhD; Sally Ann Johnston[3], BSc; Stephen R Smith[1,2,3], PhD; Alison Zucca[1,2], PhD

[1]School of Medicine and Public Health, The University of Newcastle, Callaghan, NSW, Australia

[2]Hunter Medical Research Institute, New Lambton Heights, NSW, Australia

[3]Hunter New England Local Health District, John Hunter Hospital, New Lambton Heights, NSW, Australia

Corresponding Author:
Rebecca Wyse, PhD
School of Medicine and Public Health
The University of Newcastle
University Dr
Callaghan, NSW, 2308
Australia
Phone: 61 02 40420272
Email: rebecca.wyse@health.nsw.gov.au

Abstract

Background: The successful conduct of health and medical research is largely dependent on participant recruitment. Effective, yet inexpensive methods of increasing response rates for all types of research are required. QR codes are now commonplace, and despite having been extensively used to recruit study participants, a search of the literature failed to reveal any randomized trial investigating the effect of adding a QR code on qualitative research response rates.

Objective: This study aimed to collect data on rates of response, consent, and decline among patients with cancer, and the average time taken to respond following randomization to receive either a QR code or no QR code on the patient consent form for a qualitative research study.

Methods: This was a pilot randomized controlled trial (RCT) embedded within a qualitative research study. In total, 40 eligible patients received a recruitment pack for the qualitative study, which included an information statement, a consent form, and an addressed, stamped envelope to return their consent form. Patients were randomized 1:1 to the control (standard recruitment pack only) or intervention group (standard recruitment pack including modified consent form with a QR code).

Results: In total, 27 out of 40 patients (age: mean 63.0, SD 14.8 years; 45% female) responded to the consent form. A lower proportion of the QR code group (60%) responded (odd ratio [OR] 0.57, 95% CI 0.14-2.37; *P*=.44), compared to 75% of the standard recruitment group. However, a higher proportion of the QR group (35%) consented (OR 1.84, 95% CI 0.41-8.29; *P*=.43), compared to the standard recruitment group (20%). A lower proportion of the QR group (25%) declined (OR 0.34, 95% CI 0.09-1.38; *P*=.13) relative to the standard recruitment group (55%). The mean response time of the QR code group was 16 days (rate ratio [RR] 0.79, 95% CI 0.47-1.35; *P*=.39) compared to 19 days for the standard recruitment group. None of the age-adjusted analyses were statistically significant.

Conclusions: This underpowered pilot study did not find any evidence that offering an option to respond through a QR code on a patient consent form for a qualitative study increased the overall patient response rate (combined rate of consent and decline). However, there was a nonsignificant trend, indicating that more patients who received the QR code consented compared to those who did not receive the QR code. This study provides useful preliminary data on the potential impact of QR codes on patient response rates to invitations to participate in qualitative research and can be used to inform fully powered RCTs.

Trial Registration: OSF Registries 10.17605/OSF.IO/PJ25X; https://doi.org/10.17605/OSF.IO/PJ25X

(JMIR Form Res 2025;9:e66681) doi:10.2196/66681

XSL•FO
RenderX

KEYWORDS

QR code; qualitative research; cancer; randomized controlled trial; RCT; patient recruitment; consent forms; response rates

Introduction

The successful conduct of health and medical research depends on effective recruitment strategies. Low recruitment rates affect both quantitative and qualitative research designs [1-3], and there is a need to identify effective and affordable methods that are minimally burdensome. For example, a 2019 study reported the costs of 14 recruitment strategies among patients with chronic diseases and reported that 20% of the study budget was spent on recruitment [4]. Given the recent widespread adoption of QR codes [5-8], they offer potential as a valuable and cost-effective recruitment tool. A number of randomized controlled trials (RCTs) have investigated the use of QR codes to recruit participants for quantitative surveys [9]. However, the evidence of the effectiveness of QR codes in increasing survey recruitment is equivocal, with earlier studies [10,11], tending to find no impact of QR codes on response rates, and later studies tending to find a small but significant increases [9,12]. Authors of these studies note that earlier research should be interpreted with caution, given the widespread use and familiarity with QR codes, which is relatively recent [9], and has increased rapidly following the COVID-19 epidemic [13]. A 2024 large RCT (N=5550) that evaluated the effect of including a QR code in a printed and mailed recruitment to a web-based survey found a small but significant increase in survey participation (+1.31%) [12]. It also found that including QR codes resulted in higher participation rates from some harder-to-reach groups, including younger people and single people [12].

However, there is little research about the effect of QR codes on consent rates to other research forms, such as recruitment to research trials or qualitative research, which often require ongoing or more maintained or intense commitment than a one-off survey. A 2024 study randomized recruitment methods ([Twitter rebrand as] X, Facebook [Meta] or QR code displayed on a poster) to a clinical trial and found little to no engagement with the QR code relative to the other recruitment strategies; however, this may have been due to the format (ie, poster) rather than the technology [8].

A literature search failed to identify any study that empirically tests the use of QR codes among people invited to participate in qualitative health research. Given the dearth of studies of this nature, the primary objective of this pilot RCT was to determine whether adding a QR code to a participant consent form for a qualitative research study increased the participant consent rate (ie, the proportion of patients who responded and consented to participate). Secondary objectives were to collect data regarding differences in (1) the proportion of patients who responded (either to consent or decline; this outcome was registered in the Open Science Framework as the proportion of patients who did not respond, but for ease of reporting and understanding, the inverse is reported instead [ie, the proportion who did respond])), (2) the proportion of patients who responded and declined to participate, and (3) the average time (in days) taken to respond measured from the date the recruitment packs were sent.

Methods

Study Design

This was a pilot RCT embedded within a qualitative study with patients with colorectal cancer [14]. The purpose of the qualitative study was to determine the views of patients with colorectal cancer about a Digital Health Intervention (“RecoverEsupport”) [15] to enhance their recovery from surgery. The qualitative study required participants to review a newly developed web-based support program and to provide feedback in a recorded telephone-based interview (approximately 20 minutes) that would be scheduled at a time convenient to them. The interview asked them to provide feedback about the relevance and ease of understanding of the program, as well as the information they could recall and the likely impact of the program.

Participants

Participants were recruited from a Surgical Department at a public hospital in a noncapital metropolitan region of Australia. The Colorectal Cancer Liaison Nurse identified eligible patients from medical records. The eligibility criteria for this pilot RCT were consistent with the eligibility criteria for the qualitative study. Patients were invited to participate in the qualitative study through a recruitment pack containing a printed consent form and a prepaid return envelope (mailed on 1 November 2021). Reminder packs were sent to all nonresponders 3 weeks later (22 November 2021).

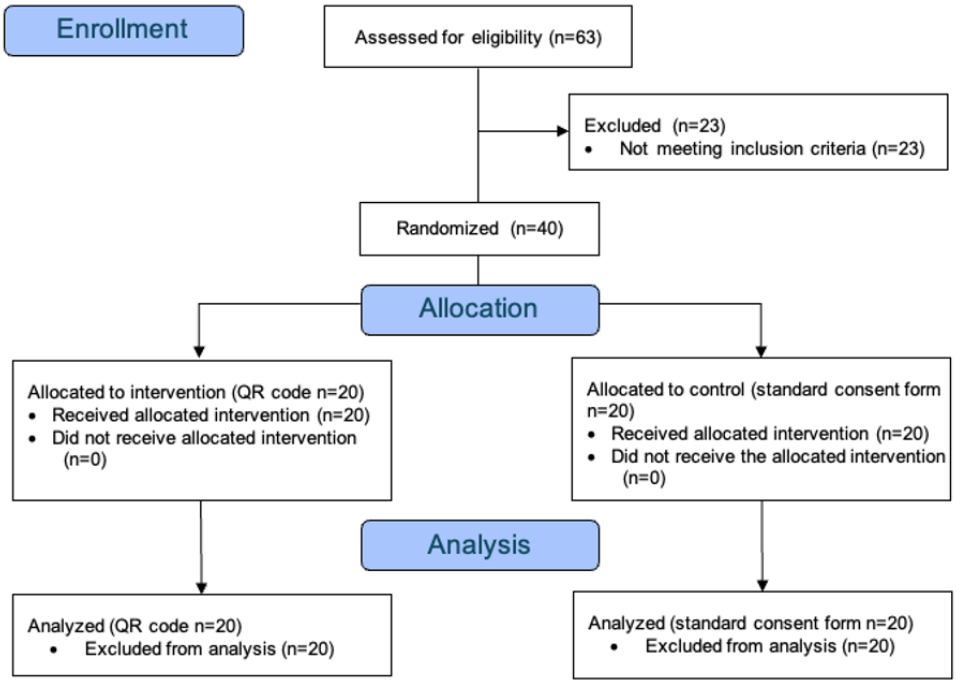
For the qualitative study in which the QR code study was embedded, the patient eligibility criteria were as follows: (1) a diagnosis of colorectal cancer, (2) being 18-90 years old, and (3) having undergone a bowel resection for colorectal cancer within the last 6 months (excluding stoma reversals and secondary bowel surgeries).

Allocation

An independent statistician randomly allocated the patients in a 1:1 ratio using a random number function in Microsoft Excel to the intervention or control group (Figure 1).

XSL•FO
RenderX

Figure 1. CONSORT Flowchart.



Standard Recruitment (Control)

These patients received the standard consent form (without the QR code). The form included 2 tick boxes for patients to indicate their decision to either consent or decline to participate in the qualitative study.

QR Code Group (Intervention)

These patients received an identical consent form to the control group, except this version included a QR code and instructions for accessing the web-based consent form to provide their consent. Scanning the QR code directed the participant to a web-consent-form, identical to the standard consent form, hosted within a REDCap (Research Electronic Data Capture; Vanderbilt University) database. REDCap automatically recorded the date of completion once the patient responded the web-based consent form.

Outcomes

The primary outcome was the proportion of patients who consented to participate in the qualitative study (through any channel, QR code, or mail). Secondary outcomes included (1) the proportion of patients who responded (either to consent or decline), (2) the proportion of patients who responded and declined to participate in the qualitative study, and (3) the average time in days taken to respond (measured from 1 November 2021 until either the postmark date on the return envelope or the date the web-based consent form was completed).

Statistical Analysis

Descriptive statistics are presented as counts (%) for categorical variables or mean (SD) for continuous variables. An intention-to-treat approach was used. Unadjusted and age-adjusted logistic regression models were fitted to estimate the odds ratio (OR) and 95% CIs. The between-group difference for the time taken to respond was analyzed using negative binomial instead of Poisson regression modeling (due to overdispersion) and unadjusted and age-adjusted rate ratios (RRs) are presented. Assumptions for the logistic regression model and negative binomial regression model were checked and found to be reasonable.

Ethical Considerations

The pilot was approved by the Ethics Committees of the Hunter New England Local Health District (2019/ETH00869) and the University of Newcastle (H-2015-0364) and was prospectively registered [16] and is reported in accordance with CONSORT (Consolidated Standards of Reporting Trials) guidelines [17]. Each invited patient received a patient information sheet inviting them to participate in the qualitative research, assuring them that their participation was entirely voluntary and that any information they provided would remain confidential. No incentive was offered for patients to participate. All project staff members were bound by confidentiality agreements. Collected data are stored in a deidentified format.

Results

Overview

Of the 63 patients assessed for eligibility, 23 were ineligible, most commonly because they did not have internet access (7/23 patients). The mean age was 58.60 (SD 17.38) years for the QR code, 67.40 (SD 10.35) years for the standard recruitment, totaling 63 (SD 14.8) years for both groups. In total, 45% (n=18) of the sample was female; 40% (n=8) for the QR code and 50% (n=10) for the standard recruitment groups were female. The outcomes are presented below and in Table 1 in sequential order (ie, overall response rate, consent rate, and decline rate).

XSL•FO
RenderX

Table 1. Pilot outcomes response rates and time to respond to the invitation to participate in a qualitative research study for participants randomized to receive a quick response code (intervention) or no quick response code (standard recruitment control).

	QR[a] code (Intervention) (n=20), (%)	Standard recruitment (Control) (n=20) (%)	Unadjusted effect estimate (95% CI)	*P* value	Adjusted for age effect estimate (95% CI)	*P* value
Responded (overall response rate)[a]	12 (60)	15 (75)	0.50 (0.13-1.93)	0.3145	0.57 (0.14-2.37)	.44
Responded and consented (primary outcome)[a]	7 (35)	4 (20)	2.15 (0.52-9.00)	0.2930	1.84 (0.41-8.29)	.43
Responded and declined (secondary outcome)[a]	5 (25)	11 (55)	0.27 (0.07-1.04)	0.0577	0.34 (0.09-1.38)	.13
Response time in days (Secondary outcome)[b]	16.33 (11.12)	19.27 (12.40)	0.85 (0.50-1.43)	0.5361	0.79 (0.47-1.35)	.39

[b]Odds ratio.

[c]Rate ratio.

Overall Response Rate

A total of 27 patients responded to the consent form, either to consent or decline; 12/20 (60%) from the QR code group and 15/20 (75%) from the standard recruitment group. The unadjusted effect estimate was 0.50 (95% CI 0.13 to 1.93, *P*=.31), indicating that the odds of the intervention group responding were 50% lower compared to the control group. After adjusting for age, the OR was 0.57 (95% CI 0.14 to 2.37, *P*=.44). Although the results suggest a lower overall response rate in the QR code group relative to the standard recruitment group, neither analysis was significant.

Proportion of Patients Who Responded and Consented (Primary Outcome)

Overall, 11/40 (27.5%) patients consented to participate in the qualitative study. Seven out of 40 consenting patients were from the QR code group (representing 35% of this group), and 4 were from the standard recruitment group (representing 20% of this group). The unadjusted OR was 2.15 (95% CI 0.52-9.00, *P*=.29), indicating that the odds of patients in the intervention group consenting was 2.15 times the odds among control patients, although this was not statistically significant. After adjusting for age, the OR was 1.84 (95% CI 0.41-8.29). However, neither was significant.

Proportion of Patients Who Responded and Declined to Participate

A total of 16 out of the 40 patients declined to participate in the qualitative study; 5 declining patients were in the QR code group (representing 25% of this group), and 11 were in the standard recruitment group (representing 55% of this group). The unadjusted OR of the QR code group declining to participate was 0.27 (95% CI 0.07-1.04), which was borderline significant (*P*=.06). The age-adjusted OR was 0.34 (95% CI 0.06-1.87, *P*=.21) and was not statistically significant.

Average Time (in Days) Taken to Respond

The average response time was 16 days (SD 11) days for the QR code group and 19 (SD 12) days for the standard recruitment group. The unadjusted incident RR for the QR code group compared to the standard recruitment group was 0.85 (95% CI 0.5-1.43, *P*=.54), indicating that the response time of the QR code group was 15% lower than the response time for the standard recruitment group. After adjusting for age, the RR was 0.79 (95% CI 0.47-1.35), indicating the QR code group response time was 21% lower than the control group. However, neither was statistically significant.

Discussion

Principal Findings

Given the dearth of studies of this nature, the aim of this pilot RCT was to determine the effect on response rates when a QR code was added to a participant consent form for a qualitative research study. Specifically, this pilot investigated the participant response rate, consent rate (primary outcome), rate of decline, as well as the average time taken to respond. Out of 40 patients invited to participate in a qualitative research study were randomized to receive either a standard consent form (control) or a consent form that included a QR code (intervention). The overall response rate (ie, either to consent or decline) was lower in the intervention group relative to the control group, however, the intervention consent rate was higher, the decline rate was lower, and the average time to respond to the invitation to participate was shorter in the intervention group (16 vs 19 days). None of the adjusted estimates of between-group differences were statistically significant. However, our results suggest a trend of patients in the QR code group having higher odds of consenting and lower odds of declining (relative to the control group), but lower odds of responding at all (ie, either to consent or decline). These results should be investigated further in a fully powered RCT.

The overall pattern of results is interesting, despite being nonsignificant. It suggests that the addition of a QR code to a consent form for qualitative research resulted in higher consent rates and lower rates of decline. Both outcomes are desirable for researchers who often spend a large portion of their research budget on recruitment. The shorter response time in the intervention group also suggests that the use of a QR code in recruitment materials could confer benefits to researchers in

XSL•FO
RenderX

terms of more timely information about likely recruitment outcomes.

This is the first pilot RCT to investigate the effect of adding a QR code to a consent form within the context of recruitment of patients with cancer to a qualitative study. It is important to consider the rapidly changing context in which QR code research has taken place when comparing these results to the existing literature. Over the past 5 years, responses to the COVID-19 pandemic have made QR code levels of awareness, familiarity, and skill in the community is now much higher. The more recent RCTs testing the addition of QR codes to surveys, showed small but significant increases, of approximately 1-2% in consent rates [9,12]. The magnitude of the changes observed in the current pilot are much greater (15% increase in consent rates), despite being nonsignificant due to under-powering. This may be due to the targeted nature of the study (ie, people who had previously had colorectal cancer surgery, compared with a general population statewide public health survey) [12] and due to differences in the specific nature of what they were being asked to do (ie, review an intervention designed to support patients with colorectal cancer going through surgery and provide feedback in a 1:1 interview, vs complete a web-based public health survey).

The randomized design is a strength; however, a few limitations should be considered when interpreting the findings of this term. Despite the randomization, due to the small sample size, the groups were not evenly distributed in age, which was subsequently adjusted for. The low sample size is another limitation, although it is still within the recommended range for pilot studies [18]. Therefore, as a next step, a fully powered RCT is recommended. Furthermore, the study did not collect data about whether the participants actually used the QR to respond (only overall response rates) or the characteristics of those users who responded using this channel. This would be valuable information to collect and analyze in future studies to better understand the impact of QR codes on the diversity of recruited participants. Given the context in which this research occurred, it is unclear the extent to which these findings would generalize to the general population, younger people or both groups. Future RCTs should also allow for subgroup analysis to determine the specific effects of QR codes on recruiting different age groups.

Conclusion

Despite the nonsignificance of the findings in this underpowered pilot trial, this research question warrants further investigation. Given their ease of use and inexpensive application, using QR codes on consent forms may hold promise for improving recruitment to qualitative research studies. As recruitment often consumes a large proportion of research budgets, even a small percentage of change could have important benefits for researchers and research programs. QR codes may be a low-risk strategy to complement other attempts to maximize research response rates, although a fully powered RCT is required to confirm this. Further research using larger and more diverse samples is warranted. This pilot study provides important preliminary data on which such studies could be based.

Acknowledgments

The authors acknowledge the patients included in this research, the research assistants of the Health Behavior Research Collaborative (HBRC) and the University of Newcastle (UoN); and Hunter Medical Research Institute (HMRI) staff members who supported this research. The research was supported by NHMRC, UoN, and HMRI grants. The content is solely the responsibility of the authors.

This research was funded as part of a National Health and Medical Research Council Partnership Grant (NHMRC APP1169822). RW is supported by a Heart Foundation postdoctoral fellowship (ID: 102156) and a Cancer Institute NSW Early Career Fellowship (ECF1355).

Data Availability

The datasets generated during this study or analysis of the data are available from the corresponding author on reasonable request and pending ethical approval.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT (Consolidated Standards of Reporting Trials) extension for pilot and feasibility studies checklist.
[PDF File (Adobe PDF File), 91 KB - formative_v9i1e66681_app1.pdf]

References

1. Kasenda B, von Elm E, You J, Blümle A, Tomonaga Y, Saccilotto R, et al. Prevalence, characteristics, and publication of discontinued randomized trials. JAMA 2014 Mar 12;311(10):1045-1051 [FREE Full text] [doi: 10.1001/jama.2014.1361] [Medline: 24618966]

XSL•FO
RenderX

2. Bernardez-Pereira S, Lopes RD, Carrion MJM, Santucci EV, Soares RM, de Oliveira Abreu M, Methodological Evaluation of clinical TriAls Study Group. Prevalence, characteristics, and predictors of early termination of cardiovascular clinical trials due to low recruitment: insights from the ClinicalTrials.gov registry. Am Heart J 2014 Aug;168(2):213-9.e1. [doi: 10.1016/j.ahj.2014.04.013] [Medline: 25066561]
3. Spratling R. Recruitment of medically fragile children and adolescents: lessons learned from qualitative research. J Pediatr Health Care 2013 Jan;27(1):62-65. [doi: 10.1016/j.pedhc.2012.08.001] [Medline: 23084469]
4. Kakumanu S, Manns BJ, Tran S, Saunders-Smith T, Hemmelgarn BR, Tonelli M, et al. Cost analysis and efficacy of recruitment strategies used in a large pragmatic community-based clinical trial targeting low-income seniors: a comparative descriptive analysis. Trials 2019 Oct 07;20(1):577 [FREE Full text] [doi: 10.1186/s13063-019-3652-5] [Medline: 31590686]
5. Biddle N, Gray M, Sollis K. The ANU centre for social research and methods. The use of QR codes to identify COVID-19 contacts and the role of data trust and data privacy. 2021 Nov 5. URL: https://tinyurl.com/438cey75 [accessed 2024-12-01]
6. Cohney S, Cheong M. COVID Down Under: where did Australia?s pandemic apps go wrong? arXiv:2212.12178 2023 Jun 23 [FREE Full text] [doi: 10.1109/ethics57328.2023.10154912]
7. Grekousis G, Liu Y. Digital contact tracing, community uptake, and proximity awareness technology to fight COVID-19: a systematic review. Sustain Cities Soc 2021 Aug;71:102995 [FREE Full text] [doi: 10.1016/j.scs.2021.102995] [Medline: 34002124]
8. Brodie K, Madden LL, Rosen CA. Applications of quick response (QR) codes in medical education. J Grad Med Educ 2020;12(2):138-140 [FREE Full text] [doi: 10.4300/JGME-D-19-00516.1] [Medline: 32322342]
9. Endres K, Heiden EO, Park K. Surv methods insights field SMIF. The Effect of quick response (QR) codes on response rates and respondent composition: results from a statewide experiment. 2024 Aug 05. URL: https://surveyinsights.org/?p=19389 [accessed 2024-12-01]
10. Marlar J. Do Quick Response Codes Enhance or Hinder Surveys?.: Gallup; 2018 Aug 30. URL: https://news.gallup.com/opinion/methodology/241808/quick-response-codes-enhance-hinder-surveys.aspx [accessed 2024-12-02]
11. Lugtig P, Luiten A. Do shorter stated survey length and inclusion of a QR code in an invitation letter lead to better response rates? Surv Methods Insights Field. 2021 Feb 12. URL: https://surveyinsights.org/?p=14216 [accessed 2021-02-12]
12. Endres K, Heiden EO, Park K, Losch ME, Harland KK, Abbott A. Experimenting with QR codes and envelope size in push-to-web surveys. J Surv Stat Methodol 2023 Apr 26;12(4):893-905 [FREE Full text] [doi: 10.1093/jssam/smad008]
13. Gostin I. Forbes Communications Council. How the pandemic saved the QR code from extinction. 2021 May 25. URL: https://tinyurl.com/mka68j87 [accessed 2024-12-02]
14. Zucca A, Mansfield E, Sanson-Fisher R, Wyse R, Johnston SA, Fakes K, et al. Perceived provision of perioperative information and care by patients who have undergone surgery for colorectal cancer: a cross-sectional study. Int J Environ Res Public Health 2022 Nov 18;19(22):15249 [FREE Full text] [doi: 10.3390/ijerph192215249] [Medline: 36429966]
15. Wyse R, Smith S, Zucca A, Fakes K, Mansfield E, Johnston SA, et al. Effectiveness and cost-effectiveness of a digital health intervention to support patients with colorectal cancer prepare for and recover from surgery: study protocol of the RecoverEsupport randomised controlled trial. BMJ Open 2023 Mar 06;13(3):e067150 [FREE Full text] [doi: 10.1136/bmjopen-2022-067150] [Medline: 36878662]
16. Wyse R, Zucca A, Norton G, Johnston SA. The Effect on study consent rates of adding a QR code to patient consent forms: a randomised controlled Pilot Study. 2021 Sep 23. URL: https://doi.org/10.17605/OSF.IO/PJ25X [accessed 2021-09-23]
17. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 2010 Aug;63(8):834-840 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.02.005] [Medline: 20346629]
18. Kunselman AR. A brief overview of pilot studies and their sample size justification. Fertil Steril 2024 Jun;121(6):899-901. [doi: 10.1016/j.fertnstert.2024.01.040] [Medline: 38331310]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
OR: odds ratio
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
RR: rate ratio

XSL•FO
RenderX

Edited by A Mavragani; submitted 19.09.24; peer-reviewed by F Temidayo, M Tandon; comments to author 08.11.24; revised version received 01.12.24; accepted 09.12.24; published 21.02.25.

Please cite as:
Wyse R, Forbes E, Norton G, Viana Da Silva P, Fakes K, Johnston SA, Smith SR, Zucca A
Effect on Response Rates of Adding a QR Code to Patient Consent Forms for Qualitative Research in Patients With Cancer: Pilot Randomized Controlled Trial
JMIR Form Res 2025;9:e66681
URL: https://formative.jmir.org/2025/1/e66681
doi:10.2196/66681
PMID:

©Rebecca Wyse, Erin Forbes, Grace Norton, Priscilla Viana Da Silva, Kristy Fakes, Sally Ann Johnston, Stephen R Smith, Alison Zucca. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Testing and Iterative Improvement of the CEN ISO/TS 82304-2 Health App Quality Assessment: Pilot Interrater Reliability Study

Anna-Lena Frey[1], PhD; Diana Matei[1], MEng; Ben Phillips[1], BSc; Adam McCabe[1], BSc; Rachel Fuller[1], BSc; Begoña Laibarra[2], MSc; Laura Alonso[2], BSc; Victor de la Hoz[2], MEng; Carme Pratdepadua Bufill[3], MSc; Berta Llebot Casajuana[3], BSc; Giuseppe D'Avenio[4], MSc; Pier Angelo Sottile[5], MSc; Laura Melania Rocchi[6], BA; Matteo Errera[6]; Yasmine Laaissaoui[7], MSc; Michael Cardinal[7], PharmD; Menno Kok[8], PhD; Petra Hoogendoorn[9], MSc

[1]Organisation for the Review of Care and Health Apps (ORCHA), Daresbury, United Kingdom

[2]Software Quality Systems, Getxo, Spain

[3]TIC Social Health Foundation, Ministry of Health, Government of Catalonia, Barcelona, Spain

[4]National Center for Innovative Technologies in Public Health, National Institute of Health, Rome, Italy

[5]BI Health Srl, Rome, Italy

[6]BaasBox Srl, Rome, Italy

[7]TherAppX, Granby, QC, Canada

[8]EIT Health Belgium and The Netherlands, Rotterdam, The Netherlands

[9]National eHealth Living Lab, Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands

Corresponding Author:
Petra Hoogendoorn, MSc
National eHealth Living Lab
Department of Public Health and Primary Care
Leiden University Medical Center
Albinusdreef 2
Leiden, 2333 ZA
The Netherlands
Phone: 31 654341785
Email: a.p.y.hoogendoorn@lumc.nl

Abstract

Background: With the increasing use of health apps and ongoing concerns regarding their safety, effectiveness, and data privacy, numerous health app quality assessment frameworks have emerged. However, assessment initiatives experience difficulties scaling, and there is currently no comprehensive, consistent, internationally recognized assessment framework. Therefore, health apps often need to undergo several quality evaluations to enter different markets, leading to duplication of work. The CEN ISO/TS 82304-2 health app assessment seeks to address this issue, aiming to provide an internationally accepted quality evaluation through a network of assessment organizations located in different countries.

Objective: This study aimed to develop and evolve the draft CEN ISO/TS 82304-2 assessment handbook and developer guidance by testing them across organizations in several countries.

Methods: Assessment organizations from 5 countries were engaged to evaluate 24 health apps using the evolving CEN ISO/TS 82304-2 assessment across 3 evaluation rounds. The information submitted by a given health app developer was evaluated by 2 assessment organizations, and interrater reliability was examined. In addition, app developers and assessors were asked to report how much time they spent on information collation or evaluation and to rate the clarity of the developer guidance or assessor handbook, respectively. The collected data were used to iteratively improve the handbook and guidance between rounds.

Results: The interrater reliability between assessment organizations improved from round 1 to round 2 and stayed relatively stable between rounds 2 and 3, with 80% (55/69) of assessment questions demonstrating moderate or better (Gwet AC1>0.41) agreement in round 3. The median time required by developers to prepare the assessment information was 8 hours and 59 minutes (IQR 5.7-27.1 hours) in round 3, whereas assessors reported a median evaluation time of 8 hours and 46 minutes (IQR 7.1-11.0 hours). The draft guidance and handbook were generally perceived as clear, with a median round-3 clarity rating of 1.73 (IQR 1.64-1.90) for developers and 1.78 (IQR 1.71-1.89) for assessors (0=“very unclear”, 1=“somewhat unclear”, and 2=“completely clear”).

XSL•FO
RenderX

Conclusions: To our knowledge, this is the first study to examine the consistency of health app evaluations across organizations located in different countries. Given that the CEN ISO/TS 82304-2 guidance and handbook are still under development, the interrater reliability findings observed at this early stage are promising, and this study provided valuable information for further refinement of the assessment. This study marks an important first step toward establishing the CEN ISO/TS 82304-2 assessment as a consistent, cross-national health app evaluation. It is envisioned that the assessment will ultimately help avoid duplication of work, prevent inequities by facilitating access to smaller markets for developers, and build trust among users, thereby increasing the adoption of high-quality health apps.

(JMIR Form Res 2025;9:e64565) doi:10.2196/64565

KEYWORDS

health apps; mobile health; digital health; quality evaluation; assessment framework; interrater reliability

Introduction

Background

Health apps, which include both medical and wellness apps, have the potential to greatly benefit patients, the public, health care professionals, and the broader health care system [1,2]. They can increase the effectiveness of health care services by supporting behavior change and improving clinical outcomes, enhance efficiency by preventing hospitalization and freeing up resources, and allow for remote monitoring and treatment and, as a result, can have a large positive socioeconomic impact [1,2].

However, alongside these potential benefits, there are several risks associated with the use of low-quality health apps. Previous research has found that many health apps demonstrate data privacy or security risks [3-6], raise clinical safety concerns [7], provide incorrect medical information [8,9], or are not supported by efficacy or effectiveness evidence [4,5,10].

With >350,000 health apps on the market [11], it is difficult for health care professionals and the public to distinguish between beneficial and potentially harmful or ineffective apps. This issue is exacerbated by the fact that readily available metrics such as user ratings and download numbers are not correlated with health app quality [12]. Moreover, concerns regarding effectiveness, safety, usability, and data privacy can be a barrier to the widespread adoption of health apps [13-15], thereby preventing the realization of their potential benefits.

Health app assessments can help address this issue by providing transparent quality information to the public, health care professionals, payers, medical societies, and health care organizations. However, there is currently no comprehensive and consistent assessment that is internationally recognized. While the certification provided under the European Union (EU) Medical Device Regulation (MDR) is valid across EU member states, a large proportion of health apps are not independently evaluated by notified bodies under the MDR. This is because many apps are not classified as medical devices or fall into risk class I, for which app developers only need to declare conformity based on a self-assessment that is not externally validated [16,17]. Furthermore, there are important aspects, such as data privacy, that are not usually independently examined as part of the MDR certification, and research shows that many health apps on the European market do not meet relevant data protection requirements [18]. Finally, the only output provided by the MDR assessment is a CE (European conformity) mark. Consequently, even for health apps that have been examined under the MDR, stakeholders need to carry out additional evaluations to gather more detailed quality information that is required to inform decision-making.

Partly as a result of these limitations of the MDR and similar regulations in other regions, a range of organizations have developed their own assessment frameworks to guide health app–related decision-making. This includes health authorities, charitable and academic institutions, patient and medical professional associations, and commercial companies [19,20]. The resulting fragmented assessment landscape leads to duplicated work for both health systems and app developers if a given app needs to undergo several assessments to evidence its quality for different markets and varying types of organizations.

The CEN ISO/TS 82304-2 health app assessment, which was developed with funding from the European Commission, seeks to address this issue. It aims to provide a comprehensive and consistent quality evaluation that is recognized across Europe and beyond without duplicating the legal tasks of notified bodies under the MDR [21,22].

In this context, a notable challenge is scaling the assessment to potentially cover a substantial proportion of the >350,000 health apps on the market [11]. Notably, a recent review reported that only a median of 45 evaluated digital health technologies, comprising mostly health apps, were available in the assessment libraries of 24 digital health assessment organizations [20]. This limited availability of evaluated apps is likely due to several factors. First, some assessments that require developers to actively submit their apps for evaluation may not offer sufficient incentives to lead to a substantial number of app submissions.For example, incentives may be insufficient if the assessments are not associated with a large enough increase in app users, are not integrated into health care system decision-making (eg, for procurement or reimbursement), or provide reimbursement payment conditions or postprescription procedures that are unappealing (eg, causing delays in patient access to the app) [23-27]. Second, app developers may not be able (due to limited funds and staff in start-ups and small companies) or willing (in light of lacking incentives) to invest the time and money required to meet the quality requirements of individual schemes, such as those requiring new clinical studies, which may affect the number of apps submitted to and passing particular assessments [26,28]. Third, difficulties in scaling the assessment

XSL•FO
RenderX

process may play a role. A review focusing specifically on national health app evaluation initiatives found that even the front-runners in this area, such as Germany, Belgium, and the United Kingdom, are struggling with efficient assessment implementation. This may partly be due to the use of centralized approaches to app evaluation, which can create bottlenecks [29]. It has been proposed that international collaboration and the use of distributed evaluation bodies is needed to address this issue [29]. Following this approach, the CEN ISO/TS 82304-2 assessment is planned to be implemented across a network of assessment organizations in Europe and beyond to achieve efficient scaling while ensuring consistent assessment results across countries (and providing sufficient incentives, as explored in our recent discrete choice experiment [25]).

Objectives

This study aimed to develop and evolve the CEN ISO/TS 82304-2 assessment handbook and developer guidance by testing the assessment across organizations located in different countries. For this purpose, assessment organizations from 5 countries were engaged to evaluate 24 diverse health apps using information submitted by the app developers. In total, 2 organizations assessed each app, and their interrater reliability was examined. In addition, app developers and assessors were asked to report how much time they spent collating or evaluating the assessment information, respectively, and to rate the clarity of the assessment instructions. The collected interrater reliability, time, and clarity rating data were used to iteratively improve the CEN ISO/TS 82304-2 assessment handbook and guidance across 3 evaluation rounds.

Methods

CEN ISO/TS 82304-2 Assessment Framework

The CEN ISO/TS 82304-2 assessment framework is intended to provide an independent, evidence-based, international evaluation of health app quality for a range of health care stakeholders. The framework was created using a Delphi methodology with input from 83 experts from 8 stakeholder groups, including medical organizations, health care authorities, and health app developers across 6 continents [21].

It should be noted that, while the CEN ISO/TS 82304-2 quality requirements are applicable worldwide, some requirements are influenced by local regulations (which is the case for other digital health assessment frameworks as well [29]). This necessitates minor amendments to the assessment across jurisdictions. Given that the project as part of which this study was conducted was funded by the European Commission, the initial CEN ISO/TS 82304-2 testing and implementation efforts were focused on Europe. Therefore, steps were taken to ensure that the CEN ISO/TS 82304-2 criteria are in line with European regulations and standards (as described further in the Discussion section). Future work will aim to support the alignment of the assessment with regulations in other regions.

The CEN ISO/TS 82304-2 framework consists of 5 subsections containing between 7 and 28 questions each: “Product information,” “Healthy and safe,” “Easy to use,” “Secure data,” and “Robust build.” In total, the assessment comprises 81 questions, 12 of which serve to collect background data about the app, mostly to enable the display of this information in a related health app quality label. As the latter questions are answered in free-text or multiple-selection format and do not affect the quality rating of the app, they were not considered in the interrater reliability analysis reported in this paper. The remaining 69 questions (quality requirements) have either 2 (“yes” or “no”) or 3 (“yes,” “no,” or “not applicable”) response options. A list of all assessment questions can be found in the supplement of the CEN ISO/TS 82304-2 framework development paper [21].

Final responses to the assessment questions are determined by an assessment organization based on documentation and other information that is provided by health app developers as part of the assessment process. Examples of the documentation that developers are asked to submit alongside providing access to their apps include evidence of the societal and health benefits of the app, an analysis of health risks, standard operating procedures (or relevant excerpts thereof), and privacy policies.

To facilitate the assessment process, detailed instructions were developed during this study to indicate what documentation is acceptable and what subrequirements need to be met for a given assessment question to be answered with “yes.” These instructions were outlined in a draft developer guidance document and a draft app assessor handbook and were tested and evolved across evaluation rounds (see the App Assessment Procedure section for details). As the handbook is still under development, it is not presented in this paper; however, it will likely be published once the first operational version is finalized.

The final output of the CEN ISO/TS 82304-2 assessment is a quality label [30] that summarizes the assessment results for each quality domain and for the app overall using letter grading (E to A) and color coding (red, orange, yellow, light green, and dark green), similar to the EU energy label. This allows health care professionals and the public to quickly determine whether a given app meets their needs and preferences. For more detailed information, a full assessment report is also available [22].

It should be noted that, as part of this study, no quality labels were issued given that the CEN ISO/TS 82304-2 assessment guidance and handbook were under development.

App Selection

As mentioned previously, conducting the CEN ISO/TS 82304-2 assessment requires health app developers to submit documentation that is not publicly available. Therefore, developers were actively involved in this study, as is described further in the App Assessment Procedure section. This also allowed for the collection of feedback on the assessment process from the app developer perspective.

To thoroughly test the evolving assessor handbook, developer guidance, and assessment process, a variety of health apps were sought to be included in this study, mimicking real-life app diversity. For this purpose, the assessors involved in this study compiled a list of health app development companies to contact, aiming to cover a wide range of company sizes, countries of origin, condition areas, and app functionalities. In addition, this study was advertised on a dedicated project website where

XSL•FO
RenderX

developers had the opportunity to get in touch to express their interest in participating.

To ensure that a given app could be evaluated by any assessor (see the App Assessment Procedure section for details), only apps available in English were eligible for inclusion in the study. However, in 2 cases, it was discovered, after the developers had collated all the assessment information, that their apps did not meet this criterion. Given that considerable effort had already been exerted at this stage, these apps were included, allocating them to assessors with relevant language knowledge or using automated translation tools where needed. No other eligibility criteria were applied apart from a willingness of the app developers to provide the necessary assessment information as well as access to their apps (or, as an exception, relevant screenshots or video demonstrations if access could not be granted, eg, due to geographical restrictions).

The assessment took place across 3 rounds, and the initial aim was to evaluate 8 health apps per round, with funding being available for the evaluation of 24 apps in total. However, due to developers' competing commitments, the amount of effort involved, and the relatively short project timelines, a number of app development companies dropped out of the study after initially agreeing to take part. In round 1, timelines did not permit the recruitment of replacements. In subsequent rounds, more companies than needed were initially recruited to account for potential dropouts. The final sample consisted of the intended total of 24 health apps, which were distributed across the assessment rounds as follows: 3 (12%) apps in round 1 (3 dropouts), 7 (29%) apps in round 2 (7 dropouts and 1 moved to round 3), and 14 (58%) apps in round 3 (10 dropouts).

As mentioned previously, no quality labels were provided for the evaluated apps as part of this study given that the aim was to develop, test, and evolve the CEN ISO/TS 82304-2 assessment guidance and handbook to arrive at an operational version that is suitable for use in quality certification, with sufficiently consistent assessment results. Therefore, no letter grades were calculated or shared. The main benefits offered to participating health app development companies were the receipt of feedback for the individual assessment questions (based on the respective draft versions of the handbook), which could be used internally for product improvements; a chance to familiarize themselves with the CEN ISO/TS 82304-2 assessment; and the offer of a reduced fee if they decided to submit their apps to the operational version of the CEN ISO/TS 82304-2 certification once the latter is rolled out.

App Assessment Procedure

The app evaluations were conducted between November 2022 and April 2024 by assessment organizations with diverse backgrounds from different countries to approximate the full scope of real-life assessment conditions. Specifically, participating organizations were located in the United Kingdom, Italy, Spain, Canada, and France and consisted of 3 private assessment organizations and 2 health authorities, the latter of which each used a subcontractor. In the final round, 1 assessment organization had to drop out as the assessors involved in the study had left the company. Other than that, the same organizations were involved in all 3 evaluation rounds.

Across all organizations, a total of 12 individual assessors were involved in the study, for 10 (83%) of whom background information was available (the remaining n=2, 17%, were based at the organization that had to drop out in the last round). The median previous experience that assessors had with evaluating the quality of health apps using any assessment framework was 19.5 (IQR 0.0-48.0) months. With regard to previous exposure to the CEN ISO/TS 82304-2 assessment in particular, 40% (4/10) of the assessors reported having previously used the CEN ISO/TS 82304-2 assessment to evaluate health apps, and 20% (2/10) of the assessors had been involved in the development of the framework. The remaining 40% (4/10) of the assessors had no previous familiarity with the CEN ISO/TS 82304-2 assessment.

The app assessment procedure across all 3 evaluation rounds was as follows. At the start of each round, the app developers who had agreed to participate in the study were supplied with the draft CEN ISO/TS 82304-2 assessment guidance for round 1, 2, or 3, as applicable. They were asked to collate and submit the required information, including relevant documentation, and grant the assessors access to their health app. In addition, developers were asked to report, for each question, how long it took them to gather the relevant information and how clear they found the associated assessment guidance (red="very unclear," yellow="somewhat unclear," and green="completely clear," coded as 0, 1, and 2, respectively, for the data analysis). Moreover, additional feedback was collected from app developers via individual interviews that were conducted via video calls after the submission of the assessment information but before the receipt of the evaluation results (see the Feedback From Developers section for details).

The assessment information submitted by a given app developer was passed on to 2 of the 5 assessment organizations, which were pseudorandomly allocated in such a way as to achieve diverse organization pairings across apps. A range of expertise was needed to complete the evaluations as the CEN ISO/TS 82304-2 assessment covers clinical, usability, privacy, and technical domains in depth. Therefore, some assessment organizations engaged different assessors to evaluate each of the assessment subsections, whereas others appointed a lead assessor who consulted subject matter experts when needed. The median number of apps evaluated by individual assessors across all 3 rounds with an evolving handbook was 9.5 (IQR 7.0-11.0).

As part of the app evaluation, assessors examined the submitted evidence, which included documentation and screenshots as well as relevant screens within the app itself based on the CEN ISO/TS 82304-2 draft handbook for round 1, 2, or 3 and recorded their responses ("yes," "no," or "not applicable") for each question. Evaluations were conducted independently (ie, different assessment organizations did not have access to each other's responses). Assessors also provided evaluation times and handbook clarity ratings for each of the 81 assessment questions.

After the app evaluations within a given round, meetings were held among the assessors to discuss any disagreements. For each app, a consensus decision on a final response for each

XSL•FO
RenderX

assessment question was reached, which was passed on to the respective app developers. The information gathered during these discussions, together with the interrater reliability results, time and clarity ratings, and the feedback collected from developers, was used to amend and evolve the assessor handbook and developer guidance between rounds (and after the final round). Due to these iterative changes, the results are reported separately for each round.

Data Analysis

Interrater reliability was determined for the agreement between the 2 assessment organizations based on their independent responses recorded before any discussion. For this purpose, Gwet AC1 values [31,32] were calculated for each assessment question in MATLAB (version 9.10.0.1710957 [R2021a]; The MathWorks, Inc). The Gwet AC1 was used as it has been shown to provide stable interrater reliability coefficients without being substantially affected by response prevalence or marginal probability (as is, eg, the case for the Cohen κ). In addition, it is suitable for designs that are not fully crossed over, such as the one used in this study, in which *different pairs* of raters evaluated the varying apps [33].

Gwet AC1 values can be interpreted as follows using Altman benchmark ranges [33]: <0.20="poor agreement," 0.21-0.40= "fair agreement," 0.41-0.60="moderate agreement," and >0.60="good agreement." The percentage of questions falling into each of these categories within each subsection of the CEN ISO/TS 82304-2, as well as across the entire assessment, was reported.

There were some instances in which assessors did not provide evaluation responses for individual questions, possibly because they were uncertain about which option to select (where coinciding clarity ratings were low). These instances were omitted from the analysis, as were a small number of cases in which an assessor responded with “not applicable” to a particular question even though this was not a valid response option. In addition, for 1 app in round 3, the evaluation responses from 1 assessment organization were not available due to technical issues. Therefore, this app was excluded from the interrater reliability analysis. Moreover, due to a miscommunication, 1 app in round 1 was evaluated by 3 assessment organizations instead of 2. As the assessors of one of these organizations were able to see the evaluation answers of another organization, the former’s responses were excluded from the interrater reliability (but not the time or clarity rating) analysis as a precaution to maintain full independence of responses. In subsequent rounds, additional measures were taken to ensure that access to other assessors’ responses was not possible.

For the time and clarity ratings collected from app developers and assessors, summary values were computed in MATLAB. Specifically, for each rating set (from a given rater for a particular app), the sum of the time ratings was calculated across all *questions* within a given section and across the entire assessment. The medians and IQRs of these values were reported as an indication of the central tendency and dispersion of the data *across rating sets.* A similar calculation was applied to the clarity ratings except that, for the latter, the average rather than the sum was computed across the questions within a given section before examining medians and IQRs across rating sets.

It should be noted that some assessors and developers provided time or clarity ratings for <75% of the questions within a particular section or across the entire assessment. These cases were excluded from the relevant section summary or the total assessment summary, respectively, as not enough data were available to arrive at a summary value that was likely to represent the rater’s time investment or perceived clarity (sample sizes are reported in the Results section). Where ratings were missing for <75% of questions, median imputation was applied using the available ratings for a given *question* before calculating the sum or average across all questions for a particular rating set. When calculating the summary values across the entire assessment (for cases in which <75% of all ratings were missing), median imputation was applied to *all* missing values, including to those within the sections that were excluded from the section summary (based on the aforementioned criterion). This was necessary for the calculation of the total amount of time spent on the assessment, for which time ratings were *summed* across *all* questions. The same methodology was applied to the clarity ratings for consistency.

Ethical Considerations

Approval by an ethics or scientific research committee was deemed unnecessary according to Dutch national regulations, since the study was not subject to the Medical Research Involving Human Subjects Act, and according to the guidelines of the Central Committee on Research Involving Human Subjects. The study did not require ethics approval as no personal or sensitive data were collected and all activities performed by assessors and developers lay strictly within their professional competence and job responsibilities. Moreover, no identifiable information is disclosed about individual health apps in this paper.

Results

App Characteristics

Of the 24 apps included across the 3 assessment rounds, 8 (33%) were developed by micro-sized start-ups with <10 employees, 12 (50%) were developed by small businesses with 10 to 49 employees, 2 (8%) were developed by medium-sized companies with 50 to 249 employees, and 2 (8%) were developed by large organizations with ≥250 employees. Most companies (22/24, 92%) were based in Europe, that is, in the Netherlands (6/22, 27%), Belgium (4/22, 18%), and France (3/22, 14%), as well as in Bulgaria, Denmark, Finland, Greece, Italy, Portugal, Spain, Switzerland, and the United Kingdom (1/22, 5% each). The remaining companies (2/24, 8%) had headquarters in the United States.

The sample included both disease-specific and more general health or health system service support apps, with the former targeting a wide range of condition areas such as diabetes, asthma, cardiovascular disease, mental health, chronic pain, nicotine dependence, cognitive impairment, hepatitis C, and colorectal conditions. Most apps had more than one intended use, including health system service support (3/24, 12%),

XSL•FO
RenderX

information provision (12/24, 50%), disease prevention (eg, through behavior change; 7/24, 29%), self-management (16/24, 67%), monitoring (15/24, 62%), treatment (4/24, 17%), communication with health care professionals (14/24, 58%), calculation of values that are likely to affect health care decisions (6/24, 25%), and research support (3/24, 12%). Overall, 46% (11/24) of the apps were classified as medical devices under the MDR.

Interrater Reliability Between Assessment Organizations

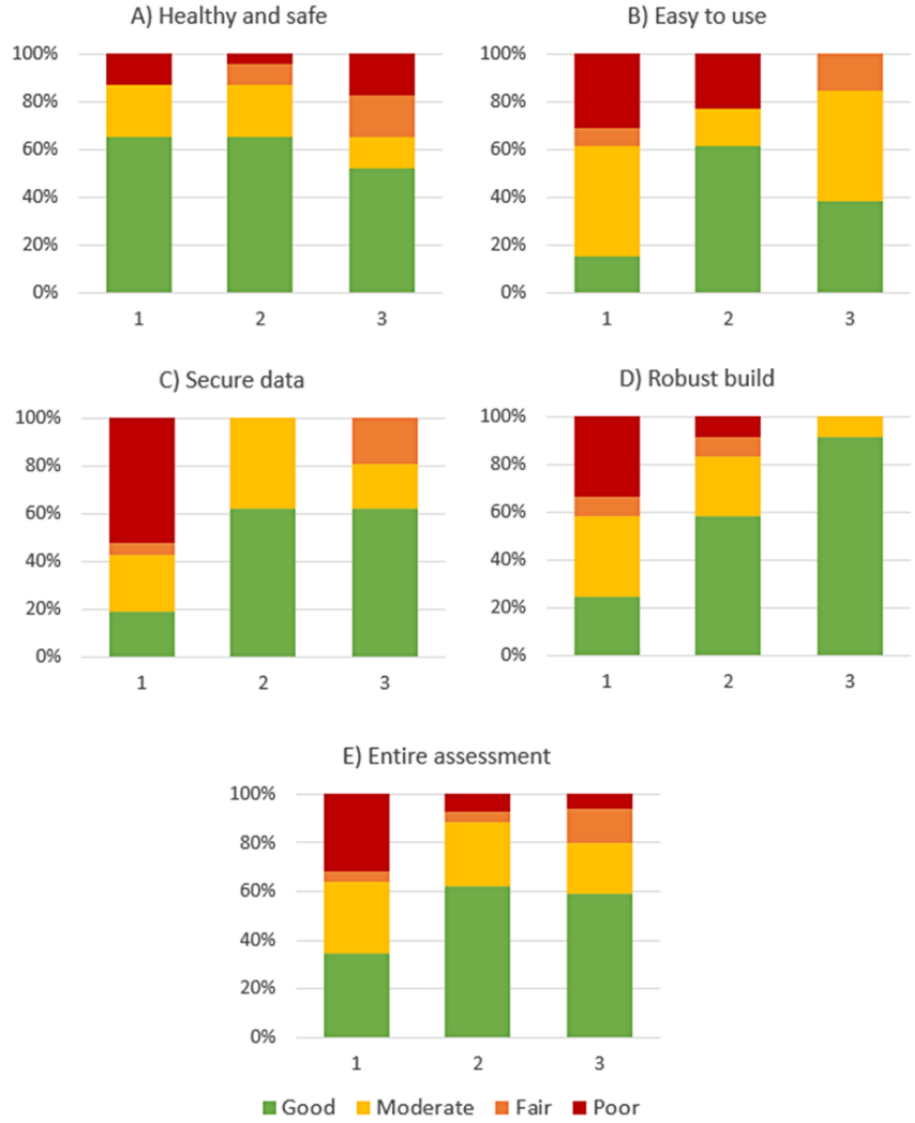
For round 1, a median Gwet AC1 value of 0.54 (IQR 0.14-1) was observed across all 69 CEN ISO/TS 82304-2 assessment questions included in the analysis, indicating moderate overall agreement between assessment organizations. Breaking this finding down further revealed that the interrater reliability was poor (<0.20) for 32% (22/69) of the questions, fair (0.21-0.40) for 4% (3/69) of the questions, moderate (0.41-0.60) for 29% (20/69) of the questions, and good (>0.60) for 35% (24/69) of the questions.

In round 2, the overall interrater reliability between assessment organizations was good, with a median Gwet AC1 value of 0.63 (IQR 0.48-0.84). Agreement was poor (<0.20) for 7% (5/69) of the questions, fair (0.21-0.40) for 4% (3/69) of the questions, moderate (0.41-0.60) for 26% (18/69) of the questions, and good (>0.60) for 62% (43/69) of the questions.

In round 3, a median Gwet AC1 value of 0.67 (IQR 0.42-0.85) indicated good overall agreement between organizations. Interrater reliability was poor (<0.20) for 6% (4/69) of the questions, fair (0.21-0.40) for 14% (10/69) of the questions, moderate (0.41-0.60) for 20% (14/69) of the questions, and good (>0.60) for 59% (41/69) of the questions.

A visual representation of the interrater reliability findings by CEN ISO/TS 82304-2 subsection can be found in Figure 1.

Figure 1. Percentage of questions falling into each interrater reliability category for the agreement between assessment organizations (Gwet AC1≤0.20="poor agreement," 0.21-0.40="fair agreement," 0.41-0.60="moderate agreement," and >0.6="good agreement"). Results are shown for the following assessment sections (A) "Healthy and safe" (23 questions), (B) "Easy to use" (13 questions), (C) "Secure data" (21 questions), and (D) "Robust build" (12 questions), as well as for (E) the Entire assessment (69 questions). Note that the "Product information" section only contained free-text and multiple-response options and, therefore, was not included in the interrater reliability analysis.



XSL•FO
RenderX

Time and Clarity Ratings

Assessor Ratings

The median evaluation time across all CEN ISO/TS 82304-2 questions reported by the assessors was 10 hours and 34 minutes (IQR 9.5-13.1 hours) in round 1 (n=5 rating sets), 8 hours and 33 minutes (IQR 4.9-9.0 hours) in round 2 (n=11 rating sets), and 8 hours and 46 minutes (IQR 7.1-11.0 hours) in round 3 (n=19 rating sets). The amount of time spent on each CEN ISO/TS 82304-2 subsection is presented in Table 1. It should be noted that these ratings include only the time it took to arrive at a response to the assessment questions based on the submitted evidence but not the time dedicated to other activities, such as meetings.

The median assessor handbook clarity rating across all questions was 1.46 (IQR 1.44-1.65) in round 1 (n=6 rating sets), 1.72 (IQR 1.53-1.88) in round 2 (n=10 rating sets), and 1.78 (IQR 1.71-1.89) in round 3 (n=12 rating sets), where 0="very unclear", 1="somewhat unclear", and 2="completely clear". Clarity ratings broken down by subsection can be found in Table 2.

Table 1. Evaluation time reported by the assessors across rounds for each subsection of the CEN ISO/TS 82304-2 assessment.

Section and round	Number of rating sets	Evaluation time (hours), median (IQR)
Product information		
1	5	0.6 (0.5-0.7)
2	11	0.5 (0.2-0.5)
3	19	0.4 (0.3-0.7)
Healthy and safe		
1	5	3.6 (3.2-4.3)
2	11	3.2 (1.7-3.4)
3	19	3.1 (2.1-3.7)
Easy to use		
1	5	1.9 (1.7-3.1)
2	11	1.5 (0.8-1.8)
3	19	1.8 (1.3-2.3)
Secure data		
1	5	2.9 (2.6-3.4)
2	11	2.0 (1.2-2.3)
3	19	2.6 (1.0-3.3)
Robust build		
1	4	1.5 (1.2-1.9)
2	10	1.1 (0.7-1.4)
3	18	1.1 (0.1-1.7)

XSL•FO
RenderX

Table 2. Handbook clarity ratings reported by the assessors across evaluation rounds for each subsection of the CEN ISO/TS 82304-2 assessment.

Section and round	Number of rating sets	Assessor clarity rating (0=very unclear, 1=somewhat unclear, and 2=completely clear), median (IQR)
Product information		
1	6	2 (2-2)
2	9	2 (2-2)
3	14	2 (2-2)
Healthy and safe		
1	6	1.54 (1.5-1.57)
2	10	1.84 (1.29-1.96)
3	12	1.88 (1.7-1.97)
Easy to use		
1	6	1.54 (1.31-1.54)
2	10	1.62 (0.62-1.77)
3	9	1.54 (1.17-1.94)
Secure data		
1	6	1.5 (1.24-1.67)
2	10	1.9 (1.62-2)
3	8	1.74 (1.52-2)
Robust build		
1	6	1.21 (1.08-1.83)
2	10	2 (1.67-2)
3	13	1.92 (1.67-2)

App Developer Ratings

The median time spent by app developers on preparing the assessment information using the evolving guidance was 2 hours in round 1 (n=1 rating set), 7 hours and 3 minutes (IQR 3.8-8.8 hours) in round 2 (n=5 rating sets), and 8 hours and 59 minutes (IQR 5.7-27.1 hours) in round 3 (n=10 rating sets). The information collation time for each CEN ISO/TS 82304-2 subsection can be found in Table 3.

The median guidance clarity rating indicated by app developers across all questions was 1.84 (IQR 1.69-1.99) in round 1 (n=2 rating sets), 1.80 (IQR 1.74-1.87) in round 2 (n=5 rating sets), and 1.73 (IQR 1.64-1.90) in round 3 (n=12 rating sets). Clarity ratings broken down by subsection are presented in Table 4.

XSL•FO
RenderX

Table 3. Information collation time reported by app developers across rounds for each subsection of the CEN ISO/TS 82304-2 assessment.

Section and round	Number of rating sets	Information collation time (hours), median (IQR)
Product information		
1	1	0.06 (—[a])
2	5	0.22 (0.16-0.31)
3	10	0.48 (0.27-0.82)
Healthy and safe		
1	1	0.5 (—)
2	5	2.14 (1.32-3.4)
3	10	4.11 (2.63-7.65)
Easy to use		
1	1	0.26 (—)
2	5	1.53 (0.8-2.22)
3	8	1.8 (0.96-3.75)
Secure data		
1	1	0.92 (—)
2	5	1.32 (1.1-2.13)
3	10	1.56 (0.99-8.5)
Robust build		
1	1	0.28 (—)
2	5	0.77 (0.38-1.46)
3	9	0.9 (0.63-3.45)

[a]Not applicable.

XSL•FO
RenderX

Table 4. Guidance clarity ratings reported by app developers across evaluation rounds for each subsection of the CEN ISO/TS 82304-2 assessment.

Section and round	Number of rating sets	Developer clarity ratings (0=very unclear, 1=somewhat unclear, and 2=completely clear), median (IQR)
Product information		
1	3	2 (1.36-2)
2	5	2 (1.93-2)
3	13	2 (1.96-2)
Healthy and safe		
1	2	1.82 (1.64-2)
2	5	1.93 (1.56-1.97)
3	12	1.87 (1.65-1.92)
Easy to use		
1	2	1.81 (1.62-2)
2	5	1.92 (1.65-2)
3	12	1.71 (1.65-1.88)
Secure data		
1	2	1.98 (1.95-2)
2	5	1.95 (1.6-1.96)
3	12	1.81 (1.67-1.95)
Robust build		
1	2	1.58 (1.25-1.92)
2	5	1.83 (1.71-1.94)
3	12	1.67 (1.33-1.88)

Feedback From App Developers

All 24 app developers took part in individual follow-up interviews, which were conducted vial video calls after the developers had submitted their assessment information but before they received the evaluation responses from the assessors. The interviews were led by MK and took between 25 and 75 minutes. On the developer side, between 1 and 5 team members attended the interviews, which in all cases included individuals who had been involved in collating the assessment information and in 92% (22/24) of the cases included the person leading this process.

When asked what improvements they would recommend for the draft version of the assessment guidance and the evolving evaluation process, developers provided several suggestions. First, in rounds 1 and 2, it was noted that some questions were mainly applicable to medical apps, causing uncertainties about whether and, if so, how they could be answered for wellness apps. It was suggested to include stratification points in the assessment to determine which questions need to be answered for which type of app (which is part of the CEN ISO/TS 82304-2 framework but had not yet been applied in the draft guidance). Second, uncertainties were reported regarding what evidence was required for some of the questions. It was indicated that the provision of example answers and document templates, as well as the opportunity to interact with assessors as needed (which was always intended for the real-world assessment implementation), would be helpful to clarify requirements. Third, it was noted, especially by companies that outsourced their software development, that it would be useful to receive more explicit information on which expertise (eg, clinical, data privacy, cybersecurity, and software development) was needed to answer each individual question to allow for a more efficient distribution of the questions across the team and subcontractors. Fourth, developers asked for guidance on how to efficiently integrate CEN ISO/TS 82304-2 into their internal processes and strategy alongside other required assessments. In this context, it was noted that there is overlap between MDR and CEN ISO/TS 82304-2 requirements (which was deliberate and necessary to cover relevant quality criteria for apps that are not classified as medical devices or, given their class-I classification, not assessed by notified bodies). It was suggested to allow for the reuse of documentation or the automatic passing of certain CEN ISO/TS 82304-2 questions if an app had already been assessed by a notified body (which was indeed always intended but requires further consideration for effective implementation).

Developers were further asked which, if any, benefits they thought the CEN ISO/TS 82304-2 assessment and the associated quality label could provide for their company. In response to this question, 67% (16/24) of the developers indicated that they expected the quality label to be an asset for marketing purposes, in addition to creating confidence and trust among users, payers, and health care providers. A further 25% (6/24) of the developers primarily mentioned receiving an independent quality

XSL•FO
RenderX

certification as a benefit in and of itself. Moreover, 21% (5/24) of the developers saw the assessment process as a valuable learning experience for the company, offering guidance on how to measure and improve the quality of their app, with 1 company reporting that they created an internal quality management system as a result of going through the assessment process.

Finally, when asked whether, knowing what they did now about the time and effort that was required, they would still have chosen to submit their app for CEN ISO/TS 82304-2 assessment, 88% (21/24) of the developers replied in the affirmative, with 8% (2/24) being unsure and 4% (1/24) indicating that they would not have submitted their app.

Discussion

Principal Findings

This study aimed to develop and evolve the CEN ISO/TS 82304-2 assessment handbook and developer guidance by testing them across a diverse range of apps and assessment organizations from different countries. Time spent on information collation or evaluation, instruction clarity, and assessment consistency were examined to inform the next steps toward establishing a robust certification process. To our knowledge, this is the first cross-national, cross-stakeholder collaboration in the field of health app assessment of this extent, diversity, and rigor.

This study revealed that the interrater reliability between assessment organizations improved from round 1 to round 2 for all assessment sections, likely reflecting both an initial practice effect and clarifications made to the assessor handbook. Between rounds 2 and 3, interrater agreement improved or stayed relatively stable for the “Robust build” and “Secure data” sections, respectively. However, unexpectedly, the “Healthy and safe” and “Easy to use” sections demonstrated somewhat lower interrater reliability in round 3 than in round 2 (which likely also drove the very slight decrease in interrater reliability observed across the entire assessment between these rounds). The evolving rigor of the handbook between rounds likely contributed to this finding, which may have affected some sections more than others depending on the changes that were made. Moreover, differences in app characteristics may have also played a role. For instance, in round 2, only 29% (2/7) of the evaluated apps were classified as medical devices, whereas in round 3, this classification applied to 43% (6/14) of the apps. The greater complexity of medical device app functionalities may have made the latter more difficult to evaluate in some (but not all) domains compared to apps that are not medical devices. In the future, to the extent possible, the assessment outcomes of notified bodies will be used to inform those CEN ISO/TS 82304-2 quality requirements that overlap with the MDR for medical device apps, which will increase the consistency of the latter’s assessment. In addition, the improvements discussed in the Learnings and Planned Improvements to the Assessment section aim to ensure consistent evaluation across all assessment sections and app types. Still, even with the round-3 draft version of the assessment handbook, 59% (41/69) of the questions across the entire assessment demonstrated good interrater reliability, with another 20% (14/69) of the questions showing moderate agreement. These results are promising, indicating that, even at this early stage, the CEN ISO/TS 82304-2 assessment already results in largely consistent outcomes across organizations located in different countries. The implications of this finding are discussed further in the Envisioned Impact of the CEN ISO/TS 82304-2 Assessment section.

The median handbook clarity ratings reported by assessors increased steadily across rounds, as was expected due to increasing familiarity with the handbook and the improvements made to the latter between rounds based on the collected data. Somewhat surprisingly, the median guidance clarity ratings collected from app developers decreased across rounds. This may have been an unintended consequence of the iterative changes (increasing rigor) that were made to the guidance. For instance, in round 3, the subrequirements that were used in the assessor handbook were added to the developer guidance with the aim of clarifying evidence needs and assessor expectations. However, this may inadvertently have caused uncertainty among developers about how to answer the associated assessment questions, with seemingly only a minority of developers having taken the individual subrequirements into account. Alternatively, especially given the small sample size, the observed differences across rounds may have been a result of variations in the included apps and development companies, which were likely to affect how easily certain assessment questions could be answered. In either case, clarification of the written guidance, giving developers the opportunity to contact the assessment organizations with clarification questions, and other improvements discussed in the Learnings and Planned Improvements to the Assessment section will likely help enhance developers’ understanding of the guidance in the future.

With regard to the time spent on the assessment process, in round 3, developers reported taking a median of 8 hours and 59 minutes (IQR 5.7-27.1 hours) to gather the assessment information, and assessors indicated spending a median of 8 hours and 46 minutes (IQR 7.1-1.0 hours) evaluating the apps. While this is a substantial time investment, it reflects the rigor of the CEN ISO/TS 82304-2 assessment, which is intended to ensure that different stakeholder groups internationally are provided with sufficient information to support their health app–related decision-making, significantly reducing the need for further evidence gathering and avoiding similar evaluations. Importantly, it should be noted that, despite the effort involved, 88% (21/24) of the app developers reported that they would submit their app for CEN ISO/TS 82304-2 assessment again, indicating that the perceived benefits of the assessment outweigh the time cost. In addition, several steps are being considered to reduce the required time for both developers and assessors in the future implementation of the assessment, as discussed in the Learnings and Planned Improvements to the Assessment section.

All in all, this study provided promising results regarding the interrater reliability between assessors at this early stage of development and testing, as well as valuable insights to support the future implementation of the CEN ISO/TS 82304-2 assessment. In the following sections, it is discussed how these results compare to previous findings; which improvements were made based on this study; what next steps are planned for the

XSL•FO
RenderX

refinement of the CEN ISO/TS 82304-2 assessment handbook, developer guidance, and process; and what benefits the widespread use of the assessment is expected to provide. Strengths and limitations of the study are also considered.

Comparison to Previous Health App Assessment Interrater Reliability Studies

Previous studies that have examined the interrater reliability of health app assessments have yielded mixed results. For several assessments, <15% of the questions demonstrated good agreement between assessors (eg, Fleiss κ>0.60 or Krippendorff α>0.667) [34-36]. Our results compare favorably to these findings, with approximately 60% of CEN ISO/TS 82304-2 questions in rounds 2 and 3 (43/69, 62%, and 41/69, 59%, respectively) showing good assessor agreement (Gwet AC1>0.60), already at this early stage of development. However, it should be noted that higher interrater reliability has been reported for several health app assessments, such as outstanding agreement (κ>0.80) for 86% of the Enlight questions [37], substantial or good agreement (κ>0.60) for 79% of the questions within the American Psychiatric Association's app evaluation framework [38], and satisfactory intraclass correlations (>0.75) for all sections of the Mobile App Rating Scale [39].

In this context, the complexity of the assessments should be considered. The aforementioned frameworks are less comprehensive in scope, rely only on publicly available information (mostly the app itself), and include less complex and often more descriptive questions than those in the CEN ISO/TS 82304-2 assessment. Therefore, it is unsurprising that good interrater reliability could be achieved relatively easily for these frameworks. In contrast, the in-depth evaluation of clinical and technical aspects provided as part of the CEN ISO/TS 82304-2 assessment requires expert assessment knowledge and careful examination of elaborate documentation submitted by developers. Consequently, it is to be expected that repeated testing and iterative improvement are required to achieve a clear handbook and good assessor agreement for all CEN ISO/TS 82304-2 questions. In line with this, a recent qualitative study found that, even for the MDR, there are significant variations in the interpretation of the included quality requirements both within and between notified bodies, which can result in inconsistent assessment outcomes [40]. This highlights the difficulty of achieving consistency for more complex digital health technology quality assessments. In this context, it is worth noting that, unlike some other quality or efficiency assessments (eg, those carried out for the EU energy label), digital health technology assessments involve evaluations carried out by assessors rather than relying only on automated measurements (eg, of energy consumption). This introduces variability, which needs to be minimized to ensure that national authorities, citizens, and health care professionals can trust that health apps are evaluated consistently. The refinements discussed in the Learnings and Planned Improvements to the Assessment section aim to further improve the CEN ISO/TS 82304-2 assessment consistency. This includes amendment of the handbook and guidance as well as considerations related to the expertise, training, and support from subject matter experts and automated tooling that can assist assessors in achieving consistent ratings.

With regard to the time required for app evaluation, the few previous interrater reliability studies that have examined this measure have reported assessment times ranging from 15 minutes to 1 hour [34,41]. These values are considerably lower than those observed in this study, with a median evaluation time of 8 hours and 46 minutes (IQR 7.1-11.0 hours) reported in round 3. Again, this difference is likely due to the reliance only on publicly available information and the lower complexity of the former assessments compared to that of the CEN ISO/TS 82304-2. Pilot studies more similar to this one, as part of which information was requested from health app developers, have reported app evaluation times more comparable to those in our study. Specifically, one study observed assessment times of 10 to 30 work hours per app (including writing the recommendation, meetings, and other communications [42]), and another study (including 2 apps) reported indicative app evaluation times of up to 6 hours (if uncertainties were encountered [43]). The time information gathered during this study will be used to inform developers of how long it is expected to take to collate the assessment information (developer workload), determine assessment pricing, and guide scalability efforts (assessor workload).

Learnings and Planned Improvements to the Assessment

On the basis of the information collected during this study and complementary studies conducted as part of the initiative that this study is a part of, improvements were made to further increase the consistency, clarity, efficiency, and value of the assessment. Several of these learnings may also be applicable to other health app evaluation initiatives.

First, to increase consistency, the handbook was further evolved to include separate sections with standardized, simplified wording. These sections provide additional guidance as to (1) the documentation needed from app developers; (2) which subrequirements the assessors need to evaluate with the provided documentation and, if applicable, with information from the app itself; and (3) which subrequirements need to be met for the assessment question to be answered with a "yes." In addition, it is being explored whether automation can be introduced to evaluate or support the evaluation of some requirements, such as testing accessibility features, examining data flow, and scanning privacy policies for relevant content. In addition to improving consistency, automation can help reduce the required assessment time and may enable the evaluation of additional quality aspects that cannot be examined manually. Moreover, in the future, the outcomes of trusted existing assessments, in particular those provided by notified bodies as part of their legal task, will, to the extent possible, be used to address the related CEN ISO/TS 82304-2 requirements, which is expected to improve consistency as well as avoid duplication of effort. Finally, consideration is being given to how best to train new assessors (eg, using practice evaluations on already assessed apps) to ensure that they demonstrate a high level of consistency before starting to conduct "real" app assessments.

Second, to increase clarity and efficiency, a fit-for-purpose platform, including subsection headings, stratification, and documentation templates, is intended to be developed to collect

XSL•FO
RenderX

assessment information from developers. Subsection headings will make the expertise needed to answer different questions (eg, medical, data privacy, or cybersecurity knowledge) more apparent. This can help developers distribute the assessment questions more effectively across the company and potential suppliers, asking relevant experts to answer the questions most suited to their expertise, thereby saving time in collating the assessment information. Stratification, carried out automatically via the platform, will more efficiently determine which assessment questions need to be answered based on the intended use and characteristics of the app, such as whether it provides health information, has already been assessed by a notified body, processes personal data, or incorporates application programming interfaces. Document templates, which interviewed developers suggested would be helpful, will further clarify the evidence requirements and allow developers to identify more quickly and effectively which information they need to provide. This, in turn, will ensure that assessors receive all and only the information needed for the evaluation in a standard format, speeding up the evaluation process and making it more robust.

Efficiency is also expected to be increased by giving developers and assessors the opportunity to interact as part of the assessment process (as was always intended). This was not the case during the study to enable the examination of developers' understanding of the guidance in the absence of outside advice and to ensure that assessors' evaluation responses were independent and based on the same information. Due to the lack of interaction during this study, developers were occasionally uncertain about how to respond to particular questions or what documentation to provide, and in some instances, assessors spent considerable amounts of time searching for relevant information, including within the apps. In the real-world implementation of the assessment, developers will have the chance to ask clarification questions, and assessors will be able to contact developers to request additional information if needed. It is expected that such interactions will reduce uncertainty and expedite the evaluation process. Future consideration will additionally be given to how to efficiently handle reassessments where changes to an app have been made that affect the responses to individual quality requirement questions and how to guide developers in efficiently integrating CEN ISO/TS 82304-2 into their internal processes and strategy.

Third, to increase the value and uptake potential of the CEN ISO/TS 82304-2 assessment, steps were taken to evolve and align it more specifically with the European context, where the initial implementation efforts are intended to take place. This included incorporation of findings of a comparative analysis with (mostly) European health technology assessment frameworks, EU-level legislation and values, international and European standards, advice from subject matter experts, literature on health app quality, and information needs reported by (mostly) European stakeholders for health app–related decision-making. Consultation of related European authorities such as the Medical Device Coordination Group, the European Data Protection Supervisor, and the EU Agency for Cybersecurity is proposed for the future to obtain feedback and ensure sustained alignment. In addition, future work will aim to align the assessment with regulations and standards in other regions as well. Notably, it has been observed that many regulations are promulgated from the EU to other regions, where they become entrenched in relevant legal frameworks (which has been termed the "Brussels effect" [44]). Therefore, by aligning the handbook with EU regulations such as the MDR and General Data Protection Regulation, many (often less rigorous) legal requirements in other regions are likely already covered, and future efforts will likely only require relatively small additions (or removal) of requirements to align with local legislation. Once the assessment is taken up in different regions, a flag displayed at the top of the CEN ISO/TS 82304-2 quality label can indicate in which region the assessment was conducted (or recognized). This is similar to the EU energy label, which is used by many countries outside the EU with some, strong, or full alignment and displays the region or country flag at the top of the label [45].

Finally, it should be noted that the initial focus of the development and testing effort was on smaller organizations and start-ups, assuming that, if the assessment process was optimized for the latter, it would also be suitable for larger companies that tend to employ more subject matter experts. However, this study revealed that larger organizations may require somewhat different guidance from that required by small or medium-sized businesses due to differing organizational structures and interrelations between products and associated documentation. For instance, in large companies, a given health app may be a supportive tool for a medical intervention that is not delivered through the app itself, and documentation for these connected products may not be easily separable. Relatedly, certain policies in place at large companies, such as data privacy provisions, may be applicable to many offered products rather than being health app specific. In addition, in some cases, larger companies may have (recently) acquired smaller app development organizations, with only some but not all documentation already having been updated to align with the larger company's standard procedures. Such scenarios may cause uncertainties within large companies regarding what documentation is suitable and relevant for the app-specific CEN ISO/TS 82304-2 assessment if the evaluation guidance is mainly aligned with the needs of smaller organizations. As only 2 large companies were involved in this pilot study, additional testing is needed and planned to calibrate the assessment guidance for use in large organizations.

Envisioned Impact of the CEN ISO/TS 82304-2 Assessment

As mentioned in the Introduction section, the need for a health app quality certification beyond that provided under the MDR seems to be widely recognized [15,19,20], but current health app assessment initiatives, including national programs, are struggling to achieve efficient, appealing, and scalable assessment implementation. This may partly be due to the use of centralized national approaches to app evaluation, which can create bottlenecks [29], as well as due to insufficient incentives for developers. It has been proposed that international collaboration and the use of distributed evaluation bodies is needed to address these issues [29]. As such, the envisioned establishment of the CEN ISO/TS 82304-2 assessment as an

XSL•FO
RenderX

efficient, effective, consistent, and scalable health app quality certification scheme that is delivered through a network of assessment organizations and is recognized across Europe and beyond, requiring only limited context-specific requirements to be added, would be a major step forward.

It is expected that the widespread availability of the CEN ISO/TS 82304-2 quality label will minimize risk and build trust among health care professionals and the public, thereby increasing the adoption of high-quality health apps. In line with this expectation, our recent research has shown that health care professionals' willingness to recommend health apps and users' intention to download them is positively influenced by the CEN ISO/TS 82304-2 quality label ([46]; Adriaanse, unpublished data).

In addition, the cross-national recognition of the CEN ISO/TS 82304-2 assessment would help address several issues associated with the current fragmented assessment landscape in which numerous organizations apply differing assessment frameworks for overlapping quality aspects [19,20]. This fragmentation leads to duplicated work both for health systems and app developers if a given app needs to be submitted to several full assessments to evidence its quality to different markets and varying types of organizations. In this regard, it is specifically problematic that the quality criteria used can differ widely across frameworks, with a recent review finding that, among 12 health system- and government-led digital health quality assessments, the criteria overlap in some domains was <50% [20]. Such inconsistencies in quality standards can cause additional work (if varying evidence needs to be prepared and submitted) and confusion among both developers and users, potentially jeopardizing trust in quality evaluations if the same health app may fail one assessment but pass another. In addition, the need for an app to undergo several assessments to enter various markets may lead to inequities. While developers may regard it as worthwhile to invest time and money to receive quality certifications in countries with large populations, this is less likely to be the case for smaller markets, for which the cost-benefit ratio of undergoing an additional assessment and addressing local evidence needs, such as a study conducted within the related country, may not be appealing. This may lead to a situation in which some high-quality apps are not available in certain regions or languages.

The CEN ISO/TS 82304-2 assessment is intended to address these issues, avoiding duplication of work, preventing inequities from arising, and building trust by serving as a "gold standard" assessment that is recognized across a wide range of countries and types of organizations. A recent paper with the European Society of Cardiology [47] and preliminary results from a comparative analysis with European health technology assessment frameworks (Hoogendoorn, unpublished data) indicate that the CEN ISO/TS 82304-2 assessment covers the vast majority of criteria that are regarded as important to evidence health app quality. By primarily relying on the CEN ISO/TS 82304-2 quality label and only carrying out additional, context-specific evaluations if the label indicates that a given app is of acceptable quality, health systems and other stakeholders can obtain substantial time and resource savings during health app–related decision-making and create streamlined processes for developers.

Strengths and Limitations

This study took an innovative and ambitious approach to health app quality assessment by engaging organizations from 5 countries in the evaluation process and examining their interrater reliability. In addition, 24 health app development companies were involved in this study, and their feedback was sought. To our knowledge, this is the first such collaboration, which lays the foundation for the consistency and scaling of the CEN ISO/TS 82304-2 assessment across Europe and beyond.

However, some limitations should be noted. First, the sample size of included apps was small (given that each app was rated by only 2 assessment organizations), limiting the generalizability of the findings. This was partly due to the funding and time constraints of the project, as well as due to developer dropouts, especially in the first assessment round. Given the small sample involved in this pilot study and the evolving nature of the guidance and handbook throughout the study and beyond, future studies with a larger sample and using the operational version of the guidance and handbook are required to provide conclusive evidence for the interrater reliability of the CEN ISO/TS 82304-2 assessment.

Second, as mentioned previously, relatively high developer dropout rates were observed, which may have distorted the findings if developers who found the guidance more difficult to understand were more likely to drop out. However, this is unlikely to have been the case as many of the dropouts occurred before developers had even received the guidance, and across all rounds, only 3 developers dropped out after they had started to collate the assessment information. In this regard, it should also be emphasized that developers were not provided with a quality label as part of this study given that the aim was to develop and refine the CEN ISO/TS 82304-2 assessment handbook and guidance before arriving at an operational version that is suitable for use in quality certification. This likely contributed to the dropouts given that, as indicated during the developer interviews, receiving a quality label to build trust and support marketing efforts is the main benefit of the assessment for developers. It is expected that dropouts will be less common in the future implementation of the assessment once a quality label is issued and developers actively seek out the assessment of their apps to gain the benefits of the label.

Third, given that this study aimed to approximate real-world conditions with a focus on app, developer, and assessor diversity, some aspects were not as tightly controlled as they could have been, such as the application of eligibility criteria or the number of individual assessors involved in the evaluation of a given app across organizations. This was deemed necessary to allow for the detection and solution of potential issues that may arise during the real-life assessment process. Moreover, this approach ensured that the gathered consistency, time, and clarity data were representative of what would be expected in the real world (with the draft handbook and guidance) without being (potentially positively) distorted by artificially imposed conditions that will not be present once the assessment is rolled out.

XSL•FO
RenderX

Conclusions

This study involved an unprecedented collaboration across organizations in 5 countries to develop, test, and evolve the CEN ISO/TS 82304-2 health app quality assessment handbook and guidance. Promising findings were obtained, especially given the early stage of development, with regard to the interrater reliability between assessors, and valuable information was gathered to support the future evolution and implementation of the assessment. These are important steps toward establishing the CEN ISO/TS 82304-2 assessment as a commonly recognized health app quality evaluation across Europe and beyond. The widespread use of the CEN ISO/TS 82304-2 assessment across countries is expected to have a substantial positive impact, avoiding duplication of work, preventing inequities by facilitating access to smaller markets for developers, enhancing the quality of apps, and building trust in health apps among health care professionals and the public. This can ultimately help increase the adoption of high-quality health apps that have the potential to improve health outcomes and health care system efficiency.

Acknowledgments

The authors are grateful to the health app developers and assessment organizations that contributed to the testing of the CEN ISO/TS 82304-2 assessment. This study was funded by the European Union. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or of the European Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

ALF programmed the MATLAB (The MathWorks, Inc) script, analyzed the data, and wrote the manuscript draft. DM and RF led the project administration, and ALF, DM, BP, and RF curated the data. DM, BP, AM, RF, BL, LA, VdlH, CPB, BLC, GD, PAS, LMR, ME, YL, and MC played key roles in the assessment data acquisition, and MK led the interviews with app developers. PH conceptualized and designed the study and provided supervision. All authors reviewed and edited the manuscript and provided final approval for its publication.

Conflicts of Interest

All authors are involved in the Label2Enable project, which supports the implementation of the CEN ISO/TS 82304-2 health app quality assessment framework and related label across Europe. In addition, at the time of this study, all authors except PAS and MK were employed by organizations that provide digital health technology assessments.

References

1. Socio-economic impact of mHealth: an assessment report for the European Union. GSMA. URL: https://www.gsma.com/iot/wp-content/uploads/2013/06/Socio-economic_impact-of-mHealth_EU_14062013V2.pdf [accessed 2023-10-20]
2. Butcher CJ, Hussain W. Digital healthcare: the future. Future Healthc J 2022 Jul 28;9(2):113-117 [FREE Full text] [doi: 10.7861/fhj.2022-0046] [Medline: 35928188]
3. Alfawzan N, Christen M, Spitale G, Biller-Andorno N. Privacy, data sharing, and data security policies of women's mHealth apps: scoping review and content analysis. JMIR Mhealth Uhealth 2022 May 06;10(5):e33735 [FREE Full text] [doi: 10.2196/33735] [Medline: 35522465]
4. Melcher J, Torous J. Smartphone apps for college mental health: a concern for privacy and quality of current offerings. Psychiatr Serv 2020 Nov 01;71(11):1114-1119. [doi: 10.1176/appi.ps.202000098] [Medline: 32664822]
5. Sander LB, Schorndanner J, Terhorst Y, Spanhel K, Pryss R, Baumeister H, et al. 'Help for trauma from the app stores?' A systematic review and standardised rating of apps for Post-Traumatic Stress Disorder (PTSD). Eur J Psychotraumatol 2020 Jan 09;11(1):1701788 [FREE Full text] [doi: 10.1080/20008198.2019.1701788] [Medline: 32002136]
6. Tangari G, Ikram M, Ijaz K, Kaafar MA, Berkovsky S. Mobile health and privacy: cross sectional study. BMJ 2021 Jun 16;373:n1248 [FREE Full text] [doi: 10.1136/bmj.n1248] [Medline: 34135009]
7. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020 Feb 01;27(2):330-340 [FREE Full text] [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
8. Faessen JP, Lucassen DA, Buso ME, Camps G, Feskens EJ, Brouwer-Brolsma EM. Eating for 2: a systematic review of dutch app stores for apps promoting a healthy diet during pregnancy. Curr Dev Nutr 2022 Jun;6(6):nzac087 [FREE Full text] [doi: 10.1093/cdn/nzac087] [Medline: 35711572]
9. van Galen LS, Xu X, Koh MJ, Thng S, Car J. Eczema apps conformance with clinical guidelines: a systematic assessment of functions, tools and content. Br J Dermatol 2020 Feb;182(2):444-453 [FREE Full text] [doi: 10.1111/bjd.18152] [Medline: 31179535]

XSL•FO
RenderX

10. Simon L, Reimann J, Steubl LS, Stach M, Spiegelhalder K, Sander LB, et al. Help for insomnia from the app store? A standardized rating of mobile health applications claiming to target insomnia. J Sleep Res 2023 Feb 27;32(1):e13642. [doi: 10.1111/jsr.13642] [Medline: 35624078]
11. Consumer health apps and digital health tools proliferate, improving quality and health outcomes for patients. IQVIA. URL: https://tinyurl.com/4haescn6 [accessed 2023-10-20]
12. Hyzy M, Bond R, Mulvenna M, Bai L, Frey A, Carracedo JM, et al. Don't judge a book or health app by its cover: user ratings and downloads are not linked to quality. PLoS One 2024 Mar 4;19(3):e0298977 [FREE Full text] [doi: 10.1371/journal.pone.0298977] [Medline: 38437233]
13. Jacob C, Sanchez-Vazquez A, Ivory C. Understanding clinicians' adoption of mobile health tools: a qualitative review of the most used frameworks. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e18072 [FREE Full text] [doi: 10.2196/18072] [Medline: 32442132]
14. Jacob C, Sezgin E, Sanchez-Vazquez A, Ivory C. Sociotechnical factors affecting patients' adoption of mobile health tools: systematic literature review and narrative synthesis. JMIR Mhealth Uhealth 2022 May 05;10(5):e36284 [FREE Full text] [doi: 10.2196/36284] [Medline: 35318189]
15. Digital Health in the WHO European region: the ongoing journey to commitment and transformation. World Health Organization. URL: https://tinyurl.com/yhs47c65 [accessed 2024-04-29]
16. van Drongelen A, de Bruijn A, Roszek B, Vonk R. Apps under the medical devices legislation. Rijksinstituut voor Volksgezondheid en Milieu. URL: https://tinyurl.com/2f4ytyre [accessed 2023-12-16]
17. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. The European Parliament and The Council of the European Union. 2017. URL: https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng [accessed 2024-04-29]
18. Fan M, Yu L, Chen S, Zhou H, Luo X, Li S. An empirical evaluation of GDPR compliance violations in android mHealth apps. In: Proceedings of the 31st International Symposium on Software Reliability Engineering. 2020 Presented at: ISSRE '20; October 12-15, 2020; Coimbra, Portugal p. 253-264 URL: https://ieeexplore.ieee.org/document/9251060 [doi: 10.1109/issre5003.2020.00032]
19. D2.1 knowledge tool 1: health apps assessment frameworks. European mHealth Hub. URL: https://mhealth-hub.org/download/d2-1-knowledge-tool-1-health-apps-assessment-frameworks [accessed 2023-10-20]
20. Frey AL, Phillips B, Baines R, McCabe A, Elmes E, Yeardsley-Pierce E, et al. Domain coverage and criteria overlap across digital health technology quality assessments: a systematic review. Health Technol 2024 Dec 20;15(1):15-27. [doi: 10.1007/S12553-024-00934-8]
21. Hoogendoorn P, Versluis A, van Kampen S, McCay C, Leahy M, Bijlsma M, et al. What makes a quality health app-developing a global research-based health app quality assessment framework for CEN-ISO/TS 82304-2: Delphi study. JMIR Form Res 2023 Jan 23;7:e43905 [FREE Full text] [doi: 10.2196/43905] [Medline: 36538379]
22. ISO/TS 82304-2:2021: health software: part 2: health and wellness apps — quality and reliability. International Organization for Standardization. URL: https://www.iso.org/standard/78182.html [accessed 2023-07-14]
23. Freitag B, Fehring L, Uncovska M, Olsacher A, Meister S. Negotiating pricing and payment terms for insurance covered mHealth apps: a qualitative content analysis and taxonomy development based on a German experience. Health Econ Rev 2024 Oct 04;14(1):81. [doi: 10.1186/s13561-024-00558-8] [Medline: 39365482]
24. Goeldner M, Gehder S. Digital health applications (DiGAs) on a fast track: insights from a data-driven analysis of prescribable digital therapeutics in Germany from 2020 to mid-2024. J Med Internet Res 2024 Aug 29;26:e59013 [FREE Full text] [doi: 10.2196/59013] [Medline: 39208415]
25. Frey AL, Leigh S, Toro C, Bufill CP, McCay C, Trupec TP, et al. Preferences and willingness to pay for health app assessments among healthcare stakeholders: a discrete choice experiment. JMIRP Preprint posted online February 17, 2024 [FREE Full text] [doi: 10.2196/preprints.57474]
26. Ataiy O, Burniki M, Sohrabi K, Groß V, Scholtes M, Schudt F. Digital health applications in Germany: explorative pilot study on the challenges of approval from the perspective of manufacturers. Digit Health 2024 Dec 19;10:20552076241302676 [FREE Full text] [doi: 10.1177/20552076241302676] [Medline: 39711743]
27. Schmidt L, Pawlitzki M, Renard BY, Meuth SG, Masanneck L. The three-year evolution of Germany's digital therapeutics reimbursement program and its path forward. NPJ Digit Med 2024 May 24;7(1):139 [FREE Full text] [doi: 10.1038/s41746-024-01137-1] [Medline: 38789620]
28. Mäder M, Timpel P, Schönfelder T, Militzer-Horstmann C, Scheibe S, Heinrich R, et al. Evidence requirements of permanently listed digital health applications (DiGA) and their implementation in the German DiGA directory: an analysis. BMC Health Serv Res 2023 Apr 17;23(1):369 [FREE Full text] [doi: 10.1186/s12913-023-09287-w] [Medline: 37069592]
29. Essén A, Stern AD, Haase CB, Car J, Greaves F, Paparova D, et al. Health app policy: international comparison of nine countries' approaches. NPJ Digit Med 2022 Mar 18;5(1):527 [FREE Full text] [doi: 10.1038/S41746-022-00573-1]
30. About CEN-ISO/TS 82304-2. Label2Enable. URL: https://label2enable.eu/about-cen-iso-ts-82304-2 [accessed 2025-01-04]
31. Haley DT, Thomas P, Petre M, De Roeck A. Using a new inter-rater reliability statistic. Faculty of Mathematics, Computing and Technology, The Open University. URL: https://oro.open.ac.uk/90234/1/TR2008_15.pdf [accessed 2024-05-30]

XSL•FO
RenderX

32. Gwet K. Kappa statistic is not satisfactory for assessing the extent of agreement between raters. Statistical Methods For Inter-Rater Reliability Assessment. 2002. URL: https://agreestat.com/papers/kappa_statistic_is_not_satisfactory.pdf [accessed 2024-04-29]
33. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC Med Res Methodol 2013 Apr 29;13(1):61 [FREE Full text] [doi: 10.1186/1471-2288-13-61] [Medline: 23627889]
34. Dona SW, Nguyen D, Angeles MR, Cooper P, Winter N, Chatterton ML, et al. Developing and testing a health app evaluation framework for organisations to recommend the best health apps for consumers in the Australian setting. Deakin University. URL: https://iht.deakin.edu.au/wp-content/uploads/sites/153/2022/07/D-HEAL-Final-Report.pdf [accessed 2023-07-14]
35. Powell AC, Torous J, Chan S, Raynor GS, Shwarts E, Shanahan M, et al. Interrater reliability of mHealth app rating measures: analysis of top depression and smoking cessation apps. JMIR Mhealth Uhealth 2016 Feb 10;4(1):e15 [FREE Full text] [doi: 10.2196/mhealth.5176] [Medline: 26863986]
36. Miró J, Llorens-Vernet P. Assessing the quality of mobile health-related apps: interrater reliability study of two guides. JMIR Mhealth Uhealth 2021 Apr 19;9(4):e26471 [FREE Full text] [doi: 10.2196/26471] [Medline: 33871376]
37. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions. J Med Internet Res 2017 Mar 21;19(3):e82 [FREE Full text] [doi: 10.2196/jmir.7270] [Medline: 28325712]
38. Lagan S, Aquino P, Emerson MR, Fortuna K, Walker R, Torous J. Actionable health app evaluation: translating expert frameworks into objective metrics. NPJ Digit Med 2020;3:100 [FREE Full text] [doi: 10.1038/s41746-020-00312-4] [Medline: 32821855]
39. Terhorst Y, Philippi P, Sander LB, Schultchen D, Paganini S, Bardus M, et al. Validation of the Mobile Application Rating Scale (MARS). PLoS One 2020 Nov 2;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
40. Baines R, Hoogendoorn P, Stevens S, Chatterjee A, Ashall-Payne L, Andrews T, et al. Navigating medical device certification: a qualitative exploration of barriers and enablers amongst innovators, notified bodies and other stakeholders. Ther Innov Regul Sci 2023 Mar 04;57(2):238-250 [FREE Full text] [doi: 10.1007/s43441-022-00463-4] [Medline: 36194368]
41. Mackey R, Gleason A, Ciulla R. A novel method for evaluating mobile apps (app rating inventory): development study. JMIR Mhealth Uhealth 2022 Apr 15;10(4):e32643 [FREE Full text] [doi: 10.2196/32643] [Medline: 35436227]
42. Jääskelä J, Haverinen J, Kaksonen R, Reponen J, Halunen K, Tokola T, et al. Digi-HTA, assessment framework for digital healthcare services: information security and data protection in health technology – initial experiences. FinJeHeW 2022 Apr 14;14(1):19-30. [doi: 10.23996/fjhw.111776]
43. Scherer J, Youssef Y, Dittrich F, Albrecht U, Tsitsilonis S, Jung J, et al. Proposal of a new rating concept for digital health applications in orthopedics and traumatology. Int J Environ Res Public Health 2022 Nov 13;19(22):14952 [FREE Full text] [doi: 10.3390/ijerph192214952] [Medline: 36429670]
44. Bradford A. The Brussels Effect: How the European Union Rules the World. Oxford, UK: Oxford Academic Press; 2012.
45. New energy efficiency labels explained. European Commission. 2019. URL: https://ec.europa.eu/commission/presscorner/detail/en/memo_19_1596 [accessed 2024-04-29]
46. Biliunaite I, van Gestel L, Hoogendoorn P, Adriaanse M. Value of a quality label and European healthcare professionals' willingness to recommend health apps: an experimental vignette study. J Health Psychol 2024 Aug 02:13591053241258205 [FREE Full text] [doi: 10.1177/13591053241258205] [Medline: 39096027]
47. Caiani EG, Kemps H, Hoogendoorn P, Asteggiano R, Böhm A, Borregaard B, et al. Standardised assessment of evidence supporting the adoption of mobile health solutions: a clinical consensus statement of the ESC regulatory affairs committee. Eur Heart J Digit Health 2024 Sep;5(5):509-523. [doi: 10.1093/ehjdh/ztae042] [Medline: 39318699]

Abbreviations

EU: European Union
MDR: Medical Device Regulation

XSL•FO
RenderX

Edited by A Mavragani; submitted 20.07.24; peer-reviewed by E Baker, J Haverinen; comments to author 19.12.24; revised version received 09.01.25; accepted 12.01.25; published 10.03.25.

Please cite as:
Frey AL, Matei D, Phillips B, McCabe A, Fuller R, Laibarra B, Alonso L, de la Hoz V, Pratdepadua Bufill C, Llebot Casajuana B, D'Avenio G, Sottile PA, Rocchi LM, Errera M, Laaissaoui Y, Cardinal M, Kok M, Hoogendoorn P
Testing and Iterative Improvement of the CEN ISO/TS 82304-2 Health App Quality Assessment: Pilot Interrater Reliability Study
JMIR Form Res 2025;9:e64565
URL: https://formative.jmir.org/2025/1/e64565
doi:10.2196/64565
PMID:

©Anna-Lena Frey, Diana Matei, Ben Phillips, Adam McCabe, Rachel Fuller, Begoña Laibarra, Laura Alonso, Victor de la Hoz, Carme Pratdepadua Bufill, Berta Llebot Casajuana, Giuseppe D'Avenio, Pier Angelo Sottile, Laura Melania Rocchi, Matteo Errera, Yasmine Laaissaoui, Michael Cardinal, Menno Kok, Petra Hoogendoorn. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Digital Translation Platform (Translatly) to Overcome Communication Barriers in Clinical Care: Pilot Study

Victor Olsavszky[1], PD, Dr med; Mutaz Bazari[2]; Taieb Ben Dai[2]; Ana Olsavszky[3]; Fabian Finkelmeier[4], Prof Dr Med; Mireen Friedrich-Rust[5], Prof Dr Med; Stefan Zeuzem[5], Prof Dr Med; Eva Herrmann[6], Prof Dr Med; Jan Leipe[7], Prof Dr Med; Florian Alexander Michael[5], Dr med; Hans von Westernhagen[2]; Olivier Ballo[4], PD, Dr med

[1]Department of Dermatology, Venereology and Allergy, University Medical Center and Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[2]Translatly UG, Offenbach, Germany

[3]Clinic for General and Visceral Surgery, Vivantes Klinikum Kaulsdorf, Berlin, Germany

[4]Department of Internal Medicine 2, Hematology/Oncology, Goethe University Hospital, Frankfurt am Main, Germany

[5]Department of Internal Medicine 1, Gastroenterology, Hepatology and Endocrinology, Goethe University Hospital, Frankfurt am Main, Germany

[6]Institute of Biostatistics and Mathematical Modeling, Goethe University Hospital, Frankfurt am Main, Germany

[7]Department of Medicine V, Division of Rheumatology, University Medical Center and Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

Corresponding Author:
Victor Olsavszky, PD, Dr med
Department of Dermatology, Venereology and Allergy
University Medical Center and Medical Faculty Mannheim
Heidelberg University
Theodor-Kutzer-Ufer 1-3
Mannheim, 68167
Germany
Phone: 49 621 383 2280
Fax: 49 621 383 3815
Email: victor.olsavszky@medma.uni-heidelberg.de

Abstract

Background: Language barriers in health care can lead to misdiagnosis, inappropriate treatment, and increased medical errors. Efforts to mitigate these include using interpreters and translation tools, but these measures often fall short, particularly when cultural nuances are overlooked. Consequently, medical professionals may have to rely on their staff or patients' relatives for interpretation, compromising the quality of care.

Objective: This formative pilot study aims to assess the feasibility of Translatly, a digital translation platform, in clinical practice. Specifically, the study focuses on evaluating (1) how health care professionals overcome language barriers and their acceptance of an on-demand video telephony platform, (2) the feasibility of the platform during medical consultations, and (3) identifying potential challenges for future development.

Methods: The study included ethnographic interviews with health care professionals and an observational pilot to assess the use of the Translatly platform in clinical practice. Translatly was developed to make real-time translation easy and accessible on both Android and iOS devices. The system's backend architecture uses Java-based services hosted on DigitalOcean. The app securely exchanges data between mobile devices and servers, with user information and call records stored in a MySQL database. An admin panel helps manage the system, and Firebase integration enables fast push notifications to ensure that health care professionals can connect with translators whenever they need to. The platform was piloted in a German university hospital with 170 volunteer nonprofessional translators, mainly medical students, supporting translation in over 20 languages, including Farsi, Dari, and Arabic.

Results: Ethnographic research conducted by interviewing health care professionals in Frankfurt am Main and other German cities revealed that current practices for overcoming language barriers often rely on family members or digital tools such as Google Translate, raising concerns about accuracy and emotional distress. Respondents preferred an on-demand translation service staffed by medically experienced translators, such as medical students, who understand medical terminology and can empathize

XSL•FO
RenderX

with patients. The observational pilot study recorded 39 requests for translation services, 16 (41%) of which were successfully completed. The translations covered 6 different languages and were carried out by a team of 10 translators. Most requests came from departments such as infectious diseases (5/16, 31%) and emergency (4/16, 25%). Challenges were identified around translator availability, with 23 (59%) total requests (N=39) going unanswered, which was further evidenced by user feedback.

Conclusions: This pilot study demonstrates the feasibility of the Translatly platform in real-world health care settings. It shows the potential to improve communication and patient outcomes by addressing language barriers. Despite its potential, challenges such as translator availability highlight the need for further development.

(JMIR Form Res 2025;9:e63095) doi:10.2196/63095

KEYWORDS

language barriers; health care communication; medical app; real-time translation; medical translation

Introduction

Language barriers in health care can have a profound impact on patient management. Patients who are not fluent in the primary language of the health care facility are at higher risk of misdiagnosis due to miscommunication during history taking or description of symptoms [1]. This can lead to inappropriate treatments and interventions, worsening the patient's condition [2]. In addition, there may be an increased risk of medical errors or even an increased risk of a serious medical event [3]. Patients may feel ignored, excluded, or misunderstood, leading to reduced satisfaction with care. Language barriers can also hinder the development of trust between the patient and the health care provider [4]. Clinic staff, by contrast, complain of frustration due to suboptimal treatment and the additional time spent with patients as a result of miscommunication [5]. To address these challenges, many health care facilities employ interpreters or multilingual staff and use written materials, simplified language, or even online translation tools [6-8]. Difficulties in communication and misunderstandings can result in additional costs [9], ultimately borne by hospitals, statutory health insurance funds, and the health care system as a whole.

In Germany, there are more than 6 million people with limited knowledge of German. More importantly, there are no or only inadequate translation services available in medical facilities. A study commissioned by the Federal Anti-Discrimination Agency summarizes the role of communication barriers as a risk of discrimination in health care, based on data from 2010 to 2020 in Germany [10]. One of the main issues raised is the lack of efficient access to professional translators for German health services. However, given that Germany is currently home to nearly 22 million people with a migrant background, and that an estimated 25% of German citizens speak 2 foreign languages in addition to German [11], this hidden potential of language skills can be reflected in health care professionals such as doctors, medical students, nurses, and medical technicians. At present, this diversity of foreign language skills is either not used at all or only used in exceptional situations that occur by chance to overcome communication obstacles in the medical sector. In addition, the shortage of health care professionals [12] will lead to a further increase in the number of people with foreign language skills in German society due to immigration in the coming years [13].

The problem of language barriers in health care is not only confined to Germany. As the world's population becomes increasingly diverse and migratory, the need to ensure effective patient-provider communication in a multilingual context is growing. Other countries with a long history of immigration have faced similar challenges in their health care systems. In Canada, for instance, the Winnipeg Regional Health Authority, a local health care organization, employs trained health interpreting staff for in-person and telephone interpreting [14]. While such language services are useful proactive policies, they are not available in all Canadian communities [15] and cannot meet the most common ad hoc interpreting needs [16]. In the United States, state laws require health care facilities to provide interpretation services for patients with limited English proficiency, but these requirements cannot be adequately met [17]. Finally, in Australia, almost half of all clinical practices make use of the translation and interpreting service, which is specifically funded by the government to address language barriers in health care [18]. While these examples provide valuable insights and potential solutions that Germany could consider in its efforts to address its own unique challenges related to language barriers, none of these possible approaches take into account the nuances of dialect or the importance of cultural context in communication. This is where technology-enabled solutions, such as Translatly (Translatly UG), play a critical role in addressing these barriers more effectively. Translatly harnesses the existing, underutilized linguistic potential within the health care workforce, offering an innovative and scalable solution to overcome communication obstacles. With more than €400 (US $436) billion spent on health care in Germany each year [19], digitally targeting these resources to overcome translation barriers could have a positive impact not only on health care delivery, but also on health care expenditure.

With this formative pilot study, we aim to assess the feasibility of Translatly, a digital eHealth platform designed to address language barriers in everyday clinical practice in Germany. Our approach focuses on implementing this digital on-demand translation platform for clinics. The main research questions of the study are as follows: (1) What are the current practices of health care professionals in overcoming language barriers and do they accept an on-demand medical translation platform? (2) Is the Translatly platform feasible for use in clinical settings to overcome language barriers? (3) What other challenges may arise in the use of Translatly that can be addressed in future studies? Translatly is a mobile and tablet app developed in Germany by a founding team of doctors, technicians, and

XSL•FO
RenderX

entrepreneurs to facilitate live translation in everyday medical practice by virtually connecting qualified language mediators with doctors and their patients. This platform allows in-house health care professionals with foreign language skills to register as translators who can be requested and engaged via video telephony when communication barriers arise during doctor-patient interactions. By providing an accessible, scalable, and efficient solution for translation, Translatly has the potential to not only improve the care and management of nonnative patients but also reduce health care costs, making it a sustainable solution for the future.

Methods

Study Design

This formative pilot study used a mixed methods approach combining qualitative ethnographic interviews and observational data collection methods to assess the feasibility of the digital translation platform Translatly in a clinical setting. To explore the extent of language barriers in health care and how they are managed in daily practice, an extensive opinion survey was conducted using an ethnographic questionnaire at the University Hospital of the Goethe University in Frankfurt am Main, as well as in private practices in the Frankfurt area and other selected cities.

The findings from these interviews were used in the design and development of the platform. An observational pilot study was then conducted to test the feasibility of the platform in real medical settings. The use of the Translatly platform was monitored and recorded without manipulating any variables, focusing on how it was integrated into the routine clinical management of patients by health care providers. The pilot study took place over 2 months, from December 1, 2022, to January 31, 2023, at the University Hospital of the Goethe University in Frankfurt am Main.

Study Participants

Participants were recruited using convenience sampling. Health care professionals, including doctors, physiotherapists, and psychotherapists, participated in interviews using an ethnographic research questionnaire. They were identified and contacted through 2 primary methods: first, practitioners at the University Hospital of the Goethe University in Frankfurt am Main were approached directly; second, health care professionals from private practices in the Frankfurt and Offenbach areas were contacted by email, followed by telephone calls to those who expressed interest. Contact details for these private practices were obtained from public directories maintained by the University Hospital of Goethe University. Additionally, practitioners in Düsseldorf and the United States were contacted through recommendations from respondents.

Locating volunteer translators within the health care workforce and matching them with medical staff presented a technical challenge. To ensure the platform's optimal function, a substantial pool of volunteer translators was needed. Recruitment focused on medical students, as they possess the medical knowledge to handle complex translations [20]. They were enrolled through a clinical elective course titled "Translatly and Digitalization in Healthcare," which also provided them with community service certification. Volunteer translators were briefed on data protection regulations, including General Data Protection Regulation (GDPR) compliance.

By the last quarter of 2022, a total of 170 volunteer translators had been registered, 90% (n=153) of whom were medical students from the University Hospital. The remaining 10% (n=17) comprised medical staff. For the pilot study, translators were recruited for the following languages: Farsi, Dari, Arabic, Italian, Urdu, Portuguese, Spanish, Turkish, Russian, Albanian, Vietnamese, Romanian, Slovenian, Pashto, Macedonian, French, Croatian, Hindi, Bosnian, Azerbaijani, Korean, and Berber.

Patients were recruited through direct contact with doctors at the University Hospital, who identified cases needing translation services. Doctors interested in participating were provided instructions on how to use the Translatly app, and patients were also supported in using the app if needed.

Data Collection

Overview

Data were collected through ethnographic interviews with health care professionals using a questionnaire, as well as observational monitoring of the Translatly platform in clinical practice during the pilot study. The interviews focused on current practices for overcoming language barriers, while the observational data recorded translation requests, languages, and involved medical departments.

The ethnographic questions were designed specifically for this study and followed the format of semistructured interviews (Multimedia Appendix 1). This approach provided a mix of open-ended questions and specific information requests, allowing for both structured inquiry and the flexibility for interviewees to express their own experiences and opinions freely. Specifically, practitioners were asked about their personal experiences, their challenges in dealing with non–German-speaking patients, and their thoughts on how to improve communication. Questions about the use of software and technology, and specific instances where translation services have been required, provide deeper insights into everyday practice and the issues involved. Typical of ethnographic interviews, the questionnaire can be used in a context where interviews are more informal and conversational. The interviews were designed to last approximately 50 minutes, with time allocated for each section of the interview guide.

The pilot study collected data on the frequency of requests for translation services, the response rate to these requests, the duration and languages of completed translations, and the hospital departments making these requests. User feedback was also collected to assess problems and challenges during the pilot phase. This information allows for a comprehensive assessment of the use of the Translatly platform in clinical practice by providing data on the demand for translation services, the availability of the translator network, and the different medical specialties requiring translations. As this feasibility study only aims to measure and observe whether Translatly can be used as a translation platform in everyday clinical practice, a comparison

XSL•FO
RenderX

group (eg, patients without a translation platform) was not required.

Instruments

Data from ethnographic interviews were collected through written notes and supplemented by audio recordings to ensure accuracy and completeness. The interviews were then transcribed into words for analysis.

The tools required for the observational pilot study were the Translatly app and a mobile device or tablet. While some medical departments at the University Hospital of the Goethe University in Frankfurt am Main provided their own tablets, this was not a requirement as the app could be used by doctors, patients, and translators on their own smartphones. Translatly was preinstalled on all devices used in the pilot study with the help of the study team. Network connectivity was provided by the university hospital's wireless network or by the mobile phone's network provider.

Analytical Techniques and Software

Qualitative Content Analysis

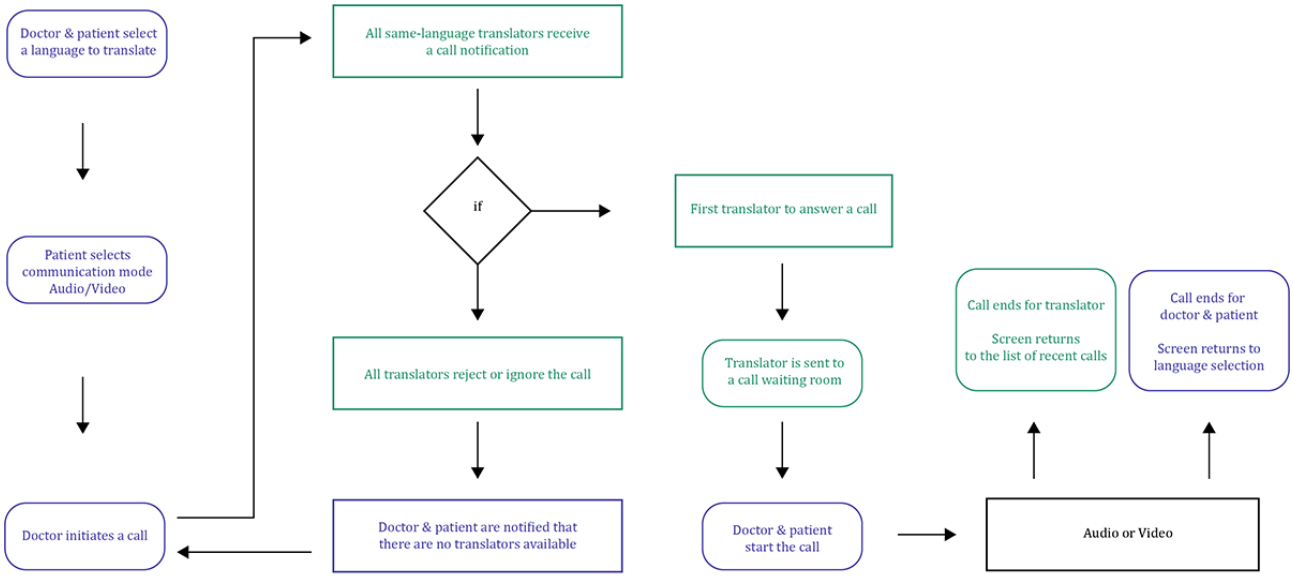
A qualitative content analysis was conducted to systematically evaluate the data from the ethnographic interviews. This analysis provided insights into clinical standards, patient documentation, and the use of translation services.

The analysis of the pilot study data was performed using Excel (version 2408 build 16.0.17928.20114, 64-bit; Microsoft Corp.), with raw data provided by the Translatly digital platform, which is methodologically described in detail below.

Operating Mode of the Translatly Digital Translation Platform

Translatly's digital translation platform is designed to provide medical facilities, such as hospitals or private practices, with an app that can be used on smartphones or tablets within the facility. With the app, doctors or other medical staff treating patients can access a pool of translators who can provide language assistance during patient consultations. The medical practitioner chooses the language to be translated and the Translatly algorithm selects and contacts the appropriate translators from the Translatly pool (Figure 1). The first translator to accept the contact request is connected directly to the treating team. The medical staff, patient, and translator communicate via Translatly's internal audio or video interface.

Figure 1. Flowchart describing the operational process of the Translatly digital translation platform. It starts with a medical practitioner (doctor) identifying the need for a translator during a patient consultation and selecting the desired language for translation using the Translatly app on a smartphone or tablet. Once the patient selects their preferred mode of communication, either audio or video, the Translatly algorithm is activated and searches its database for available translators who have registered as proficient in the required language. The algorithm sends a contact request to these translators. The translators receive the request on their devices and the first one to accept the request is connected to the doctor and the patient. If the translator is accepted, an audio or video call is initiated. If all the translators reject the request or do not respond within a certain time frame, the algorithm notifies the doctor and the patient that the search was unsuccessful. At the end of the call, the translator's screen returns to the list of recent calls and the doctor's and patient's screen returns to language selection. Doctor and patient process operations are shown in purple, while translator operations are shown in green.



Translatly's Digital Platform Development Framework

The initial steps to ensure the successful implementation of the digital platform before the full implementation of the pilot phase consisted of 3 main pillars: (1) evaluation of the feasibility of a digital interaction-translation platform in a German university hospital setting; (2) regulatory compliance research; and (3) app design and development.

Following an extensive review of the existing medical literature, the feasibility assessment (pillar 1) consisted of developing the ethnographic research questionnaire (Multimedia Appendix 1) to determine the impact of language barriers in health care and how they are managed in daily practice.

Regulatory compliance research (pillar 2) involved an in-depth assessment of relevant medical regulations and standards in Germany. In brief, our team ensured that the app complies with

XSL•FO
RenderX

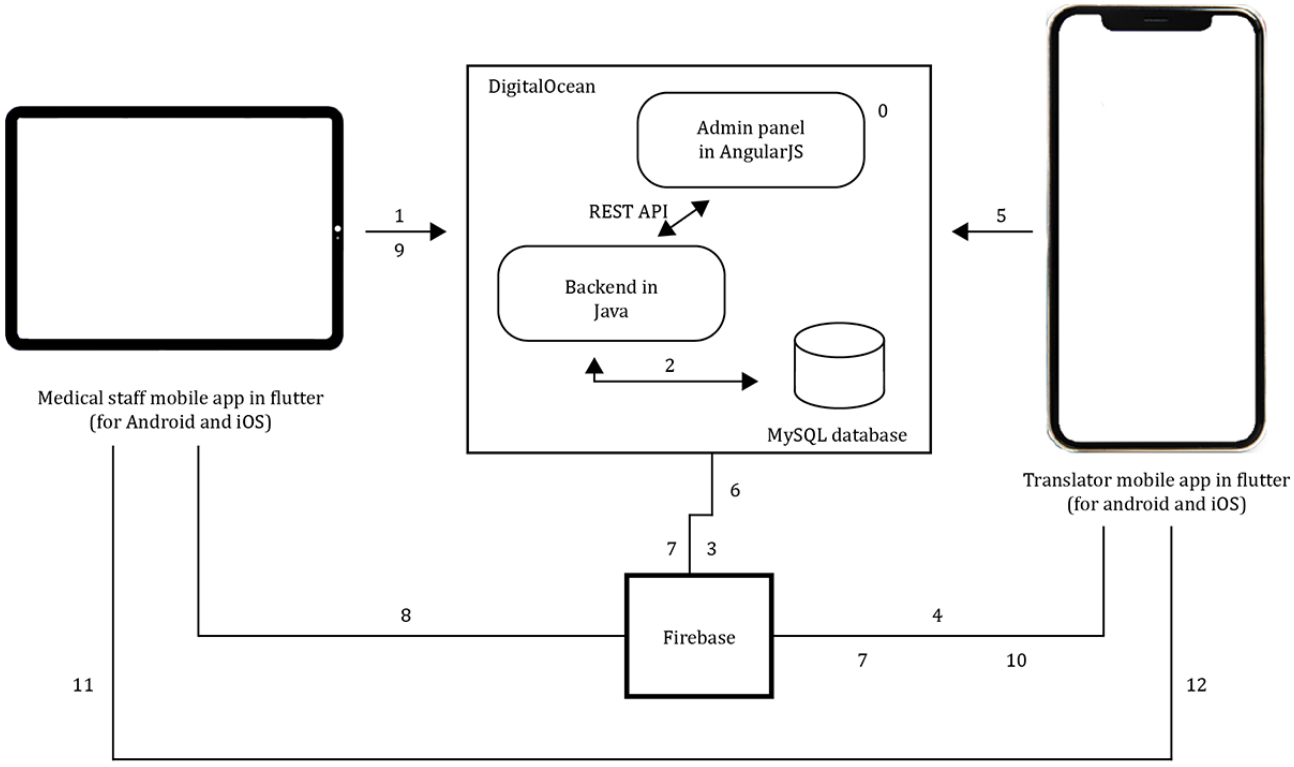
regulatory requirements for data security and privacy. In addition, we ensured that our research project was in line with the purpose of the Innovation Fund, which was established within the German Federal Joint Committee to promote new forms of health care that go beyond the existing standard of care covered by statutory health insurance in Germany [21].

Design thinking workshops, cofacilitated by members of the clinical research team and software engineering specialists, marked the start of the design and development phase (pillar 3). These sessions set the stage for a subsequent design sprint phase. During this phase, the software development team built and evaluated prototypes and integrated and validated the selected ideas for the app. The research team consisted of a multidisciplinary collective, including 2 software developers with computer science expertise, 2 clinical oncologists to provide medical insight, a clinical trial investigator to oversee research integrity, and a business engineer to manage regulatory issues.

Translatly Service System Architecture

To ensure the desired functionality of the Translatly digital interaction platform, a high-level software, database, and server architecture have been defined and implemented by our software development team (Figure 2). To achieve cross-platform compatibility on both Android (Google LLC/Alphabet Inc.) and iOS (Apple Inc.) devices, the architecture consists of 2 main mobile apps developed using the Flutter framework [22]. One app is designed specifically for medical staff, acting as a practitioner interface, while the other acts as a translator interface, supporting real-time communication via audio or video telephony.

Figure 2. Architecture and data flow of the Translatly Digital Interaction Platform. This diagram illustrates the systematic process from the initiation of a translation request to the establishment of a video telephony session between medical staff and translators. Each step (0-12) marks a critical interaction between components of the Translatly platform, ensuring an efficient and fast response to overcome language barriers in real-time health care scenarios. “0. Administration panel”: Initially developed with AngularJS for the front end, Java Spring for the backend, and integrated with a MySQL database, the administration panel is used by administrators to create, edit, and delete accounts for medical staff and translators. “1. Language request”: The operator uses the mobile app to send a Representational State Transfer (REST) request for translation services to the backend hosted on DigitalOcean. ”2. Translator list retrieval”: Upon receiving the request, the backend accesses the MySQL database to retrieve a list of available translators who speak the required language. “3. Translator notification request”: The backend sends a request to Firebase, the cloud service, to initiate a push notification to the appropriate translators. “4. Push notification to translators”: Firebase processes the request and sends a push notification to the translators' mobile devices. “5. Translator availability confirmation”: The first translator to respond sends an availability confirmation back to the backend via the REST application programming interface (API). “6. Notification of translator confirmation”: Upon confirmation, the backend updates the Firebase cloud service with the translator's availability. “7. Cancellation notification to translators”: Firebase will then send a push notification to the remaining translators to cancel the call as a translator has been found. “8. Notification of attending staff”: Firebase also notifies the medical staff that a translator has been secured for the session. “9. Meeting link to medical staff”: The medical staff member receives the meeting link from the backend after joining the call. “10. Meeting link to the translator”: Similarly, the translator receives the meeting link via a push notification from Firebase. “11. Medical staff meeting join”: The medical staff member joins the video chat meeting using the link provided. “12. Translator meeting join”: The translator also joins the video chat meeting so that the translation session can begin.



The platform’s backend is built on Java-based services, hosted on DigitalOcean (DigitalOcean Holdings, Inc.) for a stable and scalable cloud infrastructure [23]. The services can be accessed through a REST API (Representational State Transfer Application Programming Interface) [24] protected by an OAuth (Open Authorization) algorithm [25] that secures the connection

XSL•FO
RenderX

between the mobile apps and the server, ensuring efficient data transfer and real-time interactions. Moreover, a comprehensive administrative panel, originally developed with AngularJS (Angular Java Script) [26] and later migrated to ReactJS (React Java Script) [27], provides full system management and is responsible for maintaining user profiles, language preferences, and service availability of registered translators.

For data persistence, a MySQL (Oracle Corporation) database is used to store structured data, handling user information, interaction records, and call history [28]. To facilitate push notifications, we have integrated Firebase (Google LLC/Alphabet Inc.) into our system architecture [29]. This integration enables in-house health care professionals to be promptly notified and instantly connected to ongoing consultations. Real-time database and user authentication are implemented internally. Finally, Twilio (Twilio Inc.) is used as the video communication tool, known for enabling developers to build custom communication solutions [30].

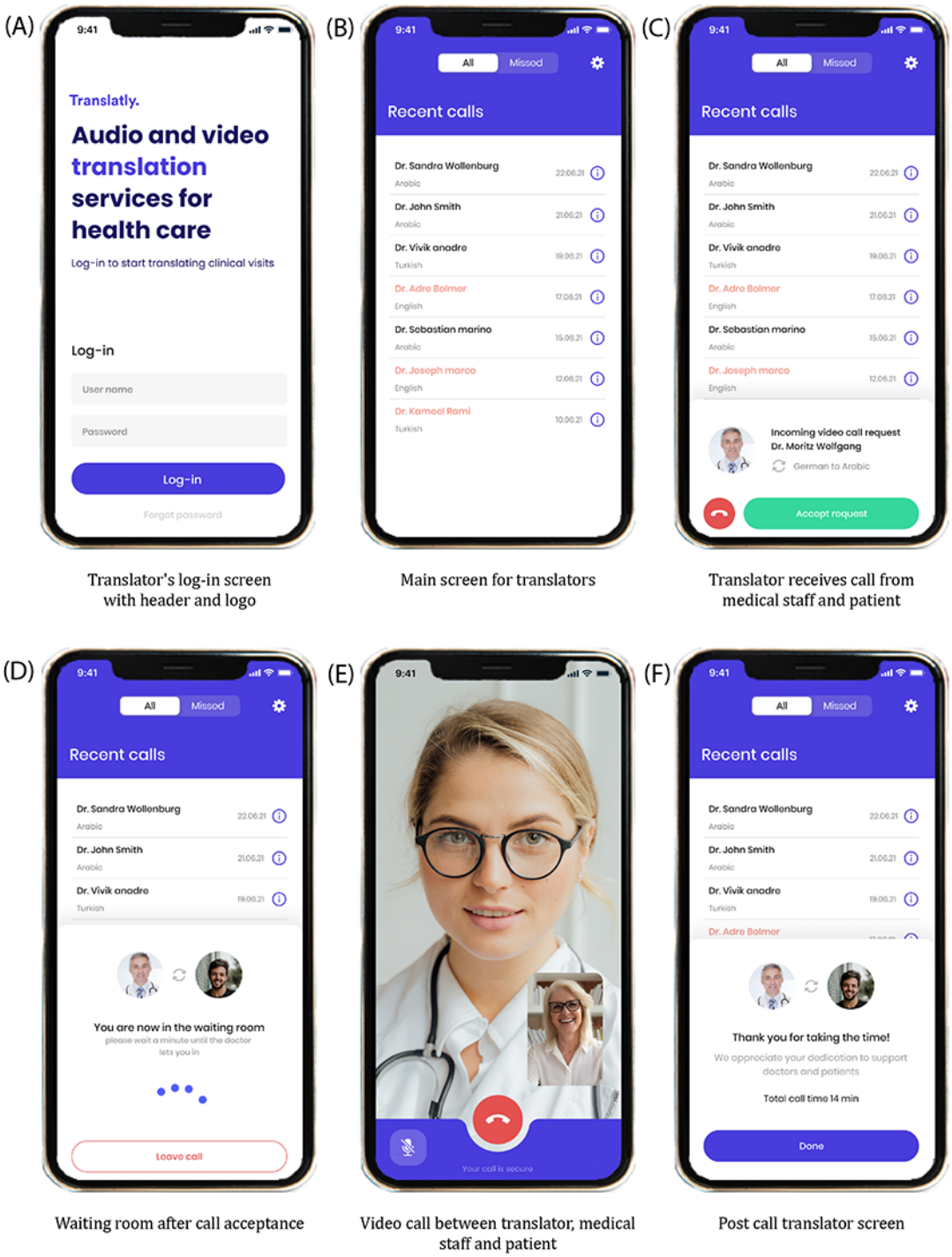
Translatly User Experience Design and Interface Functionality

To ensure the intended mode of operation, we have opted for an intuitive user interface, specifically designed to guarantee and immediately establish an optimal interaction between medical professionals and translators. Upon opening the app, medical professionals must enter their credentials provided by our team to securely access Translatly's services (Multimedia Appendix 2). After logging in, they are presented with the main app screen, which prominently features an on-demand translation tab with an extensive language selection menu. Next, the medical staff user can choose between audio and video call modes, conceivably in the presence of and in agreement with the patient. The interface then displays the status of the translator call process and notifies the medical staff when a translator has accepted the call, providing an easy transition into the call.

On the other side, translators access their dedicated portal through a similar secure log-in interface (Figure 3A). Their main screen displays recent and missed calls, with relevant details such as the caller's name and the languages translated during the respective calls (Figure 3B). An incoming call is displayed as a slide-up in-app notification with relevant information (Figure 3C). Upon acceptance, the translator is temporarily held in a virtual waiting room until the calling user agrees to start the call (Figure 3D). After the call (Figure 3E), a postcall notification provides a summary of the call (Figure 3F).

XSL•FO
RenderX

Figure 3. Selection of Translatly mobile frontend screenshots showing different stages of the user experience for translators. (A) Log-in screen with header and logo, where translators enter their credentials to access the platform. (B) The main screen displays the last calls and missed calls of the translator, including the language that was interpreted during the call and the name of the doctor. (C) Notification when a translator receives a video call request from a health care professional, indicating the languages to be translated. (D) Upon acceptance of the call, the translator is directed to a waiting room and is awaiting entry into the consultation. (E) Active video call interface used by the translator to communicate with medical staff and patient. (F) The postcall screen provides a summary of the call duration and a thank you message with an option to end the session.



Ethical Considerations

This study involving human participants has been reviewed and approved by the Medical Ethics Committee of the Goethe University, Frankfurt am Main, Germany (ethics approval 2024-1776). The study was conducted in accordance with the principles of the Declaration of Helsinki. All participants provided informed consent for their involvement in the study. Additionally, the Medical Ethics Committee has approved the secondary analysis of the pilot phase data without the need for further consent. We confirm that the Translatly team adheres to regulations, including compliance with the GDPR, the Federal Data Protection Act, the Hessian Data Protection Act, and confidentiality in accordance with § 203 of the German Criminal Code. In addition, all data collected are anonymized. No compensation was provided to participants in this study. All images used in this study are open source and publicly available

XSL•FO
RenderX

on the official Translatly website [31], with no identifiable individual participants or users depicted.

Results

Summary of Responses to Ethnographic Research Questionnaires

The ethnographic research was conducted to assess the impact of language barriers in health care, and to gain a holistic view of the challenges health care professionals face in their clinical routine when dealing with nonnative speaking patients. Respondents included specialists and therapists from maximum care hospitals or university hospitals (6/10, 60%), from private practice (3/10, 30%), or from both the private and public sectors (1/10, 10%; Multimedia Appendix 3). The majority of respondents were oncologists (4/10, 40%) from major German cities such as Offenbach am Main (2/10, 20%), Frankfurt am Main (1/10, 10%), Düsseldorf (1/10, 10%), and even from the United States (1/10, 10%). We were able to identify a number of key findings from the health care professionals we interviewed, which were later of great importance for the design and development of the app. Most respondents used both medical staff and family members (5/10, 50%), but no other formal medical translation services such as third parties or software (9/10, 90%), when facing language barriers in their workplace (Multimedia Appendix 3). Common foreign languages encountered in medical settings in Frankfurt am Main and other selected German cities were Russian, Kurdish, Arabic, French, Turkish, Afghan, Italian, Portuguese, Eritrean, Serbian, Croatian, Bulgarian, Spanish, Greek, and Chinese.

Most respondents raised concerns about accuracy and the emotional burden on relatives, with some giving specific examples of relatives hiding the truth from patients to appease them. Some professionals used digital tools such as Google Translate (Google LLC/Alphabet Inc.), but with dissatisfaction due to the lack of confidence in the accuracy of the translation and the lack of empathy associated with using a device, particularly when making difficult diagnoses. The availability of hospital interpreter pools in larger hospitals was noted but underutilized, mostly due to complex technicalities. Notably, some doctors working in private practice stated that, in the case of non–German-speaking patients, they would offer the patient a new appointment and ask them to bring an interpreter with them.

When considering either a scheduled appointment or an on-demand call, health care professionals expressed a preference for an on-demand translation service such as Translatly (8/10, 80%; Multimedia Appendix 3). The ideal translation solution was identified as a pool of medically experienced translators, suggesting that medical staff, including medical students, would be most effective due to their knowledge of medical terminology and empathy. Estimated usage of a translation app varied, with most professionals predicting usage of at least once a week. Finally, concerns about bureaucratic and logistical hurdles, the efficiency of nonmedical translators, and the overall feasibility of integrating such a system into daily medical practice were raised, despite the enthusiasm for a digital translation platform.

Evaluation of Pilot Study

The pilot study recorded 39 translation requests over a 2-month period, of which 16 (41%) were successfully completed (Table 1).

The successfully completed translations accounted for a cumulative duration of 209 minutes across 16 distinct conversations. These translation sessions were carried out by a team of 10 translators, who contributed to facilitating communication across language barriers within the hospital setting. The breakdown of languages translated through the platform included Russian, Ukrainian, Serbian/Croatian/Bosnian, Arabic, Greek, and Urdu, reflecting a wide range of linguistic needs among the patient population.

The origin of translation requests spanned several medical departments, with the Department of Infectious Diseases submitting the highest number of requests (n=5), followed by the Emergency Department with 4 requests. Internal medicine departments submitted 3 requests, while pediatrics and psychiatry departments each submitted 2 requests for translation services.

XSL•FO
RenderX

Table 1. Overview of translation requests during the pilot phase of the implementation of the digital translation platform Translatly at the Goethe University Hospital in Frankfurt am Main (December 2022-January 2023). The table summarizes the total number of on-demand translation requests, the outcome of these requests (whether completed or unanswered), the languages translated, and the medical departments from which the requests originated. The study involved health care professionals from different departments and focused on the feasibility of the Translatly platform in a clinical setting.

Description	Values (n=39), n (%)
Translation requests	
Remained unanswered	23 (59)
Successfully completed	16 (41)
Started, but not completed	0 (0)
Translated languages (n=16)	
Russian	6 (38)
Ukrainian	3 (19)
Serbian/Croatian/Bosnian	2 (13)
Arabic	2 (13)
Greek	2 (13)
Urdu	1 (6)
Medical departments of translation requests (n=16)	
Infectious diseases	5 (31)
Emergency department	4 (25)
Internal medicine departments	3 (19)
Pediatrics	2 (13)
Psychiatry	2 (13)

Challenges of Using the Translatly Platform

Several challenges were identified during the pilot study. A key issue was translator availability, with 23 out of 39 (59%) translation requests going unanswered, mainly due to translators being unavailable at the time of the request. In addition, while feedback from both medical staff and translators was generally positive—particularly regarding the simple and intuitive design of the platform—some users expressed concern about the delayed responses. Notably, some translators indicated that they had been contacted via instant messaging beforehand, to ensure availability when a translation request was made.

Taken together, the results of this study addressed the 3 primary research questions: (1) how health care professionals overcome language barriers and their acceptance of Translatly's on-demand medical translation platform, (2) the feasibility of implementing Translatly in clinical settings to overcome language barriers, and (3) the identification of potential challenges that could inform future studies.

Discussion

Principal Findings

With this formative research study, we present the first successful implementation of the digital translation platform "Translatly" in a medical setting to tackle language barriers. Ethnographic research revealed that health care professionals currently rely on family members or online tools to overcome language barriers. There was a strong preference among health care professionals for an on-demand translation service staffed by medically experienced professionals with foreign language skills. The pilot study, conducted at the Goethe University Hospital in Frankfurt am Main, Germany, demonstrated the feasibility of using Translatly in a real-world clinical setting. Translatly's solution is to virtually enable patient-doctor communication by leveraging the language skills of health care professionals. Facilitating real-time translation services during patient consultations directly addresses the problems of miscommunication in health care by providing immediate access to translators, who in this study are either medical students or trained medical professionals. Our translator pool, which is predominantly made up of migrant health professionals, not only ensures that complex medical terminology is translated properly, but can also address cultural language differences [32], thereby reducing the likelihood of misunderstandings and subsequent medical errors.

To the best of our knowledge, our digital approach to addressing the problems of miscommunication due to language barriers in health care is unique in that it seeks to combine human language skills with the capabilities of modern eHealth technology. Currently, approaches to the problem of language barriers focus heavily on so-called machine translation (MT) software, which is specifically designed to translate text or speech from one language to another without human intervention [33,34]. In fact, medical translation apps have been published or are already available in app stores. However, they either rely on MT or contain predefined health phrases [35]. The modern focus on automated translation alternatives is mainly motivated by the

XSL•FO
RenderX

fact that human translators can be costly and time-consuming [36]. This is evidenced by studies that have shown, for example, that the use of human interpreters has decreased in several health centers in the United States from 2009 to 2019 [37]. Therefore, the Translatly platform represents a novel solution that does not require translators to be tied to a specific hospital, and moreover, the intensive search for a specific translator can be quickly overcome by accessing the translator pool, which can be done by medical staff directly from the patient's side of the bed.

One of the key findings from the responses of the medical staff interviewed using our ethnographic research questionnaires was that the ideal solution proposed by the majority of interviewees would be to actually use medical staff or medical students with foreign language skills, and this was a key influence in prompting our team to look for volunteer translators to join our translator pool. However, we acknowledge that this could be a major limitation of our pilot study as the volunteer translators we recruited are not specialist interpreters but rather medical practitioners or students. Therefore, the issue of variability in the medical knowledge and language skills of the translators recruited for this study is a critical concern that could affect the quality and accuracy of translations in clinical settings. Regarding the foreign language skills of volunteer translators with a migrant background, it has been estimated that in 2023 in Germany, approximately 64.5% of people with a migrant background were first-generation migrants, and therefore native speakers of their own language [38]. Native speakers do not normally need language certificates, as they are assumed to have an innate and comprehensive understanding. A native speaker is defined as an individual who typically acquires the language as a young child and maintains it into adulthood [39]. Notably, there are linguists who distinguish between the basic and extended proficiency of native speakers [40]. With regard to the German language skills of translators from migrant backgrounds, there are data showing that in Germany 66% of migrants consider themselves to have a good or professional level of German [41] and strong language skills are required for enrollment in medical faculties. Although these statistics do not resolve the issue of variability in the medical translation skills of the recruited translators, they do support the fact that translation errors via the Translatly platform may indeed be minimal. Contrary to our findings from the ethnographic research questionnaires, other similar interviews showed that hospital staff were strongly opposed to the use of bilingual medical staff services [42]. It should be noted, however, that these interviews were conducted to assess language services for written materials and not for direct dialogue translation. By contrast, we strongly believe that while the Translatly concept does not provide an overarching solution, the use of bilingual medical staff for the Translatly platform represents a significant improvement over the already precarious situation in daily clinical routine, as evidenced by our questionnaire results and others [43], of relying on family members or untrained interpreters. Furthermore, even when official translators are present, there is still potential for mistranslation [44], making other translation alternatives necessary in some cases to improve communication with patients.

To ensure that the Translatly digital platform is user-centric, the needs of the medical staff that emerged from the questionnaires were carefully considered during the design of the operating mode and later during the development of the app. However, as the call is always initiated by the doctor in the presence of the patient, it may be unannounced and unexpected for the translator. We have therefore taken into account the possibility of some translators being unavailable during ad hoc calls and have specifically designed the operating mode and system architecture so that all translators in the pool will receive the calls if they are listed for the selected language to be interpreted. The integration of the Translatly app with Firebase [45], a widely used service for sending push notifications to a large number of users quickly and reliably, helps to facilitate the calling of all required translators from the pool. Moreover, real-time video telephony in clinical settings can be disrupted by connectivity problems due to network instability. Solutions to this issue could include switching to an audio call, which requires less data and can tolerate latency, or reconnecting with another translator via push notifications provided by the robust backend architecture with Firebase. However, these measures may not fully prevent disruptions, so offline text-based translation or rescheduling during outages would be helpful. Further time-saving is achieved through Translatly's simple, user-friendly app design. Knowing that effective and prompt communication is paramount in the medical setting, we focused on designing our app for the pilot study with simple accessibility, such as few visual elements and features, intuitive navigation, legible fonts, and a few color palettes. As our pilot study aimed to test Translatly's usability in a real-world medical setting, we had not yet included other app features such as scheduling, an integrated calendar, or chat services.

The results of the pilot study demonstrated a significant need for effective translation services in the medical setting, as evidenced by the Translatly app being used for a total of almost 3.5 hours. The data showed that over 40% (16/39) of translation requests were successfully fulfilled, highlighting the potential of the app to improve communication between health care providers and non–German-speaking patients. Another notable limitation observed during the pilot study was that 23 out of 39 (59%) translation requests went unanswered. This highlights the challenges associated with translator availability and responsiveness, particularly at peak times, which could affect the scalability of such a digital solution across larger or more diverse linguistic landscapes. As a consequence, our pilot study led us to add an optional scheduling feature and calendar to the app. To further mitigate availability issues, we are exploring the possibility of allowing translators to indicate their availability in advance. This system would prioritize periods of high demand and ensure that enough translators are on call when they are most needed. Another option would be a real-time monitoring system that tracks translator availability and dynamically adjusts call routing to match supply with demand. This would help to minimize delays and ensure that translation services are readily available when they are needed. Finally, expanding the pool of translators to include the official translators employed by hospitals and incorporating a translator selection algorithm to improve response rates or using MT-based translation options would ensure broader and more professional

XSL•FO
RenderX

coverage. Nonetheless, the diversity of languages requested, from Russian to Urdu, across a range of departments from infectious diseases to psychiatry, underlines the widespread need for effective translation services in medical settings and the successful initial broad implementation of Translatly in a variety of medical departments for patients from diverse linguistic backgrounds.

While Translatly offers many benefits, there are further inherent challenges and limitations that need to be considered. First, to address the issue of variability in the professional translation of medical conversations, more rigorous selection criteria and training protocols for translators are needed. These protocols should include targeted medical terminology training focusing on the most common conditions and procedures in the hospital departments where translation services are critical. In addition, the inclusion of role-playing scenarios with simulated doctor-patient interactions could be implemented to allow students to practice translation in a controlled environment and receive feedback on real-life challenges. Further cultural competency training would also be beneficial, with modules designed to deepen understanding of cultural nuances that may impact on communication, ensuring that translations are both linguistically accurate and culturally sensitive. Next, relying on health care professionals to perform dual roles as medical providers and translators could potentially lead to work overload and burnout. Effective workload management is essential to prevent fatigue and ensure the sustainability of translation services [46]. Therefore, we will consider assessing workload reduction in future studies on the implementation of Translatly.

Collecting robust data on user usage and satisfaction, as well as the impact on treatment outcomes will provide insights that could guide future improvements and adaptations to the platform. In addition, expanding the range of languages offered and scaling the solution to other regions and health care settings could increase its impact, making inclusive health care more accessible nationally and potentially globally. Next, Translatly's mode of operation does not address the issue of face-to-face interpreting by existing hospital interpreters. While official hospital interpreters can easily be added to Translatly's pool of translators, we continue to advocate for video interpreting, as recently a systemic review has shown that patient satisfaction with video interpreting is the same as with in-person communication [47].

Conclusions

This formative pilot study assessed the feasibility of overcoming language barriers in clinical settings through the use of the Translatly digital translation platform. The study aimed to determine how health care professionals overcome language barriers, whether they prefer and accept this on-demand translation solution, and whether it can be effectively integrated into routine clinical workflows. The platform leverages the language skills of medical students and health care professionals, enabling them to serve as on-demand volunteer translators during medical consultations via video telephony. The results showed that Translatly has significant potential to improve communication between health care providers and nonnative speaking patients, thereby improving patient outcomes and optimizing health care resources. Although this early-stage study focused primarily on feasibility, it also identified key challenges, such as translator availability, that can inform future refinements to the platform. The study did not evaluate the quality of translations, nor did it focus on the platform's architecture or mechanisms for attracting users. Rather, it laid the groundwork for future work by exploring the operational feasibility of Translatly in a real-world health care setting.

Translatly's alignment with statutory health insurance mandates in Germany positions it as a potential eHealth solution for widespread adoption [48]. By reducing language barriers, it can support better health outcomes and promote health education, particularly in diverse patient populations. A subscription-based model could address the platform's cost implications, allowing health care facilities to scale the service according to their needs. Initial costs related to technology infrastructure, training, and ongoing maintenance would be manageable, particularly in larger hospitals with robust IT systems, and would be justified by the long-term benefits of the platform, including improved workflow, improved patient outcomes, and reduced overall health care costs. Projected revenue growth of €16,852 (US $18,334) in the first year of implementation, rising to €5,933,455 (US $6,455,302) in the fourth year, supported by strategic expansion and scaling efforts, further underscores the platform's potential for sustainable financial viability within the health care sector. With further development and strategic implementation, platforms such as Translatly could become an integral part of health care systems, improving communication, patient satisfaction, and reducing health care costs.

Ultimately, Translatly has the potential to set a new standard for overcoming language barriers in health care. By enabling health care professionals to communicate effectively with nonnative speakers, the platform strengthens trust between patients and providers, reduces the emotional burden on families, and supports the delivery of high-quality, culturally competent care.

Acknowledgments

JL and VO received a grant from the Ministry for Sciences, Research and Arts, Baden-Wuerttemberg, supporting the Clinical Cooperation Unit Digital Spaces in Inflammatory Diseases at the Centre for Preventive Medicine and Digital Health, Medical Faculty of Mannheim, Heidelberg University. We thank all the registered interpreters for their valuable assistance.

Data Availability

The anonymized data supporting the findings of this study are included as Multimedia Appendix 4.

XSL•FO
RenderX

Conflicts of Interest

MB was the project manager for app design and development and TBD was responsible for software architecture and development. Both MB and TBD are working for CodeCooperation GmbH and were financially enlisted by Translatly UG. HvW and OB are owners of Translatly UG, registered at the address Frankfurter Strasse 134, 63067 Offenbach, Germany. SZ has provided consultancy and participated in speakers' bureaus for Abbvie. All other authors have no conflicts of interest.

Multimedia Appendix 1
Ethnographic research questionnaire used for the feasibility study, translated into English.
[PDF File (Adobe PDF File), 131 KB - formative_v9i1e63095_app1.pdf]

Multimedia Appendix 2
Examples of screenshots of the Translatly app user interface for medical professionals on a tablet device. (A) The log-in screen with header and logo where medical staff enter their credentials. (B) The main app screen for medical staff, showing an on-demand translation tab with a drop-down menu for language selection. (C) After selecting the language to be translated, the medical staff user can choose between audio or video call mode to initiate a translation session. (D) Screen showing the process of calling available translators once a translation request has been made. (E) Notification screen informing medical staff when a translator has accepted the call, with options to continue or start again. (F) An illustrative example of videoconferencing with a translator during medical consultation.
[PNG File , 1556 KB - formative_v9i1e63095_app2.png]

Multimedia Appendix 3
Clinical profile and characteristics of health care professionals interviewed for the feasibility assessment of the Translatly digital translation platform, conducted through an ethnographic research questionnaire. This appendix provides a comprehensive statistical summary of respondents' medical specialties, clinic types, and work locations. It also details their current practices for managing language barriers in clinical settings, including the types of translation services they use and their preferred mode of access to online translation options (on-demand or appointment-based).
[PDF File (Adobe PDF File), 97 KB - formative_v9i1e63095_app3.pdf]

Multimedia Appendix 4
Anonymized data set of translation requests from the Translatly pilot phase at the University Hospital of the Goethe University in Frankfurt am Main, Germany, covering the period from December 1, 2022, to January 31, 2023. The data include the total number of translation requests (n=39), the number of completed translations (n=16), and the number of unanswered requests (n=23). The appendix also provides detailed information on the duration of the translations, the languages translated, and the medical departments from which the requests originated. This data set is fully anonymized.
[PDF File (Adobe PDF File), 74 KB - formative_v9i1e63095_app4.pdf]

References

1. Slade S, Sergent S. Language Barrier. Treasure Island, FL: StatPearls; Apr 10, 2023:1.
2. Divi C, Koss RG, Schmaltz SP, Loeb JM. Language proficiency and adverse events in US hospitals: a pilot study. Int J Qual Health Care 2007 Apr;19(2):60-67. [doi: 10.1093/intqhc/mzl069] [Medline: 17277013]
3. Cohen AL, Rivara F, Marcuse EK, McPhillips H, Davis R. Are language barriers associated with serious medical events in hospitalized pediatric patients? Pediatrics 2005 Sep;116(3):575-579. [doi: 10.1542/peds.2005-0521] [Medline: 16140695]
4. Mustafa R, Mahboob U, Khan RA, Anjum A. Impact of language barriers in doctor - patient relationship: a qualitative study. Pak J Med Sci 2023;39(1):41-45 [FREE Full text] [doi: 10.12669/pjms.39.1.5805] [Medline: 36694787]
5. Wune G, Ayalew Y, Hailu A, Gebretensaye T. Nurses to patients communication and barriers perceived by nurses at Tikur Anbessa Specilized Hospital, Addis Ababa, Ethiopia 2018. International Journal of Africa Nursing Sciences 2020;12:100197 [FREE Full text] [doi: 10.1016/j.ijans.2020.100197]
6. Al Shamsi H, Almutairi AG, Al Mashrafi S, Al Kalbani T. Implications of language barriers for healthcare: a systematic review. Oman Med J 2020 Mar;35(2):e122 [FREE Full text] [doi: 10.5001/omj.2020.40] [Medline: 32411417]
7. Diamond LC, Schenker Y, Curry L, Bradley EH, Fernandez A. Getting by: underuse of interpreters by resident physicians. J Gen Intern Med 2009 Feb;24(2):256-262 [FREE Full text] [doi: 10.1007/s11606-008-0875-7] [Medline: 19089503]
8. Dsouza C. The power of words: how language impacts medical communication. medium. URL: https://medium.com/@kefitup/the-power-of-words-how-language-impacts-medical-communication-aa1ff2947fee [accessed 2023-10-27]
9. Hurtig RR, Alper RM, Berkowitz B. The cost of not addressing the communication barriers faced by hospitalized patients. Perspect ASHA Spec Interest Groups 2018 Jan;3(12):99-112 [FREE Full text] [doi: 10.1044/persp3.SIG12.99] [Medline: 30854467]

XSL•FO
RenderX

10. Bartig S, Kalkum D, Le HM, Lewicki A. Diskriminierungsrisiken und diskriminierungsschutz im gesundheitswesen – wissensstand und forschungsbedarf für die antidiskriminierungsforschung. Deutsches Zentrum fu¨r Integrations- und Migrationsforschung (DeZIM). 2021 Jun 01. URL: https://tinyurl.com/52pa5djj [accessed 2024-11-11]
11. Kommission E. Eurobarometer spezial 243: Die Europäer und ihre sprachen. Europäische Kommission. 2021. URL: https://www.bpb.de/system/files/pdf/1ZSNUZ.pdf [accessed 2023-10-08]
12. Reiff E, Gade C, Böhlich S. Handling the shortage of nurses in Germany: opportunities and challenges of recruiting nursing staff from abroad (IUBH Discussion Papers - Human Resources No. 3/2020). EconStar. 2022 Jul 01. URL: https://www.econstor.eu/handle/10419/222921 [accessed 2024-11-11]
13. Angenendt S, Knapp N, Kipp D. Germany is looking for foreign labour: how to make recruitment development-orientated, sustainable and fair (SWP Research Paper 3). Stiftung Wissenschaft und Politik/German Institute for International and Security Affairs. 2023 Mar 01. URL: https://www.swp-berlin.org/publications/products/research_papers/2023RP03_LookingforForeignLabour.pdf [accessed 2024-11-11]
14. Ariste R, di Matteo L. Non-official language concordance in urban Canadian medical practice: implications for care during the COVID-19 pandemic. Healthc Policy 2021 May;16(4):84-96 [FREE Full text] [doi: 10.12927/hcpol.2021.26497] [Medline: 34129480]
15. Ahmed S, Shommu NS, Rumana N, Barron GRS, Wicklum S, Turin TC. Barriers to access of primary healthcare by immigrant populations in Canada: a literature review. J Immigr Minor Health 2016 Dec;18(6):1522-1540. [doi: 10.1007/s10903-015-0276-z] [Medline: 26364053]
16. Bowen S. Language Barriers in Access to Health Care. Ottawa, ON, Canada: Health Canada; Nov 01, 2001:1-134.
17. Espinoza J, Derrington S. How should clinicians respond to language barriers that exacerbate health inequity? AMA J Ethics 2021 Feb 01;23(2):E109-E116 [FREE Full text] [doi: 10.1001/amajethics.2021.109] [Medline: 33635190]
18. Saito S, Harris MF, Long KM, Lewis V, Casey S, Hogg W, et al. Response to language barriers with patients from refugee background in general practice in Australia: findings from the OPTIMISE study. BMC Health Serv Res 2021 Sep 06;21(1):921 [FREE Full text] [doi: 10.1186/s12913-021-06884-5] [Medline: 34488719]
19. Statistisches Bundesamt. Gesundheitsausgaben im Jahr 2019 auf über 400 milliarden Euro gestiegen. Statistisches Bundesamt (Destatis). 2023. URL: https://www.destatis.de/DE/Presse/Pressemitteilungen/2021/04/PD21_167_236.html [accessed 2024-11-09]
20. Aitken G. Medical students as certified interpreters. AMA J Ethics 2019 Mar 01;21(3):E232-E238 [FREE Full text] [doi: 10.1001/amajethics.2019.232] [Medline: 30893036]
21. Bundesministerium für Gesundheit. Innovationsfonds. Bundesministerium für Gesundheit. 2023. URL: https://www.bundesgesundheitsministerium.de/service/begriffe-von-a-z/i/innovationsfonds [accessed 2023-11-06]
22. Stošović S, Stefanović D, Bogdanović M, Vukotić N. The use of the flutter framework in the development process of hybrid mobile applications. kij 2022 Sep 30;54(3):477-483. [doi: 10.35120/kij5403477s]
23. DigitalOcean. DigitalOcean® cloud computing. DigitalOcean. URL: https://www.digitalocean.com/ [accessed 2023-11-08]
24. IBM. What is a REST API? IBM. 2023. URL: https://tinyurl.com/yz5pe45h [accessed 2023-11-08]
25. Solomon S. What is OAuth? Definition, how it works, and tips for success. Frontegg. 2023 Jul 10. URL: https://frontegg.com/blog/oauth [accessed 2024-02-01]
26. AngularJS. AngularJS developer guide. AngularJS. 2022. URL: https://docs.angularjs.org/guide [accessed 2023-11-08]
27. ReactJS. React - the library for web and native user interfaces. ReactJS. 2024. URL: https://react.dev/ [accessed 2024-04-06]
28. Oracle. MySQL 8.2 reference manual. Oracle. 2023. URL: https://dev.mysql.com/doc/refman/8.2/en/ [accessed 2023-08-11]
29. Firebase. Firebase/Google. 2023. URL: https://firebase.google.com/ [accessed 2023-11-08]
30. Twilio. Twilio Inc. 2024. URL: https://www.twilio.com/en-us [accessed 2024-06-04]
31. Translatly. Translatly UG. 2024 Oct 01. URL: https://translatly.de/ [accessed 2024-11-04]
32. Bader M, Jobst LJ, Zettler I, Hilbig BE, Moshagen M. Disentangling the effects of culture and language on measurement noninvariance in cross-cultural research: The culture, comprehension, and translation bias (CCT) procedure. Psychol Assess 2021 May;33(5):375-384. [doi: 10.1037/pas0000989] [Medline: 33734754]
33. Noll R, Frischen LS, Boeker M, Storf H, Schaaf J. Machine translation of standardised medical terminology using natural language processing: a scoping review. N Biotechnol 2023 Nov 25;77:120-129 [FREE Full text] [doi: 10.1016/j.nbt.2023.08.004] [Medline: 37652265]
34. Urlaub P, Dessein E. Machine translation and foreign language education. Front Artif Intell 2022;5:936111 [FREE Full text] [doi: 10.3389/frai.2022.936111] [Medline: 35937139]
35. Panayiotou A, Gardner A, Williams S, Zucchi E, Mascitti-Meuter M, Goh AM, et al. Language translation apps in health care settings: expert opinion. JMIR Mhealth Uhealth 2019 Apr 09;7(4):e11316 [FREE Full text] [doi: 10.2196/11316] [Medline: 30964446]
36. Herrmann-Werner A, Loda T, Zipfel S, Holderried M, Holderried F, Erschens R. Evaluation of a language translation app in an undergraduate medical communication course: proof-of-concept and usability study. JMIR Mhealth Uhealth 2021 Dec 02;9(12):e31559 [FREE Full text] [doi: 10.2196/31559] [Medline: 34860678]
37. Regenstein M, Park YH, Krips M. The use of interpreters in health centers: a mixed-methods analysis. J Health Care Poor Underserved 2023;34(1):224-245. [doi: 10.1353/hpu.2023.0015] [Medline: 37464491]

XSL•FO
RenderX

38. Statistisches Bundesamt. Bevölkerung mit migrationshintergrund. Bundeszentrale für politische Bildung. 2023. URL: https://www.bpb.de/kurz-knapp/zahlen-und-fakten/soziale-situation-in-deutschland/61646/bevoelkerung-mit-migrationshintergrund/ [accessed 2024-08-17]
39. Lee J. The native speaker: an achievable model? Asian EFL Journal 2005 Jun 01;7(2):152-163 [FREE Full text]
40. Hulstijn J. In: Deusen-Scholl NSNV, editor. Language Proficiency in Native and Non-Native Speakers. Amsterdam, The Netherlands: John Benjamins Publishing Company; 2015.
41. Gazzola M. Language skills and employment status of adult migrants in Europe. In: Jean-Claude B, Hans-Jürgen K, David K, Philia T, editors. The linguistic integration of adult migrants/L'intégration linguistique des migrants adultes. Berlin, Germany: De Gruyter Mouton; 2017:297-302.
42. Regenstein M, Andres E. Hospital language service programs: a closer look at translation practices. J Health Care Poor Underserved 2014 Nov;25(4):2003-2018. [doi: 10.1353/hpu.2014.0162] [Medline: 25418255]
43. Kuo DZ, O'Connor KG, Flores G, Minkovitz CS. Pediatricians' use of language services for families with limited English proficiency. Pediatrics 2007 Apr;119(4):e920-e927. [doi: 10.1542/peds.2006-1508] [Medline: 17371933]
44. Bravo M, Lim R, Poonai N, Chen B. Clinical suspicion and language translation in the pediatric emergency department. Pediatr Emerg Care 2021 May 01;37(5):e272-e274. [doi: 10.1097/PEC.0000000000001606] [Medline: 30130342]
45. Di Filippo M, Avellone A, Belingheri M, Paladino ME, Riva MA, Zambon A, et al. A mobile app leveraging citizenship engagement to perform anonymized longitudinal studies in the context of COVID-19 adverse drug reaction monitoring: development and usability study. JMIR Hum Factors 2022 Nov 04;9(4):e38701 [FREE Full text] [doi: 10.2196/38701] [Medline: 35930561]
46. Montgomery A, Panagopoulou E, Esmail A, Richards T, Maslach C. Burnout in healthcare: the case for organisational change. BMJ 2019 Jul 30;366:l4774. [doi: 10.1136/bmj.l4774] [Medline: 31362957]
47. Joseph C, Garruba M, Melder A. Patient satisfaction of telephone or video interpreter services compared with in-person services: a systematic review. Aust Health Rev 2018 Apr;42(2):168-177. [doi: 10.1071/AH16195] [Medline: 30021688]
48. Gerke S, Stern AD, Minssen T. Germany's digital health reforms in the COVID-19 era: lessons and opportunities for other countries. NPJ Digit Med 2020;3:94 [FREE Full text] [doi: 10.1038/s41746-020-0306-7] [Medline: 32685700]

Abbreviations

AngularJS: Angular Java Script
GDPR: General Data Protection Regulation
MT: machine translation
OAuth: Open Authorization
ReactJS: React Java Script
REST API: Representational State Transfer Application Programming Interface

Edited by A Mavragani; submitted 14.06.24; peer-reviewed by A Islam, K Tekwa; comments to author 19.08.24; revised version received 24.08.24; accepted 16.10.24; published 14.03.25.

Please cite as:
Olsavszky V, Bazari M, Dai TB, Olsavszky A, Finkelmeier F, Friedrich-Rust M, Zeuzem S, Herrmann E, Leipe J, Michael FA, Westernhagen HV, Ballo O
Digital Translation Platform (Translatly) to Overcome Communication Barriers in Clinical Care: Pilot Study
JMIR Form Res 2025;9:e63095
URL: https://formative.jmir.org/2025/1/e63095
doi:10.2196/63095
PMID:

©Victor Olsavszky, Mutaz Bazari, Taieb Ben Dai, Ana Olsavszky, Fabian Finkelmeier, Mireen Friedrich-Rust, Stefan Zeuzem, Eva Herrmann, Jan Leipe, Florian Alexander Michael, Hans von Westernhagen, Olivier Ballo. Originally published in JMIR Formative Research (https://formative.jmir.org), 14.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Multilevel Intervention to Increase Patient Portal Use in Adults With Type 2 Diabetes Who Access Health Care at Community Health Centers: Single Arm, Pre-Post Pilot Study

Robin Whittemore[1*], PhD; Sangchoon Jeon[1*], PhD; Samuel Akyirem[1], RN, MRES; Helen N C Chen[2], PhD; Joanna Lipson[1], MS, MPH; Maritza Minchala[3], MPH; Julie Wagner[4], PhD

[1]School of Nursing, Yale University, Orange, CT, United States

[2]Department of Epidemiology, Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, IN, United States

[3]School of Public Health, Yale University, New Haven, CT, United States

[4]Department of Behavioral Sciences and Community Health, University of Connecticut School of Dental Medicine, Farmington, CT, United States

[*]these authors contributed equally

Corresponding Author:
Robin Whittemore, PhD
School of Nursing
Yale University
400 West Campus Drive
Orange, CT, 06477
United States
Phone: 1 860 420 9623
Email: robin.whittemore@yale.edu

Abstract

Background: Diabetes self-management education and support (DSMS) delivered via patient portals significantly improves glycemic control. Yet, disparities in patient portal use persist. Community health centers (CHCs) deliver care to anyone who needs it, regardless of income or insurance status.

Objective: This study aimed to evaluate the feasibility, acceptability, and preliminary efficacy of a multilevel intervention to increase access and use of portals (MAP) among people with type 2 diabetes (T2D) receiving health care at CHCs.

Methods: A within-subjects, pre-post design was used. Adults with T2D who were portal naive were recruited from 2 CHCs. After informed consent, participants met with a community health worker for referrals for social determinants of health, provision of a tablet with cell service, and individualized training on use of the tablet and portal. Next, a nurse met individually with participants to develop a DSMS plan and then communicated with patients via the portal at least twice weekly during the first 3 months and weekly for the latter 3 months. Data were collected at baseline, 3 months and 6 months. The primary outcome was patient activation and engagement with the portal. Secondary outcomes included technology attitudes, digital health literacy, health-related outcomes and psychosocial function.

Results: In total, 26 patients were eligible, 23 received the intervention, and one was lost to follow up. The sample was predominately Latino or Hispanic (17/22, 77%) and reported low income (19/22, 86%< US $40,000/year), low education (13/22, 59% <high school), and no health insurance (12/22, 55%). All participants had access to a Smartphone, but 91% (20/22) had never accessed a health app. The baseline hemoglobin A_{1c} level was 8.31%. Portal activation was high; 100% (22/22) of participants created a portal account and logged in within the first month. Mean participant logins per week over the first 3 months was 3.16 (SD 1.55) and 1.45 (SD 0.93) over the final 3 months; mean logins per month over the first 3 months was 12.65 (SD 6.21) and 5.79 (SD 3.74) over last 3 months. Engagement was high; 96% (20/21) logged in at least twice per month in the first 3 months and 76% (16/21) between 3 and 6 months. At 6 months, improvements were seen in technology confidence, digital health literacy, diabetes self-efficacy, and diabetes distress. Participant satisfaction with MAP was high as was intention to continue portal use. Barriers to clinical integration and recommendations for portal development were identified.

Conclusions: MAP shows promise for improving health equity in portal use for T2D. Larger, controlled studies are needed to determine how best to implement MAP in complex clinical settings and to evaluate efficacy over time.

Trial Registration: ClinicalTrials.gov NCT05180721; https://clinicaltrials.gov/study/NCT05180721

XSL•FO
RenderX

(JMIR Form Res 2025;9:e67293) doi:10.2196/67293

KEYWORDS

patient portal; mobile phone; diabetes; community health center; adults; diabetic; DM; diabetes mellitus; Type 2 diabetes; T2D; community health centers; CHCs; pilot study; feasibility; self-management; glycemic control; patient portals; social determinants of health; primary outcome; digital health; digital health literacy; health technology; health technologies; psychosocial; efficacy

Introduction

There are significant racial and ethnic disparities in the prevalence, morbidity, and mortality of type 2 diabetes (T2D). Despite medical advances and increased access to medical care, these disparities persist. Racial or ethnic minorities are more likely to have poor glycemic control [1], develop diabetes-related complications [2], and are 1.5-2.3 times more likely to die from diabetes than White individuals [3]. Further, racial and ethnic minorities were particularly affected by COVID-19, with increased risk for infection, morbidity, hospitalization, and mortality [4]. Preexisting conditions, including and especially diabetes, increase the risk for poor COVID-19 outcomes [5,6]. Thus, innovative approaches are urgently needed to address health inequities in T2D.

Patient portals provide secure web-based access to medical records with the capability of messaging providers, filling prescriptions, viewing educational materials, and accessing clinic services. Increased patient portal use has the potential to increase engagement with care and improve diabetes health outcomes. In the general population, patient portal use has been shown to increase office visits while decreasing emergency room visits and hospitalizations [7]. Patient portal use also increases patient knowledge, self-efficacy, decision-making, medication use, and preventive screening [8]. In adults with diabetes, greater portal use of secure messaging with providers led to improved glycemic control (measured based on hemoglobin A_{1c} [HbA_{1c}] levels) compared to nonusers [9-12]. Other trials have demonstrated significant reductions in HbA_{1c} through diabetes self-management education or support (DSMS) via the portal [13-15]. Thus, fostering communication with providers and diabetes self-management support are promising features of portals for adults with diabetes.

There is considerable evidence documenting disparities in patient portal use. Adults who are older, Black, Latinx, and those with low socioeconomic status and low health literacy are less likely to use patient portals as an adjunct to clinical care [16,17]. Despite increased access and popularity of patient portal use in the United States with over 92% of health care organizations offering portals in 2015 [18,19], disparities in portal use continue. More than 100 studies have demonstrated substantial disparities in portal use [17]. Concern has been raised that this well-intentioned solution for patient-centered care may actually worsen health inequities unless portal use among the underserved is increased [17,20,21]. Portal adoption interventions can successfully increase portal use. In a systematic review of interventions to increase portal use in vulnerable populations, 67% (12/18) of studies showed a significant increase in portal use, predictors of use, or reduced disparities. Free or low-cost internet access, technical training and assistance, and proactive outreach from the health care team through the portal were reported to have the strongest evidence for improving health equities in portal use and outcomes [21]. Technical training was the most effective strategy in improving patient portal logins, use of features, and secure messaging [17]. To date, no interventions on portal adoption have targeted adults with T2D of diverse races or ethnicities with limited resources, who have unique structural and social barriers to portal access, use, and diabetes self-management.

Community health centers (CHCs) play a critical role in addressing health inequities in T2D, providing care to over 27 million people in the United States [22]. The aim of CHCs is to provide affordable, high-quality, comprehensive primary care to medically underserved populations, regardless of insurance status or ability to pay for care. Most CHC patients (92%) who live in poverty or near poverty, as defined by the Federal Poverty Level, are disproportionately from racial or ethnic minority groups (total 63%: 32% Hispanic, 22% Black, 9% other minorities), and have high rates of chronic conditions compared to the general population. In 2018, 21% of adults seen at CHCs had T2D compared to 11% in the general population [23].

Social determinants of health (SDoH) are important considerations in developing interventions for adults with T2D who access care at CHCs. The SDoH Equity framework [24], developed by the World Health Organization (WHO), posits that structural determinants of health operate through intermediary and SDoH to shape health outcomes. Intermediary and social determinants include material circumstances (eg, access to tablets or the internet and food security), behavioral, and biological factors (eg, taking medications), psychosocial factors (eg, technology literacy), and the health care system (access to care). To increase patient portal use in adults with T2D accessing care at CHCs, our intervention goals were (1) to address the intermediary and SDoH by providing tablets, home internet, and technology support and (2) to improve engagement with health care by personalizing care and DSMS through known community health workers and nurses employed at the CHC.

The purpose of this study was to pilot-test a multilevel intervention to increase access and use of portals (MAP). We sought to determine MAP's feasibility, acceptability, and potential to improve outcomes over 6 months among adults with T2D who access health care at 2 CHCs. Outcomes included portal-related outcomes (portal activation [logins during the first month], portal engagement [logins over 6 months], digital health literacy, technology acceptability), health-related outcomes (HbA_{1c}, T2D self-management [medication, blood glucose monitoring, healthy eating, and physical activity], and psychosocial function [diabetes self-efficacy, autonomy support, and diabetes distress]).

XSL•FO
RenderX

Methods

Overview

A within-subjects, pre-post design was used to pilot MAP in 22 adults with T2D who were portal naïve. We developed the intervention protocol after seeking feedback from stakeholders on barriers and facilitators to patient portal use and logistics to consider optimizing the implementation of MAP in CHCs [25].

Recruitment

Participants were recruited from 2 CHCs which have been previously described [25]. The clinics are located in Connecticut which is a small and densely populated state with prominent health disparities. In Connecticut, Black residents are nearly 4 times more likely than white residents to have a diabetes-related lower extremity amputation, and among Latino individuals, the rate is nearly 3 times higher than for non-Latino White individuals [26]. Inclusion criteria for this study were as follows: (1) established patient at 1 of the 2 CHCs, (2) age 21-65 years, (3) diagnosed with T2D >6 months, (4) most recent HbA_{1c} measure >7.5%, (5) no use of the patient portal in the past year, (6) no intention of moving or changing clinic within 6 months, and (7) self-reported ability to read in English or Spanish. Exclusion criteria included cognitive impairment (≥3 incorrect answers on the Six Item Screener) [27] or current gestational diabetes.

Participants were recruited from select primary care provider panels at each clinic. A designated clinic staff member reviewed the weekly schedule for potentially eligible recruits and introduced the study to recruits in order to determine preliminary interest. If interested, a trained research assistant explained the study and determined eligibility with a screening questionnaire. If eligible, an appointment was scheduled in person or via Zoom (Zoom Video Communications) for informed consent, baseline data collection, and study enrollment. Written informed consent and data collection were completed in the language preference of the participant (English or Spanish).

Upon completion of baseline data collection, participants were scheduled for the first intervention session. All participants received standard T2D care at the CHC, which followed the guidelines for T2D management as recommended by the American Diabetes Association [28] (eg, quarterly appointments with primary care providers, medical management, and referrals to specialists as indicated). At the study clinics, trained nurses provided diabetes education individually during clinic appointments, as needed. All participants received the MAP intervention from community health workers (CHWs) and nurses employed at the clinics. Clinics were compensated for the salary of the nurses and CHWs for training, delivering the intervention, and completing study tasks.

Training and Supervising Interventionists

Before delivering MAP to study participants, CHWs and nurses were provided a 1-day, in-person, interactive training on the study protocol. They were also provided supportive supervision throughout the study (weekly to biweekly). Training covered orientation to research and goals of the study; human subjects protection; protocol and documentation; and team roles, responsibilities, and supervision. Training also covered details of the intervention content through a study manual and specific strategies for working with low-literacy or low-numeracy individuals with diabetes [29]. Such strategies include using the teach-back method; asking open-ended questions; avoiding unclear statements (“your test was positive”); keeping sessions brief; presenting small chunks of information; encouraging practice between sessions; using nonnumerical measures (eg, 1 serving butter=size of the tip of the thumb); using pictures when possible; using plain language (no jargon or acronyms); reducing the reading level of written materials; and, using a friendly tone. In addition, they were trained in principles of autonomy support. Autonomy support refers to a patient’s perception that their health care provider recognizes the person’s personal agency, encourages self-efficacy, and supports their self-care choices [30].

Intervention

The MAP intervention intentionally and directly intervened on the 4 intermediary determinants of disparities as outlined by the WHO Health Equity framework (material circumstances, psychosocial factors, behavioral and biological factors, and the health care system (Multimedia Appendix 1). All study participants received a tablet (which they were allowed to keep at the end of the study) along with internet access for the 6 months of the study.

The intervention was sequenced to first have CHWs assess SDoH needs using a questionnaire specific to each clinic (each clinic had a slightly different form already in use) and connect the participant to relevant community resources (eg, SNAP benefits). CHWs then provided training on how to gain access to the patient portal and on to mastery of the tablet and portal functionality. In addition to lack of training and lack of encouragement to use the portal, other barriers to patient use of portals addressed by CHWs included doubt about portal usefulness, lost passwords, anxiety about viewing medical information, and privacy concerns [31]. Once the technology training and social determinant referrals were completed, participants were referred to the clinic nurse (Multimedia Appendix 2).

Next, the clinic nurse contacted the participant via the portal to provide DSMS. According to the American Diabetes Association, DSMS is “the support that is required for implementing and sustaining coping skills and behaviors needed to self-manage on an ongoing basis” [32,33]. The nurse initially met individually with study participants to establish rapport, assess diabetes self-management (DSM) behaviors, and develop a DSMS plan collaboratively with the participant. Nurses were asked to communicate with patients via the portal at least twice weekly during the first 3 months and weekly for the latter 3 months, individualizing interactions based on participant needs. In these interactions, nurses assessed challenges and successes with the DSM plan of care and provided education, support, and encouragement. Nurses were also provided with a variety of electronic health education resources that could be sent to participants via the portal. Resources were assembled by the researchers from a thorough review of written and video

XSL•FO
RenderX

materials available in English and Spanish for adults with low health literacy.

Characteristics of each CHC required that MAP be integrated into clinic operations in a tailored fashion. First, one CHC did not employ CHWs and their care coordinators (who usually addressed patient social needs) were not available to deliver the MAP intervention. Therefore, at that clinic, the study nurse was trained to complete both CHW and nursing components of MAP. Second, the 2 clinics used different portal platforms. Therefore, training procedures and written materials were designed to be equivalent between, but tailored to, each clinic and its platform.

Data Collection

Data were collected and managed using REDCap (Research Electronic Data Capture; Vanderbilt University) an National Institutes of Health supported, Food and Drug Administration compliant electronic data capture application for data collection and storage. Data were collected from participants at baseline, 3 months, and 6 months including point-of-care HbA_{1c} values. Research assistants entered data at the time of data collection into the database via tablets or a computer. All participants received a gift card after each wave of data collection—US $40 at baseline, US $40 at 3 months, and US $60 at 6 months. Nurses at the clinics extracted clinical data and information technology specialists at the clinics extracted patient portal data from the electronic health record.

Measures

Overview

Demographics were collected at baseline with questions on age, sex, race or ethnicity, marital status, insurance status, and preferred language (English or Spanish). Participants reported educational attainment and indicated whether they require assistance with reading written health information (never or rarely sometimes or often or always). Participants reported annual household income and also rated their financial strain on a 4-point scale from "We have enough and can save" to "We don't have enough and we have great difficulties" [34].

Clinical Characteristics

BMI data were extracted from the electronic health record with the value most proximal to the date of the baseline assessment. Participants self-reported smoking status, and duration of T2D and completed the Charlson comorbidity index [35] which assesses the presence or absence of 20 common comorbidities.

Portal-Related Outcomes

Portal activation and engagement were recorded as the number of portal logins per participant per month. Portal activation was defined as the creation of a patient portal account and the use of the portal in the first month. The portal engagement was defined as ongoing use of the portal over the 6-month study duration. We calculated logins per week and per month. Based on previous studies, we defined portal engagement (consistent portal use) as two or more patient logins per month [36] and calculated the percentage of participants meeting this benchmark. Digital health literacy was measured with 4 items (eg, "I can use apps [like Zoom] on my cell phone, tablet or computer on my own without asking for help from someone else") on a scale from 1=strongly disagree to 5=strongly agree. Technology acceptability was measured via self-report at baseline, 3 months and 6 months with subscales including ease of use (eg, "I think it will be easy [is easy] to send a message to my provider in the patient portal," perceived usefulness ("I think that using the patient portal will help me [helps me] understand my diabetes care") and confidence (eg, "I am confident in my ability to review my health records on the patient portal") on a scale from 1=strongly disagree to 5=strongly agree [37]. Items were tailored to reference each participant's respective patient portal (MyChart or Healow) [38,39].

Health-Related Outcomes

HbA_{1c} was measured using a fingerstick point of care A1CNow+ (PTS Diagnostics). The validity of this point-of-care assessment of HbA_{1c} has been confirmed through comparisons with clinical laboratory measurements of HbA_{1c} [40,41]. Self-report data included self-management, self-efficacy, perceived autonomy support, and diabetes distress. To assess the reliability of self-report scales, α coefficients were calculated. Self-management was measured by the summary of diabetes self-care activities (eg, diet, medication adherence, blood glucose monitoring, and physical activity). Participants are asked in the last 7 days, how many days did you follow your T2D recommendations for each health behavior (from 0 to 7). Reliability and validity of English and Spanish versions have been established [42,43]. Self-efficacy was measured by the Stanford Diabetes Self-Efficacy Scale, an 8-item scale specific to T2D self-management self-efficacy. Items include confidence in exercise, interpreting blood glucose levels, and following dietary recommendations, with response options from 1=not at all confident to 5=totally confident. Reliability and validity have been established in Spanish-speaking adults [44,45], with an α coefficient of 0.92 in our sample. Perceived autonomy support was measured with the 15-item Health-Care Climate Questionnaire [46], which assesses patients' perceptions of the degree to which they experience the nurse from the CHC to be autonomy supportive versus controlling (eg, "my diabetes provider encourages me to ask questions"). Response options were on a 5-point Likert scale from 1=strongly disagree to 5=strongly agree and the α coefficient was 0.91 in our sample. Diabetes distress was assessed with the 20-item Problem Areas in Diabetes scale (PAID) [47,48] (eg, "feeling overwhelmed by your diabetes") with response options on a 5-point Likert scale from 0=not a problem to 4=serious problem. This widely used scale has evidence of strong reliability and validity; the α coefficient was 0.96 in our sample.

Feasibility and acceptability of data collection were determined as a priori based on an established framework [49]. Feasibility data included recruitment (recruits invited vs consented), attrition (number of participants who withdrew from the study or were lost to follow-up), and participant technology access and use (smartphone, computer, or tablet). Treatment fidelity was assessed by calculating percent adherence to session checklists created for the CHWs and nurses respectively. Barriers to clinical integration were documented by the

XSL•FO
RenderX

researchers during regular supervision with study nurses and research assistants.

Acceptability data included participant surveys at 3 months on satisfaction with the intervention [50] (eg, "Were the topics in the program important to you?") and participant-reported therapeutic alliance with the nurse ("How much do you like or trust or have confidence in your nurse?") [51,52] both rated from 1=not at all to 5=extremely. These questionnaires were modified from published versions to be specific to the MAP intervention protocol. At 6 months, we also asked participants about their intention to continue portal use after study completion (yes or no) and whether they would recommend MAP to a friend (yes or no).

Ethical Considerations

This study (NCT05180721) was approved by the institutional review board at Yale University (IRB# 2000031753; approved on December 21, 2022). All patients provided informed consent before participating in this study.

Data Analysis

All data were downloaded from REDCap onto a secure server. Descriptive analyses were performed to assess demographic and clinical characteristics of the sample. Distributions of outcome variables were examined for central tendency and dispersion. We estimated the Cohen *d* effect size of MAP on A1C and tested the statistical significance of the change from baseline using longitudinal models, including generalized linear mixed model (GLMM), a logistic model with random intercept (which incorporates the correlation within repeated measures), and a negative binomial model with random intercept. The coefficients of the categorial time variable (ie, baseline, 3 months, and 6 months) represent the average change of HbA_{1c} at 3 and 6 months from baseline. The GLMM included all participants with data at baseline and at least one postintervention value and missing data was handled using the maximum likelihood approach. The repeatedly measured secondary outcomes were analyzed with the same approach using GLMM, with each outcome analyzed individually. We used the GLMM identify link function for continuous variables, the logit link function for binary variables, and the log link function for count variables. Residuals were assessed for normality assumption and variables were transformed with an appropriate form when the normality assumption did not hold. Log-in data from one participant was excluded due to errors encountered in downloading it from the portal that our IT expert could not resolve.

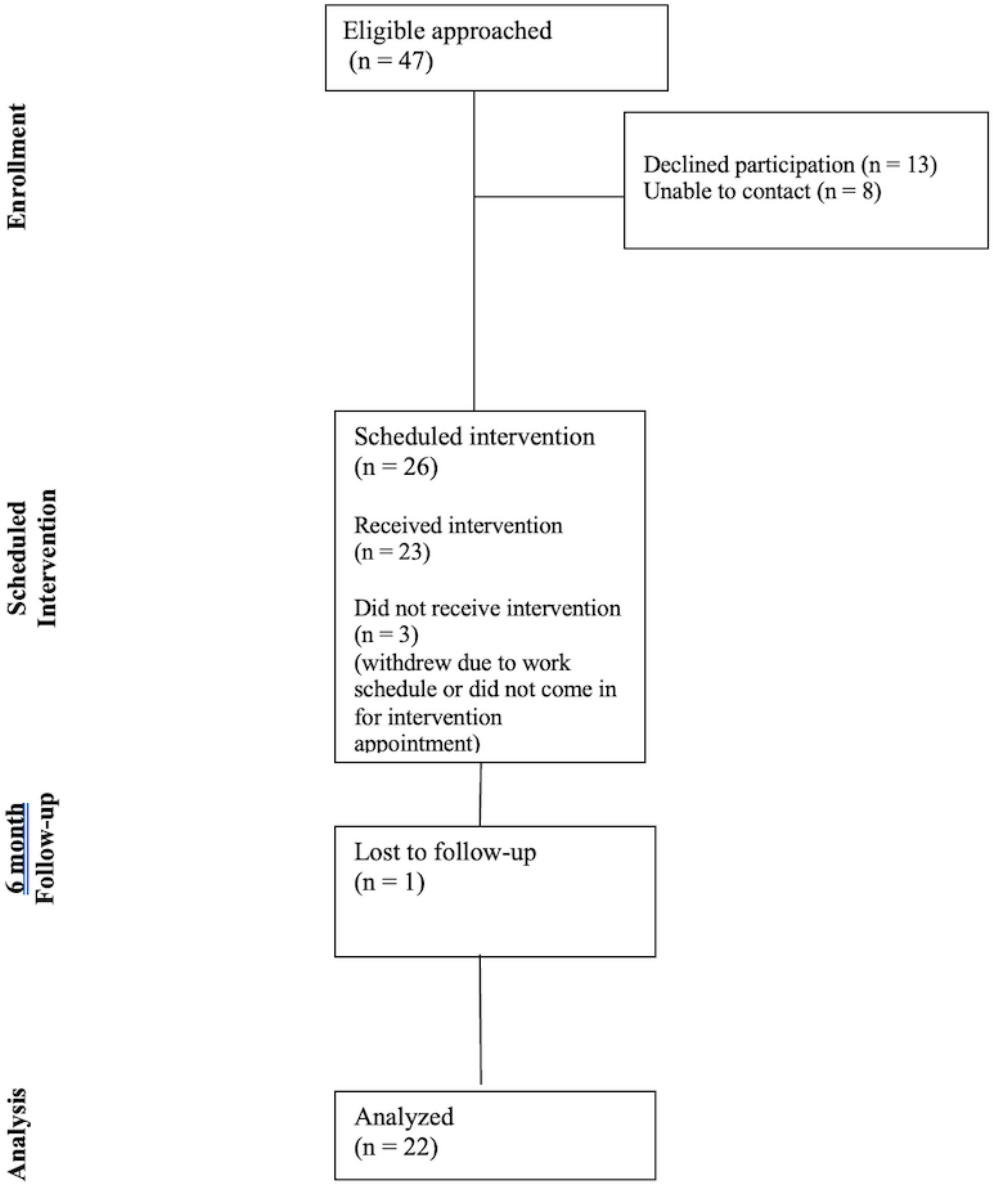
Results

Sample

Recruitment and data collection took place from May 2023 to July 2024. A total of 47 participants were approached about the study, 26 were eligible and provided informed consent, 23 received the intervention and one was lost to follow up (refer to Figure 1 for CONSORT [Consolidated Standards of Reporting Trials] diagram). The sample of completers (n=22) was recruited from 2 clinics, 68% (n=15) from 1 clinic, 32% (n=7) from the other clinic. Refer to Table 1 for demographic characteristics. The sample was predominately Latino or Hispanic (7/22, 77%) had a mean age of 56.32 (SD 10.93) years, 73% (16/22) were female, and 55% (12/22) were married or partnered. The majority reported low-income (86% <US $40,000/year, 19/22), low educational attainment (59%, 13/22 less than high school graduates), and no health insurance (55%, 12/22). All participants had access to a smartphone, but 91% (20/22) had never accessed a health app. Yet, they reported some confidence in using apps, setting up video chats, solving basic technical issues, and using a tablet (mean score 3.1, SD 1.0 with a scale ranging from 1 to 5). They also reported positive perceptions of portal ease of use, usefulness, and confidence in using the portal. The mean duration of diabetes was 11.75 (SD 9.09) years, 36% (8/22) were on insulin at baseline, 14% (3/22) were current smokers, 32% (7/22) reported a history of depression, 59% (13/22) reported severe diabetes distress, 68% (15/22) reported a history of hypertension, and baseline HbA_{1c} was 8.31% (1.05%; measured using the A1CNow system). In total, 73% (16/22) of participants had a baseline HbA_{1c} level greater than the recommended 7.0% for adults with T2D and 73% (16/22) rated their health as fair or poor while 27% (6/22) rated it as good or very good. Mean BMI was 32.21 (SD 5.98), with 23% (5/22) of the sample overweight and 68% (15/22) were categorized as obese. Other baseline clinical data are reported in Table 2 (Data extracted from medical records except for HbA_{1c}, which was collected by researchers for study assessment).

XSL•FO
RenderX

Figure 1. Consolidated Standards of Reporting Trials (CONSORT) diagram.



XSL•FO
RenderX

Table 1. Participant demographics (n=22).

Variables	Values
Site	
Percentage of site 1, n (%)	15 (68)
Age in years, mean (SD)	56.32 (10.93)
Sex	
Percentage of females, n (%)	16 (73)
Duration diabetes in years, mean (SD)	11.75 (9.09)
Charleston comorbidity index, mean (SD)	3.41 (2.32)
Married or partnered (vs not), n (%)	12 (55)
Ethnicity	
Percentage of Hispanic, n (%)	17 (77)
Race, n (%)	
Percentage of White	10 (45)
Percentage of Black	4 (18)
Percentage of other	8; of them, 6 reported Latino/Hispanic (36)
Employment, n (%)	
Percentage of working part/full time (vs others)	8 (36)
Education, n (%)	
Percentage of high school graduates or more	9 (39)
Annual income, n (%)	
Less than US $40,000	19 (86)
Financial difficulties, n (%)	
Percentage of difficulties or great difficulties	15 (68)
Preferred language, n (%)	
Spanish	14 (64)
Spanish–not able to converse in English	12 (55)
Assistance with reading health information	
Sometimes or often or always	10 (46)
Health insurance, n (%)	
No insurance	12 (55)
Self-reported use of community resources, n (%)	
Food assistance	7 (32)
Medication access	6 (27)
Housing	4 (18)
Health insurance	4 (18)
Transportation	2 (9)
Utilities	2 (9)
Childcare or employment	0 (0)
Home internet stability, n (%)	
No home internet or unsure	3 (15)
Fair	6 (27)
Good	7 (32)
Very good	6 (27)

XSL•FO
RenderX

Variables	Values
Home cell phone data stability, n (%)	
Unsure	3 (15)
Fair	6 (27)
Good	7 (32)
Very good	6 (27)

Table 2. Participant clinical characteristics at baseline (n=22).

Variables	Mean (SD)	Recommended values	Meeting recommendations, n (%)
Hemoglobin A_{1c}, %	8.31 (1.50)	<7	6 (28)
BMI	32.31 (5.98)	18.5-24.9	2 (9)
Systolic BP[a], mm Hg	124.57 (14.53)	<130	13 (62)
Diastolic BP, mm Hg	74.48 (7.90)	<80	16 (76)
Cholesterol, mg/dL	161.33 (44.59)	<200	17 (81)
LDL[b], mg/dL	76.57 (35.73)	<100	16 (78)
HDL[c], md/dL	53.00 (26.73)	>40	10 (63)
Triglycerides, mg/dL	164.48 (119.24)	<150	15 (71)

[a]BP: blood pressure.

[b]LDL: low-density lipoprotein.

[c]HDL: high-density lipoprotein.

Protocol Implementation

SDOH needs were assessed before beginning the technology training with 36% (8/22) of participants requiring a referral. Referrals needed were for assistance with utilities (n=4), transportation (n=3), food (n=3), and insurance (n=2). Training on the use of the portal averaged 84.55 (SD 49.32) minutes, with a range between 30-180 minutes. The portal training protocol implementation was high at 85% (524/616 tasks) fidelity across all protocol items and participants. Participants reported moderate confidence in using the portal after the training session with a mean confidence of 2.3 (SD 0.93) on a 3 point-scale (low, moderate, and high confidence). The majority of participants completed the portal training in one session; however, 4 participants (4/22, 18%) required technology support or additional training due to technical challenges using the portal. Nurse protocol implementation was high at 84% (222/264; 12 weeks × 22 participants) at 3 months; but decreased to 52% (137/264) at 6 months with fewer messages sent by nurses over time.

Barriers to integrating MAP into routine clinical care that was noted during regular supervision were varied and included the following: (1) turnover of CHWs, (2) MAP nurses not embedded in the participant's clinical team, (3) portal or clinic constraints on messaging from patient to PCP, (4) difficulty tracking those portal messages from nurses to patients that were not opened and acknowledged, (5) cumbersome and error-prone portal features for transmitting glucose data from patient to clinician, and (6) few high-quality diabetes education videos suitable for low-literacy, low-numeracy, and Spanish-speaking individuals that could be delivered via the portal.

Acceptability

Program satisfaction was high at 3 and 6 months (mean 4.03, SD 0.81 and mean 4.14, SD 0.49 on a 5-point scale respectively). All participants indicated that they would recommend the MAP intervention to a friend. Participants also reported high autonomy support from nurses at 3 months which increased at 6 months. At 6 months, 100% (22/22) of participants reported that they would continue to use the portal for diabetes care and 100% (22/22) felt like the portal was helpful for their diabetes care.

Portal Outcomes

Refer to Table 3 for outcomes over time. Portal activation was high with 100% (22/22) of participants creating a portal account and logging into the portal in the first month (n=21). Mean logins per week was 3.16 (SD 1.55) over the first 3 months and 1.45 (SD 0.93) between 3 and 6 months (Figure 2). Mean portal logins per month were 12.65 (SD 6.21) for the first 3 months and 5.79 (SD 3.74) for the last 3 months. Participant engagement (as defined by the literature) was high, with 96% (20/21) of participants logging into the portal at least twice per month in the first 3 months and 76% (16/21) of participants meeting this benchmark between 3 and 6 months. At baseline, participants perceived that the portal would be easy to use (mean score 4.0, SD 0.55 on a 5-point scale) and useful (mean score 4.25, SD 0.44) with no significant change over time. There was a significant increase in technology confidence over 6 months (P<.05), with a trend for increased digital health literacy at 6 months (P=.08).

XSL•FO
RenderX

Table 3. Change over time for portal and health outcomes (n=22).

	Possible Range	Baseline	3 months	6 months
Portal-related outcomes				
Portal logins/week[a]	—[b]	—	3.16 (3.18)	1.45 (1.91)
Portal logins/month[a]	—	—	12.65 (6.21)	5.79 (3.74)
Portal engagement[a]	—	—	96% met benchmark	76% met benchmark
Technology confidence	1-5	3.81 (0.86)	4.00 (0.73)	4.38 (0.56)[c]
Digital health literacy	1-5	3.15 (1.03)	3.23 (0.93)	3.59 (0.87)[c]
Portal perceived ease of use	1-5	4.00 (0.55)	3.89 (0.78)	4.08 (0.67)
Portal perceived usefulness	1-5	4.25 (0.44)	4.15 (0.660	4.31 (0.51)
Health-related outcomes				
Hemoglobin A_{1c}, %	—	8.31 (1.65)	8.09 (1.64)	8.23 (1.35)
PAID[d]	0-100	43.58 (30.51)	29.09 (23.03)[c]	25.45 (24.88)[c]
DSM[e] Self-Efficacy	1-5	3.14 (1.01)	3.44 (1.17)	3.68 (1.03)[c]
Healthcare Climate Questionnaire	1-5	4.15 (0.78)	4.14 (0.83)	3.95 (1.07)
SDSCA[f]	0-7	—	—	—
Diet	—	2.82 (2.91)	4.00 (2.54)	3.68 (2.80)
Exercise	—	2.61 (2.23)	3.23 (2.11)	3.41 (2.29)
Glucose checking	—	4.66 (2.49)	4.95 (2.44)	5.41 (2.21)
Footcare	—	5.30 (1.01)	1. 1.06)	1.79 (.94)
Medication	—	6.65 (1.00)	6.90 (0.31)	6.71 (0.73)
Use of community resources	0-9	1.32 (1.52)	1.32 (1.55)	1.36 (1.65)

[a]n=21.

[b]Not applicable.

[c]significant change from baseline at a 5% significance level.

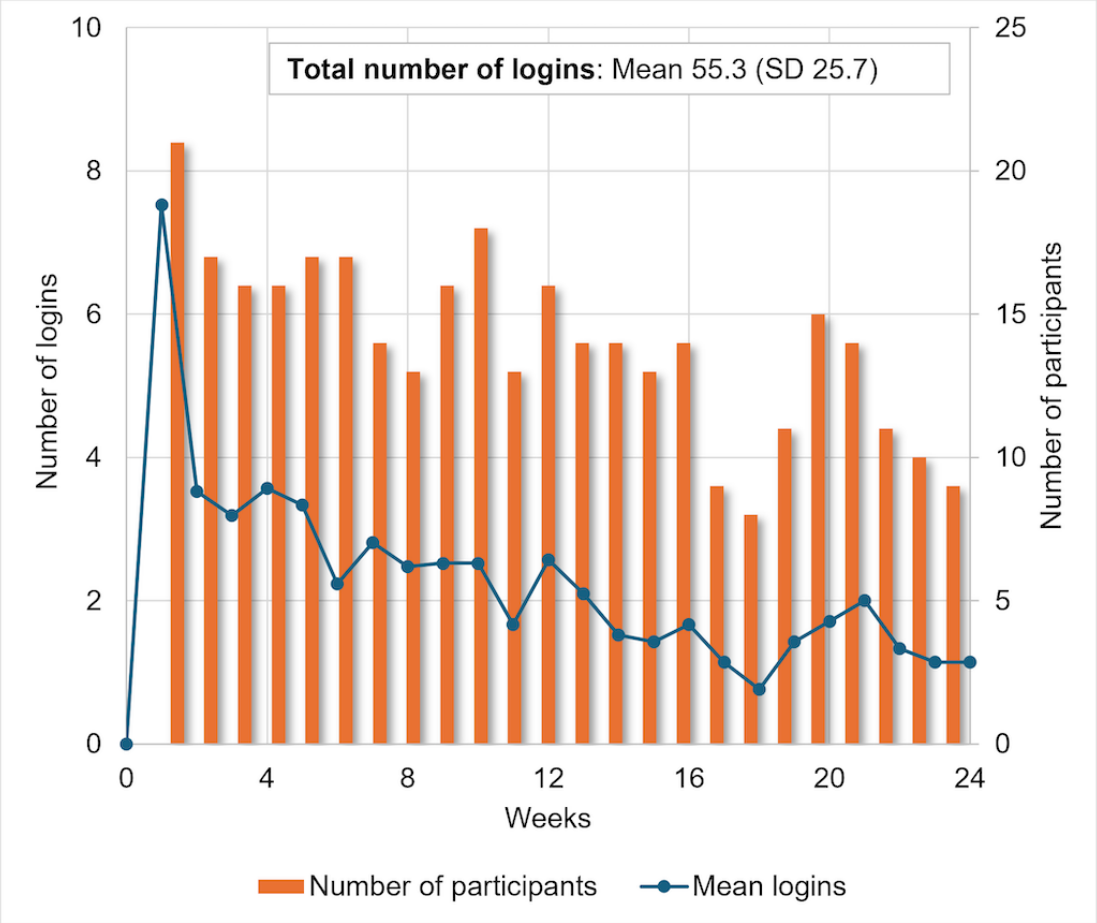
[d]PAID: Problem Areas in Diabetes (diabetes distress).

[e]DSM: diabetes self-management.

[f]SDSCA: Summary of Diabetes Self-Care Behaviors; the Healthcare Climate questionnaire measures autonomy support.

XSL•FO
RenderX

Figure 2. Portal logins over time.



Health Outcomes

In the mixed effect model, diabetes distress (PAID) significantly decreased from baseline to 3 months; it continued to decrease significantly from 3 months to 6 months (P<.01; Table 3). The proportion of clinically elevated diabetes distress (ie, PAID>40) was 50% (11/22) at baseline, and it decreased to 36.4% (8/22) at 3 months (P=.20) and 22.7% (5/22) at 6 months (P=.03). Diabetes self-efficacy increased from baseline to 3 months and continued to increase until 6 months when it became statistically significant. A similar pattern of significant change from baseline to 6 months was found for confidence in using technology and digital health literacy. HbA_{1c} was slightly decreased at 3 months with a small effect size (Cohen d=–0.17) but returned to baseline level at 6 months (Cohen d=–0.07). There was no significant change in perceived autonomy support or diabetes self-care behaviors.

Discussion

Principal Findings

This pilot study evaluated an intervention to increase patient portal use in adults with T2D who access health care at CHCs. We reached a sample of adults of diverse races or ethnicities, low income, and low educational attainment with low use of the patient portal. The MAP intervention was specifically designed to address multiple levels of challenges identified by CHC patients and providers to patient portal usage. Thus, to enhance successful implementation, participants received a tablet and data plan, SDOH screening, training in patient portal navigation with ongoing support, and diabetes education and support by nurses. Overall, feasibility, acceptability, and improvement in portal and health outcomes of clinical significance were demonstrated yet integration into clinical care was challenging.

First, we demonstrated high feasibility by tailoring implementation strategies to meet the unique needs of each clinic. We also demonstrated low attrition, high patient portal activation (22/22, 100%), and high patient portal engagement over 6 months. High participant activation is not surprising as we had dedicated staff available to train participants in portal use and assist with any challenges. Although the measurement of patient portal engagement is not standardized in the literature, a common metric is the frequency of logins per month. Consistent portal use has been defined as two or more patient logins per month according to a recent systematic review of the measurement of patient portal use [36]. In another study, super-users were defined as those logging in twice or more per month [53]. In our study, consistent portal use was demonstrated in 96% of participants over 3 months, 76% of participants between 3 and 6 months, and 87% over a 6-month period among 21 patients. Thus, the results of our portal activation and engagement are encouraging. Factors that likely contributed to these positive outcomes were technology training or support for the duration of the study, participants' development of a relationship with a clinic nurse, and consistent outreach by the

XSL•FO
RenderX

nurse over 6 months via the portal to provide DSMS. While the provision of a tablet and a data plan may have also contributed to portal engagement over time in this study, we judge that increasing access to smartphones combined with free internet sites in community settings (eg, libraries and coffee shops) may be sufficient for future implementation efforts. Decline over time in consistent use may be partially explained by less consistent protocol implementation by nurses from 3 to 6 months and lack of full integration of MAP into clinic workflow.

Engagement over time was supported by improvements in other portal-related outcomes. Whereas there was no change in how participants viewed the portal per se (perceived usefulness and perceived ease of use), participant perceptions of their own interactions with the technology did improve. Increases were observed in both technology confidence ("I am confident in my ability to use MyChart") and digital health literacy ("I can solve, or figure out how to solve, basic technical issues on my cell phone, computer, or other device"). Those improvements reached statistical significance at 6 months suggesting that individuals may need more than 3 months of supported use of the tablet and the portal to create change in confidence and digital health literacy. Also encouraging was the high level of acceptability including participant satisfaction with the MAP intervention and high intention to continue use of the portal after the MAP study ended.

Second, some clinically important health-related outcomes significantly improved in our small sample. The decrease in diabetes distress at 3 months, with a further decrease at 6 months, is important, given the body of research showing that diabetes distress is robustly associated with lower self-management and may be associated with higher HbA_{1c} [54,55]. The MAP intervention's reduction in diabetes distress is consistent with findings from other interventions. A meta-analysis of 41 studies testing eHealth interventions for diabetes self-management found that such interventions are effective in reducing diabetes distress [56]. Our rate of elevated diabetes distress (59%) is higher than a meta-analysis of 55 studies, which showed an overall prevalence of 36% of diabetes distress in people with type 2 diabetes, with higher rates in predominantly female samples and those with more comorbidities [57]. Our small sample size, comprised of participants with elevated HbA_{1c} and who enrolled in a treatment study may account for this high rate of distress. Alternatively, diabetes self-management may be exceptionally distressing in the context of high SDoH needs, as is common among patients at CHCs. In this way, MAP may be particularly beneficial for adults with T2D who access health care at CHCs. One study in T2D found that patients who had high diabetes distress at baseline had a greater increase in self-management and a decrease in HbA_{1c} from mHealth DSMS compared to their low-distress counterparts [58].

Diabetes self-efficacy increased at 3 months but only became statistically significant at 6 months, which may reflect our small sample size since larger studies have shown increases by 3 months [59]. Alternatively, this sample with elevated HbA_{1c}, high diabetes distress, and high social needs may require 6 months of intervention to increase self-efficacy. Diabetes self-efficacy has been shown in a systematic review to be an important predictor of self-care behaviors [60] including for example medication adherence [61]. The improvement in diabetes self-efficacy observed in our study may prove important for downstream clinical outcomes since interventions that improve self-efficacy have been shown to directly and indirectly improve glycemic control [62].

We observed that HbA_{1c} and DSM improved at 3 and 6 months relative to baseline values but these improvements were not statistically significant in our small sample. Research on the effectiveness of patient portals on glycemic outcomes has produced inconsistent results [13,14,63-65]; yet, observational studies with adequate sample sizes do in fact provide consistent evidence that patient portal use has a beneficial impact on HbA_{1c} [66-68]. The consistency of portal use is also associated with benefits for HbA_{1c}. In a study that examined the consistency of patient portal use among adults with T2D (N=95,043), an increasing number of calendar months of patient portal use was associated with a significant decrease in HbA_{1c} levels [69]. In addition, several studies have highlighted the role of patient portals in improving diabetes self-management behaviors. A retrospective cohort study involving over 100,000 adults with diabetes found that when patients are given access to the patient portal via a mobile device, their adherence to oral antihyperglycemic medications improves significantly [66].

In addition to the small sample size, our study may not have produced significant improvements in HbA_{1c} and DSM because of the low degree of integration of MAP into clinical care. We ensured that MAP was delivered in a clinical setting by embedded clinicians, yet we encountered a number of barriers to a deeper level of clinical integration. For example, in one clinic, participants were not able to send portal messages directly to their PCPs because that portal functionality had not yet been activated clinic-wide. As another example, at the other clinic, the MAP nurse was not part of the patient's own clinical team. Whereas MAP nurses did send monthly progress notes to the patients' respective PCPs, it is unknown how the PCPs made use of the reports. In our study, one nurse was bilingual in English or Spanish; however, not all were and one participant had Polish as a first language. In these situations, translation of messages was required. Thus, MAP's good metrics of implementation were only made possible with creative approaches and "work arounds" to clinic and portal idiosyncrasies. It remains untested how effectively MAP could improve glycemic control and DSM when fully integrated into clinical care.

Health care providers and adults with diabetes in CHCs continue to face barriers that limit their use of portals. Well-documented challenges include low health- and technical literacy, lack of regular access to internet-connected mobile phones, limited language concordance between providers and patients, and other SDoH-related issues [25,70]. CHCs also face the challenge of lacking standardized implementation strategies for rolling out portals and supporting patients to use portals. The findings from the current pilot study provide preliminary evidence of what may work in CHC settings to mitigate these barriers and improve portal use among adults with T2D.

XSL•FO
RenderX

Clinical Implications

Our findings have several clinical implications. First, having clinicians recommend the use of the patient portal and having dedicated staff to provide training and technical support to patients in CHCs can be useful in improving portal use. Racial and ethnic differences in portal use attenuate when clinicians offer portals to all patients irrespective of their racial and ethnic background [71]. In addition, a recent systematic review to summarize research priorities and best practices in patient portals found that studies that report high rates of portal use often have dedicated staff to enroll and assist patients [72]. Evidence-based techniques to educate socially vulnerable patients, such as screening for health literacy, using the teach-back method, and using bilingual messaging, were used in our intervention and can also be adopted by clinicians in CHCs. Many CHCs incorporate nursing case management for patients with chronic conditions. Nurse case management communication could also be expanded to use the patient portal for ongoing check-ins, support, and patient education, which may enhance portal engagement and health outcomes. In a systematic review of the effectiveness of patient education through patient portals, significant improvements in knowledge, self-management, health behaviors, mental health, and select health outcomes were demonstrated [73].

Lastly, clinic administrators at CHCs could work with their electronic health record support team to eliminate the need for email accounts to activate the patient portal. Patients who access care at CHCs may not have email accounts, which creates an additional barrier to portal use among CHC patients [74]. Alternative approaches such as the use of phone numbers to authenticate users should be encouraged, as the majority of patients in CHCs have access to mobile phones.

Limitations and Future Research

Findings should be interpreted with caution given study limitations, the primary of which is a design without a comparison group or randomization to treatment. Future research should test MAP using a more rigorous study design. Implementation research should identify strategies to more fully integrate interventions like MAP into complex clinical settings. Common metrics of portal use should be used as outcomes, in addition to study-specific indicators, in order to facilitate comparison across studies. Our sample size at each clinic was insufficient to test for differences between clinics in sample characteristics or differences in outcomes. Study samples should also be large enough to be powered to test for clinically meaningful improvements in HbA_{1c} and to include adequate representation of medically underserved individuals so that intervention effects on disparities in portal use—both ameliorating and exacerbating—can be examined. A longer duration of follow-up would allow for investigation of the durability of any treatment effect.

There is also a need for research to develop portal platforms and functionality that better meet the needs of patients and clinicians. This could include, for example, graphical and pictorial data displays, audiovisual capabilities, and libraries of patient education materials. Beyond the patient portal, our preparation for the intervention implementation revealed a paucity of high-quality, patient-friendly diabetes educational materials (pictorial, video, interactive, and gaming) that are designed for low-literacy, low-numeracy, and Spanish-speaking individuals. Whereas a wide variety of materials exist, few if any meet all the above criteria, and those that try tend to be extremely cursory.

Conclusions

In this pilot study among people with diabetes receiving care at CHCs, the MAP intervention produced high activation and engagement in portal use as well as meaningful improvement in psychosocial outcomes and promising changes in clinical outcomes. Numerous challenges were also identified that can be addressed in future research. Increasing portal use specifically in health disparities populations may be one key to ameliorating disparities in diabetes outcomes.

Acknowledgments

National Institute on Minority Health and Health Disparities (R21MD016949).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Multilevel intervention to increase access and use of portals (MAP) components aligned with the WHO (World Health Organization) Health Equity framework.
[DOCX File , 15 KB - formative_v9i1e67293_app1.docx]

Multimedia Appendix 2
Sequence and timing of multilevel intervention to increase access and use of portals (MAP).
[DOCX File , 14 KB - formative_v9i1e67293_app2.docx]

XSL•FO
RenderX

References

1. Kirk J, D'Agostino RBJ, Bell RA, Passmore LV, Bonds DE, Karter AJ, et al. Disparities in HbA1c levels between African-American and non-hispanic white adults with diabetes: a meta-analysis. Diabetes Care 2006;29(9):2130-2136 [FREE Full text] [doi: 10.2337/dc05-1973] [Medline: 16936167]
2. Osborn CY, de Groot M, Wagner JA. Racial and ethnic disparities in diabetes complications in the northeastern United States: the role of socioeconomic status. J Natl Med Assoc 2013;105(1):51-58 [FREE Full text] [doi: 10.1016/s0027-9684(15)30085-7] [Medline: 23862296]
3. Spanakis EK, Golden SH. Race/ethnic difference in diabetes and diabetic complications. Curr Diab Rep 2013;13(6):814-823 [FREE Full text] [doi: 10.1007/s11892-013-0421-9] [Medline: 24037313]
4. Yancy CW. COVID-19 and African Americans. JAMA 2020;323(19):1891-1892. [doi: 10.1001/jama.2020.6548] [Medline: 32293639]
5. Yan Y, Yang Y, Wang F, Ren H, Zhang S, Shi X, et al. Clinical characteristics and outcomes of patients with severe covid-19 with diabetes. BMJ Open Diabetes Res Care 2020;8(1):e001343 [FREE Full text] [doi: 10.1136/bmjdrc-2020-001343] [Medline: 32345579]
6. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, the Northwell COVID-19 Research Consortium, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA 2020;323(20):2052-2059 [FREE Full text] [doi: 10.1001/jama.2020.6775] [Medline: 32320003]
7. Reed ME, Huang J, Brand RJ, Neugebauer R, Graetz I, Hsu J, et al. Patients with complex chronic conditions: health care use and clinical events associated with access to a patient portal. PLoS One 2019;14(6):e0217636 [FREE Full text] [doi: 10.1371/journal.pone.0217636] [Medline: 31216295]
8. Han HR, Gleason KT, Sun CA, Miller HN, Kang SJ, Chow S, et al. Using patient portals to improve patient outcomes: systematic review. JMIR Hum Factors 2019;6(4):e15038 [FREE Full text] [doi: 10.2196/15038] [Medline: 31855187]
9. Devkota B, Salas J, Sayavong S, Scherrer JF. Use of an online patient portal and glucose control in primary care patients with diabetes. Popul Health Manag 2016;19(2):125-131. [doi: 10.1089/pop.2015.0034] [Medline: 26237200]
10. Chung S, Panattoni L, Chi J, Palaniappan L. Can secure patient-provider messaging improve diabetes care? Diabetes Care 2017;40(10):1342-1348. [doi: 10.2337/dc17-0140] [Medline: 28807977]
11. Harris SM, Joyce H, Miller A, Connor C, Amiel SA, Mulnier H. The attitude of healthcare professionals plays an important role in the uptake of diabetes self-management education: analysis of the barriers to uptake of type 1 diabetes education (BUD1E) study survey. Diabet Med 2018. [doi: 10.1111/dme.13704] [Medline: 29882990]
12. Kuo A, Dang S. Secure messaging in electronic health records and its impact on diabetes clinical outcomes: a systematic review. Telemed J E Health 2016;22(9):769-777. [doi: 10.1089/tmj.2015.0207] [Medline: 27027337]
13. Tang PC, Overhage JM, Chan AS, Brown NL, Aghighi B, Entwistle MP, et al. Online disease management of diabetes: engaging and motivating patients online with enhanced resources-diabetes (EMPOWER-D), a randomized controlled trial. J Am Med Inform Assoc 2013;20(3):526-534 [FREE Full text] [doi: 10.1136/amiajnl-2012-001263] [Medline: 23171659]
14. Ralston JD, Hirsch IB, Hoath J, Mullen M, Cheadle A, Goldberg HI. Web-based collaborative care for type 2 diabetes: a pilot randomized trial. Diabetes Care 2009;32(2):234-239 [FREE Full text] [doi: 10.2337/dc08-1220] [Medline: 19017773]
15. Alturkistani A, Qavi A, Anyanwu PE, Greenfield G, Greaves F, Costelloe C. Patient portal functionalities and patient outcomes among patients with diabetes: systematic review. J Med Internet Res 2020;22(9):e18976 [FREE Full text] [doi: 10.2196/18976] [Medline: 32960182]
16. Anthony DL, Campos-Castillo C, Lim PS. Who Isn't using patient portals and why? Evidence and implications from a national sample Of US adults. Health Aff (Millwood) 2018;37(12):1948-1954. [doi: 10.1377/hlthaff.2018.05117] [Medline: 30633673]
17. Grossman L, Masterson Creber RM, Benda N, Wright D, Vawdrey D, Ancker J. Interventions to increase patient portal use in vulnerable populations: a systematic review. J Am Med Inform Assoc 2019;26(8-9):855-870 [FREE Full text] [doi: 10.1093/jamia/ocz023] [Medline: 30958532]
18. American Heart Association. Individuals' Ability to Electronically Access Their Hospital Medical Records, Perform Key Tasks Is Growing. URL: https://www.aha.org/system/files/research/reports/tw/16jul-tw-healthIT.pdf [accessed 2021-02-04]
19. Adler-Milstein J, DesRoches CM, Kralovec P, Foster G, Worzala C, Charles D, et al. Electronic health record adoption in US hospitals: progress continues, but challenges persist. Health Aff (Millwood) 2015;34(12):2174-2180. [doi: 10.1377/hlthaff.2015.0992] [Medline: 26561387]
20. Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018;25(8):1080-1088 [FREE Full text] [doi: 10.1093/jamia/ocy052] [Medline: 29788380]
21. Lorenc T, Petticrew M, Welch V, Tugwell P. What types of interventions generate inequalities? Evidence from systematic reviews. J Epidemiol Community Health 2013;67(2):190-193. [doi: 10.1136/jech-2012-201257] [Medline: 22875078]
22. The number of patients continues to grow at community health centers. National Association of Community Health Centers. 2017. URL: https://www.nachc.org/the-number-of-patients-continues-to-grow-at-community-health-centers/ [accessed 2020-10-01]

XSL•FO
RenderX

23. Community health center chart book. National Association of Community Health Centers. 2019. URL: http://www.nachc.org/wp-content/uploads/2019/01/Community-Health-Center-Chartbook-FINAL-1.28.19.pdf [accessed 2020-10-01]
24. Solar O, Irwin AA. A conceptual framework for action on the social determinants of health. Social Determinants of Health Discussion Paper.: World Health Organiation; 2010. URL: https://www.who.int/sdhconference/resources/ConceptualframeworkforactiononSDH_eng.pdf [accessed 2021-02-01]
25. Akyirem S, Wagner J, Chen HN, Lipson J, Minchala M, Cortez K, et al. Recommendations to address barriers to patient portal use among persons with diabetes seeking care at community health centers: interview study with patients and health care providers. JMIR Diabetes 2024;9:e58526 [FREE Full text] [doi: 10.2196/58526] [Medline: 39284181]
26. Becker AL. Health Disparities in Connecticut: Causes, Effects, and What We Can Do. 2020. URL: https://www.cthealth.org/wp-content/uploads/2020/01/Health-disparities-in-Connecticut.pdf [accessed 2025-02-28]
27. Callahan CM, Unverzagt FW, Hui SL, Perkins AJ, Hendrie HC. Six-item screener to identify cognitive impairment among potential subjects for clinical research. Med Care 2002;40(9):771-781. [doi: 10.1097/00005650-200209000-00007] [Medline: 12218768]
28. American Diabetes Association Professional Practice Committee. Introduction and methodology: standards of care in diabetes-2024. Diabetes Care 2024;47(Supplement_1):S1-S4. [doi: 10.2337/dc24-SINT] [Medline: 38078587]
29. Osborn CY, Cavanaugh K, Kripalani S. Strategies to address low health literacy and numeracy in diabetes. Clinical Diabetes 2010;28(4):171-175. [doi: 10.2337/diaclin.28.4.171]
30. Robinson SA, Zocchi MS, Netherton D, Ash A, Purington CM, Connolly SL, et al. Secure messaging, diabetes self-management, and the importance of patient autonomy: a mixed methods study. J Gen Intern Med 2020;35(10):2955-2962 [FREE Full text] [doi: 10.1007/s11606-020-05834-x] [Medline: 32440998]
31. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
32. Diabetes self-management education and support toolkit. Centers for Disease and Control. 2020. URL: https://www.cdc.gov/diabetes-toolkit/php/about-dsmes/index.html [accessed 2025-02-28]
33. Powers MA, Bardsley J, Cypress M, Duker P, Funnell MM, Fischl AH, et al. Diabetes self-management education and support in type 2 diabetes. Diabetes Educ 2017;43(1):40-53. [doi: 10.1177/0145721716689694] [Medline: 28118121]
34. Paras PP-E PR. El rostro de la pobreza: la inseguridad alimentaria en el Distrito Federal. Rev Este Pais 2004;158:45-50.
35. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [doi: 10.1016/0021-9681(87)90171-8] [Medline: 3558716]
36. Beal LL, Kolman JM, Jones SL, Khleif A, Menser T. Quantifying patient portal use: systematic review of utilization metrics. J Med Internet Res 2021;23(2):e23493 [FREE Full text] [doi: 10.2196/23493] [Medline: 33629962]
37. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989;13(3):319-340. [doi: 10.2307/249008]
38. Whittemore R, Vilar-Compte M, De La Cerda S, Delvy R, Jeon S, Burrola-Méndez S, et al. A self-management randomized controlled pilot trial for low-income adults with type 2 diabetes in Mexico City. Curr Dev Nutr 2020;4(5):nzaa074 [FREE Full text] [doi: 10.1093/cdn/nzaa074] [Medline: 32368713]
39. Wagner J, Berthold SM, Buckley T, Bermudez-Millan A, Ha T, Scully M, et al. Remote peer learning between US and cambodian lay health workers to improve outcomes for cambodians with type 2 diabetes: a pilot study. Int J Behav Med 2020;27(5):609-614. [doi: 10.1007/s12529-020-09896-3] [Medline: 32435878]
40. Moskowitz KW, Mercer K, Suscha E, Shafai M, Lindke C, Xie C, et al. Multi-Center Accuracy Assessment of A1CNow® + : A Disposable System for Monitoring Hemoglobin A1c. 2017 Presented at: American Diabetes Association; June 2017; 1701 N BEAUREGARD ST, ALEXANDRIA, VA 22311-1717, United States p. A-231 URL: https://ptsdiagnostics.com/wp-content/uploads/2018/08/lit_001815_r0_abstract_multi_center_accuracy_a1cnow_system.pdf
41. Nathan DM, Griffin A, Perez FM, Basque E, Do L, Steiner B. Accuracy of a point-of-care hemoglobin A1c assay. J Diabetes Sci Technol 2019;13(6):1149-1153 [FREE Full text] [doi: 10.1177/1932296819836101] [Medline: 30943787]
42. Toobert DJ, Hampson SE, Glasgow RE. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. Diabetes Care 2000;23(7):943-950. [doi: 10.2337/diacare.23.7.943] [Medline: 10895844]
43. Vincent D, McEwen MM, Pasvogel A. The validity and reliability of a Spanish version of the summary of diabetes self-care activities questionnaire. Nurs Res 2008;57(2):101-106. [doi: 10.1097/01.NNR.0000313484.18670.ab] [Medline: 18347481]
44. Ritter PL, Lorig K, Laurent DD. Characteristics of the Spanish- and English-Language self-efficacy to manage diabetes scales. Diabetes Educ 2016;42(2):167-177. [doi: 10.1177/0145721716628648] [Medline: 26846336]
45. Victorson D, Banas J, Smith J, Languido L, Shen E, Gutierrez S, et al. eSalud: designing and implementing culturally competent ehealth research with latino patient populations. Am J Public Health 2014;104(12):2259-2265. [doi: 10.2105/AJPH.2014.302187] [Medline: 25320901]
46. Williams GC, Freedman ZR, Deci EL. Supporting autonomy to motivate patients with diabetes for glucose control. Diabetes Care 1998;21(10):1644-1651. [doi: 10.2337/diacare.21.10.1644] [Medline: 9773724]

XSL•FO
RenderX

47. Beléndez M, Hernández-Mijares A, Marco J, Domínguez JR, Pomares FJ. Validation of the Spanish version of the problem areas in diabetes (PAID-SP) scale. Diabetes Res Clin Pract 2014;106(3):e93-e95. [doi: 10.1016/j.diabres.2014.09.012] [Medline: 25451898]
48. Polonsky WH, Anderson BJ, Lohrer PA, Welch G, Jacobson AM, Aponte JE, et al. Assessment of diabetes-related distress. Diabetes Care 1995;18(6):754-760. [doi: 10.2337/diacare.18.6.754] [Medline: 7555499]
49. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
50. Wagner J, Bermudez-Millan A, Damio G, Segura-Perez S, Chhabra J, Vergara C, et al. Community health workers assisting Latinos manage stress and diabetes (CALMS-D): rationale, intervention design, implementation, and process outcomes. Transl Behav Med 2015;5(4):415-424 [FREE Full text] [doi: 10.1007/s13142-015-0332-1] [Medline: 26622914]
51. Budman SH, Soldz S, Demby A, Feldstein M, Springer T, Davis MS. Cohesion, alliance and outcome in group psychotherapy. Psychiatry 1989;52(3):339-350. [doi: 10.1080/00332747.1989.11024456] [Medline: 2772092]
52. Wagner J, Bermudez-Millan A, Buckley T, Buxton OM, Feinn R, Kong S, et al. A randomized trial to decrease risk for diabetes among cambodian Americans with depression: intervention development, baseline characteristics and process outcomes. Contemp Clin Trials 2021;106:106427 [FREE Full text] [doi: 10.1016/j.cct.2021.106427] [Medline: 33957272]
53. Toscos T, Daley C, Heral L, Doshi R, Chen YC, Eckert GJ, et al. Impact of electronic personal health record use on engagement and intermediate health outcomes among cardiac patients: a quasi-experimental study. J Am Med Inform Assoc 2016;23(1):119-128 [FREE Full text] [doi: 10.1093/jamia/ocv164] [Medline: 26912538]
54. Gonzalez JS, Krause-Steinrauf H, Bebu I, Crespo-Ramos G, Hoogendoorn CJ, Naik AD, GRADE Research Group. Emotional distress, self-management, and glycemic control among participants enrolled in the glycemia reduction approaches in diabetes: a comparative effectiveness (GRADE) study. Diabetes Res Clin Pract 2023;196:110229 [FREE Full text] [doi: 10.1016/j.diabres.2022.110229] [Medline: 36549506]
55. Skinner TC, Joensen L, Parkin T. Twenty-five years of diabetes distress research. Diabet Med 2020;37(3):393-400. [doi: 10.1111/dme.14157] [Medline: 31638279]
56. Fernández-Rodríguez R, Zhao L, Bizzozero-Peroni B, Martínez-Vizcaíno V, Mesas AE, Wittert G, et al. Are e-health interventions effective in reducing diabetes-related distress and depression in patients with type 2 diabetes? A systematic review with meta-analysis. Telemed J E Health 2024;30(4):919-939. [doi: 10.1089/tmj.2023.0374] [Medline: 38010739]
57. Perrin NE, Davies MJ, Robertson N, Snoek FJ, Khunti K. The prevalence of diabetes-specific emotional distress in people with type 2 diabetes: a systematic review and meta-analysis. Diabet Med 2017;34(11):1508-1520. [doi: 10.1111/dme.13448] [Medline: 28799294]
58. Clark TL, Gallo L, Euyoque JA, Philis-Tsimikas A, Fortmann A. Does diabetes distress influence clinical response to an mHealth diabetes self-management education and support intervention? Diabetes Educ 2020;46(3):289-296 [FREE Full text] [doi: 10.1177/0145721720913276] [Medline: 32228288]
59. Young HM, Miyamoto S, Dharmar M, Tang-Feldman Y. Nurse coaching and mobile health compared with usual care to improve diabetes self-efficacy for persons with type 2 diabetes: randomized controlled trial. JMIR Mhealth Uhealth 2020;8(3):e16665 [FREE Full text] [doi: 10.2196/16665] [Medline: 32130184]
60. Qin W, Blanchette JE, Yoon M. Self-efficacy and diabetes self-management in middle-aged and older adults in the United States: a systematic review. Diabetes Spectr 2020;33(4):315-323 [FREE Full text] [doi: 10.2337/ds19-0051] [Medline: 33223769]
61. Vluggen S, Hoving C, Schaper NC, De Vries H. Psychological predictors of adherence to oral hypoglycaemic agents: an application of the ProMAS questionnaire. Psychol Health 2020;35(4):387-404 [FREE Full text] [doi: 10.1080/08870446.2019.1672873] [Medline: 31588778]
62. Trief PM, Teresi JA, Eimicke JP, Shea S, Weinstock RS. Improvement in diabetes self-efficacy and glycaemic control using telemedicine in a sample of older, ethnically diverse individuals who have diabetes: the IDEATel project. Age Ageing 2009;38(2):219-225. [doi: 10.1093/ageing/afn299] [Medline: 19171951]
63. Grant RW, Wald JS, Schnipper JL, Gandhi TK, Poon EG, Orav EJ, et al. Practice-linked online personal health records for type 2 diabetes mellitus: a randomized controlled trial. Arch Intern Med 2008;168(16):1776-1782 [FREE Full text] [doi: 10.1001/archinte.168.16.1776] [Medline: 18779465]
64. McCarrier KP, Ralston JD, Hirsch IB, Lewis G, Martin DP, Zimmerman FJ, et al. Web-based collaborative care for type 1 diabetes: a pilot randomized trial. Diabetes Technol Ther 2009;11(4):211-217 [FREE Full text] [doi: 10.1089/dia.2008.0063] [Medline: 19344195]
65. van Vugt M, de Wit M, Sieverink F, Roelofsen Y, Hendriks SH, Bilo HJG, et al. Uptake and effects of the e-vita personal health record with self-management support and coaching, for type 2 diabetes patients treated in primary care. J Diabetes Res 2016;2016:5027356 [FREE Full text] [doi: 10.1155/2016/5027356] [Medline: 26955640]
66. Graetz I, Huang J, Muelly ER, Fireman B, Hsu J, Reed ME. Association of mobile patient portal access with diabetes medication adherence and glycemic levels among adults with diabetes. JAMA Netw Open 2020;3(2):e1921429 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.21429] [Medline: 32074289]
67. Lau M, Campbell H, Tang T, Thompson DJS, Elliott T. Impact of patient use of an online patient portal on diabetes outcomes. Can J Diabetes 2014;38(1):17-21. [doi: 10.1016/j.jcjd.2013.10.005] [Medline: 24485208]

XSL•FO
RenderX

68. Reed M, Huang J, Graetz I, Brand R, Hsu J, Fireman B, et al. Outpatient electronic health records and the clinical care and outcomes of patients with diabetes mellitus. Ann Intern Med 2012;157(7):482-489 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00004] [Medline: 23027319]
69. Zocchi M, Robinson S, Ash A, Vimalananda VG, Wolfe HL, Hogan TP, et al. Patient portal engagement and diabetes management among new portal users in the veterans health administration. J Am Med Inform Assoc 2021;28(10):2176-2183 [FREE Full text] [doi: 10.1093/jamia/ocab115] [Medline: 34339500]
70. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and facilitators to online portal use among patients and caregivers in a safety net health care system: a qualitative study. J Med Internet Res 2015;17(12):e275 [FREE Full text] [doi: 10.2196/jmir.4847] [Medline: 26681155]
71. Richwine C. Progress and persistent disparities in patient access to electronic health information. JAMA Health Forum 2023;4(11):e233883 [FREE Full text] [doi: 10.1001/jamahealthforum.2023.3883] [Medline: 37948063]
72. Lyles CR, Nelson EC, Frampton S, Dykes PC, Cemballi AG, Sarkar U. Using electronic health record portals to improve patient engagement: research priorities and best practices. Ann Intern Med 2020;172(11 Suppl):S123-S129 [FREE Full text] [doi: 10.7326/M19-0876] [Medline: 32479176]
73. Johnson AM, Brimhall AS, Johnson ET, Hodgson J, Didericksen K, Pye J, et al. A systematic review of the effectiveness of patient education through patient portals. JAMIA Open 2023;6(1):ooac085 [FREE Full text] [doi: 10.1093/jamiaopen/ooac085] [Medline: 36686972]
74. Casillas A, Abhat A, Mahajan A, Moreno G, Brown AF, Simmons S, et al. Portals of change: how patient portals will ultimately work for safety net populations. J Med Internet Res 2020;22(10):e16835 [FREE Full text] [doi: 10.2196/16835] [Medline: 33094732]

Abbreviations

CHC: community health center
CHW: community health worker
CONSORT: Consolidated Standards of Reporting Trials
DSM: diabetes self-management
DSMS: diabetes self-management education and support
GLMM: generalized linear mixed model
HbA_{1c}: hemoglobin A_{1c}
MAP: multilevel intervention to increase access and use of portals
PAID: Problem Areas in Diabetes scale
REDCap: Research Electronic Data Capture
SDoH: social determinants of health
T2D: type 2 diabetes
WHO: World Health Organization

Edited by A Mavragani; submitted 07.10.24; peer-reviewed by Y Lu, J Njeru, AN Ali; comments to author 14.02.25; revised version received 20.02.25; accepted 21.02.25; published 25.03.25.

Please cite as:
Whittemore R, Jeon S, Akyirem S, Chen HNC, Lipson J, Minchala M, Wagner J
Multilevel Intervention to Increase Patient Portal Use in Adults With Type 2 Diabetes Who Access Health Care at Community Health Centers: Single Arm, Pre-Post Pilot Study
JMIR Form Res 2025;9:e67293
URL: https://formative.jmir.org/2025/1/e67293
doi:10.2196/67293
PMID:40131327

©Robin Whittemore, Sangchoon Jeon, Samuel Akyirem, Helen N C Chen, Joanna Lipson, Maritza Minchala, Julie Wagner. Originally published in JMIR Formative Research (https://formative.jmir.org), 25.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Patients' and Health Care Professionals' Expectations of Virtual Therapeutic Agents in Outpatient Aftercare: Qualitative Survey Study

Diana Immel[1], DipFH; Bernhard Hilpert[2,3,4], MSc; Patricia Schwarz[5], MSc; Andreas Hein[5], PhD; Patrick Gebhard[2], PhD; Simon Barton[1*], PhD; René Hurlemann[1*], MSc, MD, PhD

[1]Department of Psychiatry, School of Medicine and Health Sciences, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

[2]Affective Computing Group, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, German Research Center for Artificial Intelligence, Kaiserslautern, Germany

[3]Leiden Institute of Advanced Computer Science, Leiden University, Snellius Gebouw, Leiden, The Netherlands

[4]Department of Computer Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[5]Assistance Systems and Medical Device Technology, Department for Health Services Research, School of Medicine & Health Sciences, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

[*]these authors contributed equally

Corresponding Author:
Diana Immel, DipFH
Department of Psychiatry
School of Medicine and Health Sciences
Carl von Ossietzky Universität Oldenburg
Hermann-Ehlers-Str. 7
Oldenburg, 26160
Germany
Phone: 49 441961517
Fax: 49 44196151599
Email: diana.immel@uni-oldenburg.de

Abstract

Background: Depression is a serious mental health condition that can have a profound impact on the individual experiencing the disorder and those providing care. While psychotherapy and medication can be effective, there are gaps in current approaches, particularly in outpatient care. This phase is often associated with a high risk of relapse and readmission, and patients often report a lack of support. Socially interactive agents represent an innovative approach to the provision of assistance. Often powered by artificial intelligence, these virtual agents can interact socially and elicit humanlike emotions. In health care, they are used as virtual therapeutic assistants to fill gaps in outpatient aftercare.

Objective: We aimed to explore the expectations of patients with depression and health care professionals by conducting a qualitative survey. Our analysis focused on research questions related to the appearance and role of the assistant, the assistant-patient interaction (time of interaction, skills and abilities of the assistant, and modes of interaction) and the therapist-assistant interaction.

Methods: A 2-part qualitative study was conducted to explore the perspectives of the 2 groups (patients and care providers). In the first step, care providers (n=30) were recruited during a regional offline meeting. After a short presentation, they were given a link and were asked to complete a semistructured web-based questionnaire. Next, patients (n=20) were recruited from a clinic and were interviewed in a semistructured face-to-face interview.

Results: The survey findings suggested that the assistant should be a multimodal communicator (voice, facial expressions, and gestures) and counteract negative self-evaluation. Most participants preferred a female assistant or wanted the option to choose the gender. In total, 24 (80%) health care professionals wanted a selectable option, while patients exhibited a marked preference for a female or diverse assistant. Regrading patient-assistant interaction, the assistant was seen as a proactive recipient of information, and the patient as a passive one. Gaps in aftercare could be filled by the unlimited availability of the assistant. However, patients should retain their autonomy to avoid dependency. The monitoring of health status was viewed positively by both groups. A biofeedback function was desired to detect early warning signs of disease. When appropriate to the situation, a sense of humor in the assistant was desirable. The desired skills of the assistant can be summarized as providing structure and

XSL•FO
RenderX

emotional support, especially warmth and competence to build trust. Consistency was important for the caregiver to appear authentic. Regarding the assistant–care provider interaction, 3 key areas were identified: objective patient status measurement, emergency suicide prevention, and an information tool and decision support system for health care professionals.

Conclusions: Overall, the survey conducted provides innovative guidelines for the development of virtual therapeutic assistants to fill the gaps in patient aftercare.

(JMIR Form Res 2025;9:e59527) doi:10.2196/59527

KEYWORDS

socially interactive agent; e-mental health; mental illness; mental disorder; depression; major depressive disorder; suicide prevention; suicidal ideation; outpatient aftercare; artificial intelligence; virtual therapeutic assistant; public health; digital technology; digital intervention; digital health care

Introduction

Background

As stated by the World Health Organization, depressive disorders are among the most serious mental illnesses. They represent a serious loss of societal potential through premature mortality. It is estimated that 4.4% of the global population experiences depressive disorders. This is an increase of 18% compared to more than a decade ago [1]. The lifetime prevalence of depressive disorders is estimated to be between 16% and 20% [2]. Depressive disorders represent the most common mental health condition associated with suicide. It is estimated that 60% to 70% of individuals experience suicidal ideation during depressive episodes. The risk of suicide is 30 times higher than that observed in the general population. The mortality rate within the first year is approximately 3.4 times that of the general population. Furthermore, the probability of a depressive episode recurring within the first year is approximately 70%. In addition, approximately 21% of patients are readmitted to the hospital within 1 year of discharge [3]. The initial 6-month period following discharge is of particular significance with regards to the potential for relapse and suicide [4,5]. In contrast with the recommendations set forth in the German S3 guideline for unipolar depression, it is observed that 90% of patients diagnosed with major depression do not receive sufficient outpatient aftercare following psychiatric hospitalization [2]. This elucidates the elevated risk of relapse and suicide during the initial year [3]. Patients with restricted access to social, economic, or structural resources are particularly susceptible to these risks [2]. This conspicuous deficiency in outpatient care results in the forfeiture of potential therapeutic benefits that could otherwise be achieved through inpatient care. Conversely, enhancing outpatient care for patients with depression has been demonstrated to result in improved survival rates. In the context of this study, the term health care professionals refers to psychiatrists, psychotherapists, nurses and social workers who treat patients after they have been discharged from the hospital. The responsibility for patients with suicidal tendencies represents the aspect of working with depressed individuals that evokes the strongest emotional responses from this group of health care professionals and those who have previously cared for such patients. The experience of losing of a patient to suicide can have a substantial and lasting negative emotional impact. The risk of suicide among practitioners increases substantially following the suicide of a patient. In one study, 10% of participants, who were health care professionals, reported having suicidal thoughts themselves [6,7]. Thus, solutions that enhance the security of particularly susceptible patients, such as effective posthospital psychiatric follow-up, can consequently enhance the lives of patients and their families, as well as health care professionals, by minimizing occupational stress.

Before discharge, health care professionals need to assess the patients' resilience in this regard and share responsibility. The decision-making process is often influenced by other factors, such as insurance coverage. Similarly, the factors that may lead to individual crisis situations are difficult to predict; therefore, postdischarge risk is also a professional burden for caregivers [8,9]. Although some patients continue to receive outpatient care after discharge, the sudden change in their situation poses several problems. One of the most prevalent problems seems to be the lack of a daily structure in time and tasks throughout the day, as provided by the hospital during inpatient care. Instead, patients are mostly responsible for their own daily structure, a task that many struggle with as their goal setting and persistence are not necessarily stable enough to overcome these hurdles. Conversely, in moments of self-doubt and struggle, they now have limited access to timely support. In the hospital, a safe environment with a wide range of support is available, especially in critical situations. However, this changes after discharge [10]. Outpatient psychiatric aftercare, when available, is limited and does not cover all critical situations that prevail in suicidal moments [11]. The limited number of health care professionals, together with time and emotional resource constraints and unclear responsibilities, pose a serious problem for patients. In addition, another source of support regarding social companionship, such as friends or family [12], may not always be available. More importantly, relatives are not professionally trained to deal with critical and potentially suicidal behavior. Some patients may even feel ashamed. In general, patients are left alone after discharge.

It appears that a gap exists in the system for outpatient care in critical situations, particularly with regards to the availability of health care professionals, the definition of responsibilities, and the lack of qualifications in many cases [2]. In the absence of in-person assistance, face-to-face support has declined and an increased demand for online support has been observed. This indicates a necessity for the development of secure, safe, and effective digital tools. Nevertheless, research indicates that internet-based cognitive behavioral therapies have a relatively modest impact on patients with suicidal tendencies [13]. Patients

XSL•FO
RenderX

who continue to experience suicidal thoughts and behaviors also present a unique challenge for electronic mental health (e-mental health) services. Therefore, this group is explicitly excluded from e-mental health services [14]. A variety of interventions are available, including web-based psychoeducation, online counseling, SMS text messaging aftercare, email therapy, and computer-assisted cognitive behavioral therapy [15,16]. However, these interventions fail to address patients' need for attachment, or for the formation of a social, empathic relationship. This has been described as the most important variable in determining the effectiveness of psychotherapy [17,18]. Consequently, the lack of a therapeutic relationship is a substantial criticism expressed by patients when evaluating internet-based cognitive behavioral therapy self-help methods [19].

It is thus evident that an intermediate professional solution is required to address the gap in care for patients with depression and suicidality, with the objective of providing support to both this susceptible patient population and their health care professionals. The solution should be readily accessible without overwhelming amounts of information, focus on depression-related cognitive and emotional impairments, and include suicide prevention. It should also be noted that using such solution, people in underserved rural or economically less developed regions could be better reached, overcoming socio-spatial barriers. Furthermore, the stress on the therapists in charge could be greatly reduced, thus contributing to staff stability in psychiatric institutions.

Using Socially Interactive Agents for Outpatient Aftercare

Socially interactive agents (SIAs) can engage in verbal and nonverbal interactions as dialogue partners, exhibiting emotional and social capabilities and performing a diverse social support and monitoring functions [20-23], including provision of emotional support to users [24], monitoring of patient safety [25] and simulation of social skills to emulate social relationships, such as active listening, mimicry, gestures and emotion models [26-29]. Given the crucial role of social relationships in coping with psychological distress [30], there is the potential for SIAs to enhance technology-assisted outpatient care by emulating human interactions and tasks [31].

To date, SIAs have been used in a variety of settings as virtual coaches and assistants. Unified platforms for SIAs as coaches have been developed [32-34] and used for stress management purposes [35]. However, they have also been used in applied contexts as training partners for social skills [36] and empathy training [37]. In clinical research, initial approaches to the use of empathic conversational agents in the context of mental health have been proposed [38]. Furthermore, the use of relational agents has been demonstrated to have a positive impact on hospitalized patients with depressive disorders [39]. These findings suggest that these approaches may offer promising possibilities for the use of a virtual therapeutic assistant (VTA) as an assistant in a supportive, gap filling role in outpatient aftercare. It is unclear which settings a VTA could be used in and what role it should play. There is also a lack of clarity regarding the design of the interface for interaction with patients or health care professionals.

In SIAs, embodiment is of substantial consequence with regard to the manner in which they are perceived. While technology-mediated solutions can have a supportive effect, as previously outlined, communicating with a disembodied, faceless machine can impede positive outcomes, even when a human is on the other end. Conversely, embodiment in SIAs has been demonstrated to influence perceptions [40-42], increase trust [43,44], facilitate the formation of an emotional connection, and enhance overall relationship quality [45].

With regards to interpersonal interaction, a substantial body of research has examined timing, skills, and behavior of SIAs. It has been demonstrated that SIAs can evoke emotional responses that are pertinent to interpersonal interaction and attachment. The timing of SIAs behavior has been shown to be of critical importance in the formation of human-agent relationships [46,47]. In a job interview setting, SIAs were found to elicit feelings of shame to the same extent as human interviewers [48]. Moreover, analogous effects have been demonstrated with respect to trust [49] and rapport through backchanneling behavior [22,23,50], the elicitation of empathetic behavior in human physicians when dealing with virtual patients [51], and the elicitation of prosocial behavior in interpersonal interaction settings [52-54].

An embodied, affective SIA has the potential to facilitate novel avenues for relationship-based, technology-mediated outpatient aftercare. Nevertheless, its application to date has been confined to the context of scientific studies or hospitalization. The context of outpatient aftercare, which is particularly sensitive due to its applied nature, requires a meticulous and precise implementation of the VTA for it to have the intended supportive effect.

Research Questions

Overview

This study presents a comprehensive analysis of the expectations held by both patients and health care professionals regarding the design of a proposed tool. The objective of the survey was to ascertain both expectations regarding the structural framework of how a VTA might interact with patients and regarding the interaction itself.

Role

The initial stage of the process entailed an evaluation of the potential integration of a VTA within the existing framework of outpatient care. It is necessary to establish the role that both groups anticipate a VTA will play in the context of outpatient care.

- Research question (RQ) 1: What is the role of a VTA in the context of outpatient care?

Embodiment

A substantial body of research and practical applications outlined previously indicates that the favorable outcomes of SIA interaction are contingent upon the manner of embodiment within the interaction interface.

XSL•FO
RenderX

- RQ 2: What are the embodiment preferences for a VTA?

Interaction Timing

The objective of this study was to address the expectations of patients and health care professionals regarding the interaction between a VTA and a patient. As no previous testing of this framework has been conducted, it is unclear what frequency and timing of interaction might prove beneficial for the patient.

- RQ 3: When should the VTA interact with the patient?

Interaction Abilities

Although a VTA offers numerous advantages to the patient, the specific interaction capabilities that the VTA should possess to be beneficial for treatment remain uncertain.

- RQ 4: What interaction capabilities should the VTA have to assist patients in their treatment?

Interaction Behavior

The following section will present the observed interaction behaviors and behaviors associated with the use of the aforementioned skills during the interaction.

- RQ 5: What interaction behaviors should the VTA use to facilitate patient treatment?

Interaction Between VTAs and Health Care Professionals

For the VTA to be considered an integral part of the treatment framework, it would be beneficial for it to provide useful information to the health care professionals themselves.

- RQ 6: How might the framework be designed in a way that would allow health care professionals at all levels (ie, psychiatrists, psychotherapists, nurses, and social workers) to benefit from the implementation of a VTA?

Methods

Overview

To investigate these questions, a 2-part qualitative study was conducted. The objective was to explore the perspectives of the 2 groups that were most directly involved (health care professionals and patients) and to assess their expectations of a VTA in the context of treatment. Initially, a group of health care professionals were queried about their expectations regarding the potential role of the assistant (RQ 1), assistant-patient interface (RQ 2), timing of the interaction (RQ 3), interaction skills (RQ 4), interaction behavior (RQ 5), and assistant–health care interaction (RQ 6). Following a comprehensive analysis of the responses, a qualitative interview was conducted with a group of patients undergoing treatment. This will provide a 360-degree perspective of the requirements that a socially interactive VTA should fulfill in a specific clinical setting to address the identified gap in outpatient aftercare. The methods and results are presented in parallel for the sake of comparability.

Participants

In the initial phase of the study, data were gathered from a sample of health care professionals (n=30; 6 male and 24 female participants) employed in various outpatient settings within the field of outpatient care in social psychiatry in Oldenburg (Germany). All health care professionals participated voluntarily, which was ensured by providing them detailed information about the objectives and implementation of the study. No data were excluded. The age of the therapists ranged from 18 to 24 years (n=1), 25 to 34 years (n=13), 35 to 44 years (n=6), 45 to 54 years (n=4), 55 to 64 years (n=5), and ≥65 years (n=1), and their professions included psychiatrist (n=3), psychotherapist (n=10), social worker (n=15), psychiatric nurse (n=1), and physician (n=1).

The patient sample consisted of individuals (n=20; 11 male and 9 female participants) who were inpatients at the Karl-Jaspers Hospital in Oldenburg and were diagnosed with depression. All patients participated on a voluntary basis, which was ensured by providing detailed information about the objectives and implementation of the study. No data were excluded from the analysis. The age of the participants ranged from 19 to 56 years (mean 40.85, SD 12.46 years). Patients were administered the Beck Depression Inventory II (mean 24.3, SD 13.45; Beck et al [55]) to obtain an overview of their current depression status. To assess trust, the Yamagishi and Yamagishi General Trust Scale (mean 3.19, SD 0.72; [56]) was used, which comprises 7 statements, including “Most people are basically honest” with a rating scale from 1 (low) to 5 (high agreement), Cronbach α=0.83.

Ethics Approval

Before data collection, approval was obtained from the Medical Ethics Committee of the University of Oldenburg (2022-005). Participants provided written informed consent to participate in this study. Data were anonymized. The participants received no compensation for this study.

Procedure

At an informative meeting of health care professionals in our clinic, general information about the purpose of the study and details of the survey were delivered. Following the provision of a link via email, they were invited to complete a qualitative web-based questionnaire (implemented in Qualtrics) comprising a combination of multiple-choice and open questions (described in detail in subsequent sections).

The patient cohort was recruited through the clinical routine of their daily inpatient therapeutic examinations at the same psychiatric hospital. A semistructured qualitative interview was conducted in a confidential session. Following the introduction, a trained interviewer proceeded to elucidate the objective and structure of the forthcoming interview. It was emphasized that the responses provided by the participants would not be shared with any of the other patients and therapeutic staff.

Materials

To ensure that all participants had the same level of knowledge about VTAs, a standardized description was first provided, which was adapted from the description set out by Hilpert et al [42,44]. The questionnaire was constructed in accordance with the 6 distinct expectations (corresponding to the 6 RQs) pertaining to a VTA in outpatient aftercare. The questions

XSL•FO
RenderX

themselves were divided into 5 blocks, each addressing a specific area of inquiry: general expectations, specific situations, usefulness, communication and data handling, and demographics. The web-based questionnaire and the interview were comparably structured and comprised the same set of questions. However, the terminology and specific questions were tailored to align with the perspective of each respective group.

Health Care Questionnaire

The questionnaire for health care professionals consisted of several sections. The initial section of the questionnaire addressed general inquiries pertaining to the health care professionals' anticipated interface, capabilities, interaction behavior, and timing (eg, how and when to contact and how and when to support) of the VTA. In addition, it encompassed broader expectations related to the role of the VTA in the therapeutic process. Second, 2 particular scenarios in which patients may find themselves were delineated, accompanied by corresponding inquiries. In the initial scenario, health care professionals were instructed to envision a typical day in the life of an outpatient, beginning to end. Regarding competencies, the inquiries centered on the diagnostic abilities and subject matter that the VTA should possess to obtain a comprehensive understanding of the patient's condition. The question pertaining to interaction timing addressed the optimal timing for the VTA to interact with the patient, while the questions concerning interaction behavior sought to ascertain the types of behaviors that health care professionals deemed beneficial. The second scenario pertains to an emergency situation that may arise during a day after discharge. The inquiry commenced with queries pertaining to the VTA's capacity to procure external assistance for the patient and how the VTA could facilitate this. Subsequently, the participants were asked to indicate the circumstances under which they would interact with the VTA and to describe the helpful types of behavior. Third, the participants were requested to evaluate the utility of a VTA in relation to a series of therapeutic functions, including medication planning and crisis intervention. Fourth, we investigated their perceptions of their own interaction with the VTA, specifically regarding the type of information they would find beneficial if the VTA were to provide it to health care professionals involved in outpatient care. Finally, the participants were asked to provide demographic information.

Patient Interview

The introductory information and briefing remained consistent throughout the process. As the interview with the patient sample was conducted after the analysis of the health care professionals responses, some of the questions were adapted based on the findings from this initial sample. This was done to elicit more detailed and nuanced responses and to avoid potentially vague responses. The overall structure of the questionnaire remained consistent. The questionnaire used in the interview encompassed the elements of RQ 1 to 5, excluding the interaction between the assistant and the therapist (as this did not impact the patients). The questionnaire was divided into 6 blocks: general expectations, assistant-patient-interaction, assistant-health care professional-interaction, feelings of the patient and demographics.

The interview commenced with a series of general inquiries pertaining to patients' expectations of a VTA's competencies, interaction style (including the modalities of contact, forms of support, and the VTA's presence in the patient's daily life), and a few more detailed questions concerning appearance. Subsequently, a description of 2 particular scenarios in which patients might find themselves was provided, accompanied by corresponding questions. Each specific situation was introduced with a subjective assessment of how the patients would feel in that particular situation. The questions were largely similar to those posed to the health care professionals, but the patients were also asked to provide an explanation for their responses. The section on usefulness was identical to that used with the health care professional sample.

In accordance with the recommendations for open research practices [57], all materials can be found on the OSF website [58].

Data Analysis

To extract meaningful insights for each of the RQs in this qualitative dataset, a 4-step process based on grounded theory was used to analyze the responses given by both groups of interest [59]. This was executed following the guidelines described by Hansen and Grimshaw-Aagaard [60]. In the initial phase, the responses from each group of interest were transcribed (for the patient interview data) and organized by question in 2 distinct datasets, one for each group. Second, 3 independent raters classified the responses according to content and relevance to the RQs. Third, the raw data and extracted categories were compared for each group of interest (health care professionals and patients), identifying both similarities and differences between the 2 groups. Finally, this comparative dataset was classified to extract clusters and subclusters of explanatory value for each RQ.

Results

This section presents the findings of the analysis of the health care professional and patient surveys. A joint analysis of both health care professional and patient responses is presented for each RQ.

RQ 1: Role of the VTA

When queried about the role of the VTA, health care professionals primarily referenced the role of a "good friend." It was emphasized that the VTA should be "designed in such a way that you enjoy the interaction, otherwise it's probably not going to be used." A similar argument was put forth with regard to the role of a social worker, which was described as "something between a good friend (feeling of closeness) and a social worker (knowledge about the disease and how to help)." Conversely, the explicit role of a VTA as a "therapist" was reported, with the addition that the VTA should be a figure with "distance to the family" or "not a substitute for family or friends," stating that "these should be present in real life or newly built." The VTA was recommended as an adjunct to existing therapy, with "more of an additional supportive role,"

XSL•FO
RenderX

with an emphasis on supporting the patient "in phasing out therapy." In a less role-specific description, the VTA was referred to as a "companion" or "extended inner support, more like a part of oneself that is currently too weak to provide self-support." Health care professionals also described the VTA as a "coach" or "someone the patient can confide in, but who remains independent." A more abstract idea was a "good spirit that no one else sees, who sees the positive and can sometimes be very direct but warm."

Patients referenced the role of a therapist, "a known person," as well as frequent references to a "good friend," "being there for me like a best friend," "asking questions about everyday life (instead of 'psychological' questions)." The wish for additional support in outpatient care is clear: "to work out goals and important topics with therapists, which the assistant knows about."

In general, both groups anticipated a system that integrates expertise and represents a figure with whom they can establish a relationship. In conjunction with the primary recommendation of the therapist's good friend, the role of a peer counselor would also be conceivable.

RQ 2: Embodiment of the Assistant-Patient Contact

Regarding the topic of embodiment, we inquired about the participants' expectations regarding general appearance characteristics of VTA. As previous research [40] has shown a clear pattern of human embodiment, we also asked about preferences for the assistant's age and gender. The question regarding general appearance was posed in an open-ended format, with the objective of obtaining an unbiased report from the participants. Most of the patients (14/20, 70%) related their responses to the human appearance of the assistant. Some described nonhuman appearance and characteristics, to which age and gender do not apply. Regarding general appearance, the responses of both health care professionals and patients who endorsed a humanlike appearance were classified and summarized into 2 relevant clusters that play a role in the perception of the assistant: *physical appearance* and *expressiveness*.

In *physical appearance*, the health care professionals' interview focused on 3 features: personal characteristics, clothing, and animation style. For physical characteristics, they tended to be the least risky, such as "medium height," "normal weight," "friendly face," and "calm speech." The same pattern was observed for clothing, as shown by responses such as "[the assistant] wears clothes in pleasant pastel shades, nothing too dark or too colorful." The third feature could be defined as animation style, with health care professionals agreeing that the assistant should appear "as realistic as possible" with "no cartoonish features." Regarding gender, 80% (24/30) of the health care professionals wanted a selectable option. The remaining participants recommended either a female assistant (3/30, 10%) or a diverse (3/30, 10%) assistant, while male gender was not explicitly recommended. Regarding age, a "middle-aged" assistant was recommended, as it should "exude life experience," "not too young for older people." Other recommendations were "age depending on the patient, maybe a bit older" or that the age should be "chosen by the patient."

For the *expressiveness* cluster, they focused on facial expressions and gestures that should be convincing enough for the assistant to be able to "respond to feelings and show empathy" as well as "natural movements."

A different set of priorities for the *physical appearance* of a VTA was observed by the patients. While they wanted the assistant to have "an upper body and hands, but no legs," it should "not have a perfect figure," with the same mentioned for the face. Clothing was not mentioned. For animation style, they reported a tendency toward a more animated style "like in a computer game" or "like in science fiction movies." However, in contrast to health care professionals, patients expressed a desire for familiarity, as the assistant should look like a known person, for example, the caregiver. Regarding gender, there was a clear and predominant preference of all participants for a female or diverse character. The reasons provided for this preference were "I am a woman myself," "I can talk better to women," "Navi, Siri are also women" and "it is easier to build trust." In cases where patients did not explicitly prefer a female character, responses such as "voice is more important (soft, calm voice)" were given. It should be emphasized that none of the patients explicitly asked for a male embodiment for their VTA. Regarding the age of the VTA, "middle-aged," "older," but also contradictory answers, such as "between teenager and young adult" were provided. However, there was a tendency to favor a middle-aged figure to older figure. For *expressiveness*, it was desired that the assistant should be able to use facial expressions and gestures "so that emotions can also be clearly expressed" (clear facial expressions). Overall, they wanted "friendly" and "laughing" expressions.

Finally, there were some health care professionals who preferred a nonhuman appearance, citing concerns about "negative transference" or being "triggered" by a human figure. One participant illustrated:

> *I would like it if the character didn't necessarily try to look like a human or even a therapist. I would be afraid that the difference would be too great and the character would look unbelievable or even ridiculous. I would therefore suggest that anthropomorphized, sympathetic avatars be created that patients could easily cast in a positive light.* [Psychotherapist #35-44, male]

Suggestions for this alternative, nonhuman appearance included a talking smiley instead of a real human face, or a "wise but humorous owl." A few patients had different nonhuman suggestions, such as "blue water on the beach," "self-selectable (robot, human, animal...)," "between emoji and human" and "can give itself a shape."

These findings indicate that both groups anticipated an imperfect or neutral-looking but emotionally expressive human embodiment for their VTA, with a tendency to avoid male characters.

RQ 3: Interaction Timing

The third RQ pertained to the timing of the interaction: when should the assistant interact with the patient? The joint analysis of the responses from both groups revealed 2 main clusters of

XSL•FO
RenderX

situations in which the VTA could potentially interact with patients.

The first cluster was designated *situation-independent or proactive* and refers to behavior that does not depend on the socioemotional context of the patient at the time of interaction. Situation-independent or proactive interaction timing does not necessitate a request for assistance from the patient or the automatic recognition of an ongoing crisis situation by the assistant. This encompasses any form of interaction that could be proactively initiated by the assistant, including both scheduled contact and unscheduled check-in messages. For example, health care professionals proposed that the assistant could contact patients in the morning to "help start the day" and in the evening to "reflect on the day." In between, at "agreed times" and "at random times." Patients corroborated this assertion by requesting regular contact "maybe 1-2 times a day," "daily contact," "regularly in the morning," "check in 3-4 times a day."

The second cluster was designated *situation-dependent* and represents a complementary counterpart to the first. It includes all instances wherein the interaction is triggered by a response to the patient's current therapeutic or socioemotional circumstances. This could manifest as the patient requesting contact, for example, to talk about a feeling at any time of the day, or even in crisis situations, such as a panic attack. It could also be triggered by monitoring the patient's current stress indicators (eg, heart rate and lack of interaction) to detect an emergency situation and trigger a *situation-dependent* interaction in the assistant. Responses that were clustered as *situation-dependent* differed between health care professionals in the frequency with which this should occur. For example, patients mentioned "more often immediately after hospitalization" and that the assistant should contact them "on holidays/at weekends." At this point, patients would like to be able to personalize the assistant's presence according to their own needs. Patients would also like to be notified "as soon as he notices that I am not feeling well" or "if I have been inactive for a long time."

Overall, health care professionals' responses to the "when" question exhibited considerable variability. Regular or daily contact, fixed times, and contact when the patient is inactive were all represented to a similar extent. Although patients emphasized "unstructured" times when they expected to feel alone, there were no substantial differences between the responses of patients and health care professionals. These numerous variabilities may indicate the importance of an individual rhythm of contact.

RQ 4: Interaction Abilities

This RQ was concerned with interaction skills: what competencies should the VTA possess to support patients in their treatment? The analysis revealed 2 main clusters of interaction skills that a VTA should possess in general: the capacity to furnish structural support and the capacity to furnish emotional support. Given the considerable overlap between the clusters extracted from the joint analysis of the responses of both groups of interest, it was determined that there was no additional benefit in reporting the responses of health care professionals and patients separately in this subsection.

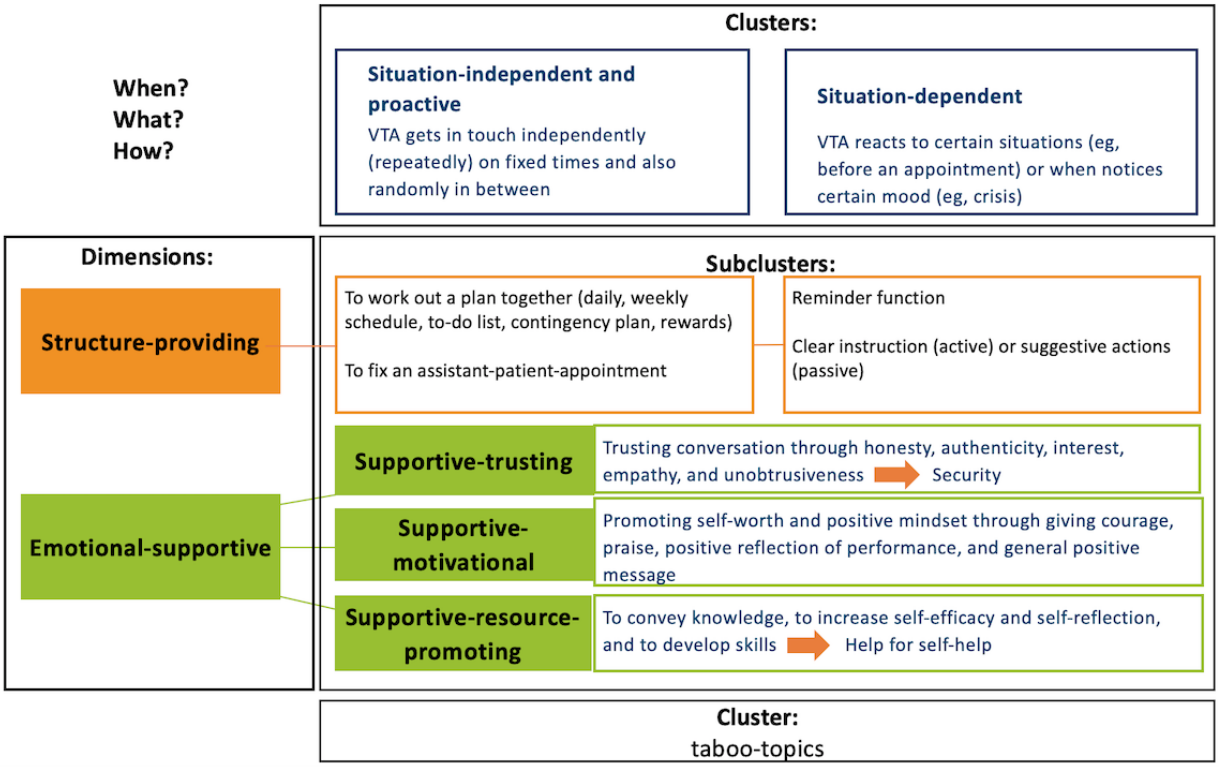
The first cluster was defined as *structure-providing* capabilities, encompassing all the interaction capabilities of the VTA that facilitate structure for the patient regarding behavior (proposing action strategies or step-by-step plans for upcoming situations, monitoring the patient's emotional state, and providing feedback) and time management (eg, structuring the day and reminder functions) to enable coping with daily challenges or tasks.

The second cluster was described as *emotional-supportive*, comprising skills that assist patients in managing their emotions in a relational context. This cluster comprised 3 subclusters: *supportive-trusting*,*supportive-motivational*, and *supportive-resource-promoting.* The capacity to be *supportive-trusting* enables the assistant to facilitate trusting conversations by being honest, authentic, interested, sensitive, nonintrusive, and reassuring. The capacity to be *supportive-motivational* enables the assistant to promote a positive attitude and increase self-esteem by acknowledging courage, praising achievements, and generally providing positive messages. The capacity to be supportive-resource-promoting enables the VTA to expand knowledge, increase self-efficacy and self-reflection, and develop and apply skills. Rather than giving instructions, this cluster emphasizes the promotion of patient autonomy through the development of self-management skills.

The results of RQs 3 and 4 provide a 2-dimensional conceptual space for interpreting participants' responses. The first dimension characterizes the situational aspect of interaction timing and encompasses the clusters of *situation-independent or proactive* and *situation-dependent* expectations regarding the VTA. The second dimension pertains to the function of the interaction and encompasses the 2 clusters of expectations: one pertaining to *structure-providing* skills and the other to *emotional-supportive* skills. The *emotional-supportive* cluster may be further subdivided into 3 facets: *supportive-trusting*, *supportive-motivational*, and *supportive-resource-promoting*. This resulted in a total of 8 response subclusters, which can be used to categorize the interaction and the interactional behavior displayed. In addition, a further category of responses was identified as *taboo topics* representing a set of issues that should not be discussed. Figure 1 provides an overview of the dimensions and clusters with their subclusters, including individual definitions.

XSL•FO
RenderX

Figure 1. An overview of a 2D concept with clusters and subclusters in the interaction abilities of a virtual therapeutic assistant (VTA) in outpatient care. The definitions used to interpret the responses of participants (N=50, including health care professionals and patients with depression) in a semistructured qualitative study.



RQ 5: Interaction Behavior

Overview

This RQ was concerned with interaction behaviors that the VTA should display to provide effective support to patients undergoing treatment. The responses are presented separately for each of the 8 subclusters of the conceptual space as described in the RQ 4: Interaction Abilities section. However, the questionnaire included a specific section of questions pertaining to an emergency situation or breakdown, the responses to which differed considerably from those provided in the general sections. Therefore, crisis behavior is reported in a separate section below.

Regarding *situation-independent* or *proactive structure-providing* behavior, health care professionals responded that the assistant could collaborate with the patient to develop a plan for the structure of the day, that is, "suggestions for organizing the day," "checking off to-do lists" as well as "validating mood, suggesting daily activities." Patients indicated a preference for regular contact, including "daily contact" and "welcoming words." In addition, patients expressed a desire for a mood or habit tracker as well as daily advice.

For *situation-dependent structure-providing* behavior (ie, in situations of distress or instability), health care professionals recommended reminders, concrete instructions, or suggestions for action. These included "behavioral training," "reminder of positive activities," and "taking medication." In addition, patients requested task reminders, eg, "appointments," instructions on specific markers (eg, "sleep quality") and "psychoeducation."

Overall, health care professionals overwhelmingly stated that the assistant should remind patients of the strategies they have learned and provide them with positive support. Patients indicated a preference for the assistant to provide guidance and support through the provision of helpful suggestions and emotional support. There were no major differences between the responses of the 2 groups. However, health care professionals repeatedly emphasized the importance of providing positive patient-emancipating support, whereas patients increasingly indicated that they would prefer the assistant to inquire about their condition.

Regarding *situation-independent* or *proactive supportive-trusting* behavior, the health care professionals indicated that authenticity was a crucial element: "no good coaxing," "moderate criticism," and "avoid empty phrases." Moreover, the VTA should monitor the patient's emotional state and inquire about their well-being, while avoiding excessive intrusion ("don't ask the same probing questions every day" and "don't lecture, don't put pressure"). The patients requested a personal connection and a relational memory to "refer to past episodes." In addition, patients requested displays of empathy and humor, as well as affirmations of positivity and emotional support "as empathetic as possible" or "help to keep my head above water."

In regard to *situation-dependentsupportive-trusting* behavior, health care professionals emphasized the importance of honesty. A "perceived deterioration" in mood, "early warning signs" or "exceeding critical levels" should be communicated to the patient first. Moreover, the VTA should demonstrate interest by "listening when the patient tells you something on his/her

XSL•FO
RenderX

own" and "keep the conversation going." In addition, the assistant may proactively "engage the patient in conversation." After specific incidents, the VTA may inquire about the patient's experience by asking, "How did it go?" The assistant should demonstrate empathy and understanding, taking "negative or sad feelings seriously" and refraining from trivializing them. Unsolicited "funny" advice should be avoided. The patient should be queried as to whether they would prefer to discuss a "particular topic" in more detail or "at another time." The patient should be "relieved" and "reassured" as much as possible. Patients in this cluster expressed a desire for honesty, interest, "being there for me like a best friend," "empathy/understanding" and a sense of security "trusting that he will respond when I contact him."

Overall, the responses were not markedly divergent, with both groups viewing the support worker as a primary source of assistance. Health care professionals had a strong focus on "avoiding harm" in anything the assistant would do, while patients' responses reflected a fear of being left alone.

For the subsequent cluster, "situation-independent or proactive supportive-motivational behavior, health care professionals advised the presentation of positive or "thought-provoking" messages, "positive images," and "music." In addition, it was recommended that the daily routine be reflected upon by asking positive questions, such as "What are five things that have been good in the last week?" after which the VTA could summarize the successes to the patient. On the patient side, this cluster included promoting a positive attitude, "being supportive and encouraging," self-esteem-boosting content, such as "responding to the patient's stress and needs," "not being annoying or condescending," and general positive messages.

In the *situation-dependent* part of the *supportive-motivational* cluster, health care professionals mentioned "encouraging," "promoting positive thoughts" and "praising" in relation to upcoming activities or occasions. In addition, patients mentioned "encouraging, praising" and "reminding of learned structures" to reflect positively on their performance.

Overall, the responses converged on the request for positive supportive behaviors, with the difference that health care professionals asked for proactive and patients for reactive behaviors.

In the *supportive-resource-promoting* cluster, health care professionals most frequently mentioned "psychoeducation" as a situation-independent behavior, with the addition that "reference to online forums should be avoided." In addition, they suggested that calming techniques, such as "breathing exercises, mindfulness exercises" could be included. At this juncture, the transition to situation-dependent recommendations is fluid, as *supportive-resource-promoting* behaviors frequently prove effective for both types of timing. The health care professionals suggested that the assistant should help to "focus on the positive," "address brooding thoughts," and "motivate, show understanding." As an example, for panic attacks, breathing exercises could be used to help. For behaviors that are exclusively mentioned as situation-dependent, health care professionals mentioned providing explanations, for example, "in bad phases, the changeability of feelings and mood phases" could be explained with reference to individual examples from the patient's past. As this type of self-help promotion was considered particularly important by the health care professionals, an exemplary list is provided in Textbox 1.

Textbox 1. Examples of how a virtual therapeutic assistants can be used to promote self-help for patients with depression in the outpatient setting, which is considered particularly important by health care professionals (n=30) in a semistructured qualitative study.

Examples for self-help promotion:

- "Where could your bad mood come from?"
- "What has helped with low mood so far?"
- "What self-stabilization options do you have?"
- "Who could you call?"

Situation-independent reminders should also include the following:

- Knowing about or instructing, if not already available, the patient's skills, strategies, and stability factors
- Methods for dealing with circles of thought and rumination
- Defining emergency contacts together (making them quickly accessible)
- Informing about outpatient help offers (self-help groups, outpatient treatment, telephone counseling, specialists, and social psychiatric service)
- Building up activities (eg, maintaining social contacts) and providing inspiration for activities
- Supporting eating behavior (only if desired) and sleep hygiene
- Homework support

Patients in the *supportive-resource-promoting* cluster requested specific forms of assistance and guidance, including "offering autogenic training," psychoeducation, and dialogues—enabling conversations between the assistant and the patient as a situation-independent behavior. However, monologues and input from the assistant, such as "nice topics" and "providing facts and clear lines" were also requested. In the situational dimension, patients expressed a desire for biofeedback, skills and competencies—showing ways to solve problems," relational

XSL•FO
RenderX

memory—"referring to past episodes," and active dialogue—"talking to me about what is bothering me."

Overall, regarding *emotional-supportive* dimension and its 3 subclusters, practitioners indicated that they would use specific strategies and interactional behaviors to facilitate conversations through honesty, authenticity, interest, empathy, and unobtrusiveness to build trust and convey safety. To motivate, positive self-esteem and minimum standards should be achieved through encouragement, praise, positive reflection on achievements, and general positive messages. The objective is to provide knowledge, enhance self-efficacy and self-reflection, and cultivate skills. Patients tended to mention more functional characteristics at this point, whereas practitioners were more likely to cite the assistant's capacity for empathic understanding.

A further category of responses was identified, namely a cluster of *tabootopics* that should not be discussed "because the questions might trigger and the patient would be left alone with negative feelings." These included suicidal thoughts, weight gain, and possible crises that could not be detected, "risk of not being able to detect this in an individual way, or inadequate response to suicidality." Nevertheless, some health care professionals advocated for the inclusion of all topics and the elimination of taboos, stating "I think there should be no taboos. Social interaction should play a big role, sexuality can also be mentioned." Additional responses on this page indicated that: "In general, the character should address many topics so that the patient thinks about his condition" and has to "show his colors."

A similar list of topics was identified in the patient responses. Topics that are deemed irrelevant or potentially "thought triggering" should be excluded from discussion. Furthermore, certain intimate topics should be avoided, as respondents indicated: "For many topics I would need a person."

In general, when queried about the appropriate conduct of the assistant, both respondents concurred that the assistant should refrain from exerting undue pressure or authority and should avoid engaging in repetitive or critical interactions. There is a distinction in the manner of presentation, with patients describing their preference for an assistant who is not "annoying," "pointing out mistakes," "using empty phrases," and "scolding." Health care professionals have indicated that the assistant should not be "pushy," "lecturing," "trivializing," and "admonishing."

Crisis Behavior

When queried as to how the assistant could facilitate a secure treatment environment during a crisis, health care professionals posited that the assistant could facilitate the establishment of contacts (personal and professional), "motivate [the] patient to turn to friends/family/therapists," continue communication by "maintaining dialogue with the person." Additional strategies included providing reassurance to the patient "Keep calm! Remind them of what has been agreed in advance," assuring them of support, "reassuring them" by "announcing supportive next steps and being as transparent as possible" and providing positive support: "Be transparent, do not deny that it is only digital support, only do what has been discussed with the patient beforehand and what they have agreed to."

A noteworthy response tendency emerged from the analysis of the practitioner questionnaire: "motivate the patient to seek help." In response to this, we explicitly asked the patients how the assistant could most likely motivate them to seek help. The respondents indicated that the assistant should refrain from calling for help prematurely and should instead afford the opportunity to ask for help first. However, half of the respondents indicated that the assistant should directly request assistance, indicating a preference for contacting the therapists rather than friends or family members. In the context of a crisis situation, patients also expressed a desire for the assistant to maintain objectivity, to be "factual/informative about the situation (what benefit of seeking help)" and to "communicate mood clearly (visually)." Patients also wanted the assistant to be authoritative: "Just say: It's time to get help" or "I'm convinced that help is appropriate now (as an announcement)." However, they also wanted to be motivated by the assistant's reassuring behavior "by using a reassuring explanatory address (not 'You have to do it now,' but 'Maybe you should...')." In addition, the concept of a "panic button" was proposed, along with the display of assistance contacts, the presentation of support alternatives, and the provision of psychoeducation.

Overall, a wide range of behavioral options for the crisis situation was desired, especially by patients. In contrast to the patients, authoritarian behavior was not desired by the health care professionals. However, there was a tendency toward a preference for the system to be able to call for help without the patient having to give consent in a crisis, showing affinity toward decision-making authority. This may reflect the need for patients in crisis to be able to relinquish responsibility or to find guidance in the often-difficult decision-making process about whether and when a call for help is appropriate and necessary.

RQ 6: Interaction Between VTA and Health Care Professionals

This RQ aimed to investigate the relationship between the VTA and the health care professionals themselves. What modifications could be made to the framework to ensure that health care professionals at all levels could benefit from the implementation of a VTA? When queried on this matter, the health care professionals indicated that this should be done at the patient's request, if the patient is able to provide consent. It is recommended that a preestablished "limit" be set regarding when the practitioner should be contacted. Moreover, the assistant should "obtain the patient's permission beforehand." In such instances, the assistant is permitted to "report suicidal thoughts, progressive deterioration of mood, or documented deterioration of mood over a period of 1-2 weeks," "if a constructive discussion is not possible." A further cluster of considerations pertains to instances of potential danger to self and others, "if the virtual character cannot assess the further danger with certainty," "with threat or lack of dissociation of self—or harm to others," or with "report of suicidal thoughts (even once)." In particular, in the absence of clear dissociation of suicidal thoughts, it was proposed that the assistant could

XSL•FO
RenderX

potentially dial the emergency number in the event of inactivity and a lack of response from the patient for a predefined period, in the case of an unannounced exit from the program, or in the event of inactivity for a period exceeding 48 hours. The final 2 clusters were identified as imminent when the "SOS button" was pressed or in the case of "external stressful events and crisis situations." At this juncture, the patients casted their votes on the question of whether the therapist should be informed without the patient's consent. The votes were one in favor and one against. A single individual posited that the assistant should be permitted to establish contact without delay, if the patient has consented. In total, 3 patients indicated a preference for multiple inquiries before action. This reinforces the necessity for individualized approaches and step-by-step plans that can be triggered according to the circumstances.

Discussion

Principal Findings

In the first part of the research findings, we examined how a VTA could be used as part of ongoing outpatient care. This includes the role (RQ1) of a VTA, particularly in the potential of filling the treatment gap, after discharge from the hospital. The respondents anticipated a system that would integrate expertise and facilitate the establishment of relationships between individuals. In conjunction with a primary recommendation from a trusted confidant of the therapist, the role of a peer counselor was also considered. Results on VTA embodiment (RQ2) indicated the need for such a counselor and provided practical implications, such as an imperfect or neutral-looking but emotionally expressive human embodiment, with a tendency to avoid male characters. In the second part of the findings, we evaluated patient and health care professionals' expectations regarding the interaction between a VTA and a patient. The results on interaction timing (RQ3) were considerable variables, reflecting the importance of an individual rhythm. The results obtained were delineated into 2 main clusters: *situation-independent* and *situation-dependent*. Regarding interaction abilities (RQ4) that the VTA should offer to support patients during treatment, 2 main clusters were also identified. *Structure-providing* included behavior for structuring and time management. *Emotional-supportive* included 3 subclusters *supportive-trusting*,*supportive-motivational*, and *supportive-resource-promoting.* The results regarding interactive behavior (RQ5) are presented separately for each of the subclusters identified in RQ3 and RQ4 and include explicit, competence-related, or emotion-stabilizing behavioral instructions. In addition, another cluster was identified as *taboo topics*, which includes content that should not be discussed.

The results pertaining to the interface between the assistant and the treating professionals (RQ6) are particularly relevant in determining the content that should be shared, the frequency, the timing, and patients consent.

Role

Expectations of the role of the assistant strongly reflect the treatment gap—both groups recognized the potential of a VTA as a supportive anchor following hospital discharge, not constrained by availability of health care professionals, but still equipped with professional techniques to navigate challenging circumstances (that relatives may not possess), while simultaneously serving as a projection screen for personal attachment. This leads us to conclude that a VTA has the potential to address the treatment gap, at least in part, and thus enhance outpatient aftercare.

Embodiment

This survey replicates previous findings [40] showing the need for an embodiment of the VTA and provides further insight into the practical implementation of this perspective. At first glance, the patients' expressed desire for an imperfect figure that could be interpreted as an indication of the "uncanny Valley effect" [61]. This phenomenon describes the abrupt decline in the acceptance of robots when they become too similar to humans. Nevertheless, the statements explicitly refer to aesthetic features, such as the (imperfect) figure and suggest that the assistant should not evoke negative social comparisons or negative self-evaluations, which can be a symptom in patients with depression. In this context, patients express fears of doubting their own self-esteem [62]. A common clinical presentation among individuals with depressive disorders is a perceived lack of self-esteem [63]. It would be beneficial to consider whether the VTA could be helpful in demonstrating how to make self-evaluation more positive, to show solution strategies. Notable examples include the chat robot Webot, which uses self-learning mechanisms through the presentation of short narratives about itself [64], or the home robot Jibo, which uses positive psychology techniques to assist individuals through the day [65].

A striking tendency in our data was the recommendation regarding the gender of the virtual assistant. Here, neither health care professionals nor patients explicitly recommended a male character; rather, a female character or a character with selectable gender was the preferred option. In other words, when explicitly asked to select a gender for a VTA, both health care professionals and patients avoided choosing a male character. Considering the highly sensitive target group of inpatient depressive patients on which our data were based, this recommendation suggests that the figure of the assistant should not convey any sense of threat. Other indications from the results regarding appearance included a "soft, calm voice" and "pleasant pastel shades, nothing too dark..." It can be concluded that men are perceived as more threatening according to the specific needs of the target group with depression. Statistical evidence indicates that instances of domestic violence against males and same-sex domestic violence are perceived as less threatening to society than instances of domestic male-on-female violence [66]. Research indicates that domestic violence has a profound impact on individuals and families and it is a substantial risk factor for mental illness [67]. In light of the observed tendency to perceive women as more empathetic, a phenomenon that has been documented in previous research on VTAs [68], it is plausible that the gender recommendation may have a dual impact here. It can be posited that both patients and health care professionals subconsciously had a preference for a female figure due to the aforementioned associations and the underlying motive to establish trust and security. Conversely, research on stereotypes in the form of dominance and gender in the VTA

XSL•FO
RenderX

indicates that male-associated behavior in female VTAs that does not align with the stereotype is met with rejection [41]. This indicates that it is crucial for gender-related conduct to align with the character's physical appearance. With respect to the specific target groups under consideration in this study, the impact of human-agent interaction in this regard should be examined, particularly in light of the notable gender-related outcomes observed here. Furthermore, these findings could be extended to other applications of SIAs where gender is a pivotal factor.

Some patients expressed a desire for the VTA to be modeled on their personal therapist. In light of the potential for technical advancement to enable the realistic modeling of avatars based on specific individuals [69], this option may be regarded as a realistic prospect. The therapeutic advantages (eg, availability of assistance) or disadvantages (eg, increased dependence on the therapist) would be highly specific to the particular therapeutic function or therapeutic relationship in question. As robots and intelligent systems begin to blur the boundaries between physical and virtual reality, there is an increased risk of negative emotions and thoughts being transferred to an assistant that has been modeled on someone from the patient's life. It remains to be seen whether intelligent systems will develop the capacity to counteract such effects in the future. This would require careful research [70]. However, the desire to recreate a known person as their VTA reflects patients' fear of being left alone and without trusted support after leaving the hospital. This was a common theme throughout this survey.

Regarding external expressiveness, patients and health care professionals indicated that the assistant should be capable of communicating through facial expressions and gestures. It is recommended that the face, upper body, arms, and hands should be visible. In addition to voice and content, body language is also of importance. This indicates that the affective functions of the assistant should be multimodal to be able to act emphatically. In crisis situations, patients may engage with such a system with their undivided attention. In contrast to other contextual studies, such as those examining a driving assistant where users have only a divided attention available for interaction with the assistant, it is desired for VTAs to be less distracting (only one visible head; [41). In addition, it is demonstrated that participants want to allocate more attention to the assistant and want it to be expressive and empathetic. The research on nonverbal synchrony in therapeutic sessions indicates that high levels of nonverbal synchrony are associated with a higher quality of the therapeutic relationship and greater symptom reduction. Consequently, the design of an assistant capable of nonverbal synchrony with the patient serves as an additional indicator of therapeutic efficacy [71]. Inferentially, through an assumption of an outward appearance with corresponding expressiveness, participants in both survey groups expressed a desire for this function. However, it is of particular importance, particularly in the context of mental health, to design expressive VTA behavior in accordance with the specific requirements of the health care target group [72].

In conclusion, the findings indicate that embodiment is a substantial factor, but its implementation must be approached with caution to ensure a beneficial impact.

Timing

In response to the question of timing, the 2 main clusters were identified as *situation-independent* and *situation-dependent*. The results of the *situation-independent* cluster may also be summarized as proactive behavior on the part of the assistant. Proactive behavior on the part of the patient was not mentioned by any of the respondents. In general, this demonstrates that the assistant is regarded as an activating medium and the patients are perceived as passive recipients of data by the therapists and by themselves, except in situations where requests are situation related. Here, the particular requirements of the specific depressive group are evident. This indicates that the initial 6-month period following discharge is a critical period for the risk of relapse and suicide [4]. The reintegration into to the home environment frequently results in incomplete remission, and the persistence of symptoms further complicates the ability to independently coordinate the various stakeholders in sociopsychiatric aftercare. The accompanying psychosocial restrictions, such as a lack of daily structure or cognitive limitations, make the necessity for a situation-independent, activating figure apparent. The almost unlimited availability of activating resources of such a system appears to be of great functional advantage at this point. Therefore, the unconstrained accessibility of activating resources within an attachment-oriented system gives rise to concerns that are not seen in the natural boundaries of human-human interactions. In light of the aforementioned limitations, a situation-independent activating assistant appears to be a logical solution for patients with depression during aftercare. However, there is a risk that unrestricted availability of the attachment-oriented assistant may result in dependency among patients. For example, a previous study [73] on attachment-oriented robots demonstrated that a system that provides emotional attachment in addition to factual competencies and addresses loneliness may potentially foster dependency in patients. It is yet to be determined whether a system designed to counteract this can be developed in a way that incorporates factual competencies at all levels into a training mode that is inherently geared toward resource building, thereby creating space for development and independence.

The frequency-related results, that is, the question of when the assistant and patient should interact, demonstrate the necessity for a more individualized system. The need for individualization can be regulated in 5 different ways: at the request of the patient, according to content, according to time, when the patient's condition changes and before outpatient therapy or appointments.

Regarding frequency of interactions, these may occur daily at designated times, on a weekly or monthly basis, or at other intervals. Interactions surrounding particular events may also be structured temporally in accordance with scheduled appointments. Human-interaction therapy sessions in a social psychiatric context are typically conducted at varying frequencies. At the outset of a treatment agreement, a potential frequency is typically established, contingent on the patient's needs and the therapist's availability. Research has shown that participants, especially those with lower levels of education, preferred a VTA to a human after discharge from hospital because they could control the pace of information [74].

XSL•FO
RenderX

Furthermore, the frequency of therapy appointments tends to increase gradually, contingent upon the efficacy of the therapy and the availability of approved cost coverage. In this context, the frequency of therapy has an impact on the success of therapy. Consequently, researchers advise therapists to treat patients as often as possible, particularly during the initial phase of therapy [75]. At this juncture, a virtual system can bridge gaps in care due to its unlimited availability and supplementary therapies that are more limited in scope. Conversely, it is of the utmost importance to prevent the patient from becoming dependent on the therapist or on a digital system. It is therefore imperative that a digital system, which is capable of maintaining unlimited availability, simultaneously instills a sense of autonomy and independence to prevent dependency. Should the patient wish to interact, this would necessitate the individual activation of the system by the patient. Content-dependent interactions refer to the dialogue or content that the patient shares with the system, either verbally or by answering questions. The individualization of the content of the dialogue represents a key aspect of this process. Interactions that occur when the patient's emotional state changes can be proactively initiated by the patient. It is similarly conceivable that this could form the basis of a digital phenotyping monitoring system. The concept of being "seen" by a system that responds to crises, provides assistance, or alerts to early warning symptoms was generally viewed positively. New interventions that take advantage of the quantification of clinical markers through condition monitoring can contribute to clinical improvement [76,77]. Our results indicate user acceptance of such monitoring at this point, as well as needs and preferences.

Interaction Abilities

Regarding the interaction skills that the assistant should be able to offer to provide support to patients undergoing treatment, 2 broad clusters were identified. On the one hand, the skills that could be classified as structured assistance were designated as the *structure-providing* cluster. In contrast, the skills that could be assigned to emotional support were classified as the *emotional-supportive* cluster. Both of these main clusters could be defined in a cross-dimensional manner, thereby allowing for the identification of a higher-order meaning. In this context, we focused on 2 sociocognitive dimensions that are particularly relevant for the description of individuals: warmth and competence [78,79]. These 2 variables have been identified by Fiske et al [78] as the 2 most important factors by which people evaluate encounters. The character traits that individuals form about others when they spontaneously gather impressions can be summarized in these 2 basic dimensions [80]. The dimension of warmth is defined as a set of traits that are perceived to indicate intentions of friendliness, helpfulness, sincerity, trustworthiness, and morality. In contrast, the competence dimension traits are associated with perceptions of abilities, including intelligence, skill, creativity, and efficacy [78]. Our data point to these sociocognitive dimensions within the 2 main clusters.

In addition, the data delineates the interaction skills, which include the virtual assistant's emotions, personality, and social skills. In conjunction with the character's appearance, these skills influence the user's assessment of the character's believability. The term "believability" in the context of VTAs is defined as the capacity to be perceived as a genuine entity [81]. For this to occur, the character must act in a manner that is consistent with its stated goals, personality, and state of mind [82]. In this context, consistency is defined as the coherence between the characters' verbal and nonverbal behaviors as well as its physical appearance. In particular, our data substantiate the hypothesis that users of virtual assistants within the subcluster *supportive-trusting* desire consistency in the assistant's behavior. This desire was described as "authentic" as well as socially interactive and competent in the sense that the assistant's emotional expressions and statements should be appropriate to the situation. It is therefore evident that the VTA must be capable of modifying its social conduct in accordance with the circumstances to appear credible and pertinent. Our findings corroborate those of previous research, which have highlighted the significance of believability in virtual assistants [81]. In addition, our data indicate that believability can be classified as a subcluster of the warmth dimension, and can be seen as a subconstruct of the 2 sociocognitive dimensions of warmth and competence.

Interaction Behavior

In contrast to human therapeutic interactions, which always entail an exchange about a momentary, subjective state and require the patient to disclose personal content, a digital monitoring system maintains a pervasive and comprehensive impression that may not necessitate an active decision by the patient to communicate. At this juncture, users of a digital monitoring system are disclosing a considerable amount of personal information by sharing data that is not actively shared, thereby relinquishing both responsibility and control over the sharing process. The data indicates that, regarding the variety of tools and forms of intervention, there is no substantial bias toward data sharing among the 2 groups of respondents. The primary objective is to identify a figure who can effectively combine competence with warmth, thereby fostering the most robust possible trust between the patient and the figure. As evidenced by psychotherapy research, trust and shame play a major impact in the efficacy of therapeutic interventions [83,84]. Assisting patients in modifying their behaviors also necessitates transparency regarding their shortcomings. Concurrently, individuals divulge a considerably more intimate array of information to computers than to humans [85,86]. Despite the well-established fact that data is identifiable, indelible, and, in certain circumstances, obtainable [87], computers do not elicit the same social inhibitions as humans. In the context of human–computer interaction, particularly in the context of digital assistants, it is pertinent to inquire into the underlying mechanisms that facilitate this phenomenon. Against this background, it seems plausible that crucial psychometric inquiries in psychiatric diagnostics may be more effectively elucidated through human-machine interaction. It is beyond dispute that the protection of patient data must be accorded with the highest priority. In its social function, shame serves as a reflection of the affected person's behavior and has developmental potential. Conversely, shame entails the participation of at least 2 individuals, thereby indicating that the observed transgression was inadvertent. In this way, shame

XSL•FO
RenderX

functions as a feedback mechanism, protecting against extremes or loss of control and leading to inclusive behavior in interaction partners. In social interactions, the successful overcoming of shame can facilitate the formation of attachment on both sides. However, existential shame, which is defined as the perception of oneself as generally flawed, can result in social withdrawal and, ultimately, the onset of depressive symptomatology [88,89]. It is, therefore, possible to hope that deficiencies in care in human-machine interaction can also be addressed at the level of communication. A solution to social psychiatric diagnostics that is free from shame can help to prevent crises, particularly in situations of critical transition, such as after hospitalization. At the same time, it should be noted that a shame-free interaction with a computer will not replace the challenges that people experience in social encounters.

Assistant-Health Care Professional Interaction

This final inquiry relates to the interface between the assistant and the health care professionals, that is, modifications to the framework to ensure that health care professionals benefit from the implementation of a VTA. Moreover, this could be regarded as a control function to guarantee that the assistant does not unintentionally impede the ongoing treatment.

The results from all cluster dimensions demonstrate substantial overlap across the full spectrum of social psychiatric care, which raises the question of how such a support system could potentially transform the landscape of social psychiatric care. For example, it could result in a shift in social values within the domain of social psychiatry or contribute to an increased reliance on digital technologies for the provision of care. Accordingly, the societal impact of an intelligent, attachment-oriented assistance system on both human-human and human-robot relationships must be explored and designed to be participatory at all levels of use [90]. Evaluations need to take place at all levels of development in this process, so that all targeted professional groups, as well as patient advocacy organizations, can provide input [76]. In particular, possible short- and long-term effects on the individual patient's identity, freedom of choice, and self-perception need to be explored.

Two potential outcomes emerge at the interaction interface between therapist and assistant, as evidenced by the overall results. On the one hand, there is the possibility of receiving information about the patient's health status via a ubiquitous computer, which the patient would be disinclined to share with the perhaps new, still unknown therapist for reasons already discussed. The primary reasons for this were insufficient time or a lack of confidence on the part of the patient. Conversely, the alarm system serves to alert the therapist or the rescue system in the event of indications of suicidal ideation. These 2 points present the potential for addressing existing care and enhancing the safety of patient follow-up for those with depression [91]. Concurrently, the frequently arduous period of follow-up care will be less onerous for treatment providers. The enhanced objectivity of the assessment of patients with depression who are to be monitored provides a basis for decision-making in the context of treatment.

It would be beneficial to ascertain which groups of patients are particularly at risk of relapse following hospitalization. Those from socioeconomically disadvantaged backgrounds are frequently inadequately served. Psychiatric care is underrepresented in rural, economically underdeveloped regions [1]. Furthermore, the risk of relapse is elevated in individuals who are young at the onset of illness, female, older, have somatic comorbidity, are single, and have a general lack of psychosocial support determine a high susceptibility to relapse [2]. During the course of the pandemic, the rise in the incidence of mental health disorders occurred concurrently with substantial disruptions to the provision of mental health services [92]. A deficiency in the provision of essential health services was observed, including those pertaining to mental health and suicide prevention, as reported by the World Health Organization. While the situation showed some improvement by the end of the pandemic, it still demonstrates the vulnerability of a health care system to unforeseen crises. Health care professionals caring for these susceptible people could benefit professionally from a VTA system that helps to make outpatient care safer. The effect sizes, considering the different groups of people affected, remain an open RQ.

Limitations and Directions for Future Research

This study has several limitations. First, the recruitment and questioning of the 2 groups of participants was carried out in different ways, with the questioning being adapted both to the specific groups and to the specific situations. This could have led to distortion. Second, the results showed that the answer option "diverse" to the gender question was misunderstood. Future studies should pay attention to clearer formulation. In addition, psychometric data were used solely for descriptive purposes, despite the potential for exploring the impact of individual factors on outcomes, a direction that merits attention in future research. The study's scope is further limited by its inclusion of health care professionals and patients from Oldenburg, Germany, precluding the generalizability of its results. To enhance the relevance and external validity of the findings, subsequent studies should aim to encompass more representative samples of the general population. Furthermore, this study exclusively examined participants' attitudes toward an SIA, neglecting to investigate its potential effects on patients. Consequently, further research is necessary to investigate the impact of such a system and to ensure the benefits for those affected at all levels of the system.

Conclusions

Main conclusions can be derived from the study related to the role of VTAs, their embodiment, the timing of interaction, interaction skills and behavior, and interaction between the assistant and health professionals.

Role

The VTA was identified as a valuable supplementary component within the outpatient aftercare system. The VTA is able to address deficiencies in care due to its extensive repertoire of therapeutic skills and its capacity to establish a rapport with patients where social workers, therapists, and family members are unable to do so.

XSL•FO
RenderX

Embodiment

A multimodal function for empathic communication (ie, voice, facial expressions, and gestures) is expected. Patients may be more likely to engage with an assistant who appears imperfect, as this may help to prevent negative self-evaluation, particularly in patients with depression who lack self-esteem. The assistant may be able to assist in the development of solution strategies to enhance a more positive self-image. The gender of the assistant should be either female or selectable, but not male. This preference is thought to be based on negative associations with the male gender, such as domestic violence, and the perception of the female gender as more empathetic.

Timing of Interaction

The findings indicated that both patients and health care professionals anticipate the assistant to engage in proactive behavior. Therefore, the assistant is regarded as an activator of information, whereas patients are viewed as passive recipients. The expectations regarding the timing of specific behaviors can be differentiated into 2 categories: situation-independent and situation-dependent behaviors.

Interaction Skills and Behavior

Two core skills were identified as being essential for both health care professionals and patients: the capacity to provide structural support and the capacity to provide emotional support. These abilities correspond to the dimensions of interpersonal social perception, namely competence and warmth. An agent who exhibits these 2 abilities through appropriate behavior at the appropriate time could serve as a valuable addition to outpatient aftercare. For each subcluster and appropriate timing, the results provide an extensive list of suggestions for appropriate behavior.

Interaction Between Assistant and Health Professional

The results demonstrate the potential for therapists to obtain otherwise unavailable information regarding patients' health status, as well as the feasibility of an emergency system in the event of suicidal risk. These innovations have the potential to address existing gaps in aftercare, enhance safety during this critical period, reduce the burden on health care professionals, and, by providing an objective measure of the patient's health status, facilitate health care professional's decision-making.

Summary of Conclusions

In summary, our findings offer a promising framework for the development of a VTA to support patients with depression during aftercare and support the work of health care professionals.

Acknowledgments

This research was funded by the UBIDENZ project (13GW0568A) and partly sponsored by the Hybrid Intelligence project (024.004.022). The artificial intelligence–supported DeepL translation tool was used for translation of certain expressions used in the final version of this paper. The authors would like to thank Dr Klaus Bajorat, Maira Keller, Anna Lea Reinwarth, Daniel Faber, Madina Khalillaeva, David Mietzner, and Rachana Mahadevan for their help with creating the experimental setup, data collection, and analysis.

The authors would like to thank their collaborators for their contributions: Janet Wessler, PhD, Tanja Schneeberger, PhD, Jan Alexanderson, PhD, Nicklas Linz, PhD, Norbert Helff, Dipl, Arno Elmer, PhD, and Elisabeth André, PhD.

Data Availability

Following the recommendations for open research practices [50], questionnaires and experimenter instruction can be found on the OSF website [51]. The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Depression and other common mental disorders: global health estimates. World Health Organization. URL: https://policycommons.net/artifacts/546082/depression-and-other-common-mental-disorders/1523689/ [accessed 2025-09-03]
2. Unipolare depression. Nationale Versorgungs Leitlinie. 2015. URL: https://register.awmf.org/assets/guidelines/nvl-005l_S3_Unipolare-Depression_2023-07.pdf [accessed 2024-04-29]
3. Wiegand HF, Saam J, Marschall U, Chmitorz A, Kriston L, Berger M, et al. Challenges in the transition from in-patient to out-patient treatment in depression. Dtsch Arztebl Int 2020 Jul 06;117(27-28):472-479 [FREE Full text] [doi: 10.3238/arztebl.2020.0472] [Medline: 33050996]
4. Baldessarini RJ, Tondo L, Pinna M, Nuñez N, Vázquez GH. Suicidal risk factors in major affective disorders. Br J Psychiatry 2019 Jul 11;215(4):1-6. [doi: 10.1192/bjp.2019.167] [Medline: 31292010]

XSL•FO
RenderX

5. Chung DT, Ryan CJ, Hadzi-Pavlovic D, Singh SP, Stanton C, Large MM. Suicide rates after discharge from psychiatric facilities: a systematic review and meta-analysis. JAMA Psychiatry 2017 Jul 01;74(7):694-702 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.1044] [Medline: 28564699]
6. Castelli Dransart DA, Heeb J, Gulfi A, Gutjahr EM. Stress reactions after a patient suicide and their relations to the profile of mental health professionals. BMC Psychiatry 2015 Oct 28;15(1):265 [FREE Full text] [doi: 10.1186/s12888-015-0655-y] [Medline: 26511910]
7. Kleespies PM, Penk WE, Forsyth JP. The stress of patient suicidal behavior during clinical training: incidence, impact, and recovery. Prof Psychol Res Pr 1993;24(3):293-303. [doi: 10.1037//0735-7028.24.3.293]
8. Cushway D, Tyler P. Stress in clinical psychologists. Int J Soc Psychiatry 1996 Jun 01;42(2):141-149. [doi: 10.1177/002076409604200208] [Medline: 8811398]
9. Deutsch CJ. Self-reported sources of stress among psychotherapists. Prof Psychol Res Pr 1984;15(6):833-845. [doi: 10.1037//0735-7028.15.6.833]
10. Sfetcu R, Musat S, Haaramo P, Ciutan M, Scintee G, Vladescu C, et al. Overview of post-discharge predictors for psychiatric re-hospitalisations: a systematic review of the literature. BMC Psychiatry 2017 Jun 24;17(1):227 [FREE Full text] [doi: 10.1186/s12888-017-1386-z] [Medline: 28646857]
11. Ho TP. The suicide risk of discharged psychiatric patients. J Clin Psychiatry 2003 Jun 15;64(6):702-707. [doi: 10.4088/jcp.v64n0613] [Medline: 12823086]
12. Mahmoud AS, Elsayed Berma A, Saleh Gabal SA. Relationship between social support and the quality of life among psychiatric patients. J Psychiatry Psychiatric Disord 2017;01(02):57-75. [doi: 10.26502/jppd.2572-519x008]
13. Büscher R, Torok M, Terhorst Y, Sander L. Internet-based cognitive behavioral therapy to reduce suicidal ideation: a systematic review and meta-analysis. JAMA Netw Open 2020 Apr 01;3(4):e203933 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3933] [Medline: 32343354]
14. Eichenberg C. E-Mental Health Anwendungen für depressive und suizidale Menschen: eine Übersicht. Psychother Forum 2019 Nov 14;23(3-4):111-119. [doi: 10.1007/S00729-019-00123-3]
15. Klein JP, Knaevelsrud C, Bohus M, Ebert DD, Gerlinger G, Günther K, et al. [Internet-based self-management interventions : quality criteria for their use in prevention and treatment of mental disorders]. Nervenarzt 2018 Nov 20;89(11):1277-1286. [doi: 10.1007/s00115-018-0591-4] [Medline: 30128736]
16. Miller I, Peake E, Strauss G, Vierra E, Koepsell X, Shalchi B, et al. Self-guided digital intervention for depression in adolescents: feasibility and preliminary efficacy study. JMIR Form Res 2023 Nov 22;7:e43260 [FREE Full text] [doi: 10.2196/43260] [Medline: 37991839]
17. Bowlby J. Attachment theory and its therapeutic implications. Adolesc Psychiatry 1978;6:5-33. [Medline: 742687]
18. Grawe K, Donati R, Bernauer F. Psychotherapie im Wandel: Von der Konfession zur Profession. Göttingen, Germany: Hogrefe Publishing; 1994.
19. Urech A, Krieger T, Möseneder L, Biaggi A, Vincent A, Poppe C, et al. A patient post hoc perspective on advantages and disadvantages of blended cognitive behaviour therapy for depression: a qualitative content analysis. Psychother Res 2019 Nov 31;29(8):986-998. [doi: 10.1080/10503307.2018.1430910] [Medline: 29385964]
20. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]
21. Gebhard P, Schneeberger T, Andre E, Baur T, Damian I, Mehlmann G, et al. Serious games for training social skills in job interviews. IEEE Trans Games 2019 Dec;11(4):340-351. [doi: 10.1109/tg.2018.2808525]
22. Gratch J, Wang N, Gerten J, Fast E, Duffy R. Creating rapport with virtual agents. In: Proceedings of the 7th International Working Conference on Intelligent Virtual Agents. 2007 Presented at: IVA '07; September 17-19, 2007; Paris, France p. 125-138 URL: https://link.springer.com/chapter/10.1007/978-3-540-74997-4_12 [doi: 10.1007/978-3-540-74997-4_12]
23. Gratch J, Kang SH, Wang N. Using social agents to explore theories of rapport and emotional resonance. In: Gratch J, Marsella S, editors. Social Emotions in Nature and Artifact. Oxford, UK: Oxford Academic Press; 2013:181-197.
24. Darcy A, Daniels J, Salinger D, Wicks P, Robinson A. Evidence of human-level bonds established with a digital conversational agent: cross-sectional, retrospective observational study. JMIR Form Res 2021 May 11;5(5):e27868 [FREE Full text] [doi: 10.2196/27868] [Medline: 33973854]
25. Dias LP, Barbosa JL, Feijó LP, Vianna HD. Development and testing of iAware model for ubiquitous care of patients with symptoms of stress, anxiety and depression. Comput Methods Programs Biomed 2020 Apr;187:105113. [doi: 10.1016/j.cmpb.2019.105113] [Medline: 31607411]
26. Schneeberger T, Reinwarth AL, Wensky R, Anglet MS, Gebhard P, Wessler J. Fast friends: generating interpersonal closeness between humans and socially interactive agents. In: Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents. 2023 Presented at: IVA '23; September 19-22, 2023; Würzburg, Germany p. 1-8 URL: https://dl.acm.org/doi/10.1145/3570945.3607302 [doi: 10.1145/3570945.3607302]
27. Gebhard P. ALMA: a layered model of affect. In: Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems. 2005 Presented at: AAMAS '05; July 25-29, 2005; Amsterdam, The Netherlands p. 29-36 URL: https://dl.acm.org/doi/10.1145/1082473.1082478 [doi: 10.1145/1082473.1082478]

XSL•FO
RenderX

28. Gebhard P, Schneeberger T, Baur T, André E. MARSSI: model of appraisal, regulation, and social signal interpretation. In: Proceedings of the 17th International Conference on Autonomous Agents and Multi Agent Systems. 2018 Presented at: AAMAS '18; July 10-15, 2018; Stockholm-Uppsala, Sweden p. 497-506 URL: https://dl.acm.org/doi/10.5555/3237383.3237458 [doi: 10.5555/3237383.3237458]
29. Zhang JE, Hilpert B, Broekens J, Jokinen JP. Simulating emotions with an integrated computational model of appraisal and reinforcement learning. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '24; May 11-16, 2024; Honolulu, HI p. 1-12 URL: https://dl.acm.org/doi/full/10.1145/3613904.3641908
30. Davis M, Hadiks D. Nonverbal aspects of therapist attunement. J Clin Psychol 1994 May;50(3):393-405. [doi: 10.1002/1097-4679(199405)50:3<393::aid-jclp2270500311>3.0.co;2-t] [Medline: 8071445]
31. Lugrin B. Introduction to socially interactive agents. In: Lugrin B, Pelachaud C, Traum D, editors. The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics. New York, NY: Association for Computing Machinery; 2021:1-20.
32. Schneeberger T, Hirsch A, König C, Gebhard P. Impact of virtual environment design on the assessment of virtual agents. In: Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents. 2019 Presented at: IVA '19; July 2-5, 2019; Paris, France p. 148-150 URL: https://dl.acm.org/doi/10.1145/3308532.3329455 [doi: 10.1145/3308532.3329455]
33. Nicora ML, André E, Berkmans D, Carissoli C, D'Orazio T, Fave AD. A human-driven control architecture for promoting good mental health in collaborative robot scenarios. In: Proceedings of the 30th IEEE International Conference on Robot & Human Interactive Communication. 2021 Presented at: RO-MAN '21; August 8-12, 2021; Vancouver, BC p. 285-291 URL: https://ieeexplore.ieee.org/document/9515315 [doi: 10.1109/ro-man50785.2021.9515315]
34. Bhuvaneshwara C, Anglet M, Hilpert B, Chehayeb L, Meyer AK, Don DW, et al. MITHOS-mixed reality interactive teacher training system for conflict situations at school. In: Proceedings for the 2023 Conference on International Society of the Learning Sciences. 2023 Presented at: ISLS '23; June 10-15, 2023; Montreal, QC p. 43 URL: https://tinyurl.com/vv9e8a5y
35. Schneeberger T, Sauerwein N, Anglet MS, Gebhard P. Stress management training using biofeedback guided by social agents. In: Proceedings of the 26th International Conference on Intelligent User Interfaces. 2021 Presented at: IUI '21; April 14-17, 2021; College Station, TX p. 564-574 URL: https://dl.acm.org/doi/10.1145/3397481.3450683 [doi: 10.1145/3397481.3450683]
36. Ali MR, Van Orden K, Parkhurst K, Liu S, Nguyen VD, Duberstein P, et al. Aging and engaging: a social conversational skills training program for older adults. In: Proceedings of the 23rd International Conference on Intelligent User Interfaces. 2018 Presented at: IUI '18; March 7-11, 2018; Tokyo, Japan p. 55-66 URL: https://dl.acm.org/doi/10.1145/3172944.3172958 [doi: 10.1145/3172944.3172958]
37. Aylett RS, Louchart S, Dias J, Paiva A, Vala M. FearNot! – an experiment in emergent narrative. In: Proceedings of the 5th International Working Conference on Intelligent Virtual Agents. 2005 Presented at: IVA '05; September 12-14, 2005; Kos, Greece p. 305-316 URL: https://link.springer.com/chapter/10.1007/11550617_26 [doi: 10.1007/11550617_26]
38. Morris RR, Kouddous K, Kshirsagar R, Schueller SM. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. J Med Internet Res 2018 Jun 26;20(6):e10148 [FREE Full text] [doi: 10.2196/10148] [Medline: 29945856]
39. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a relational agent by hospital patients with depressive symptoms. Interact Comput 2010 Jul 01;22(4):289-298 [FREE Full text] [doi: 10.1016/j.intcom.2009.12.001] [Medline: 20628581]
40. Kim K, Boelling L, Haesler S, Bailenson J, Bruder G, Welch GF. Does a digital assistant need a body? The influence of visual embodiment and social behavior on the perception of intelligent virtual agents in AR. In: Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality. 2018 Presented at: ISMAR '18; October 16-20, 2018; Munich, Germany p. 105-114 URL: https://ieeexplore.ieee.org/document/8613756 [doi: 10.1109/ismar.2018.00039]
41. Wessler J, Schneeberger T, Christidis L, Gebhard P. Virtual backlash: nonverbal expression of dominance leads to less liking of dominant female versus male agents. In: Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents. 2022 Presented at: IVA '22; September 6-9, 2022; Faro, Portugal p. 1-8 URL: https://dl.acm.org/doi/10.1145/3514197.3549682 [doi: 10.1145/3514197]
42. Hilpert B, da Silva CA, Christidis L, Bhuvaneshwara C, Gebhard P, Nunnari F, et al. Avatar visual similarity for social HCI: increasing self-awareness. arXiv Preprint posted online August 23, 2024 [FREE Full text] [doi: 10.2139/ssrn.4797708]
43. Weitz K, Schiller D, Schlagowski R, Huber T, André E. "Do you trust me?": increasing user-trust by integrating virtual agents in explainable AI interaction design. In: Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents. 2019 Presented at: IVA '19; July 2-5, 2019; Paris, France p. 7-9 URL: https://dl.acm.org/doi/10.1145/3308532.3329441 [doi: 10.1145/3308532.3329441]
44. Hilpert B, Schneeberger T, Gebhard P. Employing virtual agents for building trust in driving automation: a qualitative pilot study. In: Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction. 2023 Presented at: FAccT '23; June 12-15, 2023; Honolulu, HI URL: https://www.researchgate.net/publication/349924456_Employing_Virtual_Agents_for_Building_Trust_in_Driving_Automation_A_Qualitative_Pilot_Study

XSL•FO
RenderX

45. Loveys K, Sebaratnam G, Sagar M, Broadbent E. The effect of design features on relationship quality with embodied conversational agents: a systematic review. Int J of Soc Robotics 2020 Sep 24;12(6):1293-1312. [doi: 10.1007/s12369-020-00680-7]
46. Gebhard P, Schneeberger T, Mehlmann G, Baur T, André E. Designing the impression of social agents' real-time interruption handling. In: Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents. 2019 Presented at: IVA '19; July 2-5, 2019; Paris, France p. 19-21 URL: https://dl.acm.org/doi/10.1145/3308532.3329435 [doi: 10.1145/3308532.3329435]
47. Zhao R, Sinha T, Black AW, Cassell J. Socially-aware virtual agents: automatically assessing dyadic rapport from temporal patterns of behavior. In: Proceedings of the 16th International Conference on Intelligent Virtual Agents. 2016 Presented at: IVA '16; September 20-23, 2016; Los Angeles, CA p. 218-233 URL: https://link.springer.com/chapter/10.1007/978-3-319-47665-0_20 [doi: 10.1007/978-3-319-47665-0_20]
48. Schneeberger T, Scholtes M, Hilpert B, Langer M, Gebhard P. Can social agents elicit shame as humans do? In: Proceedings of the 8th International Conference on Affective Computing and Intelligent Interaction. 2019 Presented at: ACII '19; September 3-6, 2019; Cambridge, UK p. 164-170 URL: https://ieeexplore.ieee.org/document/8925481 [doi: 10.1109/acii.2019.8925481]
49. Rheu M, Shin JY, Peng W, Huh-Yoo J. Systematic review: trust-building factors and implications for conversational agent design. Int J Hum Comput Interact 2020 Sep 02;37(1):81-96. [doi: 10.1080/10447318.2020.1807710]
50. DeVault D, Artstein R, Benn G, Dey T, Fast E, Gainer A. SimSensei kiosk: a virtual human interviewer for healthcare decision support. In: Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems. 2014 Presented at: AAMAS '14; May 5-9, 2014; Paris, France p. 1061-1068 URL: https://dl.acm.org/doi/10.5555/2615731.2617415
51. Deladisma AM, Cohen M, Stevens A, Wagner P, Lok B, Bernard T, Association for Surgical Education. Do medical students respond empathetically to a virtual patient? Am J Surg 2007 Jun;193(6):756-760. [doi: 10.1016/j.amjsurg.2007.01.021] [Medline: 17512291]
52. Kramer P. Frames of Reference for Pediatric Occupational Therapy. Thousand Oaks, CA: Lippincott Williams & Wilkins; 2018.
53. Paiva A, Leite I, Boukricha H, Wachsmuth I. Empathy in virtual agents and robots: a survey. ACM Trans Interact Intell Syst 2017 Sep 19;7(3):1-40. [doi: 10.1145/2912150]
54. Chehayeb L, Bhuvaneshwara C, Anglet M, Hilpert B, Meyer AK. MITHOS: interactive mixed reality training to support professional socio-emotional interactions at schools. arXiv Preprint posted online September 2, 2024 [FREE Full text] [doi: 10.48550/arXiv.2409.12968]
55. Beck Depression Inventory II (BDI-II). APA PsycTests. URL: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft00742-000 [accessed 2024-04-29]
56. Yamagishi T, Yamagishi M. Trust and commitment in the United States and Japan. Motiv Emot 1994 Jun;18(2):129-166. [doi: 10.1007/bf02249397]
57. Wessler J, Schneeberger T, Hilpert B, Alles A, Gebhard P. Empirical research in affective computing: an analysis of research practices and recommendations. In: Proceedings of the 9th International Conference on Affective Computing and Intelligent Interaction. 2021 Presented at: ACII '21; September 28-October 1, 2021; Nara, Japan p. 1-8 URL: https://ieeexplore.ieee.org/document/9597418 [doi: 10.1109/acii52823.2021.9597418]
58. Immel D. Materials to a survey about virtual therapeutic assistants in outpatient aftercare - expectations of patients and health care professionals. Open Science Framework. URL: https://osf.io/umwkf/ [accessed 2025-04-03]
59. Glaser BG, Strauss AL. Grounded Theory: Strategien Qualitativer Forschung. Berlin, Germany: Verlag Hans Huber; 2010.
60. Hansen C, Grimshaw-Aagaard S. Developing a guideline for reporting and evaluating grounded theory research studies. Int J Health Sci 2018:64-76. [doi: 10.15640/ijhs.v6n1a8]
61. Mori M. Bukimi no tani [the uncanny valley]. CiNii. URL: https://cir.nii.ac.jp/crid/1370013168736887425 [accessed 2025-09-03]
62. Battle J. Relationship between self-esteem and depression. Psychol Rep 1978 Jun 01;42(3 PT 1):745-746. [doi: 10.2466/pr0.1978.42.3.745] [Medline: 674498]
63. Sowislo JF, Orth U. Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies. Psychol Bull 2013 Jan;139(1):213-240. [doi: 10.1037/a0028931] [Medline: 22730921]
64. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
65. Jeong S, Alghowinem S, Aymerich-Franch L, Arias K, Lapedriza A, Picard R. A robotic positive psychology coach to improve college students' wellbeing. In: Proceedings of the 29th IEEE International Conference on Robot and Human Interactive Communication. 2020 Presented at: RO-MAN '20; August 31-September 4, 2020; Naples, Italy p. 187-194 URL: https://ieeexplore.ieee.org/document/9223588 [doi: 10.1109/ro-man47096.2020.9223588]
66. Maurya BR. An overview on domestic violence. Asian J Multidiscip Res 2021;10:12. [doi: 10.5958/2278-4853.2021.01139.3]

XSL•FO
RenderX

67. Devakumar D, Palfreyman A, Uthayakumar-Cumarasamy A, Ullah N, Ranasinghe C, Minckas N, et al. Mental health of women and children experiencing family violence in conflict settings: a mixed methods systematic review. Confl Health 2021 Oct 15;15(1):74 [FREE Full text] [doi: 10.1186/s13031-021-00410-4] [Medline: 34654456]
68. Tanaka H, Nakamura S. The acceptability of virtual characters as social skills trainers: usability study. JMIR Hum Factors 2022 Mar 29;9(1):e35358 [FREE Full text] [doi: 10.2196/35358] [Medline: 35348468]
69. da Silva CA, Hilpert B, Bhuvaneshwara C, Gebhard P, Nunnari F, Tsovaltzi D. Visual similarity for socially interactive agents that support self-awareness. In: Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents. 2023 Presented at: IVA '23; September 19-22, 2023; Würzburg, Germany p. 1-3 URL: https://dl.acm.org/doi/10.1145/3570945.3607329 [doi: 10.1145/3570945.3607329]
70. Fiske A, Henningsen P, Buyx A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. J Med Internet Res 2019 May 09;21(5):e13216 [FREE Full text] [doi: 10.2196/13216] [Medline: 31094356]
71. Ramseyer F, Tschacher W. Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome. J Consult Clin Psychol 2011 Jun;79(3):284-295. [doi: 10.1037/a0023419] [Medline: 21639608]
72. Reinwarth AL, Schneeberger T, Nunnari F, Gebhard P, Altmann U, Wessler J. Look what I Made it do - the ModelIT method for manually modeling nonverbal behavior of socially interactive agents. In: Proceedings of the 25th International Conference on Multimodal Interaction. 2023 Presented at: ICMI '23; October 9-13, 2023; Paris, France p. 200-204 URL: https://tinyurl.com/nhbysh68 [doi: 10.1145/3610661.3616549]
73. Cresswell K, Cunningham-Burley S, Sheikh A. Health care robotics: qualitative exploration of key challenges and future directions. J Med Internet Res 2018 Jul 04;20(7):e10410 [FREE Full text] [doi: 10.2196/10410] [Medline: 29973336]
74. Bickmore TW, Pfeifer LM, Jack BW. Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. In: Proceedings of the 2009 SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented at: CHI '09; April 4-9, 2009; Boston, MA p. 1265-1274 URL: https://dl.acm.org/doi/10.1145/1518701.1518891 [doi: 10.1145/1518701.1518891]
75. Erekson DM, Lambert MJ, Eggett DL. The relationship between session frequency and psychotherapy outcome in a naturalistic setting. J Consult Clin Psychol 2015 Dec;83(6):1097-1107. [doi: 10.1037/a0039774] [Medline: 26436645]
76. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit Med 2019 Sep 6;2(1):88 [FREE Full text] [doi: 10.1038/s41746-019-0166-1] [Medline: 31508498]
77. Sachan D. Self-help robots drive blues away. Lancet Psychiatry 2018 Jul;5(7):547. [doi: 10.1016/s2215-0366(18)30230-x]
78. Fiske ST, Cuddy AJ, Glick P. Universal dimensions of social cognition: warmth and competence. Trends Cogn Sci 2007 Feb;11(2):77-83. [doi: 10.1016/j.tics.2006.11.005] [Medline: 17188552]
79. Harris LT, Fiske ST. Dehumanizing the lowest of the low: neuroimaging responses to extreme out-groups. Psychol Sci 2006 Oct 01;17(10):847-853. [doi: 10.1111/j.1467-9280.2006.01793.x] [Medline: 17100784]
80. Fischer K, Jungermann H. Rarely occurring headaches and rarely occurring blindness: is rarely=rarely? The meaning of verbal frequentistic labels in specific medical contexts. J Behav Decis Making 1996;9(3):153-172. [doi: 10.1002/(sici)1099-0771(199609)9:3<153::aid-bdm222>3.3.co;2-n]
81. Niewiadomski R, Demeure V, Pelachaud C. Warmth, competence, believability and virtual agents. In: Proceedings of the 10th International Conference on Intelligent Virtual Agents. 2010 Presented at: IVA '10; July 12-16, 2010; Philadelphia, PA p. 285 URL: https://link.springer.com/chapter/10.1007/978-3-642-15892-6_29 [doi: 10.1007/978-3-642-15892-6_29]
82. Rosis FD, Pelachaud C, Poggi I, Carofiglio V, Carolis BD. From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. Int J Hum Comput Interact 2003 Jul;59(1-2):81-118. [doi: 10.1016/s1071-5819(03)00020-x]
83. Black RS, Curran D, Dyer KF. The impact of shame on the therapeutic alliance and intimate relationships. J Clin Psychol 2013 Jun 04;69(6):646-654. [doi: 10.1002/jclp.21959] [Medline: 23382093]
84. Leach MJ. Rapport: a key to treatment success. Complement Ther Clin Pract 2005 Nov;11(4):262-265. [doi: 10.1016/j.ctcp.2005.05.005] [Medline: 16290897]
85. Joinson AN, Woodley A, Reips U. Personalization, authentication and self-disclosure in self-administered internet surveys. Comput Human Behav 2004 Nov 18;23(1):35. [doi: 10.1016/s0747-5632(04)00168-2]
86. Morris GE, Kennedy C. Personal finance questions elicit slightly different answers in phone surveys than online. Pew Rearch Centre. URL: https://www.pewresearch.org/short-reads/2017/08/04/personal-finance-questions-elicit-slightly-different-answers-in-phone-surveys-than-online/ [accessed 2025-09-03]
87. Chen A. Cambridge Analytica and our lives inside the surveillance machine. The New Yorker. 2018. URL: https://www.newyorker.com/tech/annals-of-technology/cambridge-analytica-and-our-lives-inside-the-surveillance-machine [accessed 2025-09-03]
88. Gilbert P. The relationship of shame, social anxiety and depression: the role of the evaluation of social rank. Clin Psychol Psychother 2000 Jul;7(3):174-189. [doi: 10.1002/1099-0879(200007)7:3<174::aid-cpp236>3.0.co;2-u]
89. Lutwak N, Ferrari JR. Shame-related social anxiety: replicating a link with various social interaction measures. Anxiety Stress Coping 2007 May 29;10(4):335-340. [doi: 10.1080/10615809708249307]

XSL•FO
RenderX

90. Rheinberger D, Baffsky R, McGillivray L, Zbukvic I, Dadich A, Larsen ME, et al. Examining the feasibility of implementing digital mental health innovations into hospitals to support youth in suicide crisis: interview study with young people and health professionals. JMIR Form Res 2023 Nov 16;7:e51398. [doi: 10.2196/51398] [Medline: 37971790]
91. Lv N, Kannampallil T, Xiao L, Ronneberg CR, Kumar V, Wittels NE, et al. Association between user interaction and treatment response of a voice-based coach for treating depression and anxiety: secondary analysis of a pilot randomized controlled trial. JMIR Hum Factors 2023 Nov 06;10:e49715 [FREE Full text] [doi: 10.2196/49715] [Medline: 37930781]
92. Lee E, Agustines D, Woo BK. Selection bias in digital conversations on depression before and during COVID-19. JMIR Form Res 2023 Nov 20;7:e42545 [FREE Full text] [doi: 10.2196/42545] [Medline: 37983077]

Abbreviations

RQ: research question
SIA: socially interactive agent
VTA: virtual therapeutic assistant

Edited by A Mavragani; submitted 15.04.24; peer-reviewed by A AL-Asadi; comments to author 07.11.24; revised version received 25.12.24; accepted 03.02.25; published 26.03.25.

Please cite as:
Immel D, Hilpert B, Schwarz P, Hein A, Gebhard P, Barton S, Hurlemann R
Patients' and Health Care Professionals' Expectations of Virtual Therapeutic Agents in Outpatient Aftercare: Qualitative Survey Study
JMIR Form Res 2025;9:e59527
URL: https://formative.jmir.org/2025/1/e59527
doi:10.2196/59527
PMID:

©Diana Immel, Bernhard Hilpert, Patricia Schwarz, Andreas Hein, Patrick Gebhard, Simon Barton, René Hurlemann. Originally published in JMIR Formative Research (https://formative.jmir.org), 26.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

AI-Enhanced Virtual Reality Self-Talk for Psychological Counseling: Formative Qualitative Study

Moreah Zisquit[1], MA; Alon Shoa[2], MSc; Ramon Oliva[3], PhD; Stav Perry[1], BA; Bernhard Spanlang[4], PhD; Anat Brunstein Klomek[1], Prof Dr; Mel Slater[3], Prof Dr; Doron Friedman[2], Prof Dr

[1]Baruch Ivcher School of Psychology, Reichman University, Herzliya, Israel

[2]Sammy Ofer School of Communications, Reichman University, Herzliya, Israel

[3]Event Lab, University of Barcelona, Barcelona, Spain

[4]Kiin S.L., Barcelona, Spain

Corresponding Author:
Moreah Zisquit, MA
Baruch Ivcher School of Psychology
Reichman University
Baruch Ivcher School of Psychology
Ha'universita 8
Herzliya
Israel
Phone: 972 523091406
Email: Zisquit.moreah@post.runi.ac.il

Abstract

Background: Access to mental health services continues to pose a global challenge, with current services often unable to meet the growing demand. This has sparked interest in conversational artificial intelligence (AI) agents as potential solutions. Despite this, the development of a reliable virtual therapist remains challenging, and the feasibility of AI fulfilling this sensitive role is still uncertain. One promising approach involves using AI agents for psychological self-talk, particularly within virtual reality (VR) environments. Self-talk in VR allows externalizing self-conversation by enabling individuals to embody avatars representing themselves as both patient and counselor, thus enhancing cognitive flexibility and problem-solving abilities. However, participants sometimes experience difficulties progressing in sessions, which is where AI could offer guidance and support.

Objective: This formative study aims to assess the challenges and advantages of integrating an AI agent into self-talk in VR for psychological counseling, focusing on user experience and the potential role of AI in supporting self-reflection, problem-solving, and positive behavioral change.

Methods: We carried out an iterative design and development of a system and protocol integrating large language models (LLMs) within VR self-talk during the first two and a half years. The design process addressed user interface, speech-to-text functionalities, fine-tuning the LLMs, and prompt engineering. Upon completion of the design process, we conducted a 3-month long exploratory qualitative study in which 11 healthy participants completed a session that included identifying a problem they wanted to address, attempting to address this problem using self-talk in VR, and then continuing self-talk in VR but this time with the assistance of an LLM-based virtual human. The sessions were carried out with a trained clinical psychologist and followed by semistructured interviews. We used applied thematic analysis after the interviews to code and develop key themes for the participants that addressed our research objective.

Results: In total, 4 themes were identified regarding the quality of advice, the potential advantages of human-AI collaboration in self-help, the believability of the virtual human, and user preferences for avatars in the scenario. The participants rated their desire to engage in additional such sessions at 8.3 out of 10, and more than half of the respondents indicated that they preferred using VR self-talk with AI rather than without it. On average, the usefulness of the session was rated 6.9 (SD 0.54), and the degree to which it helped solve their problem was rated 6.1 (SD 1.58). Participants specifically noted that human-AI collaboration led to improved outcomes and facilitated more positive thought processes, thereby enhancing self-reflection and problem-solving abilities.

Conclusions: This exploratory study suggests that the VR self-talk paradigm can be enhanced by LLM-based agents and presents the ways to achieve this, potential pitfalls, and additional insights.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e67782) doi:10.2196/67782

KEYWORDS

virtual human; large language model; virtual reality; self-talk; psychotherapy; artificial intelligence; AI

Introduction

Background

Overview

Access to mental health services is a critical issue on a global scale. Major depression is the leading cause of living with disability for many years and the fourth leading cause of disability-adjusted life years worldwide [1]. Furthermore, research indicates that >20% of people will experience mental illness in their lifetime [2]. Unfortunately, the current clinical workforce is insufficient, with approximately 9 psychiatrists per 100,000 people in high-income countries [3]. This shortfall in care has prompted interest in technology as a potential solution to bridge the gap between the need for treatment and the capacity to deliver it, particularly through the development of conversational agents or multipurpose virtual assistants [4]. These artificial intelligence (AI) agents can be integrated into virtual reality (VR) to simulate a personal, realistic therapeutic environment to reduce feelings of unnaturalness for the patient.

However, creating a virtual therapist presents significant challenges. Despite advancements in generative AI, the reliability of AI in sensitive therapeutic roles remains uncertain [4]. Some researchers argue that AI-based therapists should be used as tools to support both patients and therapists rather than serving as direct replacements for human therapists [5]. Building on this perspective, we propose the integration of an AI agent into the VR counseling experience as explored by Osimo et al [6] and Slater et al [7]. In this framework, participants embody 2 avatars: one representing themselves as a patient and the other as a counselor, allowing an "outside-looking-in" perspective that facilitates self-directed advice and reflection. However, sometimes self-conversation can become stuck, and the patient needs outside help. We contend that large language models (LLMs), while not yet fully capable of replacing human therapists, can function in this more limited role as an effective aid to help the conversation along, thereby enriching the self-talk process. This paper details the integration of an LLM-based agent into the VR self-talk paradigm. We outline the iterative design and development process, share the lessons learned, and present the first successful version, culminating in a small qualitative study in the context of motivational interviewing [8].

Self-Talk in VR

In times of crisis, individuals may find that their capacity to analyze issues and help themselves is limited. Their ability to gain insight is constrained, and it often feels like there is only one solution or, in more dire circumstances, "no way out." Interestingly, contemplating the problem as an onlooker or from a friend's perspective can improve problem-solving capabilities, a phenomenon known as Solomon's paradox [9]. The paradox states that people reason more wisely regarding other people's problems than they do about their own. Furthermore, a straightforward linguistic change from using "me or I" to "you" has also been shown to increase the psychological distance from personal problems, consequently alleviating the distress they cause and fostering cognitive flexibility [10]. Hence, a relatively straightforward shift in language and perception can enhance problem-solving capabilities.

Both findings are at the base of the VR self-conversation paradigm, first introduced by Osimo et al [6] (Figure 1). The virtual environment comprises a consultation room and 2 avatars. One avatar represents the participants as themselves, and the other represents the participants as the counselor. The participant is able to swap between the 2 avatars, thus enabling a unique experience of talking to a representation of oneself. The VR self-talk experience is based on the ability of participants to switch in and out of virtual bodies in VR. Multisensory integration between the virtual body and real body yields a strong illusion of ownership of the virtual body [7]. The virtual body moves with the participant's movements, and the participants can see their virtual body from the first-person perspective, and in a virtual mirror. The participants (as well as their avatars) are seated, and the illusion is based on the upper body—hand and head tracking (head is visible only in the mirror; Figure 2, top).

XSL•FO
RenderX

Figure 1. Screenshot of the counseling scenario interface. Top right inset shows a woman wearing a virtual reality headset (not a study participant). She is embodied as a female avatar reflected in a virtual mirror, with a same-sex counselor avatar seated across from her. A virtual control panel in front of her gives her instructions regarding how to proceed with the self-counseling experience and features an artificial intelligence help button.

The initial study compared 2 counselor avatars; 1 group interacted with an avatar resembling Sigmund Freud, while the other engaged with an additional avatar resembling themselves. The results indicated that using the self-talk paradigm proved beneficial in finding more satisfying solutions and improved mood, with the Freud group exhibiting better outcomes. Slater et al [7] showed that body swapping led to superior results compared to a prescripted and animated Freud virtual body and that body ownership of the Freud body was an important contributor.

Since then, these results have been replicated in multiple domains. VR self-talk has been shown to improve motivation for weight loss in individuals with obesity when they conversed with a future self who had lost and maintained weight over a hypothetical 5 years [11]. Convicted offenders spoke to their future selves, reducing self-defeating behaviors, such as alcohol consumption and violent behavior toward others [12]. An additional study showed that using VR self-talk when the "counselor" is an avatar of a leading athlete, such as Serena Williams or LeBron James, could improve adherence to an exercise regime (D Levy, unpublished data, June 2021). All these examples imply that using VR self-talk enhances cognitive flexibility, subsequently facilitating behavioral changes that previously appeared unattainable. Importantly, the design of such paradigms requires care and attention; a study using stereoscopic video revealed that besides the potential benefits of self-talk in VR, under some circumstances, the experience can also "backfire" for some specific populations [13].

Despite the success of VR self-talk, Doron Friedman, Prof Dr, and Mel Slater, Prof Dr, report (email, March 20, 2019) that

XSL•FO
RenderX

people occasionally tend "to get stuck" in the conversation and cannot provide a suitable solution or advice. Sometimes, participants want to end the session or do not know how to continue after a certain point. Often, people run out of advice from the counselor's point of view. This is where AI can play an important role. Using LLMs, we can provide individuals with the external help and ideas they need to continue making progress in their self-talk sessions. This use of AI can also be considered an intermediate step toward automated therapists.

Automated Dialogue

With advancements in machine learning, natural language processing, and deep neural networks, AI has become more capable of performing complex tasks and understanding human language [14]. One of the most notable developments is the emergence of conversational agents, such as ChatGPT, based on LLMs [15]. LLMs have seen tremendous advances in recent years, thanks to increased computational power and the availability of vast datasets for training. Models today, with many billions of parameters, can generate remarkably humanlike text and engage in dialogue while demonstrating some reasoning capabilities [16]. However, significant challenges remain in making these models more aligned with human values, interpreting instructions correctly, and generating factually accurate statements [17]. Current models may generate plausible but incorrect or nonsensical text. Thus, while LLMs today are impressive and valid for specific applications, they require close monitoring and oversight before being deployed into sensitive real-world settings, such as the clinical psychology domain [18].

Embodied AI applications have increasing relevance in mental health applications ranging from social robots to artificially intelligent virtual agents [19]. These applications aim to improve the quality of care and control expenditures [20]. In addition, they are also an important pathway through which the availability of therapeutic treatments may expand [21]. While AI mental health technologies continue to advance rapidly, a significant challenge remains in their successful implementation within clinical settings, where both practitioners and patients face barriers to adoption [22].

The psychological implications of the representation of the virtual therapist have been studied for decades [23], showing mixed results; sometimes, a human appearance enhanced the effectiveness of an application, while at other times, it did not. A meta-analysis by Weber et al [24] demonstrated that adding a face (as opposed to just voice or text) was more significant than the effect of realism; that is, there was greater gain in impact from having a face than from making that face more photographically or behaviorally realistic. Demeure et al [25] evaluated the impact of an agent's representation (in a nonimmersive environment) and perceived emotion over the perceived social believability in the agent. They found that appropriate emotions conveyed through the agent's body, mainly related to the sense of competence and warmth, could lead to higher believability.

LLMs can be the missing piece to the puzzle of VR in therapeutic mental health treatment. In this study, we examined whether LLMs can support participants in their self-talk when they run out of advice and understand how we can use an AI character to help people feel safe and not judged. We describe lessons learned during the development of this unique protocol, which involves a range of advanced techniques such as body ownership illusions [6], self-talk in VR, and LLM-based virtual humans [26]. Finally, we evaluate the final version of the system with a qualitative study.

Iterative Design and Pilot Studies: Lessons Learned

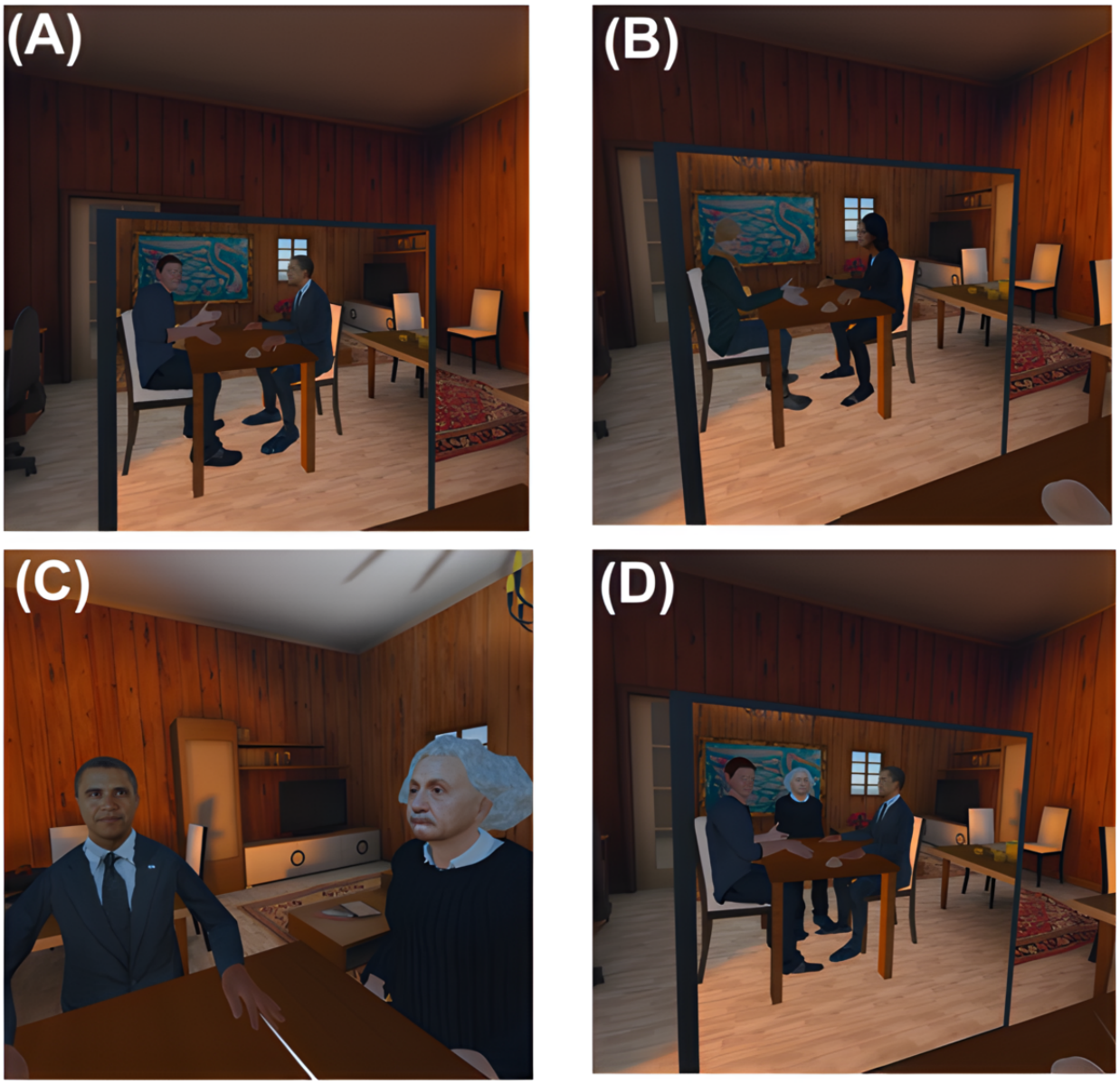
Integrating LLM into VR self-talk requires text-to-speech and speech-to-text functionalities. To address this, we used MILO [27]. The system transcribes participant voice responses from within the VR session to text, which then serves as input for the LLM. Once the LLM generates a response, the text is converted to speech and played through the Unity application.

Initially, the AI was represented as a "help" button with voice functionality (Figure 1). Participants were instructed that a virtual counselor was listening to their conversation, and they could receive its input if they felt stuck and did not know how to continue the self-conversation. We have evaluated this as part of a study comparing VR self-talk with a physical-world equivalent setup, the empty chair technique from Gestalt therapy [28]. However, of the 11 participants in this pilot study, only 1 (9%) engaged with AI. It turned out that a short psycho-education tutorial regarding basic emotion regulation skills followed by VR self-talk resulted in an overwhelming experience; all participants were highly engaged in the session and did not remember they could ask AI for "help." The minimal AI representation within the virtual environment, as only a button, can explain why it was ignored. In additional pilot evaluations, the combination of less-than-perfect speech recognition with less-than-perfect dialogue capabilities (refer to the information presented subsequently) often resulted in low-quality generated responses from the agent.

To make the AI agent more salient in the experience, we replaced the clickable button with a virtual human (Figure 2, bottom). This required addressing the repetition of texts; a by-product of the VR self-talk paradigm is that each text is repeated twice. First, it is spoken by the participant live and recorded, and next, it is played back by the corresponding avatar to be experienced by the participant after the body switch. When there was a third character in the scene, these repetitions became confusing. Consequently, we restricted access to the AI only from the "counselor" avatar and instructed participants that interaction with the AI avatar was recommended only between speech turns.

XSL•FO
RenderX

Figure 2. Screenshots from the counseling scenario. (A and B) Self-talk virtual reality (VR) as seen through the virtual mirror, featuring male and female avatars seated across from Barack Obama and Michelle Obama avatars, accordingly (both depicting the first VR experience without artificial intelligence). (C) Male participant's view, seated across from Barack Obama, with Einstein avatar representing artificial intelligence present, second VR experience. (D) Virtual mirror view showing male avatar, Barack Obama avatar, and Einstein avatar.



Furthermore, significant progress was made in LLMs over the 2 years of development. Initially, we used a specific model developed in the laboratory. This was based on the 7B parameters version of the GPT-J model [29], fine-tuned on the 2 volumes of published counseling and psychotherapy data from Alexander Street Press. The volumes are searchable collections of transcripts containing real counseling and therapy sessions and first-person narratives illuminating the experience of mental illness and treatment. The 2 volumes contain 3500 session transcripts and >700,000 utterances between a counselor and a patient. We fine-tuned the model with an 80% to 20% train-test split. While this model was state of the art at the time of the early sessions, rapid developments in LLMs rendered it obsolete; therefore, we replaced it with a pretrained model by OpenAI (described in Materials and Equipment in Methods section).

Our LLM interface provides an interface for a human operator, which can be used by the experimenters (refer to the study by Shoa et al [27] for full details). The ongoing automatically transcribed conversation appears in a text window. The operator can select parts of the text or even modify the text and send specific parts to the voice playback in the application. We have also explored the possibility of allowing the experimenter to decide when to intervene in the conversation.

To allow for a realistic experience, we designed and implemented the AI avatar to be gaze activated; that is, the virtual human representing the AI played an idle and silent animation loop and only spoke after the participant, while in the body of the counselor, stared at it for a duration of 1 second. Such gaze activation requires careful tuning—if the gaze duration is too short or the gaze area too broad, false positives

XSL•FO
RenderX

might occur, whereas otherwise, the activation becomes unnatural. In our postexperimental interviews, none of the participants complained about this type of activation.

Another challenge was response latency. The delay includes the following: (1) waiting for the participant to finish talking (indicated by clicking a controller button), (2) gaze activation of the AI avatar, (3) speech recognition in the cloud, (4) latency of one LLM predication, mostly depending on server configuration and context length, and (5) cloud-based text-to-speech conversion. To mitigate latency, the system can be used in a continuous mode; that is, the system is prompted to keep generating responses after each counselor utterance based on the ongoing conversation, regardless of whether the AI agent was activated or not.

Therefore, the consequence is that latency mostly depends on LLM prediction call latency. When using laboratory-based LLMs running from our own server, we could control the latency. The overall round trip would take several seconds, and this was acceptable in the context of a counseling experience. Using OpenAI resulted in depending on OpenAIs' response time. In our study, there were no major delays, and none of the participants in the postexperiment interview (as described subsequently) complained about latency.

Finally, we note that due to limitations in AI tools in many languages, we had to carry out the sessions in English rather than the local native language. Hence, the participants were selected to participate in the study based on their level of English speaking, ranging between conversational level English and native. Language support limitations include all parts of the pipeline: recognition, dialogue, and generation quality.

Methods

Overview

The system was iteratively refined and tested as described before until pilot studies indicated that it operated smoothly. Consequently, we carried out a qualitative explorative study.

Because the overall experience was very rich and overwhelming (as described subsequently), we opted for an exploratory study, carrying out careful sessions, including the presence of a trained clinical psychologist. Although it was made clear to the participants that the system was experimental and this was not a real counseling session, it was clear from the responses that participants behaved almost like in a "real" counseling session, and most participants went through a psychologically meaningful session, as evident from in-depth postexperiment interviews (refer to the Results section).

Participants

The study population comprised 14 anonymous university students (female participants: n=5, 36%) aged 19 to 26 years (mean 23.43, SD 2.21 years). Participants who self-declared as English speakers, had no history of epilepsy, and were not pregnant at the time of the study were invited to participate in the study as part of their credit requirements for their BA psychology degree. In total, 3 (21%) of the participants were not included in the final sample due to miscellaneous technical problems that came up during their session. Given the limited sample size, data saturation was not reached.

Materials and Equipment

Overview

The VR environment featured a consultation room. The VR simulation was developed with the Unity game engine (version 2020.3.20f; Unity Technologies). The VR headset Quest 2 (Meta Platforms, Inc), including its handheld controllers for embodiment (upper body tracking), was used. Quest 2 has a single liquid crystal display panel for each eye, with a display resolution of 1832×1920 pixels. The refresh rate of the panel is 120 Hz. Its weight is approximately 500 g, and it has a head strap that ensures comfort during prolonged use. In addition, Quest 2 delivers a comprehensive 6 dfs, providing participants with both rotational and positional tracking capabilities.

The software includes a VR self-talk experience and integration of an LLM-based agent based on MILO. In the exploratory study, we used Barack Obama and Michelle Obama, sex matched, as the counselor avatars. The AI avatar was Albert Einstein. The participants were initially embodied in a generic sex-matched avatar (Figure 2). The first part of the VR experience was intended to strengthen the virtual body ownership illusion. The participants engaged in a short "embodiment" exercise, alternately moving both hands, looking around, and seeing themselves in the mirror. Following the embodiment exercise, the participants were asked to describe a current life issue. When they finished, the participants pressed a button, transitioning them to the counselor avatar. Previous studies found that it is best to model the counselor avatar based on famous inspiring persons; we selected Barack Obama in our studies. Next, the participants pressed a button and listened to a replay of what they just said, spoken by the avatar representing themselves as patients. A prompt instructed them to respond "like a counselor," and their responses were recorded. The participants' voice pitches were changed in the recording to avoid having it sound exactly like the participant, as described in the study by Osimo et al [6]. Once completed, participants returned to their look-alike avatar to continue the conversation in this iterative manner.

We used 2 VR self-talk implementations; one was a commercial product called ConVRself (Kiin Tech) used during the iterative design, and the other was a research version with very similar design and functionality used during the study.

For LLM, we used GPT-3.5 using the following prompt: "You will now act as a motivational interviewer with a lot of experience in interpersonal psychotherapy. I am a counselor and will turn to you for advice while speaking to my patient. Sometimes, I will get stuck and address you, and in this case, you should try to act as an expert counselor and say something that would help me progress the session as best as possible. Please do not mention that you are an artificial intelligence in your replies. Act as a real person."

Semistructured Interviews

Participants' experiences and thoughts regarding the AI were assessed using semistructured interviews that lasted

XSL•FO
RenderX

approximately 15 minutes, and the audio of the interview was recorded. Due to the exploratory nature of this complex experience, we opted for the flexibility offered by semistructured interviews. We constructed 14 questions in advance, such as, "Please describe this experience as if you were describing it to a friend who was not here," (Multimedia Appendix 1). The questions were presented individually to the participants, and additional questions were added based on the participants' responses. The interviews were conducted by one of the main authors (MZ), a female clinical psychologist and PhD candidate, with appropriate training. There was no relationship between the interviewer and the participants before the study commencement, and the participants were informed of her professional expertise.

The actual conversation of the participants was logged by default as part of the system, including the AI comments. However, we opted not to analyze the conversations due to the privacy of the participants, and the content was removed.

Applied thematic analysis identifies and describes implicit and explicit ideas within the data, consequently linking them into the theoretical framework of the study [30]. The analysis was thus conducted using a bottom-up approach in which the answers provided by the participants determined the main themes rather than the questions asked. In addition, at the start of the interview, participants were asked to provide a description of the experience, to elicit unconstrained participant impressions.

Ethical Considerations

The ethical considerations regarding implementing AI in psychological settings are paramount [31]. The development process and exploratory study outlined in this paper aimed to address preliminary questions concerning effectiveness and validation. In addition, in our study, a clinical psychologist provided professional oversight, which is typically recommended for AI implementation. The study received ethics approval from the institutional review board of Reichman university (P_2022145). Participants received course credit for their participation in the study and gave written informed consent. All data protection procedures were adhered to. All participants provided informed consent, which explicitly outlined that their responses would be recorded and kept as part of the study data but would be saved anonymously and that they would interact with AI during the study, receiving advice from it regarding a personal problem.

Procedure

Upon arrival at the laboratory, the participants were given instructions about the experiment and completed a consent form. They were told that they would have the opportunity to talk about a personal problem that caused them an average level of distress (on a scale of 1-10, a problem rated between 4 and 7). Then, they wrote a short sentence describing the problem in their words. Importantly, although it was made clear to the participants that the session was experimental and should not be considered psychological counseling, our sessions were carried out in the presence of a trained clinical psychologist in a private laboratory setting in which only the participant and the psychologist were present. Next, the participants donned the VR headset and performed the VR experience in which they conversed freely between their sex-matched avatars and matching Obama characters (Michelle for female participants and Barack for male participants). After a back-and-forth conversation between the participant and Obama avatar, the participants chose when to end the session.

After the first session, the participants answered two questions regarding their experience as follows: (1) Please rate this experience on a scale of 1 to 10 (1=not useful at all and 10=very useful) and (2) On a scale of 1 to 10, how helpful was this self-conversation for solving your problem?

Next, it was explained to them that they would perform a similar session, only this time, another third avatar, depicted as Albert Einstein, would be seated in the virtual consultation room. They were told that it would be listening to their conversation and that they should turn to him during the conversation to receive his input but only from the embodied counselor's perspective. The gaze activation was explained, and they were instructed to turn to Einstein only from the Obama avatar. The participants proceeded to conduct the AI-enhanced iterative VR self-conversation, and when they were done, the semistructured interview was conducted.

Results

Overview

A total of 5 (45%) of the 11 participants chose to speak about a problem related to work or school. Another 5 (45%) participants chose to speak about relationship problems, and 1 (9%) participant chose to speak about difficulty managing stress.

The semistructured interviews were transcribed and analyzed using the thematic coding method by 2 of the coauthors. The transcripts were not reviewed by the participants. A total of 5 (45%) participants preferred the session with the AI, 4 (36%) preferred the session without it, and the remaining 2 (18%) did not address this. In total, 7 (64%) participants said that the AI's attitude toward them was positive, 1 (9%) said it was neutral, and the remaining 3 (27%) did not address this.

The participants were asked 2 questions in the break between the first session (without AI) and the second (with AI). On average, the usefulness of the session was rated 6.9 (SD 0.54); the median value was 7 (IQR 7-7). On average, the degree to which it helped solve their problem was rated 6.1 (SD 1.58), with a median value of 6 (IQR 6-7). At the end of the interview, the participants were asked to rate the extent to which they would like to come back for another session on a scale of 1 to 10; the mean response was 8.3 (SD 1.55), and the median was 8.5 (IQR 7-9), indicating a very high level of satisfaction from the experience.

The applied thematic analysis qualitative analysis revealed 4 high-level themes, which were derived from the data. These 4 themes are outlined subsequently. The participants did not give feedback on the findings outlined below.

Quality of Advice

From the thematic analysis of the participants' responses, a layered or complex pattern emerged regarding the perceived

XSL•FO
RenderX

quality of AI-generated advice. On the one hand, participants acknowledged the AI's ability to provide insightful, well-structured guidance:

> *...Einstein brought a lot of emotional considerations to the two-way problem I didn't think about. It was interesting...It sounds like if I would listen to him like everything would work out.* [S6]

> *For my questions it was very, very beneficial. It was. It was able to provide very thorough and clear and insightful points and information regarding specifically related to my problem related to the career path and choosing a career path. And it was really cool to see how the interaction with the computer algorithm was able to really contribute to my own self-reflection.* [S8]

They viewed the advice as emotional while sustaining a thorough analysis of the situation. However, sometimes this very sophistication created a feeling of uncanniness, with participants noting that it was "inhumanly smart." Some participants pointed toward the notable tension between the logical precision of AI and the inherently "messy" nature of emotional problems, suggesting that the AI's responses tended to be overly orderly or systematic:

> *I think that on the one hand it was good, and he was saying smart things, but on the other hand it was a little bit too logical and technical, and sometimes emotional problems are a little more complex.* [S4]

> *...Good as I said, but it was a little bit surface level.* [S7]

> *Like robotic smart. Like human, but inhuman smart. You know what I'm saying? Like he was too good of a psychologist. I don't know how to say it. It felt like a nice thing to say, but feels like I don't know how to describe it too, like, kitsch...The AI surprised me...Yeah, the AI was kind of freaking yeah. Like it felt. It felt like too good. I don't know how even to describe it...Yeah, but also, I mean it was like in between. It was like really compassionate and understanding and like what the perfect answer from a psychologist should be. But the perfect answer is like not, not human to some degree. It also felt like he just blabbered a bunch of information, like instantly to me, like, Oh yeah, it's OK to be compassionate and it's OK to be. You shouldn't be perfect all the time, like, in one sentence.* [S13]

There was also evidence pointing to potential overtrust in AI responses, which has been mentioned before [32]. This might suggest that careful consideration of AI authority is necessary in therapeutic settings due to increased credibility attributed to AI-reorganized versions of the participants' own thoughts:

> *I think he just had a way of wording the same ideas that I had in my head, but sort of organizing them to a more sophisticated or organized manner which made them more believable or reliable, seem more reliable.* [S11]

In addition, some of the "complaints" could also be a reasonable approach to a human counselor:

> *S6: ...that it was not really practical, but it was like another part that you need to think about.*
>
> *Question: What wasn't practical about it?*
>
> *S6: Like for my problem. I need an answer, it's like a yes or no question and he really talked about like what it made you feel and maybe rephrase the problem, and there's no way to really rephrase it.*

Self-AI Collaboration

In some cases, participants highlighted the complementary nature of the combination between the AI advice and their own input, suggesting a unique form of alliance where AI and human intelligence united and enhanced each other's strengths. The AI's tendency to a more structured approach was balanced effectively with the more emotionally attuned self-counsel, creating what one participant even termed as a "perfect" mixture. This combination led to a more comprehensive experience. As mentioned, rather than viewing AI as a replacement for traditional therapy, its optimal role, as the participants experienced, was as a tool that enhanced their self-reflection and personal insight:

> *...I liked the mix of the AI and myself as the counselor...I think the advice I gave myself was taking into account more like an emotional feeling and maybe even a little bit too much. And he was a little too technical. So maybe a mixture of them? Yeah, a mixture of them would be perfect.* [S4]

> *I think in terms of advice, it was similar to what I said, although the AI was able to provide a more detailed and well-structured answer. The AI's examples and ideas were a bit more beneficial.* [S8]

> *And I liked the mix of him and Michelle together, that I, as Michelle gave a bit of a softer input and he was like, do this and this and this and this and that together was like, they filled each other...No, I think they were really complementary to each other. I benefited very much from both of them. I'm really grateful. I got to talk to myself...OK, more minds better.* [S9]

Believability and Attitude Toward AI

The analysis suggested a notable dichotomy between the cognitive aspect and the perceptual issues of the experience. While the content and reasoning capabilities of the AI were valuable and convincingly humanlike, the nonverbal and visual elements created "breaks in presence" [33]. These elements were mainly manifested in voice quality, robotic movements, and lack of natural pauses in speech turns:

> *...yeah, the sound and obviously the, the graphics obviously. But in terms of the syntax and the structure of how the answers were provided and delivered then it was fairly good. Fairly satisfying.* [S8]

> *I feel like a mixture, like the words that he was saying felt like they made sense, they were human-like, but*

XSL•FO
RenderX

> *the way that he was telling them and like it was very technical...* [S4]
>
> *The conversation felt like very human and Einstein Avatar was a little bit robotic and felt like scripted and just a little bit like not related...No, the look was OK.* [S5]
>
> *I'm asking a question and then we'll just speak like we talk and talk and talk and talk...Humans usually wait a little bit.* [S7]
>
> *Technical stuff, which is like the character will move a little bit differently than human, and that's pretty much.* [S7]

However, a phenomenon of gradual acceptance regarding the participants' ability to adapt to the limitations of technology aligned with established research on place illusion in immersive VR environments [34]. In other words, the participants wanted to believe what they were experiencing in such a way that they were willing to ignore the technical imperfections that created disbelief. This process of adaptation was evident in the participants' ability to shift their focus from the robotic characteristics to the contextual value of the dialogue. In this way, the content served as a compensatory factor, allowing participants to overcome initial reservations regarding the artificial nature of the interaction. While technical refinements and improvements in nonverbal behavior are called for, the therapeutic value of the AI-advice might be more dependent on the quality of the dialogue rather than the behavioral human mimicry:

> *I think it's still not 100% human like but it did feel like I was talking to a very insightful person because at the end of the day I was I was focusing more on the information that was given to me by the algorithm rather than the actual feeling and and the the sound of the the interaction itself.* [S8]
>
> *...at once I got over the metallic voice, yeah, it felt a bit more natural...It's like a fake character. It's an avatar, but I think when you get in, in into it, it's become less and less weird. Less and less, Yeah.* [S9]

The Use of Celebrity Avatars

The data suggest that the selection of well-known figures as therapeutic avatars creates complex psychological dynamics that can either enhance or impede the experience. Participants' reactions demonstrated that the personal associations with these celebrity avatars substantially influenced their engagement with the AI and with their own counselor persona. For instance, responses to the Obama avatar highlighted the importance of perceived authority and strength, while reactions to Einstein often centered on his scientific reputation and intellectual credibility. However, these preexisting associations could also act as barriers, with some participants expressing a preference for avatars they could "relate to and love more." This finding suggests that the effectiveness of celebrity avatars in therapeutic settings may be highly individualized, with the same figure potentially eliciting different responses based on personal preconceptions. The data particularly emphasized how the avatar's identity can affect the reception of therapeutic advice, with some participants explicitly attributing the perceived value of the guidance to the celebrity's reputation rather than the content itself:

> *And I think Obama is quite a good character to use, actually. I feel that she's confident. She's a woman, I guess like a strong woman. She's very powerful. It seems like it's very fun to sit around her and like look at her and talk to her.* [S5]
>
> *That's the point, Talking to Obama, even if, like, I don't agree with him that much. But like, yeah, it's a cool person to talk to.* [S7]
>
> *I think I would change Barack Obama's character to someone that I love. As I said, one of the reasons I felt more comfortable with Einstein is because that's a character I can relate to and love more.* [S10]
>
> *Again, maybe it's just because it's Einstein. I mean if like a human being would say something like that to me but would phrase it differently and, you know, have different facial expressions like not like the computer Einstein, then I would feel differently about the answer.* [S13]
>
> *...Maybe because it was Einstein. Like maybe if it was someone else maybe I wouldn't think about like it was but just because it was him and I know that he was like really big scientist so he might have like great things to say.* [S14]

Discussion

Principal Findings

Mental health and well-being are major challenges worldwide, and the demand for counseling is much greater than the supply. Mental health systems worldwide are under immense strain, and in the face of growing need, many underqualified or pseudoprofessional services have appeared. As a result, an increasing number of stakeholders are turning to AI-based solutions, including LLM-driven tools, as a potential quick fix (this includes Israel's minister of science [January 3, 2025, personal communication]). However, we advocate a more measured approach, emphasizing rigorous evaluation and clear delineation of clinical versus nonclinical use. Our intervention explicitly targets nonclinical populations (ie, focusing on mental health improvement rather than formal therapy) and relies on AI only as an adjunct to self-talk rather than a stand-alone therapeutic agent. Such caution is vital given the nascent state of these technologies and the ethical implications tied to safety, efficacy, and user well-being. Accordingly, we strongly support further research and stepwise validation, coupled with the ongoing involvement of trained professionals before any large-scale implementation.

Technologies, such as VR and AI, may be a part of the solution. VR self-talk is unique in suggesting counseling without the need of either a human in the loop or AI and has been shown in the past to be beneficial and effective. Adding an LLM assistant is a natural next step; nevertheless, introducing an LLM-based AI counselor assistant into VR is not trivial, and

XSL•FO
RenderX

our iterative design process has revealed several important lessons.

Integrating multiple challenging technologies, such as VR, body tracking, dialogue, voice recognition, and speech generation into a meaningful psychological experience proved arduous. While each of these technologies separately has made impressive progress recently, accumulated problems or errors easily resulted in an unusable experience. In the case described here, it seems that during 2023, the underlying AI technologies, specifically speech recognition and LLM-based dialogue, passed an essential threshold, increasing the probability of a successful and meaningful experience regardless of the complex nature of the technological design.

While both VR and AI, individually and in combination, present significant opportunities for counseling, their widespread deployment in real-world settings remains challenging [35]. The contribution of each of these components as well as their combined contribution needs to be carefully evaluated [36]. Industry trends indicate that integrating AI models into a wide range of applications is becoming viable, for example, applications such as Tess and Woebot [37]. VR is gradually becoming more accessible, though its availability remains somewhat limited. We propose that the experience of "interacting with yourself in the third person," facilitated by VR and virtual body–ownership illusions, offers unique benefits that are unlikely to be replicated on 2D screens. In addition, the integration of LLMs into text and voice-based interactions is expected to become more widely accessible. Therefore, we hope that this study, despite its challenges and complexities, can serve as a foundation for further advancements in this field.

On the basis of the themes that emerged from the semistructured interviews, we can conclude that participants generally responded positively to the AI advice, with reactions ranging from mixed to favorable. None of the participants evaluated it as poor or as obstructive to the self-talk process. Considering the complexity of the experience, this result is very encouraging. VR self-talk, when performed correctly, is a very powerful experience, though it can be confusing even without the addition of AI. To address this, we implemented a gradual protocol, allowing the participants to accustom themselves to the VR self-talk before introducing the AI agent in a subsequent session. This protocol was necessary for our preliminary research. However, future studies could benefit from a research design such as randomized controlled trials that could compare self-talk with and without AI, providing a deeper understanding of the intricate nature of this experience and the added value of AI to it.

The overarching goal of this endeavor was to design and implement an AI agent within a self-talk experience. We envisioned the AI advice as complementing the participants' own self-guidance rather than replacing their need to think, reflect, and reason with themselves regarding the issue at hand. This vision was in alignment with the participants' responses. They found it straightforward to implement the AI advice into their self-dialogue and quickly adapted to the flow of the conversation. Our results further support the perspective that the integration of technology into counseling has the potential to enhance skills and abilities that humans already possess [16].

In our case, an "AI" button was not enough, and we opted to integrate a gaze-activated avatar to embody the AI. While participants commented that the voice and animation of the avatar were not completely realistic, several of the participants indicated that they were able to overcome such limitations and focus on the social interaction as well as the content of the conversation.

Finally, most of the participants pointed to the need to improve design elements, such as voice, body language, avatar appearance, and character selection. There was a relative consensus among participants regarding the need for improvements in this area to enhance the overall experience. This is not surprising as the level of realism has been shown to have a strong impact on affective responses of participants [26,27]. Now that we have successfully integrated multiple technologies and created a seamless psychological experience, future development and testing should focus on refining and improving these design aspects.

Limitations

This study has several limitations that should be addressed. Although multiple pilot studies were conducted throughout the iterative design process, including approximately 30 additional participants, the final sample consisted of only 11 participants. The relatively small sample size underscores the necessity for further research with larger and more diverse populations that could confirm these initial results. However, as emphasized throughout this paper, this is a novel and promising application of technology, and every exploration must begin somewhere. Due to the psychological focus of this study, further research should explore specific participant characteristics that could influence the effectiveness of self-talk in VR. Furthermore, research could target particular symptoms of psychopathology or specific diagnoses to improve the understanding of potential clinical applications.

Another limitation is the bias of novelty. The use of VR and AI in a psychological setting could have potentially fascinated the participants, leading to an inflated sense of efficacy and distorting the true impact of the experience. In addition, semistructured interviews could have contributed to social desirability and acquiescence bias, as participants might have felt inclined to provide responses they expected would be favorable to the researcher. Further research could explore ways to mitigate these identified biases.

Technological limitations are the lack of systematic analysis of the model's responses and their adherence to the therapeutic principles outlined in the prompt. The quality of the model's response is influenced by both the prompt and the internal architectures of the LLM itself, in our case GPT-3.5, which has proven highly capable but is proprietary and continuously evolving. The initial study used a relatively simple prompt, instructing the AI to adopt the role of an experienced motivational or interpersonal psychotherapist. While this approach was based on established research [38], it did not undergo formal pilot testing specifically targeting the fidelity

XSL•FO
RenderX

of the model's alignment with the therapeutic technique. This decision was made, in part, due to the exploratory nature of this study, which aimed to assess the potential use of LLMs in VR-based self-counseling rather than conduct an in-depth evaluation of their adherence to specific therapeutic techniques. Nonetheless, we acknowledge that future work would benefit from piloting more refined prompts, including detailed instructions for eliciting key therapeutic elements.

The field faces significant challenges in ensuring replicability and consistency, particularly when relying on such models. As a result, the research community must consider how to establish more robust methods for quality control and model evaluation to ensure the ongoing utility and ethical deployment of LLMs in therapeutic settings. Recommendations include clinical evaluation of the LLMs developed and their output; interdisciplinary collaboration; and attending to risk assessment, transparency, and bias [36]. We were able to implement some of the recommendations in our initial study, such as clinical evaluation of output and interdisciplinary collaboration, and further studies could improve on others.

Conclusions

The introduction of new technologies could potentially transform psychotherapy, giving rise to numerous potential challenges, limitations, and ethical considerations that should be addressed. Currently, clinical expertise is necessary for best practice; however, that may evolve in the future, and even now, monitoring and validating can be more cost-effective for certain patients compared to traditional psychotherapy. Given our encouraging results, we suggest that the paradigm of AI-enhanced VR self-talk may be ready for further research on a larger sample from the general population and studies with clinical populations. In addition, our work suggests there are numerous potential opportunities for integrating AI into VR wellness, extending beyond just "automated therapist agents."

Acknowledgments

The authors thank Peleg Shinar, bachelor of arts, humanities and social sciences research assistant, who assisted during the iterative design process. This research was supported by the Horizon 2020 Information and Communication Technologies (ICT) project - Leadership in Enabling and Industrial Technologies - SOCRATES (#951930) and European Union's Horizon 2020 program H2020-FETPROACT-2020-2 Research and Innovation - GuestXR (#101017884). MS is supported by the ERC Advanced Grant MoTIVE (#742989).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the confidential nature of the content, but numerical results will be made available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Complete interview protocol used in this study.
[DOCX File , 17 KB - formative_v9i1e67782_app1.docx]

References

1. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2197-2223 [FREE Full text] [doi: 10.1016/S0140-6736(12)61689-4] [Medline: 23245608]
2. Trautmann S, Rehm J, Wittchen HU. The economic costs of mental disorders: do our societies react appropriately to the burden of mental disorders? EMBO Rep 2016 Sep;17(9):1245-1249 [FREE Full text] [doi: 10.15252/embr.201642951] [Medline: 27491723]
3. Oladeji BD, Gureje O. Brain drain: a challenge to global mental health. BJPsych Int 2016 Aug;13(3):61-63 [FREE Full text] [doi: 10.1192/s2056474000001240] [Medline: 29093905]
4. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
5. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct 09;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]
6. Osimo SA, Pizarro R, Spanlang B, Slater M. Conversations between self and self as Sigmund Freud--a virtual body ownership paradigm for self counselling. Sci Rep 2015 Sep 10;5:13899 [FREE Full text] [doi: 10.1038/srep13899] [Medline: 26354311]

XSL•FO
RenderX

7. Slater M, Neyret S, Johnston T, Iruretagoyena G, Crespo M, Alabèrnia-Segura M, et al. An experimental study of a virtual reality counselling paradigm using embodied self-dialogue. Sci Rep 2019 Jul 29;9(1):10903 [FREE Full text] [doi: 10.1038/s41598-019-46877-3] [Medline: 31358846]
8. Hettema J, Steele J, Miller WR. Motivational interviewing. Annu Rev Clin Psychol 2005;1:91-111 [FREE Full text] [doi: 10.1146/annurev.clinpsy.1.102803.143833] [Medline: 17716083]
9. Grossmann IE, Kross E. Exploring Solomon's paradox: self-distancing eliminates the self-other asymmetry in wise reasoning about close relationships in younger and older adults. Psychol Sci 2014 Aug;25(8):1571-1580 [FREE Full text] [doi: 10.1177/0956797614535400] [Medline: 24916084]
10. Kross E, Ayduk O. Self-distancing: theory, research, and current directions. In: Advances in Experimental Social Psychology. New York, NY: Academic Press; 2017.
11. Anastasiadou D, Herrero P, Vázquez-De Sebastián J, Garcia-Royo P, Spanlang B, Álvarez de la Campa E, et al. Virtual self-conversation using motivational interviewing techniques to promote healthy eating and physical activity: a usability study. Front Psychiatry 2023;14:999656 [FREE Full text] [doi: 10.3389/fpsyt.2023.999656] [Medline: 37151962]
12. van Gelder JL, Cornet LJ, Zwalua NP, Mertens EC, van der Schalk J. Interaction with the future self in virtual reality reduces self-defeating behavior in a sample of convicted offenders. Sci Rep 2022 Feb 10;12(1):2254 [FREE Full text] [doi: 10.1038/s41598-022-06305-5] [Medline: 35145177]
13. Landau D, Hasler B, Golland Y, Huebbe B, Idan O, Magnat M, et al. A self-compassion experience in immersive video: opportunities and pitfalls. PsyArXiv Preprint posted online on November 28, 2022 [FREE Full text] [doi: 10.31234/osf.io/z4q9f]
14. Wei J, Tay Y, Bommasani R, Raffel C, Zoph B, Borgeaud S, et al. Emergent abilities of large language models. arXiv Preprint posted online on June 15, 2022 [FREE Full text]
15. Black S, Biderman S, Hallahan E, Anthony Q, Gao L, Golding L, et al. GPT-NeoX-20B: an open-source autoregressive language model. arXiv Preprint posted online on April 14, 2022 [FREE Full text]
16. Wei J, Wang X, Schuurmans D, Bosma M, Ichter B, Xia F, et al. Chain-of-thought prompting elicits reasoning in large language models. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. 2022 Presented at: NIPS'22; November 28-December 9, 2022; New Orleans, LA.
17. Menick J, Trebacz M, Mikulik V, Aslanides J, Song F, Chadwick M, et al. Teaching language models to support answers with verified quotes. arXiv Preprint posted online on March 21, 2022 [FREE Full text]
18. Hagendorff T. Machine psychology: Investigating emergent capabilities and behavior in large language models using psychological methods. arXiv Preprint posted online on March 24, 2023 [FREE Full text]
19. Fiske A, Henningsen P, Buyx A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. J Med Internet Res 2019 May 09;21(5):e13216 [FREE Full text] [doi: 10.2196/13216] [Medline: 31094356]
20. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. J Med Internet Res 2020 Jul 13;22(7):e16021. [doi: 10.2196/16021] [Medline: 32673216]
21. Holohan M, Buyx A, Fiske A. Staying curious with conversational AI in psychotherapy. Am J Bioeth 2023 May;23(5):14-16 [FREE Full text] [doi: 10.1080/15265161.2023.2191059] [Medline: 37130403]
22. Holmes EA, Ghaderi A, Harmer CJ, Ramchandani PG, Cuijpers P, Morrison AP, et al. The Lancet Psychiatry Commission on psychological treatments research in tomorrow's science. Lancet Psychiatry 2018 Mar;5(3):237-286 [FREE Full text] [doi: 10.1016/s2215-0366(17)30513-8]
23. Cresswell K, Cunningham-Burley S, Sheikh A. Health care robotics: qualitative exploration of key challenges and future directions. J Med Internet Res 2018 Jul 04;20(7):e10410 [FREE Full text] [doi: 10.2196/10410] [Medline: 29973336]
24. Weber S, Weibel D, Mast FW. How to get there when you are there already? Defining presence in virtual reality and the importance of perceived realism. Front Psychol 2021;12:628298 [FREE Full text] [doi: 10.3389/fpsyg.2021.628298] [Medline: 34025504]
25. Demeure V, Niewiadomski R, Pelachaud C. How is believability of a virtual agent related to warmth, competence, personification, and embodiment? Presence Teleoperators Virtual Environ 2011;20(5):431-448 [FREE Full text] [doi: 10.1162/PRES_a_00065]
26. Slater M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philos Trans R Soc Lond B Biol Sci 2009 Dec 12;364(1535):3549-3557 [FREE Full text] [doi: 10.1098/rstb.2009.0138] [Medline: 19884149]
27. Shoa A, Oliva R, Slater M, Friedman D. Sushi with Einstein: enhancing hybrid live events with LLM-based virtual humans. In: Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents. 2023 Presented at: IVA '23; September 19-22, 2023; Würzburg, Germany URL: https://doi.org/10.1145/3570945.3607317 [doi: 10.1145/3570945.3607317]
28. Zisquit M, Klomek AB, Shoa A, Slater M, Spanlang B, Friedman D. Intrapersonal emotion regulation training in virtual reality: embodying self-talk. Virtual Real 2025 (forthcoming).

XSL•FO
RenderX

29. Bell IH, Nicholas J, Alvarez-Jimenez M, Thompson A, Valmaggia L. Virtual reality as a clinical tool in mental health research and practice . Dialogues Clin Neurosci 2022 Apr 01;22(2):169-177 [FREE Full text] [doi: 10.31887/dcns.2020.22.2/lvalmaggia]
30. Guest G, MacQueen KM, Namey EE. Applied Thematic Analysis. Thousand Oaks, CA: SAGE Publications; 2012.
31. Imel ZE, Caperton DD, Tanana M, Atkins DC. Technology-enhanced human interaction in psychotherapy. J Couns Psychol 2017 Jul;64(4):385-393 [FREE Full text] [doi: 10.1037/cou0000213] [Medline: 28318277]
32. Hitron T, Morag NM, Erel H. Implications of AI bias in HRI: risks (and opportunities) when interacting with a biased robot. In: Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction. 2023 Presented at: HRI '23; March 13-16, 2023; Stockholm, Sweden URL: https://doi.org/10.1145/3568162.3576977 [doi: 10.1145/3568162.3576977]
33. Garau M, Widenfeld HR, Antley A, Friedman D, Brogni A, Slater M. Temporal and spatial variations in presence: a qualitative analysis. In: Proceedings of the 7th Annual International Workshop on Presence. 2004 Presented at: PRESENCE 2004; October 13-15, 2004; Valencia, Spain. [doi: 10.1162/pres.17.3.293]
34. Spanlang B, Normand J, Borland D, Kilteni K, Giannopoulos E, PomÃ©s A, et al. How to build an embodiment lab: achieving body representation illusions in virtual reality. Front Robot AI 2014 Nov 27;1 [FREE Full text] [doi: 10.3389/frobt.2014.00009]
35. Bell IH, Pot-Kolder R, Rizzo A, Rus-Calafell M, Cardi V, Cella M, et al. Advances in the use of virtual reality to treat mental health conditions. Nat Rev Psychol 2024 Jul 10;3(8):552-567 [FREE Full text] [doi: 10.1038/s44159-024-00334-9]
36. Stade EC, Stirman SW, Ungar LH, Boland CL, Schwartz HA, Yaden DB, et al. Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. Npj Ment Health Res 2024 Apr 02;3(1):12 [FREE Full text] [doi: 10.1038/s44184-024-00056-z] [Medline: 38609507]
37. Stephens TN, Joerin A, Rauws M, Werk LN. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. Transl Behav Med 2019 May 16;9(3):440-447 [FREE Full text] [doi: 10.1093/tbm/ibz043] [Medline: 31094445]
38. Reynolds L, McDonell K. Prompt programming for large language models: beyond the few-shot paradigm. arXiv Preprint posted online on February 15, 2021 [FREE Full text] [doi: 10.1145/3411763.3451760]

Abbreviations

AI: artificial intelligence
LLM: large language model
VR: virtual reality

Edited by A Mavragani; submitted 21.10.24; peer-reviewed by F Shojaei, B Meyer; comments to author 14.12.24; revised version received 18.01.25; accepted 10.02.25; published 02.04.25.

Please cite as:
Zisquit , Shoa A, Oliva R, Perry S, Spanlang B, Brunstein Klomek A, Slater M, Friedman D
AI-Enhanced Virtual Reality Self-Talk for Psychological Counseling: Formative Qualitative Study
JMIR Form Res 2025;9:e67782
URL: https://formative.jmir.org/2025/1/e67782
doi:10.2196/67782
PMID:

© Moreah Zisquit, Alon Shoa, Ramon Oliva, Stav Perry, Bernhard Spanlang, Anat Brunstein Klomek, Mel Slater, Doron Friedman. Originally published in JMIR Formative Research (https://formative.jmir.org), 02.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Factors Impacting Mobile Health Adoption for Depression Care and Support by Adolescent Mothers in Nigeria: Preliminary Focus Group Study

Lola Kola[1,2], PhD; Tobi Fatodu[1], MPH; Manasseh Kola[3], BEng; Bisola A Olayemi[1], MBBS; Adeyinka O Adefolarin[1,4], PhD; Simpa Dania[5], MBBS; Manasi Kumar[6], PhD; Dror Ben-Zeev[7], PhD

[1]WHO Collaborating Centre for Research and Training in Mental Health, Neurosciences and Drug and Alcohol Abuse, Department of Psychiatry, College of Medicine, University of Ibadan, Ibadan, Nigeria

[2]Centre for Global Mental Health, Kings College London, London, United Kingdom

[3]School of Computer Science, University of Leicester, Leicester, United Kingdom

[4]Department of Health Promotion and Education, College of Medicine, Faculty of Public Health, University of Ibadan, Ibadan, Nigeria

[5]Association for Reproductive and Family Health (ARFH), Ibadan, Nigeria

[6]Department of Population Health, School of Medicine, Institute for Excellence in Health Equity, New York University, New York, NY, United States

[7]Department of Psychiatry and Behavioral Sciences, BRiTE Center, University of Washington, Seattle, WA, United States

Corresponding Author:
Lola Kola, PhD
WHO Collaborating Centre for Research and Training in Mental Health, Neurosciences and Drug and Alcohol Abuse, Department of Psychiatry
College of Medicine
University of Ibadan
P.M.B 3017 G.P.O
Ibadan, 23401
Nigeria
Phone: 234 (2) 2411768
Email: lola_kola2004@yahoo.com

Abstract

Background: Mobile health (mHealth), the use of mobile technology in health care, is increasingly being used for mental health service delivery even in low- and middle-income countries to scale up treatment, and a variety of evidence supports their potential in different populations.

Objective: This study aims to use the Social Cognitive Theory (SCT) as a lens to explain knowledge of mHealth use for mental health care, personal behavioral capabilities, and the external social contexts that can impact the adoption of an mHealth app for depression care among perinatal adolescents in Nigeria.

Methods: At the preliminary stage of a user-centered design (UCD), 4 focus group discussions were conducted among 39 participants: 19 perinatal adolescents with a history of depression and 20 primary care providers. Guided by the SCT, a popular model used for predicting and explaining health behaviors, we documented participants' knowledge of mHealth use for health purposes, advantages, and challenges to the adoption of an mHealth app by young mothers, and approaches to mitigate challenges. Data collection and analysis was an iterative process until saturation of all topic areas was reached.

Results: The mean age for young mothers was 17.3 (SD 0.9) years and 48 (SD 5.8) years for care providers. Mistrust from relatives on mobile phone use for therapeutic purposes, avoidance of clinic appointments, and sharing of application contents with friends were some challenges to adoption identified in the study population. Supportive personal factors and expressions of self-efficacy on mobile app use were found to be insufficient for adoption. This is because there are social complications and disapprovals that come along with getting pregnant at a young age. Adequate engagement of parents, guardians, and partners on mHealth solutions by care providers was identified as necessary to the uptake of digital tools for mental health care in this population.

Conclusions: The SCT guided the interpretations of the study findings. Young mothers expressed excitement at the use of mHealth technology to manage perinatal depression. Real-life challenges, however, need to be attended to for successful

XSL•FO
RenderX

implementation of such interventions. Communications between care providers and patients' relatives on the therapeutic use of mHealth are vital to the success of a mHealth mental health management plan for depression in young mothers in Nigeria.

(JMIR Form Res 2025;9:e42406) doi:10.2196/42406

KEYWORDS

adolescent perinatal depression; primary care; mHealth app; user centered design; smartphone; human-centered design; HCD; depression; postpartum; perinatal; postnatal; teenage; adolescent; youth; low-middle-income countries; LMIC; middle income; adoption; acceptability; mobile health; mHealth; mobile app; women's health; mental health; depressive

Introduction

The World Health Organization (WHO) strongly recommends the rapid widespread use of technology as a possible solution for transforming pressing public health services [1]. Mobile Health (mHealth), the use of mobile technology in health care, is increasingly being used for mental health service delivery even in low- and middle-income countries to scale up treatment, and a variety of evidence supports their potential in different populations [2]. Mobile phones are ubiquitous globally and in low-income countries with evidence of use in different population groups [3]. Recent reports indicate their high use among the youth perinatal population in Nigeria [4]. In Nigeria, digital research has been successfully conducted for perinatal mental care within the WHO Mental Health Global Action Programme (mhGAP) task-shifting initiative in primary care [4-6] to address barriers to care in patients. Task shifting, which is a process of delegating tasks to less specialized health care workers where appropriate, is recommended by the WHO to bridge the treatment gap in global mental health [7]. Digital health care presents avenues to further scale-up care in the process of task shifting [1]. Technology-based studies for perinatal care in Nigeria have been implemented to increase pregnant women's access to care, but several knowledge gaps exist on factors that influence the use and adoption of mHealth in adolescent mothers [8,9]. Addressing these gaps is crucial for increasing access in a vulnerable population. Perinatal adolescents, often unmarried, might have their pregnancies disputed by the father of their child and could also have little or no social support from family members [8,9] necessitating the need for age and context-specific health care for this vulnerable population.

Perinatal depression, a condition common in pregnant women, is higher among adolescent mothers than older mothers [10,11] occurring during pregnancy and up to one year after childbirth. Untreated perinatal depression is a risk for negative health outcomes for mothers and their infants [12]. There are limited resources for managing mental disorders in primary care in low-income countries [13] and access for young mothers is even more limited by multiple age-related access barriers, which often translate to self and social stigma [8]. The complexities of hindrances to care experienced by this population indicate the need for practical and flexible interventions tailored to meet their unique needs [14].

Mobile phone apps, which are software programs designed to work on mobile devices, can serve as mHealth tools used in real-time for mental health care [15,16]. Mobile phone apps offer easy and affordable digital access to mental health information and solutions and can be used to increase internet-based access to mental health services. Evidence supports the involvement of users in the designing and developing of technology-based interventions to increase their potential for success and uptake [17]. User-centered designs [18] allow users to participate in iterative developments and design processes of mHealth tools, for reflection of users' needs in treatment designs. The use of mHealth even in low-resource settings [2] and the willingness of young mothers to engage in mobile phone use for health care services [4] present several opportunities for the design of technology-based interventions that attend to their unique needs [19].

In this study, we explored the benefits, barriers, and mitigating factors to mHealth adoption for depression by adolescent mothers in Nigeria through a theoretical lens. Theories play a crucial role in explaining health behaviors, and in digital health care they have provided frameworks for predicting and influencing human behaviors in the context of health and well-being. [20,21]. In this study, however, we used the Social Cognitive Theory (SCT), a widely recognized theory of behavior change [22] in explaining and understanding participants' knowledge of mHealth use for mental health care (personal experiences), personal behavioral capabilities of adolescent mothers to engage with a mobile phone app for depression care (behavioral factors), and external social context (environmental factors) that may impact their expected health outcomes [23].

Focus group discussions (FGDs) were carried out with adolescent mothers as part of the second stage of a user-centered design (UCD) of a mHealth app that was designed to complement face-to-face care for perinatal depression within routine primary care. The question guide explored participants' knowledge of mHealth use, the perceived advantages and challenges of application use, and approaches to mitigate identified challenges. The research questions were as follows:

1. What is the knowledge of mHealth use for mental health care among young mothers and care providers in primary care?
2. What are the potential barriers and facilitators of mobile app adoption in this population? What challenges could limit its use?
3. What are the possible strategies to mitigate the challenges to mHealth apps adoption by young mothers?

Methods

Participants and Study Setting

The study was conducted in Ibadan, southwest Nigeria using FGDs: 2 FGDs of young mothers (n=19), and 2 separate ones

XSL•FO
RenderX

among care providers (n=20). The venue was a seminar room at the Department of Psychiatry, University of Ibadan that was chosen because it was private and easily accessible using public transportation. All 4 FGDs took place in a 2-week period in February 2020. Participants were recruited from the database of an ongoing trial on perinatal depression in primary care [24]. A total of 30 participants aged between 16 and 19 years had completed treatment in an ongoing trial (Responding to the challenge of Adolescent Perinatal Depression [RAPiD]) [24], and those in remission from clinical depression as indicated by their hospital records were selected. Only 22 of them had phone numbers that rang through on the mobile network. The phone numbers of the remaining 8 were no longer available on the mobile networks. Even though all 22 young mothers who were contacted by phone indicated their availability for the FGDs, 3 participants later called to indicate their unavailability on the days chosen for the discussions, leaving 19 participants who attended the FGDs. On the part of the care providers, 20 primary care providers from the same trial's intervention arm were invited, and all accepted and participated in the FGDs. All participants were briefed on the purpose of the FGDs when they were contacted on the phone. The young mothers and the clinicians were selected from clinics in the 11 local government areas in Ibadan. On the days of the FGDs, the facilitator gave additional detailed information on the purpose of the FGDs. All participants signed an informed consent form after giving verbal consent to participate in the study. In total, 3 groups had 10 participants each, while the fourth (a young mother's group) had 9 participants.

FGD Procedure

The average duration for the FGDs was 90 minutes. Author LK, a medical sociologist, moderated all 4 groups using predetermined open-ended questions. Author TF was the timekeeper and note-taker in all the discussions. Both LK and DA had no previous contact or interactions with study participants. Following the introductions, the facilitator asked the initial question on knowledge of mobile phone use for health purposes. Participants who indicated such knowledge were asked to describe the interventions they were aware of. The facilitator probed on the mode of delivery of such interventions identified, the information content, and the disease conditions they were used for. Participants were then asked about their familiarity with mobile phone apps, and irrespective of the answer given, the facilitator showed all participants a video-based mobile phone app on an Infinix Hot 7 Pro Android phone. This demonstration ensured that participants understood what an mHealth app was before asking them related questions. All FGDs were audiotaped (for research purposes only), transcribed verbatim, and analyzed shortly after they were completed. Therefore, data collection and analysis became an iterative process until saturation of all topic areas was reached. Responses of participants were deidentified, and the transcribed data was not shared with anyone outside the investigating team.

Data Analysis

The data were transcribed by 2 experienced transcriptionists TA (a research assistant) and DA (co-author). Authors LK and DA conducted thematic coding of the transcripts manually along with the areas of interest. Specifically, LK created the initial code, which the authors reviewed. Identical subcodes were merged while redundant ones were removed. The interview transcripts were then read to understand and familiarize by LK and DA and independently coded to highlight texts relevant to the derived themes. The process led to a descriptive account focused on identifying critical variables for each heading [25].

SCT was used to explain the interpretations of positive supports for behavior changes in areas of personal, behavioral, and environmental factors [26]. In this investigation, personal factors (eg, previous knowledge of mHealth risks and benefits) were identified as prerequisites for behavior change but not the sole factor for the needed changes. Also, given the current dearth of mHealth use in mental health care in Nigeria and in the study population of young mothers, the point of inquiry on previous knowledge of use focused on all known areas of use within general health care. Second, the behavioral capability of adolescent mothers, their beliefs in their ability (self-efficacy) to engage in mobile phone app use for health purposes to carry out desired actions and produce desired effects, was assessed within the present social context on the roles and autonomy of adolescent girls as foundational to their likelihood of adopting and maintaining mHealth use [27,28]. Third, we documented outcome expectancies of engaging in mobile app use through participants' perceptions of the advantages and disadvantages of such action with the aim of identifying motivating and mitigating factors to assist in achieving desired results [29].

Ethical Considerations

The University College Hospital, University of Ibadan, ethical review board approved the study in accordance with the 1964 Declaration of Helsinki and its later amendments (UI/EC/19/0165). In this study, only perinatal adolescents aged 16-19 years who could give consent to participate, and primary health care providers over the age of 18 years were recruited. Study data from observations and audio recordings during the FGDs were stored as an mp4 file on a password-protected encrypted computer at the Department of Psychiatry, University of Ibadan. Group facilitators took notes that strictly contained only deidentified information. All the participants approached were given full information on study purposes and procedures and told that their participation was voluntary, before they were invited to participate. The informed consent form had two original copies for those who agreed to participate: one of these was stored securely in the research file and a copy was given to the participant. All participants gave voluntary responses to interview questions and were told they could decline to answer any questions that they chose. The adolescent girls received a total sitting and transport allowance of the equivalent of US $10, while the primary care providers received US $15.

Results

The mean age for young mothers was 17.3 (SD 0.9) years; they were predominantly single (14/19, 73.7%) with a few full cohabiting (4/19, 21.1%). The age of mothers who had some high school education was 10.6 (SD 1.9) years: 8 out of 19 (42.1%) of them were still students, 2 out of 19 (10.5%) of these were first-year university undergraduates); and 11 out of 19

XSL•FO
RenderX

(57.8%) others had dropped out of school and engaged in petty trading. The care providers were middle-aged (mean 48, SD 5.8 years) majorly female (18/19, 95%), and with significant (mean 22.2, SD 4.7 years) work experience. A summary of the demographic information of participants is presented in Tables 1 and 2. Our open-ended question guide unveiled several responses according to the themes presented in Tables 1 and 2 below.

Table 1. Demographic information of women participants.

Variables	Values
Marital status, n (%)	
Single	14 (74)
Married	1 (5)
Cohabiting	44 (21)
Occupation, n (%)	
Students	8 (42)
Petty trading	11 (58)
Age of perinatal adolescent participants (year), mean (SD)	17.3 (0.9)
Years of education of adolescent participants, mean (SD)	10.6 (1.9)

Table 2. Demographic information of care provider participants.

Variable	Values
Sex of care providers, n (%)	
Male	1 (5)
Female	19 (95)
Total	20 (100)
Marital status, n (%)	
Married	20 (100)
Age of care provider participants (year), mean (SD)	48 (5.8)
Years of work experience, mean (SD)	22.2 (4.7)

Factors Related to mHealth Use

The perinatal adolescent groups had knowledge of mHealth use for health interventions in reproductive health care and mental health care. The latter knowledge was related to their participation in the ongoing RAPID study [24] and its mHealth component. Group members commonly reported receiving SMS on reproductive health promotion and prevention programs:

> *I receive regular text messages telling me about contraceptive use and different types*
>
> *I know about text messages advising people to space their children*
>
> *I don't know who sends it but I get regular messages in Yoruba (a language in Nigeria) telling me to visit my nearest clinic to ask about family planning methods.*

However, the adolescent mothers were of two minds about the usefulness of health information (SMS) in reproductive care. On the one hand, they saw it as not useful since stigma would prevent them from taking advantage of such a program because of their age:

> *I would be too self-conscious to go and ask for such services, also I don't think health care providers will attend to people unmarried especially young girls who want contraceptives.*

On the other hand, some agreed that the SMS provided useful information that was relevant for the future.

The majority in the perinatal adolescent groups identified their sources of knowledge of mHealth use for mental health care as through mobile phone calls or SMS for follow-up from care providers during their participation in the ongoing RAPID project [24]:

> *Matron would call me to remind me about clinic appointments for depression care on the day before the appointment dates.*
>
> *The nurse sent me a message to remind me to take my baby for immunisation because when she called the day before then I had complained of not having much energy.*
>
> *I got a call from the nurse…she asked how I was doing, the calls from her always lifts my mood.*
>
> *Matron called me the first evening I visited the clinic..she said she wanted to be sure I was okay…she*

XSL•FO
RenderX

> *also told me not to forget the homework she gave me as part of the treatment for my depression.*

All the care providers reported their use of mobile phone apps and SMS in various ways in their practice in maternal and child health care:

> *…we use a mobile app given to use by the WHO for child immunisation data cordination and we also have other one from a project on family planning which helps us in self learning.*

All of the care providers had wide experiences in using phone-to-phone calls for service provision and the WhatsApp (Meta) platform for supervision and support of health interventions.

> *We call doctors and matrons who are our clinical supervisors on the mobile phone when needed and staff members also chat on group Whatsapp platforms for work purposes.*

Perinatal Adolescents and Mobile App Use

Both groups of participants (perinatal adolescents and primary health care providers) had positive comments on the appropriateness of mobile app use in the perinatal adolescent population for depression care, and the abilities of this population to engage its use for depression care. One care provider commented:

> *…you know that young people like to explore phones so delivering health services to them through this avenue increases their likelihood of engaging in such care. More than nine out of ten of these girls can read and write so accessing health care on the mobile phone is what they will be able to do.*

Many of the young mothers noted it could help them to reduce the times they might have to go to the clinics because of the long waiting times they often experience:

> *Many times at the clinic, I get tired and just want to go home because the crowd can be many…so this kind of app, will help someone like me since it could reduce clinic visits.*

A few saw it as a way to overcome access stigma related to being pregnant at a young age and cost-related barriers to health care.

> *Going to the clinic when I was pregnant was always troubling for me because I felt so self-conscious of my being young compared to adult mothers… so I skipped many clinic days.*

Majority of the young mothers were excited about the thoughts of a mobile phone app with short videos for depression self-care, and a few also believed that it could be used by young mothers to improve users' moods and to prevent boredom. They generally found the use of video for treatment on the application very appealing, noting that it will help them recollect their discussions with the nurses at the clinic, which would help consolidate the actions they need to improve mental health:

> *It will keep me from being bored (shows excitement)…I like the idea because videos will be remembered easily…and If I have such app, I can also use it to remember anything I forget that the nurse said*
>
> *I watch short videos on face book, so such an app like this will be easy for me.*

Many young mothers reported watching short video clips on Facebook (Meta) for leisure. Care providers also agreed that

> *…a health app would help reinforce treatment from the clinics and help young mothers with self-care.*

Reports of both groups on mobile app use within the social context of perinatal adolescents gave very strong indications of the need to get parents, guardians, and partners to be on board and be supportive of its use. The care providers said:

> *Some parents might think the young mothers are using their phone as an excuse not to help with housework rather than for accessing their depression treatment.*

In the same vein, young mothers' perceptions were similar:

> *If my mother thinks I am staying on the phone too much and texting my boyfriend, she controls the use of my phone*
>
> *…if I have an app for treatment, my mother would need to be told from the clinic for her because if she sees me staying long on the phone, she could think that I am just playing games and she can seize the phone for that reason.*

One care provider had concerns about cases of severe depression that could result in functional limitations, thereby impacting the ability to use the application:

> *…a few might be too depressed to even use the app or even spend time looking at the treatment on it.*

Outcome Expectations of Application Adoption Among Young Mothers

In general, the two participant groups made constant reference to the positive outcome expectancies associated with engaging in mobile app use for the management of depression in this population. Some members in the perinatal adolescent groups showed visible excitement at the possibility of accessing health care through audiovisual materials on a mobile phone app by talking in high-pitched voices:

> *if I has an app like that when I was sick, I would have gotten better quicker because this would have increase the treatment I received.*
>
> *An app with videos as treatment will be good… it could help with feelings of sadness and how to take any needed action.*
>
> *I like watching videos on Facebook…they helped me when I felt down, so this is good because it will help to prevent boredom and sadness.*

Although the care providers were generally positive on the outcome expectation of application use on the mental health of perinatal adolescents, 2 of the older nurses were skeptical, given the possibility of increased treatment default rates [11] among

XSL•FO
RenderX

young mothers who replace clinic appointments with application use:

> *…if the app treatment is not properly presented to patients as what they should use along with the face-to-face clinic visits, some might not come as they should.*

This will be more so if a treatment plan for depression modify clinic visits inconsideration for frequency of application use.

> *You know adolescents don't like going to the clinic for appointments, my fear is that they will now not come at all if they have an app for self care.*

Also, another care provider pointed out the possibility of worsening symptoms if patients are unable to use the application, for example, as a result of severe depression. Some care providers have negative consequences of application misuse if patients share treatment contents with people with similar symptoms*:*

> *The patients would need to be warned against sharing the app with friends or relatives with similar symptoms to avoid severe consequences.*

Mitigating Challenges to mHealth Adoption in Young Mothers

Participants identified 5 action plans to mitigate challenges to young mothers' mHealth use. Both groups of participants agreed on the need for care providers to educate patient's relatives on its usefulness in young mothers' depression care:

> *If the nurse tells my mother that a mobile phone app is for treatment purposes, she will even be the one to encourage me to use it more so I can get well quickly.*
>
> *My grandmother that I live with will cooperate with whatever they say from the clinic regarding my health.*

The care providers generally agreed that mHealth use should be included as part of the total management plan for perinatal depression to remove possible misconceptions from patients who many see as an alternative to face-to-face care. Care providers and young mothers also suggested the use of phone calls from care providers to encourage and remind young mothers of application use. The majority of the care providers identified the need to warn adolescent mothers on a regular basis about the harms that could result from sharing the mHealth app with relatives and friends with similar symptoms of depression. A few of the care providers pointed out it would be necessary to block young mothers' access to mHealth apps after treatment to guide against misuse. Both caregiver and young mother FGD groups wanted the application attractive and interactive in order to engage young mothers and encourage usage.

Discussion

Principal Findings

Interventions based on behavioral theory use have proven to be more effective than theoretical approaches [30,31] allowing for a deeper understanding of the subjective dimension of the process [32].

To our knowledge, this is the first UCD approach assessing the feasibility of mobile app-based depression intervention for adolescent mothers, while also identifying factors to mitigate possible challenges to use by young mothers in Africa. Our study population (mean 17.3, SD 0.9 years) was adequately educated (mean 10.6, SD 1.9) contrary to findings in previous studies where perinatal adolescents were mostly uneducated [33,34].

On the whole, young mothers' and health workers responded positively to a video-based mHealth app used to manage adolescent perinatal depression [35]. This response could be an inference that there may be a high rate of adoption of such interventions in the future. A significant strength of the study is the participation of key stakeholders who were part of a recent trial. Contrary to a previous report on insufficient knowledge of mHealth use among clinicians in Nigeria [36], this study's primary health care providers were knowledgeable and currently engaged in mHealth use for different health purposes.

Perinatal adolescents had a sense of personal agency (self-efficacy) with respect to using the mHealth app, which was indicative of them having the skills and abilities necessary for performing the behavior under a variety of circumstances. This self-efficacy in young mothers is ultimately linked to their general level of education in young people in Nigeria, which is a result of basic universal education in the country [37].

Young mothers and care providers noted several positive outcome expectations as advantages of application use. Both groups agreed that this mode of digital health service provision could keep young mothers in contact with care as they navigate personal- and clinical-level barriers to care [38]. The general response supported the enthusiasm of young people for technology, which increases the likelihood of their engagement with it and its usefulness for depression care in the group. This finding is supported by a recent survey of the current use of mobile phones among young people in Nigeria, which reported the broad willingness of young mothers to engage in technology use for mental health care [4]. Based on the values and expectations of the young mothers' who reported watching videos on Facebook to entertain themselves, the idea of a video-based application therefore presented something familiar with their current pattern of entertainment, thereby increasing the likelihood of future use.

The young mothers identified the influence of social support in mediating behavior change [39,40] identifying both parents, partners, and health care providers as key stakeholders [24] in a mHealth intervention design for this population with depression. The importance of warning young mothers of the dangers of application sharing and uninstalling the mobile app after completion of treatment were potential ways to prevent misuse of mobile apps by patients. As in previous findings, the risks associated with patients sharing electronic health information with other users were an identified challenge [41].

Study Finding Through an SCT Lens

Unlike other cognitive development theories, SCT incorporates a 3-way relationship between individuals, their self-regulating process of efficacy, and their social circumstances. In this study,

XSL•FO
RenderX

we understand that supportive personal factors and the ability to engage in application use for depression care are insufficient for adoption given the limitations that can occur on the social level. Misunderstanding from parents and partners of its purposes for entertainment rather than for therapy is a main hindrance. This might be more so the case if adolescent mothers spend time away from participating in household activities due to mHealth use, or if parents suspect it is being used for other reasons such as "texting of boyfriend" or "playing games." This necessitated the need for care providers to adequately engage parents, guardians, and partners on the health benefits of mHealth solutions for mental health care and to prepare them for assisting young mothers with future challenges that they might encounter in its use. Such applications in the future also can entail tools for health care providers, parents, and partners, which can help mitigate clinical- and family-level stresses. A combined intervention strategy of mHealth and face-to-face care is a model that provides an opportunity to reduce parents' misconceptions about application use through interactions between care providers and patients' relatives.

Limitations

Data were collected from participants who were part of an ongoing clinical trial on perinatal depression that used mHealth and their current knowledge might have influenced their responses. For this reason, this study result might not be generalizable to views of all care providers or perinatal adolescents. Also, the transferability of the study finding is limited by the fact that young mothers in low-resource settings are majorly students or low-income earners and therefore might not all be able to afford smartphones if after a trial stage the intervention is found to be successful. Despite these limitations, this study provides useful information on how to mitigate identified barriers to adopting an mHealth app for adolescent perinatal depression in primary care.

Conclusions

Young mothers expressed excitement at the use of mHealth technology to manage perinatal depression. Real-life challenges, however, need to be attended to for successful implementation of such interventions. We found that care providers and relatives of young mothers are vital stakeholders in the success of a mHealth mental health management plan to manage depression in young mothers. Our results reinforce the need for management plans that combine face-to-face care in adopting mHealth in young mothers' mental health interventions.

Conflicts of Interest

DB-Z has financial interests in Merlin LLC and FOCUS technology and has provided consultation services to Trusst Health, K Health, Boehringer Ingelheim, eQuility, Deep Valley Labs, Butler Hospital, and Otsuka Pharmaceuticals.

References

1. Recommendations on digital interventions for health system strengthening. World Health Organization. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2016-06-06]
2. Naslund JA, Aschbrenner KA, Araya R, Marsch LA, Unützer J, Patel V, et al. Digital technology for treating and preventing mental disorders in low-income and middle-income countries: a narrative review of the literature. Lancet Psychiatry 2017;4(6):486-500 [FREE Full text] [doi: 10.1016/S2215-0366(17)30096-2] [Medline: 28433615]
3. Statistics ITUs reports on mobile cellular subscriptions. Sustainable Development Goals. 2019. URL: https://www.itu.int/en/ITU-D/Statistics/pages/stat/default.aspx [accessed 2025-02-05]
4. Kola L, Abiona D, Adefolarin AO, Ben-Zeev D. Mobile phone use and acceptability for the delivery of mental health information among perinatal adolescents in Nigeria: survey study. JMIR Ment Health 2021;8(1):e20314 [FREE Full text] [doi: 10.2196/20314] [Medline: 33496678]
5. Gureje O, Oladeji BD, Araya R, Montgomery AA, Kola L, Kirmayer L, et al. Expanding care for perinatal women with depression (EXPONATE): study protocol for a randomized controlled trial of an intervention package for perinatal depression in primary care. BMC Psychiatry 2015;15:136 [FREE Full text] [doi: 10.1186/s12888-015-0537-3] [Medline: 26122982]
6. Kola L, Abiona D, Oladeji BD, Ayinde O, Bello T, Gureje O. Theory-driven development of a mobile phone supported intervention for adolescents with perinatal depression. Soc Psychiatry Psychiatr Epidemiol 2022;57(6):1201-1210. [doi: 10.1007/s00127-021-02198-3] [Medline: 34779878]
7. Patel V, Saxena S, Lund C, Thornicroft G, Baingana F, Bolton P, et al. The lancet commission on global mental health and sustainable development. Lancet 2018;392(10157):1553-1598. [doi: 10.1016/S0140-6736(18)31612-X] [Medline: 30314863]
8. Kola L, Bennett IM, Bhat A, Ayinde OO, Oladeji BD, Abiona D, et al. Stigma and utilization of treatment for adolescent perinatal depression in Ibadan Nigeria. BMC Pregnancy Childbirth 2020;20(1):294 [FREE Full text] [doi: 10.1186/s12884-020-02970-4] [Medline: 32410586]
9. Kumar M, Huang KY. Impact of being an adolescent mother on subsequent maternal health, parenting, and child development in Kenyan low-income and high adversity informal settlement context. PLoS One 2021;16(4):e0248836 [FREE Full text] [doi: 10.1371/journal.pone.0248836] [Medline: 33793592]
10. Kingston D, Heaman M, Fell D, Chalmers B, Maternity Experiences Study Group of the Canadian Perinatal Surveillance System, Public Health Agency of Canada. Comparison of adolescent, young adult, and adult women's maternity experiences and practices. Pediatrics 2012;129(5):e1228-e1237. [doi: 10.1542/peds.2011-1447] [Medline: 22529278]

XSL•FO
RenderX

11. Gureje O, Oladeji BD, Montgomery AA, Araya R, Bello T, Chisholm D, et al. High- versus low-intensity interventions for perinatal depression delivered by non-specialist primary maternal care providers in Nigeria: cluster randomised controlled trial (the EXPONATE trial). Br J Psychiatry 2019;215(3):528-535. [doi: 10.1192/bjp.2019.4] [Medline: 30767826]
12. Nove A, Matthews Z, Neal S, Camacho AV. Maternal mortality in adolescents compared with women of other ages: evidence from 144 countries. Lancet Glob Health 2014;2(3):e155-e164 [FREE Full text] [doi: 10.1016/S2214-109X(13)70179-7] [Medline: 25102848]
13. Gureje O, Lasebikan VO. Use of mental health services in a developing country. Results from the Nigerian survey of mental health and well-being. Soc Psychiatry Psychiatr Epidemiol 2006;41(1):44-49. [doi: 10.1007/s00127-005-0001-7] [Medline: 16341828]
14. Leplatte D, Rosenblum KL, Stanton E, Miller N, Muzik M. Mental health in primary care for adolescent parents. Ment Health Fam Med 2012;9(1):39-45. [Medline: 23277797]
15. Ben-Zeev D, Kaiser SM, Brenner CJ, Begale M, Duffecy J, Mohr DC. Development and usability testing of FOCUS: a smartphone system for self-management of schizophrenia. Psychiatr Rehabil J 2013;36(4):289-296 [FREE Full text] [doi: 10.1037/prj0000019] [Medline: 24015913]
16. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014;39(5):356-364 [FREE Full text] [Medline: 24883008]
17. Gudmundsdottir GB, Vasbø KB. Methodological Challenges When Exploring Digital Learning Spaces in Education. Cham: Springer; 2014.
18. Abras CM, Preece J. User-centered design. In: Encyclopedia of Human-Computer Interaction. Thousand Oaks, CA: Sage Publications; 2004:445-456.
19. Hussain-Shamsy N, Shah A, Vigod SN, Zaheer J, Seto E. Mobile health for perinatal depression and anxiety: scoping review. J Med Internet Res 2020;22(4):e17011 [FREE Full text] [doi: 10.2196/17011] [Medline: 32281939]
20. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
21. Naslund JA, Shidhaye R, Patel V. Digital technology for building capacity of nonspecialist health workers for task sharing and scaling up mental health care globally. Harv Rev Psychiatry 2019;27(3):181-192 [FREE Full text] [doi: 10.1097/HRP.0000000000000217] [Medline: 30958400]
22. Bandura A. Social cognitive theory of self-regulation. Organ Behav Hum Decis Process 1991;50(2):248-287. [doi: 10.1016/0749-5978(91)90022-l]
23. McAlister AL, Perry CL, Parcel GS. How individuals, environments, and health behaviors interact: social cognitive theory. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice. Hoboken, NJ: Jossey-Bass; 2008:169-188.
24. Gureje O, Kola L, Oladeji BD, Abdulmalik J, Ayinde O, Zelkowitz P, et al. Responding to the challenge of adolescent perinatal depression (RAPiD): protocol for a cluster randomized hybrid trial of psychosocial intervention in primary maternal care. Trials 2020;21(1):231 [FREE Full text] [doi: 10.1186/s13063-020-4086-9] [Medline: 32106885]
25. Ritchie J, Lewis J, Nicholls CM. Qualitative Research Practice: A Guide for Social Science Students and Researchers. Washington, DC: SAGE Publications; 2003.
26. Wong C, Monaghan M. Behavior change techniques for diabetes technologies. In: Diabetes Digital Health. Amsterdam, Netherlands: Elsevier; 2020:65-75.
27. Maddux JE, Kleiman E, Gosselin JT. Self-Efficacy. New York, NY: Guilford Press; 2012.
28. Meltzer-Brody S, Bledsoe-Mansori SE, Johnson N, Killian C, Hamer RM, Jackson C, et al. A prospective study of perinatal depression and trauma history in pregnant minority adolescents. Am J Obstet Gynecol 2013;208(3):211.e1-211.e7 [FREE Full text] [doi: 10.1016/j.ajog.2012.12.020] [Medline: 23246315]
29. Reesor L, Vaughan EM, Hernandez DC, Johnston CA. Addressing outcomes expectancies in behavior change. Am J Lifestyle Med 2017;11(6):430-432 [FREE Full text] [doi: 10.1177/1559827617722504] [Medline: 29302249]
30. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
31. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010;31:399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
32. Stacey FG, James EL, Chapman K, Courneya KS, Lubans DR. A systematic review and meta-analysis of social cognitive theory-based physical activity and/or nutrition behavior change interventions for cancer survivors. J Cancer Surviv 2015;9(2):305-338 [FREE Full text] [doi: 10.1007/s11764-014-0413-z] [Medline: 25432633]
33. Chiazor IA, Ozoya MI, Idowu AE, Udume M, Osagide M. Teenage pregnancy: the female adolescent dilemma. Int J Sci Commer Humanit 2017;5(1):70-82 [FREE Full text]
34. Fleming N, O'Driscoll T, Becker G, Spitzer RF, CANPAGO COMMITTEE. Adolescent pregnancy guidelines. J Obstet Gynaecol Can 2015;37(8):740-756. [doi: 10.1016/S1701-2163(15)30180-8] [Medline: 26474231]
35. Kenny R, Dooley B, Fitzgerald A. Developing mental health mobile apps: exploring adolescents' perspectives. Health Informatics J 2016;22(2):265-275 [FREE Full text] [doi: 10.1177/1460458214555041] [Medline: 25385165]

XSL•FO
RenderX

36. Owolabi BS, Odugbemi TO, Odeyemi KA, Onigbogi OO. mHealth: knowledge and use among doctors and nurses in public secondary health-care facilities of Lagos, Nigeria. J Clin Sci 2018;15(1):27. [doi: 10.4103/jcls.jcls_41_17]
37. Uko-Aviomoh E, Okoh E, Omatseye BOJ. Universal basic education (UBE) in Nigeria: an appraisal. Education 2007;127(4):558-569 [FREE Full text]
38. Kruse C, Betancourt J, Ortiz S, Valdes Luna SM, Bamrah IK, Segovia N. Barriers to the use of mobile health in improving health outcomes in developing countries: systematic review. J Med Internet Res 2019;21(10):e13263 [FREE Full text] [doi: 10.2196/13263] [Medline: 31593543]
39. Luszczynska A, Cieslak R. Mediated effects of social support for healthy nutrition: fruit and vegetable intake across 8 months after myocardial infarction. Behav Med 2009;35(1):30-38. [doi: 10.3200/BMED.35.1.30-38] [Medline: 19297302]
40. Skarin F, Olsson LE, Friman M, Wästlund E. Importance of motives, self-efficacy, social support and satisfaction with travel for behavior change during travel intervention programs. Transp Res F: Traffic Psychol Behav 2019;62:451-458. [doi: 10.1016/j.trf.2019.02.002]
41. Hao J, Gao B. Advantages and disadvantages for nurses of using social media. Journal of Primary Health Care and General Practice 2017;1(1):2 [FREE Full text]

Abbreviations

FGD: focus group discussions
mHealth: mobile health
mhGAP: Mental Health Global Action Programme
RAPiD: Responding to the challenge of Adolescent Perinatal Depression
SCT: Social Cognitive Theory
UCD: user-centered design
WHO: World Health Organization

Edited by A Mavragani; submitted 06.09.22; peer-reviewed by K Atmore; comments to author 01.11.23; revised version received 16.11.23; accepted 20.11.23; published 09.04.25.

Please cite as:
Kola L, Fatodu T, Kola M, Olayemi BA, Adefolarin AO, Dania S, Kumar M, Ben-Zeev D
Factors Impacting Mobile Health Adoption for Depression Care and Support by Adolescent Mothers in Nigeria: Preliminary Focus Group Study
JMIR Form Res 2025;9:e42406
URL: https://formative.jmir.org/2025/1/e42406
doi:10.2196/42406
PMID:

©Lola Kola, Tobi Fatodu, Manasseh Kola, Bisola A Olayemi, Adeyinka O Adefolarin, Simpa Dania, Manasi Kumar, Dror Ben-Zeev. Originally published in JMIR Formative Research (https://formative.jmir.org), 09.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Monthly Cycling of Food Insecurity in Latinas at Risk for Diabetes: Methods, Retention, and Sample Characteristics for a Microlongitudinal Design

Angela Bermúdez-Millán[1*], MPH, PhD; Rafael Pérez-Escamilla[2*], PhD; Sofia Segura-Pérez[3*], MS, RD; James Grady[1*], DrPH; Richard S Feinn VI[4*], PhD; Hanako Agresta[1*], BSc; Dean Kim[1*], BA; Julie Ann Wagner[5*], PhD

[1]Department of Public Health Sciences, School of Medicine, UConn Health, 263 Farmington Avenue, Farmington, CT, United States

[2]Department of Social and Behavioral Sciences, Yale School of Public Health, New Haven, CT, United States

[3]Hispanic Health Council, Hartford, CT, United States

[4]Frank H. Netter School of Medicine, Quinnipiac University, Hamden, United States

[5]Division of Behavioral Sciences and Community Health, School of Dental Medicine, UConn Health, Farmington, CT, United States

[*]all authors contributed equally

Corresponding Author:
Angela Bermúdez-Millán, MPH, PhD
Department of Public Health Sciences, School of Medicine, UConn Health, 263 Farmington Avenue, Farmington, CT, United States

Abstract

Background: Food insecurity (FI) is a risk factor for type 2 diabetes (T2D) that disproportionately affects Latinas. We conducted a microlongitudinal study to examine the relationship of monthly cycling of FI and diabetes risk factors.

Objective: This study aimed to determine the quantitative methodology, recruitment and retention strategies, predictors of retention across time, and baseline sample demographics.

Methods: Participants were adult Latinas living in Hartford, Connecticut who were recruited through a community agency, invited to participate if they were receiving Supplementary Nutrition Assistance Program (SNAP) benefits, screened positive for FI using the 2-item Hunger Vital Sign Screener, and had elevated risk factors for T2D using the American Diabetes Association risk factor test. Using a microlongitudinal design, we collected data twice per month for 3 months (week 2, which is a period of food budget adequacy; and week 4, which is a period of food budget inadequacy) to determine if the monthly cycling of FI was associated with near-term diabetes risk (fasting glucose, fructosamine, and glycosylated albumin) and long-term risk (BMI, waist circumference, and glycated hemoglobin) markers. We determined whether household food inventory, psychological distress, and binge eating mediated associations. We examined Health Action Process Approach model constructs. To assess the relationship between monthly cycling of FI with diabetes risk markers, we used repeated measures general linear mixed models. To assess the role of mediators, we performed a causal pathway analysis.

Results: Participant enrollment was from April 1, 2021 to February 21, 2023. A total of 87 participants completed 420 assessments or a mean of 4.83 (SD 2.02) assessments. About half (47/87, 54%) of the sample self-identified as Puerto Rican, mean age was 35.1 (SD 5.8) years, with 17.1 (SD 11.6) years in the mainland United States. Just under half (41/87, 47.1%) spoke Spanish only, 69% (60/87) had no formal schooling, and 31% (27/87) had less than eighth grade education. Modal household size was 4 including 2 children; 44.8% (39/87) were not living with a partner. About half (47/87, 54%) were unemployed, 63.2% (55/87) reported a monthly income <US $1000, and 63.2% (55/87) used food pantries. In total, 61 participants (70.1%) completed all 6 assessments. On Pearson correlation analysis, having internet at home and having a tablet at home were associated with a higher number of completed assessments.

Conclusions: This study demonstrated how FI cycles over the month and whether and to what degree the cycling itself is related to the risk for T2D development, as well as the evidence for some putative mechanisms of this association that can serve as future intervention targets including SNAP disbursement schedules.

(JMIR Form Res 2025;9:e66970) doi:10.2196/66970

KEYWORDS

food insecurity; monthly cycling; type 2 diabetes risk; quantitative methods; Latinas; endocrinology; nutrition; nutrition assistance; micro-longitudinal design

XSL•FO
RenderX

Introduction

Latinas have nearly double the rates of type 2 diabetes (T2D; 11.3%) compared with non-Hispanic white women (6.1%) [1]. This health inequity is not fully understood. Prediabetes, a condition in which a person's blood glucose levels are higher than normal but do not reach the cutoff for diabetes, is a significant risk factor for developing T2D. In the United States, more than one-third of Latinas meet at least 1 American Diabetes Association criterion of prediabetes, and those younger than 45 years merit immediate attention for the prevention of advancing to full diabetes [2].

A risk factor for T2D that disproportionately affects Latinas is food insecurity (FI). A novel risk factor is the monthly cycling of food insecurity, an issue that has been notably understudied. FI is the limited or uncertain availability of nutritionally adequate and safe foods or the ability to acquire acceptable foods in socially acceptable ways [3,4]. In 2023, 13.5% (18.0 million) of US households were food insecure at least some time during the year [5]. Rates of FI were higher than the national average for households with children and particularly households with children headed by single adults, Hispanic-headed households, and households in principal cities.

FI has been associated with overweight or obesity, cardiometabolic disease, and T2D [6-8]. FI may influence diabetes risk in several ways. First, it may affect diet by promoting dependence on inexpensive, highly palatable foods that are energy dense, which are associated with weight gain and promote the development of chronic conditions [9]. Second, women in food insecure households may have elevated mental distress as a result of their central role in food acquisition and preparation for their families. The rates of depression and anxiety are higher among food insecure mothers, and among Latinos living in poverty [10]. Third, FI may affect emotional eating [11,12], the tendency to eat or overeat in response to negative emotions. Emotional eating has repeatedly been linked to overweight and obesity due to cycles of restriction and binge eating [12], resulting in fluctuations of meal size and frequency that contribute to weight gain.

A useful theoretical framework for describing the sociocognitive processes involved in health behaviors, such as fruit and vegetable intake, is the Health Action Process Approach (HAPA) model [13]. HAPA constructs include risk awareness, outcome expectancy, self-efficacy, intention, action planning, and coping planning. Understanding these constructs among Latinas in relation to FI may inform behavioral approaches for this population.

Most studies of FI have used cross-sectional or multiwave longitudinal designs, referencing the past 30 days, 3, 6, or 12 months, and have not captured the weekly ebb and flow of access to food, which may be crucial for understanding the role of FI as a diabetes risk factor, and for designing interventions. Yet, it has been hypothesized that among households that receive Supplemental Nutrition Assistance Program (SNAP) benefits, FI is not stable but cycles over the course of the month [14]. SNAP benefits are distributed in the beginning of the month and may not be sufficient to attenuate or prevent FI by the end of the month [15]. A small number of studies have shown that SNAP benefits lasting less than a month is a predictor of poor dietary quality and health outcomes [16] but this has not been studied in our population of interest. Our hypothesis is that FI cycling over the course of a month is detrimental to women's health and is a risk factor for T2D.

Descriptive and formative research is crucial to address this potentially serious public health problem. Therefore, this innovative study aims to document FI occurrence over the month and test whether FI is associated with increases in diabetes risk markers in Latinas. Using a microlongitudinal design to capture the disbursement cycle of SNAP benefits, we collected data twice per month, the second week of the month which we expected to be a period of food budget adequacy and the fourth week of the month, which we expected to be a period of food budget inadequacy. We repeated this schedule for 3 months.

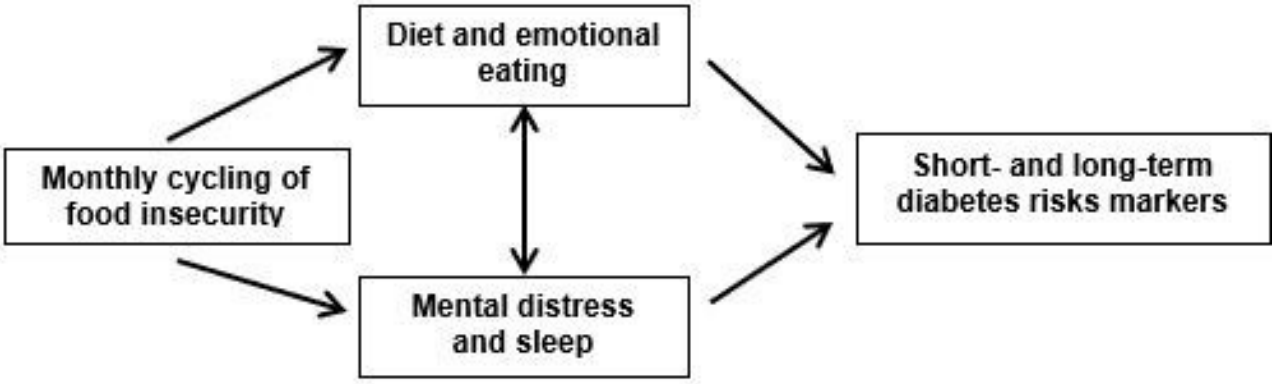
The first aim of the study was to determine if the monthly cycling of FI is associated with diabetes risk biomarkers. We hypothesized that compared with week 2 (food secure: period of food budget adequacy), week 4 (food insecure: period of food budget inadequacy) will be characterized by higher near-term diabetes risk biomarkers, that is, fasting glucose, fructosamine, and glycosylated albumin. We also hypothesized that compared with participants with little change in FI, participants with greater weekly changes in FI will have higher long-term metabolic risk markers for T2D (BMI, waist circumference, and glycated hemoglobin, HbA1c).

The second aim of the study was to describe the monthly cycling of household food inventory, eating behaviors, HAPA constructs, and distress. We hypothesized that compared with week 2 (food secure: period of food budget adequacy), week 4 (food insecure: period of food budget inadequacy) will be characterized by more eating disturbance (emotional eating and binge-eating), poorer diet quality (fat, sugar, fruit, and vegetables), and increased psychological distress (depression and anxiety symptoms; Figure 1). Changes in HAPA constructs over time are described. To our knowledge, this is the first microlongitudinal study to investigate FI cycling over time.

Longitudinal designs with multiple assessments can pose research burden on participants and are difficult to conduct in low-resourced populations. The aims of this paper are to report our quantitative methodology, to describe our recruitment and retention strategies, to examine predictors of retention across the 6 timepoints, and to report baseline demogrraphics for the sample.

XSL•FO
RenderX

Figure 1. Conceptual framework for the Food Insecurity Cycling study (Hartford, Connecticut, enrollment 2021 - 2023).



Methods

Design

This study used a microlongitudinal design with 6 observations across 3 months.

Participants

Eligibility criteria are shown in Textbox 1.

Textbox 1.

Inclusion criteria

- Age 18 - 44 years.
- Self-identify as Latina.
- Self-identify as a woman.
- Hartford area resident.
- Currently receiving Supplementary Nutrition Assistance Program (SNAP) benefits.
- Read and write in Spanish or English.
- Access to a smart phone to complete assessments.
- Screen positive for food insecurity per validated 2-item food insecurity screen [17,18] and at high risk for diabetes per standard criteria [19,20] (a score ≥4 indicates elevated risk for Hispanics). We studied adult Latinas of childbearing years because data suggest that they may prioritize nutrition for their children over their own nutrition [21].

Exclusion criteria

- Currently or recently pregnant or planning to become pregnant in the next 3 months (because of potential changes to nutrition benefits with pregnancy).
- Self-reported diabetes, kidney, or liver disease (because of nutritional requirements for management of those conditions).
- Positive screen for current substance abuse or dependence (to decrease attrition).
- Self-reported vasovagal syncope while undergoing phlebotomy (ie, history of feeling light-headed, dizziness, hot, sweaty, or fainting, to reduce medical urgencies associated with home phlebotomy).
- Living at a shelter (because women in a shelter do not have their own reliable food storage space).
- Those with previously undiagnosed diabetes who were identified at baseline as having unequivocal extant diabetes were excluded due to differences in cardiometabolic function and referred for treatment.

Setting

As a US state, Connecticut is predominantly white, but its capital city of Hartford is largely Latino. Consistent with national data, Latinos in Connecticut have a significantly higher rate of diabetes than their white counterparts [22]. The study was led and coordinated by UConn Health and implemented in close partnership with the Hispanic Health Council (HHC) that used the study's 2 bilingual and bicultural community health workers (CHWs). HHC is a community-based research, service, training, and advocacy organization headquartered in Hartford. This institutional partnership was built from a trusting relationship built over years of successful collaboration between UConn Health, Yale School of Public Health and HHC co-designing and implementing research studies using CHWs [23].

Recruitment

A comprehensive staff training program was developed and delivered for HHC recruiters and data collectors, including a core manual to support fidelity to the assessment procedures

XSL•FO
RenderX

and study protocol. CITI-trained, bilingual and bicultural recruiters distributed a recruitment flyer across HHC programs serving the study's target population. The recruitment flyer briefly described the study and informed intersted individuals to call the recruiters. Programs included the Santa Marquez Special Supplemental Nutrition Program for Women, Infant, and Children Program (WIC) Center, which serves to safeguard the health of low-income women, infants, and children up to age 5 who are at nutritional risk. The recruitment flyer was also posted in the HHC social media platforms (ie, Hispanic Health Council Facebook, Instagram, and Twitter pages). HHC staff also distributed flyers at health fairs, Hartford mobile markets, community health agencies, local convenience stores, and local daycare centers and it was mentioned on Spanish-language radio. Recruitment also took place at local food pantries that serve the target community. Foodshare is the largest antihunger organization in the Greater Hartford area, with a network of food pantries, meal programs, and mobile foodshare sites.

Screening and Enrollment

Eligibility screening was conducted by the CHW recruiters. Due to local rates of COVID-19 infection during the recruitment, most screening took place by phone, and some in a private room at the HHC. These sessions were scheduled through a virtual scheduling calendar accessible only to team personnel. The recruiters asked questions ascertaining the inclusion and exclusion criteria.

Screening Measures

Demographic and clinical inclusion and exclusion criteria were self-reported. The 2-item Hunger Vital Sign FI screener was used to detect likely FI [17]. The 2 questions were "In the past 3 months, have you or other adults in your household worried whether your food would run out before you got money to buy more?" and "In the past 3 months, has the food you or other adults in your household bought just didn't last and you didn't have money to get more?" Response options were "yes," or "no." A response of "yes" to either question met study inclusion criteria for food insecurity [17].

The American Diabetes Association risk test was used to determine elevated risk for developing T2D. It includes seven self-report questions (total score of 0 - 11) on age, gender, gestational diabetes mellitus, family history of diabetes, high blood pressure, physical activity, and obesity (based on BMI through a weight-height chart). The American Diabetes Association uses a cutoff of 5 or higher to indicate increased for diabetes. In this study, since all participants endorsed FI, which is itself a novel risk factor for diabetes, a score of ≥4 was required for inclusion [19,20].

The 4-item CAGE-AID (CAGE Adapted to Include Drugs) screener [24] was used to detect likely alcohol and substance abuse. A score of 2 is recommended as a positive screen for alcohol or substance abuse and indicates a potential problem that warrants further assessment by a health care professional; a score of 2 or higher excluded a recruit from participation.

Assessment Procedures and Incentives

After obtaining consent to participate and HIPAA (Health Insurance Portability and Accountability Act) authorization, physical assessments were conducted in the participant's preferred language (English or Spanish), at the HHC. Participants had their weight, height, waist, and hip circumferences measured and were asked to provide fasting blood samples. Assessments were performed early morning (ie, before 9 am) to obtain fasting blood samples (30 ml). Study phlebotomist from HHC took blood samples from an antecubital vein, using sterile technique. Samples were then promptly transported by trained personnel, compliant with HIPAA and biohazard regulations, in an ice chest to UConn Health Clinical Research Center for processing and analysis. There, blood samples were aliquoted. One aliquot went to UConn Health Hospital Clinical Laboratory for immediate assessment of glucose. The remaining aliquots went to the UConn Health Clinical Research Center where they were centrifuged and stored at −80°C until analyzed for all other assays described below.

Following the baseline assessment, participants provided data at 2 time points per month for 3 months, in order to capture the disbursement cycle of SNAP benefits. Participants received cash compensation for participation, US $15 at each timepoint for each blood draw and US $10 for surveys, totaling US $150.

Retention Efforts

Because the temporal cycling of food insecurity over time is central to our research question, retention of this difficult-to-reach population was of utmost importance. Several community-engaged efforts were made to retain participants including the following: (1) Socializing the participant to the research endeavor for the community and describing the importance of each data collection timepoint; (2) Maintaining updated contact information for participants and a designated secondary contact person; (3) Speaking in the participant's preferred language and fostering good rapport between CHWs and participants; (4) responding fully, genuinely, and in a timely manner to any participant concerns; (5) phone and text reminders for study appointments; and, (6) offering financial compensation for time spent on study activities.

Measures

Next, participants were scheduled for a baseline assessment. Due to COVID-19 social distancing mandates, the majority of the survey instruments were collected over the phone by trained bilingual interviewers who entered participant verbal responses into REDCap (Research Electronic Data Capture) [25] on large-screen electronic tablets. For all self-report surveys, response options referenced the preceding 7-days. Refer to Table 1 for data collection timeline.

XSL•FO
RenderX

Table . Measures and assessment schedule.

Predictor	Month 1[a]		Month 2[a]		Month 3[a]	
	Week 1	Week 4	Week 1	Week 4	Week 1	Week 4
Food security[b] (past 7 days timeframe)	✓	✓	✓	✓	✓	✓
Covariates						
Demographics	✓	N/A[c]	N/A	N/A	N/A	N/A
SNAP[d] benefits	✓	✓	✓	✓	✓	✓
Mediators (past 7 days timeframe)						
Nutrition[e]	✓	✓	✓	✓	✓	✓
Distress (symptoms of anxiety and depression)	✓	✓	✓	✓	✓	✓
Sleep quality	✓	✓	✓	✓	✓	✓
Outcomes (short-term)						
Fasting blood glucose and insulin	✓	✓	✓	✓	✓	✓
Fructosamine	✓	✓	✓	✓	✓	✓
Glycosylated albumin	✓	✓	✓	✓	✓	✓
Outcomes (long-term)						
BMI	✓	N/A	N/A	N/A	N/A	✓
Waist and hip circumference	✓	N/A	N/A	N/A	N/A	✓
HbA1c	✓	N/A	N/A	N/A	N/A	✓

[a]SNAP benefits are made available on the electronic benefits transfer card from the 1st to the 3rd of every month, based on the first letter of the client's last name. Data for week 1 were collected the week after food purchasing, during food abundancy; Week 4 data were collected before the next scheduled SNAP benefit disbursement, during food scarcity.

[b]15-item US Household Food Security Survey Module.

[c]N/A: not applicable.

[d]SNAP: Supplementary Nutrition Assistance Program.

[e]Nutrition measures included food frequency questionnaire, household food inventory, HAPA model constructs, emotional eating, and binge eating.

Participant Characteristics at Baseline

Demographics included age (18 - 24 years), language (English only, English and Spanish, or Spanish only), education (no formal schooling, eighth grade or less, some high school, high school graduate or GED [General Educational Development high school equivalency examination], some college, associates degree, or finished 4 years of college), employment (working full time or part time, unemployed, student, or homemaker), relationship status (living with a spouse or partner or not living with a spouse or partner), household composition (mother, father, mother-in-law, father-in-law, boyfriend/ or spouse, own children, sisters/, brothers, or others), and years living in the United States (free response). Financial strain was measured on a 5-point Likert scale [23] from 1= "We have enough and we can save" to 5 = "We don't have enough and we have great difficulties." Digital connectivity was measured with questions regarding whether the participant had internet and access to a computer and a tablet in the home, with response options yes or no.

Food Insecurity

Household food insecurity was assessed using a 15-item adapted and validated version of the original 18-item US Household Food Security Survey Module for pregnant Latinas [26,27]. The module captures the experience of anxiety regarding household food supplies, inadequate food quality and reduced food intake. This measure is validated in English and Spanish.

Nutrition Assistance

Participants reported how much in food stamps (SNAP) benefits they received per month, whether benefits were received once

XSL•FO
RenderX

or twice per month, how many weeks the benefits last, whether and how much others in the household receive SNAP, and whether and how much the participant received from WIC.

Food Frequency Questionnaire

Diet quality was measured with the Spanish and English versions of the Block Fat, Sugar, Fruitt, Vegetable Screener, a 55-item food frequency questionnaire shown to have good reliability and validity [28,29]. Participants were instructed how to access and enter data through a link sent to them through email. Visual aids were uploaded in REDCap as images to help participants with portion size estimations [30].

Emotional Eating

The Emotional Eating subscale of the Three Factor Eating Questionnaire R18-V2 [31,32] assesses emotional eating. This subscale was validated in English and Spanish among Latinos of various Latino heritages [33]. The Spanish-language version [34] of the Questionnaire on Eating and Weight Patterns - 5 is a screening instrument for binge eating. It assesses eating - disorder–behaviors and attitudes that align with DSM - 5 diagnostic criteria [35].

Psychological Distress

The validated Spanish version 8-item scale of the Patient-Reported Outcomes Measurement Information System emotional anxiety scale (Short Form 8a) assesses anxiety symptoms [36,37]. The validated Spanish version of the Personal Health Questionnaire Depression Scale (PHQ-8, ie, the PHQ-9 without the suicidality item) [38,39] assesses depressive symptoms. The suicidality item was omitted to avoid CHWs being required to respond to mental health urgencies in the field.

Sleep Quality

Sleep quality was measured with the Pittsburgh Sleep Quality Index (PSQI) [40]. The PSQI consists of 19 items that produce a global sleep quality score as well as individual items such as usual bedtime, wake-up time, actual hours slept, number of minutes to fall asleep, and nighttime awakenings. Global PSQI score was estimated for each participant as described previously, and as analyzed by our [41] team.

Health Action Process Approach Constructs

Questions were structured to assess motivational and volitional phases of the HAPA model as they pertain to eating behaviors. Specific items were adapted from previous studies of Latinas of childbearing age [42,43]. Four items tapped into risk awareness, for example, “If you continue with your current eating pattern, how likely is it that you will develop diabetes?” Response options were very unlikely, unlikely, likely, and very likely. Five items tapped into outcome expectancy, for example, “How true is it if you are able to improve and maintain healthier eating patterns, you will prevent diabetes?” Response options were not at all true, somewhat true, mostly true, and very true. Four items tapped into self-efficacy, for example, “I am certain or confident that I am able to improve my eating patterns, even if I have to make a detailed plan to have appropriate food available (on a food budget).” Response options were not at all true, somewhat true, mostly true, and very true. Three items tapped into intentions, for example, “During the next month, do you intend to increase the amount of vegetables you eat?” Response options were very unlikely, unlikely, likely, and very likely. Four items tapped into action planning, for example, “Do you have a plan of when you will go about eating more fruits and vegetables?” Response options were yes or no. Three items tapped into coping planning, for example, “Do you have a plan of how to incorporate more fruits and vegetables into your daily routine?” Response options were not at all, somewhat, or definitely.

Household Food Inventory

Participants completed a revised version of a home food inventory that has been validated in low-income Hispanic or Latinx households [44-46]. Participants completed the inventory themselves through REDCap [25]. Response options for each food were yes or no. Participants were instructed how to access and enter data through a link sent to them through email.

Anthropometrics

Weight was measured using a Seca Digital Scale and height was measured through a Seca Portable Stadiometer. Height was measured to the nearest 1 cm with the participant’s head positioned in the Frankfurt horizontal plane. Weight was measured to the nearest 0.1 kg. The average of 2 trials was used for both height and weight, unless measures were separated by a minimum of 1 cm or 0.5 kg, in which case, 2 more measures are taken. BMI (kg/m^2) was calculated from height and weight. Waist circumference was measured twice at the umbilicus and hip circumference at the buttocks (nearest 0.5 cm).

Glucometrics

Fasting glucose reflects the body’s ability to regulate glucose in the absence of any glycemic challenge (ie, food). It is a strong predictor of incident T2D. Fasting glucose was analyzed at the UConn Health clinical laboratory. Fasting insulin was analyzed at the UConn Health Clinical Research Center with a kit from Siemens Healthcare Diagnostics (Los Angeles, California). It has a reference range median of 8.9 uIU/mL, sensitivity of 2 uIU/mL, and inter- and intrarun coefficients of variation (CVs) of 2.5 and 3.6, respectively. Fructosamine reflects the physiology of glucose in the extracellular space and has a short half-life, reflecting glycemia over 2 - 4 weeks, ideal to address glycemia over intervals of several weeks. Fructosamine measures the fraction of total serum proteins that have undergone glycation. Assays were run at the UConn Health Clinical Research Center using a kit from ARUP laboratories (Salt Lake City, Utah) with an assay range 170 - 285 µmol/L. Glycosylated albumin is formed by glycation of serum albumin. Several large, prospective studies have found that fructosamine and glycated albumin, a protein formed when glucose (sugar) binds to albumin, is a test that reflects short-term glycemia, predict conversion to T2D [47] with some studies demonstrating this effect even after controlling for baseline fasting glucose and HbA1c [48]. Glycated albumin was analyzed in serum at the UConn Health Clinical Research Center using a kit from MyBioSource Inc, San Diego, California. The assay has a range of 0.626 - 20 mg/mL, with a sensitivity of 0.1 mg/mL. Intra- and interassay CVs are less than 15%.

XSL•FO
RenderX

Glycemia

HbA_{1c} was measured in the UConn clinical laboratory using high pressure liquid chromatography. In this laboratory, HbA_{1c} shows the following coefficients of variation for normal and high values: level 1 mean=5.51%, CV=3.3 based on n=320, and level 2 mean=9.01%, CV=3.2 based on n=304.

Insulin Resistance

Glucose was reported in mg/dL and insulin was reported in uiU/mL. Log transformed homeostatic model assessment of insulin resistance (logHOMA-IR, Homeostatic Model Assessment of Insulin Resistance) was calculated from fasting glucose and insulin values according to the standard formula: log(fasting glucose × fasting insulin)/405 [49].

Sample Size and Power Analysis

Considering insulin resistance as our primary outcome, a previous study by our team provided data for a power analysis based on insulin resistance among persons at risk for T2D [50]. In that study, the SD for logHOMA-IR was approximately 0.3 and we expected to see a difference in logHOMA-IR of 0.4 between FI and food security, and a difference of 0.5. Our targeted sample size of n=100 would have 99% power to detect a main effect of FI, 96% power for mental distress and 87% power for the (FI × distress) interaction in a regression model using a 2-sided α of 0.05.

Planned Data Analysis

Summary statistics will be used initially to present the results from the categorical and continuous variables. Skewness, kurtosis, and potential outliers will be examined and if needed transformations or categorical variables will be created. The internal consistency of self-report measures will be assessed by Cronbach α.

To assess the relationship between monthly cycling of FI with diabetes risk markers, we will use repeated measures general linear mixed models. With 3 months of data, collected at weeks 2 and 4, depending on the distribution of for each outcome variable, the most appropriate distribution and link function will be chosen. In separate analyses it will be possible to divide the sample by frequency of benefits (ie, once vs twice monthly benefits) and include a dummy variable to evaluate the effects of frequency of benefits in the statistical models. It will also be possible to divide the sample into food secure, intermittent and persistent FI based on longitudinal studies using this approach [51].

To assess the hypothesis that psychological distress, diet, and eating disturbance mediate the relationship between FI and diabetes risk, we will perform a causal pathway analysis. We are interested in examining whether there is a change in our mediators over time, and if so, whether this change significantly leads to a change in diabetes risk markers. For this we will conduct dynamic mediation using latent change (difference) score mediation analysis [52,53].

For food frequency measures, linear regression models will compare food security status on household food availability. Spearman rank-order correlations of total inventoried availability of fruit and vegetable servings with the food frequency questionnaire will be conducted. We plan to use “GENMOD” or “MIXED” procedures in SAS (SAS Institute), or similar, for regression models and MPLUS software (Muthén & Muthén), or similar, for the mediation causal pathway analysis.

Ethical Considerations

If a recruit was not excluded, then recruiters obtained verbal consent, collected contact information, and scheduled the first visit at HHC. During this first visit, the recruiters obtained written informed consent. Participants were given a signed, dated copy of the informed consent form for their records and the recruiter kept the original, signed, dated copy. All verbal and written study materials, including recruitment materials and informed consent forms, were provided to the participants in their preferred language (English or Spanish). Self-report measures and their response options were pilot tested for readability, understanding, logic, flow, cultural relevance, and face validity by 5 Spanish speakers and 5 English speakers at the Hispanic Health Council. The final assessment battery was deemed appropriate by the research team. This protocol was approved by the UConn Health Institutional Review Board Study 21-045OS-1 and HHC agreed for UConn’s Health Institutional Review Board to be the institutional review board of record. All procedures were conducted according to the Declaration of Helsinki. Incentives were calculated to compensate participants for the time spent in study-related activities but were not coercive in a sample that screens positive for food insecurity (refer to Assessment Procedures and Incentives).

Results

As can be seen in the diagram in Figure 2, 845 potential participants were screened. The majority (n=685) were not eligible, the most common exclusions being because they did not receive SNAP benefits (n=229), were outside of the age range (n=155), had extant diabetes (n=90), or did not have elevated risk for type 2 diabetes (n=83). Of the 160 participants who were eligible, 73 declined with the most common reasons for declining being lack of time, lack of transportation, and concerns about COVID-19 pandemic. A total of 89 participants were consented and completed baseline blood assays but 2 withdrew immediately afterwards; 87 of them went on to be enrolled.

As can be seen in Table 2, just over half (54%) of the sampled self-identified as Puerto Rican, mean age was 35.1 (5.8) years, with 17.1 (11.6) years in the mainland United States. Just under half (47.1%) spoke Spanish only, 69% had no formal schooling and 31% had less than eighth grade education. Modal household size was 4 including 2 children, and 44.8% were not living with a partner. About half (54%) were unemployed, 63.2% reported a monthly income <US $1000 and 63.2% used food pantries.

Overall, participants completed 420 assessments or mean=4.83 (SD=2.02) assessments. Of the 87, 70.1% (n=61) completed all 6 assessments. Results of Spearman’s correlations show that higher number of assessments completed was associated with

XSL•FO
RenderX

having internet in the home (r=0.27, p=.01) and there was a statistical trend for having a tablet in the home (r=-0.20, p=.07).

XSL•FO
RenderX

Figure 2. CONSORT diagram for the food insecurity cycling microlongitudinal study Hartford, Connecticut, enrollment 2021 - 2023.

XSL•FO
RenderX

Table . Demographic characteristics (n=87).

Characteristics	Values
Ethnicity n (%)	
Puerto Rican	47 (54)
Non-Puerto Rican	40 (46)
Age, mean (SD), range (min-max)	35.1 (5.8) (18-44)
People Living in Home, n (%)	
1	1 (1.1)
2	7 (8)
3	21 (24.1)
4	24 (27.6)
5	23 (26.4)
6+	11 (12.6)
Children in home, n (%)	
0	5 (5.7)
1	16 (18.4)
2	30 (34.5)
3	24 (27.6)
4+	12 (13.7)
Years in United States, mean (SD), range (min-max)	17.1 (11.6) (0.3-43.3)
Language spoken, n (%)	
English and Spanish	46 (52.9)
Spanish only	41 (47.1)
Education, n (%)	
No formal schooling	60 (69)
Eighth grade or less	27 (31)
Relationship status, n (%)	
Living with spouse or partner	48 (55.2)
Not living with partner	39 (44.8)
Employment status, n (%)	
Employed	39 (44.8)
Unemployed	47 (54)
No answer	1 (1.1)
Household income ($), n (%)	
0-1000	55 (63.2)
1000 or more	32 (36.8)
Financial strain, n (%)	
Have enough and save	13 (14.9)
Have just enough	48 (55.2)
Not enough, with difficulties	25 (28.7)
No answer	1 (1.1)
Financial assistance, n (%)	
Cash	3 (3.4)

XSL•FO
RenderX

Characteristics	Values
Food Pantries	55 (63.2)
Soup Kitchen	5 (5.7)
Supplemental security income	15 (17.2)
Title 19/Medicaid	54 (62.1)
Section 8 housing	16 (18.4)
Nutrition for women, infants, children and children (WIC)	35 (40.2)
Other (housing and diapers)	4 (4.6)
Digital connectivity, n (%)	
Internet in the home	82 (94.3)
Tablet in the home	39 (44.8)
Computer in the home	37 (42)

Discussion

Principal Findings

The main findings reported here are that the study successfully recruited and retained a sample of the target population. There was excellent participation over time with 71% of the sample completing all 6 assessment timepoints. Participants were able to complete assessments using only their smart phones by cell phone service. Nonetheless, higher completion of assessments was associated with having internet and a tablet at home. This may reflect ease of connectivity (rather than access to connectivity) or a familiarity with technology that was advantageous. Other indicators of socioeconomic status (income and educational attainment) were not associated with number of assessments completed. As in other parts of the United States, digital inequity in Connecticut persists across race and ethnicity, income and education [54]. The surprisingly high rate of internet observed in this sample of Latinas who screened positive for food insecurity was likely due to assistance during COVID-19 pandemic for internet connection in the home. Future studies may achieve even higher rates of participation by providing tablets or internet service to participants, but the expense and ethical dilemmas of terminating internet service at the end of the study makes this approach potentially problematic.

This study was conducted during the COVID-19 pandemic. Social distancing requirements presented numerous challenges especially for recruitment, retention, and data collection. The study team remained flexible regarding format for assessments (telephone, Cisco WebEx, or in person when it was safe to do so) and strived to balance standardization with participant preference and evolving public health guidelines.

The innovative study protocol presented here will provide a comprehensive description of how FI is related to SNAP distribution, as well as, importantly, whether and to what degree FI cycling itself is related to risk for T2D. Furthermore, the study will provide early evidence for some putative mechanisms of this association, including contextual mechanisms such as household food availability, psychosocial mechanisms such as symptoms of depression and anxiety, and behavioral mechanisms such as distress-related binge eating. Each of these findings may suggest potential intervention targets.

Novelty

Food insecurity is a known risk factor for onset of new T2D [55]. However, we postulate that food insecurity is not static; rather, it varies according to access to food which is largely driven by the reductions in the amount of money available from social assistance programs (including SNAP) over the course of the month in some instances with families finding themselves with hardly any economic resources left for food or other basic needs by the last week of the benefits monthly cycle. This cycling has been hypothesized in the literature [56] yet microlongitudinal data regarding FI cycling is not available, as far as we know (eg, [57]). Furthermore, we propose that this cycling is related not only to household food inventory, but also to distress and distress-related eating behaviors. Our group has documented that FI is associated cross-sectionally with higher symptoms of depression, anxiety, and stress hormones as well as lower sleep quality [41]. Yet it remains unknown whether FI and emotional distress covary over time.

Another novel contribution of this study is the procedure for the household food inventory. We taught participants to complete the inventory checklist themselves using REDCap [25]. Participants were instructed how to access and enter data through a link sent to them through email. Lessons learned regarding the feasibility of the method, and its fidelity, will be important to inform future studies using household food inventories with hard-to-reach populations, or studies when home visits by researchers are not possible due to pandemic, natural disaster, or other situation that disrupts home visits. We pilot tested items and omitted the items in the checklist that were not understood. However, were unable to conduct true fidelity checks to compare the participant responses with CHW responses, because the university protocols were not allowing home visits for research purposes. If found to be feasible and reliable, this method may also be more efficient for research teams and perhaps preferable for participants.

XSL•FO
RenderX

Limitations

There are several limitations to this study. There is no experimental manipulation, so, whereas the longitudinal design can establish temporal precedence, no conclusions about causation can be drawn. This study was conducted during the COVID-19 pandemic. SNAP benefits since early 2020 have been affected by temporary pandemic-related benefit increases; an adjustment to the Thrifty Food Plan, upon which SNAP benefits are based, and higher-than-normal cost-of-living adjustments to reflect high food price inflation. Thus, findings may be influenced by changes to SNAP benefits during COVID-19 pandemic. COVID-19 pandemic also delayed and then paused the study, thus hindering recruitment. Data collection therefore took longer than anticipated and we fell short of our enrollment target by 13 participants. Future research with larger sample sizes should explore interactions between the exposure (FI) and mediators (psychological distress, diet, and eating disturbance) on outcomes (biomarkers of diabetes risk). Finally, the generalizability of our findings will need to be tested in other populations (eg, men, non-Latinos, or other parts of the country).

Limitations are balanced by the strengths of this study. The microlongitudinal design allows investigation of the temporal unfolding of covariance between FI and outcomes of interest over time. Home-based methods, particularly household food inventories, greatly improve contextual validity of assessments. The study affords the opportunity to confirm or assess the validity of certain measures in a Latina population, such as the Block Fat, Sugar, Fruit, and Vegetable Screener food frequency questionnaire. The population is high-risk and difficult-to-reach. Our procedures leverage the rigor and resources of a university medical center with the experience, knowledge, and reach of a community-based organization. Use of CHWs for recruitment, retention, and data collection improves trust and legitimacy in the target population. Multilevel bio-behavioral-psychosocial-contextual assessments are a particular strength. If our hypotheses are not refuted and this line of research goes on to establish causal pathways, interventions to disrupt the cycling of FI may prove beneficial for Latina health.

Conclusions

Results of this study may have potential policy implications. In 2022, 375,400 Connecticut residents, or 10% of the state population, received SNAP benefits [58]. More than 50% of recipients were in households with children and more than 37% were working families. Findings of this study may have implications for SNAP policy, particularly amount and timing of SNAP benefit distribution. Qualitative findings show that SNAP benefits are often inadequate for participants to purchase nutrient-rich foods that are sufficient for 1 month of household intake, particularly for participants at the higher end of the SNAP income distribution. Two studies have found that participant responses to food security surveys depend upon when in the SNAP cycle they are queried, with higher reports of FI when SNAP benefits are most salient [56,59,60]. This suggests considerable variability in food security within the month.

Inadequate benefit levels may contribute to cyclical eating patterns, where participants increase food consumption when benefits are received and reduce food consumption when benefits are depleted [14]. It may also lead to the purchase of quite unhealthy ultraprocessed foods and sugar sweetened beverages that have a long shelf-life, rather than fresh fruits and vegetables that are more likely to spoil over the month. In another study, individuals were not receiving benefits at baseline and then received 3 months of benefits. At baseline, food expenditures were not cyclical and were not related to impulsivity. However, total food expenditures became cyclical after households started receiving benefits, with more impulsive shoppers demonstrating a more accentuated benefit cycle [15]. The current monthly distribution of benefits could be replaced with a biweekly distribution to address cyclical food security, food expenditures, and eating patterns. Such a disbursement would provide SNAP recipients with greater flexibility.

Acknowledgments

This study was funded by the US National Institute of Diabetes and Digestive and Kidney Diseases 1R21DK122312-01A1. We thank the staff from the Hispanic Health Council for their assistance implementing this study.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

ABM contributed to conceptualization, funding acquisition, investigation, methodology, project administration, supervision, writing original, and writing review. JAW managed methodology, writing original, and writing review. RPE handled methodology, supervision, and writing review. JG and RF conducted formal analysis, resources, and writing review. HOA and DK assisted with formal analysis, visualization, and writing review. SSP handled project administration, resources, supervision, and writing review. All authors approve the final version. All authors agree to be accountable for all aspects of the work.

References

1. National diabetes statistics report centers for disease control and prevention. 2022. URL: https://www.cdc.gov/diabetes/data/statistics/statistics-report.html [accessed 2025-03-19]

XSL•FO
RenderX

2. Avilés-Santa ML, Pérez CM, Schneiderman N, et al. Detecting prediabetes among Hispanics/Latinos from diverse heritage groups: does the test matter? Findings from the Hispanic Community Health Study/Study of Latinos. Prev Med 2017 Feb;95:110-118. [doi: 10.1016/j.ypmed.2016.12.009] [Medline: 27956225]
3. Anderson SA. Core indicators of nutritional state for difficult-to-sample populations. J Nutr 1990 Nov;120(11):1555-1598. [doi: 10.1093/jn/120.suppl_11.1555]
4. Pérez-Escamilla R. Food and nutrition security definitions, constructs, frameworks, measurements, and applications: global lessons. Front Public Health 2024;12:1340149. [doi: 10.3389/fpubh.2024.1340149] [Medline: 38550323]
5. Rabbitt MP, Reed-Jones M, Hales LJ, Burke MP. Household food security in the united states in 2023. Economic Research Service, US Department of Agriculture. 2024 Sep ERR-337. [doi: 10.32747/2024.8583175.ers]
6. Martin MA, Lippert AM. Feeding her children, but risking her health: the intersection of gender, household food insecurity and obesity. Soc Sci Med 2012 Jun;74(11):1754-1764. [doi: 10.1016/j.socscimed.2011.11.013] [Medline: 22245381]
7. Leung CW, Epel ES, Ritchie LD, Crawford PB, Laraia BA. Food insecurity is inversely associated with diet quality of lower-income adults. J Acad Nutr Diet 2014 Dec;114(12):1943-1953. [doi: 10.1016/j.jand.2014.06.353] [Medline: 25091796]
8. Fitzgerald N, Hromi-Fiedler A, Segura-Pérez S, Pérez-Escamilla R. Food insecurity is related to increased risk of type 2 diabetes among Latinas. Ethn Dis 2011;21(3):328-334. [Medline: 21942166]
9. Laraia BA. Food insecurity and chronic disease. Adv Nutr 2013 Mar 1;4(2):203-212. [doi: 10.3945/an.112.003277] [Medline: 23493536]
10. Hromi-Fiedler A, Bermúdez-Millán A, Segura-Pérez S, Pérez-Escamilla R. Household food insecurity is associated with depressive symptoms among low-income pregnant Latinas. Matern Child Nutr 2011 Oct;7(4):421-430. [doi: 10.1111/j.1740-8709.2010.00266.x] [Medline: 20735732]
11. Nguyen-Rodriguez ST, Chou CP, Unger JB, Spruijt-Metz D. BMI as a moderator of perceived stress and emotional eating in adolescents. Eat Behav 2008 Apr;9(2):238-246. [doi: 10.1016/j.eatbeh.2007.09.001] [Medline: 18329603]
12. Rasmusson G, Lydecker JA, Coffino JA, White MA, Grilo CM. Household food insecurity is associated with binge-eating disorder and obesity. Int J Eat Disord 2018 Dec 19. [doi: 10.1002/eat.22990] [Medline: 30565270]
13. Schwarzer R. Self-regulatory processes in the adoption and maintenance of health behaviors. J Health Psychol 1999 Mar;4(2):115-127. [doi: 10.1177/135910539900400208] [Medline: 22021474]
14. Leung CW, Hoffnagle EE, Lindsay AC, et al. A qualitative study of diverse experts' views about barriers and strategies to improve the diets and health of Supplemental Nutrition Assistance Program (SNAP) beneficiaries. J Acad Nutr Diet 2013 Jan;113(1):70-76. [doi: 10.1016/j.jand.2012.09.018] [Medline: 23260725]
15. Valluri S, Mason SM, Peterson HH, French SA, Harnack LJ. Trends in cyclical food expenditures among low-income households receiving monthly nutrition assistance: results from a prospective study. Public Health Nutr 2021 Feb;24(3):536-543. [doi: 10.1017/S136898002000405X] [Medline: 33059779]
16. Whiteman ED, Chrisinger BW, Hillier A. Diet quality over the monthly supplemental nutrition assistance program cycle. Am J Prev Med 2018 Aug;55(2):205-212. [doi: 10.1016/j.amepre.2018.04.027] [Medline: 29935945]
17. Hager ER, Quigg AM, Black MM, et al. Development and validity of a 2-item screen to identify families at risk for food insecurity. Pediatrics 2010 Jul;126(1):e26-e32. [doi: 10.1542/peds.2009-3146] [Medline: 20595453]
18. Gundersen C, Engelhard EE, Crumbaugh AS, Seligman HK. Brief assessment of food insecurity accurately identifies high-risk US adults. Public Health Nutr 2017 Jun;20(8):1367-1371. [doi: 10.1017/S1368980017000180] [Medline: 28215190]
19. Ward ED, Hopkins WA, Shealy K. Evaluation of the use of a diabetes risk test to identify prediabetes in an employee wellness screening. J Pharm Pract 2022 Aug;35(4):524-527. [doi: 10.1177/0897190021997010] [Medline: 33622071]
20. Scanlan AB, Maia CM, Perez A, Homko CJ, O'Brien MJ. Diabetes risk assessment in Latinas: effectiveness of a brief diabetes risk questionnaire for detecting prediabetes in a community-based sample. Diabetes Spectr 2018 Feb;31(1):31-36. [doi: 10.2337/ds16-0051] [Medline: 29456424]
21. Olson CM. Food insecurity in women. Top Clin Nutr 2005 Oct;20(4):321-328. [doi: 10.1097/00008486-200510000-00005]
22. Mortality tables, statewide by race/ethnicity, 2005-2007 2010. 2011 Mar 3. URL: http://www.ct.gov/dph/lib/dph/hisr/hcqsar/mortality/xls/AAMR_State_RaceEthn_aggregate_2005-2007c.xls [accessed 2025-03-19]
23. Wagner J, Bermudez-Millan A, Damio G, et al. Community health workers assisting Latinos manage stress and diabetes (CALMS-D): rationale, intervention design, implementation, and process outcomes. Transl Behav Med 2015 Dec;5(4):415-424. [doi: 10.1007/s13142-015-0332-1] [Medline: 26622914]
24. Hinkin CH, Castellon SA, Dickson‐Fuhrman E, Daum G, Jaffe J, Jarvik L. Screening for drug and alcohol abuse among older adults using a modified version of the CAGE. American J Addict 2001 Oct;10(4):319-326 [FREE Full text] [doi: 10.1111/j.1521-0391.2001.tb00521.x]
25. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
26. Hromi-Fiedler A, Bermúdez-Millán A, Melgar-Quiñonez H, Pérez-Escamilla R. Psychometric properties of an adapted version of the U.S. household food security survey module for assessing food insecurity among low-income pregnant Latinas. J Hunger Environ Nutr 2009 Jan 1;4(1):81-94. [doi: 10.1080/19320240802706866] [Medline: 20046913]

XSL•FO
RenderX

27. Hromi-Fiedler A, Bermúdez-Millán A, Segura-Pérez S, Damio G, Pérez-Escamilla R. Adaptation of the U.S. food security survey module for low-income pregnant Latinas: qualitative phase. J Hunger Environ Nutr 2009 Jan 1;4(1):62-80. [doi: 10.1080/19320240802706841] [Medline: 20046909]
28. Lalonde I, Graham M, Slovinec-D'Angelo M, Beaton L, Brown J, Block T. Validation of the block fat/sugar/fruit/vegetable screener in a cardiac rehabilitation setting. J Cardiopulm Rehabil Prev 2008;28(5):340. [doi: 10.1097/01.HCR.0000336162.26720.95]
29. Block G, Gillespie C, Rosenbaum EH, Jenson C. A rapid food screener to assess fat and fruit and vegetable intake. Am J Prev Med 2000 May;18(4):284-288. [doi: 10.1016/s0749-3797(00)00119-7] [Medline: 10788730]
30. Boushey CJ, Spoden M, Zhu FM, Delp EJ, Kerr DA. New mobile methods for dietary assessment: review of image-assisted and image-based dietary assessment methods. Proc Nutr Soc 2017 Aug;76(3):283-294. [doi: 10.1017/S0029665116002913] [Medline: 27938425]
31. Karlsson J, Persson LO, Sjöström L, Sullivan M. Psychometric properties and factor structure of the Three-Factor Eating Questionnaire (TFEQ) in obese men and women. Results from the Swedish Obese Subjects (SOS) study. Int J Obes Relat Metab Disord 2000 Dec;24(12):1715-1725. [doi: 10.1038/sj.ijo.0801442] [Medline: 11126230]
32. Stunkard AJ, Messick S. The three-factor eating questionnaire to measure dietary restraint, disinhibition and hunger. J Psychosom Res 1985;29(1):71-83. [doi: 10.1016/0022-3999(85)90010-8] [Medline: 3981480]
33. Jáuregui-Lobera I, García-Cruz P, Carbonero-Carreño R, Magallares A, Ruiz-Prieto I. Psychometric properties of Spanish version of the Three-Factor Eating Questionnaire-R18 (Tfeq-Sp) and its relationship with some eating- and body image-related variables. Nutrients 2014 Dec 4;6(12):5619-5635. [doi: 10.3390/nu6125619] [Medline: 25486370]
34. Toolkit P. Questionnaire on eating and weight patterns - adult. 2024 Apr 17. URL: https://www.phenxtoolkit.org/protocols/view/651201 [accessed 2025-03-19]
35. Yanovski SZ, Marcus MD, Wadden TA, Walsh BT. The questionnaire on eating and weight patterns-5: an updated screening instrument for binge eating disorder. Int J Eat Disord 2015 Apr;48(3):259-261. [doi: 10.1002/eat.22372] [Medline: 25545458]
36. Irwin DE, Stucky B, Langer MM, et al. An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. Qual Life Res 2010 May;19(4):595-607. [doi: 10.1007/s11136-010-9619-3] [Medline: 20213516]
37. Craig BM, Reeve BB, Cella D, Hays RD, Pickard AS, Revicki DA. Demographic differences in health preferences in the United States. Med Care 2014 Apr;52(4):307-313. [doi: 10.1097/MLR.0000000000000066] [Medline: 24374420]
38. Gilmer TP, Walker C, Johnson ED, Philis-Tsimikas A, Unützer J. Improving treatment of depression among Latinos with diabetes using project Dulce and IMPACT. Diabetes Care 2008 Jul;31(7):1324-1326. [doi: 10.2337/dc08-0307] [Medline: 18356401]
39. Merz EL, Malcarne VL, Roesch SC, Riley N, Sadler GR. A multigroup confirmatory factor analysis of the Patient Health Questionnaire-9 among English- and Spanish-speaking Latinas. Cultur Divers Ethnic Minor Psychol 2011 Jul;17(3):309-316. [doi: 10.1037/a0023883] [Medline: 21787063]
40. Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
41. Bermúdez-Millán A, Pérez-Escamilla R, Segura-Pérez S, et al. Psychological distress mediates the association between food insecurity and suboptimal sleep quality in Latinos with type 2 diabetes mellitus. J Nutr 2016 Oct;146(10):2051-2057. [doi: 10.3945/jn.116.231365] [Medline: 27489004]
42. Hromi-Fiedler A, Chapman D, Segura-Pérez S, et al. Barriers and facilitators to improve fruit and vegetable intake among WIC-eligible pregnant Latinas: an application of the health action process approach framework. J Nutr Educ Behav 2016;48(7):468-477. [doi: 10.1016/j.jneb.2016.04.398] [Medline: 27373861]
43. Martinez-Brockman JL, Harari N, Segura-Pérez S, Goeschel L, Bozzi V, Pérez-Escamilla R. Impact of the lactation advice through texting can help (LATCH) trial on time to first contact and exclusive breastfeeding among WIC participants. J Nutr Educ Behav 2018 Jan;50(1):33-42. [doi: 10.1016/j.jneb.2017.09.001] [Medline: 29325660]
44. Hearst MO, Fulkerson JA, Parke M, Martin L. Validation of a home food inventory among low-income Spanish- and Somali-speaking families. Public Health Nutr 2013 Jul;16(7):1151-1158. [doi: 10.1017/S1368980012004508] [Medline: 23034097]
45. Fulkerson JA, Nelson MC, Lytle L, Moe S, Heitzler C, Pasch KE. The validation of a home food inventory. Int J Behav Nutr Phys Act 2008 Nov 4;5:55. [doi: 10.1186/1479-5868-5-55] [Medline: 18983668]
46. Kong A, Fitzgibbon M, Hubbard C, Campbell RT, Kessee N, Schiffer L. Validation of a self-report home food availability checklist against in-home food inventories conducted in low-income Black/African American and Hispanic/Latinx households with preschool-age children. Appetite 2022 May 1;172:105964. [doi: 10.1016/j.appet.2022.105964] [Medline: 35124159]
47. Zaccardi F, Kurl S, Pitocco D, Ronkainen K, Laukkanen JA. Serum fructosamine and risk of type 2 diabetes mellitus among middle-age Finnish men: a 23-year population-based prospective study. Acta Diabetol 2015 Feb;52(1):161-166. [doi: 10.1007/s00592-014-0625-8] [Medline: 25026924]
48. Juraschek SP, Steffes MW, Miller ER III, Selvin E. Alternative markers of hyperglycemia and risk of diabetes. Diabetes Care 2012 Nov;35(11):2265-2270. [doi: 10.2337/dc12-0787] [Medline: 22875225]

XSL•FO
RenderX

49. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia 1985 Jul;28(7):412-419. [doi: 10.1007/BF00280883] [Medline: 3899825]
50. Wagner JA, Bermúdez-Millán A, Buckley TE, et al. Community-based diabetes prevention randomized controlled trial in refugees with depression: effects on metabolic outcomes and depression. Sci Rep 2023 May 30;13(1):8718. [doi: 10.1038/s41598-023-35738-9] [Medline: 37253820]
51. Garcia J, Hromi-Fiedler A, Mazur RE, et al. Persistent household food insecurity, HIV, and maternal stress in peri-urban Ghana. BMC Public Health 2013 Mar 11;13:215. [doi: 10.1186/1471-2458-13-215] [Medline: 23497026]
52. Preacher KJ. Advances in mediation analysis: a survey and synthesis of new developments. Annu Rev Psychol 2015 Jan 3;66:825-852. [doi: 10.1146/annurev-psych-010814-015258] [Medline: 25148853]
53. Selig JP, Preacher KJ. Mediation models for longitudinal data in developmental research. Res Hum Dev 2009 Jun 5;6(2-3):144-164. [doi: 10.1080/15427600902911247]
54. State of connecticut administrative services. State Digital Equity Plan. 2024. URL: https://portal.ct.gov/das/-/media/das/ctedtech/meetings/2024/9-9-24_de_updates.pdf [accessed 2025-03-19]
55. Seligman HK, Bindman AB, Vittinghoff E, Kanaya AM, Kushel MB. Food insecurity is associated with diabetes mellitus: results from the National Health Examination and Nutrition Examination Survey (NHANES) 1999-2002. J Gen Intern Med 2007 Jul;22(7):1018-1023. [doi: 10.1007/s11606-007-0192-6] [Medline: 17436030]
56. Gregory CA, Todd JE. SNAP timing and food insecurity. PLoS ONE 2021;16(2):e0246946. [doi: 10.1371/journal.pone.0246946] [Medline: 33630866]
57. Kinsey EW, Dupuis R, Oberle M, Cannuscio CC, Hillier A. Chronic disease self-management within the monthly benefit cycle of the Supplemental Nutrition Assistance Program. Public Health Nutr 2019 Aug;22(12):2248-2259. [doi: 10.1017/S1368980019001071] [Medline: 31104648]
58. Priorities CoBaP. A closer look at who benefits from SNAP: state-by-state fact sheets. 2023. URL: https://www.cbpp.org/research/a-closer-look-at-who-benefits-from-snap-state-by-state-fact-sheets#Connecticut [accessed 2025-03-19]
59. Gregory CA, Smith TA. Salience, food security, and SNAP receipt. J Policy Anal Manage 2019;38(1):124-154. [Medline: 30572412]
60. Pérez-Escamilla R, Ferris AM, Drake L, et al. Food stamps are associated with food security and dietary intake of inner-city preschoolers from Hartford, Connecticut. J Nutr 2000 Nov;130(11):2711-2717. [doi: 10.1093/jn/130.11.2711] [Medline: 11053511]

Abbreviations

CAGE-AID: CAGE Adapted to Include Drugs
CHW: community health workers
CV: coefficient of variation
FI: Food insecurity
GED: General Educational Development high school equivalency exam
HAPA: Health Action Process Approach
HbA1c: glycated hemoglobin
HHC: Hispanic Health Council
HIPAA: Health Insurance Portability and Accountability Act
logHOMA-IR: Log transformed homeostatic model assessment of IR
PHQ: Personal Health Questionnaire
PSQI: Pittsburgh Sleep Quality Index
REDCap: Research Electronic Data Capture
SNAP: Supplementary Nutrition Assistance Program
T2D: type 2 diabetes
WIC program: Women, Infants, and Children program

Edited by A Mavragani; submitted 27.09.24; peer-reviewed by MM Lee, R Bipat; revised version received 18.02.25; accepted 20.02.25; published 28.03.25.

Please cite as:
Bermúdez-Millán A, Pérez-Escamilla R, Segura-Pérez S, Grady J, Feinn VI RS, Agresta H, Kim D, Wagner JA
The Monthly Cycling of Food Insecurity in Latinas at Risk for Diabetes: Methods, Retention, and Sample Characteristics for a Microlongitudinal Design
JMIR Form Res 2025;9:e66970
URL: https://formative.jmir.org/2025/1/e66970
doi:10.2196/66970

XSL•FO
RenderX

© Angela Bermúdez-Millán, Rafael Pérez-Escamilla, Sofia Segura-Pérez, James Grady, Richard S Feinn VI, Hanako Agresta, Dean Kim, Julie Ann Wagner. Originally published in JMIR Formative Research (https://formative.jmir.org), 28.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Improving Health and Well-Being of People With Post–COVID-19 Consequences in South Africa: Situation Analysis and Pilot Intervention Design

Nicole Audrey Glover[1], MBChB; Farzana Sathar[1], MPH; Pride Mokome[1], BCur; Nkululeko Mathabela[1], BA; Sipokazi Taleni[1], PGDip; Sarah Alexandra van Blydenstein[2], CertPulm(SA)Phys; Anna-Maria Mekota[3], Dr rer nat; Salome Charalambous[1,4], PhD; Andrea Rachow[3,5,6], MD; Olena Ivanova[3,6], PhD

[1]Implementation Research Division, The Aurum Institute, 29 Queens Road, Parktown, Johannesburg, South Africa

[2]Department of Internal Medicine, Chris Hani Baragwanath Academic Hospital, University of the Witwatersrand, Johannesburg, South Africa

[3]Institute of Infectious Diseases and Tropical Medicine, Ludwig-Maximilians-Universität University Hospital, Munich, Germany

[4]School of Public Health, University of the Witwatersrand, Johannesburg, South Africa

[5]German Research Centre for Environmental Health, Helmholtz Zentrum München, Unit Global Health, Neuherberg, Germany

[6]German Center for Infection Research, Partner Site Munich, Munich, Germany

Corresponding Author:
Nicole Audrey Glover, MBChB
Implementation Research Division, The Aurum Institute, 29 Queens Road, Parktown, Johannesburg, South Africa

Abstract

Background: Multisystemic complications post–COVID-19 infection are increasingly described in the literature, yet guidance on the management remains limited.

Objectives: This study aimed to assess the needs, preferences, challenges, and existing interventions for individuals with post–COVID-19 symptoms. Based on this, we aimed to develop a context-adapted intervention to improve the overall health and well-being of individuals with post–COVID-19 complications.

Methods: We conducted a cross-sectional mixed-methods situation analysis assessing the needs, preferences, challenges, and existing interventions for patients with post–COVID-19 symptoms. We collected data through questionnaires, semistructured in-depth interviews, and focus group discussions (FGDs) from individuals diagnosed with COVID-19 within the previous 18-month period and health care providers who managed patients with COVID-19 in both inpatient and outpatient settings. Quantitative data were summarized using descriptive statistics, qualitative data were transcribed, and deductive analysis focused on suggestions for future interventions. Findings guided the development of a group intervention.

Results: We conducted 60 questionnaires, 13 interviews, and 3 FGDs. Questionnaires showed limited knowledge of post–COVID-19 complications at 26.7% (16/60). Of those who received any rehabilitation for COVID-19 (19/60, 31.7%), 94.7% (18/19) found it helpful for their recovery. Just over half (23/41, 56%) of those who did not receive rehabilitation reported that they would have liked to. The majority viewed rehabilitation as an important adjunct to post–COVID-19 care (56/60, 93.3%) and that support groups would be helpful (53/60, 88.3%). Qualitative results highlighted the need for mental health support, structured post–COVID-19 follow-up, and financial aid in post–COVID-19 care. Based on the insights from the situation analysis, the theory of change framework, and existing post–COVID-19 evidence, we designed and conducted a pilot support group and rehabilitation intervention for individuals with post–COVID-19 complications. Our main objective was to assess the change in physical and psychological well-being pre- and postintervention. The intervention included 8 weekly themed group sessions supplemented by home tasks. Effectiveness of the intervention was evaluated by questionnaires pre- and postintervention on post–COVID-19 symptoms, quality of life with the EuroQoL 5-Dimension 5-Level, short Warwick-Edinburgh Mental Wellbeing Scale, and physical function by spirometry and 1-minute sit-to-stand test. We also assessed the feasibility and acceptability of the intervention by questionnaires and semistructured in-depth interviews. The intervention outcome analysis is yet to be conducted.

Conclusions: Insights from patients and health care providers on the characteristics of post–COVID-19 complications helped guide the development of a context-adapted intervention program with potential to improve health and well-being post–COVID-19.

(JMIR Form Res 2025;9:e58436) doi:10.2196/58436

XSL•FO
RenderX

KEYWORDS

post–COVID-19; rehabilitation; support; quality of life; group care; well-being; South Africa; COVID-19; situation analysis; pilot; intervention; context-adapted; physical health; mental health; cross-sectional; mixed method; questionnaire; in-depth; interviews; survey; focus group; quantitative; qualitative; support group; hospital; patients; health care workers; health worker

Introduction

Most people who have had SARS-CoV-2 infection, causing COVID-19 disease, recover fully; however, a proportion experience longer-term complications resulting in persistent symptoms and functional impairment [1,2]. Terms used to describe this syndrome include long COVID-19, post-acute COVID-19 syndrome, or post–COVID-19 condition; however, the definitions remain heterogeneous [1-3]. Post–COVID-19 complications can affect multiple body systems including pulmonary, cardiovascular, neurological, and musculoskeletal systems, where over 50 long-term symptoms have been identified [1,4]. The prevalence of post–COVID-19 complications varies substantially, from 6% up to 80%, depending on definitions used, hospitalization status, and duration of follow-up [2,4,5]. Common symptoms include fatigue, headache, arthralgia and myalgia, dyspnea and notably neurocognitive, mental health, and sleep disorders [2,6,7]. In addition to the symptoms, objective assessments of lung, cardiac, and neurological function demonstrate long-term pathology [1,2,8]. These complications are reported to impact activities of daily living (ADL), quality of life, and ability to work [9,10].

Limited data exists in African countries on post-COVID-19 morbidity and its characterization [2,4,6]. A study conducted in South Africa demonstrated COVID-19-related symptoms in 66.7% of people 3 months posthospital discharge [11], which was supported in another study by Jassat et al [12]. Further descriptions of post–COVID-19 complications in the South African context have come from our previous study, demonstrating pathology in hospitalized post–COVID-19 individuals, where up to 1-month posthospital discharge 85.2% had abnormal chest–computed tomography, 49.5% had abnormal spirometry, 79% had abnormal diffusing capacity of the lung, and functional limitation was evidenced by reduced median 6-minute walking test (300, IQR 210-400 m) and elevated St George's Respiratory Questionnaire scores (median 21.6, IQR 9.2-48.3) [13]. Despite these findings, sample sizes are small and the breadth of data and characterization of post–COVID sequelae in these cohorts is limited.

In addition to the multisystem sequelae recognized post–COVID-19, recovery can be complicated by preexisting or newly developed chronic diseases such as cardiovascular disease, diabetes, HIV, and tuberculosis (TB) [14]. Understanding these multisystem complications is integral to directing health services effectively [9]. Due to the complexity of recovery post–COVID-19, a holistic approach should be taken for management. Furthermore, the burden of disease and local health resources differ across regions [2,6], demanding the evaluation of context-specific assessment tools and interventions, to improve long-term recovery post–COVID-19 [15-17].

The World Health Organization (WHO) and National Institute for Health and Care Excellence (NICE) guidelines recommend multidisciplinary and patient-centered care for post–COVID-19 complications [3,18]. In a randomized control trial conducted by Liu et al [19], a respiratory rehabilitation program was shown to improve respiratory function, quality of life, and anxiety. These findings are reflected in another study where post–COVID-19 rehabilitation showed improvement in lung function and a reduction in anxiety levels [20].

Considering this multidisciplinary approach in resource-constrained settings, such as low- and middle-income countries (LMICs), assisted self-management programs and peer support groups become pertinent. This was evidenced by Wright et al [21], where a self-management program guided by weekly educational support sessions demonstrated improvement in patient self-efficacy and mental well-being post–COVID-19 [21,22]. Self-management techniques and patient education with other chronic lung diseases, including post-TB lung disease and bronchiectasis, have been noted to empower patients, improve patient confidence, competence, and overall quality of life [23,24]. However, few studies are available on context-adapted and culturally appropriate tools and interventions for the management of post–COVID-19 in LMICs, the importance of which was highlighted by Jassat et al [25].

This paper aims to describe the findings from a situation analysis on post–COVID-19 rehabilitation services and the development of a context-adapted pilot intervention aimed to improve overall physical and mental health post–COVID-19 conducted in Johannesburg, South Africa.

Methods

Both the situation analysis and intervention have been conducted. Below we outline the methods used for the situation analysis and framework for intervention development.

Situation Analysis

We conducted a mixed-method situation analysis study with our main aim to explore the post–COVID-19 management landscape by assessing the needs, preferences, challenges, and existing interventions and strategies for patients with post–COVID-19 complications in Johannesburg, South Africa.

We conducted this study at the Chris Hani Baragwanath Academic Hospital (CHBAH) and Tembisa Provincial Tertiary Hospital (Tembisa), Johannesburg, South Africa. These health care facilities are located in townships in the City of Johannesburg and Ekurhuleni districts and cater to the general population. They are public health care facilities funded by the Gauteng Department of Health. Participants were recruited from outpatient medical and rehabilitation departments at the respective facilities, where patients who had COVID-19 disease were invited to participate, and health care workers who had managed patients with COVID-19 were invited to participate.

XSL•FO
RenderX

Potential participants were approached in the waiting areas of clinics, where prescreening procedures assessing COVID-19 history were done followed by full informed consent if potential participants were willing. The study was conducted between December 2022 and February 2023, which was 11-months after the last clear reported COVID-19 wave in South Africa [26].

We collected quantitative data through a cross-sectional standardized questionnaire-based interview (see Multimedia Appendix 1) administered by the research staff with individuals who were previously diagnosed with COVID-19. The questionnaire covered topics on post–COVID-19 symptoms, experience with rehabilitation programs, challenges, preferences, and suggestions for potential interventions. A total of 2 research assistants (male and female) and a research nurse (female) were trained and collected the survey data. We used descriptive statistics to summarize the data in frequencies, mean and median, using Microsoft Excel.

The qualitative study included semistructured in-depth interviews with health care workers who manage COVID-19 patients and focus group discussions (FGDs) with individuals who had post–COVID-19 symptoms. Separate interviews were used with health care workers to allow in-depth conversation on each discipline, and FGDs were used with patients to allow a broad conversation flow of COVID-19 reflection and to encourage group sharing (see Multimedia Appendices 2 and 3). Interviews and FGDs were facilitated by the research team (3 females and 1 male), in English, except 2 FGD's conducted in a combination of English, isiZulu, and Sepedi. The duration of the interviews and FGDs ranged from 10-minutes to 90-minutes depending on participant availability and willingness to engage in the content of the questions. A guide was used for the interviews focusing on health care worker reflection on managing COVID-19 patients, challenges and perceived need for rehabilitation; and a guide was used in the FGDs to discuss patient post–COVID-19 experiences and rehabilitation needs and challenges. Both guides allowed for flexibility of the discussion. Interviews and FGDs were recorded, transcribed, coded, and deductive thematic analysis was carried out focusing on main concepts for the intervention development.

Quantitative and qualitative data were triangulated through an iterative process to ensure the coherence of findings and obtain more insights on the desirable content and format of the intervention.

Intervention Development

The creation of the intervention followed the steps described in the Group Intervention framework [27]. We ensured a suitable context and setting for the intervention. Following this, we designed the group intervention addressing outcomes and measurements, sampling into groups, quantity of groups, and structure of the intervention. Finally, during the delivery of the intervention, we introduced team briefings and group notes to describe and reflect on the dynamics within each group and session, position of group participants and leaders, and communication.

The data collected in the situation analysis and available evidence in post–COVID-19 rehabilitation fed into the content and delivery format of the intervention. The design process, implementation, and evaluation plan for the intervention are described in more detail in the Results section. The outcomes of the intervention are being analyzed and will be presented in a separate paper.

Ethical Considerations

We obtained ethical approval for the situation analysis from Wits Human Research Ethics Committee (230405), and the 2 facilities, CHBAH and Tembisa Hospital. All participants gave written informed consent before participation. The study data are deidentified to protect participant confidentiality. Participants received R 300 (US $16.85) cash reimbursement for each study visit covering costs for travel and time.

Results

Situation Analysis and Needs Assessment

Quantitative Results

We conducted 60 questionnaires with individuals who had COVID-19; Table 1 shows the findings. We had 21.7% (13/60) male participants, with a median age of 49.5 (IQR 39.5-60) years. Knowledge of post–COVID-19 sequelae was limited, with only 26.7% (16/60) having heard of these complications. From the 60 participants, 31.7% (19/60) received any rehabilitation for COVID-19 in either inpatient or outpatient setting, where physiotherapy was the most common (17/19, 89.5%), followed by dietician (4/19, 21.1%), speech therapy and audiology (3/19, 15.8%), psychology (2/19, 10.5%), and occupational therapy (1/19, 5.3%). Of those who received rehabilitation, all received rehabilitation within 6 weeks of COVID-19 infection and 94.7% (18/19) found it helpful for recovery. Of those who did not receive any rehabilitation, 56.1% (23/41) reported that they would have liked to receive rehabilitation, notably psychological support and counseling (14/23, 60.9%) and physiotherapy (8/23, 34.8%). Majority of the participants felt that rehabilitation was an important adjunct to COVID-19 recovery (56/60, 93.3%) and said that they would find a support group helpful for their recovery post–COVID-19 (53/60, 88.3%).

XSL•FO
RenderX

Table . Characteristics and questionnaire responses from participants (N=60). Quantitative data collected in situation analysis at Chris Hani Baragwanath Academic Hospital and Tembisa Provincial Tertiary Hospital from December 2022 to February 2023.

Characteristic	Participants (N=60)
Sex (male), n (%)	13 (21.7)
Age (years), median (IQR)	49.5 (39.5-60)
Knowledge of post–COVID-19 complications[a], n (%)	16 (26.7)
Received rehabilitation for acute COVID-19 or post–COVID-19 complications, n (%)	19 (31.7)
Inpatient rehabilitation	14 (73.7)
Outpatient rehabilitation	5 (26.3)
Physiotherapy	17 (89.5)
Dietician	4 (21.1)
Speech therapy and audiology	3 (15.8)
Psychology	2 (10.5)
Occupational therapy	1 (5.3)
Received rehabilitation for acute COVID-19 or post–COVID-19 complications and found it helpful for recovery, n (%)	18 (30)
Did not receive rehabilitation and would have liked to, n (%)	23 (38.3)
Psychology or counseling requested	14 (23.3)
Physiotherapy requested	8 (13.3)
Main goals for recovery post–COVID-19, n (%)	
To carry out activities of daily living better	34 (56.7)
To be able to exercise or have better physical capacity	32 (53.3)
Improved mental health or coping strategies	25 (41.7)
Return to work	21 (35)
Connection with people and to be able to care for family	20 (33.3)
Feel rehabilitation is an important adjunct to care post–COVID-19, n (%)	56 (93.3)
Would find a support group helpful for recovery post–COVID-19, n (%)	53 (88.3)

[a]Knowledge of post–COVID-19 complications was assessed by asking "Do you know what long-COVID or post-COVID-19 condition is?" and allowing participants to answer yes or no and elaborate on their answer.

Qualitative Results

We conducted 13 interviews with health care providers and 3 FGDs with a total of 20 post–COVID-19 patients. The 3 main themes that emerged from the thematic analysis for potential interventions post–COVID-19 included the need for mental health support, structured post–COVID-19 follow-up, and the need for financial aid.

Need for mental health support was further explained by the request to focus on psychological well-being and support groups. Both health care providers and patients reported that psychological well-being was neglected during the pandemic and should be a focus in post–COVID-19 care. One participant reported:

> *I think we focus more on the physical healing than the psychological healing. So, most people who had Covid, they are not okay. They had post-Covid depression or post-Covid anxiety. So, I think that there is a lot of mental health [care] that needs to be done post-Covid.* [Female, health care provider]

Health care providers and patients who had opportunities to have debriefing sessions and support groups, further expanded on the importance of incorporating support groups into the mental health support post–COVID-19:

> *They really must get help. From our side… sitting together talking, [about our experience], it's a good thing. It's a helping thing, it's a very, very good thing. Thank you.* [Male, patient in FGD]

This importance of sharing experiences and leaning on additional support from others in the group context during post–COVID-19 care was further supported by a health care provider, saying:

> *… looking at a rehab unit where… people feel comfortable with coming in… sharing their stories to each other, and their success stories, because that's how you then motivate each other. It's not coming*

XSL•FO
RenderX

> *from someone that may not understand it, you're hearing it from patient to patient. And that might be nice.* [Female, health care provider]

In addition to the mental health support, structured medical post–COVID-19 follow-up care was also reported as an important part of post–COVID-19 care, where both patients and health care providers should have a structured follow-up program, rehabilitation, and referral to specialists depending on their complications. The importance of recognizing each individual and the treatment that they require was emphasized:

> *We are not statistics, we are human beings, we are sick, and we are affected by COVID, and I think there must be a way to create a rehab for us [to have] proper check-up for us, because she is telling [us] about her lungs and I am telling you about my eyes and what are they doing? So, in six months or one year you must take this person to a specialist for check-ups.* [Female, health care provider]

Finally, financial aid was recognized as an important area for support post–COVID-19, largely expressed by patients and recognized by health care workers. Patients reported that their daily functioning is worse compared with their functioning before COVID-19. Financial assistance was requested for those who can no longer work due to their complications post–COVID-19:

> *… he said he is no longer going to work. His salary is no longer the same and also our sister here, she her health is no longer the same. I think they are supposed to approve her for grant.* [Female, patient in FGD]

Taking into account our findings from both the quantitative and qualitative data, there was a strong emphasis on mental health support, support groups, and structured medical and rehabilitation follow-up post–COVID-19.

Design and Implementation of the Intervention

Using the information obtained from the situation analysis and existing evidence on post–COVID-19 rehabilitation, we developed an in-person support group program for people with post–COVID-19 complications titled CoPilot. The main study objective was to assess the change in physical and psychological well-being in patients with post–COVID-19 sequelae before and after the pilot intervention. To guide the program and evaluation we present a theory of change (Figure 1). The main intended impact of the intervention is to improve the overall health and well-being of COVID-19 survivors.

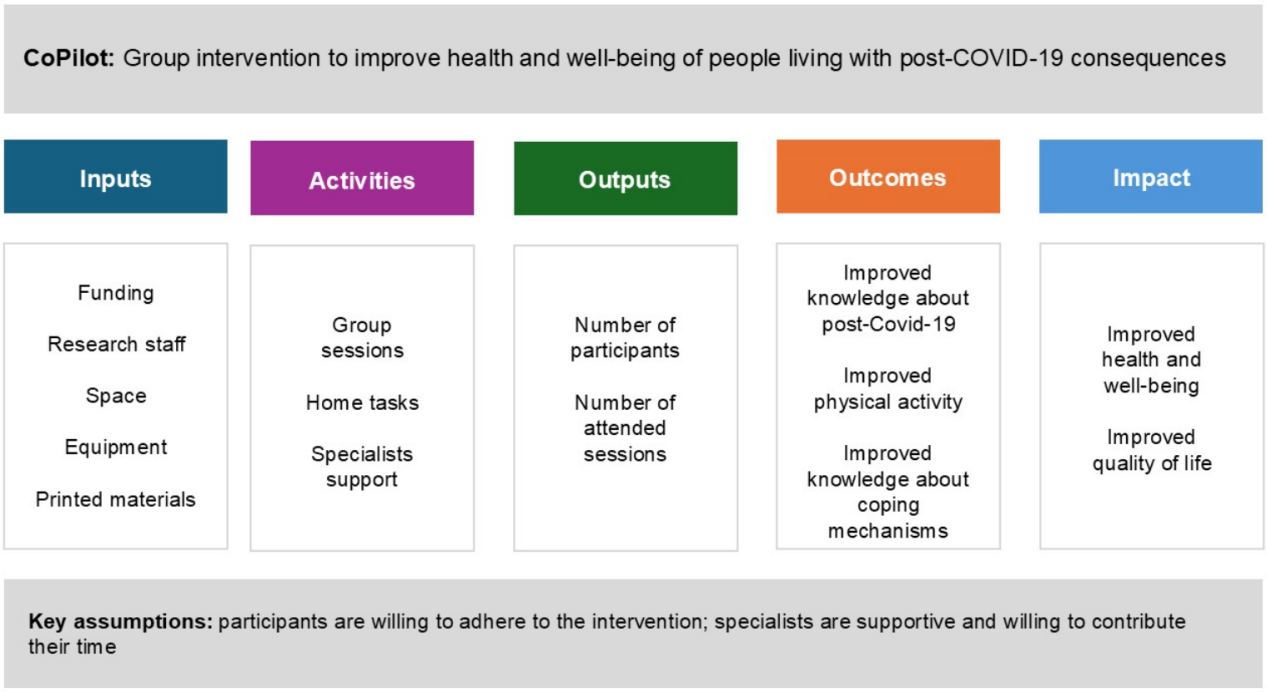
The target population for the intervention was individuals who had been diagnosed with COVID-19 and were experiencing ongoing post–COVID-19 complications according to the WHO post–COVID-19 condition definition [3]. The participants were recruited from out-patient medical and rehabilitation departments at the health care facility, including both patients and health care providers, and further participants were recruited by snowball sampling. Eligibility criteria included: self-reported or diagnosed previous COVID-19; adults aged 18-years and older; reporting at least one symptom lasting for more than 2 months, at least 3 months since COVID-19, which could not otherwise be explained by another condition; and willingness to provide written informed consent and to follow study procedures. Reported symptoms could include, but were not limited to, fatigue, breathlessness, pain or discomfort, impaired sleep, or impaired usual activity [2]. Participants with known severe medical or psychiatric conditions which affected ability to consent or participate, or required a higher level of care were excluded.

Group formats in rehabilitation have been shown to strengthen curative factors like universality and altruism, to provide peer and social support, to share coping strategies, and to practice gratitude [28,29]. Taking this into account, with the strong findings for group support in post–COVID-19 rehabilitation from our situation analysis, we designed and conducted a group-based intervention. The program consisted of weekly in-person group sessions over eight-weeks complimented by home-tasks and self-management exercises, with the option to adapt tasks to each person's abilities. We had 7 groups where each group consisted of 8 to 10 participants. Our target sample size was 60, which was informed by evidence that sample sizes of 24 - 50 are sufficient to estimate the key parameters of an efficacy and feasibility trial [30]. Sessions were guided by the predesigned schedule and a duration of one-hour, but with flexibility for participant experiences and needs. The core of the sessions involved research staff facilitating discussion around participants' post–COVID-19 experiences, challenges and self-management strategies whilst encouraging group support and education on post–COVID-19 complications. Select sessions included invited rehabilitation specialists (eg, psychologist, physiotherapist, and dietician) for additional expertise. If any participant missed a session, they were contacted telephonically to explain the weekly tasks. If participants missed a session and were not contactable, they were regarded as lost-to-follow-up.

The content of the intervention was guided by the findings from our situation analysis, the WHO Brochure "Support for rehabilitation: self-management after COVID-19-related illness [31] and from the HOPE intervention for people with post–Covid-19 conducted by Wright et al [21]. The detailed description of the session's content, weekly themes, evidence that informed the themes are shown in Table 2 along with the planned duration of the session and home-tasks.

XSL•FO
RenderX

Figure 1. Theory of change for CoPilot, demonstrating the post–COVID-19 intervention development plan adapted to the theory of change framework for individuals with post–COVID-19 complications.



XSL•FO
RenderX

Table . Intervention modules and content.

Module	Content	Evidence-informing theme	Format: 1 hour	Home-task
Week 1: post–COVID-19 complications introduction	• Welcome and introductions. • Overview of the program. • Pacing and goal setting. • Breathing exercise introduction.	• Quantitative findings on lack of post–COVID-19 knowledge. HOPE intervention[a]. Supported by WHO[b] brochure.	• Group session facilitated by research staff – discussions, home tasks.	• Self-test: how are you feeling? • Goal setting: SMARTER tool (general)[c]. • Long–COVID-19 education. • Breathing exercise.
Week 2: symptoms and physical activity	• Main symptom complaints. • Exercise routine and demonstration. • Setting physical goals. • Share experience and symptom management. • Fatigue management.	• Quantitative findings on lack of post-COVID-19 knowledge, and goals for post-COVID-19 recovery to be able to carry out activities of daily living better and have better physical capacity. Qualitative finding on providing post-COVID-19 symptom follow-up. Supported by HOPE intervention and WHO brochure.	• Group session with physiotherapist – discussion, exercise demonstration, home tasks.	• Exercises self-moderated. • SMARTER goal setting (physical ability focus).
Week 3: coping strategies	• Prioritizing and taking breaks. • Relaxation. • Mindfulness practices to manage stress.	• Quantitative findings on request for psychological support, goal for post–COVID-19 recovery of better coping strategies. Qualitative findings on need for mental health support.	• Group session facilitated by research staff – discussions, guided meditation.	• SMARTER goal setting (relaxation focus). • Relaxation and mindfulness tasks. • Gratitude diary.
Week 4: support and grief	• Debriefing. • Grief process and tracker. • Finding support. • Self-compassion and prioritization. • Asking for and accepting help.	• Quantitative findings on request for psychological support, goal for post–COVID-19 recovery of better coping strategies, and better connection with people and family. Qualitative findings on need for mental health support.	• Group session with psychologist – discussion and home tasks.	• Grief tracker. • Speaking to family or a friend. • Self-compassion task. • Self-test: how are you feeling?
Week 5: sleep and mindfulness	• Sleep disorder information. • Tips for better sleep. • Mundane task focusing.	• Quantitative finding of goal for post–COVID-19 recovery for better coping strategies. Supported by HOPE intervention.	• Group session facilitated by research staff – discussions, home tasks.	• Sleep diary. • SMARTER goal setting (sleep). • Mundane task focusing.
Week 6: move and eat better	• Importance of keeping active. • Eating well for physical and mental health. • Continuing exercises.	• Quantitative findings on goals for post–COVID-19 recovery to be able to carry out activities of daily living better and have better physical capacity. Supported by HOPE intervention and WHO brochure.	• Group session with dietician – discussion and home tasks.	• SMARTER goal setting (nutrition). • Home exercises and stretching. • What makes you happy?

XSL•FO
RenderX

Module	Content	Evidence-informing theme	Format: 1 hour	Home-task
Week 7: happiness and strengths	• Long COVID-19 and mental health education. • Identifying strengths and how to use them. • Managing setbacks. • Reviewing, what makes you happy?	• Quantitative findings on request for psychological support, goal for post–COVID-19 recovery of better coping strategies. Qualitative findings on need for mental health support.	• Group session facilitated by research staff – discussions and home tasks.	• Gratitude diary. • SMARTER goal setting (strengths and happiness). • Long–COVID-19 and mental health education.
Week 8: looking forward	• Hopes and aspirations. • Keeping in touch with support. • Review of the program. • Planning pleasant activities.	• Quantitative findings on request for psychological support, goal for post–COVID-19 recovery of better coping strategies, and better connection with family. Qualitative findings on need for mental health support. Supported by HOPE intervention.	• Group session facilitated by research staff – discussions and home tasks.	• Self-test: how are you feeling? • SMARTER goal setting (hopes for the future). • Moving forward with hope.

[a]HOPE intervention refers to Hope Programme for Long COVID-19 [21].

[b]WHO: World Health Organization.

[c]The SMARTER goal-setting tool is used to define goals following the criteria: specific, measurable, achievable, relevant, time-bound, exciting, and reviewable.

Monitoring and Evaluation Plan

Monitoring of the intervention was done using a participant registry, discussions at each session to review understanding of the content and tasks, and completion of a questionnaire to assess adherence to tasks and subjective progress.

The evaluation of the intervention was performed by measuring effectiveness (pre-and postintervention assessment) and a process evaluation (including feasibility and acceptability) of the group intervention. The main effectiveness outcomes included: (1) lung function, (2) mental well-being, and (3) quality of life. Data collection included baseline information on demographics and knowledge of COVID-19 and sequelae. The pre-post-intervention assessments included questionnaires on COVID-19 symptoms, fatigue and sleepiness scales [21,32,33], quality of life using the EuroQoL 5-Dimension 5-Level (EQ-5D-5L) [34], short Warwick-Edinburgh Mental Wellbeing Scale to describe mental well-being [35], lung function testing by means of spirometry, and functional capacity by 1-minute sit-to-stand test [36]. The process evaluation was guided by the Medical Research Council framework to assess both the implementation process (how the project is delivered), what is delivered (fidelity, dose, adaptations, or reach) and contextual factors influencing the implementation and acceptability of the intervention [37]. We used mixed methods to assess feasibility and acceptability of the intervention through a questionnaire and semistructured interviews with participants to elaborate on the design of the intervention, the potential barriers, and facilitators. Data from the intervention will be analyzed using descriptive statistics producing frequencies, means, and proportions of the different variables of interest. Outcomes will be analyzed using paired sample *t* tests. Qualitative process evaluation data will be transcribed and analyzed by thematic analysis.

Discussion

Principal Findings

To the best of our knowledge, this is the first study to develop and implement a context-adapted group intervention for post–COVID-19 survivors in South Africa. It was aimed to address debilitating consequences of COVID-19 and provide people with the tools to manage their symptoms to improve every-day quality of life and potentially health outcomes including mental and lung health. This situation analysis and intervention development highlighted 2 main points: first, engaging with patients and health care workers on their perspectives on post–COVID-19 care is an integral component in developing management strategies for post–COVID-19 complications, and second, that mental health support should strongly be considered in rehabilitation guidelines post–COVID-19 while considering support groups as a component.

Engaging with health care providers and patients' perspectives on post–COVID-19 complications and the management thereof is a crucial step in guideline development. The cornerstone to this is understanding the patient experience of post–COVID-19 complications and leveraging on existing and innovative strategies for care [21,25]. Fatigue is a common symptom reported post–COVID-19; however, it has been described by patients to have physical, mental, and emotional components [6,21], thereby illustrating that understanding the problem on a deeper level can help direct health services toward patient needs, as well as develop suitable assessment measures for

XSL•FO
RenderX

patient-reported outcomes [9]. Understanding patient and health care provider perspectives and communication between them has been shown to be important in HIV care, specifically relating to social, mental health, and health care system challenges [38]. In addition, coworking and partnership with shared responsibility in care between health care provider and patient with a focus on self-management principles were recognized as key features in the management process of other chronic lung diseases [24]. Taking into account the health resource limitations in LMICs and the strain on insufficient rehabilitation services poses another challenge to the management of post–COVID-19 complications, emphasizing the need for context-adapted guidelines [39].

The importance of the need for mental health support was a major theme that came through during our situation analysis. Mental health symptoms have been reported in up to 21% of people post–COVID-19 and can include anxiety, depression, sleep disorders, and neurocognitive disorders [4,6]. In a study by Rofail et al [9], it was reported that despite no clear psychological diagnosis, most impacts on daily life were not attributed to any singular symptom, but rather to the holistic experience of COVID-19 infection, further supporting the need to support psychological processing in recovery. Findings from our situation analysis clearly showed the gaps in psychological support for patients and health care providers post–COVID-19 and that these services are sought after as an adjunct to post–COVID-19 care. More specifically, support groups were suggested as helpful platforms to engage in psychological support, facilitating connection and shared experiences. This focus in care has been shown to be valued and helpful in support of patients with other chronic lung conditions [24].

One limitation is that this was a narrow population, where participants were only recruited from 2 facilities; however, to overcome this we recruited from multiple departments and recruited both health care providers and patients in order to increase variation. Another limitation was the heterogeneity of the post–COVID-19 sequelae definition, where interpretation conducted by the study team was to meet the WHO post–COVID-19 condition definition. This could result in bias of either overestimating or underestimating post–COVID-19 consequences, depending on alternate definitions used in other contexts. Despite this, consistency was maintained throughout recruitment for the situation analysis and planned intervention, and symptoms were clearly documented.

The strengths of the study include the context and population of the study where post–COVID-19 rehabilitation has not been studied previously. Another strength includes the mixed-methods approach for both the situation analysis and intervention in order to obtain deeper insight into the rehabilitation post–COVID-19.

Conclusion

Understanding the characteristics and burden of post–COVID-19 complications, particularly from the patient and health care provider perspectives, helps to direct post–COVID-19 management principles to ensure a more holistic recovery. As more people recover from COVID-19, it becomes crucial to maximize on the opportunity to better understand postviral complications and adapted health service allocation, particularly in LMICs. We believe that our group intervention holds a potential to improve health and well-being of people living with post–COVID-19 complications in Johannesburg, and its effectiveness and feasibility will be analyzed in the next steps.

Acknowledgments

We would like to acknowledge Ency Mahlase, an additional member of the research team, who assisted with analyzing the data. We are grateful to the Bavarian State Ministry of Science and Arts for their funding of this project. We thank Chris Hani Baragwanath Academic Hospital and Tembisa Tertiary Hospital for welcoming us to work at their site and supporting us in our participant recruitment and follow-up. We thank all the participants for their time and support. Generative artificial intelligence was not used in any part of the development of this manuscript.

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

NAG and OI contributed to the literature search. OI, FS, AMM, SC, and AR contributed to the study funding acquisition. NAG, OI, FS, SC, SAB, and AR contributed to the conceptualization, methodology, and project administration. NAG, OI, AR, ST, PM, and NM contributed to the data analysis and interpretation. NAG, OI, and FS contributed to the intervention design. NAG and OI drafted the manuscript, and all authors reviewed the manuscript critically for intellectual content. All coauthors approved the final draft of the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire guide.
[DOCX File, 35 KB - formative_v9i1e58436_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Focus group discussion guide.
[DOCX File, 17 KB - formative_v9i1e58436_app2.docx]

Multimedia Appendix 3
Clinical staff interview guide.
[DOCX File, 19 KB - formative_v9i1e58436_app3.docx]

References

1. Groff D, Sun A, Ssentongo AE, et al. Short-term and long-term rates of postacute sequelae of SARS-CoV-2 infection: a systematic review. JAMA Netw Open 2021 Oct 1;4(10):e2128568. [doi: 10.1001/jamanetworkopen.2021.28568] [Medline: 34643720]
2. O'Mahoney LL, Routen A, Gillies C, et al. The prevalence and long-term health effects of long COVID among hospitalised and non-hospitalised populations: a systematic review and meta-analysis. EClinicalMedicine 2023 Jan;55:101762. [doi: 10.1016/j.eclinm.2022.101762] [Medline: 36474804]
3. A clinical case definition of post COVID-19 condition by a Delphi consensus. World Health Organization. URL: https://iris.who.int/bitstream/handle/10665/345824/WHO-2019-nCoV-Post-COVID-19-condition-Clinical-case-definition-2021.1-eng.pdf?sequence=1&isAllowed=y [accessed 2023-08-22]
4. Lopez-Leon S, Wegman-Ostrosky T, Perelman C, et al. More than 50 long-term effects of COVID-19: a systematic review and meta-analysis. Sci Rep 2021 Aug 9;11(1):16144. [doi: 10.1038/s41598-021-95565-8] [Medline: 34373540]
5. Global Burden of Disease Long COVID Collaborators, Wulf Hanson S, Abbafati C, et al. Estimated global proportions of individuals with persistent fatigue, cognitive, and respiratory symptom clusters following symptomatic COVID-19 in 2020 and 2021. JAMA 2022 Oct 25;328(16):1604-1615. [doi: 10.1001/jama.2022.18931] [Medline: 36215063]
6. Natarajan A, Shetty A, Delanerolle G, et al. A systematic review and meta-analysis of long COVID symptoms. Syst Rev 2023 May 27;12(1):88. [doi: 10.1186/s13643-023-02250-0] [Medline: 37245047]
7. Anaya JM, Rojas M, Salinas ML, et al. Post-COVID syndrome. A case series and comprehensive review. Autoimmun Rev 2021 Nov;20(11):102947. [doi: 10.1016/j.autrev.2021.102947] [Medline: 34509649]
8. Parotto M, Gyöngyösi M, Howe K, et al. Post-acute sequelae of COVID-19: understanding and addressing the burden of multisystem manifestations. Lancet Respir Med 2023 Aug;11(8):739-754. [doi: 10.1016/S2213-2600(23)00239-4] [Medline: 37475125]
9. Rofail D, McGale N, Podolanczuk AJ, et al. Patient experience of symptoms and impacts of COVID-19: a qualitative investigation with symptomatic outpatients. BMJ Open 2022 May 2;12(5):e055989. [doi: 10.1136/bmjopen-2021-055989] [Medline: 35501077]
10. Tran VT, Riveros C, Clepier B, et al. Development and validation of the long coronavirus disease (COVID) symptom and impact tools: a set of patient-reported instruments constructed from patients' lived experience. Clin Infect Dis 2022 Jan 29;74(2):278-287. [doi: 10.1093/cid/ciab352] [Medline: 33912905]
11. Dryden M, Mudara C, Vika C, et al. Post-COVID-19 condition 3 months after hospitalisation with SARS-CoV-2 in South Africa: a prospective cohort study. Lancet Glob Health 2022 Sep;10(9):e1247-e1256. [doi: 10.1016/S2214-109X(22)00286-8] [Medline: 35961348]
12. Jassat W, Cohen C, Tempia S, et al. Risk factors for COVID-19-related in-hospital mortality in a high HIV and tuberculosis prevalence setting in South Africa: a cohort study. Lancet HIV 2021 Sep;8(9):e554-e567. [doi: 10.1016/S2352-3018(21)00151-X] [Medline: 34363789]
13. Glover N, Sathar F, Charalambous S, Ivanova O, Riess F, Shambhu RR, et al. Respiratory outcomes post-hospitalisation for symptomatic covid-19 disease in patients from a tertiary facility, South Africa. Presented at: Proceedings of the 7th South African TB Conference; Sep 13-16, 2022; Durban, South Africa.
14. Galal I, Hussein A, Amin MT, et al. Determinants of persistent post-COVID-19 symptoms: value of a novel COVID-19 symptom score. Egypt J Bronchol 2021 Dec;15(1). [doi: 10.1186/s43168-020-00049-4]
15. Munblit D, Nicholson T, Akrami A, et al. A core outcome set for post-COVID-19 condition in adults for use in clinical practice and research: an international Delphi consensus study. Lancet Respir Med 2022 Jul;10(7):715-724. [doi: 10.1016/S2213-2600(22)00169-2] [Medline: 35714658]
16. Alwan NA. Surveillance is underestimating the burden of the COVID-19 pandemic. Lancet 2020 Sep 5;396(10252):e24. [doi: 10.1016/S0140-6736(20)31823-7] [Medline: 32861312]
17. Demeco A, Marotta N, Barletta M, et al. Rehabilitation of patients post-COVID-19 infection: A literature review. J Int Med Res 2020 Aug;48(8):300060520948382. [doi: 10.1177/0300060520948382] [Medline: 32840156]
18. COVID-19 rapid guideline: managing the long-term effects of COVID-19. National Institute for Health and Care Excellence. 2024. URL: https://www.nice.org.uk/guidance/ng188/chapter/5-Management [accessed 2024-02-08]
19. Liu K, Zhang W, Yang Y, Zhang J, Li Y, Chen Y. Respiratory rehabilitation in elderly patients with COVID-19: a randomized controlled study. Complement Ther Clin Pract 2020 May;39:101166. [doi: 10.1016/j.ctcp.2020.101166] [Medline: 32379637]

XSL•FO
RenderX

20. Nambi G, Abdelbasset WK, Alrawaili SM, et al. Comparative effectiveness study of low versus high-intensity aerobic training with resistance training in community-dwelling older men with post-COVID 19 sarcopenia: a randomized controlled trial. Clin Rehabil 2022 Jan;36(1):59-68. [doi: 10.1177/02692155211036956] [Medline: 34344230]
21. Wright H, Turner A, Ennis S, et al. Digital peer-supported self-management intervention codesigned by people with long COVID: mixed methods proof-of-concept study. JMIR Form Res 2022 Oct 14;6(10):e41410. [doi: 10.2196/41410] [Medline: 36166651]
22. Oates GR, Niranjan SJ, Ott C, et al. Adherence to pulmonary rehabilitation in COPD: a qualitative exploration of patient perspectives on barriers and facilitators. J Cardiopulm Rehabil Prev 2019 Sep;39(5):344-349. [doi: 10.1097/HCR.0000000000000436] [Medline: 31348127]
23. Blackstock FC, Evans RA. Rehabilitation in lung diseases: "Education" component of pulmonary rehabilitation. Respirology 2019 Sep;24(9):863-870. [doi: 10.1111/resp.13582] [Medline: 31099164]
24. Kelly CA, Tsang A, Lynes D, Spencer S. "It's not one size fits all": a qualitative study of patients' and healthcare professionals' views of self-management for bronchiectasis. BMJ Open Respir Res 2021 Mar;8(1):e000862. [doi: 10.1136/bmjresp-2020-000862] [Medline: 33664124]
25. Jassat W, Reyes LF, Munblit D, et al. Long COVID in low-income and middle-income countries: the hidden public health crisis. Lancet 2023 Sep 30;402(10408):1115-1117. [doi: 10.1016/S0140-6736(23)01685-9] [Medline: 37652071]
26. Jassat W, Abdool Karim SS, Mudara C, et al. Clinical severity of COVID-19 in patients admitted to hospital during the omicron wave in South Africa: a retrospective observational study. Lancet Glob Health 2022 Jul;10(7):e961-e969. [doi: 10.1016/S2214-109X(22)00114-0] [Medline: 35597249]
27. Hoddinott P, Allan K, Avenell A, Britten J. Group interventions to improve health outcomes: a framework for their design and delivery. BMC Public Health 2010 Dec 31;10:800. [doi: 10.1186/1471-2458-10-800] [Medline: 21194466]
28. Wood AM, Maltby J, Gillett R, Linley PA, Joseph S. The role of gratitude in the development of social support, stress, and depression: two longitudinal studies. J Res Pers 2008 Aug;42(4):854-871. [doi: 10.1016/j.jrp.2007.11.003]
29. Larsen DJ, King RL, Stege R, Egeli NA. Hope in a strengths-based group activity for individuals with chronic pain. Couns Psychol Q 2015 Apr 3;28(2):175-199. [doi: 10.1080/09515070.2015.1007444]
30. Teresi JA, Yu X, Stewart AL, Hays RD. Guidelines for Designing and Evaluating Feasibility Pilot Studies. Med Care 2022 Jan 1;60(1):95-103. [doi: 10.1097/MLR.0000000000001664] [Medline: 34812790]
31. Support for rehabilitation self-management after COVID-19-related illness. : World Health Organization Regional Office for Europe; 2021 URL: https://www.who.int/publications/m/item/support-for-rehabilitation-self-management-after-covid-19-related-illness [accessed 2023-03-01]
32. Shahid A, Wilkinson K, Marcu S, Shapiro CM. Fatigue severity scale (FSS). In: In: STOP, THAT and One Hundred Other Sleep Scales: Springer; 2011:167-168. [doi: 10.1007/978-1-4419-9893-4_35]
33. Johns M. About the ESS. The Epworth Sleepiness Scale. URL: https://epworthsleepinessscale.com/about-the-ess/ [accessed 2023-11-09]
34. EQ-5D-5L. EuroQoL. 2024. URL: https://euroqol.org/information-and-support/euroqol-instruments/eq-5d-5l/ [accessed 2023-06-10]
35. The Warwick-Edinburgh Mental Wellbeing Scales - WEMWBS. Warwick Medical School. 2021. URL: https://warwick.ac.uk/fac/sci/med/research/platform/wemwbs [accessed 2023-06-20]
36. Vaidya T, Chambellan A, de Bisschop C. Sit-to-stand tests for COPD: a literature review. Respir Med 2017 Jul;128:70-77. [doi: 10.1016/j.rmed.2017.05.003] [Medline: 28610673]
37. Moore GF, Audrey S, Barker M, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015 Mar 19;350:h1258. [doi: 10.1136/bmj.h1258] [Medline: 25791983]
38. Bogart LM, Chetty S, Giddy J, et al. Barriers to care among people living with HIV in South Africa: contrasts between patient and healthcare provider perspectives. AIDS Care 2013;25(7):843-853. [doi: 10.1080/09540121.2012.729808] [Medline: 23061894]
39. Ngeh EN, Chigbo NN, Whitehouse Z, et al. A report on the development of COVID-19 guidelines for rehabilitation professionals in African settings. Pan Afr Med J 2021;38:129. [doi: 10.11604/pamj.2021.38.129.26311] [Medline: 33912299]

Abbreviations

ADL: activities of daily living
CHBAH: Chris Hani Baragwanath Academic Hospital
EQ-5D-5L: EuroQoL 5-Dimension 5-Level
FGD: focus group discussion
LMIC: low- and middle-income country
NICE: National Institute for Health and Care Excellence
TB: tuberculosis
WHO: World Health Organization

XSL•FO
RenderX

Edited by A Mavragani; submitted 16.03.24; peer-reviewed by A AL-Asadi, J Schulze, S Markham; revised version received 17.02.25; accepted 02.03.25; published 10.04.25.

Please cite as:
Glover NA, Sathar F, Mokome P, Mathabela N, Taleni S, van Blydenstein SA, Mekota AM, Charalambous S, Rachow A, Ivanova O Improving Health and Well-Being of People With Post–COVID-19 Consequences in South Africa: Situation Analysis and Pilot Intervention Design
JMIR Form Res 2025;9:e58436
URL: https://formative.jmir.org/2025/1/e58436
doi:10.2196/58436

© Nicole Audrey Glover, Farzana Sathar, Pride Mokome, Nkululeko Mathabela, Sipokazi Taleni, Sarah Alexandra van Blydenstein, Anna-Maria Mekota, Salome Charalambous, Andrea Rachow, Olena Ivanova. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impact of Acute Respiratory Infections on Medical Absenteeism Among Military Personnel: Retrospective Cohort Study

Premikha M[1], MPH, MBBS; Jit Khong Goh[1,2], MPH, MBBS; Jing Qiang Ng[1], MBBS; Adeliza Mutalib[1], MSc; Huai Yang Lim[1], MPH, MBBS

[1]Force Health Group, HQ Medical Corps, Singapore Armed Forces, 701 Transit Road, #02-03, Singapore, Singapore

[2]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

Corresponding Author:
Premikha M, MPH, MBBS
Force Health Group, HQ Medical Corps, Singapore Armed Forces, 701 Transit Road, #02-03, Singapore, Singapore

Abstract

Background: Acute respiratory infections (ARI) are a significant challenge in military settings due to close communal living, which facilitates the rapid transmission of pathogens. A variety of respiratory pathogens contribute to ARI, each varying in prevalence, severity, and impact on organizational productivity. Understanding and mitigating the impact of ARI is critical for optimizing the health of military personnel and maintaining organizational productivity.

Objective: This retrospective study of surveillance data aims to identify pathogens causing ARI among servicemen and determine which pathogens contribute most to medical absenteeism, defined as the combined duration of the issued medical certificate and light duty.

Methods: From September 2023 to August 2024, anonymous nasopharyngeal swabs (BioFire FilmArray Respiratory Panel) were collected from Singapore Armed Forces servicemen presenting with ARI symptoms after a doctor's consultation at a local military camp's medical centre. The presence of fever and duration of medical certificate and light duty were self-reported by Singapore Armed Forces servicemen.

Results: A total of 1095 nasopharyngeal swabs were collected, of which 608 (55.5%) tested positive. The most common respiratory pathogen was human rhinovirus/enterovirus (HRV/HEV) in 303 (27.7%) individuals. The highest proportions of fever were observed in servicemen with influenza (62.8%, 27/43), SARS-CoV-2 (34.3%, 12/35), and parainfluenza (31.6%, 12/38). The odds of patients with influenza that have fever was 5.8 times higher than those of patients infected with HRV/HEV (95% CI 2.95 - 11.40, P<.001). The median duration of medical certificate, light duty, and medical absenteeism were 0 (IQR 0), 2 (IQR 2) and 2 (IQR 0) days, respectively. The odds of patients with influenza having a medical certificate with duration ≥1 day was 5.34 times higher than those in patients with HRV/HEV (95% CI 2.63 - 10.88, P<.001). No significant differences in the duration of medical absenteeism were found between HRV/HEV and other pathogens.

Conclusions: Compared to HRV/HEV, influenza infections were significantly associated with longer medical certificate duration. Nonetheless, there were no significant differences in the overall duration of medical absenteeism across pathogens, as servicemen infected with other pathogens were given light duty instead. These findings emphasize the need for pathogen-agnostic ARI measures. While influenza vaccinations are already mandatory for servicemen in local military camps, encouraging additional public health measures (eg, mask-wearing among symptomatic servicemen, COVID-19 vaccinations, therapeutics) can further reduce ARI incidence, minimize the duration of medical absenteeism, and mitigate the impact on organizational productivity.

(JMIR Form Res 2025;9:e69113) doi:10.2196/69113

KEYWORDS

respiratory infections; military; epidemiology; public health; surveillance

Introduction

Background

Acute respiratory infections (ARI) pose a significant challenge in military settings like the Singapore Armed Forces (SAF), where close quarters living and shared communal spaces, equipment, and furniture could facilitate the rapid transmission of respiratory pathogens [1,2]. Singapore's high humidity and tropical climate further exacerbate these risks, promoting the survival and spread of respiratory viruses [3]. These conditions not only increase ARI risk and impact the health of military personnel but also reduce organizational productivity due to medical absenteeism [1]. Despite ongoing surveillance, gaps remain in understanding how specific ARI pathogens impact medical absenteeism and readiness in high-density military environments.

XSL•FO
RenderX

A comprehensive understanding of the full spectrum of ARI pathogens is critical. While influenza is known for causing significant morbidity and mortality, HRV is typically the most prevalent pathogen and can lead to chronic infections and secondary bacterial infections, particularly in immunocompromised individuals [4]. A population-wide study in the United States found that the hospitalization burden was higher for HRV/HEV compared to influenza among adults aged <65 years [5,6]. Additionally, bacterial pathogens such as *Streptococcus pneumoniae* and *Mycoplasma pneumoniae* can act as primary agents or secondary infections [7,8]. A 2016 study in Singapore military recruits further emphasized the diversity of ARI pathogens, with 41% of 2647 recruits presenting with febrile respiratory illness (FRI) were predominantly caused by adenovirus and influenza viruses [9]. Among the remaining nonfebrile ARI individuals, rhinovirus accounted for 47% [9]. These findings highlight the need for expanded surveillance to capture the full spectrum of pathogens, beyond influenza to assess their contribution to the ARI burden in military settings.

Furthermore, it is unclear whether current influenza vaccination efforts sufficiently address the impact of influenza in military settings. Historical and contemporary data underscore the disproportionate impact of influenza on military populations, with attack rates reaching as high as 37 - 45% during outbreaks [10]. During the 1918 influenza pandemic, military personnel living in confined spaces faced significantly higher morbidity and mortality rates compared to civilian populations [10]. Recent outbreaks further highlight the heightened vulnerability of military personnel; a 2013 influenza A (H1N1) outbreak in a Swiss military camp infected 31% of soldiers, while a 2009 influenza A outbreak in a Peruvian Navy ship affected 22% of crew members [11,12]. Recognizing influenza as a significant cause of ARI with substantial morbidity, the SAF has conducted annual Northern Hemisphere influenza vaccination drives since 2015, achieving over 97% coverage in 2023 [2]. Additionally, the SAF conducts Southern Hemisphere influenza vaccinations for high-risk servicemen (eg, health care workers, medically vulnerable individuals), and strongly encourages them to receive additional doses of the COVID-19 vaccines, aligning with national guidelines [13]. Despite these measures, there is limited understanding of whether SAF's influenza vaccination programme sufficiently mitigate influenza's prevalence and severity in the SAF. These efforts are also insufficient to address the broader ARI burden, as evidenced by the prevalence of other ARI pathogens, highlighting the need for pathogen-specific data to understand the impact of various ARI pathogens on medical absenteeism.

Another critical gap is the limited understanding of the differential impact of febrile and nonfebrile ARI on medical absenteeism in military settings. The FRI, characterized by the presence of fever (≥38°C), serves as a marker of severity and reflects a robust immune response to more virulent pathogens [14]. Influenza and SARS-CoV-2 often cause FRI, whereas HRVs typically cause nonfebrile illnesses [3,15,16]. This distinction is crucial not only for clinical management but also for informing public health strategies aimed at controlling the spread of more severe FRI. However, findings from the Singapore military recruit study in 2016 showed that ARI cases without fever constituted the majority (59%) of respiratory illnesses and contributed significantly to the overall burden, emphasizing the need to account for both FRI and nonfebrile ARI in public health strategies [9]. Clinical management for ARI is mostly supportive, making vaccines critical for certain pathogens such as influenza and SARS-CoV-2. Recent events such as the 2023 ARI outbreak in the Philippines military that affected 20% of military students due to low influenza vaccine coverage [17], highlight the importance of maintaining high vaccine uptake. However, they also underscore the need to explore additional public health measures to reduce the broader ARI burden in military settings.

Study Objectives

Therefore, this study aimed to address these gaps by reviewing the results of the respiratory pathogen surveillance study from a local military camp to answer the following questions:

1. What is the distribution of pathogens responsible for ARI among SAF servicemen?
2. Which pathogens had the greatest impact on duration of medical absenteeism and how can this inform public health interventions to mitigate the impact?

By addressing these questions, this study seeks to determine if additional public health measures need to be instituted at local camps, beyond the existing influenza vaccinations. Understanding and mitigating the impact of ARI is critical for reducing medical absenteeism and optimizing organizational productivity.

Methods

Study Design and Recruitment

This retrospective study of surveillance data analyzed ARI cases from September 2023 to August 2024 at the medical centre of a local military camp in Singapore, the primary health care facility for servicemen. The camp's population primarily consists of young adult males, both conscripted and regular military personnel, who live in high-density accommodations conducive to ARI transmission. The medical centre typically manages 100-150 servicemen per day with acute conditions. The study used a convenience sampling approach, with recruitment conducted solely at the medical centre. Doctors screened servicemen presenting with ARI symptoms (eg, cough, sore throat, rhinorrhea) during routine consultations. Eligible servicemen were provided a standardized explanation of the study and invited to participate voluntarily. Those who agreed to participate were directed to the study team for data collection. Inclusion criteria included servicemen who (1) presented with at least one ARI symptom and (2) be willing to provide nasopharyngeal swabs for diagnostic testing. Exclusion criteria included servicemen diagnosed with other acute concurrent illnesses or those unwilling to provide swabs.

Anonymized nasopharyngeal swabs were collected from all eligible participants by the study team and tested using the BioFire FilmArray Respiratory Panel (BioFire Diagnostics), a multiplex polymerase chain reaction assay capable of detecting 22 respiratory pathogens. These pathogens were grouped into:

XSL•FO
RenderX

(1) coronaviruses (coronavirus 229E, HKU1, NL63, OC43), (2) SARS-CoV-2, (3) human metapneumovirus, (4) HRV/HEV, (5) influenza (influenza A, A/H1, A/H3, A/H1-2009, B), (6) parainfluenza (parainfluenza 1, 2, 3, 4), (7) respiratory syncytial virus, (8) adenovirus, and (9) bacteria (ie, *Bordetella parapetussis*, *Bordetella pertussis*, *Chlamydia pneumoniae*, *Mycoplasma pneumoniae*) for analysis. Pathogen testing was conducted immediately after swab collection using our mobile BioFire systems, which were serviced annually to minimize testing errors. The presence of fever (≥38°C) and duration of medical certificate and light duty were recorded through self-reported surveys administered directly by the surveillance team to ensure consistency in data collection. Our outcome of interest was duration of medical absenteeism, defined as the combined duration of medical certificate and light duty, representing the total period during which servicemen were unable to perform their primary jobs at full capacity. Servicemen with medical certificates were fully exempted from their duties, while those on light duty were partially exempted.

Statistical Analysis

Descriptive statistics were presented as mean (SD) for continuous variables and distribution (n, %) for categorical variables. Invalid swabs, coinfections, and bacterial cases were excluded from the analysis. Coinfections were defined as the presence of more than one pathogen detected in a single swab. Excluding coinfections aimed to minimize potential bias and ensured clearer attribution of outcomes (eg, fever, duration of medical absenteeism) to individual pathogens. Bacterial cases, which included those positive for bacterial pathogens such as *Mycoplasma pneumoniae*, were excluded due to the small number of cases detected during the study period.

Comparative analyses were performed to examine the differences in the prevalence of fever, duration of medical certificate, light duty, and duration of medical absenteeism among different pathogens using Pearson's χ^2 test of independence. Logistic regression was then performed for the outcomes of interest: (1) medical certificate (medical certificate ≥1 day), (2) light duty (>1 day), and (3) medical absenteeism (>1 day) to quantify outcome differences across pathogens. Human rhinovirus/human enterovirus was selected as the reference group due to its high prevalence in the study population, contributing to the stability of regression estimates. While "negative" cases could serve as an alternative reference group, they introduce heterogeneity, as "negative" indicates the absence of detectable pathogens by the BioFire Respiratory Panel, rather than the absence of infection. The Hosmer-Lemeshow goodness-of-fit test was used to ensure that all models characterized the data distribution accurately. Data were analyzed using Stata (version 17.0, StatCorp LLC) at a 5% significance level.

Ethical Considerations

This study was granted exemption from ethical review by the institutional review board (IRB) of the Defence Science Organisation–SAF (DSO-SAF IRB; reference no.: 0006/2024), as it was retrospective in nature, with only nonidentifiable parameters extracted and all data anonymized prior to recording and analysis. The IRB exemption allowed for secondary analysis without additional consent. All data were deidentified to ensure privacy and confidentiality, and no personally identifiable information or images of participants were included in this manuscript. No compensation was provided to the study participants.

Results

ARI Pathogen Distribution

From September 2023 to August 2024, a total of 1095 servicemen with ARI symptoms were included in the study. The median age of participants was 20 (IQR 2.0, range 17-31) years, with 89.9% of participants aged 18 - 21 years, reflecting a relatively homogenous age distribution. Additionally, 99.8% of participants were male. Of the 1095 nasopharyngeal swabs collected, 608 (55.5%) tested positive for at least one respiratory pathogen. The distribution of pathogens is shown in (Table 1). The most common pathogen was HRV/HEV (n=303, 27.7%). There were 69 (6.3%) coinfections; specifically, HRV/HEV co-occurred with other pathogens in 50 cases (72.4%), influenza co-occurred in 6 cases (8.7%), and both HRV/HEV and influenza co-occurred in 7 cases (10.1%). Coinfections were excluded from further analysis to minimize bias and ensure a clearer attribution of the study outcomes to specific pathogens. Additionally, *Mycoplasma pneumoniae* (n=3) and adenovirus (n=1) were excluded, along with 31 (2.83%) invalid swabs.

XSL•FO
RenderX

Table . Nasopharyngeal swab test results among servicemen with acute respiratory infections in the Singapore Armed Forces from September 2023 to August 2024.

Nasopharyngeal swab test results	Number of swabs (N=1095), n (%)[a]
Positive	608 (55.5)[b]
Coronavirus[c]	68 (3.2)
SARS-CoV-2	35 (5.8)
Human metapneumovirus	37 (3.4)
Human rhinovirus/enterovirus	303 (27.7)
Influenza	43 (3.9)
Parainfluenza	38 (3.5)
Respiratory synctial virus	11 (1.0)
Adenovirus	1 (0.1)
Bacteria[d]	3 (0.3)
Coinfections	69 (6.3)
Negative	456 (41.6)
Invalid	31 (2.8)

[a]The percentages do not add up to 100% due to rounding.

[b]The percentages of all pathogens in the Positive test results category add up to 55.5%.

[c]Coronavirus includes seasonal coronaviruses (229E, HKU1, NL63, OC43) but does not include SARS-CoV-2, which is analyzed and reported as a separate pathogen group.

[d]*Bordetella parapetussis*, *Bordetella pertussis*, *Chlamydia pneumoniae*, *Mycoplasma pneumoniae.*

Comparing the Prevalence of Fever (Severity) Across Respiratory Pathogen Groups

Among the 991 servicemen included for further analysis, fever was observed among 212 (21.4%). The highest proportions of fever were observed in servicemen with influenza (27/43, 62.8%), SARS-CoV-2 (12/35, 34.3%) and parainfluenza (12/38, 31.6%) (Table 2). There were significant differences in the proportions of servicemen with fever across the various pathogens (χ^2_7=65.9, *P*<.001). The odds of patients with influenza having fever were 5.8 times higher than those in patients infected with HRV/HEV (95% CI 2.95 - 11.40; *P*<.001). There were no statistically significant associations between the other pathogen types and fever.

Table . Prevalence of fever among servicemen with specific respiratory pathogen groups from a retrospective study on acute respiratory infections among servicemen from the Singapore Armed Forces from September 2023 to August 2024.

Respiratory pathogen groups	Febrile cases/total infected (N=991), n (%)
Influenza	27/43 (62.8)
SARS-CoV-2	12/35 (34.3)
Parainfluenza	12/38 (31.6)
Human metapneumovirus	11/37 (29.7)
Human rhinovirus/enterovirus	70/303 (23.1)
Coronavirus[a]	13/68 (19.1)
Negative	66/456 (14.5)
Respiratory synctial virus	1/11 (9.1)

[a]Coronavirus includes seasonal coronaviruses (229E, HKU1, NL63, OC43) but does not include SARS-CoV-2, which is analyzed and reported as a separate pathogen group.

Impact of ARI on Organizational Productivity

The median durations of medical certificates, light duty and medical absenteeism were 0 (IQR 0), 2 (IQR 2) and 2 (IQR 0) days, respectively.

- Medical certificate: There were significant differences in the proportions of servicemen with medical certificates ≥1 day across the various pathogens (χ^2_7=29, *P*<.001). The odds of patients with influenza having medical certificates ≥1 day was 5.34 times that of HRV/HEV patients (95% CI 2.63 - 10.88; *P*<.001).

XSL•FO
RenderX

- Light duty: There were no significant differences in the proportion of servicemen with light duty >1 day across the various pathogens (χ^2_7=13.3, *P*=.07). However, patients with influenza had significantly lower odds of having light duty >1 day compared to that of HRV/HEV patients (odds ratio [OR] 0.33, 95% CI 0.17 - 0.64); *P*=.001).
- Medical absenteeism: No significant differences in medical absenteeism were found between HRV/HEV and other pathogens (Table 3).

Table . Results of logistic regression of medical certificate ≥1 day, light duty >1 day and medical absenteeism >1 day on pathogens (n=991).

Outcomes and pathogens	Odds ratio (95% CI)	*P* value
Medical certificate ≥1 day		
Human rhinovirus/enterovirus	Reference	_[a]
Coronavirus	1.57 (0.75 - 3.31)	.23
SARS-CoV-2	1.69 (0.65 - 4.38)	.28
Human metapneumovirus	1.28 (0.47 - 3.51)	.63
Influenza	5.34 (2.63 - 10.88)	*<.001*
Parainfluenza	1.24 (0.45 - 3.40)	.68
Respiratory syncytial virus	1.82 (0.38 - 8.78)	.46
Negative	1.10 (0.69 - 1.74)	.69
Light duty >1 day		
Human rhinovirus/enterovirus	Reference	_[a]
Coronavirus	0.93 (0.54 - 1.61)	.80
SARS-CoV-2	0.60 (0.30 - 1.22)	.16
Human metapneumovirus	0.83 (0.41 - 1.69)	.61
Influenza	0.33 (0.17 - 0.64)	*.001*
Parainfluenza	0.78 (0.39 - 1.56)	.48
Respiratory syncytial virus	0.89 (0.25 - 3.10)	.85
Negative	0.91 (0.67 - 1.24)	.56
Medical absenteeism >1 day		
Human rhinovirus/human enterovirus	Reference	_[a]
Coronavirus	1.31 (0.68 - 2.56)	.41
SARS-CoV-2	0.68 (0.32 - 1.46)	.32
Human metapneumovirus	0.97 (0.44 - 2.15)	.94
Influenza	1.03 (0.48 - 2.19)	.94
Parainfluenza	0.87 (0.40 - 1.88)	.73
Respiratory syncytial virus	1.30 (0.30 - 6.64)	.67
Negative	0.98 (0.70 - 1.38)	.91

[a]Not applicable.

Discussion

Principal Findings

This study provides critical insights into the epidemiology of ARI in a military setting, specifically among SAF servicemen in Singapore. It addresses a key gap in understanding the contribution of different ARI pathogens to medical absenteeism and highlights actionable strategies to mitigate ARI-related disruptions in military contexts. Over the study period, HRV/HEV was the most frequently detected pathogen (27.7%), aligning with previous research findings. Influenza was significantly associated with higher odds of fever, longer medical certificate duration, but the overall duration of medical absenteeism was similar to that of HRV/HEV. Even after adjusting for fever, influenza remained significantly associated with longer medical certificate duration (OR 2.81, 95% CI 1.28 - 6.13, *P*=.01) compared to HRV/HEV, suggesting that other ARI symptoms (eg, sore throat, rhinorrhea), contribute to its more severe clinical presentation. This distinction is important as it provides evidence that influenza leads to more severe illness overall, beyond fever, compared to other ARI pathogens and makes servicemen more likely to receive medical certificates than light duties. Conversely, servicemen infected

XSL•FO
RenderX

with other ARI pathogens—perceived as less severe and not significantly associated with fever—are more likely assigned to light duties instead. This reflects clinical practice, where servicemen with more severe ARI are allowed to recuperate at home for a longer duration, while those with milder ARI are required to return to work earlier but allowed to perform only light duties. These findings reinforce the importance of continuing robust influenza vaccination programmes in the SAF.

Despite these findings, our study demonstrated no significant differences in duration of medical absenteeism across pathogens, reflecting that ARIs, regardless of pathogen type, contribute substantially to absenteeism and operational disruptions. This highlights the need for pathogen-agnostic strategies targeting ARI prevention and management, rather than focusing solely on influenza in the SAF. The importance of nonpharmaceutical intervention (NPI) is further underscored by historical examples, such as the 1996 influenza A (H3N2) outbreak aboard a US Navy ship [18]. Despite 95% of the crew being vaccinated against influenza, the antigenic distinctness of the circulating strain led to an infection rate of 42%, demonstrating the limitations of vaccination alone and the critical need for complementary measures such as isolation protocols, improved ventilation, and rapid response strategies to mitigate influenza outbreaks in confined military settings [18].

Globally, ARIs beyond influenza have consistently disrupted organizational productivity in military settings, emphasising the universal challenges posed by communal living and intense operational demands. Among troops of the Polish Military Contingent deployed overseas from 2003 to 2005, ARIs were the leading cause of outpatient morbidity, with incidence rates ranging from 45.6-54.8 cases per 100 soldiers, significantly disrupting training schedules and productivity [19]. Similarly, respiratory illnesses affected 69.1% of US military personnel deployed to Iraq and Afghanistan from 2003 to 2004, which often resulted in missed patrols and transient decreases in operational efficiency [20]. Notably, the incidence of ARI doubled during the combat phase, reflecting the heightened susceptibility of military personnel under physically and mentally demanding conditions [20]. Additionally, a systematic review of 46 military studies in 17 countries observed a higher incidence of COVID-19 in military personnel on deployment and those who had overseas exposure, amplifying the risks faced by military personnel where operational stressors and novel environmental exposures exacerbate the burden of ARIs, including COVID-19 [21]. These findings align with our study results, which showed no significant differences in medical absenteeism across pathogens, underscoring the need to adopt prevention strategies targeting the broader spectrum of ARIs rather than focusing primarily on influenza.

A multipronged approach to ARI prevention, integrating vaccination programmes, targeted NPIs and the strategic use of therapeutics, is essential for military personnel. Vaccination has proven highly effective; for instance, the H1N1-2009 influenza vaccination programme in Singapore reduced the weekly cases by 54% among military personnel post vaccination [22]. However, vaccines alone are insufficient, as demonstrated by our study findings. The US military has successfully implemented cohorting strategies in basic training facilities, dividing personnel into smaller, stable groups to limit the exposure and spread of ARI [1]. Complementing these NPIs, personal protective measures such as mandatory mask-wearing during outbreaks or in enclosed, high-density environments, coupled with educational campaigns on respiratory hygiene, could further reduce the ARI transmission [1]. Similarly, Poland's military experienced peaks in ARI during troop rotations, highlighting the need for stricter isolation protocols and improved acclimatization strategies for personnel entering new environments [19]. Insights from the systematic review on COVID-19 in military settings also emphasised the utility of digital tools, such as contact tracing applications, in effectively managing outbreaks [21]. Therapeutics are another critical component of ARI management. A study in Singapore during the H1N1 influenza outbreak in 2009 demonstrated that early use of oseltamivir significantly reduced the risk of hospitalization and severe outcomes, even in young and healthy military populations [23]. Enhanced vaccination coverage for pathogens with available vaccines (ie, influenza, SARS-CoV-2), rapid isolation of infected personnel, and timely therapeutic interventions collectively contribute to reducing the overall ARI burden. By focusing on overall respiratory health rather than targeting specific pathogens, these strategies will help create a more resilient military workforce.

The findings of our study, while focused on a single SAF military camp, have broader implications for the military and other high-density living settings globally. The consistent alignment of our findings with other military studies, such as those conducted in US, Polish, and Singaporean military contexts, highlights the relevance of our findings across varied geographic and operational contexts. Our study findings are generalizable to other SAF camps or military units with similar profiles, where young men live and work together in close quarters, as these shared conditions contribute significantly to ARI transmission. Additionally, these findings may offer insights for surveillance and ARI prevention measures in other high-density living environments, such as dormitories, where communal living and close interactions similarly increase the risk of ARI outbreaks. Expanding surveillance efforts to multiple camps with different demographics and training intensities, as well as integrating findings with civilian studies, could further enhance the applicability of these findings to both military and nonmilitary populations.

Strengths

The study used data from comprehensive surveillance with the BioFire FilmArray Respiratory Panel, which enabled the detection of a wide range of respiratory pathogens. The large sample size allowed for accurate determination of pathogen prevalence in the military context. Additionally, the study's focus on medical absenteeism bridges the gap between clinical outcomes and organizational impact, making the results actionable for health policy and military planning.

Limitations

First, excluding coinfections from the analysis may have overlooked the complexity of pathogen interactions. Coinfections often exacerbate illness severity or prolong recovery time, and their exclusion may have led to an

XSL•FO
RenderX

underestimation of the duration of MA. However, the number of coinfections in our study was relatively small, and their exclusion is unlikely to have significantly impacted the overall results or conclusions. Second, the small number of bacterial ARI cases limited our understanding of their impact on organizational productivity. Bacterial pathogens, such as *Mycoplasma pneumoniae*, are known to cause severe symptoms and potentially longer MA. Without sufficient data, their role in contributing to productivity losses remains unclear. Third, the reliance on self-reported data for fever and medical certificates or light duty duration introduces the potential for recall bias. Lastly, while our study is generalizable to SAF units or camps with similar profiles (ie, young men in close quarters), its findings may not apply to camps with different demographics or training intensities.

Future studies should address these limitations by including coinfection cases to better understand the interplay between different pathogens and their combined impact on ARI severity and duration of medical absenteeism. Expanding surveillance to multiple military camps and similar high-density living settings would enhance generalizability of findings and account for variations in living conditions, vocations, and operational structures. This broader approach could help identify high-risk groups or settings and guide the implementation of targeted interventions. Additionally, longitudinal studies tracking the same cohort over multiple ARI seasons could provide insights into the effectiveness of preventive measures, such as influenza vaccinations, NPIs, and the impact of digital health tools such as the upcoming Data Analytics for Soldier Health (DASH 2.0) system in the SAF. The integration of iDASH2.0 into future electronic medical records could enable real-time monitoring of ARI trends, identification of seasonal trends in pathogen prevalence, and assessing waning immunity post vaccination. Finally, studies focusing on bacterial ARI cases should explore their clinical and operational impact using advanced diagnostic tools and longitudinal study designs, paving the way for antibiotic stewardship programmes.

Conclusion

This study highlights the significant burden of ARIs in military settings, with influenza contributing to longer medical leave compared to noninfluenza pathogens, yet no differences in the overall duration of medical absenteeism across pathogens. These findings underscore the need for a multipronged approach to ARI prevention that goes beyond vaccination programmes to include targeted nonpharmaceutical interventions and therapeutics. By addressing the identified gaps and building on the strengths of this study, future research should explore the impact of coinfections and bacterial ARIs, seasonal trends, and the long-term effectiveness of preventive measures across multiple military camps in Singapore. This will enhance generalizability and inform robust, evidence-based policies to optimize the health and productivity of servicemen. Ultimately, integrating vaccination programmes, NPIs, and therapeutics into a cohesive ARI management framework will strengthen the SAF's resilience against respiratory threats and ensure sustained organizational productivity.

Acknowledgments

No generative AI tools were used in the drafting, analysis, or writing of this manuscript.

Data Availability

The dataset analyzed during this study is not publicly available due to the sensitive nature of military-related health data but are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: HYL, PM
Data curation: PM, JQN
Formal analysis: PM
Investigation: JQN, AM
Methodology: HYL, PM, JKG
Project administration: AM, JQN
Resources: JKG, AM, JQN
Supervision: HYL, JKG
Writing – original draft: PM
Writing – review & editing: PM, JKG, JQN, AM, HYL

Conflicts of Interest

None declared.

References

1. Sanchez JL, Cooper MJ, Myers CA, et al. Respiratory infections in the U.S. military: recent experience and control. Clin Microbiol Rev 2015 Jul;28(3):743-800. [doi: 10.1128/CMR.00039-14] [Medline: 26085551]

XSL•FO
RenderX

2. Ho ZJM, Hwang YFJ, Lee JMV. Emerging and re-emerging infectious diseases: challenges and opportunities for militaries. Mil Med Res 2014;1(1):21. [doi: 10.1186/2054-9369-1-21] [Medline: 25722877]
3. Burbank AJ. Risk factors for respiratory viral infections: a spotlight on climate change and air pollution. J Asthma Allergy 2023;16:183-194. [doi: 10.2147/JAA.S364845] [Medline: 36721739]
4. Jacobs SE, Lamson DM, St George K, Walsh TJ. Human rhinoviruses. Clin Microbiol Rev 2013 Jan;26(1):135-162. [doi: 10.1128/CMR.00077-12] [Medline: 23297263]
5. Herold S, Becker C, Ridge KM, Budinger GRS. Influenza virus-induced lung injury: pathogenesis and implications for treatment. Eur Respir J 2015 May;45(5):1463-1478. [doi: 10.1183/09031936.00186214] [Medline: 25792631]
6. Zimmerman RK, Balasubramani GK, D'Agostino HEA, et al. Population-based hospitalization burden estimates for respiratory viruses, 2015-2019. Influenza Other Respir Viruses 2022 Nov;16(6):1133-1140. [doi: 10.1111/irv.13040] [Medline: 35996836]
7. Lalbiaktluangi C, Yadav MK, Singh PK, et al. A cooperativity between virus and bacteria during respiratory infections. Front Microbiol 2023;14:1279159. [doi: 10.3389/fmicb.2023.1279159] [Medline: 38098657]
8. Helou M, Mahdi A, Daoud Z, et al. Epidemiology of community-acquired respiratory tract infections in patients admitted at the emergency departments. Trop Med Infect Dis 2022 Sep 8;7(9):233. [doi: 10.3390/tropicalmed7090233] [Medline: 36136644]
9. Lau YF, Koh WHV, Kan C, et al. Epidemiologic analysis of respiratory viral infections among Singapore military servicemen in 2016. BMC Infect Dis 2018 Mar 12;18(1):123. [doi: 10.1186/s12879-018-3040-x] [Medline: 29529993]
10. Kiliç S, Gray GC. Nonpharmaceutical interventions for military populations during pandemic influenza. Turk Silahli Kuvvetleri Koruyucu Hekim Bul 2007;6(4):285-290 [FREE Full text] [Medline: 18516249]
11. Jeger V, Dünki A, Germann M, et al. H1N1 outbreak in a Swiss military boot camp--observations and suggestions. Swiss Med Wkly 2011;141:w13307. [doi: 10.4414/smw.2011.13307] [Medline: 22144395]
12. Centers for Disease Control and Prevention (CDC). Outbreak of 2009 pandemic influenza A (H1N1) on a Peruvian Navy ship - June-July 2009. MMWR Morb Mortal Wkly Rep 2010 Feb 19;59(6):162-165 [FREE Full text] [Medline: 20168295]
13. COVID-19 vaccination. Ministry of Health-Singapore. 2024 May 29. URL: https://www.moh.gov.sg/covid-19/vaccination [accessed 2024-08-05]
14. Puro V, Fusco FM, Lanini S, Nisii C, Ippolito G. Risk management of febrile respiratory illness in emergency departments. New Microbiol 2008 Apr;31(2):165-173 [FREE Full text] [Medline: 18623980]
15. Paules C, Subbarao K. Influenza. Lancet 2017 Aug 12;390(10095):697-708. [doi: 10.1016/S0140-6736(17)30129-0] [Medline: 28302313]
16. Guan WJ, Ni ZY, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720. [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
17. Velasco JM, Valderama MT, Diones PC, et al. Outbreak of influenza and SARS-CoV-2 at the Armed Forces of the Philippines Health Service Education and Training Center, September 25-October 10, 2023. MSMR 2024 May 20;31(5):9-15 [FREE Full text] [Medline: 38847656]
18. Earhart KC, Beadle C, Miller LK, et al. Outbreak of influenza in highly vaccinated crew of US Navy ship. Emerg Infect Dis 2001;7(3):463-465. [doi: 10.3201/eid0703.010320] [Medline: 11384530]
19. Korzeniewski K, Nitsch-Osuch A, Konior M, Lass A. Respiratory tract infections in the military environment. Respir Physiol Neurobiol 2015 Apr;209:76-80. [doi: 10.1016/j.resp.2014.09.016] [Medline: 25278277]
20. Sanders JW, Putnam SD, Frankart C, et al. Impact of illness and non-combat injury during Operations Iraqi Freedom and Enduring Freedom (Afghanistan). Am J Trop Med Hyg 2005 Oct;73(4):713-719 [FREE Full text] [Medline: 16222015]
21. Gwee SXW, Chua PEY, Pang J. SARS-CoV-2 transmission in the military during the early phase of the pandemic-a systematic analysis. Int J Environ Res Public Health 2022 Jun 16;19(12):7418. [doi: 10.3390/ijerph19127418] [Medline: 35742662]
22. Lee VJ, Tan CH, Yap J, et al. Effectiveness of pandemic H1N1-2009 vaccination in reducing laboratory confirmed influenza infections among military recruits in tropical Singapore. PLoS ONE 2011;6(10):e26572. [doi: 10.1371/journal.pone.0026572] [Medline: 22053196]
23. Lee VJ, Yap J, Cook AR, et al. Oseltamivir ring prophylaxis for containment of 2009 H1N1 influenza outbreaks. N Engl J Med 2010 Jun 10;362(23):2166-2174. [doi: 10.1056/NEJMoa0908482] [Medline: 20558367]

Abbreviations

ARI: acute respiratory infection
FRI: febrile respiratory illness
HEV: human enterovirus
HRV: human rhinovirus
HRV/HEV: human rhinovirus/human enterovirus
IRB: institutional review board
NPI: nonpharmaceutical intervention

XSL•FO
RenderX

OR: odds ratio
SAF: Singapore Armed Forces

Edited by A Mavragani; submitted 22.11.24; peer-reviewed by C Dalton, E Lau, TK Er; revised version received 04.02.25; accepted 05.02.25; published 18.04.25.

Please cite as:
M P, Goh JK, Ng JQ, Mutalib A, Lim HY
Impact of Acute Respiratory Infections on Medical Absenteeism Among Military Personnel: Retrospective Cohort Study
JMIR Form Res 2025;9:e69113
URL: https://formative.jmir.org/2025/1/e69113
doi:10.2196/69113

© Premikha M, Jit Khong Goh, Jing Qiang Ng, Adeliza Mutalib, Huai Yang Lim. Originally published in JMIR Formative Research (https://formative.jmir.org), 18.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Perceptions and Experiences of Developing Prediabetes Awareness and Participating in the National Diabetes Prevention Program Among Hispanic US Adults: Qualitative Study

Kevin L Joiner[1], PhD; Mackenzie Adams[2], MPH; Edith Beltrán Minehan[3], PhD; Matthew J O'Brien[4], MD, MSc

[1]School of Nursing, University of Michigan, 400 N Ingalls St, Ann Arbor, MI, United States

[2]School of Public Health, University of Michigan, Ann Arbor, MI, United States

[3]School of Academic Advancement, Madison Area Technical College, Madison, WI, United States

[4]Institute of Public Health and Medicine, Feinberg School of Medicine, Northwestern University, Evanston, IL, United States

Corresponding Author:
Kevin L Joiner, PhD
School of Nursing, University of Michigan, 400 N Ingalls St, Ann Arbor, MI, United States

Abstract

Background: Prediabetes conveys an increased risk for subsequently developing type 2 diabetes (T2D). The National Diabetes Prevention Program (DPP) is a widely available intensive behavioral intervention that decreases the risk of developing T2D in adults with prediabetes. Data are needed to inform approaches to increase prediabetes awareness and National DPP participation. Few studies have explored perceptions and experiences of prediabetes diagnosis and National DPP participation, and none have focused on Hispanic adults and participation in the National DPP as implemented by a community-based organization.

Objective: This study aims to explore perceptions and experiences of developing prediabetes awareness and participating in the National DPP among Hispanic US adults.

Methods: The sample was recruited from participants in the National DPP as implemented in Spanish by a community-based organization in the upper Midwest. Semistructured interviews were conducted by telephone in April and May 2021. A qualitative descriptive approach was used. Data from the interviews were reviewed, coded, and integrated into themes to reflect the narratives elicited in the interviews.

Results: A total of 16 interviews were conducted. The mean age of the participants was 46 (SD 6, range 34 - 55) years. Most (n=15) identified as female. The majority (n=15) reported having been born in Mexico. More than two-thirds (n=11) had a level of educational attainment of high school completion or less. Nearly half (n=7) reported not having health insurance. Qualitative description resulted in the emergence of four main themes: (1) processing the news of having prediabetes, (2) deciding on treatment for T2D primary prevention, (3) valuing language and cultural congruence in the National DPP, and (4) appreciating action-oriented knowledge gained during National DPP participation. Participants described the emotional impact of becoming aware of having prediabetes. National DPP lifestyle coaches' outreach and recruitment efforts on a local radio program and a Facebook Live (Meta Platforms) broadcast helped raise awareness of prediabetes and influence attitudes toward participation in the National DPP. Values and cultural beliefs appeared to contribute to perceptions and experiences of participating in the National DPP. Participants were inclined to share information about the National DPP with others in their community.

Conclusions: This study presents some of the first evidence exploring perceptions and experiences of developing prediabetes awareness and participating in the National DPP among Hispanic US adults. The findings can inform approaches to increase prediabetes awareness and National DPP participation among Hispanic US adults.

(JMIR Form Res 2025;9:e66964) doi:10.2196/66964

KEYWORDS

prediabetes; Hispanic; prevention; diabetes; diabetes prevention; awareness; older adult; elder; aging; diabetes prevention program; diabetic; type 2 diabetes; type 1 diabetes; endocrinology; diabetes mellitus; community-based; qualitative study; interview; United States

Introduction

Prediabetes is a state of elevated blood glucose that conveys an increased risk for subsequently developing type 2 diabetes (T2D) [1]. In the United States, approximately 97.6 million adults have prediabetes (38% of the entire population) [2]. Among adults with prediabetes, an estimated 29% will develop T2D within 3 years, and 70% will develop T2D within their lifetime without an effective strategy for primary prevention [3,4]. Lifestyle

XSL•FO
RenderX

modification is an effective primary prevention strategy, as demonstrated in a landmark clinical trial of adults with prediabetes and overweight or obesity. Participants randomized to receive the Diabetes Prevention Program (DPP) intensive lifestyle intervention experienced a 58% lower risk of developing T2D over 3 years than the placebo group [5].

Based on this promising evidence, the Centers for Disease Control and Prevention is leading an effort to offer the National DPP, which is derived from the evidence-based intensive lifestyle intervention [6]. The National DPP is implemented in-person and online across the country by community-based organizations, for-profit businesses, health care systems, and insurers [6,7]. The out-of-pocket cost of participating in the National DPP is covered through Medicare, commercial insurance, or employer wellness programs [6]. There are also organizations that enable participation without any out-of-pocket cost through support from philanthropic funds and grants [8].

Despite the high risk that prediabetes portends for subsequent development of T2D and the large-scale effort to provide access to an effective T2D primary prevention strategy, less than 1% of adults with prediabetes have participated in the National DPP [8]. Compared with White non-Hispanic adults with prediabetes, adults with prediabetes of underrepresented racial and ethnic groups are less likely to participate in the National DPP [7,8]. Hispanic adults have the lowest uptake rate into the National DPP among major racial and ethnic groups. Despite making up 15.4% of the US adult population with prediabetes, Hispanic adults only constitute 9% of individuals who have participated in the National DPP [2,8].

Racial and ethnic disparities in National DPP participation are likely multifactorial, stemming from patient-, provider-, and system-level factors. National survey data suggest that only about 20% of adults who meet the glycemic criteria for prediabetes are aware that they have the condition [2]. Even when adults are aware of having prediabetes, evidence suggests that many may not receive counseling from a health care provider about their subsequent risk for developing T2D and the preventive benefits participation in an intensive lifestyle intervention might offer [9]. Factors hindering effective provider-patient discussions about prediabetes and the National DPP include limited educational materials, time constraints during clinic visits, provider perceptions of potential financial constraints to participation, and provider perceptions of patient low self-efficacy for lifestyle change [10-14]. An analysis of 2016 and 2017 National Health Interview Survey data found that less than 5% of adults with diagnosed prediabetes received a referral to the National DPP [15].

Qualitative data from Hispanic adults with prediabetes who have participated in the National DPP are needed to understand their perceptions and experiences about becoming aware of having prediabetes, as well as making an informed choice about initiating and continuing participation in the National DPP. Previous research in the context of T2D prevention in Hispanic adults has explored perceptions regarding the risk of developing diabetes and preferences for prediabetes treatment options [16,17]. In addition, qualitative studies have provided insight into perceptions and experiences of participation in the National DPP as implemented by an integrated health care system among adults with diverse ethnic backgrounds [18,19]. Research in the setting of the National DPP implemented by a community-based organization can augment these earlier qualitative findings.

The objective of this study was to explore perceptions and experiences of prediabetes awareness and National DPP participation among Hispanic adults. The findings from this study will inform future interventions to increase prediabetes awareness and National DPP participation in a high-risk population while helping promote racial and ethnic equity in the prevention of T2D.

Methods

Overview

A qualitative descriptive approach was used [20]. This approach is intended to systematically summarize individual participants' specific and unique perspectives and experiences and report them in detailed descriptions that are closely linked to the words of the participants. Semistructured interviews were used to collect in-depth perspectives and experiences from the participants, allowing their responses to be considered within context [21]. Specifically, follow-up questions were included to elicit more detail about how perspectives and experiences might be influenced by language and cultural factors [22]. The research team was led by the first author, who is a nurse practitioner with a PhD in nursing and formal training and experience in qualitative methods. Additional team members included a research assistant with a PhD in Spanish and a research assistant with a master's degree in public health, both of whom had experience but no formal training in qualitative methods.

The sample was recruited from among current and completed participants of the National DPP, which was implemented in Spanish by a community-based organization, the National Kidney Foundation of Michigan, in Kent County, an urban county in West Michigan (April and May 2021). According to the US 2020 Census, the population of Kent County, Michigan, was approximately 650,000, with 11.7% of residents identifying as having Hispanic or Latino ethnicity [23].

The eligibility criteria were being 18 years or older, identifying as having Hispanic or Latino ethnicity, having a history of prediabetes, and participating in the National DPP. Participating in the National DPP was defined as attending at least one session. There were no exclusion criteria. One of the lifestyle coaches who facilitated the National DPP in Spanish, distributed study flyers in English and Spanish through email to National DPP participants. Those interested in learning more about the study completed an online screener in which they provided their contact information. Individuals who supplied their contact information in the screener were reached by telephone by the first author, who speaks fluent Spanish and identifies as being non-Hispanic White. The first author was unknown to the participants. Those who agreed to participate were subsequently enrolled in the study.

After obtaining verbal informed consent, the first author offered to conduct the semistructured interview in English or Spanish,

XSL•FO
RenderX

depending on the participant's preference. All of the participants chose to complete their interview in Spanish. Textbox 1 presents the semistructured interview guide in English. Sociodemographic and clinical data collected included age, gender identity, country of birth, English language proficiency, language spoken at home, educational attainment, insurance status, and annual household income.

Textbox 1. Interview guide.

Interview questions:

- When did you first learn you had prediabetes?
- Before your diagnosis, what did you know about prediabetes?
- What do you know about the consequences of prediabetes?
- What was your reaction when you were diagnosed?
- Tell me about what it was like when you were diagnosed. How did it make you feel?
- How did you learn about the National Diabetes Prevention Program (DPP)? Did you know someone who had joined?
- Which organization presented the National DPP that you participated in?
- How did your diagnosis influence your decision to participate in the National DPP?
- What made you want to join the National DPP? What gave you motivation?
- If you were to try to convince someone to join the program, what would you tell them?
- Did you encounter obstacles or barriers to participating? Did you have to make any sacrifices to be able to attend? Did you need to find childcare, miss work, or find transportation? How did you overcome these barriers?
- Was there any aspect of the program that made it easier for you to join?
- Which aspects of the National DPP did you like the most or find the most helpful?
- Which aspects of the National DPP did you not like or not find helpful?
- What do you think you learned or gained from the program?

A research assistant, a native Spanish speaker of Mexican heritage who identifies as having Hispanic ethnicity, transcribed the audio recordings of the interviews and translated them from Spanish into English. The research assistant, well-acquainted with the Spanish-language dialects used by the participants, took special care in the translation. The translated transcripts were reviewed by the first author and updated for accuracy when needed. Independently, the research assistant and the other research assistant, who identifies as having non-Hispanic White ethnicity and race, read all of the transcripts and created a list of possible codes with a description for each code. Based on the list of the codes and their descriptions, a codebook was developed by the 2 research assistants and the first author through consensus. The research assistants then independently applied the codes to text segments using ATLAS.ti (Scientific Software Development) computer-aided qualitative data analysis software [24]. The rationale for using computer-aided qualitative data analysis software was to assist in applying the codes and organizing and managing the textual data. As new concepts were identified inductively, additional codes were added to the codebook. After all the transcripts were coded, discrepancies in coding were identified and discussed by the research assistants and the first author. Any differences in coding were adjudicated by the first author and resolved through consensus. Using coding reports, patterns were identified, and a consensus was reached on the themes to be reported. During the analysis, data reduction tables were used to facilitate the iteration of concepts and the development of themes. In addition, audit trails were used to document analytical processes. Thematic saturation was achieved when additional evidence no longer changed the overall description of the themes. Standards for Reporting Qualitative Research endorsed by the EQUATOR (Enhancing the Quality and Transparency of Health Research) Network were followed [25].

Ethical Considerations

The study was approved by the institutional review Board of the University of Michigan (HUM00192113). Informed verbal consent was obtained from all participants before the interviews were conducted. All transcripts were deidentified during transcription, before analysis. Unique identifiers were used to anonymize quotes from individual participants. Each study participant received a US $20 Amazon gift card as compensation for their time spent participating.

Results

Overview

A total of 16 participant interviews were conducted. Table 1 presents the characteristics of the participants. The mean age was 46 (SD 6, range 34 - 55). Most (n=15) identified as female. The majority (n=15) reported being born in Mexico. More than two-thirds (n=11) had a level of educational attainment of high school completion or less. Nearly half (n=7) reported not having health insurance.

Qualitative description resulted in the emergence of four main themes: (1) processing the news of having prediabetes, (2) deciding on treatment for T2D primary prevention, (3) valuing

XSL•FO
RenderX

language and cultural congruence in the National DPP, and (4) appreciating action-oriented knowledge gained during National DPP participation.

Table . Participant demographics (N=16).

Characteristic	Values
Age (years), mean (SD; range)	46 (6; 34-55)
Gender, n	
Male	1
Female	15
Country of birth, n	
Mexico	15
Puerto Rico	1
English proficiency, n	
I don't speak English	3
Not well	11
Well	1
Very well	1
Language spoken at home, n	
Spanish	13
Spanish and English	3
Educational attainment, n	
Elementary school (grades 1 - 8)	2
High school (grades 9 - 12)	9
College or graduate school	5
Insurance status, n	
Private insurance	3
Public insurance	4
No insurance	7
Annual household income (US $), mean (range)	35,500 (17,000-96,000)

[a]Not applicable.

Theme 1: Processing the News of Having Prediabetes

Participants shared a range of unique perceptions and experiences upon becoming aware that they had prediabetes, often through specific events. Commonly, individuals were informed by their health care provider based on the results of a glycated hemoglobin (HbA_{1c}) screening test performed during a routine physical exam. Other participants reported developing awareness of having prediabetes based on a point-of-care HbA_{1c} test performed at a health fair. In one instance, a participant discovered her prediabetes following a pregnancy complicated by gestational diabetes:

> *My prediabetes started… well, 9 years ago. I was pregnant. My son is now 9 years old. Back then, they told me I had gestational diabetes. And after that, I went back to the doctor after giving birth and all of that. Then they told me that I was in the prediabetes [range]. And, I still have prediabetes now.* [Participant 16]

Participants who learned about their prediabetes during a routine physical exam or a visit to a health fair described the emotions they experienced while processing the news of having prediabetes. One woman described how she was caught off guard by the diagnosis of prediabetes during a routine physical examination:

> *I reacted with, well, surprise and also, a little bit of sadness and anger. Because one thinks they're okay until they get checked. All the time, every year, I get my physical, and before, it never turned out that way… that I had prediabetes.* [Participant 11]

For some participants, their diagnosis was a chance to begin understanding prediabetes and making lifestyle changes, including attempting to "eat healthier" and "exercise more." The words of one woman conveyed the sentiment of the participants:

XSL•FO
RenderX

> *I felt distressed. Sad. But this distress and this sadness, I focused it in a positive direction.* [Participant 6]

Several participants explained that the news of having prediabetes was followed by concern about their future quality of life if they were to develop T2D. Nearly all the participants had at least one close family member with T2D, which informed their views of what it would be like to live with diabetes. For example, one woman explained:

> *Yes, I knew [about T2D] because my mother-in-law died from it. And a lot of family on my husband's side [have T2D]. And I knew that they had to inject insulin and check their blood and things like that. I knew a little.* [Participant 6]

In one case, a participant reported that her health care provider outlined a potential future medication scenario for her if she went on to develop T2D:

> *He told me that I was prediabetic and that if I kept going, the next step would be diabetes, and that I would have to take the medicines that are necessary, and possibly, after that, I would have to take insulin. And that is when I said "No… I don't want that." I had to do something with myself so I would not get to the point of having diabetes.* [Participant 2]

Theme 2: Deciding on Treatment for T2D Primary Prevention

Many participants reported being concerned about the risk of subsequently developing T2D and considering possible paths to T2D primary prevention. The only options presented by the health care provider were typically losing weight through eating healthier meals and being more physically active or taking medication. When asked if the individual who had delivered their prediabetes diagnosis had offered the National DPP or any other support for achieving modest weight loss, a few participants reported that they were referred to a nutritionist. However, most participants reported not being offered any support for healthy lifestyle changes and weight loss. To many, at the time of their prediabetes diagnosis, it seemed that the only option was to try lifestyle changes on their own. One woman described her experience choosing between making lifestyle changes and taking medication:

> *I first went to my appointment, and they drew blood. Then, later, I came back for the results. It was then that the doctor told me what was going on and that I should try to get better, lose weight, and eat healthier. Because the doctor knows I am very reluctant to take medicines. I said: "No, I will do it on my own, lose weight, and do it.* [Participant 2]

One woman explained:

> *It was not my doctor [from who I heard about the National DPP]. My doctor would only say: "You need this [to eat healthier and lose weight]. You need that [to take a medication]." She never told me that there is a diabetes prevention program.* [Participant 1]

Participants expressed that the prospect of having to eat healthier and lose weight on their own, without support, seemed overwhelming. One woman shared her frustration:

> *I want to lose weight, but I cannot. Meanwhile, for example, during Holy Week, my clothes did not fit me, and I thought: What can I do? Sometimes, I do not have money to buy new clothes, and it's very unflattering.* [Participant 7]

Another woman stated:

> *You see that we want to lose weight, and we try to diet. But that only lasts, what, 15 days to a month.* [Participant 8]

The majority of participants learned about the National DPP through public health campaigns after being diagnosed with prediabetes. For example, most participants reported finding out about the National DPP through a local Spanish-language program on the radio or a local broadcast on Facebook Live in which a media personality spoke with a local lifestyle coach. What was immediately appealing to many study participants was how the lifestyle coach prepared them for initiating and participating in the National DPP. The lifestyle coach discussed information about the National DPP that was essential so that they could make an informed choice about participation, including that it would teach them about topics such as "how to eat healthy food" and "how to count calories." The lifestyle coach also discussed how the National DPP was a free, group-based program. One woman recalled her experience finding out about the National DPP:

> *I heard about it [the National DPP] on the radio. It was on the program La Poderosa. She [the local lifestyle coach] was saying that peoples' bodies are becoming larger because we have bad habits. And … what caught my attention was that it [T2D] is caused by the decisions that people make … and that people need counseling, some motivation. And I said: "Oh, that sounds interesting!" … I wrote down the number, called, and registered myself.* [Participant 6]

Another woman described her experience finding out about the program:

> *I responded to the video that she [the lifestyle coach] made on El Informador [Facebook Live program]. She said that she was going to teach us how to take care of ourselves better so that we wouldn't develop it [T2D]. And we would be changing our lifestyle and eating healthier. And that she would support us.* [Participant 8]

In addition, a couple of participants who became aware that they had prediabetes at a health fair reported that the National DPP had been offered to them as an opportunity for T2D primary prevention.

Theme 3: Valuing Language and Cultural Congruence in the National DPP

Participants reported that an important factor in their engagement with the National DPP was the linguistic and cultural tailoring for Hispanic adults. The fact that the program

XSL•FO
RenderX

was in Spanish was described as essential. In addition to the language, participants expressed their appreciation that the program was culturally salient in other ways. One participant shared that the National DPP offered her access to information that she felt that she was missing from visits with her health care provider. She expressed that this was especially important since many Hispanic women like herself may not visit a health care provider to manage their health. She explained:

> *To tell you the truth, as a Hispanic, you take things differently. One does not go to the doctor just because. You go to the doctor when you are sick and when you feel bad. You do not go to get a physical or anything. So, I learned a lot in it [the National DPP] about what you should look for in the labels. What you must do and all that, I learned it there.* [Participant 11]

When considering how being Hispanic might present unique challenges or opportunities to engage with the National DPP, one woman identified her country of origin as impacting her relationship with food choices. She explained:

> *I'm Mexican, and we are used to eating… no not eating, ingesting, too much. We're accustomed to beans, rice, meat… steak, chicken, hamburgers, or whatever it may be, …. and always tortillas. We ingest a lot of carbohydrates. … And one eats and eats, and one doesn't pay attention to how much one eats. What I learned from the women [the lifestyle coaches] has been of great importance. I've learned a lot. Thank God.* [Participant 12]

Some participants perceived that being an immigrant created challenges related to gaining weight and eating behaviors. One woman explained:

> *I should not gain any more weight because it is very easy here [in the US] to gain weight. That was the only thing I thought. That I won't gain weight.* [Participant 14]

Another woman explained:

> *Because the problems of Hispanic people are different from the problems of people who were born here. Having left your country, the pain of leaving your parents, your culture, all of that, is painful, and that also causes depression and not eating appropriately.* [Participant 6]

Theme 4: Appreciating Action-Oriented Knowledge Gained During National DPP Participation

Many participants appreciated the action-oriented knowledge they gained during National DPP participation, which promoted their attempts at healthy lifestyle change. One woman explained:

> *From the beginning, when we started with the teacher, she taught us that we need to be detectives of nutrients and food labels. I had never paid attention to that before. I had tried to look for things with less sugar and stuff, but never stopped to look at ingredient by ingredient and the amount of fat. This is what we also studied in class. I tell you, for me, that was something new… She also taught us about exercise, and she showed us what we could do at home.* [Participant 7]

Another woman stated:

> *I've seen nutritionists before, and I was the type of person who wanted to be told, 'Ok, in the morning, you will eat this; at noon, you will eat that; and at night, you will eat this.' But no one tells you that … She [the lifestyle coach] is teaching us how to read food labels to find out about calories, sugars, and fats.* [Participant 8]

A participant recalled the early challenges of learning about prediabetes and reported that understanding biological mechanisms offered in the National DPP enhanced his motivation to make healthy lifestyle changes. He noted:

> *For me, at first, it was very complicated. But then after that, when relating how one eats to how sugars rise, and knowing that the pancreas regulates sugars, it's then when one starts to understand why, before eating and after eating, one's sugar levels are different.* [Participant 11]

Participants discussed how they perceived that they were able to carry over the focus on the empowerment of the National DPP to other family members.

> *To tell you the truth, it [the National DPP] did help me because I have … small children. … I do not want what happened to me to happen to them. I mean, it [the National DPP] helps you. It helps you. It helps make your family better. They [the lifestyle coaches] help you to take care of your family.* [Participant 3]

Some of the participants seemed to feel confident sharing the knowledge that they gained from the National DPP with others in the community. One woman recounted:

> *In my experience, I have gotten to know more people and told them, when they say "prediabetes," you have to take it seriously…and try to change your way of living and lifestyle.* [Participant 3]

Another woman explained:

> *There should be more, like, more programs. Like more publicity. Give us more information.… They should give programs like this [the National DPP] more promotion. Also, tell people more so they know about it. Because the truth is, one doesn't know about them. Look after this, the truth is that I sent a lot of people there. I told them: "Go. Go. It's very important."* [Participant 1]

Discussion

Principal Findings

This qualitative study presents some of the first evidence exploring perceptions and experiences of developing prediabetes awareness and participating in the National DPP among Hispanic US adults. The findings emphasize the emotional impact of becoming aware of having prediabetes and highlight the significance of dynamic outreach and recruitment by

XSL•FO
RenderX

National DPP lifestyle coaches. Outreach by National DPP lifestyle coaches, who leveraged local community forums, including a local radio program and a Facebook Live broadcast, was pivotal in raising awareness of prediabetes and influencing attitudes toward participation in the National DPP. Values and cultural beliefs among US Hispanic communities appeared to contribute to perceptions and experiences of participating in the National DPP. Some participants expressed a strong inclination to share information about the National DPP with others in their community.

Participants vividly remembered the moment they found out that they had prediabetes and the emotions evoked by the diagnosis. Emotions such as surprise, distress, and concern about their future quality of life were commonly reported. This resonates with findings in other communities where similar emotional reactions to developing prediabetes awareness have been documented [26]. The familiarity with T2D and the way T2D had disrupted the lives of close family members and friends was also consistently and clearly communicated, which complements previous research on culturally relevant issues among Hispanic adults in the United States [27]. When reflecting on their views before joining the National DPP and how participating had affected them, the participants highlighted a previous lack of knowledge regarding prediabetes and a previous feeling of lack of control over their trajectory toward developing T2D. For many, the diagnosis of prediabetes served as a catalyst for participating in the National DPP and taking proactive measures such as healthier eating and increased physical activity. These findings highlight the importance of providing clear, easily understood, complete information to enhance individuals' optimism regarding the risk that prediabetes portends for developing T2D and the effectiveness of the National DPP in mitigating this risk.

Guidelines recommend primary care providers play an essential role in T2D primary prevention through prediabetes screening, counseling, and support for initiating National DPP participation [28,29]. However, the findings suggest that primary care providers often fail to effectively counsel Hispanic adults with prediabetes about their risk for developing T2D and the benefits of the National DPP [9,15]. This gap could be attributed to limited awareness among primary care providers about the National DPP and the community-based organizations implementing it [13,14]. Participants echoed this concern, indicating that many did not learn about the National DPP from their health care providers but through other channels.

The role of trusted community figures and effective communication through local media emerged as a key factor in participants' decisions to join the National DPP. Information disseminated through appearances by lifestyle coaches on a local Spanish-language radio program, and a Facebook Live broadcast provided essential insights into the National DPP, making it accessible and appealing. This aligns with previous research that highlights the positive impact of social media on health promotion and National DPP recruitment, particularly within Hispanic communities where the use of social media is prevalent [30,31].

Participants valued the National DPP implemented by the community-based organization for its linguistic and cultural congruence, noting that it resonated with their Hispanic heritage. Cultural factors like traditional diets and the challenges faced by immigrants in adopting healthier lifestyles were common themes. The National DPP's emphasis on action-oriented knowledge, teaching participants about reading food labels, and understanding the biological mechanisms of prediabetes, empowered them to make informed health choices. This sense of control that they gained extended beyond their own personal health, influencing family members and potentially enhancing health consciousness in the community.

Participants' appreciation of the National DPP extended to its broader impacts on family and community. Promotion through family and friends was also highlighted as a valuable channel for reaching more people, including men who might be hesitant about participating or facilitating their spouse's participation. Participants suggested that more supportive efforts and more comprehensive promotion could help inform others in their community about prediabetes and combat fatalism, which was viewed as a cultural belief that might have an impact on how treatments such as the National DPP for primary prevention of T2D are viewed [32].

Offering the National DPP at no out-of-pocket cost helped to alleviate socioeconomic barriers, making the program accessible to those with fewer resources [33]. Lifestyle coaches marketing directly to adults with prediabetes could help overcome potential financial concerns that health care providers may have regarding patient National DPP participation [34]. This approach holds promise for reaching underserved communities and warrants further exploration of coordinated interventions between health care systems and community outreach.

Limitations

This study has several limitations. The Hispanic adults with prediabetes who took part in the study had participated in the National DPP as implemented by a community-based organization, which was in Spanish and culturally tailored for the Hispanic community. Therefore, this study does not explore perceptions and experiences of Hispanic adults with prediabetes when the National DPP is not accessible in the community in a linguistically and culturally tailored manner. The community-based organization was supported by philanthropic and grant funding, which allowed Hispanic adults with prediabetes to participate in the National DPP without any out-of-pocket cost. This study included an overrepresentation of women and individuals who identify as Mexican American compared with the overall Hispanic adult population in the United States [35]. The overwhelming majority of all National DPP participants (75%) nationwide are women [8]. Notwithstanding these limitations, the qualitative description offers clinicians, researchers, and policymakers, insight into the perceptions and experiences of an at-risk population currently underserved by the National DPP, upon which future studies may expand. Further research in this area should consider sampling from a larger number of National DPP sites in geographically dispersed areas. Expanded representation of adults of diverse racial, ethnic, and cultural backgrounds, as

XSL•FO
RenderX

well as male adults, who likely have unique needs and challenges related to being reached and enrolled in the National DPP, are also opportunities for future research.

Future Directions

A community-based intervention for Hispanic adults that strengthens prediabetes awareness and provides counseling about participating in the National DPP may increase uptake into the National DPP. The American Diabetes Association and the Centers for Disease Control and Prevention advocate for patient-centered approaches in primary care to strengthen awareness of prediabetes and support decisions about National DPP participation [6,29]. The potential for local communication channels, including radio programs and social media broadcasts, to amplify health care services' National DPP recruitment and enrollment efforts at scale is also important [36]. In total, 66% of Hispanic adults use Facebook [31,37]. Having an interview with a National DPP lifestyle coach on a radio program or a broadcast on Facebook Live leverages community forums and requires much less investment of resources than health care services calling many people to ask them to join the National DPP [38]. Interventions that foster joint efforts through multiple communication channels have the potential to make the path from prediabetes diagnosis to National DPP participation smoother and more effortless and reach underrepresented racial and ethnic groups [30].

Conclusions

Efforts to increase National DPP participation among Hispanic adults with prediabetes should be tailored to these populations' unique social, cultural, and linguistic needs. This study is helpful for primary care providers, community organizations, and researchers aiming to boost prediabetes awareness and National DPP participation. Participants in this study described their perceptions and experiences of being diagnosed with prediabetes and what they gained from participating in the National DPP, which provides insights for future efforts to advance T2D primary prevention. Overall, participants reported finding the National DPP valuable and well-suited to their needs. First, the enrollment information should be a multipronged effort, including coming through respected, vibrant communication forums that serve the needs of the community. Second, a National DPP lifestyle coach who is a community member would be trusted easily; however, if a lifestyle coach would like to be helpful to a Hispanic community, there must be an awareness of the cultural, social, and linguistic issues pertaining to this community. Finally, it is important to note that the empowerment that the National DPP brings to individuals and families has its core in the community. Empowerment comes from managing the diagnosis of prediabetes with lifestyle changes that affect and impact family and loved ones. Hispanic adults with prediabetes who participate in the National DPP don't just decrease their own risk for developing T2D; they also become passionate and engaged leaders who have the potential to improve others' lives in the community.

Acknowledgments

This study would not have been possible without the support of the National Kidney Foundation of Michigan. The authors acknowledge the use of ChatGPT for editing and refining language in manuscript preparation and take full responsibility for the content and interpretations.

Data Availability

The dataset generated and analyzed during this qualitative research study is not publicly available due to concerns about participant privacy and the sensitive nature of the information collected.

Conflicts of Interest

None declared.

Checklist 1
Standards for Reporting Qualitative Research (SRQR) checklist.
[DOCX File, 22 KB - formative_v9i1e66964_app1.docx]

References

1. Tabák AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M. Prediabetes: A high-risk state for diabetes development. Lancet 2012 Jun 16;379(9833):2279-2290. [doi: 10.1016/S0140-6736(12)60283-9] [Medline: 22683128]
2. National Diabetes Statistics Report 2024. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/diabetes/php/data-research/index.html [accessed 2024-09-27]
3. Nathan DM, Davidson MB, DeFronzo RA, et al. Impaired fasting glucose and impaired glucose tolerance: implications for care. Diabetes Care 2007 Mar;30(3):753-759. [doi: 10.2337/dc07-9920] [Medline: 17327355]
4. Gerstein HC, Santaguida P, Raina P, et al. Annual incidence and relative risk of diabetes in people with various categories of dysglycemia: A systematic overview and meta-analysis of prospective studies. Diabetes Res Clin Pract 2007 Dec;78(3):305-312. [doi: 10.1016/j.diabres.2007.05.004] [Medline: 17601626]

XSL•FO
RenderX

5. Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 2002 Feb 7;346(6):393-403. [doi: 10.1056/NEJMoa012512] [Medline: 11832527]
6. National Diabetes Prevention Program. Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/diabetes-prevention/index.html [accessed 2024-07-08]
7. Cannon MJ, Ng BP, Lloyd K, Reynolds J, Ely EK. Delivering the National Diabetes Prevention Program: Assessment of enrollment in in-person and virtual organizations. J Diabetes Res 2022;2022:2942918. [doi: 10.1155/2022/2942918] [Medline: 35118160]
8. Gruss SM, Nhim K, Gregg E, Bell M, Luman E, Albright A. Public health approaches to Type 2 diabetes prevention: the US National Diabetes Prevention Program and beyond. Curr Diab Rep 2019 Aug 5;19(9):78. [doi: 10.1007/s11892-019-1200-z] [Medline: 31385061]
9. Cunningham A, Li E, Silverio A, Mills G, Han J, Waters A. Patient and provider prediabetes knowledge, attitudes, and behavior in a large urban family medicine practice. Ann Fam Med 2022 Apr 1;20(20 Suppl 1):2748. [doi: 10.1370/afm.20.s1.2748] [Medline: 36857128]
10. Madrigal L, Haardörfer R, Kegler MC, et al. A structural equation model of CFIR inner and outer setting constructs, organization characteristics, and national DPP enrollment. Implement Sci Commun 2023 Nov 17;4(1):142. [doi: 10.1186/s43058-023-00522-3] [Medline: 37978574]
11. Papalii M, Ely E. 333-OR: What motivates people to enroll in the National Diabetes Prevention Program Lifestyle Change Program. Diabetes 2023 Jun 20;72(Supplement_1). [doi: 10.2337/db23-333-OR]
12. AuYoung M, Moin T, Richardson CR, Damschroder LJ. The Diabetes Prevention Program for Underserved Populations: A brief review of strategies in the real world. Diabetes Spectr 2019 Nov;32(4):312-317. [doi: 10.2337/ds19-0007] [Medline: 31798288]
13. Tseng E, Greer RC, O'Rourke P, et al. National Survey of Primary Care Physicians' Knowledge, Practices, and Perceptions of Prediabetes. J Gen Intern Med 2019 Nov;34(11):2475-2481. [doi: 10.1007/s11606-019-05245-7] [Medline: 31502095]
14. Tseng E, Greer RC, O'Rourke P, et al. Survey of primary care providers' knowledge of screening for, diagnosing and managing prediabetes. J Gen Intern Med 2017 Nov;32(11):1172-1178. [doi: 10.1007/s11606-017-4103-1] [Medline: 28730532]
15. Ali MK, McKeever Bullard K, Imperatore G, et al. Reach and use of Diabetes Prevention Services in the United States, 2016-2017. JAMA Netw Open 2019 May 3;2(5):e193160. [doi: 10.1001/jamanetworkopen.2019.3160]
16. Joiner KL, Sternberg RM, Kennedy CM, Fukuoka Y, Chen JL, Janson SL. Perception of risk for developing diabetes among foreign-born Spanish-speaking US Latinos. Diabetes Educ 2016 Aug;42(4):418-428. [doi: 10.1177/0145721716646204] [Medline: 27150605]
17. O'Brien MJ, Moran MR, Tang JW, et al. Patient perceptions about prediabetes and preferences for diabetes prevention. Diabetes Educ 2016 Dec;42(6):667-677. [doi: 10.1177/0145721716666678] [Medline: 27621093]
18. Ritchie ND, Phimphasone-Brady P, Sauder KA, Amura CR. Perceived barriers and potential solutions to engagement in the National Diabetes Prevention Program. ADCES in Practice 2021 Jan;9(1):16-20. [doi: 10.1177/2633559X20966275]
19. Harrison CR, Phimphasone-Brady P, DiOrio B, et al. Barriers and facilitators of National Diabetes Prevention Program engagement among women of childbearing age: a qualitative study. Diabetes Educ 2020 Jun;46(3):279-288. [doi: 10.1177/0145721720920252] [Medline: 32597384]
20. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240X(200008)23:4<334::AID-NUR9>3.0.CO;2-G]
21. DeJonckheere M, Vaughn LM. Semistructured interviewing in primary care research: A balance of relationship and rigour. Fam Med Com Health 2019 Mar;7(2):e000057. [doi: 10.1136/fmch-2018-000057]
22. Hispanic Americans' trust in and engagement with science. Pew Research Center. 2022. URL: https://www.pewresearch.org/science/2022/06/14/hispanic-americans-trust-in-and-engagement-with-science/ [accessed 2024-09-27]
23. U.S. Census Bureau. Quick Facts: Kent County, Michigan. 2024. URL: https://www.census.gov/quickfacts/fact/table/kentcountymichigan/PST04022424 [accessed 2025-02-03]
24. Friese S. Qualitative Data Analysis with ATLAS.ti, 3rd edition: SAGE Publications Ltd; 2019.
25. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251. [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
26. Skoglund G, Nilsson BB, Olsen CF, Bergland A, Hilde G. Facilitators and barriers for lifestyle change in people with prediabetes: A meta-synthesis of qualitative studies. BMC Public Health 2022 Mar 21;22(1):553. [doi: 10.1186/s12889-022-12885-8] [Medline: 35313859]
27. Caban A, Walker EA. A systematic review of research on culturally relevant issues for Hispanics with diabetes. Diabetes Educ 2006;32(4):584-595. [doi: 10.1177/0145721706290435] [Medline: 16873596]
28. Pronk NP, Remington PL, Community Preventive Services Task Force. Combined diet and physical activity promotion programs for prevention of diabetes: Community preventive services task force recommendation statement. Ann Intern Med 2015 Sep 15;163(6):465-468. [doi: 10.7326/M15-1029] [Medline: 26168073]
29. ElSayed NA, Aleppo G, Bannuru RR, et al. 3. Prevention or Delay of Diabetes and Associated Comorbidities: Standards of Care in Diabetes—2024. Diabetes Care 2024 Jan 1;47(Supplement_1):S43-S51. [doi: 10.2337/dc24-S003]

XSL•FO
RenderX

30. Jalkanen K, Järvenpää R, Tilles-Tirkkonen T, et al. Comparison of communication channels for large-scale type 2 diabetes risk screening and intervention recruitment: empirical study. JMIR Diabetes 2021 Sep 9;6(3):e21356. [doi: 10.2196/21356] [Medline: 34499036]
31. Americans' social media use. Pew Research Center. 2024. URL: https://www.pewresearch.org/internet/2024/01/31/americans-social-media-use/ [accessed 2024-09-27]
32. Moreira T, Hernandez DC, Scott CW, Murillo R, Vaughan EM, Johnston CA. Susto, Coraje, y Fatalismo: Cultural-bound beliefs and the treatment of diabetes among socioeconomically disadvantaged Hispanics. Am J Lifestyle Med 2018 Jan;12(1):30-33. [doi: 10.1177/1559827617736506] [Medline: 29541008]
33. Improving access to diabetes prevention. Center for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/diabetes/health-equity/improving-access-type-2-diabetes-prevention-program.html [accessed 2024-08-05]
34. Thomas TW, Golin CE, Kinlaw AC, et al. Did the 2015 USPSTF abnormal blood glucose recommendations change clinician attitudes or behaviors? A mixed-method assessment. J Gen Intern Med 2022 Jan;37(1):15-22. [doi: 10.1007/s11606-021-06749-x] [Medline: 33826060]
35. Hispanic/Latino health. U.S. Department of Health and Human Services Office of Minority Health. 2024. URL: https://minorityhealth.hhs.gov/hispaniclatino-health [accessed 2024-07-17]
36. Mora J, Romo R, Dempsey S, et al. Engaging the community served: A U.S. Cancer Center'S Facebook live cancer awareness campaign for Spanish-speaking Latinos during COVID-19. Cancer Causes Control 2023 Dec;34(12):1037-1042. [doi: 10.1007/s10552-023-01755-2] [Medline: 37490139]
37. Social media use in 2021. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2024-09-27]
38. Philis-Tsimikas A, Fortmann A, Lleva-Ocana L, Walker C, Gallo LC. Peer-led diabetes education programs in high-risk Mexican Americans improve glycemic control compared with standard approaches: A Project Dulce promotora randomized trial. Diabetes Care 2011 Sep;34(9):1926-1931. [doi: 10.2337/dc10-2081] [Medline: 21775748]

Abbreviations

DPP: Diabetes Prevention Program
EQUATOR: Enhancing the Quality and Transparency of Health Research
HbA_{1c}: glycated hemoglobin
T2D: type 2 diabetes

Edited by A Mavragani; submitted 27.09.24; peer-reviewed by B Gibson, TB Loeb; revised version received 26.12.24; accepted 27.12.24; published 11.02.25.

Please cite as:
Joiner KL, Adams M, Minehan EB, O'Brien MJ
Perceptions and Experiences of Developing Prediabetes Awareness and Participating in the National Diabetes Prevention Program Among Hispanic US Adults: Qualitative Study
JMIR Form Res 2025;9:e66964
URL: https://formative.jmir.org/2025/1/e66964
doi:10.2196/66964

© Kevin L Joiner, Mackenzie Adams, Edith Beltrán Minehan, Matthew J O'Brien. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Disease Awareness in Patients With Type 2 Diabetes: Analysis of Baseline Data From the SMART-Finder Observational Study

Christian Mueller[1], PhD; Thomas Neusser[1], PhD; Inga Thate-Waschke[1], PhD; Julia Nowicki[2], BSc; Tomasz Plominski[2], MSc; Regine Griesinger[3]; Stefanie Kessner[3]; Stephan Martin[4], MD

[1]Pharmaceuticals Medizin, Pharmaceuticals, BAYER Vital GmbH, K56, Leverkusen, Germany

[2]smartpatient GmbH, Munich, Germany

[3]Emmes Biopharma Germany GmbH, Munich, Germany

[4]Verbund Katholischer Kliniken Duesseldorf (VKKD) und Westdeutsches Diabetes- und Gesundheitszentrum (WDGZ), Duesseldorf, Germany

Corresponding Author:
Christian Mueller, PhD
Pharmaceuticals Medizin, Pharmaceuticals, BAYER Vital GmbH, K56, Leverkusen, Germany

Abstract

Background: Chronic kidney disease (CKD) is a common comorbidity of type 2 diabetes mellitus (T2DM). Data on the determination of CKD-related biomarkers among patients with T2DM in a real-life setting within Germany are limited.

Objective: We aimed to determine the prevalence of CKD and risk factors, availability of urine albumin-to-creatinine ratio (UACR) and estimated glomerular filtration rate (eGFR) values, treatment satisfaction, and quality of life among patients with T2DM in Germany.

Methods: SMART-Finder is a retrospective and prospective, observational, digital, patient-centered cohort study being performed as part of the routine use of an adherence-supporting app. This baseline analysis' observation period was from August to November 2023. Patients with T2DM in Germany who actively used the MyTherapy app; allowed push notifications; and documented use of diabetes medications, renin-angiotensin system inhibitors, finerenone, and/or blood glucose test strips were eligible for inclusion. Study materials (background information, electronic consent form, and laboratory and electronic questionnaires) were provided to eligible patients via app push notifications. Participants completed an electronic case report form that included questions on their blood pressure; their most recent UACR, eGFR, and glycated hemoglobin (HbA_{1c}) values in the past 12 months; the EQ-5D-5L; and the Diabetes Treatment Satisfaction Questionnaire. The primary outcome was the proportion of patients with a UACR of ≥30 mg/g.

Results: Of 9527 invited eligible patients, 101 completed the electronic case report form (male: n=61; female: n=40; age: mean 54.2, SD 11.4 y). Of these, 1 female patient and 5 male patients reported their UACR values; 3 (all male) had a UACR of ≥30 mg/g. The remaining 95 patients reported that their health care professionals had not provided UACR measurements. Only 9 (8.9%) patients were aware of their latest eGFR values (3 patients: 15 - 44 mL/min/1.73 m^2; 6 patients: 45 - 89 mL/min/1.73 m^2), 90 provided HbA_{1c} values (80 patients: ≥6.0%), 46 had a systolic blood pressure of ≥130 mm Hg, and 83 reported former or current nephrotoxic medication intake. The mean EQ-5D-5L index score was 0.7 (SD 0.3; range –0.1 to 1.0; 50 patients). The mean Diabetes Treatment Satisfaction Questionnaire score was 28.8 (SD 6.8; range 9.0-36.0; 49 patients).

Conclusions: Patients with T2DM who were using an adherence-supporting app in Germany lacked awareness of CKD-related biomarkers but had high knowledge of self-manageable biomarkers (eg, blood pressure, serum fasting glucose, and HbA_{1c} values). Our results suggest that treating physicians either do not test for UACRs and eGFRs or do not inform patients about the results. Nonadherence to diagnostic testing guidelines and a lack of physician-patient communication put patients at risk. Another reason for this health literacy imbalance may be the focus on HbA_{1c} instead of kidney comorbidity in patient education material. Future goals for diabetes management must include guideline-compliant testing of CKD-related biomarkers and open physician-patient communication.

International Registered Report Identifier (IRRID): RR2-10.2196/44996

(JMIR Form Res 2025;9:e60246) doi:10.2196/60246

KEYWORDS

app-based documentation; chronic kidney disease; CKD; disease awareness; MyTherapy; type 2 diabetes; type 2 diabetes mellitus; urine albumin-to-creatinine ratio screening; patient; observational study; Germany; quality of life; treatment; therapy; physician-patient communication

XSL•FO
RenderX

Introduction

Chronic kidney disease (CKD) is a common comorbidity of type 2 diabetes mellitus (T2DM), which is a leading cause of death and disability worldwide [1-8]. Patients with CKD show persistently decreased estimated glomerular filtration rates (eGFRs) or persistently elevated urine albumin excretion [9-11], which may progress to end-stage renal disease [12]. A high-normal urine albumin-to-creatinine ratio (UACR) is associated with a significantly increased risk of all-cause mortality [13].

Effective therapeutic approaches that can influence the development of CKD in patients with T2DM are available [14-16]. The current consensus guideline of the American Diabetes Association (ADA) and the Kidney Disease: Improving Global Outcomes (KDIGO) Work Group recommends renin-angiotensin system inhibitors, sodium-glucose cotransporter 2 inhibitors (SGLT2is), statins, and metformin as first-line drug therapies for patients with T2DM and CKD [17]. Glucagon-like peptide-1 receptor agonists are recommended if SGLT2is and metformin do not achieve sufficient glycemic control. For patients with persistent albuminuria despite standard treatment with glucose-lowering and antihypertensive medications, a nonsteroidal mineralocorticoid receptor antagonist is recommended in addition to baseline therapy [17].

The early identification of patients with diabetes who are at increased risk for renal disease is highly recommended [18,19]. Therefore, the eGFR (based on serum creatinine) and albuminuria (based on the UACR) should be assessed once per year in patients with T2DM [20,21]. As the UACR demonstrates a high degree of within-individual variability among individuals, regularly collecting samples for UACR testing may improve capacity for monitoring changes over time in clinical and research settings [22]. However, data on the determination of UACRs in a real-life setting and the prevalence of CKD among patients with T2DM in Germany are limited [23-26].

The SMART-Finder study was designed to determine the prevalence of CKD and risk factors, availability of UACR and eGFR values, satisfaction with treatment, and quality of life among patients with T2DM using an adherence-supporting app in Germany [27]. In this analysis of baseline data from the SMART-Finder study, the primary objective was to assess the proportion of patients with T2DM who had an elevated UACR (≥30 mg/g; albumin-to-creatinine ratio [ACR] stages A2 and A3). As secondary objectives, the proportion of patients without UACR test data was determined, blood pressure and nephrotoxic comedication status were assessed, and quality of life was analyzed. Disease awareness and treatment satisfaction were assessed as further secondary objectives.

Methods

Study Design

As previously described [27], the recruitment of patients and documentation of data for this retrospective and prospective, observational, digital, patient-centered cohort study were carried out without the assistance of a health care professional (HCP), as part of the routine use of the MyTherapy app (smartpatient GmbH) by patients with T2DM in Germany. The app helps patients manage their treatment-related tasks and access information on other health topics, including comorbidities, by using their mobile device. The app provides the user with a daily "To do" list of tasks, including those regarding medication intake, and asks the user to confirm whether each task has been completed [27]. Patients who documented the following medications or medical devices within the app were considered eligible for participation during initial selection: metformin, acarbose, dipeptidyl-peptidase-4 inhibitor, glucagon-like peptide-1 receptor agonist, renin-angiotensin system inhibitor, SGLT2i, basal insulin, nonsteroidal mineralocorticoid receptor antagonist, sulfonylurea, and blood glucose test strips.

To ensure that patients were properly informed, all eligible patients who actively used the app and allowed push notifications during the app's regular operation received content via push notifications during an awareness phase, which lasted from August 2 to 15, 2023. This content comprised 3 videos about the scope and goals of the SMART-Finder study, the risk of CKD in patients with T2DM, and the impairment of kidney function in patients with CKD; a study description page; a document for collecting laboratory values; and an electronic consent form for explaining, in detail, which data were used for this study and obtaining consent for the use of pseudonymized data. After completion of the consent form, an electronic case report form (eCRF), the EQ-5D-5L, and the Diabetes Treatment Satisfaction Questionnaire (DTSQ) were provided in the "Discover" section of the app (patients were notified about this via push notifications as well) on August 18, 21, and 24, 2023, respectively. The eCRF included questions on patients' residential region, age, gender (self-reported as 1 of 3 options: male, female, or diverse), weight, height, smoking status, last systolic and diastolic blood pressures, serum fasting glucose level, and most recent available laboratory values within the last 12 months (UACR, eGFR, and glycated hemoglobin [HbA_{1c}]). If patients did not know these laboratory values, they were encouraged to request them from their treating physician.

The observation period for the baseline analysis was from August 18, 2023, until the cutoff date (November 20, 2023). Patients who signed the informed consent form and completed at least the eCRF were eligible for baseline analysis.

Ethical Considerations

Ethics approval for this study was granted by the Ethics Committee of the North Rhine Medical Association (Ärztekammer Nordrhein) on October 13, 2022 (approval number: 2022263). All patients were offered the informed consent form; of the 103 who signed the form, 2 withdrew consent. All patient data were handled in a pseudonymized manner, and the privacy of patients was ensured. Upon completion of the baseline eCRF (regardless of EQ-5D-5L or DTSQ completion), patients received a €10 (US $10.80) voucher for Shop Apotheke as compensation for their time spent on data documentation.

XSL•FO
RenderX

Statistical Analysis

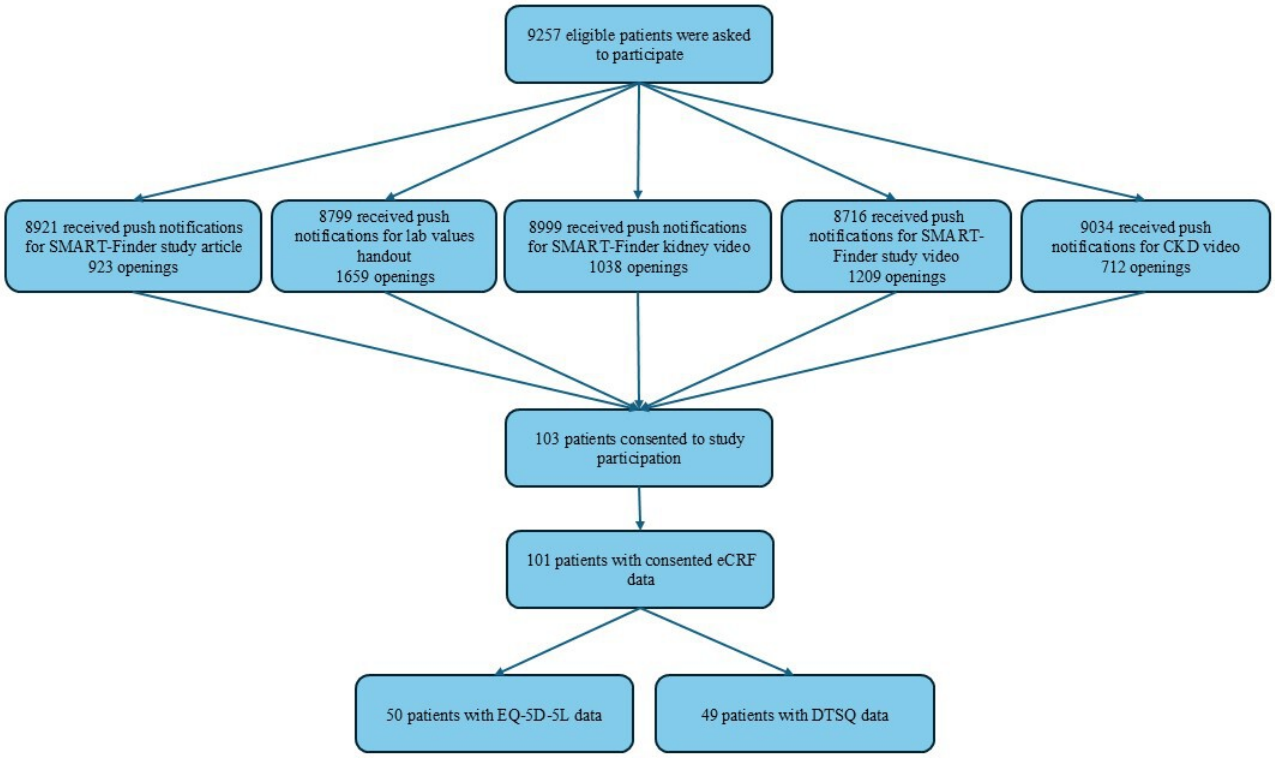
Data were analyzed descriptively with SAS (release 9.4 or higher; SAS Institute Inc). UACRs were categorized based on KDIGO ACR stages [6], with the addition of a fourth level indicating nephrotic-range albuminuria [28,29] (stage A1: UACR of <30 mg/g; stage A2: UACR of 30 - 300 mg/g; stage A3: UACR of 301 - 3000 mg/g; stage A4: UACR of >3000 mg/g). Serum fasting glucose levels were categorized based on the ADA diagnostic criteria for diabetes and prediabetes [30], and eGFRs were reported by KDIGO category [6]. The EQ-5D-5L index score was calculated by using a value set for the German general population [31].

Results

Study Population

During the awareness campaign, push notifications for study- and indication-related content were sent out to a cohort of patients with T2DM (Figure 1), whereof 8799 to 9527 patients received these push notifications. Of these 8799 to 9527 patients, 103 provided consent, and 2 withdrew consent. A total of 101 patients completed the eCRF and were therefore included in the baseline analysis.

Figure 1. Patient recruitment and eligibility in the SMART-Finder observational study. Push notifications were sent out to all eligible patients with T2DM (N=9527). The number of patients receiving each push notification could be influenced by factors such as internet connectivity or patients having their devices switched off at the time of the notification. Therefore, 8921 (93.6%), 8799 (92.4%), 8999 (94.5%), 8716 (91.5%), and 9034 (94.8%) of the 9527 eligible patients received push notifications for the article on the SMART-Finder study, the laboratory values handout, the video on kidney function in CKD, the video on the SMART-Finder study, and the video on risk of CKD, respectively. CKD: chronic kidney disease; DTSQ: Diabetes Treatment Satisfaction Questionnaire; eCRF: electronic case report form; T2DM: type 2 diabetes mellitus.



Of the 101 participating patients with T2DM, 61 were male, and 40 were female. The study population had a mean age of 54.2 (SD 11.4; range 22 - 76) years, a mean weight of 103.9 (SD 21.7; range 45 - 149) kg, and a mean BMI of 34 (SD 7.0; range 16.9 - 54.7) kg/m^2. Further, 26 patients were smokers, 43 were former smokers, and 32 had never smoked.

Laboratory Values and Blood Pressure

Of the 101 patients in the analyzed cohort, only 1 female patient and 5 male patients reported their UACR values (Table 1); 3 of the patients (all male) had a UACR of ≥30 mg/g. The remaining 95 (94.1%) patients reported that their HCPs had not provided UACR measurements. For eGFR values, a similar finding was observed—only 9 (8.9%) of the 101 patients were aware of their latest values.

XSL•FO
RenderX

Table . Classification of UACRs[a] in female and male patients.

	Total patients (N=101), n (%[b])	Patients stratified by UACR, n (%[b])				
		Missing data (n=95)	<30 mg/g (n=3)	30 - 300 mg/g (n=2)	301 - 3000 mg/g (n=1)	>3000 mg/g (n=0)
Female	40 (39.6)	39 (41)	1 (33)	0 (0)	0 (0)	0 (0)
Male	61 (60.4)	56 (59)	2 (67)	2 (100)	1 (100)	0 (0)

[a]UACR: urinary albumin-to-creatinine ratio.

[b]Percentages were calculated by using the n values in the column headings as the denominator.

In contrast to laboratory values provided by HCPs, self-determined test results were more commonly reported. In total, 65 (64.4%) patients provided their serum fasting glucose levels, and 90 (89.1%) patients provided their latest HbA_{1c} values (Table 2).

Table . eGFR[a], HbA_{1c}[b,c], and serum fasting glucose[d] at baseline.

	Patients (N=101), n (%)
eGFR at baseline, mL/min/1.73m^2	
Missing	92 (91.1)
15 - 29	1 (1)
30 - 44	2 (2)
45 - 59	2 (2)
60 - 89	4 (4)
HbA_{1c} at baseline, %	
Missing	11 (10.9)
<6.0	9 (8.9)
6.0 - 6.5	29 (28.7)
6.6 - 7.0	16 (15.8)
7.1 - 7.5	16 (15.8)
7.6 - 8.0	6 (5.9)
8.1 - 8.5	3 (3)
8.6 - 9.0	8 (7.9)
>9.0	3 (3)
Serum fasting glucose summary at baseline, mg/dL	
Missing	36 (35.6)
<100	7 (6.9)
100 - 126	38 (37.6)
>126	20 (19.8)

[a]eGFR: estimated glomerular filtration rate.

[b]HbA_{1c}: glycated hemoglobin.

[c]To convert HbA_{1c} to mmol/mol, multiply by 10.93 and subtract 23.5 mmol/mol.

[d]To convert serum fasting glucose to mmol/L, divide by 18.018.

Systolic and diastolic blood pressures are presented in Table 3. In total, 31 (30.7%) patients had a high-normal systolic blood pressure, 15 (14.9%) patients showed systolic hypertension, and 13 (12.9%) patients showed diastolic hypertension according to the 2023 European Society of Hypertension guidelines for the management of arterial hypertension (assuming office-based blood pressure measurements) [32].

XSL•FO
RenderX

Table . Systolic and diastolic blood pressures at baseline.

	Patients (N=101), n (%)
Systolic blood pressure at baseline, mm Hg	
≤119	18 (17.8)
120 - 129	37 (36.6)
130 - 139	31 (30.7)
140 - 159	13 (12.9)
160 - 179	2 (2)
Diastolic blood pressure at baseline, mm Hg	
55 - 59	4 (4)
60 - 64	4 (4)
65 - 69	7 (6.9)
70 - 74	14 (13.9)
75 - 79	11 (10.9)
80 - 84	36 (35.6)
85 - 89	12 (11.9)
90 - 99	9 (8.9)
100 - 109	4 (4)

Medication Usage

Use of SGLT2is, the nonsteroidal mineralocorticoid receptor antagonist finerenone, hypertension medication (defined by anatomic therapeutic chemical codes), and nephrotoxic medication (as listed by Patel and Sapra [33]) was reported by 44, 0, 80, and 83 patients, respectively. All patients confirmed their T2DM diagnosis at baseline. Further, 81 reported hypertension (77 with treated hypertension and 4 with untreated hypertension) and 12 reported CKD as comorbidities. Reported comedication data for all patients are presented in Multimedia Appendix 1.

Quality of Life and Treatment Satisfaction

A total of 50 patients completed the EQ-5D-5L. The mean EQ-visual analog scale score was 64.4 (SD 24.4; range 3.0 - 99.0; on a scale of 0 - 100, where 0 is the worst and 100 is the best health state possible), and the mean EQ-5D-5L index score was 0.7 (SD 0.3; range –0.1 to 1.0). The DTSQ was completed by 49 patients, with a mean DTSQ score of 28.8 (SD 6.8; range 9.0 - 36.0) at baseline.

Discussion

Principal Results and Comparison With Prior Work

With the SMART-Finder study, we are investigating the self-reporting of UACR, eGFR, and HbA_{1c} values in a cohort of patients with T2DM using an adherence-supporting app in Germany. This baseline analysis provides insights into the implementation of treatment guidelines and HCP-patient interactions.

We observed a low participation rate in our cohort of patients, who were taking responsibility for their health status by regularly tracking adherence. Of the 9527 patients who were contacted, only 712 (7.5%) to 1659 (17.4%) patients engaged with at least one aspect of the awareness campaign by clicking on a video or the written information content in the app. The final participation rate—8.4% (101 of 1209 patients who saw the SMART-Finder study video)—is comparable to those observed in other awareness campaigns; for example, 10.2% (804/7865) and 10.1% (802/7920) of invited participants completed a survey on a public health campaign for increasing kidney health awareness in a Canadian province [34].

In contrast to activities that can be self-managed by patients (blood pressure measurement and testing of HbA_{1c} and serum fasting glucose levels), about which our cohort appears to be well informed, our results suggest a dramatic lack of awareness of renal function status in this population at risk of CKD. This is of particular concern, since almost half of the cohort (46/101, 45.5%) had a systolic blood pressure of ≥130 mm Hg, and 82.2% (83/101) reported former or current intake of nephrotoxic medication. Further, more than 90% of the patients were unaware of their eGFR (92/101, 91.1%) and UACR (95/101, 94.1%), but 89.1% (90/101) of the patients knew their HbA_{1c} level, suggesting that treating physicians either do not test for UACRs and eGFRs or do not inform patients about the results. Both a lack of adherence to guidelines for diagnostic testing and a lack of HCP-patient communication put patients at risk. Another reason for this imbalance in health literacy may be the focus on HbA_{1c}, rather than on kidney comorbidity, in patient education material.

Limitations

The presented baseline analysis has some limitations. There is a selection bias, as only users of the adherence-supporting app, that is, those who seemed to have experience with using

XSL•FO
RenderX

technology, were able to participate. Additionally, the unconventional method of data collection, with patients requesting data from HCPs, may have introduced reporting bias from the HCPs. Further, due to the low participation rate and the lack of UACR and eGFR values, our findings on CKD risk in the observed cohort of patients may not be generalizable to all patients with T2DM. However, the missing data show a lack of awareness in a high-risk population. Even though the annual eGFR and albuminuria testing rates in Germany are 96.5% and 77.2%, respectively, our results suggest that among German patients with diabetes, knowledge regarding their kidney function and disease is still insufficient [35].

Conclusions

Baseline data from the SMART-Finder study, which were derived from patients with T2DM who were using an adherence-supporting app in Germany, showed a dramatic lack of awareness of kidney-related biomarkers (UACR and eGFR) but revealed high knowledge of HbA_{1c} values. As only informed patients are empowered patients, future goals for diabetes management must include guideline-compliant testing of CKD-related biomarkers and open communication between HCPs and patients.

Acknowledgments

This study was sponsored by Bayer Vital GmbH, Leverkusen, Germany. It was performed in collaboration with smartpatient GmbH, Munich, Germany (the provider of the MyTherapy app)—a subsidiary of Shop Apotheke, one of the largest internet-based pharmacies in Germany—and EMMES Biopharma Germany, who assisted with the statistical analysis. The funding organization was the sponsor of this study. Bayer provided support for designing and conducting the study; managing data collection and interpretation; and managing manuscript preparation, review, and approval. The authors would like to thank Claire Mulligan, PhD (Beacon Medical Communications Ltd, Brighton, United Kingdom) for editing the final draft of this manuscript (funded by Bayer Vital GmbH, Leverkusen, Germany).

Data Availability

Data sharing can be requested from the sponsor of this study, and it will be managed according to internal standard operating procedures and the General Data Protection Regulation of the European Union.

Authors' Contributions

CM and SK had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. CM and TN contributed to the concept and design of the study. All authors contributed to the acquisition, analysis, or interpretation of data. CM drafted the manuscript. ITW, SM, TN, JN, RG, and SK critically reviewed the manuscript for important intellectual content. SK and RG contributed to statistical analysis. CM and TN obtained funding. TP, JN, and TN provided administrative, technical, or material support. SM supervised the study.

Conflicts of Interest

CM, TN, and ITW are employees of Bayer Vital GmbH. JN and TP are employees of smartpatient GmbH. RG and SK are employees of EMMES Biopharma Germany. SM received consultancy fees from Bayer Vital GmbH.

Multimedia Appendix 1
Reported concomitant medications.
[DOCX File, 46 KB - formative_v9i1e60246_app1.docx]

References

1. Song MK, Davies NM, Roufogalis BD, Huang THW. Management of cardiorenal metabolic syndrome in diabetes mellitus: a phytotherapeutic perspective. J Diabetes Res 2014;2014:313718. [doi: 10.1155/2014/313718] [Medline: 24818164]
2. Galicia-Garcia U, Benito-Vicente A, Jebari S, et al. Pathophysiology of type 2 diabetes mellitus. Int J Mol Sci 2020 Aug 30;21(17):6275. [doi: 10.3390/ijms21176275] [Medline: 32872570]
3. Roden M, Shulman GI. The integrative biology of type 2 diabetes. Nature 2019 Dec;576(7785):51-60. [doi: 10.1038/s41586-019-1797-8] [Medline: 31802013]
4. Stumvoll M, Goldstein BJ, van Haeften TW. Type 2 diabetes: principles of pathogenesis and therapy. Lancet 2005 Apr;365(9467):1333-1346. [doi: 10.1016/S0140-6736(05)61032-X] [Medline: 15823385]
5. Ronco C, Haapio M, House AA, Anavekar N, Bellomo R. Cardiorenal syndrome. J Am Coll Cardiol 2008 Nov 4;52(19):1527-1539. [doi: 10.1016/j.jacc.2008.07.051] [Medline: 19007588]
6. Kidney Disease: Improving Global Outcomes (KDIGO) Diabetes Work Group. KDIGO 2020 clinical practice guideline for diabetes management in chronic kidney disease. Kidney Int 2020 Oct;98(4S):S1-S115. [doi: 10.1016/j.kint.2020.06.019] [Medline: 32998798]

XSL•FO
RenderX

7. Perkovic V, Agarwal R, Fioretto P, et al. Management of patients with diabetes and CKD: conclusions from a “Kidney Disease: Improving Global Outcomes” (KDIGO) controversies conference. Kidney Int 2016 Dec;90(6):1175-1183. [doi: 10.1016/j.kint.2016.09.010] [Medline: 27884312]
8. GBD 2021 Diabetes Collaborators. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2023 Jul 15;402(10397):203-234. [doi: 10.1016/S0140-6736(23)01301-6] [Medline: 37356446]
9. Wilmer WA, Rovin BH, Hebert CJ, Rao SV, Kumor K, Hebert LA. Management of glomerular proteinuria: a commentary. J Am Soc Nephrol 2003 Dec;14(12):3217-3232. [doi: 10.1097/01.asn.0000100145.27188.33] [Medline: 14638920]
10. Chen TK, Knicely DH, Grams ME. Chronic kidney disease diagnosis and management: a review. JAMA 2019 Oct 1;322(13):1294-1304. [doi: 10.1001/jama.2019.14745] [Medline: 31573641]
11. Chronic Kidney Disease Prognosis Consortium, Matsushita K, van der Velde M, et al. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. Lancet 2010 Jun 12;375(9731):2073-2081. [doi: 10.1016/S0140-6736(10)60674-5] [Medline: 20483451]
12. Hashmi MF, Benjamin O, Lappin SL. End-Stage Renal Disease: StatPearls Publishing; 2023. [Medline: 29763036]
13. Mahemuti N, Zou J, Liu C, Xiao Z, Liang F, Yang X. Urinary albumin-to-creatinine ratio in normal range, cardiovascular health, and all-cause mortality. JAMA Netw Open 2023 Dec 1;6(12):e2348333. [doi: 10.1001/jamanetworkopen.2023.48333] [Medline: 38113044]
14. Garcia Sanchez JJ, Thompson J, Scott DA, et al. Treatments for chronic kidney disease: a systematic literature review of randomized controlled trials. Adv Ther 2022 Jan;39(1):193-220. [doi: 10.1007/s12325-021-02006-z] [Medline: 34881414]
15. Bao W, Zhang M, Li N, Yao Z, Sun L. Efficacy and safety of finerenone in chronic kidney disease associated with type 2 diabetes: a systematic review and meta-analysis of randomized clinical trials. Eur J Clin Pharmacol 2022 Dec;78(12):1877-1887. [doi: 10.1007/s00228-022-03408-w] [Medline: 36273065]
16. Shiau CH, Tsau LY, Kao CC, et al. Efficacy and safety of sodium-glucose cotransporter-2 inhibitors in patients with chronic kidney disease: a systematic review and meta-analysis. Int Urol Nephrol 2024 Apr;56(4):1359-1381. [doi: 10.1007/s11255-023-03789-6] [Medline: 37752340]
17. de Boer IH, Khunti K, Sadusky T, et al. Diabetes management in chronic kidney disease: a consensus report by the American Diabetes Association (ADA) and Kidney Disease: Improving Global Outcomes (KDIGO). Kidney Int 2022 Nov;102(5):974-989. [doi: 10.1016/j.kint.2022.08.012] [Medline: 36202661]
18. Kovesdy CP. Epidemiology of chronic kidney disease: an update 2022. Kidney Int Suppl (2011) 2022 Apr;12(1):7-11. [doi: 10.1016/j.kisu.2021.11.003] [Medline: 35529086]
19. Gheith O, Farouk N, Nampoory N, Halim MA, Al-Otaibi T. Diabetic kidney disease: world wide difference of prevalence and risk factors. J Nephropharmacol 2015 Oct 9;5(1):49-56. [Medline: 28197499]
20. American Diabetes Association. 11. Microvascular complications and foot care: standards of medical care in diabetes-2019. Diabetes Care 2019 Jan;42(Suppl 1):S124-S138. [doi: 10.2337/dc19-S011] [Medline: 30559237]
21. Cosentino F, Grant PJ, Aboyans V, et al. 2019 ESC guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD. Eur Heart J 2020 Jan 7;41(2):255-323. [doi: 10.1093/eurheartj/ehz486] [Medline: 31497854]
22. Rasaratnam N, Salim A, Blackberry I, et al. Urine albumin-creatinine ratio variability in people with type 2 diabetes: clinical and research implications. Am J Kidney Dis 2024 Jul;84(1):8-17.e1. [doi: 10.1053/j.ajkd.2023.12.018] [Medline: 38551531]
23. Hagnäs M, Sundqvist H, Jokelainen J, et al. The prevalence of chronic kidney disease and screening of renal function in type 2 diabetic patients in Finnish primary healthcare. Prim Care Diabetes 2020 Dec;14(6):639-644. [doi: 10.1016/j.pcd.2020.05.005] [Medline: 32527663]
24. Manski-Nankervis JAE, Thuraisingam S, Lau P, et al. Screening and diagnosis of chronic kidney disease in people with type 2 diabetes attending Australian general practice. Aust J Prim Health 2018 Jul;24(3):280-286. [doi: 10.1071/PY17156] [Medline: 29807557]
25. Chu L, Bhogal SK, Lin P, et al. AWAREness of diagnosis and treatment of chronic kidney disease in adults with type 2 diabetes (AWARE-CKD in T2D). Can J Diabetes 2022 Jul;46(5):464-472. [doi: 10.1016/j.jcjd.2022.01.008] [Medline: 35739044]
26. Chu CD, McCulloch CE, Banerjee T, et al. CKD awareness among US adults by future risk of kidney failure. Am J Kidney Dis 2020 Aug;76(2):174-183. [doi: 10.1053/j.ajkd.2020.01.007] [Medline: 32305206]
27. Mueller C, Schürks M, Neußer T, et al. Identification of patients with elevated urine albumin-to-creatinine ratio levels in a type 2 diabetes mellitus cohort based on data submitted by patients via a smartphone app (SMART-Finder): protocol for an observational study. JMIR Res Protoc 2023 Apr 5;12:e44996. [doi: 10.2196/44996] [Medline: 37018036]
28. Kodner C. Diagnosis and management of nephrotic syndrome in adults. Am Fam Physician 2016 Mar 15;93(6):479-485. [Medline: 26977832]
29. Kuritzky L, Toto R, Van Buren P. Identification and management of albuminuria in the primary care setting. J Clin Hypertens (Greenwich) 2011 Jun;13(6):438-449. [doi: 10.1111/j.1751-7176.2010.00424.x] [Medline: 21649844]
30. American Diabetes Association. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes-2021. Diabetes Care 2021 Jan;44(Suppl 1):S15-S33. [doi: 10.2337/dc21-S002] [Medline: 33298413]

XSL•FO
RenderX

31. Ludwig K, Graf von der Schulenburg JM, Greiner W. German value set for the EQ-5D-5L. Pharmacoeconomics 2018 Jun;36(6):663-674. [doi: 10.1007/s40273-018-0615-8] [Medline: 29460066]
32. Mancia G, Kreutz R, Brunström M, et al. 2023 ESH guidelines for the management of arterial hypertension The Task Force for the management of arterial hypertension of the European Society of Hypertension: endorsed by the International Society of Hypertension (ISH) and the European Renal Association (ERA). J Hypertens 2023 Dec 1;41(12):1874-2071. [doi: 10.1097/HJH.0000000000003480] [Medline: 37345492]
33. Patel JB, Sapra A. Nephrotoxic Medications: StatPearls Publishing; 2023. [Medline: 31985937]
34. Ryz K, Tangri N, Verrelli M, et al. A before and after cross-sectional analysis of a public health campaign to increase kidney health awareness in a Canadian province. BMC Res Notes 2015 Nov 20;8:695. [doi: 10.1186/s13104-015-1662-2] [Medline: 26590133]
35. Kellerer M, Wanner C. Kidney diseases in type 2 diabetes mellitus : overview and implementation of guidelines, position papers and practical recommendations for diagnostics and monitoring. Inn Med (Heidelb) 2023 Dec;64(12):1211-1217. [doi: 10.1007/s00108-023-01610-1] [Medline: 37955643]

Abbreviations

ACR: albumin-to-creatinine ratio
ADA: American Diabetes Association
CKD: chronic kidney disease
DTSQ: Diabetes Treatment Satisfaction Questionnaire
eCRF: electronic case report form
eGFR: estimated glomerular filtration rate
HbA_{1c}: glycated hemoglobin
HCP: health care professional
KDIGO: Kidney Disease: Improving Global Outcomes
SGLT2i: sodium-glucose cotransporter 2 inhibitor
T2DM: type 2 diabetes mellitus
UACR: urine albumin-to-creatinine ratio

Edited by A Mavragani; submitted 07.05.24; peer-reviewed by JC Galle, JH Shubrook; revised version received 08.01.25; accepted 09.01.25; published 18.02.25.

Please cite as:
Mueller C, Neusser T, Thate-Waschke I, Nowicki J, Plominski T, Griesinger R, Kessner S, Martin S
Disease Awareness in Patients With Type 2 Diabetes: Analysis of Baseline Data From the SMART-Finder Observational Study
JMIR Form Res 2025;9:e60246
URL: https://formative.jmir.org/2025/1/e60246
doi:10.2196/60246

© Christian Mueller, Thomas Neusser, Inga Thate-Waschke, Julia Nowicki, Tomasz Plominski, Regine Griesinger, Stefanie Kessner, Stephan Martin. Originally published in JMIR Formative Research (https://formative.jmir.org), 18.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Very Low–Carbohydrate Program in Adults With Metabolic Dysfunction–Associated Steatotic Liver Disease and Phospholipase Domain–Containing Protein 3 Risk Genotype: Pre-Post Intervention Study

Laura R Saslow[1], PhD; Jamie Krinock[1], BS; Alison O'Brien[1], MPH; Kaitlyn Raymond[1], BS; Hovig Bayandorian; Judith T Moskowitz[2], MPH, PhD; Jennifer Daubenmier[3], PhD; Antonino Oliveri[4], PhD; Deanna J Marriott[5], PhD; Dina H Griauzde[6,7], MD, MSc; Elizabeth K Speliotes[4,8], MPH, MD, PhD

[1]Department of Health Behavior and Biological Sciences, School of Nursing, University of Michigan, Ann Arbor, MI, United States

[2]Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[3]Institute of Holistic Health Studies, San Francisco State University, San Francisco, CA, United States

[4]Division of Gastroenterology, Department of Internal Medicine, University of Michigan Medical School, Medical Sciences Building II, Room 4741, Ann Arbor, MI, United States

[5]Department of Systems, Populations and Leadership, School of Nursing, University of Michigan, Ann Arbor, MI, United States

[6]VA Ann Arbor Healthcare System, Ann Arbor, MI, United States

[7]Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI, United States

[8]Gilbert S Omenn Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, United States

Corresponding Author:
Elizabeth K Speliotes, MPH, MD, PhD
Division of Gastroenterology, Department of Internal Medicine, University of Michigan Medical School, Medical Sciences Building II, Room 4741, Ann Arbor, MI, United States

Abstract

Background: Insulin resistance and the G allele of rs738409 interact to create a greater risk of metabolic dysfunction–associated steatotic liver disease.

Objective: This study aims to confirm that one promising way to reduce insulin resistance is by following a very low–carbohydrate (VLC) dietary pattern.

Methods: Adults with rs738409-GG or -CG with liver steatosis and elevated liver function tests, were taught an ad libitum VLC diet, positive affect and mindful eating skills, goal setting, and self-monitoring and given feedback and coaching for 4 months. We measured liver steatosis, anthropometric, serum metabolic diet adherence, and quality of life measures.

Results: In this small pilot trial, of the 11 participants enrolled, 9 (82%) participants completed outcomes. All 11 participants viewed at least 1 session of the intervention, and 8 (73%) participants viewed at least half of the sessions. Among the 9 participants who provided 4-month self-report information, intervention satisfaction was high (mean 6.22, 95% CI 5.58-6.85), with 5 (56%) participants rating the intervention the top score, and 4 (44%) participants reporting they did not plan to stop following the VLC diet. Across participants with a 4-month hepatic liver fat percent measurement, the percent change in liver fat was −33.17% (95% CI −86.48 to 20.14), and in only the participants who were adherent to the eating pattern, the percent change in liver fat was −53.12% (95% CI −71.25 to −34.99). Amongst participants with a 4-month hepatic liver fat percent measurement, 6 out of 8 (75%) participants were considered responders, with a relative decline in liver fat ≥30%, and of the 9 participants with a 4-month body weight, 9 (100%) participants lost ≥5% of their body weight. There were no serious adverse events.

Conclusions: Results suggest the feasibility, acceptability, and preliminary efficacy of the VLC intervention in adults with higher genetic risk for metabolic dysfunction–associated steatotic liver disease, although there is a need for further studies given the small sample size and the high risk of substantial biases in this small pilot study.

(JMIR Form Res 2025;9:e60051) doi:10.2196/60051

KEYWORDS

metabolic dysfunction–associated steatotic liver disease; ketogenic diet; low carbohydrate; adult; genotype; insulin; insulin resistance; metabolic dysfunction; dietary pattern; type 2 diabetes; T2DM; single-arm pilot trial; liver function test; genome; non-alcoholic fatty liver disease

XSL•FO
RenderX

Introduction

Metabolic dysfunction–associated steatotic liver disease (MASLD) is an increasingly prevalent disease afflicting about 25% of US adults and more than 50% of people with type 2 diabetes [1]. MASLD is caused by the deposition of excess fat in the liver, not due to alcohol, and can lead to advanced liver disease in the form of inflammation (metabolic dysfunction–associated steatohepatitis), fibrosis/cirrhosis (scarring), and hepatocellular carcinoma (liver cancer) [2]. Advanced MASLD is present in more than 90% of severely obese individuals [3]. It is associated with a shorter lifespan [4] and is expected to become the leading indication for liver transplantation in the United States [5].

Several dietary approaches have been suggested to treat MASLD, but no single approach has been found to be superior. For example, time-restricted eating and low-calorie, low-fructose, low-carbohydrate, very low–carbohydrate (VLC), and Mediterranean diets have all been studied in the context of MASLD [6-8].

We and others have shown that MASLD is heritable and the strongest common risk factor is having the I148M variant for the patatin-like phospholipase domain–containing protein 3 (PNPLA3) gene [9]. We have previously shown that individuals who carry this allele have a multiplicative risk of developing MASLD if they also have insulin resistance. The largest gene-environment meta-analysis study to date included approximately 15,000 individuals of diverse ancestry and demonstrated that rs738409-GG individuals had a 57% increased risk of MASLD if their insulin levels were in the highest versus lowest quartile of insulin, whereas rs738409-CC individuals had only a 32% increased risk, suggesting the presence of gene-environment interaction. Despite being at high risk of liver disease, individuals with rs738409-GG or -CG and insulin resistance are not being specifically targeted by programs aiming to reduce insulin resistance.

One promising way to reduce insulin resistance is through the use of a VLC eating pattern, also known as a ketogenic or “keto” eating pattern, which is a very reduced carbohydrate, moderate protein, and higher fat eating pattern. A VLC eating pattern has been found to be effective for reducing insulin resistance, body weight, inflammation, intrahepatic lipid content, and de novo fatty acid synthesis and fatty liver deposits, with some preliminary research showing the benefit of a VLC for MASLD [10-14]. On the other hand, high-carbohydrate, low-fat diets are related to greater liver inflammation or damage [15,16].

The National Institutes of Health’s obesity-related behavioral intervention trials model for behavioral intervention development [17] encourages preliminary testing of an intervention in a subpopulation before conducting a full-scale clinical trial. In this research, we built on previous research findings demonstrating that a VLC eating pattern can be taught using a web-based program that includes positive affect skills, which aim to increase the frequency with participants experience positive emotions, to improve intervention adherence and satisfaction [18,19]; mindful eating, to help participants cope with emotional and hedonic-driven eating, drivers of weight gain [20-22]; and other strategies including goal setting and self-monitoring; personalized feedback and support from a coach; social support; reminders; and booster messages [23].

We hypothesized that a VLC eating pattern might be able to achieve MASLD improvement or even reversal in adults with steatosis or mild fibrosis, especially in the higher-risk subpopulations of rs738409-GG or -CG individuals. As a first step, in this small pilot trial of adults with MASLD, we examined intervention feasibility, intervention acceptability, and physical and patient-reported outcomes of rs738409-GG or -CG individuals.

Methods

Study Design

This was a single-arm pilot trial of an online, VLC dietary intervention.

Ethical Considerations

The research was approved by the University of Michigan Institutional Review Board (HUM00154361) and registered at ClinicalTrials.gov (NCT05010070). The study followed the ethical standards of the responsible committee on human experimentation (institutional and national) and abided by the Helsinki Declaration of 1975, as revised in 2000. All participants provided informed consent and could withdraw from the study at any point. No writing assistance was used for this manuscript. Recruitment began on August 31, 2021, and the outcomes of the trial were finalized by October 3, 2022.

Participant Recruitment and Screening

Potential participants were contacted based on their rs738409 genotype in the Michigan Genomics Initiative (MGI), a research initiative that collects and genotypes blood samples and allows researchers to link this genetic information to patients. As part of the initiative, participants had already provided informed consent for broad research purposes [24]. Participants interested in our pilot study were referred to a webpage that described the trial and linked to an online screening survey (Qualtrics). The screening survey included questions used to assess eligibility. If individuals passed this, they were sent a video describing the goals, procedure, and pros and cons of participating in the trial. If they continued to express interest in the trial, they consented, baseline measures were collected, and then they were enrolled in the trial. Inclusion criteria included a baseline magnetic resonance imaging with liver steatosis but not cirrhosis, having been identified based on information from the MGI database, and having elevated liver function tests. Exclusion criteria included non-MASLD causes of elevated liver function tests; use of exogenous insulin; planned or history of weight loss surgery; active substance use or untreated mental health condition that could pose a safety risk; advanced medical conditions including as current chemotherapy, heart failure, or kidney failure; type 1 diabetes; Cushing syndrome; adherence to a vegan or dietary vegetarian; pregnant or planning to get pregnant in the next 6 months; currently enrolled in a weight loss program or other investigative study that might conflict with this research; taking medications known to cause weight gain or loss; or recent decompensation/hospitalization.

XSL•FO
RenderX

Intervention

The online intervention included 16 weekly emails with a video and handouts, as well as text messages and email-based coaching. At baseline, participants received a body weight scale (Bodytrace) and urine ketone strips. Participants were also sent VLC cookbooks by mail 3 times throughout the intervention [25-27]. We taught participants to follow a VLC eating pattern, consisting of 20 - 35 net (nonfiber) grams of carbohydrates per day. At week 6, we encouraged participants to become physically active for at least 150 minutes of moderate-intensity physical activity per week as well as to sleep enough to feel well rested. The intervention also taught skills related to feeling more positive affect and eating mindfully [28]. Throughout, we asked participants to track their dietary intake and to step on the body weight scale at least once a week. The program's email-based coach provided feedback and support and we sent near daily, unique text messages that provided reminders of the intervention content.

Data Collection

At baseline, participants were asked to self-weigh using the scale, complete a fasting blood draw and magnetic resonance imaging at Michigan Medicine, complete an online survey (Qualtrics); and complete a 2-day dietary recall with a trained dietitian by phone. At 4 months, at the end of the intervention, these were repeated. Prior to each week's didactic session, we asked participants questions about their dietary adherence and experiences. Procedures associated with the research were either covered by the study or the participants' insurance.

Outcome Measures

Intervention Feasibility

Trial and Intervention Retention

We defined trial retention as the percentage of participants who completed the postintervention outcomes divided by the total number of trial participants enrolled. We assessed intervention retention based on intervention participation: (1) viewing at least 1 session, and (2) being active in the program, defined as viewing at least 8 out of 16 (50%) of the sessions.

Dietary Adherence

At baseline and 4 months, participants took part in two 24-hour dietary recalls. We considered participants adherent to the eating pattern if their net carbohydrates at 4 months were below 60 grams of carbohydrates per day and their overall calories did not increase by more than 400 kcal per day (as this was ad libitum, in which participants are asked to eat when hungry and stop when full, which typically leads to lower hunger).

Intervention Acceptability

General Intervention Satisfaction

We asked participants, "How would you rate your overall satisfaction with the program?" Response options ranged from 1="not at all satisfied" to 7="very satisfied." To assess potential acceptability, we also asked participants to answer, "How long can you see yourself following your assigned diet?"

Intervention Skills Satisfaction

Participants rated their satisfaction with the positive affect skills and, separately, the mindfulness skills. Response options ranged from 1="don't include them, they were not helpful" to 7="you must include them, they were very helpful."

Qualitative Feedback

At 4 months, using open-ended questions, we asked participants about their experiences with the intervention and recommendations for improvement.

Patient-Centered Outcomes

Change in Liver Percent Fat and Fibrosis

We used magnetic resonance imaging to assess the percent fat of the left and right liver lobes, which we then combined. We also used magnetic resonance imaging to quantify the fibrosis in the liver. We examined what percent of participants lost at least 30% of their liver fat, a clinically significant threshold [29]. We a priori registered a reduction in liver fat as our primary outcome of the trial. We therefore examined this outcome for all participants and for adherent participants.

Body Weight Changes

Participants self-weighed throughout the trial and this weight was sent automatically to the trial's staff through the scale's connection to a cellular network. Mean change in weight and BMI (using baseline self-reported height) was calculated at 4 months compared with baseline. Mean percent weight loss was defined as (weight at 4 months – baseline weight) / (baseline weight) *100. We examined what percent of participants lost at least 5% of their body weight, a clinically significant threshold [30].

Blood Test Changes

We assessed liver function with aspartate aminotransferase and alanine aminotransferase, glucose control with glycated hemoglobin (HbA_{1c}) and fasting glucose, fasting insulin, Homeostatic Model Assessment for Insulin Resistance (HOMA-IR), and lipids (triglycerides, high-density lipoprotein or HDL, and low-density lipoprotein or LDL), and inflammation (with C-reactive protein). We calculated the mean change at 4 months compared with baseline.

Self-Rated Change in Health

We asked participants to rate how their health changed over the intervention by answering the questions, "How much do you think your health has changed as a result of participating in this program?" (Response options ranged from 1="very much worse" to 7="very much better").

Changes in Chronic Liver Disease Questionnaire (CLDQ)

We measured change in MASLD-related quality of life with the chronic liver disease questionnaire (CLDQ), which assesses 6 dimensions, rated from 1 to 7, with a higher number reflecting worse symptoms: abdominal symptoms (bloating, discomfort, and pain), activity (trouble with eating, physical ability), emotional (poor mood, sleep, and ability to concentrate), fatigue,

XSL•FO
RenderX

systemic symptoms (body pain, dry mouth, itching, and muscle cramps), and worry about MASLD [31].

Changes in Psychological Factors

We assessed psychological changes from baseline to 4 months for positive affect with the Scale of Positive and Negative Experience [32], mindful eating with the reliance on hunger and satiety cues subscale of the Intuitive Eating Scale-2 [33], and stress-based eating with the Palatable Eating Motives Scale coping subscale [34].

Statistical Analysis

For statistical significance, we followed the guidance for the use of frequentist inferential statistics in public health [35]. We completed an intention-to-treat analysis with all available data and analyses that were limited to only adherent individuals, with no imputation for missing data. Means and 95% CIs were calculated for continuous variables.

Results

Sample Characteristics

Of the 97 individuals who expressed interest in study participation and were screened for study eligibility, 55 did not meet inclusion criteria (including, as the top reasons for ineligibility: 13 were taking insulin, 11 had prior weight loss surgery or were taking weight loss medications, 8 had normal liver function tests, 6 had cirrhosis of the liver, 3 had heart failure or another related heart condition), 12 were eligible based on the prescreening survey but did not reply to follow-up requests, and 5 were eligible but declined to participate. Eleven participants enrolled in the trial.

At baseline, participants were 38 - 77 years old (mean 55.54 years, SD 11.85 years), 8 out of 11 (72.7%) participants were male, most were White and not Hispanic (10/11, 90.9%), with 1 being White and Hispanic. Of the 11 participants, 7 (63.6%) participants were rs738409-GG and 4 (36.4%) participants were rs738409-CG. Other baseline values are in Table 1.

XSL•FO
RenderX

Table . Baseline for all participants (n=11).

Outcomes	Values
Age (years), mean (SD)	55.54 (11.85)
Sex (male), n (%)	8 (73)
Race (White and not Hispanic), n (%)	10 (91)
rs738409-GG (vs rs738409-CG), n (%)	7 (63.6)
Weight (kg), mean (95% CI)	96.19 (83.24-109.14)
BMI (kg/m^2), mean (95% CI)	33.79 (29.60-37.98)
Liver lobe fat (%), mean (95% CI)	23.10 (14.25-37.98)
Liver lobe fibrosis, mean (95% CI)	2.46 (2.03-2.89)
AST[a] (U/L), mean (95% CI)	49.91 (24.85-74.97)
ALT[b] (U/L), mean (95% CI)	78.36 (36.90-119.83)
HbA_{1c}[c] (mmol/mol), mean (95% CI)	40.22 (35.86-44.58)
Glucose (mg/dL), mean (95% CI)	101.55 (91.44-111.65)
Insulin (mIU/mL), mean (95% CI)	21.15 (14.01-28.28)
HOMA-IR[d], mean (95% CI)	5.44 (3.51-7.37)
Triglycerides (mg/dL) mean (95% CI)	127.82 (87.70-167.94)
HDL[e] (mg/dL), mean (95% CI)	48.00 (35.56-60.44)
LDL[f] (mg/dL), mean (95% CI)	84.82 (63.13-106.51)
C-reactive protein (mg/dL), mean (95% CI)	0.79 (0.24-1.34)
CLDQ[g], mean (95% CI)	
Overall	2.62 (1.93-3.31)
Abdominal symptoms	2.49 (1.65-3.32)
Activity	2.33 (1.55-3.12)
Emotional function	2.60 (1.80-3.39)
Fatigue	2.98 (2.12-3.84)
Systemic symptoms	2.43 (1.67-3.19)
Worry about MASLD[h]	2.75 (1.78-3.73)
Positive affect, mean (95% CI)	24.18 (20.99-27.37)
Mindful eating, mean (95% CI)	2.88 (2.48-3.28)
Stress-based eating, mean (95% CI)	1.70 (1.24-2.17)
Energy, kilocalories, mean (95% CI)	1901.92 (1547.20-2256.64)
Carbohydrates (g), mean (95% CI)	192.40 (139.66-245.15)
Fiber (g), mean (95% CI)	17.23 (12.18-22.28)
Net (Non-fiber) Carbohydrates (g), mean (95% CI)	175.17 (123.97-226.37)
Total fat (g), mean (95% CI)	85.99 (67.55-104.43)
Monounsaturated fat (g), mean (95% CI)	30.38 (23.13-37.63)
Polyunsaturated fat (g), mean (95% CI)	17.26 (11.37-23.15)
Saturated Fat (g), mean (95% CI)	30.88 (23.74-38.02)
Protein (g), mean (95% CI)	80.31 (67.57-94.06)

[a]AST: aspartate aminotransferase.

[b]ALT: alanine aminotransferase.

XSL•FO
RenderX

[c]HbA_{1c}: glycated hemoglobin.

[d]HOMA-IR: Homeostatic Model Assessment for Insulin Resistance.

[e]HDL: high-density lipoprotein.

[f]LDL: low-density lipoprotein.

[g]CLDQ: chronic liver disease questionnaire.

[h]MASLD: metabolic dysfunction–associated steatotic liver disease.

Intervention Feasibility

Trial and Intervention Retention

Of the 11 participants who were enrolled, 9 (82%) participants completed study outcomes. All 11 participants viewed at least 1 session of the intervention, and 8 (73%) participants viewed at least half of the sessions.

Dietary Adherence

Nine participants completed the pre- and post-intervention dietary recalls. Changes in nutrient intake are shown in Table 2. In Table 3, we show results for just the 6 participants who were adherent to the eating pattern.

XSL•FO
RenderX

Table . Change in outcomes from baseline to post for all participants with outcome data. For all outcomes n=9, except for liver lobe fat % (n=8), liver lobe fibrosis (n=6), and C-reactive protein (n=4).

Outcomes	Baseline (mean, 95% CI)	Post (mean, 95% CI)	Percent change (mean, 95% CI)	Change (mean, 95% CI)
Weight (kg)	96.19 (83.24 to 109.14)	85.50 (74.63 to 96.37)	−10.88 (−13.06 to −8.69)	−10.69 (−13.37 to −8.00)
BMI (kg/m^2)	32.93 (28.03 to 37.83)	29.26 (25.29 to 33.22)	−10.85 (−13.01 to −8.69)	−3.68 (−4.74 to −2.62)
Liver lobe fat (%)	19.13 (12.22 to 26.04)	12.06 (5.64 to 18.48)	−33.17 (−86.48 to 20.14)	−7.06 (−15.50 to 1.37)
Liver lobe fibrosis	2.41 (2.04 to 2.78)	2.09 (1.59 to 2.59)	−13.22 (−31.39 to 4.95)	−0.33 (−0.82 to 0.17)
AST[a] (U/L)	50.22 (18.48 to 81.97)	41.33 (2.88 to 79.78)	−24.30 (−40.89 to −7.70)	−8.89 (−19.52 to 1.74)
ALT[b] (U/L)	79 (26.35 to 131.65)	54.44 (−4.97 to 113.86)	−35.16 (−56.14 to −14.17)	−24.56 (−49.79 to 0.68)
HbA_{1c}[c] (mmol/mol)	40.22 (35.86 to 44.58)	36.33 (32.83 to 39.83)	−9.40 (−13.50 to −5.30)	−3.89 (−5.75 to -2.03)
Glucose (mg/dL)	105.56 (96.32 to 114.79)	95.56 (84.86 to 106.25)	−9.55 (−15.79 to −3.30)	−10.00 (−16.33 to -3.67)
Insulin (mIU/mL)	22.36 (14.01 to 30.7)	14.49 (9.04 to 19.94)	−30.71 (−55.17 to −6.25)	−7.87 (−15.44 to −0.29)
HOMA-IR[d]	5.88 (3.69 to 8.07)	3.54 (2.11 to 4.97)	−36.75 (−60.54 to −12.96)	−2.34 (−4.30 to −0.38)
Triglycerides (mg/dL)	123.89 (86.9 to 160.87)	107.89 (64.33 to 151.45)	−14.32 (−25.91 to −2.74)	−16.00 (−28.74 to −3.26)
HDL[e] (mg/dL)	50.67 (35.63 to 65.7)	47.78 (37.57 to 57.99)	−2.50 (−14.87 to 9.88)	−2.89 (−9.62 to 3.84)
LDL[f] (mg/dL)	80.11 (55.84 to 104.38)	83.44 (65.01 to 101.88)	9.33 (−9.12 to 27.78)	3.33 (−9.75 to 16.42)
C-reactive protein (mg/dL)	1.55 (0.59 to 2.52)	1.18 (0.36 to 2.00)	−23.58 (−59.94 to 12.77)	−0.38 (−1.07 to 0.32)
CLDQ[g]				
Overall	2.33 (1.66 to 3.01)	1.68 (1.47 to 1.89)	−20.47 (−41.16 to 0.22)	−0.65 (−1.40 to 0.09)
Abdominal symptoms	2.67 (1.55 to 3.78)	1.37 (1.07 to 1.67)	−33.15 (65.25 to −1.05)	−1.30 (−2.48 to −0.11)
Activity	2.33 (1.49 to 3.17)	1.81 (1.16 to 2.47)	−9.88 (−45.53 to 25.77)	−0.52 (−1.58 to 0.55)
Emotional function	2.08 (1.58 to 2.58)	1.54 (1.21 to 1.87)	−22.55 (−39.55 to −5.55)	−0.54 (−1.11 to 0.03)
Fatigue	2.72 (1.62 to 3.83)	1.71 (1.13 to 2.29)	−27.05 (−49.50 to −4.59)	−1.01 (−2.02 to −0.01)
Systemic symptoms	2.31 (1.39 to 3.24)	1.82 (1.37 to 2.27)	−11.94 (−36.28 to 12.40)	−0.49 (−1.17 to 0.19)
Worry about MASLD[h]	2.18 (1.43 to 2.92)	1.82 (1.17 to 2.47)	−8.20 (−32.58 to 16.18)	−0.36 (−1.11 to 0.40)
Positive affect	25.56 (23.45 to 27.66)	26.78 (24.58 to 28.98)	5.01 (−1.11 to 11.13)	1.22 (−0.26 to 2.70)
Mindful eating	2.93 (2.44 to 3.41)	3.52 (3.24 to 3.79)	26.77 (−2.03 to 55.57)	0.59 (0.16 to 1.03)
Stress-based eating	1.69 (1.1 to 2.29)	1.31 (0.88 to 1.73)	−18.55 (−39.01 to 1.90)	−0.39 (−0.81 to 0.03)
Energy (kilocalories)	1752.25 (1401 to 2103.51)	1585.48 (1231.09 to 1939.88)	−6.97 (−29.49 to 15.55)	−166.77 (−551.21 to 217.67)
Carbohydrates (g)	171.37 (117.21 to 225.53)	88.53 (29.72 to 147.35)	−51.07 (−68.51 to −33.63)	−82.84 (−122.62 to −43.05)
Fiber (g)	17.35 (10.91 to 23.8)	18.88 (12.64 to 25.12)	22.28 (−28.19 to 72.76)	1.53 (−6.41 to 9.46)
Net (Non-fiber) Carbohydrates (g)	154.02 (102.54 to 205.5)	69.65 (13.66 to 125.65)	−58.62 (−76.36 to −40.89)	−84.36 (−123.63 to −45.09)
Total fat (g)	77.77 (60.71 to 94.83)	100.01 (71.62 to 128.4)	31.45 (−7.46 to 70.37)	22.24 (−4.99 to 49.47)
Monounsaturated fat (g)	28.2 (20.46 to 35.93)	34.43 (25.12 to 43.73)	25.24 (−4.70 to 55.18)	6.23 (−2.36 to 14.82)
Polyunsaturated fat (g)	13.95 (10.19 to 17.72)	18.47 (12.15 to 24.78)	41.43 (−22.53 to 105.39)	4.51 (−1.94 to 10.96)
Saturated fat (g)	28.53 (20.59 to 36.48)	38.48 (25.74 to 51.22)	38.58 (−5.55 to 82.71)	9.95 (−2.82 to 22.72)
Protein (g)	79.81 (63.11 to 96.51)	87.04 (72.31 to 101.76)	13.98 (−13.11 to 41.07)	7.22 (−13.07 to 27.51)

[a]AST: aspartate aminotransferase.

[b]ALT: alanine aminotransferase.

[c] HbA_{1c}: glycated hemoglobin.

[d]HOMA-IR: Homeostatic Model Assessment for Insulin Resistance.

XSL•FO
RenderX

[e]HDL: high-density lipoprotein.

[f]LDL: low-density lipoprotein.

[g]CLDQ: chronic liver disease questionnaire.

[h]MASLD: metabolic dysfunction–associated steatotic liver disease.

XSL•FO
RenderX

Table . Change in outcomes from baseline to post for participants who were adherent to the eating pattern. For all outcomes n=6, except for C-reactive protein (n=3).

Outcomes	Baseline, mean (95% CI)	Post, mean (95% CI)	Percent change, mean (95% CI)	Change, mean (95% CI)
Weight (kg)	96.83 (83.93 to 109.74)	85.15 (74.36 to 95.94)	−11.99 (−14.33 to −9.65)	−11.69 (−14.59 to −8.78)
BMI (kg/m^2)	34.05 (26.76 to 41.34)	29.92 (23.83 to 36.01)	−11.94 (−14.21 to −9.66)	−4.13 (−5.47 to −2.8)
Liver lobe fat (%)	21.67 (10.77 to 32.57)	10.92 (3.05 to 18.79)	−53.12 (−71.25 to −34.99)	−10.75 (−16.38 to −5.12)
Liver lobe fibrosis	2.35 (1.76 to 2.94)	1.88 (1.49 to 2.28)	−17.73 (−37.41 to 1.94)	−0.47 (−0.98 to 0.05)
AST[a] (U/L)	41.67 (31.01 to 52.32)	26.33 (20.79 to 31.87)	−33.94 (−52.93 to −14.95)	−15.33 (−25.72 to −4.95)
ALT[b] (U/L)	71 (35.6 to 106.4)	32.33 (22.52 to 42.15)	−47.49 (−70.67 to -24.31)	−38.67 (−70.38 to −6.95)
HbA_{1c}[c] (mmol/mol)	38.33 (34.48 to 42.18)	35.33 (31.6 to 39.07)	−7.73 (−13.15 to −2.32)	−3.00 (−5.10 to −0.90)
Glucose (mg/dL)	104.33 (90.18 to 118.49)	96.83 (80.54 to 113.12)	−7.36 (−15.76 to 1.05)	−7.50 (−15.65 to 0.65)
Insulin (mIU/mL)	25.3 (13.27 to 37.33)	15.08 (6.04 to 24.12)	−40.52 (−72.62 to −8.42)	−10.22 (−21.53 to 1.10)
HOMA-IR[d]	6.63 (3.43 to 9.82)	3.77 (1.45 to 6.09)	−44.02 (−77.02 to −11.03)	−2.86 (−5.79 to 0.08)
Triglycerides (mg/dL)	126.67 (64.35 to 188.98)	111.5 (38.45 to 184.55)	−13.59 (−32.80 to 5.61)	−15.17 (−36.73 to 6.40)
HDL[e] (mg/dL)	50.83 (25.58 to 76.08)	47.33 (30.03 to 64.64)	−2.60 (−23.00 to 17.80)	−3.50 (−14.46 to 7.46)
LDL[f] (mg/dL)	77 (50.1 to 103.9)	86.5 (63.52 to 109.48)	16.92 (−8.88 to 42.72)	9.50 (−5.70 to 24.70)
C-reactive protein (mg/dL)	1.83 (0.71 to 2.95)	1.43 (-0.57 to 3.44)	−17.16 (−74.65 to 40.33)	−0.40 (−1.71 to 0.91)
CLDQ[g]				
Overall	2 (1.62 to 2.38)	1.7 (1.38 to 2.02)	−12.21 (−38.44 to 14.02)	−0.30 (−0.85 to 0.24)
Abdominal symptoms	2.44 (1.07 to 3.82)	1.28 (0.87 to 1.69)	−31.67 (−81.09 to 17.76)	−1.17 (−2.61 to 0.28)
Activity	1.89 (1.17 to 2.61)	1.67 (0.73 to 2.61)	−2.92 (−56.30 to 50.46)	−0.22 (−1.41 to 0.96)
Emotional function	1.88 (1.6 to 2.15)	1.65 (1.17 to 2.12)	−12.96 (−30.42 to 4.50)	−0.23 (−0.54 to 0.09)
Fatigue	2.12 (1.02 to 3.21)	1.57 (0.8 to 2.33)	−18.50 (−48.68 to 11.67)	−0.55 (−1.41 to 0.31)
Systemic symptoms	2 (1.36 to 2.64)	1.83 (1.26 to 2.4)	−3.96 (−39.18 to 31.27)	−0.17 (−0.75 to 0.42)
Worry about liver disease	1.93 (0.99 to 2.88)	2.03 (1.02 to 3.05)	5.42 (−7.76 to 18.59)	0.10 (−0.22 to 0.42)
Positive affect	25.83 (22.42 to 29.24)	27.33 (24.24 to 30.42)	6.28 (−3.59 to 16.15)	1.50 (−0.87 to 3.87)
Mindful eating	3.03 (2.62 to 3.43)	3.56 (3.1 to 4.01)	18.12 (5.48 to 30.75)	0.53 (0.15 to 0.90)
Stress-based eating	1.71 (0.75 to 2.67)	1.29 (0.66 to 1.92)	−20.79 (−39.39 to −2.20)	−0.42 (−0.85 to 0.01)
Energy (kilocalories)	1836.48 (1298.33 to 2374.62)	1385.79 (933.4 to 1838.17)	−24.28 (−38.40 to −10.16)	−450.69 (−777.83 to −123.54)
Carbohydrates (g)	169.25 (111.67 to 226.82)	60.25 (39.73 to 80.76)	−64.14 (−72.48 to −55.79)	−109.00 (−150.88 to -67.13)
Fiber (g)	19.11 (9.35 to 28.86)	16.86 (7.49 to 26.22)	−5.84 (−61.83 to 50.15)	−2.25 (−13.11 to 8.61)
Net (Non-fiber) Carbohydrates (g)	150.14 (95.45 to 204.83)	43.39 (28.69 to 58.08)	−70.04 (−79.77 to −60.31)	−106.75 (−153.97 to −59.54)
Total fat (g)	84.12 (58.22 to 110.01)	90.09 (49.99 to 130.19)	5.53 (−30.18 to 41.24)	5.97 (−23.95 to 35.90)
Monounsaturated fat (g)	30.1 (17.68 to 42.52)	32.04 (17.99 to 46.1)	7.89 (−29.08 to 44.86)	1.94 (−9.69 to 13.57)
Polyunsaturated fat (g)	15.64 (10.15 to 21.14)	14.96 (6.91 to 23)	−9.79 (−35.91 to 16.34)	−0.69 (−3.89 to 2.52)
Saturated fat (g)	30.58 (17.87 to 43.3)	35.17 (18 to 52.34)	17.23 (−32.59 to 67,06)	4.58 (−12.01 to 21.18)
Protein (g)	86.15 (60.68 to 111.63)	80.55 (61.78 to 99.33)	−4.24 (−23.72 to 15.24)	−5.60 (−25.7 to 14.51)

[a]AST: aspartate aminotransferase.

[b]ALT: alanine aminotransferase.

[c]HbA_{1c}: glycated hemoglobin.

[d]HOMA-IR: Homeostatic Model Assessment for Insulin Resistance.

XSL•FO
RenderX

[e]HDL: high-density lipoprotein.

[f]LDL: low-density lipoprotein.

[g]CLDQ: chronic liver disease questionnaire.

Intervention Acceptability

General Intervention Satisfaction

Of the 9 participants who provided 4-month self-report information, intervention satisfaction was high (mean 6.22, 95% CI 5.58-6.85), with 5 (56%) participants rating the intervention the top score. Only 1 (11%) participant reported that they would stop the assigned eating pattern as soon as the study was over, with 4 (44%) participants reporting that they intended to continue it for at least another few months, and 4 (44%) participants stating that they did not plan to ever stop following it.

Intervention's Skills Satisfaction

The 9 participants who provided 4-month self-report information rated their satisfaction with the positive affect skills (mean 4.78, 95% CI 03.53-6.03), with 1 (11%) participant rating them the top score. Participants rated their satisfaction with the mindfulness skills (mean 4.44, 95% CI 3.09-5.79), with 1 (11%) participant rating them the top score.

Qualitative Feedback

Participants noted barriers that reduced their adherence to and motivation for the intervention, including lack of time, carbohydrate cravings, fatigue, vacations, and social situations. They mentioned factors that supported their adherence to and motivation for the intervention, including accountability by being in the intervention, the program's supportive coach, feeling that the "diet was manageable," weight loss success, and an overall desire to improve their own health. Opinions were mixed about the positive affect and mindful eating skills: one person said that they did not need them and another thought that the "skill building around mindfulness is essential."

Participants noted positive health changes such as weight loss, increased energy, and a changed attitude about food and the "need/desire to eat food that is not good for me." They noted that their family, friends, and physicians were impressed with their weight loss, health changes, and ability to stick with the program. Some comments included, "My physician noticed I lost weight and said I looked much better. She also encouraged me to continue this after the program is over." "My primary care provider reported he was happy for the change and that my blood tests show I am now healthier than ever."

Secondary Patient-Centered Outcomes

Adverse Events

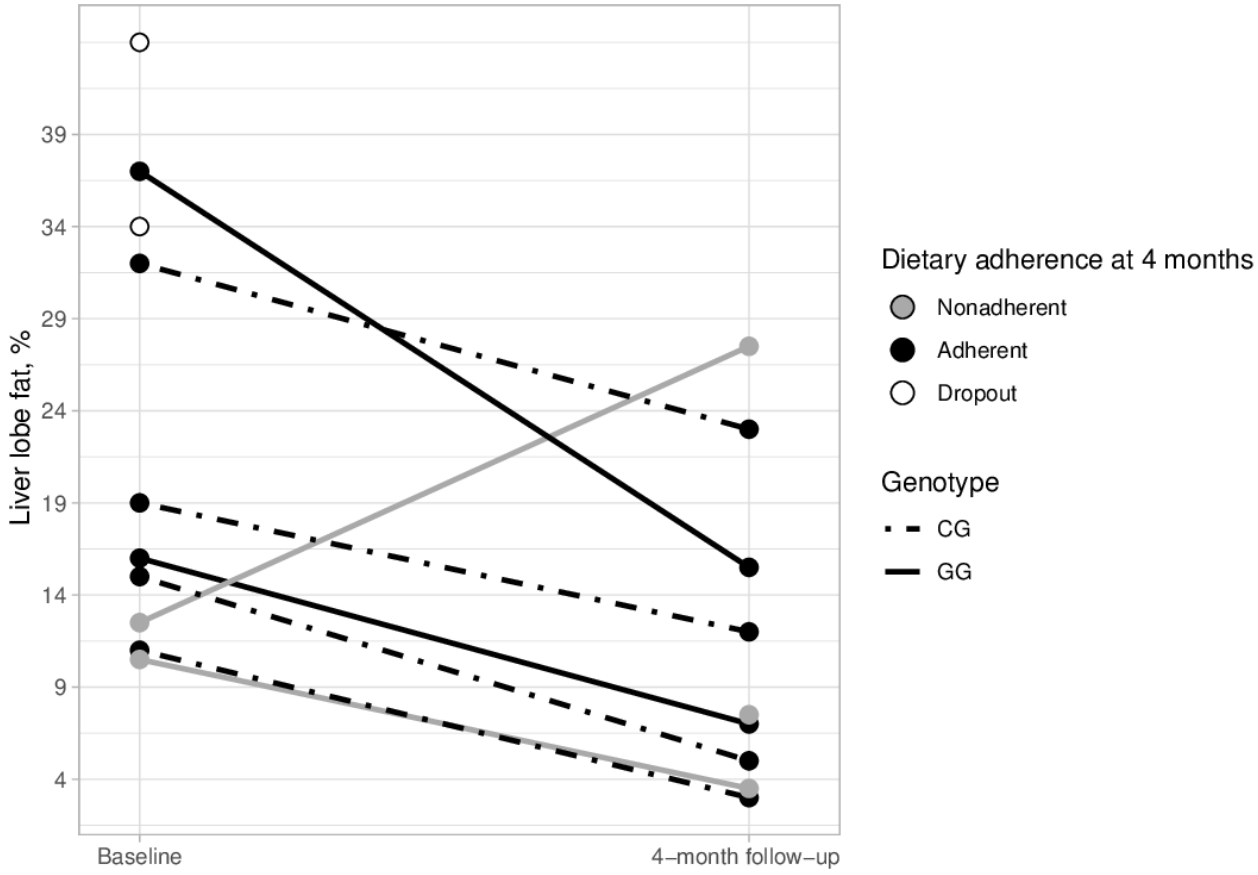
No serious adverse events were reported, but adverse events, possibly unrelated, included diarrhea (later discovered to be a Clostridioides difficile infection), a flare-up of dermatitis, and a broken foot.

Change in Liver Percent Fat, Body Weight, and Blood Tests

In Figure 1 we show the changes in liver lobe fat percent across participants. In Table 2 we show the changes with values for all available participants. For the participants with this outcome, the percent change in liver fat was −33.17% (95% CI −86.48 to 20.14). In Table 3 we show these changes for only the participants who were adherent to the eating pattern. For the adherent participants, the percent change in liver fat was −53.12% (95% CI −71.25 to −34.99).

XSL•FO
RenderX

Figure 1. Change in liver lobe fat percent from baseline to post across participants.



Achievement of ≥30% Liver Fat Loss

Amongst participants with a 4-month hepatic liver fat percent measurement, 6 out of 8 (75%) participants were considered responders, with a relative decline in liver fat ≥30%, with one person not counted losing 28%.

Achievement of ≥5% Body Weight Loss

Amongst participants with a 4-month body weight, 9 out of 9 (100%) of participants lost at least 5% of their body weight.

Self-Rated Change in Health

Two participants reported their health as having gotten a little better, 6 reported that their health was much better, and 1 reported that their health was very much better.

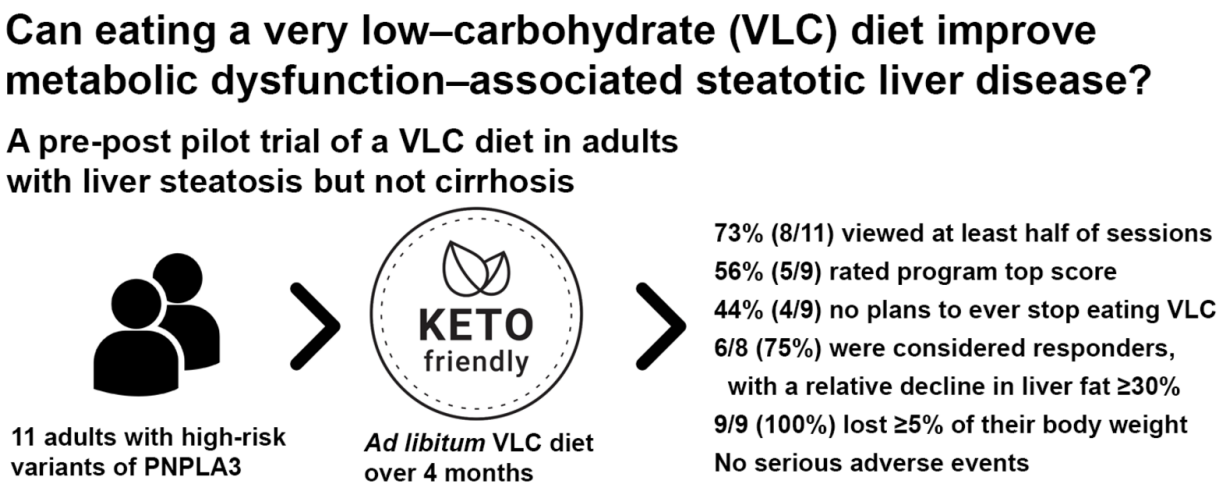
Changes in Chronic Liver Disease Questionnaire and Psychological Factors

Across all participants (Table 2) and for participants who were adherent (Table 3), changes were in the expected, salutary direction.

A summary of the study design and results is shown in Figure 2.

XSL•FO
RenderX

Figure 2. Graphical table of contents. PNPLA3: phospholipase domain–containing protein 3; VLC: very low–carbohydrate.



Discussion

Principal Findings

Results suggest the feasibility, acceptability, and preliminary efficacy of the VLC intervention in adults with higher genetic risk for MASLD, although there is a need for further studies given the small sample size and the high risk of substantial biases in this small pilot study. One concern of the VLC diet is its tolerability. Here we show that the diet was well tolerated with 5 out of 9 participants rating it at the highest satisfaction score. Only 1 out of 9 said they would stop the assigned dietary pattern as soon as the study was over. In contrast, 4 out of 9 said they would continue to follow the diet indefinitely. This is consistent with data where close to 60% of individuals with relapsing multiple sclerosis continued on a VLC or low-carb diet following a trial of the diet, particularly if they had lost a lot of weight and had decreased fatigue [36]. The mindfulness training was less helpful to participants with only 1 out of 9 rating it with the top score. They noted challenges with adhering to the diet as time, carbohydrate cravings, fatigue, vacations, and social situations which need to be addressed as part of a larger program to change eating patterns.

Several previous trials have examined the impact of a VLC eating pattern for adults with MASLD or metabolic dysfunction–associated steatohepatitis, and they found similar results. For example, in a 1-month trial of a VLC diet with 27 participants with MASLD, left hepatic lobe volume decreased by 20% [37]. In a 3-month trial of a Spanish and Mediterranean–adapted VLC diet in 12 men, liver steatosis improved in 93% of participants, and complete fatty liver regression was observed in 21% of participants [12]. In a 6-month trial of a VLC diet with 5 participants, 4 (80%) showed improvements in liver histology, steatosis, inflammatory grade, and fibrosis [10]. In a 2-month trial of a very low-calorie, VLC diet with 20 participants with obesity, the prevalence of liver steatosis (proton density fat fraction>5.4%) decreased from 70% to 30%. Likewise, 2 other nonrandomized, pre-post trials of hypocaloric VLC diets, a 6.5-month [38] and an 8-week trial, both showed improvements in liver steatosis [39].

Five previous trials, with a total of 179 people, have randomized with MASLD to a Mediterranean diet and measured liver-based fat with magnetic resonance imaging. These varied in length of time, with 1 being 1.5 months long, 2 being 3 months long, and 3 being 6 months long. Overall, they have found an average of 12% - 60% relative reduction in hepatic fat percentage [40-44]. Our results and the results of previous VLC trials suggest that a VLC, ketogenic diet may be at least as effective as a Mediterranean diet, although future randomized trials are needed to assess this. Our study is also in line with a previous study where carriers of PNPLA3 G risk allele were able to lose substantial liver fat on a low carbohydrate diet [45].

XSL•FO
RenderX

Limitations

One of our major limitations is the fact that our sample size was small. Our recruitment was limited by the fact that we reached out to adults who had already been genotyped by the MGI. Instead, if we had genotyped interested participants, it may have been easier to recruit more participants. We also required that some of the study-related tests, which could be justified as medical expenses as part of regular health care, including the baseline magnetic resonance imaging and most of the blood tests, be covered by the participants' health insurance. However, not all potential participants were willing to add these charges to their insurance, because it would still incur costs to the participants. In addition, the trial was conducted during the COVID-19 pandemic, which may have influenced who was willing or able to participate or their ability to adhere to dietary instructions.

Moreover, this trial's generalizability is limited, as most participants were White, non-Hispanic men. Future recruitment efforts should focus on recruiting a more diverse population. This trial lacked a control group to which participants could have been randomized to, and thus statements cannot be made about causality.

Another important limitation is duration. As with other trials in this area, participants were tracked for less than a year, which does not allow for the understanding of the longer-term impacts of this nutritional approach nor whether participants may maintain dietary adherence longer term [10,12,37-44]. A growing body of literature including our paper showed that both insulin resistance and the PNPLA3 risk allele can drive the progression of liver disease [46,47]. A limitation of our study is that whether PNPLA3-driven liver fat accumulation causes liver insulin resistance cannot be determined here. Finally, the lack of a PNPLA3 CC control group limits our ability to determine whether the outcomes would be even more improved in the G versus C carriers.

Conclusions and Future Directions

Thus, results supported the feasibility, acceptability, and preliminary efficacy of the VLC intervention in adults thought to have a higher genetic risk for MASLD. Overall, there is a need for longer-term studies to assess the sustainability of benefits and potential long-term risks associated with this way of eating. Moreover, future studies should aim to include a more diverse population, explore the biological mechanisms by which the VLC diet affects MASLD, consider the role of the PNPLA3 genotype, and integrate omics technologies to identify modulated biomarkers and pathways, thereby enhancing the generalizability and depth of the findings.

Acknowledgments

We thank our dedicated participants, Lara Khadr for her help with the figure, Samuel Handelman for his help with the initial grant, and the Michigan Genomics Initiative. For this research, LRS and EKS were supported by R03 DK124741 and P30 DK089503, and Core Services was used, supported by grant DK089503 to the University of Michigan. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

LRS and EKS were involved in the conceptualization, project administration, and methodology of the study. AO and JK were involved in the project administration of the study. KR was the dietary coach. LRS completed the analyses with help from DJM. All authors reviewed and assisted with writing revisions. The authors read and approved the final manuscript.

Conflicts of Interest

LRS' partner, HB, is an inventor of the software used in this study, which purchased a software services agreement for its use. The other authors have no conflicts of interest to declare.

References

1. Amiri Dash Atan N, Koushki M, Motedayen M, et al. Type 2 diabetes mellitus and non-alcoholic fatty liver disease: a systematic review and meta-analysis. Gastroenterol Hepatol Bed Bench 2017;10(Suppl1):S1-S7. [Medline: 29511464]
2. Lazo M, Hernaez R, Eberhardt MS, et al. Prevalence of nonalcoholic fatty liver disease in the United States: the third national health and nutrition examination survey, 1988-1994. Am J Epidemiol 2013 Jul 1;178(1):38-45. [doi: 10.1093/aje/kws448] [Medline: 23703888]
3. Crespo J, Fernández-Gil P, Hernández-Guerra M, et al. Are there predictive factors of severe liver fibrosis in morbidly obese patients with non-alcoholic steatohepatitis? Obes Surg 2001 Jun 1;11(3):254-257. [doi: 10.1381/096089201321336548]
4. Portillo P, Yavuz S, Bril F, Cusi K. Role of insulin resistance and diabetes in the pathogenesis and treatment of nonalcoholic fatty liver disease. Curr Hepatology Rep 2014 Jun;13(2):159-170. [doi: 10.1007/s11901-014-0229-3]
5. Charlton MR, Burns JM, Pedersen RA, Watt KD, Heimbach JK, Dierkhising RA. Frequency and outcomes of liver transplantation for nonalcoholic steatohepatitis in the United States. Gastroenterology 2011 Oct;141(4):1249-1253. [doi: 10.1053/j.gastro.2011.06.061] [Medline: 21726509]

XSL•FO
RenderX

6. Katsagoni CN, Papachristou E, Sidossis A, Sidossis L. Effects of dietary and lifestyle interventions on liver, clinical and metabolic parameters in children and adolescents with non-alcoholic fatty liver disease: a systematic review. Nutrients 2020 Sep 19;12(9):2864. [doi: 10.3390/nu12092864] [Medline: 32961669]
7. Parra-Vargas M, Rodriguez-Echevarria R, Jimenez-Chillaron JC. Nutritional approaches for the management of nonalcoholic fatty liver disease: an evidence-based review. Nutrients 2020 Dec 17;12(12):3860. [doi: 10.3390/nu12123860] [Medline: 33348700]
8. Saeed N, Nadeau B, Shannon C, Tincopa M. Evaluation of dietary approaches for the treatment of non-alcoholic fatty liver disease: a systematic review. Nutrients 2019 Dec 16;11(12):3064. [doi: 10.3390/nu11123064] [Medline: 31888132]
9. Speliotes EK, Yerges-Armstrong LM, Wu J, et al. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. PLoS Genet 2011 Mar;7(3):e1001324. [doi: 10.1371/journal.pgen.1001324] [Medline: 21423719]
10. Tendler D, Lin S, Yancy WS Jr, et al. The effect of a low-carbohydrate, ketogenic diet on nonalcoholic fatty liver disease: a pilot study. Dig Dis Sci 2007 Feb;52(2):589-593. [doi: 10.1007/s10620-006-9433-5] [Medline: 17219068]
11. Mardinoglu A, Wu H, Bjornson E, et al. An integrated understanding of the rapid metabolic benefits of a carbohydrate-restricted diet on hepatic steatosis in humans. Cell Metab 2018 Mar 6;27(3):559-571. [doi: 10.1016/j.cmet.2018.01.005] [Medline: 29456073]
12. Pérez-Guisado J, Muñoz-Serrano A. The effect of the Spanish ketogenic mediterranean diet on nonalcoholic fatty liver disease: a pilot study. J Med Food 2011;14(7-8):677-680. [doi: 10.1089/jmf.2011.0075] [Medline: 21688989]
13. Sevastianova K, Santos A, Kotronen A, et al. Effect of short-term carbohydrate overfeeding and long-term weight loss on liver fat in overweight humans. Am J Clin Nutr 2012 Oct;96(4):727-734. [doi: 10.3945/ajcn.112.038695] [Medline: 22952180]
14. Hudgins LC, Hellerstein M, Seidman C, Neese R, Diakun J, Hirsch J. Human fatty acid synthesis is stimulated by a eucaloric low fat, high carbohydrate diet. J Clin Invest 1996 May 1;97(9):2081-2091. [doi: 10.1172/JCI118645] [Medline: 8621798]
15. Solga S, Alkhuraishe AR, Clark JM, et al. Dietary composition and nonalcoholic fatty liver disease. Dig Dis Sci 2004 Oct;49(10):1578-1583. [doi: 10.1023/b:ddas.0000043367.69470.b7] [Medline: 15573908]
16. Kang H, Greenson JK, Omo JT, et al. Metabolic syndrome is associated with greater histologic severity, higher carbohydrate, and lower fat diet in patients with NAFLD. Am J Gastroenterol 2006 Oct;101(10):2247-2253. [doi: 10.1111/j.1572-0241.2006.00719.x] [Medline: 17032189]
17. Czajkowski SM, Powell LH, Adler N, et al. From ideas to efficacy: the ORBIT model for developing behavioral treatments for chronic diseases. Health Psychol 2015 Oct;34(10):971-982. [doi: 10.1037/hea0000161] [Medline: 25642841]
18. Carr A, Cullen K, Keeney C, et al. Effectiveness of positive psychology interventions: a systematic review and meta-analysis. J Posit Psychol 2021 Nov 2;16(6):749-769. [doi: 10.1080/17439760.2020.1818807]
19. Fredrickson BL, Joiner T. Positive emotions trigger upward spirals toward emotional well-being. Psychol Sci 2002 Mar;13(2):172-175. [doi: 10.1111/1467-9280.00431] [Medline: 11934003]
20. Mason AE, Epel ES, Aschbacher K, et al. Reduced reward-driven eating accounts for the impact of a mindfulness-based diet and exercise intervention on weight loss: data from the SHINE randomized controlled trial. Appetite 2016 May 1;100:86-93. [doi: 10.1016/j.appet.2016.02.009] [Medline: 26867697]
21. Koenders PG, van Strien T. Emotional eating, rather than lifestyle behavior, drives weight gain in a prospective study in 1562 employees. J Occup Environ Med 2011 Nov;53(11):1287-1293. [doi: 10.1097/JOM.0b013e31823078a2] [Medline: 22027541]
22. Lazarevich I, Irigoyen Camacho ME, Velázquez-Alva MDC, Zepeda Zepeda M. Relationship among obesity, depression, and emotional eating in young adults. Appetite 2016 Dec 1;107:639-644. [doi: 10.1016/j.appet.2016.09.011] [Medline: 27620648]
23. Willmott TJ, Pang B, Rundle-Thiele S, Badejo A. Weight management in young adults: systematic review of electronic health intervention components and outcomes. J Med Internet Res 2019 Feb 6;21(2):e10265. [doi: 10.2196/10265] [Medline: 30724736]
24. Zawistowski M, Fritsche LG, Pandit A, et al. The Michigan genomics initiative: a biobank linking genotypes and electronic clinical records in Michigan medicine patients. Cell Genom 2023 Feb 8;3(2):100257. [doi: 10.1016/j.xgen.2023.100257] [Medline: 36819667]
25. MarcAurele L, Volo L. Low Carb Yum 5-Ingredient Keto: 120+ Easy Recipes: Houghton Mifflin Harcourt; 2020:222.
26. Ketchum C. The Everyday Ketogenic Kitchen: 150+ Inspirational Low-Carb, High-Fat Recipes to Maximize Your Health: Victory Belt Publishing Inc; 2017:384.
27. Krampf M. The Wholesome Yum Easy Keto Cookbook: 100 Simple Low Carb Recipes: 10 Ingredients or Less: Harmony Books; 2019:256.
28. Saslow LR, Moskowitz JT, Mason AE, et al. Intervention enhancement strategies among adults with type 2 diabetes in a very low-carbohydrate web-based program: evaluating the impact with a randomized trial. JMIR Diabetes 2020 Sep 9;5(3):e15835. [doi: 10.2196/15835] [Medline: 32902391]

XSL•FO
RenderX

29. Stine JG, Munaganuru N, Barnard A, et al. Change in MRI-PDFF and histologic response in patients with nonalcoholic steatohepatitis: a systematic review and meta-analysis. Clin Gastroenterol Hepatol 2021 Nov;19(11):2274-2283. [doi: 10.1016/j.cgh.2020.08.061] [Medline: 32882428]
30. Blackburn G. Effect of degree of weight loss on health benefits. Obes Res 1995 Sep;3 Suppl 2:211s-216s. [doi: 10.1002/j.1550-8528.1995.tb00466.x] [Medline: 8581779]
31. Younossi ZM, Guyatt G, Kiwi M, Boparai N, King D. Development of a disease specific questionnaire to measure health related quality of life in patients with chronic liver disease. Gut 1999 Aug;45(2):295-300. [doi: 10.1136/gut.45.2.295] [Medline: 10403745]
32. Diener E, Wirtz D, Biswas-Diener R, et al. New measures of well-being. In: Assess Well-Being Springer: Springer; 2009:247-266. [doi: 10.1007/978-90-481-2354-4_12]
33. Tylka TL, Kroon Van Diest AM. The intuitive eating scale-2: item refinement and psychometric evaluation with college women and men. J Couns Psychol 2013 Jan;60(1):137-153. [doi: 10.1037/a0030893] [Medline: 23356469]
34. Burgess EE, Turan B, Lokken KL, Morse A, Boggiano MM. Profiling motives behind hedonic eating. preliminary validation of the palatable eating motives scale. Appetite 2014 Jan;72:66-72. [doi: 10.1016/j.appet.2013.09.016] [Medline: 24076018]
35. Rovetta A, Mansournia MA, Vitale A. For a proper use of frequentist inferential statistics in public health. Glob Epidemiol 2024 Dec;8:100151. [doi: 10.1016/j.gloepi.2024.100151] [Medline: 39021384]
36. Merkel M, Velez-Carrasco W, Hudgins LC, Breslow JL. Compared with saturated fatty acids, dietary monounsaturated fatty acids and carbohydrates increase atherosclerosis and VLDL cholesterol levels in LDL receptor-deficient, but not apolipoprotein E-deficient, mice. Proc Natl Acad Sci U S A 2001 Nov 6;98(23):13294-13299. [doi: 10.1073/pnas.231490498] [Medline: 11606787]
37. Schiavo L, Pilone V, Rossetti G, Barbarisi A, Cesaretti M, Iannelli A. A 4-week preoperative ketogenic micronutrient-enriched diet is effective in reducing body weight, left hepatic lobe volume, and micronutrient deficiencies in patients undergoing bariatric surgery: a prospective pilot study. Obes Surg 2018 Aug;28(8):2215-2224. [doi: 10.1007/s11695-018-3145-8] [Medline: 29502279]
38. Belopolsky Y, Khan MQ, Sonnenberg A, Davidson DJ, Fimmel CJ. Ketogenic, hypocaloric diet improves nonalcoholic steatohepatitis. J Transl Int Med 2020 Mar;8(1):26-31. [doi: 10.2478/jtim-2020-0005] [Medline: 32435609]
39. Rinaldi R, De Nucci S, Castellana F, et al. The effects of eight weeks' very low-calorie ketogenic diet (VLCKD) on liver health in subjects affected by overweight and obesity. Nutrients 2023 Feb 6;15(4):825. [doi: 10.3390/nu15040825] [Medline: 36839183]
40. Ryan MC, Itsiopoulos C, Thodis T, et al. The Mediterranean diet improves hepatic steatosis and insulin sensitivity in individuals with non-alcoholic fatty liver disease. J Hepatol 2013 Jul;59(1):138-143. [doi: 10.1016/j.jhep.2013.02.012] [Medline: 23485520]
41. Properzi C, O'Sullivan TA, Sherriff JL, et al. Ad libitum mediterranean and low-fat diets both significantly reduce hepatic steatosis: a randomized controlled trial. Hepatology 2018 Nov;68(5):1741-1754. [doi: 10.1002/hep.30076] [Medline: 29729189]
42. Marin-Alejandre BA, Abete I, Cantero I, et al. The metabolic and hepatic impact of two personalized dietary strategies in subjects with obesity and nonalcoholic fatty liver disease: the fatty liver in obesity (FLiO) randomized controlled trial. Nutrients 2019 Oct 22;11(10):2543. [doi: 10.3390/nu11102543] [Medline: 31652512]
43. Abbate M, Mascaró CM, Montemayor S, et al. Energy expenditure improved risk factors associated with renal function loss in NAFLD and MetS patients. Nutrients 2021 Feb 15;13(2):629. [doi: 10.3390/nu13020629] [Medline: 33672073]
44. George E, Reddy A, Nicoll A, et al. Impact of a Mediterranean diet on hepatic and metabolic outcomes in non-alcoholic fatty liver disease: the MEDINA randomized controlled trial. J Gastroenterol Hepatol 2021;36(3):160-161. [doi: 10.1111/liv.15264]
45. Sevastianova K, Kotronen A, Gastaldelli A, et al. Genetic variation in PNPLA3 (adiponutrin) confers sensitivity to weight loss-induced decrease in liver fat in humans. Am J Clin Nutr 2011 Jul;94(1):104-111. [doi: 10.3945/ajcn.111.012369] [Medline: 21525193]
46. Barata L, Feitosa MF, Bielak LF, et al. Insulin resistance exacerbates genetic predisposition to nonalcoholic fatty liver disease in individuals without diabetes. Hepatol Commun 2019 Jul;3(7):894-907. [doi: 10.1002/hep4.1353] [Medline: 31334442]
47. Luukkonen PK, Qadri S, Ahlholm N, et al. Distinct contributions of metabolic dysfunction and genetic risk factors in the pathogenesis of non-alcoholic fatty liver disease. J Hepatol 2022 Mar;76(3):526-535. [doi: 10.1016/j.jhep.2021.10.013] [Medline: 34710482]

Abbreviations

CLDQ: chronic liver disease questionnaire
MASLD: metabolic dysfunction–associated steatotic liver disease
MGI: Michigan Genomics Initiative
PNPLA3: phospholipase domain–containing protein 3

XSL•FO
RenderX

VLC: very low–carbohydrate

Edited by A Rovetta; submitted 30.04.24; peer-reviewed by D Kaminska, GP Caviglia; revised version received 24.09.24; accepted 10.10.24; published 10.01.25.

Please cite as:
Saslow LR, Krinock J, O'Brien A, Raymond K, Bayandorian H, Moskowitz JT, Daubenmier J, Oliveri A, Marriott DJ, Griauzde DH, Speliotes EK
A Very Low–Carbohydrate Program in Adults With Metabolic Dysfunction–Associated Steatotic Liver Disease and Phospholipase Domain–Containing Protein 3 Risk Genotype: Pre-Post Intervention Study
JMIR Form Res 2025;9:e60051
URL: https://formative.jmir.org/2025/1/e60051
doi:10.2196/60051

© Laura R Saslow, Jamie Krinock, Alison O'Brien, Kaitlyn Raymond, Hovig Bayandorian, Judith T Moskowitz, Jennifer Daubenmier, Antonio Oliveri, Deanna J Marriott, Dina H Griauzde, Elizabeth K Speliotes. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Continuous Glucose Monitoring Among Patients With Type 1 Diabetes in Rwanda (CAPT1D) Phase I: Prospective Observational Feasibility Study

Jason Baker[1], MD; Giacomo Cappon[2], PhD; Jean Claude Habineza[3,4], MD; Corey H Basch[1], MPH, EdD; Steven Mey[1], BS; Diana L Malkin-Washeim[1], PhD; Christian Schuetz[1], MD, PhD; Niyonsenga Simon Pierre[5], MSc; Etienne Uwingabire[3], BSN; Alvera Mukamazimpaka[3], BSc; Paul Mbonyi[3], MPH; Sandhya Narayanan[1], EdD

[1]Diabetes Empowerment International, New York, NY, United States

[2]Department of Information Engineering, University of Padova, Padova, Italy

[3]Rwanda Diabetes Association, Kigali, Rwanda

[4]School of Health Sciences, University of Rwanda, Kigali, Rwanda

[5]Rwanda Biomedical Center, Kigali, Rwanda

Corresponding Author:
Sandhya Narayanan, EdD
Diabetes Empowerment International, New York, NY, United States

Abstract

Background: The development of minimally invasive continuous glucose monitoring systems (CGMs) has transformed diabetes management. CGMs have shown clinical significance by improving time in the euglycemic range, decreasing rates of hypoglycemia, and improving hemoglobin A_{1c} (HbA_{1c}). In Rwanda, CGMs are currently not routinely used, and no clinical studies of CGM use were identified in the literature.

Objective: This study aims to determine the impact and feasibility of real-time CGM use among people living with type 1 diabetes (T1D) in Rwanda through assessment of sensor use, time in range, rates of hypoglycemia and hyperglycemia, HbA_{1c}, and rates of diabetes-related hospitalizations over time.

Methods: The Continuous Glucose Monitoring Among Patients with Type 1 Diabetes in Rwanda (CAPT1D) study is a single-arm, prospective observational study conducted at the Rwandan Diabetes Association clinic in Kigali, Rwanda, aiming to assess the impact and feasibility of CGM use in Rwanda. A cohort of 50 participants diagnosed with T1D were enrolled. Participants were at least 21 years old, undergoing multiple daily insulin therapy, and not currently pregnant. Phase I of the study was conducted over 12 months, using the Dexcom G6 CGM. Phase II and Phase III extended CGM use for an additional 6 months respectively, using the next-generation Dexcom G7 CGM. Here, we report the quantitative results of the Phase I study.

Results: Participants used the sensor for >80% of the time throughout the study period. A significant increase in time in range was observed within 3 months and sustained over 12 months. HbA_{1c} decreased significantly in 3 months and stayed lower throughout the 12-month period. Mean HbA_{1c} levels decreased by 2.8% at 6 months (P<.001) and 3.2% at 12 months (P<.001). A total of 12 diabetes-related hospitalizations were reported during the study period. No cases of diabetic ketoacidosis or episodes of severe hypoglycemia occurred.

Conclusions: Significant and meaningful improvements in key glycemic indices indicate the potential feasibility and impact of a CGM among people living with T1D in Rwanda. Future studies could be designed to include pre- and postintervention analysis to determine the effectiveness in terms of complications and costs.

(JMIR Form Res 2025;9:e64585) doi:10.2196/64585

KEYWORDS

CGM; feasibility; type 1 diabetes; self-management; complications; Rwanda; time in range; hypoglycemia; continuous glucose monitoring; single-arm; feasibility study; diabetes; diabetes mellitus; euglycemic; HbA1c; observational study; quantitative; adult; health; public health; health informatics

XSL•FO
RenderX

Introduction

The International Diabetes Federation reports that there are 537 million adults aged 20‐79 years living with diabetes worldwide, with estimates that this number will increase to 700 million by 2045 [1]. There are more than 19 million adults with diabetes living in the African region [1], and in the International Diabetes Federation's latest report, Rwanda had a prevalence of diabetes registered at 2.7% (168,900/6,358,500) [2]. For those people with diabetes classified as having type 1 diabetes (T1D), the only treatment is through the administration of insulin exogenously. In addition, people with type 2 diabetes may need intensive insulin therapy with disease progression or poor management with noninsulin therapies. In both cases, glucose monitoring is necessary [3] to prevent or treat acute (hypoglycemia, hyperglycemia, and diabetic ketoacidosis [DKA]) or chronic (microvacular and macrovascular) complications, and more frequent monitoring has been shown to have better diabetes outcomes [4].

Traditional methods of glucose monitoring among people with diabetes have been through finger-stick blood sugar checks, referred to as self-monitoring blood glucose (SMBG), involving pricking one's finger with a small needle at different times during the day. The blood drawn is then transferred to a test strip which is read by a blood glucose monitor to display the glucose value at that point in time. There may be several obstacles related to SMBG—an invasive method that can be accompanied by discomfort and inconvenience [5,6] and relies on glucose meters that vary in accuracy [3]. Furthermore, SMBG provides very limited insight into glycemic trends, as only a few data points per day are available for decision-making [6].

In the absence of regular information regarding glucose levels, the primary way to understand how well diabetes is managed is through a blood test known as the hemoglobin A_{1c} (HbA_{1c}) test, which measures the percentage of glycated hemoglobin in red blood cells. This is an average over a 3‐ to 4-month period and provides a retrospective evaluation of therapeutic and self-management efficacy [7]. As such, it does not provide meaningful information regarding frequent or sustained high or low blood glucose values. Therefore, it is not the best tool to assist people with diabetes with treatment interventions for daily fluctuations in blood glucose [8].

The development of minimally invasive continuous glucose monitoring systems (CGMs) has transformed diabetes management [9]. CGMs provide continuous readings throughout the day and provide meaningful insight to people with diabetes regarding the time in range (TIR), as well as alerts when the glucose values veer from the range into higher (time above range [TAR]) or lower (time below range [TBR]) levels. These alerts then allow for immediate treatment action, thereby reducing the time spent in the times outside of the euglycemic range [8].

CGMs have shown clinical significance by improving time in the euglycemic range, decreasing rates of hypoglycemia and HbA_{1c} levels [10], and lowering rates of emergency department visits and hospitalizations for hypoglycemia [11]. Improved outcomes have also been shown when a CGM is introduced in the first year after diagnosis with the improvements sustained over time, for as long as 7 years [12].

In Rwanda, CGMs are currently not routinely used, and no clinical studies of CGM use were identified in the literature. The study objectives were to determine the impact and feasibility of real-time CGM use among people living with T1D in Rwanda through assessment of sensor use; TIR; HbA_{1c}; and rates of hypoglycemia, diabetes-related hospitalizations (DRHs), and DKA.

Methods

Study Design

The Continuous Glucose Monitoring Among Patients with Type 1 Diabetes in Rwanda (CAPT1D) study is a single-arm, prospective observational study conducted at the Rwandan Diabetes Association (RDA) clinic in Kigali, Rwanda, aiming to assess the impact and feasibility of CGM use in Rwanda.

The study protocol involved a run-in phase with blinded CGM (Dexcom G6, Dexcom Inc), during which participants received instructions in August 2022. Subsequently, continuous glucose monitoring was unblinded, and participants underwent a 12-month period during which they visited the RDA clinic every 20 days to upload CGM data onto the Dexcom Clarity (Dexcom Inc) platform. Participants were queried for and self-reported episodes of DKA (date, symptoms, treatment), instances of hypoglycemia, and DRHs at each study visit.

Study Sample and Recruitment

A convenience sample of established patients of the RDA was selected using a 2-stage process. First, the RDA study staff created a list of patients who met the inclusion criteria, as follows: older than 21 years, not pregnant, and treated with multiple daily doses of insulin. Second, this list was polled to identify those participants who were willing and able to commute to the RDA clinic as frequently as needed for orientation, to change sensors, and to upload data. Individuals were informed prior to enrollment that they would not be paid to participate in the study, but that transportation costs to and from the clinic will be reimbursed. The first 50 participants who agreed to participate were then enrolled in the study. The sample size was determined by the capacity of the staff and the availability of equipment. Phase I of the study was conducted over 12 months, using the Dexcom G6 CGM. Phase II and Phase III extended CGM use for an additional 6 months respectively, using the next-generation Dexcom G7 CGM. The number of participants was reduced to 47 when 3 participants withdrew from the study at month 6. Thereafter, the study sample was further reduced to 42 as a result of 5 participants joining another study where they received additional education possibly biasing the results of CAPT1D. To handle the fact that several patients were removed from month 6 on, statistical analysis was adjusted with the last observation carried forward method, that is, the last available measurement for a patient is carried forward to later time points. Here we report the quantitative results of the Phase I study.

XSL•FO
RenderX

Ethical Considerations

Ethical approval was obtained from the University of Rwanda School of Medicine Ethical Committee, and the study was registered with the institutional review board under number #271/CMHSIRB/2022. Participants were provided with all details regarding the study's objectives, procedures, potential risks, and benefits before requesting consent. At this time, any questions were addressed, and privacy and confidentiality policies were explained. A consent form outlining the study's scope, potential risks, the voluntary nature of participation, and the right to withdraw at any time was provided and signed by each participant. Additionally, consent to use photographs and other recorded material in published material was obtained. To this end, all data collected for this study have been fully anonymized and deidentified to ensure the privacy and confidentiality of all participants, and to prevent linkage of the data to individual participants. Participants were compensated varied amounts for their transportation to and from the clinic, provided light snacks and beverages during clinic visits, and received the CGMs for no cost. There is no identification of individual participants/users in any aspect of this manuscript.

Statistical Methods

CGM data were preprocessed and analyzed using AGATA [13], an open-source software tool that computes glucose control metrics in accordance with the definitions of the International Consensus on CGM-derived metrics for clinical trials [14]. Initially, CGM data were preprocessed by aligning each data point on a uniform 5-minute time grid and linearly interpolating data gaps up to 30 minutes.

Subsequently, AGATA was used to extract a set of well-established metrics [14] quantifying glucose control, including 5 metrics for glucose variability (mean, SD, coefficient of variation, glucose management indicator, and glucose risk index); 6 metrics describing the percentage of time spent in specific glucose ranges (level 1 and level 2 TBR [TBR_{L1} and TBR_{L2}], defined as <70 mg/dL and <55 mg/dL respectively; level 1 and level 2 TAR [TAR_{L1} and TAR_{L2}], defined as >180 mg/dL and >250 mg/dL, respectively; TIR, defined as 70-180 mg/dL; and time in tight range [TITR], defined as 70-140 mg/dL); and a metric quantifying CGM use, namely the percentage of time the CGM was active in the considered period ($\%CGM_{USE}$).

To align CGM data to HbA_{1c} samples collected every 3 months, each metric has been calculated by aggregating CGM data over 3-month periods and compared with the baseline computed during the initial run-in session at month 0.

Hereafter, for the sake of brevity, we will label M0 the baseline run-in session, while we will use M3, M6, M9, and M12 to represent the metric values computed using data from M0 to month 3, month 3 to month 6, and so forth.

Results were presented as the mean (SD) if the metric distribution followed a Gaussian distribution (determined by the Lillefors test at a significance level of 5%), or as the median (IQR) if not. Changes were evaluated by comparing M0 versus M6, M6 versus M12, and M0 versus M12 using a 2-tailed *t* test with a significance level of 5% for Gaussian-distributed metrics. Alternatively, the Wilcoxon rank sum test with a significance level of 5% was used.

Results

The results are illustrated in Table 1. Participants used the sensor for >80% of the time throughout the study period. TIR was defined as 70 - 180 mg/dL. The absolute change from baseline to 6 months for TIR was an increase of 9.9% (P<.001), and from baseline to 12 months, the absolute change was an increase of 14.7% (P<.001). Thus, the significant improvement in TIR at month 6 was maintained (indeed increased) at 12 months. TITR was defined as 70 - 140 mg/dL. The absolute change from baseline to 6 months for TITR was an increase of 7.1%, and that from baseline to 12 months was an increase of 10.4% (P<.001) (Figure 1). Time in hypoglycemia or TBR was defined as <70 mg/dL. From baseline to 6 months, this increased by 0.7%, which was not statistically significant (P=.31), and from baseline to 12 months, there was an increase of 2.5% (P<.001). The TAR is defined as >180 mg/dL. The absolute change from baseline to 6 months for TAR was a decrease of 9.7% (P<.01), and that from baseline to 12 months was a decrease of 15.3% (P<.001). Mean HbA_{1c} levels decreased by 2.8% at 6 months (P<.001) and 3.2% at 12 months (P<.001) (Figure 2). A total of 12 DRHs were reported during the study period. More detailed analysis of these hospitalizations is underway. No cases of DKA or episodes of severe hypoglycemia occurred.

XSL•FO
RenderX

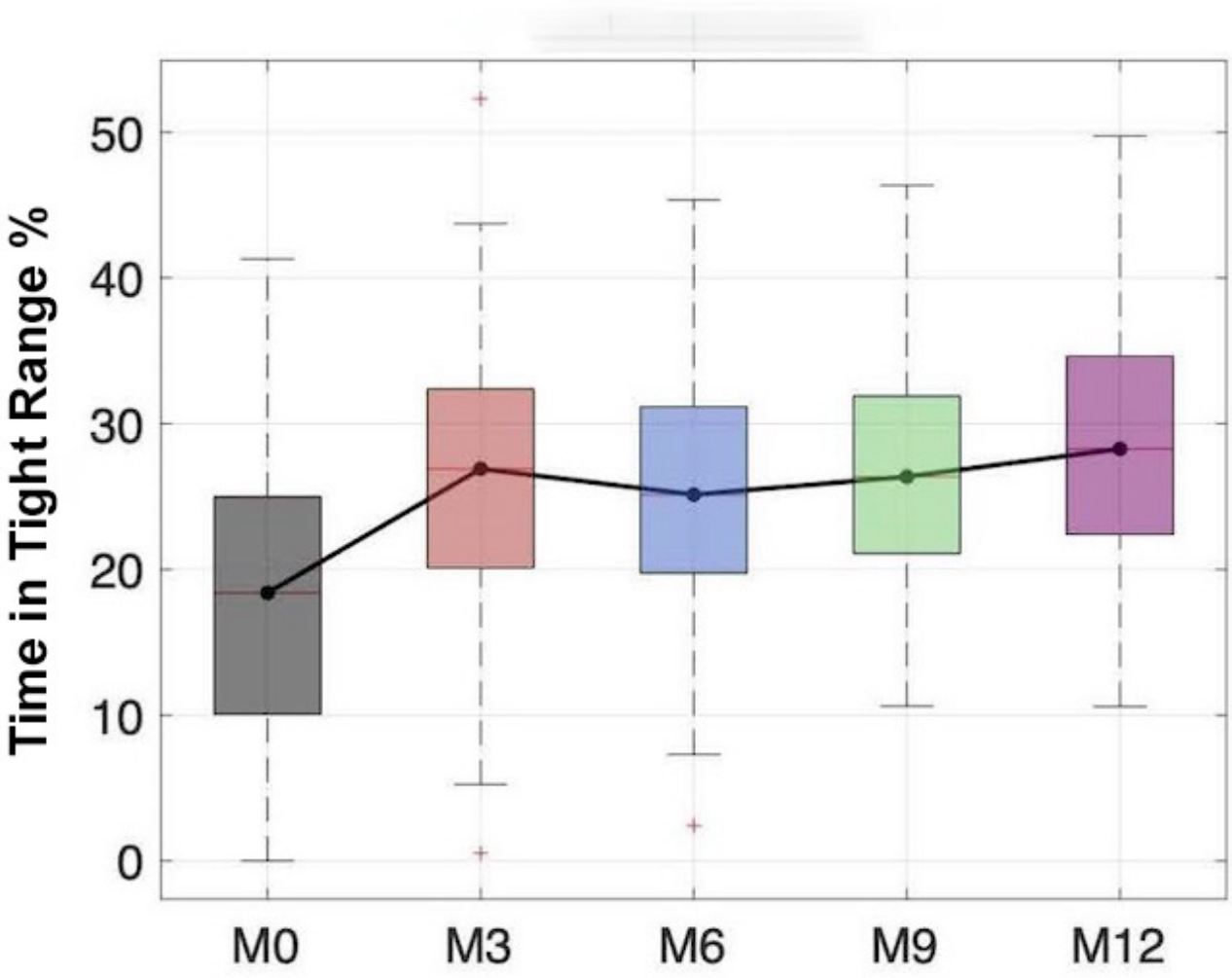
Table . Glycemic indicators at baseline and at 3, 6, 9, and 12 months from the Continuous Glucose Monitoring Among Patients with Type 1 Diabetes in Rwanda study, a single-arm, prospective observational study conducted at the Rwandan Diabetes Association clinic in Kigali, Rwanda (2022 - 2023).

Metrics	Baseline (N=46)		M3 (N=47)		M6 (N=47)		M9 (N=42)		M12 (N=42)		*P* value (baseline vs M12)	*P* value (baseline vs M6)	*P* value (M6 vs M12)
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)			
Percentage of sensor data	—[a]	99.8941 (98.7353 - 100)	—	84.4136 (77.8819 - 91.8981)	—	83.8927 (73.1684 - 93.2301)	—	86.5239 (81.3619 - 95.2238)	—	88.0305 (73.5031 - 95.7562)	<.001	<.001	.49
Time in range (%)	30.9998 (15.0695)	—	42.4303 (14.683)	—	40.8986 (12.8338)	—	44.0248 (12.4312)	—	45.6591 (12.0551)	—	<.001	<.001	.08
Time in tight range (%)	18.098 (9.5095)	—	25.8216 (10.056)	—	25.2415 (8.7807)	—	27.1875 (8.7808)	—	28.4829 (8.6731)	—	<.001	<.001	.08
Time below range (%)	—	2.369 (0.1768 - 6.4643)	—	2.146 (1.2232 - 6.1304)	—	3.0203 (1.054 - 6.4066)	—	2.9182 (1.2722 - 7.7068)	—	4.8789 (1.602 - 8.9319)	<.001	.31	.17
Time above range (%)	64.1901 (17.409)	—	53.5695 (16.3889)	—	54.5215 (14.4949)	—	50.9759 (13.8959)	—	48.8406 (13.6259)	—	<.001	<.001	.06
Mean glucose (mg/dL)	238.9856 (53.6842)	—	—	193.5024 (181.2758 - 232.7176)	209.5659 (40.2107)	—	200.1775 (34.7531)	—	194.077 (32.4736)	—	<.001	<.001	.05
Glucose management indicator	9.0265 (1.2841)	—	—	7.9386 (7.6461 - 8.8766)	8.3228 (0.96184)		8.0982 (0.83129)	—	7.9523 (0.77677)	—	<.001	<.001	.05
Glucose risk index	—	99.5992 (76.2825 - 100)	75.5983 (19.3713)	—	78.7051 (17.9064)	—	—	81.6141 (57.678 - 90.9717)	73.6757 (19.2475)	—	<.001	<.001	.21
Coefficient of variation	41.107 (10.4761)	—	42.7715 (7.7309)	—	43.8174 (7.1405)	—	44.2991 (6.5068)	—	44.6247 (6.6217)	—	.07	.15	.58
SD (mg/dL)	94.2823 (19.0105)	—	87.6943 (17.0699)	—	90.4366 (15.6548)	—	88.2446 (17.5455)	—	86.2282 (17.1804)	—	<.001	.29	.23
Hypoglycemic events per week (n)	—	4.4336 (0.70961 - 9.1829)	—	4.6668 (1.9445 - 7.2531)	—	4.978 (2.3529 - 7.7781)	—	4.9002 (2.6445 - 9.1781)	6.4021 (4.3053)	—	.14	.50	.38
Hyperglycemic events per week (n)	16.2707 (4.8316)	—	16.5623 (4.4013)	—	16.5226 (4.8881)	—	17.6488 (4.882)	—	16.7118 (5.7817)	—	.70	.80	.87

[a]Not applicable.

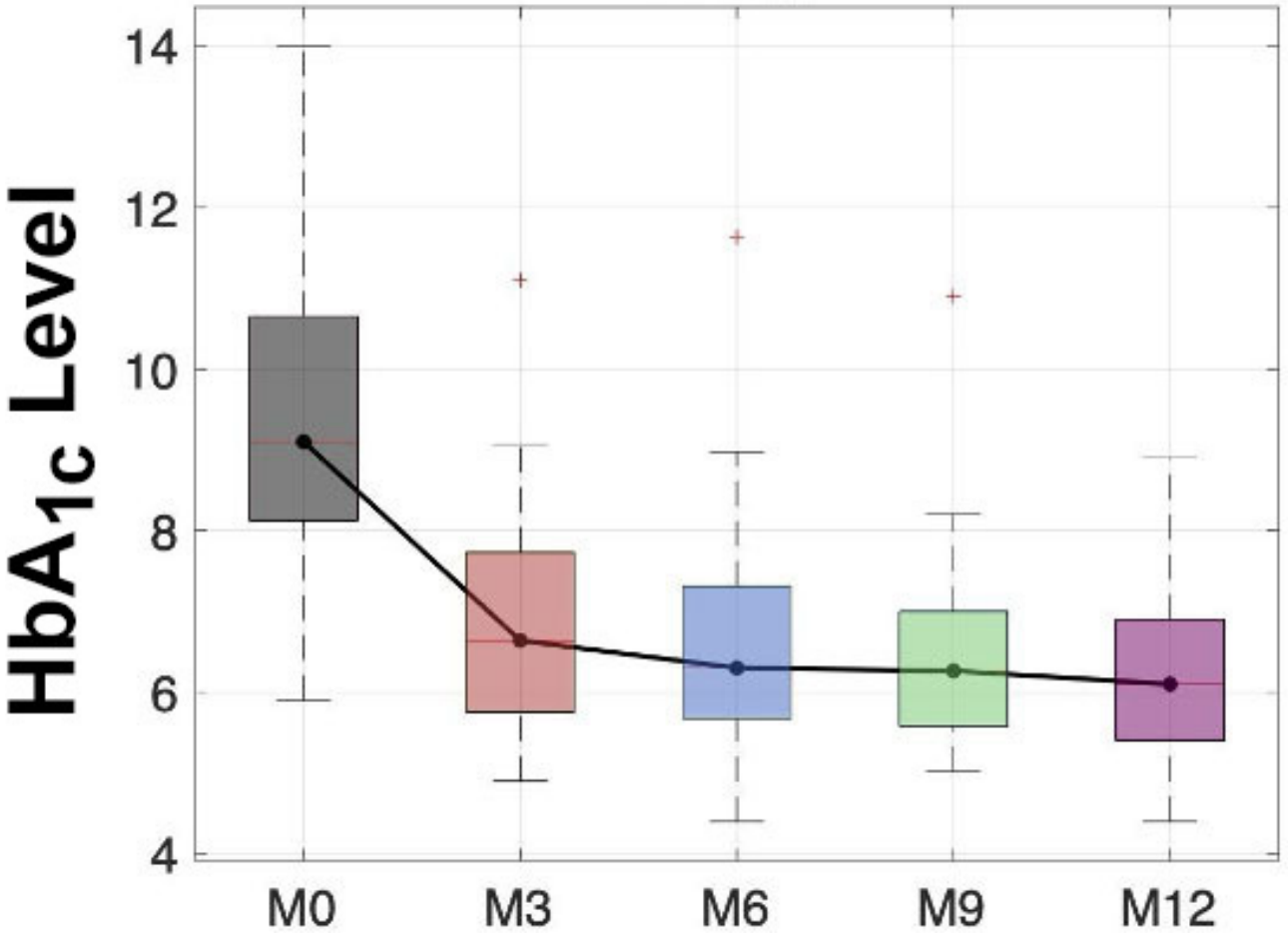
XSL•FO
RenderX

Figure 1. Box plot distributions of time in tight range were obtained at M0 (gray), M3 (red), M6 (light blue), M9 (green), and M12 (violet) from the Continuous Glucose Monitoring Among Patients with Type 1 Diabetes in Rwanda study, a single-arm, prospective observational study conducted at the Rwandan Diabetes Association clinic in Kigali, Rwanda (2022 - 2023). Median values are highlighted and connected by a black solid line.



XSL•FO
RenderX

Figure 2. Box plot distributions of HbA_{1c} were obtained at M0 (gray), M3 (red), M6 (light blue), M9 (green), and M12 (violet) from the Continuous Glucose Monitoring Among Patients with Type 1 Diabetes in Rwanda study, a single-arm, prospective observational study conducted at the Rwandan Diabetes Association clinic in Kigali, Rwanda (2022 - 2023). Median values are highlighted and connected by a black solid line. HbA_{1c}: hemoglobin A_{1c}.



Discussion

The findings from Phase I of the CGM use among individuals with T1D in Rwanda provide insight into the impact and feasibility of real-time CGM use. As noted previously, the key objectives of this study included evaluating sensor use; TIR; HbA_{1c} levels; and rates of hypoglycemia, DRH, and DKA. Study participants demonstrated high levels of sensor use (80% of the study period). As noted in the results, TIR improved with increased levels at 6 and 12 months. HbA_{1c} levels also improved with marked decreases at 6 and 12 months. Of note, there was a mild, nonstatistically significant increase in TBR, which is indicative of hypoglycemia, but there was also a statistically significant reduction in TAR at 6 and 12 months. There was no report of severe hypoglycemia or DKA events, and only 12 DRHs were reported during the study. In sum, these data indicate a low level of adverse effects.

The low attrition rate, the high percentage of sensor use, and the willingness of participants to return to the clinic every 20 days for data uploads indicated the feasibility of continuous glucose monitoring in Rwanda among patients with T1D. The findings from this study augment those from other studies conducted in African countries showing that continuous glucose monitoring is a feasible alternative to SMBG [15,16].

In a study aiming to assess the feasibility of CGM use in Malawi, one of the challenges identified was the inability of participants to change their own sensors [16]. However, in this study, participants were given sufficient training so that they could change their sensors on their own. Additionally, participants visited the clinic frequently per study protocol, providing ample opportunity for study staff to help provide guidance on sensor removal and insertion.

A systematic review of 22 randomized controlled trials comparing continuous glucose monitoring and SMBG found a significant reduction in HbA_{1c} with the CGM group [17]. Although this study was not interventional, the observed results showed a similar significant reduction in HbA_{1c} with the use of a CGM.

Further, the findings from this study are aligned with other studies showing a reduction in HbA_{1c} with the use of a CGM [10,18] and consistent with prior studies showing that the use of a CGM increased TIR, a desirable outcome for people with diabetes[18].

XSL•FO
RenderX

There were several limitations of this study that limited validity and generalizability. The single-arm observational design lacks a control group, which is consistent with a feasibility study but nonetheless limits between-group comparisons, the ability to isolate the effect of CGM use, the ability to conduct complete comparison analyses, and the ability to generalize findings.

While we were able to glean insight from this convenience sample, the sample was small at the outset, and the further reduction due to attrition impacts statistical power and effect size. Our use of the last observation carried forward method for missing data could potentially lead to bias.

As with any study that uses self-reported information, bias can be introduced and accuracy can be compromised. Notwithstanding these limitations, the length of this study, conducted over a 1-year period, provided a meaningful amount of data points for analysis. These preliminary results can serve as a foundation for future studies that can include control groups and interventions. This work is particularly relevant given the dearth of studies to date related to CGM use in people with T1D in Rwanda.

The results suggest that implementing CGM technology in T1D patients in Rwanda is a feasible and promising method for improving glycemic outcomes among people with T1D. Within 3 months, results indicate improvements in glycemic management, with the improvements sustained over the 1-year study period, as evidenced by better TIR, TITR, and TAR and lower HbA_{1c} levels. These outcomes indicate potential for better diabetes self-management, thereby impacting rates of diabetes-related complications and their associated personal and social costs. Results from this study can be used to conduct more rigorous interventional studies to confirm the beneficial impact of a CGM on health outcomes for people with diabetes.

Data Availability

Data are available upon reasonable request to the corresponding author.

Conflicts of Interest

SM and SN are employees of Dexcom and own company shares. JB is a speaker for Dexcom and a consultant for Sanofi (as a speaker) and Insulet (on the advisory board).

References

1. IDF Diabetes Atlas. International Diabetes Federation. URL: https://diabetesatlas.org/data/en/ [accessed 2024-07-13]
2. IDF Members: Rwanda. International Diabetes Federation. URL: https://idf.org/our-network/regions-and-members/africa/members/rwanda/ [accessed 2024-07-13]
3. American Diabetes Association Professional Practice Committee. 7. Diabetes technology: standards of care in diabetes-2024. Diabetes Care 2024 Jan 1;47(Suppl 1):S126-S144. [doi: 10.2337/dc24-S007] [Medline: 38078575]
4. Miller KM, Beck RW, Bergenstal RM, et al. Evidence of a strong association between frequency of self-monitoring of blood glucose and hemoglobin A1c levels in T1D exchange clinic registry participants. Diabetes Care 2013 Jul;36(7):2009-2014. [doi: 10.2337/dc12-1770] [Medline: 23378621]
5. Patton SR. Adherence to glycemic monitoring in diabetes. J Diabetes Sci Technol 2015 May;9(3):668-675. [doi: 10.1177/1932296814567709] [Medline: 25591853]
6. Didyuk O, Econom N, Guardia A, Livingston K, Klueh U. Continuous glucose monitoring devices: past, present, and future focus on the history and evolution of technological innovation. J Diabetes Sci Technol 2021 May;15(3):676-683. [doi: 10.1177/1932296819899394] [Medline: 31931614]
7. Cappon G, Vettoretti M, Sparacino G, Facchinetti A. Continuous glucose monitoring sensors for diabetes management: a review of technologies and applications. Diabetes Metab J 2019 Aug;43(4):383-397. [doi: 10.4093/dmj.2019.0121] [Medline: 31441246]
8. Dovc K, Battelino T. Time in range centered diabetes care. Clin Pediatr Endocrinol 2021;30(1):1-10. [doi: 10.1297/cpe.30.1] [Medline: 33446946]
9. Rodbard D. Continuous glucose monitoring: a review of recent studies demonstrating improved glycemic outcomes. Diabetes Technol Ther 2017 Jun;19(S3):S25-S37. [doi: 10.1089/dia.2017.0035] [Medline: 28585879]
10. Nemlekar PM, Hannah KL, Green CR, Norman GJ. Association between adherence, A1C improvement, and type of continuous glucose monitoring system in people with type 1 diabetes or type 2 diabetes treated with intensive insulin therapy. Diabetes Ther 2024 Mar;15(3):639-648. [doi: 10.1007/s13300-023-01529-8] [Medline: 38289464]
11. Karter AJ, Parker MM, Moffet HH, Gilliam LK, Dlott R. Association of real-time continuous glucose monitoring with glycemic control and acute metabolic events among patients with insulin-treated diabetes. JAMA 2021 Jun 8;325(22):2273-2284. [doi: 10.1001/jama.2021.6530] [Medline: 34077502]
12. Champakanath A, Akturk HK, Alonso GT, Snell-Bergeon JK, Shah VN. Continuous glucose monitoring initiation within first year of type 1 diabetes diagnosis is associated with improved glycemic outcomes: 7-year follow-up study. Diabetes Care 2022 Mar 1;45(3):750-753. [doi: 10.2337/dc21-2004] [Medline: 35018417]
13. Cappon G, Sparacino G, Facchinetti A. AGATA: a toolbox for automated glucose data analysis. J Diabetes Sci Technol 2024 Sep;18(5):1109-1121. [doi: 10.1177/19322968221147570] [Medline: 36602030]

XSL•FO
RenderX

14. Battelino T, Alexander CM, Amiel SA, et al. Continuous glucose monitoring and metrics for clinical trials: an international consensus statement. Lancet Diabetes Endocrinol 2023;11(1):42-57. [doi: 10.1016/S2213-8587(22)00319-9]
15. McClure Yauch L, Velazquez E, Piloya-Were T, Wainaina Mungai L, Omar A, Moran A. Continuous glucose monitoring assessment of metabolic control in east African children and young adults with type 1 diabetes: a pilot and feasibility study. Endocrinol Diabetes Metab 2020 Jul;3(3):e00135. [doi: 10.1002/edm2.135] [Medline: 32704558]
16. Gomber A, Valeta F, Coates MM, et al. Feasibility of continuous glucose monitoring in patients with type 1 diabetes at two district hospitals in Neno, Malawi: a randomised controlled trial. BMJ Open 2024 May 6;14(5):e075554. [doi: 10.1136/bmjopen-2023-075554] [Medline: 38719319]
17. Teo E, Hassan N, Tam W, Koh S. Effectiveness of continuous glucose monitoring in maintaining glycaemic control among people with type 1 diabetes mellitus: a systematic review of randomised controlled trials and meta-analysis. Diabetologia 2022 Apr;65(4):604-619. [doi: 10.1007/s00125-021-05648-4] [Medline: 35141761]
18. Beck RW, Riddlesworth T, Ruedy K, et al. Effect of continuous glucose monitoring on glycemic control in adults with type 1 diabetes using insulin injections: the DIAMOND randomized clinical trial. JAMA 2017 Jan 24;317(4):371-378. [doi: 10.1001/jama.2016.19975] [Medline: 28118453]

Abbreviations

CAPT1D: Continuous Glucose Monitoring Among Patients with Type 1 Diabetes in Rwanda
CGM: continuous glucose monitoring system
DKA: diabetic ketoacidosis
DRH: diabetes-related hospitalization
HbA_{1c}: hemoglobin A_{1c}
RDA: Rwandan Diabetes Association
SMBG: self-monitoring blood glucose
T1D: type 1 diabetes
TAR: time above range
TBR: time below range
TIR: time in range
TITR: time in tight range

Edited by A Mavragani; submitted 21.07.24; peer-reviewed by I Milani, JA Aloi; revised version received 21.11.24; accepted 25.11.24; published 21.01.25.

Please cite as:
Baker J, Cappon G, Habineza JC, Basch CH, Mey S, Malkin-Washeim DL, Schuetz C, Simon Pierre N, Uwingabire E, Mukamazimpaka A, Mbonyi P, Narayanan S
Continuous Glucose Monitoring Among Patients With Type 1 Diabetes in Rwanda (CAPT1D) Phase I: Prospective Observational Feasibility Study
JMIR Form Res 2025;9:e64585
URL: https://formative.jmir.org/2025/1/e64585
doi:10.2196/64585

© Jason Baker, Giacomo Cappon, Jean Claude Habineza, Corey H Basch, Steven Mey, Diana L Malkin-Washeim, Christian Schuetz, Niyonsenga Simon Pierre, Etienne Uwingabire, Alvera Mukamazimpaka, Paul Mbonyi, Sandhya Narayanan. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Developing Assessments for Key Stakeholders in Pediatric Congenital Heart Disease: Qualitative Pilot Study to Inform Designing of a Medical Education Toy

Neda Barbazi[1], MArch; Ji Youn Shin[1], PhD; Gurumurthy Hiremath[2], MD; Carlye Anne Lauff[1], PhD

[1]Department of Design Innovation, College of Design, University of Minnesota, Twin Cities, Minneapolis, MN, United States

[2]Division of Cardiology, Department of Pediatrics, Medical School, University of Minnesota, Twin Cities, Minneapolis, MN, United States

Corresponding Author:
Carlye Anne Lauff, PhD
Department of Design Innovation, College of Design
University of Minnesota, Twin Cities
13 Rapson Hall, 89 Church Street SE
Minneapolis, MN, 55455-0213
United States
Phone: 1 724 889 5511
Fax: 1 612 625 7525
Email: carlye@umn.edu

Abstract

Background: Congenital heart disease (CHD) is a birth defect of the heart that requires long-term care and often leads to additional health complications. Effective educational strategies are essential for improving health literacy and care outcomes. Despite affecting around 40,000 children annually in the United States, there is a gap in understanding children's health literacy, parental educational burdens, and the efficiency of health care providers in delivering education.

Objective: This qualitative pilot study aims to develop tailored assessment tools to evaluate educational needs and burdens among children with CHD, their parents, and health care providers. These assessments will inform the design of medical education toys to enhance health management and outcomes for pediatric patients with CHD and key stakeholders.

Methods: Through stakeholder feedback from pediatric patients with CHD, parents, and health care providers, we developed three tailored assessments in two phases: (1) iterative development of the assessment tools and (2) pilot testing. In the first phase, we defined key concepts, conducted a literature review, and created initial drafts of the assessments. During the pilot-testing phase, 12 participants were recruited at the M Health Fairview Pediatric Specialty Clinic for Cardiology—Explorer in Minneapolis, Minnesota, United States. We gathered feedback using qualitative methods, including cognitive interviews such as think-aloud techniques, verbal probing, and observations of nonverbal cues. The data were analyzed to identify the strengths and weaknesses of each assessment item and areas for improvement.

Results: The 12 participants included children with CHD (n=5), parents (n=4), and health care providers (n=3). The results showed the feasibility and effectiveness of the tailored assessments. Participants showed high levels of engagement and found the assessment items relevant to their education needs. Iterative revisions based on participant feedback improved the assessments' clarity, relevance, and engagement for all stakeholders, including children with CHD.

Conclusions: This pilot study emphasizes the importance of iterative assessment development, focusing on multistakeholder engagement. The insights gained from the development process will guide the creation of tailored assessments and inform the development of child-led educational interventions for pediatric populations with CHD.

(JMIR Form Res 2025;9:e63818) doi:10.2196/63818

KEYWORDS

assessment; congenital heart disease; children health literacy; health education; caregiving; patient-centered care; design; qualitative pilot; children health; educational interventions

XSL•FO
RenderX

Introduction

Background

Congenital heart disease (CHD) is a heart abnormality present at birth, requiring intensive medical care and often leading to life-threatening complications [1]. It affects around 40,000 newborns annually in the United States and 1% globally, posing lifelong challenges for children, families, and health care systems [1-4]. Despite significant advancements in diagnosis and treatment, CHD remains a chronic condition that demands ongoing care and specialized educational resources [5,6]. Tailored education is critical for addressing the needs of children with CHD, providing accessible educational support for parents, and equipping health care providers with efficient educational strategies. However, no tools currently exist to understand children's health literacy, assess parental educational burdens, or measure health care providers' efficiency in delivering education [7,8].

Children with CHD face unique barriers in understanding their condition due to developmental challenges, such as difficulty grasping abstract medical concepts or relating them to their lived experiences. These limitations can hinder their ability to adhere to treatment plans, actively participate in care, or respond effectively during emergencies and transitions to adult care [4,6,9-11]. Despite these challenges, pediatric patients are often excluded from health literacy efforts, as most resources are designed for parents or caregivers. This exclusion highlights the critical need for age-appropriate tools that empower children to engage with their care actively, ultimately improving adherence and long-term health outcomes [7,12-16].

Parents, meanwhile, bear a significant educational burden. They must interpret medical jargon, simplify it for their child, and act as intermediaries with health care providers, all while managing the emotional strain of caregiving and the cognitive load of understanding complex medical information. This burden can increase stress, reduce caregiving effectiveness, and impact family well-being. Low parental health literacy further compounds these challenges, as it is linked to medical errors and poorer health outcomes for children [17-24]. Furthermore, health care providers face the challenge of balancing clear communication with time and resource constraints, often struggling to deliver education tailored to the needs of both children and parents [11,25-27]. Addressing these gaps is essential for improving communication, care coordination, and health outcomes for children with CHD and their families [7,28].

Study Objectives

As part of a multiphase research project, this pilot study aims to close these gaps by developing tailored assessment tools for key stakeholders in pediatric CHD education. This research comprises two phases: (1) iterative development of the assessment tools and (2) pilot testing to create, refine, and validate 3 CHD-specific assessments based on stakeholder feedback in a real-world setting. We gathered feedback using cognitive interviews and observations to assess each assessment item [29-34]. We iteratively developed face-to-face and computer-based assessments based on this feedback from all stakeholders, including children. The assessments measure changes in children's knowledge, parental educational burdens, and health care provider efficiency before and post interventions. The results show that developing and refining these assessments are essential before introducing our medical education toy. The objectives of this study are 3-fold:

1. To assess changes in knowledge of children with CHD before and after educational interventions to empower pediatric engagement in health care innovation.
2. To measure the educational burden experienced by parents before and after interventions, focusing on their needs and challenges in navigating CHD-related educational responsibilities.
3. To assess the efficiency and effectiveness of health care providers in communicating essential information to children with CHD and their families to improve care coordination and patient outcomes.

Prior Work

Assessments Developed for Pediatric Patients With CHD, Parents, and Health Care Providers

Assessments are essential for evaluating the health status and practices of pediatric patients, their parents, and health care providers across various domains [35-37]. These domains include health-related quality of life, emotional well-being, physical or psychological health, social support, and behavioral problems. For pediatric patients aged 8-18 years, assessments often involve parent proxies completing tools for younger children [38-43]. Commonly used tools in this context include the Pediatric Quality of Life Inventory (PedsQL), Pediatric Cardiac Quality of Life Inventory (PCQLI) [44-47], Child Health Questionnaire (CHQ), and Child Behavior Checklist (CBCL). However, there is a notable gap in health literacy measures for children younger than 9 years [48]. While some tools, such as Food Label Literacy for Applied Nutrition Knowledge (FLLANK) questionnaire [49], exist, none are specifically tailored to disease populations such as those with CHD.

Parental well-being and caregiving burden assessments are also critical for understanding and supporting effective caregiving. However, the specific burden related to education remains underexplored. Existing tools, such as the Parenting Stress Index/Parental Stress Scale (PSI/PSS) and the Family Impact Module (FIM) of the PedsQL, focus on parental stress, caregiving difficulty, and family functioning [50-53]. Similarly, tools such as the Caregiver Health Self-Assessment, which aims to improve the caregiver-provider dyadic relationship (commonly for older adult patients), do not address the educational needs of caregivers of patients with CHD [54-56].

For health care providers, efficiency assessments—measuring the ability to minimize wasted time and maximize outcomes—remain an ongoing challenge despite progress in evaluating physician and hospital care effectiveness [57,58]. While tools such as Key Performance Indicators (KPIs), Lean Six Sigma (LSS), and Performance Metrics are widely used to evaluate provider and hospital performance, their primary focus is on patient care outcomes (eg, patient satisfaction). These methods often overlook the educational challenges faced by

XSL•FO
RenderX

health care providers themselves [59-63]. Addressing these gaps can enhance our understanding of the needs and well-being of CHD stakeholders, including children, enabling the development of more targeted and effective educational interventions. More details about these assessments are available in Multimedia Appendix 1.

Empowering Pediatric Engagement in Health Care Interventions

Engaging children effectively in health care interventions and assessments is challenging due to their unique developmental hurdles [64-67]. Younger children face difficulties participating because of limited attention spans and cognitive abilities, requiring reliance on parental feedback as a proxy for designing and testing interventions [68,69]. To address these challenges, studies suggest using interviews, focus groups, and activity-based methods [70-74].

Prior studies in design, human-computer interaction, and health care domains indicate that adolescents aged 13-17 years engage well in interviews and focus groups. In contrast, younger children, particularly those aged 4-12 years, benefit more from creative techniques such as activity-based methods [72-78]. Furthermore, studies highlight the importance of social factors, such as ongoing support, in alleviating children's potential stress during research activities [79]. Empowering children through age-appropriate strategies, coupled with family facilitation, enhances their sense of ownership in health care decision-making. These approaches not only improve engagement but also strengthen the research process by addressing children's developmental needs and promoting their active participation [79-83].

Methods

We developed and refined 3 CHD-specific assessments to evaluate children's knowledge, parental educational burdens, and health care provider efficiency in delivering education. This process included defining key concepts, conducting a literature review, and designing initial assessment tools. These assessments were subsequently pilot-tested with stakeholders at the M Health Fairview Pediatric Specialty Clinic for Cardiology—Explorer in Minneapolis, Minnesota, United States, to ensure validity and relevance.

Assessment Development

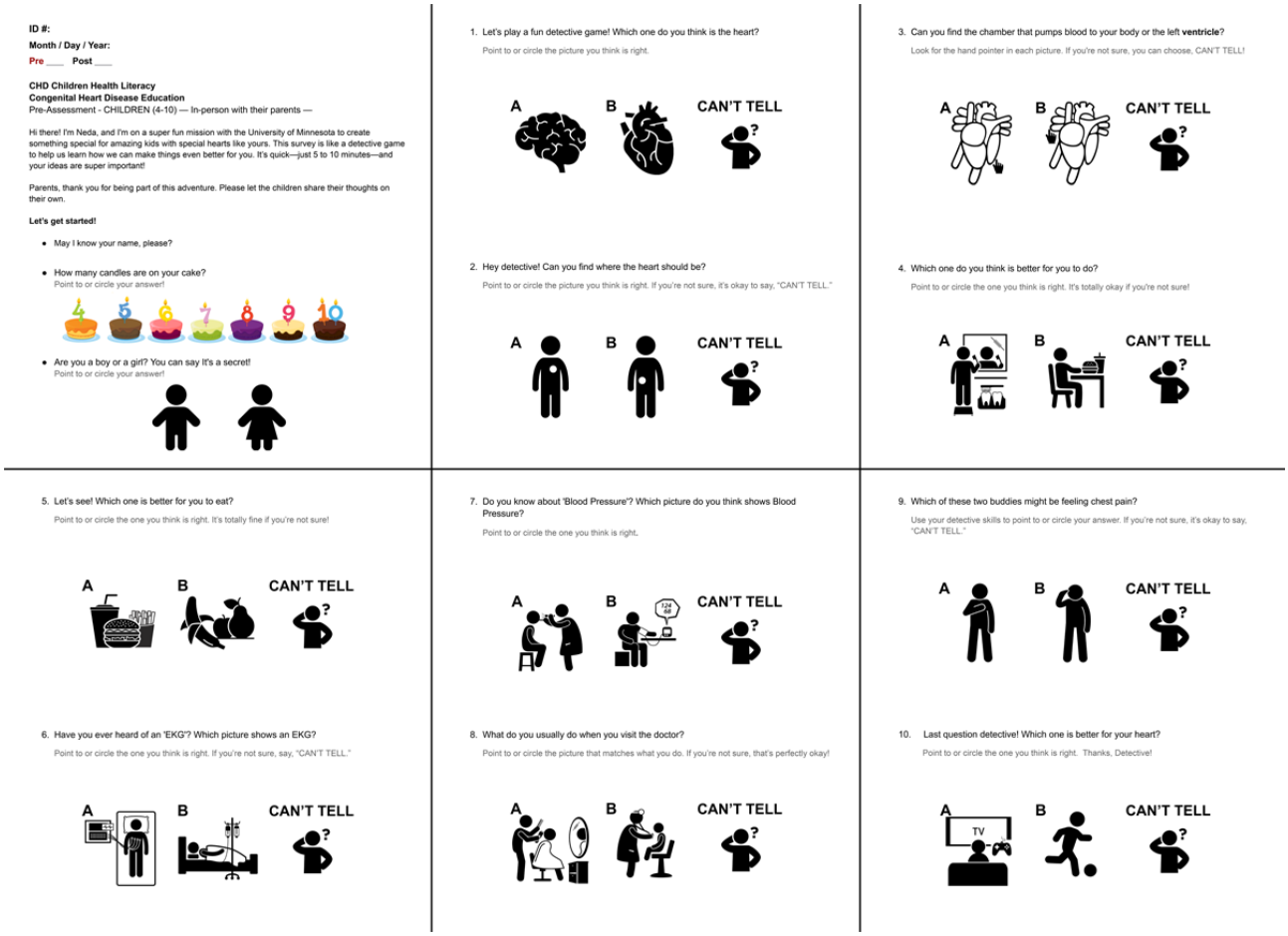
CHD Health Literacy Children Assessment

We developed the CHD Health Literacy Children Assessment (CHD-HLCA) to evaluate CHD health literacy in children aged 4-10 years through pre- and posteducational intervention. This tool draws inspiration from the FLLANK questionnaire [84] and incorporates storytelling techniques to enhance engagement [85]. The assessment consists of 10 questions featuring simple black and white icons without color or intricate details to reduce visual distractions and avoid potential psychological and physiological influences of color. Instead, children use colored markers to answer, fostering their engagement and enthusiasm during the assessment process [86,87].

This self-report assessment is administered with the assistance of the research team or parents, particularly for younger children. The assessment covers various knowledge dimensions, including understanding CHD, preparing for doctor visits, and practicing self-care [49]. Questions offer 2 comparison options to minimize complexity, using simple icons and illustrations to reduce distractions and prevent cognitive overload. A "can't tell" option is also included to ensure that children feel comfortable expressing uncertainty. Furthermore, a Likert scale question with Smiley Face Likerts assesses their general knowledge about their heart [88], a common method for children's surveys. The assessment is conducted face-to-face with parental involvement, fostering a supportive environment where children can freely express themselves through drawing, crafting, or pointing to icons. It takes approximately 5-10 minutes per child and evaluates 4 key domains: Conceptual Knowledge, Comprehension, Appraisal, and Application/function specific to CHD (Figure 1 [84]).

XSL•FO
RenderX

Figure 1. Final children's assessment, inspired by the Food Label Literacy for Applied Nutrition Knowledge (FLLANK) questionnaire [84].



CHD Parental Educational Burden Assessment

The CHD Parental Educational Burden Assessment (CHD-PEBA) is a 30-question assessment designed to evaluate parental educational responsibilities. It uses Likert scales, multiple-choice questions, matrix format, and open-ended responses to gather both quantitative and qualitative data on parental challenges. Inspired by Neuro-QoL and the Caregiver Health Self-Assessment Questionnaire [54,89,90], this self-report assessment examines parental understanding, coping mechanisms, and support needs in educating children about CHD. The assessment evaluates parents' knowledge of CHD, their perception of their children's understanding of the condition, and their confidence and stress levels in providing educational support. It also explores preferred information sources, information-seeking behaviors, and the time and effort parents dedicate to educating their children. Demographic data are collected to contextualize responses. Administered on the web, the survey takes approximately 5-10 minutes to complete. It serves as both a pre-and postintervention tool, measuring changes in parental educational burden over time.

CHD Healthcare Provider Educational Efficiency Assessment

The self-reported CHD Healthcare Provider Educational Efficiency Assessment (CHD-HEEA) health care providers' assessment is a 20-question self-report assessment to evaluate health care providers' educational efficiency in educating children with CHD and their families. It assesses the practices of pediatric cardiologists, fellows, nurses, and child life specialists through various question formats, including Likert scales, matrix questions, and multiple-choice and open-ended responses. Inspired by tools such as the Physician Task Checklist and Lean Six Sigma [60,63,91], the assessment examines educational practices, challenges, and strategies. Key areas of evaluation include the time and effort spent on educating children with CHD and families, perceptions of current educational methods, use of preappointment materials and self-education resources, encouragement of questions from caregivers and children, and strategies to streamline education while maintaining information accessibility. It also collects some demographic data to provide context for the responses. Administered on the web, the survey takes approximately 5-10 minutes to complete. It serves as a pre- and postassessment tool to evaluate changes in health care provider practices after educational interventions to optimize care coordination and patient education. Additional details about the 3 assessments are provided in Multimedia Appendix 2.

Pilot Testing of Assessment Tools

During the pilot phase, we tested 3 different assessments with CHD stakeholders at the M Health Fairview Pediatric Specialty Clinic for Cardiology—Explorer.

Study Population

The study defined specific criteria for each stakeholder group. Children with CHD were required to be between 4 and 10 years

XSL•FO
RenderX

of age, diagnosed by a health care provider, and fluent in English. Parents or guardians had to speak English and be confirmed as caregivers by a health care provider. Health care providers must be actively involved in CHD care and fluent in English.

Participants in this pilot study included children with CHD (n=5), their parents (n=4), and health care providers (n=3). Among the children were four 7-year-olds (3 males and 1 female) and one 5-year-old female (n=5). The parental and caregiver group comprised 5 mothers aged 25-44 years, including 2 White/Caucasian and 3 African American/Black participants, all with bachelor's degrees and reporting middle to high-level incomes (n=5). However, one of the parents did not complete the assessment, resulting in a final parental group size (n=4). The health care provider group included 2 pediatric cardiology fellows and 1 physician assistant student, aged 25-44 years, with 1-5 years of experience in pediatric cardiology (n=3).

Study Recruitment and Informed Consent

The medical team facilitated communication with potential children with CHD and their families through purposive sampling. Parents signed parental permission forms for their children and children provided assent. We administered the children's assessment while their parents were present but ensured that the parents did not guide their answers. After providing their own informed consent, parents received a link to complete the computer-based parental assessment. They could complete it at the clinic using the researcher's laptop or later at home. Parents shared their experiences and provided feedback on the assessment, including its clarity and length, either during or after completion. Health care providers also participated by completing their assessments and providing feedback on their experiences. All participants received a gift card for their time. We transcribed participants' input or feedback and anonymized results and transcripts to protect privacy.

Data Collection

We used qualitative methods, including cognitive interviews, observational notes, and interactive techniques, to gather feedback from children, parents, and health care providers [29,30,32,34]. These methods captured participants' impressions, behaviors, and engagement, providing rich data for iterative refinements of the assessments.

Logistics and Setup

To accommodate the clinic setting and time constraints, we conducted parent and child assessments simultaneously. Parents began slightly later than their children to ensure that the children were comfortable and understood the process before starting. This approach minimized parental influence on children's responses while saving time. After completing their assessments, children engaged in painting activities, allowing parents to complete their assessments without distractions. For health care providers, assessments were scheduled flexibly, either on the web or in person, to accommodate their busy schedules.

Cognitive and Observational Methods

Cognitive interviews and verbal probing revealed how participants interpreted and responded to survey questions. As a qualitative method widely used in survey design, cognitive interviews helped identify ambiguities, cognitive challenges, and difficulties in understanding. Participants verbalized their thoughts concurrently (using the think-aloud technique) or retrospectively (recalling and explaining their thought processes after completing the assessment) [29-34]. Observational techniques captured nonverbal cues, such as hesitation, frustration, or excitement, along with body language and engagement levels. These observations were particularly valuable for younger children, who often struggled with abstract concepts or articulating their thoughts [76]. Combining these methods provided detailed feedback that informed refinements to our assessments.

Engaging Children

We used developmentally appropriate and interactive methods to engage children effectively. A researcher read questions aloud to younger children, with parents assisting if the child felt uncomfortable interacting with the research team. To make the activity engaging, we incorporated storytelling and asked children to role-play as detectives solving questions. Prompts such as "Hey, detective! Let's figure out which, what, or where!" encouraged participation. Children used markers and craft materials, such as multicolor pom-poms, to point to their answers, making the process interactive and enjoyable. For older children who could read independently, we encouraged self-guided responses and asked clarifying questions to explore their thought processes. The think-aloud technique and verbal probing provided deeper insights into their reasoning [70-74]. To ensure that children understood the questions rather than guessing, we rephrased questions after they provided answers. For example, if a child pointed to an icon for chest pain, we asked, "Do you know where your chest is?" and used physical prompts to confirm their understanding. This approach validated their answers and often prompted children to share personal experiences, enriching the feedback [66,67,76,83].

Refining the Assessments

We iteratively refined the assessments, tailoring questions to each stakeholder group by focusing on their thoughts, understanding, relevance, completeness, survey length, and overall experience.

For the children's assessment, we prioritized meeting the developmental needs of children aged 4-10 years based on criteria provided by health care providers. These criteria outlined essential knowledge for this age group while addressing the additional challenges faced by younger or newly diagnosed children. Testing revealed that younger children often struggled with multistep questions, so we simplified these into single, clear actions. Language and visuals were adjusted for clarity while retaining enough detail to engage older children. Older children provided feedback on question relevance and reflected on how they might have responded when younger, offering insights that shaped refinements.

Each testing round informed adjustments to ensure that the questions were clear, precise, and engaging for all stakeholders in CHD pediatric care. These assessments now support pre- and postassessment stages to validate interventions, including our

XSL•FO
RenderX

ongoing study of medical educational toys for CHD pediatric populations. Table 1 outlines the development and testing process, with sample prompts provided in Multimedia Appendix 3.

Table 1. Iterative development and pilot testing of congenital heart disease (CHD) assessments.

Step	Description	Stakeholders involved	Outcomes
1. Define key concepts	Define health literacy, parental educational burden, and health care provider efficiency	Research team, pediatric cardiologist	Established key concepts for the assessment framework
2. Conduct literature review	Review existing best practices and assessments	Research team	Insights for developing new assessments
3. Develop initial assessments	Draft assessments for children, parents, and health care providers	Research team, pediatric cardiologist	Initial assessments ready for review
4. Pilot testing—first iteration	Test initial assessments with children with CHD, parents, and health care providers	Children with CHD, parents, and health care providers	Feedback on relevance, clarity, and appropriateness
5. Analyze feedback	Analyze feedback from pilot testing	Research team, pediatric cardiologist	Identify strengths, weaknesses, and improvements
6. Refine assessments	Modify assessments based on feedback	Research team	Improved assessments
7. Pilot testing—further iterations	Conduct additional testing and refinement	Children with CHD, parents, and health care providers	Continuous improvement and validation
8. Final analysis	Analyze all data and feedback to finalize assessments	Research team, pediatric cardiologist	Validated assessments for implementation

Data Analysis

Our research team held weekly meetings to analyze assessment data and stakeholder feedback. Using qualitative methods, including thematic analysis, we identified common themes from feedback and observational data [92,93]. The first author coded the data to highlight relevant themes, focusing on the assessment's reliability, engagement, and responsiveness to meet all stakeholders' needs. During each round of pilot testing, we prioritized feedback using four criteria: (1) its potential impact on the assessment's effectiveness, (2) alignment with intended outcomes, (3) feasibility of incorporating changes, and (4) stakeholder preferences. Feedback was categorized by feasibility, clarity, relevance to educational needs, and participant engagement, then ranked by the frequency and significance of reported issues. We implemented revisions iteratively, prioritizing aspects that most improved comprehension and ease of response. This systematic process enhanced the assessments' accuracy and relevance across stakeholder groups. During weekly meetings, the team reviewed how each piece of feedback aligned with the established criteria and worked collaboratively to resolve discrepancies. This iterative process led to the development of a set of codes such as "strengths," "weaknesses," and "areas of improvement." These ongoing discussions and thematic analyses refined the assessments and informed adjustments for subsequent pilot tests [94].

Ethical Considerations

The study received ethical approval from the University of Minnesota institutional review board (STUDY00020670). The medical team facilitated communication with potential participants—children with CHD and their families—through purposive sampling. Informed consent was obtained from all participants: parents also signed parental permission forms for their children, and children provided assent. Participants were informed about the study's purpose, procedures, and their right to withdraw at any time. Data were anonymized and deidentified during transcription and analysis, with all personal information securely stored in adherence to institutional guidelines. Participants received a gift card as compensation for their time, ensuring fairness and transparency.

Results

We identified consistent themes across three assessments: (1) assessment engagement, (2) relevance and structure of assessment, and (3) opportunities of assessments. These themes guided iterative revisions before each new pilot test. In this section, we summarize the findings from (1) children's assessments, (2) parents' assessments, and (3) health care providers' assessments. Some feedback from key stakeholders on 3 survey experiences can be found in Multimedia Appendix 2.

Assessments Engagement

Both children (C) and parents (P) actively engaged in the children's assessment process. Parents were surprised by their children's interest and enjoyment, with one parent expressing, "I wasn't sure she'd even talk. Wow!" It appeared that the interactive elements, such as choosing pens or markers and storytelling, not only boosted curiosity and made them concentrate on doing the assessment but also eased the anxiety and fear of children. For instance, C3 quickly transitioned from nervousness about another medical procedure to excitement upon seeing colorful markers, exclaiming, "Yay, I can paint here in the doctor's office!" Initially, she hid under the bed due to fear when we entered the examination room. However, she relaxed and became comfortable after seeing the markers and being invited to answer the questions using different colors. Even children like C2, initially focused on their iPad, became

XSL•FO
RenderX

engaged, asking, “Can I use all colors? I can answer anything like this!”

Parents appreciated the playful and instructive design of the assessments, with one commenting, “It’s like teaching by itself through playtime. She’s telling me all you've asked her.” After finishing their assessments, we provided paper to children like C4, whose mother was also being assessed. This engagement prompted them to draw, becoming so engrossed to continue that they requested us to stay there even after the doctor’s arrival (Figure 2). Initially, parental influence impacted C1’s responses, but gentle interventions and prioritizing children’s uninfluenced responses over accuracy fostered independent responses in other pilot tests. Through face-to-face interaction using cognitive interviews akin to semistructured interviews [95,96], we collected less biased data from children rather than their parents, maintained attention, and reduced parental influence.

Figure 2. Children’s interactions with children’s assessments and parental responses to their assessments at the Explorer clinic.

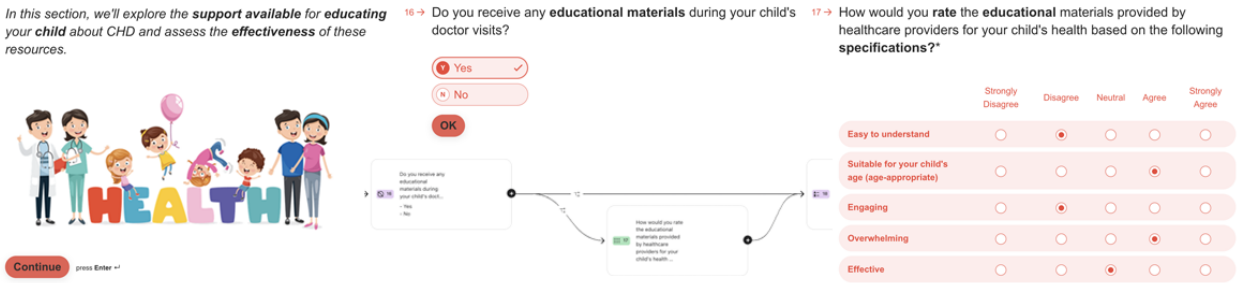
Similarly, in the clinic, parents provided feedback after their children’s assessments through a think-aloud and verbal probing method [29,32] while answering the computer-based parents’ assessment using Typeform (Typeform SL). During the assessment, P4 noted, “You know what to ask.” P3 mentioned that they would change only some of the wording. However, despite this feedback, P3 described the survey as “clear, easy, feeling good; finally, somebody asks!” This feedback highlighted the ease of use, clear directions, and meeting the participants’ needs, facilitating effective data collection. The health care providers’ assessment also received positive feedback for its engaging interface, with H1 noting, “Much better than usual surveys,” and H3 stating, “...really interactive with photos and video, and buttons. It’s easy....” Such feedback underscores the effectiveness of engaging participants and facilitating data collection.

Relevance and Structure of Assessment

The effectiveness of engaging all stakeholders through assessments depends on their relevance and well-structured design. Maintaining high relevance across all assessments, we noticed an interesting trend involving the Smiley Face Likert question during the children’s assessment. Despite our efforts to adjust the question wording, children consistently chose the happiest face in the initial version of the assessment. They emphasized their positive feelings associated with smiley faces by saying, “I chose it because I like it!” We expected this issue but still tested the smiley face Likert question due to its common use in children’s surveys. To improve engagement and reliability, we found that comparison questions, structured like a right or wrong game format, were more effective. Children showed high concentration levels and often asked, “Is this right?” They felt like participating in a game, increasing their involvement in the assessment process. Incorporating feedback from our medical team and parents highlighting the assessment’s educational value, we replaced the Likert question with a prompt offering a choice between physical activity and screen time (Figure 1 [84]), addressing another relevant habit for children with CHD. We also adjusted the question order to observe response variations, albeit with limited reliability due to its implementation with only participants C2 and C3 immediately after the initial version.

P2, P3, P4, and P5 actively engaged with the questions in the parents’ assessment, finding them clear, easy, and relevant. However, P2 noted assumptions in specific questions regarding the prior receipt of educational material during doctor visits. They proposed a preliminary question to confirm whether educational material was used before assessing its appropriateness (Figure 3). They asked, “How do you know if we got any educational material before asking how good it was? It’d make more sense to check if we received any and then ask about it.” Interestingly, when explicitly asked about this question, P3, a follow-up patient, did not express the same concern. Another participant suggested improving the context of the question about comfort levels during clinic visits (follow-ups or surgeries), which we addressed in the later version. We condensed the questionnaire to 25 items in the latest parent assessments by adding a matrix question.

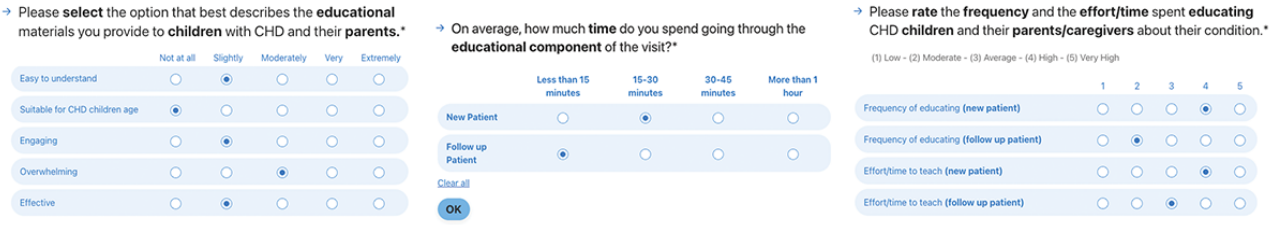
Figure 3. Congenital Heart Disease Parental Educational Burden Assessment (CHD-PEBA) logic model question.



XSL•FO
RenderX

In response to initial feedback on the health care providers' assessment, H1 proposed categorizing questions differently for new and follow-up patients. All participants provided positive feedback regarding the matrix questions (Figure 4). H2 preferred to keep the existing matrix questions but suggested reordering them for better clarity and efficiency. They emphasized that combining related aspects into single questions helps respondents compare options and find the correct answer more easily: "Having one question for each of these aspects would make the questionnaire longer and harder to compare when answering; it just seems more practical this way." Furthermore, H2 recommended adding a specific demographic—vulnerable parents—to measure parental CHD information instead of using a general category. These suggestions were integrated into version 03, reducing the length to 15 items.

Figure 4. Congenital Heart Disease Healthcare Provider Educational Efficiency Assessment (CHD-HEEA) matrix questions.



Opportunities of Assessments

We used patients' wait time between their initial screenings and the meeting with their physician to complete the assessments. This integration not only streamlined care coordination and reduced wait times but also engaged families and alleviated worries, as noted by P4, "It keeps us good busy!" This suggests that our design intervention (an educational toy) can be implemented during patients' wait time as well. The assessments also doubled as educational tools for children, engaging them in learning about their condition and easing anxiety. Parents found solace in sharing burdens, while health care providers gained insights into their practices and sought potential solutions. These assessments exceeded their initial purposes, offering education, support, and reflection opportunities for child-led approaches in pediatric care.

Discussion

Principal Results

During face-to-face interviews, we identified areas for refinement in assessment items, such as clarifying the context of health care experiences for children's comfort-level rating or specifying the type of visit for health care providers. Observational notes highlighted instances where assumptions in specific questions, such as the prior receipt of educational material during doctor visits, caused confusion among interviewers. To alleviate this issue, we proposed preliminary questions. Despite these challenges, participants generally interpreted the majority of survey items as intended, with occasional adjustments to wording choices for better comprehension. Furthermore, we changed the order of questions based on feedback and observations to improve the structure of the assessment.

The pilot testing evaluated the feasibility and effectiveness of the tailored assessments. Iterative revisions further improved their clarity and appropriateness. We ensured the validity of the assessments through a collaborative and iterative approach, incorporating perspectives from all stakeholders. As a result, the assessments offer a comprehensive understanding of educational needs and burdens, providing valuable insights to guide the development of targeted educational interventions.

Comparison With Prior Work

The tailored assessments developed in this pilot study address a gap in the existing literature by focusing on the educational needs of pediatric patients with CHD, their parents, and health care providers [7,8,27]. While existing assessment tools mainly measure health-related quality of life and emotional well-being, they offer limited insight into health literacy and the unique educational challenges associated with CHD [38-56]. Moreover, these assessments are primarily developed quantitatively and lack qualitative insights into educational needs and burdens. We used qualitative methods to explore how respondents interpret, understand, and respond to specific survey items. This approach offered a more comprehensive understanding of the questions that assessments should address [29-32,93-97].

Through collaboration with stakeholders, including children, we developed and refined assessments to evaluate children's knowledge, parental educational burdens, and health care provider efficiency in pediatric CHD care. Including children in the development process was crucial due to the lack of tailored educational materials for children with CHD and the complexities involved in assessing this population [64-67,98]. While parents were present during the children's assessments to provide comfort, their involvement was carefully managed to ensure that it did not interfere with the children's meaningful participation. This approach contrasts with prior methods, where parents often act as proxies, potentially biasing the results [68,69]. Feedback from all stakeholders was collected post assessment to gain insights into their experiences and further improve the assessments.

We conducted pilot tests for two reasons: (1) to refine the assessments before involving more stakeholders, specifically children, and (2) to ensure the assessments were well adapted to meet all stakeholders' needs before the actual study [99,100]. This pilot study revealed critical gaps in existing tools and advanced the methodology for developing educational assessments. Emphasizing collaborative, iterative, and direct engagement with children, parents, and health care providers

XSL•FO
RenderX

ultimately leads to the design of tailored assessments. These assessments are crucial for informing the development of effective, child-led educational interventions for pediatric populations with CHD, demonstrating feasibility before broader implementation.

The assessments developed in this study have potential applications in routine clinical practice, offering health care providers a structured tool to assess and enhance CHD-related health literacy among pediatric patients and their families. By integrating these assessments into preappointment resources or waiting room activities, health care providers can identify educational gaps early and address them proactively. Furthermore, the assessments could complement existing educational interventions, providing feedback on the effectiveness of child- and family-centered resources to improve health literacy. This integration would not only inform the design of more effective child- and family-centered resources but also support continuous improvement of educational interventions. This approach aligns with broader goals in pediatric health care to support lifelong health management through early literacy interventions.

Limitations

Although the pilot testing provided valuable insights, several limitations are acknowledged. First, despite efforts to minimize parental influence, their presence may have still impacted some children's responses. Future research should explore strategies to reduce this influence further and ensure more independent responses from children. Second, the sample distribution was not spread evenly across age groups, with only 1 child aged 5 years and 4 children aged 7 years. This uneven distribution may affect the representativeness of the findings. Future studies should aim for a more balanced age distribution. Third, conducting assessments for both children and parents at the clinic during their visit might have influenced their responses, as they were exposed to medical procedures. Furthermore, the researcher's presence during the assessment of parents and children could have influenced their feedback. Finally, while general accessibility was considered, specific adaptations for educational disabilities were not within the scope of this pilot study.

Conclusions

This pilot study aims to improve educational interventions and care coordination through tailored assessments designed by stakeholder feedback, including affected children. The findings emphasize the importance of interactive and qualitative methods that foster multistakeholder engagement and ensure question relevance. These assessments can potentially enhance child-led interventions and improve outcomes for patients with CHD from childhood through to adult care. However, the iterative development process highlighted several challenges, such as managing parental presence to encourage independent responses from children and the logistical complexities of recruiting diverse participants within clinical settings. It is also important to ensure that assessments for younger children are read aloud without influencing their responses. Integrating these assessments into real-world clinical workflows requires adaptability to avoid disrupting patient care. Addressing these practical challenges will be essential to scaling and sustaining the use of these assessments in diverse health care contexts.

Acknowledgments

The authors gratefully acknowledge the funding provided by the Kusske Design Initiative (KDI) and Dr Gwenyth Fischer, MD. Their sincere thanks extend to the M Health Fairview Explorer Clinic and all pilot participants. Finally, they would like to thank contributions from former students Grace Rubas, Jessica Jenkins, Jonathan Jakubas, Levi Skelton, and Andy Thai.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Existing assessment tools.
[DOCX File , 21 KB - formative_v9i1e63818_app1.docx]

Multimedia Appendix 2
Three congenital heart disease–specific assessments.
[DOCX File , 3612 KB - formative_v9i1e63818_app2.docx]

Multimedia Appendix 3
Sample prompts and feedback from data collection.
[DOCX File , 23 KB - formative_v9i1e63818_app3.docx]

References

1. Centers for Disease Control and Prevention. Data and statistics on congenital heart defects. 2023. URL: https://www.cdc.gov/heart-defects/data/index.html [accessed 2023-09-22]

XSL•FO
RenderX

2. GBD 2017 Congenital Heart Disease Collaborators. Global, regional, and national burden of congenital heart disease, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Child Adolesc Health 2020;4(3):185-200 [FREE Full text] [doi: 10.1016/S2352-4642(19)30402-X] [Medline: 31978374]
3. Liu Y, Chen S, Zühlke L, Black GC, Choy M, Li N, et al. Global birth prevalence of congenital heart defects 1970-2017: updated systematic review and meta-analysis of 260 studies. Int J Epidemiol 2019;48(2):455-463 [FREE Full text] [doi: 10.1093/ije/dyz009] [Medline: 30783674]
4. Chong LSH, Fitzgerald DA, Craig JC, Manera KE, Hanson CS, Celermajer D, et al. Children's experiences of congenital heart disease: a systematic review of qualitative studies. Eur J Pediatr 2018;177(3):319-336. [doi: 10.1007/s00431-017-3081-y] [Medline: 29327140]
5. Lee A, Bailey B, Cullen-Dean G, Aiello S, Morin J, Oechslin E. Transition of care in congenital heart disease: ensuring the proper handoff. Curr Cardiol Rep 2017;19(6):55. [doi: 10.1007/s11886-017-0859-5] [Medline: 28477319]
6. Moons P, De Volder E, Budts W, De Geest S, Elen J, Waeytens K, et al. What do adult patients with congenital heart disease know about their disease, treatment, and prevention of complications? A call for structured patient education. Heart 2001;86(1):74-80 [FREE Full text] [doi: 10.1136/heart.86.1.74] [Medline: 11410567]
7. Barbazi N, Shin JY, Hiremath G, Lauff CA. Exploring Health Educational Interventions for Children with Congenital Heart Disease: A Scoping Review. JMIR Pediatr Parent 2024 Jan 24;8:e64814 [FREE Full text] [doi: http://dx.doi.org/10.2196/64814]
8. Wray J, Maynard L. The needs of families of children with heart disease. J Dev Behav Pediatr 2006;27(1):11-17. [doi: 10.1097/00004703-200602000-00002] [Medline: 16511363]
9. Veldtman GR, Matley SL, Kendall L, Quirk J, Gibbs JL, Parsons JM, et al. Illness understanding in children and adolescents with heart disease. Heart 2000;84(4):395-397 [FREE Full text] [doi: 10.1136/heart.84.4.395] [Medline: 10995408]
10. Goossens E, Fieuws S, Van Deyk K, Luyckx K, Gewillig M, Budts W, et al. Effectiveness of structured education on knowledge and health behaviors in patients with congenital heart disease. J Pediatr 2015;166(6):1370-1376.e1. [doi: 10.1016/j.jpeds.2015.02.041] [Medline: 25841537]
11. Burns J, Higgins C, Ganigara M, Kalivas B, Basken A. Health literacy in CHD. Cardiol Young 2022;32(9):1369-1372. [doi: 10.1017/S1047951122001020] [Medline: 35611812]
12. van der Heijden Z, de Gooijer F, Camps G, Lucassen D, Feskens E, Lasschuijt M, et al. User requirements in developing a novel dietary assessment tool for children: mixed methods study. JMIR Form Res 2024;8:e47850 [FREE Full text] [doi: 10.2196/47850] [Medline: 38300689]
13. Francis-Oliviero F, Loubières C, Grové C, Marinucci A, Shankland R, Salamon R, et al. Improving children's mental health literacy through the cocreation of an intervention and scale validation: protocol for the CHILD-Mental Health literacy research study. JMIR Res Protoc 2023;12:e51096 [FREE Full text] [doi: 10.2196/51096] [Medline: 37796588]
14. Gupta A, Cafazzo JA, IJzerman MJ, Swart JF, Vastert S, Wulffraat NM, et al. Genomic health literacy interventions in pediatrics: scoping review. J Med Internet Res 2021;23(12):e26684 [FREE Full text] [doi: 10.2196/26684] [Medline: 34951592]
15. Shin JY, Holtz B. Identifying opportunities and challenges: how children use technologies for managing diabetes. 2020 Presented at: IDC '20: Interaction Design and Children; 2020 June 24; London, United Kingdom p. 495-507 URL: https://dl.acm.org/doi/10.1145/3392063.3394444
16. Shin JY, Holtz BE. Towards better transitions for children with diabetes: user experiences on a mobile health app. 2019 Presented at: Proceedings of the 18th ACM International Conference on Interaction Design and Children; 2019 June 15; Boise, ID. [doi: 10.1145/3311927.3325319]
17. DeHoff BA, Staten LK, Rodgers RC, Denne SC. The role of online social support in supporting and educating parents of young children with special health care needs in the United States: a scoping review. J Med Internet Res 2016;18(12):e333 [FREE Full text] [doi: 10.2196/jmir.6722] [Medline: 28007689]
18. Shin JY, Kedroske J, Vue R, Sankaran R, Chaar D, Churay T, et al. Design considerations for family-centered health management: preliminary findings with pediatric BMT patients. 2018 Presented at: Proceedings of the 17th ACM Conference on Interaction Design and Children; 2018 June 22; Trondheim, Norway. [doi: 10.1145/3202185.3210781]
19. Chaar D, Shin JY, Mazzoli A, Vue R, Kedroske J, Chappell G, et al. A mobile health app (Roadmap 2.0) for patients undergoing hematopoietic stem cell transplant: qualitative study on family caregivers' perspectives and design considerations. JMIR Mhealth Uhealth 2019;7(10):e15775 [FREE Full text] [doi: 10.2196/15775] [Medline: 31651402]
20. Rodts ME, Unaka NI, Statile CJ, Madsen NL. Health literacy and caregiver understanding in the CHD population. Cardiol Young 2020;30(10):1439-1444. [doi: 10.1017/S1047951120002243] [Medline: 32746956]
21. Tennant R, Allana S, Mercer K, Burns CM. Exploring the experiences of family caregivers of children with special health care needs to inform the design of digital health systems: formative qualitative study. JMIR Form Res 2022;6(1):e28895 [FREE Full text] [doi: 10.2196/28895] [Medline: 34989692]
22. Shin JY, Okammor N, Hendee K, Pawlikowski A, Jenq G, Bozaan D. Development of the socioeconomic screening, active engagement, follow-up, education, discharge readiness, and consistency (SAFEDC) model for improving transitions of care: participatory design. JMIR Form Res 2022;6(4):e31277 [FREE Full text] [doi: 10.2196/31277] [Medline: 35412461]

XSL•FO
RenderX

23. Hölgyesi Á, Luczay A, Tóth-Heyn P, Muzslay E, Világos E, Szabó AJ, et al. The impact of parental electronic health literacy on disease management and outcomes in pediatric type 1 diabetes mellitus: cross-sectional clinical study. JMIR Pediatr Parent 2024;7:e54807 [FREE Full text] [doi: 10.2196/54807] [Medline: 38506893]
24. Sakamoto M, Ishikawa H, Suzuki A. Evaluation of parents' use of a child health care information app and their health literacy: cross-sectional study. JMIR Pediatr Parent 2024;7:e48478 [FREE Full text] [doi: 10.2196/48478] [Medline: 38623727]
25. Castor C, Lindkvist RM, Hallström IK, Holmberg R. Health care professionals' experiences and views of eHealth in pediatric care: qualitative interview study applying a theoretical framework for implementation. JMIR Pediatr Parent 2023;6:e47663 [FREE Full text] [doi: 10.2196/47663] [Medline: 37851500]
26. Lindström NB, Pozo RR. Perspectives of nurses and doulas on the use of information and communication technology in intercultural pediatric care: qualitative pilot study. JMIR Pediatr Parent 2020;3(1):e16545 [FREE Full text] [doi: 10.2196/16545] [Medline: 32181748]
27. Arya B, Glickstein JS, Levasseur SM, Williams IA. Parents of children with congenital heart disease prefer more information than cardiologists provide. Congenit Heart Dis 2013;8(1):78-85 [FREE Full text] [doi: 10.1111/j.1747-0803.2012.00706.x] [Medline: 22891764]
28. Van Orne J. Care coordination for children with medical complexity and caregiver empowerment in the process: a literature review. J Spec Pediatr Nurs 2022;27(3):e12387. [doi: 10.1111/jspn.12387] [Medline: 35671389]
29. Willis GB, Artino AR. What do qur respondents think we're Asking? using cognitive interviewing to improve medical education surveys. J Grad Med Educ 2013;5(3):353-356 [FREE Full text] [doi: 10.4300/JGME-D-13-00154.1] [Medline: 24404294]
30. Balza JS, Cusatis RN, McDonnell SM, Basir MA, Flynn KE. Effective questionnaire design: how to use cognitive interviews to refine questionnaire items. J Neonatal Perinatal Med 2022;15(2):345-349 [FREE Full text] [doi: 10.3233/NPM-210848] [Medline: 35034911]
31. Kedroske J, Koblick S, Chaar D, Mazzoli A, O'Brien M, Yahng L, et al. Development of a national caregiver health survey for hematopoietic stem cell transplant: qualitative study of cognitive interviews and verbal probing. JMIR Form Res 2020;4(1):e17077 [FREE Full text] [doi: 10.2196/17077] [Medline: 32012037]
32. Eccles DW, Arsal G. The think aloud method: what is it and how do I use it? Qual Res Sport Exerc Health 2017;9(4):514-531. [doi: 10.1080/2159676x.2017.1331501]
33. Presser S, Couper MP, Lessler JT, Martin E, Martin J, Rothgeb JM, et al. Methods for testing and evaluating survey questions. Public Opin Q 2004;68(1):109-130. [doi: 10.1093/poq/nfh008]
34. Collins D. Pretesting survey instruments: an overview of cognitive methods. Qual Life Res 2003;12(3):229-238. [doi: 10.1023/a:1023254226592] [Medline: 12769135]
35. Mingjie Z, Li W, Jianxin Z. Developing a computer system for health assessment. 2009 Presented at: Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Huma; 2009 November 26; New York, NY p. 303-306. [doi: 10.1145/1655925.1655981]
36. Morris M, Intille S. HCI challenges in health assessment. 2005 Presented at: CHI '05 Extended Abstracts on Human Factors in Computing Systems; 2005 April 7; New York, NY. [doi: 10.1145/1056808.1057122]
37. Ricci M, Cimini A, Gli?ovi? K, Jiménez RJ, Medina-García R, Steinböck M, et al. Quality of life assessment methodology in TeNDER project. 2022 Presented at: Proceedings of the 15th International Conference on Pervasive Technologies Related to Assistive Environments; 2022 June 29; New York, NY. [doi: 10.1145/3529190.3534743]
38. Ayers TS, Sandler IN, West SG, Roosa MW. Children?s Coping Strategies Checklist. 1991. URL: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft42024-000 [accessed 2024-12-13]
39. ScienceDirect Topics. Child behavior checklist—an overview. URL: https://www-sciencedirect-com.ezp2.lib.umn.edu/topics/psychology/child-behavior-checklist [accessed 2024-03-12]
40. Haley SM, Coster WI, Kao YC, Dumas HM, Fragala-Pinkham MA, Kramer JM, et al. Lessons from use of the pediatric evaluation of disability inventory: where do we go from here? Pediatr Phys Ther 2010;22(1):69-75 [FREE Full text] [doi: 10.1097/PEP.0b013e3181cbfbf6] [Medline: 20142708]
41. Horner-Johnson W, Krahn G, Andresen E, Hall T, Rehabilitation ResearchTraining Center Expert Panel on Health Status Measurement. Developing summary scores of health-related quality of life for a population-based survey. Public Health Rep 2009;124(1):103-110 [FREE Full text] [doi: 10.1177/003335490912400113] [Medline: 19413032]
42. Hullmann SE, Ryan JL, Ramsey RR, Chaney JM, Mullins LL. Measures of general pediatric quality of life: Child Health Questionnaire (CHQ), DISABKIDS Chronic Generic Measure (DCGM), KINDL-R, Pediatric Quality of Life Inventory (PedsQL) 4.0 Generic Core Scales, and Quality of My Life Questionnaire (QoML). Arthritis Care Res (Hoboken) 2011;63 Suppl 11:S420-S430 [FREE Full text] [doi: 10.1002/acr.20637] [Medline: 22588762]
43. Child Outcomes Research Consortium. Strengths and Difficulties Questionnaire (SDQ). URL: https://www.corc.uk.net/outcome-experience-measures/strengths-and-difficulties-questionnaire-sdq/ [accessed 2024-12-20]
44. Marino BS, Tomlinson RS, Wernovsky G, Drotar D, Newburger JW, Mahony L, Pediatric Cardiac Quality of Life Inventory Testing Study Consortium. Validation of the pediatric cardiac quality of life inventory. Pediatrics 2010;126(3):498-508 [FREE Full text] [doi: 10.1542/peds.2009-2973] [Medline: 20805147]

XSL•FO
RenderX

45. Pearson. Pediatric Evaluation of Disability Inventory. URL: https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Developmental-Early-Childhood/Pediatric-Evaluation-of-Disability-Inventory/p/100000505.html [accessed 2024-12-20]
46. Cleveland Clinic Children's. Pediatric Cardiac Quality of Life Inventory (PCQLI). URL: https://my.clevelandclinic.org/pediatrics/medical-professionals/pediatric-cardiac-quality-life-inventory [accessed 2024-12-20]
47. PedsQL TM (Pediatric Quality of Life Inventory TM). URL: https://www.pedsql.org/about_pedsql.html [accessed 2024-12-20]
48. Health Literacy Tool Shed. URL: https://healthliteracy.tuftsmedicine.org/ [accessed 2024-12-20]
49. Reynolds JS, Treu JA, Njike V, Walker J, Smith E, Katz CS, et al. The validation of a food label literacy questionnaire for elementary school children. J Nutr Educ Behav 2012;44(3):262-266. [doi: 10.1016/j.jneb.2011.09.006] [Medline: 22406011]
50. Child Outcomes Research Consortium. Parental Stress Scale (PSS). URL: https://www.corc.uk.net/outcome-experience-measures/parental-stress-scale-pss/ [accessed 2024-12-20]
51. American Psychological Association. Parenting Stress Index. URL: https://www.apa.org/pi/about/publications/caregivers/practice-settings/assessment/tools/parenting-stress [accessed 2024-12-20]
52. Ríos M, Zekri S, Alonso-Esteban Y, Navarro-Pardo E. Parental stress assessment with the Parenting Stress Index (PSI): a systematic review of its psychometric properties. Children (Basel) 2022;9(11):1649 [FREE Full text] [doi: 10.3390/children9111649] [Medline: 36360377]
53. Varni JW, Sherman SA, Burwinkle TM, Dickinson PE, Dixon P. The PedsQL family impact module: preliminary reliability and validity. Health Qual Life Outcomes 2004;2:55 [FREE Full text] [doi: 10.1186/1477-7525-2-55] [Medline: 15450120]
54. HealthinAging. org. Caregiver Self Assessment Questionnaire. URL: https://www.healthinaging.org/tools-and-tips/caregiver-self-assessment-questionnaire [accessed 2024-12-20]
55. Epstein-Lubow G, Gaudiano BA, Hinckley M, Salloway S, Miller IW. Evidence for the validity of the American Medical Association's caregiver self-assessment questionnaire as a screening measure for depression. J Am Geriatr Soc 2010;58(2):387-388. [doi: 10.1111/j.1532-5415.2009.02701.x] [Medline: 20370867]
56. Hanmer J, Jensen RE, Rothrock N, HealthMeasures Team. A reporting checklist for HealthMeasures' patient-reported outcomes: ASCQ-Me, Neuro-QoL, NIH Toolbox, and PROMIS. J Patient Rep Outcomes 2020;4(1):21 [FREE Full text] [doi: 10.1186/s41687-020-0176-4] [Medline: 32215788]
57. de Brantes F. Measuring Provider Efficiency Version 1. 2004. URL: https://www.commonwealthfund.org/sites/default/files/documents/___media_files_publications_other_2004_dec_measuring_provider_efficiency__version_1_0__a_collaborative_multi_stakeholder_effort_measurproviderefficiency1_12312004_pdf.pdf [accessed 2004-12-03]
58. Palmer S, Torgerson DJ. Economic notes: definitions of efficiency. BMJ 1999;318(7191):1136 [FREE Full text] [doi: 10.1136/bmj.318.7191.1136] [Medline: 10213735]
59. Bailie R, Bailie J, Larkins S, Broughton E. Editorial: continuous quality improvement (CQI)—advancing understanding of design, application, impact, and evaluation of CQI approaches. Frontiers in Public Health. 2017. URL: https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2017.00306 [accessed 2024-02-24]
60. Henrique DB, Godinho Filho M. A systematic literature review of empirical research in Lean and Six Sigma in healthcare. Total Qual Manag Bus Excell 2018;31(3-4):429-449. [doi: 10.1080/14783363.2018.1429259]
61. Hussey PS, de Vries H, Romley J, Wang MC, Chen SS, Shekelle PG, et al. A systematic review of health care efficiency measures. Health Serv Res 2009;44(3):784-805 [FREE Full text] [doi: 10.1111/j.1475-6773.2008.00942.x] [Medline: 19187184]
62. O'Neill SM, Hempel S, Lim YW, Danz MS, Foy R, Suttorp MJ, et al. Identifying continuous quality improvement publications: what makes an improvement intervention 'CQI'? BMJ Qual Saf 2011;20(12):1011-1019 [FREE Full text] [doi: 10.1136/bmjqs.2010.050880] [Medline: 21727199]
63. Rathi R, Vakharia A, Shadab M. Lean Six Sigma in the healthcare sector: a systematic literature review. Mater Today Proc 2022;50:773-781 [FREE Full text] [doi: 10.1016/j.matpr.2021.05.534] [Medline: 35155129]
64. Shepard L, Kagan SL, Wurtz E. Principles and recommendations for early childhood assessments. National Education Goals Panel. 1998. URL: https://eric.ed.gov/?id=ED416033 [accessed 2024-03-16]
65. Committee on the Science of Children Birth to Age 8: Deepening and Broadening the Foundation for Success, Board on Children, Youth, and Families, Institute of Medicine. In: National Research Council. Transforming the Workforce for Children Birth Through Age 8: A Unifying Foundation. Washington, DC: National Academies Press; 2015.
66. Banker A, Lauff C. Usability testing with children: history of best practices, comparison of methods & gaps in literature. URL: https://dl.designresearchsociety.org/drs-conference-papers/drs2022/researchpapers/225 [accessed 2022-06-25]
67. Vang PC, Lauff CA. Reflections on data collection during toy prototype development in a design studio course. 2024 Presented at: Proceedings of the 23rd Annual ACM Interaction Design and Children Conference; 2024 June 20; New York, NY. [doi: 10.1145/3628516.3659422]
68. Koutná V, Blatný M, Jelínek M. Concordance of child self-reported and parent proxy-reported posttraumatic growth in childhood cancer survivors. Cancers (Basel) 2021;13(16):4230 [FREE Full text] [doi: 10.3390/cancers13164230] [Medline: 34439384]

XSL•FO
RenderX

69. Smriti D, Kao TSA, Rathod R, Shin JY, Peng W, Williams J, et al. Motivational interviewing conversational agent for parents as proxies for their children in healthy eating: development and user testing. JMIR Hum Factors 2022;9(4):e38908 [FREE Full text] [doi: 10.2196/38908] [Medline: 36206036]
70. Angell C, Alexander J, Hunt JA. 'Draw, write and tell': a literature review and methodological development on the 'draw and write' research method. J Early Child Res 2014;13(1):17-28. [doi: 10.1177/1476718x14538592]
71. McWhirter J. The draw and write technique as a versatile tool for researching children's understanding of health and well-being. Int J Health Promot Educ 2014;52(5):250-259. [doi: 10.1080/14635240.2014.912123]
72. Teela L, Verhagen LE, van Oers HA, Kramer EEW, Daams JG, Gruppen MP, et al. Pediatric patient engagement in clinical care, research and intervention development: a scoping review. J Patient Rep Outcomes 2023;7(1):32 [FREE Full text] [doi: 10.1186/s41687-023-00566-y] [Medline: 36988738]
73. Wang Q, Hay M, Clarke D, Menahem S. Adolescents' drawings of their cardiac abnormality. Cardiol Young 2011;21(5):556-561. [doi: 10.1017/S1047951111000461] [Medline: 21554829]
74. Noonan RJ, Boddy LM, Fairclough SJ, Knowles ZR. Write, draw, show, and tell: a child-centred dual methodology to explore perceptions of out-of-school physical activity. BMC Public Health 2016;16:326 [FREE Full text] [doi: 10.1186/s12889-016-3005-1] [Medline: 27080384]
75. Flynn R, Walton S, Scott SD. Engaging children and families in pediatric health research: a scoping review. Res Involv Engagem 2019;5:32 [FREE Full text] [doi: 10.1186/s40900-019-0168-9] [Medline: 31700676]
76. Druin A. The role of children in the design of new technology. Behav Inf Technol 2002;21(1):1-25. [doi: 10.1080/01449290110108659]
77. Sanders EBN, Brandt E, Binder T. A framework for organizing the toolstechniques of participatory design. 2010 Presented at: Proceedings of the 11th Biennial Participatory Design Conference; 2010 November 29; New York, NY p. 195-198. [doi: 10.1145/1900441.1900476]
78. Read JC, MacFarlane S. Using the fun toolkit and other survey methods to gather opinions in child computer interaction. 2006 Presented at: Proceedings of the 2006 Conference on Interaction Design and Children; 2006 June 9; New York, NY p. A. [doi: 10.1145/1139073.1139096]
79. Bradbury-Jones C, Taylor J. Engaging with children as co-researchers: challenges,counter-challenges and solutions. Int J Soc Res Methodol 2013;18(2):161-173. [doi: 10.1080/13645579.2013.864589]
80. Abdelghani R, Law E, Desvaux C, Oudeyer PY, Sauzéon H. Interactive environments for training children's curiosity through the practice of metacognitive skills: a pilot study. : Association for Computing Machinery; 2023 Presented at: Proceedings of the 22nd Annual ACM Interaction Design and Children Conference; 2023 June 23; New York, NY p. 495-501. [doi: 10.1145/3585088.3593880]
81. Boberg M, Karapanos E, Holopainen J, Lucero A. PLEXQ: Towards a Playful Experiences Questionnaire. New York, NY: Association for Computing Machinery; 2015 Presented at: Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play; 2015 October 7; London. [doi: 10.1145/2793107.2793124]
82. Kellett M, Forrest R, Dent N, Ward S. 'Just teach us the skills please, we'll do the rest': empowering ten ‐ year ‐ olds as active researchers. Children Soc 2006;18(5):329-343. [doi: 10.1002/chi.807]
83. Druin A. Cooperative inquiry: developing new technologies for children with children. Proc SIGCHI Conf Hum Factors Comput Syst New York, NY. New York, NY: Association for Computing Machinery; 1999 Presented at: Proceedings of the SIGCHI conference on Human Factors in Computing Systems; 1999 May 20; Pittsburgh, PA. [doi: 10.1145/302979.303166]
84. Yale-Griffin Prevention Research Center. Food Label Literacy for Applied Nutrition Knowledge (FLLANK). URL: https://yalegriffinprc.griffinhealth.org/products-resources/prc-products/food-label-literacy [accessed 2023-12-20]
85. Stals S, Voysey I, Baillie L. Let's make this fun!: activities to motivate children teens to complete questionnaires. New York, NY: Association for Computing Machinery; 2024 Presented at: ACMIEEE International Conference on Human-Robot Interaction; 2024 March 15; Boulder, CO. [doi: 10.1145/3610978.3640675]
86. Gaines KS, Curry ZD. The inclusive classroom: the effects of color on learning and behavior. J Fam Consum Sci Educ 2011;29(1):46-57 [FREE Full text]
87. Barbazi N, Wang CX. Hum factors Aging Spec Needs. Perceiving through colors: visual supports for children with autism 2023;88:104-112. [doi: 10.54941/ahfe1003667]
88. Hall L, Hume C, Tazzyman S. Five degrees of happinessffective smiley face likert scales for evaluating with children. New York, NY: Association for Computing Machinery; 2016 Presented at: Proceedings of the 15th International Conference on Interaction Design and Children; 2016 June 24; Manchester, UK p. 311-321. [doi: 10.1145/2930674.2930719]
89. National Cancer Institute at the National Institutes of Health. Health measures. URL: https://healthcaredelivery.cancer.gov/healthmeasures/ [accessed 2024-12-20]
90. HealthMeasures. Neuro-QoL. 2024. URL: https://www.healthmeasures.net/explore-measurement-systems/neuro-qol [accessed 2024-02-17]
91. Stein DM, Wrenn JO, Stetson PD, Bakken S. What "to-do" with physician task lists: clinical task model development and electronic health record design implications. AMIA Annu Symp Proc 2009;2009:624-628 [FREE Full text] [Medline: 20351930]

XSL•FO
RenderX

92. Braun V, Clarke V. Thematic analysis. In: APA Handbook of Research Methods in Psychology Vol 2. Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological. Washington, DC: American Psychological Association; 2012:57-71.
93. Creswell JW, Creswell JD. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Washington DC: SAGE Publications; 2017.
94. Martin B, Hanington B. Universal Methods of DesignWays to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions. Gloucester MA: Rockport Publishers; 2012.
95. Creswell JW, Poth CN. Qualitative Inquiry and Research Design: Choosing Among Five Approaches. Washington, DC: SAGE Publications; 2016.
96. DeJonckheere M, Vaughn LM. Semistructured interviewing in primary care research: a balance of relationship and rigour. Fam Med Community Health 2019;7(2):e000057 [FREE Full text] [doi: 10.1136/fmch-2018-000057] [Medline: 32148704]
97. de Souza RH, Dorneles CF. Searchingranking questionnaires: an approach to calculate similarity between questionnaires. New York, NY: Association for Computing Machinery; 2019 Presented at: Proceedings of the ACM Symposium on Document Engineering 2019; 2019 September 26; Berlin, Germany p. 1-9. [doi: 10.1145/3342558.3345390]
98. Coyne I. Children's participation in consultations and decision-making at health service level: a review of the literature. Int J Nurs Stud 2008;45(11):1682-1689. [doi: 10.1016/j.ijnurstu.2008.05.002] [Medline: 18706560]
99. Hassan ZA, Schattner P, Mazza D. Doing a pilot study: why is it essential? Malays Fam Physician 2006;1(2-3):70-73 [FREE Full text] [Medline: 27570591]
100. Schroder C, Medves J, Paterson M, Byrnes V, Chapman C, O'Riordan A, et al. Development and pilot testing of the collaborative practice assessment tool. J Interprof Care 2011;25(3):189-195. [doi: 10.3109/13561820.2010.532620] [Medline: 21182434]

Abbreviations

CBCL: Child Behavior Checklist
CHD: congenital heart disease
CHD-HEEA: Congenital Heart Disease Healthcare Provider Educational Efficiency Assessment
CHD-HLCA: Congenital Heart Disease Health Literacy Children Assessment
CHD-PEBA: Congenital Heart Disease Parental Educational Burden Assessment
CHQ: Child Health Questionnaire
FIM: Family Impact Module
FLLANK: Food Label Literacy for Applied Nutrition Knowledge
KPI: Key Performance Indicator
LSS: Lean Six Sigma
PCQLI: Pediatric Cardiac Quality of Life Inventory
PedsQL: Pediatric Quality of Life Inventory
PSI/PSS: Parenting Stress Index/Parental Stress Scale

Edited by A Mavragani; submitted 30.06.24; peer-reviewed by F Yang, LB Carey; comments to author 13.11.24; revised version received 02.12.24; accepted 02.12.24; published 27.01.25.

Please cite as:
Barbazi N, Shin JY, Hiremath G, Lauff CA
Developing Assessments for Key Stakeholders in Pediatric Congenital Heart Disease: Qualitative Pilot Study to Inform Designing of a Medical Education Toy
JMIR Form Res 2025;9:e63818
URL: https://formative.jmir.org/2025/1/e63818
doi:10.2196/63818
PMID:

©Neda Barbazi, Ji Youn Shin, Gurumurthy Hiremath, Carlye Anne Lauff. Originally published in JMIR Formative Research (https://formative.jmir.org), 27.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effect of the Reassured Self-Compassion–Based School Program on Anxiety, Video Game Addiction, and Body Image Among Rural Female Adolescents: Retrospective Study

Areeg Zuair[1], PhD

Community Health Nursing Department, College of Nursing, Taibah University, Medina, Saudi Arabia

Corresponding Author:
Areeg Zuair, PhD
Community Health Nursing Department
College of Nursing
Taibah University
Prince Naif Ibn Abdulaziz Rd
Medina, 42351
Saudi Arabia
Phone: 966 594800400
Email: Azuair@taibahu.edu.sa

Abstract

Background: The COVID-19 pandemic has amplified mental health challenges among adolescents, particularly in rural areas with limited access to services. In response, the Saudi government launched mental health campaigns and mandated schools to implement mental health programs. However, the effectiveness of these programs remains largely unreported.

Objective: This study aims to determine the prevalence of anxiety disorder symptoms, video game addiction, and body image dissatisfaction, as well as to examine the effect of a school-based program, The Reassured Self, grounded in compassion-focused therapy, on anxiety symptoms, video game addiction, and body image dissatisfaction among rural adolescent females in Saudi Arabia.

Methods: A retrospective secondary analysis of pre-post intervention data was used. All female middle school students (N=133; age: mean 13.7, SD 1.01 years) in grades 1-3 from a rural setting were included, with no exclusion criteria. Participants were recruited as part of a school-mandated mental health program. Data were collected at baseline (2 weeks before the intervention) and 2-3 weeks post intervention during school hours in a quiet classroom setting using teacher-supervised, printed surveys. Survey completion was voluntary, and students exhibiting high distress based on post data analysis were referred to the school health counselor for support. The intervention consisted of 3 sessions (30-35 minutes each) delivered over 2 weeks. Validated Arabic versions of the Spence Children's Anxiety Scale, Game Addiction Scale, and Body Image Discrepancy Assessment were used to measure outcomes 2 weeks post intervention (April to June 2024). Data were analyzed using linear mixed-effects regression models. Data were analyzed using linear mixed effects regression models.

Results: Among the students, 25.8% (32/124) were identified as having high anxiety levels, with 19.3% (24/124) falling into the clinical concern or very high clinical concern categories. Additionally, 26% (32/123) were classified as problem gamers, while 14% (18/123) were categorized as addicted gamers. A smaller proportion, 5.1% (6/117), expressed a strong desire for thinness. Compared to pre-intervention levels, students exhibited significant reductions in anxiety symptoms (mean difference –4.11, 95% CI –6.98 to –1.24; *P*=.005) and video game addiction (mean difference –1.62, 95% CI –2.83 to –0.42; *P*=.009) following the program. However, changes in body image dissatisfaction were minimal and not statistically significant (mean difference 0.067, 95% CI –0.046 to 0.18; *P*=.057). High satisfaction scores, with a mean of 28.49 out of 35 (SD 6.31), indicated strong acceptability and cultural alignment of the intervention. High satisfaction scores indicated strong acceptability and cultural alignment with the intervention.

Conclusions: The results highlight the potential for compassion-focused school programs to address mental health disparities in underserved adolescent populations. Future research should explore the long-term effects and scalability of such culturally adapted interventions.

(JMIR Form Res 2025;9:e68840) doi:10.2196/68840

XSL•FO
RenderX

KEYWORDS

adolescents; rural; compassion-focused therapy; mental health; Saudi Arabia; school

Introduction

The COVID-19 pandemic has profoundly impacted adolescent mental health worldwide, notably increasing anxiety disorder symptoms in this population [1-3]. With societies implementing lockdowns, enforcing social distancing, and facing unprecedented uncertainties, adolescents were exposed to substantial stressors that affected their psychological well-being [1-3]. A meta-analysis across 29 samples and 80,879 youth, conducted post the pandemic, highlighted a global increase in anxiety, with clinically significant symptoms nearly doubling, from 11.6% prepandemic to 20.5% post the pandemic [1]. This surge underscores the pandemic's enduring psychological effects on adolescents' mental health. Adolescence, a crucial developmental phase, often marks the onset of mental health disorders, with 50% emerging by age 14 and 75% by the age of 24 years [4]. Without timely intervention, these conditions can have long-lasting impacts on well-being and life trajectories.

Female adolescents have been disproportionately affected in the post–COVID-19 era, facing higher anxiety prevalence and greater symptom severity, necessitating targeted interventions that address their unique mental health needs [1,5]. Modifiable risk factors for anxiety disorders include game addiction and body image dissatisfaction, both of which significantly affect female adolescents [3,6]. Studies report an increase in gaming addiction during and after the pandemic, with prevalence estimates ranging between 2.3% and 29.4% [6]. Research shows a bidirectional relationship where excessive gaming escalates anxiety, and anxious adolescents use gaming as a coping mechanism [7]. Although many studies focus on male adolescents, emerging evidence indicates that female adolescents are similarly at risk for gaming addiction and associated mental health concerns [8,9]. This highlights the need for gender-specific research and interventions. Reports also indicate an increase in disordered eating and body image issues since the COVID-19 outbreak, partly due to increased exposure to appearance-focused social media and heightened internalization of social pressures [3,5].

In Saudi Arabia, the mental health landscape among adolescents is concerning, with 40.1% experiencing at least one lifetime disorder, 26.84% affected by anxiety, and 7.06% affected by eating disorders [10]. Female adolescents, particularly those in rural areas, are at higher risk for anxiety and body dissatisfaction, with a significantly higher anxiety prevalence than males [11]. For instance, 14.1% (424/3008) of Saudi females are at risk for generalized anxiety disorder, compared to 11.4% (342/3007) of males [11].

Rural female adolescents face additional barriers. Limited access to mental health services and cultural stigma exacerbate untreated mental health disorders, with only 0.5% of at-risk individuals in Saudi receiving treatment [11]. This calls for culturally appropriate, accessible interventions tailored to the specific needs of rural female adolescents in Saudi Arabia. Low-intensity mental health interventions that target anxiety, game addiction, and body image issues are critical to address these unmet needs.

Currently, a gap exists in the Saudi literature regarding school-based mental health and game addiction interventions for adolescents. Despite research on gaming addiction prevalence and related mental health issues, there is limited research on intervention programs, particularly those tailored to Saudi adolescents and female-specific needs [12]. Additionally, comprehensive programs that address both gaming addiction and associated mental health issues like anxiety are scarce, with most research focusing on treatment rather than prevention [8,12]. Also, our search failed to identify any mental health school program among Saudi adolescents. Additionally, there is a notable gap in research on the use of self-compassion interventions in populations outside of Western, educated, industrialized, rich, and democratic countries [13]. In Saudi Arabia, specifically, no reports were found on community- or school-based programs for female adolescents using any mental health intervention methods [10,12-16]. This study aims to fill this gap by providing valuable data on an underserved population: rural female adolescents, who often lack access to mental health services. Designed within the Saudi cultural context, this school-based intervention addresses interconnected issues of anxiety, game addiction, and body image dissatisfaction. The program emphasizes self-reassurance skills, empowering adolescents to manage their mental health independently—an approach particularly relevant in a cultural setting where mental health stigma may discourage individuals from seeking external help [15,17]. By fostering resilience, this intervention addresses critical mental health needs within this underserved population, contributing to the limited research in the area and offering insights that could inform future health policies and practices in Saudi Arabia.

To address these gaps, this study aimed to identify the prevalence of anxiety disorder symptoms, game addiction, and body image dissatisfaction among rural female adolescents; evaluate the efficacy of a compassion-focused therapy (CFT) school-based mental health intervention, The Reassured Self, in reducing anxiety disorder symptoms, game addiction, and body image dissatisfaction among rural female adolescents; and measure the acceptability of The Reassured Self intervention among rural female adolescents.

Methods

Design and Participants

This study used a retrospective design using existing data collected through a predesigned program implemented in a public school setting for quality improvement purposes (see Multimedia Appendix 1 for the checklist used to report the results of this retrospective database study, as recommended by the International Society for Pharmacoeconomics and Outcomes Research Task Force [18]). A pre-post design with no control group was used to evaluate the efficacy of a school-based mental health program, The Reassured Self. The

XSL•FO
RenderX

intervention was conducted with middle school female students in an underserved rural area in the Medina region, Saudi Arabia. Saudi Arabian schools are required to offer health programs to students; however, reports on the efficacy and acceptability of these programs are limited. This study documents a partnership between a non-profit organization and a school to develop a mental health program and evaluates its efficacy and acceptability for future development.

Most rural areas in Saudi Arabia have a single girls' school. This project was implemented for all female middle school students (N=133) in grades 1-3, with no exclusion criteria, as the program was mandatory for all students (see Multimedia Appendix 2 for details on data completeness for intervention participants).

Intervention Program: The Reassured Self

The Reassured Self is a culturally adapted, low-intensity intervention program designed for Saudi Arabian students. Based on the principles of self-compassion, the program integrates cultural and religious values with evidence-based practices to promote mental and emotional resilience. It addresses anxiety, game addiction, and body image dissatisfaction through tailored strategies. The program was delivered to middle school female students in a classroom environment, fostering peer support and shared learning experiences. Each class received three 30- to 35-minute sessions within 2 weeks. To ensure consistency, a trained nurse with experience on delivering the program to other schools delivered the same program to all classes. The intervention content focused on 5 core components, as outlined in Table 1.

Table 1. Detailed overview of the topics and descriptions covered in The Reassured Self intervention program[a].

	Topic	Description
1	Foundational concepts of self-compassion	• Self-blaming vs Self-reassurance: Introduces the difference between self-blame and self-compassion, guiding students to develop a supportive inner dialogue. • Religious integration: Emphasizes Islamic principles written in the Quran and Hadith about compassion, assurance, self-blaming, and gratitude. Morning and evening supplications (Athkar) were incorporated for self-affirmation. • Examples: “O reassured soul, return to your Lord well-pleased and pleasing to Him” (Surah Al-Fajr, 89:27-28) to emphasize fostering inner peace and self-assurance. • “And I swear by the self-blaming soul” (Surah Al-Qiyamah, 75:2) to highlight self-reflection without destructive self-blame. Gratitude practices (Shukr): Students participated in gratitude journaling exercises, drawing on the verse: “If you are grateful, I will certainly give you more” (Surah Ibrahim, 14:7).
2	Building resilience	• Challenge as growth: Teaches students to view challenges as opportunities for growth and resilience by using Quranic verses about patience, effort, and striving (Sabr and Sa’y). For example: “And that there is not for man except that [good] for which he strives” (Surah An-Najm, 53:39).
3	Addressing game addiction	• Understanding dopamine and addiction: Explains how game addiction provides a temporary dopamine boost, leading to a false sense of achievement. • Healthier coping strategies: Encourages real-life connections, exercises, and activities as healthier sources of satisfaction.
4	Reframing bullying	• Understanding bullying behavior: Frames bullying as often stemming from the bully’s own self-blame, helping students empathize and avoid internalizing harmful behaviors. • Responding with resilience: Guides students to respond to negativity with self-respect and compassion, fostering an empathetic school environment.
5	Practical applications and exercises	• Self-compassion exercises: Includes activities like gratitude practices, physical self-care routines, and stress-management tools such as using stress-relief items. • Exercise as self-care: Promotes physical exercise as a form of self-respect and care, enhancing both mental and physical well-being. • Social skills for conflict resolution: Involves role-playing scenarios to provide practical guidance on maintaining self-respect in interactions with others, including handling criticism or bullying.

[a]The program integrates principles of self-compassion and Islamic values to foster emotional resilience, reduce anxiety, address game addiction, and promote healthy coping strategies. Each topic focuses on specific aspects of self-reassurance, resilience building, and practical exercises to enhance mental well-being in a culturally sensitive manner.

Measurements and Data Collection

Data were collected at 2 time points: baseline (2 weeks before the intervention) and postintervention (2-3 weeks after the intervention). Assessments were conducted during school hours in a quiet classroom setting. The teacher distributed the printed surveys (paper-and-pencil format) to students, who were given 45 minutes to complete the questionnaires. The teacher remained available in the classroom to answer any questions. After the follow-up data collection, students exhibiting high levels of psychological distress, identified through their questionnaire responses, were referred to the school health counselor at that time for additional support.

Anxiety Symptoms

To assess clinically significant levels of anxiety symptoms, the Spence Children's Anxiety Scale (SCAS)–Arabic version was administered [19]. The SCAS is a 38-item questionnaire rated

XSL•FO
RenderX

on a 4-point scale (0-3; never-always), with total scores reflecting the sum of responses [20]. This scale encompasses symptoms of the most prevalent *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders* [Fifth Edition]) anxiety disorders and is widely used in both clinical practice and research [21]. The SCAS includes 6 subscales that measure different dimensions of anxiety: separation anxiety, social phobia, obsessive-compulsive symptoms, panic or agoraphobia, physical injury fears, and generalized anxiety. Total anxiety was calculated by summing the scores from the individual subscales. In this study, the SCAS demonstrated excellent internal consistency (α=0.94). *t*-Scores and cutoffs for the SCAS are available on the SCAS website and are categorized by age and gender [22]. For girls aged 12-15 years, a total score cutoff of ≥60 (84th percentile) suggests clinically significant anxiety symptoms.

Video Game Addiction

A newly modified Arabic translation of the 7-item version of the Game Addiction Scale (GAS) was used to measure video game addiction [23]. Responses were given on a 5-point scale (1="never" to 5="very often"), with a total score range of 7-35. The original Arabic version of the GAS served as the basis for this translation [24]. In this study, the GAS-Arabic demonstrated good internal consistency (α=0.81).

Video Game Addiction Status

To classify participants based on their level of video game engagement, responses to the core 4-item criteria for addiction (withdrawal, relapse, conflict, and problems) were evaluated according to methods recommended in recent literature [25]. Participants were divided into 3 groups: addicted gamers, problem gamers, and engaged gamers. Respondents who scored 3 or higher on all 4 items were classified as Addicted Gamers. Those who scored 3 or higher on 2 or 3 of the 4 items were categorized as problem gamers, indicating some signs of problematic gaming behavior. Respondents who did not meet the criteria for either addicted or problem gamers were classified as engaged gamers, reflecting low engagement without signs of problematic behavior.

Body Image Dissatisfaction

Body image dissatisfaction was assessed using a series of silhouettes representing 4 different body sizes [26]. Students selected the silhouette that best represented their ideal body size and the one that most resembled their current body size. The difference between these selections (ideal vs current) indicated the level of body image discrepancy, with scores ranging from –3 to +3. A negative score suggested a desire for thinness (preference for a smaller body size), while a positive score indicated a desire for fullness (preference for a larger body size) [27]. Participants were categorized as follows: high drive for thinness (scores less than –1.5), low drive for thinness (scores between –1.5 and 0), and no drive for thinness (scores of 0 or higher).

Program Reaction

Participant feedback was collected using the Kirkpatrick model's level 1 evaluation 2-3 weeks post the program. The level 1 evaluation, also known as Reaction, assesses participants' immediate responses to a training or intervention [28]. It captures how participants felt about the experience, evaluating satisfaction, perceived ease and clarity, engagement (perceived effectiveness), and perceived relevance of the content [28]. A 7-item Likert scale was used, with responses ranging from 1 (totally disagree) to 5 (totally agree), resulting in a total score range of 7-35.

Statistical Analysis

Overview

Descriptive statistics (means, SDs, and frequencies) were calculated for all variables. A series of linear mixed models were used to examine the effects of the intervention on 3 dependent variables—anxiety symptoms, video game addiction, and body image discrepancy—measured at 2 time points (before and post the intervention). The models included time as a fixed effect and student ID as a random effect to account for within-subject variability across repeated measures. Restricted maximum likelihood estimation was used to fit the models. Grade levels were tested as fixed effects, but neither demonstrated significant associations with the outcomes or interactions with time. Consequently, these variables were excluded from the model. Pairwise comparisons of the estimated marginal means were conducted using least significant difference adjustment. The assumption of homoscedasticity (constant variance) appeared mostly satisfied, and the residuals were approximately normally distributed. Linearity between time and each dependent variable was reasonable, as no strong nonlinear patterns were observed. Also, correlation analyses were conducted to identify potential confounding variables, including age, baseline anxiety, body image discrepancy, and gaming addiction.

Participants with valid data on anxiety symptoms, video game addiction, or body image discrepancy at either baseline or follow-up were included in the analysis. Linear mixed models use all available data, regardless of missing outcome data, under the assumption that data are missing at random. Data were found to be missing completely at random, as indicated by Little's MCAR test (χ^2_{135}=110.5, *P*=.94). No cases were missing more than 75% of data, which is the recommended threshold for analyzing the SCAS.

Power Analysis

A post hoc power analysis was conducted using G*Power (version 3.1.9.7) to evaluate the achieved power for detecting changes in anxiety levels, game addiction, and body image discrepancy across pre- and postintervention time points. The paired samples *t* test was selected as the appropriate statistical test for the pre-post design, focusing on differences between dependent means. Effect sizes (dz) were calculated as 0.25, 0.24, and 0.076 based on the observed mean differences and SDs of the difference scores for anxiety, game addiction, and body image discrepancy, respectively.

The analysis was performed with an alpha level of 0.05 and total sample sizes of 133, 132, and 130 participants for anxiety, game addiction, and body image discrepancy, respectively. The results indicated that the study had an 89% and 86% probability of detecting a true effect for anxiety and game addiction,

XSL•FO
RenderX

respectively. However, the post hoc power for body image discrepancy was approximately 13.8%, indicating that the study had very low power to detect such a small effect. This suggests that the study may not have been adequately powered to identify significant changes in body image discrepancy.

Ethical Considerations

Given the retrospective nature of the study, all students participated in the program as it was mandated by the school health unit as part of the standard curriculum. Written consent was obtained from students specifically for their participation in completing the surveys associated with the program. Information about the surveys, including their purpose and voluntary nature, was communicated to the students in an age-appropriate manner. This included verbal explanations by teachers and written summaries that were easy to understand. Students were assured that completing the surveys was voluntary and would not affect their grades or school standing. To encourage participation, all students received a small gift (a relaxation squeeze toy) after completing the program satisfaction survey. Parental consent was not required, as the program itself was a required school activity, and the decision to waive parental consent for the surveys followed ethical guidelines for school-based health programs. This approach was approved by the University of Taibah Nursing College Ethical Review Committee (IRB No. TUCN-REC 11 2024). To ensure the privacy and confidentiality of participants, all data from school project records were securely stored and deidentified prior to analysis.

Results

Students' Characteristics

Table 2 presents the descriptive statistics for the study variables, including age, anxiety symptoms, video game addiction, and body image discrepancy. A total of 133 students participated in the program. However, only 93% (124/133) completed the preanxiety symptoms survey, 92% (123/133) completed the video game addiction, and 88% (117/133) completed the body image discrepancy assessment (see Multimedia Appendix 2 for details on data completeness for intervention participants). The mean age of participants was 13.72 (SD 1.01) years. Total anxiety symptoms had a mean score of 47.60 (SD 22.02), indicating a moderate level of anxiety across the sample. Among the anxiety subscales, panic or agoraphobia symptoms were the highest, with a mean of 9.38 (SD 6.76), followed by obsessive-compulsive symptoms (mean 8.79, SD 4.54) and generalized anxiety symptoms (mean 8.66, SD 4.55). Physical injury fears (mean 7.64, SD 3.75) and separation anxiety (mean 6.31, SD 3.88) were also prominent. Social phobia symptoms had a mean of 6.82 (SD 4.12), indicating a relatively lower prevalence compared to other anxiety types.

Table 2. Descriptive statistics for age, anxiety symptoms, video game addiction, and body image discrepancy among rural adolescent female students[a].

Descriptive statistics	Mean (SD) values
Age (years; n=133)	13.72 (1.01)
Total anxiety symptoms (n=124)	47.60 (22.02)
Generalized anxiety symptoms	8.66 (4.56)
Physical injury fears symptoms	7.64 (3.75)
Separation anxiety symptoms	6.31 (3.88)
Panic agoraphobia symptoms	9.38 (6.76)
Social phobia symptoms	6.82 (4.12)
Obsessive-compulsive symptoms	8.79 (4.54)
Video game addiction (n=123)	15.57 (6.67)
Body image discrepancy (n=117)	–0.197 (0.87)

[a]Descriptive statistics summarizing the mean and SD for age, anxiety symptoms, video game addiction, and body image discrepancy among rural adolescent female students who participated in The Reassured Self intervention program.

The mean score for video game addiction was 15.57 (SD 6.67), indicating a moderate level of engagement or potential addiction to gaming. The body image discrepancy score had a mean of –0.20 (SD 0.87), suggesting a slight drive for thinness among the sample, with negative values indicating a preference for a thinner body size compared to current perception. Correlations between baseline variables showed a significant, moderate positive relationship between anxiety and gaming addiction (r=0.457, *P*<.001). Age and body image discrepancy demonstrated weak, nonsignificant correlations with other variables.

Characterizing Mental Health Status

The distribution of anxiety symptoms among students reveals that a substantial portion experience elevated or clinically concerning levels across various anxiety types (Table 3). Specifically, 25.8% (32/124) of students had a high anxiety level, and 19.3% (24/124) fall within the clinical concern or very high clinical concern category, indicating that a significant subset may require attention. Generalized anxiety is particularly prevalent, with 33.1% (41/124) of students classified in the clinical concern or very high clinical concern categories. In addition, fears related to physical injury show one of the most severe patterns, with 33.9% (42/124) of students in the very high clinical concern category, highlighting this as a critical

XSL•FO
RenderX

area for mental health support. Separation anxiety also shows a high prevalence, with 33.9% (42/124) of students in the clinical or very high clinical concern categories. Symptoms of panic and agoraphobia are notable as well, with 28.2% (35/124) of students in the very high clinical concern category, indicating a significant impact. Obsessive-compulsive symptoms present another area of concern, with 21% (26/124) of students in the very high clinical concern category. In contrast, social phobia has a lower prevalence of severe concern, with only 2.4% (3/124) of students in the very high clinical concern category.

Table 3. Frequency and percentage distribution of anxiety levels, video game status, and body image dissatisfaction among rural adolescent female students.

Anxiety levels (N=124)[a]	Normal range	Elevated anxiety	Clinical concern	Very high clinical concern
Anxiety symptoms, n (%)	92 (74.2)	8 (6.5)	5 (4)	19 (15.3)
Generalized anxiety, n (%)	61 (49.2)	22 (17.7)	25 (20.2)	16 (12.9)
Physical injury fears, n (%)	38 (30.6)	20 (16.1)	24 (19.4)	42 (33.9)
Separation anxiety, n (%)	55 (44.4)	27 (21.8)	16 (12.9)	26 (21)
Panic and agoraphobia, n (%)	54 (43.5)	18 (14.5)	17 (13.7)	35 (28.2)
Social phobia, n (%)	93 (75)	12 (9.7)	16 (12.9)	3 (2.4)
Obsessive-compulsive, n (%)	38 (30.6)	30 (24.2)	30 (24.2)	26 (21)
Video game status (N=123)[a], n (%)	Engaged gamers: 73 (59)	Problem gamers: 32 (26)	Addicted gamers: 18 (14)	Addicted gamers: 18 (14)
Body image dissatisfaction (N=117)[a], n (%)	No drive for thinness: 77 (65.8)	Low drive for thinness: 34 (29.1)	High drive for thinness: 6 (5.1)	High drive for thinness: 6 (5.1)

[a]Total number of completed surveys.

Overall, the findings suggest that while some anxiety symptoms present moderate concern, specific areas—particularly physical injury fears, separation anxiety, and obsessive-compulsive symptoms—are more prevalent at clinically concerning levels, underscoring the need for focused mental health support in these domains.

Regarding body image dissatisfaction, the results indicate a slight drive for thinness among most of the students. Approximately 29.1% (34/117) of students exhibit a moderate inclination toward weight reduction, while a smaller proportion, 5.1% (6/117), express a strong desire for thinness.

Characterizing Video Game Status

The data on video game status among students reveals varying levels of gaming involvement, with a significant portion showing signs of problematic or addictive behavior (Table 3). Overall, 59% (73/123) of the students are categorized as engaged gamers, indicating regular gaming without notable adverse effects. However, 26% (32/123) are identified as problem gamers, suggesting a level of gaming that may interfere with daily responsibilities or personal well-being. Additionally, 14% (18/123) of students fall into the addicted gamers category, indicating a more severe level of dependency that likely impacts various aspects of their lives.

Effectiveness of The Reassured Self Program

The Reassured Self intervention demonstrated a positive effect in reducing anxiety levels among students, as shown by a significant decrease in anxiety scores from preintervention to follow-up (see Table 4). The effect size, with a mean difference of –4.11, was small but statistically significant (*P*=.005). Additionally, the intervention had a beneficial impact on reducing video game addiction, with an estimated mean reduction of –1.62 in game addiction scores, which was also statistically significant (*P*=.009). In contrast, although the intervention led to a slight decrease in the drive for thinness, this change was not statistically significant (*P*=.057), suggesting a minimal effect on body image dissatisfaction. Overall, these findings indicate that the intervention was effective in reducing both anxiety and game addiction among students, while its impact on body image discrepancy remains inconclusive.

Table 4. Changes in anxiety symptoms, video game addiction, and body image dissatisfaction from baseline to follow-up[a].

Outcome	Baseline, mean (SE) [95% CI]	Follow-up, mean (SE) [95% CI]	Within group differences change from follow-up–baseline	
			Mean difference [95% CI]	*P* value
Anxiety symptoms (n=133)	47.137 (1.93) [43.03 to 50.95]	43.32 (2.13) [38.81 to 47.25]	–4.11 [–6.98 to –1.24]	.005
Video game addiction (n=132)	15.45 (0.59) [14.28 to 16.63]	13.83 (0.61) [12.63 to 15.03]	–1.62 [–2.83 to –0.42]	.009
Body image dissatisfaction (n=130)	–0.19 (0.08) [–0.35 to 0.038]	–0.125 (0.07) [–0.27 to 0.02]	0.067 [–0.046 to 0.18]	.057

[a]This table presents the results of linear mixed-effects models, which accounted for repeated measures within individuals, to evaluate changes in anxiety symptoms, video game addiction, and body image dissatisfaction from baseline to follow-up among participants of The Reassured Self intervention program. Adjustments for multiple comparisons were made using the least significant difference method.

XSL•FO
RenderX

Program Reaction

The program received positive feedback across multiple dimensions, as reflected in the high mean scores for various aspects of participant satisfaction (Table 5). The overall satisfaction sum score ranged from 7 to 35, with a mean of 28.49 (SD 6.31), indicating a generally favorable response. The overall results suggest that participants responded positively to the program, finding it engaging, clear, and psychologically beneficial, with a strong likelihood to recommend it to others.

Table 5. Descriptive statistics for reaction measures of The Reassured Self program (N=123)[a].

Measure	Mean (SD) values
Overall satisfaction score	28.49 (6.31)
Perceived ease and clarity	4.41 (0.88)
Enjoyment and satisfaction	4.16 (1.14)
Psychological benefits gained	4.14 (1.11)
Program relevance and value	3.9 (1.29)
Teaching effectiveness	3.97 (1.13)
Engagement and attention retention	4.15 (1.09)
Likelihood of recommending the program	4.2 (1.20)

[a]This table presents the mean scores and SDs for various aspects of participant satisfaction with The Reassured Self program. The consistently high scores across all measures reflect positive feedback and strong acceptability of the intervention.

Discussion

Principal Findings

This study conducted a retrospective analysis of data collected from female adolescents who participated in a CFT-based school program, The Reassured Self, in a rural setting. The program, collaboratively implemented by a nonprofit organization and an academic institution, aimed to reduce anxiety, video game addiction, and body image dissatisfaction among rural Saudi female adolescents, addressing critical mental health challenges within this population. The findings revealed a high prevalence of anxiety and video game addiction, consistent with national averages. The intervention effectively reduced anxiety and video game addiction symptoms but showed no improvement in body image dissatisfaction, highlighting its potential as a health education tool.

Anxiety Reduction

The overall high anxiety symptom prevalence of 25.8% (32/124) and the high severe anxiety symptoms of 19.3% (24/124) in our sample are comparable to global postpandemic figures (20.5%) [1] and align closely with the recent national prevalence of lifetime anxiety disorders among adolescents (26.84%) [10]. Our results underscore the urgent need for interventions targeting adolescent anxiety in underserved areas. Recent literature argues that school well-being programs should consider children's and adolescents' anxiety levels as a key metric for measuring program impact [29]. The Reassured Self program demonstrated a significant reduction in anxiety symptoms, suggesting that a culturally adapted CFT approach can effectively address anxiety. This aligns with recent meta-analyses and systematic reviews, which consistently show that compassion-based interventions reduce anxiety and psychological distress among adolescents [30-32].

This study's findings also align with recent evidence on compassion-based school interventions. For instance, Maratos et al [29] evaluated a Compassionate Mind Training (CMT-Pupils) intervention among Year 7 students aged 11-12 and found significant improvements in anxiety levels compared to the control group. Similar to our intervention, CMT-Pupils emphasized emotional regulation and prosocial behaviors through compassion-based practices, including mindfulness and self-compassion. Their study highlighted the importance of promoting kindness, emotional awareness, and classroom inclusivity, which aligns with the core mechanisms of The Reassured Self program.

Further supporting the importance of self-compassion in adolescent mental health, Hammad et al [16] conducted a cross-sectional study in Saudi Arabia that explored the relationship between perceived stress and self-compassion among adolescents during the COVID-19 pandemic. Their findings revealed a significant negative correlation (r=–0.460, P<.001) between perceived stress and self-compassion levels, suggesting that higher self-compassion is associated with lower levels of stress. These results underscore the role of self-compassion as a protective factor for adolescent mental health, reinforcing the theoretical foundation of The Reassured Self program, which aims to reduce anxiety by fostering self-compassion and emotional regulation. CFT is rooted in evolutionary psychology and attachment theory, incorporating mindfulness through activities designed to foster self-compassion toward oneself and others, particularly in relation to self-criticism, self-blame, and shame [17,31,33]. Recommended techniques and exercises to develop self-compassion include self-talk and understanding the evolved nature of the human mind, along with the origins of self-criticism and self-blame [17]. Therefore, the core focus of The Reassured Self program is to introduce students to the concepts of self-blaming and self-reassurance, fostering a supportive inner dialogue and emotional regulation.

However, while The Reassured Self program demonstrated reductions in anxiety, our study did not directly measure

XSL•FO
RenderX

emotional regulation or self-compassion processes—key mechanisms observed in Petrocchi et al [33] and other self-compassion studies. Tools such as the Forms of Self-Criticizing/Attacking and Self-Reassuring Scale, developed by Gilbert et al [34] (the founder of CFT), provide validated measures for assessing self-criticism and self-reassurance. Incorporating these tools in future studies would enable a more comprehensive evaluation of how The Reassured Self program works. Specifically, measuring changes in emotional regulation, self-reassurance, and self-criticism could elucidate the mechanisms driving the reduction in anxiety and gaming addiction. As literature suggests, a decrease in self-criticism and self-blame is strongly associated with a reduction in anxiety [17,31,33]. Including these intermediate outcomes will help validate the program's effectiveness and provide deeper insights into its psychological processes. Overall, our findings suggest that culturally relevant, compassion-based programs can provide a viable solution for reducing anxiety in rural, underserved contexts. By addressing both the reduction of anxiety symptoms and fostering emotional regulation, such interventions hold promise as scalable and effective mental health solutions for adolescents. Also, our approach is supported by recent findings that advocate for brief, targeted interventions to effectively engage adolescents, particularly in an era where social media usage has impacted attention spans [35]. A recent meta-analysis of brief school mental health interventions found that programs lasting fewer than 4 sessions or 250 minutes resulted in significant, albeit small, improvements in students' mental health when assessed 1 month, 6 months, and 1 year post intervention [35].

Reduction in Video Game Addiction

The high baseline prevalence of video game addiction (14.6%, 18/123) among rural female adolescents in our study underscores the need for targeted mental health interventions in this area. This prevalence is close to global averages, which show approximately 12% of adolescents are at risk for problematic gaming, highlighting the specific challenges faced by this population [36]. However, the prevalence in our study is lower than that in the recent national report for the Madinah region (28%), and the overall national prevalence of 20.45% (897/4431) [37].

The program’s focus on self-compassion and emotional resilience appears to have positively impacted gaming behaviors, as evidenced by a significant decrease in video game addiction scores. Furthermore, the significant correlation between anxiety and gaming addiction underscores their interconnectedness, aligning with existing research that highlights gaming as a coping mechanism for anxiety [7].

This finding supports research suggesting that interventions emphasizing emotional regulation and coping skills such as exercise can reduce problematic gaming behaviors [38]. For example, a recent study investigating the relationship between internet gaming disorder, sleep quality, self-compassion, physical activity, and psychological distress among young adults aged 18-24 years found that internet gaming disorder was positively associated with interpersonal stress and depressive symptoms. Additionally, self-compassion and exercise behavior were found to mediate the relationship between depressive symptoms and sleep quality [39].

By providing alternative coping mechanisms, The Reassured Self may help reduce adolescents' reliance on gaming to manage emotional distress, addressing a critical mental health concern in this demographic. However, as discussed earlier, the lack of measurement of changes in emotional regulation, or adapting exercise behaviors, which could be assessed using validated tools, limits the conclusions that can be drawn about the program's overall efficacy.

Body Image Dissatisfaction

The notable prevalence of body image dissatisfaction, with 5.1% (6/117) of students showing a strong drive for thinness, reveals another area for potential intervention. Although the program’s impact on body image dissatisfaction was minimal and nonsignificant, the slight reduction in drive for thinness suggests that longer or more targeted interventions may more effectively address body image issues. Body image is deeply influenced by social and cultural factors, and this program provides a foundation for future interventions to build upon by incorporating more specific strategies to address body image concerns in rural contexts. While CFT has shown success in improving body image in other studies [30], the short duration of our program may have limited its impact on this complex issue. Additionally, this study found an average of 1 size discrepancy among the students, aligning with other research indicating that rural students may have low body image discrepancy [40,41]. Future iterations of the program may need to incorporate more targeted components to address body image issues specific to the cultural context.

Program Reaction

The integration of Islamic values and culturally sensitive practices, such as gratitude and self-compassion, likely enhanced the intervention's acceptability and effectiveness. Student feedback indicated high levels of satisfaction regarding the program's ease and clarity, perceived effectiveness, and content relevance. The concept of self-compassion, rooted in Islamic culture and religious practice, fosters self-reassurance and reduces self-blame [16]. Cultural and contextual considerations played a crucial role in the program’s design and implementation, which aligned with current recommendations for culturally adapting mental health interventions through self-compassion [17]. In addition, the program's success in a rural setting addresses a critical gap in mental health services for underserved populations and supports the need for global mental health interventions [42,43].

Limitations and Recommendations

This study’s findings have important implications for designing school-based mental health programs in rural or underserved areas. To ensure the effective delivery and sustainability of such interventions, several strategies should be prioritized: (1) training mental health professionals, educators, and community facilitators is critical for delivering the intervention effectively. (2) Partnerships with local organizations and schools can enhance resource availability and facilitate recruitment in resource-limited settings. (3) Leveraging community

XSL•FO
RenderX

partnerships: the collaboration demonstrated in this study—between a nonprofit organization, an academic institution, and a local school—provides a viable model for delivering mental health interventions in underserved areas. This approach aligns with broader recommendations to leverage community partnerships to enhance mental health support in schools [43,44]. (4) Advocacy efforts should focus on investments in culturally adapted, school-based mental health interventions to ensure they can reach underserved populations and address mental health disparities in context-specific and meaningful ways. (5) Future interventions should explore digital education methods, such as webinars, web-based workshops, and gaming-based mental health curricula, to enhance accessibility and engagement, particularly among rural students [45,46]. (6) Future research should investigate the intervention's long-term effects, including its impact on academic performance (eg, grades, school engagement) and social functioning (eg, relationships with peers, family members, and teachers). Understanding whether mental health improvements translate into better academic and social outcomes would provide a more holistic evaluation of the program's effectiveness. (7) Further exploration of specific cultural factors, such as family dynamics, social stigma, and rural-urban differences, is necessary to understand their contributions to the program's effectiveness. Adapting the intervention for broader cultural and socioeconomic contexts is critical for expanding its applicability.

However, several limitations should be considered when interpreting these results. The retrospective design and lack of a control group limit causal inferences. Additionally, the short follow-up period (2-3 weeks) may not capture long-term effects or potential relapses [17]. The specific demographic focus on rural Saudi female adolescents may also limit generalizability to other populations. To address these limitations and build on this study, future research should (1) incorporate randomized controlled trial designs with active control groups to establish causality; (2) extend follow-up periods to evaluate the sustainability of intervention benefits over time; (3) compare The Reassured Self program with other evidence-based approaches, such as cognitive-behavioral therapy and mindfulness-based stress reduction, to assess relative effectiveness and adaptability; (4) examine the specific components of the program and their relative contributions to outcomes to refine and optimize the intervention; and (5) investigate the interplay between cultural factors and program outcomes to ensure effective adaptations for different cultural and socioeconomic contexts.

Conclusions

Adolescents in rural Saudi Arabia face significant mental health challenges, with high rates of anxiety, video game addiction, and, to a lesser extent, body image dissatisfaction. This study's findings underscore the potential of a culturally adapted CFT program, The Reassured Self, to address these issues. Collaboratively implemented by a non-profit organization and an academic institution, the program successfully reduced anxiety symptoms and problematic gaming behaviors among rural female adolescents, providing a promising model for school-based mental health interventions in underserved areas.

The integration of Islamic values and culturally sensitive practices, such as gratitude and self-compassion, likely enhanced the program's acceptability and relevance, as reflected in the high levels of student satisfaction with the intervention's content and delivery. The Reassured Self demonstrates the feasibility and effectiveness of a culturally sensitive, compassion-focused intervention for adolescent mental health in rural Saudi Arabia. The promising findings of The Reassured Self program highlight its potential for widespread implementation in schools and communities, particularly in underserved areas. By addressing the unique challenges faced by this underserved population, the program contributes valuable insights for global mental health initiatives aimed at reducing disparities in mental health access and support. Future efforts to refine and scale the intervention can inform broader applications in diverse cultural and socioeconomic contexts.

Acknowledgments

The author would like to express sincere gratitude to the Ministry of Education School Health Department in Medina and the participating school for their invaluable support.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Checklist for Reporting Results of Retrospective Database Studies from the ISPOR Task Force.
[PDF File (Adobe PDF File), 54 KB - formative_v9i1e68840_app1.pdf]

Multimedia Appendix 2
Data Completeness for Intervention Participants.
[PDF File (Adobe PDF File), 36 KB - formative_v9i1e68840_app2.pdf]

XSL•FO
RenderX

References

1. Racine N, McArthur BA, Cooke JE, Eirich R, Zhu J, Madigan S. Global prevalence of depressive and anxiety symptoms in children and adolescents during COVID-19: a meta-analysis. JAMA Pediatr 2021;175(11):1142-1150 [FREE Full text] [doi: 10.1001/jamapediatrics.2021.2482] [Medline: 34369987]
2. Triplett NS, Woodard GS, Johnson C, Nguyen JK, AlRasheed R, Song F, et al. Stakeholder engagement to inform evidence-based treatment implementation for children's mental health: a scoping review. Implement Sci Commun 2022;3(1):82. [doi: 10.1186/s43058-022-00327-w] [Medline: 35906675]
3. Rodrigue C, Rodgers RF, Carbonneau N, Bégin C, Dion J. COVID-19-related distress, body image, and eating behaviors: a cross-sectional explanatory model. BMC Psychol 2024;12(1):117 [FREE Full text] [doi: 10.1186/s40359-024-01613-z] [Medline: 38431696]
4. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. Arch Gen Psychiatry 2005;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
5. Firoozjah MH, Shahrbanian S, Homayouni A, Hower H. Comparison of eating disorders symptoms and body image between individual and team sport adolescent athletes during the COVID-19 pandemic. J Eat Disord 2022;10(1):119 [FREE Full text] [doi: 10.1186/s40337-022-00644-4] [Medline: 35962428]
6. Han TS, Cho H, Sung D, Park MH. A systematic review of the impact of COVID-19 on the game addiction of children and adolescents. Front Psychiatry 2022;13:976601 [FREE Full text] [doi: 10.3389/fpsyt.2022.976601] [Medline: 36061296]
7. Kaya A, Türk N, Batmaz H, Griffiths MD. Online gaming addiction and basic psychological needs among adolescents: the mediating roles of meaning in life and responsibility. Int J Ment Health Addict 2023:1-25 [FREE Full text] [doi: 10.1007/s11469-022-00994-9] [Medline: 36643385]
8. Bumozah HS, Al-Quwaidhi AJ, Al-Ghadeeb R. Prevalence and risk factors of internet gaming disorder among female secondary school students in Al-Ahsa, Kingdom of Saudi Arabia. Cureus 2023;15(6):e40375 [FREE Full text] [doi: 10.7759/cureus.40375] [Medline: 37456410]
9. Alghamdi MH, Alghamdi MM. Prevalence of internet gaming disorder among intermediate and high school students in Albaha, Saudi Arabia: a cross-sectional study. Cureus 2023;15(4):e37115 [FREE Full text] [doi: 10.7759/cureus.37115] [Medline: 37153233]
10. Altwaijri Y, Kazdin AE, Al-Subaie A, Al-Habeeb A, Hyder S, Bilal L, et al. Lifetime prevalence and treatment of mental disorders in Saudi youth and adolescents. Sci Rep 2023;13(1):6186 [FREE Full text] [doi: 10.1038/s41598-023-33005-5] [Medline: 37061556]
11. Alhabeeb AA, Al-Duraihem RA, Alasmary S, Alkhamaali Z, Althumiri NA, BinDhim NF. National screening for anxiety and depression in Saudi Arabia 2022. Front Public Health 2023;11:1213851 [FREE Full text] [doi: 10.3389/fpubh.2023.1213851] [Medline: 37441650]
12. Basurrah AA, Di Blasi Z, Lambert L, Murphy M, Warren MA, Setti A, et al. The effects of positive psychology interventions in Arab countries: a systematic review. Appl Psychol Health Well Being 2023;15(2):803-821. [doi: 10.1111/aphw.12391] [Medline: 35856920]
13. Kotera Y, Kirkman A, Beaumont J, Komorowska MA, Such E, Kaneda Y, et al. Self-compassion during COVID-19 in non-WEIRD countries: a narrative review. Healthcare (Basel) 2023;11(14):2016 [FREE Full text] [doi: 10.3390/healthcare11142016] [Medline: 37510457]
14. Mariee AA, Abd Elhamid EA, Al-Dhmashi ASF, Alshahrani FF, El-Said HDA. Integrating of mental health interventions in school health in Saudi Arabia: policy brief. Int J Nov Res Healthc Nurs 2024;11(1):303-306. [doi: 10.5281/zenodo.10868098]
15. Chatterjee K, Srivastava K, Prakash J, Dangi A. Positive mental health for all serving the under-served. Ind Psychiatry J 2023;32(2):234-239 [FREE Full text] [doi: 10.4103/ipj.ipj_132_22] [Medline: 38161446]
16. Hammad SS, Alzhrani MD, Almulla HA. Adolescents' perceived stress of COVID-19 and self-compassion in Saudi Arabia: a cross-sectional study. Int J Nurs Sci 2023;10(2):215-220 [FREE Full text] [doi: 10.1016/j.ijnss.2023.03.008] [Medline: 37095851]
17. Neff KD. Self-compassion: theory, method, research, and intervention. Annu Rev Psychol 2023;74(1):193-218 [FREE Full text] [doi: 10.1146/annurev-psych-032420-031047] [Medline: 35961039]
18. Motheral B, Brooks J, Clark MA, Crown WH, Davey P, Hutchins D, et al. A checklist for retrospective database studies--report of the ISPOR task force on retrospective databases. Value Health 2003;6(2):90-97 [FREE Full text] [doi: 10.1046/j.1524-4733.2003.00242.x] [Medline: 12641858]
19. Ali N, Snell T. Spence children's anxiety scale ? Arabic. 1994. URL: https://www.scaswebsite.com/arabic/ [accessed 2025-02-05]
20. Spence SH. Structure of anxiety symptoms among children: a confirmatory factor-analytic study. J Abnorm Psychol 1997;106(2):280-297. [doi: 10.1037/0021-843x.106.2.280]
21. Galán-Luque T, Serrano-Ortiz M, Orgilés M. Factor structure and psychometric properties of the spence children's anxiety scale: a 25-Year systematic review. Child Psychiatry Hum Dev 2023. [doi: 10.1007/s10578-023-01566-1] [Medline: 37490107]

XSL•FO
RenderX

22. Spence SH. Using the SCAS-Child T-Scores. URL: https://www.scaswebsite.com/portfolio/scas-child-t-scores/ [accessed 2025-02-05]
23. Lemmens JS, Valkenburg PM, Gentile DA. The internet gaming disorder scale. Psychol Assess 2015;27(2):567-582. [doi: 10.1037/pas0000062] [Medline: 25558970]
24. Asaad T, Morsy KE, Hasan H, El Din HM, El Meguid MA. Game addiction scale for adolescents: Arabic version validation. Addict Disord Their Treat 2019;18(6):223-228. [doi: 10.1097/adt.0000000000000167]
25. Brunborg GS, Hanss D, Mentzoni RA, Pallesen S. Core and peripheral criteria of video game addiction in the game addiction scale for adolescents. Cyberpsychol Behav Soc Netw 2015;18(5):280-285 [FREE Full text] [doi: 10.1089/cyber.2014.0509] [Medline: 25826043]
26. Simeon DT, Rattan RD, Panchoo K, Kungeesingh KV, Ali AC, Abdool PS. Body image of adolescents in a multi-ethnic caribbean population. Eur J Clin Nutr 2003;57(1):157-162. [doi: 10.1038/sj.ejcn.1601515] [Medline: 12548311]
27. Latiff AA, Muhamad J, Rahman RA. Body image dissatisfaction and its determinants among young primary-school adolescents. J Taibah Univ Med Sci 2018;13(1):34-41 [FREE Full text] [doi: 10.1016/j.jtumed.2017.07.003] [Medline: 31435300]
28. Kirkpatrick D, Kirkpatrick J. Evaluating Training Programs: The Four Levels. San Francisco: Berrett-Koehler Publishers; 2006.
29. Maratos FA, Wood W, Cahill R, Tronco Hernández YAT, Matos M, Gilbert P. A mixed-methods study of compassionate mind training for pupils (CMT-Pupils) as a school-based wellbeing intervention. Mindfulness 2024;15(2):459-478. [doi: 10.1007/s12671-024-02303-y]
30. Turk F, Waller G. Is self-compassion relevant to the pathology and treatment of eating and body image concerns? A systematic review and meta-analysis. Clin Psychol Rev 2020;79:101856. [doi: 10.1016/j.cpr.2020.101856] [Medline: 32438284]
31. Vidal J, Soldevilla JM. Effect of compassion-focused therapy on self-criticism and self-soothing: a meta-analysis. Br J Clin Psychol 2023;62(1):70-81 [FREE Full text] [doi: 10.1111/bjc.12394] [Medline: 36172899]
32. Tali E, Potharst ES, de Bruin EI, Utens EMWJ. Self-compassion and anxiety in adolescents with and without anxiety disorder. Children (Basel) 2023;10(7):1181 [FREE Full text] [doi: 10.3390/children10071181] [Medline: 37508678]
33. Petrocchi N, Ottaviani C, Cheli S, Matos M, Baldi B, Basran JK, et al. The impact of compassion-focused therapy on positive and negative mental health outcomes: results of a series of meta-analyses. Clin Psychol Sci Pract 2024;31(2):230-247. [doi: 10.1037/cps0000193]
34. Gilbert P, Clarke M, Hempel S, Miles J, Irons C. Criticizing and reassuring oneself: an exploration of forms, styles and reasons in female students. Br J Clin Psychol 2004;43(Pt 1):31-50. [doi: 10.1348/014466504772812959] [Medline: 15005905]
35. Cohen KA, Ito S, Ahuvia IL, Yang Y, Zhang Y, Renshaw TL, et al. Brief school-based interventions targeting student mental health or well-being: a systematic review and meta-analysis. Clin Child Fam Psychol Rev 2024;27(3):732-806. [doi: 10.1007/s10567-024-00487-2] [Medline: 38884838]
36. Boniel-Nissim M, Tynjälä J, Gobiņa I, Furstova J, van den Eijnden RJJM, Marino C, et al. Adolescent use of social media and associations with sleep patterns across 18 European and North American countries. Sleep Health 2023;9(3):314-321 [FREE Full text] [doi: 10.1016/j.sleh.2023.01.005] [Medline: 36804326]
37. Alfaleh A, Alzaher A, Alkattan A, Alabdulkareem K, Ibrahim MH. Prevalence of video gaming disorder in Saudi Arabia: a school-based national study. J Egypt Public Health Assoc 2024;99(1):20 [FREE Full text] [doi: 10.1186/s42506-024-00165-9] [Medline: 39164516]
38. King DL, Delfabbro PH, Perales JC, Deleuze J, Király O, Krossbakken E, et al. Maladaptive player-game relationships in problematic gaming and gaming disorder: a systematic review. Clin Psychol Rev 2019;73:101777. [doi: 10.1016/j.cpr.2019.101777] [Medline: 31707185]
39. Wong MYC, Yuan GF, Liu C, Lam SKK, Fung HW. The relationship between internet gaming disorder, sleeping quality, self-compassion, physical activity participation and psychological distress: a path analysis. Glob Ment Health (Camb) 2024;11:e67. [doi: 10.1017/gmh.2024.36] [Medline: 39220110]
40. Błaszczyk-Bębenek E, Jagielski P, Nowak R, Ostachowska- Gąsior A. Dissatisfaction and discrepancy in body-image perceptions among adolescents aged 10–17 years from rural areas of southern poland – a cross-sectional study. Ann Agric Environ Med 2024. [doi: 10.26444/aaem/193461]
41. Paria B, Dwivedi M, Roy SK. A cross-sectional study on body image perception and self-esteem among adolescent girls in urban and rural Areas of Kolkata, West Bengal, India. J Clin Diagn Res 2023:LC11-LC14. [doi: 10.7860/jcdr/2023/62667.18697]
42. Barnett ML, Gonzalez A, Miranda J, Chavira DA, Lau AS. Mobilizing community health workers to address mental health disparities for underserved populations: a systematic review. Adm Policy Ment Health 2018;45(2):195-211 [FREE Full text] [doi: 10.1007/s10488-017-0815-0] [Medline: 28730278]
43. Mongelli F, Georgakopoulos P, Pato MT. Challenges and opportunities to meet the mental health needs of underserved and disenfranchised populations in the United States. Focus (Am Psychiatr Publ) 2020;18(1):16-24 [FREE Full text] [doi: 10.1176/appi.focus.20190028] [Medline: 32047393]

XSL•FO
RenderX

44. Castillo EG, Ijadi-Maghsoodi R, Shadravan S, Moore E, Mensah MO, Docherty M, et al. Community interventions to promote mental health and social equity. Curr Psychiatry Rep 2019;21(5):35 [FREE Full text] [doi: 10.1007/s11920-019-1017-0] [Medline: 30927093]
45. van Lotringen C, Lusi B, Westerhof GJ, Ludden GDS, Kip H, Kelders SM, et al. The role of compassionate technology in blended and digital mental health interventions: systematic scoping review. JMIR Ment Health 2023;10:e42403 [FREE Full text] [doi: 10.2196/42403] [Medline: 37027207]
46. Jenson C, Wolff SF, Milkovich LM. Effects of a preventive mental health curriculum embedded into a scholarly gaming course on adolescent self-esteem: prospective matched pairs experiment. JMIR Serious Games 2023;11:e48401 [FREE Full text] [doi: 10.2196/48401] [Medline: 38059568]

Abbreviations

CFT: compassion-focused therapy
CMT-Pupils: Compassionate Mind Training
DSM-5: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
GAS: Game Addiction Scale
SCAS: Spence Children's Anxiety Scale

Edited by A Mavragani; submitted 15.11.24; peer-reviewed by R Ghammem, A Alenezi; comments to author 12.12.24; revised version received 06.01.25; accepted 28.01.25; published 19.02.25.

Please cite as:
Zuair A
Effect of the Reassured Self-Compassion–Based School Program on Anxiety, Video Game Addiction, and Body Image Among Rural Female Adolescents: Retrospective Study
JMIR Form Res 2025;9:e68840
URL: https://formative.jmir.org/2025/1/e68840
doi:10.2196/68840
PMID:

©Areeg Zuair. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Machine Learning–Based Prediction of Delirium and Risk Factor Identification in Intensive Care Unit Patients With Burns: Retrospective Observational Study

Ryo Esumi[1*], MD, PhD; Hiroki Funao[2*], MSc; Eiji Kawamoto[1], MD, PhD; Ryota Sakamoto[3], PhD; Asami Ito-Masui[1], MD, PhD; Fumito Okuno[4], MD; Toru Shinkai[4], MD, PhD; Atsuya Hane[4], MD; Kaoru Ikejiri[4], MD; Yuichi Akama[1], MD, PhD; Arong Gaowa[1], PhD; Eun Jeong Park[1], PhD; Ryo Momosaki[5], MD, PhD; Ryuji Kaku[6], MD, PhD; Motomu Shimaoka[1], MD, PhD

[1]Department of Molecular Pathobiology and Cell Adhesion Biology, Mie University Graduate School of Medicine, Mie University, Tsu, Japan

[2]Department of Practical Nursing, Mie University Graduate School of Medicine, Tsu city, Japan

[3]Department of Medical Informatics, Mie University Hospital, Tsu city, Japan

[4]Department of Emergency and Critical Care Center, Mie University Hospital, Tsu city, Japan

[5]Department of Rehabilitation Medicine, Mie University Graduate School of Medicine, Tsu city, Japan

[6]Department of Anesthesiology, Mie University Hospital, Tsu city, Japan

[*]these authors contributed equally

Corresponding Author:
Eiji Kawamoto, MD, PhD
Department of Molecular Pathobiology and Cell Adhesion Biology, Mie University Graduate School of Medicine
Mie University
Edobashi 2-174
Tsu, 5140001
Japan
Phone: 81 0592321111
Email: a_2.uk@mac.com

Abstract

Background: The incidence of delirium in patients with burns receiving treatment in the intensive care unit (ICU) is high, reaching up to 77%, and has been associated with increased mortality rates. Therefore, early identification of patients at high risk of delirium onset is essential for improving treatment strategies.

Objective: This study aimed to create a machine learning model for predicting delirium in patients with burns during their ICU stay using patient data from the first day of ICU admission and identify predictive factors for ICU delirium in patients with burns.

Methods: This study focused on 82 patients with burns aged ≥18 years who were admitted to the ICU at Mie University Hospital for ≥24 hours between January 2015 and June 2023. In total, 70 variables were measured in patients upon ICU admission and used as explanatory variables in the ICU delirium prediction model. Delirium was assessed using the Intensive Care Delirium Screening Checklist every 8 hours after ICU admission. A total of 10 different machine learning methods were used to predict ICU delirium. Multiple receiver operating characteristic curves were plotted for various machine learning models, and the area under the curve (AUC) for each was compared. In addition, the top 15 risk factors contributing to delirium onset were identified using Shapley additive explanations analysis.

Results: Among the 10 machine learning models tested, logistic regression (mean AUC 0.906, SD 0.073), support vector machine (mean AUC 0.897, SD 0.056), k-nearest neighbor (mean AUC 0.894, SD 0.060), neural network (mean AUC 0.857, SD 0.058), random forest (mean AUC 0.850, SD 0.074), adaptive boosting (mean AUC 0.832, SD 0.094), gradient boosting machine (mean AUC 0.821, SD 0.074), and naïve Bayes (mean AUC 0.827, SD 0.095) demonstrated the highest accuracy in predicting ICU delirium. Specifically, 24-hour urine output (from ICU admission to 24 hours), oxygen saturation, burn area, total bilirubin level, and intubation upon ICU admission were identified as the major risk factors for delirium onset. In addition, variables, such as the proportion of white blood cell fractions, including monocytes; methemoglobin concentration; and respiratory rate, were identified as important risk factors for ICU delirium.

XSL•FO
RenderX

Conclusions: This study demonstrated the ability of machine learning models trained using vital signs and blood data upon ICU admission to predict delirium in patients with burns during their ICU stay.

(JMIR Form Res 2025;9:e65190) doi:10.2196/65190

KEYWORDS

burns; delirium; intensive care unit; machine learning; prediction model; artificial intelligence; AI

Introduction

Background

Delirium is a significant complication in patients in the intensive care unit (ICU) and is recognized as an urgent medical need requiring treatment and prevention. Delirium is defined as acute brain dysfunction associated with underlying conditions characterized by fluctuating bouts of impaired consciousness, attention, and cognition. This condition is frequently observed in patients admitted to the ICU, with delirium occurring in 10% to 50% of patients [1]. Delirium is an independent predictor of poor outcomes, and there is currently no established specific treatment, making early diagnosis and prevention critically important [2]. In particular, the incidence of delirium in patients with burns can reach 77% [3], with reports indicating that 30% of patients who develop delirium respond effectively to prevention and treatment [4]. These facts underscore the importance of identifying patients at high risk of delirium and implementing preventive measures.

In recent years, research on ICU delirium prediction using artificial intelligence technology has advanced, with a particular focus on the application of various machine learning algorithms. These algorithms, such as random forest (RF), support vector machine (SVM), and gradient boosting, can be used to develop predictive models for ICU delirium [5]. While these algorithms themselves do not identify relevant features, they can be combined with feature importance analysis techniques such as Shapley additive explanations (SHAP) to determine which variables contribute most significantly to the predictions. This approach allows for the analysis of large volumes of patient data with speed and accuracy beyond human capacity. However, research on the prediction of ICU delirium in patients with burns remains underdeveloped, and very few studies have been conducted in this field.

Objectives

This study aimed to demonstrate whether it is possible to predict ICU delirium in patients with burns using machine learning. Specifically, we hypothesized that a machine learning model using clinical data such as vital signs and blood test results could predict delirium in ICU patients. The null hypothesis was that these models would not be superior to random chance in predicting ICU delirium.

While ICU delirium prediction potential for multiple machine learning models, as well as the model with the highest accuracy, was assessed, the second objective of this study was to identify risk factors for ICU delirium in patients with burns and contribute to the development of more effective prevention and treatment strategies.

This ICU delirium prediction approach using machine learning has the potential to support the early detection of ICU delirium in patients with burns and, ultimately, improve patient outcomes.

Methods

Patient Demographics and Data Collection

This was a retrospective observational study focused on predicting delirium in patients with burns admitted to the ICU. This study included 82 patients with burns aged ≥18 years who were admitted to the Mie University Hospital ICU for ≥24 hours between January 2015 and June 2023. The sample size of 82 patients was determined by including all patients with burns who were admitted to the ICU within the study period.

Patients were retrospectively included based on the inclusion criteria, which ensured the comprehensive capture of all eligible cases during the study period. This approach minimized selection bias and allowed for a representative sample of the population of patients with burns in our ICU.

Physiological, biochemical, and clinical data collected from these patients upon ICU admission were used to extract 70 explanatory variables. This study aimed to develop a model using these 70 variables to predict delirium onset during ICU admission, assess its accuracy, and identify the risk factors contributing to each model.

The Definition, Diagnostic Criteria, and Standard Assessment of ICU Delirium

ICU delirium, also known as ICU psychosis, is an acute, fluctuating change in consciousness and cognition that occurs frequently in patients who are critically ill. ICU delirium is characterized by disturbances in attention, awareness, and cognitive function. These disturbances are often temporary and reversible but can lead to prolonged ICU stays, increased morbidity and mortality, and long-term cognitive impairments if not properly managed.

The standard assessment of ICU delirium involves the use of validated diagnostic tools to ensure accurate detection and timely intervention. In total, 2 widely recognized tools are the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU) and the Intensive Care Delirium Screening Checklist (ICDSC).

The CAM-ICU is a structured diagnostic tool specifically designed for use in the ICU setting. It is based on the Confusion Assessment Method and has been modified for the critical care environment. The CAM-ICU assesses four key features: (1) acute onset of mental status changes or a fluctuating course, (2) inattention, (3) disorganized thinking, and (4) altered level of consciousness. A positive CAM-ICU diagnosis requires the

XSL•FO
RenderX

presence of both features 1 and 2 and either feature 3 or 4. This tool is favored for its ease of use and quick administration, making it suitable for frequent assessments.

The ICDSC is another tool used to screen for delirium in ICU patients, consisting of an 8-item checklist that assesses various cognitive and behavioral symptoms associated with delirium. The items are (1) altered level of consciousness, (2) inattention, (3) disorientation, (4) hallucinations or delusions, (5) psychomotor agitation or retardation, (6) inappropriate speech or mood, (7) sleep and wake cycle disturbances, and (8) symptom fluctuation. Each item is scored based on its presence within the previous 24 hours, and a total score of ≥4 indicates the presence of delirium. The ICDSC provides a comprehensive overview of the patient's condition over a longer period than the CAM-ICU.

In this study, delirium was assessed using the internationally recognized ICDSC every 8 hours after ICU admission. Delirium was diagnosed when the ICDSC score was ≥4 points. We opted for a binary classification approach because this method simplifies the model's output to either *delirium present* or *delirium absent*, facilitating smoother decision-making in clinical settings. Another reason for this choice was the diversity in patient conditions in real-world clinical environments, where a few specific atypical cases can skew the predictions of regression models. Binary classification is less susceptible to the influence of such outliers, enabling the development of a more robust model. This method allowed for accurate determination of the presence and severity of delirium.

Development of Machine Learning Models

In total, 10 different algorithms were used to develop the machine learning models: logistic regression (LR), RF, SVM, neural network, k-nearest neighbor (KNN), decision tree, naïve Bayes, adaptive boosting (AdaBoost), gradient-boosting machine (GBM), and linear discriminant analysis (LDA). These models were selected based on the area under the receiver operating characteristic curve (AUC) to compare the accuracy of delirium prediction.

Ethical Considerations

This study was approved by the ethics committee of Mie University (H2020-164), ensuring compliance with ethical guidelines for clinical research. This study underwent ethical review and was approved as a clinical research study. For informed consent, we applied the opt-out method, allowing participants to refuse the inclusion of their data in this study. If participants chose to opt out, their data were excluded from the analysis. To protect privacy and confidentiality, all study data were fully anonymized. Data extracted from electronic medical records were deidentified, ensuring that individual patients could not be identified. No compensation was provided to participants. As the study used an opt-out approach, participants had the right to refuse the inclusion of their data without any consequences.

Feature Importance Analysis Using SHAP

In this study, we used SHAP to interpret the feature importance of our machine learning models. SHAP provides a unified measure of each variable's contribution to the prediction outcome. By applying SHAP, we identified the top 15 predictors of ICU delirium across all models, which allowed for a detailed understanding of the factors most associated with delirium onset. The SHA*P* values helped clarify the influence of variables such as 24-hour urine output, total bilirubin (T-bil) levels, and respiratory rate on the predictive performance of the models. SHAP analysis also enabled clinicians to visualize and interpret the risk factors, enhancing the practical applicability of the model's predictions.

Data Preprocessing for Machine Learning

Handling Missing Values

In this study, missing values were not handled. This approach ensured that all features had complete data, thus preserving the dataset's integrity without excluding any records.

Data Splitting

The dataset was loaded from a CSV file using the *Pandas* library. It was then split into explanatory variables and the target variable, where the target variable was delirium. The data were divided into 2 subsets: 80% was allocated for training the model, and 20% was reserved for internal validation. To evaluate the model's performance, the dataset was further split into folds using stratified k-fold cross-validation, which ensures that the proportion of classes remains consistent across folds.

Data Standardization

The explanatory variables in both the training and test sets were standardized using *StandardScaler* from the *sklearn* library to ensure that each feature had a mean of 0 and an SD of 1. This preprocessing step is crucial for models sensitive to feature scaling.

Model Selection and Initialization

Several machine learning models were initialized for this study, including SVM, neural network, KNN, decision tree, naïve Bayes, AdaBoost, GBM, LDA, LR, and RF. Each model was imported from the *sklearn* or *SciPy* library and initialized with appropriate default parameters.

Hyperparameter Tuning

For specific models, such as SVM, neural network, KNN, decision tree, and RF, hyperparameter tuning was performed using grid search with cross-validation. The *GridSearchCV* class from the *sklearn* library was used to search for the best combination of hyperparameters, optimizing the model's performance based on the AUC metric.

Model Training and Evaluation

Each model was trained using stratified k-fold cross-validation to ensure robust performance evaluation. For models with hyperparameter grids defined, grid search was applied to find the best hyperparameter set. The best-performing model from the grid search was then used to fit the data in each fold. The model's performance was evaluated using the AUC score on the test subset within each fold.

XSL•FO
RenderX

Statistical Analysis

Continuous variables are reported as medians with IQRs and were compared using the Mann-Whitney *U* test or the Kruskal-Wallis test, depending on the number of groups. Categorical variables are expressed as counts and percentages, with comparisons made using the Fisher exact test or the chi-square test, as appropriate. All statistical analyses were conducted using the SPSS software (version 21; IBM Corp), and statistical significance was defined as $P<.05$.

To evaluate differences in model performance, pairwise 2-tailed *t* tests were applied to the AUC scores using the *ttest_ind* function from the *scipy.stats* module. In addition, a 1-way ANOVA was conducted using the *f_oneway* function to identify statistically significant differences across models.

Model performance results were visualized through bar charts displaying the mean AUC scores along with their SDs for each model. These visualizations were created using the *Matplotlib* library. A heat map was also generated to illustrate pairwise *P* values from the *t* tests, offering a clear representation of the statistical significance of the differences between models. The ANOVA *P* value was similarly calculated and displayed.

Results

Comparative Analysis of ICU Delirium in Patients With Burns

Compared to patients without delirium (Table 1), patients with burns with ICU delirium were older (age: median 77.0, IQR 69.5-84.5 years vs median 60.5, IQR 37.5-73.0 years; $P<.001$), were more likely to have airway burns (12/32, 38% vs 8/50, 16%; $P=.03$), and experienced longer ICU stays (median 2.5, IQR 1.8-11.2 days vs 2.0, IQR 1.0-3.0 days; $P=.009$). The mortality rate was significantly higher in the delirium group (7/32, 22% vs 0%; $P=.001$), and these patients more frequently required intubation (20/32, 62% vs 4/50, 8%; $P<.001$). The delirium group also had greater burn areas (median 16%, IQR 9.75%-34% vs 8%, IQR 4.6%-17%; $P=.007$) and burn indexes (median 9.5, IQR 4.7-25.1 vs median 4.0, IQR 1.0-10.0; $P=.002$).

XSL•FO
RenderX

Table 1. Characteristics of patients with burns with intensive care unit (ICU) delirium—comparison of explanatory factors between patients with and without deliriuma.

	Burns with delirium (n=32)	Burns without delirium (n=50)	*P* value
Age (y), median (IQR)	77.0 (69.5-84.5)	60.5 (37.5-73.0)	*<.001* [b]
Sex (male), n (%)	20 (63)	32 (64)	.90
Height (cm), median (IQR)	160.5 (150.0-170.0)	162.2 (158.0-168.0)	.42
Weight (kg), median (IQR)	54.4 (44.4-63.3)	60.2 (50.5-67.1)	.07
BMI (kg/m^2), median (IQR)	21.2 (19.0-23.4)	23.1 (19.7-24.9)	.06
Airway burns, n (%)	12 (38)	8 (16)	*.03* [b]
Length of ICU stay (days), median (IQR)	2.5 (1.8-11.2)	2.0 (1.0-3.0)	*.009* [b]
Mortality rate, n (%)	7 (22)	0 (0)	*.001* [b]
Intubation, n (%)	20 (62)	4 (8)	*<.001* [b]
Burn area (%), median (IQR)	16 (9.75-34)	8 (4.6-17)	*.007* [b]
Burn index, median (IQR)	9.5 (4.7-25.1)	4.0 (1.0-10.0)	*.002* [b]
WBC[c] (10^3 per µL), median (IQR)	11.9 (9.0-18.5)	8.8 (6.5-11.3)	*.006* [b]
RBC[d] (10^6 per µL), median (IQR)	4.3 (3.7-5.3)	4.4 (4.0-4.9)	.51
Hemoglobin (g/dL), median (IQR)	13.5 (11.9-15.9)	14.6 (14.0-15.9)	*.04* [b]
Hematocrits (%), median (IQR)	40 (35-44)	43 (41-46.7)	*.009* [b]
MCV[e] (fL), median (IQR)	92.0 (89.9-93.8)	92.3 (87.7-95.6)	.95
MCH[f] (pg), median (IQR)	30.8 (30.0-32.4)	31.4 (29.7-32.9)	.71
MCHC[g] (%), median (IQR)	33.8 (33.2-34.6)	33.9 (33-34.6)	.77
Platelet count (10^3 per µL), median (IQR)	205.5 (165.0-314.8)	249.5 (202.2-302.0)	.50
Neutrophils (%), median (IQR)	77 (73.9-83.7)	74.4 (64.6-80.9)	*.04* [b]
Lymphocytes (%), median (IQR)	15.8 (8.4-17.9)	16.8 (12.6-23.6)	*.02* [b]
Monocytes (%), median (IQR)	6.6 (5.2-7.7)	6.6 (5.7-7.2)	.73
Eosinophils (%), median (IQR)	0.6 (0.4-1.4)	1.2 (0.6-1.9)	*.04* [b]
Basophils (%), median (IQR)	0.4 (0.2-0.5)	0.4 (0.2-0.5)	.84
Neutrophil count (10^3 per µL), median (IQR)	8985 (6265-10,830)	6840.0 (4449.2-9029.0)	*.03* [b]
Lymphocyte count (10^3 per µL), median (IQR)	1550 (1130-1943)	1550.0 (1178.2-2145.0)	.55
Monocyte count (10^3 per µL), median (IQR)	767 (478-993)	575.0 (447.5-767.0)	*.04* [b]
Eosinophil count (10^3 per µL), median (IQR)	70.0 (37.5-111.0)	90.0 (40.0-167.7)	.22
Basophil count (10^3 per µL), median (IQR)	38.5 (30.0-62.5)	37.0 (20.0-50.0)	.13
APTT[h] (seconds), median (IQR)	30.8 (25.9-36.8)	27.4 (24.9-29.9)	*.02* [b]
PT[i] (seconds), median (IQR)	12.2 (11.5-13.6)	11.3 (10.8-12.1)	*.001* [b]
PT (%)[j], median (IQR)	94.8 (77-104.1)	105.4 (97-116.9)	*.001* [b]
PT-INR[k], median (IQR)	1.0 (0.9-1.1)	0.9 (0.9-1.0)	*.002* [b]
Fibrinogen (mg/dL), median (IQR)	299.0 (251.2-374.2)	272.5 (225.2-327.7)	.20
D-dimer (µg/mL), median (IQR)	4.0 (1.5-8.5)	1.0 (0.2-8.2)	*.02* [b]
pH, median (IQR)	7.3 (7.3-7.4)	7.4 (7.3-7.4)	*.04* [b]

XSL•FO
RenderX

	Burns with delirium (n=32)	Burns without delirium (n=50)	*P* value
PCO_2[l] (mm Hg), median (IQR)	40.0 (31.7-45.0)	36.5 (33.2-39.7)	.28
PO_2 (mm Hg), median (IQR)	152.5 (93.5-322.7)	123.0 (83.2-166.0)	*.03* [b]
SO_2[m] (%), median (IQR)	99.1 (98.3-99.6)	97.6 (94.9-99)	*.01* [b]
HCO_3^-[n] (mmol/L), median (IQR)	23.1 (18.9-25.6)	23.1 (20.5-24.4)	.76
Anion gap (mmol/L), median (IQR)	10.4 (4.4-17.9)	13.1 (10.4-17.2)	.24
O_2-Hb[o] (%), median (IQR)	96.0 (92.5-97.3)	95.9 (90.8-97.1)	.59
CO-Hb[p] (%), median (IQR)	1.6 (1.1-3.4)	1.7 (0.7-3.4)	.75
MetHb[q] (%), median (IQR)	0.9 (0.5-1.1)	0.4 (0.3-0.9)	*.004* [b]
Ionized calcium (mmol/L), median (IQR)	1.1 (1.0-1.1)	1.1 (1.0-1.1)	.60
Lactate (mmol/L), median (IQR)	3.0 (1.5-4.9)	2.3 (1.8-3.3)	.45
TP[r] (g/dL), median (IQR)	6.4 (5.9-6.7)	7.0 (6.2-7.5)	*.006* [b]
Albumin (g/dL), median (IQR)	3.6 (3.1-3.9)	4.1 (3.2-4.4)	*.004* [b]
BUN[s] (mg/dL), median (IQR)	16.9 (12.8-22.7)	13.4 (11.0-19.0)	.01
Creatinine (mg/dL), median (IQR)	0.8 (0.6-1.0)	0.7 (0.5-0.8)	*.03* [b]
eGFR[t] (mL per minute per 1.73 m^2), median (IQR)	67.2 (47.8-87.2)	83.7 (59.9-101.0)	*.02* [b]
Sodium (mmol/L), median (IQR)	139.0 (138.0-141.0)	140.0 (138.0-141.0)	>.99
Potassium (mmol/L), median (IQR)	4.0 (3.7-4.5)	3.9 (3.7-4.4)	.58
CL[u] (mmol/L), median (IQR)	105.0 (102.0-107.0)	104.0 (103.0-106.7)	.57
Calcium (mg/dL), median (IQR)	8.5 (8.2-8.9)	8.9 (8.4-9.2)	*.009* [b]
AST[v] (U/L), median (IQR)	39.5 (22.7-60.5)	30.5 (24.2-38.0)	.30
ALT[w] (U/L), median (IQR)	23.0 (14.0-43.5)	20.5 (14.5-29.0)	.46
LDH[x] (U/L), median (IQR)	294.5 (216.7-630.7)	255.5 (197.7-303.0)	.05
ALP[y] (U/L), median (IQR)	215.5 (154.2-263.5)	177.0 (91.5-221.7)	.05
T-bil[z] (mg/dL), median (IQR)	0.9 (0.7-1.3)	0.6 (0.4-0.8)	*<.001* [b]
Glucose (mg/dL), median (IQR)	147.0 (120.5-179.2)	136.0 (113.2-180.5)	.64
CPK[aa] (U/L), median (IQR)	249.5 (96.5-645.7)	129.0 (80.0-243.0)	.06
AMY[ab] (U/L), median (IQR)	85.5 (57.7-134.0)	80.5 (61.0-103.5)	.46
CRP[ac] (mg/dL), median (IQR)	0.4 (0.0-5.5)	0.1 (0.0-1.85)	*.03* [b]
Daily urinary output (mL), median (IQR)	412.5 (215.0-908.2)	1493.0 (985.0-1977.0)	*<.001* [b]
Respiratory rate (breaths per minute), median (IQR)	19.5 (17.7-21.0)	16.0 (14.0-20.0)	*.01* [b]
sBP[ad] (mmHg), median (IQR)	146.5 (115.5-160.2)	141.0 (124.2-157.0)	.84
dBP[ae] (mmHg), median (IQR)	81.0 (64.7-88.5)	74.0 (65.5-86.5)	.99
HR[af] (beats per minute), median (IQR)	97.0 (81.0-113.5)	88.5 (81.2-99.0)	.12

XSL•FO
RenderX

	Burns with delirium (n=32)	Burns without delirium (n=50)	*P* value
BT[ag] (°C), median (IQR)	37.0 (36.3-37.5)	37.0 (36.8-37.4)	.23

[a]This table presents the demographic, clinical, and laboratory characteristics of adult patients with burns who were admitted to the ICU at Mie University Hospital between January 2015 and June 2023. This study compared patients who developed ICU delirium with those who did not, highlighting differences in vital signs, burn severity, inflammatory markers, coagulation parameters, organ function indicators, and other physiological variables. The findings aimed to identify key risk factors for ICU delirium and provide insights into the pathophysiology and early prediction of delirium in patients with burns who are critically ill. Statistical comparisons were conducted using median and IQR values, with a *P* value of <.05 considered statistically significant.

[b]Statistically significant at *P*<.05.

[c]WBC: white blood cell count.

[d]RBC: red blood cell count.

[e]MCV: mean corpuscular volume.

[f]MCH: mean corpuscular hemoglobin.

[g]MCHC: mean corpuscular hemoglobin concentration.

[h]APTT: activated partial thromboplastin time.

[i]PT: prothrombin time.

[j]PT percentage.

[k]PT-INR: prothrombin time international normalized ratio.

[l]PCO_2: partial pressure of carbon dioxide.

[m]SO_2: saturation of oxygen.

[n]HCO_3–: bicarbonate.

[o]O_2-Hb: oxygenated hemoglobin.

[p]CO-Hb: carboxyhemoglobin.

[q]MetHb: methemoglobin.

[r]TP: total protein.

[s]BUN: blood urea nitrogen.

[t]eGFR: estimated glomerular filtration rate.

[u]CL: chloride.

[v]AST: aspartate aminotransferase.

[w]ALT: alanine aminotransferase.

[x]LDH: lactate dehydrogenase.

[y]ALP: alkaline phosphatase.

[z]T-bil: total bilirubin.

[aa]CPK: creatine phosphokinase.

[ab]AMY: amylase.

[ac]CRP: C-reactive protein.

[ad]sBP: systolic blood pressure.

[ae]dBP: diastolic blood pressure.

[af]HR: heart rate.

[ag]BT: body temperature.

Laboratory parameters further demonstrated significant differences. The delirium group showed higher white blood cell counts (median 11.9, IQR 9.0-18.5 per µL vs median 8.8, IQR 6.5-11.3 per µL; *P*=.006), lower hemoglobin (median 13.5, IQR 11.9-15.9 g/dL vs median 14.6, IQR 14.0-15.9 g/dL; *P*=.04), and lower hematocrit levels (median 40%, IQR 35%-44% vs median 43%, IQR 41%-46.7%; *P*=.009). Neutrophil percentages were elevated (median 77%, IQR 73.9%-83.7% vs median 74.4%, IQR 64.6%-80.9%; *P*=.04), whereas lymphocyte (median 15.8%, IQR 8.4%-17.9% vs median 16.8%, IQR 12.6%-23.6%; *P*=.02) and eosinophil (median 0.6%, IQR 0.4%-1.4% vs median 1.2%, IQR 0.6%-1.9%; *P*=.04) percentages were lower in the delirium group. Absolute neutrophil counts were higher (median 8985, IQR 6265-10,830 cells per µL vs median 6840, IQR 4449.2-9029.0 cells per µL; *P*=.03), as were monocyte counts (median 767, IQR 478-993 cells per µL vs median 575.0, IQR 447.5-767.0 cells per µL; *P*=.04).

Coagulation parameters revealed that activated partial thromboplastin time was prolonged in the delirium group (median 30.8, IQR 25.9-36.8 seconds vs median 27.4, IQR 24.9-29.9 seconds; *P*=.02). Prothrombin time (PT) and PT percentage also differed significantly (median 12.2, IQR 11.5-13.6 seconds vs median 11.3, IQR 10.8-12.1 seconds with *P*=.001 and median 94.8%, IQR 77%-104.1% vs median 105.4%, IQR 97%-116.9% with *P*=.001, respectively), as did the PT international normalized ratio (median 1.0, IQR 0.9-1.1 vs median 0.9, IQR 0.9-1.0; *P*=.002). D-dimer levels were

XSL•FO
RenderX

elevated (median 4.0, IQR 1.5-8.5 mg/L vs median 1.0, IQR 0.2-8.2 mg/L; *P*=.02).

Acid-base and oxygenation parameters showed a lower pH (median 7.3, IQR 7.3-7.4 vs median 7.4, IQR 7.3-7.4; *P*=.04) and higher PO_2 (median 152.5, IQR 93.5-322.7 mmHg vs median 123.0, IQR 83.2-166.0 mmHg; *P*=.03) in the delirium group. Oxygen saturation was also significantly higher (median 99.1%, IQR 98.3%-99.6% vs median 97.6%, IQR 94.9%-99%; *P*=.01), as was methemoglobin (MetHb; median 0.9%, IQR 0.5%-1.1% vs median 0.4%, IQR 0.3%-0.9%; *P*=.004).

Biochemical markers indicated lower total protein (median 6.4, IQR 5.9-6.7 g/dL vs median 7.0, IQR 6.2-7.5 g/dL; *P*=.006) and albumin (median 3.6, IQR 3.1-3.9 g/dL vs median 4.1, IQR 3.2-4.4 g/dL; *P*=.004) levels in the delirium group. Creatinine levels were higher (median 0.8, IQR 0.6-1.0 mg/dL vs median 0.7, IQR 0.5-0.8 mg/dL; *P*=.03), and estimated glomerular filtration rate was lower (median 67.2, IQR 47.8-87.2 mL per minute per 1.73 m^2 vs median 83.7, IQR 59.9-101.0 mL per minute per 1.73 m^2; *P*=.02). Serum calcium was lower (median 8.5, IQR 8.2-8.9 mg/dL vs median 8.9, IQR 8.4-9.2 mg/dL; *P*=.009), and T-bil levels were elevated (median 0.9, IQR 0.7-1.3 mg/dL vs median 0.6, IQR 0.4-0.8 mg/dL; *P*<.001). C-reactive protein levels were significantly higher in the delirium group (median 0.4, IQR 0.0-5.5 mg/dL vs median 0.1, IQR 0.0-1.85 mg/dL; *P*=.03).

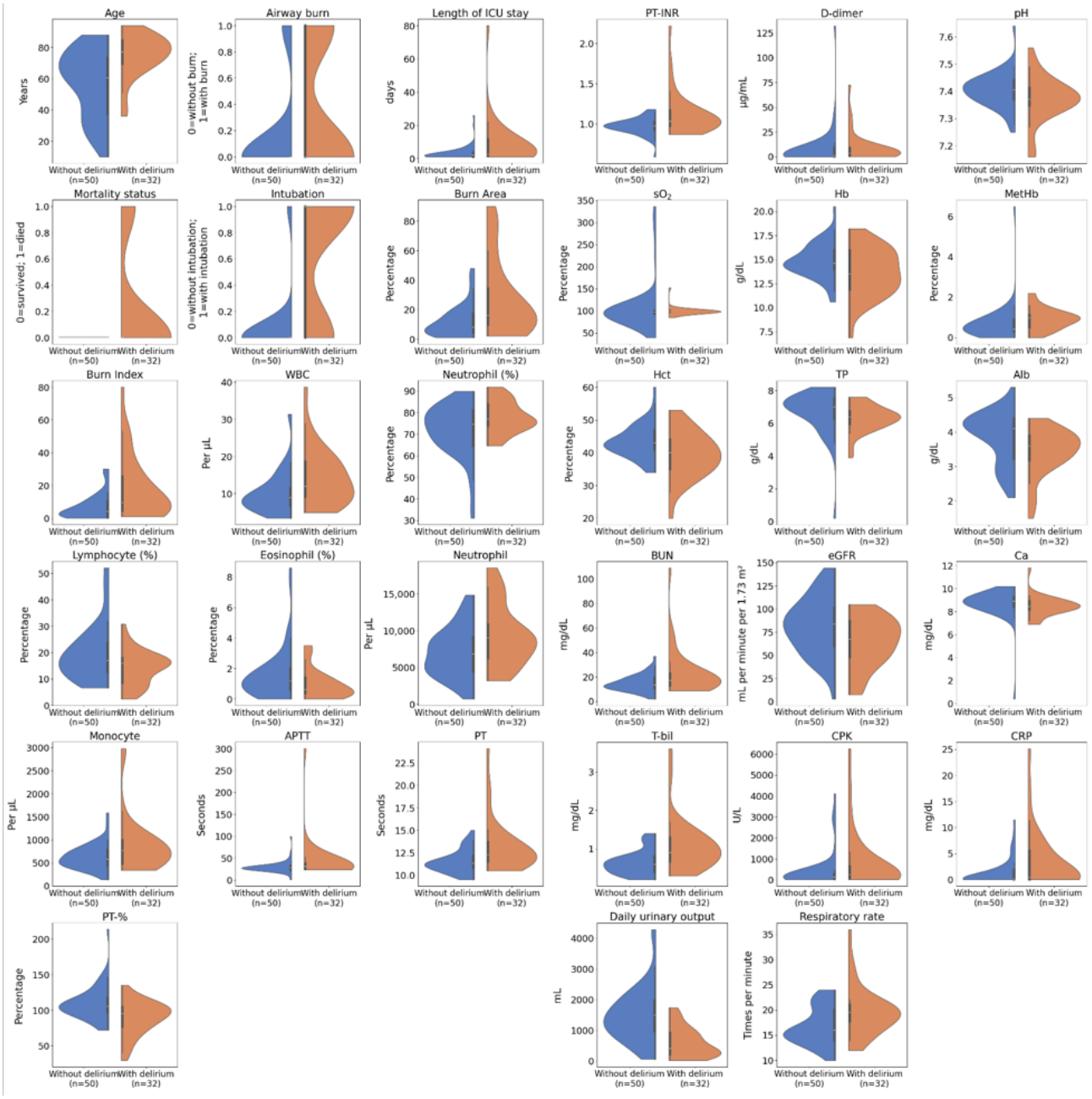
Finally, daily urinary output was markedly reduced in the delirium group (median 412.5, IQR 215.0-908.2 mL vs median 1493.0, IQR 985.0-1977.0 mL; *P*<.001), and respiratory rates were higher (median 19.5, IQR 17.7-21.0 breaths per minute vs median 16.0, IQR 14.0-20.0 breaths per minute; *P*=.01).

Visual Examination of the Data Using Violin Plots

Next, the distribution of the data between the 2 groups (those with and without ICU delirium) was visually examined using violin plots. As shown in Figure 1, violin plots visually represent the density and range of the data, allowing for the identification of data dispersion and bimodality, that is, the shape of the distribution. Specifically, it became evident that urine output decreased proportionally, serving as a risk factor for ICU delirium.

XSL•FO
RenderX

Figure 1. Comparative analysis of explanatory factors for intensive care unit (ICU) delirium in patients with burns using violin plots. This figure illustrates the distribution of key clinical and laboratory variables among 2 groups of ICU patients with burns: those who developed delirium (with delirium) and those who did not (without delirium). Violin plots visualize the distribution and density of each variable, highlighting differences in factors such as age, burn area, total bilirubin (T-bil), methemoglobin (MetHb) levels, daily urinary output, and respiratory rate. These comparisons emphasize the significance of various explanatory factors in predicting ICU delirium, aiding in the identification of potential risk indicators for clinical decision-making. Alb: albumin; APTT: activated partial thromboplastin time; BUN: blood urea nitrogen; Ca: calcium; CPK: creatine phosphokinase; CRP: C-reactive protein; eGFR: estimated glomerular filtration rate; Hb: hemoglobin; Hct: hematocrit level; PT-%: prothrombin time percentage; PT: prothrombin time; PT-INR: PT international normalized ratio; sO2: oxygen saturation; TP: total protein; WBC: white blood cell count.



Machine Learning Model Evaluation for Delirium Prediction

In this study, we evaluated 10 different machine learning models for delirium prediction using 2 performance metrics: the AUC and Matthews correlation coefficient (MCC). The results are summarized in Table 2 and Figure 2.

XSL•FO
RenderX

Table 2. Performance comparison of machine learning models across various metrics[a].

Model	AUC[b], mean (SD)	MCC[c], mean (SD)	True positives	True negatives	False positives	False negatives	Accuracy	Precision	Recall	F_1-score
Support vector machine	0.897 (0.056)	0.157 (0.202)	0.6	10.0	0.0	5.8	0.647	0.400	0.095	0.150
Neural network	0.857 (0.058)	0.522 (0.238)	4.2	8.4	1.6	2.2	0.757	0.750	0.652	0.664
K-nearest neighbor	0.894 (0.060)	0.529 (0.311)	3.4	9.6	0.4	3.0	0.793	0.733	0.524	0.605
Decision tree	0.729 (0.033)	0.417 (0.059)	4.2	7.6	2.4	2.2	0.720	0.655	0.624	0.634
Naïve Bayes	0.827 (0.095)	0.411 (0.351)	3.0	9.2	0.8	3.4	0.744	0.683	0.462	0.533
AdaBoost[d]	0.832 (0.094)	0.493 (0.126)	4.0	8.4	1.6	2.4	0.756	0.753	0.614	0.651
Gradient boosting machine	0.821 (0.074)	0.486 (0.026)	3.8	8.6	1.4	2.6	0.768	0.796	0.586	0.651
Linear discriminant analysis	0.660 (0.114)	0.337 (0.163)	3.8	7.4	2.6	2.6	0.684	0.601	0.590	0.583
Logistic regression	*0.906* [e] *(0.073)*	*0.625 (0.162)*	4.8	8.6	1.4	1.6	*0.818*	0.797	*0.743*	*0.755*
Random forest	0.850 (0.074)	0.505 (0.217)	3.4	9.2	0.8	3.0	0.757	*0.811*	0.529	0.617

[a]This table presents the predictive performance of 10 machine learning models trained on clinical and laboratory data from patients with burns admitted to the intensive care unit (ICU) at Mie University Hospital. Model performance was evaluated using the mean area under the curve, the mean Matthews correlation coefficient, accuracy, precision, recall, and F_1-score, providing a comprehensive assessment of each model's ability to predict ICU delirium onset. The table also reports the true positive, true negative, false positive, and false negative counts, offering insights into each model's sensitivity and specificity. Logistic regression demonstrated the highest predictive performance (area under the curve=0.906), whereas decision tree and linear discriminant analysis showed relatively lower predictive power. These findings highlight the potential of machine learning in early risk stratification for ICU delirium, emphasizing the importance of selecting an optimal predictive model for clinical application.

[b]AUC: area under the curve.

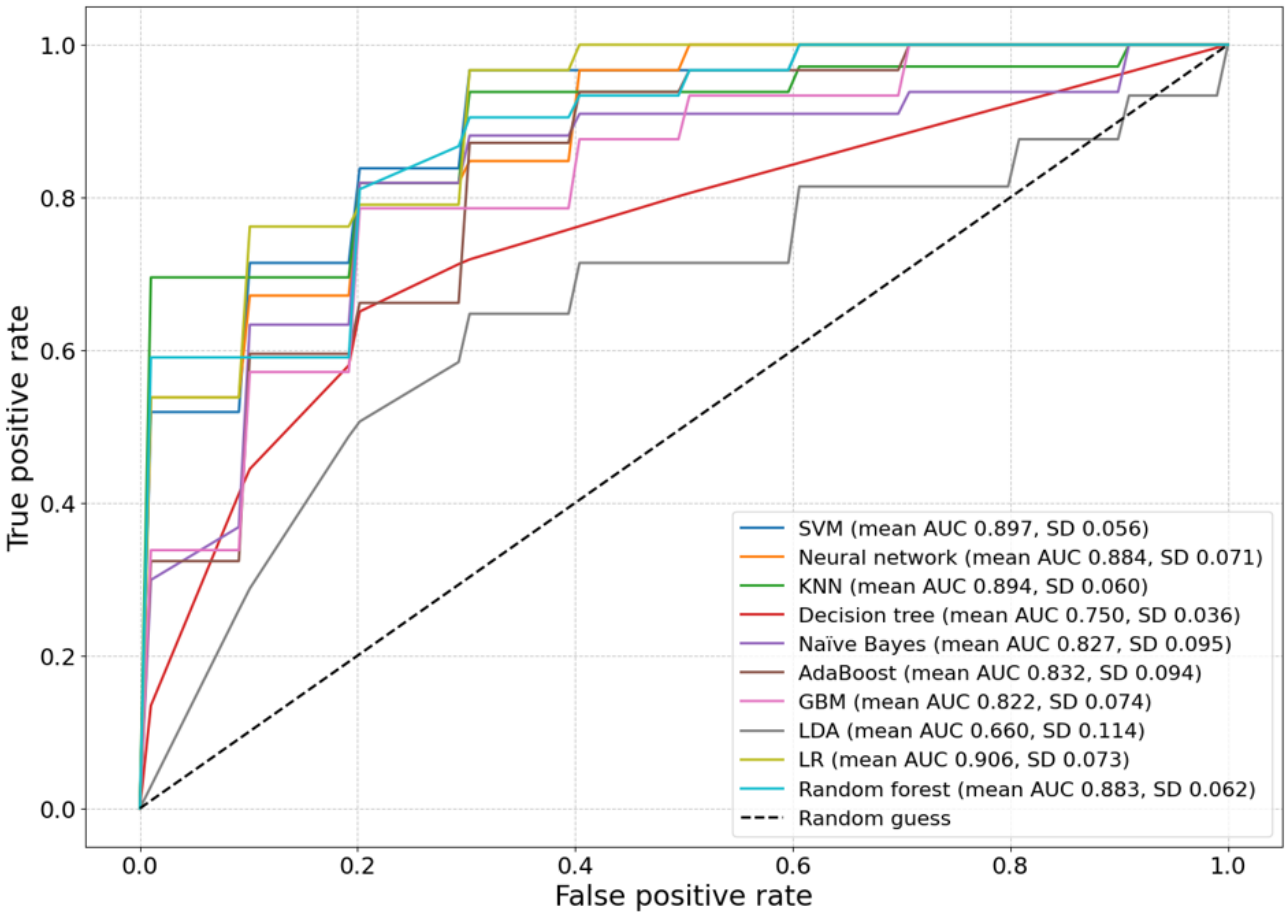
[c]MCC: Matthews correlation coefficient.

[d]AdaBoost: adaptive boosting.

[e]Values in italics indicate the best performance for each metric. Logistic regression exhibited the highest AUC, MCC, accuracy, recall, and F_1 -score, demonstrating superior overall performance.

XSL•FO
RenderX

Figure 2. Receiver operating characteristic (ROC) curves for machine learning models predicting intensive care unit (ICU) delirium. This figure shows ROC curves for 10 machine learning models, illustrating sensitivity and 1-specificity from 5-fold stratified cross-validation. The legend presents mean area under the curve (AUC) values with SDs, indicating predictive accuracy. Logistic regression (LR; mean AUC 0.906, SD 0.073) achieved the highest performance, whereas decision tree and linear discriminant analysis (LDA) performed poorly. The diagonal line (AUC 0.5) represents random guessing. Models with curves near the top left corner demonstrate superior predictive ability for ICU delirium in patients with burns. AdaBoost: adaptive boosting; GBM: gradient-boosting machine; KNN: k-nearest neighbor; SVM: support vector machine.



The LR model showed the best overall performance, achieving the highest mean AUC of 0.906 (SD 0.073) and the highest mean MCC of 0.625 (SD 0.162). This suggests that LR provides a good balance between discriminative power and classification accuracy for delirium prediction. The SVM and KNN models also showed strong discriminative ability, with mean AUC values of 0.897 (SD 0.056) and 0.894 (SD 0.060), respectively. However, the SVM model's relatively low MCC score of 0.157 (SD 0.202) indicates that it may struggle with precise classification, possibly due to class imbalance or suboptimal threshold selection. Neural network and RF demonstrated comparable performance, with mean AUC values of 0.857 (SD 0.058) and 0.850 (SD 0.074), respectively. Their MCC scores were also similar, suggesting consistent performance across both metrics. The decision tree model and LDA showed the lowest performance among the evaluated models, with mean AUC values of 0.729 (SD 0.033) and 0.660 (SD 0.114), respectively. This suggests that these models may not capture the complexity of the delirium prediction task as effectively as the other algorithms. It is noteworthy that, while some models (eg, SVM) achieved high AUC scores, their corresponding MCC scores were relatively low. This discrepancy highlights the importance of using multiple evaluation metrics to gain a comprehensive understanding of model performance, especially in potentially imbalanced classification tasks such as delirium prediction. In summary, LR emerged as the most promising model for delirium prediction in this comparative analysis. However, the strong performance of several other models, such as KNN and neural network, suggests that ensemble methods or model stacking could potentially yield further improvements in predictive accuracy (Table 2).

Performance Metrics of Machine Learning Models in ICU Delirium Prediction

On the basis of the provided classification performance metrics, LR showed the highest overall performance in terms of accuracy (0.818), precision (0.797), recall (0.743), and F_1-score (0.755), indicating a robust balance between sensitivity and specificity. Although KNN showed slightly higher accuracy (0.793) than most models, its recall was lower (0.524), which may limit its utility in detecting positive cases.

Among ensemble methods, GBM and AdaBoost performed well, with GBM achieving a relatively high precision (0.796), reflecting its ability to minimize false positives. RF also yielded strong precision (0.811) but showed lower recall (0.529), suggesting a higher risk of missing positive cases.

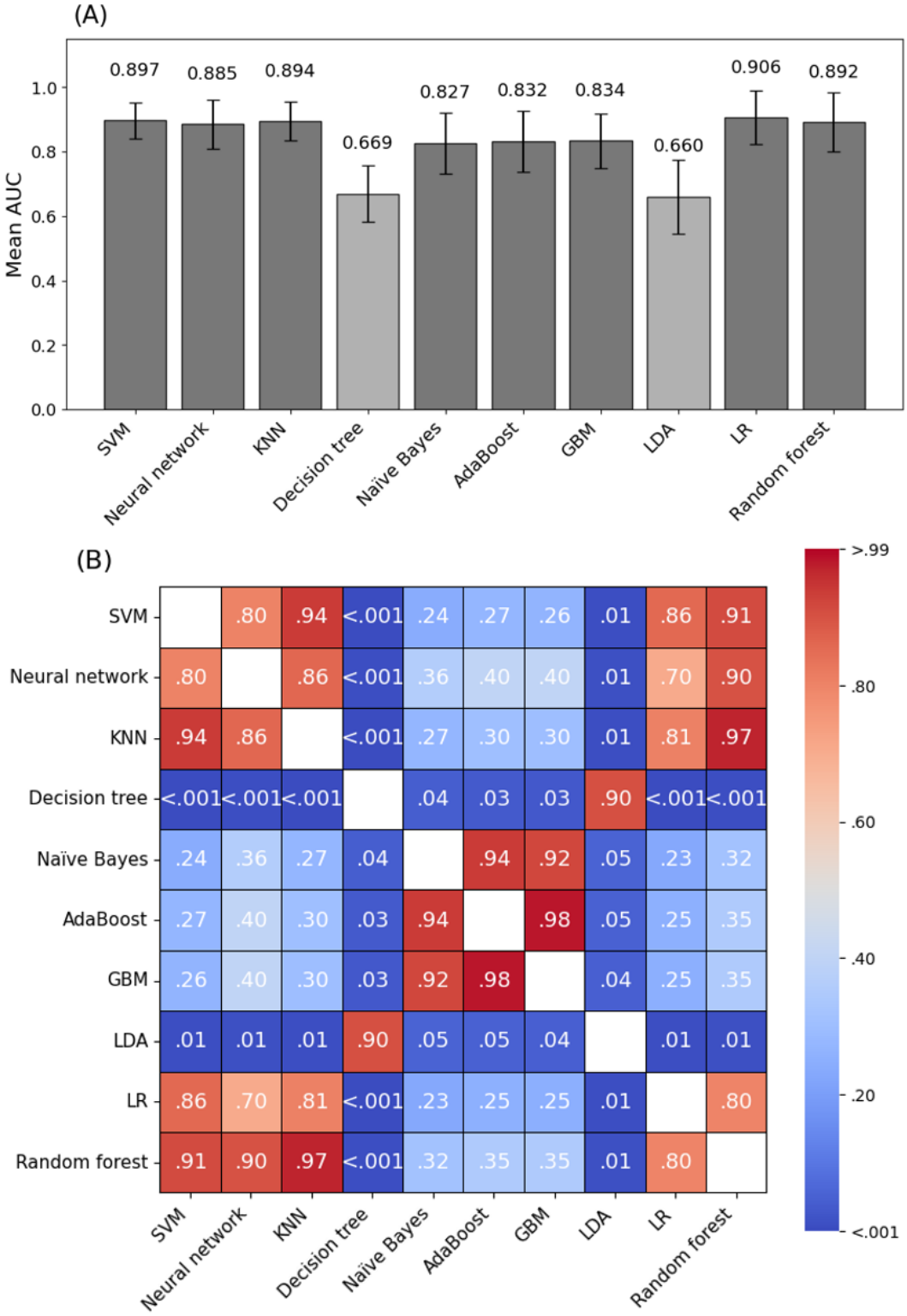
XSL•FO
RenderX

Neural network models achieved a competitive balance across all metrics, notably F_1-score (0.664), which highlights their consistency in handling both sensitivity and specificity. However, SVM and LDA demonstrated weaker recall, suggesting that these models are less suited for tasks requiring high sensitivity.

We then compared the mean AUCs across the models to evaluate their overall discriminative ability. When compared with decision tree, SVM (*P*=.002), neural network (*P*=.005), KNN (*P*=.003), LR (*P*=.003), and RF (*P*=.005) showed a significantly better performance. Naïve Bayes (*P*=.04), AdaBoost (*P*=.03), and GBM (*P*=.03) also demonstrated a significantly better performance than that of decision tree. LDA showed a comparable performance to that of decision tree (*P*=.99; Figure 3).

XSL•FO
RenderX

Figure 3. (A) Mean area under the curve (AUC) scores and (B) statistical significance across machine learning models. This figure compares the predictive performance of 10 machine learning models for intensive care unit (ICU) delirium in patients with burns. The bar graph shows mean AUC scores with SDs from 5-fold stratified cross-validation. Among the models, logistic regression (LR; AUC=0.906) performed best, whereas decision tree and linear discriminant analysis (LDA) had lower AUCs. The heat map displays pairwise *P* values from 2-sample 2-tailed t tests, with darker cells indicating significant differences (*P*<.05). This figure highlights model performance and statistical differences, aiding in selecting effective classifiers for ICU delirium prediction. AdaBoost: adaptive boosting; GBM: gradient-boosting machine; KNN: k-nearest neighbor; SVM: support vector machine.



Similarly, when compared with LDA, SVM (*P*=.006), neural network (*P*=.01), KNN (*P*=.007), LR (*P*=.007), and RF (*P*=.009) exhibited a significantly better performance. GBM (*P*=.04) and AdaBoost (*P*=.047) also showed a significantly better performance than that of LDA, whereas naïve Bayes showed a marginally better performance (*P*=.05). Decision tree demonstrated a comparable performance to that of LDA (*P*=.99).

In summary, LR was the most balanced model across all metrics, making it a preferred choice for clinical applications where both

XSL•FO
RenderX

precision and recall are critical. Ensemble methods such as GBM and AdaBoost offer strong alternatives, particularly for optimizing precision (Table 2).

Identification of High-Risk Factors for ICU Delirium Using SHAP Analysis

We used SHAP analysis to identify the top 15 high-risk factors for ICU delirium in each validated machine learning model (Multimedia Appendix 1). This comprehensive analysis revealed distinct sets of risk factors across different models, providing valuable insights into the complex nature of ICU delirium. Key findings included the following:

- LR (best overall performance; AUC=0.906; MCC=0.625): identified daily urine output, eosinophil count, age, and fibrinogen levels as key risk factors.
- Neural network (AUC=0.857; MCC=0.522): highlighted lactate dehydrogenase levels, daily urine output, inhalation injury, neutrophil count, and platelet count as key risk factors.
- SVM (AUC=0.897; MCC=0.157): uniquely identified burn area and length of ICU stay as significant factors.
- KNN (AUC=0.894): emphasized neutrophil and monocyte percentages as key risk factors.
- Decision tree: despite its lower performance, it identified endotracheal intubation and D-dimer levels as risk factors.
- Other models: consistently highlighted hematological parameters, daily urine output, and age across multiple models. Novel factors such as MetHb levels (LDA) and anion gap (RF) were also identified.

Discussion

Machine Learning–Based Prediction of ICU Delirium in Patients With Burns: LR Performance and SHAP Analysis for Risk Factor Identification

Our study demonstrated that LR effectively predicted ICU delirium in patients with burns using clinical data, including vital signs and blood biomarkers. Among the machine learning models, LR achieved the highest predictive accuracy, confirming that ICU delirium risk can be assessed through computational modeling (Figures 2 and 3).

Furthermore, SHAP analysis identified key risk factors associated with ICU delirium, highlighting T-bil, MetHb, daily urine output, and leucocyte fractions as novel predictors alongside established factors such as burn area and tracheal intubation (Multimedia Appendix 1). These findings provide crucial insights for developing targeted prevention and treatment strategies, emphasizing the importance of respiratory, renal, hepatic, and inflammatory function management in patients with burns who are critically ill.

The Role of T-Bil Levels in Predicting Delirium in ICU Patients With Burns: Connections With Cholestasis and Inflammation

T-bil levels were identified as a risk factor for the development of delirium in ICU patients with burns. However, the direct relationship between burns and cholestasis remains unclear. Cholestasis often occurs after burns, and patients with burns who have increased bilirubin levels without a corresponding increase in alkaline phosphatase and gamma-glutamyl transferase levels face a higher risk of mortality. Furthermore, intrahepatic cholestasis is observed in half of patients with severe burns [6]. Cholestasis is also associated with hypoxic hepatitis [7]. In addition, hypovolemic shock observed in severe burns may be involved in the elevation of T-bil levels [8]. Increases in interleukin (IL)-6 and tumor necrosis factor-α, which are observed in the early stages of severe burns [9], have been reportedly associated with hyperbilirubinemia [10] and organ dysfunction [11]. Inflammatory cytokines such as IL-6 and tumor necrosis factor-α may reduce the expression of bile transporters on the canalicular membrane of hepatocytes, leading to an increase in T-bil levels [12,13]. These research findings support the validity of our study, which identified an increase in T-bil levels as an important risk factor for predicting ICU delirium in patients with burns.

MetHb Level as a Novel Indicator of ICU Delirium in Patients With Burns: Insights From Sepsis and Hemolysis Research

MetHb level was identified as an explanatory factor. No previous studies have clearly demonstrated the involvement of MetHb level as a risk factor for ICU delirium in patients with burns. However, several studies have shown the involvement of MetHb in delirium in patients with sepsis, which, similarly to burns, can cause a cytokine storm. As is well known, patients with sepsis have a high incidence of delirium, and in these patients, nitric oxide is released into the bloodstream due to ischemia-reperfusion stimulation. Nitric oxide is converted into MetHb and nitrates; as a result, the concentration of MetHb in the blood is a useful marker for the onset of sepsis or septic shock [14]. However, the molecular mechanisms through which MetHb causes delirium in patients with sepsis remain largely unknown [15]. In patients with severe conditions, such as trauma or infection, intracellular hemolysis may occur, leading to anemia. Anemia, which develops relatively early in severe conditions, is thought to result from damaged red blood cells processed by the reticuloendothelial system. Acute hemolysis leads to an increase in free hemoglobin in the blood. Subsequently, free hemoglobin and heme are released into the circulatory system, and the wound interstitium is rapidly converted into MetHb by oxidants. The increase in MetHb levels is more pronounced in the ischemia-reperfusion areas, where activated macrophages and neutrophils accumulate [16]. Therefore, MetHb produced by ischemia-reperfusion injury and hemolysis, as observed in severe conditions such as sepsis, may affect leucocyte cell adhesion, phagocytic ability, and metabolic activation and may be involved in ICU delirium. Previous research has shown a stronger correlation among the total amount of hemoglobin [17,18], red blood cell count [19], and delirium. Thus, hemoglobin and MetHb levels in the blood are important factors influencing delirium.

The Impact of Decreased Urine Output on Delirium Risk in ICU Patients With Burns: Insights From SHAP Analysis

Using SHAP analysis, we identified that a decreased daily urine output within 24 hours of ICU admission is a risk factor for

XSL•FO
RenderX

ICU delirium in patients with burns. Interestingly, daily urine output was identified as an important risk factor in 50% (5/10) of the machine learning models evaluated (Multimedia Appendix 1). The finding that a decrease in 24-hour urine volume is a risk factor for delirium in patients with burns seems reasonable considering that acute kidney injury can potentially increase the risk of delirium 10-fold in patients who are critically ill [20]. For example, urine output is an important indicator of renal function and hydration status. In patients with burns, a decrease in urine output may indicate insufficient renal perfusion or dehydration. Both factors contribute to the development of delirium [21]. Furthermore, patients with burns often require multiple medications, such as sedatives and analgesics. Decreased urine output can affect the metabolism and excretion of these drugs, potentially leading to the accumulation of psychoactive substances that can induce or exacerbate delirium [22]. In addition, a decrease in urine output can lead to electrolyte imbalance, which is known to cause neurological dysfunction and delirium. Electrolyte abnormalities such as hyponatremia or hypernatremia can occur in patients with burns because of fluid shifts and inadequate fluid replacement [23].

Leukocyte Biomarkers as Indicators of Delirium in ICU Patients With Burns: The Role of Inflammatory Response

Our study identified the number or proportion of neutrophils and monocytes in the leucocyte fraction as risk factors for delirium development (Multimedia Appendix 1). This suggests that exposure of leucocytes to a cytokine storm due to excessive stress from burns contributes to delirium in patients with burns. Numerous studies have used leucocyte biomarkers to diagnose delirium in the past [24]. Inflammatory biomarkers and brain-specific metabolic biomarkers have been extensively studied in delirium, and inflammatory cytokines and activation markers of astrocytes and glial cells (IL-6, IL-8, IL-10, tumor necrosis factor-α, C-reactive protein, and S100β protein levels) positively correlate with longer duration of delirium, severity of delirium, and higher in-hospital mortality [25]. In addition, elevated levels of IL-8 and S100β protein have been associated with increased mortality in patients with delirium [26]. In a mouse model with delirium, the infiltration of bone marrow–derived monocytes into the blood-brain barrier [27] and activation of microglia were observed [28]. Although the number and proportion of neutrophils and monocytes were identified as risk factors for ICU delirium, these leucocytes may be involved in the production of inflammatory cytokines and contribute to the onset of ICU delirium.

Respiratory Rate as a Predictor of ICU Delirium in Patients With Burns: New Insights and Implications

In our study, an increase in the respiratory rate was a risk factor for ICU delirium. Patients with ICU delirium had a higher median respiratory rate of 19.5 (IQR 17.7-21.0) breaths per minute compared to 16.0 (IQR 14.0-20.0) breaths per minute in patients without delirium (P=.01; Table 1). To our knowledge, no studies have clearly established a link between delirium and respiratory rate. Delirium is generally recognized as a common complication in patients with respiratory failure in the ICU. The incidence of delirium in the ICU ranges from 10% to 78%, with most cases occurring in patients receiving mechanical ventilation. This suggests a significant overlap between respiratory complications and the occurrence of delirium; however, a direct correlation between an increase in respiratory rate and delirium has not been explicitly stated [29]. Inhalation injuries occur in approximately one-third of burn hospital admissions and contribute to a high mortality rate (50%) in patients with burns. Therefore, an increase in the respiratory rate may be associated with carbon monoxide poisoning and chemical tracheobronchitis due to the inhalation of toxic combustion products and generally correlates with a higher mortality rate. Unfortunately, many patients with burns receive high-concentration oxygen therapy from emergency teams before being transported to the ICU or emergency room. Therefore, carboxyhemoglobin levels are often adjusted to lower levels, and PO_2 is frequently high during treatment, which is why carboxyhemoglobin and PO_2 were not identified as risk factors in our model. Therefore, it might be appropriate to consider respiratory rate as a potential risk factor for ICU delirium in future studies [30,31].

Advancing ICU Delirium Research: The Prediction of Delirium in ICU Patients Model and the Need for Machine Learning Approaches in Patients With Burns

In the field of delirium research, the Prediction of Delirium in ICU Patients (PRE-DELIRIC) model is considered a seminal study [32]. In the ICU, the PRE-DELIRIC model uses 10 identified risk factors (age, Acute Physiology and Chronic Health Evaluation II score, admission group, coma, infection, metabolic acidosis, use of sedatives and morphine, blood urea nitrogen, and emergency admission) and predicts delirium with an AUC of 0.87 (95% CI 0.85-0.89) within 24 hours of ICU admission [32]. Furthermore, the model by Lanzhou University [33] heavily relies on patients' detailed past medical histories, making data collection challenging in busy clinical settings such as emergency rooms and ICUs, where obtaining comprehensive patient histories, diagnoses, and treatments can be difficult [34]. Therefore, it is crucial to establish machine learning models that can accurately predict conditions with multifactorial risk factors, such as ICU delirium, using data that are easily obtainable during emergency department visits, such as vital signs and blood data. Despite this need, a delirium prediction model for ICU patients with burns using machine learning has not yet been developed. Predictive models for diseases such as ICU delirium, which involve numerous risk factors, stand to benefit significantly from machine learning's capability to perform multifactorial analyses, surpassing traditional biostatistical methods. Our proposed machine learning model can more effectively evaluate complex interactions among multivariate data, which is essential for accurately predicting conditions with multifactorial risk factors such as ICU delirium. Therefore, although our study had a small number of cases, it is considered valuable for the development of a machine learning–based ICU delirium prediction model and the identification of risk factors.

Comprehensive Model Evaluation for ICU Delirium

When evaluating the performance of different models, it is important to consider not only the AUC but also other metrics,

XSL•FO
RenderX

such as accuracy, precision, recall, and F_1-score. In a specialized medical environment such as the ICU, some metrics may become more important than others when dealing with specific diseases. For example, precision is important if avoiding false detections of delirium is crucial. Conversely, if it is vital to avoid missing cases of delirium, recall should be emphasized. In our study, we considered all these metrics comprehensively and selected the model that best suited the objectives of the research and clinical demands.

Among the 10 models we adjusted, the LR model was found to be the most balanced and high performing. Notably, it demonstrated the highest values across key metrics, including AUC, MCC, accuracy, and F_1-score, suggesting that it provides the most reliable predictions from various perspectives.

Strengths

The explanatory variables in our study were based on blood data collected immediately upon the arrival of patients with burns at the emergency outpatient clinic. Therefore, there was a time lag between the collection of these data and the collection of data at the time of delirium diagnosis using the ICDSC. However, we believe that our model, which accurately predicts the development of delirium during an ICU stay based on blood data and vital signs at the time of patient arrival, can be easily interpreted by clinicians and has high general applicability. This allows clinicians to predict the incidence of ICU delirium, which significantly affects the prognosis of severe burns from early in the patient's hospitalization, enabling the initiation of early interventions for patients at high risk of ICU delirium.

Limitations

Demerit of Binary Classification for Delirium

A binary classification model was used for predicting delirium, which simplified the model's output to either *delirium present* or *delirium absent*. This approach facilitated decision-making in clinical settings by providing clear, binary outcomes. Conversely, using a regression model would require predicting specific ICDSC scores and interpreting these scores to assess patient status. Thus, binary classification often offers greater practicality in busy clinical environments due to its straightforward interpretation. However, regression analysis may be more appropriate for predicting continuous outcomes, such as the ICDSC score. This method offers detailed information by producing continuous values, allowing for a nuanced understanding of the severity of ICU delirium, ranging from mild to severe. For instance, the difference in delirium between an ICDSC score of 3 and 4 might be minimal, whereas the difference between scores of 1 and 7 indicates substantially different symptoms. Therefore, future research should investigate the benefits of developing regression models that predict ICDSC scores.

Model Selection Criteria for Small Datasets

This study examined delirium in patients with burns who were critically ill admitted to the ICU and did not include patients with missing data, resulting in a small number of cases. In analyses with small datasets, such as in our 82 cases, it is generally recommended to avoid overfitting and choose a simpler model. In this study, we were cautious about using complex models (such as deep neural networks or RF with many trees) that are prone to overfitting with limited data. Therefore, simple models are generally recommended for studies with few cases. Typically, models with regularization effects (such as LR) are effective in preventing overfitting. In our study, LR demonstrated a high accuracy (0.818), indicating that overfitting was well controlled during model creation. Linear SVMs have also been proposed to prevent overfitting. In addition, LDA has linear boundaries, is computationally fast, and can sometimes provide relatively stable results even with a small amount of data.

Limitations and Future Directions for External Validation

The performance of our model was discussed using data from patients in the ICU in a single hospital; however, an external validation was not conducted. Therefore, our study is limited by its single-institution setting. Future research could strengthen the reliability of our delirium prediction model for patients with burns in the ICU by conducting external validation using datasets from other hospitals.

Enhancing Predictive Power: Inclusion of Diverse Variables

Our machine learning model incorporated 70 explanatory variables, including patients' vital signs and blood biomarkers. However, unlike the PRE-DELIRIC model, which includes medication history and environmental factors, our study did not consider these variables. Expanding the range of input variables in future research may further enhance the predictive accuracy of delirium risk assessment.

Conclusions

Our study underscores the clinical utility of machine learning in predicting ICU delirium in patients with burns, demonstrating that LR provided the highest predictive accuracy among the tested models. Using SHAP analysis, we identified both well-established and novel risk factors, such as T-bil, MetHb, urine output, and leucocyte fractions, offering new insights into the complex pathophysiology of delirium. These findings suggest that early identification of patients at high risk using readily available clinical data upon ICU admission could facilitate proactive intervention strategies, potentially reducing morbidity and improving patient outcomes.

Beyond its immediate clinical applications, this study highlights the need for integrating machine learning into real-world ICU decision-making systems. Traditional delirium prediction models such as PRE-DELIRIC rely on a limited set of predefined variables, whereas machine learning models can dynamically incorporate diverse clinical parameters, enabling real-time risk stratification. This adaptability is particularly relevant in burn care, where patients exhibit highly variable and rapidly evolving physiological changes.

Moving forward, the integration of multi-institutional external validation is crucial to ensure the generalizability and robustness of our predictive model across diverse ICU settings. In addition, further research should explore the mechanistic pathways linking

XSL•FO
RenderX

identified risk factors with delirium onset, which could pave the way for personalized prevention and treatment strategies. Ultimately, this study provides a foundation for the next generation of delirium risk prediction models, emphasizing the potential of artificial intelligence–driven clinical decision support to enhance patient care in critical care medicine.

Acknowledgments

This study was supported by the Japan Society for the Promotion of Science (grants JP 22K06910, 22K09117, 22K09138, 19KK0224, 18K08916, and 21K09015). The authors want to thank their colleagues in the intensive care unit at Mie University Hospital. They also appreciate Dr Sakamoto, a PhD candidate in engineering, and his colleagues for their help in creating the machine learning model and Shapley additive explanations analysis. In addition, the authors benefited from the cooperation of the Dr Shimaoka Laboratory for Molecular Biology in biostatistics and molecular biological descriptions. They would like to express their gratitude to Dr Sakamoto and his team, who are experts in machine learning, for their invaluable contributions to improving the accuracy of the authors' machine learning models. In addition, they extend their thanks to the members of Dr Shimaoka Laboratory for Molecular Biology for their insightful comments on the manuscript, which significantly enhanced its quality.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the need to protect participant confidentiality but are available from the corresponding author on reasonable request. In addition, the Python code used for the machine learning analyses, including modifications and optimizations conducted using generative artificial intelligence, is provided in Multimedia Appendices 2-14. Further details regarding the code can be obtained by contacting the corresponding author.

Authors' Contributions

RE and EK were primarily responsible for manuscript preparation and drafting of the original text. MS oversaw the overall direction and planning of the research. The other coauthors, RS, AIM, FO, TS, AH, KI, YA, HF, AG, EJP, RM, and RK, provided supervision and were instrumental in shaping the research, analysis, and manuscript refinement.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comparison of feature Shapley additive explanations (SHAP) values across multiple machine learning models. The 15 most influential features in each model's predictions are displayed. A higher SHA*P* value indicates a stronger impact of the feature on the model's predictions.
[DOCX File , 23 KB - formative_v9i1e65190_app1.docx]

Multimedia Appendix 2
Python code used in this study for generating the violin plots comparing explanatory factors for intensive care unit delirium in patients with burns.
[DOCX File , 17 KB - formative_v9i1e65190_app2.docx]

Multimedia Appendix 3
Python code used in this study for generating the receiver operating characteristic curves illustrating machine learning model performance in predicting intensive care unit delirium.
[DOCX File , 15 KB - formative_v9i1e65190_app3.docx]

Multimedia Appendix 4
Python code used in this study for generating the bar graph and heat map comparing machine learning model performance in predicting intensive care unit delirium.
[DOCX File , 16 KB - formative_v9i1e65190_app4.docx]

Multimedia Appendix 5
Adaptive boosting code for Shapley additive explanations analysis.
[DOCX File , 14 KB - formative_v9i1e65190_app5.docx]

XSL•FO
RenderX

Multimedia Appendix 6
Decision tree code for Shapley additive explanations analysis.
[DOCX File , 15 KB - formative_v9i1e65190_app6.docx]

Multimedia Appendix 7
K-nearest neighbor code for Shapley additive explanations analysis.
[DOCX File , 15 KB - formative_v9i1e65190_app7.docx]

Multimedia Appendix 8
Linear discriminant analysis code for Shapley additive explanations analysis.
[DOCX File , 18 KB - formative_v9i1e65190_app8.docx]

Multimedia Appendix 9
Light gradient-boosting machine code for Shapley additive explanations analysis.
[DOCX File , 15 KB - formative_v9i1e65190_app9.docx]

Multimedia Appendix 10
Logistic regression code for Shapley additive explanations analysis.
[DOCX File , 17 KB - formative_v9i1e65190_app10.docx]

Multimedia Appendix 11
Naïve Bayes code for Shapley additive explanations analysis.
[DOCX File , 14 KB - formative_v9i1e65190_app11.docx]

Multimedia Appendix 12
Neural network code for Shapley additive explanations analysis.
[DOCX File , 17 KB - formative_v9i1e65190_app12.docx]

Multimedia Appendix 13
Random forest code for Shapley additive explanations analysis.
[DOCX File , 15 KB - formative_v9i1e65190_app13.docx]

Multimedia Appendix 14
Support vector machine code for Shapley additive explanations analysis.
[DOCX File , 16 KB - formative_v9i1e65190_app14.docx]

References

1. Watt J, Tricco AC, Talbot-Hamon C, Pham B, Rios P, Grudniewicz A, et al. Identifying older adults at risk of delirium following elective surgery: a systematic review and meta-analysis. J Gen Intern Med 2018 Apr 26;33(4):500-509 [FREE Full text] [doi: 10.1007/s11606-017-4204-x] [Medline: 29374358]
2. Devlin JW, Skrobik Y, Gélinas C, Needham DM, Slooter AJ, Pandharipande PP, et al. Clinical practice guidelines for the prevention and management of pain, agitation/sedation, delirium, immobility, and sleep disruption in adult patients in the ICU. Crit Care Med 2018 Sep;46(9):e825-e873 [FREE Full text] [doi: 10.1097/CCM.0000000000003299] [Medline: 30113379]
3. Agarwal V, O Neill PJ, Cotton BA, Pun BT, Haney S, Thompson J, et al. Prevalence and risk factors for development of delirium in burn intensive care unit patients. J Burn Care Res 2010;31(5):706-715. [doi: 10.1097/bcr.0b013e3181eebee9]
4. Guo Z, Liu J, Li J, Wang X, Guo H, Ma P, et al. Postoperative delirium in severely burned patients undergoing early escharotomy. J Burn Care Res 2017;38(1):e370-e376. [doi: 10.1097/bcr.0000000000000397]
5. Hur S, Ko RE, Yoo J, Ha J, Cha WC, Chung CR. A machine learning-based algorithm for the prediction of intensive care unit delirium (PRIDE): retrospective study. JMIR Med Inform 2021 Jul 26;9(7):e23401 [FREE Full text] [doi: 10.2196/23401] [Medline: 34309567]
6. de Tymowski C, Dépret F, Soussi S, Nabila M, Vauchel T, Chaussard M, et al. Contributing factors and outcomes of burn-associated cholestasis. J Hepatol 2019 Sep;71(3):563-572. [doi: 10.1016/j.jhep.2019.05.009] [Medline: 31152758]
7. Jäger B, Drolz A, Michl B, Schellongowski P, Bojic A, Nikfardjam M, et al. Jaundice increases the rate of complications and one-year mortality in patients with hypoxic hepatitis. Hepatology 2012 Dec 24;56(6):2297-2304. [doi: 10.1002/hep.25896] [Medline: 22706920]

XSL•FO
RenderX

8. Soussi S, Dépret F, Benyamina M, Legrand M. Early hemodynamic management of critically ill burn patients. Anesthesiology 2018 Sep;129(3):583-589. [doi: 10.1097/ALN.0000000000002314] [Medline: 29958240]
9. Csontos C, Foldi V, Pálinkas L, Bogar L, Röth E, Weber G, et al. Time course of pro- and anti-inflammatory cytokine levels in patients with burns--prognostic value of interleukin-10. Burns 2010 Jun;36(4):483-494. [doi: 10.1016/j.burns.2009.10.009] [Medline: 20045261]
10. Jones A, Selby PJ, Viner C, Hobbs S, Gore ME, McElwain TJ. Tumour necrosis factor, cholestatic jaundice, and chronic liver disease. Gut 1990 Aug 01;31(8):938-939 [FREE Full text] [doi: 10.1136/gut.31.8.938] [Medline: 2387521]
11. Sprague AH, Khalil RA. Inflammatory cytokines in vascular dysfunction and vascular disease. Biochem Pharmacol 2009 Sep 15;78(6):539-552 [FREE Full text] [doi: 10.1016/j.bcp.2009.04.029] [Medline: 19413999]
12. Whiting JF, Green RM, Rosenbluth AB, Gollan JL. Tumor necrosis factor-alpha decreases hepatocyte bile salt uptake and mediates endotoxin-induced cholestasis. Hepatology 1995 Oct;22(4 Pt 1):1273-1278. [doi: 10.1016/0270-9139(95)90639-8] [Medline: 7557881]
13. Green RM, Whiting JF, Rosenbluth AB, Beier D, Gollan JL. Interleukin-6 inhibits hepatocyte taurocholate uptake and sodium-potassium-adenosinetriphosphatase activity. Am J Physiol Gastrointest Liver Physiol 1994 Dec 01;267(6):G1094-G1100. [doi: 10.1152/ajpgi.1994.267.6.g1094]
14. Ohashi K, Yukioka H, Hayashi M, Asada A. Elevated methemoglobin in patients with sepsis. Acta Anaesthesiol Scand 1998 Jul 31;42(6):713-716. [doi: 10.1111/j.1399-6576.1998.tb05306.x] [Medline: 9689279]
15. Atterton B, Paulino MC, Povoa P, Martin-Loeches I. Sepsis associated delirium. Medicina (Kaunas) 2020 May 18;56(5):240 [FREE Full text] [doi: 10.3390/medicina56050240] [Medline: 32443606]
16. Schaer DJ, Buehler PW, Alayash AI, Belcher JD, Vercellotti GM. Hemolysis and free hemoglobin revisited: exploring hemoglobin and hemin scavengers as a novel class of therapeutic proteins. Blood 2013 Feb 21;121(8):1276-1284 [FREE Full text] [doi: 10.1182/blood-2012-11-451229] [Medline: 23264591]
17. Kijima E, Kayama T, Saito M, Kurosaka D, Ikeda R, Hayashi H, et al. Pre-operative hemoglobin level and use of sedative-hypnotics are independent risk factors for post-operative delirium following total knee arthroplasty. BMC Musculoskelet Disord 2020 May 02;21(1):279 [FREE Full text] [doi: 10.1186/s12891-020-03206-4] [Medline: 32359366]
18. Liu YM, Huang H, Gao J, Zhou J, Chu HC. Hemoglobin concentration and post-operative delirium in elderly patients undergoing femoral neck fracture surgery. Front Med (Lausanne) 2021 Jan 5;8:780196 [FREE Full text] [doi: 10.3389/fmed.2021.780196] [Medline: 35071265]
19. Prasko J, Horacek R, Krnacova B, Latalova K. Delirium as a complication of the surgical intensive care. Neuropsychiatr Dis Treat 2016 Sep;Volume 12:2425-2434. [doi: 10.2147/ndt.s115800]
20. Zipser CM, Deuel J, Ernst J, Schubert M, Weller M, von Känel R, et al. Predisposing and precipitating factors for delirium in neurology: a prospective cohort study of 1487 patients. J Neurol 2019 Dec 13;266(12):3065-3075. [doi: 10.1007/s00415-019-09533-4] [Medline: 31520105]
21. Pang H, Kumar S, Ely EW, Gezalian MM, Lahiri S. Acute kidney injury-associated delirium: a review of clinical and pathophysiological mechanisms. Crit Care 2022 Aug 27;26(1):258 [FREE Full text] [doi: 10.1186/s13054-022-04131-9] [Medline: 36030220]
22. Wan R, McKenzie CA, Taylor D, Camporota L, Ostermann M. Acute kidney injury as a risk factor of hyperactive delirium: a case control study. J Crit Care 2020 Feb;55:194-197. [doi: 10.1016/j.jcrc.2019.10.013] [Medline: 31756583]
23. Arieff AI, Massry SG. Calcium metabolism of brain in acute renal failure. Effects of uremia, hemodialysis, and parathyroid hormone. J Clin Invest 1974 Feb 1;53(2):387-392. [doi: 10.1172/jci107571]
24. Toft K, Tontsch J, Abdelhamid S, Steiner L, Siegemund M, Hollinger A. Serum biomarkers of delirium in the elderly: a narrative review. Ann Intensive Care 2019 Jul 01;9(1):76 [FREE Full text] [doi: 10.1186/s13613-019-0548-1] [Medline: 31263968]
25. Khan BA, Perkins AJ, Prasad NK, Shekhar A, Campbell NL, Gao S, et al. Biomarkers of delirium duration and delirium severity in the ICU. Crit Care Med 2020 Mar;48(3):353-361 [FREE Full text] [doi: 10.1097/CCM.0000000000004139] [Medline: 31770149]
26. Khan BA, Perkins AJ, Gao S, Hui SL, Campbell NL, Farber MO, et al. The confusion assessment method for the ICU-7 delirium severity scale: a novel delirium severity instrument for use in the ICU. Crit Care Med 2017 May;45(5):851-857 [FREE Full text] [doi: 10.1097/CCM.0000000000002368] [Medline: 28263192]
27. Terrando N, Eriksson LI, Ryu JK, Yang T, Monaco C, Feldmann M, et al. Resolving postoperative neuroinflammation and cognitive decline. Ann Neurol 2011 Dec 21;70(6):986-995 [FREE Full text] [doi: 10.1002/ana.22664] [Medline: 22190370]
28. Feng X, Valdearcos M, Uchida Y, Lutrin D, Maze M, Koliwad SK. Microglia mediate postoperative hippocampal inflammation and cognitive decline in mice. JCI Insight 2017 Apr 06;2(7):e91229 [FREE Full text] [doi: 10.1172/jci.insight.91229] [Medline: 28405620]
29. Fu X, Wang L, Wang G, Liu X, Wang X, Ma S, et al. Delirium in elderly patients with COPD combined with respiratory failure undergoing mechanical ventilation: a prospective cohort study. BMC Pulm Med 2022 Jul 09;22(1):266 [FREE Full text] [doi: 10.1186/s12890-022-02052-5] [Medline: 35810306]
30. Won YH, Cho YS, Joo SY, Seo CH. Respiratory characteristics in patients with major burn injury and smoke inhalation. J Burn Care Res 2022 Jan 05;43(1):70-76. [doi: 10.1093/jbcr/irab085] [Medline: 34142710]

XSL•FO
RenderX

31. Żwierełło W, Piorun K, Skórka-Majewicz M, Maruszewska A, Antoniewski J, Gutowska I. Burns: classification, pathophysiology, and treatment: a review. Int J Mol Sci 2023 Feb 13;24(4):3749 [FREE Full text] [doi: 10.3390/ijms24043749] [Medline: 36835171]
32. van den Boogaard M, Pickkers P, Slooter AJ, Kuiper MA, Spronk PE, van der Voort PH, et al. Development and validation of PRE-DELIRIC (PREdiction of DELIRium in ICu patients) delirium prediction model for intensive care patients: observational multicentre study. BMJ 2012 Feb 09;344(feb09 3):e420 [FREE Full text] [doi: 10.1136/bmj.e420] [Medline: 22323509]
33. Chen Y, Du H, Wei BH, Chang XN, Dong CM. Development and validation of risk-stratification delirium prediction model for critically ill patients: a prospective, observational, single-center study. Medicine (Baltimore) 2017 Jul 17;96(29):e7543-e7590 [FREE Full text] [doi: 10.1097/MD.0000000000007543] [Medline: 28723773]
34. Green C, Bonavia W, Toh C, Tiruvoipati R. Prediction of ICU delirium: validation of current delirium predictive models in routine clinical practice. Crit Care Med 2019 Mar;47(3):428-435. [doi: 10.1097/CCM.0000000000003577] [Medline: 30507844]

Abbreviations

AdaBoost: adaptive boosting
AUC: area under the receiver operating characteristic curve
CAM-ICU: Confusion Assessment Method for the Intensive Care Unit
GBM: gradient-boosting machine
ICDSC: Intensive Care Delirium Screening Checklist
ICU: intensive care unit
IL: interleukin
LDA: linear discriminant analysis
LR: logistic regression
MCC: Matthews correlation coefficient
MetHb: methemoglobin
PRE-DELIRIC: Prediction of Delirium in Intensive Care Unit Patients
PT: prothrombin time
RF: random forest
SHAP: Shapley additive explanations
SVM: support vector machine
T-bil: total bilirubin

Edited by A Mavragani; submitted 08.08.24; peer-reviewed by J Nalepa; comments to author 19.09.24; revised version received 06.01.25; accepted 03.02.25; published 05.03.25.

Please cite as:
Esumi R, Funao H, Kawamoto E, Sakamoto R, Ito-Masui A, Okuno F, Shinkai T, Hane A, Ikejiri K, Akama Y, Gaowa A, Park EJ, Momosaki R, Kaku R, Shimaoka M
Machine Learning–Based Prediction of Delirium and Risk Factor Identification in Intensive Care Unit Patients With Burns: Retrospective Observational Study
JMIR Form Res 2025;9:e65190
URL: https://formative.jmir.org/2025/1/e65190
doi:10.2196/65190
PMID:39895101

©Ryo Esumi, Hiroki Funao, Eiji Kawamoto, Ryota Sakamoto, Asami Ito-Masui, Fumito Okuno, Toru Shinkai, Atsuya Hane, Kaoru Ikejiri, Yuichi Akama, Arong Gaowa, Eun Jeong Park, Ryo Momosaki, Ryuji Kaku, Motomu Shimaoka. Originally published in JMIR Formative Research (https://formative.jmir.org), 05.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Crisis Communication About the Maui Wildfires on TikTok: Content Analysis of Engagement With Maui Wildfire–Related Posts Over 1 Year

Jim P Stimpson[1,2], PhD; Aditi Srivastava[3]; Ketan Tamirisa[4]; Joseph Keawe aimoku Kaholokula[5], PhD; Alexander N Ortega[6], PhD

[1]Department of Health Economics, Systems, and Policy, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX, United States

[2]Peter O'Donnell Jr. School of Public Health, University of Texas Southwestern Medical Center, Dallas, TX, United States

[3]Frisco High School, Frisco, TX, United States

[4]Washington University in St. Louis, St Louis, MO, United States

[5]Department of Native Hawaiian Health, John A. Burns School of Medicine, University of Hawai i at Mānoa, Honolulu, HI, United States

[6]Thompson School of Social Work & Public Health, University of Hawai i at Mānoa, Honolulu, HI, United States

Corresponding Author:
Jim P Stimpson, PhD
Department of Health Economics, Systems, and Policy, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX, United States

Abstract

Background: The August 2023 wildfire in the town of Lāhainā on the island of Maui in Hawai i caused catastrophic damage, affecting thousands of residents, and killing 102 people. Social media platforms, particularly TikTok, have become essential tools for crisis communication during disasters, providing real-time crisis updates, mobilizing relief efforts, and addressing misinformation. Understanding how disaster-related content is disseminated and engaged with on these platforms can inform strategies for improving emergency communication and community resilience.

Objective: Guided by Social-Mediated Crisis Communication theory, this study examined TikTok posts related to the Maui wildfires to assess content themes, public engagement, and the effectiveness of social media in disseminating disaster-related information.

Methods: TikTok posts related to the Maui wildfires were collected from August 8, 2023, to August 9, 2024. Using TikTok's search functionality, we identified and reviewed public posts that contained relevant hashtags. Posts were categorized into 3 periods: during the disaster (August 8 to August 31, 2023), the immediate aftermath (September 1 to December 31, 2023), and the long-term recovery (January 1 to August 9, 2024). Two researchers independently coded the posts into thematic categories, achieving an interrater reliability of 87%. Engagement metrics (likes and shares) were analyzed to assess public interaction with different themes. Multivariable linear regression models were used to examine the associations between log-transformed likes and shares and independent variables, including time intervals, video length, the inclusion of music or effects, content themes, and hashtags.

Results: A total of 275 TikTok posts were included in the analysis. Most posts (132/275, 48%) occurred in the immediate aftermath, while 76 (27.6%) were posted during the long-term recovery phase, and 24.4% (n=67) were posted during the event. Posts during the event garnered the highest average number of likes (mean 75,092, SD 252,759) and shares (mean 10,928, SD 55,308). Posts focused on "Impact & Damage" accounted for the highest engagement, representing 36.8% (4,090,574/11,104,031) of total likes and 61.2% (724,848/1,184,049) of total shares. "Tourism Impact" (2,172,991/11,104,031, 19.6% of likes; 81,372/1,184,049, 6.9% of shares) and "Relief Efforts" (509,855/11,104,031, 4.6% of likes; 52,587/1,184,049, 4.4% of shares) were also prominent themes. Regression analyses revealed that videos with "Misinformation & Fake News" themes had the highest engagement per post, with a 4.55 coefficient for log-shares (95% CI 2.44-6.65), while videos about "Tourism Impact" and "Relief Efforts" also showed strong engagement (coefficients for log-likes: 2.55 and 1.76, respectively).

Conclusions: TikTok is an influential tool for disaster communication, amplifying both critical disaster updates and misinformation, highlighting the need for strategic content moderation and evidence-based messaging to enhance the platform's role in crisis response. Public health officials, emergency responders, and policy makers can leverage TikTok's engagement patterns to optimize communication strategies, improve real-time risk messaging, and support long-term community resilience.

(JMIR Form Res 2025;9:e67515) doi:10.2196/67515

XSL•FO
RenderX

KEYWORDS

social media; public health; disasters; Hawaii; media; post; communication; disaster; disaster communication; wildfire; information; dissemination; engagement; content analysis; content; metrics; misinformation; community; support

Introduction

Background

The wildfire that began on August 8, 2023, in the town of Lāhainā on Maui island in Hawai i brought widespread destruction, disrupted the lives of residents, destroyed homes, killed 102 people, and altered the environment [1-4]. Wildfires are increasingly becoming a serious public health concern due to their devastating impacts on the environment, communities, and health infrastructure [5-7]. The growing frequency and intensity of wildfires have been, in part, attributed to climate change, which creates hotter and drier conditions conducive to fire outbreaks [8-10]. Understanding public perceptions and engagement with these disasters is crucial for enhancing response, recovery, and resiliency [11].

In the digital age, social media platforms have emerged as critical tools for crisis communication, particularly during natural disasters and public health emergencies, providing situational updates, relief information, emotional support, and avenues for public mobilization [12-19]. TikTok has emerged as a key platform for public engagement during crises, offering a unique short-form video format that encourages personal storytelling and emotional connection [20-26]. TikTok's algorithm promotes high-engagement posts, which amplifies diverse perspectives. Previous studies, such as Sharma and De Choudhury's [20] exploration of TikTok in public health, underscore its value in disseminating information and fostering community resilience. However, limited research has focused on its role during natural disasters, especially compared with platforms like Facebook or Instagram [13,14]. The rise of mis and disinformation on social media also poses challenges, which potentially complicates disaster response and relief efforts [27-31].

Despite TikTok's increasing influence, few studies have focused on its role in communicating and shaping public discourse during natural disasters. One study analyzed TikTok videos from official accounts featured in the COVID-19 information hub, focusing on how credible information was disseminated to combat misinformation during the pandemic, highlighting the platform's role in delivering accessible, authoritative public health messaging [32]. Another examined trending TikTok content during the early days of the US COVID-19 pandemic, emphasizing how the platform facilitated social connection and coping through creative and relatable content during the "shelter in place" period [33]. A third study evaluated TikTok videos tagged with #ClimateChange to assess the platform's use in raising awareness and driving discussions on environmental issues [34]. Collectively, these studies underscore TikTok's growing importance as a tool for disseminating information, fostering public engagement, and influencing perceptions on global challenges but more information is needed about its use during natural disasters.

This study applies Social-Mediated Crisis Communication (SMCC) theory to understand how TikTok was used to disseminate and engage with information about the August 2023 Maui wildfires. SMCC theory explains how people seek, share, and respond to crisis information in digital environments [35]. SMCC theory posits that during a crisis, individuals rely on traditional media, social media influencers, and peer networks to make sense of unfolding events. The theory also emphasizes the role of organizational crisis communicators, but in the case of decentralized platforms like TikTok, crisis-related information is largely driven by user-generated content rather than official sources. This shift highlights the need to analyze which themes and message characteristics drive engagement on social media during disasters [35,36].

Applying SMCC theory, this study examined how content themes, video features (eg, video length, music, and effects), and hashtags influence engagement metrics, including likes and shares. Specifically, we assessed whether certain types of crisis messages align with SMCC theory's key information-sharing mechanisms. Instructing information refers to messages that provide practical guidance or factual updates, such as evacuation notices and safety alerts. Adjusting information includes messages that help users emotionally process the crisis, including personal narratives and community responses. Bolstering information consists of messages that reinforce community solidarity, fundraising efforts, and expressions of resilience. Prior research indicates that adjusting and bolstering messages often garner higher engagement on social media platforms, as users seek emotional support and opportunities to contribute to recovery efforts [36-38]. For instance, a study analyzing social media interactions during crises found that content providing emotional support and community solidarity received substantial user engagement [37]. Conversely, while instructing information—such as safety alerts and factual updates—is crucial for public safety, it tends to receive comparatively lower engagement, especially on entertainment-focused platforms like TikTok [23-26]. Research on TikTok's user engagement patterns suggests that content aligning with entertainment and emotional resonance is more likely to engage users than purely informational posts [38].

By applying SMCC theory, this study seeks to identify which crisis communication strategies gain traction on TikTok, offering insights into how social media platforms shape public understanding and response during natural disasters. The findings have implications for emergency responders, public health officials, and community organizations aiming to optimize social media engagement in future crises. Natural disasters, like the Maui wildfire, are typically seen as external and uncontrollable, which places greater emphasis on the need for empathetic and supportive communication rather than blame attribution [35-38]. Understanding how social media users adapt these strategies in real time contributes to advancing best practices for disaster communication [12-15].

XSL•FO
RenderX

Research Questions

Guided by SMCC theory, this study examined public engagement with TikTok content related to the Maui wildfire to understand its role in digital health communication during disasters. Specifically, we address three key questions: (1) how do TikTok posts related to the Maui wildfire align with SMCC's instructing, adjusting, and bolstering information-sharing mechanisms and what are the implications for public health messaging during crises; (2) which content themes generate the highest engagement among TikTok users throughout different phases of the disaster and how do these themes contribute to public health awareness, misinformation spread, and community resilience; and (3) how does engagement evolve over time from the immediate disaster response to the long-term recovery period and what does this reveal about the role of social media for public health communication after the immediate danger of crisis is past? By examining these research questions, this study contributes to the growing body of literature on digital health communication in disaster contexts, focusing on TikTok as an influential platform shaping disaster-related communication strategies. Understanding public engagement patterns across distinct temporal phases provides actionable insights for public health agencies, emergency responders, and policy makers to strengthen community resilience, improve evidence-based health messaging, and optimize the use of social media platforms for disaster response and recovery.

Methods

Data Collection

We collected TikTok posts about the Maui wildfire from August 8, 2023, to August 9, 2024, covering both the immediate aftermath of the wildfire and the 1-year anniversary. This timeframe was chosen to align with the phases emphasized in SMCC theory, which highlights the evolving nature of crisis communication on social media, emphasizing both the acute crisis period and the extended postcrisis discourse that shapes public understanding of response and recovery [35]. Posts were identified using a combination of disaster-related hashtags, including #MauiStrong and #MauiWildfires, and were sampled until thematic saturation was achieved [39].

To ensure diverse and representative content, posts were selected at random from the top search results for each hashtag or hashtag combination, prioritizing those with high visibility. For each hashtag, all available posts were thoroughly reviewed and included if they were relevant to the wildfire. Duplicate posts were removed to prevent redundancy and posts unrelated to the wildfire were excluded. After reaching thematic saturation, the final data set consisted of 275 unique posts, which was sufficient to capture the breadth of content shared about the wildfire [39].

Measures

The outcome variables for this study were likes and shares, which represent the primary engagement metrics for TikTok posts. Likes captured the number of times users expressed approval of a post, and shares indicated how often a post was redistributed by users to others on the platform. To address outliers in these engagement metrics, we applied a log transformation to both likes and shares. This transformation helped normalize the distribution of the variables and reduced the influence of extreme values, ensuring a more robust analysis. The log-transformed values were used as dependent variables in the regression models to evaluate factors influencing engagement.

Posts were coded with a variety of themes, and each post could be coded with multiple themes. Two researchers independently coded the posts to ensure reliability and minimize bias. Discrepancies in coding were resolved through discussion and consensus. The level of agreement between the 2 researchers was measured using the κ reliability statistic, which resulted in a high agreement rate of 87%. This high level of interrater reliability confirms the consistency and accuracy of the thematic classification [40].

We identified several major themes in the TikTok posts related to the Maui wildfire, each reflecting different aspects of the event and its aftermath. One prominent theme was "Impact & Damage", with posts documenting the destruction caused by the wildfires. These posts often included powerful visuals, such as videos and images of affected areas, illustrating the extent of the devastation. The "Personal Narratives & Interviews" theme featured individuals sharing their experiences of living through the wildfire. These posts were often deeply emotional, with appeals for continued attention to the needs of those affected. Personal stories ranged from immediate reactions during the wildfire to reflections on the ongoing recovery process, amplifying the voices of those directly impacted by the disaster. The theme of "Community, Solidarity, & Tribute" was also widely represented. These posts conveyed messages of support and solidarity from people around the world. Many stories highlighted local heroes, community efforts, and the resilience of those affected by the wildfire. A common subtheme within these posts was the presentation of "before the fire" scenes, which offered a sentimental and solemn perspective on what was lost and underscored the emotional impact of the disaster.

Another key theme was "Relief Efforts", which encompassed content aimed at promoting fundraisers, donations, and other forms of assistance. Posts in this category highlighted the work of both local and national organizations providing aid to affected communities. They also documented the progress (or lack thereof) in recovery efforts, which served as calls to action for further support from the public. These posts were integral in mobilizing resources and raising awareness about ongoing needs in the aftermath of the wildfire. Posts related to "Government & Policy Responses" critiqued or praised the governmental handling of the wildfire. Discussions centered on the effectiveness of emergency preparedness and response efforts, with users offering their perspectives on the adequacy of relief efforts, resource allocation, and communication from officials during and after the crisis. "Tourism Impact" was another important theme.

As Maui is a major tourist destination and most tourists to the island visit Lāhainā, many posts discussed how the wildfires affected the tourism industry. Both travelers and locals shared their thoughts on the changing landscape and economy,

XSL•FO
RenderX

reflecting on how the wildfires disrupted the flow of visitors and altered the island's economic prospects. Posts categorized under "Environmental & Climate Commentary" explored the role of climate change in exacerbating wildfires. Users discussed how changing weather patterns and prolonged droughts have contributed to the frequency and intensity of such natural disasters. Many of these posts also advocated for policy changes aimed at mitigating future risks, calling for stronger environmental protections and climate action to prevent similar events from occurring.

Finally, posts under the "Informational" theme provided detailed overviews of the wildfire. These posts explained the causes, effects, and other key details of the disaster, often serving as educational content for users looking to understand the broader context and implications of the wildfire. These posts helped disseminate factual and descriptive information about the event. In contrast, the theme of "Misinformation & Fake News" was also present. Posts categorized as "Misinformation" were identified based on 2 criteria: posts that discussed and attempted to debunk misinformation about the fires by explaining why the claims were inaccurate and posts that directly promoted misinformation, with the accuracy of their claims verified through fact-checking conducted by our research team. Any post meeting either of these criteria was classified as "Misinformation."

We included several other measures to assess how structural and temporal characteristics of posts influenced engagement levels (Likes and Shares). Video length, representing the duration of each TikTok post, was measured in seconds and included in the analyses as a continuous variable. A squared term for video length was also included to account for potential nonlinearity in its relationship with engagement metrics. Additionally, the presence of music or effects was included as a binary variable, capturing whether a post used TikTok's music or special effects features. Event timing was categorized into 3 periods: during the disaster (August 8 to August 31, 2023), the immediate aftermath (September 1 to December 31, 2023), and the long-term recovery (January 1 to August 9, 2024). Posts were coded based on their publication date relative to these time frames. This segmentation aligns with SMCC's theoretical framework, enabling insights into how content types and engagement patterns shift over time [35].

We categorized individual hashtags into thematic groups to analyze engagement patterns related to the Maui wildfire. Each hashtag group was designed to capture distinct aspects of public discourse and response. Posts were assigned to groups based on predefined hashtags, with a binary variable indicating whether a post contained at least 1 relevant hashtag. The "Support" group included hashtags such as #MauiStrong, #PrayForMaui, and #MauiSupport, capturing expressions of solidarity and general support for those affected. The "Relief" group encompassed hashtags like #DonateForMaui, #Fundraiser, #MauiRelief, #DisasterRelief, and #RedCross, reflecting calls for donations, fundraising efforts, and relief activities. The "WildfireContent" group included hashtags such as #MauiWildfires, #Wildfire, #WildfireAwareness, and #FirePrevention, focusing on wildfire-specific content, including awareness and prevention efforts. The "SafetyAlerts" group comprised hashtags like #StaySafe and #EmergencyAlert, emphasizing safety information and emergency updates. The "MauiLocations" group included hashtags referencing specific locations such as #Lahaina, #Kaanapali, #WestMaui, and #Hawaii, highlighting geographically specific content. The "AdvocacyAndUpdates" group encompassed hashtags such as #EnvironmentalJustice, #NewsUpdate, and #BreakingNews, reflecting advocacy-related posts and updates on unfolding events. Each group was assigned a binary variable, where a post was flagged as belonging to a group if it contained at least one of the associated hashtags. Logical functions in spreadsheet software were used to implement this classification efficiently, which ensured consistent assignment of posts to their respective groups.

Data Analysis

Engagement metrics were analyzed to assess immediate reactions and the reach of content to viewers [41,42]. The analysis strategy began with descriptive statistics to provide an overview of the distribution of engagement metrics (likes and shares) across key categories, such as time intervals and hashtag groups. We summarized engagement metrics by time interval, showing the number of posts, average likes and shares, and their distribution across the event duration, immediate aftermath, and long-term recovery phases. Crosstabulations of hashtags and content themes were conducted to examine how engagement metrics varied across different combinations of hashtag groups and content themes. These tables found in Multimedia Appendix 1 highlighted the total and relative contributions of each category to overall engagement.

The multivariable analyses included linear regression models predicting the log-transformed likes and shares. Independent variables included time intervals, video length (and its squared term), the presence of music or effects, content themes, and hashtag groups. This approach assessed the relative contribution of these variables to engagement metrics while adjusting for potential confounding factors. To test the robustness of the results, we replicated the linear regression models using robust regression, which is less sensitive to outliers and provides an alternative estimation strategy. The results of the robustness check in Multimedia Appendix 2 were largely consistent with the primary analyses, confirming the stability of key findings. This study adheres to the STROBE (Strengthening the Reporting of Observational studies in Epidemiology) guidelines for reporting observational studies.

Ethical Considerations

The University of Hawai i's Office of Research Compliance determined that the study does not meet the definition of human participants research and therefore does not require institutional review board approval or oversight.

Results

The analysis shown in Table 1 examined social media engagement metrics across the event time intervals. A total of 275 observations were included in the analysis. Most posts (132/275, 48%) occurred in the immediate aftermath, while 76 (27.6%) were posted during the long-term recovery phase, and

XSL•FO
RenderX

24.4% (n=67) were posted during the event. Posts during the event garnered the highest average number of likes, with a mean of 75,092 (SD 252,759). Engagement declined in subsequent periods, with posts in the "Immediate Aftermath" receiving an average of 40,641 likes, and posts in the "Long-Term Recovery" averaging 8685 likes. Across all intervals, the average number of likes was 40,203. Similarly, posts during the event achieved the highest average number of shares, with a mean of 10,928 (SD 55,308). This figure dropped to 3111 in the "Immediate Aftermath" and further declined to 493 in the "Long-Term Recovery" period. The overall average number of shares across all intervals was 4292.

Table . Engagement metrics (likes and shares) by time interval. Content analysis of 275 TikTok posts from August 8, 2023, to August 9, 2024.

	Value, n (%)	Likes, mean (SD)	Shares, mean (SD)
During event (August 8 to August 31, 2023)	67 (24.4)	75,092 (252,759)	10,928 (55,308)
Immediate aftermath (September 1 to December 31, 2023)	132 (48.0)	40,641 (184,994)	3111 (10,365)
Long-term recovery (January 1 to August 9, 2024)	76 (27.6)	8685 (27,244)	493 (1,207)
Total	275 (100.0)	124,418	14,532

Table 2 presents the results from the linear regression analysis of log-transformed likes. Video length was positively associated with likes, with the linear term (coefficient=0.01, *P*=.05) indicating a slight increase in engagement for longer videos. However, the quadratic term (coefficient=0.00, *P*=.03) suggests diminishing returns for extremely long videos. Event timing showed that engagement was significantly lower during the immediate aftermath of the disaster (coefficient=−0.96, *P*=.01) and during the long-term recovery phase (coefficient=−1.47, *P*<.001) compared with the disaster period. Videos featuring music or effects had a marginally lower engagement, though this relationship was not statistically significant (coefficient=−0.56, *P*=.09). Regarding content themes, several categories significantly increased engagement. Videos focused on "Impact & Damage" (coefficient=1.55, *P*=.01), "Informational" content (coefficient=1.32, *P*=.04), "Misinformation & Fake News" (coefficient=3.24, *P*<.001), "Personal Narratives & Interviews" (coefficient=1.66, *P*=.02), "Relief Efforts" (coefficient=1.62, *P*=.02), and "Tourism Impact" (coefficient=2.53, *P*<.001) all received higher levels of likes compared with the reference group. In contrast, themes such as "Environmental & Climate Commentary" and "Government & Policy Response" did not significantly impact engagement. For hashtags, the "Support" and "WildfireContent" groups were associated with significantly higher engagement, with coefficients of 1.90 (*P*<.001) and 1.87 (*P*<.001), respectively, compared with the reference group. However, the "MauiLocations" and "Relief" hashtags did not exhibit significant effects.

XSL•FO
RenderX

Table . Multivariable linear regression of log-transformed likes. Content analysis of 275 TikTok posts from August 8, 2023, to August 9, 2024.

		Coefficient	*P* value	95% CI
Video length		0.01	.05	0.00 to 0.02
Video length square		0.00	.03	0.00 to 0.00
Event timing				
	During disaster	—[a]	—	—
	Immediate aftermath	−0.96	.01	−1.71 to −0.21
	Long-term recovery	−1.47	<.001	−2.31 to −0.63
Music and effects		−0.56	.09	−1.21 to 0.09
Content theme				
	Community, Solidarity, & Tribute	—	—	—
	Environmental & Climate Commentary	0.04	.97	−2.19 to 2.27
	Government & Policy Response	0.95	.23	−0.61 to 2.51
	Impact & Damage	1.55	.01	0.33 to 2.77
	Informational	1.32	.04	0.05 to 2.58
	Misinformation & Fake News	3.24	<.001	1.22 to 5.26
	Personal Narratives & Interviews	1.66	.02	0.32 to 3.00
	Relief Efforts	1.62	.02	0.23 to 3.00
	Tourism Impact	2.53	<.001	1.02 to 4.05
Hashtag group				
	AdvocacyAndUpdates	—	—	—
	MauiLocations	0.00	>0.99	−0.89 to 0.90
	Relief	−0.93	.12	−2.09 0.24
	Support	1.90	<.001	0.97 to 2.84
	WildfireContent	1.87	<.001	0.95 to 2.78
Constant		6.02	<.001	4.55 to 7.49

[a]Not applicable (reference category).

Table 3 presents the linear regression analysis for log-transformed shares. Video length was not significantly associated with shares. The Immediate Aftermath period was negatively associated with shares (coefficient=−0.82, *P*=.04), and long-term recovery showed an even stronger negative effect (coefficient=−1.51, *P*<.001), both indicating fewer shares during these periods compared with the baseline (during event). Content themes also had significant effects on shares. Specifically, “Impact & Damage,” “Informational,” and “Misinformation & Fake News” content themes had large positive associations with shares (coefficients of 2.46, 1.86, and 4.55, respectively, all with *P* values <.01), suggesting that content related to these themes is shared more frequently. “Personal Narratives & Interviews,” “Relief Efforts,” and “Tourism Impact” also showed significant positive associations with shares, further emphasizing the popularity of content related to these themes. Finally, certain hashtag groups influenced shares. “Relief” had a negative impact on shares (coefficient=−1.68, *P*=.01), while “Support” and “WildfireContent” both had positive effects (coefficients of 2.28 and 1.64, respectively, both with *P* values <.01), indicating that these types of hashtags contribute to higher share rates relative to the reference group.

XSL•FO
RenderX

Table . Multivariable linear regression of log-transformed shares. Content analysis of 275 TikTok posts from August 8, 2023, to August 9, 2024.

		Coefficient	*P* value	95% CI
Video length		0.01	.08	0.00 to 0.02
Video length square		0.00	.08	0.00 to 0.00
Event timing				
	During disaster	—[a]	—	—
	Immediate aftermath	−0.82	.04	−1.60 to −0.04
	Long-term recovery	−1.51	<.001	−2.39 to −0.64
Music and effects		−0.43	.21	−1.11 to 0.25
Content theme				
	Community, Solidarity, & Tribute	—	—	—
	Environmental & Climate Commentary	0.44	.71	−1.89 to 2.77
	Government & Policy Response	1.62	.05	0.00 to 3.25
	Impact & Damage	2.46	<.001	1.19 to 3.73
	Informational	1.86	.01	0.54 to 3.18
	Misinformation & Fake News	4.55	<.001	2.44 to 6.65
	Personal Narratives & Interviews	2.11	<.001	0.72 to 3.51
	Relief Efforts	2.38	<.001	0.94 to 3.82
	Tourism Impact	2.96	<.001	1.37 to 4.54
Hashtag group				
	AdvocacyAndUpdates	—	—	—
	MauiLocations	0.19	.70	−0.75 to 1.12
	Relief	−1.68	.01	−2.89 to −0.46
	Support	2.28	<.001	1.30 to 3.25
	WildfireContent	1.64	<.001	0.68 to 2.60
Constant		2.70	<.001	1.17 to 4.23

[a]Not applicable (reference category).

Discussion

Principal Findings

This study provides novel insights into how TikTok serves as a platform for public engagement during and after a natural disaster, using the 2023 Maui wildfire as a case study. Guided by SMCC theory, we examined how content themes, hashtags, and video features influence user engagement metrics, such as likes and shares. Our findings contribute to the growing body of literature on social media's role in crisis communication while offering unique contributions that expand upon prior studies [43-49].

This study reveals that TikTok serves as a dynamic platform for real-time information dissemination and public engagement during and after natural disasters. Posts emphasizing themes such as "Impact & Damage" and "Personal Narratives & Interviews" garnered the highest engagement metrics, illustrating the platform's capacity to emotionally and visually document the crisis. Content highlighting tourism and economic impacts also performed strongly, reflecting the broader public's interest in understanding the disaster's ripple effects. However, posts addressing critical but less emotional themes, such as "Government & Policy Response" and "Relief Efforts," attracted relatively lower engagement, highlighting a potential gap in how these topics resonate with TikTok audiences.

Engagement trends varied across the disaster timeline. While a prior study on the Maui wildfire focused on the immediate aftermath of the disaster, we analyzed engagement over a 1-year period, encompassing both crisis and postcrisis phases [38]. This extended time frame enabled us to explore the evolution of public engagement and highlighted key shifts in user priorities over time. Posts shared during the crisis phase generated the highest engagement, underscoring the public's urgent demand

XSL•FO
RenderX

for real-time updates. Engagement declined in the immediate aftermath and long-term recovery periods, reflecting waning public attention. These patterns emphasize the importance of strategically timing communication to sustain public interest and engagement throughout the recovery phase. The observed temporal trends also underscore the need for innovative strategies to re-engage audiences as the crisis progresses into recovery [35].

Engagement was predominantly driven by themes that conveyed the emotional impact and visual magnitude of the disaster, reinforcing the role of emotionally resonant content in shaping public discourse [37,38]. This pattern aligns with SMCC theory, which emphasizes the importance of adjusting information in crisis communication—content that helps audiences process the event and fosters emotional connection [35]. Moreover, informational posts about safety alerts and evacuation notices garnered relatively low engagement despite their critical importance for public safety. In contrast, adjusting and bolstering information, such as personal narratives and fundraising efforts, received significantly higher engagement, reflecting the public's preference for emotionally resonant and actionable content. These findings suggest that while instructing information is vital during the early crisis phase, sustained engagement may rely on content that fosters emotional connection and community solidarity [35,36].

Our results underscore the importance of visually compelling and emotionally resonant content in driving engagement during disasters [50-52]. This aligns with a recent study that found audiovisual vividness was a key predictor of engagement on TikTok during the same wildfire [38]. Both studies highlight the role of survivor-led narratives in capturing public attention, suggesting that user-generated content may resonate more strongly than traditional news media posts. However, while prior work used a digital storytelling framework to analyze the role of narrative persuasion and visual storytelling, our study extends the literature by applying SMCC theory to examine how instructing, adjusting, and bolstering messages align with engagement patterns [35-38]. This distinction situates our findings within the broader literature on crisis communication strategies.

Furthermore, our study highlights the persistent challenge of misinformation during disasters [21,27-29,53]. Posts categorized under misinformation themes, while limited in number, garnered substantial engagement, a finding that aligns with a recent study that found that posts leveraging emotional storytelling and high audiovisual vividness were particularly effective at drawing audience engagement [38]. This finding underscores the broader concern in public health regarding the virality of misinformation on social media, where false claims about disaster response, relief efforts, or health risks can undermine public trust in official sources [27-31]. Proactive strategies are needed to combat misinformation and amplify credible sources during crisis events. Integrating insights from both SMCC theory and the digital storytelling literature, our findings suggest that pairing information with compelling visual narratives may enhance its reach and impact, particularly on entertainment-driven platforms like TikTok.

Finally, our study contributes to the ongoing discourse on the role of video features in driving engagement [25,26,50,51]. The platform's algorithm-driven reach and short-form video format provide unique opportunities to disseminate impactful messages that resonate with diverse audiences [34,38,54]. Consistent with prior research, we find that shorter videos with clear and emotionally engaging messages are more likely to capture public attention [55]. These findings align with broader trends in TikTok engagement metrics and highlight the need for tailored communication strategies that align with platform-specific user behaviors. Tailored strategies to amplify these themes could enhance the platform's utility in disaster response [20].

Limitations

While 2 researchers independently coded the posts to ensure reliability and minimize bias, the interpretation of themes still involves subjective judgment. Even with our high interrater reliability, some residual bias may persist. The 1-year timeframe of the study provides a snapshot but might miss evolving trends over a longer duration. Additionally, the focus on TikTok, with its distinct user demographics and content styles, limits the generalizability of the findings to other social media platforms such as X (formerly Twitter), Facebook, or Instagram. The unique characteristics of TikTok's user base might not reflect broader social media trends. Furthermore, language and cultural barriers could have resulted in the underrepresentation of posts in languages other than English or those deeply rooted in specific cultural contexts, limiting the diversity of perspectives analyzed.

We acknowledge that TikTok's algorithmic influence on post visibility is a significant factor, as the platform curates content based on user behavior, preferences, and engagement, potentially creating biases in the posts available for analysis. As a result, our data set may disproportionately reflect content with high algorithmic visibility rather than a fully representative sample of all relevant posts. Additionally, regional and language biases in content discovery may affect the results, as TikTok's algorithms and search functionalities prioritize content tailored to specific geographic regions and linguistic preferences. This could result in the underrepresentation of videos from certain regions or in languages less prominent on the platform. The impact of paid promotion or boosted content is another limitation. Videos that received paid promotions may have artificially higher visibility and engagement, skewing the data toward commercially sponsored content. Our study did not distinguish between organically popular posts and those boosted through paid promotion, which could influence the analysis of engagement trends.

Implications for Future Research and Practice

The findings have important implications for disaster communication and public health response strategies. Understanding how engagement evolves over time and informs messaging that sustains public interest beyond the acute crisis phase. Disaster communicators can leverage TikTok's ability to amplify emotionally resonant content while addressing misinformation through targeted campaigns. Future research should explore cross-platform comparisons to identify complementary roles of different social media platforms.

XSL•FO
RenderX

Additionally, research on the effectiveness of public health interventions in countering misinformation and improving digital health literacy is needed to develop best practices for social media–based disaster communication.

Conclusion

In conclusion, this study builds upon prior work by integrating theoretical insights from SMCC theory with empirical analyses of TikTok engagement during a major natural disaster. TikTok is an influential tool for real-time communication, emotional support, and advocacy during natural disasters. By examining themes related to the Maui wildfire, this study demonstrated how users respond to and interact with crisis-related content in real time. While TikTok's strengths lie in its ability to visually and emotionally connect with audiences, addressing challenges such as misinformation and sustaining engagement throughout the recovery phase is essential. Strengthening digital crisis communication strategies, particularly in sustaining engagement beyond the immediate response phase, is critical for improving disaster preparedness and recovery efforts. Future research should further investigate best practices for leveraging social media in disaster preparedness, response, and community resilience.

Acknowledgments

JPS, JKK, and ANO were supported by the National Institute on Minority Health and Health Disparities (NIMHD) at the National Institutes of Health (NIH) under award number R61MD019939. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIMHD. The funders had no role in the study design, data analysis, decision to publish, or preparation of the manuscript.

Data Availability

The data underlying the results presented in the study are publicly available from TikTok.

Authors' Contributions

All authors contributed to the study's conception and design. Data collection and analysis were performed by AS and KT. JPS wrote the first draft of the manuscript. All authors contributed to subsequent drafts of the manuscript. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Distribution of likes and shares.
[DOCX File, 34 KB - formative_v9i1e67515_app1.docx]

Multimedia Appendix 2
Multivariable robust regression of log-transformed likes and shares.
[DOCX File, 24 KB - formative_v9i1e67515_app2.docx]

References

1. Farrant DN, Trauernicht C, Kagawa-Viviani A, Giambelluca TW, D'Antonio CM. On the anniversary of the Maui fires, a call for indigenous land care to mitigate future disasters. Proc Natl Acad Sci USA 2024 Sep 24;121(39):e2401045121. [doi: 10.1073/pnas.2401045121]
2. Welch L, Ramones M. 102 angels looking down on us: Maui marks 1 year since wildfire disaster. Hawaii News Now. URL: https://www.hawaiinewsnow.com/2024/07/09/1-year-anniversary-maui-wildfires-approaches-events-take-place-honor-victims [accessed 2024-07-09]
3. Suresh A, Tu L, Stanford FC. The 2023 Maui wildfires and the mental health effects of climate-induced relocation. Am J Public Health 2024 Nov;114(11):1184-1187. [doi: 10.2105/AJPH.2024.307815] [Medline: 39236278]
4. Holliday R, Krishnamurti LS, Jordan SE, Sia MA, Brenner LA, Monteith LL. The health and social impacts of the Maui wildfires: post-disaster care from a sociocultural lens. Hawaii J Health Soc Welf 2024 Mar;83(3):85-87. [doi: 10.62547/KXYZ1785] [Medline: 38456158]
5. Balch JK, Bradley BA, Abatzoglou JT, Nagy RC, Fusco EJ, Mahood AL. Human-started wildfires expand the fire niche across the United States. Proc Natl Acad Sci U S A 2017 Mar 14;114(11):2946-2951. [doi: 10.1073/pnas.1617394114] [Medline: 28242690]
6. Wen J, Heft-Neal S, Baylis P, Boomhower J, Burke M. Quantifying fire-specific smoke exposure and health impacts. Proc Natl Acad Sci U S A 2023 Dec 19;120(51):e2309325120. [doi: 10.1073/pnas.2309325120] [Medline: 38085772]

XSL•FO
RenderX

7. Gould CF, Heft-Neal S, Johnson M, Aguilera J, Burke M, Nadeau K. Health effects of wildfire smoke exposure. Annu Rev Med 2024 Jan 29;75:277-292. [doi: 10.1146/annurev-med-052422-020909] [Medline: 37738508]
8. Bowman D, Kolden CA, Abatzoglou JT, et al. Wildfires: climate change, deforestation, and the aridification of wildlands. Science 2020;370(6519):eabb5365. [doi: 10.1126/science.abb5365]
9. Xu R, Yu P, Abramson MJ, et al. Wildfires, global climate change, and human health. N Engl J Med 2020 Nov 26;383(22):2173-2181. [doi: 10.1056/NEJMsr2028985] [Medline: 33034960]
10. Lambrou N, Kolden C, Loukaitou-Sideris A, Anjum E, Acey C. Social drivers of vulnerability to wildfire disasters: a review of the literature. Landsc Urban Plan 2023 Sep;237:104797. [doi: 10.1016/j.landurbplan.2023.104797]
11. Hertelendy AJ, Howard C, Sorensen C, et al. Seasons of smoke and fire: preparing health systems for improved performance before, during, and after wildfires. Lancet Planet Health 2024 Aug;8(8):e588-e602. [doi: 10.1016/S2542-5196(24)00144-X] [Medline: 39122327]
12. Alexander DE. Social media in disaster risk reduction and crisis management. Sci Eng Ethics 2014 Sep;20(3):717-733. [doi: 10.1007/s11948-013-9502-z] [Medline: 24306994]
13. Houston JB, Hawthorne J, Perreault MF, et al. Social media and disasters: a functional framework for social media use in disaster planning, response, and research. Disasters 2015 Jan;39(1):1-22. [doi: 10.1111/disa.12092] [Medline: 25243593]
14. Lachlan KA, Spence PR, Lin X, Najarian K, Del Greco M. Social media and crisis management: CERC, search strategies, and Twitter content. Comput Human Behav 2016 Jan;54:647-652. [doi: 10.1016/j.chb.2015.05.027]
15. Saroj A, Pal S. Use of social media in crisis management: a survey. Int J Disaster Risk Reduct 2020 Sep;48:101584. [doi: 10.1016/j.ijdrr.2020.101584]
16. Clark N, Boersma K, Bonati S, et al. Exploring the impacts of social media and crowdsourcing on disaster resilience. Open Res Eur 2021;1:60. [doi: 10.12688/openreseurope.13721.3] [Medline: 38380383]
17. Ogie RI, James S, Moore A, Dilworth T, Amirghasemi M, Whittaker J. Social media use in disaster recovery: a systematic literature review. Int J Disaster Risk Reduct 2022 Feb;70:102783. [doi: 10.1016/j.ijdrr.2022.102783]
18. Lam NSN, Meyer M, Reams M, et al. Improving social media use for disaster resilience: challenges and strategies. Int J Disaster Risk Reduct 2023 Oct 2;16(1):3023-3044. [doi: 10.1080/17538947.2023.2239768]
19. Seneviratne K, Nadeeshani M, Senaratne S, Perera S. Use of social media in disaster management: challenges and strategies. Sustainability 2024 Jan;16(11):4824. [doi: 10.3390/su16114824]
20. Sharma R, De Choudhury M. TikTok and the evolving public health landscape. Nat Med 2021;27(1):8-10. [doi: 10.1038/s41591-020-01170-1]
21. Kirkpatrick CE, Lawrie LL. TikTok as a source of health information and misinformation for young women in the United States: survey study. JMIR Infodemiology 2024 May 21;4:e54663. [doi: 10.2196/54663] [Medline: 38772020]
22. Comp G, Dyer S, Gottlieb M. Is TikTok the next social media frontier for medicine? AEM Educ Train 2021 Jul;5(3). [doi: 10.1002/aet2.10532] [Medline: 34095694]
23. Kong W, Song S, Zhao YC, Zhu Q, Sha L. TikTok as a health information source: assessment of the quality of information in diabetes-related videos. J Med Internet Res 2021 Sep 1;23(9):e30409. [doi: 10.2196/30409] [Medline: 34468327]
24. Samuel L, Kuijpers K, Bleakley A. TherapyTok for depression and anxiety: a quantitative content analysis of high engagement TikTok videos. J Adolesc Health 2024 Jun;74(6):1184-1190. [doi: 10.1016/j.jadohealth.2024.02.002] [Medline: 38493396]
25. Lau N, Srinakarin K, Aalfs H, Zhao X, Palermo TM. TikTok and teen mental health: an analysis of user-generated content and engagement. J Pediatr Psychol 2025 Jan 1;50(1):63-75. [doi: 10.1093/jpepsy/jsae039] [Medline: 38950415]
26. Xiong M, Young AT, Fernandez A, Jones JS, Kirby AM. Understanding breast cancer surgery through Tiktok: an analysis of video content and viewer engagement. Ann Surg Oncol 2024 Jul;31(7):4182-4184. [doi: 10.1245/s10434-024-15268-y] [Medline: 38592623]
27. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 9;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
28. Lewandowsky S, Ecker UKH, Cook J. Beyond misinformation: understanding and coping with the “post-truth” era. J Appl Res Mem Cogn 2017;6(4):353-369. [doi: 10.1016/j.jarmac.2017.07.008]
29. Starbird K, Maddock J, Orand M, Achterman P, Mason RM. Rumors, false flags, and digital vigilantes: misinformation on twitter after the 2013 boston marathon bombing. Presented at: iConference 2014; Mar 4-7, 2014; Berlin. [doi: 10.9776/14308]
30. George JO, Elayan S, Sykora M, et al. The role of social media in building pandemic resilience in an urban community: a qualitative case study. Int J Environ Res Public Health 2023 Sep 4;20(17):6707. [doi: 10.3390/ijerph20176707] [Medline: 37681847]
31. Xie L, Pinto J, Zhong B. Building community resilience on social media to help recover from the COVID-19 pandemic. Comput Human Behav 2022 Sep;134:107294. [doi: 10.1016/j.chb.2022.107294] [Medline: 35431426]
32. Li Y, Guan M, Hammond P, Berrey LE. Communicating COVID-19 information on TikTok: a content analysis of TikTok videos from official accounts featured in the COVID-19 information hub. Health Educ Res 2021 Jul 12;36(3):261-271. [doi: 10.1093/her/cyab010] [Medline: 33667311]
33. Unni Z, Weinstein E. Shelter in place, connect online: trending TikTok content during the early days of the U.S. COVID-19 pandemic. J Adolesc Health 2021 May;68(5):863-868. [doi: 10.1016/j.jadohealth.2021.02.012] [Medline: 33824071]

XSL•FO
RenderX

34. Basch CH, Yalamanchili B, Fera J. #Climate change on TikTok: a content analysis of videos. J Community Health 2022 Feb;47(1):163-167. [doi: 10.1007/s10900-021-01031-x] [Medline: 34545460]
35. Jin Y, Austin L, Liu BF. Social‐Mediated crisis communication research: how information generation, consumption, and transmission influence communication processes and outcomes. In: Coombs WT, Holladay SJ, editors. The Handbook of Crisis Communication, 1st edition: Wiley; 2022:151-167. [doi: 10.1002/9781119678953]
36. Lauran N, Kunneman F, Van de Wijngaert L. Connecting social media data and crisis communication theory: a case study on the chicken and the egg. J Risk Res 2020 Oct 2;23(10):1259-1277. [doi: 10.1080/13669877.2019.1628097]
37. Tian Y, Yang J. Deny or bolster? A comparative study of crisis communication strategies between Trump and Cuomo in COVID-19. Public Relat Rev 2022 Jun;48(2):102182. [doi: 10.1016/j.pubrev.2022.102182] [Medline: 35368968]
38. Liu W, Zhao X, Zhan M(, Hernandez S. Streaming disasters on TikTok: examining social mediated crisis communication, public engagement, and emotional responses during the 2023 Maui wildfire. Public Relat Rev 2024 Dec;50(5):102512. [doi: 10.1016/j.pubrev.2024.102512]
39. Saunders B, Sim J, Kingstone T, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907. [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
40. Gisev N, Bell JS, Chen TF. Interrater agreement and interrater reliability: key concepts, approaches, and applications. Res Social Adm Pharm 2013;9(3):330-338. [doi: 10.1016/j.sapharm.2012.04.004] [Medline: 22695215]
41. McCay-Peet L, Quan-Haase A. A model of social media engagement: user profiles, gratifications, and experiences. In: The SAGE Handbook of Social Media Research Methods: SAGE; 2017:199-214.
42. Neiger BL, Thackeray R, Van Wagenen SA, et al. Use of social media in health promotion: purposes, key performance indicators, and evaluation metrics. Health Promot Pract 2012 Mar;13(2):159-164. [doi: 10.1177/1524839911433467] [Medline: 22382491]
43. Muniz-Rodriguez K, Ofori SK, Bayliss LC, et al. Social media use in emergency response to natural disasters: a systematic review with a public health perspective. Disaster Med Public Health Prep 2020 Feb;14(1):139-149. [doi: 10.1017/dmp.2020.3] [Medline: 32148219]
44. Mitcham D, Taylor M, Harris C. Utilizing social media for information dispersal during local disasters: the communication hub framework for local emergency management. Int J Environ Res Public Health 2021 Oct 14;18(20):10784. [doi: 10.3390/ijerph182010784] [Medline: 34682529]
45. Simon T, Adini B, El-Hadid M, Goldberg A, Aharonson-Daniel L. The race to save lives: demonstrating the use of social media for search and rescue operations. PLoS Curr 2014 Nov 6;6:ecurrents.dis.806848c38f18c6b7b0037fae3cd4edc5. [doi: 10.1371/currents.dis.806848c38f18c6b7b0037fae3cd4edc5] [Medline: 25685618]
46. Gómez-Casillas A, Gómez Márquez V. The effect of social network sites usage in climate change awareness in Latin America. Popul Environ 2023 May 1;45(2):7. [doi: 10.1007/s11111-023-00417-4] [Medline: 37152891]
47. Meng Y, Chung D, Zhang A. The effect of social media environmental information exposure on the intention to participate in pro-environmental behavior. PLoS ONE 2023;18(11):e0294577. [doi: 10.1371/journal.pone.0294577] [Medline: 37972040]
48. Maran DA, Begotti T. Media exposure to climate change, anxiety, and efficacy beliefs in a sample of Italian university students. Int J Environ Res Public Health 2021 Sep 4;18(17):9358. [doi: 10.3390/ijerph18179358] [Medline: 34501946]
49. Shan S, Zhao F. Social media-based urban disaster recovery and resilience analysis of the Henan deluge. Nat Hazards (Dordr) 2023 May 23:1-29. [doi: 10.1007/s11069-023-06010-0] [Medline: 37360801]
50. Chen Q, Min C, Zhang W, Ma X, Evans R. Factors driving citizen engagement with government TikTok accounts during the COVID-19 pandemic: model development and analysis. J Med Internet Res 2021 Feb 4;23(2):e21463. [doi: 10.2196/21463] [Medline: 33481756]
51. Cheng Z, Li Y. Like, comment, and share on TikTok: exploring the effect of sentiment and second-person view on the user engagement with TikTok news videos. Soc Sci Comput Rev 2024 Feb;42(1):201-223. [doi: 10.1177/08944393231178603]
52. Basch CH, Hillyer GC, Yalamanchili B, Morris A. How TikTok is being used to help individuals cope with breast cancer: cross-sectional content analysis. JMIR Cancer 2022 Dec 6;8(4):e42245. [doi: 10.2196/42245] [Medline: 36472899]
53. Basch CH, Basch CE, Hillyer GC, Meleo-Erwin ZC. Social media, public health, and community mitigation of COVID-19: challenges, risks, and benefits. J Med Internet Res 2022 Apr 12;24(4):e36804. [doi: 10.2196/36804] [Medline: 35380539]
54. Chavda J. Who U.S. adults follow on TikTok. Pew Research Center. URL: https://www.pewresearch.org/data-labs/2024/10/08/who-u-s-adults-follow-on-tiktok/ [accessed 2025-02-19]
55. Dai X, Wang J. Effect of online video infotainment on audience attention. Humanit Soc Sci Commun 2023;10(1):421. [doi: 10.1057/s41599-023-01921-6]

Abbreviations

SMCC: Social-Mediated Crisis Communication
STROBE: Strengthening the Reporting of Observational studies in Epidemiology

XSL•FO
RenderX

Edited by A Mavragani; submitted 14.10.24; peer-reviewed by KP Hocevar, R Peng; revised version received 29.01.25; accepted 30.01.25; published 04.03.25.

Please cite as:
Stimpson JP, Srivastava A, Tamirisa K, Kaholokula JK, Ortega AN
Crisis Communication About the Maui Wildfires on TikTok: Content Analysis of Engagement With Maui Wildfire–Related Posts Over 1 Year
JMIR Form Res 2025;9:e67515
URL: https://formative.jmir.org/2025/1/e67515
doi:10.2196/67515

© Jim P Stimpson, Aditi Srivastava, Ketan Tamirisa, Joseph Keawe aimoku Kaholokula, Alexander N Ortega. Originally published in JMIR Formative Research (https://formative.jmir.org), 4.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Fast Healthcare Interoperability Resources (FHIR)–Based Interoperability Design in Indonesia: Content Analysis of Developer Hub's Social Networking Service

Lukman Heryawan[1], PhD; Yukiko Mori[2], MD, PhD; Goshiro Yamamoto[3], PhD; Naoto Kume[2], PhD; Lutfan Lazuardi[4], PhD; Anis Fuad[4], DEA; Tomohiro Kuroda[2], Prof Dr

[1]Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Sekip utara BLS 21, Sleman, Indonesia

[2]Division of Medical Information Technology and Administration Planning, Kyoto University Hospital, Kyoto University, Kyoto, Japan

[3]Preemptive Medicine and Lifestyle Related Disease Research Center, Kyoto University Hospital, Kyoto, Japan

[4]Health Policy and Management, Faculty of Medicine Public Health and Nursing, Universitas Gadjah Mada, Sleman, Indonesia

Corresponding Author:
Lukman Heryawan, PhD
Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Sekip utara BLS 21, Sleman, Indonesia

Abstract

Background: Interoperability in health care is a critical aspect for the exchange of health information. The Fast Healthcare Interoperability Resources (FHIR) framework has become widely adopted to provide interoperable data exchange in the health care industry. The COVID-19 pandemic has demonstrated the significance of interoperable data in tracking patients who have contracted the virus and keeping track of the vaccinated population. Indonesia is one of the many countries that have implemented interoperable data systems to track patients with COVID-19, and it has aspirations to expand the system to other use cases, particularly in the primary health care setting. The primary health care providers in Indonesia include *Puskesmas* (community health centers) and private clinics.

Objective: To promote interoperable health data exchange in the primary health care sector, the Indonesian government has launched the Satusehat project. The goal of the Satusehat platform is to make health data in Indonesia interoperable and exchangeable between health care organizations, particularly *Puskesmas* and private clinics.

Methods: For a successful implementation of the Satusehat platform in *Puskesmas* and private clinics, it is crucial to understand the challenges that may arise. This study analyzed the pain points of the Satusehat platform based on a content analysis of the Satusehat Social Networking Service Telegram group messages. The study revealed the pain points and suggested existing approaches to address them, which can be used as a proposed design of interoperability for *Puskesmas* and private clinics, making it easier for these organizations to adopt the Satusehat platform.

Results: The pain points identified in this study include issues with the FHIR server, problems with FHIR profile selection, and the mapping of electronic medical record data into standardized data, such as mapping into the Systematized Nomenclature of Medicine Clinical Terminology. The results show that the value of the mapping issue is 37, profile issue 9, and server issue 61. Among the 3 categories, server issues had the highest population, followed by mapping issues and then profile issues. To address these issues, the study proposed practical approaches, including a federated architecture for the FHIR server instead of a centralized architecture, an FHIR writer and FHIR viewer system inspired by the Standardized Structured Medical Record Information eXchange system in Japan, and an FHIR conversion framework that integrates with our FHIR writer and FHIR viewer system.

Conclusions: These proposed solutions can help resolve the pain points identified in the study and help the advancement of the Satusehat platform implementation in *Puskesmas* and private clinics in Indonesia. We believed that the proposed solutions have potential to be adopted to other countries with similar issues when conducting nationwide project in health care interoperability design.

(JMIR Form Res 2025;9:e51270) doi:10.2196/51270

KEYWORDS

interoperability; Fast Healthcare Interoperability Resources (FHIR); COVID-19; Satusehat; Indonesia; interoperability design; social networking; health information; content analysis; private clinic

XSL•FO
RenderX

Introduction

Background

Health information exchange (HIE) involves the transfer of electronic health data between different health information systems, including the electronic medical record (EMR) system, the electronic health record (EHR) system, and the personal health record (PHR) system. Notably, when the data are controlled by a health facility, an EMR system transitions into an EHR system. Conversely, if the data are managed by the patient or consumer themselves, it transforms into a PHR system [1]. To facilitate this exchange, HIE requires a data exchange or data interoperability framework that provides a standardized format for transferring health data [2]. One of the organizations that provides such interoperability frameworks is Health Level Seven (HL7) [3]. HL7 is a nonprofit organization that develops interoperability frameworks, such as HL7 FHIR (Fast Healthcare Interoperability Resources). It is a newer standard that provides a modern and flexible approach to exchanging health information. It uses a Representational State Transfer web services approach and supports both structured and unstructured data [4]. There have been several studies and reports on the implementation of EHR systems using the FHIR standard in different countries. One study, focused on the implementation of an FHIR-based EHR in Japan, highlighted some of the challenges faced during the implementation process, including lack of standardization and compatibility between different systems, difficulty of data migration, and the need for improved communication and cooperation between health care providers [5]. Another study provided the significance of standardization, data quality, and the need for technical support. The study also offered recommendations for improving the implementation and adoption of FHIR-based EHR systems [6,7].

The Ministry of Health (MoH) manages the EHR system in Indonesia, while the EMR system is independently managed by medical providers, such as hospitals, *Puskesmas* (community health centers), and private clinics. The PHR system, on the other hand, is provided by the MoH to the citizens of Indonesia [8]. In 2022, the expansion of EMR implementation to encompass all types of health care facilities was enacted through MoH decree number 24/2022. This regulation is a part of the digital health transformation blueprint released by the MoH [9].

A significant driver in the adoption of EMR, EHR, and PHR in Indonesia has been the COVID-19 pandemic. The PeduliLindungi app, developed as a tool for tracking, tracing, and monitoring during the pandemic, has become a crucial part of this initiative. This app has been used by more than 100 million users, making it interoperable with hundreds of laboratories and 17 telemedicine service providers. Most recently, through the MoH of the Republic of Indonesia decree number HK.01.07/MENKES/1559/2022, health data management is now under the MoH. Data from the PeduliLindungi app have been migrated to the Satusehat platform, as shown in Figure 1, which will integrate the PHR system and EMR system into a single nationwide EHR system [10]. The Satusehat Mobile platform enables the public to track COVID-19 vaccinations, immunization records, and medication information from EMRs of various health care facilities connected to this platform, using the FHIR HL7 standard. Furthermore, the EMR systems will be able to exchange and share data with ease through the Satusehat platform.

Legal basis for the implementation of the Satusehat system in Indonesia has been further strengthened with the enactment of the Health Law number 17/2023. This legislation specifically references the National Health Information System, which is interpreted as a system based on the Satusehat platform. The forthcoming derivative laws are expected to lay out the structural and operational details, ensuring a coherent and efficient implementation of these systems across the country.

The Satusehat platform represents the cornerstone of Indonesia's nationwide EHR system [11]. This platform aims to facilitate the exchange and sharing of health data, ultimately leading to better patient care and outcomes. In Indonesia, the Satusehat platform places a strong emphasis on primary health care. This is caused by the referral health care system in place, where most patients' health data begin with their first visit to a primary health care facility such as a *Puskesmas* or private clinic. These facilities are typically the first point of contact for patients seeking medical attention. If the primary health care facilities are unable to provide the necessary treatment or care, the patient will be referred to a higher level of care, such as a hospital. This referral system is depicted in Figure 2, which shows the progression of patients through different levels of care within the Indonesian health care system [12].

XSL•FO
RenderX

Figure 1. Satusehat platform's architecture. PeduliLindungi app as a PHR is integrated to FHIR server with view only. On the other hand, hospital information system with data from EMR could view or write with other health care applications. The FHIR server which is handled by the Ministry of Health acts as a bridge between health care digital instances. EMR: electronic health record; FHIR: Fast Healthcare Interoperability Resources; MoH: Ministry of Health; PHR: personal health record.

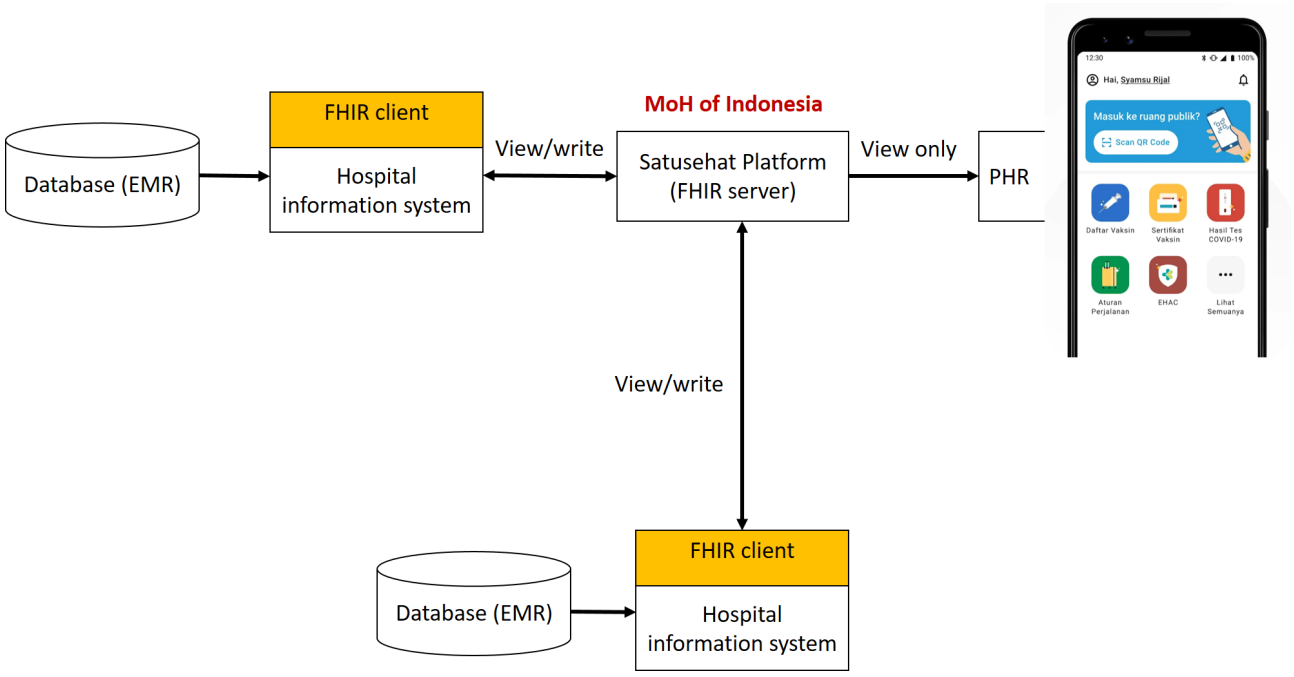
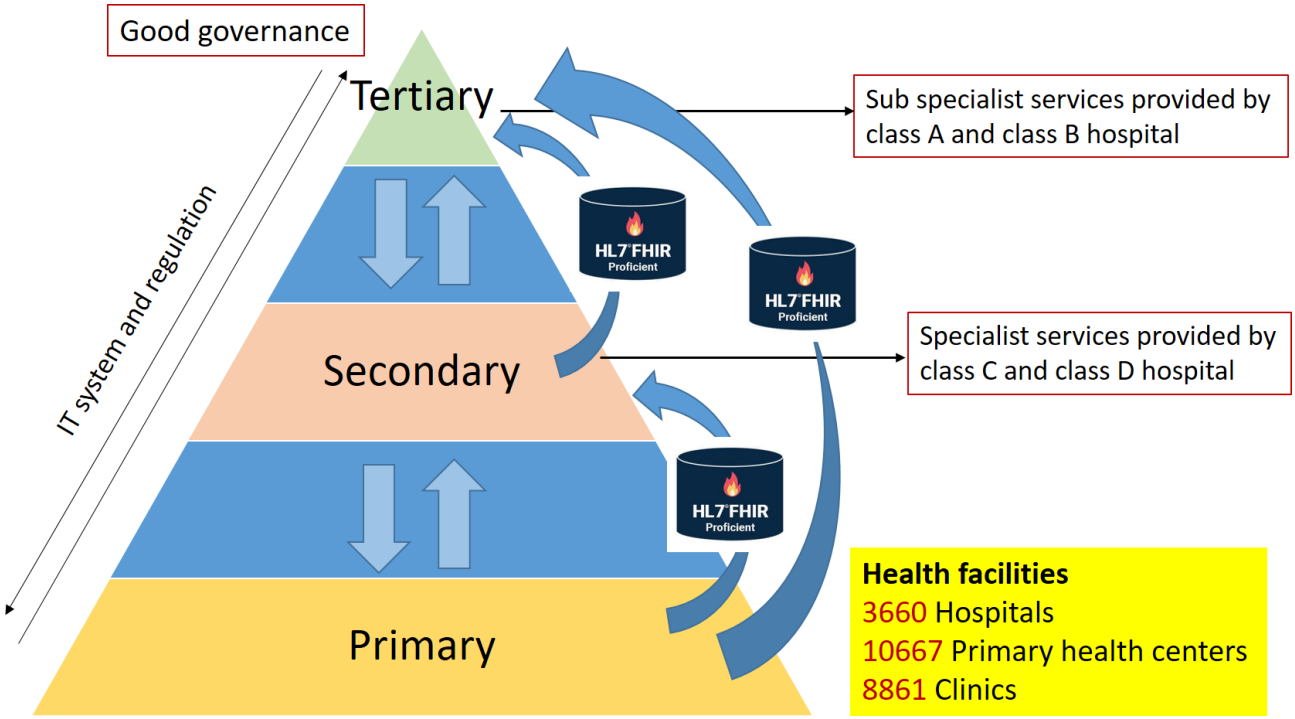


Figure 2. Indonesia's health care tiered referral system. Primary care facilities such as *Puskesmas* or private clinics as the base and entry point of health care service. Patient progress or referred to higher level, including secondary of tertiary facilities, if needed. This system is implemented with HL7 FHIR between tiers, thorough IT system, and regulation. HL7 FHIR: Health Level Seven Fast Healthcare Interoperability Resources.

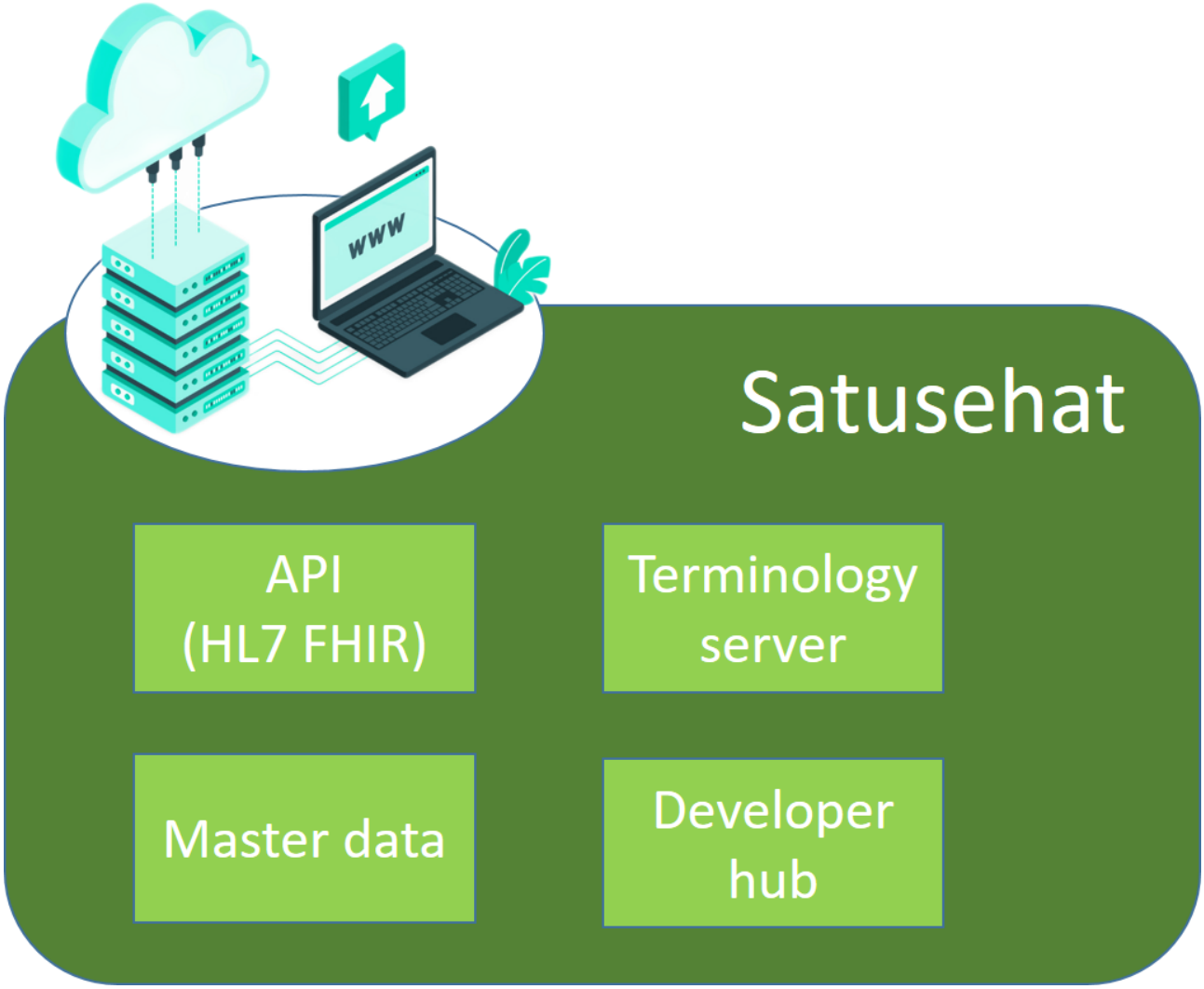


The Satusehat platform consists of 4 primary components: the FHIR server, the terminology server, the master data server, and the developer hub as shown in Figure 3. The FHIR server is responsible for providing an HL7 FHIR application programming interface (API) for health data exchange. The terminology server serves as a database, providing standardized data references for diagnoses using ICD-10 (*International Statistical Classification of Diseases, Tenth Revision*), observations using Logical Observation Identifiers Names and Codes (LOINC), and other clinical terms using Systematized Nomenclature of Medicine Clinical Terminology (SNOMED-CT). The master data server is a database that contains the nationwide identification numbers of patients or citizen ID, practitioner data, health facilities data, pharmacies data, financing data, and health service data. The developer hub provides a sandbox environment for developers to test the FHIR

XSL•FO
RenderX

API offered by the MoH. Developers can experiment and test data exchange scenarios, such as immunization, using the provided FHIR API. The developer hub also includes community tools, such as a video-sharing platform, a website, and a Telegram group, to provide support and assistance to developers while they test the Satusehat FHIR API [13].

Figure 3. Satusehat platform components that consist of FHIR server, terminology server, master data server, and developer hub. FHIR is responsible for health data exchange, terminology server provides diagnoses terminology standardization, master data stores patient and health care instances data (practitioners and infrastructures), and developer hub is for support center for developers when using FHIR application programming interface. HL7 FHIR: Health Level Seven Fast Healthcare Interoperability Resources.



To ensure a successful implementation of the Satusehat platform in *Puskesmas* and private clinics, it is crucial to identify potential pain points in integrating EMR into the platform from a developer's perspective. Understanding the underlying architecture and structure of the FHIR standard is necessary for effective integration with EHR systems [14]. Another reason for pain points is the lack of resources, as implementing FHIR standards can be resource-intensive, requiring time and expertise. Developers may face constraints in terms of budget and man power, which can make it difficult to effectively implement the standard [15]. In addition, the lack of support from other stakeholders, including health care providers and technical experts, can also pose challenges for developers in implementing FHIR standards [16]. The Indonesian MoH is currently in the trial phase of the Satusehat platform. The Digital Transformation Office (DTO) of the MoH conducts workshops and training sessions for EMR system developers and vendors [17]. DTO also provides a website that provides step-by-step tutorials on how to connect and exchange data from EMR systems to the Satusehat platform using the FHIR framework, including the standardized terminologies used for mapping health data to standardized terminologies, such as LOINC for observations, *ICD-10* for diagnoses, and SNOMED-CT for other health terminologies. To provide interactive support, DTO also provides a Telegram group that developers can join to ask for support when integrating their EMR systems into the Satusehat platform. Based on the activities in the Telegram group, which mostly involve questions about how to integrate EMR systems into the Satusehat platform, we conclude that there are some pain points when integrating EMR systems into the Satusehat platform. This conclusion came from analysis through the Telegram group including a question and answer session with the DTO representative.

XSL•FO
RenderX

Objectives

The purpose of this study is to identify the challenges and difficulties encountered by developers that are present in situ during the integration process of EMR systems into the Satusehat platform. To accomplish this, we will analyze the communication content in the Telegram group of the Satusehat developer hub, which serves as an interactive community platform for developers. By exploring the pain points communicated by the developers in this forum, we aim to gain a deeper understanding of the barriers and obstacles encountered during the integration process. Once the pain points have been revealed, the study proposes practical approaches that can be used to ensure a successful integration into the Satusehat platform. These approaches are then combined to form an interoperability design of an FHIR-based system for EMR in *Puskesmas* and private clinics. The study is organized into 4 sections. The first section introduces the Satusehat platform, related work, and how pain points will be revealed. The second section describes the content analysis methods used in the study. The third section presents the results, and the fourth section discusses the practical approaches that will be proposed as an interoperability design to ensure a smoother integration into the Satusehat platform.

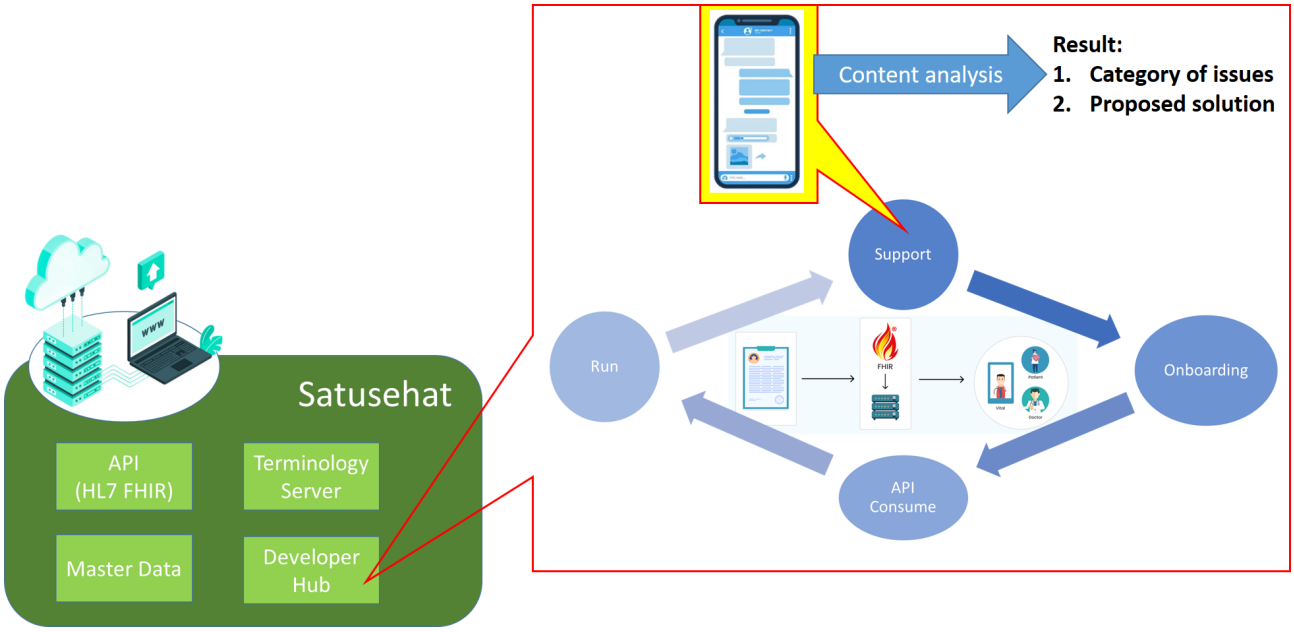
Methods

Study Design

To uncover the pain points that are addressed in the developer hub, we conducted a content analysis of the Telegram group's contents. We decided to conduct the analysis to the specific telegram group because the group is officially managed by the DTO. It is the most active group that contains developer discussions about EMR integration to the Satusehat platform. Content analysis is a research method used to make inferences by interpreting and coding written, spoken, or visual material [18]. The analysis was conducted at a Telegram group that consists of 1067 active members to this day. This study focused on analyzing conversations from December 20, 2022, the date the group was created, until January 29, 2023. The end date was chosen because the amount of extracted data is sufficient to support the experiment, which is content analysis. Extending the time span of chat extraction could potentially enhance the volume of data collected and strengthen the robustness of our conclusions, as the amount of data sampled in an experiment highly depends on the subject of the study [19]. In this instance, our focus was on Telegram as a social network service. A total of 219 chats were chosen during this period. However, we selected only 107 chats related to our research purpose. These chats primarily involved developers discussing issues encountered while integrating their EMR systems with the Satusehat platform using FHIR standards. In our research, we focused on issue-related chats, such as "http server error 500," "what FHIR profile should be used for chief complaints," and "SNOMED-CT or LOINC for mapping blood glucose." We excluded nonissue chats that contained simple responses such as "Ok" or "thank you." To collect the chats, we used the Telegram menu to export them into HTML files, which we later transferred to an Excel file. The collected chats were later manually labeled by 2 of the researchers (LH and ADW), who are trained to categorize the data according to agreed and specific rules for these issue-related chats. The labeler was chosen with taking the research design, the amount of data being analyzed, and the complexity of the coded classification process of the study into account. In addition, the coding process is ensured to be consistent and reliable with involving 2 researchers including the author and his colleagues (LH and ADW) with a background in information technology and understanding of the Satusehat platform. This established intercoder reliability in order to achieve valid interpretation of the data [20]. After conducting our content analysis and categorizing the issues, we developed a proposed approach for each issue category as shown in Figure 4. These proposed approaches are feasible incorporated into the existing Satusehat platform architecture, potentially resulting in an interoperability design that is less burdensome for developers. By less burdensome, we mean that the design can be easily implemented by novice developers, which are commonly found in private clinics and *Puskesmas* (community health centers). This design can be referred to as the FHIR-based Satusehat platform interoperability design for private clinics and *Puskesmas*. By reducing the burden of the interoperability design, the integration of EMR systems into the Satusehat platform in Indonesia will be smoother. It is important to note that there are other factors that can influence the smooth implementation beyond the technical scope this study. For the sake of the study's conciseness, we will focus only on the technical issues from a developer's perspective.

XSL•FO
RenderX

Figure 4. The approach method of the study. Given the developer hub, as the support and assistance forum for developers to test FHIR, the Telegram group chat content is extracted. Then, content analysis is conducted that resulted in categorization of issues and proposed solutions for each issues. API: application programming interface; HL7 FHIR: Health Level Seven Fast Healthcare Interoperability Resources.



Ethical Considerations

Public Telegram group of the Satusehat developer hub contents was used for content analysis in this research. There were no any extra compensation given to conduct the analysis since the content itself is public. We acknowledge the ethical concerns that follows while analyzing data that may raise privacy issues. An original study relating to framework that can be used by researchers to consider ethical issues in web-based research also included how privacy policy and content that were analyzed on specific web-based platform should be reviewed. The collected data have to be accessible to public, focused to the content, and not the individual who posted the content [21]. In addition, taking into account the public nature of the group and also Telegram's privacy policy relating to public chats, we believe that group members implicitly consented to having their data accessible publicly [22]. Following the logics provided in the framework and Telegram's privacy policy, the exemption of additional ethics review is feasible for this study. This study was reviewed by the Medical and Health Research Ethics Committee Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada-Dr. Sardjito General Hospital, Indonesia (KE/FK/0204/EC/2023).

The focus of this study is to analyze the content of the public developer hub and not on the individual who posted in the developer hub. Therefore, anonymization is applied to the data that are collected and presented in this study to ensure group participant's data privacy and confidentiality. According to other literatures, it is stated that waiver of consent could be requested when no identifying information is presented. Also, permission of further use of the data is permitted if the data are freely available on public forum [23]. Considering aforementioned points, acquiring consent was not necessary for this research. However, it is important to get consent in general for research using data from people. In our case, it was not possible to get informed consent due to the scope of our research.

Results

Overview

The results of our study are divided into 2 sections for clarity and organization. The first section outlines the pain points and issues with the Satusehat platform, which were identified through content analysis of the Satusehat Social Networking Service Telegram group messages. These pain points include issues with the FHIR server, difficulties in selecting the appropriate FHIR profile, and challenges in mapping EMR data into standardized data formats such as SNOMED-CT terminology [24]. Implementing and maintaining an FHIR server may be challenging due to the interoperability, security concerns, and scalability. Selecting the appropriate FHIR profile gives difficulties in aligning with specific use cases, balancing flexibility, and considering long-term maintenance. Mapping EMR data into standardized data formats such as SNOMED-CT could be challenging due to data complexity, granularity, updates, and achieving semantic interoperability. The second section of our study presents proposed designs that aim to address the pain points identified in the first section. We believe that the proposed designs will help overcome the challenges that may arise during the implementation of the Satusehat platform in primary health care settings, such as *Puskesmas* and private clinics. These proposed designs include a federated architecture for the FHIR server to ensure reliable and distributed server load management, an FHIR writer and viewer system inspired by the successful Standardized Structured Medical Record Information eXchange (SS-MIX) system in Japan [25], and an FHIR conversion framework that integrates with our FHIR writer and FHIR viewer system.

XSL•FO
RenderX

Pain Points

Based on content analysis of the Satusehat Social Networking Service Telegram group messages, we identified 3 main types of issues faced by developers: server issues, profile issues, and mapping issues as illustrated in Multimedia Appendix 1. Among the 3 categories, server issues had the highest population with 61 chats, followed by mapping issues with 37 identified chats and then profile issues with 9 chats. Our data collection method allowed us to focus on the issues and concerns that developers faced when integrating their EMR systems with Satusehat platform using FHIR standards.

Proposed Design

The Satusehat platform is built on a centralized architecture, which is illustrated in Figure 5. To ensure authentication, the platform uses an OAuth server [14]. Once a user is authenticated, they can access the FHIR resources available on the platform through the FHIR API. However, based on the developers' perspective, connecting to the Satusehat platform can be challenging. We have identified this issue as a server issue. Sometimes, the Satusehat platform server fails to respond to the user's request, resulting in errors such as HTTP Error 401—Unauthorized, HTTP status 403—Forbidden, HTTP status 404—Not Found, and HTTP status 500—Internal Server Error [26].

To overcome this issue, we propose a federated server architecture instead of a centralized server architecture [27], as shown in Figure 6. In this approach, the FHIR resources are stored in local or regional servers that are subordinate to the central server. Periodically, the central server inquiries about the FHIR resources and copies them to the central server. This approach ensures high reliability for both the central and local or regional servers, resulting in a more robust exchange of FHIR resources. By implementing the federated server architecture, the exchange of FHIR resources will be more reliable, making it easier for authenticated users to post, get, put, and patch the FHIR resources from the Satusehat platform.

The issue of mapping is a significant challenge faced by developers working on the Satusehat platform. Health care facilities in Indonesia use various local terminologies, which are often in Indonesian or local tribal languages, such as Javanese. In contrast, standardized terminologies such as *ICD-10* for diagnosis and *ICD-9-CM* (*International Classification of Diseases, Ninth Revision, Clinical Modification*) for medical procedures are in non-Indonesian languages, such as English [28]. This presents a significant challenge in mapping the various local terminologies with their many variations into a single standardized terminology in English. We believe that this issue is not only unique to Indonesia but also affects other countries whose national language is not English, such as Japan, Taiwan, and Thailand [29]. To overcome this challenge, we propose an approach that is similar to the one used in Japan. The Japanese approach uses the FHIR conversion framework [30]. In this approach, the conversion is done only on the terminology code and not on the terminology text. The written terminology still uses the local language, such as Japanese, but the code follows the standardized code. This approach reduces the complexity of the mapping task significantly. The FHIR conversion framework is illustrated in Figure 7. By implementing this approach, we can ensure that the Satusehat platform can handle the mapping of various local terminologies into standardized terminologies in a more efficient and effective way.

In addition, automating the code conversion process is much simpler than the text conversion process because the variations in code are relatively small. This approach not only reduces the workload for developers but also ensures the accuracy and consistency of the mapping process. The last issue concerns the challenge of selecting and applying the correct FHIR profiles to EMR data. While there is an implementation guide available to guide users in selecting the appropriate profiles for specific use cases, such as outpatient medical resumes and immunization records, manually selecting and applying the profiles can still be difficult. To address this issue, we propose a solution that involves feeding EMR data into a system that automatically selects the correct FHIR profile and builds the content of that profile based on the input data. Similar approaches have been implemented in countries that have standardized their health data using HL7 standards, such as Japan's use of HL7 V 2.5 to standardize various health data into a storage system called the SS-MIX [28]. The SS-MIX system has been successful in facilitating HIE in Japan. Since both SS-MIX and FHIR are from HL7, we believe that it would be feasible to adopt the SS-MIX system and modify it to use HL7 FHIR structures instead of HL7 V 2.5 structures. This is a practical solution as there are already open structures and guidelines available to implement SS-MIX. We simply need to modify the structure to use HL7 FHIR data.

XSL•FO
RenderX

Figure 5. Satusehat current architecture. The platform was built on centralized architecture and database [27]. Authenticated institutions or individuals are able to access the Fast Healthcare Interoperability Resources in the central database. Writer identifies this as the root of server issue stated in developer hub.

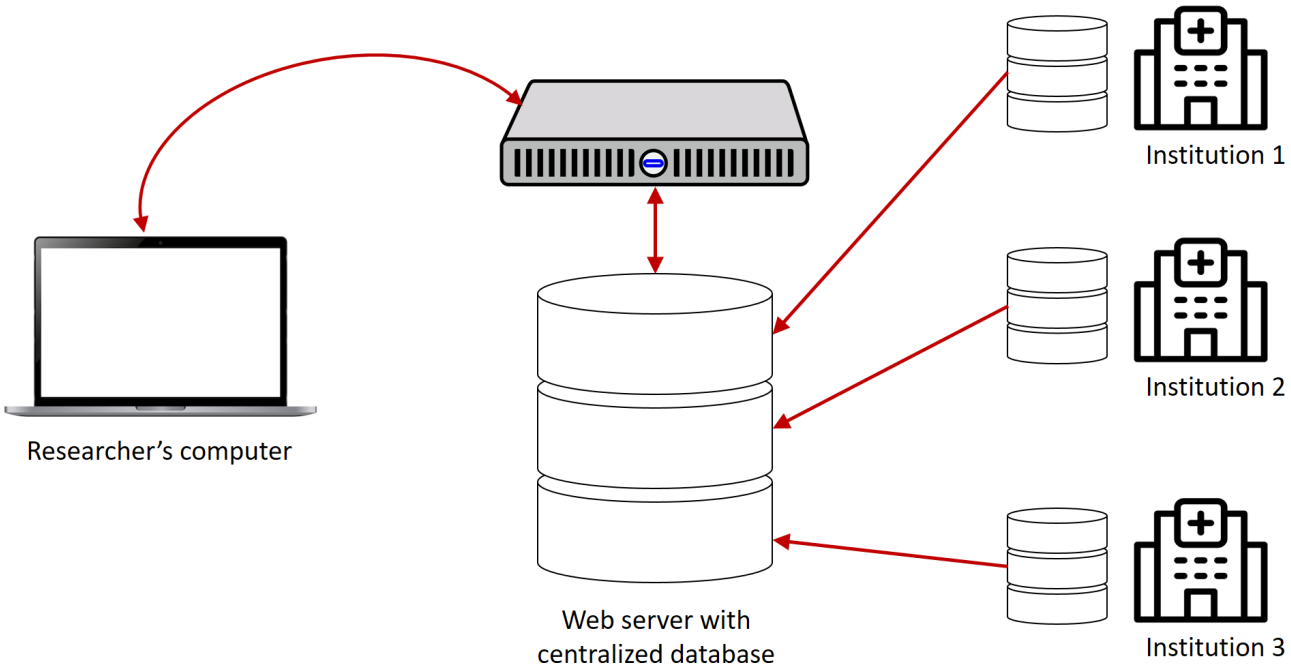
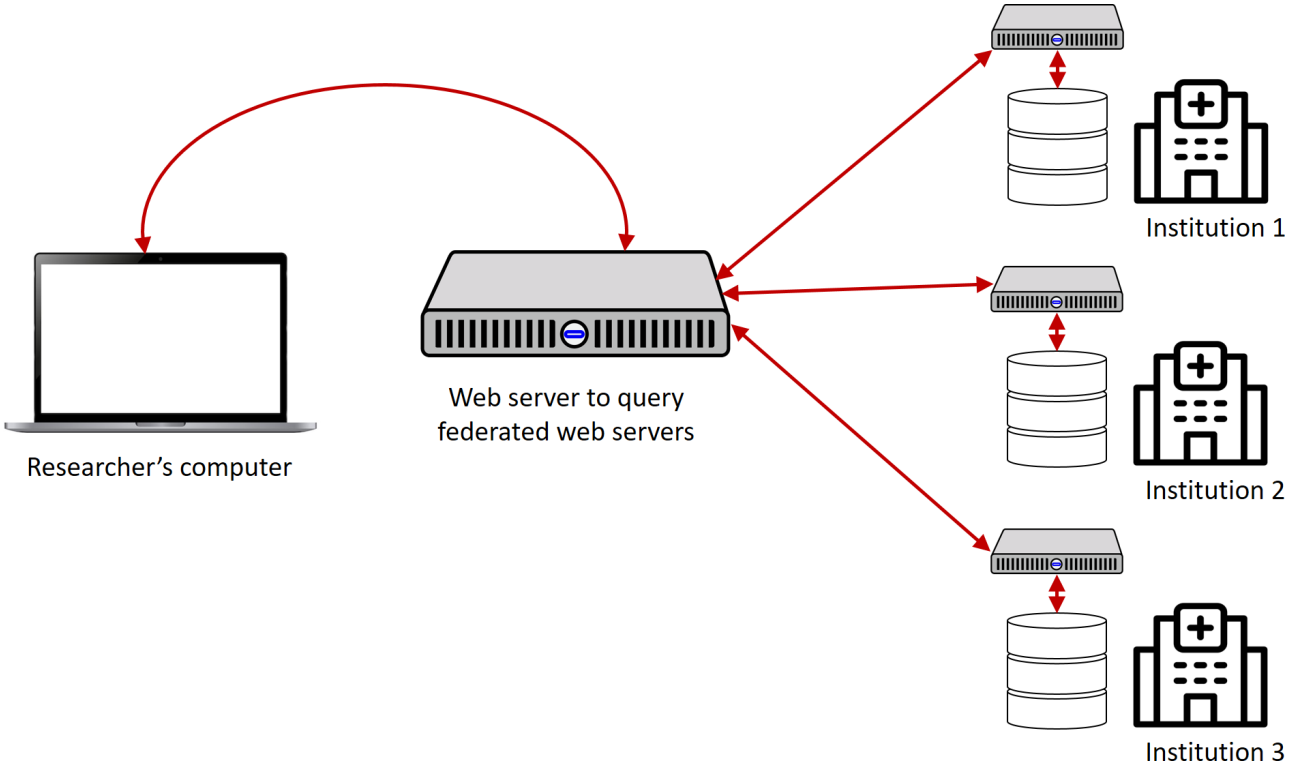
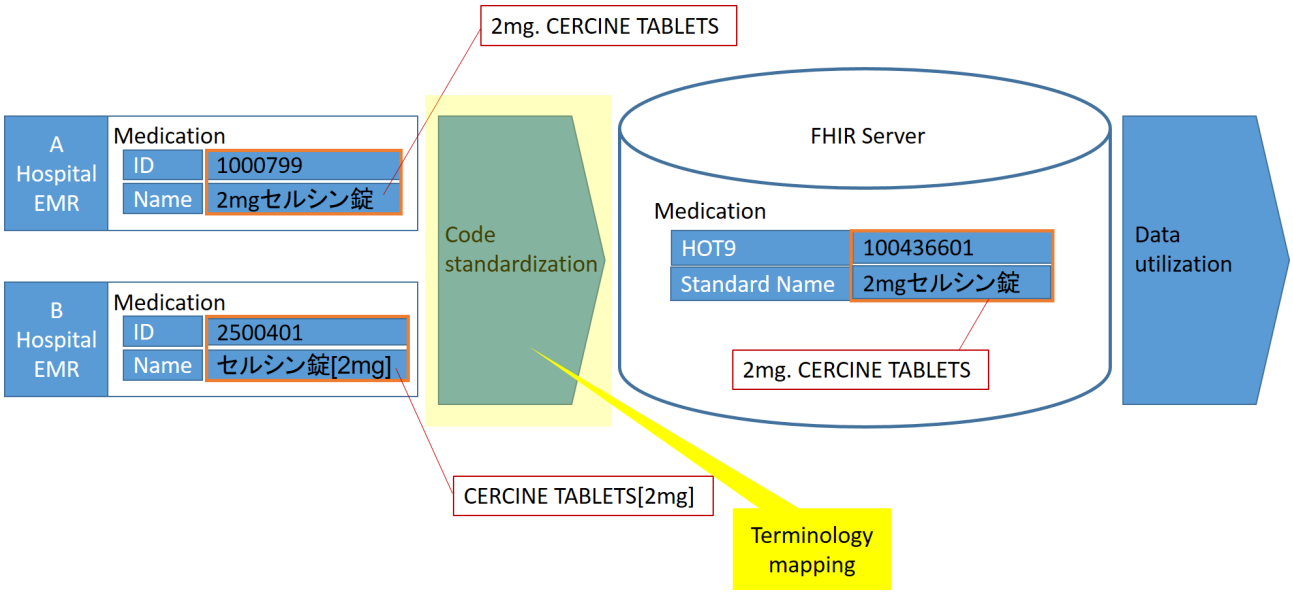


Figure 6. Our proposed solution: Satusehat with federated architecture [27]. The central server periodically fetch and add Fast Healthcare Interoperability Resources that are stored in local servers with the aim to enhance reliability and robustness of the platform.



XSL•FO
RenderX

Figure 7. FHIR conversion framework to address mapping issues [14]. Japanese terminology for medication data, in the name attribute (2mg.CERCINE TABLETS) is preserved in hospital EMR to ensure readability. The corresponding standardized code (such as *ICD-10* [*International Statistical Classification of Diseases, Tenth Revision*]) is mapped into the FHIR server with terminology mapping. EMR: electronic medical record; FHIR: FHIR: Fast Healthcare Interoperability Resources

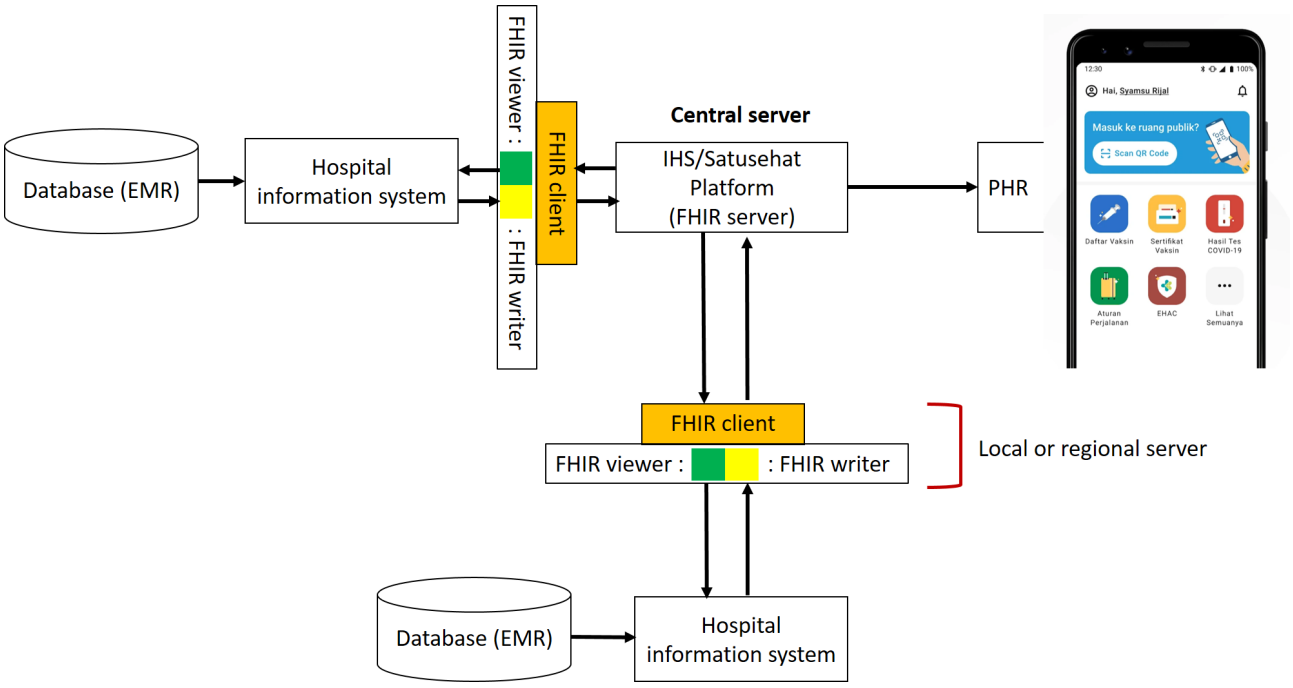


Discussion

Principal Findings

Server issue, mapping issue, and profile issue are identified as the most frequent problems that developers faced when integrating EMR and Satusehat with FHIR protocol. The proposed design for an interoperable FHIR-based Satusehat platform tailored to the needs of *Puskesmas* (community health centers) and private clinics is depicted in Figure 8, which illustrates the combined approaches that have led to a more reliable and user-friendly platform. The design integrates several key features, including a federated architecture that distributes the data, a data mapping system that ensures consistency across diverse health care terminologies, and an automated selection and implementation of appropriate FHIR profiles for EMR data.

Figure 8. Our proposed design for FHIR-based Satusehat platform. It consists of federated architecture between central server and hospital information system in local or regional server as FHIR client, local terminology system mapping, and automated FHIR profile selection. Fast Healthcare Interoperability Resources; IHS: Indonesia Health Services; PHR: personal health record.



To address the mapping and profile issues related to data, we propose an FHIR writer and viewer system that combines the FHIR conversion framework and the SS-MIX system. This system can be deployed on both local and regional servers to ensure accessibility for health care facilities such as *Puskesmas*, hospitals, and private clinics. To use this system, hospital

XSL•FO
RenderX

information systems or EMR systems feed data into the FHIR writer subsystem. The FHIR writer then converts the local code into standardized code and selects the appropriate profile to be inserted with the standardized data. The FHIR profile, filled with standardized data, is then stored or prepared in the local or regional server and is ready to be queried by a central server. If health care facilities want to view patient data from their own facilities, their EMR system can make a request to the local or regional server to view the patient data. If the data are not available in the local or regional server, the server continues the request to the central server to retrieve the data using FHIR API. In this way, the local or regional server acts as an FHIR client, and the central server acts as an FHIR server in terms of the client-server connection [29]. Our proposed FHIR writer and viewer system provides a standardized approach to data mapping and FHIR profile selection, ensuring consistency and ease of use for health care facilities. It also allows for efficient querying of patient data across multiple facilities, increasing the interoperability and accessibility of health data. The issue of server architecture is a critical consideration in developing an efficient and reliable health information system. In particular, the ability to handle large volumes of data and ensure seamless data exchange across different health care providers is crucial. The approach to addressing this challenge lies in the use of federated architecture. Federated architecture offers a scalable and flexible solution to server issues by distributing the load across multiple servers [30]. This ensures that the server load is manageable even during peak periods, such as when there is an automatic processing of EMR data. By spreading the load, federated architecture also enhances the reliability of the system, making it more robust and resilient. Moreover, federated architecture is an ideal solution for countries such as Indonesia, which has a vast archipelago with numerous islands separated by sea. With federated architecture, health facilities in remote areas can be connected to a centralized health information system, enabling the efficient exchange of health data and improving patient care. To fully leverage the benefits of federated architecture, an automated FHIR writer or viewer system can be implemented. This system enables the automatic conversion of local codes to standardized codes, the selection of appropriate FHIR profiles, and the storage and retrieval of patient data from local and regional servers. The system can be easily deployed and customized to suit the specific needs of *Puskesmas*, private clinics, and hospitals. By using this approach, health care providers can efficiently access patient data, resulting in improved health care outcomes for patients. Although our proposed design is based on a small amount of chats in the Satusehat Telegram group, we recognize that this group is the most interactive channel for developers seeking support in integrating their EMR systems into the Satusehat platform. Therefore, despite the limited data, we believe that most of the issues related to integrating EMR systems into Satusehat using the FHIR framework have been recorded in the Satusehat Telegram group. By analyzing and addressing these issues, we were able to develop a design that effectively addresses the common challenges faced by developers. We also recognize that ongoing monitoring and analysis of the Satusehat Telegram group will be critical for identifying and addressing new issues that arise as the platform continues to evolve. Our proposed design for the FHIR-based Satusehat platform can be a valuable solution for other countries facing similar challenges in implementing HIE systems. While we drew inspiration from the design of Japan's HIE system, which shares similar characteristics with Indonesia, such as a non-English language and multiple islands, our approach is adaptable and scalable to meet the needs of other countries. By collaborating and sharing knowledge, countries facing similar issues can learn from each other to enhance the implementation of FHIR-based platforms such as Satusehat. This approach could foster the development of a regional-level HIE system in ASEAN countries or even beyond, incorporating other Asian countries such as Japan, Taiwan, and India, which have already implemented HIE systems [31]. By working together, we are able to create a more efficient and effective HIE system that benefits all participating countries.

Limitations

The content analysis in the study was limited into specific time frame (December 20, 2022, to January 29, 2023) and a small size sample of 107 chats. It was originated from the developer hub's Telegram group that was managed by the DTO of Indonesia MoH. The telegram also addresses only the technical pain points from developers' perspective to integrate their EMR to Satusehat platform. Also, the proposed approaches based on the identified pain points have not been implemented to the existing Satusehat platform.

Future Works

After gaining an understanding of the challenges faced by the Satusehat platform and proposing an interoperability design for the platform, our next step is to develop an FHIR writer or viewer system that is tailored to the Satusehat platform. To accomplish this, we plan to collaborate with an EMR system developer for the purpose of collecting EMR data. These data will be used to train and test the FHIR writer or viewer system, to determine whether the EMR data can be standardized based on the FHIR-based Satusehat platform, and to determine whether it can be viewed back by the same EMR system. We are also interested in collaborating with researchers from Japan, whose approach to building the SS-MIX system has inspired our proposed design. By working together, we aim to create an FHIR-compliant version of SS-MIX that can not only be used in the Satusehat platform in Indonesia but also has the potential to be implemented in Japan. With this joint collaboration, the FHIR-based SS-MIX can facilitate cross-country border HIE by ensuring interoperability and seamless exchange of health data between the 2 countries.

Conclusions

For ensuring successful implementation of the Satusehat platform, this paper proposes content analysis as a way to identify pain points that are faced by developers during development. Analyzing Satu Sehat developers' discussions on social networking system, we have identified 3 top pain points about Satusehat platform: server issues, data mapping, and difficulties managing user profiles. We have showed how using federated server architecture, an FHIR writer and an FHIR viewer, and an FHIR conversion framework could address the

XSL•FO
RenderX

mentioned issues. The proposed architecture for solving pain points gives several suggestions as a blueprint on how to reduce developer workload when committing integration, increasing the robustness of the resources exchange, and ensure accuracy and consistency on mapping and also profiling process. All of this will further help the advancement of Satusehat platform implementation in *Puskesmas* (community health centers) and private clinics in Indonesia.

Acknowledgments

This research was conducted without the use of ChatGPT or any similar generative artificial intelligence model for ideation, method execution, manuscript writing, and referencing process. This work is the result of collaborative works and within the authors. The information and sources that are included in this work have been carefully checked for accuracy. This research was made possible by the collaboration with Medical informatics Lab in Kyoto University, Japan, and Software and Data Engineering Lab in Universitas Gadjah Mada, Indonesia. This work was supported by the Department of Computer Science and Electronics, Universitas Gadjah Mada, under the Publication Funding Year 2025.

Data Availability

The datasets generated during and/or analyzed during this study are not publicly available due to the Satusehat privacy policy but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Most prominent pain points perceived by developers when testing Fast Healthcare Interoperability Resources from the Telegram group (developer hub).
[PNG File, 29 KB - formative_v9i1e51270_app1.png]

References

1. Sonoda T. Evolution of electronic medical record solutions. Fujitsu Sci Tech J 2011 Jan 1;47:19-27.
2. Ishigure Y. Trends, standardization, and interoperability of healthcare information. NTT Tech Rev 2011;9(4):57-62. [doi: 10.53829/ntr201104gls]
3. Joyia GJ, Akram MU, Akbar CN, Maqsood MF. Evolution of health level-7: a survey. Presented at: Proceedings of the 2018 International Conference on Software Engineering and Information Management; Jan 4, 2018; Casablanca, Morocco p. 118-123. [doi: 10.1145/3178461.3178480]
4. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. J Am Med Inform Assoc 2016 Sep;23(5):899-908. [doi: 10.1093/jamia/ocv189] [Medline: 26911829]
5. Xiao D, Song C, Nakamura N, Nakayama M. Development of an application concerning fast healthcare interoperability resources based on standardized structured medical information exchange version 2 data. Comput Methods Programs Biomed 2021 Sep;208:106232. [doi: 10.1016/j.cmpb.2021.106232] [Medline: 34174764]
6. Faulkenberry JG, Luberti A, Craig S. Electronic health records, mobile health, and the challenge of improving global health. Curr Probl Pediatr Adolesc Health Care 2022 Jan;52(1):101111. [doi: 10.1016/j.cppeds.2021.101111] [Medline: 34969611]
7. Dullabh P, Heaney-Huls K, Hovey L, et al. The technology landscape of patient-centered clinical decision support—where are we and what is needed? Stud Health Technol Inform 2022 Jun 6;290:350-353. [doi: 10.3233/SHTI220094] [Medline: 35673033]
8. Harahap NC, Handayani PW, Hidayanto AN. The challenges in integrated personal health record adoption in Indonesia: a qualitative analysis of regulatory perspectives. Presented at: 2021 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS); Oct 2021; Jakarta, Indonesia p. 169-174. [doi: 10.1109/ICIMCIS53775.2021.9699353]
9. Ministry of Health of the Republic of Indonesia. Blueprint for digital health transformation strategy 2024. 2021. URL: https://dto.kemkes.go.id/ENG-Blueprint-for-Digital-Health-Transformation-Strategy-Indonesia%202024.pdf [accessed 2024-02-25]
10. Chong HY, Allotey PA, Chaiyakunapruk N. Current landscape of personalized medicine adoption and implementation in Southeast Asia. BMC Med Genomics 2018 Oct 26;11(1):94. [doi: 10.1186/s12920-018-0420-4] [Medline: 30367635]
11. Saputra Y, Ashila MN, Muliarini P. Readiness and acceptance of electronic medical records among health professionals in indonesia. Presented at: Proceedings of International Conference of Graduate School on Sustainability; Nov 4-5, 2022; Malang, Indonesia p. 81-92.

XSL•FO
RenderX

12. Handayani PW, Saladdin IR, Pinem AA, Azzahro F, Hidayanto AN, Ayuningtyas D. Health referral system user acceptance model in Indonesia. Heliyon 2018 Dec;4(12):e01048. [doi: 10.1016/j.heliyon.2018.e01048] [Medline: 30603690]
13. Postman. URL: https://www.postman.com/satusehat/satusehat-public/overview [accessed 2023-02-19]
14. Lee YL, Lee HA, Hsu CY, Kung HH, Chiu HW. Implement an international interoperable PHR by FHIR—a Taiwan innovative application. Sustainability 2020;13(1):198. [doi: 10.3390/su13010198]
15. Harbi A. Health care expert's readiness to implement National Unified Medical Records (NUMR) system in the United Arab Emirates: a qualitative study. IJCAI 2021;45(5). [doi: 10.31449/inf.v45i5.3358]
16. Bender D, Sartipi K. HL7 FHIR: an agile and restful approach to healthcare information exchange. Presented at: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems; Jun 2013; Porto, Portugal p. 326-331.
17. DTO kementerian kesehatan—home page. dto.kemkes.go.id. URL: https://dto.kemkes.go.id/ [accessed 2023-02-19]
18. Renz SM, Carrington JM, Badger TA. Two strategies for qualitative content analysis: an intramethod approach to triangulation. Qual Health Res 2018 Apr;28(5):824-831. [doi: 10.1177/1049732317753586] [Medline: 29424274]
19. Kim H, Jang SM, Kim SH, Wan A. Evaluating sampling methods for content analysis of Twitter data. Soc Media + Soc 2018 Apr;4(2):205630511877283. [doi: 10.1177/2056305118772836]
20. O'Connor C, Joffe H. Intercoder reliability in qualitative research: debates and practical guidelines. Int J Qual Methods 2020 Jan 1;19:160940691989922. [doi: 10.1177/1609406919899220]
21. Cilliers L, Viljoen K. A framework of ethical issues to consider when conducting internet-based research. SA J Inform Manage 2021;23(1):a1215. [doi: 10.4102/sajim.v23i1.1215]
22. Telegram. Telegram privacy policy. URL: https://telegram.org/privacy/id?setln=en [accessed 2024-04-23]
23. Tripathy JP. Secondary data analysis: ethical issues and challenges. Iran J Public Health 2013 Dec;42(12):1478-1479. [Medline: 26060652]
24. Döring M. Evaluating an approach for mapping FHIR profiles to research protocols. Doctoral dissertation.: Hochschulbibliothek der Technischen Hochschule Köln; 2018.
25. Japan Association for Medical Informatics. SS-MIX2 standardized storage' explanation of the structure and guidelines for implementation ver. 1.2. 2014. URL: https://www.jami.jp/jamistd/docs/SSMIX2/descript-implemglonSS-MIX2_V1.2.pdf [accessed 2023-02-19]
26. TIBCO. TIBCO product documentation. docs.tibco.com. URL: https://docs.tibco.com/pub/bwpluginhl7/8.1.0/doc/html/GUID-20E40FAF-6F0B-41D6-8257-3387607178DA.html [accessed 2023-02-19]
27. Vesteghem C, Brøndum RF, Sønderkær M, et al. Implementing the FAIR Data Principles in precision oncology: review of supporting initiatives. Brief Bioinform 2020 May 21;21(3):936-945. [doi: 10.1093/bib/bbz044] [Medline: 31263868]
28. Quan H, Li B, Saunders LD, et al. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. Health Serv Res 2008 Aug;43(4):1424-1441. [doi: 10.1111/j.1475-6773.2007.00822.x] [Medline: 18756617]
29. Aljunid SM, Srithamrongsawat S, Chen W, et al. Health-care data collecting, sharing, and using in Thailand, China mainland, South Korea, Taiwan, Japan, and Malaysia. Value Health 2012;15(1 Suppl):S132-S138. [doi: 10.1016/j.jval.2011.11.004] [Medline: 22265060]
30. Hisashi Osanai—effort to internationalize a FHIR implementation guide in Japan | devdays. www.youtube.com. 2022. URL: https://www.youtube.com/watch?v=HwAxcdKfbxY [accessed 2023-02-19]
31. Dornan L, Pinyopornpanish K, Jiraporncharoen W, Hashmi A, Dejkriengkraikul N, Angkurawaranon C. Utilisation of electronic health records for public health in Asia: a review of success factors and potential challenges. Biomed Res Int 2019;2019:7341841. [doi: 10.1155/2019/7341841] [Medline: 31360723]

Abbreviations

API: application programming interface
DTO: Digital Transformation Office
EHR: electronic health record
EMR: electronic medical record
FHIR: Fast Healthcare Interoperability Resources
HIE: health information exchange
HL7: Health Level Seven
***ICD-10*:** *International Statistical Classification of Diseases, Tenth Revision*
***ICD-9-CM*:** *International Classification of Diseases, Ninth Revision, Clinical Modification*
LOINC: Logical Observation Identifiers Names and Codes
MoH: Ministry of Health
PHR: personal health record
SNOMED-CT: Systematized Nomenclature of Medicine Clinical Terminology
SS-MIX: Standardized Structured Medical Record Information eXchange

XSL•FO
RenderX

Edited by A Mavragani; submitted 27.07.23; peer-reviewed by B Senst, G Grieve, J Frimpong; revised version received 20.05.24; accepted 26.05.24; published 21.04.25.

Please cite as:
Heryawan L, Mori Y, Yamamoto G, Kume N, Lazuardi L, Fuad A, Kuroda T
Fast Healthcare Interoperability Resources (FHIR)–Based Interoperability Design in Indonesia: Content Analysis of Developer Hub's Social Networking Service
JMIR Form Res 2025;9:e51270
URL: https://formative.jmir.org/2025/1/e51270
doi:10.2196/51270

© Lukman Heryawan, Yukiko Mori, Goshiro Yamamoto, Naoto Kume, Lutfan Lazuardi, Anis Fuad, Tomohiro Kuroda. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Informal Caregivers Connecting on the Web: Content Analysis of Posts on Discussion Forums

Michelle L Foster[1], MPH; Chinenye Egwuonwu[1,2], MPH, MD; Erin Vernon[3], MA, PhD; Mohammad Alarifi[1,4], MS, PhD; M Courtney Hughes[1], MS, PhD

[1]School of Health Studies, Northern Illinois University, DeKalb, IL, United States

[2]Department of Internal Medicine, Piedmont Athens Regional, Athens, GA, United States

[3]Department of Economics, Seattle University, Seattle, WA, United States

[4]Radiological Sciences Department, College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia

Corresponding Author:
M Courtney Hughes, MS, PhD
School of Health Studies, Northern Illinois University, DeKalb, IL, United States

Abstract

Background: About 53 million adults in the United States offer informal care to family and friends with disease or disability. Such care has an estimated economic value of US $600 million. Most informal caregivers are not paid nor trained in caregiving, with many experiencing higher-than-average levels of stress and depression and lower levels of physical health. Some informal caregivers participate in web-based forums related to their caregiving role.

Objective: This study aimed to explore how informal caregivers use easy-to-access caregiving web-based forums, including the types of information they share and seek from others. It also aimed to gain insights into the informal caregiver experience from the content these informal caregivers posted.

Methods: The study population consisted of participants who posted on 5 web-based forums for informal caregivers between February and April 2024. Researchers extracted the first 6 responses to the first 20 questions and comments to appear posted by the informal caregivers in each of the 5 forums, removing any individually identifying information. We used a codebook thematic analysis approach to examine the data with Dedoose (SocioCultural Research Consultants). Researchers independently read all posts and coded the data. The author group discussed the codes, reiteratively refined them, and identified themes within the data.

Results: The data consisted of 100 initial posts and 600 responses. Over half of the initial posts included specific questions, with the remaining initial posts sharing experiences or reflections. Posts ranged in length from a sentence to more than 500 words. Domains identified included handling interpersonal challenges, navigating complicated systems, gathering tactical coping strategies, managing emotions, and connecting with others in similar situations. Negative interpersonal interactions were mentioned 123 times, with 77 posts describing challenging situations with extended family. Posters inquired about accessing resources, with health care and health insurance included 51 times, while legal and financial concerns were addressed 124 times. Caregiving challenges were mentioned hundreds of times, including discussion of hygiene (n=18), nutrition (n=21), and desire for a caregiving break (n=47). Posters expressed emotion in their comments 180 times, which included 32 mentions of guilt and 26 mentions of positive emotion. The importance of web-based group support was mentioned 301 times.

Conclusions: Informal caregivers play an essential role in society. Many experience multifaceted challenges related to their caregiving role, and some turn to the internet for community. Accessing web-based discussion forums is a low-barrier method for informal caregivers to connect with others who may be experiencing similar emotions and challenges. Gaining a greater understanding of the ways informal caregivers seek advice and offer support to one another provides insight into the challenges they face. The domains identified on these forums may be helpful, as clinicians provide information to care recipients and their informal caregivers along their health journeys.

(JMIR Form Res 2025;9:e64757) doi:10.2196/64757

KEYWORDS

informal caregivers; family caregivers; discussion forum; caregiver support; support group; social support; caregiver navigation; content analysis; adults; United States; informal care; codebook; thematic analysis; web-based discussion; web-based forums; clinicians; medical care; peer-to-peer support; web-based communities; caregivers

XSL•FO
RenderX

Introduction

About 53 million informal caregivers assist individuals in need of medical care and activities of daily living in the United States [1]. The unpaid value of this care was about US $600 billion, up from a 2017 estimate of about US $470 billion value of unpaid informal care. Informal caregivers are generally unpaid and untrained, and historically, these caregivers performed household tasks and personal care activities in contrast to formal caregivers who were part of the health system [2]. Research shows that more than 58% of informal caregivers are responsible for medical or nursing tasks for their care recipients [3].

Informal caregiving is associated with both benefits and burdens for caregivers. Many informal caregivers experience higher levels of stress and depression and have lower levels of physical health than their counterparts [4-6]. Care recipients with poor levels of physical and mental health were associated with worse caregiving experiences [5,6]. Furthermore, informal caregivers with older age or mental health issues have been shown to have worse caregiving experiences [5,7].

Most informal caregivers have internet access and use it to obtain health information [8]. There is significant research regarding the use of web-based tools among informal caregivers [9-18]. Some research used web-based surveys to ask about informal caregiver needs [10,11,17], while others evaluated the use or content of various purpose-built social support service interventions [13-15,18]. The qualitative research in this space has focused on specific topics related to caregiving, such as well-being during COVID-19 [19], advance care planning [20], and developing the identity of an informal caregiver [21]. Many other qualitative analyses have focused on particular diagnoses, such as care recipients with Alzheimer disease or dementia [22-25], posttraumatic stress disorder [26], stroke [27,28], and cancer [29-31]. Qualitative research about general topics addressed in informal caregiver forums focused only on 1 discussion platform (Reddit) and was evaluated with the use of artificial intelligence [32].

The previous studies investigating informal caregivers' use of web-based forums have found that informal caregivers use web-based support groups and forums to find emotional support [13-15,19,20,22-28,30], request information [14,20,23,26-28,30], discuss symptoms and conditions [23,24,27,28], interact with other caregivers [13,15,19-21,23,26,30], and talk about the challenges of caregiving [15,19,20,22,23,25,26]]. Additionally, all existing studies but one [27] evaluated data from 1 specific web-based forum rather than more than 1 forum. The study herein aims to help fill the gap in research that examines how informal caregivers use caregiving web-based forums that are not restricted to just 1 forum. This is also the first study to examine the forums that appear first when a caregiver searches for chronic illness caregiver support forums on Google. This approach can shed light on how caregivers interact with one another across multiple caregiving web-based forums.

This study explores how informal caregivers used a variety of popular, easy-to-access web-based forums. Specifically, we sought to discover the types of information individuals posting to web-based informal caregiver forums shared and sought from others. Insights into the informal caregiver experience from exploring multiple discussion forums can inform health care providers and other community groups so they may better design interventions and policies responsive to caregiver needs.

Methods

Study Population

The study population consists of participants who posted on publicly accessible web-based forums for informal caregivers between February and April 2024. Descriptions of the forums (AgingCare [33], Caregiver Action Network [34], Facebook Dementia Caregivers Support Group [35], Facebook Empowering Caregivers Support Group [36], and Mayo Clinic Caregivers Forum [37]) are presented in Table 1. We selected these forums because they appeared as the top 5 forums in a web-based Google search using the phrase "caregiver support forum and chronic illness."

XSL•FO
RenderX

Table . Description of web-based caregiver forums from each forum's website.

Forum name	Description	Details
AgingCare [33]	"Connects families who are caring for aging parents, spouses, or other elderly loved ones with the information and support they need to make informed caregiving decisions."	• Caregivers seek mutual support for medical challenges that occur with aging such as dementia • Caregivers offer valuable suggestions based on firsthand experience managing these conditions
Caregiver Action Network[a] [34]	"The nation's leading family caregiver organization working to improve the quality of life for the more than 90 million Americans who care for loved ones with chronic conditions, disabilities, or disease, as well as those supporting the living needs of older adults."	• Caregivers seek support for general, day-to-day issues in caregiving • Caregivers share personal experiences while caring for their loved ones with disabilities or disease
Facebook Dementia Caregivers Support Group [35]	"A group for caregivers taking care of Dementia and Alzheimer's afflicted loved ones. Designed to stimulate conversation and ensure all members a safe haven in posting personal feelings."	• Caregivers seek support for general challenges with caregiving • Caregivers seek support for caring for loved ones with neurocognitive behavioral changes and degenerative diseases such as delirium and dementia
Facebook Empowering Caregivers Support Group [36]	"A public group created so that there is a place for family members to turn to when in need of mental, emotional, and informational support when it comes to taking care of aging family members."	• Caregivers seek support for the day-to-day challenges of aging family members • Caregivers seek support on specific information such as health system challenges and legal issues
Mayo Clinic Connect [37]	"An online community, connecting patients and family caregivers with each other. You do not have to be a Mayo Clinic patient or caregiver to join the conversations."	• Many forum members are caregivers of patients with chronic medical conditions, such as cancer, who receive treatment at the Mayo Clinic, though this is not required • Caregivers typically seek support for medical challenges, such as medication side effects

[a]Caregiver Action Network's caregiver forum was on its website at the time of the research. It has since moved to Facebook.

Data Analysis

Two researchers (MA and CE) extracted the first 6 responses from the first 20 questions and comments to appear posted by the caregivers in each of the 5 forums. We chose this method based on the passive or unobtrusive method, which involves techniques that allow researchers to collect data without actively engaging with participants or influencing the conversation [38]. We did not extract questions and responses that did not contain comments relating to caregiver support (spam posts) because they were irrelevant to the research topic. To protect the identity of all posters on the web-based forum, we removed any names that appeared in the comments.

We used a subjective approach to gain insights into the caregiver forum participants' perceptions of their experiences. Specifically, we adopted an approach based on grounded theory using an iterative inductive process for our qualitative thematic analysis. As such, the number of comments and posts collected were adjusted iteratively until 2 authors (MLF and CE) separately believed that thematic saturation within forums' comments and posts was reached [39]. Similarly, the authors began with 2 forums and then added the next highest publicly accessible forum listed on Google until they achieved thematic saturation across forums. Using a codebook thematic analysis approach to examine the data [40] with Dedoose (SocioCultural Research Consultants), 2 researchers (MLF and CE) read through all the posts and coded the data. The entire author group discussed the codes, refined them reiteratively, and identified themes within the data. The author group also resolved any coding disagreements and resulting frequencies of each theme through discussions and collective agreements. They also examined the data to ensure that the chosen themes accurately reflected the data. Additionally, our author group comprised researchers with varying caregiving experiences. However, none of the researchers have intense caregiving experience that would deeply influence the research.

Ethical Considerations

The institutional review boards at Northern Illinois University and Seattle University determined the study to be exempt from institutional review board review in accordance with federal regulation criteria. The data used in this study were collected from forums that are easy to access and generally open to the public. All data were deidentified before analysis, including removing usernames and other naming data to maintain anonymity and confidentiality. There was no contact between the researchers and the posters.

XSL•FO
RenderX

Results

Overview

The data consisted of 100 initial posts and 600 responses to the initial posts or conversations stemming from the initial posts. Over half of the initial posts included specific questions, while the others shared reflections or experiences. The length of the initial posts and responses varied from a few with just a short sentence to others that were a couple of paragraphs. A few stretched beyond 500 words—a full page of single-spaced text. Response lengths were similar, with ranges from a few words to more than 500.

Posters (our term for individuals posting anything, including initial posts and responses) to the forums discussed care recipients with various diseases and cited dementia and cancer the most. We found some variation in the care recipient's disease based on the forum used. The Mayo Clinic Caregivers Forum included more clinical posts, with posters sharing specific disease states. In the Facebook Dementia Caregivers Support Group, some posters asked about what medications had worked for other care recipients and for strategies to obtain particular treatments, with almost all having to do with dementia. In the other forums, the care recipient's diseases often were not apparent, with many posters omitting such details.

Initial posts often included significant details describing the poster's life with their care recipient. Some posters responded with advice or feedback on the situation, while others shared their own experiences with their care recipients. Below, we present 5 key domains that resulted from the themes.

Domains

Handling Interpersonal Challenges

In these forums, posters shared their varying daily interactions, ranging from toileting and feeding to medication management and household chores. These interactions sometimes led to interpersonal challenges with the person receiving care or others involved in the household, with such negative interactions recorded 123 times in these posts. One poster noted his care recipient wife was upset with how he did things around the house.

> *She is very exacting on how she wants things done. I am hard of hearing and she gets upset when she has to repeat herself. If I guess at what she wants and get it wrong, oh well, you get the picture. I have to control myself better and be there for her. I have to swallow my pride and do what is right for her. It is not easy.* [Mayo Clinic Connect]

In total, 10 posts described challenges related to changing relationship roles. Husbands and wives recounted working to feed their spouses and care for them throughout the day. The posters shared their feelings, which often included frustration and sadness about these changes.

> *I lost an equal parent, as well as my romantic and "relational" partner, partly. It's a loss that is fully understood only by other spousal caregivers. It hasn't gotten any easier after 6 years.* [Caregiver Action Network]

Some posters also described negative situations with extended family members, with such interactions reported 77 times. Posters shared that some siblings were only concerned about financial arrangements after death, while other siblings were described as expressing negative attitudes and overall dissatisfaction with care.

> *She is very difficult to deal with. She questions everything and does not trust me to care for our parents in a manner that is in THEIR best interest. Without details suffice to say my husband and I gave up our dreams to take care of my parents but she berates me, tries to put my other siblings against me, says I lie, and accuses me of taking things.* [Facebook Empowering Caregivers Support Group]

Other posters described a lack of family support for the care recipient and feeling isolated as the main caregiver. This was highlighted 33 times in these posts, with some posters, noting that they did not get breaks or have family visitors who could relieve the ever-present demands of caregiving. Other posters wrote that this constant responsibility negatively impacted their relationships with their own spouses and children.

> *My husband and kids are not getting the best version of me. I can't stand having sex now because I spend so long caring for my parents the idea of sharing intimacy with my husband is actually repulsive to me. I don't enjoy anything that I'm doing.* [AgingCare]

Navigating Complicated Systems

Posters asked and responded to questions about accessing resources related to health care, living arrangements, and financial and legal systems. They mentioned health care and health insurance systems in 51 posts. Other posters responded by sharing their experiences with the health care system as a guide for others and sometimes suggesting various social services. Information on accessing home health aides, hospice care, and various medicines was requested 47 times in posts. Ten discussed necessary steps for care recipients to qualify for nursing home access through Medicaid, such as spending down funds or putting assets into a trust.

> *I can't wait until my mom is approved for Medicaid. She's in the nursing home. We [are] just waiting on that. I spent all her money. Most of it went to the nursing home. Also when I sent information to Medicaid I sent copies of all checks to show where her money went.* [Facebook Dementia Caregivers Support Group]

> *I have been her sole caregiver for about 5 years and I know caring for her now is beyond my capabilities. What do people do when they can no longer keep their parent in the home? Money is a huge issue, her insurance doesn't want to cover anything. I don't know what to do at this point. She needs more than I can give her.* [Facebook Empowering Caregivers Support Group]

XSL•FO
RenderX

Posters expressed concern about the quality of care hospice or nursing home workers provided to their care recipients 23 times and asked for feedback on how to advocate. They described their situation and asked for advice on what to do when the nursing care was not working out as expected.

> *The nurses have asked that when I come to visit I help them get her to the shower ... I can't be there all the time to make sure she gets up, takes a shower, just participates with life? The CNAs try to help, the nurses have told me that they can't force her because they could get into trouble.* [AgingCare]

Posters responded with messages stating how they understand these caregivers' concerns. They offered ideas on how they had dealt with similar situations and shared details about similar struggles they faced in working with care homes or hospice providers.

Posters described legal and financial concerns related to caregiving 124 times and asked others to weigh in on their options. Posters sometimes asked how to work within the legal system to help their loved ones. Some individuals receiving care wanted the informal caregiver to inherit a greater portion of assets upon their death, and posters asked about how to make the care recipient's wishes clear. Respondents offered suggestions to these legal and financial concerns, often urging posters to consult with professionals.

> *If your mom is of sound mind and wants to make this change, that's her decision. With the no-contest order and the doctor's assessment, legal challenges from your siblings might not hold up. Considering your exhaustion, it might be worth seeking legal advice to prepare, but ultimately, your well-being is the priority.* [Facebook Empowering Caregivers Support Group]

Gathering Tactical Coping Strategies

Posters described hundreds of challenges related to their caregiving roles. They asked questions about how to address these concerns, and respondents offered ideas or shared how they had responded in similar circumstances.

One common theme was difficulties related to hygiene among those receiving care, highlighted in 18 posts. Posters described challenges related to body odors and getting care recipients to shower or bathe. One poster asked for help with their care recipient who was touching foods in their kitchen pantry, which caused the poster concern about contamination of foods due to dirty hands. Fellow posters pitched in with various ideas.

> *Find a lock for the pantry but leave snacks out for him on the counter that are just for him. My Mom goes prowling for snacks at night so we leave some out for her.* [Facebook Dementia Caregivers Support Group]

> *I keep extra bags of chips because I can't eat out of the ones already opened. I have to constantly remind my [loved one] to wash her hands so I know if I'm not here they are probably not getting washed and she takes her dentures in and out of her mouth which totally grosses me out.* [Facebook Dementia Caregivers Support Group]

Concern about a loved one eating too much sugar or other nutrition concerns were mentioned in 21 posts. Respondents said that they experienced the same problems and highlighted that desire for sweets is common among people with dementia. Other posters responded with a variety of recipes and food ideas.

Posters expressed a desire for a break from caregiving duties 47 times. Respondents offered helpful tips, such as enlisting community or church volunteers. Enrollment in palliative or hospice care was mentioned in 55 posts, and posters shared their experiences engaging in those services.

> *It wasn't until she was in home hospice that they were ever able conquer her incredibly high anxiety. We certainly realized no course of care is ever perfect and always knew palliative care was her only choice given her situation. We never regretted the care she received.* [Mayo Clinic Connect]

Managing Emotions

Posters expressed emotions in their comments 180 times. Some described themselves as being "stressed," "overwhelmed," "weary," or "experiencing burnout." They talked about their emotions, which often included anger, depression, sadness, and resentfulness. Posters mentioned feelings of guilt 32 times, and some described guilt for having negative feelings about caregiving and their loved ones.

> *I feel so guilty for the anger and resentment I feel. I always thought in our 50's we would be traveling and enjoying ourselves and there is none of that. Whenever I think he is finally improving enough that we can do something enjoyable, he comes up with a new problem or complaint, which leads to more Dr. appointments, tests and procedures ... I feel sorry for myself and wonder what horrible things I did to deserve the life I have now.* [Caregiver Action Network]

Posters said they cried frequently 7 times. They described facing challenging decisions about their loved one's care and the emotional toll of such choices.

> *My heart is breaking because I'm not sure if I'm ready to let him go but at the same time I know I can't keep living like this, I'm feeling guilty cuz it feels like I'm giving up on him ... This is the hardest decision I've ever had to deal with but I know it's the best decision, but then why do I feel so guilty?* [Mayo Clinic Connect]

Isolation was another common feeling described in 33 posts. Some lived in rural areas where there were few community groups or other social outlets. Others described a lack of connection to friends and other family members and difficulty engaging in employment.

> *My friendships are starting to be strained, because they don't understand why "I'm hardly ever available" and quite frankly I'm starting to get bitter how they can just stay out as long as they want*

XSL•FO
RenderX

> *without worry. Romantic relationships are nonexistent, as most girls immediately become disinterested when they find out about my home situation. And in general I've found it harder to relate with other people, as my only social contact is in class or work.* [Caregiver Action Network]

Posters shared negative emotions 123 times and reported anger about the way their care recipient's medical condition had impacted their lives, calling the situation unfair. Posters also described helping their care recipient deal with negative emotions 40 times, including anger about their diagnosis. Some posters detailed navigating emotional outbursts from the care recipient that were emotionally hurtful to the poster.

> *The last day that my Dad was home before going to a hospice facility, he kicked me when I was trying to clean him after a BM. He also punched at me and said "I'm going to beat you up!" Which wasn't his normal personality- he was the sweetest, gentlest Dad usually. I know he didn't feel good and he was embarrassed to need that much help, but I'm human and it still hurt my feelings.* [Facebook Empowering Caregivers Support Group]

A few posters expressed negative emotions toward themselves—saying they were "horrible" or a "monster" when describing difficult decisions involved in caregiving. This was often expressed along with frustration. One poster detailed how her husband had been put on hospice, then taken off, and lived more than 4 more years facing many health challenges. "WHY IS HE YET ALIVE[?]" she asked. Another described her internal wishes.

> *I pray every day my mother has a heart attack or a massive stroke.* [AgingCare]

While the majority of posts with emotion had a negative tone, 26 were positive. For example, some posters expressed fulfillment in their caregiving duties. Others said they were proud of the care they had been able to offer their loved ones, and some noted personal growth during the caregiving experience.

> *Over the years, I can see a huge change in myself. For one, I am more patient now. However, the goal line of ENOUGH patience keeps moving further away! I believe that my husband is providing me an opportunity to become a better person.* [Mayo Clinic Connect]

Connecting With Others in Similar Situations

Many posters expressed a desire to connect with others in similar situations, with web-based group support mentioned 301 times. As specific posters communicated more with one another, their comments took on a more familiar tone. A few even offered each other direct assistance.

> *If you are a man of faith, use it. If not, you may want to consider it and there are those who can help you with that. I'll help you with that if I can. There is LOTS of help out there and there are LOTS of people that understand and are ready to be your friend and your lifeline.* [Caregiver Action Network]

> *Know that there are people that care. Like [name redacted] said, he will help, I will too, we are both reaching out to you, you are not alone.* [Caregiver Action Network]

Spiritual conversations appeared in posts 36 times. In some instances, the posters shared a positive spiritual statement as an initial post. In other instances, respondents promised to pray for the initial poster regarding the challenging situation or their care recipient's medical condition.

> *As a caregiver myself to my husband, I've never felt more useless in my life. Feels like all you can do is watch and wait. There are days I'm afraid to go to sleep. I do pray for healing over your husband, and peace that surpasses all understanding for you. Bless you both.* [Mayo Clinic Connect]

Posters mentioned feeling lonely and shared their challenges with that across 20 posts. They brought up hobbies they no longer engaged in and not having time for friends or socialization. While respondents said that they had similar difficulty making time for themselves, they often prescribed self-care to one another.

> *Coping with the isolation of being a 24/7 caregiver is very challenging. It's crucial to seek support from friends, family, or here with caregiving friends who understand your journey. Schedule regular breaks, explore local resources, delegate tasks, and set boundaries, which are essential for your well-being.* [Facebook Empowering Caregivers Support Group]

Posters also commiserated about the many tasks involved in caregiving. They described their day-to-day activities and the challenges associated with them. Respondents said they do the same things and appreciated knowing others were facing similar circumstances.

> *I feel like I am forever walking uphill. Frankly, one of the most helpful things I'm finding on here is that I am not alone. I don't really need anyone to tell me what to do or fix this ... just to validate my feelings.* [Caregiver Action Network]

Discussion

Principal Findings

Our qualitative analysis of posts across 5 web-based caregiver forums showed that informal caregivers participate in these communities to share and receive moral and practical support. These forums functioned as a space for caregivers to disclose their daily struggles with those in similar circumstances. This was true across forums that offered more anonymity (AgingCare, Caregiver Action Network, and Mayo Clinic Caregivers Forum) but was also seen in the 2 Facebook groups that had the potential to be tied to caregivers' personal lives.

Our research identified 5 key domains explaining why people posted on these caregiver forums. Within these themes were consistent comments about being heard and supported,

XSL•FO
RenderX

highlighting the positive benefits that popular publicly accessible web-based caregiver forums can provide. Caregivers described putting significant time and effort into their caregiving, and their posts and stories showed how the role was a critical part of their lives. While the 5 domains we identified were similar to those found in prior research on web-based caregiver support that focused on particular topics related to caregiving, one website, or on caregivers of people with a specific diagnosis [17-30], this analysis demonstrated the consistency of these themes across several caregiver forums.

Many caregivers in this analysis experienced significant interpersonal tensions, including resentment and guilt when relating to the care recipient or other family members. Caregiving is often characterized by intense, intimate relationships with unexpected role changes and needs [41]. Research shows that spousal caregiving, in particular, can be more intense, with the informal caregiver experiencing feelings of depression and guilt [42,43]. Research also shows that caregivers caring for individuals younger than them may find more emotional, physical, and financial difficulties than their counterparts [44]. It is possible that caregivers anticipate caring for someone younger than them for a longer time, making it more difficult than caring for someone older. Similarly, our research indicates many caregivers are experiencing challenges with their relationships. These caregivers are reaching out on the web to vent and share their experiences with others, and this research shows that there are many places on the web where caregivers may find community among fellow caregivers likely to understand the relational intricacies and difficult situations they experience.

Informal caregivers used web-based forums to aid in navigating the maze of legal, financial, and health care systems. Previous research shows that informal caregivers experience significant difficulty identifying resources in the fragmented and complex health system and find a lack of consistency in access to community and home care [45]. In addition, there is insufficient community and health infrastructure for sharing resources with the caregiver community [45]. Caregivers use forums for aid related to logistical issues of caregiving in areas including legal, financial, and health care topics [23,32]. In our research, the problem-solving required to navigate these systems often stumped many caregivers and left them feeling frustrated and isolated. While respondents were frequently able to provide helpful steps to guide the original poster toward the desired outcome, our findings support the need for additional informal caregiver support toward navigating the necessary solutions within health and community systems.

In our study, informal caregivers expressed persistent concerns about managing practical daily matters. The range of such daily demands on caregivers is vast, with tasks such as feeding and bathing, paying bills and shopping, caring for property and financial matters, and offering encouragement and companionship. Informal caregivers who help with a greater number of activities of daily living or independent activities of daily living experience more caregiver burden [46]. Additionally, caregiving often includes more medical-focused tasks; yet, practical daily tasks were the dominant type mentioned by posters in our study. Research indicates that support with medical-focused tasks from effective interventions and health care providers can help informal caregivers improve their caregiving experience [47,48]. Web-based caregiver forums are likely a better source for advice for practical daily tasks than medical-focused tasks, given that providing advice on the latter involves having medical expertise.

Every web-based forum we reviewed included numerous caregivers who expressed gratitude and encouragement related to connecting with others walking a similar path in life. Whether it was offering a step-by-step suggestion on how to address a problem, thanking each other for sharing, or simply saying, “Hang in there,” it was clear that the posters in our study appreciated the connection the forum provided. Similarly, a previous study analyzing web-based caregiver forums during COVID-19 identified a common theme of “celebrating” the virtual way of life [19]. In that study, posters highlighted the sense of community in the forum and how engaging with the site had helped them with caregiving during the pandemic.

Our analysis highlighted web-based forums as places where caregivers shared thoughts and sensitive caregiving details that could be considered socially inappropriate—things they might not say in an interview or to someone they know—such as comments on family disagreements, personal hygiene, and the caregiver asking “Why is he yet alive” about her spouse. This finding is consistent with another study that found participants in web-based caregiver groups felt safe sharing thoughts that were not socially acceptable [49]. Additionally, research shows that reciprocity, altruism, and empathy have a greater influence on knowledge sharing on web-based forums for nonhealth professionals than health professionals [50]. These motivations help explain the candid statements made by posters in our study and suggest that the informal caregivers want to help one another.

Many of the posts on the caregiver web-based forums in our analysis were emotionally charged, and posters seemed to use these communities as a way to vent negative emotions. Spousal caregivers, in particular, have been found to share more negative emotion and anxiety than other informal caregivers [22]. Furthermore, previous research found that informal caregivers who share negative aspects of caregiving on the web may ask others to confirm that their feelings are shared [32]. One reason that sharing frustrations on the web may help informal caregivers is that the negative emotions can be posted as soon as they experience the frustration [51].

Our study comes after the COVID-19 pandemic, a challenging time for informal caregivers, who often experienced increased burden from illness for themselves and their care recipient, had to navigate pandemic-related policies like medical office closings, and experienced social isolation [52,53]. Web-based discussion forums provided a medium for connecting with others and gaining information during the pandemic. Older adults, in particular, became more comfortable with using the internet and other technologies during the pandemic, with digital connectivity playing an increased role in their lives even after the pandemic [54]. With many older adults caring for spouses, siblings, and friends, access to web-based caregiver forums is

XSL•FO
RenderX

a tool that aging adults, for one, can continue to give and receive support.

Limitations

Our study had some limitations. Due to the anonymity of the web-based forum posters, information on their descriptive characteristics could not be obtained. Additionally, the web-based caregiver forums we examined were based in the United States. While the internet is available globally, it is likely that the posters were primarily based in the United States and experiencing caregiving in that country. This may limit the generalizability of the study findings on a global scale. For example, many health system challenges highlighted in the study are specific to those in the United States. In addition, this research captures caregivers' posts at one point in time and may not be generalizable in the future, as the caregiver landscape changes over time. Finally, it is possible that some posters were not informal caregivers, even though the content seemed to be posted by informal caregivers.

Conclusions

This paper contributes meaningfully to the current state of the art by providing a rich understanding of informal caregiving through an innovative lens, that of web-based forums. From this work, researchers and practitioners can be better informed about the needs of informal caregivers beyond what such caregivers may mention in interviews or medical appointments. By illuminating the lived experiences shared in web-based forums, the study advances knowledge about caregiving.

Informal caregivers play an essential role in society. Each year, informal caregivers in the United States provide as much as 36 billion hours of care to adults with limitations in daily activities, and such caregiving is linked to significant emotional, physical, and financial burdens to informal caregivers [55]. Since most informal caregivers have internet access [8], accessing web-based discussion forums is a low-barrier method for them to reach a community of people who may be experiencing similar emotions and challenges. With the number of American people aged 65 years and older projected to jump from 53 million [1] to nearly 95 million by 2060, the need for informal caregiving is projected to grow substantially [56]. Our study provides insights into how informal caregivers use web-based caregiver forums and can inform health care providers and community groups who provide information, education, and support to patients and their informal caregivers. In summary, web-based caregiver forums enable informal caregivers to use modern technology to help meet their support needs and, in the process, help to show them that they are not alone in facing caregiving challenges.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. AARP and National Alliance for Caregiving. Caregiving in the United States 2020. AARP. 2020. URL: https://www.aarp.org/pri/topics/ltss/family-caregiving/caregiving-in-the-united-states/ [accessed 2025-01-09]
2. Reinhard SC, Young HM, Levine C, Kelly K, Choula RB, Accius J. Home alone revisited: family caregivers providing complex care. AARP. 2019. URL: https://www.aarp.org/pri/topics/ltss/family-caregiving/home-alone-family-caregivers-providing-complex-chronic-care/ [accessed 2024-12-30]
3. Keeton VF, Trask J, Whitney R, Bell JF. Overburdened and underprepared: medical/nursing task performance among informal caregivers in the United States. J Gerontol Nurs 2020 Sep 1;46(9):25-35. [doi: 10.3928/00989134-20200811-05] [Medline: 32845344]
4. Coughlin SS, Datta B. Predictors of unhealthy physical and mental days among informal cancer caregivers: results from the 2019 Behavioral Risk Factor Surveillance System survey. Support Care Cancer 2022 Mar;30(3):2163-2171. [doi: 10.1007/s00520-021-06635-6]
5. Pristavec T. The burden and benefits of caregiving: a latent class analysis. Gerontologist 2019 Nov 16;59(6):1078-1091. [doi: 10.1093/geront/gny022] [Medline: 29659788]
6. Pinquart M, Sörensen S. Correlates of physical health of informal caregivers: a meta-analysis. J Gerontol B Psychol Sci Soc Sci 2007 Mar;62(2):126-137. [doi: 10.1093/geronb/62.2.p126] [Medline: 17379673]
7. Petry SE, Lara L, Boucher NA. Older caregivers: who they are and how to support them. J Aging Soc Policy 2024 Jul 3;36(4):589-602. [doi: 10.1080/08959420.2022.2051683] [Medline: 35290168]
8. Shaffer KM, Chow PI, Cohn WF, Ingersoll KS, Ritterband LM. Informal caregivers' use of internet-based health resources: an analysis of the health information national trends survey. JMIR Aging 2018 Dec 18;1(2):e11051. [doi: 10.2196/11051] [Medline: 31518244]
9. Sitges-Maciá E, Bonete-López B, Sánchez-Cabaco A, Oltra-Cucarella J. Effects of e-Health training and social support interventions for informal caregivers of people with dementia—a narrative review. Int J Environ Res Public Health 2021 Jul 21;18(15):7728. [doi: 10.3390/ijerph18157728]

XSL•FO
RenderX

10. Denham AMJ, Wynne O, Baker AL, et al. An online survey of informal caregivers' unmet needs and associated factors. PLoS ONE 2020;15(12):e0243502. [doi: 10.1371/journal.pone.0243502] [Medline: 33301483]
11. Kim H, Lee S, Cheon J, Hong S, Chang M. A comparative study to identify factors of caregiver burden between baby boomers and post baby boomers: a secondary analysis of a US online caregiver survey. BMC Public Health 2018 May 2;18(1):579. [doi: 10.1186/s12889-018-5488-4] [Medline: 29720119]
12. Anderson JG, Hundt E, Dean M, Keim-Malpass J, Lopez RP. "The church of online support": examining the use of blogs among family caregivers of persons with dementia. J Fam Nurs 2017 Feb;23(1):34-54. [doi: 10.1177/1074840716681289] [Medline: 27920340]
13. Blusi M, Kristiansen L, Jong M. Exploring the influence of internet-based caregiver support on experiences of isolation for older spouse caregivers in rural areas: a qualitative interview study. Int J Older People Nurs 2015 Sep;10(3):211-220. [doi: 10.1111/opn.12074] [Medline: 25425070]
14. Bateman DR, Brady E, Wilkerson D, Yi EH, Karanam Y, Callahan CM. Comparing crowdsourcing and friendsourcing: a social media-based feasibility study to support Alzheimer disease caregivers. JMIR Res Protoc 2017 Apr 10;6(4):e56. [doi: 10.2196/resprot.6904] [Medline: 28396304]
15. Diefenbeck CA, Klemm PR, Hayes ER. "Anonymous meltdown": content themes emerging in a nonfacilitated, peer-only, unstructured, asynchronous online support group for family caregivers. Comput Inform Nurs 2017 Dec;35(12):630-638. [doi: 10.1097/CIN.0000000000000376] [Medline: 28742535]
16. McKechnie V, Barker C, Stott J. The effectiveness of an internet support forum for carers of people with dementia: a pre-post cohort study. J Med Internet Res 2014 Feb 28;16(2):e68. [doi: 10.2196/jmir.3166] [Medline: 24583789]
17. Tanis M, Das E, Fortgens-Sillmann M. Finding care for the caregiver? Active participation in online health forums attenuates the negative effect of caregiver strain on wellbeing. Communications 2011;36(1). [doi: 10.1515/comm.2011.003]
18. Horrell B, Stephens C, Breheny M. Online research with informal caregivers: opportunities and challenges. Qual Res Psychol 2015 Jul 3;12(3):258-271. [doi: 10.1080/14780887.2015.1040318]
19. Lindroos A, Douglas-Smith N. Exploring informal caregivers' well-being during COVID-19 through online discussion forums. Scand J Caring Sci 2024 Mar;38(1):104-113. [doi: 10.1111/scs.13199] [Medline: 37522268]
20. Menke M, Wagner AJM, Kinnebrock S. Communicative care in online forums: how burdened informal caregivers seek mediated social support. Int J Commun 2020;14:21 [FREE Full text]
21. Andréasson F, Andreasson J, Hanson E. Developing a carer identity and negotiating everyday life through social networking sites: an explorative study on identity constructions in an online Swedish carer community. Ageing Soc 2018 Nov;38(11):2304-2324. [doi: 10.1017/S0144686X17000551]
22. Donnellan WJ, Warren JG. Emotional word use in informal carers of people living with dementia: linguistic analysis of online discussion forums. JMIR Aging 2022 Jun 17;5(2):e32603. [doi: 10.2196/32603] [Medline: 35713942]
23. Du Y, Dennis B, Liu J, et al. A conceptual model to improve care for individuals with Alzheimer's disease and related dementias and their caregivers: qualitative findings in an online caregiver forum. J Alzheimers Dis 2021;81(4):1673-1684. [doi: 10.3233/JAD-210167] [Medline: 33967054]
24. Scharett E, Madathil KC, Lopes S, et al. An investigation of the information sought by caregivers of Alzheimer's patients on online peer support groups. Cyberpsychol Behav Soc Netw 2017 Oct;20(10):640-657. [doi: 10.1089/cyber.2017.0274]
25. Kruk B. 'I can't bear the thought that he might not recognise me': personal narratives as a site of identity work in the online Alzheimer's support group. Commun Med 2015;12(2-3):273-286. [doi: 10.1558/cam.18453] [Medline: 29048867]
26. Ferrell EL, Russin SE, Hardy RM. Informal caregiving experiences in posttraumatic stress disorder: a content analysis of an online community. J Community Psychol 2019 May;47(4):757-771. [doi: 10.1002/jcop.22151] [Medline: 30592051]
27. Lobo EH, Johnson T, Frølich A, et al. Utilization of social media communities for caregiver information support in stroke recovery: an analysis of content and interactions. PLoS One 2022;17(1):e0262919. [doi: 10.1371/journal.pone.0262919] [Medline: 35081150]
28. De Simoni A, Shanks A, Balasooriya-Smeekens C, Mant J. Stroke survivors and their families receive information and support on an individual basis from an online forum: descriptive analysis of a population of 2348 patients and qualitative study of a sample of participants. BMJ Open 2016 Apr 6;6(4):e010501. [doi: 10.1136/bmjopen-2015-010501] [Medline: 27053271]
29. Washington KT, Oliver DP, Benson JJ, et al. Factors influencing engagement in an online support group for family caregivers of individuals with advanced cancer. J Psychosoc Oncol 2020;38(3):235-250. [doi: 10.1080/07347332.2019.1680592] [Medline: 31690247]
30. Benson JJ, Oliver DP, Washington KT, et al. Online social support groups for informal caregivers of hospice patients with cancer. Eur J Oncol Nurs 2020 Feb;44:101698. [doi: 10.1016/j.ejon.2019.101698] [Medline: 31816508]
31. Ihrig A, Renner T, Muck T, et al. Online support groups offer low-threshold backing for family and friends of patients with prostate cancer. Eur J Cancer Care (Engl) 2019 Mar;28(2):e12982. [doi: 10.1111/ecc.12982] [Medline: 30569592]
32. Shoults CC, Rutherford MW, Kemp AS, et al. Analysis of caregiver burden expressed in social media discussions. Int J Environ Res Public Health 2023 Jan 20;20(3):1933. [doi: 10.3390/ijerph20031933] [Medline: 36767297]
33. Aging Care Caregiver Forum. AgingCare. URL: https://www.agingcare.com/caregiver-forum [accessed 2025-1-9]

XSL•FO
RenderX

34. Caregiver Action Network Community. Facebook. URL: https://www.facebook.com/groups/201714486219468 [accessed 2024-07-27]
35. Dementia Caregivers Support Group. Facebook. URL: https://www.facebook.com/groups/672984902717938/ [accessed 2024-07-27]
36. Empowering Caregivers. Facebook. URL: https://www.facebook.com/groups/empoweringcaregivers/ [accessed 2024-07-27]
37. Caregivers: discussions. Mayo Clinic Connect. URL: https://connect.mayoclinic.org/group/caregivers/ [accessed 2024-07-27]
38. Shaw EK. The use of online discussion forums and communities for health research. Fam Pract 2020 Sep 5;37(4):574-577. [doi: 10.1093/fampra/cmaa008] [Medline: 32050014]
39. Lim WM. What is qualitative research? An overview and guidelines. Aust Mark J 2024;Jul 25:14413582241264619. [doi: 10.1177/14413582241264619]
40. Clarke V, Braun V. Using thematic analysis in counselling and psychotherapy research: a critical reflection. Couns Psychother Res 2018 Jun;18(2):107-110. [doi: 10.1002/capr.12165]
41. Schulz R, Eden J, Committee on Family Caregiving for Older Adults, Board on Health Care Services, Health and Medicine Division, National Academies of Sciences, Engineering, and Medicine. Families Caring for an Aging America: National Academies Press (US); 2016. [doi: 10.17226/23606]
42. Baumbach A, Hughes MC, Liu Y. Challenges and coping strategies in transitioning from caregiving to widowhood: a systematic review. Res Aging 2024;46(9-10):535-547. [doi: 10.1177/01640275241254396] [Medline: 38742924]
43. Hansen T, Slagsvold B. The psychological effects of providing personal care to a partner: a multidimensional perspective. Health Psychol Res 2013 Apr 18;1(2):e25. [doi: 10.4081/hpr.2013.e25] [Medline: 26973910]
44. Huo M, Kim K. Lasting impact of relationships on caregiving difficulties, burden, and rewards. J Soc Pers Relat 2023 Jan;40(1):288-310. [doi: 10.1177/02654075221118095]
45. Kim B, Wister A, O'dea E, Mitchell BA, Li L, Kadowaki L. Roles and experiences of informal caregivers of older adults in community and healthcare system navigation: a scoping review. BMJ Open 2023 Dec 9;13(12):e077641. [doi: 10.1136/bmjopen-2023-077641] [Medline: 38070939]
46. Riffin C, Van Ness PH, Wolff JL, Fried T. Multifactorial examination of caregiver burden in a national sample of family and unpaid caregivers. J Am Geriatr Soc 2019 Feb;67(2):277-283. [doi: 10.1111/jgs.15664] [Medline: 30452088]
47. Sugihara Y, Sugisawa H. Influence of medical care tasks on subjective burden and gain among older adults' family caregivers: structural equation modeling for testing the role of formal and informal support. BMC Geriatr 2023 Oct 6;23(1):628. [doi: 10.1186/s12877-023-04348-5] [Medline: 37803249]
48. Hughes MC, Afrin S, Hamlish T. Effectiveness of skill-building interventions for informal caregivers of adults with cancer: a systematic review. J Cancer Educ 2023 Apr;38(2):390-397. [doi: 10.1007/s13187-022-02236-x] [Medline: 36372814]
49. Daynes-Kearney R, Gallagher S. Online support groups for family caregivers: scoping review. J Med Internet Res 2023 Dec 13;25:e46858. [doi: 10.2196/46858] [Medline: 38090796]
50. Zhang X, Liu S, Deng Z, Chen X. Knowledge sharing motivations in online health communities: a comparative study of health professionals and normal users. Comput Human Behav 2017 Oct;75:797-810. [doi: 10.1016/j.chb.2017.06.028]
51. White MH, Dorman SM. Online support for caregivers. Analysis of an internet Alzheimer mailgroup. Comput Nurs 2000;18(4):168-176. [Medline: 10939186]
52. Hughes MC, Liu Y, Baumbach A. Impact of COVID-19 on the health and well-being of informal caregivers of people with dementia: a rapid systematic review. Gerontol Geriatr Med 2021;7:23337214211020164. [doi: 10.1177/23337214211020164] [Medline: 34104687]
53. Liu Y, Hughes MC, Wang H. Financial train, health behaviors, and psychological well-being of family caregivers of older adults during the COVID-19 pandemic. PEC Innov 2024 Dec;4:100290. [doi: 10.1016/j.pecinn.2024.100290] [Medline: 38799257]
54. Martin W, Collett G, Bell C, Prescott A. Ageing, the digital and everyday life during and since the Covid-19 pandemic. Front Psychol 2023;14:1168340. [doi: 10.3389/fpsyg.2023.1168340] [Medline: 37829072]
55. Reinhard SC, Caldera S, Houser A, Choula R. Valuing the invaluable: 2023 update. ARCH National Respite Network and Resource Center.: AARP; 2023. URL: https://archrespite.org/wp-content/uploads/2023/03/valuing-the-invaluable-2023-update.doi_.10.26419-2Fppi.00082.006.pdf [accessed 2025-01-09]
56. Fact sheet: how the build back better framework will support the sandwich generation. The White House. 2021. URL: https://www.whitehouse.gov/briefing-room/statements-releases/2021/09/21/fact-sheet-how-the-build-back-better-framework-will-support-the-sandwich-generation/ [accessed 2024-12-27]

XSL•FO
RenderX

Edited by A Mavragani; submitted 28.07.24; peer-reviewed by B Davis, DB Loo, J Johnson; revised version received 06.12.24; accepted 08.12.24; published 17.01.25.

Please cite as:
Foster ML, Egwuonwu C, Vernon E, Alarifi M, Hughes MC
Informal Caregivers Connecting on the Web: Content Analysis of Posts on Discussion Forums
JMIR Form Res 2025;9:e64757
URL: https://formative.jmir.org/2025/1/e64757
doi:10.2196/64757

© Michelle L Foster, Chinenye Egwuonwu, Erin Vernon, Mohammad Alarifi, M Courtney Hughes. Originally published in JMIR Formative Research (https://formative.jmir.org), 17.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Planned Behavior in the United Kingdom and Ireland Online Medicine Purchasing Context: Mixed Methods Survey Study

Bernard D Naughton[1,2], MPharm, PhD

[1]School of Pharmacy and Pharmaceutical Sciences, Trinity College Dublin, University of Dublin, Dublin, Ireland

[2]Centre for Pharmaceutical Medicine Research, School of Cancer and Pharmaceutical Science, Kings College London, London, United Kingdom

Corresponding Author:
Bernard D Naughton, MPharm, PhD
School of Pharmacy and Pharmaceutical Sciences
Trinity College Dublin, University of Dublin
College Green, Dublin 2
Dublin, D02 PN40
Ireland
Phone: 353 18962824
Email: bernard.naughton@tcd.ie

Abstract

Background: Online medicine purchasing is a growing health care opportunity. However, there is a scarcity of available evidence through a behavioral lens, which addresses why consumers buy medicines online. Governments try to influence online medicine purchasing behavior using health campaigns. However, there are little data regarding specific online medicine purchasing behaviors to support these campaigns.

Objective: The theory of planned behavior explains that perceived behavioral control (PBC), attitudes, and norms contribute to intentions, leading to behaviors. This study challenges these assumptions, by testing them in an online medicine purchasing context. We asked: What is the role of attitudes, norms, and PBC in an online medicine purchasing context.

Methods: An anonymous online snowball convenience sample survey, including open and closed questions concerning online medicine purchasing, was implemented. The data were thematically analyzed until data saturation. The emerging themes were applied to each individual response, as part of a case-by-case narrative analysis.

Results: Of the 190 consumers from the United Kingdom and Ireland who consented to participate in the study, 46 participants had purchased medicines online, 9 of which were illegal sales. Of the 113 participants who demonstrated an intention to purchase, 42 (37.2%) completed a purchase. There were many cases in which participants demonstrated an intention to buy medicines online, but this intention did not translate to a purchasing behavior (71/190, 37.4%). Reasons for consumers progressing from intention to behavior are suggested to be impacted by PBC and attitudes. Qualitative data identified access to medicine as a factor encouraging online medicine purchasing behaviors and a facilitator of behavior transition. Despite understanding the importance of why some medicines required a prescription, which is described as an example of legal and health norms, and despite suspicion and concern categorized as negative attitudes in this paper, some participants were still buying products illegally online. Risk reduction strategies were performed by 17 participants (17/190, 9%). These strategies facilitated a transition from intention to behavior.

Conclusions: The study results indicate that a consumer's intention to buy does not automatically translate to a purchasing behavior online; instead, a transition phase exists. Second, consumers followed different pathways to purchase and used risk reduction practices while transitioning from an intention to a behavior. Finally, owing to the covert nature of online medicine purchasing, norms do not appear to be as influential as PBC and attitudes in an online medicine purchasing setting. Understanding how a consumer transitions from an intention to a behavior could be useful for researchers, health care professionals, and policymakers involved in public health campaigns. We encourage future research to focus on different consumer behavior pathways or ideal types, rather than taking a blanket approach to public health campaigns.

(JMIR Form Res 2025;9:e55391) doi:10.2196/55391

KEYWORDS

planned behavior; consumer behavior; perceived behavioral control; attitudes; online purchasing; medicine

XSL•FO
RenderX

Introduction

Background

Authenticity is an important consideration when buying any product online, especially medicines, as inappropriate access and poor-quality medicines can have catastrophic implications for human health [1], with some sales causing death [2,3] including several coroner report cases [4]. Furthermore, online medicine purchasing has increased in recent times due to increased convenience, access, and an improved consumer experience [5,6]. From 1999 to 2003 in the United States, there was an increase in online sales of prescription drugs from $160 million to $3.2 billion [7], and the act of purchasing medicine online is likely to increase in line with internet access. In the United Kingdom in 2020, 96% of households had internet access, up from 2017, when 90% of UK households had internet access, and 2006, when only 57% of households had internet access [8]. These increases in internet access have also influenced the rate at which goods and services are bought online, with 87% of adults having bought goods or services online in 2020, up from 53% in 2008 [8].

Despite the dangers and increasing rates of online medicine purchases, there has been a scarcity of research to examine online medicine purchasing behavior. This paper examined online medicine purchasing through a behavioral model. The theory of planned behavior literature describes that past behavior, attitudes, perceived behavioral control (PBC), and norms affect a consumer's intention and that this intention leads to a behavior. We challenge this assumption in the online medicine purchasing context and propose that there are unexplored contributing factors that affect a consumer's likelihood of moving from an intention to a behavior. We did not aim to consider specific process variables such as the role of internal and external drivers on the translation from intention to behavior. Instead, we argued the presence of a transition phase that exists between an intention and a behavior and the core factors that specifically influence this transition. Although we see the merit in more granular research, it was not within the scope of this paper to pursue this.

Theory of Planned Behavior

Overview

This article examined human behavior in an online medicine purchasing context. Through a theory of planned behavior lens, we asked: What does the relationship between intentions and behaviors look like when buying medicine online? The theory of planned behavior is a social cognitive model in psychology that concerns intentional behavior; it was constructed by Ajzen in 1985 [9] and 1991 [10] and was originally known as the theory of reasoned action [11-15]. The theory identifies 3 determinants of intentions: (1) the attitude toward the behavior, which can be either positive or negative; (2) subjective norms, which are a person's perceptions of the social expectations to adopt a behavior; and (3) PBC, which is a person's perception of how easy or hard it might be to perform a behavior [16]. Since the book published by Fishbein and Ajzen in 2010 [13], this theory has also been known as the reasoned action approach. Although this theory was constructed over 50 years ago, it is still used in contemporary times, with Hagger et al [17] applying the reasoned action approach to health behavior in 2018 and contributing "past behavior" to the framework [18].

Theory of Planned Behavior: Strengths and Limitations

The theory of planned behavior is the most influential and highly cited model for the research and prediction of human behavior and is therefore a suitable theory to apply in this study [19]. However, like most behavioral theories, it has its limitations and criticisms in the literature when used in qualitative research [19]. One of the main limitations, which we addressed in this study, is the intention-behavior gap, as several studies consider intentions to be a poor indicator of behaviors [18,20] even over short time frames. The theory also does not take into consideration emotions or irrationality. The absence of emotion is important, especially when dealing with the purchase of health products as one's health or the health of a loved one can have significant emotional impacts [21]. The lack of a focus on irrational behavior is an area of the model on which this study touched [19].

Online Medicine Purchasing Context

The at-distance medicine sales context is different to in-person sales for a variety of reasons. In an at-distance sales environment, attitudes, which are affected by suspicion or distrust, are augmented due to diminished seller-consumer accountability and a lack of physical interaction between a consumer and a health care professional, a lack of trust between the consumer and seller, and a fear of online fraud, often due to reports of rogue online traders [22-25]. Furthermore, norms are silenced at the point of sale, as the purchase usually occurs in isolation and is removed from direct and indirect social influence due to consumers commonly purchasing and administering the medicine covertly. Finally, PBC is augmented in an online context, due to the ability to buy products that may be prohibited or restricted, resulting in some patients buying illegal medicines online that they would not otherwise have access to. This context causes an augmentation of PCB when compared to in-person purchasing. We wished to understand if the transition from intention to behavior is as straightforward as explained in the literature or if there is a further unexplored nuance when applied to the online medicine purchasing context [26].

The constructs of planned behavior are relevant to the purchase of counterfeit luxury goods [27-29]. There are also many studies in the medical sciences literature that investigate issues concerning online pharmacies [5,22,30-38]. Based upon our literature review, there is a scarcity of published research concerning online medicine purchasing behavior research examined through a theoretical lens. This lack of research may be due to difficulties accessing patients who are willing to admit to buying medicines online for fear their activity may be illegal.

As part of a preprint report of a qualitative study conducted in the United Kingdom in 2023, Almomani and colleagues [39] interviewed 20 participants who had bought prescription-only medicines (POM) from the internet after 2019. They indeed acknowledged the reluctancy of consumers to disclose such activity because of its illegal connotations. The same research

XSL•FO
RenderX

group also explored consumers' beliefs by reviewing news media articles [40]. During analysis, they applied the theory of planned behavior model and observed behavioral beliefs (risks and benefits of purchasing), social beliefs (social influencing factors [eg, health care professionals and influencers]), and control beliefs (facilitators of purchasing [eg, accessibility, medicine shortages]) as influencing the likelihood of people buying POMs online. Hall and Antonopoulos [41] also conducted some behavioral research in this context. They used a virtual ethnography approach that provided qualitative evidence of the link between falsified medicine and online sales through the advancement of the field of demand dynamics and provided qualitative data that support their propositions and the findings of this paper.

This work is insightful but could be built upon through a more in-depth use of theoretical lenses. The rise of internet purchasing, the effect of poor-quality medicines on health [2], and the risks associated with buying potent medicines on the web [34,37] make this research and context a challenging area of unique and increasing importance in online consumer behavior research, where qualitative data are difficult to attain. The aim of this study was to gain a broad understanding of online medicine purchasing behaviors through the lens of planned behavior. In doing so, we questioned the role of attitudes, norms, and PBC in an online medicine purchasing context and how they affect the intention-behavior gap.

Methods

Study Design

This study used a mixed methods survey study design that posed open and closed questions to 190 members of the public. Due to the illegal nature of some online medicine purchases and the challenges of obtaining data from this population, an online convenience snowball sample survey was implemented. This approach was chosen to encourage anonymous participation. This included broad open-ended and closed-ended questions regarding the purchase of medicines online using Google Forms (Multimedia Appendix 1). The theory was developed subsequent to data collection [42]. This mixed methods survey was designed according to the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist [43].

Data Analysis

The study involved a qualitative relational analysis that used thematic evaluation and inductive reasoning to identify themes to help better understand the relationship between intentions and behaviors in the online medicine purchasing context. Quantitative analysis used percentages to depict descriptive quantitative data and was generated from closed answer questions. This quantitative data supported the qualitative inductive reasoning, which proposed a novel transition phase between intentions and behaviors and the factors affecting this transition. Based on the emerging themes, a case-by-case narrative analysis was performed to further understand how these themes affected individual consumer purchasing journeys. This involved analyzing each case to understand the role that each theme played in their transition from intention to behavior. Percentages were rounded to 1 decimal place.

Sampling Methodology

The recruitment approach was intended for scale and breadth using an established methodology [44,45] involving a convenience snowballing sampling method. Convenience sampling has its advantages: It is a useful tool for exploratory research in areas where there is little or no existing literature or where resources are limited. Convenience sampling can also generate large amounts of data in a short space of time and can identify hard-to-reach populations, such as willing participants who have purchased medicines online [46]. The survey contained closed questions to record the cohort demographics and open questions concerning the purchase of online medicines to facilitate qualitative data analysis. The data were collected in spring 2018 prior to the COVID-19 pandemic, which is a study limitation and should be considered within the context of that era. The inclusion criteria in this study were English-speaking internet users as per the methodology in [45].

Study Recruitment

This was an open survey disseminated via social media. Therefore, this study was specifically representative of social media users and may not be representative of the wider public. The survey took approximately 10 minutes to complete, and the data were stored in an online database. No personally identifiable information was gathered. An invite to this survey was posted internally by researchers' academic affiliations and the lead researcher's professional LinkedIn and X (previously known as Twitter) and personal Facebook accounts. Reposting was encouraged to promote a snowballing effect. The survey contained a review step that allowed participants to go back to amend their responses. Cookies and IP checks were not used in this study. Only completed surveys were analyzed. The survey was posted and reposted through social media outlets for 4 weeks by the lead researcher, his contacts, and their connections. Payments or incentives were not provided to participants.

Study Pilot and Consent

An anonymous, online, mixed methods survey, with a randomized response recording, was developed by the lead researcher and piloted with members of the research team and professional colleagues of the lead researcher to improve the survey questions. This included a review to ensure the online survey operated as intended, the questions were clear, and the survey was free of typos or grammatical errors (Multimedia Appendix 1). As part of the consenting process, the participants read study information and ticked a box consenting to being involved in the research. (Multimedia Appendix 1).

Product and Sale Type

A free-text option in the survey allowed participants to identify the medicine or medicine type that they purchased online. These medicines were separated into the 3 most common UK legal classes: POM, pharmacy-only (P) medicines, and general sales list medicines. Some medicine classes differ based on the strength, quantity, formulation, and country in which they are purchased. Therefore, the medicines mentioned in this study were listed according to the most frequent class for each product in the United Kingdom. The medicines type, whether clinical questions were asked at the time of purchase, and whether a

XSL•FO
RenderX

prescription was required were recorded to understand if the medicine was purchased legally or illegally. For the purposes of this study, an illegal sale was classified as one in which consumers were not asked for a prescription or asked medical questions when buying products that were considered as POM within the purchasing country.

A P medicine is one that does not require a prescription but can only be sold in a pharmacy. A POM medicine is one that requires a prescription and can only be supplied in a pharmacy. In cases where medicine class was uncertain (such as when the drug was listed as "Hair loss," which could be a P medicine (minoxidil) or POM medicine (finasteride), the drug was classified as P.

Ethical Considerations and Approval

This research was classified according to the National Institute for Health and Care Research guidelines. It was reviewed and received ethical approval from Keele University (reference number ERP2369). Participants provided informed consent to participate. Study data were anonymous. Participants did not receive compensation for participation.

Results

This study involved a survey that generated qualitative and quantitative results that informed the contributions of this paper. The key themes were subjective norms, attitudes, and PBC.

Demographics

This study attracted 227 respondents, of which 190 were living in the United Kingdom or Ireland and consented to the study (Table 1). The largest group (78/190, 41.1%) of participants was aged between 30 years and 39 years. There were more female respondents than male respondents (Table 1). The majority (135/190, 71.1%) of respondents lived in the United Kingdom, and 28.9% (55/190) lived in Ireland (Table 1). The study also included respondents with a variety of employment statuses and education levels. The majority (170/190, 89.5%) of respondents were employed full time or part time. Full-time students (13/190, 6.8%) and the unemployed (7/190, 3.7%) were also represented (Table 1). About one-third of participants (55/190, 28.9%) had health insurance, 64.7% (123/190) of participants lived in a country that provided free health care, and 6.3% (12/190) of participants did not have health insurance and did not live in a country that provided free health care.

XSL•FO
RenderX

Table 1. Demographic characteristics of the study participants (N=190).

Characteristics	Results, n (%)
Gender	
Male	67 (35.3)
Female	123 (64.7)
Age (years)	
19-29	64 (33.7)
30-39	78 (41.1)
40-49	29 (15.3)
50-59	14 (7.4)
60-69	4 (2.1)
70-79	1 (0.5)
Country	
United Kingdom	135 (71.1)
Ireland	55 (28.9)
Employment status	
Full time	144 (75.8)
Part time	26 (13.7)
Full-time student	13 (6.8)
Unemployed	7 (3.7)
Education level	
Primary school	1 (0.5)
Secondary school	24 (12.6)
Undergraduate	85 (44.7)
Postgraduate	66 (34.7)
PhD	14 (7.4)
Health insurance status	
Free national health care	123 (64.7)
Health insurance	55 (28.9)
Pay for health care when required	12 (6.3)

Medicines Purchased

Of the 190 study participants, 46 had purchased medicines online. These medicines included a wide variety such as antibiotics, antimalarials, controlled drugs such as diazepam, and cosmetic medicines such as minoxidil for hair loss and Orlistat for weight loss (Multimedia Appendix 2). As all study participants were from the United Kingdom and Ireland, these medicines were categorized according to UK medicine classifications.

Health and Legal Norms

The participants in this study were a largely well-educated population from the United Kingdom and Ireland (Table 1) who understood why prescriptions were needed. They could identify doctors, pharmacists, dentists, opticians, and nurses or a mixture of all 5 as the main professionals legally permitted to supply a prescription or drugs. The majority (189/190, 99.5%) of respondents knew that all medicines did not require a prescription and were aware of reasons why some medicines required a prescription and others did not. Respondents, which are numbered (#), gave many legitimate reasons why prescriptions were needed, for example:

> *Monitor drug usage,...ensure correct medications are given.* [#159]
>
> *To legally obtain drugs.* [#152]
>
> *To prevent risk to patients.* [#141]
>
> *To allow you to legally obtain medicines.* [#79]

Although the norm within the cohort was an understanding that prescriptions were required for the safe supply of medicine, of the 46 respondents who bought medicines online, 18 of those respondents (39%) purchased POMs (according to the UK classification). Of those 18 participants, 17 either had health insurance or free health care provided by the state. Of the 18

XSL•FO
RenderX

participants who bought POM online, 78% (17/18) were not asked for a prescription when purchasing these POM online, and 50% (9/18) were not asked for a prescription or asked medical questions when buying POM, making these 9 sales likely to be illegal. In summary, an interesting contradiction existed. Despite an understanding of the safety reasons for why prescriptions were needed, some participants still completed an illegal purchase of a medicine online without a prescription.

Attitudes: Cost, Convenience, and Access

Participants were asked whether they would consider buying medicines online, and 59% (112/190) said that they would. In another question, participants were asked "On a scale of 1 to 10 (with 1 being unlikely and 10 being highly likely), how likely are you to buy a medicine online?" The total cohort gave an average mean response of 4.2, and those who had previously bought medicines online described a high likelihood, a mean average of 7.6, of buying medicines in the future.

There were several reasons in this study why consumers purchased their medicines online. Participants selected from a predetermined list of reasons for positive or negative attitudes that motivated consumers to buy medicine online, or they could free-type their response. The 2 reasons that contributed to positive attitudes toward online medicine purchasing behavior were convenience (18/46, 39%) and cost (22/46, 48%). A free-text space in this question resulted in 17 other responses. These responses largely focused on medicine access issues or due to their general practitioner (GP) not being willing to supply the medicine:

> *The specific medication that is most useful for my condition is not available in my country.* [#162]
>
> *Dr will only prescribe two weeks of treatment, despite chronic condition.* [#66]

The point relating to a GP's unwillingness to supply (diazepam) was mentioned in this study a second time, when the same participant stated:

> *I was desperate as local GP will only prescribe very short-term treatment in low doses.* [#66]

Furthermore, despite buying and taking medicines bought online, 26% (12/46) of participants had negative attitudes toward online medicine purchasing. They explained that they were concerned that the medicine they purchased could be fake, yet these participants progressed to purchase.

PBC: Risk Reduction Measures

The participants demonstrated that they had the autonomy and capacity to mitigate the risks associated with online medicine purchasing. In Table 2, we see a variety of practices used by our study participants to do so.

Table 2. Risk reduction practices used by this study population to support the online purchase of medicines.

Risk reduction practice	Example quotes (participant number)
No checks	• "None, just had bought from it (online site) many times before but not medicines" (#34) • "The product is manufactured from the same supplier that a prescription could have given me. They are cheaper, and a prescription is not required in this case. So, the safety of the website is not a factor to me." (#137) • "I assumed it was safe." (#59) • "None, because the product works (this is bad I know)" (#24) • "Word of mouth reviews" (#4)
Recommendations and reviews	• "By looking for recommendations on online forums" (#74) • "Asked friends where they purchased from" (#3)
Using a well-known or reputable website or company	• "It was the website of a pharmacy I used to shop when I lived there." (#140) • "Bought through an online pharmacy that is linked to a high street pharmacy. They asked a variety of health-related questions." (#94) • "Bought from Amazon, reliable source" (#183)
Other risk-mitigating methods	• "Google search or company registration number check" (#158) • "I conducted extensive research to ensure that the website was genuine, authentic, regulated and safe." (#162) • "Research the company I was buying from. Check their address on Google earth, recognise the name of the company, ensure site was safe (Https) and not a redirect" (#144) • "Checked their registration from link to gov (government) website" (#37) • "Register with regulator GMP (Good Manufacturing Practice) approved" (#175)

Practices to support a move from an intention to a purchasing behavior are categorized in this study as a form of PBC. From the cohort of 46 participants who purchased medicines online, there were a variety of methods used by respondents to support the purchasing behavior. These risk reduction practices are categorized as themes, represented as quotes, and summarized in Table 2. These practices include reading reviews and using a well-known or trusted website among other methods.

Case-by-Case Narrative Analysis

As part of our case-by-case narrative analysis, which focused on the transition between intentions and behavior, we identified behavioral pathways or, as we call them, behavioral ideal types. In Table 3, we see that all participants in our study could be categorized into 6 patterns of behaviors or ideal types, which all involved an intention, an attitude (positive or negative), the presence or absence of risk reduction practices, and the presence

XSL•FO
RenderX

or absence of the completed behavior. Although norms play an important role in a consumer's intention to buy medicine online, they do not appear, in our data, to influence the transition from an intention to a purchasing behavior. This is possibly due to a lack of social influence at the point of purchase, as online medicine purchasing does not require social interaction with a pharmacist or other customers.

Table 3. The 6 ideal types identified in this study that describe the transition between intentions and behavior, mediated by attitudes and perceived behavioral control based on behavioral pathways revealed as part of the case-by-case narrative analysis (N=190).

Consumer ideal type	Intention	Attitude	Perceived behavioral control	Behavior	Participants, n (%)
1	Present	Unassessed	Unassessed	Absent	71 (37.4)
2	Present	Absence of a negative attitude	No risk reduction practices	Present	20 (10.5)
3	Present	Negative attitude	No risk reduction practices	Present	5 (2.6)
4	Present	Negative attitude	Risk reduction practices	Present	4 (2.1)
5	Present	Absence of a negative attitude	Risk reduction practices	Present	13 (6.8)
6	Absent	Unassessed	Unassessed	Absent	77 (40.5)

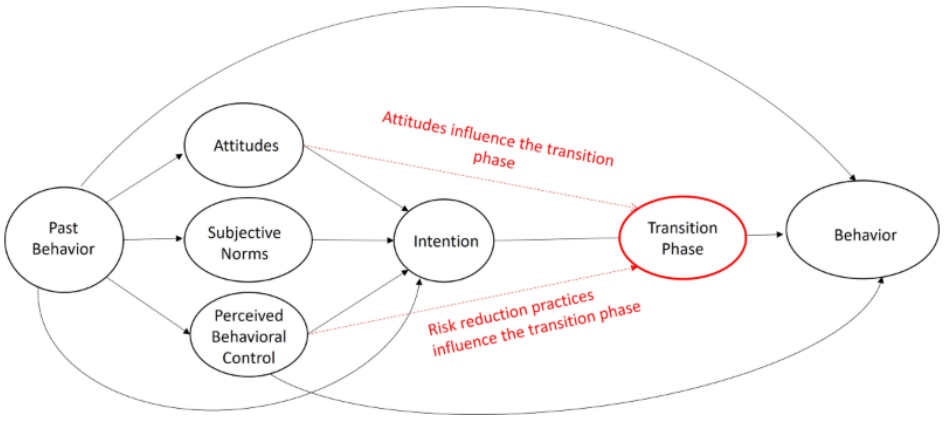
The case-by-case narrative analysis involved a review of each complete individual survey response to identify purchasing pathways from intentions to a purchasing behavior. On some consumers' pathways from intention to behavior, they encountered a negative attitude postintention, and in some ideal types, they used practices to reduce risk. Consumers commonly had either no intention to purchase and did not (77/190, 40.5%; ideal type 6) or had intentions to purchase and did not (71/190, 37.4%; ideal type 1). These individuals were not asked questions regarding attitudes or PBC as they had no experience with online medicine purchasing. There was also evidence that participants with intentions to purchase had no negative attitudes toward the purchasing behavior; therefore, they did not need to implement a risk reduction practice and completed the purchase (20/190, 10.5%; ideal type 2). There were also participants (ideal type 3), who had an intention to purchase had a negative attitude toward the behavior, and, without reducing the risk, completed the purchase (5/190, 2.6%). There were also 4 participants (4/190, 2.1%) who had an intention to purchase, experienced a negative attitude, and used a risk reduction practice before completing the purchasing behavior (ideal type 4). Ideal type 5 (13/190, 6.8%) included individuals who had an intention to purchase, appeared to demonstrate no negative attitudes, implemented a risk reduction practice in any case, and completed the purchase. Ideal type 3 is an example of irrational behavior, as these study participants had an intention to purchase, then had negative attitudes toward the behavior, and, without taking any steps to manage that attitude, they still purchased. When asked "What measures, if any, did or do you take, to make sure the website that you are buying from is or was safe?", one participant replied "None because the product works (this is bad I know)' which typifies irrational behavior" (participant #24).

Discussion

Themes

This study aimed to gain a broad understanding of online medicine purchasing behaviors through the theory of planned behavior lens with a focus on how intentions lead to behaviors and the intention-behavior gap. The dominant themes running through this paper concerned (1) health and legal norms; (2) attitudes toward online purchasing (positive: convenience and cost; negative: concern or suspicion); and (3) PBC, which presented as practices to minimize risk. The study also proposes that a transition stage (Figure 1) exists between intentions and behavior, which appears, in our data, to be influenced by attitudes and risk reduction practices but less so by norms.

Figure 1. An adapted theory of planned behavior model with red lines representing this study's proposed contributions, showing attitudes and perceived behavioral control as mediators of the transition from intention to behavior.



XSL•FO
RenderX

Theoretical Contribution

This paper identified qualitative and quantitative patterns within our data that support the use of the theory of planned behavior in this context [9,10] and described the influence of attitude, PBC, and norms on intentions and behavior. However, within this context, we suggest that the relationship between intention and behavior contains more nuance than is covered by Ajzen [9,10] and others [13,17]. We therefore adapted the model by Hagger et al [17] (Figure 1) to include our propositions, which are highlighted in red. Other studies support our proposal that intentions do not automatically translate to behaviors [26]. However, this paper reasons that a transition stage exists, where positive attitudes toward purchasing, such as convenience and cost, promote a transition from intention to behavior. Negative attitudes, which can emerge after the initial intention due to concern or suspicion, can inhibit a transition from intention to behavior [47], whereas PBC in the form of using risk reduction practices, such as reading reviews to mitigate negative attitudes, plays an important role in an actor's successful psychological transition from an intention to a purchasing behavior. These propositions are supported in part by the work of Sniehotta [26] and aim to help address the intention to behavior gap in the literature (Figure 1).

This proposed transition phase is interesting and suggests that intentions can change before completing a purchase. If this transition stage exists, then there is an opportunity to influence consumers between the intention and behavior stages, which is an opportunity for public health campaigners or health care professionals to encourage or discourage practices that influence the transition phase.

Norms: Health and Legal

There are several subjective norms that may affect online medicine purchasing behaviors. Norms differ in an online context because consumers have access to products that they would not normally have access to, due to availability issues or laws that prevent their purchase without a legal and valid prescription. In our study, the 2 most prominent norms that emerged included legal and health norms. There was no evidence from this study context that social norms negatively affected a participant's transition from intention to behavior. This may be due to the standard medical treatment setting, where consumers are influenced by health care professionals to conform to norms during diagnosis and medicine supply, or by fellow consumers within stores. However, this is not the case online. Remote purchasing of medicine online is a step removed from the influences in a traditional purchasing environment, whereby the patient is without pressure to conform to social norms as they do not see the clinician or pharmacist face to face and are free to request whatever medicine they wish, in whatever quantity they desire. The effects of being refused a request online are unlikely to be as impactful as being refused in person by a health care professional who may know you and your family or have full access to your personal details and records. Therefore, it is understandable that the consumers in this study that transitioned from an intention to a behavior did not raise any negative normative influences throughout the transition.

If the medicine received from an online pharmacy is legitimate, the risks relate to addiction, inadequate medical supervision, and poor condition management. If the medicine is poor quality or falsified, there are additional risks such as side effects, overdose, and death. Buying medicine online is not a problem if the website is legitimate. Although several consumers answered yes to "At any stage were you ever concerned that the product(s) you were buying may have been fake, falsified, substandard or of poor quality?", 7 of these (#3, #4, #24, #66, #74, #168, #182) still progressed to purchasing a product. From this evidence, we propose that the usual societal norms and legal medicine access norms in the online context do not appear to inhibit the transition from intention to behavior.

Attitudes: Cost, Convenience, and Access

We present, in this paper, that attitudes toward purchasing behavior were influenced by cost, convenience, and access, which is also described in other studies [22].

Hall and Antonopoulos [41] described cases of patients purchasing medicines online due to inconvenience or because they were struggling to see a GP, when it was possible to see the GP:

> *After trying to get a doctor's appointment for several weeks with no luck, I gave up and decided to go online to buy the drug I wanted.*

The idea that consumers were taking control of the situation themselves due to cost and convenience was also covered in another study by Almomani et al [22]:

> *There is no need to make an appointment with your GP. Well, it takes a couple of weeks to get an appointment in our surgery.* [2QL]

Medicine access was a qualitative theme that commonly emerged from the survey data, which also correlates with a study by Cicero and Ellis [32] and is argued to impact attitudes. The 2 reasons for buying tramadol online in the empirical study by Cicero and Ellis were (1) difficulty finding a doctor to prescribe the opioid medicine and (2) doctors who would not prescribe enough. Both points are access-related issues that affect PBC, appear to be augmented in the context of online addictive drug purchasing, and increase intentions to purchase. In the case of addictive medicines, access appears to be a key driver of intentions. Another example from this study was related to diazepam, and in that case, the participant who bought the diazepam online also explained that they were motivated to do so because their doctor refused to supply enough of the medicine.

In this case, the participant went on to buy this medicine online from a website that did not require an online consultation or a prescription. This is a case of a patient bypassing the GP to obtain medicines, which demonstrates autonomy. This is a common case that is also reflected in quotes gathered by Hall and Antonopoulos in their 2016 study [41].

The issue of limited access affects the participant's attitude toward the behavior and may be a driver to regain PBC that encourages online purchasing behavior which appears to be common in this study and the wider literature.

XSL•FO
RenderX

PBC and Risk Reduction Practices

We propose that PBC affects the psychological transition from intention to behavior. Consumers can assert their own control in a purchasing context to mitigate against risks that develop postintention and prebehavior, such as a consumer noticing something suspicious about an online pharmacy. Their PBC allows them to check the website URL against a list of approved websites or check other customer reviews to mitigate their negative attitudes toward that purchase [48-51].

Hall and Antonopoulos [41] identified the importance of finding a trusted illegal source.

In order to find a "Reputable" supplier, a consumer commonly relied on the use of risk reduction practices before purchase, which is a form of PBC. These practices included their own research, the company's reputation, and online reviews (Table 2). We reason that postintention risk reduction practices influence consumer transitions from intention to behavior. This is supported by evidence from Filieri et al [48] and Jin et al [49], who discussed the role of product reviews in online decision-making. When consumers encounter convincing information to mitigate identified risks, such as online reviews, they tend to complete the transition from initial intention to behavior [48]. Of course, there will be exceptions to this transition scenario, whereby consumers' intentional and logical behaviors are overtaken by impulsiveness behaviors [52], which is covered by consumer ideal type 3 in Table 3.

Behavioral Pathways: Public Health Policy Implications

A consumer's transition from intention to behavior, depicted as consumer ideal types in Table 3, acts as a lens through which to see consumer purchasing behavior patterns. This study has its limitations; however, these ideal types could be useful tools to help understand how different consumer groups transition from an intention to a behavior. This may assist with targeting individual psychological processes associated with norms, negative attitudes, PBC, and the facilitation of behavioral change through targeted public health campaigns. The current public health campaign policies concerning the purchase of falsified medicine online but do not appear to consider the different attitudes and PBC practices that affect individuals. Instead, public health campaigns either send an educational message about the risks of buying medicines online or try to incite fear into customers to deter them from the online purchase of medicine. Viewing consumers through different ideal types will provide public health campaign marketers with a new structured strategy to help them target specific behavior types. For example, consumers who demonstrate a negative attitude toward the online purchasing behavior but tend to continue from intention to behavior when adopting a risk reduction practice are unlikely to find a public health campaign that incites fear or suspicion as an effective deterrent. In those cases, these individuals would be best served by public health campaigns that describe safe and effective risk reduction practices such as explaining ways to verify the legitimacy of the website or the medicine they are purchasing.

Retail Pharmacy Marketing Implications

For legitimate online pharmacies, it may be useful to create marketing approaches that develop upon the principles described in this paper. These approaches may include the reassurance of consumers with negative attitudes and sign-posting reliable risk reduction practices to help reassure their customers. Reassurance may include regular blogs or posts concerning the risks of buying from illegal pharmacies; risk reduction practices to support patients, which may include searching for privacy or security statements [25]; or methods to verify online pharmacies or medicines they receive from pharmacies. These practices could be monitored using consumer metrics to understand which strategies were most effective at converting intentions to behaviors [53]. This argument may also hold true for marketers in other industries, as they too could run campaigns to raise awareness of illegal websites and ways to verify the legitimacy of an online retailer or the products they sell.

Limitations

This study is based on a relatively small sample of participants from the United Kingdom and Ireland and could benefit from being repeated as part of a large, multicenter, triangulated, mixed methods study that includes interviews, focus groups, and observations across different regions. In addition, the data were collected in 2018; therefore, the findings should be considered in light of that era. This study was also based on social media users, who may not represent a suitable cross-section of society. However, considering the sensitive nature of the subject, it can be difficult to recruit participants to such a study. This study is a first step to identifying the presence of certain factors facilitating or inhibiting a consumer's transition from an intention to a behavior in an online medicine purchasing context.

In conclusion, intentions do not always directly translate to a behavior in an online medicine purchasing context, which is an important consideration for researchers in this area. Instead, we suggest that the transition is mediated by PBC and attitudes. In many cases, those buying medicines online were concerned that their medicine may be a fake but still proceeded to purchase, which is concerning from a public health perspective. In cases where a complete transition from intention to behavior did occur, it was often in the absence of a negative attitude or the presence of a negative attitude coupled with a practice to reduce risk. The arguments in this article support the presence of a transition phase when buying medicine online. It is hoped that this paper's contribution will help public health campaigners and online pharmacy marketers to successfully target specific behavior types and support the use of consumer risk mitigation tools.

Data Availability

Primary study data are available on request.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
The questionnaire applied in this study.
[DOCX File , 19 KB - formative_v9i1e55391_app1.docx]

Multimedia Appendix 2
Medicines purchased by respondents in this study.
[DOCX File , 15 KB - formative_v9i1e55391_app2.docx]

References

1. Cinelli MD, LeBoeuf RA. Keeping it real: How perceived brand authenticity affects product perceptions. J Consum Psychol 2019 Jul 05;30(1):40-59. [doi: 10.1002/jcpy.1123]
2. Philby C. Death of medical student Sarah Houston after taking banned slimming drug Dinitrophenol highlights dangers of buying pills online, warns Government. Independent. 2013 Apr 23. URL: http://www.independent.co.uk/news/uk/home-news/death-of-medical-student-sarah-houston-after-taking-banned-slimming-drug-dinitrophenol-highlights-8584597.html [accessed 2025-02-02]
3. Forster K. Young mother dies after 'taking dodgy diet pills sold on Facebook'. Independent. 2017 Jul 05. URL: http://www.independent.co.uk/news/health/diet-pills-facebook-dead-young-mother-24-years-old-lucero-garza-mexico-monterrey-avitia-cobrax-a7824206.html [accessed 2025-02-02]
4. Aronson JK, Ferner RE, Richards GC. Deaths attributed to the use of medications purchased online. BMJ Evid Based Med 2022 Feb 21;27(1):60-64. [doi: 10.1136/bmjebm-2021-111759] [Medline: 34933931]
5. Fittler A, Ambrus T, Serefko A, Smejkalová L, Kijewska A, Szopa A, et al. Attitudes and behaviors regarding online pharmacies in the aftermath of COVID-19 pandemic: At the tipping point towards the new normal. Front Pharmacol 2022 Dec 23;13:1070473 [FREE Full text] [doi: 10.3389/fphar.2022.1070473] [Medline: 36642991]
6. Pantano E, Priporas C. The effect of mobile retailing on consumers' purchasing experiences: A dynamic perspective. Computers in Human Behavior 2016 Aug;61:548-555. [doi: 10.1016/j.chb.2016.03.071]
7. Jena AB, Goldman DP. Growing Internet use may help explain the rise in prescription drug abuse in the United States. Health Aff (Millwood) 2011 Jun;30(6):1192-1199 [FREE Full text] [doi: 10.1377/hlthaff.2011.0155] [Medline: 21565838]
8. Internet access – households and individuals, Great Britain: 2017. Office for National Statistics. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2017 [accessed 2025-02-02]
9. Ajzen I. From Intentions to Actions: A Theory of Planned Behavior. In: Kuhl J, Beckmann J, editors. Action Control. SSSP Springer Series in Social Psychology. Berlin, Germany: Springer; 1985:11-39.
10. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211 [FREE Full text] [doi: 10.1016/0749-5978(91)90020-T]
11. Ajzen I, Fishbein M. Understanding Attitudes and Predicting Social Behavior. Upper Saddle River, NJ: Prentice Hall; 1980.
12. Bamberg S, Ajzen I, Schmidt P. Choice of travel mode in the theory of planned behavior: the roles of past behavior, habit, and reasoned action. Basic and Applied Social Psychology 2010 Jun 07;25(3):175-187. [doi: 10.1207/s15324834basp2503_01]
13. Fishbein M, Ajzen I. Predicting and Changing Behavior: The Reasoned Action Approach. New York, NY: Psychology Press; 2010.
14. Gold GJ. Review of predicting and changing behavior: the reasoned action approach. The Journal of Social Psychology 2011 Apr 21;151(3):382-385. [doi: 10.1080/00224545.2011.563209]
15. Sheppard BH, Hartwick J, Warshaw PR. The theory of reasoned action: a meta-analysis of past research with recommendations for modifications and future research. Journal of Consumer Research 1988 Dec;15(3):325-343. [doi: 10.1086/209170]
16. Ajzen I, Madden TJ. Prediction of goal-directed behavior: attitudes, intentions, and perceived behavioral control. Journal of Experimental Social Psychology 1986;22(5):453-474. [doi: 10.1016/0022-1031(86)90045-4]
17. Hagger MS, Polet J, Lintunen T. The reasoned action approach applied to health behavior: role of past behavior and tests of some key moderators using meta-analytic structural equation modeling. Soc Sci Med 2018 Sep;213:85-94. [doi: 10.1016/j.socscimed.2018.07.038] [Medline: 30064092]
18. McEachan R, Taylor N, Harrison R, Lawton R, Gardner P, Conner M. Meta-analysis of the reasoned action approach (RAA) to understanding health behaviors. Ann Behav Med 2016 Aug 11;50(4):592-612 [FREE Full text] [doi: 10.1007/s12160-016-9798-4] [Medline: 27169555]
19. Ajzen I. The theory of planned behaviour: reactions and reflections. Psychol Health 2011 Sep;26(9):1113-1127. [doi: 10.1080/08870446.2011.613995] [Medline: 21929476]

XSL•FO
RenderX

20. Kor K, Mullan BA. Sleep hygiene behaviours: an application of the theory of planned behaviour and the investigation of perceived autonomy support, past behaviour and response inhibition. Psychol Health 2011 Sep;26(9):1208-1224. [doi: 10.1080/08870446.2010.551210] [Medline: 21678170]
21. Conner M, Armitage CJ. Extending the theory of planned behavior: a review and avenues for further research. J Applied Social Pyschol 2006 Jul 31;28(15):1429-1464. [doi: 10.1111/j.1559-1816.1998.tb01685.x]
22. Almomani H, Raza A, Patel N, Donyai P. Reasons that lead people to buy prescription medicines on the internet: a systematic review. Front Pharmacol 2023 Aug 31;14:1239507 [FREE Full text] [doi: 10.3389/fphar.2023.1239507] [Medline: 37719862]
23. Benedicktus RL, Brady MK, Darke PR, Voorhees CM. Conveying trustworthiness to online consumers: reactions to consensus, physical store presence, brand familiarity, and generalized suspicion. Journal of Retailing 2010 Dec;86(4):322-335. [doi: 10.1016/j.jretai.2010.04.002]
24. Hoffman DL, Novak TP, Peralta M. Building consumer trust online. Commun. ACM 1999 Apr;42(4):80-85. [doi: 10.1145/299157.299175]
25. Pan Y, Zinkhan GM. Exploring the impact of online privacy disclosures on consumer trust. Journal of Retailing 2006 Jan;82(4):331-338. [doi: 10.1016/j.jretai.2006.08.006]
26. Sniehotta FF. Towards a theory of intentional behaviour change: plans, planning, and self‐regulation. British J Health Psychol 2010 Dec 24;14(2):261-273. [doi: 10.1348/135910708x389042]
27. Gistri G, Romani S, Pace S, Gabrielli V, Grappi S. Consumption practices of counterfeit luxury goods in the Italian context. J Brand Manag 2009 Feb 6;16(5-6):364-374. [doi: 10.1057/bm.2008.44]
28. Phau I, Teah M, Lee A. Targeting buyers of counterfeits of luxury brands: a study on attitudes of Singaporean consumers. J Target Meas Anal Mark 2009 Mar 27;17(1):3-15. [doi: 10.1057/jt.2008.25]
29. Wilcox K, Kim HM, Sen S. Why do consumers buy counterfeit luxury brands? Journal of Marketing Research 2009 Apr 01;46(2):247-259. [doi: 10.1509/jmkr.46.2.247]
30. Bate R, Jin G, Mathur A. In whom we trust: the role of certification agencies in online drug markets. The B. E. Journal of Economic Analysis & Policy 2013;14(1):111-150. [doi: 10.1515/bejeap-2013-0085]
31. Cho S, Park HJ, Lee JH, Do J, Heo S, Jo JH, et al. Determination of anabolic-androgenic steroid adulterants in counterfeit drugs by UHPLC-MS/MS. J Pharm Biomed Anal 2015 Jul;111:138-146. [doi: 10.1016/j.jpba.2015.03.018] [Medline: 25880245]
32. Cicero TJ, Ellis MS. Health outcomes in patients using no-prescription online pharmacies to purchase prescription drugs. J Med Internet Res 2012 Dec 06;14(6):e174 [FREE Full text] [doi: 10.2196/jmir.2236] [Medline: 23220405]
33. Fittler A, Bősze G, Botz L. Evaluating aspects of online medication safety in long-term follow-up of 136 Internet pharmacies: illegal rogue online pharmacies flourish and are long-lived. J Med Internet Res 2013 Sep 10;15(9):e199 [FREE Full text] [doi: 10.2196/jmir.2606] [Medline: 24021777]
34. Gaudiano MC, Borioni A, Antoniella E, Valvo L. Counterfeit Adderall containing aceclofenac from internet pharmacies. J Forensic Sci 2016 Jul 06;61(4):1126-1130. [doi: 10.1111/1556-4029.13095] [Medline: 27364297]
35. Liang BA, Mackey TK, Lovett KM. Illegal "no prescription" internet access to narrow therapeutic index drugs. Clin Ther 2013 May;35(5):694-700. [doi: 10.1016/j.clinthera.2013.03.019] [Medline: 23597722]
36. Orizio G, Merla A, Schulz PJ, Gelatti U. Quality of online pharmacies and websites selling prescription drugs: a systematic review. J Med Internet Res 2011 Sep 30;13(3):e74 [FREE Full text] [doi: 10.2196/jmir.1795] [Medline: 21965220]
37. Rodomonte AL, Gaudiano MC, Antoniella E, Lucente D, Crusco V, Bartolomei M, et al. Counterfeit drugs detection by measurement of tablets and secondary packaging colour. J Pharm Biomed Anal 2010 Oct 10;53(2):215-220. [doi: 10.1016/j.jpba.2010.03.044] [Medline: 20547026]
38. Wang T, Hoag SW, Eng ML, Polli J, Pandit NS. Quality of antiretroviral and opportunistic infection medications dispensed from developing countries and Internet pharmacies. J Clin Pharm Ther 2015 Feb;40(1):68-75 [FREE Full text] [doi: 10.1111/jcpt.12226] [Medline: 25381836]
39. Almomani H, Patel N, Donyai P. Reasons that lead people to end up buying fake medicines on the internet: qualitative interview study. JMIR Form Res 2023 Feb 16;7:e42887 [FREE Full text] [doi: 10.2196/42887] [Medline: 36795460]
40. Almomani H, Patel N, Donyai P. News media coverage of the problem of purchasing fake prescription medicines on the internet: thematic analysis. JMIR Form Res 2023 Mar 21;7:e45147 [FREE Full text] [doi: 10.2196/45147] [Medline: 36943354]
41. Hall A, Antonopoulos GA. Fake Meds Online: The Internet and the Transnational Market in Illicit Pharmaceuticals. London, England: Palgrave Macmillan; 2016.
42. Corbin J, Strauss A. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA: SAGE Publications Inc; 2008.
43. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
44. Middleton A, Morley KI, Bragin E, Firth HV, Hurles ME, Wright CF, et al. Attitudes of nearly 7000 health professionals, genomic researchers and publics toward the return of incidental results from sequencing research. Eur J Hum Genet 2016 Jan 29;24(1):21-29 [FREE Full text] [doi: 10.1038/ejhg.2015.58] [Medline: 25920556]

XSL•FO
RenderX

45. Middleton A, Bragin E, Parker M, Study D. Finding people who will tell you their thoughts on genomics-recruitment strategies for social sciences research. J Community Genet 2014 Oct 18;5(4):291-302 [FREE Full text] [doi: 10.1007/s12687-014-0184-2] [Medline: 24535681]
46. Baltar F, Brunet I. Social research 2.0: virtual snowball sampling method using Facebook. Internet Research 2012 Jan 27;22(1):57-74. [doi: 10.1108/10662241211199960]
47. Broniarczyk SM, Griffin JG. Decision difficulty in the age of consumer empowerment. J Consum Psychol 2014 May 24;24(4):608-625. [doi: 10.1016/j.jcps.2014.05.003]
48. Filieri R, Hofacker CF, Alguezaui S. What makes information in online consumer reviews diagnostic over time? The role of review relevancy, factuality, currency, source credibility and ranking score. Computers in Human Behavior 2018 Mar;80:122-131. [doi: 10.1016/j.chb.2017.10.039]
49. Jin L, Hu B, He Y. The recent versus the out-dated: an experimental examination of the time-variant effects of online consumer reviews. Journal of Retailing 2014 Dec;90(4):552-566. [doi: 10.1016/j.jretai.2014.05.002]
50. Khare A, Labrecque LI, Asare AK. The assimilative and contrastive effects of word-of-mouth volume: an experimental examination of online consumer ratings. Journal of Retailing 2011 Mar;87(1):111-126. [doi: 10.1016/j.jretai.2011.01.005]
51. Senecal S, Nantel J. The influence of online product recommendations on consumers' online choices. Journal of Retailing 2004 Jan;80(2):159-169. [doi: 10.1016/j.jretai.2004.04.001]
52. Baumeister RF. Yielding to temptation: self‐control failure, impulsive purchasing, and consumer behavior. Journal of Consumer Research 2002 Mar;28(4):670-676. [doi: 10.1086/338209]
53. Zenetti G, Klapper D. Advertising effects under consumer heterogeneity – the moderating role of brand experience, advertising recall and attitude. Journal of Retailing 2016 Sep;92(3):352-372. [doi: 10.1016/j.jretai.2016.02.004]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
GP: general practitioner
P: pharmacy-only
PBC: perceived behavioral control
POM: prescription-only medicines

Edited by A Mavragani; submitted 11.12.23; peer-reviewed by A Fittler, D Dinh; comments to author 02.05.24; revised version received 23.05.24; accepted 05.01.25; published 21.02.25.

Please cite as:
Naughton BD
Planned Behavior in the United Kingdom and Ireland Online Medicine Purchasing Context: Mixed Methods Survey Study
JMIR Form Res 2025;9:e55391
URL: https://formative.jmir.org/2025/1/e55391
doi:10.2196/55391
PMID:

©Bernard D Naughton. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Applications of Self-Driving Vehicles in an Aging Population

Sara Shu[1,2], DO; Benjamin K P Woo[2,3,4], MD

[1]Department of Community Internal Medicine, Geriatrics and Palliative, Mayo Clinic, 200 First Street SW, Rochester, MN, United States

[2]University of California, Los Angeles, Los Angeles, CA, United States

[3]Asian American Studies Center, University of California, Los Angeles, Los Angeles, CA, United States

[4]Chinese American Health Promotion Laboratory, University of California, Los Angeles, Los Angeles, CA, United States

Corresponding Author:
Sara Shu, DO
Department of Community Internal Medicine, Geriatrics and Palliative, Mayo Clinic, 200 First Street SW, Rochester, MN, United States

Abstract

The proportion of older adult drivers is increasing and represents a growing population that must contemplate reducing driving and eventually stopping driving. The advent of self-driving vehicles opens vast possibilities with practical and far-reaching applications for our aging population. Advancing technologies in transportation may help to overcome transportation barriers for less mobile individuals, transcend social and geographical isolation, and improve resource and medical access. Herein, we propose various applications and benefits that self-driving vehicles have in maintaining independence and autonomy specifically for our aging population to preserve aging.

(JMIR Form Res 2025;9:e66180) doi:10.2196/66180

KEYWORDS

self-driving; driverless; driver; autonomous vehicles; car; transportation; mobility; travel; vehicle; driving; artificial intelligence; gerontology; geriatric; older; elderly; aging; healthy aging; older adult; autonomy; independence; aging in place; health equity

Introduction

In the United States, driving remains a symbol of independence and is an important means to maintain a connection to and identity within the community. For many community-dwelling older adults, private vehicles remain the primary mode of transportation. In the United States, there are over 55 million licensed automobile drivers over 65 years of age, comprising over 25% of the total driving population [1]. Motor vehicle collisions are among the most common causes of nonfatal and fatal injuries in the United States in older adults [2,3]. Older adults have higher mortality compared to a younger population with similar motor vehicle collision-related injuries [4]. The association between increased dangers of driving and collisions in the older adult population is well-known. With increasing age comes impairments in vision and visuospatial and reaction times, all of which make it more dangerous and difficult to navigate safely [5]. Primary care providers and associated interdisciplinary teams are in a unique position to assess the fitness of older adults to drive safely and provide alternative transportation recommendations [6,7]. Developments in technology may also provide tools to assist with the assessment of driving safety. For example, speech data analysis during interactions with voice assistants has been suggested to accurately predict future accident experiences in older drivers [8].

Environmental factors such as congestion of roads and the distance to the locations of goods and services can influence driving reduction and cessation in older adults [9]. Even medical events requiring hospitalization have been shown to correlate with immediate or eventual driving reduction or cessation, which ultimately compromises the autonomy of individuals [10]. Moreover, driving cessation has been associated with declines in general health and mental, physical, social, and cognitive function, and has even been associated with increased risk of admission to long-term care facilities and mortality [11,12]. In this paper, we aim to provide an updated review of current and proposed applications of self-driving vehicles, particularly through the integration and application of mobility as a service to assist with prolonging the autonomy of aging persons, as well as review the limitations and future directions that have yet to be explored.

Towards Age-Friendly Communities

Over the past several decades, there has been a push to develop communities and city infrastructure to be more supportive of an aging population in areas such as housing, transportation, and access to health care and social services. There is great heterogeneity in the driving challenges and concerns of older adult drivers in different geographical settings: urban, suburban, and rural [13]. Differences in community infrastructure have shown varying densities in the incidence of fatal crashes for older drivers. For example, the majority of fatal crashes occurred

XSL•FO
RenderX

on rural roads, while suburban and urban roads conferred a different likelihood of crash location and mechanisms, such as multivehicle crashes, as well as those within intersections versus multilane roads [14]. The movement towards age-friendly communities has driven the development of the World Health Organization Global Network for Age-Friendly Cities and Communities to provide a framework of considerations when developing age-friendly environments [15]. These include the development of technologies, services, and systems to transcend physical and social barriers that can promote the inclusion and continued integration and engagement of older adults in the community. As there are increasing needs in the rapidly growing older adult population, there will be an increased need to develop innovative strategies to meet them.

Self-driving vehicles hold promise in maintaining autonomy and maintaining independence to allow living in and remaining at home for as long as possible. When the option of driving autonomy is exhausted, one begins to recruit assistance from social networks comprising relatives, neighbors, and friends. However, as social networks organically decrease among an aging population, socially isolated individuals may not have the luxury to call on such assistance. The reliance on public transportation as a primary mode of transportation is not always feasible nor uniformly available across the United States. Extreme temperatures and climates during the winter and summer months may furthermore be a barrier for older adults to use public transportation. Volunteer transportation programs, senior services, or service operations that charge a small fee are available but are met with similar limitations in access depending on the local infrastructure and individual socioeconomic status. Therefore, finding a reliable and affordable means of transportation remains a great challenge in our aging society.

Applications of Self-Driving Vehicles

Self-driving (also known as autonomous) vehicles present a promising alternative and emerging transportation option for our older adult population [16,17]. Semiautonomous vehicles already have applications for use as cognitive-assistive devices for community-dwelling older adults owing to features such as onboard navigation, parking assistance, lane departure, and collision avoidance technologies [18]. As fully autonomous vehicle technology continues to develop, the applications are far reaching. By freeing the need to engage another community-dwelling adult or one in the social network, the availability of transportation options will inadvertently increase. In metropolitan areas where self-driving vehicles have been launched as mobility services, many individuals are already using these mobility services for commuting to and from work. The use of this technology in the older population can help to close the divide in populations otherwise socially isolated and enable them to once again be independent through their community and be able to engage in the community-based or group activities they once enjoyed. It would significantly relieve the caregiver's burden of having to provide transportation to and from medical appointments, the grocery store, barber shop, or other social functions and leisure activities. Similar mobility service efforts are seen in the European Union and Asia, where working towards cooperative, connected, and automated mobility is being optimized to embolden digital connectivity between vehicles, the transportation infrastructure, and other road users. These so-called cooperative intelligent transport systems are the foundation for the widespread use and applications of these technologies to provide services for otherwise homebound community-dwelling older adults with limitations that would preclude them from independent and safe utilization of their own motor vehicles.

As the market is already saturated with home food and medical supply delivery services, self-driving vehicles can eventually be equipped to undertake these tasks, thereby relieving the current workforce to engage in other areas of need. As much of the home delivery services rely on assistance from community-dwelling adults, these individuals would now have increased caregiving capacities for their aging community and loved ones, to serve in capacities where technology is unable to provide assistance. Some of the country's largest ride-share companies are already pioneering the driverless vehicle space by partnering with home-based care agencies [19]. By increasing the capacity and affordability of these services, the use of these technologies can help to provide solutions to building more equitable age-friendly, disability-friendly communities.

Similar advantages of self-driving vehicles can be extended to mobility-isolated persons or those with disabilities. In order to optimize accessibility to these populations, car designs should be optimized to be age-friendly and wheelchair accessible, such as being equipped with a ramp or capacity for the vehicle to lower itself to the curbside. Vehicles such as Toyota Sienna's "Autono-MaaS" (Autonomous-Mobility as a Service) vehicles are already incorporating these features in partnership with transportation companies to provide increased transportation options for those with disabilities. Companies such as May Mobility and Via have launched new and free self-driving shuttle services called "Accessibili-D" for the local inhabitants of Detroit, Michigan aged 65 years and older or those with disabilities [20]. These on-demand shuttle services have deployed similar services in Texas, Arizona, and throughout Michigan and have expanded into retirement communities. The "goMARTI" (Minnesota's Autonomous Rural Transit Initiative) in Grand Rapids, Minnesota, was the first pilot to provide free, on-demand rides with the American Disabilities Act (ADA)-compliant vehicles to improve mobility to those otherwise mobility-isolated in rural areas and during inclement weather [21].

With the continued development of artificial intelligence (AI) technology, medical capabilities could eventually be installed in autonomous vehicles to be able to detect and respond to medical emergencies. They could be equipped with emergency medical supplies in addition to sensors and technology that could detect the need to redirect to the nearest medical center, connect with a remote telemedicine provider, or provide basic equipment such as oxygen or epinephrine pens when indicated. Vehicles could even be engineered to detect mood fluctuations including depressed, anxious, or agitated states, and be equipped with robotic animal interventions for temporal therapy or connect individuals to telemedicine assistance or crisis response teams. These so-called autonomous mobile clinics equipped

XSL•FO
RenderX

with AI software, medical diagnostic detection, equipment, and telemedicine capabilities could reshape and facilitate prompt medical care and can even help to bring medical care to once-difficult-to-reach, more rural populations [22,23]. Epidemiologically conscious uses of self-driving vehicles have already been piloted in the days of the COVID-19 pandemic, such as contactless goods transportation and transporting patients in isolation to prevent exposure to other high-infection risks. Nevertheless, such widespread integration and the adoption of clinical applications of AI have yet to be subject to thorough clinical validation and approval by the appropriate regulatory authorities [24].

Challenges and Limitations

The limitations of self-driving vehicles are many, and as far as our imaginations can carry us considering the tremendous potential these technologies have to offer, so too do foreseen issues and concerns arise. Hypothetical issues of the unknown have guided various expert opinions. Driving principles in regard to road user safety are expected to be upheld; however, there are countless scenarios by which these self-driving vehicles must navigate tensions between obeying the traffic code and avoiding human harm. This highlights the importance of the translation of these frameworks into the engineering of such self-driving vehicles to guarantee consistency and safety in behavior that is driven by automated technology [25]. Moreover, in addition to predicting the unpredictability of road interactions, passenger safety is needed as well. For example, in caring for an older adult with dementia, one with limited mobility, or one with mental health concerns, it is important to have the assurance of safety for the otherwise vulnerable adult, such as precluding drop-off into an unfavorable environment. For example, a person with compromised cognitive abilities may not recognize that it may be unsafe to be outside for extended periods of time in a heatwave, or a person with impulsivity may not comprehend the dangers of going to a casino. Such ethical dilemmas tread the fine line between autonomy and paternalism, especially when such technology works to expand the independence and autonomy of an increasingly diverse, vulnerable population. Whether autonomous vehicles should have the ability to override the decisions of their passengers under the pretense of responsibility for safety over vulnerable populations or risk the repercussions for invading the right to autonomy has yet to be explored.

The implications of AI software on passenger privacy in autonomous vehicles will always be a concern. As the software of these autonomous vehicles could potentially be targets of cyberattacks, the kind of protections that would be placed on ensuring the safety of its passengers is important. Moreover, the use of these technologies would result in varying perceptions and attitudes from different potential users, all of which are also important. Various small-scale studies have captured such perceptions in the older adult population, including in persons with dementia [26-31]. It is important, therefore, to address targeted user populations' concerns about the cost, functionality, performance, and trustworthiness of the automated vehicles to suit their needs. Social determinants of health can also affect the accessibility of these technologies and have the potential to further marginalize underserved communities and populations; therefore, affordability needs to be considered. By partnering with local, state, and national departments of transportation and agencies on aging, solutions are needed for the promotion of equity, accessibility, and inclusion of all the mobility needs of older adults as well as individuals with disabilities and caregivers, regardless of the socioeconomic status.

Conclusion

In summary, the advent of self-driving, or autonomous, vehicles can have beneficial effects on how individuals are able to age while ensuring autonomy. We are presently at a unique crossroads of personalizing this developing technology to facilitate and optimize independence for the older adult population and maintain autonomy for as long as possible. Engaging the proper stakeholders may enable the rapid development of these technologies to promote a more age-friendly society.

Conflicts of Interest

None declared.

References

1. Federal Highway Administration, Department of Transportation (US). Distribution of licensed drivers 2022. : FHWA URL: https://www.fhwa.dot.gov/policyinformation/statistics/2022/ [accessed 2024-06-07]
2. Platts-Mills TF, Hunold KM, Esserman DA, Sloane PD, McLean SA. Motor vehicle collision-related emergency department visits by older adults in the United States. Acad Emerg Med 2012 Jul;19(7):821-827. [doi: 10.1111/j.1553-2712.2012.01383.x] [Medline: 22724382]
3. Dellinger AM, Stevens JA. The injury problem among older adults: mortality, morbidity and costs. J Safety Res 2006;37(5):519-522. [doi: 10.1016/j.jsr.2006.10.001] [Medline: 17140602]
4. Perdue PW, Watts DD, Kaufmann CR, Trask AL. Differences in mortality between elderly and younger adult trauma patients: geriatric status increases risk of delayed death. J Trauma 1998 Oct;45(4):805-810. [doi: 10.1097/00005373-199810000-00034] [Medline: 9783625]
5. Anstey KJ, Windsor TD, Luszcz MA, Andrews GR. Predicting driving cessation over 5 years in older adults: psychological well-being and cognitive competence are stronger predictors than physical health. J Am Geriatr Soc 2006 Jan;54(1):121-126. [doi: 10.1111/j.1532-5415.2005.00471.x] [Medline: 16420208]

XSL•FO
RenderX

6. Pomidor A, editor. Clinician's Guide to Assessing and Counseling Older Drivers, 4th edition: The American Geriatrics Society; 2019. URL: https://geriatricscareonline.org/toc/clinicians-guide-to-assessing-and-counseling-older-drivers-4th-edition/B047/ [accessed 2024-04-08]
7. Hill LJN, Pignolo RJ, Tung EE. Assessing and counseling the older driver: a concise review for the generalist clinician. Mayo Clin Proc 2019 Aug;94(8):1582-1588. [doi: 10.1016/j.mayocp.2019.03.023] [Medline: 31378232]
8. Yamada Y, Shinkawa K, Kobayashi M, et al. Using speech data from interactions with a voice assistant to predict the risk of future accidents for older drivers: prospective cohort study. J Med Internet Res 2021 Apr 8;23(4):e27667. [doi: 10.2196/27667] [Medline: 33830066]
9. Vivoda JM, Heeringa SG, Schulz AJ, Grengs J, Connell CM. The influence of the transportation environment on driving reduction and cessation. Gerontologist 2017 Oct 1;57(5):824-832. [doi: 10.1093/geront/gnw088] [Medline: 27342439]
10. Gaulton TG, Neuman MD, Brown RT, Betz ME. Association of hospitalization with driving reduction and cessation in older adults. J Am Geriatr Soc 2021 Aug;69(8):2231-2239. [doi: 10.1111/jgs.17178] [Medline: 33864381]
11. Chihuri S, Mielenz TJ, DiMaggio CJ, et al. Driving cessation and health outcomes in older adults. J Am Geriatr Soc 2016 Feb;64(2):332-341. [doi: 10.1111/jgs.13931] [Medline: 26780879]
12. Qin W, Xiang X, Taylor H. Driving cessation and social isolation in older adults. J Aging Health 2020 Oct;32(9):962-971. [doi: 10.1177/0898264319870400] [Medline: 31431103]
13. Payyanadan RP, Lee JD, Grepo LC. Challenges for older drivers in urban, suburban, and rural settings. Geriatrics (Basel) 2018 Mar 22;3(2):14. [doi: 10.3390/geriatrics3020014] [Medline: 31011058]
14. Zwerling C, Peek-Asa C, Whitten PS, Choi SW, Sprince NL, Jones MP. Fatal motor vehicle crashes in rural and urban areas: decomposing rates into contributing factors. Inj Prev 2005 Feb;11(1):24-28. [doi: 10.1136/ip.2004.005959] [Medline: 15691985]
15. The global network for age-friendly cities and communities: looking back over the last decade,looking forward to the next. : World Health Organization; 2018 URL: https://www.who.int/publications/i/item/WHO-FWC-ALC-18.4 [accessed 2025-04-08]
16. Abdi S, de Witte L, Hawley M. Emerging technologies with potential care and support applications for older people: review of gray literature. JMIR Aging 2020 Aug 11;3(2):e17286. [doi: 10.2196/17286] [Medline: 32780020]
17. Shu S, Woo BKP. Use of technology and social media in dementia care: current and future directions. World J Psychiatry 2021 Apr 19;11(4):109-123. [doi: 10.5498/wjp.v11.i4.109] [Medline: 33889536]
18. Knoefel F, Wallace B, Goubran R, Sabra I, Marshall S. Semi-autonomous vehicles as a cognitive assistive device for older adults. Geriatrics (Basel) 2019 Nov 16;4(4):63. [doi: 10.3390/geriatrics4040063] [Medline: 31744041]
19. Donlan A, Home Health Care News. Driverless cars finding new lane in home-based care. 2020 Mar 1. URL: https://homehealthcarenews.com/2020/03/driverless-cars-finding-new-lane-in-home-based-care/ [accessed 2025-04-08]
20. Hope G, IOT World Today. Self-driving shuttle for elderly, disabled residents launches in detroit. 2024 Jun 12. URL: https://www.iotworldtoday.com/transportation-logistics/self-driving-shuttle-for-elderly-disabled-residents-launches-in-detroit [accessed 2025-04-08]
21. Bellan R, Tech Crunch. May mobility, via launch wheelchair-accessible autonomous shuttle service. 2022 Sep 28. URL: https://techcrunch.com/2022/09/28/may-mobility-via-launch-wheelchair-accessible-autonomous-shuttle-service [accessed 2025-04-08]
22. Liu S, Kong A, Huang Y, Liu X. Autonomous mobile clinics. Bull World Health Organ 2022 Sep 1;100(9):527-527A. [doi: 10.2471/BLT.22.288985] [Medline: 36062246]
23. Eliot L, Forbes. The problems of medical emergencies while inside a self-driving car are going to be challenging. 2020 Mar 17. URL: https://www.forbes.com/sites/lanceeliot/2020/03/17/the-problems-of-medical-emergencies-while-inside-a-self-driving-car-are-going-to-be-challenging/?sh=5179515a1670 [accessed 2025-04-08]
24. van Hartskamp M, Consoli S, Verhaegh W, Petkovic M, van de Stolpe A. Artificial intelligence in clinical health care applications: viewpoint. Interact J Med Res 2019 Apr 5;8(2):e12100. [doi: 10.2196/12100] [Medline: 30950806]
25. D'Amato A, Dancel S, Pilutti J, et al. Exceptional driving principles for autonomous vehicles. J of Law and Mobility 2022;1(2) [FREE Full text]
26. Haghzare S, Stasiulis E, Delfi G, et al. Automated vehicles for people with dementia: a "tremendous potential" that "has ways to go"-reports of a qualitative study. Gerontologist 2023 Jan 24;63(1):140-154. [doi: 10.1093/geront/gnac115] [Medline: 35926470]
27. Doulabi S, Hassan HM, Li B. Senior Americans' perceptions, attitudes, and safety concerns toward autonomous vehicles (AVs). J Safety Res 2023 Feb;84:218-231. [doi: 10.1016/j.jsr.2022.10.022] [Medline: 36868650]
28. Oxley J, Charlton J, Logan D, O'Hern S, Koppel S, Meuleners L. Safer vehicles and technology for older adults. Traffic Inj Prev 2019;20(sup2):S176-S179. [doi: 10.1080/15389588.2019.1661712] [Medline: 31674855]
29. Faverio M, Pew Research Center. Older americans more wary than younger adults about prospect of driverless cars on the road. 2022 Oct 24. URL: https://www.pewresearch.org/short-reads/2022/10/24/older-americans-more-wary-than-younger-adults-about-prospect-of-driverless-cars-on-the-road/ [accessed 2024-07-06]

XSL•FO
RenderX

30. Zandieh R, Acheampong RA. Mobility and healthy ageing in the city: Exploring opportunities and challenges of autonomous vehicles for older adults' outdoor mobility. Cities 2021 May;112:103135. [doi: 10.1016/j.cities.2021.103135]
31. Lajunen T, Sullman MJM. Attitudes toward four levels of self-driving technology among elderly drivers. Front Psychol 2021;12:682973. [doi: 10.3389/fpsyg.2021.682973] [Medline: 34248785]

Abbreviations

ADA: American Disabilities Act
AI: artificial intelligence

Edited by A Mavragani; submitted 05.09.24; peer-reviewed by M Ayuso, T Lajunen; revised version received 02.03.25; accepted 21.03.25; published 28.04.25.

Please cite as:
Shu S, Woo BKP
Applications of Self-Driving Vehicles in an Aging Population
JMIR Form Res 2025;9:e66180
URL: https://formative.jmir.org/2025/1/e66180
doi:10.2196/66180

© Sara Shu, Benjamin KP Woo. Originally published in JMIR Formative Research (https://formative.jmir.org), 28.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Assessing the Current Limitations of Large Language Models in Advancing Health Care Education

JaeYong Kim[1*]; Bathri Narayan Vajravelu[2*], MBBS, MPH, PhD

[1]School of Pharmacy, Massachusetts College of Pharmacy and Health Sciences, Boston, MA, United States

[2]Department of Physician Assistant Studies, Massachusetts College of Pharmacy and Health Sciences, 179 Longwood Avenue, Boston, MA, United States

[*]all authors contributed equally

Corresponding Author:
Bathri Narayan Vajravelu, MBBS, MPH, PhD
Department of Physician Assistant Studies, Massachusetts College of Pharmacy and Health Sciences, 179 Longwood Avenue, Boston, MA, United States

Abstract

The integration of large language models (LLMs), as seen with the generative pretrained transformers series, into health care education and clinical management represents a transformative potential. The practical use of current LLMs in health care sparks great anticipation for new avenues, yet its embracement also elicits considerable concerns that necessitate careful deliberation. This study aims to evaluate the application of state-of-the-art LLMs in health care education, highlighting the following shortcomings as areas requiring significant and urgent improvements: (1) threats to academic integrity, (2) dissemination of misinformation and risks of automation bias, (3) challenges with information completeness and consistency, (4) inequity of access, (5) risks of algorithmic bias, (6) exhibition of moral instability, (7) technological limitations in plugin tools, and (8) lack of regulatory oversight in addressing legal and ethical challenges. Future research should focus on strategically addressing the persistent challenges of LLMs highlighted in this paper, opening the door for effective measures that can improve their application in health care education.

(JMIR Form Res 2025;9:e51319) doi:10.2196/51319

KEYWORDS

large language model; generative pretrained transformer; health care education; health care delivery; artificial intelligence; LLM; ChatGPT; AI

Introduction

Artificial intelligence (AI) as a field of computer science research aims to maximize the development of software tools that capacitate machine-based simulation of human intelligence within defined parameters [1]. Often described as the pinnacle of information technology of this century, its integration into the greater boundaries of human infrastructure is expected to both fundamentally and permanently revolutionize the ongoing information age.

Large language models (LLMs) represent deep learning architectures called transformer networks [2]. It relies on neural networks that discerns the relationships within and between sequential data. Generative AI models, such as the generative pretrained transformers (GPT), collectively operate under such deep learning neural networks, thereby allowing for the training, processing, and analysis of large quantity of complex data to be possible at an exceptional rate and accuracy [3].

With the power of NVIDIA's graphics processing units, OpenAI announced a 175 billion parameter language model (GPT-3.0) to the public in June of 2020 [4]. GPT-3.0 has quickly garnered international attention for its ability to summarize, translate, classify, and engage in real time, generating detailed and human-like response to user queries. Since its release, a plethora of domain specific LLMs (such as Med-PaLM) and alternative models (such as Gemini and Claude) with advancements in multimodal capabilities have emerged [5], significantly enhancing the breadth of publicly accessible LLMs.

In May 2024, OpenAI introduced ChatGPT 4o (omni), an advanced iteration built on the GPT 4 model, which continues to establish new standards for LLMs. It incorporated various advancements over its predecessors and contemporary models, including improved performance in complexity, specialization, multilingual capabilities, and resource optimization [6]. Furthermore, OpenAI's integration of multimodal tools within the GPT interface has enabled the model to comprehend and generate responses based on both visual and verbal inputs. Specifically, the integration of DALL-E, a text-to-image model [7], along with the advanced data analysis feature [8], has expanded the traditionally text-based nature of LLMs into a more versatile, information integrating modality.

XSL•FO
RenderX

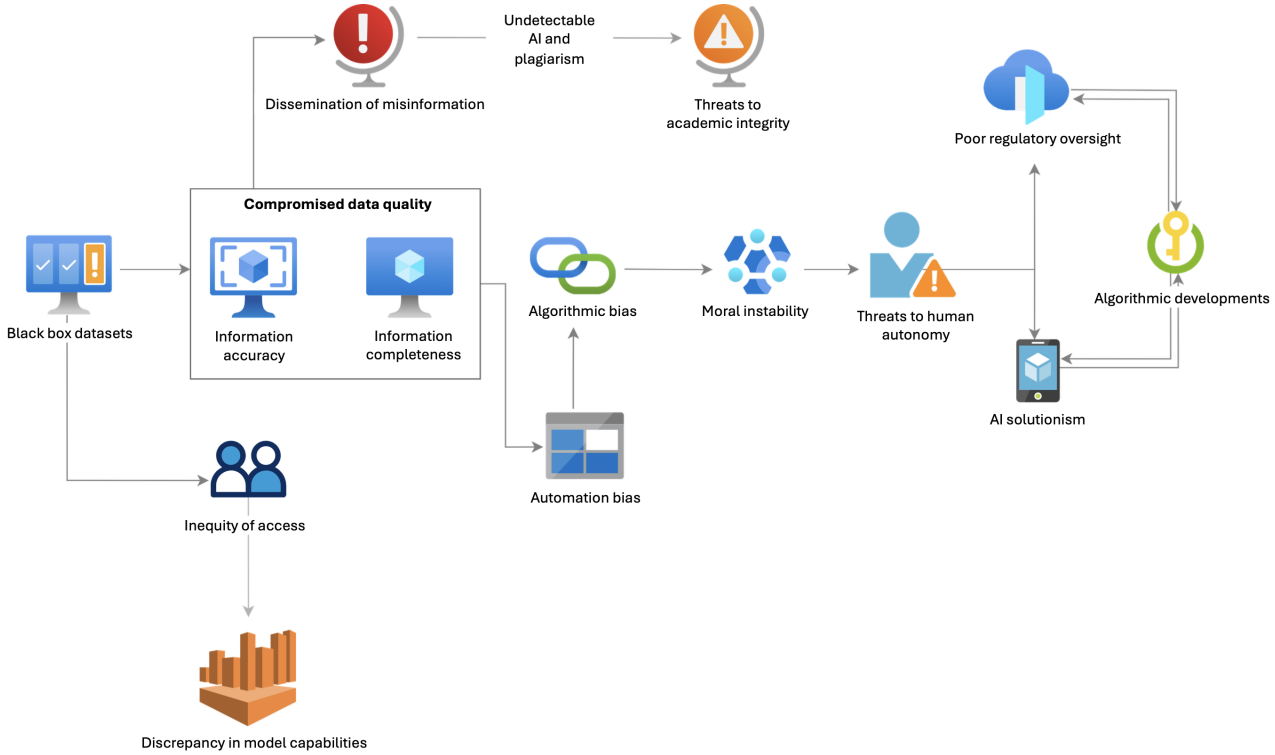
Despite the ongoing "AI race" that continues to raise the bar for generative AI technology, its integration into health care education poses significant and multifaceted challenges. In recent years, the shift toward embracing AI in education has gained more momentum, as highlighted by New York City's decision to rescind its previous ban on ChatGPT [9]. To this end, the primary objective of this review is to identify potential risks and propose effective mitigation strategies that developers of LLMs should adopt to facilitate their successful integration into health care education.

Risks of Large Language Models in Health Care Education

Overview

This section evaluates the current risks and limitations associated with the use of LLMs in health care education. It aims to highlight specific areas of concern that must be addressed at both individual and systemic levels. Figure 1 provides a summary of these issues through a flowchart.

Figure 1. Assessing risks and limitations of large language models in health care education. AI: artificial intelligence.



Threats to Academic Integrity

LLMs, like GPT-4o, provide individualized learning assistance tailored to the specific user demands. The increasing power of LLM's ability to process nuanced language understanding, along with enhanced problem thinking skills make it incredibly enticing for students to exploit its use under academic settings, both knowingly and unknowingly [10]. With increasing reliance on virtual learning platforms and learning management systems in health care education [11], AI-generated texts in particular compromise academic integrity by blurring the distinction between human and machine ingenuity.

Despite the actuation of detection software tools designed to combat AI-generated plagiarism (such as GPTZero), technological limitations and high error rates of these AI content detectors have been well-demonstrated [12-15]. In July 2023, OpenAI officially terminated its own AI detection software, AI Classifier, citing high false positives and inconsistent performance [16]. Comparative studies further affirm that neither human evaluators nor AI detectors are effective in identifying AI-generated medical literature [13]. Furthermore, generative texts can be easily modified to evade AI detection through simple grammatical adjustments, such as with the addition of adverbs or the use of synonyms [17,18]. Unsurprisingly, the widespread availability of AI-based plagiarism removal tools, such as Writehuman, leverages these strategies to convert AI-generated texts to appear more human-like, which may inadvertently discourage personal authenticity. Therefore, LLM's flexibility and immediacy in generating original, user-specific contents render the detection of AI-based plagiarism increasingly impractical [17].

The growing inability to differentiate between human and machine-generated writing poses significant threats to the principles of academic integrity, namely "honesty, trust, fairness, respect, and responsibility" [19]. For instance, the ethical controversy surrounding Elsevier's AI author policy, which led to a publication featuring an AI-generated introductory sentence, underscores the already pervasive exploitation of LLMs in academic writing and research [20]. The long-term repercussions of incorporating LLMs in education, such as the risk of over-reliance, erosion of critical thinking and problem-solving abilities, the deterioration of writing and summarization abilities, challenges associated with verifying information, and many more, have been well discussed [21,22]. In an era of infobesity, health care educators and LLM developers must establish clear policies and expectations that promote transparency and foster

XSL•FO
RenderX

dialogue about the use of AI-integrated technology, while actively working to minimize the associated risks.

Dissemination of Misinformation and Risks for Automation Bias

LLMs are engineered to generate a response by predicting the most probable response from input strings of its users. However, the details of both the data sources and the quality of data, along with specific parameters that are used to train LLMs, continue to remain undisclosed [23-25]. Such "black box" nature of LLMs pose challenges in assessing their reliability and robustness in health care applications [26-28], where information accuracy in health care knowledge is directly associated with the effectiveness of patient outcomes [29]. Any inaccuracies or misinformation generated by these models can manifest into inappropriate treatment strategies, misdiagnoses and professional incompetence that collectively lower the quality of health care delivery [30].

Despite receiving reinforcement learning from repetitive user feedback [31] and significant algorithmic improvements made to address AI hallucinations [32], both ChatGPT-Free and ChatGPT-Plus continue to remain susceptible to disseminating misinformation [24,33-35] and knowledge fabrication [36]. In a comprehensive meta-analysis assessing the performance of ChatGPT-3.5 in medical inquiries, the overall accuracy was found to be 56% [37]. Similarly, in a cross-sectional study comparing ChatGPT-3.5 and ChatGPT-4o with 284 physician-developed medical queries, only 50% of the responses were accurate [28]. Mounting evidence from domain-specific research [25,28,38-42] increasingly highlight significant concerns surrounding the accuracy and reliability of model data, pointing to their consistently poor performance in the context of health care education and clinical decision making.

Furthermore, LLM generated contents are highly vulnerable to committing source-based plagiarism, as it readily produces fabricated or inaccurate reference when asked to provide one [23,32,43]. This makes it impossible for users to reliably track and retrieve source information for verification [44] . Another persistent issue in the programmed nature of ChatGPT is its tendency to present information using bullet points. Although this format appears effective, it can inadvertently spread logical fallacies and misinformation. The hierarchical structure can misrepresent the significance of information, leading to inconsistencies that could unintentionally violate clinical practice guidelines [39].

To mitigate such concerns, OpenAI integrated third-party web plugins for GPT-4o, enabling the model to access the internet in real-time [45]. However, such strategy aggravates automation bias [27], where overreliance on AI capabilities increase the risks for AI solutionism. The dangers of AI solutionism have been highlighted by instances where AI-generated medical literature has successfully deceived human experts in broader, nonspecialized subject fields [14]. As a result, both health care students and clinicians may struggle to fulfill their responsibility to critically evaluate authentic knowledge from unverified information, impeding information seeking and processing abilities [10].

Without stringent regulatory frameworks and transparent disclosure of its training datasets, LLMs will struggle to gain the credibility necessary for integration into health care education and clinical practice. For this purpose, the development of a specialized, scaled-down and health care optimized LLMs, such as Google's Med-PaLM 2, will be essential. Mitigation steps toward enhancing the reliability LLMs in health care education include prioritizing transparency in training datasets, improving information accuracy within smaller datasets, ensuring the deidentification of medical data, and making algorithmic advancements to minimize AI hallucinations.

Challenges With Information Completeness and Consistency

Information completeness measures the extent to which a dataset is comprehensive. Data are considered complete when they encompass all required and relevant data fields, without any omissions [46]. LLMs, despite being built on vast body of knowledge, are susceptible to generating incomplete or partial representations of their knowledge dataset. Consequently, this leads to inconsistencies in the quality and comprehensiveness of their outputs [42].

In a study using a 3-point Likert-scale to measure completeness of ChatGPT's response to 180 medical queries, GPT-3.5 only scored 53.3% in terms of answer comprehensiveness [28]. Similarly, when GPT-4o's responses to clinical case questions were evaluated using the same scale, the model scored 53.3% in answer completeness [47]. In a different study, majority of the correct medical responses generated by GPT-3.5 were labeled as incomplete, where the omission of decision-making cut-offs, treatment strategies, and durations [38] greatly undermined information completeness. These findings were corroborated in another study, exhibited by GPT-3.5's lack of insights into treatment efficacy and age-or patient-specific recommendations, resulting in a 45% score in comprehensive output [48]. It is important to note that, when queried individually about such specific components, ChatGPT demonstrates understanding of relevant body of knowledge but fails to perform systemic and comprehensive integration of them under complex user requests.

As a natural extension of these observations, LLMs, such as the ChatGPT series, exhibit limited proficiency in generating outputs based on complex and information-rich inputs. This inverse relationship between input quantity and output quality has been well demonstrated by ChatGPT's tendency to produce ambiguous response when addressing lengthy medical queries that involve complex clinical context and nuance [49,50]. Furthermore, the GPT series were found to lack human-like understanding required for medical training, often leading to absurd responses [34]. The model's inability to fully grasp and accurately synthesize complex medical information undermines their reliability, leading to a phenomenon called "model overconfidence" [44] that generate imprecise and erroneous outputs. This supports the narrative that LLMs require robust oversight to fine-tune data consistency for the model to perform reliably [19].

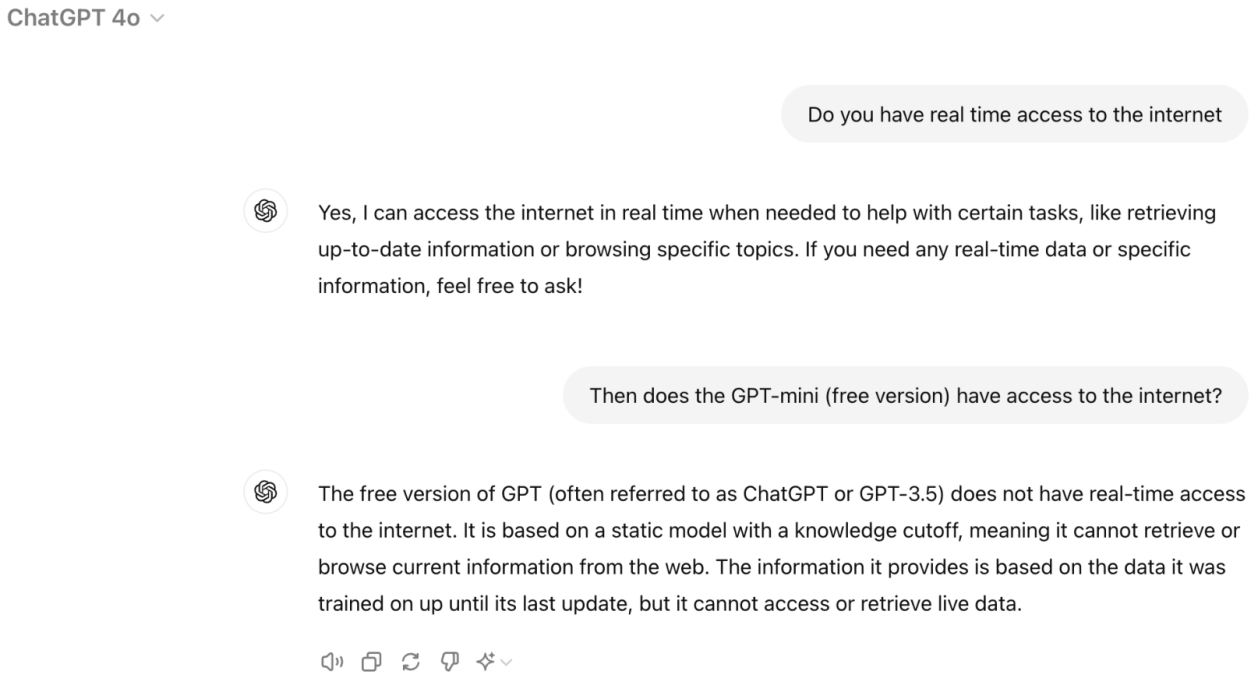
XSL•FO
RenderX

Along with information accuracy, information completeness is a critical component of the overall quality of information generated by LLMs. Developers of LLMs must refine the model's ability to consistently produce comprehensive and contextually accurate outputs based on improving both the quality and quantity of data parameters. Addressing these deficiencies is imperative in augmenting the applicability of LLMs in health care education.

Inequity of Access

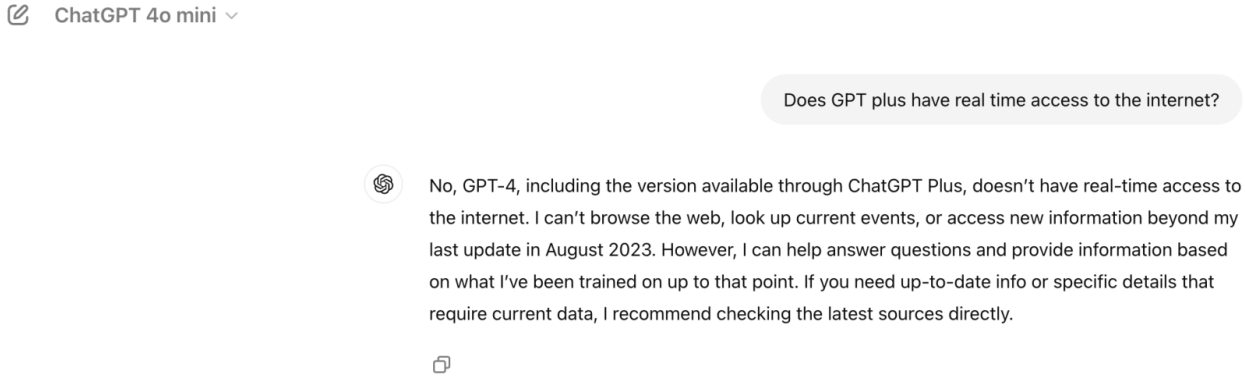
Monetization of certain LLMs, such as seen by Open AI's distribution of ChatGPT-Plus by a monthly subscription service [51], raise ethical concerns with regards to fair and equal accessibility to information. The vast majority of LLMs, with the exceptions of Gemini [52] and ChatGPT-Plus (Figure 2), lack real-time web access and, therefore, are incapable of retrieving information beyond their trained datasets. This undermines both the quality and consistency of the information that LLMs present, which renders the knowledge databases of most accessible LLMs presently outdated [9]. Figure 3 demonstrates how misinformation proliferates from outdated databases.

Figure 2. ChatGPT model differences in internet access capabilities.



XSL•FO
RenderX

Figure 3. Understanding misinformation risks in ChatGPT-Free.



Generative AI tools, like most other automated technological infrastructures, have inherent disparate impacts and thereby are less accessible to those whose training data are unavailable or inadequate for a specific language other than English (disparities in both language [53] and cultural proficiency [54]), those without necessary technology for access and limited digital literacy [55] those with physical disability and or impairment [56], and those with intellectual disability and or psychiatric impairments [57]. The monetary approach to the production of a premium service of GPT introduces economic disparity to the above list, where those that cannot afford subscription fees (US $20/month for ChatGPT-Plus [51]) are at a risk of aggravating digital divide disparity [58].

By this notion, standard search engines greatly outcompete LLM's use when it comes to both data reliability, uniformity, and completeness [26]. In other words, LLMs that provide incomplete or superficial information, often with significant delays in information currency [27], offer limited value in the fast-paced realms of health care education, research, and clinical practice.

A review (n=21) was conducted to compare information accuracy between the premium and free versions of ChatGPT using studies from PubMed and Google Scholar (Tables 1and 2). Inclusion criteria were based on analysis focused on health care education and clinical management, categorizing literature by subject matter to calculate mean accuracy differences. Categories were carefully defined to minimize selection bias, distinguishing between subjects such as "question banks" and "board exams," acknowledging GPT-Plus's expansive data access compared with that of GPT-Free.

XSL•FO
RenderX

Table . Comparative analysis of accuracy discrepancies: GPT-3.5 versus GPT-4.0.

Subject and category of comparison	Study	Reported percentage difference in accuracy levels between GPT-4.0 versus GPT-3.5 in corresponding literature, %	Accuracy levels (%) by category (n=7), mean (SD)
Board, entrance, and licensing exams	[59-65]	29.1, 27, 20, 22.2, 29.6, 29, 21.3	25.46 (4.14)
Question banks, mock exams, self-assessments	[66-72]	17.7, 10, 22, 23.8, 20.2, 16.5, 18.7	18.41 (4.47)
Clinical case, clinical questions, referencing	[73-79]	12.3, 27.8, 27, 10, 10.42, 13.6, 8.6	15.67 (8.17)

Table . Cross-tabulation of performance accuracy of GPT versions.[a]

	Studies reporting an accuracy of ≥70% (+), n	Studies reporting an accuracy of <70% (–), n	Total, n
ChatGPT-4.0 (PLUS)	18	3	21
ChatGPT-3.5 (FREE)	2	19	21
Total	20	22	42

[a]$P<0.001$. P was calculated as follows: $P = ([a+b]![c+d]![a+c]![b+d]!)/a!b!c!d!n!$

This review demonstrates that ChatGPT-Plus significantly outperforms the free version in information accuracy, with statistical significance confirmed by Fischer exact test ($P<0.001$) at a 95% CI. The study sets a 70% accuracy threshold for binary classification, highlighting a profound performance capability between the 2 versions. These findings raise significant concern about the potential for a digital divide, driven by the differential AI capabilities between the paid and free versions of LLMs.

Without a large-scale regulatory measure that dictate the quantifiable discrepancies in model capabilities between its free and premium editions, the ongoing rapid evolvement of AI technology will continue to create a stark polarity in equitable resource distribution [80]. In other words, development may further aggravate discrepancies in knowledge access and information availability in resource-constrained environments, particularly jeopardizing information accessibility in low-income demographics or users of 3rd world countries [55].

Risks of Algorithmic Bias

LLMs propagate algorithmic biases rooted in their training data, often leading to medically inaccurate and discriminatory responses. While no algorithm is designed to be innately and or deliberately discriminatory [81], they inevitably inherit and amplify existing sociocultural and historical biases embedded in their training [82,83]. In health care education and clinical practice, where racial misinformation persists [84], the integration of LLMs risks perpetuating biases, potentially leading to clinical errors and malpractice with significant repercussions.

LLMs, such as the ChatGPT series, Gemini, and Claude, have been demonstrated to perpetuate race-based biases in medicine, which unfortunately reflect and exacerbate existing health care disparities [85]. These models often recommend inconsistent treatment strategies for patients of different racial backgrounds, rooted in flawed assumptions about biological variations, such as differences in pain tolerance [86] and kidney function [87]. Furthermore, the GPT series has been criticized for promoting demographic stereotypes, which disproportionately associate certain diseases with specific races, ethnicities, and genders [88].

Furthermore, research has highlighted the influence of racial classification on the outputs generated by ChatGPT models. For instance, when users identified their race in questioning about HIV, ChatGPT-4o provides more detailed and supportive responses for White and Asian groups, while generating often overlooked or generalized responses by American Indians, Alaska Natives, and Pacific Islanders, reflecting bias in its training data [89]. These findings underscore the alarming need for developers of LLMs to re-establish bias-free training database to prevent reinforcement of discrimination.

Integration of LLMs to current practices and delivery of health care cannot proceed without significant algorithmic refinements and regulatory oversights that filter out both historical and existing biases. Having a high potential to act as a supplementary diagnostic tools and decision aids [50], the use of LLMs will impact health care education and, consequently, clinical outcomes. Furthermore, in an increasingly digitalized world, LLM's direct access or possible integration to electronic health record software can exacerbate existing discrimination. This can contribute to the polarization within social infrastructure on multiple levels.

Exhibition of Moral Instability

Moral competence in professionalism is grounded in the "knowledge, skills, and attitudes" required to address ethical issues [90]. Health care professionals regularly confront such challenges due to threats that substantiate ethical values and integrity [91]. Hence, ethics education in health care is crucial in preparing students to navigate ethically challenging workplace scenarios [92]. Therefore, it is vital to assess how comparable

XSL•FO
RenderX

LLMs are to humans in exerting stability and moral soundness when confronted with ethically challenging scenarios.

The lack of a firm moral stance in LLMs like the GPT series, coupled with their tendency to dispense moral advice, has been critically evaluated [93]. Research reveals that GPT-3.5 often generates contradictory responses to identical ethical dilemmas, offering recommendations that are shallow. It was also found that GPT corrupts user's moral competence by influencing user judgement, thereby undermining human autonomy. Similarly, while GPT-4o has shown success in identifying complex ethical dilemmas in medicine, it exhibits limited proficiency to fully encode the depth of real-world ethical challenges, particularly lacking understanding of "relational complexities and context-specific values" [94].

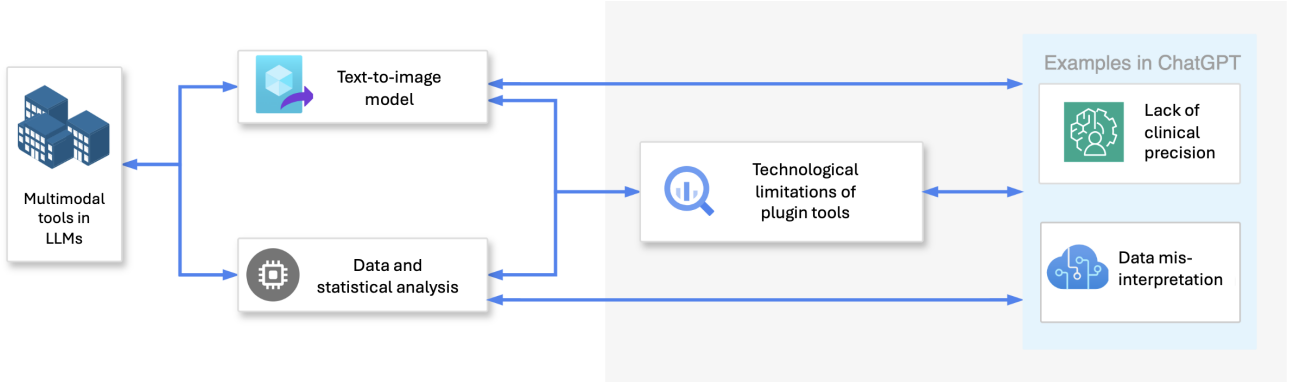
On a more positive note, ChatGPT-3.5's high accuracy in correctly answering bioethics questions [95,96] supports its possible use as an assistance or a reference tool in clinical decision-making. It accentuates a potential for GPT's ability to accurately address challenging contextual scenarios requiring high social intelligence and a firm grasp of ethical theories.

Technological Limitations in Plugin Tools

Overview

The rapid advancement of LLMs has driven the assimilation of multimodal technologies, such as text-to-image models and data analytics tools. However, their direct application in health care education presents numerous challenges, which are summarized in the flowchart shown in Figure 4.

Figure 4. Exploring technological limitations in multimodal plugin tools for LLMs. LLM: large language model.



Inadequate Image-Generating Capacity

The integration of the text-to-image generation model such as DALL-E and Stable Diffusion 3 into LLMs has enabled the generation of original and photorealistic images based on user descriptions [7]. This capability allows for multimodal input from a single user interface representing a significant advancement in AI technology. While the potential for such tools to advance health care education is recognized, practical applications seem distant, which require advancements to meet the rigorous standards required to train health professionals.

The most prominent technological deficiency of DALL-E and other built-in text-to-image AI models is its struggle with accurate text generation within images [97], which leads to production of inaccurate or nonsensical content. The illiteracy of these models diminishes its utility in health care education, particularly in subjects such as physiology and pharmacology, where accurate labeling and annotation in visual learning materials are critical.

Furthermore, the evaluation of DALL-E's limited capabilities in generating accurate medical imagery across various specialties has been highlighted. While DALL-E 2 can produce realistic radiographs, it poorly depicts pathological abnormalities such as tumors or fractures [98]. Furthermore, its generative capabilities for more complex imaging modalities, such as CT, MRI, and ultrasound, were found to be erroneous. In a similar study, DALL-E 3's ability to generate ECG tracings was assessed, revealing that the depicted waveforms were neither physiologically accurate nor interpretable [99]. In dermatology research, DALL-E 2 successfully and correctly illustrated only 20% of prevalent inflammatory skin conditions [100]. Such findings underscore the current limitations of text-to-image models, which lack both the understanding and capability to generate complex and pathologically accurate diagrams required for health care applications.

These evaluations collectively suggest that text-to-image models such as DALL-E, by design, is optimized for visual creativity and authenticity, but fall short of achieving clinical precision. Even if future iterations of DALL-E manage to address both textual and contextual complexities more effectively, concerns remain regarding about its performance in both consistency accuracy. The extent to which DALL-E can effectively combine "concepts, attributes, and styles" [101] for reliable representations for health care education remains a complex area for algorithmic development.

Poor Data Analysis Skills and Statistis Powerhouse

An emerging feature of LLMs is the integration of external plugins that enable advanced statistical functions and data analytics. Initially, LLMs, such as the early iterations of the GPT series, were criticized for their poor arithmetic capabilities [102]. ChatGPT's advanced data analysis feature (formerly known as code interpreter or python sandbox) can now perform a wide range of tasks that involve data analysis, statistical analysis, mathematical calculations, programming, file manipulation, text processing, to name a few [8]. The assimilation of such robust data-analytic tools to natural

XSL•FO
RenderX

language processing is expected to broaden LLM's applicability in health care education, research, and clinical practice.

However, when evaluated against biostatistics questions derived from the Oxford Handbook of Medical Statistics, the GPT series, encompassing both premium and free iterations, demonstrated a mean accuracy rate of 55% [103]. In this study, GPT-3.5 consistently failed in analysis of variance, χ^2 test, and sample size calculations. In a similar study, GPT-3.5's accuracy in general statistical analysis was found to be only 50% [104] indicating a significant technological gap. This limitation is primarily driven by the model's tendency to employ inappropriate statistical tests that engender data misinterpretation. Furthermore, GPT-4o's poor performance in advanced statistical methods for epidemiological studies has been highlighted, with authors cautioning against its use beyond intermediate levels in data analysis [105].

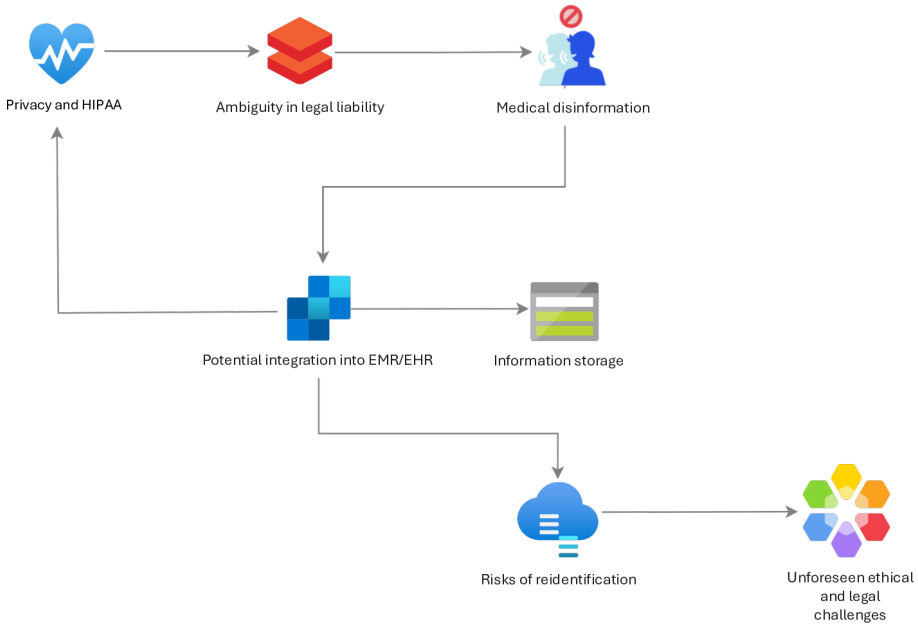
To transform LLMs into a more powerful tool for information integration in health care education, their development must advance alongside improvements in data analytics and presentation capabilities. Incorporating more specialized analytic tools with enhanced model accuracy will significantly ease the performance and interpretation of analytical results, thereby supporting diverse research efforts in the health care field.

Legal and Ethical Challenges of Integrating Large Language Models in Health Care Education

Overview

The rapid advancement of LLMs in the lack of strict regulatory or legal frameworks, has led to unforeseen bioethical challenges in both health care education and clinical management. These emerging issues are summarized in Figure 5, which provides a visual overview of the complexities that need to be addressed.

Figure 5. Navigating legal and ethical challenges in integrating large language models into health care education. EHR: electronic health record; EMR: electronic medical record; HIPAA: Health Insurance Portability and Accountability Act.



Privacy and Health Insurance Portability and Accountability Act

Biomedical libraries or medical research databases, which collects and stores valuable and comprehensive health information, is a highly useful data for training LLMs [106]. Although patient data is stringently protected through deidentification, reidentification of personal information has been made possible through cross-referencing with other databases [107].

Furthermore, LLM's interaction with patient or health care professionals could, in principle, collect and analyze personal health records, such as but not limited to medical histories and sensitive lab values into its own database [44]. The foreseeable integration of LLMs to both electronic health records and electronic medical records greatly heightens privacy concerns [108], as it involves exchange of information between 2 databases. This can drastically reduce the efforts made to anonymize confidential medical records. This has a potential to violate health insurance portability and accountability act and pose legal challenges as to how best to ensure the privacy regulations and protection of private data without compromising technological integrity and use of LLMs.

Ambiguity in Determining Legal Liability

Since its release, LLMs raised controversies over legal accountability for the absence of regulatory measures in protecting patients against clinical malpractice [44]. The

XSL•FO
RenderX

question fundamentally lies in who should be held liable for the potential harm that LLMs pose through its inaccurate clinical decisions or propagation of medical misinformation.

Noting the fact that OpenAI currently disclaims all legal responsibilities over the potential harm its generated contents may circulate [109], it is therefore inherently ambiguous and challenging to determine how legal liabilities and frameworks should be established in events of medical malpractice or errors that involve GPT usage. Such ambiguity highlights the urgency for legal guidelines to protect both patients and clinicians, while underscoring the need for AI licensing or development of domain specialized LLMs with added regulatory measures.

Medical Disinformation

LLMs can be intricately manipulated by malicious users to produce and disseminate medical disinformation: a phenomenon better known as "medical deepfakes" [110]. With LLM's increasing advancements in both written and visual realism, unethical exploitation is possible on both individual and collective levels, for example, falsifying personal medical records or creating fraudulent high-impact medical journals [111]. Therefore, the integration of LLMs into health care education and practice presents unprecedented regulatory and legal challenges, particularly in determining liability for medical forgery and falsification. Much like how other forms of AI deepfakes are criminalized [112], there needs an urgent legislative oversight to prevent the emergence of health care fraud.

Discussion

LLMs have great potential to augment and elevate health care education. To do so, developers must demonstrate heightened standards in ensuring security, accuracy, transparency, equity and sustainability of AI models to establish long-lasting reliability with human users. Furthermore, different stakeholders, especially those in bioethics, legislative, and regulatory bodies must also contribute relentlessly to systemically minimize both foreseeable and long-term repercussions of incorporating AI into the deeper boundaries of everyday human lives.

Therefore, future research endeavors must be oriented toward strategic mitigation of persistent flaws and limitations that this paper has attempted to address as few. With robust and constructive understanding of its current flaws, it opens many new avenues for research that can collectively remediate current shortcomings for its application in health care education. Such efforts will also allow for a greater control over this technology, thus allowing a more symbiotic relationship with generative AI technology.

Much like how the internet and computers have permanently altered human life [113], it is conceivable that future iterations of LLMs could become indispensable tools tailored for intellectual and professional productivity, akin to a Swiss army knife, in the expanding information age.

Conflicts of Interest

None declared.

References

1. Wang F, Preininger A. AI in health: state of the art, challenges, and future directions. Yearb Med Inform 2019 Aug;28(1):16-26. [doi: 10.1055/s-0039-1677908] [Medline: 31419814]
2. Large language models explained. NVIDIA. URL: https://www.nvidia.com/en-us/glossary/large-language-models/ [accessed 2024-08-29]
3. Yang X, Wang Y, Byrne R, Schneider G, Yang S. Concepts of artificial intelligence for computer-assisted drug discovery. Chem Rev 2019;119(18) [FREE Full text]
4. Hines K. History of ChatGPT: a timeline of the meteoric rise of generative AI chatbots. SearchEngine Journal. 2023. URL: https://www.searchenginejournal.com/history-of-chatgpt-timeline/488370/ [accessed 2025-01-09]
5. Gupta P. What is Google Bard, and how does it fare against ChatGPT? EasyInsights. 2023. URL: https://easyinsights.ai/blog/google-bard-everything-you-need-to-know-google-bard-vs-chat-gpt/ [accessed 2025-01-09]
6. Hello GPT-4o. OpenAI. URL: https://openai.com/index/hello-gpt-4o/ [accessed 2025-01-09]
7. DALL·E 3. OpenAI. URL: https://openai.com/index/dall-e-3 [accessed 2025-01-09]
8. How to use ChatGPT's advanced data analysis feature. MIT Sloan Teaching & Learning Technologies. URL: https://mitsloanedtech.mit.edu/ai/tools/data-analysis/how-to-use-chatgpts-advanced-data-analysis-feature/ [accessed 2025-01-09]
9. Fütterer T, Fischer C, Alekseeva A, et al. ChatGPT in education: global reactions to AI innovations. Sci Rep 2023 Sep 15;13(1):15310. [doi: 10.1038/s41598-023-42227-6] [Medline: 37714915]
10. Kufel J, Bargieł-Łączek K, Kocot S, et al. What as machine learning, artificial neural networks and deep learning?-examples of practical applications in medicine. Diagn (Basel) 2023 Aug 3;13(15):2582. [doi: 10.3390/diagnostics13152582] [Medline: 37568945]
11. Mahdavi Ardestani SF, Adibi S, Golshan A, Sadeghian P. Factors influencing the effectiveness of e-learning in healthcare: a fuzzy ANP study. Healthcare (Basel) 2023 Jul 16;11(14):2035. [doi: 10.3390/healthcare11142035] [Medline: 37510477]
12. Habibzadeh F. GPTZero performance in identifying artificial intelligence-generated medical texts: a preliminary study. J Korean Med Sci 2023 Sep 25;38(38):e319. [doi: 10.3346/jkms.2023.38.e319] [Medline: 37750374]

XSL•FO
RenderX

13. Gao CA, Howard FM, Markov NS, et al. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. NPJ Digit Med 2023 Apr 26;6(1):75. [doi: 10.1038/s41746-023-00819-6] [Medline: 37100871]
14. Cheng SL, Tsai SJ, Bai YM, et al. Comparisons of quality, correctness, and similarity between chatGPT-generated and human-written abstracts for basic research: cross-sectional study. J Med Internet Res 2023 Dec 25;25:e51229. [doi: 10.2196/51229] [Medline: 38145486]
15. Bellini V, Semeraro F, Montomoli J, Cascella M, Bignami E. Between human and AI: assessing the reliability of AI text detection tools. Curr Med Res Opin 2024 Mar 3;40(3):353-358. [doi: 10.1080/03007995.2024.2310086]
16. Edwards B. OpenAI discontinues its AI writing detector due to “low rate of accuracy”. Ars Technica. 2023. URL: https://arstechnica.com/information-technology/2023/07/openai-discontinues-its-ai-writing-detector-due-to-low-rate-of-accuracy/ [accessed 2025-01-09]
17. Guleria A, Krishan K, Sharma V, Kanchan T. ChatGPT: ethical concerns and challenges in academics and research. J Infect Dev Ctries 2023 Sep 30;17(9):1292-1299. [doi: 10.3855/jidc.18738] [Medline: 37824352]
18. Alser M, Waisberg E. Concerns with the usage of chatGPT in academia and medicine: a viewpoint. Am J Med Open 2023 Jun;9:100036. [doi: 10.1016/j.ajmo.2023.100036] [Medline: 39035060]
19. Kadayam Guruswami G, Mumtaz S, Gopakumar A, Khan E, Abdullah F, Parahoo SK. Academic integrity perceptions among health-professions’ students: a cross-sectional study in the middle east. J Acad Ethics 2023;21(2):231-249. [doi: 10.1007/s10805-022-09452-6] [Medline: 35815317]
20. Crotty D. The latest “crisis” - is the research literature overrun with ChatGPT- and LLM-generated articles? The Scholarly Kitchen. 2024. URL: https://scholarlykitchen.sspnet.org/2024/03/20/the-latest-crisis-is-the-research-literature-overrun-with-chatgpt-and-llm-generated-articles/ [accessed 2025-01-09]
21. Mehmet F. ChatGPT: bullshit spewer or the end of traditional assessments in higher education? JALT 2023;6(1) [FREE Full text] [doi: 10.37074/jalt.2023.6.1.9]
22. Kasneci E, Sessler K, Küchemann S, et al. ChatGPT for good? On opportunities and challenges of large language models for education. Learn Individ Differ 2023 Apr;103:102274. [doi: 10.1016/j.lindif.2023.102274]
23. Anderson N, Belavy DL, Perle SM, et al. AI did not write this manuscript, or did it? Can we trick the AI text detector into generated texts? The potential future of ChatGPT and AI in sports & exercise medicine manuscript generation. BMJ Open Sport Exerc Med 2023;9(1):e001568. [doi: 10.1136/bmjsem-2023-001568] [Medline: 36816423]
24. Lee H. The rise of chatgpt: exploring its potential in medical education. Anat Sci Ed 2024 Jul;17(5):926-931. [doi: 10.1002/ase.2270]
25. Saenger JA, Hunger J, Boss A, Richter J. Delayed diagnosis of a transient ischemic attack caused by ChatGPT. Wien Klin Wochenschr 2024 Apr;136(7-8):236-238. [doi: 10.1007/s00508-024-02329-1] [Medline: 38305909]
26. Khowaja SA, Khuwaja P, Dev K, Wang W, Nkenyereye L, Fortino G. ChatGPT needs SPADE (sustainability, privacy, digital divide, and ethics) evaluation: a review. TechRxiv. Preprint posted online on Nov 9, 2023. [doi: 10.36227/techrxiv.22619932.v4]
27. Nguyen T. ChatGPT in medical education: a precursor for automation bias? JMIR Med Educ 2024 Jan 17;10:e50174. [doi: 10.2196/50174] [Medline: 38231545]
28. Goodman RS, Patrinely JR, Stone CA Jr, et al. Accuracy and reliability of chatbot responses to physician questions. JAMA Netw Open 2023 Oct 2;6(10):e2336483. [doi: 10.1001/jamanetworkopen.2023.36483] [Medline: 37782499]
29. Kisekka V, Giboney JS. The effectiveness of health care information technologies: evaluation of trust, security beliefs, and privacy as determinants of health care outcomes. J Med Internet Res 2018 Apr 11;20(4):e107. [doi: 10.2196/jmir.9014] [Medline: 29643052]
30. Ngo E, Patel N, Chandrasekaran K, Tajik AJ, Paterick TE. The importance of the medical record: a critical professional responsibility. J Med Pract Manage 2016;31(5):305-308. [Medline: 27249883]
31. Gallifant J, Fiske A, Levites Strekalova YA, et al. Peer review of GPT-4 technical report and systems card. PLOS Dig Health 2024 Jan;3(1):e0000417. [doi: 10.1371/journal.pdig.0000417] [Medline: 38236824]
32. Goddard J. Hallucinations in ChatGPT: a cautionary tale for biomedical researchers. Am J Med 2023 Nov;136(11):1059-1060. [doi: 10.1016/j.amjmed.2023.06.012] [Medline: 37369274]
33. Abd-Alrazaq A, AlSaad R, Alhuwail D, et al. Large language models in medical education: opportunities, challenges, and future directions. JMIR Med Educ 2023 Jun 1;9:e48291. [doi: 10.2196/48291] [Medline: 37261894]
34. Waisberg E, Ong J, Masalkhi M, Lee AG. Large language model (LLM)-driven chatbots for neuro-ophthalmic medical education. Eye (Lond) 2024 Mar;38(4):639-641. [doi: 10.1038/s41433-023-02759-7] [Medline: 37749374]
35. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: implications in scientific writing. Cureus 2023 Feb;15(2):e35179. [doi: 10.7759/cureus.35179] [Medline: 36811129]
36. Zielinski C, Winker M, Aggarwal R, et al. Chatbots, ChatGPT, and scholarly manuscripts: WAME recommendations on ChatGPT and chatbots in relation to scholarly publications. Open Access Maced J Med Sci 2023 Jul 28;11(A):83-86. [doi: 10.3889/oamjms.2023.11502]
37. Wei Q, Yao Z, Cui Y, Wei B, Jin Z, Xu X. Evaluation of ChatGPT-generated medical responses: a systematic review and meta-analysis. J Biomed Inform 2024 Mar;151:104620. [doi: 10.1016/j.jbi.2024.104620]

XSL•FO
RenderX

38. Yeo YH, Samaan JS, Ng WH, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. Clin Mol Hepatol 2023 Jul;29(3):721-732. [doi: 10.3350/cmh.2023.0089]
39. Walker HL, Ghani S, Kuemmerli C, et al. Reliability of medical information provided by ChatGPT: assessment against clinical guidelines and patient information quality instrument. J Med Internet Res 2023 Jun 30;25:e47479. [doi: 10.2196/47479] [Medline: 37389908]
40. Ali MJ. ChatGPT and lacrimal drainage disorders: performance and scope of improvement. Ophthalmic Plast Reconstr Surg 2023;39(3):221-225. [doi: 10.1097/IOP.0000000000002418] [Medline: 37166289]
41. Van Bulck L, Moons P. What if your patient switches from Dr. Google to Dr. ChatGPT? A vignette-based survey of the trustworthiness, value, and danger of ChatGPT-generated responses to health questions. Eur J Cardiovasc Nurs 2024 Jan 12;23(1):95-98. [doi: 10.1093/eurjcn/zvad038]
42. Cankurtaran RE, Polat YH, Aydemir NG, Umay E, Yurekli OT. Reliability and usefulness of ChatGPT for inflammatory bowel diseases: an analysis for patients and healthcare professionals. Cureus 2023 Oct;15(10):e46736. [doi: 10.7759/cureus.46736] [Medline: 38022227]
43. Athaluri SA, Manthena SV, Kesapragada V, Yarlagadda V, Dave T, Duddumpudi RTS. Exploring the boundaries of reality: investigating the phenomenon of artificial intelligence hallucination in scientific writing through ChatGPT references. Cureus 2023 Apr;15(4):e37432. [doi: 10.7759/cureus.37432] [Medline: 37182055]
44. Wang C, Liu S, Yang H, Guo J, Wu Y, Liu J. Ethical considerations of using ChatGPT in health care. J Med Internet Res 2023 Aug 11;25:e48009. [doi: 10.2196/48009] [Medline: 37566454]
45. ChatGPT plugins. OpenAI. URL: https://openai.com/index/chatgpt-plugins/ [accessed 2025-01-09]
46. Chen H, Hailey D, Wang N, Yu P. A review of data quality assessment methods for public health information systems. Int J Environ Res Public Health 2014 May 14;11(5):5170-5207. [doi: 10.3390/ijerph110505170] [Medline: 24830450]
47. Hatia A, Doldo T, Parrini S, et al. Accuracy and completeness of ChatGPT-generated information on interceptive orthodontics: a multicenter collaborative study. J Clin Med 2024 Jan 27;13(3):735. [doi: 10.3390/jcm13030735] [Medline: 38337430]
48. Lakdawala N, Channa L, Gronbeck C, et al. Assessing the accuracy and comprehensiveness of ChatGPT in offering clinical guidance for atopic dermatitis and acne vulgaris. JMIR Dermatol 2023 Nov 14;6:e50409. [doi: 10.2196/50409] [Medline: 37962920]
49. Homolak J. Opportunities and risks of ChatGPT in medicine, science, and academic publishing: a modern Promethean dilemma. Croat Med J 2023 Feb;64(1):1-3. [doi: 10.3325/cmj.2023.64.1]
50. Deng J, Heybati K, Park YJ, Zhou F, Bozzo A. Artificial intelligence in clinical practice: a look at ChatGPT. Cleve Clin J Med 2024 Mar 1;91(3):173-180. [doi: 10.3949/ccjm.91a.23070] [Medline: 38429005]
51. GPT-4 is openai's most advanced system, producing safer and more useful responses. OpenAI. URL: https://openai.com/gpt-4 [accessed 2024-03-06]
52. Google Gemini: what is it, and how does it work? XDA. URL: https://www.xda-developers.com/google-bard/ [accessed 2024-05-01]
53. Zhang X, Li S, Hauer B, Shi N, Kondrak G. Don't trust ChatGPT when your question is not in English: a study of multilingual abilities and types of LLMs. Presented at: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing; Dec 6-10, 2023; Singapore. [doi: 10.18653/v1/2023.emnlp-main.491]
54. Lai VD, Ngo NT, Veyseh APB, et al. ChatGPT beyond English: towards a comprehensive evaluation of large language models in multilingual learning. arXiv. Preprint posted online on Apr 12, 2023. [doi: 10.48550/arXiv.2304.05613]
55. Wang X, Sanders HM, Liu Y, et al. ChatGPT: promise and challenges for deployment in low- and middle-income countries. Lancet Reg Health West Pac 2023 Dec;41:100905. [doi: 10.1016/j.lanwpc.2023.100905] [Medline: 37731897]
56. Kuzdeuov A, Nurgaliyev S, Varol HA. ChatGPT for visually impaired and blind. TechRxiv. Preprint posted online on May 3, 2023. [doi: 10.36227/techrxiv.22047080]
57. A social robot connected with ChatGPT to improve cognitive functioning in ASD subjects. Frontiers. URL: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1232177/full [accessed 2025-01-09]
58. Adhikari K, Naik N, Hameed BZ, Raghunath SK, Somani BK. Exploring the ethical, legal, and social implications of ChatGPT in urology. Curr Urol Rep 2024 Jan;25(1):1-8. [doi: 10.1007/s11934-023-01185-2] [Medline: 37735339]
59. Takagi S, Watari T, Erabi A, Sakaguchi K. Performance of GPT-3.5 and GPT-4 on the Japanese medical licensing examination: comparison study. JMIR Med Educ 2023 Jun 29;9:e48002. [doi: 10.2196/48002] [Medline: 37384388]
60. Meyer A, Riese J, Streichert T. Comparison of the performance of GPT-3.5 and GPT-4 with that of medical students on the written German medical licensing examination: observational study. JMIR Med Educ 2024 Feb 8;10:e50965. [doi: 10.2196/50965] [Medline: 38329802]
61. Jang D, Yun TR, Lee CY, Kwon YK, Kim CE. GPT-4 can pass the Korean national licensing examination for Korean medicine doctors. PLOS Dig Health 2023 Dec;2(12):e0000416. [doi: 10.1371/journal.pdig.0000416] [Medline: 38100393]
62. Farhat F, Chaudhry BM, Nadeem M, Sohail SS, Madsen D. Evaluating large language models for the national premedical exam in India: comparative analysis of GPT-3.5, GPT-4, and Bard. JMIR Med Educ 2024 Feb 21;10:e51523. [doi: 10.2196/51523] [Medline: 38381486]

XSL•FO
RenderX

63. Oh N, Choi GS, Lee WY. ChatGPT goes to the operating room: evaluating GPT-4 performance and its potential in surgical education and training in the era of large language models. Ann Surg Treat Res 2023;104(5):269. [doi: 10.4174/astr.2023.104.5.269]
64. Nakajima N, Fujimori T, Furuya M, et al. A comparison between GPT-3.5, GPT-4, and GPT-4V: can the large language model (ChatGPT) pass the Japanese board of orthopaedic surgery examination? Cureus 2024 Mar;16(3):e56402. [doi: 10.7759/cureus.56402] [Medline: 38633935]
65. Rojas M, Rojas M, Burgess V, Toro-Pérez J, Salehi S. Exploring the performance of ChatGPT versions 3.5, 4, and 4 with vision in the Chilean medical licensing examination: observational study. JMIR Med Educ 2024 Apr 29;10:e55048. [doi: 10.2196/55048] [Medline: 38686550]
66. Moshirfar M, Altaf AW, Stoakes IM, Tuttle JJ, Hoopes PC. Artificial intelligence in ophthalmology: a comparative analysis of GPT-3.5, GPT-4, and human expertise in answering StatPearls questions. Cureus 2023 Jun;15(6):e40822. [doi: 10.7759/cureus.40822] [Medline: 37485215]
67. Ali R, Tang OY, Connolly ID, et al. Performance of ChatGPT and GPT-4 on neurosurgery written board examinations. Neurosurgery 2023 Dec 1;93(6):1353-1365. [doi: 10.1227/neu.0000000000002632] [Medline: 37581444]
68. Brin D, Sorin V, Vaid A, et al. Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments. Sci Rep 2023 Oct 1;13(1):16492. [doi: 10.1038/s41598-023-43436-9] [Medline: 37779171]
69. Gill GS, Blair J, Litinsky S. Evaluating the performance of ChatGPT 3.5 and 4.0 on StatPearls oculoplastic surgery text- and image-based exam questions. Cureus 2024 Nov;16(11):e73812. [doi: 10.7759/cureus.73812] [Medline: 39691123]
70. Ali R, Tang OY, Connolly ID, et al. Performance of ChatGPT, GPT-4, and Google Bard on a neurosurgery oral boards preparation question bank. medRxiv. Preprint posted online on 2023. [doi: 10.1101/2023.04.06.23288265]
71. Taloni A, Borselli M, Scarsi V, et al. Comparative performance of humans versus GPT-4.0 and GPT-3.5 in the self-assessment program of American Academy of Ophthalmology. Sci Rep 2023 Oct 29;13(1):18562. [doi: 10.1038/s41598-023-45837-2] [Medline: 37899405]
72. Lee GU, Hong DY, Kim SY, et al. Comparison of the problem-solving performance of ChatGPT-3.5, ChatGPT-4, Bing Chat, and Bard for the Korean emergency medicine board examination question bank. Medicine (Abingdon) 2024;103(9):e37325. [doi: 10.1097/MD.0000000000037325]
73. Al-Ashwal FY, Zawiah M, Gharaibeh L, Abu-Farha R, Bitar AN. Evaluating the sensitivity, specificity, and accuracy of ChatGPT-3.5, ChatGPT-4, Bing AI, and Bard against conventional drug-drug interactions clinical tools. Drug Healthc Patient Saf 2023;15:137-147. [doi: 10.2147/DHPS.S425858] [Medline: 37750052]
74. Agharia S, Szatkowski J, Fraval A, Stevens J, Zhou Y. The ability of artificial intelligence tools to formulate orthopaedic clinical decisions in comparison to human clinicians: An analysis of ChatGPT 3.5, ChatGPT 4, and Bard. J Orthop 2024 Apr;50:1-7. [doi: 10.1016/j.jor.2023.11.063] [Medline: 38148925]
75. Lechien JR, Briganti G, Vaira LA. Accuracy of ChatGPT-3.5 and -4 in providing scientific references in otolaryngology–head and neck surgery. Eur Arch Otorhinolaryngol 2024 Apr;281(4):2159-2165. [doi: 10.1007/s00405-023-08441-8]
76. Balasanjeevi G, Surapaneni KM. Comparison of ChatGPT version 3.5 & 4 for utility in respiratory medicine education using clinical case scenarios. Respir Med Res 2024 Jun;85:101091. [doi: 10.1016/j.resmer.2024.101091] [Medline: 38657295]
77. Liang R, Zhao A, Peng L, et al. Enhanced artificial intelligence strategies in renal oncology: iterative optimization and comparative analysis of GPT 3.5 versus 4.0. Ann Surg Oncol 2024 Jun;31(6):3887-3893. [doi: 10.1245/s10434-024-15107-0] [Medline: 38472675]
78. Momenaei B, Wakabayashi T, Shahlaee A, et al. Assessing ChatGPT-3.5 versus ChatGPT-4 performance in surgical treatment of retinal diseases: a comparative study. Ophthalmic Surg Lasers Imaging Retina 2024 Aug;55(8):481-482. [doi: 10.3928/23258160-20240227-02] [Medline: 38531015]
79. Samaan JS, Rajeev N, Ng WH, et al. ChatGPT as a source of information for bariatric surgery patients: a comparative analysis of accuracy and comprehensiveness between GPT-4 and GPT-3.5. Obes Surg 2024 May;34(5):1987-1989. [doi: 10.1007/s11695-024-07212-6] [Medline: 38564173]
80. Singh OP. Artificial intelligence in the era of ChatGPT - Opportunities and challenges in mental health care. Indian J Psychiatry 2023 Mar;65(3):297-298. [doi: 10.4103/indianjpsychiatry.indianjpsychiatry_112_23] [Medline: 37204980]
81. Adams-Prassl J, Binns R, Kelly-Lyth A. Directly discriminatory algorithms. Mod Law Rev 2023 Jan;86(1):144-175. [doi: 10.1111/1468-2230.12759] [Medline: 37065788]
82. Hoffman S, Podgurski A. Artificial intelligence and discrimination in health care. Yale J Health Policy Law Ethics 2020;19(3) [FREE Full text]
83. Amin KS, Forman HP, Davis MA. Even with ChatGPT, race matters. Clin Imaging 2024 May;109:110113. [doi: 10.1016/j.clinimag.2024.110113]
84. Hamed S, Bradby H, Ahlberg BM, Thapar-Björkert S. Racism in healthcare: a scoping review. BMC Public Health 2022 May 16;22(1):988. [doi: 10.1186/s12889-022-13122-y] [Medline: 35578322]
85. Omiye JA, Lester JC, Spichak S, Rotemberg V, Daneshjou R. Large language models propagate race-based medicine. NPJ Digit Med 2023 Oct 20;6(1):195. [doi: 10.1038/s41746-023-00939-z] [Medline: 37864012]

XSL•FO
RenderX

86. Hoffman KM, Trawalter S, Axt JR, Oliver MN. Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. Proc Natl Acad Sci U S A 2016 Apr 19;113(16):4296-4301. [doi: 10.1073/pnas.1516047113] [Medline: 27044069]
87. Tsai JW, Cerdeña JP, Goedel WC, et al. Evaluating the impact and rationale of race-specific estimations of kidney function: estimations from U.S. NHANES, 2015-2018. E Clin Med 2021 Dec;42:101197. [doi: 10.1016/j.eclinm.2021.101197] [Medline: 34849475]
88. Zack T, Lehman E, Suzgun M, et al. Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study. Lancet Dig Health 2024 Jan;6(1):e12-e22. [doi: 10.1016/S2589-7500(23)00225-X]
89. Yang Y, Liu X, Jin Q, Huang F, Lu Z. Unmasking and quantifying racial bias of large language models in medical report generation. Commun Med (Lond) 2024 Sep 10;4(1):176. [doi: 10.1038/s43856-024-00601-z] [Medline: 39256622]
90. Koskenvuori J, Stolt M, Suhonen R, Leino-Kilpi H. Healthcare professionals' ethical competence: a scoping review. Nurs Open 2019 Jan;6(1):5-17. [doi: 10.1002/nop2.173] [Medline: 30534390]
91. Iyalomhe GBS. Medical ethics and ethical dilemmas. Niger J Med 2009;18(1):8-16. [Medline: 19485140]
92. Andersson H, Svensson A, Frank C, Rantala A, Holmberg M, Bremer A. Ethics education to support ethical competence learning in healthcare: an integrative systematic review. BMC Med Ethics 2022 Mar 19;23(1):29. [doi: 10.1186/s12910-022-00766-z] [Medline: 35305627]
93. Krügel S, Ostermaier A, Uhl M. ChatGPT's inconsistent moral advice influences users' judgment. Sci Rep 2023 Apr 6;13(1):4569. [doi: 10.1038/s41598-023-31341-0] [Medline: 37024502]
94. Balas M, Wadden JJ, Hébert PC, et al. Exploring the potential utility of AI large language models for medical ethics: an expert panel evaluation of GPT-4. J Med Ethics 2024 Feb;50(2):90-96. [doi: 10.1136/jme-2023-109549]
95. Franco D'Souza R, Mathew M, Louis Palatty P, Surapaneni KM. Teaching humanism with humanoid: evaluating the potential of ChatGPT-4 as a pedagogical tool in bioethics education using validated clinical case vignettes. Int J Ethics Educ 2024 Apr 30. [doi: 10.1007/s40889-024-00190-4]
96. Chen J, Cadiente A, Kasselman LJ, Pilkington B. Assessing the performance of ChatGPT in bioethics: a large language model's moral compass in medicine. J Med Ethics 2024 Feb;50(2):97-101. [doi: 10.1136/jme-2023-109366]
97. Singh G, Deng F, Ahn S. Illiterate DALL-E learns to compose. arXiv. Preprint posted online on Oct 17, 2021. [doi: 10.48550/arXiv.2110.11405]
98. Adams LC, Busch F, Truhn D, Makowski MR, Aerts H, Bressem KK. What does DALL-E 2 know about radiology? J Med Internet Res 2023 Mar 16;25:e43110. [doi: 10.2196/43110] [Medline: 36927634]
99. Zhu L, Mou W, Wu K, Zhang J, Luo P. Can DALL-E 3 reliably generate 12-Lead ECGs and teaching illustrations? Cureus 2024 Jan;16(1):e52748. [doi: 10.7759/cureus.52748] [Medline: 38384621]
100. Cheraghlou S. Evaluating dermatologic domain knowledge in DALL-E 2 and potential applications for dermatology-specific algorithms. Int J Dermatol 2023 Oct;62(10):e521-e523. [doi: 10.1111/ijd.16683] [Medline: 37057623]
101. DALL·E 2. OpenAI. URL: https://openai.com/index/dall-e-2 [accessed 2025-01-09]
102. Why is chatgpt bad at math? Baeldung. 2023. URL: https://www.baeldung.com/cs/chatgpt-math-problems [accessed 2025-01-09]
103. Ignjatović A, Stevanović L. Efficacy and limitations of ChatGPT as a biostatistical problem-solving tool in medical education in Serbia: a descriptive study. J Educ Eval Health Prof 2023;20:28. [doi: 10.3352/jeehp.2023.20.28] [Medline: 37840252]
104. Ordak M. ChatGPT's skills in statistical analysis using the example of allergology: do we have reason for concern? Healthc (Basel) 2023 Sep 15;11(18):2554. [doi: 10.3390/healthcare11182554] [Medline: 37761751]
105. Huang Y, Wu R, He J, Xiang Y. Evaluating ChatGPT-4.0's data analytic proficiency in epidemiological studies: a comparative analysis with SAS, SPSS, and R. J Glob Health 2024 Mar 29;14:04070. [doi: 10.7189/jogh.14.04070] [Medline: 38547497]
106. McKay F, Williams BJ, Prestwich G, Bansal D, Treanor D, Hallowell N. Artificial intelligence and medical research databases: ethical review by data access committees. BMC Med Ethics 2023 Jul 8;24(1):49. [doi: 10.1186/s12910-023-00927-8] [Medline: 37422629]
107. Rothstein MA. Is deidentification sufficient to protect health privacy in research? Am J Bioeth 2010 Sep;10(9):3-11. [doi: 10.1080/15265161.2010.494215] [Medline: 20818545]
108. Baumgartner C. The potential impact of ChatGPT in clinical and translational medicine. Clin Transl Med 2023 Mar;13(3):e1206. [doi: 10.1002/ctm2.1206] [Medline: 36854881]
109. Terms of use. OpenAI. URL: https://openai.com/policies/terms-of-use [accessed 2024-05-03]
110. Cohen IG. What should ChatGPT mean for bioethics? Am J Bioeth 2023 Oct;23(10):8-16. [doi: 10.1080/15265161.2023.2233357] [Medline: 37440696]
111. Else H. Abstracts written by ChatGPT fool scientists. Nature New Biol 2023 Jan;613(7944):423. [doi: 10.1038/d41586-023-00056-7] [Medline: 36635510]
112. Mai KT, Bray S, Davies T, Griffin LD. Warning: humans cannot reliably detect speech deepfakes. PLoS ONE 2023;18(8):e0285333. [doi: 10.1371/journal.pone.0285333] [Medline: 37531336]
113. Hoehe MR, Thibaut F. Going digital: how technology use may influence human brains and behavior. Dialogues Clin Neurosci 2020 Jun;22(2):93-97. [doi: 10.31887/DCNS.2020.22.2/mhoehe] [Medline: 32699509]

XSL•FO
RenderX

Abbreviations

AI: artificial intelligence
GPT: generative pretrained transformer
LLM: large language model

Edited by A Mavragani; submitted 28.07.23; peer-reviewed by E Vashishtha, S Fazeli; revised version received 31.08.24; accepted 03.09.24; published 16.01.25.

Please cite as:
Kim J, Vajravelu BN
Assessing the Current Limitations of Large Language Models in Advancing Health Care Education
JMIR Form Res 2025;9:e51319
URL: https://formative.jmir.org/2025/1/e51319
doi:10.2196/51319

© JaeYong Kim, Bathri Narayan Vajravelu. Originally published in JMIR Formative Research (https://formative.jmir.org), 16.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Promoting Digital Health Data Literacy: The Datum Project

Daniel Powell[1*], BCL, JD; Laiba Asad[1*], BCL, JD; Elissa Zavaglia[2], MSc, PhD; Manuela Ferrari[2,3], MHS, PhD

[1]Faculty of Law, McGill University, Montreal, QC, Canada

[2]The Douglas Research Centre, 6875 Boulevard LaSalle (Perry C3 E-3102), Montreal, QC, Canada

[3]Department of Psychiatry, McGill University, Montreal, QC, Canada

[*]these authors contributed equally

Corresponding Author:
Manuela Ferrari, MHS, PhD
The Douglas Research Centre, 6875 Boulevard LaSalle (Perry C3 E-3102), Montreal, QC, Canada

Abstract

With the increased use of digital health innovations in Canadian health care, educating health care users, professionals, and researchers on the ethical challenges and privacy implications of these tools is essential. The Datum project, funded by the Fondation Barreau du Quebec, was created to help these actors better understand legal and ethical issues regarding the collection, use, and disclosure of digital health data for the purposes of scientific research, thereby enhancing literacy around data privacy. The project consists of a multimedia website divided into legislation and policy documents and narrative-based video content. Users can access the core legislation and policies governing the collection and use of health care data geared toward researchers and health practitioners. Users can also view the narrative-based video content explaining key concepts related to digital health data. The Datum project makes an original contribution to the field of law and ethics in health science research by using novel approaches, such as learning health systems and data banks, to improve equity in health care delivery and by generating multimedia content aimed at encouraging health care users to become better consumers and supporting the collective use of their data. The Datum project also promotes digital literacy as a digital communication tool, which has the significant potential to improve health outcomes, bridge the digital divide, and reduce health inequities.

(JMIR Form Res 2025;9:e60832) doi:10.2196/60832

KEYWORDS

health data; digital data; medical records; legislation; ethics; knowledge dissemination; learning health system; data bank

The Digital Health Landscape in Canada

The landscape of digital health data in Canada is multifaceted, encompassing issues of access, cultural relevance, policy implementation, and ethical considerations in using digital health tools in medicine [1]. There is a growing recognition of the potential for digital health innovations to meet the health care needs of Canadians, notwithstanding the challenges and complexities of integrating digital health data in health systems [2]. Public health research highlights the ethical challenges and privacy implications of using digital public health tools, emphasizing the need to find solutions while harnessing their potential [3]. Gasser et al [3] investigated the impact of using big data for health research and surveillance on the public demand for transparency, trust, and fairness. Such ethical challenges may be overcome by balancing risks and benefits to individuals and populations while educating consumers on this topic.

The Datum project aimed to support health care users, professionals, and researchers to better understand the legal and ethical implications of digital health data. There was a need for accessible ways of understanding the collection, storage, and uses of data and how data can be leveraged to meet the objectives of scientific research. Multimedia content and a library of resources were developed to achieve this end.

Background: Genesis of the Datum Project

The word “datum,” the singular of data, originates from the Latin verb *dare*, meaning “to give.” As an act of giving, data may be viewed as not only the result or object of research but also an object arising from a relational context. When patients consent to participate in a research study, they agree to “give” researchers information about their health and health outcomes. The process of collecting, using, and aggregating this information turns it into data that involves the interests of those who originally shared it.

Digital health data are revolutionizing medicine, driving personalized and measurement-based care approaches, informing evidence-based practices, enabling population health management, facilitating early intervention services, supporting research and innovation, and transforming ways of delivering and monitoring health care [4,5]. Analyses of health data reveal the patient’s unique characteristics, risk and protective factors, and treatment responses while providing information on service

XSL•FO
RenderX

outcomes [6]. This allows providers to develop more tailored and precise medical interventions, including targeted therapies and personalized treatment plans while improving the quality of services [7,8].

Influenced by the growing importance of digitalized data, recent legislative and regulatory changes require researchers to carefully consider the relational aspects of data and how data about people is collected, processed, and used. The European Union, through the General Data Protection Regulation, spearheaded early efforts to address these issues, establishing a global standard regarding the responsibilities of those who act as the custodians of data [9]. Recent legislation in the Canadian province of Quebec (eg, Law 25) recognizes the need for enhanced privacy and confidentiality policies, a clearly informed consent process for research participants, and governance frameworks outlining these policies [10].

This rapidly evolving issue calls for enhanced literacy around digital health data, empowering patients, caregivers, and providers to engage with data policies, ask questions, and increase their understanding of their privacy rights. Literacy about data privacy enables individuals to make informed decisions about the extent to which they share their personal information and to advocate for their privacy rights within the broader data governance framework [11].

It is important to empower health care users to support the collective use of data, which extends beyond individual interests, the economic value of data, and commercial use. The current shift toward broader considerations of social and collective value includes introducing democratic forms of governing data production [12]. Moreover, the digital economy should be relational such that individuals are interconnected within population-based data relations [12]. For example, learning health systems align data, technology, and care by continuously aggregating and analyzing ongoing health care data to improve future services, thereby creating a continuous feedback cycle for learning and quality improvement [13,14].

Content of the Datum Website

The Datum project, a multimedia website hosted at McGill University, offers users easy access to core legislation and policy documents that offer an overview of the legislative and policy frameworks governing the collection and use of health care data in Canada and a multimedia section with narrative-based video content capturing realistic exchanges between patients and providers on the collection, use, and sharing of digital health data [15].

Narrative-Based Video Content

Each video animation is set in a real-world context where health care users may be asked to share their digital health data (Figure 1). For example, doctors generally gather intake information from patients seen at medical appointments, including their medical history and symptoms, then upload these data into the patients' electronic health records. Patients who agree to participate in research studies also sign consent forms. Health care users may also identify and share concerns about sharing their personal data through dialogue with an authority figure within the health care system. Their interactions and the responses they receive demonstrate that health care providers appreciate their concerns.

By focusing on the questions of health care users regarding the collection, use, retention, and disclosure of their data in such instances, each video highlights the importance of empowering users to control their data, notably through consent and authorization, data access and transparency, revocation of consent, and data security and privacy. Through these videos, the Datum project hopes to empower health care users to make informed decisions by developing literacy in data privacy and understanding the implications of sharing digital health data in health care and research contexts. The scenarios presented also emphasize the need for collaboration and dialogue between different actors, as well as the advantages of sharing digital health data, such as informing research, empowering future research, and improving health outcomes and the delivery of health care services.

The videos build on each other to provide a comprehensive overview of the legal and ethical implications of using digital health data. The first video, "Introduction to Digital Health Data," focuses on foundational concepts and addresses the difference between data and information. The first and second videos, which situate the characters in both clinical and research contexts, explain some foundational concepts with which health care users may be unfamiliar, such as digital health data and the difference between information and data, as well as primary and secondary uses of digital health data. The third and fourth videos present more complex systems, specifically learning health systems and data banks, where actors in the research and clinical contexts interact with each other and work together to improve health care systems and generate new knowledge. These later videos illustrate the relational aspects of data, specifically how a health care user's data can be used for broader societal purposes. In order to develop learning health systems and data banks which rely on relational databases, health care users have a crucial need to understand the scope and potential of their consent to improve health care.

XSL•FO
RenderX

Figure 1. Image of 4 videos from the Datum multimedia library.

Content Library

As a complement to accessible multimedia content for health care users, the Datum website also features a content library, organized according to legislation and policy documents (Figure 2). This section of the website targets researchers and health care practitioners.

Figure 2. From image of how to select the legislation and policy documents Datum libraries.

XSL•FO
RenderX

Legislation

The regulation of personal information in federalist states like Canada is governed by a complex of provincial and federal rules. The legislation page describes the division of these rules between provincial and federal authorities. At the federal level, the use, collection, and processing of personal information for commercial purposes is subject to the federal Personal Information Protection and Electronic Documents Act. In general, privacy regulation is separated according to rules for private and for public data. Health care data, which may be subject to these privacy regulations, is typically subject to additional rules. The applicable legislation for each province and territory distinguishes three categories of personal information: (1) health information (personal health care information), (2) personal information held by public bodies, and (3) personal information held by private entities.

Policy Documents

The policy documents page provides additional resources for user consultation regarding the collection, use, and disclosure of personal information in the health care context. These materials were selected from the publications of reputable government sources and serve as a starting point for future research.

Feedback and Future Directions

When the Datum website was launched in September 2023, we circulated a survey to gather feedback regarding the content and accessibility of the website. User feedback suggested the utility of incorporating interactive elements into the website to increase user engagement, to allow professionals to contribute to the content library, and as ways of sharing the website with a broader audience.

Ethical Considerations

The Datum project involves the use of fictional characters in the scenarios described in the narrative-based video content. The video content is designed to highlight foundational concepts in the literature and is based on research experience. It does not recommend specific medical procedures or decisions but rather presents interactions between these fictional characters. In addition, usability testing was used to gather feedback to improve the content of the Datum website. Survey responses were not identifiable. No personal information was collected from participants to ensure privacy and confidentiality.

Discussion

Contribution

The Datum project makes an original contribution to the field of law and ethics in health science research. The primary contribution of the project involved the development of literacy related to digital health data through accessible multimedia content and a focused information repository. The presentation of multimedia content transforms traditional privacy practices in the medical sphere into learning opportunities for multiple stakeholders. Digital communication tools, such as mobile health apps, telemedicine, and web-based health information resources have demonstrated a significant potential to improve health and digital literacy, ultimately leading to better health outcomes [16].

Building on the idea that digital communication tools can effectively educate and empower health care users [16], the Datum project particularly seeks to address the data-related concerns of health care users engaged in decision-making processes within the health care context. Users may be unaware of the legislative and institutional measures that protect their digital health data or the various potential uses of their data, whether at the individual or collective levels. Indeed, findings from a survey conducted by Canada Health Infoway, entitled "What Canadians Think: Exploring Public Trust in a Digital Data Ecosystem," revealed that only 3 in 10 respondents had an awareness of Canadian privacy laws protecting their personal health information. Moreover, nearly all participants indicated that, before consenting to the use of their personal health information, they wanted to be informed of how their personal health information would be used and who would have access to it. Surprisingly, a majority of participants expressed readiness to share their personal health information for secondary purposes [17].

Further evidence supporting the promotion of digital literacy in health data, for patients to become better consumers of health care, is rapidly emerging. Informed and activated patients may effectively facilitate positive health outcomes at a lower cost [18], while digital health literacy has also been found to have a significant potential to improve health outcomes, bridge the digital divide, and reduce health inequities [19]. Studies have shown that increasing the numbers of individuals who use professional health websites is an effective measure for improving digital health literacy [20]. Additionally, research has indicated that the better their digital health literacy skills, the more individuals are able to function in the digital environment, exchange data with care providers, and actively participate in the co-design and co-delivery of health services [21]. Our rapidly evolving digital environment underlines the importance of early education in supporting the promotion of digital health literacy as an integral part of school curricula [22]. Finally, research shows that the engagement of health care service users as partners in identifying health research and service improvement priorities has led to the optimization of patient-centered health services [23].

Limitations

Certain limitations in the Datum project should be acknowledged, especially considering the rapid rate of digital transformation. For example, the project content has yet to touch on the impact of using artificial intelligence for digital health data and challenges that the use of artificial intelligence in data modeling may create. As well, the content of the Datum project does not address the concept of data ownership, suggesting future directions for the project.

Conclusions

The evidence from various studies supports the importance of promoting digital literacy concerning health data. The empowerment of health data users is achieved through improved health outcomes, knowledge on how to bridge the digital divide,

XSL•FO
RenderX

active participation in health services, and the optimization of patient-centered care. We hope that the Datum project will empower patients to become better consumers of health care services and research.

Conflicts of Interest

None declared.

References

1. Sekercioglu F, Hamid S. Ontario's digital health vision in the post-COVID-19 pandemic era: a Canadian perspective. J Int Health Sci Manag 2023;9(17):15-22. [doi: 10.48121/jihsam.1132918]
2. Paré G, Leaver C, Bourget C. Diffusion of the digital health self-tracking movement in Canada: results of a national survey. J Med Internet Res 2018 May 2;20(5):e177. [doi: 10.2196/jmir.9388] [Medline: 29720359]
3. Gasser U, Ienca M, Scheibner J, Sleigh J, Vayena E. Digital tools against COVID-19: taxonomy, ethical challenges, and navigation aid. Lancet Digit Health 2020 Aug;2(8):e425-e434. [doi: 10.1016/S2589-7500(20)30137-0] [Medline: 32835200]
4. Abernethy A, Adams L, Barrett M, et al. The promise of digital health: then, now, and the future. NAM Perspect 2022;2022. [doi: 10.31478/202206e] [Medline: 36177208]
5. Walshe K, Rundall TG. Evidence-based management: from theory to practice in health care. Milbank Q 2001;79(3):429-457. [doi: 10.1111/1468-0009.00214] [Medline: 11565163]
6. Ozomaro U, Wahlestedt C, Nemeroff CB. Personalized medicine in psychiatry: problems and promises. BMC Med 2013 May 16;11:132. [doi: 10.1186/1741-7015-11-132] [Medline: 23680237]
7. Hsin H, Fromer M, Peterson B, et al. Transforming psychiatry into data-driven medicine with digital measurement tools. NPJ Digit Med 2018;1:37. [doi: 10.1038/s41746-018-0046-0] [Medline: 31304319]
8. Hulsen T, Jamuar SS, Moody AR, et al. From big data to precision medicine. Front Med (Lausanne) 2019;6:34. [doi: 10.3389/fmed.2019.00034] [Medline: 30881956]
9. Complete guide to GDPR compliance. General Data Protection Regulation (GDPR) EUE. URL: https://gdpr.eu [accessed 2024-12-05]
10. Principaux changements aux lois sur la protection des renseignements personnels. Comission d'accès à l'information (CAI). 2024. URL: https://www.cai.gouv.qc.ca/protection-renseignements-personnels/sujets-et-domaines-dinteret/principaux-changements-loi-25 [accessed 2024-12-05]
11. Wissinger CL. Privacy literacy: from theory to practice. Comm Info Lit 2017;11(2):378-389 [FREE Full text] [doi: 10.15760/comminfolit.2017.11.2.9]
12. Viljoen S. A relational theory of data governance. Yale Law J 2021;131(2):573-654 [FREE Full text]
13. Austin EJ, LeRouge C, Lee JR, et al. A learning health systems approach to integrating electronic patient-reported outcomes across the health care organization. Learn Health Syst 2021 Oct;5(4):e10263. [doi: 10.1002/lrh2.10263] [Medline: 34667879]
14. Ferrari M, Iyer S, LeBlanc A, Roy MA, Abdel-Baki A. A rapid-learning health system to support implementation of early intervention services for psychosis in Quebec, Canada: protocol. JMIR Res Protoc 2022 Jul 19;11(7):e37346. [doi: 10.2196/37346] [Medline: 35852849]
15. Ludic Mind Studio. Datum: developing literacy about digital health data. McGill University. 2024. URL: https://www.mcgill.ca/datum [accessed 2024-12-05]
16. Fitzpatrick PJ. Improving health literacy using the power of digital communications to achieve better health outcomes for patients and practitioners. Front Digit Health 2023;5:1264780. [doi: 10.3389/fdgth.2023.1264780] [Medline: 38046643]
17. Understanding canadians' attitudes, utilization, perceptions and expectations around digital health privacy. Canada Health Infoway. 2024. URL: https://insights.infoway-inforoute.ca/digital-health-privacy-survey [accessed 2024-12-05]
18. Batalden M, Batalden P, Margolis P, et al. Coproduction of healthcare service. BMJ Qual Saf 2016 Jul;25(7):509-517. [doi: 10.1136/bmjqs-2015-004315] [Medline: 26376674]
19. Yang K, Hu Y, Qi H. Digital health literacy: bibliometric analysis. J Med Internet Res 2022 Jul 6;24(7):e35816. [doi: 10.2196/35816] [Medline: 35793141]
20. Liu S, Zhao H, Fu J, et al. Current status and influencing factors of digital health literacy among community-dwelling older adults in Southwest China: a cross-sectional study. BMC Public Health 2022 May 17;22(1):996. [doi: 10.1186/s12889-022-13378-4] [Medline: 35581565]
21. Palumbo R, Nicola C, Adinolfi P. Addressing health literacy in the digital domain: insights from a literature review. K 2022 Dec 19;51(13):82-97. [doi: 10.1108/K-07-2021-0547]
22. Reddy P, Sharma B, Chaudhary K. Digital literacy: A review of literature. International Journal of Technoethics (IJT) 2020 Jul;11(2):65-94. [doi: 10.4018/IJT.20200701.oa1]
23. Murphy F, Noonan MM, Munro D. Conversations and action; combining world cafés with experience-based-co-design to support women to breastfeed. ARCH Women Health Care 2019 Feb 13;2(1):1-4. [doi: 10.31038/AWHC.2019213]

XSL•FO
RenderX

Edited by A Mavragani; submitted 23.05.24; peer-reviewed by B Sterner; revised version received 21.10.24; accepted 28.10.24; published 03.01.25.

Please cite as:
Powell D, Asad L, Zavaglia E, Ferrari M
Promoting Digital Health Data Literacy: The Datum Project
JMIR Form Res 2025;9:e60832
URL: https://formative.jmir.org/2025/1/e60832
doi:10.2196/60832

© Daniel Powell, Laiba Asad, Elissa Zavaglia, Manuela Ferrari. Originally published in JMIR Formative Research (https://formative.jmir.org), 3.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Development of the Big Ten Academic Alliance Collaborative for Women in Medicine and Biomedical Science: “We Built the Airplane While Flying It”

Maya S Iyer[1,2*], MD, MEd; Aubrey Moe[3*], PhD; Susan Massick[4*], MD; Jessica Davis[5*], BS; Megan Ballinger[6*], PhD; Kristy Townsend[7*], PhD

[1]Department of Pediatrics, The Ohio State University College of Medicine, 420 Hamilton Hall, 1645 Neil Avenue, Columbus, OH, United States
[2]Nationwide Children's Hospital, Columbus, OH, United States
[3]Department of Psychiatry and Behavioral Health, The Ohio State University College of Medicine, Columbus, OH, United States
[4]Department of Dermatology, The Ohio State University College of Medicine, Columbus, OH, United States
[5]Office of Faculty Affairs, The Ohio State University College of Medicine, Columbus, OH, United States
[6]Department of Internal Medicine, The Ohio State University College of Medicine, Columbus, OH, United States
[7]Department of Neurosurgery, The Ohio State University College of Medicine, Columbus, OH, United States
[*]all authors contributed equally

Corresponding Author:
Maya S Iyer, MD, MEd
Department of Pediatrics, The Ohio State University College of Medicine, 420 Hamilton Hall, 1645 Neil Avenue, Columbus, OH, United States

Abstract

Women-identifying and women+ gender faculty (hereto described as women+ faculty) face numerous barriers to career advancement in medicine and biomedical sciences. Despite accumulating evidence that career development programming for women+ is critical for professional advancement and well-being, accessibility of these programs is generally limited to small cohorts, only offered to specific disciplines, or otherwise entirely unavailable. Opportunities for additional, targeted career development activities are imperative in developing and retaining women+ faculty. Our goal was the development of a new collaborative of Big Ten Academic Alliance (BTAA) institutions to support gender equity for women+ faculty in medicine and biomedical sciences, with two initial aims: (1) hosting an inaugural conference and establishing a foundation for rotation of conference hosts across BTAA schools, and (2) creating an infrastructure to develop programming, share resources, conduct environmental scans, and promote networking. In 2022, leaders from The Ohio State University College of Medicine Women in Medicine and Science envisioned, developed, and implemented a collaborative named CommUNITYten: The Big Ten Academic Alliance for Women in Medicine and Biomedical Science. Conference program development occurred through an iterative and collaborative process across external and internal task forces alongside industry partners. We developed a fiscal model to guide registration fees, budget tracking, and solicitation of conference funding from academic and industry sponsors. Attendees completed postconference surveys assessing speaker or workshop effectiveness and suggestions for future events. Finally, we developed an environmental scan survey to assess gender equity needs and existing programming across BTAA institutions. In June 2024, The Ohio State University hosted the inaugural CommUNITYten conference in Columbus, Ohio, featuring 5 keynote presentations, 9 breakout sessions, and networking opportunities across one and a half days of curated programming. Nearly 180 people attended, with representation from 9 BTAA institutions, 6 industry companies, staff, and trainees. Postconference surveys showed 50% (n=27) of respondents were likely to attend another in-person conference and suggested future conference topics. The environmental scan survey launched in October 2024. We successfully established the CommUNITYten collaborative and hosted the inaugural conference. Establishing key stakeholders from each BTAA institution, obtaining sponsorship, and detailed conference planning and partnerships were critical in ensuring realization of this collaborative. The conference brought together leaders, faculty, staff, trainees, and industry partners from across the country and met the initial goal of networking, sharing resources, and building community for women+ faculty. These efforts lay a robust foundation for the BTAA CommUNITYten collaborative to foster ongoing collaboration, innovation, and progress in the years to come. Given the importance of steady improvements, this viewpoint may further guide the efforts of other individuals, groups, and leadership supporting women+ as they consider approaches and strategies advocating for gender equity at the national level.

(JMIR Form Res 2025;9:e65561) doi:10.2196/65561

XSL•FO
RenderX

KEYWORDS

collaborative; gender equity; women in medicine; women in science; biomedical science; women; women+; gender; medicine; university; faculty; accessibility; career; equity; networking; opportunity; retaining; programming; Big Ten Academic Alliance; BTAA; academic alliance

Introduction

Women-identifying and women+ gender faculty (women+ is inclusive of different identities pertaining to gender, hereto described as women+ faculty) in medicine and biomedical sciences face numerous barriers to career advancement. These barriers include fewer invitations to keynote speaking engagements, reduced academic promotion rates, decreased leadership opportunities, and disparities in research funding and other research metrics [1-4]. For women with intersectional identities, these barriers multiply [5]. There are few prominent medicine-based national career development programs in the United States targeted to women, such as the Association of American Medical Colleges (AAMC) Early-Career and Mid-Career programs, the Hedwig van Ameringen Executive Leadership in Academic Medicine (ELAM) program, the University of Michigan's Rudi Ansbacher Advancing Women in Academic Medicine Leadership Scholars Program, and Harvard's Career Advancement and Leadership Skills for Women, with metrics demonstrating postparticipation career success and advancement [6-8]. Women who participate in these career development programs are more likely than men and nonparticipant women to be promoted to associate professor, and as likely as men and more likely than nonparticipant women to be promoted to full professor within 10 years [9]. Career development program participation also improves retention of women in the academic workforce and leadership opportunities, while expanding participants' networks and increasing visibility, both internally and externally at their institutions [10].

However, career development programs such as those from the AAMC and ELAM can be cost prohibitive and are often limited to one to two participants per institution [10]. Additionally, these programs do not accommodate faculty in medicine-adjacent fields such as biomedical engineering, veterinary medicine, and nursing. Attendance at shorter duration gender-specific conferences for women broadly in the medical and biomedical fields may provide similar benefits through delivery of content, networking, and professional support that are needed for career advancement [11]. Moreover, bridging the divide among health-related professions and bringing together women+ faculty across these fields may create a more effective cohort of professionals who can work synergistically to advance gender equity across academic institutions.

In 2022, leaders (all authors) from The Ohio State University College of Medicine's (OSUCOM's) Women in Medicine and Science (WIMS) created the Big Ten Academic Alliance (BTAA) collaborative for women in medicine and biomedical sciences with the goal of developing a network of women+ faculty in the BTAA to collaborate, share resources, provide support, and connect in a community that could be stronger together in combating gender barriers in these disciplines. Herein, we detail how we, as the leaders of this collaborative, developed a group of women+ individuals and respective institutions across the BTAA; how we crafted the inaugural conference; and how we are actively conducting an environmental scan survey and planning continuity of programming through conference host handoff, sharing of materials, and webinars.

Collaborative Development

Background and Overview

As a member of the Big Ten Conference, The Ohio State University (OSU) felt as if it was the natural place to start planning a collaborative to bring together peer institutions across the BTAA, an academic outcropping of this sports network. The BTAA was founded in 1958 [12]. Other BTAA conferences already exist in disciplines such as neuroscience; lipids; music education; women+ in technology; and Women in Science, Technology, Engineering, and Mathematics. The BTAA integrates across an existing network of deans, with central coordination and support in Rosemont, Illinois. The BTAA currently includes 14 universities. We chose to cast a broad net for the collaborative development and included institutions from all 19 prior, current, and future universities in the BTAA (Table 1). This BTAA collaborative of women in medicine and biomedical sciences would enable similar-sized universities to share effective programming and resources and work together to address academic gender inequity.

XSL•FO
RenderX

Table . The BTAA[a] institutions.

	Member of BTAA	Representatives on ETF (n)[b,c]
Indiana University	✓	4
Michigan State University	✓	3
Northwestern University	✓	3
Pennsylvania State University	✓	3
Purdue University	✓	2
Rutgers University	✓	1
The Ohio State University	✓	4
University of California, Los Angeles	✓[d]	1
University of Maryland	✓	2
University of Nebraska-Lincoln	✓	3
University of Chicago	✓[c]	1
University of Illinois	✓	2
University of Iowa	✓	1
University of Michigan	✓	2
University of Minnesota	✓	2
University of Oregon	✓[d]	0
University of Southern California	✓[d]	1
University of Washington	✓[d]	1
University of Wisconsin-Madison	✓	1

[a]BTAA: Big Ten Academic Alliance.

[b]ETF: external task force.

[c]Historical member.

[d]New BTAA member.

As we generated the vision and mission for this initiative, we benefited by learning from prior speakers hosted by OSUCOM WIMS, as well as from internal advisors. Key suggestions from our advisors were to ensure early discussions with planning partners, include broad representation of organizing bodies, gather data that could be published as part of the initiative, and listen to key stakeholders. We differentiated ourselves from existing groups, such as the BTAA Women in Science, Technology, Engineering, and Mathematics and the AAMC Group on Women in Medicine and Science (GWIMS) by focusing on support of gender equity for faculty across medicine and biomedical sciences, with additional shared resources and interactions among trainees, staff, and relevant industry members.

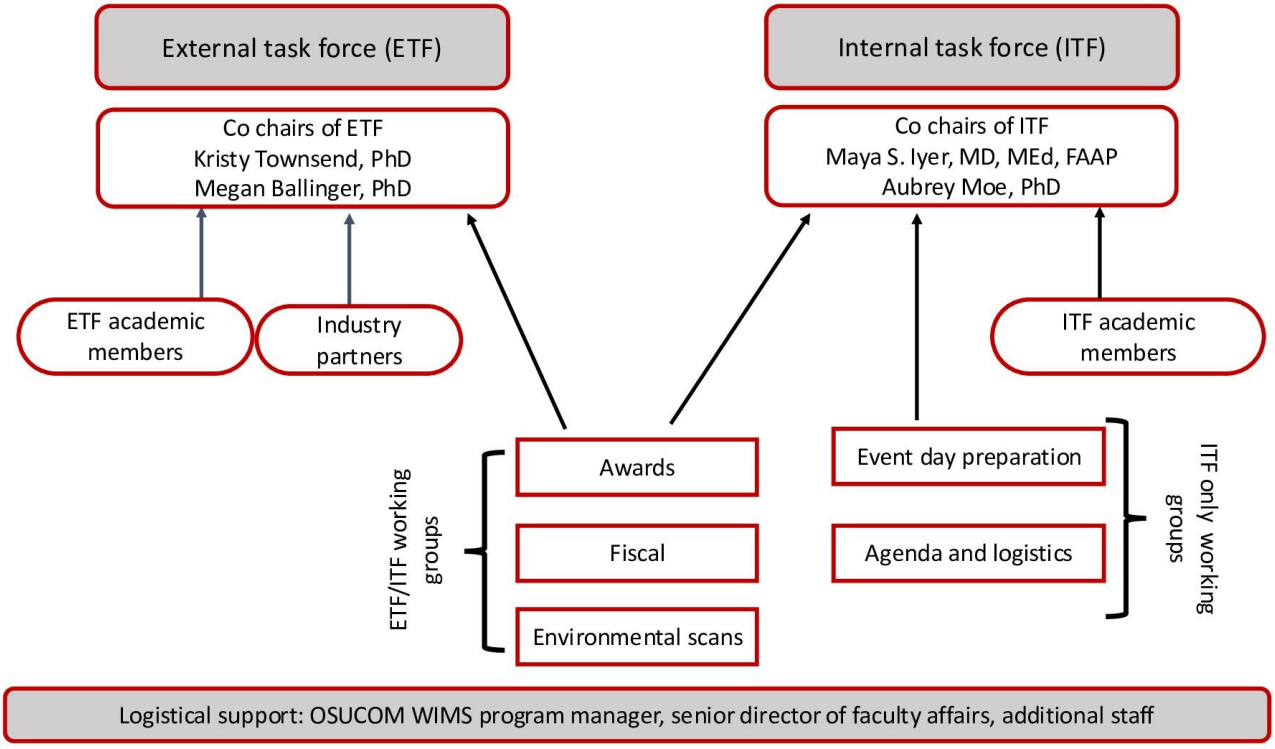
We also envisioned that once the CommUNITYten collaborative was formed, this group could work together on initiatives that would benefit the BTAA's future growth toward gender equity. Given the persistent national trend of women being promoted to full professor less frequently than men [1,13], we hope that this collaborative will provide future opportunities for networking, obtaining external talk and letters of evaluation for promotion, while also cultivating opportunities for networking, mentorship, collaboration, and sponsorship. Finally, we prioritized inclusivity and cultivating a partnership of interested industry companies, allies, staff, and leaders. Women faculty cannot solve the problems of gender inequity alone; we need the involvement and support of the entire community.

To solicit membership to this collaborative's leadership group, OSUCOM WIMS reached out to faculty representatives of the AAMC GWIMS to obtain contacts at the 19 BTAA institutions. Not all BTAA schools are part of GWIMS, so we also searched publicly available web pages dedicated to supporting women faculty at these institutions to find individuals who may be interested in planning the inaugural conference. OSUCOM WIMS also hosted a lunch roundtable at the 2022 AAMC Lead, Serve, Inspire annual meeting to gather additional participants [14]. In addition, we also relied on personal networks and connections to identify faculty members involved in initiatives that would align with the conference objectives. We were ultimately able to obtain at least one name and current contact information from 18 of the 19 BTAA institutions (Table 1). We included these individuals in all future email correspondence and meeting invitations. We also engaged with four industry representatives to bolster the networking component of this collaborative. These individuals together formed our external task force (ETF). We also created an internal task force (ITF)

XSL•FO
RenderX

of faculty from OSU for development and planning for the inaugural conference. Figure 1 displays the CommUNITYten organizational chart.

Figure 1. Reporting structure of BTAA CommUNITYten collaborative. BTAA: Big Ten Academic Alliance; ETF: external task force; ITF: internal task force.



About ETF

The ETF began meeting by Zoom (Zoom Communications, Qumu Corporation) in June 2023. The ETF was led by two coauthors (MB and KT). During these calls, we refined the collaborative's mission, solicited speaker votes for the initial conference, and discussed and voted on logistics such as registration fees, the cadence of ETF meetings, solicitation of sponsorships, and communication tactics.

Industry Partners

In many cases, industry companies in pharma, biotech, or medicine are ahead of academics in their diversity, inclusion, and equity efforts. As we spoke with our four initial industry planning partners (Abbott, Bayer, Entrada, and Novartis), we learned that their efforts to educate on neurodiversity and to support women employee groups represented opportunities for academia to learn from industry. Furthermore, since academic-industry partnerships are key to successful academic medicine and biomedical research, and since these relationships can often be informal and dominated by men, we wanted to include industry attendees to bolster the networking goals of the conference. Industry could also benefit from learning about resources and best practices for gender equity in academic settings, and also benefit from networking opportunities for recruitment of potential employees. As part of our partnership with industry, we planned to have one panel discussion at the inaugural conference to learn more about industry equity initiatives, career opportunities, and translational research across academic-industry partnerships.

About ITF

An ITF was specifically established to plan for the inaugural CommUNITYten conference and to execute key deliverables and tasks related to event logistics. This task force was led by two coauthors (MSI and AM). Given that OSUCOM WIMS led the establishment of the CommUNITYten collaborative, the ITF members were solicited from existing WIMS members, but invites were also extended to working groups from our academic sponsors and ETF members. The ITF met monthly and was further divided into working groups. This approach permitted work delegation to specific individuals and allowed for identification of specific leaders for each working group, providing leadership opportunities. A description of the five working groups is below. The ITF also worked closely with the WIMS' program coordinator (coauthor JD), OSUCOM staff, and marketing teams.

ITF Working Groups

Agenda and Logistics

This working group focused on aspects of agenda planning, speaker communications, and coordination of timing for multiple tracks of speakers, seminars, and events during the day-and-a-half conference. Engagement and ongoing coordination with other working groups (fiscal and event day preparation) were critical to maintaining alignment of the agenda with CommUNITYten's goals and budgetary resources.

XSL•FO
RenderX

Event Day Preparation

This working group oversaw event logistics and worked directly under the WIMS program coordinator (coauthor JD). Specific items managed by this group included coordination of speaker transportation, audio or visual access and support, creation of program materials (including printed products and conference swag) and providing conference day assistance to attendees.

ITF and ETF Joint Working Groups

Awards

This group was led by two OSUCOM WIMS faculty members and had two additional BTAA members. Given the underrecognition of women recipients among societal and specialty-based awards [15-17], this working group intentionally developed three awards for the inaugural conference to recognize trainees and faculty at all academic ranks: (1) catalyst award, (2) rising star, and (3) luminary. The catalyst award is granted to a trainee (postdoctoral fellow, clinical fellow, or resident) who has demonstrated excellence, drive, and dedication in promoting women+ individuals in medicine and biomedical science or research. The rising star award recognizes early- or midcareer faculty (within 10 y of a full-time faculty position or rank of assistant professor) who have similar advanced women in their fields. The luminary award recognizes senior faculty members who have substantially contributed to the promotion and advancement of women+ individuals within or external to the nominee's institution. Through an iterative process, this group developed these award descriptions, announcements, and scoring rubrics. These awards were presented at the inaugural conference.

Fiscal

This working group developed the fiscal model to support the CommUNITYten conference and was coled by OSUCOM's senior director of faculty affairs and coauthor KT in conjunction with the entire ETF. Starting with an initial US $20,000 investment from OSUCOM WIMS, the fiscal working group developed a registration model to generate needed revenue (with a fee schedule based on input and votes from the ETF), and solicited academic and industry sponsorships. Fundraising for the inaugural CommUNITYten conference involved a 3-pronged approach: (1) internal solicitation of host institution's colleges and centers or institutes; (2) external solicitation of industry, nonprofit, or community organizations; and (3) academic sponsorship from other BTAA schools. Together, the fundraising from registrations and sponsorships would form the available budget for conference. Since we were fundraising as we were building the conference program, there was an iterative process whereby later fiduciary decisions (catering choices, for example) were dictated closer to the conference date and based on projections of confirmed incoming funds. By June 2024, we had raised approximately US $130,000 for the CommUNITYten collaborative.

To solicit sponsors, we generated a sponsor package that included an overview of the intent of the meeting and what would be gained by sponsoring across several options of sponsorship tiers (Multimedia Appendix 1). An Excel (Microsoft Corp) file was maintained to track sponsor calls, verbal commitments, and follow-up correspondence, or the "income" for the meeting. This file was compared to an Excel outline of expected expenditures (which would become a "budget," once figures were finalized). The goal was to increase income above the numbers in the projected budget, in time for the deadline to finalize space and catering contracts. The last component of this effort was generating a dedicated account at OSU to hold funds separate from the college operating budget. This would allow for easy transfer of funds to the next host institution. We prioritized transparency in our efforts, with detailed recordkeeping and fiscal responsibility that was shared with the ETF. As an example of fiscal responsibility, our initial keynote choice had a prohibitively high speaker fee, and we now hope to have this speaker for a webinar in the future using remaining carryover funds now that the inaugural conference is over.

Environmental Scan Surveys

This working group was led by two coauthors (MB and SM) in conjunction with the entire ETF. The group met monthly for 6 months to develop an institutional review board–approved research study protocol to ascertain current trends and faculty experiences across the BTAA institutions related to gender equity in medicine and biomedical science. The purpose for conducting a multi-institutional survey was to ask questions specifically on workplace climate, career promotion, leadership development, discrimination or bias, retention, mentorship, sponsorship, scholarship, pay equity, care responsibilities, and lactation or reproductive care services. This survey was piloted among non-BTAA faculty members to ensure understanding. The survey was launched in October 2024 and sent to conference participants or registrants, members of the ETF and other BTAA faculty contacts. Responses will be analyzed, and aggregated data will be shared among the participating BTAA institutions. Survey findings may identify opportunities for improvement, strategic interventions, and targeted programming that can benefit faculty across disciplines and shape programming initiatives for upcoming BTAA CommUNITYten conferences, future CommUNITYten collaborative work, and potentially a partnership with the BTAA and AAMC GWIMS going forward.

CommUNITYten 2024 Inaugural Conference

The inaugural "CommUNITYten: Big Ten Academic Alliance Conference for Women in Medicine and Biomedical Science" was held on Friday June 7 and Saturday June 8 at the Ohio Union in Columbus, Ohio. Day 1 had 171 attendees and day 2 had 78 attendees, out of 220 total registrants. The final agenda for the conference is listed in Multimedia Appendix 2. The conference included 5 keynote presentations, 9 breakout sessions, coffee breaks with interactive activities (mindfulness, conversations surrounding people of difference, and a session with our Buckeye Paws service dogs), and a networking event on the first evening of the conference. Breakout sessions included presentations on topics relevant to the collaborative mission (eg, gender equity, developing a professional narrative, and effective science communication) as well as panel events on publishing, allyship from men, and using social media as a

XSL•FO
RenderX

tool to establish professional presence. Finally, we awarded 1 catalyst, 1 luminary, and 2 rising star awards.

We conducted postconference evaluations (N=54 respondents). Among the respondents, on a 5-point Likert scale, 50% (n=27) reported that they would be likely to attend a future in-person BTAA CommUNITYten conference. In addition, 50% (n=27) stated that they would be likely to attend a webinar series offered by BTAA CommUNITYten. Nearly 60% (n=31) of respondents heard about the conference from a direct invitation. Open-ended comments showed that future topics should include negotiation; setting boundaries; navigating career changes; lesbian, gay, bisexual, transgender, or queer issues; creating networks; and clinical and translational research. Areas for improvement included shortening the conference to a one-day event, going beyond cis-normative topics, and including more time for networking.

Challenges and Lessons Learned

The initial challenge in creating the CommUNITYten collaborative was "building the plane while flying it." Our fiscal model was dependent on funds generated from registration costs, sponsorships, and the seed money from OSUCOM WIMS—all of which were unknowns when we initially embarked on this process. As a result, we had to establish our incoming funds flow simultaneously as we solicited speakers and panelists. We could not build a website until we had a set agenda. In addition, communicating with our ETF was challenging at the tail end of the academic year due to conflicts with scheduled meetings; loss of contacts (ie, individuals leaving their institutions due to retirement and promotions); and other academic obligations such as commencements, other conferences, and personal vacations.

We were fortunate to have the OSUCOM WIMS program coordinator (coauthor JD) serve as a project manager across all task forces and working groups. However, we recommend hosts of similar events bring on a full-time equivalent to support a host institution in planning a conference of this scale. Textbox 1 provides information on all the aspects of the conference that our program coordinator oversaw, including tasks carried out with our coleads, working group members, and support staff. Table 2 presents conference attendees' institutional affiliations.

XSL•FO
RenderX

Textbox 1. Conference planning and execution logistics. Logistical items listed were largely carried out by The Ohio State University College of Medicine (OSUCOM) Women in Medicine and Science's program coordinator for the inaugural CommUNITYten conference, with some items carried out by coleads, OSUCOM support staff, or working group members.

Conference planning

- Create an online file system (Microsoft Teams) to share information across internal and external partners.
- Schedule internal task force (ITF), external task force (ETF), marketing, and coleads meetings; send out reminder emails, meeting agendas, meeting notes, and postmeeting tasks.
- Research and secure space and catering orders.
- Preview facility or venue and work with venue event planner to confirm number of rooms, room layouts, technology needs, day of logistics, and build a timeline to meet venue deadlines.
- Collaborate with coleads and ETF and ITF on go/no go decisions such as event date or times, registration platform and wording, continuing medical education, fiscal model, Big Ten Academic Alliance (BTAA) logo usage, etc.
- Advertise through numerous channels, such as emails, campus-wide university newsletter, OSUCOM newsletters, and Twitter. Externally, the coleads advertised the conference through the Executive Leadership in Academic Medicine EDGE newsletter.
- Work with venue IT team for audio or video needs.
- Decide on and secure decorations (photo backdrop, centerpieces, or stage décor).
- Secure photographer.

Fiscal

- Schedule meetings for Women in Medicine and Science director and associate director to meet with The Ohio State University deans, center directors, and other leaders to request sponsorship.
- Meet with coleads and fiscal working group leads to review finances.
- Secure sponsor information: logo, marketing contact, fiscal contact, and contact for dissemination of information post conference.

Speakers or agenda

- Collate and track list of potential speakers, as determined by ETF, across price ranges and topics.
- Follow up on emails from coleads to potential speakers asking and securing availability, interest, and fees.
- Develop speaker intake forms (name, credentials, talk title and objectives, contact info, IT needs, travel needs, and payment forms).
- Track speaker information from intake through postevent payment process.
- Work with speakers on contracts and send to legal for approval.
- Set up logistics meetings with panelist groups and their moderators to understand their needs and desired speaker introductions.
- Set up logistics meetings with speakers and session presenters on talk time, topic, format, and logistics.
- Book speaker travel (lodging, flights, and transportation to the venue) and create individual speaker itineraries.
- Finalize agenda.
- Create day of speaker itineraries including travel, meetings, green room locations, and arrange shuttles or transport to and from the hotel to venue.
- Collect items for speaker gift bags to bring to the conference.
- Process speaker payments and honorariums, including receipts from speakers and create expense reports for travel reimbursement.

Marketing

- Create a comprehensive communication plan: Work with host institution marketing teams to create marketing language, create save the date PowerPoint (Microsoft Corp) slides, flyers, and email templates.
- Once approval is received from BTAA to use logo; work with BTAA representative to secure swag delivery before conference.
- Create registration site (Cvent) that includes deadlines, refund policy, option for single day registration versus full conference only, what information you want to gather from participants, what registration confirmation emails should look like, ability for registered participants to use an app or have a separate postregistration website, what sponsor logos will look like, what reports you will to need to monitor registration.
- Get input from coleads and event day logistics working group on aesthetic of Cvent site.
- Work with marketing team regularly to create and update QR codes, URL short links, content for social media posts, websites, and develop and pull daily reports from Cvent or conference registration website.
- Collaborate with marketing and event day preparation working group to install communications on electronic boards across institution.
- Add communications to electronic boards across institution.

XSL•FO
RenderX

- Create name tags.

Working groups (awards, environmental scan, event day preparation, fiscal, and speaker or agenda and logistics)

- Set up monthly group meetings, create or gather agendas, send reminder emails, and meeting notes and tasks.
- Develop a survey tool to capture award nominations and deadlines.
- Design and order plaques for the award recipients.
- Determine when and how to recognize awardees in event schedule with speaker or agenda working group.
- Market call for nomination emails to registered attendees and internally to participating colleagues.

Day-of-conference logistics

- Create run of show with day of assignments (décor, registration desk, speaker movement, monitor catering, audio-visual, space issues, and postevent clean-up).
- Inventory and gather supplies to take to venue (banners, signage, BTAA swag, speaker bags, BTAA notebooks, and pens).
- Create postevent surveys for feedback with coleads.
- Create rolling slide decks and master slide decks for each day, collaborate with audio-visual team at venue to ensure slide decks run smoothly.

Postconference

- Create and send thank you emails to speakers with coleads.
- Distribute postevent surveys for feedback for those that may have left early.
- Process venue or service-provider payments.
- Set up debrief meetings for ITF and ETF.
- Collate event survey data to send to coleads for analysis.

Table . Conference attendees' institutional affiliations.

Group or institution	Friday June 7, 2024 (n)	Saturday June 8, 2024 (n)
Industry	6	0
Indiana University	3	4
Northwestern University	6	2
Pennsylvania State University	1	1
Rutgers University	2	2
The Ohio State University	134	56
University of Illinois	2	0
University of Maryland	3	3
University of Michigan	6	4
Univeristy of Nebraska	3	3
Non–Big Ten Academic Alliance institution	5	3
Total	171	78

After hosting the conference, we learned several lessons to improve future conference gatherings. First, future conferences in this series could benefit from selecting annual conference titles and themes to align content around specific goals. Many speakers expressed a desire for this information to tailor their talks to align with the conference's intentions and goals beyond the broad gender equity focus. In addition, we learned the importance of engaging the institutional marketing team from the outset. Their support is crucial for successful outreach, within the institution and to external networks via LinkedIn. Before launching the registration platform, having a complete program agenda and speaker bios helps convert website visitors to registrants when they can see the full conference content. We also learned that shortening the conference to 1 day and avoiding programming on a weekend could boost attendance. Finally, if possible, we found that a 2-year timeline is needed to secure sponsorships and plan logistics associated with the in-person meeting.

XSL•FO
RenderX

Future Directions

One of the pillars in the development of this conference was to establish a network of like-minded women+ interested in bolstering gender equity in medicine and biomedical science. In order to provide equitable leadership opportunities and accessibility of the in-person conference geographically, the ETF decided that the host institution would rotate every other year for in-person conferences. In the interim nonconference years, a virtual webinar series will continue the momentum and the networks established by this collaborative, with allocated budget voted on by the ETF and ITF from our rollover funds. We anticipate that each BTAA institution will be responsible for sponsoring a speaker for the interim year webinars. Additionally, we hope that results from the environmental scan, once completed, will be presented in upcoming meetings or publications and used as a needs assessment for future CommUNITYten conferences in determining conference themes, workshops, sessions, and speakers.

In summary**,** the BTAA CommUNITYten collaborative successfully launched in 2024. Establishing key stakeholders from each BTAA institution, obtaining sponsorship, and detailed conference planning and partnerships were critical in ensuring realization of this group. The inaugural conference brought together leaders, faculty, staff, trainees, and industry partners from across the country and met the initial goal of networking and developing interinstitutional community. Our future initiatives of conducting environmental scans and sharing resources are in progress, as well as development of a national webinar series. These efforts lay a robust foundation for the BTAA CommUNITYten collaborative to foster ongoing collaboration, innovation, and progress in the years to come.

Acknowledgments

We would like to thank Dean Carol Bradford, MD, MS, and Vice Dean Tatiana Oberyszyn, PhD, for their unwavering support of this initiative. We would also like to thank Kathy Tober, PhD, for her logistical support of the conference, for coleading the Fiscal Task Force, and for overseeing all financial aspects of the initial event. In addition, we would like to thank Alex Hildebrandt, MA, Kim Knight, BA, and the members of the College of Medicine Marketing Team for their excellent support of this event. The coauthors wish to thank all members of the internal task force, external task force, and working groups that made the inaugural conference successful. Finally, we acknowledge the key contributions of our inaugural sponsors. Generative artificial intelligence was not used in any portion of writing this paper.

Authors' Contributions

MSI did the conceptualization, data curation, formal analysis, funding acquisition, methodology, supervision, writing of the original draft, and reviewing and editing of the writing. AM handled writing of the original draft, and reviewing and editing of the writing. SM handled the reviewing and editing of the writing. JD worked on the project administration, writing of the original draft, and reviewing and editing of the writing. MB wrote the original draft and reviewed and edited the writing. KT assisted in conceptualization, data curation, formal analysis, funding acquisition, methodology, supervision, writing of the original draft, and reviewing and editing of the writing. MSI and AM served as co-first authors and MB and KT served as co-senior authors for this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sponsorship tiers.
[DOCX File, 21 KB - formative_v9i1e65561_app1.docx]

Multimedia Appendix 2
Conference agenda.
[PDF File, 7174 KB - formative_v9i1e65561_app2.pdf]

References

1. Richter KP, Clark L, Wick JA, et al. Women physicians and promotion in academic medicine. N Engl J Med 2020 Nov 26;383(22):2148-2157. [doi: 10.1056/NEJMsa1916935] [Medline: 33252871]
2. Carr PL, Raj A, Kaplan SE, Terrin N, Breeze JL, Freund KM. Gender differences in academic medicine: retention, rank, and leadership comparisons from the National Faculty Survey. Acad Med 2018 Nov;93(11):1694-1699. [doi: 10.1097/ACM.0000000000002146] [Medline: 29384751]
3. Amrein K, Langmann A, Fahrleitner-Pammer A, Pieber TR, Zollner-Schwetz I. Women underrepresented on editorial boards of 60 major medical journals. Gend Med 2011 Dec;8(6):378-387. [doi: 10.1016/j.genm.2011.10.007] [Medline: 22153882]

XSL•FO
RenderX

4. Marcelin JR, Bares SH, Fadul N. Improved infectious diseases physician compensation but continued disparities for women and underrepresented minorities. Open Forum Infect Dis 2019 Feb;6(2):ofz042. [doi: 10.1093/ofid/ofz042] [Medline: 30815507]
5. Verduzco-Gutierrez M, Wescott S, Amador J, Hayes AA, Owen M, Chatterjee A. Lasting solutions for advancement of women of color. Acad Med 2022 Nov 1;97(11):1587-1591. [doi: 10.1097/ACM.0000000000004785] [Medline: 35731581]
6. Helitzer DL, Newbill SL, Morahan PS, et al. Perceptions of skill development of participants in three national career development programs for women faculty in academic medicine. Acad Med 2014 Jun;89(6):896-903. [doi: 10.1097/ACM.0000000000000251] [Medline: 24871241]
7. The Hedwig van Ameringen Executive Leadership in Academic Medicine. Drexel University College of Medicine. URL: https://drexel.edu/medicine/academics/womens-health-and-leadership/elam/ [accessed 2023-12-11]
8. Woitowich NC, Graff SL, Swaroop M, Jain S. Gender-specific conferences and symposia: a putative support structure for female physicians. J Womens Health (Larchmt) 2020 Aug;29(8):1136-1141. [doi: 10.1089/jwh.2019.7859] [Medline: 31697577]
9. Chang S, Guindani M, Morahan P, Magrane D, Newbill S, Helitzer D. Increasing promotion of women faculty in academic medicine: impact of national career development programs. J Womens Health (Larchmt) 2020 Jun;29(6):837-846. [doi: 10.1089/jwh.2019.8044] [Medline: 32466701]
10. Chang S, Morahan PS, Magrane D, et al. Retaining faculty in academic medicine: the impact of career development programs for women. J Womens Health (Larchmt) 2016 Jul;25(7):687-696. [doi: 10.1089/jwh.2015.5608] [Medline: 27058451]
11. Salem V, McDonagh J, Avis E, Eng PC, Smith S, Murphy KG. Scientific medical conferences can be easily modified to improve female inclusion: a prospective study. Lancet Diabetes Endocrinol 2021 Sep;9(9):556-559. [doi: 10.1016/S2213-8587(21)00177-7] [Medline: 34332698]
12. History of the Big Ten Alliance. Big Ten Academic Alliance. URL: https://btaa.org/about/history [accessed 2024-05-15]
13. 2018-2019: the state of women in academic medicine: exploring pathways to equity. AAMC. 2018. URL: https://www.aamc.org/data-reports/faculty-institutions/report/state-women-academic-medicine [accessed 2024-05-15]
14. Learn serve lead: the AAMC annual meeting. AAMC. URL: https://www.aamc.org/career-development/events/learn-serve-lead [accessed 2024-12-07]
15. Jacobs JW, Bibb LA, Allen ES, et al. Women and non-white people among Lasker Award recipients from 1946 to 2022: cross sectional study. BMJ 2023 May 17;381:e074968. [doi: 10.1136/bmj-2023-074968] [Medline: 37197781]
16. Neikirk K, Vue Z, Vue N, et al. Disparities in funding for Nobel Prize awards in medicine and physiology across nationalities, races, and gender. J Cell Physiol 2024 Jul;239(7):e31157. [doi: 10.1002/jcp.31157] [Medline: 38225913]
17. Silver JK, Slocum CS, Bank AM, et al. Where are the women? the underrepresentation of women physicians among recognition award recipients from medical specialty societies. PM R 2017 Aug;9(8):804-815. [doi: 10.1016/j.pmrj.2017.06.001] [Medline: 28606837]

Abbreviations

AAMC: Association of American Medical Colleges
BTAA: Big Ten Academic Alliance
ELAM: Executive Leadership in Academic Medicine
ETF: external task force
GWIMS: Group on Women in Medicine and Science
ITF: internal task force
OSU: The Ohio State University
OSUCOM: The Ohio State University College of Medicine
WIMS: Women in Medicine and Science

Edited by A Mavragani; submitted 25.08.24; peer-reviewed by D Helitzer, GS Booth; revised version received 14.11.24; accepted 14.11.24; published 23.01.25.

Please cite as:
Iyer MS, Moe A, Massick S, Davis J, Ballinger M, Townsend K
Development of the Big Ten Academic Alliance Collaborative for Women in Medicine and Biomedical Science: "We Built the Airplane While Flying It"
JMIR Form Res 2025;9:e65561
URL: https://formative.jmir.org/2025/1/e65561
doi:10.2196/65561

XSL•FO
RenderX

© Maya S Iyer, Aubrey Moe, Susan Massick, Jessica Davis, Megan Ballinger, Kristy Townsend. Originally published in JMIR Formative Research (https://formative.jmir.org), 23.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Applying Critical Discourse Analysis to Cross-Cultural Mental Health Recovery Research

Yasuhiro Kotera[1,2], PhD; Riddhi Daryanani[3], MSc; Oliver Skipper[3], MSc; Jonathan Simpson[3], MSc; Simran Takhi[1], MSc; Merly McPhilbin[1], MSc; Benjamin-Rose Ingall[1], MSc; Mariam Namasaba[3], PhD; Jessica Jepps[3], MSc; Vanessa Kellermann[3], MSc; Divya Bhandari[4], MSc; Yasutaka Ojio[5], PhD; Amy Ronaldson[3], PhD; Estefania Guerrero[6], MSc; Tesnime Jebara[3], PhD; Claire Henderson[3], PhD; Mike Slade[1,7], PhD; Sara Vilar-Lluch[8], PhD

[1]Institute of Mental Health, School of Health Sciences, University of Nottingham, Nottingham, United Kingdom
[2]Center for Infectious Disease Education and Research, Osaka University, Suita, Japan
[3]Health Service and Population Research Department, King's College London, London, United Kingdom
[4]Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, United States
[5]Department of Community Mental Health & Law, National Institute of Mental Health, Tokyo, Japan
[6]Department of Social Psychology and Quantitative Psychology, University of Barcelona, Barcelona, Spain
[7]Faculty of Nursing and Health Sciences, Nord University, Namsos, Norway
[8]School of English, University of Nottingham, Nottingham, United Kingdom

Corresponding Author:
Yasuhiro Kotera, PhD
Institute of Mental Health
School of Health Sciences
University of Nottingham
Triumph Road
Nottingham, NG7 2TU
United Kingdom
Phone: 44 01158231294
Email: yasuhiro.kotera@nottingham.ac.uk

Abstract

The purpose of this paper is to demonstrate how critical discourse analysis (CDA) frameworks can be used in cross-cultural mental health recovery research. CDA is a qualitative approach that critically appraises how language contributes to producing and reinforcing social inequalities. CDA regards linguistic productions as reflecting, consciously or unconsciously, the narrators' understandings of, or attitudes about, phenomena. Mental health recovery research aims to identify and address power differentials, making CDA a potentially relevant approach. However, CDA frameworks have not been widely applied to mental health recovery research. We adapted established CDA frameworks to our cross-cultural mental health recovery study. The adapted methodology comprises (1) selecting discourses that indicate positive changes and (2) considering sociocultural practices informed by relevant cultural characteristics identified in our previous research, without placing value judgments. Our adapted framework can support cross-cultural mental health recovery research that uses CDA.

(JMIR Form Res 2025;9:e64087) doi:10.2196/64087

KEYWORDS

critical discourse analysis; cross-cultural mental health recovery research; linguistic analysis; social inequality; mental health; recovery research; language; social inequalities; qualitative analytical approach; linguistic expressions; discourse; analysis; framework; inequalities; CDA

Critical Discourse Analysis and Mental Health

Critical discourse analysis (CDA) is a qualitative analytical approach that critically appraises how language contributes to the production and reproduction of social inequalities through the examination of authentic uses of language [1,2]. CDA considers that linguistic expressions reflect the speakers' and writers' conscious or unconscious perceptions or opinions towards phenomena [1,2]. CDA used in mental health research

XSL•FO
RenderX

is grounded in the belief that people's mental illness experiences are represented in the language they use [3]. This belief underscores the value of analyzing various forms of textual expression, such as first-person narratives and other linguistic representations. CDA can offer profound insights into the personal accounts of mental illness experiences and the societal perceptions surrounding specific mental illnesses.

The application of CDA to mental health has become an important area of research. Informed by corpus linguistics approaches (software-assisted examination of large collections of digitized texts), scholars have delved into media depictions of mental health conditions, as evidenced by Karaminis et al's [4] examination of autism portrayal in the British press. Likewise, Balfour [5] explored schizophrenia representations in the British press, while Price [6] scrutinized the perpetuation of mental health stereotypes resulting from media portrayals. The critical approach of CDA turns linguistic analysis into an engaged social enterprise; in these studies, it reveals engrained prejudices towards mental health conditions reproduced in mainstream media and how stereotypical representations have evolved over the years.

Moreover, researchers have turned their attention to the rich landscape of digital informal interactions among service users and nonprofessional caregivers. Digital forums and support groups have become significant spaces where individuals share their experiences, exchange information, and provide emotional support related to illnesses. For instance, based on corpus methods, Hunt and Brookes [7] shed light on discussions surrounding anorexia, depression, and diabulimia, while Jones et al [8] appraise the experiences of individuals with psychosis and bipolar disorder. Corpus-informed CDA is actively used to identify patterns of language use in large text-data collections, which deepens understanding of how language shapes our perceptions of mental health and illness.

First-person narratives of mental health experiences are another emerging field. This includes investigations into autobiographical accounts, such as stylistic studies on the experiences of voice hearing narrated by individuals with schizophrenia [9,10]. These stylistic studies have provided refined typologies of voices in auditory hallucinations in schizophrenia, bringing new insights into the phenomenology of the condition. Additionally, CDA studies have evaluated naturally occurring narratives, such as the examination of men's accounts of depression [11]. CDA research on naturally occurring narratives highlights the importance of lived-experience insights in gaining a more nuanced understanding of living with a mental health condition, both from individual and social perspectives. CDA studies, and discourse studies more generally, have also examined the construction of the self in interaction—eg, Meehan and MacLachlan [12] on schizophrenia, and Kłosińska and Ziółkowska [13] on dementia. Studies on self-identity construction among individuals with dementia, for example, have provided evidence that a sense of self is maintained in the later stages of dementia. This is achieved through the reliance on previous social identities and lexical repertoires to give meaning to the present [13]. A tenet of CDA is that discourse is ideological. In studying narratives of illness and recovery, this may involve examining how dominant discourses around identity (eg, gender, ethnicity) can influence the experience of illness, potentially affecting treatment and recovery (eg, normative ideologies of masculinity may underpin men's experience of depression [11]).

CDA has been employed extensively within mental health research. Nevertheless, the specific application of CDA within one area of mental health research, namely "mental health recovery research," remains largely unexplored to date [14]. In this paper, we will present one example of using established CDA frameworks in our cross-cultural mental health recovery study [15].

Mental Health Recovery

In mental health research, mental health recovery has been receiving attention [16]. Mental health recovery is "a deeply personal, unique process of changing one's attitudes, values, feelings, goals, skills, and/or roles" and "a way of living a satisfying, hopeful, and contributing life even within the limitations caused by illness" [17]. Approaches oriented towards mental health recovery, known as "recovery-oriented approaches," have been implemented in mental health services globally, and their positive impacts have been reported, such as empowerment, hope, autonomy, and reduced stigma in service users [18,19], as well as positive attitudes and understanding towards service users in mental health staff [20]. Recovery-oriented approaches are also helpful in understanding the mental health experiences of people with mental disorders such as bipolar disorders [21,22]. Owing to these benefits, mental health recovery has been embedded in many national policies [23-29].

Both CDA and mental health recovery value highlighting and addressing inequalities. CDA's fundamental principle is to identify inequalities in linguistic and social practices [30]. Similarly, mental health recovery aims to address power differentials in mental health practice and places the person living with a mental health experience at the center [31]. Mental health recovery is deeply rooted in social justice principles, empowering marginalized voices, reducing stigma and discrimination, and promoting equality [32]. This overlap between CDA and mental health recovery research suggests their compatibility, and yet the application of CDA to this field remains relatively unexplored. The current status of psychiatric discourse is diagnostic-driven rather than person-centered, marginalizing service users' lived experience rather than empowering them, biologically deficit-focused rather than strength-based, and regarding service users as passive patients rather than as individuals with unique needs, goals, and human rights [33]. Research on mental health recovery, empowered by CDA, has the potential to drive cultural changes in the current psychiatric discourse.

Our Study

Recovery colleges (RCs) are one of the recovery-oriented approaches that are actively used globally [34] and are regarded as a mental health innovation [35]. RCs are learning-based

XSL•FO
RenderX

support systems offering information, social support, and skill development for people with mental health symptoms, carers, and staff [36]. RCs are open to anyone interested in mental health and recovery, such as service users, carers, and professionals. People can directly self-enroll in courses offered at an RC. RCs are often free to use (eg, those offered by NHS) or may charge a modest fee (eg, £3 [equivalent to US$ 3.73 as of January 28, 2025] in some RCs in Japan). The duration and length of RC courses vary (eg, a one-off 2-hour session or 5 weekly 3-hour mindfulness sessions). RCs are advertised through various channels such as mental health services, digital platforms (eg, social media), community outreach (eg, flyers), and word of mouth [15,34]. Since the first RC was developed in England in 2009, there have been 221 RCs established in 28 countries within 15 years [34].

Our recent RC study was informed by CDA frameworks to be able to highlight textual emphases in the RC promotional texts in Japan and England and compared them [15]. We explored how RCs are promoted to the public in Japan and England and compared textual and thematic differences from cross-cultural perspectives. The textual emphases found in each country were in line with the cultural characteristics associated with the RC operational model [37], namely collectivism versus individualism and long-term orientation versus short-term orientation [38]. The texts in RCs in Japan emphasized collectivism (eg, "learn together") and long-term orientation (eg, highlighting the presence of difficulties now), whereas those in RCs in England emphasized Individualism (eg, "self-management") and short-term orientation (eg, focusing on skill acquisition) [15].

Specific recommendations were made for the RC operational model, which comprises 12 components [37]. The RC operational model was aligned with individualism and short-term orientation [38]. Our findings indicate some of the components may need to change in order to include under-recognized cultural characteristics such as collectivism and long-term orientation. People in Japan see the RC promotional texts that emphasize collectivism and long-term orientation and attend an RC expecting to engage with activities associated with collectivism and long-term orientation. RCs in Japan need to adapt their operations to meet those expectations, which work disadvantageously to achieve some of the components. For example, one of the components—component 3—evaluates whether an RC actively enquires about the student's individual needs. This activity is more accepted in individualism than in collectivism. In collectivism, actively enquiring about individual needs can be regarded as rude because expressing individual needs may violate group harmony [39]. Therefore, we proposed that component 3 should include collectivism to ensure RCs oriented to this cultural characteristic are included.

Critical Discourse Analysis in Recovery Research

Our cross-cultural recovery research was informed by established CDA frameworks [1,2,40]. CDA involves textual analysis [1]. In our study, we examined whether the promotional texts of 61 RCs in England and those of all the 13 RCs in Japan reflected any cultural characteristics in their descriptions of recovery and support provided. The analysis was supported by corpus linguistics tools, which involve a computer-assisted examination of digitized texts (eg, [41,42]). In the analysis, we (1) considered the contexts of production and reception [38] and (2) interpreted the results in relation to sociocultural practices (Table 1). We adapted 2 stages of traditional CDA to better fit cross-cultural mental health recovery research: "selection of the discourse" and "consideration of the sociocultural practice" (see Table 1 for the original versions and adaptations).

Since our study considered RCs, our "selection of the discourse" focused on texts highlighting positive changes rather than inequalities (ie, RCs promotional material), as informed by Bartlett's and Martin's positive approach [43,44]. Traditionally, CDA studies predominantly focus on highlighting inequalities and discriminatory practices [46]. This approach is not fully applicable to recovery research that aims not only to highlight individuals experiencing difficulty and societal injustice encompassing social, political, and rights aspects around mental health and recovery [47] but also to underscore the positive impacts on people living with mental health symptoms [19]. Many people living with mental health symptoms, and those who work with them, are well aware of inequalities [48]. In addition to recognizing inequalities, they need changes that address inequalities [35]. Therefore, we have employed a positive approach and considered textual productions aimed at promoting inclusive mental health recovery practices (ie, RCs).

The second adaptation was made when considering the relevance of textual findings for sociocultural practice. The "consideration of sociocultural practice" was informed by cross-cultural findings. In our previous study [38], we used Hofstede's cross-cultural indexes [45] to identify cultural characteristics that were associated with the RC operational model. These cultural characteristics informed how we interpreted the results of the textual analysis, which revealed that different themes were emphasized in RCs in Japan and England. Drawing our interpretation on empirical cross-cultural research also allowed us to reduce potential value judgments. Cross-cultural theories aim to *understand* different cultures, thus seeking to explain cultural differences without placing value judgments on them [49]. In highlighting inequalities reproduced in texts, many CDA studies have critically evaluated those aspects of advantaged groups that perpetuate discrimination. This explicit critical stance has led to criticisms of being biased for formulating value judgments according to a predefined agenda [50,51]. These evaluative assessments, however, can conflict with the aim of cross-cultural theories to understand different cultures. Therefore, our interpretation of textual differences was informed by empirical cross-cultural research [38] and a cross-cultural theory [45] to move away from value judgments.

The implications of adopting these 2 adaptations—"selection of the discourse" and "consideration of the sociocultural practice"—resulted in tangible recommendations for positive changes. Specific changes were proposed in the discussion section of our study paper, such as the suggestion for component 3 to include collectivism as noted above [15].

XSL•FO
RenderX

Table 1. Adapted version of the analytical framework for critical discourse analysis [1,2,40].

Stage of analysis	Description	How it is addressed in our work
Preparation (1): select the discourse	Select a discourse according to your research interests. Originally focused on addressing social injustices and inequalities, (critical) discourse analysis is also applied to discourses that promote positive social change [43,44] **Original version** Select a discourse related to injustice or inequality in society.	Selection of recovery colleges (RCs) promotional texts, a relatively new mental health support system that promotes individual empowerment and recovery through learning in community.
Preparation (2): data gathering	Select data sources, consider any ethical implications involved in data gathering, and prepare the data for analysis.	Identification of RCs in England and Japan, and retrieval of relevant descriptions from their information websites, notably focusing on those presenting RCs and recovery to the public. Translation of Japan RCs descriptions into English. Preparation of 2 .TXT files[a] (Japan RCs and England RCs datasets) to be used with the software Sketch Engine.
Text analysis (micro-level)	According to research interests: Identify the major underlying themes and subthemes. Examine linguistic choices used to represent social actors or events. Examine the stance taken by the author or speaker. Examine whether the text includes references to other texts (intertextuality).	Corpus linguistics-based analysis supported with the software Sketch Engine to retrieve keywords (single and multi-words) and identification of key themes for each dataset. Examination of concordances (keywords in context) guided by the RQs (ie, construal of RCs and recovery in England and Japan RCs promotional texts). Focus on both the portrayal of the RCs and the role attribution to service users.
Discourse practice (meso-level)	Examine the contexts of production and reception of the text. Consider the goal of the text, who has produced it, and the putative audience.	The context of production (RCs as mental health intervention) and the values of producers (RC managerial staff) and putative audience (RC students) have been studied [38]. For the purpose of this study (ie, construal of RC and recovery), the linguistic analysis has not considered characteristics of the textual register (promotional texts of medical services).
Sociocultural practice (macro-level)	Consider relevant sociocultural or historical factors (ie, context of production) that have conditioned the text. Consider whether the text reflects any sociocultural values. **Original version** Examine social relations that control the production of the text; in addition, examine the reciprocal relations (how the texts affect social practices and structures). How do social practices inform the arguments in the text? How does the text in turn influence social practices?	Examination of RCs as a new mental health support system and its main underpinning philosophies. Interpretation of the main themes emerging from the linguistic analysis based on Hofstede's cultural dimensions theory [45] and Kotera et al's [38] study on the impact of culture on the RC operational model.

[a]TXT files: text files.

Conclusions

As both CDA and mental health recovery research share the common goal of identifying and addressing inequalities, our adaptation of established CDA frameworks to cross-cultural mental health recovery research represents a crucial step forward. This paper offered valuable insights to researchers seeking to explore recovery in cross-cultural contexts using CDA. We hope our approach will foster a more evidence-based understanding of recovery and reduce inequalities.

Acknowledgments

We would like to thank Nigel Henderson who helped facilitate the completion of RC surveys in Scotland. We thank the RECOLLECT Lived Experience Advisory Panel (LEAP) who provided input into the design of the survey and interpretation of results. MS acknowledges the support of NIHR Nottingham Biomedical Research Centre.

Conflicts of Interest

None declared.

References

1. Fairclough N. Critical Discourse Analysis. 2nd ed. London, England: Routledge; 2010.

XSL•FO
RenderX

2. Fairclough N. Language and Power. 3rd ed. London, England: Routledge; 2014.
3. Demjén Z, Atkins S, Semino E. Researching Language and Health. London: Routledge; 2023.
4. Karaminis T, Gabrielatos C, Maden-Weinberger U, Beattie G. Portrayals of autism in the British press: a corpus-based study. Autism 2023;27(4):1092-1114 [FREE Full text] [doi: 10.1177/13623613221131752] [Medline: 36317371]
5. Balfour J. Representing Schizophrenia in the Media. London, England: Routledge; 2023.
6. Price H. The Language of Mental Illness. Cambridge, England: Cambridge University Press; 2022.
7. Hunt D, Brookes G. Corpus, Discourse and Mental Health. London, England: Bloomsbury Academic; 2020.
8. Jones S, Atanasova D, Dodd S, Flowers S, Rosala-Hallas A, Robinson H, et al. Use of an online forum for relatives of people with psychosis and bipolar disorder: mixed methods study. JMIR Ment Health 2022;9(10):e35837 [FREE Full text] [doi: 10.2196/35837] [Medline: 36264621]
9. Demjén Z, Semino E. Stylistics: Mind style in an autobiographical account of schizophrenia. In: Brookes G, Hunt D, editors. Analysing Health Communication: Discourse Approaches. London: Palgrave; 2021.
10. Demjén Z, Semino E. Henry's voices: the representation of auditory verbal hallucinations in an autobiographical narrative. Med Humanit 2015;41(1):57-62. [doi: 10.1136/medhum-2014-010617] [Medline: 25505160]
11. Galasinski D. Men's Discourses of Depression. Basingstoke, England: Palgrave Macmillan; 2008.
12. Meehan T, MacLachlan M. Self construction in schizophrenia: a discourse analysis. Psychol Psychother 2010;81(2):131-142. [doi: 10.1348/147608307x256777]
13. Kłosińska U, Ziółkowska J. Discourses of work life in narrative identities of people with advanced dementia. Int J Geriatr Psychiatry 2023;38(5):e5921. [doi: 10.1002/gps.5921] [Medline: 37173832]
14. Mancini MA, Rogers R. Narratives of recovery from serious psychiatric disabilities: a critical discourse analysis. Crit Approach Discourse Anal Disciplines 2015;1(2):35-40. [doi: 10.4135/9781849208765.n8]
15. Kotera Y, Miyamoto Y, Vilar-Lluch S, Aizawa I, Reilly O, Miwa A, et al. Cross-cultural comparison of recovery college implementation between Japan and England: corpus-based discourse analysis. Int J Ment Health Addict 2024 [FREE Full text] [doi: 10.1007/s11469-024-01356-3]
16. Slade M, Amering M, Farkas M, Hamilton B, O'Hagan M, Panther G, et al. Uses and abuses of recovery: implementing recovery-oriented practices in mental health systems. World Psychiatry 2014;13(1):12-20 [FREE Full text] [doi: 10.1002/wps.20084] [Medline: 24497237]
17. Anthony WA. Recovery from mental illness: the guiding vision of the mental health service system in the 1990s. Psychosocial Rehabil J 1993;16(4):11-23. [doi: 10.1037/h0095655]
18. Ellison ML, Belanger LK, Niles BL, Evans LC, Bauer MS. Explication and definition of mental health recovery: a systematic review. Adm Policy Ment Health 2018;45(1):91-102. [doi: 10.1007/s10488-016-0767-9] [Medline: 27709376]
19. Dell NA, Long C, Mancini MA. Models of mental health recovery: an overview of systematic reviews and qualitative meta-syntheses. Psychiatr Rehabil J 2021;44(3):238-253. [doi: 10.1037/prj0000444] [Medline: 33734781]
20. McPherson P, Lloyd-Evans B, Dalton-Locke C, Killaspy H. A systematic review of the characteristics and efficacy of recovery training for mental health staff: implications for supported accommodation services. Front Psychiatry 2021;12:624081 [FREE Full text] [doi: 10.3389/fpsyt.2021.624081] [Medline: 34054593]
21. Jagfeld G, Lobban F, Humphreys C, Rayson P, Jones SH. How people with a bipolar disorder diagnosis talk about personal recovery in peer online support forums: corpus framework analysis using the POETIC framework. JMIR Med Inform 2023;11:e46544 [FREE Full text] [doi: 10.2196/46544] [Medline: 37962520]
22. Jagfeld G, Lobban F, Marshall P, Jones SH. Personal recovery in bipolar disorder: systematic review and 'best fit' framework synthesis of qualitative evidence - a POETIC adaptation of CHIME. J Affect Disord 2021;292:375-385 [FREE Full text] [doi: 10.1016/j.jad.2021.05.051] [Medline: 34139411]
23. Department of Health and Ageing. Fourth National Mental Health Plan: An Agenda for Collaborative Government Action in Mental Health 2009. Canberra: Commonwealth of Australia; 2009.
24. Mental Health Commission of Canada. Changing Directions, Changing Lives: The Mental Health Strategy for Canada. Calgary: Mental Health Commission of Canada; 2012.
25. Department of Health Social Services and Public Safety of Northern Ireland. In: Revised Service Framework for Mental Health and Wellbeing 2018-2021. Belfast: Department of Health Social Services and Public Safety; 2018.
26. New Freedom Commission on Mental Health. Achieving the promise: transforming mental health care in America report. U.S. Department of Health and Human Services. 2003. URL: https://dmh.mo.gov/media/pdf/achieving-promise-transforming-mental-health-care-america-report [accessed 2025-02-06]
27. No health without mental health: a cross-government mental health outcomes strategy for people of all ages. HM Government. 2011. URL: https://assets.publishing.service.gov.uk/media/5a7c348ae5274a25a914129d/dh_124058.pdf [accessed 2025-02-06]
28. Slade M, Adams N, O'Hagan M. Recovery: past progress and future challenges. Int Rev Psychiatry 2012;24(1):1-4 [FREE Full text] [doi: 10.3109/09540261.2011.644847] [Medline: 22385420]
29. National Mental Health Development Plan No. 1 (2018-2037). Strategy and Planning Division, Department of Mental Health. Department of Mental Health-Thailand. Nonthaburi: Department of Mental Health-Thailand; 2019. URL: https://dmh-elibrary.org/items/show/450 [accessed 2025-01-23]

XSL•FO
RenderX

30. Blommaert J, Bulcaen C. Critical discourse analysis. Annu Rev Anthropol 2000;29(1):447-466 [FREE Full text] [doi: 10.1146/annurev.anthro.29.1.447]
31. Slade M, Longden E. Empirical evidence about recovery and mental health. BMC Psychiatry 2015;15(1):285 [FREE Full text] [doi: 10.1186/s12888-015-0678-4] [Medline: 26573691]
32. Slade M. Personal Recovery and Mental Illness: A Guide for Mental Health Professionals. Cambridge: Cambridge University Press; 2009.
33. Crowe M. Psychiatry and/or recovery: a critical analysis. Int J Ment Health Nurs 2022;31(6):1542-1551 [FREE Full text] [doi: 10.1111/inm.13072] [Medline: 36379908]
34. Hayes D, Hunter-Brown H, Camacho E, McPhilbin M, Elliott RA, Ronaldson A, et al. Organisational and student characteristics, fidelity, funding models, and unit costs of recovery colleges in 28 countries: a cross-sectional survey. Lancet Psychiatry 2023;10(10):768-779. [doi: 10.1016/s2215-0366(23)00229-8]
35. Whitley R, Shepherd G, Slade M. Recovery colleges as a mental health innovation. World Psychiatry 2019;18(2):141-142 [FREE Full text] [doi: 10.1002/wps.20620] [Medline: 31059628]
36. Kotera Y, Ronaldson A, Hayes D, Hunter-Brown H, McPhilbin M, Dunnett D, et al. Cross-cultural insights from two global mental health studies: self-enhancement and ingroup biases. Int J Ment Health Addiction 2024. [doi: 10.1007/s11469-024-01307-y]
37. Toney R, Knight J, Hamill K, Taylor A, Henderson C, Crowther A, et al. Development and evaluation of a recovery college fidelity measure. Can J Psychiatry 2019;64(6):405-414 [FREE Full text] [doi: 10.1177/0706743718815893] [Medline: 30595039]
38. Kotera Y, Ronaldson A, Hayes D, Hunter-Brown H, McPhilbin M, Dunnett D, et al. 28-country global study on associations between cultural characteristics and recovery college fidelity. Npj Ment Health Res 2024;3(1):46. [doi: 10.1038/s44184-024-00092-9] [Medline: 39379618]
39. Kotera Y, Ozaki A, Miyatake H, Tsunetoshi C, Nishikawa Y, Kosaka M, et al. Qualitative investigation into the mental health of healthcare workers in Japan during the COVID-19 pandemic. Int J Environ Res Public Health 2022;19(1):568 [FREE Full text] [doi: 10.3390/ijerph19010568] [Medline: 35010828]
40. Mullet DR. A general critical discourse analysis framework for educational research. J Adv Acad 2018;29(2):116-142. [doi: 10.1177/1932202x18758260]
41. Hunston S. Corpora in Applied Linguistics. Cambridge: Cambridge University Press; 2002.
42. Flowerdew L. Corpus-based discourse analysis. The Routledge Handbook of Discourse Analysis. London: Routledge; 2023.
43. Bartlett T. Positive discourse analysis 1. In: Flowerdew J, Richardson JE, editors. The Routledge Handbook of Critical Discourse Studies. London: Routledge; 2017.
44. Martin JR. Positive discourse analysis: solidarity and change. Revista Canaria Estudios Ingleses 2004;49:179-200. [doi: 10.25145/j.recaesin]
45. Hofstede G, Minkov M. Values Survey Module 2013 Manual. Geert Hofstede BV. URL: https://geerthofstede.com/wp-content/uploads/2016/07/Manual-VSM-2013.pdf [accessed 2025-01-23]
46. Liu K, Guo F. A review on critical discourse analysis. Theory Pract Lang Stud 2016;6(5):1076. [doi: 10.17507/tpls.0605.23]
47. Llewellyn-Beardsley J, Rennick-Egglestone S, Callard F, Crawford P, Farkas M, Hui A, et al. Characteristics of mental health recovery narratives: systematic review and narrative synthesis. PLoS One 2019;14(3):e0214678 [FREE Full text] [doi: 10.1371/journal.pone.0214678] [Medline: 30921432]
48. Monk-Cunliffe J. How can we overcome health inequalities in psychiatry? BJPsych Bull 2023;47(4):228-230. [doi: 10.1192/bjb.2023.49] [Medline: 37401250]
49. Anjum G, Aziz M. Advancing equity in cross-cultural psychology: embracing diverse epistemologies and fostering collaborative practices. Front Psychol 2024;15:1368663 [FREE Full text] [doi: 10.3389/fpsyg.2024.1368663] [Medline: 38638521]
50. Widdowson HG. Discourse analysis: a critical view. Lang Literature 1995;4(3):157-172. [doi: 10.1177/096394709500400301]
51. Breeze R. Critical discourse analysis and its critics. Pragmatics 2011;21(4):493-525 [FREE Full text] [doi: 10.1075/prag.21.4.01bre]

Abbreviations

CDA: critical discourse analysis
RC: recovery college
TXT file: text file

XSL•FO
RenderX

Edited by A Mavragani; submitted 08.07.24; peer-reviewed by C Entwistle; comments to author 20.09.24; revised version received 04.10.24; accepted 11.10.24; published 21.02.25.

Please cite as:
Kotera Y, Daryanani R, Skipper O, Simpson J, Takhi S, McPhilbin M, Ingall BR, Namasaba M, Jepps J, Kellermann V, Bhandari D, Ojio Y, Ronaldson A, Guerrero E, Jebara T, Henderson C, Slade M, Vilar-Lluch S
Applying Critical Discourse Analysis to Cross-Cultural Mental Health Recovery Research
JMIR Form Res 2025;9:e64087
URL: https://formative.jmir.org/2025/1/e64087
doi:10.2196/64087
PMID:

©Yasuhiro Kotera, Riddhi Daryanani, Oliver Skipper, Jonathan Simpson, Simran Takhi, Merly McPhilbin, Benjamin-Rose Ingall, Mariam Namasaba, Jessica Jepps, Vanessa Kellermann, Divya Bhandari, Yasutaka Ojio, Amy Ronaldson, Estefania Guerrero, Tesnime Jebara, Claire Henderson, Mike Slade, Sara Vilar-Lluch. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Are Dating App Algorithms Making Men Lonely and Does This Present a Public Health Concern?

Eric Balki[1], BSc, MSc, MA, PhD

Faculty of Health and Medicine, Department of Health, Lancaster University, Lancaster, United Kingdom

Corresponding Author:
Eric Balki, BSc, MSc, MA, PhD
Faculty of Health and Medicine
Department of Health
Lancaster University
Innovation One, Sir John Fisher Dr, Bailrigg
Lancaster, LA1 4AT
United Kingdom
Phone: 44 1524 593763
Email: e.balkhi@lancaster.ac.uk

Abstract

During the pandemic, dating apps emerged as essential platforms connecting users amid social isolation, experiencing rapid growth in engagement and profile creation. This paper examines the evolution of these apps, highlighting their shift from facilitating offline encounters to promoting match accumulation for revenue. In particular, the study investigates gender disparities, addictive behaviors, and algorithmic match throttling that disproportionately impact men's psychological well-being. Drawing on evidence linking dating app use to increased depression and anxiety, the analysis calls for regulatory intervention to eliminate pay-for-advantage models and ensure fair, healthy user experiences, thereby mitigating adverse public health outcomes.

(JMIR Form Res 2025;9:e70594) doi:10.2196/70594

KEYWORDS

dating apps; mental health; men; algorithm; anxiety; depression; loneliness

This paper deals with issues revolving around heterosexual dating, which currently forms the largest marketplace in dating apps [1]. During the pandemic with frequent lockdowns, dating apps provided an opportunity for people to connect [2]. In a short span of time, there was a surge in dating profiles on these apps. Just like Peloton, the home exercise machine company that found a way of appealing to users to use stationary bikes and treadmills, dating apps attracted a vast range of young adults looking for a connection or relationship, at a time when isolation and loneliness was high [3].

At the start of the pandemic, the way the apps were working was quite simple. When two users accepted each other's profile, they had a match. Browsing through and evaluating other user's profiles was the central activity, eventually leading to an offline encounter where possible [4]. Dating apps, however, slowly started to not facilitate offline encounters but rather feed into the user's desire of accumulating matches [5].

Even before the pandemic, more than half of the users on dating apps reported not going on in-person dates, and this percentage was especially higher for women [6], with them looking for validation, gratification, and a boost in self-worth instead [7]. It is very likely this has worsened even further, with a large majority of users not going on in-person, offline dates [8].

Even those who are now in relationships find it hard to quit these apps—keeping them for gratification purposes [9], especially in women who accumulate matches to satisfy the need for belongingness or as a replacement for relational or sexual intimacy, getting value from social acceptance and approval [10].

In the last two years, the problem has exacerbated significantly. Women report being inundated with so many matches that it is hard from them to make a decision on choosing someone to engage with properly [11]. At the same time, men are on the opposite end of the spectrum, where they get very few responses and must purchase expensive, paid features and subscriptions, which allow them to get a limited amount of priority over nonpaying members [12].

Men continue to form the largest group of paid subscribers to dating apps [12]. Dating apps are focused on maximizing profit activities, especially as many of them are now owned by publicly listed companies [13]. Losing a subscriber means losing revenue, so it is conjectured that dating apps are motivated to keep their

XSL•FO
RenderX

paying customers as long as they possibly can, as a successful offline encounter mediated by the dating app would lead to the loss of that subscriber [12]. Unlike other social media platforms, there is no upside in successful encounters from a revenue-generating perspective. Men are also more likely to quit a dating app if they have found someone in person [12].

Dating apps are like casinos in a way, in that they have to strategize where the reward needs to be—just enough to keep users coming back for more, but the reward cannot be so high that users walk away and not return, which can increase addictive behavior [14]. They have gamified the process of meeting a partner, invoking "gambler" tendencies, thus leading to behaviors that are addictive or even compulsive [15]. Now, sophisticated artificial intelligence (AI) tools are being deployed by dating apps to maximize subscriber revenue, which is potentially causing excessive harm to the psychological profiles of a large number of men and hindering the development of healthy relationships, which could become a large public health concern [16]. Thus, there is both algorithmic and natural throttling of the mutual acceptance matching process on these apps, which is driving an increasing frustration in men as well as impacting their female users [16].

This is the fallacy in the approach by dating app companies, as the pool of paid subscribers will eventually, and inevitably, fall as the frustration leads users to find alternative ways of finding dates.

It has become extremely important for researchers to study the psychological traits and well-being of users of dating apps [17], especially with the changes and advances in algorithms and use of AI. As things stand, there are very few studies that have looked at dating app use; the algorithms being deployed by such companies; match throttling; and the associated psychological outcomes, especially in men.

There is a need for studies that move away from simply associating correlations between dating app use and psychological outcomes, including mere use associated with decreased well-being. Authors need to distinguish which specific activities, features, or experiences may be leading to undesired effects. While it may not always be possible to understand how the backend of the apps works or what algorithms may be at play, real-time testing with users could reveal basic issues revolving around authentic matching [18]. The design of such studies will have to be carefully considered to ensure the cross-sectional design does not end up ruling out self-selection, that is, attracting lonely individuals and those in fear of being single [19].

The majority of the population are reliant on dating apps to meet prospective partner, and such apps are having an impact on their mental health; these factors may very well make this into a public health concern. The Centers for Disease Control and Prevention's [20] Household Pulse Survey estimates that depression and anxiety are now affecting 34.2% of the population, making them the two most common mental health disorders in the United States, and to some extent, the increased numbers have been associated with dating app use [21].

Some studies have linked dating apps directly with higher depression and anxiety, especially the frequency and length of use [17]. Symptoms from these disorders can further impact social skills, making the formation of healthy relationships even more difficult, thus leading to a downward spiral in the populace [18].

This is an urgent problem that needs regulatory action from the Western governments, where dating app algorithms should provide fair conditions to all users. Match throttling and pay-for-advantage models should be disallowed as they have a disproportionately negative impact on the health of young men.

If such action is not taken, then over time we will have a population of young, productive men who grow tired of using dating apps to find a partner and come out feeling lonely, depressed, and anxious, thus impacting in other areas of their life and their ability to be productive. Therefore, rather than helping people find partners, dating apps may be failing the very users they rely upon for their income.

Acknowledgments

This work was not funded.

Authors' Contributions

The author listed has made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflicts of Interest

None declared.

References

1. Bandinelli C, Gandini A. Dating apps: the uncertainty of marketised love. Cultural Sociology 2022 Jan 10;16(3):423-441. [doi: 10.1177/17499755211051559]
2. Gibson AF. Exploring the impact of COVID-19 on mobile dating: critical avenues for research. Soc Personal Psychol Compass 2021 Nov;15(11):e12643 [FREE Full text] [doi: 10.1111/spc3.12643] [Medline: 34899975]
3. Weldon M. The Time We Have: Essays on Pandemic Living. Evanston, IL: Northwestern University Press; 2024.

XSL•FO
RenderX

4. Cummings JJ, Mays K. Trait motivational reactivity as a predictor of online dating app behavior. Computers in Human Behavior 2021 Aug;121:106775. [doi: 10.1016/j.chb.2021.106775]
5. Bandinelli C, Bandinelli A. What does the app want? a psychoanalytic interpretation of dating apps' libidinal economy. Psychoanal Cult Soc 2021 Apr 19;26(2):181-198. [doi: 10.1057/s41282-021-00217-5]
6. Timmermans E, Courtois C. From swiping to casual sex and/or committed relationships: exploring the experiences of Tinder users. The Information Society 2018 Mar 08;34(2):59-70. [doi: 10.1080/01972243.2017.1414093]
7. Wimark T, Hedlund D. From individuals to emotional drones: technology-driven change in the collective conditioning of intimacy. New Media & Society 2021 Nov 18;26(1):236-252. [doi: 10.1177/14614448211057108]
8. Toma CL. Online dating and psychological wellbeing: a social compensation perspective. Curr Opin Psychol 2022 Aug;46:101331. [doi: 10.1016/j.copsyc.2022.101331] [Medline: 35349878]
9. Sharabi LL, Timmermans E. Why settle when there are plenty of fish in the sea? Rusbult's investment model applied to online dating. New Media & Society 2020 Jul 10;23(10):2926-2946. [doi: 10.1177/1461444820937660]
10. Cross SE, Madson L. Models of the self: self-construals and gender. Psychol Bull 1997 Jul;122(1):5-37. [doi: 10.1037/0033-2909.122.1.5] [Medline: 9204777]
11. Thomas MF, Binder A, Stevic A, Matthes J. 99+ Matches but a spark ain't one: adverse psychological effects of excessive swiping on dating apps. Telematics and Informatics 2023 Mar;78:101949. [doi: 10.1016/j.tele.2023.101949]
12. O'Gorman K, Wilson MJ, Seidler Z. Swiping into the void: tears, fears, and motivations of young men on dating apps. In: Seilder Z, editor. Masculinities and Mental Health in Young Men. Cham, Switzerland: Palgrave Macmillan, Springer Nature; 2024:229-262.
13. Nickenig A. The market of love: dating economies from early modern matchmaking to Tinder. In: Schomacher E, Söffner J, editors. The Culture of Money: Implications for Contemporary Economics. London, United Kingdom: Routledge; 2024.
14. Mackinnon L. Love, games and gamification: gambling and gaming as techniques of modern romantic love. Theory, Culture & Society 2022 Mar 17;39(6):121-137. [doi: 10.1177/02632764221078258]
15. Isisag A. Mobile dating apps and the intensive marketization of dating: gamification as a marketizing apparatus. In: Bagchi R, Block L, Lee L, editors. Advances in Consumer Research, Volume 47. Duluth, MN: Association for Consumer Research; 2019:135-141.
16. Fletcher R, Tzani C, Ioannou M. The dark side of artificial intelligence – risks arising in dating applications. Assessment and Development Matters 2024 Mar 01;16(1):17-23. [doi: 10.53841/bpsadm.2024.16.1.17]
17. Holtzhausen N, Fitzgerald K, Thakur I, Ashley J, Rolfe M, Pit SW. Swipe-based dating applications use and its association with mental health outcomes: a cross-sectional study. BMC Psychol 2020 Mar 04;8(1):22 [FREE Full text] [doi: 10.1186/s40359-020-0373-1] [Medline: 32127048]
18. Braithwaite S, Holt-Lunstad J. Romantic relationships and mental health. Curr Opin Psychol 2017 Feb;13:120-125. [doi: 10.1016/j.copsyc.2016.04.001] [Medline: 28813281]
19. Lauckner C, Truszczynski N, Lambert D, Kottamasu V, Meherally S, Schipani-McLaughlin AM, et al. "Catfishing," cyberbullying, and coercion: an exploration of the risks associated with dating app use among rural sexual minority males. Journal of Gay & Lesbian Mental Health 2019 Mar 29;23(3):289-306. [doi: 10.1080/19359705.2019.1587729]
20. Chin K, Edelstein R, Vernon PA. Attached to dating apps: attachment orientations and preferences for dating apps. Mobile Media & Communication 2018 Jun 18;7(1):41-59. [doi: 10.1177/2050157918770696]
21. National Center for Health Statistics. Anxiety and depression. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nchs/covid19/pulse/mental-health.htm [accessed 2024-12-25]

Abbreviations

AI: artificial intelligence

Edited by A Mavragani; submitted 26.12.24; peer-reviewed by V Sideropoulos; comments to author 06.02.25; revised version received 25.02.25; accepted 11.03.25; published 07.04.25.

Please cite as:
Balki E
Are Dating App Algorithms Making Men Lonely and Does This Present a Public Health Concern?
JMIR Form Res 2025;9:e70594
URL: https://formative.jmir.org/2025/1/e70594
doi:10.2196/70594
PMID:

XSL•FO
RenderX

©Eric Balki. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Quantifying the Regional Disproportionality of COVID-19 Spread: Modeling Study

Kenji Sasaki[1], PhD; Yoichi Ikeda[1], PhD; Takashi Nakano[1,2], PhD

[1]Center for Infectious Disease Education and Research, Osaka University, Co-creation BLDG. D88-1, 2-1 Yamadaoka, Suita, Osaka, Japan
[2]Research Center for Nuclear Physics, Osaka University, Ibaraki, Osaka, Japan

Corresponding Author:
Kenji Sasaki, PhD
Center for Infectious Disease Education and Research, Osaka University, Co-creation BLDG. D88-1, 2-1 Yamadaoka, Suita, Osaka, Japan

Abstract

Background: The COVID-19 pandemic has caused serious health, economic, and social consequences worldwide. Understanding how infectious diseases spread can help mitigate these impacts. The Theil index, a measure of inequality rooted in information theory, is useful for identifying geographic disproportionality in COVID-19 incidence across regions.

Objective: This study focused on capturing the degrees of regional disproportionality in incidence rates of infectious diseases over time. Using the Theil index, we aim to assess regional disproportionality in the spread of COVID-19 and detect epicenters where the number of infected individuals was disproportionately concentrated.

Methods: To quantify the degree of disproportionality in the incidence rates, we applied the Theil index to the publicly available data of daily confirmed COVID-19 cases in the United States over a 1100-day period. This index measures relative disproportionality by comparing daily regional case distributions with population proportions, thereby identifying regions where infections are disproportionately concentrated.

Results: Our analysis revealed a dynamic pattern of regional disproportionality in the confirmed cases by monitoring variations in regional contributions to the Theil index as the pandemic progressed. Over time, the index reflected a transition from localized outbreaks to widespread transmission, with high values corresponding to concentrated cases in some regions. We also found that the peaks in the Theil index often preceded surges in confirmed cases, suggesting its potential utility as an early warning signal.

Conclusions: This study demonstrated that the Theil index is one of the effective indices for quantifying regional disproportionality in COVID-19 incidence rates. Although the Theil index alone cannot fully capture all aspects of pandemic dynamics, it serves as a valuable tool when used alongside other indicators such as infection and hospitalization rates. This approach allows policy makers to monitor regional disproportionality efficiently, offering insights for early intervention and targeted resource allocation.

(JMIR Form Res 2025;9:e59230) doi:10.2196/59230

KEYWORDS

infectious disease; COVID-19; epidemiology; public health; SARS-CoV-2; pandemic; inequality measure; information theory; Kullback-Leibler divergence

Introduction

The COVID-19 pandemic has caused serious health problems and has had major economic and social consequences worldwide. It has highlighted the need to understand regional disparities in infection rates to strengthen public health responses since infection dynamics are influenced by factors such as population density, socioeconomic conditions, and health care infrastructure [1,2]. Numerous indicators and models have been proposed to address the problem, and mechanisms for the spread of the infection and intervention measures to control the pandemic have been studied [3-9].

Several recent studies have investigated regional differences in COVID-19 prevalence [10-13]. Differences in the prevalence rates between regions highlight the need to understand regional inequalities in pandemic response strategies. Effectively addressing these disparities requires accurate quantification and understanding of regional disproportionalities in daily confirmed COVID-19 cases.

In the field of economics, various indicators have been developed to measure resource and income inequality, including an index proposed by Theil, which incorporated information theory [14]. Manz and Mansmann [15] have demonstrated the importance of using inequality indices for monitoring changes in geographic inequality; for instance, the Theil index was used to track geographic disproportionality over time during the COVID-19 pandemic, providing important insights for public health policy.

XSL•FO
RenderX

The aim of this paper is to quantify the interregional disproportionality in the number of confirmed cases using the Theil index, which mathematically corresponds to the Kullback-Leibler (KL) divergence in information theory [16]. The Theil index is an effective method of measuring the degree of disproportionality and objectively assessing biases in the interregional distribution of infected individuals.

Methods

Overview

We analyzed the time trends of daily COVID-19–confirmed cases in the United States over 1100 days since the first reported case on January 21, 2020 [17]. Data are taken from the COVID-19 data repository at the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University [18]. US state population data were obtained from the US Census Bureau website [19]. Population changes due to migration, births, and deaths were not considered in the analysis.

The Disproportionality Measure: Theil Index

The Theil index is commonly applied in various fields including economics, sociology, and information theory. It quantifies the relative differences between various components of a dataset. In the context of regional analysis of the confirmed cases, the Theil index can be employed to evaluate the distribution of infected individuals across different regions. In this study, we use the Theil index to identify regions with disproportionate numbers of confirmed cases relative to their population size.

The discrete form of the Theil index is expressed as:

$$T=\sum_{i=1}^{N} t_i=\sum_{i=1}^{N} p_i \ln \frac{p_i}{q_i}$$

where N is the total number of regions being considered and ln is the natural logarithm. The Theil index T is composed of a sum of t_i which is a partial contribution from region i. The discrete probability distribution, p_i, in region i is defined as the ratio of daily confirmed cases region i to the total confirmed cases across all regions for that day. Similarly, the population ratio, q_i, in region i is the ratio of the population in region i to the total population across all regions.

The Theil index, which is mathematically related to the KL divergence, is a nonsymmetric metric that measures the relative entropy or informational difference between two distributions. It is sensitive to the interregional variations in the distribution of the confirmed cases, with its maximum value attained when the confirmed cases are concentrated in areas with the smallest population proportion. Consequently, the index tends to exhibit higher values when a small number of regions account for a large share of the confirmed cases, and conversely, lower values when the confirmed cases are more evenly distributed across regions. Notably, it remains nonnegative and reaches a minimum value of 0 only when the two distributions are identical. Therefore, applying the Theil index to the time-series data of the confirmed cases, and monitoring changes in the index over time, we quantified the degree of spread of COVID-19 cases and assessed whether the confirmed cases were disproportionately concentrated relative to the regional population sizes over time.

Ethical Considerations

This study used publicly available, deidentified COVID-19 data from CSSE at Johns Hopkins University [18], and therefore, additional ethics approval and informed consent were not required. The aggregated data ensured privacy and confidentiality, and no direct human participants were involved; thus, no compensation was provided. No identifiable information appears in any images or materials.

Results

To address fluctuations in the Theil index caused by data aggregation inconsistencies during holidays across different regions, the 7-day average of confirmed COVID-19 cases was used instead of raw data.

Figure 1 illustrates a 2-axis graph showing the time trends of the Theil index (left axis) and the number of confirmed cases in a logarithmic scale (right axis, logarithmic scale). The horizontal axis represents the number of days elapsed (denoted by d in the text) since the date of the first reported case in the United States.

In Figure 1, there are eight notable surges of the confirmed cases, occurring at approximately d=80 (first), 180 (second), 350 (third), 450 (fourth), 580 (fifth), 720 (sixth), 900 (seventh), and 1080 (eighth), respectively. The presence of multiple peaks in the Theil index indicates that infected individuals were concentrated in specific regions during the period, and the degree of this concentration can be assessed by examining the numerical values. However, it is important to note that changes in the Theil index simply indicate the degree of regional disproportionality in the confirmed cases rather than absolute increases or decreases in the number of infected individuals. Therefore, this indicator is most effective when interpreted in conjunction with actual trends in the number of confirmed cases.

Before the first peak, the number of confirmed cases was quite low, and the Theil index fluctuated erratically. As d increased near the first peak, the Theil index gradually decreased, reaching a local minimum around d=120. This suggested that the initially localized epidemic began to spread throughout the US during the early stages of the global pandemic. Similar trends were observed during subsequent surges, such as slight a increase in the Theil index before the peak, followed by a decrease. This could be seen as a precursor to a surge in the number of infected individuals. This finding aligns with previous research by Ikeda, Sasaki, and Nakano [7].

The following examples provide interesting insights; when the Theil index value was high and the number of confirmed cases was low (d=60, 550, etc), it indicated that the infectious disease was localized and beginning to spread to various regions. Conversely, when the index was low and the number of confirmed cases was high (d=750, etc), it indicated that there was no obvious epicenter of the infectious disease, with the number of confirmed cases increasing relatively and evenly across different regions.

The contributions to the Theil index from each region (t_i), calculated from the number of cases on each date, were arranged

XSL•FO
RenderX

in chronological order and visualized using a heatmap, as shown in Figure 2. Regions with a high proportion of confirmed cases are represented in red, while regions with a low proportion are colored blue. Notably, there are long intervals between the deep red patches in some regions such as California, Florida, and New York. Particularly, the periods of intense infection represented by these deep red patches were not repeated at short intervals. This phenomenon is of great importance in infectious disease management. Once a major epidemic in an area has subsided, the interval between subsequent outbreaks provides an opportunity to rebuild the health care infrastructure and implement preventive measures before the occurrence of the next epidemic.

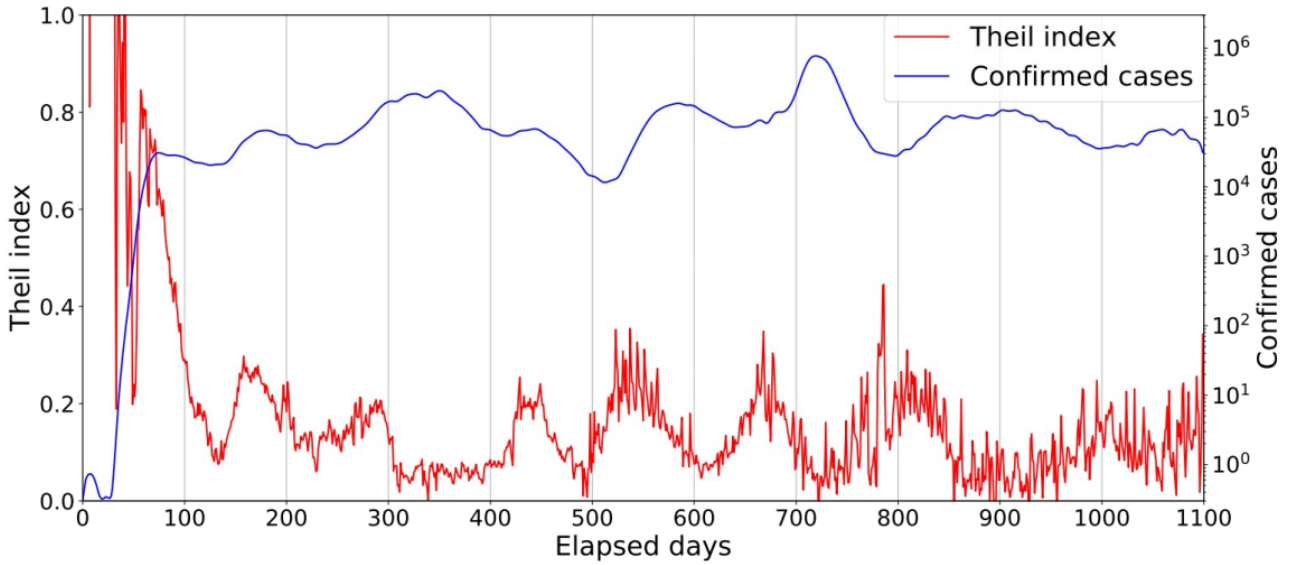
Based on the observations from Figure 2, the epicenter of infectious diseases as indicated by the red patches alternates between New York, California, and Florida. This insight is crucial for understanding the underlying mechanisms of the spread of infectious diseases in the future. Furthermore, after d=750, both the red and blue colors fade over time, indicating the absence of a single epicenter, and a widespread outbreak of COVID-19. This pattern suggests the ineffectiveness of countermeasures against the spread of infectious diseases under these circumstances.

Figure 3A shows the contributions to the Theil index by region at d=60. The horizontal axis in the figure shows the state code (as listed in Multimedia Appendix 1). There is a significant contribution to the Theil index from New York State compared to the other regions. Figure 3B shows that at this point confirmed cases were highly localized in these regions.

There were relatively large negative contributions to the Theil index from California, Florida, and Texas, which were regions with high population ratios. It is interesting to note that there was little risk of infection in these regions at that point; however, the number of infected individuals rapidly increased following the concentration of confirmed cases in New York.

Figure 4 shows the contributions to the Theil index from each region at d=550 and 750. At d=550 shown in Figure 4A, the Theil index reaches a peak, and the trend of confirmed cases is increasing. This suggests that a new epidemic is emerging, mainly in Florida and Louisiana. However, their contributions are significantly smaller compared to New York at d=60, as seen in Figure 3. This indicates that regional disproportionality is much less pronounced than in the early stage of the COVID-19 pandemic. It is also interesting to look at data on d=750 as shown in Figure 4B, when confirmed cases in the United States are at their maximum. Although several regions show large contributions to the Theil index, the epicenter of COVID-19 is no longer obvious, suggesting that COVID-19 cases are uniformly distributed across the country.

Figure 1. Time trends of the Theil index on the left axis and the 7-day average number of the confirmed cases on the right axis on a logarithmic scale are shown in the red and blue curves, respectively. The horizontal axis is the number of days elapsed since January 21, 2020.



XSL•FO
RenderX

Figure 2. Partial contributions to the Theil index from each region, t_i, are displayed in a heatmap over time. The horizontal axis represents the number of days elapsed since January 21, 2020. The vertical axis shows the names of states in the United States. The positive (high concentration of incidences) and negative (low concentration) contributions to the Theil index correspond to deep red and blue colors, respectively.

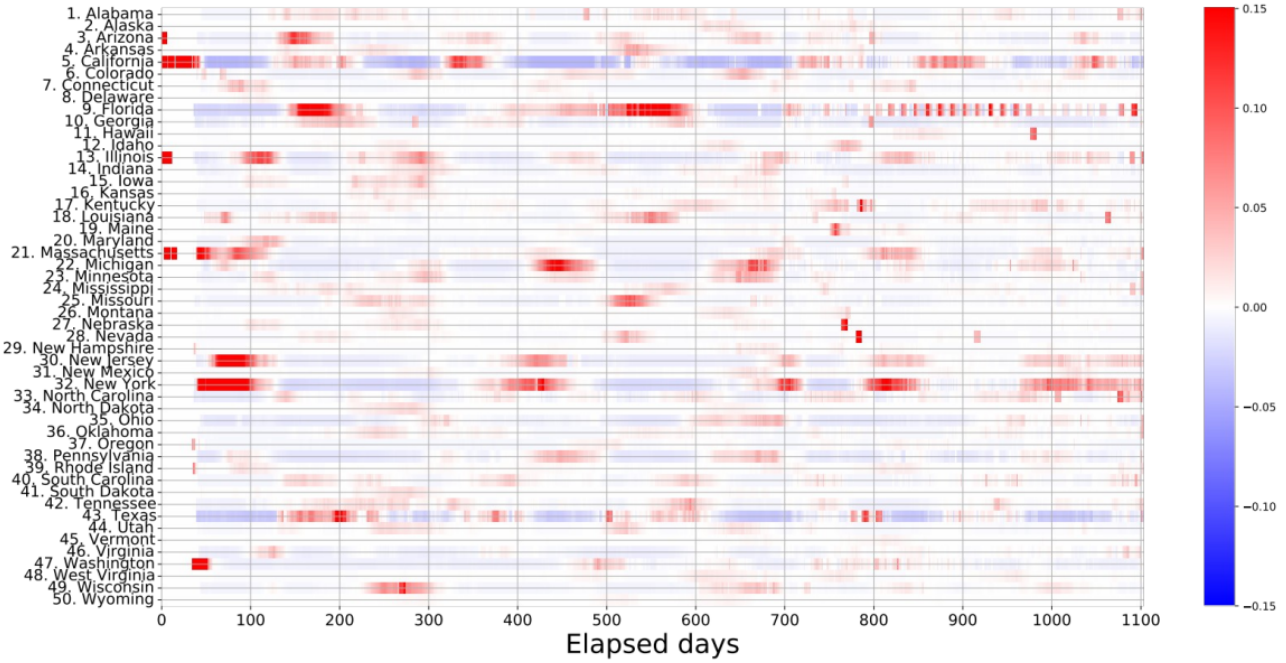
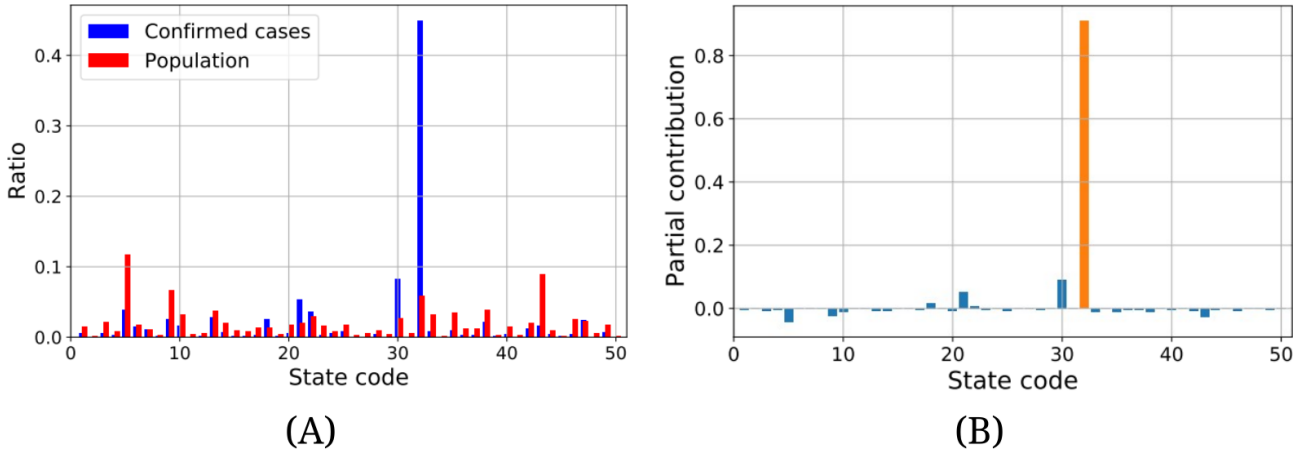
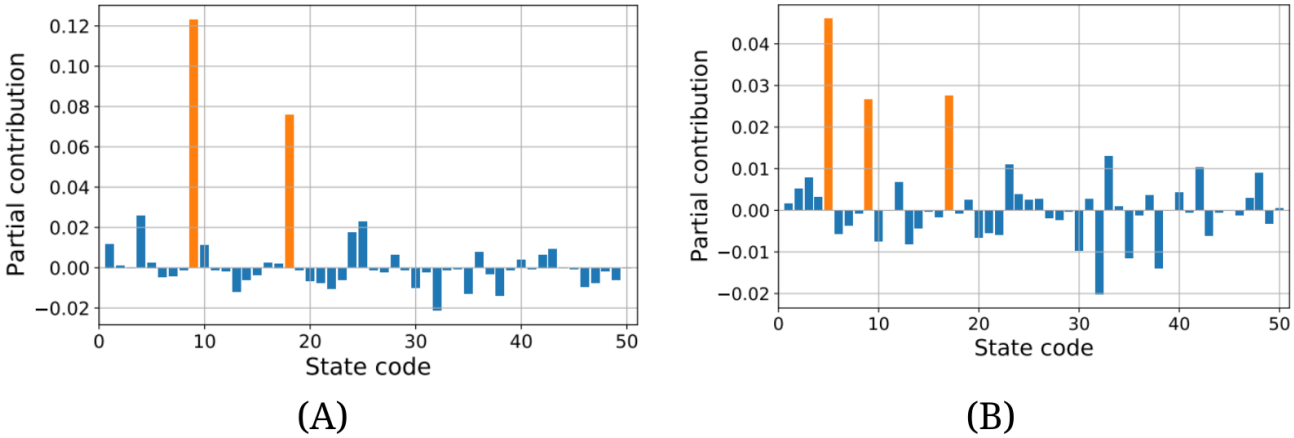


Figure 3. Partial contributions to the Theil index from each region, t_i, at d=60. The horizontal axis shows the state code given in Multimedia Appendix 1. (A) Comparison of the distribution of the confirmed cases and population. The vertical axis shows the ratio of a part to the whole region for populations and for the confirmed cases. (B) Contributions to the Theil index from each region. The vertical axis shows the strength of the contribution to the Theil index. The significantly high value of partial contribution to the Theil index is highlighted in orange.



XSL•FO
RenderX

Figure 4. Partial contributions of the Theil index from each region, t_i at a specific date. The vertical axis shows the strength of contribution to the Theil index. The horizontal axis shows the state code given in Multimedia Appendix 1. (A) Contributions of the Theil index at d=550. (B) Contributions of the Theil index at d=750. The significantly high values of partial contribution to the Theil index are highlighted in orange.



Discussion

This study demonstrates the utility of the Theil index for quantifying regional disproportionalities in the distribution of COVID-19 cases. It offers an intuitive and efficient approach to identifying hotspots and monitoring the spread of infection. However, certain limitations may affect result interpretation.

The accuracy of the analysis depends on data quality; factors such as underreporting, delays in case confirmation, and regional differences in testing capacity may introduce biases into case counts. These issues could potentially impact the calculated Theil index and the assessment of regional disproportionalities.

Additionally, this study focuses primarily on confirmed cases rather than new infections, limiting its capacity to predict future spread. Therefore, the Theil index alone may not be sufficient for determining the timing and location of public health interventions, such as isolation measures. To support comprehensive policy-making, it should be used alongside other indicators, such as infection rates, hospitalization rates, and health care capacity.

Conventional spatiotemporal analysis methods [20,21] are widely used in epidemiology and public health to track infectious disease spread and visualize infection clusters over time in specific regions. These established tools effectively detect geographical clusters, identify areas with unusually high incidence, and reveal disease hotspots within defined spatial ranges.

In contrast, our method offers two distinct advantages. First, an increase in the Theil index acts as a precursor to a surge in the number of infected individuals. Second, it quantifies regional disproportionalities in incidence rates at any given time. Unlike conventional methods that emphasize physical distance and spatial proximity, our approach treats regions as discrete categories to calculate incidence rate disproportionalities. Although simple, this approach provides an intuitive way to identify epicenters at a lower computational cost compared to spatiotemporal scanning, enabling us to detect early surges in confirmed cases and pinpoint regions with concentrated infections.

For instance, the concentration of COVID-19 cases in New York at d=60 as shown in Figure 3A and B, cannot be overlooked when considering infection control. The lockdown was implemented in New York City [22] and coincided with a period when the contribution to the Theil index was concentrated in New York State. Although it is challenging to assess the direct impact of lockdown using the Theil index alone, the timing appears appropriate based on the pattern of concentration of confirmed cases.

Integrating our method with additional data sources, such as mobility patterns and health care capacity, will enhance pandemic response strategies, particularly for early intervention and efficient resource allocation.

In conclusion, this study demonstrates the application of the Theil index in quantifying regional disproportionalities in confirmed cases and monitoring their evolution over time. By analyzing confirmed case data in the United States, we have identified patterns of disproportionalities, specified epicenters, and characterized localized outbreaks.

Continued monitoring and analysis of regional differences in COVID-19 transmission remain essential, especially considering emerging variants and evolving public health responses. Our findings highlight the importance of understanding the regional dynamics of infected individuals for effective pandemic response interventions.

Incorporating the findings of this study will help policy makers refine strategies and address the diverse needs of different regions, ultimately increasing the effectiveness of pandemic response efforts and mitigating the impact of future health crises.

Lastly, the decomposability of the Theil index makes it possible to quantify and compare disproportionality in groups with specific characteristics, such as age, vaccination coverage, and health care accessibility. Identifying these disproportionalities will provide important insights for future pandemic responses.

XSL•FO
RenderX

Acknowledgments

We thank the members of the Division of Scientific Information and Public Policy (SiPP) at the Center for Infectious Disease Education and Research (CiDER) Osaka University for useful discussions. This research was supported by The Nippon Foundation—Osaka University Project for Infectious Disease Prevention.

Data Availability

Data were derived from public resources.

Conflicts of Interest

None declared.

Multimedia Appendix 1
US states codes.
[PNG File, 58 KB - formative_v9i1e59230_app1.png]

References

1. Lopez L, Hart LH, Katz MH. Racial and ethnic health disparities related to COVID-19. JAMA 2021 Feb 23;325(8):719-720. [doi: 10.1001/jama.2020.26443] [Medline: 33480972]
2. Adhikari S, Pantaleo NP, Feldman JM, Ogedegbe O, Thorpe L, Troxel AB. Assessment of community-level disparities in coronavirus disease 2019 (COVID-19) infections and deaths in large US metropolitan areas. JAMA Netw Open 2020 Jul 1;3(7):e2016938. [doi: 10.1001/jamanetworkopen.2020.16938] [Medline: 32721027]
3. Perkins TA, España G. Optimal control of the COVID-19 pandemic with non-pharmaceutical interventions. Bull Math Biol 2020 Sep 4;82(9):118. [doi: 10.1007/s11538-020-00795-y] [Medline: 32888118]
4. Nakano T, Ikeda Y. Novel indicator to ascertain the status and trend of COVID-19 spread: modeling study. J Med Internet Res 2020 Nov 30;22(11):e20144. [doi: 10.2196/20144] [Medline: 33180742]
5. Buchy P, Buisson Y, Cintra O, et al. COVID-19 pandemic: lessons learned from more than a century of pandemics and current vaccine development for pandemic control. Int J Infect Dis 2021 Nov;112:300-317. [doi: 10.1016/j.ijid.2021.09.045] [Medline: 34563707]
6. Stenseth NC, Dharmarajan G, Li R, Shi ZL, Yang R, Gao GF. Lessons learnt from the COVID-19 pandemic. Front Public Health 2021;9:694705. [doi: 10.3389/fpubh.2021.694705] [Medline: 34409008]
7. Ikeda Y, Sasaki K, Nakano T. A new compartment model of COVID-19 transmission: the broken-link model. Int J Environ Res Public Health 2022 Jun 3;19(11):6864. [doi: 10.3390/ijerph19116864] [Medline: 35682447]
8. Pandemic preparedness. Nature. 2022 Oct 26. URL: https://www.nature.com/collections/jaacfgeief [accessed 2024-03-14]
9. Sasaki K, Ikeda Y, Nakano T. The effects of behavioral restrictions on the spread of COVID-19. Rep 2022;5(4):37. [doi: 10.3390/reports5040037]
10. Jinjarak Y, Ahmed R, Nair-Desai S, Xin W, Aizenman J. Accounting for global COVID-19 diffusion patterns, January-April 2020. Econ Disaster Clim Chang 2020;4(3):515-559. [doi: 10.1007/s41885-020-00071-2] [Medline: 32901228]
11. Mollalo A, Vahedi B, Rivera KM. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. Sci Total Environ 2020 Aug 1;728:138884. [doi: 10.1016/j.scitotenv.2020.138884] [Medline: 32335404]
12. Villanustre F, Chala A, Dev R, et al. Modeling and tracking Covid-19 cases using big data analytics on HPCC system platform. J Big Data 2021;8(1):33. [doi: 10.1186/s40537-021-00423-z] [Medline: 33614394]
13. Yue H, Hu T. Geographical detector-based spatial modeling of the COVID-19 mortality rate in the continental United States. Int J Environ Res Public Health 2021 Jun 25;18(13):6832. [doi: 10.3390/ijerph18136832] [Medline: 34202168]
14. Theil H. Economics and Information Theory: North-Holland Publishing Company: Amsterdam, North-Holland; 1967.
15. Manz KM, Mansmann U. Inequality indices to monitor geographic differences in incidence, mortality and fatality rates over time during the COVID-19 pandemic. PLoS ONE 2021;16(5):e0251366. [doi: 10.1371/journal.pone.0251366] [Medline: 33984055]
16. Kullback S, Leibler RA. On information and sufficiency. Ann Math Statist 1951 Mar;22(1):79-86. [doi: 10.1214/aoms/1177729694]
17. Holshue ML, DeBolt C, Lindquist S, et al. First case of 2019 novel coronavirus in the United States. N Engl J Med 2020 Mar 5;382(10):929-936. [doi: 10.1056/NEJMoa2001191] [Medline: 32004427]
18. COVID-19 data repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. Github. URL: https://github.com/CSSEGISandData/COVID-19 [accessed 2024-03-14]
19. National population totals and components of change: 2020-2023. US Census Bureau. URL: https://www.census.gov/data/datasets/time-series/demo/popest/2020s-national-total.html [accessed 2024-03-14]
20. Kulldorff M, Heffernan R, Hartman J, Assunção R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. PLoS Med 2005 Mar;2(3):e59. [doi: 10.1371/journal.pmed.0020059] [Medline: 15719066]

XSL•FO
RenderX

21. Desjardins MR, Hohl A, Delmelle EM. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: detecting and evaluating emerging clusters. Appl Geogr 2020 May;118:102202. [doi: 10.1016/j.apgeog.2020.102202] [Medline: 32287518]
22. Huang Y, Li R. The lockdown, mobility, and spatial health disparities in COVID-19 pandemic: a case study of New York City. Cities 2022 Mar;122:103549. [doi: 10.1016/j.cities.2021.103549] [Medline: 35125596]

Abbreviations

CSSE: Center for Systems Science and Engineering
KL: Kullback-Leibler

Edited by A Mavragani; submitted 06.04.24; peer-reviewed by A Inostrosa-Psijas, Z Guo; revised version received 04.11.24; accepted 06.11.24; published 03.01.25.

Please cite as:
Sasaki K, Ikeda Y, Nakano T
Quantifying the Regional Disproportionality of COVID-19 Spread: Modeling Study
JMIR Form Res 2025;9:e59230
URL: https://formative.jmir.org/2025/1/e59230
doi:10.2196/59230

© Kenji Sasaki, Yoichi Ikeda, Takashi Nakano. Originally published in JMIR Formative Research (https://formative.jmir.org), 3.1.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Increasing COVID-19 Testing and Vaccination Uptake in the Take Care Texas Community-Based Randomized Trial: Adaptive Geospatial Analysis

Kehe Zhang[1,2], MS; Jocelyn V Hunyadi[1,2], MPH; Marcia C de Oliveira Otto[3], PhD; Miryoung Lee[4], PhD; Zitong Zhang[1], MS, PhD; Ryan Ramphul[3], PhD; Jose-Miguel Yamal[1], PhD; Ashraf Yaseen[1], PhD; Alanna C Morrison[3], PhD; Shreela Sharma[3,5], PhD; Mohammad Hossein Rahbar[3,6,7], PhD; Xu Zhang[7], PhD; Stephen Linder[8], PhD; Dritana Marko[9], MSc, MD; Rachel White Roy[10], MPH, PhD; Deborah Banerjee[11], PhD; Esmeralda Guajardo[12], MA; Michelle Crum[13], PhD; Belinda Reininger[4], DrPH; Maria E Fernandez[14], PhD; Cici Bauer[1,2], MS, PhD

[1]Department of Biostatistics and Data Science, School of Public Health, The University of Texas Health Science Center at Houston, 1200 Pressler St., RAS-E819, Houston, TX, United States

[10]Harris County Public Health, Houston, TX, United States

[11]Houston Health Department, Houston, TX, United States

[12]Cameron County Public Health, San Benito, TX, United States

[13]Department of Preventive Medicine and Population Health, School of Medicine, The University of Texas at Tyler, Tyler, TX, United States

[14]Department of Health Promotion and Behavior Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, United States

[2]Center for Spatial-Temporal Modeling for Applications in Population Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, United States

[3]Department of Epidemiology, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, United States

[4]Department of Epidemiology, School of Public Health, The University of Texas Health Science Center at Houston, Brownsville, TX, United States

[5]Center for Health Equity, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, United States

[6]Biostatistics/Epidemiology/Research Design Component, Center for Clinical and Translational Sciences, The University of Texas Health Science Center at Houston, Houston, TX, United States

[7]Department of Internal Medicine, Division of Clinical and Translational Sciences, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX, United States

[8]Department of Management, Policy and Community Health, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, United States

[9]Department of Quantitative and Qualitative Health Sciences, University of Texas School of Public Health San Antonio, San Antonio, TX, United States

Corresponding Author:
Cici Bauer, MS, PhD
Department of Biostatistics and Data Science, School of Public Health, The University of Texas Health Science Center at Houston, 1200 Pressler St., RAS-E819, Houston, TX, United States

Abstract

Background: Geospatial data science can be a powerful tool to aid the design, reach, efficiency, and impact of community-based intervention trials. The project titled *Take Care Texas* aims to develop and test an adaptive, multilevel, community-based intervention to increase COVID-19 testing and vaccination uptake among vulnerable populations in 3 Texas regions: Harris County, Cameron County, and Northeast Texas.

Objective: We aimed to develop a novel procedure for adaptive selections of census block groups (CBGs) to include in the community-based randomized trial for the Take Care Texas project.

Methods: CBG selection was conducted across 3 Texas regions over a 17-month period (May 2021 to October 2022). We developed persistent and recent COVID-19 burden metrics, using real-time SARS-CoV-2 monitoring data to capture dynamic infection patterns. To identify vulnerable populations, we also developed a CBG-level community disparity index, using 12 contextual social determinants of health (SDOH) measures from US census data. In each adaptive round, we determined the priority CBGs based on their COVID-19 burden and disparity index, ensuring geographic separation to minimize intervention "spillover." Community input and feedback from local partners and health workers further refined the selection. The selected CBGs were then randomized into 2 intervention arms—multilevel intervention and just-in-time adaptive intervention—and 1 control arm, using covariate adaptive randomization, at a 1:1:1 ratio. We developed interactive data dashboards, which included

XSL•FO
RenderX

maps displaying the locations of selected CBGs and community-level information, to inform the selection process and guide intervention delivery. Selection and randomization occurred across 10 adaptive rounds.

Results: A total of 120 CBGs were selected and followed the stepped planning and interventions, with 60 in Harris County, 30 in Cameron County, and 30 in Northeast Texas counties. COVID-19 burden presented substantial temporal changes and local variations across CBGs. COVID-19 burden and community disparity exhibited some common geographical patterns but also displayed distinct variations, particularly at different time points throughout this study. This underscores the importance of incorporating both real-time monitoring data and contextual SDOH in the selection process.

Conclusions: The novel procedure integrated real-time monitoring data and geospatial data science to enhance the design and adaptive delivery of a community-based randomized trial. Adaptive selection effectively prioritized the most in-need communities and allowed for a rigorous evaluation of community-based interventions in a multilevel trial. This methodology has broad applicability and can be adapted to other public health intervention and prevention programs, providing a powerful tool for improving population health and addressing health disparities.

(JMIR Form Res 2025;9:e62802) doi:10.2196/62802

KEYWORDS

COVID-19 testing; COVID-19 vaccination; study design; community-based interventions; geospatial analysis; public health; social determinants of health; data dashboard

Introduction

The COVID-19 pandemic created a public health crisis, prompting discussions on how to best mitigate its far-reaching effects in a timely and efficient manner. From the start of the pandemic (March 2020) to May 2021 (the beginning of this study), there were over 30 million cumulative COVID-19 cases [1], over 2 million hospitalizations, and over half a million deaths in the United States alone [2]. During natural and health-related disasters (eg, pandemics and epidemics), vulnerable populations have historically endured a disproportionate level of negative health outcomes [3-7]. Extensive research has revealed similar impacts of the COVID-19 pandemic on vulnerable populations, particularly older adults, racial and ethnic minority groups, individuals who are immunocompromised, those with chronic conditions, and individuals with lower socioeconomic status and education attainment [8-17]. For example, a higher risk of severe COVID-19 symptoms, hospitalization, and mortality is associated with the presence of underlying medical conditions and older age [8,9]. Further, multiple studies have assessed population vulnerability through the Centers for Disease Control and Prevention (CDC) Social Vulnerability Index (SVI). For example, COVID-19 risk increases with the overall SVI score and the scores for specific SVI domains, notably those for the minority status and language domains [10,11,16]. Specifically, Hispanic and Black individuals have experienced higher rates of infection, hospitalization, and mortality when compared to non-Hispanic White individuals [12-14,18]. Racial and ethnic minority groups tend to be more susceptible to COVID-19, partially due to economic and social factors such as limited financial resources; insecure housing; higher housing density; reliance on public transportation; and additional inequities in nutrition, jobs, environments, and access to health care and health-related resources [13,14,19,20]. These findings underscore the importance and complex interplay of social determinants of health (SDOH) with respect to COVID-19 outcomes.

Health disparities research has consistently shown that geography—where individuals live, work, and play—matters, and it can greatly influence mortality, morbidity, life expectancy, and various aspects of health [21]. Geographic disparities in COVID-19 health outcomes have also been well documented. Previous studies have demonstrated county-level disparities in SARS-CoV-2 testing, infection, and fatality rates [22-24]. Research conducted by Das et al [25] reported geographic disparities in COVID-19 incidence across various zip code tabulation areas (ZCTAs) in the greater St. Louis area, Missouri. Das et al [25] also reported that the strength of the associations between the area proportion of the population working in agriculture and COVID-19 incidence varied by different ZCTAs. In a separate study, Bauer et al [26] developed a novel approach to identifying testing disparities by census block groups (CBGs), revealing significant geographical disparities in the context of smaller areas. Further, geographical disparities may not be constant and instead shift over time. For example, from 2019 to 2020, temporal trends in confirmed COVID-19 incidence and mortality differed between the most and least vulnerable counties in the United States [27]. Another study reported significant changes in the spatial-temporal patterns of the health care accessibility, test positivity rate of confirmed COVID-19 cases, and case fatality ratio in Texas [28].

Effective and real-time identification of vulnerable, at-risk populations that are characterized by diverse characteristics or reside within specific geographical regions is essential to address and eliminate disparities in COVID-19 health outcomes. In the context of health care, where resources are often limited and must be efficiently and equitably allocated, questions regarding site or population selection become paramount. Achieving this goal requires (1) the identification of metrics to accurately measure the multifaceted nature of disparities with available datasets and (2) a robust method to summarize all metrics and provide a meaningful output with practical use. The latter has been achieved through various approaches, such as assigning weights to each metric, which are often provided by domain experts, and constructing an index from the weighted sum of all metrics.

XSL•FO
RenderX

This paper presents an innovative protocol for adaptive geospatial selections in a community-based intervention program—the Take Care Texas project—funded by the National Institutes of Health (NIH) Rapid Access to Diagnostics for Underserved Populations (RADx-UP) program. Our protocol leveraged a data-driven approach for the rapid adaptation and deployment of multilevel, just-in-time adaptive interventions to improve COVID-19 testing and vaccination uptake in Texas. In the *Methods* section, we focus on 2 key areas. First, we outline the criteria and metrics for identifying the vulnerable populations via SDOH and geographical measures. Second, we detail the methodology for selecting the local communities based on CBGs, which involved using a phased, multistage approach. Designed for real-time prioritization, our protocol addresses the immediate needs of communities most affected by the changing landscape of the COVID-19 crisis.

Methods

Study Region and Populations

Our Take Care Texas project includes 3 geographically distinct sites within Texas, each with unique demographic characteristics (Multimedia Appendix 1). Southeast Texas, represented by the Houston area within Harris County, has an ethnically diverse population (N=4.73 million), covering both urban and suburban communities. This region serves as a significant residence for various racial and ethnic minority groups and is recognized for having one of the largest metropolitan economies in the United States. The South Texas site, represented by Cameron County (population: N=423,029) in the Rio Grande Valley, is primarily Hispanic, with a considerable proportion of the population (22.6%) living below the federal poverty line, exceeding the national average (11.5%) [29]. The Northeast Texas (NETX) region covers an extensive area, including the Smith, Gregg, Henderson, Van Zandt, Anderson, Wood, and Rains counties. NETX is characterized by a mix of small cities and rural communities, with a lower population density when compared to other sites. Anderson County was excluded from this study due to insufficient SARS-CoV-2 infection data coverage at the time of analysis. The demographic distribution across these regions reflects a wide socioeconomic and cultural spectrum, offering a comprehensive understanding of the diverse population within Texas. There are 2144 CBGs in Harris County, 222 CBGs in Cameron County, and 349 CBGs from the NETX region, according to the 2010 census data.

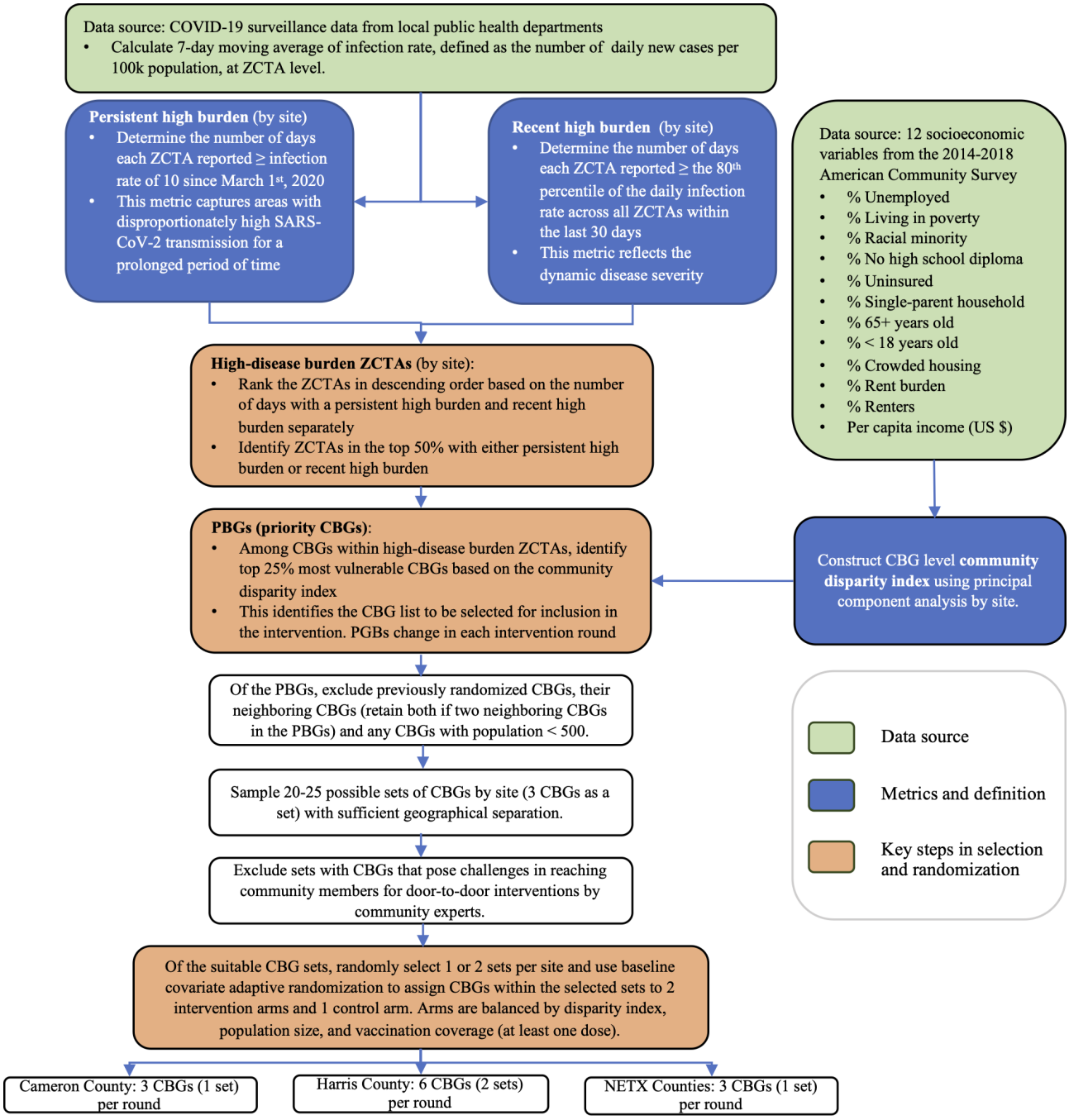
Overview of Priority CBG Selection Process

The process for identifying priority CBGs (PBGs) in this study is presented in Figure 1. The process began with the identification of geographic areas experiencing significant COVID-19 burden, using SARS-CoV-2 surveillance data and population estimates at the ZCTA level. We then identified CBGs within these areas that exhibited high socioeconomic disparities, using the community disparity index we constructed to achieve the desired spatial granularity. Detailed descriptions for determining disease burden and community disparities are provided in the *Persistent and Recent COVID-19 Burdens* and *CBG-Level Community Disparity Index* sections.

In response to the dynamic nature of the COVID-19 pandemic, our identification and selection process was conducted monthly with a near–real-time data feed from the local public health department for timely and effective intervention planning. Table 1 illustrates the stepped planning and intervention calendar. Between May 2021 and October 2022, a total of 10 rounds of CBG selection were conducted for each of the three study sites. In each round, among the CBGs within the identified high–disease burden ZCTAs, we selected the most vulnerable CBGs based on the community disparity index (ie, CBGs in the top 25% for this metric). We then excluded previously randomized CBGs, their neighbors (ie, CBGs that share a common geographic boundary), and any CBGs with small populations (total population<500) to form the PBG list, which served as the sampling universe. From the PBGs, we sampled geographically nonadjacent sets of CBGs (3 CBGs as a set), ensuring suitable geographic separation to mitigate potential intervention "spillover," since the study goal was to compare 2 distinct intervention strategies to a control arm. We sought feedback from the community health workers and staff in the three regions, to determine if the potential CBGs presented safety, geography, and access concerns, by showing maps of the areas and then discussing the areas and documenting concerns. When there was lack of clarity about an area, we reached out via phone calls and emails to community experts in our advisory groups (health department leaders, city leaders, and local nonprofits serving the areas) to obtain their feedback as well. CBGs that posed challenges in reaching community members (such as CBGs with primarily gated communities or business districts), which would prevent implementing the intervention protocol (eg, door-to-door visits), were excluded from the sampling frame. This collaborative approach ensured that our selection process was informed by both data-driven analysis and community and implementer insights. The final selected CBGs within each set were then assigned to 1 of 2 intervention groups (multilevel intervention and just-in-time adaptive intervention) or a control group through baseline covariate adaptive randomization, balancing disparity index and population size, as well as vaccination coverage when such data were available. These CBGs then underwent a 1-month planning phase, followed by a 2-month intervention period. In each round, this study selected and randomized 6 CBGs (2 sets) for Harris County due to its larger geographic area and population, 3 CBGs for Cameron County, and 3 CBGs for NETX counties.

XSL•FO
RenderX

Figure 1. Workflow of the adaptive geospatial selection of CBGs for intervention in the Take Care Texas study. The selection of CBGs for intervention was conducted each round. CBG: census block group; NETX: Northeast Texas; PBG: priority census block group; ZCTA: zip code tabulation area.



XSL•FO
RenderX

Table . The stepped planning and intervention calendar by month. Each round, selected census block groups were randomized into 1 control arm and 2 intervention arms—multilevel intervention (MILI) and just-in-time adaptive intervention (JITAI).

	2021								2022					
	May	June	July	August	September	October	November	December	January	February	March	April	May	June
Round 1														
Control	✓[a]	✓[b]	✓[b]											
MILI	✓[a]	✓[b]	✓[b]											
JITAI	✓[a]	✓[b]	✓[b]											
Round 2														
Control		✓[a]	✓[b]	✓[b]										
MILI		✓[a]	✓[b]	✓[b]										
JITAI		✓[a]	✓[b]	✓[b]										
Round 3														
Control			✓[a]	✓[b]	✓[b]									
MILI			✓[a]	✓[b]	✓[b]									
JITAI			✓[a]	✓[b]	✓[b]									
Round 4														
Control				✓[a]	✓[b]	✓[b]								
MILI				✓[a]	✓[b]	✓[b]								
JITAI				✓[a]	✓[b]	✓[b]								
Round 5														
Control					✓[a]	✓[b]	✓[b]							
MILI					✓[a]	✓[b]	✓[b]							
JITAI					✓[a]	✓[b]	✓[b]							
Round 6														
Control						✓[a]	✓[b]	✓[b]						
MILI						✓[a]	✓[b]	✓[b]						
JITAI						✓[a]	✓[b]	✓[b]						
Round 7														
Control									✓[a]	✓[b]	✓[b]			
MILI									✓[a]	✓[b]	✓[b]			
JITAI									✓[a]	✓[b]	✓[b]			
Round 8														
Control										✓[a]	✓[b]	✓[b]		
MILI										✓[a]	✓[b]	✓[b]		
JITAI										✓[a]	✓[b]	✓[b]		

XSL•FO
RenderX

	2021								2022					
	May	June	July	August	September	October	November	December	January	February	March	April	May	June
Round 9														
Control											✓[a]	✓[b]	✓[b]	
MILI											✓[a]	✓[b]	✓[b]	
JITAI											✓[a]	✓[b]	✓[b]	
Round 10														
Control												✓[a]	✓[b]	✓[b]
MILI												✓[a]	✓[b]	✓[b]
JITAI												✓[a]	✓[b]	✓[b]

[a]Planning months.

[b]Intervention months.

Persistent and Recent COVID-19 Burdens

The persistent and recent high COVID-19 burdens were quantified by using SARS-CoV-2 surveillance data obtained from local public health departments. We computed a 7-day moving average for infection rate, which was defined as the daily new cases per 100,000 population. For each ZCTA, the persistent disease burden was evaluated based on the number of days on which ≥10 new cases per 100,000 population were reported, as per the common COVID-19 risk level threshold [30], starting from March 1, 2020, up to the respective analysis date for each selection round. To evaluate the recent disease burden, we first calculated the 80th percentile of daily new cases per 100,000 population across all ZCTAs within each study site, using data from the 30 days preceding each analysis date. We then counted the number of days on which the infection rate for each ZCTA exceeded this 80th percentile threshold. In each selection round and for each site, we ranked the ZCTAs, in descending order, based on their number of days with (1) a persistent high burden and (2) a recent high burden. The ranking was performed for each metric separately, and the ZCTAs in the top 50% for either metric were identified as high–disease burden ZCTAs.

CBG-Level Community Disparity Index

We developed the CBG-level community disparity index due to the lack of spatial granularity in existing indices, such as the SVI, which did not meet our study's needs. Using 12 SDOH measures from the 2014 - 2018 American Community Survey 5-year estimates [31], our community disparity index covered the socioeconomic, demographic, and housing dimensions of SDOH (Multimedia Appendix 2). It integrated economic stability factors (employment and poverty rates, per capita income, education level, and health insurance coverage), demographic characteristics (percentages of individuals younger than 18 years, individuals older than 65 years, individuals in single-parent households, and racial minority individuals), and housing factors (percentage of renter-occupied units, rent burden, and crowded housing).

We used principal component analysis (PCA) to construct the community disparity index, similar to the approach used by Kolak et al [32]. PCA is a data-driven approach that synthesizes multiple pieces of information into one combined measure, while simultaneously handling high correlation and minimizing information loss [32-35]. We kept the principal components with eigenvalues exceeding 1, following the Kaiser criteria. The scores of these selected principal components were weighted according to the variance they accounted for and then summed, with higher scores representing a greater community disparity. The resulting indices were then ranked and scaled from 0 to 1 within each study region, providing a relative ranking of CBGs. All data were processed and analyzed with R statistical software (v4.1.2; R Foundation for Statistical Computing) [36].

Interactive Data Dashboards

We developed ArcGIS (Environmental Systems Research Institute Inc) and R Shiny (Posit PBC) data dashboards that provided key geospatial information and relevant COVID-19 data for this study. Feedback was received from the dashboard's end users, including community health workers. The dashboards included maps that showed community disparity indices, PBGs, intervention-prioritized CBGs, COVID-19 testing sites, road maps, and SARS-CoV-2 infection and testing rates at the CBG and ZCTA levels. These tools enabled community health workers to identify the boundaries of the priority areas and access the relevant COVID-19 data for effective on-ground intervention messaging and referral to needed services. Both dashboards were designed to be interactive to facilitate strategic planning efforts and were password protected for internal team use only. The dashboard was especially useful during the planning phase and facilitated the creation of a community profile document, which community health workers used to share real-time data and deliver messaging tailored to the communities.

Ethical Considerations

The research protocol was approved by the UTHealth Committee for the Protection of Human Subjects

XSL•FO
RenderX

(HSC-SPH-20 - 1372). Data user agreements were established between UTHealth and the public health departments of Harris County, Cameron County, and NETX counties. This study did not involve direct contact with participants, as data were aggregated at the CBG level, so no individual-level informed consent was required. Data were deidentified, and all analyses followed data privacy guidelines.

Results

Table 2 presents summary statistics of the population density and 12 SDOH variables used to construct the community disparity index for CBGs in Harris County, Cameron County, and NETX counties, in comparison to the whole state of Texas. Harris County, with the highest population density (2600 per km^2) and per capita income (US $33,500), also has the highest percentage of renters (44.9%). Cameron County, on the other hand, has a higher percentage of crowded housing (11.2%), individuals without high school diplomas (35.1%), uninsured individuals (28.0%), people living in poverty (30.9%), and racial minority individuals (90.0%). NETX has a higher percentage of people aged 65 years and older (18.2%) but a much lower proportion of racial minority individuals compared to Texas overall (35% vs 57.7%).

Between March 2021 and March 2023, we identified 5 distinct COVID-19 waves, with 3 occurring within the intervention period of our study, specifically around August 2021, December 2021, and July 2022 (Multimedia Appendix 3). The overall infection rates varied across the three study regions, highlighting the importance of site-specific burden assessment. The temporal trends of SARS-CoV-2 infection rates differed among the sites when the pandemic started but became more similar during the intervention period, that is, from May 2021 to October 2022. Harris County consistently reported higher infection rates than those reported by the other two regions throughout this study.

Figure 2 displays maps of the CBG-level community disparity index, highlighting substantial variations across and within the study regions. The leading four PCA components captured a significant portion of the variance in the 12 SDOH measures. For instance, in Harris County, these components explained 70.96% of the variance, with the first principal component alone accounting for 43.45%, and were primarily influenced by socioeconomic factors (percentage of individuals with no high school diploma, percentage of individuals living under poverty, per capita income, percentage of uninsured individuals, and percentage of racial minority individuals). Consequently, we derived the standardized index scores by using the first four PCA components.

The maps of the COVID-19 burden and community disparity index shared similar geographic patterns, with high COVID-19 burdens often found in ZCTAs with elevated community disparity (Multimedia Appendix 4). However, distinctive patterns also existed. For instance, during the first selection round for Harris County in May 2021, several ZCTAs located in the western and eastern parts of the county experienced high COVID-19 burden, despite their low community disparity scores. This highlights the complex interplay between the pandemic impact and SDOH, as COVID-19 can affect diverse communities, including those traditionally considered less socially vulnerable. Therefore, it is crucial to leverage both SDOH data and COVID-19 surveillance data to dynamically identify and prioritize areas for interventions.

Over the 17-month period from May 2021 to October 2022, we selected and randomized 120 CBGs for intervention across 10 rounds, with 60 CBGs in Harris County, 30 in Cameron County, and 30 in NETX counties (Table 1). The population size of the randomized CBGs ranged from 615 to 11,321. The combination of the community disparity index and real-time COVID-19 surveillance data allowed for the identification and selection of areas—those with the highest disease burden among the most vulnerable populations—for the adaptive, multilevel, just-in-time intervention. This study also observed substantial changes in the PBG list across different rounds, with more CBGs identified in the later rounds of selection, corresponding to increased infection rates during the pandemic, as shown in Figure 3.

Multimedia Appendix 5 illustrates the R Shiny and ArcGIS dashboards that we developed to provide a dynamic visualization of the data from this study. These interactive dashboards, along with the community profiles we developed, were used by the intervention team to tailor their strategies. By leveraging these tools, the team was able to adapt culturally relevant messages and materials, enhance door-to-door education by community health workers in selected CBGs, and better inform social marketing campaigns to promote COVID-19 testing and vaccination efforts.

XSL•FO
RenderX

Table . Summary statistics of the American Community Survey variables (2014-2018 5-year estimates) [31] used in community disparity index creation at the census block group (CBG) level for 3 regions in Texas and Texas overall.

	Texas regions			Texas overall
	Cameron County (CBGs: n=222)	Harris County (CBGs: n=2144)	NETX[a] counties (CBGs: n=349)	
Population (n)				27,900,000
Mean (CV[b] [%])	1900 (83.7)	2150 (83.5)	1540 (48.6)	—[c]
Median (IQR)	1550 (967-2350)	1770 (1210-2470)	1370 (1060-1890)	—
Population density (per km^2)				43.1
Mean (SD [%])	1420 (86.5)	2600 (97.5)	546 (136.5)	—
Median (minimum, maximum)	1180 (345, 2130)	2020 (1180, 3090)	209 (41.4, 870)	—
Crowded housing (%)				4.79
Mean (CV [%])	11.2 (74.0)	6.28 (123.8)	3.45 (141.4)	—
Median (IQR)	9.84 (4.82-16.3)	3.68 (0-9.46)	1.82 (0-5.24)	—
Individuals without a high school diploma (%)				16.8
Mean (CV [%])	35.1 (45.1)	21.5 (84.1)	16.7 (74.3)	—
Median (IQR)	36.0 (23.2-47.8)	17.0 (5.65-34.5)	14.1 (7.93-21.5)	—
Renters (%)				38.1
Mean (CV [%])	36.0 (56.4)	44.9 (65.1)	32.2 (64.9)	—
Median (IQR)	34.0 (20.9-46.1)	40.2 (20.5-66.5)	27.4 (17.2-43.7)	—
Rent burden (%)				44.4
Mean (CV [%])	46.8 (45.9)	43.7 (52.2)	39.7 (61.6)	—
Median (IQR)	46.4 (33.3-60.9)	45.5 (29.0-58.2)	39.1 (22.2-56.3)	—
Individuals younger than 18 years (%)				26.2
Mean (CV [%])	29.1 (29.6)	25.3 (36.2)	23.7 (36.4)	—
Median (IQR)	30.1 (23.6-35.4)	25.8 (19.5-31.5)	23.0 (18.2-29.0)	—
Individuals older than 65 years (%)				12.0
Mean (CV [%])	15.5 (69.0)	11.3 (66.7)	18.2 (53.7)	—
Median (IQR)	13.2 (9.67-18.4)	9.99 (6.28-14.5)	17.2 (11.0-23.5)	—
Individuals in single-parent households (%)				14.7
Mean (CV [%])	18.3 (62.5)	17.2 (83.3)	14.0 (92.8)	—
Median (IQR)	17.1 (9.94-25.8)	14.4 (6.23-25.3)	10.8 (4.72-20.9)	—
Unemployed individuals (%)				5.41
Mean (CV [%])	7.05 (91.1)	6.45 (91.4)	5.65 (103.5)	—
Median (IQR)	5.85 (2.60-9.76)	5.16 (2.54-8.89)	4.39 (1.47-7.87)	—
Per capita income (US $)				30,100
Mean (CV [%])	16,400 (45.3)	33,500 (78.2)	25,800 (41.8)	—
Median (IQR)	14,500 (10,700-20,200)	24,200 (16,700-40,700)	25,400 (17,600-30,700)	—
Uninsured individuals (%)				17.1
Mean (CV [%])	28.0 (38.4)	20.9 (67.3)	17.6 (54.8)	—
Median (IQR)	28.4 (20.2-35.3)	19.5 (9.25-30.9)	16.0 (10.6-23.4)	—
Individuals living in poverty (%)				14.3
Mean (CV [%])	30.9 (46.9)	16.2 (81.2)	16.6 (71.9)	—

XSL•FO
RenderX

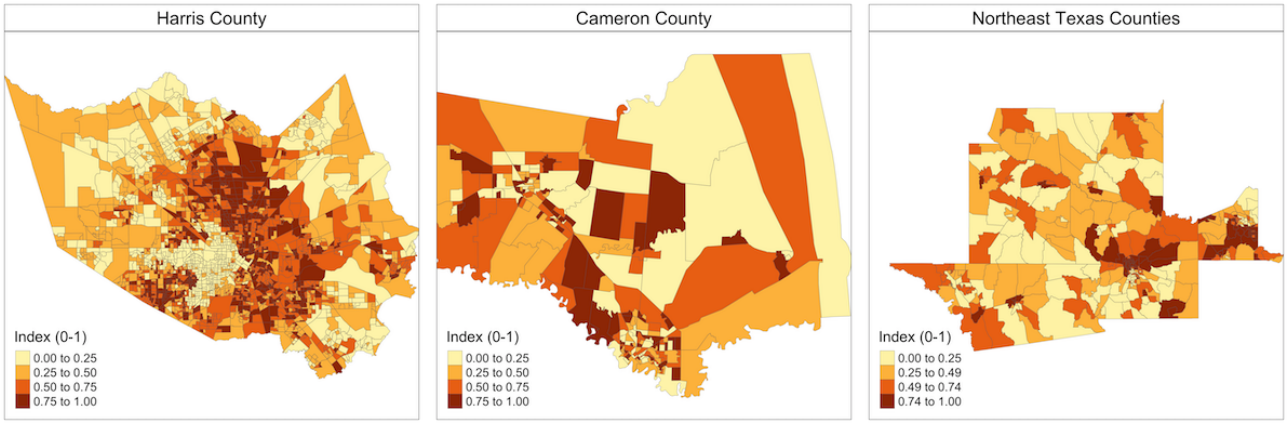
	Texas regions			Texas overall
	Cameron County (CBGs: n=222)	Harris County (CBGs: n=2144)	NETX[a] counties (CBGs: n=349)	
Median (IQR)	31.1 (19.1-42.1)	13.2 (5.63-23.9)	14.5 (7.93-22.4)	—
Racial minority individuals (%)				57.7
Mean (CV [%])	90.0 (15.6)	69.3 (40.0)	35.0 (77.2)	—
Median (IQR)	94.2 (87.3-98.4)	79.7 (46.5-94.0)	26.1 (14.4-50.9)	—

[a]NETX: Northeast Texas.

[b]CV: coefficient of variation.

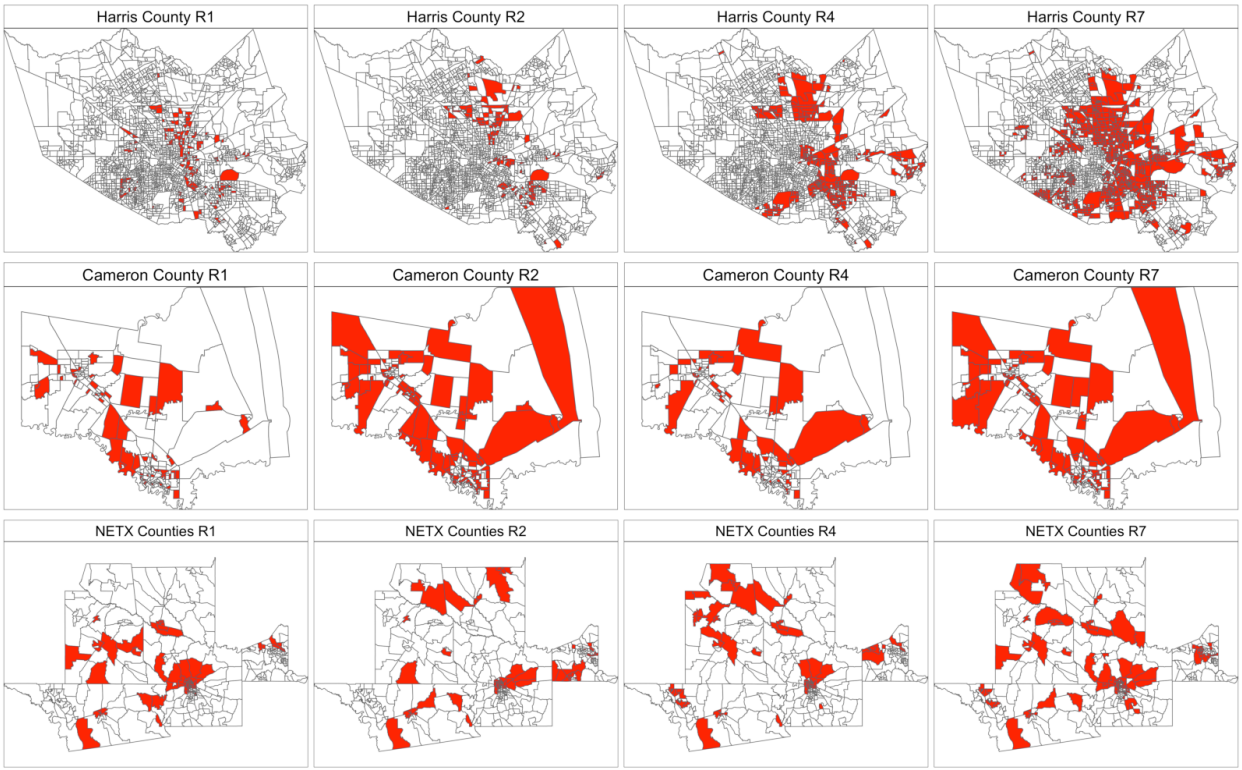
[c]Not applicable.

Figure 2. Census block group–level community disparity index developed for the community-based intervention by study region. The community disparity index was constructed by using principal component analysis to combine 12 social determinants of health measures from the American Community Survey 2014-2018 5-year estimates [31]. The index was scaled to range from 0 to 1 within each study area, with higher index values indicating greater community disparity.



XSL•FO
RenderX

Figure 3. Priority CBGs for the community-based intervention by study region and selection rounds: highlighted CBGs were identified as candidates for inclusion in R1, R2, R4, and R7. CBG: census block group; NETX: Northeast Texas; R: round.



Discussion

Principal Findings

In this study, we used geospatial data science methods to develop a protocol for the adaptive selection of 120 CBGs in 3 regions of Texas, which occurred across 10 rounds over a 17-month period (May 2021 to October 2022). This protocol, which was developed for the Take Care Texas project, integrated real-time dynamic data on SARS-CoV-2 infection rates and contextual SDOH to identify and prioritize the CBGs most in need of tailored activities and resources to improve COVID-19 testing and vaccination. The adaptive selection of CBGs prioritized the most vulnerable communities and allowed for the rigorous evaluation of community-based interventions in a multilevel trial.

Strengths and Limitations

Our protocol demonstrated several unique strengths in integrating public health surveillance data and geospatial data science. First, our study offered spatial granularity at the level of CBGs, which has rarely been seen in other studies [11,16,37]. Such granularity could potentially result in more tailored interventions that are best suited to the local communities for which they were developed. Second, we incorporated both the reported SARS-CoV-2 infection rates and the community SDOH, reflecting the socioeconomic and housing conditions that contributed to health disparities. This allowed us to prioritize communities experiencing not only high disease burden but also high social vulnerability that could potentially impact community resilience and the ability to adapt to challenges presented by the pandemic. Third, the dynamic nature of our protocol, particularly the initial selection of high–disease burden ZCTAs, allowed for real-time adaptation of the selection strategies in response to the changing pandemic, thereby enabling us to prioritize the most in-need and most vulnerable communities effectively. Finally, the protocol was strengthened by the input from and collaboration among community members and frontline community health workers, which were critical in designing and implementing interventions. The final CBG selection was made by community partners considering on-the-ground intervention feasibility. In the context of this study and site selection, the input from our community partners and community health workers supplemented the data-driven approaches [38]. Although multicriteria decision-making methods have been previously used for site selection problems [39-42], including the integration of objective criteria with experts' qualitative input to determine site suitability, some multicriteria decision-making techniques have been critiqued for arbitrary and irrational rankings of alternatives [43] and questionable mathematical validity [44]. By contrast, previous

XSL•FO
RenderX

work has shown that PCA is a robust approach for integrating various vulnerability and health indicators and providing reliable unequal weighting [33,34] to help guide the selection of local communities for health care interventions. Future studies should continue to investigate best practices for engaging communities in trial design and implementation.

Our protocol also presented several limitations. First, the COVID-19 pandemic highlighted significant challenges in data reporting due to the absence of established surveillance data infrastructures for COVID-19 and other rapidly spreading communicable diseases. The state reporting requirements for data elements and definitions changed over time, complicating the consistent use of data across study regions. In our study, we directly collaborated with individual county public health departments and processed and harmonized data across different study regions, which were time-consuming processes. Second, the time lag commonly seen in surveillance databases posed challenges in accurately identifying current hot spots, and incomplete demographic data hindered the assessment of health equity throughout the COVID-19 pandemic by delaying the analysis of COVID-19 health outcomes by subgroups and across more granular geographic scales [45,46]. The Congressional Research Service's 2020 report highlighted the need to modernize public health reporting systems to improve the timeliness and accuracy of public health data reporting [45]. However, differences in data standards, technical capacity, and regional policy priorities, particularly in the early stages of the COVID-19 pandemic, limited effective data sharing across study regions. Third, our protocol required accurate and complete address information for geographical granularity in identifying the high–disease burden priority areas, and when address information was missing, we may not have been able to include the SARS-CoV-2 case records in determining the PBGs. Finally, finding geographically separated CBGs in later rounds of selection could be difficult in small counties with fewer CBGs, such as Cameron County. Traditionally, administrative boundaries are used to define spatial units, yet alternative approaches may better capture actual community layouts. For example, Tuson et al [47] developed the Spatial Targeting Algorithm to generate flexible priority regions not restricted by administrative boundaries. However, while flexible boundaries may better capture neighborhoods, they may not provide SDOH information, which is typically accessible only at the administrative level. Additional work is needed to develop more flexible spatial units, with or without administrative boundaries as a constraint, for population-based research.

Study Implications

Despite these challenges, our study provided valuable protocols for data processing and analytical support to an intervention team working with local public health departments and organizations. It demonstrated a useful approach to ensuring that the most vulnerable communities participated in a multilevel intervention trial for increasing COVID-19 testing and vaccination uptake. Our proposed protocol could be adapted to other population-based randomized trials for which rapid assessment and allocation of finite resources are imperative. For example, geospatial approaches have been used to assess cancer screening rates among a vulnerable population in Ontario, Canada [48] and determine priority clinics to allocate funding for improving up-to-date colorectal cancer screening rates [49]. The integration of PCA and the real-time nature of our protocol can also be adapted to other population-based trials. When tackling multifaceted and complex problems, our protocol demonstrates that PCA is effective at robustly incorporating a wide range of health indicators. Further, the real-time nature of our protocol allows for ongoing, continuous assessment and adjustment in allocation of resources, provided that accurate and timely data are available. This could be applied to study recruitment protocols or other intervention-based trials.

Conclusion

The selection of optimal sites for health care and public health interventions, especially under limited resources, is a crucial challenge. Our protocol contributes to existing research by integrating disease surveillance and SDOH data, the robust PCA method, and community engagement to identify locations where COVID-19 testing interventions for vulnerable populations are most needed. The proposed data-driven approach could be adapted to other disease control and prevention programs to improve population health and reduce disease burden in areas with disparities.

Acknowledgments

This work was supported by the National Institutes of Health (NIH) Rapid Access to Diagnostics for Underserved Populations (RADx-UP) initiative (NIH 3UL1TR003167-02S1, UL1TR003167-02S6, and U01TR004355). This project is partially supported by UTHealth Houston Center for Clinical and Translational Sciences/Clinical and Translational Awards. We acknowledge the important contributions of the community health workers and staff of Take Care Texas in providing expertise, valuable feedback, and guidance during the revision. These staff include Ibette Cavazos, Lisa Mitchell-Bennett, Maribel Sifuentes, Carol Killingley, Carlton Allen, and LuBeth Perez. We also thank all the officials of the counties for providing their support to the Take Care Texas project.

Authors' Contributions

MCOO, ML, ACM, SS, MC, BR, MEF, and CB conceptualized the study design. KZ, MCOO, ML, ZZ, RR, JMY, AY, SL, DM, RWR, DB, EG, MC, and CB collected and processed the data. KZ, JVH, MCOO, ZZ, and CB analyzed and visualized the data. KZ, JVH, and CB wrote the initial draft of the manuscript. KZ, JVH, MCOO, ML, JMY, ACM, SS, MHR, XZ, DM, MC, BR, MEF, and CB edited and revised the manuscript. ML, MC, BR, MEF, and CB supervised the study and acquired funding.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Map of Texas counties with 3 study regions highlighted. Harris County (yellow) has the most ethnically diverse population in Texas. Cameron County (blue), located on the southern border of Texas, is primarily Hispanic. A large proportion of the population in Northeast Texas (NETX; orange) is White.
[PNG File, 494 KB - formative_v9i1e62802_app1.png]

Multimedia Appendix 2
Metrics for identifying vulnerable populations, definitions, data sources, and available spatial scale.
[DOCX File, 24 KB - formative_v9i1e62802_app2.docx]

Multimedia Appendix 3
Temporal trends of weekly COVID-19 cases per 100,000 population in Harris County, Cameron County, and Northeast Texas (NETX) counties, with vertical lines indicating the start and end of the planning and intervention periods for the Take Care Texas project. Historical Texas COVID-19 data were obtained from the Texas Department of State Health Services.
[PNG File, 350 KB - formative_v9i1e62802_app3.png]

Multimedia Appendix 4
Maps of the census block group–level community disparity index developed in this protocol, overlaid with high–disease burden zip code tabulation areas (shown in red geographical boundaries) in the first round of selection (May 2021) for the three study regions.
[PNG File, 1603 KB - formative_v9i1e62802_app4.png]

Multimedia Appendix 5
Screenshots of the (A) R shiny dashboard and (B) ArcGIS dashboard developed for the Take Care Texas study. These data dashboards were designed to be interactive to best facilitate strategic planning efforts and were password protected for internal team use only.
[PDF File, 233 KB - formative_v9i1e62802_app5.pdf]

References

1. WHO COVID-19 dashboard. World Health Organization. URL: https://covid19.who.int [accessed 2023-11-29]
2. COVID data tracker. Centers for Disease Control and Prevention. URL: https://covid.cdc.gov/covid-data-tracker [accessed 2023-11-29]
3. Paramanathan P, Abbas M, Huda SA, Huda S, Mortazavi M, Taravati P. Comparing racial health disparities in pandemics a decade apart: H1N1 and COVID-19. Future Healthc J 2021 Nov;8(3):e722-e728. [doi: 10.7861/fhj.2021.0030] [Medline: 34888475]
4. Quinn SC, Kumar S, Freimuth VS, Musa D, Casteneda-Angarita N, Kidwell K. Racial disparities in exposure, susceptibility, and access to health care in the US H1N1 influenza pandemic. Am J Public Health 2011 Feb;101(2):285-293. [doi: 10.2105/AJPH.2009.188029] [Medline: 21164098]
5. D'Adamo A, Schnake-Mahl A, Mullachery PH, Lazo M, Diez Roux AV, Bilal U. Health disparities in past influenza pandemics: a scoping review of the literature. SSM Popul Health 2023 Mar;21:101314. [doi: 10.1016/j.ssmph.2022.101314] [Medline: 36514788]
6. Howell J, Elliott JR. Damages done: the longitudinal impacts of natural hazards on wealth inequality in the United States. Soc Probl 2019 Aug 14;66(3):448-467. [doi: 10.1093/socpro/spy016]
7. Climate change and social vulnerability in the United States: a focus on six impacts. United States Environmental Protection Agency. 2021. URL: www.epa.gov/cira/social-vulnerability-report [accessed 2023-11-28]
8. Sanyaolu A, Okorie C, Marinkovic A, et al. Comorbidity and its impact on patients with COVID-19. SN Compr Clin Med 2020;2(8):1069-1076. [doi: 10.1007/s42399-020-00363-4] [Medline: 32838147]
9. Souch JM, Cossman JS. A commentary on rural-urban disparities in COVID-19 testing rates per 100,000 and risk factors. J Rural Health 2021 Jan;37(1):188-190. [doi: 10.1111/jrh.12450] [Medline: 32282964]
10. Karaye IM, Horney JA. The impact of social vulnerability on COVID-19 in the U.S.: an analysis of spatially varying relationships. Am J Prev Med 2020 Sep;59(3):317-325. [doi: 10.1016/j.amepre.2020.06.006] [Medline: 32703701]
11. Khazanchi R, Beiter ER, Gondi S, Beckman AL, Bilinski A, Ganguli I. County-level association of social vulnerability with COVID-19 cases and deaths in the USA. J Gen Intern Med 2020 Sep;35(9):2784-2787. [doi: 10.1007/s11606-020-05882-3] [Medline: 32578018]

XSL•FO
RenderX

12. Millett GA, Jones AT, Benkeser D, et al. Assessing differential impacts of COVID-19 on black communities. Ann Epidemiol 2020 Jul;47:37-44. [doi: 10.1016/j.annepidem.2020.05.003] [Medline: 32419766]
13. Aleligne YK, Appiah D, Ebong IA. Racial disparities in coronavirus disease 2019 (COVID-19) outcomes. Curr Opin Cardiol 2021 May 1;36(3):360-366. [doi: 10.1097/HCO.0000000000000847] [Medline: 33657019]
14. Mackey K, Ayers CK, Kondo KK, et al. Racial and ethnic disparities in COVID-19-related infections, hospitalizations, and deaths : a systematic review. Ann Intern Med 2021 Mar;174(3):362-373. [doi: 10.7326/M20-6306] [Medline: 33253040]
15. Upshaw TL, Brown C, Smith R, Perri M, Ziegler C, Pinto AD. Social determinants of COVID-19 incidence and outcomes: a rapid review. PLoS One 2021 Mar 31;16(3):e0248336. [doi: 10.1371/journal.pone.0248336] [Medline: 33788848]
16. Biggs EN, Maloney PM, Rung AL, Peters ES, Robinson WT. The relationship between social vulnerability and COVID-19 incidence among Louisiana census tracts. Front Public Health 2021 Jan 20;8:617976. [doi: 10.3389/fpubh.2020.617976] [Medline: 33553098]
17. Bauer C, Zhang K, Lee M, et al. Census tract patterns and contextual social determinants of health associated with COVID-19 in a Hispanic population from South Texas: a spatiotemporal perspective. JMIR Public Health Surveill 2021 Aug 5;7(8):e29205. [doi: 10.2196/29205] [Medline: 34081608]
18. Tai DBG, Shah A, Doubeni CA, Sia IG, Wieland ML. The disproportionate impact of COVID-19 on racial and ethnic minorities in the United States. Clin Infect Dis 2021 Feb 16;72(4):703-706. [doi: 10.1093/cid/ciaa815] [Medline: 32562416]
19. Caraballo C, Ndumele CD, Roy B, et al. Trends in racial and ethnic disparities in barriers to timely medical care among adults in the US, 1999 to 2018. JAMA Health Forum 2022 Oct 7;3(10):e223856. [doi: 10.1001/jamahealthforum.2022.3856] [Medline: 36306118]
20. Radley DC, Baumgartner JC, Collins SR, Zephyrin LC, Schneider EC. Achieving racial and ethnic equity in U.S. health care: a scorecard of state performance. The Commonwealth Fund. 2021 Nov 18. URL: https://www.commonwealthfund.org/publications/scorecard/2021/nov/achieving-racial-ethnic-equity-us-health-care-state-performance [accessed 2023-11-29]
21. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Population Health and Public Health Practice, Committee on Community-Based Solutions to Promote Health Equity in the United States. The state of health disparities in the United States. In: Baciu A, Negussie Y, Geller A, Weinstein JN, editors. Communities in Action: Pathways to Health Equity: National Academies Press; 2017. URL: https://www.ncbi.nlm.nih.gov/books/NBK425844 [accessed 2023-12-07] [doi: 10.17226/24624] [Medline: 28418632]
22. Deb Nath N, Khan MM, Schmidt M, Njau G, Odoi A. Geographic disparities and temporal changes of COVID-19 incidence risks in North Dakota, United States. BMC Public Health 2023 Apr 20;23(1):720. [doi: 10.1186/s12889-023-15571-5] [Medline: 37081453]
23. Khan MM, Odoi A, Odoi EW. Geographic disparities in COVID-19 testing and outcomes in Florida. BMC Public Health 2023 Jan 11;23(1):79. [doi: 10.1186/s12889-022-14450-9] [Medline: 36631768]
24. Jackson SL, Derakhshan S, Blackwood L, et al. Spatial disparities of COVID-19 cases and fatalities in United States counties. Int J Environ Res Public Health 2021 Aug 4;18(16):8259. [doi: 10.3390/ijerph18168259] [Medline: 34444007]
25. Das P, Igoe M, Lenhart S, et al. Geographic disparities and determinants of COVID-19 incidence risk in the greater St. Louis Area, Missouri (United States). PLoS One 2022 Sep 28;17(9):e0274899. [doi: 10.1371/journal.pone.0274899] [Medline: 36170339]
26. Bauer C, Li X, Zhang K, et al. A novel Bayesian spatial-temporal approach to quantify SARS-CoV-2 testing disparities for small area estimation. Am J Public Health 2023 Jan;113(1):40-48. [doi: 10.2105/AJPH.2022.307127] [Medline: 36516388]
27. Neelon B, Mutiso F, Mueller NT, Pearce JL, Benjamin-Neelon SE. Spatial and temporal trends in social vulnerability and COVID-19 incidence and death rates in the United States. PLoS One 2021 Mar 24;16(3):e0248702. [doi: 10.1371/journal.pone.0248702] [Medline: 33760849]
28. Park J, Michels A, Lyu F, Han SY, Wang S. Daily changes in spatial accessibility to ICU beds and their relationship with the case-fatality ratio of COVID-19 in the state of Texas, USA. Appl Geogr 2023 May;154:102929. [doi: 10.1016/j.apgeog.2023.102929] [Medline: 36960405]
29. U.S. Census Bureau QuickFacts: Cameron County, Texas. United States Census Bureau. 2023. URL: https://www.census.gov/quickfacts/cameroncountytexas [accessed 2023-06-09]
30. Harvard Global Health Institute, Edmond J. Safra Center for Ethics. Key metrics for COVID suppression: a framework for policy makers and the public. Global Epidemics. 2020 Jul 1. URL: https://globalepidemics.org/wp-content/uploads/2020/09/key_metrics_and_indicators_v5-1.pdf [accessed 2025-01-28]
31. 2014-2018 ACS 5-year estimates. United States Census Bureau. URL: https://www.census.gov/programs-surveys/acs/technical-documentation/table-and-geography-changes/2018/5-year.html [accessed 2023-06-09]
32. Kolak M, Bhatt J, Park YH, Padrón NA, Molefe A. Quantification of neighborhood-level social determinants of health in the continental United States. JAMA Netw Open 2020 Jan 3;3(1):e1919928. [doi: 10.1001/jamanetworkopen.2019.19928] [Medline: 31995211]
33. Niu Y, Li Z, Gao Y, et al. A systematic review of the development and validation of the heat vulnerability index: major factors, methods, and spatial units. Curr Clim Change Rep 2021;7(3):87-97. [doi: 10.1007/s40641-021-00173-3] [Medline: 34745843]

XSL•FO
RenderX

34. Pathak S, Wang H, Seals K, Adusumilli NC, Holston D. Self-assessed health status and obesity vulnerability in rural Louisiana: a cross-sectional analysis. PLoS One 2023 Jun 16;18(6):e0287181. [doi: 10.1371/journal.pone.0287181] [Medline: 37327219]
35. Lieberman-Cribbin W, Tuminello S, Flores RM, Taioli E. Disparities in COVID-19 testing and positivity in New York City. Am J Prev Med 2020 Sep;59(3):326-332. [doi: 10.1016/j.amepre.2020.06.005] [Medline: 32703702]
36. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2022. URL: https://www.R-project.org/ [accessed 2025-01-28]
37. Desjardins MR, Hohl A, Delmelle EM. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: detecting and evaluating emerging clusters. Appl Geogr 2020 May;118:102202. [doi: 10.1016/j.apgeog.2020.102202] [Medline: 32287518]
38. Cromley E, Kleinman LC, Ramos MA, et al. A community-engaged approach to select geographic areas for interventions to reduce health disparities. Prog Community Health Partnersh 2011;5(3):299-305. [doi: 10.1353/cpr.2011.0033] [Medline: 22080778]
39. Mardani A, Hooker RE, Ozkul S, et al. Application of decision making and fuzzy sets theory to evaluate the healthcare and medical problems: a review of three decades of research with recent developments. Expert Syst Appl 2019 Dec;137:202-231. [doi: 10.1016/j.eswa.2019.07.002]
40. Boyacı A, Şişman A. Pandemic hospital site selection: a GIS-based MCDM approach employing Pythagorean fuzzy sets. Environ Sci Pollut Res Int 2022 Jan;29(2):1985-1997. [doi: 10.1007/s11356-021-15703-7] [Medline: 34357491]
41. Tripathi AK, Agrawal S, Gupta RD. Comparison of GIS-based AHP and fuzzy AHP methods for hospital site selection: a case study for Prayagraj City, India. GeoJournal 2022;87(5):3507-3528. [doi: 10.1007/s10708-021-10445-y] [Medline: 34075269]
42. Sotoudeh-Anvari A. The applications of MCDM methods in COVID-19 pandemic: a state of the art review. Appl Soft Comput 2022 Sep;126:109238. [doi: 10.1016/j.asoc.2022.109238] [Medline: 35795407]
43. Asadabadi MR, Chang E, Saberi M. Are MCDM methods useful? A critical review of Analytic Hierarchy Process (AHP) and Analytic Network Process (ANP). Cogent Engineering 2019 May 31;6(1). [doi: 10.1080/23311916.2019.1623153]
44. Zhü K. Fuzzy analytic hierarchy process: fallacy of the popular methods. Eur J Oper Res 2014 Jul;236(1):209-217. [doi: 10.1016/j.ejor.2013.10.034]
45. Sekar K, Napili A. Tracking COVID-19: U.S. public health surveillance and data. : Congressional Research Service; 2020 R46588 URL: https://crsreports.congress.gov/product/pdf/R/R46588 [accessed 2025-01-28]
46. Krieger N, Gonsalves G, Bassett MT, Hanage W, Krumholz HM. The fierce urgency of now: closing glaring gaps in US surveillance data on COVID-19. Health Affairs. 2020 Apr 14. URL: https://www.healthaffairs.org/do/10.1377/forefront.20200414.238084/full/ [accessed 2024-02-02]
47. Tuson M, Yap M, Kok MR, et al. Improving the efficiency of geographic target regions for healthcare interventions. Int J Environ Res Public Health 2022 Nov 9;19(22):14721. [doi: 10.3390/ijerph192214721] [Medline: 36429437]
48. Lofters AK, Gozdyra P, Lobb R. Using geographic methods to inform cancer screening interventions for South Asians in Ontario, Canada. BMC Public Health 2013 Apr 26;13:395. [doi: 10.1186/1471-2458-13-395] [Medline: 23622426]
49. Zhan FB, Liu Y, Yang M, et al. Using GIS to identify priority sites for colorectal cancer screening programs in Texas health centers. Prev Chronic Dis 2023 Mar 2;20:E10 [FREE Full text] [doi: 10.5888/pcd20.220205] [Medline: 36862605]

Abbreviations

CBG: census block group
CDC: Centers for Disease Control and Prevention
NETX: Northeast Texas
NIH: National Institutes of Health
PBG: priority census block group
PCA: principal component analysis
RADx-UP: Rapid Access to Diagnostics for Underserved Populations
SDOH: social determinants of health
SVI: Social Vulnerability Index
ZCTA: zip code tabulation area

XSL•FO
RenderX

Edited by A Mavragani; submitted 31.05.24; peer-reviewed by JA Yaun, L Zhu; revised version received 24.10.24; accepted 01.11.24; published 11.02.25.

Please cite as:
Zhang K, Hunyadi JV, de Oliveira Otto MC, Lee M, Zhang Z, Ramphul R, Yamal JM, Yaseen A, Morrison AC, Sharma S, Rahbar MH, Zhang X, Linder S, Marko D, Roy RW, Banerjee D, Guajardo E, Crum M, Reininger B, Fernandez ME, Bauer C
Increasing COVID-19 Testing and Vaccination Uptake in the Take Care Texas Community-Based Randomized Trial: Adaptive Geospatial Analysis
JMIR Form Res 2025;9:e62802
URL: https://formative.jmir.org/2025/1/e62802
doi:10.2196/62802

© Kehe Zhang, Jocelyn V Hunyadi, Marcia C de Oliveira Otto, Miryoung Lee, Zitong Zhang, Ryan Ramphul, Jose-Miguel Yamal, Ashraf Yaseen, Alanna C Morrison, Shreela Sharma, Mohammad Hossein Rahbar, Xu Zhang, Stephen Linder, Dritana Marko, Rachel White Roy, Deborah Banerjee, Esmeralda Guajardo, Michelle Crum, Belinda Reininger, Maria E Fernandez, Cici Bauer. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Investigating Reddit Data on Type 2 Diabetes Management During the COVID-19 Pandemic Using Latent Dirichlet Allocation Topic Modeling and Valence Aware Dictionary for Sentiment Reasoning Analysis: Content Analysis

Meghan Nagpal[1,2], MSc, PEng; Niloofar Jalali[3], PhD; Diana Sherifali[4], PhD, RN, CDE; Plinio Morita[1,2,3], PhD, PEng; Joseph A Cafazzo[1,2,5,6], PhD, PEng

[1]Institute of Health Policy, Management, & Evaluation, Dalla Lana School of Public Health, University of Toronto, Health Sciences Building, 4th Floor, Toronto, ON, Canada

[2]Centre for Digital Therapeutics, Techna Institute, University Health Network, Toronto, ON, Canada

[3]School of Public Health & Health Systems, University of Waterloo, Kitchener, ON, Canada

[4]School of Nursing, McMaster University, Hamilton, ON, Canada

[5]Department of Computer Science, University of Toronto, Toronto, ON, Canada

[6]Institute of Biomedical Engineering, Faculty of Applied Science & Engineering, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Meghan Nagpal, MSc, PEng
Institute of Health Policy, Management, & Evaluation, Dalla Lana School of Public Health, University of Toronto, Health Sciences Building, 4th Floor, Toronto, ON, Canada

Abstract

Background: Type 2 diabetes (T2D) is a chronic disease that can be partially managed through healthy behaviors. However, the COVID-19 pandemic impacted how people managed T2D due to work and school closures and social isolation. Moreover, individuals with T2D were at increased risk of complications from COVID-19 and experienced worsened mental health due to stress and anxiety.

Objective: This study aims to synthesize emerging themes related to the health behaviors of people living with T2D, and how they were affected during the early stages of the COVID-19 pandemic by examining Reddit forums dedicated to people living with T2D.

Methods: Data from Reddit forums related to T2D, from January 2018 to early March 2021, were downloaded using the Pushshift API; support vector machines were used to classify whether a post was made in the context of the pandemic. Latent Dirichlet allocation topic modelling was performed to identify topics of discussion across the entire dataset and a subsequent iteration was performed to identify topics specific to the COVID-19 pandemic. Sentiment analysis using the VADER (Valence Aware Dictionary for Sentiment Reasoning) algorithm was performed to assess attitudes towards the pandemic.

Results: From all posts, the identified topics of discussion were classified into the following themes: managing lifestyle (sentiment score 0.25, 95% CI 0.25-0.26), managing blood glucose (sentiment score 0.19, 95% CI 0.18-0.19), obtaining diabetes care (sentiment score 0.19, 95% CI 0.18-0.20), and coping and receiving support (sentiment score 0.34, 95% CI 0.33-0.35). Among the COVID-19–specific posts, the topics of discussion were coping with poor mental health (sentiment score 0.04, 95% CI −0.01 to0.11), accessing doctor and medications and controlling blood glucose (sentiment score 0.14, 95% CI 0.09-0.20), changing food habits during the pandemic (sentiment score 0.25, 95% CI 0.20-0.31), impact of stress on blood glucose levels (sentiment score 0.03, 95% CI −0.03 to 0.08), changing status of employment and insurance (sentiment score 0.17, 95% CI 0.13-0.22), and risk of COVID-19 complications (sentiment score 0.09, 95% CI 0.03-0.14). Overall, posts classified as COVID-19–related (0.12, 95% CI 0.01-0.15) were associated with a lower sentiment score than those classified as nonCOVID (0.25, 95% CI 0.24-0.25). This study was limited due to the lack of a method for assessing the demographics of users and verifying whether users had T2D.

Conclusions: Themes identified from Reddit data suggested that the COVID-19 pandemic significantly influenced how people with T2D managed their disease, particularly in terms of accessing care and dealing with the complications of the virus. Overall, the early stages of the pandemic negatively impacted the attitudes of people living with T2D. This study demonstrates that social media data can be a qualitative data source for understanding patient perspectives.

(JMIR Form Res 2025;9:e51154) doi:10.2196/51154

XSL•FO
RenderX

KEYWORDS

diabetes; diabetes mellitus; DM; COVID-19; pandemics; social media; health behavior; health knowledge; attitudes; practice; self-management; patient-generated health data; perspective; T2DM

Introduction

Background

Diabetes is a serious metabolic condition in which the body experiences elevated blood glucose levels, which can result in serious complications such as cardiovascular disease, kidney disease, stroke, eye disease, foot ulcers, nerve damage, and amputation. The World Health Organization [1] states that high blood glucose levels are the third leading cause of premature mortality. Type 2 diabetes (T2D) is characterized by insulin resistance or insufficient production of insulin. Research suggests that the risks of complications in people living with T2D can be mitigated through self-management, including glucose levels monitoring, optimized nutrition, regular physical activity, and taking prescribed medications [2,3].

Managing T2D During the COVID-19 Pandemic

With the emergence of the COVID-19 pandemic in March 2020, nearly the entire global population was affected by social distancing measures that included business closures, remote schooling and work, prohibition of large crowds, limited socialization, and virtual outpatient diabetes care delivery. Beyond the fear and anxiety that manifested [4], individuals with T2D were at an increased risk of complications from COVID-19, as the virus thrives in an environment of high blood glucose [5]. T2D was considered among the most prevalent chronic conditions where patients were at serious risk of hospitalization or death if they contracted the virus [6]. This was primarily due to increased inflammation from obesity and insulin resistance, and prevalence of other comorbidities such as hypertension, cardiovascular disease, dyslipidemia, and older age [7]. Because T2D is mostly managed by lifestyle behaviors, COVID-19 restriction measures reduced physical activity and increased other unhealthy behaviors among those living with T2D and other chronic conditions [4,8]. Moreover, there was a concern about limited access to health care for people with T2D, as telemedicine replaced many in-person visits [9]. Overall, people with T2D exhibited worsening glucose levels during lockdown periods [10].

Rationale and Study Objectives

Given that the pandemic impacted the lifestyle behaviors of individuals with T2D, it is reasonable to assume that there were changes in how they managed their condition, which contributed to added stress and anxiety. Many people with T2D use social medial forums to discuss how they manage their condition through sharing information on diet, symptoms, research findings, and recipes, while seeking peer support [11]. Social media also serves as a public data source to gauge sentiment and topics of discussion during the initial lockdown period. The objectives of this study were to synthesize emerging themes from peer discussion on Reddit, a social media forum with dedicated communities for individuals coping with T2D. Using data harnessed through the Pushshift application programming interface (API), this study aimed to understand how their attitudes and diabetes management evolved during the early stages of the pandemic. Reddit was chosen over other social networking platforms due to the segregation of special interest communities, such as those focused on diabetes, and the availability of the Pushshift API as an open-source tool for data scraping.

Methods

Overview

Support vector machines were used to classify Reddit posts related to the COVID-19 pandemic. Latent Dirichlet allocation (LDA) topic modeling [12] and sentiment analysis using the VADER (Valence Aware Dictionary for Sentiment Reasoning) algorithm [13] were performed to achieve the study objectives. Algorithms were executed through Python scripts in the Jupyter Notebook platform version 6 (Linux Foundation), with final extractions outputted as CSV files and further analyzed through Microsoft Excel. No predefined protocol was used for this experiment, given the novelty of social media analysis at the time of the study.

Data Source and Collection

For this study, 3 communities on Reddit were examined: r/type2diabetes, r/diabetes_t2, and r/diabetes [14-16]. From the r/diabetes community, only posts tagged with the “flair” and “type 2 diabetes” were included. Raw data of 100,887 posts between January 1, 2018, to March 5, 2021, from these 3 communities were collected from Reddit in March 2021 via the Pushshift application programming interface (API) using Python scripts in Jupyter Notebook [17]. Data was exported as CSV files and examined in Microsoft Excel. Posts with less than 5 words, posts with the label “removed” and those containing special characters were removed from the dataset using Microsoft Excel filtering tools to avoid bias in the algorithm.

Classification of Posts

Manual examination of the dataset found that the first post about the COVID-19 pandemic was written on February 28, 2020, and coincidentally, the last post about the pandemic in the dataset was made on February 28, 2021. In total, 48,988 posts were published during this period. Within the dataset, posts were manually searched for terms related to the COVID-19 pandemic . These terms included “COVID”, “coronavirus”, “pandemic”, “social distancing”, “lockdown”, “quarantine”, “toilet paper”, “unemploy” and “unemployed”, “work and working from home”, “telehealth”, “vaccine”, “sanitizer”, and “mask.”

Posts that contained these terms in the text body were manually evaluated for context and labeled as COVID or nonCOVID. In total, 9803 posts were manually classified, of which 2065 were labeled as COVID and 7738 as nonCOVID; these posts were subsequently classified using support vector machines. An additional 818 posts made in the context of the COVID-19 pandemic were identified, bringing the total number of

XSL•FO
RenderX

pandemic-specific posts to 2883. The remaining unclassified posts from the identified pandemic period were labeled as nonCOVID.

Data Analysis

The LDA topic modeling algorithm [18] was applied in two iterations to identify topics of discussion by obtaining clustering words belonging to a single topic. A total of 85,266 posts analyzed, with 2682 being specific to the COVID-19 pandemic. The first iteration of topic modeling was performed on the entire processed dataset of 85,266 posts made between January 1, 2018, and March 5, 2021. The second iteration was performed on a subset of this dataset comprising the 2682 posts specifically related to the COVID-19 pandemic.

Sentiment analysis was performed to understand the subjective emotions or sentiment associated with each post, using a normalized compound score between −1 and 1. The compound score thresholds for classifying text into 3 sentiment categories were as per the VADER algorithm [19]: positive (≥0.05), neutral (between −0.05 and 0.05), and negative (≤−0.05). Further statistical analysis was performed with R software version 4 (R Foundation for Statistical Computing), to calculate mean sentiment scores and 95% CI.

Ethical Considerations

According to the University of Toronto's research ethics guidelines, ethical approval was not sought for this study as per exemptions in Section 1, as Reddit was assumed to be a public data source with no reasonable expectation of privacy [20]. Moreover, there was no direct interaction between the researchers and participants; therefore, ethical approval was not necessary, as per Section 2 of the guidelines [20]. The data used for qualitative analysis were scraped directly from Reddit. As users do not need authentication to view Reddit forums, it was assumed that the users who made these posts were aware that they would be displayed publicly. As this is a publicly available dataset, participants were not compensated.

Data for this study was collected through the Pushshift API. Over 100 published research studies have already used Reddit data extracted through this API [21]. Proferes et al [22] conducted a systematic analysis of 727 manuscripts that used Reddit as a data source and found that fewer than 15% mentioned undergoing any form of ethical review.

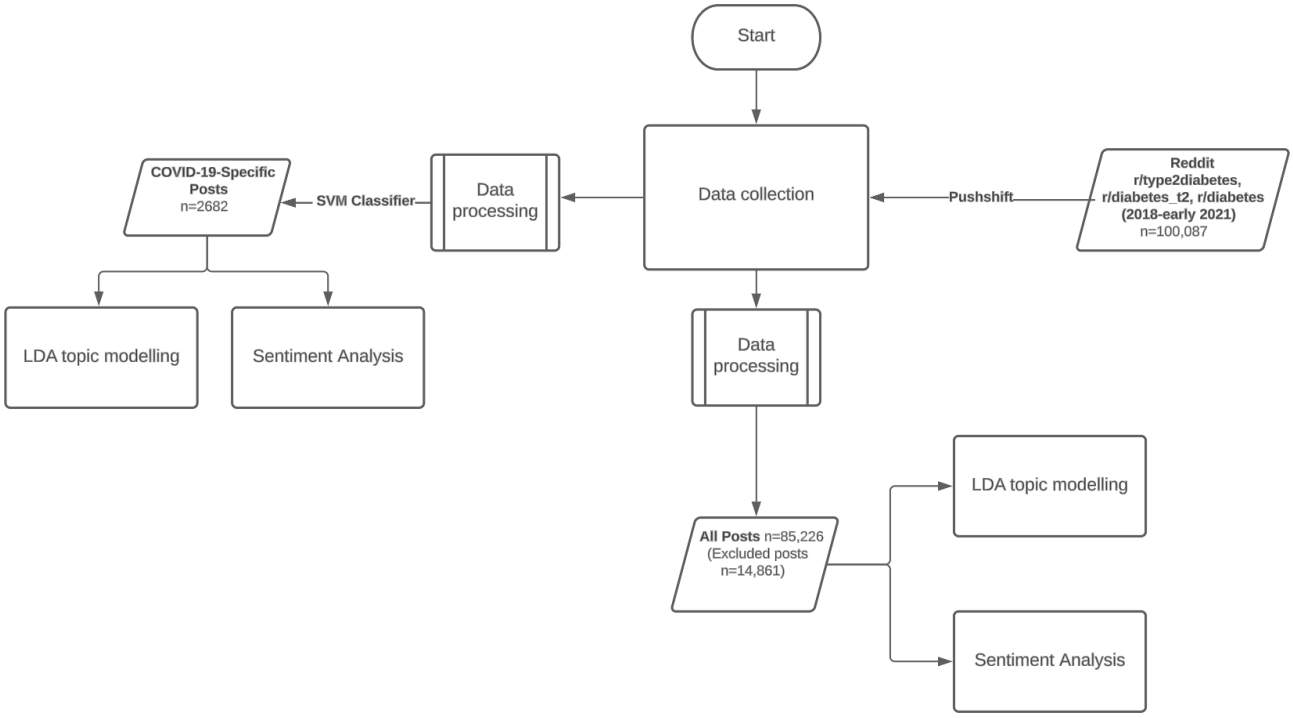
While the usernames of the post authors' were obtained, they were assumed to be pseudonyms and not their actual names. However, we acknowledge that some users may have integrated their real names into their usernames; in this study, the usernames were removed to deidentify the data. Reddit does not provide identifiable characteristics of individual users, such as their names, genders, or geographical locations. However, it is possible that some users may include identifiable information within their posts. In this study, such identifiable information was not extracted or analyzed. Since the posts could not be traced to any individual, no compensation was provided.

Results

Topics of Discussion

The final processed dataset consisted of 85,266 posts made between January 1, 2018, and March 3, 2021. The study methodology is shown in Figure 1. Data were collected from Reddit diabetes communities through the Pushshift API and classified as COVID or nonCOVID. LDA topic modelling and sentiment analysis was conducted on processed text. The topics were manually labeled and categorized into broader themes as shown in Table 1.

Figure 1. Study methodology conducted in early 2021. LDA: latent Dirichlet allocation; SVM: support vector machine.



XSL•FO
RenderX

Table . Words extracted from topic modeling of processed data of all Reddit posts in diabetes communities between January 2018–March 2021.

Themes and topic names	Words
Theme 1: Managing blood glucose	
Using blood glucose meters	meter, insurance, reading, check, strip, libre, sensor, finger, cover, cgm, free, pay, app, stick, test, track, buy, freestyle, monitor, difference
Controlling blood glucose	glucose, blood, high, glucose, level, normal, low, range, control, number, spike, check, average, mgdl, raise, point, fast, affect, measure, test
Controlling HbA_{1C} levels	ac, month, weight, year, lose, start, med, diagnose, metformin, week, diet, stop, lb, change, medication, drop, loss, exercise, pound, control
Rationale for glucose variability	insulin, type, body, resistance, gt, disease, people, medication, risk, cure, increase, reverse, cell, case, pancreas, term, treatment, long, produce, diabetic
Theme 2: Managing lifestyle	
Exercising for diabetes care	work, time, exercise, walk, bit, long, pretty, day, start, minute, run, ill, thing, week, stress, home, stay, big, lot, hit
Timing meals and snacks	fast, hour, meal, bg, metformin, morning, time, number, week, mg, night, eat, effect, day, start, low, dinner, breakfast, sleep, reading
Eating low carbohydrate or ketogenic diet	diet, low, keto, carb, fat, body, protein, calorie, lot, liver, carbs, healthy, process, cholesterol, reduce, energy, high, carbohydrate, intake, muscle
Controlling carbohydrates and alcohol	eat, carbs, food, drink, water, cut, meal, carb, avoid, lot, spike, soda, fiber, drinking, limit, small, alcohol, gram, sweet, ate
Options for food	bread, rice, add, cheese, egg, potato, meat, cream, fruit, veggie, chicken, taste, pasta, recipe, snack, butter, nut, coffee, salad, milk
Theme 3: Coping and receiving support	
Attitudes towards disease	change, life, great, lot, thing, work, dr, make, learn, hope, luck, live, easy, control, love, healthy, manage, hard, lifestyle, sound
Caregiving for family and history	feel, time, family, year, felt, thought, happen, care, make, deal, mom, friend, turn, kind, super, shit, die, guess, hard, suck
Theme 4: Obtaining diabetes care	
Interacting with care team	doctor, test, result, advice, talk, doc, give, diagnose, endo, read, appointment, wait, endocrinologist, follow, diagnosis, time, thought, adjust, mention, idea
Educating for self-management	people, post, health, question, medical, understand, research, information, study, read, answer, agree, link, support, comment, patient, diabetic, base, opinion, group
Managing complications/comorbidities	issue, problem, symptom, pain, foot, effect, bad, eye, year, damage, heart, experience, kidney, stomach, hand, stop, happen, shot, neuropathy, vision

A subset of 2682 posts written between February 28, 2020, and February 28, 2021, was classified as COVID-specific. The topics were manually labeled as shown in Table 2.

XSL•FO
RenderX

Table . Words extracted from topic modeling of processed data from all COVID–related Reddit posts in diabetes communities from 28 February 2020 to 28 February 2021.

Topic Name	Words
Coping with poor mental health	feel, day, week, work, weight, lose, walk, bit, symptom, bad, lot, gym, night, happen, end, great, covid, felt, ill, diagnosis
Accessing doctor and medications and controlling blood glucose	doctor, test, ac, year, insulin, month, metformin, diagnose, low, glucose, diet, week, med, exercise, change, level, stop, reading, check, medication
Changing food habits during the pandemic	eat, food, carbs, lot, meal, low, carb, diet, hard, hour, thing, water, keto, fast, cut, easy, glucose, add, stuff, rice
Impact of stress on blood glucose levels	blood, glucose, high, time, stress, number, long, exercise, bg, body, problem, morning, normal, control, sleep, make, level, effect, change, kind
Changing status of employment and insurance	work, home, hospital, today, time, year, talk, give, advice, wait, insurance, visit, guess, state, order, meter, live, strip, situation, friend
Risk of COVID complications	covid, type, health, care, risk, sick, control, issue, virus, disease, diabetic, question, pandemic, mask, hand, case, infection, wear, patient, home

Sentiment Analysis

Using the VADER algorithm, the compound sentiment score of each post was determined. The sentiment across all posts in the dataset was positive with a mean compound sentiment score of 0.23 (95% CI 0.23-0.24).

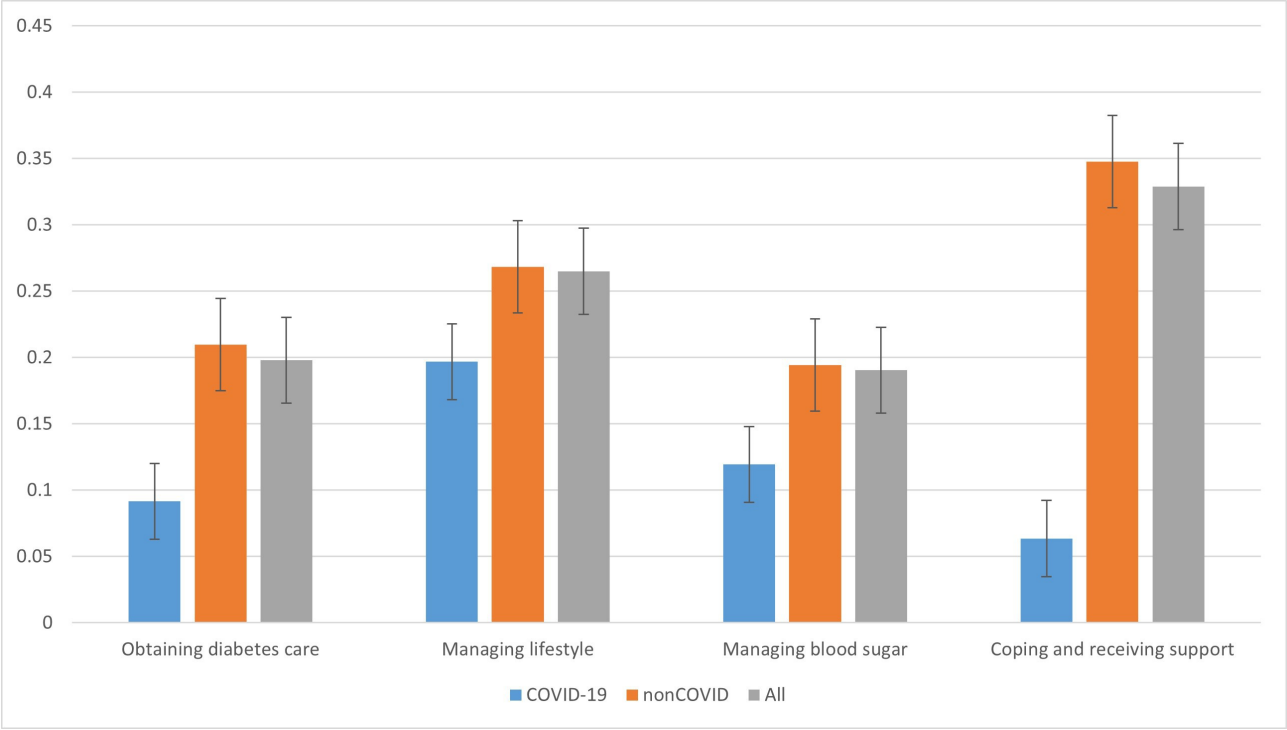
Comparison of Sentiment of COVID Versus NonCOVID Posts

The sentiment was compared between posts that were classified as COVID and nonCOVID. The mean compound scores of COVID-related posts remained within the positive threshold (sentiment score 0.12, 95% CI 0.01-0.15) but was lower than that of nonCOVID-related posts (sentiment score 0.25, 95% CI 0.24-0.25).

Sentiment by Theme

Coping and receiving support had the highest mean compound score (sentiment score 0.34, 95% CI 0.33-0.35), followed by managing lifestyle (sentiment score 0.25, 95% CI 0.25-0.26), obtaining diabetes care (sentiment score 0.19, 95% CI 0.18 - 0.20), and managing blood glucose (0.19, 95% CI 0.18 - 0.19) (Figure 2). All mean compound scores fall in the threshold to be classified as positive, with coping and receiving support demonstrating the strongest positive intensity score.

Figure 2. Mean compound sentiment score of all posts, sorted by theme and classified as covid-related, noncovid-related, or combined.



XSL•FO
RenderX

Sentiment by Theme Amongst COVID-Classified and NonCOVID Posts

Mean compound scores were further analyzed among posts classified as COVID-related or as nonCOVID-related. Coping and receiving support had the highest mean compound score (sentiment score 0.35, 95% CI 0.33 - 0.36), followed by managing lifestyle (sentiment score 0.27, 95% CI 0.26 - 0.28), obtaining diabetes care (sentiment score 0.21, 95% CI 0.20 - 0.22), and managing blood glucose (0.19, 95% CI 0.18 - 0.20). However, for COVID-related posts, coping and receiving support had the lowest mean compound score (sentiment score 0.06, 95% CI −0.02 to 0.13), which was slightly above the threshold for positive classification. Ranked from lowest to highest, obtaining diabetes care had the second-lowest mean compound score (sentiment score 0.09, 95% CI 0.05 - 0.13), followed by managing blood glucose (sentiment score 0.12, 95% CI 0.08 - 0.16), and managing lifestyle (sentiment score 0.20, 95% CI 0.15 - 0.24).

Sentiment by COVID-Specific Topics

Sentiment was further analyzed across the topics generated for the COVID-related posts (Table 3).

Table . Mean compound sentiment score of COVID-related posts sorted by topic.

Topic	Compound sentiment score, mean (95% CI)
Accessing doctor and medications and controlling blood sugar	0.14 (0.09 - 0.20)
Changing status of employment and insurance	0.17 (0.13 - 0.22)
Changing food habits during the pandemic	0.25 (0.20 - 0.31)
Coping with poor mental health	0.04 (−0.01 to 0.11)
Risk of COVID-19 complications	0.09 (0.03 - 0.14)
Impact of stress on blood glucose levels	0.03 (–0.03 to 0.08)

Impact of stress on blood glucose levels and coping with poor mental health had mean compound scores that were within the threshold of being classified as neutral (sentiment score 0.03, 95% CI −0.03 to 0.08 and 0.04, 95% CI −0.01 to 0.11, respectively). These were followed by risk 0f COVID complications (sentiment score 0.09, 95% CI 0.03 - 0.14), accessing doctor and medications and controlling blood glucose (sentiment score 0.14, 95% CI 0.09 - 0.20), and changing status of employment and insurance (sentiment score 0.17, 95% CI 0.13 - 0.22), ranked from lowest to highest. Finally, changing food habits had the highest mean compound score (sentiment score 0.25, 95% CI 0.20 - 0.31).

Discussion

Our study confirmed that T2D subreddits were used for peer support to discuss themes related to diabetes management in depth. Oyebode and Orji [23] and Griffis et al [24] found that the themes for discussion of diabetes management on online platforms focused on nutrition, lifestyle, symptoms, treatments, research, risk factors, and supplements or remedies. Additionally, we found that topics of discussion also included glycemic control and support-seeking behaviors. The overall emerging themes identified in this study include health behaviors such as glycemic control, exercise, and nutrition, along with attitudes towards diabetes, interactions with family and caregivers, management of comorbidities, and educational resources. Topics of discussion were specific to the COVID-19 pandemic and were generally associated with lower sentiment, with specific discussions focusing on glycemic and lifestyle management under unique circumstances, including changing food habits, increased stress, deteriorated mental health, difficulties in accessing health care, and fear of complications due to COVID-19. Additional themes of discussion included access to diabetes care during the pandemic and the impact on employment and insurance during this period.

Topics Discussed on Reddit Communities

Topics discussed on Reddit communities for people managing T2D included glycemic control, exercise, and nutrition, indicating that users were discussing health behaviors. Additionally, users discussed their attitudes toward their disease, interactions with their families and caregivers, management of comorbidities, and shared educational resources. This indicated that online forums provide opportunities for holistic discussions that consider the person as a whole, including the social and psychological factors impacting the management of diabetes. The possibilities for such discussions are endless, which may not be captured in structured clinical settings.

COVID-Specific Discussions on Reddit Forums

Topics extracted from COVID-specific posts were about accessing medical care, the impact of stress on glycemic control, changing food habits, poor mental health, and fear of complications due to COVID. This was in line with a literature review of conventional qualitative studies that assessed the impact of the pandemic on people with T2D [8]. Additionally, we found that the changing status of employment and insurance status was also a topic of discussion among the COVID-specific posts. These data supports our finding that discussions among people with T2D on Reddit offer insights from a holistic perspective, considering various aspects of a person's life related to their disease.

Sentiment Analysis

For all posts in the dataset, the mean compound sentiment score as per the VADER algorithm was 0.23 (95% CI, 0.23 - 0.24). As this score is above the threshold of 0.05, we conclude that, on an average, posts in the Reddit communities for people with diabetes were positive.

A significant difference in sentiment was observed between the COVID-related and nonCOVID posts. Overall, the mean

XSL•FO
RenderX

compound score of posts labeled as "nonCOVID" was 0.24 (95% CI 0.24 - 0.25), while posts labeled as COVID had a mean compound score of 0.12 (95% CI 0.01 - 0.15), which was significantly lower. While we cannot conclude that the COVID-labeled posts were overall negative in tone, their sentiment intensity was lower.

Among the nonCOVID posts, coping and receiving support was associated with the highest mean compound score (sentiment score 0.35, 95% CI 0.33 - 0.36). However, for COVID-related posts, coping and receiving support had the lowest mean compound score (sentiment score 0.06, 95% CI −0.01 to 0.13), slightly above the threshold to be classified as positive. This indicated that the pandemic negatively impacted people with T2D.

While specifically analyzing the topics generated only among the COVID-related posts, changing food habits (sentiment score 0.25, 95% CI 0.20 - 0.31) had the highest mean compound score, indicating that even though eating patterns were impacted during the pandemic, there was peer support and camaraderie in discussions about food choices. Conversely, the impact of stress on blood glucose levels and coping with poor mental health had the lowest mean compound scores (sentiment score 0.03, 95% CI −0.03 to 0.08 and 0.05, 95% CI −0.01 to 0.11, respectively). These low sentiment scores suggest that the pandemic resulted in anxiety that affected glycemic control and resulted in poorer mental health outcomes.

Impact of the Pandemic on Managing T2D

Our study identified key themes for managing T2D during the pandemic from Reddit communities and the associated sentiment toward these themes.

Access to Diabetes Care During the Pandemic

During the COVID-19 pandemic, accessing care was perceived as a significant barrier to managing T2D, as more clinical visits were through web-based platforms. Khader, Jabeen, and Namoju [25] found that in a study of 1582 participants in India, the frequency of clinical visits was reduced in 87.28% of participants, and 87.81% of participants did not have access to health services. Al-Sofiani et al [26] found that among 568 participants in the Arab Gulf region, a lack of communication with the health care providers was associated with higher odds of depression and anxiety among people with diabetes. These findings were further confirmed through our sentiment analysis. Among all posts, the mean compound sentiment score for posts related to the theme of obtaining diabetes care was 0.19 (95% CI 0.18 - 0.20); however, among the COVID-specific posts, this score significantly dropped to 0.09 (95% CI 0.05 - 0.13).

Further, topic modeling among COVID-specific posts also identified access to doctor and controlling blood glucose as a topic for discussion. While barriers to accessing health care providers existed, another barrier including the fear of acquiring a COVID-19 infection and avoiding hospitalization in potentially dangerous situations was identified. Further research should examine the impact of reduced in-person health care visits among people with T2D.

Impact on Employment During the Pandemic

Efforts to curb the spread of the virus resulted in employers requiring employees to work from home [27]. As this was a disruption to daily life, Reddit discussions suggested that this disruption impacted the health behaviors of individuals managing T2D.

In addition to work-from-home measures, the pandemic also resulted in the loss of employment due to the shutting down of businesses to curb the spread of the virus. Unemployment not only affected income and standard of living but also decreased their sense of purpose [28], potentially impacting health behaviors due to increased stress. Furthermore, many unemployed people were also impacted by changes to their health insurance caused by job loss [29,30]. From Reddit discussions, it could be inferred that loss of insurance created barriers to obtaining medications, blood glucose meters, and test strips, ultimately affecting glycemic control for people with T2D.

Glycemic Control During the Pandemic

Managing blood glucose was a major theme for discussion across our dataset. Glycemic control is a major component of self-management of T2D and was expected to be a major theme of discussion among peers. Our analysis found that specifically during the pandemic, users of the Reddit communities for managing T2D reported that the pandemic-related stress affected their blood glucose levels. Laboratory studies have demonstrated that psychological stressors are linked to hyperglycemia [31-33].

Overall, managing blood glucose was the theme with the lowest mean sentiment score (0.19, 95% CI 0.18 - 0.19), which was further reduced among the COVID-related posts (0.12, 95% CI 0.08 - 0.16). This correlation suggested that attitudes toward managing blood glucose are associated with lower sentiment compared to other themes and that the pandemic further increased concerns about glycemic management. Specifically, the impact of stress on blood glucose levels had the lowest sentiment score (0.03, 95% CI −0.03 to 0.08), suggesting that users had increased anxiety about glycemic control.

Additionally, the topic associated with accessing a doctor also included terms related to controlling blood glucose. This suggested that reduced access to a health care provider may have affected glycemic control in people with T2D through limited A1c monitoring or medication access during the pandemic. Changes to employment were associated with reduced access to health insurance, which further resulted in reduced access to medications, blood glucose meters, and test strips, affecting glycemic control.

Lifestyle Management During the Pandemic

In our dataset, lifestyle management was the most frequent theme of discussion, with a positive mean sentiment score of 0.25 (95% CI 0.25 - 0.26). Among the COVID-specific posts, this sentiment score dropped to 0.20 (95% CI 0.15 - 0.24) but was overall positive. Changing food habits during the pandemic was associated with the highest mean sentiment score (0.25, 95% CI 0.20 - 0.31), indicating that increased unhealthy food

XSL•FO
RenderX

consumption may have been a coping mechanism associated with positive emotions through the stressful time.

Interestingly, the terms "gym" and "walk" were included among the COVID-specific posts under the topic coping with poor mental health. This suggests that reduced physical activity was attributed to gym closures in the initial months of the pandemic and people relied on walking as a means of physical activity. Furthermore, physical activity is also another means of coping with stressors and many people were unable to be physically active while staying at home. Further investigation by clinicians is needed to help people with T2D cope during changes such as pandemic-related lockdowns, so that they could continue to maintain a healthy lifestyle during stressful times.

Mental Health During the Pandemic

The mental health impact of this pandemic were expected to be long-term due to the extreme measures necessary to prevent the spread of the virus and the resulting economic implications [34]; people with T2D were no exceptions to this. In our study, we found that while coping and receiving support was associated with the highest mean sentiment overall in our dataset (sentiment score 0.33, 95% CI 0.33 - 0.35), among the COVID-19–related posts, it was associated with the lowest mean sentiment (sentiment score 0.06, 95% CI −0.01 to 0.13). This finding suggests that while Reddit forums for managing T2D have been helpful overall for people with the disease, the pandemic negatively impacted how people coped and their attitudes toward their disease. Specifically, among the COVID-19–related posts, discussions about coping with poor mental health and risk of COVID complications were associated with lower mean sentiment scores (0.05, 95% CI −0.01 to 0.11 and 0.09, 95% CI 0.03-0.14, respectively). Although our study demonstrated that people with T2D used peer support to cope with the stressors of the pandemic, it also highlights the negative impact caused by these psychological stressors. Moreover, the implications of potentially being exposed to the virus posed an additional source of anxiety, given the risk of serious COVID-19 complications among people with T2D. As discussed in this paper, the implications of poor mental health and anxiety regarding COVID-19 exposure affected glycemic control and lifestyle management, both of which are necessary to prevent complications from diabetes. Furthermore, the anxiety related to COVID-19 exposure served as a barrier to seeking care at urgent care centres and hospitals when necessary.

Summary of Findings

This study builds upon the work of Nagpal et al [11], which suggests that public social media forums can serve as a form of patient-generated health data for individuals with chronic conditions such as T2D. Social media forums such as Reddit provide a platform for individuals with chronic diseases to discuss their condition, particularly during times of collective chaos, such as the global COVID-19 pandemic; the insights generated from these discussions provide a valuable source of qualitative knowledge. Consistent with the findings from conventional qualitative studies, this study found that health behaviors of people with T2D were impacted during the pandemic, along with declines in mental health [8], as reflected in the sentiment scores. In particular, the pandemic raised concerns about COVID-19 exposure and access to health care. However, our analysis also indicated that employment and insurance were impacted during the pandemic, affecting how people managed their diabetes care. This finding suggests that social media data can be a valuable source of qualitative data to capture health behaviors during particular periods, such as the global COVID-19 pandemic. As the world continues to evolve due to geopolitics, wars, diseases, and climate change, social media can be a source of qualitative data to gauge attitudes toward chronic disease management.

Limitations of the Study

While topic modeling and sentiment analysis were performed on the dataset, researchers did not manually examine each post for context and cannot comment on the quality of discussions within the forum. Additionally, we cannot confirm that all users who contributed posts were people living with T2D. This study only examined 3 Reddit communities mentioned earlier, and did not include other subreddits related to COVID-19 or mental health.

Furthermore, there is no information about the demographics of users specific to the diabetes forums of Reddit, and we assumed that their demographics were similar to those of general Reddit users. Under this assumption, there may be a sampling bias in our findings, as Reddit users are predominantly male and within the 18 - 49 age demographic [35,36]. By country, the United States has the largest number of Reddit users, with other users mostly residing in high-income, English-speaking countries [37]. However, we cannot determine the specific demographics of the users in the Reddit diabetes communities and have assumed that the stated demographics represent the entire Reddit platform.

Given that COVID-19 implications such as number of cases, lockdown policies, and public attitudes toward the pandemic varied by geographical areas, even within the same country, it was beyond the scope of this study to correlate our findings with these factors, as we did not have geographical information of the users.

Conclusion

Themes of Discussion

Using LDA topic modeling from 85,266 posts made between January 2020 and early March 2021 on Reddit forums for people managing T2D, we found 14 topics of discussion that could be assigned into four themes: managing blood glucose, managing lifestyle, coping and receiving support, and obtaining diabetes care. These findings are in line with recommended health behaviors for T2D management that include glycemic control, healthy nutrition, increasing physical activity, reducing sedentary time, and taking prescribed medications. Additionally, discussions encompassed attitudes toward the disease, interactions with family members and caregivers, managing comorbidities, and the sharing of educational resources. These findings validate that Reddit forums for diabetes management not only serve as a support network for people with T2D, but also facilitate discussions that are holistic in nature and extend beyond treating the symptoms of the disease. Among COVID-19–specific posts, topics of discussion included coping

XSL•FO
RenderX

with poor mental health, accessing doctor and medications and controlling blood glucose, changing food habits during the pandemic, impact of stress on blood glucose levels, changing status of employment and insurance, and risk of COVID-19 complications

Impact on Attitudes Toward T2D Management

The posts classified as COVID-19–related had significantly lower mean sentiment scores than those classified as nonCOVID-19–related, suggesting that the pandemic negatively affected attitudes toward T2D management. Among posts classified as COVID-19–related, topics related to mental health and the impact of stress on glycemic control were associated with lower sentiment scores, whereas posts relating to changing food habits were associated with higher sentiment scores.

Significance of Findings

Our study demonstrates that social media analysis is a simpler and effective method for understanding the behaviors and attitudes of people managing chronic diseases such as diabetes, compared to other qualitative inquiry methods. Topics of discussion and sentiment can potentially be generated in real time and over a long period, directly capturing the perspectives of the patient; this is unlike clinical visits, where data is generated from the perspective of the clinician and only presents a snapshot of the patient's condition. The benefit of social media posts is that they offer infinite possibilities for discussion and provide a more holistic perspective, potentially capturing insights that would have otherwise been overlooked in structured surveys or interviews.

This study has implications on both macro- and microlevels. On a macro scale, extracting data from social media forums can provide an epidemiological overview of behavioral trends of people with T2D, potentially guiding public health policy and formulating treatment guidelines, particularly during public health emergencies such as the COVID-19 pandemic.

Acknowledgments

This study was conducted as part of MSN's master's thesis project and funding was provided by the Centre for Digital Therapeutics. No generative AI was used for this paper.

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: MN, NJ, PM, DS, JAC
Data curation: MN, NJ
Formal analysis: MN, NJ, PM
Funding acquisition: JAC
Investigation: MN
Project administration: JAC
Supervision: JAC
Validation: DS
Writing – original draft: MN
Writing – review & editing: MN,
MN received guidance on the methodology from NJ and PPM. DS serves as the clinical advisor for this study. All listed authors approved the final version of this paper.

Conflicts of Interest

None declared.

References

1. Global Health Risks: Mortality and Burden of Disease Attributable to Selected Major Risks: World Health Organization; 2009. URL: https://www.who.int/publications/i/item/9789241563871 [accessed 2025-02-09]
2. Sherifali D, Viscardi V, Bai JW, Ali RMU. Evaluating the effect of a diabetes health coach in individuals with type 2 diabetes. Can J Diabetes 2016 Feb;40(1):84-94. [doi: 10.1016/j.jcjd.2015.10.006] [Medline: 26827684]
3. Diabetes Canada Clinical Practice Guidelines Expert Committee. Diabetes Canada 2018 Clinical practice guidelines for the prevention and management of diabetes in Canada. Can J Diabetes 2018;42(Suppl1):S1-325.
4. Mental Health and COVID-19. World Health Organization. 2021 Aug 23. URL: https://www.who.int/teams/mental-health-and-substance-use/covid-19 [accessed 2025-02-09]
5. COVID-19 and diabetes. International Diabetes Federation. 2021 Aug 23. URL: https://www.idf.org/aboutdiabetes/what-is-diabetes/covid-19-and-diabetes/1-covid-19-and-diabetes.html [accessed 2025-02-09]

XSL•FO
RenderX

6. Corrao S, Pinelli K, Vacca M, Raspanti M, Argano C. Type 2 diabetes mellitus and COVID-19: a narrative review. Front Endocrinol (Lausanne) 2021;12:609470. [doi: 10.3389/fendo.2021.609470] [Medline: 33868163]
7. Rajpal A, Rahimi L, Ismail-Beigi F. Factors leading to high morbidity and mortality of COVID-19 in patients with type 2 diabetes. J Diabetes 2020 Dec;12(12):895-908. [doi: 10.1111/1753-0407.13085] [Medline: 32671936]
8. Nagpal MS, Jalali N, Sherifali D, Morita PP, Cafazzo JA. Managing type 2 diabetes during the COVID-19 pandemic: scoping review and qualitative study using systematic literature review and reddit. Interact J Med Res 2024 Aug 8;13:e49073. [doi: 10.2196/49073] [Medline: 39116432]
9. Amsah N, Md Isa Z, Ahmad N, Abdul Manaf MR. Impact of COVID-19 pandemic on healthcare utilization among patients with type 2 diabetes mellitus: a systematic review. Int J Environ Res Public Health 2023 Mar 4;20(5):4577. [doi: 10.3390/ijerph20054577] [Medline: 36901588]
10. Eberle C, Stichling S. Impact of COVID-19 lockdown on glycemic control in patients with type 1 and type 2 diabetes mellitus: a systematic review. Diabetol Metab Syndr 2021 Sep 7;13(1):95. [doi: 10.1186/s13098-021-00705-9] [Medline: 34493317]
11. Nagpal MS, Barbaric A, Sherifali D, Morita PP, Cafazzo JA. Patient-Generated data analytics of health behaviors of people living with type 2 diabetes: scoping review. JMIR Diabetes 2021 Dec 20;6(4):e29027. [doi: 10.2196/29027] [Medline: 34783668]
12. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003 Mar 1;3:993-1022. [doi: 10.1162/jmlr.2003.3.4-5.993]
13. vaderSentiment 3.3.2. PyPI. URL: https://pypi.org/project/vaderSentiment [accessed 2021-08-24]
14. R/type2diabetes. Reddit. URL: https://www.reddit.com/r/type2diabetes/ [accessed 2021-08-24]
15. R/diabetes_t2. Reddit. URL: https://www.reddit.com/r/diabetes_t2/ [accessed 2021-08-24]
16. R/diabetes. Reddit. URL: https://www.reddit.com/r/diabetes/ [accessed 2021-08-24]
17. Pushshift reddit API documentation. GitHub. URL: https://github.com/pushshift/api [accessed 2024-08-24]
18. Blei DM. Probabilistic topic models. Commun ACM 2012 Apr 1;55(4):77-84. [doi: 10.1145/2133806.2133826]
19. Bonta V, Kumaresh N, Janardhan N. A comprehensive study on lexicon based approaches for sentiment analysis. AJCST 2019;8(S2):1-6. [doi: 10.51983/ajcst-2019.8.S2.2037]
20. Activities exempt from human ethics review. University of Toronto - Division of the Vice-President, Research & Innovation. URL: https://research.utoronto.ca/ethics-human-research/activities-exempt-human-ethics-review [accessed 2024-05-30]
21. Baumgartner J, Zannettou S, Keegan B, Squire M, Blackburn J. The pushshift reddit dataset. proceedings of the international AAAI conference on web and social media, 14(1), 830-839. 2020 Presented at: Proceedings of the International AAAI Conference on Web and Social Media p. 830-839. [doi: 10.1609/icwsm.v14i1.7347]
22. Proferes N, Jones N, Gilbert S, Fiesler C, Zimmer M. Studying reddit: a systematic overview of disciplines, approaches, methods, and ethics. Social Media + Society 2021 Apr;7(2):205630512110190. [doi: 10.1177/20563051211019004]
23. Oyebode O, Orji R. Detecting factors responsible for diabetes prevalence in Nigeria using social media and machine learning. Presented at: 2019 15th International Conference on Network and Service Management (CNSM); 2019; Halifax, NS, Canada p. 1-4. [doi: 10.23919/CNSM46954.2019.9012679]
24. Griffis H, Asch DA, Andrew Schwartz H, Ungar L, Buttenheim AM, Barg FK, et al. Using social media to track geographic variability in language about diabetes: infodemiology analysis. JMIR Diabetes 2020;5(1):e14431. [doi: 10.2196/14431]
25. Khader MA, Jabeen T, Namoju R. A cross sectional study reveals severe disruption in glycemic control in people with diabetes during and after lockdown in India. Diabetes Metab Syndr 2020;14(6):1579-1584. [doi: 10.1016/j.dsx.2020.08.011] [Medline: 32858476]
26. Al-Sofiani ME, Albunyan S, Alguwaihes AM, Kalyani RR, Golden SH, Alfadda A. Determinants of mental health outcomes among people with and without diabetes during the COVID-19 outbreak in the Arab Gulf Region. J Diabetes 2021 Apr;13(4):339-352. [doi: 10.1111/1753-0407.13149] [Medline: 33351239]
27. Savić D. COVID-19 and work from home: digital transformation of the workforce. Grey Journal (TGJ) 2020 Jun 1;16(2):101-104 [FREE Full text]
28. Unemployment and mental health. Institute for Work & Health. 2009 Aug 24. URL: https://www.iwh.on.ca/sites/iwh/files/iwh/reports/iwh_issue_briefing_mental_health_2009.pdf [accessed 2021-08-24]
29. Gangopadhyaya A, Garrett AB. Unemployment, health insurance, and the COVID-19 recession. SSRN Journal 2024 Apr 1. [doi: 10.2139/ssrn.3568489]
30. Garrett AB, Gangopadhyaya A. How the COVID-19 recession could affect health insurance coverage. SSRN Journal 2020 May 4. [doi: 10.2139/ssrn.3598558]
31. Goetsch VL, VanDorsten B, Pbert LA, Ullrich IH, Yeater RA. Acute effects of laboratory stress on blood glucose in noninsulin-dependent diabetes. Psychosom Med 1993;55(6):492-496. [doi: 10.1097/00006842-199311000-00004] [Medline: 8310109]
32. Faulenbach M, Uthoff H, Schwegler K, Spinas GA, Schmid C, Wiesli P. Effect of psychological stress on glucose control in patients with type 2 diabetes. Diabet Med 2012 Jan;29(1):128-131. [doi: 10.1111/j.1464-5491.2011.03431.x] [Medline: 21883440]

XSL•FO
RenderX

33. Bruce DG, Chisholm DJ, Storlien LH, Kraegen EW, Smythe GA. The effects of sympathetic nervous system activation and psychological stress on glucose metabolism and blood pressure in subjects with type 2 (non-insulin-dependent) diabetes mellitus. Diabetologia 1992 Sep;35(9):835-843. [doi: 10.1007/BF00399929] [Medline: 1397778]
34. Iqbal SZ, Li B, Onigu-Otito E, Naqvi MF, Shah AA. The Long-Term Mental Health Effects of COVID-19. Psychiatr Ann 2020 Dec;50(12):522-525. [doi: 10.3928/00485713-20201103-01]
35. Dean B. Reddit usage and growth statistics: how many people use reddit in 2021. Backlinko. URL: https://backlinko.com/reddit-users [accessed 2021-08-24]
36. Agrawal A. The user demographics of reddit: the official app. Medium. 2021 Aug 24. URL: https://medium.com/@sm_app_intel/the-user-demographics-of-reddit-the-official-app-7e2e18b1e0e1 [accessed 2025-02-09]
37. Reddit: traffic by country. Statista. URL: https://www.statista.com/statistics/325144/reddit-global-active-user-distribution/ [accessed 2021-08-24]

Abbreviations

API: Application Programming Interface
LDA: Latent Dirichlet Allocation
T2D: Type 2 Diabetes
VADER: Valence Aware Dictionary for Sentiment Reasoning

Edited by A Mavragani; submitted 22.07.23; peer-reviewed by C Porter, P Khorasani; revised version received 31.12.24; accepted 02.01.25; published 21.02.25.

Please cite as:
Nagpal M, Jalali N, Sherifali D, Morita P, Cafazzo JA
Investigating Reddit Data on Type 2 Diabetes Management During the COVID-19 Pandemic Using Latent Dirichlet Allocation Topic Modeling and Valence Aware Dictionary for Sentiment Reasoning Analysis: Content Analysis
JMIR Form Res 2025;9:e51154
URL: https://formative.jmir.org/2025/1/e51154
doi:10.2196/51154

© Meghan Nagpal, Niloofar Jalali, Diana Sherifali, Plinio Morita, Joseph A Cafazzo. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

COVID-19 Public Health Communication on X (Formerly Twitter): Cross-Sectional Study of Message Type, Sentiment, and Source

Sana Parveen[1], PhD; Agustin Garcia Pereira[2], MSc; Nathaly Garzon-Orjuela[1], PhD; Patricia McHugh[3], PhD; Aswathi Surendran[4], PhD; Heike Vornhagen[2], PhD; Akke Vellinga[1], PhD

[1]School of Public Health, Physiotherapy and Sports Science, University College Dublin, Belfield, Dublin 4, Ireland, Dublin, Ireland

[2]Data Science Institute, University of Galway, Galway, Ireland

[3]J.E. Cairnes School of Business & Economics, University of Galway, Galway, IrelandIreland

[4]School of Medicine, Ollscoil na Gaillimhe – University of Galway, Galway, Ireland

Corresponding Author:
Sana Parveen, PhD
School of Public Health, Physiotherapy and Sports Science, University College Dublin, Belfield, Dublin 4, Ireland, Dublin, Ireland

Abstract

Background: Social media can be used to quickly disseminate focused public health messages, increasing message reach and interaction with the public. Social media can also be an indicator of people's emotions and concerns. Social media data text mining can be used for disease forecasting and understanding public awareness of health-related concerns. Limited studies explore the impact of type, sentiment and source of tweets on engagement. Thus, it is crucial to research how the general public reacts to various kinds of messages from different sources.

Objective: The objective of this paper was to determine the association between message type, user (source) and sentiment of tweets and public engagement during the COVID-19 pandemic.

Methods: For this study, 867,485 tweets were extracted from January 1, 2020 to March 31, 2022 from Ireland and the United Kingdom. A 4-step analytical process was undertaken, encompassing sentiment analysis, bio-classification (user), message classification and statistical analysis. A combination of manual content analysis with abductive coding and machine learning models were used to categorize sentiment, user category and message type for every tweet. A zero-inflated negative binomial model was applied to explore the most engaging content mix.

Results: Our analysis resulted in 12 user categories, 6 message categories, and 3 sentiment classes. Personal stories and positive messages have the most engagement, even though not for every user group; known persons and influencers have the most engagement with humorous tweets. Health professionals receive more engagement with advocacy, personal stories/statements and humor-based tweets. Health institutes observe higher engagement with advocacy, personal stories/statements, and tweets with a positive sentiment. Personal stories/statements are not the most often tweeted category (22%) but have the highest engagement (27%). Messages centered on shock/disgust/fear-based (32%) have a 21% engagement. The frequency of informative/educational communications is high (33%) and their engagement is 16%. Advocacy message (8%) receive 9% engagement. Humor and opportunistic messages have engagements of 4% and 0.5% and low frequenciesof 5% and 1%, respectively. This study suggests the optimum mix of message type and sentiment that each user category should use to get more engagement.

Conclusions: This study provides comprehensive insight into Twitter (rebranded as X in 2023) users' responses toward various message type and sources. Our study shows that audience engages with personal stories and positive messages the most. Our findings provide valuable guidance for social media-based public health campaigns in developing messages for maximum engagement.

(JMIR Form Res 2025;9:e59687) doi:10.2196/59687

KEYWORDS

public health communication; surveillance; COVID-19; SARS-CoV-2; coronavirus; respiratory; infectious; pulmonary; pandemic; public health messaging; healthcare information; social media; tweets; text mining; data mining; social marketing; infoveillance; intervention planning

XSL•FO
RenderX

Introduction

Public health communication is the scientific research, strategic transmission, and critical evaluation of health information to promote public health [1]. Public health communication initiatives can result in change by increasing awareness, boosting knowledge and forming attitudes when initiatives are well-planned, meticulously carried out, and sustained over time [2].

Social media can be used to quickly disseminate focused public health messages, increasing message reach and interaction with the general public [3,4]. Identifying and understanding information needs, false information, hate speech and discrimination, adherence to precautions, and where concerns lay, aids in the customization of public health strategy and eventually, the development of more informed interventions [5].

Social media can be a valuable resource for learning about people's emotions, concerns and exchanging information. This was shown for instance, when Ebola broke out in Nigeria and public health institutions assisted in containing the Ebola outbreak by tracking social media interactions and disseminating accurate information about the illness [6]. Social media allows public health institutions to track outbreaks in real time and Twitter (rebranded as X in 2023) has been frequently used as a communication tool [7]. The features and status of disease outbreaks can be predicted and explained using information from social media sites and user-generated information has supported the development of early response methods [8].

Social media data text mining can be used for disease forecasting and understanding public awareness of health-related concerns [8]. However, it is still unclear how different social media messages are shared and interpreted or whether different sources (individuals or institutions) communicate efficiently.

During the COVID-19 pandemic [9], social media successfully informed and increased public awareness about this new phenomenon [10]. However, there were considerable differences in the preferred social media platforms, message formats and source sender types [11].

An important concern during the COVID-19 pandemic was the spread of misinformation on social media [12]. Research shows that promoting more messages from reputable, authoritative sources on social media is one of the best ways to prevent misinformation [13].

Examining the content of social media messages provides valuable and timely insights regarding public awareness levels and their needs [14]. While there have been many studies analyzing the tweets on various health issues, there has been limited studies to explore the impact of type, sentiment and source of posts on engagement in public health communication. This study used tweets sent during the pandemic to explore 3 research questions:

- Which sources of information are effective in public health communication?
- What are the most effective types of messages in public health communication?
- Which message type should different sources use to improve engagement?

Methods

Data Collection

Data were extracted from Twitter using a Python script to communicate with Twitter Rest API (application programming interface) using the "Search" endpoint. The "query" parameter was used to filter the results based on the "has:geo" tag, the "place_country:UK"/ "place_country:IE" tag (for Ireland and the United Kingdom), and "lang:en" tag (for English). The "start_time" and "end_time" parameters were used to filter the posts from January 1, 2020 to March 31, 2022.

A basic search was conducted with phrases such as "coronavirus," "SARS-CoV-2," and "COVID-19" on Twitter. Using an iterative method, keywords were added and removed in order to find the most suitable for the data search. The final list of 10 keywords for data extraction included "COVID-19," "pandemic," "SARS-CoV-2," "coronavirus," "SARS-CoV-2 virus," "social distancing," "self-isolation," "self-quarantine," "quarantine," and "new variant." The total number of tweets extracted totaled 867,485.

Data Analysis

Due to the extensive volume of data, a systematic approach was adopted using a 4-step analytical process including sentiment analysis, user-classification, message classification and statistical analysis (Table 1). A combination of manual coding and machine learning (ML) models were used for sentiment, user and message classification. This allowed for multiple rounds of coding to increase the robustness of our results. This methodology is also used by Kummervold et al in their study which shows that by utilizing machine learning models, they could almost exactly match the accuracy of a single human coder when it came to tweet classification. Their research indicates that this automated method, which is dependable and accurate, may be able to guide potentially useful and essential interventions while also freeing up important time and resources for carrying out similar analyses [15].

XSL•FO
RenderX

Table . Phase wise description of the data analysis process.

Phase	Tasks	Methods	Outcome
Data collection	• Identification of keywords related to COVID-19	• Identification of 15 keywords. • After discussion among research team, reduced to 10 keywords for data extraction.	• Final keywords
	• Data extraction	• An academic researcher access was applied for with Twitter. • 867,485 tweets extracted.	
Sentiment analysis	• Allocate sentiment to each tweet	• Random selection of 260 tweets for sentiment allocation (positive, negative, and neutral). • A total of 3 coders manually assigned sentiments to 1 set of 260 random tweets. • Using majority voting, out of 260 tweets, coders agreed on 247 tweets (2 or 3 coders voted the same). • Kappa statistic calculated between all coders and the models. • RoBERTa-base model (highest agreement) applied to the whole dataset.	• Manual coded tweets • Kappa statistic results • Tweets sentiment assigned—full dataset
User classification	• Identify user categories	• User categories identified based on literature on social media supported public health interventions. • Definition of 5 user categories: health institute, health professional, influencer, researcher, and public.	• User categories
	• Coding user profiles	• Abductive coding by 4 Manual coders on 1000 randomly selected tweets to identify new user categories. • After 3 rounds of manual coding, most discrepancies or disagreements resolved through discussion. • Using majority voting, out of 250 tweets, coders agreed on 165 tweets (at least 3 coders voted the same). • A total of 3 ML models selected to assign user categories to same set of 250 tweets used in Round 3 of manual coding. • Bag of words were defined for each user category for user categorization by the ML models. • Accuracy score and F_1-score calculated between ML models and manual coders. • SetFit model (highest agreement) applied to the whole dataset.	• User classification coding—Round 1, 2, and 3 • Bag of words • Accuracy and F_1-score • Tweets users assigned—full dataset

XSL•FO
RenderX

Phase	Tasks	Methods	Outcome
Message classification	• Identify message categories	• Message categories defined based on review of public health communication literature. • Definition of 5 message categories: humor, shock/disgust, informative/educational, opportunistic, and personal stories.	• Message categories
	• Coding tweets	• Abductive coding by 5 manual coders on randomly selected tweets in 3 rounds: 100 tweets in round 1, 100 tweets in round 2, and 250 tweets in round 3. • After 2 rounds of manual coding, discrepancies or disagreements resolved through discussion. • A total of 6 final message categories after addition and reduction of categories over 3 rounds. • Using majority voting, out of 250 tweets, coders agreed on 203 tweets (3 or more coders voted the same). • GPT-3 and Setfit model to assign message categories to set of 250 tweets from the final round using 6 message categories. • Accuracy score and F_1-score calculated between models and manual coder. • SetFit model (highest agreement) applied to the whole dataset.	• Manual coding—Round 1, 2, and 3 • Tweets message assigned—Tweet users assigned
Statistical analysis	• Engagement calculation	• Exclusion of zero-follower users (n=537). • Calculation of engagement (sum of likes, replies, retweets and quoted tweet count divided by the respective user's followers)	
	• Zero-inflation model	• Exclusion of public user group (outside of objective of research) due to 87% zero-engagement. • Application of zero-inflated Poisson and zero-inflated negative binomial model due to remainder of 26% zero-engagement. • Selection of zero-inflated negative binomial model (best fit) with informative/educational message type and neutral sentiment as comparator.	• Final model results

Our cross-sectional observational study aims to provide guidance for social media-based public health campaigns in developing messages for maximum engagement. The study adhered to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist for cross-sectional studies (Checklist 1). Public health interventions are mostly run by governmental agencies or health institutes but can also involve collaborations with researchers, influencers, artists, etc.

XSL•FO
RenderX

Therefore, it is important to study how the public engages with different types of messages from various sources. To ensure a clear distinction and understanding of public engagement, tweets by the general public were removed from our final analysis to focus on engagement received by tweets from all other sources.

Sentiment Analysis

A subset of 260 tweets was randomly extracted for manual annotation by three coders with sentiment labels - positive, negative or neutral (Multimedia Appendix 1). Subsequently, we used 3 ML sentiment analysis models to detect the sentiment for the same tweet subset, which were: distilbert-base-uncased-finetuned-sst-2-english [16], finite automata/bertweet-base-sentiment-analysis [17], and RoBERTa-base for Sentiment Analysis [18].

The kappa statistic was calculated between the 3 coders and between each coder and model. The model with the highest agreement with manual coders based on accuracy score [19] and F_1-score [20] was applied to the full dataset.

User Classification

A directed content analysis with abductive coding was used to explore which sources of information are effective in public health communication. User categorization was based on the profile description and categorized according to an adaptation of the user categories from Cole-Lewis [21]: health institute, health professional, influencer, researcher and public.

User coding was performed in 3 cycles. The profile description of 1000 randomly selected tweets were classified by 4 coders with an overlap of 10% for double coding, that is, coders independently assigned codes to the same set of tweets. The 1000 profile descriptions were categorized into 6 distinct categories with a directed content analysis approach [22] with an additional category “others.” This “others” category was further defined through exploring the profile descriptions, resulting in a total of 12 categories.

The second round aimed to ensure consistency and accuracy between the coders. In total, 100 randomly selected profile descriptions of tweets were allocated to 12 user categories by each coder (Table 2) and agreements, disagreements and potential additions to the categories were discussed. A “bag of words” was compiled for each user category, providing descriptive insights into the characteristics of the users (Table 2 and Multimedia Appendix 2).

XSL•FO
RenderX

Table . User category and bag of words used for classification.

Index	User category	Profile keywords–"bag of words" (Version 6)
1	Health institute	ICGP[a]; WHO[b]; NHS[c]; public health; ECDC[d]; HSE[e]; HPSC[f]; clinic; hospital
2	Health professional	health worker; GP[g]; consultant; nurse; MD[h]; specialist; physician; clinician; surgeon
3	University/Researcher	university; researcher; PhD; student; scientist; academic professor; academia; principal lecturer
4	Influencer	influencer; blogger; vlogger; coach; YouTube
5	Teacher	teacher; teach; school; education; children
6	Politician	politics; government; governor; cabinet; council; minister; councilor; ambassador; MP;[i] Secretary of state; Fine Gael; TD[j]; Mayor
7	Sports	football; rugby; run; swim; tennis; exercise; sports club; pickleball
8	Journalist	journalist; report; news; reporter; columnist; reviewer; media; correspondent; editor
9	Charity	charity; church; ngo; foundation; donations
10	Public	community; union; group; nature; lover; adventure; travel; live; life; world; freedom; farm; pet; cat; dog; walks
11	Known personality	views my own, "True" in verified status
12	Artist	artist; actor; actress; music; writer; singer; photography; movie; sing; play; cinema, orchestra

[a]ICGP: Irish College of General Practitioners.

[b]WHO: World Health Organization.

[c]NHS: National Health Service.

[d]ECDC: European Centre for Disease Prevention and Control.

[e]HSE: Health Service Executive

[f]HPSC: Health Protection Surveillance Centre.

[g]GP: General Practitioner.

[h]MD: Doctor of Medicine.

[i]MP: Member of Parliament.

[j]TD: Teachta Dála.

In the third round, 250 randomly selected profile descriptions were coded by 4 coders and compared with 3 different ML models using the same bag of words: Lbl2TransformerVec model (unsupervised) [23], SGRank model [24], and SetFit model [23].

The ML models classified many influencers as "Public," which was rectified by allocating influencer to any user with more than 3000 followers. The 3 models' performance was evaluated using the profile descriptions of 250 tweets (third round of coding) from 4 coders. Majority voting was used to generate a final coding for the 250 profile descriptions and filtering out profiles that had less than 3 coders agreement. A final dataset of 165 profile descriptions was left after this filtering. The 3 models performance was compared with the manual allocation and the final classification was based on the majority allocation.

Message Classification

Similar to the user classification, a directed content analysis with abductive coding was applied. A review of literature resulted in the identification of 5 message categories from a study by Gough et al [25], which are humor, shock/disgust, educational/informative, personal stories, and opportunistic. The dataset was divided into 2 subsets—public tweets and nonpublic tweets—to determine which type of messages the public engages with most.

A total of 3 manual coding cycles were applied, starting the allocation of 100 random nonpublic tweets to the five message categories and an "other" category. In addition, 2 additional categories emerged, namely, fear and advocacy.

In Round 2, a total of 5 coders allocated 100 tweets to 7 message categories and discussed agreements and disagreements leading to a refinement of the message categories. Fear-based messaging was combined with shock/disgust, personal stories were combined with personal statements and a "not enough information" category was added. In the final round, 250 tweets were categorised into 7 message categories, which are humor, shock/disgust/fear-based, educational/informative, personal

XSL•FO
RenderX

stories/statements, opportunistic, advocacy, and not enough info, and compared to 2 ML models (GPT-3 [OpenAI] and SetFit; Multimedia Appendix 3).

Statistical Analysis

For each tweet, engagement was calculated as the sum of likes, replies, retweets, and quoted tweet count divided by the respective user's follower count [26,27].

For 87% of the tweets, the engagement was zero due to the lack of followers or likes, replies, or quotes (mainly public users). The final dataset excluded this user category resulting in the reduction to 26% zeros. Zero-inflated Poisson and zero-inflated negative binomial models were applied.

Ethical Considerations

Only publicly available data were extracted for this study. All personal identifiable information was deidentified during the data cleaning process.

Results

The total number of tweets extracted was 867,485, which reduced to 802,042 after deleting duplicates. Majority of the tweets (729,619) came from the United Kingdom, 72,350 from Ireland, and 73 were without a location.

The follower count for the dataset showed a wide range, from a maximum of 14,065,098 followers to 0 followers, with an average of 4296 followers. Accounts (users) without followers were investigated to identify inactive accounts or bots and 537 users were removed from the dataset.

Public user category tweets (430,760) were removed from the final dataset as they were not a user category of interest. The final dataset included 370,745 tweets.

Sentiment Analysis

Out of 260 tweets, coders agreed on 247 tweets (this is when 2-3 coders voted the same), which were compared with the accuracy score and F_1-score (weighted) of the 3 ML models. The RoBERTa-base (Model 3) performed the best and was used to assign sentiment to all tweets (Table 3).

Sentiment was negative for 138,379 tweets (37.3%), positive for 84,939 (22.9%), and neutral for 147,427 tweets (39.8%). Positive sentiment tweets had the highest engagement (29.3%), followed by negative (26.7%) and neutral (20.4%; Table 4).

Table . Accuracy and F_1-score for sentiment models with machine learning models.

Model	Accuracy	F_1-score (weighted)
1–Distilbert-base-uncased-finetuned-sst-2-english	0.68	0.61
2–Finite automata/bertweet-base-sentiment-analysis	0.58	0.61
3–RoBERTa-base	0.74	0.76

Table . Frequency and engagement for each sentiment category.

Sentiment	Engagement, %	Frequency, %
Positive	29.3	22.9
Negative	26.7	37.3
Neutral	20.4	39.8

User Classification

Model 3 (assisted) achieved an accuracy score of 0.73 when 4 to 5 coders were in agreement, and achieved an accuracy score of 0.77 when there was at least one match with one coder—192/250 tweets (Table 5).

Health professionals (8%) significantly outnumbered health institutes (1%), while the number of influencers (32%) was more than twice that of journalists (15%) and politicians (13%). Artists (6%) outnumbered individuals associated with sports (2%), and teachers (3%) were one-third in comparison to university/researchers (14%), see Table 6.

Table 7 shows the frequency and percentage of engagement for each user category and provides valuable insights into their impact and effectiveness in engaging audiences. Health professionals have the highest level of engagement (15%), followed by university/researchers (13%) despite a lower frequency (8%). Even though influencers have high frequency (32%) it does not translate into high engagement with influencers having a lower percentage of engagement at 12%.

Journalists and politicians account for 15% and 13%, respectively, with 8% engagement each. Teachers, known personalities. and charities received engagement of 6%, 4%, and 3%, respectively, at lower frequencies. Sports and health Institutes have the lowest engagement each at 1%.

XSL•FO
RenderX

Table . Accuracy and F_1-score for user classification with machine learning models.

Model	Accuracy	F_1-score (weighted)
1-Lbl2TransformerVec	0.26	0.21
2-SGRank	0.43	0.43
3-SetFit	0.69	0.67
3-SetFit (assisted)	0.73	0.73

Table . User categories and their frequency.

User categories	Frequency, n	Frequency, %
Influencer	119,711	32.3
Journalist	54,384	14.7
University/Researcher	50,275	13.6
Politician	47,694	12.9
Health professional	28,963	7.8
Artist	21,730	5.9
Known personality	15,095	4.1
Teacher	12,019	3.2
Charity	11,788	3.2
Sports	5659	1.5
Health institute	3427	0.9

Table . Frequency and engagement for each user category excluding public.

User category	Engagement, %	Frequency, %
Health institute	0.7	0.9
Sports	1.4	1.5
Charity	3.2	3.2
Teacher	5.5	3.2
Known personality	4.1	4.1
Artist	6.1	5.9
Health professional	15.2	7.8
Politician	7.8	12.9
University/Researcher	12.7	13.6
Journalist	8.2	14.7
Influencer	11.5	32.3

Message Classification

The 2 models' performance was evaluated using 250 tweets from 5 coders. Majority voting was used to generate a final coding for the 250 tweets, and filtering out tweets that had less than 3 coders agreement. In addition, 11 tweets were removed because they were identified by the coders as "Not enough information" to code. A final dataset of 203 tweets was left after this filtering. SetFit outperformed GPT-3 and this model achieved 80% accuracy when considering a match with at least 1 coder (192/239 tweets; Table 8).

Personal stories/statements have the highest engagement at 27%, but are not the most frequent tweeted category (22%; Table 9). Shock/disgust/fear-based messages (32%) have 21% engagement. Informative/educational messages have high frequencies (33%) and have 16% engagement. Advocacy messages (8%) have 9% engagement and humor and opportunistic messages have engagements of 4% and 0.5%, and low frequencies 5% and 1%, respectively (Table 10).

XSL•FO
RenderX

Table . Accuracy and F_1-score for user classification with machine learning (ML) models.

	4 and 5 coders agreement		3, 4, or 5 coders agreement	
	Accuracy	F_1-score (weighted)	Accuracy	F_1-score (weighted)
GPT-3	0.60	0.60	0.51	0.49
SetFit	0.74	0.74	0.64	0.62

Table . Frequency and engagement for each message type.

Message type	Engagement, %	Frequency, %
Opportunistic	0.5	0.7
Humor	3.7	4.7
Advocacy	9.1	8.5
Personal stories/statements	26.7	21.8
Shock/disgust/fear-based	21	31.9
Informative/educational	15.5	32.5

Table . Message types frequency and engagement percentage.

Message category	Example tweet	Frequency, n	Engagement, %
Informative/ educational	Coronavirus Daily Update: As at 06 Mar 2022, in the Isle of Man there have been 23328 confirmed cases. #coronavirus #iom #coronaupdate	120,317	15.5
Shock/disgust/ fear-based	@Mysturji @AntacsB @Keir_Starmer General strike? We’re already basically on one. Take to the streets? And die of a pandemic?	118,238	21
Personal stories/ statements	Celebrating my end of Self-isolation period with some improvised gluten free macaroni and, er, fussili and cheese. #norecipe #foodie #satisfied @ Dublin, Ireland	80,881	26.7
Advocacy	‚ÄòEach person who has died in this pandemic is a loved person, a life gone too soon and a family torn apart.' Hold a public inquiry into the Government’s handling of the Covid-19 pandemic #Covid19 - Sign the Petition!	31,340	9.1
Humor	The government says we need to exercise social distancing, stay indoors as much as possible and behave like other people might be carrying a disease like this is all new, but I‚Äôve been doing it since about 2002 #introvert #hermit	17,302	3.7
Opportunistic	In light of #pandemic financial challenges for families, I considered the 10% property tax base increase on Waterford #households trying to recover from the pandemic morally wrong in 2020. Yesterday I voted against #LPT 10% increase & my party again!	2667	0.5

XSL•FO
RenderX

Statistical Analysis

A zero-inflated negative binomial model was applied with informative/educational tweets and neutral sentiment as reference categories. Health professionals received more engagement with advocacy, personal stories/statements and humor. Health Institutes observe higher engagement with advocacy, personal stories/statements and tweets with a positive sentiment (Multimedia Appendix 4).

Journalists and teachers observe higher engagement with advocacy, personal stories/statements and humor while artists have more engagement with shock/disgust/fear-based messages and positive sentiment. Charity organizations, universities, and researchers have more engagement with advocacy and personal stories/statements but not with shock/disgust/fear tweets.

Known personalities have most engagement with advocacy and humor-based, which is opposite for sports entities. Sports entities have more engagement with personal stories/statements and positive sentiment. Politicians and influencers have more engagement with advocacy, personal stories/statements and humor, while they should avoid opportunistic tweets. Table 11 shows the message type and sentiment each user category should tweet and avoid to get maximum engagement.

Table . Findings of zero-inflated model.

User category	Message type to post	Message type to avoid
Health Institute	Advocacy, personal stories/statements, and positive sentiment	Opportunistic, shock/disgust/fear-based, and negative sentiment
Artist	Shock/disgust/fear-based and positive sentiment	Humor and opportunistic
Charity	Advocacy and personal stories/statements	Humor and negative sentiment
Journalist	Advocacy, personal stories/statements and humor	Shock/disgust/fear-based and negative sentiment
Known personality	Advocacy and humor	Shock/ disgust/fear-based
Teacher	Advocacy, personal stories/statements and humor	Shock/disgust/fear-based and negative sentiment
Health professional	Advocacy and personal stories/statements	Opportunistic and shock/disgust/fear-based
Influencer	Advocacy, personal stories/statements and Humor	Opportunistic and negative sentiment
Sports	Personal stories/statements and positive sentiment	Shock/disgust/fear-based
Politician	Advocacy, personal stories/statements, and humor	Opportunistic and negative sentiment
University/Researcher	Advocacy and personal stories/statements	Shock/disgust/fear-based and negative sentiment

Discussion

Principal Findings

This analysis of type, sentiment, and source of COVID-19–related tweets showed what type of tweets have the most engagement for each user group. Overall, personal stories and positive messages have most engagement, even though not for every user group; known persons and influencers have most engagement with humorous tweets. Journalists and teachers were found to have flexibility in their posting strategies, with advocacy, personal stories/statements, and humor being effective for them as well. Artists garnered the most engagement with shock/disgust/fear-based and positive sentiment posts, while charity organizations and universities/researchers had advocacy and personal stories/statements as more engaging message types. Known personalities and sports entities also displayed varying engagement for message types and sentiments, with the former having advocacy and humor-based posts more engaging, while the latter should focus on personal stories/statements and positive sentiment. Politicians and influencers, on the other hand, exhibited higher engagement with advocacy, personal stories/statements, and shock/disgust/fear-based messages, while avoiding opportunistic content.

Positive sentiments in public health messages typically evoke feelings of hope, encouragement, and trust among users, leading to increased sharing behavior [28]. Users are more inclined to engage with and share positive messages that resonate with them emotionally, as they perceive such content as uplifting and supportive [28]. This explains the high engagement received by positive sentiment tweets despite having lower frequency, as shown in Table 4 in this study. Our initial manual categorization ensured a reliable baseline for sentiment analysis in this study. The frequency of negative sentiment tweets and their engagement reflects the widespread anxiety, fear, and uncertainty during the pandemic [29]. The high engagement of tweets with negative sentiments further emphasizes the need to consider emotional content in information dissemination on social media platforms such as Twitter. In this context, future studies should explore the specific impact of tweets by analyzing the responses they generate, including retweets with quotes and replies. This approach would provide additional insights and allow us to evaluate whether tweets achieve their intended impact.

The findings of this study revealed distinct patterns of public engagement across different user categories and message types. Trusted sources are important in shaping public behavior and engagement with health information during crises [30]. Overall,

XSL•FO
RenderX

in the case of users, health professionals received high engagement during the pandemic whereas health institutes received the lowest engagement, maybe reflecting their different use of messages. This highlights the importance of using specific message types for each user category to achieve engagement, as recommended by our study. In addition, health experts, such as general practitioners, communicate health information with greater credibility and persuasiveness than nonexperts or institutes [31].

Internet-based health communities and social media platforms influence public health behaviors and engagement levels [32]. The definition and expansion of sources (user categories) in this study provides a framework to develop specific messaging by user group. Our findings reiterate the importance of understanding audience and tailoring engagement strategies based on category-specific behaviors for optimal engagement [33-35].

Comparison With Previous Work

Most studies have focused on the content of the tweets to understand public reactions or sentiment [36,37], trends during the COVID-19 pandemic [38,39], or vaccinations [40-42].

In one study, Twitter data were explored using machine learning to examine how public opinions and discussions changed throughout the COVID-19 epidemic [36]. Trends in social media discussions during the pandemic were explored using sentiment analysis and topic modeling, which produced useful information about the public discussion around the pandemic [38]. These studies focused mainly on the content of tweets and their engagement but did not explore their association with sources, message type, or sentiments.

The classification of tweets based on types enriched the analysis by capturing the multifaceted nature of communication during the pandemic and provided an understanding of how different message categories influence public engagement and sharing behavior. This study also added to the existing literature by introducing 2 new message categories—advocacy and personal statements—while modifying the existing categories to enable application and provide guidance in framing future public health communications.

Public health literature on message types is very limited. A scoping review on the health risk communication with the public during a pandemic found a lack of studies on the modes of communication [43]. One study discussed the framing of effective COVID-19 messages to connect individuals to authoritative content, emphasising the importance of positive and gain-framed messages [44]. Similarly, personal experiences increased the salience of public health messaging, particularly in promoting sanitation and hygiene practices [45]. Public health messaging during the lockdown in New Zealand showed the importance of consistent messaging principles such as transparency, timeliness, empathy, and clarity [46]. None of these studies used defined message categories and the definition and recommendation of message types for different user categories is a major contribution of our study. This will help content creators, particularly health intervention planners, in choosing the right mix of message and sentiment type to increase their engagement.

A similar study of social media messages explored account type and message structure, taking elements such as hashtags, hyperlinks, mentions, and any images or videos into account but only counting retweets as engagement [8]. They found that tweets with hashtags, videos, and pictures were retweeted more often, while tweets with links had fewer retweets. Furthermore, tweets with sentiment were more frequently retweeted than tweets with neutral sentiments. In our study, the user profile and engagement were explored through engagement metrics such as likes, retweets, reply count, and quote count. We found health institutes to be the least engaged user category, while Xie et al [8] found national health authorities received more engagement when compared with provincial accounts. However, their analysis was limited to the study of the official (national and provincial) public health agencies' Sina Weibo posts only (a China-only microblogging platform).

This study's novel approach lies in the examination of engagement for a 2-year period across different user categories for different message types and sentiment, providing insights into public's response toward different messages and sources. The volume of data and methodology used allows us to provide insights that were not addressed in the existing body of work. This differs from the current studies with similar research objectives, such as a comparative study between Poland and Jordan [47] on social media's role focused on the disparities in platform choices and message efficacy. Another study categorized COVID-19-related tweets into themes to understand public sentiment. The study identified 5 themes for message categories, which were general information, health information, expressions, humor and others, but used a small dataset [48]. Furthermore, a study examined Canadian public health tweets, revealing that tweets promoting action garnered more engagement than purely informational ones. However, retweets were used as the measure for engagement, unlike our study, where we took into consideration a more comprehensive method for calculating engagement [49]. This is another strength of our study, where we compared engagement across message types and user categories instead of solely depending on likes for comparison.

Our findings suggest a correlation between message type, sentiment, source credibility, and engagement. Our study shows that audience engages with personal stories and positive messages the most. Also, with varied users, different types of messages yield engagement. Our study provides guidance for social media–based public health campaigns for developing messages for maximum engagement.

Limitations

We included just 3 factors (sentiment, user type, and message type) for the analysis, which was leading to variance in our analysis. Other factors not recorded or captured may also influence engagement, for instance, time or day of posts, hashtags, images, etc. For our study, we also excluded tweets from the public user category, which was almost 50% of the dataset as it was not required to address this study's research questions.

XSL•FO
RenderX

Conclusion

Our study provides a framework to develop social media messages according to sentiment and message type for different users. Health professionals and institutes and other users can build on the results to improve effective communication through social media channels.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sentiment analysis.
[XLSX File, 44 KB - formative_v9i1e59687_app1.xlsx]

Multimedia Appendix 2
User classification.
[XLSX File, 89 KB - formative_v9i1e59687_app2.xlsx]

Multimedia Appendix 3
Message classification.
[XLSX File, 80 KB - formative_v9i1e59687_app3.xlsx]

Multimedia Appendix 4
Results of zero-inflated model.
[XLSX File, 11 KB - formative_v9i1e59687_app4.xlsx]

Checklist 1
STROBE checklist for cross-sectional studies.
[DOC File, 94 KB - formative_v9i1e59687_app5.doc]

References

1. Bernhardt JM. Communication at the core of effective public health. Am J Public Health 2004 Dec;94(12):2051-2053. [doi: 10.2105/ajph.94.12.2051] [Medline: 15569948]
2. Hornik RC, editor. Public Health Communication: Evidence for Behavior Change: Routledge; 2002. [doi: 10.4324/9781410603029]
3. Plackett R, Kaushal A, Kassianos AP, et al. Use of social media to promote cancer screening and early diagnosis: scoping review. J Med Internet Res 2020 Nov 9;22(11):e21582. [doi: 10.2196/21582] [Medline: 33164907]
4. Khan Y, Tracey S, O'Sullivan T, Gournis E, Johnson I. Retiring the flip phones: exploring social media use for managing public health incidents. Disaster Med Public Health Prep 2019 Dec;13(5-6):859-867. [doi: 10.1017/dmp.2018.147] [Medline: 31218988]
5. Jang H, Rempel E, Roth D, Carenini G, Janjua NZ. Tracking COVID-19 discourse on twitter in North America: infodemiology study using topic modeling and aspect-based sentiment analysis. J Med Internet Res 2021 Feb 10;23(2):e25431. [doi: 10.2196/25431] [Medline: 33497352]
6. Carter M. How twitter may have helped Nigeria contain ebola. BMJ 2014 Nov 19;349:g6946. [doi: 10.1136/bmj.g6946] [Medline: 25410185]
7. Bartlett C, Wurtz R. Twitter and public health. J Public Health Manag Pract 2015;21(4):375-383. [doi: 10.1097/PHH.0000000000000041] [Medline: 24356087]
8. Xie J, Liu L. Identifying features of source and message that influence the retweeting of health information on social media during the COVID-19 pandemic. BMC Public Health 2022 Dec;22(1):805. [doi: 10.1186/s12889-022-13213-w]
9. Ciotti M, Ciccozzi M, Terrinoni A, Jiang WC, Wang CB, Bernardini S. The COVID-19 pandemic. Crit Rev Clin Lab Sci 2020 Sep;57(6):365-388. [doi: 10.1080/10408363.2020.1783198] [Medline: 32645276]
10. Al-Dmour H, Masa'deh R, Salman A, Abuhashesh M, Al-Dmour R. Influence of social media platforms on public health protection against the COVID-19 pandemic via the mediating effects of public health awareness and behavioral changes: integrated model. J Med Internet Res 2020 Aug 19;22(8):e19996. [doi: 10.2196/19996] [Medline: 32750004]

XSL•FO
RenderX

11. Gan CCR, Feng S, Feng H, et al. #WuhanDiary and #WuhanLockdown: gendered posting patterns and behaviours on Weibo during the COVID-19 pandemic. BMJ Glob Health 2022 Apr;7(4):e008149. [doi: 10.1136/bmjgh-2021-008149] [Medline: 35414567]
12. Kouzy R, Abi Jaoude J, Kraitem A, et al. Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on twitter. Cureus 2020 Mar 13;12(3):e7255. [doi: 10.7759/cureus.7255] [Medline: 32292669]
13. Llewellyn S. Covid-19: how to be careful with trust and expertise on social media. BMJ 2020 Mar 25;368:m1160. [doi: 10.1136/bmj.m1160] [Medline: 32213480]
14. Lenoir P, Moulahi B, Azé J, Bringay S, Mercier G, Carbonnel F. Raising awareness about cervical cancer using twitter: content analysis of the 2015 #SmearForSmear campaign. J Med Internet Res 2017 Oct 16;19(10):e344. [doi: 10.2196/jmir.8421] [Medline: 29038096]
15. Kummervold PE, Martin S, Dada S, et al. Categorizing vaccine confidence with a transformer-based machine learning model: analysis of nuances of vaccine sentiment in twitter discourse. JMIR Med Inform 2021 Oct 8;9(10):e29584. [doi: 10.2196/29584] [Medline: 34623312]
16. Sanh V, Debut L, Chaumond J, Wolf T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv. Preprint posted online on Oct 2, 2019. [doi: 10.48550/arXiv.1910.01108]
17. Pérez JM, Rajngewerc M, Giudici JC, et al. Pysentimiento: a python toolkit for opinion mining and social NLP tasks. arXiv. Preprint posted online on Jun 17, 2021. [doi: 10.48550/arXiv.2106.09462]
18. Loureiro D, Barbieri F, Neves L, Espinosa Anke L, Camacho-collados J. TimeLMs: diachronic language models from twitter. Presented at: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics; May 22-27, 2022; Dublin, Ireland URL: https://aclanthology.org/2022.acl-demo [doi: 10.18653/v1/2022.acl-demo.25]
19. Huilgol P. Accuracy vs F1-score. Medium. 2019. URL: https://medium.com/analytics-vidhya/accuracy-vs-f1-score-6258237beca2 [accessed 2025-03-07]
20. Kundu R. F1 score in machine learning: intro & calculation. V7labs. 2022. URL: https://www.v7labs.com/blog/f1-score-guide [accessed 2025-03-07]
21. Cole-Lewis H, Pugatch J, Sanders A, et al. Social listening: a content analysis of e-cigarette discussions on twitter. J Med Internet Res 2015 Oct 27;17(10):e243. [doi: 10.2196/jmir.4969] [Medline: 26508089]
22. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
23. Schopf T, Braun D, Matthes F. Semantic label representations with lbl2vec: a similarity-based approach for unsupervised text classification. In: Web Information Systems and Technologies: Springer, Cham; 2023:59-73. [doi: 10.1007/978-3-031-24197-0_4]
24. Eldallal A, Barbu E. BibRank: automatic keyphrase extraction platform using metadata. Information 2023;14(10):549. [doi: 10.3390/info14100549]
25. Gough A, Hunter RF, Ajao O, et al. Tweet for behavior change: using social media for the dissemination of public health messages. JMIR Public Health Surveill 2017 Mar 23;3(1):e14. [doi: 10.2196/publichealth.6313] [Medline: 28336503]
26. Semiz G, Berger PD. Determining the factors that drive twitter engagement-rates. ABR 2017;5(2) [FREE Full text] [doi: 10.14738/abr.52.2700]
27. Katie Sehl KM. Engagement rate calculator. IndiKit. 2024. URL: https://www.indikit.net/document/371-engagement-rate-calculator [accessed 2025-03-07]
28. Voorveld HAM, van Noort G, Muntinga DG, Bronner F. Engagement with social media and social media advertising: the differentiating role of platform type. J Advert 2018 Jan 2;47(1):38-54. [doi: 10.1080/00913367.2017.1405754]
29. Melton CA, White BM, Davis RL, Bednarczyk RA, Shaban-Nejad A. Fine-tuned sentiment analysis of COVID-19 vaccine-related social media data: comparative study. J Med Internet Res 2022 Oct 17;24(10):e40408. [doi: 10.2196/40408] [Medline: 36174192]
30. Latkin CA, Dayton L, Miller JR, et al. Behavioral and attitudinal correlates of trusted sources of COVID-19 vaccine information in the US. Behav Sci (Basel) 2021 Apr 20;11(4):56. [doi: 10.3390/bs11040056] [Medline: 33924118]
31. Jucks R, Thon FM. Better to have many opinions than one from an expert? Social validation by one trustworthy source versus the masses in online health forums. Comput Human Behav 2017 May;70:375-381. [doi: 10.1016/j.chb.2017.01.019]
32. Chen X. Online health communities influence people's health behaviors in the context of COVID-19. PLOS ONE 2023;18(4):e0282368. [doi: 10.1371/journal.pone.0282368] [Medline: 37053287]
33. Lustria MLA, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
34. Maibach EW. In: Parrott RL, editor. Designing Health Messages: Approaches from Communication Theory and Public Health Practice: SAGE Publication; 1995. [doi: 10.4135/9781452233451]
35. Lutkenhaus RO, Jansz J, Bouman MP. Tailoring in the digital era: Stimulating dialogues on health topics in collaboration with social media influencers. Digit Health 2019;5:2055207618821521. [doi: 10.1177/2055207618821521] [Medline: 30729023]

XSL•FO
RenderX

36. Xue J, Chen J, Hu R, et al. Twitter discussions and emotions about the COVID-19 pandemic: machine learning approach. J Med Internet Res 2020 Nov 25;22(11):e20550. [doi: 10.2196/20550] [Medline: 33119535]
37. Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. PLoS ONE 2020;15(9):e0239441. [doi: 10.1371/journal.pone.0239441] [Medline: 32976519]
38. Boon-Itt S, Skunkan Y. Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study. JMIR Public Health Surveill 2020 Nov 11;6(4):e21978. [doi: 10.2196/21978] [Medline: 33108310]
39. Worrall AP, Kelly C, O'Neill A, et al. Online search trends influencing anticoagulation in patients with COVID-19: observational study. JMIR Form Res 2021 Aug 31;5(8):e21817. [doi: 10.2196/21817] [Medline: 34292865]
40. Liu S, Liu J. Understanding behavioral intentions toward COVID-19 vaccines: theory-based content analysis of tweets. J Med Internet Res 2021 May 12;23(5):e28118. [doi: 10.2196/28118] [Medline: 33939625]
41. Hussain A, Tahir A, Hussain Z, et al. Artificial intelligence-enabled analysis of public attitudes on Facebook and Twitter toward COVID-19 vaccines in the United Kingdom and the United States: observational study. J Med Internet Res 2021 Apr 5;23(4):e26627. [doi: 10.2196/26627] [Medline: 33724919]
42. Lyu JC, Han EL, Luli GK. COVID-19 vaccine-related discussion on twitter: topic modeling and sentiment analysis. J Med Internet Res 2021 Jun 29;23(6):e24435. [doi: 10.2196/24435] [Medline: 34115608]
43. Berg SH, O'Hara JK, Shortt MT, et al. Health authorities' health risk communication with the public during pandemics: a rapid scoping review. BMC Public Health 2021 Jul 15;21(1):1401. [doi: 10.1186/s12889-021-11468-3] [Medline: 34266403]
44. Pattison AB, Reinfelde M, Chang H, et al. Finding the facts in an infodemic: framing effective COVID-19 messages to connect people to authoritative content. BMJ Glob Health 2022 Feb;7(2):e007582. [doi: 10.1136/bmjgh-2021-007582] [Medline: 35131808]
45. Pakhtigian EL, Downs-Tepper H, Anson A, Pattanayak SK. COVID-19, public health messaging, and sanitation and hygiene practices in rural India. J Water Sanit Hyg Dev 2022 Nov 1;12(11):828-837. [doi: 10.2166/washdev.2022.134]
46. Officer TN, McKinlay E, Imlach F, Kennedy J, Churchward M, McBride-Henry K. Experiences of New Zealand public health messaging while in lockdown. Aust N Z J Public Health 2022 Dec;46(6):735-737. [doi: 10.1111/1753-6405.13297] [Medline: 36190208]
47. Abuhashesh MY, Al-Dmour H, Masa'deh R, et al. The role of social media in raising public health awareness during the pandemic COVID-19: an international comparative study. Informatics (MDPI) 2021;8(4):80. [doi: 10.3390/informatics8040080]
48. Karmegam D, Mapillairaju B. What people share about the COVID-19 outbreak on Twitter? An exploratory analysis. BMJ Health Care Inform 2020 Nov;27(3):e100133. [doi: 10.1136/bmjhci-2020-100133] [Medline: 33214193]
49. Slavik CE, Buttle C, Sturrock SL, Darlington JC, Yiannakoulias N. Examining tweet content and engagement of Canadian public health agencies and decision makers during COVID-19: mixed methods analysis. J Med Internet Res 2021 Mar 11;23(3):e24883. [doi: 10.2196/24883] [Medline: 33651705]

Abbreviations

API: application programming interface
ML: machine learning
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by A Mavragani; submitted 19.04.24; peer-reviewed by E Willis, R Chandrasekaran; revised version received 18.12.24; accepted 18.12.24; published 19.03.25.

Please cite as:
Parveen S, Pereira AG, Garzon-Orjuela N, McHugh P, Surendran A, Vornhagen H, Vellinga A
COVID-19 Public Health Communication on X (Formerly Twitter): Cross-Sectional Study of Message Type, Sentiment, and Source
JMIR Form Res 2025;9:e59687
URL: https://formative.jmir.org/2025/1/e59687
doi:10.2196/59687

© Sana Parveen, Agustin Garcia Pereira, Nathaly Garzon-Orjuela, Patricia McHugh, Aswathi Surendran, Heike Vornhagen, Akke Vellinga. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Latino Parents' Reactions to and Engagement With a Facebook Group–Based COVID-19 Vaccine Promotion Intervention: Mixed Methods Pilot Study

Anna I González-Salinas[1], BSc, MBA, MM; Elizabeth L Andrade[1], BA, MPH, DrPH; Lorien C Abroms[1], BSc, MA, DSC; Kaitlyn Gómez[2], BA; Carla Favetto[1]; Valeria M Gómez[1], BS; Karen K Collins[1], PhD

[1]George Washington University, Washington, DC, United States

[2]California State University, Fullerton, CA, United States

Corresponding Author:
Anna I González-Salinas, BSc, MBA, MM
George Washington University
950 New Hampshire Avenue
Washington, DC, 20052
United States
Phone: 1 (202) 994 7400
Email: agonzalez985@gwmail.gwu.edu

Related Article:

Correction of: https://formative.jmir.org/2024/1/e51331

(JMIR Form Res 2025;9:e76107) doi:10.2196/76107

In "Latino Parents' Reactions to and Engagement With a Facebook Group–Based COVID-19 Vaccine Promotion Intervention: Mixed Methods Pilot Study" (JMIR Form Res 2024;8:e51331) the authors made one update.

The Acknowledgements section has been changed from:

> *This research was funded by the National Institutes of Health Community Engagement Alliance Against COVID-19 Disparities (NIH CEAL; grant 1OT2HL156812-01; multiple Principal Investigators: Mark Edberg and ELA), and the George Washington University Institute for Data, Democracy, and Politics (internal grant; multiple Principal Investigators: LCA and ELA).*

To:

> *This research was, in part, funded by the National Institutes of Health (NIH) Agreement OT2HL158287. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the NIH. This research was also funded, in part, by The George Washington University Institute for Data, Democracy, and Politics (multiple Principal Investigators LCA and ELA).*

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 19.04.25; this is a non–peer-reviewed article;accepted 22.04.25; published 28.04.25.

Please cite as:
González-Salinas AI, Andrade EL, Abroms LC, Gómez K, Favetto C, Gómez VM, Collins KK
Correction: Latino Parents' Reactions to and Engagement With a Facebook Group–Based COVID-19 Vaccine Promotion Intervention: Mixed Methods Pilot Study
JMIR Form Res 2025;9:e76107
URL: https://formative.jmir.org/2025/1/e76107
doi:10.2196/76107
PMID:

XSL•FO
RenderX

©Anna I González-Salinas, Elizabeth L Andrade, Lorien C Abroms, Kaitlyn Gómez, Carla Favetto, Valeria M Gómez, Karen K Collins. Originally published in JMIR Formative Research (https://formative.jmir.org), 28.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Correction: Evaluating Bard Gemini Pro and GPT-4 Vision Against Student Performance in Medical Visual Question Answering: Comparative Case Study

Jonas Roos[1], MD; Ron Martin[2], MD; Robert Kaczmarczyk[3], MD

[1]Department of Orthopedics and Trauma Surgery, University Hospital of Bonn, Venusberg-Campus 1, 53127, Bonn, Germany

[2]Department of Plastic and Hand Surgery, Burn Center, BG Clinic Bergmannstrost, Halle (Saale), Germany

[3]Department of Dermatology and Allergy, Technical University of Munich, Munich, Germany

Corresponding Author:
Jonas Roos, MD
Department of Orthopedics and Trauma Surgery, University Hospital of Bonn, Venusberg-Campus 1, 53127, Bonn, Germany

Related Article:

Correction of: https://formative.jmir.org/2024/1/e57592

(JMIR Form Res 2025;9:e71664) doi:10.2196/71664

In "Evaluating Bard Gemini Pro and GPT-4 Vision Against Student Performance in Medical Visual Question Answering: Comparative Case Study" (JMIR Form Res 2024;8:e57592) the authors noted two errors.

Multimedia Appendices 1 and 2 contained identical files. Therefore, the Multimedia Appendix 1 file has been changed to the correct supplementary inference code.

Additionally, the "query" column of Multimedia Appendix 2 has been removed.

The corrections will appear in the online version of the paper on the JMIR Publications website on February 11, 2025, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Multimedia Appendix 1
Inference code.
[ZIP File, 2 KB - formative_v9i1e71664_app1.zip]

Multimedia Appendix 2
Responses from the models.
[ZIP File, 492 KB - formative_v9i1e71664_app2.zip]

Submitted 23.01.25; this is a non–peer-reviewed article;accepted 28.01.25; published 11.02.25.

Please cite as:
Roos J, Martin R, Kaczmarczyk R
Correction: Evaluating Bard Gemini Pro and GPT-4 Vision Against Student Performance in Medical Visual Question Answering: Comparative Case Study
JMIR Form Res 2025;9:e71664
URL: https://formative.jmir.org/2025/1/e71664
doi:10.2196/71664

© Jonas Roos, Ron Martin, Robert Kaczmarczyk. Originally published in JMIR Formative Research (https://formative.jmir.org), 11.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Acceptability and Initial Adoption of the Heart Observation App for Infants With Congenital Heart Disease: Qualitative Study

Elin Hjorth Johansen[1,2], RN, MSCN; Elin Børøsund[3,4], RN, PhD; Ingeborg Martinsen Østen[1], RN, MSc; Henrik Holmstrøm[2,5], MD, PhD; Anne Moen[6], RN, PhD

[1]Neonatal Intensive Care Unit, Division of Children and Adolescent Medicine, Oslo University Hospital, Oslo, Norway

[2]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway

[3]Department of Digital Health Research, Division of Medicine, Oslo University Hospital, Oslo, Norway

[4]Department of Nursing and Health Sciences, Faculty of Health and Social Sciences, University of South-Eastern Norway,, Drammen, Norway

[5]Department of Cardiology, Division of Children and Adolescent Medicine, Oslo University Hospital, Oslo, Norway

[6]Department of Public Health Science, Institute of Health and Society, Faculty of Medicine, University of Oslo, Oslo, Norway

Corresponding Author:
Elin Hjorth Johansen, RN, MSCN
Neonatal Intensive Care Unit
Division of Children and Adolescent Medicine
Oslo University Hospital
E2 4th Floor
Sognsvannsveien 20
Oslo, 0372
Norway
Phone: 47 98676884
Email: ehjorth@ous-hf.no

Related Article:

Correction of: https://formative.jmir.org/2023/1/e45920

(JMIR Form Res 2025;9:e70280) doi:10.2196/70280

In "Acceptability and Initial Adoption of the Heart Observation App for Infants with Congenital Heart Disease: Qualitative Study" (JMIR Form Res 2023; 7:e45920) the authors noted that author Elin Hjorth Johansen's affiliations need to be adjusted.

The following affiliation was added for author Elin Hjorth Johansen:

[2]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway

The correction will appear in the online version of the paper on the JMIR Publications website on January 10, 2025, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 18.12.24; this is a non–peer-reviewed article;accepted 19.12.24; published 10.01.25.

Please cite as:
Hjorth Johansen E, Børøsund E, Martinsen Østen I, Holmstrøm H, Moen A
Correction: Acceptability and Initial Adoption of the Heart Observation App for Infants With Congenital Heart Disease: Qualitative Study
JMIR Form Res 2025;9:e70280
URL: https://formative.jmir.org/2025/1/e70280
doi:10.2196/70280
PMID:

XSL•FO
RenderX

©Elin Hjorth Johansen, Elin Børøsund, Ingeborg Martinsen Østen, Henrik Holmstrøm, Anne Moen. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Assessing and Improving Data Integrity in Web-Based Surveys: Comparison of Fraud Detection Systems in a COVID-19 Study

Stephen Bonett[1], RN, PhD; Willey Lin[1], MB; Patrina Sexton Topper[1], BSN, MS, PhD; James Wolfe[1], MS; Jesse Golinkoff[1], MPH; Aayushi Deshpande[2], MPhil; Antonia Villarruel[1], RN, PhD; José Bauermeister[1], MPH, PhD

[1]School of Nursing, University of Pennsylvania, Philadelphia, PA, United States

[2]Department of Psychology, Ashoka University, Sonepat, India

Corresponding Author:
Stephen Bonett, RN, PhD
School of Nursing
University of Pennsylvania
418 Curie Boulevard
Philadelphia, PA, 19104
United States
Phone: 1 2155734299
Email: stepdo@nursing.upenn.edu

Related Article:

Correction of: https://formative.jmir.org/2024/1/e47091/

(JMIR Form Res 2025;9:e76462) doi:10.2196/76462

In "Assessing and improving data integrity in web-based surveys: Comparison of fraud detection systems in a COVID-19 study" (JMIR Form Res 2024;8:e47091) the authors made one change.

The Acknowledgments section has been updated from:

> *This work was supported by funding from the National Institutes of Health Agreement OT2HL16156 as part of the Community Engagement Alliance Against COVID-19 Disparities, PhillyCEAL.*

to:

> *This research was, in part, funded by the National Institutes of Health (NIH) Agreement OT2HL158287. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the NIH.*

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 23.04.25; this is a non–peer-reviewed article;accepted 25.04.25; published 30.04.25.

Please cite as:
Bonett S, Lin W, Sexton Topper P, Wolfe J, Golinkoff J, Deshpande A, Villarruel A, Bauermeister J
Correction: Assessing and Improving Data Integrity in Web-Based Surveys: Comparison of Fraud Detection Systems in a COVID-19 Study
JMIR Form Res 2025;9:e76462
URL: https://formative.jmir.org/2025/1/e76462
doi:10.2196/76462
PMID:

XSL•FO
RenderX

©Stephen Bonett, Willey Lin, Patrina Sexton Topper, James Wolfe, Jesse Golinkoff, Aayushi Deshpande, Antonia Villarruel, José Bauermeister. Originally published in JMIR Formative Research (https://formative.jmir.org), 30.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Medical Misinformation in AI-Assisted Self-Diagnosis: Development of a Method (EvalPrompt) for Analyzing Large Language Models

Troy Zada[1*], BASc; Natalie Tam[1*], BASc; Francois Barnard[1], BASc; Marlize Van Sittert[2], BS; Venkat Bhat[3,4], MSc, MD; Sirisha Rambhatla[1], PhD

[1]Department of Management Sciences and Engineering, University of Waterloo, 200 University Avenue West, Waterloo, ON, Canada

[2]Faculty of Law, University of Toronto, Toronto, ON, Canada

[3]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[4]Interventional Psychiatry Program, St. Michael's Hospital, Unity Health Toronto, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Sirisha Rambhatla, PhD
Department of Management Sciences and Engineering, University of Waterloo, 200 University Avenue West, Waterloo, ON, Canada

Abstract

Background: Rapid integration of large language models (LLMs) in health care is sparking global discussion about their potential to revolutionize health care quality and accessibility. At a time when improving health care quality and access remains a critical concern for countries worldwide, the ability of these models to pass medical examinations is often cited as a reason to use them for medical training and diagnosis. However, the impact of their inevitable use as a self-diagnostic tool and their role in spreading health care misinformation has not been evaluated.

Objective: This study aims to assess the effectiveness of LLMs, particularly ChatGPT, from the perspective of an individual self-diagnosing to better understand the clarity, correctness, and robustness of the models.

Methods: We propose the comprehensive testing methodology evaluation of LLM prompts (EvalPrompt). This evaluation methodology uses multiple-choice medical licensing examination questions to evaluate LLM responses. Experiment 1 prompts ChatGPT with open-ended questions to mimic real-world self-diagnosis use cases, and experiment 2 performs sentence dropout on the correct responses from experiment 1 to mimic self-diagnosis with missing information. Humans then assess the responses returned by ChatGPT for both experiments to evaluate the clarity, correctness, and robustness of ChatGPT.

Results: In experiment 1, we found that ChatGPT-4.0 was deemed correct for 31% (29/94) of the questions by both nonexperts and experts, with only 34% (32/94) agreement between the 2 groups. Similarly, in experiment 2, which assessed robustness, 61% (92/152) of the responses continued to be categorized as correct by all assessors. As a result, in comparison to a passing threshold of 60%, ChatGPT-4.0 is considered incorrect and unclear, though robust. This indicates that sole reliance on ChatGPT-4.0 for self-diagnosis could increase the risk of individuals being misinformed.

Conclusions: The results highlight the modest capabilities of LLMs, as their responses are often unclear and inaccurate. Any medical advice provided by LLMs should be cautiously approached due to the significant risk of misinformation. However, evidence suggests that LLMs are steadily improving and could potentially play a role in health care systems in the future. To address the issue of medical misinformation, there is a pressing need for the development of a comprehensive self-diagnosis dataset. This dataset could enhance the reliability of LLMs in medical applications by featuring more realistic prompt styles with minimal information across a broader range of medical fields.

(JMIR Form Res 2025;9:e66207) doi:10.2196/66207

KEYWORDS

ChatGPT; health care; LLM; misinformation; self-diagnosis; large language model

XSL•FO
RenderX

Introduction

Background

Large language models (LLMs) have grown in popularity with an ever-expanding list of applications due to their efficiency and accessibility [1,2]. With their increased usage, LLMs are gaining user trust [3], partly due to the anthropomorphic responses produced by models such as GPT-4o, even though they can generate misinformation at scale [4,5]. Recent reports highlight the inability of differentiating truth from misinformation, and the potential collapse of health care systems, as major disruptors on the horizon [6]. This emphasizes the urgent need to develop solutions to ensure the delivery of factual information.

In health care, self-diagnosis through web searches has become a widespread practice and is especially important for underserved communities [7-10], which means that the prospective usage of LLMs in this domain is inevitable. However, relying solely on online searches for health information can result in severe misinformation as content on social media often spreads more rapidly than scientific knowledge [11]. Inaccurate content, conspiracy theories, and false claims are all forms of misinformation which can impact public perceptions, alter behaviors, and reduce trust in health care systems [12]. Moreover, the ongoing global shortage of health care workers [13-16] has driven government entities and health care organizations to explore the use of LLMs as health care assistants and expertise replacements for diagnosis and education [17-23]. Thus, there is a need to study the quality and reliability of LLM-generated responses to health care–related questions.

Prior Work

Recent work has focused on analyzing ChatGPT across various industries, including its application within health care. ChatGPT is a natural language processing model distinctive for its narrative response style to user input [24,25]. Studies have assessed its performance on examinations [20,24,26] and its utility as a self-studying tool [24,27,28], leveraging its ability to provide tailored responses and immediate feedback. Furthermore, ChatGPT has demonstrated potential in assisting research and academic writing by enhancing efficiency and mitigating gaps in researcher knowledge [29]. However, the increased usage of ChatGPT raises significant ethical concerns regarding plagiarism, bias, transparency, inaccuracy, and health equity [30-34].

ChatGPT has also demonstrated superior performance in answering medical questions compared to other LLMs [35]. Research has explored its applications in medical education, including its effectiveness on licensing examinations, tailored learning experiences, and comprehension of complex medical concepts and clinical reasoning [20,24,27,33,36]. Other areas of study have focused on identifying inefficiencies and inaccuracies within clinical workflows, medical research, and diagnoses, with the objective of integrating LLMs to optimize documentation, triage, and clinical data management procedures [33,37-44]. Additionally, investigations into diagnostic assistance have integrated patient questionnaires and medical imaging with LLMs [43,45-48]. Despite the apparent high performance of LLMs in health care, they remain inferior compared to the judgment of human clinicians [49].

In summary, prior work has generally evaluated the trustworthiness of LLMs [50], along with specifically examining their performance in medical situations. These methods have used the multiple-choice questions from the United States Medical Licensing Exam (USMLE) [51] to evaluate the capability of LLMs in attaining scores near the passing threshold of 60% [20,52]. However, these evaluations inaccurately depict the capability of LLMs for self-diagnosis. In practical situations where individuals use technology to self-diagnose, they would not include answers when posing questions and would not provide the same level of information as the examination questions.

Goal of This Study

This study critically examines the performance of LLMs in responding to health care–related questions. To achieve this, we propose evaluation of LLM prompts. This evaluation procedure contains detailed guidelines to assess ChatGPT's response to open-ended questions and validate the robustness of these responses using a sentence dropout method. This 2-staged approach, to our knowledge, is the first comprehensive strategy aimed at better understanding LLM responses and their implications for medical misinformation. We hypothesize that LLMs, particularly ChatGPT-4.0 (referred to as GPT-4.0), are currently unsuitable for self-diagnosis purposes since a significant portion of responses will be ambiguous or incorrect. In particular, this hypothesis is validated if GPT-4.0 surpasses a minimum threshold of 60% [20] for each of the following three questions:

1. Are the responses clear? This question can be answered by analyzing the response consistency.
2. Are the responses genuinely correct? This question can be answered by identifying the responses classified as correct by all assessors.
3. Are the responses robust? This question can be answered by conducting an ablation study on the correct responses.

Methods

Study Design

In this section, the considerations and preparations for the decided dataset are first specified. Then, the assessor procedures and guidelines are discussed along with the process used to analyze the output answers from ChatGPT. Finally, the complete testing methodology is introduced. The overall process uses both nonmedical and medical experts for assessment and is segmented into two experiments: (1) ChatGPT responses on USMLE step 1 open-ended prompts and (2) ChatGPT robustness and ablation study analysis. The GPT-4.0 responses were generated for each of the USMLE questions using Python, with the code and datasets available in the EvalPrompt (evaluation of large language model prompts) repository [53].

XSL•FO
RenderX

Dataset Considerations and Preparations

Overview

The USMLE [51] dataset consists of 3 steps undertaken by medical students throughout their program. Each step is a test consisting of multiple-choice, single-answer, and no-justification questions. Particularly, in this work, the questions are extracted from step 1 since the step 2 and step 3 questions are medically complex for the general population. Furthermore, only textual questions were kept, resulting in a dataset containing 94 single-answer questions that would be used for prompting ChatGPT. Textbox 1 [51] displays a sample of an extracted step 1 question.

Textbox 1. Sample question 3.1 is directly extracted from the United States Medical Licensing Exam step 1 test. This question is a multiple-choice question with a correct answer of (D).

Question: In a sample of 100 individuals, the mean leukocyte count is 7500/mm³, with a standard deviation of 1000/mm³. If the leukocyte counts in this population follow a normal (gaussian) distribution, approximately 50% of individuals will have which of the following total leukocyte counts?

(A) 5500–9500/mm³

(B) <6500/mm³ or >8500/mm³

(C) 6500–8500/mm³

(D) <7500/mm³

(E) >9500/mm³

Answer: (D) <7500/mm³

Baseline ChatGPT Answer Analysis

The first experiment established the foundation for all testing. From the initial multiple-choice questions, each question was transformed into an open-ended question to accurately simulate the circumstances of an individual interacting with ChatGPT. This process was accomplished by removing the multiple-choice options and replacing any instance of "which of the following" with "what." An example transformation is provided in Textbox 2 [51], where the original question from Textbox 1 was transformed into an open-ended question.

Textbox 2. Transformed United States Medical Licensing Exam step 1 test question based on the original Textbox 1 question. The transformation involves removing the (A)-(E) options and replacing the text "which of the following" with "what."

Question: In a sample of 100 individuals, the mean leukocyte count is 7500/mm³, with a standard deviation of 1000/mm³. If the leukocyte counts in this population follow a normal (gaussian) distribution, approximately 50% of individuals will have what total leukocyte counts?

ChatGPT Robustness and Ablation Study Analysis

The second experiment was built upon the foundation established in the first experiment by using only the questions that were categorized as correct (C) by all assessors. Using this subset of data, new questions were iteratively generated by removing a sentence from the question and prompting ChatGPT with the modified question. This process is repeated across all the sentences of a question, excluding the final question sentence indicated by the "what...?" structure. For instance, if a question has 3 sentences, then 3 versions of the same question would be generated, where each question variation has 2 sentences. The first version would remove the first sentence, the second version would remove the second sentence, and the third version would remove the third sentence. The ablation study aimed to simulate how a self-diagnosing individual may overlook information within their health-related searches.

Each of the open-ended modified questions was then provided to ChatGPT using the OpenAI application programming interface to aggregate the LLM responses for the questions. Specifically, the *gpt-4* model was used for GPT-4.0, along with specific parameters to ensure consistency among the responses and that each question would be answered independently. For instance, the model's temperature was set to zero to ensure the responses were focused and deterministic. Details of the complete GPT-4.0 model and corresponding parameters are provided in Multimedia Appendix 1, along with Textbox 3 [51] displaying the GPT-4.0 response to the open-ended Textbox 2 question.

Textbox 3. GPT-4.0 response to United States Medical Licensing Exam step 1 test question based on the transformed Textbox 2 question. The assessor can then compare the GPT-4.0 answer with the original multiple-choice options to evaluate the correctness of the response.

Answer: Approximately 50% of individuals will have total leukocyte counts of 7500/mm³ or less. Explanation: In a normal distribution, the mean value is the point at which approximately 50% of the data falls below and 50% falls above. Therefore, in this case, where the mean leukocyte count is 7500/mm³, approximately 50% of individuals will have leukocyte counts of 7500/mm³ or less.

Assessor Procedures and Guidelines

The 94 open-ended questions served as prompts for ChatGPT, where 3 nonexperts and 3 experts evaluated the generated responses. Nonexperts were used to simulate the self-diagnosis process, while experts were used to assess the response quality.

Each assessor was provided with detailed procedures and guidelines as outlined in Multimedia Appendix 2 to ensure the response categorization process was standardized among the

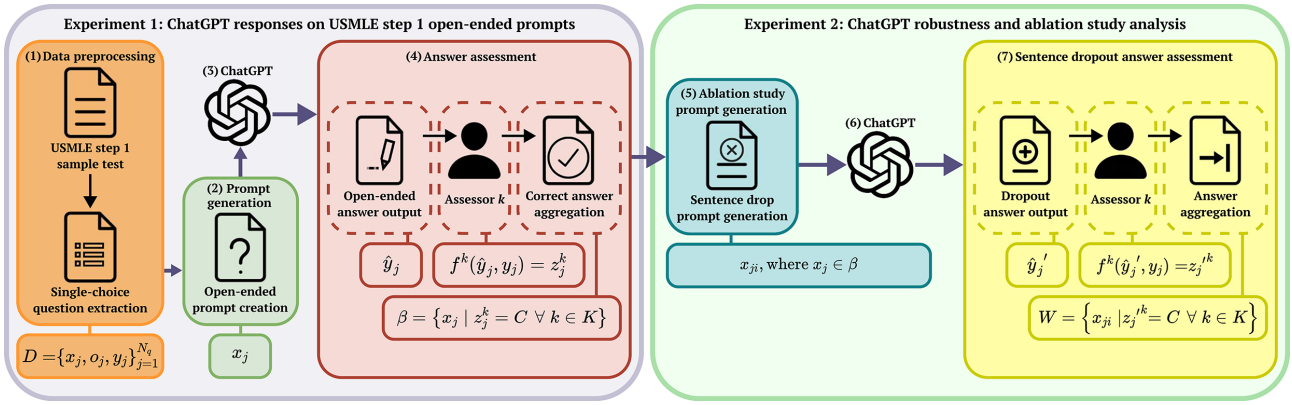
XSL•FO
RenderX

assessors. Assessors were instructed to categorize the responses as either C, partially correct (PC), incorrect (I), or ambiguous (A), and were provided with detailed examples to distinguish between the categories. Moreover, assessors were directed to evaluate the responses independently to ensure no bias.

Testing Methodology

We develop the testing methodology EvalPrompt, as summarized in Figure 1. In total, Nq=94 questions were selected from the USMLE step 1 test. Each question contained a question portion xj, answer options oj, and a correct answer yj, altogether forming the dataset of multiple-choice questions D=xj,oj,yjj=1Nq. Subsequently, each question, xj, was extracted and transformed into an open-ended variation which was then presented to ChatGPT.

Figure 1. EvalPrompt summary. (1) A subset of 94 USMLE step 1 [51] questions consisting of multiple-choice, single-answer questions are selected. (2) These 94 questions are modified to produce open-ended prompts. (3) The open-ended prompts are processed through the ChatGPT API (4) ChatGPT's answers are recorded and presented to *k* independent assessors to categorize as either correct, partially correct, incorrect, or ambiguous. (5) The categorizations classified as correct by all assessors are aggregated to formulate a new dataset for sensitivity analysis. (6) The prompts produced from the iterative sentence dropout are processed through ChatGPT. (7) The independent assessors categorize ChatGPT's answers. These categorizations are then aggregated based on the agreement where all assessors categorized the answer as correct. The resulting dataset, W, is the subset of USMLE question variations that ChatGPT answered correctly. USMLE: United States Medical Licensing Exam.



After collecting and processing the answers from ChatGPT, 3 nonexpert and 3 expert assessors independently evaluated each ChatGPT answer, y^j. In total, K=6 assessors were used. The assessors were instructed to provide a label z∈ Z;Z=C,PC,I,A denoting whether the answer, in comparison to the ground truth, yj, was C, PC, I, or A, respectively. The assessment process can be expressed as a function fky^j,yj=zjk for the kth assessor, where the assessor would use the correct answer to the question along with ChatGPT's answer to categorize the response.

After the k assessors finished the categorization process, the questions categorized as C by all k assessors were aggregated to define a subset dataset, β. Once the resulting dataset, β=xj|zjk=C∀k∈ K was generated, the first experiment was concluded.

The second experiment was built upon the dataset and assessments completed during the first experiment. An ablation study was conducted over the questions within dataset β via an iterative sentence dropout by performing i-1 iterations over an open-ended prompt xj, where i is the number of sentences in the prompt. For the ith iteration, the ith sentence was removed from the prompt before running the prompt through ChatGPT, as expressed by xji, where xj∈ β. Each ith sentence was iteratively removed and processed except for the final sentence which contained the question sentence. The final sentence of the question was mandatory to include to ensure ChatGPT provided an appropriate response.

After processing the sentence dropout questions through ChatGPT, the same 6 assessors evaluated the responses, y^j`. The assessment process can again be expressed as a function fky^j`,yj=zj`k for the kth assessor, where the assessor would use the correct answer to the question along with ChatGPT's answer to categorize the response. This generated the resulting dataset, W, defined as W=xji|zj`k=C∀k∈ K, being the subset of USMLE question variations for the certainly correctly prompts.

Ethical Considerations

Since the aim of this study is to analyze ChatGPT and not human subjects, a research ethics board review was not required. The evaluation of ChatGPT's responses was carried out on a volunteer basis, and all assessors were informed that contributing to the experiments would not result in any safety or privacy risks.

Results

Are the Responses Clear?

To determine if the GPT-4.0 responses are clear, response categorizations across assessors were compared. As displayed in Figure 2, categorization disparities exist across the nonexpert and expert assessors with few alignments across assessors as depicted in Table 1. Namely, the number of C categorizations ranges from 39 to 51, while the number of I categorizations ranges from 12 to 39. This wide range of values for the categories immediately depicts the uncertainty across assessors due to each assessor having varying backgrounds and levels of medical expertise.

Analyzing the categorization for each question independently, many discrepancies exist among the assessors. As detailed in Table 2, on average 52% ([(52+46)/2]/94, SD of 3) of the

XSL•FO
RenderX

responses in experiment 1 and 73% ([(119+104)/2]/152, SD of 7.5) of the responses in experiment 2 were categorized identically by the assessors. For example, USMLE question 54.1 was categorized by the nonexperts as PC, A, and C, and by the experts as I, PC, and PC, respectively. The inconsistency in categorizations depicts that LLM responses do not yet have a single apparent answer, but rather are still subject to interpretation depending on the individual.

Figure 2. Experiment 1: ChatGPT responses on USMLE step 1 open-ended prompts. Individual categorizations for the nonexpert and expert assessors, where each bar represents an individual's categorizations. The categorizations drastically vary across assessors, where some assessors categorized more responses as correct while other assessors categorized more responses as incorrect on the same dataset. USMLE: United States Medical Licensing Exam.

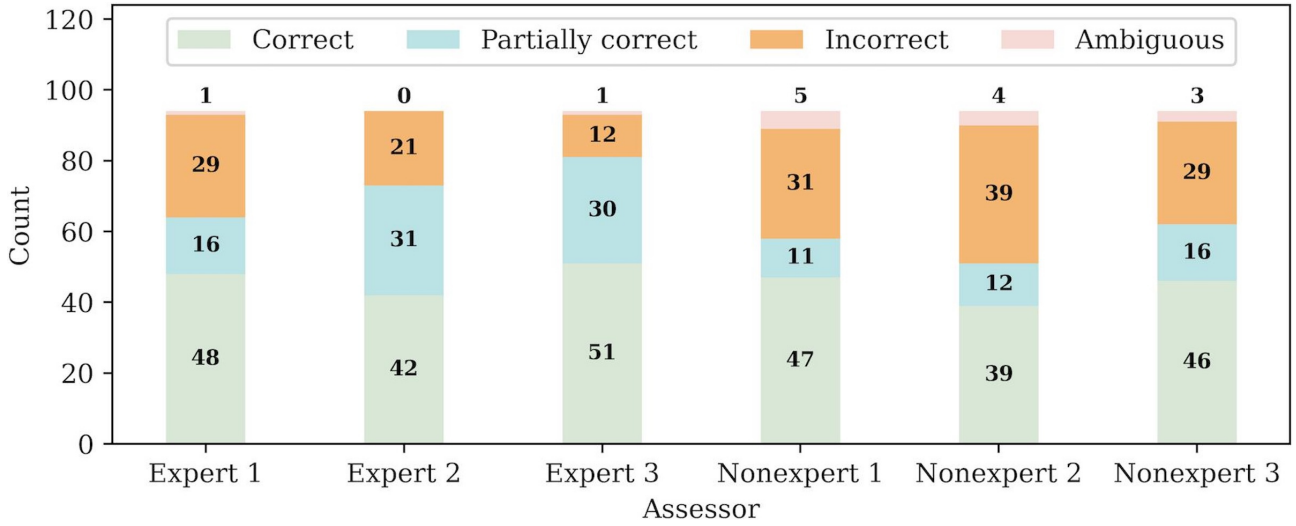


Table . Experiment 1: ChatGPT responses on USMLE[a] step 1 open-ended prompts. Overall response categorization between nonexpert and expert assessors. The top left section indicates the number of instances, regardless of correctness, where all nonexpert and expert assessors categorized a question identically. The bottom right section indicates the total number of questions categorized as correct by all nonexpert and expert assessors.

	Expert			
	Identical	Mismatch	Correct	Other
Nonexpert				
Identical	32	20	__[b]	—
Mismatch	14	28	—	—
Correct	—	—	29	4
Other	—	—	6	55

[a]USMLE: United States Medical Licensing Exam.

[b]Not applicable.

Table . Number of identical categorizations among all individuals within the 2 groups for the 2 experiments. Experiment 1 had 94 questions in total, with 52 and 46 of the questions being categorized the same among the nonexpert and expert assessors, respectively. Similarly, experiment 2 had 152 responses in total spanning 29 unique questions, with 119 and 104 of the questions being categorized the same among the nonexpert and expert assessors, respectively.

	Experiment 1		Experiment 2	
	Nonexpert	Expert	Nonexpert	Expert
Correct	33	35	111	96
Partially correct	1	4	0	1
Incorrect	18	7	8	7
Ambiguous	0	0	0	0
Overall agreement	52	46	119	104

The nonexpert and expert assessors reached an identical conclusion for 52 and 46 of the questions for experiment 1, respectively, as detailed in Table 2. However, the groups collectively could only reach an identical conclusion for 32 of the questions as listed in Table 1. Since only 34% (32/94) of the responses were consistent across all assessors, the 60% threshold could not be met indicating that the responses are unclear. The decrease in the collective number of identically

XSL•FO
RenderX

classified responses suggests that the GPT-4.0 responses are still too ambiguous for assessors to reach appropriate conclusions. The responses are not obvious enough such that anyone, regardless of their background and expertise, can reach the same conclusion.

Are the Responses Genuinely Correct?

To determine the number of GPT-4.0 responses that are genuinely correct, the responses categorized as C by all nonexpert and expert assessors were analyzed and compared. As provided in Table 1, the experiment 1 categorizations where all assessors agreed are listed. In particular, 29 of the 94 responses were categorized as C, meaning that GPT-4.0 is certainly correct 31% (29/94) of the time.

The limited number of C responses indicates that GPT-4.0 is not often factual. As portrayed in Table 2, nonexpert and expert assessors classified 33 and 35 of the responses as C, respectively, even though collectively 29 responses were considered C. Since only 31% (29/94) of the responses were considered C across all assessors, the 60% threshold could not be met indicating that the responses are mostly incorrect. Hence, even though both groups approximately categorized the same number of responses as C, there are still many responses that are either A or I.

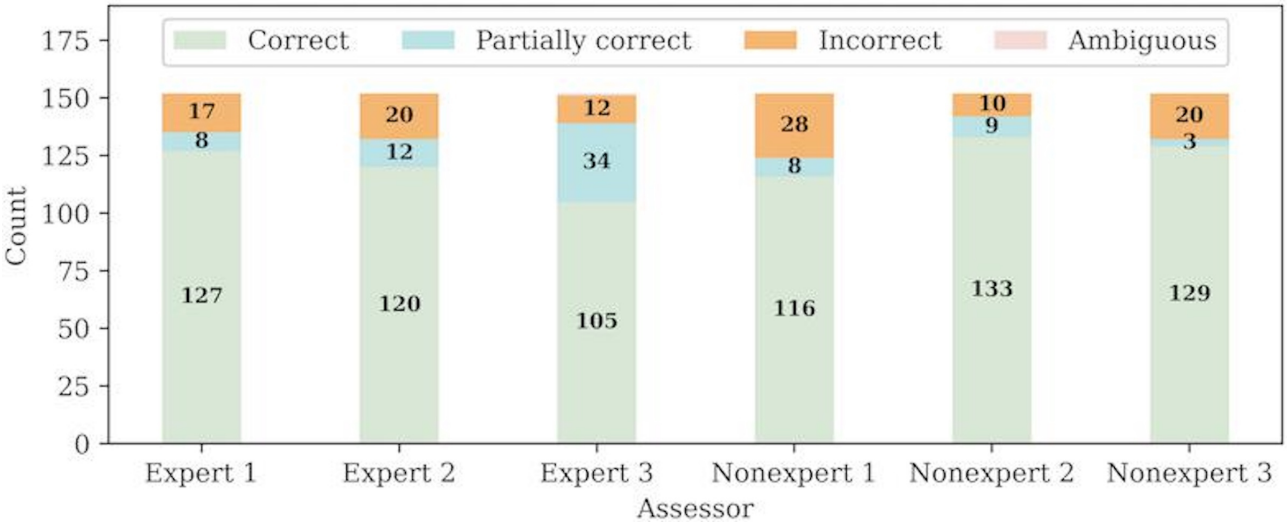
Are the Responses Robust?

To determine if the GPT-4.0 responses are robust, an ablation study was conducted on the 29 correct responses. Reprompting GPT-4.0 with similar variations of the correct questions tested its ability to attain the correct answer even with information missing. This process aimed to simulate the self-diagnosis process since each individual would prompt LLMs with varying levels of information; some individuals would provide extensive details, while others may provide limited information. Thus, assessing GPT-4.0's robustness.

Table 2 provided the categorization details of the ablation study, consisting of 29 unique questions with 152 question variations. On average 68% ([(111+96)/2]/152, SD of 7.5) of the responses continue to be categorized as C even after removing information for each group. These results are far greater than GPT-4.0's accuracy on the initial 94 questions, which was correct only 31% (29/94) of the time. Moreover, Figure 3 depicts the categorizations for each of the 6 assessors. Assessors categorized the experiment 2 dataset C much more frequently than the experiment 1 dataset. On average, 80% ([(127+120+105+116+133+129)/6]/152, SD of 9.34) of the assessors categorized the questions as C, depicting GPT-4.0's robustness on answers that are certainly correct.

Table 3 also displays the assessment similarities between the nonexpert and expert assessors for experiment 2. A total of 92 of the 152 questions were categorized C by all nonexpert and expert assessors, meaning that 61% (92/152) of the responses are certainly correct. Since 61% (92/152) of the responses for the sentence dropout experiment were categorized as C across all assessors, the 60% threshold is met indicating that the responses are robust. In other words, if GPT-4.0 correctly answers a question, it is likely to correctly answer a similar variation of the question again, even if some information is missing.

Figure 3. Experiment 2: ChatGPT robustness and ablation study analysis. Individual categorization for the nonexpert and expert assessors, where each bar represents an individual's categorizations. Out of the 152 questions, the assessors on average categorized 122 of the questions as correct, depicting its robustness.



XSL•FO
RenderX

Table . Experiment 2: ChatGPT robustness and ablation study analysis. Overall response categorization between nonexpert and expert assessors. The top left section indicates the number of instances, regardless of correctness, where all nonexpert and expert assessors categorized a question identically. The bottom right section indicates the total number of questions categorized as correct by all nonexpert and expert assessors.

	Expert			
	Identical	Mismatch	Correct	Other
Nonexpert				
Identical	95	27	—[a]	—
Mismatch	9	21	—	—
Correct	—	—	92	22
Other	—	—	4	34

[a]Not applicable.

Discussion

Principal Results

The hypothesis that ChatGPT is currently unsuitable for self-diagnosis is proved. From each of the 3 addressed assessments with a passing threshold of 60%, GPT-4.0 marginally only passed one. The analysis indicates that GPT-4.0 is generally unclear and incorrect when providing medical information. However, when GPT-4.0 provides correct responses, it remains robust enough to continue answering these questions accurately even when some information is missing. Table 4 summarizes the results of the evaluation procedure for GPT-4.0.

Table . Results summarization of the evaluation procedure for GPT-4.0. With a minimum threshold of 60%, the GPT-4.0 responses are deemed generally unclear and incorrect, while exhibiting robustness when providing correct answers.

	Definition	Success rate (%)	Pass or fail
Are the responses clear?	Agreement reached between all assessors	34	Fail
Are the responses genuinely correct?	Responses categorized as correct by all assessors	31	Fail
Are the responses robust?	Responses continue to be categorized as correct by all assessors	61	Pass

Implications

Recognizing that the GPT-4.0 responses are often ambiguous, its medical advice should be accepted with caution since these responses can vary widely in interpretation depending on the individual. Additionally, since only a small number of responses were found to be genuinely correct, LLMs still require improvement before they can reliably be used in a medical setting. Without these improvements, LLMs risk misinforming individuals.

Even though ChatGPT cannot currently be used for self-diagnosis, substantial evidence indicates that LLMs are continuously improving, suggesting their potential future use in health care systems. As EvalPrompt was also conducted on an earlier GPT-3.5 version, as detailed in Multimedia Appendix 3, the significant difference between the model performances proves that LLMs continue to improve as time progresses.

With ChatGPT being highly acclaimed for its success in passing medical examinations, researchers have proposed using ChatGPT in areas such as medical education and medical report creation [17,20]. However, ChatGPT's ability to answer examination questions does not inherently equate to genuine medical comprehension and proficiency. Instead, using ChatGPT in these medical settings can undermine health care systems since ChatGPT's overconfidence can result in misinforming individuals.

Unlike practicing clinicians, ChatGPT lacks formal testing and accreditation for its abilities. It has not undergone accredited medical education or licensing, lacks approval for clinical practice, and has not demonstrated the necessary understanding or skill set to support its claims. While clinicians face severe consequences for errors, such as medical malpractice charges or license revocation, ChatGPT lacks such liability. Thus, relying on ChatGPT before it becomes accurate, consistent, and robust poses a significant risk of misleading health care practitioners and the general public.

Limitations

This study has signified a substantial advancement in the quality of medical advice that LLMs can provide and their potential utility in the health care industry. Although the USMLE questions provide a solid foundation for simulating the self-diagnosis process via LLMs, the prompts contain extensive detail. In other words, an ordinary individual typically would not input paragraphs of information when self-diagnosing and may lack explicit technical health care knowledge. Therefore, future research can explore using a dataset of self-diagnosis questions that contains less detailed information to accurately assess the capability of LLMs. As a result, developing larger and more realistic self-diagnosing datasets can enhance the

XSL•FO
RenderX

training of LLMs, in turn improving its performance. Additionally, LLMs are inclined to excel in tasks well-represented in the training data, potentially leading to lower performance in niche problems [17]. This raises equity concerns, as questions concerning underrepresented groups may endure poor performance. Although not explicitly examined in this study, future research could generate datasets with diverse question designs across medical specializations to ensure all fields are represented in training.

Conclusions

While LLMs make headlines for passing medical licensing examinations and are consequently being considered as candidates to train the next generation of health care professionals, it is evident that LLMs' capabilities are (understandably) modest at this time. More importantly, this misplaced trust in these systems can lead to reliance on their use for self-diagnosis by the public. In constructing EvalPrompt for assessing the capabilities of LLMs in medical contexts, the extent of misinformation inversely correlates to the LLMs' performance. LLMs that offer unclear and inaccurate responses are more likely to misinform individuals. Ultimately, this investigation presents challenges for machine learning researchers to build more transparent artificial intelligence–powered models capable of reasoning and responding responsibly, while also highlighting the need for a dataset that incorporates more realistic prompt styles.

Acknowledgments

Sirisha Rambhatla, Ph.D., would like to acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (RGPIN-2022-03512).

Data Availability

The step 1 United States Medical Licensing Exam dataset along with the GPT-4.0 answers analyzed during this study are available in the EvalPrompt repository [53]. The datasets for assessor categorizations analyzed during this study are not publicly available due to ChatGPT only being analyzed in aggregate, but are available from the corresponding author on reasonable request.

Conflicts of Interest

VB is supported by an Academic Scholar Award from the University of Toronto Department of Psychiatry and has received research funding from the Canadian Institutes of Health Research, Brain & Behavior Foundation, Ontario Ministry of Health Innovation Funds, Royal College of Physicians and Surgeons of Canada, Department of National Defence (Government of Canada), New Frontiers in Research Fund, Associated Medical Services Inc Healthcare, American Foundation for Suicide Prevention, Roche Canada, Novartis, and Eisai.

Multimedia Appendix 1
The selected ChatGPT-4.0 model along with the corresponding parameters.
[DOCX File, 24 KB - formative_v9i1e66207_app1.docx]

Multimedia Appendix 2
Procedures and guidelines were provided to the assessors to evaluate the LLM responses. LLM: large language model.
[DOCX File, 143 KB - formative_v9i1e66207_app2.docx]

Multimedia Appendix 3
Categorization results for ChatGPT-3.5 in comparison to ChatGPT-4.0.
[DOCX File, 413 KB - formative_v9i1e66207_app3.docx]

References

1. Shahsavar Y, Choudhury A. User intentions to use ChatGPT for self-diagnosis and health-related purposes: cross-sectional survey study. JMIR Hum Factors 2023 May 17;10:e47564. [doi: 10.2196/47564] [Medline: 37195756]
2. Taecharungroj V. "What can ChatGPT do?" analyzing early reactions to the innovative AI Chatbot on Twitter. BDCC 2023 Mar;7(1):35. [doi: 10.3390/bdcc7010035]
3. Salah M, Alhalbusi H, Ismail MM, Abdelfattah F. Chatting with ChatGPT: decoding the mind of Chatbot users and unveiling the intricate connections between user perception, trust and stereotype perception on self-esteem and psychological well-being. Curr Psychol 2024 Mar;43(9):7843-7858. [doi: 10.1007/s12144-023-04989-0]
4. Pan Y, Pan L, Chen W, Nakov P, Kan MY, Wang WY. On the risk of misinformation pollution with large language models. arXiv. Preprint posted online on Oct 26, 2023. [doi: 10.48550/arXiv.2305.13661]
5. Weidinger L, Uesato J, Rauh M, et al. Taxonomy of risks posed by language models. Presented at: FAccT '22; Jun 21-24, 2022; Seoul, Republic of Korea. [doi: 10.1145/3531146.3533088]

XSL•FO
RenderX

6. Disruptions on the horizon. Policy Horizons Canada. 2024. URL: https://horizons.service.canada.ca/en/2024/disruptions/index.shtml [accessed 2025-02-18]
7. Goyder C, McPherson A, Glasziou P. Self diagnosis. BMJ 2009 Nov 11;339:b4418. [doi: 10.1136/bmj.b4418]
8. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: an analysis of health information seeking behavior among US adults. Cogent Social Sciences 2017 Jan 1;3(1):1302785. [doi: 10.1080/23311886.2017.1302785]
9. Swire-Thompson B, Lazer D. Public health and online misinformation: challenges and recommendations. Annu Rev Public Health 2020 Apr 2;41:433-451. [doi: 10.1146/annurev-publhealth-040119-094127] [Medline: 31874069]
10. White RW, Horvitz E. Experiences with web search on medical concerns and self diagnosis. AMIA Annu Symp Proc 2009 Nov 14;2009:696-700. [Medline: 20351943]
11. El Mikati IK, Hoteit R, Harb T, et al. Defining misinformation and related terms in health-related literature: scoping review. J Med Internet Res 2023 Aug 9;25:e45731. [doi: 10.2196/45731] [Medline: 37556184]
12. Okoro YO, Ayo-Farai O, Maduka CP, Okongwu CC, Sodamade OT. A review of health misinformation on digital platforms: challenges and countermeasures. Int J Appl Res Soc Sci 2024 Mar 8;6(1):23-36. [doi: 10.51594/ijarss.v6i1.689]
13. Boniol M, Kunjumen T, Nair TS, Siyam A, Campbell J, Diallo K. The global health workforce stock and distribution in 2020 and 2030: a threat to equity and “universal” health coverage? BMJ Glob Health 2022 Jun;7(6):e009316. [doi: 10.1136/bmjgh-2022-009316] [Medline: 35760437]
14. Kuehn BM. Clinician shortage exacerbates pandemic-fueled “mental health crisis”. JAMA 2022 Jun 14;327(22):2179. [doi: 10.1001/jama.2022.8661]
15. Michel JP, Ecarnot F. The shortage of skilled workers in Europe: its impact on geriatric medicine. Eur Geriatr Med 2020 Jun;11(3):345-347. [doi: 10.1007/s41999-020-00323-0] [Medline: 32328964]
16. Turale S, Nantsupawat A. Clinician mental health, nursing shortages and the COVID-19 pandemic: crises within crises. Int Nurs Rev 2021 Mar;68(1):12-14. [doi: 10.1111/inr.12674] [Medline: 33891772]
17. Garg RK, Urs VL, Agarwal AA, Chaudhary SK, Paliwal V, Kar SK. Exploring the role of ChatGPT in patient care (diagnosis and treatment) and medical research: a systematic review. Health Promot Perspect 2023;13(3):183-191. [doi: 10.34172/hpp.2023.22] [Medline: 37808939]
18. Horesh A. Using ChatGPT to study medicine: learn the basics. FutureDoctorAI. 2023 Mar. URL: https://futuredoctor.ai/chatgpt [accessed 2025-02-18]
19. Iftikhar L, Iftikhar MF, Hanif MI. DocGPT: impact of ChatGPT-3 on health services as a virtual doctor. EC Paediatrics 2023 Mar;12(3):45-55 [FREE Full text]
20. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023 Feb;2(2):e0000198. [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
21. Lee H. The rise of ChatGPT: Exploring its potential in medical education. Anat Sci Educ 2024;17(5):926-931. [doi: 10.1002/ase.2270] [Medline: 36916887]
22. Primack D. Here come the robot doctors. Axios. 2023 Jan. URL: https://www.axios.com/2023/01/18/chatgpt-ai-health-care-doctors [accessed 2025-02-18]
23. Sedaghat S. Early applications of ChatGPT in medical practice, education and research. Clin Med (Lond) 2023 May;23(3):278-279. [doi: 10.7861/clinmed.2023-0078] [Medline: 37085182]
24. Gilson A, Safranek C, Huang T, et al. How does ChatGPT perform on the medical licensing exams? The implications of large language models for medical education and knowledge assessment. medRxiv. Preprint posted online on Dec 26, 2022. [doi: 10.1101/2022.12.23.22283901]
25. Scott K. Microsoft teams up with openai to exclusively license GPT-3 language model. The Official Microsoft Blog. 2020 Sep. URL: https://blogs.microsoft.com/blog/2020/09/22/microsoft-teams-up-with-openai-to-exclusively-license-gpt-3-language-model [accessed 2025-02-18]
26. Choi JH, Hickman KE, Monahan A, Schwarcz DB. ChatGPT goes to law school. SSRN Journal 2022;71(3):387-400. [doi: 10.2139/ssrn.4335905]
27. Sallam M, Salim NA, Barakat M, Al-Tammemi AB. ChatGPT applications in medical, dental, pharmacy, and public health education: a descriptive study highlighting the advantages and limitations. Narra J 2023 Apr;3(1):e103. [doi: 10.52225/narra.v3i1.103] [Medline: 38450035]
28. van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting CL. ChatGPT: five priorities for research. Nature New Biol 2023 Feb;614(7947):224-226. [doi: 10.1038/d41586-023-00288-7] [Medline: 36737653]
29. Zhai X. ChatGPT user experience: implications for education. SSRN Journal 2022. [doi: 10.2139/ssrn.4312418]
30. Arora A, Arora A. The promise of large language models in health care. The Lancet 2023 Feb;401(10377):641. [doi: 10.1016/S0140-6736(23)00216-7]
31. Biswas SS. Role of Chat GPT in public health. Ann Biomed Eng 2023 May;51(5):868-869. [doi: 10.1007/s10439-023-03172-7] [Medline: 36920578]
32. Liebrenz M, Schleifer R, Buadze A, Bhugra D, Smith A. Generating scholarly content with ChatGPT: ethical challenges for medical publishing. Lancet Digit Health 2023 Mar;5(3):e105-e106. [doi: 10.1016/S2589-7500(23)00019-5] [Medline: 36754725]

XSL•FO
RenderX

33. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023 Mar 19;11(6):887. [doi: 10.3390/healthcare11060887] [Medline: 36981544]
34. Ufuk F. The role and limitations of large language models such as ChatGPT in clinical settings and medical journalism. Radiology 2023 May;307(3):e230276. [doi: 10.1148/radiol.230276] [Medline: 36880943]
35. Li J, Dada A, Puladi B, Kleesiek J, Egger J. ChatGPT in healthcare: a taxonomy and systematic review. Comput Methods Programs Biomed 2024 Mar;245:108013. [doi: 10.1016/j.cmpb.2024.108013] [Medline: 38262126]
36. Holmes J, Liu Z, Zhang L, et al. Evaluating large language models on a highly-specialized topic, radiation oncology physics. Front Oncol 2023;13:1219326. [doi: 10.3389/fonc.2023.1219326] [Medline: 37529688]
37. Agrawal M, Hegselmann S, Lang H, Kim Y, Sontag D. Large language models are few-shot clinical information extractors. Presented at: The 2022 Conference on Empirical Methods in Natural Language Processing; Dec 7-11, 2022; Abu Dhabi, United Arab Emirates. [doi: 10.18653/v1/2022.emnlp-main.130]
38. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. J Med Syst 2023 Mar 4;47(1):33. [doi: 10.1007/s10916-023-01925-4] [Medline: 36869927]
39. Harrer S. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. EBioMedicine 2023 Apr;90:104512. [doi: 10.1016/j.ebiom.2023.104512] [Medline: 36924620]
40. Jayakumar P, Moore MG, Furlough KA, et al. Comparison of an artificial intelligence-enabled patient decision aid vs educational material on decision quality, shared decision-making, patient experience, and functional outcomes in adults with knee osteoarthritis: a randomized clinical trial. JAMA Netw Open 2021 Feb 1;4(2):e2037107. [doi: 10.1001/jamanetworkopen.2020.37107] [Medline: 33599773]
41. Patel SB, Lam K. ChatGPT: the future of discharge summaries? Lancet Digit Health 2023 Mar;5(3):e107-e108. [doi: 10.1016/S2589-7500(23)00021-3] [Medline: 36754724]
42. Tan XJ, Cheor WL, Lim LL, Ab Rahman KS, Bakrin IH. Artificial intelligence (AI) in breast imaging: a scientometric umbrella review. Diagnostics (Basel) 2022 Dec 9;12(12):3111. [doi: 10.3390/diagnostics12123111] [Medline: 36553119]
43. Xue VW, Lei P, Cho WC. The potential impact of ChatGPT in clinical and translational medicine. Clin Transl Med 2023 Mar;13(3):e1216. [doi: 10.1002/ctm2.1216] [Medline: 36856370]
44. Yang X, Chen A, PourNejatian N, et al. A large language model for electronic health records. NPJ Digit Med 2022 Dec 26;5(1):194. [doi: 10.1038/s41746-022-00742-2] [Medline: 36572766]
45. Horvat N, Veeraraghavan H, Nahas CSR, et al. Combined artificial intelligence and radiologist model for predicting rectal cancer treatment response from magnetic resonance imaging: an external validation study. Abdom Radiol (NY) 2022 Aug;47(8):2770-2782. [doi: 10.1007/s00261-022-03572-8] [Medline: 35710951]
46. Pun FW, Leung GHD, Leung HW, et al. Hallmarks of aging-based dual-purpose disease and age-associated targets predicted using PandaOmics AI-powered discovery engine. Aging (Albany NY) 2022 Mar 29;14(6):2475-2506. [doi: 10.18632/aging.203960] [Medline: 35347083]
47. Rao A, Kim J, Kamineni M, Pang M, Lie W, Succi MD. Evaluating ChatGPT as an adjunct for radiologic decision-making. medRxiv. Preprint posted online on Feb 7, 2023. [doi: 10.1101/2023.02.02.23285399] [Medline: 36798292]
48. Wang S, Zhao Z, Ouyang X, Liu T, Wang Q, Shen D. Interactive computer-aided diagnosis on medical image using large language models. Commun Eng 2024 Sep 17;3(1):133. [doi: 10.1038/s44172-024-00271-8] [Medline: 39284899]
49. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. Nature New Biol 2023 Aug;620(7972):172-180. [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
50. Huang Y, Sun L, Wang H, Wu S, Zhang Q, Li Y, et al. TrustLLM: trustworthiness in large language models. arXiv. Preprint posted online on Jan 10, 2024. [doi: 10.48550/arXiv.2401.05561]
51. Step 1 sample test questions. USMLE. 2022 Jun. URL: https://www.usmle.org/exam-resources/step-1-materials/step-1-sample-test-questions [accessed 2025-02-18]
52. Ziaei R, Schmidgall S. Language models are susceptible to incorrect patient self-diagnosis in medical applications. arXiv. Preprint posted online on Sep 17, 2023. [doi: 10.48550/arXiv.2309.09362]
53. Zada T, Tam N, Rambhatla S. EvalPrompt: analyzing large language models for self-diagnosis. GitHub. 2022. URL: https://github.com/criticalml-uw/EvalPrompt [accessed 2025-02-18]

Abbreviations

A: ambiguous
C: correct
EvalPrompt: evaluation of large language model prompts
I: incorrect
LLM: large language model
PC: partially correct
USMLE: United States Medical Licensing Exam

XSL•FO
RenderX

Edited by A Mavragani, D Chartash; submitted 06.09.24; peer-reviewed by D Lungeanu, GK Gupta, Y Huang; revised version received 29.01.25; accepted 29.01.25; published 10.03.25.

Please cite as:
Zada T, Tam N, Barnard F, Van Sittert M, Bhat V, Rambhatla S
Medical Misinformation in AI-Assisted Self-Diagnosis: Development of a Method (EvalPrompt) for Analyzing Large Language Models
JMIR Form Res 2025;9:e66207
URL: https://formative.jmir.org/2025/1/e66207
doi:10.2196/66207

© Troy Zada, Natalie Tam, Francois Barnard, Marlize Van Sittert, Venkat Bhat, Sirisha Rambhatla. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

YouTube User Traffic to Paired Epilepsy Education Videos in English and Spanish: Comparative Study

Luna Kimahri Varela, MS; Stephanie Horton, MHA; Ahmed Abdelmoity, MD; Jean-Baptiste Le Pichon, MD, PhD; Mark A Hoffman, PhD

Children's Mercy Hospital Kansas City, Children's Mercy Research Institute, 2401 Gillham Road, Kansas City, MO, United States

Corresponding Author:
Mark A Hoffman, PhD
Children's Mercy Hospital Kansas City, Children's Mercy Research Institute, 2401 Gillham Road, Kansas City, MO, United States

Abstract

Background: Effectively managing epilepsy in children necessitates the active engagement of parents, a factor that is reliant on their understanding of this neurological disorder. Widely available, high-quality, patient-focused, bilingual videos describing topics important for managing epilepsy are limited. YouTube Analytics is a helpful resource for gaining insights into how users of differing backgrounds consume video content.

Objective: This study analyzes traffic to paired educational videos of English and Spanish versions of the same content. By examining the use patterns and preferences of individuals seeking information in different languages, we gained valuable insights into how language influences the use of clinical content.

Methods: Physician experts created epilepsy management videos for the REACT (Reaching Out for Epilepsy in Adolescents and Children Through Telemedicine) YouTube channel about 17 subjects, with an English and Spanish version of each. The Children's Mercy Kansas City neurology clinic incorporated these into the department's educational process. YouTube Analytics enabled analysis of traffic patterns and video characteristics between September 2, 2021, and August 31, 2023.

Results: The Spanish group had higher engagement and click-through rates. The English versions of all videos had 141,605 total impressions, while impressions for the Spanish versions totaled 156,027. The Spanish videos had 11,339 total views, while the English videos had 3366. The views per month were higher for the Spanish videos (mean 472, SD 292) compared to the English set (mean 140, SD 91; P<.001). The two groups also differed in search behavior and external traffic sources, with WhatsApp driving more traffic to the Spanish videos than the English versions (94 views compared to 1). The frequency of search terms used varied by language. For example, "tonic clonic" was the most frequent term (n=372) resulting in views for English videos, while "tipos de convulsiones" (types of convulsions) was the most common expression (n=798) resulting in views for Spanish videos. We noted increased monthly views for all videos after adding tags on YouTube. Before tagging, the mean number of views per month for the English-language group was 61 (SD 28), which increased to 220 (SD 53) post tagging. A similar trend can be observed in the Spanish-language group as well. Before tagging, the mean number of monthly views was 201 (SD 71), which increased to 743 (SD 144) after tagging.

Conclusions: This study showed high traffic for Spanish video content related to epilepsy in a set of paired English/Spanish videos. This highlights the importance of bilingual health content and optimizing video content based on viewer preferences and search behavior. Understanding audience engagement patterns through YouTube Analytics can further enhance the dissemination of clinical video content to users seeking content in their primary language, and tagging videos can have a substantial impact on views.

(JMIR Form Res 2025;9:e56720) doi:10.2196/56720

KEYWORDS

epilepsy; patient education; informatics; social media; biomedical research; social determinants of health; accessibility; engagement; comparative analysis; clinical videos; English; Spanish; neurological disorder; YouTube; bilingual; audience engagement; clinical knowledge

XSL•FO
RenderX

Introduction

Video Health Content

Access to informative and reliable clinical content is crucial for managing complex clinical conditions, especially for individuals in rural areas or with limited access to specialist care [1]. Epilepsy, a common chronic condition, highlights the challenges faced by patients needing access to clinical specialists. Children and youth with epilepsy and their families need accurate and trustworthy information to manage their condition effectively [2].

Social media platforms, particularly YouTube, have become significant sources of information for many people. With billions of users monthly, YouTube has established its presence in over 100 countries worldwide and offers localized versions in 80 languages to serve a diverse global population [3]. However, poor-quality health videos on YouTube can be misleading, but they are still promoted via YouTube's search algorithm that prioritizes popularity, relevance, and view history over content quality and accuracy [4]. This presents a significant challenge for individuals who seek health information without proper guidance or supervision, as they are increasingly exposed to unverified and partially misleading content. The consequences of such exposure can include the promotion of unhealthy habits and activities [5]. A 2023 study about video-sharing platforms revealed a strong negative correlation between high-quality content and popularity regarding videos providing accurate information on cancer and nutrition [6]. Out of 1200 videos, the most reliable videos on YouTube were from universities, professional organizations, and nonprofit physician groups, while unreliable videos included content related to treatments that scored low on the Journal of American Medical Association scale for quality medical content [7]. Analysis of video content using the DISCERN instrument yielded similar results for diabetes content [8]. High-quality health care educational resources developed by experts on widely accessible platforms such as YouTube offer an approach to mitigating these concerns. Through analytics available through YouTube, we can understand the factors influencing viewership and engagement with clinical videos.

REACT

Through the REACT (Reaching Out for Epilepsy in Adolescents and Children Through Telemedicine) project, Children's Mercy Kansas City has developed online educational resources for families and children and youth with epilepsy [9]. One of these resources includes online videos covering different aspects of managing epilepsy and seizures, such as general content about seizures, types of seizures, seizure precautions, alternative treatments, and frequently asked questions. The goal was to provide access to patients and their families, primarily those who spoke English or Spanish. This design provided an opportunity to compare the use of videos with the same clinical content. A preliminary analysis of YouTube Analytics traffic patterns for one of these videos ("What is an absence seizure?") showed high use for the Spanish version [7]. This paper presents a comprehensive traffic analysis of the REACT YouTube channel, comparing the paired English and Spanish versions of otherwise identical content.

Methods

Epilepsy Video Development and Deployment

Videos for epilepsy-related topics prioritized by the clinical team were developed in English and Spanish. A multistep approach was taken when translating the videos from English to Spanish. First, the bilingual research team translated the scripts into Spanish. The scripts were reviewed by bilingual and culturally competent staff before recording the videos. The resulting videos were uploaded to the REACT YouTube channel. The videos for each topic are considered as episodes with two versions.

YouTube Analytics

YouTube Analytics for the REACT channel were accessed to evaluate online traffic patterns to the video content [10]. No personal identifiers were available or sought from YouTube. We did not attempt to correlate patient-prescribed views with the traffic patterns evaluated in this work. GraphPad Prism 9.5.1 (GraphPad Software, Inc) was used to quantify video characteristics, with continuous variables presented as means with SDs and ranges.

Tags are descriptive keywords used on YouTube to improve the discoverability of a video. In August 2022, the study team added tags to each of the videos. To analyze the impact of these tags, total monthly views were separated into a pretagging period (September 2021 to August 2022) and a posttagging period (September 2022 to August 2023).

Impressions are the number of times a video's thumbnail is presented to a user on YouTube. The click-through rate was provided by YouTube Advanced Analytics and represents the percentage of REACT channel thumbnails that were clicked by viewers.

Ethical Considerations

All study data used were anonymous.

Results

Epilepsy Video Development and Deployment

Clinical experts developed video content for 17 topics related to treating epilepsy and seizures. The Children's Mercy Kansas City neurology clinic adopted these videos as part of the patient departure educational process. Children and youth with epilepsy and their families were "prescribed" to view at least one video after each visit. The information and prescribed video were printed in the educational department and given to patients so they could scan a QR code that led them to a REDCapÒ survey. This survey included demographic information, a link to the prescribed YouTube video, and a pre- and postsurvey. Children and youth with epilepsy and families also had the option to view more than one video if desired. The first videos were uploaded to the REACT YouTube channel on August 7, 2020, and all videos were made publicly available by September 2, 2021 [11].

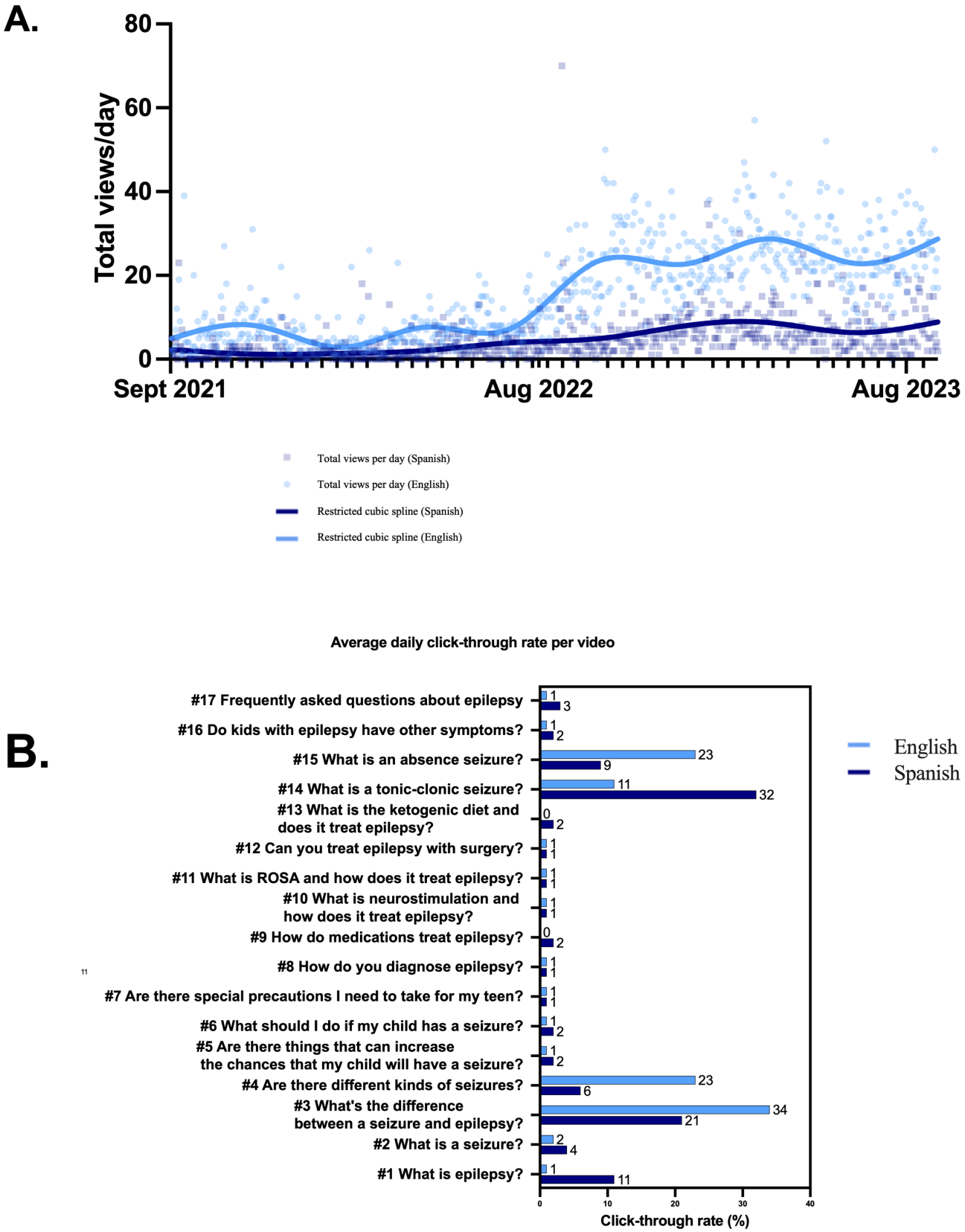
XSL•FO
RenderX

YouTube Analytics

We evaluated traffic patterns between September 2, 2021, when all videos were available, and August 31, 2023. In the aggregate, views per day for the Spanish videos surpassed the English versions (Figure 1A). Compared to the English videos, traffic to the Spanish-language videos was significantly higher ($P<.001$) based on click-through rate (Figure 1B). Some episodes, such as 4 and 14, exhibited different click-through rates between the two languages.

XSL•FO
RenderX

Figure 1. Daily views, by language, of epilepsy YouTube channel. (1A) The total daily views of the REACT (Reaching Out for Epilepsy in Adolescents and Children Through Telemedicine) YouTube channel from September 2, 2021, to August 31, 2023. The x-axis represents the total daily views, and a smoothing spline curve is included. (1B) Average daily click-through rate per video grouped by language. ROSA: robotic stereotactic assistance.

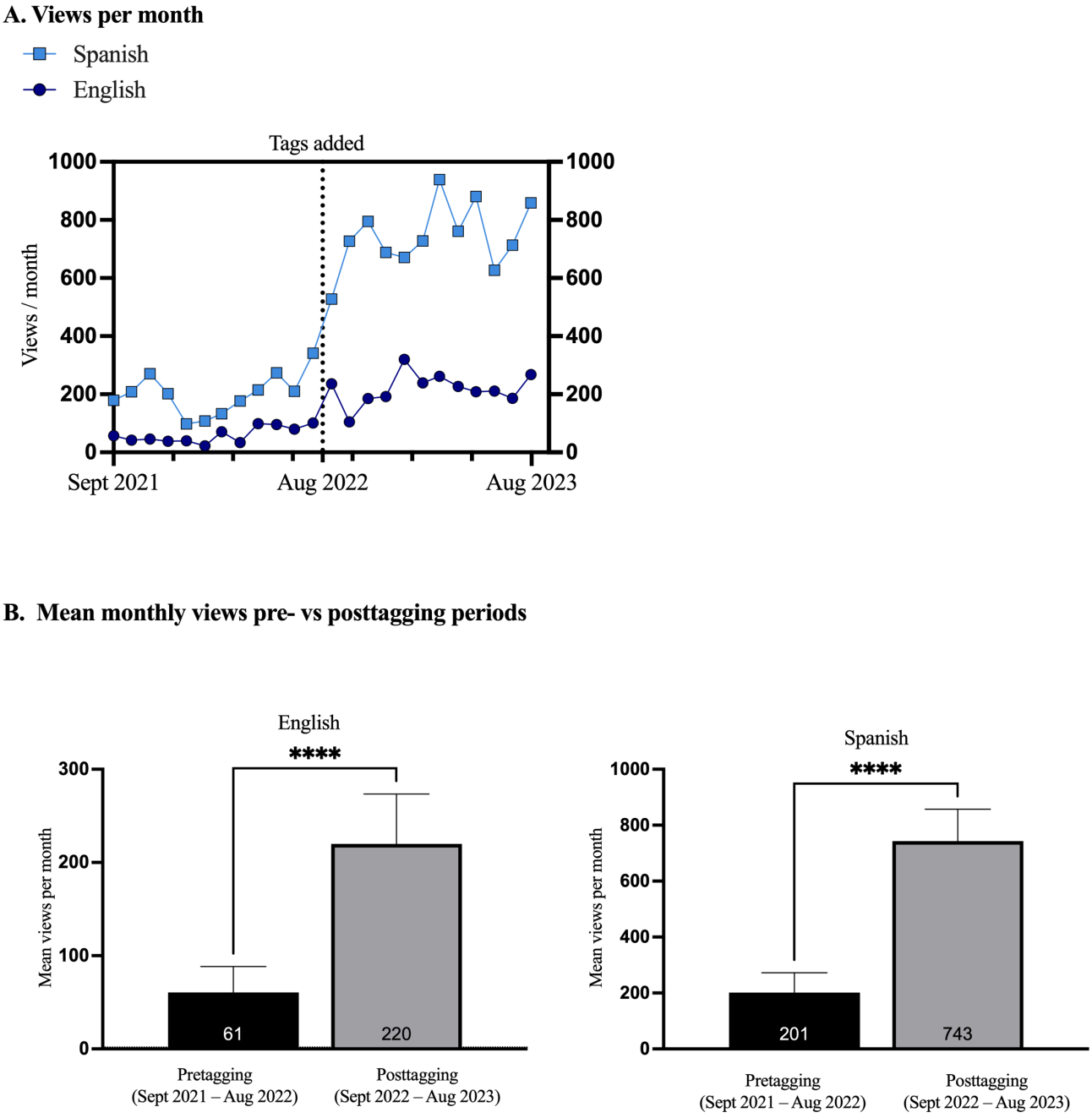


The addition of tags correlates with higher traffic (Figure 2A). The highest total monthly views for the English-language group of videos was 320 in January 2023, while the Spanish-language group of videos' highest monthly views were 939 in March 2023. To further assess the impact of the tags, the total monthly views were divided into two periods: pretagging period (September 2021 to August 2022) and posttagging period (September 2022 to August 2023). In the pretagging period, the

XSL•FO
RenderX

views per month for the English-language group increased after tagging (pretagging mean 61, SD 28; posttagging mean 220, SD 53). A similar trend was observed in the Spanish-language group as well (pretagging mean 201, SD 71; posttagging mean 743, SD 114). Both the English- and Spanish-language videos showed significant increases in traffic after the addition of tags (P<.001; Figure 2B).

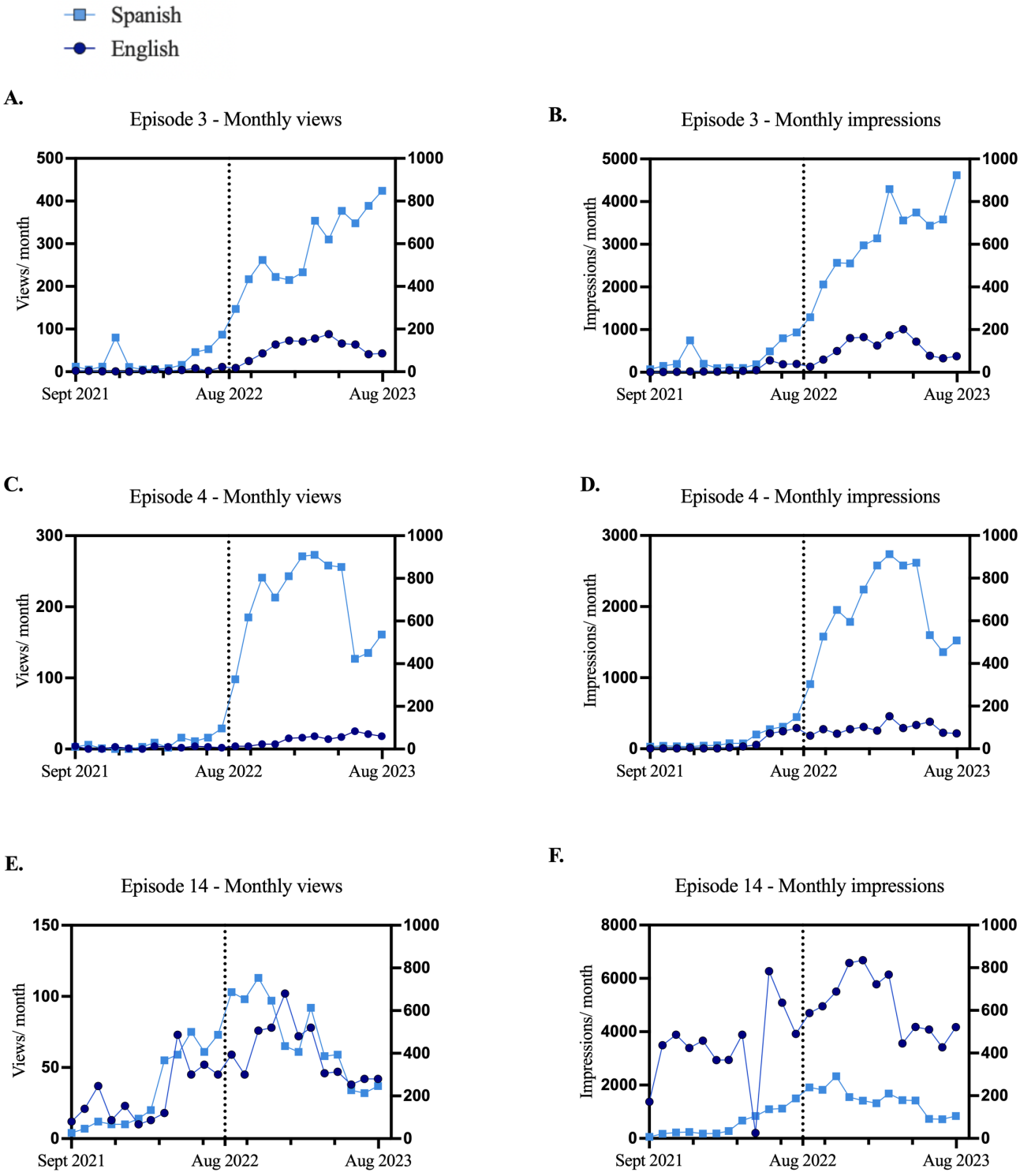
Figure 2. Impact of tagging: Total views per month grouped by language. (2A) Total views per month. Monthly views for all videos on the REACT (Reaching Out for Epilepsy in Adolescents and Children Through Telemedicine) YouTube channel were collected from September 2, 2021, to August 31, 2023, and grouped by language. The dotted line represents the addition of tags. Each data point indicates the total views per language group for that month. (2B) Mean monthly views pre- versus posttagging periods. Comparison of mean total monthly views before and after adding tags in August 2022 for English- and Spanish-language group videos. Paired *t* tests showed a significant difference in views. ****P<.001.



Traffic to individual episodes demonstrated differing patterns. The Spanish versions of videos 3 and 4 showed a strong association between impressions and views (Figure 3A-D). For video 14, impressions and views were similar for the Spanish version, but for the English version, impressions increased earlier (Figure 3E-F). In the English category, video 14 had the highest impressions (n=106,936) and views (n=1087; Table 1). In the Spanish category, video 15 had the highest impressions (n=49,039), followed closely by video 3 (n=41,892). However, video 3 had more views (n=3839) than video 14.

XSL•FO
RenderX

Figure 3. Monthly impressions and views for videos 3, 4, and 14. (3A, C, and E) Monthly views per month across time for videos 3, 4, and 14 from September 2021 to August 2023. (3B, D, and F) Monthly impressions across time for videos 3, 4, and 14.



XSL•FO
RenderX

Table . Impressions, click-through rate, and views for each epilepsy video.

	Video title		Total impressions, n		Click-through rate		Total views, n	
	English title	Spanish title	English	Spanish	English	Spanish	English	Spanish
1	What is epilepsy?	¿Que es la epilepsia?	454	806	4	3	352	106
2	What is a seizure?	¿Qué es una convulsión?	718	3523	4	5	144	255
3	What's the difference between a seizure and epilepsy?	¿Cuál es la diferencia entre la epilepsia y una convulsión?	7727	41,892	7	7	706	3839
4	Are there different kinds of seizures?	¿Hay diferentes tipos de convulsiones?	4327	25,084	2	8	192	2557
5	Are there things that can increase the chances that my child will have a seizure?	¿Hay cosas que pueden aumentar las posibilidades de que tenga una convulsión?	341	902	5	4	51	66
6	What should I do if my child has a seizure?	¿Qué debo hacer si mi hijo tiene una convulsión?	1434	781	1	5	73	75
7	Are there special precautions I need to take for my teen?	¿Se deben tomar precauciones si mi hijo adolescente tiene epilepsia?	276	945	3	6	35	80
8	How do you diagnose epilepsy?	¿Cómo se diagnostica la epilepsia?	318	1519	3	5	33	111
9	How do medications treat epilepsy?	¿Cómo tratan los medicamentos la epilepsia?	1894	543	2	2	62	25
10	What is neurostimulation and how does it treat epilepsy?	¿Qué es la neuroestimulación y cómo trata la epilepsia?	355	2241	7	5	47	131
11	What is ROSA[a] and how does it treat epilepsy?	¿Qué es ROSA y como trata la epilepsia?	326	606	4	4	25	35
12	Can you treat epilepsy with surgery?	¿Se puede tratar la epilepsia con cirugía?	214	1481	4	4	22	74
13	What is the ketogenic diet and does it treat epilepsy?	¿Qué es la dieta Ketogénica?	1106	439	4	4	79	26

XSL•FO
RenderX

	Video title		Total impressions, n		Click-through rate		Total views, n	
	English title	Spanish title	English	Spanish	English	Spanish	English	Spanish
14	What is a tonic-clonic seizure?	¿Cómo cuido a mi hijo con convulsiones tónico´-clónicas?	106,936	23,727	1	4	1087	1249
15	What is an absence seizure?	¿Qué es una crisis de ausencia?	13,319	49,039	2	4	304	2562
16	Do kids with epilepsy have other symptoms?	Otros síntomas de epilepsia	950	1968	2	4	65	115
17	Frequently Asked Questions about Epilepsy	Nuestra experta responde a sus preguntas frecuentes	910	531	6	4	89	33
Totals	—[b]	—	141,605	156,027	1	6	3366	11,339

[a]ROSA: robotic stereotactic assistance.

[b]Not applicable.

While the location of the viewer was only available for 2096 (14.3%) of the 14,705 views, we observed that countries with predominantly non–English-speaking populations tended to have higher view counts reported by YouTube than primarily English-speaking countries. Users in the United States generated the most views (n=884) for the Spanish language, followed by Mexico (n=859) and Argentina (n=110; Multimedia Appendix 1).

The results from the analysis of YouTube search terms used to find English and Spanish videos are presented in Table 2. The top searches that resulted in views for English videos were “tonic,” “clonic,” or “tonic-clonic.” The term “absence seizure” also generated strong interest, resulting in 118 views. Spanish videos had more views for the comparable search term “crisis de ausencia” or “absence seizure,” with 527 views. In comparison, the terms “convulsiones tonico clonico” or “tonic clonic seizures” only garnered 118 views. The search term “tipos de convulsiones” or “types of seizures” resulted in the most views for the Spanish-language group, with 798 views. These findings suggest differences in viewers’ search behavior when seeking content in different languages on YouTube.

XSL•FO
RenderX

Table . Most frequent YouTube search terms leading users to REACT (Reaching Out for Epilepsy in Adolescents and Children Through Telemedicine) content.

Search terms	Views, n
English language	
“tonic clonic” or “clonic” or tonic”	372
“Absence Seizure”	64
“what is a seizure”	17
“difference between seizure and epilepsy”	16
“seizure”	15
“types of seizures”	10
“epilepsy seizure”	8
“different types of seizures”	7
“difference between epilepsy and seizure”	5
“what is epilepsy”	5
Spanish language	
“tipos de convulsiones” (types of seizures)	798
“crisis de ausencia” (absence seizures)	527
“convulsiones tonico clonicas” (tonic-clonic seizures)	118
“que es una convulsión” (what is a seizure)	43
“convulsiones” (seizures)	40
“epilepsia” (epilepsy)	31
“que es la epilepsia” (what is epilepsy)	25
“como es una convulsión” (how is a seizure)	13
“convulsiones en adultos” (seizure’s in adults)	13
“convulsiones y epilepsia” (seizures and epilepsy)	13

Table 3 presents a comprehensive list of external traffic sources used to find video content on the REACT YouTube channel. English viewers predominantly accessed the content through the Children’s Mercy website, with 443 views, compared to the Spanish count of 151. Spanish-language content was embedded on an epilepsy education website maintained by the government of Chile and generated 121 views. Some of the Spanish-language video views originated from a Spanish-language occupational health education consultancy group’s website. Spanish-language viewers relied more heavily on WhatsApp as their preferred means of finding content, with 94 views compared to only 1 among English-language viewers.

XSL•FO
RenderX

Table . Traffic sources used by viewers to find REACT (Reaching Out for Epilepsy in Adolescents and Children Through Telemedicine) content.

Traffic source	Views, n	
	English language	Spanish language
Children's Mercy website	443	161
Google	201	73
YouTube	42	35
Yahoo	2	3
Instagram	1	0
WhatsApp	1	94
bing.com	1	0
live.com	1	5
telegram.org	1	0
revistachilenadeepilepsia.cl	0	121
WhatsApp Business	0	2
precoinprevencion.com	0	10
Facebook	0	12

Much of the traffic to the REACT channel originated from sources outside the YouTube platform. These include internet searches, direct URL entries, bookmarks, or embedded links. External traffic sources accounted for 38,172 (12.8%) of the 297,632 impressions. However, despite the relatively lower number of impressions, these external sources, especially WhatsApp, contributed significantly to the overall viewership. In contrast, videos directly delivered by YouTube accounted for 259,460 (87.2%) impressions. Further analysis examined views by internal versus external sources, grouped by language. The Spanish-language groups contributed 8697 views internally and 2642 views externally to the YouTube platform, while the English videos generated 2508 views from within YouTube and 858 from external sources.

We investigated the impact of YouTube algorithmic suggestions on driving views to REACT content. These suggestions were presented as pop-ups after a video was completed or while users were browsing the website. Within the English group, we observed that videos discussing atonic seizures, tonic-clonic seizures, and the differences between seizures and epilepsy received the highest number of views, with 61, 31, and 14 views, respectively (Multimedia Appendix 2). The Spanish videos addressing convulsions and absence seizures garnered the most attention from these algorithmic recommendations, with 146 and 96 views, respectively.

Discussion

YouTube reaches a larger audience than videos delivered through self-publication to isolated websites, for example, a university website [12]. This study evaluated the viewership of paired Spanish and English educational videos on the YouTube platform and noted substantial traffic to the Spanish-language versions relative to most of the English versions. One notable aspect of our work is the access to detailed analytics for videos with identical clinical content in English and Spanish, a distinguishing feature from other studies evaluating YouTube content for chronic disease [13]. The availability of such analytics provides valuable insights into the use patterns for clinical video content, allowing for informed content design and development strategies. YouTube analytics has substantial, yet underused, value to the developers of clinical content. While the primary intended use of YouTube metrics is to monitor and optimize ad views, we demonstrate that there is utility to these metrics for health education. By leveraging data-driven approaches, the quality and relevance of clinical content can be improved. This analysis enables the promotion of health equity by tailoring content and using search engine optimization features such as tagging to increase visibility.

Our analysis demonstrates a distinct difference in audience engagement between English- and Spanish-language videos providing the same content. Engagement is quantified by the number of views a video generates as well as other metrics, including view time, which we did not evaluate in this study. Overall, the Spanish-language group consistently exhibited higher levels of engagement, as evidenced by increased views and impressions. The modes of interaction differed; for example, Spanish speakers were more likely to use WhatsApp to access the content, consistent with the popularity of this messaging app among Spanish-speaking groups [14,15]. This finding highlights the importance of developing bilingual clinical content to meet the needs of diverse populations.

Furthermore, our study explored the correlation between impressions and views per month, providing insights into variations in monthly engagement across different videos.

The videos with the highest viewing rates were related to the clinical appearance of a seizure ("What's the difference between a seizure and epilepsy?" "Are there different kinds of seizures?" "What is a tonic-clonic seizure?" "What is an absence seizure?"). This is in line with the public misconceptions of what a seizure looks like [16] and further emphasizes the importance of

XSL•FO
RenderX

education regarding the many possible phenomenology of epileptic seizures [17].

While data about audience demographics was limited, our analysis showed that users in countries with predominantly non–English-speaking populations tended to have higher reported view counts of the Spanish versions of a video than users in primarily English-speaking countries. This suggests a global demand for clinical content in languages other than English, highlighting the need for bilingual or multilingual approaches to reach diverse audiences. The search behavior of viewers also differed between languages, as reflected in the top search terms leading to views for English and Spanish videos. Understanding these differences can inform content creators about the keywords and topics that resonate with their audiences, enabling them to optimize their content accordingly. For example, search terms most often leading users to a video can be added as tags.

External traffic sources (non-YouTube) significantly drove viewership. This highlights the importance of leveraging external platforms, such as websites and social media, to promote and distribute clinical video content posted on YouTube. WhatsApp was a frequent source of traffic to the Spanish videos, consistent with common use of this app in non–English-speaking countries [18]. Tailoring strategies to effectively reach and engage different target audiences based on language preferences and behaviors is crucial for maximizing the impact of clinical content dissemination. Our findings also demonstrate that the implementation of tags significantly improved the visibility and accessibility of the videos, leading to a substantial increase in total monthly views. This is consistent with previous research that has shown the effectiveness of tags in enhancing the discoverability of content on YouTube [19].

Finally, our analysis of suggestions generated by the YouTube platform showed variations in viewer preferences between the English and Spanish groups. This underscores the influence of recommendations in directing viewers toward specific content related to seizures and epilepsy. Content creators should consider these recommendations to increase viewership and engagement when developing and promoting their videos.

This study has known limitations. We focused solely on YouTube to evaluate clinical video viewership, and we acknowledge that other platforms may exhibit different use patterns and levels of audience engagement. The analysis was limited to English and Spanish videos, excluding other languages that may generate significant viewership. User privacy settings influence whether their location is visible in analytics, limiting access to granular geographic data and other demographic data (race, gender, age). Furthermore, the study did not explore the long-term effects of video viewership on health outcomes or behavior change. Lastly, we did not consider the influence of external factors, such as marketing or promotion efforts, on video viewership.

We demonstrated the effectiveness of paired educational videos in reaching a global audience well beyond the intended focus of the in-clinic recommendations. Techniques such as tags improve the discoverability and viewership of clinical videos on YouTube. The availability of detailed analytics provides valuable insights into audience engagement and use patterns, enabling informed content design and development strategies. Bilingual or multilingual approaches are essential for reaching diverse populations and promoting health equity. Leveraging external (outside of the YouTube platform) traffic sources and optimizing video content based on search behavior and viewer preferences can further enhance the impact of clinical video dissemination, for example, top YouTube searches that resulted in views related to the differences between types of seizures. The development of paired videos enabled us to perform side-by-side comparisons of traffic patterns and video characteristics on a specific health topic. Through this work, we elucidated practices to improve clinical content available on the YouTube platform for future health content creators seeking to support non–English-speaking users.

Acknowledgments

This work was funded in part by the Health Resources and Services Administration Awareness and Access to Care for Children and Youth with Epilepsy award H98MC332390100. We appreciate the contributions of the Division of Neurology experts who developed the video content and the marketing team of Children's Mercy Kansas City for the video editing. We thank Rose Reynolds for the helpful discussions.

Data Availability

The datasets generated or analyzed during this study are available at [20].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Total view count by country.
[DOCX File, 14 KB - formative_v9i1e56720_app1.docx]

Multimedia Appendix 2

XSL•FO
RenderX

Top 10 traffic sources to REACT (Reaching Out for Epilepsy in Adolescents and Children Through Telemedicine) from other YouTube content grouped by language.
[DOCX File, 16 KB - formative_v9i1e56720_app2.docx]

References

1. Cassidy L, Quirke MB, Alexander D, et al. Integrated care for children living with complex care needs: an evolutionary concept analysis. Eur J Pediatr 2023 Apr;182(4):1517-1532. [doi: 10.1007/s00431-023-04851-2] [Medline: 36780041]
2. Ellis N, Upton D, Thompson P. Epilepsy and the family: a review of current literature. Seizure 2000 Jan;9(1):22-30. [doi: 10.1053/seiz.1999.0353] [Medline: 10667959]
3. YouTube by the numbers secondary YouTube by the numbers 2023. YouTube Blog. URL: https://blog.youtube/press [accessed 2025-03-03]
4. Osman W, Mohamed F, Elhassan M, Shoufan A. Is YouTube a reliable source of health-related information? A systematic review. BMC Med Educ 2022 May 19;22(1):382. [doi: 10.1186/s12909-022-03446-z] [Medline: 35590410]
5. Nour MM, Nour MH, Tsatalou OM, Barrera A. Schizophrenia on YouTube. Psychiatr Serv 2017 Jan 1;68(1):70-74. [doi: 10.1176/appi.ps.201500541] [Medline: 27524369]
6. Basch CH, Menafro A, Mongiovi J, Hillyer GC, Basch CE. A content analysis of YouTube videos related to prostate cancer. Am J Mens Health 2017 Jan;11(1):154-157. [doi: 10.1177/1557988316671459] [Medline: 27694554]
7. Demirci A, Başar H. YouTube is an unreliable source of information about delayed ejaculation treatment. Urologia 2024 May;91(2):403-412. [doi: 10.1177/03915603231204077] [Medline: 37818833]
8. Birch EM, Leziak K, Jackson J, Dahl E, Niznik CM, Yee LM. Content quality of YouTube videos about gestational diabetes: systematic evaluation. JMIR Diabetes 2022 Apr 7;7(2):e30156. [doi: 10.2196/30156] [Medline: 35389355]
9. Le Pichon JB, Horton S, Abdelmoity O, et al. The use of virtual tools in narrowing the impact of health disparities in neurology. Front Pediatr 2022 Oct 14;10:1028833. [doi: 10.3389/fped.2022.1028833] [Medline: 36313873]
10. AlHasan A. The power of YouTube videos for surgical journals. Surgery 2023 Sep;174(3):744-746. [doi: 10.1016/j.surg.2023.05.042] [Medline: 37419760]
11. Children's Mercy KC Neurology-REACT YouTube page. URL: https://www.youtube.com/@reactcmh [accessed 2025-02-28]
12. Topps D, Helmer J, Ellaway R. YouTube as a platform for publishing clinical skills training videos. Acad Med 2013 Feb;88(2):192-197. [doi: 10.1097/ACM.0b013e31827c5352] [Medline: 23269305]
13. Zaila KE, Osadchiy V, Anderson AS, Eleswarapu SV, Mills JN. Popularity and worldwide reach of targeted, evidence-based internet streaming video interventions focused on men's health topics. Transl Androl Urol 2020 Jun;9(3):1374-1381. [doi: 10.21037/tau-20-580] [Medline: 32676422]
14. Anderson E, Koss M, Castro Luque AL, Garcia D, Lopez E, Ernst K. WhatsApp-based focus groups among Mexican-origin women in Zika risk area: feasibility, acceptability, and data quality. JMIR Form Res 2021 Oct 28;5(10):e20970. [doi: 10.2196/20970] [Medline: 34709185]
15. Díaz EN, Le Q, Campos D, et al. Central American parents' preferences for content and modality for a family-centered intervention to promote healthful energy balance-related behaviors of their preschool-age children. Int J Environ Res Public Health 2022 Apr 21;19(9):5080. [doi: 10.3390/ijerph19095080] [Medline: 35564475]
16. Kerson JF, Kerson TS, Kerson LA. The depiction of seizures in film. Epilepsia 1999 Aug;40(8):1163-1167. [doi: 10.1111/j.1528-1157.1999.tb00836.x] [Medline: 10448833]
17. Kilinç S, Campbell C. "It shouldn't be something that's evil, it should be talked about": a phenomenological approach to epilepsy and stigma. Seizure 2009 Dec;18(10):665-671. [doi: 10.1016/j.seizure.2009.09.001] [Medline: 19819737]
18. Rosen AO, Holmes AL, Balluerka N, et al. Is social media a new type of social support? Social media use in Spain during the COVID-19 pandemic: a mixed methods study. Int J Environ Res Public Health 2022 Mar 26;19(7):3952. [doi: 10.3390/ijerph19073952] [Medline: 35409634]
19. Zhou R, Xia D, Wan J, Zhang S. An intelligent video tag recommendation method for improving video popularity in mobile computing environment. IEEE Access 2019 Dec;8. [doi: 10.1109/ACCESS.2019.2961392]
20. Data from: youtube user traffic to paired epilepsy education videos in english and spanish: comparative study. SHARE. 2024. URL: https://scholarlyexchange.childrensmercy.org/data_sets/3/ [accessed 2025-03-13]

Abbreviations

REACT: Reaching Out for Epilepsy in Adolescents and Children Through Telemedicine

XSL•FO
RenderX

Edited by A Mavragani; submitted 24.01.24; peer-reviewed by J Huber; revised version received 08.11.24; accepted 10.11.24; published 13.03.25.

Please cite as:
Varela LK, Horton S, Abdelmoity A, Le Pichon JB, Hoffman MA
YouTube User Traffic to Paired Epilepsy Education Videos in English and Spanish: Comparative Study
JMIR Form Res 2025;9:e56720
URL: https://formative.jmir.org/2025/1/e56720
doi:10.2196/56720

© Luna Kimahri Varela, Stephanie Horton, Ahmed Abdelmoity, Jean-Baptiste Le Pichon, Mark A Hoffman. Originally published in JMIR Formative Research (https://formative.jmir.org), 13.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Privacy Concerns Versus Personalized Health Content—Pregnant Individuals' Willingness to Share Personal Health Information on Social Media: Survey Study

Haijing Hao[1], PhD; Yang W Lee[2], PhD; Marianne Sharko[3], MD; Qilu Li[3], MS; Yiye Zhang[3], PhD

[1]Bentley University, Boston, MA, United States

[2]Northeastern University, Boston, MA, United States

[3]Weill Cornell Medicine, 575 Lexington Ave, New York, NY, United States

Corresponding Author:
Yiye Zhang, PhD
Weill Cornell Medicine, 575 Lexington Ave, New York, NY, United States

Abstract

Background: Often lacking immediate access to care providers, pregnant individuals frequently turn to web-based sources for information to address their evolving physical and mental health needs. Social media has gained increasing prominence as a source of news and information despite privacy concerns and unique risks posed to the pregnant population.

Objectives: This study investigated the extent to which patients may be willing to disclose personal health information to social media companies in exchange for more personalized health content.

Methods: We designed and deployed an electronic survey to pregnant individuals worldwide electronically in 2023. We used the classical Internet Users' Information Privacy Concerns (IUIPC) model to examine how privacy concerns modulate pregnant individuals' behaviors and beliefs regarding risk and trust when using social media for health purposes. Results were analyzed using partial least squares structural equation modeling.

Results: Among 317 respondents who initiated the survey, 84% (265/317) of the respondents remained in the study, providing complete responses. Among them, 54.7% (145/265) indicated willingness to provide their personalized health information for receiving personalized health content via social media, while 26% (69/265) were uncertain and 19.3% (51/265) were opposed. Our estimated IUIPC model results are statistically significant and qualitatively align with the classic IUIPC model for the general population, which was previously found in an e-commerce context. The structural model revealed that privacy concerns (IUIPC) negatively affected trusting beliefs (β=−0.408; P<.001) and positively influenced risk beliefs (β=0.442; P<.001). Trusting beliefs negatively impacted risk beliefs (β=−o.362; P<.001) and positively affected the intention to disclose personal health information (β=o.266; P<.001). Risk beliefs negatively influenced the intention to disclose (β=−0.281; P<.001). The model explained 41.5% of the variance in the intention to disclose personal health information (R^2=0.415). In parallel with pregnant individuals' willingness to share, we find that they have heightened privacy concerns and their use of social media for information seeking is largely impacted by their trust in the platforms. This heightened concern significantly affects both their trusting beliefs, making them less inclined to trust social media companies, and their risk beliefs, leading them to perceive greater risks in sharing personal health information. However, within this population, an increase in trust toward social media companies leads to a more substantial decrease in perceived risks than what has been previously observed in the general population.

Conclusions: We find that more than half of the pregnant individuals are open to sharing their personal health information to receive personalized content about health via social media, although they have more privacy concerns than the general population. This study emphasizes the need for policy regarding the protection of health data on social media for the pregnant population and beyond.

(JMIR Form Res 2025;9:e60862) doi:10.2196/60862

KEYWORDS

privacy concerns; trust; pregnancy; health information seeking; pregnant women; maternal; maternity; childbearing; web-based information; health information; mental health; internet; social support; technology; mobile health; mHealth; digital health; health informatics; social media

XSL•FO
RenderX

Introduction

Background

Advancements in web-based technology have significantly enhanced opportunities for digital social support and web-based health information seeking [1,2]. The trend of searching for health information on the web has gained momentum [3-5], particularly since the COVID-19 pandemic where health care services were overburdened to address individual questions [6,7]. Instead, patients gathered on the web, using social media in particular, for guidance and resources to empower themselves, cope with stress, and strengthen positive social support. This health information seeking behavior is particularly prevalent among pregnant patients [8]. Pregnancy is a vulnerable period that exposes patients to heightened anxiety, depression, and stress, leading to adverse maternal, infant, and family outcomes, disproportionately affecting disadvantaged families [9,10]. The negative impact can be mitigated by interventions from health care providers [11,12]. However, disparities in access to health care, health literacy, socioeconomic status, and neighborhood characteristics strangle equitable access to clinical interventions [12-14]. As a result, many patients resort to web-based information seeking and social support to relieve stressors and resolve individual needs during pregnancy [15]. In fact, people of childbearing age are one of the most active users in the digital space in the United States [16,17]. An estimated three-quarters of the US pregnant population seek pregnancy and birth information on the web, and there is increasing IT acceptance for data sharing [18,19]. Social media is an overwhelmingly popular source for pregnancy information, with myriad of social media accounts and influencers distributing a wide range of information and offering venues for discussion [8,20]. For example, research showed that web-based support and interventions improved health outcomes among pregnant patients with gestational diabetes mellitus, and individualized health interventions resulted in lower odds of developing gestational diabetes mellitus [21].

Yet, privacy and risk concerns are important to consider when using digital health technologies for pregnancy [22,23]. Research has shown that individuals often possess both risk and privacy concerns when turning to social media for health information [24-28]. For example, pregnancy privacy concerns have heightened since the overturn of Dobbs v Jackson by the Supreme Court in the United States, which had protected an individual's right to abortion. This has resulted in variable US state laws criminalizing certain reproductive health care, amplifying privacy concerns for pregnant individuals seeking web-based information about abortions and storing reproductive health information on personal applications [29]. However, despite significant privacy concerns, given patients' overwhelming health needs and the lack of clinical resources to address them, social media remains as a popular source for seeking health information and digital social support.

Objectives

Given the increasing need for information and limited access to prompt responses from health care providers regarding pregnancy-related inquiries, there is a notable tension between privacy concerns and the demand for personalized health communication. Our study aims to fill a significant research gap by applying the Internet Users' Information Privacy Concerns (IUIPC) model to understand pregnant individuals' privacy concerns when seeking health information and social support on social media. While prior studies have used the IUIPC model to examine privacy issues across various digital technologies and populations, none have specifically focused on pregnant individuals—a group that is particularly active on the web and yet vulnerable due to heightened health needs and privacy risks. By investigating how privacy concerns, trusting beliefs, and risk perceptions influence pregnant individuals' willingness to disclose personal health information for personalized content, our research contributes to the literature by providing new insights into the information disclosure behaviors of this unique population on social media. This enhances the understanding of privacy dynamics in digital health contexts and informs strategies for health care providers, policy makers, and social media companies to better support pregnant individuals' needs.

Methods

Study Design

This is a survey study. We developed and deployed a survey on Prolific in March 2023. Prolific is a web-based platform that connects researchers with a pool of prescreened study participants globally. Earlier studies have shown that participants recruited from Prolific provide high-quality data regarding users' perceptions of software and digital platforms [30,31]. The survey questions were hosted on Qualtrics (Qualtrics), a survey software. The survey comprises 41 questions: 22 questions focus on the conceptual model, with 7-point Likert scales, and the remaining questions focus on demographic information, social media questions related to health, and 3 quality-check questions. The survey questions are listed in Table S1 in Multimedia Appendix 1.

Settings

Participants were recruited using Prolific's recruitment method based on inclusion criteria: (1) aged at least 18 years or older, and (2) currently or recently pregnant within 2 years prior to the time of the survey. We used Prolific's prescreening filters to identify eligible participants, including age, gender, and pregnancy status to align with our inclusion criteria. Eligible participants identified through the prescreening process received an invitation to participate in our study via Prolific's messaging system.

The invitation included a brief overview of the study's purpose, estimated time commitment, and the compensation offered. Participants were excluded if their responses in the survey were incomplete or of poor quality, as determined by the quality-check questions. The first quality check is based on age, using the following questions: Q29_1 "What is your age?—Age" and Q14_1 "What is your birth year?—Year," which were placed in 2 different sections of the survey. If the difference between the answers to these 2 questions is more than 2 years, we excluded that participant. The second quality check is based on the Collection construct from the IUIPC model, using the

XSL•FO
RenderX

following questions: Q18_1 “It usually bothers me when health-conscious social media companies ask me for personal health information.” and Q18_5 “It usually doesn’t bother me when health-conscious social media companies ask me for personal health information.” These 2 opposite questions were placed in the survey, and if participants provide similar responses to both (eg, strongly disagree or strongly agree for both), they were excluded. However, if both answers are neutral (neither agree nor disagree), the participants were retained. The third quality check is based on the Risk-Belief construct from the IUIPC model, using the following questions: Q20_1 “In general, it would be risky to give my health information to health-conscious social media companies.” and Q20_5 “I would feel safe giving my health information to health-conscious social media companies.” The same strategy as the second quality check applies to this pair of questions.

Ethical Considerations

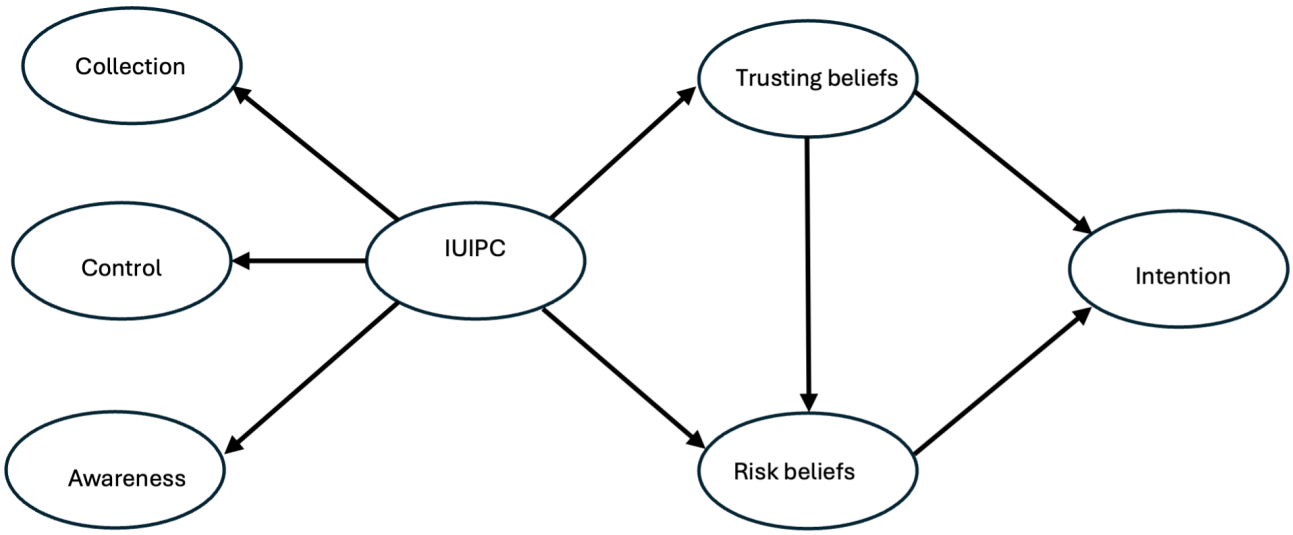
Our study was approved as exempt under institutional review board category II by Bentley University (no. 2023011). Upon accessing the survey link, participants were presented with an informed consent form detailing the study’s objectives, procedures, potential risks, benefits, and the voluntary nature of participation. Participants had to electronically agree to the consent form before proceeding to the survey questions. The study data are fully anonymous. Each participant received a US $1 reward for answering the questions per the site’s rules. Compensation was managed through Prolific’s system to ensure timely and secure payment.

Variables and Data Sources

The survey draws on a conceptual model, the IUIPC model by Malhotra et al [32]. IUIPC model was designed to evaluate whether concerns about information privacy sway an individual’s intent to adopt digital technology in the e-commerce context [32]. Since then, numerous studies have delved into various populations’ privacy concerns across various digital technologies based on the IUIPC model [26,33-38]. However, limited knowledge exists on the perspective of pregnant individuals’ privacy concerns when seeking health information on social media. Thus, we aim to uncover how information privacy concerns may influence pregnant individuals’ intentions to reveal personal health data to receive personalized pregnant-related health content through social media using IUIPC.

Figure 1 depicts the structure of the basic theoretical IUIPC model, consisting of 5 major constructs: collection, control, awareness, trusting beliefs, and risk beliefs [32]. Collection, control, and awareness are 3 first-order constructs, which are directly measured through survey items. IUIPC is a second-order construct, or a latent construct, which is estimated by the 3 first-order constructs. In this study, the collection construct refers to the extent of an individual’s privacy concerns about whether it is worth revealing their personal information to social media companies. The control construct measures the degree to which individuals feel they have control over their personal data and know how the data will be collected, used, and shared. The awareness construct gauges the level of a user’s privacy concern regarding their awareness of social media companies’ transparency and the proper handling of their information. Trusting beliefs refer to the degree to which people believe an entity is trustworthy in protecting their personal information, whereas risk beliefs refer to the expectation of a potential loss when sharing personal information with the entity. In this study, the personal information is personal health information and the entity is social media companies. IUIPC integrates trusting beliefs and risk beliefs to elucidate why a user may have an intention to adopt a new technology for certain benefits. In the context of this study, IUIPC reflects social media users’ concerns about social media companies’ collection of personal information, the users’ control over the collected information, and the users’ awareness of how the collected information is used, which in turn affects users’ intentions to reveal personal health information to receive customized health information through social media.

Figure 1. Internet Users’ Information Privacy Concerns (IUIPC) with 5 major constructs: collection, control, awareness, trusting beliefs, and risk beliefs.



XSL•FO
RenderX

Research Hypotheses

Previous research has demonstrated that users with greater privacy concerns regarding social media tend to have lower trust toward these platforms [36,38]. Therefore, we hypothesize the following:

- Hypothesis 1: For pregnant individuals, IUIPC will have a negative effect on their trusting beliefs.

There is often an inverse relationship between trusting beliefs and risk beliefs because trust implies a firm belief in the reliability of something, while risk indicates the possibility of negative outcomes. When users exhibit higher privacy concerns toward social media, it implies that they perceive a greater risk associated with its use [36,38]. Thus, we propose the following 2 hypotheses:

- Hypothesis 2: For pregnant individuals, IUIPC will have a positive effect on their risk beliefs.
- Hypothesis 3: For pregnant individuals, their trusting beliefs will have a negative effect on their risk beliefs.

Research has shown that trust in a specific technology positively influences individuals' intention to explore that technology [39]. Pregnant individuals will share their personal information in exchange for health information through social media if they trust the companies and find the specific health content valuable enough to justify the privacy trade-off. Previous studies have also shown that trust plays a crucial role in adopting social media [33,34,40]. Therefore, we further hypothesize the following:

- Hypothesis 4: For pregnant individuals, trusting beliefs have a positive effect on their intention to reveal personal health information to social media companies to receive personalized health content through social media.

A previous study on risk beliefs concerning social media suggested that a perception of risk by users made them less likely to adopt social media [38]. If pregnant individuals perceive that sharing personal health information with social media companies for health purposes is risky, then this perception will result in a lower intention to use. Thus, we hypothesize the following:

- Hypothesis 5: For pregnant individuals, risk beliefs have a negative effect on their intention to reveal personal health information to social media companies in exchange for personalized health content through social media.

Measurement Model and Structural Equation Model Evaluation

We used partial least squares structural equation modeling to estimate the IUIPC model and test our hypotheses. Partial least squares structural equation modeling is a variance-based structural equation modeling (SEM) technique suitable for exploratory research and complex models with latent constructs, particularly when the sample size is small and data that may not meet the strict assumptions of covariance-based SEM [41]. We estimated the IUIPC model and assessed the measurement model results according to the following 4 criteria: unidimensionality, item reliability, construct reliability, and convergent validity [39,40]. We used Stata SE 17, and the Stata add-on package, plssem (StataCorp) [42] for the analysis. We reported the *P* values associated with each path coefficient in the structural model to indicate the statistical significance of the hypothesized relationships with significance at the .05 level ($P<.05$). CIs were determined using bootstrapping with 5000 resamples. Dimensionality is the number of constructs that a group of items or a group of survey questions reflects in a formative measurement model. We conducted principal component factor analysis to examine eigenvalues and kept only the constructs whose eigenvalues are >1. Item reliability refers to the extent to which an observed item can reliably measure its corresponding construct. We assume reliability when the standardized loading of an item on their constructs is >0.7. Construct reliability refers to the degree to which the items of a construct consistently and accurately measure that construct. Two commonly used metrics to evaluate construct reliability are Cronbach α and Dillon-Goldstein ρ, and their generally acceptable values are >0.7. Convergent validity evaluates the extent to which a set of items designed to measure the same construct are strongly related to one another. Average variance extracted (AVE) measures the average amount of variance in the items explained by the construct. AVE values should be >0.5, meaning that the construct explains 50% or more of the variance in its indicators. We used the R^2 value and the goodness of fit to evaluate the predictive power of IUIPC, which is a structural model. The acceptable threshold for goodness of fit is 0.36 [43]. Finally, variance inflation factor values are used to check for multicollinearity, with a common threshold of 2.5.

Results

Descriptive Statistics

The data collection lasted for 2 days until we reached the goal of 300 responses. A total of 317 respondents answered the survey. After removing incomplete surveys and responses that did not meet our quality criteria based on the 3 quality check questions, 83.6% (265/317) of the respondents remained in the study. Table 1 shows the sociodemographic information of the respondents. The largest racial group is White (164/265, 61.9%), and the second largest racial group is African American (82/265, 30.9%). Nearly 75% (198/265) of the respondents have a bachelor's degree or above, which matches our expectations and is consistent with previous studies showing that participants on Prolific tend to have higher levels of education than the general population [44].

Most participants (256/265, 96.7%) responded that they receive their pregnancy health information from their doctors, and 69.4% (184/265) of the participants mentioned the web or social media as pregnancy health information sources. Regarding specific social media platforms, 74.7% (198/265) used Instagram, 69.4% (184/265) used Facebook, 55.5% (147/265) used TikTok, and 30.1%(81/265) used Twitter (Now X). A small percentage (9/265, 3.4%) responded that they used PatientsLikeMe, a health-focused social media platform. An additional 4.2% (11/265) named other social media platforms including Reddit, LinkedIn, YouTube, Pinterest, and WhatsApp. Finally, 0.74% (2/265) of participants did not use social media.

XSL•FO
RenderX

Overall, 99.2% (263/265) of respondents used 1 or more social media platforms. Among all 265 participants, 54.7% (145/265) answered that they were comfortable sharing their personal health information in order to receive customized health content on social media, whereas 26% (69/265) were uncertain and 19.3% (51/265) responded that they were not comfortable sharing.

Table . Descriptive statistics of the 265 participants who were enrolled in the survey study.

Participant characteristics		Values, n (%)
Race		
	Asian	8 (3)
	African American	82 (30.9)
	Multirace	5 (1.9)
	White	164 (61.9)
	Other	6 (2.3)
Ethnicity		
	Hispanic	14 (5.3)
	Non-Hispanic	228 (86)
	Prefer not to say	23 (8.7)
Marital status		
	Married	126 (47.6)
	In a relationship	111 (41.9)
	Single[a]	28 (10.6)
Education		
	High school	23 (8.7)
	Some college	44 (16.6)
	Bachelor's degree	141 (53.2)
	Master's degree	47 (17.7)
	Doctoral degree	2 (0.8)
	Professional degree	8 (3)

[a]Single includes separated, divorced, widowed, and single.

Outcome Data

Of those who were willing to share their personal health information to receive personalized health content, all of them used 1 or more popular social media platforms such as Facebook, Instagram, TikTok, or Twitter. Of those who responded that they were not comfortable sharing, 47% (24/51) cited privacy concerns, and equally, 47% (24/51) cited a lack of trust in social media companies and influencers. One respondent commented on the political environment as a concern for sharing. There were no significant differences in ethnicity, race, or education level (bachelor's degree or more vs up to high school) in the responses. Among users of the top four social media platforms, 56.1% (111/198) of Instagram users, 56.5% (104/184) of Facebook users, 64% (94/147) of TikTok users, and 65.4% (53/81) of Twitter (now X) users were comfortable sharing personal health information in exchange for more customized social media content. There was a significant difference in willingness between people who use and do not use TikTok, and between those who do and do not use Twitter (now X).

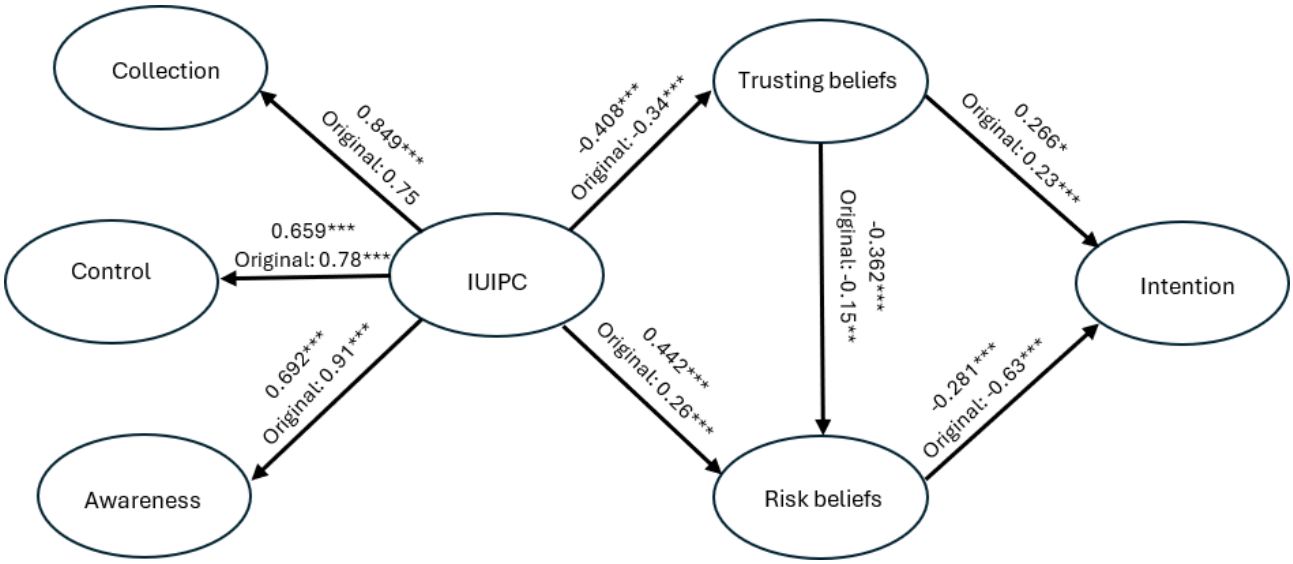
Measurement Model Result

Overall, the measurement model results demonstrate that our constructs exhibit strong reliability and validity, confirming their appropriateness for use in the structural analysis. The collection, control, and awareness constructs all had eigenvalues >1. Thus, they were kept in the model as unidimensional constructs. The loadings and constructs' AVE values are shown in Tables S2 and S3 in Multimedia Appendix 1. As Table S2 in Multimedia Appendix 1 shows, most of the bolded diagonal block loadings are >0.7 for the corresponding constructs, which demonstrate the items' strong reliability. Although a few items on the second-order construct, IUIPC, are slightly <0.7, we still accept this model result because these loadings are not trivial, ranging from 0.44 to 0.58. In addition, they are still statistically significant. Furthermore, as Figure 2 shows, among the estimated coefficients for the structural model, all the path coefficients are statistically significant and are in the directions we expected, which confirms that those items are meaningful for our model. In Table S2 in Multimedia Appendix 1, we can see that the 2 commonly used metrics for evaluating construct reliability, Cronbach α and Dillon-Goldstein ρ, are >0.7, above

XSL•FO
RenderX

the acceptable threshold value, indicating good construct reliability of our model. As Table S3 in Multimedia Appendix 1 shows, all constructs' AVE values are >0.5 except for the IUIPC, the second order construct. This AVE value is <0.5 because of the lower loadings of the control and awareness items on IUIPC, as shown in Table S2 in Multimedia Appendix 1. Finally, the model's composite reliability scores, as measured by Dillon-Goldstein ρ of all constructs (Table S2 in Multimedia Appendix 1), are >0.7, indicating good internal consistency of the IUIPC model.

Figure 2. Internet Users' Information Privacy Concerns (IUIPC) model with estimated path coefficients all statistically significant and are in the directions we expected, confirming that those items are meaningful for our model. The asterisks denote statistical significance based on *P* values. *P<.05, **P<.01. ***P<.001



Structural Model Result

Our estimated IUIPC model results and the 5 path coefficients of the structural model support all 5 hypotheses that we constructed. The IUIPC model's average R^2 value is 0.4146, indicating a moderate effect. The absolute goodness of fit is 0.4923, which is >0.36, the acceptable threshold [43]. All variance inflation factor values were <2.5, the commonly suggested threshold value (Table S4 in Multimedia Appendix 1), indicating little concern for multicollinearity in our structural model. As Figure 2 shows, all path coefficients are statistically significant and in the directions as we expected, and most of them are qualitatively consistent with the original IUIPC model estimates from the study by Malhotra et al [32].

Discussion

Interpretation

Our survey of 265 pregnant individuals globally finds that more than half of pregnant individuals are open to sharing their personal health information to receive personalized content about health via social media, although they have more privacy concerns than the general population. In the original study by Malhotra et al [32], the coefficient between the collection construct and the IUIPC was not statistically significant (β=0.75). However, we observed significance in this coefficient in our model (β=0.849; *P*<.001). This finding was supported by another study by Zeng et al [38], which examined the privacy concerns on social media platforms among the general population and found this coefficient to be at a similar significant level to ours (β=0.36; *P*<.001), albeit smaller. This may indicate that pregnant individuals are more careful about sharing their health-related information with social media companies, reflecting their privacy concerns. In this study, the loadings for both the control and awareness constructs are slightly lower than those in the original study by Malhotra et al [32]. Both our study and the study by Zeng et al [38] have loadings of control and awareness that are lower than results from the classic model by Maholtra et al [32]. This potentially indicates that today, social media users' information control and social media users' awareness of the social media company's practices have a strong impact on their IUIPC concerns, although their quantitative effects are lower than those previously found in the e-commerce setting.

Our model results suggest that the impact of trusting beliefs on pregnant individuals' intention to adopt social media for health purposes is greater than that found in the study by Malhotra et al [32]. The IUIPC in our model negatively affects trusting beliefs, evidenced by a path coefficient of β=−0.408 (*P*<.001), which is consistent with but stronger than the original model's path coefficient of β=−0.34 (*P*<.001). This difference may suggest that pregnant individuals' privacy concerns regarding trust are more pronounced than those found in the classical model [32]. The IUIPC construct in our model positively influences risk beliefs, with a coefficient of β=0.442 (*P*<.001), which is larger than the coefficient in the original model (β=0.26; *P*<.001). This suggests that, in addition to trust, pregnant individuals' privacy concerns also have a stronger influence on their risk beliefs than those of the general population in the e-commerce context. In our model, trusting beliefs negatively impact risk beliefs, with a coefficient of β=−0.362 (*P*<.001), which is stronger than the original model's coefficient of β=−0.15 (*P*<.001). This suggests that if a pregnant individual's trust in a social media company increases, their

XSL•FO
RenderX

perceived risks will decrease more than in the general population in the e-commerce context as previously found [32]. Trusting beliefs positively influence pregnant individuals' intention to reveal health information to receive customized health content through social media with a coefficient of β=0.266 (*P*<.001), which is larger than the original model (β=0.23; *P*<.001). Conversely, risk beliefs have a negative effect on this intention with a coefficient of β=−0.281 (*P*<.001), which is smaller in magnitude than the original model (β=−0.63; *P*<.001). This is consistent with a previous study, which showed that trust is positively associated with adoption intention while privacy concerns are negatively associated with the adoption intention of technology in health care [45].

Our study adds to our current understanding that pregnant individuals are increasingly turning to social media for health information and support, despite significant privacy and risk concerns. For health care providers, this underscores the importance of integrating digital tools into their practice to meet patients where they are seeking information. By establishing a trustworthy web-based presence and providing accurate, evidence-based health information through social media, providers can better address the informational needs of pregnant patients. Educating patients about the potential risks associated with sharing personal health information on the web is also crucial. This includes guidance on maintaining privacy, recognizing credible sources, and understanding how personal data might be used by third parties. Enhancing the patient-provider relationship through open communication about privacy concerns can reduce the likelihood of patients seeking information from less reliable sources. For policy makers, our findings confirm a pressing need to safeguard personal health information on the web and in social media. Developing robust regulatory frameworks that govern how social media companies collect, use, and share personal health data is essential. Updating and enforcing privacy laws to encompass social media platforms can provide additional protections for sensitive health information. Policy makers can also educate individuals about digital privacy, helping pregnant individuals make informed decisions about sharing personal health information on the web.

Our results suggest that pregnant individuals are concerned about their personal health data being collected by social media companies, about losing control of their health information collected by social media companies, and about how their health information will be used. Privacy and confidentiality are paramount, as pregnant individuals may inadvertently disclose sensitive information that could be misused and harm them. Therefore, prior to collecting personal health information, social media companies with health-related content should explain clearly that the purpose of data collection is to create personalized health content for each user's benefit. Furthermore, social media companies should guarantee that patients can easily access any information they have provided to ensure users' rights to control their data. Users should also be informed, in clear language, of the companies' privacy policies, including if and how their health data will be used. This transparency and control over their data may help build trust between pregnant individuals and the social media companies.

Limitations

This study has limitations. First, since we distributed the survey to participants globally, various cultural differences, religions, and government policies may influence the perceptions of the participants. Second, our sample size was small and limited to web-based survey participants, which may have introduced a built-in bias toward the digital space. Our sample, recruited from Prolific, likely comprises individuals who are more digitally savvy compared with the general population, as reflected in their higher education levels and potentially greater web-based literacy, which may introduce selection bias and affect the generalizability of our findings. Future studies will recruit larger samples while specifically target participants with low web-based health literacy to ensure adequate representation of their perspectives and investigate whether there is any difference in the types of personal health data that participants are willing to share. Relatedly, while we did ask each participant which social media platforms they use, our small sample size did not allow for extensive platform-specific comparisons. In addition, our survey did not specify any social media or web-based health forum specifically but referred to general social media for health. Considering that privacy policies of various health-related social media platforms differ, users' trusting beliefs and risk beliefs may differ across social media platforms. Furthermore, we acknowledge while our SEM allows for the testing of directional hypotheses, causality cannot be definitively established due to the cross-sectional nature of the data. Future studies should investigate additional factors such as sociodemographic moderating effects that may influence pregnant individuals' intentions to use social media or web-based health platforms for health information seeking and communication. Moreover, future studies may investigate the benefit or cost of health information social media despite privacy concerns and consider extending our study by exploring how other vulnerable populations such as patients with chronic illnesses or mental health conditions navigate privacy and trust issues in the digital space. Finally, future research can consider designing and developing customized constructs to add to the classical IUIPC model. This would allow for examining how these tailored constructs impact pregnant patients' privacy concerns on social media.

Conclusions

In this study, we examined pregnant individuals' willingness to disclose personal health information in exchange for more personalized health content on social media. Based on a classical model of IUIPC, we designed and deployed a survey for pregnant patients on the web. As found in previous studies such as those by Malhotra et al [32] and Zeng et al [38], our findings align with this consensus that higher privacy concerns negatively impact trust and positively influence risk perceptions in digital contexts. We found that pregnant individuals exhibit more pronounced social privacy concerns than those identified in the classic IUIPC model. This heightened concern significantly affects both their trusting beliefs, making them less inclined to trust social media companies, and their risk beliefs, leading them to perceive greater risks in sharing personal health information. However, within this population, an increase in trust toward social media companies leads to a more substantial

XSL•FO
RenderX

decrease in perceived risks than what has been previously observed in the general population. Thus, social media companies delivering personalized-related health content should prioritize building trust with their users. This involves clearly informing patients about what health data are collected, providing them with easy access to their own health data, keeping them informed about how their health data are used, and including information about privacy protection policies. Given pregnant individuals' needs for web-based health information seeking and the level of openness toward information sharing found from this study, we argue that more regulations are needed to mitigate users' privacy and risk concerns and protect their data against exploitation on social media.

Acknowledgments

We sincerely thank Maya Daiter, Yiliang Zhou, and Haiyu Zheng for deploying the surveys used in the study. We sincerely thank Dr Sergio Venturini for his patience and thorough explanation of the usage of the plssem package, developed by Dr Sergio Venturini and Dr Mehmet Mehmetoglu, for our model estimation.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

HH initiated the study, designed the survey, analyzed the data, conducted model estimation and analyses, wrote and revised the manuscript. YWL provided advice on the paper's structure and assisted with revising the model and writing the paper. MS contributed a clinical perspective, reviewed and revised the manuscript, and provided references for related health care studies. QL assisted in data collection, data cleaning, literature review, and paper format polishing. YZ initiated the study, provided funding for data collection, designed and deployed the survey, analyzed data and model, and wrote and revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File, 26 KB - formative_v9i1e60862_app1.docx]

References

1. Hao H, Levkoff S, Wang W, et al. Studying online support for caregivers of patients with Alzheimer's disease in China: a text-mining approach to online forum in China. Int J Healthc Inf Syst Inform 2020;15:1-17. [doi: 10.4018/IJHISI.2020100101]
2. Hao H, Zhang K. The voice of Chinese health consumers: a text mining approach to web-based physician reviews. J Med Internet Res 2016 May 10;18(5):e108. [doi: 10.2196/jmir.4430] [Medline: 27165558]
3. Jia X, Pang Y, Liu LS. Online health information seeking behavior: a systematic review. Healthcare (Basel) 2021;9(12):1740. [doi: 10.3390/healthcare9121740]
4. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec;34(4):268-283. [doi: 10.1111/hir.12192] [Medline: 29045011]
5. Chen J, Wang Y. Social media use for health purposes: systematic review. J Med Internet Res 2021 May 12;23(5):e17917. [doi: 10.2196/17917] [Medline: 33978589]
6. Chatwin J, Butler D, Jones J, James L, Choucri L, McCarthy R. Experiences of pregnant mothers using a social media based antenatal support service during the COVID-19 lockdown in the UK: findings from a user survey. BMJ Open 2021 Jan 17;11(1):e040649. [doi: 10.1136/bmjopen-2020-040649] [Medline: 33455927]
7. Neely S, Eldredge C, Sanders R. Health information seeking behaviors on social media during the COVID-19 pandemic among American social networking site users: survey study. J Med Internet Res 2021 Jun 11;23(6):e29802. [doi: 10.2196/29802] [Medline: 34043526]
8. Lee JY, Lee E. What topics are women interested in during pregnancy: exploring the role of social media as informational and emotional support. BMC Pregnancy Childbirth 2022 Jun 26;22(1):517. [doi: 10.1186/s12884-022-04842-5] [Medline: 35754031]
9. Dunkel Schetter C, Tanner L. Anxiety, depression and stress in pregnancy: implications for mothers, children, research, and practice. Curr Opin Psychiatry 2012 Mar;25(2):141-148. [doi: 10.1097/YCO.0b013e3283503680] [Medline: 22262028]
10. Walsh K, McCormack CA, Webster R, et al. Maternal prenatal stress phenotypes associate with fetal neurodevelopment and birth outcomes. Proc Natl Acad Sci U S A 2019 Nov 26;116(48):23996-24005. [doi: 10.1073/pnas.1905890116] [Medline: 31611411]

XSL•FO
RenderX

11. Peahl AF, Zahn CM, Turrentine M, et al. The Michigan plan for appropriate tailored health care in pregnancy prenatal care recommendations. Obstet Gynecol 2021;138:593. [doi: 10.1097/AOG.0000000000004531]
12. Panaite V, Bowersox NW, Zivin K, Ganoczy D, Kim HM, Pfeiffer PN. Individual and neighborhood characteristics as predictors of depression symptom response. Health Serv Res 2019 Jun;54(3):586-591. [doi: 10.1111/1475-6773.13127] [Medline: 30829408]
13. Zhang Y, Tayarani M, Wang S, et al. Identifying urban built environment factors in pregnancy care and maternal mental health outcomes. BMC Pregnancy Childbirth 2021 Dec;21(1):599. [doi: 10.1186/s12884-021-04056-1]
14. Adhikari K, Patten SB, Williamson T, et al. Neighbourhood socioeconomic status modifies the association between anxiety and depression during pregnancy and preterm birth: a community-based Canadian cohort study. BMJ Open 2020 Feb 10;10(2):e031035. [doi: 10.1136/bmjopen-2019-031035] [Medline: 32047008]
15. Bedaso A, Adams J, Peng W, Sibbritt D. The relationship between social support and mental health problems during pregnancy: a systematic review and meta-analysis. Reprod Health 2021 Jul 28;18(1):162. [doi: 10.1186/s12978-021-01209-5] [Medline: 34321040]
16. Harpel T. Pregnant women sharing pregnancy-related information on Facebook: web-based survey study. J Med Internet Res 2018 Mar 22;20(3):e115. [doi: 10.2196/jmir.7753] [Medline: 29567636]
17. Lupton D. The use and value of digital media for information about pregnancy and early motherhood: a focus group study. BMC Pregnancy Childbirth 2016 Jul 19;16(1):171. [doi: 10.1186/s12884-016-0971-3] [Medline: 27435182]
18. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016 Mar 28;16:65. [doi: 10.1186/s12884-016-0856-5] [Medline: 27021727]
19. Rahdar S, Montazeri M, Mirzaee M, Ahmadian L. The relationship between e-health literacy and information technology acceptance, and the willingness to share personal and health information among pregnant women. Int J Med Inform 2023 Oct;178:105203. [doi: 10.1016/j.ijmedinf.2023.105203] [Medline: 37688834]
20. Smith M, Mitchell AS, Townsend ML, Herbert JS. The relationship between digital media use during pregnancy, maternal psychological wellbeing, and maternal-fetal attachment. PLoS One 2020;15(12):e0243898. [doi: 10.1371/journal.pone.0243898] [Medline: 33326465]
21. Cheng Z, Hao H, Tsofliou F, Katz MD, Zhang Y. Effects of online support and social media communities on gestational diabetes: a systematic review. Int J Med Inform 2023 Dec;180:105263. [doi: 10.1016/j.ijmedinf.2023.105263] [Medline: 37907014]
22. Wanni Arachchige Dona S, Angeles MR, Nguyen D, Cooper P, Sweet L, Peeters A. An assessment of ten popular pregnancy applications (apps) available for women in Australia. Int J Med Inform 2024 Apr;184:105354. [doi: 10.1016/j.ijmedinf.2024.105354] [Medline: 38325121]
23. Valencia SA, Barrientos Gómez JG, Gómez Ramirez MC, et al. Evaluation of a telehealth program for high-risk pregnancy in a health service provider institution. Int J Med Inform 2023 Nov;179:105234. [doi: 10.1016/j.ijmedinf.2023.105234] [Medline: 37776668]
24. Wijayanti RP, Handayani PW, Azzahro F. Intention to seek health information on social media in Indonesia. Procedia Comput Sci 2022;197:118-125. [doi: 10.1016/j.procs.2021.12.125]
25. Petersen C, Lehmann CU. Social media in health care: time for transparent privacy policies and consent for data use and disclosure. Appl Clin Inform 2018 Oct;9(4):856-859. [doi: 10.1055/s-0038-1676332] [Medline: 30485880]
26. Li Y, Wang X. Seeking health information on social media: a perspective of trust, self-determination, and social support. J Organ End User Comput 2018;30:1-22. [doi: 10.4018/JOEUC.2018010101]
27. Fan K, Jiang W, Li H, Yang Y. Lightweight RFID protocol for medical privacy protection in IoT. IEEE Trans Ind Inf 2018;14(4):1656-1665. [doi: 10.1109/TII.2018.2794996]
28. Grande D, Mitra N, Iyengar R, et al. Consumer willingness to share personal digital information for health-related uses. JAMA Netw Open 2022 Jan 4;5(1):e2144787. [doi: 10.1001/jamanetworkopen.2021.44787] [Medline: 35072717]
29. Clayton EW, Embí PJ, Malin BA. Dobbs and the future of health data privacy for patients and healthcare organizations. J Am Med Inform Assoc 2022 Dec 13;30(1):155-160. [doi: 10.1093/jamia/ocac155] [Medline: 36048014]
30. Douglas BD, Ewell PJ, Brauer M. Data quality in online human-subjects research: comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. PLoS One 2023;18(3):e0279720. [doi: 10.1371/journal.pone.0279720] [Medline: 36917576]
31. Peer E, Rothschild D, Gordon A, Evernden Z, Damer E. Data quality of platforms and panels for online behavioral research. Behav Res 2021;54(4):1643-1662. [doi: 10.3758/s13428-021-01694-3]
32. Malhotra NK, Kim SS, Agarwal J. Internet Users' Information Privacy Concerns (IUIPC): the construct, the scale, and a causal model. Inf Syst Res 2004 Dec;15(4):336-355. [doi: 10.1287/isre.1040.0032]
33. Adhikari K, Panda RK. Users' information privacy concerns and privacy protection behaviors in social networks. J Glob Mark 2018 Mar 15;31(2):96-110. [doi: 10.1080/08911762.2017.1412552]
34. Mohamed N, Ahmad IH. Information privacy concerns, antecedents and privacy measure use in social networking sites: evidence from Malaysia. Comput Human Behav 2012 Nov;28(6):2366-2375. [doi: 10.1016/j.chb.2012.07.008]
35. Harborth D, Pape S. JonDonym users' information privacy concerns. In: Janczewski LJ, Kutyłowski M, editors. ICT Systems Security and Privacy Protection: Springer International Publishing; 2018:170-184.

XSL•FO
RenderX

36. Harborth D, Pape S. How privacy concerns, trust and risk beliefs, and privacy literacy influence users' intentions to use privacy-enhancing technologies: the case of Tor. SIGMIS Database 2020 Jan 21;51(1):51-69. [doi: 10.1145/3380799.3380805]
37. Pape S, Ivan A, Harborth D, et al. Re-evaluating internet users' information privacy concerns: the case in Japan. AIS Trans Replic Res 2020. [doi: 10.17705/1atrr.00061]
38. Zeng M, Lin S, Armstrong D. Are all internet users' information privacy concerns (IUIPC) created equal? Trans Rep Res 2020;6. [doi: 10.17705/1ATRR.00046]
39. Mcknight DH, Carter M, Thatcher JB, et al. Trust in a specific technology: an investigation of its components and measures. ACM Trans Manage Inf Syst 2011;2:1-25. [doi: 10.1145/1985347.1985353]
40. Cain JA, Imre I. Everybody wants some: collection and control of personal information, privacy concerns, and social media use. New Media Soc 2022 Dec;24(12):2705-2724. [doi: 10.1177/14614448211000327]
41. Mehmetoglu M, Venturini S. Structural Equation Modelling with Partial Least Squares Using Stata and R: CRC Press; 2021.
42. Venturini S, Mehmetoglu M. plssem: a Stata package for structural equation modeling with partial least squares. J Stat Soft 2019;88(8). [doi: 10.18637/jss.v088.i08]
43. Wetzels MGM, Odekerken-Schröder GJ, van Oppen C. Using PLS path modeling for assessing hierarchical construct models: guidelines and empirical illustration. MIS Q 2009;33(1):177. [doi: 10.2307/20650284]
44. Samiotis IP, Qiu S, Lofi C, Yang J, Gadiraju U, Bozzon A. An analysis of music perception skills on crowdsourcing platforms. Front Artif Intell 2022;5:828733. [doi: 10.3389/frai.2022.828733] [Medline: 35774636]
45. Dhagarra D, Goswami M, Kumar G. Impact of trust and privacy concerns on technology acceptance in healthcare: an Indian perspective. Int J Med Inform 2020 Sep;141:104164. [doi: 10.1016/j.ijmedinf.2020.104164] [Medline: 32593847]

Abbreviations

AVE: average variance extracted
IUIPC: Internet Users' Information Privacy Concerns
SEM: structural equation modeling

Edited by A Mavragani; submitted 23.05.24; peer-reviewed by A Jain, L Ahmadian, RS Gomaa Mahmoud, U Sinha, Z Hou; revised version received 19.11.24; accepted 27.11.24; published 10.02.25.

Please cite as:
Hao H, Lee YW, Sharko M, Li Q, Zhang Y
Privacy Concerns Versus Personalized Health Content—Pregnant Individuals' Willingness to Share Personal Health Information on Social Media: Survey Study
JMIR Form Res 2025;9:e60862
URL: https://formative.jmir.org/2025/1/e60862
doi:10.2196/60862

© Haijing Hao, Yang W Lee, Marianne Sharko, Qilu Li, Yiye Zhang. Originally published in JMIR Formative Research (https://formative.jmir.org), 10.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Instagram Posts Promoting Colorectal Cancer Awareness: Content Analysis of Themes and Engagement During Colorectal Cancer Awareness Month

Aditi Srivastava[1]; Jim P Stimpson[2], PhD

[1]Frisco High School, Frisco, TX, United States

[2]Department of Health Economics, Systems, and Policy, Peter O'Donnell Jr. School of Public Health, The University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX, United States

Corresponding Author:
Jim P Stimpson, PhD
Department of Health Economics, Systems, and Policy, Peter O'Donnell Jr. School of Public Health, The University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX, United States

Abstract

Background: Colorectal cancer (CRC) is a leading cause of cancer-related deaths worldwide, with early detection and screening being critical for reducing mortality. Social media platforms like Instagram offer a unique opportunity to raise awareness about CRC, particularly during designated awareness months. However, there is limited research on the effectiveness of CRC-related content on Instagram.

Objective: This study aims to examine how Instagram is used to raise awareness about CRC during Colorectal Cancer Awareness Month by analyzing the thematic content and engagement metrics of related posts. The research seeks to identify the prevalent themes, assess audience interaction with these messages, and highlight areas for improvement in leveraging Instagram as a tool for cancer awareness campaigns.

Methods: A total of 150 Instagram posts were collected based on their use of specific hashtags related to CRC awareness (#colorectalcancer, #colorectalcancerawareness, #colorectalcancerawarenessmonth) during March 2024. The text and images in the posts were categorized into themes such as screening and early detection, symptoms, general awareness, risk factors, individual's experiences, representation of racial and ethnic minoritized communities, and representation of women. Engagement metrics, including the number of likes and comments, were also analyzed. Two researchers independently coded the posts, achieving high interrater reliability (Cohen κ=0.93).

Results: Organizational accounts were more active, contributing 82% (n=123) of the 150 posts, compared to 18% (n=27) from individual users. The most frequently mentioned theme was screening and early detection, which made up 37.3% (n=56) of all posts. General awareness came in second at 19.3% (n=29), and risk factors came in third at 12% (n=18). Posts about individual experiences and general awareness received the highest engagement, indicating the effectiveness of personal narratives and broad informational content. Themes related to symptoms and representation of racial and ethnic minoritized communities and women were underrepresented.

Conclusions: This study highlights the potential of Instagram as a platform for promoting CRC awareness, particularly through posts about screening and early detection and personal experiences. However, there is a need for more inclusive and diverse content to ensure a broader reach and impact.

(JMIR Form Res 2025;9:e63344) doi:10.2196/63344

KEYWORDS

social media; colorectal neoplasms; early detection of cancer; public health; health inequities; harnessing; Instagram; colorectal cancer; colorectal cancer awareness; content analysis; cancer-related deaths; detection; screening; mortality; post; early detection

Introduction

Colorectal cancer (CRC) remains one of the most prevalent and deadly forms of cancer worldwide [1,2]. Despite the clear benefits of early detection and screening for reducing mortality rates, many individuals remain unaware of the importance of regular screenings and the risk factors associated with CRC [3-5]. March is designated as Colorectal Cancer Awareness Month, a period dedicated to increasing public knowledge about CRC, promoting early detection and screening, and ultimately reducing the incidence and mortality of this disease [6]. During this month, various health organizations, advocacy groups, and individuals spread awareness through different communication channels, including social media.

XSL•FO
RenderX

Instagram, a visual-centric platform with over a billion users, offers a unique opportunity to disseminate health information widely and engage with a diverse audience [7-9]. The platform's use of hashtags allows for the categorization and amplification of posts, making it easier for users to find and share relevant content. In the context of health awareness campaigns, hashtags such as #colorectalcancer, #colorectalcancerawareness, and #colorectalcancerawarenessmonth can help consolidate posts under a common theme, enhancing the visibility of key messages [10].

Instagram has been shown to be instrumental in disseminating cancer prevention material. A study of Brazilian Instagram accounts during a nationwide cervical cancer campaign revealed an emphasis on secondary prevention measures, such as Pap screenings, and primary prevention strategies, like the human papillomavirus (HPV) vaccine, received less attention [11]. A descriptive study of Instagram posts related to breast cancer found that the most common posts were individual stories or posts discussing treatment [12]. A separate investigation examined the portrayal of skin cancer narratives on Instagram and underscored the platform's potential for sharing personal experiences and fostering prevention efforts and awareness [13]. A study that searched specific hashtags on Instagram for posts about acute myeloid leukemia found that the most frequently posted type of content by patients were health updates, and a quarter of these posts described symptom experiences [14]. Finally, a content analysis was conducted on information related to CRC on Instagram using the hashtag #colorectalcancer and found the most common posts supported someone with CRC, provided an individual story, or discussed treatment [15]. Collectively, these studies highlight Instagram's potential as a tool for raising cancer awareness and promoting preventive behaviors, while emphasizing the need for further research to identify the most impactful types of cancer prevention messaging and to optimize social media platforms for effective dissemination of cancer prevention information [16-20].

Despite the potential of Instagram to raise awareness about CRC, there is limited research on how effectively the platform is used for this purpose, particularly during dedicated cancer awareness months. This study aims to fill this gap by examining the engagement of Instagram posts related to CRC during Colorectal Cancer Awareness Month. By analyzing the engagement metrics of these posts, this research seeks to understand the current state of CRC awareness on Instagram and identify areas to improve cancer communication strategies. By analyzing Instagram posts based on specific hashtags and thematic content, the findings offer insights into the types of messages that resonate most with the audience and highlight the strengths and weaknesses of current CRC awareness efforts on social media. Ultimately, this research contributes to the broader goal of enhancing public health campaigns and increasing the effectiveness of social media as a tool for health promotion.

Methods

Data Collection

The data for this study were collected from Instagram, focusing specifically on public posts related to CRC published from March 1 to 31, 2024, which is Colorectal Cancer Awareness Month. These dates were chosen to capture the peak of awareness activities and messages. Using Instagram's search functionality, the top 150 posts containing hashtags related to CRC were identified and collected. The hashtags used for this search included #colorectalcancer, #colorectalcancerawareness, #colorectalcancerawarenessmonth, and combinations of these hashtags. To ensure comprehensive coverage, a supplementary search was conducted using additional terms such as "CRC," "Colon cancer," and "Colorectal Cancer." This supplementary search confirmed that the top posts identified through these terms consistently included 1 or more of the primary hashtags, indicating that the initial hashtag-based approach effectively captured relevant posts. This process minimized the risk of excluding significant content while aligning with the study's focus on thematic content and engagement during Colorectal Cancer Awareness Month. The number of likes and comments on the posts served as a gauge of their popularity, ensuring a representative sample of both highly engaged content and content with little engagement.

Given that institutional accounts often repost the same content multiple times during awareness campaigns, we implemented a systematic approach to handle duplicate posts. If a post was identical in both text and hashtags and originated from the same account, only the first instance was included in the dataset to prevent overrepresentation. However, if a repost contained minor variations in wording or hashtag use, it was retained to capture potential differences in engagement strategies. Engagement metrics (likes and comments) were analyzed separately for each retained post, rather than aggregating them across duplicates, as each post had the potential to generate unique interactions. This approach ensured that our dataset reflected a diverse range of content while minimizing redundancy and maintaining the integrity of engagement analysis.

Coding Process

This study employed a deductive content analysis approach to analyze both the text and images of Instagram posts related to CRC awareness. The thematic categories used for coding were predefined based on existing literature and theoretical frameworks in health communication [21-23]. The deductive approach was intentionally chosen to assess the prevalence and resonance of these theoretically supported themes, as reflected by engagement metrics (eg, likes, comments). This structured framework allowed for a systematic evaluation of whether established health communication messages were effectively disseminated through Instagram posts during Colorectal Cancer Awareness Month.

The coding framework was developed based on a review of existing literature on health communication and social media and common themes in CRC awareness [11-15]. Screening and early detection posts emphasized the importance of regular

XSL•FO
RenderX

screenings and early detection methods. Symptom posts provided information on the symptoms of CRC.

General awareness posts were aimed at raising general awareness about CRC through encouraging others to post about it or stating general statistics. Risk factor posts discussed factors that increase the risk of developing CRC and, conversely, preventative lifestyle measures one can take to reduce their risk. Individuals' experience posts shared personal stories and experiences of individuals affected by CRC. Because racial and ethnic minoritized groups and women face significant barriers to screening and care due to socioeconomic issues, limited health care access, and cultural or systemic biases, we examined posts on racial and ethnic minoritized communities and women to uncover messaging gaps and improve future awareness efforts' reach and impact [1,3]. Therefore, the representation of racial and ethnic minoritized community posts highlighted the vulnerability of racial and ethnic minoritized communities to CRC, and the representation of women posts focused on how CRC affects women and the risks it poses. Irrelevant posts were not related to CRC or awareness activities. Finally, other posts contained relevant information about CRC but did not fit into the predefined categories.

Captions and images contained in the Instagram posts were coded to align with the predefined thematic categories. Each post was examined holistically, with its image and caption assessed collectively to determine the dominant theme. Images were categorized based on their primary visual content and captions were analyzed alongside images to provide context and ensure consistency in thematic classification. If the image and caption conveyed different themes, coders determined which element carried the primary message, prioritizing captions for text-heavy posts and images for visually dominant content. In cases where thematic classification was unclear, discrepancies were resolved through discussion between researchers to ensure consistency and accuracy in coding.

To ensure reliability and minimize bias, a systematic coding process was implemented. One researcher (AS) initially coded all posts in the dataset, applying the predefined thematic categories based on existing literature and the objectives of the study. To validate the coding and minimize potential bias, a second researcher (JPS) independently coded a random sample of 10% of the posts. Discrepancies in coding between the 2 researchers were resolved through discussion and consensus, refining the thematic framework as needed. The level of agreement between the 2 researchers for the text and image content of the posts was measured using the κ reliability statistic, which resulted in a high agreement rate of 93%. This high level of interrater reliability confirms the consistency and accuracy of the thematic classification.

Analysis of Content and Engagement

Hashtags and post content were analyzed separately as they serve distinct functions in social media communication. Hashtag analysis was conducted to understand how users label and promote their posts within the broader discourse of CRC awareness, which helped identify trending topics, user engagement strategies, and the reach of campaign-related messaging. In contrast, analyzing the textual content (captions) and images within posts to determine the thematic focus of the message itself allowed us to assess whether the themes promoted through hashtags aligned with the actual messages in the post captions and images.

Our analysis began by describing the distribution of hashtags used for our search during Colorectal Cancer Awareness Month and the engagement of Instagram posts in terms of posts, likes, comments, and hashtag counts. Then, we described the content of the Instagram posts by hashtag, thematic category, and user profile (individual versus organizational account). This study adheres to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for reporting observational studies.

Ethical Considerations

The University of Texas Southwestern Medical Center Institutional Review Board determined that the study does not meet the definition of human subjects research and therefore does not require institutional review board approval or oversight (Y1-23-0390). As the data used are publicly available on Instagram, no anonymization or deidentification was required.

Results

This study analyzed a total of 150 Instagram posts related to CRC during Colorectal Cancer Awareness Month. These posts were categorized based on specific hashtags and the captions and image content of each post were classified into various thematic categories to understand the messaging and engagement trends.

Table 1 presents the distribution and engagement metrics of the posts across different hashtags. Posts tagged with #colorectalcancer received the highest engagement among the posts that were singularly tagged, amassing 9751 likes and 560 comments from 16 posts. This indicates a substantial interaction level, suggesting that this hashtag effectively attracts attention and engagement from users. The hashtag #colorectalcancerawarenessmonth was the most frequently used, appearing in 45 posts. These posts collectively garnered 6065 likes and 188 comments, reflecting their widespread use and moderate engagement. #colorectalcancerawareness was utilized in 31 posts, which received 3712 likes and 66 comments. This shows a lower engagement compared to the other hashtags, despite its considerable use. Posts with multiple tags were the most common, with 58 instances. These posts achieved the highest overall engagement, accumulating 17,729 likes and 120 comments. The use of multiple hashtags likely expanded the reach and visibility of the posts.

XSL•FO
RenderX

Table . Hashtag distribution and engagement of 150 Instagram posts during Colorectal Cancer Awareness Month, March 1-31, 2024.

Hashtag	Posts, n	Likes, n	Comments, n	Hashtag count, n
#colorectalcancer	16	9751	560	107
#colorectalcancerawareness-month	45	6065	188	132
#colorectalcancerawareness	31	3712	66	100
Multiple tags	58	17,729	120	414

Table 2 classifies the content of the Instagram posts by hashtag and thematic category. Screening and early detection was the most prevalent theme, featuring in 56 posts. This underscores a strong focus on promoting regular screenings and the importance of early detection in preventing the severity and mortality associated with CRC. General awareness was the theme in 29 posts, which highlighted efforts to spread general knowledge about CRC, including statistics and awareness activities. Risk factors were discussed in 18 posts, educating users about the various factors that increase the risk of developing CRC. Posts about the symptoms of CRC appeared in 12 instances, providing crucial information on recognizing the signs of the disease. Individual's experiences were shared in 16 posts, offering personal narratives and testimonies from individuals affected by CRC, which can be powerful in humanizing the disease and encouraging others to take preventive actions. The representation of racial and ethnic minoritized communities and women was notably limited, with only 3 and 2 posts, respectively. This indicates a gap in addressing the specific risks and experiences of these groups, suggesting a need for more inclusive messaging. The other category included 14 posts that were relevant but did not fit into the predefined categories, reflecting a diversity of information and perspectives related to CRC. Only 1 post was classified as irrelevant, demonstrating that most posts were pertinent to the topic.

Table . Content classification of 150 Instagram posts by hashtag during Colorectal Cancer Awareness Month, March 1-31, 2024.

Hashtag	Screening or early detection (n=56), n	Symptoms (n=12), n	General awareness (n=29), n	Risk factors (n=18), n	Individual's experience (n=16), n	Representation of racial and ethnic minority communities (n=3), n	Representation of women (n=2), n	Other (n=14), n	Irrelevant (n=1), n
#colorectalcancer (n=16)	5	1	1	1	4	0	1	2	1
#colorectalcancerawarenessmonth (n=45)	21	3	4	5	5	1	1	5	0
#colorectalcancerawareness (n=31)	13	3	6	5	2	1	0	1	0
Multiple tags (n=59)	17	5	18	7	5	1	0	6	0

Table 3 categorizes the 150 Instagram posts by user profile type—individual users (denoted as "Person") and organizational accounts (denoted as "Organization"). The distribution of these categories was analyzed across the 2 user profile types. Organizational accounts were more active overall, contributing 82% (n=123) of the total posts, compared to 18% (n=27) from individual users. Screening and early detection emerged as the predominant category overall, with 37.3% (56/150) of all posts falling into this category. This category was significantly more common among organizational accounts, constituting 41.4% (51/123) of their posts, compared to 19% (5/27) of posts from individual users. Posts related to symptoms accounted for 8% (12/150), with a nearly equal distribution between individual and organizational accounts. The general awareness category comprised 19% (n=29) of the total posts, with individual users and organizational accounts contributing 19% (5/27) and 19.5% (24/123) of their posts, respectively. Risk factors were addressed in 12% (18/150) of the posts. Both individual users and organizational accounts showed interest in this category, with 19% (5/27) of individual users' posts and 10.5% (13/123) of organizational accounts' posts mentioning risks. Posts discussing individuals' experiences accounted for 10.6% (16/150) of the total, with greater representation from individual users at 19% (5/27) compared to organizational accounts. Posts representing racial and ethnic minority communities were notably sparse, comprising only 2% (3/150) of the total posts. Individual users contributed 7% (2/27) and organizational accounts just 0.8% (1/123). Similarly, posts focusing on the representation of women were minimal, representing 1.3% (2/150) of the total posts, with no contributions from individual users and 1.6% (2/123) from organizational accounts. The other category included 9.3% (14/150) of the total posts, with individual users

XSL•FO
RenderX

contributing 7% (2/27) and organizational accounts contributing 9.7% (12/123). Lastly, irrelevant posts constituted 0.6% (1/150) of the total, all from individual users at 4% (1/27), with no contributions from organizational accounts.

Table . Content classification of 150 Instagram posts by user profile during Colorectal Cancer Awareness Month, March 1-31, 2024.

	Person (n=27), n (%)	Organization (n=123), n (%)	Total (N=150), n (%)
Screening and early detection	5 (19)	51 (41.4)	56 (37.3)
Symptoms	2 (7)	10 (8.1)	12 (8)
General awareness	5 (19)	24 (19.5)	29 (19.3)
Risk factors	5 (19)	13 (10.5)	18 (12)
Individual's experience	5 (19)	10 (8.1)	15 (10)
Representation of racial and ethnic minoritized communities	2 (7)	1 (0.8)	3 (2)
Representation of women	0 (0)	2 (1.6)	2 (1.3)
Other	2 (7)	12 (9.7)	14 (9.3)
Irrelevant	1 (4)	0 (0)	1 (0.6)

Discussion

Principal Findings

The goal of our research study was to explore the thematic content and engagement patterns of Instagram posts shared during Colorectal Cancer Awareness Month, with a focus on understanding how key messages about CRC prevention and awareness are conveyed and received. We analyzed the top 150 Instagram posts, and this methodological approach ensured a robust analysis of the content and engagement of CRC-related posts on Instagram, providing insights into the effectiveness of social media as a tool for health awareness campaigns. Our findings highlighted the diverse approaches used to raise awareness about CRC on Instagram.

A key finding was the dominance of themes related to screening and early detection, which reflects the goal of Colorectal Cancer Awareness Month to raise awareness about CRC and motivate individuals to participate in screening, and also suggests that public health campaigns are effectively leveraging Instagram to promote these critical aspects of CRC prevention [2,3,15]. The significant engagement metrics, particularly the high number of likes and comments associated with the hashtag #colorectalcancer, suggest that this specific tag resonates well with the Instagram audience. Additionally, the use of multiple hashtags in posts appeared to enhance engagement, suggesting that combining hashtags may increase the visibility and interaction of the posts.

While the themes of general awareness and risk factors were well represented, the relatively limited number of posts addressing symptoms and individual experiences points to potential gaps in the current messaging strategy. Personal stories can be particularly powerful in motivating behavior change and should be utilized more extensively [22-24]. Furthermore, the underrepresentation of posts targeting racial and ethnic minoritized communities and women highlights a critical gap in the current approach. Given the disparities in CRC incidence and outcomes among different demographic groups, it is essential to tailor messages to address these disparities and ensure more inclusive outreach [1,3].

This study supports the existing literature on the effectiveness of visual content in attracting user interaction [10-15]. However, discrepancies may exist regarding the types of messages that are most effective. For instance, while some studies suggest that statistical information is persuasive, the current findings highlight the potential impact of personal narratives, which may be underutilized in existing campaigns [24].

Implications for Public Health Practice

The results of this study have several implications for public health practice. First, there is a need for more tailored campaigns that specifically address the risks and experiences of marginalized groups and women [25]. Developing culturally sensitive and inclusive content can help bridge the current gaps in outreach [26]. Second, incorporating more individual experiences and testimonies in posts can create more relatable and impactful messages, potentially increasing awareness and encouraging preventive behaviors [27]. Lastly, public health campaigns should consider using multiple and specific hashtags to maximize reach and engagement [28].

Limitations and Future Research

This study has several strengths, including a comprehensive analysis of the top posts on a significant awareness month and high interrater reliability in content classification, which ensures consistency in the findings. However, there are also limitations. The study is limited to 1 social media platform (Instagram) and 1 specific month, which may not capture the full spectrum of awareness activities and engagement throughout the year on other social media platforms. Instagram was chosen due to its visual-centric nature, which is particularly well suited for health awareness campaigns, and its algorithmic features that facilitate content discovery and engagement through hashtags. Additionally, the overlap between Instagram and Facebook, both owned by Meta, often results in similar campaign content across the 2 platforms, providing a representative snapshot of Meta's ecosystem. Moreover, the cross-sectional nature of the

XSL•FO
RenderX

study limits the ability to assess changes over time or the long-term impact of these posts.

We did not assess the credibility and quality of the Instagram posts, which was a deliberate decision, as the study aimed to explore thematic content and user engagement rather than verify the accuracy or reliability of the posts. The credibility of posts is important for assessing the public health value of messages [29]. Future research should consider assessing the quality and credibility of health-related social media content to better understand its reliability and influence on public health outcomes.

Future research should consider longitudinal studies to evaluate the impact of social media campaigns over time. Comparative studies across different social media platforms would also be beneficial to identify the most effective channels for health promotion. Furthermore, research focusing on the development and testing of tailored messages for underrepresented groups is necessary to enhance the inclusivity and effectiveness of these campaigns. Finally, while this study focused on predefined themes, we recognize the value of inductive approaches, such as topic modeling, for identifying emergent themes or uncovering patterns directly from the data. Future research could complement our findings by employing such methods to explore new dimensions of social media health communication.

Conclusion

In conclusion, this study highlights the potential of Instagram as a platform for promoting CRC awareness, particularly for screening and early detection. While certain themes effectively engage users and align with the goals of awareness campaigns, critical gaps remain in inclusivity and representation, particularly for marginalized communities. To maximize the impact of future campaigns, messaging strategies should be refined and incorporate a broader range of themes, ensuring that social media platforms like Instagram are leveraged more effectively to raise awareness, reduce disparities, and promote preventive health behaviors.

Acknowledgments

The American Cancer Society provided funding for JPS' research (award DBG-23-1155771-01-HOPS). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funder. The funder had no role in the study design, data analysis, decision to publish, or preparation of the manuscript.

Data Availability

The datasets analyzed during this study are publicly available on Instagram.

Authors' Contributions

All authors contributed to the study's conception and design. Data collection and analysis were performed by AS. JPS wrote the first draft of the manuscript. All authors contributed to subsequent drafts of the manuscript. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Tan JY, Yeo YH, Ng WL, Fong ZV, Brady JT. How have US colorectal cancer mortality trends changed in the past 20 years? Int J Cancer 2024 Aug 1;155(3):493-500. [doi: 10.1002/ijc.34926] [Medline: 38525799]
2. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74(3):229-263. [doi: 10.3322/caac.21834] [Medline: 38572751]
3. Cronin KA, Scott S, Firth AU, et al. Annual report to the nation on the status of cancer, part 1: national cancer statistics. Cancer 2022 Dec 15;128(24):4251-4284. [doi: 10.1002/cncr.34479] [Medline: 36301149]
4. Doubeni CA, Fedewa SA, Levin TR, et al. Modifiable failures in the colorectal cancer screening process and their association with risk of death. Gastroenterology 2019 Jan;156(1):63-74. [doi: 10.1053/j.gastro.2018.09.040] [Medline: 30268788]
5. Smith RA, Fedewa S, Siegel R. Early colorectal cancer detection-current and evolving challenges in evidence, guidelines, policy, and practices. Adv Cancer Res 2021;151:69-107. [doi: 10.1016/bs.acr.2021.03.005] [Medline: 34148621]
6. National Colorectal Cancer Awareness Month. Colorectal Cancer Alliance. URL: https://colorectalcancer.org/get-involved/advocate/colorectal-cancer-awareness-month [accessed 2025-02-12]
7. Pantel HJ, Kleiman DA, Kuhnen AH, Marcello PW, Stafford C, Ricciardi R. Has National Colorectal Cancer Awareness Month increased endoscopy screening rates and public interest in colorectal cancer? Surg Endosc 2021 Jan;35(1):398-405. [doi: 10.1007/s00464-020-07413-x] [Medline: 32016518]
8. Smith K. 50 incredible Instagram statistics you need to know. Brandwatch. 2019. URL: https://www.brandwatch.com/blog/instagram-stats/ [accessed 2025-02-12]

XSL•FO
RenderX

9. Dixon SJ. Number of monthly active Instagram users from January 2013 to December 2021. Statista. 2024. URL: https://www.statista.com/statistics/253577/number-of-monthly-active-instagram-users/ [accessed 2025-02-12]
10. Gatewood J, Monks SL, Singletary CR, Vidrascu E, Moore JB. Social media in public health: strategies to distill, package, and disseminate public health research. J Public Health Manag Pract 2020;26(5):489-492. [doi: 10.1097/PHH.0000000000001096] [Medline: 32732723]
11. Vicente EP, De Faria SEE, Almeida ABL, et al. Cervical cancer prevention on Instagram: content and social interaction analysis of Brazilian accounts. Asian Pac J Cancer Prev 2022 Sep 1;23(9):3043-3049. [doi: 10.31557/APJCP.2022.23.9.3043] [Medline: 36172667]
12. Basch CH, MacLean SA. Breast cancer on Instagram: a descriptive study. Int J Prev Med 2019;10(1):166. [doi: 10.4103/ijpvm.IJPVM_36_19] [Medline: 32133084]
13. Gomaa B, Houghton RF, Crocker N, Walsh-Buhi ER. Skin cancer narratives on Instagram: content analysis. JMIR Infodemiology 2022;2(1):e34940. [doi: 10.2196/34940] [Medline: 37113805]
14. Parker C, Zomer E, Liew D, Ayton D. Characterising experiences with acute myeloid leukaemia using an Instagram content analysis. PLoS ONE 2021;16(5):e0250641. [doi: 10.1371/journal.pone.0250641] [Medline: 33939746]
15. Basch CH, MacLean SA. Colorectal cancer on Instagram: a content analysis. J Consum Health Internet 2019 Oct 2;23(4):378-383. [doi: 10.1080/15398285.2019.1684149]
16. Pellino G, Simillis C, Qiu S, et al. Social media and colorectal cancer: a systematic review of available resources. PLoS ONE 2017;12(8):e0183031. [doi: 10.1371/journal.pone.0183031] [Medline: 28832603]
17. Plackett R, Kaushal A, Kassianos AP, et al. Use of social media to promote cancer screening and early diagnosis: scoping review. J Med Internet Res 2020 Nov 9;22(11):e21582. [doi: 10.2196/21582] [Medline: 33164907]
18. Xu S, Markson C, Costello KL, Xing CY, Demissie K, Llanos AA. Leveraging social media to promote public health knowledge: example of cancer awareness via Twitter. JMIR Public Health Surveill 2016;2(1):e17. [doi: 10.2196/publichealth.5205] [Medline: 27227152]
19. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85. [doi: 10.2196/jmir.1933] [Medline: 23615206]
20. Laranjo L, Arguel A, Neves AL, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256. [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
21. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
22. Wu J, Benjamin EJ, Ross JC, Fetterman JL, Hong T. Health messaging strategies for vaping prevention and cessation among youth and young adults: a systematic review. Health Commun 2024 May 14:1-19. [doi: 10.1080/10410236.2024.2352284] [Medline: 38742648]
23. Zhao YC, Zhao M, Song S. Online health information seeking among patients with chronic conditions: integrating the health belief model and social support theory. J Med Internet Res 2022 Nov 2;24(11):e42447. [doi: 10.2196/42447] [Medline: 36322124]
24. Hinyard LJ, Kreuter MW. Using narrative communication as a tool for health behavior change: a conceptual, theoretical, and empirical overview. Health Educ Behav 2007 Oct;34(5):777-792. [doi: 10.1177/1090198106291963] [Medline: 17200094]
25. Agarwal V. Editorial: communicating for social justice in health contexts: creating opportunities for inclusivity among marginalized groups. Front Commun 2022 Sep 30;7:1008533. [doi: 10.3389/fcomm.2022.1008533]
26. Häfliger C, Diviani N, Rubinelli S. Communication inequalities and health disparities among vulnerable groups during the COVID-19 pandemic - a scoping review of qualitative and quantitative evidence. BMC Public Health 2023 Mar 6;23(1):428. [doi: 10.1186/s12889-023-15295-6] [Medline: 36879229]
27. de Graaf A, Sanders J, Hoeken H. Characteristics of narrative interventions and health effects: a review of the content, form, and context of narratives in health-related narrative persuasion research. RCR 2016;4:88-131. [doi: 10.12840/issn.2255-4165.2016.04.01.011]
28. Ghahramani A, Prokofieva M, de Courten MP. Health promotion campaigns using social media: association rules mining and co-occurrence network analysis of Twitter hashtags. BMC Public Health 2025 Jan 7;25(1):67. [doi: 10.1186/s12889-024-21255-5] [Medline: 39773431]
29. Etta RE, Babatunde AO, Okunlola PO, et al. The assessment of TikTok as a source of quality health information on human papillomavirus: a content analysis. Cureus 2024 Dec;16(12):e75419. [doi: 10.7759/cureus.75419] [Medline: 39791048]

Abbreviations

CRC: colorectal cancer
HPV: human papillomavirus
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

XSL•FO
RenderX

Edited by A Mavragani; submitted 17.06.24; peer-reviewed by D Surian, Y Sarikhani; revised version received 22.01.25; accepted 29.01.25; published 19.02.25.

Please cite as:
Srivastava A, Stimpson JP
Instagram Posts Promoting Colorectal Cancer Awareness: Content Analysis of Themes and Engagement During Colorectal Cancer Awareness Month
JMIR Form Res 2025;9:e63344
URL: https://formative.jmir.org/2025/1/e63344
doi:10.2196/63344

© Aditi Srivastava, Jim P Stimpson. Originally published in JMIR Formative Research (https://formative.jmir.org), 19.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Quality Assessment of Medical Institutions' Websites Regarding Prescription Drug Misuse of Glucagon-Like Peptide-1 Receptor Agonists by Off-Label Use for Weight Loss: Website Evaluation Study

Rie Oyama[1], MD; Tsuyoshi Okuhara[1], MPH, PhD; Emi Furukawa[2], MD, PhD; Hiroko Okada[1], MPH, PhD; Takahiro Kiuchi[1], MD, PhD

[1]Department of Health Communication, School of Public Health, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
[2]University Hospital Medical Information Network, The University of Tokyo Hospital, Tokyo, Japan

Corresponding Author:
Rie Oyama, MD
Department of Health Communication
School of Public Health
Graduate School of Medicine, The University of Tokyo
7-3-1 Hongo
Bunkyo-ku
Tokyo, 113-8655
Japan
Phone: 81 3 5800 6549
Fax: 81 3 5689 0726
Email: rooyama-oss@umin.ac.jp

Abstract

Background: Misuse of glucagon-like peptide-1 receptor agonists (GLP-1RAs) has emerged globally as individuals increasingly use these drugs for weight loss because of unrealistic and attractive body images advertised and shared on the internet.

Objective: This study assesses the quality of information and compliance with Japan's medical advertising guidelines on the websites of medical institutions that prescribe GLP-1RAs off-label for weight loss.

Methods: Websites were identified by searching Google and Yahoo! by using keywords related to GLP-1RAs and weight loss in August 2024. The quality of information on these websites was assessed using the DISCERN instrument. To comply with Japan's medical advertising guidelines, we evaluated whether the 5 mandatory items for advertisements of self-paid medical treatments involving the off-label use of drugs were stated and whether there were any exaggerated claims. The content of the exaggerated advertisements was categorized into themes.

Results: Of the 87 websites included, only 1 website stated all 5 mandatory items. Websites listing "ineligible for the relief system for sufferers from adverse drug reactions" had the lowest percentage at 9% (8/87), while 83% (72/87) of the websites listed exaggerated advertisements. Approximately 69% (60/87) of the websites suggested that no exercise or dietary therapy was required, 24% (21/87) suggested that using GLP-1RAs is a natural and healthy method, and 31% (27/87) of the websites provided the author's personal opinions on the risks of using GLP-1RAs. The mean total DISCERN score for all 87 websites was 32.6 (SD 5.5), indicating low quality. Only 1 website achieved a good rating, and 9 websites were rated as fair. The majority of the websites were rated as poor (72 websites) or very poor (5 websites).

Conclusions: We found that the quality of information provided by the websites of medical institutions prescribing GLP-1RAs off-label for weight loss was very low and that many websites violated Japan's medical advertising guidelines. The prevalence of exaggerated advertisements, which may lead consumers to believe that they can lose weight without dietary or exercise therapy, suggests the risk of GLP-1RA misuse among consumers. Public institutions and health care providers should monitor and regulate advertisements that violate guidelines and provide accurate information regarding GLP-1RAs, obesity, and weight loss.

(JMIR Form Res 2025;9:e68792) doi:10.2196/68792

XSL•FO
RenderX

KEYWORDS

prescription drug misuse; GLP-1 receptor agonists; off-label use; weight loss; information quality; DISCERN; web-based information; information provision; misinformation; advertising guidelines; exaggerated advertisements

Introduction

Prescription drug misuse is an emerging global public health issue [1-3]. According to a previous study [4], misuse is defined as the therapeutic use of legal drugs in a manner other than as prescribed, and abuse is defined as the use of illegal drugs for nontherapeutic purposes such as psychological pleasure and weight loss. The misuse and abuse of drugs can lead to health problems. In the United States and European countries, misuse of opioid analgesics, benzodiazepines, over-the-counter cough suppressants containing dextromethorphan and codeine, and psychostimulants have been observed [1-3,5]. Further, in Japan, misuse of over-the-counter cough suppressants, pain relievers, antianxiety medications, and sedative drugs has been observed [6,7]. However, misuse of glucagon-like peptide-1 receptor agonists (GLP-1RAs) has become a problem in recent years [8-10].

GLP-1RAs are the most effective drugs for obesity treatment [11]. Semaglutide (Wegovy), a type of GLP-1RA, was approved by the US Food and Drug Administration (FDA) as an obesity treatment drug in 2021 [12]. Since then, its market share has rapidly expanded [13]. Conversely, multiple news media outlets have reported concerns about the misuse of GLP-1RAs [14,15]. Due to unrealistic and attractive body images that are advertised by celebrities on social media, people without obesity allegedly obtain and use GLP-1RAs illicitly for weight loss purposes [16]. Adverse events of GLP-1RA misuse have also been reported [8,9]. A previous study showed that the use of semaglutide in patients with overweight/obesity may be associated with the incidence of nonarteritic anterior ischemic optic neuropathy, a severe ophthalmologic condition that can lead to blindness [17]. Another previous study showed that GLP-1RA users for weight management had an increased risk of serious gastrointestinal disorders compared to bupropion-naltrexone users, with adjusted hazard ratios for pancreatitis, bowel obstruction, and gastroparesis reported as 9.09, 4.22, and 3.67, respectively [18]. However, the use of GLP-1RAs for weight management is a relatively new therapeutic approach, and the long-term side effects associated with their use remain unknown. Additionally, the effectiveness and potential adverse effects of using GLP-1RAs for further weight loss in individuals without obesity also remain unknown. This increased demand has caused a global shortage of GLP-1RAs [16]. The European Medicines Agency has announced recommendations by manufacturers, health care professionals, and citizens for the proper use of GLP-1RAs [19].

In Japan, Wegovy was approved for the treatment of obesity on March 27, 2023. According to the Optimal Use Promotion Guideline by the Japan's Ministry of Health, Labor, and Welfare, patients eligible for obesity treatment are "those with BMI ≥35 kg/m² or those with BMI ≥27 kg/m² who also have 2 or more health conditions related to obesity and who do not achieve sufficient results from diet and exercise therapy" [20]. The guidelines also strictly prohibit the use of GLP-1RAs for weight loss purposes without obesity as well as institutional criteria for prescribing GLP-1RAs under insurance-covered medical care [20,21]. Japan's Ministry of Health, Labor, and Welfare released a "Request for cooperation due to tight inventory of GLP-1 receptor agonist" [22], and the Japan Diabetes Society [23] and Japan Society for the Study of Obesity [24] have made recommendations regarding the off-label use of GLP-1RAs.

However, several private clinics in Japan prescribe GLP-1RAs for off-label weight loss. Japan's National Consumer Affairs Center and Consumer Affairs Agency warns citizens against the use of GLP-1RAs for weight loss purposes because of the possibility of medical institutions prescribing GLP-1RAs through advertisements that are prohibited by guidelines, inadequate physician consultations and explanations, and inadequate systems to manage adverse events [25,26]. Websites of medical institutions in Japan are regulated by the Medical Care Act [27] and the Ministry of Health, Labor and Welfare's "Regulations on Advertisements of Hospitals, etc under the Medical Care Act" (Japan's medical advertising guidelines) [28], which prohibit false or exaggerated advertisements. The reason for this is "because false advertisements may cause patients to lose the opportunity to receive appropriate medical consultation or receive inappropriate medical treatments by giving them information that is significantly different from the facts, etc"[28].

In previous studies on health information quality, these issues were assessed as part of the reliability of health information. The DISCERN instrument developed by Charnock et al [29] has been used in many studies to assess the publication reliability and the quality of the written health information (Multimedia Appendix 1). The validity and reliability of the DISCERN instrument were ensured, and in addition to the original English language version, Chinese [30], Spanish [31], Brazilian Portuguese [32], and Japanese [33] versions have been developed [29]. Previous studies using the DISCERN instrument have examined the quality of information on websites regarding posttraumatic stress disorder [34], breast cancer [35], nasopharyngeal carcinoma [36], and cosmetic injectability [37]. These studies showed that the quality of individual websites varied considerably: the quality of information was higher for academic and government institutions that operate websites and lower for commercial institutions [34-37]. A systematic review showed that in research using the DISCERN instrument, none found the mean DISCERN score of all websites to be excellent, and the information quality of the majority of websites ranged from fair to very good [38].

In recent years, most people have searched the internet for health information [39]. The internet usage rate for individuals in Japan in 2022 was 84.9% and exceeded 90% when limited to the 10-50 years age range [40]. For cosmetic treatments, 95% of the patients search for information on the internet before entering a doctor's office [41]. Hence, if medical institutions' websites

XSL•FO
RenderX

on GLP-1RAs contain misleading, exaggerated, or unreliable information, these websites should be immediately assessed and improved. However, to our knowledge, no study has investigated the quality of information that people without obesity obtain from websites when they consider obtaining GLP-1RAs for weight loss. This study aims to assess compliance with medical advertising guidelines and the quality of information on the websites of medical institutions that prescribe GLP-1RAs off-label for weight loss in Japan.

Methods

Data Collection

Search Terms

We used a Japanese-language search string input into Google Japan [42] and Yahoo Japan [43], which has the largest search engine market shares in Japan, in August 2024 [44]. The GLP-1RAs approved by the FDA for obesity treatment are liraglutide, semaglutide, and tirzepatide that are sold in Japan under the product names Rybelsus, Ozempic, Victoza, Saxenda, and Mounjaro. The search keywords comprised 6 in total (Rybelsus diet, Ozempic diet, Victoza diet, Saxenda diet, Mounjaro diet, and GLP-1 diet), which included each product name combined with diet (used specifically to mean weight loss in Japan), as well as the comprehensive term GLP-1 diet. Wegovy was excluded from the search terms because it was the newest drug and was not available in the general market in Japan. Saxenda is approved in Japan for the treatment of diabetes under the name Victoza, but private clinics that use the drug off-label advertise it under the name Saxenda; so, it was included in the search terms.

All words were entered in the search window, one keyword at a time, and 20 websites were reviewed per search engine for each term. This is because a previous study showed that 97.2% of the users only view the first 1-10 websites of search results, and 2% click on links on the second page of the search results and beyond [45]. Google Chrome's incognito browser that would have all history deleted was used, and personal accounts were not used.

Inclusion and Exclusion Criteria

The inclusion criteria were as follows: websites written in Japanese; websites created by medical institutions (with medical department labels); websites clearly stating that GLP-1RAs were being sold; and the medical institutions operating the websites were not accredited as facilities capable of treating obesity using GLP-1RAs by the Japanese Circulation Society, Japan Diabetes Society, or Japan Endocrine Society. The exclusion criteria were as follows: duplicate websites, government or academic statements, PDF documents, news reports, affiliate advertisements, and websites that were unclear whether they prescribe GLP-1RAs. The first author (RO) reviewed all the websites in the search results and extracted eligible websites.

Data Extraction

RO recorded the names and URLs of all the websites, names of the medical institutions operating the websites, the medical departments they represented, and the types of GLP-1RAs they sold. If not listed, it was recorded as unknown. Medical institutions can represent multiple medical departments in Japan. Therefore, if several departments were listed, all were recorded. All data were recorded in a Microsoft Excel 2021 spreadsheet by RO.

Data Assessment

First, all identified websites were assessed by an internal medicine physician (RO) for compliance with the medical advertising guidelines in Japan and quality by using the DISCERN instrument. Second, the third author (EF, an internal physician) was trained through a 1-hour meeting and a 1-hour pilot assessment, and then EF independently assessed approximately 20% (20/87) of the total websites. These 20 websites were randomly selected using Microsoft Excel to generate random numbers. Finally, they assessed the agreement between the coders of the 20 websites. The results were recorded using a Microsoft Excel 2021 spreadsheet. More precise information is provided below.

Assessment of Compliance With Medical Advertising Guidelines

In the case of noninsured medical treatments involving off-label use of drugs, it is prohibited to advertise on websites without explicitly stating that "the drugs are not approved, the acquisition route, information on whether there are other domestically approved drugs with the same ingredients and performance, information on the safety of the products in other countries, and they are ineligible for the relief system for sufferers from adverse drug reactions," in Japan's medical advertising guidelines [28]. All the websites were assessed based on whether these 5 items were stated.

Japan's medical advertising guidelines also prohibit exaggerated advertisements [28]. The guidelines [28] define exaggerated advertisements as advertisements that are not necessarily false but that misleads citizens by misrepresenting facts such as about the size of the facility, staffing, or the contents of the medical care provided. Moreover, "misleading citizens" is defined as "it is sufficient to say, as a matter of common sense, that there is a difference between the 'impression' or 'expectation' that citizens perceive from the advertisements and the actual contents, and it is not necessary to prove that the advertisements mislead or to show that the advertisements have actually misled citizens" [27]. Therefore, since taking GLP-1RAs is considered an auxiliary therapy for exercise and dietary therapy in the treatment of obesity, statements such as "you can lose weight with GLP-1RAs alone," "you can lose weight even if you cannot exercise," or "you do not need dietary therapy" are considered exaggerated advertisements. The coders carefully read the guidelines and assessed each statement by using a binary (0/1) scale to indicate the presence or absence of exaggerated advertisements. RO then analyzed the contents of the exaggerated advertisements and categorized them into themes according to the specific examples in the guidelines [28].

Quality Assessment Using the DISCERN Instrument

The DISCERN instrument in the Japanese version [33] was used for the quality assessment of information. The DISCERN

XSL•FO
RenderX

instrument consists of 16 items, each of which is assessed on a 5-point Likert scale (1=not at all, 5=completely). Section 1 (questions 1-8) addresses website reliability, section 2 (questions 9-15) addresses the quality of information about treatment options, and section 3 (question 16) presents the overall assessment. The total score ranged from 16 to 80, with higher scores indicating higher quality information. Several previous studies ranked the quality of websites based on their total score as follows: excellent=80-63, good=62-51, fair=50-39, poor=38-27, very poor=26-16 [35,46].

Statistical Analysis

Data are presented as means for continuous variables and counts with frequency percentages for categorical variables. Cohen κ statistic was used to measure the intercoder reliability for compliance with medical advertising guidelines and DISCERN scores between the first and second coders. Linear regression models were used to investigate the association between the specialties of the clinic operating the website and the total DISCERN score. *P* values were 2-sided, and $P<.05$ was considered statistically significant. All analyses were conducted using the R version 4.3.1 (June 16, 2023; R Foundation for Statistical Computing).

Ethical Considerations

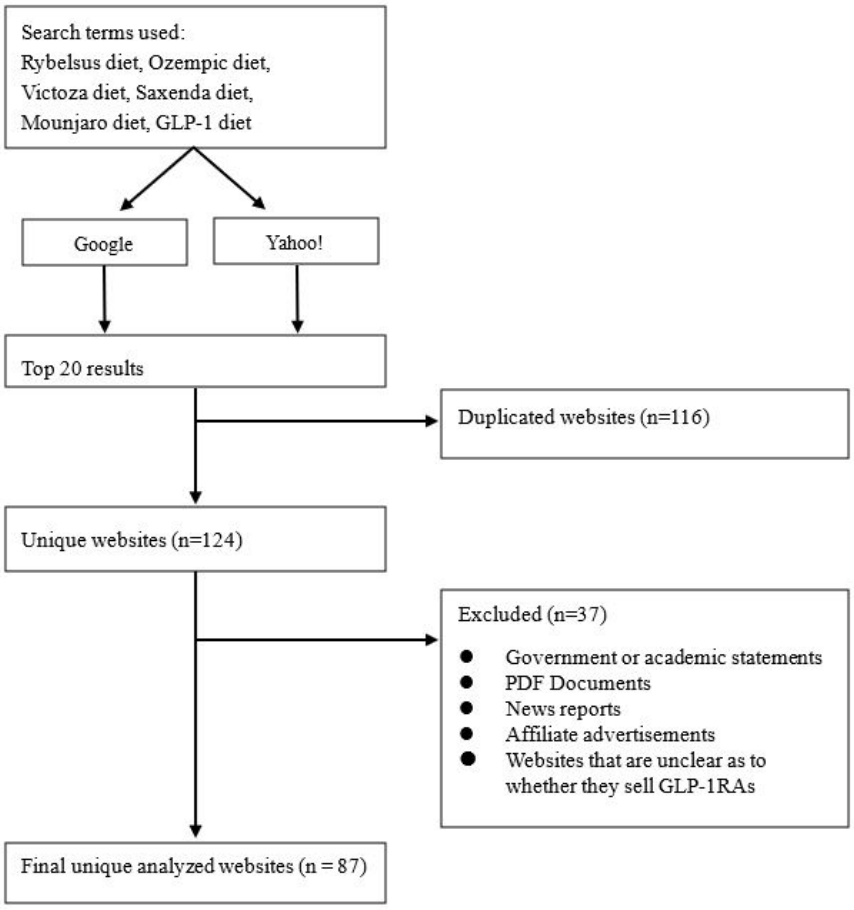
This study did not involve human or animal participants, and all data were collected from publicly available web-based sources. This study was granted an exemption from ethics approval by the ethics review committee of the Graduate School of Medicine, University of Tokyo.

Results

Characteristics of the Included Websites

In total, 240 websites were identified. After the removal of intrasearch term duplicates and those that met the exclusion criteria, 87 websites were used for the analysis (Figure 1). A total of 67 clinics operated the 87 websites. This was because 8 clinics had multiple different GLP-1RA diet-related articles. Table 1 shows that the most common departments, as far as could be determined from the websites, were dermatology (22/120, 18.3%), cosmetic dermatology (21/120, 17.5%), and internal medicine (19/120, 15.8%); 17 clinics were unaware of their departments. The percentages of advertisements with Ozempic, Rybelsus, Mounjaro, Victoza, and Saxenda were 18.5% (30/162), 29.1% (47/162), 12.3% (20/162),14.2% (23/162), and 25.9% (42/162), respectively. Approximately 18% (16/87) of the websites mentioned selling sodium-glucose cotransporter 2 inhibitor along with GLP-1RAs, and 12% (10/87) mentioned selling metformin.

Figure 1. Search process flowchart. GLP-1: glucagon-like peptide-1; GLP-1RA: glucagon-like peptide-1 receptor agonist.



XSL•FO
RenderX

Table 1. Characteristics of the websites regarding prescription drug misuse of glucagon-like peptide-1 receptor agonists by off-label use for weight loss.

	Values, n (%)
Specialty[a] (n=120[b])	
Dermatology	22 (18.3)
Cosmetic dermatology	21 (17.5)
Internal medicine	19 (15.8)
Cosmetic surgery	6 (5)
Urology	6 (5)
Allergology	5 (4.2)
Other	24 (20)
Unknown	17 (14.2)
Types of advertised drugs[c] (n=162[d])	
Ozempic	30 (18.5)
Rybelsus	47 (29)
Mounjaro	20 (12.3)
Victoza	23 (14.2)
Saxenda	42 (25.9)

[a]Medical departments represented by medical institutions.

[b]Total number of mentions of specialty.

[c]Types of glucagon-like peptide-1 receptor agonists that medical institutions advertised for sale.

[d]Total number of mentions of drugs.

Compliance With Medical Advertising Guidelines

The intercoder reliability was acceptable (average Cohen κ=0.667). Table 2 shows the compliance with medical advertising guidelines. Only one of the 87 websites stated the 5 required items. The items with the highest percentage of listing were "acquisition route" and "information on the safety of the drugs in other countries," both at 79% (69/87). The item with the lowest percentage of listing was "ineligible for the relief system for sufferers from adverse drug reactions" at 9% (8/87).

Table 2. Compliance with advertising guidelines of the websites regarding prescription drug misuse of glucagon-like peptide-1 receptor agonists by off-label use for weight loss (N=87).

	Values, n (%)
Content that should be included[a]	
The drug is unapproved	43 (49)
Acquisition route	69 (79)
Information on whether there are other domestically approved drugs with the same ingredients and performance	24 (28)
Information on the safety of the drug in other countries	69 (79)
Ineligible for relief system for sufferers from adverse drug reactions	8 (9)
Exaggerated advertisements[b]	72 (83)

[a]Explicit statements required under Japan's medical advertising guidelines for noninsured medical treatments involving off-label drug use.

[b]Advertisements containing expressions that could mislead citizens, as determined by coders under Japan's medical advertising guidelines.

Table 2 shows that 83% (72/87) of the websites listed exaggerated advertisements. Table 3 shows the content of the exaggerated advertisements: of the 87 websites, 69% (60/87) conveyed messages indicating no exercise and dietary therapy were required, 24% (21/87) conveyed messages indicating that using GLP-1RAs is a natural and healthy method, and 31% (27/87) provided the website author's subjective expression of the risks of using GLP-1RAs.

XSL•FO
RenderX

Table 3. Content of exaggerated advertising regarding prescription drug misuse of glucagon-like peptide-1 receptor agonists by off-label use for weight loss (N=87)[a].

Themes	Illustrative quotes	Values, n (%)
Induction by nonscientific information (no diet or exercise therapy required)	*…Exercise is also unnecessary. Exercise is less effective in losing weight.* *…Medical dieting is a treatment that suppresses appetite and promotes weight loss by taking a medicine called GLP-1 receptor agonist, which is used in the treatment of diabetes, without diet restrictions or hard exercise regimen.*	60 (69)
Induction by nonscientific information (natural and healthy methods)	*…By managing portion sizes, you can reduce calorie intake and engage in a natural and controlled dietary regimen.* *…To achieve weight loss in a healthy manner.* *…It also has fat-burning and pancreas-protecting effects, making it a healthy way to lose weight.*	21 (24)
Subjective expressions of risks	*…Since the delayed gastric emptying effect is not long-lasting, there is little need to consider it dangerous.* *…I, as the director, have also used it and had no side effects other than slight heartburn, which may have caused mild anorexia and slight weight loss.*	27 (31)
Statements that you can always get the drug only through web-based medical consultation	*…Everything from making an appointment and consultation to prescribing a drug is done online, so there is no need to go out.* *…You can see us from anywhere in the country via our web-based clinic. [...] There is no consultation fee; GLP-1 injections will be mailed to you.*	20 (23)

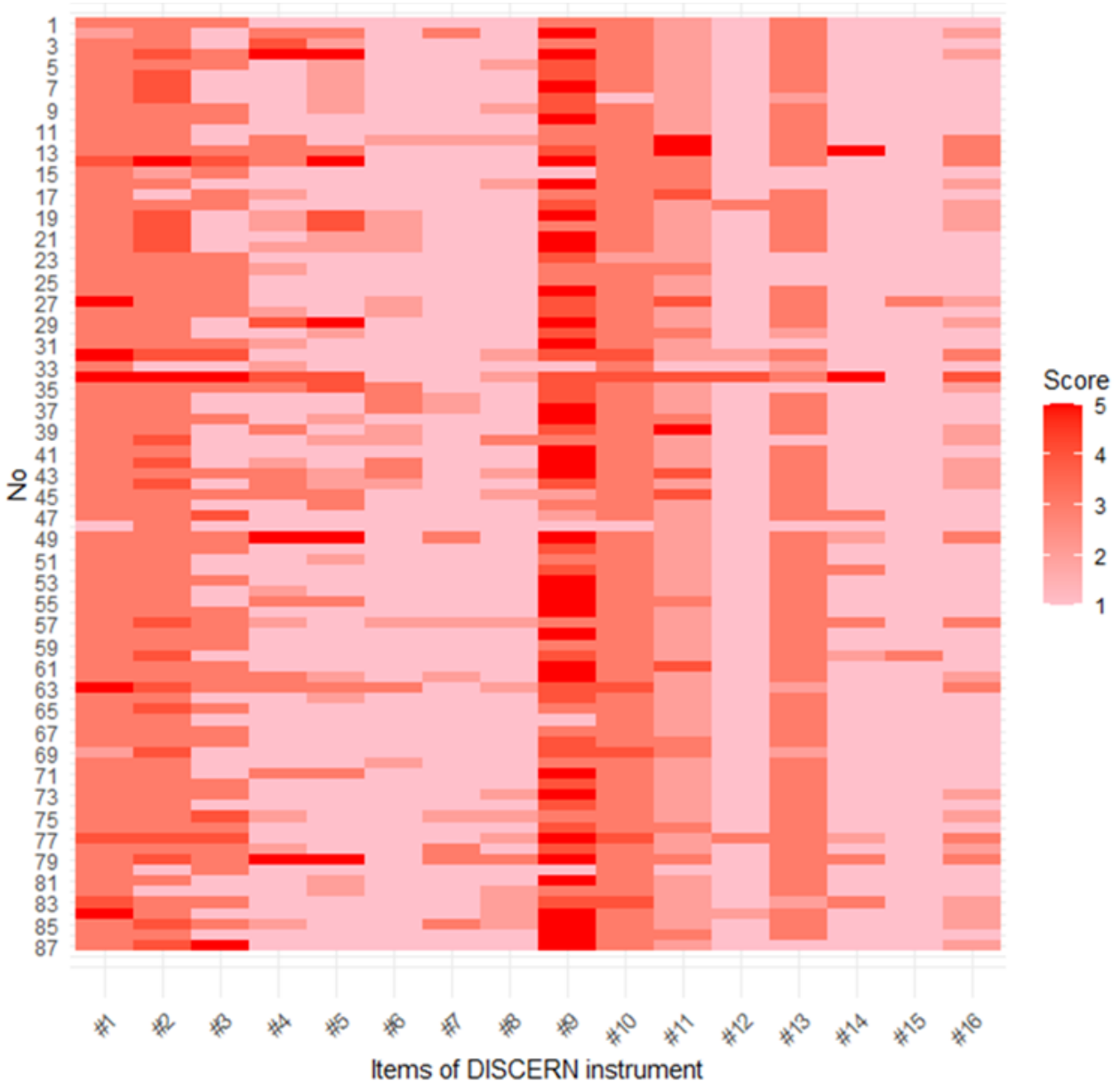
[a]Advertisements containing expressions that could mislead citizens, as determined by coders under Japan's medical advertising guidelines.

Quality Assessment Using DISCERN Instrument

Intercoder reliability was acceptable (average Cohen κ=0.663). Figure 2 shows the DISCERN scores for all 87 websites across each item. Table 4 presents the mean score for each of the 16 DISCERN instrument criteria and the total DISCERN score. The mean total DISCERN score for all 87 websites was 32.6 (SD 5.5). Only 1 website scored more than 51 points, which is good; 9 websites scored between 39 and 50 points, which is fair; 72 websites were poor; and 5 websites were very poor. The items that scored a mean of at least 3 points were criterion 9 (does it describe how each treatment works?) with a mean score of 3.96 (SD 1.10), criterion 2 (does it achieve its aims?) with a mean score of 3.18 (SD 0.70), criterion 1 (are the aims clear?) with a mean score of 3.10 (SD 0.57), and criterion 10 (does it describe the benefits of each treatment?) with a mean score of 3.01 (SD 0.42). The items that scored a mean of less than 2 points were criterion 15 (does it provide support for shared decision-making?), criterion 12 (does it describe what would happen if no treatment was used?), criterion 7 (does it provide details of additional sources of support and information?), criterion 8 (does it refer to the areas of uncertainty?), criterion 14 (is it clear that there may be more than one possible treatment choice?), criterion 6 (is this balanced or unbiased?), criterion 4 (is it clear what sources of information were used to compile the publication?), and criterion 5 (is it clear when the information used or reported in the publication was produced?). The mean score for criteria 16, which assessed the overall quality, was 1.49 (SD 0.73). No statistically significant association was found between the specialties and the total DISCERN score (all $P>.05$).

XSL•FO
RenderX

Figure 2. DISCERN scores for each item: #1 to #16 represent the criteria numbers of the DISCERN instrument. No: identification numbers for individual websites for analysis.



XSL•FO
RenderX

Table 4. Mean score for each of the 16 DISCERN instrument criteria[a].

Criterion number	Criteria	Mean score (SD)
1	Provides clear aims	3.10 (0.57)
2	Achieves its aims	3.18 (0.71)
3	Provides relevant information	2.21 (1.15)
4	Provides sources of information	1.70 (1.07)
5	Provides information production date	1.73 (1.16)
6	Balanced and unbiased	1.28 (0.58)
7	Provides additional sources of information	1.18 (0.52)
8	Refers to areas of uncertainty	1.24 (0.48)
9	Describes how each treatment works	3.97 (1.10)
10	Describes benefits of each treatment	3.01 (0.42)
11	Describes risks of each treatment	2.36 (0.79)
12	Describes what would happen if no treatment is used	1.10 (0.46)
13	Describes how treatment affects quality of life	2.70 (0.67)
14	Makes clear that there may be more than one possible treatment choice	1.24 (0.76)
15	Supports shared decision-making	1.05 (0.30)
16	Overall quality	1.49 (0.73)

[a]Each DISCERN item is assessed on a 5-point Likert scale (1=not at all, 5=completely). Total score = 32.55 (SD 5.53).

Discussion

Principal Findings

This study assesses compliance with Japanese medical advertising guidelines, exaggerated advertisement classification, and DISCERN scores of Japanese websites of medical institutions that prescribe GLP-1RAs off-label for weight loss, which Japanese consumers may refer to when considering weight loss with GLP-1RAs.

Many websites potentially violated Japan's medical advertising guidelines, and very few websites listed all the items required for advertising GLP-1RAs off-label use. In particular, only 9% (8/87) of the websites stated that using GLP-1RAs off-label was not covered by the Adverse Reactions Relief System for Drugs. In recent years, problems related to cosmetic therapy for weight loss have increased in Japan [47], so it is problematic that consumers have limited exposure to information indicating that the off-label use of GLP-1RAs is not covered by the Adverse Reactions Relief System for Drugs. Furthermore, concerns have been raised that GLP-1RAs may be obtained without a prescription and are at risk of misuse and abuse in inappropriate quantities [8]. However, research on who is actually misusing and what kinds of adverse effects occur because of misuse remains scarce. Generally, it has been reported that distinguishing between therapeutic errors, unintentional misuse, and intentional abuse of drugs is challenging, and the content of the reports tends to vary [5]. As mentioned above, in Japan, because it is not covered by the Adverse Reactions Relief System for Drugs, it may be difficult for consumers who have experienced adverse events due to the misuse of GLP-1RAs to claim compensation.

Approximately 83% (72/87) of the websites included exaggerated advertisements, and much of their content may mislead consumers into assuming that one can lose weight by using GLP-1RAs without exercise or dietary therapy. However, since participants in clinical trials evaluating the efficacy of obesity treatment drugs received medication therapy as an auxiliary to lifestyle interventions, the message that the use of GLP-1RAs for weight loss does not require exercise or dietary therapy is unscientific [48]. Additionally, according to another previous study, such messages may evoke a "compensatory health belief" in consumers that "if an alternative behavior is adopted, it is not necessary to engage in healthy behaviors or refrain from unhealthy ones." [49,50]

In this study, the mean of the total DISCERN scores of the analyzed websites was 32.56, which qualified as poor. Individually, most of the websites qualified as poor, only 1 website qualified as good, 9 websites qualified as fair, 72 websites qualified as poor, and 5 websites qualified as very poor. The mean score of criterion 16, the overall quality rating, was also low (score=1.49). These results were extremely low compared to those of previous studies that used the DISCERN instrument to assess the quality of health information [34-36,38]. For each question, criterion 15 (shared decision-making), criterion 12 (if no treatment is used), criterion 7 (additional information sources), criterion 8 (uncertainty), criterion 14 (possible treatment choice), and criterion 6 (balanced and unbiased) scored low, and criterion 9 (how each treatment works), criterion 2 (achieves its aims), and criterion 1 (provides clear aim) scored high. This trend was consistent with that of previous studies [34-36]. Unlike previous studies that compared DISCERN scores by the type of website operator (academic, government, or commercial agencies), only private clinic

XSL•FO
RenderX

websites that prescribe GLP-1RAs for off-label use were assessed in this study. Therefore, the DISCERN score was significantly lower in this study, possibly because the purpose of the websites was to sell GLP-1RAs and not to provide accurate information about obesity or obesity treatment to the public. In other words, these websites were advertisements designed to appeal to consumers about the effectiveness of GLP-1RAs for weight loss and not to draw their attention to other treatment options or shared decision-making with other health care providers or family members. Nevertheless, a previous study on consumer judgments of the reliability of web-based health information showed that website owners' authority had a positive impact on reliability and credibility [51]. Accordingly, if medical institutions representing specialties such as internal medicine were the owners of these websites, many consumers might trust the content of the websites. In summary, the problems with the information on medical institution websites identified by this study are not only unscientific and in violation of medical advertising guidelines but are also problematic from the perspective of public health and health promotion.

In this study, we observed that there were few public institution websites providing information on obesity treatment for consumers. To the best of our knowledge, no study has comprehensively compared the overall quality of web-based medical information between Japan and other countries. However, a previous study comparing palliative care websites in the United States and Japan reported that Japanese websites were of lower quality than the American ones [52]. A study focusing on the quality of web-based medical information within Japan has generally reported low quality [53]. This issue may be influenced by differences in the medical advertising regulations between Japan and other countries. Future studies should aim to provide a more comprehensive and international assessment of the quality of medical information.

Practical Implications

The results of this study have several implications. Regarding future studies, researchers should assess the reliability of media content, including medical institutions' websites that prescribe drugs for off-label use to contribute to better control of prescription drug misuse in the future. Many studies have investigated the consumer risk factors and preventive interventions for drug misuse. For example, a study investigating the misuse of over-the-counter drugs showed that health literacy, low education level, and misunderstanding of over-the-counter medicine instructions were correlated with the incidence of misuse [5]. Another study showed that the misuse of prescription stimulants for weight loss was positively associated with being female, lacking insurance, and having a history of lifetime misuse but not past-year misuse of prescription sedatives or tranquilizers [54]. Limited health literacy has been reported to be correlated with overweight/obesity in the first place [55], greater difficulty in the medication compliance, and higher mortality [56]. Thus, efforts to enhance consumer health literacy may be equally important as well as the content reliability of medical institutions' websites. In particular, studies on the risk of GLP-1RA misuse or off-label use are scarce and should be investigated.

Regarding future practices, this study suggests that governments should continuously monitor and provide corrective actions on websites, including unscientific content, to promote drug sales. A previous study in Japan showed that the number and volume of newspaper advertisements for dietary supplements that might mislead consumers decreased following the collective action of Japan's Consumer Affairs Agency [57]. Therefore, government monitoring and regulations are also important to address misleading information provided by medical institutions concerning prescription drug misuse. Furthermore, there are no uniform international guidelines for medical advertising. The permissibility of over-the-counter drug advertisements and advertisements for medical services differs from country to country. However, with the recent proliferation of social networking service platforms, misinformation or information with commercial intent about medicine can spread around the world in a short period of time. Therefore, it is also important to establish international medical advertising guidelines. Second, medical institutions and health care providers should recognize the current problem of GLP-1RA misuse and provide accurate information on the internet and through in-person explanations to individuals who are interested in weight loss. Additionally, medical institutions, academic organizations, and ministries should improve their websites with a focus on search engine optimization to ensure that high-quality medical information appears at the top of search engine results [58]. Internet-based companies, including Google, should be encouraged to prioritize the display of high-quality medical information. Third, medical institutions, health care providers, academic institutions, and ministries should inform citizens that current information on GLP-1RAs on websites is unreliable and that there is no miracle drug that is risk-free and absolutely effective. If individuals without obesity wish to lose weight, they should consult a trusted specialist such as a primary care physician to understand the risks and proceed with the therapy best suited for them.

Limitations and Strengths

This study has several limitations. First, this study only analyzed Japanese-language websites. The approval status of GLP-1RAs and medical advertising regulations differ from country to country. Thus, the generalizability of the results of this study to other countries is limited. Second, this study assesses the search results for web browsers and websites when viewed with web browsers. The results and website display contents may differ when using a smartphone. The results may have been affected by websites and search algorithms that were optimized for smartphones. Third, this study only used the DISCERN instrument for quality assessment. Therefore, this study was unable to discuss the reliability items other than those evaluated by DISCERN. There are multiple other quality assessment indicators such as the Health on the Net Foundation Code of Conduct certification and the Journal of the American Medical Association (JAMA) benchmark criteria. However, the Health on the Net Foundation Code of Conduct is not well known in Japan, and few websites have obtained it; therefore, it was not used in this study. The JAMA benchmark has only 4 criteria. Since most of the medical institutions of this study were private clinics, it was assumed that the 2 criteria of JAMA benchmark, authorship and attribution, would not differ. Additionally,

XSL•FO
RenderX

previous studies that evaluated web-based health information reported a positive correlation between JAMA benchmark scores and DISCERN scores [59,60]. Therefore, only DISCERN was used in this study. Additionally, the DISCERN instrument was developed to enable patients to assess the quality of written information related to treatment choices and to promote the creation of high-quality health information for documents intended for patients. Thus, assessing quality in terms of medical accuracy is not the primary focus of the DISCERN instrument. However, in this study, 2 internal medicine physician coders assessed the medical accuracy when evaluating questions 6, 9, and 12 of the DISCERN instrument, thereby reflecting medical accuracy in the DISCERN score. Fourth, this study relies on a single reviewer to screen eligible websites, which raises the possibility that some potentially eligible websites may have been excluded from the analyzed dataset. Fifth, the intercoder reliability between the 2 coders was ensured to be acceptable; however, coder bias may not have been completely eliminated. Finally, the impact of the websites analyzed in this study on viewers' cognition and behavior was outside the scope of this study; future studies should examine this aspect.

Despite these limitations, to the best of our knowledge, this is the first study on the quality of information on the websites of medical institutions that prescribe GLP-1RAs for weight loss in Japan. This study provides important implications for future studies and practice on better communication to decrease prescription drug misuse of GLP-1RAs, as discussed above. The quality of the websites of medical institutions that prescribe GLP-1RAs for weight loss was very low, and many of them did not comply with the medical advertising guidelines in Japan. This study shows that many websites, including exaggerated advertisements, may promote the misconception that exercise and dietary therapy are unnecessary. This study also shows that many websites provide biased information that may lead consumers to use GLP-1RAs without presenting alternative weight loss methods other than GLP-1RAs or the option of shared decision-making with health care providers or family members. Consumers and citizens who read such information may misuse GLP-1RAs because of insufficient understanding of their side effect risks and proper usage. Public institutions must monitor and regulate advertising content that violates guidelines and misleads individuals. Health care providers need to inform consumers about the risks and proper usage of GLP-1RAs and provide accurate information related to weight loss. Beneficial medications developed for people's health should not pose health risks because of the inappropriate information provided by health care providers.

Acknowledgments

This work was supported by the Japan Society for the Promotion of Science KAKENHI (23K09595).

Data Availability

The data presented in this study are available upon reasonable request to the corresponding author.

Authors' Contributions

RO and TO conceptualized this study and performed the methodology. RO and EF performed the formal analysis. RO wrote the original draft. HO acquired the funds. TK supervised this study. All authors contributed to reviewing and editing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
DISCERN instrument.
[DOC File , 29 KB - formative_v9i1e68792_app1.doc]

References

1. Casati A, Sedefov R, Pfeiffer-Gerschel T. Misuse of medicines in the European Union: a systematic review of the literature. Eur Addict Res 2012;18(5):228-245 [FREE Full text] [doi: 10.1159/000337028] [Medline: 22572594]
2. Schifano F, Chiappini S, Miuli A, Mosca A, Santovito MC, Corkery JM, et al. Focus on over-the-counter drugs? misuse: A systematic review on antihistamines, cough medicines, and decongestants. Front Psychiatry 2021;12:657397 [FREE Full text] [doi: 10.3389/fpsyt.2021.657397] [Medline: 34025478]
3. Martins SS, Sampson L, Cerdá M, Galea S. Worldwide prevalence and trends in unintentional drug overdose: A systematic review of the literature. Am J Public Health 2015 Nov;105(11):e29-e49. [doi: 10.2105/AJPH.2015.302843] [Medline: 26451760]
4. Smith SM, Dart RC, Katz NP, Paillard F, Adams EH, Comer SD, et al. Classification and definition of misuse, abuse, and related events in clinical trials: ACTTION systematic review and recommendations. Pain 2013 Nov;154(11):2287-2296 [FREE Full text] [doi: 10.1016/j.pain.2013.05.053] [Medline: 23792283]

XSL•FO
RenderX

5. Algarni M, Hadi MA, Yahyouche A, Mahmood S, Jalal Z. A mixed-methods systematic review of the prevalence, reasons, associated harms and risk-reduction interventions of over-the-counter (OTC) medicines misuse, abuse and dependence in adults. J Pharm Policy Pract 2021 Sep 13;14(1):76 [FREE Full text] [doi: 10.1186/s40545-021-00350-7] [Medline: 34517925]
6. Hashimoto T, Kaneda Y, Ozaki A, Hori A, Tsuchiya T. Eleven-year trend of drug and chemical substance overdose at a local emergency hospital in Japan. Cureus 2022 Dec;14(12):e32475 [FREE Full text] [doi: 10.7759/cureus.32475] [Medline: 36644086]
7. Matsumoto T. Distributed research report: survey of drug-related psychiatric disorders in psychiatric hospitals in Japan. Department of Drug Dependence Research.: National Institute of Mental Health; 2020. URL: https://www.ncnp.go.jp/nimh/yakubutsu/report/pdf/J_NMHS_2020.pdf [accessed 2024-12-22]
8. Chiappini S, Vickers-Smith R, Harris D, Papanti Pelletier GD, Corkery JM, Guirguis A, et al. Is there a risk for semaglutide misuse? focus on the Food and Drug Administration's FDA Adverse Events Reporting System (FAERS) pharmacovigilance dataset. Pharmaceuticals (Basel) 2023 Jul 11;16(7):994 [FREE Full text] [doi: 10.3390/ph16070994] [Medline: 37513906]
9. Butuca A, Dobrea CM, Arseniu AM, Frum A, Chis AA, Rus LL, et al. An assessment of semaglutide safety based on real world data: from popularity to spontaneous reporting in Eudravigilance database. Biomedicines 2024 May 18;12(5):1124 [FREE Full text] [doi: 10.3390/biomedicines12051124] [Medline: 38791086]
10. Mahase E. Semaglutide: FDA publishes warning amid reports of patients overdosing. BMJ 2024 Jul 30;386:q1696. [doi: 10.1136/bmj.q1696] [Medline: 39079728]
11. Han SH, Safeek R, Ockerman K, Trieu N, Mars P, Klenke A, et al. Public interest in the off-label use of glucagon-like peptide 1 agonists (Ozempic) for cosmetic weight loss: a Google trends analysis. Aesthet Surg J 2023 Dec 14;44(1):60-67. [doi: 10.1093/asj/sjad211] [Medline: 37402640]
12. FDA approves new drug treatment for chronic weight management, first since 2014. USFDA. URL: https://www.fda.gov/news-events/press-announcements/fda-approves-new-drug-treatment-chronic-weight-management-first-2014 [accessed 2024-09-08]
13. Suran M. As Ozempic's popularity soars, here's what to know about semaglutide and weight loss. JAMA 2023 May 16;329(19):1627-1629. [doi: 10.1001/jama.2023.2438] [Medline: 37099334]
14. Goodman B. Poison centers see nearly 1,500% increase in calls related to injected weight-loss drugs as people accidentally overdose. CNN Health. URL: https://www.cnn.com/2023/12/13/health/semaglutide-overdoses-wellness [accessed 2024-10-21]
15. Brafman N. Ozempic: French authorities issue alert for anti-diabetic drug misused for weight loss. Le Monde 2023 Mar 2. URL: https://www.lemonde.fr/en/health/article/2023/03/02/ozempic-french-authorities-issue-alert-for-anti-diabetic-drug-misused-for-weight-loss_6017913_14.html [accessed 2024-10-21]
16. Mahase E. GLP-1 shortages will not resolve this year, EMA warns, amid concern over off-label use. BMJ 2024 Jun 28;385:q1448. [doi: 10.1136/bmj.q1448] [Medline: 38942431]
17. Hathaway JT, Shah MP, Hathaway DB, Zekavat SM, Krasniqi D, Gittinger JW, et al. Risk of nonarteritic anterior ischemic optic neuropathy in patients prescribed semaglutide. JAMA Ophthalmol 2024 Aug 01;142(8):732-739. [doi: 10.1001/jamaophthalmol.2024.2296] [Medline: 38958939]
18. Sodhi M, Rezaeianzadeh R, Kezouh A, Etminan M. Risk of gastrointestinal adverse events associated with glucagon-like peptide-1 receptor agonists for weight loss. JAMA 2023 Nov 14;330(18):1795-1797 [FREE Full text] [doi: 10.1001/jama.2023.19574] [Medline: 37796527]
19. Recommendations of the executive steering group on shortages and safety of medicinal products on shortage of glucagon-like peptide-1 (GLP-1) receptor agonists. European Medicines Agency. URL: https://www.quotidianosanita.it/allegati/allegato1719409581.pdf [accessed 2024-09-08]
20. Notes on the establishment of guidelines for the promotion of appropriate use of semaglutide (genetic recombination) preparation for the indication of obesity document in Japanese. Ministry of Health, Labour and Welfare, Japan. URL: https://www.pmda.go.jp/files/000265450.pdf [accessed 2024-09-08]
21. Hakariya H, Ohnishi M, Tanimoto T. Japan initiates market authorization of weight-loss drug semaglutide under universal health coverage, but with stringent prescription restrictions. Diabetes Obes Metab 2024 Aug;26(8):3006-3008. [doi: 10.1111/dom.15638] [Medline: 38747140]
22. Request for cooperation due to tight inventory of GLP-1 receptor agonist (Part 2) document in Japanese. Ministry of Health, Labour, and Welfare, Japan. URL: https://www.mhlw.go.jp/content/001165743.pdf [accessed 2024-11-09]
23. Position on off-label use of GLP-1 receptor agonists and GIP/GLP-1 receptor agonists document in Japanese. The Japan Diabetes Society. URL: https://www.jds.or.jp/uploads/files/document/info/jds_statement_GLP-1.pdf [accessed 2024-10-28]
24. Statement on the safe and appropriate use of obesity drugs document in Japanese. Japan Society for the Study of Obesity. URL: http://www.jasso.or.jp/data/Introduction/pdf/academic-information_statement_20231127.pdf [accessed 2024-09-08]
25. Telemedicine trouble for slimming purposes: Subscription trouble? where diabetes drugs are prescribed off-label for several months for the purpose of weight loss document in Japanese. National Consumer Affairs Center of Japan. URL: https://www.kokusen.go.jp/news/data/n-20231220_1.html [accessed 2024-09-08]

XSL•FO
RenderX

26. Before you get a cosmetic procedure, recheck to make sure you've been properly briefed and understand it again document in Japanese. Consumer Affairs Agency, Government of Japan. URL: https://www.caa.go.jp/policies/policy/consumer_policy/information/information_002/assets/information_002_240202_02.pdf [accessed 2024-09-08]
27. Medical Care Act (Japanese Law Translation System). Ministry of Justice, Japan. 1948. URL: https://www.japaneselawtranslation.go.jp/ja/laws/view/4006 [accessed 2024-09-08]
28. Regulations on advertisements of hospitals under the Medical Care Act (in Japanese). Ministry of Health, Labour and Welfare, Japan. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/iryou/kokokukisei/index.html [accessed 2024-09-08]
29. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
30. Shan Y, Xing Z, Dong Z, Ji M, Wang D, Cao X. Translating and adapting the DISCERN instrument into a simplified Chinese version and validating its reliability: Development and usability study. J Med Internet Res 2023 Feb 02;25:e40733 [FREE Full text] [doi: 10.2196/40733] [Medline: 36729573]
31. Montoya A, Llopis N, Gilaberte I. Validation of the translation of an instrument to measure reliability of written information on treatment choices: a study on attention deficit/hyperactivity disorder (ADHD). Educ Health (Abingdon) 2011 Dec;24(3):577. [Medline: 22267353]
32. Logullo P, Torloni MR, de O C Latorraca C, Riera R. The Brazilian Portuguese version of the DISCERN instrument: Translation procedures and psychometric properties. Value Health Reg Issues 2019 Dec;20:172-179 [FREE Full text] [doi: 10.1016/j.vhri.2019.09.001] [Medline: 31622803]
33. Batchelor JM, Ohya Y. Use of the DISCERN instrument by patients and health professionals to assess information resources on treatments for asthma and atopic dermatitis. Allergol Int 2009 Mar;58(1):141-145 [FREE Full text] [doi: 10.2332/allergolint.08-SC-0022] [Medline: 19153541]
34. Killip SC, Kwong NKR, MacDermid JC, Fletcher AJ, Carleton NR. The quality, readability, completeness, and accuracy of PTSD websites for firefighters. Int J Environ Res Public Health 2020 Oct 19;17(20):7629 [FREE Full text] [doi: 10.3390/ijerph17207629] [Medline: 33086772]
35. Li Y, Zhou X, Zhou Y, Mao F, Shen S, Lin Y, et al. Evaluation of the quality and readability of online information about breast cancer in China. Patient Educ Couns 2021 Apr;104(4):858-864 [FREE Full text] [doi: 10.1016/j.pec.2020.09.012] [Medline: 32988687]
36. Tan DJY, Ko TK, Fan KS. The readability and quality of web-based patient information on nasopharyngeal carcinoma: quantitative content analysis. JMIR Form Res 2023 Nov 27;7:e47762 [FREE Full text] [doi: 10.2196/47762] [Medline: 38010802]
37. Patel AA, Joshi C, Varghese J, Hassan AM, Janis JE, Galiano RD. Do websites serve our patients well? a comparative analysis of online information on cosmetic injectables. Plast Reconstr Surg 2022 Apr 01;149(4):655e-668e. [doi: 10.1097/PRS.0000000000008921] [Medline: 35139065]
38. Daraz L, Morrow AS, Ponce OJ, Beuschel B, Farah MH, Katabi A, et al. Can patients trust online health information? a meta-narrative systematic review addressing the quality of health information on the internet. J Gen Intern Med 2019 Sep;34(9):1884-1891 [FREE Full text] [doi: 10.1007/s11606-019-05109-0] [Medline: 31228051]
39. Weaver JB, Mays D, Weaver SS, Hopkins GL, Eroglu D, Bernhardt JM. Health information-seeking behaviors, health indicators, and health risks. Am J Public Health 2010 Aug;100(8):1520-1525. [doi: 10.2105/AJPH.2009.180521] [Medline: 20558794]
40. Telecommunications usage trends survey (in Japanese). Ministry of Internal Affairs and Communications. URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r05/html/datashu.html#f00281 [accessed 2024-10-12]
41. Montemurro P, Porcnik A, Hedén P, Otte M. The influence of social media and easily accessible online information on the aesthetic plastic surgery practice: literature review and our own experience. Aesthetic Plast Surg 2015 Apr;39(2):270-277. [doi: 10.1007/s00266-015-0454-3] [Medline: 25697277]
42. Google Japan. URL: https://www.google.co.jp/ [accessed 2024-12-23]
43. Yahoo! Japan. URL: https://www.yahoo.co.jp/ [accessed 2024-12-23]
44. Search engine share in Japan. Ministry of Internal Affairs and Communications. URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r05/html/datashu.html#f00228 [accessed 2024-09-08]
45. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002 Mar 09;324(7337):573-577 [FREE Full text] [doi: 10.1136/bmj.324.7337.573] [Medline: 11884321]
46. Ved R, Cobbold N, Igbagiri K, Willis M, Leach P, Zaben M. Online patient information on Vagus Nerve Stimulation: How reliable is it for facilitating shared decision making? Seizure 2017 Aug;50:125-129 [FREE Full text] [doi: 10.1016/j.seizure.2017.06.009] [Medline: 28651125]
47. Trouble with online medical treatment for slimming purposes, etc document in Japanese. National Consumer Affairs Center, Japan. URL: https://www.kokusen.go.jp/pdf/n-20231220_1.pdf [accessed 2024-10-01]

XSL•FO
RenderX

48. Jastreboff AM, Aronne LJ, Ahmad NN, Wharton S, Connery L, Alves B, SURMOUNT-1 Investigators. Tirzepatide once weekly for the treatment of obesity. N Engl J Med 2022 Jul 21;387(3):205-216. [doi: 10.1056/NEJMoa2206038] [Medline: 35658024]
49. Hartmann C, Keller C, Siegrist M. Compensatory beliefs, nutrition knowledge and eating styles of users and non-users of meal replacement products. Appetite 2016 Oct 01;105:775-781. [doi: 10.1016/j.appet.2016.07.013] [Medline: 27421243]
50. Iye R, Okuhara T, Okada H, Goto E, Furukawa E, Kiuchi T. The effect of exposure to "exemption" video advertisements for functional foods: a randomized control study in Japan. Healthcare (Basel) 2022 Feb 11;10(2):345 [FREE Full text] [doi: 10.3390/healthcare10020345] [Medline: 35206959]
51. Zhang Y, Kim Y. Consumers' evaluation of web-based health information quality: Meta-analysis. J Med Internet Res 2022 Apr 28;24(4):e36463 [FREE Full text] [doi: 10.2196/36463] [Medline: 35482390]
52. Tanabe K, Fujiwara K, Ogura H, Yasuda H, Goto N, Ohtsu F. Quality of web information about palliative care on websites from the United States and Japan: comparative evaluation study. Interact J Med Res 2018 Apr 03;7(1):e7 [FREE Full text] [doi: 10.2196/ijmr.9574] [Medline: 29615388]
53. Yamamura A, Watanabe S, Yamaguchi S, Iwata K, Kimura S, Mikami Y, et al. Readability and quality of online patient resources regarding knee osteoarthritis and lumbar spinal stenosis in Japan. J Orthop Sci 2024 Sep;29(5):1313-1318. [doi: 10.1016/j.jos.2023.08.003] [Medline: 37599135]
54. Compton WM, Han B, Blanco C, Johnson K, Jones CM. Prevalence and correlates of prescription stimulant use, misuse, use disorders, and motivations for misuse among adults in the United States. Am J Psychiatry 2018 Aug 01;175(8):741-755 [FREE Full text] [doi: 10.1176/appi.ajp.2018.17091048] [Medline: 29656665]
55. Michou M, Panagiotakos DB, Costarelli V. Low health literacy and excess body weight: a systematic review. Cent Eur J Public Health 2018 Sep;26(3):234-241 [FREE Full text] [doi: 10.21101/cejph.a5172] [Medline: 30419628]
56. Fan Z, Yang Y, Zhang F. Association between health literacy and mortality: a systematic review and meta-analysis. Arch Public Health 2021 Jul 01;79(1):119 [FREE Full text] [doi: 10.1186/s13690-021-00648-7] [Medline: 34210353]
57. Okuhara T, Yokota R, Shirabe R, Iye R, Okada H, Kiuchi T, et al. Japanese newspaper advertisements for dietary supplements before and after COVID-19: a content analysis. BMJ Open 2021 Nov 23;11(11):e050898 [FREE Full text] [doi: 10.1136/bmjopen-2021-050898] [Medline: 34815281]
58. Modave F, Shokar NK, Peñaranda E, Nguyen N. Analysis of the accuracy of weight loss information search engine results on the internet. Am J Public Health 2014 Oct;104(10):1971-1978. [doi: 10.2105/AJPH.2014.302070] [Medline: 25122030]
59. Foster B, Grampp S, Ozdag Y, Tang A, Vazquez F, Mercuri J. Assessing the quality, content, and readability of online patient resources on viscosupplementation. Journal of Orthopaedic Experience & Innovation Charter Services New York d/b/a Journal of Orthopaedic Experience and Innovation; 2024 Oct 4 2024 Oct 4;5(2):1-8. [doi: 10.60118/001c.120353]
60. Ozduran E, Büyükçoban S. Evaluating the readability, quality and reliability of online patient education materials on post-covid pain. PeerJ 2022;10:e13686 [FREE Full text] [doi: 10.7717/peerj.13686] [Medline: 35880220]

Abbreviations

FDA: Food and Drug Administration
GLP-1: glucagon-like peptide-1
GLP-1RA: glucagon-like peptide-1 receptor agonist
JAMA: Journal of the American Medical Association

Edited by A Mavragani; submitted 13.11.24; peer-reviewed by E Gray, N Misra, M Agbede; comments to author 30.11.24; revised version received 08.12.24; accepted 16.12.24; published 01.01.25.

Please cite as:
Oyama R, Okuhara T, Furukawa E, Okada H, Kiuchi T
Quality Assessment of Medical Institutions' Websites Regarding Prescription Drug Misuse of Glucagon-Like Peptide-1 Receptor Agonists by Off-Label Use for Weight Loss: Website Evaluation Study
JMIR Form Res 2025;9:e68792
URL: https://formative.jmir.org/2025/1/e68792
doi:10.2196/68792
PMID:

©Rie Oyama, Tsuyoshi Okuhara, Emi Furukawa, Hiroko Okada, Takahiro Kiuchi. Originally published in JMIR Formative Research (https://formative.jmir.org), 01.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The

XSL•FO
RenderX

complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Public Health Discussions on Social Media: Evaluating Automated Sentiment Analysis Methods

Lisa M Gandy[1*], PhD; Lana V Ivanitskaya[2*], PhD; Leeza L Bacon[3*], DHA; Rodina Bizri-Baryak[2*], DHA

[1]Department of Computer Science, College of Sciences and Liberal Arts, Kettering University, Flint, MI, United States

[2]Department of Health Administration, The College of Health Professions, Central Michigan University, Mt Pleasant, MI, United States

[3]Department of Healthcare Management, Northwood University, Midland, MI, United States

[*]all authors contributed equally

Corresponding Author:
Lisa M Gandy, PhD
Department of Computer Science
College of Sciences and Liberal Arts
Kettering University
1700 University Ave
2300 AB
Flint, MI, 48504
United States
Phone: 1 9898547001
Email: lgandy@kettering.edu

Abstract

Background: Sentiment analysis is one of the most widely used methods for mining and examining text. Social media researchers need guidance on choosing between manual and automated sentiment analysis methods.

Objective: Popular sentiment analysis tools based on natural language processing (NLP; VADER [Valence Aware Dictionary for Sentiment Reasoning], TEXT2DATA [T2D], and Linguistic Inquiry and Word Count [LIWC-22]), and a large language model (ChatGPT 4.0) were compared with manually coded sentiment scores, as applied to the analysis of YouTube comments on videos discussing the opioid epidemic. Sentiment analysis methods were also examined regarding ease of programming, monetary cost, and other practical considerations.

Methods: Evaluation methods included descriptive statistics, receiver operating characteristic (ROC) curve analysis, confusion matrices, Cohen κ, accuracy, specificity, precision, sensitivity (recall), F_1-score harmonic mean, and the Matthews correlation coefficient. An inductive, iterative approach to content analysis of the data was used to obtain manual sentiment codes.

Results: A subset of comments were analyzed by a second coder, producing good agreement between the 2 coders' judgments (κ=0.734). YouTube social media about the opioid crisis had many more negative comments (4286/4871, 88%) than positive comments (79/662, 12%), making it possible to evaluate the performance of sentiment analysis models in an unbalanced dataset. The tone summary measure from LIWC-22 performed better than other tools for estimating the prevalence of negative versus positive sentiment. According to the ROC curve analysis, VADER was best at classifying manually coded negative comments. A comparison of Cohen κ values indicated that NLP tools (VADER, followed by LIWC's tone and T2D) showed only fair agreement with manual coding. In contrast, ChatGPT 4.0 had poor agreement and failed to generate binary sentiment scores in 2 out of 3 attempts. Variations in accuracy, specificity, precision, sensitivity, F_1-score, and MCC did not reveal a single superior model. F_1-score harmonic means were 0.34-0.38 (SD 0.02) for NLP tools and very low (0.13) for ChatGPT 4.0. None of the MCCs reached a strong correlation level.

Conclusions: Researchers studying negative emotions, public worries, or dissatisfaction with social media face unique challenges in selecting models suitable for unbalanced datasets. We recommend VADER, the only cost-free tool we evaluated, due to its excellent discrimination, which can be further improved when the comments are at least 100 characters long. If estimating the prevalence of negative comments in an unbalanced dataset is important, we recommend the tone summary measure from LIWC-22. Researchers using T2D must know that it may only score some data and, compared with other methods, be more time-consuming and cost-prohibitive. A general-purpose large language model, ChatGPT 4.0, has yet to surpass the performance of NLP models, at least for unbalanced datasets with highly prevalent (7:1) negative comments.

XSL•FO
RenderX

(JMIR Form Res 2025;9:e57395) doi:10.2196/57395

KEYWORDS

ChatGPT; VADER; valence aware dictionary for sentiment reasoning; LIWC-22; machine learning; social media; sentiment analysis; public health; population health; opioids; drugs; pharmacotherapy; pharmaceuticals; medications; YouTube

Introduction

The Pew Research Center [1] reports that as of 2021, 72% of Americans used social media. On a global scale, as of 2019, social media platforms were used by 1 in 3 people worldwide and by more than two-thirds of all internet users [2]. With social media users rapidly increasing, user-generated content has grown exponentially. Generated by a broad swath of global citizens, the data provides insights into a wide array of human experiences, for example, the effects of the opioid crisis on physical, mental, and social well-being.

Analysis of the opioid epidemic revealed the struggles of opioid victims, their families, and communities. This information adds value to health policy analysis.

This study examines the use of YouTube, a unique social media platform. It is a diverse medium because its primary purpose is content sharing and educating its users about relevant topics [3]. Users seek out content of interest and, comment and react. When examining health care crises and the relay of medical information, sentiment analysis may be used to gauge the response to the stimulus. By analyzing the sentiments expressed in comments, researchers can perform qualitative content analyses to explore how these reactions influence reputation, potentially affect individuals, and impact the communities involved. Furthermore, understanding the sentiment toward health-related content or issues like human suffering and societal concerns can aid policy makers in developing strategies to enhance public health.

Sentiment analysis is the most popular artificial intelligence used to mine and examine all text types in various fields of study. Sentiment analysis is a computational method that extracts sentiment from a text. Some sentiment analysis methods use rule-based lexicons such as Linguistic Inquiry and Word Count (LIWC)-22 [4]. Other sentiment analysis methods use traditional machine learning approaches such as Support Vector Machines [5] and Naive Bayes classification [6], while others use deep learning [7]. Sometimes, a hybrid approach is used between 2 or more sentiment analysis methods [8].

This study compares 3 popular sentiment analysis methods on social media data: Valence Aware Dictionary for Sentiment Reasoning (VADER), TEXT2DATA (T2D), and LIWC, and also uses the ChatGPT large language model (LLM) for sentiment analysis. A total of 2 methods (VADER and LIWC) were picked due to their previous validation and use in many published studies by scholars from different disciplines; other methods were chosen because of their user-friendly interface and no requirement of prerequisite programming skills (LIWC-22 and T2D). ChatGPT was chosen as LLM represent a game-changing technological leap in natural language processing (NLP), including sentiment analysis.

VADER [9] is a rule and lexicon-based sentiment analysis tool that, when analyzing text, returns a numeric valence (polarity) score between –1 (extremely negative) and +1 (extremely positive). The VADER lexicon is primarily built using pre-existing lexicons, but these lexicons were extended to include emoticons, acronyms, and slang commonly used in social media. Each lexical feature is assigned a score; then, the score is shifted based on the presence of punctuation marks, capital letters, and negations. The lexical scores are then averaged. VADER has been validated on multiple data types, including product and movie reviews, and has been used as a quality benchmark in numerous studies [9-11]. It has been used for sentiment analysis in a wide variety of areas, such as customer reviews and opinion mining [12,13], political discourse analysis [14,15], and mental health studies [16,17]. Although VADER has been used extensively, concerns about VADER's sensitivity to text length have been expressed [18,19]. For example, as each word in a text is assigned a polarity score, words with strong sentiment can unduly influence the overall sentiment of a short text. There are proposed methods to mitigate this phenomenon [9], such as considering the text of an entire document about a sentence. Another possible issue is that VADER is primarily used as a Python-based program, and users must be moderately proficient in Python programming to use its full capabilities.

T2D [20] is a Microsoft Excel or Google Sheets add-in with a sentiment analysis application programming interface (API). It classifies text into 5 categories: very negative, negative, neutral, positive, and very positive. The T2D website contains scant information about the corpus or the methodology used for sentiment classification. The website states that the API is "based on an NLP engine" and that the "system also contains specially prepared classification models for Twitter (rebranded as X) and other social media content, trained on billions of manually verified entries" [20]. It is a paid subscription model, but users can make 1000 API calls a month for free. The next tier is 10,000 API calls for US $27 a month, then a variety of tiers exist, with the highest tier being Enterprise at US $351 a month and Unlimited API calls. As a Microsoft Excel add-in and a user-friendly tool for those familiar with Microsoft Excel, T2D has been applied to studies that examined tourist experiences [21], compared Twitter comments with polling results [22], and analyzed digital mental health interventions [23].

LIWC-22 is a dictionary-based sentiment analysis tool with a representative score in over 100 categories. The score in each category indicates the percentage of words in the text corresponding to the particular category and, therefore, can range from 0% to 100%. Many LIWC-22 categories are organized in a hierarchical structure. The same word may be categorized into multiple categories. For instance, "celebrate" is in both the positive emotion and achievement categories.

XSL•FO
RenderX

Each category is represented internally in LIWC-22 by a dictionary with words and emoticons related to that category. A complete list of LIWC-22 categories can be found in [24]. The creators of LIWC-22 selected a vast pool of words that represented a wide range of linguistic categories and psychological dimensions, such as emotions, cognitive processes, social terms, pronouns, prepositions, and other linguistic constructs. LIWC-22 is the fifth iteration of LIWC-22 (LIWC 2001, 2007, 2015, and 2019). To create LIWC, experts in linguistics and psychology first gathered a large corpus of texts and tagged each word into a linguistic category. The resulting dictionaries underwent validation studies and then passed through a refinement phase when researchers added new vocabulary and, as meanings shifted, modified existing word-category pairings. LIWC-22 is a subscription model. Currently, users pay US $129.95 for a 3-year subscription, and shorter-term subscriptions are offered at lower prices. LIWC-22 has both a web-based app and an app for download. The users upload their file (JSON, CSV, and EXCEL), choose the LIWC-22 dictionary they would like to use, and then select the label in their data that denotes the text to be classified. Users also have the option to choose all or specific LIWC-22 categories to be evaluated.

Similar to VADER and T2D, LIWC-22 has been used as a sentiment analysis tool in many studies, such as education [25], public discourse analysis [26], and brand perception [27] among others. Many categories in LIWC-22 could represent sentiment, including positive emotion, negative emotion, anger, sadness, and anxiety. LIWC-22 includes positive tone and negative tone dimensions but also includes another Tone variable that merges the 2 dimensions into a single summary variable. The higher the number, the more positive the tone; numbers below 50 are counted as negative tone. This paper uses the LIWC-22 tone composite (summary) measure as its dictionary was more extensive than other sentiment-based LIWC-22 categories. We will refer to it as LIWC tone.

Previous efforts to evaluate sentiment analysis tools did not compare LIWC tone, VADER, and T2D. For example, Boukes et al [28] compared LIWC, SentiStrength, Pattern, Polyglot, and DANEW on economic news in Danish. These sentiment analysis tools were chosen as they were off the shelf and supported the Danish language. Hartmann et al [29] compared LIWC-22 with a host of machine learning-based sentiment analysis methods (Support Vector Machines, Neural Networks, K Nearest Neighbors, and Random Forests) but did not include VADER or T2D and focus on the marking aspect of social media data.

Recent advances in NLP have led to the development of powerful language models, such as the GPT series, including ChatGPT (GPT–3.5 and GPT–4) [30]. These models, pretrained on vast amounts of text data, have demonstrated strong performance across tasks like language translation, text summarization, and question-answering [31]. A LLM ChatGPT, in particular, has shown promise in education, health care, reasoning, text generation, human-machine interaction, and scientific research [30,32]. Despite these opportunities, challenges and ethical concerns remain, particularly regarding accuracy. The accuracy of these models depends heavily on the quality, diversity, and complexity of the training data, as well as the quality of user input [30,31]. Previous research has highlighted the importance of developing higher-order thinking skills in education, but NLP systems may struggle with the nuances of human language, leading to potential errors [30-32].

We compare the sentiment scores produced by VADER, T2D, LIWC_tone, and ChatGPT 4.0 to manually created codes. The data source is YouTube comments on videos that discussed the opioid epidemic. This data source was chosen as it was easily accessible to the authors, but also because each of the sentiment analysis classifiers mentioned has been heavily used to classify social media data in the past. As social media (and the data it creates) grows, researchers will likely analyze it using the sentiment tools discussed in this paper.

In the remainder of the paper, we discuss the methods by which data was collected and coded and will then compare the results given by VADER, T2D, LIWC_tone, and ChatGPT 4.0 to human coding. This paper reports which methods give accurate sentiment scores and discusses other practical considerations when using these tools. We will end with a discussion of the results and considerations for future research.

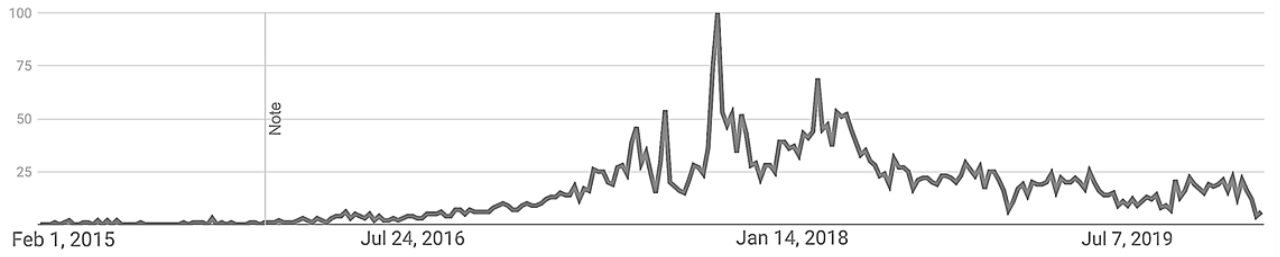
Methods

Data Sources

To evaluate sentiment analysis methods, we used secondary data from a qualitative study about the opioid crisis in the United States [33], which included manually coded sentiment from YouTube comments. To collect YouTube comments (N=8761), the term “opioid” epidemic was searched on YouTube with a date range between January 1, 2017, and December 31, 2018. The majority of the videos collected were located on CNN’s (Cable News Network) YouTube channel and the Fox News YouTube Channel. Subsequently, videos were ranked by the number of views, and the 20 most watched videos by CNN (10 newscasts) and Fox News (10 newscasts) were kept for further analysis. As Google Trends indicates (Figure 1), the chosen dates coincide with a particularly high interest in the opioid epidemic. The comments for each video were downloaded using the Netlytic website. Comments were deidentified by deleting email addresses and assigning codes such as a1, a2, and so on, to track comments on comments. All other comment information was left unaltered.

XSL•FO
RenderX

Figure 1. Google Trends results for the search term "opioid epidemic." The search was constrained to dates between February 2015 and December 2019. The y-axis indicates percent popularity.



Sentiment Measures

An inductive, iterative approach to content analysis of the data was used to obtain manual sentiment codes. Bacon [33] created a codebook to determine the sentiment of comments as positive, neutral, or negative based on the commenter's attitude toward the video or another comment. The comments were sorted by time of post and video, then coded in the context of the broader discussion. Video transcripts also informed the coding. A subset of comments was analyzed by a second coder, producing good agreement between the 2 coders' judgments (κ=0.734; 95% CI 0.57-0.89; P<.001).

To begin testing automated sentiment tools, comments with manual codes indicating neutral or unclear sentiment were excluded from the data. The remaining comments were scored using VADER, T2D's Google Sheets add-in, LIWC, and ChatGPT 4.0. Specifically, the VADER compound score, T2D sentiment score, and LIWC-22 tone score (LIWC tone), a composite measure for positive and negative tone dimensions, were used. ChatGPT 4.0, VADER, and T2D were centered around 0 and ranged from –1 to 1. Negative values indicated a negative sentiment and positive values showed a positive sentiment. LIWC tone scores ranged from 1 to 100, representing percentiles based on standardized scores from large comparison corpora [34]. They are calculated using a dictionary with words, word stems, phrases, and select emoticons built for text analysis. LIWC tone overall mean for 15 diverse corpora was 47.81 (SD 26.39). Bacon's codebook was adapted to prompt the general-purpose LLM to generate positive and negative sentiment classification (Multimedia Appendix 1 for the ChatGPT 4.0 prompt). The classification was generated on the third try after two failed attempts.

Analyses

We performed the receiver operating characteristic (ROC) curve analysis and computed descriptive statistics, confusion matrices, Cohen , accuracy, specificity, precision, sensitivity (recall), F_1-score harmonic mean, and the Matthews correlation coefficient (MCC). Relying solely on the ROC curve without considering precision and negative predictive value can lead to a misleading assessment of a model's success [35]. Although widely used in machine learning, the F_1-score also has limitations; it can vary when positive and negative classes are exchanged, potentially distorting its interpretation [36]. In addition, the F_1-score does not account for correctly classified negative and positive samples, drawing criticism for diverging from more intuitive metrics like accuracy and losing significance when class labels are reversed [36]. In contrast, MCC provides a more balanced evaluation, achieving its highest values of –1 or +1 only when the classifier performs well across all 4 key rates of the confusion matrix: sensitivity, specificity, precision, and negative predictive value [35,36].

Finally, to reveal misclassification patterns by model, we analyzed the content of the comments marked as false positives and false negatives from both NLP and LLM models and compared them with true positives and true negatives identified through manual coding. We summarized possible reasons behind misclassifications and provided representative comments as illustrations.

Ethical Considerations

Data were collected from a social media platform (YouTube) where data are publicly available. However, all sentiment evaluation methods were performed at a macro scale and not at the user level. In addition, social media profile information is not shared in the data provided in this article.

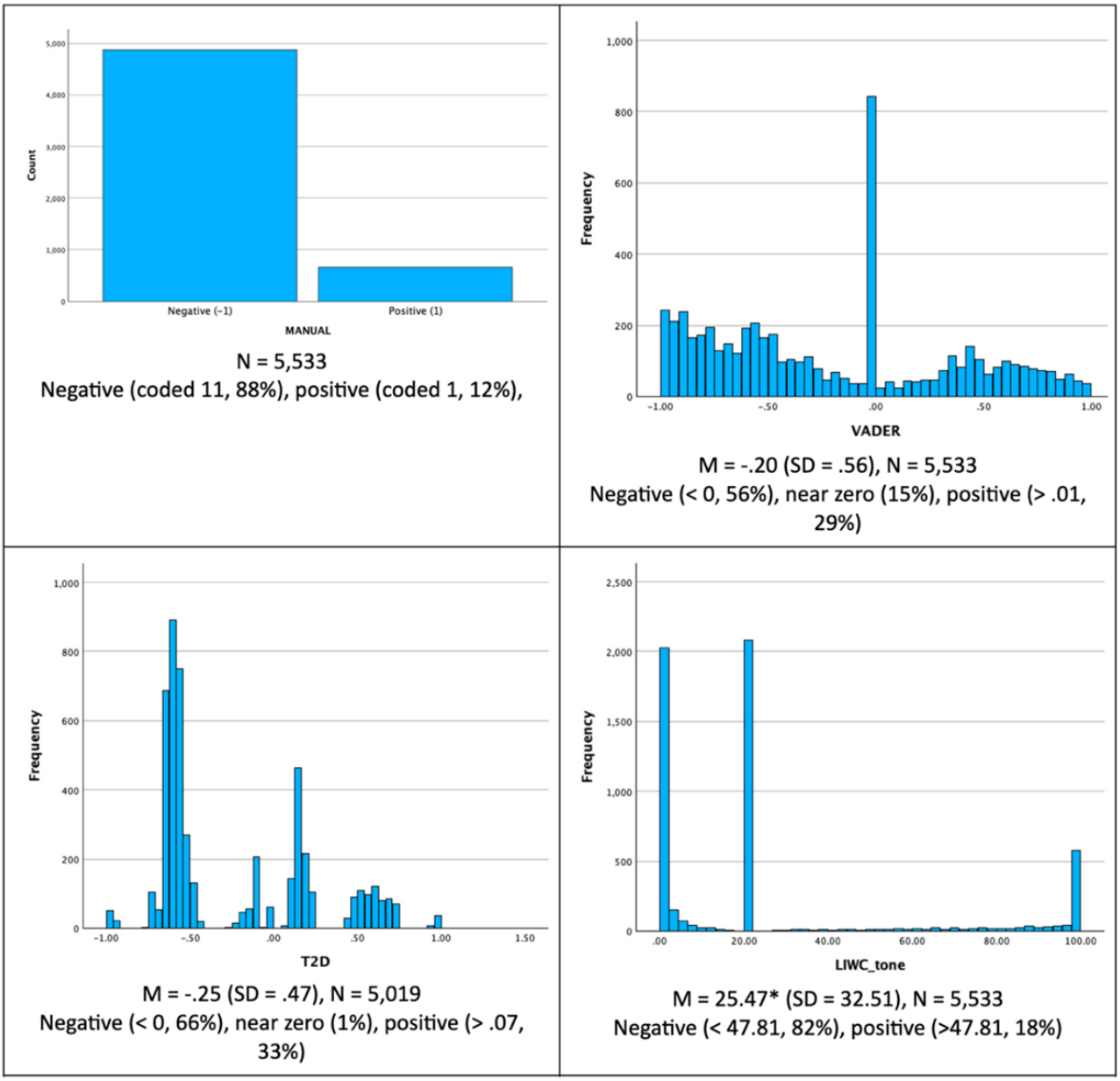
Results

Manual coding placed 63.2% of comments into either positive or negative categories; the remainder were neutral, or their sentiment could not be ascertained. Only positively or negatively classified comments were used for further analyses (N=5533). Positive comments were much less common (79/662, 12%) than negative (4286/4,871, 88%) comments.

VADER, LIWC tone, and ChatGPT 4.0 were able to classify all comments. However, for unknown reasons, T2D sentiment scores could not be calculated for 514 comments, resulting in 9% of missing values. T2D is a "black box" system, and documentation has not been released. As shown in Figure 2, LIWC tone analysis most closely matched the high prevalence of negative comments (88%) in manually coded data: 82% of LIWC tone scores fell below the mean of 47.81. ChatGPT 4.0 overestimated negative sentiment, classifying 98% of the comments as negative. In comparison, only 56% of VADER and 66% of T2D scores were negatively scored. VADER score distribution had a mode around 2 and was the most continuous compared with T2D and LIWC. VADER assigned near-0 scores to 15% of comments, T2D had 1% of near-zero data but did not score 9% of comments, and LIWC tone scores were very unevenly distributed with 3 modes (at scale's endpoints and 23.23).

XSL•FO
RenderX

Figure 2. A comparison of Valence Aware Dictionary for Sentiment Reasoning, TEXT2DATA, and Linguistic Inquiry and Word Count tone score distributions for 5533 comments classified as either negative or positive using manual coding. *LIWC tone scores below 47.81 are considered negative. T2D: TEXT2DATA.



As an overall measure of discrimination, we used the ROC curve [37]. Discrimination is the ability of a measure to discern between social media comments that are manually coded as negative or positive. We were interested in evaluating continuously measured VADER, T2D, and LIWC tone scores as predictors of positive and negative sentiment in YouTube comments; manual coding is an outcome against which the 3 models were assessed. ROC curve analysis was not performed for ChatGPT 4.0 because it was only generated as a binary measure. The area under the ROC curve is a measure of the overall discriminatory ability of the binomial logistic regression model, that is, the ability of a chosen sentiment scoring method to classify comments into the 2 groups of our dichotomous dependent variable, a manually coded sentiment where negative is assigned a value –1 and positive is given a value of 1. ROC curve analysis is most suitable for balanced analyses. While we are interested in correctly classifying positives and negatives, our dataset is unbalanced due to the high prevalence of negative comments.

As shown in Figure 3, ROC curves with red lines above the blue straight reference line indicate discrimination; the further above the reference line, the better. The area under the ROC curve is equivalent to the concordance probability [38]. Figure 2 shows an excellent level of discrimination, according to Hosmer et al [39], for VADER's ability to classify manually coded negative comments (the area under the ROC curve was 0.800, 95% CI 0.78-0.82), followed by T2D and LIWC tone. T2D and LIWC tone demonstrated acceptable discrimination, with the areas under the curve of 0.770 and 0.747, respectively.

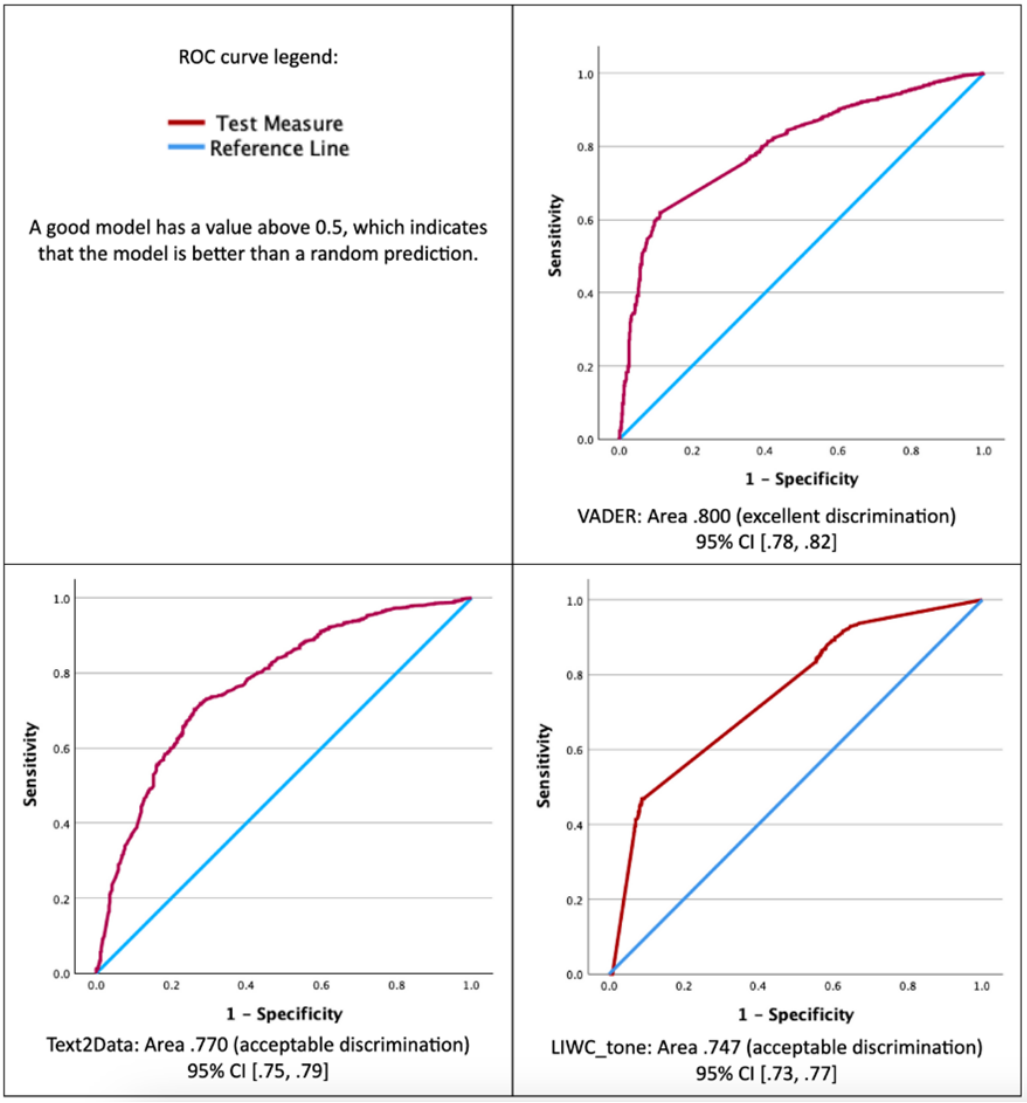
LIWC tone, according to Figure 4, performs the same or worse across the entire range of true positive rates (sensitivity), except at the short stretch of higher sensitivity rates where it is superior to T2D. Overall, VADER performs better than the 2 other

XSL•FO
RenderX

sentiment analysis systems. Next, we tested if VADER performed better for longer strings of text. Figure 4 compares VADER results or all comments, regardless of length, and for longer comments (>100 characters and >200 characters long). VADER indeed performed better when short comments were excluded. Confusion matrices are given in Tables 1-4.

Table 5 shows 7 measures of model performance compared with manual coding that are Cohen , accuracy, specificity, precision, sensitivity (recall), F_1-score, and MCC.

Figure 3. Receiver operating characteristic analyses: a comparison of Valence Aware Dictionary for Sentiment Reasoning, TEXT2DATA, and Linguistic Inquiry and Word Count tone scores' ability to classify comments that were manually coded as either negative (–1) or positive (1).



XSL•FO
RenderX

Figure 4. Receiver operating characteristic curve comparisons for Valence Aware Dictionary for Sentiment Reasoning, TEXT2DATA, and Linguistic Inquiry and Word Count Tone.

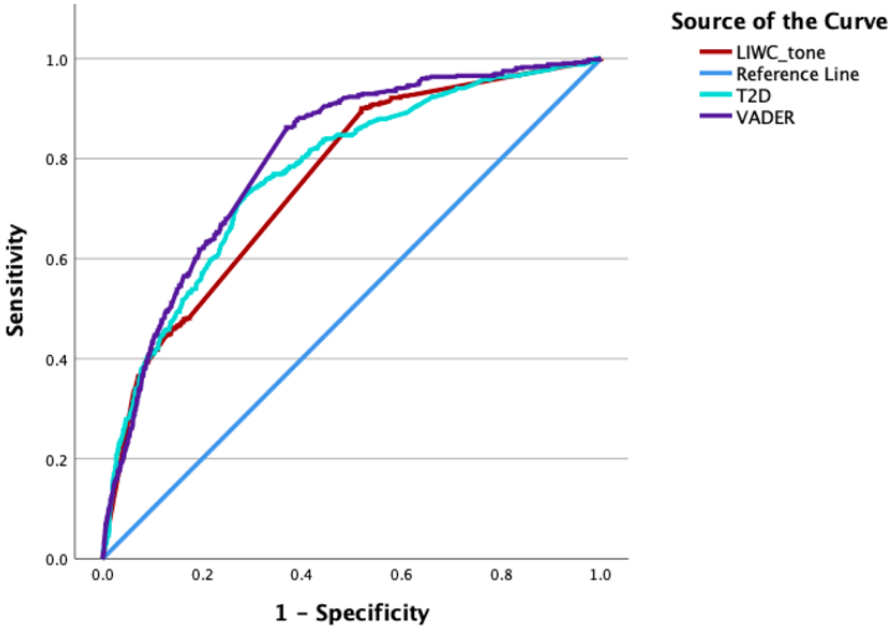


Table 1. A confusion matrix for binary coded Valence Aware Dictionary for Sentiment Reasoning, which are negative (–1) and positive (1) sentiments.

VADER[a]	Manual coding, n		Total, N
	–1	1	
–1	3693	230	3923
1	1178	432	1610
Total	3923	662	5533

[a]VADER: Valence Aware Dictionary for Sentiment Reasoning.

Table 2. A confusion matrix for binary coded t2d: negative (–1) and positive (1) sentiment.

T2D[a]	Manual coding, n		Total, n
	–1	1	
–1.00	3218	145	3363
1.00	1277	379	1656
Total	4495	524	5019

[a]T2D: TEXT2DATA.

Table 3. A confusion matrix for binary coded LIWC_tone: negative (–1) and positive (1) sentiment.

LIWC_tone[a]	Manual coding, n		Total, n
	–1	1	
–1	4214	373	4515
1	729	289	1018
Total	4871	662	5533

[a]LIWC: Linguistic Inquiry and Word Count.

XSL•FO
RenderX

Table 4. A confusion matrix for binary coded ChatGPT 4.0: negative (–1) and positive (1) sentiments.

ChatGPT 4.0	Manual coding, n, negative (–1); positive (1)		Total, n
	–1	1	
–1	4814	610	5424
1	57	52	109
Total	4871	662	5533

Table 5. A comparison of Valence Aware Dictionary for Sentiment Reasoning, TEXT2DATA, Linguistic Inquiry and Word Count tone, and ChatGPT 4.0 against manual coding.

Model	Cohen κ (95% CI; *P* value)	Accuracy	Specificity	Precision	Sensitivity (recall)	F_1-score	MCC[a]
VADER[b]	0.254 (95% CI 0.23- 0.28; <.001)	0.74	0.94	0.26	0.65	0.38	0.29
T2D[c]	0.225 (95% CI (0.2-.25; <.001)	0.72	0.96	0.23	0.72	0.35	0.3
LIWC[d]_tone	0.233 (95% CI 0.20-0.26; <.001)	0.80	0.92	0.28	0.47	0.34	0.24
ChatGPT 4.0	0.105 (95% CI 0.1-0.31; <.001)	0.88	0.88	0.50	0.08	0.13	0.16

[a]MCC: Matthews correlation coefficient.

[b]VADER: Valence Aware Dictionary for Sentiment Reasoning.

[c]T2D: TEXT2DATA.

[d]LIWC: Linguistic Inquiry and Word Count.

First, Cohen is used to examine the classification of YouTube comments into positive and negative sentiments. κ Values show fair agreement of manual coding with VADER (κ=0.254, 95% CI 0.23-0.28; *P*<.001), LIWC tone (κ=0.233, 95% CI 0.20-0.26; *P*<.001), and T2D (κ=0.26, 95% CI 0.20-0.25; *P*<.001). Overall, NLP demonstrated better agreement than LLM (κ=0.105, 95% CI 0.1-0.31; *P*<.001) for ChatGPT 4.0 indicating especially poor agreement of manual coding [40].

Second, ChatGPT 4.0 achieved the highest accuracy score of 88%, followed by LIWC tone (80%), VADER (74%), and T2D (72%). Third, the proportion of accurately detected negatives or specificity varied from 96% for T2D to 94% for VADER, 92% for LIWC tone, and 88% for ChatGPT 4.0. Fourth, precision reveals how well a model accurately makes positive predictions [41]. Precision values indicate that positive predictions were unlikely to be very accurate: ChatGPT 4.0 was at 50%, as compared with even lower values for LIWC tone (28%), VADER (26%), and T2D (23%).

Fifth, sensitivity, recognized as recall, detects true positives within a confusion matrix, with T2D achieving 72%, VADER achieving 65%, LIWC tone reaching 47%, and ChatGPT 4.0 achieving only 8% [41]. Sixth, the F_1-score is the harmonic mean of precision and sensitivity (recall), offering an evaluation of true positives and positive predictive values was below 50% for all models: VADER (at 0.38) was closely followed byT2D (0.35), LIWC tone (0.34), whereas ChatGPT's F_1-score stood out as the lowest (0.13). Finally, MCC was calculated by measuring the sensitivity, specificity, precision, and negative predictive value to evaluate the model performance [36]. None of the MCCs reported in Table 5 reached a strong correlation level; they varied from 0.30 (T2D) to 0.29 (VADER), 0.24 (LIWC tone), and the lowest value of 0.16 for ChatGPT 4.0.

Large discrepancies observed between the automated classifications and human coding may be explained by specific linguistic features, contextual nuances, model limitations, and other factors. To uncover any underlying patterns contributing to misclassifications and determine whether specific errors are more prevalent in 1 model over another, we analyzed comments classified as false positives and false negatives by LLM and NLP models.

Table 6 shows how sentiment misclassification varied by model. None of the models stood out as being excellent at avoiding both false positives and false negatives; however, ChatGPT 4.0 performed especially poorly. It classified even the most obviously positive comments as negative sentiment, leading to a very high rate (92%) of false negatives.

An analysis of false positives did not produce model-specific patterns. Sarcastic comments expressing feigned empathy for drug users were frequently misidentified by all models as false positives, for example, when the commenters talked about people dying but also added phrases such as "*White addiction, i truly love it. Now lets see what happens. Lmbao!*" [comment ID: 756]. Other common false positives involved discussions about marijuana as a gateway drug, criticisms of both Democrats and Republicans, racist remarks, and the attribution of blame to drug users.

Across all 4 models, discussions about the legalization of marijuana were misclassified as false negatives, as illustrated by this comment,

> *The opioid crisis was huge here in Florida. Look at what happened to it when they legalized marijuana for medical purposes, opioid deaths dropped! Look at the states that have it as recreational, opioid use*

XSL•FO
RenderX

and crime drop 20% in just the first year! [comment ID: 3795]

Our examination of false negatives also revealed model-specific, thematic patterns. VADER tended to misclassify political posts and comments on drug legalization as false negatives. T2D had difficulties with mentions of safe injection sites and support for Donald Trump. LIWC tone was prone to misclassifying short comments, those with emojis, as well as references to kratom, CBD (Cannabidiol) oil, and marijuana.

Finally, Table 7 summarizes the observed differences between sentiment analysis methods applied to our unbalanced social media dataset.

As shown in Table 7, T2D operates as a black box with minimal technical documentation, posing challenges for academic analysis. Unlike the other methods, VADER requires a moderate level of programming skill. All methods have low or no monetary cost (VADER is free) except for T2D, which charges per transaction, potentially making it cost-prohibitive for large datasets. Based on our findings, Table 7 provides other important considerations when selecting a model.

Table 6. Misclassified comments by model: false positives, false negatives, and representative comments.

Method, compared with manual coding	False positives, %	False negatives, %	Representative comments
VADER[a]	30.02	34.74	FP[b]: *Doctor's have become legalized dope dealers, they tried to get my 17 nephew to take opioids. They gave him a 2 month supply. Luckily he didn't take it because he'd seen what it can do. He's still playing football no thanks to the doc and big pharma* [comment ID: 70]; FN[c]: *Yes! It [kratom] got me off pain meds... pain medication* [comment ID: 2470]
T2D[d]	28.40	27.67	FP: *Big Phama! Big Insurance! Doctors get a cut for each pill script filled! We're worth more dead, than alive*! [comment ID 71]; FN: *Wish this war on opioids started earlier so many people gone. Very grateful for President Trump* [comment ID: 1483]
LIWC[e]_tone	14.97	56.34	FP: *Sue the opioid companies? What like the CIA[f]? LMFAO #CNNISFAKENEWS* [comment ID: 5872]; FN: K*ratom is the best maintenance program you could ever discover. Was on oxy for years, then methadone for years, then I found Kratom. It helps with pain, cravings, and has a bonus effect of reducing the craving for alcohol...* [comment ID: 2478]
ChatGPT 4.0	1.17	92.14	FP: *This isn't helping anyone. The medical industry is already terrified to prescribe these drugs, you people who don't take opioids have no idea of what damage you are doing to your medical system until you need to use it…* [comment ID: 909]; FN: *I have used MJ In the past to get off of Opioids and Xanax and Vicodin.. Yes, it does work and it also helps with Pain very well. It should be LEGAL even for NonMedical Use! It helps MANY problems and Various Ailments!* [comment ID: 3071]

[a]VADER: Valence Aware Dictionary for Sentiment Reasoning.

[b]FP: false positive.

[c]FN: false negative.

[d]T2D: TEXT2DATA.

[e]LIWC: Linguistic Inquiry and Word Count.

[f]CIA: Central Intelligence Agency.

XSL•FO
RenderX

Table 7. Considerations for selecting sentiment analysis methods when using social media datasets that are unbalanced toward negatives.

Method	Implementation type	Programming skill	Cost	Additional cons	Additional pros
Manual coding	Coding by a trained researcher	None	Time for creating codebook, manual coding	Retraining for intercoder reliability	Reaches good agreement when 2 humans code comments, accurate coding of ambiguous comments (sarcasm, etc)
VADER[a]	Rule-based dictionary	Moderate	Free	Not user-friendly for beginners, may code sarcastic comments as false positives.	Low runtime, fair agreement with manual codes, excellent discrimination, and performance can be improved by excluding short comments (<100 characters)
T2D[b]	Black box	None	Paid API[c]: Free up to 1000 transactions, US $27/month for 10,000 transactions.	Might not code all data, implementation is not defined (black box), may code sarcastic comments as false positives.	Low learning curve, self-contained within the same spreadsheet, high runtime, fair agreement with manual codes, acceptable discrimination
LIWC[d]_tone	Rule-based dictionary across multiple LIWC dimensions	None	Academic license: US $55 (1-year license) US $129 (3-year license)	May misclassify sarcastic comments as false positives and short comments as false negatives.	Includes a contextualizer that highlights words in reported dimensions, low learning curve, low runtime, accurate estimation of prevalence of negative versus positive sentiment, fair agreement with manual codes, acceptable discrimination
ChatGPT 4.0	Large language model	None	US $20/month	Requires prompt design, might not be consistent between iterations, may not be responsive due to high API usage, poor agreement with manual codes, inaccurate estimation of prevalence of negative vs positive sentiment, low MCC[e] may code sarcastic comments as false positives, hallucination possible.	Low runtime through the OpenAI API can provide reasoning for classification

[a]VADER: Valence Aware Dictionary for Sentiment Reasoning.

[b]T2D: TEXT2DATA.

[c]API: application programming interface.

[d]LIWC: Linguistic Inquiry and Word Count.

[e]MCC: Matthews correlation coefficient.

Discussion

Principal Findings

This study involved a comparison of manual sentiment coding to 4 automated sentiment analysis methods, namely VADER, T2D, LIWC_tone, and ChatGPT 4.0. We aimed to assess the efficacy of these sentiment analysis techniques in categorizing comments as either positive or negative sentiment in YouTube comments.

YouTube and other social media platforms are valuable repositories of comments and reviews on topics relevant to various organizations and stakeholders, such as businesses, public policy analysis, and politicians. Our corpus, comments on the US opioid crisis, was manually analyzed to reveal the struggles of opioid epidemic victims, their families, and communities, the issues of value to health policy analysis [33]. Like many other datasets of interest to social media researchers, it was skewed toward negative sentiment (7:1) and contained relatively few positive comments. Social media discussions often lean in the negative direction, such as Facebook posts on vaccine hesitancy [42] or long Covid discussions on Twitter (rebranded as X) [43]. Moreover, identifying social media content with negative effects is particularly valuable because it is more likely to be shared than positive content [44,45].

Our analysis showed that accurate positive sentiment classification can be important and policy-relevant when applied to negatively skewed data. The discussion of the opioid epidemic was primarily about human suffering but it was not all negative. Some YouTube users, for example, praised kratom as a means of overcoming opioid addiction. A subset of comments highlighted positive experiences of overcoming addiction and mixed reactions to health policies, such as safe injection sites and marijuana legalization. Marijuana, for example, can be discussed as a gateway drug (negative sentiment) and also from a harm reduction perspective as a substitute for opioids and

XSL•FO
RenderX

other harmful street drugs (positive sentiment). Classifying the data by sentiment enables researchers to explore diverse perspectives of the digital public, potentially leading to health policy insights. Given the importance of sentiment analysis in unbalanced datasets, this study offers valuable guidance for social media researchers on the pros and cons of several available methods.

Overall, VADER performed best on the ROC curve analysis, demonstrating excellent discriminatory capabilities compared with LIWC tone (acceptable) and T2D (acceptable). VADER's performance improved when short comments were excluded, a finding that verified VADER's sensitivity to text length suggested by Nair et al [18] and Tymann et al [19]. However, the ROC analysis has limitations when applied to data with an unbalanced count of negative and positive comments. Also, it could not be computed for binary ChatGPT 4.0 data. LIWC tone performed better than other automated models when estimating the prevalence of negative and positive sentiment in our data.

A comparison of Cohen κ values indicated that the NLP models (VADER, followed by LIWC and T2D) showed only fair agreement of manual coding, whereas ChatGPT 4.0 had poor agreement. While all models performed better than chance when predicting the dominant class, leading to higher precision, their level of agreement with manual coding is not exceptionally high. Moreover, variations in accuracy, specificity, precision, and sensitivity (recall), F_1-score, and MCC did not suggest a single superior model. All models evaluated had relatively low F_1–scores, which serve as overall prediction performance measures that combine precision and recall. F_1-score was below 50% for all automated models and especially low for ChatGPT 4.0. The same pattern was observed for MCC: none of the models resulted in a significant correlation (T2D's value was the highest at 0.30), but ChatGPT 4.0 performed the worst. While we cannot endorse a particular best model, our analysis suggests caution when using ChatGPT 4.0 to classify sentiment in an unbalanced dataset.

A total of 5 sets of instructions were uploaded to train ChatGPT 4.0 to conduct sentiment analysis, incorporating the codebook used by manual coders. The file upload process could have been more convenient, and 2 out of 3 attempts resulted in errors during file reading. LLM sentiment analysis was only accessible through a paid subscription, as the free version could not handle uploading and coding thousands of comments or formatting the results for further statistical analysis and comparison. Despite this, ChatGPT 4.0, like NLP models, seemed to rely on identifying words labeled as positive or negative in the codebook to classify comments, which might have led to misclassifications of sarcastic comments.

ChatGPT 4.0 outperformed the NLP models as an LLM in only 2 measures: accuracy and precision. It required specific coding instructions expressed as prompts, which we derived from the manually created codebook. Even though the same codebook was used for manual coding and designing ChatGPT 4.0 prompts, their level of agreement could have been better. Considering generation failures, poor agreement with manual coding, and the need for a paid subscription, ChatGPT 4.0 may be different from the model of choice for social media researchers looking to perform sentiment analysis on unbalanced datasets. Our comparison of ChatGPT 4.0 to 3 NLP models indicated that the general-purpose LLM has yet to surpass the performance of traditional NLP models, at least for unbalanced datasets with highly prevalent (7:1) negative comments.

Across all models, false negatives were associated with discussions on the legalization of marijuana and the observed reduction in mortality in states with more permissive drug policies. The primary issue with NLP models may stem from their reliance on pre-existing dictionaries to classify sentiment in a way that is not target specific. They cannot interpret the nuances of certain words within specific contexts the way a human can. To mitigate this, LIWC-22 and similar NLP models may require the creation of tailored dictionaries to better grasp the particular meanings of words in relevant contexts. Even then, NLP models may never be able to differentiate between negative statements of fact and negative sentiment with a specific target.

Not only NLP models but perhaps also LLM, tended to link any drug-related vocabulary with negative sentiment, failing to consider the context or nuance, particularly in discussions about overcoming the opioid crisis. This highlights a significant limitation of dictionary-based NLP models: their inability to accurately classify positive comments or recognize positive aspects of complex, contentious discussions compared with the accuracy of manual coding. Misclassification was observed for comments with sarcasm, leading models to mistake feigned empathy for genuine concern. Sentiment classification is complex and requires a deep understanding of the issues to interpret social media discourse accurately.

Manual coding remains the most reliable method for detecting sentiments when analyzing complex topics on social media, especially for unbalanced datasets. In addition to being time-intensive, it has other limitations. According to Krippendorff [46], texts often have multiple meanings. Manual coding is an interpretive process that may only sometimes match the commenter's intent, even when there is a good interrater agreement. On the other hand, automated sentiment classification may not consistently align with human judgments due to sentiment's inherently subjective and context-dependent nature, lowering reliability [47]. Fair reliability can be considered appropriate in cases involving complex or subjective tasks, exploratory research, resource constraints, qualitative analysis, or research contexts where a high level of agreement is not a primary objective [47]. Researchers must carefully weigh the trade-offs between achieving higher reliability and the practical constraints specific to their research.

Future research should explore whether general-purpose and fine-tuned LLMs and NLP models demonstrate comparable discriminatory performance in social media samples where positive comments are more prevalent than negative ones. Researchers should also test different methods for prompting general-purpose LLMs to follow complex manual codebook instructions.

XSL•FO
RenderX

Conclusions and Recommendations

We offer suggestions for VADER, T2D, LIWC tone, and ChatGPT 4.0 applications in the semantic classification of social media, specific to an unbalanced dataset with a high prevalence of negative comments. None of the automated models emerged as a clear leader. With caution, we recommend a no-cost tool, VADER, due to its excellent discrimination, according to our ROC curve analysis, which improves when the comments are at least 100 characters long. VADER requires some programming skills and may underestimate the prevalence of negative comments in unbalanced datasets. LIWC tone may be useful for social media researchers studying negative emotions, public worries, or dissatisfaction when they need to accurately estimate the prevalence of positive versus negative comments in their datasets. Researchers using T2D must know that it may only score some data and, compared with other NLP methods, can be time-consuming and cost prohibitive. ChatGPT 4.0 did not demonstrate superior performance. While the use of general-purpose LLMs is promising, it remains to be determined how to translate manual codebook instructions into prompts best to achieve superior classification results.

Data Availability

A codebook with representative comments has been added through Multimedia Appendix 1, as well as a copy of the ChatGPT 4.0 prompts used in sentiment analysis, and a list of videos that evoked YouTube comments.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Representative code book and the prompt used for ChatGPT sentiment analysis. A list of videos from which YouTube comments were extracted are also included.
[PDF File (Adobe PDF File), 228 KB - formative_v9i1e57395_app1.pdf]

References

1. Social Media Fact Sheet. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2024-10-02]
2. Ortiz-Ospina E. The rise of social media. Our World in Data. URL: https://ourworldindata.org/rise-of-social-media [accessed 2024-10-02]
3. Khan ML, Malik A. Researching Youtube: Methods, Tools, and Analytics. In: The Sage Handbook of Social Media Research Methods. Los Angeles, Washington DC: Sage Publications; 2022:651-663.
4. Lyu S, Ren X, Du Y, Zhao N. Detecting depression of Chinese microblog users text analysis: Combining Linguistic Inquiry Word Count (LIWC) with culture and suicide related lexicons. Front Psychiatry 2023;14:1121583 [FREE Full text] [doi: 10.3389/fpsyt.2023.1121583] [Medline: 36846219]
5. Liu JX, Leu JS, Holst S. Stock price movement prediction based on Stocktwits investor sentiment using FinBERT and ensemble SVM. PeerJ Comput Sci 2023;9:e1403 [FREE Full text] [doi: 10.7717/peerj-cs.1403] [Medline: 37346695]
6. Xu F, Pan Z, Xia R. E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework. Information Processing & Management 2020;57(5):102221 [FREE Full text] [doi: 10.1016/j.ipm.2020.102221]
7. Subramanian M, Easwaramoorthy Sathiskumar V, Deepalakshmi G, Cho J, Manikandan G. A survey on hate speech detection and sentiment analysis using machine learning and deep learning models. Alexandria Engineering Journal 2023;80:110-121 [FREE Full text] [doi: 10.1016/j.aej.2023.08.038]
8. Appel O, Chiclana F, Carter J, Fujita H. A hybrid approach to the sentiment analysis problem at the sentence level. Knowledge-Based Systems 2016;108:110-124 [FREE Full text] [doi: 10.1016/j.knosys.2016.05.040]
9. Hutto C, Gilbert E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. ICWSM 2014;8(1):216-225 [FREE Full text] [doi: 10.1609/icwsm.v8i1.14550]
10. Alaei AR, Becken S, Stantic B. Sentiment analysis in tourism: capitalizing on big data. J Travel Res 2019;58:175-191 [FREE Full text] [doi: 10.1177/0047287517747753]
11. Ribeiro FN, Araújo M, Gonçalves P, André Gonçalves M, Benevenuto F. SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods. EPJ Data Sci 2016;5(1):1-29 [FREE Full text] [doi: 10.1140/epjds/s13688-016-0085-1]
12. Botchway RK, Jibril AB, Kwarteng MA, Chovancova M, Oplatková ZK. A review of social media posts from UniCredit bank in Europe: a sentiment analysis approach. 2019 Presented at: Proceedings of the 3rd International Conference on Business and Information Management; 2020 July 17th-19th; Ireland. [doi: 10.1145/3361785.3361814]
13. Fasha EF, Keikhosrokiani P, Asl MP. Opinion mining using sentiment analysis: a case study of readers' response on Long Litt Woon's The Way Through the Woods in Goodreads. In: Advances on Intelligent Informatics and Computing. Cham: Springer International Publishing; 2021.

XSL•FO
RenderX

14. Bestvater SE, Monroe BL. Sentiment is not stance: target-aware opinion classification for political text analysis. Polit Anal 2022;31(2):235-256 [FREE Full text] [doi: 10.1017/pan.2022.10]
15. Nithyanand R, Schaffner B, Gill P. Online political discourse in the Trump era. ArXiv Preprint posted online on November 17, 2017 [FREE Full text] [doi: 10.5860/choice.45-0602]
16. Loveys K, Crutchley P, Wyatt EA, Coppersmith G. Small but mighty: affective micropatterns for quantifying mental health from social media language. 2017 Presented at: Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology - from Linguistic Signal to Clinical Reality; 2016 June 16; California, USA URL: https://doi.org/10.18653/v1/w17-3110 [doi: 10.18653/v1/w17-3110]
17. Valdez D, Ten Thij M, Bathina K, Rutter LA, Bollen J. Social media insights into US mental health during the COVID-19 pandemic: longitudinal analysis of Twitter data. J Med Internet Res 2020;22(12):e21418 [FREE Full text] [doi: 10.2196/21418] [Medline: 33284783]
18. Nair AJ, Veena G, Vinayak A. Comparative study of twitter sentiment on covid-19 tweets. : IEEE; 2021 Presented at: Proceedings of 5th International Conference on Computing Methodologies and Communication; 2021 April 08-10; Erode, India URL: https://doi.org/10.1109/ICCMC51019.2021.9418320 [doi: 10.1109/iccmc51019.2021.9418320]
19. Tymann K, Lutz M, Palsbröker P, Gips C. GerVADER-A German Adaptation of the VADER Sentiment Analysis tool for Social Media Texts. 2021 Presented at: Lernen, Wissen, Daten, Analysen; 2024 September 23-25; Germany p. 178-189.
20. TEXT2DATA. URL: https://text2data.com/sentiment-analysis-api [accessed 2024-10-02]
21. Gao S, Hao J, Fu Y. The application and comparison of web services for sentiment analysis in tourism. 2015 Presented at: Proceedings of the 12th International Conference on Service Systems and Service Management (ICSSSM); 2015 June 22-24; Guangzhou p. 1-6 URL: https://doi.org/10.1109/ICSSSM.2015.7170341 [doi: 10.1109/icsssm.2015.7170341]
22. Ramesh D. The correlation between Twitter sentiments and polling results for the 2016 Presidential race. South Carolina Junior Academy of Science. 2016. URL: https://scholarexchange.furman.edu/scjas/2016/all/273 [accessed 2024-10-02]
23. Tan B, Lim S, Qi Y, Miao C. A comparison of web services for sentiment analysis in digital mental health interventions. : Springer International Publishing; 2022 Presented at: Proceedings of International Conference on Human-Computer Interaction; 2023 July 23–28; Copenhagen, Denmark p. 389-407. [doi: 10.1007/978-3-031-05061-9_28]
24. Boyd RL, Ashokkumar A, Seraj S, Pennebaker JW. The development and psychometric properties of LIWC-22. The University of Texas at Austin. 2022. URL: https://www.liwc.app [accessed 2024-10-02]
25. Sell J, Farreras IG. LIWC-ing at a century of introductory college textbooks: have the sentiments changed? Procedia Computer Science 2017;118:108-112 [FREE Full text] [doi: 10.1016/j.procs.2017.11.151]
26. Smirnova A, Laranetto H, Kolenda N. Ideology through sentiment analysis: a changing perspective on Russia and Islam in NYT. Discourse & Communication 2017;11(3):296-313 [FREE Full text] [doi: 10.1177/1750481317699347]
27. Olagunju T, Oyebode O, Orji R. Exploring key issues affecting African mobile eCommerce applications using sentiment and thematic analysis. IEEE Access 2020;8:114475-114486 [FREE Full text] [doi: 10.1109/access.2020.3000093]
28. Boukes M, van de Velde B, Araujo T, Vliegenthart R. What's the Tone? Easy doesn't do it: analyzing performance and agreement between off-the-shelf sentiment analysis tools. Commun Methods Meas 2019;14(2):83-104 [FREE Full text] [doi: 10.1080/19312458.2019.1671966]
29. Hartmann J, Huppertz J, Schamp C, Heitmann M. Comparing automated text classification methods. Int J Res Mark 2019;36(1):20-38 [FREE Full text] [doi: 10.1016/j.ijresmar.2018.09.009]
30. Liu Y, Han T, Ma S, Zhang J, Yang Y, Tian J, et al. Summary of ChatGPT-related research and perspective towards the future of large language models. Meta-Radiology 2023;1(2):100017 [FREE Full text] [doi: 10.1016/j.metrad.2023.100017]
31. Baskara R, Mukarto M. Exploring the implications of chatgpt for language learning in higher education. IJELTAL 2023;7:343-358 [FREE Full text]
32. Fuchs K. Exploring the opportunities and challenges of NLP models in higher education: is Chat GPT a blessing or a curse? Front Educ 2023;8:1166682 [FREE Full text] [doi: 10.3389/feduc.2023.1166682]
33. Bacon LL. Opioid epidemic: CNN and Fox News Youtube viewers' stories, claims, and solutions. URL: https://tinyurl.com/yemxye4f [accessed 2024-10-02]
34. Cohn MA, Mehl MR, Pennebaker JW. Linguistic markers of psychological change surrounding September 11, 2001. Psychol Sci 2004;15(10):687-693. [doi: 10.1111/j.0956-7976.2004.00741.x] [Medline: 15447640]
35. Chicco D, Jurman G. The matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. BioData Min 2023;16(1):4 [FREE Full text] [doi: 10.1186/s13040-023-00322-4] [Medline: 36800973]
36. Chicco D, Jurman G. The advantages of the matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 2020;21(1):6 [FREE Full text] [doi: 10.1186/s12864-019-6413-7] [Medline: 31898477]
37. Hilbe JM. Logistic Regression Models. New York, NY: Chapman and Hall/CRC; 2009.
38. Gonen M. Analyzing receiver operating characteristic curves with SAS. United States: SAS Institute; 2007:362.
39. Hosmer D, Lemeshow S, Sturdivant R. Applied Logistic Regression. Chicester, UK: Wiley; 2013.
40. Laerd Statistics. URL: https://statistics.laerd.com/premium/spss/ck/cohens-kappa-in-spss-7.php [accessed 2002-10-24]

XSL•FO
RenderX

41. Sharma DK, Chatterjee M, Kaur G, Vavilala S. Deep learning applications for disease diagnosis. Deep Learning for Medical Applications with Unique Data 2022;1:31-51 [FREE Full text] [doi: 10.1016/b978-0-12-824145-5.00005-8]
42. Aleksandric A, Anderson HI, Melcher S, Nilizadeh S, Wilson G. Spanish facebook posts as an indicator of COVID-19 vaccine hesitancy in Texas. Vaccines (Basel) 2022;10(10):1713 [FREE Full text] [doi: 10.3390/vaccines10101713] [Medline: 36298580]
43. Martin S, Chepo M, Deom N, Khalid AF, Vindrola-Padros C. #LongCOVID affects children too: a Twitter analysis of healthcare workers sentimentdiscourse about Long COVID in childrenyoung people in the UK. medRxiv Preprint posted online on July 25, 2022 [FREE Full text] [doi: 10.1101/2022.07.20.22277865]
44. Schöne JP, Garcia D, Parkinson B, Goldenberg A. Negative expressions are shared more on twitter for public figures than for ordinary users. PNAS Nexus 2023;2(7):pgad219. [doi: 10.1093/pnasnexus/pgad219] [Medline: 37457891]
45. Schöne J, Parkinson B, Goldenberg A. Negativity spreads more than positivity on twitter after both positive and negative political situations. Affect Sci 2020;2(4):379-390 [FREE Full text] [doi: 10.31234/osf.io/x9e7u]
46. Krippendorff K. Commentary: A dissenting view on so-called paradoxes of reliability coefficients. Annals of the International Communication Association 2016;36(1):481-499 [FREE Full text] [doi: 10.1080/23808985.2013.11679143]
47. Provoost S, Ruwaard J, van Breda W, Riper H, Bosse T. Validating automated sentiment analysis of online cognitive behavioral therapy patient texts: an exploratory study. Front Psychol 2019;10:1065 [FREE Full text] [doi: 10.3389/fpsyg.2019.01065] [Medline: 31156504]

Abbreviations

API: application programming interface
LIWC: Linguistic Inquiry and Word Count
LLM: large language model
MCC: Matthews correlation coefficient
NLP: natural language processing
ROC: receiver operating characteristic
T2D: TEXT2DATA
VADER: Valence Aware Dictionary and Sentiment Reasoning

Edited by A Mavragani; submitted 15.02.24; peer-reviewed by AR Ashraf, DB Margolin; comments to author 25.06.24; revised version received 31.08.24; accepted 20.09.24; published 08.01.25.

Please cite as:
Gandy LM, Ivanitskaya LV, Bacon LL, Bizri-Baryak R
Public Health Discussions on Social Media: Evaluating Automated Sentiment Analysis Methods
JMIR Form Res 2025;9:e57395
URL: https://formative.jmir.org/2025/1/e57395
doi:10.2196/57395
PMID:

©Lisa M Gandy, Lana V Ivanitskaya, Leeza L Bacon, Rodina Bizri-Baryak. Originally published in JMIR Formative Research (https://formative.jmir.org), 08.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Understanding Health-Related Discussions on Reddit: Development of a Topic Assignment Method and Exploratory Analysis

Garrett J Chan[1], MD; Mark Fung[2], MD, PhD; Jill Warrington[2], MD, PhD; Sarah A Nowak[2], PhD

[1]Departments of Pathology and Laboratory Medicine, University of California, San Francisco, San Francisco, CA, United States

[2]Larner College of Medicine, University of Vermont, Burlington, VT, United States

Corresponding Author:
Sarah A Nowak, PhD
Larner College of Medicine
University of Vermont
89 Beaumont Ave
Burlington, VT, 05405
United States
Phone: 1 802 656 0359
Email: sarah.nowak@med.uvm.edu

Abstract

Background: Social media has become a widely used way for people to share opinions about health care and medical topics. Social media data can be leveraged to understand patient concerns and provide insight into why patients may turn to the internet instead of the health care system for health advice.

Objective: This study aimed to develop a method to investigate Reddit posts discussing health-related conditions. Our goal was to characterize these topics and identify trends in these social media–based medical discussions.

Methods: Using an initial query, we collected 1 year of Reddit posts containing the phrases “get tested” and “get checked.” These posts were manually reviewed, and subreddits containing irrelevant posts were excluded from analysis. This selection of posts was manually read by the investigators to categorize posts into topics. A script was developed to automatically assign topics to additional posts based on keywords. Topic and keyword selections were refined based on manual review for more accurate topic assignment. Topic assignment was then performed on the entire 1-year Reddit dataset containing 347,130 posts. Related topics were grouped into broader medical disciplines. Analysis of the topic assignments was then conducted to assess condition and medical topic frequencies in medical condition–focused subreddits and general subreddits.

Results: We created an automated algorithm to assign medical topics to Reddit posts. By iterating through multiple rounds of topic assignment, we improved the accuracy of the algorithm. Ultimately, this algorithm created 82 topics sorted into 17 broader medical disciplines. Of all topics, sexually transmitted infections (STIs), eye disorders, anxiety, and pregnancy had the highest post frequency overall. STIs comprised 7.44% (5876/78,980) of posts, and anxiety comprised 5.43% (4289/78,980) of posts. A total of 34% (28/82) of the topics comprised 80% (63,184/78,980) of all posts. Of the medical disciplines, those with the most posts were psychiatry and mental health; genitourinary and reproductive health; infectious diseases; and endocrinology, nutrition, and metabolism. Psychiatry and mental health comprised 26.6% (21,009/78,980) of posts, and genitourinary and reproductive health comprised 13.6% (10,741/78,980) of posts. Overall, most posts were also classified under these 4 medical disciplines. During analysis, subreddits were also classified as *general* if they did not focus on a specific health issue and *topic-specific* if they discussed a specific medical issue. Topics that appeared most frequently in the top 5 in general subreddits included addiction and drug anxiety, attention-deficit/hyperactivity disorder, abuse, and STIs. In topic-specific subreddits, most posts were found to discuss the topic of that subreddit.

Conclusions: Certain health topics and medical disciplines are predominant on Reddit. These include topics such as STIs, eye disorders, anxiety, and pregnancy. Most posts were classified under the medical disciplines of psychiatry and mental health, as well as genitourinary and reproductive health.

(JMIR Form Res 2025;9:e55309) doi:10.2196/55309

XSL•FO
RenderX

KEYWORDS

digital health; internet; open data; social networking; social media

Introduction

Background

With over 3.8 billion people regularly sharing information about their lives on social media platforms [1], social media has become a ubiquitous part of daily life. In addition to building relationships with other people, social media helps users find information on disciplines that they are concerned about or interested in. Social media platforms have become a commonly used way to discuss problems and seek advice [2-4].

Given its prevalent use, social media content may provide insight into individual attitudes and opinions about health. The anonymous nature of certain social media platforms allows users to choose to discuss potentially embarrassing subjects without an expectation of privacy. Deidentified information can be used for research purposes to understand social behavior as it is impossible to associate a specific comment with a single person.

Analysis of social media content has other strengths compared to more traditional methods such as surveys. For example, surveys may be influenced by social desirability bias, and respondents may not be as open when discussing sensitive or potentially embarrassing conditions [5] such as drug use or sexually transmitted infections (STIs). Several studies have demonstrated that internet-based platforms can circumvent or mitigate this limitation. For example, as Google search history reflects an active process (seeking information) rather than stated beliefs, it may be more accurate than surveys for some stigmatized beliefs in particular [6]. People have a variety of motivations for using social media, including information seeking, information sharing, and expression of opinion [2]. While previous work has also shown that online content becomes more shared, or viral, when it invokes strong, activating emotions, either positive or negative [7], some studies have suggested that individuals may be more honest with sharing their thoughts online.

An advantage of the use of social media to learn about people's behaviors and sentiments is that we can obtain large volumes of data in a short period and user-generated content can be processed quickly [8,9]. Social media websites count among the most trafficked ones in the United States. Facebook is the 6th most popular website in the United States at the time of writing, Reddit is the 7th, Instagram is the 18th, and Twitter (subsequently rebranded X) is the 33rd. The variety of topics discussed on social media is wide-ranging, and these topics include information about people's health concerns as users can pose and answer questions about health-related conditions on internet forums.

Reddit has over 270 million monthly users [10,11] who can post first-person accounts of illness and interactions with the health care system and leave comments on other users' posts about their experiences with similar illnesses. At the time of writing, Reddit ranks 17th in global engagement [12]. In addition, 42.8% of Reddit site traffic comes from search engines rather than direct visits to the site. For comparison, only 12.4% of Twitter traffic comes from searches, which may mean that a greater proportion of visits to Reddit comes from internet users doing a general search for information rather than choosing to go to a specific website. Therefore, many internet users reach Reddit from search engines, making the site an important source of information to both Reddit users and nonusers.

While the spectrum of discussion about health on social media is broad, studies of sites such as Reddit, Twitter, and Facebook have been able to examine disease- and health concern–specific discussions. This includes detection of opioid addiction based on Twitter posts [13] and discussion of colorectal cancer, breast cancer, and diabetes through analysis of Twitter posts and Facebook groups [14]. One study also categorized user data from Twitter posts into 4 specific disease-related topics: *flu*, *depression*, *pregnancy*, and *eating disorders* [15]. Other studies on Reddit have sought to detect depression [16], detect suicidality among opioid users [17], and characterize weight loss discussions [18]. These works have demonstrated the ability to select specific topics from social media data and draw conclusions about how these topics are encountered in the social media sphere. Multiple studies have characterized language markers to evaluate discussions on medical topics, giving each medical topic a unique pattern of markers [19]. In all, each medical topic can be thought of as having a social media "gene" that conceptualizes it as having a recognizable identity on social media that can be tracked and analyzed [20].

The specific system architecture and features of different social media platforms make each more optimally suited for different types of studies (Table 1). For example, differences in information organization, post length, and rules for content creation can influence the analysis. The use of Reddit, for example, is well suited for evaluation of stigmatized topics and topic-based conversations. On Reddit, posts are organized into groups by topic, and users can interact through specific forums, called subreddits, to find discussions related to the subject they are interested in. A post's visibility in a subreddit is determined by user upvotes and downvotes. In addition, posts may be removed by moderators if they violate subreddit-specific community rules [21]. Therefore, subreddits are a dedicated space for people to address a topic with the expectation of pertinent discussion. Previous work has noted that the topic-based organization of Reddit content is appropriate for studying specific health topics in greater detail [22]. Furthermore, some subreddits are question and answer–based. On the subreddit *r/askdocs*, users can seek answers about medical topics from health professionals who have been verified by subreddit moderators. While these subreddits do not coalesce around a certain topic, they allow users to ask questions to content experts anonymously, a function difficult to find on other social media platforms.

XSL•FO
RenderX

Table 1. Comparison of select social media platforms by content, organization, level of anonymity, moderation, and number of articles on PubMed.

	Reddit	Twitter	Facebook
Content length	40,000-character limit	280-character limit	No limit
Content organization and sharing	Users search for posts of interest; posts organized in topic-based subreddits	On the basis of accounts that a user follows; users can search for common themes using hashtags	Often based on real-life connections
Anonymity	May be completely anonymous	May or may not be anonymous	Real-name culture encouraged
Moderation	User moderated; subreddits have content guidelines	Platform moderated	Platform moderated
PubMed articles, N	277	5769	5030

Despite discussion forums such as Reddit having features that distinguish them from platforms such as Twitter and Facebook in important ways, comparatively little research has been conducted investigating health beliefs and discussions on Reddit. A number of text analysis studies on social media platforms have been conducted—at the time of writing, there are 5769 articles on PubMed discussing Twitter and 5030 about Facebook. However, only 277 articles have been published discussing Reddit [23]. For an internet user searching for information on a specific subject, Reddit represents a useful starting point because of its topic-focused organization.

Objectives

We report a text analysis of health concerns and topics reflected in users' Reddit posts during 2019. Given the previous studies on using social media to study health and medicine, we wanted to continue this work while looking at medical topics on another platform. In addition to the relative lack of study compared to other social media platforms, we chose to investigate Reddit because it has features that differentiate it from heretofore more frequently studied social media platforms: specific, topic-based discussions; greater anonymity; and user moderators. We wanted to capture posts from a variety of health-relevant subreddits discussing seeking or receiving testing and diagnostic advice.

Methods

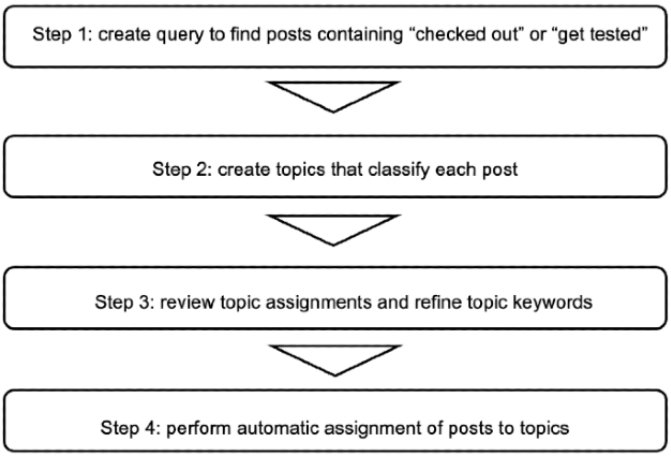
Ethical Considerations

The RAND Corporation was consulted regarding institutional review board approval. The RAND Human Subject Protection Committee (the RAND Institutional Review Board) determined that the project did not involve human participants as defined by the regulations of Title 45 of the Code of Federal Regulations Part 46.102(f) and, therefore, was not subject to further review. We did not explicitly obtain a waiver of informed consent; however, the determination of the absence of human participants implies a waiver of the need for documented informed consent. In addition, all data were anonymous, and obtaining informed consent would require identifying individuals, which would also be a violation of the Reddit terms of service.

Overview

We conducted our analysis in 4 steps (Figure 1) designed to balance false positives and false negatives. In step 1, we developed and refined a query to obtain posts about medical testing and advice about seeking diagnoses. In step 2, topics were created to classify the posts. Subreddits that clearly contained irrelevant posts were identified and excluded to improve the efficiency of the topic creation process. In step 3, posts assigned to the developed topics were reviewed, and the keywords used for topic assignment were refined. The process for identifying subreddits with relevant posts to be included in the analysis was refined as well. In step 4, keywords were used to do a final automatic assignment to a set of posts from a 12-month period. The 4 steps are shown in more detail in Figure 2 and are described in the following sections.

Figure 1. Overview of analysis steps—the overall methodology for gathering posts, creating topics, and automatically assigning posts to topics.

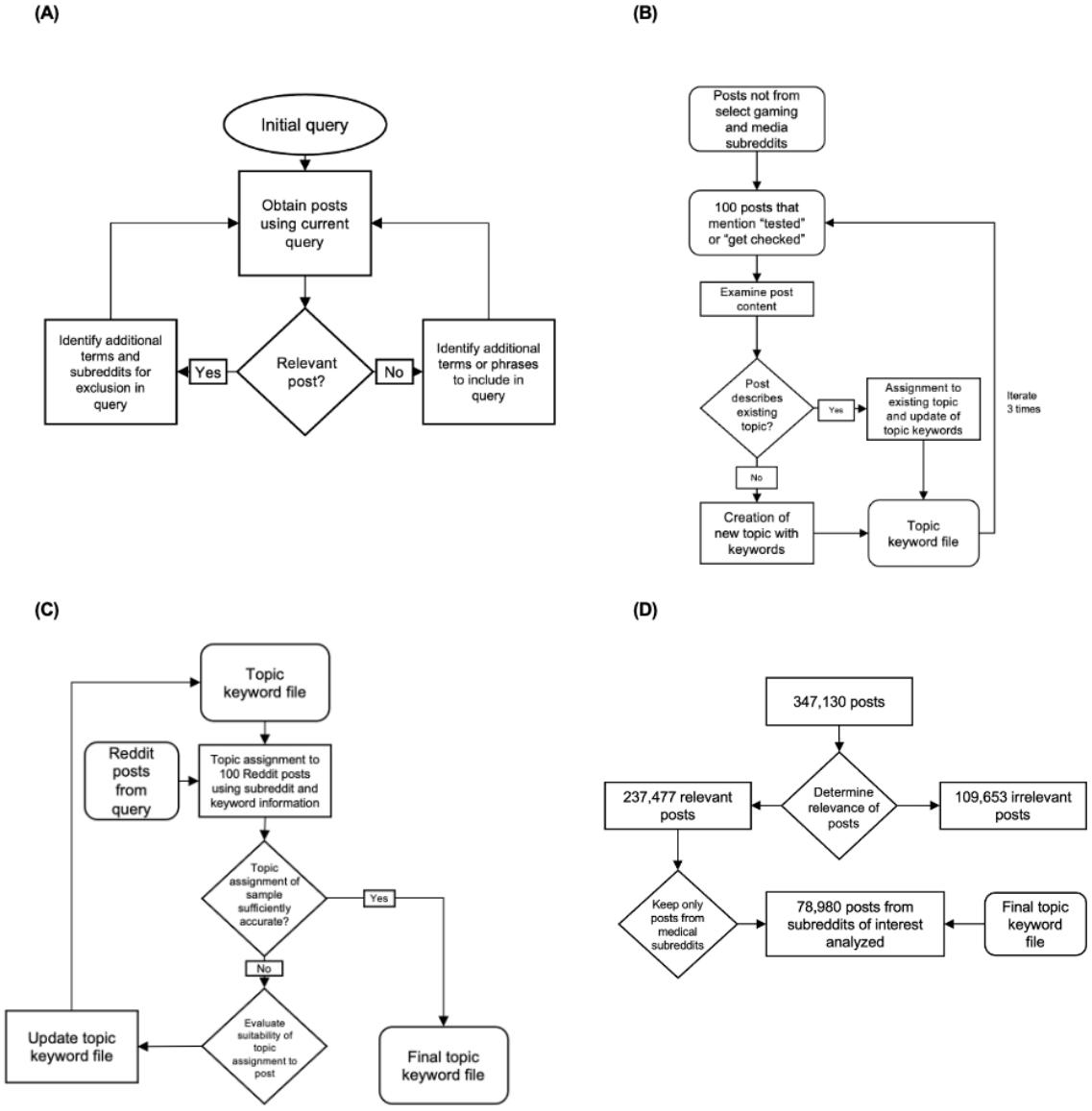


XSL•FO
RenderX

Step 1: Query Creation and Refinement

First, we captured posts from Reddit that were relevant to testing and diagnosis. We developed a query in Crimson Hexagon's Monitor application programming interface to obtain posts that contained variations of the phrases "getting tested" or "getting checked out" and excluded posts from subreddits that would be unlikely to be relevant to medicine or health care (Figure 2A and Table S1 in Multimedia Appendix 1). The query run was as follows: "((get OR got) AND ((test AND tested)OR(check OR checked)) AND for)~4" (Figure 2A). Posts gathered using this query included either the words "get" or "got" in conjunction with "test," "tested," "check," or "checked" appearing within 4 words of each other. The posts collected were created between January 1, 2019, and December 31, 2019, and included the text content of the post, its title, and its date. A *post* could be the body of text created by the original poster or a comment from a user on an original post. In both cases, the content of a post during our analysis included the title of the original post to provide additional context.

Figure 2. Detailed flow diagrams of analysis steps 1 to 4. (A) Step 1: selection of Reddit posts. Posts from January 1, 2019, to December 31, 2019, were considered in this study. (B) Step 2: identification of topic keywords. Topics were manually created from three 100-post samples, and keywords were assigned to each topic. (C) Step 3: topic assignment. Topics were manually assigned to additional 100-post samples based on the topic keyword file. (D) Step 4: automatic assignment of posts for analysis. The algorithm using the final topic keyword file was run on 78,980 posts chosen from the overall corpus of 347,130 posts.



To limit our false-positive rate, we first excluded subreddits tabulated on *r/listofsubreddits* (accessed November 18, 2019) that would be most irrelevant to health-related discussions (Table S1 in Multimedia Appendix 1). Posts not from these subreddits were included in the topic creation step.

Step 2: Topic Creation

In this step, we manually read through 100 posts and labeled each with a de novo topic (Figure 2B). Topic creation was guided by the existing taxonomy in the *International Classification of Diseases, 10th Revision* (*ICD-10*) [24]. Topics were sorted into medical disciplines to optimize the analysis by clustering topics into larger groups.

XSL•FO
RenderX

These disciplines corresponded to chapter titles in the *ICD-10*. We then designated keywords that corresponded to each of these topics. For example, for the topic of STIs, keywords included "antibiotics," "gonorrhea," "chlamydia," and "bacterial vaginosis." Additional keywords were chosen from the corresponding article topic on UpToDate (Wolters Kluwer N.V.) [25], a reference for health care providers covering the diagnosis, pathophysiology, and treatment of a broad spectrum of medical conditions. In addition, several topics related to the health care system but not to a disease or topic were created, such as medical billing. Finally, for posts that did not fit with any of the aforementioned topics, a *nonmedical* topic was devised. Posts that corresponded to an existing topic were assigned to that topic.

The topics and disciplines that we formulated corresponded to the diagnoses and chapters in the *ICD-10*. A list of topics with their corresponding category is available in Table S2 in Multimedia Appendix 1. This categorization system is similar to the categorization of diseases and topics that health care providers follow. Therefore, our topics and medical disciplines were validated by an existing and accepted hierarchy of disease classification. We chose to use the *ICD-10* over other controlled vocabulary systems such as the Medical Subject Headings thesaurus of the National Library of Medicine and the Logical Observation Identifiers Names and Codes database because the *ICD-10* is the system that most physicians are in contact with and is the most targeted at covering the diagnosis of diseases. In contrast, the Medical Subject Headings thesaurus is a broader system for indexing publications that encompasses anatomy, organisms, and drugs, among many other topics.

During this process, additional irrelevant subreddits and terms were excluded from the query. In total, 3 iterations of keyword selection, each performed on a 100-post set, were completed. This generated our initial topic keyword file.

Step 3: Topic Refinement and Determination of False-Positive Rate as Means of Validating Topic Assignment

To begin automatically assigning topics to posts (Figure 2C), we created a script in R (R Foundation for Statistical Computing) that first put each post into tidy format using the *tidyverse* package [26]. The script then assigned a single topic to each post based on words found in the post and on the keyword file. With further manual analyses of more posts, we added more keywords to each topic with the goal of improving future topic assignment in later iterations.

We read through 100-post samples to add additional keywords to topics. After each run, we manually validated the accuracy of the topic assignments and iterated through this review process for 15 runs. Our topic assignments were reviewed by SAN and GC. Keyword classifications were reviewed by SAN and GC, as well as by MF and JW, 2 clinical pathologists. Accuracy was determined by reading the title and content of the posts and determining whether the topic captured the essence of the post. Topic assignment was designated as either correct or incorrect. We maintained objectivity by blinding the topic reviewer to each post's topic as assigned by the script before manually assigning a topic to that post. An example of a post and its assigned topic is as follows:

> *POST: I find that when I go through bouts of anxiety the mornings are usually the hardest, no appetite, tired etc, but as the day goes on I find I feel better, even to the extent by evening time I feel "normal."*
>
> *ASSIGNED TOPIC: Anxiety*

We then determined how many posts' topics had been correctly assigned per run and calculated an accuracy rate. During this process, we improved our script by allowing multi-word keywords and lemmatizing keywords. In addition, more subreddits were excluded to improve accuracy (Table S3 in Multimedia Appendix 1).

In addition, our algorithm excluded irrelevant posts based on irrelevant terms and subreddits determined in step 2. During the manual review process, we reviewed which posts were determined to be irrelevant by the algorithm to determine the accuracy and efficacy of using those exclusion methods [27-29]. Then, in each iteration, we calculated the false-positive rate, defined as the number of relevant posts divided by the total number of posts subtracted from 1 [30].

Posts were defined as relevant to health care if the main subject of the post related to seeking a diagnosis for a medical condition, discussion of treatment and management of a medical condition, or medical billing. An example of such a post is as follows:

> *Licensed skincare practitioner here. Get it checked. SKs [seborrheic keratoses] can itch and flake. Also may be BCC.*

An example of an irrelevant post is as follows:

> *Him going around checking if his wife cleaned properly is not something that should be rationalised....*

Step 4: Analysis of Reddit Data From January 1, 2019, to December 31, 2019, and Subreddit Analysis

To restrict our final analysis to posts mostly likely to be relevant to health-related discussion, we further limited the posts in the final analysis to those from certain subreddits. We specifically selected the 77 health-related subreddits posted on *r/listofsubreddits* and manually read through these subreddits to check for relevance to health discussions. In addition, we chose 10 subreddits with the most posts from our data sample to gather a broad range of Reddit discussion and another 8 health-related subreddits not posted on *r/listofsubreddits* that we determined contained relevant posts. In total, we obtained 95 subreddits.

We used the topic keyword file to automatically assign topics and medical disciplines to the set of Reddit posts that matched the query described previously (Figure 2D). Of note, we found that only posts from 89% (85/95) of the subreddits were assigned to a topic, indicating that the other 10 subreddits did not contain any health-related discussion as determined by the algorithm. We manually read through 300 posts to determine the accuracy rate. The topic assignments then underwent further analysis to determine the distribution of topics and medical disciplines,

XSL•FO
RenderX

distribution of posts in subreddits, and distribution of topics in each subreddit. Finally, we selected quotes from our quantitative results to provide human-relatable insights that illustrated what was discussed in each discipline, as previously done in other social media–mining research [27,31,32].

Results

Step 1: Determination of Relevant Posts, Subreddit Limitation, and Query Refinement

In total, 347,130 posts from January 1, 2019, to December 31, 2019, were downloaded from Reddit. Of those 347,130 posts, 109,653 (31.59%) belonged to subreddits that were determined to be irrelevant, and 237,477 (68.41%) came from relevant subreddits. Irrelevant subreddits were found to be those pertaining to games, popular media, and technology (Table S1 in Multimedia Appendix 1).

In step 1 (Figure 2A), the exclusion of irrelevant subreddits from our query decreased the false-positive rate of irrelevant posts from 59% to 29% (calculated from a random sample of 100 posts). We did not attempt to further decrease the false-positive rate in this step to avoid overexcluding subreddits and posts in the manual review step.

Steps 2 and 3: Topic Creation and Validation

A total of 82 topics were created through the inductive workflow outlined previously and sorted into 17 medical disciplines. Each topic was assigned to only 1 medical discipline. We validated our choices of topics and medical disciplines by comparing these categorizations to the diseases and medical topics delineated in the *ICD-10*. Of the 82 topics we created, 80 (98%) had a corresponding topic in the *ICD-10*; the exceptions were the *medical billing* and *nonmedical* topics. The accuracy of the topic assignments in this step was 56.3% (SD 4.92%; SEM 4.62%).

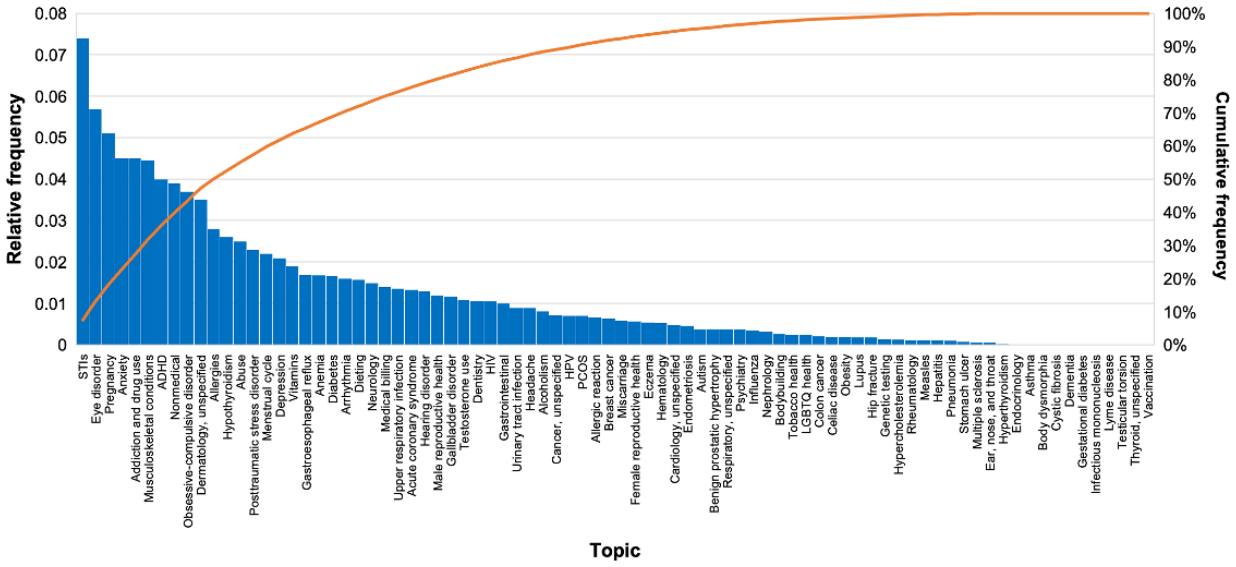
Step 4: Topics and Medical Disciplines

In step 4, subreddits to include in the final analysis were identified. While our objective in step 1 was to reduce the false-positive rate without being too restrictive (eliminating only subreddits least likely to contain medical discussion), our objective in this step was to keep only subreddits most likely to contain medical discussions. In total, 95 subreddits were chosen. These subreddits contained 78,980 posts.

Limiting posts to only relevant subreddits further decreased the false-positive rate from 29% to 15.5%. The accuracy rate in this step was 62.5% (SD 1.59%; SEM 0.793%), which was higher than in steps 1 and 3. Therefore, limiting posts to mainly medically relevant subreddits improved the relevance and accuracy compared to only excluding irrelevant subreddits.

The relative frequency of each topic was calculated (Figure 3). STIs, eye disorders, anxiety, and pregnancy had the highest frequency at 7.44% (5876/78,980), 5.6% (4423/78,980), 5.43% (4289/78,980), and 5.09% (4020/78,890) of all posts, respectively. Interestingly, 10 topics had a 0% relative frequency: asthma, body dysmorphic disorder, cystic fibrosis, dementia, gestational diabetes, infectious mononucleosis, Lyme disease, testicular cancer, thyroid (unspecified), and vaccination. Nonmedical posts made up 3.87% (3057/78,890) of all posts. If including the *nonmedical* topic, 80% (63,184/78,890) of posts were categorized within the first 34% (28/82) of topics. The mean relative frequency of all topics was 1.22% (SD 0.15%).

Figure 3. Relative frequency of all topics in descending order as determined by the topic assignment algorithm. The cumulative frequency of all topics is also shown. ADHD: attention-deficit/hyperactivity disorder; HPV: human papillomavirus; LGBTQ: lesbian, gay, bisexual, transgender, and queer; PCOS: polycystic ovary syndrome; PTSD: posttraumatic stress disorder; STI: sexually transmitted infection.



As 34% (28/82) of the topics comprised 80% (63,184/78,890) of posts, this points to a skewed distribution of topics. In total, 61% (17/28) of those topics belonged to the top 4 medical disciplines (psychiatry and mental health; genitourinary and reproductive health; infectious diseases; and endocrinology, nutrition, and metabolism).

Topics were grouped into 17 broader medical disciplines (Table 2) based on the categorization scheme of the chapter codes in

XSL•FO
RenderX

the *ICD-10*, as described previously. The most discussed disciplines were psychiatry and mental health; genitourinary (related to the urinary and reproductive systems) and reproductive health; infectious diseases; and endocrinology, nutrition, and metabolism, whose posts, respectively, constituted 26.6% (21,009/78,980), 13.6% (10,741/78,980), 8.95% (7069/78,980), and 8.91% (7037/78,980) of all analyzed posts. Psychiatry and mental health comprised most posts. Interestingly, most of the other medical disciplines were discussed with similar relative frequencies. The next most common medical discipline was hematology and immunology with a 5.88% (4644/78,890) relative frequency. The medical discipline with the lowest relative frequency was otolaryngology (related to medicine dealing with the ear, nose, and throat) at 0.57% (450/78,890).

Table 2. The relative frequency of the broader medical disciplines, into which the medical topics were sorted, shown in descending order.

Medical discipline	Relative frequency
Psychiatry and mental health	0.266
Genitourinary and reproductive health	0.136
Infectious diseases	0.0895
Endocrinology, nutrition, and metabolism	0.0891
Hematology and immunology	0.0588
Ophthalmology	0.0572
Musculoskeletal and connective tissue conditions	0.0476
Gastroenterology	0.0426
Dermatology	0.0410
Nonmedical	0.0387
Neurology	0.0373
Cardiovascular	0.0353
Pulmonology	0.0183
Medical; other	0.0160
Neoplasms	0.0157
Dentistry	0.0095
Otolaryngology	0.0022

The most common topics in the 4 most discussed medical disciplines are shown in Table 3. Within the medical discipline of psychiatry and mental health, the 3 most common topics were anxiety, addiction and drug use, and attention-deficit/hyperactivity disorder (ADHD) with relative frequencies of 20.42% (4289/21,009), 17.3% (3635/21,009), and 15.1% (3172/21,009), respectively. In the medical discipline of infectious diseases, the most common topic by a significant margin was STIs, making up 83.12% (5876/7069) of infectious disease posts. HIV and influenza made up 11.01% (778/7069) and 3.3% (233/7069) of posts, respectively. In genitourinary and reproductive health, pregnancy was the most common topic, comprising 37.43% (4020/10,741) of genitourinary and reproductive health posts, followed by menstrual cycle (1665/10,741, 15.5%) and male reproductive health (591/10,741, 5.5%). Finally, posts within the medical discipline of endocrinology, nutrition, and metabolism were more evenly distributed across topics. Hypothyroidism had the greatest number of posts (2013/7037, 28.61%), followed by vitamins (1534/7037, 21.8%) and diabetes (1316/7037, 18.7%).

XSL•FO
RenderX

Table 3. Relative frequency of topics on Reddit, as determined by our algorithm, within medical disciplines. The breakdown of the relative frequency of topics in the 4 most prevalent medical disciplines is as follows: psychiatry and mental health; genitourinary and reproductive health; infectious diseases; and endocrinology, nutrition, and metabolism. The medical disciplines are also shown in descending order by prevalence.

Medical discipline and topic	Relative frequency
Psychiatry and mental health	
Anxiety	0.203
Addiction and drug use	0.169
ADHD[a]	0.151
Obsessive-compulsive disorder	0.139
Abuse	0.094
PTSD[b]	0.087
Depression	0.079
Alcoholism	0.031
Autism	0.010
Psychiatry	0.0090
Tobacco use	0.0085
LGBTQ[c] health	0.0051
Body dysmorphic disorder	0.0015
Dementia	0.0013
Genitourinary and reproductive health	
Pregnancy	0.376
Menstrual cycle	0.160
Male reproductive health	0.088
Testosterone use	0.081
Urinary tract infection	0.068
HPV[d]	0.053
Miscarriage	0.044
Female reproductive health	0.042
Endometriosis	0.035
Benign prostatic hypertrophy	0.028
Nephrology; unspecified	0.024
Gestational diabetes	0
Testicular torsion	0
Infectious diseases	
STIs[e]	0.831
HIV	0.118
Influenza	0.039
Measles	0.013
Infectious mononucleosis	0
Lyme disease	0
Vaccination	0
Endocrinology, nutrition, and metabolism	
Hypothyroidism	0.286
Vitamins	0.216

XSL•FO
RenderX

Medical discipline and topic	Relative frequency
Diabetes	0.188
Dieting	0.178
PCOS[f]	0.078
Bodybuilding	0.029
Obesity	0.022
Hyperthyroidism	0.002
Endocrinology	0.001
Thyroid; unspecified	0

[a]ADHD: attention-deficit/hyperactivity disorder.

[b]PTSD: posttraumatic stress disorder.

[c]LGBTQ: lesbian, gay, bisexual, transgender, and queer.

[d]HPV: human papillomavirus.

[e]STI: sexually transmitted infection.

[f]PCOS: polycystic ovary syndrome.

Posts per Subreddit

Of the 95 subreddits identified in step 4, a total of 10 (11%) did not contain any posts that were assigned to our topics and were excluded, leaving 85 final subreddits. These 85 subreddits contained 78,980 posts (Table 4). The subreddits with the most posts were those that we describe as *general*, or non–topic-specific, subreddits. The top 10 subreddits with the most posts included 6 general subreddits: *r/askreddit*, *r/askdocs*, *r/amitheasshole*, *r/medical*, *r/legaladvice*, and *r/advice*. The most popular subreddit was *r/askreddit* with 26.7% (21,086/78,980) of posts, followed by *r/askdocs* with 12.02% (9496/78,980) of posts. A total of 61.93% (48,912/78,890) of the posts analyzed were found in the top 5 subreddits by number of posts, and 79.91% (63,112/78,890) were found in the top 15 subreddits.

XSL•FO
RenderX

Table 4. Subreddits by number of posts analyzed showing the relative differences in the number of posts in each subreddit (N=78,980).

Subreddit	Posts analyzed, n (%)
r/askreddit	21,086 (26.7)
r/askdocs	9496 (12.02)
r/amitheasshole	5473 (6.93)
r/relationship_advice	5322 (6.74)
r/relationships	2775 (3.51)
r/medical	2495 (3.16)
r/skincareaddiction	2268 (2.87)
r/legaladvice	2227 (2.82)
r/advice	2150 (2.72)
r/teenagers	1974 (2.5)
r/ADHD	1876 (2.38)
r/babybumps	1775 (2.25)
r/Testosterone	1746 (2.21)
r/STD	1345 (1.7)
r/Anxiety	1215 (1.54)
r/Depression	1193 (1.51)
r/pregnant	1166 (1.48)
r/birthcontrol	1102 (1.4)
r/childfree	966 (1.22)
r/Hypothyroidism	880 (1.11)
r/fitness	778 (0.99)
r/MentalHealth	741 (0.94)
r/Diabetes	699 (0.89)
r/Dentistry	602 (0.76)
r/science	489 (0.62)
r/diagnoseme	467 (0.59)
r/Infertility	446 (0.56)
r/SuicideWatch	429 (0.54)
r/Optometry	410 (0.52)
r/Ibs	391 (0.5)
r/CrohnsDisease	376 (0.48)
r/Cancer	349 (0.44)
r/Fibromyalgia	318 (0.4)
r/ChronicPain	300 (0.38)
r/MultipleSclerosis	279 (0.35)
r/Aspergers	272 (0.34)
r/bpd	235 (0.3)
r/SleepApnea	230 (0.29)
r/BipolarReddit	201 (0.25)
r/cfs	188 (0.24)
r/askdoctorsmeeee	175 (0.22)
r/askdoctors	150 (0.19)

XSL•FO
RenderX

Subreddit	Posts analyzed, n (%)
r/Rheumatoid	127 (0.16)
r/medicine	125 (0.16)
r/Eczema	120 (0.15)
r/Asthma	118 (0.15)
r/InterstitialCystitis	117 (0.15)
r/Nutrition	115 (0.15)
r/Dementia	86 (0.11)
r/askscience	84 (0.11)
r/Healthcare	80 (0.1)
r/StopSmoking	79 (0.1)
r/Dysautonomia	77 (0.1)
r/Psoriasis	75 (0.09)
r/Keratoconus	72 (0.09)
r/Pharmacy	59 (0.07)
r/Vegetarian	52 (0.07)
r/Massage	48 (0.06)
r/KidneyStones	47 (0.06)
r/GravesDisease	40 (0.05)
r/Menieres	32 (0.04)
r/Pilonidalcyst	31 (0.04)
r/Posture	30 (0.04)
r/Paleo	27 (0.03)
r/Gastroparesis	24 (0.03)
r/Hemophilia	22 (0.03)
r/Psychology	17 (0.02)
r/TroubledTeens	16 (0.02)
r/PancreaticCancer	13 (0.02)
r/Amblyopia	11 (0.01)
r/LactoseIntolerant	9 (0.01)
r/AlternativeHealth	8 (0.01)
r/NaturalBeauty	7 (0.01)
r/PublicHealth	6 (0.01)
r/Transplant	5 (0.01)
r/CareGiverSupport	3 (<0.01)
r/hepc	3 (<0.01)
r/chd	2 (<0.01)
r/Flu	2 (<0.01)
r/Juicing	1 (<0.01)
r/Pescetarian	1 (<0.01)
r/UKHealthcare	1 (<0.01)
r/Dystonia	1 (<0.01)
r/ItsNeverLupus	1 (<0.01)
r/MacularDegeneration	1 (<0.01)

XSL•FO
RenderX

General Subreddits Versus Topic-Specific Subreddits

Subreddits that we defined as *general* were those that did not focus on a specific health topic. These 15 subreddits were *r/askreddit*, *r/askdocs*, *r/amitheasshole*, *r/relationship_advice*, *r_relationships*, *r/medical*, *r/legaladvice*, *r/advice*, *r/teenagers*, *r/fitness*, *r/science*, *r/diagnoseme*, *r/askdoctorsmeeee*, *r/askdoctors*, and *r/medicine*. Compared to topic-specific subreddits, a wider variety of topics was represented in these subreddits, and no one topic dominated over other topics in each subreddit. We identified the top 5 topics by number of posts in each subreddit. Among the general subreddits, these topics comprised, on average, 41.1% of posts (SD 12.9%; 95% CI 28.2%-54%). For comparison, among topic-specific subreddits, the top 5 topics on average comprised 70.3% of posts (SD 38.3%; 95% CI 32%-100%).

In 67% (10/15) of the general subreddits, STIs was one of the top 5 topics (Table 5). Notably, the second most prevalent topic in *r/science* was human papillomavirus, whereas the most prevalent topic in *r/healthcare* (not one of the top 15 general subreddits) was medical billing. Interestingly, we also noted that the top 5 topics in *r/askreddit*, which is not specifically targeted toward health issues, were musculoskeletal conditions, eye disorders, ADHD, addiction and drug use, and allergies. For comparison, in *r/askdocs*, the most prevalent topics were STIs, dermatology, musculoskeletal conditions, allergies, and anxiety. Mental health topics were common in the general subreddits. Anxiety appeared as a top-5 topic in 53% (8/15) of the general subreddits. In those 8 subreddits, an average of 6.27% (SD 0.017%) of all posts were assigned to the topic of anxiety. Addiction and drug use appeared as a top-5 topic in 33% (5/15) of the general subreddits, and an average of 10.9% (SD 5.30%) of all posts in those 5 subreddits were assigned to that topic. On the other hand, we found that topic-specific subreddits often had the topic of the same name as the most assigned topic (eg, the topic of ADHD would be the most assigned topic in the subreddit *r/ADHD*; Table 6).

XSL•FO
RenderX

Table 5. The top 5 topics by number of posts in the top 15 general subreddits (subreddits without a specific health-related topic). A description of the subreddit and the total number of posts in the subreddit are provided (N=78,980).

Subreddit	Description	Posts, n (%)	Top 5 topics in subreddit
r/askreddit	Open question-and-answer forum	21,086 (26.7)	• Musculoskeletal conditions (n=1673 posts) • Eye disorder (n=1556 posts) • ADHD[a] (n=1043 posts) • Addiction and drug use (n=1001 posts) • Allergies (937 n=posts)
r/askdocs	Questions and answers with verified medical professionals	9496 (12.02)	• STIs[b] (n=748 posts) • Dermatology; unspecified (n=646 posts) • Musculoskeletal conditions (n=598 posts) • Allergies (n=465 posts) • Anxiety (n=435 posts)
r/amItheasshole	Users post stories to “find out if [they] were wrong”	5473 (6.93)	• STIs (n=637 posts) • Anxiety (n=461 posts) • Abuse (n=400 posts) • Nonmedical (n=353 posts) • Pregnancy (n=334 posts)
r/relationship_advice	Relationship-based discussions	5322 (6.74)	• STIs (n=1190 posts) • Abuse (n=530 posts) • Dating (n=497 posts) • Obsessive-compulsive disorder (n=374 posts) • Anxiety (n=358)
r/relationships	Interpersonal relationship–based discussion	2775 (3.51)	• STIs (n=420 posts) • Abuse (n=390 posts) • Obsessive-compulsive disorder (n=281 posts) • Dating (n=248 posts) • Anxiety (n=222 posts)
r/medical	Medical question–based discussion	2495 (3.16)	• Dermatology; unspecified (n=235 posts) • STIs (n=175 posts) • Musculoskeletal conditions (n=173 posts) • Allergies (n=151 posts) • Eye disorder (n=114 posts)
r/legaladvice	Forum to ask “simple legal questions”	2227 (2.82)	• Addiction and drug use (n=401 posts) • Medical billing (n=373 posts) • Musculoskeletal conditions (n=110 posts) • PTSD[c] (n=109 posts) • STIs (n=102 posts)
r/advice	Advice forum	2134 (2.7)	• Obsessive-compulsive disorder (n=186 posts) • STIs (n=184 posts) • Anxiety (n=168 posts) • Addiction and drug use (n=163 posts) • ADHD (n=144 posts)
r/teenagers	Intended for users aged 13-19 years to discuss age-appropriate subjects	1974 (2.5)	• Dermatology; unspecified (n=591 posts) • ADHD (n=154 posts) • Eye disorder (n=101 posts) • Depression (n=91 posts) • Anxiety (n=73 posts)
r/fitness	Physical fitness–based discussion	778 (0.99)	• Musculoskeletal conditions (n=152 posts) • Dieting (n=77 posts) • Bodybuilding (n=69 posts) • Gallbladder disorder (n=46 posts) • Testosterone use (n=37 posts)

XSL•FO
RenderX

Subreddit	Description	Posts, n (%)	Top 5 topics in subreddit
r/science	Forum for new scientific research	489 (0.62)	• Addiction and drug use (n=48 posts) • HPV[d] (n=29 posts) • Anxiety (n=27 posts) • STIs (n=23 posts) • Obsessive-compulsive disorder (n=19 posts)
r/diagnoseme	Medical question–based discussion	479 (0.61)	• Dermatology; unspecified (n=52 posts) • STIs (n=42 posts) • Allergies (n=38 posts) • Musculoskeletal conditions (n=28 posts) • Hypothyroidism (n=25 posts)
r/askdoctorsmeeee	Forum for medical questions	175 (0.22)	• Dermatology; unspecified (n=14 posts) • Musculoskeletal conditions (n=14 posts) • Allergies (n=12 posts) • Arrhythmia (n=9 posts) • Hypothyroidism (n=9 posts)
r/askdoctors	Forum for medical advice	150 (0.19)	• STIs (n=13 posts) • Dermatology; unspecified (n=11 posts) • Allergies (n=8 posts) • Anxiety (n=8 posts) • Gallbladder disorder (n=8 posts)
r/medicine	Forum for physicians to discuss medicine	125 (0.16)	• Addiction and drug use (n=18 posts) • Pregnancy (n=8 posts) • Anemia (n=6 posts) • Gallbladder disorder (n=5 posts) • HIV (n=5 posts)

[a]ADHD: attention-deficit/hyperactivity disorder.

[b]STI: sexually transmitted infection.

[c]PTSD: posttraumatic stress disorder.

[d]HPV: human papillomavirus.

XSL•FO
RenderX

Table 6. The top 5 topics by number of posts in the top 15 topic-specific subreddits (describing a certain health-related topic). A description of the subreddit and the total number of posts in the subreddit are provided (N=78,980).

Subreddit	Description	Posts, n (%)	Top 5 topics in each subreddit
r/skincareaddiction	Broad forum to discuss skin	2268 (2.87)	• Eye disorder (n=1226 posts) • Dermatology; unspecified (n=273 posts) • PCOS[a] (n=225 posts) • Allergies (n=87 posts) • Eczema (n=73 posts)
r/testosterone	Testosterone replacement therapy and testosterone levels	1746 (2.21)	• Vitamins (n=370 posts) • Testosterone use (n=264 posts) • Hypothyroidism (n=171 posts) • Pregnancy (n=105 posts) • Anemia (n=103 posts)
r/babybumps	Pregnancy-related discussion	1775 (2.25)	• Pregnancy (n=819 posts) • Musculoskeletal conditions (n=69 posts) • Anxiety (n=58 posts) • Menstrual cycle (n=58 posts) • STIs[b] (n=49 posts)
r/ADHD	Forum for people with ADHD[c] to exchange stories and strategies	1807 (2.29)	• ADHD (n=944 posts) • Addiction and drug use (n=199 posts) • Obsessive-compulsive disorder (n=112 posts) • Depression (n=91 posts) • Anxiety (n=87 posts)
r/STD	Forum to discuss STDs[d]	1611 (2.04)	• STIs (n=1158 posts) • HIV (n=75 posts) • Anxiety (n=54 posts) • HPV[e] (n=48 posts) • Respiratory infection (n=32 posts)
r/Pregnant	Forum to discuss pregnancy	1166 (1.48)	• Pregnancy (n=494 posts) • Menstrual cycle (n=92 posts) • Hematology (n=50 posts) • STIs (n=47 posts) • Anxiety (n=40 posts)
r/Depression	Support for those struggling with a depressive disorder	1193 (1.51)	• Depression (n=255 posts) • Obsessive-compulsive disorder (n=236 posts) • Anxiety (n=129 posts) • Addiction and drug use (n=940 posts) • ADHD (n=63 posts)
r/Anxiety	Support for those struggling with anxiety	1215 (1.54)	• Anxiety (n=628 posts) • Obsessive-compulsive disorder (n=165 posts) • Arrhythmia (n=55 posts) • PTSD[f] (n=53 posts) • Eye disorder (n=31 posts)
r/birthcontrol	Forum to discuss birth control	1102 (1.4)	• Menstrual cycle (n=521 posts) • Pregnancy (n=180 posts) • STIs (n=70 posts) • Anxiety (n=26 posts) • Musculoskeletal conditions (n=24 posts)
r/Childfree	Forum for those who choose to not have children	966 (1.22)	• Pregnancy (n=580 posts) • Menstrual cycle (n=58 posts) • Male reproductive health (n=32 posts) • Abuse (n=22 posts) • Anxiety (n=18 posts)

XSL•FO
RenderX

Subreddit	Description	Posts, n (%)	Top 5 topics in each subreddit
r/Hypothyroidism	Forum for those with hypothyroidism	779 (0.99)	• Hypothyroidism (n=519 posts) • Anemia (n=69 posts) • Dieting (n=29 posts) • Vitamins (n=23 posts) • Anxiety (n=19 posts)
r/MentalHealth	Moderated forum to discuss mental health issues	741 (0.94)	• Obsessive-compulsive disorder (n=208 posts) • Anxiety (n=126 posts) • Depression (n=73 posts) • ADHD (n=57 posts) • Addiction and drug use (n=43 posts)
r/Dentistry	Forum for dental professionals	683 (0.86)	• Dentistry (n=418 posts) • Dermatology; unspecified (n=28 posts) • Allergies (n=24 posts) • STIs (n=23 posts) • Musculoskeletal conditions (n=19 posts)
r/Diabetes	Forum for those living with diabetes	702 (0.89)	• Diabetes (n=491 posts) • Eye disorder (n=35 posts) • Anemia (n=23 posts) • Dieting (n=19 posts) • Vitamins (n=15 posts)
r/Infertility	Forum for those dealing with infertility for all reasons	446 (0.56)	• Pregnancy (n=110 posts) • Menstrual cycle (n=34 posts) • Vitamins (n=33 posts) • Hypothyroidism (n=29 posts) • Endometriosis (n=22 posts)

[a]PCOS: polycystic ovary syndrome.

[b]STI: sexually transmitted infection.

[c]ADHD: attention-deficit/hyperactivity disorder.

[d]STD: sexually transmitted disease.

[e]HPV: human papillomavirus.

[f]PTSD: posttraumatic stress disorder.

Medical Disciplines Divided by Subreddit

We also analyzed the breakdown of posts in each subreddit differentiated by medical discipline (Figure S1 in Multimedia Appendix 1). The relative frequency of medical disciplines in each subreddit was calculated by dividing the number of posts in a medical discipline by the total number of analyzed posts in that subreddit. Our analysis showed that many subreddits included posts about psychiatry and mental health. This mirrors the breakdown of medical disciplines overall as most posts belonged to that discipline. Some subreddits predominantly contained posts belonging to other medical disciplines. In those cases, the medical discipline often correlated with the subreddit (eg, in *r/hypothyroidism*, most posts belonged to the endocrinology, nutrition, and metabolism medical discipline). However, most subreddits comprised posts from multiple medical disciplines.

Discussion

Most Prevalent Medical Topics and Disciplines

By using markers of medical topics identified from broader samples of user data to categorize social media posts into specific diseases, previous studies have been able to predict diagnoses using social media information [13,14,20]. We accomplished a similar goal in our study by identifying topic keywords and using those to assign topics to posts. From there, our analysis revealed which disciplines were most prevalent across Reddit. Similarly to previous studies, we also used keywords to categorize this form of data into medical topics, but in this case, we did this on a broader scale, sorting data into all possible medical disciplines and generally not excluding posts that did not meet predetermined categories.

The 4 medical disciplines with the greatest proportion of posts were psychiatry and mental health; genitourinary and reproductive health; infectious diseases; and endocrinology, nutrition, and metabolism. We note the fact that psychiatry and mental health had approximately twice as many posts as the next highest medical discipline, genitourinary and reproductive health (21,009/79,980, 26.6% vs 10,741/78,980, 13.6% of posts) as another sign that discussion of mental health is common on Reddit.

Psychiatry and mental health, as a discipline, was the most prevalent. In addition to posts about anxiety, this medical discipline comprised posts about depression, addiction, alcoholism, and tobacco use. There were many posts about drug use, drug testing at work, and how drugs would affect other

XSL•FO
RenderX

health conditions. Overall, the medical discipline of psychiatry and mental health emphasized the amount of concern that people have about mental health topics and that they were willing to discuss their issues online. The large proportion of posts being about mental health raises the question of whether there is insufficient access to mental health services, whether online discussion of mental health is predominant over other health topics because of anonymity, or both.

Within this discipline, anxiety was another medical topic that frequently appeared, making up 5.43% (4289/78,980) of all posts and being in the top 5 topics in both general and topic-specific subreddits. Interestingly, anxiety was one of the most frequent topics not only in subreddits related to mental health but also in *r/PancreaticCancer*, *r/STD*, and *r/juicing*. Anxiety is a broad topic. Posts in this topic ranged from those about sharing experiences fighting melanoma to anxiety due to marijuana-triggered hallucinations to broad descriptions of "head pressure, headache, jaw pain, throat tension." Because the description of anxiety is open to interpretation, capturing which posts discuss anxiety is difficult and will result in a variety of posts that mention anxious feelings and worry, but the main subject of the post may not necessarily be anxiety.

We note that the most prevalent medical disciplines and topics can entail feelings of shame and embarrassment, be highly sensitive, or evoke a need for secrecy [33-35]. STIs was the most common topic, making up 7.44% (5876/78,980) of all posts. STIs had a significantly higher relative frequency than the next most common topic, eye disorders. Posts on this topic included many situational descriptions of suspected STIs. An example of an STI post is as follows:

> *[These] past couple of weeks, my girlfriend (we are in a monogamous relationship) found out that she was having light stringy brown discharge midway through her period cycles...So she went to the doctor and the doctor has prescribed her 1g Azithromycin (one time) and a 1 week course of Metrogyl 400 gms...The doctor also adviced [sic] her that I should take only the Azithromycin as I could have the infection as well (no blood tests or similar were done, the doctor did a physical examination)...*

The medical discipline of endocrinology, nutrition, and metabolism, which included posts about hypothyroidism, vitamins, and diabetes, among others, presented an interesting category. Posts discussed checking vitamin D levels and supplements. However, many posts also overlapped with the topic of pregnancy, such as checking vitamin levels during pregnancy and hormone levels. Therefore, it was difficult to parse whether all these posts were truly about endocrinology, nutrition, and metabolism. We also noticed that there were posts mentioning thyroid problems or diseases that were not specific to hypo- or hyperthyroidism. While these posts were more often categorized under hypothyroidism, the endocrine, nutrition, and metabolism medical discipline more accurately captured them under a broader category.

Overall, Reddit holds advantages over other social media platforms because its culture trends toward anonymity. Authors have discussed how the anonymity of communication over the internet reduces social desirability bias and social anxiety. Users are more disinhibited to discuss things openly [5]. Other studies have suggested that the internet is useful for collecting self-reports on risky health behaviors such as alcohol use, drug use, and smoking. Respondents accurately self-reported behaviors online, even ones that were undesirable, because of the perception of privacy [36]. It is established that social media sites with a culture of anonymity allow people to talk about sensitive issues online without fear of social desirability bias. Our study leverages that culture to gather a broader look at all medical disciplines rather than a predetermined set of medical topics. Our methods classified user data into those topics iteratively and broadly into all possible medical disciplines. Therefore, our findings expand on the literature by highlighting the most prevalent disciplines of conversation in a platform-wide study.

Interestingly, the top 5 topics may not be considered top health topics by health care providers. In a recent metric published by Blue Cross Blue Shield (Blue Cross Blue Shield Health Index), the top 10 health topics that had the greatest impact on quality of life in the United States did not include any of the topics that we mentioned previously. Therefore, our study may show that there is a different set of online topics, or it may reinforce that mining internet data reveals a different set of health concerns from those traditionally discussed or considered to be major health issues.

The Distribution of Topics

The aforementioned skewed distribution of topics suggests to us that conversation on Reddit is concentrated in a minority of discussion topics. This correlates with our supposition that users, when discussing health topics, overwhelmingly post about issues related to mental health, reproductive health, and infectious diseases, as well as endocrine, metabolic, and nutritional issues. The relative paucity of other topics may suggest that these are not discussed either in the general subreddits we focused on or in the topic-specific subreddits in the selection of health-related subreddits provided by Reddit but may be discussed in other subreddits.

As previously noted, 10 topics had no posts assigned to them. This meant that some topics that were created inductively during manual post review did not have any posts assigned to them during automatic topic assignment in the 1-year dataset. This invites us to investigate why certain topics would not have posts assigned to them. One reason is that posts in the subreddits included in the final analysis may not have been assigned to those topics. Another reason is that posts about a certain topic, for example, vaccination, may have also discussed measles, influenza, or human papillomavirus and, subsequently, were assigned to other topics, such as upper respiratory infection and STIs.

In general, users appeared more willing to talk about subjects that would otherwise be uncomfortable to discuss in person. This is in line with other studies that have examined how people are more willing to talk about chronic diseases on social media. They are also willing to share their disease experience in detail. Previous work has found that, on certain disease-oriented subreddits, users freely discuss disease-anatomy relationships,

XSL•FO
RenderX

such as cancer-lymph and asthma-lung relationships. The authors write that this openness on subreddits may stem from a desire to find others who share their experience of the disease [24] and users feel reassured by the subreddits' shared experiences.

Giving Advice

A subset of subreddits include those where users can post a question with the expectation that a subject matter expert will provide an answer while also being open to public input. On subreddits such as *r/askdocs* and *r/askdoctorsmeeee*, the experts are stated to be health professionals verified by the subreddit moderators. In a way, these accounts are analogous to the quantified influence described in the study by Koo et al [37] and the power accounts in the study by Sugawara et al [38], showing that, similarly to Twitter, medical advice about testing and diagnosis may be heavily influenced by several accounts on Reddit.

r/askdocs has approximately 270,000 members, whereas *r/askdoctorsmeeee* currently has approximately 14,400 members. In these subreddits, the most prevalent topics were similar to those found in other subreddits. We believe that many people turn to these forums for advice and that these subreddits are representative of the kinds of conversations that users have elsewhere on the site. However, because of the availability of experts on certain subreddits, Reddit users may be more discerning about the type of advice they expect to receive.

Limitations

We initially wanted to determine whether there was any correlation between medical topics that were mentioned on Reddit and medical tests that were mentioned on Reddit. However, our review of the Reddit posts demonstrated that users, in general, did not always mention specific medical tests, instead referring or alluding to a medical condition or a broad class of diseases, such as autoimmune diseases or STIs. At times, the content of a post would simply talk about "something being wrong," making it difficult to classify into a topic.

We noted other limitations, some of which we were able to address. Some posts were included that only had phrases such as "you should get that checked out," which were too vague to categorize under any topic. However, this problem was mitigated by using posts only from the 95 subreddits, where the frequency of such comments was lower. This was because those select subreddits, being more focused on their specific diseases or topics, had fewer vague comments and tended to have more thoughtful, directed discussion.

It is possible that keyword selection introduced bias into our topic assignments. For example, for the topic of benign prostatic hypertrophy, keywords included "difficulty urinating" and "difficulty peeing," but these keywords could have also been applicable to a topic such as nephrology (unspecified). However, we mitigated these biases by having multiple reviewers with backgrounds in clinical medicine and correlating our topic assignments with words in UpToDate. This ensured that the keywords we chose would lead to the least overlap between topics.

We also noted that some posts had too much jargon to be easily classifiable into any one topic. In future studies, they should be categorized in the *nonmedical* topic. Some posts, upon manual review, covered multiple health topics. These were the most difficult posts of all. As our algorithm did not have a way of tagging posts with multiple topics, these posts were assigned to only 1 topic. Future refinement of our algorithm could create a tagging system to identify all the major topics in a post, thereby circumventing this issue.

Future Applications

Our study provides insights into health discussions on Reddit. Our results are generalizable to other social media platforms. Facebook groups and Twitter hashtags allow for topic-based forums where people can find community discussion on health issues similarly to subreddits. Previous studies on Facebook and Twitter topic analyses have shown that social media data can be collected to better understand public sentiment on specific health topics [13-15,19,39,40]. The topic analysis methods we outlined could be applied to other social media platforms, which may show differences in what topics are discussed. Therefore, future platform-wide studies are needed.

We found that topics such as STIs and mental health issues are especially prevalent. These are potentially subjects that patients are not often discussing with their health care providers. As previously discussed, people tend to avoid potentially stigmatizing, embarrassing, and uncomfortable conversation topics, and STIs and mental health often entail such feelings. Therefore, our findings reinforce the need for health care systems to have open discussions about these issues. Health care providers establish a framework for these conversations through currently available questionnaires. The Patient Health Questionnaire–9 is a tool that providers can use to assess for depressive symptoms and mood. Asking for a sexual history during patient visits using the 5 P approach (partners, practices, protection from STIs, past history of STIs, and pregnancy) is a systematic way of starting a conversation about a patient's sexual history and concerns about STIs and pregnancy, topics that were found to be prevalent in our study. In addition, prior work has shown that, when patients seek health information online, the patient-physician relationship can be improved if the physician asks about this information and discusses it with the patient [41]. Our study supports that physicians and patients should discuss health information that the patient finds online.

Health-related discussions on social media and their effects on health policy remain an active area of research. Broadly, health-related discussions on social media have been cited as a potential impactor on disease surveillance and health care policy, practices, and delivery [42,43]. More recently, the relationship between social media and health policies has been extensively discussed in contexts such as COVID-19 [39,40,44] and mental health [45-49]. Further studies on the impact of social media on health interventions and health outcomes are needed [40]. It has been suggested that social media can play a role in health policy beyond increasing awareness of health issues [50]. Given the volume and wide variety of health topics discussed on Reddit, future health policies may need to address prevalent beliefs promulgated by health-related discussions on social

XSL•FO
RenderX

media sites. Guarding against health misinformation remains an important issue on social media and an active area of study [51-53]. Therefore, we support policies that encourage informed use of social media and discourage the spread of misinformation. Governance and education about obtaining health information online is crucial.

Conclusions

To investigate how people use Reddit to talk about seeking medical advice, we conducted a targeted review of posts that talked about "getting tested" and "checked out." Using an inductive approach to sort these posts into different medical topics, we found that discussion centered on several medical topics. They were further grouped into medical disciplines, and the infectious disease, reproductive, mental health, and endocrine categories had the greatest prevalence.

Reddit proved to be conducive to gathering information on what people discuss when talking about getting checked or tested. We took a platform-wide approach to this study, aiming to examine posts from a spectrum of subreddits related to health care and medicine. We believe that this approach is unique as it aimed to classify posts into all possible medical disciplines. In addition, based on our review of the literature, this is also one of the few studies conducted on Reddit to understand health discussions. Reddit provides similar advantages to those of other social media platforms in its large user base, wealth of information, and exchange of ideas. However, Reddit is additionally useful for studying communication about health concerns because of several of its features. Subreddits provide structured forums to further sort where discussions are happening on the platform. Several subreddits feature "expert" users, fostering a question-and-answer culture. Most importantly, users may favor Reddit because of its anonymity, which allows them to ask about and discuss potentially embarrassing topics in detail. In this way, Reddit's unique features prove it to be a viable source of research information that is worth exploring.

This study was a preliminary examination of the discussions occurring on Reddit. Future work on Reddit would likely take a more focused approach. Further exploration would include answering specific questions about certain topics encountered on Reddit and their prevalence in certain subreddits. Ultimately, these findings could be translated into actionable items, such as improving physician-patient interactions and laboratory management practices. Overall, the culture, structure, and wealth of information on Reddit make it a valuable source for continued investigation of health-related discussions.

Acknowledgments

The authors would like to thank the Department of Pathology and Laboratory Medicine at the University of Vermont Robert Larner College of Medicine and the RAND Corporation for their support. No generative artificial intelligence was used in the writing of this manuscript.

Data Availability

The datasets generated and analyzed during this study are available at the Reddit_public_data repository [54].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Analysis of our Reddit dataset.
[DOCX File , 156 KB - formative_v9i1e55309_app1.docx]

References

1. We Are Social. URL: https://wearesocial.com/blog/2020/01/digital-2020-3-8-billion-people-usesocial-media [accessed 2020-01-30]
2. Kim Y, Sohn D, Choi SM. Cultural difference in motivations for using social network sites: a comparative study of American and Korean college students. Comput Hum Behav 2011 Jan;27(1):365-372. [doi: 10.1016/j.chb.2010.08.015]
3. Whiting A, Williams D. Why people use social media: a uses and gratifications approach. Qual Mark Res 2013 Aug 30;16(4):362-369. [doi: 10.1108/QMR-06-2013-0041]
4. Moturu ST, Yang J, Liu H. Quantifying utility and trustworthiness for advice shared on online social media. In: Proceedings of the International Conference on Computational Science and Engineering. 2009 Presented at: CSE 2009; August 29-31, 2009; Vancouver, BC. [doi: 10.1109/CSE.2009.461]
5. Joinson A. Social desirability, anonymity, and internet-based questionnaires. Behav Res Methods Instrum Comput 1999 Sep;31:433-438. [doi: 10.3758/BF03200723]
6. Stephens-Davidowitz S. Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are. New York City, NY: William Morrow; 2017.
7. Berger J, Milkman KL. What makes online content viral? J Mark Res 2012 Apr 01;49(2):192-205. [doi: 10.1509/jmr.10.0353]

XSL•FO
RenderX

8. Shatz I. Fast, free, and targeted: Reddit as a source for recruiting participants online. Soc Sci Comput Rev 2016 May 19;35(4):537-549. [doi: 10.1177/0894439316650163]
9. Kubina M, Varmus M, Kubinova I. Use of big data for competitive advantage of company. Procedia Econ Financ 2015;26:561-565. [doi: 10.1016/s2212-5671(15)00955-7]
10. Robards B. Belonging and neo-tribalism on social media site Reddit. In: Hardy A, Bennett A, Robards B, editors. Neo-Tribes. Cham, Switzerland: Palgrave Macmillan; Sep 29, 2018.
11. Reddit statistics For 2024 (Demographics, usage and traffic data). Foundation Marketing. URL: https://foundationinc.co/lab/reddit-statistics/ [accessed 2022-09-14]
12. Amazon Alexa. Amazon Web Services. URL: https://www.alexa.com/siteinfo/reddit.com#card_overlap [accessed 2022-09-13]
13. Fan Y, Zhang Y, Ye Y, Li X, Zheng W. Social media for opioid addiction epidemiology: automatic detection of opioid addicts from Twitter and case studies. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017 Presented at: CIKM '17; November 6-10, 2017; Singapore, Singapore. [doi: 10.1145/3132847.3132857]
14. De la Torre-Díez I, Díaz-Pernas FJ, Antón-Rodríguez M. A content analysis of chronic diseases social groups on Facebook and Twitter. Telemed J E Health 2012;18(6):404-408. [doi: 10.1089/tmj.2011.0227] [Medline: 22650380]
15. Prieto VM, Matos S, Álvarez M, Cacheda F, Oliveira JL. Twitter: a good place to detect health conditions. PLoS One 2014 Jan 29;9(1):e86191 [FREE Full text] [doi: 10.1371/journal.pone.0086191] [Medline: 24489699]
16. Ren L, Lin H, Xu B, Zhang S, Yang L, Sun S. Depression detection on Reddit with an emotion-based attention network: algorithm development and validation. JMIR Med Inform 2021 Jul 16;9(7):e28754 [FREE Full text] [doi: 10.2196/28754] [Medline: 34269683]
17. Yao H, Rashidian S, Dong X, Duanmu H, Rosenthal RN, Wang F. Detection of suicidality among opioid users on Reddit: machine learning-based approach. J Med Internet Res 2020 Nov 27;22(11):e15293 [FREE Full text] [doi: 10.2196/15293] [Medline: 33245287]
18. Liu Y, Yin Z. Understanding weight loss via online discussions: content analysis of Reddit posts using topic modeling and word clustering techniques. J Med Internet Res 2020 Jun 08;22(6):e13745 [FREE Full text] [doi: 10.2196/13745] [Medline: 32510460]
19. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, et al. Facebook language predicts depression in medical records. Proc Natl Acad Sci U S A 2018 Oct 30;115(44):11203-11208 [FREE Full text] [doi: 10.1073/pnas.1802331115] [Medline: 30322910]
20. Merchant RM, Asch DA, Crutchley P, Ungar LH, Guntuku SC, Eichstaedt JC, et al. Evaluating the predictability of medical conditions from social media posts. PLoS One 2019 Jun 17;14(6):e0215476 [FREE Full text] [doi: 10.1371/journal.pone.0215476] [Medline: 31206534]
21. Ways to become a moderator. Reddit. URL: https://mods.reddithelp.com/hc/en-us/articles/360001745332-Ways-to-become-a-moderator [accessed 2023-02-28]
22. Park A, Conway M, Chen AT. Examining thematic similarity, difference, and membership in three online mental health communities from reddit: a text mining and visualization approach. Comput Human Behav 2018 Jan;78:98-112 [FREE Full text] [doi: 10.1016/j.chb.2017.09.001] [Medline: 29456286]
23. reddit - search results - PubMed. PubMed. URL: https://pubmed.ncbi.nlm.nih.gov/?term=reddit&sort_order=asc&size=200 [accessed 2022-09-14]
24. The ICD-10 classification of mental and behavioural disorders : diagnostic criteria for research. World Health Organization. URL: https://iris.who.int/handle/10665/37108 [accessed 2022-12-03]
25. UpToDate: trusted, evidence-based solutions for modern healthcare. Wolters Kluwer. URL: https://www.wolterskluwer.com/en/solutions/uptodate [accessed 2023-07-15]
26. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the Tidyverse. J Open Source Softw 2019 Nov 21;4(43):1686. [doi: 10.21105/joss.01686]
27. Foufi V, Timakum T, Gaudet-Blavignac C, Lovis C, Song M. Mining of textual health information from Reddit: analysis of chronic diseases with extracted entities and their relations. J Med Internet Res 2019 Jun 13;21(6):e12876 [FREE Full text] [doi: 10.2196/12876] [Medline: 31199327]
28. Nowak SA, Chen C, Parker AM, Gidengil CA, Matthews LJ. Comparing covariation among vaccine hesitancy and broader beliefs within Twitter and survey data. PLoS One 2020 Oct 8;15(10):e0239826 [FREE Full text] [doi: 10.1371/journal.pone.0239826] [Medline: 33031405]
29. Eldh AC, Årestedt L, Berterö C. Quotations in qualitative studies: reflections on constituents, custom, and purpose. Int J Qual Methods 2020 Nov 02;19:1-6. [doi: 10.1177/1609406920969268]
30. Patton MQ. Qualitative Research & Evaluation Methods: Integrating Theory and Practice. Thousand Oaks, CA: SAGE Publications; 2014.
31. Roman P, Rodriguez-Arrastia M, Molina-Torres G, Márquez-Hernández VV, Gutiérrez-Puertas L, Ropero-Padilla C. The escape room as evaluation method: a qualitative study of nursing students' experiences. Med Teach 2020 Apr;42(4):403-410. [doi: 10.1080/0142159X.2019.1687865] [Medline: 31738615]

XSL•FO
RenderX

32. Du C, Lee W, Moskowitz D, Lucioni A, Kobashi KC, Lee UJ. I leaked, then I Reddit: experiences and insight shared on urinary incontinence by Reddit users. Int Urogynecol J 2020 Feb;31(2):243-248. [doi: 10.1007/s00192-019-04165-8] [Medline: 31832718]
33. Stuart GS, Grimes DA. Social desirability bias in family planning studies: a neglected problem. Contraception 2009 Aug;80(2):108-112. [doi: 10.1016/j.contraception.2009.02.009] [Medline: 19631784]
34. Morisky DE, Ang A, Sneed CD. Validating the effects of social desirability on self-reported condom use behavior among commercial sex workers. AIDS Educ Prev 2002 Oct;14(5):351-360. [doi: 10.1521/aeap.14.6.351.24078] [Medline: 12413181]
35. Davis CG, Thake J, Vilhena N. Social desirability biases in self-reported alcohol consumption and harms. Addict Behav 2010 Apr;35(4):302-311. [doi: 10.1016/j.addbeh.2009.11.001] [Medline: 19932936]
36. Crutzen R, Peters GJ, Portugal SD, Fisser EM, Grolleman JJ. An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. J Adolesc Health 2011 May;48(5):514-519. [doi: 10.1016/j.jadohealth.2010.09.002] [Medline: 21501812]
37. Koo PK, Majdandzic A, Ploenzke M, Anand P, Paul SB. Global importance analysis: an interpretability method to quantify importance of genomic features in deep neural networks. PLoS Comput Biol 2021 May 13;17(5):e1008925 [FREE Full text] [doi: 10.1371/journal.pcbi.1008925] [Medline: 33983921]
38. Sugawara Y, Narimatsu H, Hozawa A, Shao L, Otani K, Fukao A. Cancer patients on Twitter: a novel patient community on social media. BMC Res Notes 2012 Dec 27;5:699 [FREE Full text] [doi: 10.1186/1756-0500-5-699] [Medline: 23270426]
39. Monselise M, Chang CH, Ferreira G, Yang R, Yang CC. Topics and sentiments of public concerns regarding COVID-19 vaccines: social media trend analysis. J Med Internet Res 2021 Oct 21;23(10):e30765 [FREE Full text] [doi: 10.2196/30765] [Medline: 34581682]
40. Valdez D, Ten Thij M, Bathina K, Rutter LA, Bollen J. Social media insights into US mental health during the COVID-19 pandemic: longitudinal analysis of Twitter data. J Med Internet Res 2020 Dec 14;22(12):e21418 [FREE Full text] [doi: 10.2196/21418] [Medline: 33284783]
41. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
42. Chen J, Wang Y. Social media use for health purposes: systematic review. J Med Internet Res 2021 May 12;23(5):e17917 [FREE Full text] [doi: 10.2196/17917] [Medline: 33978589]
43. Patrick M, Venkatesh RD, Stukus DR. Social media and its impact on health care. Ann Allergy Asthma Immunol 2022 Feb;128(2):139-145. [doi: 10.1016/j.anai.2021.09.014] [Medline: 34555532]
44. Lin CY, Broström A, Griffiths MD, Pakpour AH. Investigating mediated effects of fear of COVID-19 and COVID-19 misunderstanding in the association between problematic social media use, psychological distress, and insomnia. Internet Interv 2020 Sep;21:100345 [FREE Full text] [doi: 10.1016/j.invent.2020.100345] [Medline: 32868992]
45. Latha K, Meena KS, Pravitha MR, Dasgupta M, Chaturvedi SK. Effective use of social media platforms for promotion of mental health awareness. J Educ Health Promot 2020 May 28;9:124 [FREE Full text] [doi: 10.4103/jehp.jehp_90_20] [Medline: 32642480]
46. Mazuz K, Yom-Tov E. Analyzing trends of loneliness through large-scale analysis of social media postings: observational study. JMIR Ment Health 2020 Apr 20;7(4):e17188 [FREE Full text] [doi: 10.2196/17188] [Medline: 32310141]
47. Onyeaka H, Firth J, Ajayi KV, Muoghalu C, Holmes K, Nkemjika S, et al. Association between social media use and health promotion among individuals with depression and anxiety: insights from the 2017–2020 Health Information National Trends Survey. J Mood Anxiety Disord 2023 Jun;1:100006. [doi: 10.1016/j.xjmad.2023.100006]
48. Plackett R, Blyth A, Schartau P. The impact of social media use interventions on mental well-being: systematic review. J Med Internet Res 2023 Aug 11;25:e44922 [FREE Full text] [doi: 10.2196/44922] [Medline: 37565693]
49. Vandenbosch L, Fardouly J, Tiggemann M. Social media and body image: recent trends and future directions. Curr Opin Psychol 2022 Jun;45:101289. [doi: 10.1016/j.copsyc.2021.12.002] [Medline: 35030460]
50. Ghahramani A, de Courten M, Prokofieva M. "The potential of social media in health promotion beyond creating awareness: an integrative review". BMC Public Health 2022 Dec 21;22(1):2402 [FREE Full text] [doi: 10.1186/s12889-022-14885-0] [Medline: 36544121]
51. Allcott H, Boxell L, Conway J, Gentzkow M, Thaler M, Yang D. Polarization and public health: partisan differences in social distancing during the coronavirus pandemic. J Public Econ 2020 Nov;191:104254 [FREE Full text] [doi: 10.1016/j.jpubeco.2020.104254] [Medline: 32836504]
52. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. J Med Internet Res 2021 Jan 20;23(1):e17187 [FREE Full text] [doi: 10.2196/17187] [Medline: 33470931]
53. Wang ML, Togher K. Health misinformation on social media and adolescent health. JAMA Pediatr 2024 Feb 01;178(2):109-110. [doi: 10.1001/jamapediatrics.2023.5282] [Medline: 38079180]
54. Reddit_data_public. Google Drive. URL: https://drive.google.com/drive/folders/1gkyl8COxEeumbMZP-CFa2cbBnV1pNGt1 [accessed 2025-01-16]

XSL•FO
RenderX

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
***ICD-10*:** *International Classification of Diseases, 10th Revision*
STI: sexually transmitted infection

Edited by A Mavragani; submitted 08.12.23; peer-reviewed by J Mistry; comments to author 26.08.24; revised version received 10.10.24; accepted 28.10.24; published 29.01.25.

Please cite as:
Chan GJ, Fung M, Warrington J, Nowak SA
Understanding Health-Related Discussions on Reddit: Development of a Topic Assignment Method and Exploratory Analysis
JMIR Form Res 2025;9:e55309
URL: https://formative.jmir.org/2025/1/e55309
doi:10.2196/55309
PMID:39879094

©Garrett J Chan, Mark Fung, Jill Warrington, Sarah A Nowak. Originally published in JMIR Formative Research (https://formative.jmir.org), 29.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Person-Specific Analyses of Smartphone Use and Mental Health: Intensive Longitudinal Study

Merve Cerit[1], MS; Angela Y Lee[2], MA; Jeffrey Hancock[2,3], PhD; Adam Miner[4], MS, PsyD; Mu-Jung Cho[5], PhD; Daniel Muise[6], PhD; Anna-Angelina Garròn Torres[2], BA; Nick Haber[1,7], PhD; Nilam Ram[2,8], PhD; Thomas N Robinson[9], MD, MPH; Byron Reeves[2], PhD

[1]Graduate School of Education, Stanford University, Stanford, CA, United States
[2]Department of Communication, Stanford University, Stanford, CA, United States
[3]Cyber Policy Center, Stanford University, Stanford, CA, United States
[4]Department of Psychiatry and Behavioral Science, Stanford University, Stanford, CA, United States
[5]Center for Survey Research, Research Center for Humanities and Social Sciences, Academia Sinica, Taipei, Taiwan
[6]Screenlake Inc, San Mateo, CA, United States
[7]Department of Computer Science, Stanford University, Stanford, CA, United States
[8]Department of Psychology, Stanford University, Stanford, CA, United States
[9]Departments of Pediatrics, Medicine and Epidemiology & Population Health, Stanford University, Stanford, CA, United States

Corresponding Author:
Merve Cerit, MS
Graduate School of Education
Stanford University
520 Galvez Mall
Stanford, CA, 94305
United States
Phone: 1 650 723 21 46
Email: mervecer@stanford.edu

Abstract

Background: Contrary to popular concerns about the harmful effects of media use on mental health, research on this relationship is ambiguous, stalling advances in theory, interventions, and policy. Scientific explorations of the relationship between media and mental health have mostly been found null or have small associations, with the results often blamed on the use of cross-sectional study designs or imprecise measures of media use and mental health.

Objective: This exploratory empirical demonstration aims to answer whether mental health effects are associated with media use experiences by (1) redirecting research investments to granular and intensive longitudinal recordings of digital experiences to build models of media use and mental health for single individuals over the course of 1 year, (2) using new metrics of fragmented media use to propose explanations of mental health effects that will advance person-specific theorizing in media psychology, and (3) identifying combinations of media behaviors and mental health symptoms that may be more useful for studying media effects than single measures of dosage and affect or assessments of clinical symptoms related to specific disorders.

Methods: The activity on individuals' smartphone screens was recorded every 5 seconds when devices were in use over 1 year, resulting in a dataset of 6,744,013 screenshots and 123 fortnightly surveys from 5 adult participants. Each participant contributed between 0.8 and 2.7 million screens. Six media use metrics were derived from smartphone metadata. Fortnightly surveys captured symptoms of depression, attention-deficit/hyperactivity disorder, state anxiety, and positive affect. Idiographic filter models (p-technique canonical correlation analyses) were applied to explore person-specific associations.

Results: Canonical correlations revealed substantial person-specific associations between media use and mental health, ranging from r=0.82 (P=.008) to r=0.92 (P=.03). The specific combinations of media use metrics and mental health dimensions were different for each person, reflecting significant individual variability. For instance, the media use canonical variate for 1 participant was characterized by higher loadings for app-switching, which, in combination with other behaviors, correlated strongly with a mental health variate emphasizing anxiety symptoms. For another, prolonged screen time, alongside other media use behaviors, contributed to a mental health variate weighted more heavily toward depression symptoms. These within-person correlations are among the strongest reported in this literature.

XSL•FO
RenderX

Conclusions: Results suggest that the relationships between media use and mental health are highly individualized, with implications for the development of personalized models and precision smartphone-informed interventions in mental health. We discuss how our approach can be extended generally, while still emphasizing the importance of idiographic approaches. This study highlights the potential for granular, longitudinal data to reveal person-specific patterns that can inform theory development, personalized screening, diagnosis, and interventions in mental health.

(JMIR Form Res 2025;9:e59875) doi:10.2196/59875

KEYWORDS

media use; mental health; mHealth; uHealth; digital health; precision mental health; idiographic analysis; person-specific modeling; p-technique; longitudinal study; precision interventions; smartphones; idiosyncrasy; psychological well-being; canonical correlation analysis; United States

Introduction

Over the last decade, researchers have conducted extensive studies involving hundreds of thousands of participants to understand the relationship between media use and mental health. Despite significant investments, findings remain ambiguous, often reporting null or small associations [1-5]. This lack of clarity undermines progress in theory development, public policy, and clinical interventions. Researchers have proposed solutions, including granular tracking of media use [6-8], longitudinal study designs [9,10], the inclusion of at-risk populations [11-13], and person-specific analytics to account for individual differences [14,15].

A central challenge in this research is the problem of idiosyncrasy—the substantial variability in how individuals use media and how it affects their mental health [16,17]. In psychology, the opportunity to reclaim idiosyncratic variance has been defined as one of identifying idiographic filters for nomothetic interests [18]. The core argument is that the conventional practice of averaging out idiosyncrasy is neither theoretically satisfactory nor practically useful. We explore if and how an idiographic filter approach might enhance this research about media use and mental health. While researchers often seek universal rules to inform public policies or clinical interventions, such generalizations obscure the reality that individuals' media experiences are inherently personal and shaped by unique combinations of content and context. Aggregate analyses, which average out these differences, are insufficient and often misleading. Similarly, mental health symptom clusters frequently cut across traditional diagnostic boundaries, further emphasizing the need for individualized approaches. Addressing this challenge requires an idiographic perspective, which preserves individual variability and analyzes person-specific patterns rather than treating them as noise. By adopting an idiographic approach, research can uncover meaningful insights into the unique configurations of media use and mental health for each person.

To address this challenge, we used p-technique canonical correlation analysis (pCCA) [19], a method designed to analyze within-person relationships over time. This approach creates idiographic filters [20] that highlight how core constructs, such as media use and mental health metrics, manifest differently for each participant. Unlike traditional methods that assume consistency across individuals, pCCA uncovers the specific patterns and dynamics unique to each person based on the common set of variables. For example, media use heavily characterized by frequent app-switching may correlate with symptoms of hyperactivity for 1 individual, while prolonged social screen time may have significant weight in describing the changes in symptoms of depression for another. By prioritizing individual-level insights, this method provides a pathway to more personalized interventions (Multimedia Appendix 1 [2-4,7,15-19,21-50]).

This study uses granular longitudinal data collected through the Human Screenome Project [21,51], which recorded smartphone activity every 5 seconds for 1 year, resulting in over 6.7 million screenshots and 123 surveys across 5 participants. Mental health metrics, including symptoms of depression [52], attention-deficit/hyperactivity disorder (ADHD) [22], state anxiety [53], and positive affect were assessed fortnightly. Using pCCA, we explored person-specific associations between media use and mental health to address the following questions.

- Can granular longitudinal recordings reveal associations between media use and mental health that differ from the null or small effects reported in aggregated studies?
- Can new metrics, such as fragmentation of media use, explain person-specific mental health effects and advance media psychology theory?
- Can combinations of mental health symptoms provide better insights into media effects than single measures of dosage or individual clinical symptoms?

By focusing on individual experiences and variability over time, this study offers a novel approach to understanding the relationship between media use and mental health, emphasizing the potential of person-specific data to inform theory development and precision interventions.

Methods

Procedure and Participants

This study was conducted in the United States, leveraging a longitudinal design to assess smartphone use and its relationships with mental health. Data were collected via a custom smartphone app and biweekly participant surveys, enabling the integration of behavioral data with mental health measures. Participants were recruited via the Qualtrics research panel as part of a larger research project on assessing digital experiences with smartphone data [51]. Eligibility criteria included residing in the United States, being the sole user of an Android phone, and providing informed consent and Health

XSL•FO
RenderX

Insurance Portability and Accountability Act (HIPAA) authorization.

Participants installed a custom study app on their smartphones, enabling an unobtrusive collection of smartphone screens. The app required no active user management and operated transparently, with a status bar icon indicating its activity. Screenshots, along with operating system metadata like phone battery state and foreground app were captured and recorded automatically every 5 seconds when the screen was active.

Our study selected a focused sample that satisfied the minimum requirements for robust within-person analysis and allowed for a demonstration of our methodology, prioritizing long time series from a small number of participants over larger aggregate datasets so that we might uncover detailed person-specific insights and enable future scalability through computational methods [7,21,23]. We conducted a detailed analysis of 5 participants, randomly chosen from a group of 36 individuals that met several eligibility requirements. First, they had contributed at least 800,000 screenshots, reflecting an average of at least 3 hours of daily smartphone use over the year. This measure was chosen to capture use data that is close to the national average [54-56]. The goal was to ensure that participants had sufficient smartphone use to demonstrate relationships with mental health. Second, the selected participants had completed at least 20 biweekly surveys, creating a sufficient longitudinal record for each participant that exceeded the number of variables in the pCCA [57]. Finally, the selected participants had reported at least 1 biweekly period where their mental health symptoms reached clinically significant thresholds [22,52,58,59].

Ethical Considerations

This study adhered to national regulations and institutional policies, including the Declaration of Helsinki (2013), and was approved by Stanford University's institutional review board under protocol (#56430). All procedures were reviewed and classified as minimal risk.

Participants provided electronic informed consent and HIPAA authorization before enrollment. The consent process included detailed information about the study's goals, procedures, and privacy protections, and participants confirmed their consent using an e-signature interface. To ensure privacy and confidentiality, all data were encrypted during transfer and securely stored on Stanford's high-risk personal identifiable information and protected health information–compliant servers. Screenshots and metadata were deidentified prior to analysis, with access to raw screenshots restricted to a small subset of researchers under strict institutional protocols. Only aggregated, deidentified data were analyzed and reported to prevent the identification of individual participants. Participants were compensated US $15 for each fortnightly survey and data submission, with potential earnings of up to US $390 for completing the 1-year study.

Data were collected unobtrusively using a custom Screenomics app that captured screenshots every 5 seconds when the device was active. The app included an icon on the status bar to ensure transparency. Although nearly all screenshots were processed through automated methods, a small subset could be viewed by researchers in specific analyses. To mitigate risks, research staff were trained as mandated reporters under Stanford University and California guidelines. Any concerning content signaling potential legal, medical, or social risks was escalated to the Principal Investigators for appropriate action. Protocols were in place to refer participants to professionals or services as needed, ensuring their privacy while addressing potential risks to their well-being.

No images or supplementary materials in this manuscript include identifiable participant information. All visual data were anonymized to ensure privacy, and no consent was required for image inclusion as all images were deidentified.

All research staff completed mandatory HIPAA and human subjects training. Data security measures included strong passwords, restricted database access, and frequent reminders of privacy policies. These procedures ensured the highest level of data integrity and participant confidentiality.

Measures

Mental Health Measures

Participants completed fortnightly surveys over a year to assess symptoms of depression, and ADHD, in addition to state anxiety and positive affect. This extended time frame allowed the study of fluctuations across weeks or months [24-29].

Depression symptoms were measured using the 10-item Center for Epidemiologic Studies Depression scale [52]. Participants rated the frequency of symptoms (eg, "I felt I could not shake off the blues") on a 4-point Likert scale. Scores range from 0 to 60, with scores above 16 indicating clinical risk [59]. These scores were used to calculate the number of fortnight participants who were at clinical risk for depression (Table 1). The CES-D scale is in the public domain.

State anxiety was assessed with the 20-item State Anxiety subscale of the State-Trait Anxiety Inventory [60]. Items (eg, "I am tense; I am worried") were rated on a 4-point Likert scale. Scores range from 20 to 80, with scores over 40 indicating moderate to high anxiety [58]. This threshold was used to calculate fortnights at clinical risk (Table 1). A license was obtained from Mind Garden, Inc.

Positive affect was measured using a single item: "Looking back over the last two weeks, I felt happy," rated on a 100-point slider scale from "strongly disagree" to "strongly agree."

Hyperactivity and inattention symptoms were measured fortnightly using the 18-item Adult ADHD Self-Report Scale [22,61]. Participants rated hyperactivity and inattention on a 5-point Likert scale. Subscale scores range from 0 to 36. A score of 9 or more on the first 6 items (4 inattention and 2 hyperactivity) indicates clinical risk for adult ADHD [62]. These thresholds were used to calculate risk fortnights (Table 1). A license was obtained from NYU TOV Licensing, New York University.

XSL•FO
RenderX

Table 1. Participant data and demographics[a].

	Participant A (n=26)	Participant B (n=25)	Participant C (n=20)	Participant D (n=25)	Participant E (n=27)
Total screens	824,111	1,191,958	897,514	1,106,164	2,724,266
Age (year)	62	22	37	20	24
Sex	Female	Female	Female	Female	Female
Race	White	White	Multiracial Hispanic	White	Black
Region	Midwest	Midwest	Northeast	Pacific	Northeast
Residence type	Rural	Suburban	Suburban	Suburban	Urban
Highest educational level	2-year college	High school or GED[b]	Some college	Some college	Some College
Marital status	Divorced or separated	Single or never married	Married	Single or never married	Married
Income level (US $)	15,000-24,999	14,999 or less	25,000-29,999	75,000-99,999	14,999 or less

[a]The table presents the demographic characteristics of participants in the study, as collected through an initial onboarding survey. The survey was administered before the commencement of the study to gather pertinent information regarding participant backgrounds. “n” is the count of fortnightly surveys completed by each participant throughout the study year, while “Total screens” is the cumulative number of screenshots recorded from each participant’s smartphone during the study year.

[b]GED: General Educational Development.

Media Use Metrics

Six metrics for smartphone media use were derived from the screenshots and operating system metadata recorded continuously at 5-second intervals over the study year. Metadata enabled us to map app names and categories using Google Play Market classifications [63,64]. All metrics were averaged daily within biweekly periods, defined as the 14 days preceding each survey, except session duration calculated as the average screen-on interval. Missing data were addressed under the “missing at random” assumption.

Screen time was measured as the daily average number of screenshots captured during screen-on periods. The data collection app captured screenshots every 5 seconds when the screen was on. For each fortnight, the total screenshots were divided by the number of active data collection days.

Social screen time was measured as the daily average of screenshots recorded while using social apps. Social apps were those classified as “social” by Google Play Market categorization [63,64]. Social apps included common social media apps (eg, TikTok and Instagram), certain messaging apps (eg, Telegram and Text Free), and social network or dating apps (eg, Tagged and Grindr). Similar to the screen time measure, we calculated this metric by dividing the total number of social app screenshots by the number of active data collection days.

The number of sessions of smartphone use indicated average daily sessions, defined as intervals between screen on or off events. For each fortnight, we calculated the daily average number of sessions by dividing the total number of sessions by the number of days the participant actively used their phone. A higher number of sessions (eg, ten 30-second sessions) relative to the duration of each session (eg, one 5-minute session) indicates a more fragmented digital experience.

Session duration was calculated as the time interval between the initiation of a session and its termination. For each fortnight, we summed the duration of all sessions that occurred within that period and then computed the mean. This provided an average session duration metric for the specified time frame.

The number of unique apps was calculated as the daily average of distinct apps used, calculated by dividing the count of unique apps within a period by the number of active days.

The number of app switches indicated the daily average number of times individuals switched between apps and was calculated as a measure of fragmentation. For each biweekly period, we counted how many times individuals moved from app to app (ie, from Facebook to Zoom=1 switch) and divided that number by the number of active smartphone data collection days. Please see Multimedia Appendix 1 [2-4,7,15-19,21-50] for more discussion on theories and measures examining media use and mental health over time and the origins of our measures.

Data Analysis

Canonical correlation analysis (CCA) is a classical technique used to obtain the maximal correlation between objects that are represented by 2 sets of variables [65]. Adopting the Nesselroade et al idiographic filter approach [18], we use pCCA to model each individual’s multivariate repeated measures data. The observed correlation between an individual’s media use and mental health is not static; rather, it depends on the chosen coordinate system, that is, how observed variables are projected onto latent constructs of media use and mental health. Unlike traditional factorial analyses that assume a uniform variable projection across individuals, the idiographic filter approach allows person-specific rotations of the coordinate system.

pCCA does this by locating the specific rotation where the linear correlation between the latent media use and mental health canonical variates is mutually maximized. The size and significance of the maximum correlation (ρ) are used to answer our first research question: Does the analysis of single individuals offer different conclusions than the small or null effects reported in the across-individual analyses dominating the literature? The configurations of the basis vectors x and y

XSL•FO
RenderX

are interpreted to address our second and third questions: Can we identify combinations of media use metrics and mental health metrics that might advance theories about media effects?

Given A_x (27 occasions × 6 media use variables) and A_y (27 occasions × 5 mental health variables) for an individual, the data are projected onto canonical variates, $z_x = A_x x$ and $z_y = A_y y$, the correlation of which is given by the following equation.

where x and y are chosen to maximize ρ, achieved via singular value decomposition [19,66]. We use permutation tests to ensure observed correlations are not simply artifacts of random chance. Specifically, for each participant, we perform 1000 permutations by randomly shuffling the values within the columns of A_x or A_y, disrupting any temporal ordering between media use and mental health. This process generates a null distribution for ρ, representing the range of chance correlations expected in the absence of a systematic relationship. Comparing the observed ρ to this null distribution provides an exact P value based on the empirical data, thereby avoiding reliance on theoretical distributions that could inflate P values. Histograms of these permutations are presented in Multimedia Appendix 2. All pCCA and permutation test analyses were done using the *cca* package [67] in R (R Foundation for Statistical Computing).

Results

The following results include an analysis of a full year of media use and mental health data (6,744,013 screenshots and 123 fortnightly surveys) provided by 5 individuals. The background of each participant and descriptive statistics about media use and mental health are in Tables 1-3.

Table 2. Descriptive statistics for participants' mental health measures throughout 1 year[a].

Participants	Depression symptoms (0-60)		State anxiety (20-80)		ADHD[b] symptoms (0-72)			Positive affect (0-100), mean (SD)
	Mean (SD)	Clinical risk periods, n	Mean (SD)	Clinical risk periods, n	Inattention (0-36), mean (SD)	Hyperactivity (0-36), mean (SD)	Clinical risk periods, n	
A	9.23 (5.31)	1	44.23 (4.43)	22	1.88 (3.24)	2.04 (3.32)	0	53.85 (17.54)
B	24.80 (5.07)	24	51.84 (10.15)	21	12.72 (1.99)	8.72 (1.79)	13	62.12 (19.81)
C	12.40 (4.77)	4	45.05 (14.25)	13	10.10 (2.25)	5.70 (0.80)	3	65.35 (5.43)
D	12.80 (6.10)	7	45.84 (8.76)	17	5.36 (2.27)	2.76 (1.39)	0	57.00 (15.21)
E	14.78 (7.06)	4	55.00 (8.98)	27	4.52 (4.37)	4.34 (4.36)	0	16.85 (27.98)

[a]The table shows the means and SD for each mental health measure, as well as the clinical risk associated with symptomatology exceeding certain thresholds. Clinical risk periods are calculated as the number of fortnightly surveys in which participants' responses exceeded the clinical threshold for each measure.

[b]ADHD: attention-deficit/hyperactivity disorder.

Table 3. Descriptive statistics for participants' media use measures throughout 1 year.

Participant	Screen time (hours per day), mean (SD)	Social screen time (hours per day), mean (SD)	Number of sessions (per fortnight), mean (SD)	Session duration (minutes per day), mean (SD)	Number of unique apps (per day), mean (SD)	Number of app switches (per day), mean (SD)
A	2.99 (0.51)	0.41 (0.17)	57.84 (8.12)	3.10 (0.49)	4.08 (0.39)	210.27 (32.79)
B	4.15 (1.10)	0.37 (0.12)	32.76 (3.18)	7.59 (1.86)	3.64 (0.37	206.87 (18.96)
C	4.18 (0.95)	0.15 (0.23)	57.81 (8.89)	4.33 (0.85)	7.29 (2.50)	324.88 (52.73)
D	4.18 (1.05)	0 (0)	90.90 (25.28)	2.87 (0.89)	4.54 (0.61)	273.22 (54.38)
E	9.55 (2.03)	0 (0)	9.45 (2.10)	62.62 (14.75)	1.50 (0.59)	39.51 (14.48)

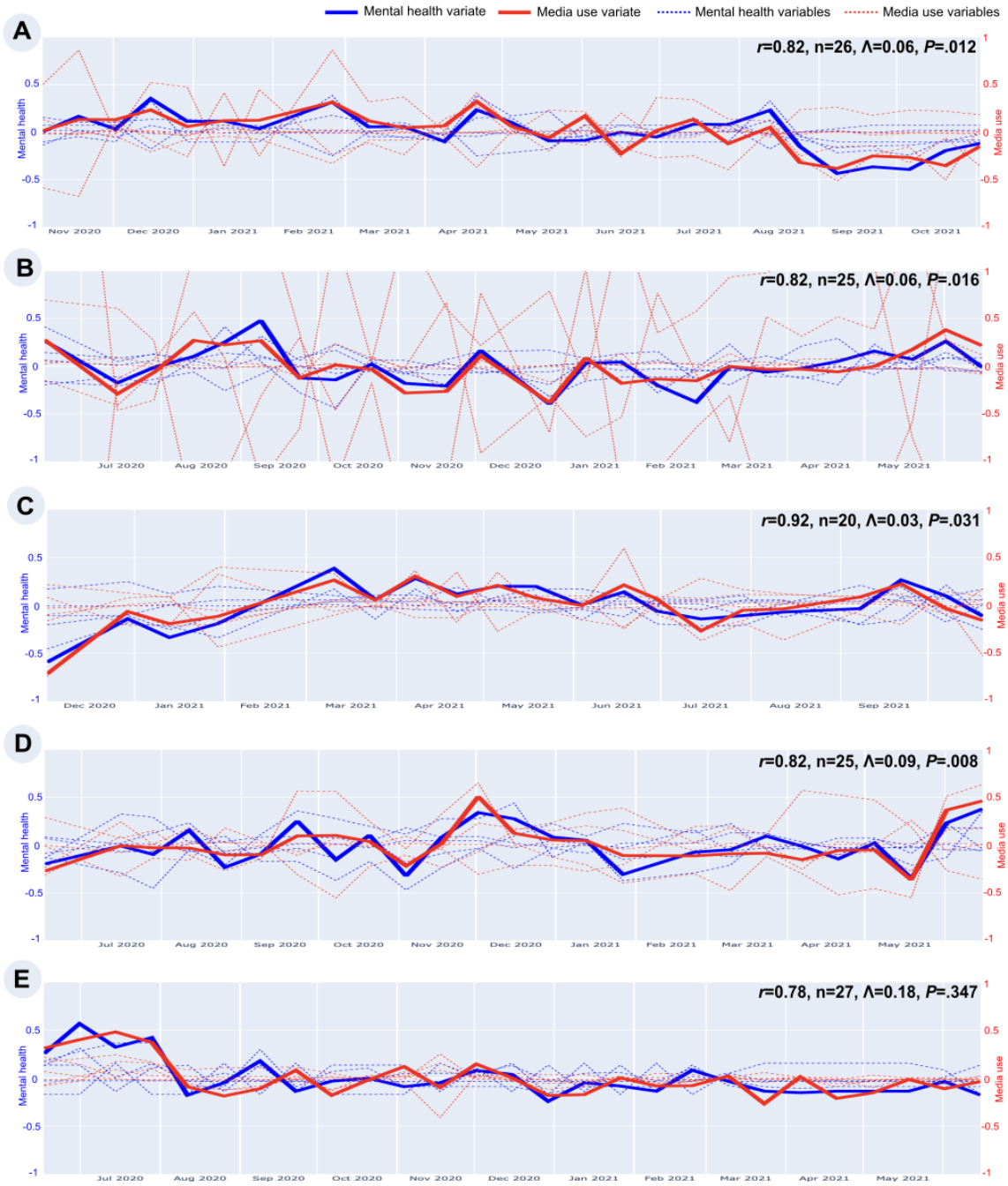
The main objective of this study was to assess whether analysis of media use metrics derived from super-intensive observational data passively collected from the smartphones individuals use in their daily lives could unearth relationships between media use and mental health within individuals that have previously been difficult to observe using analyses aggregated across individuals. Related to that objective, we first report the within-person associations between the media use and mental health latent variables that were obtained separately for each participant using pCCA.

The main results of this study are reported in Figure 1, which shows correlations and plots of the fortnightly media use and mental health variables across 1 year for each participant. The observed p-technique canonical correlations are large, ranging from 0.78 to 0.92. These values provide a compelling answer to our primary research question: for 4 of the 5 participants in

XSL•FO
RenderX

this study, fluctuations in canonical media use variate were significantly and tightly coupled with fluctuations in a canonical mental health variate, more so than for any previous associations observed in this literature [3,5].

Figure 1. Canonical variate plot for each participant (A-E). This figure presents the results of pCCA, highlighting within-person associations between media use and mental health metrics for five female Android phone users in the United States. The data were collected longitudinally as part of the Human Screenome Project, spanning 1 year (from July 2020 to October 2021). For each participant, the figure shows the temporal alignment of canonical media use variate (red solid line) and mental health variate (blue solid line), derived from granular measures recorded fortnightly. Individual metrics of media use (dotted red lines, eg, app-switching frequency, social screen time) and mental health (dotted blue lines, eg, depression, state anxiety) are also plotted in the background for comparison. Canonical correlations (r) for each participant range from 0.78 to 0.92, indicating strong covariation. Wilks statistic (Λ), *P* values, and sample size (n) are displayed for statistical significance. Permutation tests were conducted to validate correlations by generating null distributions based on shuffled data. Participant A shows a strong positive association between mental health and media use, with both variates generally moving together over time. Participant B exhibits a similar pattern, though with more pronounced fluctuations. Participant C has the highest correlation, indicating tightly coupled trends in media use and mental health across the year. Participant D also shows a strong association, though with some deviations at certain time points. In contrast, Participant E presents a weaker and statistically non-significant correlation, suggesting that media use and mental health follow more independent patterns in this case. These findings illustrate the unique within-person dynamics of media use and mental health over time, emphasizing the importance of person-specific models in understanding this relationship.



The composition of media use and mental health variates was unique for each participant, reflecting person-specific relationships and highlighting the heterogeneity in how media use impacts mental health. While individual metrics (eg,

XSL•FO
RenderX

app-switching, screen time) varied considerably over time, the canonical variates showed consistent and strong covariation for four of the 5 participants. These findings underscore the use of pCCA in identifying tightly coupled, within-person associations that are otherwise obscured in aggregated analyses, paving the way for more individualized theoretical insights and interventions.

The figures below highlight the unique person-specific relationships between media use metrics and mental health measures. These findings demonstrate the variability in how media use correlates with mental health for each participant, emphasizing the value of individualized approaches.

For participant A (Figure 2), the media use behavior most related to greater mental health symptoms over time was primarily shorter sessions (–0.93), and secondarily a higher number of sessions (0.51) and shorter screen time (–0.48). This suggests that the fluctuations in media use most related to mental health for this person are quick-checking behaviors on the smartphone. In complement, the configuration of mental health symptoms and positive affect that was most related to fluctuations in media use was characterized by lower state anxiety (–0.85) and lower levels of ADHD symptoms for both inattention (–0.59) and hyperactivity (–0.52), and lower positive affect (–0.71). Generally, for participant A, quick checking on the smartphone was associated with lower positive affect, and lower state anxiety and ADHD symptoms.

Figure 2. Canonical correlation analysis of participant A's media use and mental health. This figure illustrates the results of pCCA for participant A, based on 1 year of longitudinal data. Media use (left column) includes 6 variables (eg, session duration, number of sessions, and screen time), while mental health (right column) includes 5 variables (eg, state anxiety, depression symptoms, and positive affect). Arrows from the canonical variates (center) to observed variables indicate the direction (positive or negative) and strength of associations, with values closer to ±1.0 indicating stronger relationships. The canonical correlation (r=0.82; P=.01) highlights a significant within-person relationship, showing that quick-checking behaviors (eg, shorter session duration and more frequent sessions) were associated with reduced anxiety and ADHD symptoms but also diminished positive affect.

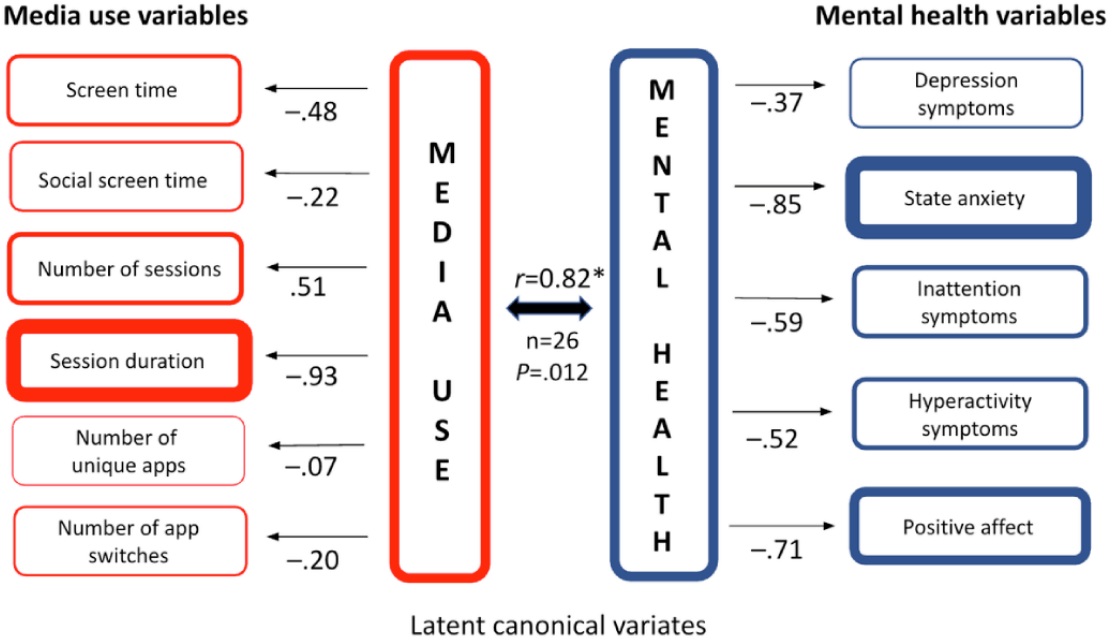
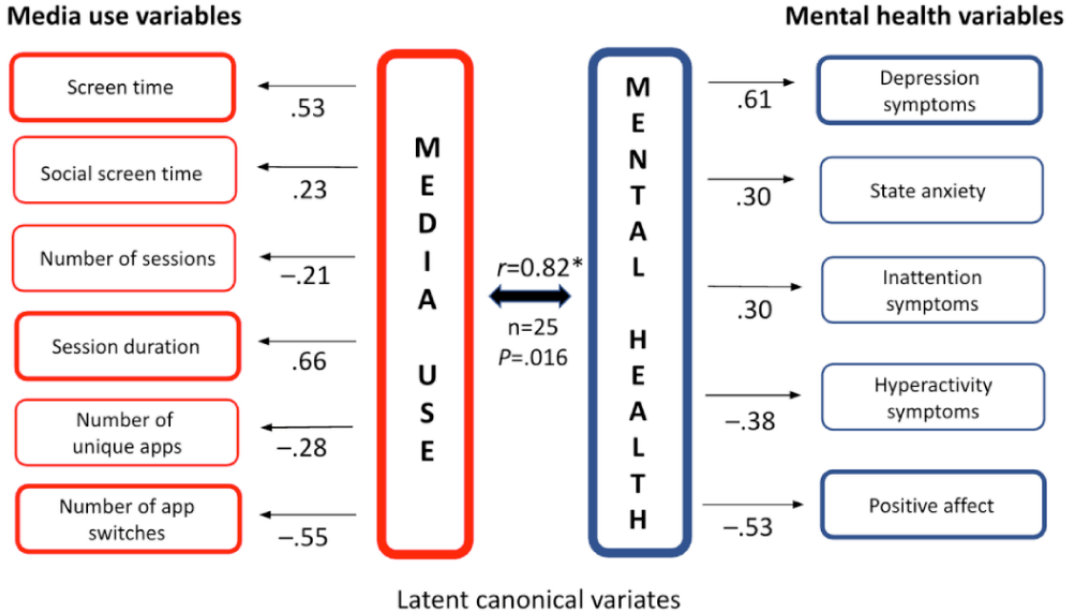


Figure 3 shows that for participant B, the media use behaviors most related to greater mental health symptoms over time were characterized by longer sessions (0.66) and secondarily by more screen time (0.53) and little app switching (–0.55). We characterize this combination as focused smartphone use. The mental health symptoms most related to media use were characterized by depression symptoms (0.61) and lower positive affect (–0.53). It is also notable that this participant's mental health changes, especially true for depression symptoms, had a stronger association with media use changes than for state anxiety. The media use canonical variate for participant B was driven by fluctuations in total screen time rather than social app use, specifically.

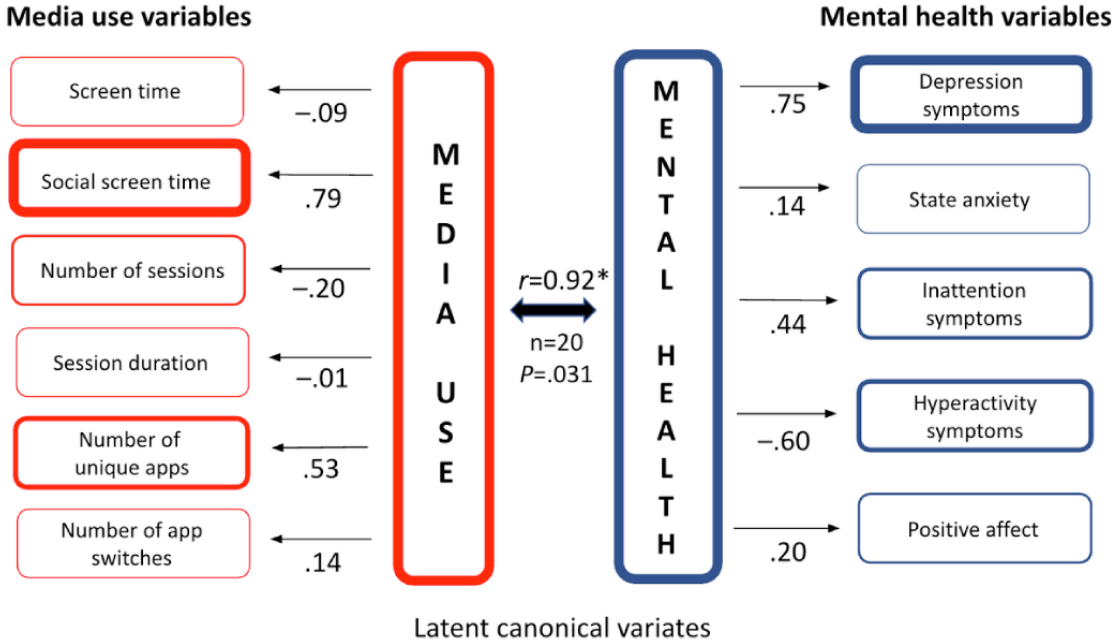
XSL•FO
RenderX

Figure 3. Canonical correlation analysis of participant B's media use and mental health. This figure illustrates the results of pCCA for participant B, based on 1 year of longitudinal data. Media use (left column) includes 6 variables (eg, session duration, screen time, and app-switching intensity), while mental health (right column) includes 5 variables (eg, depression symptoms, state anxiety, and positive affect). Arrows from the canonical variates (center) to observed variables indicate the direction (positive or negative) and strength of associations, with values closer to ±1.0 indicating stronger relationships. The canonical correlation (r=0.82; P=.02) highlights a significant within-person relationship, showing that focused smartphone use—characterized by longer session duration, higher screen time, and minimal app-switching—was associated with increased depression symptoms and reduced positive affect.



For participant C (Figure 4), the media use canonical variate shows the strongest loadings for the number of social screen time (distinct from total screen time) at 0.79. Secondarily, there is a high loading for the use of unique apps (0.53). The mental health canonical variate shows a strong loading associated with depression symptoms (0.75). Additionally, there is a noteworthy negative loading in relation to hyperactivity symptoms (–0.60). The much larger loading for social screen time in comparison to total screen time shows the value of separating these uses when examining relationships with mental health. This distinction is often part of discussions about media effects (ie, more worry about intensive social app use and depression than about the effects of all media use). As for many media effects, however, the value of the distinction is person-specific.

Figure 4. Canonical correlation analysis of participant C's media use and mental health. This figure illustrates the results of pCCA for participant C, based on 1 year of longitudinal data. Media use (left column) includes 6 variables (eg, social screen time, number of unique apps, session duration), while mental health (right column) includes 5 variables (eg, depression symptoms, hyperactivity symptoms, and positive affect). Arrows from the canonical variates (center) to the observed variables indicate the direction (positive or negative) and strength of associations, with values closer to ±1.0 reflecting stronger relationships. The canonical correlation (r=0.92; P=.03) underscores a significant within-person relationship, with social screen time (0.79) emerging as the strongest media use factor, closely tied to increased depressive symptoms (0.75) and reduced hyperactivity symptoms (–0.60).



XSL•FO
RenderX

Participant D (Figure 5) did not use any social apps during the entire year (ie, there are no screens from any of the hundreds of social apps in the Google Play store). Their media use canonical variate is characterized by fewer app switches (–0.77), extended session durations (0.71), and a reduced number of sessions overall (–0.57). This pattern indicates longer, focused interactions. The mental health canonical variate was characterized primarily by reduced positive affect (–0.58), with no large loadings associated with any of the mental health symptoms. For participant D, longer and more focused smartphone sessions had a more singular definition (diminished positive affect). However, the number and duration of the sessions were still highly correlated with smartphone use features (r=0.82; P=.008).

There was no statistically significant association between the media and mental health canonical variates for Participant E (Figure 6).

Figure 5. Canonical correlation analysis of participant D's media use and mental health. This figure presents the results of pCCA for participant D, based on longitudinal data collected over 1 year. Media use (left column) includes 6 variables (eg, session duration, number of sessions, and app switching), while mental health (right column) includes 5 variables (eg, positive affect, state anxiety, and ADHD symptoms). The arrows between latent canonical variates (center) and observed variables represent the direction (positive or negative) and strength of associations, with values closer to ±1.0 indicating stronger relationships. The canonical correlation (r=0.82; P=.008) underscores a significant within-person association, highlighting that longer session durations and reduced app-switching behaviors were most associated with lower positive affect for participant D.

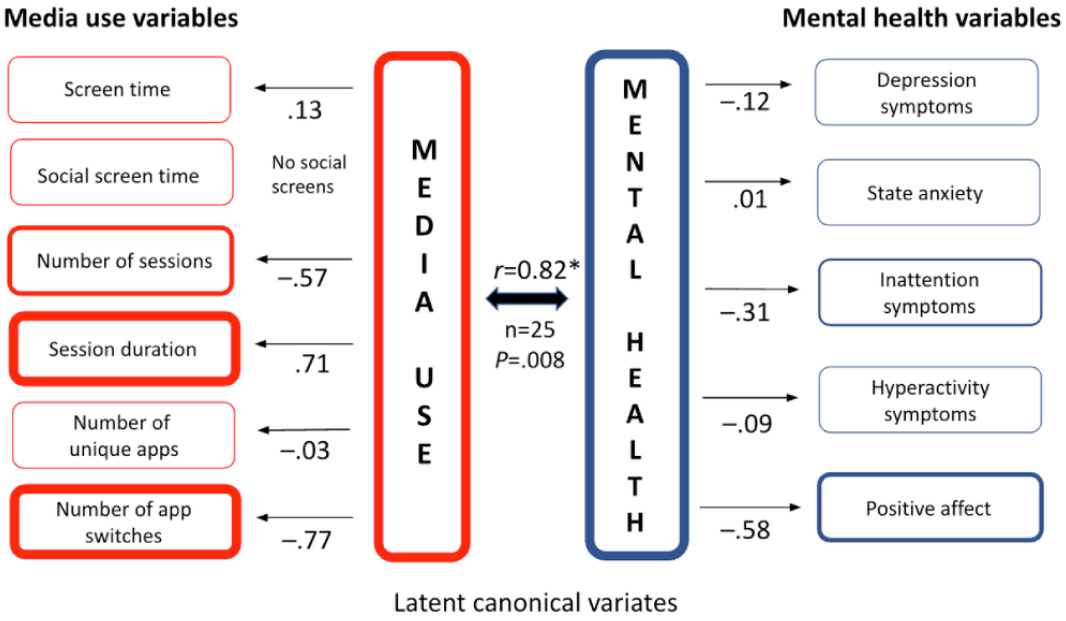
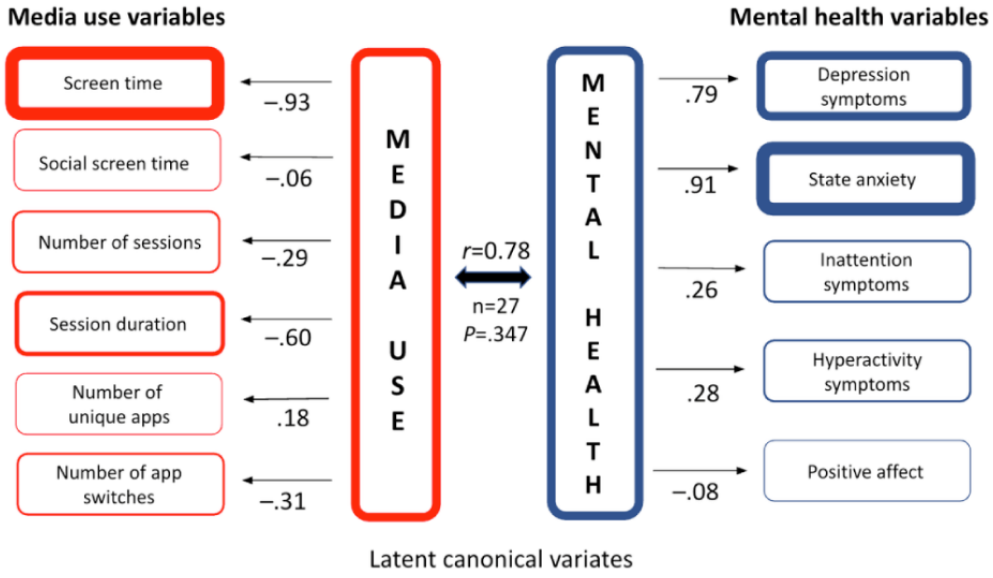


Figure 6. Canonical correlation analysis of participant E's media use and mental health. This figure illustrates the results of pCCA for participant E, based on 1 year of longitudinal data. Media use (left column) is represented by 6 variables, such as screen time, session duration, and number of app switches, while mental health (right column) is represented by 5 variables, including depression symptoms, state anxiety, and positive affect. Arrows connecting latent canonical variates (center) to observed variables indicate the direction (positive or negative) and strength of associations, with values closer to ±1.0 reflecting stronger associations. The canonical correlation (r=0.78; P=.35) indicates a nonsignificant association.



XSL•FO
RenderX

Discussion

Principal Findings

The results highlight the person-specific associations between media use and mental health. Canonical correlations revealed substantial covariation (0.82 to 0.92) over time for 4 out of 5 individuals. Simultaneously, as hypothesized, the specific combinations of media dosage metrics and dimensions of mental health that defined this covariation were different for each person (Multimedia Appendix 1 [2-4,7,15-19,21-50]. The approach supports the idea of a common association between media use and mental health while accommodating idiosyncrasy in how that association manifests for individuals. The couplings between media use and mental health for each person offer insights into the complexity of this relationship, directly addressing the research questions posed in the introduction.

In summary, our conclusions are the following.

- As research investments shift toward intensive, granular longitudinal data about individuals' digital experiences, this study demonstrates the potential to construct distinct, person-specific models of media use and mental health, revealing insights that transcend traditional aggregated analyses.
- The study's innovative approach, combining various dosage metrics of fragmented media use, advances person-specific theorizing in media psychology, offering potential explanations for the impact on mental health and emphasizing the significance of tailored approaches.
- By identifying individualized combinations of mental health symptoms, the study underscores the necessity of considering nuanced factors beyond single measures of dosage and affect or clinical symptoms associated with specific disorders.

Causality and Theory

The substantial p-technique canonical correlations observed in this study reflect within-person associations between media use and mental health but do not establish causality. Media use metrics may act as markers responding to mental health changes, antecedents influencing mental health, or reflections of confounding variables. A causal explanation likely involves complex mechanisms interacting over different time domains and individual histories, which this analysis does not aim to disentangle.

Past research has explored single mechanisms over varying time frames, such as the introduction of Facebook on campuses (months/years) [68], ecological momentary assessment studies linking self-reported media use with mental health (days/weeks) [69], and cohort analyses over decades [70]. Additionally, momentary exposures, like a single unflattering social media post, might amplify depression symptoms, particularly during periods of heightened vulnerability. Interactions between short- and long-term domains—for example, a negative social media post during an extended depressive episode—could further complicate causal pathways.

Combining mechanisms into a single causal conclusion for media use and mental health is beyond the scope of this project and likely any single study, given the heterogeneity observed across individuals. The participants in this study suggest that no universal explanation applies across diverse media behaviors, content, and time domains. Instead, multiple mechanisms, each tailored to specific contexts, are likely needed. Controlled experiments targeting distinct media experiences [71,72] and idiographic approaches like ours can help identify mechanisms for further exploration.

The within-person associations in this project emphasize the need for new theories addressing the accumulation of similar media experiences over time, regardless of specific content. Such mechanisms echo the concept of resonance in social reality theory [73,74], where outcomes arise from the alignment of digital and real-life environments. For instance, increased screen use during a depressive episode may reinforce social isolation, creating a feedback loop that intensifies mental health symptoms [75]. These findings highlight the importance of reexamining how repetitive media experiences influence mental health over extended periods.

Interventions

The p-technique canonical correlations observed in this project (Figure 1) suggest that results obtained for a participant over 1 year may generalize to future months. Acute changes in media use could serve as markers for triggering mental health interventions without requiring frequent or intrusive direct assessments. Interventions could integrate patient and clinician access to data through dashboards alerting users to mental health risks and offering tailored recommendations. For instance, systems might limit social app usage for some individuals or promote it for others based on identified risks. Person-specific thresholds for intervention provide a more accurate approach than generalized thresholds derived from population averages [76].

The key to these interventions is precision timing and targeted actions that connect continuous monitoring of screen activity with information designed to improve mental health. Screen-based systems might rival the diagnostic accuracy of traditional methods for some disorders. Studies estimate, for example, that up to 60% of people with depression are never diagnosed [77,78], and even those connected to care often face delays in appointments, diagnosis, and treatment initiation [79]. Screen summaries might accelerate care by offering behavioral insights prior to intake or as indicators of treatment effectiveness.

Determining the clinical use of idiographic tracking will require research to establish optimal baseline collection periods, which may vary by individual. While longer data collection can enhance accuracy, even shorter periods could yield valuable insights. For example, smartphone monitoring during the weeks between initial clinician contact and the first appointment, or in critical postdischarge periods, could inform preliminary evaluations and adaptive interventions. These monitoring systems could feed data into clinician dashboards and support the implementation of just-in-time adaptive interventions [80]. Importantly, such tools are intended to augment rather than replace direct patient-clinician interactions, which require robust support resources.

XSL•FO
RenderX

Limitations of the Study

This longitudinal study, conducted with a small, nonclinical sample of female Android phone users, was not intended to provide population-level conclusions about the prevalence of media and mental health relationships nationally or within subgroups. Avoiding generalization at this stage helps circumvent the small sample fallacy [81]. Our intent was to first understand the unique media dosage and mental health indicators important for individuals without prematurely investing in a research strategy aimed at extensive between-subject variability that we believe has resulted in considerable ambiguity in this literature. This bottom-up inductive approach celebrates idiosyncrasy and prioritizes individual models as a first step toward building models that might apply to larger groups of people. We have successfully demonstrated that it is possible to find strong, individualized relationships with granular measures and longitudinal data.

Our analysis focused on a limited set of variables: 6 media use metrics and 5 mental health measures. Many additional factors could influence within-person associations between media use and mental health. Notably, data collection coincided with the height of the COVID-19 pandemic. Although none of the participants contracted the virus, the pandemic's broader effects may have shaped their mental health and media use behaviors. Future analyses could examine how contextual factors, such as global or personal life events, moderate these relationships. For example, exploring interactions between media use, sleep patterns, and mental health outcomes [82,83], or identifying whether media use acts as a supportive resource [84,85] or an additional stressor [86,87] could inform context-aware models that better capture individual contingencies.

Future Work and Conclusions

While new computational methods enable the collection and analysis of large datasets, they often focus on aggregated measures, obscuring person-specific longitudinal insights and yielding ambiguous population-level conclusions. Our method emphasizes the importance of a person-specific approach applicable to new datasets. A key recommendation for extending this investigation is to maintain this individual focus and compare people based on similarities between models derived for single individuals, rather than relying solely on priori groupings based on demographics or other psychosocial characteristics. When idiographic models are available for many individuals, researchers can use clustering, latent class, and mixture models to quantify and examine between-person similarities and differences [88-90]. One such method suggested by Ram et al [88] involves calculating Tucker congruence coefficients [91] that compare parameter estimates (eg, the canonical variates) for each pair of individuals in a larger sample. Importantly, this approach entails specifying and modeling each person's data separately, as done in this study, before identifying shared patterns. Rather than assuming a priori commonalities (eg, based on sex or personality), this bottom-up paradigm uncovers the full range of person-specific solutions, which can then inform theory, policy, and interventions [18,88,92,93].

Future work will expand on the metrics presented here, integrating richer variables through our computational assay [20] and identifying screen content characteristics—such as words, objects, emotional arousal, and environmental features—using advanced artificial intelligence models. These granular data can help refine our understanding of how different media use patterns related to mental health. While traditional screen time measures, often derived from self-reports, have been criticized for their imprecision [94], this study suggests that accurate, granular recordings of screen behavior can reveal strong associations for some individuals. Similar increments in the precision of assessing media use have paid off in the study of other media effects; for example, exposure to political information [95], advertising [94], and food content in digital media [30]. Our ongoing work also includes open-sourcing the most up-to-date Screenomics app and toolset [96].

The idiographic filter approach, producing individual models 1 person at a time, may initially seem to risk overfocusing on small samples. However, there may not be a shorter route to useful answers about media use and mental health. This is especially true if we want to retain media use and mental health as important and psychologically complex categories of digital and non-digital life. The present results suggest that to advance research addressing media use and mental health problems generally, we may need to seek answers applicable to one person at a time; that is, we must embrace idiosyncrasy, not ignore it.

Acknowledgments

This project was partially funded by an internal Stanford University seed grant from the Stanford Medical School (Spectrum SPADA).

Data Availability

All deidentified data used and reported in this study, along with the code for analysis and visualizations, are available from the corresponding author upon reasonable request.

Authors' Contributions

TNR, NR, and BR did the inception and design of the study; MC, AYL, and BR drafted the manuscript and did data interpretation; NR, MC, and NH did data analysis and strategy; AM and JH advised on literature review, mental health assessments, and

XSL•FO
RenderX

interpretations; DM and AAGT organized data acquisition and data management. All authors have edited the manuscript in multiple versions and accept responsibility for the entire content of this manuscript.

Conflicts of Interest

DM is a founder of Screenlake Inc, a startup company focusing on smartphone data collection and analysis for the purpose of market intelligence. His involvement in the present study was the design and management of data collection during his doctoral studies, which took place prior to the founding of the company. JH consults for YouTube and Google on issues related to social media and receives compensation for these services. However, his contribution to this paper is solely his own work. The other authors declare no competing interests.

Multimedia Appendix 1
Supplementary Information.
[DOCX File , 31 KB - formative_v9i1e59875_app1.docx]

Multimedia Appendix 2
Permutation distribution histograms. This figure displays histograms of the permutation distributions of 1,000 permuted canonical correlations for each participant (A to E). To assess the statistical significance of the observed canonical correlations, we conducted permutation tests using Wilks' lambda by randomly permuting each individual's data, creating a null distribution. This null distribution represents correlations expected by chance—meaning the values that would arise if there were no systematic relationship between media use and mental health variables and any observed associations were purely random. The histograms display these chance-based distributions under the null hypothesis, with observed canonical correlations indicated by red dashed lines to evaluate their statistical significance. The test statistic shown in each histogram is Wilks' lambda. For Participant E, the result is nonsignificant (P> 0.05), suggesting that the observed correlation cannot be confirmed as statistically significant. For the other participants, the observed correlations are confirmed by the permutation test.
[PNG File , 5001 KB - formative_v9i1e59875_app2.png]

References

1. Sohn SY, Rees P, Wildridge B, Kalk NJ, Carter B. Prevalence of problematic smartphone usage and associated mental health outcomes amongst children and young people: a systematic review, meta-analysis and GRADE of the evidence. BMC Psychiatry 2019;19(1):356 [FREE Full text] [doi: 10.1186/s12888-019-2350-x] [Medline: 31779637]
2. Kushlev K, Leitao MR. The effects of smartphones on well-being: theoretical integration and research agenda. Curr Opin Psychol 2020;36:77-82. [doi: 10.1016/j.copsyc.2020.05.001] [Medline: 32563707]
3. Hancock J, Liu SX, Luo M, Mieczkowski H. Psychological well-being and social media use: a meta-analysis of associations between social media use and depression, anxiety, loneliness, eudaimonic, hedonic and social well-being. SSRN J 2022:84. [doi: 10.2139/ssrn.4053961]
4. Marciano L, Ostroumova M, Schulz PJ, Camerini AL. Digital media use and adolescents' mental health during the Covid-19 pandemic: a systematic review and meta-analysis. Front Public Health 2021;9:793868 [FREE Full text] [doi: 10.3389/fpubh.2021.793868] [Medline: 35186872]
5. Nawi A, Zameran FA. Exploring the relationship between smartphone use and adolescent well-being: a systematic literature review. In: Rajakumar G, Du KL, Vuppalapati C, Beligiannis GN, editors. Intelligent Communication Technologies and Virtual Mobile Networks. Singapore: Springer; 2023.
6. Harari GM, Müller SR, Aung MS, Rentfrow PJ. Smartphone sensing methods for studying behavior in everyday life. Curr Opin Behav Sci 2017;18:83-90. [doi: 10.1016/j.cobeha.2017.07.018]
7. Brinberg M, Bodie GD, Solomon DH, Jones SM, Ram N. Examining individual differences in how interaction behaviors change over time: a dyadic multinomial logistic growth modeling approach. Psychol Methods 2023. [doi: 10.1037/met0000605] [Medline: 37561491]
8. Mao C, Bayer JB, Ross MQ, Rhee L, Le HTK, Mount J, et al. Perceived vs. observed mHealth behavior: a naturalistic investigation of tracking apps and daily movement. Mobile Media Commun 2023;11(3):526-548. [doi: 10.1177/20501579221149823]
9. Barnett ML, Gonzalez A, Miranda J, Chavira DA, Lau AS. Mobilizing community health workers to address mental health disparities for underserved populations: a systematic review. Adm Policy Ment Health 2018;45(2):195-211 [FREE Full text] [doi: 10.1007/s10488-017-0815-0] [Medline: 28730278]
10. Chancellor S, De Choudhury M. Methods in predictive techniques for mental health status on social media: a critical review. NPJ Digit Med 2020;3:43 [FREE Full text] [doi: 10.1038/s41746-020-0233-7] [Medline: 32219184]
11. Torous J, Choudhury T, Barnett I, Keshavan M, Kane J. Smartphone relapse prediction in serious mental illness: a pathway towards personalized preventive care. World Psychiatry 2020;19(3):308-309 [FREE Full text] [doi: 10.1002/wps.20805] [Medline: 32931109]

XSL•FO
RenderX

12. Orben A, Przybylski AK, Blakemore SJ, Kievit RA. Windows of developmental sensitivity to social media. Nat Commun 2022;13(1):1649 [FREE Full text] [doi: 10.1038/s41467-022-29296-3] [Medline: 35347142]
13. Orben A, Blakemore SJ. How social media affects teen mental health: a missing link. Nature 2023;614(7948):410-412. [doi: 10.1038/d41586-023-00402-9] [Medline: 36788371]
14. Beyens I, Pouwels JL, van Driel II, Keijsers L, Valkenburg PM. Social media use and adolescents' well-being: developing a typology of person-specific effect patterns. Commun Res 2021;51(6):691-716. [doi: 10.1177/00936502211038196]
15. Valkenburg PM, Meier A, Beyens I. Social media use and its impact on adolescent mental health: an umbrella review of the evidence. Curr Opin Psychol 2022;44:58-68 [FREE Full text] [doi: 10.1016/j.copsyc.2021.08.017] [Medline: 34563980]
16. Bryant J, Oliver MB. Media Effects: Advances in Theory and Research. New York, NY: Routledge; 2009.
17. McLeod JM, Reeves B. On the nature of mass media effects. In: Withey SB, Abeles RP, editors. Television and Social Behavior: Beyond Violence and Children/A Report of the Committee on Television and Social Behavior, Social Science Research Council. Oxfordshire, United Kingdom: Routledge; 2013.
18. Nesselroade JR, Gerstorf D, Hardy SA, Ram N. Focus article: idiographic filters for psychological constructs. Meas Interdiscip Res Perspect 2007;5(4):217-235. [doi: 10.1080/15366360701741807]
19. Thompson B. Canonical Correlation Analysis: Uses and Interpretation. Washington, DC: Sage Publications; 1984.
20. Ram N, Haber N, Robinson TN, Reeves B. Binding the person-specific approach to modern AI in the human screenome project: moving past generalizability to transferability. Multivariate Behav Res 2024;59(6):1211-1219. [doi: 10.1080/00273171.2023.2229305] [Medline: 37439508]
21. Reeves B, Robinson T, Ram N. Time for the Human Screenome Project. Nature 2020;577(7790):314-317. [doi: 10.1038/d41586-020-00032-5] [Medline: 31942062]
22. Kessler RC, Adler L, Ames M, Demler O, Faraone S, Hiripi E, et al. The World Health Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population. Psychol Med 2005;35(2):245-256. [doi: 10.1017/s0033291704002892] [Medline: 15841682]
23. Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HYK, Chen R, et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell 2012;148(6):1293-1307 [FREE Full text] [doi: 10.1016/j.cell.2012.02.009] [Medline: 22424236]
24. Rozgonjuk D, Levine JC, Hall BJ, Elhai JD. The association between problematic smartphone use, depression and anxiety symptom severity, and objectively measured smartphone use over one week. Comput Hum Behav 2018;87:10-17. [doi: 10.1016/j.chb.2018.05.019]
25. Alloy LB, Uroševi ć S, Abramson LY, Jager-Hyman S, Nusslock R, Whitehouse WG, et al. Progression along the bipolar spectrum: a longitudinal study of predictors of conversion from bipolar spectrum conditions to bipolar I and II disorders. J Abnorm Psychol 2012;121(1):16-27 [FREE Full text] [doi: 10.1037/a0023973] [Medline: 21668080]
26. McInnis MG, Andreassen OA, Andreazza AC, Alon U, Berk M, Brister T, et al. Strategies and foundations for scientific discovery in longitudinal studies of bipolar disorder. Bipolar Disord 2022;24(5):499-508 [FREE Full text] [doi: 10.1111/bdi.13198] [Medline: 35244317]
27. Keller MB, Lavori PW, Mueller TI, Endicott J, Coryell W, Hirschfeld RM, et al. Time to recovery, chronicity, and levels of psychopathology in major depression. a 5-year prospective follow-up of 431 subjects. Arch Gen Psychiatry 1992;49(10):809-816. [doi: 10.1001/archpsyc.1992.01820100053010] [Medline: 1417434]
28. Crocetti E, Klimstra T, Keijsers L, Hale WW, Meeus W. Anxiety trajectories and identity development in adolescence: a five-wave longitudinal study. J Youth Adolesc 2009;38(6):839-849. [doi: 10.1007/s10964-008-9302-y] [Medline: 19636785]
29. Buecker S, Luhmann M, Haehner P, Bühler JL, Dapp LC, Luciano EC, et al. The development of subjective well-being across the life span: a meta-analytic review of longitudinal studies. Psychol Bull 2023;149(7-8):418-446. [doi: 10.1037/bul0000401]
30. Ram N, Yang X, Cho MJ, Brinberg M, Muirhead F, Reeves B, et al. Screenomics: a new approach for observing and studying individuals' digital lives. J Adolesc Res 2020;35(1):16-50 [FREE Full text] [doi: 10.1177/0743558419883362] [Medline: 32161431]
31. Humphreys L. The Qualified Self: Social Media and the Accounting of Everyday Life. Cambridge, MA: MIT Press; 2018.
32. Conway CC, Forbes MK, Forbush KT, Fried EI, Hallquist MN, Kotov R, et al. A hierarchical taxonomy of psychopathology can transform mental health research. Perspect Psychol Sci 2019;14(3):419-436 [FREE Full text] [doi: 10.1177/1745691618810696] [Medline: 30844330]
33. Nock MK, Kessler RC, Franklin JC. Risk factors for suicide ideation differ from those for the transition to suicide attempt: the importance of creativity, rigor, and urgency in suicide research. Clin Psychol Sci Pract 2016;23(1):31-34. [doi: 10.1037/h0101734]
34. Nock MK, Banaji MR. Prediction of suicide ideation and attempts among adolescents using a brief performance-based test. J Consult Clin Psychol 2007;75(5):707-715 [FREE Full text] [doi: 10.1037/0022-006X.75.5.707] [Medline: 17907852]
35. Robinaugh DJ, Brown ML, Losiewicz OM, Jones PJ, Marques L, Baker AW. Towards a precision psychiatry approach to anxiety disorders with ecological momentary assessment: the example of panic disorder. Gen Psychiatr 2020;33(1):e100161 [FREE Full text] [doi: 10.1136/gpsych-2019-100161] [Medline: 32175524]

XSL•FO
RenderX

36. Chahal R, Ho TC, Miller JG, Borchers LR, Gotlib IH. Sex-specific vulnerability to depressive symptoms across adolescence and during the COVID-19 pandemic: the role of the cingulum bundle. JCPP Adv 2022;2(1):e12061 [FREE Full text] [doi: 10.1002/jcv2.12061] [Medline: 35572852]
37. Hardoon DR, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: an overview with application to learning methods. Neural Comput 2004;16(12):2639-2664. [doi: 10.1162/0899766042321814] [Medline: 15516276]
38. Orben A. Teenagers, screens and social media: a narrative review of reviews and key studies. Soc Psychiatry Psychiatr Epidemiol 2020;55(4):407-414. [doi: 10.1007/s00127-019-01825-4] [Medline: 31925481]
39. Schimmack U, Krause P, Wagner GG, Schupp J. Stability and change of well being: an experimentally enhanced latent state-trait-error analysis. Soc Indic Res 2009;95(1):19-31. [doi: 10.1007/s11205-009-9443-8]
40. Liu X, Ping S, Gao W. Changes in undergraduate students' psychological well-being as they experience university life. Int J Environ Res Public Health 2019;16(16):2864 [FREE Full text] [doi: 10.3390/ijerph16162864] [Medline: 31405114]
41. Odgers CL, Jensen MR. Annual research review: adolescent mental health in the digital age: facts, fears, and future directions. J Child Psychol Psychiatry 2020;61(3):336-348 [FREE Full text] [doi: 10.1111/jcpp.13190] [Medline: 31951670]
42. Yoon S, Kleinman M, Mertz J, Brannick M. Is social network site usage related to depression? A meta-analysis of Facebook-depression relations. J Affect Disord 2019;248:65-72. [doi: 10.1016/j.jad.2019.01.026] [Medline: 30711871]
43. Houben M, Van Den Noortgate W, Kuppens P. The relation between short-term emotion dynamics and psychological well-being: a meta-analysis. Psychol Bull 2015;141(4):901-930. [doi: 10.1037/a0038822] [Medline: 25822133]
44. Meier A, Reinecke L. Computer-mediated communication, social media, and mental health: a conceptual and empirical meta-review. Commun Res 2020;48(8):1182-1209. [doi: 10.1177/0093650220958224]
45. Vanden Abeele MMP, Halfmann A, Lee EWJ. Drug, demon, or donut? Theorizing the relationship between social media use, digital well-being and digital disconnection. Curr Opin Psychol 2022;45:101295. [doi: 10.1016/j.copsyc.2021.12.007] [Medline: 35123383]
46. Goodman LS, Gilman A. In: Brunton LL, Knollman BC, editors. Goodman and Gilman's The Pharmacological Basis of Therapeutics. New York, NY: McGraw Hill; 2022.
47. Schleidgen S, Friedrich O, Gerlek S, Assadi G, Seifert J. The concept of “interaction” in debates on human-machine interaction. Humanit Soc Sci Commun 2023;10(1). [doi: 10.1057/s41599-023-02060-8]
48. Orben A, Meier A, Dalgleish T, Blakemore SJ. Mechanisms linking social media use to adolescent mental health vulnerability. Nat Rev Psychol 2024;3(6):407-423. [doi: 10.1038/s44159-024-00307-y]
49. Lee AY, Katz R, Hancock J. The role of subjective construals on reporting and reasoning about social media use. Soc Media Soc 2021;7(3):205630512110353. [doi: 10.1177/20563051211035350]
50. Yeykelis L, Cummings JJ, Reeves B. The fragmentation of work, entertainment, e-mail, and news on a personal computer: motivational predictors of switching between media content. Media Psychol 2017;21(3):377-402. [doi: 10.1080/15213269.2017.1406805]
51. Reeves B, Ram N, Robinson TN, Cummings JJ, Giles CL, Pan J, et al. Screenomics: A framework to capture and analyze personal life experiences and the ways that technology shapes them. Hum Comput Interact 2021;36(2):150-201 [FREE Full text] [doi: 10.1080/07370024.2019.1578652] [Medline: 33867652]
52. Radloff L. The CES-D Scale: A self-report depression scale for research in the general population. Appl Psychol Meas 1977;1(3):385-401 [FREE Full text] [doi: 10.1177/014662167700100306]
53. Marteau TM, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). Br J Clin Psychol 1992;31(3):301-306. [doi: 10.1111/j.2044-8260.1992.tb00997.x] [Medline: 1393159]
54. Wilkinson D. Screen time trends in the age of COVID-19. SimpleTexting. 2021. URL: https://simpletexting.com/blog/screen-time-survey [accessed 2023-10-22]
55. Flynn J. 20 Vital Smartphone Usage Statistics (2023): facts, data, and trends on mobile use in the U.S. Zippia.com. Zippia. 2023. URL: https://www.zippia.com/advice/smartphone-usage-statistics/ [accessed 2024-02-03]
56. Kerai A. Cell phone usage statistics: mornings are for notifications. Reviews.org. URL: https://www.reviews.org/mobile/cell-phone-addiction/ [accessed 2024-02-03]
57. Pezeshki A, Scharf LL, Azimi-Sadjadi MR, Lundberg M. Empirical canonical correlation analysis in subspaces. 2004 Presented at: Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers,; November 7-10, 2004; Pacific Grove, CA p. 994-997. [doi: 10.1109/acssc.2004.1399288]
58. Julian LJ. Measures of anxiety: State-Trait Anxiety Inventory (STAI), Beck Anxiety Inventory (BAI), and Hospital Anxiety and Depression Scale-Anxiety (HADS-A). Arthritis Care Res 2011;63:S467-S472 [FREE Full text] [doi: 10.1002/acr.20561] [Medline: 22588767]
59. Vargas-Terrones M, Nagpal TS, Perales M, Prapavessis H, Mottola MF, Barakat R. Physical activity and prenatal depression: going beyond statistical significance by assessing the impact of reliable and clinical significant change. Psychol Med 2021;51(4):688-693. [doi: 10.1017/S0033291719003714] [Medline: 32102723]
60. Spielberger CD, Gorsuch RL, Lushene RE. State-Trait Anxiety Inventory (STAI): Test Manual for Form X. Washington, DC: Consulting Psychologists Press; 1968.
61. Adler L, Cohen J. Diagnosis and evaluation of adults with attention-deficit/hyperactivity disorder. Psychiatr Clin North Am 2004;27(2):187-201. [doi: 10.1016/j.psc.2003.12.003] [Medline: 15063992]

XSL•FO
RenderX

62. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. Arch Gen Psychiatry 2005;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
63. JoMingyu. Google Play scraper. Github. URL: https://github.com/JoMingyu/google-play-scraper [accessed 2024-02-05]
64. Rhee L, Bayer JB, Lee DS, Kuru O. Social by definition: how users define social platforms and why it matters. Telematics Inform 2021;59:101538. [doi: 10.1016/j.tele.2020.101538]
65. Hotelling H. Relations Between Two Sets of Variates. In: Kotz S, Johnson NL, editors. Breakthroughs in Statistics: Methodology and Distribution. New York, NY: Springer; 1992.
66. Weenink D. Canonical correlation analysis. 2003 Presented at: Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam ;25?99; 2003; University of Amsterdam p. 81-99 URL: https://uvafon.hum.uva.nl/archive/2003/2003-ifaproc25-Weenink.pdf
67. González I, Déjean S. Package CCA: Canonical correlation analysis. R Foundation for Statistical Computing. 2021. URL: https://cran.r-project.org/package=CCA [accessed 2024-04-04]
68. Braghieri L, Levy R, Makarin A. Social media and mental health. Am Econ Rev 2022;112(11):3660-3693. [doi: 10.1257/aer.20211218]
69. Müller SR, Peters H, Matz SC, Wang W, Harari GM. Investigating the relationships between mobility behaviours and indicators of subjective well–being using smartphone–based experience sampling and GPS tracking. Eur J Pers 2020;34(5):714-732. [doi: 10.1002/per.2262]
70. Johnson AL. Changes in mental health and treatment, 1997-2017. J Health Soc Behav 2021;62(1):53-68. [doi: 10.1177/0022146520984136] [Medline: 33480305]
71. Bailey ER, Matz SC, Youyou W, Iyengar SS. Authentic self-expression on social media is associated with greater subjective well-being. Nat Commun 2020;11(1):4889 [FREE Full text] [doi: 10.1038/s41467-020-18539-w] [Medline: 33024115]
72. Brailovskaia J, Delveaux J, John J, Wicker V, Noveski A, Kim S, et al. Finding the "sweet spot" of smartphone use: reduction or abstinence to increase well-being and healthy lifestyle?! An experimental intervention study. J Exp Psychol Appl 2023;29(1):149-161. [doi: 10.1037/xap0000430] [Medline: 35389685]
73. Gerbner G, Gross L, Signorielli N, Morgan M. Aging with television: images on television drama and conceptions of social reality. J Commun 1980;30(1):37-47. [doi: 10.1111/j.1460-2466.1980.tb01766.x] [Medline: 7372841]
74. Clark JL, Green MC. Self-fulfilling prophecies: perceived reality of online interaction drives expected outcomes of online communication. Personal Individual Differ 2018;133:73-76. [doi: 10.1016/j.paid.2017.08.031]
75. Boers E, Afzali MH, Newton N, Conrod P. Association of screen time and depression in adolescence. JAMA Pediatr 2019;173(9):853-859 [FREE Full text] [doi: 10.1001/jamapediatrics.2019.1759] [Medline: 31305878]
76. Molenaar PCM, Campbell CG. The new person-specific paradigm in psychology. Curr Dir Psychol Sci 2009;18(2):112-117. [doi: 10.1111/j.1467-8721.2009.01619.x]
77. Williams SZ, Chung GS, Muennig PA. Undiagnosed depression: a community diagnosis. SSM Popul Health 2017;3:633-638 [FREE Full text] [doi: 10.1016/j.ssmph.2017.07.012] [Medline: 29349251]
78. Handy A, Mangal R, Stead TS, Coffee RL, Ganti L. Prevalence and impact of diagnosed and undiagnosed depression in the United States. Cureus 2022;14(8):e28011 [FREE Full text] [doi: 10.7759/cureus.28011] [Medline: 36134073]
79. Boerema AM, Ten Have M, Kleiboer A, de Graaf R, Nuyen J, Cuijpers P, et al. Demographic and need factors of early, delayed and no mental health care use in major depression: a prospective study. BMC Psychiatry 2017;17(1):367 [FREE Full text] [doi: 10.1186/s12888-017-1531-8] [Medline: 29145820]
80. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
81. Poulton EC. Small sample fallacy. In: Behavioral Decision Theory: A New Approach. Cambridge, MA: Cambridge University Press; 1994.
82. Li X, Buxton OM, Lee S, Chang AM, Berger LM, Hale L. Sleep mediates the association between adolescent screen time and depressive symptoms. Sleep Med 2019;57:51-60 [FREE Full text] [doi: 10.1016/j.sleep.2019.01.029] [Medline: 30897456]
83. Guerrero MD, Barnes JD, Chaput JP, Tremblay MS. Screen time and problem behaviors in children: exploring the mediating role of sleep duration. Int J Behav Nutr Phys Act 2019;16(1):105 [FREE Full text] [doi: 10.1186/s12966-019-0862-x] [Medline: 31727084]
84. Zhang K, Kim K, Silverstein NM, Song Q, Burr JA. Social media communication and loneliness among older adults: the mediating roles of social support and social contact. Gerontologist 2021;61(6):888-896. [doi: 10.1093/geront/gnaa197] [Medline: 33284972]
85. Lisitsa E, Benjamin KS, Chun SK, Skalisky J, Hammond LE, Mezulis AH. Loneliness among young adults during COVID-19 pandemic: the mediational rolea of social media use and social support seeking. J Soc Clin Psychol 2020;39(8):708-726. [doi: 10.1521/jscp.2020.39.8.708]

XSL•FO
RenderX

86. Zhao N, Zhou G. Social media use and mental health during the COVID-19 pandemic: moderator role of disaster stressor and mediator role of negative affect. Appl Psychol Health Well Being 2020;12(4):1019-1038 [FREE Full text] [doi: 10.1111/aphw.12226] [Medline: 32945123]
87. Seo J, Lee CS, Lee YJ, Bhang SY, Lee D. The type of daily life stressors associated with social media use in adolescents with problematic internet/smartphone use. Psychiatry Investig 2021;18(3):241-248 [FREE Full text] [doi: 10.30773/pi.2020.0060] [Medline: 33735548]
88. Ram N, Brose A, Molenaar PCM. Dynamic factor analysis: modeling person-specific process. In: The Oxford Handbook of Quantitative Methods. New York, NY: Oxford University Press; 2013:441.
89. van der Nest G, Lima Passos V, Candel MJJM, van Breukelen GJP. An overview of mixture modelling for latent evolutions in longitudinal data: modelling approaches, fit statistics and software. Adv Life Course Res 2020;43:100323. [doi: 10.1016/j.alcr.2019.100323] [Medline: 36726256]
90. Gates KM, Lane ST, Varangis E, Giovanello K, Guskiewicz K. Unsupervised classification during time-series model building. Multivariate Behav Res 2017;52(2):129-148 [FREE Full text] [doi: 10.1080/00273171.2016.1256187] [Medline: 27925768]
91. Lorenzo-Seva U, ten Berge JMF. Tucker's congruence coefficient as a meaningful index of factor similarity. Methodology 2006;2(2):57-64. [doi: 10.1027/1614-2241.2.2.57]
92. Brose A, Ram N. Within-person factor analysis: Modeling how the individual fluctuates changes across time. In: Mehl MR, editor. Handbook of Research Methods for Studying Daily Life. New York, NY: Guilford Press; 2012:459.
93. Yang X, Ram N, Lougheed JP, Molenaar PCM, Hollenstein T. Adolescents' emotion system dynamics: network-based analysis of physiological and emotional experience. Dev Psychol 2019;55(9):1982-1993 [FREE Full text] [doi: 10.1037/dev0000690] [Medline: 31464499]
94. Lee J, Hamilton JT, Ram N, Roehrick K, Reeves B. The psychology of poverty and life online: natural experiments on the effects of smartphone payday loan ads on psychological stress. Inf Commun Soc 2022;26(14):2775-2796. [doi: 10.1080/1369118x.2022.2109982]
95. Muise D, Hosseinmardi H, Howland B, Mobius M, Rothschild D, Watts DJ. Quantifying partisan news diets in web and TV audiences. Sci Adv 2022;8(28):eabn0083 [FREE Full text] [doi: 10.1126/sciadv.abn0083] [Medline: 35857498]
96. Yee AZH, Yu R, Lim SS, Lim KH, Dinh TTA, Loh L, et al. ScreenLife Capture: an open-source and user-friendly framework for collecting screenomes from android smartphones. Behav Res Methods 2023;55(8):4068-4085 [FREE Full text] [doi: 10.3758/s13428-022-02006-z] [Medline: 36289177]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
CCA: canonical correlation analysis
HIPAA: Health Insurance Portability and Accountability Act
pCCA: p-technique canonical correlation analysis

Edited by A Mavragani; submitted 24.04.24; peer-reviewed by A AL-Asadi, T Seibers; comments to author 20.08.24; revised version received 15.10.24; accepted 17.12.24; published 26.02.25.

Please cite as:
Cerit M, Lee AY, Hancock J, Miner A, Cho MJ, Muise D, Garròn Torres AA, Haber N, Ram N, Robinson TN, Reeves B
Person-Specific Analyses of Smartphone Use and Mental Health: Intensive Longitudinal Study
JMIR Form Res 2025;9:e59875
URL: https://formative.jmir.org/2025/1/e59875
doi:10.2196/59875
PMID:39808832

©Merve Cerit, Angela Y Lee, Jeffrey Hancock, Adam Miner, Mu-Jung Cho, Daniel Muise, Anna-Angelina Garròn Torres, Nick Haber, Nilam Ram, Thomas N Robinson, Byron Reeves. Originally published in JMIR Formative Research (https://formative.jmir.org), 26.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

The AI Reviewer: Evaluating AI's Role in Citation Screening for Streamlined Systematic Reviews

Jamie Ghossein[1*], MD; Brett N Hryciw[2*], MD; Tim Ramsay[3,4], PhD; Kwadwo Kyeremanteng[2,4,5], MD

[1]Interdepartmental Division of Critical Care Medicine, University of Toronto, Toronto, ON, Canada

[2]Division of Critical Care, Department of Medicine, University of Ottawa, 501 Smyth Road, Ottawa, ON, Canada

[3]Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

[4]Clinical Epidemiology, Ottawa Hospital Research Institute, Ottawa, ON, Canada

[5]Institute du Savoir Montfort, Montfort Hospital, Ottawa, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Brett N Hryciw, MD
Division of Critical Care, Department of Medicine, University of Ottawa, 501 Smyth Road, Ottawa, ON, Canada

(JMIR Form Res 2025;9:e58366) doi:10.2196/58366

KEYWORDS

article screening; artificial intelligence; systematic review; AI; large language model; LLM; screening; analysis; reviewer; app; ChatGPT 3.5; chatbot; dataset; data; adoption

Introduction

Systematic reviews are regarded as one of the highest form of evidence in medical research and are vital for answering clinical questions [1]. However, the conventional systematic review methodology is time-consuming, particularly the manual screening of articles for pertinence [2]. The exponential increase in biomedical literature presents a challenge for researchers to remain updated. Artificial intelligence (AI) has shown promise in various fields [3], with large language models (LLMs) specifically offering capabilities to interpret complex text, which can be leveraged in the systematic review process [4]. We conducted a pilot feasibility study evaluating 5 distinct LLMs in an existing systematic review dataset.

Methods

Overview

We compared 5 commonly used LLMs to screen citations from a previously published systematic review on trauma hemorrhage, originally screened by two human reviewers [5]. Of the 1186 total citations, 21 (1.8%) were included for full-text review and 1165 (98.2%) were excluded. We randomly selected 100 excluded citations using Microsoft Excel. Hence, 121 citations (n=21, 17.4% included and n=100, 82.6% excluded) were tested against predefined eligibility criteria using ChatGPT 3.5 (version September 25, 2023), ChatGPT 4 (version September 25, 2023), Google Bard (version 1.15; released on September 2, 2023), Meta Llama 2 (70b parameters, version 2.1.1; released on October 10, 2023), and Claude AI 2 (version 1.3; released on July 11, 2023). We used descriptive statistics to evaluate sensitivity, specificity, and overall accuracy.

Ethical Considerations

All citations were taken from publicly available, previously published literature. No personal or patient-level data were used, and no identifiers were included. Formal research ethics board approval was therefore not required.

Results

Among the 121 total citations, the LLMs' sensitivity (correctly identifying included citations) ranged from 57% to 100%, and specificity (correctly excluding noneligible citations) ranged from 18% to 79%. ChatGPT 3.5 achieved the highest sensitivity (100%) and the highest specificity (79%). Full results are shown in Table 1.

Table . Performance metrics of large language models in citation screening for systematic reviews, including sensitivity, specificity, and accuracy.

Large language model	Sensitivity, %	Specificity, %	Accuracy, %
ChatGPT 3.5	100	79	83
ChatGPT 4	95	66	72
Google Bard	100	71	77
Meta Llama 2 (70b parameters)	95	18	34
Claude AI 2	57	77	73

XSL•FO
RenderX

Discussion

In this pilot assessment, selected LLMs demonstrated high sensitivity for identifying relevant studies, with ChatGPT 3.5 and Google Bard reaching 100%. Notably, the specificity varied widely, ranging from as low as 18% for Meta Llama 2 to 79% for ChatGPT 3.5. While some LLMs can be remarkably sensitive for screening articles within our sample, excluding irrelevant citations remains a challenge for certain LLMs. These findings suggest that AI-driven LLMs could be poised to support the screening phase, potentially replacing the second human reviewer and streamlining the often labor-intensive study screening process.

The sample size of 121 citations is a limitation, and findings may not be generalizable to other systematic reviews or inclusion and exclusion criteria. Larger studies, ideally with multiple runs of the same citations, are necessary to capture the probabilistic variability inherent to LLMs. As we only ran each citation through a given LLM once, multiple runs or "prompt engineering" strategies could yield more consistent or refined outcomes when evaluating LLMs. Nonetheless, our study offers a novel approach by directly comparing the performance of multiple LLMs, thus providing insight into how different architectures perform on the same dataset. Future research should explore repeated runs to assess LLM consistency, implement advanced prompt engineering, and investigate the explainability of LLM results.

LLMs have previously been demonstrated to effectively generate Boolean queries for a systematic review literature search [1]. As LLMs evolve further, it is conceivable that they could entirely manage the title and abstract screening. This progress can eventually lead to a fully automated review process, where AI might oversee the search strategy, title and abstract screening, full-text review, data analysis and synthesis, and even drafting and publication. Such automation would epitomize a living systematic review, ensuring evidence is continuously updated as soon as new research is published. As transparency and accountability concerns may arise, a robust ethical framework will be paramount as we navigate the advancements of this technology [6].

Conflicts of Interest

None declared.

References

1. Wang S, Scells H, Koopman B, Zuccon G. Can ChatGPT write a good boolean query for systematic review literature search? Presented at: SIGIR '23: The 46th International ACM SIGIR Conference on Research and Development in Information Retrieval; Jul 23-27, 2023; Taipei, Taiwan p. 1426-1436. [doi: 10.1145/3539618.3591703]
2. Tsafnat G, Glasziou P, Karystianis G, Coiera E. Automated screening of research studies for systematic reviews using study characteristics. Syst Rev 2018 Apr 25;7(1):64. [doi: 10.1186/s13643-018-0724-7] [Medline: 29695296]
3. Hryciw BN, Fortin Z, Ghossein J, Kyeremanteng K. Doctor-patient interactions in the age of AI: navigating innovation and expertise. Front Med (Lausanne) 2023 Aug 30;10:1241508. [doi: 10.3389/fmed.2023.1241508] [Medline: 37711734]
4. Kanjee Z, Crowe B, Rodman A. Accuracy of a generative artificial intelligence model in a complex diagnostic challenge. JAMA 2023 Jul 3;330(1):78-80. [doi: 10.1001/jama.2023.8288] [Medline: 37318797]
5. Ghossein J, Fernando SM, Rochwerg B, Inaba K, Lampron J, Tran A. A systematic review and meta-analysis of sample size methodology for traumatic hemorrhage trials. J Trauma Acute Care Surg 2023 Jun 1;94(6):870-876. [doi: 10.1097/TA.0000000000003944] [Medline: 36879398]
6. Jha D, Durak G, Sharma V, Keles E, Cicek V, Zhang Z, et al. A conceptual algorithm for applying ethical principles of AI to medical practice. arXiv. Preprint posted online on Jan 3, 2025. [doi: 10.48550/arXiv.2304.11530]

Abbreviations

AI: artificial intelligence
LLM: large language model

Edited by A Mavragani; submitted 13.03.24; peer-reviewed by J Deng, J Mistry, Q Khraisha; revised version received 26.01.25; accepted 29.01.25; published 28.03.25.

Please cite as:
Ghossein J, Hryciw BN, Ramsay T, Kyeremanteng K
The AI Reviewer: Evaluating AI's Role in Citation Screening for Streamlined Systematic Reviews
JMIR Form Res 2025;9:e58366
URL: https://formative.jmir.org/2025/1/e58366
doi:10.2196/58366

XSL•FO
RenderX

©Jamie Ghossein, Brett N Hryciw, Tim Ramsay, Kwadwo Kyeremanteng. Originally published in JMIR Formative Research (https://formative.jmir.org), 28.3.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Engaging Stakeholders With Professional or Lived Experience to Improve Firearm Violence Lexicon Development

Nicole Cook[1], MPA, PhD; Frances M Biel[1], MPH; Kerry Ann Bet[1], MPH; Marion R Sills[1], MPH, MD; Ali Al Bataineh[2], PhD; Pedro Rivera[1], MS; Anna R Templeton[1], DNP; Natalie Cartwright[2], PhD

[1]OCHIN, PO Box 5426, Portland, OR, United States

[2]Norwich University, Northfield, VT, United States

Corresponding Author:
Nicole Cook, MPA, PhD
OCHIN, PO Box 5426, Portland, OR, United States

Abstract

Framing the public health burden of firearm violence should include people with secondary exposure to firearm violence beyond acute bodily injury, yet such data are limited. Electronic health record clinical notes, when leveraged through natural language processing (NLP), are a potential source of data on firearm exposure. As part of NLP lexicon development, diverse stakeholders were engaged to identify keywords, and our findings demonstrated that engaging diverse stakeholders adds valuable input to NLP development.

(JMIR Form Res 2025;9:e68105) doi:10.2196/68105

KEYWORDS

firearm violence; firearm injury; stakeholder engagement; natural language processing; electronic health record

Introduction

Exposure to firearm violence includes witnessing a shooting, being threatened by a firearm, losing a loved one to gun violence, and sustaining injuries from firearms [1]. Such exposure is associated with adverse physical and behavioral impacts. To better understand health sequelae following firearm violence exposure, some researchers recommend including these exposures in ongoing surveillance and research [2,3]. Although structured data (eg, diagnostic codes) for exposure to firearm violence are largely unavailable, unstructured electronic health record (EHR) data represent a potential source of information collected in the clinical care process [4]. To explore the applicability of such data in surveillance and research, we developed a natural language processing (NLP) pipeline for using unstructured EHR fields, including clinical notes, to identify those with exposure to firearm violence and subsequently understand more about the health impacts of such exposure. A crucial part of formative work on NLP lexicon development is effectively engaging a broad representation of stakeholders to contribute lexicon terms that might not otherwise be known to study teams [5]. In this research letter, we describe the process and outcomes of engaging stakeholders in NLP lexicon development.

Methods

Study Design

We assembled a stakeholder advisory committee (SAC) comprised of patients, advocates, clinicians, researchers, and others with lived and/or professional experience that includes exposure to firearm violence. The SAC structure was designed to include community advocates, patients with lived experience, practicing clinicians, clinical researchers, data scientists, and clinical informaticists. SAC members were recruited via prior collaboration with a study team member, word of mouth, and presentations to patients and clinicians at web-based workgroup meetings. The final SAC composition included 12 members—5 community advocates and/or patients with lived experience, 1 physician, 2 clinician researchers, 2 clinical informaticists, and 2 data scientists.

In April 2024, SAC members reviewed the lexicon of identified keywords that was developed by the study team [4] and informed by MacPhaul et al's [6] NLP study, which investigated firearm injury intent. Afterward, SAC members suggested additional terms from their lived or professional experience that may indicate exposure to firearm violence. The review of terms and suggestion of new terms were done via an asynchronous, interactive SAC meeting.

New terms identified by the SAC were searched across EHR clinical notes from 7,103,301 patients who were receiving primary or behavioral health care at community-based health centers in the OCHIN multistate network [7]. A maximum of

XSL•FO
RenderX

30 random notes per keyword were reviewed by a study team member to determine whether a note indicated exposure to firearm violence from a powder mechanism [8]. Descriptive statistics were used to summarize results in a table for review, discussion, and adjudication by the study team and SAC.

Ethical Considerations

The Norwich University institutional review board deemed this study exempt. Stakeholders agreed to guide this study and received compensation for participating in monthly meetings; they were not research participants. Stakeholders completed a professional services agreement, which outlined roles and responsibilities, and a nondisclosure agreement prior to joining the SAC. The EHR study data used by the study team for NLP development were a limited dataset. SAC members were not provided access to the limited dataset.

Results

The initial lexicon of identified keywords provided by the study team to the SAC included 19 terms, of which 13 were used in pilot work and 6 were additionally identified by the study team [4]. SAC members identified 35 additional keywords that were not included in the first iteration of the lexicon. Of the 35 terms, 27 had at least one mention in at least one clinical note. Of the 585 clinical notes reviewed, 5 contained 4 SAC-identified terms—*bbs*, *buckshot*, *firing*, and *metal pole*—and possibly indicated exposure to firearm violence. The study team met to perform the final adjudication for including or excluding each term. Two terms (*buckshot* and *firing*) were determined to have sufficient contextual information to indicate true exposure (Table 1). Notes that included these terms were added to the NLP testing and training dataset (Figure 1).

XSL•FO
RenderX

Table . Manual review of electronic health record (EHR) clinical notes (N=585) containing stakeholder advisory committee (SAC)–identified firearm lexicon terms from 300 ambulatory health care clinics.

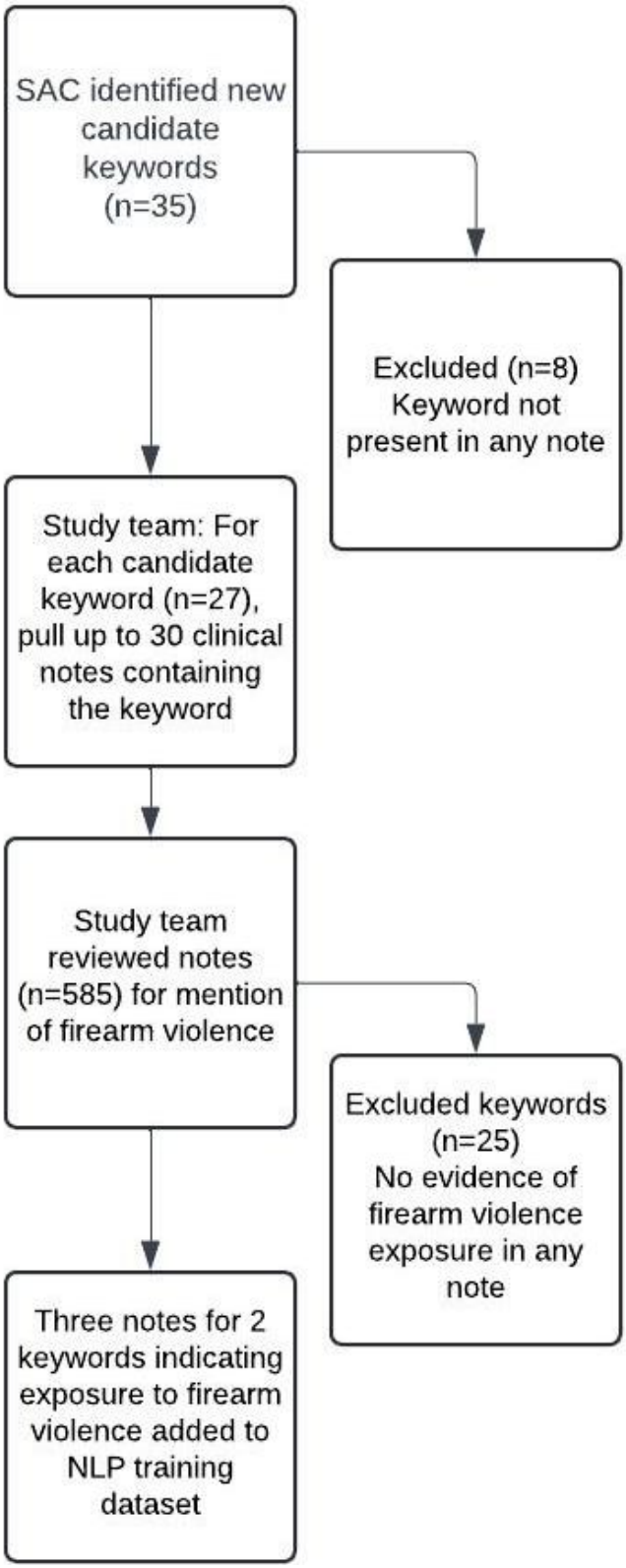
SAC-identified term	Notes in EHR with keyword, n	Clinical notes reviewed[a,b], n	Notes with possible indication of exposure to firearm violence, n	Text portion reviewed by study team to determine exposure	Final determination that reviewed clinical notes indicated firearm violence exposure
buckshot	1	1	1	• "My dad had just been shot in the back, he had buckshot in him."	Yes
firing	13	13	2	• "He reports there was a firing of fire arms at his apartment complex this last weekend." • "A few days later, the same brother was apparently responsible for firing a gun into the house; the bullet traversed two rooms and came to rest very close to the patient's son."	Yes
metal pole	2	2	1	• "Pt believes she is being tracked to be murdered, pt sees others are potential threats, as people who are trying to kill her-…reports pt threatening another wielding a metal pole"	No
bbs	30	30	1	• "Cl said he had a gun and he was not going to jail. (He had a BB gun, broken, threw it in the woods)"	No

[a]All notes for terms that had less than 30 notes were reviewed. For terms with >30 notes, a random sample of notes was reviewed.

[b]Additional stakeholder advisory committee–identified terms with clinical notes that did not indicate firearm violence exposure include *Banger*, *draco*, *semiautomatic*, *gat*, *toaster*, *drill*, *stick*, *strap*, *trigger*, *heat*, *heater*, *iron*, *metal*, *nina*, *nine*, *piece*, *pole*, *rod*, *burner*, *cap*, *carrying*, *69*, and *ammunition*.

XSL•FO
RenderX

Figure 1. Diagram of SAC keyword term review for firearm violence exposure. NLP: natural language processing; SAC: stakeholder advisory committee.



Discussion

Effectively engaging varied stakeholders in NLP lexicon development led to the identification of 2 additional terms (*buckshot* and *firing*) that were previously not considered by the study team. Novel clinical notes with these keywords were input into the training and testing dataset to enhance NLP model performance. Although most words suggested by the SAC did not indicate true exposure, the exercise demonstrated that periodic engagement of different stakeholder advisors in artificial intelligence and machine learning research is an important strategy to reduce data bias [9,10].

XSL•FO
RenderX

Acknowledgments

The authors are grateful to Angela Williams, Patricia Devine Wilder, Jennifer Obermeyer, Jelena MacLeod, and the rest of the stakeholder advisory committee for their work driving efforts toward trauma-informed care that supports collection and identification of firearm violence exposure data in patient electronic health records.

The research reported in this work was powered by PCORnet. PCORnet has been developed with funding from the Patient-Centered Outcomes Research Institute (PCORI) and conducted with the Accelerating Data Value Across a National Community Health Center Network (ADVANCE) Clinical Research Network (CRN). ADVANCE is a clinical research network in PCORnet led by OCHIN in partnership with Health Choice Network, Fenway Health, University of Washington, and Oregon Health & Science University. ADVANCE's participation in PCORnet is funded through the PCORI Award RI-OCHIN-01-MC.

This work was supported by the AIM-AHEAD Coordinating Center, funded by the National Institutes of Health (NIH). The research reported in this publication was supported by the Office of the Director, NIH Common Fund, under award number 1OT2OD032581, OTA: 1OT2OD032581. The work is solely the responsibility of the authors and does not necessarily represent the official view of AIM-AHEAD or the NIH.

Conflicts of Interest

None declared.

References

1. Schumacher S, Kirzinger A, Presiado M, Valdes I, Mollyann B. Americans' experiences with gun-related violence, injuries, and deaths. KFF. 2023 Apr 11. URL: https://www.kff.org/other/poll-finding/americans-experiences-with-gun-related-violence-injuries-and-deaths/ [accessed 2025-04-07]
2. Kaufman EJ, Delgado MK. The epidemiology of firearm injuries in the US: the need for comprehensive, real-time, actionable data. JAMA 2022 Sep 27;328(12):1177-1178. [doi: 10.1001/jama.2022.16894] [Medline: 36166012]
3. Cook N, Sills M. Tracking all injuries from firearms in the US. JAMA 2023 Feb 14;329(6):514. [doi: 10.1001/jama.2022.21997] [Medline: 36786795]
4. Cook N, Biel FM, Cartwright N, Hoopes M, Al Bataineh A, Rivera P. Assessing the use of unstructured electronic health record data to identify exposure to firearm violence. JAMIA Open 2024 Nov 4;7(4):ooae120. [doi: 10.1093/jamiaopen/ooae120] [Medline: 39498385]
5. Wan R, Kim J, Kang D. Everyone's voice matters: quantifying annotation disagreement using demographic information. In: Proceedings of the 37th AAAI Conference on Artificial Intelligence: Association for the Advancement of Artificial Intelligence; 2023:14523-14530. [doi: 10.1609/aaai.v37i12.26698]
6. MacPhaul E, Zhou L, Mooney SJ, et al. Classifying firearm injury intent in electronic hospital records using natural language processing. JAMA Netw Open 2023 Apr 3;6(4):e235870. [doi: 10.1001/jamanetworkopen.2023.5870] [Medline: 37022685]
7. Health IT and EHR solutions - community health care. OCHIN. URL: https://ochin.org/ [accessed 2025-04-07]
8. Pulcini CD, Goyal MK, De Souza HG, et al. A firearm violence research methodologic pitfall to avoid. Acad Emerg Med 2022 Sep;29(9):1140-1145. [doi: 10.1111/acem.14491] [Medline: 35332976]
9. Zou J, Schiebinger L. Ensuring that biomedical AI benefits diverse populations. EBioMedicine 2021 May;67:103358. [doi: 10.1016/j.ebiom.2021.103358] [Medline: 33962897]
10. Vishwanatha JK, Christian A, Sambamoorthi U, Thompson EL, Stinson K, Syed TA. Community perspectives on AI/ML and health equity: AIM-AHEAD nationwide stakeholder listening sessions. PLOS Digit Health 2023 Jun 30;2(6):e0000288. [doi: 10.1371/journal.pdig.0000288] [Medline: 37390116]

Abbreviations

EHR: electronic health record
NLP: natural language processing
SAC: stakeholder advisory committee

XSL•FO
RenderX

Edited by A Mavragani; submitted 29.10.24; peer-reviewed by M Elbattah, R Khoury; revised version received 18.03.25; accepted 19.03.25; published 21.04.25.

Please cite as:
Cook N, Biel FM, Bet KA, Sills MR, Al Bataineh A, Rivera P, Templeton AR, Cartwright N
Engaging Stakeholders With Professional or Lived Experience to Improve Firearm Violence Lexicon Development
JMIR Form Res 2025;9:e68105
URL: https://formative.jmir.org/2025/1/e68105
doi:10.2196/68105

© Nicole Cook, Frances M Biel, Kerry Ann Bet, Marion R Sills, Ali Al Bataineh, Pedro Rivera, Anna R Templeton, Natalie Cartwright. Originally published in JMIR Formative Research (https://formative.jmir.org), 21.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Guideline-Incorporated Large Language Model-Driven Evaluation of Medical Records Using MedCheckLLM

Marc Cicero Schubert; Stella Soyka, MD; Wolfgang Wick, MD; Varun Venkataramani, MD, PhD

[1]Department of Neurology, University Hospital Heidelberg, Im Neuenheimer Feld 400, Heidelberg, Germany

Corresponding Author:
Varun Venkataramani, MD, PhD

Abstract

Abstract: The study introduces MedCheckLLM, a large language model–driven framework that enhances medical record evaluation through a guideline-in-the-loop approach by integrating evidence-based guidelines.

(JMIR Form Res 2025;9:e53335) doi:10.2196/53335

KEYWORDS

large language models; AI; electron medical records; checklists; LLM; language model; NLP; natural language processing; records; documentation; documents; framework; conceptual; machine learning; artificial intelligence; evidence; evaluate; evaluation; guideline; health care

Introduction

Large language models (LLMs) have demonstrated enormous potential in assessing complex datasets in health care across many applications [1,2]. One underexplored area is their application for the reliable evaluation of medical documents. The automated evaluation of these documents has the potential to enhance patient safety. The system's reasoning process must be (1) transparent and comprehensible to human evaluators and (2) guided by established medical guidelines proven to increase patient safety [3].

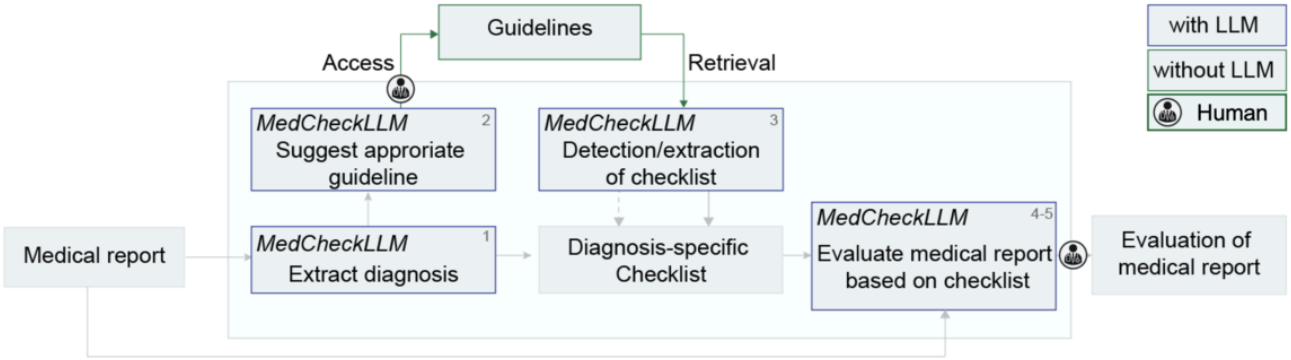
In this study, we introduce a framework that consists of a multistep approach for medical record evaluation that incorporates guidelines into the evaluation process (ie, guideline-in-the-loop). Our proposed algorithm, MedCheckLLM, is an LLM-driven, structured reasoning mechanism designed to automate the evaluation of medical records against evidence-based guidelines. The guidelines are deterministically accessed and returned to the LLM as input without further model fine-tuning. This strict separation of LLM and guidelines is expected to increase the validity and interpretability of the evaluations. The approach's step-by-step structure could improve transparency in clinical applications. The primary objective of this research is to introduce the conceptual framework and assess its feasibility.

Methods

The MedCheckLLM algorithm begins by extracting a patient's diagnosis from the medical report (Figure 1). Based on the diagnosis, it suggests an appropriate guideline. A human medical expert makes the final guideline selection. Guidelines are then accessed independently of the LLM's mechanisms using programmatically built interfaces for guideline retrieval. Subsequently, guidelines are provided as input to the LLM and are either identified as already formatted in a usable checklist or converted into a checklist. This diagnosis-specific checklist is used to assess the medical report by the LLM, with a final verification by a human medical expert. To test this approach, we used expert-validated simulated medical reports (simulated dataset) and physician-generated medical reports (physician dataset). Performance was analyzed for patient histories with headaches using guidelines from the International Headache Society and the physician dataset for four further neurological diagnoses (ie, border zone infarction, meningitis, neuromyelitis optica, and subarachnoid haemorrhage). The validity of this method was further analyzed by evaluating generated doctor's notes with a correct diagnosis compared to doctor's notes with a false diagnosis. The LLMs, GPT-4 and Claude-3 were used for testing (see Multimedia Appendix 1).

XSL•FO
RenderX

Figure 1. Workflow of MedCheckLLM. A medical report including medical history, diagnosis and treatment is provided as input. First, the LLM identifies the given diagnosis. Second, it suggests a medical guideline for evaluation of the medical report, with a human medical expert making the final selection. Then, independently of the LLM, the selected guideline is accessed, and diagnosis-specific text is extracted and inputted into the LLM. Subsequently, the LLM determines whether the input guideline text is in checklist format; if not, it extracts a checklist. Using this diagnosis-specific checklist, the LLM evaluates the medical report based on the diagnosis-specific checklist. Finally, a human expert assesses the LLM evaluation. Dashed arrow: Checklist extraction instead of detection of checklist format. Blue box: Component uses an LLM. Green Box: Components do not use an LLM.



Results

We evaluated the medical report analysis conducted by MedCheckLLM for various headache diagnoses. In the simulated dataset, MedCheckLLM (based on GPT-4 and Claude-3, Table 1) extracted the specified diagnosis correctly in 100% of cases from a list of 61 possible diagnoses from The International Classification of Headache Disorders-3 [4]. The model suggested existing evidence-based guidelines in 70.59% (12/17) of medical reports and detected the format of the guidelines as checklists in 100% of the cases (N=17). MedCheckLLM accurately evaluated 87% (67/ 77) of checklist items. Performance on the physician dataset showed an accurate evaluation in 77.4% (24/ 31) of checklist items (Table 2). It identified incorrect diagnoses where the stated diagnosis did not align with the content of the doctor's letters in 94.1% (16/17) of cases, while it correctly recognized 100% (N=17) of letters with matching diagnoses.

Table . Performance of MedCheckLLM on the simulated dataset.

Elements of algorithmic structure	GPT-4 performance, % (n/N)	Claude-3 performance, % (n/N)	Explanation of specific task of each element
Extracting stated diagnosis	100 (17/17)	100 (17/17)	Extract the diagnosis that is stated in the medical report
Suggestion of existing guidelines	70.6 (12/17)	58.8 (10/17)	Suggest a guideline that should be used to evaluate the medical report
Detection of checklist	100 (17/17)	100 (17/17)	Detect whether the accessed guidelines are in a structured checklist-criteria format
Evaluation of diagnostic criteria (checklist items)	87 (67/77)	83.8 (62/74)	Assess whether the criteria listed in the checklist are met in the medical report
Evaluation of letters with correct diagnosis (clinical descriptions and diagnosis match)	100 (17/17)	94.1 (16/17)	Assess whether the diagnosis stated in the medical report aligns with the clinical descriptions
Evaluation of letters with false diagnosis (clinical descriptions and diagnosis do not match)	91.4 (16/17)	91.4 (16/17)	Evaluate whether the diagnosis that is stated in the medical report fits the clinical descriptions

XSL•FO
RenderX

Table . Performance of MedCheckLLM on the physician dataset.

Element of algorithmic structure	Stroke	Meningitis	Neuromyelitis optica	Subarachnoid hemorrhage
Extracting stated diagnosis	Yes	Yes	Yes	Yes
Suggestion of existing guidelines[a]	Yes, applicable	Yes, partially applicable	Yes, applicable	Yes, partially applicable
Creation of checklist, level of detail[a]	Yes, moderate detail	Yes, moderate detail	Yes, thorough detail	Yes, minimal detail
Evaluation of diagnostic criteria, % (n/N)	100 (7/7)	66.7 (4/6)	87.5 (7/8)	60 (6/10)

[a]The responses were classified as yes, and partially applicable,applicable, or minimal, moderate, or thorough detail.

Discussion

The framework of MedCheckLLM represents a promising approach for a comprehensive, guideline-anchored review of electronic health records. It holds the potential to function as a quality assurance framework throughout patient care due to its advantages of separate partitioning of the LLM and the guidelines, rather than training guidelines into an LLM. The flexibility of this approach allows for immediate implementation of guideline updates or the option to implement customized protocols for subgroups of patients. Due to the checklist-based approach, each item can be verified individually, thus increasing the algorithm's interpretability, which is crucial in health care settings [5]. Due to the LLM's subpar guideline suggestion capability, medical experts are integrated at this step to ensure that established guidelines are used. Further research is essential to advance the development of LLM-driven methods for extracting checklists from unstructured guidelines, as well-structured guidelines are crucial for detailed, high-quality checklists. Further, this framework facilitates improved data mining practices in electronic health records [6]. In the future, it is crucial to address privacy concerns to ensure the ethical application of these powerful tools in real-world clinical settings [7-9].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Large language models used in this study.
[PDF File, 68 KB - formative_v9i1e53335_app1.pdf]

References

1. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit Health 2023 Feb;2(2):e0000198. [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
2. Schubert MC, Wick W, Venkataramani V. Performance of Large Language Models on a Neurology Board-Style Examination. JAMA Netw Open 2023 Dec 1;6(12):e2346721. [doi: 10.1001/jamanetworkopen.2023.46721] [Medline: 38060223]
3. Thomassen O, et al. The effects of safety checklists in medicine: a systematic review. Acta Anaesthesiol Scand 2014 Jan;58(1):5-18. [doi: 10.1111/aas.12207] [Medline: 24116973]
4. Headache Classification Committee of the International Headache Society (IHS) The International Classification of Headache Disorders, 3rd edition. Cephalalgia 2018 Jan;38(1):1-211. [doi: 10.1177/0333102417738202] [Medline: 29368949]
5. Amann J, et al. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 2020 Nov 30;20(1):310. [doi: 10.1186/s12911-020-01332-6] [Medline: 33256715]
6. Jiang LY, Liu XC, Nejatian NP, et al. Health system-scale language models are all-purpose prediction engines. Nature New Biol 2023 Jul;619(7969):357-362. [doi: 10.1038/s41586-023-06160-y] [Medline: 37286606]
7. Meskó B. The Impact of Multimodal Large Language Models on Health Care's Future. J Med Internet Res 2023 Nov 2;25:e52865. [doi: 10.2196/52865] [Medline: 37917126]
8. Dorr DA, Adams L, Embí P. Harnessing the Promise of Artificial Intelligence Responsibly. JAMA 2023 Apr 25;329(16):1347-1348. [doi: 10.1001/jama.2023.2771] [Medline: 36972068]
9. Wang C, Liu S, Yang H, Guo J, Wu Y, Liu J. Ethical Considerations of Using ChatGPT in Health Care. J Med Internet Res 2023 Aug 11;25:e48009. [doi: 10.2196/48009] [Medline: 37566454]

XSL•FO
RenderX

Abbreviations

LLM: large language model

Edited by A Mavragani; submitted 03.10.23; peer-reviewed by D Chrimes, P Han; revised version received 02.11.24; accepted 17.11.24; published 24.04.25.

Please cite as:
Schubert MC, Soyka S, Wick W, Venkataramani V
Guideline-Incorporated Large Language Model-Driven Evaluation of Medical Records Using MedCheckLLM
JMIR Form Res 2025;9:e53335
URL: https://formative.jmir.org/2025/1/e53335
doi:10.2196/53335

© Marc Cicero Schubert, Stella Soyka, Wolfgang Wick, Varun Venkataramani. Originally published in JMIR Formative Research (https://formative.jmir.org), 24.4.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Research Letter

Exploring Public Sentiment on the Repurposing of Ivermectin for COVID-19 Treatment: Cross-Sectional Study Using Twitter Data

Angga Prawira Kautsar[1,2,3*], MBA; Rano Kurnia Sinuraya[1,3,4*], PhD; Jurjen van der Schans[1,5], PhD; Maarten Jacobus Postma[1,3,5], PhD; Auliya A Suwantika[3,4,6], PhD

[1]Unit of Global Health, Department of Health Sciences, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
[2]Department of Pharmaceutics and Pharmaceutical Technology, Faculty of Pharmacy, Universitas Padjadjaran, Sumedang, Indonesia
[3]Center of Excellence in Higher Education for Pharmaceutical Care Innovation, Universitas Padjadjaran, Bandung, Indonesia
[4]Department of Pharmacology and Clinical Pharmacy, Faculty of Pharmacy, Universitas Padjadjaran, Sumedang, Indonesia
[5]Department of Economics, Econometrics and Finance, Faculty of Economics and Business, University of Groningen, Groningen, Indonesia
[6]Center for Health Technology Assessment, Universitas Padjadjaran, Universitas Padjadjaran, Bandung, Indonesia
[*]these authors contributed equally

Corresponding Author:
Angga Prawira Kautsar, MBA
Unit of Global Health, Department of Health Sciences
University Medical Center Groningen
University of Groningen
Antonius Deusinglaan 1
Groningen, 9713 AV
The Netherlands
Phone: 31 0503611111
Email: angga.prawira@unpad.ac.id

Abstract

A sentiment analysis of 5051 Twitter posts from January 2022 found that 53.4% of them expressed positive views on ivermectin as a COVID-19 treatment, 35.6% of them were neutral, and 11% of them were negative, highlighting the polarized public perception and the need for careful interpretation of social media data in health communication.

(JMIR Form Res 2025;9:e50536) doi:10.2196/50536

KEYWORDS

COVID-19; ivermectin; sentiment analysis; Twitter; social media; public health; misinformation; geolocation analysis

Introduction

As the COVID-19 pandemic evolves, the scientific community confronts the limitations of vaccines due to emerging viral mutations that potentially decrease vaccine efficacy [1], thus necessitating a parallel investigation into additional therapeutic agents. Ivermectin, a well-established, safe drug, has emerged as a repurposed drug candidate due to preliminary studies suggesting its antiviral properties against SARS-CoV-2 in vitro [1,2]. Nonetheless, the scientific debate remains vigorous, with discussions on the drug's appropriate formulation and dosing for potential COVID-19 prophylaxis and treatment [3].

Simultaneously, ivermectin's role has garnered widespread attention on social media, reflecting the public's quest for alternative treatments. Twitter (now X), a hub for real-time public discourse, has become a fertile ground for divergent views on COVID-19 treatment [4,5]. This sentiment analysis focuses on Twitter discussions about ivermectin, showing public opinion that, while not devoid of misinformation risks, these discussions offer an alternative lens to understand the societal pulse on this contentious topic [6]. By examining the sentiments expressed on Twitter, we aim to add nuance to the ongoing discourse, acknowledging the platform's influence on public perception and its implications for health communication strategies.

Methods

Overview

This cross-sectional examined the COVID-19–related sentiments on ivermectin using Twitter data. Primary data were collected from Twitter posts about ivermectin worldwide from January 15 to 22, 2022. We searched posts with the keyword

XSL•FO
RenderX

"ivermectin" and retrieved raw data with various variables. Data cleaning involved lowercasing capital letters, eliminating retweet symbols, removing punctuation marks, and other preprocessing steps to ensure data accuracy [6]. People's sentiments were determined by creating a corpus (body of text) and loading a lexicon dictionary based on the positive and negative words. The sentiment score was calculated with the Bing method using a range of –6 to +6 and considered 3 sentiment types: positive, neutral, and negative [7,8]. The Bing lexicon was chosen for its proven effectiveness and simplicity in extensive dataset analysis. Frequency analysis of single terms (unigrams), word pairs (bigrams), and bigram networks helped identify frequently mentioned terms and explore relationships between words [7,8]. Despite its simplicity, this approach is as robust as more complex methods, as it accurately categorizes sentiments and identifies patterns through unigram and bigram frequency analysis [6]. This computational analysis–focused methodology offers valuable insights into public sentiment, complementing traditional clinical evaluation without extensive statistical validation. Data were mined using Twitter's limited application programming interface (version 2) and analyzed using RStudio (R Foundation) and relevant packages for visualization.

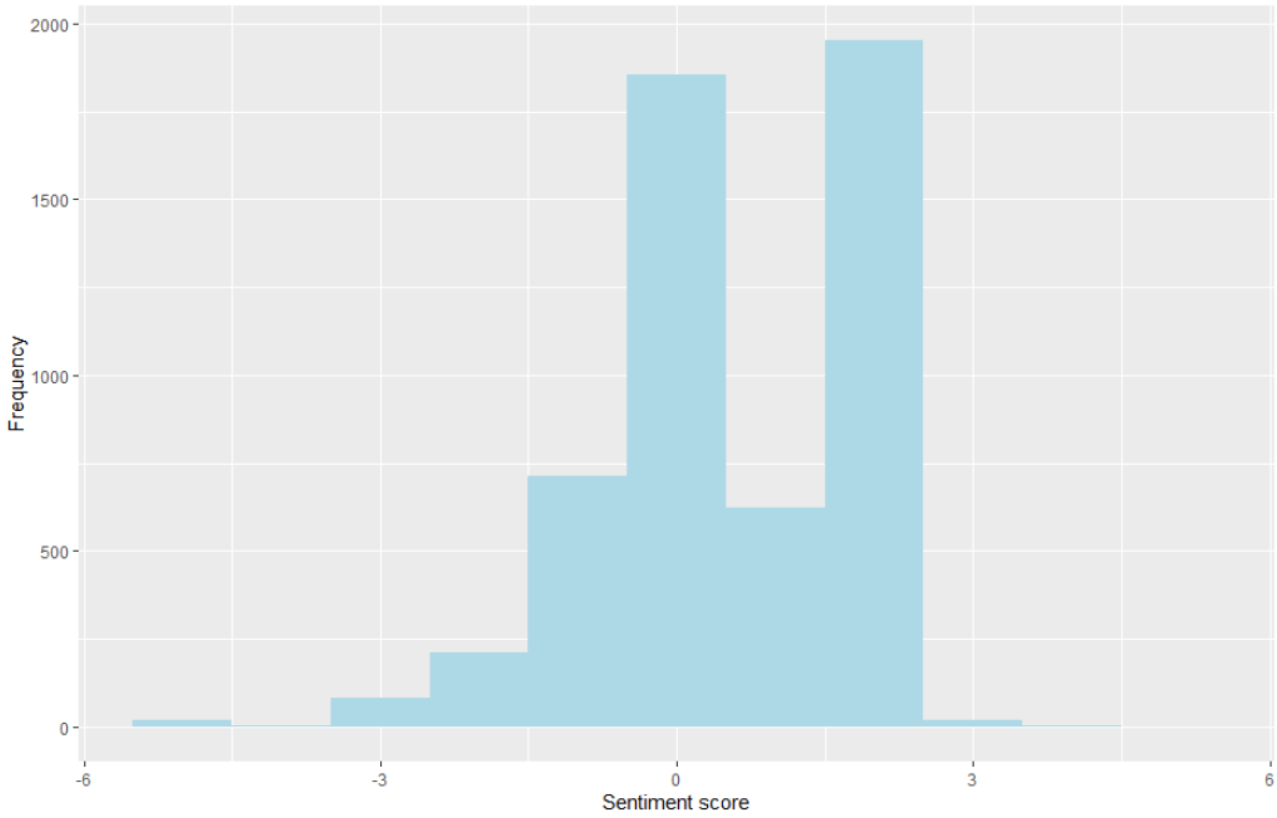
Ethical Considerations

There was no direct connection with Twitter users. We removed all individual data, usernames, IDs, and tweets in the manuscript and supporting material. Therefore, ethical approval was not required.

Results

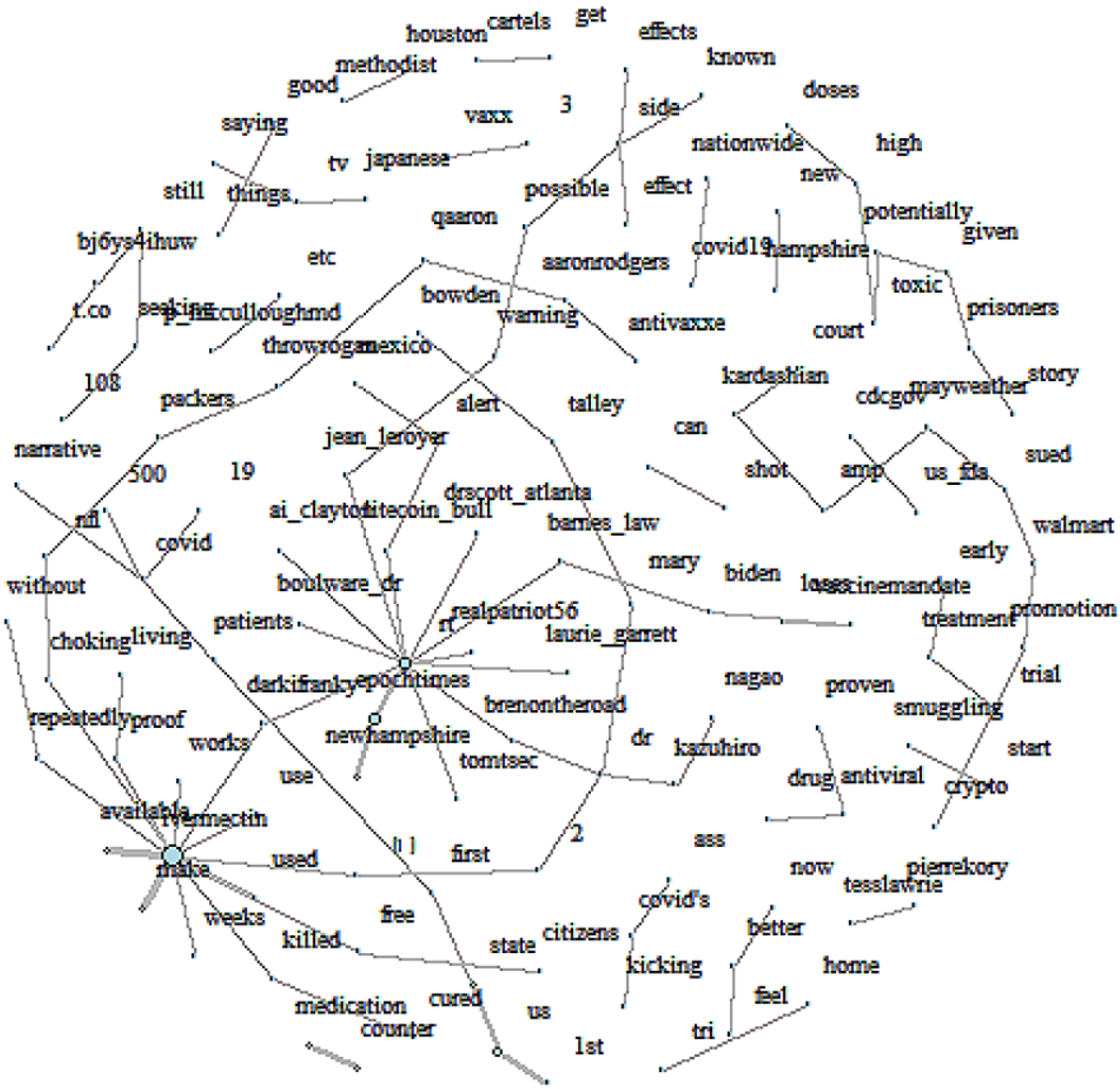
In total, 5051 ivermectin-related tweets underwent sentiment analysis, revealing a prevailing positive sentiment (53.4%), followed by neutral (35.6%) and negative (11%) sentiments among the analyzed tweets (Figure 1). The analysis identified frequently mentioned positive terms such as "medication," "potentially," and "treatment," while negative terms like "poised," "smuggling," and "toxic" were less common. Notably, analysis of word pairings unveiled strong associations with positive sentiments, further supporting the notion of ivermectin's potential efficacy in COVID-19 treatment. Examining these word pairings facilitated in-depth exploration of sentiment patterns within the bigram network, revealing significant connections and relationships in the analyzed tweets (Figure 2).

Figure 1. Distribution of sentiment score.



XSL•FO
RenderX

Figure 2. Bigram count network related to tweets ivermectin.



Discussion

Principal Findings

Our analysis shows that tweets about repurposing ivermectin for COVID-19 treatment are predominantly positive. Terms like “medication,” “potentially,” and “treatment” frequently appeared in positive contexts, reinforcing this finding. Phrases such as “ivermectin works,” “available ivermectin,” and “ivermectin medication” were strongly associated with positive sentiment. These results align with previous findings showing ivermectin’s popularity on Twitter for COVID-19 treatment [4,5]. However, some negative sentiment was observed, particularly concerning warnings from the Food and Drug Administration and the limited clinical evidence supporting ivermectin’s efficacy in COVID-19 prevention and treatment [9].

Twitter data have proven valuable in monitoring public responses to the COVID-19 pandemic, as evidenced by a study of millions of SARS-CoV-2–related tweets [10]. This study not only identified dominant topics like new cases, death rates, and preventive measures, but also explored the geographic distribution of sentiments through tweet-embedded geolocation data (code available in Multimedia Appendix 1).

Limitations

Of note, the findings are based on Twitter data, which may not represent the entire population’s sentiments. Therefore, researchers conducted cross-validation of the sources.

Conclusions

This sentiment analysis highlights the polarized perception of ivermectin in COVID-19–related discourse and reflects broader public health debates. Given the regulatory advisories and the limited clinical evidence, these public sentiments must be interpreted cautiously.

XSL•FO
RenderX

Acknowledgments

We acknowledge receiving the Ministry of Higher Education, Science, and Technology Overseas Postgraduate Education scholarship (DIKTI BPP-LN grant T/915/D3.2/KD.02.01/2019) awarded to AK. Despite this financial support, we have diligently maintained transparency and objectivity in our work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Code in R.
[ZIP File (Zip Archive), 5 KB - formative_v9i1e50536_app1.zip]

References

1. Jiang S. Don't rush to deploy COVID-19 vaccines and drugs without sufficient safety guarantees. Nature 2020 Mar 16;579(7799):321-321. [doi: 10.1038/d41586-020-00751-9] [Medline: 32179860]
2. Caly L, Druce JD, Catton MG, Jans DA, Wagstaff KM. The FDA-approved drug ivermectin inhibits the replication of SARS-CoV-2 in vitro. Antiviral Res 2020 Jun;178:104787 [FREE Full text] [doi: 10.1016/j.antiviral.2020.104787] [Medline: 32251768]
3. The WP. NIH trial may settle debate over ivermectin as a covid-19 treatment. The Washington Post. URL: https://www.washingtonpost.com/health/2021/04/08/ivermectin-covid-drug/ [accessed 2023-03-06]
4. de Melo T, Figueiredo CMS. Comparing news articles and tweets about COVID-19 in Brazil: sentiment analysis and topic modeling approach. JMIR Public Health Surveill 2021 Feb 10;7(2):e24585 [FREE Full text] [doi: 10.2196/24585] [Medline: 33480853]
5. Garcia K, Berton L. Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. Appl Soft Comput 2021 Mar;101:107057 [FREE Full text] [doi: 10.1016/j.asoc.2020.107057] [Medline: 33519326]
6. Hill S, Scott R. Developing an Approach to Harvesting, Cleaning, and Analyzing Data from Twitter Using R. Inf Sys Educ J 2017;15(3):15.
7. Liu B, Hu M, Cheng J. Opinion observer: analyzing and comparing opinions on the Web. 2005 Presented at: WWW '05: Proceedings of the 14th international conference on World Wide Web; May 10-14, 2005; Chiba, Japan p. 342-351. [doi: 10.1145/1060745.1060797]
8. Liu B. Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge, United Kingdom: Cambridge University Press; 2020.
9. Why You Should Not Use Ivermectin to Treat or Prevent COVID-19. Food and Drug Administration. URL: https://nycourts.gov/reporter/webdocs/Why-You-Should-Not-Use-Ivermectin-to-Treat-or-Prevent-COVID-19.pdf [accessed 2022-03-01]
10. Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, et al. Twitter discussions and emotions about the COVID-19 pandemic: machine learning approach. J Med Internet Res 2020 Nov 25;22(11):e20550 [FREE Full text] [doi: 10.2196/20550] [Medline: 33119535]

Edited by A Mavragani; submitted 05.07.23; peer-reviewed by R Gore, J Plasek, A Naser, B Cao; comments to author 16.11.23; revised version received 12.02.24; accepted 09.09.24; published 27.03.25.

Please cite as:
Kautsar AP, Sinuraya RK, van der Schans J, Postma MJ, Suwantika AA
Exploring Public Sentiment on the Repurposing of Ivermectin for COVID-19 Treatment: Cross-Sectional Study Using Twitter Data
JMIR Form Res 2025;9:e50536
URL: https://formative.jmir.org/2025/1/e50536
doi:10.2196/50536
PMID:

©Angga Prawira Kautsar, Rano Kurnia Sinuraya, Jurjen van der Schans, Maarten Jacobus Postma, Auliya A Suwantika. Originally published in JMIR Formative Research (https://formative.jmir.org), 27.03.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Monitoring Nystagmus in a Patient With Vertigo Using a Commercial Mini-Infrared Camera and 3D Printer: Cost-Effectiveness Evaluation and Case Report

Hiroyuki Sakazaki[1], SLP, MPAS; Masao Noda[1,2*], MD, MBA, PhD; Yumi Dobashi[1], MT; Tatsuaki Kuroda[1,3], MD; Reiko Tsunoda[1], MD, PhD; Hiroaki Fushiki[1*], MD, PhD

[1]Department of Otolaryngology, Mejiro University Ear Institute Clinic, 320 Ukiya, Iwatsuki-ku, Saitama-shi, Saitama, Japan

[2]Department of Otolaryngology and Head and Neck Surgery, Jichi Medical University, Tochigi, Japan

[3]Kuroda ENT Clinic, Yatsushiro-shi, Kumamoto, Japan

[*]these authors contributed equally

Corresponding Author:
Masao Noda, MD, MBA, PhD
Department of Otolaryngology, Mejiro University Ear Institute Clinic, 320 Ukiya, Iwatsuki-ku, Saitama-shi, Saitama, Japan

Abstract

Background: Observing eye movements during episodic vertigo attacks is crucial for accurately diagnosing vestibular disorders. In clinical practice, many cases lack observable symptoms or clear findings during outpatient examinations, leading to diagnostic challenges. An accurate diagnosis is essential for timely treatment, as conditions such as benign paroxysmal positional vertigo (BPPV), Ménière's disease, and vestibular migraine require different therapeutic approaches.

Objective: This study aimed to develop and evaluate a cost-effective diagnostic tool that integrates a mini-infrared camera with 3D-printed goggles, enabling at-home recording of nystagmus during vertigo attacks.

Methods: A commercially available mini-infrared camera (US $25) was combined with 3D-printed goggles (US $13) to create a system for recording eye movements in dark conditions. A case study was conducted on a male patient in his 40s who experienced recurrent episodic vertigo.

Results: Initial outpatient evaluations, including oculomotor and vestibular tests using infrared Frenzel glasses, revealed no spontaneous or positional nystagmus. However, with the proposed system, the patient successfully recorded geotropic direction-changing positional nystagmus during a vertigo attack at home. The nystagmus was beating distinctly stronger on the left side down with 2.0 beats/second than the right side down with 1.2 beats/second. Based on the recorded videos, a diagnosis of lateral semicircular canal-type BPPV was made. Treatment with the Gufoni maneuver effectively alleviated the patient's symptoms, confirming the diagnosis. The affordability and practicality of the device make it particularly suitable for telemedicine and emergency care applications, enabling patients in remote or underserved areas to receive accurate diagnoses.

Conclusions: The proposed system demonstrates the feasibility and utility of using affordable, accessible technology for diagnosing vestibular disorders outside of clinical settings. By addressing key challenges, such as the absence of symptoms during clinical visits and the high costs associated with traditional diagnostic tools, this device offers a practical solution for real-time monitoring and accurate diagnosis. Its potential applications extend to telemedicine, emergency settings, and resource-limited environments. Future iterations that incorporate higher-resolution imaging and automated analysis could further enhance its diagnostic capabilities and usability across diverse patient populations.

(JMIR Form Res 2025;9:e70015) doi:10.2196/70015

KEYWORDS

dizziness; vertigo; smartphone; BPPV; telemedicine; 3D-printer

Introduction

Vertigo is a common ailment experienced at all ages, with an estimated lifetime prevalence of 3% - 10% [1]. It is symptomatic of peripheral vestibular, central, and psychogenic disorders and circulatory dysfunction, as well as life-threatening conditions, such as cerebrovascular disease, making accurate diagnosis clinically significant [2,3]. Even when not life-threatening, episodic vertigo attacks accompanied by nausea and vomiting can substantially impair an individual's quality of life [4].

Recently, the use of telemedicine has expanded across various fields owing to the rapid spread of information and communication technology, the widespread use of smartphones and tablets, and advancements in data communication

XSL•FO
RenderX

technology [5]. The COVID-19 epidemic further accelerated the adoption of telemedicine in clinical care [6]. Consequently, the use of smartphones for telemedicine in vertigo treatment has increased [7-10]. For example, Shah et al [11] used smartphone-recorded eye movement videos to diagnose benign paroxysmal positional vertigo (BPPV). Additionally, we developed a recording and analysis system for video head impulse testing using smartphones [12]; this system was found to be effective when field physicians collaborated with remote specialists, especially in emergencies.

Observing eye movements during vertigo attacks is critical for accurately diagnosing recurrent vestibular disorders [13-15]. Young et al [15] demonstrated that patient-initiated monitoring of vestibular events is feasible and can lead to rapid and accurate diagnoses of episodic vestibular disorders. The characteristics of nystagmus observed in conditions such as Ménière's disease, vestibular migraine, and BPPV can aid in diagnosis, even in remote settings. This approach is also valuable in emergency departments, bridging the gap between frontline physicians and specialists [12]. Continuous monitoring of nystagmus may help differentiate between Ménière's disease and vestibular migraine based on variations in nystagmus duration [14]. However, in clinical practice, symptoms often resolve before examination, making diagnosis challenging. Independent at-home recording of eye movements during vertigo attacks provides critical diagnostic information, benefiting both patients and physicians. Although smartphone recordings have been used to diagnose conditions such as Ménière's disease [16], they may not always adequately capture nystagmus due to suppression by the patient's gaze. Phillips et al [17] demonstrated a case in which the continuous ambulatory vestibular assessment (CAVA) device, which does not require a smartphone, successfully recorded nystagmus throughout an entire Ménière's attack. Their study highlighted the device's utility in predicting vertigo attack onset and diagnosing Ménière's disease [17].

By integrating a commercially available mini-infrared camera with 3D-printed goggles, we devised a method that allows patients to record eye movements during vertigo attacks in darkness at home and share the recordings with their physicians. Herein, we report a case demonstrating the diagnostic utility of this method.

Methods

Ethical Considerations

This study was approved by the Medical Research Ethics Committee of Mejiro University (approval number Medical 21 - 005). Written informed consent was obtained from the patient after providing a sufficient explanation regarding the protection of personal information. All data collected in this study were anonymized, ensuring that no personally identifiable information was included in the analysis or publication. CARE (Case Report) EQUATOR (Enhancing the Quality and Transparency of Health Research) guidelines were followed. No financial or material compensation was provided to the participant.

Recruitment Procedures

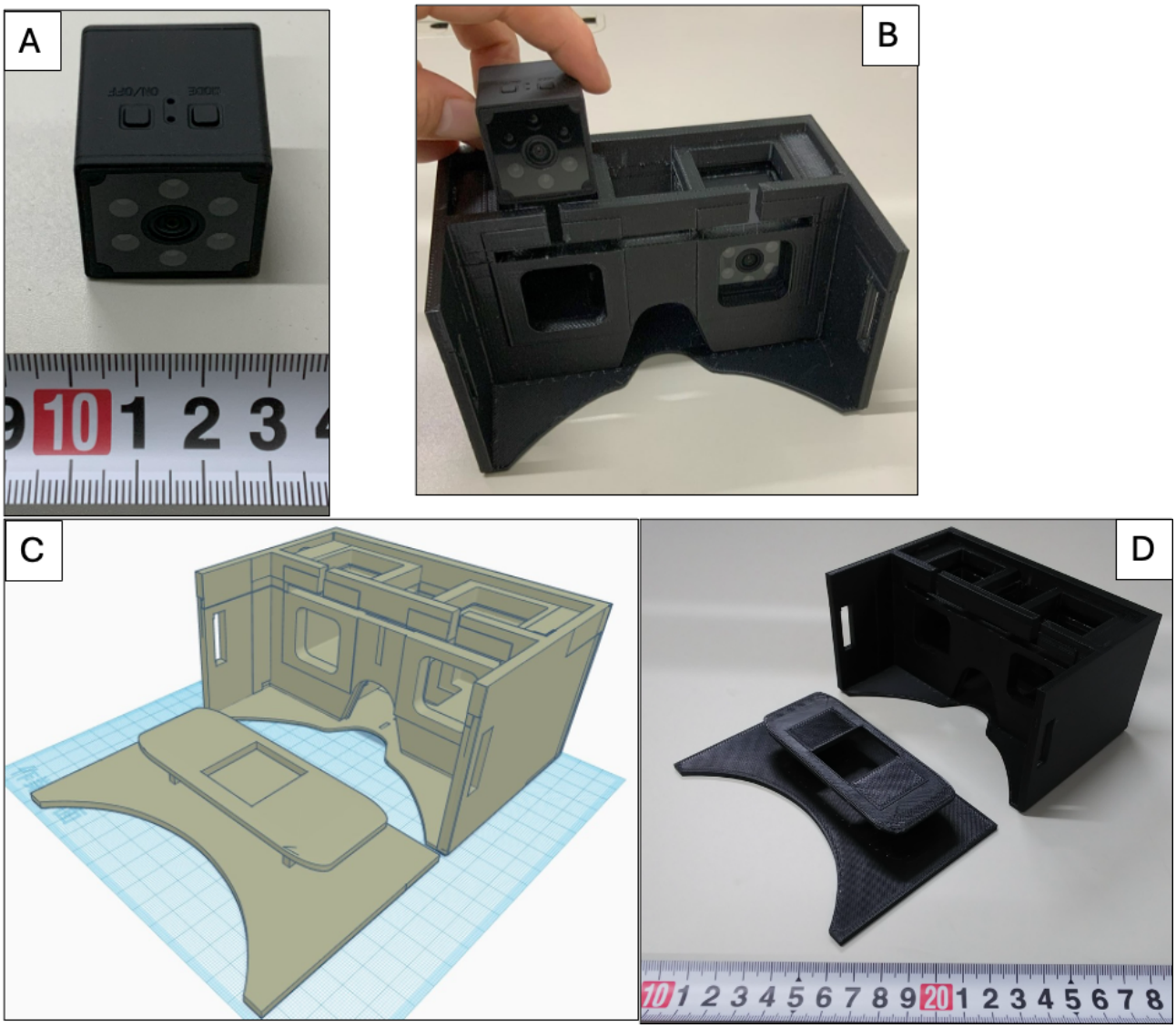
Participants for this study were recruited from the otolaryngology outpatient clinic at Mejiro University, a tertiary center specializing in the diagnosis and treatment of vestibular disorders. The patient included in this case report was referred to the clinic after experiencing recurrent episodes of vertigo that had gone undiagnosed in a primary care setting. During the recruitment process, patients presenting with vertigo as their primary symptom were screened based on their medical history, physical examination, and routine vestibular tests. Those who were unable to obtain a definitive diagnosis through standard diagnostic procedures, such as infrared video Frenzel glasses or vestibular function tests, were considered eligible for inclusion. The patient selected for this study provided written informed consent after receiving a detailed explanation of the study's objectives, the operation of the device, and measures to protect personal information.

Development of the Device

A mini-infrared camera equipped with an infrared imaging mode was selected to capture spontaneous nystagmus under low-light conditions (Wireless Mini Camera Model: Q15; Shenzhen Aobo Technology Co, Ltd; Figure 1A). The mini-infrared camera, costing approximately US $25, boasts a resolution of 1920×1080 pixels, a shooting angle of 150°, video encoding of H.264, and dimensions of 32.5 mm in width, 33.0 mm in height, and 34.2 mm in depth. Using Wi-Fi to start and stop recordings, patients can simultaneously and conveniently produce video and audio recordings using their smartphones.

XSL•FO
RenderX

Figure 1. The mini-infrared camera and handmade goggles designed for recording eye movements during vertigo attacks. This system was developed as part of a study to enhance diagnostic support for patients with episodic vertigo at Mejiro University Ear Institute Clinic. (A) A mini-infrared camera designed for portability and affordability. (B) Goggles featuring slots to securely attach the wireless mini-infrared camera on each side, ensuring stable recording during vertigo episodes. (C) A computer-aided design (CAD) diagram of the goggles, optimized for 3D printing compatibility. (D) The final output of the goggles produced using a 3D printer with a filament.



Slots were placed inside the goggles on the left and right sides, and the center of the mini-infrared camera was inserted in the slot aligned with the line of sight (Figure 1B). Both eyes were simultaneously photographed using 2 mini-infrared cameras. We designed the goggles using Tinkercad, a web-based 3D modeling tool, and an AFINIA H+1 3D printer (Microboards Technology Inc; Figure 1C). Black filament AFINIA ABS Premium Plus (Microboards Technology Inc) was used for printer output to preserve the darkness (Figure 1D). Eighteen pairs of goggles were produced, with filament costs of approximately US $13 per pair. The construction data for the 3D printer are available for download [18]. These data can be used to create goggles compatible with other mini-infrared cameras.

Operation of the Device

To operate the mini-infrared camera, the patient installed an application (HDSPCAM APP) on his smartphone or tablet. He paired the smartphone/tablet with the mini-infrared camera and selected the infrared imaging mode. During vertigo attacks, he placed the mini-infrared camera in the goggles, pressed the start button on the paired smartphone or tablet, and placed the goggles on his face. The captured images were automatically stored via Wi-Fi on the patient's smartphone or tablet (Figure 2).

The handmade goggles were used by the patient to record eye movements during vertigo episodes at home under dark conditions. The videos were wirelessly transferred to and stored on a smartphone for later analysis by physicians. This system was used in a case study of a male patient in his 40s with recurrent episodic vertigo.

XSL•FO
RenderX

Figure 2. Use of the handmade goggles.

Results

Case Presentation

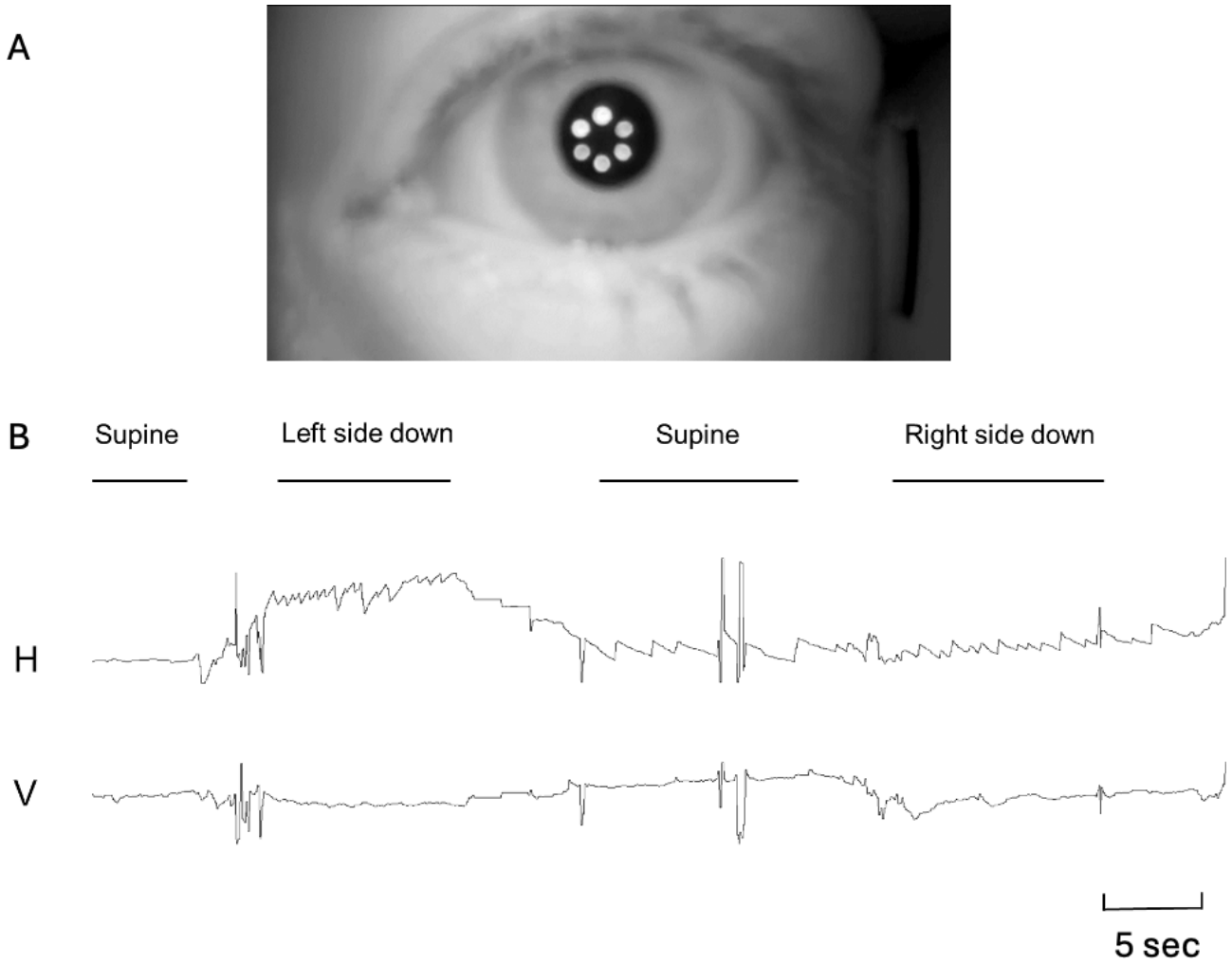
A man in his 40s with episodic vertigo visited our clinic. Initial assessments using infrared video Frenzel glasses showed no spontaneous or positional nystagmus, and oculomotor-vestibular function tests revealed no abnormalities; therefore, a definitive diagnosis was not reached. We rented him our self-made 3D-printed goggles and a mini-infrared camera to record the patient's eye movements during vertigo. We also asked him to verbally record his posture when he changed positions (eg, supine, lower left, supine, and lower right).

Assessment Results

On reviewing the videos, geotropic direction-changing positional nystagmus was observed, and lateral semicircular canal-type BPPV was diagnosed (Figure 3). The nystagmus in the first 5 seconds after the change in head position was beating distinctly stronger on the left side down with 2.0 beats/second than on the right side down with 1.2 beats/second. The affected ear was identified from the nystagmus's appearance, direction, and intensity in different positions using the audio recordings. Treatment with the Gufoni maneuver alleviated the vertigo and nystagmus.

XSL•FO
RenderX

Figure 3. Eye movements recorded by the patient during a vertigo attack. Geotropic direction-changing positional nystagmus was observed. (**A**) A still image extracted from the video recording, captured in a dark environment to avoid nystagmus suppression. (**B**) Because video clips cannot be shown in this paper, we instead present the waveforms of the eye movements. The ocular position was measured using the corneal reflex method to illustrate the recorded nystagmus. OpenCV (Intel Corp) was used for image processing. H: horizontal component. V: vertical component.



Discussion

Principal Findings

Episodic vertigo attacks often pose diagnostic challenges due to their transient nature and the absence of symptoms during clinical examinations. In this study, we developed a cost-effective diagnostic tool that combines a mini-infrared camera with 3D-printed goggles, enabling patients to record eye movements during vertigo attacks in dark conditions at home. Using this system, we successfully diagnosed lateral semicircular canal-type BPPV in a patient with recurrent vertigo, which had previously gone undiagnosed in standard outpatient evaluations. This approach demonstrates the feasibility and utility of our device in providing accurate diagnostic support for patients experiencing episodic vertigo. The system is also effective for patients whose vertigo symptoms resolve before clinical visits, allowing for the capture of critical diagnostic information at home during active episodes.

The goggles were printed with an adhesive filament to secure the mini-infrared camera and maintain their shape despite their complex structure. However, each printing required approximately 10 hours. Optimizing the design of the goggles' interior and improving printer performance might shorten production time. Alternatively, eye movements during vertigo attacks can be captured via video recording using a smartphone. Using this method, Kıroğlu and Dağkıran [16] diagnosed Ménière's disease in patients with vertigo attacks. For vertigo diagnosis, confirming spontaneous nystagmus under nongazing conditions is crucial [19]. However, when recording with a smartphone, nystagmus may be suppressed in bright conditions, as people tend to gaze at ambient light or the lens of the mini-infrared camera, making it difficult to capture eye movements in the dark. Melliti et al [20] attempted to record nystagmus using a smartphone lens adapter but reported that only approximately half (49%) of the patients successfully recorded nystagmus during the study period. Additionally, the study highlighted that significant time and effort were required to adequately explain the device's usage to patients. In contrast, the device used in this study allows for recording in dark conditions, overcoming the limitations of ambient light and enabling more reliable recordings for a wider range of patients. This feature holds promise for improving diagnostic outcomes by providing clear recordings of nystagmus during vertigo attacks. While Phillips et al [14] introduced an excellent system for continuous monitoring to investigate the pathophysiology

XSL•FO
RenderX

of Ménière's disease and vestibular migraine, practical clinical settings often require affordable and accessible tools, such as our system, to differentiate between BPPV and spontaneous vertigo.

In addition to its affordability compared to conventional diagnostic tools, the system demonstrated in this study offers significant benefits in terms of accessibility and usability. The mini-infrared camera used in this device costs approximately US $25, while the 3D-printed goggles can be produced for US $13 per unit. These low costs make the device not only economical for health care systems but also user-friendly for patients, allowing them to independently record diagnostic findings during vertigo attacks. This accessibility enables the capture of observations that might not have been possible with traditional equipment, particularly for patients who face challenges in attending repeated clinical visits. Such findings are crucial for early diagnosis and timely treatment, improving patient outcomes and reducing the burden on health care systems. Similarly, studies on comparable systems, such as the CAVA system, have reported cost savings of between £631 (US $791) and £1305 (US $1635.90) per patient by eliminating unnecessary tests and streamlining the diagnostic process [21]. This highlights the potential of community-based, patient-centered diagnostic tools to transform vertigo care, making it more efficient and accessible for all.

Previous studies used a deep learning system [22] and a smartphone application [23] to detect nystagmus. Several innovative applications for the diagnosis and management of vertigo have been developed [23], and combining these applications with other devices has the potential to enhance diagnostic accuracy and usability. Advances in these technologies and their integration with other devices will facilitate the assessment of smartphone-captured eye movements at home and improve the detection of nystagmus.

Future developments could further optimize the system for broader patient populations, such as older individuals or those with manual dexterity challenges, by integrating automated features and caregiver support options. These enhancements would not only expand the usability of the device but also strengthen its role in community-based and remote medical care, making it a versatile tool for vertigo diagnostics. As a potential adaptation to telemedicine, remote specialists have used eye movements captured on a smartphone to evaluate BPPV. Compared to infrared cameras, smartphones are more portable and less expensive. Commercially available infrared cameras with Wi-Fi and 3D-printed goggles can be used for web-based medical care by dizziness specialists and physicians in emergency medicine.

Limitations

Despite its promising utility, this study has several limitations. First, the findings are based on a single case, which limits the generalizability of the results. Second, the low resolution of the mini-infrared camera made it difficult to analyze the torsional components of nystagmus due to unclear iris patterns. However, the direction and intensity of nystagmus were easily determined, making the device a practical diagnostic tool.

Third, the system requires patients to operate the device during vertigo attacks, which may pose challenges for older individuals or those with impaired dexterity. Additionally, it relies on the user's ability to keep their eyes open and to ensure that the device is adequately charged and ready for use during an episode of vertigo. These operational requirements may limit its applicability for certain patient populations, particularly those who experience severe vertigo or lack manual dexterity.

Conclusion

In this study, we demonstrate the feasibility of using a low-cost, commercially available mini-infrared camera combined with 3D-printed goggles to accurately diagnose vestibular disorders such as BPPV during vertigo attacks. This approach highlights the potential for integrating affordable and accessible technologies into telemedicine, providing a valuable tool for remote, real-time diagnostic support. By enabling patients to record eye movements at home, this system bridges gaps in traditional outpatient care and empowers both patients and physicians to achieve timely and precise diagnoses. Future advancements in device performance and usability could further expand its applications in diverse health care settings, particularly for older populations or those in resource-limited regions.

Acknowledgments

We thank Editage (www.editage.com) for English language editing. This work was supported by the JST (Japan Science and Technology Agency; grant JPMJPF2403).

Data Availability

The data that support the findings of this study are available from the corresponding author (MN) upon reasonable request.

Conflicts of Interest

None declared.

References

1. Murdin L, Schilder AGM. Epidemiology of balance symptoms and disorders in the community: a systematic review. Otol Neurotol 2015 Mar;36(3):387-392. [doi: 10.1097/MAO.0000000000000691] [Medline: 25548891]

XSL•FO
RenderX

2. Kim JS, Newman-Toker DE, Kerber KA, et al. Vascular vertigo and dizziness: diagnostic criteria. J Vestib Res 2022;32(3):205-222. [doi: 10.3233/VES-210169] [Medline: 35367974]
3. Strupp M, Brandt T. Diagnosis and treatment of vertigo and dizziness. Dtsch Arztebl Int 2008 Mar;105(10):173-180. [doi: 10.3238/arztebl.2008.0173] [Medline: 19629221]
4. Mendel B, Bergenius J, Langius A. Dizziness symptom severity and impact on daily living as perceived by patients suffering from peripheral vestibular disorder. Clin Otolaryngol Allied Sci 1999 Aug;24(4):286-293. [doi: 10.1046/j.1365-2273.1999.00261.x] [Medline: 10472462]
5. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2021;2:100117. [doi: 10.1016/j.sintl.2021.100117] [Medline: 34806053]
6. Ganjali R, Jajroudi M, Kheirdoust A, Darroudi A, Alnattah A. Telemedicine solutions for clinical care delivery during COVID-19 pandemic: a scoping review. Front Public Health 2022;10:937207. [doi: 10.3389/fpubh.2022.937207] [Medline: 35937265]
7. Noda M, Kuroda T, Nomura A, Ito M, Yoshizaki T, Fushiki H. Smartphone-assisted medical care for vestibular dysfunction as a telehealth strategy for digital therapy beyond COVID-19: scoping review. JMIR Mhealth Uhealth 2023 Sep 11;11:e48638. [doi: 10.2196/48638] [Medline: 37695671]
8. Hoppes CW, Lambert KH, Whitney SL, Erbele ID, Esquivel CR, Yuan TT. Leveraging technology for vestibular assessment and rehabilitation in the operational environment: a scoping review. Bioengineering (Basel) 2024 Jan 25;11(2):117. [doi: 10.3390/bioengineering11020117] [Medline: 38391603]
9. Barahim Bastani P, Saber Tehrani AS, Badihian S, et al. Self-recording of eye movements in amyotrophic lateral sclerosis patients using a smartphone eye-tracking app. Digit Biomark 2024;8(1):111-119. [doi: 10.1159/000538992] [Medline: 39015513]
10. Parker TM, Badihian S, Hassoon A, et al. Eye and head movement recordings using smartphones for telemedicine applications: measurements of accuracy and precision. Front Neurol 2022;13:789581. [doi: 10.3389/fneur.2022.789581] [Medline: 35370913]
11. Shah MU, Lotterman S, Roberts D, Eisen M. Smartphone telemedical emergency department consults for screening of nonacute dizziness. Laryngoscope 2019 Feb;129(2):466-469. [doi: 10.1002/lary.27424] [Medline: 30478898]
12. Kuroda T, Kuroda K, Fushiki H. Development of a prototype video head impulse test system using an iPhone for screening of peripheral vestibular dysfunction. Digit Biomark 2023;7(1):150-156. [doi: 10.1159/000534543] [Medline: 37928503]
13. Sung KB, Lee TK, Furman JM. Abnormal eye movements in dizzy patients. Neurol Clin 2005 Aug;23(3):675-703. [doi: 10.1016/j.ncl.2005.01.006] [Medline: 16026671]
14. Phillips JS, Newman JL, Cox SJ. Clinical techniques and technology: vestibular telemetry. Otolaryngol Head Neck Surg 2021 Nov;165(5):751-753. [doi: 10.1177/0194599821993411] [Medline: 33650899]
15. Young AS, Lechner C, Bradshaw AP, et al. Capturing acute vertigo: a vestibular event monitor. Neurology (ECronicon) 2019 Jun 11;92(24):e2743-e2753. [doi: 10.1212/WNL.0000000000007644] [Medline: 31092626]
16. Kıroğlu M, Dağkıran M. The role of mobile phone camera recordings in the diagnosis of Meniere's disease and pathophysiological implications. J Int Adv Otol 2020 Apr;16(1):18-23. [doi: 10.5152/iao.2019.6605] [Medline: 32066548]
17. Phillips JS, Newman JL, Cox SJ, FitzGerald J. Nystagmus during an acute Ménière's attack: from prodrome to recovery. Int J Audiol 2021 Jan;60(1):70-74. [doi: 10.1080/14992027.2020.1799252] [Medline: 32731785]
18. Handmade goggles designed for recording eye movements. OSF. URL: https://osf.io/rvxw7/ [accessed 2025-02-19]
19. Halmagyi GM, McGarvie LA, Strupp M. Nystagmus goggles: how to use them, what you find and what it means. Pract Neurol 2020 Dec;20(6):446-450. [doi: 10.1136/practneurol-2020-002513] [Medline: 33115786]
20. Melliti A, van de Berg M, van de Berg R. Capturing nystagmus during vertigo attacks using a smartphone: adherence, characteristics, pearls and pitfalls. J Neurol 2023 Dec;270(12):6044-6056. [doi: 10.1007/s00415-023-11965-y] [Medline: 37653139]
21. Phillips JS, Newman J. Quantifying the direct cost benefits of vestibular telemetry using the CAVA system to diagnose the causes of dizziness. Cost Eff Resour Alloc 2023 Jan 13;21(1):2. [doi: 10.1186/s12962-022-00413-9] [Medline: 36639653]
22. Wagle N, Morkos J, Liu J, et al. aEYE: a deep learning system for video nystagmus detection. Front Neurol 2022;13:963968. [doi: 10.3389/fneur.2022.963968] [Medline: 36034311]
23. van Bonn SM, Behrendt SP, Pawar BL, Schraven SP, Mlynski R, Schuldt T. Smartphone-based nystagmus diagnostics: development of an innovative app for the targeted detection of vertigo. Eur Arch Otorhinolaryngol 2022 Dec;279(12):5565-5571. [doi: 10.1007/s00405-022-07385-9] [Medline: 35451613]

Abbreviations

BPPV: benign paroxysmal positional vertigo
CARE: Case Report
CAVA: continuous ambulatory vestibular assessment
EQUATOR: Enhancing the Quality and Transparency of Health Research

XSL•FO
RenderX

Edited by A Mavragani; submitted 12.12.24; peer-reviewed by JS Phillips, T Miwa; revised version received 31.01.25; accepted 03.02.25; published 27.02.25.

Please cite as:
Sakazaki H, Noda M, Dobashi Y, Kuroda T, Tsunoda R, Fushiki H
Monitoring Nystagmus in a Patient With Vertigo Using a Commercial Mini-Infrared Camera and 3D Printer: Cost-Effectiveness Evaluation and Case Report
JMIR Form Res 2025;9:e70015
URL: https://formative.jmir.org/2025/1/e70015
doi:10.2196/70015

© Hiroyuki Sakazaki, Masao Noda, Yumi Dobashi, Tatsuaki Kuroda, Reiko Tsunoda, Hiroaki Fushiki. Originally published in JMIR Formative Research (https://formative.jmir.org), 27.2.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Short Paper

Identifying the Minimal Clinically Important Difference in Emotion Regulation Among Youth Using the JoyPop App: Survey Study

Jaidyn Charlton[1], MA; Ishaq Malik[1], MA; Angela M Ashley[1], HBSc; Amanda Newton[2], PhD; Elaine Toombs[1], PhD; Fred Schmidt[1], PhD; Janine V Olthuis[3], PhD; Kristine Stasiuk[4], HBSW; Tina Bobinski[4], MSW; Aislin Mushquash[1], PhD

[1]Department of Psychology, Lakehead University, Thunder Bay, ON, Canada

[2]Department of Pediatrics, University of Alberta, Edmonton, AB, Canada

[3]Department of Psychology, University of New Brunswick, Fredericton, NB, Canada

[4]Dilico Anishinabek Family Care, Fort William First Nation, ON, Canada

Corresponding Author:
Aislin Mushquash, PhD
Department of Psychology
Lakehead University
955 Oliver Road
Thunder Bay, ON
Canada
Phone: 1 8073438010 ext 8771
Email: aislin.mushquash@lakeheadu.ca

Abstract

Background: The minimal clinically important difference (MCID) is an important threshold to consider when evaluating the meaningfulness of improvement following an intervention. The JoyPop app is an evidence-based smartphone app designed to improve resilience and emotion regulation. Information is needed regarding the JoyPop app's MCID among culturally diverse youth.

Objective: This study aims to calculate the MCID for youth using the JoyPop app and to explore how the MCID may differ for a subset of Indigenous youth.

Methods: Youth (N=36; aged 12-18 years) were recruited to use the JoyPop app for up to 4 weeks as part of a larger pilot evaluation. Results were based on measures completed after 2 weeks of app use. The MCID was calculated using emotion regulation change scores (Difficulties in Emotion Regulation–Short Form [DERS-SF]) and subjective ratings on the Global Rating of Change Scale (GRCS). This MCID calculation was completed for youth overall and separately for Indigenous youth only.

Results: A significant correlation between GRCS scores and change scores on the DERS-SF supported face validity (r=–0.37; P=.04). The MCID in emotion regulation following the use of the JoyPop app for youth overall was 2.80 on the DERS-SF. The MCID for Indigenous youth was 4.29 on the DERS-SF. In addition, most youth reported improved emotion regulation after using the JoyPop app.

Conclusions: These MCID findings provide a meaningful threshold for improvement in emotion regulation for the JoyPop app. They provide potential effect sizes and can aid in sample size estimations for future research with the JoyPop app or e-mental health technologies in general. The difference between overall youth and Indigenous youth MCID values also highlights the importance of patient-oriented ratings of symptom improvement as well as cultural considerations when conducting intervention research and monitoring new interventions in clinical practice.

(JMIR Form Res 2025;9:e64483) doi:10.2196/64483

KEYWORDS

mHealth; mobile health; app; psychometrics; emotion regulation; Indigenous mental health; Indigenous youth; mental health interventions; resilience; clinical psychology; adolescent mental health; mental health; JoyPop; pediatrics; mobile phone

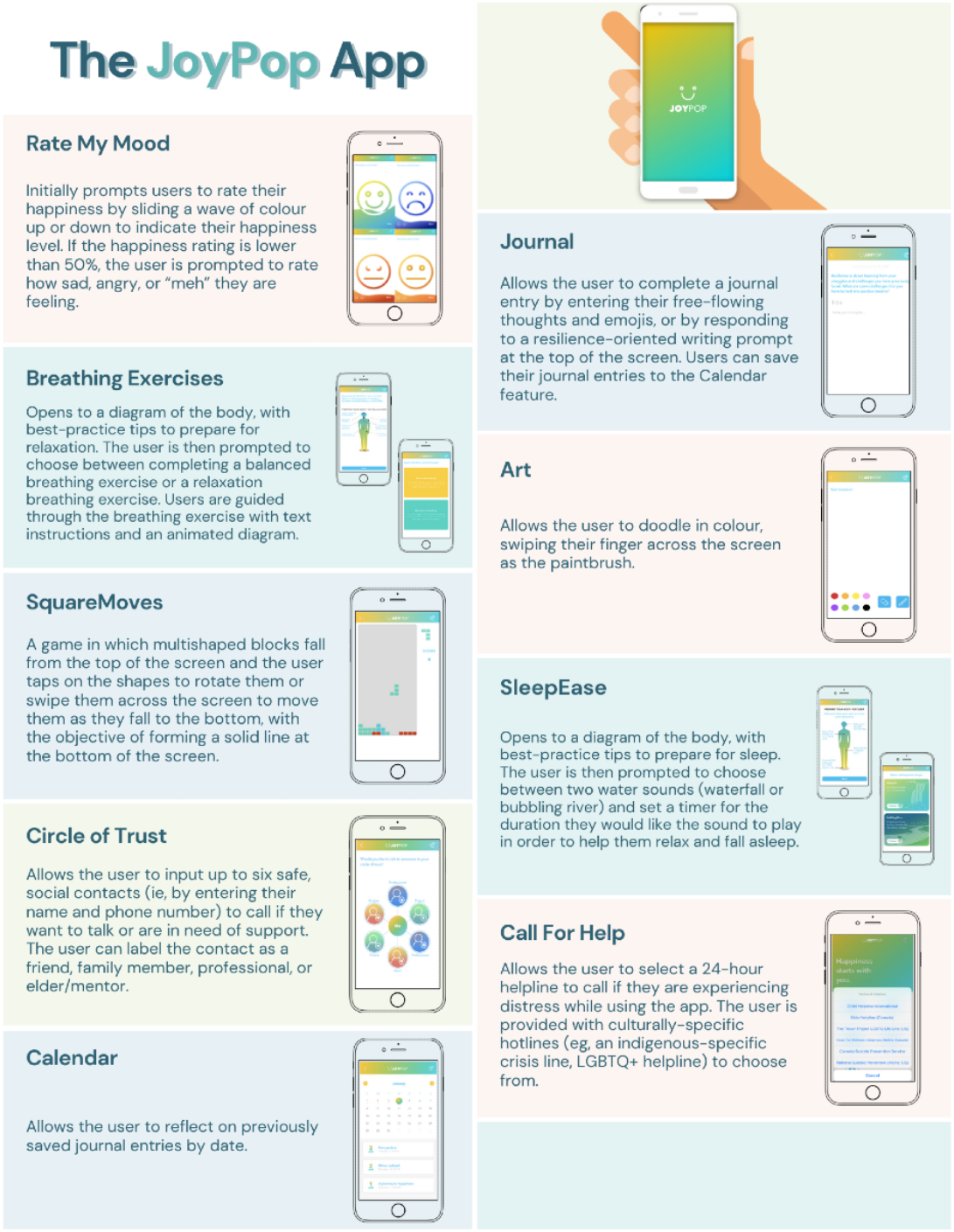
XSL•FO
RenderX

Introduction

The minimal clinically important difference (MCID) is the smallest change in an outcome (eg, symptom or ability) that patients consider meaningful to continue using an intervention [1]. These patient-derived scores serve as thresholds for outcome improvement when evaluating new interventions [1] and inform researchers on potential effect sizes in clinical trials.

The JoyPop app is a mobile health (mHealth) tool with a growing evidence base supporting its use among youth (Figure 1) [2]. JoyPop app use is associated with improved emotion regulation [2], and service providers, adolescents, and young adults view it as an acceptable, useful, and feasible tool to support existing mental health services [3,4]. Research to determine how much change in emotion regulation is clinically meaningful to app users is needed. Additionally, identifying the MCID among Indigenous youth, who experience greater mental health difficulties [5], is required.

Figure 1. JoyPop app features.



We asked youth about the emotion regulation change that was important to them before and after using the JoyPop app. We calculated the MCID for all youth and Indigenous youth. We also explored the importance of patient perspectives and how MCIDs may differ between populations to provide patient-oriented recommendations for future research. Given

XSL•FO
RenderX

this study's exploratory nature, we did not have a priori hypotheses.

Methods

Participants and Procedure

All youth (aged 12-18 years) who used the JoyPop app while receiving mental health services from the 2 largest mental health organizations in Northwestern Ontario were eligible to participate; 36 participated. Staff (ie, the youth's clinician) from the organizations informed them about the study. Staff sent the research team contact information for youth who were interested. Youth were contacted to receive further information. Interested youth attended an orientation session and downloaded the app onto their own device or were provided with a refurbished phone (containing just the JoyPop app) to use during the study. Participants were encouraged to use the app at least twice daily for 4 weeks and were sent 2 daily reminders to use the app through email and text. Participants completed measures for a larger pilot study at baseline (time 1; N=36), after 2 weeks (time 2; n=31), and after 4 weeks (time 3; n=25). Compensation was provided for each time point completed (up to CAD $80 [US $54]). If the JoyPop app was downloaded to participants' devices, they could continue using it after study completion. Given the pilot study's small sample size and attrition by time 3, we used time-1 and time-2 data for calculating the MCID.

Measures

GRCS

The Global Rating of Change Scale (GRCS) is a single-item measure assessing the degree of change individuals felt occurred during or after an intervention [6]. The GRCS asked participants "As a result of using the JoyPop app, has your ability to understand and manage your emotions been..." Participants responded using an 11-point Likert scale (–5="a very great deal worse" to 0="no change" to +5="a very great deal better") at time 2. At time 1, we asked participants how much change in their ability to regulate their emotions they "would like to see" to deem using the app worthwhile.

DERS-SF

The Difficulties in Emotion Regulation Scale–Short Form (DERS-SF) [7,8] is an 18-item questionnaire assessing emotion regulation abilities (eg, "When I was upset, I became out of control") rated on a 5-point Likert scale (1="almost never" to 5="almost always"). Total scores range from 18 to 80, with higher scores indicating greater difficulties. The DERS-SF has satisfactory reliability and validity [7,8]. The reliability among participants in this study was excellent (time-1 α=.90; time-2 α=.94).

Statistical Analysis

We used the anchor-based method: patient-reported scores on outcome measures were compared with a more subjective measure of change ("anchor" or cutoff) [1]. Time-1 GRCS scores showed that most participants reported wanting to feel +2 "somewhat better" emotion regulation from using the app (Table 1). Therefore, +2 was used as the "anchor" at time 2 when determining the MCID. We calculated the mean change score on the DERS-SF (outcome measure) for participants who rated +2 on the GRCS ("anchor")—this became the MCID. We ran descriptive statistics to calculate the proportion of participants who scored at or above the MCID. We completed these analyses for the total sample and separately for Indigenous participants. Correlation tested face validity.

Table 1. Frequencies of Global Rating of Change Scale (GRCS) ratings.

GRCS rating[a]	Time 1[b] (N=36), n (%)	Time 2 (n=31), n (%)
Emotion regulation worsened (rated from –1 to –5)	4 (11)	2 (6)
No change in emotion regulation (rated 0)	4 (11)	5 (16)
Emotion regulation was almost the same, hardly better at all (rated +1)	3 (8)	8 (26)
Emotion regulation was somewhat better (rated +2)	13 (36)	10 (32)
Emotion regulation was much better (rated from +3 to +5)	12 (33)	6 (19)

[a]The percentage of youth that rated each item on the GRCS (N=36). Participants (aged 12-18 years) were recruited for the pilot study from 2 mental health organizations in Northwestern Ontario.

[b]GRCS ratings at time 1 (N=36) reflect how much change youth prospectively wanted to see from using the JoyPop app. GRCS ratings at time 2 (n=31) reflect how much change youth actually experienced after 2 weeks of using the JoyPop app.

Ethical Considerations

The study was reviewed and approved by research ethics boards at Lakehead University (ROMEO #1468701) and the collaborating mental health organizations. Study information was reviewed with all youth. All participants provided informed consent. Data only contained the ID number provided at study onset. The list linking participant ID numbers to participant names was kept separately from the study dataset and was destroyed after data collection. Youth could withdraw at any time without penalty.

Results

Descriptive Statistics

Participant demographics (Table 2) and descriptive statistics for emotion regulation and global ratings of change (Table 3) are shown. Time-2 GRCS scores showed that many participants reported improved emotion regulation after using the app for 2 weeks (Table 1). A significant negative correlation between time-2 GRCS ratings and DERS-SF change scores provided face validity that the GRCS can indicate a meaningful patient-rated change in emotion regulation (r=–0.37; P=.04).

XSL•FO
RenderX

Table 2. Youth[a] participant demographics.

Variables	Value (N=36)
Ethnicity, n (%)	
White	9 (25)
Indigenous	25 (69)
Mixed	1 (3)
South Asian	1 (3)
Age (years), mean (SD)	14.47 (1.75)
Gender identity, n (%)	
Girl	22 (61)
Boy	5 (14)
Agender or nonbinary	7 (19)
Transgender	2 (6)

[a]Participants (N=36; aged 12-18 years) were recruited for the pilot study from 2 mental health organizations in Northwestern Ontario.

Table 3. Mean and SD values of the Difficulties in Emotion Regulation Scale–Short Form (DERS-SF) and Global Rating of Change Scale (GRCS) overall and for Indigenous youth.

Participants[a] and time point	Participants, n	DERS-SF[b], mean (SD)	GRCS[c], mean (SD)
All youth			
Time 1	36	59.64 (14.03)	1.75 (1.54)
Time 2	31	55.84 (16.11)	1.48 (1.44)
Indigenous youth			
Time 1	25	58.44 (15.92)	1.64 (1.29)
Time 2	21	53.19 (17.49)	1.43 (1.36)

[a]Participants (aged 12-18 years) were recruited for the pilot study from 2 mental health organizations in Northwestern Ontario.

[b]Higher scores indicate greater emotional dysregulation and lower scores indicate improvement.

[c]Higher scores indicate higher subjective improvement. GRCS ratings at time 1 reflect how much change youth prospectively wanted to see from using the JoyPop app. GRCS ratings at time 2 reflect how much change youth actually experienced after 2 weeks of using the JoyPop app.

MCID

The mean change score on the DERS-SF for participants who rated +2 “somewhat better” on the GRCS at time 2 was –2.80 (SD 9.22), translating to an MCID of 2.80. Using this threshold, 45% (14/31) of participants met or exceeded this change after 2 weeks, regardless of their time-2 GRCS rating (Table 4). For Indigenous youth who rated +2 “somewhat better” on the GRCS, the mean change score on the DERS-SF was –4.29 (SD 8.96), indicating that they experienced a 4.29-point improvement in emotion regulation after using the app for 2 weeks and had an MCID of 4.29. Using this threshold, 43% (9/21) of Indigenous youth met or exceeded this change after 2 weeks.

Table 4. Frequencies of Global Rating of Change Scale (GRCS) ratings of those who met or exceeded the minimal clinically important difference on the Difficulties in Emotion Regulation Scale–Short Form.

GRCS rating[a]	Youth (n=14), n (%)
Emotion regulation worsened (rated from –1 to –5)	0
No change in emotion regulation (rated 0)	1 (7)
Emotion regulation was almost the same, hardly better at all (rated +1)	5 (36)
Emotion regulation was somewhat better (rated +2)	5 (36)
Emotion regulation was much better (rated from +3 to +5)	3 (21)

[a]Participants (aged 12-18 years) were recruited for the pilot study from 2 mental health organizations in Northwestern Ontario. This table represents only the youth who reached the minimal clinically important difference of 2.80 or exceeded it (14/31, 45%) and what they rated on the GCRS at time 2.

XSL•FO
RenderX

Discussion

The MCID in emotion regulation was 2.80 for all youth and 4.29 for Indigenous youth. These provide preliminary effect sizes for future research as benchmarks for meaningful patient-oriented change, rather than solely basing findings on statistical significance, which lacks practicality or patient-oriented relevance [9]. Future study designs could recognize that youth generally need at least a 2.80-point improvement in their emotion regulation on the DERS-SF to consider the JoyPop app worthwhile. These findings can also inform future sample size calculations for JoyPop research [10,11].

Evaluating the MCID among Indigenous youth allowed us to consider their unique cultural experiences and provided insight into the effectiveness and appropriateness of the app for this subgroup. Our results suggest that Indigenous youth need to see a greater improvement in emotion regulation to deem the JoyPop app worthwhile and to start feeling "somewhat better." Conducting MCID research with different groups, beyond ethnicity, can provide information on distinct change thresholds that interventions should aim for as opposed to "one-size-fits-all" approaches.

The findings also highlight the value of pairing outcome measures with patient-rated measures of change [12]. Not all participants who met or exceeded the MCID threshold subjectively rated an improvement in their emotion regulation (Table 4). This suggests that many, but not all, youth notice improvements in their emotion regulation, and that some youth require a higher threshold of change (>2.80) to notice improvement. This underscores the importance of measuring how individuals perceive their symptoms and abilities in relation to an intervention. If a young person does not perceive improvement from using an intervention, they are unlikely to continue, even if outcome data show improvement. We recommend subjective experience ratings (eg, GRCS) alongside outcome measures (eg, DERS-SF) when evaluating mHealth interventions.

MCID research typically involves participants reporting on what happened following an intervention. Previous research has relied on clinician judgment and clinical theory to determine their "anchor" on which to base the MCID [13]. We included a novel prospective question at time 1 to explore what youth would want to experience to inform the choice of anchor for our sample. Most youth wanted to see a minimum "somewhat better" improvement in their emotion regulation, aligning with previous research using +2 as an anchor [13]. This is useful in promoting the JoyPop app to youth and service providers as it highlights continuity between clinician and client judgment.

Despite encouraging preliminary findings, this study has limitations. The GRCS has received criticism as a single-item measure, which lacks robust psychometric properties [14]. Single-item measures do not capture the complexity of experience and can introduce differences in interpretation between participants [15]. However, single-item measures are informative and valid when used appropriately, such as concurrently with multi-item measures (eg, alongside the DERS-SF) or when they ask straightforward questions that do not require multiple items (eg, how much change in emotion regulation youth experienced) [14]. Another limitation is that the MCIDs were calculated based on a small sample, which limits generalizability. Nevertheless, these findings provide valuable contributions to the JoyPop app's evidence base. Through cumulative longitudinal evidence, the MCIDs can be improved and expanded on for different populations (eg, adolescent vs young adult; differing ethnicities), thus demonstrating a client-centered approach to research and mHealth app implementation [16].

Acknowledgments

We would like to thank the youth who informed the initial design of the JoyPop app and those who continue to inform its evolution through participation in past and current research. We would also like to acknowledge our community partner organizations for their ongoing support of our collaborations.

Data Availability

The dataset generated and analyzed is publicly available due to our team's adherence to the Ownership, Control, Access, and Possession (OCAP) standards set by the First Nations Information Governance Centre. Questions about study datasets can be directed to the corresponding author and principal investigator (ARM).

Conflicts of Interest

The principal investigator (ARM) recently acquired intellectual property ownership rights for the JoyPop app (in June 2024). Before this time, intellectual property rights were owned by other researchers and institutions. ARM did not own any intellectual property rights when the study was developed, when data were collected, when data were analyzed, or when the initial draft of the manuscript was written.

References

1. Copay AG, Subach BR, Glassman SD, Polly DW, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. Spine J 2007;7(5):541-546. [doi: 10.1016/j.spinee.2007.01.008] [Medline: 17448732]

XSL•FO
RenderX

2. MacIsaac A, Mushquash AR, Mohammed S, Grassia E, Smith S, Wekerle C. Adverse childhood experiences and building resilience with the JoyPop app: evaluation study. JMIR Mhealth Uhealth 2021;9(1):e25087 [FREE Full text] [doi: 10.2196/25087] [Medline: 33393908]
3. Malik I, Perez A, Toombs E, Schmidt F, Olthuis JV, Charlton J, et al. Female youth and mental health service providers' perspectives on the JoyPop™ app: a qualitative study. Front Digit Health 2023;5:1197362 [FREE Full text] [doi: 10.3389/fdgth.2023.1197362] [Medline: 37829596]
4. Mushquash AR, Pearson ES, Waddington K, MacIsaac A, Mohammed S, Grassia E, et al. User perspectives on a resilience-building app (JoyPop): qualitative study. JMIR Mhealth Uhealth 2021;9(7):e28677 [FREE Full text] [doi: 10.2196/28677] [Medline: 34255696]
5. Firestone M, Smylie J, Maracle S, McKnight C, Spiller M, O'Campo P. Mental health and substance use in an urban first nations population in Hamilton, Ontario. Can J Public Health 2015;106(6):e375-e381 [FREE Full text] [doi: 10.17269/cjph.106.4923] [Medline: 26680428]
6. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther 2009;17(3):163-170 [FREE Full text] [doi: 10.1179/jmt.2009.17.3.163] [Medline: 20046623]
7. Gratz KL, Roemer L. Multidimensional assessment of emotion regulation and dysregulation: development, factor structure, and initial validation of the difficulties in emotion regulation scale. Journal of Psychopathology and Behavioral Assessment 2004;26(1):41-54. [doi: 10.1023/B:JOBA.0000007455.08539.94]
8. Kaufman EA, Xia M, Fosco G, Yaptangco M, Skidmore CR, Crowell SE. The Difficulties in Emotion Regulation Scale-Short Form (DERS-SF): validation and replication in adolescent and adult samples. J Psychopathol Behav Assess 2015;38(3):443-455. [doi: 10.1007/s10862-015-9529-3]
9. Anvari F, Lakens D. Using anchor-based methods to determine the smallest effect size of interest. Journal of Experimental Social Psychology 2021;96:104159. [doi: 10.1016/j.jesp.2021.104159]
10. Neely JG, Karni RJ, Engel SH, Fraley PL, Nussenbaum B, Paniello RC. Practical guides to understanding sample size and minimal clinically important difference (MCID). Otolaryngol Head Neck Surg 2007;136(1):14-18. [doi: 10.1016/j.otohns.2006.11.001] [Medline: 17210326]
11. McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. JAMA 2014;312(13):1342-1343. [doi: 10.1001/jama.2014.13128] [Medline: 25268441]
12. Cook CE. Clinimetrics corner: the minimal clinically important change score (MCID): a necessary pretense. J Man Manip Ther 2008;16(4):E82-E83 [FREE Full text] [doi: 10.1179/jmt.2008.16.4.82E] [Medline: 19771185]
13. Radomski AD, Bagnell A, Curtis S, Hartling L, Newton AS. Examining the usage, user experience, and perceived impact of an internet-based cognitive behavioral therapy program for adolescents with anxiety: randomized controlled trial. JMIR Ment Health 2020;7(2):e15795 [FREE Full text] [doi: 10.2196/15795] [Medline: 32022692]
14. Allen MS, Iliescu D, Greiff S. Single item measures in psychological science. European Journal of Psychological Assessment 2022;38(1):1-5. [doi: 10.1027/1015-5759/a000699]
15. Sedaghat AR. Understanding the minimal clinically important difference (MCID) of patient-reported outcome measures. Otolaryngol Head Neck Surg 2019;161(4):551-560. [doi: 10.1177/0194599819852604] [Medline: 31159641]
16. McGrath P, Wozney L, Rathore SS, Notarianni M, Schellenberg P. Toolkit for e-mental health implementation. Mental Health Commission of Canada. 2018. URL: https://mentalhealthcommission.ca/resource/toolkit-for-e-mental-health-implementation/ [accessed 2024-09-28]

Abbreviations

DERS-SF: Difficulties in Emotion Regulation Scale–Short Form
GRCS: Global Rating of Change Scale
MCID: minimal clinically important difference
mHealth: mobile health

Edited by A Mavragani; submitted 18.07.24; peer-reviewed by A Serlachius; comments to author 16.08.24; revised version received 28.08.24; accepted 24.09.24; published 23.01.25.

Please cite as:
Charlton J, Malik I, Ashley AM, Newton A, Toombs E, Schmidt F, Olthuis JV, Stasiuk K, Bobinski T, Mushquash A
Identifying the Minimal Clinically Important Difference in Emotion Regulation Among Youth Using the JoyPop App: Survey Study
JMIR Form Res 2025;9:e64483
URL: https://formative.jmir.org/2025/1/e64483
doi:10.2196/64483
PMID:39847426

XSL•FO
RenderX

©Jaidyn Charlton, Ishaq Malik, Angela M Ashley, Amanda Newton, Elaine Toombs, Fred Schmidt, Janine V Olthuis, Kristine Stasiuk, Tina Bobinski, Aislin Mushquash. Originally published in JMIR Formative Research (https://formative.jmir.org), 23.01.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Short Paper

Assessing the Impact of a Virtual Reality Cognitive Intervention on Tennis Performance in Junior Tennis Players: Pilot Study

Joaquin A Anguera[1], PhD; Aleem Choudhry[2], BS, MBA; Michael Seaman[2], BS, MBA; Dominick Fedele[2], BS

[1]Neuroscape, Department of Neurology and Psychiatry, Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, United States

[2]Mastermind Sports, Naperville, IL, United States

Corresponding Author:
Joaquin A Anguera, PhD
Neuroscape
Department of Neurology and Psychiatry, Weill Institute for Neurosciences
University of California, San Francisco
675 Nelson Rising Lane
San Francisco, CA, 94158
United States
Phone: 1 415 702 4322
Email: joaquin.anguera@ucsf.edu

Abstract

Background: There is evidence that cognitive training interventions can positively impact executive functions, and that some studies have demonstrated that athletes typically exhibit greater accuracy and faster response times on select cognitive tasks. While the engagement of executive functions is suggested to be part of high-level sporting activities, it is unclear whether such training approaches could directly benefit athletic performance.

Objective: The objective of this study was to evaluate the impact of a combined virtual reality (VR)– and tablet-based cognitive training intervention on adolescent tennis players' performance. Here, we examined differences in Universal Tennis Rating (UTR) between players who supplemented their regular tennis training with a cognitive training intervention and a group that continued regular tennis training alone. This custom cognitive training program targeted specific cognitive control abilities including attention, working memory, and goal management.

Methods: Data were collected from a cohort of tennis players in a randomized controlled trial design led by the dedicated research team. Participants (N=23, age: mean 14.8, SD 2.4 years) from the Czech Lawn Tenis Klub (Prague, Czech Republic) were invited to participate in this study. These individuals were randomized into an intervention + training-as-usual group (n=13) or training-as-usual group (control group; n=10), with the change in UTR score being the primary metric of interest.

Results: There was no difference in UTR between the 2 groups at baseline (intervention: mean 8.32, SD 2.7; control: mean 7.60, SD 2.3). Following the treatment period, individuals in the intervention group showed a significant improvement in their UTR (an increase of 0.5; t_{12}=4.88, P<.001) unlike the control group (an increase of 0.02; t_9=1.77, P=.12). On comparing the change in UTR (posttraining UTR minus pretraining UTR) attained by each group, we found that the intervention group had a 38% greater improvement in UTR than the control group. An analysis of covariance revealed a significantly greater improvement in UTR for the intervention group than for the control group ($F_{1,20}$=8.82, P=.008).

Conclusions: The present findings suggest that training cognitive abilities through an immersive visual platform may benefit athletic performance, including tennis. Such a result warrants careful consideration, given the known difficulties in evidencing far transfer not only in cognitive studies but also in athletic activities. These preliminary pilot findings suggest that the Mastermind Cognitive Training program may be a viable tool for supplementing athletic training practices, although this result warrants further investigation and replication. However, many questions remain unanswered, and further work is needed to better understand the potential utility and mechanisms underlying potential effects of such a platform.

(JMIR Form Res 2025;9:e66979) doi:10.2196/66979

XSL•FO
RenderX

KEYWORDS

executive function; serious games; cognitive training; performance enhancement; athletes; sport; pilot study; VR; virtual reality; serious game; tennis; adolescents; teenagers; youth; randomized controlled trial; players

Introduction

The engagement of cognitive control abilities, defined here as attention, working memory, and multitasking [1], is thought to be a critical part of high-level athletic performance. This idea stems from research showing that both amateur and elite athletes typically exhibit greater accuracy and faster response times on select cognitive tasks [2,3], and that one's underlying cognitive abilities can predict future athletic achievement [4-6]. However, the existing literature on whether enhancing these abilities through cognitive training can improve athletic performance is limited [7-9]. Numerous examples in the cognitive training literature provide reasons why so-called "far transfer" (eg, the ability to apply learning from one context to a different, dissimilar context [10]) is rarely observed [11,12]. However, those positive findings following the use of specific cognitive training platforms suggest that such an approach could possibly be beneficial [13-16]. Such positive effects have been observed in those platforms using closed-loop algorithms, an approach that changes the level of difficulty on a trial-by-trial basis to challenge a given individual appropriately for a given training experience [1,17-19]. This approach has shown generalizable benefits to not only untrained tasks assessing cognitive abilities (including attention, working memory, and multitasking) [16,20-24] but also so-called "real-world" constructs including mood [14,15,25-27] and even academic abilities [28]. Here, we describe the preliminary findings of a proof-of-concept randomized controlled study involving advanced adolescent tennis players who supplemented their regular tennis training with a cognitive training intervention in comparison to a group that continued regular tennis training alone.

Methods

Ethical Considerations

For this project, all study procedures were conducted in accordance with protocols approved by the Fakultní Nemocnice Brno (Brno, Czech Republic; 14-110123/EK). Participant assent and parental consent were obtained prior to study participation in person by the study team. All study data were deidentified to ensure privacy and confidentiality protections were in place. No compensation was given for participating in this study.

Participants

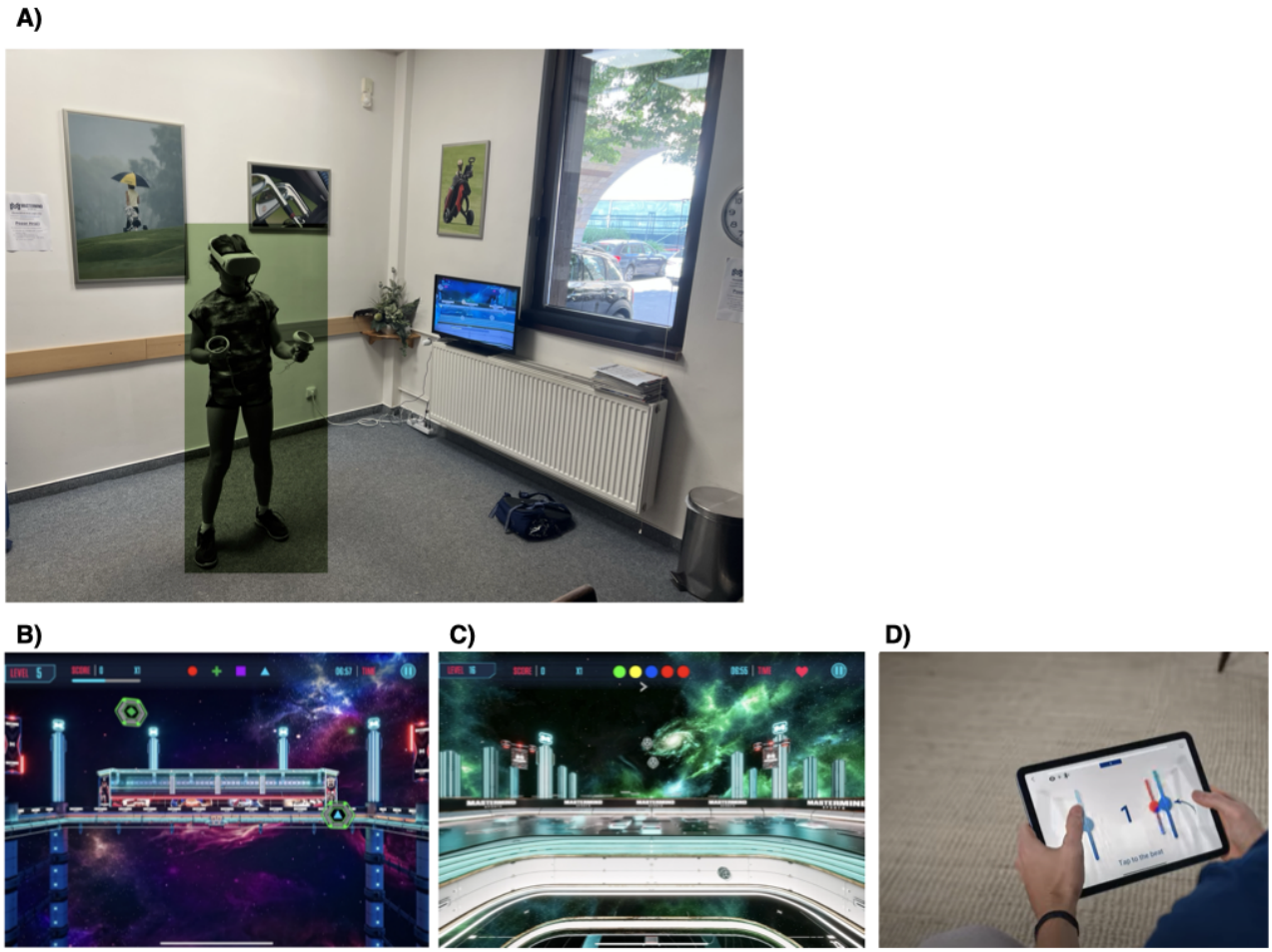
A total of 59 tennis players (age: mean 16.7, SD 3.6 years; 34 male) from the Czech Lawn Tenis Klub (Prague, Czech Republic) were invited to participate in this study (and provided a full overview of the study procedures and time requirements). See Multimedia Appendix 1 for more details.

Intervention

The Mastermind Cognitive Training program used a virtual reality (VR) headset (Oculus Quest 2 by Meta) to deliver a custom cognitive training program targeting specific cognitive control abilities including attention, working memory, and goal management. The VR component comprised 3 modules, each focusing on a different aspect of cognitive control: a multiple object–tracking module aimed at engaging attention (with increasing distraction), a spatial span module to engage working memory, and a visual search module with task switching requirements to engage multitasking abilities. In each case, participants used VR hand-held controllers (Figure 1A) to respond to stimuli presented in a video game–like environment, in some cases "shooting" specific target stimuli while avoiding others (Figure 1B), or trying to catch certain colored targets in a given order stimuli while avoiding nontargets (Figure 1C). Each module was adaptive, adjusting task difficulty based on performance. In addition, participants completed 20 minutes of tablet-based training (iPad) on tasks related to rhythmic timing (where participants were asked to tap on the screen to the beat of music being played [28-31]; Figure 1D). The intervention group was asked to complete 24 program sessions over the course of 10 weeks, with each session consisting of 15 minutes of VR exercises and 20 minutes of tablet-based exercises, totaling approximately 14 hours of training. See Multimedia Appendix 1 for more details.

XSL•FO
RenderX

Figure 1. The Mastermind Cognitive Training program. (A) Participant training during the Mastermind Cognitive Training program using virtual reality. Participants were tennis players from the Czech Lawn Tenis Klub (Prague, Czech Republic) who trained between January and November 2023. Participants used VR hand-held controllers to respond to stimuli presented in a video game–like environment, in some cases "shooting" specific target stimuli while avoiding others or trying to catch certain colored targets in a given order stimuli while avoiding nontargets. (B) the virtual reality attention module. (C) The virtual reality multitasking module. (D) The tablet rhythm training module where participants tapped in a rhythmic fashion to stimuli presented on the screen.



Measure

The primary outcome measure was the change in Universal Tennis Rating (UTR), a scale from 1.00 (eg, a beginner player) to 16.50 (eg, a professional player; for context, the current #1 player in the world is rated 16.05), which promotes fair and competitive play [32]. Each player's UTR was collected before training and after the completion of the 24-session training period, which an interassessment interval of approximately 4 months. See Multimedia Appendix 1 for more details about this measure.

Results

Of the 59 consenting individuals, 9 withdrew upon learning of the study requirements, 15 failed to complete the stipulated training schedule and posttraining assessments due to scheduling conflicts with travel and academic commitments, 8 opted out due to moving out of the area or changing tennis clubs, and 4 discontinued participation due to unrelated injuries requiring physical rehabilitation. Thus, the final sample that completed the stipulated training and outcome assessments consisted of 23 participants: 13 in the intervention group (3 female, aged mean 15.0, SD 2.4) and 10 (7 female, aged mean 14.6, SD 2.5) in the training-as-usual group (see Figure 2 for the CONSORT [Consolidated Standards of Reporting Trials] diagram and Multimedia Appendix 2 for CONSORT checklist). To test for training effects on UTR (Table 1), we conducted an analysis of covariance (ANCOVA) with posttraining UTR as the dependent variable, pretraining UTR as the covariate, and the group as the fixed factor as in our previous work [14,20,33,34]. The analysis revealed a significant group difference in UTR improvement over time ($F_{1,20}$=8.82, P=.008; Figure 3), indicating that the UTR of the intervention group improved significantly more than that of the control group. On comparing the change in UTR (posttraining UTR minus pretraining UTR) attained by each group, the intervention group had a 38% greater improvement in UTR than the control group. Post hoc analysis of the within-group changes in UTR using 2-tailed, paired-samples t tests for each group separately to test for significant differences between each testing session revealed a significant improvement by the intervention group (increase of 0.50; t_{12}=4.88, P<.001) unlike the control group (increase of 0.20; t_9=1.77, P=.12). Critically, there was no difference in UTR at baseline between the 2 groups (independent-samples t test: t_{21}=.68, P=.50).

XSL•FO
RenderX

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.

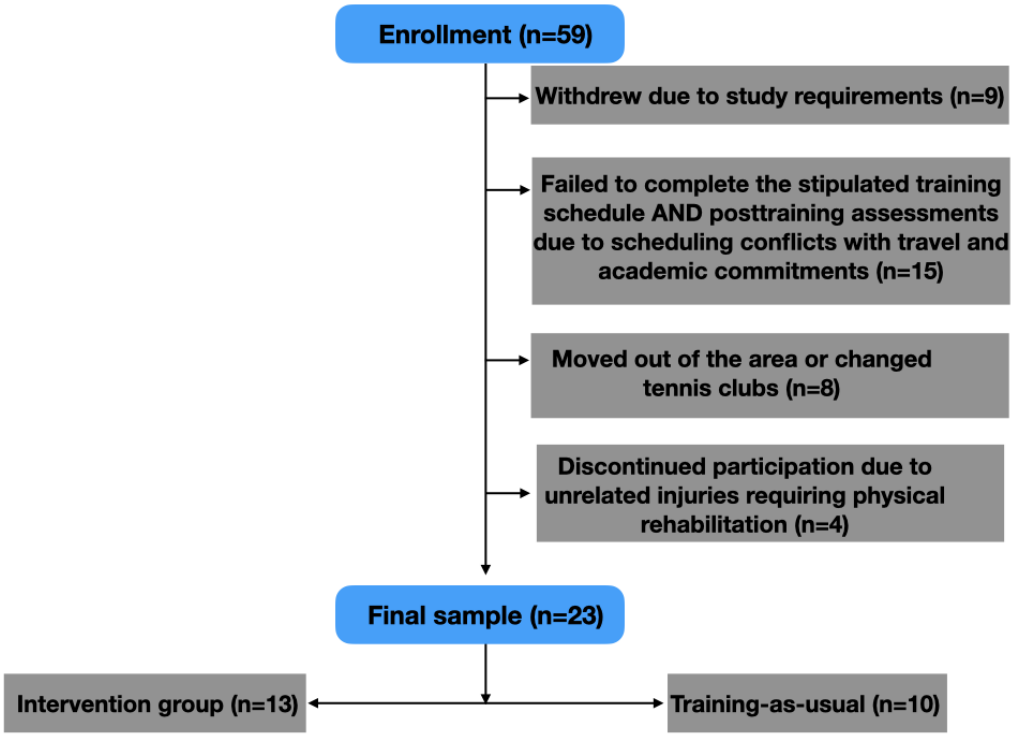
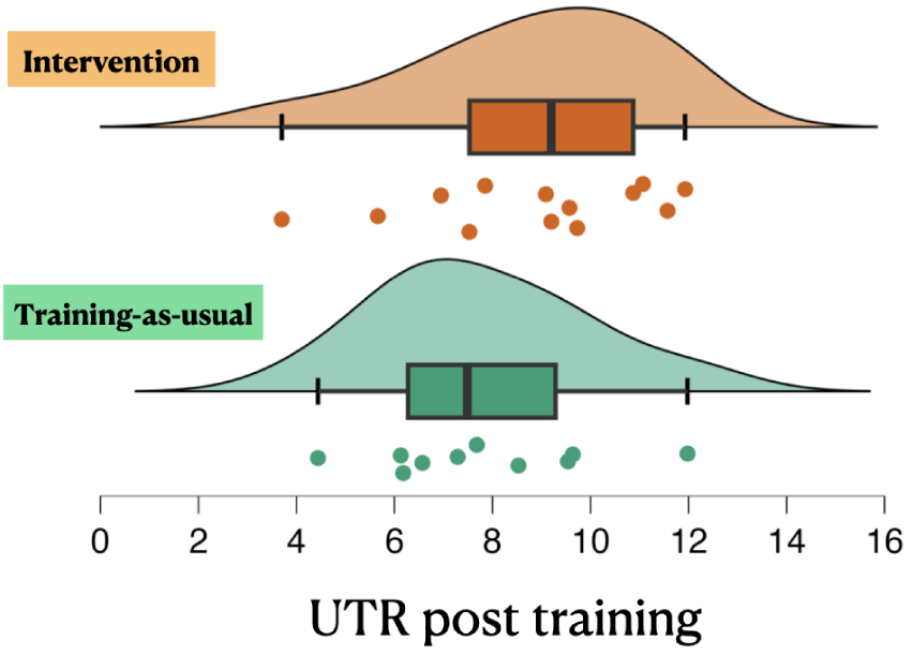


Table 1. Universal Tennis Rating (UTR) by group and timepoint.

	Pretraining UTR, mean (SD)	Posttraining UTR, mean (SD)	Change (posttraining UTR–pretraining UTR)
Intervention	8.32 (2.7)	8.83 (2.4)	0.50 (0.37)
Treatment-as-usual	7.60 (2.3)	7.79 (2.2)	0.20 (0.37)

Figure 3. Universal Tennis Rating (UTR) by group. The UTR is a scale from 1.00 (eg, a beginning player) to 16.50 (eg, a professional player), which promotes fair and competitive play. Each player's UTR was recorded before training and after the completion of the 24-session training period (approximately 4 months). Posttraining UTR scores for the intervention and training-as-usual groups visualized via a raincloud plot with difference-adjusted 95% CIs of the mean. An analysis of covariance (controlling for pretraining performance) revealed that the improvement observed by the intervention group was significantly greater than that of the training-as-usual group ($F_{1,20}$=8.82, P=.008).



XSL•FO
RenderX

Discussion

Principal Findings

The primary goal of this proof-of-concept pilot study was to investigate the effects of the Mastermind Cognitive Training program in adolescent tennis players. These findings suggest that supplementing regular training with cognitive training can lead to meaningful improvements in real-world athletic performance. It has been suggested that visual-based training (including the use of VR and/or tablet platforms) is more likely to benefit trained athletes than novices [35]. Indeed, this theory has empirical support: "real-world" improvements have been observed, including performance in college-level baseball players [36,37] as well as table tennis performance [38] when coupling such training with their regular sports training [38]. In conjunction with those studies, the present findings suggest that training cognitive abilities through an immersive visual platform may benefit performance on athletic endeavors, including tennis. Such a result warrants careful consideration, given the known difficulties in evidencing far transfer not just in cognitive studies, but especially in athletic activities [7]. That said, our findings do align with those of the few positive studies demonstrating that cognitive training can result in some form of on-field improvements [8,9]. See Multimedia Appendix 1 for a further discussion of practical considerations.

Limitations

Several limitations should be considered with the present findings. First, the most important factor to consider with this study is the small sample size, especially considering the large number of individuals who did not complete the study. It should be noted that this study was not designed to be anything more than a pilot study intended to gauge the feasibility of implementing the Mastermind Cognitive Training program following the unexpected opportunity to work with adolescent tennis players. Future work must involve a larger trial of participants to replicate these findings to provide supporting evidence for the present interpretations. Second, the very real possibility exists that participation in the intervention generated a placebo effect through the novel use of a VR platform. Given the pervasive problem of placebo effects in such work [39], it is critical that future endeavors account for these effects in a carefully designed manner. Finally, it is unknown what aspect of the training program led to the improvement in UTR. The absence of cognitive control outcome measures to assess whether, and to what extent, this training program led to cognitive changes hinders a deeper understanding of what possibly drove the differential improvement in UTR. Future work is warranted to better understand the potential underlying mechanisms driving such change.

Conclusions

This exploratory proof-of-concept study investigated the potential benefits of a novel cognitive intervention platform in a sample of healthy adolescent tennis athletes. It represents the first step toward a number of data-driven studies demonstrating the utility and benefits of the use of this platform. These pilot findings suggest that this platform may be beneficial for adolescent tennis athletes looking to improve their sporting performance by supplementing their existing training protocols with a targeted cognitive training experience. However, many questions remain unanswered, and further work is needed to better understand the potential utility and mechanisms underlying potential effects of such a platform.

Acknowledgments

We would like to thank the participants from the Czech Lawn Tenis Klub, their coaches, and their families for being a part of this study. We thank C Gallen for her insights with the study design and suggestions regarding data preprocessing, C Walsh for providing suggestions for the manuscript, and P Wais and T Zanto for insights on figure creation and use. Note that no generative artificial intelligence was used in any portion of the manuscript writing.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization and design: JAA, MS, AC, and DF
Data collection: AC
Data processing and analysis: JAA and AC
Writing—original draft: JAA
Writing—review and editing: JAA, MS, AC, and DF

Conflicts of Interest

JAA is a paid consultant for MasterMind Sports, MS and DF are employees of MasterMind Sports. AC is a board member of MasterMind Sports.

Multimedia Appendix 1

XSL•FO
RenderX

Additional material.
[DOCX File , 22 KB - formative_v9i1e66979_app1.docx]

Multimedia Appendix 2
CONSORT-EHEALTH (V 1.6.1) checklist.
[PDF File (Adobe PDF File), 2950 KB - formative_v9i1e66979_app2.pdf]

References

1. Anguera JA, Gazzaley A. Video games, cognitive exercises, and the enhancement of cognitive abilities. Curr Opin Behav Sci 2015 Aug;4:160-165. [doi: 10.1016/j.cobeha.2015.06.002]
2. Mann D, Williams AM, Ward P, Janelle CM. Perceptual-cognitive expertise in sport: a meta-analysis. J Sport Exerc Psychol 2007 Aug;29(4):457-478. [doi: 10.1123/jsep.29.4.457] [Medline: 17968048]
3. Voss MW, Kramer AF, Basak C, Prakash RS, Roberts B. Are expert athletes 'expert' in the cognitive laboratory? A meta‐analytic review of cognition and sport expertise. Appl Cogn Psychol 2009 Jun 17;24(6):812-826. [doi: 10.1002/acp.1588]
4. Vestberg T, Gustafson R, Maurex L, Ingvar M, Petrovic P. Executive functions predict the success of top-soccer players. PLoS One 2012 Apr 4;7(4):e34731 [FREE Full text] [doi: 10.1371/journal.pone.0034731] [Medline: 22496850]
5. Vestberg T, Reinebo G, Maurex L, Ingvar M, Petrovic P. Core executive functions are associated with success in young elite soccer players. PLoS One 2017 Feb 8;12(2):e0170845 [FREE Full text] [doi: 10.1371/journal.pone.0170845] [Medline: 28178738]
6. Mangine G, Hoffman JR, Wells AJ, Gonzalez AM, Rogowski JP, Townsend JR, et al. Visual tracking speed is related to basketball-specific measures of performance in NBA players. J Strength Cond Res 2014 Sep;28(9):2406-2414. [doi: 10.1519/JSC.0000000000000550] [Medline: 24875429]
7. Walton CC, Keegan RJ, Martin M, Hallock H. The potential role for cognitive training in sport: more research needed. Front Psychol 2018 Jul 3;9:1121 [FREE Full text] [doi: 10.3389/fpsyg.2018.01121] [Medline: 30018585]
8. Romeas T, Guldner A, Faubert J. 3D-Multiple Object Tracking training task improves passing decision-making accuracy in soccer players. Psychol Sport Exerc 2016 Jan;22:1-9. [doi: 10.1016/j.psychsport.2015.06.002]
9. Hirao T, Masaki H. Modulation of spatial attentional allocation by computer-based cognitive training during lacrosse shooting performance. Front Psychol 2017 Jan 4;8:2271 [FREE Full text] [doi: 10.3389/fpsyg.2017.02271] [Medline: 29354085]
10. Sala G, Gobet F. Cognitive training does not enhance general cognition. Trends Cogn Sci 2019 Jan;23(1):9-20. [doi: 10.1016/j.tics.2018.10.004] [Medline: 30471868]
11. Zelinski EM. Far transfer in cognitive training of older adults. Restor Neurol Neurosci 2009 Jan 01;27(5):455-471. [doi: 10.3233/rnn-2009-0495]
12. Simons DJ, Boot WR, Charness N, Gathercole SE, Chabris CF, Hambrick DZ, et al. Do "Brain-Training" programs work? Psychol Sci Public Interest 2016 Oct;17(3):103-186. [doi: 10.1177/1529100616661983] [Medline: 27697851]
13. Anguera JA, Brandes-Aitken AN, Antovich AD, Rolle CE, Desai SS, Marco EJ. A pilot study to determine the feasibility of enhancing cognitive abilities in children with sensory processing dysfunction. PLoS One 2017 Apr 5;12(4):e0172616 [FREE Full text] [doi: 10.1371/journal.pone.0172616] [Medline: 28380008]
14. Anguera JA, Gunning FM, Areán PA. Improving late life depression and cognitive control through the use of therapeutic video game technology: A proof-of-concept randomized trial. Depress Anxiety 2017 Jun 03;34(6):508-517 [FREE Full text] [doi: 10.1002/da.22588] [Medline: 28052513]
15. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
16. Ziegler DA, Simon AJ, Gallen CL, Skinner S, Janowich JR, Volponi JJ, et al. Closed-loop digital meditation improves sustained attention in young adults. Nat Hum Behav 2019 Jul 03;3(7):746-757 [FREE Full text] [doi: 10.1038/s41562-019-0611-9] [Medline: 31160812]
17. Mishra J, Gazzaley A. Closed-loop cognition: the next frontier arrives. Trends Cogn Sci 2015 May;19(5):242-243 [FREE Full text] [doi: 10.1016/j.tics.2015.03.008] [Medline: 25818939]
18. Mishra J, Anguera J, Gazzaley A. Video games for neuro-cognitive optimization. Neuron 2016 Apr 20;90(2):214-218 [FREE Full text] [doi: 10.1016/j.neuron.2016.04.010] [Medline: 27100194]
19. Mishra J, Anguera JA, Ziegler DA, Gazzaley A. A cognitive framework for understanding and improving interference resolution in the brain. Prog Brain Res 2013;207:351-377 [FREE Full text] [doi: 10.1016/B978-0-444-63327-9.00013-8] [Medline: 24309262]
20. Anguera JA, Volponi JJ, Simon AJ, Gallen CL, Rolle CE, Anguera-Singla R, et al. Integrated cognitive and physical fitness training enhances attention abilities in older adults. NPJ Aging 2022 Aug 30;8(1):12 [FREE Full text] [doi: 10.1038/s41514-022-00093-y] [Medline: 36042247]

XSL•FO
RenderX

21. Mishra J, Sagar R, Parveen S, Kumaran S, Modi K, Maric V, et al. Closed-loop digital meditation for neurocognitive and behavioral development in adolescents with childhood neglect. Transl Psychiatry 2020 May 18;10(1):153 [FREE Full text] [doi: 10.1038/s41398-020-0820-z] [Medline: 32424253]
22. Mishra J, de Villers-Sidani E, Merzenich M, Gazzaley A. Adaptive training diminishes distractibility in aging across species. Neuron 2014 Dec 03;84(5):1091-1103 [FREE Full text] [doi: 10.1016/j.neuron.2014.10.034] [Medline: 25467987]
23. Rolle CE, Anguera JA, Skinner SN, Voytek B, Gazzaley A. Enhancing spatial attention and working memory in younger and older adults. J Cogn Neurosci 2017 Sep;29(9):1483-1497 [FREE Full text] [doi: 10.1162/jocn_a_01159] [Medline: 28654361]
24. Wais PE, Arioli M, Anguera-Singla R, Gazzaley A. Virtual reality video game improves high-fidelity memory in older adults. Sci Rep 2021 Jan 28;11(1):2552 [FREE Full text] [doi: 10.1038/s41598-021-82109-3] [Medline: 33510315]
25. Gunning FM, Anguera JA, Victoria LW, Areán PA. A digital intervention targeting cognitive control network dysfunction in middle age and older adults with major depression. Transl Psychiatry 2021 May 04;11(1):269 [FREE Full text] [doi: 10.1038/s41398-021-01386-8] [Medline: 33947831]
26. Pratap A, Homiar A, Waninger L, Herd C, Suver C, Volponi J, et al. Real-world behavioral dataset from two fully remote smartphone-based randomized clinical trials for depression. Sci Data 2022 Aug 27;9(1):522 [FREE Full text] [doi: 10.1038/s41597-022-01633-7] [Medline: 36030226]
27. Pratap A, Renn BN, Volponi J, Mooney SD, Gazzaley A, Arean PA, et al. Using mobile apps to assess and treat depression in Hispanic and Latino populations: fully remote randomized clinical trial. J Med Internet Res 2018 Aug 09;20(8):e10130 [FREE Full text] [doi: 10.2196/10130] [Medline: 30093372]
28. Zanto TP, Giannakopoulou A, Gallen CL, Ostrand AE, Younger JW, Anguera-Singla R, et al. Digital rhythm training improves reading fluency in children. Dev Sci 2024 May 09;27(3):e13473 [FREE Full text] [doi: 10.1111/desc.13473] [Medline: 38193394]
29. Johnson V, Hsu W, Ostrand AE, Gazzaley A, Zanto TP. Multimodal sensory integration: Diminishing returns in rhythmic synchronization. J Exp Psychol Hum Percept Perform 2020 Oct;46(10):1077-1087. [doi: 10.1037/xhp0000833] [Medline: 32730071]
30. Nandi B, Ostrand A, Johnson V, Ford TJ, Gazzaley A, Zanto TP. Musical training facilitates exogenous temporal attention via delta phase entrainment within a sensorimotor network. J Neurosci 2023 Mar 28;43(18):3365-3378. [doi: 10.1523/jneurosci.0220-22.2023]
31. Zanto TP, Padgaonkar NT, Nourishad A, Gazzaley A. A tablet-based assessment of rhythmic ability. Front Psychol 2019 Nov 1;10:2471 [FREE Full text] [doi: 10.3389/fpsyg.2019.02471] [Medline: 31736843]
32. How UTR Rating Works. UTR Sports. 2023. URL: https://www.utrsports.net/blogs/news/how-utr-works [accessed 2024-09-25]
33. Anguera JA, Boccanfuso J, Rintoul JL, Al-Hashimi O, Faraji F, Janowich J, et al. Video game training enhances cognitive control in older adults. Nature 2013 Sep 05;501(7465):97-101 [FREE Full text] [doi: 10.1038/nature12486] [Medline: 24005416]
34. Hsu W, Rowles W, Anguera JA, Zhao C, Anderson A, Alexander A, et al. Application of an adaptive, digital, game-based approach for cognitive assessment in multiple sclerosis: observational study. J Med Internet Res 2021 Jan 20;23(1):e24356 [FREE Full text] [doi: 10.2196/24356] [Medline: 33470940]
35. Wilkins L, Gray R. Effects of stroboscopic visual training on visual attention, motion perception, and catching performance. Percept Mot Skills 2015 Aug 01;121(1):57-79. [doi: 10.2466/22.25.pms.121c11x0]
36. Deveau J, Ozer DJ, Seitz AR. Improved vision and on-field performance in baseball through perceptual learning. Curr Biol 2014 Feb 17;24(4):R146-R147 [FREE Full text] [doi: 10.1016/j.cub.2014.01.004] [Medline: 24556432]
37. Clark JF, Ellis JK, Bench J, Khoury J, Graman P. High-performance vision training improves batting statistics for University of Cincinnati baseball players. PLoS One 2012 Jan 19;7(1):e29109 [FREE Full text] [doi: 10.1371/journal.pone.0029109] [Medline: 22276103]
38. Rezaee M, Ghasemi A, Momeni M. Visual and athletic skills training enhance sport performance. Eur J Exp Biol 2012;2(6):2243-2250 [FREE Full text]
39. Boot WR, Simons DJ, Stothart C, Stutts C. The pervasive problem with placebos in psychology: why active control groups are not sufficient to rule out placebo effects. Perspect Psychol Sci 2013 Jul 09;8(4):445-454. [doi: 10.1177/1745691613491271] [Medline: 26173122]

Abbreviations

ANCOVA: analysis of covariance
CONSORT: Consolidated Standards of Reporting Trials
UTR: Universal Tennis Rating
VR: virtual reality

XSL•FO
RenderX

Edited by A Mavragani; submitted 02.10.24; peer-reviewed by OR Runswick; comments to author 31.10.24; revised version received 08.01.25; accepted 16.01.25; published 07.02.25.

Please cite as:
Anguera JA, Choudhry A, Seaman M, Fedele D
Assessing the Impact of a Virtual Reality Cognitive Intervention on Tennis Performance in Junior Tennis Players: Pilot Study
JMIR Form Res 2025;9:e66979
URL: https://formative.jmir.org/2025/1/e66979
doi:10.2196/66979
PMID:39918854

©Joaquin A Anguera, Aleem Choudhry, Michael Seaman, Dominick Fedele. Originally published in JMIR Formative Research (https://formative.jmir.org), 07.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX